THE UNIVERSITY OF SYDNEY

Assignment Coversheet – GROUP ASSIGNMENT

Please fill in your details below. Use one form for each group assignment.

Assignment Details:

Assignment Title		How did the mobile device market change?			
Assignment number		Assignment 2			
Unit of Study Tutor		Shubhi Goel			
Group or Tutorial ID		RE04-G6			
Due Date	11/09/2022	Submission Date	11/9/2022	Word Count	4177

Membership

Group Name/Number	G6				
Family Name	Given Name (s)	Student Number (SID)	Unikey	Contribution + percentage	Signature
Pi	Ruochen	500055496	rupi2638	Project manager, part of the report, 20%	Ruochen Pi
Hao	Zhang	500388349	hzha6248	Analyst, part of the report,20%	Hao Zhang
Zhang	Jiaqi	510637510	jzha9054	Analyst, part of the report, 20%	Jiaqi Zhang
Hu	Jianhua	520256961	jihu9398	Analyst, part of the report, 20%	Jianhua Hu
Dong	Bokang	510626743	bdon4410	Programmer, part of the report,20%	Dong Bokang

Declaration:

1. I understand that all forms of plagiarism and unauthorised collusion are regarded as academic dishonesty by the university, resulting in penalties including failure of the unit of study and possible disciplinary action.
2. I have completed the **Academic Honesty Education Module** on Canvas.
3. I understand that failure to comply with the Academic Dishonesty and Plagiarism in Coursework Policy can lead to the University commencing proceedings against me for potential student misconduct under Chapter 8 of the ***University of Sydney By-Law 1999*** (as amended).
4. This work is substantially my own, and to the extent that any part of this work is not my own I have indicated that it is not my own by acknowledging the source of that part or those parts of the work.
5. The assessment has not been submitted previously for assessment in this or any other unit, or another institution.
6. I acknowledge that the assessor of this assignment may, for the purpose of assessing this assignment may:
   a. Reproduce this assignment and provide a copy to another member of the school; and/or
   b. Use similarity detection software (which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking).
7. I have retained a duplicate copy of the assignment.

Please type in your group number here to acknowledge this declaration:	Group 06

How did the mobile device market change?

Ruochen Pi
University of Sydney
Shanghai, China
rupi2638@uni.sydney.edu.au

Jiaqi Zhang
University of Sydney
Wuhan, China
jzha9054@uni.sydney.edu.au

Bokang Dong
University of Sydney
Nanjing, China
bdon4410@uni.sydney.edu.au

Hao Zhang
University of Sydney
Guangdong, China
hzha6248@uni.sydney.edu.au

Jianhua Hu
University of Sydney
Zhejiang, China
jihu9398@uni.sydney.edu.au

***Abstract*—As technology evolves day by day and mobile devices are updated generation after generation, there is no one company that is always strong and no one product that is always hot. Some groundbreaking new mobile devices are introduced to the market, and some of them come from big companies and some from small ones. This report analyzes how the mobile device market has changed over the past few decades.**

Keywords—Mobile device, Visualization, Tableau, Predictions, Trends, Performance, CPU

I. INTRODUCTION

In the past decades, many mobile devices have been developed. According to the continuous upgrading of technology, some new mobile device types have replaced the previously popular mobile device types. By analyzing the data of mobile devices in the past decades, the team obtains interactive and visualizations to classify the data of these mobile devices, analyze the changing trend and predict future data.

II. DATA PROCESS

This section is the pre-study data processing section, including the data source introduction and pre-processing process.

A. Data Source

The original dataset file has three tables, of which the first two tables, "Normalized Product Data" and "Model-Company", need to be concerned. The first table, "Normalized Product Data", contains 3163 rows of data and has 16 columns of features. These 16 columns of features contain some information on mobile devices, such as RAM, capacity, CPU clock, and Display size. The second table "Model-Company" contains some blank rows. Before removing these blank rows, the table contains 5807 rows of data and 5 columns of features. These 5 columns of features contain the company ID, release date, year, name, and ID of the mobile device corresponding to the release of these mobile devices.

