Setup

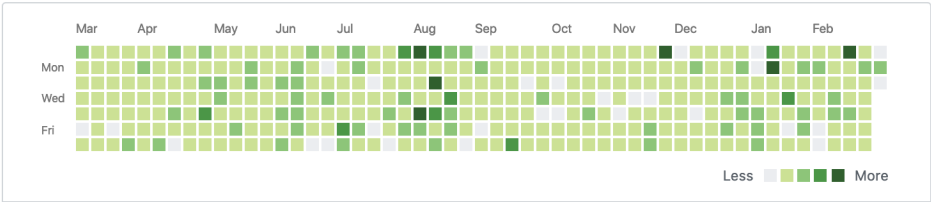
1. Go to www.github.com and make a free account
2. Make sure you have a recent version (v1.1 or later) of RStudio https://www.rstudio.com/products/rstudio/download/#download
3. Keep www.happygitwithr.com open
4. Download these slides via: https://github.com/kuriwaki/github-demo/raw/master/presentation-slides/kuriwaki_github.pdf

Happy Git and GitHub for the useR

Jenny Bryan, the STAT 545 TAs, Jim Hester

Let's Git started

2,412 contributions in the last year



Introduction to git for social science students

(not software developers)

Shiro Kuriwaki

March 5, 2019

Thanks for having me

About me

- G-4 in Government
- American Politics, elections and representation
- Before: A year in political data analytics (where I learned git from Annie Wang)

Thanks for having me

About me

- G-4 in Government
- American Politics, elections and representation
- Before: A year in political data analytics (where I learned git from Annie Wang)
- I do some software development,
- but most of my work is applied ("substantive")

Thanks for having me

About me

- G-4 in Government
- American Politics, elections and representation
- Before: A year in political data analytics (where I learned git from Annie Wang)
- I do some software development,
- but most of my work is applied ("substantive")

My perspective

- Version control is mandatory for programmers (and professional data scientists)

Thanks for having me

About me

- G-4 in Government
- American Politics, elections and representation
- Before: A year in political data analytics (where I learned git from Annie Wang)
- I do some software development,
- but most of my work is applied ("substantive")

My perspective

- Version control is mandatory for programmers (and professional data scientists)
- but does it make sense for *applied* researchers
- who work with datasets that are

Thanks for having me

About me

- G-4 in Government
- American Politics, elections and representation
- Before: A year in political data analytics (where I learned git from Annie Wang)
- I do some software development,
- but most of my work is applied ("substantive")

My perspective

- Version control is mandatory for programmers (and professional data scientists)
- but does it make sense for *applied* researchers
- who work with datasets that are **large**,

Thanks for having me

About me

- G-4 in Government
- American Politics, elections and representation
- Before: A year in political data analytics (where I learned git from Annie Wang)
- I do some software development,
- but most of my work is applied ("substantive")

My perspective

- Version control is mandatory for programmers (and professional data scientists)
- but does it make sense for *applied* researchers
- who work with datasets that are **large**, **unstructured**,

Thanks for having me

About me

- G-4 in Government
- American Politics, elections and representation
- Before: A year in political data analytics (where I learned git from Annie Wang)
- I do some software development,
- but most of my work is applied ("substantive")

My perspective

- Version control is mandatory for programmers (and professional data scientists)
- but does it make sense for *applied* researchers
- who work with datasets that are **large**, **unstructured**, **prone to change**,

Thanks for having me

About me

- G-4 in Government
- American Politics, elections and representation
- Before: A year in political data analytics (where I learned git from Annie Wang)
- I do some software development,
- but most of my work is applied ("substantive")

My perspective

- Version control is mandatory for programmers (and professional data scientists)
- but does it make sense for *applied* researchers
- who work with datasets that are **large**, **unstructured**, **prone to change**, **with collaborators**

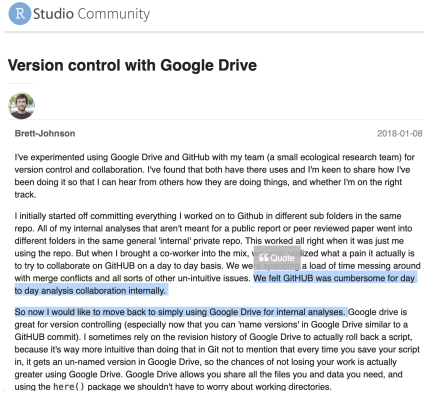
Setting Expectations: Is it worth it?

