Twitter

2024-08-21

Twitter Sentiment Analysis Data-set

This is an entity-level sentiment analysis data-set of twitter. Given a message and an entity, the task is to judge the sentiment of the message about the entity. There are three classes in this data-set: Positive, Negative and Neutral. We regard messages that are not relevant to the entity (i.e. Irrelevant) as Neutral.

ID <int>	Topic <chr>	Sentiment <chr>
2401	Borderlands	Positive
2401	Borderlands	Positive
2401	Borderlands	Positive
2401	Borderlands	Positive
2401	Borderlands	Positive
2401	Borderlands	Positive
2402	Borderlands	Positive
2402	Borderlands	Positive
2402	Borderlands	Positive
2402	Borderlands	Positive

1-10 of 10,000 rows | 1-3 of 4 columns   Previous 1 2 3 4 5 6 ... 1000 Next

Dimension of the data-set

```
[1] 74682     4
```

Variables in the data-set

```
[1] "ID"        "Topic"     "Sentiment" "Text"
```

Description of the variables in the dataset

```
'data.frame':   74682 obs. of  4 variables:
 $ ID       : int  2401 2401 2401 2401 2401 2401 2402 2402 2402 2402 ...
 $ Topic    : chr  "Borderlands" "Borderlands" "Borderlands" "Borderlands" ...
 $ Sentiment: chr  "Positive" "Positive" "Positive" "Positive" ...
 $ Text     : chr  "im getting on borderlands and i will murder you all ," "I am coming to the borders and I will kill you all," "im getting on borderlands and i will kill you all," "im coming on borderlands and i will murder you all," ...
```

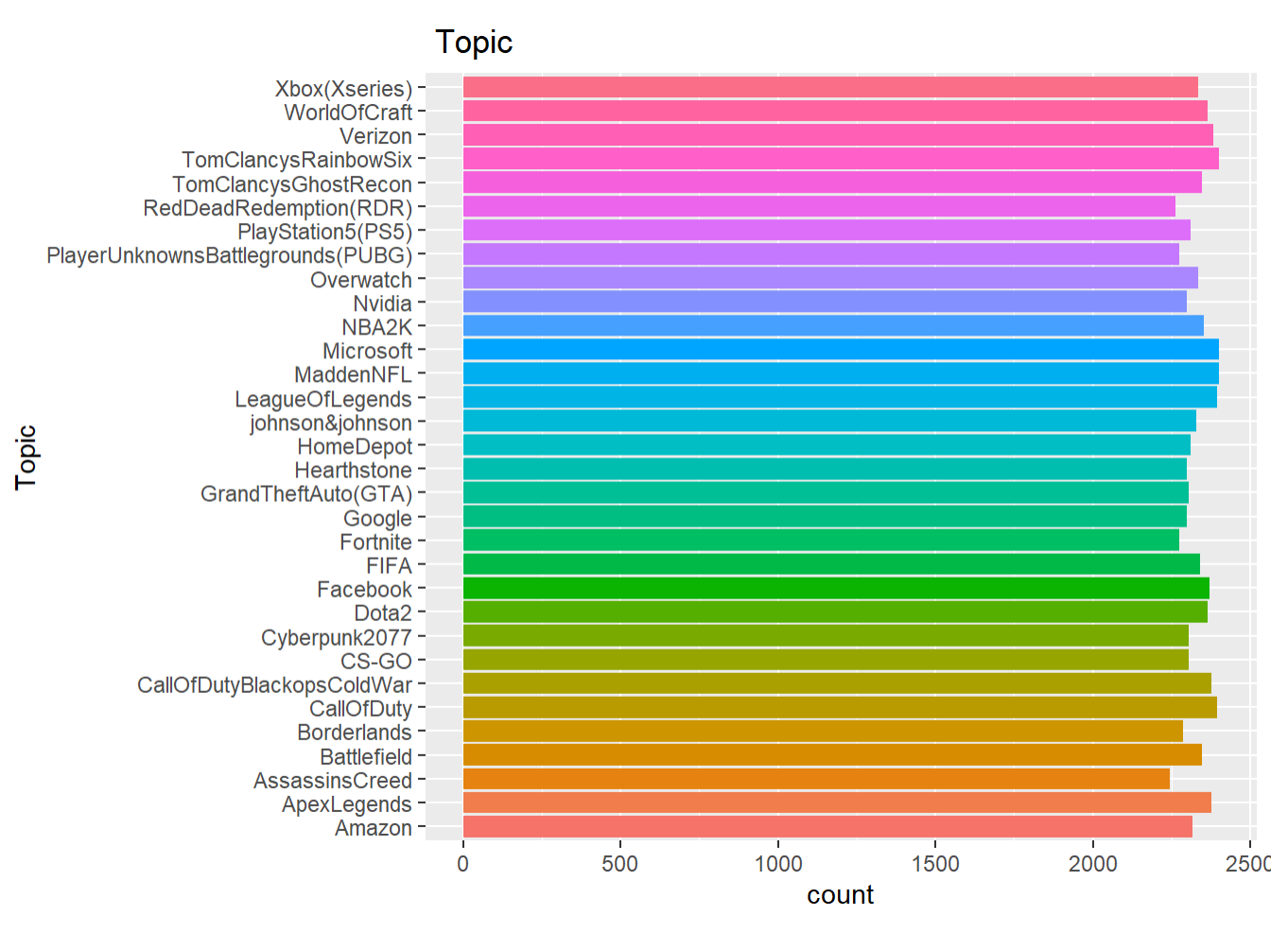
Summary of the data-set

```
       ID           Topic              Sentiment              Text          
 Min.   :    1   Length:74682       Length:74682       Length:74682      
 1st Qu.: 3195   Class :character   Class :character   Class :character  
 Median : 6422   Mode  :character   Mode  :character   Mode  :character  
 Mean   : 6433                                                           
 3rd Qu.: 9601                                                           
 Max.   :13200                                                           
```