B. Pre-processing

The blank lines in the original dataset need to be processed first. After using python to remove the blank rows in the second table, the table has 2904 rows of data. Then use the corresponding feature to connect the data of the two tables for subsequent visualization. The data in the first table "Normalized Product Data" has been standardized, so it is no longer standardized.

III. VISUALIZATION - CLASSIFICATION

This section group gives this dataset a post-analysis classification.

A. Hardware performance

After viewing the columns (features) of mobile devices in the first table "Normalized Product Data", the features reflect the information of two mobile devices, one is about the performance of the mobile device, and the other is about the external information of the mobile device. The team can visualize the two kinds of information and analyze which mobile devices have similar characteristics, and then classify the mobile devices according to the visualization figures. The team combined and visualized different features belonging to the same kind of information to find common features to define types of mobile devices. The two combinations with obvious distribution rules are "CPU Clock-RAM Capacity" and "Display Diagonal-Pixel Density". In general, the features in the above two combinations belong to the same information, and the distribution rule can be found after visualization to help define mobile device types. Therefore, the team finally chose to use the above two combinations to classify mobile devices differently.

For the first combination "CPU Clock-RAM Capacity", take "CPU Clock" as the x-axis and "RAM Capacity" as the y-axis, and then use the size of the average sum of the two data (hereinafter referred to as performance) to represent the color depth on the figure to get figure 3.1.1 below.

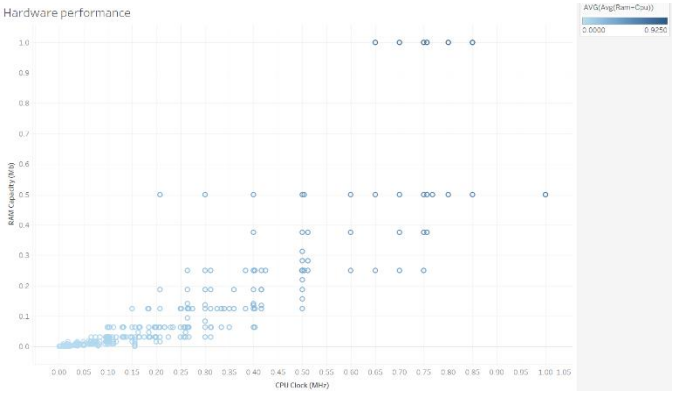

Figure 3.1.1 Hardware Performance

After using these two features as the x and y axes, there is obvious stratification. Then use the color depth in the figure to represent the size of the performance value to help classify mobile devices, because it is difficult to distinguish the dividing line among several categories if the size is used to represent. According to the visualisation, python is used to classify the data according to the range with performance less than 0.165 as level C, performance greater than or equal to 0.165 and less than 0.48 as level B, and the rest as level A. The following figure 3.1.2 is obtained by visualizing the classified data.

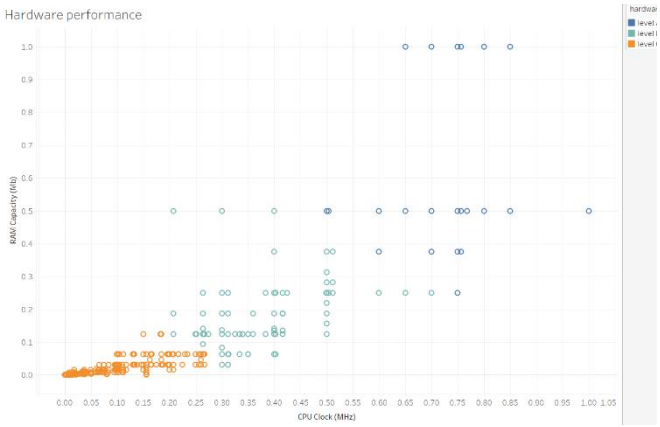

Figure 3.1.2 Hardware Performance (after classification)

B. Display Performance

For the other combination "Display Diagonal-Pixel Density", the team used a similar method to visualize and draw figure 3.2.1. From the distribution of points in figure 3.2.1, it can be roughly seen that when only one of the features of Display Diagonal and Pixel Density is considered, the points can be divided into two parts. Therefore, the team uses this combination to divide mobile devices into four categories: high resolution-large screen, high resolution-small screen, low resolution-large screen, and low resolution-small screen.

