INTERNAL PRACTICAL

AIM:

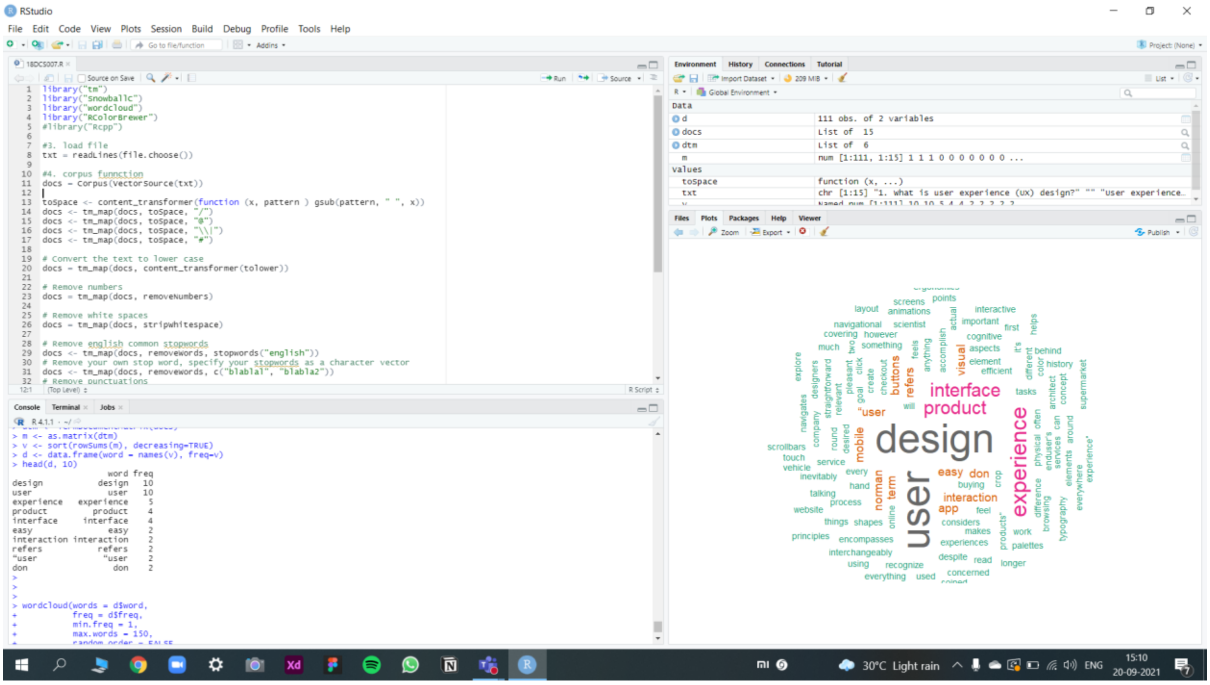
To implement basic functions and commands in R Programming. To build WordCloud, a text mining method using R for easy to understand and better visualization than a data table.