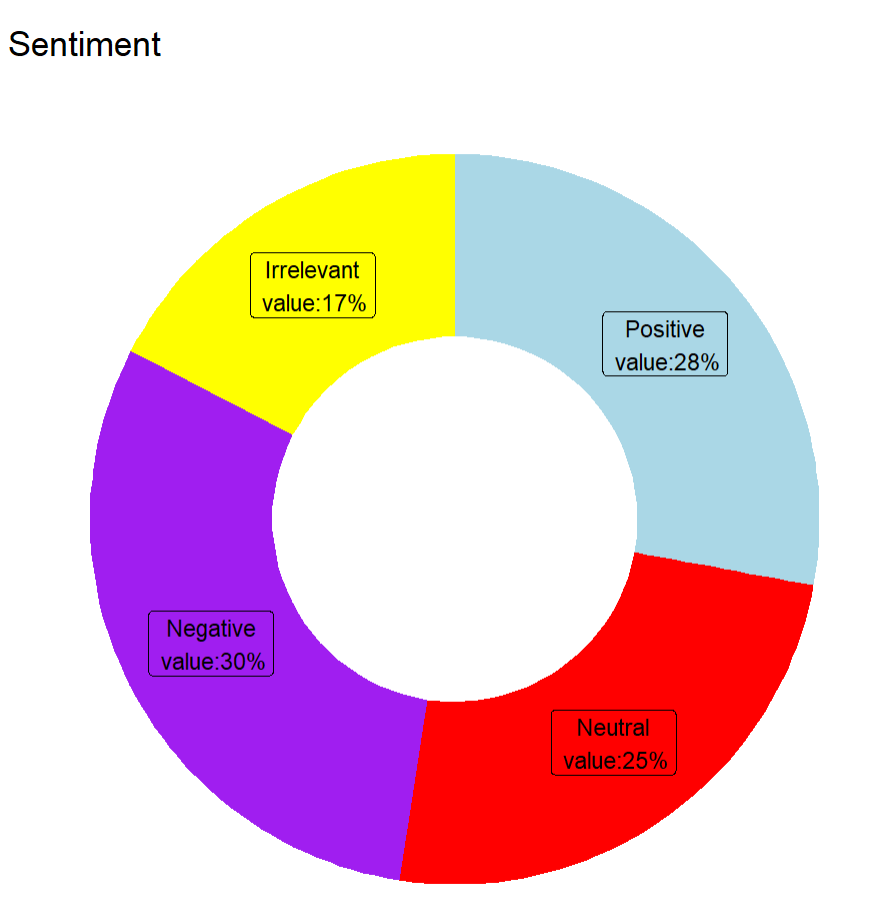
Checking for the missing values

```
[1] 0
```

