

**36-315**  
**Statistical Graphics and Visualization**  
**Spring 2016**

**Instructor:** Sam Ventura ([sventura@stat.cmu.edu](mailto:sventura@stat.cmu.edu))  
Office: Baker Hall 232A  
Office Hours: Mondays, 2pm – 3pm

**Teaching Assistants:** Peter Elliot  
Office Hours: Mondays, 4pm – 5pm, Baker Hall 132A

Kayla Frisoli  
Office Hours: Tuesdays, 1:30pm – 2:30pm, Porter Hall 117

Lauren Miller  
Office Hours: Tuesdays, 4:30pm – 5:30pm, Porter Hall 117

**Textbooks:** None of these are required.  
*ggplot2: Elegant Graphics for Data Analysis*, by Hadley Wickham  
*Graphics for Statistics and Data Analysis with R*, by Keen. CRC Press, 2010.  
*Interactive and Dynamic Graphics for Data Analysis*, by Coke and Swayne. Springer,  
*Introductory Statistics with R*, by Peter Dalgaard. Springer, 2002/2004.  
*Using R for Introductory Statistics*, by John Verzani, Chapman & Hall, 2004.  
*R Graphics*, by Paul Murrell. Chapman & Hall, 2006.  
*Visualizing Data*, by William S. Cleveland. Hobart Press, 1993.  
*Envisioning Information*, by Edward R Tufte. Graphics Press, 1990.  
*The Visual Display of Quantitative Information*,  
by Edward R Tufte. Graphics Press, 1983.

**Lectures** Monday and Wednesday, 12:30pm – 1:20pm, DH 1212  
**Labs** Friday, 12:30pm – 1:20pm, BH 140 Wing

**Weekly assignments:** Homeworks are due on Wednesday at noon, unless otherwise instructed.  
Labs are due on Friday at 6:30pm, unless otherwise instructed.  
Short oral presentations will be conducted during lab (more details below).

**Lab Exams:** There will be two lab exams throughout the semester (more details below)

**Final Project:** There is no final exam. There will be a final course group project,  
which will culminate in an oral presentation (during the final weeks of class)  
and a paper (due during finals week).

**Prerequisites:** 36202 or 36208 or 36226 or 88250 or 36309 or 36625 or 70208 or 36303 or 36225.

## **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 computer-generated animations. Each student will be 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. 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, GitHub, and other tools to promote reproducible research and allow others to build from your work

### **2. Creating 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

### **3. Writing About Statistical Graphics.**

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

### **4. Speaking 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

### **5. Critiquing 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' data visualizations

## 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.

Lecture attendance is not mandatory, but if you miss class, you are responsible for the material covered during the lectures you have not attended. Students are expected to take notes and follow along with example problems in class. Some course notes and example code will be posted on the course website.

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 are designed to take 45 minutes to complete. The instructor and TAs are in lab to help you, so please feel free to ask questions when you need assistance. Additionally, please feel free to talk 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 lab exams.

While students are permitted (and encouraged) to ask questions during lab, questions about the lab assignment will not be answered after the lab session has ended. Any emails or discussion board questions about lab assignments will not be answered whatsoever.

Lab attendance is mandatory. It is also required that you attend your own lab section and are in the correct room.

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.

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 a single .html file, unless otherwise specified.

3. **Oral Evaluations.** During lab, a random group of students will be asked to give a short, one-on-one oral presentation of a single graphic to the instructor. The graphic will either be provided to all students before lab or be one of the graphics produced by students in lab.

These are for the benefit of the students and will be graded on completion only. If you are not in (the correct) lab when you are chosen to give a presentation, you will lose credit. Students are expected to take notes on the feedback that the instructor provides during and after these presentations in order to improve their data visualization presentation skills.

4. **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 homework and/or lab exams.

Homeworks are due on Wednesdays at noon, submitted through the course website. Students should submit a single .Rmd file and a single .html file, unless otherwise specified.

5. **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.

6. **Lab Exams.** There are two lab exams throughout the semester. Specific details about the content and format of the lab exams will be available closer to the exam dates. The exams will be on Wednesday, 3/2 (right before Spring Break) and Wednesday, 4/13 (right before Carnival). Please plan accordingly.
7. **Final 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. A public group presentation of the analysis will occur during the last two weeks of classes. More details will be available in April.

### 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:

Final Project and Paper	25%
Oral Evaluations (completion only)	10%
Average Lab Score	15%
Average Lab Exam Score	25%
Average Homework Score	25%

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

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

- No assignments are dropped when calculating mid-semester grades.
- The lowest lab score and homework score are not included in the calculation of final grades.

### 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 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.

## Administrative Procedures and Logistics

- **Lectures.** Use common courtesies: 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 is allowed only for course-related purposes.
- **Class web page.** The syllabus, lab assignments, homework assignments, solutions, assigned readings, any supplementary material, grades, and announcements (in fact, pretty much everything) for this course can be found on the course web page on Blackboard: [www.cmu.edu/blackboard](http://www.cmu.edu/blackboard). **Please check Blackboard regularly.**
- **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 TA's during office hours.

**Questions concerning the current 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 students name and Andrew ID on front page. 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. university athletic event, job interview) or **extreme circumstances** (e.g. hospitalization) 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 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 232 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.** Cheating, plagiarism and unauthorized assistance on homework or exams will be dealt with in accordance with the academic integrity policy in place at Carnegie Mellon University, as described here: <http://www.cmu.edu/academic-integrity/index.html>
- **Disability Services.** If you have a disability and need special 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 8-2013, 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: <http://www.cmu.edu/hr/eos/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 questions will not be answered.** Please direct these questions to the course discussion board (see below).

**Emails to the TAs will not be answered.**

- **Discussion Board:** All questions about the homework and the notes should be directed to Blackboard's discussion board. 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.

Students are expected to subscribe to all discussion board 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.