HAL
open science

Vasco: Interactive Tool for Early-Exploration of Data

Miriam Luque Chipana, Himanshu Verma, Florian Evéquoz, Pierre Vanhulst, Denis Lalanne

To cite this version:

Miriam Luque Chipana, Himanshu Verma, Florian Evéquoz, Pierre Vanhulst, Denis Lalanne. Vasco: Interactive Tool for Early-Exploration of Data. 31e conférence francophone sur l'Interaction Homme-Machine (IHM 2019), Dec 2019, Grenoble, France. pp.4:1-7, 10.1145/3366551.3370343 . hal-02388850

HAL Id: hal-02388850

https://hal.science/hal-02388850v1

Submitted on 2 Dec 2019

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.

Vasco: Interactive Tool for Early-Exploration of Data

Vasco : un outil interactif pour l'exploration précoce des données

Miriam Luque-Chipana
Univ. of Fribourg
Fribourg, Switzerland
mmluquec@gmail.com

Pierre Vanhulst
Univ. of Fribourg
Fribourg, Switzerland
pierre.vanhulst@unifr.ch

Himanshu Verma
HES-SO Valais-Wallis
Sierre, Switzerland
himanshu.verma@acm.org

Denis Lalanne
Univ. of Fribourg
Fribourg, Switzerland
denis.lalanne@unifr.ch

Florian Evéquoz
HES-SO Valais-Wallis
Sierre, Switzerland
florian.evequoz@hevs.ch

ABSTRACT

We describe Vasco, a data visualization tool for inexperienced users. Vasco is designed to allow and promote early exploration of data, targeting users without experience in the design of visualizations and data analysis. Vasco structures the interface to select easily data and create visualizations, using panels and cards. Vasco automatically generates the best graphics according to the selection of variables and data morphology. In addition, Vasco helps give the user control and organization with

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

IHM '19 Adjunct, December 10–13, 2019, Grenoble, France

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-7027-1/19/12.
https://doi.org/10.1145/3366551.3370343

multiple workspaces. Finally, a controlled study, that compares the usability of Vasco with Voyager 2, shows that users find helpful the presence of the dimensions data, plus the iterative nature of the exploration, help users understand the data visualization process, make them feel more confident and more performant.

CCS CONCEPTS

• **Human-centered computing** → **Visualization tools**; **Early exploration**.

KEYWORDS

Visualization tools; data visualization; early exploration.

RÉSUMÉ

Nous décrivons Vasco, un outil de visualisation de données pour les utilisateurs inexpérimentés. Vasco est conçu pour permettre et promouvoir l'exploration précoce des données, en ciblant les utilisateurs sans expérience dans la conception de visualisations et l'analyse des données. Vasco structure l'interface pour sélectionner facilement les données et créer des visualisations à l'aide de panneaux et de cartes. Vasco génère automatiquement des graphiques adéquats en fonction des variables sélectionnées et de la morphologie des données. De plus, Vasco permet aux utilisateurs de contrôler et d'organiser leur processus de conception grâce à des espaces de travail multiples. Enfin, une étude contrôlée, qui compare l'utilisabilité de Vasco avec celle de Voyager 2, montre que la présentation constante de la taxonomie des données et la nature itérative de l'exploration aident les utilisateurs à comprendre le processus de visualisation des données, à se sentir plus confiants et plus performants.

MOTS CLÉS

Outils de visualisation ; visualisation de données ; exploration précoce.

INTRODUCTION

Data visualizations are direct representations of the data. They allow the comprehension of the data, the generation of hypotheses and the representation of an idea to communicate[2]. Most tools that offer assistance with data visualization usually require users to have training and prior knowledge of the data. To choose relevant visual representations, inexperienced users also need to understand their data and their attributes, a task that requires early exploration of the data. We believe these users would benefit from an automatic visualization presentation that is generated based on data selections grouped by attributes.

