## NRES 898 Special Topics: Sampling, data management and visualization 1 credit SPRING 2016

*Instructor*: Dr. Drew Tyre, atyre2@unl.edu, 472-4054 HH 416, East Campus

*Time: asynchronous except for office hours Location: online (see below for details)*

*Office hours:* M 11-12pm, Tuesday 2-5pm or *by appointment*

### Description/Rationale

Wilson et al (2014) said “…recent studies have found that scientists typically spend 30% or more of their time developing software. However, 90% or more of them are primarily self-taught, and therefore lack exposure to basic software development practices …” and then documented a number of high profile retractions of scientific papers due to software errors. Natural resource scientists in the 21st century need to have a basic grasp of tools for developing sampling designs, entering data with Quality Assurance practices, validating data with Quality Control practices, documenting data with metadata tools, and manipulating data for visualization and analysis. An increasing number of journals require submission of both data and code to back up research results. Funding agencies are beginning to require formal data management plans as part of grant proposals, and those plans are expected to provide for sharing of data and code. Ensuring that these artifacts actually correspond to the results in a manuscript is not a trivial task.

This course will empower students with a range of basic tools to implement best practices for scientific computing. Students will practice with the scientific workflow that follows from the design of the sampling scheme, through generation of the data in the field or lab, up to the point of analysis.

This is a test of what happens when you edit a docx file that is stored in a git repostority.

Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, et al. (2014) Best Practices for Scientific Computing. PLoS Biol 12(1): e1001745. doi:10.1371/journal.pbio.1001745

### Prerequisites

Graduate standing or permission of instructor

### Learning Objectives

*By the end of the course the students will be able to:*

1. Set up a project directory using principles of reproducible research
2. Draw a stratified random sample from a population of observational units
3. Develop data entry techniques that use QA/QC procedures
4. Document data with metadata tools
5. Fuse data from different datasets together for a single analysis
6. Reshape data to produce informative visualizations
7. Automate data workflows with simple programming constructs using R
8. Produce submission ready documents exemplifying the principles of reproducible research

*Students will have the opportunity to*

1. Critically discuss modern literature on data manipulation, analysis, and reproducible research.
2. Use a range of software tools to do sampling design, data entry, metadata entry, QA/QC, data manipulation and visualization.

### Instructor’s role in course

This course is offered fully online, in an asynchronous fashion. I will have online “office hours” using Adobe Connect on Monday between 11 and 12, and Tuesday between 2 and 5 pm. At other times I will read items posted in the discussion groups within 24 hours of posting during the week, and within 48 hours over the weekend. I will respond if needed to clarify a question or comment, answer a question, or clarify a response by another course participant. If you send in a question via email, I will generally repost it in a discussion thread and answer it there. You will get your answer much more quickly by posting in the discussion thread directly.

I am also available by telephone at the number above during the week, generally between 8:30 am and 4:30 pm. If I am not at my desk, I typically respond to voice mail messages within 2 hours.

## Textbooks

We will use the lessons available at software-carpentry.org/lessons.html for reading material.

## Hardware and Software requirements

Any modern pc or mac should be able to run all of the software we will use in this course. An internet connection is needed to access course materials on Blackboard, and submit assignments.

Participating in the online help sessions requires an internet connection, a microphone, and speakers. The integrated microphone and speakers on a laptop are typically *not* sufficient, but a cheap pair of earbuds will reduce or eliminate issues with feedback. A USB headset that combines headphones and a microphone is optimal. A webcam, either integrated or standalone, also helps interaction online because I can see when you’re not getting it even if you’re not telling me in words!

We will use the Bash shell, R/RStudio, Git and SQLite. All of the software we will use is available as open source. Links to software are on BlackBoard in the Getting Ready module. Video tutorials on downloading and installing the software are available in the Getting Ready module.

## Login instructions

http://my.unl.edu

Log in using your Bb username and password. If you don’t know what this is, please contact the UNL Help Desk (see below).

## Technical or administrative help

UNL HELP DESK

Phone: 472-3970 or toll-free (866) 472-3970

E-mail: helpdesk@unl.edu

Walk-in: room 105 of the 501 Building

(501 Stadium Drive) on the west edge of City Campus, 7:30 a.m. – 11:30 p.m.

