Data Exploration Project

Mohit Sainani

April 23, 2019

Prologue

GSMArena is a gadget review website with a focus on cellular and mobile devices. I believe to have come across the site a long time ago through the now defunct iGoogle start-page. It was among the first few tech blogs among the likes of How-To Geek, XDA Developers, Android Police, and AnandTech that had invoked in me the sort of enthusiasm for gadgets and devices that I have carried on ever since.

Introduction

Over the years, GSMArena has compiled a rich database of mobile device specifications. The decision to scrape off their website - despite there being numerous data sets out in the wild to run exploratory analysis upon - was based upon multiple factors. Some of them are:

- an excuse to demonstrate (read : hone, improve) my data mining skills
- experiment with alternative looping techniques in R using the purrr library
- most importantly, the nostalgia attached with working with such a dataset

You can find the code used for scraping device specifications here on my GitHub.

Data Wrangling

This project has been entirely performed in the R (ver 3.5.3) programming language. We start off with reading the exported json file (created after data scrape-ing) and converting it into the familiar row-column structure. Each data frame element from the list is changed into the wide format. After this operation, each row represents specifications of one device.

```
ll <- readLines("gsm.json") %>% fromJSON()

long_to_wide <- function(df){
  as.data.frame(t(df$val)) %>%
    `colnames<-`(paste0(df$type, "_", df$sub_type) %>%
                   str_trim(side = "both"))
}

gsm <- ll$devices %>% flatten() %>% map_df(long_to_wide)
```

The initial dataset comprises of 8980 rows x 86 columns

Comparing this dataset with the GSMArena repository, we find that we managed to capture data for 93.5% (8980 out of 9601) devices listed on the website.

We restrict the dataset to about 24 columns which fall under the scope of the proposed questions.

Next, we'll clean some of the data columns that we know will be required for answering the suggested questions. Regex is used for pattern matching, and the R tidyr and dplyr libraries are used for wrangling purposes. The tasks performed are:

1. Device weight in grams
   The *body_weight* column is a string which has information in both grams and ounces. We extract the weight in grams and cast it as a numeric value
2. Body dimensioms
   Extract the length, breadth, and width (in millimeters) of each device from the the *body_dimensions* column
3. Display size
   Extract the display size (diagonal) of the device in inches
4. Battery capacity
   Extract the battery capacity (in mili ampere hours, mAh)
5. Display resolution
   Extract the screen resolution as horizontal and vertical pixel counts, and save them in respective fields
6. Price
   Separate the price amount and price currency into different variables
7. RAM
   Get the RAM from *memory_internal* field and convert it into gigabytes of memory
8. Bluetooth version
   Get the version of bluetooth implemented in each device
9. Camera lense count
   Collpase the *main_camera* fields into a new column indication the count of main camera lense(s)
10. Year announced
    Obtain the year in which the device was first announced
11. Screen-to-body Ratio
    Although GSMArena provides this metric for devices announced in the last few years, we'd prefer to obtain this number for all devices. It may be calculated based on the diagnoal screen size, the display resolution of the screen, and length and width of the phone. Simple use of Pythagoras' theorem yields the desired value.

Data Checking

Our dataset consists of all sorts of devices including tablets and smartwatches. These were discovered using filters based on screen size, available cellular networks, and year of announcement. We know that the first modern smartwatch was announced in 2011[1]. Also, tablets usually tend to be over 7 inches in screen size[2].

Removing smartwatches and tablet devices

Since we are only concerned with smart (or dumb/feature) *phones* we can filter for devices that match certain criteria:

