GOVT PhD Math Camp Practice Day 1

Theodore Landsman

Georgetown University

August 16 2021

Quick Lecture: Habits of Effective Social Scientists

- We're going to be spending most of this session walking you through software installation and troubleshooting.

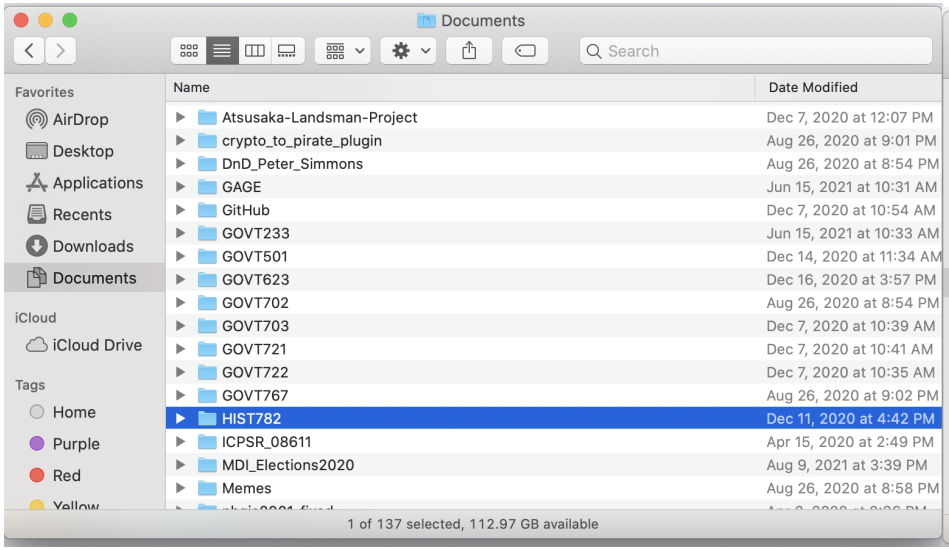
Quick Lecture: Habits of Effective Social Scientists

- We're going to be spending most of this session walking you through software installation and troubleshooting.
- But first, I wanted to do a quick lecture on how to be an effective social scientist.

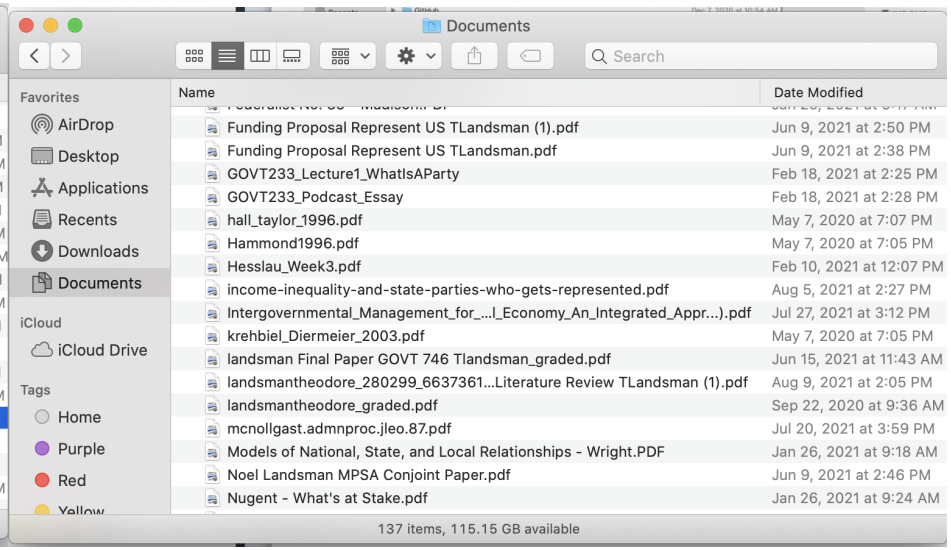
Quick Lecture: Habits of Effective Social Scientists

- We're going to be spending most of this session walking you through software installation and troubleshooting.
- But first, I wanted to do a quick lecture on how to be an effective social scientist.
- As academics, we have to structure our own work and time. Some professors will tell you that there is one perfect way to do this: E.G: Work in short bursts, POMODORO, etc. The goal for this presentation is merely to demonstrate workflow habits that will help keep you sane and organized as you figure out how generate motivation for one of these strategies.

Organize your File system!



(a) Right Way



(b) Wrong Way

Organize your File system! Cont.

- Whether you are using your personal file system, a cloud service, or both, files should be organized into projects with subfolders for different types of files (e.g: data, scripts, etc)

Organize your File system! Cont.

