Assignment 1: Reproducibility, Workflow, Version Control

Student Name

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: Reproducale practices are becoming the norm in data analytics because as the field of data management and analysis grows, and pools of data continue to grow as well, it is important to be able to see the process that produced a product, particularly if analyzed data is being shared. Moreover, it is important to be able to easily retrace steps taken in an analysis, especially when the analysis needs to be modified.

Question

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

> Answer: I have no prior experience with data analytics or Git, but I have some familiarity with R. I have used R during my assistantship, but the code was written for me, and I learned very specific code for the analysis I was doing.

Question

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

> Answer: Despite my minimal experience with R, I feel confident to begin the semester because I know both of the professors will teach this course well. Both Kateri and John communicate well and are genuinely interested in our progress, which are essential to my understanding of the material.

Question

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

> Answer: I don't feel apprehensive about any part of the course, but I understand that it will be frustrating at points, particularly when it comes to trouble shooting.

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/jakegreif/Environmental_Data_Analytics