4. Version Tracking in GitHub

Data Wrangling and Visualization

Prof. Jack Reilly

F2025

In progress

Readings & Reference Material

- "Happy Git with R" https://happygitwithr.com/
- **The Plain Person's Guide to Plain Text Social Science**, ch 2 (https://plain-text.co/index.html)
- Bryan, Jennifer. "Excuse me, do you have a moment to talk about version control?"
- Wickham, Hadley. "Tidy Data".
- Kuriwaki, Shuri. "Git for Students in the Social Sciences: A Pitch

Computer Work

If you do not have one already, create a GitHub account and obtain a GitHub Personal Access Token.

1. Register a GitHub Account.
2. (Optional) Install Git. *Note: this step is only optional if you use the GitHub Desktop Client; if you use any other client, you must install Git at the command line.*
3. Install a Git Client. I strongly recommend you use the GitHub Desktop Client, which comes with git built-in, so that you do not have to install git separately (and can thus skip step 2).
4. Get a Personal Access Token and connect your GitHub Desktop Client to GitHub.
5. **(Optional)** Start an initial repo.[1] We will do this together in workshop this week, as well, but you can get a head start if you like.

---

[1]Jargon alert: repo is short for "repository".

Data Work

Setup

1. Create a GitHub repository for this assignment. (Don't forget a readme file and a .gitignore file!) It's OK if it is a private repository.
2. Generate a Quarto document (like last week). Title it `DWV Assignment 4.qmd`. Answer all questions for this assignment in this quarto document.
3. Practice good file management: keep all documents for this assignment in your assignment folder (github repo) dedicated to just this assignment. It is OK - it's advisable, actually - to have the datasets themselves in a separate folder.[2]

Part 1. Lines

1. Draw a y=4x+3 on a plot that ranges from -20 to +20 on both x and y.
2. Label the line appropriately next to itself.
3. Color the line in an interesting color! And make it dashed.
4. **Bonus** Use a combination of the above options (plus others, if you feel like looking them up) to create the wildest looking series of lines on a graph that you can. (Remember, you can overplot lines!)

Part 2. Protected Lands

Consider the `protected_lands.csv` dataset. This dataset contains information about a sample of countries sourced from the World Bank. `Protected Lands` represents the terrestrial protected land of a country as a percentage of total land area. GDP is represented in the dataset on a per capita (`gdp_percap`) as well as total basis (with total being measured in billions - `tot_gdp_billions`).

1. Find the mean, standard deviation, and range of each of the three variables: protected lands, GDP per capita, and total GDP.
2. Create a scatterplot where "protected lands" is on the Y axis and GDP per capita is on the X axis. Place a title on the graph and labels on the X and Y axis appropriately.
3. Create a scatterplot where "protected lands" is on the Y axis and total GDP is on the X axis. Place a title on the graph and labels on the X and Y axis appropriately.
4. Regress "protected lands" on GDP per capita.
5. where "protected lands" is on the Y axis and GDP per capita is on the X axis. Then, overlay the plot with a line representing the slope.

---

[2]I recommend that you have a folder on your computer where all of your assignments for the class are kept. Inside this folder, you should have a folder for each assignment; you can also have a folder that stores the data. The data folder should not be entered into a github repo; the individual assignments folders can be (or need to be, depending on the assignment.)

6. Regress "protected land" on total GDP.

Part 3. National Election Studies

1. Using `nes2000subset_age.csv`, regress partisanship on age.
2. Plot the resulting regression line, and properly label all axes. (Hint: don't forget to rescale the X axis!)
3. Color the line! Label the line next to itself with "OLS Line"

Submission

1. Render your quarto file to .html and .docx.
2. **Push** your completed assignment, including your .html and .docx files, to your GitHub repo.
3. Either:

   - Ensure the repo is public, or
   - Share the repo with me (I'm `@jacklreilly` on GitHub).

4. On Blackboard, under assginment 4, upload a link to the repo so that I can click straight through to it.