

Statistics 98T
Data Visualization
Location: TBA
Time: TBA

Instructor: Amelia McNamara
Email: amelia.mcnamara@stat.ucla.edu
Website: www.stat.ucla.edu/~amelia.mcnamara
Office: MS 8105G
Office Hours: TBA

More and more, the products of statistical analysis that we encounter in our everyday lives come in the form of data visualizations. Visualizations have the advantage of being easier to interpret for many people, but they also give the impression of being a form of absolute truth. As with any presentation method, though, visualizations can be manipulated by their creators to show the story they are trying to tell. In this class, we will explore the many methods of information visualization, and develop intuition for when data graphics are not telling the truth. Throughout the class, you will be expected to respond to visualizations in the form of written response, but we will discuss the methods for analyzing graphics and the vocabulary for responding to them.

This course will draw upon methods from statistics, graphic design, geographic information systems, and computer science. However, we will approach these ideas from the ground up, and they will all be framed in the context of visualization. As a result, you can expect to come away from this course with a basic understanding of concepts from all of the aforementioned disciplines, as well as one place where they fit together.

Course Objectives

By the end of this course, you should be able to:

- Explain, in your own words, what a data visualization is representing.
- Provide an explanation of why the message portrayed in the visualization might be correct, but also:
- Identify possible sources of bias, problems with the analysis or assumptions behind the visualization.
- Identify and critique data visualizations “in the wild.”

Class policies

Learning philosophy

This is not a typical lecture-based class. You must be an active participant in class discussion, which means coming to class prepared (having read the readings and completed the reading forum response). At the beginning of the quarter, we will agree on a set of classroom norms, to which we will all adhere. For example, I like the philosophy of “step up/step down.” This norm means that if you are someone who tends to be very active in class discussions and always have something to contribute, “step down” a little to allow others the space to participate. Conversely, if you are someone who does not typically contribute to group discussions, try to “step up” a little to stretch yourself as a participant.

Academic Integrity

I do not tolerate academic dishonesty. When you provide written work to me, I expect it to be entirely your own original product, with the exception of quoted and properly cited sources. If I find work that I believe to be plagiarized, I will take the issue to the Dean of Students. As such, it's in your best interest to understand the definition and consequences of plagiarism. The Office of the Dean has a great handout on "before you begin that paper" which we will discuss at the start of class.

Communication

If you have questions for me, the best way to contact me is by emailing me at amelia.mcnamara@stat.ucla.edu or coming to my weekly office hours. If your question is something that might be of interest to other students, please feel free to post it on the CCLE discussion forum in the "questions" thread. I will monitor and respond to questions on the forum, but if you see a question that you know the answer to, please respond directly to your fellow students.

Late work

Late work will not be accepted unless you have discussed it with me beforehand, or you provide me with a doctor's note. If you know you will have trouble completing an assignment on the required schedule, please contact me as soon as possible.

Disability accommodations

I am committed to providing assistance to help you be successful in this course. Reasonable accommodations are available for students with disabilities. However, it is important that you register with the Office for Students with Disabilities as soon as possible, to arrange accommodations through their office. The OSD phone number is (310) 825-1501 (or (310) 206-6083 TTY) and they are located in the basement of Murphy Hall (A255 Murphy).

Required Materials

1. Course Reader (featuring excerpts from the books listed in the Reading List, below).
2. Data journal. This should be a small notebook, easy to carry around in your everyday life. We'll be recording places where we see data visualizations (billboards, newspapers, television commercials, textbooks) and critiquing them in the journal. It can be blank, ruled, or squared, to your preference. My choice would be the Moleskine Cahier Journal, Plain.

Assignments

Reading Forum Responses (15% of your final grade)

Each week, we will be reading and discussing texts about visualization, as well as examining exemplary visualizations (both good and bad) chosen from a variety of sources. To kick off the in-class discussion, you will submit a response to the reading or visualizations before class in a forum on the class CCLE website. Guiding questions will be posted on the forum, but will include questions like “what assumptions were made in the creation of this graphic?” “are there any obvious biases visible?” “is there anything in the visualization or reading that you don’t agree with?” (this must be supported by data) and “do you have any questions about this reading or graphic?”

