intro.R

wnarifin

Tue Oct 3 12:33:42 2017

```
# Introduction to R
```

```
# Author: Wan Nor Arifin
```

```
# Outlines

# - RStudio Interface
# - Function, Library & Object
# - Read data
# - Handle data
# - Basic analysis

# RStudio Interface

## The windows
# 1. Script
# 2. Console
# 3. Environment & History
# 4. Files & others

## Tasks
# - Set the working directory (Files)
# - Install packages a.k.a libraries (Packages)
#   - psych, lavaan, MVN, semTools, semPlot
# - Open a new R script
#   - type all commands/functions here
#   - comments, start with "#"
#   - run all commands by Ctrl+Enter

# Function, Library, Object

## Function
#function(), think of MS Excel function

## Library
library(psych)

##Object
# - name assigned on left side of "<-" / "="
# - variable, data (data frame, matrix, list)
x <- 1
y = 2
z = x + y
z  #type object name, you'll get the value
```

```
## [1] 3
```

```r
# Read data

#We have these files:
# - cholest.csv
# - cholest.sav
# - cholest.dta
# - cholest.xlsx
#Always make sure that you set the working directory first!
data.csv = read.csv("cholest.csv")  #most natural way to open data in R
library(foreign)  #library to read .sav (SPSS) and .dta (STATA) files
data.sav = read.spss("cholest.sav", to.data.frame = TRUE)  #SPSS
data.dta = read.dta("cholest.dta")  #STATA
library(readxl)  #library to read excel files, must install first
data.xls = read_excel("cholest.xlsx", sheet = 1)

# Handle data

## Basics
str(data.csv)  #Basic info
```

```
## 'data.frame':    80 obs. of  5 variables:
##  $ chol    : num  6.5 6.6 6.8 6.8 6.9 7 7 7.2 7.2 7.2 ...
##  $ age     : int  38 35 39 36 31 38 33 36 40 34 ...
##  $ exercise: int  6 5 6 5 4 4 5 5 4 6 ...
##  $ sex     : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ categ   : int  0 0 0 0 0 0 0 0 0 0 ...
```

```r
dim(data.csv)  #Dimension (row/case column/variable)
```

```
## [1] 80  5
```

```r
names(data.csv)  #Variable names
```

```
## [1] "chol"     "age"      "exercise" "sex"      "categ"
```

```r
## View data
head(data.csv)  #View data, first 6 rows
```

```
##   chol age exercise sex categ
## 1  6.5  38        6   1     0
## 2  6.6  35        5   1     0
## 3  6.8  39        6   1     0
## 4  6.8  36        5   1     0
## 5  6.9  31        4   1     0
## 6  7.0  38        4   1     0
```

```r
tail(data.csv)  #View data, last 6 rows
```

```
##    chol age exercise sex categ
## 75  9.4  45        4   0     2
## 76  9.5  52        4   0     2
## 77  9.6  35        4   0     2
## 78  9.8  43        3   0     2
## 79  9.9  47        3   0     2
## 80 10.0  44        3   0     2
```

```r
data.csv  #View all
```

```
##    chol age exercise sex categ
## 1   6.5  38        6   1     0
## 2   6.6  35        5   1     0
## 3   6.8  39        6   1     0
## 4   6.8  36        5   1     0
## 5   6.9  31        4   1     0
## 6   7.0  38        4   1     0
## 7   7.0  33        5   1     0
## 8   7.2  36        5   1     0
## 9   7.2  40        4   1     0
## 10  7.2  34        6   1     0
## 11  7.3  38        6   1     0
## 12  7.3  40        5   1     0
## 13  7.3  40        4   1     0
## 14  7.3  28        5   1     0
## 15  7.3  37        5   1     0
## 16  7.4  38        4   1     0
## 17  7.4  49        5   1     0
## 18  7.4  29        5   1     0
## 19  7.5  40        3   1     0
## 20  7.6  38        5   1     0
## 21  7.6  34        5   1     0
## 22  