Column diagram for differnt topics vs count

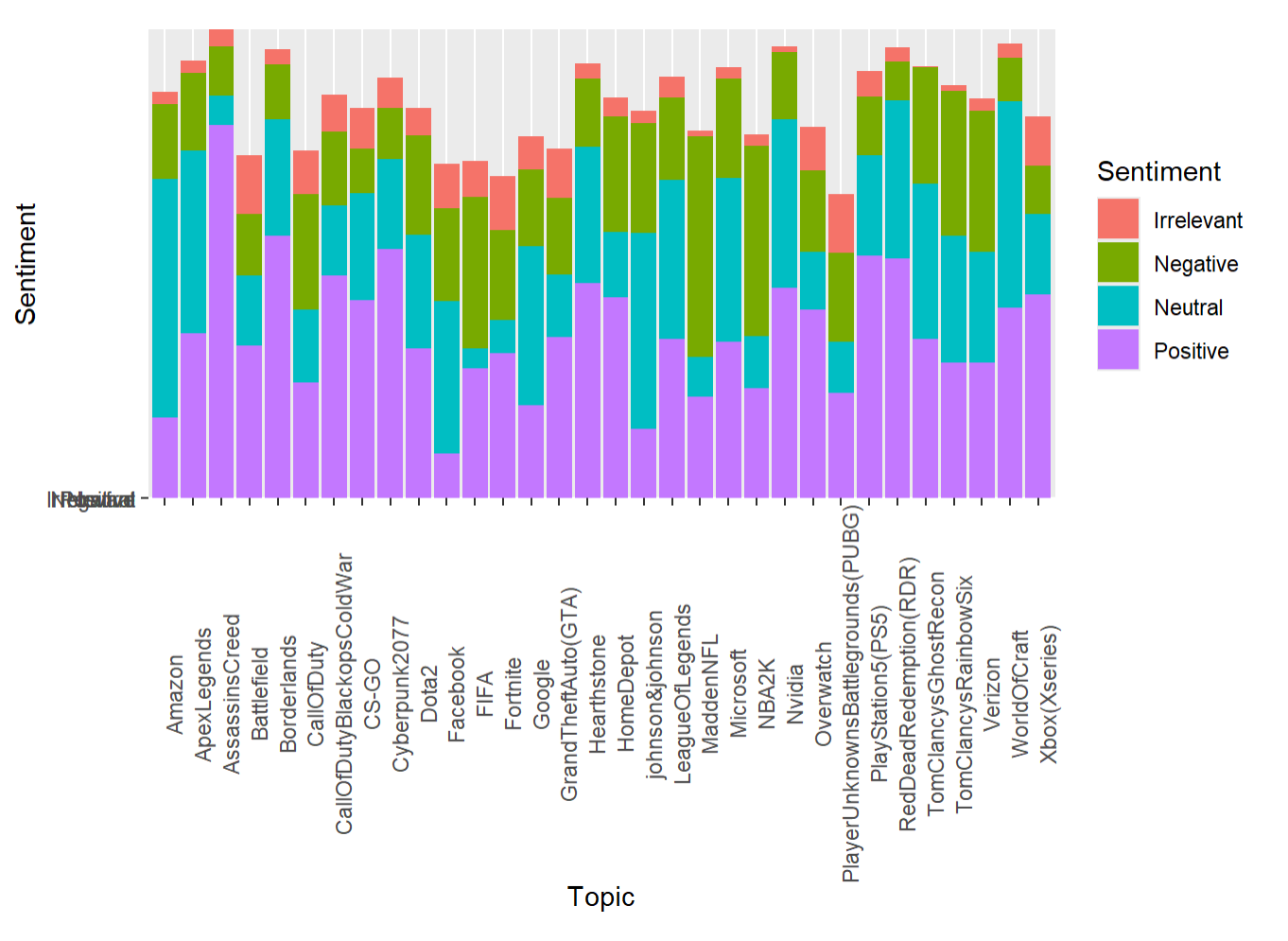


Donut chart for Sentiment (Positive,Neutral,Negative,Irrelevant)



Bar chart for different levels of sentiment (Positive, Neutral, Negative, Irrelevant) vs. different topics

