![KIT logo]
Karlsruhe Institute of Technology

Interactive Visualization of Correlations in High-Dimensional Streams

Bachelor's Thesis of

Yimin Zhang

at the Department of Informatics
Institute for Program Structures and Data Organization (IPD)

Reviewer: Prof. Dr.-Ing. Klemens Böhm
Second reviewer:
Advisor: Ph.D. Edouard Fouché

01. Mar 2019 – 01. Jun 2019

Karlsruher Institut für Technologie
Fakultät für Informatik
Postfach 6980
76128 Karlsruhe

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.
PLACE, DATE

. . . .
(Yimin Zhang)

Abstract

A fundamental task of Data Mining is to estimate the correlation between the attributes of a data set. Knowing the relationship between a set of variables, one can infer useful knowledge about external, a priori unknown outcomes.

In contrast to static data, the data is often available as a stream, i.e., it is an infinite, ever evolving sequence of observations. Concepts learned at a certain time cannot be expected to hold in the future. Therefore, correlation estimation should be a continuous process.

Also, the data is often high-dimensional, i.e., it contains hundreds or thousands of dimensions. Besides the computational burden to estimate the correlation between many subsets, it becomes difficult for a human observer to extract knowledge from the results. The task becomes even more difficult if one considers correlations between more than two variables, because the size of the result increases exponentially. In our thesis, we only focus on pairwise relationships and the correlations between more than two variables may remain to be discovered in the future work.

The aim of this bachelor thesis is to develop a graphical interface for data scientists, dedicated to the visualization of correlation in user-given data streams. With this interface, available for example as a web-service, users provide their own data sets. Then, the system's back-end estimates the correlations, and provide several visualizations of the results, for example, via force-directed graphs. Users will be able to interact in several ways with the interface, by setting parameters to tune the visualization. We evaluate the benefits of our interface and determine which visualization is the most appropriate visualization, depending on what one is looking for, via controlled user studies.

Zusammenfassung

Die Abschätzung der Korrelation von Attribute in einer Datenmenge ist einer der grundlegenden Aufgaben von Data Mining. Wenn man die Beziehung von Variablen kennt, dann kann man einige nützliche Ausgaben über zusätzliche und unbekannte Informationen folgern.

Normalerweise sind die Daten als Datenfluss verfügbar, d.h. Es ist unendlich und sogar evolutionär. Die zur-zeitigen schon erkennende Begriffe und Resultaten kann man in der Zukunft nicht mehr benutzen. Deshalb muss die Abschätzung der Korrelation ständig werden.

Ein anderes Problem ist die hohen Dimensionen. Die Daten enthalten oft mehr als 100 oder sogar 1000 Dimensionen, sodass es ist schwierig für das Rechnen der Daten. Es ist auch schwer für ein Mensch um Daten zu analysieren. Wenn es um die Korrelation über mehr als zwei Variablen geht, wächst das Rechnen der Datenmenge exponentiell an. Unsere Arbeit konzentriert sich zur Zeit nur auf Korrelation über zwei Variablen und das Problem auf die Korrelationen über mehr als zwei Variablen steht noch in der zukünftigen Arbeiten aus.

Das Ziel dieser Arbeit ist die Entwicklung von einer graphischen Schnittstelle für die Daten Wissenschaftler, um die Korrelation der Daten zu visualisieren. Mit dieser Schnittstelle, als zum Beispiel Web-Service, laden die Benutzer selbst Datenmenge hoch. Danach wird das Backend von System die Korrelationen von Attribute berechnen und einige Visualisierungen von Daten ausgeben, zum Beispiel, durch Force-Directed-Graph. Es ist auch möglich für die Benutzer mit der Schnittstelle Parameter aufzustellen, um die Visualisierung zu verbessern. Zum Schluss bewerten wir die Vorteile und Nachteile dieser Schnittstelle und bestimmen die beste Visualisierung anhand verschiedene Situationen und Forderungen durch einige Anwendungsfälle.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Correlation analysis aims at discovering and summarizing the relationship between the attributes of a data set. Knowing the relationship between a set of variables, one can infer useful knowledge about external, a priori unknown outcomes.

For example, we can measure the stock relationship via correlation coefficients in stock markets that may undergo large fluctuations of stock prices. Figure 1.1 represents the correlation coefficients among the five funds' returns from 1980 to 1998. [9] Correlation coefficients describe the extent to which asset returns "move together." Correlation coefficients range in value between negative one (completely negatively correlated) and positive one (completely positively correlated), while a correlation of zero means that there is no correlation between two attributes. Such a correlation monitoring system could be of great help to a financial analysis.

Fund	U.S. Stock	European Stock	Pacific Stock	U.S. Bond	U.S. Money Market
U.S. Stock	1.00	.59	.33	.29	−.05
European Stock		1.00	.53	.22	−.13
Pacific Stock			1.00	.14	−.10
U.S. Bond				1.00	.14
U.S. Money Market					1.00

Figure 1.1: Correlations Among the Five Funds' Returns, Monthly Returns, from 1980 to 1998

As an investor, we can see that the stock markets are positive correlated between each other, which indicates a similar behavior during this period. As a result, it may make less sense when we want to maximize our wealth by investing different stock markets, if our current stock market reaches its valley. Also, if we are fully aware of our current stock market, we are likely to predict the behavior of other stock markets due to the positive correlation.

Analyzing the correlation between different attributes helps us to understand their relationship. In general, the correlations between attributes remain same or change gradually. If the correlation structure changes brutally, which often indicates a sudden peak or valley, we can infer that one of the attributes may have an enormous change or the relationship between them may differ thoroughly. Figure 1.2 shows the dynamic correlation between stock market index and the crude oil price. [3] During the period 1987 - 2009, 6 important events occurred:

- Iraq invasion in Kuwait/first war in Iraq

- Asian economic crisis

- Housing market boom

- Second war in Iraq

- Chinese economic growth

- Global financial crisis

These events signed the brutal changes in the correlation between markets and oil price, which are printed in blue circles in the Figure 1.2.

1.2 Challenges

The challenges of analyzing high-dimensional data streams is twofold: the evolving nature of streams and the high-dimensionality.

1.2.1 The evolving nature of streams

In contrast to static data, the data is often available as a stream, i.e., it is an infinite, ever evolving sequence of observations. As the concepts learned at a certain time cannot be expected to hold in the future, correlation analysis should be a continuous process.

1.2.2 The high-dimensionality

Also, the data is often high-dimensional, i.e., it contains hundreds or thousands of dimensions. In the case of streams with many dimensions, it is difficult to extract actionable insights from the correlation matrix, as the number of pairs of attributes increases quadratically and the coefficients evolve over time in unforeseen ways. For the pairwise correlation analysis of any data steam with n components, one need to compute the correlation between $\dfrac{n * (n - 1)}{2}$ pairs, i.e., with $n = 5$, one needs to compute 10 pairs, with $n = 10$, one needs to compute 45 pairs. Therefore, it becomes difficult to visually keep track of correlation and impossible to understand the result of the correlation analysis as the number of attributes increases. The Figure 1.3 shows the visualization of different numbers of attributes,

1.3 Goal of the thesis

The goal of this thesis is to propose and evaluate new tools for the interactive visualization of correlation in high-dimensional streams. In this thesis, we are going to answer the following three questions:

- What visualization methods are the most appropriate to visualize correlation?

- How can one visualize the evolution of correlation results?

- What are the desirable features of a correlation monitoring interface?

To answer these questions, we first search for some interactive visualization methods and then develop an interface to visualize correlations. This interface can be available in a browser and should be user friendly, which means that users provide their own data sets and the system's back-end calculates the correlations and then provides the visualization of the results, for example, via force-directed graphs. Users are also able to interact in several ways with this interface, such as setting parameters to tune the visualization. At last, we evaluate the benefits of our interface systematically via controlled user studies and discuss the advantages and disadvantages of each visualization method when meet up with different scenarios. Based on the results, we discover the appropriate visualization methods of correlations in high-dimensional streams and the desirable features of the interactive interface.

1.4 Thesis outline

The process of investigation can be split into three parts, that is performing a literature review, developing a graphical interface and evaluating this interface, which also leads to the three parts of this thesis, the visualization methods, the design and implementation of interface, and the evaluation of this interface.

In Chapter 2, the first part of the thesis, we introduce some state-of-art methods for correlation visualization. Section 2.1 introduces the correlation matrix, which is the standard tool of performing the correlation analysis. In Section 2.2, Section 2.3 and Section 2.4, some other examples of data visualization are discussed.

In Chapter 3, we give an overview of parallel coordinates and scatterplot matrices as examples of multidimensional visualization techniques. Also, we introduce an interactive Framework for Exploring and Understanding Multivariate Correlations "FEXUM", which uses Force-Directed-Graph as the visualization method for the data set.

