36-315 Statistical Graphics and Visualization
Spring 2014

Instructor: Rebecca Nugent
Baker Hall 232C
rnugent@stat.cmu.edu
http://www.stat.cmu.edu/∼rnugent
Office Hours: Mon 3-4pm

Teaching Assistants:
Mingyu Tang, Cong Lu
mingyut@stat.cmu.edu, congl@stat.cmu.edu
Office Hours: Tue 3-4pm, Mon 10am-11am; both in Wean 8110

Class Meetings: Mondays, Wednesdays 12:30-1:20pm, Porter 125C

Lab Meetings: Fridays, 12:30-1:20pm, Baker 140E/F

Website: http://www.cmu.edu/blackboard

Prerequisites: any of 36-202, 36-208, 36-226, 36-303, 36-309, 36-625, 88-250

Textbooks:

- *Graphics for Statistics and Data Analysis with R*. Kevin J. Keen. CRC Press, 2010.
- *Interactive and Dynamic Graphics for Data Analysis*. Cook & Swayne. Springer, 2007. (Recommended)

General Course Plan: Graphical displays of quantitative information take on many forms to 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. The class will also cover some principles of visual perception and estimation. We will start with univariate and bivariate data, looking at some commonly used graphs and, after discussing their advantages/disadvantages, then turning to more sophisticated tools. We will then explore some three-dimensional tools, group structure/clustering, and projections of higher dimensional data. As time permits, the course will consider some more advanced graphical models such as statistical maps, networks, and the usage of icons. Each student will required to engage in a project 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.

Course Objectives:

1. Demonstrate how to use different graphical displays to visualize a data set and its characteristics
2. Use principles of visual perception and estimation to generate effective graphical displays
3. Develop written and verbal communication skills for discussing the information presented in different graphs and their appropriateness; present graphs and visualization appropriately in a poster session.
4. Effectively use R, a widely-used statistical package, to generate graphs/displays to visualize data

Course Work: Your grade in this course will be determined by homework assignments, lab assignments, two lab exams, and a final graphics visualization project.

- Weekly homework assignments will be due at the beginning of class (12:30pm) on Wednesdays. Assignments should be submitted electronically to the Blackboard Digital Dropbox and in class (B-W is fine for paper version). Convert your HW to a pdf before submitting on Blackboard. Late homeworks are not accepted (exceptions may be made depending on circumstances; instructor permission required in advance). Note that the HW deadline is the beginning of class. There is a grace period of 10 minutes to account for printer mishaps, etc.

  Homework Format: name on front page; questions answered in order; all answers marked and labeled; *just circling answers on R output or attaching graphs with no explanation is not acceptable; answers should be written up in context of problem.* Graphs should be as close to the corresponding problem as possible. Deviating from this format may result in loss of points.

  If you do not staple (or paper clip) your homework together, we will take off points.

- The labs are designed to give you practice working with data/class material. They are mandatory; attendance is taken and part of the lab grade. You will receive the lab during class and will turn in your work via Blackboard. We are aware that not everyone has the same R programming skills; if you do not finish your lab by the end of class, you will have until 5pm that day to finish it.

- The two lab exams will be a combination of a take-home exam and an in-class exam. You will be given a data set and a set of scenarios a week in advance. You have the entire week to explore the data and determine appropriate graphical visualization techniques to describe the data and answer the questions. You then have one lab period to generate your graphs for randomly chosen questions and describe them and justify their use. More details will follow.

- The final project will be group-based and will require constructing graphs to visualize a data set and answer questions about that data set. The project will be presented in a poster to the class and other statisticians. You will need to be able to describe your graphs, their construction, and, most importantly, what information they give and why they were chosen. More details will follow.

Grading policy: You are encouraged to discuss homework problems and lab assignments with your fellow students, however the work you submit must be your own. Acknowledge any help received on your assignments or labs. Copied work will receive no credit. Late assignments will not be accepted. Your lowest homework grade and your lowest lab grade will be dropped (group project labs are included). **Please come talk to me if there are difficulties; problems/conflicts must be discussed IN ADVANCE.** Cheating/copying on exams results in a zero for the exam and a letter to your dean. Do your own work. Final grades will be computed with the following weights:

Homeworks	.35
Labs	.15
Lab Exam 1	.15
Lab Exam 2	.15
Final Project	.20

You have one week from the day an assignment, exam, etc is handed back in class to bring any grading issues, comments, complaints, etc to the attention of the instructor. Please note that if you are absent the day something is handed back, you will not receive an extension unless arrangements have been made in advance with the instructor.

Final letter grades will be determined as usual: [90,100] = A, [80,89] = B, [70,79] = C, [60,69] = D, [< 60] = R. Grades may be curved at the instructor's discretion (but it is unlikely for this class).

Computing: The statistical computing package we will use in this course is R. R is available on many campus computers, and you may download a free version from `www.r-project.org`. You may also use the nearly-identical (but not free) program called S+, available on all campus computers. You can obtain a free temporary version from `myandrew`. This version is good for 1 year; you can keep renewing the license as long as you are a CMU student.