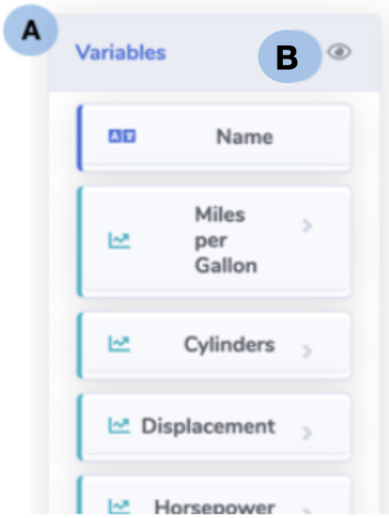

Figure 1: Vasco Interface, Variable Panel(A), Visualization dimension information button(B)

Vasco attempts to help these inexperienced users during the data comprehension process, through exploration based on visualization. Vasco uses univariate analysis to describe the data and help the user to find patterns that exist within the variable. To this end, Information Cards display sparklines and the data characteristics (mean, mode, average) to help the user understand the data; to continue with the exploration process, Vasco allows the user to select the variables (get from the data) and generates a list of visualizations from the combination of the selected variable and the remaining variables. For the user to be able to focus on specific questions, Vasco allows the creation of specific visualizations using the selection of variables grouped in a card without the specification of the visualization channels. The interface of Vasco relies on dedicated color schemes and icons to facilitate the identification of the type of each variable in a dataset. By improving their understanding of the data, users of Vasco are also better informed about the relevance of the associations between each dimension of their data. At the moment, Vasco is not focused on the customization of visualizations and provides no mean to change the specifications.

STATE OF THE ART

The tools for generating visualizations differ in the approach and process that the user must follow to create visualizations. Data Illustrator[3] focuses on the design of the visualization, the user composes the visualization and configures its parameters to be able to add data. RawGraphs[5] has a predefined list of visualizations, the users choose the visualization and then add the data, the data is restricted by the chosen visualization. In both cases, creating and choosing the most suitable visualization requires to know about the visualization design and the data. Tools such as Datawrapper[1], Data Voyager[9], Voyager 2[10] and Tableau[4, 7] require the specification of dimensions and display marks. Data Wrapper generates a visualization and then allows you to change the visualization type, dimensions and marks. Data Voyager, Voyager 2 and Tableau use a ranking system to decide which is the best visualization based on the selected data dimensions. Once the visualization is generated it is possible to change the visualization type. Data Voyager and Voyager 2 have a mixed-initiative, that use a recommendation system to generate visualizations related to the visualization created by the user.

VASCO PLATFORM

Vasco is built on NodeJs and uses CompassQl[9], Vega, Vega-Lite[6] to generate visualizations. Vasco uses CompassQl to choose the most relevant visualization. Vasco is divided into three components (1) data upload, (2) analyzer and (3) report. The first component, Data Upload allows the user to upload data in CSV format files. The second component, the Analyzer is in charge of generating the visualizations and creating a list of the dimensions together with relevant attributes(average, media, moda, sparklines) for each dimension. The analyzer uses CompassQl to generate a list of Vega-Lite

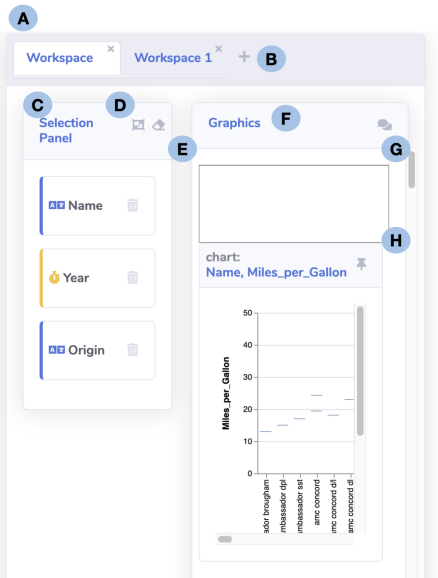

Figure 2: Vasco Interface, Workspace(A), Add workspace button(B), Selection Panel(C), Creation Card Button(D), Clear the selection panel button(E), Graphic panel(F), Comments button(G), Comments box(H) .

Figure 3: Data types icons and colors

specifications, and then render them using Vega. The last component is the Report, it allows the grouping and export of the visualizations.

Vasco Mode of Use

Vasco is focused on exploring data. To this end, the platform has two modes of use: (1) exploratory mode and (2) creation mode. The exploratory mode generates a list of views for each dimension, these views are formed from the combination of the selected variable with the remaining dimensions. The selection of multiple dimensions in exploratory mode avoids the redundancy of the visualizations. The second mode allows you to create specific views by grouping dimensions. It is not necessary to specify the axes and display marks.