Chapter 4 is the main part of this thesis. Section 4.1 is the design of the interface and Section 4.2 describes the implementation of this interface.

We introduce the evaluation of the developed interface in Chapter 5. Chapter 5 contains not only the experimental settings of controlled user studies for evaluation in Section 5.1, but also the results of controlled user studies and feedbacks in Section 5.4.

In the end, we summarize the whole thesis.

a) Oil Exporting Countries
Time varying correlation between Brazilian stock market index (Bovespa) and crude oil.

b) Oil Importing Countries
Time varying correlation between USA stock market index (Dow Jones Industrial) and crude oil.

Time varying correlation between Canadian stock market index (S&P/TSX 60) and crude oil.

Time varying correlation between Netherlands stock market index (AEX General) and crude oil.

Time varying correlation between Mexican stock market index (MXIPC 35) and crude oil.

Time varying correlation between German stock market index (DAX 30) and crude oil.

Figure 1.2: Dynamic correlation between stock market index and the crude oil price

(a) Correlation matrix of 5 attributes

(b) Correlation matrix of 10 attributes

(c) Correlation matrix of 40 attributes
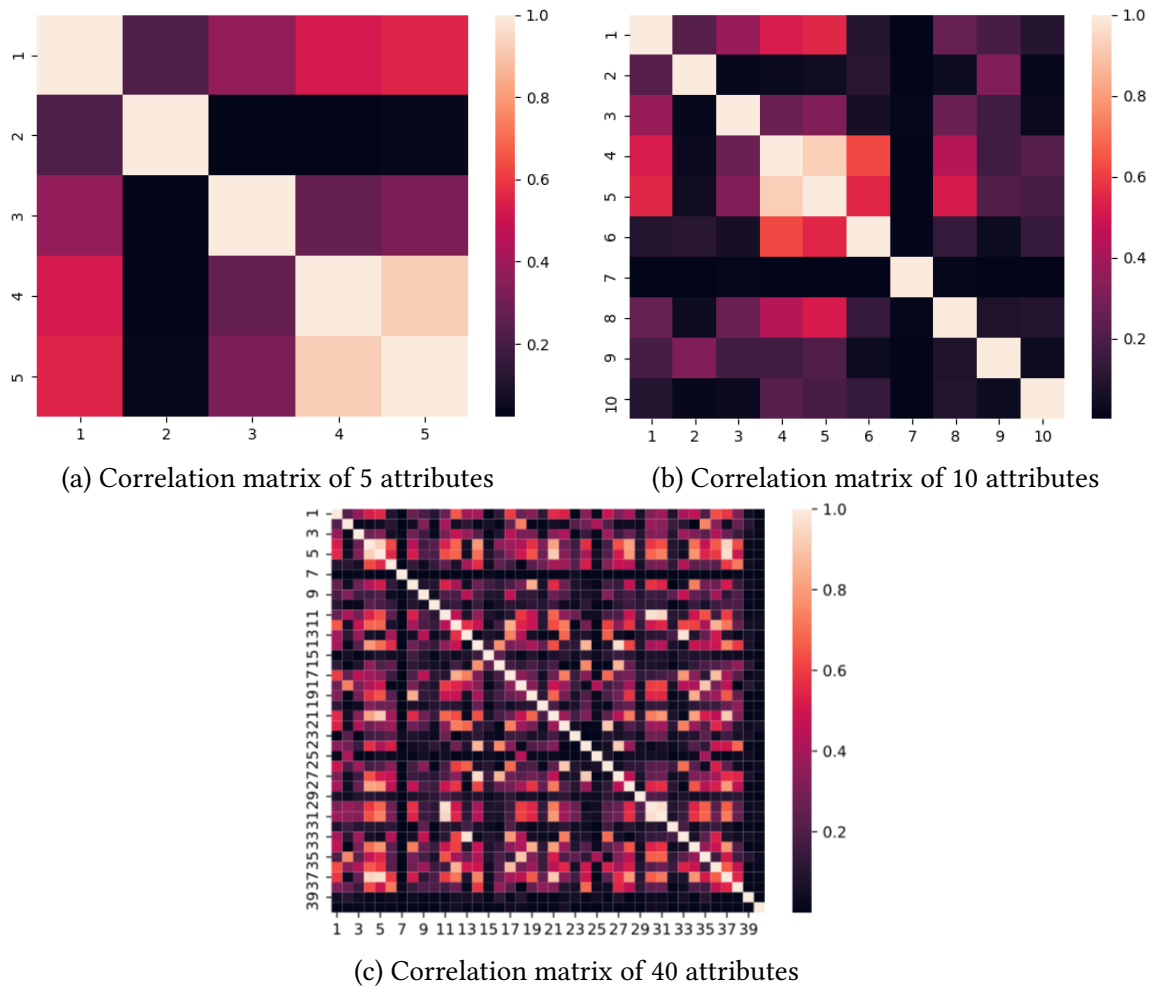
Figure 1.3: Visualizations of different numbers of attributes

2 Visualization Methods

In this chapter, we first introduce the correlation matrix in Section 2.1 and then introduce some useful interactive visualization methods for correlations in Section 2.2, Section 2.3 and in Section 2.4.

2.1 Correlation Matrix

The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient [8], known as "Pearson's correlation coefficient". It is obtained by dividing the covariance of the two variables by the product of their standard deviations. The population correlation coefficient $\rho_{X,Y}$ between two random variables X and Y and standard deviations σ_X and σ_Y is defined as

$$\rho_{X,Y} = \mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{2.1}$$

where cov means covariance, and corr is a widely used alternative notation for the correlation coefficient. The Pearson correlation is defined only if both of the standard deviations are finite and non-zero.

The standard tool of correlation analysis is the computation of a correlation matrix, which is used to investigate the dependence between multiple variables at the same time.

2.2 Heatmap

A common visualization is the heatmap[10], which is originated in 2D displays of the values in a data matrix. The Figure 2.1 is a heatmap of a correlation matrix, in which the variables with strong correlation (high values) are printed in light colour and those with low correlation are in dark colour.

2.3 Force-Directed Graph

Also, Force-Directed Graph[5] is a useful visualization, which assigns forces among the set of edges and the set of nodes of a graph drawing. The purpose of it is to position the nodes of a graph in two-dimensional or three-dimensional space so that all the edges are of more or less equal length and there are as few crossing edges as possible. In such a
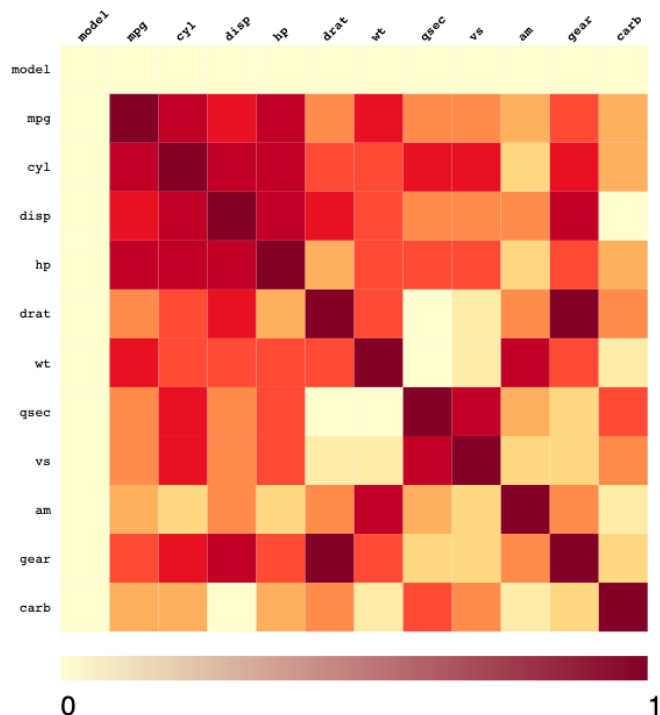
Figure 2.1: Correlation Matrix of a data set in heatmap

simulation, the forces are applied to the nodes, pulling them closer together or pushing them further apart. This can be used to simulate the relationship of different attributes throughout the time, in which the force is the representation of correlation matrix. The Figure 2.2 shows an example of Force-Directed Graph, which actually uses the same data set in Figure 2.1.

