Assignment 1: Reproducibility, Workflow, Version Control

Caroline Reents

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 more and more people are working in teams on projects across different machines. By creating reproducible data, another person on another machine can understand the logic you followed in order to come to a conclusion. By being able to follw that logic, you can come to a consensus on whether or not that is a valid conclusion.

Question

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

> Answer: I took the R course my first year and used R for thaat class project as well as on my class project for social science surveys. Other than that I do not have much more experience with R. I have never used git before.

Question

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

> Answer: I feel pretty confident about the interface of R/R studio and running basic statistical tests.

Question

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

> Answer: I am apprehensive about the connection between git and R, it is a little confusing to me, but I am getting the hang of it.

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/c-reents/Environmental_Data_Analytics