SYLLABUS & COURSE INFORMATION

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

PUBH 7462, SECTION 001

Advanced Programming and Data Analysis in R
Spring 2019

COURSE & CONTACT INFORMATION

Credits: 2
Meeting Days: Friday
Meeting Time: 2:30 p.m. – 4:25 p.m.
Meeting Place: Bruininks Hall 530A

Instructor: Julian Wolfson (he/him/his)
Office Address: Mayo A453
Office Phone: (612) 625-9514
Fax (612) 626-0660
E-mail: julianw@umn.edu
Office Hours: TBD

COURSE DESCRIPTION

This course is intended for students, both within and outside the School of Public Health, who are relatively proficient with R, and are looking to improve their coding and data analysis skills. The emphasis of the course will be on learning tools and techniques which are useful to students who will be doing non-trivial statistical programming and/or data analysis in either a research or production environment. The course will focus on techniques for organizing, processing, and presenting data; in-depth knowledge of statistical modeling techniques is not required.

COURSE PREREQUISITES

PUBH 7461 (Exploring and Visualizing Data in R) or equivalent experience in R programming with instructor consent.

COURSE GOALS & OBJECTIVES

Upon completing this course, students should be able to:

- Understand and use common coding style and conventions in R
- Use Github for version control
- Understand and produce work that is reproducible using RMarkdown
- Use R to interact with databases and conduct basic data merges
- Use Shiny to create interactive tables and graphics in R
- Use R to visualize spatial data (mapping)
- Write and document a basic R package

METHODS OF INSTRUCTION AND WORK EXPECTATIONS

This course is a mixture of lecture, demonstration and hands on activities, with time spent outside of class on programming activities. Students are expected to attend class, participate in class discussions, and complete the assigned homework and project. **Working together on homework assignments is permitted, but each student is expected to independently write-up homework assignments, using their own computing and in their own words.**

Course Workload Expectations
Advanced Programming and Data Analysis in R is a 2-credit course. The University expects that for each credit, you will spend a minimum of three hours per week attending class or comparable online activity, reading, studying, completing assignments, etc. over the course of a 15-week term. Thus, this course requires approximately 90 hours of effort spread over the course of the term to complete the required material.

Learning Community

Many School of Public Health courses ask students to work in teams or discussion groups. We do not come to our courses with identical backgrounds and experiences and building on what we already know about collaborating, listening, and engaging is critical to successful professional, academic, and scientific engagement with topics.

In this course, students are expected to engage with each other in respectful and thoughtful ways.

In group work, this can mean:

- Setting expectations with your groups about communication and response time during the first week of the semester (or as soon as groups are assigned) and contacting the TA or instructor if scheduling problems cannot be overcome.
- Setting clear deadlines and holding yourself and each other accountable.
- Determining the roles group members need to fulfill to successfully complete the project on time.
- Developing a rapport prior to beginning the project (what prior experience are you bringing to the project, what are your strengths as they apply to the project, what do you like to work on?)

In group discussion, this can mean:

- Respecting the identities and experiences of your classmates.
- Avoid broad statements and generalizations. Group discussions are another form of academic communication and responses to instructor questions in a group discussion are evaluated. Apply the same rigor to crafting discussion posts as you would for a paper.
- Consider your tone and language, especially when communicating in text format, as the lack of other cues can lead to misinterpretation.

Like other work in the course, all student to student communication is covered by the Student Conduct Code (https://z.umn.edu/studentconduct).

COURSE TEXT & READINGS

There is no required text and no formal readings for this course; students will learn from hands-on in-class exercises and by consulting help documentation and online sources (e.g., DataCamp, StackOverflow). However, you may find the following additional references useful:

- Hadley Wickham. "ggplot 2: Elegant graphics for data analysis". Available through U of M (http://link.springer.com.ezp3.lib.umn.edu/book/10.1007%2F978-0-387-98141-3)
- Hadley Wickham. "R for Data Science". (http://r4ds.had.co.nz/)
- Jenny Bryan. "Happy Git and GitHub for the useR". (http://happygitwithr.com)
- Yihui Xie. "Dynamic Documents with R & knitr". (https://www.amazon.com/dp/1482203537/)
- Garrett Grolemund. "Hands-On Programming with R". (http://shop.oreilly.com/product/0636920028574.do#)
- Norman Matloff. "The Art of R Programming". (https://www.nostarch.com/artofr.htm)
- Hernan G. Resnizky. "Learning Shiny". (https://www.packtpub.com/application-development/learning-shiny)
- R Shiny (https://shiny.rstudio.com/)
- R Markdown (http://rmarkdown.rstudio.com)

