

36 - 315: Statistical Graphics and Visualization

Course Policies and Syllabus, Summer 2017

Instructor:	Ben LeRoy (bpleroy@stat.cmu.edu), OH : TBD
Graduate TA:	Alden Green, OH: TBD
Course Website:	Canvas
Lectures:	MWF 10:30AM - 11:50AM, Baker Hall 235A
Lab:	TTh 10:30AM - 11:50AM, College of Fine Arts 317
Weekly Assignments:	Two a week, due 11:59 pm the night before lab (due on Monday/Wednesday)
Lab Exam:	There will be a Lab Exam June 9th
Static Graphics Project:	Group project involving a poster presentation
Interactive Graphics Project:	Final group project oral presentation and paper
Prerequisite:	36202 or 36208 or 36226 or 88250 or 36309 or 36625 or 70208 or 36303 or 36225.
Textbooks:	None of these are required. <i>ggplot2: Elegant Graphics for Data Analysis</i> , by Hadley Wickham <i>Graphics for Statistics and Data Analysis with R</i> , by Keen. CRC Press, 2010. <i>The Functional Art</i> , by Alberto Cairo, 2014 <i>Creating More Effective Graphs</i> , by Naomi B. Robbins, 2013. <i>Introductory Statistics with R</i> , by Peter Dalgaard. Springer, 2002/2004. <i>Using R for Introductory Statistics</i> , by John Verzani, Chapman & Hall, 2004. <i>R Graphics</i> , by Paul Murrell. Chapman & Hall, 2006. <i>Visualizing Data</i> , by William S. Cleveland. Hobart Press, 1993. <i>Envisioning Information</i> , by Edward R Tufte. Graphics Press, 1990. <i>Tidy Text Mining with R</i> , by Julia Silge and David Robinson.

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.

OBJECTIVES

1. Demonstrate Understanding 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. Speak About Statistical Graphics.

- Give eloquent oral presentations of statistical graphics to both technical and non-technical audiences
- Enhance public speaking skills when presenting data visualizations
- Demonstrate interesting features of datasets via interactive graphics

5. 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.** There will be three lectures per week - Monday, Wednesday, and Friday. Time for breaks will be incorporated into the lectures to separate up topics. Attendance is expected for the entirety of the course. There may be short participation and reading quizzes during lectures.

2. **Assignments.** Assignment problems provide you with the opportunity to learn, practice, and test your knowledge and understanding of the material. You will have assignments due twice a week at 11:59 pm the day before each lab (11:59 pm Monday, and 11:59 pm Wednesday) unless otherwise noted. Students should submit a single .Rmd file, unless otherwise specified. All material found in the homework may show up in later assignments and/or the lab exam.

Each assignment will have a *lab* portion and a *homework* portion. The *lab* portion is expected to be completed in lab and is done for completion.

3. **Labs.** There are two labs per week - Tuesday and Thursday - during the usual class meeting time. The first portion of each homework is designed to be completed in class, and will be graded on competition. Attendance is **mandatory** to receive a completion grade.

4. **Code.** All code should be written in R and RMarkdown. Students should follow one of two popular style guidelines: (1) [Googles R Style Guide](#) or (2) [Hadley Wickhams Advanced R Style Guide](#). Students should specify what style guide they are using at the top of their submitted code and assignment. If a students 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 (Thursday, June 15).

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 will occur on June 22nd. More details will be available after mid-session.
7. **Interactive Graphics Project.** There will be a final project. Groups of students will be assigned a dataset to analyze. A public group presentation of the analysis will occur during our regularly scheduled final exam period. More details will be available near the end of the semester.
8. **Cheating and plagiarism.** You may work together on homework, but all R code and written solutions must be your own. CMUs Academic Integrity Policy can be found at cmu.edu/academic-integrity. If you cheat or plagiarize on any assignment, quiz, or exam, you will receive a grade of zero; for quizzes and exams, I will also notify the Office of Student Affairs and the dean of your college. A second offense will result in university disciplinary action. *Using solutions or answer keys from previous versions of this course, or from any website, or statistics major friends you might find and presenting it as your own is cheating and will be treated as such.*

GRADING POLICIES

1. **Grades.** The usual grading scheme will be used (**A:** 90 - 100, **B:** 80 - 90, **C:** 70 - 80, **D:** 60 - 70, **F:** Below 60). Final letter grades may be adjusted at the discretion of the instructor.

