Sentiment Analysis in R

॥वसुधैव कुटुम्बकम्॥

SYMBIOSIS
INSTITUTE OF TECHNOLOGY

Group no. - 12

Aditya Kuleyar	20070123003
Aditi Saxena	20070123004
Rewaa Mital	20070122110

Abstract:

In the current times, critic reviews of pc video games like fall guys, valorant,etc are aggregated by websites like metacritic. These reviews are a way for the players or users to share their video game experience, which in turn is useful for the company to make future updates in the games and also for other users who are interested in playing the game. Classifying reviews into positive and negative sentiment is necessary to derive useful insights from a big set of reviews. A computational study called sentiment analysis is used to extract subjective information from texts. We have performed sentiment analysis on the critic reviews of the pc game, Fall Guys, found on the website Metacritic.

Introduction:

Fall Guys is a free-to-play platform battle royale game developed by Mediatonic. The game involves up to 60 players who control jellybean-like creatures and compete against each other in a series of randomly selected mini-games, such as obstacle courses or tag. Currently this game has more than 50 million active users.

Fall guys has often been compared to many other pc and mobile games. Like any other video game, the reviews have been at variance with each other. However what matters is that through analysis of these reviews the company and users can compare this game with its competitors. The user will be able to gauge whether playing this game will be worth his while or not,what level of joy are people getting after playing this game and how likable it is by other players.

Whereas the company will use these reviews to its advantage to make the game better by working on the negative reviews and increasing the positive reviews. Consequently as the game will start getting more positive reviews it will start gaining more popularity. So, basically the critic reviews will act as a motivation for many other users to try out the game with the correct information in hand.

We are using the website Metacritic to find these reviews.Metacritic is regarded as the foremost online review aggregation site for the video game industry.As there are multiple reviews on the website, in order to make better decisions and act swiftly, it is necessary to categorise these reviews as good or negative based on sentiment. By categorising evaluations according to their mood, a firm and its users can more effectively assess both good and negative critic feedback and make decisions that better suit their needs.
In this study, unstructured data from Metacritic's reviews of the video game Fall Guys was taken. It has been pre-processed to gauge review sentiment and has been filtered to eliminate noisy data.

About the dataset:

This dataset was extracted from metacritic and captured all user reviews, from all languages, up to August 26th, 2020 from both pc and playstation 4 platforms.

The dataset contains metadata of the review (date, username, platform, amount of votes the review received, score) and the text of the review itself.

```
'data.frame':   353 obs. of  8 variables:
 $ username      : chr  "Locke_VI" "Crisener" "EnigmaXtreme" "nitorita" ...
 $ review_type   : chr  "user" "user" "user" "user" ...
 $ published_date: chr  "Aug 25, 2020" "Aug 25, 2020" "Aug 25, 2020" "Aug 23, 2020" ...
 $ score         : int  3 8 7 7 10 4 8 8 5 10 ...
 $ votes         : int  0 0 0 0 0 0 0 0 0 0 ...
 $ review_text   : chr  "Fall Guys was a game I was really looking forward to when it was first announced, and there is dece
nt potential"| __truncated__ "You gotta love it. The aesthetics are very colorful and charming, and the gameplay brings a lo
t of joy and fun."| __truncated__ "When we covered Fall Guys last week during it's beta, we talked about how fun the game wa
s with its crazy cours"| __truncated__ "I like how Fall Guys adapted Mario Party and made it support up to 60 players. It's
 a nice little casual game t"| __truncated__ ...
 $ profile_url   : chr  "https://www.metacritic.com/user/Locke_VI" "https://www.metacritic.com/user/Crisener" "https://www.m
etacritic.com/user/EnigmaXtreme" "https://www.metacritic.com/user/nitorita" ...
 $ platform      : chr  "pc" "pc" "pc" "pc" ...
```

Packages used:

tm:

Used for text-mining in R.

Useful for data import, corpus handling and pre processing.

