HAL
open science

Showing Data about People: A Design Space of Anthropographics

Luiz Morais, Yvonne Jansen, Nazareno Andrade, Pierre Dragicevic

To cite this version:

Luiz Morais, Yvonne Jansen, Nazareno Andrade, Pierre Dragicevic. Showing Data about People: A Design Space of Anthropographics. IEEE Transactions on Visualization and Computer Graphics, In press. hal-02931257v1

HAL Id: hal-02931257

https://hal.science/hal-02931257v1

Submitted on 5 Sep 2020 (v1), last revised 7 Sep 2020 (v2)

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.

Showing Data about People: A Design Space of Anthropographics

Luiz Morais, Yvonne Jansen, Nazareno Andrade, and Pierre Dragicevic

Abstract—When showing data about people, visualization designers and data journalists often use design strategies that presumably help the audience relate to those people. The term *anthropographics* has been recently coined to refer to this practice and the resulting visualizations. Anthropographics is a rich and growing area, but the work so far has remained scattered. Despite preliminary empirical work and a few web essays written by practitioners, there is a lack of clear language for thinking about and communicating about anthropographics. We address this gap by introducing a conceptual framework and a design space for anthropographics. Our design space consists of seven elementary design dimensions that can be reasonably hypothesized to have some effect on prosocial feelings or behavior. It extends a previous design space and is informed by an analysis of 105 visualizations collected from newspapers, websites, and research papers. We use our conceptual framework and design space to discuss trade-offs, common design strategies, as well as future opportunities for design and research in the area of anthropographics.

Index Terms—Anthropographics, design space, empathy, compassion, prosocial behavior.

1 Introduction

When communicating data about people, information designers and data journalists regularly create visualizations[1] meant to foster an emotional connection with the persons whose data are represented. Figure 1 shows one example where the reader can see the story and personal information of each person who died in a public mass shooting in the USA. Meanwhile, the visualization in Figure 2 conveys the hardship of the life of refugees by narrating the story of S.W.G., a 26 years old refugee who left Pakistan and spent 651 days before arriving at his final destination. Both visualizations were crafted to bring readers closer to the persons whose data are visualized.

The practice of visualizing data about people in a way that helps the audience relate has been called *anthropographics* [1], [2]. Boy and colleagues [1] coined this term as an abbreviation for *anthropomorphized data graphics*—visualizations with human-shaped symbols. The meaning was then extended to *"visual strategies to make the connection between data and the humans behind them more direct and, hopefully, more empathic"* [2]. Similarly, the term *data humanism* was coined to refer to a range of visualization design practices intended to promote humanistic values [3]. In this article, we reconcile these different views by defining anthropographics as:

> *visualizations that represent data about people in a way that is intended to promote prosocial feelings (e.g., compassion or empathy) or prosocial behavior (e.g., donating or helping).*

- *Luiz Morais is with Universidade Federal de Campina Grande, Brazil, and Inria and Université Paris-Saclay, France. E-mail: luizaugustomm@gmail.com*
- *Yvonne Jansen is with CNRS and Sorbonne Université, France.*
- *Nazareno Andrade is with Universidade Federal de Campina Grande.*
- *Pierre Dragicevic is with Inria and Université Paris-Saclay, France.*

Manuscript received August 15 2019; revised September 5, 2020.

1. In this article, we employ *visualization* in a broad sense that also includes infographics.

Like Boy and colleagues [1], this definition refers to a class of visualizations rather than a set of design strategies, and like Bertini [2], it generalizes beyond the use of human-shaped symbols. Based on findings that empathy is not necessarily conducive to helping behavior [4], the definition considers prosocial behavior as a potential design goal, as well as other prosocial feelings such as compassion, which is presumably more directly connected to prosocial behavior [4]. Prosocial feelings and behavior can be directed towards the very people represented in the visualization, or towards other people who share a connection with them (e.g., survivors, descendants, or people sharing identity characteristics).

Visualization designers have explored many strategies for creating anthropographics. A popular strategy is to use human-shaped symbols [1] (such as in Figure 1). Other strategies include the use of text annotations to make each person appear unique [1] (for example, see the top left of Figure 2), the use of visual metaphors (such as a red bar chart to symbolize blood or death), or the representation of persons as individual marks rather than aggregated data [5], [6] (such as in Figure 1). So far, only a few of these strategies have been empirically tested, and even though initial studies have been mostly inconclusive [1], [7], [8], there is an enormous potential for information visualization research.

Anthropographics is a rich and growing area, but the work so far has been scattered. As such, the visualization community presently lacks a precise language for discussing anthropographics, as well as an overview of design strategies and opportunities for design. The goal of this paper is to contribute to fill this gap by proposing a design space and conceptual framework that are meant to help researchers and practitioners reason and communicate about anthropographics. We aim at providing conceptual foundations and a language to facilitate the design, critique, comparison, and empirical evaluation of anthropographics.

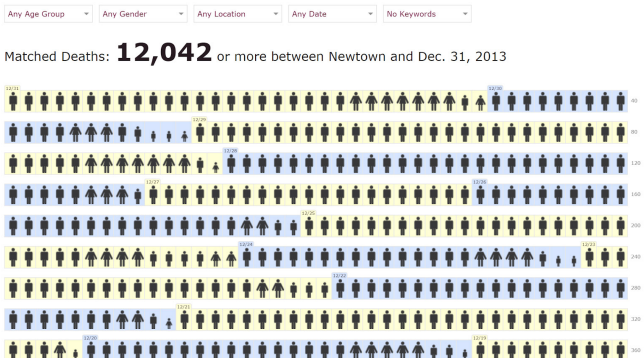

Fig. 1: *How Many People Have Been Killed by Guns in USA Since Newtown's attack*. This visualization shows people who died due to a gun shot between the Newtown attack in 2012, and December, 2013. Each icon is a person. Sex and age are shown through the icon's shape. Additional information about each person can be obtained by clicking on their icon. Source: Slate [9].

Our framework extends a previously proposed design space [1]. The primary focus of this previous work was, however, on reporting empirical studies. In contrast, the present article is fully dedicated to laying out a comprehensive design space of anthropographics. Specifically, we introduce useful basic terminology, extend and refine the earlier design space, and provide and discuss a range of illustrations and examples.

Our design space consists of seven design dimensions that can be reasonably believed to have some effect on prosocial feelings or behavior, namely: *granularity*, *specificity*, *coverage*, *authenticity*, *realism*, *physicality*, and *situatedness*. These dimensions were identified by examining a collection of 105 communicative visualizations that convey data about people, some of which are anthropographics and some of which are not (that is, some are conventional statistical charts). After describing the design space and the conceptual framework it is based on, we identify recurrent combinations of dimensions (that is, families of visualization designs), and discuss opportunities for future research and design.

Our conceptual framework is meant to be *descriptive*, *generative* and *evaluative* [10]: it has been devised to help designers think more clearly about existing and possible new designs. However, it is not meant to be *prescriptive* [10]: it cannot help predict what designs will work best, and cannot prescribe what designs to use. More studies are needed before any such predictive attempts can be made. Frameworks such as ours are, however, important to:

- help *researchers* design empirical studies to estimate the effects that such charts have on people, and
- help *designers* consider different alternative designs by making design dimensions explicit and by emphasizing differences across the continuum of each dimension.

While we conceived the design space by analyzing a corpus of anthropographic visualizations, we believe its design dimensions are also relevant to consider in the context of any visualization design problem in which data about people are to be visualized.

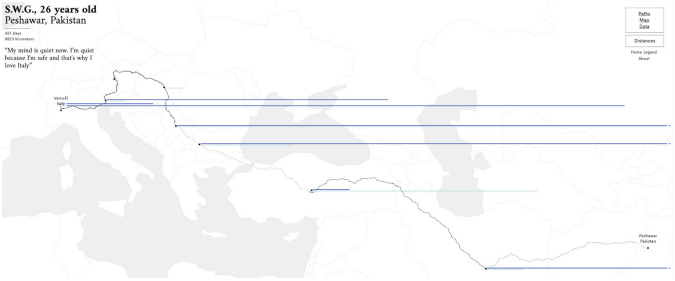

Fig. 2: *Stories Behind a Line* shows stories of seven refugees who fled from their home country to a new place. This figure shows the story of an immigrant who left Pakistan towards Italy. The stroke patterns in the path correspond to different means of transport (e.g., by foot or by car); the dark blue line represents the days of stay in a city while the light blue ones represent days of travel. Source: storiesbehindaline.com [11]

2 Related Work

In this section we discuss previous related work, both within visualization and outside.

2.1 Visual Embellishment in Visualization

Visual embellishment consists of adding non-essential imagery to visualizations in order to make them more memorable, more persuasive or more engaging [12]. Visual embellishment can take different forms, such as seemingly gratuitous uses of the third dimension, but most frequently it consists of embedding pictures with charts. Although visual embellishment is a controversial practice [13], recent studies find that the inclusion of pictures can help make charts more memorable [12], [14], [15], though it can also make them more difficult to process [14], [16]. One study found that the use of pictorial symbols to represent units of data can help people remember information and can encourage them to inspect visualizations more closely, apparently without any negative impact on legibility [16]. This suggests that when showing data about people, the use of anthropomorphic symbols (as in Figures 1, 6, and 12) may have benefits on memorability and engagement without clear downsides. However, the benefits of anthropomorphic symbols on prosocial feelings or behavior have not been studied, with the exception of the study by Boy and colleagues [1] we already mentioned, and a more recent study [7].

Although past research on visual embellishment overlaps with the present work, its focus is on *various* possible benefits of a *specific* design strategy (visual embellishment). In contrast, our work is on *specific* benefits (promotion of prosocial feelings or behavior) of *various* design strategies. Again, the area of anthropo*graphics* encompasses a range of design strategies, not all of which involve the use of anthropo*morphic* symbols [1], [2] (see for example, Figures 2, 7, and 11). Conversely, the use of anthropomorphic symbols does not automatically make a visualization an anthropographic. For example, such symbols can be used to show how many people live in different cities or how many prefer to eat cheese over chocolate; although the use of such symbols may have several benefits such as on memorability, comprehension or engagement, it is not intended to promote prosocial feelings or behavior in these cases.

2.2 Narrative Visualization and Storytelling

Narrative media such as newspaper articles and live presentations are increasingly mixed with visualizations, be they static, animated, or interactive [17], [18], [19], [20]. There has been considerable interest in visualization research in trying to map strategies for telling stories with visualizations. An early study examined existing practices and identified seven genres of narrative visualization (i.e., magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and video) [17]. Subsequent research discussed narrative visualizations in terms of the rhetorical [21] and cinematographic [22] techniques employed, sequences and transitions [23], types of presentation they are used in [18], general design patterns [20], and workflows involved in creating them [24]. Other contributions include novel types of narrative visualizations [25], empirical studies [26] and authoring tools (e.g., [27]).

One clear area of overlap between this previous body of work and the present article is the focus on communicative visualizations. However, because not all communicative visualizations are narrative visualizations [24], our focus is in one sense broader. Our focus is also much narrower in other respects, as we are interested in visualizations that convey data about people and in techniques that are thought to promote prosocial feelings or behavior. Another area of overlap is methodological — much of the previous work on narrative visualization analyzes existing designs to derive a conceptual framework, taxonomy, or design space. The present article does this too, but again with a different focus. Our design space is also at a lower level of description than most design spaces on narrative visualization, which typically focus on rhetorical and narrative techniques.

A special case within the area of narrative visualization is the use of personal narrative visualizations [28]. Whereas personal visualizations in general are concerned with tools to help people understand data about themselves [29], personal narrative visualizations are "visual data representations that tell stories about personal experiences from the point of view of the narrator" [28]. They may be used either as memorabilia and tools of self-reflection, or with the intent to share personal information and stories with others, as done by Perin [30], [31]. Such visualizations are about people and can share some design strategies with anthropographics. However, they are only anthropographics according to our definition if their authors created them with the intent to instill compassion for themselves in others.

2.3 Work in Psychology

Although empathy and compassion have been the subject of a lot of research in psychology, they have no widely accepted definition. Most previous work however considers that empathy and compassion are different constructs [32], [33], [34]. Empathy has been defined as *"the act of experiencing what you believe others are experiencing"* [33], while compassion has been defined as *"the feeling that arises in witnessing another's suffering and that motivates a subsequent desire to help"* [35]. We adopt these two definitions in the present article. While most visualization researchers and practitioners only use the term "empathy" in the context of anthropographics, they are likely also concerned about compassion, and probably often use the term "empathy" to refer to compassion.

