4 Shiny Application

Exercise 4.1

1. Create a shiny application which plots the histogram of a set of samples. The $n = 100$ samples will be generated from a univariate normal distribution with mean $m = 0$ and standard deviation $\sigma = 1$. You must control the number of bins of the histogram with a slider starting to 1 until 50. The first bin starts with the minimum value of the samples and the last one with the largest value.
2. Add the theoretical pdf of the samples over the histogram.
3. Add a slider (starting from 10 to 10 000) to choose the number of generated samples.
4. Add an input numeric control to select the mean m between -10 and 10 with the step 0.5.
5. Add a list of choices to select whether you want only the histogram, only the theoretical pdf, or both (the default choice is both).

Exercise 4.2

1. Create a file "helpers.R" containing the function "connectTwitter <- function() " which establishes the connection with Twitter. You can load this function with "source("helpers.R")".
2. Explain each line of the following R code :

```
1  library(twitteR)
2  library(tm)
3  # It is assumed that a connection with Twitter is established.
4  processedWord <- "datascience"
5  nbTweets <- 200
6  tweets <- searchTwitter(processedWord,n=nbTweets)
7  tweetTxt <- sapply(tweets, function(x) x$getText())
8  tweetTxt <- sapply(tweetTxt,
9                     function(row) iconv(row,
10                                         from="latin1", to="ASCII",
11                                         sub="")
12                    )
13 myCorpus <- Corpus(VectorSource(tweetTxt))
14 tdm <- TermDocumentMatrix(myCorpus,
15                            control = list(removePunctuation = TRUE,
16                               stopwords = c(stopwords("english")),
17                               removeNumbers = TRUE,
18                               tolower = TRUE)
19                            )
20 m <- as.matrix(tdm)
21 word_freqs <- sort(rowSums(m), decreasing=TRUE)
22 dm <- data.frame(word=names(word_freqs), freq=word_freqs)
```

   Describe carefully the content of *dm*.

3. Plot the cloud of 10 significant words collected in *dm* by using the function "wordcloud" from the package "wordcloud". What is the interest of this cloud ?
4. Insert your cloud of words into a shiny application. The code to plot the cloud of points must be inserted in the file "helpers.R" as the function "cloud_twitter <- function(processedWord, N = 100, n = 10)" where N is the number of processed tweets.

5. Add a textInput control to choose the keyword on which the cloud of words is based. From this way, you can change “datascience” by an other word : big data, Polytech, etc.
6. Add a sliderInput to choose the number of words plotted in the cloud between 5 until 30 words. The default value is 10 words.