



Raul de Araújo Lima

**A Question-Oriented Visualization
Recommendation System for Data Exploration**

Dissertação de Mestrado

Dissertation presented to the Programa de Pós-graduação em
Informática da PUC-Rio in partial fulfillment of the requirements
for the degree of Mestre em Informática.

Advisor: Prof^a. Simone Diniz Junqueira Barbosa

Rio de Janeiro
February 2020



Raul de Araújo Lima

A Question-Oriented Visualization Recommendation System for Data Exploration

Dissertation presented to the Programa de Pós-graduação em Informática da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Informática. Approved by the Examination Committee.

Prof^a. Simone Diniz Junqueira Barbosa

Advisor

Departamento de Informática – PUC-Rio

Prof. Alberto Barbosa Raposo

Departamento de Informática – PUC-Rio

Prof. Hélio Cortês Vieira Lopes

Departamento de Informática – PUC-Rio

Rio de Janeiro, February 20th, 2020

All rights reserved.

Raul de Araújo Lima

Bachelor's in Computer Science (2018) at the Federal University of Ceará (UFC). Worked for the BTG Pactual lab from March 2018 to June 2018, and for the Software Engineering Laboratory (LES) from October 2018 to March 2019 at PUC-Rio.

Bibliographic data

Lima, Raul de Araújo

A Question-Oriented Visualization Recommendation System for Data Exploration / Raul de Araújo Lima; advisor: Simone Diniz Junqueira Barbosa. – Rio de Janeiro: PUC-Rio, Departamento de Informática, 2020.

v., 125 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui bibliografia

1. Visualização de Informação. 2. Recomendação de Visualizações. 3. Ferramenta de Visualização. I. Barbosa, Simone Diniz Junqueira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

To my family.

Acknowledgments

I thank God for the life and the redemption He has given to us through Jesus Christ, in the power of the Holy Spirit.

I thank all my family through my parents Adailton and Jackeline, for all the support, encouragement, and all the education they have given me.

I thank my colleagues and friends, “battle companions”, for the company that brought me peace and health. I am particularly grateful to those who were closest: Micaele, Rômulo, André, Alysson, João Vitor, Sérgio, Dieinison, Rodrigo, Pedro Henrique, André Brandão, Anderson, Dalai, Luisa, Lauro.

I thank my advisor, professor Simone Diniz Junqueira Barbosa, for all the learning, for the patience, for the dedication, for the supernatural brownies, and everything else. I also thank professor Hélio Cortês Vieira Lopes, for all the proximity and support.

I thank the priests through whom I received the real presence of Jesus Christ in the Holy Eucharist, as well as His Forgiveness and Mercy in the Sacrament of Reconciliation during the time I was in the city of Rio de Janeiro. I am particularly grateful to the priests Thiago Azevedo, Wallace Prado and Rafael Martins of the Holy Angels Parish (Leblon); priests Mauro and Lincoln of the Our Lady of the Immaculate Conception Parish (Gávea); and to priest Alexandre Pacioli, rector of the Sacred Heart of Jesus Church at PUC-Rio. I also thank all the brothers and sisters I have met and with whom I have the grace to share the joy of faith.

I thank all the professionals in the Informatics Department at PUC-Rio.

I thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for partially financing this research under grant 130814/2018-0.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Abstract

Lima, Raul de Araújo; Barbosa, Simone Diniz Junqueira (Advisor). **A Question-Oriented Visualization Recommendation System for Data Exploration.** Rio de Janeiro, 2020. 125p. Dissertação de mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The increasingly rapid growth of data production and the consequent need to explore them to obtain answers to a wide range of questions have promoted the development of tools to facilitate the manipulation and construction of data visualizations. These tools should allow users to effectively explore data, communicate information accurately, and enable more significant knowledge gain through data. However, building useful data visualizations is not a trivial task: it may involve a large number of decisions that often require experience from their designer. To facilitate the process of exploring datasets through the construction of visualizations, we developed VisMaker, a software tool which uses a set of rules to determine appropriate visualizations for a certain selection of variables. In addition to allowing the user to define visualizations by mapping variables onto visualization channels, VisMaker presents visualization recommendations organized through questions constructed based on the variables selected by the user, trying to facilitate the understanding of the visualization recommendations and assisting the exploratory process. To evaluate VisMaker, we carried out two studies comparing it with another tool that exists in the literature, one aimed at solving questions and the other at data exploration. We analyzed some aspects of the use of the tools. We collected feedback from the participants, through which we were able to identify the advantages and disadvantages of the recommendation approach we proposed, raising possible improvements for this type of tool.

Keywords

Information Visualization; Visualization Recommendation; Visualization Tool.

Resumo

Lima, Raul de Araújo; Barbosa, Simone Diniz Junqueira. **Um Sistema de Recomendação de Visualizações Orientado a Perguntas para Exploração de Dados**. Rio de Janeiro, 2020. 125p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O crescimento cada vez mais acelerado da produção de dados e a decorrente necessidade de explorá-los a fim de se obter respostas para as mais variadas perguntas têm promovido o desenvolvimento de ferramentas que visam a facilitar a manipulação e a construção de gráficos. Essas visualizações devem permitir explorar os dados de maneira efetiva, comunicando as informações com precisão e possibilitando um maior ganho de conhecimento. No entanto, construir boas visualizações de dados não é uma tarefa trivial, uma vez que pode requerer um grande número de decisões que, em muitos casos, exigem certa experiência por parte de seu projetista. Visando a facilitar o processo de exploração de conjuntos de dados através da construção de visualizações, nós desenvolvemos a ferramenta VisMaker, que utiliza um conjunto de regras para construir as visualizações consideradas mais apropriadas para um determinado conjunto de variáveis. Além de permitir que o usuário defina visualizações através do mapeamento entre variáveis e dimensões visuais, o VisMaker apresenta recomendações de visualizações organizadas através de perguntas construídas com base nas variáveis selecionadas pelo usuário, objetivando facilitar a compreensão das visualizações recomendadas e auxiliando o processo exploratório. Para a avaliação do VisMaker, nós realizamos dois estudos comparando-o com o Voyager 2, uma ferramenta de propósito similar existente na literatura. O primeiro estudo teve foco na resolução de perguntas enquanto que o segundo esteve voltado para a exploração de dados em si. Nós analisamos alguns aspectos da utilização das ferramentas e coletamos os comentários dos participantes, através dos quais pudemos identificar vantagens e desvantagens da abordagem de recomendação que propusemos, levantando possíveis melhorias para esse tipo de ferramenta.

Palavras-chave

Visualização de Informação; Recomendação de Visualizações;
Ferramenta de Visualização.

Table of contents

1	Introduction	14
1.1	Problem, Question, Objectives and Research Scope	16
2	Background	18
2.1	Visualization Effectiveness	18
2.2	Recommender Systems	20
2.3	Ontologies	21
3	Related Works	23
3.1	Visualization Taxonomies and Ontologies	23
3.2	Visualization Construction and Recommender Systems	27
3.2.1	Rule-based Visualization Recommendation Systems	27
3.2.2	Visualization Recommender Systems based on Machine Learning	30
4	The VisMaker Tool	32
4.1	Objective	32
4.2	Requirements	32
4.2.1	Functional Requirements	32
4.2.2	Non-functional Requirements	33
4.3	Technologies Used by VisMaker	33
4.4	VisMaker's User Interface	34
4.4.1	Variables Panel (A)	35
4.4.2	Visual Dimensions Panel (B)	37
4.4.3	Bookmark Panel	37
4.4.4	Questions Panel (D)	37
4.5	Supported Visualizations	38
4.6	Mappings between Data, Questions, and Visualizations	40
4.7	VisMaker's Recommender Engine	45
5	Evaluation	47
5.1	Study 1: Question Answering	47
5.1.1	Procedure	48
5.1.2	Study Execution	49
5.1.3	Tasks	50
5.1.3.1	The CAPES-RJ Task	50
5.1.3.2	The WEATHER Task	51
5.1.4	Post-task Questionnaire	51
5.1.5	Participants	53
5.1.6	Study 1 Findings	53
5.1.6.1	Questionnaire Responses	54
5.1.6.2	Tasks' Correctness	56
5.1.6.3	Generated Charts	63
5.1.6.4	Participants' Feedback about the Tools	65
5.1.7	Discussions	69

5.2	Study 2: Data Exploration	71
5.2.1	Participants	72
5.2.2	Study 2 Findings	72
5.2.2.1	Questionnaire Responses	73
5.2.2.2	Participants' Feedback about the Tools in Study 2	75
5.2.3	Discussions	77
5.3	Answering our Research Subquestions	78
6	Conclusions	81
6.1	Contributions	81
6.2	Future Work	82
A	Informed Consent Form (in Portuguese)	87
B	CAPES-RJ Task Questions	88
C	WEATHER Task Questions	89
D	User Profile Questionnaire (in Portuguese)	90
E	VisMaker Questionnaire (in Portuguese)	91
F	Voyager 2 Questionnaire (in Portuguese)	94
G	Transcription of participants' steps in each experiment (in Portuguese)	97

List of figures

Figure 1.1	The Anscombe's quartet and its data visualizations (Munzner, 2014).	14
Figure 1.2	Cartogram: World population in 2018. Adapted from: https://worldmapper.org/maps/population-year-2018/ . Accessed in: October 2, 2019.	15
Figure 2.1	Visualization channels rank proposed by Cleveland and McGill (1984) (Munzner, 2014).	19
Figure 2.2	Visualization channels ranked by effectiveness according to data and channel type (Munzner, 2014).	19
Figure 3.1	Relationships between object of study, data, data models, and conceptual models (Tory and Möller, 2002).	24
Figure 3.2	The main high-level concepts defined by the visualization ontology (Duke et al., 2004).	25
Figure 3.3	Some VISO concepts (Polowinski and Voigt, 2013).	26
Figure 3.4	Tableau's main interface (Mackinlay et al., 2007).	28
Figure 3.5	The ViSC interface de Sousa and Barbosa (2014).	28
Figure 3.6	Voyager's main interface (Wongsuphasawat et al., 2016).	29
Figure 3.7	The Compass visualization recommendation engine (Wongsuphasawat et al., 2016).	29
Figure 3.8	Main interface of Voyager 2 (Wongsuphasawat et al., 2017).	30
Figure 4.1	Vega-lite specification examples	35
Figure 4.2	VisMaker main interface	36
Figure 4.3	VisMaker variables panel	36
Figure 4.4	VisMaker bookmark buttons	37
Figure 4.5	VisMaker bookmark panel	38
Figure 4.6	VisMaker questions example	39
Figure 4.7	VisMaker supported charts	40
Figure 4.8	The Compass visualization recommendation engine Wongsuphasawat et al. (2016).	45
Figure 4.9	A comparison between Compass engine (Wongsuphasawat et al., 2016) and VisMaker engine.	46
Figure 5.1	Steps of experiment sections.	50
Figure 5.2	Profile questionnaire responses of Study 1.	53
Figure 5.3	Questionnaire responses of Study 1.	55
Figure 5.4	Errors per tools in Study 1.	61
Figure 5.5	A typical chart constructed through Voyager to solve WEATHER T2 in Study 1.	61
Figure 5.6	A typical chart constructed through VisMaker to solve WEATHER T2 in Study 1.	62
Figure 5.7	The moment when participant P14 builds a correct view but finds the wrong answer for CAPES-RJ T1 in Study 1.	62

Figure 5.8	The temporal filter bug in VisMaker.	63
Figure 5.9	Errors per $\langle dataset, tool \rangle$ in Study 1.	63
Figure 5.10	Examples of charts constructed to solve CAPES-RJ T1.	64
Figure 5.11	Examples of charts constructed to solve CAPES-RJ Q1.	64
Figure 5.12	Examples of charts constructed to solve CAPES-RJ T1 in Study 1.	65
Figure 5.13	Profile questionnaire responses of Study 2.	72
Figure 5.14	Questionnaire responses of Study 2.	74

List of tables

Table 2.1	Recommendation techniques (Burke, 2002).	21
Table 4.1	VisMaker functional requirements	33
Table 4.2	VisMaker non-functional requirements	34
Table 4.3	Automatic marks rules (Mackinlay et al., 2007)	41
Table 4.4	Mapping between variable types and Vega-lite types (Wongsuphasawat et al., 2016)	41
Table 4.5	Attribute combinations along corresponding analytic tasks and visualizations (Srinivasan et al., 2019)	42
Table 4.6	Mappings between variable types and chart types used in VisMaker	42
Table 4.7	Mappings between data, questions and visualizations used in VisMaker	44
Table 5.1	Participants per study.	47
Table 5.2	CAPES-RJ variables	50
Table 5.3	WEATHER variables.	51
Table 5.4	Characteristics of the participants in the first study.	54
Table 5.5	Statistical significance of questionnaire responses in Study 1	56
Table 5.6	Correctness of the constructed visualizations in Study 1	56
Table 5.7	Correctness of the written answers in Study 1	58
Table 5.8	Characteristics of the participants in the second study.	72
Table 5.9	Statistical significance of scores in Study 2	75

“In all truth I tell you, unless a wheat grain falls into the earth and dies, it remains only a single grain; but if it dies it yields a rich harvest.”

Jesus Christ, Gospel of John.

1

Introduction

The increasingly rapid growth of data production and the resulting need to exploit them in order to obtain knowledge and answers to a wide range of questions have promoted the development of tools to facilitate the data exploration process through the construction of data visualizations. In addition to data visualizations, data summaries (*i.e.*, mean, variance, correlation, etc.) can make it possible to explore and analyze datasets. However, there are situations in which a data visualization enables greater and more efficient information gain than using a brief summary of the data (Munzner, 2014). To illustrate this, Anscombe (1973) created a set of four datasets, named the “Anscombe’s quartet”. Although they have exactly the same summaries (mean, variance, correlation, and linear regression line), they are quite different, which is evident when their data points are plotted on a Cartesian plane (Figure 1.1).

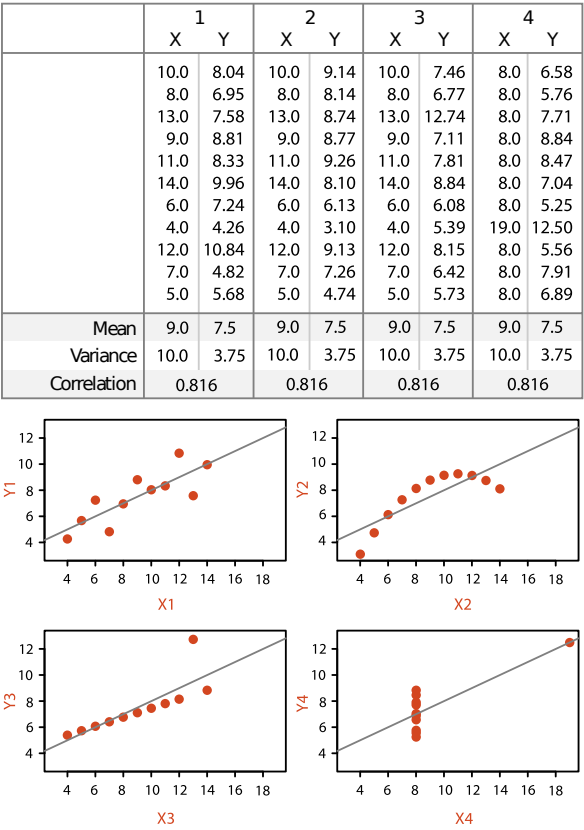


Figure 1.1: The Anscombe’s quartet and its data visualizations (Munzner, 2014).

According to Pinker (1990), each data visualization communicates a set of mathematical values through objects with visual dimensions (*i.e.*, length, position, etc.) that correspond to the respective values. Since data visualizations aim “*to aid our understanding of data by leveraging the human visual systems highly tuned ability to see patterns, spot trends, and identify outliers*” (Heer et al., 2010), they should enable readers to effectively explore datasets, communicating information more accurately, and enabling greater knowledge gain through data.

Figure 1.2 shows a cartogram visualization about world population data for 2018. In this chart, the area of each country is distorted to represent the quantitative data for its population, so countries with a larger population will appear with a larger area compared to countries with a smaller population.

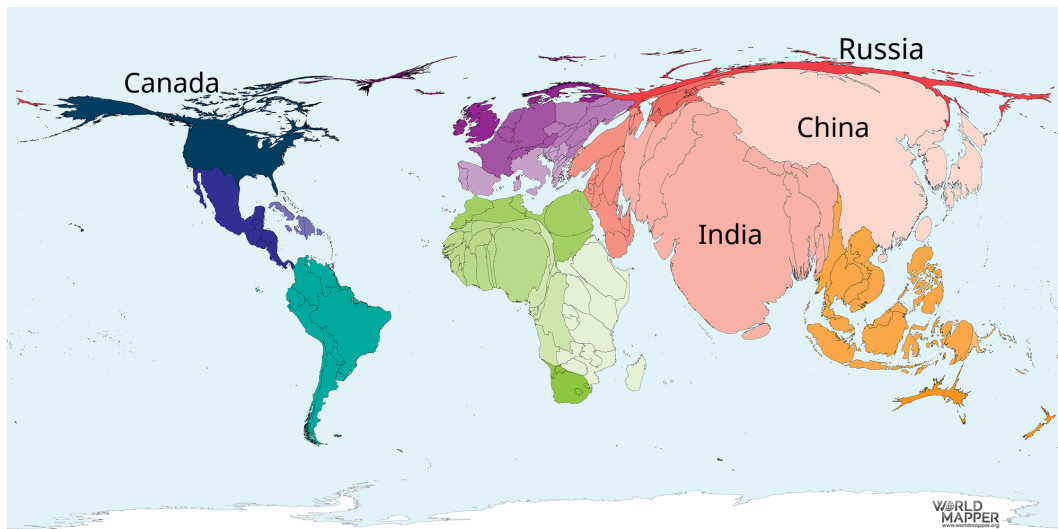


Figure 1.2: Cartogram: World population in 2018. Adapted from: <https://worldmapper.org/maps/population-year-2018/>. Accessed in: October 2, 2019.

In this figure, we can see and be impressed by the huge area occupied by China and India and the tiny area that makes Russia and Canada virtually disappear. Although this visualization does not present quantitative values, it communicates information regarding extreme numbers of demographic concentration in a very efficient and memorable way, making the reader, who would not necessarily be interested in knowing the exact values, aware of the main disparities between some countries.

According to Heer et al. (2010), “*well-designed visual representations can replace cognitive calculations with simple perceptual inferences and improve comprehension, memory, and decision making*”. In addition, the mapping between data and visual dimensions are the “DNA” of all visualizations and various visualization types can be constructed by varying these mappings (Heer

et al., 2010). However, the visualization construction process requires a large number of decisions that involve the specification of which question to ask and what data to use, besides the selection of the most appropriate visual encodings to present the data (Heer et al., 2010). The latter are, in many cases, the main barrier of the construction of good data visualizations, especially by information visualization (InfoVis) novices (Grammel et al., 2010).

1.1

Problem, Question, Objectives and Research Scope

Considering the importance of exploring and communicating statistical data through visualizations, and the difficulties involved in building good data visualizations, this work aims to develop a visualization building and recommendation tool. This tool should be simple, practical, and efficient, enabling less experienced users and InfoVis novices to be able to effectively explore and communicate their data, receiving recommendations for visualizations by identifying the question(s) they wish to answer.

In this context, we define our main research question as follows: ***How can we help data exploration users or InfoVis novices to build useful data visualizations by recommending visualization data to explore datasets?***

Many works have been developed in the Information Visualization field. Some of them discussed about visualization effectiveness, proposing ranks of different visual encoding, sometimes called visual channels, when used to communicate some types of statistical variables (Cleveland and McGill, 1984; Mackinlay, 1986; Munzner, 2014). Other works proposed visualization ontologies or taxonomies that can help develop intelligent visualization systems (Tory and Möller, 2002; Amar et al., 2005; Duke et al., 2004; Polowinski and Voigt, 2013).

More recently, some works presented tools for constructing or recommending data visualizations through rules based on the state-of-the-art of visualization effectiveness (de Sousa and Barbosa, 2014; Gotz and Wen, 2009; Wongsuphasawat et al., 2016, 2017). None of these works, however, take into account the question the user wants to answer, or even the visualization task they must perform on the data. Instead, they merely present to the user a list of recommended visualizations, often without any contextualization. Therefore, in our work we consider, in addition to selecting variables, the possible questions that may help the user when recommending data visualizations, as a way to improve recommendations and usability. Seeking to explore and better understand the importance of presenting the questions the user intends to answer

through visualizations, we formulate the following research subquestions:

- SQ1. *How can a question about data determine the visualization type, including the mapping of data variables onto visual dimensions?*
- SQ2. *How do users interact with the questions presented by the tool via the recommendations?*
- SQ3. *How can questions improve visualization recommendations?*

To answer these questions, we developed VisMaker, a software tool through which the users can build data visualizations and explore a selected dataset. Inspired by Voyager 2 (Wongsuphasawat et al., 2017), VisMaker allows users to construct data visualizations by mapping variables onto visual dimensions, while the system recommends different visualizations grouped by questions through the addition of unselected variables.

2 Background

In this chapter, we present some important concepts for this research. Section 2.1 presents some relevant studies in the field of visualization effectiveness. Section 2.2 introduces the concept of Recommender Systems, describing the main techniques used in this area. The concept of Ontology and its use are presented in Section 2.3.

2.1 Visualization Effectiveness

The main purpose of a visualization is to communicate data (Pinker, 1990). For such, there are visualization channels, that is, visual components that encode data (*i.e.*, position, color, size, etc.). The mapping between data and visualization channels is a significant challenge for designing good visualizations, as certain channels can present specific data types more efficiently than others, so understanding visualization occurs more effectively.

Cleveland and McGill (1984) wrote one of the pioneering works in the evaluation of perceptual elements in information visualization tasks. They listed a set of ten elements and conducted an empirical study to define the ranking of elements in quantitative data visualizations. Mackinlay (1986) performed a similar ranking for quantitative, ordinal, and nominal attributes. Figure 2.1 shows this latter ranking.

As we can see in Figure 2.1, there is greater effectiveness in using the position channel to present any type of attribute. The use of hues, although not highly effective in communicating quantitative data, was considered the second most effective when used to present categorical data.

More recently, Munzner (2014) discussed the effectiveness of each visualization channel, demonstrating, in line with previous works, that some channels have higher effectiveness and promote more significant degrees of perception and understanding in reading visualizations.

Figure 2.2 presents a ranking of several visualization channels according to their effectiveness in presenting ordinal and categorical data (Munzner, 2014). We can once again notice the greater effectiveness made possible by the position channel, listed as position on common or unaligned scales for

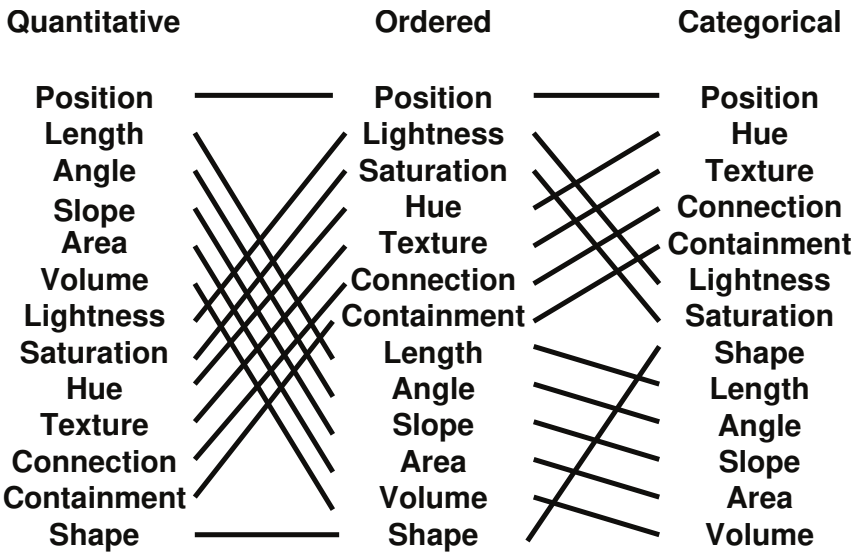


Figure 2.1: Visualization channels rank proposed by Cleveland and McGill (1984) (Munzner, 2014).

ordinal data, and spatial region for categorical data. In this ranking, the size, angle and area channels appear practically in the same order as the ranking proposed by Mackinlay (1986) for quantitative data. This is expected since this second ranking considers that ordinal data is encoded through channels of magnitude, *i.e.*, quantitative. For categorical data, the motion channel appears as an alternative in interactive tools, and the hue appears in second place.

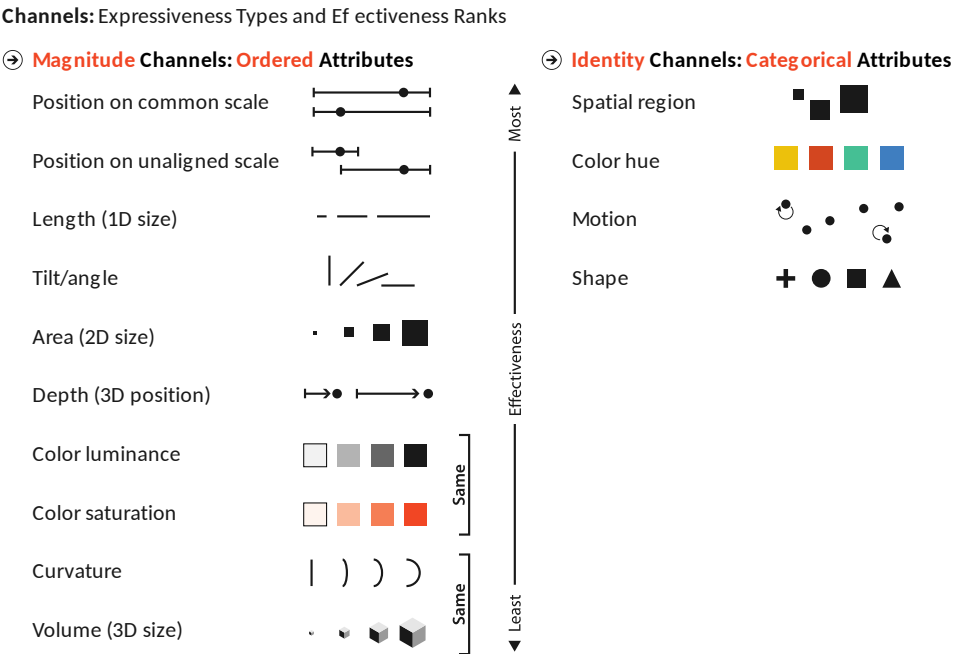


Figure 2.2: Visualization channels ranked by effectiveness according to data and channel type (Munzner, 2014).

Using the rankings presented in these works is quite common among visualization recommender tools, which generally use them to produce visual-

ization recommendations. In this work, a similar procedure was used to define the mappings we should use to recommend visualizations based on user variable selection.

2.2

Recommender Systems

Recommender Systems have become increasingly popular, and their purpose is mainly to assist users in performing everyday tasks such as e-commerce (Schafer et al., 1999), watching movies or series (Gomez-Urbe and Hunt, 2015), and listening to music (Jacobson et al., 2016).

According to Good et al. (1999), “*information filtering recommenders look at the syntactic and semantic content of items to determine which are likely to be of interest or value to a user.*”. Burke (2007) defined recommender systems as personalized information agents that provide suggestions for items that may be useful to the user. Similarly, Ricci et al. (2011) argue that recommender systems are software tools and techniques that provide suggestions for items that are most likely to be of interest to a particular user. These definitions are quite similar and refer basically to the ability of a system to suggest items that suit the user’s interest. Burke (2002) proposed a classification of recommendation techniques, presented in Table 2.1.

In the *Collaborative* techniques, perhaps the most commonly used, the system aggregates user ratings of items, identifies user rating patterns, and makes recommendations based on those patterns. Thus, if a particular item was rated highly by most users, there is a high possibility that it is recommended for other users or for users who are somehow close to the users who rated it highly.

Systems that make *Content-based* recommendations seek to identify the relationship between the user and key item characteristics and then extrapolate that relationship to other items to identify how useful they would be to that user. Thus, if user u rated item i_1 highly, and item i_2 is quite similar to item i_1 , then it is very reasonable that user u also rated item i_2 highly.

The *Demographic* recommendation technique resembles the collaborative technique, but it is based on finding a demographic similarity among users to recommend the most useful items to them. Therefore, knowledge of user u ’s ratings of items is not required.

Utility-based recommender systems require a way to determine which items best fit the needs of each user. They use this metric to rank items and recommend them. Finally, the *Knowledge-based* recommendation technique requires identifying how each item is associated with the user’s needs and

Table 2.1: Recommendation techniques (Burke, 2002).

Technique	Background	Input	Process
Collaborative	Ratings from \mathbf{U} of items in \mathbf{I}	Ratings from \mathbf{u} of items in \mathbf{I}	Identify users in \mathbf{U} similar to \mathbf{u} , and extrapolate from their ratings of \mathbf{i}
Content Based	Features of items in \mathbf{I}	\mathbf{u} 's ratings of items in \mathbf{I}	Generate a classifier that fits \mathbf{u} 's rating behavior and use it on \mathbf{i}
Demographic	Demographic information about \mathbf{U} and their ratings of items in \mathbf{I}	Demographic information about \mathbf{u}	Identify users that are demographically similar to \mathbf{u} , and extrapolate from their ratings of \mathbf{i}
Utility-based	Features of items in \mathbf{I}	A utility function over items in \mathbf{I} that describes \mathbf{u} 's preferences	Apply the function to the items and determine \mathbf{i} 's rank
Knowledge-based	Features of items in \mathbf{I} . Knowledge of how these items meet a user's needs	A description of \mathbf{u} 's needs or interests	Infer a match between \mathbf{i} and \mathbf{u} 's need

then recommending the items that best match their expectations.

In this research, we constructed VisMaker, a knowledge-based recommender system that uses a set of predefined rules to recommend data visualizations based on the user selection of variables.

2.3 Ontologies

An ontology is an explicit specification of a conceptualization (Gruber, 1993). In this context, the term conceptualization refers to an abstract model that defines some aspects of the real world, especially with regard to the concepts and relationships to be modeled, and the definition of explicit specification refers to the lack of ambiguities in the model, which makes it processable by both humans and machines (Staab and Studer, 2010).

Staab and Studer (2010) stated that the use of ontologies has gained much space, and several applications in fields such as knowledge management, information integration, cooperative information systems, information retrieval, and even e-commerce have been developed. In recommender systems, domain ontologies, which define the concepts related to data domain, have been used to add a semantic dimension to items and user profiles, improving the recommendations provided by content-based systems (Staab and Studer, 2010).

In our work, we do not define or use ontologies, but only simpler structures that make it possible to carry out the process of recommending visualizations. Besides that, in the InfoVis field, some recent works have proposed the development of categorizations and ontologies that could provide, especially for automated systems, an effective way to represent data visualizations and manipulate them. We present some of these works in Chapter 3. In this work, we

3

Related Works

Many works have been developed in the Information Visualization field, ranging from studies that identify the perceptual properties of visual elements, through categorizations of charts and analysis tasks, to the development of tools to assist non-expert users in building good visualization or exploration of datasets. This chapter presents two categories of work related to this research: Section 3.1 presents works that have defined visualization taxonomies or ontologies. Section 3.2, in turn, discuss some visualization recommendation tools. These works, in addition to the visualization effectiveness studies presented in the previous chapter, help guide the development of this research in the field of information visualization.

3.1

Visualization Taxonomies and Ontologies

The development of visualization taxonomies helps to better organize and understand the state of research in this field. Ontologies, in turn, enable the construction or recommendation of data visualizations by automated systems, since they must provide an unambiguous language so that machines can also use it. Some works have proposed visualization taxonomies or ontologies and are essential for this research, as they help us identify paths for the automatic recommendation of data visualizations.

