Git for Research in Economics

Giacomo Battiston   Jared Gars

3 June 2021
LEAP Development Coffee

Version Control

Research in economics heavily relies on coding in all phases of projects:

1. data collection: e.g., python for scraping, python and R for NLP
2. econometric analysis: e.g. Stata and R
3. writing: e.g. Latex

Managing the workflow requires keeping track of code versions:

▶ avoiding conflicts and keeping track of most recent version
▶ documenting changes for future self (or future collaborators)

Widespread shift in economics towards reproducible results

$\rightarrow$ we need a clean and quick way to share code with external users

What we need

In managing versions, we need to:

- ▶ synchronize code with collaborators/multiple workstations
- ▶ back-up projects (not just files)
- ▶ track changes
- ▶ store comments
- ▶ manage code experiments

What we often do

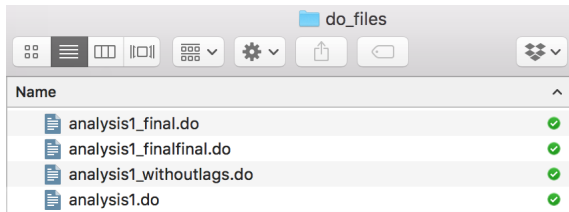Keep code and data in a Dropbox folder.

What we need

In managing versions, we need to:

- ▶ synchronize code with collaborators/multiple workstations ✓
- ▶ back-up for projects (not just files) ✓/✗

What we often do

Keep code and data in a Dropbox folder.

Apply "hacks" to deal with version control:

What we often do

Keep code and data in a Dropbox folder.

Apply "hacks" to deal with version control:

```
*SPECICATION IN THE DRAFT
*reg price weight i.year i.month, r

*TENTATIVE SPECIFICATION ADDED ON OCTOBER 3, 2021: CHECK BEFORE SUBMISSION
reg price weight i.year i.week, r
```

What we often do

Keep code and data in a Dropbox folder.

Apply "hacks" to deal with version control:

```
*SPECICATION IN THE DRAFT
*reg price weight i.year i.month, r

*TENTATIVE SPECIFICATION ADDED ON OCTOBER 3, 2021: CHECK BEFORE SUBMISSION
reg price weight i.year i.week, r
```

Easy to mistake, difficult to keep track of why we did stuff (technical debt).

What we need

In managing versions, we need to:

- ▶ synchronize code with collaborators/multiple workstations ✓
- ▶ back-up for projects (not just files) ✓/✗
- ▶ track changes ✗
- ▶ store comments ✗
- ▶ manage code experiments ✗

Git and GitHub

Git is a widely software for version control.

- ▶ it keeps (restorable) versions of your whole project
- ▶ compares versions of the same code (differences in lines)
- ▶ keeps track of comments with changes
- ▶ finds conflicts
- ▶ allows branching

Git works locally, synchronization requires a shared repository.

GitHub is a commonly used repository for Git.

References

Jesus Fernandez-Villaverde, 2020. Version Control: Git. *Computational Methods for Economists.*
https://www.sas.upenn.edu/~jesusfv/Chapter_HPC_5_Git.pdf

Frank Pinter, 2019. Git: A Guide for Economists. *Presentation.* https://www.frankpinter.com/notes/git-for-economists-presentation.pdf

Matteo Sostero, 2018. Git Better: Collaborative project management using Git and GitHub. *Presentation.*
https://matteosostero.com/files/slides_git.pdf

Francis DiTraglia, 2015. git-for-economists. *Git Repository.*
https://github.com/fditraglia/git-for-economists