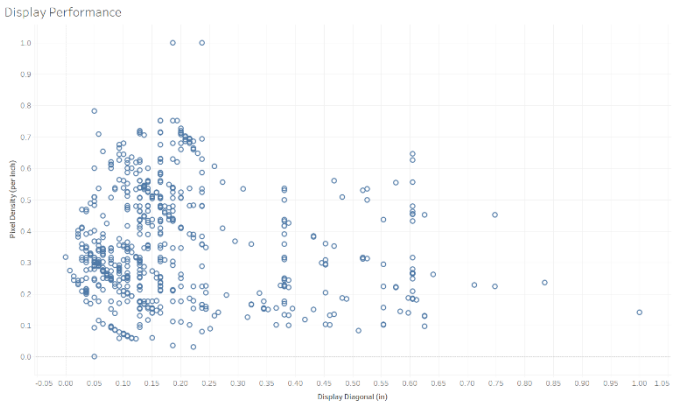

Figure 3.2.1 Display Performance

C. Evaluation and modification

After the mobile devices are divided into four categories by analyzing the distribution rule of points in the visualization obtained from the "Display Diagonal-Pixel Density", the data is divided and classified using python. When the Pixel Density is less than 0.41 and the Display Diagonal is less than 0.3, it is divided into the low resolution-small screen; when the Pixel Density is greater than or equal to 0.41 and the Display Diagonal is less than 0.3, it is divided into the high resolution-small screen; when the Pixel Density is less than 0.41 and the Display Diagonal is greater than or equal to 0.3, it is divided into the low resolution-large screen; and the rest is divided into the high resolution-large screen. Similarly, the data after classification will be visualized (figure 3.2.2). Different colors will be used to represent different classes to more intuitively see which class a point belongs to. It is not intuitive to use size or different shapes to represent.

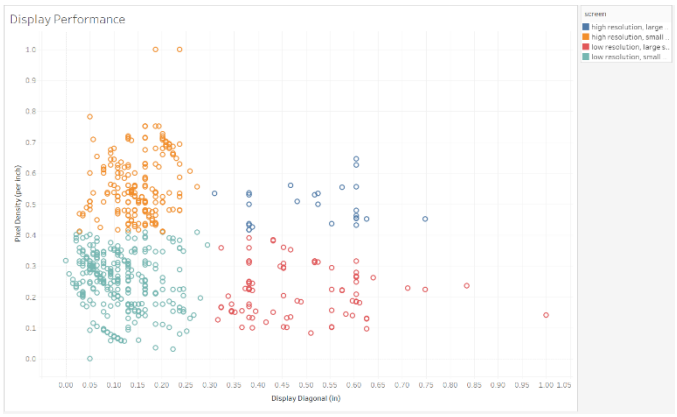

Figure 3.2.2 Display Performance (after classification)

The above four visualization related figures are the final results of the team after evaluation and modification, which is not the idea at the beginning of visualization. Before deciding to use these two combinations to get the visualization results, the team tried to combine multiple related features into one, and then used these new features to make a visualization in tableau to observe whether the mobile device was obviously distributed into multiple divisible parts, where the size and shape represented a new feature was also tried to use. However, when the team showed the visualizations to other students outside the team, they thought that they could not well distinguish the different types of mobile devices. Finally, the team decided to combine two features belonging to the same kind of information and then use colors to represent different categories to create a more understandable visualization.

IV. VISUALIZATION - TRENDS

Based on the classification results from the previous question, mobile devices are classified according to hardware performance and display. When analyzing trends, these two properties will be visualized and observed.