2.4 Bar Graph

A bar graph [4] is a graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. It can be either vertical or horizontal. We use vertical bar graph as a visualization in our developed interface. In a vertical bar graph, the x-Axis shows the specific categories being compared, and the y-Axis represents a measured value, which is the correlation value in our situation. Bar graphs provide a visual presentation of categorical data, which are usually qualitative. The Figure 2.3 is an example of bar graph from Mike Bostock. [6] This chart shows the relative frequency of letters in the English language.
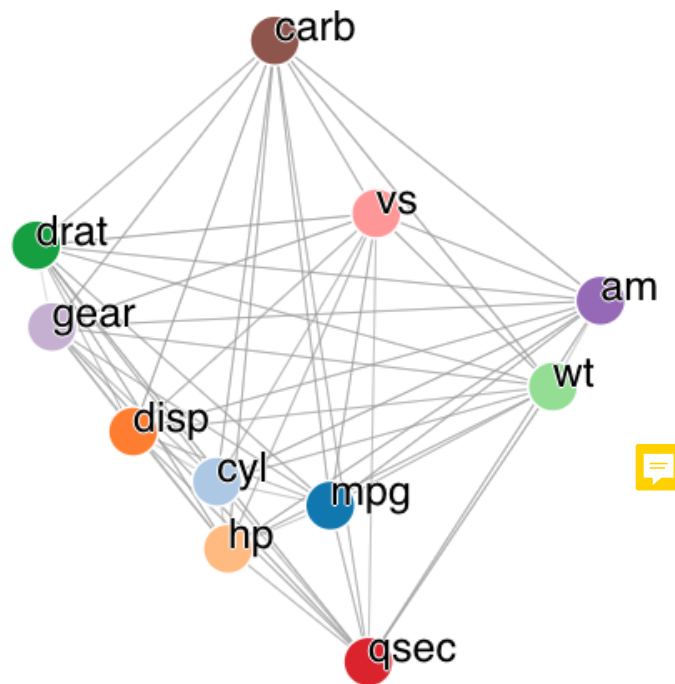
Figure 2.2: Force-Directed Graph
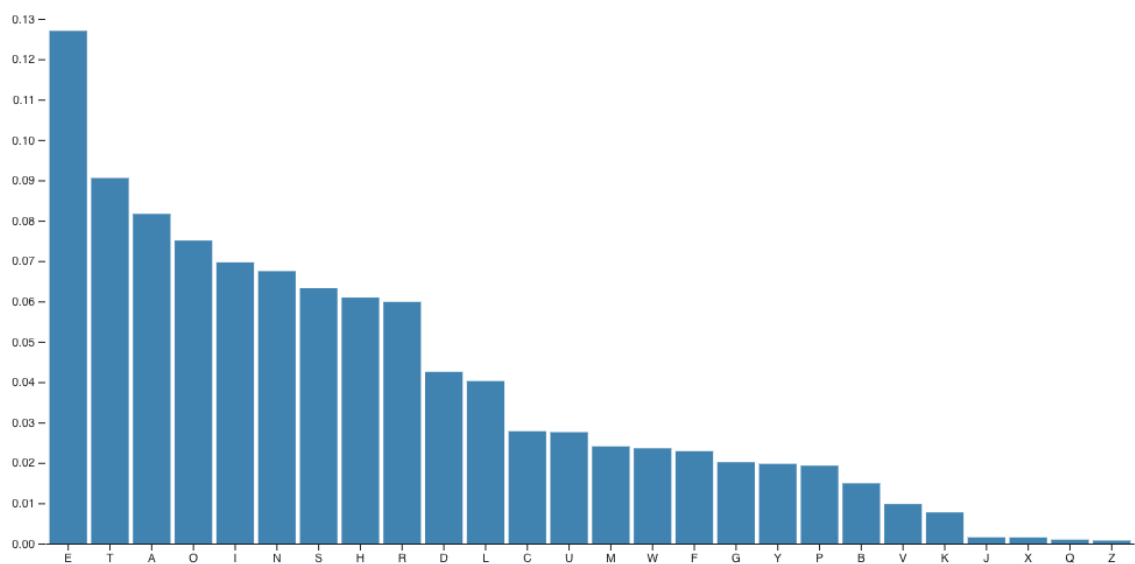
Figure 2.3: Bar Graph

3 Related Work

With the exponentially increasing amount of acquired multivariate data, several multi-dimensional visualization techniques have been proposed during the last decades. Parallel coordinates and scatterplot matrices are widely used to visualize multi-dimensional data sets. But these visualization techniques are insufficient when the number of dimensions grows. To solve this problem, different approaches to select the best views or dimensions in advance have been proposed in the last years.
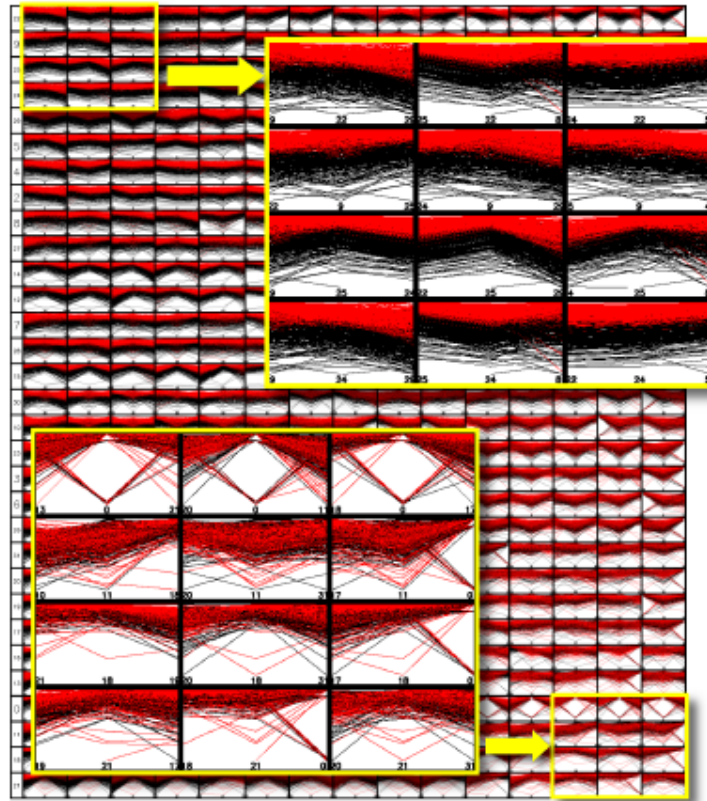
Figure 3.1: Parallel coordinates matrices for the data set

Georgia Albuquerque et al. [1] presented three new methods to explore multivariate data sets: a parallel coordinates matrix (Figure 3.1), in analogy to the well-known scatterplot matrix, a class-based scatterplot matrix that aims at finding good projections for each class pair (Figure 3.2), and an importance aware algorithm to sort the dimensions of scatterplot and parallel coordinates matrices. They aims at providing a visualization of the

whole data set, not the correlations of attributes in this data set. As we focus on the correlations only between each two attributes in our thesis, it is no need for us to have parallel coordinates in our framework.
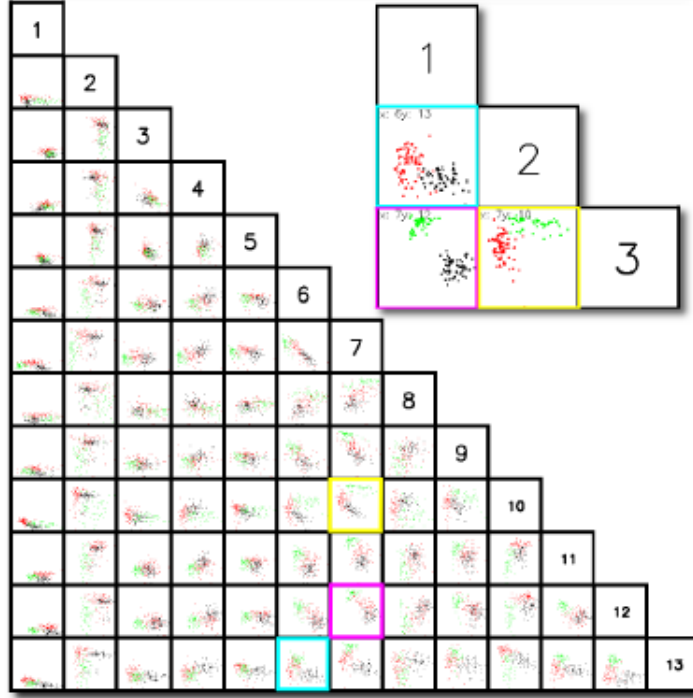

Figure 3.2: Class-based scatterplot matrices for the data set

In Section 1.1, we have pointed out the use and importance of correlation analysis between different attributes, which helps people to understand the relationship of attributes in a data set. Therefore, our interactive interface aims at providing a visualization of correlations for people.
As a high-dimensional data set may contain many redundant features, feature selection becomes an essential step for correlation analysis. Louis Kirsch et al. [2] developed an interactive Framework for Exploring and Understanding Multivariate Correlations (FEXUM) to simultaneously visualize all feature correlations to the target and pairwise correlations using Force-Directed-Graph. This visualization provides a layout in which a smaller distance between two features denotes a greater redundancy. In Figure 3.3, nodes represent features and weighted edges represent distances. FEXUM provides users with an understanding of how features interact with each other in terms of redundancy so that they can easily find influential value ranges in the analysis view and make feature selection.
In our developed interface, we give the users an overview of all correlations in the data set. Instead of choosing a target attribute to focus on, our system represents the whole correlations of the data set. Unlike FEXUM, Force-Directed-Graph is not the only visualization method in our system. Heat map and Bar graphs are alternative visualization

method for the users so that the users can choose the most suitable visualization method in their opinion.

