Lab experiment - 7

Name: Prithviraj Guntha

Reg. No.: 20BRS1188

Subject: Essentials of data analytics

Subject code: CSE3506

Professor: A Sheik Abdullah

Slot: L55+L56

Perform sentiment analysis for a text dataset and also plot the wordcloud for the same.

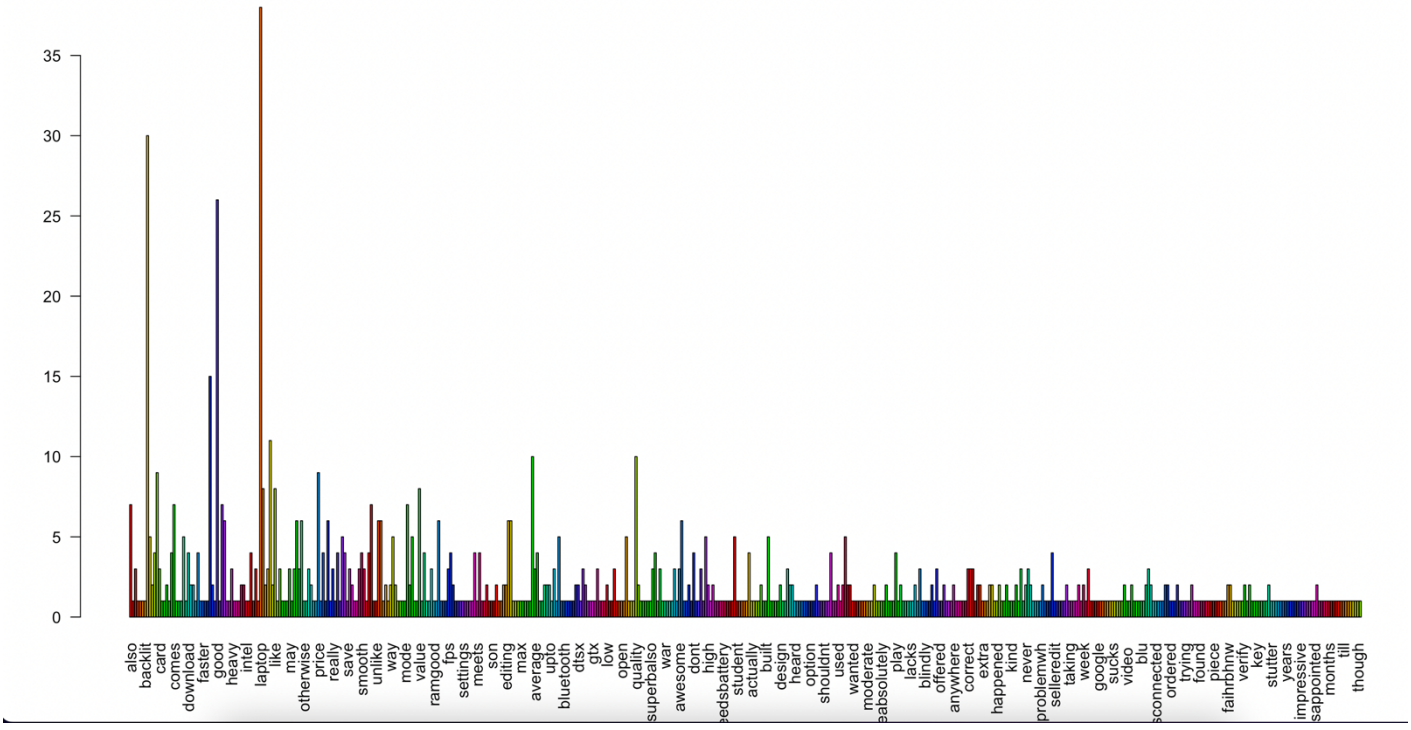
We will take a dataset that contains the reviews for a laptop from amazon and perform the sentiment analysis.

