

# Group Homework

*Tigran Sedrakyan*

*Nareg Barsoumian*

*Nane Vardanyan*

*Kevork Sulahian*

*Hovhannes Margaryan*

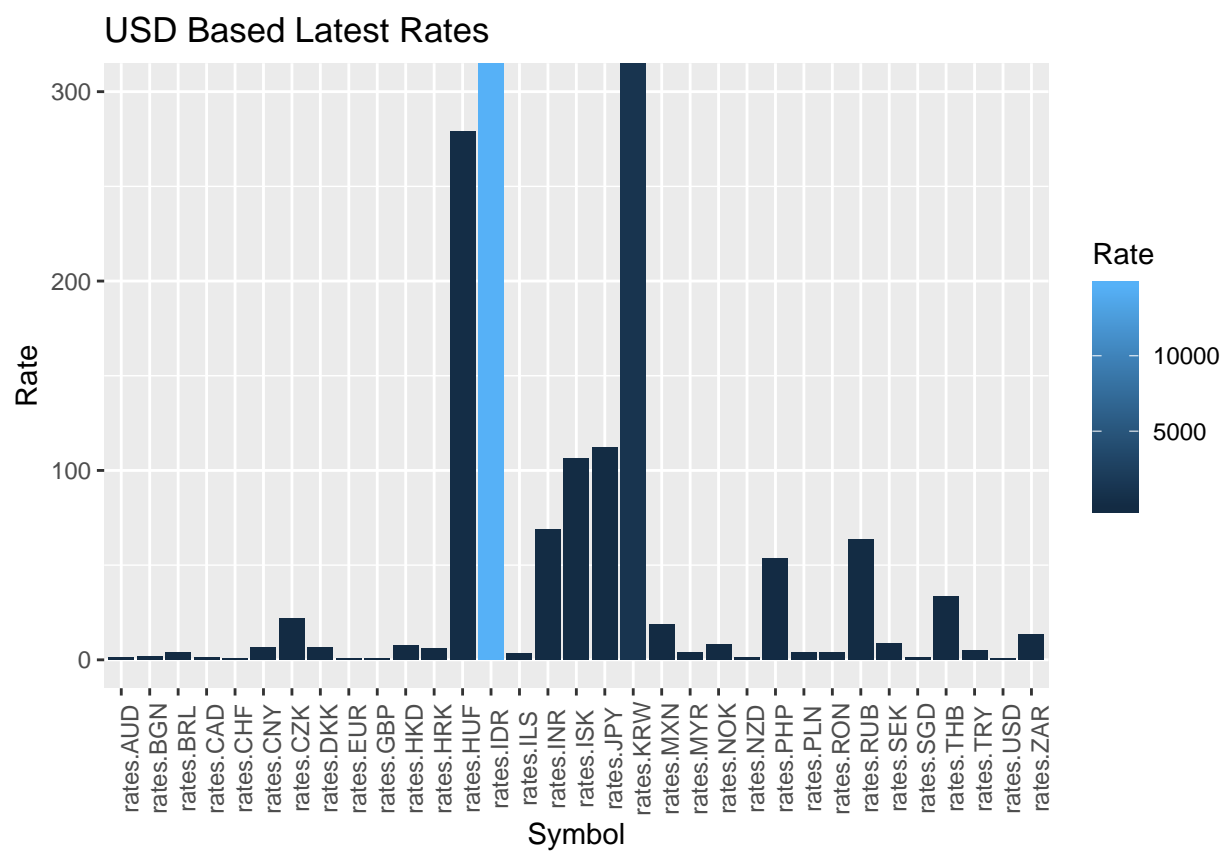
*July 21, 2018*

## Quantitative Analysis

### Getting the data

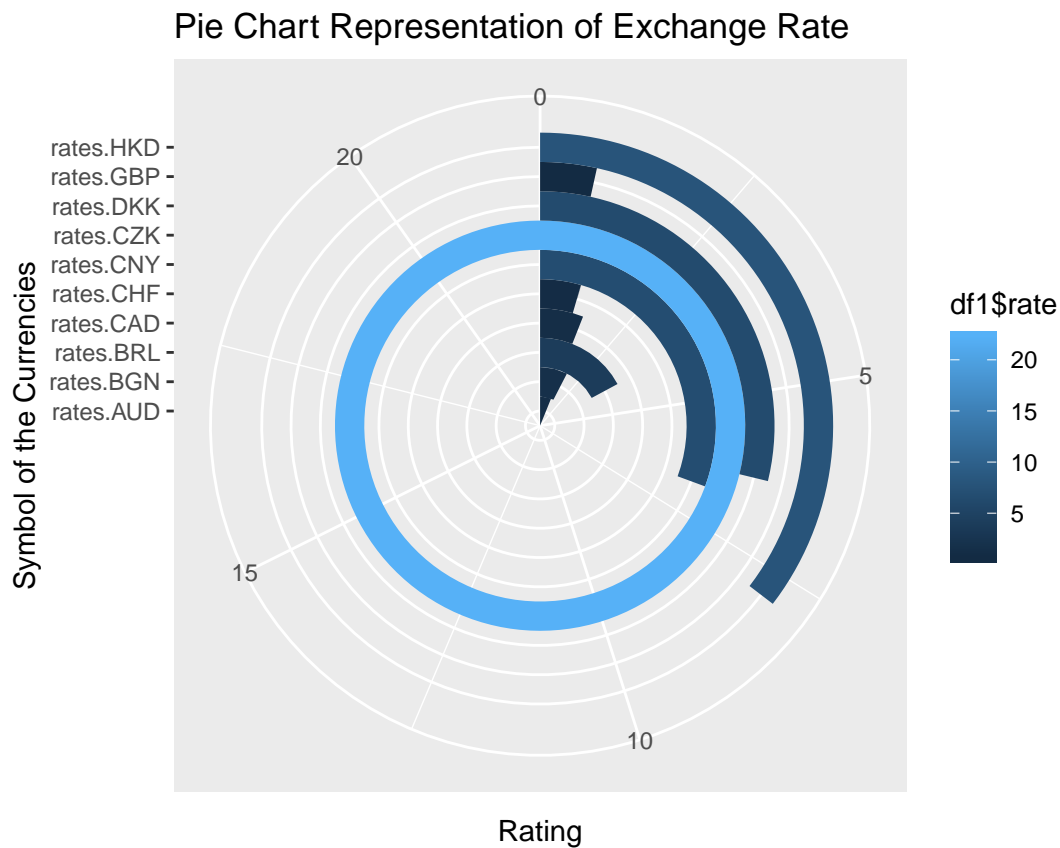
For getting the quantitative data, we chose API requests to [Exchange Rates API](#). This API provides exchange rates of currencies for any day since 1999. We made several types of date requests, but for any of them the base currency was chosen the USD.

