Computational Bootcamp 1: Software Installation

Ankushi Mitra

Department of Government

Georgetown University

August 14, 2023

What We'll Be Covering Overall

1. File management, software installation

What We'll Be Covering Overall

1. File management, software installation
2. Basics of R: writing code, creating objects, thinking in matrices

What We'll Be Covering Overall

1. File management, software installation
2. Basics of R: writing code, creating objects, thinking in matrices
3. R: working with datasets

What We'll Be Covering Overall

1. File management, software installation
2. Basics of R: writing code, creating objects, thinking in matrices
3. R: working with datasets
4. More R: data cleaning, visualization

What We'll Be Covering Overall

1. File management, software installation
2. Basics of R: writing code, creating objects, thinking in matrices
3. R: working with datasets
4. More R: data cleaning, visualization
5. LaTex: producing documents with Markdown and Overleaf

What We'll Be Covering Today

1. Intro to statistical programming

What We'll Be Covering Today

1. Intro to statistical programming
2. Good data analysis practices

What We'll Be Covering Today

1. Intro to statistical programming
2. Good data analysis practices
3. File Management

What We'll Be Covering Today

1. Intro to statistical programming
2. Good data analysis practices
3. File Management
4. Installing R, Stata, LaTeX

Why Learn Statistical Programming?

- Excel is powerful and useful...but won't get you all the way

Why Learn Statistical Programming?

- Excel is powerful and useful...but won't get you all the way
- Replicability & Collaboration

Why Learn Statistical Programming?

- Excel is powerful and useful...but won't get you all the way
- Replicability & Collaboration
- Traceable errors

Why Learn Statistical Programming?

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

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

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

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
- State of the discipline

Why Learn Statistical Programming?

- The "Worm Wars"

Why Learn Statistical Programming?

- The "Worm Wars"
  - Miguel and Kremer. 2003. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.".
  - Michael Kremer wins the Nobel Prize.

Why Learn Statistical Programming?

- The "Worm Wars"
  - Miguel and Kremer. 2003. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.".
  - Michael Kremer wins the Nobel Prize.
  - Results disappear after correcting coding errors (Aiken et al. 2015).

Why R?

- Open source for individual users

Why R?