Vasco Interface

The Vasco interface is divided into 4 sections: (1) Load the data, (2) Variable Panel(Figure 1 A), (3) Workspace(Figure 2 A), (4) Report. Vasco is composed of cards that allow interaction, display information and visualizations. The first section Load the Data, allows the user to upload the CSV files selected by the users. The second section, the Variables Panel shows the list of dimensions found in the data. Information Cards are used to present the dimension, the cards change according to the type of data of the variable. The types of data(Figure 3) are classified in quantitative, nominal, temporal. Vasco assigns an icon and a different color for each type of data. The information cards(Figure 4) are divided into two parts: (1) the header and (2) the body. The header contains the data type icon and the name of the dimension; the body of the card shows additional information such as the mean, fashion, average and sparkline that is created as a summary of the dimension. The third section Workspace is divided into two subsections: (1) Selection Panel(Figure 2 C) and (2) Graphics Panel(Figure 2 F). The user can create multiple Workspaces, creating also new selection and graphics panels. The user must drag the cards from the variable panel to the selection panel to create visualizations. When Vasco is in exploration mode, the list of graphics generated for the variable will be displayed in the graphics panel.

To create a specific visualization, Vasco uses creation cards that allow to group dimensions. The creation cards(Figure 5) are divided into two parts: (1) the header and (2) the body. The first part is intended to give feedback to the user on the visualization to be created. The icons of the data types are shown with their respective colors, when the user adds a dimension inside this card the color of the icons change to lead if it is not possible to continue adding a dimension of the same type; in this way Vasco seeks to help the users to create a visualization without having to worry about the configuration of the axes and marks of the visualization. The graphics panel shows the created visualizations, uses graphics cards to show the visualizations. Graphics cards(Figure 6) are divided into two parts:(1) the header and (2) the display. The header of the card contains the name of the

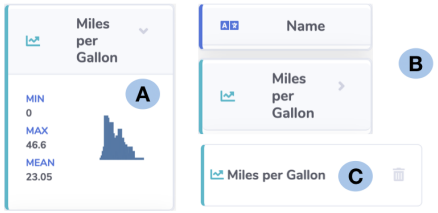

Figure 4: Information card with additional information displayed(A), Information card without additional information(B), Information card when is in the selection panel(C)

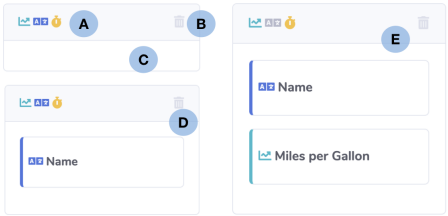

Figure 5: Empty creation card(A), Delete creation card button(B), Empty body of the creation card(C), Creation card feedback without restriction of variables (D), Creation card feedback restriction of variables (E)

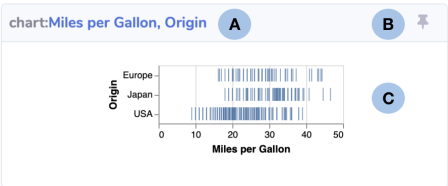

Figure 6: Name of the variables involved in the generation of the graph (A), Add to the report button(B), Visualization(C)

dimensions that have been used when generating the display; additionally it has a button to add the display to the report. The last part of the interface is the Report which allows the grouping of important visualizations.

EVALUATION AND RESULTS

We conducted a user study to compare Vasco with Voyager 2, with a specific focus on early exploration. Through this comparison, we intended to study the generation of fully manual visualizations through semi-automatic process.

Study Protocol

We recruited twelve participants without experience in data analysis and visualization design. The users had no prior experience with Voyager 2 and the datasets. We used the same databases that the authors of Voyager 2 used in their study. The study session was designed to compare Vasco and Voyager 2. We used two different databases, but they shared the same data types. The order of use of the tools was alternated among the participants. The study started with 5 minutes of training on the tool, then the users were asked to: "explore the dataset, get an idea of what it was about and what information they think is relevant"; they were free to use the tool during 20 minutes. The process was repeated for each tool. At the end of the session, an interview was performed and the user was asked to complete a post-study survey.

Results

We present the results obtained in the study, which focuses on early exploration by covering dimensions as part of the exploration process, the use of creation cards to generate specific visualizations and the feedback obtained in interviews and surveys. We use linear mixed-models that allow us to estimate the effect of the visualization tool taking into account the variance of data set and individual performance. We made the following assumptions regarding Vasco:

(1) Information cards increase the understanding and coverage of variables. To evaluate the coverage of the dimensions we include the visualization tool as a fixed effect and the interaction with variables as the dependent variable, under the relation of the interaction with the variables as a function of the Vasco or Voyager2 tool. It was found that this model was significant $t(9)=2.61$, $p=.03$. Vasco affected the interaction with the variables by increasing the coverage of the variables in 1.7 variables more than Voyager 2 with a standard error of 0.9. The information sheets in Vasco try to summarize the information of the dimensions, in the same way, Voyager presents the summarized visualizations. While Voyager 2 starts with a list of summarized views, Vasco starts with an empty workspace. Some of the users stated that the initial visualization

of Voyager 2 made them feel overwhelmed and confused because they didn't understand the relationship with dimensions. This limited the user's interaction with the variables

