Statistical Graphics and Communication

COURSENUMBER: Fall 2020

Teaching Team

Instructor
Kim Hochstedler
(hochsted@umich.edu)
Office: M4141, SPH II

Insert photo of TA. ☺

Teaching Assistant
Name

Insert photo of TA. ☺

Teaching Assistant
Name

Insert photo of TA. ☺

Teaching Assistant
Name

Lectures, Labs, and Office Hours

	Monday	Tuesday	Wednesday	Thursday	Friday
9:00 am					
10:00 am	Lecture	TA Office Hour	Lecture		TA Office Hour
11:00 am					
12:00 pm					
1:00 pm					
2:00 pm	TA Office Hour	Kim's Office Hour		Lab	
3:00 pm					

*****All TA Office Hours will be located in ROOM NUMBER.**

Textbooks: None required. Useful references listed below.
Advanced R, by Hadley Wickham (https://adv-r.hadley.nz/index.html)
ggplot2: Elegant Graphics for Data Analysis, by Hadley Wickham
Introductory Statistics with R, by Peter Delgaard
R Cookbook, by Paul Teetor

Lab Exam: There will be one lab exam on Wednesday, October 7th.

Static Graphics Project: After mid-semester, there will be a group project involving a poster presentation of static graphics. Presentations will take place on November 20th.

Interactive Graphics Project: There is no final exam. There will be a final project, which will culminate in an oral presentation (due during finals week).

Pre-requisites: BIOSTAT 600 or another introductory undergraduate statistics course.

Course Website: www.umich.instructure.com/courses/mycoursenumber.html
Course Discussion Board: piazza.com/umich/fall2020/coursenumber/home

Course Description

There are many ways to graphically display quantitative information that help biostatisticians and clinical collaborators understand the conclusions drawn from data and statistical models. This course will introduce first- and second-year graduate students in biostatistics to state-of-the-art graphical displays and appropriate ways to communicate statistical results. Students will learn how to create graphical displays of data in RStudio and how to compile those displays into professional reports. Through in-class practice, a midterm, and a final project, students will also develop skills to critique and speak about graphics in a statistically sound manner. Each student will be required to complete a midterm and projects using graphical methods to understand data from real health science problems.

Course Goals

1. **Understand the fundamentals of data and reproducible data analysis.**
    - Distinguish between data types.
    - Write easily readable and reproducible code to explore and analyze data sets using R, R Studio, and R Markdown.
2. **Create statistical graphics.**
    - Learn strategies to develop easily readable and understandable statistical graphics.
    - Incorporate statistical information (i.e. results of statistical tests) into data visualizations.
    - Create both static and interactive graphics.
3. **Think critically about statistical graphics.**
    - Complete objective analyses of data to form thoughtful and meaningful conclusions.
    - Understand the extent to which you should draw conclusions and make claims about statistical graphics and analyses.
    - Review others' statistical graphics objectively and professionally.
    - Describe the advantages and disadvantages of a given data visualization choice.
4. **Write about statistical graphics.**
    - Accurately and concisely describe statistical graphics.
    - Incorporate appropriate statistical language into written descriptions of graphics.
5. **Speak about statistical graphics.**
    - Give oral presentations of statistical graphics to both technical and non-technical audiences.
    - Demonstrate the features of interactive graphics.

Course Competencies

- **Evidence-based Approaches to Public Health-3:** Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate.
- **Evidence-based Approaches to Public Health-4:** Interpret results of data analysis for public health research, policy, or practice.
- **Evidence-based Approaches to Public Health-19:** Communicate audience-appropriate public health content, both in writing and through oral presentation.

Course Components

1. **Lectures:** Lectures will cover the main topics of the course. Printed lecture outlines (to be annotated and filled in during lecture) will be available for you at the beginning of class. You are responsible for any announcements and materials covered in lectures.
   Most (but not all) lecture outlines and example code will be posted on the course website.

2. **Labs:** Lab sessions are designed to add context and allow hands-on practice of topics covered in lecture. These sessions allow you to begin working on lab assignments for the upcoming week. The instructor and TA(s) are in lab to help you, so please ask questions when you need assistance. Additionally, feel free to talk with other students, ask other students for help, and help other students with the lab.
   You are encouraged to use either the computers in the computer cluster, or to use your own personal computer during lab.
   Lab attendance is mandatory. Please see additional details under "Oral Evaluations" below.

