Course Syllabus

✎ **Edit**

Instructor: Noah Simon (email nrsimon@uw.edu, office HSB F650; phone 685-3375)

Office hours: Monday 1:30-2:30 (or by appointment).

Course Time/Location:

Monday/Wednesday 11:30am - 1:20pm, HST T530

Course Description:

There have been many new technological advances and funding initiatives for biomedical data acquisition and storage. With this new wealth of data, we now have the potential to ask and answer many pivotal questions in public health, biology, and medicine. However there is a long pipeline from data to knowledge, and it requires a heterogeneous collection of skills. The art of traversing this entire pipeline to pose and answer scientific questions from data has come to be known as "Data Science."

This course provides an introduction to biomedical data science with an emphasis on statistical perspectives: the process of collecting, organizing, and integrating information toward extracting knowledge from data in public health, biology and medicine. The course covers four critical aspect of data science: (i) data acquisition, and common data models; (ii) data transformation and integration (wrangling); (iii) interactive and dynamic exploratory data analysis; and (iv) data visual analytics. Students will learn how to apply these ideas and create reproducible data analyses using a powerful open-source tool chain (R-studio, ggplot, github, etc). There will also be a central focus on statistical principles that guide all aspects of healthcare data analytics including the relevance of sampling schemes on potential generalizability, and the impact of both data quality/completeness and potential confounding factors on the reliability of scientific conclusions.

Students will leave this course with 3 things: 1) a conceptual overview of the ideas in biomedical data science, 2) a toolset for the acquisition, analysis, and reporting of biomedical data, and 3) hands-on experience asking and answer scientific questions with modern, relevant biomedical data.

Learning Objectives:

Upon completion of this course, students will be able to:

- Use a powerful open-source tool chain to:

1. Extract and manipulate data from common biomedical data repositories
2. Use simple numeric and visual summaries to explore biomedical data
3. Create reproducible analyses and reports

- Communicate conclusions to a general scientific audience
- Understand and judge statistical factors that determine how much the reliability of scientific conclusions (sampling biases, data quality/completeness, confounding)

- Understand the ideas behind some of the modern deep learning tools.

At the end of this course students will have basic competence in full-stack data science for moderate-scale biomedical data.

Outline of Course Topics:

- Intro to R (with R-studio)
- Overview of various biomedical data and tools for collecting such data (omics data, electronic health records, surveys)
- Data wrangling (data integration, data quality control, data cleaning, and query tools such as SQL) in R (with dplyR and RSQL)
- Basic methods for processing unstructured data such as imaging and text (natural language processing)
- Evaluate uncertainty using data resampling techniques
- Data models, summarization, and exploratory data analysis
- Healthcare data visualization (with ggplot2)
- Reproducible analyses and report generation (with github and R-markdown)
- Statistical issues which affect the reliability of scientific conclusions (Sampling, Confounding, and Inference)
- Intro to deep learning

Methods of Grading:

Student performance will be assessed using 4 bi-weekly homeworks (45%), a significant culminating project (45%), and in class warmups/participation (10%). The project will begin early in the quarter. As students learn the various aspects of data science, they will apply that understanding to their projects. The students will submit their projects as both a written report and a "flash talk" (short oral presentation). The projects will be graded on 3 criteria:

- How well the student employed the tools of data science to attack the question they pose
- The quality of the written report
- The quality of the oral presentation

In assessing the "quality" of the products, clarity of exposition, graphics, and overall polish will be taken into account.

Readings:

There will be no required texts. The following supplemental texts are suggested:

Grolemund G, Wickham H, "R for Data Science."

Wickham H, "ggplot2: Elegant Graphics for Data Analysis."

Horton, N and Kleinman K, "Using R and RStudio for Data Management, Statistical Analysis and Graphics, Second Edition."

Academic Integrity Statement - Students at the University of Washington (UW) are expected to maintain the highest standards of academic conduct, professional honesty, and personal integrity.

The UW School of Public Health (SPH) is committed to upholding standards of academic integrity consistent with the academic and professional communities of which it is a part. Plagiarism, cheating, and other misconduct are serious violations of the University of Washington Student Conduct Code (WAC 478-120). We expect you to know and follow the university's policies on cheating and plagiarism, and the **SPH Academic Integrity Policy** **(http://sph.washington.edu/students/academicintegrity/)**. Any suspected cases of academic misconduct will be handled according to University of Washington regulations. For more information, see the University of Washington Community Standards and Student Conduct website.

UW Disability Statement – Access and Accomodations

Your experience in this class is important to me. If you have already established accommodations with Disability Resources for Students (DRS), please communicate your approved accommodations to me at your earliest convenience so we can discuss your needs in this course.

If you have not yet established services through DRS, but have a temporary health condition or permanent disability that requires accommodations (conditions include but not limited to; mental health, attention-related, learning, vision, hearing, physical or health impacts), you are welcome to contact DRS at 206-543-8924 or **uwdrs@uw.edu (mailto:uwdrs@uw.edu)** or **disability.uw.edu. (http://depts.washington.edu/uwdrs/)** DRS offers resources and coordinates reasonable accommodations for students with disabilities and/or temporary health conditions. Reasonable accommodations are established through an interactive process between you, your instructor(s) and DRS. It is the policy and practice of the University of Washington to create inclusive and accessible learning environments consistent with federal and state law.

UW Inclusion Statement

The UW School of Public Health seeks to ensure all students are fully included in each course. We strive to overcome systemic racism by creating an environment that reflects community and mutual caring, while we ally with others in combating all forms of social oppression. This is a work in progress, as transformation is rarely a fully-completed project. In this course, we will look for opportunities to improve our performance as we seek to break down institutional racism. This can include course readings, class interactions, faculty performance, and/or the institutional environment. We encourage students to talk to your faculty member and/or the program director if you have concerns about classroom climate. **DCinfo@uw.edu (mailto:DCinfo@uw.edu)** is a resource for students with classroom climate concerns.

Religious Accommodations

Washington state law requires that UW develop a policy for accommodation of student absences or significant hardship due to reasons of faith or conscience, or for organized religious activities. The UW's policy, including more information about how to request an accommodation, is available at Religious Accommodations Policy

(https://registrar.washington.edu/staffandfaculty/religious-accommodations-policy/).

Accommodations must be requested within the first two weeks of this course using the Religious Accommodations Request form

(https://registrar.washington.edu/students/religious-accommodations-request/).

Course Summary:

Date	Details	
Sun Sep 29, 2019	📝 **Homework 0** **(https://canvas.uw.edu/courses/1318370/assignments/5026483)**	due by 11:59pm
Tue Oct 8, 2019	📝 **Project Preference** **(https://canvas.uw.edu/courses/1318370/assignments/5041540)**	due by 11:59pm
Sun Oct 13, 2019	📝 **Homework 1** **(https://canvas.uw.edu/courses/1318370/assignments/5041428)**	due by 5pm