Tory and Möller (2002) defended the idea that visualizations about a given data set may vary according to the conceptions that each builder or reader has at a given moment about the data. They proposed a high-level taxonomy that attempts to unite the concepts of scientific visualization and information visualization by categorizing by the data model. They also highlight user participation in the construction of visualizations, since their preconceptions and interpretations about the object of study influence how to present the data. Figure 3.1 presents this relationship.

Wehrend and Lewis (1990) proposed a classification of visualization operations that are independent of particular application domains. More recently, Amar et al. (2005) presented a categorization of low-level analytical activities commonly performed by users of information visualization tools.

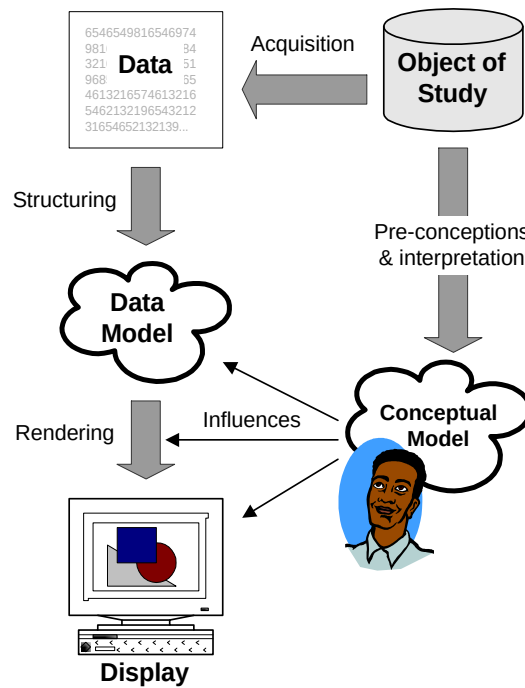


Figure 3.1: Relationships between object of study, data, data models, and conceptual models (Tory and Möller, 2002).

They conducted a study with five datasets and collected approximately 200 sample questions that a group of students used to analyze this data. From these answers, the authors developed a taxonomy that brings together 10 tasks most commonly used in data analysis. They are listed and described below:

1. **Retrieve Value:** find certain attribute values for specific data cases;
2. **Filter:** given a set of concrete conditions, find the data instances that satisfy those conditions;
3. **Compute Derived Value:** compute an aggregate value (*e.g.*, mean, sum) for a given set of data cases;
4. **Find Extremum:** find data cases that have extreme values in certain attributes;
5. **Sort:** sort a set of data cases according to some ordinal metric;
6. **Determine Range:** given a set of data cases and an attribute of interest, find the range of values;
7. **Characterize Distribution:** given a set of data cases and a quantitative attribute of interest, characterize the distribution of the values of that attribute values over the set;

8. **Find Anomalies:** identify any anomalies in a given data set against a specific expectation;
9. **Cluster:** find similar attribute groupings for a given set of data cases; and
10. **Correlate:** given a set of data cases and two attributes, determine useful relationships between the values of those attributes in the set.

The high-level visualization ontology proposed by Duke et al. (2004) is divided in four concept groups:

- **task and use:** why are the data being visualized, and what is being done with the visualization?
- **representation:** how are the data presented?
- **process:** how can the transformation from data to representation be expressed in terms of a given set of primitives?
- **data:** how are the data processed through the various stages of the transformation process expressed and organized?

Figure 3.2 presents the main high-level concepts defined by the ontology proposed in Duke et al. (2004).

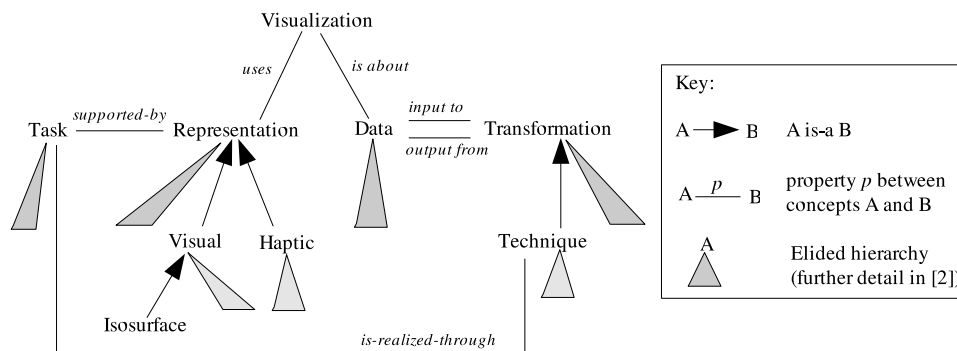


Figure 3.2: The main high-level concepts defined by the visualization ontology (Duke et al., 2004).

More recently, Voigt and Polowinski (2011) undertook a systematic literature review on visualization terminologies, taxonomies, and ontologies, and pointed to the need to create an ontology that would unify concepts and address the inconsistencies and ambiguities of terms that exist in the InfoVis field. The authors presented an initial specification of a visualization ontology that culminated in the construction of the VISO¹ (Polowinski and

¹<http://purl.org/viso/>

Voigt, 2013), a visualization ontology divided in seven modules that define the concepts related to attributes and graphic representations (GRAPHIC), the variables and data structures (DATA), the human aspects of the visualization (*e.g.*, tasks, actions and operations) (ACTIVITY), the relationships and rankings between visualization types found in the literature (FACTS), and some particular concepts that describe the visualization context in a specific domain (SYSTEM, USER, and DOMAIN).

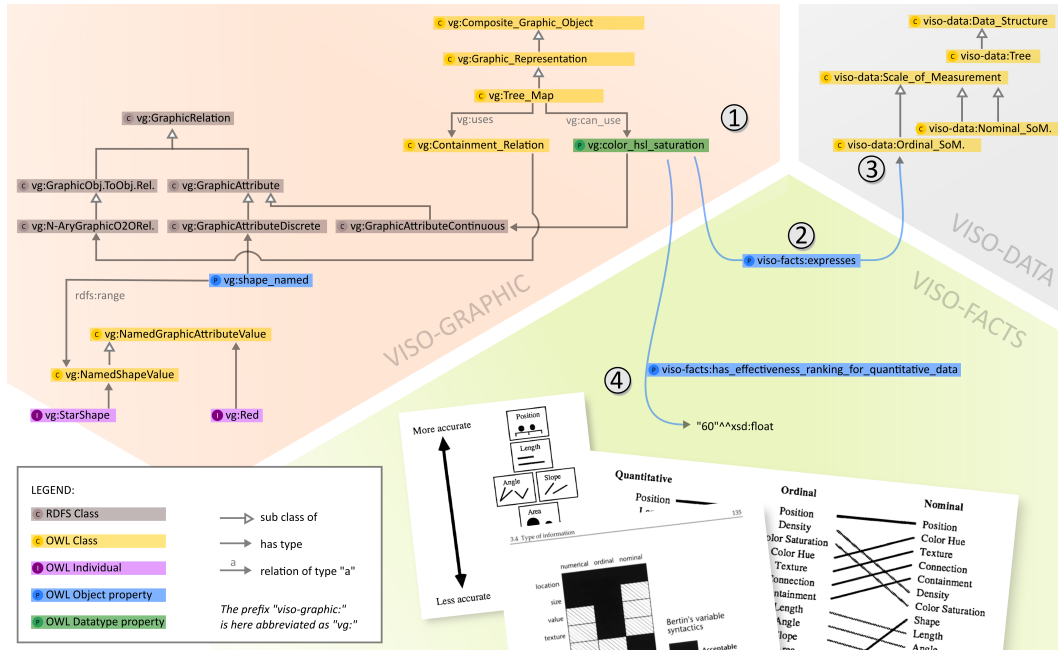


Figure 3.3: Some VISO concepts (Polowinski and Voigt, 2013).

Figure 3.3 presents some terms defined by the three main modules of VISO (GRAPHIC, DATA, and FACTS). In the figure, the attribute `vs:color_hsl_saturation` (1) can be used to express (`viso-facts:expresses`) (2) an ordinal scale of measurement (`viso-data:Ordinal_SoM`). In addition, this attribute is assigned with the value “60” of effectiveness for quantitative data on an ordinal scale.

de Sousa and Barbosa (2014) used the taxonomies proposed by Tory and Möller (2002) and Amar et al. (2005), together with the ontology presented by Duke et al. (2004), to construct an ontology that maps questions and tasks onto suitable visualization types. This approach made it possible to develop a tool to support inexperienced users in building visualizations by identifying the question the user would like to answer about the data, and by presenting alternative questions and their respective visualizations during the visualization recommendation process.

3.2

Visualization Construction and Recommender Systems

Among the works that addressed the visualization recommendation problem, two main approaches are commonly used: (1) the use of rules, such as ontologies, which define the mapping between data and suitable visualizations to present them, taking into consideration or not the identification of tasks or questions that the user intends to perform or answer through visualization; and (2) building optimization or machine learning models that attempt to “learn” and identify the relationship between data, metadata, and the visualization types that best fit your presentation.

3.2.1

Rule-based Visualization Recommendation Systems

Some tools that enable the construction of statistical data visualizations are well known. Microsoft Excel² and LibreOffice Calc³, for example, allow users to build charts using data from a spreadsheet.

One of the first tools to specify and automatic design effective graphical representations was the APT system, proposed by Mackinlay (1986). APT uses a set of primitive graphical languages and expressiveness and effectiveness criteria. More recently, tools such as Tableau⁴ (Stolte et al., 2002) have been developed to enable the construction of data visualizations more efficiently. With Tableau, the users can specify the visualization by defining which visualization channels will be responsible for presenting each selected attribute. Its main interface is shown in Figure 3.4. Tableau also provides recommendations for different visualization types, through the *Show Me* interface (Mackinlay et al., 2007).

Gotz and Wen (2009) introduced a visualization recommender system guided by user behavior. The system consists of two phases: pattern detection and visualization recommendation. In the first phase, the system attempts to identify user interaction patterns. In the second phase, the system uses the detected patterns to infer the visual task that the user intends to perform on the data, recommending alternative visualizations. The work concluded that, while users eventually ignored very “subtle” recommendations, more “aggressive” recommendations could interrupt their cognitive process, requiring greater transparency in the recommendation process.

²<https://products.office.com/pt-br/excel>

³<https://pt-br.libreoffice.org/descubra/calc/>

⁴<https://www.tableau.com/>

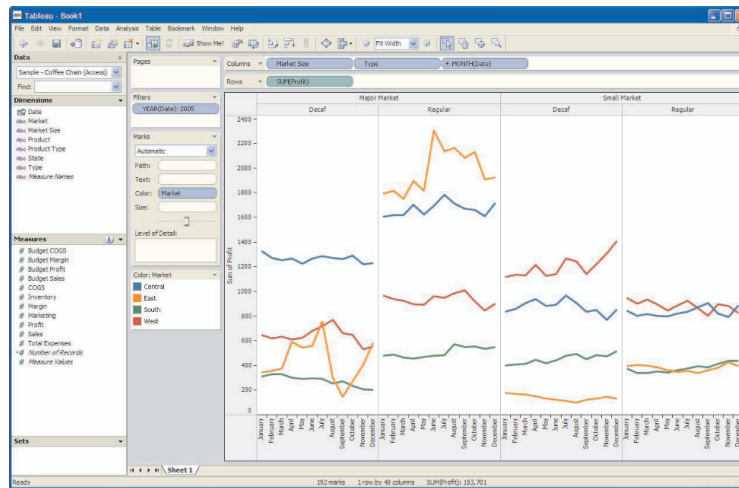


Figure 3.4: Tableau's main interface (Mackinlay et al., 2007).

The VizDeck tool, proposed by Perry et al. (2013), automatically recommends a set of visualizations based on statistical properties of the data. Visualizations can be fetched by a search field through which the user can enter visualization types or attribute names by filtering the visualizations.

de Sousa and Barbosa (2014) developed the ViSC tool, whose main interface is shown in Figure 3.5. The tool comprises a recommender system to support chart construction for statistical data. ViSC uses an ontology that defines the task that the user can perform on the data, and maps it onto questions and visualizations that could be used to answer it. The system works as follows: first, the user selects which data will be visualized; the system presents a visualization deemed appropriate for the user and lists some questions related to the selected data and the current visualization, each linked to a new one visualization.

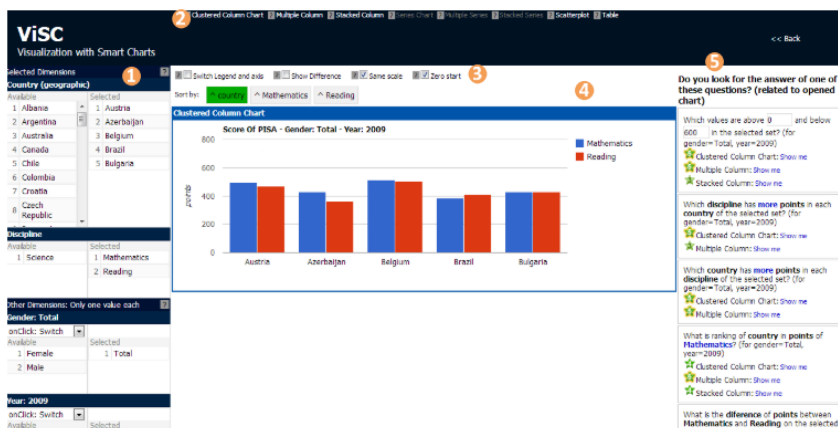


Figure 3.5: The ViSC interface de Sousa and Barbosa (2014).

More recently, Wongsuphasawat et al. (2016) presented the Voyager tool (Figure 3.6), whereby users can select attributes of a given dataset receiving

visualization recommendations considering the selected attributes. One of the objectives of the tool is to enable the process of data exploration through browsing, skipping the need for the user to define the mapping between data and visualization channels.



Figure 3.6: Voyager's main interface (Wongsuphasawat et al., 2016).

Figure 3.7 presents the process followed by the *Compass* visualization recommendation engine, used by Voyager to recommend data visualizations. When the user selects some attributes to compose a visualization (A), the system generates new variable combinations by adding new attributes to previously selected ones, enumerating and applying mathematical transformations on attributes (B). After that, the system defines the recommended visualizations using Vega-Lite grammar (C) and presents a clustered list of possible ranked visualizations using perceptual effectiveness metrics (D).

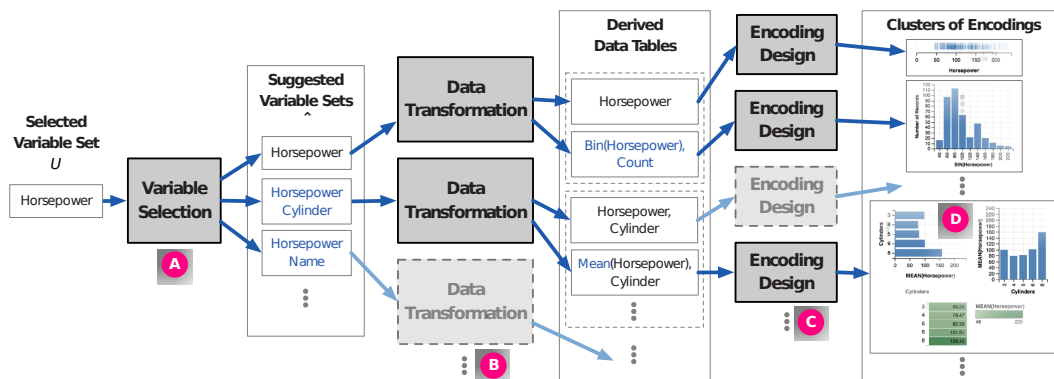


Figure 3.7: The Compass visualization recommendation engine (Wongsuphasawat et al., 2016).

When comparing Voyager and PoleStar, a simple implementation of the Tableau's interface, the authors concluded that, while Voyager works very

well to help users explore datasets, PoleStar can be more effective when the user already knows the dataset and has some experience in building data visualizations.

As an evolution of Voyager, Wongsuphasawat et al. (2017) presented Voyager 2,⁵ whose main interface is shown in Figure 3.8. In addition to recommending data visualizations based on user variables selection, Voyager 2 allows users to specify a visualization defining the mapping between variables and visual dimensions.

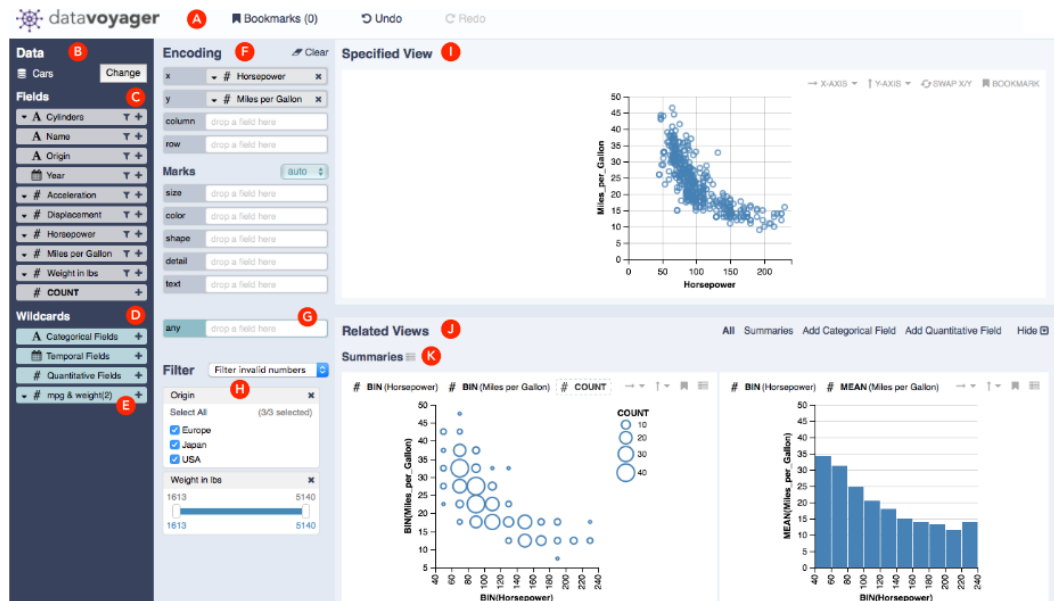


Figure 3.8: Main interface of Voyager 2 (Wongsuphasawat et al., 2017).

3.2.2

Visualization Recommender Systems based on Machine Learning

Dibia and Demiralp (2018) formulated visualization projection as a translation problem between data specifications and visualization specifications. The authors trained a Machine Learning (ML) model based on Neural Networks with a set of visualization specifications defined in Vega-Lite grammar. The work demonstrated the feasibility of this approach through the results obtained by the model and performed the integration of the model with a web tool that recommends a list of visualizations for a given user-specified dataset.

Luo et al. (2018) proposed DeepEye, a system for automatic data visualization that addresses three issues: (i) **visualization recognition**: given a visualization, identifying whether it is “good” or “bad”; (ii) **visualization ranking**: given two visualizations, identify which one is the best; and (iii) **visualization selection**: given a dataset, find the top- k visualizations. The

⁵<https://vega.github.io/voyager/>

authors collected 42 datasets from different real-world domains and listed for each one some possible data visualizations. Next, a group of 100 students labeled each visualization as “good” or “bad” and, for each pair of “good” visualizations, identified which one would be the best. These annotations were used to train and test some ML models, including a Decision Tree algorithm to perform the visualization recognition task, and a Learn to Rank algorithm along with a partial ranking approach to rank and select visualizations.

Hu et al. (2019) presented the construction of a dataset of 120,000 tables related to a particular data visualization and developed an ML model to classify and then recommend visualizations for certain attribute combinations. The tables and their visualizations were obtained from the Plotly Community Portal.⁶ The authors followed a process to remove duplicate tables and, to avoid model bias, they collected only one table per user. They applied some classification algorithms to perform five classification tasks: two at the visualization level and three at the attribute level.

As we describe in the next chapter, VisMaker is a new tool that, similar to Voyager 2, uses a combination of manual and automatic visualization specifications. The significant difference of VisMaker is that the presentation and recommendation of visualizations are organized by questions that aim to facilitate the understanding of the recommendations, helping the user to make and test hypotheses about a given dataset during data exploration, thus reducing their “*visual mapping barrier*” Grammel et al. (2010).

⁶<https://plot.ly/feed/>

4

The VisMaker Tool

In this chapter, we discuss the main design aspects we have taken into account when developing VisMaker. We also introduce the VisMaker’s user interface and the mappings between data, questions and visualization types that we used in the visualization recommendation process.

4.1

Objective

We developed VisMaker to make it easier to explore datasets through data visualizations. We have taken into account the “*visual mapping barrier*” (Grammel et al., 2010) that many users face when they are constructing data visualization. To facilitate the understanding of recommendations, VisMaker organizes the recommended visualizations in groups of questions, aiming to add some semantic value to the recommendations. We developed a mechanism to relate possible questions with variable types, helping users to recognize what visualizations are in accordance with their interests and to know what the visualizations precisely mean.

4.2

Requirements

To guide the development of VisMaker, we defined some functional and non-functional requirements beforehand. They are presented in the next subsections.

4.2.1

Functional Requirements

Table 4.1 lists and describes the VisMaker functional requirements, from uploading the dataset to exporting the visualization. They have been identified through inspection and use of existing systems and tools.

ID	Name	Description	Priority
FR01	Load generic datasets	The system must allow the user to upload any dataset in CSV format	High
FR02	Change variable types	The system must allow the user to change the type of each dataset variable loaded	High
FR03	Construct visualizations	The system must allow the user to construct appropriate visualizations of the uploaded dataset	High
FR04	Apply data transformations	The system must allow the user to apply data transformations (<i>e.g.</i> , aggregate functions) on the variables of the uploaded dataset	High
FR05	Recommend visualizations	The system should recommend new data visualizations according to the variable selections by the user	High
FR06	Apply data filters	The system must allow the user to apply filters on the values of the selected variables	Medium
FR07	Export visualizations	The system must allow the user to export the constructed visualizations by saving them to a file on their computer	Low

Table 4.1: VisMaker functional requirements

4.2.2

Non-functional Requirements

Table 4.2 presents and describes the non-functional requirements of the VisMaker tool.

4.3

Technologies Used by VisMaker

We developed VisMaker as a Web application, using frameworks like Vue.js¹ and BootstrapVue². We also used Vega-lite³ (Satyanarayan et al., 2016), a specification language for constructing and embedding data visualizations.

¹<https://vuejs.org/>

²<https://bootstrap-vue.js.org/>

³<https://vega.github.io/vega-lite/>

ID	Name	Description
NFR01	Usability	The system should be easy to use, allowing inexperienced users to be able to generate good data visualizations
NFR02	Performance	The system should quickly generate and display the data visualizations

Table 4.2: VisMaker non-functional requirements

Vega-lite was developed to be a high-level specification language to facilitate the visualization construction process through the low-level grammar Vega, presented by Satyanarayan et al. (2015). According to Wongsuphasawat et al. (2016), Vega-lite “*consists of a set of mappings between visual encoding channels and (potentially transformed) data variables*”. Vega-lite specifications are defined as JSON (JavaScript Object Notation) objects, which are compiled into Vega specifications.

The Vega-lite unit specifications, which VisMaker uses, can describe a Cartesian plot. Each specification is defined as a four-tuple:

$$unit := (data, transforms, mark, encoding)$$

where *data* defines the data source, *transforms* defines the data transformations, *mark* determines the geometric object used to visually encode the data records, and *encoding* determines the mappings between data attributes and visual channels.

Figure 4.1 presents two Vega-lite specifications about a stock market dataset⁴ and their respective visualizations. In (A) we define a line chart (“*mark*”: “*line*”) that presents the average of price along the years. In (B) there is a distinction of action symbols by adding the symbol field to the color channel, defined by the encoding object. We highlight all the encoding channels to make it easier to understand the charts specification.

4.4

VisMaker’s User Interface

Figure 4.2 presents the VisMaker user interface after having loaded the Iris dataset and added the `sepal_length` variable to the X-axis. In the figure, panel (A) presents all the dataset variables and their types. Panel (B) shows all the supported visual dimensions, represented as dropdown widgets. To define the visualization that will be presented in panel (C), users drag the variables

⁴<https://github.com/vega/datalib/blob/master/test/data/stocks.csv>

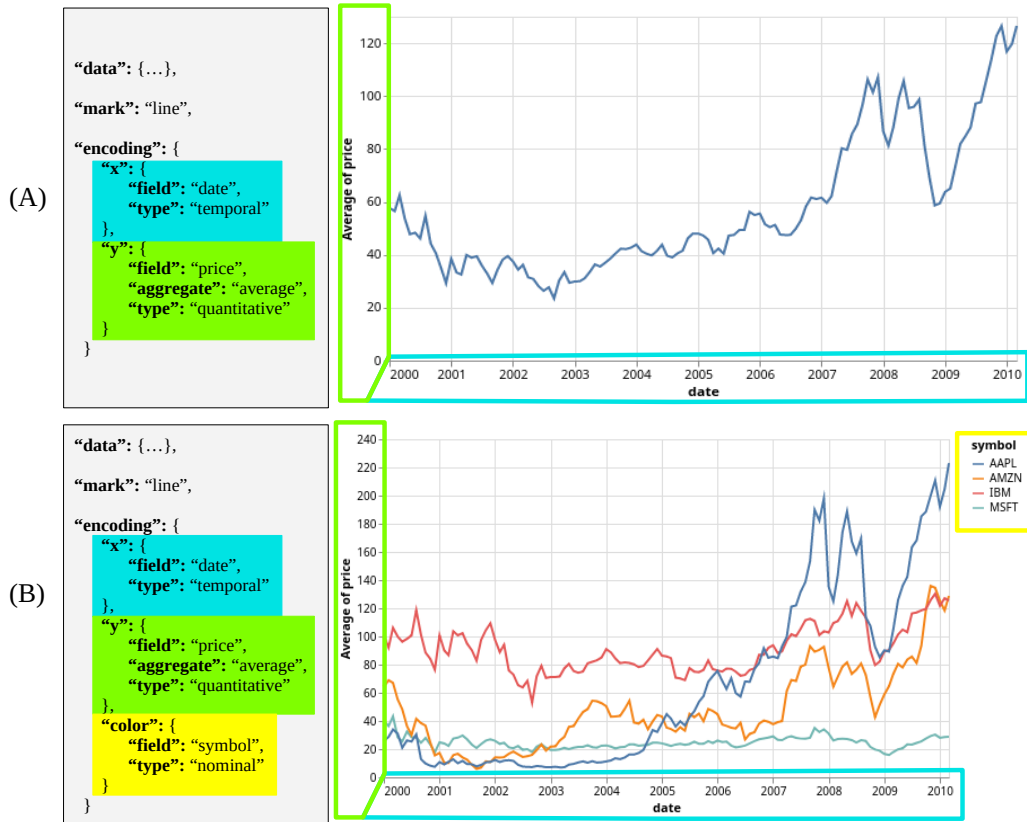


Figure 4.1: Vega-lite specification examples

of interest from panel (A) and drop them to the respective field in panel (B). Panel (D) presents a list of related questions and visualizations constructed by adding new variables to the current variable selection.

After loading a dataset, panels (C) and (D) present, respectively, an empty chart and a list of questions and charts created by selecting a single variable, as no variable has yet been selected. In addition, only the X and Y axes fields in panel (B) are available to receive mapped variables, since the other visual dimensions require mapping at least one variable to the X or Y axis. In the following, we present each panel in more detail.

4.4.1

Variables Panel (A)

VisMaker currently considers only four variable types: quantitative (Q), nominal (N), ordinal (O), and temporal (T). We defined a color scheme that visually distinguishes the presentation of each variable type: blue for quantitative, green for nominal, and yellow for temporal variables. Figure 4.3 presents the VisMaker variables panel when the Iris dataset is loaded. This dataset has four quantitative variables: `sepal_length`, `sepal_width`, `petal_length`,

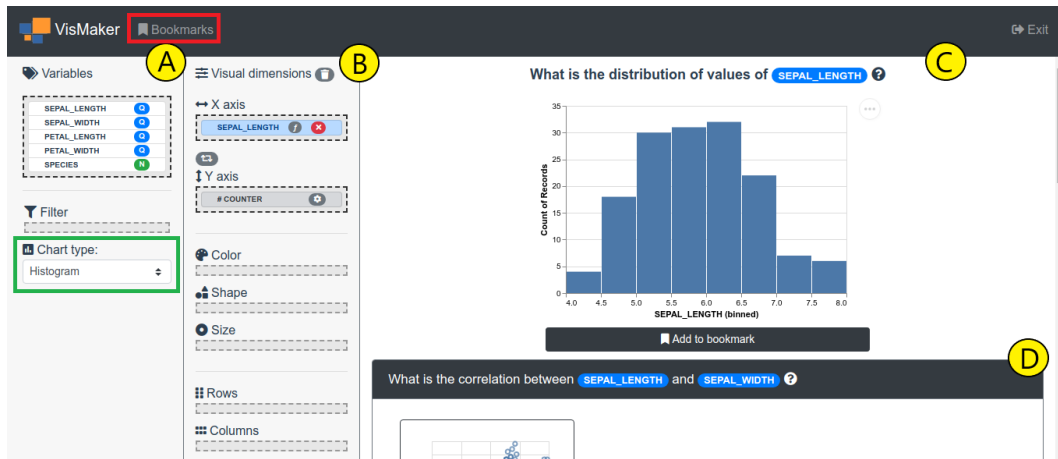


Figure 4.2: VisMaker main interface

`petal_width`; and one nominal variable: `species`.

By default, when loading a dataset, VisMaker automatically assigns the quantitative type to any variable that has numeric values. The nominal type is assigned to variables that have textual values. The temporal type is automatically assigned when a column has values in date format. To avoid erroneous inferences, the ordinal type is never assigned automatically. Through the variables panel, users can change the variable types by clicking on the colored button with the letter that identifies the variable type.

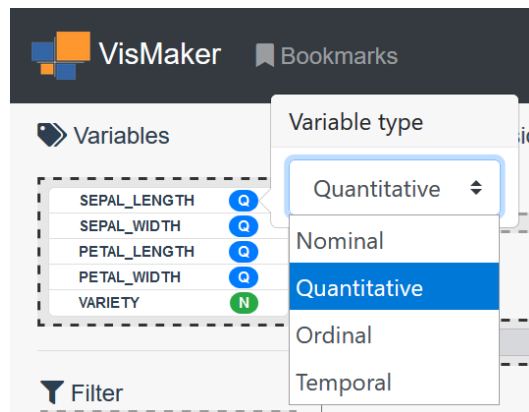


Figure 4.3: VisMaker variables panel

Figure 4.2 highlighted the chart type selection field and the “Bookmark” menu through the green and red rectangles, respectively. The chart type selection field can be used to select a chart type when more than one chart type is available for the mapping defined by the user. We describe the Bookmark menu in more details in Section 4.4.3.

4.4.2

Visual Dimensions Panel (B)

VisMaker provides a set of seven visual dimensions (X axis, Y axis, color, shape, size, rows, and columns), which represent visualization channels and can be used to specify the main visualization (panel (C)) by using the Vega-Lite⁵ grammar (Satyanarayan et al., 2016).

4.4.3

Bookmark Panel

Wongsuphasawat et al. (2016) proposed a bookmark menu through which users can save visualizations to access them more easily when using the Voyager tool in later sessions. VisMaker also has a similar menu. To add a visualization to the bookmark list, users can use the “Add to bookmark” button that can be accessed either through the main visualization panel (Figure 4.4 (A)) or through the questions panel (Figure 4.4 (B)).