Research suggests that empathy tends to lead to behaviors to reduce one's own suffering, while compassion tends to produce behaviors that reduce others' suffering [35]. Similarly, compassion is a positive predictor of prosocial behavior [36], while empathy *"is either not predictive or negatively predictive of prosocial actions"* [34]. One finding from psychology that is directly relevant to our work is the "identifiable victim effect", according to which people tend to empathize and donate more to vividly identifiable victims than for less identifiable individuals [37]. A related finding is the "compassion fade effect", according to which the level of compassion towards people decreases as the number of suffering individuals increases [38]. These findings and their implications for anthropographics design will be discussed further as we introduce the dimensions of our design space.

2.4 Studies in HCI and Visualization

The field of virtual reality (VR) has explored whether immersive environments can affect people's prosocial feelings and behavior. Calvert and colleagues [39] have examined whether showing a story using a VR technology vs. a monitor could affect empathy. Kandaurova and Lee [40] conducted a similar experiment in the context of charitable giving. They measured empathy, prosocial behavior (donation of time and money), and other factors such as guilt and responsibility. Overall, VR studies suggest that immersive environments can promote prosocial feelings and behavior. Closer to information visualization, Ivanov and colleagues [41] argued that showing victims of a tragedy as people in a virtual reality environment could promote empathy since *"the level of fidelity [of such environments] has the potential to enable a kind of direct face-to-face relationship with visualizations"*. In robot-human interaction research, anthropomorphism has also been shown to facilitate empathy [42].

In information visualization, only a handful of studies have investigated prosocial feelings or behavior. Kennedy and Hill [43] explored emotional engagement with data and visualizations, and observed that participants became compassionate with migrants after seeing a visualization about the topic. Erlandsson and colleagues [44] investigated to what extent people allocate money according to the visualization that is presented. Participants were shown different datasets on cancer death risks, and results suggest that people tend to donate more to projects when the data is worrisome. To the best of our knowledge, only three recent studies investigated how visualization design affects prosocial feelings [1], [7], [8]. Although different anthropographic design strategies were tested, the results were mostly inconclusive. We will discuss these studies in more detail in section 6.2, using the terminology from our design space.

3 Methodology

The starting point of our work was the design space of anthropographics by Boy and colleagues [1]. We extended this design space based on an analysis of a corpus of visualizations, and based on data visualization and psychology research, as well as blog posts from practitioners. This section

describes how we collected our corpus of visualizations, and how we used this corpus to inform our design space.

3.1 Scope

We collected a total of 105 visualizations that convey data about people (see a static version of our corpus on OSF[2] or the interactive website[3] for an updated list). The collection process started with visualization repositories (e.g., flowing-data.com and dataphys.org/list) and proceeded through a snowball sampling to include blogs and newspapers. We also included visualizations from the additional material provided by Boy et al. [1]. Finally, the collection was complemented by charts found in social media, and by visualizations from our own practice. Three criteria were considered to include a visualization:

1) *The visualization shows data about people*: examples include statistics about people who died due to a gunshot, the path taken by refugees to escape from war, or the characteristics of women who fought breast cancer.
2) *The data visualized is about real people*. This condition led us to exclude, for example, simulation-based visualizations where imaginary individuals are created to convey life expectancy data or causes of death.[4]
3) *The visualization was published with a communicative intent*. This criterion includes infographics from newspapers, for example, but excludes visualizations produced for purely analytic purposes, for which there is no communicative intent.

The resulting set of 105 visualizations was compiled between October 5, 2018, and July 30, 2019. Among all visualizations, 29% are from blog posts, 27% from newspapers or magazines, 21% from project web pages, 10% from public exhibitions, 10% from books or reports, and 3% from academic papers. The collection includes both interactive and static visualizations. It also includes both anthropographics and non-anthropographics (i.e., visualizations that were likely not designed to promote prosocial feelings or behavior). Doing so allowed us to lay out a more comprehensive design space, by considering the full continuum between anthropographics and non-anthropographics.

3.2 Development of the Design Space

We started with the set of dimensions proposed by Boy et al. [1], and progressively iterated over them by characterizing each visualization from our collection according to each of the dimensions. This process helped us establish the characteristics and boundaries of each of the dimensions, and identify gaps. In parallel to defining and refining the dimensions of the design space, we developed a conceptual framework and set of elementary definitions (described in Sections 4.1 and 4.2) in order to have a firm conceptual ground on which to rest our final design space.

As our corpus contains a broader range of visualizations than initially considered by Boy and colleagues, the dimensions that could not properly fit the corpus or could not be easily operationalized (in the sense of being difficult to define) were removed. For example, we removed *unit grouping* which only applies to unit visualizations and thus would not be applicable to many visualizations in our corpus (see the justifications in section 4.7). During the process, we also came up with new dimensions, partly inspired from past literature in psychology and data visualization.

After a seemingly stable set of definitions and dimensions was established and the corpus of visualizations had been fully categorized by the first author, we performed a multi-coder evaluation. We collectively wrote a code book, which the last three authors used to classify a random sample of 17 visualizations. This partial coding made apparent that some of the dimensions and their levels could be interpreted in different ways. We computed Fleiss Kappa coefficients for each dimension and found that for four out of eight initial dimensions, these were below 0.56 (moderate agreement). Consequently, we discussed difficulties and discrepancies in the coding and iterated one last time on the design space which led to the removal of one dimension and updated concept definitions. The first author then recoded all visualizations in the corpus. The codebook, the codes and the Kappa scores are available on OSF.[2]

The resulting design space has a total of seven dimensions. It shares three dimensions with Boy and colleagues [1] (see section 4.7 for more details), and includes four new dimensions.

4 Design Space

In the next sections, we first introduce basic terminology and a conceptual framework that will serve as a foundation for our design space, and then describe and motivate each design space dimension.

4.1 General Visualization Concepts

We assume for simplicity that all **datasets** are flat tables [45]. A **data item** (or simply **item**) is "an individual entity that is discrete", and which corresponds to a row in the table [45]. In our framework, a data item always corresponds either to *a person* or *a group of persons*. Meanwhile, a **data attribute** (or simply **attribute**) is "some specific property that can be measured, observed, or logged" [45]. Examples include the salary, height, or name of a person, or the average salary or height of a group of persons.

Visualizations are human-readable representations of data items and data attributes. Here too, our terminology largely follows previous literature. The main building block of a visualization is the **mark**, which is an element that represents a data item [45], [46], [47], [48]. A mark can either consist of a single graphical primitive (e.g., a point, line, or area [49]) or a combination thereof (e.g., a glyph or an icon) [50]. Meanwhile, **perceptual channels**[5] are properties of a mark that can be varied to convey the value of data attributes. Common examples include elementary graphical attributes such as size, value, color or shape [49].

2. Supplementary material on OSF: osf.io/wfgmp
3. Corpus: luizaugustomm.github.io/anthropographics
4. See, for example, *Visualizing smoking risk*: www.stubbornmule.net/2010/10/visualizing-smoking-risk/

5. This term generalizes the notion of visual channel [45] or visual variable [51] to non-visual data representations [52].

4.2 New Concepts

We assume that a visualization is created by a designer with an intent, i.e., a message to convey. For any such visualization, we define the **reference population** as the set of all persons who are the subject of the visualization's message, i.e., all the persons the visualization designer chose to tell a story about. For example, if a visualization tells a story about WWII casualties, we can assume that the reference population is the set of all people who died due to WWII, even if not all of these people are represented in the visualization, and even if not all of these people are known. Thus, the reference population is a conceptual set, which is not necessarily fully visible in the visualization, and about which we typically can only speculate.

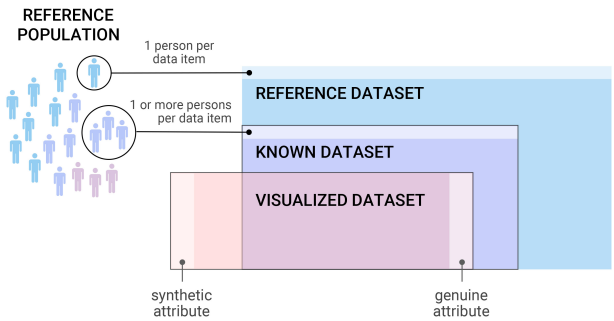

Fig. 3: The three conceptual datasets we use to capture the message and information conveyed by an anthropographic. CC-BY 4.0. Source: osf.io/zhq9x/.

In our conceptual framework, the message and the information conveyed by a visualization involve three datasets (see Figure 3). The first one, the **reference dataset**, is purely conceptual as it consists of all the people from the reference population – each mapped to an item, plus all of their characteristics, each mapped to an attribute. An example of a reference dataset is all migrants who died in the Middle East between 2015 and 2019, together with all the information possibly imaginable about them. In the rest we will assume that the reference dataset is given, and thus the designer has no control over it.

The **known dataset** is the subset of the reference dataset available to the designer, possibly aggregated in such a way that each data item corresponds to a group of people instead of a single person. It consists of a finite number of items and attributes, for example, the data about migrants who are known to have died in the Middle East between 2015 and 2019, as reported in the Missing Migrants dataset[6].

Finally, the **visualized dataset** consists of all the information — items and attributes — that is represented in the visualization, or in other words, all the information that can potentially be extracted from the visualization. It is a subset of the known dataset with potentially extraneous information added by the designer, for example, for aesthetic reasons or for storytelling purposes. We formalize this by introducing the notions of genuine and synthetic attributes.

Genuine attributes are attributes from the visualized dataset that originate from the known dataset. For example, a genuine attribute could be the age of the migrants who died in the Middle East between 2015 and 2019, as reported in the Missing Migrants dataset. On the other hand, **synthetic attributes** are attributes in the visualized dataset that do not originate from the known dataset. For example, a designer may assign an arbitrary gender to each person (through the use of a male or female silhouette) or may give them fictitious names (e.g., Figure 12). Synthetic attributes may occasionally match the attributes from the reference dataset, either by accident or because the designer made an informed guess. However, we generally cannot assume synthetic attributes to reflect reality, since they are by definition not known. Note that it is not necessarily visible in a visualization whether an attribute is genuine or synthetic.

Another way to characterize attributes is by their distinctiveness, which captures how much information they convey. The **distinctiveness** of an attribute or a set of attributes is the extent to which it allows to distinguish people or groups of people from each other. When data items are individuals, attributes with low distinctiveness are ones whose values are generally shared by many people (e.g., sex, age, or country of origin). Meanwhile, attributes with high distinctiveness are ones that few people have in common (e.g., full names or photographs). Many attributes lie somewhere in the middle of the distinctiveness continuum. When data items are groups of people (e.g., demographic groups or countries), distinctiveness refers to the extent to which the attributes allow to distinguish those groups from each other.

Finally, attributes can be represented in a visualization in two major ways. **Encoded attributes** are attributes from the visualized dataset that are mapped to perceptual channels. One example is mapping people's age to the height of bars. In contrast, **literal attributes** are attributes from the visualized dataset that are presented either in a literal form or using written prose or numerals. One example is fully spelling out somebody's age, e.g., with a label stating "22 years old". Some attributes such as names or photos do not easily lend themselves to visual encoding, and thus are almost always represented as literal attributes.

The rest of this section describes the dimensions that make up our design space. We illustrate the dimensions with a series of minimalistic visualizations, all of which assume the same known dataset consisting of ten persons and four attributes: whether each person prefers cats or dogs, their gender, their name, and a photo of their face (see Table 1). The reference population contains the same ten people.

