



**HAL**  
open science

## Personal Visualization: Exploring Data in Everyday Life

Charles Perin, Alice Thudt, Melanie Tory, Wesley Willett, Sheelagh  
Carpendale

► **To cite this version:**

Charles Perin, Alice Thudt, Melanie Tory, Wesley Willett, Sheelagh Carpendale (Dir.). Personal Visualization: Exploring Data in Everyday Life. 2015. hal-01623460

**HAL Id: hal-01623460**

**<https://hal.science/hal-01623460>**

Submitted on 25 Oct 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Workshop Proceedings

## **Personal Visualization: Exploring Data in Everyday Life**

October 2015

IEEE VIS 2015 conference, Chicago, IL, USA

<http://www.vis4me.com/personalvis15/>

### **Organizers**

**Charles Perin** | University of Calgary

**Alice Thudt** | University of Calgary

**Melanie Tory** | University of Victoria

**Wesley Willett** | University of Calgary

**Sheelagh Carpendale** | University of Calgary



## Content

<i>Charles Perin, Alice Thudt, Melanie Tory, Wesley Willett, Sheelagh Carpendale</i> <b>Personal Visualization: Exploring Data in Everyday Life</b>	4
<i>Darius Coelho, Ayush Kumar, Klaus Mueller</i> <b>Data Memes for Personal Visualization</b>	8
<i>Tamara Flemisch, Fateme Rajabiyazdi, Mona Hosseinkhani Loorak, Sheelagh Carpendale</i> <b>NeckLan: Language as Jewellery</b>	12
<i>Tsai Ling Fung, Kwan-Liu Ma</i> <b>Visual Characterization of Personal Bibliographic Data Using a Botanical Tree Design</b>	16
<i>Pascal Goffin, Wesley Willett, Petra Isenberg</i> <b>Sharing Information from Personal Digital Notes using Word-Scale Visualizations</b>	20
<i>Liang Gou</i> <b>Visualizing Personality Traits Derived from Social Media</b>	24
<i>Thor Herdal, Jeppe Gerner Pedersen, Søren Knudsen</i> <b>Designing Personal Visualizations for Different People: Lessons from a Study with Elite Soccer Teens</b>	29
<i>Yea-Seul Kim, Jessica Hullman</i> <b>User-driven Expectation Visualization: Opportunities for Personalized Feedback</b>	33
<i>Carrie Mah, Richard Pusch, Samuel Huron, Sheelagh Carpendale</i> <b>A Visualization Fingerprint: Comparing Nutrient Data Visually</b>	36
<i>Narges Mahyar, Sung-Hee Kim, Bum Chul Kwon</i> <b>Towards a Taxonomy for Evaluating User Engagement in Information Visualization</b>	40
<i>Heike Otten, Lennart Hildebrandt, Till Nagel, Marian Dörk, Boris Müller</i> <b>Are there networks in maps? An experimental visualization of personal movement data</b>	40

<i>Bruno Pagno, Luciana Nedel</i> <b>Everyday Visualization</b>	<b>47</b>
<i>Jonathan C. Roberts, Christopher Headleand, David Perkins, Panagiotis D. Ristos</i> <b>Personal Visualization for Learning</b>	<b>51</b>
<i>Sidharth Thakur, Paul Jones, Steve Cox</i> <b>Enabling Self-Awareness for Knowledge Workers Through Visualization of Instrumentation Data</b>	<b>55</b>
<i>Xin Tong, Diane Gromala, Lyn Bartram, Fateme Rajabiyazdi, Sheelagh Carpendale</i> <b>Evaluating the Effectiveness of Three Physical Activity Visualizations — How People Perform vs. Perceive</b>	<b>59</b>
<i>Pierre Vanhulst, Denis Lalanne</i> <b>Collaborative Visualizations of Self-impacted Data</b>	<b>63</b>
<i>Kodzo Webga, Aidong Lu</i> <b>Personal Movie Recommendation Visualization from Rating Streams</b>	<b>67</b>
<i>Corrie M. Whisner, Alexandra Porter, Nicholas Pecor, Ross Maciejewski</i> <b>A Survey of Personal Visualization in mHealth Nutrition Apps</b>	<b>71</b>

# Personal Visualization: Exploring Data in Everyday Life

Charles Perin  
University of Calgary

Alice Thudt  
University of Calgary

Melanie Tory  
University of Victoria

Wesley Willett  
University of Calgary

Sheelagh Carpendale  
University of Calgary

## 1 CONTACT DETAILS OF THE ORGANIZERS

Charles Perin  
Department of Computer Science  
University of Calgary  
Math Science Building (MS)  
2500 University Dr. NW  
Calgary, AB  
Canada T2N 1N4  
email: [charles.perin@ucalgary.ca](mailto:charles.perin@ucalgary.ca)  
web: <http://charles.perin.free.fr>

Alice Thudt  
Department of Computer Science  
University of Calgary  
Math Science Building (MS)  
2500 University Dr. NW  
Calgary, AB  
Canada T2N 1N4  
email: [alice.thudt@gmail.com](mailto:alice.thudt@gmail.com)  
web: <http://alicethudt.de/>

Melanie Tory  
Department of Computer Science  
University of Victoria  
Engineering/ Computer Science Building (ECS), Room 504  
PO Box 1700, STN CSC  
Victoria, BC  
Canada V8W 2Y2  
email: [mtory@uvic.ca](mailto:mtory@uvic.ca)  
web: <http://www.cs.uvic.ca/~mtory>

Wesley Willett  
Department of Computer Science  
University of Calgary  
Math Science Building (MS)  
2500 University Dr. NW  
Calgary, AB  
Canada T2N 1N4  
email: [wesley.willett@ucalgary.ca](mailto:wesley.willett@ucalgary.ca)  
web: <http://www.wjwillett.net/>

Sheelagh Carpendale  
Department of Computer Science  
University of Calgary  
Math Science Building (MS)  
2500 University Dr. NW  
Calgary, AB  
Canada T2N 1N4  
email: [sheelagh@cpsc.ucalgary.ca](mailto:sheelagh@cpsc.ucalgary.ca)  
web: [www.cpsc.ucalgary.ca/~sheelagh](http://www.cpsc.ucalgary.ca/~sheelagh)

## 2 ORGANIZER'S BACKGROUND

**Charles Perin** is currently a postdoc at the University of Calgary. His doctoral research, at the intersection of Human-Computer Interaction and Information Visualization, focused on exploring new interaction techniques for Information Visualization. He mostly published his work on interaction techniques at CHI [13, 15, 19] and Infovis [12, 14]. He is interested in i) designing new interaction techniques; and ii) information visualization application areas related to personal data visualization for everyone (e.g., sports monitoring, health improvement, individual decision-making).

**Alice Thudt** is currently a PhD student in Computational Media Design. She is pursuing her research at the InnoVis Group at the University of Calgary, Canada. Her research focuses on using information visualization to integrate people's personal data collections better into their everyday lives. Recent projects have explored visualizations for autobiographical storytelling [18] and the design of visualization techniques to reflect autobiographical memories [17]. She is further interested in democratizing the creation of visualization to a wider audience of non-experts [7].

**Melanie Tory** is an Associate Professor of Computer Science at the University of Victoria. Her visualization research applies to a wide variety of areas spanning natural sciences and engineering, business, and personal informatics. Recent projects have focused on visualizing personal home energy and fitness data and supporting collaboration around visualizations. Relevant papers include a characterization of personal visualization and personal visual analytics [5] an article on what motivates people to use visualizations in non-work contexts [16] and work on integrating personal data within digital calendars [6].

**Wesley Willett** is an Assistant Professor of Computer Science at the University of Calgary. His interests span information visualization, social computing, new media, and human computer interaction, and his research focuses on pairing data and interactivity to support collaboration, learning, and discovery. In particular, his work explores the intersection of visualization and computer-supported collaborative work and shows how communities, crowds, and novice users can work together build knowledge from data. Dr. Willett's recent work has explored how communities can collaborate to collect and analyze data [21, 25], how large groups of crowd workers can help process data sets [22, 24], and how visualization can help people better understand personal data like digital notes [23]. This work has earned a number of awards, including the Best Paper award at Pervasive 2010 and a Best Paper Honorable Mention at CHI 2011.

**Sheelagh Carpendale** is a Canada Research Chair in Information Visualization at the University of Calgary. Her research draws upon her combined backgrounds in computer science and visual arts, benefiting from the interdisciplinary cross-fertilization to enable the design of innovative, people-centred information technologies. By studying how people interact with information both in work and social practices, she works towards designing more natural, accessible and understandable interactive visual representations of data [8, 11, 20]. She combines information visualization, visual analytics and human-computer interaction with innovative new interaction techniques to better support the everyday practices of people who are viewing, representing, and interacting with information [2].

### 3 WORKSHOP OUTLINE

In this workshop we will discuss the emerging area of personal visualization. In this section we describe the proposed focus and motivation, the technical scope, and the main goals of the workshop.

#### 3.1 Proposed Focus and Motivation

Individuals recently began to seek how they can explore and understand the data that affect their personal lives. This includes biometric personal data such as health-related data [11], self-monitoring, sports and performance understanding and improving; and personal data such as online social networks, energy consumption, web activity, and photo collections [17]. These personal data often involve large quantities of data. Popular devices for sports monitoring record measures such as heart-rate, speed, and pace at high frequency, resulting in massive amount of data that is difficult to make sense of. As another example, online social networks can consist of thousands of posts every day from hundreds of acquaintances/friends with various degrees of intimacy. These data are relevant to our personal lives, enabling us, as individuals, to explore and make sense of important, personally relevant information [10].

Such explorations of personal data are usually carried out on the web by non-visualization experts. Commercial products propose simple designs to import collected data, usually specific to a device and featuring simple charts such as bar charts and line graphs (e.g., fitbit<sup>1</sup> and garmin connect<sup>2</sup>). However, working with large amounts of data of particular types (e.g., multi-dimensional, temporal, cyclic, incomplete, and sometimes fuzzy) is a real challenge for individuals. While the area of quantified-self recently emerged and HCI communities have followed the trend (e.g., [1, 3, 9]), exploring ways of providing individuals with new devices and interfaces for collecting personal data, few visualization researchers have started to investigate personal visualization as an area of research.

Information visualization application areas and visual analytics tools tend to focus on data important to science, finance, business and government. Such tools are designed for experts (e.g., scientists, traders, and professionals trying to make sense of their data). Personal life related data brings a new context to both visualization and visual analytics, and with that new context come new research and design challenges [2]. The visualization community has started to consider the emerging and important case of personal visualization, i.e. individual's challenges for collecting, visualizing, and making sense of personal data. Personal visualization is not *just* about visualizing personal data, but also about visualizing data of personal interest (e.g., global warming, sport, neighborhood) and about creating a personal visualization representing non-personal data (e.g., personal artistic representations of country indicators).

Understanding individual's needs in the context of their personal data and designing appropriate tools to support visualization and analysis of this data is a crucial and emergent challenge. This should consider factors such as data collection challenges, casual visualization, attractive and enjoyable interfaces, sensemaking and storytelling strategies [18], statistical reasoning for non statistical experts, and engagement. In particular, such designs should carefully consider pleasure, gamification, and engagement with the data instead of efficiency to perform low-level functional tasks quickly and accurately. The main purpose of such data understanding is generating insights, and eventually making decisions to improve one's life, the ultimate purpose of visualization.

To summarize, personal visualization is still in its infancy in the research area of visualization while individuals gather more and more personal data and need to make sense of it. For these reasons, the workshop format during the VIS conference offers the ideal set-

ting to gather the community and discuss open and important questions such as:

- How to connect to research areas such as personal informatics [1, 9] that focus on collecting personal data?
- How to provide individuals with ways of visualizing and making sense of large amounts of multi-dimensional, temporal, multi-variate data?
- How to apply the knowledge from the community to this new research topic? (e.g., uncertainty visualization, missing data, and storytelling).
- How to tackle the important challenge of sharing and privacy, crucial for personal data?
- Which application areas involve personal data? (e.g., health, well-being, environmental concerns, sport monitoring, photo collections, social networks).
- What are the differences between designing visualization for domain experts and for non-expert individuals?
- How does personal visualization relate to art? What are the particularities of e.g., data portraits [4, 26] and data sculptures [27]?
- How to assess the challenges of discoverability, gamification, engagement?
- How to evaluate tools for personal visualization? How to assess the impact on one's life?

#### 3.2 Goals and Intended Results

The main objective of this workshop is to bring together people from academia and industry to discuss future perspectives in the area of personal visualization. This workshop has two main goals: 1) gather the community working on the topic of personal visualization, and 2) converge on a research agenda for the community.

**Gather the community.** Personal visualization exist on the web, on blogs, on personal websites. We hope authors of such visualizations "in the wild", both from academia and industry will attend the workshop. Other research areas such as personal informatics, HCI, design, and art contribute to personal visualization. Specialized industries also provide platforms on the web to visualize data collected by their wearable devices, usually using simple visualizations such as pie charts, bar charts, and line graphs. We intend to bring all these communities together and demonstrate that the visualization community plays a crucial role in personal data understanding and sensemaking and has the required expertise (e.g., storytelling, literacy, infographics, uncertainty) to make well suited and dedicated visualization tools, beyond standard charts.

**Converge on a research agenda.** The workshop is designed in such a way that we will identify the main challenges and opportunities for the visualization community. We will need to establish definitions, identify open challenges, and ultimately converge on a research agenda for the community. Such challenges include challenges about data (e.g., collection, storage, structure) and are related to research areas such as personal informatics and quantified-self; challenges about engagement, gamification, enjoyment, and aesthetics; and challenges in terms of visualization, with personal visualization having different purposes than scientific/industry visualization for experts. It raises numerous problems such as sharing and privacy, evaluating such visualizations for which time and error measurements do not always make sense, and designing appealing and easy to discover/learn tools for the general public.

We intend to use the workshop website we will create if the paper is accepted (see section 4.1) as a platform for researchers and professionals to exchange ideas, and to showcase the results from the workshop. We will also build a mailing list to keep people in touch. Finally, we will ask the authors if they are interested in organizing/contributing to a special issue of a journal to publish their work on the topic. A second objective is to gather interested people

<sup>1</sup><http://www.fitbit.com>

<sup>2</sup><https://connect.garmin.com>

in writing a group article summarizing the research agenda based on the challenges and perspectives identified during the workshop.

#### 4 PARTICIPANTS AND CALL FOR PARTICIPATION

This workshop proposal is for a half-day workshop. As VIS workshops are open to all attendees, the final size of the workshop is difficult to predict. We expect a final size of about 40–80 attendees with 15–20 accepted position papers / posters / demos. In order to be invited to speak at the workshop, participants will have to first send one of the following contributions:

1. **Position papers** will be 4–6 pages length papers about a challenge for personal visualization. We will encourage, but not restrict, authors to contribute to one of the identified topics of personal visualization. Position papers should clearly identify one challenge and discuss implications for the visualization community.
2. **Posters/Demos** will be 2 pages length papers accompanied with either a A0 poster and/or an interactive demonstration. Posters and demos should present author’s work on the topic of personal visualization. We expect authors to submit either i) tools they built for themselves, for relatives, or for communities they are involved in (mainly for demonstrations); or ii) design ideas and early sketches (mainly for posters).

Workshop presenters will be selected to represent the different aspects of personal visualization that we enumerated above. We will favor diversity in order to cover the topic widely. We will select participants according to a number of criteria, which include: originality of the submission, complementary nature of research background, and quality of previous publications related to the workshop topic.

##### 4.1 Participants and Timeline

To achieve a diversity of participants, we will actively advertise the workshop to people with backgrounds in information visualization, visual analytics, HCI, interface design, cognitive psychology, and art. In addition to publishing it on relevant web pages, this includes e-mail to related mailing lists as well as personal e-mails to people known by the organizers to be working in the field or interested in the area. We will contact industries in related areas (e.g., sport monitoring devices) that demonstrated us their interest in visualization.

If the workshop is accepted, we will set up a website containing the call for participation, the timeline, and the schedule. We will also create a twitter account. We will advertise as soon as the workshop proposal is accepted. We expect to include researchers from the community with a strong interest in personal visualization to help review submissions.

The submissions will be made publicly available on the workshop website as soon as possible. We will encourage both the authors and the attendees to have a look at these submissions before the workshop.

Additionally, we will contact several persons from outside the research community we are connected with, whose work on personal visualization is recognized. We will i) ask them if they are interested in attending the workshop, and ii) ask them the permission to showcase their work, in the form of printed posters. This exhibit-like showcase will add value to the workshop and we hope it will demonstrate the importance of this new research area from outside the community and give an external eye on the topic.

This workshop highly encourages the active participation of attendees, with the first part of the workshop consisting of presentations, the second part consisting of poster viewing, interactive demonstrations, and a showcase during an extended coffee break,

and the last part consisting of group discussions about identified discussion topics involving the audience.

The tentative schedule for the workshop is:

1. **Introduction** (15–20 min) from the workshop organizers.
2. **Part I** (1h) Short presentations from the authors (exact time to be determined according to the number of submissions). Authors of position papers do a 4–8 minutes presentation and authors of demos/posters do a 2-4 minutes presentation, as they will showcase their work during the break. Initial presentations are short, similar to a fast forward session. This fast-paced introduction allows attendees to get a quick overview of the richness of the ideas but leaves the remainder of the available time to discussions and activities. Speakers will be asked to raise 1–2 interesting discussion topics at the end of their presentation. We will collect these discussion topics for **Part III**.
3. **Part II** (40min–1h) Extended coffee break during which posters/demos authors showcase their work, and during which attendees will also have time to look at the exhibition.
4. **Part III** (1h-1h30) Group discussions involving both the audience and the authors. The workshop organizers will select the best / most recurrent discussion topics from the presentations in **Part I**, and organize discussion groups based on these topics. Each group will brainstorm and then give a short (2–3 minutes) summary of their discussion.
5. **Wrap-up** (15–30 min) conclusion and discussion of the outcomes of the workshop.

##### 4.2 Requested Facilities

For this workshop we will need a room with a projection surface in the front for the initial fast forward presentations. We will need a desk in the front for speaker presentations, group presentations, and organizer’s introduction / conclusion. Because we intend to break the audience into discussion groups during **Part III**, we would need a room with several round or square tables arranged in a way that is suitable for small group discussion. The tables and chairs should be movable.

We will also need space and several desks (number to be determined) for participants to set up their demonstration on the side of the room. We will not provide additional screens or projectors to the demonstration authors. Similarly, we will need space and poster boards (number to be determined) to hang out posters during the workshop.

Finally, and optionally, it may be interesting for all attendees of the conference to be able to showcase the posters and the exhibit permanently during the conference. If there is an empty space somewhere in the building, it would be an opportunity for the authors to showcase their work on the topic.

#### REFERENCES

- [1] J. R. Brubaker and G. R. Hayes. Lost in translation: Three challenges for the collection and use of data in personal informatics. *Personal Informatics & HCI: Design, Theory, & Social Implications* (in Conjunction with ACM CHI 2011), 2011.
- [2] S. Carpendale and M. Tory. Guest editor’s introduction: Personal visualization and personal visual analytics. *IEEE Computer Graphics and Applications: Special Issue on Personal Visualization and Personal Visual Analytics*, ??(?):??, 2015.
- [3] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz. Understanding quantified-selfers’ practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’14, pages 1143–1152, New York, NY, USA, 2014. ACM.

- [4] J. Donath, A. Dragulescu, A. Zinman, F. Viégas, and R. Xiong. Data portraits. In *ACM SIGGRAPH 2010 Art Gallery*, SIGGRAPH '10, pages 375–383, New York, NY, USA, 2010. ACM.
- [5] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 21(3):420–433, March 2015.
- [6] D. Huang, M. Tory, and L. Bartram. Data in everyday life: Visualizing time-varying data on a calendar. *Poster at IEEE VIS 2014*, 2014.
- [7] S. Huron, S. Carpendale, A. Thudt, A. Tang, and M. Mauerer. Constructive visualization. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, pages 433–442, New York, NY, USA, 2014. ACM.
- [8] S. Huron, Y. Jansen, and S. Carpendale. Constructing visual representations: Investigating the use of tangible tokens. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2102–2111, Dec 2014.
- [9] I. Li, A. Dey, and J. Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 557–566, New York, NY, USA, 2010. ACM.
- [10] I. Li, A. K. Dey, and J. Forlizzi. Understanding my data, myself: Supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 405–414, New York, NY, USA, 2011. ACM.
- [11] H. MacLeod, A. Tang, and S. Carpendale. Personal informatics in chronic illness management. In *GI '13: Proceedings of the 2013 Graphics Interface Conference*, pages 149–156, 2013.
- [12] C. Perin, P. Dragicevic, and J.-D. Fekete. Revisiting bertin matrices: New interactions for crafting tabular visualizations. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2082–2091, Dec 2014.
- [13] C. Perin, F. Vernier, and J.-D. Fekete. Interactive horizon graphs: Improving the compact visualization of multiple time series. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '13, pages 3217–3226. ACM, 2013.
- [14] C. Perin, R. Vuillemot, and J.-D. Fekete. Soccerstories: A kick-off for visual soccer analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2506–2515, 2013.
- [15] C. Perin, R. Vuillemot, and J.-D. Fekete. A table!: Improving temporal navigation in soccer ranking tables. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '14, pages 887–896. ACM, 2014.
- [16] D. Sprague and M. Tory. Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations. *Information Visualization*, 11(2):106–123, Apr. 2012.
- [17] A. Thudt, D. Baur, and S. Carpendale. Visits: A spatiotemporal visualization of location histories. Earned EuroVis 2013 Best Short Paper Award, 2013.
- [18] A. Thudt, D. Baur, and S. Carpendale. Autobiographical visualizations: Challenges in personal storytelling. In *Proceedings of the DIS'14 Workshop on A Personal Perspective on Visualization and Visual Analytic*, 2014.
- [19] R. Vuillemot and C. Perin. Investigating the direct manipulation of ranking tables for time navigation. In *Proceedings of the Conference on Human Factors in Computing Systems*, CHI '15, pages 2703–2706. ACM, 2015.
- [20] J. Walny, S. Huron, and S. Carpendale. An exploratory study of data sketching for visual representation. *Computer Graphics Forum*, To appear, 2015.
- [21] W. Willett, P. Aoki, N. Kumar, S. Subramanian, and A. Woodruff. Common sense community: Scaffolding mobile sensing and analysis for novice users. In *Proceedings of the 8th International Conference on Pervasive Computing*, Pervasive'10, pages 301–318, Berlin, Heidelberg, 2010. Springer-Verlag.
- [22] W. Willett, S. Ginosar, A. Steinitz, B. Hartmann, and M. Agrawala. Identifying redundancy and exposing provenance in crowdsourced data analysis. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2198–2206, Dec. 2013.
- [23] W. Willett, P. Goffin, and P. Isenberg. Understanding digital note-taking practice for visualization. *IEEE Computer Graphics and Applications: Special Issue on Personal Visualization and Personal Visual Analytics*, ??(?)?:??, 2015.
- [24] W. Willett, J. Heer, and M. Agrawala. Strategies for crowdsourcing social data analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 227–236, New York, NY, USA, 2012. ACM.
- [25] W. Willett, J. Heer, J. Hellerstein, and M. Agrawala. CommentSpace: Structured support for collaborative visual analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3131–3140, New York, NY, USA, 2011. ACM.
- [26] R. Xiong and J. Donath. Peoplegarden: Creating data portraits for users. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, UIST '99, pages 37–44, New York, NY, USA, 1999. ACM.
- [27] J. Zhao and A. V. Moere. Embodiment in data sculpture: A model of the physical visualization of information. In *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts*, DIMEA '08, pages 343–350, New York, NY, USA, 2008. ACM.

# Data Memes for Personal Visualization

Darius Coelho, Ayush Kumar, and Klaus Mueller

Visual Analytics and Imaging Lab, Computer Science Department, Stony Brook University, NY, USA and SUNY Korea, Songdo, Korea



Figure 1. Personal visualizations (a) shows Joe's speed over multiple heats that he ran (b) shows Mike's workout routine for the week

## ABSTRACT

We introduce the concept of Data Memes as artistic visuals of data in which users can merge data visualizations with an image such that the structure of the image supports the user's intended meaning (or interpretation) of the data. Since Data Memes can represent very personal views, it is natural to employ them in the visualization of personal data. Hence, in this paper we focus on the use of Data Memes for this purpose and also explain how they can be constructed with relative ease. Data Memes as a general concept are meant to engage users with the visualization and persuade a viewer to look at the data. Communicating personal achievements is often the wish of posters in social networks, and although we have not tested this yet, we believe that Data Memes augmented with pertinent personal data have good potential to achieve higher levels of attention in these circles, and elsewhere.

**Keywords:** Chart junk, infographics, personal visualization.

## 1 INTRODUCTION

The information age is upon us and in recent years millions of terabytes of data have been generated every day. A growing chunk of this data is related to our personal lives – information about ourselves, our communities, and issues that have a personal relevance [9]. With the access to such data, people want to make sense of it, and to do so they turn to visualization. While there are a multitude of solutions available to visualize these data, they are rarely used in a personal context.

Recently researchers have been exploring the area of *Personal Visualization* and *Personal Visual Analytics*. Personal visualization caters to the wide variety of requirements of people

who are not necessarily visualization experts or analysts. These people bring personal interpretations to their data and wish to design visualizations that can represent these personal views [20]. In our paper we describe a design framework that allows users to express their opinion about the data by framing them as charts into a self-selected and self-designed visual context. This has good potential for fostering their own engagement with the data as well as that of others with whom these design products are shared.

Fig. 1 shows some examples of the personal visualizations of personal data our system can assist users in creating. Here we note that our system can also be used for other types of data, but the opportunity for users to frame these data into a personalized visual context always remains and is a hallmark of our system [5].

We have coined a new term for these types of visualizations – *Data Meme*. Wikipedia defines the notion of Meme "an idea, behavior, or style that spreads from person to person within a culture". A more recent phenomenon is that of Internet Meme. These are images augmented with text in which the creator's choice of image puts forward his/her idea about the message the text conveys. They are often reused with the same core message, and as a result gain viral dissemination. Similarly, our Data Memes are images augmented with related data in which the designer's choice of image puts forward his/her idea about the message the data convey.

Our Data Memes appeal to the masses by embracing mainly basic charts, specifically pie, line, and bar charts. Following the definition put forward by Pousman et al. [15], they fall into the set casual information visualization techniques catering to a wide variety of users. The chosen and further enhanced image that becomes part of the visualization is itself a powerful element of the visualization as it can encode a multitude of information. The image is linked to the topic of the data being displayed, with the choice of image putting forth the designer's point of view about the data. At the same time it is also capable of enhancing memorability and engaging viewers. We foresee that users may use the Data Memes to post their own data on social media or blogs in order to (1) engage viewers to consider the information or (2) put across their own point of view about the data.

Emails: {dcoelho, ayush.kumar, mueller}@sunykorea.ac.kr

Electronic proceedings of the IEEE VIS 2015 workshop  
*Personal Visualization: Exploring Data in Everyday Life*  
The authors remain the holders of the copyright



In the following, Section 2 reports on related work. Section 3 describes our system by ways of several use examples. Section 4 presents a few examples, and Section 5 ends with conclusions.

## 2 RELATED WORK

### 2.1 Visual Embellishments in Charts

Our Data Memes are significantly related to *chart junk* which are visual embellishments of charts. The notion of chart junk has received growing attention in the field of information visualization. Visualization guru Edward Tufte strongly discourages the use of visual embellishments in charts. He opines that visualizations that contain chart junk are distracting to users and make the chart difficult to read [18].

On the other hand, famous designer Nigel Holmes [13] believes that chart junk will pique a viewer's interest in the chart. The study by Bateman et al. [1] concluded that while chart junk and plain graphs were equally comprehended by users, chart junk was more memorable. Stephen Few [7] responded that the study was flawed in some ways, but then ended by stating that properly designed chart junk can be useful provided it does not lead to distractions or misinterpretations. Borkin et al. [3] investigated which aspects of a visualization make it memorable. Other authors [2][11] have made similar basic observations. Harrison et al. [8] studied the first impression that infographics have on people, concluding that colorful infographics with not much text and image areas appeal to most people. Viegas et al. [19] observed that chart junk visualizations created by artists, while memorable, can change the consumer's attitude or change the way people think.

Chart junk so far has been mainly created by artists like Peter Ørntoft [14]. He created simple but beautiful infographics by dressing a model with the chart or designing it physically and then photographing it. Nuage Vert by Evans et al. [6] highlights segments of smoke in an image to show power consumption.

All of these artist-generated chart junk creations are commonly very attractive and inspirational, but it remains difficult for main stream users to create good chart junk with existing digital tools [23]. Also the usability of a visualization is dependent on the aesthetics [4] and the style used in the design

[13]. This inspired us to create a novel design tool we call *MemViz* which allows non-artists to design charts with visual embellishments. MemViz morphs charts such that they fit well in areas of an image. For the work that is subject of this paper we use an enhanced version of our tool, appropriately called *MemeViz* that enables users to create personal visualizations of personal data.

### 2.2 Personal Visualization

Researchers have been investigating the use of visualizations for personal and casual use. Some of the work explores how both experts and non-experts can make personal data engaging and gain insights into personal data. Leung et al. [10] projected a visualization of a user's social network data and found that it triggered social interactions when viewed by others and acted as a mode of self expression. Xiong et al. [22] also visualize social network data with a flower metaphor in a manner that is easy for non-experts to comprehend. Wood [21] tries to express his personal experiences in the visualization of his cycling data. Rodgers et al. [16] observe that users were able to easily understand their energy consumption data through abstract visualizations with less detail as compared to traditional techniques. Wang et al. [20] studied the impact of design for personal visualization and observed that the abstract techniques were better at engaging the user but at the cost of determining concrete data.

## 3 DESIGNING THE DATA MEMES

Our experiments have shown that even novice users can employ MemeViz to create eye-catching visualizations that embed basic charts into an image. All can be achieved by ways of a set of dedicated algorithms and interactions which can be executed via simple mouse clicks. The user interface is shown in Fig 2. A user starts off by providing the data file and selecting the type of chart (pie, bar, line) he or she wishes to plot. Next, the tool suggests a set of relevant thematic image backdrops retrieved by a web-scale text-based image search engine (e.g., Google Images, Flickr). To form the search query our tool uses the title of the dataset and the attribute labels since these are intuitive descriptors of the data. Alternatively, the user can also call in a picture from his or her

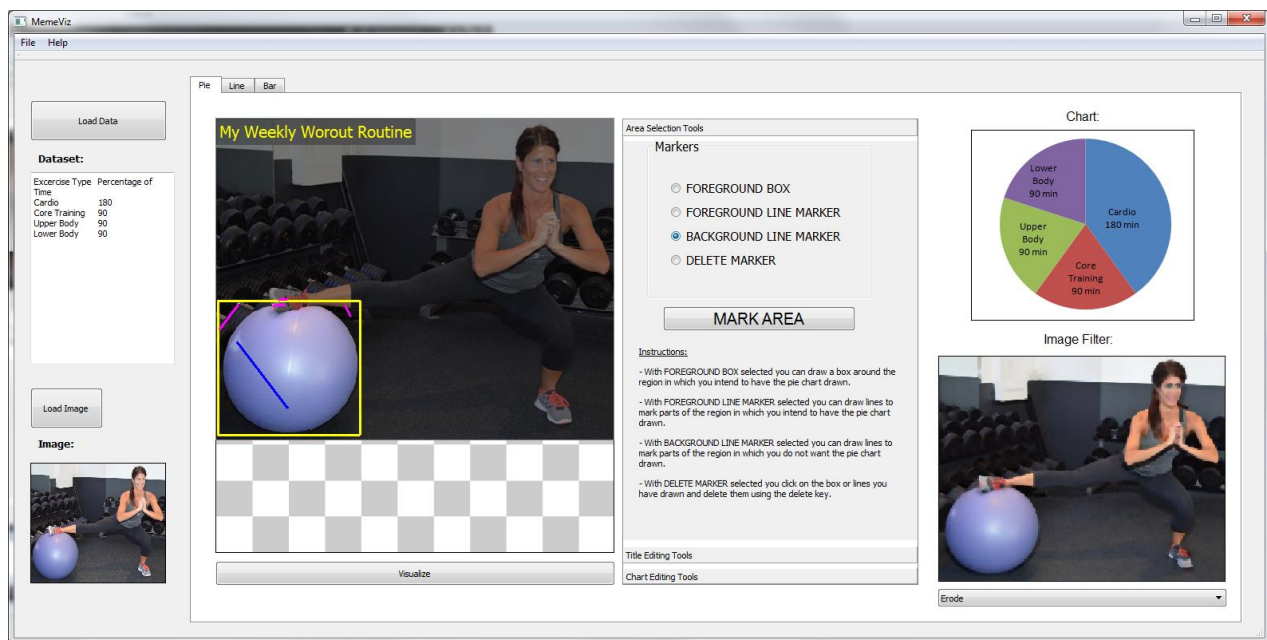


Figure 2. The MemeViz user interface which was used to create all of the Data Memes shown in this paper



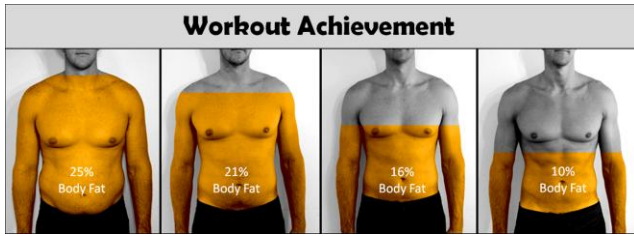


Figure 3. Marcus' body fat levels at different stages of his life.

private photo collection. The latter would likely be a more popular option for the purpose of personal visualization where users wish to post their own data to social media or blog, utilizing a personal image as a carrier.

Once the user has selected an image from the provided set, he or she is given access to an array of processes (described in the following subsections) to fuse the chart with the image. Then, after the chart has been embedded, the user can customize the visualization further by applying a diverse set of visual effects to the backdrop (filters, transparency etc.) or to the chart (color selection, glow, transparency). This customization gives the visualization an illustrative look and feel, and it also reduces image detail that might interfere with the detail of the chart. This finalizes the result.

### 3.1 Pie Charts

Pie charts and their spatial generalizations use a division of unity paradigm to show numerical proportions. To embed pie charts into images we decided to have users select a region (in most cases an object) in the image after which the tool would divide the region proportionately according to the data.

With this tool, after selecting the backdrop image, the user interactively selects a region into which to plot the chart by marking a bounding box around the region. The tool segments the region using the Grabcut algorithm [17]. The segmentation can be refined by marking highlights on the background and the foreground. The next step involves laying out the chart into the segmented region. Here, users have various options to divide and layout the chart. For pie-style charts they can divide the region radially in two ways: (1) angle-based, i.e. dividing  $360^\circ$  radially proportional to the data, and (2) area-based, i.e. dividing the area of the region radially proportional to the data. In the latter case we use a pixel count to determine the area. For fill-style charts we allow the user to specify an axis across the segmented region and then fill the region along this axis. Again, we divide the region in two ways: (1) based on the length of the axis, and (2) based on the area of each division within the segment for which we use a pixel count as well. Once the region is divided we color the divisions of the region in CIE LAB color space. This retains the texture of the region and only changes the color.

### 3.2 Line Charts

The line chart is one of the most common visualizations to plot a series of data points. To merge line charts with images, we match the line to structures in the image that have a similar trend to the line. In this case instead of looking for images with closed regions as in the fill chart task, we are now looking for images with features and edges that mimic the trend of the line chart as close as possible.

As the first step the user is asked to place the line chart over a desired region in the backdrop image. We experimented with a number of methods to perform line matching so as to place the

chart automatically in the image, but a solution remains elusive. Instead we now rely on the user to find the features visually by looking at the chart and the image side by side. Although this is not ideal we observed that users were capable of performing this task well. We have provided an interactive interface that allows users to move and scale line charts and so fit them into appropriate positions in the image. However, we specifically prevented the user from rotating the chart as it may lead to line chart trends looking more dramatic.

### 3.3 Bar Charts

Bar charts are mainly used to show comparison of values. They are also a very popular visualization paradigm. In our application, the bar charts can be merged with images by either overlaying the chart over the image or by treating objects in the image as bars. In this case we make use of techniques used to embed pie charts and overlay line charts.

Once the backdrop has been selected, the user is asked to decide if he/she wants to place the bar chart over the image or fill in objects. In the former case the system applies the same technique that was also used for the line charts where the user can place the chart using the mouse. On the other hand, if the user wishes to fill objects then we follow the segmentation technique used for the pie charts, but in this case for every bar the user must mark a region in the image.

## 4 EXAMPLES

We shall now give a few examples on what types of personal visualizations might be possible using our tool. All images were obtained using Google image search using the respective activity terms in the search string.

Fig. 1a shows a Data Meme featuring Jim, a short distance runner, who visualizes his progress in sprinting speed via a line plot embedded in a picture that shows him just after setting off. The picture was chosen since the posture at that time aligns well with the line plot. This creates a good synergy between image content and the data which is a prerequisite for a good Data Meme.

Fig. 1b shows Mike during an exercise with his workout routine allocation mapped as a pie chart onto a medicine ball. On the other hand, Fig. 3 shows Marcus at different stages of his life with the corresponding level of body fat mapped like a bar chart. The sequence of pictures was acquired during a long term "Get Lean in 180 Days" program that Marcus underwent. He uses the Data Meme to proudly showcase his amazing progress.

David, who plays an online game League of Legends, created the Data Meme of Fig. 4a with his game stats and an image of a character he plays in the game. In Fig 4b, c, d we see the Data Meme's of Maria, Betty and Anna, created from their fitness data. In Fig. 4e, we see Patrick's Data Meme visualizing the Netflix data of his most watched TV shows. Finally, Fig 4f shows Ryan's Data Meme that visualizes his dataset about gym accidents.

## 5 CONCLUSION AND FUTURE WORK

We believe that the examples we have presented do a good job in explaining the potential of Data Memes as a means to contextualize the data with an image and so provide for a more engaging viewing experience. In the case of a personal visualization this image context will often involve the user as an actor, or a well-chosen substitute or avatar. Most of the Data Memes we have created here were meant to increase viewer engagement for data usually posted just as plots or numbers on social media or blogs. In some cases, however, they can also be used to put forth a personal point of view about the data.



Figure 4. Personal visualizations created by our designers: (a) shows David's League of Legends stats (b) shows the amount of time Maria ran everyday for a week and (c) shows Betty's bike tension levels during her spinning routine (c) shows Anna's workout routine (e) shows the top 3 most watched television shows Patrick watches on Netflix and (f) show a stat about gym accidents created by Ryan.

Future work will run a user study where subjects would create Data Memes with real and even personal datasets using pictures of themselves. We will also test how engaging these visualizations are and if the user's point of view is put across in the Data Meme.

## 6 ACKNOWLEDGEMENT

This research was partially supported by NSF grants IIS 1117132 and 1527200 and the MSIP, Korea, under the "ICT Consilience Creative Program" (IITP-2015-R0346-15-1007).

## REFERENCES

- [1] Bateman, S., Mandryk, R., Gutwin, C., Genest, A., McDine, D., Brooks, C. "Useful junk? the effects of visual embellishment on comprehension and memorability of charts." *CHI 2010*, pp. 2573-82.
- [2] Borgo, R., Abdul-Rahman, A., Mohamed, F., Grant, P. W., Reppa, I., Floridi, L., Chen, M. "An empirical study on using visual embellishments in visualization." *IEEE TVCG*, 18(12), 2012, pp. 2759-2768.
- [3] Borkin, M., Vo, A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. "What makes a visualization memorable?." *IEEE TVCG*, 19(12), 2013, pp. 2306-2315.
- [4] Cawthon, N., & Moore, A. "The effect of aesthetic on the usability of data visualization." *IV'07 Int'l Conf. IEEE*, 2007, pp. 637-648.
- [5] D. Coelho, S. Ha, S. Cheng, S. Mahmood, J. Kim, K. Mueller, "MemViz: A Tool for Creating Memorable Visualizations," *IEEE VIS Poster Abstracts*, Paris, France, November 2014
- [6] Evans, H., Hansen, H., & Hagedorn, J. "Artful media: Nuage vert." *MultiMedia*, IEEE, 16(3), 2009, pp. 13-15.
- [7] S. Few. "The chartjunk debate. a close examination of recent findings", *Perceptual Edge Visual Business Intelligence Newsletter*, April, May, and June 2011
- [8] Harrison, L., Reinecke, K., & Chang, R. "Infographic Aesthetics: Designing for the First Impression." *CHI*, 2015, pp. 1187-1190.
- [9] Huang, D., Tory, M., Aseniero, B., Bartram, L., Bateman, S., Carpendale, S., Tang, A., Woodbury, R. "Personal visualization and personal visual analytics." *IEEE TVCG*, 21(3), 2015, pp. 420-433.

- [10] Leung, M., Tomitsch, M., & Vande Moere, A. "Designing a personal visualization projection of online social identity." *CHI'11 Extended Abstracts*, 2011, pp. 1843-1848.
- [11] Li, H., & Moacdieh, N.. "Is "chart junk" useful? An extended examination of visual embellishment." *Proc. Human Factors and Ergonomics Society Annual Meeting*, Vol. 58, No. 1, (2014).
- [12] Moore, A. V., Tomitsch, M., Wimmer, C., Christoph, B., & Grechenig, T. "Evaluating the effect of style in information visualization." *IEEE TVCG*, 18(12), 2012, pp. 2739-2748
- [13] Nigel Holmes, <http://nigelholmes.com/>
- [14] Peter Ørntoft, <http://www.peterorntoft.com/>
- [15] Pousman, Z., Stasko, J. T., & Mateas, M. "Casual information visualization: Depictions of data in everyday life." *IEEE TVCG*, 13(6) (2007): 1145-1152.
- [16] J. Rodgers and L. Bartram, "Exploring Ambient and Artistic Visualization for Residential Energy Use Feedback," *IEEE TVCG*, 17(12), 2011, pp. 2489-2497.
- [17] Rother, C., Kolmogorov, V., & Blake, A. "Grabcut: Interactive foreground extraction using iterated graph cuts." *ACM Transactions on Graphics* 23(3), 2004, pp. 309-314.
- [18] E. Tufte. *The Visual Display of Quantitative Information*. Cheshire (Conn.), 2001.
- [19] Viégas, F., Wattenberg, M. "Artistic data visualization: Beyond visual analytics." *Online Communities and Social Computing*, 2007, pp. 182-191.
- [20] Wang, S., Tanahashi, Y., Leaf, N., & Ma, K. L. "Design and Effects of Personal Visualizations." *CG&A*, IEEE 35(4), 2015, pp. 82-93.
- [21] Wood, J. "Visualizing Personal Progress in Participatory Sports Cycling Events." *CG&A*, IEEE, 35(4), 2015, pp.73-81.
- [22] R. Xiong and J. Donath, "PeopleGarden: Creating Data Portraits for Users," *Proc. 12th Ann. ACM Symp. User Interface Software and Technology*, 1999, pp. 37-44.
- [23] Su, Y. S. "It's easy to produce chartjunk using Microsoft® Excel 2007 but hard to make good graphs." *Computational Statistics & Data Analysis* 52(10), 2008, pp. 4594-4601.

# NeckLan: Language as Jewellery

Tamara Flemisch\*  
Technische Universität Dresden

Fateme Rajabiyazdi†  
University of Calgary

Mona Hosseinkhani Loorak‡  
University of Calgary

Sheelagh Carpendale§  
University of Calgary

## ABSTRACT

We present NeckLan, a physical visualization in form of a decorative pendant for a necklace. The encoded data consists of the languages a person can speak, their corresponding language families and the person’s proficiency for each language. This project is motivated by internationality and globalization of research laboratories and conferences. In an iterative design process we drew sketches and designed NeckLan. Furthermore, we explore the possible usages of NeckLan for different purposes.

**Index Terms:** H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles

## 1 INTRODUCTION

We designed NeckLan, a decorative pendant for a necklace or a keychain that portrays several facets of personal data. Examples of such personal data include the number, the proficiency and the type of languages spoken by a person.

People from different countries who attend international events, or are new in multi-cultural groups like research labs or companies, need a way to be engaged in conversation with others. Knowing a few personal details about others plays a key role in getting involved in conversations with them.