3. **Oral Evaluations:** In each lab session, a random selection of students will be asked to give a short, one-on-one presentation of a single graphic to the instructor. The graphic will be provided to you before lab. During the first week of the course, the instructor will give an example of what is expected for these oral presentations. No students will be required to give presentations during Lab 1.
   These oral evaluations are strictly for your benefit and will be graded on **completion only**. If you are not in lab when you are selected to give a presentation, you will lose credit. You may be selected for oral evaluations multiple times during the semester. The instructor will provide feedback on these presentations in order to help you improve their oral presentation skills.

4. **Lab Assignments (Homework):** Lab assignments give you the opportunity to apply and practice the data analysis and report writing, based off of material covered in lecture. As such, it is expected that you treat lab assignments like formal reports. All material covered in lab assignments may show up in later lab assignments and in the lab exam and projects.
   Lab assignments are designed to be started in Thursday lab sessions. Lab assignments will be released Wednesday after lecture so that you may prepare for the assignment and bring specific questions to lab if you so choose. Lab assignments are due on Wednesdays at 9:59 am, submitted through the course website as a single .Rmd file, unless otherwise specified.

5. **Code:** All code should be written in R and R Markdown, freely available at https://cran.r-project.org/. You should follow Hadley Wickham's Advanced R Style Guide (http://adv-r.had.co.nz/Style.html). If your submitted code does not adhere to this style guide, you will lose up to 10% credit on that assignment. If you are an experienced R programmer who wishes to use a different (but well-defined) style guide, please discuss this with the instructor.

6. **Lab Exam:** There is one lab exam during the semester. The lab exam is designed to assess your proficiency in R coding for statistical analysis. There will be a practice lab exam the week before the exam date (Wednesday, October 7th). More details about the content and the format of the lab exam will be made available closer to the exam date.

7. **Static Graphics (Midterm) Project:** Groups of students will select a data set to analyze with respect to a research question of interest. Your group will create a poster describing their work. A public group presentation of the posters will take place on Friday, November 20th. As part of your grade on this

project, you will evaluate your own group performance and will peer review the work of another group. More details about the expectations and guidelines for the static graphics project will be made available after Fall Break.

8. **Interactive Graphics (Final) Project:** You will select a data set to analyze with respect to a research question of interest. You will create an R Shiny web application displaying interactive graphics that serve to answer their research question. A presentation of the app will occur during the regularly scheduled final exam period. As part of your grade on this project, you will reflect on your own work and will peer review the work of another student. More details about the expectations and guidelines for the interactive graphics project will be made available after Thanksgiving Break.

What can you take away from these Course Components after this semester?

- Lecture notes will serve as a reference guide for coding and analysis for later classes and projects.
- Oral evaluation feedback (provided during lab sessions) will guide students to become better presenters.
- Lab assignments will serve as example statistical reports that students can use as (1) guides for later projects and (2) evidence of proficiency in report writing and coding to employers.
- Adherence to a style guide will allow your code to be easily read and reproduced by others.
- The lab exam will serve as evidence of your efficiency in coding and simple analyses.
- The static graphics project gives you an opportunity to make and present a scientific poster, and work effectively in a group (Hint: put this presentation on your CV!)
- The interactive graphics project gives you an opportunity to make and present an R Shiny web application (Hint: put this on your CV too!)

Grading Policies

- All numeric grades are on a scale from 0 to 100.
- Final grades will be computed according to the following weights:

Average Lab Assignment Score	50.0%
Oral Evaluations (completion only)	5.0%
Lab Exam Score	12.5%
Static Graphics Project	12.5%
Interactive Graphics Project	20.0%

- Final letter grades will be determined according to the following rules (subject to change at the instructor's discretion, though lower bound of each letter grade cut off will *not* increase).

A	≥ 90
B	[80, 90)
C	[70, 80)
D	[60, 70)
R	< 60

- The lowest lab assignment score is not included in the calculation of final grades.

Administrative Procedures and Logistics

Lectures: Much of the learning in lecture will be done through active participation, so you should be prepared to participate during class. Use common courtesy: please arrive on time, do not leave early, and do not be disruptive in class. You are responsible for any announcements and materials covered in lectures.