### Analysing the data

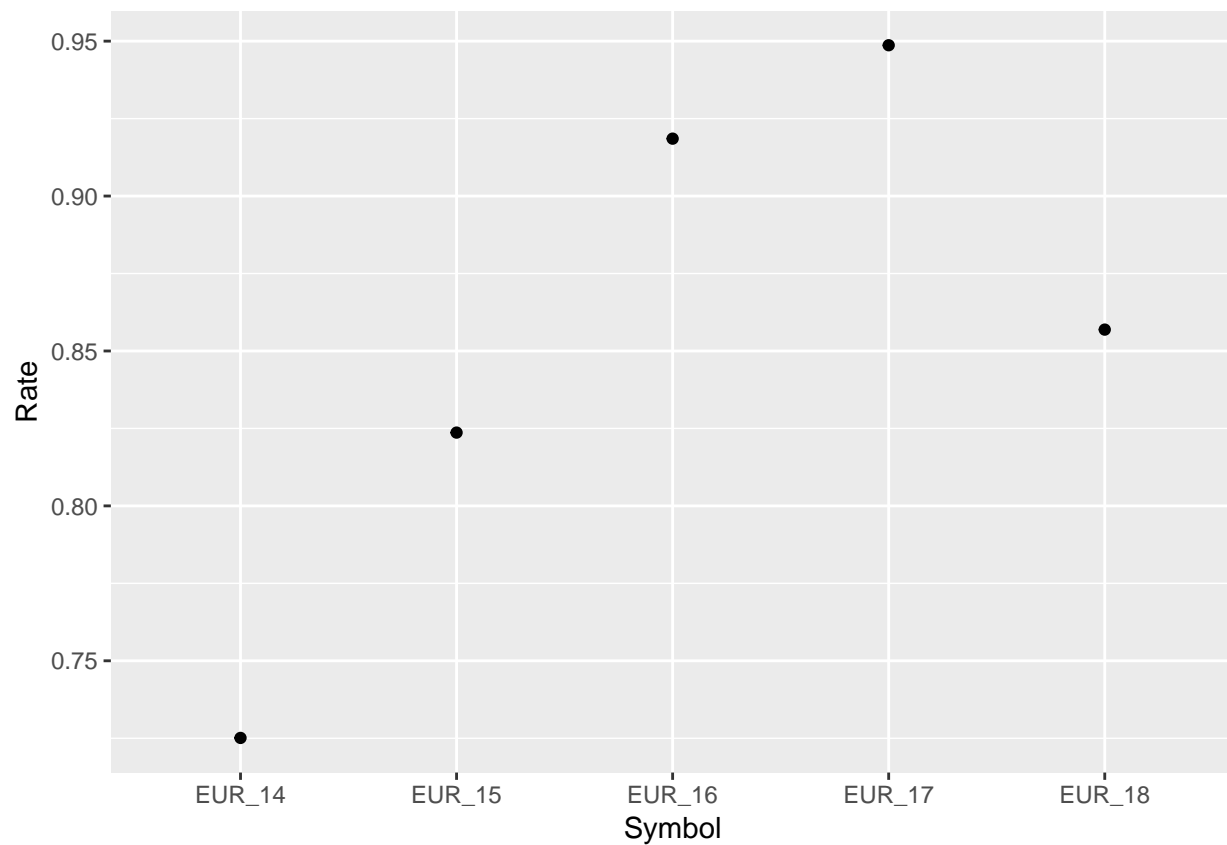


In the bar plot above we can see the latest exchange rates based on US dollar. The x-axis shows the symbols for currencies and the rates are depicted on the y-axis. According do the plot, we can say that after IDR and