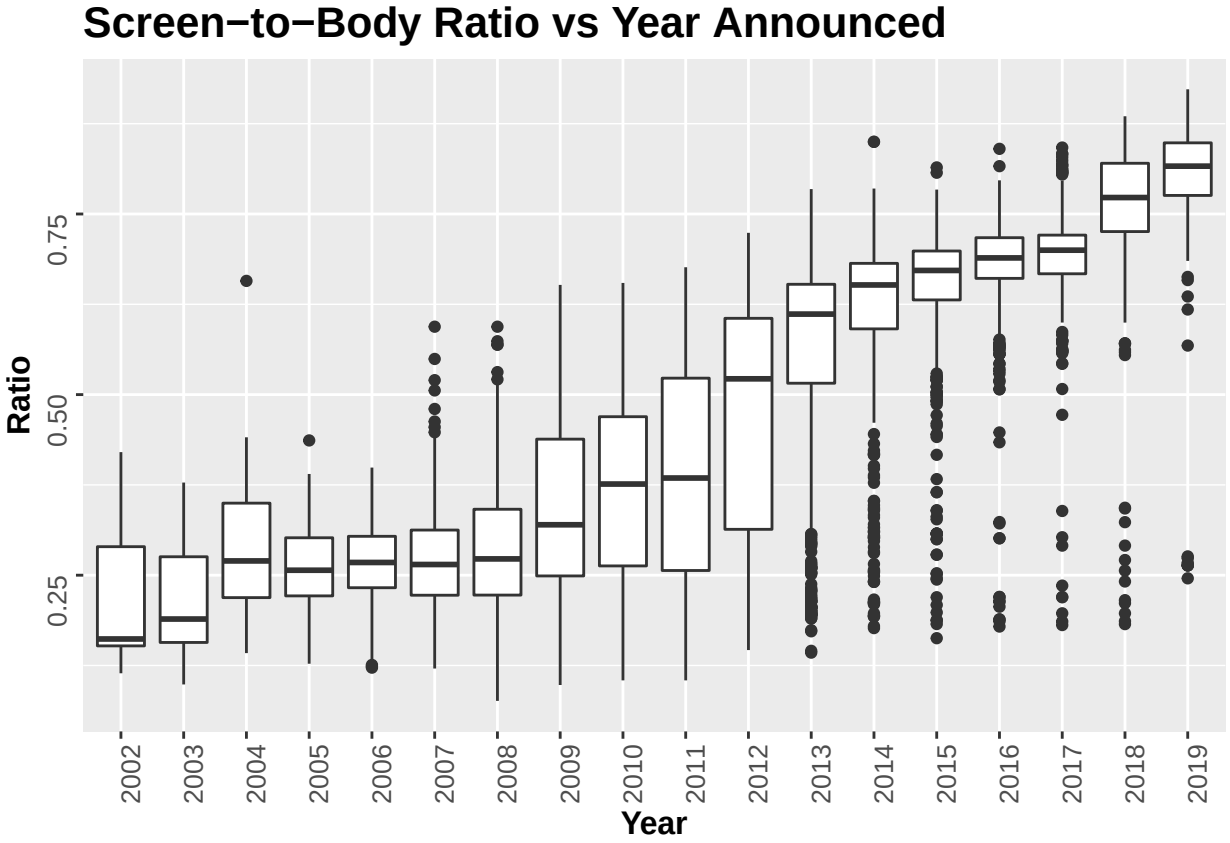
- Remove devices with *No cellular connectivity*
- Remove devices that have screens over 6.9 inches (predominantly tablets)
- Remove devices that were announced after 2011 and have screens smaller than 1.7 inches (mostly smartwatches)

[1] https://en.wikipedia.org/wiki/Smartwatch#2010s
[2] https://en.wikipedia.org/wiki/Tablet_computer#Mini_tablet

