

36-721 Statistical Graphics and Visualization

Fall 2010 - Mini 1

Instructor: Rebecca Nugent
Baker Hall 232C
(412) 268-7830
rnugent@stat.cmu.edu
<http://www.stat.cmu.edu/~rnugent>
Office Hours: Wed 9-11am

TA: Zachary Kurtz
zkurtz@stat.cmu.edu
Office Hours: by appt

Class Meetings: Tuesdays and Thursdays 1:30-2:50pm, Wean 4709

Website: <http://www.cmu.edu/blackboard>
<http://www.stat.cmu.edu/~rnugent>

Supplemental Textbook: Not required, just useful background reading.

Handbook of Data Visualization Chen, Hardle, Unwin (editors)
Other material (journal papers, etc) will be provided and posted online.

General Course Description: This course will be a broad overview of determining and visualizing the structure in data. Basic graphical principles (e.g. Tufte) will be discussed, but the primary goal is to be able to utilize data analysis/visualization tools to explore data, summarize results, and identify structure, both expected and unexpected. Material will be pulled from books including Springer's "Handbook of Data Visualization" (Chen, Hardle, Unwin editors). This book in particular also contains several chapters on how to work with different languages/software packages. After some basics, topics will be chosen based on the research interests of the class.

Course Objectives: For each scenario, we want to be able to:

1. Know what type of structure we might be expecting
2. Know which method is appropriate to use
3. Understand the underlying theory
4. Effectively choose any needed parameters
5. Implement the method and interpret the results

We would also like to be able to adapt methods or design our own visualizations to suit our particular applications.

Course Work: Your grade in this course will be determined by occasional homework assignments (60%) and a final graphics visualization project (40%).

- Assignments should be completed and submitted on paper; electronic versions will be accepted if discussed with instructor in advance.

Please see the TA or instructor during office hours for help with homework problems. Questions posed by email must be sent at least 24 hours before the time an assignment is due in order to guarantee a response.

- The project will be started during the mini and will essentially require applying the techniques learned in class to a large database. Your results and visualization will be presented to the class and turned in as a scientific report. More details will follow.

Computing: The statistical computing package that I will be using in this course is R. However, you may use other computing packages as long as your code is well documented. 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.cmu.edu. 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, beepers, and anything else that makes noise should either be turned off or silenced during class.

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.

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 are 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 accommodations in this class, please contact the instructor. You may also want to contact the Disability Resources office at 8-2013.

SCHEDEULE *subject to change*

Date	Topic	Due
Tue 8/24	Introduction and History	Intro Survey
Thu 8/26	Histograms: Theory/Properties, Choosing Bin Width	
Tue 8/31	Average Shifted Histograms: Theory, Bin Width, Connection to Kernels	
Thu 9/2	Kernel Density Estimates: Theory, Kernels, Bandwidths, Box-Percentile Plots, Beanplots, Violins, Raindrops	
Tue 9/7	1-Dim: Sectioned Density, Density Strips; Multivariate Kernels: Product Kernels, Bandwidth Selection	
Thu 9/9	Binned Kernel Density Estimates, Visualizing as Function of Bandwidth, Modes, Mode Trees, SiZeR	HW 1 due Fri 5pm
Tue 9/14	Visualizing High Dimensional Density Estimates: Level Sets, Level Set Trees, Cluster Trees, Spanning Graphs	
Thu 9/16	Glyphs/Icons: Stars, Segments, Spiders, Faces, Thermometers, Sticks	
Tue 9/21	Projections: Multidimensional Scaling	
Thu 9/23	Projections: Random Tours, Grand Tour, Projection Pursuit, Ggobi	HW 2 due Fri 5pm
Tue 9/28	Visualizing Categorical/Contingency Tables: Bar/Spine, Pie/Rose, Association Plots, Sieve Plots, Mosaic Plots	
Thu 9/30	Variants on Mosaic Plots: Reordering, Odds Ratios, Double Decker Plots, Fluctuation Plots, Trellis Displays	
Tue 10/5	Visualizing Clustering: Dendograms, Silhouettes, Shadows, Neighborhood Graphs, Stripes, Shadow Stars	
Thu 10/7	Time Series/Functional Data: Profile Plots, Cluster Calendars, Time Wheels, MultiCombs/Spike Glyphs, Spirals	HW 3 due Fri 5pm
Wed 10/13		Group Project Presentation