Data journal entries (10% of your final grade)

In addition to the reading forum responses, you will also be responsible for keeping a data journal. In your journal, you should record interesting or notable data visualizations you encounter in your everyday life. When you find a visualization, you can sketch it or paste it into the journal, and then write a few sentences about what drew you to this particular example and how it relates to what we’ve been discussing in class. I will conduct regular checks of your data journals, but the main goal is to make you more aware of data products around you, and to spark class discussion.

Final Paper (35%)

The main product of this class will be a paper, which you will submit during finals week. This will be an exploration of a topic that has been getting statistical treatment in the media during the time of the class. For example, when I was creating this syllabus the Sochi Winter Olympics were motivating a lot of data visualizations, as was Obamacare enrollment. I will pick more recent and relevant examples during the class, and give you a list to choose from. You are also welcome to propose your own idea.

In a 15-page paper (double spaced) you will focus on two or three graphics on the same subject, each telling a different side or facet to the story. Using the readings we’ve discussed all quarter, you will provide a source-backed critique of each of the graphics. Topics to focus on include data processing, color theory, scaling, axes and biases implicit in any of the above.

Of course, this paper will depend on finding appropriate data visualizations from a variety of sources on the same topic, and your text will be analyzing and critiquing those graphics. However, the 15-page length is without graphics. Please include an appendix with reproductions of all the visualizations you comment on in your paper, labeled with figure numbers and appropriately cited with source information. In the text, you will refer to the figure number, but you will not have graphics in-line with the paper.

Paper Proposal (5%)

In the third week of class, you will turn in a 1-page topic proposal. The proposal consists of

- The subject matter you will be exploring
- At least one example of a graphic you will examine (preferably, two)
- One area you already find interesting, given the topics we’ve discussed in class so far

During the fourth week of the quarter, I will post a Doodle scheduler with my availability, and you will choose a time to come discuss your topic with me.

Literature Review (20%)

In the sixth week, you will submit a literature review for your final paper. The literature review will be in the form of an annotated bibliography of sources that you expect to cite in your final paper. For each source, provide a short description of the information it contains and the major takeaways. Then, write a few sentences on how it ties into your final paper topic.

It's acceptable for your literature review to be mainly sources from the course reader, but you will need to include at least three additional sources. For the material we discussed in class, it is acceptable to refer to your class notes for the summary.

Your literature review should be about 5 pages long, and much of the material you produce will become part of your final paper.

Peer Review Activity (5%)

In the eighth week, we will do a peer review activity. Students will be put into pairs, and will exchange first drafts of their final papers. This first draft may not be the full 15 pages, but it should be at least 8 pages for this exercise. We will read and critique each others papers, and provide written feedback. This is similar to the peer review process that academic papers go through before they are published.

Final Presentation (10%)

During finals week, you will do a short (10 minute) presentation on your final paper topic. This is a great opportunity to point out visual components to your analysis that are hard to describe in your paper, as well as gather your classmates thoughts to help you finish your paper. I expect that most people will choose to do a presentation with slides, but if you prefer to use a different format, just let me know ahead of time. To keep presentations moving along quickly (in order to accommodate everyone's presentations) I will ask you to send me your completed presentation the night before you present. I will also be keeping time.

