Assignment 1: Reproducibility, Workflow, Version Control

Njeri Kara

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 data analysis ensures that data analysis can be repreated or retraced by the data scientist or an independent person from the same raw data set at a later time. This practice promotes accountability because the data analysis process is transparent, giving it credibility. It also enables the data scientist or an independent person can catch errors or mistakes that could have been made in the data analysis since the process is clear, improving the quality of data analysis and results.

Question

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

> Answer: My experience with R is in the Applied Statitics course (ENV 710) I took in fall 2018. I have no prior expereince with Git.

Question

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

> Answer: I feel confident that I will be able to pick up the main concepts and R functions to carry out the various data science pipeline components.

Question

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

> Answer: I am apprehensive about dealing with the git repository because I have no prior experience with git hub. I am also apprehensive about the collaborative nature of the course especially in dealing with bugs and trouble shooting. I have a preference for independent study, therefore I need to push myself to seek out help and work with others as part of a team.

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/Njeri-Kara/Environmental_Data_Analytics