Paramount Plus Tweet Analysis Project

By: Janise Hope

Introduction:

In my opinion, out of the ads shown at the SuperBowl, the ones for Paramount Plus were the best this year. I wanted to analyze tweets about Paramount Plus to get a sense of what people thought about the ads and the new streaming services.

Step 1: Gather Tweets

```
#Loading rtweet & dplyr libraries
library(rtweet)
library(dplyr)
library(stringr)
library(utils)
library(tidytext)
library(ggplot2)
library(wordcloud)



# Accessing my Twitter API which allows me to retrieve tweets
# key & secret are redacted for privacy
appname <- "MontyBlathers"
key <- "xxxxxxxxxxxxxxxx"
secret <- "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
twitter_token <- create_token(
  app = appname,
  consumer_key = key,
  consumer_secret = secret)

# Query for tweets containing "ParamountPlus", limiting the query to 10,000 tweets, and excluding retweets
paramount_df <- search_tweets("ParamountPlus", n = 10000, include_rts = FALSE)
```

The query returned 9,998 tweets that contained “ParamountPlus” along with other data such as screen name, source, and the time the tweet was created. These tweets were from February 8th – February 12th , 2021. Below is a snippet of the resulting data frame called paramount_df .

	user_id	status_id	created_at	screen_name	text	source	display_text_width	reply_to_status_id
1	x13175752726742671 37	x1360040899607228422	2021-02-12 01:39:53	uhgoose	thanks	Twitter for iPhone	6	x1360037972469
2	x13175752726742671 37	x1359910278314532866	2021-02-11 17:00:50	uhgoose	youre so funny laughing so hard	Twitter for iPhone	34	x1359909507040
3	x13175752726742671 37	x1360033272999337984	2021-02-12 01:09:35	uhgoose	do you like my fancam	Twitter for iPhone	21	x1359933010204
4	x4368356782	x1360040775279648769	2021-02-12 01:39:23	_marcusturner_	Can we just have a smash universe already plea...	Twitter Web App	248	x1359535790661
5	x71011111	x1360040699471798276	2021-02-12 01:39:05	WeaselRFU	We know youre on the adfree plan but what if shitlo...	Twitter for Android	105	x1360032848472
6	x13483441349275811 85	x1360039286918287363	2021-02-12 01:33:28	ChristoferRiley	streaming 3 mustwatch shows StarTrekDiscovery pi...	Twitter Web App	278	
7	x13038280646532382 78	x1360037972469497857	2021-02-12 01:28:15	Lemonade121314	Yes	Twitter for iPad	3	x1360033272999
8	x13038280646532382 78	x1359933010204844037	2021-02-11 18:31:10	Lemonade121314		Twitter for iPad	0	x1359900477941
9	x7589230495843860 52	x1360037601344770048	2021-02-12 01:26:47	technofizzz		Twitter for iPhone	0	x1359933010204
10	x13094644667199897 62	x1360037078378160129	2021-02-12 01:24:42	KyleMarston4	February 11 2021 lets forget about and the and m...	Twitter for iPhone	251	
11	x3215305726	x1360036881405104130	2021-02-12 01:23:55	PINKNOIZE	Girl cast me	Twitter for iPhone	12	
12	x12498031985813504 01	x1360035909614374915	2021-02-12 01:20:03	BrianZiman	Nothing says Welcome to like having the CBS All Ac...	Twitter Web App	245	
13	x13137392398598594 67	x1360035677296066564	2021-02-12 01:19:08	Bme3691		Twitter for iPad	23	
14	x19772031	x1360034155082481664	2021-02-12 01:13:05	nerdynattyg	So cbs all access account is commercial free But I ...	Twitter for Android	158	
15	x565858374	x1360033413974142978	2021-02-12 01:10:08	knickerson82	will your app be better once you switch cbsallacces...	Twitter for Android	75	
16	x7055668605410344 97	x1359364359542108160	2021-02-10 04:51:33	waytoomuchbeer	Yep	Twitter Web App	4	x1359363457213
17	x7055668605410344 97	x1360025113400287236	2021-02-12 00:37:09	waytoomuchbeer	Ive never said anything to them either	Twitter Web App	40	x1360024288724
18	x7055668605410344 97	x1360028125359263745	2021-02-12 00:49:07	waytoomuchbeer	Yeah Rapp seems like a nice person Cruz is not ...	Twitter Web App	63	x1360027128767
19	x7055668605410344 97	x1359285274375712769	2021-02-09 23:37:18	waytoomuchbeer	They seemed to have blocked me so I guess they...	Twitter Web App	79	x1358611413925
20	x7055668605410344 97	x1360026130065383528	2021-02-12 00:41:12	waytoomuchbeer	Yeah he did that to me too I was basically like t...	Twitter Web App	80	x1360025758568

