Please click on the file to open.



Figure 1Anupam R assignment 

Code:

#library required for data cleaning and wordcloud plotting

library(tidytext)

library(dplyr)

library(reshape2)

library(janeaustenr)

library(ggplot2)

library(wordcloud)

#janeausten books grouped by book name

austen\_dataframe <- austen\_books() %>%

group\_by(book) %>%

mutate(linenumber = row\_number())

#filtering Pride & Prejudice book data from austen book dataframe

Pride\_dataframe <- austen\_dataframe %>% filter(book=="Pride & Prejudice")

#splitting text of book to token words.

tidy\_books <- Pride\_dataframe %>% unnest\_tokens(word, text)

#Removing stopwords

#-----------------------------#

custom\_stop\_words <- tribble(

# Column names should match stop\_words

~word, ~lexicon,

# Addmiss as custom stop words

"miss", "CUSTOM"

)

# Bind the custom stop words to stop\_words

stop\_words2 <- stop\_words %>%

bind\_rows(custom\_stop\_words)

#-----------------------------#

data("stop\_words")

tidy\_books <- tidy\_books %>%

anti\_join(stop\_words2)

#we used "bing" dictionary for getting positive and negative sentiments

tidy\_books %>%

inner\_join(get\_sentiments("bing")) %>%

count(word, sentiment, sort = TRUE) %>%

acast(word ~ sentiment, value.var = "n", fill = 0) %>%

comparison.cloud(title.colors=c("red","green"),colors = c("red", "lightgreen"),

max.words = 100)