What do Gentzkow and Shapiro say?

Definitely:

> *"It will probably take you a couple days to set up a repository and learn how you want to interact with [Version Control]. You will break even on that time investment within a month or two."*[1]

but also see[2]



---

[1] Code and Data for Social Sciences: A Practioners Guide. 2014. https://perma.cc/5J9D-BTD6. Although I'm not sure about learning version control in "a couple of days" (I certainly couldn't!), I can guarantee reading their guide in its entirety *is* a time investment you'll break even on immediately.

[2] https://community.rstudio.com/t/version-control-with-google-drive/4032

My (recommended) setup

Terminology

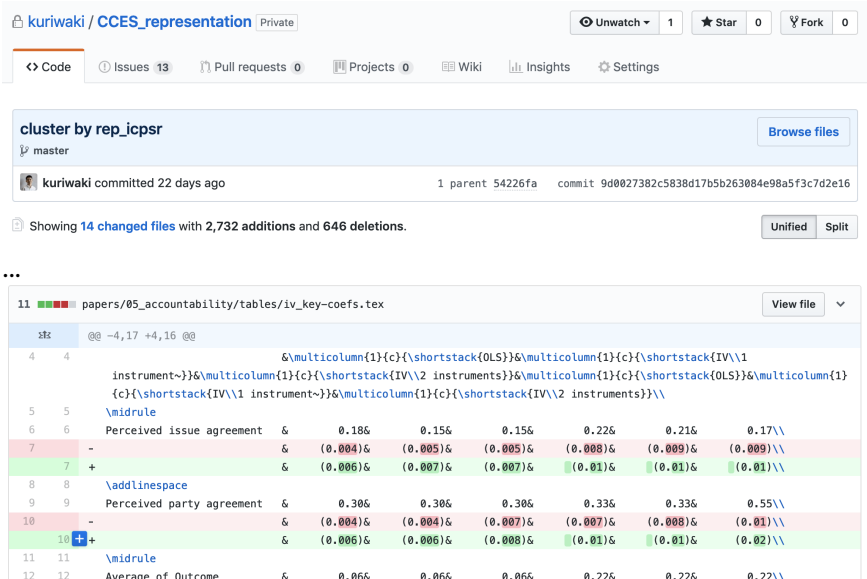
- **Git** is a particular type of software for version control (Subversion is another)
- **GitHub** is an app (recently bought by Microsoft) to host git on the web (Bitbucket is another)
- A **desktop client** is an app that connects a webhost like Github to your computer and facilitates simple tasks (here I use **RStudio**, there are many others)
- A **repository** is the fundamental unit of a version control, like a project folder.

Terminology

- **Git** is a particular type of software for version control (Subversion is another)
- **GitHub** is an app (recently bought by Microsoft) to host git on the web (Bitbucket is another)
- A **desktop client** is an app that connects a webhost like Github to your computer and facilitates simple tasks (here I use **RStudio**, there are many others)
- A **repository** is the fundamental unit of a version control, like a project folder. Do not make a repository within a repository!

