Corona NLP test

**The data file contains tweets that have been pulled from Twitter. In this dataset**

**use the text data in the "originalTweet" column and perform the following:**

**a) Convert the text corpus into tokens.**

**b) Perform stop word removal.**

**c) Count Word frequencies**

**d) Create word clouds.**

# Install and load required libraries

if (!require("tm")) install.packages("tm")

## Loading required package: tm

## Warning: package 'tm' was built under R version 4.2.3

## Loading required package: NLP

if (!require("wordcloud")) install.packages("wordcloud")

## Loading required package: wordcloud

## Warning: package 'wordcloud' was built under R version 4.2.3

## Loading required package: RColorBrewer

library(tm)  
library(wordcloud)

# Load data  
tweets <- read.csv("C:/Users/fvcds/Desktop/150$/Project/Corona\_NLP\_test.csv")

**Part a**

# Convert text to corpus  
corpus <- Corpus(VectorSource(tweets$OriginalTweet))  
  
# Clean corpus  
corpus <- tm\_map(corpus, tolower)

## Warning in tm\_map.SimpleCorpus(corpus, tolower): transformation drops documents

corpus <- tm\_map(corpus, removePunctuation)

## Warning in tm\_map.SimpleCorpus(corpus, removePunctuation): transformation drops  
## documents

corpus <- tm\_map(corpus, removeNumbers)

## Warning in tm\_map.SimpleCorpus(corpus, removeNumbers): transformation drops  
## documents

**Part b**

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

## Warning in tm\_map.SimpleCorpus(corpus, removeWords, stopwords("english")):  
## transformation drops documents

corpus <- tm\_map(corpus, stripWhitespace)

## Warning in tm\_map.SimpleCorpus(corpus, stripWhitespace): transformation drops  
## documents

# Convert corpus to document term matrix  
dtm <- DocumentTermMatrix(corpus)

**Part c**

# Calculate word frequencies  
word\_freq <- colSums(as.matrix(dtm))  
  
# Sort words by frequency  
sorted\_freq <- sort(word\_freq, decreasing = TRUE)

**Part d**

# Create word cloud  
wordcloud(names(sorted\_freq), freq = sorted\_freq, min.freq = 5, max.words = 100, random.order = FALSE, colors = brewer.pal(8, "Dark2"))