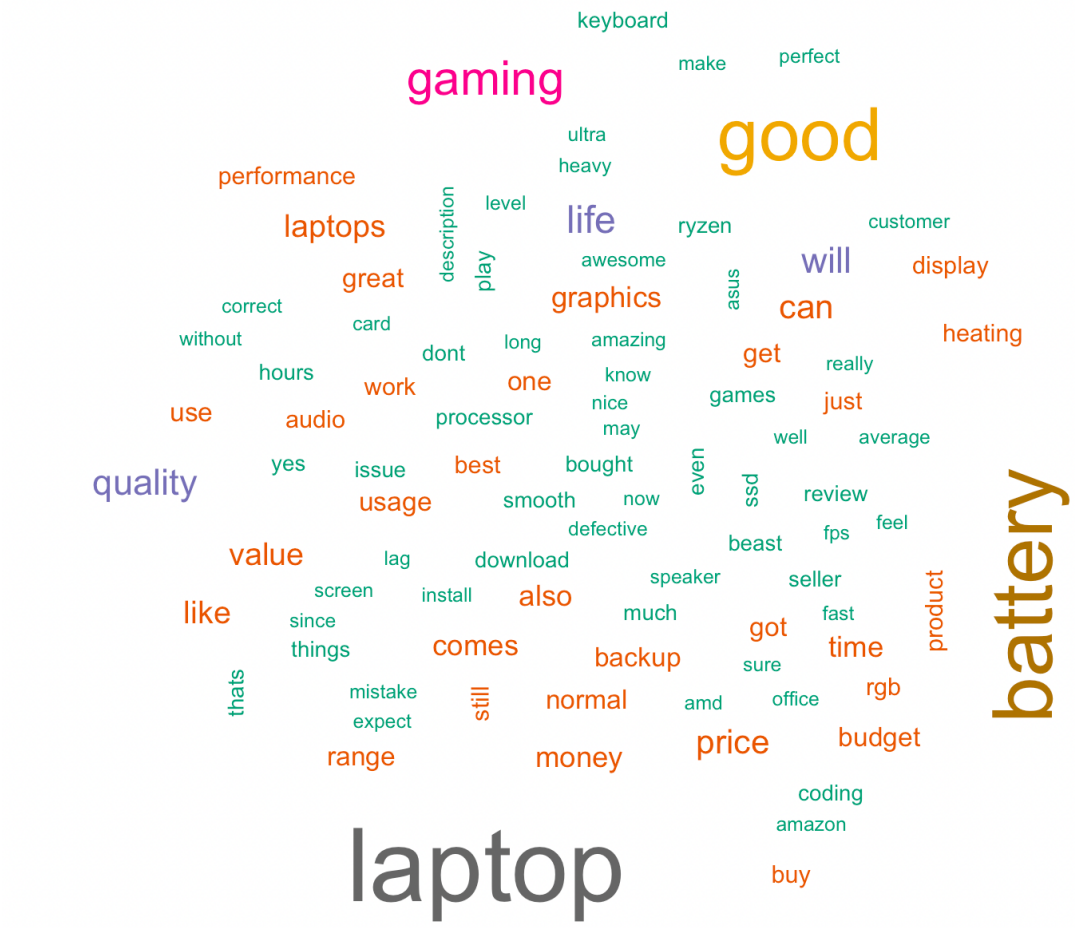
Code:

```
library(ggplot2)
library(lattice)
library(tm)
library(RColorBrewer)
library(wordcloud)
library(sentimentr)

data <-
read.csv("/Users/prithviraj/Downloads/AmazonReviews.csv")
data
str(data)

corpus <- iconv(data$Review.Content,to = "utf-8-mac")
corpus <- Corpus(VectorSource(corpus))
inspect(corpus[1:5])

corpus <- tm_map(corpus, tolower)
corpus <- tm_map(corpus, removePunctuation)
corpus <- tm_map(corpus, removeNumbers)
cleanset <- tm_map(corpus, removeWords, stopwords("english"))

inspect(cleanset[1:5])

removeslah <- function(x) gsub('\n','',x)
cleanset <- tm_map(cleanset,content_transformer(removeslah))
cleanset <- tm_map(cleanset, stripWhitespace)

tdm <- TermDocumentMatrix(cleanset)
tdm
```