(2) We expected that Voyager 2 would be preferred to answer specific questions, but that the margin of preference to Voyager 2 would not be prejudicial to Vasco. Voyager 2 offers control over the visual encodings, unlike Vasco which automatically generates the visualization and indicates what type of variables can be added. To measure this, we include the visualization tool as a fixed effect and the creation of visualizations as the dependent variable, under the relation of the creation of visualizations in function of the Vasco or Voyager2 tool. It was found that this model resulted in $t(9)=-1.7$, $p=.01$. This result argues that there is a slight preference for Voyager 2 when it comes to specification control, but the difference with Vasco is thin. The users said they would like Vasco to offer more control over the channels and customization of the tools. Voyager 2 was preferred by users, but they also said they would use Vasco to create simple visualizations. They require a mid-point that allows them to customize the visualizations after they are automatically created as Vasco does.

(3) The use of workspaces facilitates and improves the work of the user. Regarding the third hypothesis, we asked the users about their experience with workspaces. 9 of the 10 participants agreed that it helped them with the organization, and made their work more efficient. The average number of workspaces created during the sessions was 3, the highest number of workspaces used was 5 and the lowest was 2.

We finally submitted a survey to the users, that consisted of 15 questions and statements related to both platforms and rated from 0 (strongly disagree) to 5 (strongly agree). We found that 8 of the 10 users considered Vasco easy to use, while 3 of the 10 users considered Voyager 2 the same. We found a relationship between the difficulty of use and the complexity of the tools. For data exploration, 8 of the 10 prefer Vasco, while 5 of the 10 users said they would also work on Voyager 2. For the specific creation of graphics, 6 of the 10 would use Vasco, while 7 of the 10 would also use Voyager 2.

CONCLUSION AND FUTURE WORK

We introduced Vasco, a system to support early exploration of data through visualization for inexperienced users. We found that the visual presentation of information related to the attributes of the data improves their comprehension. The use of multiple workspaces helps the users organize their work and consult their previous explorations without having to recreate the visualizations. A common issue raised by our participants was the lack of control of the specifications. Further work should extend Vasco with this feature, with the aim of serving both inexperienced and expert users. Another trail to follow is the addition of annotations [8], leveraging Vasco to allow the sharing of discoveries made by analysts during their exploration of the data.

REFERENCES

[1] Datawrapper. 2012. Enrich your stories with charts, maps and tables. https://www.datawrapper.de

[2] Danyel Fisher. 2018. *Making data visual : a practical guide to using visualization for insight.* O'Reilly Media, Sebastopol, CA.

[3] Zhicheng Liu, John Thompson, Alan Wilson, Mira Dontcheva, James Delorey, Sam Grigg, Bernard Kerr, and John Stasko. 2018. Data Illustrator: Augmenting Vector Design Tools with Lazy Data Binding for Expressive Visualization Authoring. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 123, 13 pages. https://doi.org/10.1145/3173574.3173697

[4] Jock Mackinlay, Pat Hanrahan, and Chris Stolte. 2007. Show Me: Automatic Presentation for Visual Analysis. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (Nov. 2007), 1137–1144. https://doi.org/10.1109/TVCG.2007.70594

[5] Michele Mauri, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi. 2017. RAWGraphs: A Visualisation Platform to Create Open Outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter (CHItaly '17)*. ACM, New York, NY, USA, Article 28, 5 pages. https://doi.org/10.1145/3125571.3125585

[6] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2017. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)* (2017). http://idl.cs.washington.edu/papers/vega-lite

[7] Tableau. 2003. Changing the way you think about data.

[8] Pierre Vanhulst, Florian Evequoz, Raphael Tuor, and Denis Lalanne. 2018. A Descriptive Attribute-Based Framework for Annotations in Data Visualization. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics*. Springer, 143–166.

[9] K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer. 2016. Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan 2016), 649–658. https://doi.org/10.1109/TVCG.2015.2467191

[10] Kanit Wongsuphasawat, Zening Qu, Dominik Moritz, Riley Chang, Felix Ouk, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2017. Voyager 2: Augmenting Visual Analysis with Partial View Specifications. In *ACM Human Factors in Computing Systems (CHI)*. http://idl.cs.washington.edu/papers/voyager2