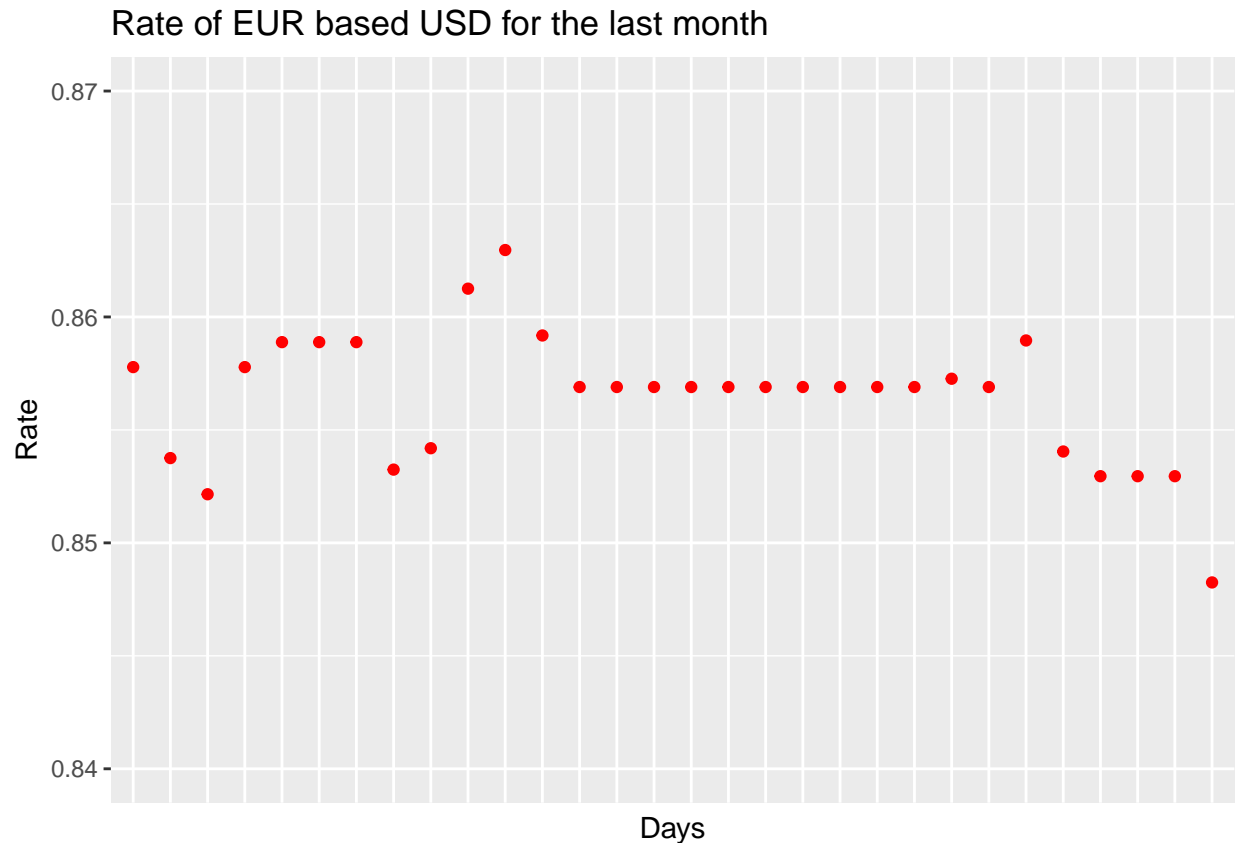
KRW, HUF and JPY have the highest rates. Moreover, GBP has the lowest rate. We can represent the same chart in a bit more elegant form.



However those chart stand for the most recent exchange rates. Let's see, however, how the rate for EUR has changed, for the last 4 years.



This plot shows a general idea about the rates of EURO based USD from 2014 to 2018. For the plot, only a single day has been chosen each year for a general idea about the behavior of the rates through years. We encounter the high rate in 2017 and the lowest 2014. We see a constant increase with the exception of last year, when rate for Euro has dropped. If we nail down the date range to the last month, we see the following tendency:



This plot shows the rate of EUR based on USD for the last month (June).

## Qualitative Analysis

### Getting the data

For qualitative analysis part of our group homework we chose to scrape data from Steam game store and service. For this purpose, we created a dataframe with 4 columns: Game Id, Game Title, Vote and Review. The first one shows the id of the game, the second one corresponds to the title, third one is “Recommended” or “Not recommended”, converted to a number (1 and -1, correspondingly). The fourth one is the review left by the user. We chose 5 games, and for each game we’ve scraped the most recent 1010 reviews, with corresponding votes. It’s worth noting that we chose the 5 most played games at the moment, according to [SteamDB](#) in the corresponding order of daily players:

1. PLAYERUNKNOWN’S BATTLEGROUNDS
2. Dota 2
3. Counter-Strike: Global Offensive
4. Warframe
5. Tom Clancy’s Rainbow Six Siege.

The game ids were extracted manually and kept in a vector over which we later iterated. The rest of the data, including titles were scraped using different techniques. The data was then saved into csv file, for easier future usage.

Of particular interest is the column Review, which we had to clean of the needless information, which could later interfered in the analysis part. With every review came the first phrase, which contained the ‘Posted:’ +

