

# Introduction to the R Statistical Environment

Sarah Hunter

*[sghunter@umich.edu](mailto:sghunter@umich.edu)*

ICPSR Day 2

# Overview

- 1 Review
- 2 Data Types
- 3 Putting Data in R
- 4 Checking Data
- 5 Cleaning Data

# Class Structure

**Lecture** Most days will be a lecture explaining the logic behind R coding and the basics of how to use the program.

**Labs** Some days will have in class labs, or a set of tasks to perform in R

## Topics Covered in this Class

- Data Manipulation
- Descriptive Statistics
- Data Visualization
- Debugging
- Linear Models and Extensions
- Model Presentation
- Simple Programming
- Special Topics if time permits

# How to Get the Most from this course

- 1 Come to class
- 2 Listen/read
- 3 **Practice**

# 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

## So, R... What is so special about it?

- Free and open-source
- User-created packages
- Flexible and wide range of uses
- Better data visualization
- Highly valued skill in academia and industry

# The not-so-great parts of R

- Steep learning curve
- 10 ways to do any one task
- Highly sensitive
- No “point-and-click”



# The Script File

Save. Your. Code.

- Saving code for future reference'
- Replication
- So you don't have to figure it out again

# R Packages

## Defined

“collection of functions, data, and documentation that extends the capabilities of base R (Wickman and Grolemond)”

Packages help you:

- Perform specific (user-created) functions
- Access certain datasets
- Create specialized plots/figures

# Today's Topic: Loading and Cleaning Data in R

# Data Files

Data comes in all types of different files. R can read the following types of data files:

- .csv
- .txt
- .sav
- .dta
- .xls
- .xlsx
- .dat
- .spss
- .Rdata

# 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	<code>read.csv("filename.csv")</code>
.dta	<code>read.dta("filename.dta")</code> [in <code>foreign</code> library]
.dta	<code>read.dta13(filename.dta)</code> [in <code>readstata13</code> library]
.txt	<code>read.table("filename.txt", sep=" ")</code>

## 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)?

## How R Stores and Reads Data: Datasets

R Stores data in vectors and matrices. (More on this later)



## How R Stores and Reads Data: Variables

Variables are slightly different:

- Variables are **nested** in datasets
- You must refer to the dataset AND the variable in order to manipulate a variable.
- How R reads these variables determines what R can do with them

# Types of Variables (According to R)

## Numeric

- Your usual continuous variables such as GDP, etc.

## Integer

- Whole numbers, such as years in school or survey response, etc.

## Character

- Non-numeric data, such as "yes" or "no"

## Checking your data is loaded correctly

Do NOT skip this step. Always make sure your data are loaded correctly by inspecting your data:

- Check the names of your data columns
- Check the dimensions of your dataset
- Check the first few rows
- Look at the summary statistics

## Real-Word Data

Data in the wild are...

- Messy
- Not what you need for your analysis
- Incomplete
- Overly complicated
- Missing
- Too much
- At the wrong level

**This is why we first need to clean data before using it, also called data management!**

## Some Vocabulary for Cleaning Data

- Subset
- Merging
- Transforming
- Recoding
- Reshaping (more on this tomorrow)

# Until Next Time