Introduction to the R Statistical Environment

Sarah Hunter

sghunter@umich.edu

ICPSR Day 3

Overview

Course Outline

Day	Topics
1	Introduction and Workflow
2	Loading and cleaning Data in R
3	Using the tidyverse for advanced data manipulation
4	Descriptive Statistics
5	Data Visualization with ggplot
6	Debugging your code
7	Common Models
8	Presenting results
9	Loops and functions

Loading Data into R

1. Identify your file format
2. Set your working directory
3. Install/Load package needed to load data (if applicable)
4. Load the data into R, creating an object
5. Check to make sure the data were loaded correctly

Common Data-loading Commands

File Extension	R Command
.csv	read.csv("filename.csv")
.dta	read.dta("filename.dta") [in `foreign` library]
.dta	read.dta13("filename.dta") [in `readstata13` library]
.sav	read.spss("filename.sav") [in `foreign` library]

Common Errors in Loading Data

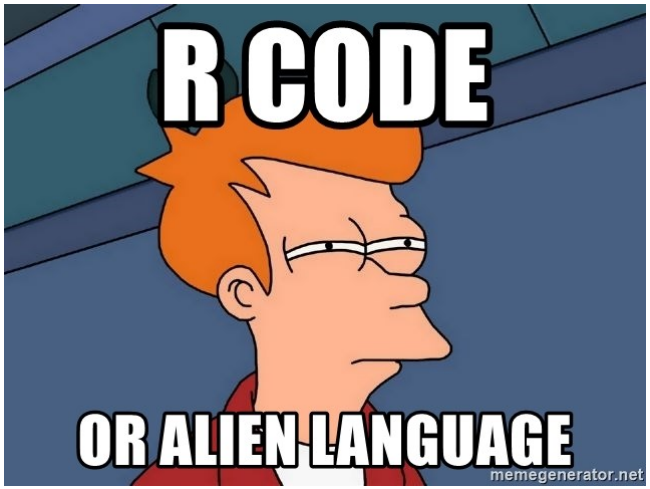
So you get an error message.

- Did you use the correct command for the data type?
- Did you set your working directory?
- Is your data saved in your working directory?
- Check the name of the data again, is it correct in R?
- Did you use the library command to load a package (if needed)?

Today's Topic: The Tidyverse

What is the Tidyverse?

A Warning



- The Tidyverse has its own syntax and structure
- All Tidyverse packages do one specific thing, but also work together seamlessly
- The Tidyverse does data importing (readr), data cleaning(dplyr and tidyr), data visualization (ggplot2), and data analytics (purrr).

Why Learn the Tidyverse?

- More efficient
- easier to link functions into one step instead of multiple
- Makes data management/wrangling so much easier

Important Tidyverse Commands

- `%>%`
- `select`
- `filter`
- `mutate`
- `pivot`

And Now...

To R!

Until Next Time