**CODE USED FOR INSTALLING PACKAGES:**

# Install

install.packages("tm") # for text mining

install.packages("SnowballC") # for text stemming

install.packages("wordcloud") # word-cloud generator

install.packages("RColorBrewer") # color palettes

# Load

library("tm")

library("SnowballC")

library("wordcloud")

library("RColorBrewer")

text <- readLines(file.choose())

In the example below, I’ll load a .txt file hosted on STHDA website:

text <- readLines(filePath)

Load the data as a corpus

# Load the data as a corpus

docs <- Corpus(VectorSource(text))

VectorSource() function creates a corpus of character vectors

**Inspect the content of the document**

inspect(docs)

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

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

docs <- tm\_map(docs, toSpace, "\\|")

**Cleaning the text**

# Convert the text to lower case

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

# Remove numbers

docs <- tm\_map(docs, removeNumbers)

# Remove english common stopwords

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

# Remove your own stop word

# specify your stopwords as a character vector

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

# Remove punctuations

docs <- tm\_map(docs, removePunctuation)

# Eliminate extra white spaces

docs <- tm\_map(docs, stripWhitespace)

# Text stemming

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

**Build a term-document matrix**

dtm <- TermDocumentMatrix(docs)

m <- as.matrix(dtm)

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

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

head(d, 10)

**Generating Word cloud**

set.seed(1234)

wordcloud(words = d$word, freq = d$freq, min.freq = 1,

max.words=200, random.order=FALSE, rot.per=0.35,

colors=brewer.pal(8, "Dark2")