A. Hardware performance

Based on the observational visualization data, it is concluded that the number of releases of high-performance hardware devices increases exponentially, while the number of releases of low-performance devices is close to zero. On the other hand, medium-performance mobile devices have started to decline in recent years after reaching their peak. From this it can be concluded that regarding the trend of mobile devices, low

performance devices were the mainstream in the market until 2007, with the development of hardware in 2008-2009, medium level devices started to replace them. Low-performance hardware is starting to be phased out of the market. Looking at the data visualization, it turns out that in 2011-2012, the number of new device releases outnumbered older mid-performance devices which were on a downward trend. After 2009, the hardware of mobile devices began to update and iterate rapidly, and the hardware performance increased rapidly. High-performance mobile devices with higher CPU clock and larger RAM became popular, resulting in medium-performance devices being taken over by new which is in an exponential growth trend.

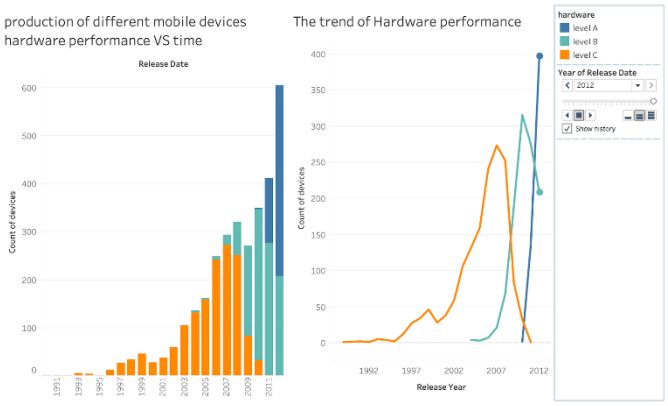

Figure 4.1 Hardware performance visualization

In order to represent trends in hardware performance, the visual interaction will use bar graphs and dot chart, as figure 4.1 shows below. The reason for using the histogram is that the height of the column reflects the difference in value and can clearly compare the data difference between different years. Helps visualize the comparison of numerical values. The interactive dot chart looks like a line chart, which in order to show the trend in the number of releases of different performance devices over time.

In this dot chart, when using Tableau to create the y-axis, the number of different types of hardware performance levels in each year, the user can clearly understand the number of production releases of different types of mobile devices in each year. Then construct the x-axis, which will represent the year of release, from 1989 to 2012. Use three colors to represent three different hardware performance levels, in ascending order from A to C. At this time, the user can see that the column in the dot graph is divided into three parts according to the year, and each dot can show how many mobile devices are released. To increase the visual interactivity, the Page Playback feature is used, which allows the user to jump to a specific page based on the year. At the same time, the marks of the previous page are also displayed, making the figure look like a line chart, as figure 4.1 shows below.

For a bar chart, the properties of the X and Y axes are the same as for the dot chart, which for consistency, the user will not be confused about the properties of the axes. On the other hand, the colors representing the hardware performance levels are also the same as the dot plot.

B. Display performance

The production of low resolution and small screen mobile devices has been increased sharply between 1989 and 2007, it rose to a high point and peaked in 2007. After 2007, there has been a steady decrease in the production of low resolution and small screen mobile devices.

Large screen mobile devices emerged at around 1997. However, there has only been a slight rise in the production of large screen with either high- or low-resolution over the next 10 years. After 2007, the production of large screen mobile finally goes up in a steady fashion.

The production of high resolution and small screen started at around 1998. It increased rapidly between 2007 and 2012. It became the most produced mobile devices each year since 2011.

Overall, the mobile device market from 1989 to 2007 is dominated by low resolution and small screen types of devices. It was taken over by high resolution and small screen devices after 2011. The production of low resolution and small screen devices is expected to continue decreasing and become obsolete.