In Subsection 1.2.1, we have discussed that the correlation analysis should be a continuous process. FEXUM enables the users to upload their own data sets and visualize them. However, in our system, the uploaded data set by users can be a data set of data stream so that the users can choose a period of time to perform the correlation analysis and visualization. Our goal is to visualize a concise but useful summary of correlations in the stream over time.

Figure 3.3: Features drawn using a force-directed graph (right), with the target highlighted in green. An analysis view of two features (left) for inspecting the correlations. (Color figure online)

4 Interface

To answer the questions we mentioned in the Section 1.3, we develop an interface in a browser available as a web-service. The entire framework is open source and available online. In Section 4.1, we describe the mock-up of this interface and its available functions. And we introduce the details of implementation in Section 4.2.

4.1 Design

Figure 4.1 is the mock-up of this interface. This interface can be used with a wide range of data sets as csv files, supplied through upload by the user. After the calculation in the back-end, the visualization of data correlation is shown in the website, for example, via a force-directed graph. In the mock-up, a sliding window with start point and step size is supposed to be used to represent the continuous process of data throughout the time. Also, we have some simple user settings, such as changing the window size, setting the minimum and maximum of correlation to be visualized.

4.2 Implementation

Our interface is developed by Javascript. WebStorm is the implementing tool for writing Javascript.

4.2.1 D3.js

D3.js (Data-Driven Documents)[7] is a data-driven JavaScript library for producing dynamic, interactive data visualizations in web browsers. It makes use of the widely implemented SVG, HTML5, and CSS standards and allows great control over the final visual result. SVG stands for Scalable Vector Graphics which is technically an XML based markup language. It is a great tool to display icons, logos, illustrations or charts, which is supported in all major browsers and requires no third-party lib because of owning the DOM interface;. In our project, we mainly use D3.js to implement the visualizations of correlations in high-dimensional data streams.
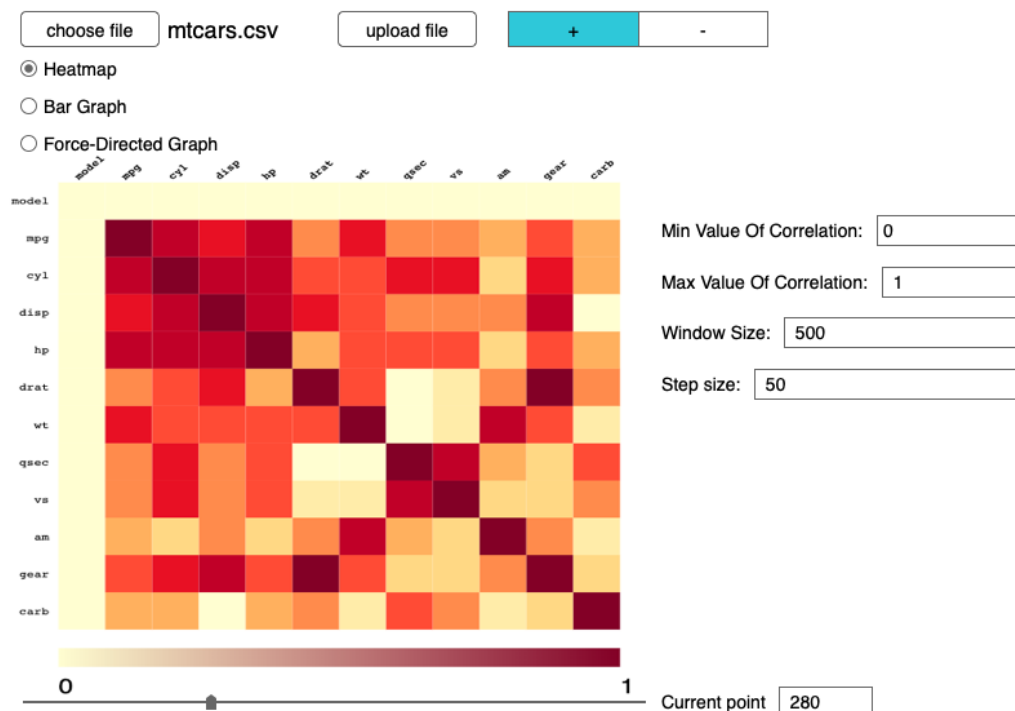
Figure 4.1: Mockup of the interface

4.2.2 Details

The Figure 4.2 is an overview of our developed interface, which is available in a browser as a web-service. Users can press the "Choose file" button to upload their own data sets.

After pressing the **upload** button, the data set will be be calculated at the back-end. If the calculation of correlations is finished, a confirm window Figure 4.3 will pop up to inform the maximal window size of the current data set to the user. The maximal window size is the whole size of the uploaded data. Also, **the step size** will be set to 1 and **the window size** will be set to the maximal window size after uploading the data set, see Figure 4.4.

Minimum is the minimal value of the correlation value, which is not filtered in the visualization graph, and it is set to 0 as default. **Maximum** is the maximal value of the correlation value, which is not filtered in the visualization graph, and it is set to 1 as default. **The window size** represents the size of the selected timestamps of the current data set. The width of the panel on the slider indicates the size of the current window. Combined with **the step size**, the user is able to get the visualization of a certain time period by sliding the sliding window. **The current point** represents the starting point of current selected group of timestamps. Therefore, it is also possible for the user to input the starting point of timestamps.

Interactive Visualization of Correlations in High-Dimensional Streams

Figure 4.2: Overview of the interface

Figure 4.3: The confirm window to inform the maxiaml window size of the current data set

As we aim at analyzing big data stream, we use 50 threads to calculate part of the correlations between each two attributes in parallel so that the calculation time would be reduced after the user press the **update** button, see Figure 4.5.

Our interface provides three visualization methods: heat map, bar graph and force directed graph. After selecting the corresponding radio button, a visualization will be drawn on the web site. The Figure 4.6 shows the three visualizations based on the same data set. Figure 4.6a is the heat map. Each box represents a pair of attributes and the color of the box represents the correlation value of this pair. The darker the box is, the bigger is the correlation value. Figure 4.6b is the bar graph, in which the height of each bar represents the correlation value of each pair of attributes. And Figure 4.6c is the force directed graph, in which the length of link between two nodes represents the correlation value of this pair of attributs. The shorter the linked distance is, the bigger is the correlation value. The user can also change the minimal and maximal correlation value they want to visualize, and slide the slider to see different time period of the current data set.

