STA 108 Discussion 1: R Basics

Yuanyuan Li (yynli@ucdavis.edu)

Installation

1. Install R:

1. Go to http://cran.freestatistics.org/.
2. Select the appropriate package based on your operating system.
3. Follow the installers instruction.

2. Install RStudio:

1. Go to https://rstudio.com/products/rstudio/download/.
2. Select the "Desktop" version.
3. Select the installer based on your operating system (usually the recommended for your system works just fine).

3. Install the R Markdown packages (Recommended for class projects):

1. Install Markdown package by typing this command in R **Console** (I will explain it in next section):

```
install.packages("rmarkdown")
```

2. Install LaTeX (TinyTeX) for PDF reports by typing this command in R **Console**:

```
install.packages("tinytex")
```

Starting out

- Click RStudio icon to run this program, and R will be open automatically in the background at the same time. You should have 4 panels. A brief description of the four windows follows:
  - **Environment, History, Build, VCS**: This window lists all Data, Values Functions entered.
  - **Files, Plots, Packages, Help**: Should you use the help command (the ?) in R, the relevant help document will be opened here.Should you plot a figure, the relevant figure will be displayed.
  - **Console**: Where you type in your commands and retrieve your output.
  - **Source**: A convenient place to store your commands, which can then be sent to the console via keyboard shortcuts or with a button. I **highly** recommend putting all of your code in the source window (in a R Script or R markdown), and transferring it to the console.
- Ways of typing commands in R:

  - Type code in Console
  - Create new script file and type code in editor (**recommended**)

- Running R code:
  - In R Console: hit *Enter* after the command line
  - In R Script editor: Highlight the part you need to run and hit the "run" symbol button on the menu; or apply the shortcut: *Ctrl/command + Return*(MAC) or *Ctrl + Enter*(Windows)

- Change working directory:
  - go to Session > Set Working Directory > to source file location or Choose Directory
  - Or use "setwd("file-path")" command

- Save R Source files:
  - click the blue square button
  - click File > Save/ Save as

Let's start coding!

1. Calculation: Use R as a calculator

```
31/4*(37-25)
```

```
## [1] 93
```

```
3^2
```

```
## [1] 9
```

```
sqrt(36)
```

```
## [1] 6
```

```
log(4)
```

```
## [1] 1.386294
```

```
cos(6)
```

```
## [1] 0.9601703
```

2. Objects

Assign values to object 'x' using any one of the following:

```r
x = 5
x <- 5
5 -> x
```

Calculation

```r
x+3
```

```
## [1] 8
```

```r
x^2
```

```
## [1] 25
```

```r
sqrt(x)
```

```
## [1] 2.236068
```

```r
y = x^2
```

Overwrite the existing objects

```r
x = 10
x = x+1
```

3. Vectors

```r
# Create a vector
c(1,3,2,4)
```

```
## [1] 1 3 2 4
```

```r
# Save the vector as 'x'
x = c(1,3,2,4)

# R applies functions to every element of a vector
x - 10
```

```
## [1] -9 -7 -8 -6
```

```r
x^2
```

```
## [1]  1  9  4 16
```

4. Some useful functions

```
mean(x) # mean
```

```
## [1] 2.5
```

```
sd(x) # standard deviation
```

```
## [1] 1.290994
```

```
var(x) # variance
```

```
## [1] 1.666667
```

```
summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    1.75    2.50    2.50    3.25    4.00
```

```
sum(x) # sum of all elements
```

```
## [1] 10
```

```
prod(x) # product of all elements
```

```
## [1] 24
```

```
length(x) # number of elements
```

```
## [1] 4
```

```
x[1:3] # the first three elements
```

```
## [1] 1 3 2
```

5. Import the data

- Read in a data set by specifying the full file path (remember to replace "\" with "/" if your file path include "\"):

```
patients101 = read.csv("~/books/108s21/Discussion/datasets/patients101.csv")
```

- Read in a data set by setting the parent folder as working directory, then use "read.csv" or "read.table" function:

```
setwd("~/books/108s21/Discussion/datasets")#set working directory to "datasets" folder
patients101 = read.csv("patients101.csv")#read the data set in the folder
```

- Use "Import Dataset" button in the menu of Environment:
  - Read .txt file or .csv file : Import Dataset>From Text(base)>choose file>open>import
  - Read Excel file : Import Dataset>From Excel>choose file>open>import

6. Accessing specific rows and columns in a dataframe

```
head(patients101)#Display the first six rows
```

```
##   age totalchol sysBP weight height sedmins      obese marriage gender
## 1  34       135   114   87.4  164.7     480      obese  married      M
## 2  60       202   154  116.8  166.0     240      obese  widowed      F
## 3  26       160   102   97.6  173.0     720      obese  married      M
## 4  49       259   118   86.7  168.4     240      obese    other      F
## 5  80       182   142   79.1  174.3      60 overweight  married      M
## 6  80       148   126   89.6  180.1     540 overweight  widowed      M
```

```
patients101[1:6,]#Display the first six rows by row index
```

