

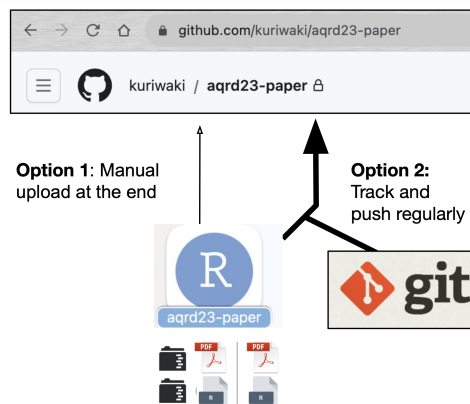
Class 12.2: Github + Synthesis

December 6, 2023

Going into this class, you should: (1) have a github account, (2) have your paper code, data, and final project template.


Goals for today + section: What is git? Create a github repo and upload files. Interact with *git* in Rstudio

You must upload your final project code and data through git. The main schematic:



- **GitHub** is a Microsoft-owned interface that visualizes and shares git
- **git** is the version tracking software

Make a Github account and repository

<https://github.com> → 

Create a new repository
 A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

Required fields are marked with an asterisk (*).

Repository template
 No template

Start your repository with a template repository's contents.

Owner * **Repository name ***
 kuriwaki / aqrd-paper
aqrd-paper is available.

Great repository names are short and memorable. Need inspiration? How about upgraded-sniffle?

Description (optional)





☐ **Public**
 Anyone on the internet can see this repository. You choose who can commit.

☒ **Private**
 You choose who can see and commit to this repository.

- Make a free Github account
- If keeping your repository private: add the teaching team as “Collaborators” (see PS-10 for our usernames)

Option 1: Manual upload at the end

Quick setup — if you've done this kind of thing before

 Set up in Desktop or  HTTPS  SSH <https://github.com/kuriwaki/aqrd-paper.git> 

Get started by [creating a new file](#) or [uploading an existing file](#). We recommend every repository include a [README](#), [LICENSE](#), and [.gitignore](#).

...or create a new repository on the command line

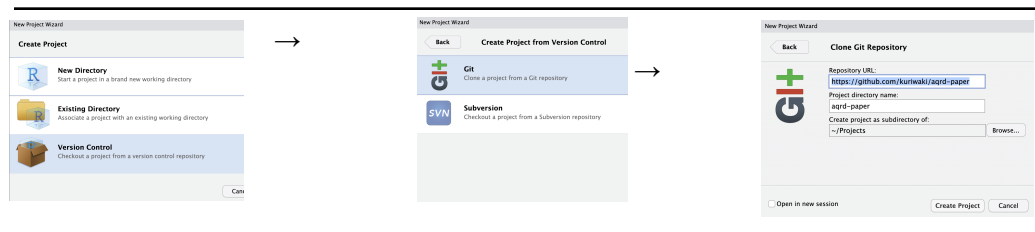
Option 2: Track and push regularly

1. Install git on your computer by downloading it (Chapter 6 of happygitwithr). Re-start Rstudio
2. Store your Github username and password to your computer (Chapter 7 of happygitwithr, while changing the `user.name` and `user.email` to your credentials).

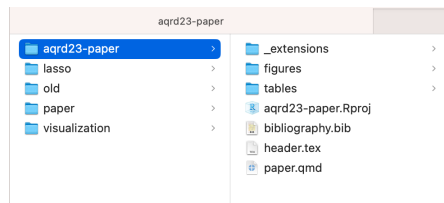
```
## install if needed (do this exactly once):
## install.packages("usethis")

library(usethis)
use_git_config(user.name = "Jane Doe", user.email = "jane@example.org")
```

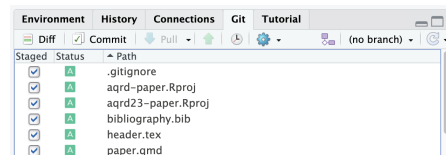
3. Create a local RStudio Project connected to the github repository



4. Move files to your local Rstudio project



5. In the git pane, add commit, then push files



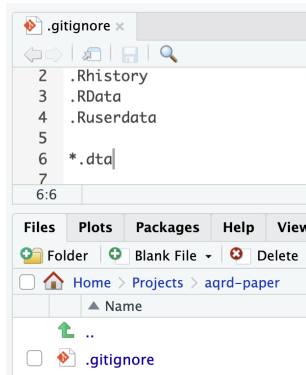
Git breaks apart the version saving process in three steps on a file-by-file basis:

1. **Add**: Preliminary step to stage the change
2. **Commit**: Permanently store the change into local git. Needs a user description.
3. **Push**: Upload the commit online (to the "remote")