Information visualization can provide an engaging and aesthetically appealing way of communicating data. Thus, representing some personal data in a simple, aesthetic, and easily accessible way has the potential to attract other people’s attention for the beginning of a new conversation. Furthermore, this allows people to know what characteristics they share with others. This could potentially give them a possible topic to talk about and it can be used as an icebreaker. In particular, knowing which languages a person speaks might provide a common ground.

Within the last few years, physical visualizations have become more popular due to the increased technical possibilities, such as 3D printing. Using real material for portraying data can increase wearability and personality of a visualization. For example Meshu [3] is a website with which people can create their own personal physical visualization based on self-chosen locations. The result is a piece jewellery that looks like a mesh. After selecting the places, people can order their personal meshes as earrings, necklace or ring and it will be 3D printed or laser cut depending on the material.

We think that having a little piece of jewellery that visualizes some personal information about people can be helpful in this regard. Thus, we designed NeckLan as a physical decorative visualization revealing the languages a person can speak together with some of its characteristics such as the level of proficiency, and the family of the language.

In the following, we will explain our approach, the design choices and the pendant’s possible usage in real life scenarios.

\*e-mail: tamara.flemisch@tu-dresden.de

†e-mail: frajabiy@ucalgary.ca

‡e-mail: hossem@ucalgary.ca

§e-mail: sheelagh@ucalgary.ca

Electronic proceedings of the IEEE VIS 2015 workshop  
*Personal Visualization: Exploring Data in Everyday Life*  
The authors remain the holders of the copyright

## 2 RELATED WORK

People record their personal daily activities ranging from exercise records, to nutrition intake, to their online interactions. Personal visualization and personal visual analytics give people opportunities to gain insights about their data that may lead to understanding themselves better [2]. One of the many benefits of personal visualization is raising general awareness. It may trigger people to reflect on themselves and to take actions such as reminding people about global warming and encouraging them to take public transport [1]. In addition, personal visualization can influence communities in various domains such as raising public awareness of elections [10] or getting to know their communities by recognizing the most common languages spoken in their city [6].

Physical representation of data has an ancient history. For thousands of years people have used tokens to represent data. Now, people are crafting physical data visualization, called data physicalizations [4] to help people understand, explore, and communicate data. Data physicalization has many possible benefits such as leveraging perception, bringing data into real life, and increasing engagement with data [4].

Data physicalization has been used for different purposes including artistic, educational, and personal self-reflection. There is a growing interest for owning jewellery with personal data. Rezaeian and Donovan designed a DNA jewellery based on a person’s DNA profile [7]. Their goal was to create an aesthetically pleasing piece of jewellery but maintain the data readability of the complex DNA structure. Furthermore, Stusak et al. [8] designed a physical visualization of people’s running activity, *Activity Sculptures*, and investigated its impact on people’s behavior. The findings of their study revealed promising acceptance of the sculptures. There are also industrial services that design and create physical representations of people’s personal data, such as Meshu [3].

Languages have always been a fascinating topic to explore and to analyze. There are many visualizations that try to convey insights about the distribution of languages, their structure, and their origin. A static visualization by López [5] shows the most common languages in the world with their number of native speakers per country. The representation is similar to a radial tree map where the upper level shows languages and the lower level represents the countries. Another example is Tube Tongues by O’Brien [6]. It shows a map of London with its tube stations as circles. The color of the circle represents what the second most spoken language around the station is. The size encodes how many people speak this particular language. The visualization as a whole shows that there are some areas in London which are more popular by some native speakers than others. However, to the best of our knowledge, there is no visualization for the languages a single person speaks. Therefore, in this paper, we designed a piece of jewellery based on the languages people know as a physical visualization to foster communication and engagement in unknown places.

## 3 JEWELLERY DESCRIPTION

In this section, we describe the iterative design as well as different parts of the jewellery that we have designed for physical and personal visualization of the languages a person can speak.



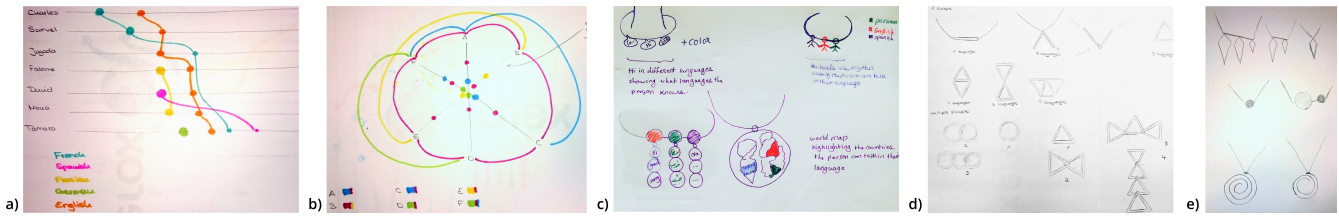


Figure 1: Design iteration with sketches, a+b) Lab-scope visualization of spoken languages, c) First necklace designs incorporating words and symbols, d + e) More abstract, geometric and jewellery-like designs

### 3.1 The Iterative Design

Since we are working in a multi-cultural laboratory environment where people know and talk in various languages, we started to think of different ways that could let people communicate more easily and get to know each other more readily. To this end, we started to sketch different alternatives, three authors were involved in sketching design possibilities and discussing about the advantages and disadvantages of each design by considering its applicability, aesthetics, and usability of the design. Figure 1 demonstrates the gradual development of our design from left to the right. The sketches towards the left are the lab-scope visualizations representing the languages people speak in our laboratory, while the ones towards the right show our personal visualization designs. Figure 4 shows our final sketch representing a pendant for the personal visualization of the languages spoken by a person.

### 3.2 NeckLan

We designed a gender neutral piece of jewellery, NeckLan, which is a wearable tangible data representation of languages that a person knows. The pendant encodes several aspects of personal data, such as the number of spoken languages, the family of each language, and the level of proficiency in each language. In this section, we describe the encoding for the representation of all data aspects, and the design of our pendant.

#### 3.2.1 Language Encoding

Every spoken language has a classification which consists of multiple levels of language families. Each family include a number of languages which have the same origin. In this paper, we only talk about the most commonly spoken languages in Canada<sup>1</sup>.



Figure 2: Four different shapes to represent language families. Indo-European: circle, Sino-Tibetan: triangle, Afro-Asiatic: square, Japonic: pentagon

We chose to use color as a visual variable for representing the spoken languages. As humans can only distinguish and name a very small number of colors and there are far more languages than could be represented by color, we chose to take advantage of two visual variables for encoding the spoken languages. Thus, we used shape as a visual variable for representing the language family and color for demonstrating languages inside a family. Therefore, we can make use of the same colors in combination with each shape which leads to a higher number of representable languages.

<sup>1</sup>Most spoken languages in Canada

We decided on different shapes to represent four language families including Indo-European (circle), Sino-Tibetan (triangle), Afro-Asiatic (square), and Japonic (pentagon) which are common languages among the most spoken languages in Canada. These shapes are shown in Figure 2.

In order to easily differentiate languages in each family, we utilized the nine different colors that Ware has suggested can be distinguished from each other [9]. The selected colors are listed as follows: Green, yellow, orange, red, pink, purple, blue, aqua and white. Table 1 shows the list of languages and their corresponding colors we selected for the purpose of this paper.

Language	Language Family	Color encoding
English	Indo-European	Yellow
French	Indo-European	Orange
Punjabi	Indo-European	Red
Spanish	Indo-European	Blue
German	Indo-European	Green
Italian	Indo-European	Pink
Portuguese	Indo-European	Purple
Persian	Indo-European	Aqua
Russian	Indo-European	White
Mandarin	Sino-Tibetan	Blue
Cantonese	Sino-Tibetan	Pink
Arabic	Afro-Asiatic	Red
Japanese	Japonic	Green

Table 1: The sample languages used and their color encoding.

#### 3.2.2 Proficiency Encoding

People have different levels of proficiency in the languages they can speak. In this paper, we considered three levels of language proficiency: basic knowledge, intermediate knowledge, and fluent. As shown in Figure 3, three triangles inside each other show that either the person is fluent in the language, or that it is his/her native language. Two nested triangles demonstrate intermediate level language skills, and a single shape shows basic level of language proficiency.



Figure 3: Encoding of language proficiency - one shape: basic knowledge, two shapes: intermediate knowledge, three shapes: fluent or native language

Another possible encoding for proficiency level is the size of the element that represents the language. However, due to having different shapes, encoding the proficiency using size makes it incomparable with other shapes. Using the stroke thickness of the elements is another option. However, both possibilities share the problem that their attributes are not discrete but continuous. That means that, we could only encode a continuous level and not three separate levels of proficiency using stroke thickness and size. Therefore, we decided to use the tangible and discrete way of using multiple shapes.

### 3.2.3 Pendant Design

For the overall design of the pendant our intention was to use a gender neutral appearance because we want it to be suitable for everybody. Additionally, we thought that a necklace in general might not be considered gender neutral and that we needed a more general design for our jewellery. Thus, our current design and layout of the single elements allows multiple ways of wearing the jewellery. It can be worn as a bracelet, a short or long necklace, or as a keychain and there are countless more options.

Furthermore, the pendant design is abstract and does not immediately reveal the encoded data, which makes it incomprehensible for people who do not know its real purpose. This is a feature that we included intentionally to explore how this piece of jewellery might encourage people to engage in conversations. For example, one person always knows his/her own languages and their representations. Therefore, he/she will recognize the same elements on other people's pendants and thus will know which languages they have in common. In contrast, people who do not know about the pendant and its goal, or those who just have not seen the shape or color of one element before might find it interesting to ask the wearer about it.

As we have a number of different colors and shapes for encoding the languages, there should be a legend available for reading and comprehending the visual representation of data. This legend could either be a mobile app, a website or a poster hanging on the wall. The clear benefit of the poster is that it attracts people's attention and fosters further conversation about the piece of jewellery.

### 3.2.4 Sample pendant

Figure 4 shows a sample pendant that consists of four language elements. The first one on top always represents the native language. In this case, the triangle shape shows us that the language is Sino-Tibetan, and in this language family, the blue color stands for the language Mandarin. So, the person's native language is Mandarin. Furthermore, we can see that he/she is also fluent in another language which is represented by a yellow circle. The encodings in Table 1 shows that this belongs to an Indo-European language family, more precisely English. The third element consists of two shapes, which means that the proficiency is intermediate. We can see that this Indo-European language is Spanish. Finally, the last element is a green pentagon with only one shape. Thus, the person has basic knowledge in another language that is neither Sino-Tibetan nor Indo-European. The shape and color of this element represent a Japanic language.

In total, the person can speak four languages. He/she is a native Mandarin speaker, is fluent in English, has intermediate knowledge in Spanish and basic knowledge in Japanese. Furthermore, he/she knows two Indo-European languages, one Sino-Tibetan and one Japanic language.

## 4 USING NECKLAN IN PRACTICE

We envision that our design could be useful for encoding and representing other personal information such as a person's favourite music genres or his/her skills in playing musical instruments. In

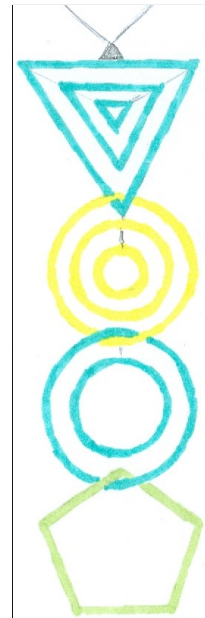


Figure 4: Sample for our design which encodes the spoken languages, their language families and the proficiency of the person for each language.

the following, we provide three main ways that we think NeckLan is useful when representing the spoken languages of an individual.

**NeckLan as an icebreaker:** For people who do not know each other well, NeckLan has the potential to engage them in a conversation. It could provide opportunities to meet new people and it would act as an icebreaker. Special events, conferences, and meetings between different companies are a few examples of using NeckLan in real life scenarios.

**Communicate shared spoken languages:** Another way of using NeckLan is to find out the languages two people are sharing and can communicate with. An interesting example would be representing the sign language and its various dialects. So, deaf people and those who can use sign languages will be able to easily recognize who they can talk to and who will not understand what they are saying.

**Customize based on personal preferences:** One of the main features of personal physical visualizations is optional customizability based on individual's preferences. In our design, we thought of two main ways the languages encoded within the necklace can be customized.

- **Country symbols:** Different people might have different signs in their mind as the symbol of a country or a language. Examples include country flag, the most popular attraction of the country, and the geographic location of the country on the map. Choosing any customized symbol allows the person to select the symbol he/she knows about the most, and this can possibly open up a visual discussion with others about the chosen symbol of the country and its language.
- **Country's most common words:** Languages belonging to different language families probably have distinct alphabets. We hypothesize that encoding one or more common words of each spoken language in the allocated borders of that language can catch the attention of people knowing the same alphabet. Moreover, people who do not know a language might find it interesting to talk to those who know that language and learn the words written on the jewellery.

## 5 CONCLUSION AND FUTURE WORK

We designed a piece of jewellery that encodes data about the languages a person speaks. We think this jewellery has the potential to help people get involved in conversations with new people. It represents the languages a person speaks, their level of proficiency, and the language family of each spoken language.

For now, we decided to use an abstract appearance but there are possibilities of incorporating symbols or words for each language. We chose the abstract design because of its aesthetics and wearability. We anticipate that if too many symbols are encoded, people will not recognize it as a piece of jewellery and therefore, they would not wear it very often. However, symbols and words could have benefits regarding engagement and attracting people's attention. We would like to run a study to see which one of these three versions is more engaging for people.

In the future, we also want to produce some pendants for a study and to see how they look in reality. This could be done with 3D printing or assembling the shapes by ourselves.

### ACKNOWLEDGEMENTS

This research was supported in part by AITF, NSERC, GRAND, Surfnet, and SMART Technologies.

### REFERENCES

- [1] J. Froehlich, T. Dillahunt, P. Klasnja, J. Mankoff, S. Consolvo, B. Harrison, and J. A. Landay. Ubigreen: Investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1043–1052, New York, NY, USA, 2009. ACM.
- [2] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 21(3):420–433, March 2015.
- [3] S. Hwang and R. Binx. Meshu - turn your places into beautiful objects. <http://meshu.io/>, 2012. Accessed: August 27, 2015.
- [4] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian, and K. Hornbæk. Opportunities and Challenges for Data Physicalization. In *CHI 2015 - Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Seoul, South Korea, Apr. 2015. ACM.
- [5] A. L. López. A world of languages - and how many speak them. <http://www.scmp.com/infographics/article/1810040/infographic-world-languages>, 2015. Accessed: August 31, 2015.
- [6] O. O'Brien. Tube tongues. <http://oobrien.com/2014/10/tube-tongues/>, 2014. Accessed: August 31, 2015.
- [7] A. Rezaeian and J. Donovan. Design of a tangible data visualization. In *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*, VINCI '14, pages 232:232–232:235, New York, NY, USA, 2014. ACM.
- [8] S. Stusak, A. Tabard, F. Sauka, R. Khot, and A. Butz. Activity sculptures: Exploring the impact of physical visualizations on running activity. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2201–2210, Dec 2014.
- [9] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.
- [10] J. Wood, D. Badawood, J. Dykes, and A. Slingsby. Ballotmaps: Detecting name bias in alphabetically ordered ballot papers. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2384–2391, Dec 2011.

# Visual Characterization of Personal Bibliographic Data Using a Botanical Tree Design

Tsai Ling Fung\*

Kwan-Liu Ma†

University of California, Davis

## ABSTRACT

This paper presents a preliminary study on egocentric visualization of a bibliographic database. The visualization design is based on a botanic tree metaphor, resulting in a visually interesting and information rich depiction of one's research career in terms of publication records. The case studies reveal both the strengths and limitations of the current design.

**Index Terms:** K.8.1 [Personal Computing]: Application Packages—Graphics; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Evaluation/methodology

## 1 INTRODUCTION

The area of personal visualization emerges as all information about individuals and their activities and behaviors goes online. Personal visualization may be done for either exploring data or facilitating personal improvement. Therefore, both general public and domain experts such as sociologists and security analysts can make use of personal visualization. In this paper, we present a study of individual researchers' professional development in terms of their publications, co-authorships, and associated scientific disputes using a personal visualization design based on a botanical tree metaphor [12].

Online bibliographic databases storing rich information are important resources to scientific researchers. These databases are also of strong interest of sociologists for studying the development of a particular discipline or patterns of academic career. For example, these patterns could be of citations for peer-reviewed journal articles among the scientists working in the same field. In one study [13], an analysis of citation networks shows that scientific debates exhibit distinct spiral patterns.

Our study uses a dataset extracted from Cornell's public scientific library at arXiv.org. This dataset is unique because it contains not only the bibliographic data for published research works, but also scientific disputes among over 400,000 researchers working in physics and related fields. The disputes are in the form of published articles for either commenting a published work or responding to particular comments. Our objective is to visually characterize each individual researcher's career path and compare one's to another's. We like to observe different career paths in terms of one's publication record and co-authorships. We also like to understand the social process due to disputes as well as the implications of the resulting challenges to one's career. Our case studies show that the botanic tree like visualization can effectively bring out particular aspects of one's professional development and also facilitate comparisons.

## 2 RELATED WORK

There has been growing research on visualization of personal content. With data surrounding our lives like medical history [11], ac-

\*e-mail:tlfung@ucdavis.edu

†e-mail:ma@cs.ucdavis.edu

Electronic proceedings of the IEEE VIS 2015 workshop  
*Personal Visualization: Exploring Data in Everyday Life*  
The authors remain the holders of the copyright

tivity logs [1, 19], social network [9], etc., we also see growing demands of tools and devices for visualization and management of the data to facilitate self-development or to gain insights. Recent research done by Huang *et al.* suggests that using visual means on personal data can encourage individuals to examine their personal life and help them promote self-awareness and self-expression, and possibly improve their lives [7]. The types of insight that may be gained are characterized in [3] based on an analysis of quantified self-presentations. As daily data such as email information [17], life events [16], and related cyber activities [6] being increasingly collected into digital formats, it's helpful to map such personal digital archives into visual representations to trigger reflection or narrative and recall of a person. By the help of media, personal visualization can also be a real-time, projected self-expression [9]. To support better personal storytelling, Thudt *et al.* mention the challenges of integrating multiple data sources, personalization, self-reliance and privacy in visualizing personal data as autobiographical visualization, which might be a potential aspect of personal visualization [15].

Our work uses a more organic visualization design, largely based on *Contact Trees* [12], introduced by Sallaberry *et al.* for studying contact diaries [5], data collected for the research of individuals' social contacts and activities. *Contact Trees* uses a botanical tree like visualization design to effectively present multidimensional data with different tree features. Other botanical tree visualizations [2, 8] also encode high-dimension data into aesthetic tree appearances. *LinkedIn Career Tree Builder* is a tool using a node-link diagram laid out into a tree like visualization for highlighting career experience [22]. There are other organic visualization designs. *PeopleGarden* by Xiong *et al.* [21] employs a flower metaphor to visualize online social interactions. *UbiFit Garden* by Consolvo *et al.* [4] shows that using pleasing visualization can help people continuously improve their ideal lifestyle. More recently, Wang *et al.* also use a flower metaphor for visualizing Facebook data [18]. Their study found that more abstract or organic visualizations could be a good trigger for insight gaining, although it could take more time to interpret the visualization.

We use bibliography data in our study. Bibliography data is commonly visualized with node-link diagrams to show the relationship between authors [14, 10]. Since we are interested in studying individual researchers in an egocentric way, a botanic tree design is adequate. Wu *et al.* introduce *PathWay* [20], also an egocentric visualization design, to present and study patterns and trends in an individual researcher's bibliographic data over a selected period of time. In particular, they study not only patterns in scholarly contributions but also scientific disputes associated with one's academic career. We use a dataset similar to theirs. However, *Pathway* is a much more abstract visualization design than our botanic visualization which shows greater details of one's publication record and enables different studies.

## 3 AN EGOCENTRIC VISUALIZATION DESIGN

Our egocentric visualization design is extended from *Contact Trees* [12] with a particular emphasis on usability. The visual representation is a 2D tree, and its usable tree features include trunk size,

trunk sides, tree branches and branch sides, sticks, leaf and its size, color, and shape, fruits, and tree roots. The case studies presented here do not use all the tree features. A web-based, interactive visual interface is provided to the users to make easier mapping of a selected data attribute to a selected tree feature.

Considering the bibliographic database we intend to study, some possible tree feature mappings to use are as follows. The tree trunk presents the tree's major shape giving a view of one's overall career. Trunk size can indicate the total number of articles authored by this researcher. Trunk side can be mapped to some natural binary division of the data. For example, in most of the trees presented here, we place an article with a DOI (Digital Online Identification) number on the right, and without on the left. The other possible division is to place all conference papers on one side, and all journal papers on the other side. Tree branch is probably best to encode time information. Branches from bottom to top of the tree can represent different periods of time over one's career. Branch side can also be used to encode information. For example, we can place regular articles on the upper side of the branch, and articles that give comments to another article or responses to comments on the lower side. Each leaf may display a particular kind of collaborative relationship that the ego has. Leaf color or size can indicate quantitative information such as paper length, number of co-authors, or number of citations received. Finally, fruit can be used to highlight a particular aspect of an article. Egocentric visualizations of the bibliographic data using some of the above mappings are shown in Figure 1.

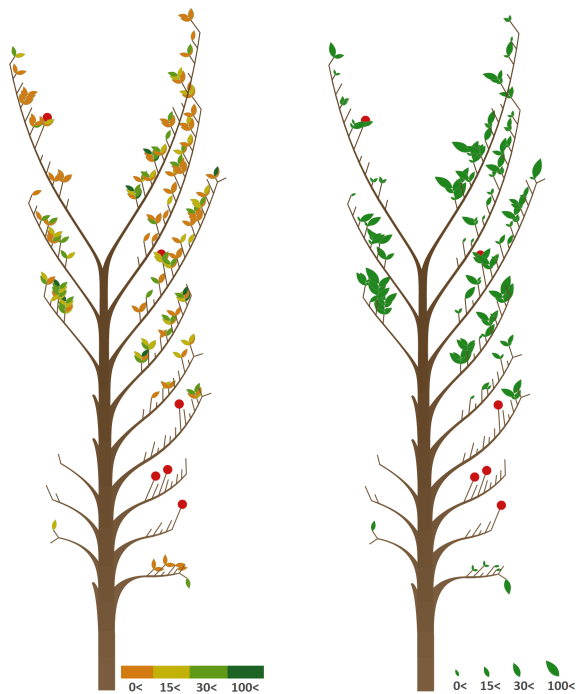


Figure 1: An egocentric visualization of one's publication record using a botanic tree metaphor. Each leaf is a co-author. The total number of publications that the co-author has is mapped to either leaf color (Left) or size (Right). An article with a fruit attached indicates somebody comments on this article.

#### 4 CASE STUDIES

We have chosen a few researchers sharing similar patterns or with contrasting patterns for our studies. The objective is to show usability and identify limitations of the current design. A more comprehensive evaluation of the botanic tree design and a sociological

study of the whole database remain to be conducted.

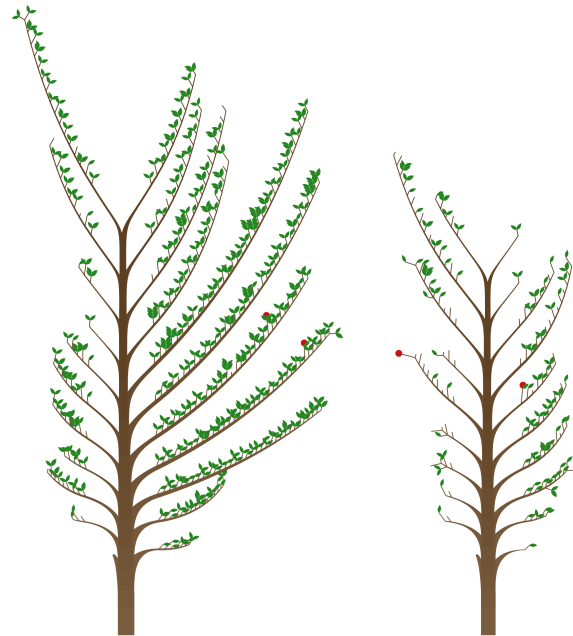


Figure 2: Two active researchers. Each branch encodes two years of publications. Going upward, the trees display their publications between 1993-2013 for the left one and 1995-2012 for the right one.

We start with two more productive researchers as depicted in Figure 2. The right side of trunk displays articles with DOI while the left side display those without. We place regular articles on the upper side of each branch, and those articles commenting others' articles or responding to comments are placed on the lower side. While both researchers look productive over their career, the one on the left has more publications than the one on the right, but only two of his articles are comments to others' works. In contrast, the researcher on the right has written ten articles to comment others' works, and he also has produced more single-author articles.

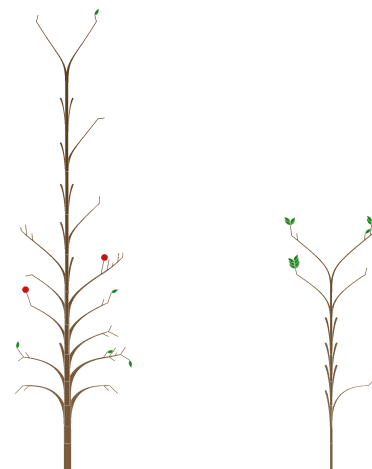


Figure 3: Two less productive researchers.

In contrast, Figure 3 displays the trees of two less productive researchers. It's interesting to see the one on the left was more active over the first half of his career, while the one on the right almost did not publish at all until much later in his career. The



one on the left has several articles commenting on others' works especially early in his career. This researcher also has very few collaborators. It's not clear if the two sets of comments he received in the middle part of his career impacted him or not.

According to [20], we know there are some researchers having very close collaborative relationships. We present two such researchers in Figure 4. As seen, their trees have similar shapes. The similarity is particular strong over the early time of their careers. We examine their similarity further by combining their trees into one as shown in Figure 5. As expected, the resulting tree become quite symmetrical and their close work relationship appears even clearer. Note that we have to change mappings. Here the upper side of each branch displays articles with DOI while the lower side shows articles without. Each leaf is still a co-author. Orange leaves are those articles commenting others' works, while green leaves are regular articles.

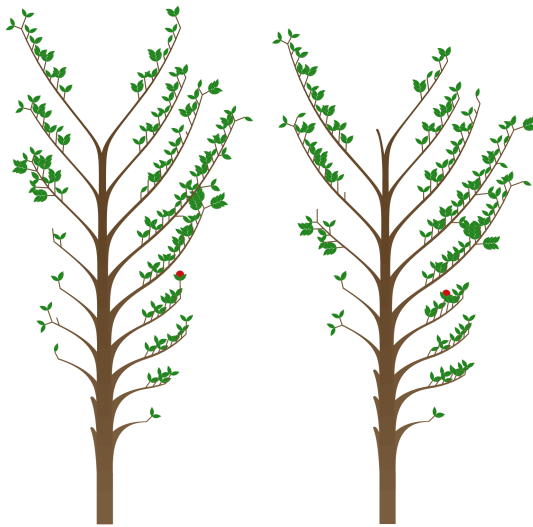


Figure 4: Two researchers with a close collaborative relationship.

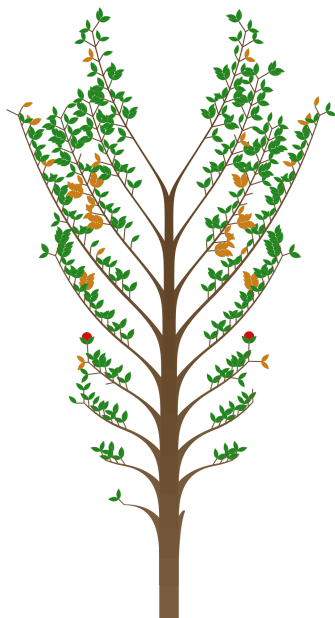


Figure 5: Visualization of the same two researchers in Figure 4 in one tree with each researcher on one side of the tree. The tree looks highly symmetrical as expected.

In the next example, we separate single-authored or first-authored articles from co-authored articles. Figure 6 shows the resulting visualization compared to a partition based on DOI. While this author clearly becomes more collaborative, he still plays significant roles in many of his publications. In particular, the number of single-authored articles do not seem to decrease as this author becomes more senior.

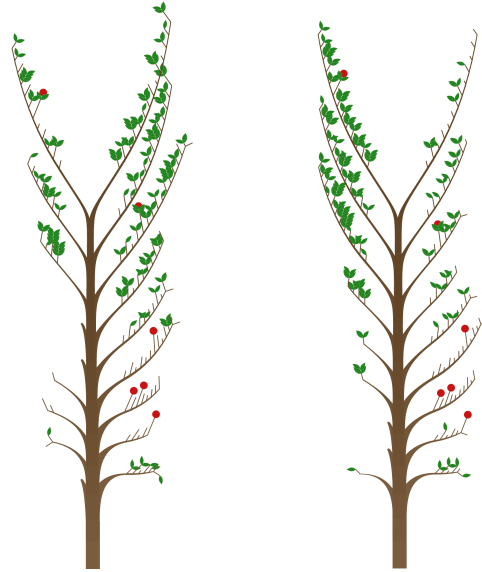


Figure 6: The same researcher in Figure 1. Left: Articles with DOI are placed on the right side of the trunk while those without DOI are on the left side. Right: Single-authored or first-authored articles are placed on the right side of the trunk while co-authored articles are on the left.

Finally, we make a visualization of a researcher's publication record in terms of article length, which is mapped to leaf size. The visualization also separates single-authored or first-authored articles from co-authored articles by trunk side. Two researchers with the identical academic period are displayed in Figure 7. The researcher on the left has largely published on his own while the one on the right has mainly published as a co-author. Both researchers have written articles of different lengths but we can see the one on the right wrote more longer articles. By the dominant color of the tree, it is clear that he has more co-authors in his publications. The other researcher has no publication with more than 10 co-authors since no dark green leaf is found in his tree. It's interesting to see both researchers become more collaborative and tend to write longer papers later in their careers.

## 5 CONCLUSION

The botanic tree design for personal visualization is effective in presenting multidimensional information. Our preliminary study on a bibliography database using this design suggests several directions for further studies:

- While our tree-like visualization looks appealing it is not without limitations. It is not scalable. We shall introduce more abstract levels of the tree for comparison and selection tasks. Furthermore, current tree features limits how many categories an attributes can have. For example, there might not be an attribute in the dataset that can be naturally mapped to the two sides of a tree.
- The usability of the visualization very much depends on how easy it is for the user to specify each tree feature mapping. Our system provides an interactive interface for users to specify



Figure 7: The same researcher in Figure 6 (left) compared to another researcher for the same academic period 1997-2013 (right). Leaf size indicates the length of a paper. Leaf color tells us if the article is co-authored or not and approximately how many co-authors. Single-authored or first-authored articles are placed on the right side of the trunk while co-authored articles are on the left. Articles with DOI are placed on the upper side of each branch while those without are on the lower side.

mapping of tree features. However, this interface remains to be evaluated and improved.

- It seems plausible to make image-level tree classification to find like egos.
- It is desirable to provide alternative organic visualization designs to the users. One design might match one particular type of data better.
- We shall engage sociologists for studying sociological aspects of one's academic career development.

We also like to enhance the bibliographic database by collecting other data such as citation numbers for conducting deeper analyses.

#### ACKNOWLEDGEMENTS

This research has been sponsored in part by the National Science Foundation via grants NSF DRL-1323214 and NSF IIS-1320229, and UC Davis RISE program.

#### REFERENCES

[1] J. S. Bauer, S. Consolvo, B. Greenstein, J. Schooler, E. Wu, N. F. Watson, and J. Kientz. Shuteye: Encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1401–1410, 2012.

[2] E. B. Chlan and P. Rheingans. A botanically inspired high-dimensional visualization with multivariate glyphs. In *Proceedings of the Sixth Joint Eurographics - IEEE TCVG Conference on Visualization*, VISSYM'04, pages 231–236, 2004.

[3] E. K. Choe, B. Lee, and M. Schraefel. Characterizing visualization insights from quantified selfers' personal data presentations. *Computer Graphics and Applications, IEEE*, 35(4):28–37, July 2015.

[4] S. Consolvo, D. W. McDonald, and J. A. Landay. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 405–414, 2009.

[5] Y.-C. Fu. Contact diaries: Building archives of actual and comprehensive personal networks. *Field Methods*, 19(2):194217, 2007.

[6] J. Hailpern, N. Jitkoff, A. Warr, K. Karahalios, R. Seseck, and N. Shkrob. Youpivot: Improving recall with contextual search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1521–1530, 2011.

[7] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 21(3):420–433, March 2015.

[8] E. Kleiberg, H. van de Wetering, and J. J. Van Wijk. Botanical visualization of huge hierarchies. In *Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, pages 87–, 2001.

[9] M. Leung, M. Tomitsch, and A. Vande Moere. Designing a personal visualization projection of online social identity. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1843–1848, 2011.

[10] S. Noel, C.-H. H. Chu, and V. Raghavan. Co-citation count vs correlation for influence network visualization. *Information Visualization*, 2(3):160–170, 2003.

[11] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman. Lifelines: Visualizing personal histories. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pages 221–227, New York, NY, USA, 1996. ACM.

[12] A. Sallaberry, Y. chih Fu, H.-C. Ho, and K.-L. Ma. Contact Trees: A Technique for Studying Personal Network Data. *CoRR*, abs/1411.0052, 2014.

[13] U. Schwed and P. Bearman. The temporal structure of scientific consensus formation. *American Sociological Review*, 75:817840, 2010.

[14] Z. Shen, M. Ogawa, S. T. Teoh, and K.-L. Ma. Biblioviz: A system for visualizing bibliography information. In *Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation - Volume 60, APVis '06*, pages 93–102, 2006.

[15] A. Thudt, D. Baur, and S. Carpendale. Autobiographical visualizations: Challenges in personal storytelling. In *In Proceedings of the DIS'14 Workshop on A Personal Perspective on Visualization and Visual Analytic*, 2014.

[16] A. Thudt, D. Baur, and S. Huron. Visual mementos: Reflecting memories with personal data. *Visualization and Computer Graphics, IEEE Transactions on*, PP(99):1–1, 2015.

[17] F. Viegas, D. Boyd, D. Nguyen, J. Potter, and J. Donath. Digital artifacts for remembering and storytelling: posthistory and social network fragments. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10 pp.–, Jan 2004.

[18] S. Wang, Y. Tanahashi, N. Leaf, and K.-L. Ma. Design and effects of personal visualizations. *Computer Graphics and Applications, IEEE*, 35(4):82–93, July 2015.

[19] J. Wood. Visualizing personal progress in participatory sports cycling events. *Computer Graphics and Applications, IEEE*, 35(4):73–81, July 2015.

[20] M. Q. Y. Wu, R. Faris, and K.-L. Ma. Visual exploration of academic career paths. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 779–786, 2013.

[21] R. Xiong and J. Donath. Peoplegarden: Creating data portraits for users. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, UIST '99, pages 37–44, 1999.

[22] N. YAU. (2010) Build your LinkedIn career tree. [online]. Available at <http://flowingdata.com/2010/11/29/build-your-linkedin-career-tree/>.

# Sharing Information from Personal Digital Notes using Word-Scale Visualizations

Pascal Goffin\*  
Inria

Wesley Willett†  
University of Calgary

Petra Isenberg‡  
Inria

## ABSTRACT

We describe how small visualizations embedded in text (word-scale visualizations) can be used to share information from and in personal notes. From our previous research, we learned that people see many opportunities for sharing personal notes, for example among a small social group. Yet, people reported that they were hesitant to share raw notes due to the notes' often disorganized structure, haphazard writing style, or due to the fact that notes may contain a number of unrelated or irrelevant pieces of information. In this paper, we discuss how word-scale visualizations can be used in a collaborative personal visualization setting—to show abstracted information from a shared set of notes in the context of personal note-taking. In particular, we discuss potential kinds of data about notes that can be shared and motivate why sharing them may be helpful. Additionally, we provide two examples that illustrate the challenges and implications of using word-scale visualizations to share data in notes. The first example describes how notes in a shared notebook can be combined with private notes. The second example shows how data from public sources can be mixed with private comments to support sharing between notes on a common topic.

**Index Terms:** Information visualization, personal visualization, digital note-taking, awareness, collaboration, word-scale visualizations.

## 1 INTRODUCTION

Digital notebooks can store a variety of different types of personal information, including recipes, health related information, results of web searches, shopping lists, or personal diaries. Digital notes are, thus, a unique and common type of personal data. They offer numerous opportunities for personal analysis—for example, examining the evolution of a person's culinary tastes or retracing a person's daily activity.

While note-taking is typically a solitary activity, for certain situations and tasks sharing the content of notes is important. Examples include a family planning their summer vacation or the members of a book club reading a book and preparing themselves for the next meeting. Many note-taking tools support content sharing. However, note-takers typically refrain from sharing their notes because they are concerned about how their notes and note-taking style will be perceived by others [10]. As a result note-authors clean, reformat, and even rewrite their notes before sharing or simply refrain from sharing their notes at all. Yet, they seem to be comfortable with sharing some information about the notes' content—including which terms in the note are also used by their collaborators[10].

Sharing information and ideas among acquaintances and family members can be a perfect way to spark discussions, prepare events, and increase awareness of other people's thoughts. For example

\*e-mail: pascal.goffin@inria.fr

†e-mail: wj@wjwillett.net

‡e-mail: petra.isenberg@inria.fr

when planning a summer vacation, several family members may use their personal digital notebooks to collect information about potential destinations, what activities or points of interest are available at these locations, or which friends live close by and could be visited. Later, the family must collaborate and consolidate these notes in order to settle on a single vacation plan.

Here, we explore the use of word-scale visualizations—small data graphics that display information associated with specific words or word combinations in text—to enrich personal notes with information from collaborators' notes. In the vacation planning scenario above, small visualizations embedded in a note about possible vacation spots could show which other family members have listed similar properties, locations, or countries and, thus, aid in the process of finding a joint destination.

In this paper, we discuss how word-scale visualizations can be used to show descriptive metadata from collaborators' notes. We specifically focus on sharing only metadata extracted from notes as a simple channel for communicating shared ideas and understanding collaborators' activities. By focusing on metadata we also free collaborators from the burden of feeling that they need to reformat and clean their notes before sharing.

Here, we explore and describe which types of information originating in notes could be shared between friends and family members. We discuss potential problems that can arise when sharing personal information and then elaborate on how word-scale visualizations could help to support sharing relevant pieces of information between members of a small social group. Finally, we illustrate our discussions with two examples.

## 2 RELATED WORK

In recent years, digital note-taking tools have become more and more available both inside and outside the workplace. Yet, general information on the characteristics of personal digital notes is relatively sparse. We base most of our knowledge on digital notes on our own previous study [10] of digital note-taking practices and their implications for visualization. We found that while notes that are digital can theoretically be easily shared and distributed, people typically felt the need to edit notes prior to sharing. Hence, we discussed how visualization can be a good way to abstract notes and prepare them for sharing. In that previous work, we discussed visualizations that can create awareness of overlapping concepts and phrases in collaborators' digital notes. Here, we discuss in particular how small-scale visualizations can integrate metadata from friends' and family's notes in-context of one's own notes.

Chuah and Roth [2] previously discussed how visualizations can provide common ground during collaboration. One example of the use of visualization to provide common ground was introduced by Brennan et al. [1]. In their work, private views of a graph visualization in a distributed work scenario are merged to show shared information and common ground. Keel [5] discusses a similar idea, using computational agents to identify information in private views that should be relayed to a larger group of collaborators. Similarly, we use word-scale visualizations to provide common ground for small social groups.

Other researchers have used visualization to create awareness about collaborators' search activity. Awareness of other people's

data and actions is an important prerequisite for common ground formation and several tools have attempted to provide mechanisms for facilitating awareness. For example, in Cambiera [4] Isenberg et al. use small visualizations (e. g. bars) as Scented Widgets [11] to show the overlap between the documents two participants had searched for and used to provide awareness of the pair’s reading histories. The authors termed this concept of explicitly visualizing collaborators’ activities “collaborative brushing and linking.” Similarly, Mayar and Tory [7] use partial merging to provide subtle awareness of other people’s work. They term this concept of merging “linked common work” and implement it in their tool CLIP. The tool includes node-link graphs in which nodes have color coded overlapping segments to represent common and different information among collaborators. This is similar to our approach, which uses word-scale visualizations to provide subtle awareness of information that is shared between collaborators’ notes. Finally, Morris and Horowitz’s SearchTogether [8] creates awareness of web search histories by showing what collaborators have searched for and by explicitly sharing information via recommendations. In contrast to these previous three tools, we use small visualizations embedded directly within the text of digital notes rather than stand-alone visualization tools.

### 3 WORD-SCALE VISUALIZATIONS AS AN APPROACH TO PRESENT SHARED INFORMATION

Here, we discuss how word-scale visualizations embedded in notes can support shared awareness among a small social group. We begin with a short definition of word-scale visualizations before discussing the types of data that can be shared.

#### 3.1 Why word-scale visualizations?

Word-scale visualizations are small data graphics that are associated with words or sets of words (which we refer to as “entities”) and integrated alongside them in the text. Word-scale visualizations are similar to sparklines [9], in that they are “small, intense, simple, word-sized graphic with typographic resolution”, but can use a wider variety of visual encodings and sizes [3]. Small word-scale visualizations can be the size of a single letter, while larger ones can approach the size of a sentence or paragraph.

Word-scale visualizations are small and can easily embed additional information in text without being obtrusive. As a result, we can use word-scale visualizations to add additional information in the context of a note-taker’s original notes. This can help note-takers better understand the overlap between their notes and their collaborators’, without requiring a separate interface.

#### 3.2 Sharing metadata and snippets

In our previous study of digital note-taking [10], we found that note-takers were generally open to sharing their notes but were worried that unedited notes would overwhelm collaborators with unimportant information. In order to facilitate sharing notes, we therefore propose using word-scale visualizations that show abstracted metadata and snippets from notes. Using only metadata and brief snippets alleviates collaborators from the burden of having to edit their notes and, thus, simplifies sharing.

We specifically discuss descriptive metadata that is created when a note is saved or edited and snippets extracted from the notes themselves.

##### 3.2.1 Technical metadata

*Where the note was created:* Sometimes digital notebooks store GPS locations when notes are created on mobile devices with location services turned on. Seeing where notes were written could, for example, allow you to see if a collaborator has visited and taken notes on a potential vacation destination. While location data can

be particularly sensitive, it is also easy to abstract (e.g. reporting only the city, country, or even continent) and obfuscate [6].

*Who authored or edited a note:* In personal notebooks, notes are typically first authored and later edited by the same person. When sharing data from notes, the author/editor information might be crucial in order to judge the potential usefulness of information given existing background information in a social group. This information could also make it easier to contact the author directly and request the full content of a note.

*When was a note created or edited:* This information can be an indicator of how recently someone has engaged him or herself with the topic of a note. In addition, this information could give additional information on the note type. In our previous work [10] we found that certain types of notes, such as project logs or running lists were edited relatively frequently while notes containing information scraps or references and copied content were rarely if ever edited by note-takers. Thus, data about note creation and editing patterns—coupled with other metadata—can provide a powerful tool for deducing the importance of a note.

*Other metadata:* There is a wealth of other technical metadata available for notes that can be the basis of a shared visualization. For example, the length of a note may hint at how much information a note contains. The note content type describes if the note has been authored by the note-taker or if it only contains clippings from public sources (what Willett et al. [10] call “references and copied content”) or if it is a mix of authored and clipped content. These clippings can be text or image clippings from websites such as Wikipedia or any other website. The tags or notebooks used to file the note might also be useful when identifying related notes.

##### 3.2.2 Extracted keywords

In order to avoid sharing the raw text from notes but still provide cues on the semantic content of notes, automatically extracted key terms from notes could serve as a data source. These terms can be keyword labels applied to the note by its author or named entities like locations, dates, and surnames that are extracted directly from the text.

User-authored keywords or entities can succinctly describe the topic or ideas of a note without disclosing much unrelated information or revealing a note’s disorganized structure. Entities can be found in the text through entity recognition and text processing. They can also be manually tagged by the original note-taker. For visualization purposes, entities or keywords can be used to cluster notes that are of interest to a specific topic. For example for planning a trip to New York City, notes containing entities such as “New York City”, “Manhattan restaurants”, or “MoMA” are relevant and may provide inspirations of what to visit and where to eat.

