COMP5048: Visual Analytics Assignment 2

Abigail Gentle
agen2864@uni.sydney.edu.au

Fei Hui
fhui3997@uni.sydney.edu.au

Shreya Kathuriya
skat9998@uni.sydney.edu.au

Rongshang Li
roli5128@uni.sydney.edu.au

†: See our demonstration video

I. Introduction

The aim of this report is to answer the given questions by analysing mobile device market data using da. Specifically, we answered Q1 in section section IV, Q2 in subsection section V-A and Q3 in subsection section V-B.

II. The Data

In order to analyse trends in the mobile device market from 1989 to 2013 the dataset contains 10 variables, some of which are actually represented by multiple variables such as length, width, and height.

- Model name and ID,
- Release Date, Year, (ranging from 1989 to 2013),
- RAM (memory) capacity (Megabyte),
- Storage capacity (Megabyte),
- CPU clock (MHz),
- Display size, (dimensions) (diagonal in inch, width and length in pixels),
- Volume (width-length-depth in mm),
- Mass (grams),
- Pixel Density (per inch), and
- Company ID.

In our initial inspection of the data, we noticed it contains many different types of mobile devices, ranging from GPS navigation devices, devices without modern displays, "palm-tops" or subnotebooks, as well as smartphones, tablets, and some notebooks. The data has all been normalised, and this made some computations difficult.

a) Pre-Processing: For pre-processing, some visuals are made based on year instead of the given "Release Year". The "Release Year" variable contains the a decimal representation of the year a device was released. To visually analyse specifications of the devices released yearly, Year was extracted from "Release Date" variable. A visualization showing the trend of number of devices released over the years can be seen in figure 1. The dataset given contained three sheets: "Normalized Product Data", "Model Company"' and 'Company ID'. These datasets were linked with each other through linking the variables "Model ID" with "Model-Company" an then "Model-Company" with "Company ID", which allowed us to extract company specific information.

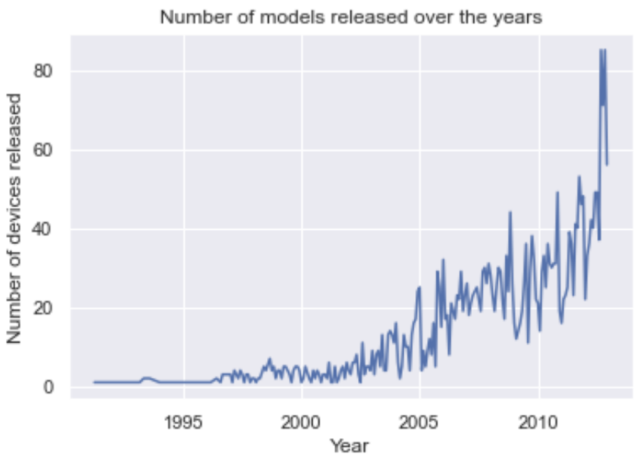

Fig. 1: Number of Releases over the Years

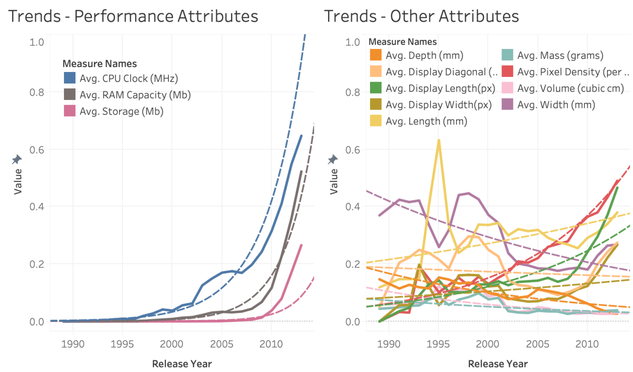

Fig. 2: Attribute Trends

III. Performance Trends

To understand how the attributes changed year over year, we first examined yearly trends (measured by average value) in all of the attributes in a line chart with exponential trend line added for each attribute. As shown in the figure below, trends in the attributes relating to the computational performance of the devices - CPU, Ram and Storage - show different changing behaviors to the other attributes. While the trend lines for the other attributes, either upward or downward trending, are relatively flat, trend lines for the performance attributes are flat up to a turning point, after which the average value

