Priya\_kumari\_lab7&8

* Examine the retail\_transaction\_survey\_data .csv in the D2L.

This contains data of retail transactions survey responses.

* The objective of this exercise is to create a word cloud of the data and the frequency count of the most commonly occurring words in this data set.
* Please use the in-class example WordCloudExample.R code and make appropriate changes to the template to generate this word cloud and the frequency of words.
* Based on the top 10 frequently occurring words, what do you think is the sentiment of this data set.
* Now change the parameter of the findFreqTerms command to find the terms that occur at least 100 times. How many words did you find that occurs at least 100 times in this data set.

library("tm")

library("SnowballC")

library("wordcloud")

library("RColorBrewer")

LOAD:

> filepath <- "retail\_transaction\_seurvey\_data.csv"

> text<-readLines(filepath)

> docs <- Corpus(VectorSource(text))

> docs <- Corpus(VectorSource(text))

> toSpace <-content\_transformer(function (x , pattern)gsub(pattern, " ", x))

> docs <- tm\_map(docs, toSpace, "/")

> docs <- tm\_map(docs, toSpace, "@")

> docs <- tm\_map(docs, toSpace, "[\\|](file://\\|\)")

Clean text:

> docs <- tm\_map(docs, content\_transformer(tolower))

> docs <- tm\_map(docs, removeNumbers)

> docs <- tm\_map(docs, removeWords,stopwords("english"))

> docs <- tm\_map(docs, removeWords, c("blabla1","blabla2"))

> docs <- tm\_map(docs, stripWhitespace)

> docs <- tm\_map(docs, stemDocument)

> set.seed(1234)

> wordcloud(words = d$word, freq = d$freq,c(8,.3),2,, min.freq = 1, max.words = 200, random.order = FALSE, rot.per = 0.35, colors=brewer.pal(8, "Dark2"))



|  |
| --- |
| barplot(d[1:10,]$freq, las = 2, names.arg = d[1:10,]$word,  + col ="orange", main ="Most frequent words",  + ylab = "Word frequencies")    > findFreqTerms(dtm, lowfreq=100)  [1] "servic" "help" "good" "phone" "sprint" "custom" "great" |
|  |
| |  | | --- | | Only 7 words above occur atleast 100 times. | |

> dtm <- TermDocumentMatrix(docs)

> m <- as.matrix(dtm)

> v <- sort(rowSums(m),decreasing = TRUE)

> d <- data.frame(word = names(v),freq=v)

> head(d, 10)

word freq

servic servic 341

custom custom 164

sprint sprint 161

great great 137

phone phone 136

good good 125

help help 115

store store 97

get get 58

time time 54

**Based on the most frequent occurring words, it looks like this text document is a customer service** (word service occurring 341 times and word custom probably stemmed from the word customer occurring 164 times) **review** ( words great, good and help clearly indicate a kind of review) **for sprint**(which occurs 161 times). **It is also clear from the word “phone” and “store” that it is store review of the phone service.**

**And the words “great”,” help” and “good” indicate that customer mostly have positive review on the service provided.**