In addition to the mere presence of a specific entity, a diverse set of statistics can be gathered around entities. One of them is the number of occurrences of an entity in a note. Others include co-occurrence patterns with other entities or entity types and semantic relationships between entities extracted from knowledge structures such as WordNet. The number of occurrences, for example, can be a good indicator of what to expect from a note in terms of information relevant to a certain topic.

## 4 EXAMPLES

We discuss two information-sharing scenarios that illustrate how word-scale visualizations embedded in notes can show some of the data mentioned in the previous section and help support information sharing in a small social group. Additionally, we provide sketches of potential word-scale visualizations.

### 4.1 Shared recipes with Roommates

In this scenario, four flatmates live in a shared apartment. They like to cook together in the evenings and on weekends, and enjoy creat-



Figure 1: This sketch is an excerpt of a recipe from the flatmates’ shared recipe notebook. Each ingredient is accompanied by a word-scale visualization depicting the number of recipes with that ingredient shared by other members of the group.

ing new menus. Each of the apprentice cooks has a notebook called recipes. All of them like to collect interesting recipes from their family and friends. Moreover, they enjoy browsing the Internet to find interesting recipes to add to their collection. The four flatmates created a shared digital notebook where they collect and share large numbers of recipes with each other.

This shared notebook provides a mechanism for sharing specific recipes, as well as for planning future meals, identifying interesting recipe pairings, and selecting sets of recipes that share ingredients. However, if the shared folder contains a large number of recipes it may be difficult to get a sense of which recipes others in the group are interested in or which have overlapping ingredients.

#### 4.1.1 Sketch

Figure 1 shows a note containing a recipe for “Zitronencake” which has been shared by one of the flatmates. A small bar chart next to each ingredient highlights how many other recipes in the shared folder contain this ingredient as well as which roommates added them. Each of the bars corresponds to a single roommate and indicates how many recipes with that ingredient he or she has contributed. Hovering over a bar opens a small window that provides links to the three recipes with that ingredient which that roommate has most recently read or added. These visualizations make it easier for the flatmates to navigate the large collection of shared notes and identify interesting recipes added by their peers that might pair well together or which use the same ingredients.

### 4.2 Book Club

Our second scenario—a book club with about ten members—differs in that no notes are shared directly. Instead, each member of the club keeps their own personal notes about the current book and those notes can be annotated to provide peripheral indicators of other members’ activity. Each member’s individual notes on the book usually contain direct quotes from the work as well as observations about the characters, places, events, and ideas mentioned in the book, written in their own words. For example, a member may take notes discussing why they think a character behaves in one way or another or describing interesting situations that appear in the book. He or she may also discuss relationships between characters or describe feelings and ideas that occurred when reading a memorable passage.

#### 4.2.1 Sketch

Figure 2 shows how a member’s personal notes about the book could be augmented to provide more information about which as-

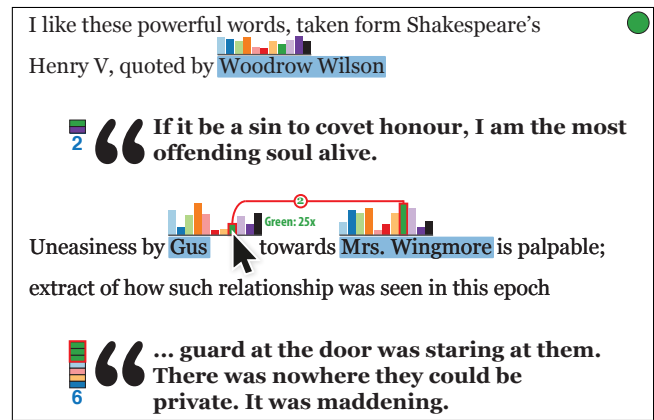


Figure 2: This sketch is an excerpt of one of the book club member’s note about the current book. The bar chart describes the number of occurrences of the character in each book club member’s current book note.

pects of the book were interesting to their fellow members. Specifically, the augmented note highlights the shared use of quotes from the book as well as how often other members used the names of important characters, places, and ideas. The quotes can be seen as data from a public source—the book—and can be presented in more detail, whereas the comments about the characters and places might be seen as more private. As a result, the note can be augmented to highlight shared mentions of those entities—letting the note-taker know which other members have something to say about these points, without directly revealing their personal thoughts on the topic. This information about shared quotes and entities can then be used to steer the conversations when the club meets in person.

In Figure 2, the colored circle in the upper right corner of the note indicates its owner—Mrs. Green. The bar charts over each character’s name depicts how many times that character was mentioned in each of the other members’ notes. Interaction can be used to show more details. By hovering over a bar, the note-taker can reveal the name of the book club member and the number of occurrences of the entity. Additionally, hovering highlights all bars belonging to Mrs. Green in the paragraph. Any characters that are mentioned in the same paragraph are visually connected via a line. The strength of this relationship—number of co-occurrences of the characters in the note’s paragraphs—is also shown on the line.

Quotes drawn from the book are marked with a large quotation mark, indicating that this complete snippet of public material is present in multiple members’ notes. The bar on the left of the big quotation mark indicates which other members included the quote as well as how many times they included it (for example, Mrs. Green included the second quote three separate times, when discussing different aspects of the book). The note-taker can also click on the bar to see the names of the characters associated with the quote.

## 5 CHALLENGES AND IMPLICATIONS

Using word-scale visualizations to share metadata and entities drawn from personal notes requires careful consideration of the content of the visualizations and openness of the collaborators. Moreover, designers must create visualizations that help to abstract note content but also do not unintentionally remove or create links between pieces of information.

## 5.1 Visualization

The design space for word-scale visualizations is very large. However, an interesting implication for word-scale visualizations in the context of notes is that notes are often edited. As a result, even though embedding visualizations in-line can minimize the degree to which the text is disrupted [3], it may break note-takers' expectations about the behavior of common commands like cut, copy, paste, or backspace. For example, what happens if the cursor is to the right of a word-scale visualization? Will it be okay to delete using backspace? As a result, the best position for the word-scale visualizations may depend on whether or not the note is likely to be edited. Consider the word-scale visualizations we placed in-line in the sketch for the recipe scenario. Since ingredients are typically given as itemized lists with plenty of empty white space to the right of each item, and recipes are less often edited than other notes [10] this placement is less likely to cause problems.

## 5.2 Privacy

The degree of privacy attached to information in a note has a wide spectrum. On one end is the raw note, that when shared with associated metadata discloses everything. The other extreme is a setting in which nothing at all is shared and all notes are kept private. In between these two endpoints we have many possibilities. Certainly there are tradeoffs between sharing and providing awareness and privacy. Word-scale visualizations can be a way to deal with this tradeoff. By the sharing metadata and entities of a note there is a level of flexibility available. Not all metadata, entities, or types of entities have to be shared if not wanted. Entities can be chosen through entity recognition and text processing. For digital notebooks, however, this means that an additional infrastructure has to be available to provide the possibility to regulate which entities and metadata are shared and to whom.

## 5.3 Sharing useful data

Word-scale visualizations provide a degree of abstraction from the original note. Yet, when they are based on metadata such as entities or keywords, word-scale visualizations might provide less information than originally contained in the note. For example, relationships can disappear or be created by inferring that two characters from a book appear in the same paragraph. In fact, the two characters might appear in unrelated sentences or even in quotes from separate books, but the abstraction of the paragraph may seem to imply a relationship. The challenge of choosing the right information to share through word-scale visualizations can be complex. This has implications for the design of the word-scale visualizations. For example, a word-scale visualization that groups entities based on co-occurrence at the sentence level could help avert incorrect assumptions about the relationship between entities, but runs a greater risk of revealing the raw content.

## 6 CONCLUSION

Word-scale visualizations provide a lightweight way of sharing important pieces of information from personal notes, and can support awareness and decision making without requiring family, friends and collaborator to reveal their raw notes. Personal digital notes are a treasure trove for personal analytics. This work represents a first attempt at helping make family, friends and collaborators more aware of one another while keeping sharing light-weight, avoiding the need for note editing prior to sharing, and placing shared data in-context to already existing personal information.

## ACKNOWLEDGEMENTS

This work is sponsored by the French Research Organization, project grant ANR-11-JS02-003 and supported by the Collaborative European Digital Archive Infrastructure project CENDARI (cendari.eu).

## REFERENCES

- [1] S. E. Brennan, K. Mueller, G. Zelinsky, I. Ramakrishnan, D. S. Warren, and A. Kaufman. Toward a multi-analyst, collaborative framework for visual analytics. In *Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST)*, pages 129–136, 2006.
- [2] M. C. Chuah and S. F. Roth. Visualizing common ground. In *Proc. of the Intl. Conference on Information Visualization (IV)*, pages 365–372. IEEE, 2003.
- [3] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Exploring the placement and design of word-scale visualizations. *IEEE TVCG*, 20(12):2291–2300, 2014.
- [4] P. Isenberg and D. Fisher. Collaborative brushing and linking for co-located visual analytics of document collections. *Computer Graphics Forum*, 28(3):1031–1038, 2009.
- [5] P. E. Keel. Collaborative visual analytics: Inferring from the spatial organization and collaborative use of information. In *Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST)*, pages 137–144, 2006.
- [6] J. Krumm. A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6):391–399, 2009.
- [7] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE TVCG*, 20(12):1633–1642, 2014.
- [8] M. R. Morris and E. Horvitz. SearchTogether: An interface for collaborative web search. In *Proceedings of the Conference on User Interface Software and Technology (UIST)*, pages 3–12. ACM, 2007.
- [9] E. R. Tufte. *Beautiful Evidence*. Graphics Press, Cheshire, CT, 2006.
- [10] W. Willett, P. Goffin, and P. Isenberg. Understanding digital note-taking practice for visualization. *IEEE Computer Graphics and Applications, Special Issue on Personal Visualization and Personal Visual Analytics*, 35(4):38–51, 2015.
- [11] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007.



# Visualizing Personality Traits Derived from Social Media

Liang Gou\*

IBM Almaden Research Center

## ABSTRACT

Leveraging the findings in Psycholinguistic studies, we now can automatically compute an individual’s personality traits, including Big 5 traits, emotions, and human basic values, from one’s linguistic footprints left on social media. To help users interpret one’s personality traits derived from social media, we are building interactive visual analytic systems with a set of visual representation to summarize the derived traits and explain the associated linguistic evidence to generate such traits. In this paper, we present the design challenges and two prototype systems for visualizing personality traits derived from social media. We also discuss future challenges for personality trait visualization.

**Keywords:** visual text analysis, personality traits, personality, social media, information visualization.

## 1 INTRODUCTION

A number of human studies have shown that one’s personality traits, such as Big 5 personality, motivation and belief, influence a person’s behavior and performance in the real world, including occupational proficiency and economic decisions [7, 18]. With the emergence of social media, hundreds of millions of people leave their digital footprints on the Internet, which can be leveraged to automatically infer various personality traits (e.g., [8, 14, 9]).

Inspired by Psycholinguistic studies (e.g., [22, 20]) and numerous efforts of applying such study results to social media analytics (e.g., [4, 8, 9]), we are developing personality analytics that can automatically derive an individual’s personality portrait from their linguistic footprints left on social media (e.g., Twitter and Facebook). Such a portrait is characterized by multiple types of personality traits that uniquely identify an individual. Each type of trait further comprises a hierarchy of sub-traits. For example, one’s Big 5 personality trait has five dimensions, each of which has five or six sub-facets. To better understand how these traits are derived, such trait data are associated with linguistic evidence extracted from text based on personality analytic models. For example, the *Agreeableness* trait from Big 5 personality is positively correlated with *Inclusive* words such as “we”, “us”, “together” and “with”.

Given the complexity of the derived traits, we are facing some challenges of designing visual analytic tools to help people explore and understand such personality traits derived from social media. In this paper, we present the unique data characteristics of personality data, outline the design challenges for exploring and explaining such data, and show several visual analytic systems designed to alleviate these challenges.

## 2 RELATED WORK

### 2.1 Personality Analytics from Text

The findings in Psycholinguistic studies (e.g., [20, 22]) and the availability of people’s online text footprints have inspired a large body of work on automatically inferring one’s personality traits from one’s text footprints. For example, there are many efforts

on predicting one’s Big 5 personality traits from various types of text, including from essays and conversation scripts [12], and social media [8]. Besides deriving one’s personality traits, there are also works on inferring other personal traits, such as political orientation [14] and emotional states [6].

Similar to the existing efforts, ours also aims at deriving one’s personality traits from text. Unlike these works, ours focuses on inferring an extended set of finer-grained personality traits and visually explaining the inferred traits to end users.

### 2.2 Visualization for Personal Trait Data

The results of personality analytics can be complex and abstract, which make them difficult for users to comprehend and consume. For example, our personality analytics currently produces a hierarchy of 50+ personality traits. One of our challenges is to find a set of suitable visual presentation to convey the derived traits and associated linguistic evidence.

Few existing work is designed and developed to visually explore and explain personality trait, although increasing amount of visual analytic tools are presented around personal data, such as physical movements, health and well-beings [11, 3]. The closest work along this line of research is visualization of emotions and sentiments derived from social media. For example, Whisper [2] traces the process of information diffusion via microblogs on various aspects including temporal trends, sentiment states and community topics. AffectAura [13] captures, predicts and visualizes emotions through a user’s valence, arousal and engagement values. PEARL [23] visualizes the emotion changes over the time and how the emotions are derived from tweets.

## 3 PERSONALITY ANALYTICS

Currently, we model five types of personality traits. Here, we provide an overview of these traits and how they are computed.

### 3.1 Five Types of Personality Traits

Table 1 summarizes the five types of personality traits and their facets that we currently derive. First, the *Big 5* personality traits capture one’s unique and enduring personality character from five aspects, each of which further includes five or six sub-facets [22]. Second, the *needs* model captures a person’s fundamental needs along 12 dimensions (6 of them are currently used in PETALS), which often influences one’s brand and product choices [7]. Third, the *values* model characterizes one’s beliefs and motivators along five major dimensions [18]. Fourth, the *emotion* reflects one’s primary emotional state (e.g., angry and joy) [16]. Finally, the *social genome* characterizes one’s social network make up (e.g., homogeneous vs. heterogeneous network). This trait can be used to determine the person’s network potential—how much and whom the person can influence within his/her own social circle.

### 3.2 Personality Trait Modeling and Analysis

Currently, we use a lexicon-based, psycholinguistic analysis to process one’s tweets and derive the first four types of traits based on the particular word uses in the tweets. Specifically, we use the LIWC [15] dictionary and its extensions to derive the Big 5 personality traits and *Values* traits, our own lexicon to infer *Needs*, and ANEW [1] to extract *Emotions*. The LIWC dictionary also includes word categories that group similar words together. For example,

\*e-mail: lgou@us.ibm.com

<b>Big 5 Personality</b> (Use: emotion and problem coping; Lexicon used: LIWC [15] and Yarkoni [22])
<i>Openness</i> (Imagination, Artistic interests, Emotionality, Adventurousness, Intellect, Liberalism)
<i>Conscientiousness</i> (Self-Efficacy, Orderliness, Dutifulness, Achievement striving, Self-discipline, Cautiousness)
<i>Extraversion</i> (Friendliness, Gregariousness, Assertiveness, Activity-level, Excitement-seeking, Cheerfulness)
<i>Agreeableness</i> (Trust, Morality, Altruism, Cooperation, Modesty, Sympathy)
<i>Neuroticism</i> (Anxiety, Anger, Depression, Self-consciousness, Immoderation, Vulnerability)
<b>Needs</b> (Use: brand and product choice, Lexicon used: our own)
<i>Ideals</i> : a desire for perfection
<i>Harmony</i> : appreciating other people, their feelings
<i>Closeness</i> : being connected to family and setting up home
<i>Self-expression</i> : discovering and asserting one’s own identity
<i>Excitement</i> : upbeat emotions, and having fun <i>Curiosity</i> : a desire to discover and grow
<b>Values</b> (Use: motivations and incentives, Lexicon used: our own)
<i>Self-transcendence</i> (universalism, benevolence)
<i>Conservation</i> (conformity, tradition, security)
<i>Self-enhancement</i> (achievement, power)
<i>Openness-to-change</i> (stimulation, self-direction)
<i>Hedonism</i>
<b>Emotion</b> (Use: interaction state, Lexicon used: ANEW [1])
Anger, Disgust, Fear, Joy, Sadness, Surprise, Trust, and Anticipation
<b>Social Genome</b> (Use: network potential, Lexicon used: none)
Social network makeup (homogeneous vs. heterogeneous network)

Table 1: Five types of traits derived from social media in PETALS.

one category called “*inclusive*” includes words like “*with*” and “*together*”. To infer the *social genome*, we compute the percentile of one’s social network who has the *similar* personality of the person. The bigger the percentile is, the more influence the person has more on her own social network, assuming that *like-minded* people influence each other more.

In general, a type of trait,  $T_i$ , in our model can be derived based on the appearance of certain words or word categories in one’s text footprints (e.g., tweets) and the frequencies of their appearance. Computationally, it can be modeled as a linear combination of relevant words and associated categories (if any):

$$T_i = f_i(LC_j(w_k)), \quad (1)$$

Here  $f_i(\cdot)$  is a linear mapping function,  $LC_j(\cdot)$  is a mapping function for word categories, and  $w_k$  is a word that is correlated with trait  $T_i$  and appears in the tweets. The correlated word categories or words, and the weights (correlational co-efficients) in these mapping functions are found by various Psycholinguistic studies (e.g., [22]). If there is no associated word category,  $LC_j(\cdot)$  then becomes a self-mapping function. The full detail on how each type of trait is derived can be found in [4, 21].

### 3.3 A Data Model for Personality Analytics

Our derived personality traits (Table 1) contain multi-dimensional values and some of them are also hierarchical. To support efficient visual representation and interactive exploration of these traits and their associated information (e.g. linguistic evidence), we use a compound graph model, TKG, to manage both trait values and other related information (Figure 1). The graph model is defined as:

$$TKG = (T, K, \rho) \quad (2)$$

The traits themselves are represented as a tree,  $T = (V_T, E_T)$ , where a set of nodes  $V_T$  denotes the traits, and a set of edges  $E_T$  denotes the relationships between traits and their sub-traits. Each trait has a value and two meta properties. The linguistic evidence

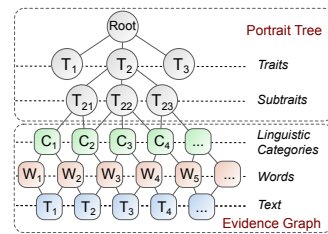


Figure 1: PETALS data model that represents all the data.

used to derive a trait may include multiple levels of information, including related word categories, words, whole tweets, and weighted links among them. This data model is represented by a  $k$ -partite graph [17],  $K = (K_1, K_2, \dots, K_k, E_K)$ , where the  $K$ th part includes a set of nodes,  $V_K$ , representing a set of evidence at a particular level (e.g., word level), and a set of edges,  $E_K$ , representing the relationships between the entities. A mapping function,  $\rho$ , links the traits to their associated evidence,  $\rho(n_T) \rightarrow n_k$ , where  $n_T \in V_T$ ,  $n_k \in V_k$ .

## 4 DESIGN CHALLENGES

Personality analytics includes two types of data: (1) five types of personality traits, including their meta properties, and (2) linguistic evidence used to derive the traits. With this unique data, we first present the design challenges and goals for personality visual analytics.

The general goal of personality visual analytics is to aid users in understanding, exploring, and assessing their psychological portraits characterized by a set of personality traits. To achieve this goal, there are four design challenges.

- DC1 **Easy Visual Exploration.** A system should use proper visual presentation to help users explore a personality profile with multiple traits. It should enable users easily interact and navigate through the hierarchical structure and also other properties, such as temporal changes over the time.
- DC2 **Accountable Visual Explanation.** A system should provide accountable visual explanation of the multi-level linguistic evidence discovery to help users understand how traits are derived. Our design must enable users to easily link analytics results to their evidence at multiple levels and facilitate the verification of the results.
- DC3 **Engaging and Enjoyable.** Personality traits are about people and artifact derived from people’s data, and the system may be used by users repeatedly. To engage users to explore their data, the system should be enjoyable and fun to use, since positive affects can improve the users’ performance in their visual tasks [10].
- DC4 **Novel Visual Metaphor.** The personality profile is a new type personal data, and a novel visual metaphor could help people digest the data. It should be
  - (a) *Intuitive.* People tend to interpret new, unfamiliar information based on their past experiences and knowledge [5].
  - (b) *Distinct.* Since users may need to examine and compare different portraits, the desired visual metaphors should help them easily distinguish one from another. This requires our design to capture the most salient traits of each person so that a user can recognize the unique characters of the person at a glance.

## 5 RESEARCH PROTOTYPES

We present two research prototypes to illustrate our design efforts alleviating the above challenges. These systems focus different aspects of the design challenges we outlined.



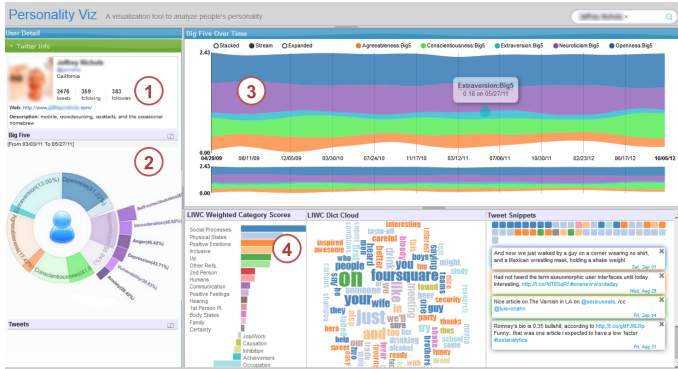


Figure 2: Big5Vis visualizes the Big 5 personality derived from tweets.

### 5.1 Big5Vis

Big5Vis is our first attempt along this line of work by visualizing the Big 5 personality traits derived from tweets. The system consists of four views, shown in 2, including (1) Basic information of a Twitter user; (2) An overview of the Big Five personality traits (including six sub-facets for each trait); (3) A timeline view of temporal changes of Big Five personality; (4) Visual evidence for the value of a personality trait or facet, including: a ranked LIWC category scores view (bottom left), a word cloud of LIWC dictionary (bottom middle) and tweet snippets related to the LIWC categories of interest (bottom right).

This system is an initial effort to visualize personality traits by considering the design challenges of supporting visual exploration (DC1) and explanation (DC2), but did not address the issues of enjoyability (DC3) and visual metaphor (DC4).

### 5.2 PETALS

To make a visual analytic tool of personality data enjoyable and engaging, we designed and developed PETALS, which uses a set of nature-inspired, interactive visual metaphors to summarize the derived traits and explain the associated linguistic evidence.

Given one’s tweets (Figure 3a), PETALS automatically derives one’s personality portrait (Figure 3b), which is visually summarized using a flower-inspired visual metaphor. Each portrait includes five types of derived personality traits, each of which forms a flower petal. The fullness of a petal indicates the person’s trait value in percentile compared to a test population. For example, 87% agreeableness means that the person is more agreeable than 87% of the population that we have analyzed<sup>1</sup>. A user can further examine the details of a particular trait and the linguistic evidence used to derive that trait, including the relevant tweets shown as *flower seeds* (Figure 3c), and word categories and words displayed along the *trailing vines* (Figure 3d).

#### 5.2.1 Personality Portrait as Flowers

Inspired by both classic Radial Space Filling (RSF) technique [19] and the beauty of a flower, we design a flower-based visual metaphor to illustrate an individual’s psychological portrait with a set of personality traits.

Since our data model represents a portrait as a tree internally, we use RSF for its initial layout. A RSF layout places the root node in the center of a circle, and distributes all the children along an arc sector (Figure 4a). The angular width of a non-leaf node is calculated based on the aggregated properties of all its children. We adopt the RSF technique for two reasons. First, it encodes a complex data structure in a compact form. Second, we can use the arc sector encoding a node to display additional information. For

<sup>1</sup>So far we have derived portraits for over 1 million people. All the trait percentile scores in this paper were computed against this population

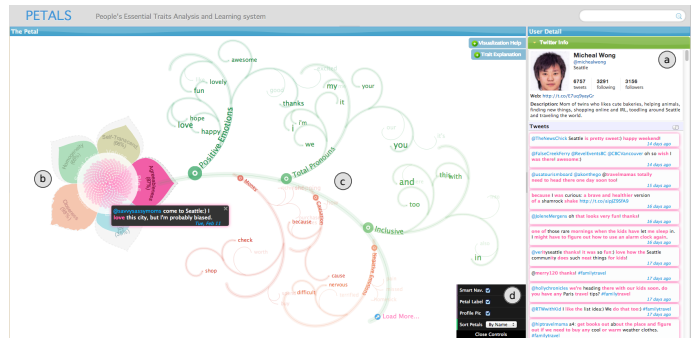


Figure 3: PETALS user interface and visualization design. (a) information panel of an individual who is being analyzed; (b) visual summary of this person’s portrait as a flower; (c) visualization of the linguistic evidence as a vine; (d) visualization control panel.

example, each flower petal encodes one type of trait (node) and additional information of the trait (e.g., meta data) (Figure 4b).

We use different visual elements to encode metadata related to each derived trait: trait type, trait value and accuracy of trait value.

The *trait type* is mapped to the color of a petal and use the “filled” area of a petal to encode the *trait value*: the higher the value is, the more filled the petal is. All petals have the same size (unfilled area shown in gray) to show the magnitude of each type of trait value. (Figure 5a). The *accuracy* of a trait value is encoded with the smoothness of a petal edge. A smooth edge indicates a high accuracy (a large amount of evidence used) in the derived value, and a jagged edge indicates otherwise (Figure 5b).

#### 5.2.2 Linguistic Evidence as Vines and Seeds

To help users understand and verify the derived personality traits, we visualize the linguistic evidence used to derive the traits. In particular, we visually encode three levels of evidence: related word categories, words, and original text.

##### Word Categories and Words as Vines

We use vines and their branches to encode multiple word categories and words that may be used to derive a trait, (Figure 6). Word categories are represented as the main branches along the trunk, and words are displayed as sub-branches under the associated category branch. The word branches are organized by their properties, such as frequency and semantic similarity.

*Visual Elements of Vines.* We also use various visual elements of a vine to encode additional information of the linguistic evidence. As show in Figure 6a, glyphs are used to represent two types of evidence: a hollow circle for a word category and a solid dot for a word. The glyph size encodes the correlational strength of the evidence with the derived trait, while the text font indicates the frequency of the evidence appears in the original text (tweets). The color of a branch indicates the type of correlation: green for positive correlation and orange for negative correlation.

##### Text Snippets as Flower Seeds

We use the flower seeds in the center of a flower to represent the original text (e.g., tweets) used to derive a particular trait (Figure 6b). The flower seeds are sorted by their time stamp. More recent tweets are placed toward the center while older ones appear near the edge of the circle. Each seed is colored the same as the associated trait.

## 6 DISCUSSION

The lessons we learned from two prototypes suggest some design considerations for visual personality analytic systems.

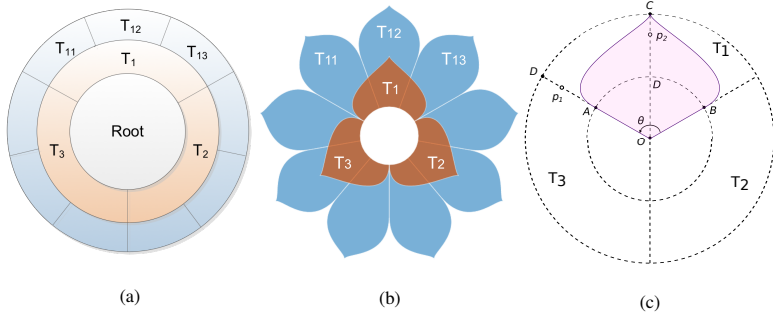


Figure 4: The visualization design of a portrait flower: (a) a RSF view of a sample profile; (b) a portrait flower view; (c) the petal primitive for the portrait flower view.

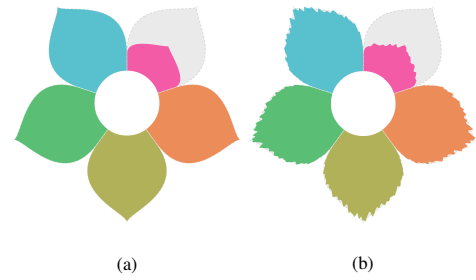


Figure 5: The visual encodings in a portrait flower: (a) the length of a petal to show a trait value; (b) the smoothness of a petal edge to show the value accuracy.

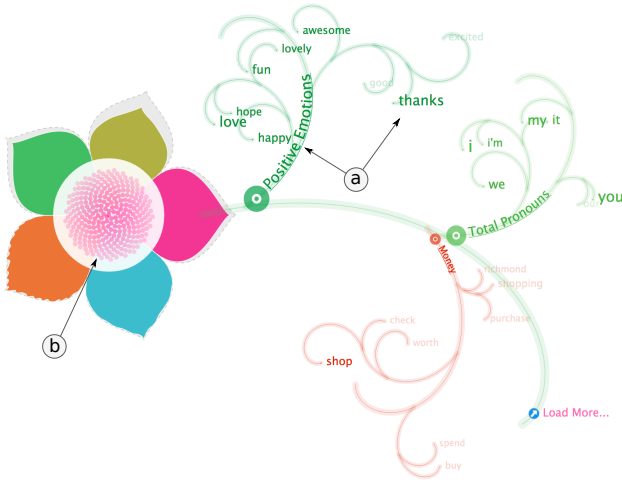


Figure 6: Visualization design for linguistic evidence: (a) categories and keywords as vines (b) text snippets as flower seeds.

### 6.1 Scalability

So far we have shown that the systems depicts five major types of personality traits. As the number of derived traits increases, we need to scale the systems to handle potentially an arbitrary number of traits or sub-traits. Since the flower metaphor is partly based on RSF, it can be scaled up to handle a large data set as a RSF does [19]. Specifically, the flower metaphor can “close” and “open” different petals encoding different traits, and almost infinitely add sub-traits as additional layers of petals. However, with the increased number of petals or layers of petals, not only will the display become cluttered, but there will also be limited space for displaying trait labels.

### 6.2 Balancing Aesthetics and Usability

During our design process, we had to constantly balance between the aesthetics and usability, as we struggled whether we should faithfully follow the visual features of a real flower or render just an abstract design. For example, we used different petal colors of a single flower to encode different types of traits to facilitate visual perception. However, such encodings are unnatural, as in nature a flower normally has the same colored petals. Another example is that we initially used the luminance of petal color instead of the “fullness” to encode trait value to maintain the shape of the flower. However, it is hard for people to perceive the luminance variations to distinguish different trait values. The vine layout with keywords makes the best use of visual metaphor but introduces visual complexity. So how should we balance between the aesthetics of the design including visual realism and the effective illustration of the information? To answer these questions, we need to conduct more

research to rigorously evaluate the pros and cons of different designs, especially their effects on users with different cognitive styles (e.g., analytic vs. intuitive).

### 6.3 Privacy Concerns

People may be concerned with the privacy issue. Although their tweets are publicly available, the derived information about their personality is sensitive and should not be disclosed publicly. Because of the sensitivity involved in personality analytics, a system must address various privacy concerns explicitly and support privacy-preserving visual analytics. One approach is to make the system more transparent to users by clearly explaining how their information is used to derive their personality traits and potential consequences of revealing such information. The above systems has made the first step in this direction by allowing users to examine the linguistic evidence used to derive a trait. As the next step, a system should allow users to explicitly mark which traits s/he is or is not willing to share and with whom. Based on user-specified preferences, the system can then decide how to visually disclose the intended information and hide others. Moreover, due to the imperfections in our personality analytics, users should also be allowed to annotate the derived traits and mark their own assessment (e.g., disagreement on the wrongfully derived traits).

## 7 CONCLUSION

In this paper, we identified the design challenges of visualizing personality traits derived from social media. We also reported the design and development of two prototype systems which help users to explore the derived traits and explain the associated linguistic evidence. We finally discussed future challenges for personality trait visualization with the lessons we learned from two prototype systems.

## REFERENCES

- [1] M. M. Bradley and P. J. Lang. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, 1999.
- [2] N. Cao, Y.-R. Lin, X. Sun, D. Lazer, S. Liu, and H. Qu. Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time. *IEEE TVCG*, 18(12):2649–2658, 2012.
- [3] S. Carpendale and M. Tory. Guest editors introduction: Personal visualization and personal visual analytics. *IEEE Computer Graphics and Applications*, 2015. 00000.
- [4] J. Chen, G. Hsieh, J. U. Mahmud, and J. Nichols. Understanding Individuals’ Personal Values from Social Media Word Use. In *Proc. of CSCW’14*, pages 405–414, 2014.
- [5] E. R. M. Cormac. *A Cognitive Theory of Metaphor*. Mit Press, 1990.
- [6] M. De Choudhury, S. Counts, and E. Horvitz. Predicting postpartum changes in emotion and behavior via social media. In *Proc. of SIGCHI ’13*, pages 3267–3276, 2013.
- [7] J. K. Ford. *Brands Laid Bare*. John Wiley & Sons, 2005.

- [8] J. Golbeck, C. Robles, M. Edmondson, and K. Turner. Predicting personality from twitter. In *Privacy, security, risk and trust in SocialCom '11*, pages 149–156, 2011.
- [9] L. Gou, M. X. Zhou, and H. Yang. KnowMe and ShareMe: Understanding Automatically Discovered Personality Traits from Social Media and User Sharing Preferences. In *Proc. of SIGCHI '14*, 2014.
- [10] L. Harrison, D. Skau, S. Franconeri, A. Lu, and R. Chang. Influencing Visual Judgment Through Affective Priming. In *Proc. of SIGCHI '13*, pages 2949–2958, 2013.
- [11] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, and S. Carpendale. Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 21(3):420–433, 2015.
- [12] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *J. Artif. Intell. Res.(JAIR)*, 30:457–500, 2007.
- [13] D. McDuff, A. Karlson, A. Kapoor, A. Roseway, and M. Czerwinski. AffectAura: an intelligent system for emotional memory. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pages 849–858, New York, NY, USA, 2012. ACM. 00021.
- [14] M. Pennacchiotti and A.-M. Popescu. A Machine Learning Approach to Twitter User Classification. In *ICWSM*, 2011.
- [15] J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, page 71, 2001.
- [16] R. Plutchik. The Nature of Emotions. *American Scientist*, 89(4):344, 2001.
- [17] T. Saaty and P. Kainen. *The Four-Color Problem: Assaults and Conquest*. McGrawHill, New York, 9, 1977.
- [18] S. H. Schwartz. *Basic Human Values: Theory, Measurement, and Applications*. Revue française de sociologie, 2006.
- [19] J. Stasko and E. Zhang. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *IEEE InfoVis 2000*, pages 57–65, 2000.
- [20] Y. R. Tausczik and J. W. Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. *J. of Lang. and Soc. Psy.*, 29(1):24–54, 2010.
- [21] J. Yang and Y. Li. Identifying User Needs from Social Media. Technical report, IBM Tech. Report, 2013.
- [22] T. Yarkoni. Personality in 100,000 words: A Large-scale Analysis of Personality and Word Use Among Bloggers. *J. of Res. in Personality*, 44(3):363–373, 2010.
- [23] J. Zhao, L. Gou, F. Wang, and M. Zhou. PEARL: An Interactive Visual Analytic Tool for Understanding Personal Emotion Style Derived from Social Media. In *IEEE VAST*, 2014. 00000.

# Designing Personal Visualizations for Different People: Lessons from a Study with Elite Soccer Teens

Thor Herdal, Jeppe Gerner Pedersen, and Søren Knudsen\*

University of Copenhagen

## ABSTRACT

This paper describes a study that sought to understand elite soccer teenagers' use of information visualizations to learn about their own sports performance, how this might motivate them to change behavior, and thus potentially improve their own performance. We specifically investigate how information visualizations support the players' data comprehension, and how their level of comprehension might depend on factors such as their general literacy, visualization literacy and maturity. We show unsurprisingly that elite soccer teenagers are able to use information visualizations to gain new information about their performance. Based on our investigation, we define a classification of the level of data comprehension. Secondly, we demonstrate a method which allows visualization researchers and practitioners to design and evaluate visualization concepts based on real data over a short time-span. Finally, we argue that designers of personal information visualizations need to consider the range of visualization literacies that are to be expected in many target populations. Our lessons provide new insights in peoples' use of visualizations and more broadly in visualization literacy. From these insights, we discuss implications for design of visualizations that consider peoples' level of visualization literacy. **Keywords:** Qualitative study, Interviews, Children, Teenagers, Visualization literacy, Sports visualization, Personal visualization.

## 1 INTRODUCTION

Much information visualization work focus on visualizations designed for people with great expertise within a particular field tasked with complex problems. These people are often discussed as domain experts in the visualization community. In personal visualizations [5], we expect that the people using our visualizations, might come from a wide range of backgrounds and social hierarchies. However, this means that there is a wealth of differences in the target population for our designs. In this paper, we aim to shed light on these differences, and explore what they might implicate for design of personal visualizations.

We do this based on findings from a qualitative study that inquired how elite soccer teenagers (boys, 13 years) might use information visualizations to learn about their own sports performance, how they might be motivated to change their behavior, and thus potentially improve their own performance. In the study, we identify different levels of comprehension between the teenagers.

We conducted the study in collaboration with the Danish Football Club FC Nordsjælland (in the following FCN) and Eye4talent. Eye4talent develops an application for soccer players to tag and evaluate personal performance data. Their application gathers data on performance specific parameters (e.g. passing, repress, finishing accuracy) and visualizes these according to the players' personal goals. The application uses the collected data to

partly show information visualizations of performance data and partly to show synchronized match video clips of the soccer field. FCN is currently evaluating the application with their youth teams. Teenagers on the team follow an elite program that combines school and training. They are highly motivated to increase their soccer skills, and so spend most of their leisure time with activities related to soccer. For example, they follow diets, watch soccer matches, play soccer console games, and reflect on their previous and upcoming matches. The teenagers have varied backgrounds, belong to different social classes, and live in different regions. They were scouted to FCN based on their soccer talent. Despite their different backgrounds, their talent and motivation is comparable. Thus, we argue that these differences enabled us to obtain a varied sample of study participants, which personal visualizations are expected to support.

Based on our qualitative study we present the following two contributions:

1. A three-level classification of people's level of data comprehension based on visualizations. We base this classification on our investigation of how information visualizations support the players' understanding of performance. The classification is introduced as a measure of analytic output, which we believe is related to the emerging visualization literacy concept [3]. We grounded this classification in data collected during design workshops, interviews and evaluations with teenagers who were motivated to improve their sports performance. Based on these findings, we suggest to a) consider a stronger focus on non-experts in designing visualizations and in talking about how they are used, and to b) identify peoples' level of understanding in visualization design, and to include people that represent the different positions along this dimension, both in designing and evaluating visualizations.
2. A method that allows early design work to be based on concrete domain-data in cases where no data exists prior to the visualizations. The method is based on study participants' own data collection. This facilitates quick and simple data collection, which allows visualization researchers and practitioners to go from concept to paper prototype evaluation within a day. The method seems particularly relevant for personal visualizations researchers.

## 2 RELATED WORK

We briefly review two strands of related work: First, we describe related sports visualization contributions. Next, we introduce work that considers visualizations in teaching or coaching situations.

### 2.1 Sports visualizations

Though people have created sports visualizations for decades, sports visualizations emerged as a subfield in information visualization and visual analytics recently. The first workshop on sports data visualization was held in conjunction with VisWeek 2013 [2]. Most importantly, Perin et al. [9] visualized soccer performance data in SoccerStories. SoccerStories provided interactive information visualizations of soccer matches, and displayed performance data according to the phases of the match

---

\* sknudsen@di.ku.dk

and visualized players' positions, movements and actions on the field. The program was intended for coaches and soccer analysts. Other contributions have also considered visualizations of soccer data (e.g., [8]). Pileggi et al. [10] visualized ice hockey matches in SnapShot. The visualizations were designed to support coaches, sports journalists, and talent scouts, in providing commentary and obtain data-based insights. While these previous contributions have focused on communicating sports data to analysts, we study how soccer players themselves might use visualizations to understand and improve their performance. This bares some similarity to other types of data logging situations in individual sports, which have become ubiquitous with the rise of smartphones (e.g., RunKeeper, EndoMondo, Nike+ Running).

## 2.2 Visualizations as part of coaching situations

We are unaware of sports visualization work that considers the communications between players and coach. However, we found similar interactions occurring in other domains. Paay et al. [12] showed the value of personal guidance in helping people quit smoking. Based on a prior study [11], they designed a mobile application which enabled people to track their smoking habits. Employees at a national smoking cessation service provided personalized weekly counseling based on data collected in the application, and thus observed a role similar to a coach. Gasser et al. [4] suggested to use social facilitation between people using mobile and web-based applications to help them manage nutrition and physical exercise activities. MacLeod et al. [7] described how patients used information visualizations to provide a way of understanding their condition within the framework provided by their doctors.

After this brief overview of related work, we continue to describe our choice of methodology.

## 3 METHOD

To study the elite teenage soccer player's use of visualizations, we set up a series of inquiries (see Figure 1). In conducting the inquiries, we recorded video or audio as appropriate to the form of inquiry. We analyzed the inquiry data with inspiration from Grounded Theory [14] and interview guidelines [6]. Next, we describe each step of our approach.

First, we conducted a first round of interviews with four key stakeholders: two team members, a coach, and the chief of talent development. We conducted these interviews to learn about the teenager's background and daily life, to obtain an overview of the goals in introducing visualizations to the elite teenage soccer players, and to establish rapport with the club management – necessary for the next step in our study.

Second, we conducted a workshop with the twelve players on the soccer team, to observe how the elite teenage soccer players used the existing solution to understand their own soccer performance. The team players watched a video of a match and tagged their own performance data using the existing mobile application. Based on the mobile application, which represented the performance specific data, we collected written summaries from the players, which they were given ten minutes to produce. The workshop led to the identification of differences in the team members' abilities to evaluate their performance, based on the current application design.

Third, based on the workshop, we conducted a second round of individual interviews with three team members, to learn more about the differences in their abilities. We identified team members which we suspected had different abilities in understanding their data. We based the selection criteria on the previously collected data. This is a form of theoretical sampling; we specifically looked for interview participants that had shown different analysis capabilities. The interviews allowed us to inquire about the team members' written soccer performance evaluations and thereby determine the player's

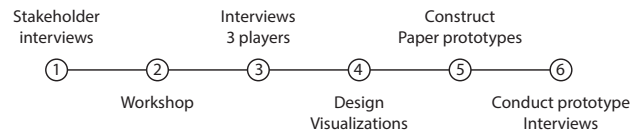


Figure 1: Study process.

level of comprehension. We performed opinion-categorization [6] of the player's ability to explain and reflect on their performance. Next, we categorized their statements in a) explanations based on data from the Eye4talent-application, and b) explanations based on the captured video. These interviews showed that the three players' level of comprehension fit in the three different categories.

Fourth, we chose to design a range of visualizations. We aimed to study the team members' abilities to understand their personal data based on these visualization designs. Using the interview insights, we created visualizations of the manually tagged data, to support the players' analysis of their own data from a football match. We sketched a range of visualizations as suggested by Tohidi et al. [15], which we condensed into two core design concepts. We based the design work on insights gathered during the previous inquiry steps.

Fifth, we selected two visualization designs, which we developed into interactive paper prototypes. We used these paper prototypes as the basis for a third round of interviews with the same three team members as previously, to once again inquire into their understanding of performance based on visualizations of personal data. Here, we were interested in observing and measuring visualization-based insights. To base these inquiries on concrete data, we repeated the data entry workshop with the team. The data for the three players that we interviewed were entered by other members of the team, to reduce potential learning effects.

Finally, we based the paper prototype interviews on concrete match data. Before the match, the interviewed players set personal goals, which we used for the paper prototype interviews. We did so to be able to correlate the players' expectations before the match to the visualizations afterwards. The manually tagged data allowed participants to compare their goals to their performance. Within a day, we collected the correlations and differences between player expectations and the tagged data in paper-prototype visualizations. We grounded and analyzed how the players used the visualizations to gain unexpected insights and how the visualizations supported their understanding of their performance.