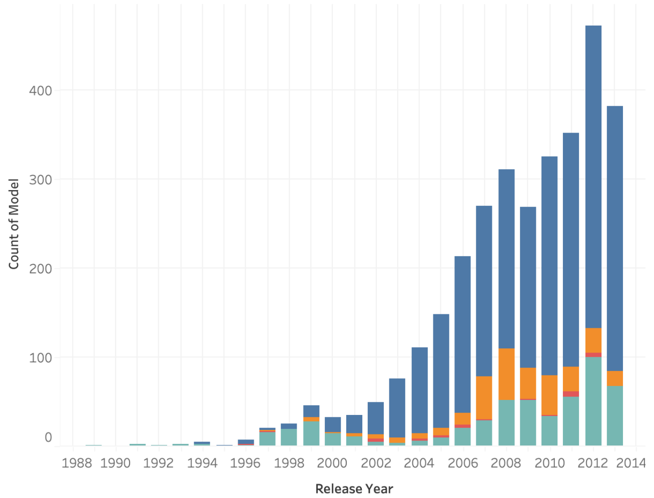

Fig. 3: We omit the colour legend which is visible in figure 3, *tall-tall* devices dominate, this is expected as most smartphones fit this category, the second largest group is *wide-wide* devices. Screens oriented in the direction of the device are clearly the most common.

experienced steep growth through year 2013.

In searching for attributes that are most important in differentiating mobile device types and can potentially be used as axes or visual variables, high variance in attributes would be preferred as they would more likely spread out the models in the visual space. However, as a result of the steep increasing trend lines of performance attributes, they are not as reliable in this aspect as the other attributes, because the variance in these attributes is more reflective of technological advancement throughout the years than differences in device types. This is evidenced by our calculation of total variance and yearly average variance, where the performance attributes showed strong total variance but significantly less so in yearly average variance. Based on this, we placed our focus on the other attributes in discerning mobile device types in the next stage.

IV. Mobile Device Types

a) Tall and Wide Devices: As we established in section II, there are many types of mobile devices in the data, and they are not easily discernible from one another by their performance attributes such as their memory or clock speed. One of the most discernible features of modern mobile devices are their displays. We know that with the rise of touch screens we began to see much larger screens, higher screen-body ratios, and greater pixel density. We aimed to use these features to identify traits of the devices we identified in our initial inspection.

We began by plotting the device length-width ratio with the screen length-width ratio in figure 4 as a simple scatter plot. We chose this arrangement because "height-to-width" ratios are commonly used in the marketing and development of mobile devices and mobile device screens, we expected them to be quite highly distinguishing.

We noted that the point 1 on the y-axis represented where devices are longer than they are wide, and that there is an extremely visually apparent separation on the x-axis from 1.2 to 1.6. By creating two reference lines at $y = 1$ and $x = 1.4$ we separated the data and colour coded these categories to use as a starting point for distinguishing mobile device types. These four categories formed the basis of our identification of mobile device types.

We then checked the quantity and trends of these simple categories in figure 3. To create this visualisation we created a stacked bar chart of the number of devices in each category by year.

This selection of axes, along with the retinal variable of the groupings, shows that *tall-tall* and *wide-wide* devices dominate, especially in more recent years. We can also see that for some time there were very few devices with screens, however around the year 2000 they began to increase in number massively. While the *tall-tall* category seems to be mostly smartphones when we perform a light inspection of the devices classified, we know that there should be separations between pocket devices, and larger tablet or notebook devices.

b) Small and Large Devices: To identify large devices from small ones we began investigating each of these four groups individually on a new scale. Looking directly at the distribution of width and screen diagonal variables in figure 6 we annotated points that marked the end of different peaks of the histogram, at which we could potentially separate devices.

We then created a new scatter plot with width and diagonal as the axes, and our established groups as the colours. We created reference lines at the points labelled on the histogram in order to inspect the devices spread of size across these groups.

In order to gain a better understanding of the *types* of devices we added this plot to an interactive dashboard† (figure 5) where we could filter by groups, and by values above or below the reference lines we added. Additionally we paired this scatter plot with a stacked bar chart representing an ordered count of the devices represented per company where the bar is stacked by the year of the product release. While many companies produce a wide variety of devices, a quick search will often reveal that most of their products are tablets, notebooks, or GPS devices. For example, we were very interested in the *wide-wide* category, which represents a very wide spread of diagonal widths, filtering to *wide-wide* devices that have a diagonal value less than 0.32, we found that the top companies were "MiTAC" and "Null" (where the null results were all pre-2000), performing a quick inspection of these points we saw that the majority of devices below our diagonal cutoff in the *wide-wide*category were GPS navigators. Hence we could classify them as such.