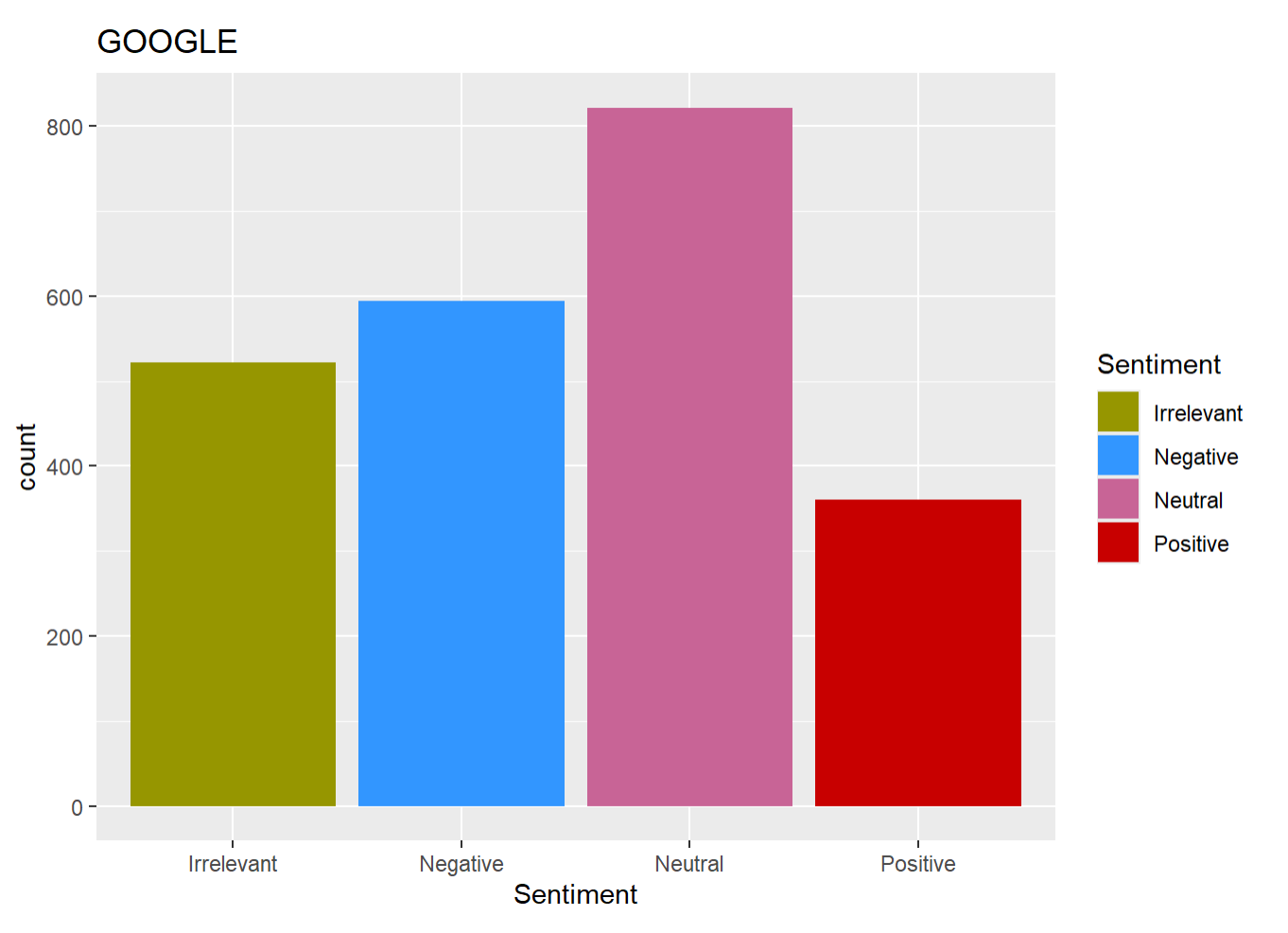
```
Warning in geom_bar(aes(fill = Sentiment), stat = "identity", positive =
"dodge"): Ignoring unknown parameters: `positive`
```

