

Assignment 1: Reproducibility, Workflow, Version Control

Felipe Raby Amadori

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: I think that reproducible practices are becoming the norm in data analytics because it allows anyone to reproduce exactly the results of a data analytics study, which is extremely important for the validation and review of these studies; moreover, it allows different groups of people to collaborate in a efficient, clear, and open way, and where you can always go back and try to fix things that went wrong along the way.

Question

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

Answer: I studied hydrology in undergrad and later in my career as a hydraulic engineer I worked using hydrologic data like precipitation, river flow, evaporation, etc., to determine return periods, frequencies, probability distributions, an use that data for designing infrastructure and/or making management decisions. I have also managed large data sets of geographic information like deep wells locations and characteristic using GIS software. My only experience using R was doing the course ENV 710 Applied Data Analysis for Environmental Sciences last semester. I have no experience with Git.

Question

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

Answer: I feel confident learning R and others programming languages. Logical systems are my comfort zone. Also, having experience managing some environmental data sets in the past (not extremely large) makes that part of the course familiar.

Question

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

Answer: Nothing makes me feel super apprehensive, but I am always a little scared of exploring new areas like new software, new ways of working, or statistical methods and approaches. It is a good scare though, mixed with excitement of learning new things. I hope the assignments will not be super time demanding.

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/fr55/Environmental_Data_Analytics.git