Week	Topic
Week 1 Jan. 25th	• **Good coding practices**
Week 2 Feb. 1st	• **Writing custom functions**
Week 3 Feb. 8th	• **Version control** • **Introduction to GitHub**
Week 4 Feb. 15th	• **R Markdown**
Week 5 Feb. 22nd	• **R Shiny**
Week 6 March 1st	• **More R Shiny**
Week 7 March 8th	• **Getting data: conversions, databases, and web scraping**
Week 8 March 15th	• **Mapping with R**
Week 9 March 22nd	**SPRING BREAK**
Week 10 March 29th	• **Interactive data visualizations (plotly)**
Week 11 April 5th	• **Text mining and sentiment analysis**
Week 12 April 12th	• **Writing and documenting R packages**
Week 13 April 19th	• **TBD**
Week 14 April 26th	• **Final Project Work Day**
Week 15 May 3rd	**FINAL PROJECT PRESENTATIONS**

Assignment and project due dates will be posted on the course Canvas site.

SPH AND UNIVERSITY POLICIES & RESOURCES

The School of Public Health maintains up-to-date information about resources available to students, as well as formal course policies, on our website at www.sph.umn.edu/student-policies/. Students are expected to read and understand all policy information available at this link and are encouraged to make use of the resources available.

The University of Minnesota has official policies, including but not limited to the following:

- Grade definitions
- Scholastic dishonesty
- Makeup work for legitimate absences
- Student conduct code
- Sexual harassment, sexual assault, stalking and relationship violence
- Equity, diversity, equal employment opportunity, and affirmative action
- Disability services
- Academic freedom and responsibility

Resources available for students include:

- Confidential mental health services
- Disability accommodations
- Housing and financial instability resources
- Technology help
- Academic support

EVALUATION & GRADING

There will be 8 homework assignments which focus on application of a major skill for the course. Students will also complete a group project in which they build a Shiny application based on the knowledge they have gained during the two course sequence. The project grade will be based on the quality of your group's app/report/presentation (50%), the quality of your peer evaluation of other projects (25%), and on your active and timely participation in the project (25%), as determined by group member feedback and by the instructor's judgement based on such things as your attendance at and participation in required meetings.

For assignment and project due dates, please see the course website.

The final grade will be determined as:

60% homework (7.5% per assignment), 40% group project

No make-up work will be allowed for missed assignments or projects.

Grading Scale
The University uses plus and minus grading on a 4.000 cumulative grade point scale in accordance with the following, and you can expect the grade lines to be drawn as follows:

% In Class	Grade	GPA
93 - 100%	A	4.000
90 - 92%	A-	3.667
87 - 89%	B+	3.333
83 - 86%	B	3.000
80 - 82%	B-	2.667
77 - 79%	C+	2.333
73 - 76%	C	2.000
70 - 72%	C-	1.667

67 - 69%	D+	1.333
63 - 66%	D	1.000
< 62%	F	

- A = achievement that is outstanding relative to the level necessary to meet course requirements.
- B = achievement that is significantly above the level necessary to meet course requirements.
- C = achievement that meets the course requirements in every respect.
- D = achievement that is worthy of credit even though it fails to meet fully the course requirements.
- F = failure because work was either (1) completed but at a level of achievement that is not worthy of credit or (2) was not completed and there was no agreement between the instructor and the student that the student would be awarded an I (Incomplete).
- S = achievement that is satisfactory, which is equivalent to a C- or better
- N = achievement that is not satisfactory and signifies that the work was either 1) completed but at a level that is not worthy of credit, or 2) not completed and there was no agreement between the instructor and student that the student would receive an I (Incomplete).

Evaluation/Grading Policy	Evaluation/Grading Policy Description
Scholastic Dishonesty, Plagiarism, Cheating, etc.	You are expected to do your own academic work and cite sources as necessary. Failing to do so is scholastic dishonesty. Scholastic dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis (As defined in the Student Conduct Code). For additional information, please see https://z.umn.edu/dishonesty The Office for Student Conduct and Academic Integrity has compiled a useful list of Frequently Asked Questions pertaining to scholastic dishonesty: https://z.umn.edu/integrity. If you have additional questions, please clarify with your instructor. Your instructor can respond to your specific questions regarding what would constitute scholastic dishonesty in the context of a particular class-e.g., whether collaboration on assignments is permitted, requirements and methods for citing sources, if electronic aids are permitted or prohibited during an exam. Indiana University offers a clear description of plagiarism and an online quiz to check your understanding (http://z.umn.edu/iuplagiarism).
Late Assignments	Late assignments will incur a penalty of 25% per day, up to a maximum of 4 days.
Attendance Requirements	
Extra Credit	Opportunities for extra credit may be available on work assigned to the entire class. However, no additional extra credit activities will be created at the end of the semester.

CEPH COMPETENCIES

Competency	Learning Objectives	Assessment Strategies

3. Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software as appropriate	- Understand and implement good coding practices - Be able to use advanced R tools to organize, summarize, and visualize data in compelling way - Be familiar with tools for software version control	- In-class coding activities - Out-of-class assignments (including DataCamp courses) - Final group project