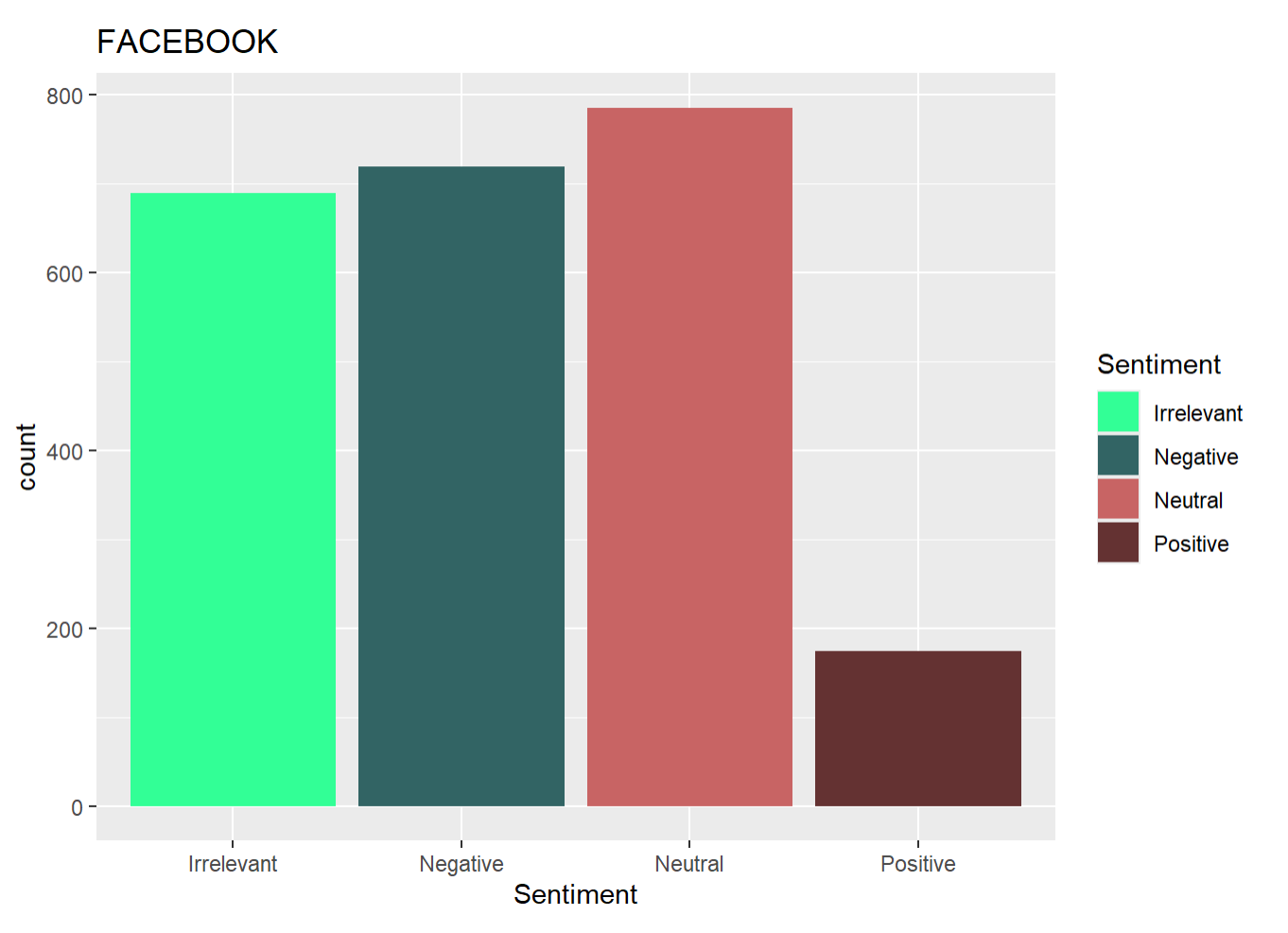


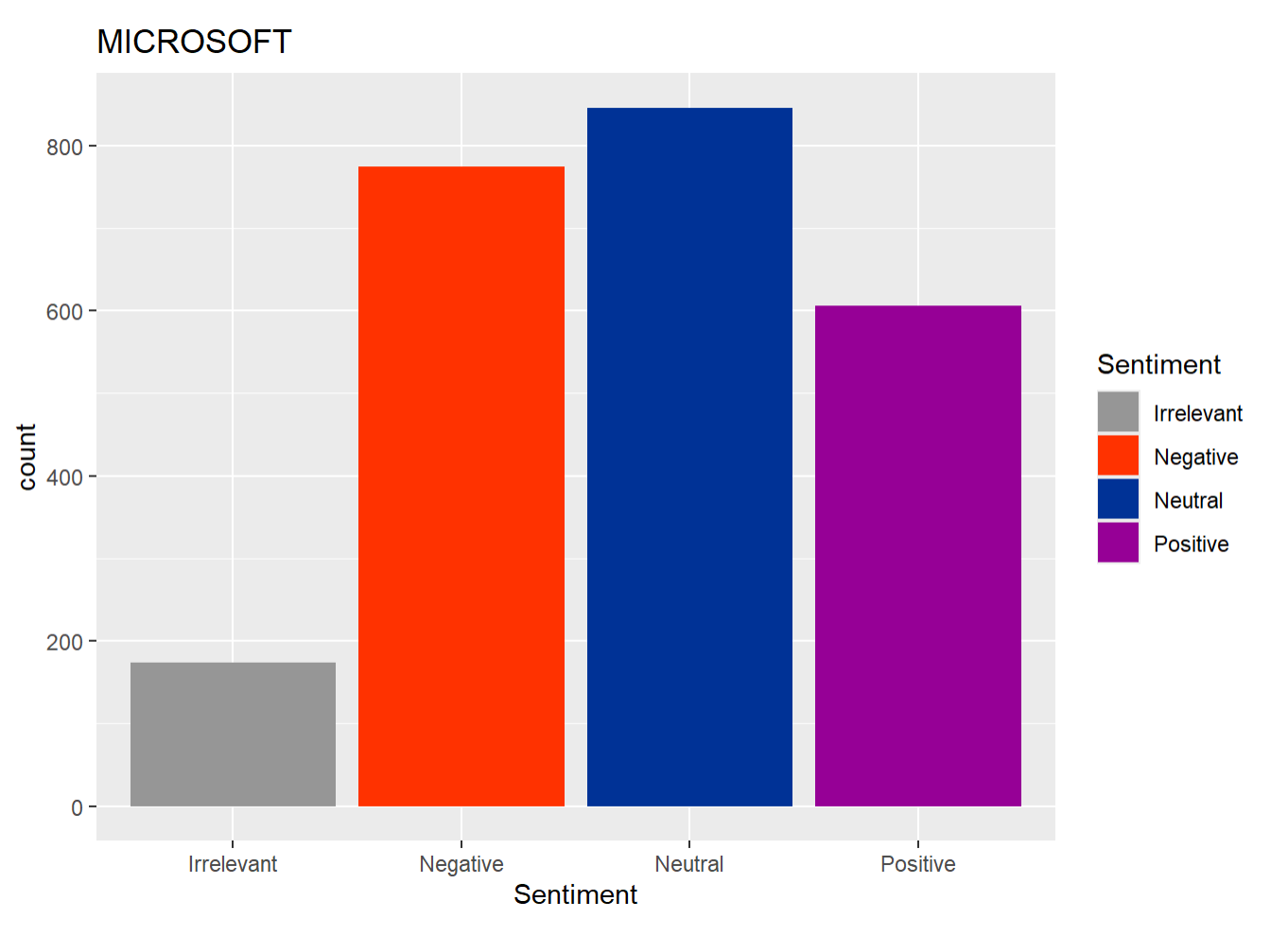
We list the Twitter topics in decreasing order of their trending popularity.

```
                    MaddenNFL                          Microsoft 
                         2400                               2400 
           TomClancysRainbowSix                         CallOfDuty 
                         2400                               2394 
              LeagueOfLegends                            Verizon 
                         2394                               2382 
                 ApexLegends          CallOfDutyBlackopsColdWar 
                         2376                               2376 
                     Facebook                              Dota2 
                         2370                               2364 
                  WorldOfCraft                              NBA2K 
                         2364                               2352 
                  Battlefield            TomClancysGhostRecon 
                         2346                               2346 
                         FIFA                          Overwatch 
                         2340                               2334 
                 Xbox(Xseries)                   johnson&johnson 
                         2334                               2328 
                        Amazon                          HomeDepot 
                         2316                               2310 
              PlayStation5(PS5)                              CS-GO 
                         2310                               2304 
                 Cyberpunk2077                GrandTheftAuto(GTA) 
                         2304                               2304 
                        Google                        Hearthstone 
                         2298                               2298 
                        Nvidia                        Borderlands 
                         2298                               2286 
         Fortnite PlayerUnknownsBattlegrounds(PUBG) 
                         2274                               2274 
       RedDeadRedemption(RDR)                     AssassinsCreed 
                         2262                               2244 
```