Exploration

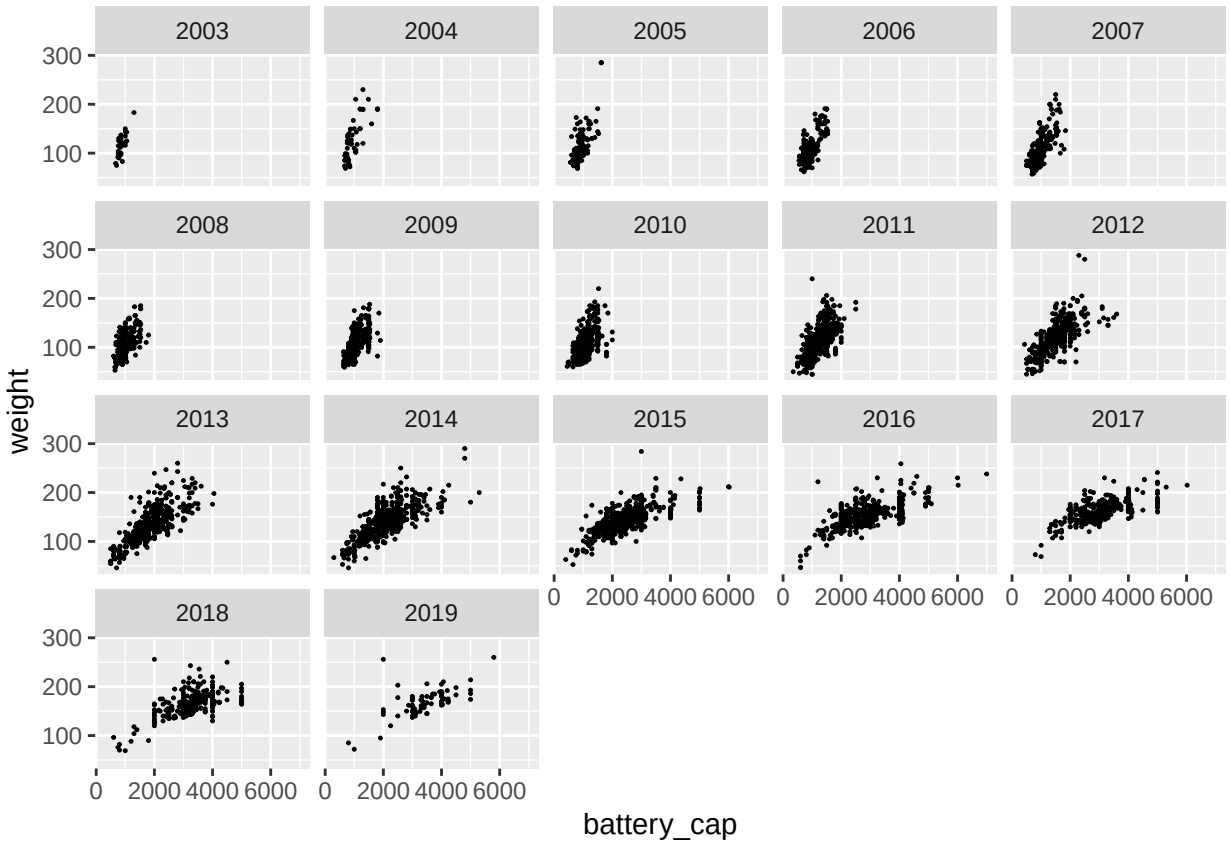
Say No to Bezels

Bezel-less display is one of the many buzz words surrounding the mobile phone industry these days. The term bezel refers to the portion of a device's front surface that is not screen estate. Most mobile phones today look more or less the same with their ubiquitous rectangular profile and a display ingrained on the front. However this wasn't always the case when companies were trying to innovate on the design and usability fronts. Some of the popular designs that have gone extinct are flip-phones, phones with slide-out or front-facing hardware keyboards, devices with track-balls, and phones resembling portable gaming consoles.



Notice how the rectangular boxes keep getting smaller in lenght as we move ahead the x-axis. This shift represents a homegenity in design which the industry seems to have embraced following in the footsteps of the first monolithic iPhone[3]. Outliers below the first quartile also witness a drop in number, signalling a decrease in feature-phone announcements as these are typically the devices that have both a display and hardware keypad, and thus a lower screen-body-ratio.

