

**36-315**  
**Statistical Graphics and Visualization**  
**Summer 2019**

<b>Instructor:</b>	Robin Dunn ( <a href="mailto:dunnr@cmu.edu">dunnr@cmu.edu</a> ) Office Hours TBD
<b>Teaching assistant:</b>	Beomjo Park Office Hours TBD
<b>Calendar:</b>	<a href="#">Google calendar link</a> (not required to use, just for your reference)
<b>Lectures:</b>	Monday, Wednesday, Friday, 12:00pm–1:20pm, PH A18B
<b>Labs:</b>	Tuesday, Thursday, 12:00pm–1:20pm, BH 140E; bring laptop
<b>Weekly assignments:</b>	Homeworks are due on Monday at 10:00pm (except homework 1) Labs are due on Tuesday and Thursday at 10:00pm (except lab 1)
<b>Midterms:</b>	The two midterms are <b>Tuesday, June 4th</b> and <b>Tuesday, June 18th</b> . There will be <b>no makeup exams</b> .
<b>Oral graphic critique:</b>	June 11th 12:00-1:20 pm
<b>Final Presentations:</b>	<b>Thursday, June 27th, 12-3pm</b> and <b>Friday, June 28th, 12-3pm</b> . Please email me if you cannot attend from 1:30-3pm.
<b>Final Report:</b>	Due <b>Friday, June 28th</b> at 10:00pm
<b>Prerequisites:</b>	36202, 36208, 36226, 88250, 36217, 36218, 70208, 36303, 36309, 36225, 36625, or 21325.

### Course Description

Graphical displays of quantitative information take on many forms as they help us to 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. They will learn to distinguish when and where particular graphics are appropriate. They will also learn how to critique and speak about graphics in a concise and statistically sound manner. Each student will be required to engage in midterms and projects using graphical methods to understand data. In addition to 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. **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

### 2. **Demonstrate an understanding of 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

### 3. **Think critically about statistical graphics/visualization and data analyses.**

- Complete objective analyses of data to form thoughtful and meaningful conclusions
- Understand the extent to which you should draw conclusions / make claims about graphics and analyses

### 4. **Write about statistical graphics.**

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

### 5. **Speak about statistical graphics and data analyses.**

- Describe stand-alone graphics using clear and concise language
- Eloquently discuss a complete data analysis, from introduction to conclusion, for non-technical and technical audiences

### 6. **Critique and reproduce 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
- Recreate (and improve upon) existing statistical graphics

- 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 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. 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 most lectures, we will also assign short small-group or individual 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, but they are monitored for your class participation grade.

Please always write your Andrew ID on anything you hand in.

**You are responsible for any announcements or material covered in class.**

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 may include lecture material, so please come prepared to take notes.

The instructor and TA 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 midterm, however.

Piazza questions about lab assignments are not guaranteed to be answered. Emailed questions about lab assignments will not be answered.

**Lab attendance is mandatory and it is your responsibility to sign in.**

Lab assignments are due at 10:00pm on the day of lab, 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 1.)

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 midterms.

You will be graded on your ability to think critically about both the data and application at hand, code graphics properly, incorporate relevant statistical methods, and write concisely about your analyses.

Homeworks are due on Mondays at 10:00pm (except homework 1, which is due on Wednesday of the second week of class) and are submitted through Canvas. 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 half of a homework assignment (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. **Oral graphic critique.** On June 11th, students will concisely and accurately present a graphic to the class, for a strict 60 seconds each. Students will be graded on both their public speaking skills and their ability to interpret the graphic. Graphics will be chosen or provided ahead of time.

To prepare for the oral graph critique, 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 be provided to all students before lab. During the first lab, the instructor or TA will give an example of what is expected for these oral presentations, and no students will be required to give presentations during Lab 1.

These in-lab oral evaluations are strictly for the benefit of the students (to prep for the oral graph critique and the final presentation). Students are expected to take notes on the feedback that the instructor provides during and after these presentations in order to improve their oral presentation skills.