6. Missing Migrants website: missingmigrants.iom.int

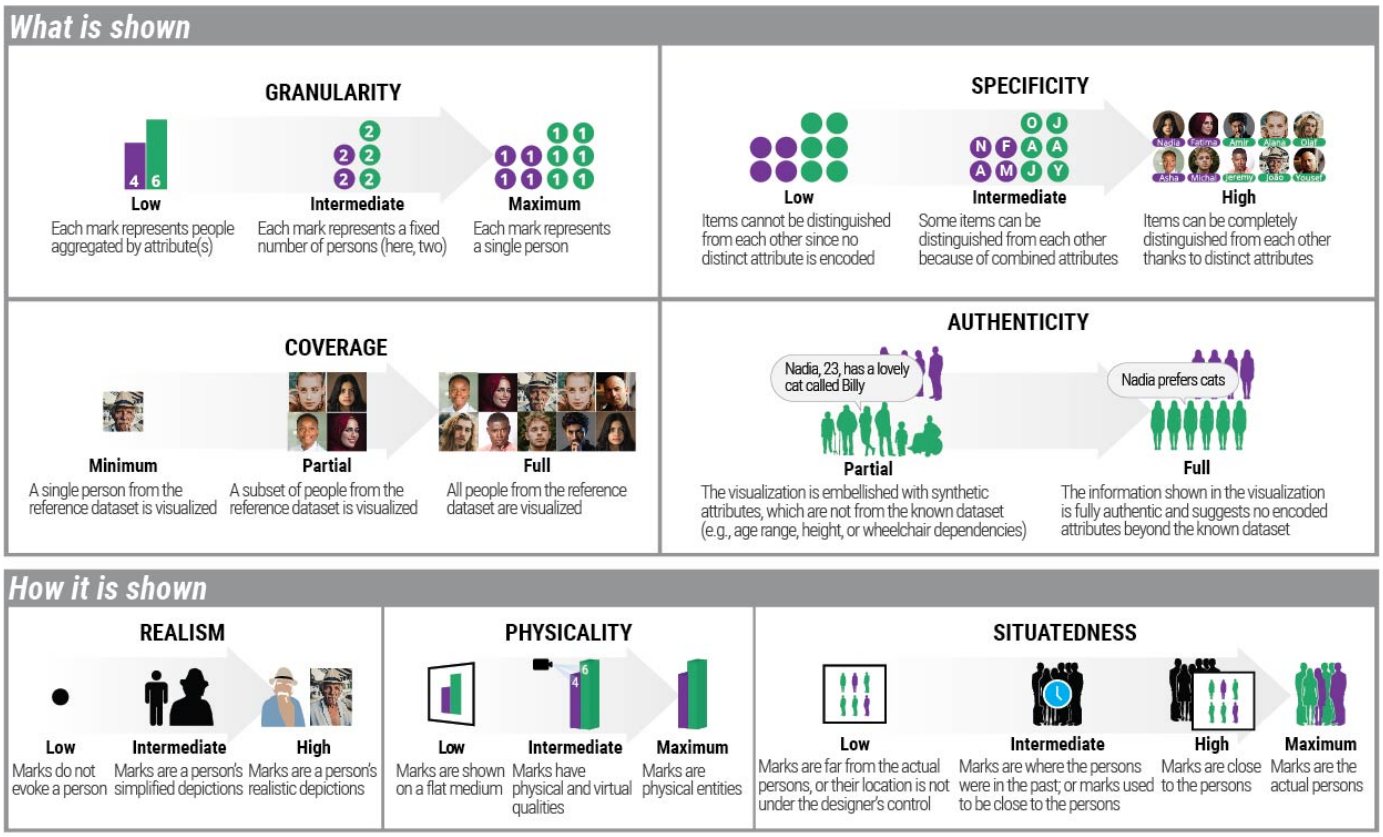

Fig. 4: The seven dimensions of the anthropographics design space. CC-BY 4.0. Source: osf.io/2zhjt/.

Name	Gender	Dogs or cats?	Photograph
Yousef	Male	Dogs	(image)
Amir	Male	Dogs	(image)
Alana	Female	Dogs	(image)
Olaf	Male	Dogs	(image)
Jeremy	Male	Dogs	(image)
João	Male	Dogs	(image)
Fatima	Female	Cats	(image)
Nadia	Female	Cats	(image)
Asha	Female	Cats	(image)
Michal	Male	Cats	(image)

TABLE 1: Fictional dataset underlying the illustrations of the design space dimensions in Figure 4.

4.3 Introducing the Design Space

The dimensions of our design space fall into two broad groups. The *what is shown* group corresponds to dimensions concerned with the information represented in the visualization (e.g., whether all people from the reference population are represented). Meanwhile, dimensions from the *how it is shown* group describe the way information is represented on the visualization (e.g., whether the marks look like people).

4.4 What is Shown

There are four dimensions related to what information and how much information is presented in a visualization.

4.4.1 Granularity

Granularity refers to the extent to which the persons in the visualized dataset are mapped to separate data items – and equivalently, to separate marks on the visualization.

In a visualization with **low granularity**, each mark corresponds to a *group of persons* who have one or more attribute values in common (see Figure 4). An example is depicted in Figure 5, where each segment of a different color represents the sum of immigrants in a period of time. Statistical charts where each mark represents a large number of people (thousands or millions) are typical examples of low-granularity visualizations. In such visualizations, length or area are commonly used to encode cardinality (for example, bar charts, bubble charts, or area charts). Position encoding is also occasionally used (for example, in line charts of population growth).

In a visualization with **intermediate granularity**, each mark maps to a *fixed number of persons*, this number being greater than one. Many so-called Isotype visualizations [53], [54] fall into this category. Figure 6 shows an example of an Isotype visualization, where each mark represents one million soldiers. Since one million is a large number, the granularity is relatively low in absolute terms, but it is higher than if people were fully aggregated according to the three categories "killed", "wounded" or "others returning home". While most Isotype-like designs show absolute counts, a variation over this design uses marks to show rounded percentages (that is, the counts from all marks sum up to 100 people).

In a visualization with **maximum granularity**, each data item corresponds to *a single person*. For example, the visualization in Figure 7 has maximum granularity because each data item stands for a different person who died by a gunshot in 2019.

Note that in the case of low granularity, the number of persons is typically encoded using perceptual channels, whereas in intermediate and maximum granularity, it is mapped to the number of marks.

Using visualizations that represent people as individual marks allows designers to convey specific information about each person, be them genuine attributes detailing those portrayed, or synthetic attributes to humanize the data (for example, some of the designs by Boy and colleagues [1]). There is also a belief that representing people using single marks can make readers empathize with the persons represented [6]. However, maximum granularity may have drawbacks with large datasets: representing thousands of people as different marks demands a large space or forces the designer to reduce each mark to a speck with little details. Nonetheless, visually compelling examples of such visualizations do exist, such as the unit visualizations introduced by Drucker and colleagues [55].

It is still necessary to investigate in which scenarios there are relevant benefits in using visualizations with higher granularity. The only study that investigated visualizations with intermediate granularity in the context of human rights was not able to confirm the hypothesis that higher granularity leads to more empathy [1]. More studies are needed to examine the role of granularity in promoting prosocial feelings or behavior.

4.4.2 Specificity

The **information specificity** of a visualization (or simply **specificity**) corresponds to *how distinctive* is the *entire set of attributes* in the visualized dataset (see section 4.2 for the definition of distinctiveness and Figure 4 for an illustration). The more the attributes allow to distinguish between data items (either individuals or groups of people) the higher the visualization's specificity. All attributes contribute to a visualization's specificity, whether they are encoded or literal, and whether they are genuine or synthetic.

In a visualization with **low specificity**, items tend to be visually very similar to each other and most of them cannot be distinguished. The Isotype visualization in Figure 6 is an example of a low-specificity visualization, where the low distinctiveness of the visualized attributes (survival status and side) contributes to making the soldiers look rather deindividualized. The gun violence visualization (Figure 7) also has somehow low specificity, because the only visualized attribute is the location where each person died. In this case, location can be thought to have relatively low distinctiveness because there are spots on the map where more than one person might have died, which makes it hard to distinguish the victims. Had the map been higher-resolution or zoomable, specificity would have been higher.

In a visualization with **intermediate specificity**, attributes in the visualized dataset allow the reader to easily distinguish many data items from each other. A typical approach is to use glyphs to represent multidimensional data about individual people. For example, in Figure 8, the visualization shows six attributes. Although each attribute has low distinctiveness, once combined together, the attributes give some sense of individuality to each person.

Finally, in a visualization with **high specificity**, the attributes in the visualized dataset allow the reader to perfectly

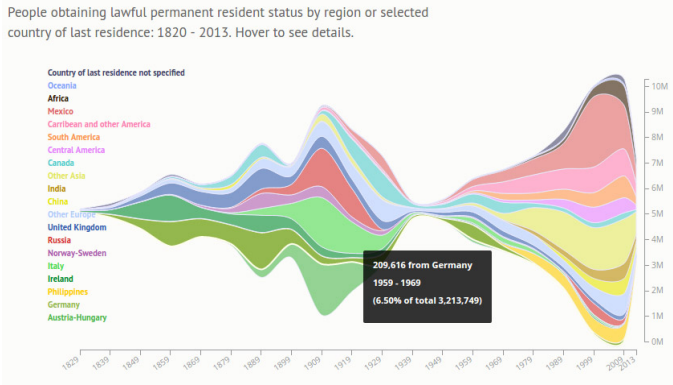

Fig. 5: *200 Years of Immigration to the US*. This stream graph shows immigrants by region from 1820 to 2015. Each color corresponds either to the country of last residence or a region. The area represents the number of immigrants who moved to the U.S. over the years. Hovering over an area shows the total number of immigrants within a period of 10 years until the selected date. Source: Insightful Interaction [56].

Fig. 6: *The Great War 1914-18*. Each Isotype mark represents 1 million soldiers. Black crosses represent soldiers who died, soldiers in red came back wounded, and soldiers in green returned without major physical injuries. The soldiers on the left are the Allies and the ones on the right belong to the Central Power. Source: LA Worrell [57].

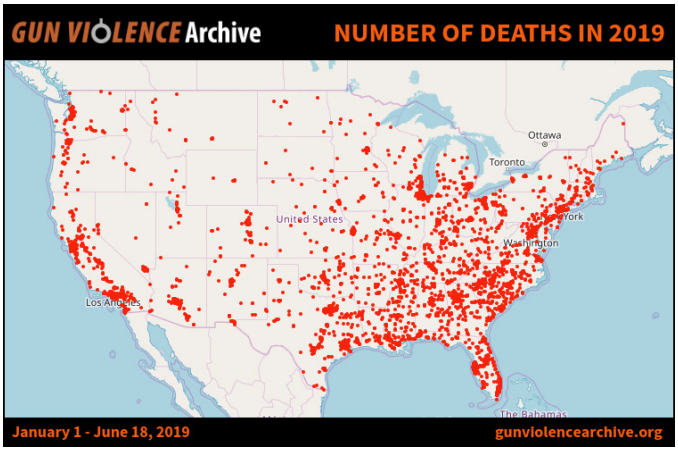

Fig. 7: *Gun violence Chart of Deaths in 2019*. Depicts all deaths caused by guns in the USA from January to June, 2019. Each red dot is a victim, and its placement indicates where the person was killed. Source: Gun Violence Archive [58].

distinguish all data items. One example is shown in Figure 9, where people's faces and other physical characteristics make each person unique. A visualization can also have high specificity when it shows a set of attributes that is so large that the data is necessarily unique to each person. In Figure 11, for example, the visualization shows the story of a child coping with an auto-immune disease. Since in our framework data items are always people or groups of people, the visualized dataset is comprised of a single data item, and thus the visualization consists of a single mark that can be thought of as an extremely complex glyph.

Studies suggest that showing pictures and details of victims is linked to a higher likelihood of donating to charity [59], [60]. Although the study by Boy et al. [1] has failed to find an effect for intermediate-specificity visualizations with intermediate granularity, it remains possible that other designs could promote prosocial feelings and behaviors. Although privacy issues often make it hard to reach high specificity in a visualization [1], it is sometimes possible to reach intermediate levels of specificity without compromising privacy, for example by showing people's first names, age, and other non-identifiable data, as in Figure 10.

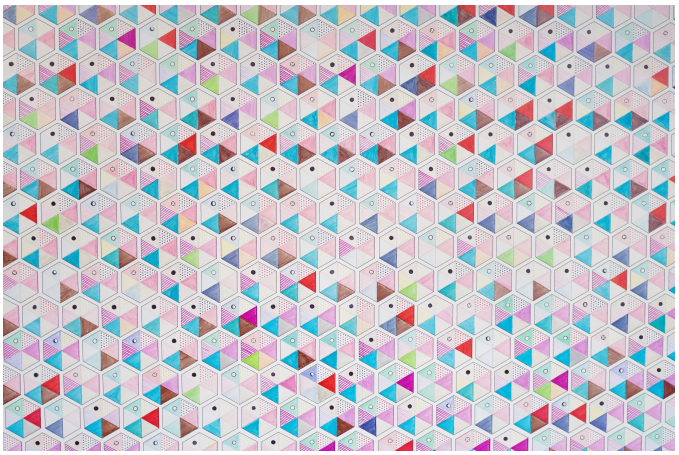

Fig. 8: *Data wallpaper*. Each glyph represents 6 questions answered by a visitor to a store, ranging from how they see the future of work to how they unlock their creativity. The designer initially installed an empty visualization template as a wallpaper in the store, and the store employees filled the hexagons as they were receiving visitor responses. Source: Giorgia Lupi [61].

4.4.3 Coverage

Coverage corresponds to the extent to which the visualized dataset includes the persons from the reference dataset. Although in many visualizations it may not be possible to identify with certainty the reference population considered by the designer and consequently the coverage, it remains possible to speculate, and more generally, to reason about coverage on a theoretical level.

In a **minimum coverage** visualization, the reference population consists of more than one person but the visualized dataset only contains data about *a single* person from that population.

In a visualization with **partial coverage**, the visualized dataset contains data about a *subset* of people from the reference dataset. The people may be selected at random or chosen according to some specific attributes. For example, the visualization in Figure 10 delivers a message about women who have breast cancer in Brazil and are in remission, but only includes data about 13 of these women. Another example of a partial coverage visualization is shown in Figure 2: "The Stories Behind a Line" starts by showing six lines, each of which contains initials of a refugee. The user can click on a line to see the story of a person who left their home seeking a better life. In this example, the six individuals were presumably chosen in an arbitrary manner, and used as illustrations in order to tell a broader story about the arduous life of thousands of refugees around the world.