Figure 4.4: After uploading a csv file
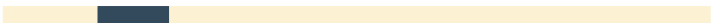
Figure 4.5: After pressing the "update" button
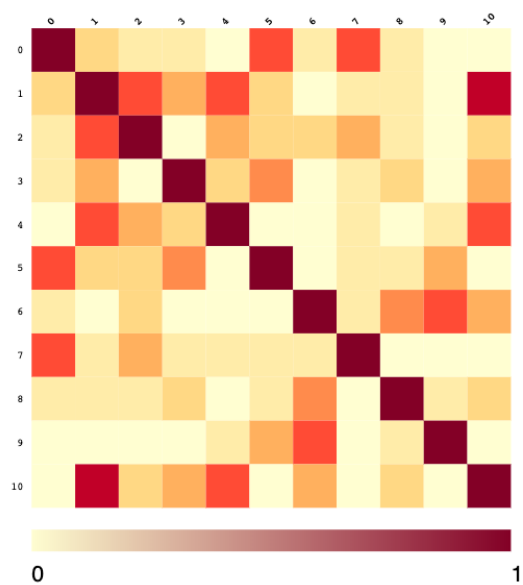
(a) Select Heatmap as visualization method
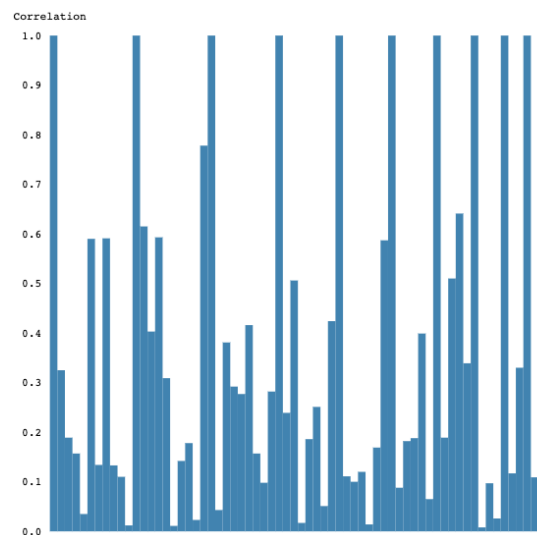
(b) Select Bar graph as visualization method
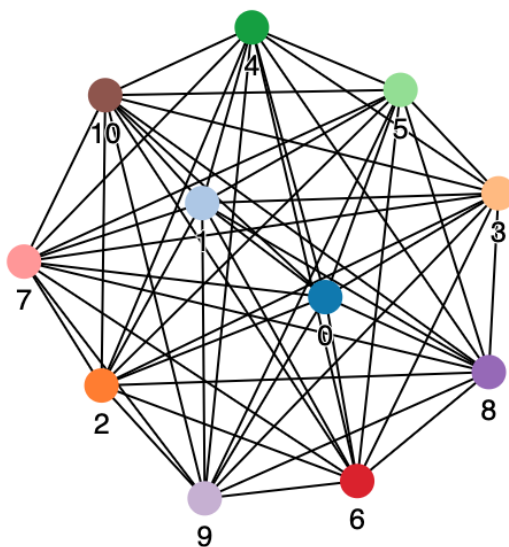
(c) Select Bar graph as visualization method

Figure 4.6: Select different visualization methods

5 Evaluation Via User Studies

5.1 Experimental Settings

The goal of the study is to compare three different visualization methods, namely so the so-called "heat map", "bar graph" and "force-direct graph". The participants are assessed with respect to different conditions, to which they will be assigned randomly. Participants are not aware of the condition they are assigned to:

- **Condition A**: Participants can only use the heat map and Figure 5.1 is its mockup

Figure 5.1: Mockup for Condition A

- **Condition B**: Participants can only use the bar graph and Figure 5.2 is its mockup

- **Condition C**: Participants can only use the force-directed graph and Figure 5.3 is its mockup

- **Condition D**: Participants can use any visualization method they want,which are mentioned in Condition A,B and C, and Figure 4.1 is its mockup

Interactive Visualization of Correlations in High-Dimensional Streams
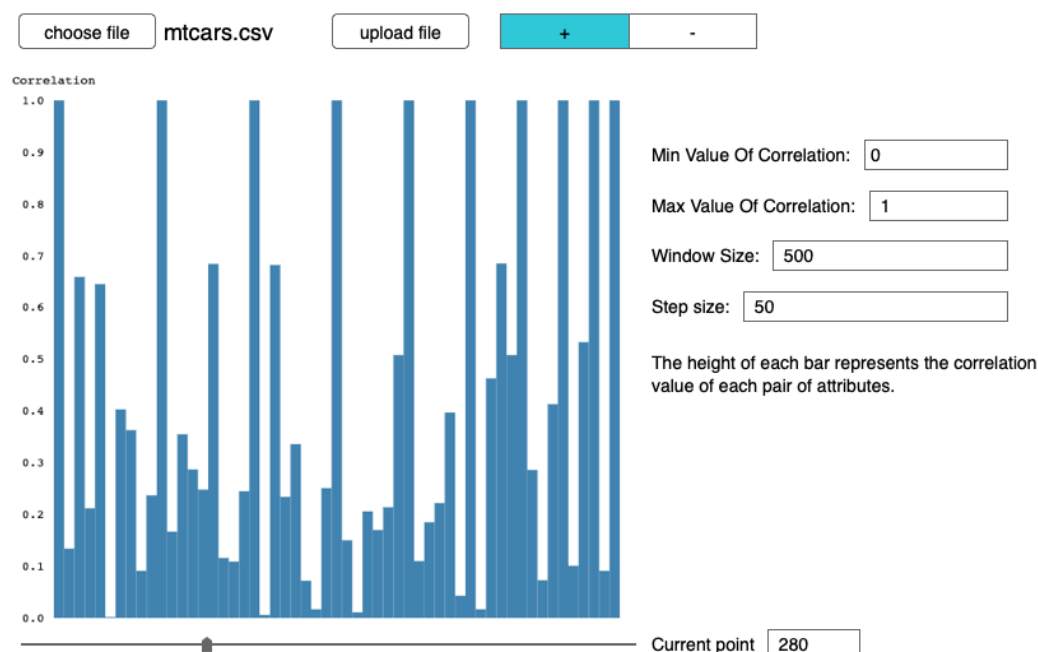
Figure 5.2: Mockup for Condition B

5.1.1 Participants Profiles

The participants we are looking for are adults who at least have the basic knowledge about browsing ~~the~~ web page, typically 14 to 40 years old. We will sample a large share of our participants from the pool of students at KIT, so we may expect basic knowledge of computer science and data analysis.

Knowledge in data analysis and correlation analysis should not be required to use the prototype for the visualization of a data set. Still, we hypothesis that participants with prior exposure to data visualization will require less time to fulfill the tasks.

5.1.2 Data Set

The participant of user study are asked to use a data set taken at random from a pool of 3 data sets in total for completing the tasks. **Data Set 1 (DS1)** has 10 attributes within 10000 timestamps. **Data Set 2 (DS2)** has 20 attributes within 2000 timestamps. **Data Set 3 (DS3)** has 40 attributes within 1000 timestamps. Obviously, the difference between DS1, DS2 and DS3 are their level of difficulty according to their number of dimensions. In practice, this means that the same question will be harder to answer with DS3 than DS2 and DS1. We define what it means for each question type to be "easy", "middle" or "hard".
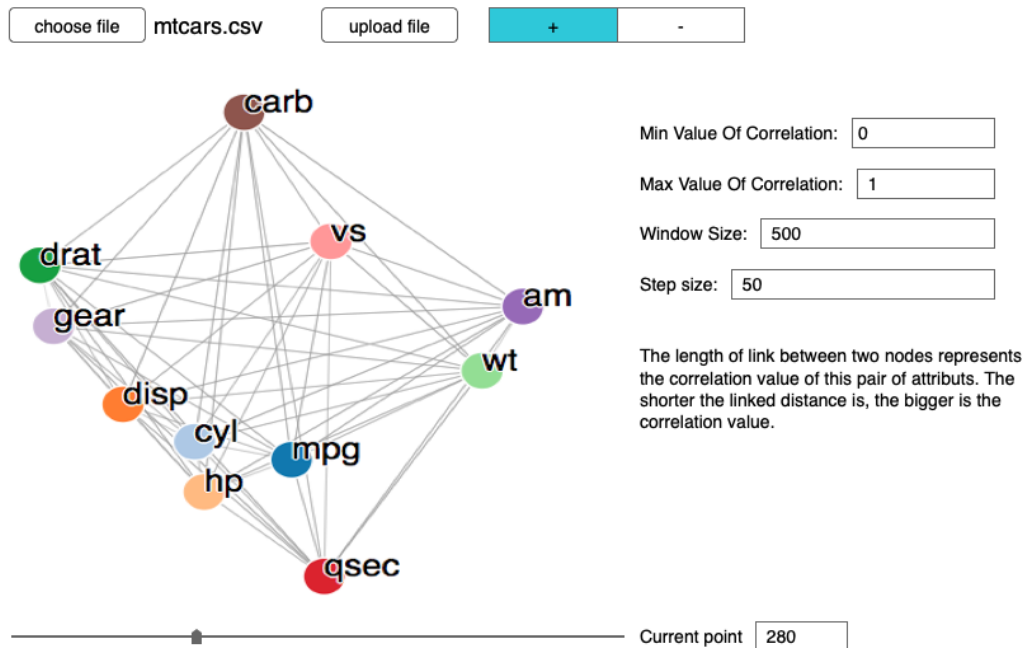
Figure 5.3: Mockup for Condition C

5.1.3 Settings of interface

The window size and step size are set to the following values, which can not be changed by the participants:

- **Data Set 1**: The window size is set to 1000 and the step size is set to 500.

- **Data Set 2**: The window size is set to 200 and the step size is set to 50.

- **Data Set 3**: The window size is set to 200 and the step size is set to 50.

During the user study, the participants are only able to change the minimum and maximum to filter the correlation values. Also, they can slide the sliding window or input the current point to get the visualization graph of current point.

