

Assignment 1: Reproducibility, Workflow, Version Control

Sebastian Bognar

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 they more efficient and can greatly increase productivity. You can create a tool or code for a certain type of analysis that can be applied to multiple datasets increasing your efficiency and production. It can also reduce errors in analysis and automate the process to a degree. Reproducible practices also allow another user to obtain the same results providing uniformity to an analysis that takes place in a team setting.

Question

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

Answer: I learned how to use R in R Statistics course at the Nicholas school and have used it for several projects that have required data analysis. I have mainly used R for water related datasets and have used it to perform some regression analysis. I have also learned how to use Matlab in Watershed Hydrology course, but have a better understanding of R than Matlab. In the Fall, I learned how to use Python in advanced GIS where we also learned how to use Git. However, I have not used Git for outside projects. I think that Git will be really helpful in the data analysis

portion of our MP Project that deals with quantifying restoration success of wetlands in North Carolina.

Question

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

Answer: I am pretty confident in using the basic functions of R. If I have any problems, I know how to troubleshoot and find the answer that I am looking for. I am pretty familiar with Git from Advanced GIS so I am pretty confident in using it as well.

Question

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

Answer: I feel that everything is pretty straight forward and currently I am not apprehensive about any components of the course.

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/seabass72/Environmental_Data_Analytics