# Computational Bootcamp 3: Datasets in R

Henry Watson

Georgetown University

8/17/22

# What We'll Be Covering Overall

1. Software installation, file management
2. Basics of R: writing code, creating objects, thinking in matrices
3. **More R: working with datasets**
4. Stata: pros & cons vs. R, working with datasets
5. LaTex: producing documents with Markdown and Overleaf

# What We'll Be Covering Today

1. Packages in R
2. Loading in datasets
3. Manipulating datasets
4. Combining datasets
5. Advanced Topics?

## Packages in R

- A collection of functions, data, and documentations which is publicly shared to enhance the functionality of R
- R functions that don't use packages are referred to as "Base R"
- Packages installed from within R

## Common Packages in R

- tidyverse : collection of R packages designed for data science
  - dplyr : "A Grammar of Data Manipulation"
  - ggplot2 : "The Grammar of Graphics"
  - tidyr : Tools to create "tidy" data
- foreign / haven / readr / readxl: packages to allow importing datasets from file formats other than text
- stargazer : formatted regression tables
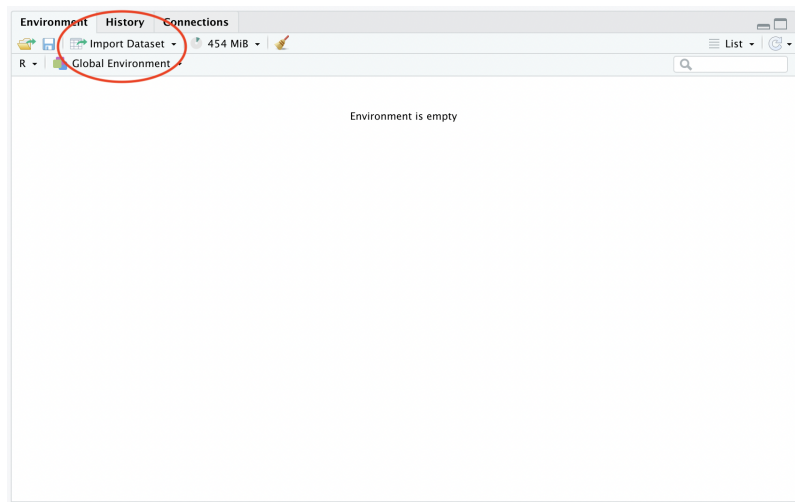- rmarkdown / tinytex : "Dynamic Documents for R"

# Installing Packages

```r
install.packages("package_name")
library(package_name)
```
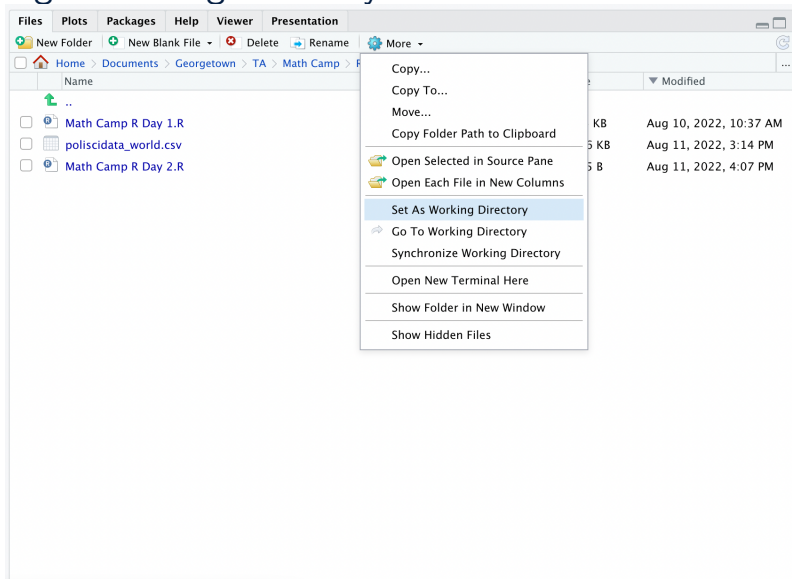
## Installing Packages

- Install packages once. *Don't* include install.packages() in Scripts
- Run library() each new session of R. *Do* include library() in Scripts

# Loading Datasets into R: Point-and-Click

# Setting a Working Directory in R: Point-and-Click

## A Note on Point-and-Click Methods

- It's okay to use point-and-click methods, especially as you're learning, with one key caveat:
- *Always copy the returned code from the Console and paste it into a Script*
- Remember, a major advantage of statistical programming is replicability, and point-and-click methods aren't replicable (unless you copy the code!)
- You will still need to install and library packages like haven to import certain data types, even if you use point-and-click