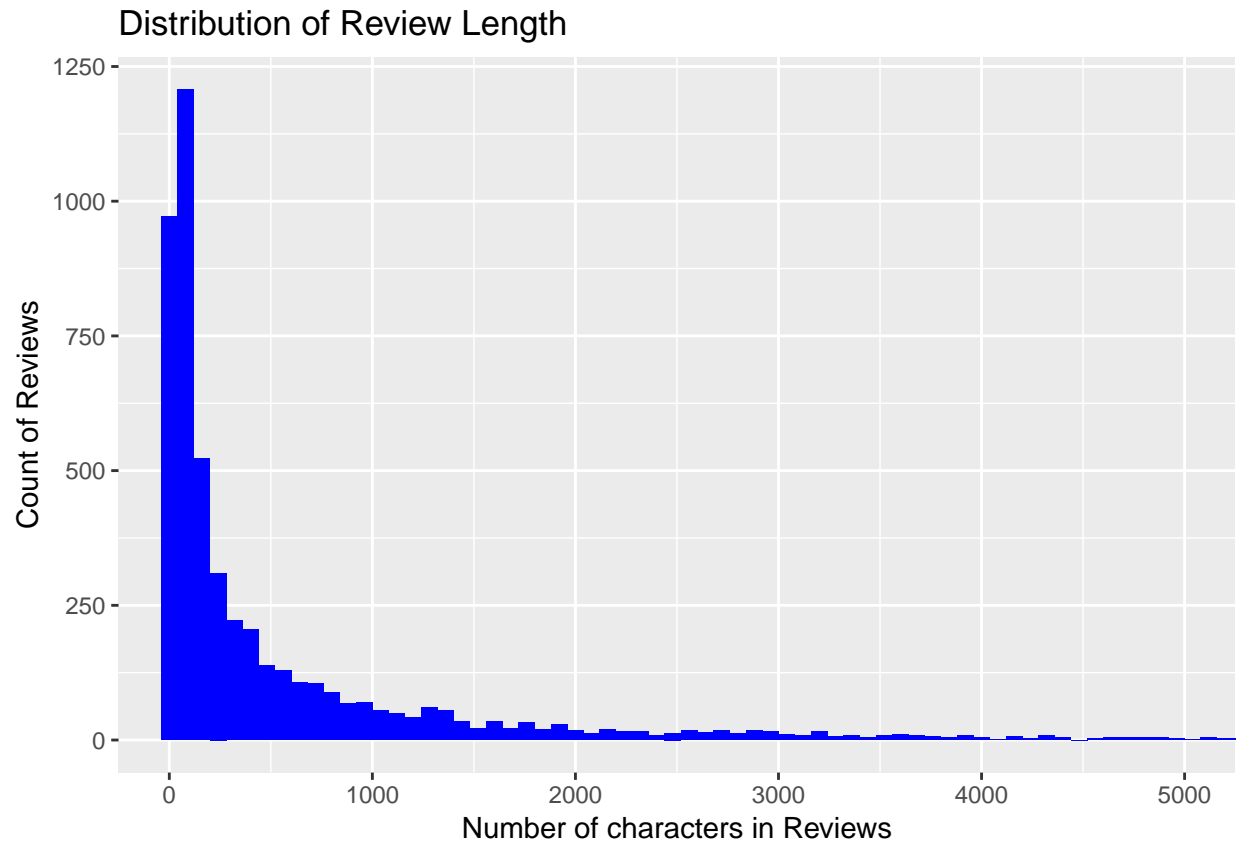
the date of posting of the review. Because we're doing pure text analysis we thought that it'd be better to get rid of this text. This was achieved using regex patterns. Whitespaces, with the exception of space, and non-ASCII characters were also removed, thus leaving pure reviews, if such were present. All of the above can be found in file which can be found in the scrape.R file attached.

In the end we got a dataframe that looks like this:

```
##      Game.Id      Game.Title
## 1  578080 PLAYERUNKNOWN'S BATTLEGROUNDS
## 2  578080 PLAYERUNKNOWN'S BATTLEGROUNDS
## 3  578080 PLAYERUNKNOWN'S BATTLEGROUNDS
## 4  578080 PLAYERUNKNOWN'S BATTLEGROUNDS
## 5  578080 PLAYERUNKNOWN'S BATTLEGROUNDS
## 6  578080 PLAYERUNKNOWN'S BATTLEGROUNDS
##
## 1
## 2
## 3
## 4 Early Access ReviewPUBG is the type of game where it starts off with not caring too much if you die
## 5
## 6
##      Vote
## 1     -1
## 2     -1
## 3     -1
## 4      1
## 5     -1
## 6     -1
```

