

Open Science and Reproducibility Principles

Zehao Zhao

There are three steps that I can do to make my project reproducible and follow the open science principles. First, I have created a GitHub repository for this project. Whenever I make a modification to my code or adding a file or documents, I need to `git add` and `git commit -m "messages"`. Therefore, each modification is recorded as the project goes, which follows the open science principles. Second, I will make citations whenever I have used a dataset from the library or internet, this will allow readers to know where the dataset comes from, and thus using their datasets to reproduce the result I have made. Third, as the project goes I will make my method and code open sources, so everyone will the dataset and my code can run it on any platform. Furthermore, I code write code on Jupiter notebook, which controls the variable that each person's computer and internet access are different.

In addition to the three fundamental steps of open science and reproducibilities mentioned above. I will try my best to use the open-source code, packages, or dataset. Whenever I wrote code, I will add short comments in the cell or the function I have implemented, or use good naming for variable names to make sure, people can reproduce intermediate results without confusion. Finally, before writing the final thesis or make my results and conclusion public. I will reproduce my code again to make sure I, myself is able to reproduce the code. My GitHub repository URL: <https://github.com/zehao-zhao/datah195a>