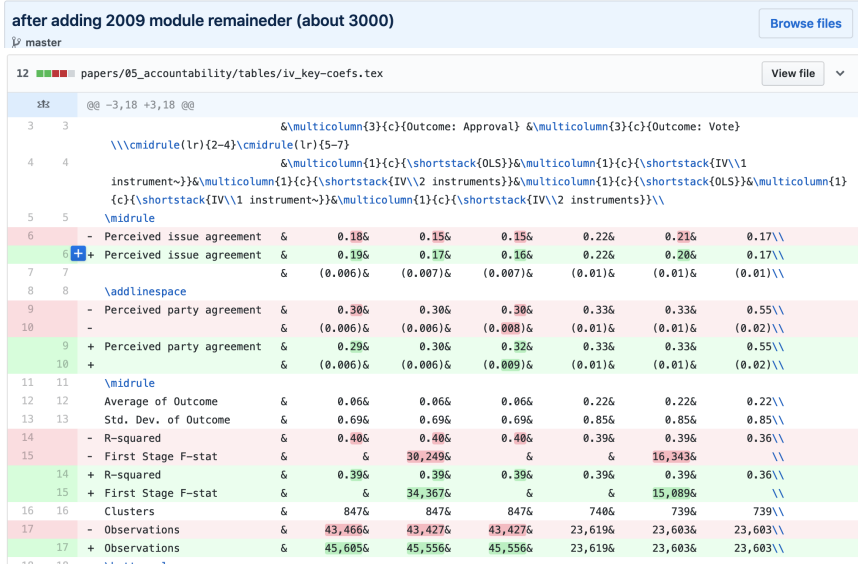
Keep Track of how your results changed

Problem: You tweak a regression specification and re-run your script, re-writing dozens of tables. You need to know how much your results changed



Keep Track of how your results changed

You collect more data and re-run the regressions. Now how did the results change?



after adding 2009 module remaineder (about 3000)

Browse files

master

12 papers/05_accountability/tables/iv_key~coefs.tex

View file

```
@@ -3,18 +3,18 @@
3  3                                  &\multicolumn{3}{c}{Outcome: Approval} &\multicolumn{3}{c}{Outcome: Vote}
                                       \\\cmidrule(lr){2-4}\cmidrule(lr){5-7}
4  4                                  &\multicolumn{1}{c}{\shortstack{OLS}}&\multicolumn{1}{c}{\shortstack{IV\\1
                                       instrument~}}&\multicolumn{1}{c}{\shortstack{IV\\2 instruments}}&\multicolumn{1}{c}{\shortstack{OLS}}&\multicolumn{1}
                                       {c}{\shortstack{IV\\1 instrument~}}&\multicolumn{1}{c}{\shortstack{IV\\2 instruments}}\\
5  5   \midrule
6     - Perceived issue agreement   &       0.18&       0.15&       0.15&       0.22&       0.21&       0.17\\
      6 + Perceived issue agreement   &       0.19&       0.17&       0.16&       0.22&       0.20&       0.17\\
7  7                                  &   (0.006)&   (0.007)&   (0.007)&   (0.01)&   (0.01)&   (0.01)\\
8  8   \addlinespace
9     - Perceived party agreement   &       0.30&       0.30&       0.30&       0.33&       0.33&       0.55\\
10    -                              &   (0.006)&   (0.006)&   (0.008)&   (0.01)&   (0.01)&   (0.02)\\
      9 + Perceived party agreement   &       0.29&       0.30&       0.32&       0.33&       0.33&       0.55\\
      10 +                            &   (0.006)&   (0.006)&   (0.009)&   (0.01)&   (0.01)&   (0.02)\\
11 11   \midrule
12 12   Average of Outcome           &       0.06&       0.06&       0.06&       0.22&       0.22&       0.22\\
13 13   Std. Dev. of Outcome         &       0.69&       0.69&       0.69&       0.85&       0.85&       0.85\\
14    - R-squared                    &       0.40&       0.40&       0.40&       0.39&       0.39&       0.36\\
15    - First Stage F-stat           &          &     30,249&          &          &     16,343&          \\
      14 + R-squared                  &       0.39&       0.39&       0.39&       0.39&       0.39&       0.36\\
      15 + First Stage F-stat         &          &     34,367&          &          &     15,089&          \\
16 16   Clusters                     &       847&       847&       847&       740&       739&       739\\
17    - Observations                 &     43,466&     43,427&     43,427&     23,619&     23,603&     23,603\\
      17 + Observations               &     45,605&     45,556&     45,556&     23,619&     23,603&     23,603\\
```

Tracking your text changes

Problem: You start writing up your paper, `draft.tex`

- The next day, you make a new draft. Do you overwrite?