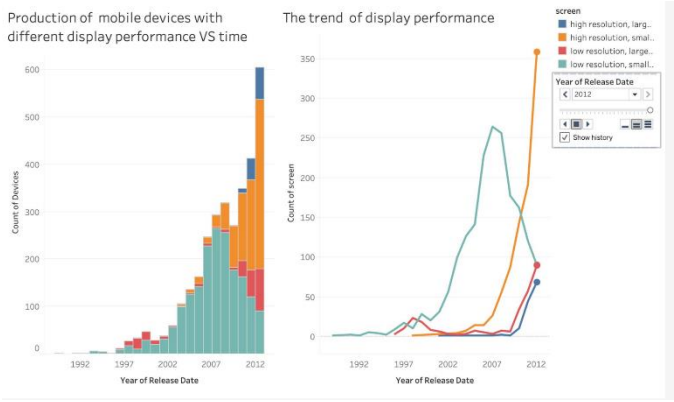

Figure 4.2 Display performance visualization

For the visualization of different screen types, we used two plots to visualize “production of different mobile devices VS time”. We decided on choosing “release date” as x-axis and “count of devices” as y-axis. Under this axes arrangement, our plot can reflect the long-run growth or decline in the production of mobile devices with respect to time. In addition, since line graph can represent how data has changed over time, it is used to observe trend components.

In terms of visual variables, we use “color” to differentiate multiple types of screens: low resolution with small screen, low resolution with large screen, high resolution with small screen and high resolution with large screen. We have also added legends in the top-right of the visualization. Color variable is essential for our visualization because it allows users to quickly identify different types of mobile devices and understand the underlying data correctly.

We used a time panel at the right side of the visualization. With this interactivity, users can observe how the production of different types of mobile devices change between 1989 and

2012. Besides, they can choose any specific period that they are interested in and see how the data is distributed during that period. Overall, this interactivity will improve the efficiency and effectiveness of our visualization.

In the second plot we want to support our findings of trend by adding a stacked bar plot that visualize "production of different mobile devices VS time". The choice of x-axis, y-axis, visual variables, and interactivity remain the same as the first plot, however, using a stacked bar plot allows user to compare composition and amount of mobile devices production in each year. Therefore, combining these two plots together can enhance the understanding of data and therefore improve the readability of our visualization.

C. *Evaluation and modification*

Self-assessment and peer-assessment will be used in these visulization. When implementing a visualization process, after completing the initial visualization, assess whether it answers the question which presents trends for both old and new devices. In the initial visualization, the team chose the bar chart to present the total number of different types of mobile devices released over time. However, after self-assessment, the histogram can be considered that the histogram does not clearly show the trend of these two types of mobile devices, but more prominent comparisons, comparisons of numerical values. Therefore, adding a line chart can effectively help visualize the fluctuation trend of hardware and display performance over time. After the self-assessment is completed, the interactive visualization will be peer-assessed. Teams pass it on to other teams and team members who are responsible for other parts of this assignment. Based on their feedback, it was concluded that the visualizations were weakly interactive which perception of data transitions is weak., and simply highlighting the same kind of data did not show the fluctuations of the trend. So, to render data transitions in the visualization, page animations were decided to be used. Using a dot graph plus a historical track, it looks like a moving "line graph" when the user clicks to play. It enhanced audience awareness of data trend transformation and increased interactivity. The modifications resulted in the final interactive visualisation and as figure 4.1 and figure 4.2 shows above.

V. VISUALIZATION – FORECASTING

A. *Hardware performance(CPU+RAM)*

i. Explanation of visualizations:

From the visualizations of Figure 4.1, we can see that the lower hardware performance was taken over by stronger devices. In particular, level C devices disappeared after 2010, and the number of level B devices started decreasing as long as level A devices occurred. From the visualization, the team believed that the number of level B devices will keep decreasing like what happened to level C devices and be taken over by level A devices. Also, the high increase rate of the number of level-A devices significantly shows that level-A devices are going to be famous devices in the future.

In general, with the development of CPU or RAM, high performance on CPU and RAM will be one of the parts of a newly emerging mobile type.