Corpus represents a collection of text documents

Snowballc:
An R interface to the C 'libstemmer' library that implements Porter's word stemming algorithm for collapsing words to a common root to aid comparison of vocabulary.
wordcloud:
Used to get a graphic representation of the most frequent words in the reviews.
ColorBrewer:
R package that contains a ready-to-use color palettes for creating beautiful graphics.
syuzhet:
The package comes with four sentiment dictionaries and provides a method for accessing the robust, but computationally expensive, sentiment extraction tool developed in the NLP group at Stanford
ggplot2:
Grammar of Graphics is a free, open-source, and easy-to-use visualization package widely used in R

Pre-processing of the data:

After reading the csv file and storing it in a corpus, we now get rid of the unnecessary information by pre-processing our data.

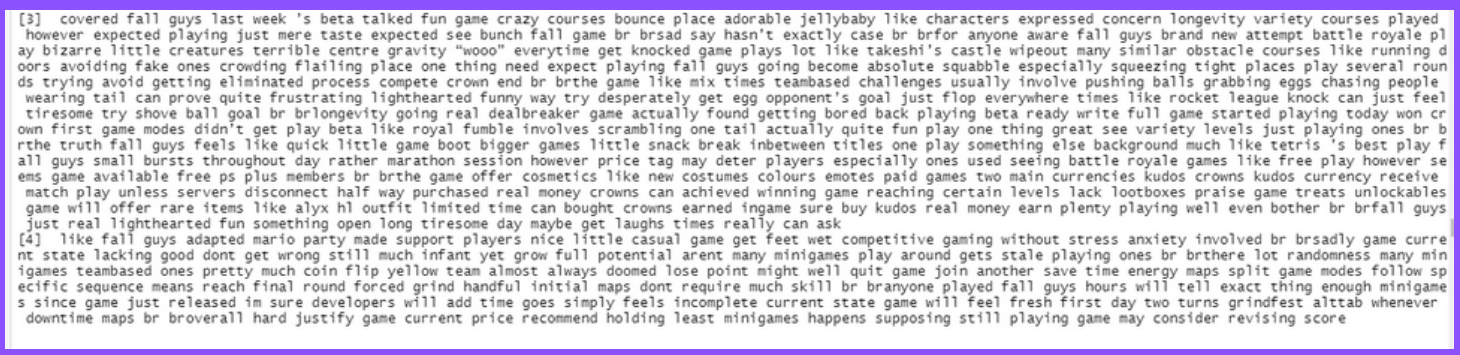
So, we begin by converting the entire text to lower case for uniformity of data to work with.

Then we remove the punctuations and whitespaces i.e., extra spaces as they are not really necessary.

We also remove numbers and stopwords from our data to reduce the size of our data by keeping what's useful.

We also remove words 'this' ,'interesting' and 'life' from the corpus later on, as they are not useful.

After pre-processing, the data in the corpus looks like:



