Individual Report

Ye Cai
The University of Sydney
Sydney, Australia
ycai6198@uni.sydney.edu.au

I. Non-Visulization Task

A. Data Understanding

Data understanding is the first step of data visualization to comprehend the background of the topic, the meaning of each attribute in the dataset, and the relationship between different sheets. This process provides me a basic understanding of the dataset our team would use to establish graphs and answer questions. I shared my understanding of the dataset and some confusions on Slack to discuss with team members (Refer to Appendix). The opinions I shared are displayed in the following aspects:

- Normalized Data

All the data except release date and names like model name and company name is normalized from 0 to 1 which means we cannot observe the meaning of the data. However, I think it brings convenience in data visualization like show more than one attribute in the line chart with the same value range of rows.

- Data Relationship

There are three sheets in the dataset and each sheet focus on different aspects of the mobile device market information. I found out that we can use model ID and company ID to create connections between sheets. In addition, I noticed that in the Model-Company sheet the methods released just from 2003 to 2008 were displayed which means we are unable to use all the data in the product data sheet to analyze the second question.

- Problems

There are two questions I observed after reading the dataset and shared with my team members. The first one is the Release Year. The data type of release year in the sheets is float but the type of year should be an integer. Secondly, I feel confused about the new ID and old ID in the Company ID sheet and I do not know which should be used to create a connection between the model-company sheet and company ID sheet.

B. Data Processing

Data processing can be divided into two parts: Release Year transform and data interpreter.

- Release Year Transform

After communicating with team members, we decided to transform the release year into an integer to ensure normal use in data visualization. I tried to use excel formulas to transform the data, but I failed as some incorrect years in the list. However, our team member Yifan successfully transform the data and shared the files with us.

- Data Interpreter

I use data interpreter function in Tableau to identify the structure of the data in the Excel file and turn it into the proper format for analysis like remove extra rows and remove null values in the initial data source.

C. Document Menagement

As different schedules of team members and Tableau does not allow collaboration work. I will share the Tableau packaged and workbook with a brief description of my visualization on Slack (Refer to Appendix). Team members could better understand my opinions and concepts expressed by charts. Through sharing documents on Slack, team members could get the content they need instantly, sharing solutions, and would increase the working efficiently. I am also responsible for the final report integration and content review to ensure the correct sequence and answers after team members completing their parts.

D. Deriving Visualization Application

As all team members know how to use Tableau to establish visualizations and I can solve almost all the problems about Tableau, so we chose Tableau as the visualization application to complete the visualization based on hypothesis. I provided several suggestions about how to improve the visualization and teach team members some complex operation on the Tableau.

E. Literature survey

To create a better visualization, I researched many examples about how to present multiple data in the dashboard and create a better interaction on the Tableau Gallery. Meanwhile, I also review the lectures about the utilize of visual variables like colors and graphics to ensure good human-computer interaction and avoid phenomenon like an optical illusion.

F. Team Communication

There are two ways to keep communication with team members: Slack and team meeting once a week. I always shared my work on Slack and welcome others to give me some feedback. On the other hand, I engaged in a weekly team meeting which is half an hour before the tutorial to discuss the solutions and ask the tutor some questions. Through these two approaches, each team member can keep connect and understanding of each other's work. Many questions and advice mentioned at the team meeting are very useful for me to provide

another aspect to consider the questions and improve the visualization of graphs.

II. Visualization Task

A. Attributes Division

Categories	Attributes
Performance of devices	RAM (memory) capacity (Megabyte)
	Storage capacity (Megabyte)
	CPU clock (MHz)
	Pixel Density (per inch)
Appearance of devices	Display size (dimensions) (diagonal in inch, width, and length in pixels)
	Volume (width-length-depth in mm)
	Mass (grams)

During many attempts of creating different charts to show the changes of all the attributes from 1989 to 2012, our team observed that there are too many attributes and colors in the same diagram which would directly increase the difficulty of analysis and grab important points. The third week of the color lecture and tutorial also mentioned that the wrong combination in the chart, such as too many colors, can cause visual confusion. I proposed to divide the attributes into two groups: performance group and appearance group. The performance group includes 4 attributes (RAM capacity, CPU clock, Storage capacity, and Pixel Density) and all the attributes are related to the performance of the devices. The other attributes which related to the size and mass of the device are classified as appearance group. Our team finally adopt this suggestion as we agree that it is a suitable and reasonable approach to divide attributes and achieve a better visualization. However, I also use Yifan's advice that only use display size, volume, and mass in the appearance group to avoid a similar situation at the beginning.

B. Visualization Method Selection

During visualization, all team members agreed that the dashboard is an excellent way to group the charts related to the same period together to better show the results.

The bar chart of the period division part I think is the most controversial part during team discussion as I think the median of a line chart is enough to analyze different periods, but Dian insisted that a maximum and minimum comparison is necessary. Finally, Niti established a cumulative bar chart that includes all these three items and was accepted by all team members as the knowledge mentioned in the Week 5 lecture, cumulative bar chart is one of the best ways to display the composition analysis.