6. As you edit your code, **commit** your modifications and keep pushing.

Pull becomes relevant when working with multiple local machines, multiple people, or branches (i.e., parallel versions)

(Not) Sharing Data



- git and Github can track datasets, but git is not designed for data
- Tracking large datasets slow things down
- One option: list file names and file name patterns to ignore in the `.gitignore` file

The example will pretend that files `.RData` and those that end with `.dta` do not exist, so they do not get tracked each time

Final Project

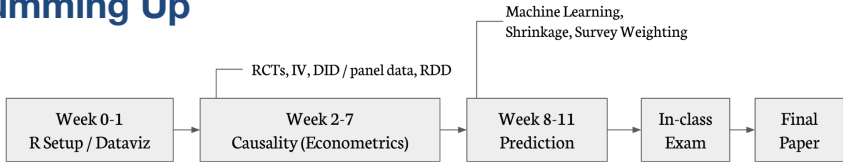
Ways to extend

- Different Data
 - Change outcome / collect newer time series
 - Different measures
 - Interesting subgroups
- Different Estimation Method
 - Different panel adjustment methods
 - Different weighting targets

FAQ

- Use your estimating equation “1 pager” option to show to me before submitting
- Appendix does not count towards 10 page limit
- Office hours

Summing Up



Experiments

A multifaceted program causes lasting progress for the very poor: Evidence from six countries

Abhijit Banerjee,^{1,2,3,4} Esther Duflo,^{1,2,3,4} Nathanael Goldberg,⁵ Dean Karlan,^{2,3,4,5,6} Robert Osei,⁷ William Parienté,^{4,8} Jeremy Shapiro,⁹ Bram Thuysbaert,^{1,10} Christopher Udry,^{3,5,6,8}

We present results from six randomized control trials of an integrated approach to improve livelihoods among the very poor. The approach combines the transfer of a productive asset with consumption support, training, and coaching plus savings encouragement and health education and/or services. Results from the implementation of the same basic program, adapted to a wide variety of geographic and institutional contexts and with multiple implementing partners, show statistically significant cost-effective impacts on consumption (fueled mostly by increases in self-employment income) and psychosocial status of the targeted households. The impact on the poor households listed at least a year after all implementation ended. It is possible to make sustainable improvements in the economic status of the poor with a relatively short-term intervention.

Observational Comparisons

Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact

Neil Malhotra^{Stanford University}
Yotam Margalit^{Columbia University}
Cecilia Hyunjung Mo^{Vanderbilt University}

What explains variation in individuals' opposition to immigration? While scholars have consistently shown cultural concerns to be strong predictors of opposition, findings regarding the labor-market competition hypothesis are highly contested. To help understand these divergent results, we distinguish between the prevalence and conditional impact of determinants of immigration attitudes. Leveraging a targeted sampling strategy of high-technology counties, we conduct a study of Americans' attitudes toward H-1B visas. The plurality of these visas are occupied by Indian immigrants, who are skilled but ethnically distinct, enabling us to measure a specific skill set (high technology) that is threatened by a particular type of immigrant (H-1B visa holders). Unlike recent aggregate studies, our targeted approach reveals that the conditional impact of the relationship in the high-technology sector between economic threat and immigration attitudes is sizable. However, labor-market competition is not a prevalent source of threat and therefore is generally not detected in aggregate analyses.

Instrumental Variables

Economic Shocks and Civil Conflict: An Instrumental Variables Approach

Estimating the impact of economic conditions on the likelihood of civil conflict is difficult because of endogeneity and omitted variable bias. We use rainfall variation as an instrumental variable for economic growth in 41 African countries during 1981-99. Growth is strongly negatively related to civil conflict: a negative growth shock of five percentage points increases the likelihood of conflict by one-half the following year. We attempt to rule out other channels through which rainfall may affect conflict. Surprisingly, the impact of growth shocks on conflict is *not* significantly different in richer, more democratic, or more ethnically diverse countries.

Time Series

Democracy, War, and Wealth: Lessons from Two Centuries of Inheritance Taxation

KENNETH SCHEVE^{Yale University}
DAVID STASAVAGE^{New York University}

In this article we use an original data set to provide the first empirical analysis of the political economy of inherited wealth taxation that covers a significant number of countries and a long time frame (1816-2000). Our goal is to understand why, if inheritance taxes are often very old taxes, the implementation of inheritance tax rates significant enough to affect wealth inequality is a much more recent phenomenon. We hypothesize alternatively that significant taxation of inherited wealth depended on (1) the extension of the suffrage and (2) political conditions created by mass mobilization for war. Using a difference-in-differences framework for identification, we find little evidence for the suffrage hypothesis but very strong evidence for the mass mobilization hypothesis. Our study has implications for understanding the evolution of wealth inequality and the political conditions under which countries are likely to implement policies that significantly redistribute wealth and income.