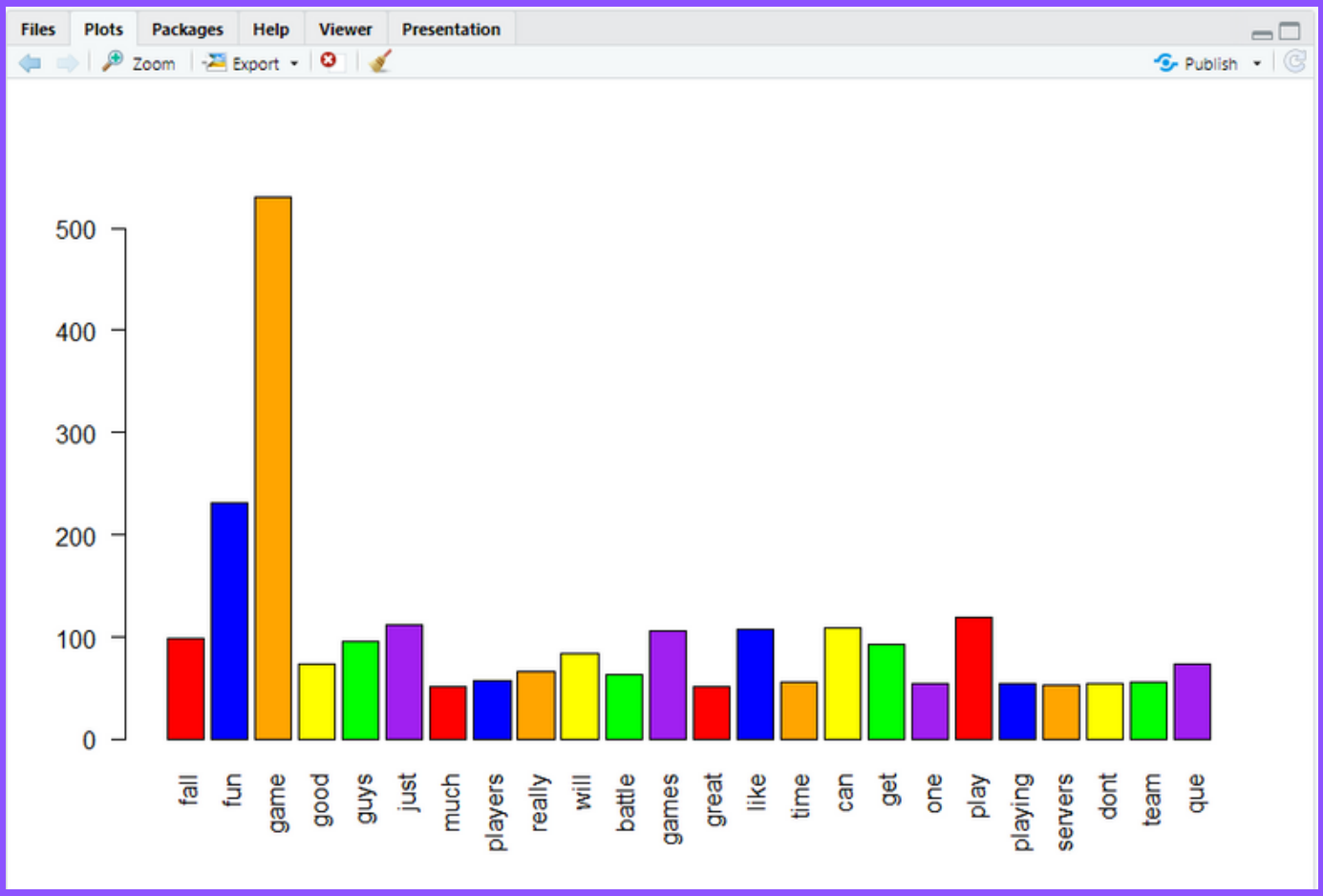
```
[3]  covered fall guys last week 's beta talked fun game crazy courses bounce place adorable jellybaby like characters expressed concern longevity variety courses played
 however expected playing just mere taste expected see bunch fall game br brsad say hasn't exactly case br brfor anyone aware fall guys brand new attempt battle royale pl
ay bizarre little creatures terrible centre gravity "wooo" everytime get knocked game plays lot like takeshi's castle wipeout many similar obstacle courses like running d
oors avoiding fake ones crowding flailing place one thing need expect playing fall guys going become absolute squabble especially squeezing tight places play several roun
ds trying avoid getting eliminated process compete crown end br brthe game like mix times teambased challenges usually involve pushing balls grabbing eggs chasing people
 wearing tail can prove quite frustrating lighthearted funny way try desperately get egg opponent's goal just flop everywhere times like rocket league knock can just feel
 tiresome try shove ball goal br brlongevity going real dealbreaker game actually found getting bored back playing beta ready write full game started playing today won cr
own first game modes didn't get play beta like royal fumble involves scrambling one tail actually quite fun play one thing great see variety levels just playing ones br b
rthe truth fall guys feels like quick little game boot bigger games little snack break inbetween titles one play something else background much like tetris 's best play f
all guys small bursts throughout day rather marathon session however price tag may deter players especially ones used seeing battle royale games like free play however se
ems game available free ps plus members br brthe game offer cosmetics like new costumes colours emotes paid games two main currencies kudos crowns kudos currency receive
 match play unless servers disconnect half way purchased real money crowns can achieved winning game reaching certain levels lack lootboxes praise game treats unlockables
 game will offer rare items like alyx hl outfit limited time can bought crowns earned ingame sure buy kudos real money earn plenty playing well even bother br brfall guys
 just real lighthearted fun something open long tiresome day maybe get laughs times really can ask
[4]  like fall guys adapted mario party made support players nice little casual game get feet wet competitive gaming without stress anxiety involved br brsadly game curre
nt state lacking good dont get wrong still much infant yet grow full potential arent many minigames play around gets stale playing ones br brthere lot randomness many min
igames teambased ones pretty much coin flip yellow team almost always doomed lose point might well quit game join another save time energy maps split game modes follow sp
ecific sequence means reach final round forced grind handful initial maps dont require much skill br branyone played fall guys hours will tell exact thing enough minigame
s since game just released im sure developers will add time goes simply feels incomplete current state game will feel fresh first day two turns grindfest alttab whenever
 downtime maps br broverall hard justify game current price recommend holding least minigames happens supposing still playing game may consider revising score
```