Analysing the other side of this dividing line (see our video for a demonstration) we see that many of the devices are tablets and notebooks. We wanted to distinguish between notebooks and tablets to see clearer trends.

To do this we repeated our procedure of using two histograms to create reference lines for our scatter plots. Figure 7

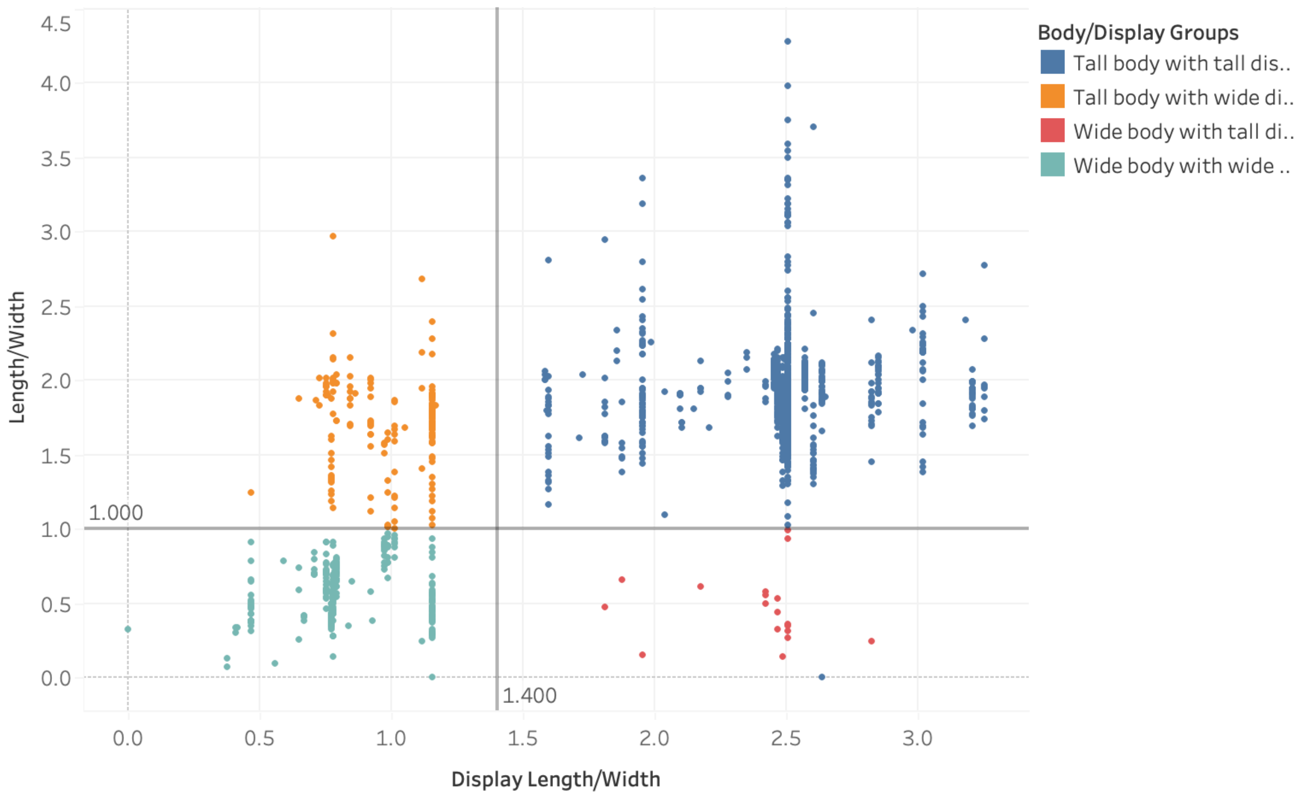

Fig. 4: Length/Width by Display-Length/Display-Width. Clockwise from the top right we define the groups as *tall-tall* devices, *wide-tall* devices, *wide-wide*devices, and *tall-wide* devices. We can distinguish horizontal and vertical devices by whether they are above or below 1 on the y-axis. Additionally, there are clusters around 16:10 and other common screen ratios.

shows the dips we chose to place reference lines on. We split depth at the first heavy dip. We chose to split the mass at two points knowing that notebooks aught to be the heaviest devices, while phones should be the lightest.

These histograms led to our final exploratory scatter plot in figure 8. We were aiming to use mass and depth to distinguish between notebooks and tablets among devices that were above our chosen diagonal reference line. The teal points represent devices that have low depth and medium mass, investigating them with an interactive hover function† indicates that most of the devices in this category are in fact tablets.