Preliminary Schedule

Week 1	What is data visualization?	Cairo, Introduction Cairo [3], Yau, Introduction Yau [19], Huff, How to talk back to a statistic Huff [9]
Week 2	How to read data graphics	Yau, Representing Data [20], Cleveland, Principles of Graph Construction [4], Fry, Ways to Read a Visualization [7]
Week 3	How to read data graphics, cont. Paper proposal due	Yau, How to Read Histograms and Use them in R [21], Wickham, 40 Years of Boxplots [16], Friendly, A Brief History of Mosaic Displays [6]
Week 4	Maps: projections, totals versus rates Meet with me to discuss final paper topics	Monmonier, Map Generalization, and Data Maps: Making Nonsense of the Census [11], Yau, Visualizing Spatial Relationships [19]
Week 5	Perception: How color, form, and size can distort our understanding	Wilkinson, Aesthetics [18], Cleveland, Graphical Perception [4], Heer and Bostock, Crowdsourcing Graphical Perception [8]
Week 6	Chark junk: when a visualization doesn't mean anything Literature review due	WTF Visualizations, [1], Tufte, Chartjunk: Vibrations, Grids, and Ducks [14]
Week 7	Big things: mulidimensional data versus "big data"	Six Provocations for Big Data [2], Cook et al, Grand Tour [5], Cleveland, Three or More Quantitative Variables [4], Wickham, et. al [17]
Week 8	Small things: small multiples, sparklines, mini-graphs Peer review of first drafts	Tufte, Small Multiples, [13], Tufte, Sparkline Theory and Practice [15]
Week 9	More complex visualizations: Bayesian statistics, data wrangling	Ottley et al, Visually Communicating Bayesian Statistics to Laypersons, [12]. Kandel et al, Research Directions in Data Wrangling, [10]
Week 10	Final oral presentations	
Finals Week	Final paper due	

Reading List

- [1] Wtf visualizations. <http://wtfviz.net/>, 2014.
- [2] danah boyd and Kate Crawford. Six provocations for big data. *Social Science Research Network*, 2011.
- [3] Alberto Cairo. *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2013.
- [4] William S Cleveland. *The elements of graphing data*. Wadsworth Advanced Books and Software, 1985.
- [5] Dianne Cook, Andreas Buja, Javier Cabrera, and Catherine Hurley. Grand tour and projection pursuit. *Journal of computational and graphical statistics*, 4(3):155–172, 1995.
- [6] Michael Friendly. A brief history of the mosaic display. *Journal of Computational and Graphical Statistics*, 2001.
- [7] Benjamin Fry. *Organic Information Design*. PhD thesis, Massachusetts Institute of Technology, May 2000.
- [8] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. *ACM Human Factors in Computing Systems (CHI)*, pages 2013–212, 2010.
- [9] Darrell Huff. *How to lie with statistics*. W W Norton & Company, 1954.
- [10] Sean Kandel, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank van Ham, Nathalie Henry Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck, and Paolo Buono. Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, 10(4):271–288, 2011.
- [11] Mark Monmonier. *How to Lie with Maps*. University of Chicago Press, 1996.
- [12] Alvitta Ottley, Blossom Metevier, Paul K J Han, and Remco Chang. Visually communicating bayesian statistics to laypersons. Technical report, Tufts University, 2012.
- [13] Edward Tufte. *Envisioning Information*. Graphics Pr, 1990.
- [14] Edward Tufte. *The Visual Display of Quantitative Information*. Graphics Pr, 2001.
- [15] Edward Tufte. Sparkline theory and practice. http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00010R, 2014.
- [16] Hadley Wickham and Lisa Stryjewski. 40 years of boxplots. 2011.
- [17] Hadley Wickham, Dianne Cook, Heike Hofmann, and Andreas Buja. Graphical inference for infovis. *IEEE transactions on visualization and computer graphics*, 16(6), 2010.
- [18] Leland Wilkinson. *The Grammar of Graphics*. Statistics and computing. Springer Science + Business Media, 2005.
- [19] Nathan Yau. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. Wiley, 2011.

- [20] Nathan Yau. *Data Points: Visualization that Means Something*. Wiley, 2013.
- [21] Nathan Yau. How to read histograms and use them in R. <http://flowingdata.com/2014/02/27/how-to-read-histograms-and-use-them-in-r/>, 2014.