W209 Data Visualization

Instructors: Fereshteh Amini, John Guerra Gómez, Andy Reagan
Course Lead: Annette Greiner
UC Berkeley School of Information

Course Structure

Fundamentals of Visualization:

1. The Purpose of Good Visualization
2. Data Structure and Visual Marks
3. Selecting Appropriate Chart Types
4. Perception and Patterns
5. Color and Contrast
6. Interactions
7. The Visualization Design Process

Specialized Topics:

8. Exploratory Analysis
9. Visualization in Python
10. Visualizing Time and Place
11. Multidimensional Data
12. Performance
13. Untangling Networks
14. Narrative

Course Description

Visualization enhances exploratory analysis as well as efficient communication of data results. This course focuses on the design of visual representations of data in order to discover patterns, answer questions, convey findings, drive decisions, and provide persuasive evidence. The goal is to give you the practical knowledge you need to create effective tools for both exploring and explaining your data. Exercises throughout the course provide a hands-on experience using relevant programming libraries and software tools to apply research and design concepts learned. (3 units)

Objectives

After completing this course, you should be able to

- Constructively critique existing visualizations, identifying issues of integrity as well as excellence.

- Analyze data using exploratory visualization.

- Design aesthetically pleasing static and interactive visualizations with perceptually appropriate forms and encodings.

- Improve your own work through usability testing and iteration, with attention to context.

- Build commonly requested types of visualizations as well as more advanced visualizations using ground-up customization.

- Create useful, performant visualizations from real-world data sources, including large and complex datasets.

- Select appropriate tools for building visualizations, and gain skills to evaluate new tools.

Prerequisites

Master of Information and Data Science students only.
DATASCI W203, Statistics for Data Science
DATASCI W205, Fundamentals of Data Engineering (may be taken concurrently with DATASCI W209. If taken concurrently, students may not drop W205 and remain in W209.)

We recommend that students either have experience with HTML, CSS, and JavaScript or have confidence in their ability to learn new programming languages quickly. If python is the only programming language you know, you will probably benefit from learning the basics of web development with JavaScript in advance.

Course Format

The course will consist of 90 minutes of asynchronous online content ("lecture") and 90 minutes of synchronous (live) online discussion per week. The discussion session offers a chance to consider the readings and lectures with classmates and the instructor. In it, we will consider new visualizations and practice critiquing them as well as each other's work. You will also have opportunities to apply the principles from lecture to your own visual representations of data.

Assignments

There are three graded assignments as well as a final project. In addition, exercises throughout the asynchronous online content serve as practice in applying concepts learned. These exercises are credit/no credit and are required; their completion prior to the corresponding live session counts toward class participation. Note that, due to limitations in the online grading system, we cannot give participation credit for late asynchronous exercise completions. The remainder of your participation grade is based on how much you contribute in live sessions (discussions and critique) and on Slack, as well as one credit/no credit assignment (Assignment 1 Part 1). For the final project, you will work with your group to produce a web-based interactive visualization that presents findings and/or allows exploration of a dataset of your choice. As part of the final project, you will deliver a midterm presentation (10% of your course grade) plus a final presentation, website, demo video, and brief final report (together, 30% of your course grade).

- Individual graded assignments (3): 15% each
- Individual participation (participation in synchronous sessions and on Slack, submissions to online asynchronous exercises, and ungraded assignments): 15%
- Final project: 40% (10% for midterm presentation, 30% for final deliverables)

Late policy: Late assignments incur a five-point penalty for submission up to one week late and a 10-point penalty for submission up to one month late. Assignments later than one month will receive no credit.

Collaboration and independence: This course requires group work in order to complete the final project. All members of the group are expected to contribute substantially and equally to the milestone deliverables as well as the final product. You may split the work up however you like, but the division must be fair, and the final project report asks for a clear description of each group member's contributions. However, homework assignments should be completed independently, and materials turned in as homework should be the result of one's own independent work.

Grading scale

93–100	A
90–92	A–
87–89	B+
83–86	B
80–82	B–
77–79	C+
73–76	C
70–72	C–
67–69	D+
63–66	D
60–62	D–
< 60	F

Readings and Tutorials

This course will use a combination of textbooks and online readings. When both paper and electronic versions of the textbooks are available, you may use either. All students are expected to do the readings before watching lectures and before all discussion sections. In a few cases, the lecture materials will not make sense if you do not have knowledge of the examples and issues from the readings.