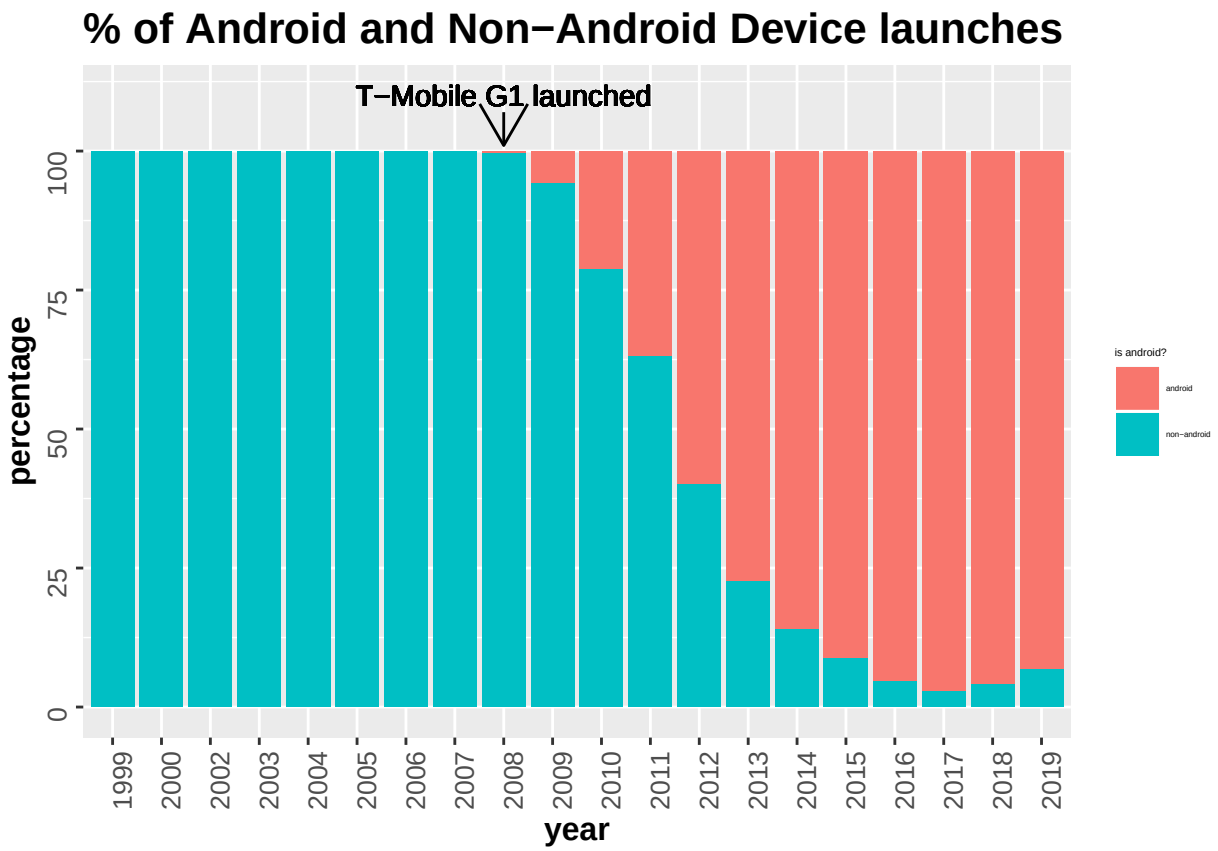
[3]Citation Pending



Part 5

The first android device (according GSMArena's database) was released in the year 2008. It was T-Mobile's G1 running Android 1.6 (Donut). It probably marked the beginning of a new era for smartphone operating systems. Since then, Android has become wildly dominant and perhaps the only OS (apart from Apple's iOS) to survive in today's market. The likes of Symbian, Microsoft's Windows Mobile, BlackBerry OS, Palms OS, and Samsung's Bada OS have all become extinct now.

We take a look at the ratio of Android and non-Android devices announced in different years:

