Introduction to R and Rstudio

### Download and install R and Rstudio

1. Go to <http://www.r-project.org>. This is the home page for the R project.
2. On this page click on the link with the words Download R.
3. There is a list of many web servers from which you can download R. Choose one of these. The one in Iowa is faster than the one in Michigan.
4. Near the top of the page will be a section titled Download and install R. Choose whether it is for Windows or for a MAC.
5. Next install Rstudio. This is an interface for R that students find easier to use.
6. Go to <http://www.rstudio.com/>
7. Click on Download Rstudio
8. Click on the Desktop iccon
9. Under Open Source Edition Click on Download Rstudio Desktop.
10. Choose the version for MAC or Windows.
11. Follow the instruction in the pop-up windows.

### Before you start

Create a folder *Stat* where you can save relevant documents. Create subfolders, e.g. *Rstudio* and *Documents* and *Data*, where you will save your computational work.

### R script

In Rstudio click on File > New File > R script. In this window you can type code and comments that you can save to your folder. When you highlight a piece of code and click *Run* it will copy the code to the *Console* window and execute. Alternatively, you can type directly into the Console window, but it won't save your text.

### R markdown (optional)

In Rstudio click on File > New File > R markdown. You can give this project a title, your name (and your PID, if you hand it in as an assignment) and choose whether you want the output saved in Word or html or pdf format. Only choose PDF output, if you have Miktex installed (also free software). In Rstudio an "untitled" page will appear with a template. First save this file to the folder you created, e.g "Rclassnotes". It will then be saved as "Rclassnotes.Rmd". The extension ".Rmd" stands for R markdown and can be opened with R studio.

## Important:

Click on Session > Set Working Directory > To Source File Location. Then Rstudio will know where to save your putout from your work. You can check your working directory anytime in the Console of Rstudio by typing

getwd()

## [1] "/Users/mariannehuebner/Desktop/Teaching/Classes\_Rmd"

## Create an R markdown document (optional)

Click on "Chunks" and then "Insert Chunk". This will result in a gray area starting with three backslashes followed by {r} and ending with three backslashes. In this grey area you can type R commands.

If you click anywhere in the grey area and then click on Chunks > Run Current Chunk, all R commands will be automatically pasted into the "Console" window and evaluated. If you only want to run one line, highlight this line, and click on Run.

You can enter text like in a Word document in the other areas.

In the toolbar "ABC" is your spell checker. Always run your spell checker after you finished.

In the toolbar the question mark has a "Markdown Quick Reference". If you click on that it will open in the lower right window with tips how to get italic, bold face, or larger text size, and more.

In the toolbar, click on "Knit Word". This will start a process that converts the current template file into a Word file which opens automatically. Do this now.

## Add R commands

A comment can be inserted following a # sign.

2 + 3 \* 5 # Note the order of operations.

## [1] 17

log(10) # Natural logarithm with base e=2.718282

## [1] 2.302585

4^2 # 4 raised to the second power

## [1] 16

3/2 # Division

## [1] 1.5

sqrt(16) # Square root

## [1] 4

abs(3-7) # Absolute value of 3-7

## [1] 4

exp(2) # exponential function

## [1] 7.389056

15 %/%4 # This is the integer divide operation

## [1] 3

# This is a comment line

Let's type a data list into R (in the script file or in an R markdown file). The letter "c" stands for "concatenate".

x = c(1, 3, 7, 9, 11, 13, 20)  
x

## [1] 1 3 7 9 11 13 20

Note that in the output the square bracket *[1]* refers to the first entry of the first line of output. Sometimes you may have longer lines. They will be broken up and numbered by first entry on that line. E.g.

[1] 1 20 30 4 555  
[6] 25 36 77

Suppose you made an error and really wanted the first number in the data list called x to be a 10. You can change the first entry This is a quick without retyping the whole data list.

x[1] = 10  
x

## [1] 10 3 7 9 11 13 20

You can check a number of things in your data list:

x[3] # the third element of x

## [1] 7

x[3:5] # the third to fifth element of x, inclusive

## [1] 7 9 11

x[x>3] # list of elements in x greater than 3

## [1] 10 7 9 11 13 20

You can also add a constant to each element in your data list or multiply by a constant:

x

## [1] 10 3 7 9 11 13 20

x+1

## [1] 11 4 8 10 12 14 21

x\*2

## [1] 20 6 14 18 22 26 40

R is case sensitive, so if you type upper case X instead of the lower case X, you will get an error message.