In the comparison part, Scatter plots were mentioned by Yifan which is also be mentioned in the exploratory data analysis lecture and have advantages in distribution analysis. After reviewing the lecture, I understood and used it to show the performance of different attributes of devices in different years (Refer to Figure 1). Yifan also posed that we can try to aggress three attributes into one scatter plot by using color and size, but I denied his idea and using the knowledge learned from visualization design and representation lecture in week 7 to explain the disadvantages of using size to distinguish the attributes. However, we also decided to choose to use scatter plots in the comparison part.

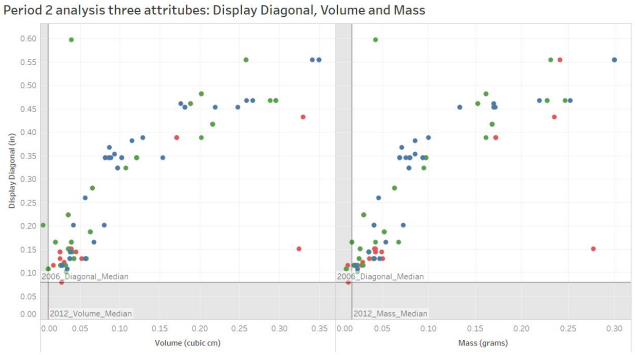

Figure 1-Scatter plots in the comparison part

In the selection part, I suggested that the table is not a good way to show the results as it is hard for customers to compare differences by long numbers such as 0.1504 and 0.1527. A bar chart may be a better choice which can both show the results of the selected devices and differences between other parameters like the median of the data in the year the device was released. The exploratory data analysis lecture in week 5 also referred to that the bar chart is suitable for attribute analysis. Consequently, our team chose the bar chart in the selection part.

C. Visualization

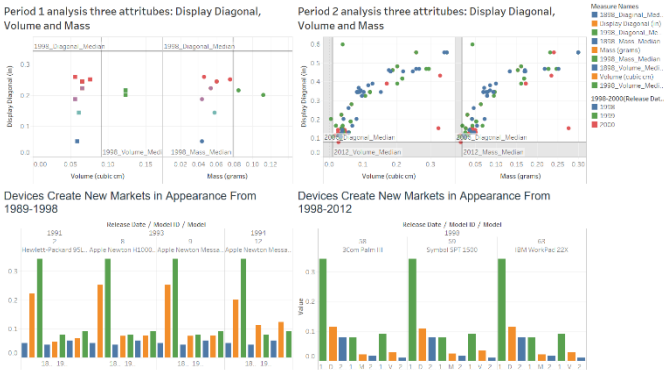

Figure 2-Dashboard: the appearance of devices

Visualization is my main job in this assignment. I fully used the knowledge learned in the unit like color utilize in the week 3 lecture, visualization design and representation in week 7, and evaluation in week 9. Charts have been finished after making a decision and grouped into dashboards (Refer to Figure 2). After I finished, I uploaded the file with a brief description of Slack, and I also explained some aspects that need attention to the team members in meetings and tutorials. My team members would give me some suggestions and shared their opinions about a better way to visualization these data. For example, after I completed the first version of the virtualization, Zeta and Niti suggested that it is unsuitable to choose all the years during the period (e.g., 2008-2012) to compare as some devices released in the later years like 2011 may learn from the previous experience of devices and we cannot consider them as the devices which tried to create a new market. This is an important suggestion that

provided a huge help in solving the questions. I modified diagrams and description documents after achieving their feedbacks and uploaded them again on Slack to ensure everyone can understand each chart clearly (Refer to Appendix).

D. User Interaction and Visual Variables

User interaction and visual variables are the most important aspects that should be considered during visualization and I discussed this part with Yifan to ensure make a suitable choice in several factors like color and size selection. We used the knowledge learned from the Fundamentals of HCI in week 4 to improve the experience of customers. For example, in the dashboards, customers could use the cards to highlight the devices from different aspects like released years (Refer to Figure 3).

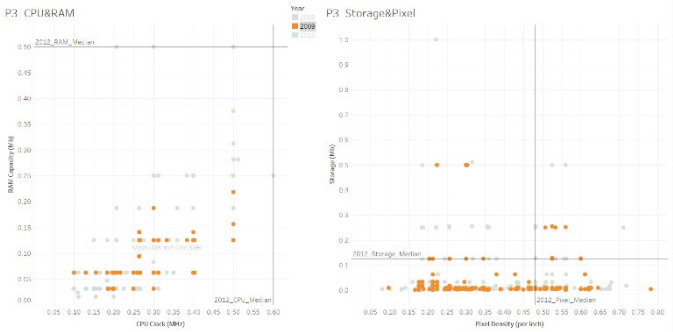

Figure 3-Dashboard: The comparison of devices in performance

III. APPENDIX