Average Homework Score	35.0%
Oral Evaluations (completion only)	5.0%
Lab Component	10.0%
Lab Exam Score	12.5%
Static Graphics Project/Presentation	12.5%
Interactive Graphics Project/Presentation	25.0%

Course grades will appear on the Canvas site. Each student is responsible for verifying his or her recorded scores on an ongoing basis.

2. **Regrades.** Although we strive for consistency and accuracy in grading, we understand that grading errors can occur.
 - We will gladly correct all errors in tabulation or overlooked material.
 - All regrading requests must be accompanied by a written statement carefully highlighting and explaining the items that were misgraded. Note that regrading requests can end in a positive, negative, or no change in points.
 - Regrade requests should be submitted to the instructor within one week of when the assignment or exam is returned.

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 computers in the computer cluster. Students are 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.

POLICIES

1. **Lectures.** 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 not allowed during lecture. You should bring your computer to Lab.*
2. **Email.** Sending email to your 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. Email questions should be sent a reasonable amount of time before a deadline. Students should not assume their emails will be answered immediately. Allow 24 hours for a response. Questions to which the answer is *Its* in the syllabus may not be answered.

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

Emails to the TAs will not be answered.

3. **Discussion Board: Piazza.** All questions about the homework and the notes should be directed to piazza discussion board that can be found at the bottom of Canvas or piazza.com/cmu/summer2017/36315. Homework-related email to the instructor or TAs 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.
4. **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.
5. **Disability services.** If a student believes that he/she has a disability and needs special accommodations, he/she should contact the instructor and the Disability Resources office: 412-268-2013, access@andrew.cmu.edu or cmu.edu/hr/eos/disability .
6. **Take care of yourself.** Diminished mental health, including significant stress, mood changes, excessive worry, or problems with eating and/or sleeping can interfere with optimal academic performance. The source of symptoms might be strictly related to your course work; if so, please speak with me. However, problems with relationships, family worries, loss, or a personal struggle or crisis can also contribute to decreased academic performance. CMU provides mental health services to support the academic success of students. Counseling and Psychological Services (CaPS) offers free, confidential services to help you manage personal challenges. Getting help is a smart and courageous thing to do - for yourself and for those who care about you. You can learn more about confidential mental health services available on campus at: cmu.edu/counseling/. Support is always available (24/7) from Counseling and Psychological Services: 412-268-2922, and Re:solve Crisis Network: 888-796-8226.

TENTATIVE SCHEDULE

Day	Date	Topic	Homework Due
Monday	May 22	Basics and Critique	
Tuesday	May 23	Lab1: R and Data Structure	
Wednesday	May 24	Grammar of Graphics / 1D Categorical	HW 1
Thursday	May 25	Lab2: 1D Categorical	
Friday	May 26	1D Categorical / 2D Categorical	
Monday	May 29	Memorial Day, No Class	HW 2
Tuesday	May 30	Lab3: 2D Categorical / Mosaics	
Wednesday	May 31	1D Continuous	HW 3
Thursday	June 1	Lab4: 1D Continuous	
Friday	June 2	2D/3D Continuous	
Monday	June 5	High D	HW 4
Tuesday	June 6	Lab5: Continuous and High D	
Wednesday	June 7	Maps	HW 5
Thursday	June 8	Lab6: Maps	
Friday	June 9	Regression & P-values	
Monday	June 12	Networks	HW 6
Tuesday	June 13	Lab7: Networks	
Wednesday	June 14	Color and Review	Hw 7
Thursday	June 15	Lab8: Lab Exam	
Friday	June 16	Text Analysis & Static Presentation Prep	HW 8 assigned
Monday	June 19	Time Series	HW 8
Tuesday	June 20	Lab9: Text and Time series	
Wednesday	June 21	Cleveland and Review	HW 9
Thursday	June 22	Lab10: Static Presentations	
Friday	June 23	Interactive Graphics Prep*	HW 10 assigned
Monday	June 26	TBD: Review and Catch Up	HW 10
Tuesday	June 27	Lab11: Interactive Graphics	
Wednesday	June 28	TBD: Group Meetings with Ben	
Thursday	June 29	Lab12:	
Friday	June 30	Interactive Presentations	

* Location: TBD (i.e. may be in a lab)