Interactive Visualization of Survey Results

by

Zheng Xu
B.Eng., Tsinghua University, 2005
M.Sc., University of Warwick, 2006

A Master’s Project Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

© Zheng Xu, 2016
University of Victoria

All rights reserved. This Project may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.
Interactive Visualization of Survey Results

by

Zheng Xu
B.Eng., Tsinghua University, 2005
M.Sc., University of Warwick, 2006

Supervisory Committee

Dr. Melanie Tory, Supervisor
(Department of Computer Science)

Dr. Margaret-Anne Storey, Departmental Member
(Department of Computer Science)
Supervisory Committee

Dr. Melanie Tory, Supervisor  
(Department of Computer Science)

Dr. Margaret-Anne Storey, Departmental Member  
(Department of Computer Science)

ABSTRACT

Conducting surveys is a common practice in industry, science, and education. Current online survey tools focus on collecting data but not on effective visualization of the results. This project report describes a visualization approach to help lay-users to dynamically explore survey results. Moreover, the report proposes that for exploratory tasks, the lay-user should focus on choosing the data that is relevant to them and the system should infer the type of charts that are appropriate for the chosen data. Based on this proposal, a dashboard style web tool was designed and implemented. This tool automatically generates visualizations that are appropriate for the current data selection, and it also has a series of interaction features such as chart reordering and brushing / linking. A pilot study was also conducted to evaluate this tool and the preliminary conclusion was that the tool could help the user to complete some survey tasks and enable exploration.
Contents

Supervisory Committee ii
Abstract iii
Table of Contents iv
List of Tables vi
List of Figures vii
Acknowledgements ix
Dedication x

1 Introduction 1
  1.1 Uses of Surveys .................................................. 1
  1.2 Problem Description ............................................... 3
  1.3 Structure of the Report .......................................... 3

2 Related Work 5
  2.1 Research on Survey Data Visualization and Exploratory Visualization 5
  2.2 Challenges Faced by Lay-users .................................. 9
  2.3 Methods of Survey Data Visualization .......................... 12

3 Design Principles 14
  3.1 Experiences with Industry Software ............................ 14
  3.2 Interface Approach ............................................... 15
  3.3 Target Users ..................................................... 16
  3.4 Supported Tasks .................................................. 16
  3.5 Interaction ....................................................... 18
List of Tables

Table 1.1 Types of questions in a survey ........................................ 2
Table 3.1 Typical survey tasks ......................................................... 17
Table 4.1 Content of tooltips upon hover on visualization elements .... 36
Table B.1 The relationships between the question types and chart types in the overview tab ......................................................... 44
Table B.2 The relationships between the question types and chart types in the query tab ......................................................... 45
# List of Figures

2.1 OpinionSeer by Wu et al.[3] .............................................. 6
2.2 Competitive relation map by Xu et al.[4] ............................. 6
2.3 Coordinated graph representing positive and negative terms by Chen et al.[5] ............................................................ 7
2.4 Hierarchical Keyword Graph by Uchida et al.[1] ..................... 7
2.5 Co-occurrence graph by Sheng et al.[2] ............................... 8
2.6 Coordinated-and-multiple-view design in Improvise by Weaver[8] 8
   (a) Visualization of simulated ion trajectory (scientific data) .... 8
   (b) Visualization of election results .................................. 8
2.7 Visual Search Engine by Boukhelifa et al.[9] .......................... 9
2.8 City‘O’Scope by Brodbeck et al.[10] ..................................... 9
2.9 Cdv tool by Dykes[11] ..................................................... 10
2.10 CommonGIS by Andrienko et al.[12] ................................... 10
2.11 The brushing function in scatterplot matrices by Becker[13] .... 11

4.1 Use case scenarios and visualizations (implemented) ............. 23
4.2 Use case scenarios and visualizations (not implemented) ....... 24
4.3 Interface screenshot of the 1st version ............................. 25
4.4 Interface screenshot of the 2nd version ............................ 27
4.5 Interface screenshot of the 3rd version ............................. 28
4.6 Overview tab view ...................................................... 29
4.7 Selectors in overview tab .............................................. 30
   (a) Survey selector ................................................. 30
   (b) Question selector ............................................. 30
4.8 Query tab with an empty chart and with ”add new chart” button available .................................................. 31
4.9 Query tab with three different charts: heat map, correlation matrix and stacked bar chart ............................... 31
4.10 Response reordering and synchronizing .................................................. 32
   (a) Dragging "Satisfied" to the next lower position in a bar chart .................. 32
   (b) Dragging "Satisfied" to the next right hand side position in a heat map ....... 32
   (c) Dragging "Satisfied" to the next lower position in a heat map ................. 32
   (d) Dragging "Satisfied" to the next lower position in a stacked bar chart ....... 32
   (e) By performing any one of the above actions, order of responses can be synchronized in all charts .................................................. 32

4.11 Examples of actions to invoke brushing .................................................. 33
   (a) Clicking the first bar in bar chart ...................................................... 33
   (b) Clicking first bar in histogram ......................................................... 33
   (c) Clicking the first bar of the first question "Q4a" in stacked bar chart ......... 33
   (d) Selecting an rectangle area in scatterplot .......................................... 33

4.12 A sample of brushing effect by select “Master Card” in “Q1”. Note that the small multiple of text response is also brushed while the bar chart of a ranking question is not brushed .................................................. 35
   (a) Overview tab after brushing .............................................................. 35
   (b) Query tab after brushing ................................................................. 35
ACKNOWLEDGEMENTS

I would like to thank:

My parents, my fiancé, my friends and my colleagues in VisID group, for supporting me in the low moments.

Dr. Melanie Tory, for mentoring, support, encouragement, and patience.

Dr. Maria-Elena Froese, for raising the project idea and supporting me in the whole implementation process.
DEDICATION

I dedicate this project report to my fiancé Simon who always supports me.
Chapter 1

Introduction

Survey is a research method used to gather information such as preferences, opinions and factual information from a sample of the population. Surveys are frequently used by organizations to review and improve quality of services or products. There are many online services and tools that facilitate creating surveys and thus have potential to generate a large amount of survey data. This project focuses on how to help non-visualization experts to construct charts from survey data. Based on literature reviews about survey data visualization and analysis of some current visualization tools, the project proposes a web-based and dashboard-style visualization tool. A pilot study was also conducted as a preliminary assessment of this tool. The Javascript code of the visualization tool is located in an open-source repository on GitHub \(^1\).

1.1 Uses of Surveys

There are a lot of scenarios where surveys are conducted. For example, companies use questionnaires to collect data about user experience on products and systems. Teachers may use quizzes and teaching / learning evaluations to collect feedback from students; governments and political parties can use surveys to get opinions from a particular demographic group about a social or political issue.