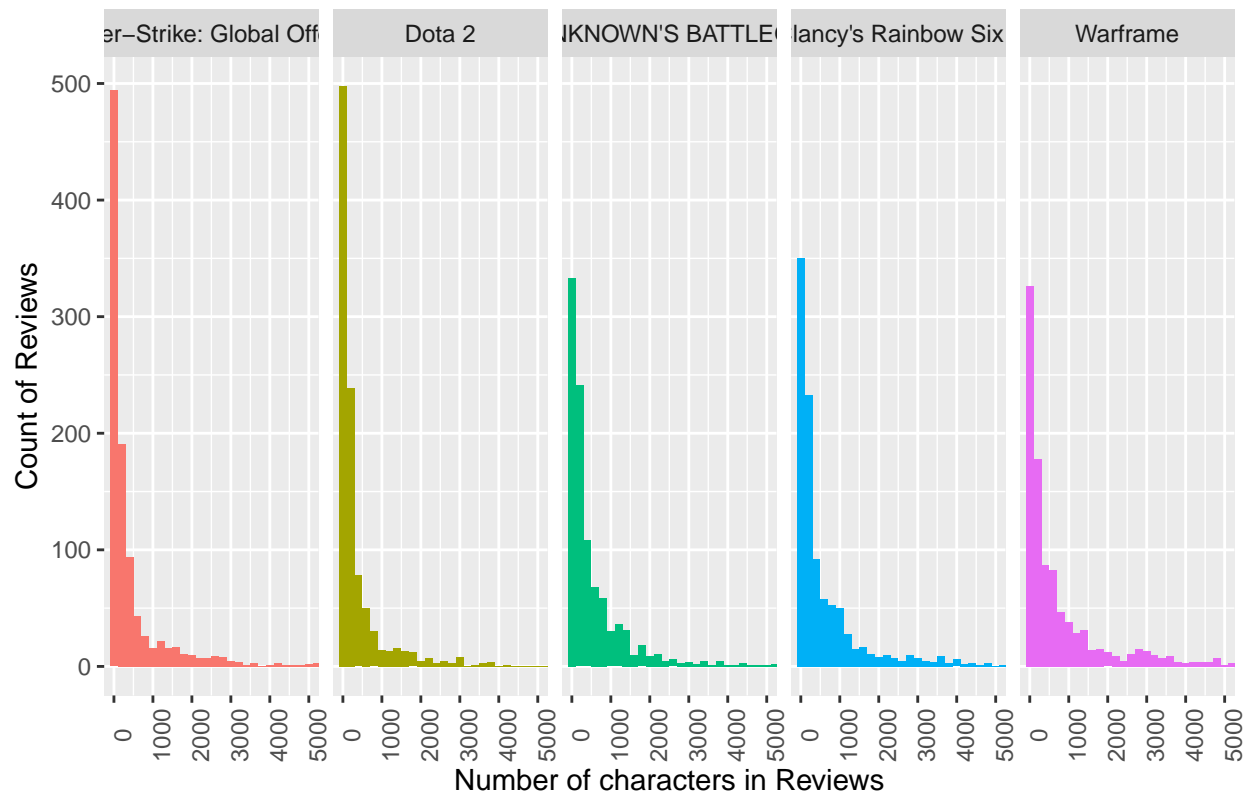
## Analysing the data

First of all, let's see how review character length is distributed.



The picture is not surprising. Only very few users spend their time on writing reviews with over 1000 character length. Mostly people manage to fit their gratitude or anger in less symbols, that's why we have more reviews with few symbols and less reviews with a lot of symbols. But how does this plot look for each game individually?

### Distribution of Review Length By Game



As we see review length distribution for each game isn't much different from that of all game altogether. Although we notice here that Counter Strike and Dota 2, in contrast to the other 3 games, are concentrated to the left, meaning that people spend the least time and characters for writing reviews for these games. We also notice that little spike around 3000 characters for Warframe. Let's see which words are the most popular in reviews of this game.