With this categorisation we devise a simple device categorisation scheme to identify phones and tablets. We do not expect that this will be 100% accurate, however it serves to neatly identify the important properties we discerned.

```
IF [Display Group Filter]
    = "Mid Mass Low Depth"
AND [Display Diagonal (in)]
    > [Diagonal Reference Line]
THEN "Tablet"
ELSEIF [Body/Display Groups]
    = "Tall body with tall display"
AND [Display Diagonal (in)]
    < [Diagonal Reference Line]
THEN "Phone"
ELSE NULL
END
```

Figure 9 represents devices broken down into many more subcategories than this previous example. We broke the "Display Body Groups" into different categories at the reference lines indicated in our interactive visualisations figure 7 and figure 5. We performed this by simply checking what types of devices existed in each region and allowing for some error. We believe this to be mostly true to the device types.

V. Mobile Device Type Trends

A. *Obsolete and being taken over*

To answer the question of what types of devices are obsolete and being taken over by others, we start from observing the following figure figure 9 of the amount of models of different types of devices produced by the aforementioned categorize method. As can be tell, smartphone and PDA took the place of notebook and palmtop start from early 2000s. And we can inform that PNA have been replaced by modern smartphones start from 2010s, as we all know modern smartphones contain GPS function which makes PNA obsolete. Also, the tablet took place of notebook and palmtop as tablet fills the need of convenient laptop-like devices.

B. Predicting Beyond 2010

Given the information we would have had if the data went up to 2010, we can see several trends. Firstly there is the general trend discussed in section III that devices are consistently getting faster, and gaining storage. Using the exponential trend line in figure 2 we can see that our curves for CPU speed, RAM (Mb), and Storage (Mb) all follow the trend. This tracks with what we know of technological development, and indicates that devices with better technology will continue to enter the market and overtake devices with worse performance.

It is also observed that devices with less storage were released less and eventually stopped over the years. In the years 2011 and 2012, 1 Mb became the standard storage. This can be seen in Figure 10.

Further there are trends specific to some of the mobile device types we extracted from the data, 2009 represented the first year a tablet was released as we defined, and then 2010 represented a huge spike in tablets being released (extrinsically we know that the iPad was released that year) we could easily expect these to be a newly emerging trend, and the trend from 2009-2010 clearly shows a

Screen to body ratios are a common buzzword in mobile phone tech, and while the normalised data makes calculating this ratio somewhat less precise than it otherwise would be, we were still able to roughly estimate it with some simple maths.

$$[\text{Display Diagonal (in)}] \Big/ \sqrt{([\text{Width (mm)}]^2 + [\text{Length (mm)}]^2)}$$

We then visualised this trend in figure 12 with a polynomial-3 trend line that closely matches the data. We can see by the trend line that screen to body ratios are expected to increase, and also that the trend matches the real results.

VI. Evaluation of Visualisation

We evaluated the visualizations by assessing if benefits have been provided through the visualizations. The evaluation method undertaken was 'analytic inspection', where team members reviewed the visualisation against a list of principles, including R. Kosara's minimal set of requirements for visualization [**1**] and principles of depicting information [**2**].

Our visualisations meet R. Kosara's minimal set of requirements as they are based on data, produce images and are both readable and recognizable. For example, to ensure