Characteristics of a survey, especially types of questions, can vary depending on its topic. For example, surveys about satisfaction mainly consist of multiple-choice questions; surveys about purchasing behaviours may contain a lot of numeric questions; surveys about preferences on products may consist of ranking questions. Table

\(^1\)“Survey Dashboard” https://github.com/xuzheng0927/SurveyDashboard
<table>
<thead>
<tr>
<th>Question type</th>
<th>Example</th>
<th>Answer Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple choice: select one or more</td>
<td>Are you a: A. Undergraduate student B. Masters student C. PhD student</td>
<td>One or more</td>
</tr>
<tr>
<td>answers from a list</td>
<td></td>
<td>categorical values</td>
</tr>
<tr>
<td>Ranking question: give ranks for</td>
<td>Rank the following teaching styles in order of your preference. A. Pure lecture B. Discussion C. Watching videos</td>
<td>One array</td>
</tr>
<tr>
<td>multiple items</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Binary question: give one yes/no</td>
<td>Will you recommend this course to other students? A. Yes B. No</td>
<td>Binary value</td>
</tr>
<tr>
<td>response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likert-scale question: give</td>
<td>How do you like flipped classroom approach? A. Like B. Neutral C. Dislike</td>
<td>One ordinal value</td>
</tr>
<tr>
<td>positive / negative feedback</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open question with numeric response</td>
<td>How many hours do you spend in readings or watching the videos before lecture?</td>
<td>One numeric value</td>
</tr>
<tr>
<td>Open question with textual response</td>
<td>Do you have any comments on the flipped approach?</td>
<td>One textual value</td>
</tr>
<tr>
<td>Open question with date / time</td>
<td>When is the last time that you read the materials or watched the videos?</td>
<td>One date / time object</td>
</tr>
<tr>
<td>response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open question with image response</td>
<td>Show a screenshot of your work done in this course.</td>
<td>One image object</td>
</tr>
</tbody>
</table>

Table 1.1: Types of questions in a survey

1.1 shows a list of question types that are frequently used.

Online survey is currently being widely used and has a large population of users; thus it could be inferred that there is a great need for effective analysis tools for survey data. There are many online services that help to create customized online surveys, such as SurveyMonkey [23], SurveyGizmo [24], KwikSurveys [25], etc. These online survey services own a large population of users, e.g., more than 20 million users are using SurveyMonkey. Most online survey services support their users to view summaries of surveys, and more importantly, to analyze the results with customized visualization of survey data. The users can also choose to download their survey data and then perform further analysis with outside tools.
1.2 Problem Description

In the last section, it is inferred that there is a great demand for effective analysis tools for survey data. Among the survey analysis tools, the visualization tool plays an important role. However, from the perspective of “lay-users”, the current survey visualization tools are limited for effective analysis.

From the websites of online survey services, it can be seen that there are many enterprise users making use of the services for business-level purposes, such as market research, investigation of customer satisfaction, etc. Based on the experience interacting with survey companies, the majority of these users are unlikely to have substantial experience in data analytics, statistics, or visualization. Thus the project aimed to design a survey visualization tool for “lay-users”, or people who have limited expertise in these topics. Lay-users have limited knowledge on information visualization and visual data analysis; their visualization choices are driven by heuristics and familiarity with visualization types [14]. Meanwhile, most of the online survey services provide static charts that only represent individual questions. These charts offer limited opportunity for deep exploration especially correlations between different questions.

Due to the above, it is common that users turn to use other visualization tools outside the online survey services. Such tools include web-based visualization tools, programming toolkits for visualization and industry visualization software. From the perspective of lay-users, these tools also have their own limitations in terms of survey data visualization: web-based tools also provide static and individual-question-based chart; programming toolkits require additional programming skills; industry software has efficiency problems dealing with large number of questions. These limitations are described in detail in the next chapter. Therefore, a visualization tool customized for survey data and lay-users is needed.

1.3 Structure of the Report

The project report contains 7 chapters. The contents of each chapter are described as follows.

Chapter 1 contains introduction of the background and the problem that this project focused on, followed by an overview of the structure of the report itself.
Chapter 2 describes the related work, lay-users’ challenges and the limitations of current solutions / approaches.

Chapter 3 describes the design principles of the project.

Chapter 4 describes the design of visualizations, user interface and interaction features in detail.

Chapter 5 describes the pilot user study against the designed tool.

Chapter 6 includes the discussions about the pilot user study, limitations of the designed tool and challenges met in this project.

Chapter 7 draws the conclusion of this project and describes future work.
Chapter 2

Related Work

2.1 Research on Survey Data Visualization and Exploratory Visualization

This section first describes some research papers in the area of survey data visualization and investigates their usability for lay-users. Since survey analysis is an exploratory process, some papers that did not specifically focus on survey data visualization, but focus on other areas of exploratory visualization, are also reviewed to see whether these methods are also applicable in survey data visualization.

There have been many research papers about survey data visualization. Their limitations can be summarized in two aspects. The first limitation is that many visualization tools are specific to certain topics, and this is not favourable for a general use. For example, Wu et al.[3] designed OpinionSeer, a wheel-style tool to visualize opinions of hotel customers (Figure 2.1), but it relies on survey datasets consisting of satisfaction rates mostly. Xu et al.[4] designed a competitive relation map to compare between competing mobile phones (Figure 2.2), but the visualization is only for visualizing competitive relations. The second limitation is that many of these visualizations have complex structures built by visualization experts; if lay-users of visualization want to use them to build charts from their own data, it is difficult for them to make visual mappings, i.e., map data elements to visualization elements[14]. For example, Chen et al.[5] used coordinated graph to visualize conflicting opinions. Uchida et al.[1] used Hierarchical Keyword Graph (Figure 2.4) to extract and visualize important keywords in free text of questionnaire data. Sheng et al.[2] used Co-occurrences graph (Figure 2.5) to present relationships between responses in choice-type
The applications all require relatively complex query abstraction and data mapping to build up the chart.

Figure 2.1: OpinionSeer by Wu et al.[3]

Figure 2.2: Competitive relation map by Xu et al.[4]

There have also been some research papers indicating that less complex charts are more favourable in survey data visualization. Shamim et al.[6] evaluated a group of opinion mining systems with visualizations. The evaluation data was collected by conducting seminars and using a web-based online questionnaire. A comprehensive set of metrics were used for the evaluation: eye pleasing, easy to understand, user-friendly, informative design, intuitive design, usefulness, comprehensiveness, comparison ability, representation style, pre-knowledge required. They concluded that the most effective visualizations were simple, easy to understand. For example, bar chart, glowing bar, tree map, line graph and pie chart were the top five visualizations in their
results. It should be noticed that most of the participants took a “Human-Computer Interaction” course before the experiment and had more knowledge about visualization techniques than lay-users, thus they were supposed to have a better sense about how to construct or interpret complex visualizations; however, the experiment results showed that they preferred visualizations with less complexity. If even this experienced group preferred simple visualizations, it should be inferred that keeping the interface simple may be even more important for lay-users.

In terms of research about exploratory visualizations, some relevant research papers were reviewed to search for the interaction features that are also favourable for survey data visualization. The charts were also evaluated to see whether they can also be applied to survey data visualization. Roberts[7] made a comprehensive survey about the applications of coordinated and multiple views. For example,
Figure 2.5: Co-occurrence graph by Sheng et al. [2]

(a) Visualization of simulated ion trajectory (scientific data)
(b) Visualization of election results

Figure 2.6: Coordinated-and-multiple-view design in Improvise by Weaver [8]

Weaver [8] used Improvise to build visualizations of scientific data and election results (Figure 2.6); Boukhelifa et al. [9] used a multi-view model to build Visual Search Engine (Figure 2.7); Brodbeck et al. [10] designed City’O’Scope, which used multiple coordinated views to analyze geo-referenced high-dimensional datasets (Figure 2.8). Meanwhile Roberts’ paper also demonstrated the wide use of brushing and linking in exploratory visualizations. For example, Dykes [11] designed the “cdv tool” (Figure 2.9) showing various geographical and statistical views that are linked together; Andrienko et al. [12] used multiple linked views with the space-time cube visualization in CommonGIS (Figure 2.10); Becker et al. [13] applied brushing to scatterplot matrices (Figure 2.11). From the above research works, it can be seen that coordinated and multiple views as well as brushing and linking are generic techniques used in exploratory visualizations. Therefore, they are also potentially helpful features
in survey data visualization, thus these techniques were applied in the tool design. While these features are useful and could be adopted, the tools themselves could not be directly used for survey data visualization because their designs are specific to other kinds of data.

