**Introduction to Data Science**
**200 - level**
**Autumn**
**(A pre/co req for 290)**
(3 Credits)

**Professor**
Dr. Steven Roberts
Office: FTR 232
Email: sr320@uw.edu
Phone: 206-866-5141

**Times & Location**
Tuesdays: 1:00-3:00; FSH 213
Thursdays: 9:30-11:00; FSH 213
On a per week basis, 3.5 hours class time, 2 hours reading and preparation, 1.5 hours Summary Sheets, 2 hours Weekly Assignments

**Website**
https://sr320.github.io/course-fish497-2018.

**Course Objectives**
In this course you will learn fundamental aspects of data acquisition, data wrangling, and visualization that are necessary for conducting biological and ecological research. By the end of the course you will be able to import data into R, perform analysis on that data, and export the results to graphs, and presentations. Upon learning basic skills, students will identify a research question and corresponding data set to analyze. By learning how to get the computer to do your work for you, you will be able to more effectively analyze complex data sets and communicate the findings of your analyses.

**Learning Goals**
Students successfully completing this course will be able to:

1. Understand how to navigate a computer file system
2. Create spreadsheets using proper formatting for downstream analysis
3. Synthesize complex data sets to address a pre-defined question.
4. Evaluate the efficiency of R code to perform basic tasks.
5. Effectively visually communicate data analysis products.

6. Develop a reproducible data analysis workflow

**Reading Materials**
All required reading material can found on the course website in the schedule. Most readings will be in R for Data Science by Garrett Grolemund and Hadley Wickham. A version of the textbook is available free online at http://r4ds.had.co.nz/. Additional readings will include material developed as part of the Data Carpentry: Ecology Curriculum and R vignettes. Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research. The mission of Data Carpentry is to provide researchers high-quality, domain-specific training covering the full lifecycle of data-driven research.

**Technology**
Students are required to provide their own laptops and to install free and open source software on those laptops (see Setup for installation instructions). Support will be provided by the instructor in the installation of required software. The Student Technology Loan Program (https://stlp.uw.edu/) also allows students to borrow laptops on a first come, first serve basis.

**Course Description**
Data management skills are needed for entering data without errors, storing it in a usable way, and extracting key aspects of the data for analysis. This course will provide an introduction to data management, manipulation, and analysis, with an emphasis on biological problems. Class will typically consist of short introductions or question & answer sessions, followed by hands on computing exercises. The course will be taught using bash, R, and Jupyter Notebooks, but the concepts learned will easily apply to all other computational work. There will be an emphasis on reproducible research which includes the use of R markdown and version control. Students will need to come to class each week having completed the assigned readings.
The course will culminate in students identifying a data set, identifying a specific research question that can be answered with the data set, and communicate the findings in an effective visual manner. Work on the these student projects will begin during week 6 of the course and students will present their work to the class during week 10.

**Prerequisite Knowledge and Skills**
Knowledge of Introductory biology. No background in programming is required.

**Assignments and Grading**
There will be 10 equally weighted assignments due on Canvas at 11:59 pm on Friday. Weekly summary sheets will be due in class on Thursday. Assignments can be found on the course

website under schedule. Weekly assignments will be based primarily on the readings and class activities during weeks 1-5. During weeks 6-10 the weekly assignments will consist of a) readings and class activities and b) questions specific to student projects. Weekly assignments will assess the students progress on Learning Goals 1,2, 3, 4 and 6. The Weekly summary sheets align with Learning Goals 1, 2, 4, 5, and 6. The Final Presentation will assess the students' competency in Learning Goals 3, 4, 5, and 6.

Weekly Assignments - 70%
Weekly Summary Sheets - 10%
Final Presentation - 20%

**Disability Accommodations**
It is crucial that all students in this class have access to the full range of learning experiences. At the University of Washington, it is the policy and practice to create inclusive and accessible learning environments consistent with federal and state law. Full participation in this course requires the following types of engagement: Working with computers in the classroom and outside of the classroom and working in small groups. If you anticipate or experience barriers to your learning or full participation in this course based on a physical, learning, or mental health disability, please immediately contact the instructor to discuss possible accommodations. A more complete description of the disability policy of the College of the Environment can be found at https://environment.uw.edu/intranet/academics/teaching/disability-accommodation/. If you have, or think you have, a temporary or permanent disability that impacts your participation in any course, please also contact Disability Resources for Students (DRS) at: 206–543–8924 V / 206–543–8925 TDD / uwdss@uw.edu / http://www.uw.edu/students/drs.

**Academic Integrity**
Passing anyone else's scholarly work (which can include written material, exam answers, graphics or other images, and even ideas) as your own, without proper attribution, is considered academic misconduct and a violation of the University of Washington Student Conduct Code (WAC 478-120). Please review the College of the Environment website on academic integrity so that you are clear on what constitutes academic misconduct. I expect that you will know and follow all university policies on cheating and plagiarism. For more information, see the College of the Environment Academic Misconduct Policy and the University of Washington Community Standards and Student Conduct website. Anyone engaging in academic misconduct will not receive credit for the course, and will be further handled according to university regulations.

**Netiquette and Communication Courtesy**
All members of the class are expected to follow rules of common courtesy in all email messages, threaded discussions and chats.

**Getting Help**

**Most importantly, if you are struggling for any reason please come talk to me and I will do my best to help.**

**Course Schedule**

Links to readings, lectures and assignments can be found on the course website at https://sr320.github.io/course-fish497-2018/schedule

| Week | Language | Lesson | Readings |
|---|---|---|---|
| Oct 2 | Excel | Data Organization in Spreadsheets (LG 2) | *Data Carpentry: Ecology Curriculum*<br>❏ Introduction to data organization<br>❏ Formatting data tables in spreadsheets<br>❏ Formatting problems<br>❏ Dates as data<br>❏ Exporting data |
| Oct 9 | Bash (define) | Navigating the Command line (LG 1) | *Data Carpentry: Ecology Curriculum*<br>❏ Introducing the Shell<br>❏ Consider the Shell<br>❏ Navigating Files and Directories |
| Oct 16 | R | Introduction to R and RStudio (LG 2 ,4) | *Data Carpentry: Ecology Curriculum*<br>❏ Getting Started<br>❏ Introduction to R<br>❏ Starting with data<br>*R Tutorial*<br>❏ Vector Reference<br>❏ (***read links at bottom***)<br>❏ Data Frame Reference<br>*R for Data Science*<br>❏ Ch. 4: Workflow: basics |
| Oct 23 | R | Working with Data (*dplyr*) (LG 1, 3, 4) | ❏ dplyr vignette<br>*Optional Resources:*<br>❏ DCEC: Analyzing data with dplyr<br>❏ R4DS Ch. 5: Data transformation |
| Oct 30 | R | Data Visualization (*ggplot*) (LG 5) | *R for Data Science*<br>❏ Ch. 3: Data visualisation |

| | | | |
|---|---|---|---|
| Nov 6 | R | Project Structure<br>(LG 2, 6) | *Karl Broman*<br>❏ Initial Steps Toward Reproducible Research<br>❏ (***also read the first link in bulleted list- "Organize your data and code"***) |
| Nov 13 | R | Version Control<br>(LG 6) | |
| Nov 20 | R | Projects<br>(LG 4, 5, 6) | |
| Nov 27 | R | Dynamic report generation with R - (knitr)<br>(LG 5) | *R for Data Science*<br>❏ Ch. 27: R Markdown |
| Dec 1 | R | Creating tidy data - (tidyr)<br>(LG 4, 5) | *R for Data Science*<br>❏ Ch. 12: Tidy data |