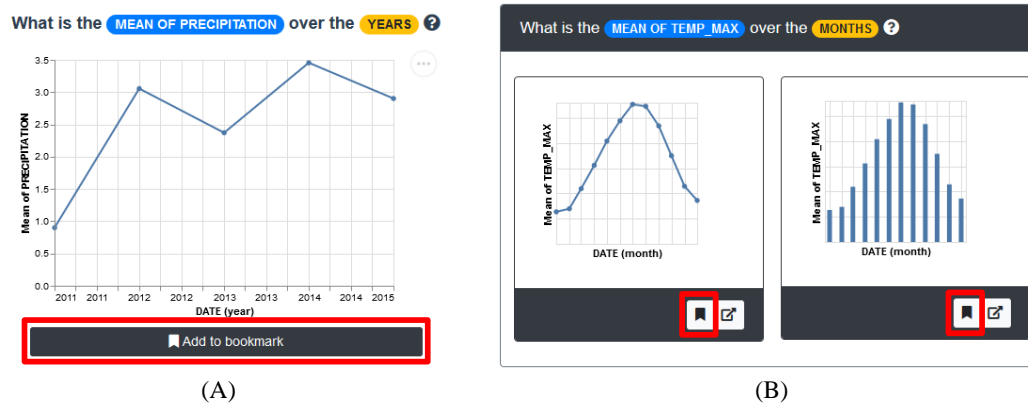


Figure 4.4: VisMaker bookmark buttons

After saving some visualizations to the bookmark list, users can access them through the Bookmarks menu, in the upper left corner of the screen, opening the bookmark modal panel illustrated in Figure 4.5.

4.4.4

Questions Panel (D)

The questions panel presents a list of cards, illustrated in Figure 4.6. The list contains a questions and a corresponding set of visualizations that can be used to answer each question. Each visualization is presented in a card that has a bookmark button (see Section 4.4.3) followed by a button that can be used to transfer the corresponding chart into the main visualization

⁵<https://vega.github.io/>

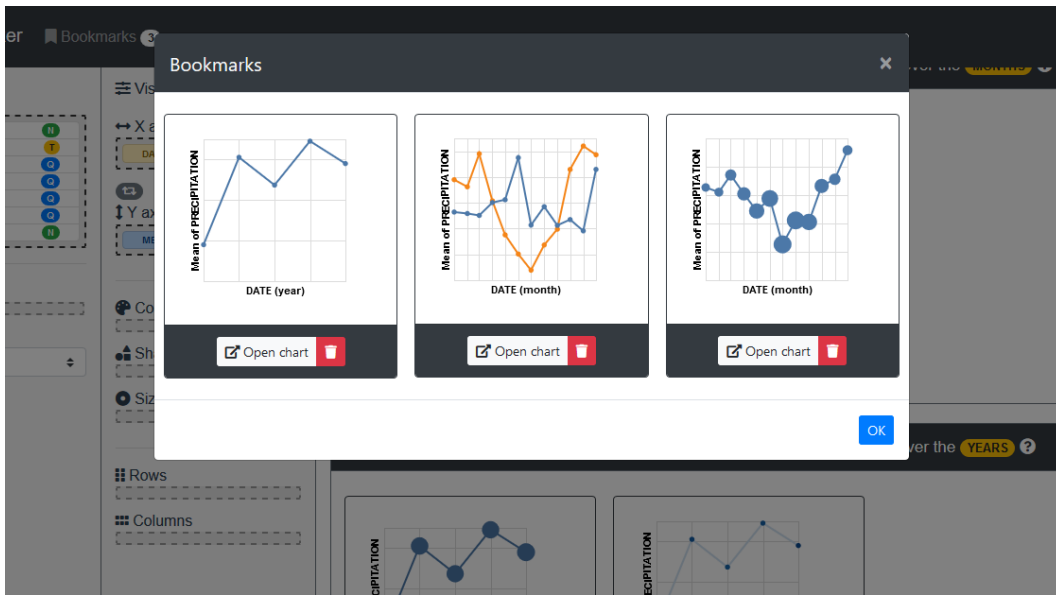


Figure 4.5: VisMaker bookmark panel

panel. In the Questions panel, we also use the color defined for each variable type, highlighting the variable names whenever they appear in a question to facilitate their identification.

4.5

Supported Visualizations

VisMaker supports eight visualization types, which have been selected taking into account the Vega-lite capabilities. They are listed and described below and illustrated in Figure 4.7:

- **Histogram:** describes the distribution of a single numeric variable, where each range of values (*bin*) is represented by a bar and the length of the bar is defined by the number of occurrences of values in that range in the dataset;
- **Bar chart:** presents the relationship between a numeric variable and a categorical variable. In this chart, categorical values are represented by a bar whose length represents the corresponding numerical value;
- **Heatmap:** relates two discrete variables across two axes in the same plane, with a color scale that expresses a numeric value related to both variables (*e.g.*, the number of co-occurrences of the values of two nominal variables of a dataset);
- **Scatter plot:** shows the relationship between two numeric variables so that, for each data instance, the value of one of the variables is mapped onto the X axis and the value of the other variable is mapped onto the Y axis of the same Cartesian plane;

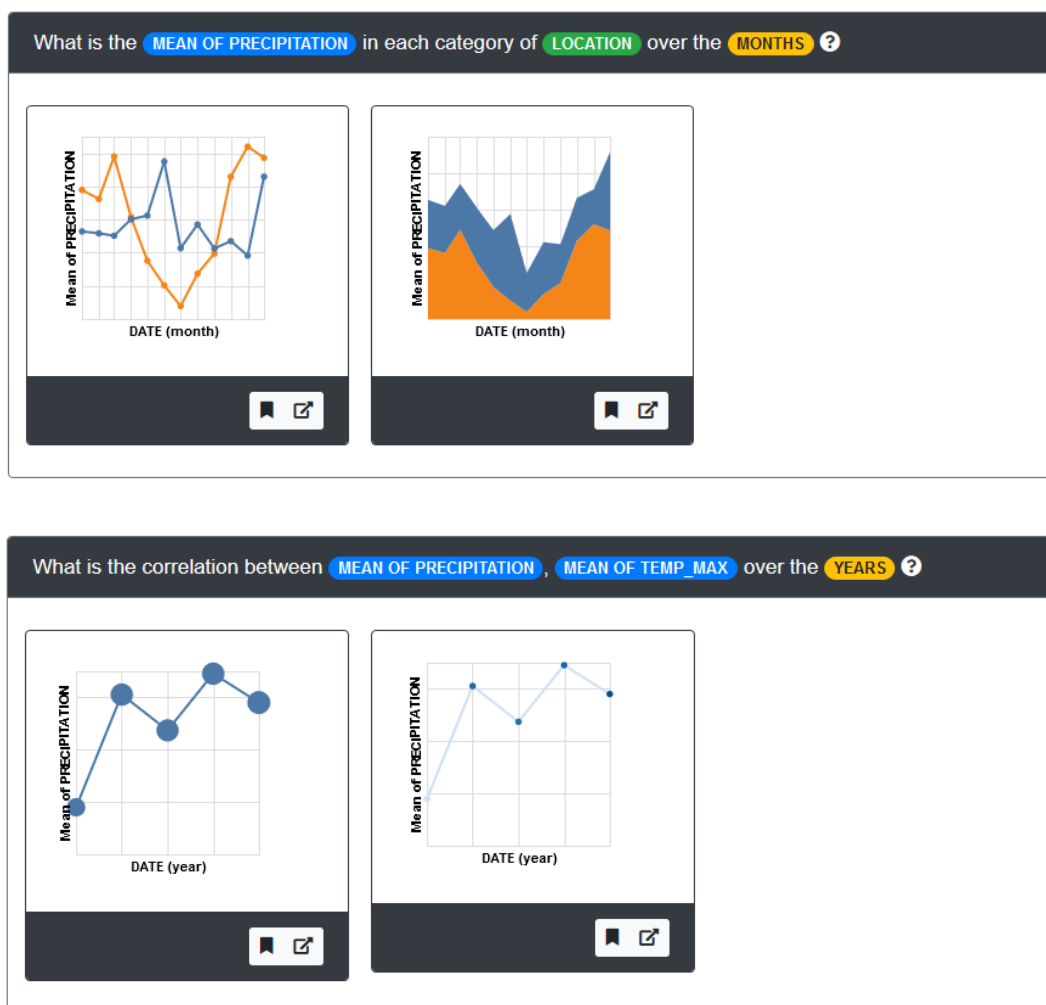


Figure 4.6: VisMaker questions example

- **Box plot:** presents an abbreviated visualization of the distribution of one or more numerical variables, representing their minimum, maximum, median, and first and third quartiles;
- **Strip plot:** similar to a box plot, it makes it possible to visualize the distribution of numeric variables in more detail, showing where each data instance is located in an axis;
- **Line chart:** displays the evolution of one or more numeric variables over an ordered dimension. It is often used to visualize time series;
- **Area chart:** is quite similar to a line chart, but the quantitative values are presented by filled regions (“under the lines”).

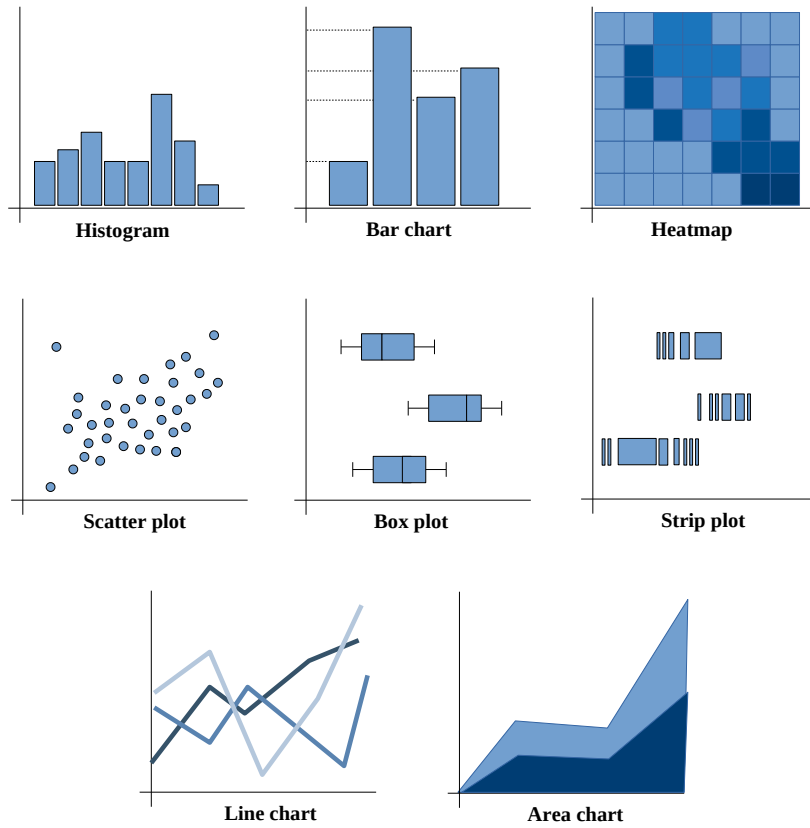


Figure 4.7: VisMaker supported charts

Some variations of these charts are also supported. For example, when you have constructed a bar chart, you can modify it to a stacked bar chart by mapping a categorical variable onto the color field. A bubble chart, in turn, can be obtained after mapping a quantitative variable to the size field when you already have a scatter plot built. In addition to these, clustered charts can also be constructed by mapping any categorical variable to the columns or rows fields.

4.6

Mappings between Data, Questions, and Visualizations

The Tableau *‘Show Me’* interface (Mackinlay et al., 2007) used a mapping between variable types to automatically identify the visualization type. Table 4.3 presents the proposed mapping for categorical (C), categorical date (Cdate), quantitative dependent (Qd) and quantitative independent (Qi) variable types.

In turn, the Voyager tool proposes a mapping between variable types and the Vega-lite mark types. Vega-lite considers quantitative (Q), nominal (N), ordinal (O), and temporal (T) data. Table 4.4 presents the proposed mapping.

Srinivasan et al. (2019) mapped some of the data analytic tasks identified by Amar et al. (2005) (see Section 3.1) onto a set of variable types to

Pane Type		Mark Type	View Type
Field	Field		
C	C	Text	Cross-tab
Qd	C	Bar	Bar view
Qd	Cdate	Line	Line view
Qd	Qd	Shape	Scatter plot
Qi	C	Gantt	Gantt view
Qi	Qd	Line	Line view
Qi	Qi	Shape	Scatter plot

Table 4.3: Automatic marks rules (Mackinlay et al., 2007)

Data Types	Mark Types
Q	tick > point > text
(O or N) \times (O or N)	point > text
Q \times N	bar > point > text
Q \times (T or O)	line > bar > point > text
Q \times Q	point > text

Table 4.4: Mapping between variable types and Vega-lite types (Wongsuphasawat et al., 2016)

understand which visualizations and data facts should be presented when the user defines a combination of attributes. These mappings are presented in Table 4.5, especially in the columns **Attribute combinations**, **Task(s)** and **Visualizations**.

Based on these mappings, we constructed two mappings that define which charts should be presented in the main visualization panel (C) according to the variable mapping defined by the user, as well as what related questions and charts will be listed in panel (D) if a new variable from the unselected variables list is added, similar to the approach presented by Wongsuphasawat et al. (2016).

VisMaker has two recommender systems that uses a set of rules constructed based on those mappings. The first one consists of determining the most appropriate chart according to the mapping between variables and visualization encodings that the user has defined. It includes the automatic addition of aggregate summaries, which can anticipate some user actions, directly generating some visualizations that require the step of adding a ‘counter’ attribute in other tools, like Voyager 2. For this recommender system, we constructed a mapping that defines which chart mark should be used according to variable type combinations for the X and Y axes. Table 4.6 presents the mappings used in VisMaker.

Attribute Combination	Example Data Fact	Task(s)	Visualization	Embellishments							
				O	S	IL	TH	RL	CH	QL	
N	Pontiac Grand Prix has highest value for Horsepower	Find Extremum	Strip plot								
	Most values for Horsepower are in the range 75 - 125	Characterize Distribution	Strip plot								
			Box plot								
			Histogram								
	Pontiac Catalina appears to be an outlier	Find Anomalies	Strip plot								
			Box plot								
C	Europe has the least number of items	Find Extremum	Bar chart								
	Number of items in US is 2.57 times the number of items in Europe	Characterize Distribution	Donut chart								
N x N	Acceleration and Displacement have a strong inverse correlation	Correlation	Scatterplot								
	Most items in the dataset have high Horsepower and low MPG	Characterize Distribution									
C x N	Average Retail Price of SUV is 1.76 times Sedan	Characterize Distribution (+Derived Value)	Bar chart								
	Japan has highest average MPG	Find Extremum (+Derived Value)	Donut chart								
		Europe has item (Fiat 128) with lowest value for Displacement	Find Extremum	Strip plot							
			Scatterplot								
C x C	The largest group of items in the dataset have Origin:Europe and Cylinders:5	Find Extremum	Stacked bar chart								
	US has most number of items. Most items in US belong to 8 for Cylinders	Characterize Distribution (+Find Extremum)	Scatterplot + Size								
N x N x N	Most items with low MPG and low Weight also have low Horsepower	Characterize Distribution	Scatterplot + Size								
C x N x N	Overall, Displacement and Weight have a strong correlation	Correlation	Scatterplot								
	Items with Origin:Japan exhibit a strong correlation between Displacement and Weight	Correlation (+Filter)	Scatterplot + Color								
	Most items with Origin: Europe have low Displacement and low Weight	Characterize Distribution (+Filter)	Scatterplot + Color								
			Scatterplot + Size								
C x C x N	datsun 1200 with lowest value for Weight has Cylinders:4 and Origin:Japan	Find Extremum	Strip plot + Color								
			Scatterplot + Color								
	Items with Origin: Japan and Cylinders: 4 have lowest AVG(Weight)	Find Extremum (+Derived Value)	Strip plot + Color								
			Scatterplot + Color								
			Scatterplot + Size								

Table 4.5: Attribute combinations along corresponding analytic tasks and visualizations (Srinivasan et al., 2019)

Variable types	Visualization types
Q	Density chart; Histogram
N	Bar chart
O	Line chart; Bar chart; Area chart
T	Line chart; Bar chart; Area chart
Q x Q	Scatter plot
Q x N	Box plot; Strip plot
Q x O	Line chart; Bar chart; Area chart
Q x T	Line chart; Bar chart; Area chart
N x N	Heatmap
N x O	Heatmap
N x T	Heatmap
O x O	Heatmap
O x T	Heatmap
T x T	Heatmap

Table 4.6: Mappings between variable types and chart types used in VisMaker

The second recommender system is used to support users explore the data. It is based on the *Compass* recommendation engine used by Voyager Wongsuphasawat et al. (2016). This system feeds the question panel by adding a new variable from the unselected variables set. It generates recommendations through a set of predefined mappings between variable types, questions, and chart types. Table 4.7 presents some of the most commonly used mappings we have defined. In VisMaker, the variable names are highlighted according to the color that identifies their type, as shown earlier in Figure 4.6, which presents an example of the second recommended question for a combination of Q x N x T variable types.

Variable types	Question	Chart types
Q x Q	1 - What is the correlation between <u>var0</u> and <u>var1</u> ?	Scatterplot
Q x N	1 - What is the distribution of values of <u>var0</u> in each category of <u>var1</u> ?	Boxplot; Strip plot; Area chart
Q x O	2 - What is the average of <u>var0</u> in each category of <u>var1</u> ?	Bar chart
Q x T	1 - What is the <u>MEAN OF var0</u> in each category of <u>var1</u> ?	Line chart; Bar chart; Area chart
	1 - What is the <u>MEAN OF var0</u> over the <u>YEARS</u> ?	Line chart; Bar chart; Area chart
	2 - What is the <u>MEAN OF var0</u> over the <u>MONTHS</u> ?	Line chart; Bar chart; Area chart
N x N	1 - What is the number of co-occurrences between each category of <u>var0</u> and <u>var1</u> ?	Stacked bar chart; Normalized stacked bar chart; Heatmap
N x O	1 - What is the number of co-occurrences between each category of <u>var0</u> and <u>var1</u> ?	Stacked bar chart; Normalized stacked bar chart; Heatmap
N x T	1 - What is the number of occurrences of each category of <u>var0</u> over the <u>YEARS</u> ?	Line chart; Stacked area chart; Heatmap
O x O	2 - What is the number of occurrences of each category of <u>var0</u> over the <u>MONTHS</u> ?	Line chart; Stacked area chart; Heatmap
O x T	1 - What is the number of co-occurrences between each category of <u>var0</u> and <u>var1</u> ?	Stacked bar chart; Heatmap
T x T	2 - What is the number of occurrences of each category of <u>var0</u> over the <u>YEARS</u> ?	Stacked bar chart; Heatmap
Q x Q x Q	1 - What is the number of co-occurrences of <u>var0</u> and <u>var1</u> ?	Stacked bar chart; Heatmap
Q x Q x N	1 - What is the correlation between <u>var0</u> , <u>var1</u> and <u>var2</u> ?	Scatterplot + color; Scatterplot + size
Q x Q x O	1 - What is the correlation between <u>var0</u> , <u>var1</u> grouped by <u>var2</u> categories?	Scatterplot + color; Scatterplot + shape
Q x Q x T	1 - What is the correlation between <u>var0</u> , <u>var1</u> grouped by <u>var2</u> categories?	Scatterplot + color; Scatterplot + size
Q x Q x T	1 - What is the correlation between <u>MEAN OF var0</u> , <u>MEAN OF var1</u> over the <u>YEARS</u> ?	Line chart + size; Line chart + color
	2 - What is the correlation between <u>MEAN OF var0</u> , <u>MEAN OF var1</u> over the <u>MONTHS</u> ?	Line chart + size; Line chart + color
Q x N x N	1 - What is the <u>MEAN OF var0</u> in each combination of <u>var1</u> and <u>var2</u> ?	Heatmap
Q x N x O	1 - What is the <u>MEAN OF var0</u> in each combination of <u>var1</u> and <u>var2</u> ?	Line chart; Stacked area chart; Heatmap
Q x N x T	1 - What is the <u>MEAN OF var0</u> in each combination of <u>var1</u> over the <u>YEARS</u> ?	Line chart; Stacked area chart
Q x O x O	2 - What is the <u>MEAN OF var0</u> in each category of <u>var1</u> over the <u>MONTHS</u> ?	Line chart; Stacked area chart
Q x O x T	1 - What is the <u>MEAN OF var0</u> in each combination of <u>var1</u> and <u>var2</u> ?	Line chart; Heatmap
Q x T x T	1 - What is the <u>MEAN OF var0</u> in each combination of <u>var1</u> and <u>var2</u> ?	Line chart; Heatmap
Q x Q x Q x Q	1 - What is the correlation between <u>var0</u> , <u>var1</u> , <u>var2</u> and <u>var3</u> ?	Scatterplot + color + size
Q x Q x Q x N	1 - What is the correlation between <u>var0</u> , <u>var1</u> and <u>var2</u> in each category of <u>var3</u> ?	Scatterplot + size + color; Scatterplot + color + shape
Q x Q x Q x O	1 - What is the correlation between <u>var0</u> , <u>var1</u> and <u>var2</u> in each category of <u>var3</u> ?	Scatterplot + size + color;
Q x Q x Q x T	1 - What is the correlation between <u>MEAN OF var0</u> , <u>MEAN OF var1</u> and <u>MEAN OF var2</u> over the <u>YEARS</u> ?	Line chart + size + color
	2 - What is the correlation between <u>MEAN OF var0</u> , <u>MEAN OF var1</u> and <u>MEAN OF var2</u> over the <u>MONTHS</u> ?	Line chart + size + color
Q x Q x Q x N	1 - What is the correlation between <u>var0</u> , <u>var1</u> , <u>var2</u> and <u>var3</u> in each category of <u>var4</u> ?	Scatterplot + size + color + shape

Table 4.7: Mappings between data, questions and visualizations used in Vis-Maker

We stored these tables at the source code level using JSON files, through which we mapped all the combinations of variables to their respective questions and charts. For example, the combination of the types of variables Q x N is shown below:

```

1 "QN": [
2   {
3     "question": "What is the distribution of...",
4     "charts": [...]
5   },
6   {
7     "question": "What is the MEAN OF...",
8     "charts": [...]
9   }
10 ]

```

4.7

VisMaker's Recommender Engine

The VisMaker's recommender engine works similar to *Compass*, the recommender engine used by Voyager and presented in (Wongsuphasawat et al., 2016).

The goal of the Compass engine is to support rapid and open-ended data exploration. It generates a significant number of visualizations using Vega-lite grammar. Besides that, Compass also prunes the space of visualization recommendations based on the user selection of variables, clustering the remaining visualizations into meaningful groups.

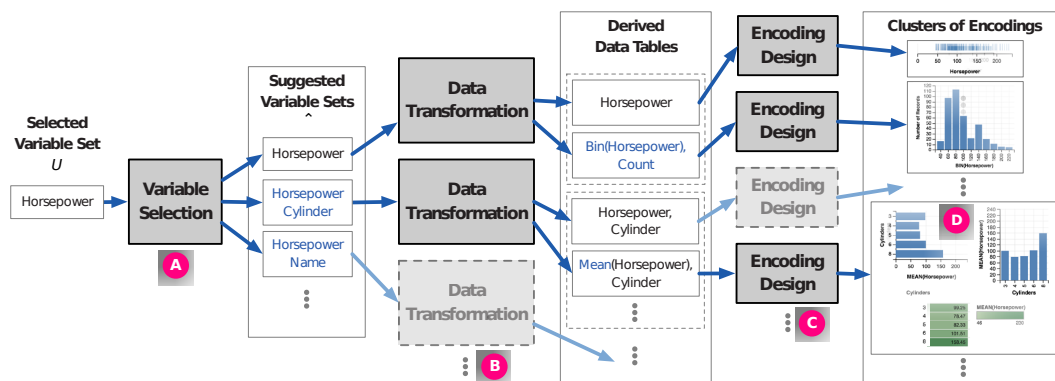


Figure 4.8: The Compass visualization recommendation engine Wongsuphasawat et al. (2016).

Inspired by the Compass recommendation engine for data exploration, the VisMaker recommendation system also performs the visualiza-

tion recommendation through the addition of unselected variables. Given a set of selected variables S , VisMaker iterates over the set of unselected variables $V = \{v_1, v_2, \dots, v_n\}$, generating a list of variable sets $\langle S \cup \{v_1\}, S \cup \{v_2\}, \dots, S \cup \{v_n\} \rangle$. The combinations of variable types are then mapped onto a list of recommended questions and visualizations as shown in Table 4.7.

Unlike Voyager 2, the VisMaker recommendation engine is “dettached” from the mapping already done by the user, i.e., VisMaker focuses only on the selected variables, but not on their mapping to channels. For instance, Vismaker does not care if the user has mapped variable A to the X-axis and will provide visualization recommendations by using variable A in several different mappings. In Voyager 2, the recommendations try to maintain the mapping that the user performed, adding new variables only to the free encoding fields.

Figure 4.9 presents the recommendations for the Iris dataset after the user has added the variables `species` and `petal_length` to X and Y axes in VisMaker (A) and in Voyager 2 (B). As we can see, the recommendations for the addition of the variable `petal_width` are quite different. While VisMaker recommends to use `species` in the color field, changing the mapping defined by the user, Voyager 2 suggests to add the `petal_width` variable to the color field, maintaining the previously defined mapping.

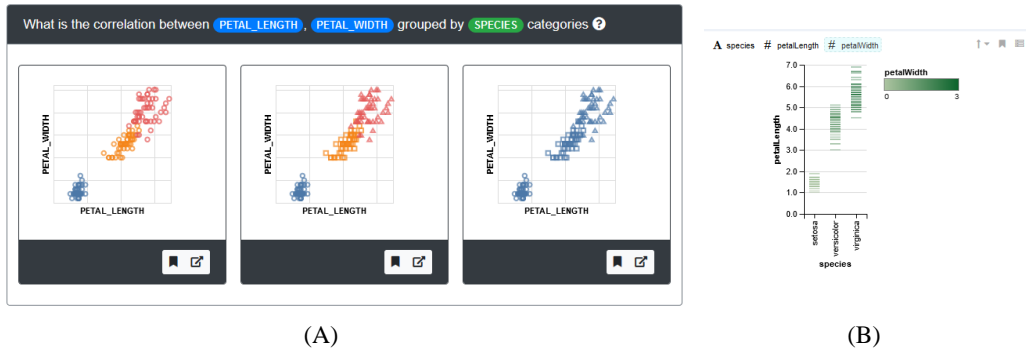


Figure 4.9: A comparison between Compass engine (Wongsuphasawat et al., 2016) and VisMaker engine.

5 Evaluation

We evaluated VisMaker aiming to answer some of our research subquestions, especially [SQ2.] *How do users interact with the questions presented by the tool via the recommendations?* and [SQ3.] *How can questions improve visualization recommendations?*

We compared VisMaker and Voyager 2 (Wongsuphasawat et al., 2017) through two different studies: Study 1 to evaluate the tools when they are used in a question answering scenario; and Study 2, designed from a limitation observed in Study 1, to evaluate them when they are used in a data exploration scenario. In each study, we divided participants into four groups according to a $\langle tool, task \rangle$ pair. Participants received two tasks, one to complete using each tool. They used the tools to construct visualizations to answer questions about data, in the case of Study 1; and to explore and gain knowledge about the data, in the case of Study 2.

Our studies involved 24 participants, all volunteers. Table 5.1 presents the distribution of participants in each study.

Study	Number of participants				
	Group 1	Group 2	Group 3	Group 4	Total
Study 1: Question Answering	4	4	4	4	16
Study 2: Data Exploration	2	2	2	2	8

Table 5.1: Participants per study.

In this chapter, we describe these two evaluation studies in detail, presenting the used datasets and the study procedures, including the tasks. We present more details about Study 1, as we reused much of its planning for Study 2.

5.1 Study 1: Question Answering

In this first study, which aimed to evaluate the use of tools in a data question answering scenario, participants were given tasks consisting of a specific dataset and a set of questions that they should answer through data visualizations. We did not set a time limit for the tasks. The session finished when all the questions had been answered by a participant.

5.1.1

Procedure

The comparison study has each participant use both VisMaker and Voyager 2 tools on different tasks. We obtained two distinct datasets and prepared, for each one, a list of questions that participants should answer through data visualizations. The study was then composed of two similar tasks, named CAPES-RJ and WEATHER, one for each dataset, which consisted of answering questions about the data using each tool freely.

We divided the participants into four distinct groups, varying the $\langle tool, task \rangle$ pair and the order in which each tool would be used. The groups were established as follows.

- **Group 1:** start using VisMaker with CAPES-RJ task followed by Voyager 2 with WEATHER task;
- **Group 2:** start using VisMaker with WEATHER task followed by Voyager 2 with CAPES-RJ task;
- **Group 3:** start using Voyager 2 with CAPES-RJ task followed by VisMaker with WEATHER task; and
- **Group 4:** start using Voyager 2 with WEATHER task followed by VisMaker with CAPES-RJ task.

As the focus of this study was not on the discoverability of the tools, we introduced participants to the use of the tools before using them. We did so by showing two video tutorials. We produced two 10-minute videos, one for each tool, using the same dataset and constructing the same visualizations to introduce the basic usage of the tools. These videos were produced in Portuguese to facilitate the understanding of the participants, all Brazilians. We published both videos on YouTube through the links: <https://youtu.be/B59hV5GG9vc> and <https://youtu.be/rTayae9gPBk>. Besides that, participants were encouraged to think-aloud during the task execution and ask questions whenever doubts about the tool arose.

We conducted a pilot study to get a sense of how long each session would take to complete, and we verified that it would take approximately 90 minutes. Next, we performed four preliminary experiments, with one participant per group. Through this preliminary study, we felt the need to change the tasks by encouraging participants to use the recommendations provided by the tools, as the participants were ignoring them.

5.1.2 Study Execution

For the execution of this study, we built an agenda through which participants could schedule a time for their session. All sessions were conducted within the university environment, using a Lenovo[®] notebook with Intel[®] Core[™] i5 processor and 8GB of RAM.

We conducted this study with 16 participants. During each session, we captured participants' spontaneous remarks and computer screen using a screen recorder software. This data is confidential and was used in the process of result analysis.

We performed each session of this experiment as follows:

1. The participant read the informed consent form, and completed the profile questionnaire (see Appendix D) by answering some questions such as the field of study and how often they read or created data visualizations.
2. The participant watched a video tutorial about the first tool they would use;
3. The participant used the first tool to complete the assigned task, according to their group;
4. The participant completed the first post-task questionnaire;
5. The participant watched a video tutorial about the second tool they would use;
6. The participant used the second tool to complete the assigned task, according to their group;
7. The participant completed the second post-task questionnaire; and, finally,
8. The participant was interviewed about using both tools to complete each task.

These steps are illustrated by Figure 5.1.

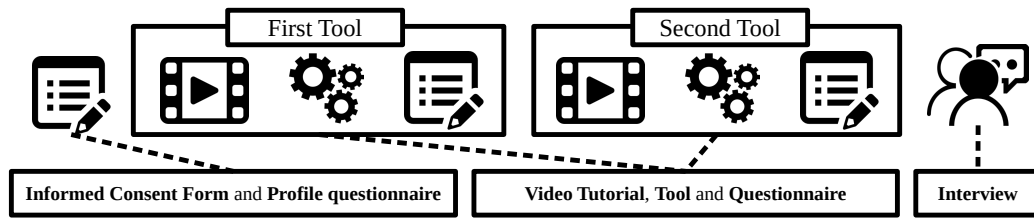


Figure 5.1: Steps of experiment sections.

In the next section, we present the details of each task used to conduct the study, describing its datasets and questions.

5.1.3 Tasks

In this section, we detail the two tasks used in this first experiment, CAPES-RJ and WEATHER.