Showing 1 to 22 of 9,998 entries, 90 total columns

Step 2: Clean the data

Before I can analyze the tweets, they must be processed. In this case, it means removing any characters that are unnecessary for analysis. In the following section of code I remove hyperlinks, mentions, punctuation & extra spaces. I also removed the terms “Paramount”, “paramount”, and “paramountplus” because those are likely to appear often in tweets.

```
#  I'm going to clean my data by removing characters that are uncessary for analysis
# The following lines of code will be using regex so I can locate and match text easily
# \S is a regex character that matches any non-white space character(i.e. everything not a tab, space, etc)
# * is a regex character that will continue to match after its found one character that is in the position of the *

#  I am removing hyperlinks.
paramount_df$text <- gsub("https\\S*", "", paramount_df$text)

#I am removing mentions( those begin with an @)
paramount_df$text <- gsub("@\\S*", "", paramount_df$text)

#I am removing ampersands
paramount_df$text <- gsub("amp", "", paramount_df$text)

#I am removing new lines & carriage returns
paramount_df$text <- gsub("[\r\n]", "", paramount_df$text)

# I am removing punctuation
paramount_df$text <- gsub("[[:punct:]]", "", paramount_df$text)

#I'm removing â€™ which is a single right quotation mark
paramount_df$text <- gsub("â€™", "", paramount_df$text)
paramount_df$text <- gsub("â€", "", paramount_df$text)
```

```
paramount_df$text <- (gsub("paramount", "", paramount_df$text))
paramount_df$text <- (gsub("paramountplus", "", paramount_df$text))
paramount_df$text <- (gsub("Paramount", "", paramount_df$text))
```

In the section of code below, I created a new object called paramount_tweets which contains all of the words from the tweets but each word is separated and forms a list. I then removed stop words from the tweets.

```
#I'm removing stop words from the tweets
paramount_tweets <- paramount_df %>%
                    select(text) %>%
                    unnest_tokens(word,text)
paramount_tweets <- paramount_tweets %>%
                    anti_join(stop_words)
```

Step 3: Create a bar chart of the most frequently used words

I created a bar chart of the most frequent words used in the tweets gathered but realized that “WynonnaEarp” was the most frequent one.