7.6  46        4   1     0
## 23  7.6  42        5   1     0
## 24  7.6  38        4   1     0
## 25  7.8  32        5   1     0
## 26  7.8  43        4   1     1
## 27  7.8  42        5   1     1
## 28  7.8  40        5   1     1
## 29  7.8  38        5   1     1
## 30  7.9  39        5   1     1
## 31  7.9  39        5   1     1
## 32  7.9  39        5   1     1
## 33  8.0  35        3   1     1
## 34  8.0  38        4   1     1
## 35  8.1  40        5   1     1
## 36  8.1  38        4   1     1
## 37  8.2  45        6   1     1
## 38  8.2  36        4   1     1
## 39  8.3  31        4   1     1
## 40  8.3  34        5   1     1
## 41  8.3  44        4   0     1
## 42  8.3  35        5   0     1
## 43  8.4  40        4   0     1
## 44  8.4  37        6   0     1
## 45  8.5  33        4   0     1
## 46  8.5  46        4   0     1
## 47  8.5  42        5   0     1
## 48  8.5  40        4   0     1
## 49  8.5  45        4   0     1
## 50  8.5  42        5   0     1
## 51  8.5  45        4   0     1
## 52  8.6  38        5   0     1
## 53  8.6  34        3   0     1
```

```
## 54  8.6  44        4   0     1
## 55  8.7  39        3   0     1
## 56  8.7  38        4   0     1
## 57  8.7  39        3   0     1
## 58  8.8  47        3   0     1
## 59  8.8  41        4   0     2
## 60  8.8  44        4   0     2
## 61  8.8  30        3   0     2
## 62  8.9  48        3   0     2
## 63  8.9  47        4   0     2
## 64  8.9  42        4   0     2
## 65  9.0  42        4   0     2
## 66  9.0  49        3   0     2
## 67  9.1  31        2   0     2
## 68  9.2  38        3   0     2
## 69  9.2  38        3   0     2
## 70  9.3  48        3   0     2
## 71  9.3  34        4   0     2
## 72  9.3  45        3   0     2
## 73  9.4  45        3   0     2
## 74  9.4  36        4   0     2
## 75  9.4  45        4   0     2
## 76  9.5  52        4   0     2
## 77  9.6  35        4   0     2
## 78  9.8  43        3   0     2
## 79  9.9  47        3   0     2
## 80 10.0  44        3   0     2
```

```
View(data.csv)  #View, graphical way

## Select specific parts of data (subsetting)
data.csv$age  #View "age" only
```

```
##  [1] 38 35 39 36 31 38 33 36 40 34 38 40 40 28 37 38 49 29 40 38 34 46 42
## [24] 38 32 43 42 40 38 39 39 39 35 38 40 38 45 36 31 34 44 35 40 37 33 46
## [47] 42 40 45 42 45 38 34 44 39 38 39 47 41 44 30 48 47 42 42 49 31 38 38
## [70] 48 34 45 45 36 45 52 35 43 47 44
```

```
data.csv["age"]
```

```
##    age
## 1   38
## 2   35
## 3   39
## 4   36
## 5   31
## 6   38
## 7   33
## 8   36
## 9   40
## 10  34
## 11  38
## 12  40
## 13  40
## 14  28
```

```
## 15  37
## 16  38
## 17  49
## 18  29
## 19  40
## 20  38
## 21  34
## 22  46
## 23  42
## 24  38
## 25  32
## 26  43
## 27  42
## 28  40
## 29  38
## 30  39
## 31  39
## 32  39
## 33  35
## 34  38
## 35  40
## 36  38
## 37  45
## 38  36
## 39  31
## 40  34
## 41  44
## 42  35
## 43  40
## 44  37
## 45  33
## 46  46
## 47  42
## 48  40
## 49  45
## 50  42
## 51  45
## 52  38
## 53  34
## 54  44
## 55  39
## 56  38
## 57  39
## 58  47
## 59  41
## 60  44
## 61  30
## 62  48
## 63  47
## 64  42
## 65  42
## 66  49
## 67  31
## 68  38
```

```
## 69  38
## 70  48
## 71  34
## 72  45
## 73  45
## 74  36
## 75  45
## 76  52
## 77  35
## 78  43
## 79  47
## 80  44
```

```
data.csv[2]
```

```