5.1.3.1 The CAPES-RJ Task

Task CAPES-RJ consists of answering a set of questions about a dataset of 4,993 records of graduate programs distributed in different cities of the state of Rio de Janeiro over the years. This dataset has seven variables, one of which is temporal, three are nominal, and three are quantitative. They are presented in Table 5.2.

Variable	Type	Description
YEAR	Temporal	The year the program was registered
CITY	Nominal	The city of the graduate program
FIELD	Nominal	The field of study of the graduate program
JURIDICAL STATUS	Nominal	The juridical status of the institution of the program
QNT MASTERS	Quantitative	The number of masters scholarships the program had when the registration was made
QNT DOCTORAL	Quantitative	The number of doctoral scholarships the program had when the registration was made
QNT POSTDOCTORAL	Quantitative	The number of postdoctoral scholarships the program had when the registration was made

Table 5.2: CAPES-RJ variables

We elaborated a set of eight questions, of which the first four could be answered freely using one of the tools (VisMaker or Voyager 2), and the last four, elaborated in a table format, should be completed by the participants using mainly the recommendations provided by the tool. The questionnaire of this task can be found in Appendix B.

5.1.3.2

The WEATHER Task

The WEATHER task is similar to the CAPES-RJ task and consists of answering a set of questions about a dataset of 2922 weather records in the cities of New York and Seattle, in the United States. This dataset has seven variables, one of which is temporal, two are nominal, and four are quantitative. They are presented in Table 5.3.

Variable	Type	Description
DATE	Temporal	The date the registration was made
LOCATION	Nominal	The city in which the registration was made
WEATHER	Nominal	A categorization of the weather that was being recorded
PRECIPITATION	Quantitative	A quantitative value for measuring precipitation
WIND	Quantitative	A quantitative value for measuring wind
TEMP_MAX	Quantitative	The maximum temperature in °C
TEMP_MIN	Quantitative	The minimum temperature in °C

Table 5.3: WEATHER variables.

As with task CAPES-RJ, we elaborated a set of eight questions, the first four to be answered freely using one of the tools (VisMaker or Voyager 2), and the last four, also elaborated in a table format, to be completed using mainly the recommendations provided by the tool. The complete questionnaire of this task can be found in Appendix C.

5.1.4

Post-task Questionnaire

After completing each task using each of the tools according to their group, participants had to answer a questionnaire. We elaborate a questionnaire to evaluate the perceived ease of use and utility based on the Technology Acceptance Model (TAM) (Davis, 1989).

Below we list all the questions in it respective questionnaire sessions. The full questionnaire can be found in Appendix E and F.

– **Ease of Use**

1. Q01 - I think the tool ease to use
2. Q02 - Learning how to use the tool was easy for me
3. Q03 - It is easy to become skillful using the tool

– **Utility**

1. Q04 - The tool makes it ease the construction of data visualizations
2. Q05 - The tool makes it more efficient the process of exploring datasets
3. Q06 - The tool helps me answer questions through the data

– **Attitude**

1. Q07 - Using the tool to explore datasets is a good idea
2. Q08 - Using the tool to help me create data visualizations is a good idea
3. Q09 - Using the tool to help me answer questions about a particular data set is a good idea

– **Self-efficacy**

1. Q10 - I felt confident using the tool
2. Q11 - I have the necessary skills to use the tool

5.1.5 Participants

As we showed earlier in Table 5.1, we conducted this study with 16 participants. At the beginning of each session, we asked each participant to fill out a form answering some questions that we considered necessary for characterizing their profile. Participants used a 7-point Likert scale, to response the following items:

1. **P01:** *I commonly perform data exploration tasks*
2. **P02:** *I commonly build data visualizations*
3. **P03:** *I can understand well several types of charts*

Figure 5.2 presents the distribution of the participants' answers for these questions in Study 1.

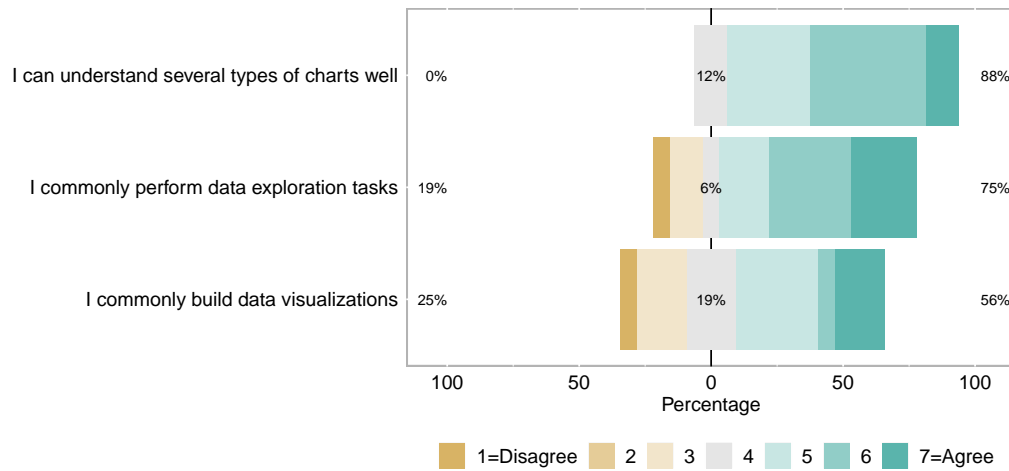


Figure 5.2: Profile questionnaire responses of Study 1.

Table 5.4 summarizes the characteristics of Study 1 participants.

5.1.6 Study 1 Findings

In this section, we present the main findings of this first study. We start presenting a little bit about the responses that the participants gave to the forms, we comment on the correctness of the tasks and the visualizations generated through the use of the tools. Finally, we present the main feedbacks given by the participants during the interview.

Participant	Sex	Degree	Field of study
P01	Male	Master's student	Data Science
P02	Female	PhD student	Data Science, Human-Computer Interaction
P03	Male	Master's student	Data Science, Optimization
P04	Male	Master's student	Data Science, Optimization
P05	Male	PhD student	Data Science, Human-Computer Interaction
P06	Female	PhD student	Data Science, Internet of Things
P07	Male	Master's student	Data Science
P08	Male	PhD student	Computer Graphics
P09	Male	PhD	Data Science, Hypertext and Hypermedia, Human-Computer Interaction, Computer Networks & Distributed Systems
P10	Female	Master's student	Psychology
P11	Male	PhD student	Software Engineering
P12	Female	PhD student	Optimization
P13	Male	Master's student	Data Bases, Data Science
P14	Male	PhD student	Data Science, Hypertext and Hypermedia, Human-Computer Interaction
P15	Male	PhD student	Data Science
P16	Male	Master's degree	Hypertext and Hypermedia

Table 5.4: Characteristics of the participants in the first study.

5.1.6.1

Questionnaire Responses

Through the questionnaires answered by the participants after using each of the tools, we were able to gather some opinions about some aspects of the tools. Figure 5.3 shows the distribution of the participants' answers to the 11 questions present in the questionnaires for each tool. Participants answered each question using a 7-level Likert scale.

We used the Wilcoxon-Mann-Whitney (Wilcoxon, 1992) test to verify the statistical significance of the answers for each tool, verifying that, although Vis-Maker received greater acceptance in most questions, the differences between the tools were only statistically significant ($p\text{-value} \geq 0.05$) for Q01, Q02, Q03, Q07 and Q10. Table 5.5 presents the corresponding p -values, and the medians of scores of each tool in each question.

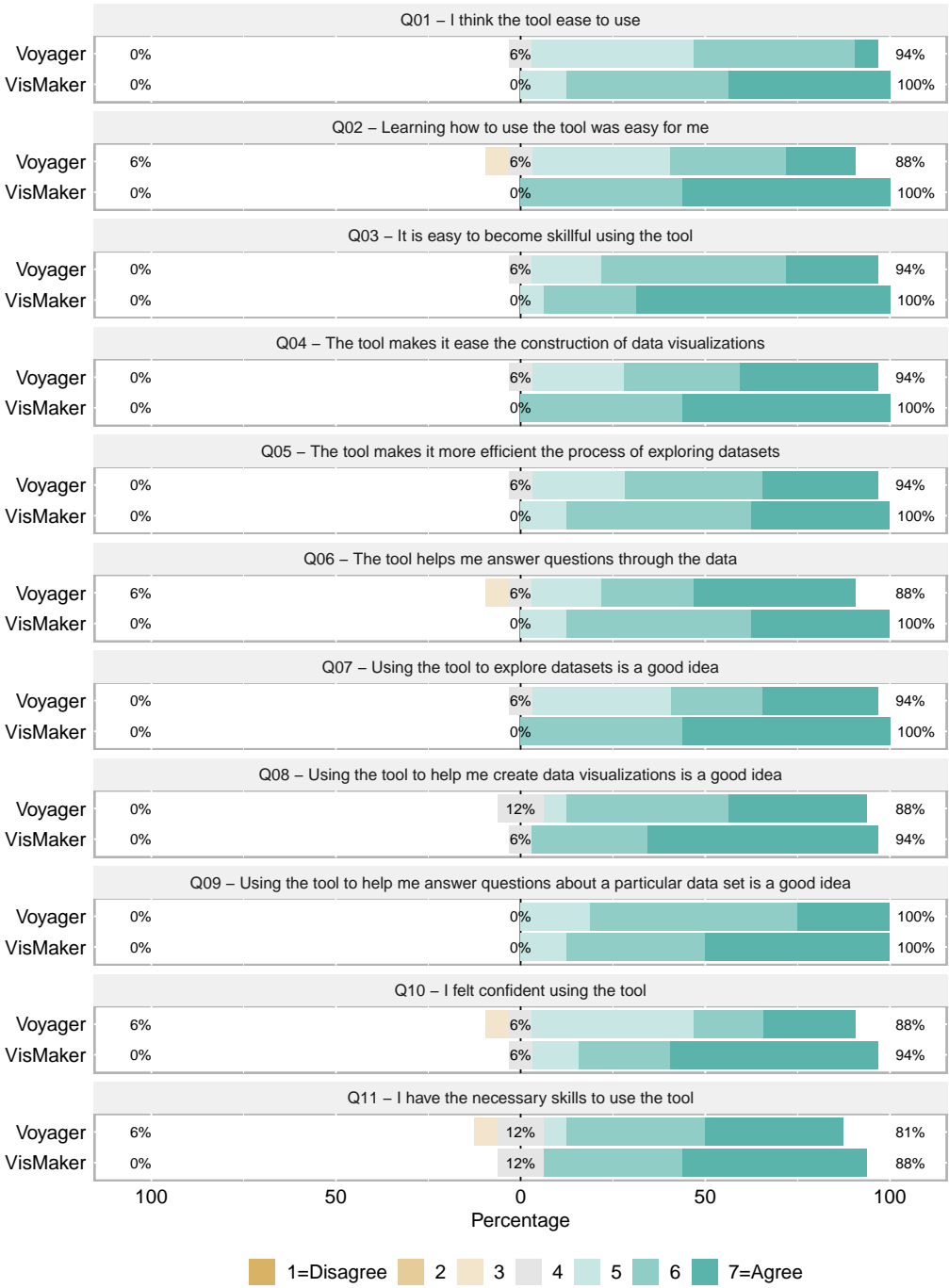


Figure 5.3: Questionnaire responses of Study 1.

Question	VisMaker median	Voyager median	<i>p</i> -value
Q01	6.0	5.5	0.005*
Q02	7.0	5.5	0.002*
Q03	7.0	6.0	0.013*
Q04	7.0	6.0	0.092
Q05	6.0	6.0	0.363
Q06	6.0	6.0	0.702
Q07	7.0	6.0	0.023*
Q08	7.0	6.0	0.145
Q09	6.5	6.0	0.195
Q10	7.0	5.0	0.038
Q11	6.5	6.0	0.382

Table 5.5: Statistical significance of questionnaire responses in Study 1

5.1.6.2
Tasks' Correctness

Regarding task correctness, we found four possible cases: (i) those who built correct visualizations and responded correctly in the form; (ii) those who built correct visualizations but provided an incorrect answer in the form; (iii) those who built incorrect visualizations but answered correctly in the form; and (iv) those who built incorrect visualizations and provided an incorrect answer in the form.

Tables 5.6 and 5.7 show the correctness of the participants' visualizations and ansers in the forms, respectively. The tables highlight the correct use of VisMaker in gray, and Voyager in blue. Errors are highlighted in red.

Group 1					Group 2					Group 3					Group 4				
Task	P01	P02	P03	P04	Task	P05	P06	P07	P08	Task	P09	P10	P11	P12	Task	P13	P14	P15	P16
CAPES-RJ	T1				WEATHER	T1				CAPES-RJ	T1				WEATHER	T1			
	T2					T2					T2					T2			
	T3					T3					T3					T3			
	T4					T4					T4					T4			
	T5					T5					T5					T5			
	T6					T6					T6					T6			
	T7					T7					T7					T7			
	T8					T8					T8					T8			
WEATHER	T1				CAPES-RJ	T1				WEATHER	T1				CAPES-RJ	T1			
	T2					T2					T2					T2			
	T3					T3					T3					T3			
	T4					T4					T4					T4			
	T5					T5					T5					T5			
	T6					T6					T6					T6			
	T7					T7					T7					T7			
	T8					T8					T8					T8			

Table 5.6: Correctness of the constructed visualizations in Study 1

The errors presented in Table 5.6 are the following, indicated by participant:

- **P04:**
 - **WEATHER - T4 - Using Voyager:** The participant forgot to apply a filter to view only Seattle city data.
- **P05:**
 - **WEATHER - T3 - Using VisMaker:** The participant understood the question differently, as if the date had to be taken into account. This made them write the answer “NY (2014)”.
- **P07:**
 - **CAPES-RJ - T1 - Using Voyager:** The participant ignored the year specified in the question.
 - **CAPES-RJ - T3 - Using Voyager:** The participant used the variable QNT DOCTORAL instead of QNT POSTDOCTORAL.
 - **CAPES-RJ - T7 - Using Voyager:** The participant used the value of SUM OF QNT MASTERS instead of COUNT.
- **P08:**
 - **WEATHER - T1 - Using VisMaker:** The participant unnecessarily used the WEATHER variable to answer the question and made a mistake.
 - **CAPES-RJ - T4 - Using Voyager:** The participant used the MAX function instead of SUM in the QNT MASTERS variable and made a mistake.
 - **CAPES-RJ - T8 - Using Voyager:** The participant was unable to add the QNT MASTERS variable to the visualization, adding it incorrectly in the filter field.
- **P13:**
 - **WEATHER - T3 - Using Voyager:** The participant used the Q2 chart to answer this question. The visualization used did not allow them to complete their answer because it presented the average values of rain precipitation distributed over the months.
 - **CAPES-RJ - T3 - Using VisMaker:** The participant used the MAX function instead of SUM in the QNT POSTDOCTORAL variable and made a mistake.

- **CAPES-RJ - T5 - Using VisMaker:** The participant even built a correct visualization, but removed the FIELD variable from the mapping and did not add it again.
- **P14:**
 - **WEATHER - T3 - Using Voyager:** The participant did not apply the MEAN function to the PRECIPITATION variable and constructed a wrong view about this question.

Group 1					Group 2					Task					Group 4				
Task	P01	P02	P03	P04	Task	P05	P06	P07	P08	Task	P09	P10	P11	P12	Task	P13	P14	P15	P16
CAPES-RJ	T1				WEATHER	T1				CAPES-RJ	T1				WEATHER	T1			
	T2					T2					T2					T2			
	T3					T3					T3					T3			
	T4					T4					T4					T4			
	T5					T5					T5					T5			
	T6					T6					T6					T6			
	T7					T7					T7					T7			
	T8					T8					T8					T8			
WEATHER	T1				CAPES-RJ	T1				WEATHER	T1				CAPES-RJ	T1			
	T2					T2					T2					T2			
	T3					T3					T3					T3			
	T4					T4					T4					T4			
	T5					T5					T5					T5			
	T6					T6					T6					T6			
	T7					T7					T7					T7			
	T8					T8					T8					T8			

Table 5.7: Correctness of the written answers in Study 1

The errors presented in Table 5.7 are the following, indicated by participant:

- **P01:**
 - **CAPES-RJ - T1 - Using VisMaker:** The participant correctly applied a filter to the YEAR variable, but a bug in the tool caused it to display the wrong value.
 - **WEATHER - T2 - Using Voyager:** A problem with the scale of the chart caused a mistake.
- **P03:**
 - **CAPES-RJ - T5 - Using VisMaker:** The participant was distracted and ended up changing the name of the areas for the value of the counter.
 - **WEATHER - T2 - Using Voyager:** A problem with the scale of the chart caused a mistake.
 - **WEATHER - T8 - Using Voyager:** A problem with the scale of the chart caused a mistake.

- **P04:**
 - **WEATHER - T2 - Using Voyager:** A problem with the scale of the chart caused a mistake.
 - **WEATHER - T4 - Using Voyager:** The participant forgot to apply a filter to view only Seattle city data.
- **P05:**
 - **WEATHER - T2 - Using VisMaker:** The participant was distracted and eventually forgot to write the month of July.
 - **WEATHER - T3 - Using VisMaker:** The participant understood the question differently, as if the date had to be taken into account. This made him write the answer “NY (2014)”.
 - **WEATHER - T2 - Using Voyager:** A problem with the scale of the chart caused a mistake.
 - **WEATHER - T4 - Using Voyager:** The participant was confused when writing the answer, noting the value of COUNT instead of the SUM OF QNT MASTERS.
- **P07:**
 - **CAPES-RJ - T1 - Using Voyager:** The participant ignored the year specified in the question.
 - **CAPES-RJ - T5 - Using Voyager:** The participant was distracted and ended up missing one of the table cells.
 - **CAPES-RJ - T6 - Using Voyager:** The participant constructed the correct visualization, but made a mistake in analyzing it.
- **P08:**
 - **WEATHER - T1 - Using VisMaker:** The participant unnecessarily used the WEATHER variable to answer the question and made a mistake.
 - **CAPES-RJ - T4 - Using Voyager:** The participant used the MAX function instead of SUM in the QNT MASTERS variable and made a mistake.
 - **CAPES-RJ - T8 - Using Voyager:** The participant was unable to add the QNT MASTERS variable to the visualization, adding it wrong in the filter field.
- **P13:**
 - **WEATHER - T2 - Using Voyager:** A problem with the scale of the chart caused a mistake.

- **WEATHER - T3 - Using Voyager:** The participant used the Q2 chart to answer this question. The visualization used does not allow him to complete his answer because it presents the average values of rain precipitation distributed over the months.
 - **CAPES-RJ - T5 - Using VisMaker:** The participant even built a correct visualization, but removed the FIELD variable from the mapping and did not add it again.
 - **CAPES-RJ - T6 - Using VisMaker:** The participant constructed the visualization correctly, but an instability in the tool caused the chart to appear distorted, causing the participant to make the mistake.
- **P14:**
- **WEATHER - T3 - Using Voyager:** The participant did not apply the MEAN function to the PRECIPITATION variable and constructed a wrong visualization for this question.
 - **CAPES-RJ - T1 - Using VisMaker:** The participant correctly applied a filter to the YEAR variable, but a bug in the tool caused it to display the wrong value.
- **P15:**
- **WEATHER - T2 - Using Voyager:** A problem with the scale of the chart caused a mistake.
- **P16:**
- **WEATHER - T2 - Using Voyager:** A problem with the scale of the chart caused a mistake.
 - **CAPES-RJ - T1 - Using VisMaker:** The participant correctly applied a filter to the YEAR variable, but a bug in the tool caused it to display the wrong value.

We grouped all the errors found in eight categories. Figure 5.4 presets a heatmap showing the occurrences of these errors in both tools

As we can see, the errors obtained through the Voyager tool are mainly related to problems with the scale of the generated charts and errors in the use of variables. Most of the problems with the charts scale occurred during the resolution of the second question (T2) of the WEATHER task. Many of the participants in groups 1 and 4, who used Voyager to solve the WEATHER task, built tiny scale visualizations that made it impossible to compare values,

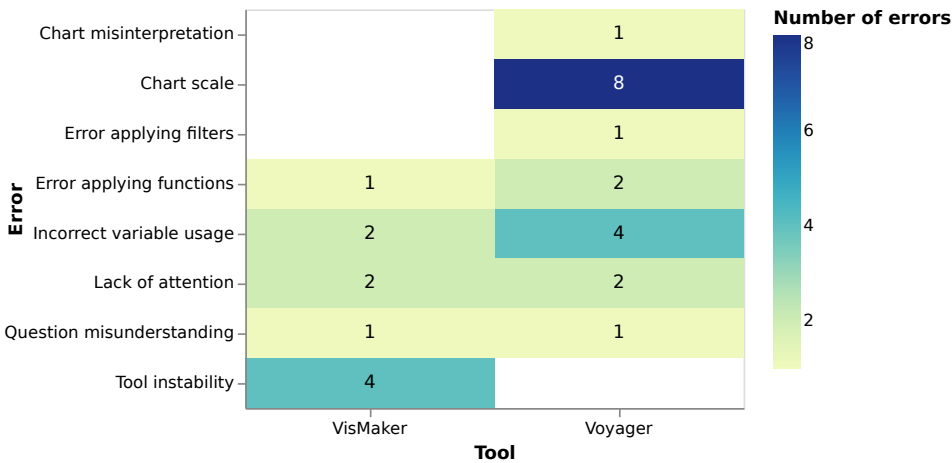


Figure 5.4: Errors per tools in Study 1.

which made them err the answer, ignoring the month of September. Figure 5.5 presents a typical chart created by Voyager users to solve WEATHER Q2.

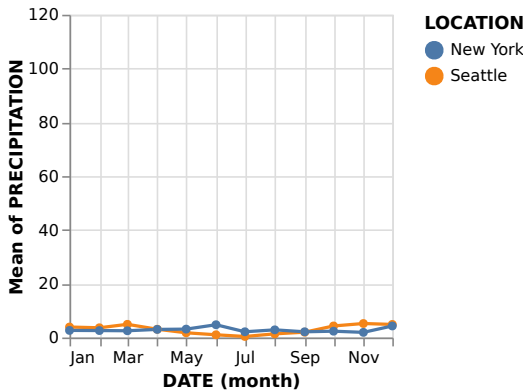


Figure 5.5: A typical chart constructed through Voyager to solve WEATHER T2 in Study 1.

Participants who used VisMaker for the same task found a chart with better scale and made no such mistake. Figure 5.6 presents this chart.

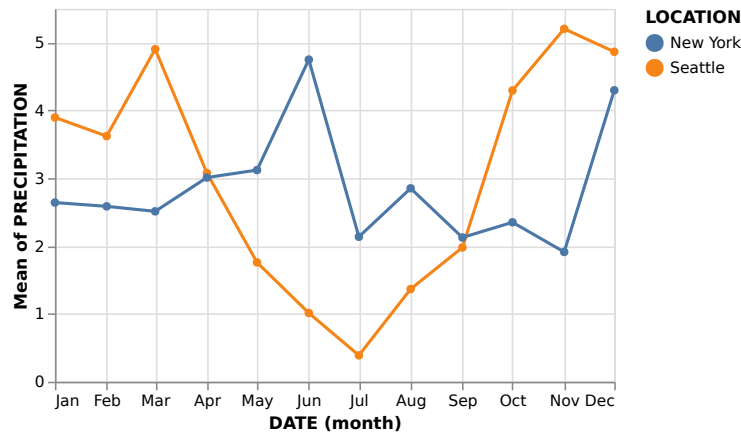


Figure 5.6: A typical chart constructed through VisMaker to solve WEATHER T2 in Study 1.

In VisMaker, some errors resulted from apparent instability in the tool. In these cases, participants constructed a visualization correctly, correctly mapping variables onto visual dimensions, but obtaining “distorted” visualizations. These errors occurred mainly in first question (T1) of the CAPES-RJ task by participants P01, P14, and P16. In these cases, participants used the DATE variable to filter only the year 2008 required by the question, and the tool presented a value considered wrong. Figure 5.7 presents a screenshot of the moment when participant P14, despite having mapped correctly, gets a visualization with the wrong answer (2,739 instead of 2,353) for CAPES-RJ T1.

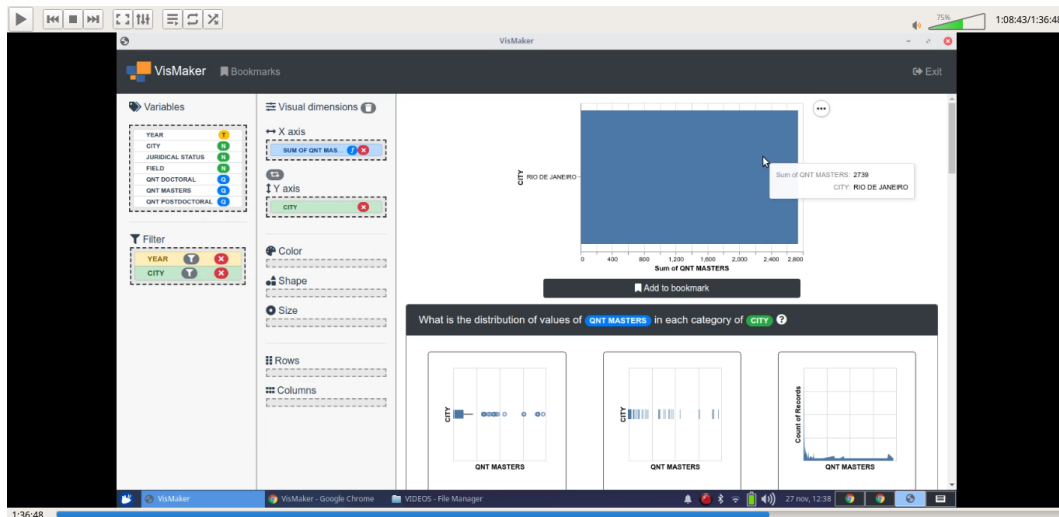


Figure 5.7: The moment when participant P14 builds a correct view but finds the wrong answer for CAPES-RJ T1 in Study 1.

Figure 5.8 presents two charts equally constructed through VisMaker (A) and Voyager (B). We can note that, although the visualizations are quite similar, the value shown in the tooltip is different in both tools, evidencing the reported bug in VisMaker.

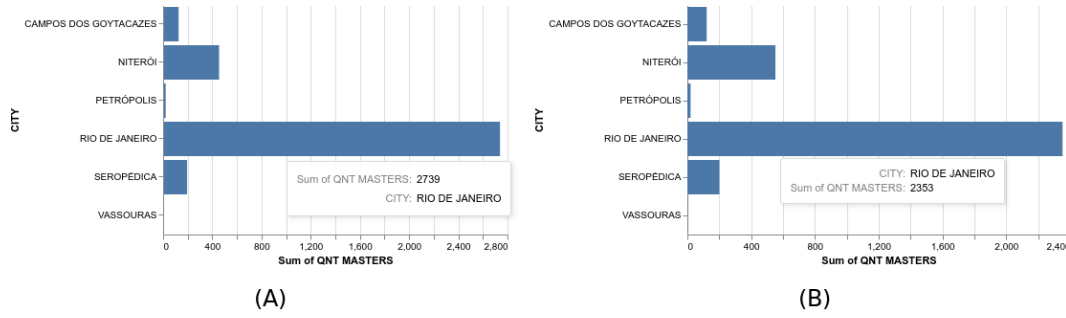
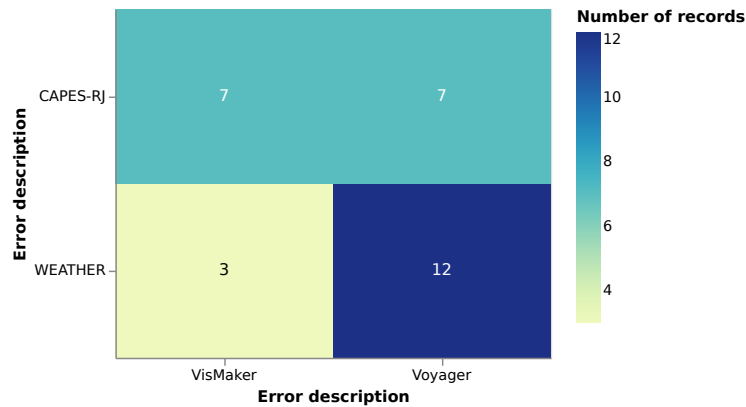


Figure 5.8: The temporal filter bug in VisMaker.

Regarding the errors that occurred in the tasks with the use of each tool, we can verify through Figure 5.9 that the participants who used VisMaker made more mistakes to perform the CAPES-RJ task. In contrast, the participants who used the Voyager made more mistakes while performing the WEATHER task. As stated earlier, most of these errors concern the filtering problem using time variables that commonly occurred in the CAPES-RJ task with VisMaker; and the scale of the charts generated through Voyager, which occurred mainly in the WEATHER task.

Figure 5.9: Errors per $\langle dataset, tool \rangle$ in Study 1.

5.1.6.3

Generated Charts

When we look at the charts that the study participants built to answer the questions, we come across some findings on how they used the tools and handled the recommended visualizations.

While some participants read the questionnaire and thought first about which filters they should apply, others first thought about which values they wanted to discover and only then which variables they should add to the mapping; only later and, if necessary, they applied filters. An example of this difference in use can be seen by comparing the views generated by participants

P14 and P13 in the first question of the CAPES-RJ task, presented by Figures 5.10 (A) and (B), respectively.

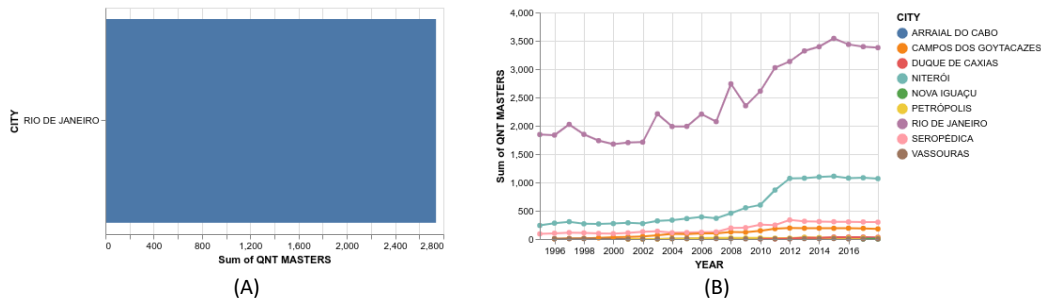


Figure 5.10: Examples of charts constructed to solve CAPES-RJ T1.

On this same question, some participants started building the visualization by inserting the variables `CITY` and `QNT MASTERS` on the X and Y axes. In these cases, the participants were faced with a doubt: “*And now? Where will I add the variable YEAR?*” Recommended by Voyager, participants P10 and P12 added the variable `YEAR` in the `COLOR` field, obtaining a visualization that, although correct, can be considered inefficient. Participant P15, in turn, did it manually through VisMaker. All of these participants, after obtaining these charts, did not look at recommendations or the alternative encodings panel. The views generated by these participants for this question are shown in Figure 5.11.

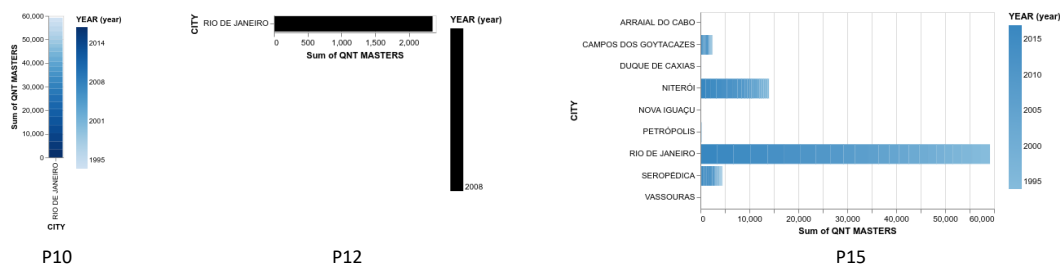


Figure 5.11: Examples of charts constructed to solve CAPES-RJ Q1.