Our study approach and intertwined analysis of the collected empirical data was inspired by Grounded Theory [14]. Throughout the study, we based decisions for next steps, on the findings emerging from analysis in the previous step. The guiding research question in the study started out as a wish to understand how the elite teenage soccer players might use of visualizations. However, we quickly identified comprehension level as an important concept, and became interested in pursuing this. By successive steps of inquiries with the team members, we aimed to saturate our empirical data about the players' information comprehension level.

## 4 FINDINGS

In the following, we present our findings in two separate sections: First, we describe our findings of differences in level of comprehension. Second, we describe how we examined how to support these different levels in the study. The primary source of these findings stem from the interviews and workshops conducted with the elite teenage soccer players.

### 4.1 Understanding Different Comprehension Levels

We analyzed the data collected from the initial workshop, in which the team members used an existing Eye4Talent application. This



application provided the players with a simple visualization design, which represented their performance data in bar charts.

By analyzing data collected at the initial workshop, we identified different levels of comprehension. At one extreme, the team members were only able to explain what they saw. At the other extreme, the team members were able to explain, interpret, and evaluate the perceived visualizations, and thus discuss potential changes in behavior, which might result in performance improvements. For example, a player that did not gain additional knowledge from the visualizations, said: *“I did four forward passes, which hit in the feet. One deep, and only one miss”*. When asked how he might improve his performance, he replied: *“there’s always room for improvement... [but,] I don’t really know how”*.

Another player read the visualizations easily. He described data about his performance with improvement proposals: *“I can see it in the numbers, I don’t have any loss of [ball] possession, because I am good at keeping the opponent away from me”*. When asked how to improve his performance, he replied: *“I need to keep training my core, because it gives me a physical advantage... also I need to be better at orientation, in order to have a better overview of where my opponent is located”*.

Naturally, we observed other team members between these two extremes. These team members were able to interpret and explain data, but did not provide suggestions for improvement. For example, when we asked a player whether his data showed good performance, he said: *“I see that I had six out of six represses. Due to my position, it’s important that I don’t lose the ball. It’s something that I have improved”*. Thus, while he was able to interpret and explain the performance data, he simply remarked that he had improved lately, and ultimately did not use the data to gather new insights. From the workshop data, we understood the varying levels of comprehension. Thus, we became interested in supporting these different levels, to help all team members to advance their data comprehension level. In the following we define these levels. First, we considered team members that provided only data explanations from visualizations, as being at comprehension level 1. Second, we considered team members that also provided explanations and interpretations of data as being at comprehension level 2. Finally, we considered team members that realized potential behavior changes to improve their performance, as being at comprehension level 3. This classification naturally serves to consider how we might support the various levels, to allow all to gain comprehension.

## 4.2 Supporting different comprehension levels

By introducing the team members to the prototype, we wanted to study how personal information visualizations might help enhance the team members’ abilities to understand their data, and thus increase their level of comprehension. We based the interviews on a paper prototype, which displayed visualizations of the individual players’ performance data for a match they played the day prior to the interviews.

We observed that the paper prototype supported a player at level 2 in evaluating his performance and contextualize it, in relation to previous and coming matches. For example, he structured his actions by color: *“I can see that I have a surprising high amount of red actions in the end of the match. This is because I am getting tired. I should have asked for a break”*, clearly considering how he might change behavior in the future, which he did not do previously. Thus, effectively, we believe that the improved visualizations helped him to better understand his performance data. In contrast, the other team members received only marginal comprehension benefits from the paper prototype.

The player at level 1 only slightly increased his understanding of the data. We believe this stemmed from confusion caused by the visualization. For example, he said: *“This one is yellow here*

*[points at his average performance], so it is displaying my repress like here [points at the timeline]. No, it doesn’t add up, I don’t get this”*. These confusions stemmed from our design choice, which we believe created a worldview gap [1] - a gap between what was being shown in the visualization and what needed to be shown, to support the players’ decision-making. The player did not gain help from the visualization, to distinguish a relation between the visualization and his experiences from the match. These misinterpretations made it difficult for him to conclude any improvements, and thereby he did not gain additional insights, in contrast to the player at level 2 described previously. The player at level 3 obtained most insights, and had few problems understanding the prototype. He used most of the prototype-based interview to draw connections between his memory of the match and the visualized data in the prototypes. He used colors to cluster his soccer actions and compare them to each other, and to his memory of the match. He also used a time visualization, to understand the phases of the match, in relation to his performance. His insights and his ability to compare his performance to the match resulted in a marginally enhanced level of data comprehension. This gave him new insights and suggestions for how to improve soccer performance. For this player, we observed a low worldview gap. We believe this stemmed from his ability to draw connections between the paper prototypes and how to improve his performance.

## 4.3 Implications

To create better premises of understanding for all comprehension levels, we suggest creating information visualizations, with a low complexity of initial visualizations. Each view has a collection of data, which summarizes the action points of a match. The players might then be able to explore additional data levels, in order to increase complexity and challenge them to find data insights.

After interviewing the team members based on the paper prototype, we observed that the gap between the levels of comprehension significantly rose compared to their previous interpretation of video data. We believe that the initial visualization design offered by the Eye4Talent application did little to support the team members’ analysis needs. We base this consideration on: first, that we noticed the players had problems elaborating on their results during the workshop, and second, that the comprehension level 2 player increased his understanding by using the visualizations during the final interviews. Going back to the early interview transcripts confirmed our suspicion. The interviews supported our belief and extended our understanding; the players explained their own performance more easily from video data, than from visualizations provided by the mobile application. All players gained a better understanding by looking at the collected video material, than by using the Eye4Talent application.

## 5 DISCUSSION

We acknowledge that our study is based on very few participants (one for each identified comprehension level). To learn about differences in comprehension, we carefully selected interview participants, who we believed represented different data comprehension levels. This allowed us to study how people at different levels might benefit from visualizations, and how to best support multiple levels in a system. Thus, our deliberate choice of participants built on theoretical sampling as argued by Strauss and Corbin [14]. With this approach, we defined comprehension level, by closely examining the player’s verbal expressions of understanding and the insights they gained during interviews.

Given our few samples, our findings should be considered as potential future directions, rather than as proofs. However, we still believe that we have shown, that with the right balance between visualization literacy and visualization complexity, all people might benefit from data visualizations.

This suggests to identify the range of visualization literacies within the group of people intended for a visualization design, and thus design for diversity within this group. Additionally, this suggestion points to the danger of considering domain experts in the design of personal visualizations. For addressing designs towards domain experts might take focus away from issues that might only appear for weak visualization readers. Comparing visualization literacy to textual literacy, designers should consider how they might lay out the text, such that both novice and experienced readers are able to take something with them. For example, in children's books, authors might provide gems for the parents. How might we do similar in visualizations? We believe that using interactions or gradually introducing complexity, might be answers to this, and obviously have been applied previously (e.g., [13]). At the same time, we acknowledge that other techniques might be fruitful. Next, we discuss our use of real data in paper prototypes, which we believe other personal visualization researchers and practitioners might benefit from.

### 5.1 Real data in paper prototypes

To ground the third round of interviews in concrete visualization designs, we constructed paper prototypes based on data from a training match played the day before the interviews. To collect match performance data, we asked individual players to tag their own or another players match performance data. The players used the Eye4Talent application, which had support for this task.

We believe that using real data minimized the worldview gap [1]. It gave the players an opportunity to explore and interpret a realistic representation of their personal data. If we had not used real data, the players might have had difficulties mapping the prototypes to their soccer performance. This might have caused difficulties in obtaining performance insights from the visualizations. By using real data, we are aware that players might have been able to express statements of their performance by memory, rather than using the visualizations. This is both an argument for basing our study on insights and an argument for the use of real data. The insights-based approach allowed us to focus on team players' unexpected discoveries, and the real data to use insights as a measurement to study the effectiveness of personal information visualizations. Additionally, the use of real data allowed the players to provide proposals for performance improvements based on the prototype. Therefore, as a motivational aid, we exposed the players to the prototype, while their memory was still fresh, and any insights they would obtain, still useful.

The paper prototype method, gave us the opportunity to quickly and easily create a design based on the team members' real data. We expected the players to contribute design improvements, but we found that they were far more interested in their performance, than the design. The Lo-Fi design resulted in some misunderstandings rather than explicit proposals of improvements. The rough edges of our paper prototypes confused some players, as they attempted to cluster separate data categories, which were drawn in the same colour. Despite the shortcomings of the paper prototype, our method was useful, and provided an efficient way to evaluate our personal visualization design concepts. In conclusion, we observed that the study participants focused on their personal data and how they could improve their performance based on the insights they obtained in the prototype.

We believe this approach to data collection might be valuable for other researchers or practitioners that aim to build paper or other Lo-Fi prototypes based on real data.

## 6 CONCLUSION

In this paper, we provided two main contributions.

First, we classified people's level of data comprehension based on visualizations, and described three main levels. At the first level,

people are able to read data from visualizations. At the second level, people use visualizations to understand previously collected data. At the third level, people consider behavior changes driven by data insights. We grounded our classification in data collected during design workshops, interviews and evaluations with teenagers who were motivated to improve their sports performance. Based on this, we suggest to first, reconsider the heavy focus on domain experts in designing visualizations and in talking about how they are used.

Secondly we suggest to identify peoples' level of understanding in visualization design, and to include people that represent the different positions along this dimension, both in designing and evaluating visualizations. Second, we described a method which allows early design work to be based on real data in cases where no data exists prior to the visualizations, and showed how this might be implemented in a concrete study. This method seems particularly relevant for personal visualizations researchers.

While many personal visualizations do not specifically aim to support children or teenagers, we still believe that the differences within this group possess relevant issues related to personal visualizations. We argue that it might even provide fertile grounds for studying differences in visualization comprehension and literacy, which seem more pronounced in this group of people.

## REFERENCES

- [1] R. Amar and J. Stasko, A knowledge task-based framework for design and evaluation of information visualizations, *INFOVIS 2004*, pp. 143–149, 2004.
- [2] R. Basole, E. Clarkson, A. Cox, C. Healey, J. Stasko, and C. Stolper, The 1st Workshop on Sports Data Visualization. [Online]. Available: <http://workshop.sportvis.com/index.html>. [Accessed: 31-Aug-2015].
- [3] J. Boy, R. a. Rensink, E. Bertini, and J. D. Fekete, A principled way of assessing visualization literacy, *IEEE TVCG*, vol. 20, no. 12, 2014.
- [4] R. Gasser, D. Brodbeck, M. Degen, J. Luthiger, R. Wyss, and S. Reichlin, Persuasiveness of a mobile lifestyle coaching application using social facilitation, *LNCS*, vol. 3962 LNCS, pp. 27–38, 2006.
- [5] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury, Personal Visualization and Personal Visual Analytics, *TVCG*, vol. 21, no. 3, pp. 420–433, 2015.
- [6] S. Kvale, *Doing Interviews*. SAGE Publications, 2008.
- [7] H. MacLeod, A. Tang, and S. Carpendale, Personal Informatics in Chronic Illness Management, *GI '13*, pp. 149–156, 2013.
- [8] C. Perin, J. Fekete, and R. Vuillemot, Table! Improving Temporal Navigation A in Soccer Ranking Tables, *ACM CHI '14*, pp. 887–896, 2014.
- [9] C. Perin, R. Vuillemot, and J.-D. Fekete, SoccerStories: A Kick-off for Visual Soccer Analysis, *IEEE TVCG*, vol. 19, no. 12, pp. 2506–15, Dec. 2013.
- [10] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko, SnapShot: Visualization to propel ice hockey analytics, *IEEE TVCG*, vol. 18, no. 12, pp. 2819–2828, 2012.
- [11] J. Paay, J. Kjeldskov, U. Brinthatparan, L. Lichon, S. Rasmussen, N. Srikandaraja, W. Smith, G. Wadley, and B. Ploderer, Quitty: Using Technology to Persuade Smokers to Quit, in *NordiCHI'14*, 2014.
- [12] J. Paay, J. Kjeldskov, M. B. Skov, N. Srikandarajah, and U. Brinthatparan, QuittyLink: Using Smartphones for Personal Counseling to Help People Quit Smoking, in *MobileHCI'15*, 2015, pp. 98–104.
- [13] E. Segel and J. Heer, Narrative visualization: telling stories with data, *IEEE TVCG*, vol. 16, no. 6, pp. 1139–48, 2010.
- [14] A. Strauss and J. Corbin, *Basics of qualitative research. Techniques and procedures for developing grounded theory*, 2. Ed. Thousand Oaks, CA: SAGE Publications, 1998.
- [15] M. Tohid, W. Buxton, R. Baecker, and A. Sellen, Getting the Right Design and the Design Right: Testing Many Is Better Than One, in *ACM CHI '06*, 2006, pp. 1243–1252.

# User-driven Expectation Visualization: Opportunities for Personalized Feedback

Yea-Seul Kim and Jessica Hullman  
University of Washington

## ABSTRACT

In this paper, we define and motivate Expectation Visualization, an interactive technique for soliciting, and presenting personalized feedback on, a user's expectation of the data. Expectation Visualization (EV) addresses the common challenge faced by designers of how to engage users with visualized data on a deeper level. We describe the design space of EV, including how it can be used for data encoded in marks and mark attributes, and describing forms of training and personalized feedback. We propose three specific applications where the benefits of EV may be particularly useful. We conclude with ideas for future research.

**Keywords:** Human-centered computing, Visualization, Information visualization.

## 1 INTRODUCTION

Traditionally interaction in visualization has been understood as user-driven manipulation of views of an existing data set: for example a user might select, sort, or filter a view [3]. Such interactions allow a user to generate views that are particularly relevant to him or her. As rich interactive visualizations become easier to author and more ubiquitous in media and other outlets, novel interactions, particularly those that make the data personally relevant to a user, are desired.

One form of "personalized" interaction that researchers have argued for is the manipulation by the user of internal (mental) representations as he or she makes sense of visualized data [10]. Relating an external visualization to one's internal representation, or mental model, can lead to better comprehension of gaps in one's knowledge [4][13]. For example mentally animating a set of small multiples showing a physics process, or comparing the small multiples to an internal representation, can lead to deeper understanding of the process [6].

Prompting users to engage in self-explanation, the process by which users explain its concept or example to themselves, may be useful as a means of guiding a user to compare internal representations to an external visualization [1]. As a form of prediction, self-explaining stimulates greater engagement with the topic. For example, it has been proposed that a system might prompt more active processing of data by asking a user to guess the direction of the trend prior to viewing the data [7].

In most examples studied in psychology [4][6][13], the benefits of self-explaining and internal visualizations come only after considerable cognitive work on the part of the user, who must mentally imagine the difference between their internal representation and the external visualization. In this work, we consider a new possibility: What if visualizations allowed people to draw their expectations of the data prior to viewing? Seeing the data along with the expectation provides a form of personalized



feedback, as it renders this gap between expectation and fact explicit. The act of drawing maintains a user's engagement, while the visual representation of the expectation against the result allows more detailed observations than may be possible through mental visualization alone.

In this paper, we define Expectation Visualization (EV), an interactive technique for soliciting and presenting personalized feedback on a user's expectations of the data. Expectation Visualization (EV) addresses the common challenge faced by designers of how to engage users with visualized data on a deeper level. In the rest of the paper, we describe the design space of EV including how EV could be applied to common visualization tasks and visual encodings. We propose specific types of applications in which the cognitive benefits of EV could be useful. We conclude by offering ideas for future work.

## 2 EXPECTATION VISUALIZATION DESIGN SPACE

We use a recent New York Times interactive graphic to motivate the design space of EV. We define the design space according to tasks and visual encodings. We discuss forms of training and presentation of personalized feedback.

### 2.1 Predicting Marks

We begin our characterization of the design space for EV by considering a recent New York Times graphic of this technique (Fig. 1). The interface presents a XY plot without any data shown, however the axes are labeled as *Parent's income percentile* and *Percent of children who attend college*. The user is encouraged to draw their guess for each income level in the chart (Fig. 1a). In the accompanying text, definitions of various visual trends are provided to help users to relate the graphical representation to their expectations (e.g.,  or ). After the user is done drawing the line, the true trend is presented as an overlay on the chart showing the user's expectation (Fig. 1c). Statistics describing how the user's guess compares to those of other users are presented in text below the chart.

In this example, the user's mental model is represented by a trend line. Two continuous variables are being considered, so a simple labeled XY plot is sufficient for the drawing interface. However other visual encodings and data types require different drawing interactions and interfaces. We envision how EV can be applied to other tasks and encodings in Table 1.

### 2.2 Extension to Mark Attributes

Visualizations are composed of marks (e.g., a bar in a bar chart, a circle in a scatterplot) and mark attributes (e.g., size, shape, color). While the above example provides interactive support for visualizing one's expectations by adding marks, it may also be possible to support interactive prediction of a variable, which will be encoded as a mark attribute. For example in a choropleth map, a user brush on color to predict the trend in a region for a continuous variable. In these examples, marks are presented and the user interacts to add value expectations (Table 2).

---

\* yeaseul1@uw.edu, jhullman@uw.edu



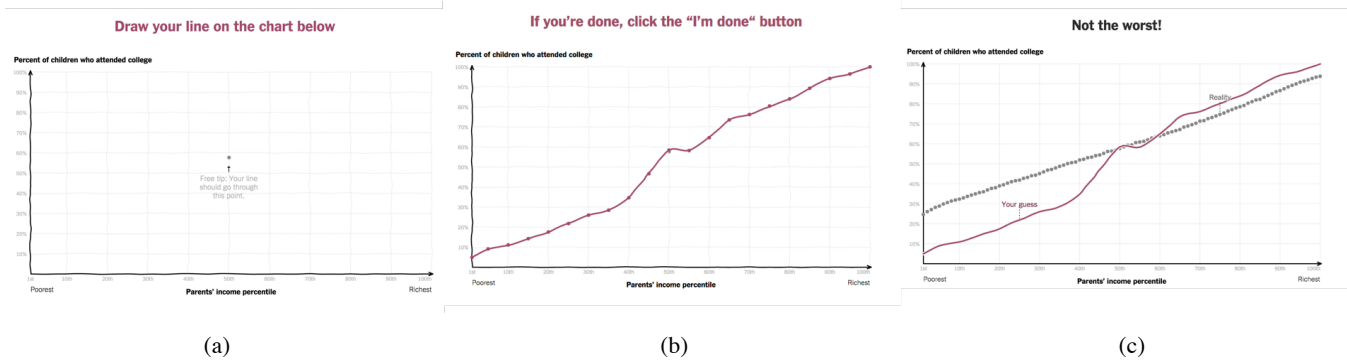


Fig. 1. A screenshot of 'You Draw It: How Family Income Predicts Children's College Changes'<sup>1</sup>.

Table 1. Possible tasks and visual encodings for predicting marks.

Task	Drawing task	Encoding
Comparison between means	Draw bars, dots	Bar chart or dot plot
Uncertainty	Draw confidence interval, coverage interval	Bar chart or dot plot with error bar
Analyze a distribution	Draw probability density function, draw quartiles, annotate expected outliers	Histogram, violin plot
Identify clusters	Draw contour, drag and drop element to the cluster	Scatter plot, dendrogram, Node-line diagram
Judge correlation/fit	Draw line of best fit	Scatterplot
Analyze connectivity	Draw edges	Node-line diagram

Table 2. Possible tasks and visual encodings for predicting mark attributes.

Task	Drawing task	Encoding
Predicting values of variables	Brush on color, drag symbols to size	Map (choropleth graduated symbols), 3D+ scatter plot
Predicting categorical membership	Brush on color, draw shape	Bar chart, scatter plot

### 2.3 Training

We anticipate that EV could be especially useful for non-expert analysts, such as news readers and students. However, it might be necessary to train such users by providing a “drawing vocabulary” that relates the expectation to the visual encodings. This vocabulary consists of sample marks and marks attribute along with definitions of what the example element represents.

For example, for the visualization presented in Figure 1, sample trend lines were given in the text (e.g. means the chances are the same for everyone). Another way to support novice users is by providing annotations directly on a plot structure to describe how to interpret the graph schema (e.g., indicating what data relationships different regions of an XY plot represent).

### 2.4 Forms of Personalized Feedback

Presenting the real data against the expectation in the same view, such as through an overlay, directly shows the user gaps between the facts and their expectation. However, personalized feedback on the user's expectation can also be presented in the form of a comparison to previous users. For example, an overlay on the

user's drawing might instead depict aggregated estimates from other users. Personalized annotations can be added directly on marks that the user predicted which vary significantly from the true data or from those derived from aggregated user data. The discrepancy between the expectation and the true data can also be quantified and presented with a description of how much the user should adjust their internalized model to accurately perceive reality.

## 3 POTENTIAL BENEFITS OF APPLICATIONS

We expect EV to facilitate deeper understanding of graphical format. Based on the more active processing of the data that prediction can stimulate, we also expect EV could increase engagement and memorability of data and concepts. Furthermore, by visualizing the discrepancy of a user's expectation and providing personalized messages, EV could be used to motivate users to change their behaviour in certain ways.

### 3.1 Deepening understanding of data and graphical schema

The canonical self-explanation approach induces active processing that stimulates deeper understanding of a concept by showing a user a representation and then prompting the user to explain it to herself [1]. In our adaption of self-explanation in EV, the user makes her prediction before seeing the visualization.

We expect that predicting before seeing the visualization will cause novice users to devote more attention to the data and graphical schema. This may result in better understanding of the semantics of reference structures of a graph (e.g., the axes, legend, etc.). It could be useful for understanding units, time span or other domain specific aspects of a dataset.

Other aspects of the data schema including the nature of the variables, or the difference between observed data and a model derived from the data, may also be better understood depending based on how the drawing interaction is realized. For example, drawing a line to visualize their expectation of a trend may help the user to perceive the variable as continuous if the line drawing interaction is smooth, and to understand that their drawing represents a model instead of a set of data observations. If the system requires the user to draw a line as a discrete set of points, this might help the user understand that the true data is a set of measurements, not a model.

<sup>1</sup> <http://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html>

### 3.2 Engagement and Memorability

We expect EV to have benefits for increasing a user's engagement with the data and increasing the memorability of the data. The act of drawing is still relatively novel compared to typical point-and-select interactions, and therefore likely to engage users' interest. This "more natural" form of interaction may be particularly attractive to novice users who are new to data analysis.

Active processing, particularly via self-testing exercises [2], can improve memory for information [7]. By presenting personalized feedback, EV makes the information more relevant to the user, which is also likely to benefit memory [8].

### 3.3 Behavioral change

Personalized messaging is an effective way of motivating one to make a desired behavioral change [9][12]. Imagine a chart in a news article describing how temperatures in the world have risen over the last century. The chart is intended to deliver information, but also to urge readers to change their behaviors that affect global warming. By using EV, the feedback presented after the user draws their expectation provides individualized messaging that directly measures their awareness of the problem. We expect that seeing the gap between reality and their expectation will make a user more likely to change their behaviour than if they had only seen the data.

## 4 FUTURE WORK

There are multiple important next steps for research and application of EV. An important next step is to further specify how EV interactions should differ for different types of visualizations and data. This may also include identifying automated techniques that can be used to more easily compare users' drawings to a data schema. For example, if the variable the user is predicting is encoded in mark shape, a computer vision algorithm may be needed to classify the shape the user draws.

One important next step is to provide experimental evidence of EV's benefit for visualization comprehension. A number of questions arise regarding how particular applications of EV will influence a user's comprehension. For example:

- How should the true data be presented against the expectation?
- How often should dynamic feedback be presented?
- What forms of training are necessary for allowing novice users to use EV applied to different visualizations?
- How useful is personalized feedback compared to the true data vs. compared to other users' expectations?
- To what extent do the benefits of EV stem from the act of drawing vs. the act of prediction?

We are beginning to develop an interactive prototype for applying EV to different chart types in order to answer these questions and others.

## 5 CONCLUSION

We define and motivate Expectation Visualization, an interactive technique for soliciting and presenting personalized feedback on a user's expectation of a data set. We describe the design space of EV including application to mark and mark attributes and describing forms of training and feedback. We propose three specific applications where the benefits of EV may be particularly useful. We conclude with a roadmap for future research of EV.

## REFERENCES

[1] M.T.H. Chi, Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in Instructional Psychology*, Hillsdale, NJ: Lawrence Erlbaum Associates 2000.

[2] G. O. Einstein, H. G. Mullet, and T. L. Harrison, The testing effect: Illustrating a fundamental concept and changing study strategies. *Teaching of Psychology*, 39(3), 190-193, 2012.

[3] J. Heer, and B. Shneiderman, Interactive dynamics for visual analysis. *Queue*, 10(2), 30, 2012.

[4] M. Hegarty, Representations in the Mind and in the World: How Cognitive Science Can Inform the Design of Visualizations, Keynote presentation, *INFOVIS '10*, 2010.

[5] M. Hegarty, S. Kriz, & C. Cate, The roles of mental animations and external animations in understanding mechanical systems. *Cognition & Instruction*, 21, 2003.

[6] M. Hegarty, Mental animation: inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1992.

[7] J. Hullman, E. Adar, and P. Shah, Benefitting infovis with visual difficulties. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12), 2213-2222, 2011.

[8] J. M. Keenan, and S. D. Baillet, Memory for personally and socially significant events. In Nicherson, R. S. (Ed.). *Attention and Performance*, VIII. Hillsdale, NJ: Erlbaum, 1980.

[9] M. W. Kreuter, D. W. Farrell, L. R. Olevitch, and L.K. Brennan, *Tailoring health messages: Customizing communication with computer technology*. Routledge, 2013.

[10] Z. Liu and J. T. Stasko, Mental models, visual reasoning and interaction in information visualization: a top-down perspective, *IEEE TVCG*, vol. 16, no. 6, 2010.

[11] H. M. Natter and D. C. Berry, Effects of active information processing on the understanding of risk information, *Appl. Cog. Psych.*, vol. 19, no. 1, 2005.

[12] C.S. Skinner, V.J. Strecher, and H. Hospers, Physicians' recommendations for mammography: do tailored messages make a difference? *Amer. J. of Pub. Health*, vol. 84, no. 1, 1994.

[13] J. G. Trafton, S. B. Trickett, and F. E. Mintz, Connecting Internal and External Representations: Spatial Transformations of Scientific Visualizations. *Foundations of Science*, vol. 10 no. 1, 2005.

# A Visualization Fingerprint: Comparing Nutrient Data Visually

Carrie Mah\*  
University of Calgary

Richard Pusch†  
University of Calgary

Samuel Huron‡  
IRI Centre Pompidou  
University of Calgary

Sheelagh Carpendale§  
University of Calgary

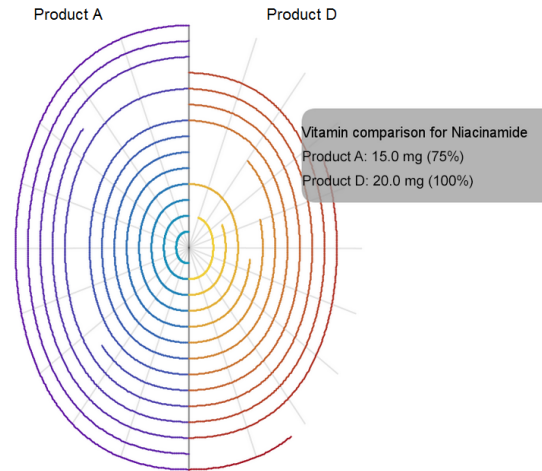


Figure 1: A prototype for the Fingerprint which represents data from Figure 3. Each vertical side represents the nutritional data for a food item. The arcs from inside to outside correspond to the data set from top to bottom. Each arc represents a single nutrient, and a full arc indicates a higher value. Selecting an arc provides more information, such as the nutrient's name and amount in each food item.

## ABSTRACT

We introduce a visualization which has been designed to create a Fingerprint of a product, and have applied this to the nutritional data in food items. This visualization allows customers to compare the data in a Nutrition Facts Panel between two different food items. It is becoming increasingly common for people to seek information so they can make more informed decisions about their diet. One such method is by reading the Nutrition Facts Panel on a product's packaging. However, the information presented is inconsistent between products, and thus, is difficult to compare. To help people make the necessary comparisons, we created the Fingerprint, a visualization designed to help customers find suitable food items by enabling direct comparison of nutritional data in food items.

**Keywords:** Personal visualizations, nutrition visualization, health, nutrition, nutrient data, nutrition facts

**Index Terms:** H.5.m [Information Interfaces & Presentation]

## 1 INTRODUCTION

We introduce Fingerprint, a visualization that has been designed to support comparisons and make it easier to recognize similarities and differences between products. Customers seek information on the nutritional data of food items and look for tools that help them make better decisions about their diets [1]. One of the deciding factors when selecting food items is knowledge about what nutrients are contained in them. Currently, one of the main sources of

this information is a food item's Nutrition Facts Panel [2, 3]. However, the information is presented in inconsistent ways; one label might refer to ounces, another to grams, and another to the percentage of daily recommended amounts for a given nutrient. As a result, customers have to translate between different measurement units, which is tedious and error-prone. By comparison, presenting information visually has the potential to increase comprehension, recollection, and retention [4]. We visualize the nutritional data to draw upon natural human abilities to process visual information [5].

Fingerprint is a visualization that has been designed to compare multiple aspects of two products; specifically, we apply this to the nutritional data in food items. The visualization presents data in a manner that resembles actual fingerprints (see Figure 1) by modifying bar charts into curves and making use of symmetry to represent two products. It organizes the data in Nutrition Facts Panels in a consistent way that visually identifies food items and helps consumers find suitable food items based on its nutritional data. This Java and Processing application is designed to ease the task when comparing the nutritional data of food items.

## 2 RELATED WORK

Our work has strong ties to personal visualizations. Personal visualization explores the creation of visualizations about data that we encounter in our personal lives [6]. Many types of tools have been designed to help people with the data in their personal lives. A few of these that include some visualization are: providing feedback in home utilities consumption [7], visualizing our musical listening habits [8], working with personal data such as exercise logs to help people understand themselves and make better decisions in their lives [9], exploring options to help people make greener transportation choices [10], and providing feedback for water usage [11]. Visualization tools are being introduced to many aspects of life such as River, which is a tool that visualizes personal activity data [12], and Visits, which visualizes travel histories as personal movement data for the purpose of reminiscing and sharing life ex-

\*e-mail: cmah@ucalgary.ca

†e-mail: rapusch@ucalgary.ca

‡e-mail: samuel.huron@cybunk.com

§e-mail: sheelagh@ucalgary.ca

periences [13]. More and more commercial tools are also available to help people keep track of their activities, such as Fitbit [14] and Nike+ [15], and their nutritional intake, such as Fooducate [16] and My Fitness Pal [17] to reach their health goals [18].

There are visualizations that display the nutritional data about food items. Often these visualizations and tools use a large database which can overwhelm a consumer and displays data in conventional rectangular bar charts. For example, DesignFood [19] is a photo book that arranges food items proportional to each item's calories, carbs, fat, and protein values. This book is limited to a few meals and does not compare items directly with each other as many items use different measurement units. Other tools include The Nutritional Understanding Tool (NUT) [20], a visualization of the nutritional data in whole food items that helps people decide what to eat by supporting comparisons between different food items to help people find food replacements. Newtrition [21] compares meals based on carbs, fat, and protein values. They also represent caloric intake for a meal and map it to equivalent amounts of physical activity. Visualizing Nutrition [22] recommends food items by ranking them based on specific amounts of nutrients the food item contains.

Other visualizations have been described as being "fingerprints", a visual mark used for identification [23, 24, 25, 26]. The Cinema Redux [23] takes movie frames every second and presents it as one image. Dawes describes the composition as a visual fingerprint, as it provides an impression of the movie's color hues and the rhythm in the editing process. Similarly, Cinematics [24] creates rings that interpret and compare movie characteristics using segments that contain color, speech, and movement during a shot. Digital City Portraits [25] are sunflower-shaped visualizations, which are described as visual fingerprints that enable people to identify cities based on their citizens' social media posts. The designer's collections in the New York Fashion Week that were visualized in minimalized color-blocked layers [26] are also referred to as fingerprints. Hinton et al. envisions these as fingerprints that present the colors of dresses and the order in which they appear [26].

### 3 DESIGNING THE FINGERPRINT VISUALIZATION

Our long range goal is to develop an application that a customer can use to compare two or more products while they are shopping in a store. There are many challenges to this, such as retrieving the data from the product bar code, or QR code if they exist, translating data to comparable units, as well as creating a suitable visualization. In this paper, we discuss the design of our prototype visualization, Fingerprint. We worked with nutritional data from multivitamins, and restricted our visualization to nutrient data and comparing two products. We start by describing the data we worked with.

#### 3.1 Nutrient Data for Comparison

It is common for food items to have a Nutrition Facts Panel, which lists information such as nutrient amounts per a specified serving. However, the data between packages is presented in inconsistent ways. As seen in Figure 2, red the nutrient values differ in measurement units between two multivitamins. This makes direct comparisons difficult, even between similar products. Labels may also differ in other ways, such as serving sizes.

Vitamins:		EACH TABLET CONTAINS: VITAMINS	
Vitamin A (Acetate)	900 mcg RAE/3,000 IU	Vitamin A (vitamin A acetate)	300 mcg/1000 IU
Beta-carotene	600 mcg	Beta-Carotene	600 mcg/1000 IU
Vitamin B1 (Thiamine mononitrate)	2.5 mg	Vitamin E (dl- $\alpha$ -tocopheryl acetate)	12.6 mg/26 IU
Vitamin B2 (Riboflavin)	2.5 mg	Vitamin C (ascorbic acid)	150 mg
Vitamin B6 (Pyridoxine HCl)	25 mg	Folate (folic acid)	0.4 mg
Vitamin B12 (Cyanocobalamin)	25 mcg	Vitamin B1 (thiamine mononitrate)	3.85 mg
Vitamin C (Ascorbic acid)	150 mg	Niacinamide	14 mg
Vitamin D3 (Cholecalciferol)	15 mcg/600 IU	Vitamin B6 (pyridoxine hydrochloride)	5 mg
Vitamin E (d-alpha tocopheryl acetate)	10 mg AT/15 IU	Vitamin B12 (cyanocobalamin)	21.6 mcg
Biotin	15 mcg	Vitamin D (cholecalciferol)	20 mcg/800 IU
Folate (Folic acid)	400 mcg	Biotin	45 mcg
Niacinamide	25 mg	Pantothenic Acid (calcium d-pantothenate)	11 mg
Pantothenic acid (Calcium d-pantothenate)	25 mg	Vitamin K1	20 mcg

Figure 2: Comparison between two multivitamin products. Nutrient values are displayed inconsistently with different measurement units.

Data Chart Sheet

LEGEND	Product A		Product B		Product C		Product D	
	mcg	mg	mcg	mg	mcg	mg	mcg	mg
Data provided								
Data converted								
Data unavailable								
VITAMINS								
Beta-Carotene (source of Vit A)	1,500	1,500	3,000	0	0	300,000	0,300	1,500
Biotin	45,000	0.045	0	5,000	0.005	200,0	0	0
Folate (folic acid)	600,00	0.600	0	100,00	0.100	400,00	0.400	400,00
Niacinamide	15,000	15,000	0	0	0	20,000	20,000	14,000
Pantothenic Acid (calcium d-pantothenate)	10,000	10,000	0	0	0	10,000	10,000	10,000
Vitamin A (vitamin A as acetate)	300,00	0.300	1,000	450,00	0.450	1,600	1,200	4,000
Vitamin B1 (thiamine mononitrate)	2,250	2,250	750,0	0	0	2,250	2,250	750,0
Vitamin B2 (riboflavin)	3,200	3,200	1,280,0	0	0	2,550	2,550	1,020,0
Vitamin B6 (pyridoxine hydrochloride)	5,000	5,000	4,1135	2,000	2,000	1,6454	3,000	3,000
Vitamin B12 (cyanocobalamin)	20,000	0.020	0	6,000	0.006	0	9,000	0.009
Vitamin C (ascorbic acid)	90,000	90,000	1,800	20,000	20,000	400,0	90,000	90,000
Vitamin D (cholecalciferol)	15,000	0.015	600,0	0	0	20,000	0.010	400,0
Vitamin E (dl- $\alpha$ tocopheryl acetate)	22,500	22,500	50,00	4,500	4,500	10,000	20,250	45,000
Vitamin K1	25,000	0.025	0	0	0	0	0	0
MINERALS								
Calcium (calcium carbonate and phosphate)	200,000	200,0	0	0	0	0	0	400,000
Chlorine (potassium chloride)	0	0	0	0	0	34,000	34,000	0
Chromium (chromium chloride)	388,000	0.0388	0	0	0	350,000	0.150	0
Copper (cupric sulfate)	1,000	1,000	0	0	0	2,000	2,000	50,000
Iodine (potassium iodide)	150,000	0.150	0	0	0	150,000	0.150	0
Iron (ferrous fumarate)	10,000	10,000	0	0	0	0	0	10,000
Magnesium (magnesium oxide)	50,000	50,000	0	0	0	100,000	100,000	0
Manganese (magnesium oxide)	5,000	5,000	0	0	0	3,500	3,500	0
Molybdenum (sodium molybdate)	45,000	0.045	0	0	0	42,000	0.042	0
Potassium (potassium chloride)	80,000	80,000	0	0	0	37,500	37,500	0
Selenium (sodium selenite)	150,000	0.015	0	0	0	37,500	0.0375	0
Zinc (zinc oxide)	7,500	7,500	0	0	0	15,000	15,000	0

Figure 3: Data Chart Sheet used for the Fingerprint prototype.

To obtain meaningful data for input to our visualization, we selected four different multivitamins and converted the data to share the same units and serving sizes between items (Figure 3). Choosing multivitamins as our data sources allowed us to compare the nutritional differences between similar products, a common task by consumers. Also, our visualization supports the wide range of diverse data points often found in multivitamins.

#### 3.2 Designing a Product Visualization

We consider the problem of designing this visualization around the context of a consumer shopping in a store, wanting to compare two products found on a shelf. Framing the problem in this way led us to develop three main design factors:

1. Represent nutritional data in an attractive manner, easily viewable and browsable on a mobile phone
2. Be able to easily compare nutritional data between food items at a glance
3. Help consumers select food items with the nutritional data that best meets their current goals

##### 3.2.1 Design Iteration 1

We initially focused on using visualizations with simple shapes, because we wanted our visualization to be easily understood. We considered using bar charts, as they have become increasingly more

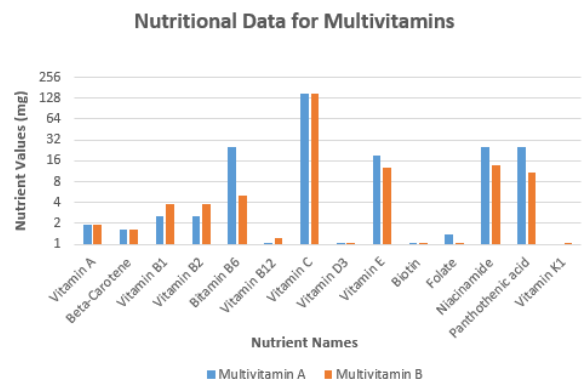


Figure 4: A double bar chart makes it difficult to compare products as a whole. The vertical axis in this chart uses a logarithmic scale.



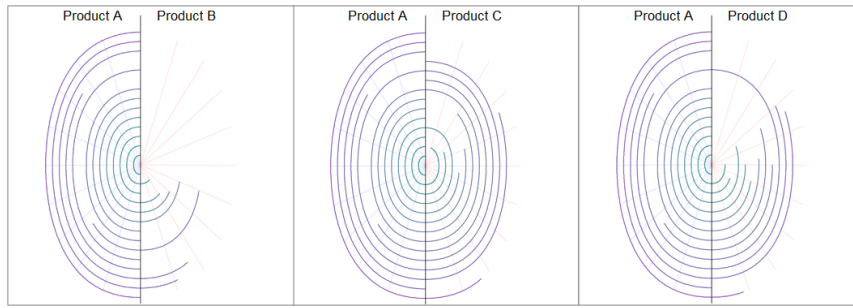


Figure 5: Comparing one multivitamin (left half) to three different multivitamins.

common and can be read by many people. However, some people struggle interpreting data from bar charts [27] and they present some unique challenges when viewing high density data. In particular, a double bar chart (see Figure 4) can be used to compare two similar sets of data, but at a glance, is difficult to draw conclusions as to how the data sets compare as a whole. On a small screen, the problems compound; a bar chart presented on a landscape aspect ratio, allows the display of many variables along the horizontal axis but with less vertical space. This results in making it harder to perceive minor differences in values. When viewing bar charts with a portrait aspect ratio, we noticed that curving the bars allowed for additional length. The elliptical curves resembled a fingerprint, an attractive shape that colloquially means a method of identification; each product can then have a distinct fingerprint associated with it.

We represent fingerprints in our system as a collection of B-spline curves. As seen in Figure 7, we modified the number of B-spline control points and their coordinates to create several possible designs that resembled a fingerprint. We moved forward with the elliptical prototype (Figure 7 right), as it filled the most space on the small screen and most resembled a human fingerprint.

### 3.2.2 Design Iteration 2

There are several ways our fingerprint design could visualize food item information. For example, each multivitamin could be represented as its own complete fingerprint, with each nutrient mapped to a curve, and the arc's length representing the proportion of that nutrient to the maximum value in the multivitamin. These multivitamins can have very different nutrient values, but if the ratios of multivitamin nutrients are similar, the differences between two fingerprints can be difficult to distinguish (Figure 6). In addition, comparing two fingerprints side-by-side means either compressing them to fit on half the screen or using two screens. Both alternatives meant that at-a-glance viewing does not lead to many meaningful conclusions about how the two products differ.

Another approach is to alternate the drawing of nutrient curves from two fingerprints moving outwards, combining the visualization into one large entity (see Figure 8 C). The final result is a sin-

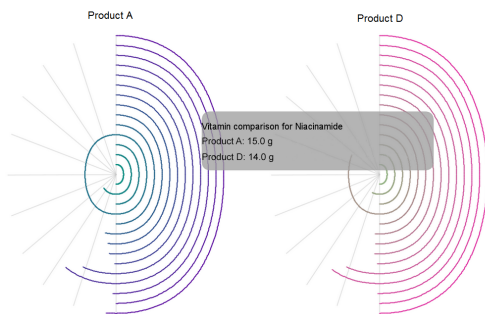


Figure 6: Comparing two food items. Calculations are based on nutrients within the item.

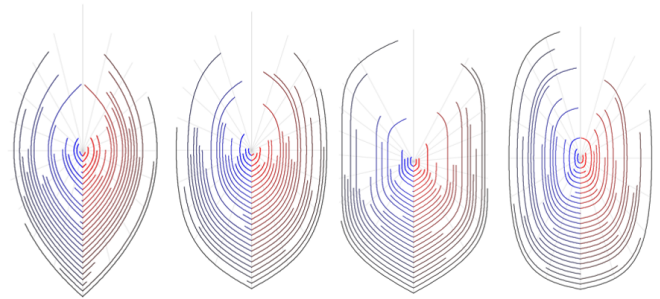


Figure 7: By varying the B-spline properties, we can change the fingerprint's overall shape.

gle fingerprint with twice as many curves as nutrients. This suffered from two main flaws: it was very difficult to parse at a glance, and direct comparisons between the same nutrient could not easily be made, as one B-spline curve would be bigger than the other.

### 3.2.3 Design Iteration 3

Our final design splits the fingerprint into two vertical halves, each representing a product (see Figures 1, 5), with zero beginning at the bottom of the center line. This symmetrical, side-by-side presentation, increases the contrast of nutrients between products, allowing for differences to be distinguished in a shorter amount of time [5].

To ease comparison between two food items, tooltips were used to display details of the nutrients, such as their name and exact amount, and used the visualization for relative comparisons. A given nutrient starts at the same height on the central zero line for each half of the visualization and curves to the left for one product (Product X) and to the right for another (Product Y). To measure

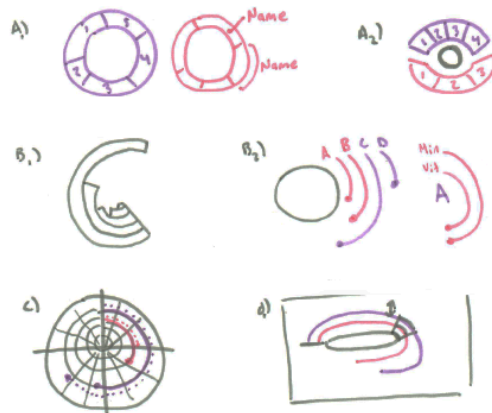


Figure 8: A series of sketches showing the progression of different arc designs.

the length of these curves, we used relative amounts. The larger of the two values is a curve that extends all the way to the vertical line at the top, while the smaller value has a curve with length proportional to the larger value. For example, if X has a nutrient with 100 mg and the same nutrient in Y has 35 mg, this nutrient's curved bar will extend all the way to the top central divider for X and Y's curve will only extend 35%. The consumer can then use a pointer to select an arc and see more specific information about a nutrient via a pop-up dialog. Reading guidelines were also created to indicate the radial axis in 10% increments. To aid with at-a-glance comprehension, a different color range was selected for each individual food item. Each nutrient within the food item is represented by a unique shade. For example, Product X uses shades of blue and Product Y uses shades of red (see Figure 1).