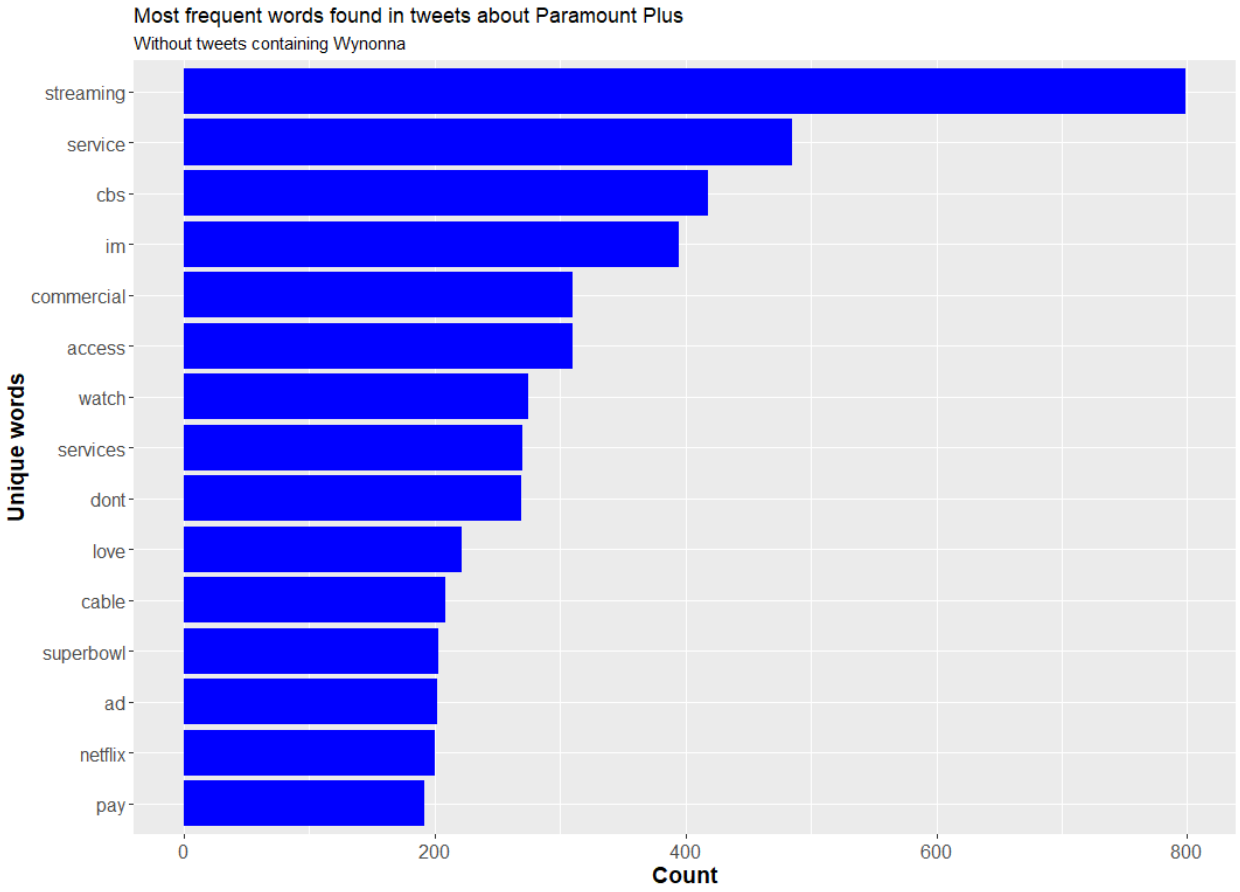
I looked up Wynonna Earp and found out that it is a show that was recently cancelled. Fans call themselves Earpers and they are mass tweeting any streaming platform to greenlight a 5th season for the show. They’ve also been using the hashtags such as #bringwynonnahome #fiveforwynonna #earpers.

I was curious to know how many people in my data frame were tweeting about Wynonna Earp. 2,318 tweets contained the term “Wynonna” and those tweets came from 634 different Twitter users.

```
# I'm creating a new dataframe that only has tweet data about tweets containing "Wynonna"

wynonna_tweets_data <- filter(paramount_df, grepl("Wynonna", text, fixed = TRUE))
# Has 2,318 tweets, I want to know how many users were tweeting to Paramount Plus about Wynonna
length(unique(wynonna_tweets_data$user_id))
# 634 users used were tweeting while"ParamountPlus" and trying to convince Paramount about Wynonna
```

In the section of code below, I created new data frame called JustParamount_tweets_data that only contained the tweets that didn’t mention Wynonna. This data frame has 7,680 tweets. I then created a bar chart of the most frequently used words in those tweets.

```r
#I'm creating a new dataframe that doesn't have tweet data about tweets containing "wynonna"
JustParamount_tweets_data <- filter(paramount_df, ! grepl("Wynonna", text, fixed = TRUE))

JustParamount_tweets <- JustParamount_tweets_data %>%
  select(text) %>%
  unnest_tokens(word,text)
JustParamount_tweets <- JustParamount_tweets %>%
  anti_join(stop_words)

#Now, I can make a bar chart of the most frequent words used in the Tweets for just Paramount tweets
JustParamount_tweets %>%
  count(word, sort = TRUE) %>%
  top_n(15) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(x = word, y = n)) +
  geom_col(fill = "blue") +
  xlab(NULL) +
  coord_flip() +
  labs(y = "Count",
       x = "Unique words",
       title = "Most frequent words found in tweets about Paramount Plus",
       subtitle = "Without tweets containing Wynonna")
```



Step 4: Word Cloud of Hashtags

I created a word cloud of hashtags to see which hashtags were commonly used in these tweets.

