

# Computational Bootcamp 1: Software Installation

Henry Watson

Georgetown University

8/15/22

# 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

- ➊ Intro to statistical programming
- ➋ Good data analysis practices
- ➌ File Management
- ➍ Installing R, Stata, LaTeX

# Why Learn Statistical Programming?

- Excel is powerful and useful...but won't get you all the way
- Replicability & Collaboration
- Traceable errors
- Large datasets
- Advanced data analysis
- Data visualization

# Organizing File Systems

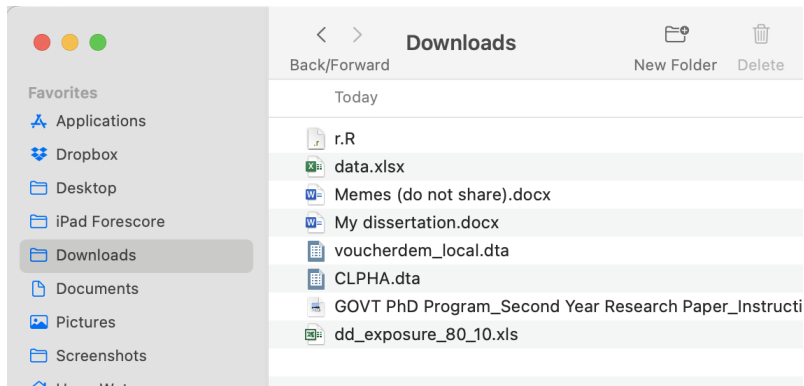
The image shows a file manager interface with a sidebar on the left and a main pane on the right. The sidebar, titled 'Favorites', lists various locations: Applications, Dropbox, Desktop, iPad Forescore, Downloads, Documents, Pictures, Screenshots, HenryWatson, Georgetown (highlighted), Recents, AirDrop, Movies, and Music. The main pane is titled 'Georgetown' and shows a list of folders. At the top of the main pane are navigation controls: back/forward arrows, a 'New Folder' button, and a 'Delete' button. The folder list has a header 'Name' and contains the following items, each preceded by a right-pointing arrow: GOVT 721 - Research in American Politics, GOVT 722 - Public Opinion and Political Behavior, GOVT 723 - State and Local Politics, GOVT 767 - Political Institutions, GOVT 774 - Justice and Immigration, GOVT 780 - Fundamentals of Political Theory, GOVT 789 - American Political Theory, GradGov, Grants, ICPSR, Math Camp, Methods Paper, Omnibus Survey, Other Applications, PPOL 628 - Text as Data, Prospectus, and Publications.

Name
> GOVT 721 - Research in American Politics
> GOVT 722 - Public Opinion and Political Behavior
> GOVT 723 - State and Local Politics
> GOVT 767 - Political Institutions
> GOVT 774 - Justice and Immigration
> GOVT 780 - Fundamentals of Political Theory
> GOVT 789 - American Political Theory
> GradGov
> Grants
> ICPSR
> Math Camp
> Methods Paper
> Omnibus Survey
> Other Applications
> PPOL 628 - Text as Data
> Prospectus
> Publications

# Organizing File Systems

- Files should be organized into projects with subfolders for different types of files (e.g: data, scripts, etc)
- This will make it easy to collaborate with others, post replication files, and type your filepaths into applications like Stata and R.
- Name your files consistently, and in ways that help you manage versions and trace your project
- Starting a project with an organized file system is much easier than cleaning it up afterward.

# Don't Throw Everything in Downloads!



# What is a file path?

- Identifies exactly where a file is saved on your computer
- Examples: Can vary based on your OS, software you're using
  - `~/Documents/Georgetown/Math Camp/data.xlsx`
  - `/Users/HenryWatson/Documents/Georgetown/Math Camp/data.xlsx`
- Working Directory: a file path to a folder
  - Tells statistical software where to look for, and save, files
- R and Stata both have built-in, point-and-click ways to identify file paths!



# What is a file extension?

- Data: `.xlsx` (Excel) ; `.xls` (old Excel) ; `.Rdata` (R data format) ; `.dta` (Stata data format)
- Text Data: `.txt` (Plain text) ; `.csv` (Comma Separated Values) ; `.tsv` (Tab Separated Values)
- Code: `.r` (R script) ; `.rmd` (R script with markdown) ; `.do` (Stata script)
- Not a comprehensive list! Just the main ones you'll be working with.