As we can see:

The entire text is in lower case.

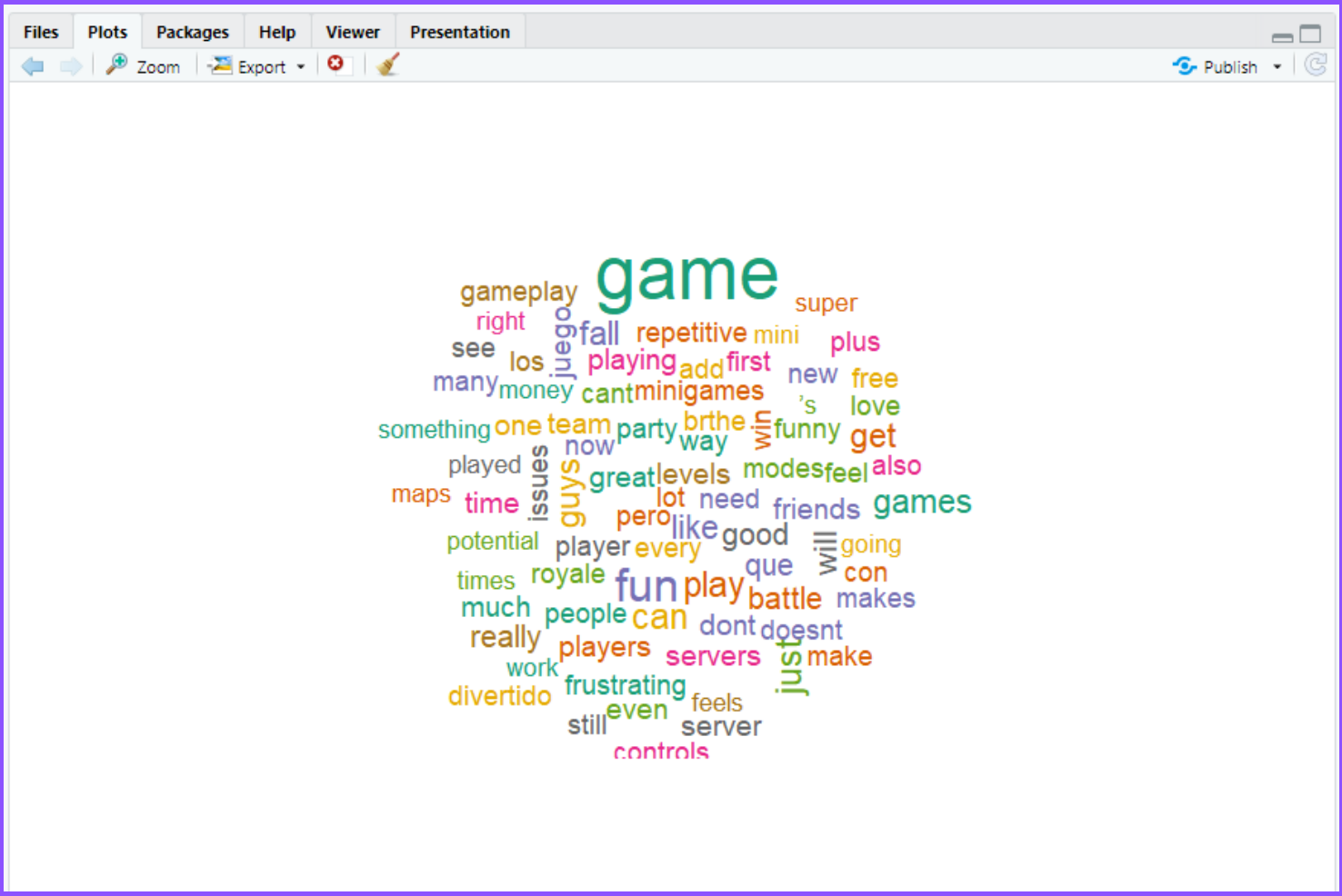
There are no punctuations, whitespaces, numbers or stopwords.