In a **full coverage** visualization, the visualized dataset contains *all* the persons from the reference dataset. Examples are Figures 5, 6 and 7.

While full coverage is the most straightforward design choice, partial coverage is necessary when datasets are incomplete (e.g., not all bodies have been found in a disaster), or it may be an appropriate choice when visualizing rich personal data that needs to be explicitly gathered from the people, or when complete coverage may violate the privacy of the persons represented (for example, in Figure 10). A related advantage of partial and minimum coverage is that

Fig. 9: *100% Montréal* is composed of 100 persons from Montréal who have been selected to represent the city population. Depending on personal questions asked during the show, the persons moved to the left or the right of the green circle. Source: Rimini Protokoll [62].

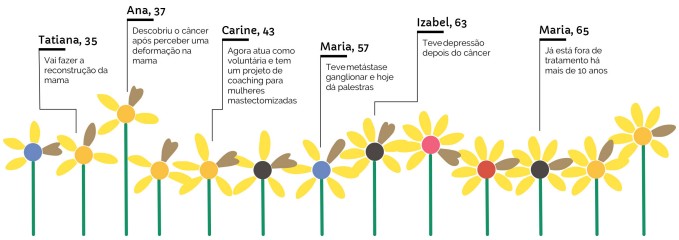

Fig. 10: *Cancer is not always the end* shows data about 13 women with breast cancer who has been in remission for years. Each flower is a woman and each part of a flower shows a characteristic of the woman such as her age or how she dealt with the treatment. Information about some women is also shown by annotations. Source: Oncoguia [63].

they reduce the number of marks, and thus leave space to show richer data about fewer individuals, for example, for storytelling purposes, where a more detailed profile of an individual serves to anchor the story being told. Furthermore, there is evidence suggesting that telling the story of a single suffering individual can better promote compassion than telling the story of an entire group [59]. Similarly, studies have suggested that as the number of suffering people increases, people feel less empathy for them and donate less [38]. Therefore, it is possible that visualizations with partial or minimum coverage could help observers be more compassionate about suffering populations. However, it remains necessary to test this hypothesis empirically.

4.4.4 Authenticity

Authenticity refers to the proportion of genuine attributes in the visualized dataset.

A visualization with **partial authenticity** contains both genuine and synthetic attributes (terms are defined in section 4.2). The more visualized attributes are synthetic, the less authentic a visualization is. The illustration of authenticity in Figure 4 shows a visualization with partial authenticity where the designer gave the marks different silhouettes to make them look unique: we can see, for example, a child, and a person in a wheelchair. The designer also annotated an individual with personal information. However, none of this information is in the known dataset (shown in Table 1) which makes these attributes *synthetic*. A real world example of a partially authentic visualization is shown in Figure 12. The only genuine attribute is whether a person lived above or below the poverty line in 2010, whereas the name and gender of the persons are synthetic attributes that were probably used to increase information specificity and make the persons look more unique, or to increase the visual appeal of the visualization.

In a visualization with **full authenticity**, all visualized attributes are genuine. In Figure 4 (authenticity dimension), the visualization with full authenticity uses realistic silhouettes like the visualization with partial authenticity. However, only two types of silhouettes are used in the fully authentic visualization, which serve to encode gender, an attribute that is in the known dataset. We will see in section 4.5.1 that realistic silhouettes like these cannot be fully authentic unless they are those of the people represented, but in this case they can be considered as (almost) fully authentic.

While full or close-to-full authenticity is the most obvious design choice, designers sometimes use synthetic attributes in visualizations about people such as in Figure 12 or in visualization 67 in the corpus. In those examples, the designers used anthropomorphic marks with different genders and ages that likely do not originate from the known dataset, possibly as an attempt to promote compassion. However, it is possible that this technique can backfire as a result of readers feeling manipulated, possibly causing them to doubt even the genuine information and the visualization as a whole. As far as we know, this technique and its possible trade-offs have never been explicitly discussed in the information visualization literature.

Fig. 11: *Bruises: the Data We Don't See* shows the progress of a child in coping with an auto-immune disease. Each petal is a day. Red dots are platelet counts (the disease destroys them). Colors represent various events such as bleeding, medications, or positive feelings. The text around the petals are the mother's notes about the day. Source: Giorgia Lupi [64].

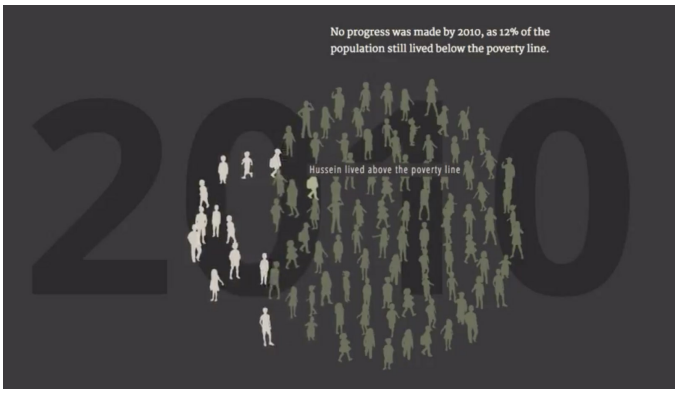

Fig. 12: *Poverty in Syria* shows the proportion of children below (lighter color) or above the poverty line (darker color) in Syria, in 2010. Each mark is a fixed number of children. Hovering over a mark shows synthetic information about a person. Source: Boy et al. [1] (screenshot of the first experiment's video from the supplementary material).

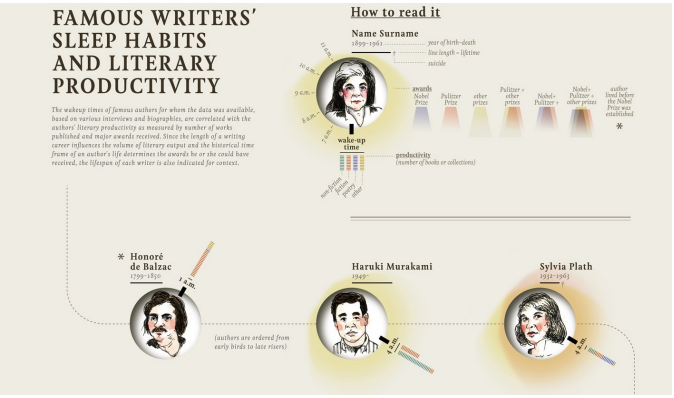

Fig. 13: A part of *Famous Writers' Sleep Habits and Literary Productivity*. Detailed drawings of famous authors are combined with illustrations of the time of day at which they rose and how productive they were throughout their lifetime. Source: Giorgia Lupi [65].

4.5 How it is shown

This section describes design dimensions that capture how information is represented. All of the dimensions here are properties of *marks*. The properties of marks are often coherent across a visualization, and thus a visualization can be characterized with respect to how its marks are represented.

4.5.1 Realism

Realism refers to the degree of resemblance of the visualization's marks to actual persons.

A visualization with **low realism** represents people or groups of people using *symbolic marks* that are non-anthropomorphic, that is, they bear no resemblance with a human. Such marks include dots, bars, abstract glyphs, or shapes that evoke inanimate objects. Figures 5, 8, and 10 are examples of low-realism visualizations.

A visualization with **intermediate realism** is made of *pictorial anthropomorphic marks*. Examples of such marks are simple icons or human silhouettes. The visualizations from Figures 1 and 12 are also visualizations with intermediate realism, since they are composed of human-shaped icons or silhouettes.

A visualization with **high realism** is made of *realistic anthropomorphic marks*, which closely resemble an actual person. An example is shown in Figure 13, where the center of all marks feature detailed drawings of famous writers. Other examples of realistic marks include 3D avatars, physical sculptures, and even real persons, such as the data physicalizations from Figures 9 and 14.

Realism interacts with authenticity: strictly speaking, realistic marks can only be considered fully authentic if they faithfully depict the people. This is not the case in the second visualization of the "authenticity" dimension in Figure 4, where the detailed silhouettes suggest additional attributes about the persons that are unknown, such as age or body mass index. However, visualizations that use many identical realistic marks or a small set of mark types to encode genuine data are almost fully authentic, since *i)* the repetitions clarify that realism is used for embellishment *ii)* the *variations* across marks encode genuine data. A counter-example is visualization 17 in the corpus, which uses many different silhouettes that could be easily interpreted as showing the actual silhouettes of the persons, although they do not.

The realism of anthropomorphic marks is best thought of as a continuum, as real persons are higher on the realism spectrum than photographs, which are themselves higher than drawings, which are in turn higher than simple icons. Designers typically use anthropomorphic marks to reinforce that the data is about real persons. It has been hypothesized that doing so could promote empathy [1], [5], [67], and that the more realistic the marks are, the more effective they could be at promoting empathy [1]. However, none of these hypotheses and potential trade-offs between realism and perceived authenticity have been experimentally confirmed.

4.5.2 Physicality

Physicality refers to the degree to which a visualization's marks are embodied in physical objects as opposed to shown on a flat display [52]. The physicality continuum is illustrated in Figure 4.

Fig. 14: *San Francisco Gay Men's Chorus representing AIDS deaths until 1993*. The men in white are the surviving members of the original San Francisco Gay Men's chorus, while the persons in black with their back turned represent the members lost to AIDS. Source: San Diego Gay and Lesbians News [66].

Fig. 15: *Of All the People in All the World: Stats with Rice* is an installation using grains of rice to show various statistics about people where one grain corresponds to one person. This image represents every person at the Aston Villa vs. Birmingham City football match on 20 April 2008. Credit: Jon Bounds, under CC 2.0 license. Source: Flickr.

In a visualization with **low physicality**, the marks are shown on a flat medium, such as a computer screen or a sheet of paper. All visualizations designed for the web or for magazines (for example, Figures 6, 12, and 13) fall in this category. The visualizations from Figures 8 and 16 also have low physicality because the former is drawn on a flat wall, while the latter is shown on a wall clock with a printed background.

Visualizations with **intermediate physicality** are characterized by marks that have both physical and virtual qualities. While we could not find an example featuring data about people, examples exist for other types of datasets. For example, the Emoto installation [68] shows tweets, of which some attributes are encoded in physical shape while others are video-projected.

In visualizations with **maximum physicality**, the marks are physical objects or actual persons. Examples are shown in

Figure 15, where the marks are grains of rice, and in Figures 9 and 14, where the marks are real persons.

Le Goc and colleagues [69] hypothesize that it is easier to empathize with people when they are represented by physical objects than when they are represented by virtual objects. Yet it is so far unknown whether using physical marks may indeed increase prosocial feelings such as empathy or compassion.

4.5.3 Situatedness

Situatedness refers to how spatially close a mark's physical presentation [71] is — or was[7] — to its physical referent [46]. In the context of anthropographics, the physical referents are the persons described by the data. The situatedness continuum is illustrated in Figure 4.

In visualizations with **low situatedness**, which we will also refer to as *non-situated*, the marks are either presented far from the persons they represent, or their physical location is not under the control of the designer. This includes all visualizations designed for magazines or the web. Figures 1 and 7 are examples of non-situated visualizations because they are displayed on computer screens which are in most cases far from the represented victims.

In a visualization with **intermediate situatedness**, the marks are either presented at a location where the persons used to be in the past, or the marks used to be in proximity to the persons they represent. The Data Wallpaper (see Figure 8) is located at this intermediate point of the continuum because the persons who are represented in the visualization provided their personal data next to the visualization, but left the exhibition space afterward.

Visualizations with **high situatedness** are made of marks that are presented close to the persons they represent. An example of this point on the continuum is the Activity Clock (see Figure 16), in which the authors installed a visualization of lab presence data in the cafeteria of the lab itself. The visualization is highly situated because the people it represents are (generally) near the visualization.

Finally, there is **maximum situatedness** where the marks are the persons they represent. This is the most extreme point on the situatedness continuum and corresponds to physical visualizations made of real people, and showing data about those people. The show 100% (see Figure 9) is a maximally-situated visualization where the persons split themselves into groups or hold signs of different colors according to the questions they are asked.

It is possible that achieving at least some degree of situatedness can, in some cases, help observers relate to the people represented. Outside of visualization, situatedness has long been thought to affect people's emotions: memorials are often placed in a location where a significant event has affected a person or a group of persons, and memorabilia can acquire emotional power by virtue of having been touched or worn by a person [73] (both are examples of intermediate situatedness). In contrast, most visualization designs are meant to be easy to replicate and share, which maximizes the number of people who can see them but also means they have low situatedness [46].

7. As we will see, we generalize Willet and colleagues' [46] notion of situatedness by also considering spatial relationships in the past.