Regarding the use of the alternative encodings panel by Voyager users, it was possible to observe that many participants used a type of bubble-based heatmap (here identified as “*bubble matrix*”) that did not satisfy them very much. These charts were constructed to fill the questions in the form of a table. If we look at the views built to fill the first table of the CAPES-RJ task, for example, we see that three types of views were most commonly built. They are shown in Figure 5.12.

In general, participants who were not satisfied with using the “bubble map” (A) looked for other visualizations and found them through the alternative encodings panel, usually a stacked bar chart (B). VisMaker users, in

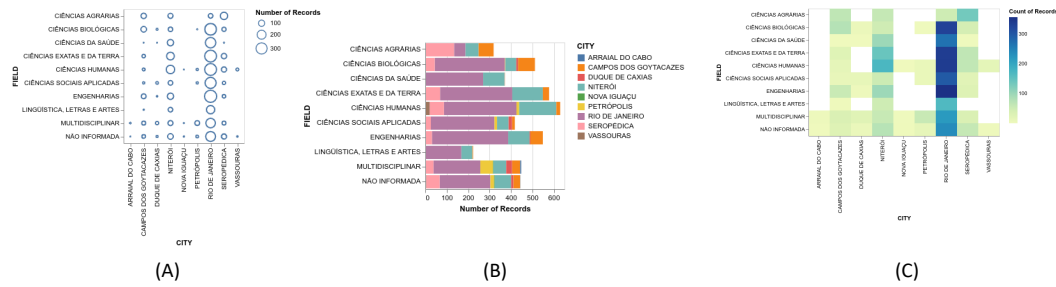


Figure 5.12: Examples of charts constructed to solve CAPES-RJ T1 in Study 1.

turn, in addition to stacked bar charts, built many heatmaps (C) and have not shown much rejection of this visualization type.

5.1.6.4

Participants' Feedback about the Tools

During the interviews, we asked two questions related to the recommendations: the first one explicitly asked about the recommendation approaches expressed through the recommended visualization panel; the second one asked about the charts obtained through both tools, if they corresponded in any way to the participants' expectations. Below we list the main feedbacks given by the participants.

1. *Considering the recommendation approaches, which tool most helped you complete the task?*

- P01: *“I found the two tools very similar in terms of recommendations. Still, for me, **VisMaker was more intuitive** because it presented a written question, and it helped me identify what I was looking for. In Voyager, I saw the recommendations and had to think a little. **There was a slight difference, but I preferred VisMaker recommendations.**”*
- P02: *“I felt **more confident with Voyager**, perhaps because **it was the second tool I used**. Although the charts were worse to analyze, I thought that building them was more comfortable with Voyager. As the questionnaire of the task was guiding me, I always looked for the variables. **In the case of VisMaker, which presented an entire question, I had more difficulty validating the variables I was looking for.** I think if I were performing an exploration without the predefined questions, it would not have happened. The coloring of the variables by type also bothered me because sometimes I thought that nothing had changed, but in fact, another variable was being used, only the same type.*

Perhaps if there were a single color per variable, it would be easier to identify.”

- **P03:** *“**Voyager helped me more.** I did well on it because there were so many recommendations, and **despite having a more cluttered interface** with more information on the screen, I just needed to find one of the visualizations that were being recommended to answer the question.”*
- **P04:** *“**I hardly used the recommendations.** I found it pretty much the same, but in Voyager I couldn’t complete the task without the recommendations. It was very useful.”*
- **P05:** *“After using VisMake, I missed the recommended questions in Voyager because I needed to interpret the charts as it only informed me of the variables he was using to build them. The way **VisMaker** presents recommendations is a summary of what visualization is, and it **seemed to me somewhat friendlier**, gave it a certain consistency that was the point I missed most in Voyager. Perhaps if you were more familiar with datasets, this confusion would not exist. VisMaker helped me a little more because **I just read the question and thought: ‘Oh, it must be one of those charts here!’**”*
- **P06:** *“**VisMaker was easier than Voyager.** Voyager had **higher information density, which made it a little complicated.** During the experiment, I made more use of VisMaker recommendations. Having the question helps, but I looked straight at the charts... When I had some difficulty, I used the questions to confirm the meaning of the charts.”*
- **P07:** *“**VisMaker helped me more** because, instead of looking for information inside the chart, I was looking for questions that resembled the question I wanted to answer. In Voyager, I had to read the chart and understand what they were reporting to know if it answered the question I had or not.”*
- **P08:** *“**The two were similar, but I preferred VisMaker** because I found its way to present a little better. [...] **The fact that VisMaker presents the questions helps**, so I liked it more.”*
- **P09:** *“**VisMaker helped because of the recommendation questions.** The questions on the questionnaire were similar to the questions presented by the tool. I think the simple fact of listing alternative possibilities with greater emphasis on variables would help in the interpretation of the charts. Perhaps the questions are not really necessary.”*

- P10: *“In general, I liked Voyager, but the VisMaker recommendation approach was better for me because it presented the questions. [...] I think VisMaker is better when I need to use many variables because the more variables I have to use, the more I have to keep asking the question I want to answer in my mind. How VisMaker presents the questions it makes it easier.”*
- P11: *“VisMaker helped me more because it already gave a better explanation of why it was using certain variables in the form of questions. Then I could go directly to the visualizations or change it to get the desired answer.”*
- P12: *“I think VisMaker helped me more because it made the chart easier to understand, it was easier.”*
- P13: *“In VisMaker, sometimes, I had to clear the mappings and start over to see if the recommendations helped me get where I wanted. I think Voyager bothered me less.”*
- P14: *“I found the recommendations of both tools quite similar, but the presentation of VisMaker is much better because the variables are highlighted by color. But I think the efficiency of the two tools is pretty much the same.”*
- P15: *“In both tools, I had questions answered by the recommendations, but in VisMaker, having a question associated with the visualization increased my confidence about what the visualization answered.”*
- P16: *“VisMaker was more specific in recommendations. Voyager shows you everything, and you don’t have a categorization or a specialization of what is being recommended... VisMaker was more specific in recommendations, and they were more accurate.”*

2. *As you read the questions, did you imagine any charts you could use to answer it? Did any of the tools better match your expectations?*

- P01: *“The charts did not always match. Voyager featured some visualizations I didn’t expect, for example, those charts that showed the amount by size of circles. In general, they were just some types of visualizations that I was unfamiliar with. VisMaker presented slightly better visualizations, as I was more familiar with them.”*

- **P03:** “In both tools I got results that roughly matched what I imagined.”
- **P04:** “Yes, I imagined some chart types. In general I could only imagine bars. But **it was great when I got heatmap visualizations**, for example.”
- **P05:** “Not always. A clear example was the last chart I built that I kept looking at a bunch of little balls. And that happened a lot in this dataset. Since the Voyager alternative encodings panel was down there, I didn’t see them and ended up using that kind of chart. I found this choice of Voyager strange. In VisMaker, I don’t remember seeing anything strange.”
- **P07:** “Yes, I imagined some visualizations I already knew, especially the most straightforward visualizations, like bar charts, for example. **VisMaker matched more. Voyager recommended me, but the first option was not so good, and I always looked for another view in alternative encodings.**”
- **P08:** “I imagined some bar charts. I think the two tools matched, but **Voyager presented some slightly stranger charts.**”
- **P11:** “Yes, I previewed some types of graphs, but I thought more about which axis I should insert each variable inside. **In Voyager, the recommendations did not match so much.** For example, I never thought of building those ball charts. **I found the VisMaker charts more natural, as I think they corresponded to the use of variables more coherently.**”
- **P12:** “In some questions, yes. Some charts I never used. In Voyager, I was a little doubt about which axis to use for the variables. **VisMaker usually put an axis for counting and that help me.**”
- **P14:** “At first, the charts didn’t match much because they presented me with the distribution of quantitative variables, and I was interested in the average or the sum. **In general, VisMaker matched more.**”
- **P15:** “Yes. I think I answered some of those questions with charts that I wouldn’t build, but as they answered the questions, I used them.”
- **P16:** “I didn’t preview many types of charts. I was trying to think more in terms of which axes I would use to map the variables. So when I was using it, I would read the question, see which variables were being measured by the question, and already trying to map to generate a visualization. **In Voyager, I had more doubts. In VisMaker, the visualizations were better.**”

5.1.7

Discussions

From Study 1, we can see that both tools enable and simplify the construction of statistical data visualizations for this purpose. With a question in mind, participants quickly identified the variables they would need to use and mapped out the necessary variables and the different viewing channels.

In general, participants experienced some difficulty while using the tools and made mistakes during the construction of the visualizations or even during the interpretation process of the constructed charts. Nonetheless, by observing the errors made by the participants, we can note that, in general, the occurrence of errors seems to be related more more to the lack of attention or haste to complete the tasks proposed in the experiment than to problems in the tools themselves. The evidence for this observation is that a participant who made a certain mistake in a particular question answered another similar question of equal or greater difficulty without making the same mistake.

Errors arising from the misinterpretation of questions, inattention when writing answers on the form, and inattention to the use of tools should not, however, take our focus off of the positive aspects and points that each tool needs to improve.

Looking into the difficulties participants had while using Voyager, we see the problem of chart scales, which caused many of them to make mistakes. Besides that, we can see through participants' feedback how difficult it was for users to find charts that met their expectations. Participant P11, for example, said: *"I never thought of building those ball charts"*, referring to the graphics obtained through Voyager. Also referring to the correspondence and familiarity with the obtained charts, P05 stated: *"Since the Voyager alternative encodings panel was down there, I didn't see them and ended up using that kind of chart"*. Some participants also commented that they found the buttons too small and that this hindered the interaction. Others commented on the lack of interactivity resources for the visualizations, especially when disturbed by the scale problem.

Regarding the problems reported by VisMaker users, they commented on the lack of feedback in the "add to bookmark" button, and the lack of interactivity in the main chart.

Regarding the recommendation approaches, Voyager has the characteristic of providing recommendations always respecting the mapping already done by the user. Thus, if the user mapped variables A and B to the X-axis and color field, respectively, Voyager recommendations will keep this mapping, proposing visualizations that will also have variable A on the X-axis and variable B in

the color field. At first, this approach can be beneficial when the user already has some experience and performs a good initial mapping. When, however, the user starts building a visualization by making a wrong choice for the mapping, all other recommendations will be affected by that choice, and the participant will have to redo the mapping or resort to the alternative encodings panel to build more efficient visualizations.

When we look into VisMaker and its distinguishing characteristic – the use of questions to facilitate and make the visualization recommendation process more effective –, we can see that the questions were beneficial to some participants but a hindrance to others. The fact that VisMaker is “detached” from the mapping already done by the user makes the charts obtained through the recommendations always similar, since they follow the mapping defined by the tool’s recommendation engine (presented earlier in Table 4.7).

Looking for the answers that users gave to the questionnaire about the tools, we see that VisMaker was significantly superior (at the significance level $\alpha = 0.05$) regarding the questions Q01, Q02, Q03, Q07, and Q10. These questions address the themes of ease of use, skill, use of the tool to explore datasets, and the confidence that the user had when using the tool, respectively.

Some feedback provided by the participants also attest to the ease that many of them felt when using VisMaker. But what factors led users to find VisMaker easier to use? What made Voyager less easy? What made users more confident using VisMaker? The answer to these questions varies for each participant. Participant P01, for example, stated that VisMaker’s recommendations were more intuitive because they presented a question that helped them identify what they were looking for. P05, in turn, stated that the questions in VisMaker made the recommendations more consistent. P06, P07, P08, P09, P10, P11, and P15 also commented directly on the facility that VisMaker provided them through the questions. Other participants, such as P12, P14, and P16, commented on other aspects, such as the highlight by type of the variables provided by VisMaker, and the vast amount of information presented in Voyager, stating that it hindered them a little.

While a high density of visualizations can cause difficulties in their interpretation and localization, the need to scroll too much through the recommendations panel may make participants tired of looking for the desired recommendation, which they are not even sure if it will appear. For some users, reading the questions helped to find the chart they were looking for, but for others it meant one more physical and mental effort, as many of them only used the recommendations and read the questions carefully when driven by the study.

5.2

Study 2: Data Exploration

In this second study, which aimed to evaluate the use of tools in a data exploration scenario, when users do not have specific questions to answer, we used the same datasets as the first study and the same session structure, presented by Figure 5.1.

The tasks that make up this study are based on the Study 1 tasks. While in Study 1 participants had to answer questions and the tasks were composed of a dataset and a set of questions, in Study 2, participants had as goal “to comprehensively explore the data” through the tools, generating useful visualizations. Since there was no explicit indicator of task completion, we limited the exploration process to 30 minutes. Thus, we performed each session of this experiment as follows:

1. The participant reads the consent agreement, and completes the profile questionnaire by answering some questions such as the field of study and how often they read or create data visualizations.
2. The participant watches a video tutorial regarding the first tool they will use;
3. **Within 30 minutes**, the participant uses the first tool to “*comprehensively explore the data*”;
4. The participant completes the first tool questionnaire;
5. The participant watches a video tutorial regarding the second tool they will use;
6. **Within 30 minutes**, the participant uses the second tool to “*comprehensively explore the data*”;
7. The participant completes the second tool questionnaire; and, finally,
8. The participant is interviewed about using both tools to complete each task.

5.2.1
Participants

As in Study 1, participants filled out an initial form through which we obtained some information about them. Figure 5.13 shows the distribution of the participants’ responses to the three questions about familiarity with performing data exploration and visualization tasks.

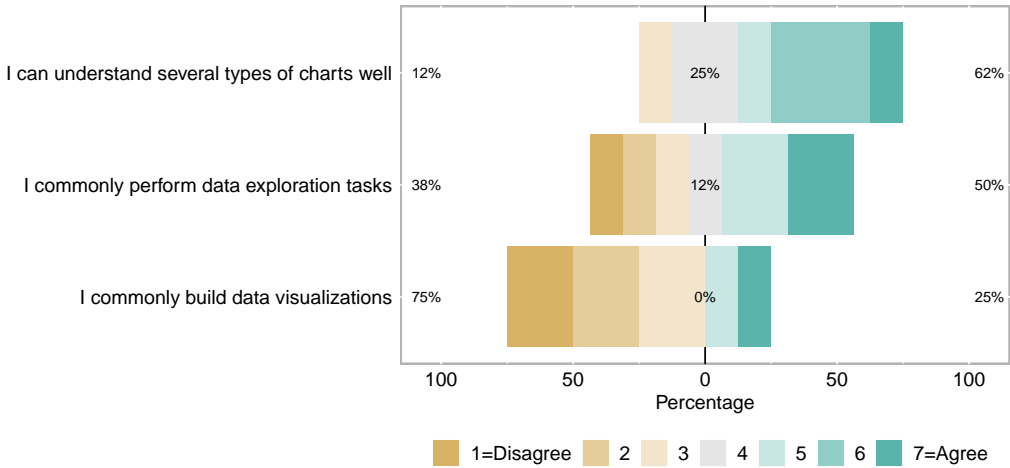


Figure 5.13: Profile questionnaire responses of Study 2.

Table 5.8 lists all eight participants in this study with some of their characteristics.

Participant	Sex	Degree	Fields of Study
P01	Female	Undergraduate	Computer Science
P02	Female	PhD student	Law
P03	Male	PhD student	Data Science, Optimization
P04	Male	Master’s student	Data Science
P05	Female	Master’s student	Software Engineering
P06	Female	Master’s student	Coputer Science
P07	Male	PhD student	Data Science
P08	Female	PhD	Optimization, Artificial Intelligence

Table 5.8: Characteristics of the participants in the second study.

5.2.2
Study 2 Findings

In this section we present the main findings of this second study. We start by presenting a little bit about the responses that the participants gave to the forms and then present the main feedback given by the participants during the interview.

5.2.2.1

Questionnaire Responses

Similar to Study 1, we applied the Wilcoxon-Mann-Whitney test to the questionnaire responses. In Study 2, we obtained statistical significance only for Q01, about ease of use. However, VisMaker obtained the highest median of the answers in most questions, which encourages us to conduct, in the future, another similar study with more participants, and therefore obtain a higher statistical power of the corresponding hypothesis test. Figure 5.14 and Table 5.9 presents these results.

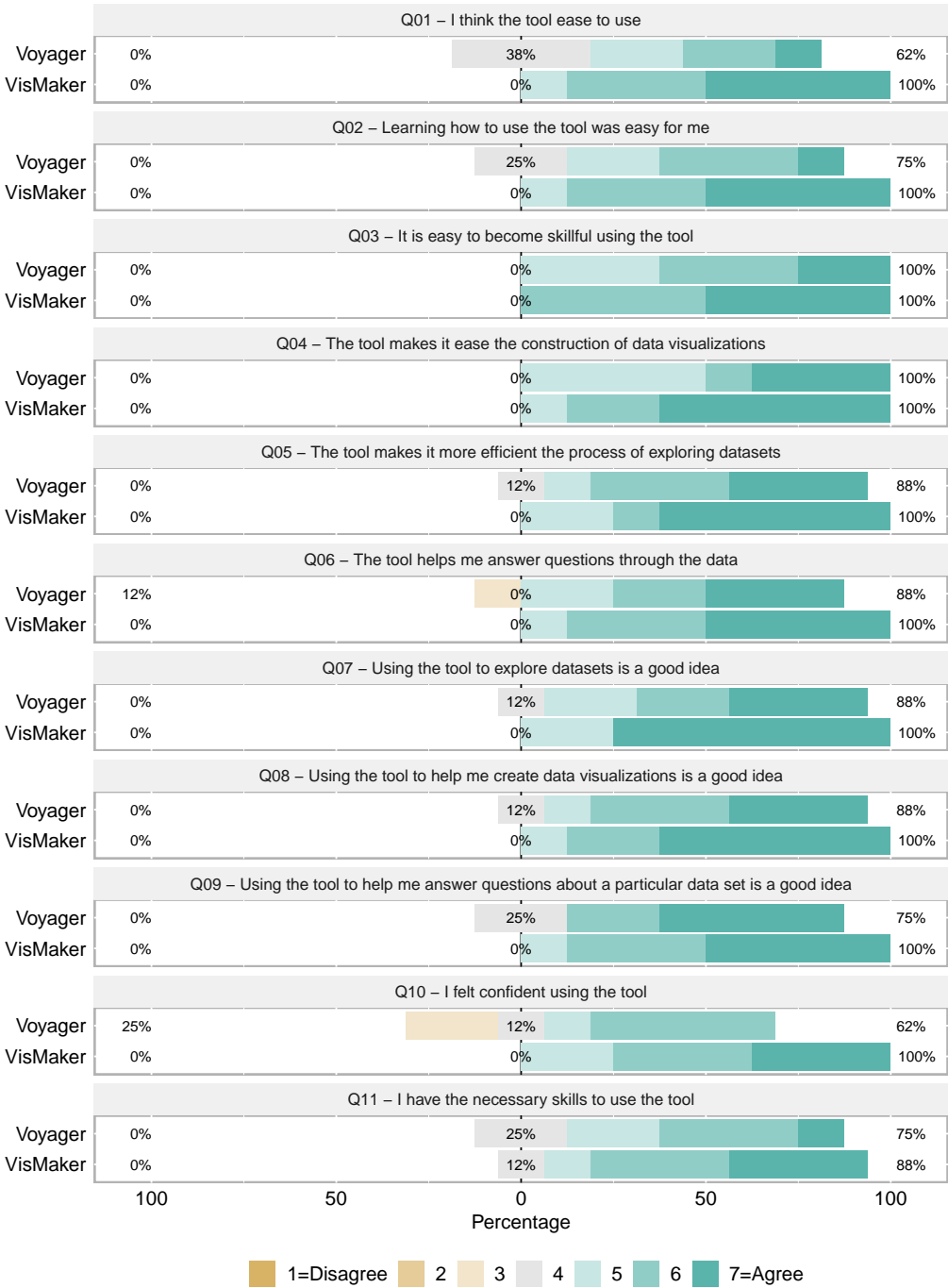


Figure 5.14: Questionnaire responses of Study 2.

Question	VisMaker median	Voyager median	p-value
Q01	6.5	5.0	0.033*
Q02	6.5	5.5	0.061
Q03	6.5	6.0	0.126
Q04	7.0	5.5	0.207
Q05	7.0	6.0	0.460
Q06	6.5	6.0	0.401
Q07	7.0	6.0	0.219
Q08	7.0	6.0	0.331
Q09	6.5	6.5	0.775
Q10	6.0	5.5	0.068
Q11	6.0	5.5	0.250

Table 5.9: Statistical significance of scores in Study 2

5.2.2.2

Participants' Feedback about the Tools in Study 2

In the interview, we included the two questions from Study 1, related to the recommendations: the first one explicitly asked about the recommendation approaches expressed through the recommended visualization panel; the second one asked about the charts obtained through both tools, if they corresponded in any way to the participants' expectations. We also asked a third question, related to the use of the questions presented in VisMaker. Below we list the main feedback given by the participants for these three interview questions.

1. *Considering the recommendation approaches, which tool most helped you complete the tasks?*

- **P01:** *"I found VisMaker easier to view the charts. The questions helped me."*
- **P02:** *"VisMaker was much easier, perhaps because it had fewer options and because I found the dataset cooler."*
- **P03:** *"VisMaker was simpler, visualizations were easier to understand, and it helped me more. Voyager has much information on the screen, and it just got a little confusing."*
- **P04:** *"I liked the recommendations because they gave new visualization ideas, and I liked the way VisMaker presents the recommendations. Voyager has much information, and it messed me up a bit."*
- **P05:** *"I think VisMaker, but not necessarily because of the questions. In general, the less information on the screen, the better for me."*

- **P06:** “*VisMaker helped me more, maybe because of the questions. Sometimes I wanted to add more information, and I didn’t know how so he would show the questions, and I knew better how to do it.*”
- **P07:** “*VisMaker was more interesting because it suggested hypotheses through questions. It is friendlier and perhaps a better option for less experienced users.*”
- **P08:** “*Watching the videos, I thought Voyager would be better but using it was easier with VisMaker, maybe because I used VisMaker later.*”

2. *During the exploration process, when you thought about the variable combinations you would like to use to build a visualization, did you imagine any chart? Did any of the tools better match your expectations?*

- **P01:** “*Yes. Sometimes I related the variables I was going to use with some charts. **I think that, in VisMaker, they corresponded more.** Voyager had many ball charts.*”
- **P02:** “*I imagined the answer, but not the way of visualization. **In VisMaker, I felt more correspondent,** I don’t know exactly why, but **I found VisMaker easier, it’s so much that I had a little difficulty in interpreting the visualizations.***”
- **P03:** “*Not at first, but after using the tools and seeing the charts they presented, I started to preview and, sometimes, see the recommendations and get ideas from them.*”
- **P04:** “*Firstly, I thought about questions that I could answer, that is my way of working with visualizations. After thinking about the question, I imagined some visualization, usually a bar or a line chart. **I think VisMaker corresponded more.***”
- **P05:** “*I imagined some bar charts. Sometimes the visualizations corresponded and sometimes not.*”
- **P06:** “*I just imagined some bar charts because that’s what is most common for me.*”
- **P08:** “*No. I think that as there is a proposal for recommendations, I was waiting to see what the tool recommended. **VisMaker’s charts was more familiar.** [...] **I don’t know if it was a coincidence, but by the time I thought something, it was there.***”

3. *Regarding the questions that VisMaker presented, you used them? Did they help you during the exploration process?*

- **P01:** *“I found VisMaker easier to view the charts. I had the questions there, and they helped me understand things better. [...] The questions helped me.”*
- **P02:** *“I looked a few times at the questions. They didn’t make much difference to me. I hardly used them.”*
- **P03:** *“I hardly used the questions.”*
- **P04:** *“I think the questions are interesting because they contextualize the recommendations, and I think it would help for less experienced users and when the user doesn’t know exactly what they are looking for.”*
- **P05:** *“Sometimes I would read the questions, and they would help me better understand the visualizations that were being recommended and to think of new visualizations that I had not yet imagined.”*
- **P06:** *“I read some questions. I looked at the variables, and if I saw that they interested me, I read the whole question.”*
- **P07:** *“Yes, sometimes I read the questions. They helped me to hypothesize about the data.”*
- **P08:** *“The questions helped a lot. It was easier and faster to understand what was in the charts, and the questions helped me raise questions and whether I really wanted to use the visualizations.”*

5.2.3 Discussions

In Study 2, we could see some of the advantages of the VisMaker’s recommendation approach, related to one of its main intended uses: supporting the data exploration process. When users do not precisely know what they are looking for, the recommended visualizations appear to present information and alternatives that can be useful for users, assisting them in acquiring knowledge.

During the exploration process, how the tools presented the recommendations can end up disturbing users. The high amount of information presented on the screen, for example, was one of the aspects highlighted with some frequency by the participants of this study. Comments like *“in general, the less information on the screen, the better for me”*, by participant P05, and whose feeling seems to be shared by P02, P03, P04, P06, and P07, who also commented about the amount of information presented by the tools, specially in Voyager.

When asked about the correspondence between the obtained visualizations and their expectations, most participants stated that they felt more familiar with the visualizations presented by VisMaker. Participant P01 explicitly

commented on the “bubble matrix” visualizations presented by Voyager, revealing no familiarity with this type of visualization. P02, in turn, stated that the visualizations built through VisMaker corresponded more to their expectations and that, despite not knowing exactly why, they had less difficulty in interpreting them. Other participants just commented that they hoped they would build more straightforward charts such as bar or line charts. Despite not having explicitly clicked on any chart during the exploration process, participant P08 commented that the visualizations obtained with VisMaker were more familiar and stated: *“whenever I thought about something, it was there”*.

It is interesting to note the answer that participant P04 gives to the second question. They describe the steps they take, starting with the question to answer until they get to the visualizations that will make it possible to obtain the answers. P04 also comments positively on the questions and how VisMaker presents the recommendations, strengthening the idea that the VisMaker recommendation approach supports users who do this same step-by-step during data exploration.

Regarding the questions presented by VisMaker, it is possible to note that more than half of the participants in this study used them in some way and gave positive feedback about them, as is the case of participant P01, who stated that *“they [the questions] helped me understand things better”*, and participant P07 who said that *“they [the questions] helped me to hypothesize about the data”*.

Although little used or even ignored by some participants, such as P02 and P03, the questions did not hinder them in carrying out their tasks. This leads us to believe that, since the participants do not need to compare the questions presented by VisMaker with a specific question they wish to answer (since they do not have it), they used the tool freely, as they wished.

5.3

Answering our Research Subquestions

The analysis of the use of tools during Study 1 allows us to answer some of our research subquestions, discussed below (in reverse order).

SQ3. How can questions improve visualization recommendations?

From the feedback of some participants, as well as from how they used the tools to complete the tasks, we can see that the attempt to identify and present the recommendations through questions can facilitate the interpretation and motivate interest in the corresponding visualizations.

When using either tool, some participants sometimes found it difficult

to map between variables and visualization channels, so they removed all the mapping they had already built and restarted the visualization construction because they understood that they had not yet concluded the task.

At other times, when part of the mapping had already been done, participants had the difficulty of having to add a new variable to the visualization and not being able to identify an appropriate channel for it. When this occurred, some participants turned to recommendations to find out where they could add such a variable. This happened with participant P04 while completing table 4 of the WEATHER task in the first study. Although they stated that they hardly used the recommendations, they stated in their feedback that *“in Voyager I couldn’t complete the task without the recommendations. It was very useful”*.

To this comment, we can add what participant P10 said when facing the same difficulty while using VisMaker: *“I think VisMaker is better when I need to use many variables because the more variables I have to use, the more I have to keep asking the question I want to answer in my mind. How VisMaker presents the questions it makes it easier”*.

In agreement with these comments, we can summarize the feedback of some other participants, who indicated that the questions presented by VisMaker during the recommendation process helped them to (i) better understand the meaning of the recommended charts; (ii) identify more quickly whether the recommended visualizations matched the participant’s questions of interest; and (iii) organize the use of variables in constructing a visualization.

From Study 2, we were able to observe another use of the questions in the recommendation process through the feedback of some participants, such as P04, who said that *“they [the questions] gave new visualization ideas”*, and P07, who said that *“VisMaker was more interesting because it suggested hypotheses through questions”*. Thus, we conclude that the questions can also help users raise hypotheses about the dataset they are exploring.

SQ2. How do users interact with the questions presented by the tool via the recommendations?

Regarding the interaction of users with the recommendation questions, de Sousa and Barbosa (2014) observed four different behaviors: (1) users who found in questions an opportunity to accomplish the task more quickly; (2) users who were not so satisfied or had doubts about the given solution and resorted to questions as a way to validate or to improve the result; (3) the user who used the questions as links to charts but who has not really read them; and (4) the user who did not interact with them. Through the VisMaker evaluation, these behaviors could also be verified, especially through Study 1,

in which users had specific questions that should be answered.

In Study 2, as there were no specific questions to be answered, the interaction with the recommendation questions was more focused on the interest of users who, at certain times, read and thought “*this can be interesting, I will use this visualization*”.

SQ1. How can a question about data determine the visualization type, including the mapping of data variables onto visual dimensions?

The relationship between visualizations and questions certainly involves the variables to which a given question refers. In this work, we defined a set of mappings we built to make this relationship possible. When appropriate, We used some mappings proposed by existing work that used the variable types to determine the most appropriate visualization types for their presentation.

The mappings between types of variables, questions, and types of visualizations that we have developed and presented in Table 4.7 compose an initial set, which is certainly not complete. The inclusion of new types of visualizations, questions, and the use of other types of data aimed at more specific contexts can contribute to having more interesting mechanisms through which users can answer questions about the data through increasingly efficient visualizations.

6

Conclusions

In this research we developed and evaluated a tool to facilitate the construction of visualizations for the exploration of data sets.

We conducted a literature review of studies in the area of information visualization and research on the main methods and tools currently used for this purpose. Many of these works were used as a basis for the development of VisMaker, a visualization building tool which includes a visualization recommender system based on the use of rules that map, for the different combinations of types of variables, a set of visualizations considered most appropriate. A major distinguishing characteristic of VisMaker was the use of questions built based on the types of variables and whose intention was to facilitate the understanding of the recommendations, raising hypotheses about the data and improving the recommendations and therefore the process of exploring the data.

We evaluated VisMaker through a comparison with the Voyager 2 tool in two controlled studies, one focused on answering specific questions and the other focused on a free data exploration scenario, without any questions to guide the participant. We analyzed the use of the tools and interviewed the study participants, obtaining feedback that allowed us not only to verify the positive and negative points of the recommendation approach proposed by VisMaker, but also to identify opportunities for improvement.

6.1

Contributions

The main contribution of this work is to support data exploration users or novices in the InfoVis area in building data visualizations and exploring data sets.

We developed VisMaker, a visualization recommender system that facilitates the construction of data visualizations and allows the exploration of generic datasets. Similar to the recommendation approach used by Voyager (Wongsuphasawat et al., 2016), VisMaker considers the variables already selected by the user and recommends new visualizations based on the addition of new variables. Inspired by Voyager 2 (Wongsuphasawat et al., 2017), VisMaker

also uses a hybrid interface, through which the user can build a visualization by defining the mapping between variables and visualization channels or select a visualization from the recommendations provided by the tool. Unlike the Voyager tools, VisMaker organizes recommendations through questions that involve the types of selected variables according to a set of rules that we have built to assign higher semantic value to the recommendations, simplifying their understanding, and assisting the data exploration process.

The two studies conducted to evaluate VisMaker revealed opportunities for improvements that can be used in future works, both related to our tool and other tools for building or recommending visualizations and exploring data sets.