Plots:
Plotting the widely used terms in the corpus.



Observation: We can see that certain positive words like ‘game’, ‘fun’ ‘play’, ‘like’ are few of the words that are used widely in the corpus. Giving us an idea that the majority customer reviews about the product might be positive.

Creating a wordcloud to visualise the frequently used words in the corpus.



Observation: Words with bigger size mean they are used frequently in the corpus.
Yet again we get to see positive words in the wordcloud.
We can infer that the majority customer reviews about the product might be positive.

Extracting sentiments from the review:

Using the get_nrc_sentiment function of the syuzhet library we will extract the sentiments from each review and store them.

It looks somewhat like this:

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive	score
1	1	4	0	2	3	3	2	3	6	9	3
2	3	10	1	0	8	0	3	7	2	16	14
3	9	11	4	7	8	9	6	9	20	19	-1
4	2	7	1	5	6	5	4	6	12	9	-3
5	0	1	0	0	2	0	0	1	0	2	2
6	6	2	4	3	2	6	0	2	9	4	-5
7	2	2	1	2	1	1	1	2	2	3	1
8	0	2	0	0	1	2	0	1	2	2	0
9	3	2	1	1	2	1	1	0	3	3	0
10	0	0	0	0	0	0	0	0	0	0	0

We have shown the result summaries of all the sentiments

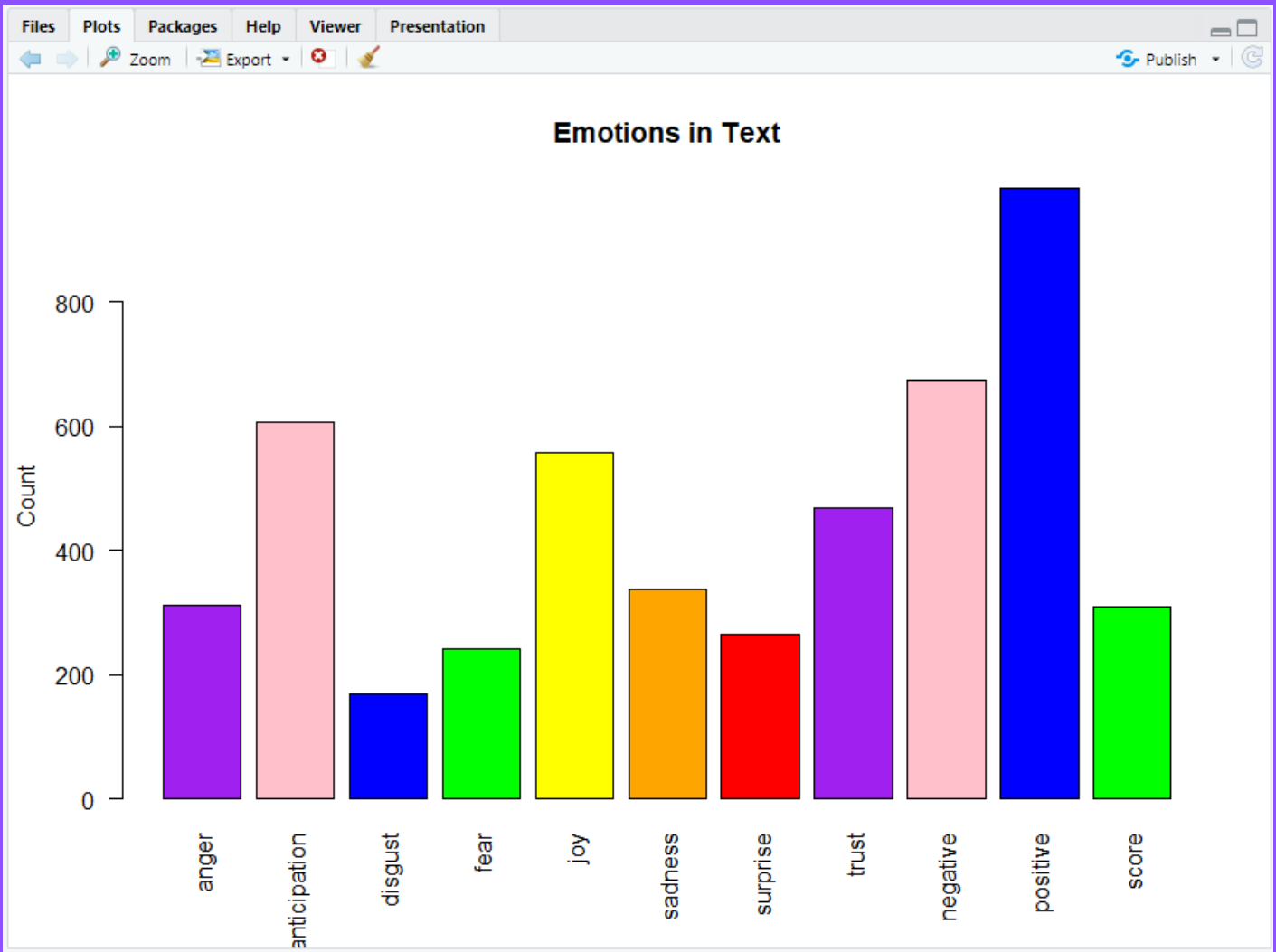
```
> summary(s)