This design offers several benefits. It does not require extra screen real estate and the fingerprint does not get more dense, as the total size and shape of the fingerprint is unchanged from Section 3.2.2. It remains easy to get an at-a-glance summary of a product's values, as each product still has an identifiable "half-fingerprint". Lastly, it leverages our ability to quickly notice differences in symmetry to compare products at a high or low level of detail [5].

#### 4 FUTURE WORK AND CONCLUSIONS

We have presented Fingerprint, a visualization that we developed to present a characterization of a product. We used it to present vitamin nutrient data details in multivitamins. It can support the comparison of two similar products, showing relative differences visually and providing specific details via tooltips.

There are many opportunities for future work. Perhaps the most important is to enable it for live real world usage. However, this would mean that product components would need to be available via a bar code. Currently details available through the bar code include varying factors such as price, suppliers and manufacturers. Alternatively, a visualization such as the Fingerprint could be included on the product package.

Another possibility for future work could include the exploration of alternate visualization designs. There are still many opportunities such as other circular or elliptical patterns [24, 25], or radial parallel coordinates [28]. Also, further explorations into the use of color would be of interest, and integrating a thumb activated pointer instead of simple tap for revealing tooltips would be important.

We have applied Fingerprint to food items, but this idea could be extended to represent other products. In the future, the Fingerprint can be improved and integrated in a variety of ways to help consumers make choices amongst different products.

#### ACKNOWLEDGEMENTS

We wish to thank David Ledo, Riane Vardeleon, Brennan Jones, David Tang, and Mike Clark. This research was supported in part by AITF, NSERC, GRAND, Surfnet, and SMART Technologies.

#### REFERENCES

- [1] M. Pollan. *In Defense of Food: An Eater's Manifesto*. Penguin Books, first edition, 2009.
- [2] J. L. Blistein and W. D. Evans. Use of nutrition facts panels among adults who make household food purchasing decisions. *JNEB*, 38(6):360–364, 2006.
- [3] J. Barreiro-Hurlé, A. Gracia, and T. de Magistris. Does nutrition information on food products lead to healthier food choices? *Food Policy*, 35(3):221–229, 2010.
- [4] L. Howard and L. Richard. Effects of text illustrations: A review of research. *ECTJ*, 30(4):195–232, 1982.
- [5] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, 2000.

- [6] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *TVCG*, 21(3):420–433, 2015.
- [7] L. Bartram, J. Rodgers, and R. Woodbury. Smart homes or smart occupants? supporting aware living in the home. In *Human-Computer Interaction-INTERACT 2011*, pages 52–64. Springer, 2011.
- [8] D. Baur, F. Seiffert, M. Sedlmair, and S. Boring. The streams of our lives: Visualizing listening histories in context. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1119–1128, 2010.
- [9] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1797–1806. ACM, 2008.
- [10] J. Froehlich, T. Dillahunt, P. Klasnja, J. Mankoff, S. Consolvo, B. Harrison, and J. A. Landay. Ubigreen: investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1043–1052. ACM, 2009.
- [11] J. Froehlich, L. Findlater, M. Ostergren, S. Ramanathan, J. Peterson, I. Wragg, E. Larson, F. Fu, M. Bai, S. Patel, et al. The design and evaluation of prototype eco-feedback displays for fixture-level water usage data. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2367–2376. ACM, 2012.
- [12] B. A. Aseniero, A. Tang, and S. Carpendale. River: Using personalisation to support reflection on personal activities. In *Proceedings of the DIS'14 Workshop on A Personal Perspective on Visualization and Visual Analytic*, page 4, 2014.
- [13] A. Thudt, D. Baur, S. Huron, and S. Carpendale. Visual mementos: Reflecting memories with personal data. *TVCG (in press)*, 2015.
- [14] Fitbit Inc. Fitbit. [www.fitbit.com](http://www.fitbit.com). Accessed: 2015-08-29.
- [15] Nike Inc. Nike+. [nikeplus.nike.com](http://nikeplus.nike.com). Accessed: 2015-08-29.
- [16] Fooducate LTD. Fooducate. [www.fooducate.com](http://www.fooducate.com). Accessed: 2015-08-29.
- [17] MyFitnessPal Inc. Myfitnesspal. [www.myfitnesspal.com](http://www.myfitnesspal.com). Accessed: 2015-08-29.
- [18] S. D. Acharya, O. U. Elci, S. M. Sereika, M. A. Styn, and L. E. Burke. Using a personal digital assistant for self-monitoring influences diet quality in comparison to a standard paper record among overweight/obese adults. *JADA*, 111(4):583–588, 2011.
- [19] R. Maceachern. *design x food*. issue, first edition, 2013. <http://issuu.com/ryanmaceachern/docs/designxfood>.
- [20] J. Dawson. Nut: A tool for understanding the nutritional content of what we eat. Technical report, University of British Columbia, 2011. [www.cs.ubc.ca/~tmm/courses/533/projects/jessica/report.pdf](http://www.cs.ubc.ca/~tmm/courses/533/projects/jessica/report.pdf).
- [21] M. Bush, S. Holbert, and N. Viswanthan. Newtrition - visualizing nutrition data. Technical report, Stanford University, 2010. [www.nikilster.com/papers/Newtrition%20-%20Visualizing%20Nutrition%20Data.pdf](http://www.nikilster.com/papers/Newtrition%20-%20Visualizing%20Nutrition%20Data.pdf).
- [22] G. Liu. Visualizing nutrients. Technical report, Ontario College of Art and Design, 2010. <http://lonelydatum.com/visualize/food>.
- [23] B. Dawes. Cinema redux. Project, Museum of Modern Art, 2004. <http://brendandawes.com/projects/cinemaredux/>.
- [24] F. Brodbeck. Cinemetrics. Project, Royal Academy of Arts, 2011. <http://cinemetrics.fredericbrodbeck.de/>.
- [25] B. Dawes. Digital city projects. Project, Everything Everywhere, 2012. <http://brendandawes.com/projects/ee>.
- [26] E. Hinton, M. Bostock, S. Carter, C. Horyn, and E. Wilson. Front row to fashion week. Project, The New York Times, 2013. [www.fpcomplete.com/wp-content/uploads/2013/08/NYTimes.pdf](http://www.fpcomplete.com/wp-content/uploads/2013/08/NYTimes.pdf).
- [27] J. Rodgers and L. Bartram. Exploring ambient and artistic visualization for residential energy use feedback. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2489–2497, 2011.
- [28] C. Tominski, J. Abello, and H. Schumann. Axes-based visualizations with radial layouts. *Proceedings of ACM symposium on applied computing*, 15(5):1242–1247, 2004.



fields. Finally, we propose a taxonomy which we adopted based on previous frameworks for information visualization. This work is an initial step towards answering the question of “How to measure engagement with an Information Visualization system? and there are many other open questions that need to be explored in more depth.

## 2 BACKGROUND

In this section, we address how engagement has been defined, measured and categorized in information visualization and other related domains.

### 2.1 Engagement in Information Visualization

#### 2.1.1 Definition

There is no clear definition of what engagement means in the Information Visualization domain. From a behavioral perspective, Boy et al. and Haroz et al. refer to engagement as users’ willingness to invest effort to explore further and gain more information from the visualization [2, 6]. Moere et al. [12] used the term engagement as the (perceived) effectiveness of visualizations, measured by higher participation/use of visualizations. We think that these definitions, along with behavior-specific measurements of them, do not highlight the granularity of engagement.

#### 2.1.2 Methods for User Engagement

Previous research mainly addressed three dimensions for engagement: **aesthetic**, **narrative**, and **interactive** [13]. Recently, researchers have reported empirical results of what triggers engagement with visualizations. Different aspects of visualizations have been evaluated to see whether the participants were engaged with the visualizations. Wood et al. suggested that using a **sketchy style** InfoVis may increase engagement. They observed participants had more positive attitudes while annotating sketchy style visualizations. Moere et al. tested **different styles of visualization** such as analytical style, magazine style, and artistic style [12] and found that users could find more fact-based and deeper insights with analytical style visualizations. Haroz et al. showed that **ISOTYPE** that uses simple pictographic elements helps users to inspect visualizations more closely [6]. At last, **Storytelling** was used to trigger user-interaction and exploration. However, the results indicated that augmenting stories with visualizations did not increase user-engagement [2].

Prior research categorize methods to engage people with InfoVis with respect to following aspects of visualizations: **Visual Encoding**, **Aesthetic**, and **Interaction Techniques**. We believe that an engagement taxonomy should not only reflect all of these methods but also include users’ goals, familiarity, interest, and background knowledge of the data. These factors can affect the willingness to explore the visualizations and eventually determine the insights that users gain [22]. In addition, we think that there are other factors that could affect user engagement particularly in the context of personal visualization. Factors such as **Display Type, Size and Input** and **Accessibility and Device Compatibility**.

While designing methods for engagement, we also need to consider the users’ context while using Personal Visualization and Personal Visual Analytics applications. Huang et al. [9] categorized current applications on Personal Visualization and Personal Visual Analytics and distinguished the designs based on the different levels of attentional demands, explorability, and actionability. For example, a sleep-related activity reminder app would require low attentional demand, low explorability, and high actionability. These different levels of design could require different levels of engagement to fulfill the goal of the application.

#### 2.1.3 Measures

Prior research provide measures of engagement. Boy et al. [2] measured low-level interactions such as hovering and clicking that were mapped to certain activities such as retrieving values, filtering, or exploring with the visualization. The total amount of these type of interactions were considered as the level of engagement. However, we suspect that the metrics used may have been inadequate and insufficient for measuring engagement. Haroz et al. gave a 3x3 grid of thumbnails that include text, plain InfoVis, and ISOTYPE InfoVis, and captured how much users clicked to get more details in 2 minutes [6]. This measure could also be inaccurate as it does not reflect user level of understanding. Wood et al. used “willingness to annotate a visualization” as an indirect measure of engagement with the visualization [21]. Though this type of measure can capture higher-level engagement, it is too intrusive so that users need to explicitly report the willingness.

### 2.2 Research on Engagement in Other Domains

Sutcliffe [20] defines user engagement as “how people are attracted to use interactive products. User engagement is a complex concept that synthesizes several influences to promote a sense of flow and fluid interaction leading to satisfying arousal and pleasurable emotions of curiosity, surprise, and joy”. Sutcliffe refers to **interaction**, **media** and **presence** as three main components of user engagement. “Interaction describes the content being communicated. Media describes how the user and the means of interaction are represented, ranging from simple cursors to icons and interactive avatars. Presence is determined by the representation of the user and how immersion is afforded by the interface on a 2D interactive surface or in a more elaborate 3D interactive world” [20].

In the context of scientific visualization, Edelson and Gordinb proposed a design framework for the creation of scientific investigation tools based on differences between scientists and science students [4]. Although the context is different, we believe their framework can be adopted for designing visualizations in more general terms. One common attribute is that they address the differences between experts and learners. Their goal is to “Take resources that enable experts to extend their knowledge and turn them into resources that enable learners to develop some of the knowledge possessed by experts by performing personally meaningful tasks”.

Hart and Angeli [7] refer to **positive affect** and **high arousal** as measures for engagement. According to them “interaction design may promote positive affect and arousal through serendipitous effects, variable pace, use of avatars and virtual environments”. They address the influence of task, context and user characteristics on perceptions of engagement as important factors.

### 2.3 Frameworks for Engagement

There are different frameworks that have been suggested for user engagement in different domains. Naps et al. framework [15] focuses on the educational impact of algorithm visualization. This framework has been used and extended for a variety of contexts (e.g. software education and collaboration [14]). Although this framework has been mostly used for pedagogy purposes, we believe there are a lot of common concepts that can be adopted for personal visualization. They argue that no matter how well visualization technology is designed, there is little educational value unless it engages learners in an active learning activity. There is a commonality between our audience and learners. Learners do not have expertise as scientist, same as novice users who are non-visualization experts who are trying to make sense of their own data.

Naps et al. [15] engagement taxonomy defines six different forms of learner engagement with visualization technology: 1) *no viewing*, 2) *viewing*, 3) *responding*, 4) *changing*, 5) *constructing*, and 6) *presenting*. We can adopt some of these levels, except no



viewing, which means that there is no visualization technology in use, and changing, which means that the system asks students for input to affect the execution of a program. Myller et al. [14] extended this taxonomy for software visualization and collaborative learning. One of their added items is: reviewing (i.e., viewing the visualization for the purpose of providing comments, suggestions, and feedback on the visualization itself or the program.) It is notable that they do not consider this taxonomy to be an ordinal scale and they mention that “ The relationships among these six forms of engagement do not form a simple hierarchical relationship.”

Another relevant taxonomy in the context of learning visualization is Bloom’s taxonomy [1]. They look at the depth of understanding and define levels while interacting with visualizations. This taxonomy has six levels as, the knowledge level, the comprehension level, the application level, the analysis level, the synthesis level, and the evaluation level. We elaborate on definitions of each level as we believe it relates to the field of InfoVis.

1. **The knowledge level:** refers to mere factual recall with no real understanding of the deeper meaning behind facts.
2. **The comprehension level:** learner is able to discern the meaning behind the facts.
3. **The application level:** learner can apply the learned material in specifically described new situations.
4. **The analysis level:** learner can identify the components of a complex problem and break the problem down into smaller parts.
5. **The synthesis level:** learner is able to generalize and draw new conclusions from the facts learned at prior levels.
6. **The evaluation level:** learner is able to compare and discriminate among different ideas and methods. By assessing the value of these ideas and methods, the learner is able to make choices based on reasoned arguments.

This taxonomy has been revised and listed as six levels of remembering, understanding, applying, analyzing, evaluating, and creating which provides a better means to evaluate visualization tools. In the following section we introduce which is very similar to this one, but has been adopted for measuring users engagement in InfoVis.

### 3 DISCUSSION AND FUTURE WORK

In this section, we conclude by addressing our proposed taxonomy and how it can capture the granularity of engagement. Then we address some of the open research issues in this field that requires further exploration. We hope that our effort will motivate other researchers to further investigate this problem and to establish a better foundation for engagement.

#### 3.1 Defining Engagement with its Degrees

As mentioned before, in order to be able to measure engagement, we first need to have a framework that encompasses different levels of engagement. Based on our survey of engagement literature in related domains, we propose the following taxonomy that defines engagement based on the level of user involvement (See Figure 1). Our taxonomy includes five levels: Expose, Involve, Analyze, Synthesize, and Decide. Each level includes former levels, but it refers to a deeper level of engagement which also correspond to deeper level of sensemaking. We adopted our taxonomy from Bloom’s digital taxonomy [1]. Similar to Bloom’s taxonomy, the degree of engagement increases from left to right as user performs higher-level cognitive tasks such as synthesizing information, and making final decisions. This classification provide a working definition of

degrees of engagement and a systematic approach to capture the individual degrees of user engagement.

1. **Expose** (Viewing): the user knows how to read data points.
2. **Involve** (Interacting): the user interacts with the visualization and manipulate the data.
3. **Analyze** (Finding Trends): the user analyze the data, finds trends, and outliers, etc.
4. **Synthesize** (Testing Hypotheses): the user is able to form and evaluate hypotheses.
5. **Decide** (Deriving Decisions): the user is able to draw final decisions based on evaluations of different hypotheses.

To better enable visualization researchers to evaluate user engagement, we capture what can be generated about the data at each level. We speculate that each level re-acquires different metrics. While, addressing metrics for each level needs further research and exploration, we believe that many of Bloom’s taxonomy concepts can be adopted here. For instance, Expose level can be measured by people’s ability to **remember**. Whereas **understand, apply, derive insights** from data are addressing deeper level of engagement. To address deeper level of engagement, there are other possible metrics such as users’ new questions and hypotheses which can address Synthesize level. This shows the users’ ability to go beyond initial questions to gain some insights [16].

#### 3.2 Open Research Questions

In this section, we bring attention to some open research questions in the context of Information Visualization. These questions can be used for future researchers to discuss the future direction of engagement research in the visualization community.

- What is the definition of engagement? This is the most critical question to be answered. Addressing other issues are determined by this definition.
- What are the levels of engagement? It is necessary to identify and distinguish between different levels of engagement.
- How to measure engagement? The inherent complexity of engagement makes measuring a challenge. What should be measured and what the appropriate metrics are, are still unclear.
- What are the differences between short-term vs long-term engagement? Do they require different metrics for measurement? There are situations (e.g. monitoring health indicators) where a user needs to interact with a visualization for longer periods of time. This issue should be reflected in any future evaluation framework because short and long term engagement may require different evaluation methods.
- How to increase the engagement level by visualization design? Although designing engaging visualizations is beyond the focus of this paper, we believe that a clear definition of engagement along with a taxonomy of different levels of engagement would be useful for visualisation researcher and designers.
- Do certain personal traits correlate with the degree of engagement with a visualization? There is also the issue of personal differences and preferences. What are the type and magnitude of personal traits on engagement? Would factors such as gender, age, background and education affect engagement?
- How can we incorporate the user engagement level into visualization models, such as knowledge generation [19] and trust building under uncertainty [18]?

## REFERENCES

- [1] B. S. Bloom. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1-2*. Longmans: McKay, 1974.
- [2] J. Boy, F. Detienne, and J.-D. Fekete. Storytelling in information visualizations: Does it engage users to explore data? In *Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems (CHI 2015)*, pages 1449–1458. ACM, 2015.
- [3] S. Consolvo, P. Klasnja, D. W. McDonald, D. Avrahami, J. Froehlich, L. LeGrand, R. Libby, K. Mosher, and J. A. Landay. Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 54–63. ACM, 2008.
- [4] D. C. Edelson and D. Gordin. Visualization for learners: A framework for adapting scientists’ tools. *Computers & Geosciences*, 24(7):607–616, 1998.
- [5] C. Fan. The future of data visualization in personal informatics tools. In *Personal Informatics in the Wild: Hacking Habits for Health & Happiness CHI 2013 Workshops*, 2013.
- [6] S. Haroz, R. Kosara, and S. L. Franconeri. Isotype visualization—working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1191–1200. ACM, 2015.
- [7] J. Hart, A. Sutcliffe, and A. De Angeli. Using affect to evaluate user engagement. In *CHI’12 Extended Abstracts on Human Factors in Computing Systems*, pages 1811–1834. ACM, 2012.
- [8] J. Heer, F. Van Ham, S. Carpendale, C. Weaver, and P. Isenberg. Creation and collaboration: Engaging new audiences for information visualization. In *Information Visualization*, pages 92–133. Springer, 2008.
- [9] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 21(3):420–433, 2015.
- [10] I. Li, A. Dey, and J. Forlizzi. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 557–566. ACM, 2010.
- [11] J. J. Lin, L. Mamykina, S. Lindtner, G. Delajoux, and H. B. Strub. Fishsteps: Encouraging physical activity with an interactive computer game. In *UbiComp 2006: Ubiquitous Computing*, pages 261–278. Springer, 2006.
- [12] A. V. Moere, M. Tomitsch, C. Wimmer, B. Christoph, and T. Grechenig. Evaluating the effect of style in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2739–2748, 2012.
- [13] S. Murray. Engaging audiences with data visualization: Communicating your research, data, and findings to people outside your field [audio podcast], Feb. 2013.
- [14] N. Myller, R. Bednarik, E. Sutinen, and M. Ben-Ari. Extending the engagement taxonomy: Software visualization and collaborative learning. *ACM Transactions on Computing Education (TOCE)*, 9(1):7, 2009.
- [15] T. L. Naps, G. Rößling, V. Almstrum, W. Dann, R. Fleischer, C. Hundhausen, A. Korhonen, L. Malmi, M. McNally, S. Rodger, et al. Exploring the role of visualization and engagement in computer science education. In *ACM SIGCSE Bulletin*, volume 35, pages 131–152. ACM, 2002.
- [16] C. North. Toward measuring visualization insight. *Computer Graphics and Applications, IEEE*, 26(3):6–9, 2006.
- [17] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. Int. Conf. on Intelligence Analysis*, volume 5, pages 2–4, 2005.
- [18] D. Sacha, H. Senaratne, B. Kwon, G. Ellis, and D. Keim. The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–1, 2015.
- [19] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. Keim. Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), Dec. 2014.
- [20] A. Sutcliffe. Designing for user engagement: Aesthetic and attractive user interfaces. *Synthesis lectures on human-centered informatics*, 2(1):1–55, 2009.
- [21] J. Wood, P. Isenberg, T. Isenberg, J. Dykes, N. Boukhelifa, and A. Slingsby. Sketchy rendering for information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2749–2758, 2012.
- [22] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaluation methods for Information Visualization*, page 4. ACM, 2008.

# Are there networks in maps?

## An experimental visualization of personal movement data

Heike Otten\*

Lennart Hildebrandt†

Till Nagel‡

Marian Dörk§

Boris Müller¶

Urban Complexity Lab  
Potsdam University of Applied Sciences

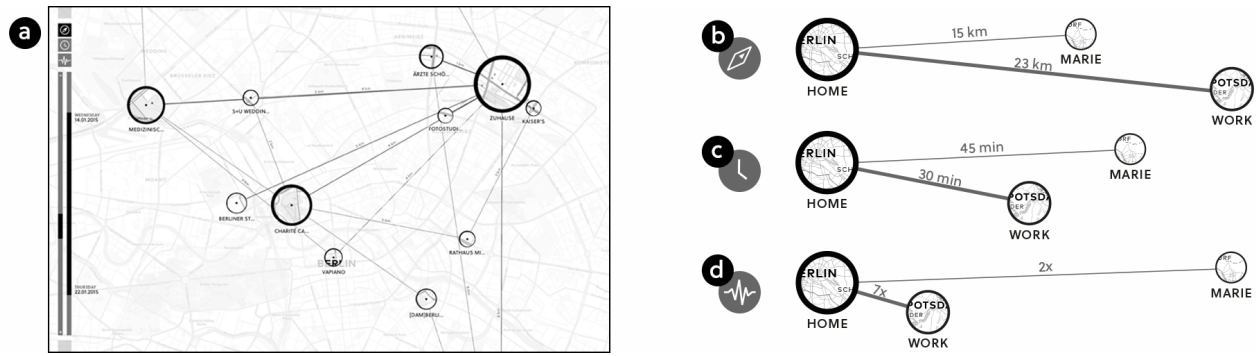


Figure 1: The Shifted Maps visualization (a) and three viewing modes for spatial (b), temporal (c), and frequency (d) networks of places.

### ABSTRACT

Shifted Maps proposes a novel visualization method to generate personal geovisualizations of individual movement data. The resulting visual appearance can be characterized as a map network consisting of visited places and their connections. The visited places are shown as circular map extracts scaled according to the time spent there and the movements between the places are represented as edges between the places. A key feature of the Shifted Maps visualization is the possibility to explore the data in three different arrangements based on geo-spatial position, travel time, and frequency of movements. By combining map and network visualizations of movement data, it becomes possible to analyze and compare spatial and temporal topologies.

**Keywords:** Geovisualization, Maps, Personal visualization, Spatiotemporal Data.

### 1 INTRODUCTION

Due to the increasing amount of collected personal data we witness a growing need for tools for analyzing these datasets. Conventional visualizations are prone to represent mass data, which tend to lack accessibility and personal relevance. Especially with the rise of the mobile phone and its tracking capabilities, more and more people use their devices to collect a wide range of personal data. While personal datasets offer the possibility to recognize small-scale patterns in individual behavior, they also increasingly result in more comprehensive and complex patterns at a larger scale and consequently require new methods to make sense of them.

\*e-mail: heike.otten@fh-potsdam.de

†e-mail: lennart.hildebrandt@fh-potsdam.de

‡e-mail: nagel@fh-potsdam.de

§e-mail: doerk@fh-potsdam.de

¶e-mail: boris.mueller@fh-potsdam.de

Electronic proceedings of the IEEE VIS 2015 workshop  
*Personal Visualization: Exploring Data in Everyday Life*  
The authors remain the holders of the copyright

With Shifted Maps we present a visualization of individual movement data, which enables people to discover recurring patterns in long-term mobility traces. In order to reveal diverse relationships between places, we developed a novel visualization consisting of a network of map extracts. We designed this technique to facilitate the representation of multiple correlations through nodes, while maintaining a reference to the spatial information through maps.

### 2 RELATED WORK

In the visualization community, there is a growing interest in visual analysis tools to support casual data visualization [6]. Over the last few years several projects started to investigate the visualization of personal movement data. The default view of the Moves app [2], which allows people to log their mobility, is a non-spatial visualization of places on a timeline using simple glyphs to represent the temporal length of trips and stays. Move-o-Scope [1] uses simple traces on a map but additionally provides a flow diagram of selected places indicating the most related places based on co-visitation. The web-based visualization Visits [8] puts map extracts on a timeline showing visited places, thus highlighting the relationship between time and space along a travelog. We aim to take these efforts further and explore the non-linear relationships between places emerging over time.

Representing spatiotemporal data is one of the core themes of geovisualization. At the most basic, this consists of showing different time snapshots of spatial data side by side or sequentially. Plenty of techniques aim to visually integrate temporal with spatial properties [3], ranging from flow maps to space time cubes. Often these techniques are used to visually analyze large amounts of aggregated movement data. We are interested in finding new approaches to explore smaller scale personal data sets. Revealing additional information such as visit duration and frequency allows people to better understand the temporality of their movement. These aspects can be harnessed to expose significant parts of one's personal environment. A number of techniques such as lenses have been introduced to emphasize certain parts of maps. A few of these techniques can be used to highlight several areas simultaneously [5,

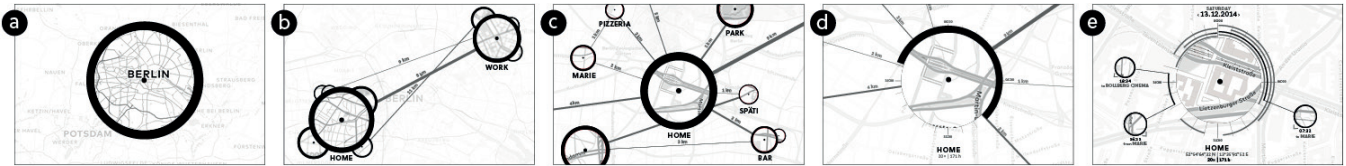


Figure 2: Semantic zoom showing more and more details. When zooming in (a to e) finer grained spatial relations emerge, then revealing the journeys between near-by places, and finally visualizing the stays at a single place (e).

7], yet the resulting distortions can cause misinterpretation of the underlying map [4].

In sum, most spatiotemporal visualizations tend to prioritize spatial over temporal patterns by displaying the underlying map as the primary representation for positioning data nodes. We aim to subvert this relationship between space and time, in order to reveal diverse data topologies in personal movement patterns.

### 3 SHIFTED MAPS - THE PROTOTYPE

Shifted Maps visualizes personal movement data. Its main purpose is revealing patterns in one’s own movement network and enabling users to gain a different, maybe more personal perspective on their own movements through their spatial surrounding. Our environmental perception is based on often visited places and frequently taken routes. The relationships between these personal places are complex, as some connections occur more frequently than others, the time needed to move between them can vary greatly, and the spatial distances between places is interwoven with networks of visitation sequences. Considering that these complex relationships are latent in all mobility datasets there is an opportunity to use existing network visualization techniques such as node-link diagrams. When visualizing movement data as node-link diagrams, geographic references and temporal sequences need to be integrated, for example, to aid the understanding of co-visitation patterns, i.e., which are the most visited places and what is the frequency of journeys between them. The design of our visualization is led by the objective of representing spatiotemporal networks of personal movements.

#### 3.1 Visualizations

Shifted Maps proposes a network of map extracts to visualize the diverse relationships between the places one traverses. The spatiotemporal networks are represented as node-link diagrams with nodes being visited places displayed as map extracts and connections between them shown as straight lines. The rationale behind such a hybrid view is to harness people’s growing familiarity with digital maps with the general versatility of node-link diagrams representing diverse topologies and relations. The map extracts are circular to correspond with a sense of perceiving the immediate spatial surrounding of a place, with a dot indicating its exact location. The area size of a circle corresponds to the time spent there, while the thickness of the ring represents the relative number of visits. Though the size of a circle may change the geospatial area it represents stays the same; it adjusts its zoom level according to the geometric size of the circle. To focus on the network structure between places, we simplify routes as straight lines with their thickness representing the number of trips.

The map extracts and the connections between them result in a graph representing the personal network of places and the movements between them. The advantage of a network of map extracts is the possibility to arrange it in different ways, while maintaining a reference back to the geographic positions through the map extracts. Shifted Maps offers three arrangements: In the geographic view, the arrangement of places is based on their geographic positions (Fig. 1a, b). In the temporal view, the places are arranged according to average time it takes to go from one

place to another (Fig. 1c). For the frequency view, places with a frequent connection move closer to each other than places with a less frequent connection (Fig. 1d). The length of the edges and the numeric labels along them respectively indicate the spatial or temporal distance, or the frequency of travels.

#### 3.2 Interactions

Though the static visualization already reveals interesting relationships, interaction with Shifted Maps enables a comparison of different places, time spans, and views. The three different views can be chosen via the buttons in the top left corner. A smooth transition between the views aims to ease the understanding of the shifting arrangements. In principle, data of any temporal length can be loaded into the visualization. A time-range slider enables the selection of a certain temporal period, making it possible to compare shorter spans or follow the growth of the network over time. Each view can be explored through a semantic zoom, which enables the user to get an overview of the connections in lower zoom levels, while revealing more detailed information about places and connections in higher zoom levels. Zooming in reveals overlapping circles making individual places and more connections visible. In parallel, details such as line labels, place names, and further information about the places are bit by bit revealed (Fig. 2). At the highest zoom level a detailed view displays the duration of stay at individual days, as well as connection details, i.e., which other places have been visited before and after the focused place.

#### 4 FIRST USES

During the iterative design process we carried out a formative evaluation, in order to test the comprehension of our visualization, its effects and general functionality. As testing the prototype with real datasets was indispensable, we recruited three participants (one female, two males, aged 24-27) with datasets of about two weeks. The tested prototype contained all main functions, except the semantic zoom. All participants were eager to explore a visualization of their own movement data and quickly understood the geographic and the time view. The frequency view, however, caused problems of comprehension, which may result from inaccuracies caused by an error in the graph layout algorithm. In general, participants remembered and recognized frequently visited places, whereas rarely visited places were partly already forgotten and took longer to identify and rediscover. The interaction techniques were largely understood and used without problems, though some of the users wished there were more interaction opportunities, i.e., interacting with the circles or lines, and additional information about the places.

#### 5 CONCLUSION AND FUTURE WORK

The rising amount of personal movement data, collected with smartphones, provides the opportunity for individuals to gain valuable insights into their own movements. Shifted Maps proposes a visualization of personal movement data with a network of map extracts. The unique potential of this technique is the possibility to see places and movements in novel ways. Three different arrangements based on the geographic, temporal, or frequential connections between places were designed to help people reveal

patterns in their movements. An early user test showed that people are able to reveal forgotten places and connections through the prototype. For the next version of Shifted Maps, we are planning to include more interaction capabilities and additional place information. Currently, we are developing a web implementation of our visualization to make it accessible for a broader audience and enable a large-scale evaluation of the new technique.

#### ACKNOWLEDGEMENTS

We thank all students from the *Doing with Data* class for their constructive feedback. This research was supported in part by HERE.

#### REFERENCES

- [1] Move-O-Scope. <https://move-o-scope.halftone.co/>. Accessed: 2015-07-30.
- [2] Moves. <https://www.moves-app.com/>. Accessed: 2015-07-30.
- [3] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale. A Review of Temporal Data Visualizations Based on Space-Time Cube Operations. In *Proc. of EuroVis*, 2014.
- [4] A. Cockburn, A. Karlson, and B. B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv.*, 41(1):2:1–2:31, Jan. 2009.
- [5] B. Jenny. Geometric distortion of schematic network maps. *Bulletin of the Society of Cartographers*, 40(1):15–18, 2006.
- [6] Z. Pousman, J. T. Stasko, and M. Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE TVCG*, 13(6):1145–1152, 2007.
- [7] G. Shoemaker and C. Gutwin. Supporting multi-point interaction in visual workspaces. In *Proc. of CHI*, pages 999–1008. ACM, 2007.
- [8] A. Thudt, D. Baur, and S. Carpendale. Visits: A Spatiotemporal Visualization of Location Histories. In *EuroVis - Short Papers*, 2013.

# Everyday Visualization

Bruno Pagno\*  
UFRGS

Luciana Nedel†  
UFRGS

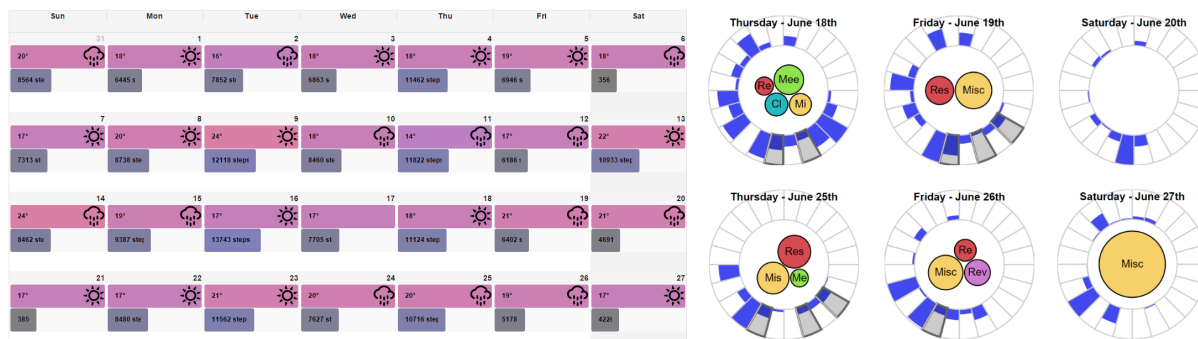


Figure 1: Everyday visualization showing a single user's data. On the left the broader calendar view, and on the right the detailed clock view.

## ABSTRACT

The use of activity monitors has increased significantly in the recent years. With the improvement of technology and the availability of better sensors, it became easier for users to watch their activities. However, there is still a lot to do. We accumulate significant amounts of data and monitor different parts of our lives, but we still miss insights about ourselves. In this paper, we propose a way to aggregate and visualize data obtained from different sensors, allowing individuals to better understand themselves. We focus on acquiring data from various sources and producing simple and casual visualizations.

## 1 INTRODUCTION

The primary objective of each person is to have a good life. The concept of *good life* may vary according to one's personality, but it is often a good balance of health, work, and personal life. Each one has just a single life, and it is not easy to achieve happiness disregarding one of these aspects. However, keeping a balanced life requires personal investment, self-knowledge, and discipline. It is not easy to manage this alone. The pursuit of a better behavior depends a lot on the knowledge of the past, and, as a consequence, on the everyday monitoring.

There is an abundance of data in our lives. We keep producing and collecting information on everything, including ourselves. Every day we found new ways to acquire data, and we are using it to improve ourselves, as well as the activities in which we take part. Thanks to Facebook we know our friends' birthday and can congratulate them in time while heart rate monitors help us to understand our instantaneous cardiac frequency and change our fitness in real-time.

There is a growing variety of sensors and different forms of capturing data from an individual. Nevertheless, there are many challenges in dealing with this kind of information. If not filtered or

presented adequately it can easily become overwhelming or uninteresting.

The goal of this project is to produce a casual customizable visualization for individuals data. It should allow the user to find knowledge about him and find ways to improve his life in specific objectives (i.e. health, work, or social life). This is a work in progress, and we intend to increase the amount of data presented and the variations/customization of the visualization tool significantly.

We collected a relevant amount of information of one single user and developed a visualization to aggregate it all. The data acquired is heterogeneous and dealing with it is a challenge by itself. We focus on the simplest and most intuitive way to show the information, using visualization metaphors present in our everyday lives, such as maps, calendars, and clocks. We present the information in different levels of detail, making it easier and quicker to understand.

The remaining of this paper is organized as follows. Section 2 presents some related works. Section 3 introduces the process used for data acquisition and management, as well as the list of data we are monitoring. Finally, Section 4 presents the visualizations we are using and the results achieved, and Section 5 the conclusions and future work.

## 2 RELATED WORK

Personal visualization has been actively discussed by researchers recently. There are many challenges from capturing sensors data to presenting useful visualizations.

Huang et al. [4] did an impressive survey on papers about personal visualization (PV) and personal visual analytics (PVA). In their work, they sort different attributes of individual visualization projects such as the effort to collect the data and who it is intended to inform. They present an extensive explanation of each attribute describing their importance and relations with each other. They also present the current scenario and challenges of PV and PVA, which go from fitting in personal routines to integrating computer analysis and evaluation. According to the authors, there is still a lot to do on PV and PVA.

Pousman et al. [11] describe casual information visualization as visualization for non-experts. They claim that many visualizations are made for professionals and that it becomes difficult for untrained individual to understand. For that same reason, casual

\*e-mail: blpagno@inf.ufrgs.br

†e-mail: nedel@inf.ufrgs.br



visualizations reach a wider set of audience. Considering that personal information is meaningful mostly to the user himself, it makes a lot of sense to make it as casual as possible.

The list of sensors that track people's activities is ever growing, as can be seen in TicTrac [15], a Web-based application that aggregates several personal information tracking software. The objective of TicTrac is similar to our own. The application is very strong on aggregating many different activities monitors but the visualizations available are still generic and simple, without refining or giving detailed information.

The availability of information on movement data makes it frequently used for user analysis. Zend et al. [16] and Grundy et al. [2] show different forms of visualization of movement data. In their case, they gathered data from lots of individuals making it possible for a different analysis. The detailed movement data may not be necessary for individuals, though, but it can be interesting to know the locations attended by the user.

Many works approach individual user data focusing on health information. Pantelopoulos and Bourbakis [10] present a review on wearable sensors for health monitoring while Schneiderman et al. [14] address the subject focusing on how visualizations can support and improve already existing health systems. Schneiderman et al. also present challenges of health visualization. However, this is focused on health visualization, which is often developed for experts and not necessarily useful to the individual himself.

Most of the user individual data is temporal. In the same way, most of the activity monitors register information according to time. For that reason, it is often a good idea to also present data in relation to time. Huang et al. [5] created an attractive solution joining user information in a regular calendar visualization. They can show user's physical activity (number of steps) or the amount of household energy used. Their solution is appealing because it displays the information to the user in a non-obtrusive manner. However, it lacks a method to visualize more details. The large amount of data displayed can be of little meaning when there is no way for the user to view detailed info.

Still regarding temporal data visualization, Kondo and Collins [7] show a technique to interact with timeline visualizations. In visualizations with many data, it becomes necessary to interact, filter and visualize the details of the information presented.

Sports often explore individual information. A previous work we developed on NBA Visualization [9] had a particular view dedicated to observing player's information. It shows attributes like points scored and rebounds of each player and his evolution each year. This individual data visualization allowed us to observe the players that improved the most each year, and in which attributes specifically.

Satyanarayan and Heer [13] present a paper about storytelling with visualizations. The authors reinforce Pousman et al. on using visualizations that are not so focused on professionals. Individual user data is an excellent source to storytelling, and it can be very interesting for end users to review their memories through storytelling with data.

The visualizations mentioned here have similar goals to ours, specially the calendar of Huang et al. [5] and TicTrac [15]. However, we intend to treat personal visualization from a broader point of view. Our idea is to aggregate much more information of a single user, and consistently present it, in a more intuitive way than we found in the previous works.

### 3 DEALING WITH DATA

Data acquisition is a crucial step for visualization because the final product is based on the information gathered. If we can find insightful data, it becomes easier to produce insightful visualizations. This is still a big challenge when we talk about personal visualization. People have too many variables that we cannot measure yet,

and the ones we can measure, often do not say much by themselves. For that reason, in this project we gathered a significant amount of information from a single user, believing that correlations between them can be of interest.

The data collected in this work until now came from various different sources and are presented below:

- Steps: data captured using a Jawbone Up3 [6] tracker. Represents the number of steps taken in a given period of time. The user wears a tracker (wrist belt) that continuously collects data through a set of sensors including a tri-axial accelerometer
- Sleep: also captured from the Jawbone Up3 tracker, it represents the number of hours slept each night. Information is tracked using the same accelerometer and a heart rate monitor. The sleep quality is logged according to the period (wake, light, deep, and REM)
- Location: information captured from Foursquare [1], the data used contains the timestamp, coordinates, the name of the city and country where the user was
- Climate: data gathered from the OpenWeatherMap API [8]. The data used includes the temperature and climate conditions on the location the user was
- Emails sent/received: data collected from Gmail through TicTrac [15]. Information obtained includes the timestamp, subject and text of each email sent and received by the user
- Tweets: also collected using TicTrac, includes the timestamp, the "@mentions" and the text of the Twitter messages sent by the user
- Appointments: collected from the Apple iCal app, the attributes used are the start and end time of the appointment, title, and description
- Work: information obtained from Hours [3] and RescueTime [12]. Hours is an iOS app for time tracking where the main goal is to track the time spent in different projects. The user should be able to inform manually all data in a very easy manner. RescueTime, for the other hand, is an automatic tool for PCs. A script executes in background and logs all the activity, organized by categories. The data obtained from both sources include a timestamp of each activity and the time spent on it

Gathering information from so many different sources lead us to some troubles. Firstly, it is important to consider that there are slight differences in the format of each source. For example, timestamps had different precision on different providers. Then, before producing any visual result, we had to standardize all the information.

Another problem that we experienced involved the period of the data acquired. TicTrac, Foursquare, iCal, and Hours readily provided information with precise timestamps. However, data captured from the Jawbone Up only allowed us to retrieve the total steps given per day. To better understand the walking pattern of the user, we needed more detailed information. Then, we had to write a script that request updated information on short intervals and log the number of steps by ourselves in a local database.

We also had to deal with different periods of data on each source. Some sources of data had been logged for longer and, consequently had data of a larger time span. To address that, we found a period where all of the sources were active. For now, we fixed this time slice of data for the following steps of our work, because we wanted to develop some visualizations with a consolidated, immutable data

set. After dealing with these difficulties, we have a lot of registers somewhat organized to start building visualizations.

It is also important to mention that each user has different personal preferences, and many of these sources are not suited for other potential users. Although we believe that there is a subset of basic “sensors” that could be always considered, a toolkit for personal visualization should provide a graceful manner to easily save and integrate data coming from different sources. Even if we are neglecting it in this work, we are aware that this problem need to be faced soon.

Up to now we captured the data using Python, Ruby scripts, and some manual maintenance. The scripts written use the available APIs of each source mentioned in this section. The resulting information is, for now, stored in several CSV files. As noted before, this setup is clearly a non-optimal setup and for that reason we intend to improve the way we capture and store information in the next iterations of our work.

The visualization produced were created to work on the Web. It is interesting to make a Web-based solution for this kind of visualization because it becomes easily accessible by the end user anytime and anywhere. We used D3js, NodeJS and basic HTML, CSS and Javascript to produce the visualization.

For future works, we intend to use a more robust Web framework setup and to deploy it on a server. That way it will be possible to deal with information of more than one user in a structured way.

#### 4 EVERYDAY VISUALIZATION

As mentioned before, our goal is to provide intuitive visualizations of personal data to everyone. As all the data used in this study is temporal data, we decided to use the *calendar* and *clock* metaphors. While the calendar provides an overview of the self-behavior along several days, the clock gives a detailed view of a single day. In this way, we show information going from a broader perspective to a more detailed one.

The broader visualization is displayed as a calendar and contains information of total steps, temperature, and climate condition of each day (see Fig. 1). The information is presented in a clear and intuitive way without many details and allows the overview of longer periods of time, such as a week or a month.

