Assignment 1: Reproducibility, Workflow, Version Control

Caroline Watson

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 are becoming the norm in data analytics because it is important to be able refer back to previous code, understand the comments, and be able to reproduce the steps, if necessary. Reproducible practices are also importnat so that others can understand your work and be able to do the same thing.

Question

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

> Answer: My previous experiences with data analytics and R comes from the Applied Statistics for the Environmental Sciences course I took last semester. This was a statistics course, so my experience in R was with the weekly lab assignments and learning how to code or use specific statistical functions during lecture. I don't have any experience with Git, but am looking forward to learning more about it!

Question

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

> Answer: I feel confident about using R because I used it last semester and feel familiar enough as to how it works.

Question

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

> Answer: I feel a bit apprehensive about the problems that I/we could run into while using R since this is different from the statistics class I took last semester. I am also not sure I fully understand the connection between R, GitHub, and files on my computer and how to update each respectively, but I'm hoping that will come with time/practice.

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/cwatson1013/Environmental_Data_Analytics