From the above, it may be inferred that there is a gap of survey data visualization tools for lay-users. To further investigate what would be an appropriate visualization method for lay-users, further analysis was performed on the challenges faced by lay-users.

### 2.2 Challenges Faced by Lay-users

In this section, the challenges faced by lay-users in survey data visualization are discussed in detail. Some guidelines for tool design are given at the end in order to tackle these challenges.

Through an exploratory laboratory study, Grammel et al.[14] discovered that the
major barriers faced by lay-users were: translating question into data attributes, designing visual mappings and interpreting the visualizations.

Specifically, a lay-user faces the following challenges when creating visualizations:

1. **Data cleaning and formatting**: the raw data of a survey may contain information that is not relevant and thus needs to be cleaned before it is read in; also, each data source will have its own designated format of output data. All these need to be manually done by the user.

2. **Query abstraction**: after getting the formatted data, the user will need to transform every query on the data into a concrete and effective representation of a certain software / tool.

3. **Data mapping**: to generate a chart, the user will need to configure the mapping relationships between data and the chart elements. For example, the user needs to map different dimensions of data into axes, colour of marks, size of marks, etc., of a chart.

4. **Chart type selection**: there can be multiple chart types that can represent the
same dataset. The user will need to decide and pick a chart type that is more effective compared to the others. However in actual fact, lay-users tend to follow their heuristics and familiarity with chart types\cite{14}, and their familiar charts could not guarantee effectiveness.

As for Challenge 1, many visualization tools already have good support on data formatting. For example, Tableau supports several input formats, which cover most cases of data files; data cleaning is generically required in any visualization tool so it is not considered to be specific to lay-users. However, lay-users will still face Challenges 2,3 and 4, which can lead to ineffective charts, i.e., the generated charts cannot reflect characteristics of data or can be misleading. In this case, the users may need extra attempts to get an effective chart. Furthermore, for the specific case of survey data visualization, another challenge would be how to deal with a large number of questions. Typically the user will need to repeatedly perform the same steps for a group of questions of the same type.

To tackle these challenges, Heer et al.\cite{15} devised several guidelines: facilitate user-friendly data input, provide automatic selection of visualization type using sensible defaults, and provide contextual information that explains which data is displayed and which encodings are used. This project followed the latter two guidelines during the design process of the new tool.
2.3 Methods of Survey Data Visualization

As mentioned in Section 1.2, it is common for users of online survey services to use outside tools for data visualization. Based on the context of survey data visualization, these tools can be categorized into three classes: web-based visualization tools, programming toolkits for visualization and industry visualization software. This section analyzes the usefulness of each class for lay-users and describes their limitations respectively.

Web-based visualization tools, e.g., ManyEyes[26], iCharts[27], etc., allow users to create their own dashboards, with only several simple steps, to visualize data using basic charts such as line charts, bar charts and pie charts. To access these tools, the user only needs a web-browser, and no extra software installation or configuration is required. These tools basically have simple interfaces for users to manipulate. These advantages lend web-based tools to a broad range of user populations. However, although their charts can present general overviews of survey data, they have limited ability to explore at a deeper level, especially the relationships between questions. Elias et al.[16] also found that although such web-based tools target lay-users, basically the users are restricted to view and share a single visible chart (based on a single question) at a time, and linking of multiple charts is not applicable. Moreover, their charts are mostly static or have very limited interactivity, which also limits their explorability.

Programming toolkits with visualization functions or features, such as Flare [28] and D3JS[29], allow people to build custom tools to visualize survey data. These toolkits provide plenty of functions for users to create customized charts with good interactivity. Elias et al.[16] and Pantazos et al.[17] investigated a series of visualization toolkits and found these tools do not target lay-users because they usually require additional programming to create the visualization. Since most lay-users are not experts in visualization programming, they are constrained from utilizing these tools without the assistance of a programmer.

Industry visualization software like Omniscope[30], Spotfire[31] and Tableau[32] can also be used for survey data visualization. Pantazos et al.[17] found such industry products are marketed towards lay-users, assisting them with predefined visualization templates. In addition, users can derive new data features by simple coding, i.e., the users only need to write some conditional statements, arithmetic or logical expressions rather than functions or complex data structures. Moreover, the included charts are
capable of presenting more complicated data structures. These tools can be ideal for creating a small number of customized charts. However, since creating every single chart requires a series of manual manipulations, when the number of datasets or questions is large, the repeated manipulations can still lead to low efficiency.

Based on the analysis of lay-users' challenges, it could be inferred that a good survey visualization tool for lay-users needs to meet the following requirements: it should provide dynamic, interactive charts based on multiple questions; it should not need any prerequisite skills (e.g., programming skills) or pre-knowledge in data visualization; it should help ensure the effectiveness of the generated charts (e.g., correctly reflecting the characteristics of data, not misleading, etc.); and it should have the ability to generate a large number of charts quickly. Therefore, none of the above solutions fully meet these requirements.
Chapter 3

Design Principles

This chapter describes the design principles of the new tool. A set of terms and definitions used related to surveys and survey data are listed in Appendix A. The conclusions of Chapter 3 were used as guidelines for the tool design.

3.1 Experiences with Industry Software

Before proceeding to the tool design phase, some industry visualization software was tested and several design principles were drawn from the perspective of survey users. They were used to create charts from two sample survey datasets: the first one was a survey of credit card feedback with 22 respondents and 11 questions; the second one was about demographics (gender, generation, etc.), health (pulse rate, blood pressure, etc.) and personal characteristics (good sense of humor, high intelligence, etc.), with 845 respondents and 20 questions. Both datasets contained two or more types of questions.

Microsoft Office Excel and Tableau were used as the testing software. Compared to web-based visualization tools, they did provide more options of chart types, e.g., Tableau provided heat map, stacked bar chart, tree map, scatter plot, etc. Although creating some of the charts required conversion or derivation of the data, it could be achieved by several simple steps, such as adding some logical or mathematical expressions. When dealing with the first survey, because there were only 11 questions, it was easy to create all the charts needed to explore the data of interest. However, when it came to the second survey, as there were more questions, the repetition of creating and modifying charts became an issue. Inappropriate selection of chart type
also happened at times. These results indicate that the new tool should provide abundant options of chart types and be able to generate a large number of charts quickly.

3.2 Interface Approach

A dashboard with appropriate visual representations and interactive functions is a good way to enhance information understanding. There have been a lot of application cases using dashboard design for data visualization. Elias et al.[16] designed Exploration Views, a system that allows novice visualization users to easily build and customize Business Intelligence information dashboards. Maldonado et al.[18] designed an interactive dashboard to visualize collaborative learning data. Aires et al.[19] used dashboard design and implemented a web tool to visualize personal information.

Two categories of solutions were considered:

- An interactive dashboard with a pre-defined set of charts that portray the survey results. This solution allows the user to dynamically change the view along with the selected data.

- A wizard style dashboard to guide the user through the creation of charts. This solution allows users to create charts in a step-by-step way and gives users more control.

Based on the characteristics of the sample survey datasets as well as the experiences with industry software described in the last section, it was found that an interactive dashboard would be more suitable for exploration. In other words, the user will frequently need to adjust his / her selection of survey questions and type of chart. If a wizard style dashboard is used, the user may need to go back and forth to adjust the settings (selected questions, type of chart, etc.) in order to get a satisfying chart. In contrast, an interactive dashboard can help the user to modify the settings in real-time.