X

## Error in eval(expr, envir, enclos): object 'X' not found

Note: If you consistenly use lower case variable names, you don't need to remember whether a name was upper or lower case.

Now you have created a data list for which you can calculate data summaries, e.g.

mean(x) # mean

## [1] 10.42857

sd(x) # standard deviation

## [1] 5.287001

IQR(x) # interquartile range

## [1] 4

summary(x) # data summary inlcuding mean, median, quartiles, minimum, maximum

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 3.00 8.00 10.00 10.43 12.00 20.00

## More on data elements in R

Basics with data lists (vectors)

### basics with data vectors  
x<-c(1,3,2,10,5)  
x

## [1] 1 3 2 10 5

sum(x)

## [1] 21

summary(x)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.0 2.0 3.0 4.2 5.0 10.0

sort(x, decreasing=T)

## [1] 10 5 3 2 1

y<-1:5  
y+2

## [1] 3 4 5 6 7

y\*2

## [1] 2 4 6 8 10

y^2

## [1] 1 4 9 16 25

length(x)

## [1] 5

length(y)

## [1] 5

x+y

## [1] 2 5 5 14 10

z<-cbind(x,y)  
z<-rbind(x,y)

Basics with data matrices

A<-matrix(c(1,2,3,4), nrow=2, ncol=2); A

## [,1] [,2]  
## [1,] 1 3  
## [2,] 2 4

rownames(A)<-c("row1", "row2")  
 colnames(A)<-c("col1", "col2")  
 dim(A)

## [1] 2 2

A<-matrix(c(1,2,3,4),byrow=TRUE);A

## [,1]  
## [1,] 1  
## [2,] 2  
## [3,] 3  
## [4,] 4

sqrt(A)

## [,1]  
## [1,] 1.000000  
## [2,] 1.414214  
## [3,] 1.732051  
## [4,] 2.000000

A^2

## [,1]  
## [1,] 1  
## [2,] 4  
## [3,] 9  
## [4,] 16

B<-matrix(c(1,2,3,4,5,6),nrow=2,ncol=3, byrow=T);B

## [,1] [,2] [,3]  
## [1,] 1 2 3  
## [2,] 4 5 6

B[order(B[,1], decreasing=T),]

## [,1] [,2] [,3]  
## [1,] 4 5 6  
## [2,] 1 2 3

B[1,2]

## [1] 2

B[2,] # second row

## [1] 4 5 6

t(B) # transpose of B

## [,1] [,2]  
## [1,] 1 4  
## [2,] 2 5  
## [3,] 3 6

B[-1,] # remove first row

## [1] 4 5 6

B[,-c(1,3)] # remove first and third column

## [1] 2 5

## Data frames in R

A data frame in R gathers together all the variables in a data set.

library(SemiPar) # if this generates and error message, do the following first: install.packages("SemiPar")  
data(fuel.frame)  
car<-fuel.frame  
names(car)

## [1] "car.name" "Weight" "Disp." "Mileage" "Fuel" "Type"

dim(car)

## [1] 60 6

car[1:3,]

## car.name Weight Disp. Mileage Fuel Type  
## 1 Eagle Summit 4 2560 97 33 3.030303 Small  
## 2 Ford Escort 4 2345 114 33 3.030303 Small  
## 3 Ford Festiva 4 1845 81 37 2.702703 Small

unique(car$car.name)[1:3] # specify the variable name after a dollar sign

## [1] Eagle Summit 4 Ford Escort 4 Ford Festiva 4  
## 60 Levels: Acura Legend V6 Audi 80 4 Buick Century 4 ... Volvo 240 4

mean(car$Mileage)

## [1] 24.58333

table(car$Type)

##   
## Compact Large Medium Small Sporty Van   
## 15 3 13 13 9 7

## Finish your R session

Before you finish your R session save your script or markdown file. If it is saved , the name is shown in black, if you made changes, the name is in red and has an asterisk.

Then you can close Rstudio (e.g. click on the red button). It will ask you whether you want to save the R work space. I never do. Since you have saved the script/markdown file, you can always pull it up and recreate what you have done. This is **reproducible**!