- Whether you are using your personal file system, a cloud service, or both, 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.

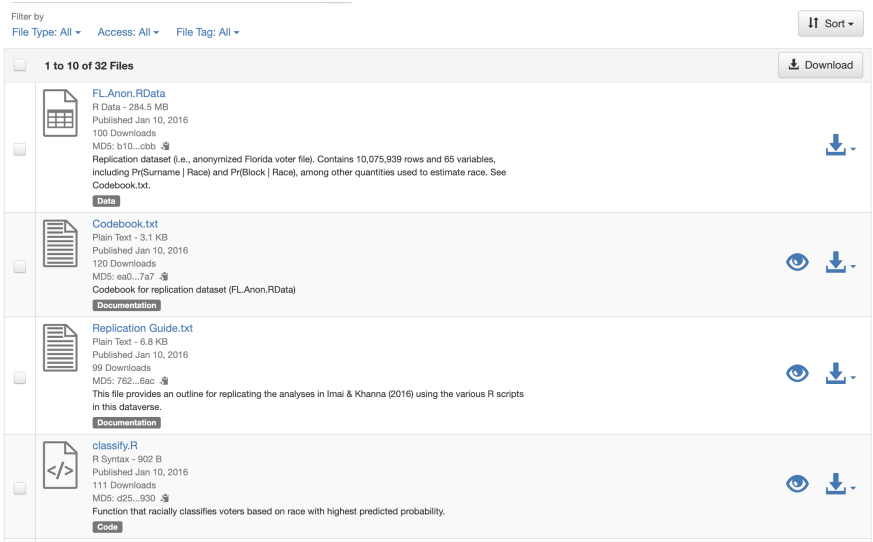
Organize your File system! Cont.

- Whether you are using your personal file system, a cloud service, or both, 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.
- Starting a project with an organized file system is much easier than cleaning it up afterward.

Goals


HARVARD Dataverse
Add Data
Search
About
User Guide
Support
Sign Up
Log In
POLITICAL ANALYSIS
Political Analysis Dataverse (Cambridge University Press)
Harvard Dataverse > Political Analysis Dataverse >
Replication Data for: Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records
Version 1.0
Khanna, Kabir; Imai, Kosuke, 2016, "Replication Data for: Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records", https://doi.org/10.7910/DVN/SVYSVF, Harvard Dataverse, V1
Cite Dataset
Learn about Data Citation Standards.
Access Dataset
Contact Owner
Share
Dataset Metrics
2,549 Downloads
Description
Replication files for Imai & Khanna (2016) "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records", including replication data, codebook, and R scripts to produce all analyses in the paper. Only academic use permitted. (2015-12-21)
Subject
Social Sciences
Keyword
Voter file, Race, Ethnicity, Prediction, Turnout
Related Publication
Forthcoming, Political Analysis doi: add DOI# here when available. If not available, please delete all text in this field before saving

(c)


Filter by
File Type: All
Access: All
File Tag: All
Sort
1 to 10 of 32 Files
Download
FL_Anon.RData
R Data - 284.5 MB
Published Jan 10, 2016
100 Downloads
MD5: b10...cbb
Replication dataset (i.e., anonymized Florida voter file). Contains 10,075,939 rows and 65 variables, including Pr(Surname | Race) and Pr(Block | Race), among other quantities used to estimate race. See Codebook.txt.
Data
Codebook.txt
Plain Text - 3.1 KB
Published Jan 10, 2016
120 Downloads
MD5: ea0...7a7
Codebook for replication dataset (FL_Anon.RData)
Documentation
Replication Guide.txt
Plain Text - 6.8 KB
Published Jan 10, 2016
99 Downloads
MD5: 762...6ac
This file provides an outline for replicating the analyses in Imai & Khanna (2016) using the various R scripts in this dataverse.
Documentation
classify.R
R Syntax - 902 B
Published Jan 10, 2016
111 Downloads
MD5: d25...930
Function that racially classifies voters based on race with highest predicted probability.
Code

(d)

File Types to Know

- Text files: .docx, .doc (Word), .txt (Plain text), .rtf (Rich Text), .md (Markdown), .tex (Latex).

File Types to Know

- Text files: .docx, .doc (Word), .txt (Plain text), .rtf (Rich Text), .md (Markdown), .tex (Latex).

- Data Files: .xls, .xlsx (Excel), .csv (Comma Seperated Values, data equivalent of plain text). .tsv (Tab Seperated Values, similar to .csv but a little more finicky), .json (java data format), .Rdata (R data format), .dta (Stata data format), .sav (SPSS file), .shp (GIS mapping data format)

File Types to Know

- Text files: .docx, .doc (Word), .txt (Plain text), .rtf (Rich Text), .md (Markdown), .tex (Latex).