- Open source for individual users
- Packages for almost everything

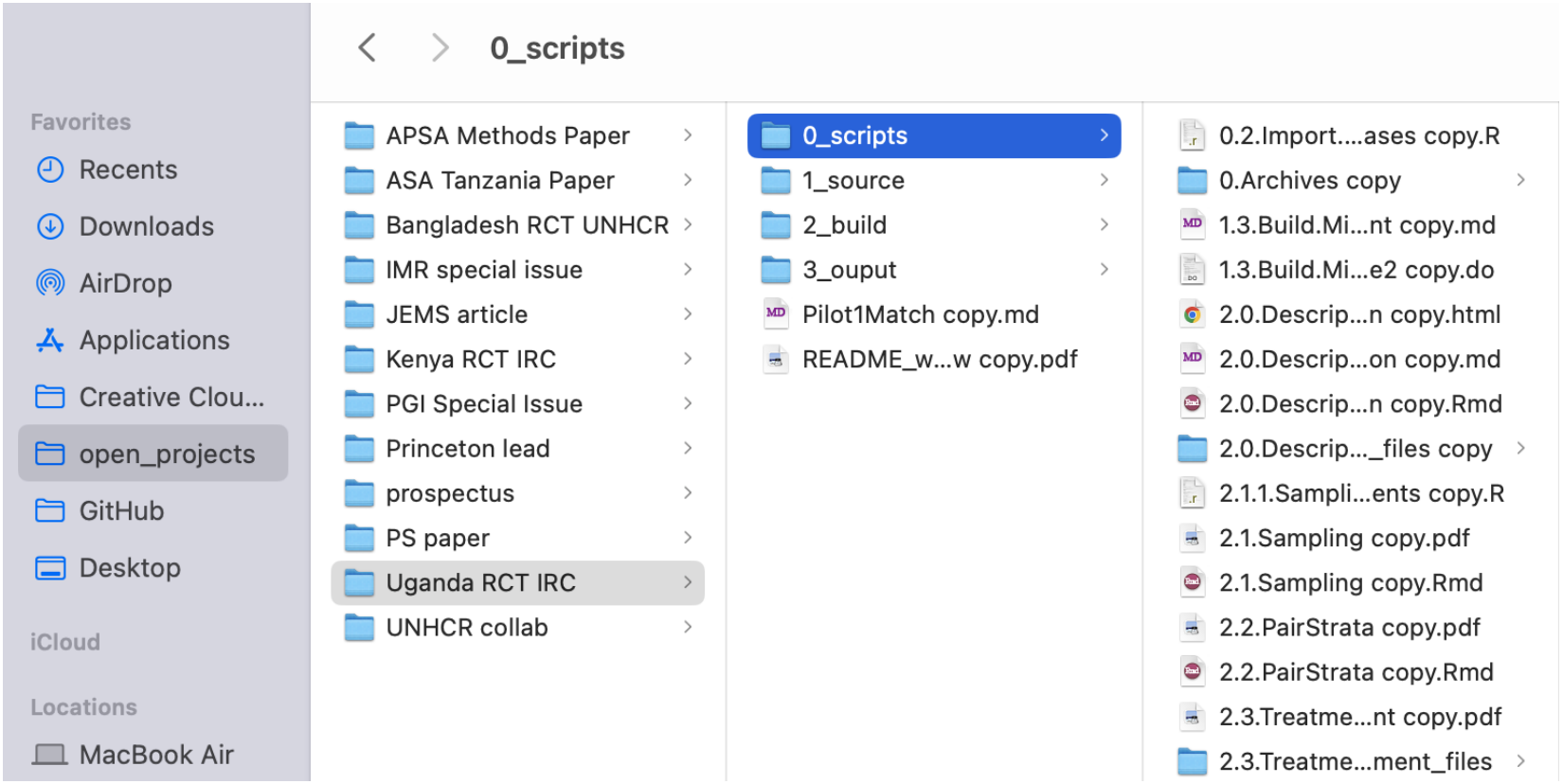
Why R?

- Open source for individual users
- Packages for almost everything
- Most popular in statistics and data science communities

Why R?

- Open source for individual users
- Packages for almost everything
- Most popular in statistics and data science communities
- Relatively easy and non-technical

Organizing File Systems



Organizing File Systems

- Files should be organized into projects with subfolders for different types of files (e.g: data, scripts, etc)

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.

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

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.

What is a file path?

- Identifies exactly where a file is saved on your computer

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
  - ~/Desktop/Edu/PhD/Math Camp 2023
  - /Users/user/Desktop/Edu/PhD/Math Camp 2023

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
  - ~/Desktop/Edu/PhD/Math Camp 2023
  - /Users/user/Desktop/Edu/PhD/Math Camp 2023
- Working Directory: a file path to a folder
  - Tells statistical software where to look for, and save, files

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
  - ~/Desktop/Edu/PhD/Math Camp 2023
  - /Users/user/Desktop/Edu/PhD/Math Camp 2023
- Working Directory: a file path to a folder
  - Tells statistical software where to look for, and save, files
- R and Stata both have ways to set and identify file paths

What is a file extension?

- Data: .xlsx (Excel) ; .xls (old Excel) ; .Rdata (R data format) ; .dta (Stata data format)

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)

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)

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

- You'll need R/Base R and a GUI, R Studio

Installation: R

- You'll need R/Base R and a GUI, R Studio
- R Studio is relatively user-friendly, has a large user community

Installation: R

Download R

The Comprehensive R Archive Network

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 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!