Finally, this tool should be easy to access; thus the dashboard was designed as an online web tool so that users only need to have an internet browser to access the dashboard.
3.3 Target Users

As mentioned in Chapter 1, this project focuses on the needs of users who are not visualization experts. Grammel et al. [14] defined such users as “InfoVis Novices”, i.e., those who are “not familiar with information visualization and visual data analysis beyond the charts and graphics encountered in everyday life. InfoVis novices can be domain experts in their area of expertise (subject matter experts) and the data they are analyzing can be from this area”. These users are not as proficient in creating charts as visual data analysts; instead they are lay-users who want to perform exploratory tasks in their own area of expertise, such as getting user opinions or ratings on a service or product.

3.4 Supported Tasks

Task abstraction allows to describe the user tasks in a non-domain-specific way. This is important when deriving visualizations from the domain-specific tasks. In this section Munzner’s method is used for task abstraction, in which tasks consist of actions and targets [21]. An action can be analyzing, searching or querying; a target can be outliers, distribution, correlation, etc.

As mentioned in Chapter 1, the new tool should be able to visualize survey data containing various types of questions, e.g., the listed types in Table 1.1. Due to time limits in this project, only the most basic types of survey questions were dealt with: multiple choice, binary, Likert-scale, numeric-response and rank-response questions. The free text questions were also included, but no attempt was made to perform visual summarization of the response text. Based on the types of questions, the next step was to get a clear image of what kinds of tasks the survey analysts, i.e., the users, want to perform against the survey.

A series of tasks (Table 3.1) the most likely to be performed were derived by setting up use cases on a sample course survey. For example, suppose the instructor teaching this course adopts a new method called the “flipped classroom approach”; there are also additional requirements of reading and watching videos. When the instructor receives the course feedback in a survey, the first thing he/she wants to see is probably an overview, e.g., is the overall rating positive or negative. If there are some questions with many negative responses, he/she may need to keep only these questions in view and do some analysis. Next, when there is a question giving general feedback on this
### Table 3.1: Typical survey tasks

<table>
<thead>
<tr>
<th>Task descriptions</th>
<th>Example</th>
<th>Task abstraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>The user wants to get an overview of the survey.</td>
<td>“I want to have a sense of the overall responses.”</td>
<td>Present trends and distribution</td>
</tr>
<tr>
<td>The user wants to see statistics only for selected questions.</td>
<td>“I only want to see students’ feedback on the pace of lecture and the responses to their frequency of watching videos.”</td>
<td>Browse data</td>
</tr>
<tr>
<td>The user wants to see correlations between two questions.</td>
<td>“I want to see if there is a correlation between how the students found the pace of the lecture and their frequencies to watch the videos and do the readings before class.”</td>
<td>Discover correlation</td>
</tr>
<tr>
<td>The user wants to see a subset of respondents’ statistics in certain questions.</td>
<td>“For the students giving negative feedback on the flipped approach, how did they feel about the pace of lecture and how did they feel about the helpfulness of instructor?”</td>
<td>Filter and browse data</td>
</tr>
<tr>
<td>The user wants to see full text responses of selected respondents.</td>
<td>“I want to see the text responses of the students that did not like the flipped classroom approach.”</td>
<td>Filter and browse data</td>
</tr>
</tbody>
</table>

course (e.g., the feedback on flipped classroom approach), the instructor may want to know from the other questions how the specific factors (e.g., how students feel about the pace of lecture) affect the general feedback; if he/she discovers a factor that has positive/negative correlation with the general feedback, he/she can make improvement on this factor with priority. The instructor may also want to select a subset of students that provided negative feedback in a certain question; then he/she can look into their statistics or text responses in the other questions to find the cause of negative feedback.

The new tool should be able to assist survey analysts to complete the listed tasks.
3.5 Interaction

As mentioned in Section 3.2, to support exploration tasks on survey data, the new tool was defined as an interactive dashboard. Below is a list of interactive scenarios and the functions that the new tool aimed to support, based on the use cases:

- The users need to create a number of charts, each of which has its own settings and corresponding chart. Therefore, the tool should allow users to modify settings of each chart separately.

- The users need to customize layout of the dashboard. For example, the users may need to put relevant charts together, enlarge an important chart or remove an irrelevant chart. Therefore, the tool should make the charts reorderable, resizable and removable.

- After creating charts, the users want to select a subset of respondents in a certain chart to see responses of this specific group in other charts. Therefore, the tool should have brushing and linking features so that the users can highlight a subset of data in a panel then see its distributions in all the other panels.

It should be noticed that except for some special occasions, the users were assumed to have little access to change the parameters of a chart (chart type, colour, value range of coordinates, etc.) since these charts were supposed to be pre-defined rather than customized.

3.6 Programming Language and Libraries

Since the new tool was defined as an online web tool, Javascript was used as the basic programming language and jQuery as DOM (Appendix A) manipulation library. D3JS[29] was used as the library of graphics because of its power to produce dynamic and interactive visualizations. Bootstrap[35] and jQuery UI[36] were used as the framework libraries since the tool should have some enhanced interaction features such as drop-down selectors, draggable widgets, tabs, etc.
3.7 Design Process

To guide implementation of the tool, a procedure with the following phases was used in this project: gathering requirements, designing low fidelity prototypes, evaluation and refinement. Among these phases, evaluation and refinement were performed iteratively until the design reached a satisfactory effect.
Chapter 4

Survey Visualization Tool Design

This chapter illustrates all the details of the tool design. The first section describes the data pre-processing phase before building the tool. The second section introduces the visualizations composed for the new tool. The third section describes the evolutionary design process of the interface design. The fourth section introduces a series of interaction features of the tool. The Javascript code of this tool is located in an open-source repository on GitHub.

4.1 Data Pre-processing

A phase of data formatting and input was conducted before building the tool. Most of the raw data from surveys is stored in spreadsheets (typically Excel files) and CSV files. Among these two, CSV files can be directly read by web scripts, but since it uses commas to separate different fields, it will cause problems when the contents of fields also contain commas (such as responses to open questions). Therefore, JSON was used as the format to store survey data.

Nevertheless, there were still two remaining problems for the CSV-JSON conversion. The first and the minor one was that not all CSV raw data provided an index for each question, e.g., “Q5” for “How do you find the pace of the lecture?” and these indices might need to be added by a script beforehand. The second and the major one was that few raw datasets contained information on the types of questions, i.e., the categories of questions (multiple choice, open response, ranking, etc.); for the purposes of exploring the tool design, the types of questions were manually added to

1 Survey Dashboard” https://github.com/xuzheng0927/SurveyDashboard
the test datasets. Then an online CSV-JSON converter can be used to transform the prepared CSV file into a JSON file.

As for raw data stored in spreadsheets, since there is not a uniform format for it, different data sources have different spreadsheet formats. Moreover, most of their formats cannot be directly interpreted as CSV or JSON, so except for case-by-case manual conversions, there was not a very workable way to deal with spreadsheets.

4.2 Visualizations

According to Munzner [21], channels (or visual variables) consist of two classes: identity channels for categorical attributes and magnitude channels for ordered attributes. Identity channels include the following, in the order of effectiveness: spatial position, colour hue, motion and shape. Magnitude channels include the following, in the order of effectiveness: position on scale, length, tilt/angle, area, depth, colour luminance, colour saturation, curvature and volume.

A survey dataset can have the following categorical attributes: wordings of questions and wordings of multiple-choice responses. It can also have the following ordered attributes: numbers of respondents, values of numeric responses and ranks. Moreover, it can also have attributes derived from the existing attributes, such as correlation coefficients between responses to two different questions, which are ordered.