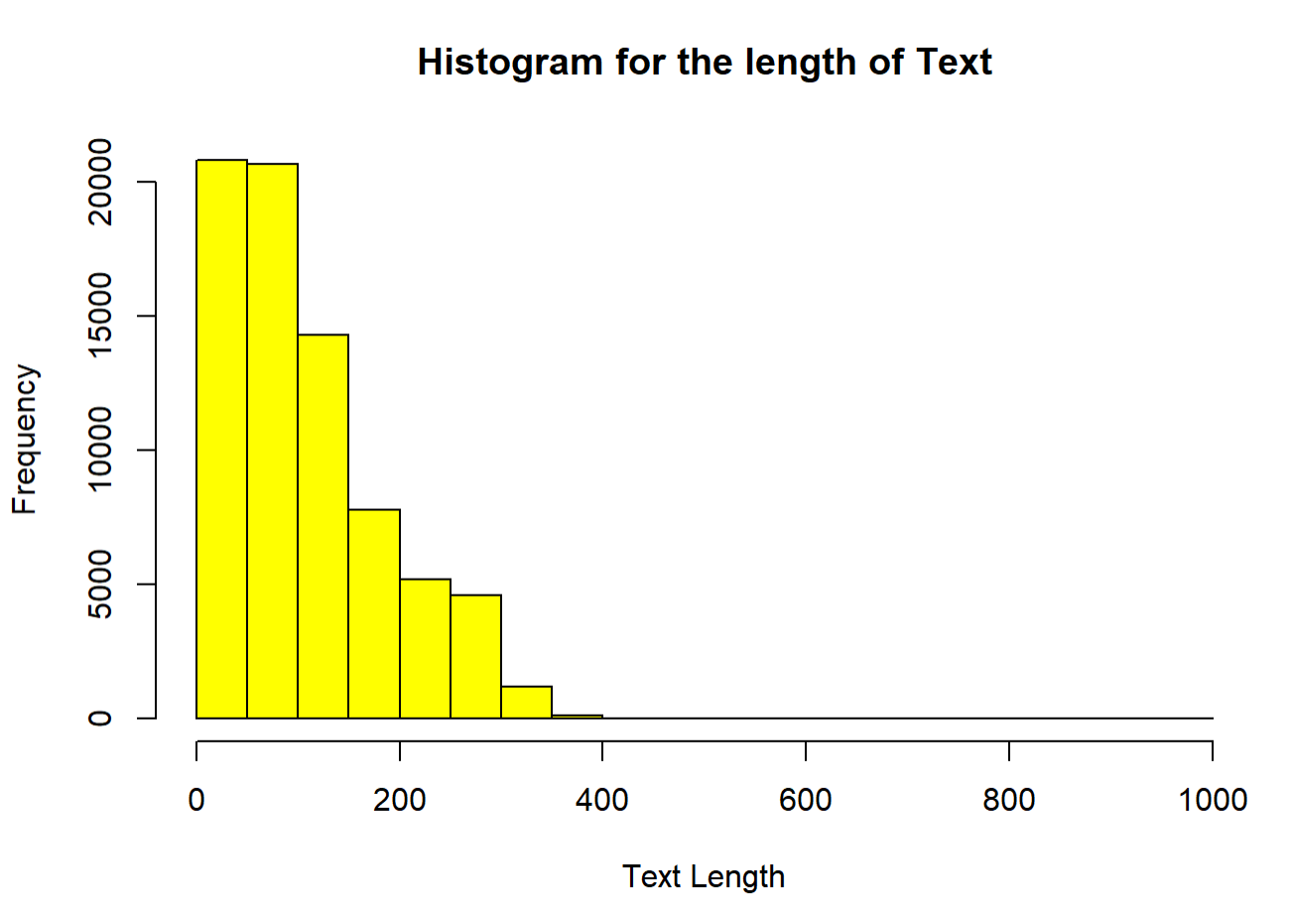
Sentiment Distribution for ‘Google’,‘Facebook’,‘Microsoft’ in Twitter