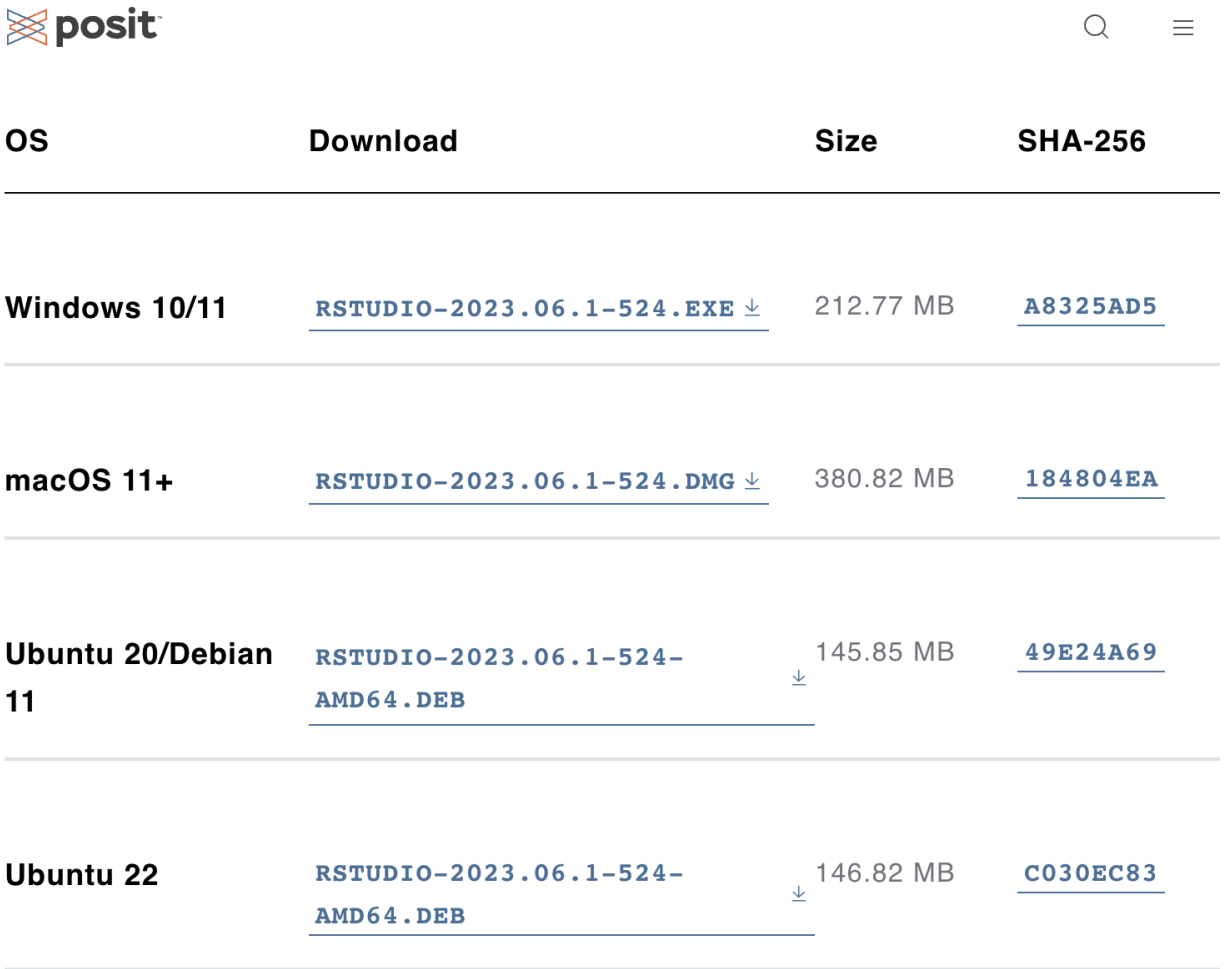
- The latest release (2022-06-23, Funny-Looking Kid) R-4.2.1.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. 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