5.2 Questionnaire

The questionnaire is divided into 3 parts. In the first part of questionnaire Section 5.2, the participants are asked to give some basic information. In the second part of questionnaire, the participants have to answer some questions using different visualization methods. Subsection 5.2.2 shows the question types and each question type will be asked three time in the user study. The last part of questionnaire Subsection 5.2.3 is the feedback of using this interface. Participants under Condition A/B/C are asked to give feedbacks to

Figure 5.4: Mockup for Condition D

the corresponding visualization methods they use, while participants under Condition D have to give feedbacks to each visualization method and the whole system.

5.2.1 Basic Information

- Field of study

- Number of semester

- Age:

18-24	25-30	More than 30

- Gender:

M	F	Do not wish to answer

- How familiar are you with the following concepts?

	1 (Unfamiliar)	2	3	4	5	6	7 (Very Familiar)
Correlation analysis							
Data analysis							
Data visualization							

5.2.2 Visualization

- How many attributes are available in this data set?

- How many pairs of attributes are available in this data set?

- What is the correlation value between ***Attribute A*** and ***Attribute B*** at ***Timestamp T***, or the probable range?

- Which pair of attributes has the biggest correlation at ***Timestamp T***?

- Which pair of attributes has the smallest correlation at ***Timestamp T***?

- The following statement is true or false: "The correlation value between ***Attribute A*** and ***Attribute B*** remains the same at ***Timestamp T1*** and at ***Timestamp T2***"?

True	False

- Which pair(s) of attributes has/have a correlation value that is not smaller than ***X*** at ***Timestamp T***?

- Which pair(s) of attributes has/have a correlation value that is not bigger than ***X*** at ***Timestamp T***?

5.2.3 Feedback

5.2.3.1 Participants Under Condition A/B/C

Participants under Condition A, B and C are asked to give feedback based on following questions:

- Please write down the visualization method you have used and rate for it.

	1 (Strongly Disagree)	2	3	4	5	6	7 (Strongly Agree)
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

- In your opinion, what are the strengths of this visualization system?

- In your opinion, what are the weaknesses of this visualization system?

- Do you have any suggestions for improving this visualization system?

5.2.3.2 Participants Under Condition D

Participants under Condition D, who are able to use either of 3 visualization methods, are asked to give feedbacks to the whole system, which is also quite similar to the one for participants under Condition A, B or C. Also, they need to give rating to each visualization methods.

- As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

- And which of the method did you use the most?

- Please rate for *Heat map*:

	1 (Strongly Disagree)	2	3	4	5	6	7 (Strongly Agree)
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

- Please rate for *Bar Graph*:

	1 (Strongly Disagree)	2	3	4	5	6	7 (Strongly Agree)
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

- Please rate for *Force-Directed Graph*:

	1 (Strongly Disagree)	2	3	4	5	6	7 (Strongly Agree)
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

- Please rate for *whole system*:

	1 (Strongly Disagree)	2	3	4	5	6	7 (Strongly Agree)
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

- In your opinion, what are the strengths of this visualization system?

- In your opinion, what are the weaknesses of this visualization system?

- Do you have any suggestions for improving this visualization system?

5.3 Script For Conducting The Experiment

The following information will be given to the participants, orally:
Thank you for participating to this experiment. My name is Yimin Zhang. I am doing my Bachelor thesis in the Institute for Program Structures and Data Organization (IPD Böhm).
The goal of the experiment is to evaluate an interface that I have developed for my thesis. First of all, please read and sign the consent form. If you have any problems during the study, please be free to talk to me.
(After signing)
Please fill in the blanks of the website about the basic information, which is the first part of the questionnaire: Section 1.
(After Part 1)
Your task from now on is to upload the data set (1/2/3) and perform the condition (A/B/C/D) to answer the questions in Part 2. You are free to use all the elements on the web page to ease the task. Please cross a question off, if you cannot answer the question. The time you need to answer for each question will also be recorded.
(After Part 2)
The last task for you is to provide your personal feedback for this interface. This is the third part of the questionnaire.
(After Part 3)
Thank you for participating in my study. If you are interested in my thesis or may have further questions, please be free to contact me using the contact information in the consent form. I wish you a nice day.

5.4 Result

…

6 Conclusion

…

Bibliography

[1] Albuquerque, G., Eisemann, M., Lehmann, D. J., Theisel, H., and Magnor, M. Quality-Based Visualization Matrices. *In Proceedings of the Vision, Modeling and Visualization* (2009), 341–350.

[2] B, L. K., Riekenbrauck, N., Thevessen, D., Pappik, M., Stebner, A., Kunze, J., Meissner, A., Shekar, A. K., and Emmanuel, M. Machine Learning and Knowledge Discovery in Databases. 404–408.

[3] Filis, G., Degiannakis, S., and Floros, C. Dynamic correlation between stock market and oil prices: The case of oil-importing and oil-exporting countries. *International Review of Financial Analysis 20*, 3 (2011), 152 – 164.

[4] Kelley, W. M.; Donnelly, R. A. The humongous book of statistics problems. *New York* (2009).

[5] Kobourov, S. G. Spring embedders and force-directed graph drawing algorithms. *eprint arXiv:1201.3011* (2012).

[6] Mike, B. Bar chart, 11 2017.

[7] Murray, S. Interactive data visualization for the web, an introduction to designing with d3. 272.

[8] Rodgers, J. L.; Nicewander, W. A. Thirteen ways to look at the correlation coefficient. *The American Statistician* (1988).

[9] Simons, K., et al. Should us investors invest overseas? *New England Economic Review* (1999), 29–40.

[10] Wilkinson, Leland; Friendly, M. The history of the cluster heat map. *The American Statistician* (2009).

7 Appendix

7.1 Consent Form

Consent Form

STUDY: Interactive Visualization of Correlations in High-Dimensional Streams

INSTITUTION: Institute for Program Structures and Data Organization (IPD Böhm), Karlsruhe Institute of Technology, Am Fasanengarten 5,76133 Karlsruhe

PURPOSE OF STUDY: The evaluation of visualization methods for correlations in high-dimensional streams

PROCEDURE: Throughout the study, you will be asked to use a prototype interface for data visualization and to answer a set of questions about your experience with the prototype. Your answers will be collected anonymously and analyzed for the purpose of the evaluation of the prototype. The time you take to solve each task will also be recorded.

RISKS: To the best of our knowledge, the tasks you are going to perform do not yield to a higher risk of harm than you would experience in everyday life.

BENEFITS: You are not likely to experience any direct benefit from the results of this study. However, the results of the study may yield to new insights in the field of data visualization.

CONFIDENTIALITY

Please do not write any identifying information.

Every effort will be made by the institution to preserve your confidentiality including the following:

- Assigning code names/numbers to participants that will be used on all research notes and documents, to preserve their anonymity
- Keeping notes, interview transcriptions, and any other identifying participant information in a locked file.

Participant data will be kept confidential except in cases where the researcher is legally obligated to report specific incidents. These incidents include, but may not be limited to, incidents of abuse and suicide risk.

CONTACT INFORMATION

If you have questions at any time about this study, or you experience adverse effects as the result of participating in this study, you are free to ask them now. If you have questions regarding your rights as a research participant, or you experience adverse effects as the result of participating in this study, you can contact with the researcher directly by telephone at 017683581082 or via the following email address uhelt@student.kit.edu.

VOLUNTATION PARTICIPATION

Your participation in this study is voluntary. It is up to you to decide whether or not to take part in this study. If you decide to take part in this study, you will be asked to sign this consent form. After you sign the consent form, you are still free to withdraw at any time and without giving a reason. Withdrawing from this study will not affect the relationship you have, if any, with the researcher and the institution. If you withdraw from the study, your data will be returned to you or destroyed.

CONSENT
I have read and I understand the provided information and have had the opportunity to ask questions. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason and without cost. I understand that I will be given a copy of this consent form. I voluntarily agree to take part in this study.

Participant's Name (printed) _____

Participant's Signature _____

Date _____

7.2 Questionnaire

7.2.1 Questionnaire For Participants under Condition A/B/C

7.2.1.1 Participants using Data Set 1

Questionnaire:

1. Basic Information
    1.1. Field of study:
    1.2. Number of semester:
    1.3. Age:

    | *18-24* | *25-30* | *More than 30* |

    1.4. Gender:

    | *M* | *F* | *Do not wish to answer* |

    1.5. How familiar are you with the following concepts?

	1	2	3	4	5	6	7
	Unfamiliar						*Very Familiar*
Correlation analysis							
Data analysis							
Data visualization							

2. Visualization
    2.1. How many attributes are available in this data set?

    2.2. How many pairs of attributes are available in this data set?

    2.3. a) What is the correlation value between **Attribute 2** and **Attribute 3** at **Timestamp T1**, or the probable range?

    b) What is the correlation value between **Attribute 5** and **Attribute 7** at **Timestamp T4001**, or the probable range?

    c) What is the correlation value between **Attribute 6** and **Attribute 8** at **Timestamp T1501**, or the probable range?