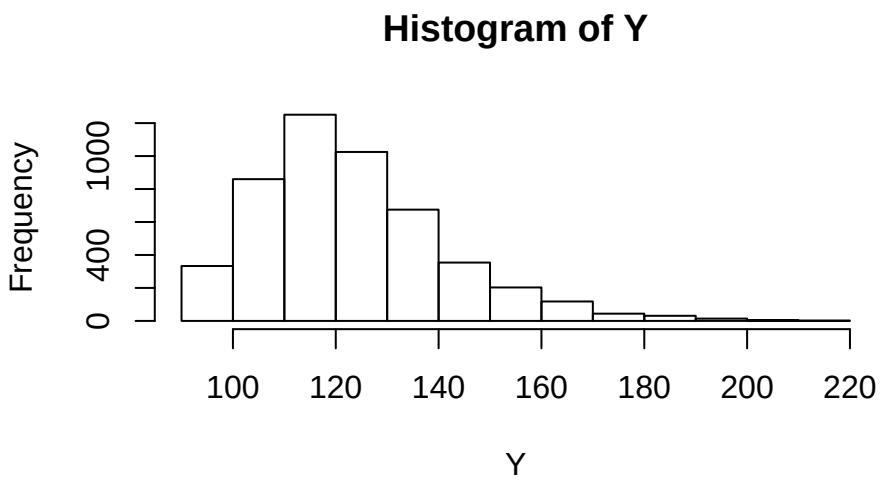
```
##   age totalchol sysBP weight height sedmins      obese marriage gender
## 1  34       135   114   87.4  164.7     480      obese  married      M
## 2  60       202   154  116.8  166.0     240      obese  widowed      F
## 3  26       160   102   97.6  173.0     720      obese  married      M
## 4  49       259   118   86.7  168.4     240      obese    other      F
## 5  80       182   142   79.1  174.3      60 overweight  married      M
## 6  80       148   126   89.6  180.1     540 overweight  widowed      M
```

```
Y = patients101$sysBP # Extract variables from dataset
Y = patients101[,3] # or Extract variables by column index
patients101[1,3]#Display the value is row 1 and column 3
```

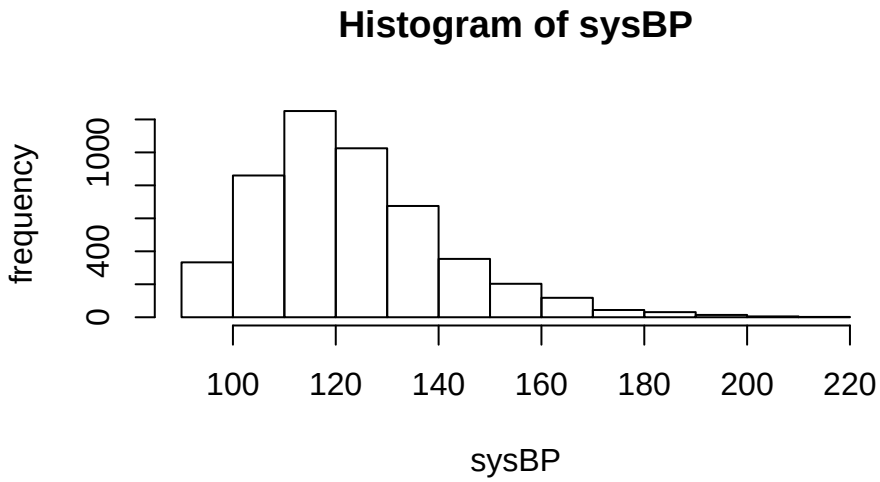
```
## [1] 114
```

7. Plot

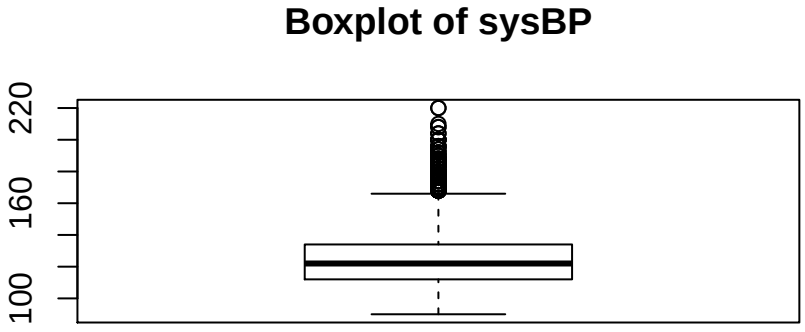
```
# histograms
hist(Y)
```



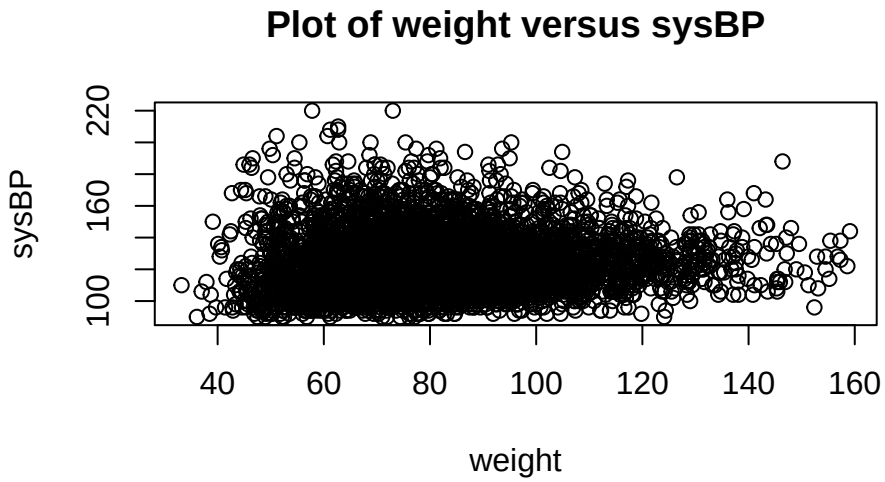

```
hist(Y, xlab = 'sysBP', ylab = 'frequency', main = 'Histogram of sysBP')
```




```
# boxplots
boxplot(Y, main = 'Boxplot of sysBP')
```




```
# scatterplots
X1 = patients101$weight
plot(X1, Y, xlab = 'weight', ylab = 'sysBP', main = 'Plot of weight versus sysBP')
```


Plot of weight versus sysBP
sysBP
100
160
220
40
60
80
100
120
140
160
weight