

STAT 324: Discussion 1

Setting Up and Intro to R and R Studio

1. Install R (a statistical programming language) from <http://cran.us.r-project.org> and R studio (a free integrated development environment) from <http://www.rstudio.com/products/rstudio/download>. There are a lot of resources to follow (YouTube or Google) if you need assistance.
2. Open RStudio (R Studio will open R within it automatically).
3. Create a folder on your computer specifically for Stat 324 files.
4. Create a new R script via the menu choice "File > New File > R Script". Save it as "Discussion1.R" in your new Stat 324 folder.
5. Take a tour of your workspace. Console: Area to run code, File Pane: Where to type code to save off your work, Bottom Right (Files, Plots, Packages, Help)

General Tips

- Up and down arrows scroll through commands in console
- R does vector calculations
- Everything is case sensitive
- Hash mark comments out (great for explaining your steps within file)
- I like to type most of my work into a script or R Markdown file (another file type that will be introduced soon) so that I can easily share and remember what I did. If I'm using R as a calculator, I just use the console.

Try some Basic Operations

In the console, type the following expressions and then enter to see that R can do basic calculator functions.

```
#everything works just as a calculator#
12*3
sqrt(36)
log(10, base = exp(1))
2^3

#use ? for help file for a function#
?log
```

R objects

We will primarily store and process information in vectors and dataframes and use functions to process those objects.

- vectors can be comprised of numeric or qualitative characteristics, but each vector can only be of one type. We will often build vectors by concatenating values using `c()`.

- dataframes are matrix-like structures, in which the columns can be of different types. Think of data frames as ‘data matrices’ with one row per observational unit but with (possibly) both numerical and categorical variables. (Many experiments are best described by data frames; the treatments are categorical, but the response is numeric.)
- There are many functions that are available in base R to process vectors, dataframes, or other data objects. You can type `?functionname` at the console prompt to get an idea of how to pass information to a desired function.

Assignment and indexing

- Almost everything needs a name
- Use `<-` or `=` to assign

Type the following lines into your R script. To run sets of lines, you can highlight the line and then select run in the top right corner. There are also keyboard shortcuts: Ctrl-Enter (Windows) or Command-Enter (Mac) to run code from your script.

Notice after you run each line the saved values in the Environment Tab. You can also check the values saved in a variable by typing that variable into the console (or running without assignment in your script).

```
#use c to concatenate elements to make a vector#
aa <- c(1,2,3)
bb = c(4,5,6)

#use matrix to make a matrix#
mat <- matrix(c(1,2,3,4), nrow = 2, ncol = 2, byrow = T)
mat #Print out what the matrix is

#all operations work component-wise on vectors and matrices#
aa*3
log(aa)
mat*3
(cc<-aa+bb)

#use [] to request pieces of a vector or matrix#
aa[1]
aa[-3]
aa[c(1,3)]

#specific element#
mat[1,2]

#entire row#
row1<-mat[1,]

#entire column#
col1<-mat[,1]
```

Some Nice Shortcuts

```
1:10
#?rep
rep(3,6)
rep(c(1,2,3), times=3)
rep(c(1,2,3), each=3)
seq(from = 1, to = 9, by = 2)
```

Entering Data

1. You can import the data from a text or csv (or other similar file)
2. You can enter data manually by typing it in with the concatenate "c" that we used above and in lecture.
3. There are some data sets that are available by default after installing R

set working directory to where you have data csv or text file () saved (save these to your new Stat324 folder)
Use Session>Set Working Directory>Choose

```
#setwd("~/Desktop/UW Stat Teaching/324/Discussion1")
#setwd("file path")
```

Reading in Sloth Data from a CSV file. Use ?read.csv to see the difference between the read.*** functions

```
#for csv file
slothdata <- read.csv("Sloth.csv", header = TRUE)
#We specify header=T since the first row has names that describe the columns
```