The *calendar* visualization is composed of various cells, each one containing the information of a single day. Each cell has four attributes: the date, the temperature, the climate condition, and the number of steps taken. The first line of each cell shows the mean temperature and the general climate condition of the day. The color of the cell reinforces the temperature information, where blue means cold, and red means hot. The climate condition is represented by an easy to recognize symbol (e.g. a sun shining, a cloud, etc.). The second line shows a bar with the total steps taken during the day. The bar’s length changes according to the number of steps, being bigger when more steps were taken. Once more we use the color of the bar to reinforce the information displayed. We use a gradient from gray to blue, where blue means more steps and gray less.

When the user wishes to know a more detailed information he can move into the *clock* visualization (see Fig. 1 and Fig. 2). Each clock is divided into 24 slices – one for each hour – and represents a day of information. Ideally, we keep the clocks aligned in a calendar setup with seven days (clocks) per line, thus making weekdays coincide on the same column and keeping the consistency between both visualizations.

The *clock* visualization is more elaborate and presents more information than the *calendar*. It is composed of two areas showing different information on the same context. The outer part of the clock shows the amount of steps taken by the user. Each slice displays the number of steps taken in the respective period of one hour. It also displays any appointment that the user had by show-

ing thicker margin and a darker color tone on the slice. The appointments reflect the data captured from the iCal app. The inner part contains some bubbles representing the amount of time spent on different projects logged by the Hours app. The bubbles have different colors that make it easier to recognize the same task in different days.

The *clock* vis offers two options to display the number of steps taken during each day. If the user wishes to see a global overview of how much he walked during the day, there is a cumulative display option where each slice contains the sum of steps until that corresponding hour of the day. On the other hand, if the user wants to see at what time of the day he walked more, there is a noncumulative display option (see Fig. 2). This customization allows for some interesting insights. For instance, only on the cumulative mode the user can view which days he walked more than the others. This customization of the visualization is crucial, as pointed out by Huang et al. [4]. Looking for avoiding excessive labels, we added a tooltip with the raw information of each slice in the clock (see Fig. 3). When the user hovers the mouse on a slice, the hour of the day and the number of steps is displayed.

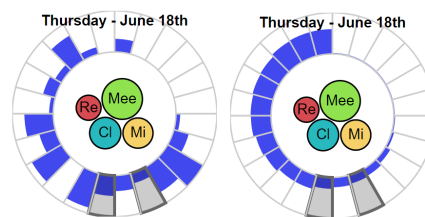


Figure 2: Clocks showing non-cumulative (left) and cumulative (right) steps information.

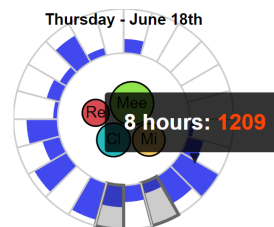


Figure 3: Tooltip on the clock visualization.

While the blue bars on the clock show the physical activity, productivity can be read looking to the inner bubbles that show the amount of time spent on each project. Projects can be configured by each user in the Hours App, but it is often used to log professional activity. The size of the bubble represents the amount of time devoted to the respective task and its size, though, is relative to the day, not global. In other words, the same bubble size on different clocks can mean different amounts of time. Each bubble also has a label and an associated color to assist the user in identifying the same task on different days. This way, it becomes possible for the user to know how he spent his time each day. For example, it is easy to identify the days with too many meetings.

Another possibility available is to visually compare two days of activity, as can be seen in Fig. 4. The information displayed is separated in different colors with transparency, enabling the user to see subtle differences in the number of steps for each moment of the day. This comparison can also be viewed in a cumulative shape.

Additionally, we developed a *map* visualization to offer a different view of the data – this time focused on location – instead of

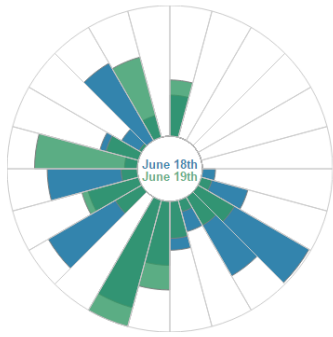


Figure 4: Comparison of two days activity.

presenting it according to time (see Fig. 5). It shows circles where the radius is proportional to the number of steps taken per day on the respective location. The color saturation is used to reinforce the number of steps per day. More red means more steps and less red mean fewer steps. Often the circles overlap each other, so we also added a tooltip that shows the average number of steps per day. As can be seen in Fig. 5 the user walked much more on the west coast than in other places. It allows for interesting insights as the user can see if his activities change when being away from the place he lives. In the specific case of Fig. 5, for instance, an inspection on the *calendar* visualization can help the user to reason that these amount of steps (more than usual) was taken during his last vacation in California.

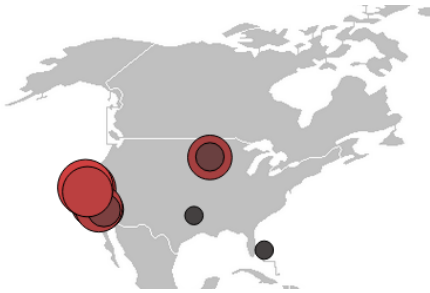


Figure 5: Map showing the number of steps taken in different places.

The *map* visualization is a nice addition because it allows the user to view his data from another perspective. Here, the location information captured from Foursquare is best presented in spatial visualizations, and can be composed with the other data – in this case, the steps – to provide insightful visualizations.

The visualizations presented in this section are still in its early steps, and there is a lot to evolve. As can be noted, we are not yet visualizing all the data we captured (e.g. sleep, e-mails exchanged, tweets). The three visualization metaphors presented here (*calendar*, *clock*, and *map*) allow the reasoning about our own behavior, but a better integration between the visualizations is still missing.

## 5 CONCLUSIONS AND FUTURE WORK

In this paper we presented a proposal for a multi-layered user-focused visualization of personal data. We also discussed the challenges of capturing and treating data from several different sources, and we showed a visualization that comprises different information from a single user in the same place.

The Everyday Visualization is still a work in progress, so we still have a long road ahead, but we think the preliminary results are promising. The technique based on the calendar and clock

metaphors can present concise information in an easy to understand casual visualization. We believe that we are taking the next step on activity monitors as they are too focused and frequently do not give us enough insight. Exploring multiple sources of information can help us to better understand ourselves and, as a consequence, to change our own behavior towards a better life.

Until now we collected a lot of information and are pursuing the best ways to present them. We are not yet visualizing all the data collected and there is still a lot to improve in the way we collect data. As on-going work, we are replacing manual capturing data methods for a real time and automatic strategy. This is probably the most important step to creating a user knowledge. The continuous flow of information will allow us to learn more about the users.

We also have to work more on the composition of different visualizations. While the various approaches used here to explore personal data work nicely, we still do not have a clear way to navigate between them. The challenge is never to let the user lose the context.

As a long term objective, we expect that with enough information on an individual, one will be able to make predictions of the user's actions, and suggest good and healthy behaviors, as a *personal assistant*.

## ACKNOWLEDGEMENTS

The authors would like to thank the Graphics, Visualization, and Interaction Lab. at UFRGS, Microsoft Brazil, CAPES and CNPq-Brazil (through graduate grants and project 485820/2012-9) for partially funding this work.

## REFERENCES

- [1] Developer.foursquare.com. foursquare for developers, 2015.
- [2] E. Grundy, M. W. Jones, R. S. Laramée, R. P. Wilson, and E. L. Shepard. Visualisation of sensor data from animal movement. In *Computer Graphics Forum*, volume 28, pages 815–822. Wiley Online Library, 2009.
- [3] Hourstimetracking.com. Hours time tracking app, 2015.
- [4] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 21(3):420–433, March 2015.
- [5] D. Huang, M. Tory, and L. Bartram. Data in everyday life: Visualizing time-varying data on a calendar. *Proc. Poster Compendium IEEE VIS*, 2014.
- [6] Jawbone. Up by jawbone — no activity tracker required, 2015.
- [7] B. Kondo and C. Collins. Dimpvis: Exploring time-varying information visualizations by direct manipulation. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2003–2012, Dec 2014.
- [8] Openweathermap.org. Weather api, 2015.
- [9] B. Pagno, L. Guedes, C. Freitas, and L. Nedel. Nbavis: Visualizing national basketball association information.
- [10] A. Pantelopoulos and N. Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(1):1–12, Jan 2010.
- [11] Z. Pousman, J. Stasko, and M. Mateas. Casual information visualization: Depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1145–1152, Nov. 2007.
- [12] Rescuetime.com. Rescuetime : Time management software for staying productive and happy in the modern workplace, 2015.
- [13] A. Satyanarayan and J. Heer. Authoring narrative visualizations with ellipsis. *Comput. Graph. Forum*, 33(3):361–370, June 2014.
- [14] B. Shneiderman, C. Plaisant, and B. Hesse. Improving healthcare with interactive visualization. *Computer*, 46(5):58–66, May 2013.
- [15] Tictrac.com. Tictrac, 2015.
- [16] W. Zeng, C.-W. Fu, S. M. Arisona, and H. Qu. Visualizing interchange patterns in massive movement data. *Computer Graphics Forum*, 32(3pt3):271–280, 2013.

# Personal Visualization for Learning

Jonathan C. Roberts\*  
Bangor University, UK

Christopher Headleand†  
Bangor University, UK

David Perkins‡  
Bangor University, UK

Panagiotis D. Ritsos§  
University of Chester, UK

## ABSTRACT

Learners have personal data, such as grades, feedback and statistics on how they fair or compare with the class. But, data focusing on their personal learning is lacking, as it does not get updated regularly (being updated at the end of a taught session) and the displayed information is generally a single grade. Consequently, it is difficult for students to use this information to adapt their behavior, and help them on their learning journey. Yet, there is a rich set of data that could be captured and help students learn better. What is required is dynamically, regularly updated personal data, that is presented to students in a timely way. Such ‘personal data’ can be presented to the student through ‘personal visualizations’ that engender ‘personal learning’. In this paper we discuss our journey into developing learning systems and our resulting experience with learners. We present a vision, to integrate new technologies and visualization solutions, in order to encourage and develop personal learning that employs the visualization of personal learning data.

**Keywords:** Learning Analytics, Personal Learning, Personal Data, Information Visualization, Teaching Analytics, Learning Support

**Index Terms:** K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

## 1 INTRODUCTION

In the Higher Education sector (especially) the concept of *Personal Learning* (PL) is gaining momentum. Within this agenda, academics are moving from *content-based learning* to *student-centric learning*, where an individual assumes responsibility for setting their own goals. While there is much rhetoric around these ideas and there are many practical ways to achieve this shift in focus, the overarching ideology is a good one: to empower the students with resources that enable them to learn better, based on their own data. Ken Robinson encourages students to find ‘their element’ [20], to find their own path and develop an innate love and genuine interest for their subject. This *learner empowerment* matches deep learning with research integration and *futures thinking* [22]. This is a two-way journey of a learner understanding themselves *personally*, their emotions, moods and sensations, and *externally* of the material world, events and circumstances.

So, how can we encourage and empower learners to become more conscious of their learning? How can we deepen learner’s understanding of what they have learnt? How can they adapt learning strategies to their own skill-set and style or see how well they are doing in comparison to others? How can they plan, manage and qualify their learning? How do we capture data – and what data – to help learners realize how they are progressing? How can we assist students using visualizations that effectively depict their progress, learning context and results?

\* e-mail: j.c.roberts@bangor.ac.uk

† e-mail: c.headleand@bangor.ac.uk

‡ e-mail: d.perkins@bangor.ac.uk

§ e-mail: p.ritsos@chester.ac.uk

Electronic proceedings of the IEEE VIS 2015 workshop  
*Personal Visualization: Exploring Data in Everyday Life*  
The authors remain the holders of the copyright

Our position is that *personal visualization* and especially *personal visual analytics* can help tackle these questions. Our focus and thoughts are born from our experiences within Higher Education, but we believe that the principles and ideas can be applied to a broader range of educational activities.

We organise the paper in four parts: (1) provide background that gives context to the personal data for learning; (2) highlight core related work; (3) focus on our experience and explain some key projects that we have been involved with, and (4) develop a vision for personal learning visualization. The aim is to motivate and excite the community to develop further the concept of *personal data visualization* for learning (Fig. 1).

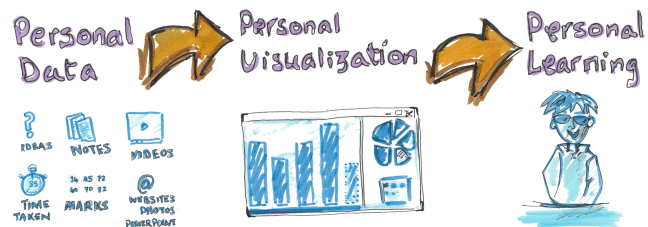


Figure 1: From personal data, personal visualization to personal learning.

## 2 BACKGROUND & CONTEXT

There are many kinds of student data that a University currently stores. These include: address, qualifications, degree major, modules or units of work to be achieved and grades. Students can view (and edit) some of this information via a ‘student portal’ and access learning resources through systems such as Moodle or Blackboard. These Learning Management Systems (LMS) provide students with access to individual grades, lecture materials, assessments and other resources the learner might require. Usually, the instructor has access to a more detailed overview of all students’ performance. It is likely, however, that grade information is stored in many places and students have to use several ‘portals’ (e.g., term-time assessment results may be stored on a separate portal to overall grades).

These systems provide a personal view, where an individual student only sees his/her grades. It is also true that these portals are becoming more feature rich and are starting to depict students grades visually. In fact, Universities are endeavoring to make the learning experience more personal, for instance:

- Strategies such as Personal Learning Plans (PLPs) are gaining popularity where students set their own goals, monitor their own progress and identify their own aims.
- Personal Learning Records (PLRs) are being introduced, where the students not only get a transcript of their grades but also receive a fully detailed document of all their subjects, learning outcomes, and extracurricular activities.
- Students are gaining more authority over their own data, and can record their own personal notes, which their tutor would receive.



However, there are several challenges with current systems. They do not capture detailed information about the student (especially information that may help them realize their own mistakes). The data granularity is poor, and only updates at major events (e.g., assessment milestones and examination results). Metadata such as notes by the student, academic and tutor, cannot be stored alongside the results. Existing systems are often not integrated and students and administrators have to utilize many different – often incompatible – tools. In order to reach a point where we can realize personal learning visualization we need to address many of these challenges.

There are obviously many different kinds of personal data that could be captured and stored. As we discuss below, at our institutions we capture attendance data, student satisfaction, and per-module evaluation data. But, it is unclear what other personal data may be useful in helping learners. In fact, some data that may seem ideal to capture, may be subject to privacy and legality restrictions.

For inspiration, let’s consider the field of exercise. Users have activity trackers like Fitbit or smartwatches (such as the Apple Watch), that capture personal data and display them through special services as daily or weekly reports of their progress. These services gamify the whole progress, by gauging how well users are doing in comparison with others. They indicate notable achievements or goals and offer digital ‘rewards’ (tags, badges etc.), that people can display on social networks, as encouragement and support. What is required for the learning domain is a similar approach: systems that automatically monitor progress, provide regular reports, and enable and encourage the student to do better.

### 3 RELATED WORK

In the context of this paper, and with an outlook to visualization, we look into related work in two dimensions: i) personal data capture and ii) personal learning visualization.

#### 3.1 Capturing personal data

Nowadays, *Educational Data Mining* (EDM) [21] and *Learning Analytics* (LA) [5] are converging. There is a shift from EDM merely informing educational practice to directly influencing it. LA is becoming a timely research topic and several researchers (including ourselves) make calls to the analytics communities to step up [4, 17]. Along with the enthusiasm of LA is the appetite for visualization (especially visual analytics). E.g., in 2015 the ‘Learning Analytics and Knowledge’ conference convenes its fifth event and holds its first workshop on Visual Approaches to Learning Analytics (visla15.org).

We already mentioned a growing interest in Personalized Learning Environments (PLEs), where a learner self-adapts their learning environment to suit their own particular learning needs [2]. We are also beginning to see data-driven systems personalized for the learner, such as recommender systems [8, 26] that help identify items such as learning resources and knowledge routes [9]. Powell and Yuan [29] argue that personal data can provide higher-level insights into how students learn. Nonetheless, while these data-driven insights are proving to be of benefit to educators, learners are often have limited access, constricting their possible reflections.

Looking at other domains, the tracking of personal data is becoming increasingly ubiquitous, helping individuals identify their habits and understand patterns in their behavior [24]. In recent years, with the advent of mobile and wearable [16] technology, and the Internet of Things [19], self tracking has become a disruptive innovation in an number of domains [25], including fitness [13] and sleep activity [7]. As these technologies become more and more pervasive, there are opportunities to employ them in educational context, in complementary fashion to traditional methods, and enrich the quality of captured data.



Figure 2: The IVY learning environment for language interpreters, showing a learning scene.



Figure 3: Picture from Technocamps, building Lego Mindstorm robots.

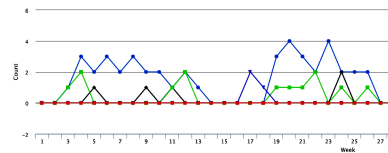


Figure 4: Attendance monitoring record of a typical student.

#### 3.2 Personalizing Learning Visualizations

The continuous increase in learning data (produced and captured) presents several visualization opportunities, in order to allow a student to reflect on their progress [6], compare themselves with their peers [23], provide greater context to acquired skills [11] or to build and depict predictive models [12]. However, there has been little research into how these visualizations can be personalized, to support a students learning style and individual needs.

Systems such as CourseVis [10], Moodle (moodle.org), Blackboard (blackboard.com) etc. all display the information in dashboards resembling the style of Google Analytics. About half of the tools discussed by Verbet et al. [27] display time spent, social interaction and tool use. In some cases students can customize data representations through these dashboard systems [3] enabling a level of self-reflection and greater understanding of personal performance.

Yet, for users to explore any information, they need to load a separate dashboards, often isolated from the learning materials and specific results and metrics (e.g., exam vs. term-time assessments). There is an opportunity to do more than merely displaying dashboards of grades, by integrating personal data, students’ learning goals, pathways and objectives.

### 4 CASE STUDIES

To develop the direction of personal learning data, we look to our own experiences for guidance, and briefly discuss four projects.

**The IVY project** [15] (Fig. 2) was funded by the European Union, and was a collaboration between six EU partners. The aim was to create a collaborative environment (The IVY-Virtual Environment) to train language interpreters. Several scenarios were developed, where students could learn interpreter-mediated communication skills, through being immersed in a 3D virtual world. Data of scenario usage, timings, repetition of loading materials etc. is captured. In the follow-up project (EVIVA) data stored included videos, recorded by the users as they used IVY-VE.

**Technocamps** (Fig. 3) was a four year outreach project designed to inspire the young people of Wales to select more Science Technology Engineering and Maths (STEM) subjects at school, and to encourage more school leavers to take STEM-based degrees or other employment. Hundreds of workshops were organized, most focused on constructionist approaches (learning by doing), using tools such as Lego Mindstorms, Minecraft, and computer programming in Scratch, Greenfoot, etc.

**Five Desing Sheets (FdS)** (fds.design) methodology, for lo-fidelity sketching and planning of visualization interfaces, has been developed and applied by the authors [18]. FdS has been used in summer schools, computer science third-year projects, and taught modules.

**Student monitoring** is a major challenge in Universities. In the UK there is a requirement by Border Agencies to monitor attendance of non-UK students. Using the OPN 2001 handheld barcode data collector we scan students' identity cards as they exit the lecture or class. This attendance data is then uploaded to a database and visualized in several ways. For instance, the information is displayed on each students' dashboard as a line plot (Fig. 4), and a visual summary of all students' attendance can be viewed in an administrator portal, for cohort monitoring.

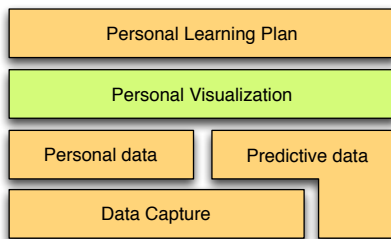


Figure 5: A tiered model, focusing on data capture, prediction, visualization. To allow students to develop Personal Learning Plans.

## 5 VISION FOR PERSONAL LEARNING VISUALIZATION

Our vision is to i) place visualization at the heart of the learning environment (rather than an afterthought), ii) put the user in control of their own learning experience, and iii) use a wide range of data as input. We propose a tiered approach where most data is captured, some is predicted, then visualized and used by students to develop their personal learning plan (see Fig. 5).

Let's consider Steve, an imaginary second year student on a computer science major. He has a set of modules that he has completed in his first year. This gives him a credit score and also a grade point average. He looks at a visualization of how well he did in the previous year, and via an interface explores his activity in specific modules. He sees his notes and realizes that he put less time into certain modules. He then makes a comparison of his results to other students, (using a map-like visualization, Fig. 6). In the visualization, he can click on better students (depicted anonymously) and explore their grades and progression paths. He then turns his attention to next year.

Looking at a visualization of predicted results, he can see how his module grades could turn out. Through this information the system helps him create a Personal Learning Plan. The plan includes visualizations of his current grades, along with courses to help him improve on certain skills. Steve does not understand some statistical analysis that was used, and refers to expanded tool-tips providing additional information, including explanatory animations. The visualizations nudge Steve to register for several additional courses, which focus on key skills that he identified he is lacking.

But Steve would not use the system unless the data can be easily captured. Data gathering must be seamless and transparent with no extra effort. We have had first hand experience of this with

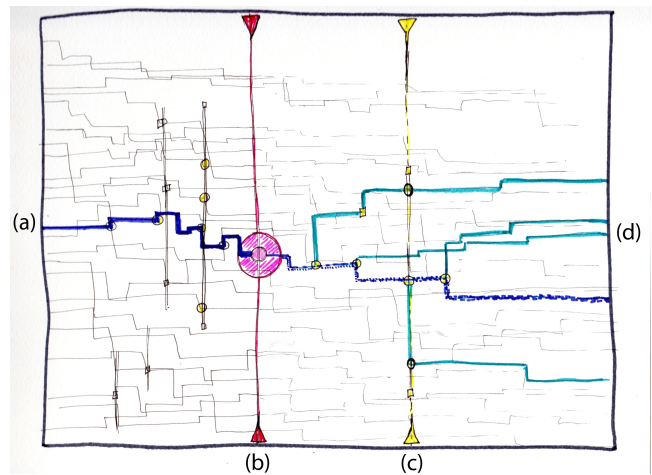


Figure 6: A sketch of a proposed students' journey. Red shows the current time, possible routes are depicted in cyan, and key moments are shown with circles.

our attendance monitoring. Originally paper-based monitoring was used. But academics saw it as a burden, and mistakes happened frequently. Nonetheless, with the scanner we are able to automate much of this process. The visualizations and the underpinning database help tutors and students see their record (and can make requests for correction should they be legitimately miss the class). It is important, therefore, that data capture is transparent and seamless, very much as dictated by ubiquitous computing [28].

There are certainly similarities between the gamification of the fitness world and our vision. Users can see their own exercise records, set goals and have competitions with their peers. This is a useful comparison because both domains have goals (get fitter, learn better) and both set targets (lose weight, improve marks). However, there are also major differences and potential pitfalls with this comparison. For instance, a significant factor is the time frame under which we observe change. Fitness has metrics that update on a minute or second scale and fairly short term goals, whereas education has (typically) longer term goals with longer term metrics. In addition, tracking mechanisms (e.g., wearables) for fitness may have matured, however similar educational monitors are less mature, and maybe (if they are intrusive) more controversial.

Another important consideration is the relative value of the information to the user and how it will impact on their lives. In 2013, Rackspace [14] published a model with six classifications of the way in which the user would regard wearable data. Over 80% of the 4000 respondents claimed that the wearable technology boosts their personal abilities. If a similar impact can be achieved within education there is an opportunity to make the collection and visualization of personal learning data have a profound influence in the domain.

Moreover, there is need to capture and visualize a wide range of data, e.g., student's or tutor's notes, observations, institution-related metrics such as attendance and National Student Survey (NSS) records, and grades of different types of assessment. In addition, HCI techniques [4] can be employed to track how students access educational content, how much time they spend studying it etc. Multimodal techniques (e.g., eye and gesture tracking, biosensors, etc.) could be used to devise more naturalistic assessments, giving insight to "overt and tacit" aspects that indicate a student's expertise on a subject [1].

Finally, the visualizations themselves need to have a profound impact. They should *nudge* or change student behavior in a positive way. They could act as suggestions or demonstrations of spe-



cific milestones within the learning journey. In fact, we imagine that the aforementioned visualization, using a map-metaphor, would act as an important reference. It would allow users to see their *learning journey*, how well they have done (compared to others), give suggestions over possible outcomes, and allow them to drill down into specific marks (see Fig. 6). This concept is inspired from the sketchbooks appearing in the Adjustment Bureau (2011) film. In our sketch, individual lines represent the progress of a student. Students' grades are shown as black lines, with time running on the x-axis. The y-axis represent an aggregated score of metrics (such as grades, time spent, notes taken etc.) Thus the flow of these lines can go up (generally doing better) and down (doing worse). The red line and circle show 'today', with the yellow bar for the next Semester and the cyan routes for grade predictions.

## 6 CONCLUSIONS

There is much research to be done to achieve our vision. Not only do we need easy ways to capture personal data, but data must be captured sensitively. There are three enabling factors that need to mature, to make personal learning a pervasive idea. Researchers need to:

- investigate good ways to capture and deliver personal data.
- develop personal visualizations, that are suitable for an individual and their needs, as well as the learning context.
- integrate personal planning and development strategies. Students need ways to plan and manage their learning, monitor their development, and take charge of their own learning progress.

In our particular case, students also have to see a clear benefit of these personal visualizations, such to persuade them to put extra effort into storing and manipulating personal data. There are also other potential challenges, where some students may feel that this is too much of an intrusion into their own personal life and activities. Consequently, data security and anonymization must be a priority.

New visualization techniques are required to display this information. The created designs must be understandable by users who may have little knowledge of specific visualization techniques. We also believe that the visualizations created need to be first-class-citizens, i.e., when these tools are created, developers should first think how the information is going to be displayed, and how a user is going to manipulate it. Too often the display of such data is an afterthought.

Personal visualization is a growing sub-area of visualization. While there are certainly many challenges ahead, there are also many benefits and opportunities. Personal learning is gaining momentum, and therefore we encourage researchers and academics to think how personal visualization can aid to achieve and enhance personal learning at whatever level of education, whether School, College or University.

## REFERENCES

- [1] P. Blikstein. Multimodal learning analytics. In *Proc. Learning Analytics and Knowledge*, LAK '13, pages 102–106. ACM, 2013.
- [2] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs. A reference model for learning analytics. *IJTEL*, 4(5-6):318–331, 2012.
- [3] A. Cuttone, M. K. Petersen, and J. E. Larsen. Four data visualization heuristics to facilitate reflection in personal informatics. In *Universal Access in Human-Computer Interaction*, pages 541–552. Springer, 2014.
- [4] A. Dix and J. Leavesley. Learning analytics for the academic: An action perspective. *J-UCS*, 21(1):48–65, jan 2015.
- [5] R. Ferguson. Learning analytics: drivers, developments and challenges. *IJTEL*, 4(5-6):304–317, 2012.
- [6] S. Govaerts, K. Verbert, J. Klerkx, and E. Duval. Visualizing activities for self-reflection and awareness. In *ICWL*, pages 91–100. Springer, 2010.
- [7] M. Kay, E. K. Choe, J. Shepherd, B. Greenstein, N. Watson, S. Con-solvo, and J. A. Kientz. Lullaby: a capture & access system for understanding the sleep environment. In *Proc. Ubiquitous Computing*, pages 226–234. ACM, 2012.
- [8] N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel, and R. Koper. Recommender systems in technology enhanced learning. In *Recommender systems handbook*, pages 387–415. Springer, 2011.
- [9] N. Manouselis and D. Sampson. Dynamic knowledge route selection for personalised learning environments using multiple criteria. In *Workshop on Intelligence and Technology in Educational Applications (ITEA2002)*, number 1, pages 448–453. IASTED, 2002.
- [10] R. Mazza and V. Dimitrova. Coursevis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *Int J Hum-comput st.*, 65(2):125 – 139, 2007.
- [11] A. Nussbaumer. Supporting self-reflection through presenting visual feedback of adaptive assessment and self-evaluation tools. In *Proc. Interactive Computer-aided Learning (ICL 2008)*, pages 24–26, 2008.
- [12] A. G. Picciano. The evolution of big data and learning analytics in american higher education. *J Asynchronous Learning Net.*, 16(3):9–20, 2012.
- [13] S. Purpura, V. Schwanda, K. Williams, W. Stubler, and P. Sengers. Fit4life: the design of a persuasive technology promoting healthy behavior and ideal weight. In *Proc. SIGCHI Human Factors in Computing Systems*, pages 423–432. ACM, 2011.
- [14] RackSpace. The human cloud: Wearable technology from novelty to production. Company Website, June 2013.
- [15] P. D. Ritsos, R. Gittins, S. Braun, C. Slater, and J. C. Roberts. Training Interpreters using Virtual Worlds. *Trans Computational Science XVIII*, 7848, May 2013.
- [16] P. D. Ritsos, J. W. Mearman, A. Vande Moere, and J. C. Roberts. Sewn with Ariadne's Thread - Visualizations for Wearable and Ubiquitous Computing. In *Workshop, Death of the Desktop Workshop, IEEE VIS 2014, Paris, France*, November 2014.
- [17] P. D. Ritsos and J. C. Roberts. Towards more Visual Analytics in Learning Analytics. In M. Pohl and J. Roberts, editors, *EuroVis Workshop on Visual Analytics*. The Eurographics Association, 2014.
- [18] J. C. Roberts, C. Headleand, and P. D. Ritsos. Sketching designs using the five design-sheet methodology. *IEEE Transactions on Visualization and Computer Graphics*, 2015.
- [19] J. C. Roberts, P. D. Ritsos, S. K. Badam, D. Brodbeck, J. Kennedy, and N. Elmqvist. Visualization beyond the desktop—the next big thing. *CG&A, IEEE*, 34(6):26–34, Nov 2014.
- [20] K. Robinson. *Finding your element: How to discover your talents and passions and transform your life*. Penguin UK, 2013.
- [21] C. Romero and S. Ventura. Educational data mining: a review of the state of the art. *Trans. SMC, Part C: App and Rev.*, 40(6):601–618, 2010.
- [22] A. Ryan and D. Tilbury. Flexible pedagogies: new pedagogical ideas. *Higher Education Academy, London*, 2013.
- [23] J. L. Santos, S. Govaerts, K. Verbert, and E. Duval. Goal-oriented visualizations of activity tracking: a case study with engineering students. In *Proc. Learning Analytics and Knowledge*, pages 143–152. ACM, 2012.
- [24] M. Swan. Sensor mania! the internet of things, wearable computing, objective metrics, and the quantified self 2.0. *JSAN*, 1(3):217–253, 2012.
- [25] M. Swan. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, 1(2):85–99, 2013.
- [26] K. Verbert, H. Drachsler, N. Manouselis, M. Wolpers, R. Vuorikari, and E. Duval. Dataset-driven research for improving recommender systems for learning. In *Proc Learning Analytics and Knowledge*, pages 44–53. ACM, 2011.
- [27] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. L. Santos. Learning analytics dashboard applications. *American Behavioral Scientist*, page 0002764213479363, 2013.
- [28] M. Weiser. Some computer science issues in ubiquitous computing. *Commun. ACM*, 36(7):75–84, July 1993.
- [29] L. Yuan, S. Powell, and J. CETIS. Moocs and open education: Implications for higher education. JISC CETIS, WP01.

# Enabling Self-Awareness for Knowledge Workers Through Visualization of Instrumentation Data

Sidharth Thakur\*  
RENCI, Chapel Hill, NC

Paul Jones†  
North Carolina State University,  
Raleigh, NC

Steve Cox‡  
RENCI, Chapel Hill, NC

## ABSTRACT

Computer activity logging tools can provide useful high-level overviews of a user’s computer usage and interaction patterns, but currently available tools usually lack specificity and can deliver only limited insights. Instrumentation tools that can capture and visualize richer activity-related data can especially benefit knowledge workers in exploring their personal context and activity histories while they undertake information analysis tasks. Moreover, providing users access to their personal activity data through engaging and effective visualizations is an incentive for them to participate in data gathering efforts. We are undertaking an effort to instrument computer workstations of knowledge workers to gather digital data to study the process of information analysis. As part of this effort we have developed visualizations to enable our user base to explore their instrumentation data and gain awareness about their personal context. This has allowed them to gain actionable insights can improve their working patterns.

**Index Terms:** H.5.2 [Information Interfaces and Presentations (e.g., HCI)]: User Interfaces—Screen design (e.g., text, graphics, color); K.8.m [Personal Computing]: Miscellaneous—

## 1 INTRODUCTION

Understanding how people interact with and use computers has been fundamental to creating more effective human-computer interfaces and technologies. A standard approach is to capture and analyze activities and actions of people during their interaction with computer interfaces, such as during casual interactions and during complex tasks such as gaming and information search and analysis. While much effort has been expended on studying behavior and requirements of people engaged in those complex activities, there is a growing need to develop tools that enable users to objectively explore their own actions and behaviors and gain insights about their personal context [1].

One class of computer users who can particularly benefit from insights derived from collecting, visualizing, and analyzing personal activities are *knowledge workers*. Very generally, a knowledge worker is someone who needs to routinely deal with, analyze, and make decisions based on complex data and information. A knowledge worker can be a student, researcher, clinician, data analyst, or anyone else who needs to derive sense from complex information.

A motivating example for instrumenting a knowledge worker’s activities is provided by insights from Stephen Wolfram about his own activities from nearly two decades of personal data on email, keystrokes and phone conversations [2]. Few people can collect such comprehensive data but there are many benefits of doing so.

\*e-mail:sthakur@renci.org

†e-mail:pjones@ncsu.edu

‡e-mail:scox@renci.org

Electronic proceedings of the IEEE VIS 2015 workshop  
*Personal Visualization: Exploring Data in Everyday Life*  
The authors remain the holders of the copyright

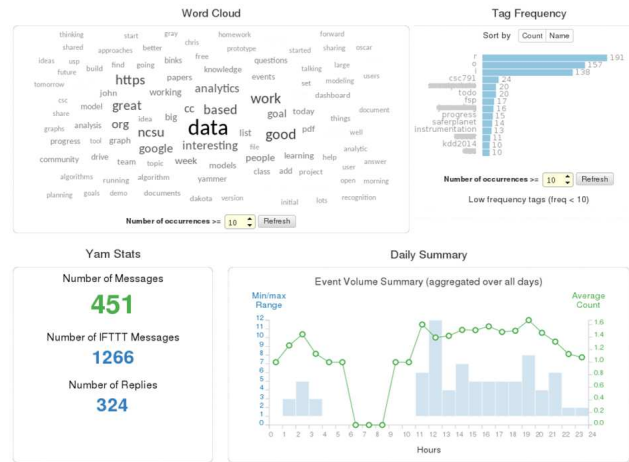


Figure 1: Visualizations of a user’s personal activity data from their organization’s social network platform (Yammer).

For example, such data can provide interesting insights about work-life balance and routines. These insights can be helpful in understanding one’s personal context and possibly even detecting and avoiding non-productive or unhealthy behaviors [3].

While activity logging and visualization tools are common, many lack specificity in information communicated or are too high-level to provide detailed, actionable information. Moreover, capturing diverse data and at large scale can be prohibitive. To address these limitations, we have developed a computer instrumentation platform that captures event-based data from personal workstations using plug-ins called *instrumentation agents*. The data contain information about applications used, messages created, and, on a higher level, tasks and actions performed. The agents capture detailed and specific information about user actions that are usually not made available by standard activity logging tools.

In this paper we describe visualizations we are developing to enable knowledge workers to explore their personal instrumentation data. The visualizations provide self-awareness, but are also a great incentive for our user base to allow us to instrument their computers. Main contributions of this work are as follows:

- We outline personal context for a knowledge worker. We use the dimensions developed by Huang et. al. [1] to describe and discuss our usage scenario.
- We discuss novel visualizations we have developed to explore personal activity data and discover interesting patterns such as trends and outliers in activity patterns.

## 2 BACKGROUND

The landscape and scope for monitoring oneself is vast. Popularization and wide availability of monitoring devices and applications on mobile devices have spurred movements such as *Quantified Self* [4]. Such efforts seek to build holistic, pervasive views of one’s

Event Protocol			
Userld	AppName	SysId (optional)	time
Unique id of user (analyst)	Name of application accessed	Id of computer system	Time stamp when server received event
ProjId	ProjVer	NetAddr	GeoLoc
Project id to distinguish data source	Version of protocol to track changes	Network / IP address	Physical location
EvtTime	EvtEndTime	EvtType	EvtDesc
Time when event was generated on user's workstation	End time stamp	Type of event (e.g., text, file access, low level such as mouse move)	Details of event

Figure 2: The core fields in our ‘event protocol’ data exchange format, which represent user activities from instrumentation agents. Optional fields are shown in gray; mandatory fields in black.

activities and focus broadly on health, physical workout, and social aspects. Our focus is more on a specific context that involves providing knowledge workers self-awareness about their activities when handling digital data.

Many software tools and applications are available that capture data on computers and provide personal histories and insights about a user’s activities. Examples include *RescueTime* [5] and *DeskTime* [6]. These types of tools attempt to guess how productive a user is being based on their application usage and web-browsing habits (using a dictionary of apps and sites classified into categories). One challenge is that such tools use a narrow definition of productivity and information presented to user generally lacks context. There has of course been vast research on computer-human interactions and behavioral patterns during computer activities but the bulk of past work has focused primarily on developing better user models and on improving understanding of user actions from a research perspective.

In this work we seek to address niche requirements of tools and applications that can provide knowledge workers better self awareness through instrumentation of their computer activities.

### 3 A KNOWLEDGE WORKER’S PERSONAL CONTEXT

Recently Huang et. al. [1] outlined a need for developing visualizations and tools focused specifically on *personal context*. According to the authors, personal context of an individual deals with requirements, motivations, and scenarios that differ from those that commonly occur within a professional context. Thus, personal context is focused on an individual’s interests, and not limited to personal data, though visualization for personal context is inclusive of analysis and visualization of personal data.

Our overall data instrumentation effort is focused on studying tradecraft of information analysis and we are also interested in developing tools that help our user base gain self-awareness of their activity patterns, habits, and context of their actions. Motivations for exploring personal data in our situation are multi-fold. For example, some users might want to analyze their work/life balance and habits based on their activity data, while others might be curious to analyze their productivity and how they compare to other users in their organization.

We briefly analyze our domain requirements based on a taxonomy proposed by Huang et. al., which they produced by analyzing literature on personal visualization and analytics. The discussion below lists the four main dimensions of personal context and their sub-dimensions (in square brackets). Levels of sub-dimensions from the taxonomy are shown in *italics*.

- **Data** [Scope, Effort, Agency]: **Scope:** Scope of our domain is

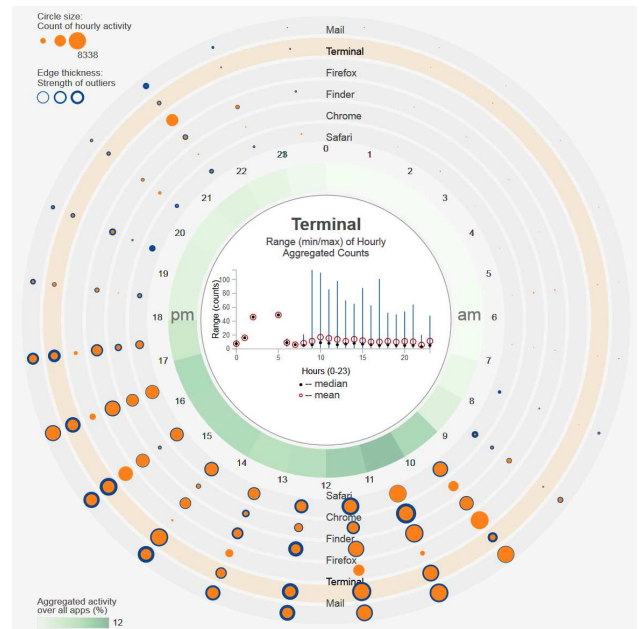


Figure 3: Visualization of aggregated hourly usage (orange bubbles) and unusual patterns (dark edges) for common applications used by a single person. The central chart shows hour-wise time series and basic statistics for the ‘Terminal’ application.

*community* of knowledge workers in our organization, which includes students, academic and industry researchers. We primarily gather event-oriented computer activity data generated during high-level interactions and activities on computers, such as application usage and web browsing. **Effort:** Effort expended can vary; some data are collected through background software agents (*sensors*), while other types of data can require significant participation on part of the user (*manual* entry of data). **Agency:** All data we collect, with exception of studies involving formal human subjects, is available to our user base.

- **Context** [Design context, Settings, Influence context]: **Design:** The tools and visualizations were designed through *participatory* collaboration with our user base. **Setting:** Currently the data are available only to our user base in a *shared, public* mode. **Influence:** The application is intended both for researchers studying knowledge workers and individual users who aim to gain self-awareness through their data.
- **Interaction** [Attentional demand, Explorability] **Attentional demand:** is *low to moderate* depending on complexity of visualization. **Explorability:** is *medium to high* based on features implemented in visualization and query interface in our data portal.
- **Insight** [Actionability, Automated analysis] **Actionability:** can vary in the range [*low, medium, high*] depending on complexity of visualizations. **Automated analysis:** We use variety of data mining and analysis tools.

### 4 INSTRUMENTATION OF PERSONAL ACTIVITY DATA

We have created a novel, versatile instrumentation framework to gather computer activity data [7]. Our goal is to instrument computer workstations used by knowledge workers to address challenges in gathering, processing, and analyzing activity log data to improve information analysis tradecraft.

Our instrumentation system collects user activity data using *instrumentation agents*, which are software applications and plug-ins deployed on workstations or through web browsers to capture specific types of information. For example, one agent captures us-

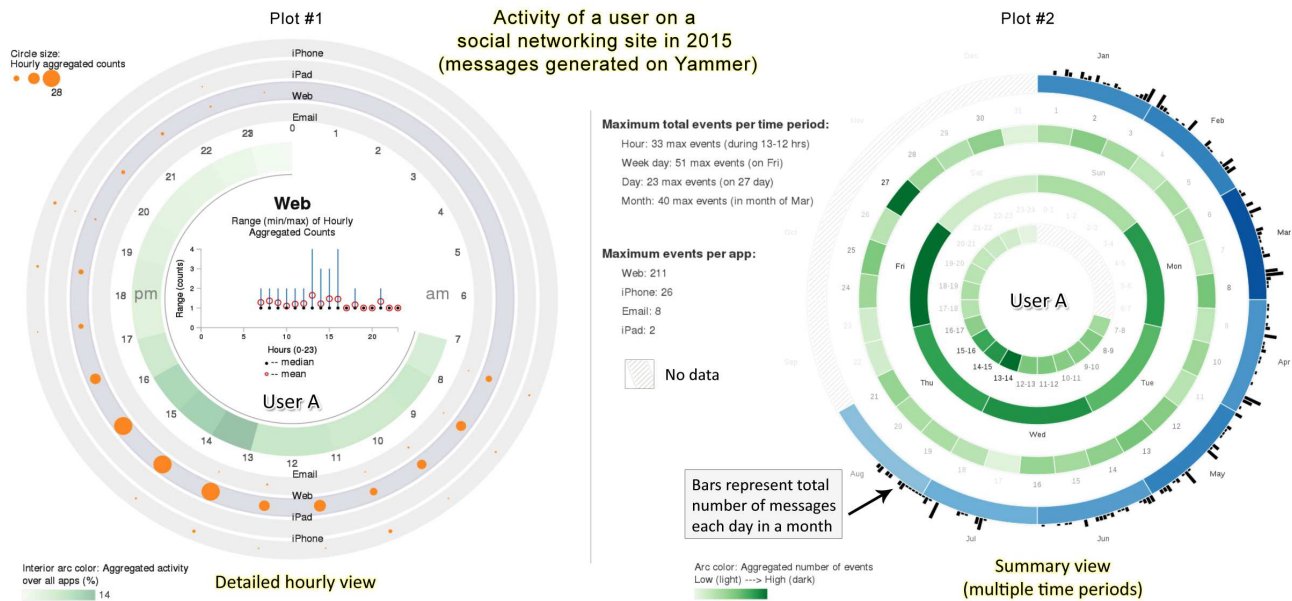


Figure 4: Visualizations showing a user's history of generating messages through their organization's social networking application (Yammer).

age of applications on Mac OS (OSX) computers, while another monitors web usage and URLs visited. Other agents capture user-generated data from third-party servers. These external data sources include messages published on an organizational social networking site (Yammer) and issues related to software engineering efforts (recorded in GitHub).

