

# Project Workflow

# Project Workflow - Motivation



# Project Workflow - Automation

- Gentzkow and Shapiro (2014): Automation
  - ▶ Automate everything that can be automated.
  - ▶ Write a single script that executes all code from beginning to end.
- Application:
  - ▶ README file
  - ▶ Master script file

# Automation - README File

- Purpose: orientation and instruction
  - ▶ Start Here
- Contents:
  - ▶ Project description
  - ▶ Instructions for replication (master script file)
  - ▶ Instructions for data access
  - ▶ Special considerations for software prep (e.g., user written Stata commands)

# Automation - README File

- Purpose: orientation and instruction
  - ▶ Start Here
- Contents:
  - ▶ Project description
  - ▶ Instructions for replication (master script file)
  - ▶ Instructions for data access
  - ▶ Special considerations for software prep (e.g., user written Stata commands)
- Example:
  - ▶ [Hollingsworth, A., Ruhm, C. J., & Simon, K. \(2017\). Macroeconomic Conditions and Opioid Abuse. Journal of Health Economics, 56, 222–233.](#)

# Automation - Master File

- Purpose: single project file that runs all individual project files
  - ▶ Roadmap for the analytic process
- Contents:
  - ▶ Series of well-annotated statements that run script files for data cleaning, analysis, table/figure creation, etc.
  - ▶ Includes commands to clear output files/tables/figures.

# Automation - Master File

- Purpose: single project file that runs all individual project files
  - ▶ Roadmap for the analytic process
- Contents:
  - ▶ Series of well-annotated statements that run script files for data cleaning, analysis, table/figure creation, etc.
  - ▶ Includes commands to clear output files/tables/figures.
- Example:
  - ▶ Goodman-Bacon, A. (2021). The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes. *American Economic Review*, 111(8): 2550-93.

# Project Workflow - Version Control

- Gentzkow and Shapiro (2014): Version Control
  - ▶ Store code and data under version control.
  - ▶ Run the whole directory before checking it back in.
- Application:
  - ▶ File repository (GitHub)



# Project Workflow - Version Control

- Purpose: avoid this

cleandata_022113.do	cleandata_022613.do	regressions.log
cleandata_022113a.do	cleandata_022613_jms.do	regressions_022413.do
chips.csv	tvdata.dta	regressions_022713_mg.do
regressions_022413.log		

- How it works:
  - ▶ Maintain one file (e.g., regressions.log).
  - ▶ File changes are tracked each time the file is "checked back in" (i.e., commit changes and push to GitHub).
  - ▶ If you screwed up, revert back to the previous commit.

# Project Workflow - Version Control

RStudio: Review Changes

Changes History main (all commits)

Search

Pull

Tutorial 6 update	k-callison <kcallison@tulane.edu>	2024-02-29	58e07a8e
Tutorial 6 update	k-callison <kcallison@tulane.edu>	2024-02-29	a8c40ea7
Tutorial 6 edit	k-callison <kcallison@tulane.edu>	2024-02-29	58e07a8e
Add tutorial 6	k-callison <kcallison@tulane.edu>	2024-02-29	476d7179
Add assignment 4	k-callison <kcallison@tulane.edu>	2024-02-26	dead8682
data assignment 2 fixes	k-callison <kcallison@tulane.edu>	2024-02-20	dac97571
Chip slides edit	k-callison <kcallison@tulane.edu>	2024-02-19	22710f07
Add Chip slides	k-callison <kcallison@tulane.edu>	2024-02-19	f3a8f29b

Commits 1-100 of 238

SHA 58e07a8e55d532f0d74f2b14144ac9e42271c41

Author k-callison <kcallison@tulane.edu>

Date (UTC) 2024-02-29 15:16

Subject Tutorial 6 edit

Parent 476d7179489dfa3ac5386f24cee0c463ecaa20eb

[assignments/tutorial\\_6.qmd](#)

[assignments/tutorial\\_6.qmd](#)

View file @ 58e07a8e

@@ -7,13 +7,15 @@ tbl-cap-location: bottom

7 7

8 8 ## Refining our Analytic Sample

9 9

10 10 In the previous tutorial, we calculated crude cancer mortality rates across Louisiana parishes from 2005 through 2019. Now we'll ....

10 10 In the previous tutorial, we calculated crude cancer mortality rates across Louisiana parishes from 2005 through 2019. Now we'll refine our method for calculating those rates by adjusting for different age distributions across parishes and over time.

11 11

12 12 Let's start by opening the .Rproj file in your 'hpam7660.Cancer.Alley' folder. Then open the Markdown document you used in the previous tutorial to create the analytic sample. We can continue working off of this same Markdown document as we refine our sample.

12 12 It's important to age-adjust our mortality rates because different parishes have different age profiles and that could bias our estimated rates. For example, suppose that parish A has a relatively high proportion of older people and parish B has a relatively high proportion of younger people. Crude cancer mortality rates would likely show higher rates among parish A than parish B. However, this may not actually be the case once we account for the fact that the older population in parish A is more prone to cancer.

13 13

14 14 ### Age-Adjusted Cancer Mortality Rates

14 14 Age-adjustment is really a way to determine how cancer mortality rates would differ between parish A and parish B if both parishes had the same population age distribution.

15 15

16 16 Let's start by opening the .Rproj file in your 'hpam7660.Cancer.Alley' folder. Then open the Markdown document you used in the previous tutorial to create the analytic sample. We can continue working off of this same Markdown document as we refine our sample in this tutorial.