Tracking your text changes

Problem: You start writing up your paper, `draft.tex`

- The next day, you make a new draft. Do you overwrite?
- Or do you call it `draft_0305.tex` ? `draft_03052019.tex`?

Tracking your text changes

Problem: You start writing up your paper, `draft.tex`

- The next day, you make a new draft. Do you overwrite?
- Or do you call it `draft_0305.tex` ? `draft_03052019.tex`?
- The next week, you find a single typo. Do you “Save As” with a new date?

Tracking your text changes

Problem: You start writing up your paper, `draft.tex`

- The next day, you make a new draft. Do you overwrite?
- Or do you call it `draft_0305.tex` ? `draft_03052019.tex`?
- The next week, you find a single typo. Do you “Save As” with a new date?
- Three weeks later, you return to your paper. Your computer indicates that the file named `draft_0305.tex` was “Last modified March 12, 2019”.

Tracking your text changes

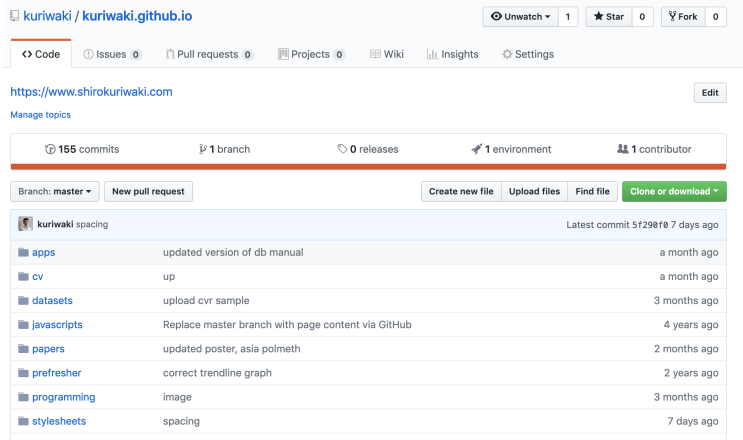
Problem: You start writing up your paper, `draft.tex`

- The next day, you make a new draft. Do you overwrite?
- Or do you call it `draft_0305.tex` ? `draft_03052019.tex`?
- The next week, you find a single typo. Do you “Save As” with a new date?
- Three weeks later, you return to your paper. Your computer indicates that the file named `draft_0305.tex` was “Last modified March 12, 2019”.

And more cool stuff like

Getting a free, customizable, add-free website

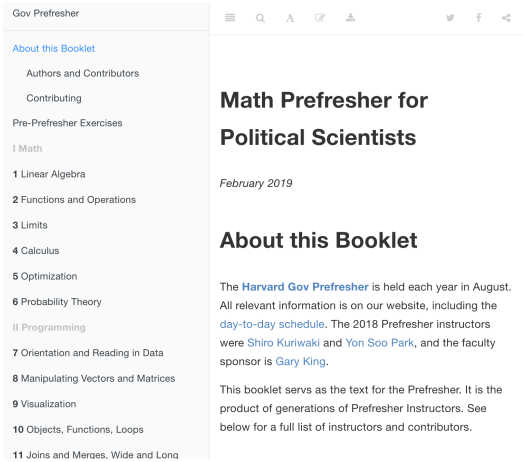
(instead of a click-and-drag Wordpress/Squarespace website)



And more cool stuff like

Work on a collaborative workbook

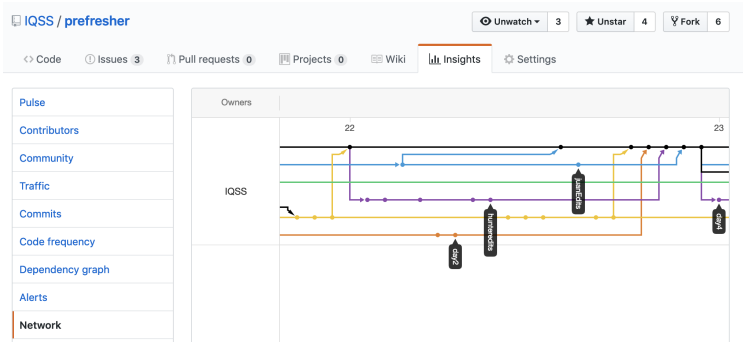
(instead of needing to add people to your Dropbox)