- Data Files: .xls, .xlsx (Excel), .csv (Comma Seperated Values, data equivalent of plain text). .tsv (Tab Seperated Values, similar to .csv but a little more finicky), .json (java data format), .Rdata (R data format), .dta (Stata data format), .sav (SPSS file), .shp (GIS mapping data format)

- Code files: .r (r script), .do (Stata script), .rmd (R script with markdown for text), .rnw (R script with latex for text), .py (python script).

Conventions for Naming Files (And Variables)

Multiple-word identifier formats

Formatting	Name(s)
`twowords`	flat case[13][14]
`TWOWORDS`	upper flat case[13]
`twoWords`	(lower) camelCase, dromedaryCase
`TwoWords`	PascalCase, UpperCamelCase, StudlyCase[15]
`two_words`	snake_case, pothole_case
`TWO_WORDS`	SCREAMING_SNAKE_CASE, MACRO_CASE, CONSTANT_CASE
`two_Words`	camel_Snake_Case
`Two_Words`	Pascal_Snake_Case
`two-words`	kebab-case, dash-case, lisp-case
`TWO-WORDS`	TRAIN-CASE, COBOL-CASE, SCREAMING-KEBAB-CASE
`Two-Words`	Train-Case,[13] HTTP-Header-Case[16]

Conventions for Naming Files (And Variables) Cont.

- snake_case_is_the_most_common_in_political_science

Conventions for Naming Files (And Variables) Cont.

- snake_case_is_the_most_common_in_political_science
- unfortunately, snake$_c$*aseCanCauseProblemsInCertainTextEditors*(*likeLatex*).

Conventions for Naming Files (And Variables) Cont.

- snake_case_is_the_most_common_in_political_science
- unfortunately, snake$_c$*aseCanCauseProblemsInCertainTextEditors*(*likeLatex*).
- camelCase and UpperCamelCase are also very common in political science.

Conventions for Naming Files (And Variables) Cont.

- snake_case_is_the_most_common_in_political_science
- unfortunately, snake$_{c}$*aseCanCauseProblemsInCertainTextEditors*(*likeLatex*).
- camelCase and UpperCamelCase are also very common in political science.
- What convention you choose is less important than deploying it consistently.

Save Things to the Cloud for Easy Collaboration and so You Never Lose Hours of Work

- Basic: Google Drive, Dropbox.

Save Things to the Cloud for Easy Collaboration and so You Never Lose Hours of Work

- Basic: Google Drive, Dropbox.
- Advanced: Github, Google filestream, Overleaf.

Save Things to the Cloud for Easy Collaboration and so You Never Lose Hours of Work

- Basic: Google Drive, Dropbox.
- Advanced: Github, Google filestream, Overleaf.
- Things you can use if you are smarter than me: Amazon AWS, Georgetown MDI servers.

Save Things to the Cloud for Easy Collaboration and so You Never Lose Hours of Work

- Basic: Google Drive, Dropbox.
- Advanced: Github, Google filestream, Overleaf.
- Things you can use if you are smarter than me: Amazon AWS, Georgetown MDI servers.
- Many of these tools have version control, so you can access earlier iterations while avoiding the dreaded myProjectFinal_FINAL_(4)_V7.pdf problem.

R-Studio is your friend

- R-Studio has a projects functionality and integration with other tools such as Github, Markdown, and Latex.

R-Studio is your friend

- R-Studio has a projects functionality and integration with other tools such as Github, Markdown, and Latex.
- R-Studio has a built in code editor that is sufficiently state-of-the-art that people use it for non-R-code (like Python).

R-Studio is your friend

- R-Studio has a projects functionality and integration with other tools such as Github, Markdown, and Latex.
- R-Studio has a built in code editor that is sufficiently state-of-the-art that people use it for non-R-code (like Python).
- R-Studio will help you install R packages, these packages make R extremely *Extensible* (more on this later) and is part of why R is taking over the data science world, with new applications to problems like mapping that are rapidly surpassing existing tools (like GIS).

Installation Instructions

- Detailed Instructions for how to install R and STATA are available here: http://blogs.commons.georgetown.edu/government-math-camp/r-bootcamp/
- To verify that you have R and RStudio properly installed, open R Studio and type the following lines into the console at the bottom of the screen, one at a time, followed by the enter key.
- `install.packages("RXKCD")`
- `library(RXKCD)`
- `getXKCD(which = 2048)`
- You should now have the same xkcd comic on the web-page displayed in Rstudio plots panel.