LAS 6292: QAQC 1

updated: 2021-03-18

Pre-Class Preparation (Instructor):

Send in an email to students:

- content of any pre-class emails.

Bring to Class:

- Snacks
- Flip charts and markers
- Dry write markers
- Tent cards for student names
- website R resources
- RStudio shortcuts PC or mac
- Cheat Sheets Rstudio base r,tidyverse, etc.

Objectives and Competencies:

After completing this lesson, the participant will be able to: Define data quality control and data quality assurance Perform quality control and assurance on their data at all stages of the research cycle

QA/QC: finding and fixing mistakes A. Plotting different ways B. Summary calculations C. Reading back D. Randomly sampling from digital file to verify E. Open Refine. F. MOST IMPORTANT: DO NOT FIX RAW DATA FILE. Txt file with changes that need to be made, then make them with a scripting language

2. Double Data Entry – gold standard for accurrace
3. Single Data Entry: how can you verify? Speech Tools: hear your data read back to you after entering it MAC: System Preferences->Accessibility PC: https://excelofftheGrid.com/excels-speak-cells-feature/

Pre-class Preparation (Students):

Readings

1. Laskowski, 2020. What to do when you don't trust your data anymore. [read online] [download pdf]

2. Pennisi, E. 2020. Spider biologist denies suspicions of widespread data fraud in his animal personality research. Science.
[read online] [download pdf]
3. Alston, J. M., and Rick, J. A.. 2020. A Beginner's Guide to Conducting Reproducible Research. Bull Ecol Soc Am 00(00):e01801. [read online] [download pdf]
4. Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK (2017) Good enough practices in scientific computing. PLoS Comput Biol 13(6): e1005510.
[read online] [download pdf]

Online Lectures:

These introductions to R and R Studio were made by Professor Ethan White (UF-WEC). They are a good overview of some R basics.

- Intro to R and RStudio.
- Navigate R and RStudio web page
- Intro to R Packages
- Expressions and Variables in R

Reminder: Don't forget to install the `Tidyverse` library; you can read how on this page. And if you want to get ahead to what we will be doing in class, you can read about how to Set up a Project in R

Computer programming is challenging to learn and teach, especially remotely. The next few sessions we will be using modified version of a style known as:

```
(1) I do it
(2) We do it
(3) You do it
```

This means you first observe me doing something in Rstudio by watching a video (or in this case you'll watch a video of Prof. Ethan White). The in class we will do some tasks together at the same time (aka "live-coding"). This is useful, because you can ask questions as you go. Then, you will work on an assignment based on the tasks we did together, which you will submit as your assignment.

Class Outline

Topic 1: Intro to Reproducibility: (15 min)

Breakout: (10 min)

Returning results: (20 min)

Break (10 min)

Coding Corrections: Intro (15 min)

Coding Corrections: Implement the Changes(60 min)

- **Key Points to teach emphasize:**

KEY MESSAGES: (1) Keep raw data raw (2) annotate. lots.

Coding Corrections: Now do it on your own (50 min)

Tools & Resources

For Advanced Users: The R package pointblank has some very advanced data validation tools for tidy datasets.

Sources

DataONE Education Module: Data Quality Control and Assurance. DataONE. Retrieved Nov12, 2012. From http://www.dataone.org/sites/all/documents/L05_DataQualityControlAssurance.pptx