Regression Discontinuity

Payments for environmental services supported social capital while increasing land management

Jennifer M. Alix-Garcia[✉], Katharine B. E. Sims[✉], Victor H. Orozco-Olvera[✉], and Sofia Bioma-Monroy[✉] [Authors Info & Affiliations](#)

Abstract
Payments for environmental services (PES) programs incentivize landowners to protect or improve natural resources. Many conservationists fear that introducing compensation for actions previously offered voluntarily will reduce social capital (the institutions, relationships, attitudes, and values that govern human interactions), yet little rigorous research has investigated this concern. We examined the land cover management and communal social capital impacts of Mexico's federal conservation payments program, which is a key example for other countries committed to reducing deforestation, protecting watersheds, and conserving biodiversity. We used a regression discontinuity (RD) methodology to identify causal program effects, comparing outcomes for PES

Prediction

Improving Worker Safety in the Era of Machine Learning (A)

It had been a week since Alex Seguro joined the new data science team in at the Occupational Safety and Health Administration (OSHA), the U.S. federal agency in charge of occupational safety. She was excited to begin her first project assignment: to join a team that would evaluate several alternative approaches to how OSHA should target its inspections to best deliver on its mission of assuring safe working conditions. The project team had just received a dataset to conduct this study (described in

Sparse Surveys

The Geography of Racially Polarized Voting: Calibrating Surveys at the District Level

SHIRO KURIWAKI^{Yale University, United States}
STEPHEN ANSOLABEHRE^{Harvard University, United States}
ANGELO DAGONEL^{Harvard University, United States}
SOICHIRO YAMAUCHI^{Independent Scholar}

Debates over racial voting, and over policies to combat vote dilution, turn on the extent to which groups' voting preferences differ and vary across geography. We present the first study of racial voting patterns in every congressional district (CD) in the United States. Using large-sample surveys combined with aggregate demographic and election data, we find that national-level differences across racial groups explain 60% of the variation in district-level voting patterns, whereas geography explains 30%. Black voters consistently choose Democratic candidates across districts, whereas Hispanic and white voters' preferences vary considerably across geography. Districts with the highest racial polarization are concentrated in the parts of the South and Midwest. Importantly, multiracial coalitions have become the norm: in most CDs, the winning majority requires support from non-white voters. In arriving at these conclusions, we make methodological innovations that improve the precision and accuracy when modeling sparse survey data.

$$Y_i = \alpha + \beta D_i + \gamma X_i + \varepsilon_i$$

where coefficients α, β, γ selected to minimize squared error $\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$

The research design changes the interpretation of the difference-in-means

- Randomize D makes the difference equivalent to avg. causal effect, by eliminating selection bias
- D compares divergence in previously parallel trends

The estimation procedure improves the precision of our estimate

Goal: Master the quantitative social science toolkit

(Write an awesome senior essay / thesis / journal article)

1. **Statistical Research Designs:**
 - Necessary to convince a skeptical audience
2. **Tools** (tidyverse, replicable project form, Github):
 - Necessary to implement your project efficiently and correctly
3. **Literature:**
 - Necessary to innovate: you need to know the forefront inside-out

(from the first day of our class in August)

Lessons I Hope You Take Away

0. Open source software and data are powerful

1. Statistics = using data you have to learn about data you don't have

What is the thing we want to observe, but can't directly?

- Would the employer have hired the person if the name on the résumé sounded Anglo-Saxon?
- Would the UK have implemented progressive taxation in 1920 had it not joined WWI?

2. What are the assumptions underlying your inference?

- Are potential outcomes Y_0 same across groups? Would the student who went to an Ivy have done better *anyways*?
- Does matching similar units by *statistical control* make the assumption more plausible?
- Is the survey representative? If not, can covariate adjustment help?

Violations of assumptions do not mean the study is completely invalid, but it introduces bias

3. Remember: Choices between statistical methods involve tradeoffs

- Internal vs. external validity
- Reducing bias vs. Reducing variance

Thanks to the guest speakers, and the teaching fellows Changwook, Robin, and Sean, and to you for your efforts and engagement. Please fill out our custom feedback survey, work hard on the final paper, and keep in touch. – Shiro

PS-10

Problem 3

I am continually thinking of ways to improve the effectiveness of the course. At the end of this month, you will evaluate the course through Yale's official course evaluation. The following survey is different: it asks you about specific aspects of the course that might not be reflected in the official course evals.

https://yalesurvey.ca1.qualtrics.com/jfe/form/SV_1T6DJsi0X1K12XI

(please take enter this survey after the last day of class on Dec 6).