

PRACTICAL-3

AIM:

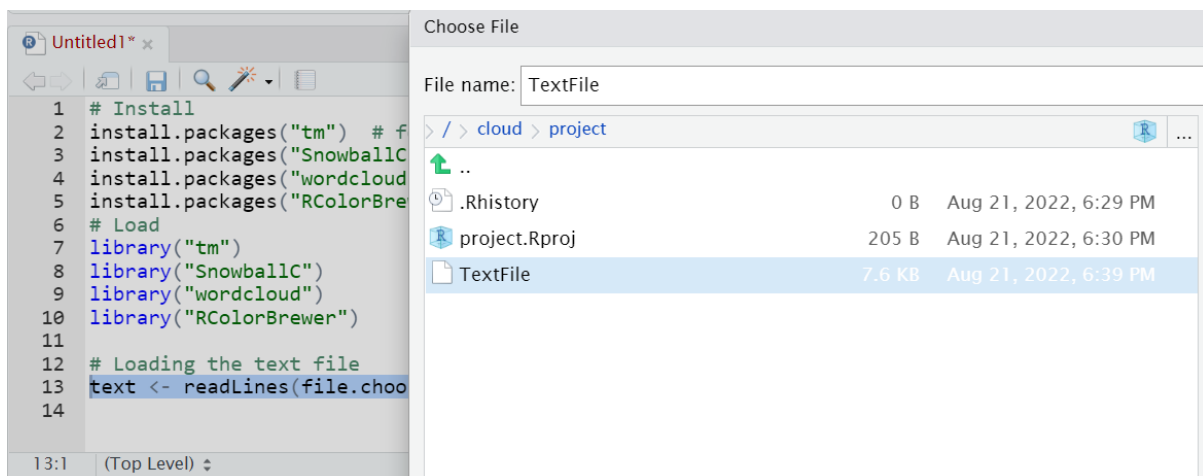
To implement basic functions and commands in R Programming. To build WordCloud, a text mining method using R for easy to understand and better visualization than a data table.

IMPLEMENTATION:

- Create a text file
- Install and load the required 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")
```

- load the text
- The text is loaded using **Corpus()** function from **text mining** (tm) package. Corpus is a list of a document (in our case, we only have one document).
- We start by importing the text file created.
- To import the file saved locally in your computer, type the following R code. You will be asked to choose the text file interactively.



Values	
text	chr [1:60] "Hello World, I am Parth." "..."

- Load the data as a corpus
- Inspect the content of the document
- Transformation is performed using `tm_map()` function to replace, for example, special characters from the text.
- Replacing “/”, “@” and “|” with space:
- the `tm_map()` function is used to remove unnecessary white space, to convert the text to lower case, to remove common stopwords like ‘the’, “we”.
- Another important preprocessing step is to make a text stemming which reduces words to their root form. In other words, this process removes suffixes from words to make it simple and to get the common origin. For example, a stemming process reduces the words “moving”, “moved” and “movement” to the root word, “move”.
- Build a term-document matrix
- Document matrix is a table containing the frequency of the words. Column names are words and row names are documents. The function `TermDocumentMatrix()` from **text mining** package can be used as follow.
- The importance of words can be illustrated as a word cloud as follow

PROGRAM CODE:

```
# 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")
```

```
# Loading the text file
```

```
text <- readLines(file.choose())
```

```
docs <- Corpus(VectorSource(text))
```

```
inspect(docs)
```

```
toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
```

```
docs <- tm_map(docs, toSpace, "/")

docs <- tm_map(docs, toSpace, "@")

docs <- tm_map(docs, toSpace, "\\")


# 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)
```

```

dtm <- TermDocumentMatrix(docs)

m <- as.matrix(dtm)

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

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

head(d, 10)

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"))

```

OUTPUT:

Data		
▶ d	165 obs. of 2 variables	
▶ docs	List of 60	
▶ dtm	List of 6	
m	num [1:165, 1:60] 1 1 1 0 0 0 0 0 0 0...	
Values		
text	chr [1:60] "Hello World, I am Parth." "..."	
toSpace	function (x, ...)	
v	Named num [1:165] 39 39 39 19 19 16 16 ...	



CONCLUSION:

By performing this practical, I learnt how to create a word cloud using R programming language.