Required Texts

— Munzner, T. (2014). *Visualization analysis and design.* CRC Press.
— Murray, Scott (2017). *Interactive data visualization for the web, 2nd Edition.* O'Reilly.*
— DATASCI W209 course packet from Study.net.

*Available online via Safari Books, http://shop.oreilly.com/product/0636920037316.do?cmp=af-strata-books-video-product_cj_0636920037316_6754088. An older version is available as well; be sure you read the 2nd edition.

Software

Assignments and exercises use:
- D3.js http://d3js.org
- Tableau Desktop https://www.tableau.com/products/desktop (a key will be provided)

Additional libraries and tools discussed in the course include:
- Matplotlib, Bokeh, Plotly, and Seaborn (Python libraries)
- Adobe Illustrator, Highcharts (a JavaScript charting library), and others

Students are expected to select additional software tools as appropriate for their final projects.

Accommodations for Students with Disabilities

Please see your section instructor as soon as possible if you need particular accommodations, and we will work out the necessary arrangements.

Scheduling Conflicts

Please notify your section instructor in writing by the second week of the term about any known or potential extracurricular conflicts (such as religious observances or employer requirements). We will try our best to help you with making accommodations but cannot promise them in all cases. In the event there is no mutually workable solution, you may be dropped from the class.

Syllabus

Part I: Fundamentals of Visualization

1. The Purpose of Good Visualization

Why and when data visualization matters, what are relevant heuristics for "good" visualizations, the evolution of web programming for data visualization.

Required Readings:
— Munzner, T. (2014). What's vis and why do it. In *Visualization analysis and design* (1–19). CRC Press.
— Tufte, E. R. (2001). Graphical excellence. In *The Visual Display of Quantitative Information,* 2nd Edition (13, 14, 24, 29, 30, 31, 37, 38, 40, 42, and 51). Cheshire, CT: Graphics Press. (In Study.net)
— Tufte, E. R. (2001). Graphical integrity. In *The Visual Display of Quantitative Information,* 2nd Edition (53, 55, 56, 57, 60, 69, 74, 76, 77). Cheshire, CT: Graphics Press. (In Study.net)
— Minard, J. Wikimedia Commons, map of French wine exports and Napoleon's Russian campaign (image files).
— Murray, S. (2017). Chapter 3. Technology Fundamentals. In *Interactive Data Visualization for the Web*, O'Reilly.

Optional:
— Few, S. (2012). Differing roles of tables and graphs. In *Show me the numbers*. Analytics Press. (In Study.net)
— JS Introduction in W3 Schools. https://www.w3schools.com/js/js_intro.asp

2. Data Structure and Visual Form

Going from data type to graphical elements, using scalable vector graphics, binding data to web elements using D3.

Required Readings:
— Card. S., Mackinlay, J., & Shneiderman, B. Information visualization. In *Readings in information visualization: Using vision to think* (17–33). (In Study.net)
— Munzner, T. (2014). Marks and channels. In *Visualization analysis and design* (95–105, stop at 5.5.2 Discriminability). CRC Press.
— Murray, S. (2017). Chapter 4. Set up. In *Interactive data visualization for the web*. O'Reilly.
— Murray, S. (2017). Chapter 5. Data. In *Interactive data visualization for the web*. O'Reilly.
— Murray, S. (2017). Chapter 6. Drawing with data. In *Interactive data visualization for the web* (stop at Making a bar chart). O'Reilly.

Optional:
— Heer, J., & Bostock, M. (2010). [Crowdsourcing graphical perception: using mechanical turk to assess visualization design](). In *Proceedings of the SIGCHI conference on human factors in computing systems* (203–212). ACM.

3. Selecting Appropriate Chart Types

Overview of chart types and uses, building basic and more complex chart types in D3.

Assignment 1, Part 1 (Credit/ No Credit) assigned.

Required Readings:
— Munzner, T. (2014). Arrange tables. In *Visualization analysis and design* (145–175). CRC Press.
— Murray, S. (2017). Chapter 6. Drawing with data. In *Interactive data visualization for the web* (finish the chapter). O'Reilly.
— Murray, S. (2017). Chapter 7. Scales. In *Interactive data visualization for the web*. O'Reilly.
— Murray, S. (2017). Chapter 8. Axes. In *Interactive data visualization for the web*. O'Reilly.