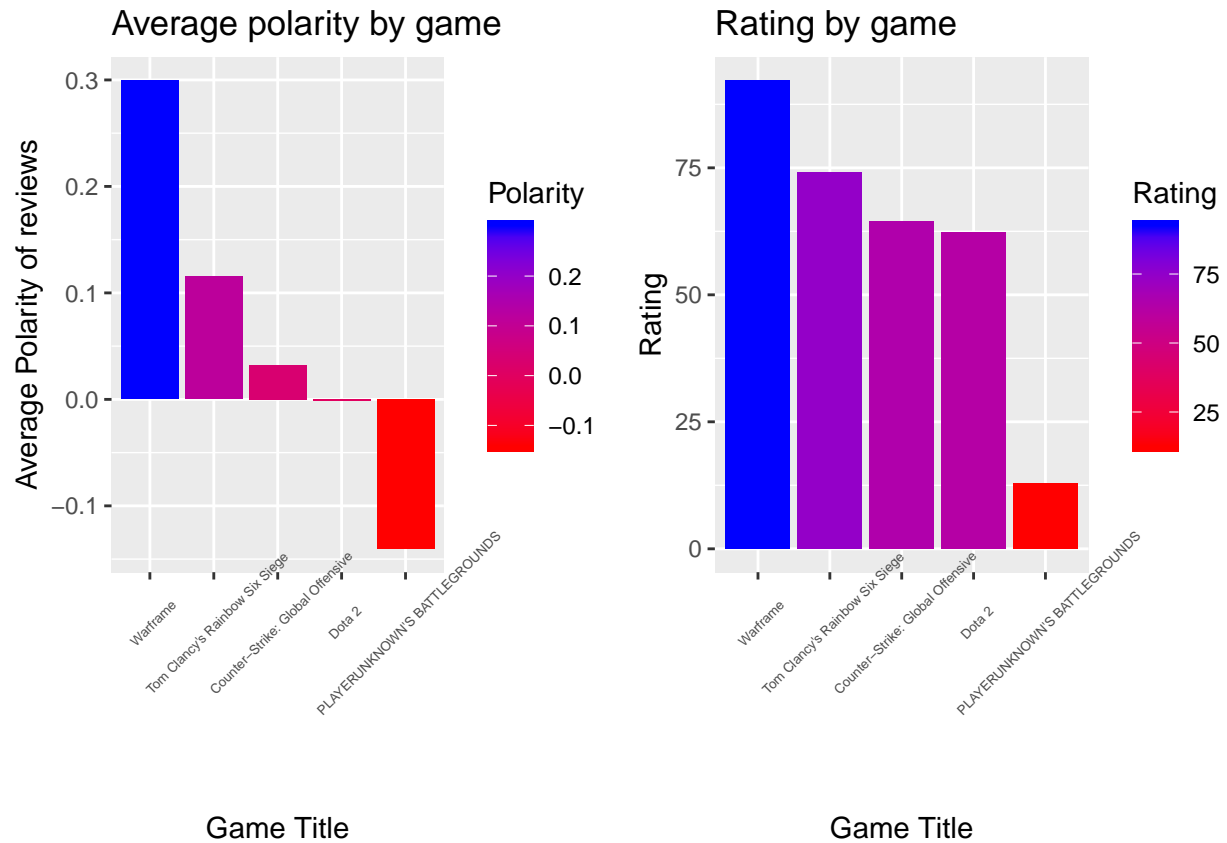
Its title 'Warframe' is predominant here. We also notice terms like play, free and good, indicating generally favourable reviews. But this wordcloud doesn't tell us much about the game compared to the other 4 games in our list. Let's see how games compare in a comparison wordcloud.