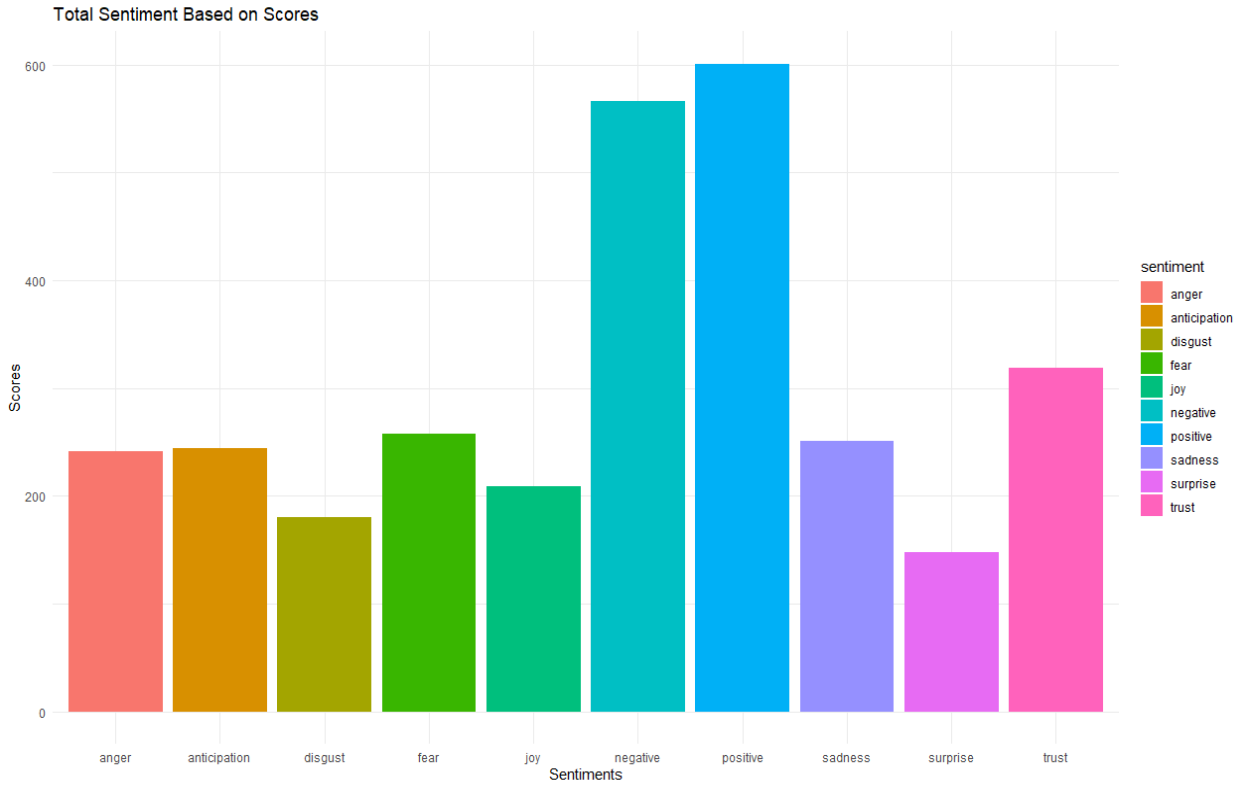
```
# Wordcloud of Hashtags
JustParamount_tweets_data$hashtags <- as.character(JustParamount_tweets_data$hashtags)
JustParamount_tweets_data$hashtags <- gsub("c\\(", "", JustParamount_tweets_data$hashtags)
set.seed(1234)
wordcloud(JustParamount_tweets_data$hashtags, min.freq=5,  scale=c(5, 1), random.order=FALSE, random.color = TRUE, rot.per=0.35,
          colors=brewer.pal(8, "Dark2"))
```

A lot of hashtags related to intellectual properties owned by ViacomCBS are mentioned like “Spongebob” and “StarTrek”. There are also some hashtags related to the superbowl which is when a series of Paramount Plus ads aired.

