Assignment 1: Reproducibility, Workflow, Version Control

Walker Grimshaw

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: Data analysis and the results of such analysis need to be reproducible so their conclusions are verifiable by other people or parties. As more data is made available across disciplines and more groups analyze this data, it will be important to scrutinize methods and results.

Question

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

> Answer: The majority of my experience with R is from John Poulsen's statistics course last fall. I had some experience before that from a one-credit course at UNC that introduced us to some of the abilities of R to analyze and visualize data. I had previously created a GitHub account when a colleague was developing a smarthphone app for logging water quality data at drinking water treatment plants, but I used it very little and did none of the software development myself.

Question

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

> Answer: I feel confident about most of the course and am excited to use R more and get an introduction to Python.

Question

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

> Answer: As we saw today, I think lectures will be difficult with so many people using their personal laptops, rather than the computers in the lab that may be more similar to one-another. I am also apprehensive about using Git. I think it will be extremely useful to learn how to effectively use, but I don't have much confidence in my current computer-server abilities.

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/wgrimshaw/Environmental_Data_Analytics