

Short Course on Statistics and Data Analysis

Brian Fannin, Mary Meyer, Jean Opsomer

May 4, 2014

Introducing us

- Brian Fannin
- Mary Meyer
- Jean Opsomer

Introducing you

Course overview

- Day 1
 1. Review of R and R-Studio
 2. Survey: recap of key characteristics
 3. Survey: introduction to R survey package
- Day 2
 4. Survey: descriptive statistics
 5. Data analysis: plotting and linear regression

Course overview (2)

- Day 3
 - 6. Data analysis: logistic regression
 - 7. Survey: plotting
- Day 4
 - 8. Survey: linear and logistic regression
 - 9. Survey: plotting and mapping your data
- Day 5
 - 10. Apply what you've learned: data analysis/survey

"Relax and embrace the chaos that is R." - Patrick Burns

www.burns-stat.com/pages/Present/infernoishR_annotated.pdf

First step

Does everyone have R and RStudio installed?

Please open RStudio.

Question: What is the area of a circle with radius = 4?

pi

```
print(pi * 4^2)
```

```
## [1] 50.27
```


Basic questions

- What version of R are you running?
- What version of RStudio do you have?
- How do you get help when you know the name of a command? Find the help page for the command: `read.csv`.
- And when you don't? Find a help page for cluster analysis

Basic answers

- `sessionInfo()` OR Menu: Tools->Global Options->General
- `rstudio::versionInfo()` OR Menu: Help->About RStudio
- Single question mark
- Double question mark

File operations

- What is your working directory?
- Change your working directory
- Within a project directory, create a new directory called "Week3"
- Delete that directory

File ops answers

```
getwd()  
setwd("C:/Users/bfannin/Dropbox/rwanda_course/BAF Files/")  
dir.create("Week3")  
unlink("Week3", recursive = TRUE)  
'?'(files)
```

Exercise

- Create the working directory you will use for the week.
- Create a new, blank script in RStudio.

source

'source' will execute all of the commands in the R script you reference. It will NOT open the script for editing.

```
source("SomefileName.R")
```

Reading data

Very common to store data in .CSV, .TXT, .TSV or similar format. 'read.table' is the generic function to read data. 'read.csv' is a convenience wrapper for 'read.table'.

```
oldDir = setwd("../Data/")

myData = read.table("Rwanda_frame.csv",
  sep = ",", stringsAsFactors = FALSE)

myData = read.csv("Rwanda_frame.csv",
  stringsAsFactors = FALSE)
setwd(oldDir)
```

More notes read.table parameters

- header - If the file has column headers these will be used as the column names of the data frame.
- stringsAsFactors - Defaults to your system option. I typically set this parameter to FALSE
- colClasses - May be useful to enforce strict interpretations of figures, i.e. return an error when character data is in a tet column
- skip - Easy to ignore preamble rows.

Binary save

```
save(myData, file = "SavedData.RData")  
load(file = "SavedData.Rdata")
```

Installing packages - 1

Several ways:

1. From a CRAN mirror
2. Locally
3. Copy someone else's library directory
4. From GitHub

The library path may be anywhere.

Installing packages - 2

1. Understand where your library is - This happens once
2. Install a package - This happens once
3. Load the package into memory - This happens many times
4. Update the package
5. Remove the package

Where is your library?

```
.libPaths()  
packageDir = .libPaths()[1]  
list.files(packageDir)
```

Install a package

```
install.packages("ggplot2")
```

Load an installed package into memory

Two ways:

1. `library` - will load the file
2. `require` - will load the file and return a value indicating whether it was loaded

```
library(MRMR)
```

```
# Will load the package and return TRUE/FALSE if the  
# package could be loaded
```

```
require(MRMR)
```

Remove a package

```
remove.packages("IdontWantThisAnymore")
```

Exercise

- Install the survey package.
- Install the sp package.
- Confirm the packages have been installed.