Fig. 16: *Activity Clock* shows the aggregated presence of persons in a lab from 8 AM to 8 PM during a 3-year period. Each bar is a 15-minute bin with the 10th and 90th percentile of the number of persons in that time. White dots and the color encode the median number of people in the corresponding time. Source: Morais and colleagues [70].

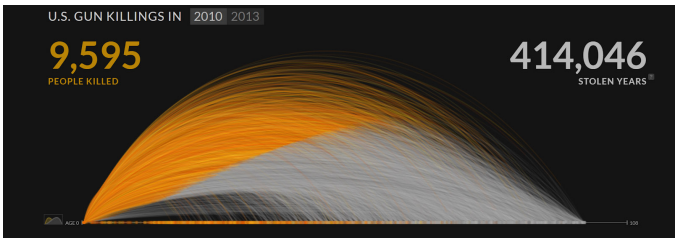

Fig. 17: *U.S. Gun Deaths in 2010 and 2013*. Each line represents a person. The orange segment corresponds to the period lived and the gray segment represents the estimated years stolen from the person. Readers can select the dataset by choosing the year and also filter by sex, age, region, and time. Source: Periscopic [72].

4.6 Interactivity and Animation

We have so far assumed static visualizations. Although many visualizations are indeed static, others are *dynamic*, i.e., they can change under the user's influence (interactive charts) or outside the user's influence (animated visualizations such as in educational videos) [74]. Among the 105 visualizations in our collection, 44 are dynamic.

A dynamic visualization can be thought of as a (potentially very large) set of static views, each of them providing a different perspective on the data. There are three major ways in which users can explore data about people by moving from view to view.

First, dynamic visualizations can let users explore different *sets of data items* — people or groups of people — over time. Figure 17, for example, shows a visualization where users can explore the persons who died by a gunshot in different years (which can be chosen by clicking on the corresponding year). Similarly, the animation from Figure 12 changes the dataset while switching from 2010 to 2016.

Second, dynamic visualizations can let users explore different *sets of attributes* for a given set of data items. In Figure 17, users can filter victims by sex, age, or region. This approach is useful when the number of perceptual channels is insufficient to display all attributes simultaneously: users

can learn more and more about people over time. In addition, in interactive visualizations, users can focus on people's attributes they care the most about.

Finally, dynamic visualizations can let users explore the same data — items and attributes — through *different representations*. In Figure 18 for example, users can smoothly transition between different representations of homelessness data such as maps, grids, and bar charts. Drucker and colleagues have demonstrated many other examples of animated transitions between representations of different granularities with their system SandDance [55], [75].

Since the different views of a dynamic visualization may have different characteristics according to our design space, interactivity and animation can offer a way for users to dynamically navigate in the anthropographic design space, both in terms of what is shown and how it is shown. The dynamic labels of visualizations from figures 1, 7, 12, and 17, for example, increase the information specificity by showing detailed data about each individual. Another dimension that can change in dynamic visualizations is granularity. For example, the chart from Figure 18 lets users change its granularity from intermediate (every dot represents 5 persons) to low (all the dots are combined to form a bar).

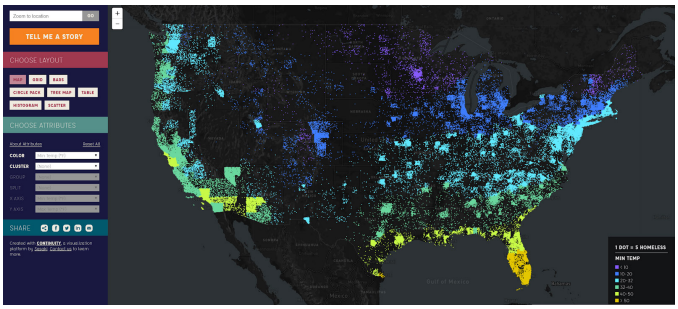

Fig. 18: *Understanding homelessness in USA*. This interactive unit visualization represents the homeless population in the USA. Each dot corresponds to five homeless people. The representations can be chosen among maps, grids, bars, etc. The viewer can also explore different factors from categories such as geographic, economic, or social. Source: Understanding Homelessness [76].

4.7 Differences with Boy and Colleagues

As we mentioned previously, our design space of anthropographics extends an earlier proposal by Boy et al. [1]. Our extension both broadens the original design space (that is, it captures a larger variety of designs) and sharpens it (that is, it makes finer distinction between related designs).

Boy et al's design space had four main dimensions:

- **Class of visualization**. This dimension distinguishes between *unit* and *aggregate* visualizations. It maps to the *granularity* dimension of our framework. In contrast with the original dimension, granularity sits on a continuum and distinguishes between two types of unit visualizations: those where each mark represent a single person, and those where each mark represents a fixed number of persons.
- **Human shape**. This dimension consists of two subdimensions:
  - *Realism* (*abstract—realistic*) directly maps to our *realism* dimension. However, our framework expands the definition of realism to also include more realistic marks such as photographs and real humans.
  - *Expressiveness* (*neutral—expressive*). We initially included expressiveness as a "how it is shown" dimension but removed it after our multi-coder evaluation because we found it hard to define, especially when considering non-anthropomorphic marks. We also realized that the expressiveness of a visualization often arises in large part from the meaning of the dataset and can be manipulated by a variety of visual design strategies like the use of metaphors, that are hard to operationalize.
- **Unit labeling**. This dimension captures three types of text annotations that can be displayed on top of marks: *generic*, *iconic* and *unique*. In our framework, it is incorporated into the more general *information specificity* dimension. Unit labeling is more specifically captured by the concept of attribute *distinctiveness*, which formalizes Boy et al.'s notion of uniqueness and generalizes it to other types of information beyond text annotations. Among other things, distinctiveness captures the use of unique anthropomorphic shapes, which was also discussed by Boy et al., but as part of the realism dimension which we see as orthogonal.
- **Unit grouping**. This dimension captures the spatial layout of the marks, such as *grid-based* or *organic*. We decided not to include this dimension in our framework because it is specific to unit visualizations and cannot be easily generalized to low-granularity visualizations. We also could not find arguments in the past literature in support of spatial layout influencing prosocial feelings or behavior.

Much of these modifications to the original design space were meant to cover a wider range of designs, and thus situate anthropomorphic unit visualizations within a larger design space of visualizations of data about people. Our conceptual framework also extends the original framework in a number of other ways, including by making a useful distinction between two classes of design space dimensions: *what is shown* (which includes two of Boy et al.'s dimensions), and *how it is shown* (which includes another dimension). Our design space also introduces two additional dimensions in each category: *coverage* and *authenticity* in the "what is shown", and *physicality* and *situatedness* in the "how it is shown", all of which we argued are relevant dimensions to consider when designing anthropographics. Finally, like Boy et al.'s work, our design space focuses on design dimensions that could plausibly promote compassion, but it includes more extensive discussions of why this should be the case, and of the underlying trade-offs. At the same time, unlike Boy et al., our work does not contribute any empirical finding.

5 Corpus Analysis

With a corpus of data visualizations and a design space to describe them, we now turn to reflect on the combinations of design choices we observed in our collection of 105 visualizations. Although our collection is not a random sample of all existing visualizations and is likely biased, it can still be informative as a proxy to what exists. For example, if there is a certain type of design we did not particularly emphasize during our search and yet it appears often in our collection, then this provides an indication that this

design has some popularity. Conversely, if there is a design (i.e., a specific combination of dimensions) that would have caught our eye but of which we found no example, then this should be an indication that this design is at best relatively uncommon. We also provide a website[8] on which the corpus can be explored using faceted search.

5.1 Distributions of Dimension Values

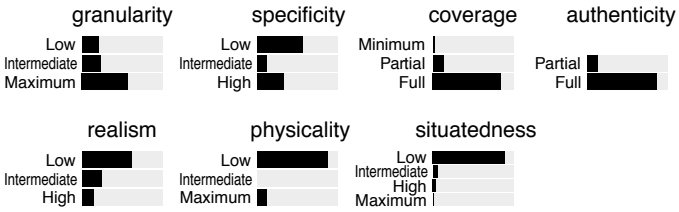

Fig. 19: Distribution of dimension values in our corpus. Each black bar indicates the number of visualizations having a specific value on a given design dimension. CC-BY 4.0. Source: osf.io/ambsk/.

Figure 19 shows the distribution of dimension values in our corpus of 105 visualizations. Given our focus on anthropographics during our example collection, it is not surprising to see a high frequency of maximum-granularity visualizations, and a good number of high-specificity visualizations. Meanwhile, there are uncommon dimension values that likely reflect a bias in the population of existing visualizations rather than a bias in our collection method. For example, minimum coverage and partial authenticity are design choices for which we have examples, but which do not appear to have received much attention by designers. Some other uncommon values are likely due to the higher technical difficulty and costs implied by these design choices: few visualizations in the corpus have high or maximum situatedness, and no visualization has intermediate physicality.

Next, we looked at the most uncommon combinations of dimension values.[9] We report on the three most uncommon combinations (all with $p < .001$). First, no visualization has both high realism and low specificity, reflecting the fact that when marks are realistic depictions of people (either photos or drawings), they can be easily distinguished from each other. Although realistic marks can in principle be duplicated to achieve low specificity, we have found no example of such a design. The second combination is intermediate granularity with high specificity, for which there is no occurrence either. High specificity seems hard to attain with lower levels of granularity, perhaps because attributes at the group level tend to be less distinctive. The third combination is partial coverage with low specificity, again with no occurrence in our corpus. It seems that when designers opt for partial coverage, this is is associated with the visualization of more distinctive attributes about the persons represented.

8. luizaugustomm.github.io/anthropographics

9. We first computed the expected number of visualizations having each possible pairwise combination of values, assuming independent dimensions. For example, if out of N visualizations, n_1 have low granularity and n_2 have high specificity, we should expect $n_{12} = n_1 \times n_2/N$ to have both. We then computed the p-value of the binomial test using the observed frequency as data and n_{12} as the null hypothesis.

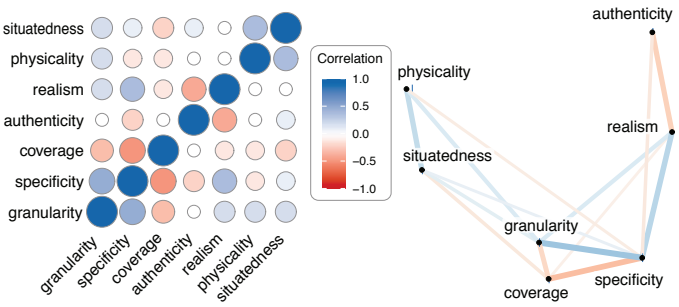

Fig. 20: Correlations between design dimensions in our corpus. The left plot is a correlation matrix, where blue circles are positive correlations and red circles are negative ones. The right plot shows the same data with a graph — the closer two nodes, the higher the absolute correlation. CC-BY 4.0. Source: osf.io/5qrkm/.

Finally, Figure 20 shows the correlations[10] between dimensions. The highest absolute correlation is between granularity and specificity ($r = 0.50$). More generally, granularity, specificity, and coverage are closely related: higher specificity is usually associated with higher granularity and lower coverage. Higher realism also tends to co-occur with high specificity and granularity. This seems to suggest that anthropographic strategies (i.e., high granularity, high specificity, high realism, low coverage) are often used in combination. Similarly, the positive association between situatedness and physicality indicates that designers often explore situated anthropographics when using physical marks. Finally, higher realism tends to be associated with higher specificity as discussed before, as well as lower authenticity, due to the occasional use of realistic depictions of people as a form of decoration. All other correlations are $|r| < 0.25$.

5.2 Families of Visualizations

The dimensions that vary the most in our corpus are realism, granularity, and specificity. Based on this, we identified six families of common designs, grouped in two large classes: non-anthropomorphic and anthropomorphic. We also identified five atypical visualization designs. Figure 21 provides an overview of the families and how they relate.

5.2.1 *Non-anthropomorphic Designs*

Non-anthropomorphic visualizations are those whose marks do not resemble a human being. We break down this large class into three common families.

Statistical charts have low granularity. Their goal is to convey patterns rather than specific information about individuals. In Figure 16, for example, observations are aggregated in a way that does not reveal data about individuals. Though statistical charts can be low in information specificity, they can also have intermediate to high specificity, like the visualization in Figure 5 which conveys rich information about each demographic group. However, having low granularity, they cannot show rich information about individuals.

10. Correlations were computed by first numerically coding each dimension value (e.g., 0 for minimum, 0.5 for intermediate and 1 for full) and then computing pairwise Pearson correlations.