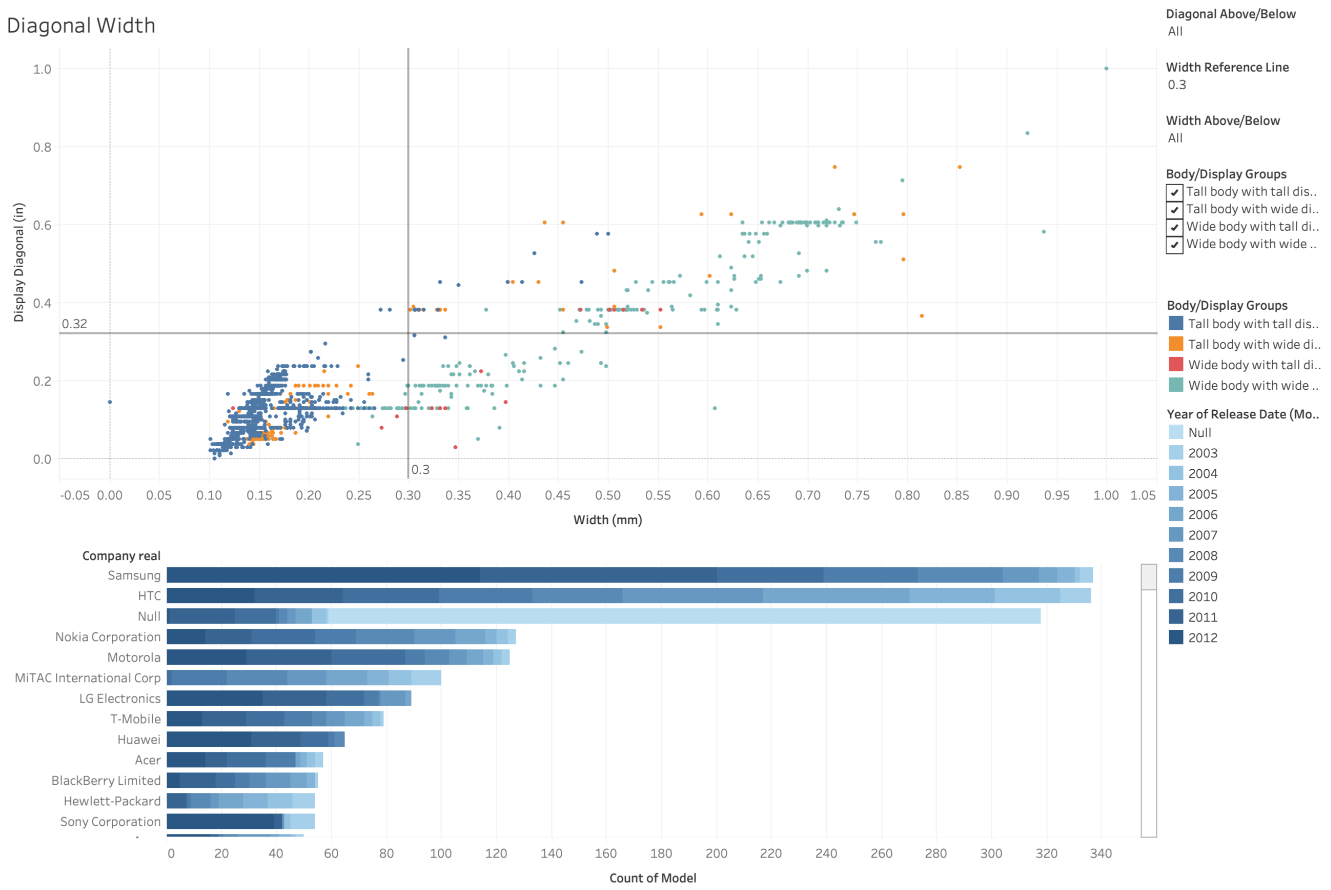

Fig. 5: Interactive dashboard[†] providing filtering for companies and our display feature groups, as well as hover values listing the devices represented at each point, and clickable links searching Wikipedia for each company.

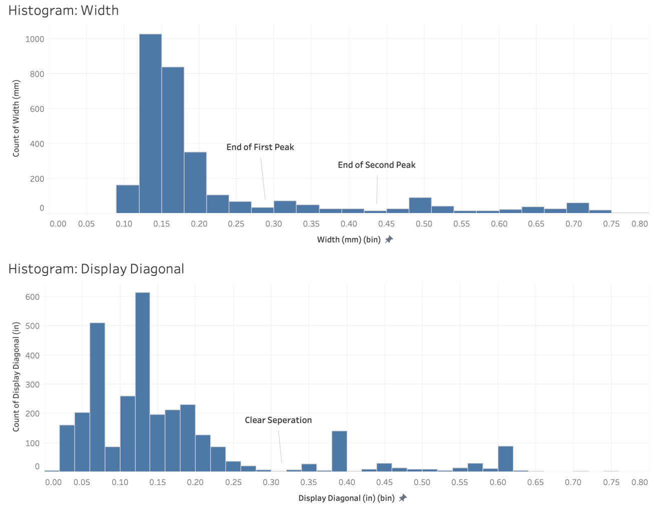

Fig. 6: Histograms of the width and diagonal variables, demonstrating dips we visually separated the data on.