Based on Munzner's effectiveness guidelines, categorical attributes are visualized using either spatial position or colour hue. Similarly, the number of respondents, values of numeric responses, ranks are presented by length. Two sets of numeric responses are presented in a coordinate graph, which consists of positions on two different scales. The only exception is when a matrix is used for cross-comparing, due to the restriction by the matrix grids, colour luminance is used to present the numbers of respondents or values of correlation efficient.

Based the above principles as well as Table 3.1, a collection of visualizations were designed to be suitable for answering the questions in the target use cases (see Figures 4.1 and 4.2). The visualizations that have been implemented (listed in Figure 4.1) in this project include:

- Bar chart (UC01-1): the length of each bar stands for the number of respondents to its corresponding response. It should be noticed that there is a special type

---

2This part of design was done by the author’s colleague, Dr. Froese
of categorical question: ranking question. If a ranking question is selected, the width of each bar stands for the average rank of its corresponding response.

- Histogram (UC01-2): the length of each bar stands for number of respondents to its corresponding bin (value range).

- Heat map (UC01-3 and UC01-4): the darkness (opposite of luminance) of a cell stands for the total number of respondents who chose the corresponding pair of answers.

- Stacked bar chart (UC01-5): the length of each bar stands for the number of respondents to its corresponding response.

- Scatter plot (UC01-7): every single point in the chart corresponds to a respondent and its position indicates the values of responses for the two selected questions.

- Correlation matrix (UC01-8): each cell of the correlation matrix uses darkness to encode the correlation coefficient of the corresponding pair of questions; the more the coefficient is, the more correlated the two questions are. Each correlation matrix is appended with a scatter plot.

Due to time limit in this project, the other three optional visualizations were not implemented (listed in Figure 4.2): UC01-3 (alternative visualization), UC01-6 and UC01-9.
<table>
<thead>
<tr>
<th>SCENARIOS</th>
<th>USER SELECTION</th>
<th>DEFAULT VISUALIZATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCDS-1 MAIN SCENARIO: Step 1: The user selects ONE categorical question. Step 2: The system displays a bar chart. Note that all the values of a question are selected automatically. The user can remove values if needed desires so.</td>
<td>Survey Name: CDS115 Midterm Survey (Fall 2016)</td>
<td>Always or nearly always</td>
</tr>
<tr>
<td>Question(s): Q2: How often do you do the readings or watch the videos BEFORE section?</td>
<td></td>
<td>Never or almost never</td>
</tr>
<tr>
<td>MCDS-2 ONE numerical question: Step 1: The user selects ONE numerical question. Step 2: The system displays a histogram.</td>
<td>Survey Name: CDS115 Midterm Survey (Fall 2016)</td>
<td></td>
</tr>
<tr>
<td>Question(s): Q6: Outside of lectures, how much time do you devote to this class?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCDS-2 TWO categorical questions with the same values: Step 1: The user selects TWO categorical questions with the same values. Step 2: The system displays a grouped bar chart. Step 3: The user can select a correlation matrix, a table, or a grouped bar chart.</td>
<td>Survey Name: CDS115 Midterm Survey (Fall 2016)</td>
<td></td>
</tr>
<tr>
<td>Question(s): Q2: How often do you do the readings or watch the videos BEFORE section?  Q4: How often do you do the readings or watch the videos EVENTUALLY?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCDS-2 TWO categorical questions with different values: Step 1: The user selects TWO categorical questions with different values. Step 2: The system displays a stacked bar chart.</td>
<td>Survey Name: CDS115 Midterm Survey (Fall 2016)</td>
<td></td>
</tr>
<tr>
<td>Question(s): Q2: How often do you do the readings or watch the videos BEFORE section?  Q4: How often do you do the readings or watch the videos EVENTUALLY?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCDS-2 MANY categorical questions with the same categorical values: Step 1: The user selects more than two categorical questions with the same values. Step 2: The system shows a stacked bar chart (horizontal).</td>
<td>Survey Name: CDS115 Midterm Survey (Fall 2016)</td>
<td></td>
</tr>
<tr>
<td>Question(s): Q2: How often do you do the readings or watch the videos BEFORE section?  Q4: How often do you do the readings or watch the videos EVENTUALLY?  Q6: How often do you affect the lectures?  Q8: How often do you do the quizzes?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCDS-2 TWO numeric questions: Step 1: The user selects two numeric questions. Step 2: The system displays a scatterplot chart.</td>
<td>Survey Name: CDS115 Midterm Survey (Fall 2016)</td>
<td></td>
</tr>
<tr>
<td>Question(s): Q6: Outside of lectures, how much time do you devote to this class?  Q11: What year are you in your program?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCDS-2 MANY numeric questions: Step 1: The user selects more than two numeric questions. Step 2: The system displays a correlation matrix.</td>
<td>Survey Name: CDS115 Midterm Survey (Fall 2016)</td>
<td></td>
</tr>
<tr>
<td>Question(s): Q6: Outside of lectures, how much time do you devote to this class?  Q11: What year are you in your program?  Q12: What term are you enrolling to get in this class?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.1: Use case scenarios and visualizations (implemented)
Figure 4.2: Use case scenarios and visualizations (not implemented)
4.3 User Interface

The interface design was an iterative and evolutionary process. Before implementing the first version, the following assumptions were made:

1. Users would like to view charts from different surveys at the same time.
2. Users would like to create separate panels; each panel has either a set of small multiples (bar chart, histogram and word cloud) for all the questions or one complex chart (heat map, scatterplot, stacked bar chart or correlation matrix).
3. Users would like to designate the time range and the location that they are interested in.
4. Users would like to select questions one by one.
5. Users would like to remove uninteresting categorical responses from the charts.

A screenshot of the first version is shown in Figure 4.3. The upper area of each panel contained selectors of survey, time range and location. If responses to a question had categorical responses, the corresponding response tags were placed under the question selector. Users could add a panel by clicking “Add a New Chart” button on the left-top corner. Clicking the rear wheel button on the left side of each panel header could hide or show all the selectors of the panel and clicking the “X” button on the right side of header could remove the panel. This design was abandoned because
of the following. First and the most important, all the existing widgets already took a lot of space and became cluttered for users to read their charts. Second, time range and location do not have much variance for a single survey, so they were not necessary for this tool either. Therefore in this phase, the following assumptions were dropped from the list: Assumption 3 “users would like to designate the time range and the location that they are interested in” and Assumption 4 “users would like to select questions one by one”.

A screenshot of the second version is shown in Figure 4.4. In order to save space, selectors of time range and location were removed, and a multi-selector was used to choose interesting questions. Two buttons “+ All” and “- All” were added to facilitate selecting and removing all questions at once. These made the layout of panels more compact. Next visualizations were implemented based on the following assumptions:

6. Before creating complex charts, users would like to have an overview; in the overview, one chart is created for every single question.

7. Users would like to see a visualization (word cloud was used here) of an open-response question as well as its full text of responses.

Based on these assumptions, by default the first survey and all its questions were selected, and a series of small multiples were displayed as the overview. Three types of charts were then used as visualizations displayed in small multiples: word cloud for open-response questions, histogram for numeric-response questions and bar chart for all the other questions. By clicking the “more” button in a word cloud chart, the user could create an independent panel showing full responses to a certain question. After testing this version on the sample datasets, it was discovered that within the word clouds generated from open-response questions, it was difficult to get useful insights such as key words showing respondents’ attitude; instead the frequently mentioned words tended to be less meaningful words such as prepositions and words that are general to a certain topic (e.g., “class” and “lecture” for a course survey). Time limits of this project prevented the integration of techniques that could extract meaningful keywords, such as modification relationships based extraction by Uchida [1] and sentence clusters based extraction by Gamon [22]. Therefore, the use of word clouds was discarded. Since the open-response questions should not be assigned empty small multiples, the full responses should be filled into the small multiples as the temporary exhibition of data. Because of the above reasons, Assumption 7 “users would like to
see a visualization of an open-response question as well as its full text of responses was temporarily unable to be realized. Nevertheless, better ways to summarize the sentiment of text responses will be investigated in the future work.