All the disparate data are converted into a standard key-value event format before being stored in a back-end data store. **Figure 2** shows the core part of our event format and some of the information we capture. The event protocol can be extended as needed by instrumentation agents. The data are then made available back to our user base through a query interface and custom visualizations.

## 5 VISUALIZATIONS OF PERSONAL ACTIVITY DATA

### 5.1 Personal Timelines (TACO)

We depict a user's aggregated data using an interactive plot that we call TACO (Timeline visualizations of Aggregated Client Observations). The plot uses a circular layout to show activities for recurring time periods such as a day, week, or month.

**Figure 3** is a TACO plot that displays events generated by a user while using most popular applications on a Mac OS computer. Each track or ring corresponds to one application, and variable-size bubbles indicate relative frequency of usage of the application. The plot shows hourly aggregated counts from the user's entire historic data. A additional plot in the center shows a time series and basic statistics for the selected application. Options are provided to normalize bubble sizes by maximum over all applications to see relative differences.

The plot depicts unusual or non-normal activity for the aggregated hourly rates to show whether an application was used more or less than usual. We measure unusual activity using the standard  $z$ -score for computing outliers. Here, for all historic values of event count for a given hour we compute how far an observed value is from the mean for that hour. The thickness of dark edges around bubbles in **Figure 3** is used to convey the strength of the outliers.

**Figure 4** shows a set of plots that visualize a user's pattern of generating messages within their organization's social networking site called Yammer over the course of a year (Jan - Aug 2015). The TACO plot on the left shows aggregated hourly rates of messages broken down by different application platforms (email, web app,

and mobile). The plot on the right in **figure 4** shows cumulative rates for all messaging platforms aggregated according to different time granularities. Darker shades of colors in the plot represent periods of high activities. The bars around the outside periphery of the plot show daily counts of messages over the different months.

### 5.2 Personal Vis Dashboard

We have also implemented a visualization dashboard that provides contextual information from a user's instrumentation data. The idea is that when a user logs into our central system, they are shown a dashboard with plots populated by their personal data. Unlike traditional dashboards, such as *RescueTime*, which provide mostly historic timeline views of user's activities, our dashboard provides more contextual information about user's actions drawn from their instrumented data.

The dashboard contains a number of different types of plot. For example, the set of figures on the right in **Figure 1** display personal metrics such as a user's word cloud and frequency of hash tag terms from the user's text-oriented data such as Yammer or GitHub messages. Other useful metrics such as rate of messaging, conversation group, and size of messages for the user are also shown.

Another plot, shown in **Figure 5**, displays a user's personal event volumes using a calendar-based visualization layout. We employ sparklines to show daily event volumes and potentially entire history of activity for a user in a single chart.

An interesting tool to explore personal application usage patterns is an 'app graph', shown in **Figure 6**. The graph displays common applications used by a user and transitions between those applications. The graph is part of a study that is investigating how people develop and use tools with workflows.

## 6 DISCUSSION AND CONCLUSIONS

Our main goal with this work is to provide knowledge workers with self-awareness that will lead to a number of *actionable insights*. Some examples of such insights that we are able to provide include:

- understanding of current personal **working patterns** and working pattern histories. This is useful both to determine when someone is likely to be available (in the office, or for teleworking), and also when they might be more likely to be



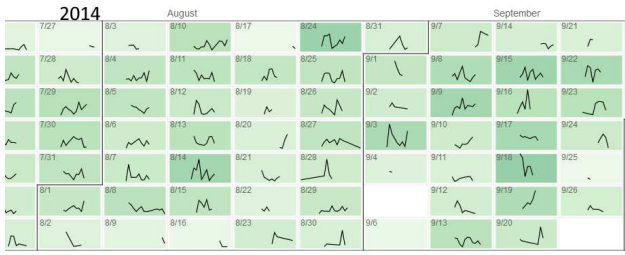


Figure 5: A partial view of a calendar-style visualization showing activity of a single user using sparklines of daily volumes.

interrupt-able, since we can see what types of work they are doing at any given time. These insights can help knowledge workers and their managers schedule meetings at appropriate times.

- generation of **activity summaries** for individuals, and also for their teams and organizations. Self-awareness metrics, when combined with visualizations such as those described in this paper, can be used to automatically generate content for progress reports, and perhaps even for performance appraisals, although clearly much care is required for this.
- usage of specific applications, which may require particular resources to be available. For instance, use of the *Terminal* application on Mac OS may indicate that someone is about to access server infrastructure, and perhaps may be about to run a computationally intensive task. This kind of information is useful for **pro-active resource allocation**.
- indications as to when people tend to get distracted by non-productive tasks. We are considering options for constructing a **focus metric** that measures the level of *flow* [8] that a knowledge worker might be experiencing. In a given time period (say one hour), some variables that could be used to construct such a focus metric include:
  - time spent active in applications (or working with documents) that are classified as *productive*,
  - frequency of transitions to other complementary / supplementary applications that continue the flow of productivity,
  - frequency of transitions to short-lived, or *distracting* applications. Reacting to e-mail (or other desktop notifications) might fall into this category.

The intuition here is that ‘100% focus’ could be approximated as a scenario in which a knowledge worker spends the entire time period active in one application that is classified as productive and with no transitions to other applications, or where all transitions are to other productive applications. Conversely, a low focus metric should be obtained when little time is spent in productive applications, or when there are many transitions to distracting applications. Exposing such a metric to users, even if it is a very crude approximation, could allow people to take action to block distractions during certain times, or to find ways to better deal with them.

To summarize, we have described several visualizations that allow knowledge workers to very quickly obtain self-awareness of their working patterns and personal context. When combined, these enable users to gain detailed insights so they can take actions to improve their working and personal lives. Our visualizations are enabled by careful instrumentation of user workstations and subsequent analysis. Research in this space is highly active in our group, and is focusing on ways to measure a person’s information analysis workflow, and to find ways in which this might be improved.

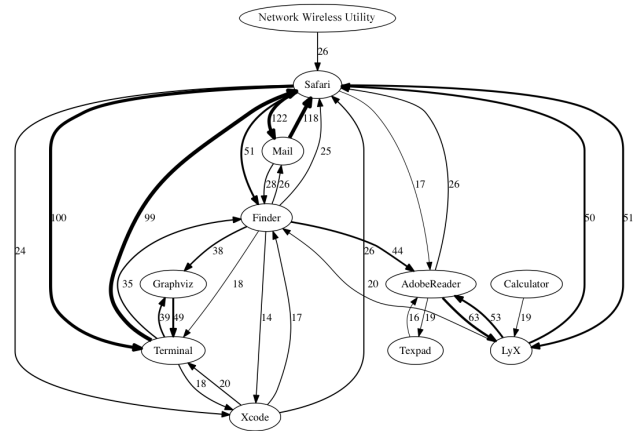


Figure 6: A directed activity transition graph from our Mac OS instrumentation agent providing preliminary insight into tool usage patterns. Thicker lines represent a larger number of transitions between pairs of applications (the number of transitions is shown alongside)

#### ACKNOWLEDGMENTS

This material is based upon work supported in whole or in part with funding from the Laboratory for Analytic Sciences (LAS). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the LAS and/or any agency or entity of the United States Government. The authors would like to thank staff at the LAS for providing funding and inspiration for much of this work.

#### REFERENCES

- [1] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury, “Personal visualization and personal visual analytics,” *IEEE Transactions, Visualization and Computer Graphics*, vol. 21, pp. 420–433, March 2015.
- [2] S. Wolfram, “The personal analytics of my life.” <https://blog.stephenwolfram.com/2012/03/the-personal-analytics-of-my-life/>, 2012. [Online; accessed 20-Aug-2015].
- [3] I. Li, A. Dey, J. Forlizzi, K. Höök, and Y. Medynskiy, “Personal informatics and hci: Design, theory, and social implications,” in *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, (New York, NY, USA), pp. 2417–2420, ACM, 2011.
- [4] D. Nafus and J. Sherman, “Big data, big questions— this one does not go up to 11: The quantified self movement as an alternative big data practice,” *International Journal of Communication*, vol. 8, p. 11, 2014.
- [5] RescueTime, “RescueTime.” <https://www.RescueTime.com/>, 2014. [Online; accessed 11-July-2014].
- [6] DeskTime, “DeskTime.” <https://www.deskttime.com/>, 2014. [Online; accessed 11-July-2014].
- [7] P. Jones, S. Thakur, and S. Cox, “A versatile platform for instrumentation of knowledge worker’s computers to improve information analysis,” 2015. [Unpublished; under review].
- [8] M. Csikszentmihalyi and M. Csikszentmihalyi, *Flow: The psychology of optimal experience*, vol. 41. HarperPerennial New York, 1991.

# Evaluating the Effectiveness of Three Physical Activity Visualizations — How People Perform vs. Perceive

Xin Tong<sup>1</sup>, Diane Gromala<sup>1</sup>, Lyn Bartram<sup>1</sup>, Fateme, Rajabiyazdi<sup>2</sup>, Sheelagh Carpendale<sup>2</sup>

<sup>1</sup>Simon Fraser University

<sup>2</sup>University of Calgary

## ABSTRACT

In this paper, two different types of personal physical activity data visualizations – a Circular Ringmap and a Virtual Pet visualization – are proposed and compared to a widely used commercial approach, the *FitBit* bar chart. A study was conducted to evaluate the effectiveness of the three visualizations based on participants' task performance and self-reported experience. The results showed that in objective terms, there were not significant differences in participants' time spent on the task and their error rates across the three visualizations. However, in subjective terms, participants did perceive the visualizations differently. The pragmatic bar chart won on several attributes (readability, comprehension, and awareness), and the Virtual Pet visualization triggered an emotional response (empathy). Although the Circular Ringmap visualization helped participants with identifying activity patterns, it was not perceived to be intuitive or helpful overall. From this study, first, we learned that how people perform a task does not equal how participants perceive. Second, decisions in designing specific visualizations related to personal data should depend on each individual's goals and contexts. Last, the role of emotion, which was triggered by the design of visualizations, is discussed.

**Keywords:** Personal Informatics, data visualization, effectiveness, physical activity, emotional feeling.

**Index Terms:** K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

## 1 INTRODUCTION

Visualizations are crucial in personal data analytics to help people gain insights, set goals, and self-reflect on their physical activity [1]. By using visualization tools, one's awareness of physical behaviors increases [3], which can help them to make smarter decisions and to support them in changing their behavior in regards to their personal health.

Visualizations designed for personal information that relate to physical activity should facilitate the process of understanding data and finding activity patterns. New methods of visualizing personal information [2] have been designed in number of research projects and activity systems. For example, *Data in Everyday Life* [7] represents data in a form of a digital calendar; *QS Spiral* shows the top of a clock dial metaphor with a circle representing a time span [3]; *Fish'n'Steps* [4] uses a social computer game to promote physical activity; and the virtual plant

in *UbiFit* [2] visualizes a garden that grows according to owners' physical activity.

Many commercial wearable devices are lowering the effort that individuals need to exert to collect and understand their data. To better engage people, *FitBit* visualizes data in multiple ways – including a bar chart (Figure 1, top) and an animated growing flower. But it is still unclear whether people can effectively identify and understand data presented in different ways, what people subjectively prefer in regards to such visualizations and the degree to which subjective preference can play a role. Further, few studies compare the objective findings with subjective experience.

While it may be sufficient for visualizations designed for use in work environments to focus on studying comparative time and efficiency, this may not be the only important factor when creating visualizations for our everyday lives. For example, while being accurate and efficient may be both necessary and sufficient for work, activities in one's personal life may be complex and affected by many other factors, such as emotional response. Therefore, in this study, we investigated people's performances in a visualization task and then explored their self-reported evaluations of visualization attributes – Readability, Comprehension, Awareness, Attractiveness and Empathy.

## 2 VISUALIZATION DESIGN

In this section, we developed two approaches of visualizing personal physical activity data, the Ringmap and the Virtual Pet visualizations. In this study, we then compared these two visualizations to the *FitBit* bar chart visualization.

*FitBit* has a commercial dashboard for people to access their personal health data online. Most of the personal data is displayed by bar charts (Figure 1, top). People can switch to different data tabs in the dashboard to assess their steps taken, distance, floors or calories burned. The interface is interactive in order to display the daily summary, and adjusted to show various ranges of time or specific periods of time.

The Ringmap visualization is a circular structure that represents one month of time (Figure 1, middle). It traces and tracks variable processes of activity data and is based on our prior visualization design – *Life Cycle* [5]. Each ring represents a day starting on the top at 12:00 am. The time is arranged like a clock and reads clockwise; the only difference is that this visualization covers an entire day rather than 12 hours. Each ring is divided into equal slots of 5 minutes. The physical activity levels are mapped onto each slot according to brightness. The brighter the slot, the more active the person was in those five minutes. When the mouse is moved over a time slot in this visualization, the exact physical activity data and time is displayed.

The particular approach that influenced the Virtual Pet visualization was motivating behavior changes by cultivating a strong internal awareness of one's physical activity through focusing on the care of pets or plants. Figure 1 (bottom) shows different physical conditions by date in a calendar view. It categorized physical activity data into five levels: extra high, high,

<sup>1</sup> {tongxint, gromala, lyn}@sfu.ca

<sup>2</sup> {f.rajabiyazdi, sheelagh}@ucalgary.ca



medium, low, and extra low. Those categories were then mapped onto the virtual pet's healthy condition or the virtual plant's growth as depicted in Figure 2. The actual five figures depicted in this visualization were designed by mapping a person's approximate activity level to possible real life scenarios. For instance, being a couch potato stands for a low level of activity, whereas being a super hero suggests an extremely high or "extra high" level of activity.

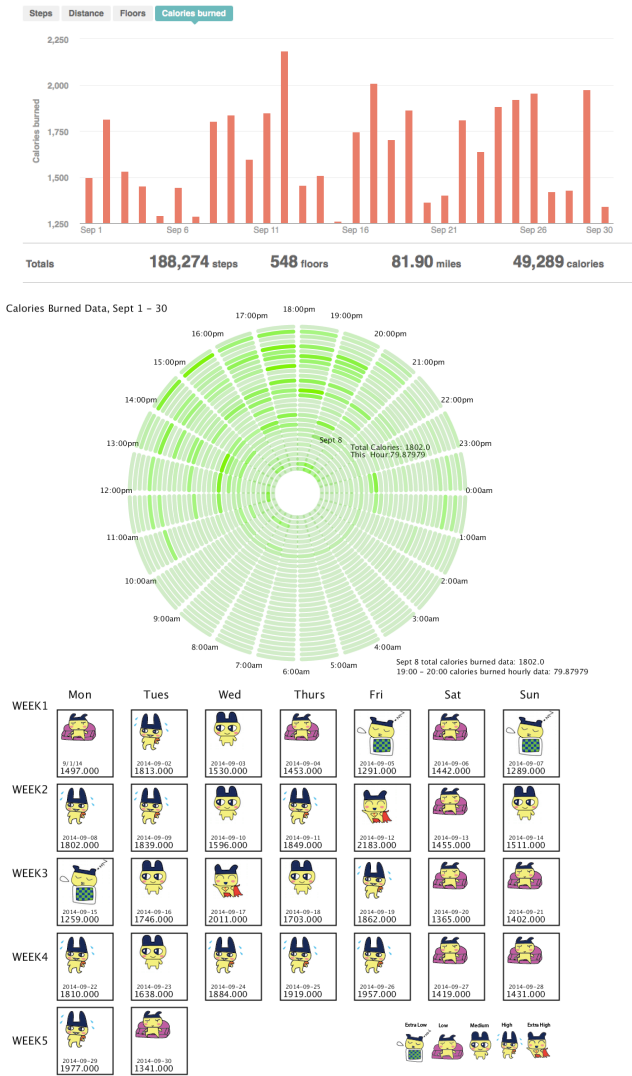


Figure 1: Three Visualizations Representing Physical Activity in the Study, Showing One-month Data Visualized in a *FitBit* Bar Chart (top); a Circular Ringmap (middle); and a Virtual Pet Calendar (bottom).



Figure 2: The actual five figures used in the Virtual Pet visualization

### 3 THE STUDY

The goal of this study was to compare the effectiveness of the bar charts, the Circular Ringmap, and the Virtual Pet calendar visualizations, and to understand how they may influence peoples' experience.

Eighteen participants (8 males and 10 females, aged from 19 to 29) were recruited among university students. Six participants each were randomly assigned to three groups; each group was provided with one visualization type. The same data set was used for all three visualization types. Next, a quantitative experiment was conducted, measuring between-groups. The definitions of five attributes were given to participants as a reference for their ratings of the visualization they interacted with in the task. The five attributes were:

- (1) Readability: to what extent could they read the visualization data.
- (2) Comprehension: to what extent did they understand what the visualization data represented in order to identify different activity levels.
- (3) Attractiveness: to what extent were they attracted by the graphic design in terms of formal qualities (color, hierarchy, data organization methods and so on), as well as the metaphors and the data mapping.
- (4) Awareness: how easy or difficult was the visualization for them to self-reflect on the activity conditions (to distinguish whether the activity level was at a high or low level in certain time periods) represented by the data visualization, and in answering the questions posed by the task.
- (5) Empathy: to what extent did the visualization evoke an emotion or feeling.

The study consisted of 3 sections: the pre-test, visualization task and post-test. Each section had a questionnaire, and the data was collected from these questionnaires. The main types of questions that were asked in the task were to: (1) Identify the data's range in a specified day in the month; (2) Identify the activity levels of a specified day in the month (from five categories: extra high/high/medium/low/extra low); (3) Identify if the activity level of the specified day or week is considered enough to maintain health (the definition of "healthy" in this task: over half of the days in a month has an activity data that is above average); (4) Explore the activity pattern presented in the visualization and when it occurs.

For the three groups, as shown in Figure 3, the time spent on task (in minutes) for the *FitBit* bar chart was ( $M = 8.17$ ,  $SD = 2.43$ ); for the Circular Ringmap ( $M = 7.89$ ,  $SD = .88$ ); and for the Virtual Pet ( $M = 7.77$ ,  $SD = .22$ ). No significant differences were found among the three groups,  $F(2, 17) = .04$ ,  $p = .96$ . The error rates were: the bar chart ( $M = .13$ ,  $SD = .13$ ), the Circular Ringmap ( $M = .23$ ,  $SD = .15$ ), and the Virtual Pet ( $M = .10$ ,  $SD = .10$ ). As shown in Figure 4, no significant differences were found among the three groups either,  $F(2, 17) = .89$ ,  $p = .46$ .

The self-reported evaluations, however, revealed differences. For the post-test self-reported evaluation, as shown in Figure 5, there were strong significant differences among three visualizations in Readability,  $F(2, 17) = 14.02$ ,  $p < .01$ , and Empathy  $F(2, 17) = 7.61$ ,  $p < .01$ . General significant differences were also identified in participants' ratings of Awareness,  $F(2, 17) = 5.86$ ,  $p < .05$ , and Comprehension,  $F(2, 17) = 5.07$ ,  $p < .05$ . However, no significant difference appears among all in terms of Attractiveness,  $F(2, 17) = .02$ ,  $p = .98$ .

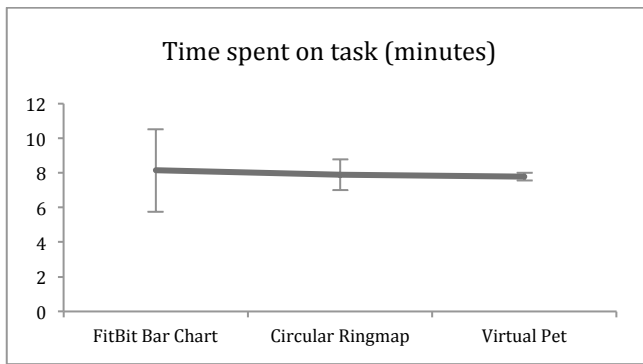


Figure 3: Time spent on task of the three visualization groups.

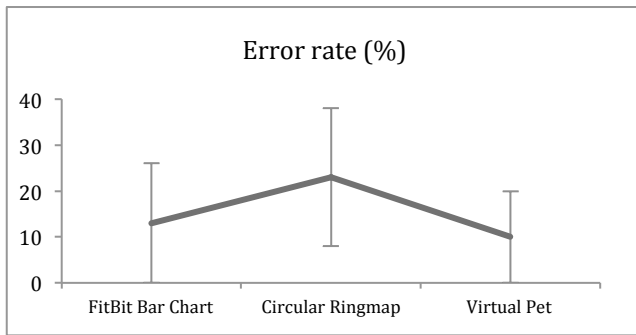


Figure 4: Error rates of the three visualization groups.

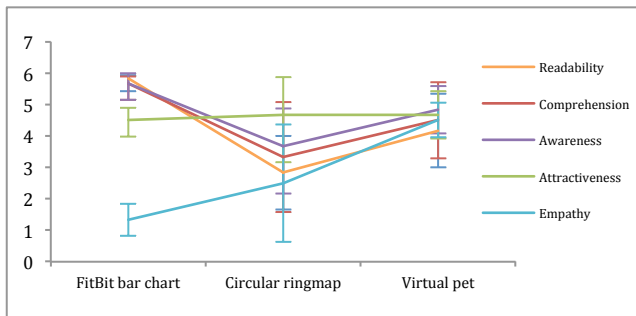


Figure 5: Self-reported evaluations of the three visualization groups.

Tukey-HSD tests were run to figure out exactly which pairs have significant differences. From the results, the *FitBit* bar chart and Circular Ringmap showed a significant difference in Comprehension and Awareness. The Virtual Pet had a strong significant difference compared with the *FitBit* bar chart in Empathy, and it also had a significant difference compared with the Circular Ringmap visualization. In terms of Readability, there were significant differences among all three groups – the *FitBit* bar chart had a strong significant difference from both the Circular Ringmap and the Virtual Pet visualizations; the Circular Ringmap and the Virtual Pet visualizations also had a significant difference.

From the results of this study, non-significant differences of task performance were found among the three visualization groups. Interestingly, however, participants' self-reported evaluations of visualization attributes from the post-test questionnaires tell us a different story. Participants perceived the *FitBit* bar chart to be the visualization that was easiest to read, comprehend, and become aware of – how physically active the data represents. The Ringmap visualization ranked last in all attributes. Although the Virtual Pet visualization was rated in the middle for Comprehension and Awareness attributes, there were

no significant differences between it and the *FitBit* bar chart. For Empathy, however, strong effects were found between the Virtual Pet and the *FitBit* bar chart, and between the Virtual Pet and the Ringmap.

#### 4 DISCUSSION

The main study results revealed that the Ringmap visualization was not helpful or intuitive enough for the participants to reinterpret. It is unclear whether the novelty of the visualization format or if representing 24 hours instead of 12 hours in a circular format played a part in the results. In comparison, the Virtual Pet visualization was considered to be the visualization that evoked the most empathy.

The Empathy attribute matters in this study, because it tells us there is some relationship between the visualization and the individual's emotions. Emotional engagement may have a role to play in visualizations that are designed for personal informatics, especially physical activity for a number of reasons. First, people are invested in attending to these visualizations over time, for personal reasons, often involving improving their health. Second, numerous studies show that to promote behavior changes implicit in physical activity data, extrinsic motivation has a short term boost but it is not sufficient enough to sustain engagement. In contrast, intrinsic motivation engages people at an emotional level and can sustain a longer time of motivation and engagement [6]. Emotional engagement is one of the critical factors that personal data visualizations use to evoke potential intrinsic motivation for lifestyle improvement; in this case, improvement of physical activity.

Moreover, interesting results were discovered when we compared the two measurements gathered from the study – the objective efficiency of performances in a visualization task and the subjective self-reported evaluations of the visualization type. In objective terms, the three groups achieved the same task performance, indicating that participants did not see differences across the three visualizations. However, in subjective terms, participants did perceive the five attributes of the visualizations differently, especially for the Readability and Empathy attributes. This is interesting, especially because the approaches to how the data was visualized varied considerably.

From this study, we learned that how well people perform does not represent how they perceive the visualizations. Therefore, we may conclude that the efficiency and errors are independent of the subjective qualities. Although this fact, the result from the study, is normally realized by the community as a common sense, which has not been demonstrated or actually investigated with concrete result findings. The valuable meaning of this independence between objective efficiency and subjective evaluation is that there can be a favorite visualization perceived by specific population of people that is really inefficient. Also, it indicates that the new visualization method – the Virtual Pet visualization – might be achievable, and that such a method does both – achieves good results for time and errors in task performance, as well as for engaging emotions.

Furthermore, according to the open questions in the post-test questionnaire regarding participants' preferences of visualizations, we summarized four important findings as design implications for future personal data visualization research:

First, to make quantified personal health data more beneficial for identifying activity patterns and for self-reflection, a large volume of data needs to be easily accessible, so that the patterns, trends and progress can be easily comprehended and monitored. Although this seems obvious, there are nuances worth mention, especially in terms of detail and pattern recognition. For instance, in the Circular Ringmap, participants reported it was easier to identify patterns when large amount of data was structured in the

ringmap shape with the detailed information that the five-minute intervals provided. In the Virtual Pet visualization, participants said they were only able to find the day-level physical activity pattern, but not on the hourly level. However, for the *FitBit* bar chart visualizations, participants reported that it was difficult for them to figure out any physical activity pattern.

Second, different visualization types afford understanding the data in diverse ways, and having multiple visualizations on offer may address the diversity of needs that people have in understanding the different things that their personal data can reveal. From the open questions in this study, participants independently also suggested that it would be helpful to combine two types of visualizations together to get a deeper and better understanding of the physical activity data over the entire month. One participant suggested using bar or line graphs for establishing an initial awareness, and for allowing him to search for specific information. The participant then suggested that the Circular Ringmap would be more useful for reflecting on activity patterns. Similarly, another participant proposed combining the *FitBit* bar chart with the Virtual Pet visualizations because it could provide both accurate information and with more engagement and enjoyment. By including more than one type of visualization, the data may be presented in more meaningful (and potentially more specific) ways, which may, in turn, increase people's engagement with the data over time.

Third, when the visualizations are in the domain of personal analytics and in the sub-domain of physical activity, exploring and incorporating factors related to emotional engagement into efficient visualizations may aid people with maintaining their motivation for changing their behaviors over time. From the study findings, an example is that the Virtual Pet visualization was deemed to be the most empathetic. This is an example of how design metaphors, such as that represented by the Virtual Pet visualization, could be useful in this regard. Moreover, in the *FitBit* bar chart visualization, new interactions can be extended and implemented in order to provide a faster exploration of activity patterns for long-term and large-scale data sets. Color-coded mechanisms for representing different levels of activity intensities is also recommended, because it could benefit people in identifying and counting active/inactive days more easily.

## 5 CONCLUSION

In this paper, we first introduced two data visualizations of people's physical activity levels in different structures, and compared them with popular and commercially used bar chart visualization. Next, we compared the effectiveness of the three visualizations in a visualization task and asked participants for self-reported evaluations of five attributes (Readability, Comprehension, Awareness, Attractiveness and Empathy).

The findings from this study proved that the efficiency and error rates of the visualization are independent from its subjective qualities and participants' affective experiences. The significance of this study lies in the fact that it provides concrete evidence that the commonly tested for factors of time and errors (or efficiency) are independent of the more subjective factors of emotional response and awareness. It opens new visualization design challenges around how to create visualizations that combine these factors, and that these factors may have an important role to play in personal data visualizations.

In fact, participants suggested approaching this challenge from both directions: (1) to incorporate more emotive factors into efficient visualizations; and (2) to increase the readability of attractive and more emotionally resonant visualizations. As visualizations move increasingly into people's everyday lives, these factors may become increasingly important.

## REFERENCES

- [1] I. Li, A. K. Dey, and J. Forlizzi, Understanding My Data, Myself: Supporting Self-reflection with Ubicomp Technologies, in *Proceedings of the 13th International Conference on Ubiquitous Computing*, New York, NY, USA, pages 405–414, 2011.
- [2] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay, Activity Sensing in the Wild: A Field Trial of Ubifit Garden, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, pages 1797–1806, 2008.
- [3] J. E. Larsen, A. Cuttone, S. L. Jørgensen, J. E. Larsen, A. Cuttone, and S. L. Jørgensen, QS Spiral: Visualizing Periodic Quantified Self Data, *Proc. CHI 2013 Workshop Pers. Inform. Wild Hacking Habits Health Happiness*, 2013.
- [4] J. J. Lin, L. Mamykina, S. Lindtner, G. Delajoux, and H. B. Strub, Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game, in *UbiComp 2006: Ubiquitous Computing*, P. Dourish and A. Friday, Eds. Springer Berlin Heidelberg, pages 261–278, 2006.
- [5] X. Tong, D. Gromala, C. Shaw, and P. Clarke, Transformation between Electronic Arts and Chronic Pain: Long-term Body Activity Data Visualisation and Pain Animation Expression, pages 75–81, 12014.
- [6] B. Burke, Gamify: How Gamification Motivates People to Do Extraordinary Things. Bibliomotion, 2014.
- [7] D. Huang, M. Tory, and L. Bartram. Data in everyday life: Visualizing time-varying data on a calendar. *Proc. IEEE VIS Poster* (2014).

# Collaborative Visualizations of Self-impacted Data

Pierre Vanhulst

Denis Lalanne

Human-IST Research Center  
University of Fribourg  
Switzerland

## ABSTRACT

This paper presents a literature review led with two purposes: understanding the impact of collaboration in visualization and analyzing the current lack of engagement of end-users with their personal data. Based on this review, we propose a few trails to follow in order to design efficient collaborative visualization systems aiming for behaviour change. We conclude by highlighting which are the unresolved challenges that will require further researches.

**Keywords:** Collaborative visualization, personal visualization.

**Index Terms:** • Human-centered computing~Visual analytics • Human-centered computing~Information visualization • Human-centered computing~Visualization systems and tools.

## 1 INTRODUCTION

The blooming of the smart meters' market allowed pervasive computing to become reality. We are now facing a crucial question that should have been answered earlier: how will those tools improve our life? There are basically two schools of thought when it comes to this: those who automate changes by machine learning and those who inform users so that they understand their personal behaviours and their consequences. While the first category requires less commitment from the users, the second allows them to reflect and understand how their daily routines affect them [6]. This realization is desirable, as it makes people less dependent on machines: humans learn and adapt their behaviours based on their personal data.

In our study, we defined the type of data gathered by quantified self devices as “self-impacted”, stressing the fact that they are not only personal, but also directly influenced by the people who produce them. Thus, “Self-impacted data” share the four following properties: they are *personal*, *continuous*, *actionable* and generally *invisible* (for example, end users usually have no idea how much energy they consume for a given household task). By opposition to quantified self data, self-impacted data also include dimensions that are not directly produced or absorbed by humans such as fuel or electricity consumption.

Using visualizations to make sense of large amounts of data is a popular research topic. While most studies focus on single users, some [12] regard “sensemaking” not only as a cognitive process, but as a social one too: under certain circumstances, groups working on a visualization may achieve better results than single individuals. This specific research area is at the crossroads of two others, “Computer Supported Collaborative Work” (CSCW) and “Information Visualization” (InfoVis). At the other side of the spectrum, “Personal Visualization & Personal Visual Analytics” (PV&PVA) have recently

been defined as new fields of research [13]. They encompass all the usages of visualizations in personal contexts for various purposes. Finally, “persuasive technologies” and “gamification” were also reviewed in an attempt to understand what motivates users to engage with such systems.

In this paper, we report a literature review of several research papers that will eventually lead to the design of a collaborative visualization system of self-impacted data. Based on this review, we also identify the challenges that stand in the way and provide guidelines to support the design of such systems.

## 2 WORK HYPOTHESES

At the origin of our research lies the assumption that collaborating while analysing a visual representation of self-impacted data would foster more relevant findings and ideas for behaviour change.

Hyp. #1: Collaborating on a shared visualization allows users to make more relevant findings.

Acquiring a relevant new behaviour, and sticking to it, is a difficult task. Studies in psychology field [3] argue that behaviour changes require years before being assimilated, and that extinct behaviours are not erased but might resurge whenever a change of context occurs. On the top of this, we assumed that users would not engage easily with visualization systems. Thus, our second hypothesis stresses the need for advanced incentives to make people use the systems and to sustain wanted behaviour changes.

Hyp. #2: New techniques are required for users to engage with the analysis of their self-impacted data and to evaluate the impact of their behaviour change.

## 3 METHODOLOGY

We reviewed 68 articles<sup>3</sup>, each of them being selected for its relevance regarding at least one of our hypotheses, using search engines such as Google Scholars or ScienceDirect. The final step of our research was to review governmental policies in the world and the scientific reports that either advocated, or not, the use of smart meters. While this search was more context-specific, field deployment reports were especially useful when it came to understanding current pitfalls of self-impacted data visualization systems.

## 4 USER ENGAGEMENT

The problem of user engagement (UE) for analysing self-impacted data is dual: first, users do not engage easily with InfoVis. Second, once they engage and find potential solutions to their problems, it is difficult to commit to these solutions. We address both aspects below.

<sup>1</sup> pierre.vanhulst@unifr.ch

<sup>2</sup> denis.lalanne@unifr.ch

<sup>3</sup> This short paper only references 25 of the 68 papers. Those not referenced covers aspects that are more technical. The full list is available at <http://human-ist.unifr.ch/?q=team/pierre-vanhulst/cvbc-guidelines-papers>

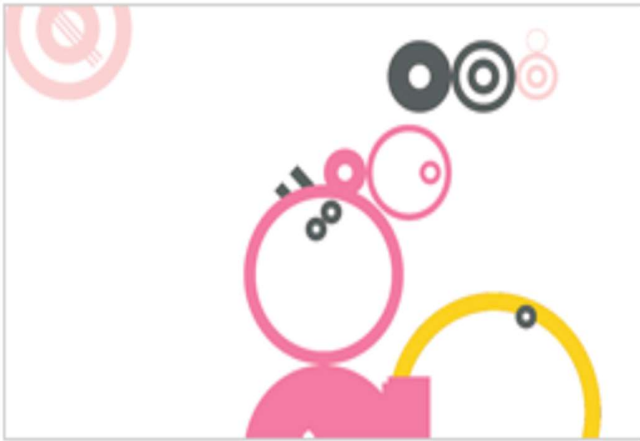


Figure 1 – Tableau Machine [18], an illustration of Casual InfoVis. This visualization displays residential activity in an aesthetically pleasant form.

Of all the contexts covered by self-impacted data, energy conservation provides the most extensive literature. Most governmental reports and studies reviewed came down to the same conclusion, albeit with few variations: so far, the benefits of smart meters are weak. Delmas et al. [8] wrote a meta-review and state that in several cases, eco-feedback could even backfire, and increase the energy use rather than decreasing it. Based on this meta-review, Buchanan et al. explain that the global energy savings offered by smart meters is of 2% [5]. The authors pointed out the lack of thorough evaluations proving the efficiency of In-Home Displays (IHDs). Rego Teixeira reported the results of several field experiments in Switzerland [20]. In one particular experiment, qualitative interviews revealed that users did not feel involved for several reasons: technical issues, lack of financial incentives (sparing energy did not save enough money) and lack of explanations regarding the system are among the most prominent barriers. During our literature reviews, we also came across experiments met with mixed results because of UE. “CommentSpace” [25] is an example of a collaborative visualization system that worked well in laboratory evaluation, but failed to attract users’ attention in practise. During its one-month field deployment, “CommentSpace” gathered 180 accounts with only 32 of them leaving 123 comments. The authors observed that incentives and labour division were necessary for users to engage with a visual analytics system.

Many solutions were proposed to this challenging lack of UE. Among these:

**Storytelling & attractive visualizations:** One is to take down the barriers that repel end users. A Netherlands governmental report regarding smart meters [23] states that simple and attractive visualizations should be used as a “first step” for “computer-illiterates” to engage with the system. To this end, the field of “**Casual InfoVis**” - a subset of InfoVis aiming at non-expert users [18] – provides several techniques that may partially apply to our research.

**Monitoring:** Westkog et al. [24] report the results of several studies, which advocate the use of feedbacks in order to catch users’ attention and make energy consumption “visible”. Darby et al. [7] second this by stating that a monitoring approach is more effective than an “information-driven” approach.

**Nudges & gamification:** Behavioural theories suggest the use of Nudges and Gamification to drive user engagement. However, these techniques suffer from various shortcomings: Rathi and Chunekar [19] highlighted “paternalism” issues or unwanted behaviour changes when using Nudges. As for Gamification, despite many positive short-term results [10], its usefulness over time is yet to be proven. Bartle also theorized that gamification is likely to become less effective as it becomes ubiquitous [2]. However, several examples show that

gamification can work over time such as “Stackoverflow”, a website where developers ask for help. Points showing expertise on various topics reward those who are the most helpful.

The sustainability of a behaviour change is also problematic to evaluate. Klasnja et al. [16] proved how difficult it is to assess the efficiency of a technology in behaviour change. First, it would require “hundreds or even thousands of people matching control group” over several months or years. Then, technology and information alone are not enough to influence an individual “in the right direction”: the success of a behaviour change is related to the socio-cultural environment of the people [17]. Existing evaluation methods of large group of users such as Randomized control trials, also fail to explain how and why users managed to change their behaviour or not. Klasnja et al. argue that early-stage technologies for behaviour change could be evaluated in a “tailored” way, depending on the intervention strategy of the technology – that is, if this technology fosters better behaviour by “self-monitoring”, “reduction”, “tunnelling”, “tailoring”, “suggesting at the right time”, “conditioning” or “surveillance” (according to Fogg’s taxonomy [9]). Each strategy uses different means to leverage user’s behaviour and these means can be evaluated without requiring a large base of evaluators. Sprague and Tory [22] argue that new metrics are to be adopted to evaluate such systems, instead of mere efficiency.

Knowing that self-impacted data appeal end users who are not expert in InfoVis, all these considerations are of prime importance to any visualization system in this context. Designers could stick to Fogg’s taxonomy and define which persuasive strategies they plan to implement, evaluating them in a “tailored” way as suggested by Klasnja et al.

## 5 EFFECTS OF COLLABORATION

In the section above, we highlighted a serious issue with user engagement in the context of self-impacted visual analytics. While literature proposed several solutions to this problem, we believe that collaboration can be an additional way to solve this issue, as well as being an efficient way to provide end users with better findings.

Collaborative visualization is defined by Isenberg et al. as “the shared use of computer-supported, interactive, visual representations of data by more than one person with the common goal of contribution to joint information processing activities” [15]. The idea that visualization is a powerful tool to support decision-making and sense making is globally accepted in the HCI community. For instance, Bresciani and Eppler [4] conclude that the use of any kind of visualization (either optimal or suboptimal) for a managerial group discussing strategy implementation offers better results than no visualization at all.

However, and contrary to the popular view, the same cannot be said of collaboration. Shepperd [21], among others, observed that collaborating does not always result in higher productivity. One of the reasons for this productivity loss is “collaboration cost” – the time people spend to discuss and coordinate themselves. Lack of perceived gains is also part of the equation. Balakrishnan et al. [1] investigated this phenomenon further and suggested that collaboration works best when each member of the group knows only a portion of the data. They are thus appealed to discuss and share their knowledge, which prevents confirmation bias that can occur in groups where each participant knows all the information.

Literature in CSCW uses the “time-space matrix”, a taxonomy based on two spatiotemporal dimensions: distance (co-located or remote collaboration) and time (synchronous or asynchronous collaboration). Of the four situations derived from this taxonomy, two were mostly studied: synchronous co-located collaboration [14] and asynchronous remote collaboration. While it is possible to imagine a social environment that promotes synchronous and co-located collaboration over self-impacted data, this study mainly focuses on remote and asynchronous platforms, similar to what most commercial solutions currently propose.



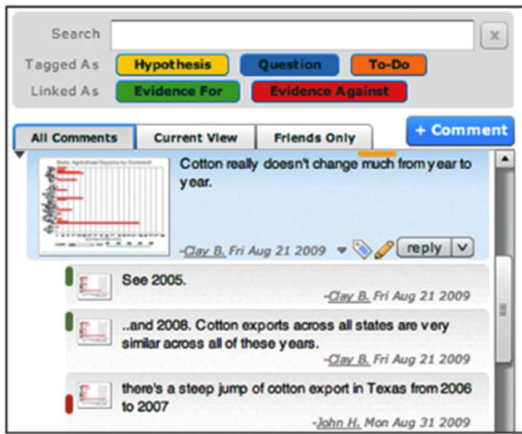


Figure 2 - CommentSpace search interface [12]. Users can create hypotheses, questions or "to-do"s. Other users can provide evidence either for or against this.

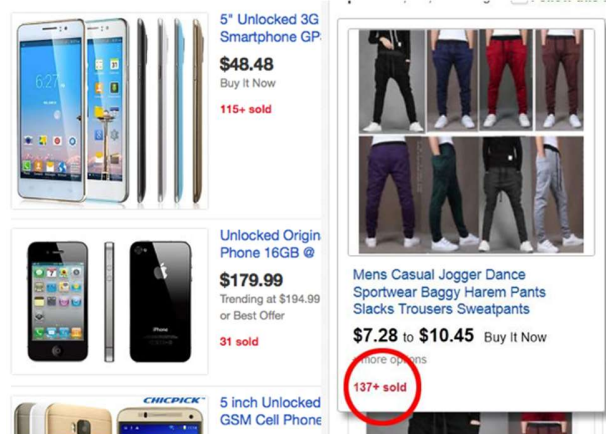


Figure 3 – Social navigation example. While searching on eBay, users are provided with several information such as how many times other users bought a given product, in order to show which are the most popular items.

Heer et al. [12] state that a fundamental aspect of collaborative visualizations is the possibility to share the current state of the visualization and discuss it. This led them to design “Sense.us”, a collaborative visualization system that put in practice their theory. With “CommentSpace” [12], Willet et al. further explored this idea by structuring discussions with predefined tags (to categorize contributions) and links (to specify whether the contribution is an evidence for or against the initial assumption). Heer and Agrawala also proposed a precious list of 24 design considerations that take into account various dimensions such as “common ground”, “incentives” or “division and allocation of work” [11].

## 6 “SHARE FINDINGS, NOT DATA”

Techniques described in the above section took into account collaboration that occurs on a shared dataset. Self-impacted data being inherently personal, they cannot be shared so easily. One of the question that emerged from these readings was whether it is possible to collaborate on a common structure from several datasets, rather than on a single dataset: concretely, how can users share experience and findings without revealing their personal data.

**Figure 4** shows a preliminary design for a 3-steps system to share findings and observations without revealing data per se. This system is “context-agnostic”, as we believe that all self-impacted data could be crossed and analysed: there might be relations between unexpected dimensions. The steps are as follows: (1) produce & gather data, (2) monitor & analyse data, (3) share ideas & adopt new behaviour.

There exist several ways to gather data (1) depending on their nature and availability, we intend to find a standardized way to import them and compare them. Being continuous, they will be characterized by their timestamp and their dimension. Once imported, the data will be displayed (2) with various visualizations as to allow end users to explore them freely. Ambient visualizations – such as power aware cords or hue changing lamps – will trigger higher user engagement, following the idea that a monitoring approach is the best way to drive people’s attention. The collaborative layer will be omnipresent: a user’s exploration of data will be structured by others’ previous findings and hypotheses using techniques described in literature such as social navigation – the fact of displaying former users’ trail over virtual objects. Then, in order to propose anonymous suggestions based on others’ experience (3), the system will gather relevant data from their own annotations. This act of sharing will be encouraged by

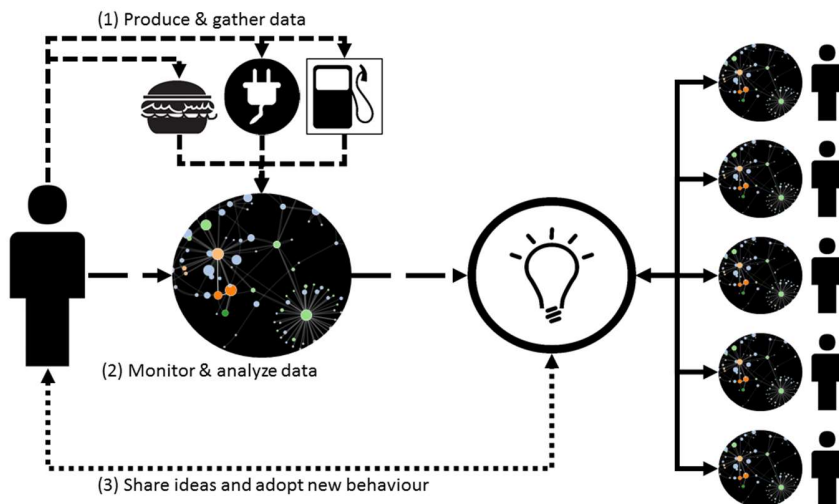


Figure 4 - Design of a collaborative visualization system for behaviour change



nudges and “gamified” mechanisms similarly to those of websites like “Stackoverflow”.

