

Juan Pablo Bautista Zambrano

Assignment 2

Gentzkow and Shapiro

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

The authors give us a manual with steps that we must consider to facilitate our work with data when doing empirical social research. They divide the manual into the following points:

1. Automation

The authors show us an orderly way of working with the data. The codes that we are using must be ordered from the beginning of the investigation, from the time we start to clean de database, until we reach the moment of running regressions, obtaining tables and graphs. We must have a directory that tells us which is the order in which the codes are run, so that if we want to replicate the results again, know the order of the steps is.

2. Version Control

The authors show us the correct way to make changes to our code files and have a control over each change. The wrong way to do it is to create new versions of the files naming them with dates or initials, this can confuse us. The correct way to do it is to create a repository that shows us the changes that have been made to a file, the date, and the author. They also tell us that we must run the entire directory to avoid possible errors that can be generated by the changes made.

3. Directories

In this chapter, they recommend us to have a general directory that contains different subdirectories by functions in case we want to modify some of the research such as analysis or a regression specification; for example, we only have to make changes in the part in which the regressions are run without having to touch the cleaning of the database. Also, the input data must be portable to avoid the files not working on another computer.

4. Keys

Here, they tell us that when working with the data it is important to know the structure of the data. Each table has a key that uniquely identifies the elements of a table and they do not take missing values; tables can have foreign keys that refer to other tables in the database. When making regressions, the different tables must be merged and for this the keys are necessary, since they will allow us to combine the tables.

5. Abstraction

In this part, the authors show us that it is better to avoid copying and pasting lines of code for a function that we want to execute several times, since we can drag errors that can affect our work. If we want to execute the same operation several, it is better to use the abstraction. Abstraction is turning specific instances of something into a general-purpose tool known as abstraction.

6. Documentation

In this chapter, the authors us when comments should be put to our scripts. The documentation serves to clarify that something is fine when at first it seems incorrect; they should also be placed to guide the reader through the operations performed. If the documentation causes confusion, comments should be kept updated, otherwise. It is better not to put them.

7. Management

In the final chapter, they recommend that when conducting research with more people, we should have a task manager so that each of those involved is clear about what tasks are assigned, the description and their status. Tools like email should not be used to assing activities because they can become unclear and ambiguous.

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

They think that these elements are important because if they are implemented in empirical research, more efficient use of the time can be made, problems that can be costly are avoided, and order is also maintained when conducting research.

Problems that each element solve:

- **Automation:** helps replication and order from the beginning of the investigation.
- **Version Control:** we have control of the changes made to the files.
- **Directories:** It helps to maintain an order, without the need to modify all the files in case of making a change in one of them
- **Keys:** helps to know the structure of the data and what are the key variables when merging tables.
- **Abstraction:** Helps to execute the same procedure several times, avoiding copying and pasting errors.
- **Documentation:** gives information to the script reader and makes clarifications.
- **Management:** When working with more people, there is an order of the responsibilities and tasks that each one has.

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.

One kind of problem that can arise if these principles are not taken into account is that there is no good communication between colleagues; If each one is not clear about what to do, then efficient work will not be done. In addition, in case of not having organized the documents or not having a control of the changes made, confusion can be generated among the researchers, and this can generate errors in the manipulation of the data, the regressions, the results and so on.

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

I plan to incorporate these solutions mainly by starting to use Git, this way I will have control of the changes I make to my tasks and documents. Also, the use of repositories and directories will allow me to have a correct order of the files. This repository method will be of great help in conducting efficient research for my master's thesis

Git

- 5. 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: It is a version control tool that allows us to manage the various changes that are made to a file. It is mainly used to make changes to files that contain code. It is useful when working with more people since everyone involved can see the changes that have been made, who did it and when.

GitHub: It is a platform that allows us to create collaborative code repositories. It is useful for working with multiple people as everyone can access the repositories and make changes.

Git and GitHub are similar in that we can control the changes we make to our files. They are different in that GitHub is an online platform that allows us to use the Git System, but we don't need GitHub to use Git.

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

One benefit of using Git is that every time I make a change to my code files, I will know at what point I made it, to what file and what change was made, avoiding errors or confusion.

The common problems that can occur if Git is not used is that we do not have adequate control over the changes made to the files. This can be very confusing when working with research colleagues, creating confusion, and perhaps generating errors for example, when organizing a database or running regressions.

- 7. 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 think that what can be a challenge at the moment is to get to know and use a new program. Whenever we start using a new tool, it takes time to become familiar with it.

To minimize these challenges, I think it would be helpful to watch YouTube videos and search for online manuals that guide me on how I should use the program, how to create a repository, directories, subdirectories and so on.

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

1. Stage: This operation is used to tell Git that we want to add changes to the history of the repository.

2. Commit: This operation tells Git that we are sure that changes made in "Stage" should be part of the repo history.

3. Pull: This operation allows us to obtain the new changes made to the repository by our research colleagues or by me on another computer

4. Push: This operation allows us to send any changes made locally to the repository. For example, if I make a change to the repository on my personal computer, the Push operation allows me to send the change to the collaborative repository on GitHub.

9. – 10. Repository

<https://github.com/JPBautista15/Titanic>

11. Course GitHub repository in my desktop