Step 5: Sentiment Analysis

Below is the sentiment analysis I performed on the tweets.

```
#I want to perform a sentiment analysis, so I'm loading the syuzhet package

library(syuzhet)

#So, I'm going to convert the tweets to ASCII
ASCII_tweets <- iconv(JustParamount_tweets, from= "UTF-8", to = "ASCII", sub="bytes")

# I'm creating the bar graph of the sentiment Analysis
ew_sentiment<-get_nrc_sentiment((ASCII_tweets))
sentimentscores<-data.frame(colSums(ew_sentiment[,]))
names(sentimentscores) <- "Score"
sentimentscores <- cbind("sentiment"=rownames(sentimentscores),sentimentscores)
rownames(sentimentscores) <- NULL
ggplot(data=sentimentscores,aes(x=sentiment,y=Score))+
  geom_bar(aes(fill=sentiment),stat = "identity")+
  theme(legend.position="none")+
  xlab("Sentiments")+ylab("Scores")+
  ggtitle("Total Sentiment Based on Scores")+
  theme_minimal()
```



There are a variety of sentiments present in the tweets. The strongest sentiment was positive with negative being a close second. In the next section, I'll investigate the differences between the positive tweets and negative tweets.

Step 6: Examining the content of Positive & Negative Tweets

I analyzed the sentiment of each tweet and created two data frames Positive_PeakPlus, for the positive tweets and Negative_PeakPlus for the negative tweets.

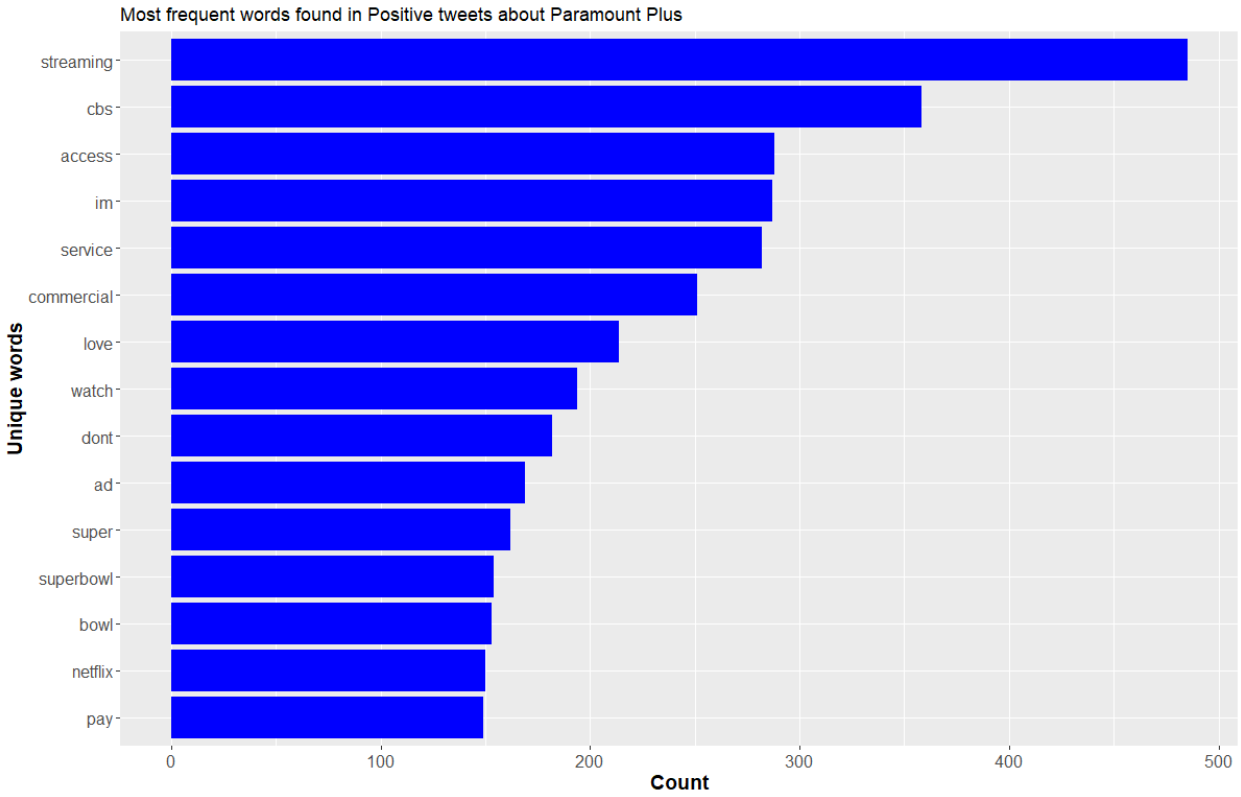
```
#Database PeakPlus has the following JustParamount_tweets_data variables: created_at, screen_name, text, & source