     anger          anticipation      disgust            fear             joy            sadness      
 Min.   : 0.0000   Min.   : 0.000   Min.   :0.0000   Min.   : 0.0000   Min.   : 0.000   Min.   : 0.0000  
 1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.0000  
 Median : 0.0000   Median : 1.000   Median :0.0000   Median : 0.0000   Median : 1.000   Median : 0.0000  
 Mean   : 0.8839   Mean   : 1.717   Mean   :0.4788   Mean   : 0.6856   Mean   : 1.581   Mean   : 0.9547  
 3rd Qu.: 1.0000   3rd Qu.: 2.000   3rd Qu.:1.0000   3rd Qu.: 1.0000   3rd Qu.: 2.000   3rd Qu.: 1.0000  
 Max.   :14.0000   Max.   :16.000   Max.   :9.0000   Max.   :18.0000   Max.   :20.000   Max.   :22.0000  
    surprise          trust           negative        positive          score        
 Min.   : 0.0000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   :-6.0000  
 1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.000   1st Qu.: 0.0000  
 Median : 0.0000   Median : 0.000   Median : 1.000   Median : 1.000   Median : 0.0000  
 Mean   : 0.7479   Mean   : 1.326   Mean   : 1.907   Mean   : 2.782   Mean   : 0.8754  
 3rd Qu.: 1.0000   3rd Qu.: 2.000   3rd Qu.: 2.000   3rd Qu.: 3.000   3rd Qu.: 2.0000  
```

We would then calculate the final score of each sentiment.

