Assignment 1: Reproducibility, Workflow, Version Control

Laurie F Muzzy

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 supports transparency in data analysis, allowing people to see the process and coding behind it, rather than only the inferences and conclusions. It also encourages collaboration. Both of these things lead to reduction in errors (hopefully!), producing better science.

Question

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

> Answer: I've used Excel and ArcMap for data analytics in undergrad and here at Duke, and independently took a class through Coursera on Probability and Data Using R (but it was pretty limited on what we learned about coding). I've had no training with Git before.

Question

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

Answer: I'm glad that I get to use my own Mac rather than the school PCs, so that lends me a tiny bit of confidence. And I am confident our well-trained, receptive professors and TAs will help us students learn a lot. I'm confident that collaboration will make us all better scientists ;)

Question

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

Answer: I don't have the best memory when it comes to using new computer programs, so I'm apprehensive that I will 'learn' something but then forget it in a crucial moment, or fall behind and have a hard time catching up (thank you for being super-aware of people getting stuck!)

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/lfm21/Environmental_Data_Analytics