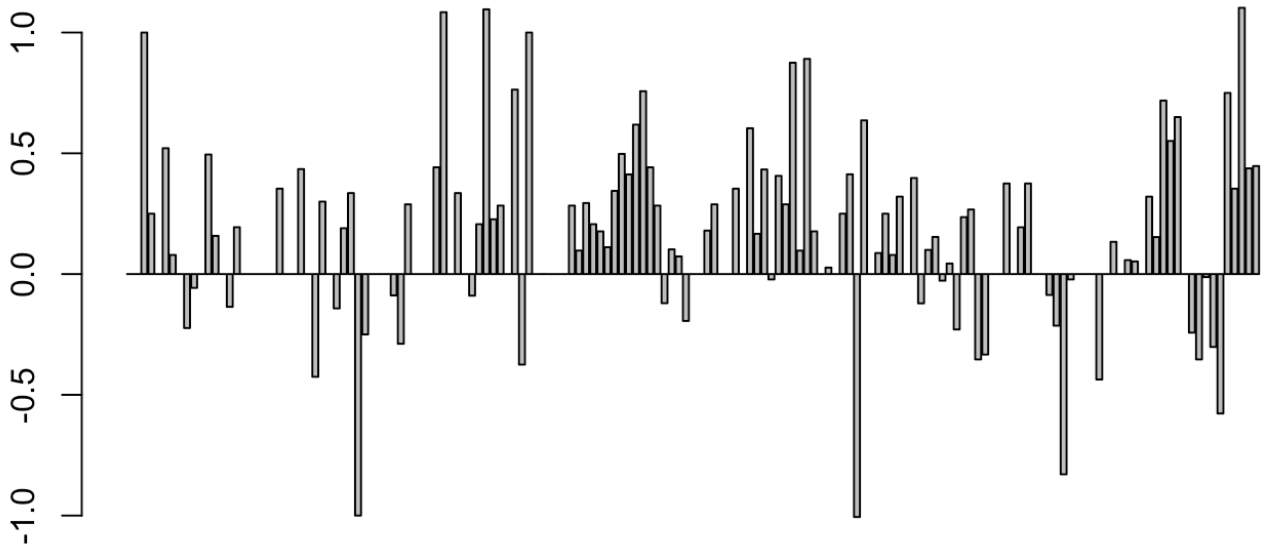
```
tdm <- as.matrix(tdm)
tdm[1:20,1:20]

w <- rowSums(tdm)

barplot(w,
        las = 2,
        col = rainbow(50))

wordcloud(words = names(w),
          freq = w,
          colors = brewer.pal(8,'Dark2'))

sent <- sentiment(data$Review.Content)
sent
barplot(sent$sentiment)
hist(sent$sentiment,col = "blue")
```

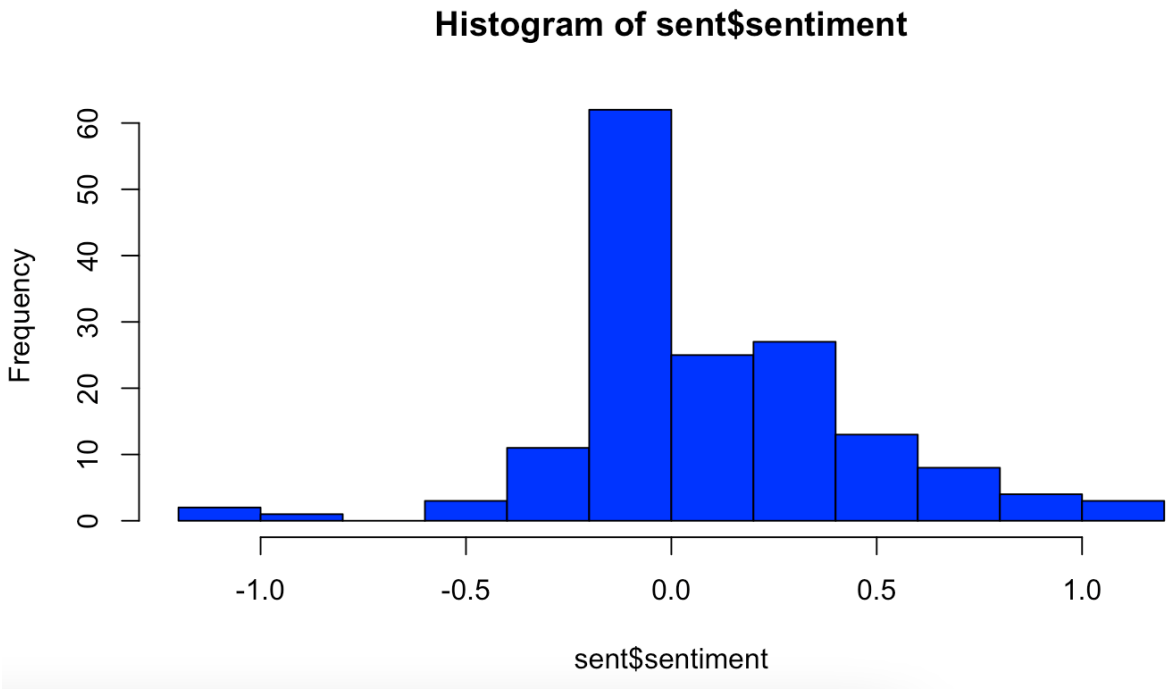
Outputs:






```
> sent <- sentiment(data$Review.Content)
> sent
     element_id sentence_id word_count sentiment
  1:          1           1          5 0.0000000
  2:          1           2         12 0.0000000
  3:          1           3          1 1.0000000
  4:          1           4          4 0.2500000
  5:          1           5         NA 0.0000000
 ---
155:         48           1          1 0.7500000
156:         49           1          2 0.3535534
157:         49           2          6 1.1022704
158:         49           3         16 0.4375000
159:         49           4          5 0.4472136
```





Inference:
After running the sentiment analysis, we have found that there are quite a few **negative** reviews for the laptop, majority are **neutral** reviews and quite a few **positive** reviews