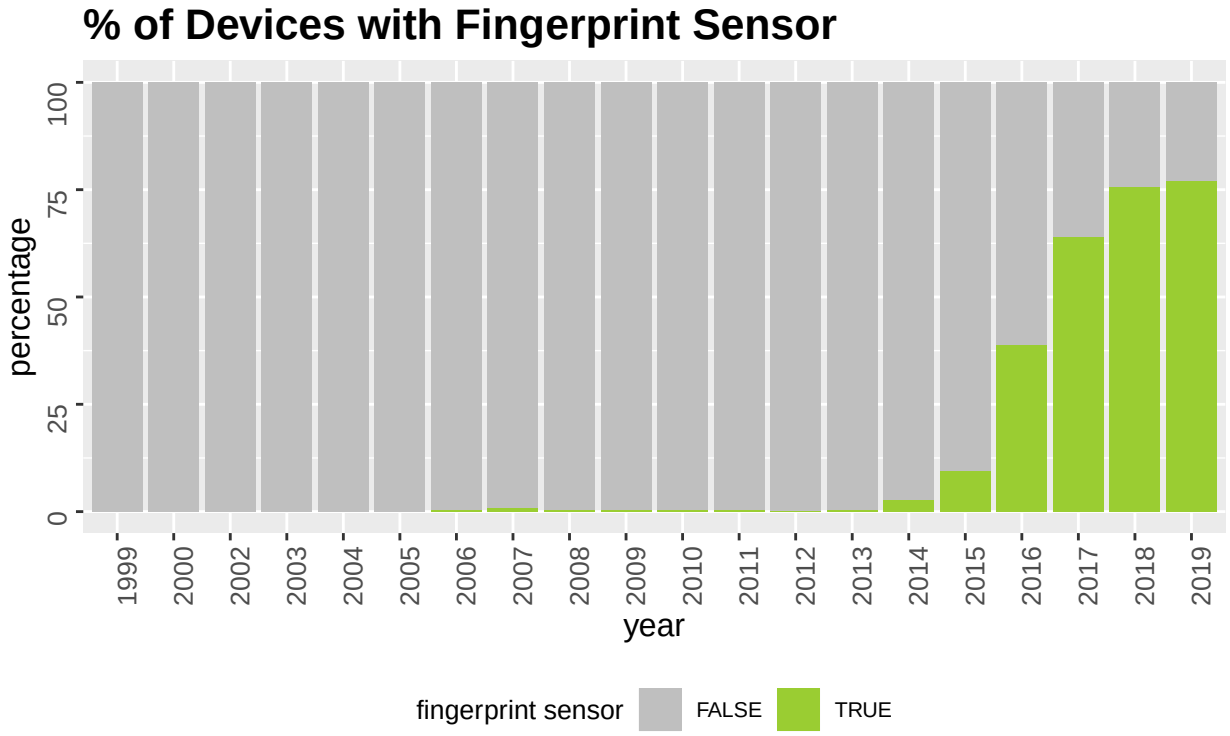


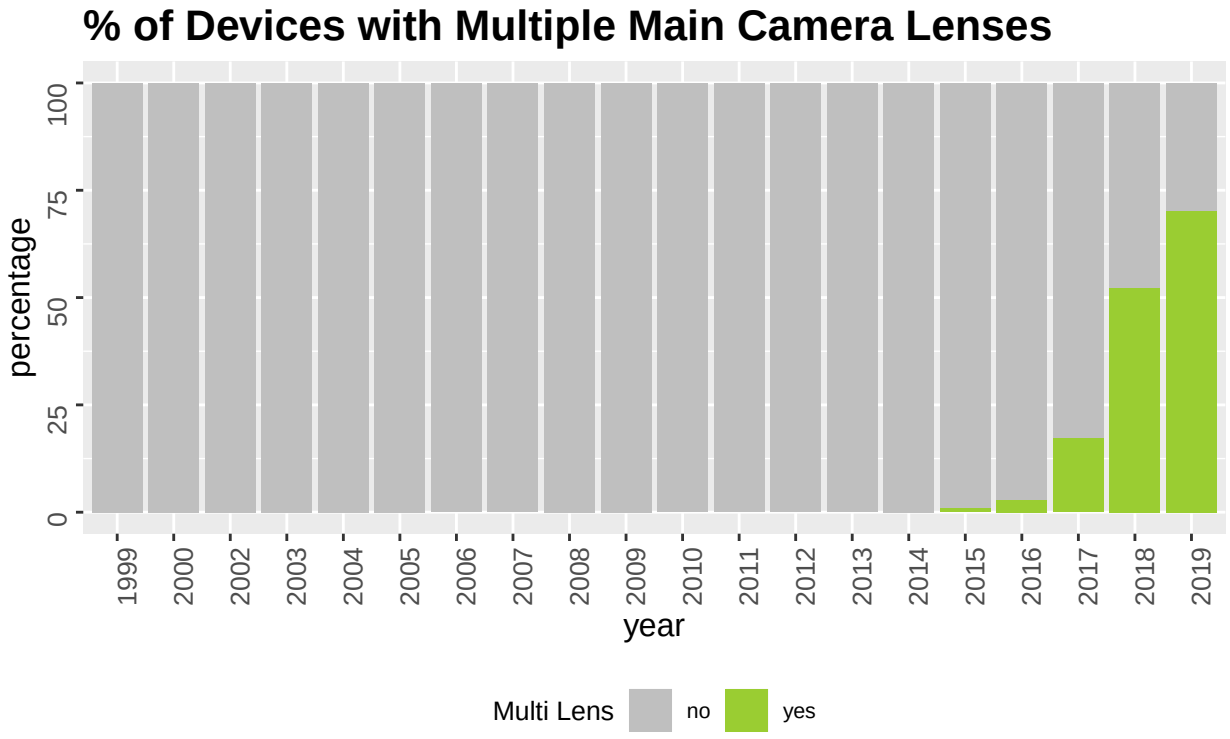
The above chart clearly displays the sort of dominance that Android has managed to gain in the smartphone market. In the last couple of years,

Answer 4

In terms of senosrs and features, mobile phones have become extremely versatile devices. What started as a device meant for serving the purpose of calling people remotely and sending text messages, has evolved into a feature-packed gadget that comes with camera(s) for photography, GPS for navigation, fingerprint readers for secure access, and accelerometers and gyro that help in motion sense gaming. A few mobile phones even come equipped with health-related sensors like heart-rate monitors, or are IP certified to be used in extereme conditions, such as underwater.

Some of these features (or trends) have become more-or-less standardized and we will take look at two of them: multiple main camera lenses, and fingerprint readers





Reflection

capture device popularity, store errors/ devices not scraped