Optional:
— Bostock, M., Ogievetsky, V., & Heer, J. (2011). [D3: Data-driven documents](). *IEEE transactions on visualization and computer graphics*, *17*(12), 2301–2309.
— Heer, J., Bostock, M., & Ogievetsky, V. (2010). [A tour through the visualization zoo](). *ACM Queue*.

4. Perception and Patterns

Mechanics of visual perception, preattentive processing, gestalt principles, perceiving multiple views, use of labels and guiding marks, scales and data transformations, occlusion and depth perception.

Assignment 1, Part 1 (Credit/ No Credit) due.
Project proposals (Credit/ No Credit) assigned.

Required Readings:
— Munzner, T. (2014). Marks and channels. In *Visualization analysis and design* (106–115). CRC Press.
— Wickham, H., Cook, D., Hofmann, H., & Buja, A. (2010). [Graphical inference for infovis]. In *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis '10), 16*(6), 973–979.
— Osmani, A. (2012). Module pattern. [In *Learning JavaScript design patterns: A JavaScript and jQuery developer's guide*]. O'Reilly Media, Inc. (Online)
— Osmani, A. (2012). Revealing module pattern. [In *Learning JavaScript design patterns: A JavaScript and jQuery developer's guide*]. O'Reilly Media, Inc. (Online)

5. Color and Contrast

Perception of color and contrast, selecting an appropriate color scheme, color in D3, techniques for highlighting.

Required Readings:
— Munzner, T. (2014). Map color and other channels. In *Visualization analysis and design* (219–235). CRC Press.
— (skim) Borland, D., & Taylor, R. M. (2007). [Rainbow color map (still) considered harmful]. In *IEEE Computer Graphics and Applications*, March/April, 14–17.

6. Interactions

Common interaction techniques, animation, navigation, overview plus detail, focus plus context, brushing and linking views, evolving interactions.

Project proposals (Credit/ No Credit) due.
Assignment 1, Part 2 (Graded) assigned.

Required Readings:
— Munzner, T. (2014). Manipulate view. In *Visualization analysis and design* (243–262). CRC Press.
— Munzner, T. (2014). Embed focus + context. In *Visualization analysis and design* (323–338). CRC Press.

— Shneiderman, B. (1996). [The eyes have it: A task by data type taxonomy for information visualizations](#). *Proceedings of the IEEE conference on visual languages*, Boulder.
— Murray, S. (2017). Chapter 9. Updates, transitions, and motion. In *Interactive data visualization for the web*. O'Reilly.
— Murray, S. (2017). Chapter 10. Interactivity. In *Interactive data visualization for the web*. O'Reilly.

Optional:
— Heer, J., & Shneiderman, B. (2012). [Interactive dynamics for visual analysis](#). *ACM Queue*, *10*(2), 30.

7. The Visualization Design Process

Iterative visualization design, design and validity, usability and evaluation methods.

Assignment 1, Part 2 (Graded) due.

Required Readings:
— Munzner, T. (2014). Analysis: Four levels of validation. In *Visualization analysis and design* (67–93). CRC Press.
— Robinson, A. C., Chen, J., Lengerich, E. J., Meyer, H. G., & MacEachren, A. M. (2005). [Combining usability techniques to design geovisualization tools for epidemiology](#). *Cartography and Geographic Information Science*, *32*(4), 243–255.
— Nielsen Norman Group. (2014). [Turn user goals into task scenarios for usability testing](#).

Optional:
— Plaisant, C. (2004). [The challenge of information visualization evaluation](#). In *Proceedings of the working conference on advanced visual interfaces*. ACM, 109–116.
— McColgin, D., Hoover, P., & Igra, M. (2016). [The DataSpace for HIV vaccine studies.](#) In *Visual Analytics Science and Technology (VAST), 2016 IEEE Conference on* (pp. 31-40). IEEE.

Part II: Specialized Topics

8. Exploratory Analysis

Charts for exploratory data analysis, designing visualization tools for exploration.

Assignment 2 (Graded) assigned.

Required Readings:
— Shelly, M. A. (1996). [Exploratory data analysis: Data visualization or torture?](#) *Infection Control & Hospital Epidemiology*, *17*(09), 605–612.