Using Fogg’s taxonomy, such a system would thus implement the “self-monitoring” and “suggestion at the right time” strategies. Surveillance – from other users – could also be implemented: at this stage of the design, we are yet to determine to what extent profiles should be private.

## 7 CONCLUSION

This paper reviewed the state-of-the-art literature according to two initial hypotheses.

Collaboration can bring better results when team members agree to share their information bias. The challenge lies in of how we are going to use this knowledge with collaborative visualization systems of self-impacted data: is it possible to generalize this observation to domains where data are personal and where there exists no clear “good” answer? How can we adapt techniques developed by Heer et al, and Isenberg et al to self-impacted data? We also assumed that it is possible to generalize techniques for all self-impacted data: this is largely arguable and experiments are required to validate this hypothesis.

This literature review also highlighted the many interesting trails to follow in order to solve user engagement problem. So far, none has been properly evaluated. This opens the door to various experiments.

Subsequent to this paper, we plan to design and develop the collaborative visualization system presented above, in order to push our hypotheses further. The focus of our research will be the sharing of findings and their application: how can an individual benefit from others’ findings and increase her understanding of her self-impacted data? The way such a system will recognize whether a finding made by another person can be applied to another, has yet to be defined.

## REFERENCES

- [1] Balakrishnan, A., Kiesler, S., Kittur, A., & Fussell, S. (2010). Pitfalls of Information Access with Visualizations in Remote Collaborative Analysis. *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 411-420). ACM.
- [2] Bartle, R. A. (2011). *Gamification: Too much of a good thing*. Retrieved July 14, 2015, from Presentation given at 'Digital Shoreditch': <http://www.scribd.com/doc/81716927/Gamification-Too-Much-Of-a-Good-Thing>
- [3] Bouton, M. (2014). Why behavior change is difficult to sustain. *Preventive medicine*, 29-36.
- [4] Bresciani, S., & Eppler, M. (2009). The Benefits of Synchronous Collaborative Information Visualization: Evidence from an Experimental Evaluation. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6), 1073-1080.
- [5] Buchanan, K., Russo, R., & Anderson, B. (2015). The question of energy reduction: The problem(s) with feedback. *Energy Policy*, 89-96.
- [6] Costanza, E., Ramchurn, S. D., & Jennings, N. R. (2012). Understanding Domestic Energy Consumption through Interactive Visualisation: a Field Study. *UbiComp '12*.
- [7] Darby, S., Liddell, C., Hills, D., & Drabble, D. (2015). *Smart Metering Early Learning Project: Synthesis report*.
- [8] Delmas, M., Fischlein, M., & Asensio, O. (2013). Information strategies and energy conservation behavior. *Energy Policy*, 61, 729-739.
- [9] Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do.
- [10] Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? — A Literature Review of Empirical Studies on Gamification. *Proceedings of the 47th Hawaii International Conference*. Hawaii.
- [11] Heer, J., & Agrawala, M. (2008). Design considerations for collaborative visual analytics. *Information Visualization*, 7, 49-62.
- [12] Heer, J., Viégas, F. B., & Wattenberg, M. (2009). Voyagers and Voyeurs: Supporting Asynchronous Collaborative Visualization. *Communications of the ACM*, 52(1), 87-97.
- [13] Huang, D., Tory, M., Aseniero, B., Bartram, L., Bateman, S., Carpendale, S., . . . Woodbury, R. (2014, September 22). Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 420-433.
- [14] Isenberg, P. (2009). Collaborative Information Visualization in Co-located Environments. Calgary.
- [15] Isenberg, P., Elmquist, N., Scholtz, J., Cernea, D., Ma, K.-L., & Hagen, H. (2011). Collaborative Visualization: Definition, Challenges and Research Agenda. *Information Visualization*, 10(4), 310-326.
- [16] Klasnja, P., Consolvo, S., & Pratt, W. (2011). How to Evaluate Technologies for Health Behavior Change in HCI Research. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3063-3072). ACM.
- [17] Owens, S., & Driffill, L. (2008). How to change attitudes and behaviours in the context of energy. *Energy Policy*, 4412-4418.
- [18] Pousman, Z., Stasko, J., & Mateas, M. (2007). Casual Information Visualization: Depictions of Data in Everyday Life. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1145-1152.
- [19] Rathi, S., & Chunekar, A. (2014). Not to buy or can be ‘nudged’ to buy? Exploring behavioral interventions for energy policy in India. *Energy Research & Social Science*, 78-83.
- [20] Rego Teixeira, A. (2014). Le « Smart-metering » en Suisse Romande : conditions sociales d’acceptation et d’appropriation. Genève.
- [21] Shepperd, J. (1993). Productivity Loss in Performance Groups: A Motivation Analysis. *Psychological Bulletin*, 113(1), 67-81.
- [22] Sprague, D., & Tory, M. (2009). Motivation and Procrastination: Methods for Evaluating Pragmatic Casual Information Visualizations. *IEEE Computer Graphics and Applications - CGA*, 29(4), 86-91.
- [23] van Elburg, H. (2014). Dutch Energy Savings Monitor for the Smart Meter.
- [24] Westskog, H., Winther, T., & Sæle, H. (2015, May 5). The Effects of In-Home Displays—Revisiting the Context. *Sustainability*, 7, 5431-5451.
- [25] Willett, W., Heer, J., Hellerstein, J., & Agrawala, M. (2011, May). CommentSpace: Structured Support for Collaborative Visual Analysis. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 3131-3140). ACM.

# Personal Movie Recommendation Visualization from Rating Streams

Kodzo Webga \*

Aidong Lu †

University of North Carolina at Charlotte

## ABSTRACT

Online ratings from consumers are important in improving user shopping experiences at almost all popular e-commerce stores. This work discusses the challenges of providing effective personal visualization of movie recommendation based on personal watch history and on-line rating streams. We present our current visualization design which layouts movies based on their statistical features, recommendation results from collaborative filtering, and the rating records. We have developed a prototype system and provided several example results to demonstrate our approach for users with different profiles. We expect that the approach of personal visualization of movie recommendation can be extended to many similar applications in online e-commerce stores and online social networks.

**Index Terms:** K.6.1 [Computer Graphics]—Application K.7.m [Information Interfaces and Presentation]—User Interfaces

## 1 INTRODUCTION

This work studies the problem of movie recommendation for individual users based on their personal watch history in on-line rating datasets. Personal movie recommendation is one popular application of rating streams and it can be extended to a number of similar scenarios, such as music and book recommendations. In fact, on-line e-commerce stores are widely available and have formed an important part of our everyday lives. The online stores often collect ratings from users and recommend similar items according to user profiles; however they seldom provide visualizations for users to browse and search recommendation results. The personal visualization of recommendations can help users digest such a large amount of data produced by our everyday life and enable individual users to organize and explore movies based on their personal watch experiences.

While recommendation problems are common in online e-commerce stores and research fields of data mining and machine learning, personal visualizations of movie recommendations are rare. Movie datasets such as Internet Movie Database (IMDb), have been visualized using graph visualization methods [7, 10, 12], which are capable of providing advanced visualization and visual analytics functions. However, they are not designed for personal visualization of general users. In the following, we describe three key challenges for personal movie recommendation from online rating datasets.

- Personal visualization - Different from visualization for professional analysts, personal visualization of movie recommendation should be intuitive for general users to understand.
- Visual recommendation - Providing advanced yet simple visual analytics functions for users to gain insights on the background recommendation mechanisms and explore movie recommendations.

---

\*e-mail: kwegbal@uncg.edu

†e-mail: aidong.lu@uncg.edu

Electronic proceedings of the IEEE VIS 2015 workshop  
*Personal Visualization: Exploring Data in Everyday Life*  
The authors remain the holders of the copyright

- Online streaming - Handling real-time rating streams at unexpected speeds, data amounts and filtering relevant information for local machines to process and store. Both the results of a recommendation and the recommendation visualization should be adjusted across time as well.

To develop a practical personal visualization system of movie recommendations, we also expect several additional challenges as follows. As the rating datasets are collected from individual users, this application has a potential privacy issue for users who prefer not to publish their user profiles but are willing to share their ratings. There can be some extreme cases requiring time-critical recommendations, especially when the new products are available to sell and the amounts are limited. As rating records often come with review comments, text analysis should also be considered in the recommendation visualization. At the end, the ambiguity issues raised from the recommendation algorithms may affect the exploration of recommendation results significantly.

In this work, we present our current visualization design for personalized movie recommendation toward addressing the three key challenges. We start by describing the application requirements and our design considerations. Then, we present our approach of embedding statistical features of movie ratings and recommendation results on a 2D layout, which is straightforward for general users to connect the relationships of movies and ratings. Specifically, each movie is located by the average rating separated by the rating columns and a distance measurement from a singular value decomposition spaces (SVD) for statistical data features. We have developed a prototype system for handling real-time rating streams and integrated interactive exploration functions for users to visualize the viewing history of the user, recommended movies, and a succinct overview of the entire movie database. We use example results for users of different watch profiles to demonstrate key interaction functions our prototype system can provide for a personal movie visualization.

The remainder of the paper is organized as follows. We first present the related work on personal visualization and recommendation visualization in section 2. Section 3 describes the application requirements and our visualization design. Section 4 provides example results and discusses our approach from different aspects of design considerations. Finally, we conclude this work and describe the future work in Section 5.

## 2 RELATED WORK

We briefly describe the related work of the personalized movie recommendation from the aspects of personal visualization and recommendation visualization.

### 2.1 Personal Visualization

Huang et al. [8] provided a survey of personal visualization and personal visual analytics (PV&PVA) from related VIS and HCI fields and they developed a taxonomy of design dimensions for a coherent vocabulary for PV&PVA. In addition, several recent work has been devoted to personal visualization. For example, Choe et al. [3] mapped the visual annotations for several insight types and discussed four areas for the design of personal visualization systems.

Wang et al. [15] presented three distinct personal visualization designs for visualizing Facebook user data and suggested the importance of an illustrative design. Our design of the personalized movie recommendation follows the guidelines from the previous work by providing a simple yet flexible visualization design.

## 2.2 Recommendation Visualization

Recommendation algorithms have been applied to quite a number of applications related to personal visualization, such as for images [5], music [2], and movies. Most approaches are from the fields of data mining and machine learning, where network visualizations are often adopted. For example, Luo et al. [11] used hyperbolic and multi-modal view to visualize the recommendation list. Kermarrec et al. [9] used SVD-like matrix factorization and PCA for global mapping of movie ratings from high dimensions to a two-dimensional space. Crnovrsanin et al. [4] proposed a task-based and information-based network representation for users to interact and visualize a recommendation list. Vlachos et al. [14] used bipartite graphs and minimum spanning trees to explore and visualize recommendation results of a movie-actor dataset. Gretarsson et al. [6] visualized recommended users in social networks with node-link diagram and grouped relevant nodes on the recommendation list in parallel layers. Our approach also adopts the node-link diagram on a newly designed two-dimensional space (SVD-based distance / rating) for general users of a movie recommendation visualization.

## 3 PERSONAL VISUALIZATION OF MOVIE RECOMMENDATION

We start with describing several aspects of the personal visualization for designing our approach. We then describe our current approach in details.

**Integration of automatic recommendation algorithms and intuitive visual interface.** To provide a personal visualization which is intuitive for users to understand and powerful to achieve advanced analysis functions, automatic recommendation algorithms should be integrated in the visual interface. It is preferred that some user interaction functions for exploring recommendation results or adjusting recommendation preferences are integrated with the background recommendation algorithms to reveal additional insights to the movie dataset.

**Personalized visualization.** As users may have very different profiles and tastes on movie selections, the visual interface should be highly customizable for adjusting the visual parameters as well as levels-of-details. It is ideal to vary the visualization by adjusting different aspects ranging from simple parameters such as the genres of movies, released years, average ratings, the number of recommendation items, to complex features of movies and recommendations according to user watch history.

**Simple and meaning visualization.** The visualization should be designed for general users who may not have previous visualization experiences. It is ideal that the visualization and interactive functions are self-explanatory. Such visualization could replace the recommendation list by providing insights between the recommended items and movies a user has watched through several simple interactions for specifying personal tastes.

We design our personal visualization of movie recommendation based on the considerations described above. For recommendation algorithms, we adopt a popular collaborative filtering approach using a singular value decomposition (SVD) [1]. The recommendation list is computed through finding movies with high ratings from users with similar profiles.

Our visualization approach modifies the node-link diagram on a 2D panel designed for personal recommendation with rating streams. Each movie is mapped to the panel according to the overall average rating for its horizontal position, and a distance between

the movie and the user in selected SVD space for its vertical position. As shown in the examples in the next section, the movies closer to the top of the space may possess stronger relationships to the user and the movies closer to the right of the space have higher average ratings.

The details of the visual parameters are set as follows. All the movie nodes are presented using circular shapes and user nodes are hidden. The size of a node is based on the number of users who have rated it. Thus larger movie nodes indicate more popular movies. The nodes are colored based on the movie category. The movies that have been watched by the user are visualized with a doughnut around the nodes and the color of doughnut is greener for a higher rating value.

To compute the recommendation list, we first map each user into the SVD feature space by multiplying the adjacency matrix of rating records  $M$  by the  $V$  component of SVD results. We then compute the cosine similarity between the active user and all other users using the results of  $M \times V$ . Next, we automatically select the list of users that have a positive similarity with the current user. We choose the movies those users rated positively ( $ratings \geq 3$ ) but have not been watched by the current user. To provide a strong connection of those movie nodes in the visualization space, we compute the minimum spanning tree of a complete graph of the recommended movies using Prim's algorithm, as shown in Figures 1 and 2.

We have developed a prototype system with two components. One is a server component that handles online rating streams and processes time consuming operations such as SVD decomposition. The other is an interface for users to visualize and analyze movie recommendations. The server component is built with Windows Communication Foundation (WCF). We used the incremental SVD package [1] to decompose real-time rating streams. Note that the details of the server description is skipped due to the paper limit and our concentration on personal visualization interface.

We have also developed several interaction functions to filter the movie nodes by the movie category, the release date, the similarity with a selected movie, and the number of recommended movies. The user can also explore these movies by clicking on IMDb page link that would take him or her directly to the IMDb page of the specific movie for further information.

## 4 RESULTS

We have tested our prototype system using the MovieLens100K dataset [13]. The dataset has 100K ratings from 1-5 and 1,682 movies from different categories (or genres) rated by 943 users. The ratings are from September 19th, 1997 through April 22nd, 1998. Each user has rated at least 20 movies during the entire period. The following shows several example results of our approach for users with different profiles.

### 4.1 Example Results

The first set of results shown in figure 1 demonstrated the variety of relevant movies for the same user in different features space of the SVD. It also highlights the personalized taste of each user.

As shown in the top left image in figure 1, we can observe in the first dimension of SVD that the first user does not like comedy movies but enjoys majority drama and romance movies. For example, the user likes the comedy-romance movies "Truth About Cats and Dogs, The 1996", and "Sleepless in Seattle (1993)". We can find a similar movie - the drama-romance movie "Phenomenon (1996)" - with similar average rating and a similar distance from the user. However, in the second and fifth dimensions of the SVD features space shown in the bottom left image in figure 1, the result reveals different set of movies that are closer to the same user. The romance and drama movies dominate this set as well. Other than the comedy-horror movie "Yang Frankenstein (1974)" that is

the closest to the user, we can find movies such as comedy-romance movie “Annie Hall (1977)”, drama movie “Gandhi 1982”, and other drama and romance movies the user have rated with high scores.

The second user in the first SVD dimension space (top right image in figure 1) does not have a preference over any movie genre. From the movies the user has rated, there is no consistency in ratings among any genre of movies. For example, we can see that the user like horror movie “Shining The (1980)”, the comedy-romance movie “Sense and Sensibility (1995)”, and the drama movie “To Kill a Mockingbird (1962)”. However, the user does not like the drama-romance movie “Phenomenon (1996)” nor drama movie “Trainspotting (1996)” etc. Overall, the user does not like comedy movies according to the rating records. This user profile is also represented by the diversity of the genre of movies in this SVD features space. A similar result can be seen in the second and fifth SVD features space (bottom right image in figure 1). In this space, the movies the user has rated spread out vertically in the 2D panel and are mixed with movies the user has not watched yet. The user may use our visualization to explore recommended movies. For example, the user may like the drama movie “Margaret’s Museum (1995)” since it has a similar average rating value and the same distance from the user with the drama movie “One Flew Over the Cuckoo’s Nest (1975)” the user likes. Our results demonstrate that personalized recommendation visualization is unique to each user.

The second set of results is for the interaction and streaming of rating records, as the recommendations need to be updated through time as well. Figure 2 shows our visualization of a user in different time frames. The first top left image shows the result at the initial time frame and the bottom left image shows a result of a set of interactions the user has performed. The user first set the movie similarity slide bar (not shown in the image) to 60 percent. Then, the user selected a reference movie to the drama movie “Field of Dreams (1989)”. This indicates that the user is interested in movies with positive ratings by at least 60 percent of similar users who also like the movie “Field of Dreams (1989)”. We can see that the system presents a number of movies fitting the taste of the user. Among which, there are three movies the user has not watched yet: a drama movie “Mr. Holland’s Opus (1995)”, a drama movie “It’s a Wonderful Life (1946)”, and a comedy movie “Birdcage (1996)”. The two images on the right of figure 2 show the results of the same user at the 10th time frame and the result of similar user interactions. These results contain more movies compared to the previous time frames. Other than the three movies the user has not rated yet at the initial frame, two additional movies are in the list for the user to explore: drama movies “Phenomenon (1996)” and “Time to Kill, A (1996)”.

## 5 CONCLUSION AND FUTURE WORK

In this work, we discuss the challenges for personal visualization of movie recommendation based on online rating streams. We provide our current visualization approach which is based on a recommendation analysis with collaborative filtering and interactive visualization on a newly designed 2D space. We demonstrate our approach with a prototype streaming system and example results for users with different profiles. Our approach can be applied to the recommendation visualization of other products with individual rating records, such as music and images.

It is challenging to design intuitive visualization for general users. While our current design requires additional improvements and user studies for its effectiveness, we preserve the similarity between movie nodes and the user from the SVD spaces. Also, we transform the locations of movies and users from high-dimensional SVD spaces to a 2D panel, which is easy to understand and interact. In the future, we will continue to explore other options to improve our design.

We will also continue to investigate SVD approaches to re-

veal additional insights of the statistical data features, as SVD approaches are widely used in automatic recommendation algorithms. We are interested in developing semantic visual analysis approaches for personal visualization, as it may significantly improve the intuitiveness of visualization design for general users. We also plan to expand our streaming platform to take online streams and develop heterogeneous multi-processing solutions to handle rating streams of various types.

## REFERENCES

- [1] C. Baker, K. Gallivan, and P. Van Dooren. Incpack software: Low-rank incremental svd methods.
- [2] D. Bogdanov, M. Haro, F. Fuhrmann, A. Xambo, E. Gomez, and P. Herrera. A content-based system for music recommendation and visualization of user preferences working on semantic notions. In *Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on*, pages 249–252, June 2011.
- [3] E. K. Choe, B. Lee, and M. Schraefel. Characterizing visualization insights from quantified selfers’ personal data presentations. *Computer Graphics and Applications, IEEE*, 35(4):28–37, July 2015.
- [4] T. Crnovrsanin, I. Liao, Y. Wuy, and K.-L. Ma. Visual recommendations for network navigation. In *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization, EuroVis’11*, pages 1081–1090, 2011.
- [5] J. Fan, D. Keim, Y. Gao, H. Luo, and Z. Li. Justclick: Personalized image recommendation via exploratory search from large-scale flickr images. *Circuits and Systems for Video Technology, IEEE Transactions on*, 19(2):273–288, Feb 2009.
- [6] B. Gretarsson, J. O’Donovan, S. Bostandjiev, C. Hall, and T. Höllerer. Smallworlds: Visualizing social recommendations. *Computer Graphics Forum*, 29(3):833–842, 2010.
- [7] B. Herr, W. Ke, E. Hardy, and K. Borner. Movies and actors: Mapping the internet movie database. In *Information Visualization, 2007. IV ’07. 11th International Conference*, pages 465–469, July 2007.
- [8] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 21(3):420–433, March 2015.
- [9] A.-M. Kermarrec and A. Moin. Data Visualization Via Collaborative Filtering. Research report, Inria, Feb. 2012.
- [10] J. Lee, G. Noh, and C. kwon Kim. Analysis and visualization of movie’s popularity and reviews. In *Big Data and Smart Computing (BIGCOMP), 2014 International Conference on*, pages 189–190, Jan 2014.
- [11] H. Luo, J. Fan, D. A. Keim, and S. Satoh. Personalized news video recommendation. In *Advances in Multimedia Modeling*, pages 459–471. Springer, 2009.
- [12] B. Nemeth, G. Takacs, I. Pílaszy, and D. Tikk. Visualization of movie features in collaborative filtering. In *Intelligent Software Methodologies, Tools and Techniques (SoMeT), 2013 IEEE 12th International Conference on*, pages 229–233, Sept 2013.
- [13] G. Research. Movielens100k: Movie rating dataset.
- [14] M. Vlachos and D. Svonava. Recommendation and visualization of similar movies using minimum spanning dendrograms. *Information Visualization*, 2012.
- [15] S. Wang, Y. Tanahashi, N. Leaf, and K.-L. Ma. Design and effects of personal visualizations. *Computer Graphics and Applications, IEEE*, 35(4):82–93, July 2015.

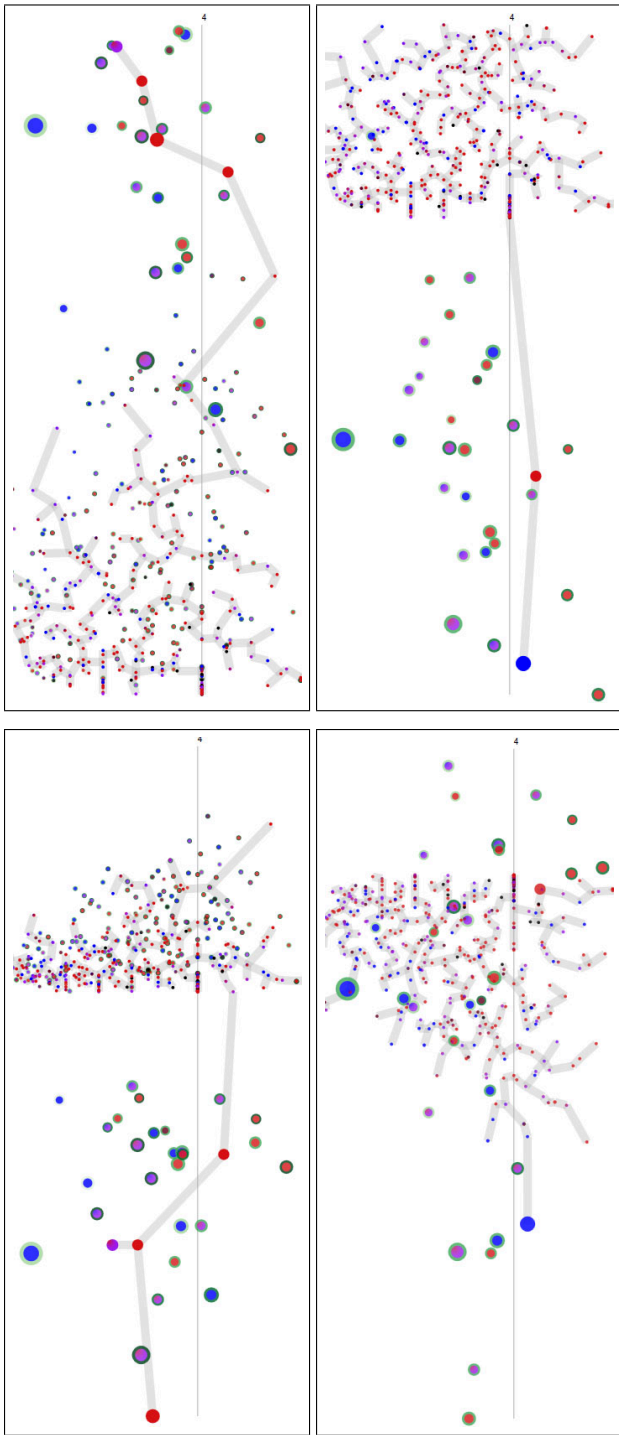


Figure 1: Example results of two users in different SVD dimensions. The images on the left show the results by using the first dimension of SVD spaces. The images on the right use the second and the fifth SVD dimensions. These dimensions can be interactively selected in our prototype system and can be improved with automatic suggestions. The recommended movies are highlighted with the tree paths.

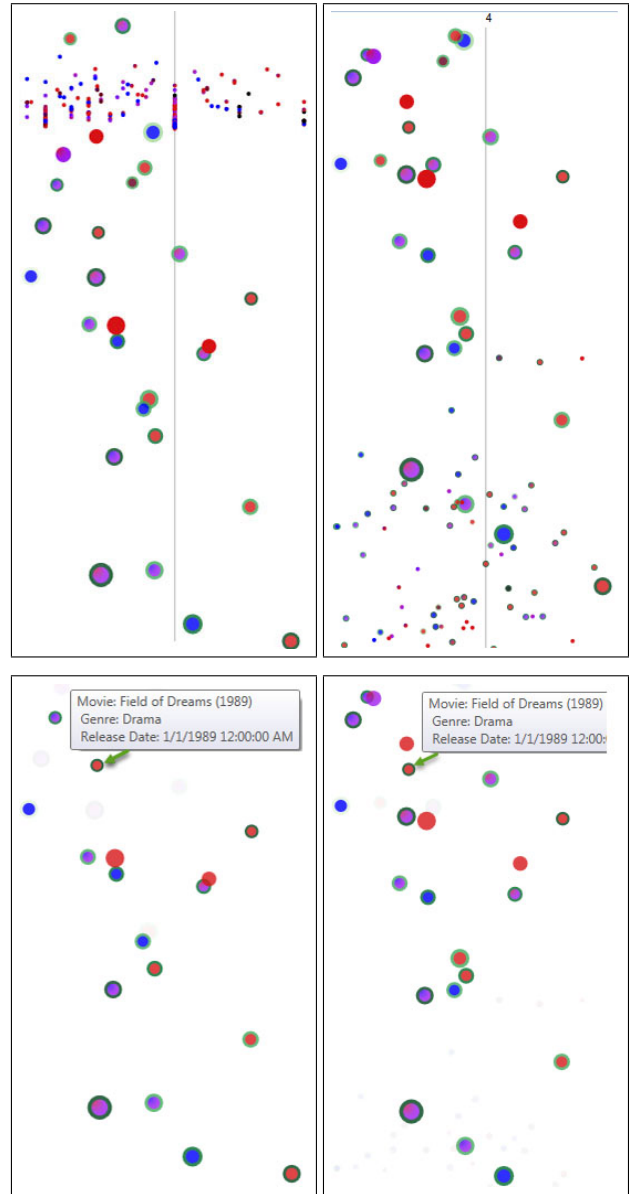


Figure 2: Example results of interaction and rating streams. The two images on the left show the results from two different time frames and the two images on the right show the interaction results. Only portions of the results are shown due to space constraint.



# A Survey of Personal Visualization in mHealth Nutrition Apps

Corrie M. Whisner, Alexandra Porter, Nicholas Pecor, and Ross Maciejewski *Senior Member, IEEE*

## ABSTRACT

One emerging means of engaging and educating individuals and communities is the combination of mobile computing, medical sensors, and communication technology (commonly referred to as mobile health or mHealth) that can enable the tracking and personal visualization of the physical health of the user. To date, hundreds of mobile apps are available for smartphone users for tracking nutrition and weight related elements of personal health. Unfortunately, few apps report on the use of evidence-based practices [4] or behavioral theory [12] in their development and/or programming and fewer still provide users with means of quickly analyzing their personal longitudinal data, relying primarily on static charts of the pie, bar or line variety. In this paper, we survey a small set of mHealth applications for diet and nutrition and identify app functions that would relate to traditional methods of assessing dietary intake and physical activity as well as advice-giving for meal planning and improvement. We discuss the use of personal visualization in this context and close with speculation on novel ways in which the community could explore visualization design for influencing personal changes.

## 1 INTRODUCTION

Obesity has reached epidemic proportions globally with over 1 out of every 3 adults being classified as obese. Overweight and obesity are linked to increased risks in chronic diseases including heart disease, diabetes, osteoporosis and some cancers. Such chronic diseases are a major contributor to the rising health care costs faced in the United States, accounting for approximately \$147 billion of medical costs in 2008 [6]. In order to combat this growing problem, new methods of communicating and educating people about health issues related to overweight and obesity are needed. The Office of the Surgeon General [1] has emphasized the need for engaging and educating communities about healthy eating habits based on the Dietary Guidelines for Americans. In fact, the nutrition and weight status objectives for Healthy People 2020 focus specifically on the promotion of healthful diets and regular physical activity as a means of improving individual health over time [2].

According to the 2015 Pew Research Center Internet and American Life Project, data suggest that a total of 62% of smartphone owners used their phones to explore health information in 2014. In 2012, exercise, diet and weight apps were among the most popular health apps with 38%, 31% and 12% of users utilizing apps to track exercise, dietary intake and body weight, respectively. Despite the increased usage of these tools there are few data to suggest that these apps are effective for achieving health-related goals, as many of the most popular apps do not report the use of evidence-based practices in their development and/or programming [4]. Furthermore, few apps take advantage of visualization. The research agenda on personal visualization by Huang et al. [9] defines several critical areas that would need to be addressed to improve the efficacy of mHealth nutrition apps, specifically the notions of enabling relevant context for reasoning with the data (i.e., what was I doing/feeling when I engaged in unhealthy eating habits?), defining appropriate baselines (i.e., how do I compare to others, normative

data or population standards in the nutrition literature), and sharing/privacy issues (i.e., some research suggests that sharing health data within a community may improve self-management [7]).

## 2 A BRIEF SURVEY OF MHEALTH NUTRITION APPS

In this paper, we present a survey of the most highly downloaded Android and Apple App Store nutrition apps from the winter of 2014. Each app was reviewed looking for functionality that enabled calorie, nutrient, fitness and weight tracking as well as goal setting, peer comparison, social network sharing, meal planning advice and incentives/gamification. Tables 1 and 2 show 15 of the most popular apps (based on number of downloads and ratings) along with a summary of the most common positive and negative comments about each app. Representative comments for the apps were extracted from the app stores to explore sentiment and topics relating to user preferences and software functionality. The sample apps shown here were collected using keywords: "nutrition, diet, health". Functionality of the apps was then broken into nine different categories that are representative of methods/approaches commonly utilized in clinical settings (calorie, nutrient, fitness and weight tracking as well as goal tracking, peer comparison, social networking/sharing, meal suggestions and gamification/incentives). Screen shots were also captured to show representative visualizations (or lack thereof) from the apps.

Of the 15 apps reviewed, the majority (n=14) were multi-functional providing functions beyond nutrient, fitness and weight tracking alone. The most common features among these apps were goal-setting (n=14; primarily weight-related goals), weight (n=15), physical activity (n=11) and calorie (n=12) tracking. Although peer support has been shown to positively impact health, only nine of the apps included a social networking feature and only 4 apps provided peer support functionality. The least common features were meal planning/recipes (n=1) and incentives/gamification (n=3).

Visualizations provided in the apps under review included line graphs, pie charts and bar graphs. Nutrition displays/images were limited to visualizing macronutrients (protein, carbohydrate and fat) in pie charts to illustrate the breakdown of these core nutrients. Vitamins and minerals (micronutrients) were frequently visualized using bar-type graphs. Apps such as MyFitnessPal allow for personalization of dietary and physical activity goals but lack educational components to assist the user in making evidence-based goals. Other dietary trackers, including FatSecret, MyFitnessPal, Lose It!, LIVESTRONG Calorie Tracker, MedHelp's Calorie Counter, and SparkPeople's Calorie Counter all offer the ability to track specific food items on a daily basis. However, the only visualization options in any of these are pie graphs based on macronutrients (protein, fat, and carbohydrates) and bar charts of calories, sugar, or other nutrients. Detailed data are presented almost exclusively through text and tables, and effective tracking requires the input of all food items consumed. While this process is simplified on these applications by making suggestions based on past records and allowing for bar code scanning, there is no way to enter general information about a user's diet that can be used to assess their nutritional behaviors. Furthermore, few of these apps allow visualization or interaction with behavior goals and the allowable changes to behavior goals may not be broad enough to meet the needs of users with additional complications beyond weight, such as cardiovascular disease, diabetes, Celiac disease, food allergies, etc. As such, we feel that an interesting avenue for personal visualization and

mHealth is focusing on behavioral methods and visual messages that may persuade users to improve their health outcomes. Future investigations in this area should aim to understand how end-users view and interpret their health data and how these visuals impact future attitudes, beliefs and behaviors about health.

### 3 FUTURE DIRECTIONS FOR VISUALIZATION AND NUTRITION

With regards to personal visualization and personal visual analytics, Huang et al. [9] discussed the need for integrating computer assisted analysis and persuasive technology. We feel that the use of visualization for persuasion and its application in personal visualization could be an interesting avenue of research. Recent work in social psychology demonstrates that social norms not only spur but also guide actions within society [8]. Such social-norms marketing campaigns have emerged as an alternative to traditional information campaigns. At the same time, work in social and cognitive science indicates that knowledge of health risks and benefits is one precondition for change [3], and research has shown that increased awareness can be associated with dietary improvement. As such, the development of health information campaigns that utilize social-norms as a marketing strategy combined with messages that promote personal efficacy could potentially demonstrate a broad impact. Currently, limited work has been done in utilizing social-norms as a means of promoting healthy eating.



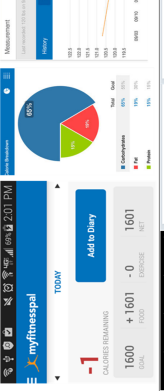
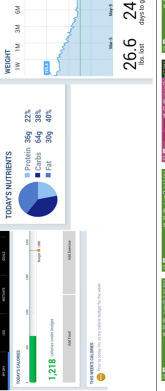

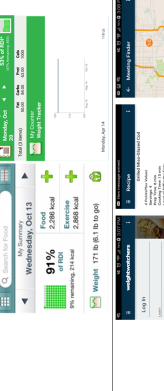


Along with social-norms, the impact of social interactions, relationships and our need to belong powerfully influence our sense of self including our motivations to adopt certain behaviors. Many evolutionary theorists postulate that humans have established such a strong and influential presence on our world specifically because of our strong social bonds and relationships. Importantly, social-norms can have either a positive or a negative impact on outcomes depending on if a person is above or below the social norm and the desired outcome [5]. Based on this, Cialdini and colleagues have developed the Focus Theory of Normative Conduct, to provide guidance on when precisely social-norms will influence behaviors [5]. The theory postulates first that social norms can only be influential if the comparison to others is salient. Further, the theory postulates two types of norms, descriptive norms (i.e., what is commonly done) and injunctive norms (i.e., what others believe ought to be done), which have differential impact on behaviors. For example, research found that combining descriptive and injunctive norms within a message can further impact behavior, in this case towel reuse among hotel guests with both generic (e.g., whole hotel) and specific (e.g., specific rooms) descriptive and injunctive norms was more influential than either individually [13]. That said, previous research exploring how individuals share health information via social media suggests that poorly designed messages can be perceived poorly and result in negative interactions (e.g., sarcastic but well intentioned, or just mean) responses from a social network or just poor responses from the individual receiving the information [10]. These results highlight the need for a more nuanced understanding of the best methods for discussing health information in social media, particularly with reference to injunctive norms as these can easily result in negative backlash based on popular media stories.


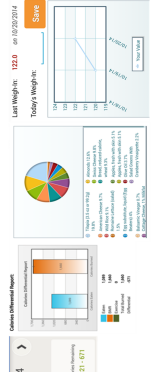
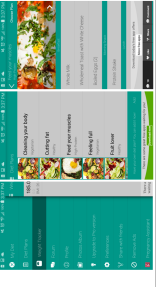
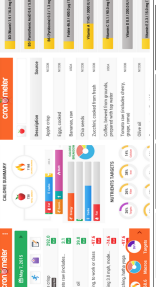
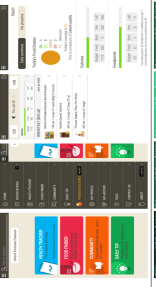
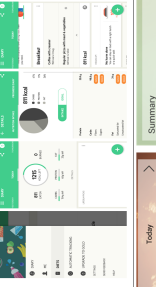
We believe an interesting future direction for personal visualization is the exploration of the use of both descriptive and injunctive norms, as compelling reference points for supporting healthier eating but with a close eye on the possible unintended consequences of message framing. For example, users could be provided a variety of healthy eating messages targeted to their specific demographics and the impact of combinations of descriptive and injunctive norms with different visualization frames could be explored. Collectively, these results could potentially highlight the many different possible frames which could be used to inspire behavioral change via creating social connections about food through data already available

from social networks. We feel such research could also consider the concept of “deceptive visualization” [11] and how one might be able to use deception to improve the adoption of healthier behaviors. As such, we feel that visualization methods for harnessing the influence of descriptive and injunctive norms along with subtle cues about social relatedness as methods for influencing behaviors could spawn a variety of new and interesting personal visualizations.

### REFERENCES

- [1] The surgeon general’s call to action to prevent and decrease overweight and obesity. Technical report, U.S. Department of Health and Human Services, 2001.
- [2] Healthy people 2020 objectives. <http://healthypeople.gov/2020/default.aspx>, 2012.
- [3] A. Bandura. A health promotion by social cognitive means. *Health Education & Behavior*, 31:143–164, 2004.
- [4] E. Breton, B. Fuemmeler, and L. Abrams. Weight loss - there is an app for that! But does it adhere to evidence-informed practices? *Translational Behavior in Medicine*, 1(4):523–529, 2011.
- [5] R. Cialdini, R. Reno, and C. Kallgren. A focus theory of normative conduct: Recycling the concepts of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58:1015–1026, 1990.
- [6] E. A. Finkelstein, J. G. Trogon, J. W. Cohen, and W. Dietz. Annual medical spending attributable to obesity: Payer- and service-specific estimates. *Health Affairs*, 28(5):w822–w831, 2009.
- [7] J. H. Frost and M. P. Massagli. Social uses of personal health information within patientslikeme, an online patient community: What can happen when patients have access to one another’s data. *Journal of Medical Internet Research*, 10(3):e15, 2008.
- [8] N. Goldstein, R. Cialdini, and V. Griskevicius. A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35:472–482, 2008.
- [9] D. Huang, M. Tory, B. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 21(3):420–433, March 2015.
- [10] S. Munson. Beyond the share button: Making social network sites work for health and wellness. *IEEE Potentials*, 30(5):42–47, 2011.
- [11] A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. How deceptive are deceptive visualizations?: An empirical analysis of common distortion techniques. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1469–1478, 2015.
- [12] H. E. Payne, V. B. Moxley, and E. MacDonald. Health Behavior Theory in Physical Activity Game Apps: A Content Analysis. *Journal of Medical Internet Research Serious Games*, 3(2), 2015.
- [13] W. Schultz, A. Khazian, and A. Zaleski. Using normative social influence to promote conservation among hotel guests. *Social Influence*, 3(1):4–23, 2008.

App Name	Rating (total reviews; mean score)	# of Downloads (android)	Functions							Positives	Negatives	Screen Shots	
			Calorie Tracker	Nutrient Tracker	Fitness Tracker	Weight Tracker	Peer Compare Network	Social Network	Meal Plan/Recipes				Incentives
MyNetDiary Calorie Counter Pro	6845; 4.4	100,000-500,000	x	x	x	x	x	x			Very detailed and easy to use, tracks calories, exercise, and nutrients	can only sync to other devices with the upgrade	
MyNetDiary Pro - Calorie Counter and Food Diary for Diet and Weight Loss	13321; 4.5	--	x	x	x	x	x	x			Gives a large variety of meals plans and also provides shopping lists for the meal plans	Doesn't track calories or nutrients. Also doesn't provide details on how to prepare the food	
Diet Point - Weight Loss	27016; 4.2	100,000-500,000				x				Extra with PRO	Simple Intuitive	Problem feature (recipe builder)	
MyFitnessPal - Calorie Counter	958985; 4.6	10,000,000-50,000,000	x	x	x	x	x	x		w/VIP	Fastest and easiest Social Tracking progress Largest food database		
Fitness Inc. - Lose It!	41409; 4.4	5,000,000-10,000,000	x	x	x	x	x	x		w/premium	Seamless connection to people, devices, information Simple, easy tool for tracking Big database Social support Challenges	Difficult to find some items Adding multiple items unnecessarily cumbersome	
My Diet Coach - Weight Loss	85748; 4.3	5,000,000-10,000,000	w/ Pro	w/ Pro	w/ Pro	x					Provides incentive to use in the form of a game - Gives tips to combat many difficulties people have such as craving sweets and exercise laziness	Game only works for females. The app doesn't actually track anything unless you upgrade. Only gives tips	
Calorie Counter - FatSecret	175452; 4.3	10,000,000-50,000,000	x	x	x	x	x	x			Allows you to share food, exercise, and weight data with your preferred professionals (dietitian, trainer, doctor, health professional)	Lacks search functionality New version more difficult Desire more settings, tracking Incomplete database	
Weight Watchers Mobile	64822; 3.9	1,000,000-5,000,000	x					online	online	online	Has a big online community, is a very well known program for weight loss/management	The app is just an extension of the web community, must pay for the online membership to use the app	

App Name	Rating (total reviews; mean score)	# of Downloads (android)	Functions								Positives	Negatives	Screen Shots	
			Calorie Tracker	Nutrient Tracker	Fitness Tracker	Weight Tracker	Peer Compare	Social Network	Meal Plan/Recipes	Incentives				Other
Noom Coach: Weight Loss Plan	157636; 4.3	10,000,000-50,000,000	x		w/ Pro	x	w/ Pro	w/ Pro				Simple and easy to use. Provides a free pedometer and weight graph	Many features can only be used with the upgrade to PRO such as Recipes, and group connections	
Sparkpeople: Calorie Counter & Diet Tracker (Weight Loss, IOS)	6861; 4.3	500,000-1,000,000	x	Online	x	x	x	x			x	Simple	More nutritional detail Incomplete database	
Diet Assistant - Weight Loss	10541; 3.9	1,000,000-5,000,000				x	x					Gives a large variety of meal plans to follow, also tracks weight loss	Doesn't track calories, nutrient intake. Doesn't provide shopping costs or explain in detail how to prepare the food, also all portion sizes are given in oz. and ml.	
CRON-O-Meter	372; 3.6	10,000-50,000	x	x	x							Has a comprehensive website, which is free to use, even without the app	Not free to download the app, but the website is free to use. Unsure if you can scan bar-codes in the app mode	
Fooducate	10975; 4.5	1,000,000-5,000,000	x	w/pro	x	x					w/pro	Very basic, provides three components: food finder, calorie tracker, and community for motivation	Doesn't track exercise with basic account	
Lifesum	36694; 4.2	1,000,000-5,000,000	X	x	x	x						Gives a good breakdown of calories and macronutrients, also provides goals for macronutrient breakdown	Provides diets plans that are confusing and hard to use, might be easier with the upgrade	
My Diet Diary Calorie Counter	17236; 4.1	1,000,000-5,000,000	x	x	x	x						Simple Effective without unnecessary features Less complex than MyFitnessPal	General bugs starting app, saving	