Course Materials, Canvas: The syllabus, lab assignments, oral evaluation graphics, assignment solutions, assigned readings, any supplementary material, grades, and announcements for the course can be found on the course web page on Canvas: www.umich.instructure.com/courses/mycoursenumber.html. **Please check Canvas regularly.**

Discussion Board: All class discussions that take place outside of lecture and lab should occur on Piazza: piazza.com/umich/fall2020/coursenumber/home

Communication: If you have questions related to the class material, lab assignments, the lab exam, or projects, feel free to ask the instructor during class or, preferably, the instructor and TA(s) during office hours. Questions concerning the current homework assignment submitted by email will not be answered. Please use email only to address administrative and logistic issues, or personal questions.

Please include COURSENUMBER in the subject line of all emails. You should expect a reply within 3 days, with the exception of Fall Break and Thanksgiving Break. Questions about lab assignments should be submitted to the course discussion board. **Emails to TAs will not be answered.**

Lab Assignment (Homework) Format: Lab assignments should have the students name and uniqname at the very top and should be written using R Markdown. Students should use Hadley Wickham's Advanced R Style Guide to write their code. All answers should be clearly labeled. Lab assignments should be treated as formal reports. This means that proper spelling and grammar is expected. Deviating from this format may result in your assignments not being graded.

You are encouraged to discuss homework problems and collaborate with your classmates (both on the discussion board and in person). However, the work you submit must be your own. This means that each student must independently write up each problem, including all code and written responses. Instances of identical, nearly identical, or copied lab assignments will be considered cheating and plagiarism.

Extensions: In general, extensions will not be granted for students (because they are behind on work, had a busy week, etc). Extensions for **reasonable academic purposes** (e.g. job interview, conference attendance) or **extreme circumstances** (e.g. family emergency) may be granted at the instructor's discretion. If you have a request for an extension, please request this at least 48 hours before the assignment is due, if possible. Students should submit proof of the issue when requesting an extension. If you require special accommodations via disability services or for religious observance, please see "Student Accommodations" below.

Regrades: If you believe a mistake was made when your assignment was graded, you must write a clear description of the issue. Please include your name, and the number of points under consideration at the top of the page. Please submit this, along with a copy of your assignment to the instructor's mailbox (SPH II) or via email **within one week of when the assignment was graded**.

Academic Integrity: The faculty and staff of the School of Public Health believe that the conduct of a student taking courses in the School should be consistent with that of a professional person. Courtesy, honesty, and respect should be shown by students toward faculty members, guest lecturers, administrative support staff, community partners, and fellow students. Similarly, students should expect faculty to treat them fairly, showing respect for their ideas and opinions and striving to help them

achieve maximum benefits from their experience in the School.

Student academic misconduct refers to behavior that may include plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials (including library materials), and aiding and abetting the perpetration of such acts. Please visit http://sph.umich.edu/student-resourses/mph-mhsa.html for the full SPH Code of Academic Integrity and further definition of these terms.

Cell Phones, Laptops, Tablets, etc: Students may bring a laptop or tablet to lectures, though handwritten lecture notes are provided. Students may utilize a personal computer during lab sessions if they choose. Please no cell phone use during class.

Phone, Audio, and Video Recording: Photo, audio, and video recordings of the course lectures, lab session, lab exam, and all other course materials are strictly prohibited.

SPH Writing Lab: The SPH Writing Lab is located in 5025 SPH II and offers writing support to all SPH students. The Lab can also help answer questions on academic integrity. To learn more or make an appointment, please visit https://sph.umich.edu/writing-lab/.

Student Well-being: SPH faculty and staff believe it is important to support the physical and emotional well-being of our students. If you have a physical or mental health issue that is affecting your performance or participation in any course, and/or if you need help connecting with university services, please contact the instructor of the Office of Graduate and Postdoctoral Studies. Please visit https://sph.umich.edu/student-life/wellness.html for more information.

Student Accommodations: Students should speak with the instructor before or during the first week of classes regarding any special needs and no later than two weeks before an exam or assignment due date. Students can also visit the SPH Office for Student Engagement and Practice for assistance in coordinating communications around accommodations. Students seeking academic accommodations should register with Services for Students with Disabilities (SSD). SSD arranges reasonable and appropriate academic accommodations for students with disabilities. Please visit https://ssd.umich.edu/topic/our-services for more information on student accommodations.