6.2

Future Work

Some of our findings were inconclusive. Because our study involved few users (especially Study 2), as future work we intend to carry out new studies with greater participation of users to better assess the statistical significance of participants' responses related to several aspects of the tools. In any case, through the studies we conducted, we have already identified opportunities to improve the recommendation approach presented in VisMaker, as follows:

- The use of domain ontologies, for example, can enable the construction of related questions that are more flexible, adaptable, and useful to the data exploration process than simply relying on the types of the selected variables.
- The application of machine learning models, such as those presented in Section 3.2.2, can improve recommendations, helping to identify the efficiency of certain types of visualizations according to the characteristics of the data to be presented.
- The implementation of a recommender system that takes users' behavior into account by identifying the choices and types of visualization they use most often can also increase the quality of recommendations, adapting them to the preferences of each user.
- Finally, the construction of a model for the recommendation of visualizations that unites all these technologies and approaches, joining them in an approach based on the behaviors of the users, may constitute valuable work to improve the visualization recommendation process.

Bibliography

- Amar, R., Eagan, J., and Stasko, J. (2005). Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 111–117. IEEE.
- Anscombe, F. J. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1):17–21.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370.
- Burke, R. (2007). Hybrid web recommender systems. In *The adaptive web*, pages 377–408. Springer.
- Cleveland, W. S. and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, pages 319–340.
- de Sousa, T. A. F. and Barbosa, S. D. J. (2014). Recommender system to support chart constructions with statistical data. In *International Conference on Human-Computer Interaction*, pages 631–642. Springer.
- Dibia, V. and Demiralp, c. (2018). Data2vis: Automatic generation of data visualizations using sequence to sequence recurrent neural networks.
- Duke, D. J., Brodlie, K. W., and Duce, D. A. (2004). Building an ontology of visualization. In *Proceedings of the conference on Visualization'04*, pages 598–7. IEEE Computer Society.
- Gomez-Uribe, C. A. and Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19.
- Good, N., Schafer, J. B., Konstan, J. A., Borchers, A., Sarwar, B., Herlocker, J., Riedl, J., et al. (1999). Combining collaborative filtering with personal agents for better recommendations. *AAAI/IAAI*, 439.

- Gotz, D. and Wen, Z. (2009). Behavior-driven visualization recommendation. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 315–324. ACM.
- Grammel, L., Tory, M., and Storey, M.-A. (2010). How information visualization novices construct visualizations. *IEEE transactions on visualization and computer graphics*, 16(6):943–952.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220.
- Heer, J., Bostock, M., Ogievetsky, V., et al. (2010). A tour through the visualization zoo. *Commun. Acm*, 53(6):59–67.
- Hu, K., Bakker, M. A., Li, S., Kraska, T., and Hidalgo, C. (2019). VizML: A Machine Learning Approach to Visualization Recommendation. In *Proceedings of the 2019 Conference on Human Factors in Computing Systems (CHI)*. ACM.
- Jacobson, K., Murali, V., Newett, E., Whitman, B., and Yon, R. (2016). Music personalization at spotify. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 373–373.
- Luo, Y., Qin, X., Tang, N., and Li, G. (2018). Deepeye: Towards automatic data visualization. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 101–112. IEEE.
- Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, 5(2):110–141.
- Mackinlay, J., Hanrahan, P., and Stolte, C. (2007). Show me: Automatic presentation for visual analysis. *IEEE transactions on visualization and computer graphics*, 13(6):1137–1144.
- Munzner, T. (2014). *Visualization analysis and design*. AK Peters/CRC Press.
- Perry, D. B., Howe, B., Key, A. M., and Aragon, C. (2013). Vizdeck: Streamlining exploratory visual analytics of scientific data.
- Pinker, S. (1990). A theory of graph comprehension. *Artificial intelligence and the future of testing*, pages 73–126.
- Polowinski, J. and Voigt, M. (2013). Viso: a shared, formal knowledge base as a foundation for semi-automatic infovis systems. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 1791–1796. ACM.

- Ricci, F., Rokach, L., and Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer.
- Satyanarayan, A., Moritz, D., Wongsuphasawat, K., and Heer, J. (2016). Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics*, 23(1):341–350.
- Satyanarayan, A., Russell, R., Hoffswell, J., and Heer, J. (2015). Reactive vega: A streaming dataflow architecture for declarative interactive visualization. *IEEE transactions on visualization and computer graphics*, 22(1):659–668.
- Schafer, J. B., Konstan, J., and Riedl, J. (1999). Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158–166.
- Srinivasan, A., Drucker, S. M., Endert, A., and Stasko, J. (2019). Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE transactions on visualization and computer graphics*, 25(1):672–681.
- Staab, S. and Studer, R. (2010). *Handbook on ontologies*. Springer Science & Business Media.
- Stolte, C., Tang, D., and Hanrahan, P. (2002). Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):52–65.
- Tory, M. and Möller, T. (2002). A model-based visualization taxonomy. *School of Computing Science, Simon Fraser University*.
- Voigt, M. and Polowinski, J. (2011). Towards a unifying visualization ontology.
- Wehrend, S. and Lewis, C. (1990). A problem-oriented classification of visualization techniques. In *Proceedings of the First IEEE Conference on Visualization: Visualization90*, pages 139–143. IEEE.
- Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in statistics*, pages 196–202. Springer.
- Wongsuphasawat, K., Moritz, D., Anand, A., Mackinlay, J., Howe, B., and Heer, J. (2016). Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE transactions on visualization and computer graphics*, 22(1):649–658.

- Wongsuphasawat, K., Qu, Z., Moritz, D., Chang, R., Ouk, F., Anand, A., Mackinlay, J., Howe, B., and Heer, J. (2017). Voyager 2: Augmenting visual analysis with partial view specifications. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2648–2659. ACM.

A

Informed Consent Form (in Portuguese)

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



TERMO DE CONSENTIMENTO

Este experimento tem por objetivo avaliar ferramentas de recomendação de visualizações de dados estatísticos. Para tanto, você será convidado a realizar duas tarefas distintas através de duas ferramentas de construção de visualizações. As duas tarefas tratam sobre dois conjuntos de dados e consistem em responder a um conjunto de perguntas sobre os dados. Você deverá utilizar as ferramentas de visualização para construir gráficos que irão possibilitar responder as perguntas apresentadas nas tarefas. Tanto as ferramentas quanto as perguntas utilizarão a língua inglesa por padrão.

Durante a realização do experimento, **SERÃO GRAVADAS A VOZ DO PARTICIPANTE, BEM COMO A SUA INTERAÇÃO COM AS FERRAMENTAS ATRAVÉS DE LOGS DE UTILIZAÇÃO E CAPTURA DE TELA.**

TODOS OS DADOS CAPTURADOS DURANTE A REALIZAÇÃO DESTE EXPERIMENTO SÃO SIGILOSOS E A IDENTIDADE DOS PARTICIPANTES NÃO SERÁ REVELADA EM MOMENTO ALGUM, NEM MESMO NA PUBLICAÇÃO DOS RESULTADOS DE PESQUISA.

O participante tem a liberdade de não participar deste experimento ou mesmo de retirar seu consentimento a qualquer momento, mesmo após o início da entrevista/coleta de dados, sem qualquer prejuízo.

() Concordo com o termo de consentimento e aceito participar do experimento.

Assinatura

B

CAPES-RJ Task Questions

TOOL: _____

NAME: _____

DATASET (CAPES-RJ.csv)**YEAR:** temporal**CITY:** nominal**JURIDICAL STATUS:** nominal**FIELD:** nominal**QNT MASTERS:** quantitative**QNT DOCTORAL:** quantitative**QNT POSTDOCTORAL:** quantitative

QUESTIONS:

1) How many masters scholarships were offered in the city of Rio de Janeiro in 2008?

R: _____

2) How many doctoral scholarships were offered in the *Ciências Humanas* field?

R: _____

3) In which city was offered the largest number of postdoctoral scholarships in the *Ciências Agrárias* field?

R: _____

4) Which knowledge field receives the most masters scholarships?

R: _____

TO SOLVE THROUGH RECOMMENDATIONS:

Most common field of study in each city			
CAMPOS DOS GOYTACAZES	NITERÓI	RIO DE JANEIRO	SEROPÉDICA

The ordered list of the three most common fields of study in each juridical status		
<i>ESTADUAL</i>	<i>FEDERAL</i>	<i>PRIVADA</i>

Most common juridical status of institutions in each city			
CAMPOS DOS GOYTACAZES	NITERÓI	RIO DE JANEIRO	VASSOURAS

Field of study with more masters scholarships in each city			
NITERÓI	PETRÓPOLIS	SEROPÉDICA	VASSOURAS

C

WEATHER Task Questions

TOOL: _____

NAME: _____

DATASET (WEATHER.csv):**LOCATION:** nominal**DATE:** temporal**PRECIPITATION:** quantitative**TEMP MAX:** quantitative**TEMP MIN:** quantitative**WIND:** quantitative**WEATHER:** nominal

QUESTIONS:

1) In which month is the highest mean of the maximum temperature?

R: _____

2) In what months is the mean of rain precipitation in Seattle less than in New York?

R: _____

3) Which city had the highest mean of rain precipitation?

R: _____

4) In which month did Seattle have the highest mean of rain precipitation?

R: _____

TO SOLVE THROUGH RECOMMENDATIONS:

The year each weather was most recorded				
DRIZZLE	FOG	RAIN	SNOW	SUN

The mean of the minimum temperature in each weather				
DRIZZLE	FOG	RAIN	SNOW	SUN

The ordered list of the three most common weather in each city	
NEW YORK	SEATTLE

The mean of the minimum temperature in each weather									
NEW YORK					SEATTLE				
DRIZZLE	FOG	RAIN	SNOW	SUN	DRIZZLE	FOG	RAIN	SNOW	SUN

D
User Profile Questionnaire (in Portuguese)

Perfil do Estudante

*Obrigatório

1. Qual o seu nome? *

2. Área de estudo *

Marque todas que se aplicam.

- ☐ Bancos de Dados
- ☐ Computação Gráfica
- ☐ Ciência de Dados
- ☐ Engenharia de Software
- ☐ Hipertexto e Multimídia
- ☐ Interação Humano-Computador
- ☐ Linguagens de Programação
- ☐ Otimização e Raciocínio Automático
- ☐ Redes de Computadores e Sistemas Distribuídos
- ☐ Teoria da Computação
- ☐ Indefinido
- ☐ Outro:

3. Costumo realizar tarefas de exploração de dados com bastante frequência *

Marcar apenas uma oval.

1234567

Discordo totalmenteConcordo totalmente

4. Costumo construir visualizações de dados com bastante frequência *

Marcar apenas uma oval.

1234567

Discordo totalmenteConcordo totalmente

5. Tenho bastante familiaridade e consigo compreender bem diversos tipos de gráficos *

Marcar apenas uma oval.

1234567

Discordo totalmenteConcordo totalmente

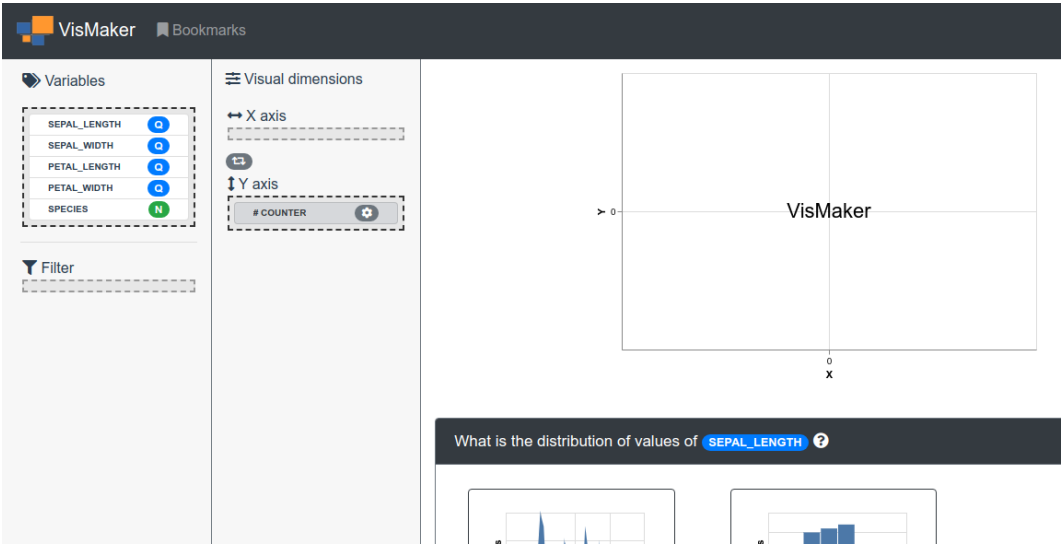
E

VisMaker Questionnaire (in Portuguese)

VisMaker

Formário para avaliação da ferramenta VisMaker

*Obrigatório



1. Qual o seu nome? *

Facilidade de Uso

2. Eu achei o VisMaker fácil de usar *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

3. Aprender a usar o VisMaker foi fácil pra mim *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

4. É fácil tornar-se hábil utilizando o VisMaker *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

Utilidade

5. O VisMaker facilita a construção de visualizações de dados *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

6. O VisMaker torna mais eficiente o processo de exploração de conjuntos de dados *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

7. O VisMaker me ajuda a responder perguntas através dos dados *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

Atitude

8. Utilizar o VisMaker para explorar conjuntos de dados é uma boa ideia *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

9. Utilizar o VisMaker para me ajudar a criar visualizações de dados é uma boa ideia *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

10. Utilizar o VisMaker para me ajudar a responder perguntas sobre um determinado conjunto de dados é uma boa ideia *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

Auto-eficácia

11. Eu me senti confiante ao utilizar o VisMaker **Marcar apenas uma oval.*

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

12. Eu tenho as habilidades necessárias para utilizar o VisMaker **Marcar apenas uma oval.*

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

Powered by



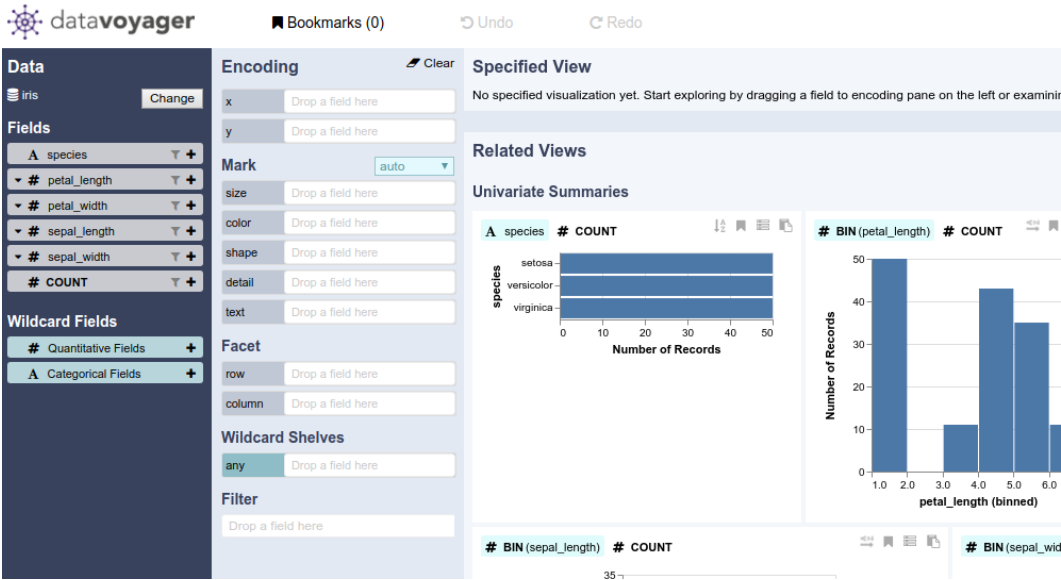
Google Forms

F
Voyager 2 Questionnaire (in Portuguese)

Voyager

Formulário para avaliação do Voyager

*Obrigatório



1. Qual o seu nome? *

Facilidade de Uso

2. Eu achei o Voyager fácil de usar *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

3. Aprender a usar o Voyager foi fácil pra mim *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

4. É fácil tornar-se hábil utilizando o Voyager *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

Utilidade

5. O Voyager facilita a construção de visualizações de dados *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

6. O Voyager torna mais eficiente o processo de exploração de conjuntos de dados *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

7. O Voyager me ajuda a responder perguntas através dos dados *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

Atitude

8. Utilizar o Voyager para explorar conjuntos de dados é uma boa ideia *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

9. Utilizar o Voyager para me ajudar a criar visualizações de dados é uma boa ideia *

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

10. **Utilizar o Voyager para me ajudar a responder perguntas sobre um determinado conjunto de dados é uma boa ideia ***

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

Auto-eficácia

11. **Eu me senti confiante ao utilizar o Voyager ***

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

12. **Eu tenho as habilidades necessárias para utilizar o Voyager ***

Marcar apenas uma oval.

	1	2	3	4	5	6	7	
Discordo totalmente	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Concordo totalmente

G

Transcription of participants' steps in each experiment (in Portuguese) GRUPO 1

P01

VisMaker - CAPES-RJ

T1 - [12:06] O participante adicionou a variável YEAR para o EIXO X. [12:16] Em seguida, visando responder a Pergunta 1 com precisão, adicionou a mesma variável ANO para o campo de filtros (isso o atrapalhou muito). [12:31] O participante se confundiu um pouco com a forma com a qual a ferramenta possibilita realizar a filtragem dos valores. [13:12] O participante adicionou a variável QNT MASTERS para o EIXO Y sem aplicar a função SUM. [13:31] O participante reabriu a configuração de filtro da variável YEAR adicionada anteriormente para verificar se havia selecionado somente o ano 2008. [13:56] O participante navegou pelo painel de perguntas relacionadas de forma rápida, verificando que todos os gráficos estavam em branco e estranhando tal fato (decorrente de um bug na ferramenta). [14:13] O participante verificou novamente a configuração de filtragem da variável YEAR. [14:29] Orientado por mim, o participante altera o *time unit* da variável YEAR no EIXO X para visualizá-la em anos (*year*). Isso não resolve o problema. [14:54] Ao reler a pergunta, o participante percebeu que deveria adicionar a variável CITY. O participante adicionou a variável CITY para o EIXO X, removendo a variável YEAR do *encoding*. [15:14] O participante adicionou a variável CITY para o campo de filtro a fim de selecionar somente os dados referentes à cidade do Rio de Janeiro. Um bug na ferramenta impediu tal utilização. [16:26] O participante alterou o gráfico para o tipo *strip plot*. [16:27] O participante navegou pela lista das perguntas, leu algumas perguntas mas não selecionou nenhuma visualização. [16:52] O participante percebe a necessidade de aplicar a função SUM na variável QNT MASTERS que estava no EIXO Y. [17:02] O participante obtém o gráfico final para a Pergunta 1. O gráfico, no entanto, fornece a resposta errada por conta da variável temporal ter sido adicionada ao campo de filtro (bug da ferramenta).

T2 - [17:44] O participante inicia a leitura da segunda pergunta e navega pela lista de perguntas procurando algo relacionado. [18:41] O participante adicionou a variável QNT DOCTORAL para o EIXO Y. [18:45] O participante adiciona a variável FIELD para o campo EIXO X. [19:01] O participante adicionou a variável FIELD para o campo de filtros. [19:17] O participante seleciona como filtro apenas o FIELD de valor CIÊNCIAS HUMANAS e a ferramenta acabou parando devido a um erro. [19:44 - 20:10] O participante refaz os passos que já havia feito antes da ferramenta ser reiniciada. [20:15] O participante altera o tipo de gráfico para *strip plot*. [20:20] O participante aplica a função SUM à variável QNT DOCTORAL que está no EIXO Y e obtém o gráfico que responde a Pergunta 2.

T3 - [20:48] O participante faz a leitura da Pergunta 3 e navega pelas perguntas relacionadas sem desfazer o mapeamento que respondia a pergunta anterior. [21:17] O participante remove a variável QNT DOCTORAL do EIXO Y. [21:21] O participante adiciona a variável QNT POSTDOCTORAL ao EIXO Y. [21:30] O participante aplica a função SUM à variável QNT POSTDOCTORAL no EIXO Y. [21:41] O participante adiciona a variável CITY ao EIXO X. [21:58] O participante adiciona altera o filtro da variável FIELD para selecionar apenas os dados referentes ao valor CIÊNCIAS AGRÁRIAS. [22:14] O participante visualiza o gráfico e finaliza a tarefa respondendo a Pergunta 3.

T4 - [22:31] O participante lê a Pergunta 4. [22:43] O participante adiciona a variável FIELD para o EIXO X. [22:47] O participante adiciona a variável QNT MASTERS ao EIXO Y. [22:51] O participante aplica a função SUM à variável QNT MASTERS no EIXO Y. [22:56] O participante remove a variável FIELD do campo de filtros. [23:04] O participante obtém a resposta da Pergunta 4.

T5 - [25:39] O participante faz a leitura da Tabela 1 e navega pela lista de perguntas. [26:32] O participante abre uma visualização referente à pergunta que adiciona a variável FIELD. [27:14] O participante encontra a pergunta que correlaciona as variáveis FIELD e CITY. [27:25] O participante abre o terceiro gráfico da pergunta encontrada (um *heatmap*) e utiliza-o para preencher a tabela.

T6 - [29:35] O participante inicia a leitura da Tabela 2 e navega pela lista de perguntas. [29:43] Após encontrar a pergunta sobre a variável JURIDICAL STATUS, o participante abre o gráfico recomendado. [30:21] O participante encontra a pergunta que correlaciona JURIDICAL STATUS e FIELD e abre o terceiro gráfico recomendado (um *heatmap*). [31:16] O participante pergunta sobre a possibilidade de “trocar a forma” do gráfico. [31:30] O participante remove a variável FIELD do EIXO Y. [31:48] O participante seleciona o segundo gráfico da pergunta que correlaciona JURIDICAL STATUS e FIELD (um gráfico de barras empilhadas normalizado) e o utiliza para preencher a tabela.

T7 - [35:50] O participante inicia a leitura da Tabela 3 e navega pelas perguntas. [36:21] O participante abre o gráfico da pergunta que adiciona a variável JURIDICAL STATUS. [36:29] O participante adiciona manualmente a variável CITY para o EIXO X. [36:39] O participante transfere a variável CITY do EIXO X para o EIXO Y. [36:44] O participante adiciona manualmente a variável JURIDICAL STATUS para o EIXO X e preenche a Tabela 3.

T8 - [38:20] O participante remove a variável JURIDICAL STATUS do EIXO X. [28:26] O participante adiciona a variável QNT MASTERS para o EIXO X. [29:30] O participante navega pelas perguntas, identifica a pergunta que relaciona as variáveis MEAN OF QNT MASTERS, CITY e FIELD e abre o gráfico recomendado. [39:25] O participante troca a função média por SUM na variável QNT MASTERS e obtém o gráfico para preencher a Tabela 4.

Voyager - WEATHER

T1 - [51:59] O participante faz a leitura da Pergunta 1. [52:32] O participante abre uma visualização recomendada que adiciona BIN(TEMP_MAX) ao EIXO X. [52:29] O participante aplica a função MEDIAN (mediana), pensando ser a função MEAN (média), na variável TEMP_MAX no EIXO X. [52:57] O participante remove todo o mapeamento e começa do início. [53:14] O participante adiciona a variável DATE para o EIXO X e muda o *time unit* para MONTH. [53:30] O participante adiciona a variável TEMP_MAX para o EIXO Y e aplica a função MEAN, obtendo o gráfico que responde a pergunta.

T2 - [54:20] O participante lê a Pergunta 2. [54:37] O participante remove a variável TEMP_MAX do EIXO Y e adiciona a variável PRECIPITATION para o EIXO Y. [54:54] O participante abre uma visualização recomendada que relaciona a MEAN OF PRECIPITATION e DATA com MONTH. [55:35] O participante adiciona a variável LOCATION para o campo COLOR. [55:44] O participante

remove a variável LOCATION do campo COLOR para adicioná-la ao campo de filtros. [55:52] O participante remove LOCATION do campo de filtros. [56:00] O participante adiciona LOCATION ao campo COLOR novamente. [56:17] Apesar do gráfico responder a pergunta, o participante pergunta se é possível dar zoom para melhor visualizar os valores e acaba errando a pergunta por causa da escala do gráfico.

T3 - [57:33] O participante inicia a leitura da Pergunta 3. [57:47] O participante troca a função da variável PRECIPITATION (ainda no EIXO Y) de MEAN para MAX. [57:55] O participante remove a variável DATE do EIXO X. [58:12] O participante adiciona a variável LOCATION para o EIXO X. [58:20] O participante remove a variável LOCATION de COLOR, verifica o gráfico e relê a pergunta. [58:37] O participante altera a função da variável PRECIPITATION para MEAN e obtém o gráfico para responder a pergunta.

T4 - [59:10] O participante inicia a leitura da Pergunta 4. [59:43] O participante navegou pelas recomendações e abriu um gráfico que relaciona as variáveis LOCATION, MEAN OF PRECIPITATION e DATE com YEAR (o mapeamento da recomendação não é bom). [59:52] O participante alterou o *time unit* da variável YEAR para MONTH. [01:00:33] O participante remove todas as variáveis e começa do início, arrasta a variável DATE para o EIXO Y e muda o *time unit* para MONTH. [01:00:42] O participante adiciona a variável PRECIPITATION para o EIXO X e aplica a função MEAN. [01:01:07] Não satisfeito com DATE no EIXO Y, o participante remove todo o mapeamento para fazê-lo novamente com PRECIPITATION no EIXO Y e DATE no EIXO X. [01:01:32] O participante adiciona LOCATION para o campo de filtro e mantém selecionado somente o valor SEATTLE e encontra a resposta para a Pergunta 4.

T5 - [01:02:30] O participante faz a leitura a tabela e verifica as recomendações. [01:03:03] O participante abre a recomendação que adiciona a variável DATE com YEAR. [01:03:28] O participante adiciona manualmente a variável WEATHER para o EIXO Y e a remove em seguida, adicionando novamente o COUNT em seu lugar. [01:03:53] O participante abre a recomendação que adiciona a variável WEATHER (gráfico de linhas) e encontra o gráfico para preencher a tabela.

T6 - [01:05:47] O participante aplica a função MEAN sobre o COUNT (isso o atrapalhou bastante). [01:06:07] O participante adiciona a variável TEMP_MIN para o EIXO X. [01:06:40] O participante removeu a variável WEATHER do campo COLOR. [01:07:05] O participante abriu uma recomendação que adiciona a variável WEATHER ao campo COLOR, voltando ao que tinha antes. [01:07:55] O participante tenta procurar novos tipos de gráficos, porém isso não o ajuda. [01:08:20] O participante aplica a função MEAN à variável TEMP_MIN, mas limpa todo o mapeamento em seguida (talvez por acidente)... Usa UNDO para refazer. [01:08:37] O participante finalmente remove o COUNT com MEAN que o estava atrapalhando, e troca a variável WEATHER de COLOR para EIXO Y, obtendo o gráfico correto.

T7 - [01:09:52] O participante faz a leitura da tabela. [01:10:33] O participante abre a recomendação que adiciona a variável WEATHER para o EIXO Y. [01:10:58] O participante abre a visualização que adiciona a variável CITY para COLOR (gráfico de barras empilhadas). [01:11:57] Não satisfeito com a visualização, o participante navega até o painel de *alternative encodings* e abre a visualização de "bubble map". [01:12:05] O participante tenta mudar o gráfico para um *heatmap*, porém retorna para o "bubble map" e preenche a tabela.

T8 - [01:13:38] O participante faz a leitura da tabela e navega pelas recomendações. [01:14:35] O participante abre a recomendação que adiciona a variável WEATHER para o EIXO Y. [01:14:43] O participante aplica a função MEAN no COUNT. Isso novamente o atrapalha e ele usa o UNDO para desfazer a ação. [01:15:34] O participante abre a recomendação que adiciona a variável TEMP_MIN ao campo COLOR. O participante demonstra dúvida sobre a utilização do contador. [01:16:57] O participante muda a variável TEMP_MIN do campo COLOR para o campo SIZE. [01:17:05] O participante adiciona a variável LOCATION para o campo COLOR. [01:17:38] O participante faz um *swap* entre as variáveis TEMP_MIN e LOCATION nos campos COLOR e SIZE, mas não gosta do resultado e desfaz. Utiliza o gráfico obtido, apesar de possuir um COUNT desnecessário, para preencher a tabela.

P02

VisMaker - CAPES-RJ

T1 - [12:12] O participante adiciona a variável CITY para o EIXO X, mudando-a logo em seguida para o campo COLOR. [12:20] O participante adiciona a variável YEAR para o EIXO X. [13:00] O participante adiciona a variável QNT MASTER para o EIXO Y. [13:16] O participante adiciona a variável CITY para o campo de filtro e aplica seleciona somente os dados onde o valor é RIO DE JANEIRO. [13:34] O participante aplica a função SUM na variável QNT MASTERS e obtém a visualização final. Orientado por mim, seleciona o *time unit* YEAR para a variável temporal YEAR.

T2 - [14:47] O participante adiciona a variável QNT DOCTORAL para o EIXO Y. [14:49] O participante aplica a função SUM à variável QNT DOCTORAL. [15:05] O participante remove a variável CITY do campo COLOR. [15:11] O participante adiciona a variável FIELD para o EIXO X. Apesar de ter feito os procedimentos corretamente, o participante esqueceu de remover a variável CITY do campo de filtro. [16:01] O participante remove a variável CITY do campo de filtros e corrige sua resposta. O participante encontrou um bug que fez com que os rótulos e valores dos eixos do gráfico principal desaparecessem após a remoção do filtro.

T3 - [16:50] O participante lê a pergunta. [17:00] O participante remove a variável QNT DOCTORAL do EIXO Y e logo em seguida adiciona a variável QNT POSTDOCTORAL em seu lugar. [17:10] O participante aplica a função SUM à variável QNT POSTDOCTORAL. [17:34] O participante move a variável FIELD do EIXO X para o campo de filtros. [17:49] O participante aplica um filtro para manter somente os dados com valor CIÊNCIAS AGRÁRIAS. [18:02] O participante adiciona por acidente a variável JURIDICAL STATUS para o EIXO X. Sua intenção era adicionar a variável CITY. [18:07] O participante corrige o erro cometido e adiciona a variável CITY para o EIXO X, obtendo a visualização final.

T4 - [18:36] O participante faz a leitura da pergunta. [18:55] O participante adiciona a variável QNT MASTERS para o EIXO Y. [19:04] O participante aplica a função SUM à variável QNT MASTERS. [19:10] O participante adiciona a variável FIELD para o EIXO X e obtém a visualização final.

T5 - [20:51] O participante adiciona a variável CITY para o EIXO Y. [21:08] O participante move a variável CITY para o EIXO X. [21:13] O participante adiciona a variável FIELD para o EIXO X.

Apesar de ter encontrado a visualização correta, o participante mostra-se surpreso ao visualizar o *heatmap* apresentado pela ferramenta. [21:34] O participante remove todo o mapeamento e reinicia. [21:41] O participante adiciona a variável FIELD para o EIXO X. [22:06] O participante adiciona a variável CITY para o EIXO Y. [22:26] O participante troca a variável FIELD do EIXO X para o campo COLOR. [22:49] O participante muda a variável CITY do EIXO Y para o campo de filtro. [22:55] O participante adiciona a variável FIELD para o EIXO Y. [23:05] O participante utiliza o botão *swap X and Y* para transferir a variável FIELD para o EIXO X. [23:51 - 26:10] O participante aplica vários filtros, um para cada valor de CITY para preencher a tabela.