    2.4. a) Which pair of attributes has the biggest correlation at **Timestamp T501**?

    b) Which pair of attributes has the biggest correlation at **Timestamp T3001**?

    c) Which pair of attributes has the biggest correlation at **Timestamp T4501**?

2.5. a) Which pair of attributes has the smallest correlation at **Timestamp T501**?

b) Which pair of attributes has the smallest correlation at **Timestamp T3001**?

c) Which pair of attributes has the smallest correlation at **Timestamp T4501**?

2.6. a) The following statement is true or false: "The correlation value between **Attribute 1** and **Attribute 6** remains the same at **Timestamp T2501** and at **Timestamp T3001**"?

          Disagree   *Agree*

b) The following statement is true or false: "The difference of correlation value between **Attribute 1** and **Attribute 8** at **Timestamp T501** and at **Timestamp T1501** is smaller than 0.1"?

          Disagree   *Agree*

c) The following statement is true or false: "The difference of correlation value between **Attribute 1** and **Attribute 2** at **Timestamp T2501** and at **Timestamp T3001** is bigger than 0.2"?

          Disagree   *Agree*

2.7. a) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.8** at **Timestamp T8001**?

b) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.5** at **Timestamp T1001**?

c) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.7** at **Timestamp T3001**?

2.8. a) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.006** at **Timestamp T1001**?

b) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.002** at **Timestamp T9001**?

c) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.002** at **Timestamp T5001**?

3. Feedback

   3.1. Please write down the visualization method you have used and rate for it.

	1 *Strongly* *Disagree*		2	3	4	5	6	7 *Strongly* *Agree*
Intuitive								
Convenient								
Interactive								
Useful								
Complicated								
Efficient								
Effective								

   3.2. In your opinion, what are the strengths of this visualization system?

   3.3. In your opinion, what are the weaknesses of this visualization system?

   3.4. Do you have any suggestions for improving this visualization system?

7.2.1.2 Participants using Data Set 2

Questionnaire:

1. Basic Information
   1.1. Field of study:_____
   1.2. Number of semester: _____
   1.3. Age:

18-24	*25-30*	*More than 30*

   1.4. Gender:

M	*F*	*Do not wish to answer*

   1.5. How familiar are you with the following concepts?

	1		2	3	4	5	6	7
	Unfamiliar							*Very Familiar*
Correlation analysis								
Data analysis								
Data visualization								

2. Visualization
   2.1. How many attributes are available in this data set?

   _____

   2.2. How many pairs of attributes are available in this data set?

   _____

   2.3. a) What is the correlation value between **Attribute 1** and **Attribute 13** at **Timestamp T801**, or the probable range?

   _____

   b) What is the correlation value between **Attribute 10** and **Attribute 18** at **Timestamp T151**, or the probable range?

   _____

   c) What is the correlation value between **Attribute 1** and **Attribute 3** at **Timestamp T901**, or the probable range?

   _____

   2.4. a) Which pair of attributes has the biggest correlation at **Timestamp T701**?

   _____

   b) Which pair of attributes has the biggest correlation at **Timestamp T1701**?

   _____

   c) Which pair of attributes has the biggest correlation at **Timestamp T1601**?

   _____

2.5. a) Which pair of attributes has the smallest correlation at **Timestamp T601**?

_____

b) Which pair of attributes has the smallest correlation at **Timestamp T551**?

_____

c) Which pair of attributes has the smallest correlation at **Timestamp T1451**?

_____

2.6. a) The following statement is true or false: "The correlation value between **Attribute 5** and **Attribute 7** remains the same at **Timestamp T801** and at **Timestamp T851**"?

    Disagree   ***Agree***

b) The following statement is true or false: "The difference of correlation value between **Attribute 1** and **Attribute 15** at **Timestamp T1001** and at **Timestamp T1501** is smaller than 0.1"?

    Disagree   ***Agree***

c) The following statement is true or false: "The difference of correlation value between **Attribute 5** and **Attribute 6** at **Timestamp T1** and at **Timestamp T51** is bigger than 0.2"?

    Disagree   ***Agree***

2.7. a) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.3** at **Timestamp T851**?

_____

b) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.3** at **Timestamp T1951**?

_____

c) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.2** at **Timestamp T601**?

_____

2.8. a) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.001** at **Timestamp T1051**?

_____

b) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.001** at **Timestamp T1901**?

_____

c) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.001** at **Timestamp T801**?

_____

3. Feedback

    3.1. Please write down the visualization method you have used and rate for it.

	1 *Strongly Disagree*		2	3	4	5	6	7 *Strongly Agree*
Intuitive								
Convenient								
Interactive								
Useful								
Complicated								
Efficient								
Effective								

    3.2. In your opinion, what are the strengths of this visualization system?

    3.3. In your opinion, what are the weaknesses of this visualization system?

    3.4. Do you have any suggestions for improving this visualization system?

7.2.1.3 Participants using Data Set 3

Questionnaire:

1. Basic Information
   1.1. Field of study:_____
   1.2. Number of semester: _____
   1.3. Age:
18-24	***25-30***	***More than 30***

   1.4. Gender:
M	***F***	***Do not wish to answer***

   1.5. How familiar are you with the following concepts?

	1 *Unfamiliar*	2	3	4	5	6	7 *Very Familiar*
Correlation analysis							
Data analysis							
Data visualization							

2. Visualization
   2.1. How many attributes are available in this data set?

   _____

   2.2. How many pairs of attributes are available in this data set?

   _____

   2.3. a) What is the correlation value between **Attribute 10** and **Attribute 20** at **Timestamp T401**, or the probable range?

   _____

   b) What is the correlation value between **Attribute 8** and **Attribute 35** at **Timestamp T751**, or the probable range?

   _____

   c) What is the correlation value between **Attribute 7** and **Attribute 25** at **Timestamp T951**, or the probable range?

   _____

   2.4. a) Which pair of attributes has the biggest correlation at **Timestamp T301**?

   _____

   b) Which pair of attributes has the biggest correlation at **Timestamp T51**?

   _____

   c) Which pair of attributes has the biggest correlation at **Timestamp T451**?

   _____

2.5. a) Which pair of attributes has the smallest correlation at **Timestamp T351**?

b) Which pair of attributes has the smallest correlation at **Timestamp T551**?

c) Which pair of attributes has the smallest correlation at **Timestamp T751**?

2.6. a) The following statement is true or false: "The correlation value between **Attribute 1** and **Attribute 37** remains the same at **Timestamp T701** and at **Timestamp T751**"?

                Disagree   *Agree*

b) The following statement is true or false: "The difference of correlation value between **Attribute 10** and **Attribute 20** at **Timestamp T551** and at **Timestamp T601** is smaller than 0.1"?

                Disagree   *Agree*

c) The following statement is true or false: "The difference of correlation value between **Attribute 27** and **Attribute 30** at **Timestamp T801** and at **Timestamp T851** is bigger than 0.2"?

                Disagree   *Agree*

2.7. a) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.95** at **Timestamp T101**?

b) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.995** at **Timestamp T801**?

c) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.0.95** at **Timestamp T1**?

2.8. a) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.001** at **Timestamp T151**?

b) Which pair(s) of attributes has/have a correlation value that is not bigger than **0** at **Timestamp T951**?

c) Which pair(s) of attributes has/have a correlation value that is not bigger than **0** at **Timestamp T551**?

3. Feedback

   3.1. Please write down the visualization method you have used and rate for it.

	1 *Strongly Disagree*		2	3	4	5	6	7 *Strongly Agree*
Intuitive								
Convenient								
Interactive								
Useful								
Complicated								
Efficient								
Effective								

   3.2. In your opinion, what are the strengths of this visualization system?

   3.3. In your opinion, what are the weaknesses of this visualization system?

   3.4. Do you have any suggestions for improving this visualization system?

7.2.2 Questionnaire For Participants under Condition D

7.2.2.1 Participants using Data Set 1

Questionnaire:

1. Basic Information
   1.1. Field of study:_____
   1.2. Number of semester: _____
   1.3. Age:

   | *18-24* | *25-30* | *More than 30* |

   1.4. Gender:

   | *M* | *F* | *Do not wish to answer* |

   1.5. How familiar are you with the following concepts?

	1	2	3	4	5	6	7
	Unfamiliar						*Very Familiar*
Correlation analysis							
Data analysis							
Data visualization							

2. Visualization
   2.1. How many attributes are available in this data set?

   _____

   2.2. How many pairs of attributes are available in this data set?

   _____

   2.3. a) What is the correlation value between **Attribute 2** and **Attribute 3** at **Timestamp T1**, or the probable range?

   b) What is the correlation value between **Attribute 5** and **Attribute 7** at **Timestamp T4001**, or the probable range?

   c) What is the correlation value between **Attribute 6** and **Attribute 8** at **Timestamp T1501**, or the probable range?