# Installation: R



[CRAN](#)  
[Mirrors](#)

[What's new?](#)  
[Search](#)

[About R](#)  
[R Homepage](#)  
[The R Journal](#)

[Software](#)  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Task Views](#)  
[Other](#)

[Documentation](#)  
[Manuals](#)  
[FAQs](#)  
[Contributed](#)

**Mirrors: R is available to download from different locations. Pick the closest to you for a faster download.**

## Download R

The Comprehensive R Archive Network

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux \(Debian, Fedora/Redhat, Ubuntu\)](#)
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

### Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2022-06-23, Funny-Looking Kid) [R-4.2.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

### Questions About R

- If you have questions about R like how to download and install the software, or what the

# Installation: R Studio

## Download R Studio

### Choose Your Version

The RStudio IDE is a set of integrated tools designed to help you be more productive with R and Python. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace.

[LEARN MORE ABOUT THE RSTUDIO IDE](#)



**RStudio Team**

RStudio's recommended professional data science solution for every team. **RStudio Team** is a bundle of RStudio's popular professional software for data analysis, package management, and sharing data products.

[Learn more about RStudio Team](#)

#### RStudio Desktop

Open Source License

**Free**

DOWNLOAD

#### RStudio Desktop

Pro

Commercial License

**\$995**

/year

BUY

#### RStudio Server

Open Source License

**Free**

DOWNLOAD

#### RStudio Workbench

Commercial License

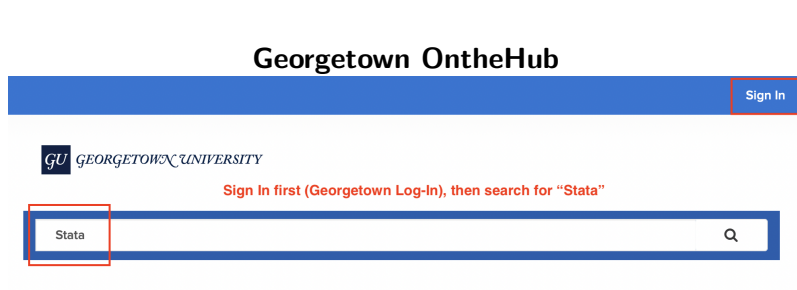
**\$4,975**

/year

(5 Named Users)

BUY

# Installation: Stata



# Installation: Stata

## Georgetown OntheHub

Stata/BE 17 ▲



Fast. Accurate. Easy to use. Stata is a complete, integrated software package that provides all your data science needs—data manipulation, visualization, statistics, and automated reporting.

Stata/BE is the basic version of Stata. Up to 2,147,583,647 observations and 2,048 variables are allowed, depending on memory. Statistical models may have up to 800 variables.

Choose a platform:

Windows

[Stata/BE 17 \(Windows\)](#)

[View Offering Restrictions](#)

Choose your platform, Add to Cart, Check Out

Free

Your “Use Case” is “Classwork”

Add to Cart

**STATA BE is provided under Georgetown University license with Stata Inc. and may be installed on Personal Machines by Students for University Business.**

# Installation: Stata

- Write down your Serial Number, Code, and Authorization Key (or take a screenshot)
- Click on the downloaded file and go through the InstallShield Installation Wizard
- Open Stata and enter your Serial Number, Code, and Authorization Key

# Installation: LaTeX

- Installing LaTeX can take a long time
- You do not need to install LaTeX to use Overleaf (although you will need to make an account on [www.overleaf.com](https://www.overleaf.com))
- My preferred way to install LaTeX, if you want it primarily for RMarkdown, is to install it through R.

# Installation: LaTeX

- Installing LaTeX can take a long time
- You do not need to install LaTeX to use Overleaf (although you will need to make an account on [www.overleaf.com](http://www.overleaf.com))
- My preferred way to install LaTeX, if you want it primarily for RMarkdown, is to install it through R (TinyTex package)
- We'll go over this again on Friday once we're more comfortable with R!



# Installation: LaTeX

```
install.packages("rmarkdown")  
  
library(rmarkdown)  
  
install.packages("tinytex")  
  
tinytex::install_tinytex()
```