T6 - [16:18] O participante mantém somente a variável FIELD no EIXO X e navega pelas recomendações da ferramenta. [26:48] O participante abre o primeiro gráfico da pergunta que relaciona as variáveis FIELD e JURIDICAL STATUS (gráfico de barras empilhadas). [27:13] O participante remove a variável CITY do campo de FILTRO. [27:21] O participante move a variável JURIDICAL STATUS para o EIXO Y. [27:26] O participante adiciona a variável FIELD para o campo COLOR. [28:17] O participante adiciona a variável FIELD para o campo de filtros. [28:33] O participante adiciona a variável JURIDICAL STATUS para o campo de filtros e remove a variável FIELD do campo de filtros, mantendo-a somente no EIXO X e preenche a tabela aplicando diversos filtros nos valores da variável JURIDICAL STATUS.

T8 - [31:39] Após navegar pelas recomendações, o participante abre o primeiro gráfico da pergunta que relaciona as variáveis FIELD e CITY, a fim de preencher logo a Tabela 4 ([33:16] O participante diz: “*eu pulei uma porque eu tentei ver uma que já estava recomendada*”). [31:50] O participante remove a variável JURIDICAL STATUS do campo de filtros. [32:20] O participante adiciona a variável QNT MASTERS para o EIXO X e aplica a função SUM logo em seguida, realizando um *swap* entre os eixos X e Y. [32:59] O participante adiciona a variável CITY para o campo de filtros e preenche a tabela aplicando vários filtros nos valores da variável CITY.

T7 - [34:22] O participante adiciona a variável JURIDICAL STATUS para o EIXO X. [34:36] O participante adiciona a variável CITY para o EIXO Y e faz um *swap* entre os eixos. [34:45] O participante tenta mover o atributo de contagem, mas a ferramenta não lhe permite tal operação. [35:06] O participante move a variável JURIDICAL STATUS do EIXO Y para o campo COLOR. [35:14] O participante se questiona sobre os campos ROWS e COLUMNS e move a variável JURIDICAL STATUS para o campo COLUMNS. [35:55] Após reiniciar o mapeamento, o participante adiciona a variável JURIDICAL STATUS para o EIXO X. [36:04] O participante adiciona a variável CITY para o campo COLUMNS, movendo-a em seguida para o campo ROWS. O participante adiciona a variável JURIDICAL STATUS para o campo COLOR, para facilitar a visualização. O participante chega a comentar “*eu poderia ter feito isso nos outros*”, referindo-se ao uso de filtros.

Voyager

T1 - [49:00] O participante adiciona a variável TEMP_MAX para o EIXO X e a variável DATE para o EIXO Y. [49:08] O participante faz um *swap* entre os eixos X e Y. [49:17] O participante seleciona o *time unit* da variável DATE como MONTH. [49:33] O participante aplica a função MEAN à variável TEMP_MAX e obtém a visualização final.

T2 - [50:15] O participante adiciona a variável DATE para o EIXO X e muda seu *time unit* para MONTH. [50:23] O participante adiciona a variável PRECIPITATION para o EIXO Y e aplica a função MEAN. [50:40] O participante adiciona a variável LOCATION para o campo COLOR. O participante tem problemas com a escala do gráfico e pergunta sobre a possibilidade de dar zoom. [52:03] O participante adiciona a variável LOCATION para o campo de filtros para responder a pergunta com precisão.

T3 - [52:37] O participante remove a variável DATE do EIXO X. A variável PRECIPITATION continua no EIXO Y. [52:45] O participante muda a variável LOCATION do campo COLOR para o EIXO X e obtém a visualização final. Mais uma vez o participante tem problema com a escala, mas isso não lhe impede de responder bem a pergunta.

T4 - [53:27] O participante adiciona a variável DATE para o EIXO X e muda seu *time unit* para MONTH. A variável PRECIPITATION continua no EIXO Y. [53:35] O participante adiciona a variável LOCATION para o campo de filtros. [53:43] O participante seleciona somente os dados referentes ao valor SEATTLE e encontra a visualização final.

T5 - [55:15] O participante adiciona a variável DATE para o EIXO Y e muda seu *time unit* para YEAR. [55:32] O participante adiciona a variável WEATHER para o EIXO X. [55:40] O participante move a variável WEATHER para o campo COLUMN. [56:30] O participante move a variável DATE para o EIXO X. [56:47] O participante move a variável DATE para o campo COLUMN e a variável WEATHER vai para o EIXO X. [57:03] O participante adiciona a variável WEATHER para o campo ROW. [57:20] O participante remove todo o mapeamento e inicia novamente. [57:37] O participante abre a recomendação que adiciona a variável WEATHER e o contador (gráfico de barras). [57:53] O participante adiciona a variável DATE para o campo COLUMN, isso gera um travamento na ferramenta. [58:10] O participante altera o *time unit* da variável DATE para YEAR e a ferramenta volta a funcionar. [58:35] O participante faz um *swap* entre as variáveis DATE e WEATHER. [58:52] O participante faz um *swap* entre os eixos X e Y, obtendo a visualização final.

T6 - [01:00:40] Sem apagar o mapeamento da Tabela 1, o participante navega pelas recomendações e abre a visualização que MEAN OF TEMP_MIN. [01:00:57] O participante move MEAN OF TEMP_MIN para o EIXO X e DATE para SIZE. [01:01:13] O participante remove a variável DATE do campo SIZE e o contador do EIXO Y. [01:01:22] O participante usa a função *undo* duas vezes seguidas para voltar a um estado anterior. [01:02:03] O participante remove a variável DATE do EIXO X e move a variável TEMP_MIN para o EIXO X. [01:02:12] O participante move a variável WEATHER para o EIXO X, remove o contador e faz um *swap* entre os eixos X e Y que ficam, respectivamente, com as variáveis WEATHER e MEAN OF TEMP_MIN. O participante encontra a visualização desejada.

T7 - [01:03:27] O participante navega pelas recomendações e abre a visualização que adiciona a variável LOCATION para o campo ROW. [01:04:17] O participante remove a variável TEMP_MIN do EIXO Y. [01:04:25] O participante move a variável LOCATION para o EIXO Y, movendo-a logo em seguida para o campo COLUMN e adicionando o contador para o EIXO Y. O participante encontra a visualização final. [01:04:42] O participante tenta aplicar funções ao contador e isso lhe atrapalha. O participante precisa remover o contador e adicioná-lo novamente para obter a visualização final.

T8 - [01:06:13] O participante navega pelas recomendações e abre a visualização que adiciona MEAN OF TEMP_MIN para o campo COLOR e encontra a visualização final, apesar de utilizar um contador desnecessário.

P03

VisMaker - CAPES-RJ

T1 - [10:55] O participante iniciou a utilização da ferramenta VisMaker e está rolando o painel de perguntas relacionadas. [11:03] O participante adicionou a variável QNT MASTERS para EIXO X. [11:44] O participante adicionou a variável YEAR para o EIXO Y. [12:41] O participante aplicou a função SUM à variável QNT MASTERS no EIXO X. [13:47] O participante removeu todo o mapeamento de variáveis. [13:55] O participante adicionou a variável QNT MASTERS para o EIXO Y. [14:03] O participante adicionou a variável YEAR para o EIXO X. [14:18] O participante aplicou a função SUM na variável QNT MASTERS no EIXO Y e encontrou um gráfico similar ao anteriormente construído, porém com uma orientação na horizontal. [15:00] Sob minha orientação, o participante aplica a *time unit year* na variável YEAR e encontra com precisão o gráfico que o faz responder a pergunta. O participante lembra que o dataset se refere ao estado do Rio de Janeiro. [15:47] O participante adiciona a variável CITY ao campo de filtros. [16:03] Após se confundir um pouco com o campo de filtros, o participante aplica o filtro que seleciona apenas os dados referentes à cidade do RIO DE JANEIRO e encontra a visualização final. Um bug fez com que os rótulos dos eixos da visualização desaparecesse após a aplicação do filtro.

T2 - [18:37] O participante adiciona a variável QNT DOCTORAL para o EIXO Y. [18:46] O participante adiciona (aparentemente sem querer após mudar de ideia) a variável FIELD para o campo de filtros e remove a variável CITY que havia ficado no campo devido à pergunta anterior. [19:02] Após se confundir novamente com a utilização do campo de filtros, o usuário aplica uma filtragem para os valores CIÊNCIAS HUMANAS da variável FIELD. [19:11] O participante aplica a função SUM à variável QNT DOCTORAL no EIXO Y, porém não compreende bem a função do contador na visualização gerada. [20:11] O participante remove a variável FIELD do campo de filtros e a adiciona para o EIXO X. [20:17] O participante aplica a função SUM à variável QNT DOCTORAL no EIXO Y e obtém a visualização que responde a pergunta.

T3 - [21:00] O participante remove a variável QNT DOCTORAL do EIXO Y, utilizada na pergunta anterior. [21:11] O participante adiciona a variável CITY para o EIXO Y. [21:19] O participante adiciona a variável FIELD para o campo de filtros. [21:27] Sem se confundir, ele aplica uma filtragem para selecionar apenas os dados de valor CIÊNCIAS AGRÁRIAS na variável FIELD. [22:02] Não convencido da resposta o participante remove a variável FIELD do EIXO X. [22:08] O participante adiciona a variável QNT POSTDOCTORAL no EIXO X. [22:22] O participante adiciona a variável FIELD ao campo de filtros e refaz o filtro anterior para CIÊNCIAS AGRÁRIAS. [22:30] O participante aplica a função SUM na variável QNT POSTDOCTORAL no EIXO X e encontra a visualização final.

T4 - [23:30] O participante adiciona a variável QNT MASTERS para o EIXO X. [23:32] O participante adiciona a variável FIELD para o EIXO Y. [23:48] O participante aplica a função SUM na variável QNT MASTERS no EIXO X e encontra a visualização que responde a pergunta.

T5 - [26:10] O participante adiciona a variável CITY ao EIXO X. [26:21] O participante adiciona a variável FIELD ao EIXO Y e encontra a visualização final sem navegar pelas recomendações. Apesar de encontrar a visualização correta, o participante se confunde e erra ao anotar as respostas na tabela.

T6 - [28:34] O participante mantém o mapeamento anterior e adiciona a variável JURIDICAL STATUS ao campo COLOR. [28:55] O participante remove a variável JURIDICAL STATUS do campo COLOR. [29:20] O participante abre o segundo gráfico que responde a pergunta sobre as co-ocorrências de CITY e FIELD (gráfico normalizado de barras empilhadas). [29:32] O participante remove a variável FIELD do campo COLOR e navega pelas perguntas relacionadas. [29:54] O participante remove a variável CITY do EIXO Y e recomeça o mapeamento do início. [30:02] O participante adiciona a variável JURIDICAL STATUS ao EIXO X. [30:35] O participante navega pelas recomendações e abre o primeiro gráfico da primeira pergunta (equivalente ao que já estava sendo visualizado) sobre a variável JURIDICAL STATUS. [31:00] O participante navega pelas recomendações e abre o terceiro gráfico da pergunta que relaciona as variáveis JURIDICAL STATUS e FIELD (*heatmap*) e encontra a visualização final.

T7 - [32:56] O participante remove a variável FIELD do EIXO Y e navega pelas recomendações. [33:14] O participante abre o primeiro gráfico da pergunta que relaciona as variáveis JURIDICAL STATUS e CITY (gráfico de barras empilhadas) e o utiliza para preencher a tabela.

T8 - [37:02] O participante adiciona a variável FIELD ao EIXO X e remove a variável JURIDICAL STATUS do EIXO Y. [37:05] O participante navega pelas recomendações e abre o gráfico da pergunta que relaciona as variáveis MEAN OF QNT MASTERS, FIELD e CITY. [37:31] O participante aplica a função MAX sobre a variável QNT MASTERS no campo COLOR. [38:41] O participante corrige e aplica a função SUM ao invés de MAX, obtendo a visualização final.

Voyager - WEATHER

T1 - [50:52] O participante abre a recomendação que adiciona a variável DATE em MONTH e o contador. [51:17] Após navegar pelas recomendações, o participante abre a recomendação que adiciona a variável MEAN OF TEMP_MAX para o campo SIZE. [51:42] O participante parece não ter gostado das visualizações e limpa todo o mapeamento. [51:50] O participante adiciona a variável TEMP_MAX para o EIXO X. [51:58] O participante navega pelas recomendações e abre o gráfico que adiciona a variável DATE em YEAR. [52:15] O participante aplica a função MEAN na variável TEMP_MAX no EIXO X. [52:31] O participante pergunta sobre a possibilidade de inverter os eixos. Orientado por mim ele consegue realizar a inversão, fazendo com que a variável DATE em YEAR vá para o EIXO X e a variável TEMP_MAX vá para o EIXO Y. [52:48] O participante altera o *time unit* da variável DATE para MONTH e encontra a visualização final.

T2 - [53:55] Após ver algumas recomendações, o participante remove a variável TEMP_MAX do EIXO Y. [54:03] O participante adiciona a variável PRECIPITATION para o EIXO Y. [54:12] O participante aplica a função MEAN à variável PRECIPITATION. [54:20] O participante adiciona a variável LOCATION para o campo SIZE e logo em seguida a transfere para o campo COLOR, obtendo a visualização que responde a pergunta. O participante pergunta sobre a possibilidade de dar um zoom na visualização e acaba errando a pergunta por causa da escala do gráfico.

T3 - [56:08] O participante remove a variável DATE do EIXO X. [56:17] O participante transfere a variável LOCATION do campo COLOR para o EIXO X, obtendo a resposta.

T4 - [56:58] O participante abre a lista de bookmarks e refaz a visualização que utilizou para responder a Pergunta 2. [57:11] O participante adiciona a variável LOCATION para o campo de filtro e seleciona somente os dados referentes a cidade de SEATTLE. Apesar de obter a visualização correta, o participante erra a resposta por conta da escala do gráfico.

T5 - [58:26] O participante adiciona a variável WEATHER para o EIXO X. [59:00] O participante adiciona a variável DATE para o EIXO Y. [59:08] O participante muda o *time unit* da variável DATE para YEAR. O participante navega pelas recomendações e abre a primeira recomendação que adiciona o COUNT para o campo SIZE, obtendo a visualização final.

T6 - [01:00:51] O participante remove a variável DATE do EIXO Y. [01:00:56] O participante abre a recomendação que adiciona BIN OF TEMP_MIN para o EIXO Y. [01:01:13] O participante aplica a função MEAN na variável TEMP_MIN e obtém a visualização final com um COUNT desnecessário.

T7 - [01:03:03] O participante abre uma recomendação da seção *alternative encodings*. [01:03:26] O participante remove a variável TEMP_MIN do campo COLOR. [01:03:35] O participante abre a recomendação que adiciona a variável LOCATION para o campo COLOR e encontra a visualização final.

T8 - [01:04:58] O participante navega pelas recomendações e abre a visualização que adiciona BIN OF TEMP_MIN para o campo ROW. [01:05:31] O participante abre uma visualização da seção *alternative encodings*. [01:05:48] O participante aplica a função MENA na variável TEMP_MIN no EIXO X e encontra a visualização final com um COUNT desnecessário.

P04

VisMaker - CAPES-RJ

T1 - [11:27] O participante faz a leitura da pergunta. [12:08] O participante adiciona a variável QNT MASTERS para o EIXO X. [12:29] O participante adiciona por acidente a variável CITY para o EIXO Y. [12:50] O participante remove a variável QNT MASTERS do EIXO X e adiciona a variável YEAR em seu lugar. [13:31] O participante abre a primeira visualização da primeira pergunta recomendada (gráfico de linhas por cidade ao longo dos anos). [13:40] O participante se engana com a visualização obtida e acha que ela pode responder a pergunta, no entanto, falta a informação sobre bolsas de mestrado. [14:21] O participante percebe que se enganou e tenta corrigir sua resposta através de um filtro. [14:30] O participante adiciona a variável QNT MASTER para o campo de filtros. [15:36] O participante decide recomeçar a tarefa pois nota que ainda não conseguiu responder a pergunta. [15:49] O participante adiciona a variável QNT MASTER para o EIXO X. [16:03] O participante adiciona a variável YEAR para o EIXO Y. [17:00] Após ser lembrado por mim sobre a estrutura do dataset, o participante aplica a função SUM na variável QNT MASTERS no EIXO X. [17:23] O participante adiciona a variável CITY para o campo de filtro. [17:40] Após se confundir um

pouco com a funcionalidade de filtro, o participante consegue aplicar um filtro para selecionar somente os dados referentes ao valor RIO DE JANEIRO e encontra a visualização final.

T2 - [19:07] O participante faz a leitura da pergunta. [19:29] O participante adiciona a variável FIELD para o campo de filtro e seleciona somente os dados com valor CIÊNCIAS HUMANS. [20:44] O participante adiciona a variável QNT DOCTORAL para o EIXO Y. [21:00] O participante aplica a função SUM à variável QNT DOCTORAL. Esta visualização, apesar de ter um COUNT desnecessário, responde a pergunta, pode ser considerada correta, mas o participante não a utiliza. [21:31] O participante abre a recomendação que apresenta o histograma da variável QNT DOCTORAL. [21:40] Não satisfeito com a visualização, o participante aplica novamente a função SUM à variável QNT DOCTORAL. [21:48] O participante remove a variável FIELD do campo de filtro e a adiciona ao EIXO Y e encontra a visualização correta. O participante desconfia de sua resposta por achar que ela não está de acordo com a visualização criada anteriormente. Tal dúvida ocorre por conta da existência de um COUNT desnecessário na visualização anterior.

T3 - [24:08] O participante faz a leitura da pergunta. [24:29] O participante adiciona a variável QNT POSTDOCTORAL para o EIXO Y. [24:33] O participante adiciona a variável CITY para o EIXO X. [24:41] O participante adiciona a variável FIELD para o campo de filtro. [24:49] O participante aplica o filtro para selecionar somente os dados com valor CIÊNCIAS AGRÁRIAS. [25:06] O participante aplica a função SUM à variável QNT POSTDOCTORAL e encontra a visualização correta.

T4 - [26:50] O participante faz a leitura da pergunta. [27:03] O participante adiciona a variável FIELD para o EIXO Y. [27:09] O participante adiciona a variável QNT MASTERS para o EIXO X. [27:15] O participante aplica a função SUM à variável QNT MASTERS e encontra a visualização final.

T5 - [29:52] O participante adicionou a variável CITY para o EIXO X. [30:04] O participante adiciona a variável FIELD para o EIXO Y e navega muito rapidamente pelas recomendações. [30:43] O participante abre o primeiro gráfico da primeira pergunta relacionada, porém a visualização não lhe agrada. O participante pergunta se é possível voltar (*undo*) e se enrola um pouco tentando obter a visualização anterior. [31:10] O participante adiciona a variável FIELD para o EIXO X e encontra a visualização final. [31:52] O participante chega a navegar pelas recomendações e até a abrir uma recomendação para tentar visualizar melhor, mas retorna para a visualização anterior (*heatmap*).

T6 - [35:06] O participante lê a tabela. [56:14] O participante adiciona a variável JURIDICAL STATUS para o EIXO X. [35:24] O participante adiciona a variável FIELD para o EIXO Y. [35:34] O participante abre o segundo gráfico das recomendações (gráfico normalizado de barras empilhadas). O participante utiliza este gráfico para começar a responder a pergunta, mas depois prefere voltar para o *heatmap*. Se confunde com isso, mas consegue.

T7 - [41:10] O participante adiciona a variável CITY para o EIXO Y. [41:20] O participante adiciona a variável JURIDICAL STATUS para o EIXO X e encontra a visualização final.

T8 - [43:24] O participante adiciona a variável FIELD para o EIXO X. [43:36] O participante adiciona a variável CITY para o EIXO Y. [44:00] O participante navega pelas recomendações e abre a visualização que adiciona a variável MEAN OF QNT MASTERS (*heatmap*). [44:06] O participante aplica a função SUM na variável QNT MASTERS e obtém a visualização final.

Voyager - WEATHER

T1 - [01:00:23] O participante lê a pergunta. [01:01:29] O participante adiciona a variável TEMP_MAX para o EIXO Y. [01:01:54] O participante adiciona a variável DATE para o EIXO X. [01:02:11] O participante abre a segunda visualização recomendada que apresenta a relação entre DATE em MONTH e MEAN OF TEMP_MAX e encontra a resposta para a pergunta.

T2 - [01:04:16] O participante adiciona a variável PRECIPITATION para o EIXO Y. [01:04:24] O participante adiciona a variável CITY para o EIXO X. [01:04:49] O participante aplica a função MEAN à variável PRECIPITATION. [01:05:46] Após se perguntar onde adicionaria a variável DATE, o participante a adiciona para o campo SIZE. [01:05:56] O participante muda o *time unit* de DATE para MONTH. [01:06:04] O participante remove a variável DATE do campo SIZE e a adiciona ao campo COLOR. [01:06:29] O participante remove a variável DATE de COLOR e a adiciona em SHAPE. [01:06:46] O participante encontra o painel de *alternative encodings* e abre a primeira recomendação (gráfico de linhas). O participante encontra o gráfico correto mas tem problemas com a escala e tenta encontrar, mudando o *mark*, um tipo de gráfico com melhor escala. A busca não é bem-sucedida. O problema da escala faz com que o participante erre a pergunta.

T3 - [01:10:23] O participante lê a pergunta. [01:10:48] O participante adiciona a variável PRECIPITATION para o EIXO Y. [01:10:56] O participante aplica a função MEAN à variável PRECIPITATION. [01:11:13] O participante adiciona a variável LOCATION para o EIXO X. Mais uma vez o participante tem problemas com a escala do gráfico.

T4 - [01:12:29] O participante lê a pergunta. [01:12:38] O participante adiciona a variável PRECIPITATION para o EIXO Y. [01:12:46] O participante adiciona a variável LOCATION para o campo de filtro e seleciona somente os dados referentes a SEATTLE. [01:12:46] O participante adiciona a variável DATE para o EIXO X e logo em seguida muda seu *time unit* para MONTH. [01:13:03] O participante abre o gráfico da segunda recomendação que aplica MEAN à variável PRECIPITATION. Novamente o participante tem problema com a escala do gráfico. O participante esquece de aplicar um filtro para selecionar somente os dados referentes à cidade SEATTLE. Isso o faz errar a questão.

T5 - [01:15:41] O participante abre a recomendação que adiciona a variável WEATHER e o contador. [01:15:49] O participante adiciona a variável DATE para o EIXO X, substituindo o contador e muda o *time unit* para YEAR. [01:16:14] O participante abriu a recomendação que adiciona o contador (*"bubble map"*) e encontrou a visualização final.

T6 - [01:17:34] O participante adiciona a variável WEATHER para o EIXO Y. [01:17:38] O participante adiciona a variável TEMP_MIN para o EIXO X e aplica a função MEAN e encontra a resposta.

T7 - [01:19:34] O participante adiciona a variável WEATHER para o EIXO X. [01:19:43] O participante adiciona a variável LOCATION para o EIXO Y. [01:20:08] O abre a primeira recomendação, que adiciona o contador (*"bubble map"*) e encontra a solução para a tabela.

T8 - [01:21:14] O participante adiciona a variável WEATHER para o EIXO Y. [01:21:23] O participante adiciona a variável MIN_TEMP para o EIXO X. [01:21:23] O participante inverte os eixos X e Y. [01:21:31] O participante abre a primeira recomendação, que aplica MEAN na variável TEMP_MIN. [01:22:13] O participante adiciona a variável LOCATION para o campo COLOR, depois para SIZE e SHAPE. [01:22:38] Não satisfeito com os resultados, o participante põe LOCATION no EIXO Y e MEAN OF TEMP_MIN vai para SHAPE e desfaz a operação em seguida. [01:23:19] O participante limpa todo o mapeamento e inicia novamente. [01:23:28] O participante adiciona a variável LOCATION no EIXO X. [01:23:44] O participante adiciona TEMP_MIN no EIXO Y, mas a remove logo em seguida. [01:23:53] O participante adiciona a variável WEATHER no EIXO Y. [01:23:54] O participante adiciona a variável TEMP_MIN para o campo SIZE, trocando-a com WEATHER logo em seguida. [01:24:09] O participante limpa o mapeamento novamente e recomeça. [01:24:26] O participante adiciona a variável LOCATION para o EIXO X. [01:24:34] O participante adiciona TEMP_MIN para o EIXO Y e aplica a função MEAN. [01:26:31] Depois de várias tentativas, o participante consegue adicionar a variável WEATHER para o campo COLUMN e encontra a resposta.

GRUPO 2

P05

VisMaker - WEATHER

T1 - [08:49] O participante adiciona a variável DATE para o EIXO X. [09:18] O participante adiciona a variável TEMP_MAX para o EIXO Y e aplica a função MEAN. [09:41] O participante seleciona o *time unit* MONTH para a variável DATE e encontra a visualização final.

T2 - [11:38] O participante adiciona a variável PRECIPITATION para o EIXO Y. [11:43] O participante adiciona a variável DATE para o EIXO X. [11:56] O participante aplica a função MEAN à variável PRECIPITATION. [11:59] O participante adiciona a variável LOCATION para o campo COLOR. [12:19] O participante verifica as recomendações, e abre o primeiro gráfico da pergunta que relaciona MEAN OF PRECIPITATION, LOCATION e DATE em MONTHS (gráfico de linhas), encontrando a visualização final. Apesar de ter encontrado a visualização correta, o participante comete um erro no momento de escrever sua resposta, esquecendo-se do mês de julho.

T3 - [13:47] O participante lê a pergunta e talvez a tenha interpretado de maneira equivocada ao perguntar “*over the years?*”. [13:55] O participante abre o gráfico da primeira pergunta relacionada que responde a pergunta “What is the MEAN OF PRECIPITATION in each category of LOCATION over the YEARS?” (gráfico de linhas). Ele utiliza esse gráfico para responder a pergunta e acaba errando a pergunta por conta dessa ambiguidade.

T4 - [14:33] O participante lê a pergunta. Com o mapeamento da questão anterior, o participante abre o primeiro gráfico da primeira pergunta, que relaciona MEAN OF PRECIPITATION, LOCATION e DATE em MONTHS (gráfico de linhas), o mesmo usado para responder a Pergunta 2 e o utiliza para responder esta pergunta.

T5 - [16:29] O participante remove a variável LOCATION do campo COLOR e adiciona a variável WEATHER em seu lugar. [17:22] O participante remove a variável PRECIPITATION do EIXO Y. [17:30] O participante adiciona a variável WEATHER no EIXO Y. [17:50] O participante altera o *time unit* da variável DATE para YEAR e, sem compreender bem o que estava acontecendo com a visualização, altera para NO TIME UNIT e volta para YEAR sem seguida. [18:48] Após remover uma das variáveis WEATHER (que estava mapeada duas vezes), o participante abre a primeira visualização da primeira pergunta recomendada (gráfico de linhas) e encontra a visualização final. [19:13] Ao ver o mapeamento obtido, o participante fala que se confundiu um pouco e pergunta “esse COUNTER saiu da onde?”.

T6 - [21:24] Sem reiniciar o mapeamento, o participante abre o primeiro gráfico da pergunta que relaciona as variáveis MEAN OF TEMP_MIN, WEATHER e DATE em YEARS (gráfico de linhas). [22:00] O participante observa a visualização e percebe que ela não responde a pergunta, questiona sobre a possibilidade de voltar (talvez um *undo*). [22:22] O participante remove a variável DATE do EIXO X. [22:37] O participante abre o primeiro gráfico da primeira pergunta que relaciona TEMP_MIN e WEATHER (*boxplot*). [22:48] O participante observa o gráfico e percebe que ele também não responde a pergunta. [23:16] O participante abre o primeiro gráfico da pergunta que relaciona MEAN OF TEMP_MIN e WEATHER (gráfico de barras) e encontra a visualização final. O participante não parece convencido e navega novamente pelas recomendações, chega a abrir novamente o *boxplot* obtido anteriormente, mas em seguida chega novamente ao gráfico de barras.

T7 - [26:34] O participante adiciona a variável WEATHER para o EIXO X. [26:48] O participante adiciona a variável LOCATION para o EIXO Y. [27:17] O participante abre o primeiro gráfico da pergunta que relaciona as variáveis WEATHER e LOCATION (gráfico de barras empilhadas na horizontal) e o utiliza para preencher a tabela.

T8 - [29:01] O participante navega pelas recomendações e abre o gráfico da pergunta que relaciona MEAN OF TEMP_MIN, WEATHER e LOCATION (*heatmap*) e o utiliza para preencher a tabela.

Voyager - CAPES-RJ

T1 - [42:08] O participante faz a leitura da pergunta. [42:16] O participante adiciona a variável QNT MASTERS no EIXO Y. [42:25] O participante adiciona a variável YEAR ao EIXO X. [42:28] O participante demonstra surpresa ao visualizar o gráfico gerado e diz que “esperava um gráfico de linhas”. [43:06] O participante aplica a função SUM à variável QNT MASTERS. [43:39] Sob minha orientação, o participante muda o *time unit* da variável YEAR para YEAR. [44:13] O participante adiciona a variável CITY ao campo de filtro e seleciona somente os dados referentes à cidade do RIO DE JANEIRO obtendo a visualização final.

T2 - [45:20] O participante lê a pergunta. [45:45] Seguindo da pergunta anterior, o participante abre o gráfico da primeira recomendação, que adiciona MEAN OF QNT DOCTORAL em SIZE. [46:35] O participante adiciona a variável FIELD para o campo de filtro e seleciona somente os dados referentes ao valor CIÊNCIAS HUMANAS. [47:33] O participante remove a variável QNT MASTERS e move MEAN OF QNT DOCTORAL para o EIXO Y. [48:06] O participante adiciona o contador para o EIXO X, substituindo a variável YEAR e utiliza esta visualização para responder a pergunta, porém

erra. O participante anotou como resposta o valor do COUNT. Além disso, esqueceu de aplicar a função SUM à variável QNT DOCTORAL.

T3 - [49:38] O participante adiciona a variável CITY para o EIXO Y. [49:46] O participante adiciona a variável QNT POSTDOCTORAL para o EIXO X e aplica a função SUM. [50:11] O participante aplica o filtro sobre a variável FIELD para manter somente os dados referentes a CIÊNCIAS AGRÁRIAS e encontra a visualização final.

T4 - [51:11] O participante adiciona a variável FIELD para o campo COLOR. [51:18] O participante adiciona a variável QNT MASTERS para o EIXO X e aplica a função SUM. [51:26] O participante adiciona o contador para o EIXO Y e utiliza a visualização construída para responder a pergunta, porém erra a resposta pois observa o eixo do contador ao invés da variável QNT MASTERS.

T5 - [53:31] O participante adiciona a variável FIELD para o campo COLOR. [54:30] O participante adiciona o contador para o EIXO X. [54:38] O participante transfere a variável FIELD de COLOR para EIXO Y. [54:55] O participante abre a recomendação que adiciona a variável CITY no campo COLOR (gráfico de barras empilhadas). [55:20] O participante faz um *swap* entre as variáveis CITY e FIELD. [55:28] O participante faz um *swap* entre a variável CITY e o contador. [56:18] O participante adiciona a variável CITY para o campo de filtros e seleciona somente os dados referentes às cidades de interesse. [57:08] O participante abre uma recomendação no painel de *alternative encodings* ("bubble map"). Ao abrir a visualização, o filtro é perdido. O participante refaz a filtragem e utiliza a visualização para preencher a tabela.

T6 - O participante abre a recomendação que adiciona a variável JURIDICAL STATUS para o campo ROW. [59:55] O participante remove a variável CITY e move o contador para o EIXO Y e chega à visualização final.

T8 - [01:02:44] O participante abre a recomendação que adiciona a variável CITY para o campo COLOR. [01:03:01] O participante remove a variável JURIDICAL STATUS do campo ROW. [01:03:09] O participante faz um *swap* entre a variável CITY e o contador. [03:03:34] O participante adiciona a variável CITY para o campo de filtros e seleciona somente os dados referentes às cidades de interesse. [01:04:08] O participante remove a variável CITY e adiciona em seu lugar a variável JURIDICAL STATUS. [01:04:16] O participante adiciona novamente a variável CITY para o EIXO X, removendo a variável FIELD e obtém a visualização final.