##    age
## 1   38
## 2   35
## 3   39
## 4   36
## 5   31
## 6   38
## 7   33
## 8   36
## 9   40
## 10  34
## 11  38
## 12  40
## 13  40
## 14  28
## 15  37
## 16  38
## 17  49
## 18  29
## 19  40
## 20  38
## 21  34
## 22  46
## 23  42
## 24  38
## 25  32
## 26  43
## 27  42
## 28  40
## 29  38
## 30  39
## 31  39
## 32  39
## 33  35
## 34  38
## 35  40
## 36  38
## 37  45
## 38  36
## 39  31
```

```
## 40  34
## 41  44
## 42  35
## 43  40
## 44  37
## 45  33
## 46  46
## 47  42
## 48  40
## 49  45
## 50  42
## 51  45
## 52  38
## 53  34
## 54  44
## 55  39
## 56  38
## 57  39
## 58  47
## 59  41
## 60  44
## 61  30
## 62  48
## 63  47
## 64  42
## 65  42
## 66  49
## 67  31
## 68  38
## 69  38
## 70  48
## 71  34
## 72  45
## 73  45
## 74  36
## 75  45
## 76  52
## 77  35
## 78  43
## 79  47
## 80  44
```

```
#In general, syntax data[row(number/name), col(number/name)]
data.csv[1:10, 2:4]  #Row 1 to 10; col 2 to 4
```

```
##    age exercise sex
## 1   38        6   1
## 2   35        5   1
## 3   39        6   1
## 4   36        5   1
## 5   31        4   1
## 6   38        4   1
## 7   33        5   1
## 8   36        5   1
## 9   40        4   1
```

```
## 10  34        6   1
```

```
data.csv[c(1,3,5,7,9), c("age", "chol")]  #Row 1,3,5,7,9; col age & chol
```

```
##   age chol
## 1  38  6.5
## 3  39  6.8
## 5  31  6.9
## 7  33  7.0
## 9  40  7.2
```

```
data.csv[data.csv["age"] == 38, c("age", "chol")]  #Row age = 38; col age & chol
```

```
##    age chol
## 1   38  6.5
## 6   38  7.0
## 11  38  7.3
## 16  38  7.4
## 20  38  7.6
## 24  38  7.6
## 29  38  7.8
## 34  38  8.0
## 36  38  8.1
## 52  38  8.6
## 56  38  8.7
## 68  38  9.2
## 69  38  9.2
```

```
data.csv[data.csv["sex"] == 1, c("sex", "chol")]  #Row Sex = 1; col sex & chol
```

```
##    sex chol
## 1    1  6.5
## 2    1  6.6
## 3    1  6.8
## 4    1  6.8
## 5    1  6.9
## 6    1  7.0
## 7    1  7.0
## 8    1  7.2
## 9    1  7.2
## 10   1  7.2
## 11   1  7.3
## 12   1  7.3
## 13   1  7.3
## 14   1  7.3
## 15   1  7.3
## 16   1  7.4
## 17   1  7.4
## 18   1  7.4
## 19   1  7.5
## 20   1  7.6
## 21   1  7.6
## 22   1  7.6
## 23   1  7.6
## 24   1  7.6
## 25   1  7.8
```

```
## 26   1  7.8
## 27   1  7.8
## 28   1  7.8
## 29   1  7.8
## 30   1  7.9
## 31   1  7.9
## 32   1  7.9
## 33   1  8.0
## 34   1  8.0
## 35   1  8.1
## 36   1  8.1
## 37   1  8.2
## 38   1  8.2
## 39   1  8.3
## 40   1  8.3
```

```
#Can also use subset(), syntax subset(data, condition, variable)
subset(data.csv, age == 38)
```

```
##    chol age exercise sex categ
## 1   6.5  38        6   1     0
## 6   7.0  38        4   1     0
## 11  7.3  38        6   1     0
## 16  7.4  38        4   1     0
## 20  7.6  38        5   1     0
## 24  7.6  38        4   1     0
## 29  7.8  38        5   1     1
## 34  8.0  38        4   1     1
## 36  8.1  38        4   1     1
## 52  8.6  38        5   0     1
## 56  8.7  38        4   0     1
## 68  9.2  38        3   0     2
## 69  9.2  38        3   0     2
```

```
subset(data.csv, age == 38, age:sex)
```

```
##    age exercise sex
## 1   38        6   1
## 6   38        4   1
## 11  38        6   1
## 16  38        4   1
## 20  38        5   1
## 24  38        4   1
## 29  38        5   1
## 34  38        4   1
## 36  38        4   1
## 52  38        5   0
## 56  38        4   0
## 68  38        3   0
## 69  38        3   0
```

```
# Basic analysis

#We use data.sav, with category labels
str(data.sav)  #numerical = num, categorical = Factor
```