PeakPlus = select (JustParamount_tweets_data, 3:6)
```

```
library(SentimentAnalysis)

#Analyzing the sentiment of each tweet
Peak_sentiment <- analyzeSentiment(PeakPlus$text)
#Selecting the Sentiment word count, and the sentiement scores according to the following sentiment dictionaries:DictionaryGI,DictionaryHE, DictionaryLM & DictionaryQDAP
Peak_sentiment <- dplyr::select(Peak_sentiment,
                                WordCount, SentimentGI, SentimentHE, SentimentLM, SentimentQDAP)

# Getting the mean sentiment score for each Tweet
Peak_sentiment <- dplyr::mutate(Peak_sentiment,
                                mean_sentiment = rowMeans(Peak_sentiment[,-1]))

# Selecting only WordCount & mean_sentiment to be in the Peak_sentiment dataframe
Peak_sentiment <- dplyr::select(Peak_sentiment,
                                WordCount,
                                mean_sentiment)

# Adding the Peak_Sentiment dataframe to the PeakPlus dataframe
PeakPlus <- cbind.data.frame(PeakPlus, Peak_sentiment)


# Converting mean sentiment scores into "positive" or "negative"
PeakPlus$sentiment <- convertToBinaryResponse(PeakPlus$mean_sentiment)


#Gathering the postive tweets into a dataframe
Positive_PeakPlus <- filter(PeakPlus, grepl("positive", PeakPlus$sentiment, fixed = TRUE ))

#Gathering the negative tweets into a dataframe
Negative_PeakPlus <- filter(PeakPlus, grepl("negative", PeakPlus$sentiment, fixed = TRUE ) )
```

