Descriptive Statistics: Getting started in R

M P Gururajan, Hina A Gokhale and Dayadeep Monder

Indian Institute of Technology Bombay, Mumbai

1 Entry and exit: The A to Z of an R session

These instructions are for working with R on a system running Linux OS, specifically, Ubuntu 18.04.5 LTS. We are using RStudio for this session – version 1.3.1073. In the console window of the RStudio (left or left bottom), if you see the following message, you have invoked R successfully:

```
R version 4.0.2 (2020-06-22) -- "Taking Off Again"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
```

As you can see, I am using R version 4.0.2.

After the session, for exiting R, we use following command:

```
q()
```

When we invoke this command (which stands for quit), R wants to know if we have to save the image of the workspace. By giving the answer, namely, (y)es, or, (n)o, you can quit with or without saving the workspace, respectively; (c)ancel is the option to cancel the quit command itself. If the workspace is saved, when R is invoked next, the previous workspace is restored. Note that when R is invoked, if the previously stored workspace is restored, R indicates so at the end of the welcome message.

One can remove such previously stored information using the following command.

```
unlink(".RData")
```

2   Getting and setting working directories

The following command fetches the current working directory:

```
getwd()

## [1] "/home/guru/Desktop/Teaching/MM217OnlineTeaching/RSessions"
```

One can switch to a specific directory using the set working directory command. Let us say, I want to switch to the `home/guru` directory. It can be achieved using the following command:

```
setwd("home/guru")
```

3   Importing csv data files

The `csv` format data files are data which is stored in the form of comma-seperated-values. They can be imported into R using the `read.csv` and stored in a variable named `GT` as follows:

```
GT = read.csv("../Data/RKNGT.csv",header=TRUE)
```

The header variable indicates that the first row of the data is the names of the variables. Note that the file name is given with full directory path.

Using the command `str` you can get the information on the data; for example, you can see the names of the variables (Sentences and Words), their type (int), the number of variables (2) and the total number of observations (100). You can print the first six or last six lines of the data using `head` and `tail` commands:

```
str(GT)

## 'data.frame': 100 obs. of  2 variables:
##  $ Sentence: int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Words   : int  27 16 7 47 39 17 9 8 10 15 ...

head(GT)

##   Sentence Words
## 1        1    27
## 2        2    16
## 3        3     7
## 4        4    47
## 5        5    39
## 6        6    17

tail(GT)

##     Sentence Words
## 95        95    28
## 96        96    15
## 97        97    24
## 98        98     9
## 99        99    18
## 100      100     6
```

Now, we are all set for our descriptive statistics sessions! Let us start!!