— Filliben, J. J. (2013). Exploratory Data Analysis. In *NIST/SEMATECH e-Handbook of Statistical Methods*. http://www.itl.nist.gov/div898/handbook/eda/eda.htm.
— Marx, V. (2013). Data visualization: Ambiguity as a fellow traveler. *Nature Methods, 10*(7), 613–615.

9. Visualization in Python

Matplotlib, Seaborn, Bokeh, Plotly, selecting the right library/tool.

Assignment 2 (Graded) due.

Required Readings:
— Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.
— Bokeh quick start. http://bokeh.pydata.org/en/latest/docs/user_guide/quickstart.html

Optional:
Barnard, L., & Mertik, M. (2015). Usability of visualization libraries for web browsers for use in scientific analysis. *International Journal of Computer Applications*, *121*(1).

10. Visualizing Time and Place

Geographic visualization; spatial units of analysis, choropleth maps, proportional symbol maps, isopleth maps, time series, temporal patterns.

Midterm project presentations in the live session.

Required Readings:
— Aigner, W., Miksch, S., Müller, W., et al. (2008). Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualizations and Computer Graphics, Jan/Feb*, 47–60.
— Kraak, M. J., & Ormeling, F. (2011). Map characteristics. In *Cartography: visualization of spatial data*. Guilford Press. (In Study.net)
— Murray, S. (2017). Chapter 14. Geomapping. In *Interactive data visualization for the web*, O'Reilly.

Optional:
— Butler, H. et al. (2008). The GeoJSON format specification.
https://tools.ietf.org/pdf/rfc7946.pdf

11. Multidimensional Data

Multivariate visualizations, small multiples, heat maps, parallel coordinates plots, other multivariate plot types.

Assignment 3 (Graded) assigned.

Required Readings:
— Few, S. (2009). Multivariate analysis. In *Now you see it* (281-294). Analytics Press. (In Study.net)
— Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., & Streit, M. (2013). LineUp: Visual analysis of multiattribute rankings. *IEEE Transactions on Visualization and Computer Graphics, 19*(12). also http://www.youtube.com/watch?v=iFqCBI4T8ks

12. Performance

Connecting to data, web frameworks, data reduction techniques, performance in D3.

Required Readings:
— Munzner, T. (2014). Reduce items and attributes. In *Visualization analysis and design* (299–321). CRC Press.
— Zakas, N. C. (2010). Responsive interfaces. *High performance JavaScript*. O'Reilly Media, Inc. (In Study.net)
— Shneiderman, B. (2008). Extreme visualization: Squeezing a billion records into a million pixels. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data* (pp. 3–12). ACM.
— Le Bek, P. (2014) Speeding D3.js checklist. Safari Books Online Blog.

Optional:
— Liu, Z., & Heer, J. (2014). The effects of interactive latency on exploratory visual analysis. *IEEE transactions on visualization and computer graphics*, *20*(12), 2122–2131.
— Hara, N. (2016). 6 tips to make your dashboard more performant. On Tableau.com.
— Mostak, T. (2013). An overview of MapD (massively parallel database). *White paper. Massachusetts Institute of Technology*.

13. Untangling Networks

Types of networks, reducing network complexity, network statistics, trees and hierarchies.

Assignment 3 (Graded) due.

Required Readings:
— Munzner, T. (2014). Arrange networks and trees. In *Visualization analysis and design*. CRC Press.
— Perer, A., & Shneiderman, B. (2008). Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis. In *Proceedings of the SIGCHI conference on human factors in computing systems* (265–274). ACM.

— Holten, D., & Van Wijk, J. J. (2009, June). [Force-directed edge bundling for graph visualization](). *Computer Graphics Forum, 28*(3), 983–990. Blackwell Publishing.

Optional:
— Wattenberg, M. (2006). [Visual exploration of multivariate graphs](). *CHI Proceedings*, Montreal, 811–819.

14. Narrative

Data visualization storytelling, types of narrative, information graphics, algorithmic approaches to annotation and narrative.

Final project presentations in live session, final project web site, video demo, and report due.

Required Readings:
— Segel, E., & Heer, J. (2010). [Narrative visualization: Telling stories with data](). *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139–1148.
— Cairo, A. (2013). Creating information graphics. In *The functional art: An introduction to information graphics and visualization*. New Riders, 112–122. (In Study.net).
— Hullman, J., Drucker, S., Riche, N. H., Lee, B., Fisher, D., & Adar, E. (2013). [A deeper understanding of sequence in narrative visualization](). *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2406–2415.