In this section of code, I created a bar chart of the most frequently used words in positive tweets.

.

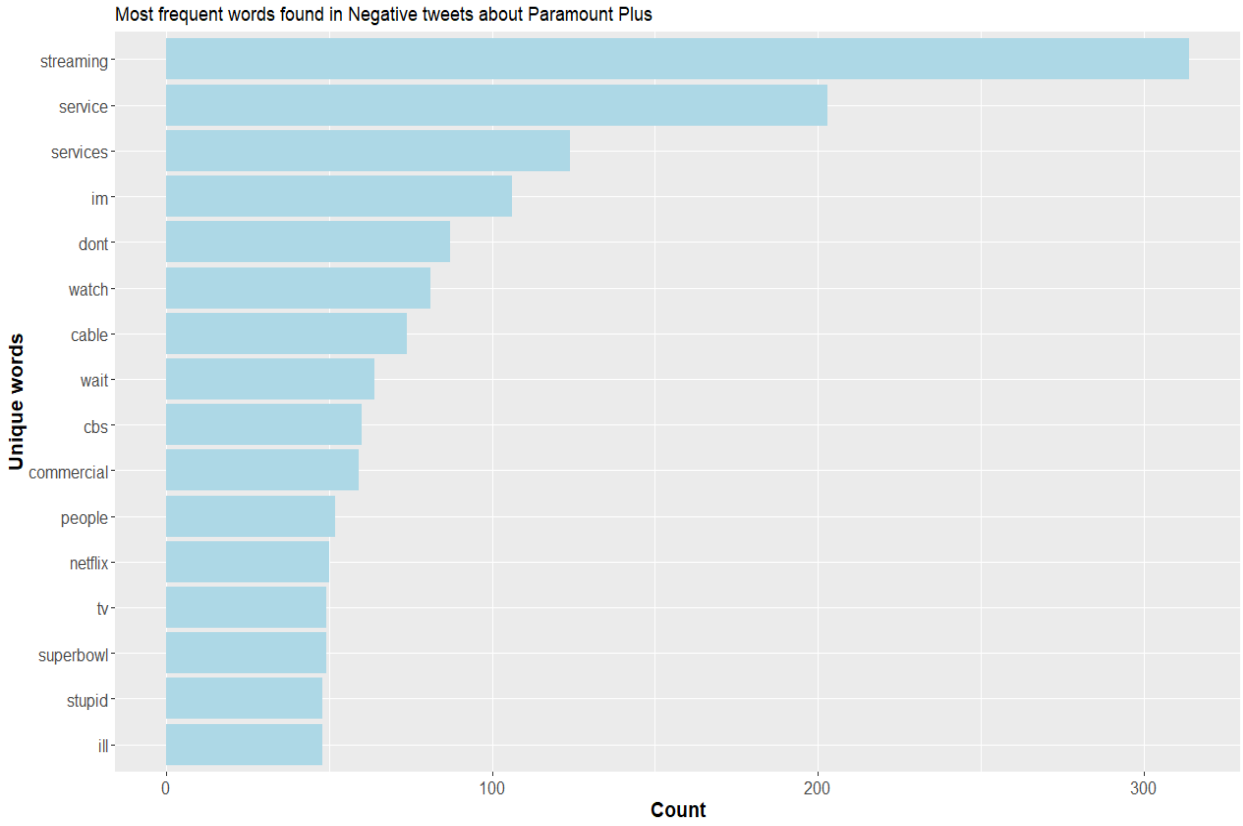
```
# Seperating tweets by indivdual words & then removing stop words
Positive_tweets <- Positive_PeakPlus %>%
  select(text) %>%
  unnest_tokens(word,text)
Positive_tweets <- Positive_tweets %>%
  anti_join(stop_words)


# Positive tweets most frequent words
Positive_tweets %>%
  count(word, sort = TRUE) %>%
  top_n(15) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(x = word, y = n)) +
  geom_col(fill = "blue") +
  theme(axis.text=element_text(size=12),
        axis.title=element_text(size=14,face="bold"))+
   xlab(NULL) +
  coord_flip() +
  labs(y = "Count",
       x = "Unique words",
       title = "Most frequent words found in Positive tweets about Paramount Plus"
      )
```



A lot of the positive tweets were related to which shows were streaming. It was either asking whether a show would be streamed or expressing excitement for a show that would be streaming. There was also some confusion on which shows will be streaming on CBS Access and which ones are on Paramount Plus. At that time, some people weren't aware that Paramount Plus was the new version of CBS All Access. In tweets in which the term love was mentioned, many Twitter users expressed either their love for Paramount Plus's Superbowl ads for the shows that were represented in the ads.

Below is the most frequent word bar chart for negative tweets and the code that created it.

```
# Doing the same to the negative tweets
Negative_tweets <- Negative_PeakPlus %>%
  select(text) %>%
  unnest_tokens(word,text)
Negative_tweets <- Negative_tweets %>%
  anti_join(stop_words)

Negative_tweets <- tokens(Negative_tweets)

Negative_tweets %>%
  count(word, sort = TRUE) %>%
  top_n(15) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(x = word, y = n)) +
  geom_col(fill = "light blue") +
  theme(axis.text=element_text(size=12),
        axis.title=element_text(size=14,face="bold"))+
  xlab(NULL) +
  coord_flip() +
  labs(y = "Count",
       x = "Unique words",
       title = "Most frequent words found in Negative tweets about Paramount Plus"
  )
```



In the negative tweets, there were a lot of users complaining about not wanting another streaming service. Many felt that with all of the new streaming services it would become just as expensive or even more than cable tv. Some users were angry about the possibility of Paramount's IPs being taken off of a different streaming service. There were others that pointed out issues they had with CBS All Access's interface and app that would make it difficult to watch shows. This shows that there are previous CBS All Access users who are concerned that the Paramount Plus app will have the same issues. In tweets that contained the word "stupid", users were either complaining about Paramount Plus's commercials or the oversaturation of the streaming market.

Main Takeaways

- Tweets about Paramount Plus are more positive than they are negative
- Positive tweets praise the ads as well as express excitement for shows streaming
- Negative tweets complain that there are two many streaming services some criticize the commercials
- Other negative tweets were about CBS All Access's app, so its important to find out why consumers believe it is difficult to use