B. *Display Performance(Display Diagonal+Pixel Density)*

i. Explanation of visualizations:

Since the team analyzed the data from 1989 to 2012， which is Figure 5.2.1, we found a relationship between Display Performance and time. Through the display performance line chart, which is Figure 5.2.2, the team can find that we found through the line chart that low resolution and small screen type is gradually being eliminated. In the future, it is high relosution and large screen type and high resolution and small screen type will occupy the future market.

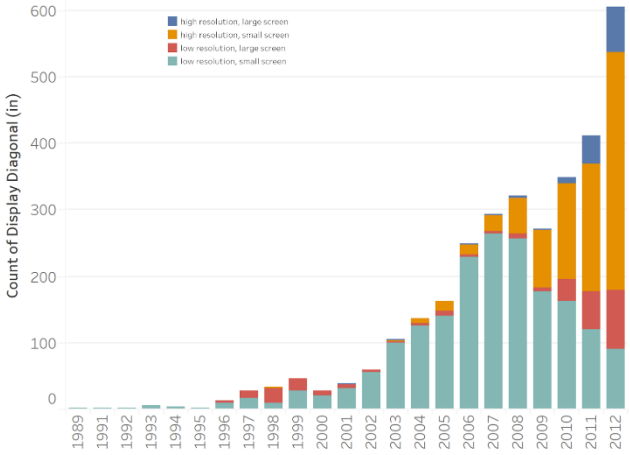

Figure 5.2.1 Display Performance Bar Chart

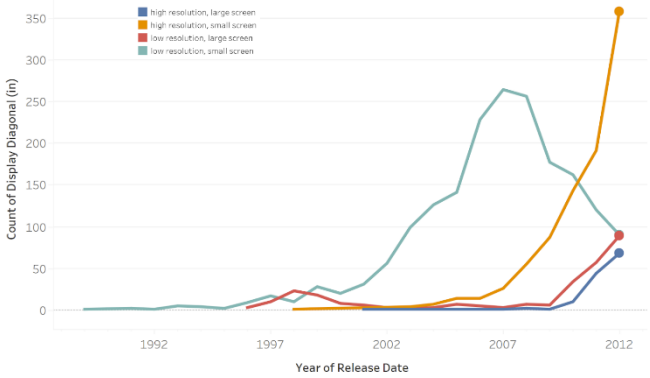

Figure 5.2.2 Display Performance Line Chart

The team took out the two properties, which are display diagonal and pixel density separately for further analysis, as shown in Figure 5.2.3. The team found that with the development of technology, the average display resolution of mobile device is on an upward trend, reaching nearly 0.48 per inch in 2022, while the average display size rises and then falls, and increases year by year after 2007. We can conclude that the pixel density of future mobile device will rise steadily and gradually, and display diagonal will rise slowly.

In general, the resolution in the future has a function of the nature of the positive correlation growth trend with time.

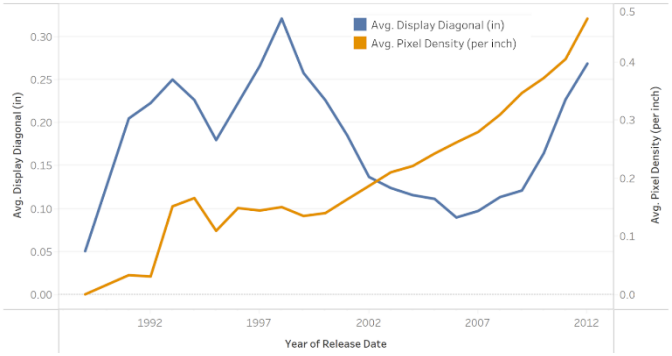

Figure 5.2.3 Display Diagonal and Pixel Density

For observing the visualization of display performance predictions, the team chose a combination of bar chart and line chart, and then created a line chart with separate analysis of display diagonal and pixel density. The conclusions were obtained by combining the three charts.

ii. Evaluation and modification:

For assessing the reliability of the method, the team assumed that only data between 1989 and 2010 could be accessed, simulated predictions for 2011 to 2013, and found that the same conclusions were obtained. In addition, the team was peer-reviewed by students in the field of visualization and received peer approval.