Students who expect to miss classes, examinations, or other assignments as a consequence of their religious observance shall be provided with a reasonable alternative opportunity to complete such academic responsibilities. It is the obligation of the student to provide faculty with two weeks' notice of the dates of religious holidays on which they will be absent. Please visit https://www.provost.umich.edu/calendar/religous_holidays.html#conflicts for the complete university policy.

Diversity, Equity, and Inclusion: I am committed to creating an equitable learning environment in this class. The goal of this course is to bring together people from diverse backgrounds (be it educational, social, regional, etc.) and promote the different perspectives that each person brings to the art of data science. Regardless of an individual's prior experience with statistical visualization and communication, students will have the tools they need to become a proficient member of the classroom community. In that community, we value the diversity of opinions and approaches to data science that everyone brings to the table. That diversity enriches everyone's educational experience and results in a creative atmosphere where all student's efforts and perspectives are valued. After learning the core principles of statistical graphics and communication, students are encouraged to adapt the course material to benefit their own needs and interests. Fellow students are expected to respectfully and constructively comment on the data science methods other students employ in their work.

Course Schedule

- Course schedule is subject to change, but the dates of the **lab exam and the poster presentation** for the static graphics project are **fixed**.

Date	Class Session	Topic	Assignment Due
9/2	Lecture 1	Course introduction, goals of graphics	Reading: "Same Stats, Different Graphs" by Matejka and Fitzmaurice
9/3	Lab 1	Introduction to R and R Markdown, style guides, dplyr	Download RStudio
9/7	Lecture 2	Goals of graphics, core components of a plot, tables vs. plots	
9/9	Lecture 3	What is data?, types of data	Lab 1 due
9/10	Lab 2	Data manipulation and R Markdown, tables	
9/14	Lecture 4	1-dimensional categorical data	
9/16	Lecture 5	1-dimensional continuous distributions	Lab 2 due
9/17	Lab 3	1-dimensional graphics	
9/21	Lecture 6	2-dimensional categorical data	
9/23	Lecture 7	2-dimensional categorical data	Lab 3 due
9/24	Lab 4	2-dimensional graphics	
9/28	Lecture 8	2-dimensional continuous data	
9/30	Lecture 9	2-dimensional continuous data **Last day of material on lab exam**	Lab 4 due
10/1	Lab 5	Practice lab exam	**Lab 5 due**
10/5	Lecture 10	2-dimensional kernel estimation	
10/7	**Lab Exam**		**Lab Exam to be completed during lab session**
10/8	No class, fall break		
10/12	No class, fall break		
10/14	Lecture 12	3-dimensional data, maps and projections	
10/15	Lab 6	3-dimensional graphics I	
10/19	Lecture 13	Clustering, PCA	
10/21	Lecture 14	Longitudinal and time series data	Lab 6 due
10/22	Lab 7	3-dimensional graphics II	Select Static Graphics Project group
10/26	Lecture 15	Format of a statistical report	
10/28	Lecture 16	Format of a statistical report	Lab 7 due
10/29	Lab 8	General report outline	Select Static Graphics Project data set
11/2	Lecture 17	Format of a statistical presentation	
11/4	Lecture 18	Use of color in graphics	Lab 8 due
11/5	Lab 9	Making a presentation in R	
11/9	Lecture 19	Format of a statistical poster	Lab 9 due

11/11	Lecture 20	Work on static graphics project	
11/12	**Static Graphics Lab Session**	Work on static graphics project	
11/16	Lecture 21	Presentation skills for a poster	
11/18	Lecture 22	Practice poster presentation	**Print your poster before class**
11/19	**Static Graphics Poster Project Presentation**		Static Graphics Project group evaluation and reflection
11/23	No class, Thanksgiving break		
11/25	No class, Thanksgiving break		
11/26	No class, Thanksgiving break		
11/30	Lecture 23	Natural language processing, text data	Select Interactive Graphics Project data set
12/2	Lecture 24	Networks	
12/3	Lab 10	Special Topics	
12/7	Lecture 25	Interactive graphics, introduction to R Shiny	
12/9	Lecture 26	Presentation skills for a web application	Lab 10 due
12/10	**Interactive Graphics Lab Session**		
12/14	Lecture 27	Work on interactive graphics project	

- Interactive Graphics Presentations will take place during **the week of 12/14 to 12/18**.
- Interactive Graphics Project evaluation and reflection **due 12/18.**

Skills Checklist

- Use this checklist as you go through the semester to keep track of the skills you have mastered.