6. **Midterms.** There are two midterm lab exams. Specific details about the content and format of the midterms will be available closer to the exam dates (Tuesday, June 4th and Tuesday, June 18th).
7. **Final report.** Each student will choose a dataset to analyze using the graphical techniques learned in the course. They will produce a well-written, statistically sound report in R Markdown describing the work. The report is due Friday, June 28th at 10:00 pm.
8. **Final presentation.** Students will present (for ~5 minutes) the findings from their final report during a class presentation, which will be created using R Studio (as opposed to

PowerPoint, Keynote etc.). Class presentations of the analyses are scheduled for the last two days of the course (Thursday, June 27 and Friday, June 28). You will be required to attend and provide feedback on your classmates' presentations.

## Grading Policies

- Final grades will be computed according to the following weights:

Class participation assignments	5%
Oral graphic critique	5%
Homework	30%
Labs	10%
Midterm 1	10%
Midterm 2	10%
Final report	20%
Final presentation	10%

- 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.
- There are 8 labs, each worth 100 points, and 5 homework assignments, each worth 200 points. However, your final lab grades will be calculated out of 700 points, up to  $700/700$ ,  $Final\ Score = \min(Total, 700) / 700$ . And, likewise, your homework will be calculated out of 900 points,  $Final\ Score = \min(Total, 900) / 900$ . Even if you miss a few points here and there, you can earn full-credit final lab and homework scores by submitting all assignments. Alternatively, you can treat this as dropping your lowest lab score and lowest half of a homework assignment. **Late homework assignments will not be accepted.**

## 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 can 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 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.

**Once again, you are responsible for any announcements or material covered in class.**

- **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/summer2019/36315>

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 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 TA 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 or during weekends.** 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, 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 (with the exception of point addition errors, clear typos, etc.), you must write a clear, detailed description of the issue. This description should include the number of points you expect to receive. Please submit this, along with a copy of your assignment, to the instructor's email ([dunnr@cmu.edu](mailto:dunnr@cmu.edu)) **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** (by Friday of the first week) 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.

Lectures: Students are expected to participate 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.

Labs: Students are encouraged to bring their laptops to labs, if they plan to use their laptops for their assignments. Again, any laptop use should pertain directly to the class.

- **Email:** Sending email to the instructor should be treated as professional communication. You should include [36-315] in your subject header. 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 TA 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 TA will not answer any email, whatsoever.) The discussion board will be checked regularly by the instructor and TA. 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, midterms, 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 (subject to change)

**Calendar:** [Google calendar link](#) (for your reference)

Day	Date	Topic	Due dates (10 pm)
Monday	May 20	Syllabus, principles of graphics	Course survey
Tuesday	May 21	Lab 1: Introduction to R	
Wednesday	May 22	Grammar of graphics, 1D categorical and continuous data	Lab 1
Thursday	May 23	Lab 2: 1D	Lab 2
Friday	May 24	Density estimation, 2D graphs (side-by-side, stacked, facets)	
Monday	May 27	Memorial Day, No Class	
Tuesday	May 28	Lab 3: 2D graphs	Lab 3
Wednesday	May 29	2D graphs, assessing independence	HW 1
Thursday	May 30	Lab 4: 2D graphs & faceting	Lab 4
Friday	May 31	2D continuous graphs (scatterplots, contours, heat maps)	
Monday	June 3	Adding a third (or higher) dimension, exam 1 prep	HW 2
Tuesday	June 4	<b>Midterm 1:</b> due at 1:20 pm	
Wednesday	June 5	Confidence intervals, linear regression	
Thursday	June 6	Lab 5: Confidence intervals, linear regression	Lab 5
Friday	June 7	Trend lines, confidence bands, time series	
Monday	June 10	Time series, color	HW 3
Tuesday	June 11	<b>Oral graphic critique</b>	
Wednesday	June 12	Networks	
Thursday	June 13	Lab 6: Time series, networks	Lab 6
Friday	June 14	High dimensional continuous data	
Monday	June 17	High dimensional continuous data, exam 2 prep	HW 4
Tuesday	June 18	<b>Midterm 2:</b> due at 1:20 pm	
Wednesday	June 19	Maps	
Thursday	June 20	Lab 7: High dimensional continuous data, maps	Lab 7
Friday	June 21	Visualizing text data	
Monday	June 24	Interactive visualizations with Shiny	HW 5
Tuesday	June 25	Lab 8: Draft of project slides	Lab 8
Wednesday	June 26	Work on projects	Final slides
Thursday	June 27	<b>Final presentations</b>	
Friday	June 28	<b>Final presentations</b>	Final report