

Assignment 1: Reproducibility, Workflow, Version Control

Rachel Bash

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: Reproducibility allows one to save multiple versions of their work, so they can go back to previous versions if necessary. Also, it allows teams more flexibility and ability to cross-check work in case mistakes were made or in case different analysis on the same dataset is required. Lastly, reproducibility allows the users to focus on the content on the data analysis by showing the workflow through all of the steps, not just the final product.

Question

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

Answer: I took a statistics class in undergrad and took John Poulsen’s Applied Statistics class. Both involved the use of R. I really liked R and it felt natural to me in John’s course. I have worked with datasets in the past, but I have never gone too in depth with analyzing them. I have never used Git.

Question

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

Answer: I feel confident in the basics of R and some exploratory data analysis. I am excited about learning data visualizations and R Shiny.

Question

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

Answer: Github was a little confusing yesterday, and I am nervous about making sure my local and my cloud are regularly up-to-date with one another, because I am nervous about merging issues.

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/rachelbash/Environmental_Data_Analytics