Before the lab exam: I can …

1-Dimensional Graphics: these skills are useful for exploratory data analysis.

- □ Make a bar plot.
- □ Make a histogram.
- □ Determine an appropriate binwidth for the data displayed in a histogram.
- □ Make a box plot.
- □ Explain why I shouldn't actually make a box plot very often.
- □ Print an aesthetically pleasing "Table1" using a package of my choice in R.
- □ Perform a t-test or chi-squared test in R.
- □ Identify when it is appropriate to use a t-test or chi-squared test.
- □ Interpret t-test and chi-squared test results

2-Dimensional Graphics: these skills are useful for exploratory data analysis, model diagnostics, and can be used in presenting analysis results.

- □ Make a stacked bar chart.
- □ Make a side-by-side bar chart.
- □ Make a scatter plot.
- □ Compute a correlation.
- □ Interpret a correlation.
- □ Perform a 2-sample t-test in R.
- □ Interpret the results of a 2-sample t-test.
- □ Print the results of a 2-sample t-test in an aesthetically pleasing way.
- □ Print an aesthetically pleasing contingency table using a package of my choice in R.

Before the static graphics project: I can…

Regression: this analysis technique is commonly used in statistical analyses.

- □ Fit a linear model in R.
- □ Fit a generalized linear model in R.
- □ Identify when it is appropriate to transform variables, and how I should transform them.
- □ Add error bars to my plots.
- □ Understand when and how to implement smoothing features in regression.
- □ Put a trend line on a scatter plot.
- □ Interpret the results of a linear regression…
    - Intercept.
    - Slope coefficient for a continuous variable.
    - Slope coefficient for a categorical variable.
    - Interaction term.

3-Dimensional Graphics and Special Topics: exploring new types of graphs helps you become fluent in R programming and ggplot2. These skills help you create more complex visual displays for reports and projects.

- □ Facet my graphs.
- □ Make a time series plot.
- □ Make a network diagram.
- □ Make a heat map.
- □ Add error bars to my plots.
- □ Include the results of statistical tests in my graphs.

***Not all special topics will be covered before the static graphics project. Please see course schedule for details, though this is subject to change.**

Before the interactive graphics project: I can …

Interactive Graphics: these skills will allow you to create an RShiny app that is useful for understanding a biomedical research question.

- □ Code an R Shiny app that actually launches.
- □ Add at least 2 interactive elements to a plot.
- □ Make multiple pages in my app.
- □ Design my R Shiny application to be aesthetically pleasing and user-friendly.
- □ Present the elements of an R Shiny app to non-technical audiences.
- □ Present the elements of an R Shiny app to technical audiences.

Throughout the semester: I can …

Skills in R: these skills will make you a better R programmer who can continue to learn new coding skills, even after this course ends.

- □ Install and use new statistical packages.
- □ Find and use R documentation and R help pages.
- □ Load data into R from multiple file formats (i.e. .csv, .sas7bdat, .txt)
- □ Export a .csv from R.
- □ Use the pipe "%>%" to manipulate data.
- □ Adhere to a style guide when coding.
- □ Document / comment my code.
- □ Find out the data type I working with using R.

Statistical Graphics (ggplot2) Skills: these skills will allow you to make professional looking statistical graphics.

- □ Change axis, legend, and graph titles.
- □ Add a caption to my graphics.
- □ Alter text size, angle, and color.
- □ Use appropriate colors and shapes in my graphics.
- □ Make my own theme for graphs I produce in ggplot2 / R.
- □ Use the annotate( ) feature to add text to my graph.

Writing and Speaking Skills: these skills will allow you to feel comfortable describing statistical graphics and statistical results in writing and in oral presentations.

- □ Identify the key take-aways from a statistical graphic, even if I did not make it.
- □ Give constructive feedback on the strengths and weaknesses of a given graphical choice.
- □ Identify potential sources of bias in graphs.
- □ Present a statistical graphic to a non-technical audience.
- □ Present a statistical graphic to a technical audience.
- □ Generate a statistical report in R Markdown.
- □ Generate presentation slides in R Markdown.