For some improvements and enhancements, the team was the first to use dimension data with the year on the x-axis and display diagonal and pixel density on the y-axis in the process of processing display diagonal and pixel density data, instead of processing it as the average of each year. Figure is as shown as 5.2.4.

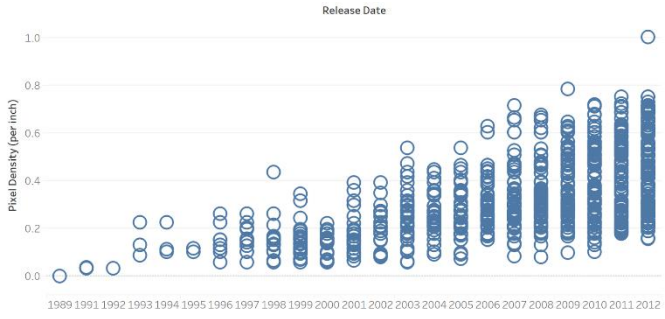

Figure 5.2.4 Display Performance Old version

iii. Justification on final visualization of display:

For Figure 5.2.1, the group uses a bar chart. release year is the x-axis and, after being processed by the group, is divided into four categories to get the counts of display performance as the y-axis. The four categories are distinguished by different colors, so that the percentage of each category in each year can be clearly understood through visualization. By moving the cursor over each category, each category can be highlighted. The interactive visualization allows for appropriate visual feedback to the user and highlights key content (Swaminathan, 2022).

For Figure 5.2.2, the team chose to use a line chart in order to show the changes in the four categories from year to year. unlike Figure 5.2.1, the team used four lines of different colors to represent the different categories, allowing the user to visibly observe the trends in each category.

For Figure 5.2.3, the group analyzed two attributes separately in order to be able to understand the two attributes that affect display performance. Their changes over time were observed using line graphs.

C. Storage

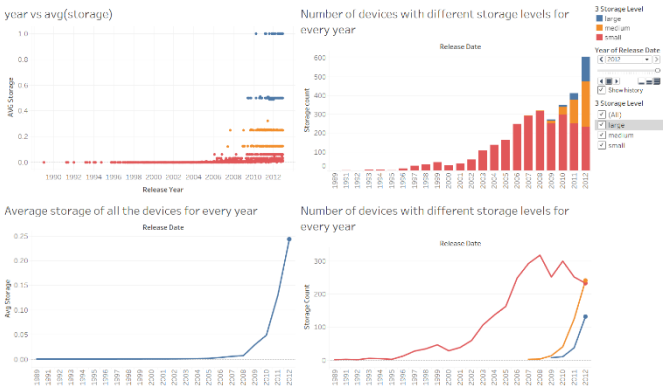

Figure 5.3.1 Dashboard for attribute “storage”

i. Explanation of visualizations:

The bottom left visualization is the average storage of all the devices for every year. The team used a line chart to show the trend of the change in storage, where the average storage had significantly increased starting in 2008. It shows that the demand for storage in the market is increasing and a larger storage device may become part of a newly emerging one.

The top left visualization is the average storage value of each device. Since the capacity is multiplied, most points in the visualization are grouped together where the highest storage group value is 1.0 and the second highest storage group value is around 0.5, the team decided to set values into three levels by groups. The “large” level represents the highest and the second-highest group, the “medium” represents the third-highest and fourth-highest group, and the rest of the storage values are signed as “small” levels.

The top right and bottom right visualizations are the number of devices with different storage levels for each year. The difference is that one is using a bar chart to show the proportion of each storage level and the other is using a line chart to show the trends of each storage level.

From the line chart, the team found that small storage devices dominate the market and keep increasing until medium storage devices occur. The decreasing rate of small storage devices is low, which means the total number of small storage devices did not decrease a lot. However, from the bar chart, the team found that the total number of devices increased very fast from 2009.

