36-315
Statistical Graphics and Visualization

Spring 2018

Instructor:	Jerzy Wieczorek (jwieczor@andrew.cmu.edu)
Lectures:	Monday and Wednesday, 12:30pm–1:20pm, DH 1212
Labs:	Friday, 12:30pm–1:20pm, BH 140CEF + HL CLSTR
Weekly Assignments:	Homeworks are due on Wednesday at 11:59pm, unless otherwise instructed. Labs are due on Friday at 6:30pm, unless otherwise instructed.
Lab Exam:	There will be one lab exam on **Friday, March 2**.
Static Graphics Project:	After mid-semester, there will be a group project involving a poster presentation of static graphics (scheduled for **Friday, April 13**).
Interactive Graphics Project:	There is no final exam. There will be a final group project involving an interactive graphic, accompanying paper, and peer critiques (due finals week).

Course Description

Graphical displays of quantitative information take on many forms as they help us understand both data and models. This course will serve to introduce the student to the most common forms of graphical displays and their uses and misuses. Students will learn both how to create these displays and how to understand them. As time permits, the course will consider some more advanced graphical methods such as interactive graphics, computer-generated animations, maps, network graphics, etc. Each student will be required to engage in projects using graphical methods to understand data collected from a real scientific or engineering experiment. In addition to two weekly lectures, there will be lab sessions where the students learn to use software to aid in the production of appropriate graphical displays.

Textbooks

	None of these are required.
Overviews:	*The Functional Art,* by Alberto Cairo. New Riders, 2012.
	Fundamental Statistical Concepts in Presenting Data, by Rafe Donahue. http://biostat.mc.vanderbilt.edu/wiki/Main/RafeDonahue
R code:	*ggplot2: Elegant Graphics for Data Analysis,* by Hadley Wickham. Springer, 2016. PDF: https://link.springer.com/book/10.1007/978-3-319-24277-4
	R Graphics Cookbook, by Winston Chang. O'Reilly, 2013.
Classics:	*Visualizing Data,* by William Cleveland. Hobart Press, 1993.
	The Elements of Graphing Data, by William Cleveland. Hobart Press, 1994.
	The Visual Display of Quantitative Information, by Edward Tufte. Graphics Press, 1983.
	Envisioning Information, by Edward Tufte. Graphics Press, 1990.
Graphic design:	*The Non-Designer's Design Book,* by Robin Williams. Peachpit Press, 2014.

Course Objectives

1. **Understand the Fundamentals of Data and Reproducible Data Analysis.**

   - Distinguish between data types
   - Write easily readable and reproducible code to explore datasets graphically
   - Master the use of R, RStudio, RMarkdown, and other tools to promote reproducible research and allow others to build from your work

2. **Create Statistical Graphics.**

   - Create easily readable and understandable statistical graphics
   - Master the use of R, RStudio, and RMarkdown to explore datasets graphically
   - Incorporate statistical information (e.g. the results of statistical tests) into elegant data visualizations
   - Create both static and interactive graphics for mass public consumption

3. **Write About Statistical Graphics.**

   - Describe statistical graphics and data visualizations in detail, but concisely
   - Incorporate appropriate statistical language into written descriptions of graphics

4. **Critique Statistical Graphics.**

   - Review others' statistical graphics objectively and academically
   - Describe the pros and cons of a given graphical choice
   - Give useful critiques, feedback, and suggestions for improvement on others' graphics

Course Components

1. **Lectures.** The main topics of the course will be covered during the lecture. You are also responsible for any additional material covered in the assigned readings, labs, and homework.

   If you miss a lecture, you are responsible for the material covered during the lecture you have not attended. Students are expected to take notes and follow along with example problems in class. Some (but not all) course notes and example code will be posted on the course website.

   During each lecture, we will also assign short small-group activities, resulting in a sketch or short writeup to be handed in at the end of class. These informal activities are designed to help you learn and will not be graded for correctness, only monitored for participation/attendance. Please always write your Andrew IDs on the sheets you hand in.

2. **Labs.** Labs are specifically designed to add context and give examples (using real-world datasets) of the concepts covered in lecture. They are also designed to prepare students for homework assignments due the following week. Labs will typically include example code to ease the introduction into new concepts.

   Labs are designed to take 45 minutes to complete. The instructor and TAs are in lab to help you, so please ask questions when you need assistance. Additionally, please discuss the lab

with other students, ask other students for help, and help other students in lab, as long as the talking is not disruptive. Talking is not allowed during the lab exam, however.

Questions about the lab assignment will not be answered after the lab session has ended. Discussion board questions about lab assignments are not guaranteed to be answered. Emailed questions about lab assignments will not be answered.

Lab attendance is mandatory. You must attend the same lab section / room each week and sign in.

Lab assignments are due at 6:30pm on the day of lab (unless otherwise specified), submitted through the course website. Students should submit a single .Rmd file and its knitted .html output file, unless otherwise specified. (This will be more clear when you complete Lab 01.)

3. **Homework**. Homework problems provide you with the opportunity to learn, practice, and test your knowledge and understanding of the material. All material found in the homework may show up in later homeworks and/or the lab exam.