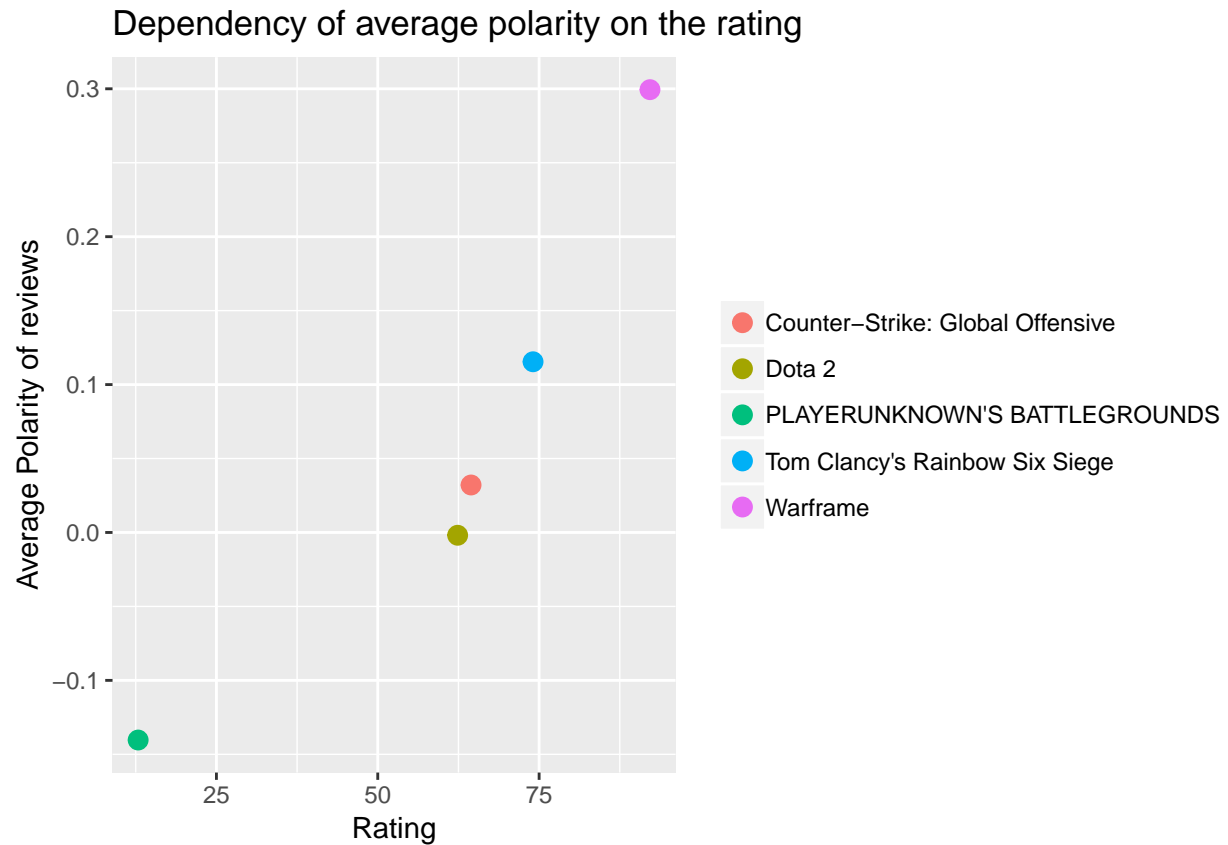


9





As expected, PUBG has the lowest polarity, which drops below zero, indicating that the reviews were mostly negative. In contrary to PUBG, is already familiar Warframe, which has pretty respectable polarity score of almost 0.3. We see approximately the same picture for rating barplot. PUBG has the lowest rating of about 12-13 out of 100. Warframe is again on the opposite side of the chart with rating of more than 90. But Polarity and Ratings from those two charts seem correlated, don't they?



They are indeed, positively correlated as we see. Increase in one of them implies the increase of the other and vice versa.

Looking at this graphs one might think that PUBG definitely cannot sell good. However, one unexplainable phenomenon to humanity is how this game manages to stay the [#1 top selling game in the world](#) for over a year, with over a million active daily players. One of the possible factors is that it's one of the first relatively big games in its [battle royale](#) genre.

We can make a conclusion that people pay more attention to the top selling charts, rather than reviews left by users just like them. This implies that if the game sells good doesn't necessarily mean that it's the best one.