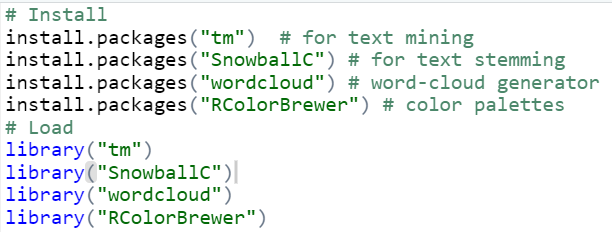
**PRACTICAL-3**

**AIM:**

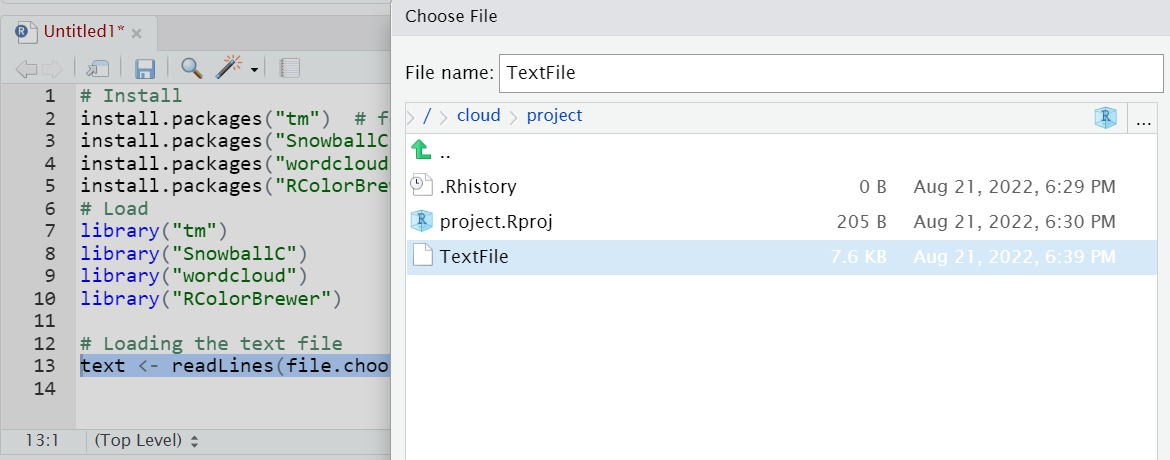
To implement basic functions and commands in R Programming. To build WordCloud, a text mining method using R for easy to understand and better visualization than a data table.

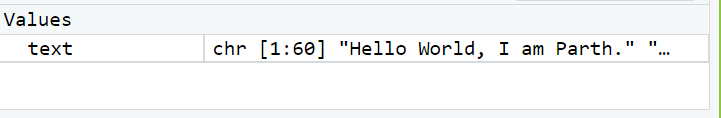
**IMPLEMENTATION:**

* Create a text file
* Install and load the required packages



* 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.





* 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)

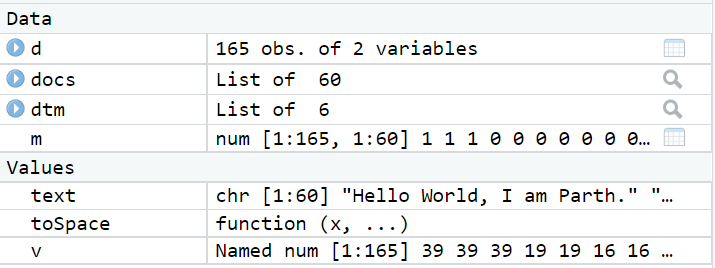
#### set.seed(1234)

#### 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:**





#### **CONCLUSION:**

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