   Homeworks are due on Wednesdays at 11:59pm, submitted through the course website. Students should submit a single .Rmd file and its knitted .html output file, unless otherwise specified.

   We will give you adequate time to work on the problems, and the graders will work hard to return your homework in a timely manner. Unfortunately, this means that **late homework will not be accepted**. Instead, the grading policy allows for the equivalent of dropping 2 homeworks (see below).

4. **Code**. All code should be written in R and RMarkdown. Students should follow one of two popular style guidelines: (1) Google's R Style Guide or (2) Hadley Wickham's Advanced R Style Guide.

   Students should specify what style guide they are using at the top of their submitted code and assignment. If a student's submitted code does not adhere to one of these two style guides, students 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 talk to the instructor.

5. **Lab Exam.** There is one lab exam during the semester. Specific details about the content and format of the lab exam will be available closer to the exam date (Friday, March 2).

6. **Static Graphics Project.** There will be a midterm project. Groups of students will be assigned a dataset to analyze. Each group will create a poster describing their work. A public group presentation of the analyses is (tentatively) scheduled for Friday, April 13. More details will be available after mid-semester.

7. **Interactive Graphics Project and Paper.** There will be a final project. Groups of students will be assigned a dataset to analyze. A group paper describing the work is due during finals week. Students will also be assigned peer critiques of other groups' work. More details will be available near the end of the semester.

Grading Policies

- All numeric grades are on a scale from 0 to 100.

- Final grades are based on exams and homework, and will be computed according to the following weights:

Homework Score	35%
Lab Score	10%
Lab Exam Score	15%
Static Graphics Project/Paper/Presentation	15%
Interactive Graphics Project/Paper/Critiques	20%
Attendance	5%

- Final letter grades will be determined according to the following rules (subject to change at the instructor's discretion):

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

- No assignments are dropped when calculating mid-semester grades.

- There are 11 labs and 12 homework assignments, each worth 100 points. However, your final lab grades will be calculated out of 1000 points, up to 1000/1000, and likewise for your final homework grades: *Final Score = min(Total, 1000) / 1000*. Even if you miss a few points here and there, you can earn full-credit final lab and homework scores by submitting all assignments. Alternately, you can simply treat this as dropping your lowest lab score and lowest 2 homework scores.

Computing

- All code for all projects must be written in R unless otherwise specified.

- All course assignments must be written in R and RMarkdown unless otherwise specified.

- Students with laptops and personal computers should download the latest versions of R and RStudio. Instructions to do this will be given during the first week of classes.

- All students should **immediately** check their university computing accounts to make sure that R and RStudio are installed. If you cannot access these resources, please notify the instructor ASAP.

- Students are encouraged to use the campus computers in the computing cluster. Students are also permitted to use their own computers during lab, though any issues arising from using personal computers (e.g. hardware, software, or operating system compatibility) are the responsibility of the student to resolve.

Administrative Procedures and Logistics

- **Lectures.** Use common courtesy: arrive on time; do not leave early; no cell-phone use allowed; do not be disruptive in class; participate in class when the instructor asks questions; etc. The use of laptops/tablets/etc is allowed only for course-related purposes.

- **Course Materials: Canvas.** The syllabus, lab assignments, homework assignments, solutions, assigned readings, any supplementary material, and grades for this course can be found on the course web page on Canvas: https://canvas.cmu.edu/. **Please check Canvas regularly.**

- **Discussion Board: Piazza.** All class discussions that take place outside of lecture and lab will occur on Piazza: https://piazza.com/cmu/spring2018/36315

  Piazza will also be used to send out course announcements. **Please regularly check the email account that you linked with Piazza.**

  All students are responsible for understanding all discussions on Piazza. Important course material, example code, etc will be distributed via Piazza.

  Discussion on Piazza should remain civil and respectful. Students are encouraged to help answer each other's general questions but should not post answers to any lab, homework, or other assignment. **Posting homework answers on Piazza will be treated as a violation of the academic integrity policy.**

- **Communication.** If you have any questions related to the class material, homework problems and exams, feel free to ask the instructor during class or, preferably, the instructor and the TAs during their office hours.

  Questions about homework submitted by email will not be answered. Please use email only to address administrative and logistic issues. You should not expect a reply within 24 hours. Questions about homework should instead be submitted to the course discussion board.

- **Homework Format.** Homeworks should have the student's name and Andrew ID at the very top/beginning. Students should specify the style guide they used to write their code (see above). Questions should be answered in order. All answers should be clearly marked and labeled. Answers should be written in the context of the problem when applicable. Proper spelling and grammar should always be used – this means using complete sentences, proper punctuation, etc. Deviating from this format may result in your assignment not being graded.

  You are encouraged to discuss homework problems and collaborate with classmates. However, the work you submit must be **your own**. This means, in particular, that each student must independently write up each problem, including all code and written responses. **Instances of identical, nearly identical, or copied homework will be considered cheating and plagiarism. The use of material from previous semesters of this course or from any other source to solve homework and exam problems is regarded as unauthorized assistance and therefore as a violation of the Carnegie Mellon University code of academic integrity.**

- **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), **extreme circumstances** (e.g. hospitalization), or religious reasons may be

granted at the instructor's discretion. If you believe you have a reasonable request for an extension, please request this at least 48 hours before an assignment is due. Students should submit proof of the issue when requesting an extension. At the top of the assignment, please clearly write that you received an extension on the assignment.

