Assignment 1: Reproducibility, Workflow, Version Control

Qianyi Xia

OVERVIEW

This exercise accompanies the lessons in Environmental Data Analytics (ENV872L) on reproducibility, workflow, and version control.

Directions

1. Change "Student Name" on line 3 (above) with your name.
2. Use the lesson as a guide. It contains code that can be modified to complete the assignment.
3. Work through the steps, **creating code and output** that fulfill each instruction.
4. Be sure to **answer the questions** in this assignment document. Space for your answers is provided in this document and is indicated by the ">" character. If you need a second paragraph be sure to start the first line with ">". You should notice that the answer is highlighted in green by RStudio.
5. When you have completed the assignment, **Knit** the text and code into a single PDF file. You will need to have the correct software installed to do this (see Software Installation Guide) Press the `Knit` button in the RStudio scripting panel. This will save the PDF output in your Assignments folder.
6. After Knitting, please submit the completed exercise (PDF file) to the dropbox in Sakai. Please add your last name into the file name (e.g., "Salk_A01_Reproducibility.pdf") prior to submission.

The completed exercise is due on Thursday, 17 January, 2018 before class begins.

1) Discussion Questions

Question

Why are reproducible practices becoming the norm in data analytics?

> Answer: Reproducible practice can recreate the final reported results accuratey and can communicate the work you have done to someone else (The Practice of Reproducible Research, 2018). Today, under a rapid expanding world of scientific data analysis, and computing, these reproducible practices become the norm because: 1)it can help to record all steps, linking the final result to the inital data, and make then all availabe to collegue or other scientists; 2)reproducibility allow the research focus to move up a level and spending more time focus on the entire scientific workflow, and later other scientists are able to use the previous practice sustainably; 3)it can be apply to almost every types of research or analysis.

Question

What are your previous experiences with data analytics, R, and Git? Include both formal and informal training.

> Answer: I have audit a course learning Java and used Github but created or saved nothing on it. Then at Duke, I took ENV 710 to learn R skills. I had good experience in studying data analysis.

Question

Are there any components of the course about which you feel confident?

> Answer: I enjoyed data analysis learning last semester, because I did bad in midterm in writing part, but I did really good in the R analysis the final exam, that makes me feel confidence in using tools for real case study. And I actually spend lot of time learning from online sources when I failed to follow everything in class.

Question

Are there any components of the course about which you feel apprehensive?

> Answer: This happened last semester, when I got stuck in one step, I kept trying to solve it for more than 5-6 hours and still can't figure out. I kind of afraid of meeting those situations.

2) GitHub

Your Repository

Provide a link below to your course repository in GitHub. Make sure you have pulled all recent changes from the course repository (https://github.com/KateriSalk/Environmental_Data_Analytics) and that you have updated your course README file.

> Answer: https://github.com/xqy1012/Environmental_Data_Analytics