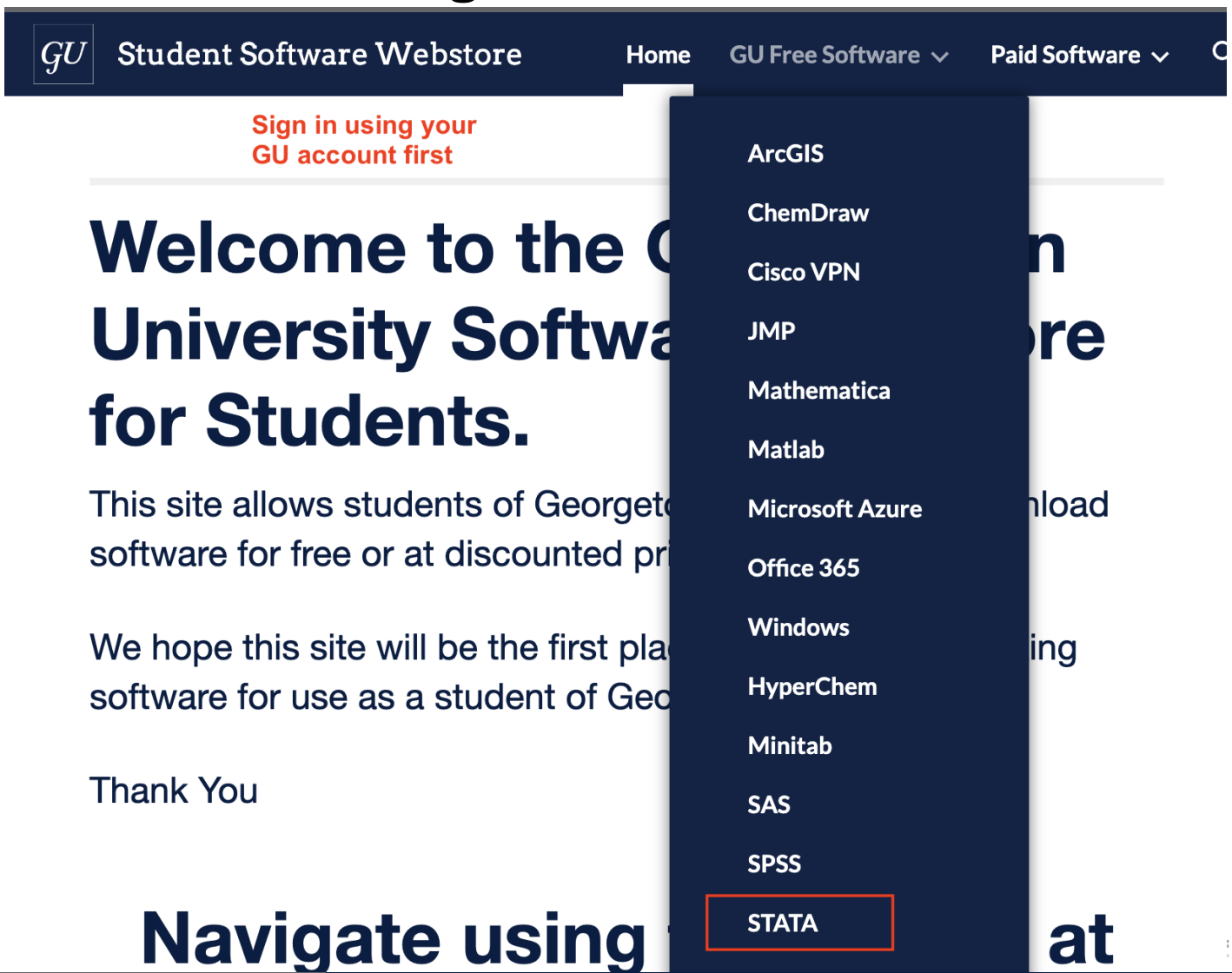
Installation: R Studio

Download R Studio



Installation: Stata

Georgetown Web Store



Installation: Stata

Students are able to use the University Site License for Stata BE. Stata SE can be used on University machines in libraries, departmental machines, and the UIS Virtual Classroom when additional variables are needed.

STATA License Request Link

Note: At the end of the Google Form, license information will be provided.

Installation: Stata

Home › Request New Service › Software › Stata License Request

Stata License Request

Request for Stata by StataCorp

Read more

Affiliation *

Student

Installation: Stata

Hi Ankushi,

This software requires UIS to collect data in regards to usage and agreement to licensing terms of the software developer. At the completion of the Google Form you will receive download and license information.

