

Assignment 1: Reproducibility, Workflow, Version Control

Xin Wang

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: Someone else or future self may view the data work some time. Reproducible work can help to recall and make clear the context, data information, analysis processing and results. If they want, data analytics work can be done again at any time. This is efficient and effectively avoids confusion.

Question

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

Answer: During undergrad I took a course called “Quantitative Methods in Ecology”, where I was trained the basic operations in R. I also had some experience in R, using certain packages for ecological research purpose (e.g. community indices calculation). That’s all.

Question

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

Answer: I feel confident about nothing but Kateri. I believe her course must be of high quality.

Question

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

Answer: I would say no, or all of them. They are the same.

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/nataliadxx/Environmental_Data_Analytics