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive	score
1	1	4	0	2	3	3	2	3	6	9	3
2	3	10	1	0	8	0	3	7	2	16	14
3	9	11	4	7	8	9	6	9	20	19	-1
4	2	7	1	5	6	5	4	6	12	9	-3
5	0	1	0	0	2	0	0	1	0	2	2
6	6	2	4	3	2	6	0	2	9	4	-5
7	2	2	1	2	1	1	1	2	2	3	1
8	0	2	0	0	1	2	0	1	2	2	0
9	3	2	1	1	2	1	1	0	3	3	0
10	0	0	0	0	0	0	0	0	0	0	0

```
> print(review_score)
      anger anticipation      disgust         fear          joy      sadness     surprise        trust 
        312          606          169          242          558          337          264          468 
   negative     positive        score 
        673          982          309 
```

Plotting sentiments:



Observation: It is evident from this graph that the positive reviews are much more than the negative ones.

Conclusion:

With new PC games being released in the market on a weekly basis, a large customer base relies on these critic reviews to base their opinion about a game.Consumer influence and trust-building are now possible thanks to online critiques.It becomes necessary to manage a huge amount of reviews and show the consumer a summary of them so they can get a better idea what all of those reviews were trying to convey. We are using sentiment analysis to help the company and other users of the game Fall Guys to make better decisions and classifying the feedback into positive and negative sentiment for a easier interpretation of the data. After performing the analysis we drew the conclusion that the critic reviews for the game are mostly positive indicating that the users are happy with game and are keen on playing it more.

Source code:

```
install.packages("tm")

install.packages("SnowballC") # for text stemming
install.packages("wordcloud") # word-cloud plot generator
install.packages("RColorBrewer") # for color palettes used in various plots
install.packages("syuzhet") # for sentiment analysis
install.packages("ggplot2") # for plotting graphs

library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
library("syuzhet")
library("ggplot2")

reviews <-read.csv(file.choose(), header = T)
str(reviews)
corpus <- iconv(reviews$review_text)
corpus <- Corpus(VectorSource(corpus))
```

```
inspect(corpus[1:5])
corpus <- tm_map(corpus, tolower)
corpus <- tm_map(corpus,
removePunctuation)
corpus <- tm_map(corpus, removeNumbers)
corpus <- tm_map(corpus, removeWords,
stopwords("english"))
corpus <- tm_map(corpus, removeWords,
c("this","interesting", "life"))
corpus <- tm_map(corpus, stripWhitespace)
inspect(corpus[1:5])
reviews_final <- corpus

tdm <- TermDocumentMatrix(reviews_final)
tdm <- as.matrix(tdm)
tdm[1:2, 1:2]
head(tdm,5)

w <- rowSums(tdm)
w <- subset(w, w>=50)
barplot(w,las=2,col
=c("red","blue","orange","yellow","green","pur
ple"),
```

```
main ="",
      ylab = "")
w <-sort(rowSums(tdm), decreasing = T)
set.seed(5678)
wordcloud(words = names(w),
        freq = w,
        max.words= 80,
        random.order = T,
        min.freq = 5,
        colors =brewer.pal(8, "Dark2"),
        random.color = T,
        scale= c(3,0.5))
sentiment_data <-
iconv(reviews$review_text)
s <- get_nrc_sentiment(sentiment_data)
s[1:10,]
head(s)
summary(s)

#Calculate Review wise Score
s$score <- s$positive - s$negative
s[1:10,]
```

```
#Checking product sentiment

#checking overall sentiment of the
product
review_score <- colSums(s[,])
print(review_score)

#Bar Plot
barplot(colSums(s),
        las = 2,
        col =
c("purple","pink","blue","green","yellow
","orange","red"),

        ylab = 'Count',
        main = 'Emotions in Text')
```