A screenshot of the third version is shown in Figure 4.5. It had a similar design to the second version except that the small multiples of open-response questions were changed into text boxes with full responses (see the small multiples of Q3 and Q7 in Figure 4.5); the independent full-response panels in the second version (see the two independent panels of Q3 and Q7 in Figure 4.4) were removed. Each panel contained either small multiples or one complex chart (e.g., heat map and stacked bar chart), as per Assumption 2 “users would like to create separate panels; each panel has either a set of small multiples for all the questions or one complex chart”. This means when creating multiple charts from the same survey, these charts would be in separate panels and operations such as survey selecting and settings would need to be performed redundantly. Moreover, this would also lead to a waste of screen space. Meanwhile, it was noticed that when creating a complex chart, there was a need to have quick access to the overview (small multiples) but this design was not very helpful in doing this. The “+All” and “-All” buttons were not utilized as often as expected either; so they were also removed from the interface. During the implementation, it was also found that an interface design in which different surveys were simultaneously displayed was rarely needed. Such a design could be useful.
Figure 4.5: Interface screenshot of the 3rd version

when multiple related surveys are conducted as part of a larger survey project. For example, a global company might conduct different surveys in its different branch sites and analyze all the survey data at the same time. However, this use case is rare. In most cases, the user only needs to analyze one single survey, thus it is not necessary to have multiple panels in the same webpage. Therefore, the following assumptions were dropped from the list: Assumption 1 “users would like to view charts from different surveys at the same time” and Assumption 2 “users would like to create separate panels; each panel has either a set of small multiples for all the questions or one complex chart”, and the interface was re-designed to view only one single survey with all charts placed in the same parent container.

In the last version, the interface contains two switchable tabs: overview and query. After opening this interface, the user will first see the overview tab (Figure 4.6). On the top of the overview tab are two selector widgets shown in Figure 4.7: survey selector (single select) and question selector (multiple select). By default the first survey and all its questions are selected. If the user changes to a different survey, questions will also be updated in the question selector. Below these two selectors are a series of small multiples, each of which shows one chart based on one selected question. By default small multiples of open-response questions are displayed in double width of the others because text responses usually take more space. Each small multiple has a header with question ID and wording, as well as a close “X” button used to remove the small multiple from the view. After a small multiple is removed, its question
will be removed from the question list too. The user can also add or remove a small multiple by changing selection in the question selector.

The user can enter the query tab by clicking its tag. On the top of the query tab is a header with survey name, same as in the survey selector in the overview tab. The query tab starts with an empty chart (with the wording “No question selected”), which has an identical question selector as in the overview tab with all questions unselected (Figure 4.8). After a new selection of questions is made, if it matches any of the cases listed in Figure 4.1, the corresponding chart will be created. A new empty chart can be added by clicking a button, the “+” icon in the right upper corner, which shows only when the query tab is visible. An example with four different types of charts is shown in Figure 4.9.

Back to the 7 assumptions during the whole design process, Assumption 1,3 and 4 were discarded because these use cases were not as common as previously assumed. Assumption 2 was discarded because too much space was taken. Assumption 7 was not implemented because there was not an appropriate visualization for open-response questions. Eventually Assumption 6 was retained and implemented, and a different feature was implemented to replace Assumption 5, which is “response reordering” to be described in the next section.
4.4 Interaction Features

According to studies of Yi et al. [20], interaction is an important part of information visualization and it provides a way to overcome the limits of representation and enhance users’ cognition. Therefore, a series of features were implemented to enhance the interactivity between the tool and users.

Chart reordering. In the overview tab, the user is able to reorder small multiples by dragging their headings. In query tab, the user can also use this function by dragging anywhere outside the chart area and moving around.

Response reordering. It is typical to see a question with responses in Likert scale such as “Satisfied”, “Neutral” and “Unsatisfied”. The software tool does not yet have the ability to tell the negativity or positivity of each response from their text contents.

In the overview tab, response reordering is available in all bar charts and is invoked by dragging any bar or text in the chart. In the query tab, response reordering is available in bar charts (working in the same way as in overview tab), heat maps and stacked bar charts: in heat maps, it is invoked by dragging text for any response and moving it either horizontally (in the first row) or vertically (in the first column); in stacked bar charts, it is invoked by dragging any text in the legend area, i.e., changing the association between colours and responses. Another important function is that when
any response reordering happens in a chart, if there are any other charts having the identical responses as the manipulated one, the reordering effect will be synchronized to these charts. The feature of synchronization is helpful to avoid repetition of the same reordering actions, especially when there are many questions within a survey. A set of examples of response reordering and synchronizing are shown in Figure 4.10.

Resizing. Panels in both the overview tab and the query tab can be resized by dragging their resizing handles in the right-bottom corner. When any panel is being resized, the visualization within will also be resized to fit the new size of the panel. Furthermore, vertical sizes of panels are restricted to be a multiple of 50 pixels to help align the panels more easily.

Brushing and linking is a technique where elements are selected and highlighted in one display (brushing), concurrently the same information in any other linked display is also highlighted (linking) [7]. In this tool, brushing and linking is invoked
(a) Dragging "Satisfied" to the next lower position in a bar chart
(b) Dragging "Satisfied" to the next right hand side position in a heat map
(c) Dragging "Satisfied" to the next lower position in a heat map
(d) Dragging "Satisfied" to the next lower position in a stacked bar chart
(e) By performing any one of the above actions, order of responses can be synchronized in all charts

Figure 4.10: Response reordering and synchronizing

when the user clicks the following: a bar in any bar chart (Figure 4.11a), a bar in any histogram (Figure 4.11b) or a bar in any stacked bar chart (Figure 4.11c), or the user selects a rectangle area in a scatter plot (either in a standalone scatter plot chart or embedded in a correlation matrix chart, e.g. Figure 4.11d). Any of the above operations specifies a categorical response or value range(s) (both denoted as “response” in the following text) as well as the selected question(s). Based on the selected question(s) and response, the tool will highlight the respondents providing the selected response in the selected question(s) in all the other charts.

The way to highlight selected respondents varies depending on the type of chart. In bar charts (except those for ranking questions) and histograms, a blue bar is
(a) Clicking the first bar in bar chart

(b) Clicking first bar in histogram

(c) Clicking the first bar of the first question "Q4a" in stacked bar chart

(d) Selecting an rectangle area in scatterplot

Figure 4.11: Examples of actions to invoke brushing
created within each original bar (in grey) and indicates the number of respondents by its length (in a bar chart) or height (in a histogram); in stacked bar charts, a thick, blue line is created above each existing bar and indicates the number of respondents by its length; in scatter plots, the point for each matched respondent is highlighted in blue colour, in text response panels, the responses given by matched respondents are highlighted in blue colour while the others are turned into grey colour. Nevertheless, in bar charts for ranking questions, although their bars can be “brushed” and invoke linking in other charts, currently they cannot be “linked” by another chart. This is because their bars indicate the average rank of responses rather than the number of respondents; the blue bars could be longer than the grey bars. Brushing also cannot be invoked in heat maps, correlation matrices (the matrix part) or text response panels and does not take effect on heat maps or correlation matrices either.