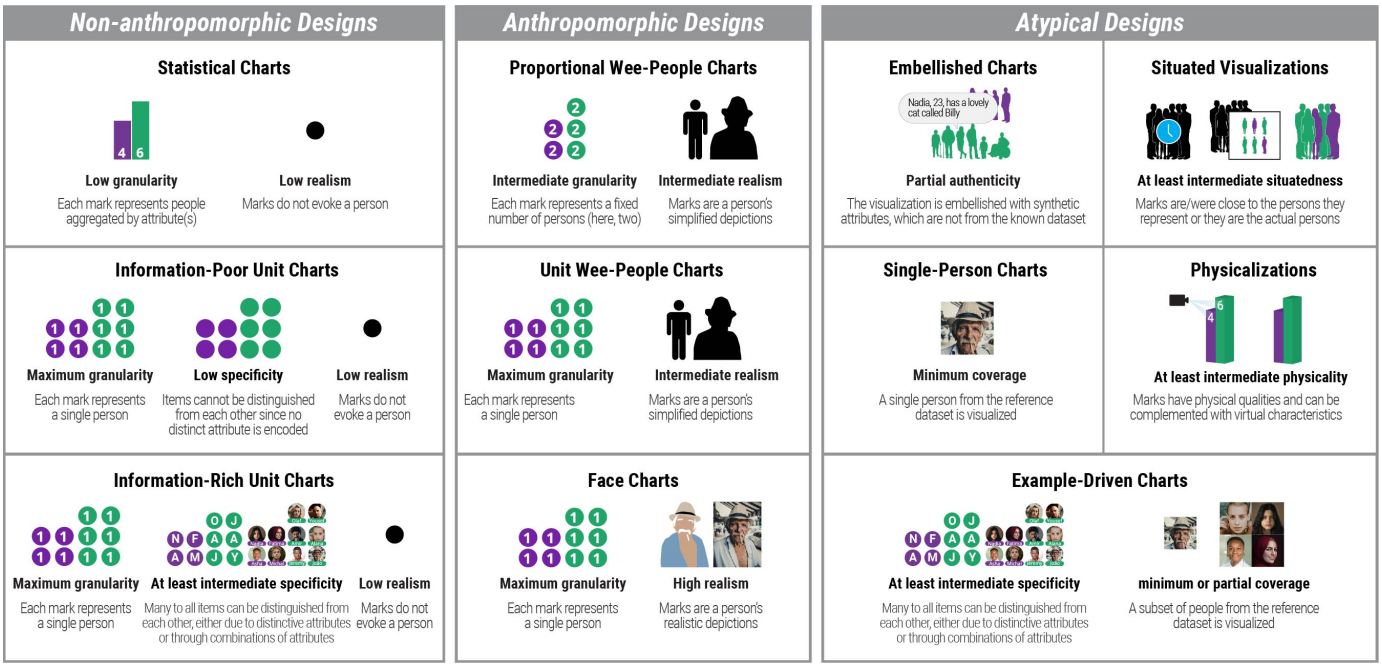

Fig. 21: An overview of the identified visualization families and the corresponding levels of the design space dimensions. CC-BY 4.0. Source: osf.io/p265q/.

Information-poor unit charts are non-anthropomorphic visualizations with maximum granularity but low information specificity. Each individual is visible but little information is conveyed about them (Figures 7 and 15). They are typically used when the intent is to convey numbers of people, or how people are distributed across a few attributes.

Information-rich unit charts have maximum granularity and intermediate to high information specificity. They present various data attributes — often both encoded or literal, to convey detailed information about each person. Examples are Figures 10 and 17.

5.2.2 Anthropomorphic Designs

Anthropomorphic visualizations are a broad class of visualizations whose marks resemble human beings. Two common variations are wee-people designs and portrait-like designs.

Proportional wee-people charts are anthropomorphic unit visualizations with intermediate realism and intermediate granularity. They usually have low information specificity and use pictorial marks to represent a fixed group of persons. Examples are Figures 6 and 12.

Unit wee-people charts are anthropomorphic unit visualizations with intermediate realism and maximum granularity. They have varying degrees of information specificity: examples 10 and 24 in the corpus have low information specificity, while the one in Figure 1 achieves a high level of specificity by providing detailed information about each individual.

Face charts are high on the realism dimension. Their marks are either photographs, drawn portraits, or the people themselves. In Figure 13, for example, the chart shows detailed faces of famous authors. Since a face is a highly distinctive attribute, all the visualizations in this category have high information specificity.

5.2.3 Atypical Designs

This section discusses less frequent design choices in the design space of anthropographics, which tap into the dimensions of authenticity, coverage, situatedness, and physicality.

Embellished charts are visualizations with partial authenticity. The same way a good storyteller can embellish a true story to make it more poignant, here the designer embellishes real data with synthetic information to enhance its impact. Embellished charts are often anthropomorphic. An example is shown in Figure 12. Another example is presented in Figure 14, where the persons who represent the men who died from AIDS are not the actual persons being represented.

Single-person charts have a single person in the visualized dataset. They are typically non-anthropomorphic and often convey rich information about the person. Figure 11 provides an example. Many examples exist outside of anthropographics, such as biographical and autobiographical visualizations [30], [31], [77]. Single-person charts can have full or minimum coverage depending on whether the message is about a specific person, or about a larger population of which one person has been taken as a representative. The visualization in Figure 11 is ambiguous in that respect, because although the data is about a person the designer knows personally, one of her stated goals was to "empower patients and families dealing with illness or disease" [64].

Example-driven charts have minimum to partial coverage. They convey a message about a population through one or a few individuals. Focusing on a small number of people creates opportunities for showing richer data since there is more space to show attributes, and it is easier to collect rich data on a few people. Thus, example-driven charts often have intermediate to high information specificity. See Figures 10 and 2 for two examples. As discussed above, example-driven charts can also be *single-person charts*.

Situated visualizations bear a spatial proximity relationship with the people they represent. In Figure 9, the marks are the persons themselves (thus they are also *face charts*). Other examples are participatory visualizations where each person crafts their own physical mark (e.g., visualizations 39 and 44 in the corpus). Situated visualizations usually cover small populations as building them can be costly, especially when the marks are physical objects or people.

Physicalizations are visualizations in which the marks are embodied by physical objects as opposed to shown on flat displays [52]. They can take on many different forms. For example, some are anthropomorphic (e.g., Figure 14) while others are not (e.g., Figure 15).

6 Discussion

In this section, we discuss opportunities for design suggested by our design space, as well as opportunities for research and limitations of our work.

6.1 Opportunities for Design

Some relevant design choices and directions are visible in our design space but appear underrepresented in our current collection of examples. We discuss four of them.

6.1.1 Hybrid Designs

One particularly promising direction is to explore designs that share some of the qualities of statistical charts and some of the qualities of anthropographics. Statistical charts are widely used in different domains and for various purposes, including for conveying facts about suffering populations. However, since they lack properties expected to promote compassion, they may be experienced by many as "cold" and "reducing people to numbers". In contrast, there is a growing community who argues that charity should be driven by reason and facts to have a real impact [78], and for whom many of the designs discussed here may appear as focusing too much on emotions. A key challenge, then, is to find designs that can appeal to both types of thinking.

One way the two approaches could be mixed is through the use of interactivity, for example, based on Harris' concept of *Near and Far* [5]. Visualizations could let readers dive deeper into personal stories by filtering or querying specific persons or attributes, while offering the possibility to explore statistical patterns through aggregated views. Those detailed views could use anthropographic design principles to help readers relate to specific individuals. The hope is that readers can understand statistical information for effective data-driven decision-making while being able to understand and empathize with the personal experience of individuals.

6.1.2 Customization and Targeting for Compassion

A person's emotional response to a visualization typically depends on their relationship with the people visualized. For example, a person may find it harder to care about the plight of distant populations than about difficulties encountered by people who live nearby and are similar to them [4]. Although this relationship is mostly outside the designer's control in terms of representation, a visualization's emotional potency may be enhanced if the designer plays with the visualization's narrative by carefully selecting which information to show to whom.

Allowing people to visualize data based on a familiar setting such as the region the user lives or incorporating data from the audience into the visualization are narrative design patterns that are intended to make users relate to the content [20]. In section 4.6, we discussed an example of a visualization that lets users select attributes and filter people by their demographic characteristics (see Figure 17). This kind of chart allows users to focus on the type of person or the characteristics they care the most about. Visualization 78 in the corpus goes further by asking users to enter their personal information, and then customizes the visualization by showing populations of people who are like them. More elaborate techniques could be used to emphasize similarities between the user and a suffering population that may otherwise appear very remote and different. For example, if the user has kids and some of the suffering people also do, this characteristic could be emphasized. If on the other hand, the user is young but the people visualized are not, the visualization may hide age information.

"Smart" techniques like the above move us away from a *user customization* approach and closer to a more controversial *user targeting* approach. While current systems require explicit interaction (e.g., people need to enter information about themselves), it is in principle possible to use personal information silently (e.g., country of residence as inferred from the IP address) in order to optimize visualizations to maximize prosocial feelings or behavior. While this is clearly a dark design pattern [79], responsible and ethical targeting may be possible and desirable. For example, a study has recently explored whether the use of short documentaries whose content depends on the viewer's place of residence can make people more compassionate [80]. Besides humanitarian applications, in a time of intense polarization and tribalism, targeted anthropographics could help individuals become more tolerant by showing them the many characteristics they share with people of a different identity or political leaning.

6.1.3 Exploring Example-Driven Charts Further

When communicating about a particular societal concern (e.g., gun fatalities), visualizations typically show data about the entire concerned population (e.g., all people who fell victim to gunshots in the USA). However, we saw that visualizations can instead focus on conveying rich information about one or a few persons (e.g., Figure 2, Figure 10, or visualization 100 in the corpus), and this could contribute to making readers more compassionate [38].

Although example-driven charts are occasionally used, it appears that a lot remains to be explored. For example, we have not found a single clear example of a minimum-coverage visualization (i.e., focusing on a single person as a representative of a larger reference population). It could also be interesting to examine how single-person charts relate to personal visualizations, where the entire reference population consists of a single person [30], [31]. While most personal visualizations are not anthropographics the way we define them (i.e., the designer's goal is rarely to evoke compassion), they may share design strategies captured by the presented design space. Another interesting direction could be to combine an example-driven approach with

the customization/targeting strategy discussed previously. For example, a visualization that uses minimum coverage could choose a representative person who shares many characteristics in common with the reader.

6.1.4 Exploring Situatedness Further

Much remains to be explored in the area of situated visualizations, especially around the notion of intermediate situatedness. Again, one way for intermediate situatedness to occur is when the marks are presented at a location where the persons used to be in the past (section 4.5.3 and visualization 2 in the corpus). This practice has received little attention, yet it can be useful to design visualizations that act as memorials, e.g., by showing data about people who perished or experienced hardship at some specific location.

In visualization 105 in the corpus, 7,000 pairs of shoes were placed on the lawn of the U.S. Capitol to symbolize the children killed during the 2012 Newtown shooting. Although the installation is emotionally powerful, it is located 300 miles away from Newtown. Would have it carried even more emotional power had it been placed exactly where the shooting occurred? Many existing memorials are situated, although they rarely visualize rich data about the people. It is possible that doing so would help people relate, although practices exist in memorial design that paradoxically do the exact opposite. For example, the tomb of the unknown soldier at Arlington has been designed to convey as little information as possible, so that anyone could relate and entertain the possibility that the soldier is a lost relative [81].

The second way intermediate situatedness can occur is when the marks used to be in proximity to the persons they represent (e.g., Figure 8). This approach also opens up a vast area of untapped possibilities. For example, anthropographic visualizations could be explored that tap into the psychological power of memorabilia that were owned, touched, or worn by people [73]. For example, would the 2012 Newtown shooting installation discussed above carry more emotional weight had it used objects that personally belonged to the victims? Because all these examples shorten the causal and (sometimes) symbolic distance between the people the data is about and the data representation [82], [83], they have the potential to make readers feel more connected to the people.

Finally, many wearable visualizations such as data jewelry and data clothing items[11] convey data about the people who wear them and thus they are fully situated [46]. Perhaps there are ways in which such objects can be used to promote prosocial feelings and behavior. For example, assuming it is possible to invent a standard visual language for conveying personal histories, persons in need may be able to share their hardships through wearable visualizations that everyone could understand at a glance. Personal wearable visualizations do, however, pose ethical issues, including privacy issues. If such visualizations become common, even the choice not to wear them could convey personal information.

6.2 Opportunities for Research

None of the design dimensions described in this article has been thoroughly explored in the visualization literature. This section underlines opportunities for empirical research.

11. See dataphys.org/list/tag/data-jewellery and dataphys.org/list/tag/data-clothing for examples.

6.2.1 Testing Basic Design Dimensions

Boy et al. [1] conducted several experiments to explore the effectiveness of design choices to promote prosocial feelings. Their experiments mostly compared proportional wee-people charts — intermediate granularity and realism — with classical statistical charts — low granularity and realism. More recently, Liem et al. [7] empirically tested anthropographic designs with maximum granularity, high specificity and partial coverage, while Morais et al. [8] tested an anthropographic design with maximum granularity, maximum physicality and intermediate situatedness.

