Assignment 1: Reproducibility, Workflow, Version Control

Kat Horan

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 practices in data analytics allow for repeated testing to confirm results, recreating a method of analysis for different data sets, data sharing with collegues, and improves transparency and integrity.

Question

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

> Answer: I have no previous experience with Git and my only prior experience with data analytics and R comes from John Poulsen's statistics class last semester.

Question

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

Answer: I feel decently capable in R now, however, the learning method in Poulsen's course felt a lot like trial by fire at the beginning. I think he expected that many of us had at least a beginners experience in R so we aside from a very brief introduction, we jumped right into the labs without much hand holding. A lot of what I learned about R's capabilities came mostly from hounding Taylor at his office hours and scouring the internet. As such, I'm not sure I have the best/most efficient coding practices and am hoping this class will help me become more proficient and provide some more structured guidance along the way.

Question

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

Answer: Scraping data, cleaning data (I've done some, but I still feel like I'm at an amateur level), and navigating Github all seem a little intimidating.

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/kdh48/Environmental_Data_Analytics/blob/master/README.md