```
## 'data.frame':    80 obs. of  5 variables:
##  $ chol    : num  6.5 6.6 6.8 6.8 6.9 7 7 7.2 7.2 7.2 ...
##  $ age     : num  38 35 39 36 31 38 33 36 40 34 ...
##  $ exercise: num  6 5 6 5 4 4 5 5 4 6 ...
##  $ sex     : Factor w/ 2 levels "female","male": 2 2 2 2 2 2 2 2 2 2 ...
##  $ categ   : Factor w/ 3 levels "Grp A","Grp B",..: 1 1 1 1 1 1 1 1 1 1 ...
##  - attr(*, "variable.labels")= Named chr  "cholesterol in mmol/L" "age in year" "duration of exercis
##   ..- attr(*, "names")= chr  "chol" "age" "exercise" "sex" ...
##  - attr(*, "codepage")= int 65001
```

```
summary(data.sav)
```

```
##       chol            age           exercise         sex       categ
##  Min.   : 6.50   Min.   :28.00   Min.   :2.000   female:40   Grp A:25
##  1st Qu.: 7.60   1st Qu.:36.00   1st Qu.:4.000   male  :40   Grp B:33
##  Median : 8.30   Median :39.00   Median :4.000               Grp C:22
##  Mean   : 8.23   Mean   :39.48   Mean   :4.225
##  3rd Qu.: 8.80   3rd Qu.:43.25   3rd Qu.:5.000
##  Max.   :10.00   Max.   :52.00   Max.   :6.000
```

```
## Numerical
library(psych)  #to use describe
describe(data.sav[c("chol","age", "exercise")])
```

```
##          vars  n  mean   sd median trimmed  mad  min max range skew
## chol        1 80  8.23 0.84    8.3    8.23 0.96  6.5  10   3.5 0.00
## age         2 80 39.48 5.13   39.0   39.47 5.19 28.0  52  24.0 0.06
## exercise    3 80  4.22 0.91    4.0    4.20 1.48  2.0   6   4.0 0.04
##          kurtosis   se
## chol        -0.84 0.09
## age         -0.49 0.57
## exercise    -0.64 0.10
```