Entering Sloth data by hand from the csv file (not recommended because easier to make typos, just included as reference).

```
Species<-rep(rep(c("ThreeToed", "TwoToed"), each=3), times=2)
WeightKg<-c(4.0,5.0, 4.6, 3.5, 3.8, 4.0, 4.7, 4.6, 5.0, 6.3, 5.0, 5.3)
Gender<-c("M", "F", "F", "M", "M", "F", "M", "M", "F", "M", "F", "M")
Treatment<-rep(c("High", "Low"), each=6)
RecoverTime<-c(10.00, 12.50, 12.30, 8.75, 10.00, 10.00, 11.80, 11.50, 12.50, 16.00, 12.50, 13.25)

slothdata.manual<-data.frame(Species, WeightKg, Gender, Treatment, RecoverTime)
```

Use View or head to look at the data that was imported. The str() function tells you the characteristics of the data.

```
#View(slothdata)
head(slothdata)
str(slothdata)

head(slothdata.manual)
str(slothdata.manual)
```

Describe the data (number of observations, number of variables, type of variables). Did R correctly recognize the types of variables when we imported? When we manually created data frame?

Manipulating Data Frames

```
#use $ to refer to single column within a data frame
slothdata$WeightKg

#use subset to look at specific data#
sloth_lowweight <- subset(slothdata, WeightKg < 4)
#View(sloth_lowweight)
#Remove # at start of line so it can run and you can see how sloth_lowweight is defined
sloth_2_fem <- subset(slothdata, Species == "TwoToed" & Gender == "F")
#View(sloth_2_fem)

#can also put logicals in [] to subset data
sloth_2Toed<-slothdata[slothdata$Species=="TwoToed",]
#This returns all columns/rows of information for those species that are "TwoToed"

sloth_2Toed_weight<-slothdata[slothdata$Species=="TwoToed", "WeightKg"]
#This returns a vector of the weights for the two-toed species
```

Graphing Characteristics of Interest

You can either graph subsets of the data independently or R can graph the combination with some functions.

```
hist(slothdata$RecoverTime, main="Recovery of all sloths observed", xlab="Time")
hist(sloth_2Toed$RecoverTime, main="Recovery time of 2-Toed sloths", xlab="Time")
stem(sloth_2Toed$RecoverTime) #Stem and Leaf Plot
stem(sloth_2Toed$RecoverTime, scale=2) #Stem and Leaf Plot

boxplot(slothdata$RecoverTime) #Boxplot of all Recovery Times
boxplot(RecoverTime~Treatment, data=slothdata) #Recovery Times broken out by treatment
boxplot(RecoverTime~Gender, data=slothdata) #Recovery Time broken out by gender

#Notice, boxplots are not great for small sets of data - hides how many data points we have
#With a small number of observations, it is actually more helpful to view the actual data values

stripchart(RecoverTime~Treatment, data=slothdata)
stripchart(RecoverTime~Gender, data=slothdata)
stripchart(RecoverTime~Species, data=slothdata)

require(ggplot2)
ggplot(data=slothdata, aes(x=Treatment, y=RecoverTime, color=Treatment))+
  geom_point()

ggplot(data=slothdata, aes(x=Treatment, y=RecoverTime, color=Gender))+
  geom_point()

#Other Interesting Graphs?
```

What kind of relationships between Recovery Time and Treatment do we see in the data?

What kind of relationships between Recovery Time and Gender do we see in the data?

What kind of relationships between Recovery Time and Species do we see in the data?

Computing Summary Measures of Interest

Useful R Functions:

`summary(x)`: min, max, median, and quantiles of vector 'x' `mean(x)`: mean of vector 'x' `sd(x)`: sample standard deviation of vector 'x' `sort(x)`: returns the numbers of vector 'x' in ascending (or descending) order `hist(x)`: histogram of vector 'x'

```
summary(sloth_2Toed_weight)
summary(sloth_2Toed)
mean(slothdata$RecoverTime)
mean(sloth_2Toed$RecoverTime)
median(sloth_2Toed$RecoverTime)
sd(sloth_2Toed$RecoverTime)
```

Additional Practice

1. Try creating an R Markdown File (HTML) and then...
2. Create Side-by-Side box plots to show the comparison of weights between the treatment groups (Treatment High vs Treatment Low).
3. Determine the mean, median, and sample sd for the weights for each treatment group.