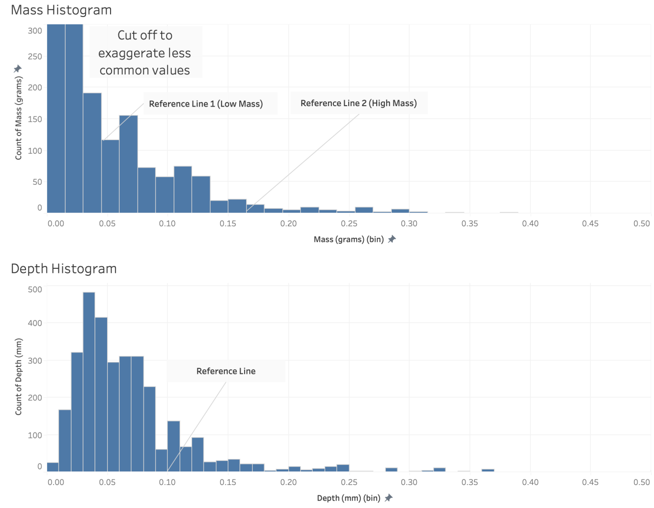

Fig. 7: Histograms for depth and mass, with chosen cut-off dips labelled.

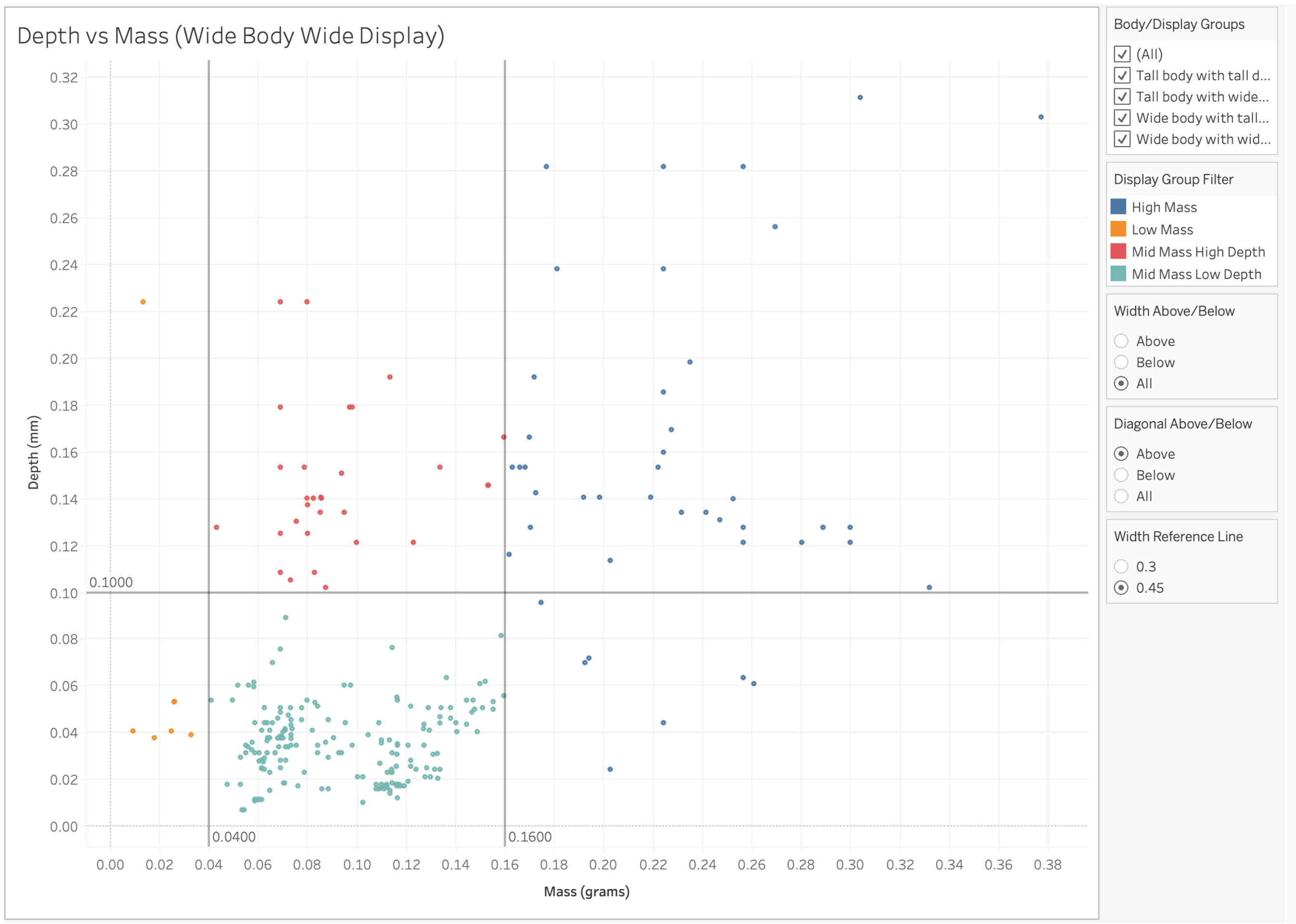

Fig. 8: Scatter plot of depth against mass, we use our reference lines defined in figure 7 in order to create depth/mass categories.

readability, based on the content in each graph, details of mobile device information, such as model, release year, and attributes relevant to the particular graphs were added to be shown when the cursor hovers over an individual point.