15 17

16 16 The first step in refining our analytic sample is to calculate age-adjusted cancer mortality rates by parish. It's important to age-adjust our mortality rates because different parishes have different age profiles and that could bias our estimated

# Project Workflow - Version Control

- Gentzkow and Shapiro (2014): Version Control
  - ▶ Store code and data under version control.
  - ▶ **Run the whole directory before checking it back in.**
    - Run your master file before each push.
- Application:
  - ▶ File repository (GitHub)

# Project Workflow - Directory Management

- Gentzkow and Shapiro (2014): Directories
  - ▶ Separate directories by function.
  - ▶ Separate files into inputs and outputs.
  - ▶ Make directories portable.
- Application:
  - ▶ Directory map in your README file

# Project Workflow - Directory Management

- Gentzkow and Shapiro directory map:

---C:/build---	---C:/analysis---
/input	/input
extract0B.xls	tvdata.dta (link to C:/build/output)
/code	/code
rundirectory.bat	rundirectory.bat
export_to_csv.stc	regressions.do
mergefiles.do	regressions_alt.do
/output	/output
tvdata.dta	fig1.eps
	fig2.eps
	tables.txt
/temp	/temp
chips.csv	regressions.log
tv.csv	regressions_alt.log

# Project Workflow - Directory Management

- Goodman-Bacon directory map:

## User-written packages (installed in the master file):

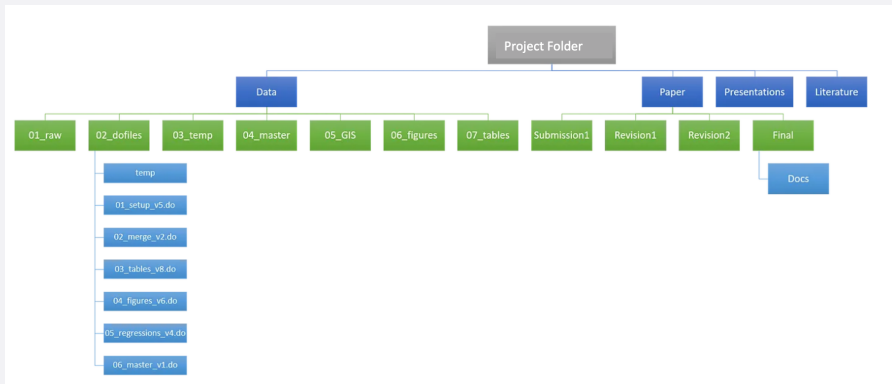
gtools, outreg2, grc1leg, cpigen, taxsim27, ranktest

## Folder Structure:

```
data/
  |--afdc/
  |--cen/
  |--datatemp/
  |--extra/
  |--vs/
  |--x/
dofile/
  |--cr/
    |--afdc/
    |--cen/
    |--pop/
    |--vs/
    |--x/
  |--an/
  |--extra/
```

# Project Workflow - Directory Management

- Example directory map:



# Project Workflow - Directory Management

- Navigating directories (Stata example):

```
clear
```

```
*** replace this with your main directory path
```

```
global projectdir "C:/Program files/Dropbox/myproject/"
```

```
global graphdir "$project/graphs"
```

```
global tabledir "$project/tables"
```



# Project Workflow - Directory Management

- Navigating directories (Stata example):

```
clear
```

```
*** replace this with your main directory path
```

```
global projectdir "C:/Program files/Dropbox/myproject/"
```

```
global graphdir "$project/graphs"
```

```
global tabledir "$project/tables"
```

- Note: be carefule using globals to define lists of control variables when using multiple .do files.

# Project Workflow - Project Management

- Gentzkow and Shapiro (2014): Management
  - ▶ Manage tasks with a task management system.
  - ▶ Email is not a task management system.
- Application:
  - ▶ GitHub Issues and Task Lists

# Project Workflow - Project Management

## Front-end work for Project Octocat #1

[Edit](#)[New issue](#)

octocat opened this issue 17 hours ago · 2 comments



octocat commented 17 hours ago · edited



### Description

We need to complete multiple tasks to make sure the front-end works smoothly and as designed.

### Tasks

- ☒ [Project Octocat] Design user interface
- ☐ [Project Octocat] Develop functionality for personalized user interactions and animations
- ☐ [Project Octocat] Optimize for mobile
- ☐ PLACEHOLDER - Integrate with back-end



saritai added the **front-end** label 17 hours ago



codercat commented 17 hours ago

Author

I just looked at the completed user interface design and it looks great! My status is that I am still working on our user interaction feature and will probably be focused on that for the next few days.



octocat commented 17 hours ago

Author

I have been breaking down all the steps it will take to integrate with the back-end and will create multiple issues for that work tomorrow!

### Assignees



octocat

### Labels



**front-end**

### Projects



None yet

### Milestone



No milestone

### Linked pull requests



Successfully merging a pull request may close this issue.

None yet

### Notifications

Customize

Unsubscribe

You're receiving notifications because you authored the thread.

2 participants



# Project Workflow - Tools

- Tools for an efficient project workflow:
  - ▶ Statistical software: R or Stata
  - ▶ Word processing software: LaTeX (Markdown/Quarto)
  - ▶ Integrated development environment (IDE): RStudio or Visual Studio Code (VS Code)
    - Stata isn't great for this
  - ▶ Version control: Git and GitHub