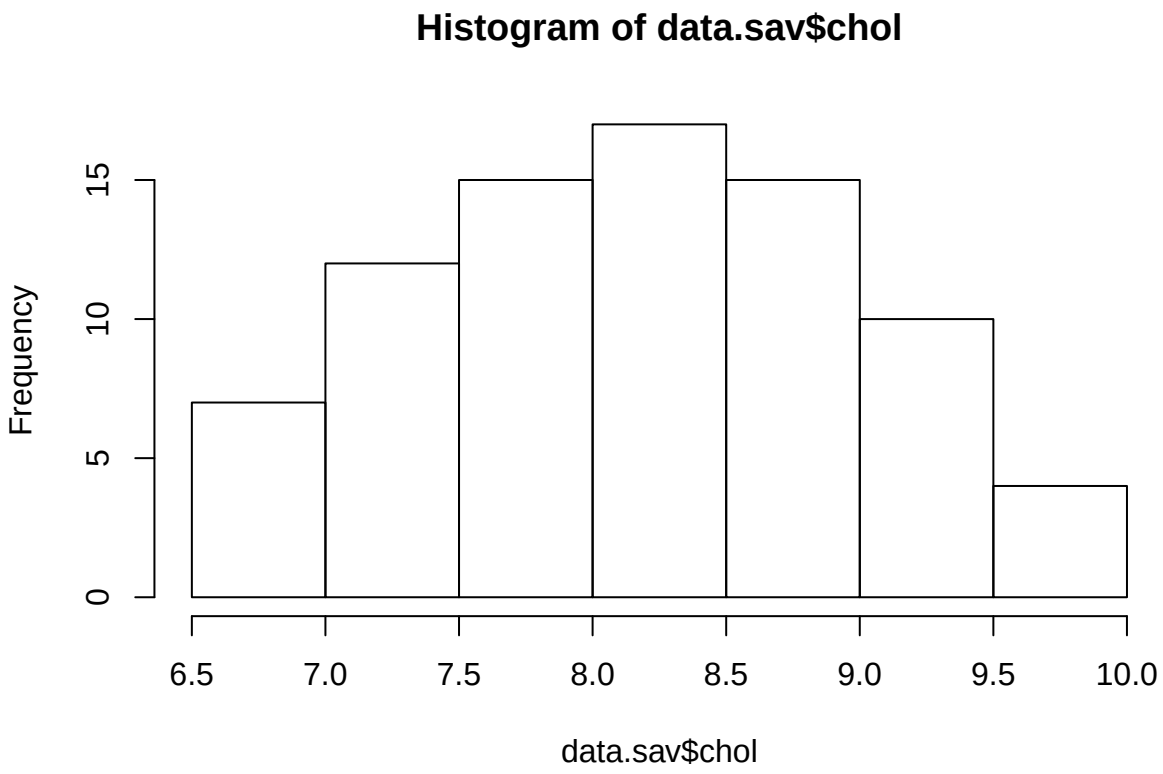
```
## Categorical
table(data.sav$sex)
```

```
##
## female   male
##     40     40
```

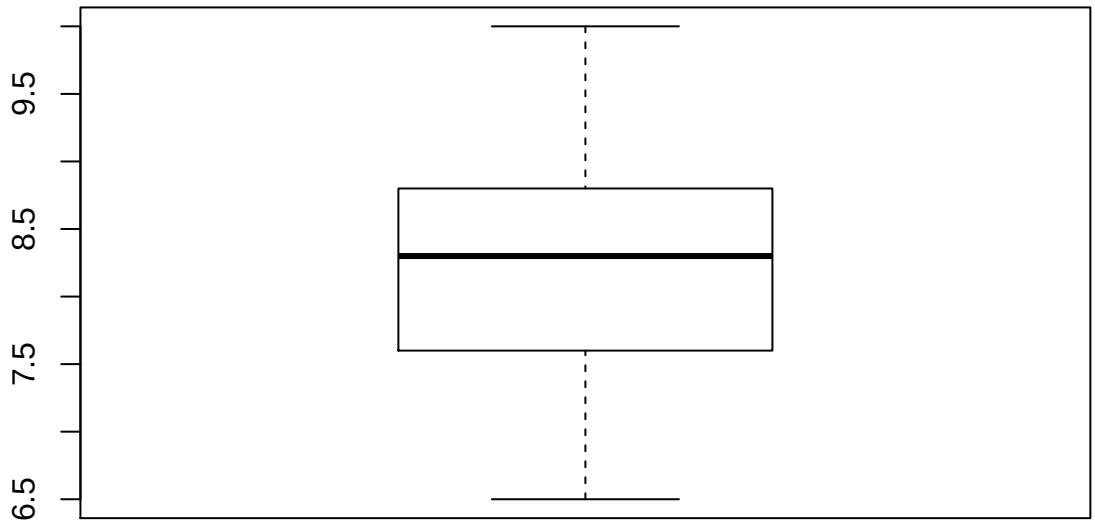
```
table(data.sav$categ)
```

```
##
## Grp A Grp B Grp C
##    25    33    22
```

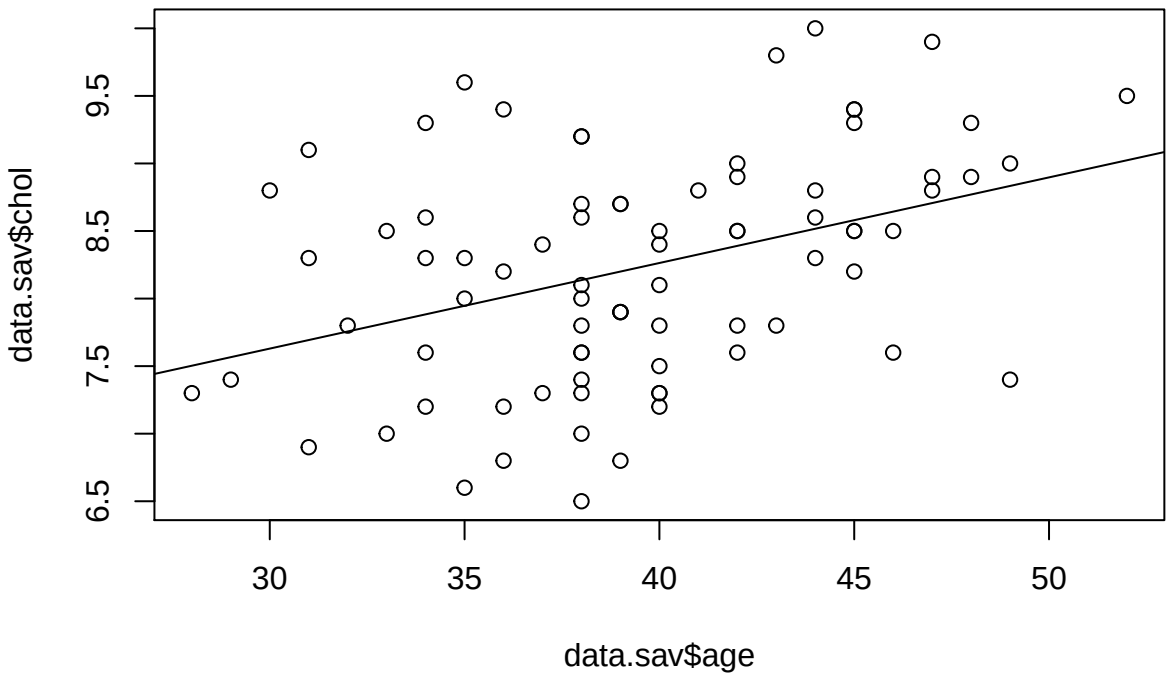
```
## Plots
#Histogram
hist(data.sav$chol)
```



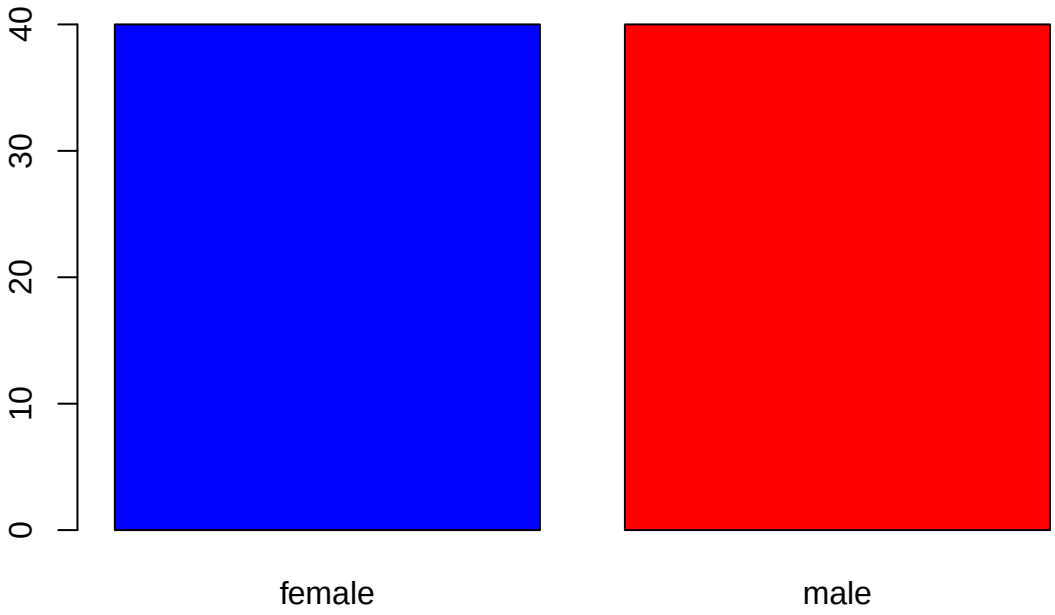

```
#Boxplot
boxplot(data.sav$chol)
```




```
#Scatter plot
plot(data.sav$age, data.sav$chol)
abline(lm(chol ~ age, data = data.sav))  #need two lines of codes
```




```r
#Bar chart
count = table(data.sav$sex)
barplot(count, col = c("blue", "red"))
```




```
# Q&A
```