**Assignment 2 – Empirical workflow**

**Due date: Wednesday, June 10th, 2020**

**Directions:** You have two days to complete the following written assignment. This assignment lays the foundation for many of the assignments that you will be required to complete over the course of the class.

**Slack**

Install slack on your phone and desktop. Create an account and log into our class slack channel. Note, this is for grade. You have two days to get enrolled, otherwise you won’t get credit for this part. It is worth **10% of your grade** and is due by end of day **June 10th**. Please do this immediately as it’s an easy 10%.

The name of the Slack channel is bogotcausalin-odk5667. The join link is here:

<https://join.slack.com/t/bogotcausalin-odk5667/shared_invite/zt-exkehag5-yTam5ufTmC3~7jd~Jr7KIA>

Change your profile with a new picture of your choosing, your name and a short description. Say hello in one of the main channels once you’re in! Be sure to tag me. This is where we will talk as a class regularly. We can also DM this way.

**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.
2. Why do Genztkow and Shapiro think these elements of modern empirical work are so important? What problems does each element solve?
3. 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.
4. How do you plan to incorporate these solutions into your own work?

**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>

I have also included a deck of slides by Frank Pinter in “Helpful stuff”. The completion of this section will satisfy the Github requirement of the course, not counting any additional assignments that use Github. You must have it done by Wednesday **June 10th** to receive the 10% credit.

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.
2. Name a benefit of using git to organize your empirical research. What types of common problems can occur if you don’t use git?
3. 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?
4. Name the four main Git operations. What does each operation do and how are is each operation different from one another?
5. 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.
6. Post a link to your repository
7. Please clone our course github repository on your desktop

<https://github.com/scunning1975/causal-inference-class>