With reference to principles of depicting information, modifications were made to visualisations so that they are more helpful in visually leading the readers to conclusions of the different mobile device types and trends. For example, in the interactive dashboard shown in figure 5, the count of model by company bar chart was added to facilitate device type identification, as companies tend to focus on/specialise in a limited number of device types. This is in line with the 'appropriateness principle' where we presented neither more nor less visual information than needed for the task of this graph, which is to discern device types.

Based on aforementioned evaluation, we believe the visualisations presented in the report and the demonstration video have provided benefits in the reader/viewer's visual analysis on the mobile device data.

VII. Conclusion

We derived our classification of types by repeatedly breaking the data into different visual elements where we could differentiate multiple groups. We used an interactive hover function to allow a human user to interpret the points in these dense plots and make decisions as to what groups were representative of what device types. By repeatedly refining our classification on the variables given we were able to identify different types of mobile devices and analyse trends of these types.

In particular we spotted that tablet devices spiked heavily in 2010, this is a very promising result for our classifications as a closer inspection of the data reveals that the first generation iPad was released that year, marking a big shift in the tablet market. We also identified the continually growing amount of smartphone devices released throughout the 2000s. These device types being correctly classified, along with the trends we established, would have proved invaluable information for anyone trying to predict the future of the mobile device market in 2010, and so we believe our visualisations succeed in finding crucial pieces of information within the data.

Our visualisations were at times quite busy, and it is possible that there were much cleaner ways to separate the types of devices, possibly by using more advanced techniques. Additionally we relied on peeking at the underlying data at times to make human judgements about the types of devices, this could have been avoided if we had found clearer clusters among the data but this is a natural step to take given that we were struggling to find other distinguishing patterns.