Furthermore, to facilitate the user to locate the chart and element where brushing is applied, in the “brushed” chart, the highlighted bar or points within the selected rectangle area are stroked in cyan colour, while the highlighted elements in all the other charts are stroked in black colour. The brushing and linking effect can be cleared by clicking the highlighted bars or in any blank area in the charts. If it is not cleared before changing the survey dataset, the brushing / linking settings will be retained and take effect the next time the same survey is selected. A sample screenshot of brushing / linking effect is shown in Figure 4.12.

Chart switching. In a panel within the query tab, when there is more than one chart type that matches the current selection of questions (e.g., two questions with the same categorical responses can be presented by either a heat map or a stacked bar chart), the tool gives a button beside the question selector. Clicking this button switches the panel to another available chart. This button will not show in the tool if only one chart type is available.

Settings. The user is able to hide all selectors to get a succinct view of the dashboard by clicking a button, the rear wheel icon on the left top corner. Clicking this button a second time will make all selectors show again so that the user can make further modifications.

Tooltip. Nearly all visualization elements have their own tooltips, which show upon mouse hover. Table 4.1 shows different visualization elements and contents of their tooltips.
Figure 4.12: A sample of brushing effect by select “Master Card” in “Q1”. Note that the small multiple of text response is also brushed while the bar chart of a ranking question is not brushed.
<table>
<thead>
<tr>
<th><strong>Type of element</strong></th>
<th><strong>Content of tooltip</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Header of small multiple</td>
<td>Full text of question</td>
</tr>
<tr>
<td>Tag of response / question (in charts)</td>
<td>Full text of response / question</td>
</tr>
<tr>
<td>Bar in a bar chart (not for a ranking question) or stacked bar chart</td>
<td>Response and number of respondents</td>
</tr>
<tr>
<td>Bar in a bar chart (for ranking question)</td>
<td>Response and average rank</td>
</tr>
<tr>
<td>Bar in a histogram</td>
<td>Value range and number of respondents</td>
</tr>
<tr>
<td>Cell in a heat map</td>
<td>Responses in two selected questions and number of respondents</td>
</tr>
<tr>
<td>Point in a scatter plot</td>
<td>Values in two selected questions</td>
</tr>
<tr>
<td>Cell in a correlation matrix</td>
<td>Correlation value</td>
</tr>
<tr>
<td>Brushed bar in a bar chart or histogram or stacked bar chart</td>
<td>Related question &amp; response and number of respondents</td>
</tr>
<tr>
<td>Brushed text response</td>
<td>Related question and response</td>
</tr>
<tr>
<td>Brushed point in a scatter plot</td>
<td>Related question and response</td>
</tr>
</tbody>
</table>

Table 4.1: Content of tooltips upon hover on visualization elements
Chapter 5

Preliminary Feedback

To acquire the preliminary feedback, a pilot study was conducted with one participant to assess the tool design. The participant was a university PhD student who did not participate the design process of this tool. She was analyzing (for her own research interests) a set of survey data related to weekly physical activity, collected using LimeSurvey, an online survey service of the university. The survey data consisted of two sources of information: the first source was the survey respondents reporting their weekly physical activity and the second source was the amount of physical activity reported by the Fitbit website[33]. Most of the questions in the survey corresponded to the physical activities of the respondents (activity at work, activity in leisure time etc.), and the last question was the physical activity reported by Fitbit. The data was collected over a time period of 8 weeks, and each respondent had 8 entries in the dataset; this is different from a typical “one entry per respondent” case. Furthermore, the respondents in this survey dataset were divided into two groups (intervention and control).

Tableau was the primary tool chosen by the participant to analyze the data, since LimeSurvey did not provide any means for visual analysis of the data. Therefore, she was asked to compare the user experience of using Tableau with that of the new tool. As for the user experience with Tableau, she mentioned that she had to manually pre-process the data into a format that was suitable for Tableau. The participant also listed the tasks that she would like to perform on her survey data:

- Compare between different user groups.
- Search correlations between different measurements (of physical activity).
See the trend of the data change based on the timelines.

See how user data and Fitbit data match each other.

By using Tableau, the participant was already able to get a comparison between groups and see a mismatch between user data and Fitbit data. Her next main interest was to find out if there was a correlation between the data submitted by the participants and by Fitbit. Moreover, she mentioned that she wanted a higher level of summary, abstraction and aggregation. She also wanted to see individual differences within the data. Overall she was happy with the visualization of Tableau and she rated the level of satisfaction as "Neutral 3".

The participant was then asked to open the new tool and perform tasks she would normally do with a survey. She noticed that she could not aggregate all the entries of the same respondent together and could not perform participant selection either: “I want to see how many people not just how many responses”, “I was trying to see individual participants and I wanted to select a set of participants”. She also mentioned a couple of times about the need to see the trend over time: “I cannot see the trend over time in the 8 weeks.” “I want to see per each user how did the measure of physical activity change over time”, “the bars are histograms, and they are not the real purpose of the data. I would prefer to see multiples timelines”. The participant then made several attempts to compare two groups and found this could be partially realized by looking at or brushing bar charts: “these two groups in terms of activity level are pretty close”.

In terms her experience with the new tool, the participant found histograms helpful in showing distribution of the data; she found the scatterplots useful in discovering outliers; she found correlation matrices useful to locate highly correlated measures. The participant said she had not thought about creating these three charts and she was happy to discover new insight from them: “this is the first time that I see a histogram of the data and I can see the Q20 and Q17 are almost normal distributions”, “(I can see some) correlation (between) Q4 and Q19 in the heat map with the scatterplot”, “Now I realize that (the activity frequency of) most people is occasionally in a week”.

The participant then summarized the tasks that she could not perform with the new tool:

• Do a comparison between two groups of users: “I want to see a correlation between groups”.

• View individual records by user name (because each user has multiple records).

• Get aggregation on a group of questions, e.g., sum of values of all “home” measures comparing with all “work” measures: “I want the aggregation part, (it) is more interesting than individual data”.

• View the trend over time, e.g., a timeline showing changes in the measures along time.

In terms of the feeling about not being able to choose the type of chart to plot, the participant felt this was less flexible than Tableau and still preferred to select chart type: “Less flexible than Tableau, I thought I could only select two questions in the query tab until you told me I can select more than two”, “I guess you can concentrate more on the data than on thinking on what type of chart is best, but of course in Tableau the charts are already recommended by the system”.

In summary, when the user felt limited by the new tool, it was basically because she expected functions that were available in Tableau but currently not available in the new tool. For example, she wanted to see individual records, trends over time etc., and she wanted to aggregate values of a group of questions. But based on the observation, the user was able to perform the survey tasks listed in Table 3.1 (e.g., to get an overview) on her data quickly and find some new insights. For example, she recognized the pattern of distribution in Q17 and Q20 in a very short time. It was also easy for her to explore her data. For example, she used the query tab and interaction features such as brushing/linking to create more complicated charts from an arbitrary combination of questions. Therefore, the user was interested in this new tool and kept trying all the features. She eventually rated her satisfaction with exploring data using the new tool as "Somewhat satisfied 4": “(it) is really cool, I like it.”
Chapter 6

Discussion

Through the pilot study, the preliminary conclusion was that the new tool was able to assist survey users with:

- Getting a quick overview of survey data.
- Creating complex charts more easily.
- Exploring survey data at a deep level to get potential new insight.

The pilot study also attempted to demonstrate another important advantage of the new tool: quickly generating a large number of charts. However, the participant did not give any feedback on this advantage. This was probably because the number of questions in her survey was not very large (20 questions). Therefore, user case studies with larger scale survey datasets are needed to validate this potential advantage. Nevertheless, it may be inferred that it is probably beneficial only for mid-size datasets, since there would be too many charts to easily look at if there are a massive number of questions.