Therefore, although the number of small storage devices did not change a lot, the proportion of them in the market decreased a lot, which is lower than 40% in 2012.

On another side, the number of medium storage devices and the number of large devices are increasing very fast. From the line chart, the team found that both of them have a highly increased rate. This shows that the devices in the later years will probably have larger storage. However, since the large storage level devices just occurred, they may need more time for development, so a newly emerging device with a larger storage level, like double storage volume to the current highest in 2012, is more likely to occur and be the famous type in a few years after 2012.

In general, larger storage devices will be one of the newly emerging mobile device types in the future.

ii. Evaluation and modification:

For the four visualizations above, the team first make some draft visualizations using different visual variables. Then we compare and contrast those results and choose the visualization which we think is the best. Then we set up some questions for the peer-reviewed by students in other groups and collect their feedback. Finally, the team analyzed the feedback and changed the visualization into the final version.

For the top left visualization of Figure 5.3.1, we first think about separating the storage into 5 levels, where the highest storage devices group is one level, and the second-highest devices group is another level. It makes sense that in the visualization, the highest storage group is far away from the second-highest storage group. However, when we made the line chart of the average storage for each storage level, which is the bottom right visualization of Figure 5.3.1, the number of devices in the small storage level is much bigger than other levels. This made the result of the line chart terrible and it was hard to tell the trend of the other four storage levels. Therefore, as one of the feedback from the peer-reviewed suggested, we decrease the number of storage to three as shown in the final visualization. Finally, we got a clear result on medium and large devices.

For the prediction of storage, the team used two line charts, which are two bottom visualizations of Figure 5.3.1, to show the future trends of storage at the beginning. The team got the results that the number of small storage devices started to decrease, and larger storage devices are increasing fast. Then the team believed that the small storage devices are not going to be taken over by larger storage devices soon since the number of small storage devices did not decrease a lot. After the peer-reviewed, the team realized that it is necessary to give the proportion of each storage level since the increase in the total number of devices cannot be shown in these two visualizations. Therefore, the team made a new visualization, which is the top right visualization of Figure 5.3.1. The team found new results that the proportion of small storage devices was decreasing very fast, which is a sign of being taken over. Using the third visualization made our result more reliable.

iii. Justification on final visualization of storage:

The team was using the release date as the x-axis for the average storage of each model, which is the top-left visualization of Figure 5.3.1, because it is easy for separating the storage into different levels. Also, the team decided to use colors for different storage levels, instead of sizes or shapes since there are only have three levels and it is easy to identify the difference.

For other visualizations, the team decided to use the year as the x-axis since it is better for showing the trends, compare to the fluctuating line chart which uses the date as the x-axis. The line chart was used for showing the trends of different storage levels and the total storage change. The line chart gives clear trends and the animation can help user to see the change every year.

The bar chart was used for showing the proportion of each storage level. The filter action help user to see the proportion change of each storage level for every year.

VI. CONCLUSION

According to the team's visualization, the number of releases of high-performance hardware devices is growing exponentially, while the number of releases of low-performance devices is close to zero. Small screen with low resolution mobile devices dominate the market before 2007, and after 2011 it is replaced by small screen with high resolution mobile devices.

For the future market, the team maintains that high-performance mobile devices will replace low-performance ones. For display performance, small screen with small resolution will eliminate in the future, and large resolution will become more and more popular. At the same time, small capacity storage is not eliminated, but they gradually become smaller occupation, the next market is occupied by large storage mobile devices.

REFERENCES

A. Swaminathan, R. Surendran, J. Joel Benjamin and S. Thiyagarajan, "Analysis and Prognosis of Covid-19 using Machine Learning Algorithms and Visualization using Tableau," 2022 7th International Conference on Communication and Electronics Systems (ICCES), 2022, pp. 1096-1103, doi: 10.1109/ICCES54183.2022.9835941.