Though the results of all three studies are largely inconclusive, it does not follow that anthropographics are ineffective. First, the effects may be small and may need more statistical power to be detected [84]. Second, many other designs remain to be tested, some of which could prove more effective. For example, many of the designs used in Boy et al. [1] and Liem et al. [7] had partial authenticity, which could have affected how people trusted the information.

There is a clear need for more empirical studies that investigate the effect of basic design dimensions, namely, granularity, specificity, authenticity, and realism. The isolated effects of these dimensions as well as their interactions will need be tested. For example, a strong sense of individuality can only be conveyed if a visualization has both maximum granularity and high information specificity, so it will be interesting to test whether this combination is particularly beneficial. Also, it will be crucial to control for authenticity and study it separately, as its effects are poorly understood.

6.2.2 Understanding Coverage

Once the effectiveness of basic design dimensions starts to be understood, it will be useful to study more elaborate strategies such as coverage. As we mentioned before, we know from psychology that that as the number of identifiable victims in a narrative increases, prosocial feelings and behavior tend to decrease [38]. Until now, no study has been conducted to test whether this so-called "compassion fade" can also occur in the context of data visualization. Keeping the reference dataset constant and manipulating the coverage dimension can be an effective way of empirically exploring compassion fade in the context of data visualization.

6.2.3 Studying Situatedness and Physicality

Situatedness and physicality may also be relevant to anthropographics and deserve attention, even though empirically studying them will likely be less practical. For example, does an intermediate-situatedness visualization that shows victims in-place evoke more compassion than a non-situated visualization that shows the data elsewhere? Can physicality add to the emotional impact or memorability of a design strategy, and in turn promote compassion? Designs like "people-as-mark" (Figure 9) and wearable physicalizations would be particularly interesting to test since they are maximum in both the situatedness and physicality dimensions. Another interesting strategy to explore is the use of objects such as shoes or backpacks to represent people, as in visualizations 105 and 112 in the corpus.

6.2.4 Investigating Ambiguity

It could be interesting to explore the effect of ambiguity on prosocial feelings and behavior. For example, when a designer does not make explicit how many persons are represented by each mark — as opposed to Figure 6, where each mark explicitly represents 1 million soldiers — coverage and granularity become ambiguous. In Figure 9, for example, each mark can either be interpreted as standing for (A) a proportion or (B) a subset of the Parisian population. If one chooses A as interpretation, the visualization would have intermediate granularity and full coverage. Conversely, if the interpretation is based on B, the visualization would have maximum granularity and partial coverage. Can the same visualization have different effects on prosocial feelings or behavior depending on how it is interpreted?

6.2.5 Ethical Considerations

An important future direction will be to better understand the ethical issues behind different anthropographic design strategies, and the trade-offs involved. For example, is the misleading use of synthetic attributes justified if it increases charitable donations and helping behavior? Similar questions arise about the use of user targeting as discussed in section 6.1.2, or of ambiguity as discussed previously. A more nuanced understanding of the different ways synthetic attributes can mislead will also be helpful. For example, a partially authentic visualization like in Figure 12 could use demographic data to derive silhouettes (including sex and age distribution) and personal information (e.g., names) that are as plausible as possible. Such a design would be less misleading than a visualization where demographics and names misrepresent the target population, but more so than a visualization where each mark conveys information about a real person whose privacy needs also to be taken into account. As another example, briefly discussed in section 4.5.1, a visualization that adds decorative details to anthropomorphic marks is likely less misleading if those details are kept constant across all marks than if they vary. A deeper understanding of ethical considerations will need to rely on empirical studies, both to understand how people interpret possibly misleading visualizations, and what people consider ethically justifiable [85]. Other aspects of ethics beyond persuasion and deception [86] will also need be examined in the context of anthropographics.

6.2.6 Understanding the Importance of Data

Our design space does not capture properties that are inherent to the data being visualized, even though they may have a considerable impact on how compassionate readers can feel. Personal connection with the data, in particular, has a dramatic impact on how people experience visualizations [87]. A visualization designer may be able to manipulate personal connection to some extent if they have control over the visualized dataset — i.e., if they can choose which persons or which attributes to visualize from the known dataset (see section 6.1.2). A designer can even exercise some control over the known dataset, since they may be able to decide which information to collect about the reference dataset and how much effort to dedicate to the process. However, if we assume the reference dataset to be given, its characteristics (such as degree of connection with the reader) are out of the designer's control and thus do not belong to a design space of visualizations.

Nevertheless, there are important questions to consider beyond visualization design, such as what reference datasets are visualized and why, and what known datasets are made available and why. Certain populations or topics may be omitted or underrepresented in available datasets, either because data does not exist or because the is a lack of interest from analysts and visualization designers [88], [89]. Consequently, readers do not have access to the underrepresented data, which might generate a cycle where it becomes less and less important [89]. Therefore, some critical approaches such as the ones proposed by Dörk and colleagues [90] may also be relevant to the design of anthropographics and should be investigated in the future.

6.2.7 Generative Use of This Design Space

Finally, to assess the generative power of our design space and assess its generative power, it would be valuable to use it to conceive new kinds of visualizations. This could be done, for example, through a study where designers are asked to generate visualizations based on the design space.

6.3 Limitations

Our goal was to identify elementary dimensions in the vast design space of anthropographics and propose a conceptual framework and terminology that can help both researchers and practitioners reflect on and communicate about essential aspects of anthropographics design. Nevertheless, this design space is far from complete.

First, our conceptual framework currently only covers flat tables (see section 4.1). Although tables are likely the most common data model, other types of datasets exist such as networks [45]. Excluding such datasets excludes all social network visualizations, for example.

Second, we restricted our scope to datasets where each item is a person or a group (see section 3). There exist datasets that contain information that can profoundly affect people, despite not containing information about people. For example, datasets on how diseases spread geographically, or data about global warming. But since there is no one in these datasets with which to empathize or feel compassion about, they may be outside the realm of anthropographics.

We also excluded other datasets and visualizations from our investigation in order to keep the scope manageable, although they could be relevant to anthropographics. In particular, we excluded datasets involving simulated people (see section 3), but these may be useful to consider in the future. Also, although our focus was on people, many non-human animals experience suffering and may need compassion as well as help [91], [92]. Our conceptual framework can be easily generalized by considering that data items can refer to other sentient beings than humans.

In choosing the seven dimensions of our design space, we focused on elementary visualization characteristics that are easy to define, manipulate in studies, and apply as a designer. Many other factors can contribute into making a visualization effective in promoting prosocial feelings, although they are more difficult to operationalize. These including aesthetics,

the use of visual metaphors, and the use of accompanying stories (see, e.g., Figure 11). The space of possibilities is vast and our design space is only a starting point.

Acknowledgments

We thank our reviewers for their insightful suggestions. The first author gratefully acknowledges the financial support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), grant number 88881.188888/2018-01. This work was partly supported by the Agence Nationale de la Recherche (ANR), grant number ANR-19-CE33-0012.

References

[1] J. Boy, A. V. Pandey, J. Emerson, M. Satterthwaite, O. Nov, and E. Bertini, "Showing people behind data: Does anthropomorphizing visualizations elicit more empathy for human rights data?" in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2017, pp. 5462–5474.

[2] E. Bertini, "Can Visualization Elicit Empathy? Our Experiments with "Anthropographics"," https://web.archive.org/web/20191113054644/https://medium.com/@FILWD/can-visualization-elicit-empathy-our-experiments-with-anthropographics-7e13590be204, 2017, [Online; accessed 13-November-2019].

[3] G. Lupi, "Data humanism, the revolution will be visualized," https://web.archive.org/web/20190915123530/https://www.printmag.com/information-design/data-humanism-future-of-data-visualization/, 2017, [Online; accessed 15-September-2019.

[4] P. Bloom, *Against empathy: The case for rational compassion*. Random House, 2017.

[5] J. Harris, "Connecting with the dots," https://web.archive.org/web/20190706114742/https://source.opennews.org/articles/connecting-dots/, 2015, [Online; accessed 6-July-2019].

[6] L. C. Rost, "A data point moves into a bar," https://web.archive.org/web/20200318143617/https://lab.dsst.io/slides/33c3/7999.html, 2016, [Online; accessed 18-March-2020].

[7] J. Liem, C. Perin, and J. Wood, "Structure and empathy in visual data storytelling: Evaluating their influence on attitude," in *Computer Graphics Forum*, 2020.

[8] L. Morais, D. Sousa, and N. Andrade, "Evaluating a situated and physical anthropographic: An in-the-wild study," 2020, unpublished pre-print. [Online]. Available: http://rgdoi.net/10.13140/RG.2.2.10719.69283

[9] C. Kirk and D. Kois, "How many people have been killed by guns since newtown?" https://web.archive.org/web/20200305065626/http://www.slate.com/articles/news_and_politics/crime/2012/12/gun_death_tally_every_american_gun_death_since_newtown_sandy_hook_shooting.html, 2013, [Online; accessed 17-July-2019].

[10] M. Beaudouin-Lafon, "Designing interaction, not interfaces," in *Proceedings of the working conference on Advanced visual interfaces*. ACM, 2004, pp. 15–22.

[11] F. Fragapane and A. Piacentini, "Stories behind a line," https://web.archive.org/web/20190719064902/http://www.storiesbehindaline.com/, 2019, [Online; accessed 19-July-2019].

[12] S. Bateman, R. L. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks, "Useful junk?: the effects of visual embellishment on comprehension and memorability of charts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 2573–2582.

[13] S. Few and P. Edge, "The chartjunk debate," *Visual Business Intelligence Newsletter, no. June*, pp. 1–11, 2011.

[14] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen, "An empirical study on using visual embellishments in visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2759–2768, 2012.

[15] M. A. Borkin, A. A. Vo, Z. Bylinskii, P. Isola, S. Sunkavalli, A. Oliva, and H. Pfister, "What makes a visualization memorable?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2306–2315, 2013.

[16] 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*. ACM, 2015, pp. 1191–1200.

[17] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 1139–1148, 2010.

[18] R. Kosara and J. Mackinlay, "Storytelling: The next step for visualization," *Computer*, vol. 46, no. 5, pp. 44–50, 2013.

[19] C. D. Stolper, B. Lee, N. H. Riche, and J. Stasko, "Emerging and recurring data-driven storytelling techniques: Analysis of a curated collection of recent stories," *Microsoft Research, Washington, USA*, 2016.

[20] B. Bach, M. Stefaner, J. Boy, S. Drucker, L. Bartram, J. Wood, P. Ciuccarelli, Y. Engelhardt, U. Koeppen, and B. Tversky, "arrative design patterns for data-driven storytelling," in *Data-Driven Storytelling*. AK Peters/CRC Press, 2018, pp. 125–152.

[21] J. Hullman and N. Diakopoulos, "Visualization rhetoric: Framing effects in narrative visualization," *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2231–2240, 2011.

[22] F. Amini, N. Henry Riche, B. Lee, C. Hurter, and P. Irani, "Understanding data videos: Looking at narrative visualization through the cinematography lens," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 1459–1468.

[23] J. Hullman, S. Drucker, N. H. Riche, B. Lee, D. Fisher, and E. Adar, "A deeper understanding of sequence in narrative visualization," *IEEE Transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2406–2415, 2013.

[24] B. Lee, N. Henry Riche, P. Isenberg, and S. Carpendale, "More than Telling a Story: A Closer Look at the Process of Transforming Data into Visually Shared Stories," *IEEE Computer Graphics and Applications*, vol. 35, no. 5, pp. 84–90, Sep. 2015. [Online]. Available: https://hal.inria.fr/hal-01158445

[25] B. Bach, N. Kerracher, K. W. Hall, S. Carpendale, J. Kennedy, and N. Henry Riche, "Telling stories about dynamic networks with graph comics," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 3670–3682.

[26] J. Boy, F. Detienne, and J.-D. Fekete, "Storytelling in information visualizations: Does it engage users to explore data?" in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 1449–1458.

[27] A. Satyanarayan and J. Heer, "Authoring narrative visualizations with ellipsis," in *Computer Graphics Forum*, vol. 33, no. 3. Wiley Online Library, 2014, pp. 361–370.

[28] S. Carpendale, A. Thudt, C. Perin, and W. Willett, "Subjectivity in personal storytelling with visualization," *Information Design Journal*, vol. 23, no. 1, pp. 48–64, 2017.

[29] D. Huang, M. Tory, B. A. 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*, vol. 21, no. 3, pp. 420–433, 2014.

[30] C. Perin, "The symmetry of my life: An autobiographical visualization," in *IEEE VIS 2017 Electronic Conference Proceedings [Poster]*, 2017.

[31] ——, "The symmetry of my life ii," in *IEEE VIS 2018 Electronic Conference Proceedings [Poster]*, 2018.