For questions about the course content or procedures, please contact your instructor.

## Assessment

### **Lab assignments 40%**

You will complete a lab assignment per week (see the Course Schedule below) that will walk you through various steps in using the software These formative assignments contribute to your grade, but generally you will earn full points simply for completing all the steps.

### **Homework assignments 40%**

There are 4 homework problems (1 per module; see the Course Schedule below); generally these will require you to do some data related tasks and summarize the results in written form. The final homework assignment is a project in which you will prepare a sampling plan, data entry and QA/QC procedures, metadata and documentation for a problem of your choice.

### **Paper discussions 20%**

**In each module there will be one or more opportunities to discuss papers either about reproducible research, sampling or visualization. These will occur online, as required by the course schedule. They will be graded on the basis of participation, with each student required to post a quote or quotes, comment on the quotes of others, and respond to any comments received on their quotations.**

### **Grade Scaling**

The total points assessed will be translated into letter grades as follows: <60 – F, 60-69 – D, 70-79 –C, 80-89 – B, 90-100 – A. I do not use minus grades, and give plus grades to scores in the top 20% of each band.

## Course Schedule

I reserve the right to deviate from this schedule if required. However, I will not make any reading or assessment due earlier; I will only postpone due dates or cancel assessments altogether.

|  |  |  |  |
| --- | --- | --- | --- |
| **Module** | **Topic** | **Week** | **Homework Due** |
| 1 | Laying the foundation: filenames and directory structures for reproducible research | 1 |  |
|  | Unleashing the beast within: using the command line and shell programming for automation and version control | 2 |  |
|  | Programming with R | 3 |  |
|  | Tying data and documentation together: knitting with code | 4 | Homework #1: principles of reproducible research and simple programming |
| 2 | Simple random samples in space and time | 5 |  |
|  | Stratified random samples in space and time | 6 |  |
|  | Entering data: Quality Assurance in spreadsheets | 7 |  |
|  | Entering data: Quality Assurance in databases | 8 |  |
|  | What is metadata and how do I enter it? | 9 | Homework #2: sampling design and entering data |
| 3 | Manipulating data with SQL for Quality Control | 10 |  |
|  | Data fusion with SQL | 11 |  |
|  | Data manipulation and fusion with other tools | 12 | Homework #3: Quality Control and data fusion |
| 4 | Principles of data visualization: cognitive limitations | 13 |  |
|  | Fusing and manipulating data to produce interesting visualizations | 14 |  |
|  | Final project | 15 |  |
|  | Final project | 16 | Homework #4: wrapping it all up with your own project |

## Course Policies

### Late Assignments

Homework assignments, proposals and the final paper may be turned in late and receive full credit as long as I have not completed grading the work from other students in the class. As soon as I have finished grading all of the submitted work, any work submitted later than that will receive a grade of zero. Depending on my workload, you might have as long as a week. However, I can finish grading quite quickly, so do not presume that you have more than one day.

All “paper summary” assignments *must* be turned in on time; no late summaries will be accepted (grade of 0).

### Student Code of Conduct

Students are expected to adhere to guidelines concerning academic dishonesty outlined in Section 4.2 of University's Student Code of Conduct ([*http://stuafs.unl.edu/dos/code*](http://stuafs.unl.edu/dos/code)). **(Optional) Instructor - Outline penalties possible for instances of dishonesty such as cheating and plagiarism here.** Students are encouraged to contact the instructor for clarification of these guidelines if they have questions or concerns. The SNR policy on Academic Dishonesty is available at [*http://snr.unl.edu/employeeinfo/information/employeehandbook-single.asp?infocode=S162*](http://snr.unl.edu/employeeinfo/information/employeehandbook-single.asp?infocode=S162).

### Netiquette

Core rules of Netiquette: <http://www.albion.com/netiquette/corerules.html>

### Students with disabilities

Students with disabilities are encouraged to contact the instructor for a confidential discussion of their individual needs for academic accommodation. It is the policy of the University of Nebraska-Lincoln to provide flexible and individualized accommodation to students with documented disabilities that may affect their ability to fully participate in course activities or to meet course requirements. To receive accommodation services, students must be registered with the Services for Students with Disabilities (SSD) office, 132 Canfield Administration, 472-3787 voice or TTY.