References

[1] R. Kosara, 2007

[2] COMP 5048 Visual Analytics, Lecture week 1

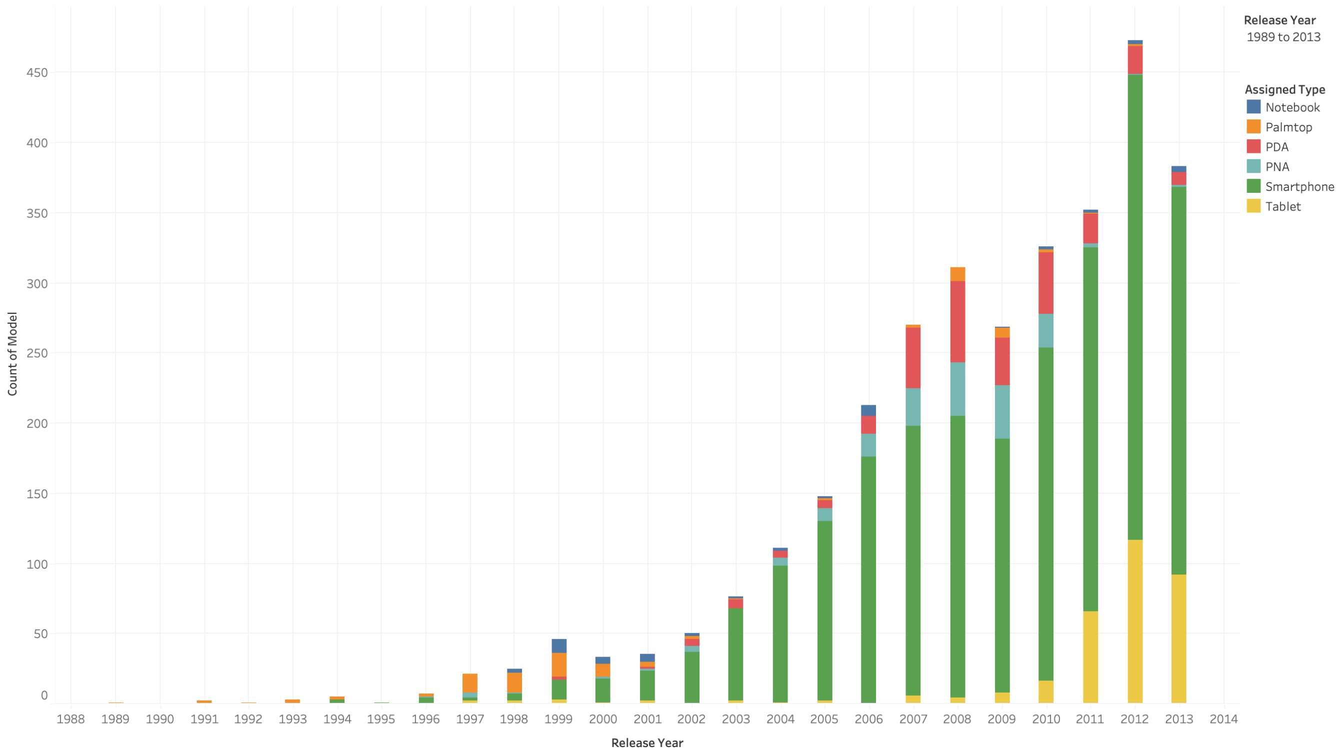

Fig. 9: Device type trends. Types are calculated using more variables (and IF ELSE statements) than figure 10 so we do not include the technical details here.

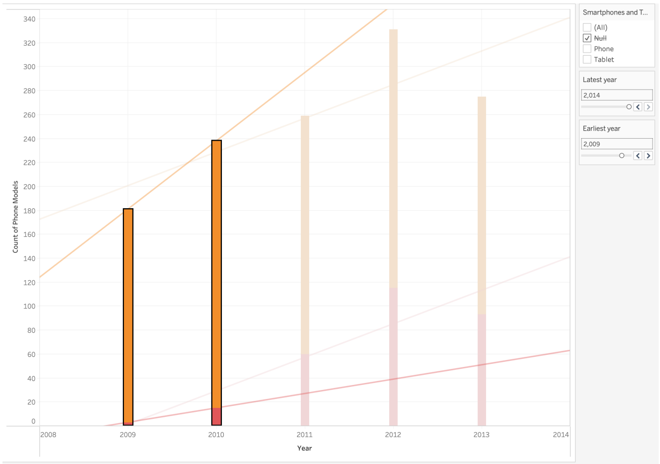

Fig. 10: Trend line calculated from 2009-2010 data, transparent data represents the true values. Orange represents quantity of phones released, red represents the quantity of tablets released

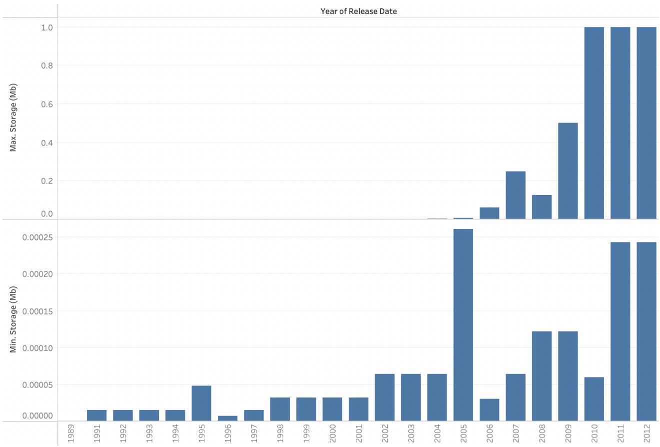

Fig. 11: Storage trend

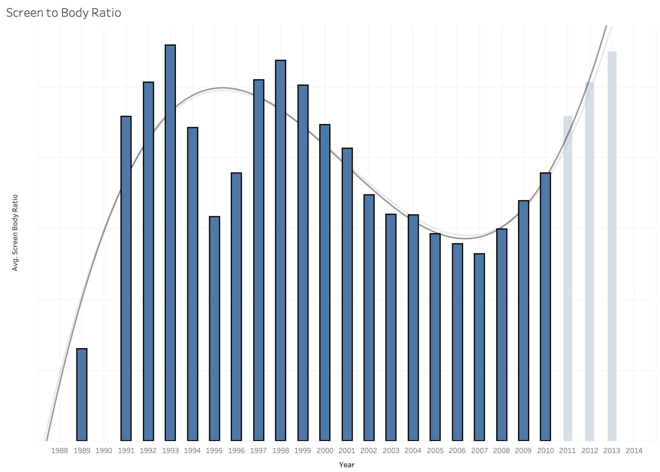

Fig. 12: Average Screen to Body Ratio per year. The trend line is based on the data up to 2010.