R References: manuals available on R website;
`http://www.stat.cmu.edu/~rnugent/teaching/introR`
Introductory Statistics with R, Peter Dalgaard; Springer-Verlag
Modern Applied Statistics with S-Plus Venables, Ripley; Springer

Laptop Policy: Students are expected to be participating in class; any laptop use during class should pertain directly to the class. Instructor reserves the right to not allow laptop use during class. When the class has a guest speaker, laptops must be turned off and put away.

Cellphones/Pagers, etc: All cellphones, pagers, and anything else that makes noise should either be turned off or silenced during class. Texting is not allowed nor is it acceptable professional behavior.

Communication: Assignments and class information will be posted on Blackboard and class website. Help with using Blackboard is available at www.cmu.edu/blackboard/help/.

Email: Sending email to your professor or teaching assistants 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. Student should not assume that their emails will be answered right away.

Academic Integrity: All students are expected to comply with the CMU policy on academic integrity. This policy is online at www.studentaffairs.cmu.edu/acad_integ/acad_int.html

Cheating, copying, etc will not be tolerated; please ask if you unsure of whether or not your actions are complying with assignment/exam instructions. Always ask if you are unsure; always default to acknowledging any help received.

Video/Audiotaping: No student may record or tape any classroom activity without the express written consent of the professor. If a student believes that he/she is disabled and needs to record or tape classroom activities, he/she should contact the Office of Equal Opportunity Services, Disability Resources to request an appropriate accommodation.

Disability Services: If you have a disability and need special accomodations in this class, please contact the instructor. You may also want to contact the Disability Resources office at 8-2013.

TENTATIVE SCHEDULE: *subject to change*

Date	**Topic**	**Due**
Mon 1/13	Introduction; Examples; Types of Data	
Wed 1/15	Graph Principles; Bad Graphs	
Fri 1/17	Lab 1: Intro to R	Intro Survey, Lab 1 Work
Mon 1/20	*No class; MLK Day*	
Wed 1/22	1-D Categorical: Bar, Spine, Pie, Rose; Comparing Distributions	HW 1
Fri 1/24	Lab 2: 1-D Categorical	Lab 2 Work
Mon 1/27	2-D Categorical: Mosaic, Double Decker, Association	
Wed 1/29	2-D Categorical: Agreement, Odds Ratios, Four Fold	HW 2
Fri 1/31	Lab 3: 2-D Categorical	Lab 3 Work
Mon 2/3	1-D Continuous: Dotplots, Stripcharts, Jitter	
Wed 2/5	1-D Continuous: Conditional Distr, Boxplots, Box-Percentile	HW 3
Fri 2/7	Lab 4: 1-D Continuous	Lab 4 Work
Mon 2/10	1-D Continuous: Histogram, ASH, Rug	
Wed 2/12	1-D Continuous: Density, KDE	HW 4
Fri 2/14	Lab 5: 1-D Continuous	Lab 5 Work
Mon 2/17	Comparing 1-D Cont: Violin, Bean, CDplots	
Wed 2/19	2-D Continuous: Scatter, Trends, Jitter, Sunflower	HW 5
Fri 2/21	Lab 6: Comparing 1-D Cont; 2-D Continuous ; **Lab Exam 1 out**	Lab 6 Work
Mon 2/24	2-D Continuous: KDE, Graphically Representing Missingness	
Wed 2/26	Review	
Fri 2/28	Exam 1	**Lab Exam 1**
Mon 3/3	Higher-Dim: Scatterplot, Ternary, Conditioning Plots	
Wed 3/5	Lab 7: 2-D Densities; Missingness; Higher Dim	HW 6/Lab 7
Fri 3/7	*No class; Mid-semester break*	
	Mon 3/10 - Fri 3/14: Spring Break	
Mon 3/17	Longitudinal Data; Time Series: Trends	
Wed 3/19	Time Series: Oscillations, Differences	HW 7
Fri 3/21	Lab 8: Longitudinal Data; Time Series	Lab 8 Work
Mon 3/24	Clustering	
Wed 3/26	Clustering	HW 8
Fri 3/28	Lab 9: Clustering	Lab 9 Work
Mon 3/31	Networks	
Wed 4/2	Networks	HW 9
Fri 4/4	Lab 10: Networks	Lab 10 Work
Mon 4/7	Icons	
Wed 4/9	Lab 11: Icons, Extras	HW 10/Lab 11 Work
Fri 4/11	*No class; Carnival*; **Lab Exam 2 out**	
Mon 4/14	Dynamic EDA	
Wed 4/16	Review	
Fri 4/18	Exam 2	**Lab Exam 2**
Mon 4/21	Special Topics	
Wed 4/23	Lab 12: Final Project Work	Lab 12
Fri 4/25	Lab 13: Final Project Work	HW 11/ Lab 13
Mon 4/28	Special Topics; Final Project	
Wed 4/30	Exam Prep; **Oral Exams** (evening)	
Fri 5/2	Public Presentation	
	Final Project Reports due Friday 5/9 5pm	