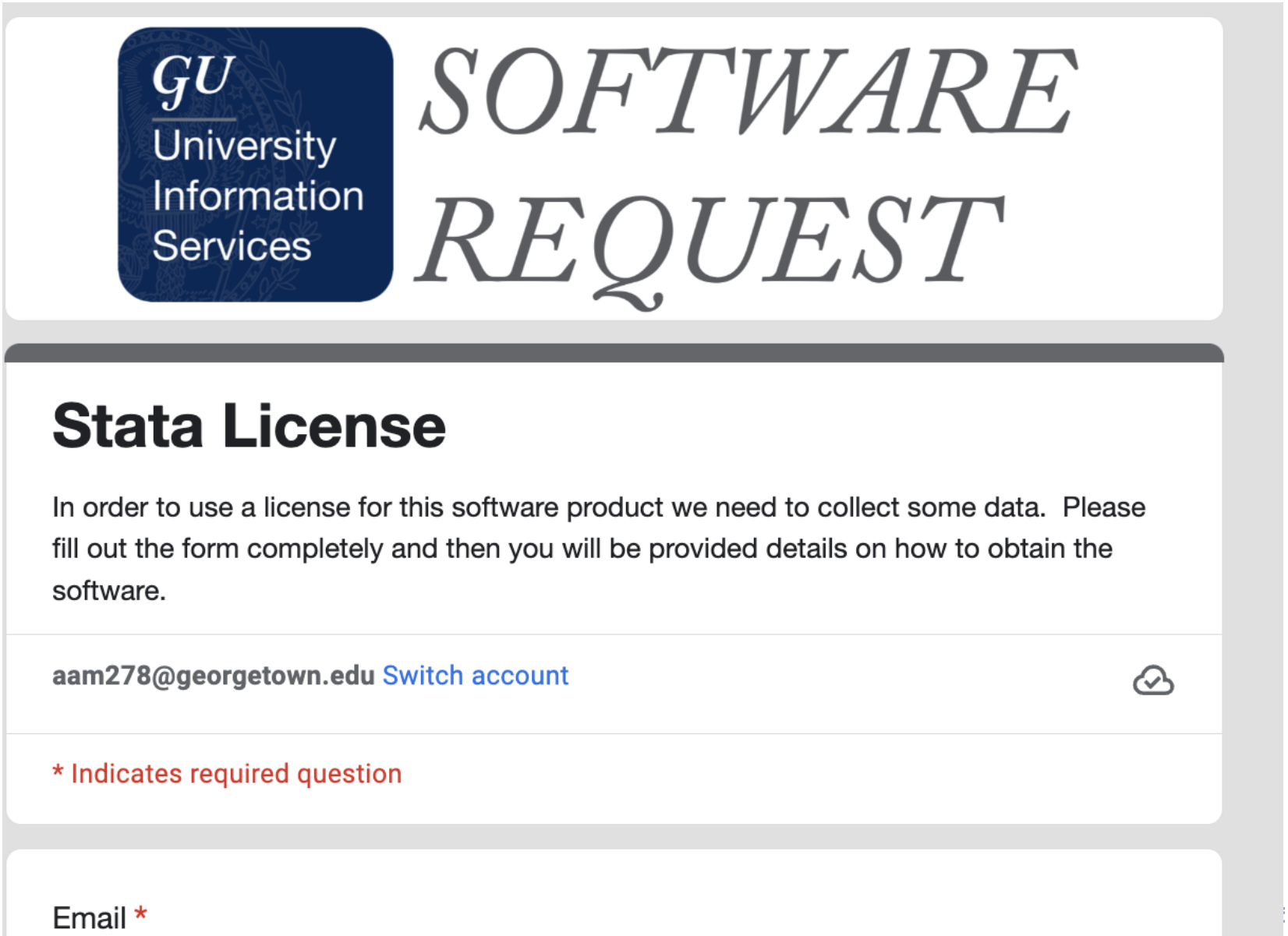
Please complete the google form below to gain access to the software.

Stata Data Collection and Vendor Agreement Form

Thank you,

Georgetown University Service Center

Installation: Stata



Installation: Stata

- Note your Serial Number, Code, and Authorization Key
- 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 (you will need to make an account on 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

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)

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)
- My preferred way to install LaTeX, if you want it primarily for RMarkdown, is to install it through R (TinyTex package)

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)
- 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()
```