Computing for the Social Sciences

MACS 30500

University of Chicago

Primer on the course

Origins

- Pre-2016
- Adapted from Jenny Bryan's Stat 545

Core objectives

- Fulfill programming requirement for certificate in computational social science
- Introduce graduate/undergraduate students to programming techniques essential for CSS
- Develop open-source resources for R learners
- Extend my R capabilities for new packages/methods (e.g. `blogdown`, `sf`, `tidytext`)

Current activity

- Two sections in fall/spring quarters
- 60 students per section
- Summer institute in social research methods
- 100K+ users of the course site in the past year

Major topics

First half

- Elementary programming techniques (e.g. loops, conditional statements, functions)
- Writing reusable, interpretable code
- Debugging programs
- Obtaining, importing, and munging data from a variety of sources
- Visualizing information
- Generating reproducible research

Second half

- Hodgepodge of applications in social science
- Performing statistical analysis
- Obtaining data via APIs and web scraping
- Text analysis
- Geospatial visualizations
- Shiny applications

Who is this class for?

Jeri

- PhD student in sociology
- Feels comfortable with regression and econometric methods in Stata
- Will be analyzing a large-scale dataset for her dissertation
- Seeks a reproducible workflow to manage all her exploratory and confirmatory analysis

Who is this class for?

Ryan

- Entering the MAPSS program
- Hasn't taken a statistics class in five years
- Expects to analyze a collection of published news articles
- Wants to understand code samples he finds online so he can repurpose them for his own work

Who is this class for?

Fernando

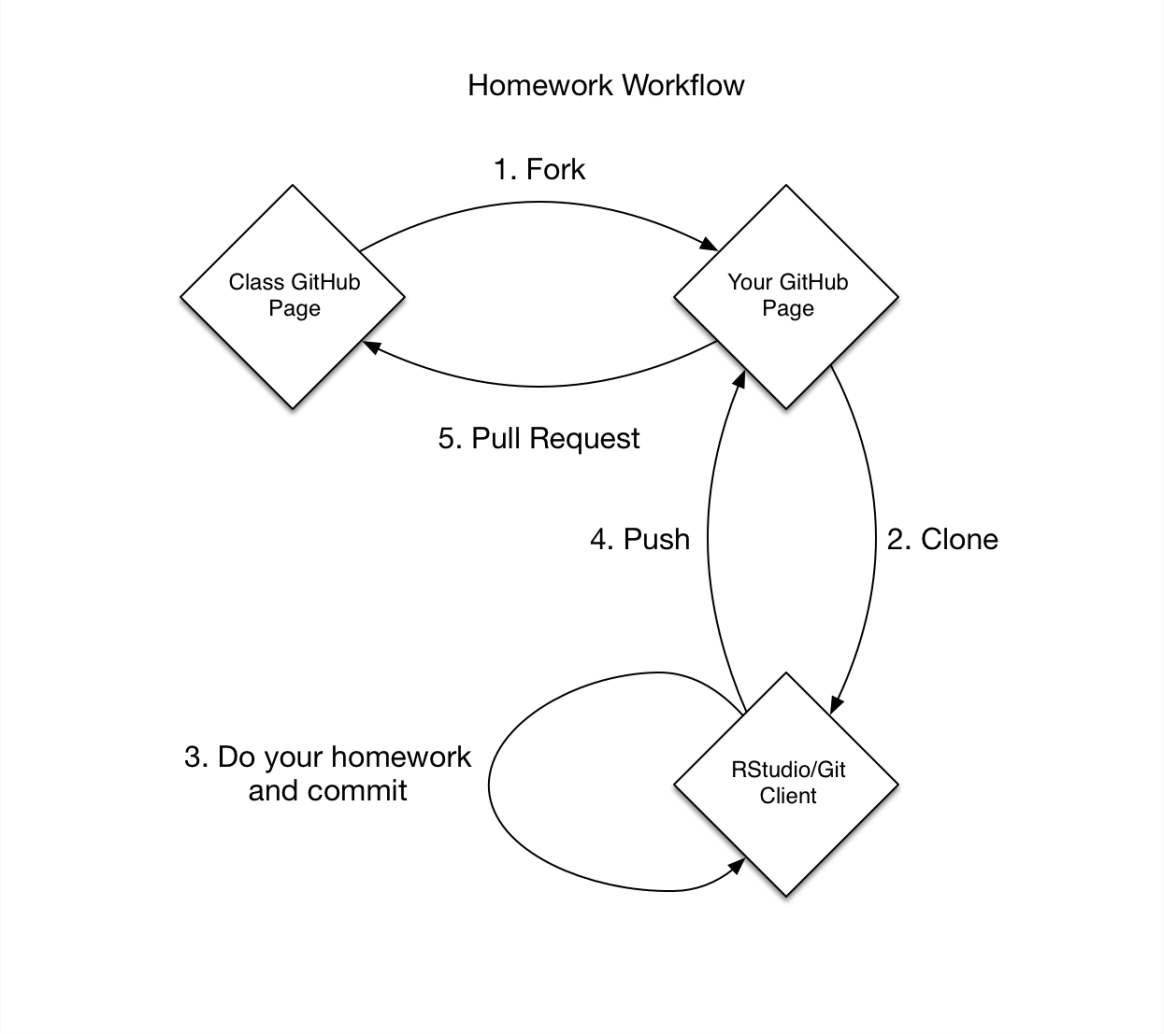
- Third-year student in the College
- Has taken MATH 15100 and 15200, plus the departmental methods course
- Wants to work as a research assistant on a project exploring the onset of civil conflict
- Needs to start contributing to a new research paper next quarter

Who is this class for?

Fang

- Born and raised in Shenzhen, China
- Background in psychology, plans to apply for doctoral programs in marketing
- Is going to run 300 experiments on Amazon MTurk in the next six months
- Expects to take courses in machine learning and Bayesian statistics which require a background in R

Homework workflow



Course site

> https://cfss.uchicago.edu

Lessons learned

- Everyone is capable of learning given sufficient resources
- Set expectations high and (most) students rise to the challenge
- As enrollment expands, automate where you can **if it is pedagogically sound**
- Tidy or base R? Good question!