Histogram of Text Length in Tweets



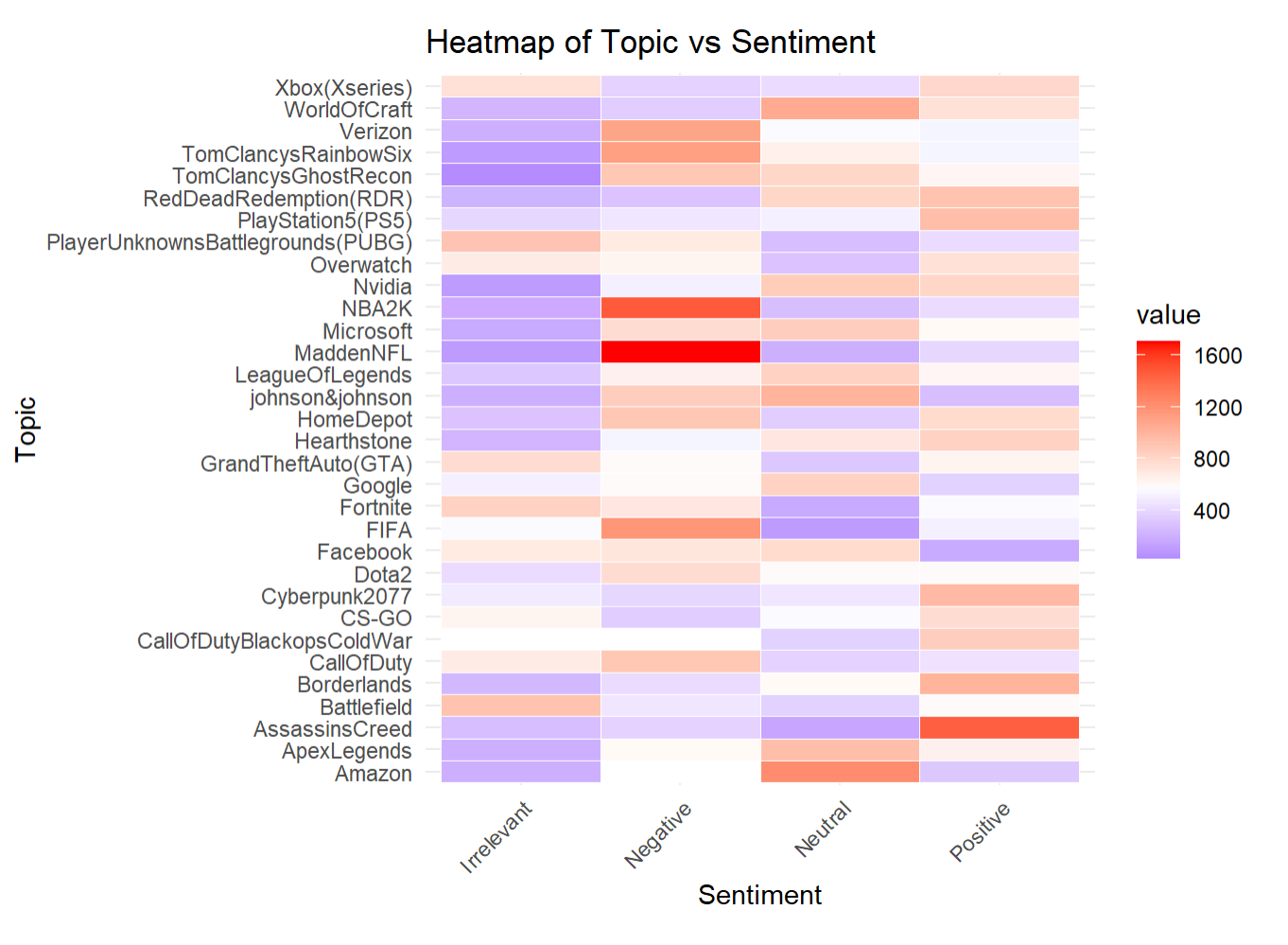
This histogram visualizes the distribution of tweet lengths based on the number of characters.

The distribution is right-skewed, with the majority of tweets having shorter text lengths.

The highest frequency of tweets falls within the 0 to 100 character range, with over 20,000 tweets in this interval. This indicates that most tweets are concise.

As text length increases, the frequency of tweets decreases sharply. Very few tweets exceed 300 characters, and tweets with lengths approaching the maximum of 1,000 characters are extremely rare.

This heatmap visualizes the relationship between sentiment (Positive, Neutral, Negative, Irrelevant) and various topics discussed on Twitter. Each cell in the heatmap represents the intensity or frequency of a particular sentiment associated with a specific topic, allowing for a quick identification of patterns and trends across different topics.



Word Cloud

In a word cloud, the size of each word indicates its frequency or importance—the larger the word, the more frequently it appears in the text.

```
Loading required package: RColorBrewer
```

```
Loading required package: NLP
```

```

Attaching package: 'NLP'
```

```
The following object is masked from 'package:ggplot2':

    annotate
```

Dominant Words: The largest words like “game,” “just,” “like,” “will,” and “good” are the most frequently mentioned in the dataset. This suggests that the tweets may be heavily focused on gaming-related discussions.

Sentiment and Topics: Words like “good,” “love,” and “great” suggest positive sentiment, while words like “fix,” “shit,” and “fucking” might indicate negative sentiment or frustration. The word “game” is central, which could imply that the primary topic of discussion is gaming.

Trends: The variety of words related to gaming, companies (e.g., “Verizon,” “Google,” “Amazon”), and social media engagement (e.g., “Facebook,” “Twitter”) indicate the topics that are trending or commonly discussed in the dataset.