[32] D. DeSteno, "Compassion and altruism: How our minds determine who is worthy of help," *Current opinion in behavioral Sciences*, vol. 3, pp. 80–83, 2015.

[33] P. Bloom, "Empathy and its discontents," *Trends in cognitive sciences*, vol. 21, no. 1, pp. 24–31, 2017.

[34] M. R. Jordan, D. Amir, and P. Bloom, "Are empathy and concern psychologically distinct?" *Emotion*, vol. 16, no. 8, p. 1107, 2016.

[35] J. L. Goetz, D. Keltner, and E. Simon-Thomas, "Compassion: an evolutionary analysis and empirical review." *Psychological bulletin*, vol. 136, no. 3, p. 351, 2010.

[36] M. O. Wilhelm and R. Bekkers, "Helping behavior, dispositional empathic concern, and the principle of care," *Social Psychology Quarterly*, vol. 73, no. 1, pp. 11–32, 2010.

[37] A. Genevsky, D. Västfjäll, P. Slovic, and B. Knutson, "Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving," *Journal of Neuroscience*, vol. 33, no. 43, pp. 17 188–17 196, 2013.

[38] D. Västfjäll, P. Slovic, M. Mayorga, and E. Peters, "Compassion fade: Affect and charity are greatest for a single child in need," *PloS one*, vol. 9, no. 6, p. e100115, 2014.

[39] J. Calvert, R. Abadla, and S. M. Tauseef, "Design and testing of a virtual reality enabled experience that enhances engagement and simulates empathy for historical events and characters," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2019, pp. 868–869.

[40] M. Kandaurova and S. H. M. Lee, "The effects of virtual reality (vr) on charitable giving: The role of empathy, guilt, responsibility, and social exclusion," *Journal of Business Research*, vol. 100, pp. 571–580, 2019.

[41] A. Ivanov, K. T. Danyluk, and W. Willett, "Exploration & anthropomorphism in immersive unit visualizations," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. LBW008.

[42] L. D. Riek, T.-C. Rabinowitch, B. Chakrabarti, and P. Robinson, "How anthropomorphism affects empathy toward robots," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 2009, pp. 245–246.

[43] H. Kennedy and R. L. Hill, "The feeling of numbers: Emotions in everyday engagements with data and their visualisation," *Sociology*, vol. 52, no. 4, pp. 830–848, 2018.

[44] A. Erlandsson, S. M. Hohle, E. Løhre, and D. Västfjäll, "The rise and fall of scary numbers: The effect of perceived trends on future estimates, severity ratings, and help-allocations in a cancer context," *Journal of Applied Social Psychology*, vol. 48, no. 11, pp. 618–633, 2018.

[45] T. Munzner, *Visualization analysis and design*. AK Peters/CRC Press, 2014.

[46] W. Willett, Y. Jansen, and P. Dragicevic, "Embedded data representations," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 461–470, 2017.

[47] S. R. Dos Santos, "A framework for the visualization of multidimensional and multivariate data," Ph.D. dissertation, University of Leeds, 2004.

[48] A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer, "Vega-lite: A grammar of interactive graphics," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 1, pp. 341–350, 2017.

[49] J. Bertin, *Sémiologie graphique: Les diagrammes-Les réseaux-Les cartes*. Paris: Gauthier-VillarsMouton & Cie, 1973.

[50] H. Senay and E. Ignatius, *Rules and principles of scientific data visualization*. Institute for Information Science and Technology, Department of Electrical Engineering, 1990.

[51] R. E. Roth, "Visual variables," *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, pp. 1–11, 2016.

[52] Y. Jansen, P. Dragicevic, P. Isenberg, J. Alexander, A. Karnik, J. Kildal, S. Subramanian, and K. Hornbæk, "Opportunities and Challenges for Data Physicalization," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. New York, NY, United States: ACM, Apr. 2015. [Online]. Available: https://hal.inria.fr/hal-01120152

[53] O. Neurath, *International picture language: The first rules of isotype*. Basic English Publ., 1936.

[54] M. Neurath, "Isotype," *Instructional science*, vol. 3, no. 2, pp. 127–150, 1974.

[55] S. Drucker and R. Fernandez, "A unifying framework for animated and interactive unit visualizations," Microsoft Research, Tech. Rep. MSR-TR-2015-65, August 2015. [Online]. Available: http://web.archive.org/web/20200318142741/https://www.microsoft.com/en-us/research/publication/a-unifying-framework-for-animated-and-interactive-unit-visualizations/

[56] T. Bronshtein, "200 years of immigration to the us," https://web.archive.org/web/20200112065959/http://insightfulinteraction.com/immigration200years.html, 2017, [Online; accessed 12-January-2020].

[57] L. A. Worrell, "The great war," https://web.archive.org/web/20200214140655/http://www.laworrell.com/blog/2015/1/14/isotype-international-system-of-typographic-picture-education, 2019, [Online; accessed 14-February-2020].

[58] Gun Violence Archive, "Gun violence chart of deaths in 2019," https://web.archive.org/web/20190714172038/https://www.gunviolencearchive.org/, 2019, [Online; accessed 14-July-2019].

[59] T. Kogut and I. Ritov, "The "identified victim" effect: An identified group, or just a single individual?" *Journal of Behavioral Decision Making*, vol. 18, no. 3, pp. 157–167, 2005.

[60] S. Lee and T. H. Feeley, "The identifiable victim effect: A meta-analytic review," *Social Influence*, vol. 11, no. 3, pp. 199–215, 2016.

[61] G. Lupi, "Collaborative data wallpaper for story," https://web.archive.org/web/20200214141001/http://giorgialupi.com/collaborative-data-wallpaper-for-story/, 2019, [Online; accessed 17-July-2019].

[62] H. Haug, S. Kaegi, and D. Wetzel, "100% Montréal," https://web.archive.org/web/20200214141952/https://www.rimini-protokoll.de/website/en/project/100-montreal, 2015, [Online; accessed 14-February-2020].

[63] L. Morais and N. Andrade, "Small data e o câncer de mama: a jornada da paciente," https://web.archive.org/web/20200214140327/http://www.oncoguia.org.br/conteudo/small-data-e-o-cancer-de-mama-a-jornada-da-paciente/12899/1195/, 2019, [Online; accessed 14-February-2020].

[64] G. Lupi, "Bruises: The data we don't see," https://web.archive.org/web/20191219033917/http://giorgialupi.com/bruises-the-data-we-dont-see, 2018, [Online; accessed 19-December-2019].

[65] G. Lupi, M. Popova, and W. MacNaughton, "Famous writers' sleep habits and literary productivity," https://web.archive.org/web/20200219160913/http://giorgialupi.com/famous-writers-sleep-habit-and-productivity, 2014, [Online; accessed 19-February-2020].

[66] T. Rawles, "Picture from 1993 reminds people of the loss of life due to aids," https://web.archive.org/web/20190331193218/https://sdgln.com/causes/2017/11/28/picture-1993-reminds-people-loss-life-due-aids, 2017, [Online; accessed 31-March-2019].

[67] S. Campbell, "Feeling numbers the rhetoric of pathos in visualization," Master's thesis, Northeastern University, 2018.

[68] M. Stefaner, D. Hemment, and S. Nand, "Emoto: Visualizing the online response to london 2012," http://web.archive.org/web/20200323155639/http://archive.stefaner.eu/projects/emoto/, 2012, [Online; accessed 23-March-2020.

[69] M. Le Goc, L. H. Kim, A. Parsaei, J.-D. Fekete, P. Dragicevic, and S. Follmer, "Zooids: Building blocks for swarm user interfaces," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 2016, pp. 97–109.

[70] L. Morais, N. Andrade, D. Sousa, and L. Ponciano, "Defamiliarization, representation granularity, and user experience: a qualitative study with two situated visualizations," *PacificVis*, 2019.

[71] B. H. Thomas, G. F. Welch, P. Dragicevic, N. Elmqvist, P. Irani, Y. Jansen, D. Schmalstieg, A. Tabard, N. A. ElSayed, R. T. Smith *et al.*, "Situated analytics," in *Immersive Analytics*. Springer, 2018, pp. 185–220.

[72] Periscopic, "U.s. gun deaths in 2010 and 2013," http://web.archive.org/web/20191204144534/https://guns.periscopic.com/?year=2013, 2013, [Online; accessed 04-December-2019].

[73] G. E. Newman and P. Bloom, "Physical contact influences how much people pay at celebrity auctions," *Proceedings of the National Academy of Sciences*, vol. 111, no. 10, pp. 3705–3708, 2014.

[74] B. Bach, P. Dragicevic, D. Archambault, C. Hurter, and S. Carpendale, "A review of temporal data visualizations based on space-time cube operations," in *Eurographics conference on visualization*, 2014.

[75] Microsoft, "Sanddance," https://web.archive.org/web/20200312151709/https://sanddance.js.org/, 2020, [Online; accessed 12-March-2020].

[76] Sasaki, "Understanding homelessness," http://web.archive.org/web/20191217013030/http://www.understandhomelessness.com/explore/, 2017, [Online; accessed 17-December-2019].

[77] B. Bach, P. Dragicevic, S. Huron, P. Isenberg, Y. Jansen, C. Perin, A. Spritzer, R. Vuillemot, W. Willett, and T. Isenberg, "Illustrative data graphics in 18th-19th century style: A case study," in *IEEE Conference on Visualization-IEEE VIS 2013*, 2013.

[78] W. MacAskill, *Doing good better: Effective altruism and a radical new way to make a difference*. Guardian Faber Publishing, 2015.

[79] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, "The dark (patterns) side of ux design," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 534.

[80] S. Concannon, N. Rajan, P. Shah, D. Smith, M. Ursu, and J. Hook, "Brooke leave home: Designing a personalized film to support public engagement with open data," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–14.

[81] J. Rosenberg, "The known unknown," https://web.archive.org/web/20191019133523/https://99percentinvisible.org/episode/the-known-unknown/, 2019, episode 344 of The 99% Invisible podcast. [Online; accessed 19-October-2019].

[82] D. Offenhuber, "Data by proxy—material traces as autographic visualizations," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 1, pp. 98–108, 2019.

[83] D. Offenhuber and O. Telhan, "Indexical visualization—the dataless information display," *Ubiquitous computing, complexity and culture*, vol. 288, 2015.
[84] J. Cohen, "The earth is round (p¡. 05)," in *What if there were no significance tests?* Routledge, 2016, pp. 69–82.
[85] J. Graham, B. A. Nosek, J. Haidt, R. Iyer, S. Koleva, and P. H. Ditto, "Mapping the moral domain," *Journal of personality and social psychology*, vol. 101, no. 2, p. 366, 2011.
[86] M. Correll, "Ethical dimensions of visualization research," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–13.
[87] E. M. Peck, S. E. Ayuso, and O. El-Etr, "Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ser. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019. [Online]. Available: https://doi.org/10.1145/3290605.3300474
[88] M. Onuoha, "On missing data sets," https://github.com/MimiOnuoha/missing-datasets, 2015, [Online; accessed 23-March-2020.
[89] C. D'Ignazio and L. F. Klein, *Data feminism*. MIT Press, 2020.
[90] M. Dörk, P. Feng, C. Collins, and S. Carpendale, "Critical infovis: exploring the politics of visualization," in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 2013, pp. 2189–2198.
[91] T. Regan and P. Singer, *Animal rights and human obligations*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[92] B. Tomasik, "The importance of wild-animal suffering," *Rel.: Beyond Anthropocentrism*, vol. 3, p. 133, 2015.

Luiz Morais is a post-doctoral fellow at Inria, studying situated and personal visualizations. He received his doctoral degree in Computer Science from the Universidade Federal de Campina Grande in 2020. His research interests include anthropographics, situated data visualization, and data physicalization.

Yvonne Jansen is a permanent Research Scientist at the Centre National de Recherche Scientifique (CNRS) since 2016. She received her doctoral degree from Université Paris Sud in 2014. She previously held a post-doctoral position at the University of Copenhagen in Denmark. Her research interests include data physicalization, situated and embedded data visualization, and methods and techniques to facilitate the transparent reporting of research outcomes.

Nazareno Andrade is a professor at the Systems and Computing Department of Universidade Federal de Campina Grande, Brazil, since 2009. Nazareno received a PhD in Electrical Engineering from the same University, and was a post-doctoral fellow at the Delft University of Technology. His research interests include anthropographics, civic technology, social computing, and music information retrieval.

Pierre Dragicevic is a permanent Research Scientist at Inria (France) since 2007, studying information visualization and human-computer interaction. He received his PhD from the Université de Nantes in 2004 and was a post-doctoral fellow at the University of Toronto from 2006—2007. His research interests include data physicalization, decision making with data visualizations, user study methodology, transparent statistical communication, design spaces and conceptual frameworks.