T5 - [01:05:23] A partir do mapeamento da tabela anterior, o participante abre a recomendação que adiciona BIN OF QNT MASTERS. [01:06:21] O participante faz um *swap* entre as variáveis JURIDICAL STATUS e BIN OF QNT MASTERS e remove JURIDICAL STATUS em seguida. [01:06:38] O participante adiciona a variável FIELD ao campo SIZE, removendo o contador. [01:06:54] O participante não compreende bem a visualização e muda a variável FIELD para o campo COLOR. [01:07:03] O participante faz um *swap* entre as variáveis FIELD e BIN OF QNT MASTERS. [01:07:11] O participante troca a função de BIN para SUM na variável QNT MASTERS. [01:08:28] O participante faz um *swap* entre as variáveis FIELD e CITY. [01:08:54] Orientado por mim, o participante muda o *mark* para o tipo *RECT*. [01:09:34] O participante adiciona a variável CITY ao campo de filtro e seleciona somente as cidades desejadas, obtendo a visualização final.

P06

VisMaker - WEATHER

T1 - [13:28] O participante navega pelas recomendações. [13:45] O participante adiciona a variável TEMP_MAX ao EIXO X. [14:41] O participante navega pelas recomendações e abre o terceiro gráfico da pergunta que relaciona as variáveis TEMP_MAX e DATE (gráfico de áreas). [15:48] Orientado por mim, o participante altera o *time unit* da variável DATE para MONTH e encontra a visualização final.

T2 - [17:52] O participante adiciona a variável PRECIPITATION ao EIXO X. [18:36] O participante adiciona variável LOCATION para o EIXO Y. [18:53] O participante faz um *swap* entre os eixos X e Y. [19:23] O participante navega pelas recomendações e abre o primeiro gráfico da pergunta que relaciona MEAN OF PRECIPITATION, LOCATION e DATE em YEARS e em seguida altera o *time unit* da variável DATE para MONTH, obtendo a visualização final.

T3 - [22:24] O participante adiciona a variável LOCATION para o EIXO X mantendo parte do mapeamento da questão anterior e encontra a visualização final.

T4 - [23:14] O participante adiciona a variável LOCATION para o campo de filtros e seleciona somente os valores referentes à cidade de SEATTLE. [23:35] O participante adiciona a variável DATE para o EIXO Y, removendo a variável PRECIPITATION, e faz um *swap* entre os eixos X e Y logo em seguida. [23:49] O participante adiciona novamente a variável PRECIPITATION para o EIXO Y e aplica a função MEAN. [24:07] O participante altera o *time unit* da variável DATE para MONTH e encontra a visualização final.

T5 - [26:22] O participante adiciona a variável WEATHER para o EIXO Y. [28:02] Após navegar pelas recomendações, o participante abre o primeiro gráfico da pergunta que relaciona WEATHER e DATE em MONTHS (gráfico de linhas). [28:08] O participante altera o *time unit* da variável DATE para YEAR e encontra a visualização final. A presença de um filtro utilizado na pergunta anterior faz com que o participante tenha de refazer a visualização.

T6 - [31:12] O participante navega pelas recomendações e abre o gráfico que adiciona a variável TEMP_MIN. [32:02] O participante move a variável WEATHER para o EIXO X, removendo a variável DATE que estava lá por conta da questão anterior. [32:43] O participante remove a variável LOCATION que estava no campo de filtros desde a Pergunta 4 e encontra a visualização final.

T7 - [26:22] Tendo o mapeamento utilizado para a Tabela 1, o participante adiciona a variável LOCATION para o campo de FILTROS. [38:09] O participante adiciona a variável WEATHER no EIXO X, removendo a variável DATE e aplica separadamente os filtros em cada cidade para preencher a tabela.

T8 - [41:31] O participante adiciona a variável TEMP_MIN para o EIXO Y e aplica a função MEAN logo em seguida. Mais uma vez, utilizando o campo de filtros com a variável LOCATION, o participante preenche a tabela através de duas visualizações distintas, uma para cada cidade.

Voyager - CAPES-RJ

T1 - [57:18] O participante lê a pergunta. [58:05] O participante abre a recomendação que adiciona BIN OF QNT MASTERS para o EIXO X. [58:55] O participante adiciona a variável YEAR para o EIXO Y. [59:12] O participante faz um *swap* entre os eixos X e Y. [01:00:02] O participante aplica a função SUM à variável QNT MASTERS. [01:00:52] O participante percebe que tem que utilizar a variável CITY e a adiciona para o EIXO Y. [01:01:08] O participante adiciona a variável CITY para o campo de filtros e seleciona somente os dados referentes à cidade do RIO DE JANEIRO. [01:01:33] O participante adiciona novamente a variável QNT MASTERS para o EIXO Y e aplica novamente a função SUM, obtendo a visualização final.

T2 - [01:02:32] O participante adiciona a variável QNT DOCTORAL para o EIXO Y. [01:02:57] O participante adiciona a variável FIELD para o campo COLOR. [01:03:05] O participante adiciona a variável FIELD para o campo de filtros e seleciona somente os dados referentes ao valor CIÊNCIAS HUMANAS. [01:03:30] O participante aplica a função SUM à variável QNT DOCTORAL. [01:04:45] O participante remove a variável YEAR do EIXO X e encontra a visualização final.

From here the transcriptions were done in less detail

T8 - O participante teve dificuldades em perceber que precisaria aplicar a função SUM à variável QNT MASTERS, mas conseguiu completar a tarefa.

P07

VisMaker - WEATHER

T1 - O participante responde essa pergunta com bastante tranquilidade.

T2 - O participante respondeu a pergunta com tranquilidade. Utilizou uma recomendação que relacionava DATE em YEARS, MEAN OF PRECIPITATION e LOCATION, depois alterou o *time unit* de DATE para MONTH e respondeu a pergunta.

T4 - O participante responde a pergunta utilizando a mesma visualização da Pergunta 2.

T3 - O participante responde a pergunta de forma rápida, apenas removendo a variável DATE do encoding.

T5 - O participante respondeu a pergunta com tranquilidade. Utilizou as recomendações, fez a leitura da pergunta e abriu o gráfico correto.

T7 - O participante responde a pergunta com tranquilidade.

T8 - Utilizando o mapeamento da Pergunta 3, o participante abre a recomendação que adiciona MEAN OF TEMP_MIN e encontra a visualização final.

T6 - Partindo da Pergunta 4, o participante remove a variável CITY e obtém um gráfico que já responde a pergunta, porém utilizando um contador desnecessário. O participante navega pelas recomendações e abre o gráfico mais preciso com a pergunta que relaciona MEAN OF TEMP_MIN e WEATHER.

Voyager - CAPES-RJ

T1 - O participante erra a pergunta por ignorar o ano requerido pela questão.

T4 - O participante responde a Pergunta 4 logo após a Pergunta 1, aproveitando seu mapeamento.

T2 - O participante respondeu a pergunta com facilidade.

T3 - Apesar da resposta correta, a visualização criada pelo participante não responde a pergunta. Está errada.

T5 - O participante encontra uma visualização correta, mas acaba errando uma das células da tabela no processo de conferir as respostas.

T8 - O participante usa o mapeamento realizado para a Tabela 1 e rapidamente constrói uma visualização que lhe permite preencher a Tabela 4.

T7 - Apesar da resposta estar correta, o participante construiu uma visualização errada utilizando SUM OF QNT MASTERS ao invés de COUNT. O participante teve que realizar alguns filtros para visualizar bem os valores.

T6 - O participante construiu uma visualização correta, mas errou ao analisá-la.

P08

VisMaker - WEATHER

T1 - O participante inicia adicionando as variáveis DATE e TEMP_MAX corretamente, mas abre a recomendação da pergunta que adiciona a variável WEATHER e utiliza esse gráfico para responder, errando a resposta.

T2 - O participante adiciona as variáveis DATE e PRECIPITATION e navega até as recomendações para encontrar a visualização final.

T4 - O participante usa a mesma visualização construída na Pergunta 2 para responder essa pergunta.

T3 - O participante remove a variável DATE e adiciona LOCATION em seu lugar no EIXO X, obtendo a visualização final.

T5 - O participante sente certa dificuldade em preencher a tabela. A presença duplicada da variável WEATHER no mapeamento o atrapalha e ele chega a construir uma visualização errada que utiliza a

variável PRECIPITATION. Depois de um tempo, o participante percebe e remove essa variável do mapeamento, obtendo a visualização correta.

T6 - O participante adiciona a variável WEATHER e abre a recomendação que adiciona a variável TEMP_MAX, em seguida, remove TEMP_MAX e adiciona TEMP_MIN em seu lugar, depois aplica a função MEAN e obtém a visualização final.

T7 - O participante adiciona a variável LOCATION, navega pelas recomendações e chega a abrir uma recomendação que já seria suficiente para responder a pergunta, mas não parece convencido disso. Reinicia o processo e navega pelas recomendações até que encontra um *heatmap* que relaciona LOCATION e WEATHER.

T8 - O participante abriu a recomendação de MEAN OF TEMP_MIN em cada WEATHER e aplicou filtros para visualizar os dados de cada cidade e preencher a tabela.

Voyager - CAPES-RJ

T1 - O participante adicionou a variável QNT MASTERS. No início tentou realizar uma filtragem por cidade. Adicionou também a variável YEAR para o campo de filtros, mas a impossibilidade de rolar para configurar o filtro fez com que ele removesse a variável CITY dos filtros. O participante aplicou filtragem sobre a variável YEAR e depois adicionou a variável CITY para o EIXO Y, obtendo a visualização final.

T2 - Apesar de confundir um pouco as variáveis (questão de atenção), o participante respondeu a pergunta com tranquilidade.

T3 - O participante responde a pergunta com tranquilidade, utilizando a variável FIELD no campo de filtros.

T4 - O participante erra a pergunta porque utiliza a função MAX na variável QNT MASTERS ao invés da função SUM.

T5 - O participante adicionou a variável FIELD e abriu a recomendação que adicionou o contador. Em seguida, adicionou a variável CITY para o campo de filtros e aplicou vários filtros para encontrar as respostas.

T6 - O participante utiliza novamente o campo de filtros para encontrar as respostas e preencher a tabela. Dessa vez os filtros são feitos na variável JURIDICAL STATUS.

T7 - Rapidamente, o participante organiza o mapeamento para, assim como nas tabelas anteriores, aplicar vários filtros e preencher essa tabela.

T8 - Como nas tabelas anteriores, o participante novamente tenta utilizar os filtros. O participante não consegue adicionar a variável QNT MASTERS. Quando percebe que precisa fazê-lo, a adiciona no campo de filtros, o que não influencia as visualizações que continuam utilizando COUNT no tamanho das barras.

GRUPO 3

P09

Voyager - CAPES-RJ

T1 - O participante adiciona corretamente as variáveis QNT MASTERS e CITY, mas sente dificuldade em adicionar a variável YEAR para a visualização. Ele a adiciona para o campo de filtros e seleciona somente o ano de 2008. Por fim, o participante adiciona o SUM OF QNT MASTERS também para o campo de texto a fim de que a visualização lhe permita visualizar o valor sem ter que passar o mouse.

T2 - O participante realiza o mapeamento com certa facilidade. Esquece, no entanto de remover a variável QNT MASTERS do campo de texto e isso quase o faz errar a questão. O participante percebe o erro e corrige a visualização, substituindo QNT MASTERS por QNT DOCTORAL.

T3 - Novamente o participante realiza o mapeamento com facilidade. Se esquece que precisa aplicar a função SUM à variável QNT POSTDOCTORAL, mas logo o faz.

T4 - O participante realiza o mapeamento com facilidade e mais uma vez se esquece de aplicar a função SUM e comenta que a aplicação de função deveria ser o padrão para valores quantitativos ao invés de mostrar a distribuição.

T5 - O participante utiliza as recomendações para encontrar a visualização correta seguindo o fluxo das recomendações.

T6 - O participante segue as recomendações e chega a encontrar uma visualização correta, mas tenta aplicar uma função de SUM no contador e acaba tendo que limpar todo o mapeamento. O participante segue novamente as recomendações e rapidamente encontra a visualização anterior. Utiliza, ainda, os *alternative encodings* para encontrar uma visualização considerada melhor.

T7 - O participante segue o fluxo das recomendações adicionando, em cada passo, cada uma das variáveis necessárias e encontra a visualização correta.

T8 - O participante tem certa dificuldade pois acaba acessando uma recomendação com função BIN OF QNT MASTERS, que dificulta sua busca por uma recomendação correta. O participante adiciona a variável CITY para o campo de filtros e realiza várias filtragens para preencher a tabela.

VisMaker - WEATHER

T1 - O participante realiza o mapeamento com certa facilidade e obtém a visualização correta.

T2 - O participante tenta adicionar duas vezes a variável LOCATION para o campo de filtros, utilizando duas filtragens diferentes, mas sem sucesso. Em seguida, move a variável LOCATION para o campo COLOR. Em seguida, adiciona a variável PRECIPITATION e aplica a função MEAN, obtendo a visualização final.

T3 - O participante, partindo da visualização construída na tabela anterior, remove a variável DATE e move a variável LOCATION para o EIXO X.

T4 - O participante utiliza a visualização construída para a Pergunta 2, abrindo-a da lista de *bookmarks*, para resolver essa questão.

T5 - O participante segue o fluxo das recomendações e encontra a visualização correta sem nenhuma dificuldade.

T6 - O participante segue da visualização anterior, adicionando, pelas recomendações, a variável MEAN OF TEMP_MIN. Depois de perceber que não era necessário utilizar a variável DATE, o participante tem certa dificuldade para encontrar a visualização correta e reinicia o mapeamento. Logo o participante consegue obter a visualização correta adicionando as variáveis requeridas.

T7 - O participante rapidamente consegue obter a visualização correta adicionando a variável ao mapeamento e utilizando as recomendações

T8 - Partindo da visualização anterior, o participante navega pelas recomendações e encontra exatamente a visualização necessária para essa tabela.

T10

Voyager - CAPES-RJ

T1 - O participante adicionou as variáveis ao mapeamento. Teve certa dificuldade em saber onde adicionaria variável YEAR. Navegou pelas recomendações e a adicionou ao campo COLOR, obtendo a visualização final.

T2 - O participante faz o mapeamento manualmente sem muita dificuldade.

T3 - O participante começa a fazer o mapeamento manualmente, sente certa dificuldade e procura pelas recomendações para finalizar a questão.

T4 - O participante realiza o mapeamento com facilidade e responde corretamente a pergunta.

T5 - O participante inicia o mapeamento e usa as recomendações e o painel de *alternative encodings* para encontrar a visualização desejada.

T6 - O participante inicia o mapeamento e usa as recomendações e o painel de *alternative encodings* para encontrar a visualização desejada.

T7 - O participante inicia o mapeamento e usa as recomendações e o painel de *alternativeencodings* para encontrar a visualização desejada. O participante ainda adiciona a variável CITY para o campo de filtros para selecionar somente as cidades de interesse.

T8 - O participante inicia o mapeamento e utiliza as recomendações para encontrar uma visualização que relaciona as 3 variáveis necessárias. Abre uma recomendação e aplica a função SUM à variável QNT MASTERS. O participante ainda aplica um filtro sobre a variável CITY para selecionar somente as cidades de interesse.

VisMaker - WEATHER

T1 - O participante adiciona as variáveis necessárias e recorre às recomendações, encontrando uma visualização correta. A leitura da pergunta parece ter lhe ajudado.

T2 - O participante adiciona parte do mapeamento e utiliza as recomendações para inserir a variável DATE, obtendo uma visualização correta.

T3 - O participante constrói todo o mapeamento manualmente com tranquilidade.

T4 - O participante constrói parte do mapeamento sozinho e utiliza as recomendações para adicionar a variável DATE na visualização, obtendo uma visualização correta.

T5 - O participante adiciona a variável WEATHER e usa as recomendações para adicionar a variável DATE, obtendo a visualização correta.

T6 - O participante adiciona a variável TEMP_MIN, navega pelas recomendações para adicionar a variável WEATHER e obtém um gráfico correto.

T7 - O participante inicia o mapeamento e usa a recomendação para adicionar a variável WEATHER.

T8 - O participante inicia o mapeamento e usa as recomendações para chegar à visualização final. Comenta não gostar muito da visualização encontrada (*heatmap*), mas a utiliza corretamente.

T11

Voyager - CAPES-RJ

T1 - O participante realiza o mapeamento com facilidade. Demora um pouco para perceber que precisaria aplicar um filtro por cidade, mas o faz.

T2 - Novamente o participante faz o mapeamento com facilidade.

T3 - O participante se enrola um pouco tentando adiciona a variável CITY para o campo de filtros, mas consegue refazer o mapeamento e chegar à visualização final.

T4 - O participante encontra a visualização correta com facilidade através do mapeamento.

T5 - O participante adiciona as variáveis manualmente e usa as recomendações para adicionar o contador.

T6 - Novamente o participante adiciona manualmente as variáveis e usa a recomendação para adicionar o contador.

T7 - Novamente o participante adiciona manualmente as variáveis e usa a recomendação para adicionar o contador. O participante ainda adiciona um filtro para selecionar somente as cidades de interesse.

T8 - O participante constrói todo o mapeamento manualmente, mas usa as recomendações para achar um *encoding* alternativo.

VisMaker - WEATHER

T1 - O participante navega pelas recomendações e abre a visualização que adiciona a variável DATE em MONTH. Adiciona desnecessariamente a variável WEATHER, removendo após um tempo. Adiciona variável TEMP_MAX para o campo COLOR e depois move-a para o EIXO X. Após isso, o participante vê as recomendações e abre a recomendação que adiciona aplica a função MEAN à variável TEMP_MAX, obtendo a resposta.

T2 - Partindo da visualização anterior, o participante adiciona a variável LOCATION para o campo COLOR e adiciona a variável PRECIPITATION para o EIXO Y, aplicando a função MEAN logo em seguida e finalizando a questão.

T3 - Com facilidade, o participante remove a variável DATE e refaz o mapeamento para as variáveis necessárias, obtendo a visualização correta.

T4 - O participante se enrola um pouco na aplicação do filtro, mas não tem grandes problemas e consegue responder bem a pergunta.

T5 - O participante altera o mapeamento, removendo o que é desnecessário e adicionando a variável WEATHER e logo acha, nas recomendações, uma visualização correta. O participante ainda aplica alguns filtros para facilitar a visualização.

T6 - O participante constrói manualmente essa visualização através do mapeamento.

T7 - O participante encontra uma visualização correta fazendo manualmente o mapeamento, mas vê as recomendações e utiliza outra visualização também correta.

T8 - O participante adiciona as variáveis necessárias para o mapeamento e usa a recomendação para obter a visualização final.

T12

Voyager - CAPES-RJ

T1 - O participante inicia o mapeamento com certa facilidade. Tenta realizar um filtro na variável YEAR, mas se enrola um pouco. Ao tentar adicionar a variável CITY, o participante acaba removendo a variável YEAR e sente dificuldade ao adicioná-la novamente. O participante navega pelas recomendações e, inspirado por uma das recomendações, adiciona a variável YEAR ao campo COLOR. Em seguida, adiciona a variável YEAR para o campo de filtros e seleciona somente o ano de 2008, obtendo a visualização final.

T2 - O participante realiza o mapeamento com certa facilidade. Tem dificuldade em perceber que ainda precisa remover a variável YEAR, advinda da resposta anterior, mas o faz e encontra a visualização correta.

T3 - O participante tem alguma dificuldade ao realizar o mapeamento e encontra uma visualização pouco comum, porém correta.

T4 - O participante realiza o mapeamento com facilidade e encontra a resposta correta.

T5 - O participante inicia o mapeamento e logo encontra a recomendações que a ajuda a finalizá-lo.

T6 - Novamente, o participante encontra a visualização correta com facilidade.

T7 - Novamente, o participante encontra a visualização correta com facilidade.

T8 - O participante sentiu bastante dificuldade e, em alguns momentos, as recomendações não o ajudaram pois vinham sempre com um contador e traziam a variável QNT MASTERS como BIN. O participante soube perceber que o contador era desnecessário e manualmente o removeu e adicionou SUM OF QNT MASTERS em seu lugar, obtendo a visualização correta.

VisMaker - WEATHER

T1 - O participante faz o mapeamento com facilidade e contra a visualização correta.

T2 - O participante novamente realiza o mapeamento com facilidade.

T3 - O participante realiza o mapeamento com facilidade.

T4 - O participante realiza o mapeamento com facilidade.

T5 - O participante navega pelas recomendações, abrindo a visualização que adiciona a variável DATE. Navega novamente até encontrar a recomendação que adiciona a variável WEATHER e responde a tabela.

T6 - O participante abre a recomendação que adiciona a variável TEMP_MIN e navega pelas recomendações até encontrar a visualização que adiciona WEATHER e MEAN OF TEMP_MIN.

T7 - O participante segue o fluxo das recomendações, sempre fazendo a leitura das perguntas e encontra a visualização correta.

T8 - O participante segue as recomendações, adiciona manualmente a variável LOCATION para o campo COLOR e abre a recomendação que relaciona as variáveis em questão, obtendo a resposta correta.

GRUPO 4

T13

Voyager - WEATHER

T1 - O participante realiza o mapeamento sem muita dificuldade e encontra a visualização final.

T2 - O participante consegue realizar o mapeamento com facilidade. Chega a construir a visualização que responderia a Pergunta 3, mas talvez ainda não a tenha lido. O participante erra a resposta por conta da escala do gráfico.

T3 - O participante utilizou a visualização da Pergunta 2 para responder a esta pergunta. Apesar de ter escrito a resposta correta, a visualização utilizada não lhe permite concluir sua resposta pois apresenta os valores de média de precipitação distribuídos ao longo dos meses. Assim, o participante errou a pergunta.

T4 - O participante utiliza a mesma visualização da Pergunta 2 e encontra a resposta correta.

T5 - O participante faz o mapeamento das variáveis DATE e WEATHER e encontra a recomendação que adiciona o contador. É interessante notar que o participante não clica no botão de especificar a recomendação, isso faz com que ele encontre as visualizações para preencher as tabelas de forma mais rápida, uma vez que não precisa refazer os mapeamentos.

T6 - O participante navega pelas recomendações, mas percebe que precisa remover a variável DATE. Ele remove ambas, DATE e WEATHER e adiciona TEMP_MIN, aplicando a função MEAN. Em seguida, adiciona a variável WETHER e encontra a visualização final.

T7 - O participante remove o mapeamento anterior, adiciona as variável LOCATION e WEATHER e encontra a visualização que adiciona o contador.

T8 - Utilizando o mapeamento da tabela anterior, o participante navega pelas recomendações e encontra a visualização que adiciona a MEAN OF TEMP_MIN e a utiliza para preencher a tabela.

VisMaker - CAPES-RJ

T1 - O participante teve certa dificuldade em perceber que precisaria utilizar a função SUM, mas consegue responder corretamente a pergunta.

T2 - O participante refaz o mapeamento e responde esta pergunta com facilidade.

T3 - O participante realiza o mapeamento de maneira tranquila, mas aplica a função MAX ao invés de SUM. Isso lhe possibilita encontrar a resposta correta, mas a visualização está errada.

T5 - O participante construiu a visualização correta em termos de mapeamento, mas removeu a variável FIELD e preencheu a tabela utilizando a visualização com informações somente sobre CITY, errando na visualização e no preenchimento.

T6 - O participante construiu corretamente a visualização, mas devido a uma instabilidade na ferramenta, o gráfico foi apresentado de forma distorcida, fazendo com que o participante errasse no preenchimento da tabela.

T7 - O participante se confunde com o termo “co-occurrences” apresentado pela ferramenta. Ele encontra visualizações corretas diversas vezes, mas parece não estar confundido disso.

T8 - O participante constrói a visualização manualmente, sem utilizar as recomendações e preenche corretamente a tabela.

T14

Voyager - WEATHER

T1 - O participante constrói o mapeamento corretamente, mas não aplica a função MEAN à variável TEMP_MAX. A visualização, apesar de não apresentar o valor exato da média, permite inferir que o mês de julho possui a maior média.

T2 - O participante encontra dificuldade em realizar o mapeamento corretamente. Sua interpretação da pergunta o faz crer que é necessário realizar um filtro sobre a variável WEATHER para considerar somente quando seu valor é RAIN. Por vezes ele tenta reiniciar a construção da visualização, até que consegue depois de resolver a Pergunta 3.

T3 - O participante não aplicou a função MEAN à variável PRECIPITATION e construiu uma visualização errada sobre essa pergunta.

T4 - O participante aplica uma filtragem sobre a visualização construída na Pergunta 2 e obtém a resposta correta para esta pergunta.

T5 - O participante teve certa dificuldade em realizar o mapeamento, mas logo encontrou uma recomendação que adicionava a variável DATE em YEAR no campo de cor e a utilizou para preencher a tabela.

T6 - Após algumas tentativas, o participante adiciona a variável WEATHER e encontra a visualização que adiciona MEAN OF TEMP_MIN, obtendo a visualização final.

T7 - Após mapear a variável WEATHER e o contador, o participante encontra a recomendação que adiciona a variável LOCATION. O participante adiciona LOCATION para o campo de filtros e realiza duas filtragens, uma para cada cidade, para preencher a tabela.

T8 - Novamente o participante realiza o mapeamento manualmente e aplica filtros em cada categoria de LOCATION para preencher a tabela.

VisMaker - CAPES-RJ

T1 - O participante teve dificuldade de compreender que era necessário aplicar a função SUM à variável QNT MASTERS. Realizou filtragem pelas variáveis YEAR e CITY. A filtragem por YEAR o fez obter uma resposta errada (bug na ferramenta).

T2 - O participante respondeu a pergunta com tranquilidade.

T3 - Novamente o participante tem uma pequena dificuldade em entender que precisa aplicar a função SUM, mas consegue responder a pergunta sem grandes problemas.

T4 - O participante responde a pergunta sem maiores dificuldades.

T5 - O participante resolve a tabela manualmente. Navegou muito rapidamente pelas recomendações, desconsiderando-as.

T6 - Novamente o participante resolve a tabela manualmente sem fazer uso das recomendações.

T7 - O participante abre a recomendação após ter adicionado a variável JURIDICAL STATUS e realiza uma filtragem nos valores de CITY para selecionar somente as cidades de interesse.

T8 - O participante navega pelas recomendações, sem interesse, até abrir a recomendação que relaciona MEAN OF QNT MASTERS, FIELD e CITY. O participante altera a função de MEAN para SUM e obtém a visualização final.

T15

Voyager - WEATHER

T1 - O participante realiza o mapeamento com tranquilidade e seleciona o *mark* para gráfico de barras.

T2 - Partindo da visualização da pergunta anterior, o participante abre a recomendação que adiciona a variável LOCATION. O participante especifica essa visualização e substitui MEAN OF TEMP_MAX por MEAN of PRECIPITATION. O gráfico obtido já responderia a pergunta, mas o participante tenta

construir uma visualização melhor. O participante constrói a visualização desejada (gráfico de linhas), mas erra a pergunta por causa da escala do gráfico.

T3 - O participante remove todo o mapeamento e adiciona manualmente as variáveis LOCATION e PRECIPITATION, aplicando a função MEAN e obtém a visualização final. A escala do gráfico novamente o atrapalha.

T4 - O participante realiza o mapeamento com tranquilidade e logo encontra a visualização que responde a pergunta. A escala do gráfico novamente o atrapalha um pouco.

T5 - O participante adiciona as variáveis WEATHER e DATE e encontra a recomendação que adiciona o contador.

T6 - Partindo da visualização anterior, o participante substitui a variável DATE por MEAN OF PRECIPITATION e encontra a visualização final ainda com um contador desnecessário.

T7 - O participante constrói a visualização muito rapidamente apenas substituindo a variável PRECIPITATION por LOCATION.

T8 - O participante se recorda do gráfico utilizado na Tabela 2, abre-o através do bookmark e adiciona a variável LOCATION para o campo COLOR, encontrando a visualização final que também possui um contador desnecessário.

VisMaker - CAPES-RJ

T1 - O participante realiza o mapeamento com facilidade, apesar de utilizar a variável temporal no campo de cor.

T2 - O participante realiza o mapeamento com facilidade.

T3 - O participante adiciona as variáveis QNT POSTDOCTORAL e CITY e tenta mapear a variável FIELD em vários campos até encontrar uma boa opção adicionando-a ao campo ROWS.

T4 - O participante constrói a visualização com facilidade.

T5 - O participante consegue obter a resposta rapidamente, mas não utiliza as recomendações.

T6 - O participante consegue obter a resposta rapidamente e utiliza as recomendações apenas para sair do *heatmap* e chegar ao gráfico de barras empilhadas.

T7 - O participante consegue obter a resposta rapidamente e utiliza as recomendações apenas para sair do *heatmap* e chegar ao gráfico de barras empilhadas.

T8 - O participante resolve a tabela com facilidade, mas não utiliza as recomendações.

T16

Voyager - WEATHER

T1 - O participante realiza o mapeamento com certa facilidade. Demora um pouco para completar a tarefa mas consegue sem problemas.

T2 - O participante sente um pouco mais de dificuldade. Adiciona a variável DATE para o campo de filtros sem sucesso. Logo depois, o participante consegue montar uma visualização que lhe permitisse responder a pergunta. A escala do gráfico o faz escrever a resposta errada.

T3 - O participante constrói o mapeamento com facilidade. Obtém a visualização correta, mas recomeça o processo talvez por uma ambiguidade na pergunta. Rapidamente, ele reconstrói a visualização e responde corretamente.

T4 - O participante constrói o mapeamento corretamente, mas esquece de aplicar o filtro. Ele percebe que precisa fazê-lo e seleciona somente os valores de LOCATION para cidade de SEATTLE.

T5 - O participante utiliza bem as recomendações e chega a encontrar um gráfico que poderia responder a tabela, mas reinicia o mapeamento a fim de encontrar uma visualização melhor. Ele adiciona a variável DATE e o contador e abre a recomendação que adiciona a variável WEATHER.

T6 - O participante realiza o mapeamento manualmente e preenche a tabela. O participante realiza novamente a visualização através das recomendações.

T7 - O participante adiciona a variável LOCATION e navega pelas recomendações até achar uma que adiciona a variável WEATHER, utilizando-a para responder a tabela. O participante ainda adiciona a variável WEATHER para o campo COLOR.

T8 - O participante abre a recomendação sobre BIN OF TEMP_MIN e altera a função para MEAN. Em seguida, adiciona a variável WEATHER para o EIXO Y e a variável LOCATION para o campo de filtros. O participante aplica filtros em cada cidade para preencher a tabela.

VisMaker - CAPES-RJ

T1 - O participante realizou uma filtragem por ano e acabou errando a resposta no formulário por conta disso (bug na ferramenta). A visualização, no entanto, está correta.

T2 - O participante realizou o mapeamento com certa facilidade. A visualização final possui um contador desnecessário.

T3 - O participante realizou o mapeamento e a filtragem com facilidade.

T4 - O participante chega a adicionar a variável QNT MASTERS para o campo de filtros. Mas corrige transferindo-a para o EIXO X.

T5 - O participante chega a construir a visualização correta utilizando as recomendações, mas não se sente seguro e a refaz. O termo “co-occurrences” pode tê-lo confundido. O participante faz o mapeamento manualmente e obtém um *heatmap*.

T6 - O participante encontra rapidamente o *heatmap* desejado através das recomendações.

T7 - Novamente o participante utiliza a recomendações e encontra rapidamente a visualização correta.

T8 - O participante adiciona as variáveis FIELD e QNT MASTERS e abre o *heatmap* que adiciona a variável CITY. O participante altera a função de MEAN para SUM OF QNT MASTERS.