**Assignment 2 – Empirical workflow**

**Due date: Wednesday, February 12th, 2021 by 11:59pm**

**Directions:**  This assignment lays the foundation for many of the assignments that you will be required to complete over the course of the class.

**Gentzkow and Shapiro**

Read Gentzkow and Shapiro “Code and Data for the Social Sciences” in the “Helpful stuff” Github directory and answer the following

1. Summarize briefly the point of chapters 2- 8 in less than one page.

*Much of what Gentzkow and Shapiro were suggesting could be categorized into to two main ideas: organization and reproducibility. Chapters 2 through 5 were focused explicitly on data management and coding tips in order to keep one’s work organized and to expedite the process of reproduction, either by you the author or by an independent third party. Chapter 6 through 8 were more generally focused, touching on concepts that can be applied to coding, but are not explicitly focused on it.*

1. Why do Genztkow and Shapiro think these elements of modern empirical work are so important? What problems does each element solve?

*They seemed to be mostly focused on an academic research workflow. So, most of their tips and tricks are aimed at keeping your work organized so it becomes easier to publish once all the work is completed. On top of that, the more organized the underlying documentation is, the easier it will be to evaluate and replicate.*

1. Give an example of the sort of problem that could arise in the course of an empirical project if someone were to fail to adopt these principles.

*Well one thing that jumps to mind that they didn’t explicitly mention is this empirical workflow would help a team avoid replicating work, particularly the task management section.*

1. How do you plan to incorporate these solutions into your own work?

*I’m going to attempt to write my code in such a way that not only is obvious what’s happening but also so that anyone who wanted to click Run on my markdown file could do so start to finish and have it output all of my figures and tables*

**Git**

These next questions concern the software “git” and “github”.  As you saw in the previous section, Gentzkow and Shapiro believe version control to be one of the pillars of contemporary empirical research.  One of the most popular methods today of version control is Git.  But Git is a bit complicated the first time one learns about it.  I encourage you to read Gentzkow and Shapiro closely, as well as google and Youtube, to learn enough to answer the following questions.  One example is this deck of slides by Grant McDermott at the University of Oregon:

[https://raw.githack.com/uo-ec607/lectures/master/02-git/02-Git.html#1 (Links to an external site.)](https://raw.githack.com/uo-ec607/lectures/master/02-git/02-Git.html#1)

1. Create a new section in the document you used to answer questions 1-4. Briefly explain what git and Github are used for, how they are similar and how they are different.

*Git is an opensource underlying protocol that enables version control. GitHub is cloud based storage system/user interface that makes the use of Git much easier.*

1. Name a benefit of using git to organize your empirical research. What types of common problems can occur if you don’t use git?

*Well for one thing it will be traceable. Meaning that you won’t have to rely on naming protocols for version control. An issue you can avoid using is you can be confident that even if you make a mistake and realize it after you’ve already saved, you can revert back to an older version of the file.*

1. What about using git is challenging for you for right now? What steps can you take to minimize those challenges such that you can adopt git for this class?

*I’ve been thinking of git as if it were a cloud storage system. So, I was confused that it wasn’t automatically reflecting the changes I’ve been making as far as file organization vs actual changes to the files themselves. I’ve broken it once I think but I’m pretty sure I have it fixed. I’m used to working with iCloud which updates everything automatically once I’ve made a change. For github I need to remember to commit the changes I make.*

1. Name the four main Git operations. What does each operation do and how are is each operation different from one another?

Stage, commit, pull, push. They are effectively stop gaps in the process to ensure you want to make changes. Staging and committing are acknowledging that you’ve made changes in your files and are confident that you want them to be added to the version history. Pull and push are just way updating the repo and getting the most up to date version of the repo.

1. The first step in your new empirical workflow is the creation of a Github repository (“repo”). You can either do this independently or do this through R functionality.  You need to create a github account, then create your first repository called “Titanic”. Initialize with a Readme and create the separate folders that we discussed in class on Monday.
2. Post a link to your repository

<https://github.com/patrick-chase/Causal-Inference>

1. Please clone our course github repository on your desktop

 I made a clone of the mixtape\_learnr. I’m pretty sure that’s what this is referencing but if it isn’t please let me know and I’ll get the correct one.