In terms of the limitations of the tool design, the first limitation is that currently it does not support quantitative questions very well; more interaction features need to be implemented to allow users to customize the charts. Second, a suitable way to visualize open-response questions needs to be investigated in the future. Third, the tool does not provide enough guide hints; as a result, the user may miss some useful interaction features as observed in the pilot study; this can be remedied by adding in some widgets to display hints when necessary.

Users may still demand some control over the type of visualization. This point could be remedied by implementing more types of visualizations (e.g. parallel coordi-
nates and group bar chart) so that one certain data selection can have more optional charts. Then the user can explore and decide which chart is the most appropriate one.

As the pilot study had only one participant, it was not comprehensive enough. A larger number of participants would be beneficial but some difficulties were met. Although some people were interested in participating, few of them confirmed to attend the user case study. Some people did have some real data and were willing to share, but their data was aggregated, i.e., only overall statistics without the independent response from each single respondent. This was possibly due to the concern of data confidentiality or privacy. Nevertheless, to test the usefulness of the new tool, participants with real data should be the focus. However, there are other options to evaluate this tool, such as a usability study. An example would be the study done by Shamim et al [6].

Other challenges met in this project include:

- The discrepancy between the expected and actual effect of the designed interface. As mentioned in Chapter 3, during the whole designing process, evaluation and refinement were performed iteratively. At the beginning stage in particular, it was frequently found that the design outcome did not give a satisfying user experience as expected. That was the reason there were several intermediate versions. It was also for this reason that the word cloud design was discarded from the visualizations.

- Non-unified formats of raw data files. The collected survey data was mostly spreadsheets from different data sources. Each of them had its own format and none of these formats was in JSON format. Therefore, these files needed to be manually converted into JSON format one by one. To solve this, there should be a way to get all survey data in a unified format (for example, CSV format) so that they can be batch processed by a script.

- Time limit versus numerous possible visualizations and interaction features. In this project, only 6 visualizations were implemented, which were insufficient to meet the diverse needs of real users. For instance, in the pilot study, the participant needed a timeline visualization. Similarly, users of this tool may need more interaction features such as a response aggregation feature and a respondent selection feature. These are all interesting directions for future work.
Chapter 7

Conclusions and Future Work

In this project, a dashboard-style and interactive web tool was designed for visualization of survey data. This tool targets lay-users of visualization and enables users to see charts by simply selecting the data they are interested in without further specification of parameters. Therefore, this tool is helpful to save time spent in generating charts, especially when the number of charts is very large. Furthermore, the overview/query tab design of this tool helps the user to get a quick overview as well as explore the data at a deep level. A preliminary pilot study was also conducted and the preliminary conclusions were as follows: this tool could help the participant to complete some survey tasks; the interactive features of this tool could enable exploration; the current tool design is limited to handle quantitative attributes; and some functions were expected by the participant but are presently unavailable in the current tool design, such as aggregating response values and visualizing data in timelines.

In terms of future work, first more visualizations need to be implemented, e.g., timeline based chart and parallel coordinates (Figure 4.2), to support more tasks. As mentioned in Section 4.3, better ways to extract keywords from text responses need to be integrated in the tool so that the visualization can better summarize the sentiment of text responses. New features such as a response aggregation feature need to be implemented to enhance the flexibility of data manipulation. Finally, a more comprehensive user study should be conducted to more thoroughly evaluate the usefulness and effectiveness of the design.
# Appendix A

## Definitions of Terms

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>The unit against which respondents give their responses; each question has a unique question ID.</td>
</tr>
<tr>
<td>Respondent</td>
<td>The person who makes responses to questions; each respondent has a unique respondent ID that corresponds to an independent record.</td>
</tr>
<tr>
<td>Survey</td>
<td>Collection of questions and responses to the questions.</td>
</tr>
<tr>
<td>Value</td>
<td>The content of a response. A value can be a text string, a number, an object, etc.</td>
</tr>
<tr>
<td>Setting</td>
<td>A selected survey and a selection of its questions.</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma-separated values (CSV), is a file format to store tabular data (numbers and text) in plain text. Each line of the file is a data record. Each record consists of one or more fields, separated by commas. The use of the comma as a field separator is the source of the name for this file format.</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation, is an open standard format that uses human-readable text to transmit data objects consisting of attribute-value pairs.</td>
</tr>
<tr>
<td>DOM</td>
<td>Document Object Model (DOM) is a cross-platform and language-independent convention for representing and interacting with objects in HTML, XHTML, and XML documents.</td>
</tr>
</tbody>
</table>
Appendix B

Design Architecture

The top level of the user interface consists of: the overview tab, the query tab, the settings button and the add-new-chart button. The user can switch between the overview tab and the query tab. The add-new-chart button only shows when the query tab is showing.

B.1 The Overview Tab

The overview tab consists of a survey selector, a question selector and a series of small multiples. Each small multiple is a chart panel and is exclusively associated with a question in the question selector. A small multiple consists of a scrollable header of the question wording, a close “X” button, a chart whose type depends on the type of question.

The relationships between the question types and the chart types are shown in Table B.1.

Clicking the settings button in the presence of the overview tab replaces the survey and question selectors with a header with the survey wording.

<table>
<thead>
<tr>
<th>Question type</th>
<th>Chart type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple choice</td>
<td>Bar chart</td>
</tr>
<tr>
<td>Numeric-response question</td>
<td>Histogram</td>
</tr>
<tr>
<td>Ranking question</td>
<td>Bar chart</td>
</tr>
<tr>
<td>Text-response question</td>
<td>Plain textarea</td>
</tr>
</tbody>
</table>

Table B.1: The relationships between the question types and chart types in the overview tab
<table>
<thead>
<tr>
<th>Question type</th>
<th>Chart type</th>
</tr>
</thead>
<tbody>
<tr>
<td>One multiple choice question</td>
<td>Bar chart</td>
</tr>
<tr>
<td>One numeric-response question</td>
<td>Histogram</td>
</tr>
<tr>
<td>One ranking question</td>
<td>Bar chart</td>
</tr>
<tr>
<td>One text-response question</td>
<td>Plain textarea</td>
</tr>
<tr>
<td>Two multiple choice questions with different sets of</td>
<td>Heat map</td>
</tr>
<tr>
<td>responses</td>
<td></td>
</tr>
<tr>
<td>Two multiple choice questions with the same set of</td>
<td>Heat map or stacked bar chart</td>
</tr>
<tr>
<td>responses</td>
<td>(switchable)</td>
</tr>
<tr>
<td>Three or more multiple choice questions with the same</td>
<td>Stacked bar chart</td>
</tr>
<tr>
<td>set of responses</td>
<td></td>
</tr>
<tr>
<td>Two numeric-response questions</td>
<td>Scatterplot</td>
</tr>
<tr>
<td>Three or more numeric-response questions</td>
<td>Correlation matrix and scatterplot</td>
</tr>
</tbody>
</table>

Table B.2: The relationships between the question types and chart types in the query tab

B.2 The Query Tab

The overview tab consists of a header of survey wording and a series of chart panels. Each chart panel consists of a question selector, a close “X” button, a chart switching “>” button and a chart; it can be non-exclusively associated with one or more questions. The relationships between the selected questions’ types and the chart types are shown in Table B.2. The chart switching button only shows when there are switchable charts.

Clicking the setting button in presence of the query tab replaces all the question selectors with headers of the corresponding question wordings. Clicking the add-new-chart button in presence of the query tab adds a new and empty chart panel into the query tab.
Bibliography


[27] “iCharts” http://icharts.net Access Date: 2015-11-07


[34] “jQuery” http://jquery.com Access Date: 2016-01-24