And more cool stuff like

Work on a collaborative workbook

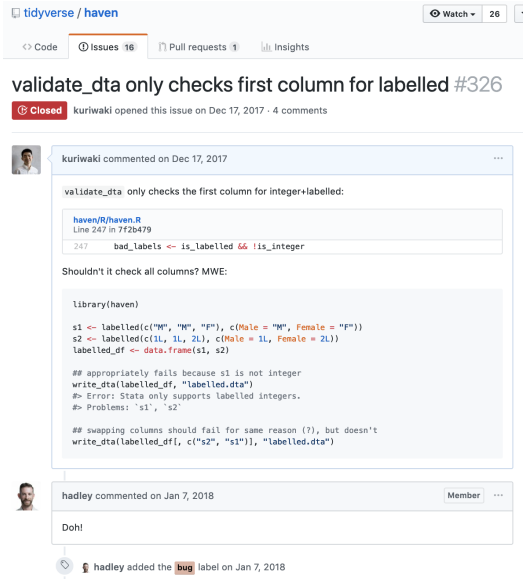
(instead of needing to add people to your Dropbox)



And more cool stuff like

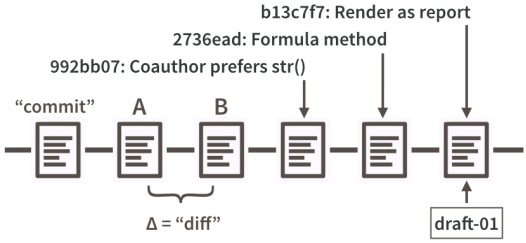
Contributing to / getting the latest on actual software packages

(Github issues is the de facto communication of open-source developers)



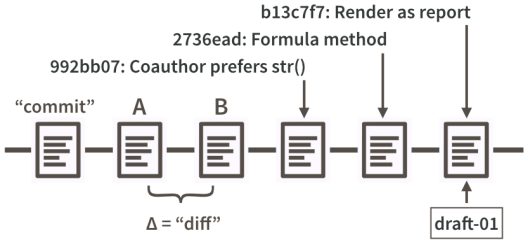
Terminology

- Files increment by **commit**s. The line-by-line changes from commits are called **diffs**.



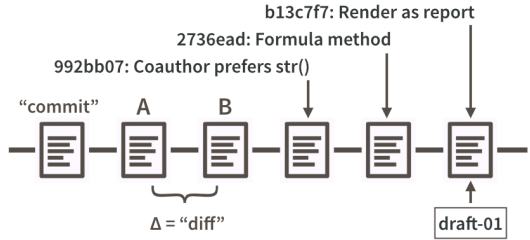
Terminology

- Files increment by **commit**s. The line-by-line changes from commits are called **diffs**.
- Commits have a human-readable **message**, and a serial code called a **SHA** (like `992bb07`).



Terminology

- Files increment by **commit**s. The line-by-line changes from commits are called **diffs**.
- Commits have a human-readable **message**, and a serial code called a **SHA** (like `992bb07`).
- At least two copies of your repository exist: the **local** on your computer, and a **remote** (hosted on Github, with URL `https://github.com/user/repo.git`), which has the name **origin**



Terminology

- Files increment by **commit**s. The line-by-line changes from commits are called **diffs**.
- Commits have a human-readable **message**, and a serial code called a **SHA** (like `992bb07`).
- At least two copies of your repository exist: the **local** on your computer, and a **remote** (hosted on Github, with URL `https://github.com/user/repo.git`), which has the name **origin**
- Once you make commits on your local, you **push** them to your remote. (The opposite of this is a **pull**)