   2.4. a) Which pair of attributes has the biggest correlation at **Timestamp T501**?

   _____

   b) Which pair of attributes has the biggest correlation at **Timestamp T3001**?

   _____

   c) Which pair of attributes has the biggest correlation at **Timestamp T4501**?

   _____

2.5. a) Which pair of attributes has the smallest correlation at **Timestamp T501**?

b) Which pair of attributes has the smallest correlation at **Timestamp T3001**?

c) Which pair of attributes has the smallest correlation at **Timestamp T4501**?

2.6. a) The following statement is true or false: "The correlation value between **Attribute 1** and **Attribute 6** remains the same at **Timestamp T2501** and at **Timestamp T3001**"?

                 Disagree   *Agree*

b) The following statement is true or false: "The difference of correlation value between **Attribute 1** and **Attribute 8** at **Timestamp T501** and at **Timestamp T1501** is smaller than 0.1"?

                 Disagree   *Agree*

c) The following statement is true or false: "The difference of correlation value between **Attribute 1** and **Attribute 2** at **Timestamp T2501** and at **Timestamp T3001** is bigger than 0.2"?

                 Disagree   *Agree*

2.7. a) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.8** at **Timestamp T8001**?

b) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.5** at **Timestamp T1001**?

c) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.7** at **Timestamp T3001**?

2.8. a) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.006** at **Timestamp T1001**?

b) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.002** at **Timestamp T9001**?

c) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.002** at **Timestamp T5001**?

3. Feedback

   3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

   ___

   And which of the method did you use the most?

   ___

   Please rate for *Heat map:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

   Please rate for *Bar Graph:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

3. Feedback
   3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

And which of the method did you use the most?

Please rate for *Heat map:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

Please rate for *Bar Graph:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

3. Feedback

   3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

   And which of the method did you use the most?

   Please rate for *Heat map:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

   Please rate for *Bar Graph:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

7.2.2.2 Participants using Data Set 2

Questionnaire:

1. Basic Information
   1.1. Field of study:_____
   1.2. Number of semester: _____
   1.3. Age:

   | *18-24* | *25-30* | *More than 30* |

   1.4. Gender:

   | *M* | *F* | *Do not wish to answer* |

   1.5. How familiar are you with the following concepts?

	1	2	3	4	5	6	7
	Unfamiliar						*Very Familiar*
Correlation analysis							
Data analysis							
Data visualization							

2. Visualization
   2.1. How many attributes are available in this data set?

   _____

   2.2. How many pairs of attributes are available in this data set?

   _____

   2.3. a) What is the correlation value between **Attribute 1** and **Attribute 13** at **Timestamp T801**, or the probable range?

   _____

   b) What is the correlation value between **Attribute 10** and **Attribute 18** at **Timestamp T151**, or the probable range?

   _____

   c) What is the correlation value between **Attribute 1** and **Attribute 3** at **Timestamp T901**, or the probable range?

   _____

   2.4. a) Which pair of attributes has the biggest correlation at **Timestamp T701**?

   _____

   b) Which pair of attributes has the biggest correlation at **Timestamp T1701**?

   _____

   c) Which pair of attributes has the biggest correlation at **Timestamp T1601**?

   _____

2.5. a) Which pair of attributes has the smallest correlation at **Timestamp T601**?

b) Which pair of attributes has the smallest correlation at **Timestamp T551**?

c) Which pair of attributes has the smallest correlation at **Timestamp T1451**?

2.6. a) The following statement is true or false: "The correlation value between **Attribute 5** and **Attribute 7** remains the same at **Timestamp T801** and at **Timestamp T851**"?
          Disagree   *Agree*

b) The following statement is true or false: "The difference of correlation value between **Attribute 1** and **Attribute 15** at **Timestamp T1001** and at **Timestamp T1501** is smaller than 0.1"?
          Disagree   *Agree*

c) The following statement is true or false: "The difference of correlation value between **Attribute 5** and **Attribute 6** at **Timestamp T1** and at **Timestamp T51** is bigger than 0.2"?
          Disagree   *Agree*

2.7. a) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.3** at **Timestamp T851**?

b) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.3** at **Timestamp T1951**?

c) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.2** at **Timestamp T601**?

2.8. a) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.001** at **Timestamp T1051**?

b) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.001** at **Timestamp T1901**?

c) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.001** at **Timestamp T801**?

3. Feedback

   3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

   And which of the method did you use the most?

   Please rate for *Heat map:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

   Please rate for *Bar Graph:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

3. Feedback

3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

And which of the method did you use the most?

Please rate for *Heat map:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

Please rate for *Bar Graph:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

3. Feedback

   3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

   And which of the method did you use the most?

   Please rate for *Heat map:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

   Please rate for *Bar Graph:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

7.2.2.3 Participants using Data Set 3

Questionnaire:

1. Basic Information
   1.1. Field of study: _____
   1.2. Number of semester: _____
   1.3. Age:

   | *18-24* | *25-30* | *More than 30* |

   1.4. Gender:

   | *M* | *F* | *Do not wish to answer* |

   1.5. How familiar are you with the following concepts?

	1	2	3	4	5	6	7
	Unfamiliar						*Very Familiar*
Correlation analysis							
Data analysis							
Data visualization							

2. Visualization
   2.1. How many attributes are available in this data set?

   _____

   2.2. How many pairs of attributes are available in this data set?

   _____

   2.3. a) What is the correlation value between **Attribute 10** and **Attribute 20** at **Timestamp T401**, or the probable range?

   _____

   b) What is the correlation value between **Attribute 8** and **Attribute 35** at **Timestamp T751**, or the probable range?

   _____

   c) What is the correlation value between **Attribute 7** and **Attribute 25** at **Timestamp T951**, or the probable range?

   _____

   2.4. a) Which pair of attributes has the biggest correlation at **Timestamp T301**?

   _____

   b) Which pair of attributes has the biggest correlation at **Timestamp T51**?

   _____

   c) Which pair of attributes has the biggest correlation at **Timestamp T451**?

   _____

2.5. a) Which pair of attributes has the smallest correlation at **Timestamp T351**?

b) Which pair of attributes has the smallest correlation at **Timestamp T551**?

c) Which pair of attributes has the smallest correlation at **Timestamp T751**?

2.6. a) The following statement is true or false: "The correlation value between **Attribute 1** and **Attribute 37** remains the same at **Timestamp T701** and at **Timestamp T751**"?

> *Disagree*   *Agree*

b) The following statement is true or false: "The difference of correlation value between **Attribute 10** and **Attribute 20** at **Timestamp T551** and at **Timestamp T601** is smaller than 0.1"?

> *Disagree*   *Agree*

c) The following statement is true or false: "The difference of correlation value between **Attribute 27** and **Attribute 30** at **Timestamp T801** and at **Timestamp T851** is bigger than 0.2"?

> *Disagree*   *Agree*

2.7. a) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.95** at **Timestamp T101**?

b) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.995** at **Timestamp T801**?

c) Which pair(s) of attributes has/have a correlation value that is not smaller than **0.0.95** at **Timestamp T1**?

2.8. a) Which pair(s) of attributes has/have a correlation value that is not bigger than **0.001** at **Timestamp T151**?

b) Which pair(s) of attributes has/have a correlation value that is not bigger than **0** at **Timestamp T951**?

c) Which pair(s) of attributes has/have a correlation value that is not bigger than **0** at **Timestamp T551**?

3. Feedback

3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

And which of the method did you use the most?

Please rate for *Heat map:*

	1 *Strongly* *Disagree*	2	3	4	5	6	7 *Strongly* *Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

Please rate for *Bar Graph:*

	1 *Strongly* *Disagree*	2	3	4	5	6	7 *Strongly* *Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

3. Feedback

   3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

And which of the method did you use the most?

Please rate for *Heat map:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

Please rate for *Bar Graph:*

	1 *Strongly Disagree*	2	3	4	5	6	7 *Strongly Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

3. Feedback

    3.1. As you are able to use all visualization methods during the research, which one of the method, in your opinion, is the most helpful method to fulfill the task?

    And which of the method did you use the most?

    Please rate for *Heat map:*

	1 *Strongly* *Disagree*	2	3	4	5	6	7 *Strongly* *Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							

    Please rate for *Bar Graph:*

	1 *Strongly* *Disagree*	2	3	4	5	6	7 *Strongly* *Agree*
Intuitive							
Convenient							
Interactive							
Useful							
Complicated							
Efficient							
Effective							