Assignment 1: Reproducibility, Workflow, Version Control

Jiaqi Li

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: Reproducibility means that an original dataset can be analyzed by others using the same or different codes as the original approach to get the same results as the original data analytics. Reproducible practices are essential and become the norm in data analytics because reproducibility ensures the accuracy of the data analytics. It makes sure that other researchers can reproduce the same results from the original data and may continue to work on the data from the existing results.

Question

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

> Answer: I learned R from ENV 710 Applied Statistics for Environmental Science course. I have never used Git before.

Question

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

Answer: Only the GLMs section.

Question

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

Answer: No.

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/Jiaqi-Li-Duke/Environmental_Data_Analytics