If you require special accommodations via disability services, please see below.

- **Regrades.** If you believe a mistake was made when your assignment was graded, you must write a clear, detailed description of the issue. Please include your name, Andrew ID, and the number of points you expect to receive at the top of the page. Please submit this, along with a printed copy of your assignment, to the instructor's mailbox (Baker Hall 132 wing) **WITHIN ONE WEEK of when the assignment was graded**.

  Regraded assignments will be processed at the end of the semester, ONLY if they have the potential to influence your final letter grade.

- **Integrity.** All students are expected to comply with the CMU policy on academic integrity: https://www.cmu.edu/student-affairs/ocsi/academic-integrity/index.html

  Always ask if you are unsure whether your actions comply with the assignment instructions. Always acknowledge any help received on assignments: list the names of the people you worked with, and cite any external sources you used. You are encouraged to discuss assignments with your classmates, but the work you submit must be your own.

  Cheating, plagiarism and unauthorized assistance on homework or exams will be dealt with in accordance with the academic integrity policy. Cheating or copying of any sort are typically grounds for failure of the course.

- **Disability Services.** If you need special disability-related accommodations in this class, please contact the instructor **immediately** to make arrangements. Special accommodations for exams must be requested **no later than one week prior to the exam.** You should also contact the Disability Resources office at 412-268-6121, request the appropriate documentation, and give a copy of this documentation to the instructor. For more information, see the Carnegie Mellon Equal Opportunity Services and Disability Resources webpage: https://www.cmu.edu/disability-resources/students/index.html

- **Cellphones, Laptops, etc.**: All cellphones and anything else that makes noise should either be turned off or silenced during class. Students are expected to participate in class. Students should not use their cell phones in class for any purpose. This includes texting and checking email. Laptops may only be used for course-related purposes. Students are encouraged to follow along and run code in class.

- **Email:** Sending email to the instructor should be treated as professional communication. Emails should have an appropriate greeting and ending; students should refrain from using any kind of shortcuts, abbreviations, acronyms, slang, etc in the email text. Emails not meeting these standards may not be answered.

  Emails about homework/lab questions will not be answered. Please direct these questions to the course discussion board.

  Emails to the TAs will not be answered.

- **Discussion/Questions:** All questions about labs, homework, and notes should be directed to Piazza. Homework-related email to the instructor will not be answered. (The TAs will

not answer any email, whatsoever.) The discussion board will be checked regularly by the instructor and TAs. That said, in order to guarantee that a question is answered in time, please allow 24 hours in advance of when an assignment is due when asking a question on the discussion board.

Students are expected to subscribe to all Piazza threads at the beginning of the semester, so that they receive email notifications when a question or answer is posted.

- **Photo, Audio, and Video Recording:** Photo, audio, and video recordings of the course lectures, course labs, lab exams, and all other course materials are strictly prohibited.

  This includes, but is not limited to: using a cell phone to take pictures of the notes, recording video and/or audio of lectures, labs, exams, and other course settings.

- **Take care of yourself:** Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

  All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

  If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 and visit their website at https://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

  If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

  CaPS: 412-268-2922

  re:solve Crisis Network: 888-796-8226

  If the situation is life threatening, call the police:

  On campus: CMU Police: 412-268-2323

  Off campus: 911

Tentative schedule

Date	Topics
	Data Visualization Basics
W 1/17	Course introduction, history, classic examples
M 1/22	Best practices and principles for statistical graphics
W 1/24	Grammar of Graphics
	Core Graphs and their associated Statistics
M 1/29	1-D categorical data
W 1/31	2-D categorical data
M 2/5	1-D continuous data
W 2/7	1-D continuous data, split by a categorical variable
M 2/12	2-D continuous data – scatterplots and trend lines
W 2/14	2-D continuous data – joint distributions
	Review
M 2/19	Further principles for statistical graphics
W 2/21	Introduce the Lab Exam
M 2/26	Review for Lab Exam
W 2/28	Miscellaneous extra topics
F 3/2	LAB EXAM
	Advanced Topics and Poster Project
W 3/7	High-dim continuous data, clustering of cases, dendrograms
M-W 3/12-14	SPRING BREAK
M 3/19	Clustering of variables
W 3/21	Graphic design and color
M 3/26	Maps and cartography – principles
W 3/28	Maps and cartography – R demos
M 4/2	Time series and longitudinal data
W 4/4	Network data
M-W 4/9-11	Office hours for poster projects
F 4/13	POSTER SESSION
	Interactive Project
M 4/16	Interaction design, UX design, intro to Shiny
W 4/18	Shiny and Plotly
M 4/23	Office hours
W 4/25	Shiny Layouts
M-W 4/30-5/22	Office hours for interactive projects
M-W 5/7-9	FINALS WEEK
F 5/11	Final projects due