PRACTICAL:

```
library("tm")
library("SnowballC")
library("wordcloud")
library("RColorBrewer")
# library("Rcpp")
# load file
txt = readLines(file.choose())
# corpus funnction
docs = Corpus(VectorSource(txt))
toSpace <- content_transformer(function (x, pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
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 white spaces
docs = tm_map(docs, stripWhitespace)
# 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("blablal", "blabla2"))
# Remove punctuations
docs <- tm_map(docs, removePunctuation)
```

```
dtm <- TermDocumentMatrix(docs)

m <- as.matrix(dtm)

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

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

head(d, 10)


wordcloud(words = d$word,

          freq = d$freq,

          min.freq = 1,

          max.words = 150,

          random.order = FALSE,

          rot.per = 0.35,

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

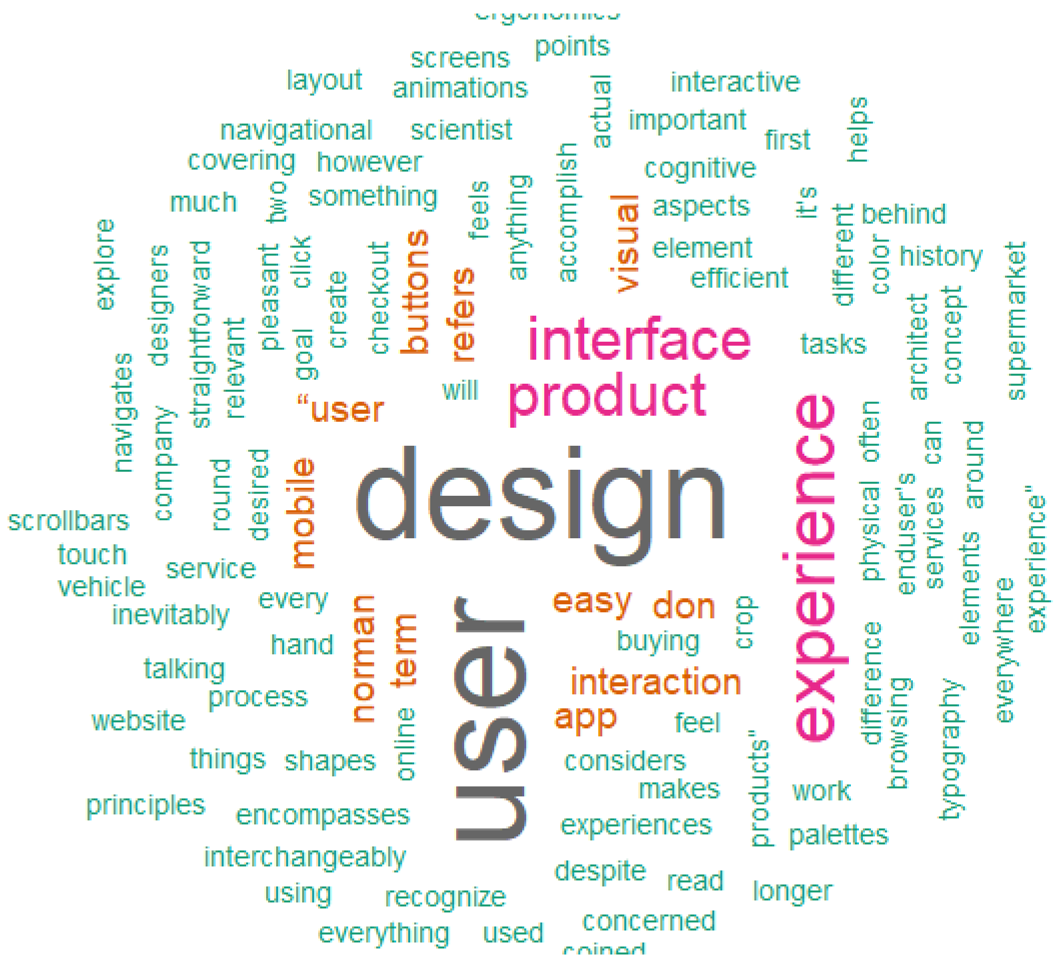
OUTPUT:

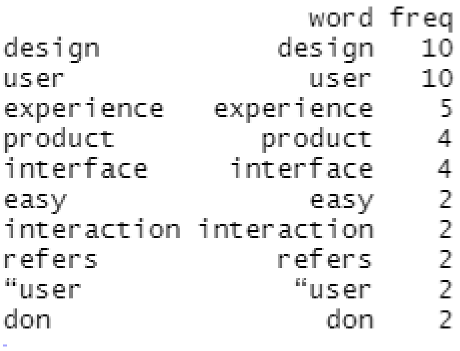


```
library("tm")
library("snowballc")
library("wordcloud")
library("RColorBrewer")
#library("Rcpp")

#3. load file
txt = readLines(file.choose())

#4. corpus funnction
docs = Corpus(VectorSource(txt))

toSpace <- content_transformer(function (x, pattern ) gsub(pattern, " ", x))
docs <- tm_map(docs, toSpace, "/")
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 white spaces
docs = tm_map(docs, stripwhitespace)

# 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
```

```
> m <- as.matrix(dtm)
> v <- sort(rowSums(m), decreasing=TRUE)
> d <- data.frame(word = names(v), freq=v)
> head(d, 10)
                   word freq
design           design   10
user               user   10
experience   experience    5
product         product    4
interface     interface    4
easy               easy    2
interaction interaction    2
refers           refers    2
“user             “user    2
don                 don    2
>
>
>
> wordcloud(words = d$word,
+           freq = d$freq,
+           min.freq = 1,
+           max.words = 150,
+           random.order = FALSE
```

```
Data
d        111 obs. of 2 variables
docs     List of  15
dtm      List of  6
m        num [1:111, 1:15] 1 1 1 0 0 0 0 0 0 0 ...
Values
toSpace  function (x, ...)
txt      chr [1:15] "1. what is user experience (ux) design?" "" "user experience…
```





```
                   word freq
design           design   10
user               user   10
experience   experience    5
product         product    4
interface     interface    4
easy               easy    2
interaction interaction    2
refers           refers    2
“user             “user    2
don                 don    2
```

CONCLUSION:

In this practical, we learnt about R and implemented Word Cloud using R.