INFO 422/I590

Data Visualization

Yong-Yeol (YY) Ahn
yyahn@iu.edu

Fall 2017. Info West 107 (M) / 109 (W)
MW 4:00pm–5:15pm.
Office hours: W 9am-10am

Assistant Instructor
TBD (TBD@iu.edu)
Office Hours: TBD

COURSE INFORMATION AND COMMUNICATION

This syllabus and other information can be found at https://yy.github.io/dviz-course/.

Announcements and other communication will be through Canvas (https://canvas.iu.edu) and Slack (https://iu-dviz-course.slack.com).

COURSE DESCRIPTION

From TV news to cutting-edge scientific papers, from a home office to the largest companies in the world, data visualization is extensively used to reveal patterns in data and to tell stories. More and more data is collected through devices and services, and more and more decisions are made through data. Data visualization is indispensible for data analysis, and thus is an essential skill for every knowledge worker. This course is an introduction to basic statistical data analysis and visualization. We will learn fundamentals of data visualization in the context of perception, integrity, design, statistics, types of data, and visualization techniques. The hands-on exercises using Python stack and D3.js will be an integral part of the course.

Relationships with S637 Information Visualization (IVMOOC): Compared with S637, this course is more geared towards fundamental statistical visualizations and exploratory data analysis, using Python stack and D3.js. Therefore, this course may be suitable for students who pursue their careers in research, developement, data analysis, and so on.

COURSE OBJECTIVES

By the end of the course, you are expected to be able to understand, explain, and manipulate basic types of data, analyze them by applying basic exploratory visualization techniques, and create basic explanatory web-based visualizations. You will also be able to evaluate the effectiveness of data visualizations based on the principles of human perception, design, types of data, and visualization techniques.

PREREQUISITES

This course's residential version is open to advanced undergraduate students as well as graduate students; online version of this course is open to graduate students. Because creating visualizations using programming languages (Python and Javascript) is an integral part of the course, it is required to have good understanding and working knowledge of programming. The prerequisites are:

- Programming foundations (I210 & I211, or equivalent courses).

It is strongly recommended to have basic understanding of mathematics, statistics, and Web. The following courses are recommended before taking this course:

- I123: Data Fluency
- I300: HCI/Interaction Design
- I308: Information Representation
- I360: Web Design

For self-assessment, visit the following link: http://bit.ly/dviz_self. Contact the instructor if you are uncertain about your background.

EXPECTATIONS AND REQUIREMENTS

You are expected to attend every class and engage in discussions. You are not allowed to use your phone or computer during the class unless asked to do so. You are expected

to read reading materials *prior* to the class. At the beginning of most class meetings, there will be an *in-class quiz* based on the assigned readings and previous materials. Near the end of the semester, there will be a more extensive quiz based on all previous course content. You are expected to complete all weekly assignments.

The final assessment will be largely based on team projects. The choice of project topic can be guided by the instructors but you will have freedom to choose projects topics. You are required to submit a final paper that contains detailed explanation of the visualization process and results as well as the visualization artifact itself (e.g. a visualization tool or a webpage) depending on the nature of your project.

BOOKS AND KEY MATERIALS

There is no required textbook, but the following books and websites are recommended.

Python and data analysis

1. Dive Into Python by Mark Pilgrim (available online)
2. An introduction to statistics (with Python) by Thomas Haslwanter (available online): this book uses Python to explain basic statistics. It also contains a succinct tutorial for Python and data visualization using Python.
3. Learnpython.org: A web-based interactive tutorial
4. Learning IPython for Interactive Computing and Data Visualization by Cyrille Rossant: Introduction to IPython as well as lots of advanced analysis

Visualization and Design

1. The Visual Display of Quantitative Information (2nd ed.) by E.R. Tufte: one of the foundational book on visualization. It contains a rich set of historical visualization, thoughtful discussion on visualization principles.
2. Atlas of Knowledge: Anyone Can Map by K. Börner: this book systematically analyzes vocabularies of visualization with a lot of great examples.
3. Visualization Analysis and Design by T. Munzner: a nice textbook that covers important topics of visualization.
4. Visual Thinking for Design by C. Ware: one of the best books on the role of perception in visualization.

D3.js

1. Interactive Data Visualization for the Web by Scott Murray
2. D3 Gallery, D3 Tutorials, and mbostock's Blocks.

POLICIES

1. *Disabilities.* Every attempt will be made to accommodate qualified students with disabilities (e.g. mental health, learning, chronic health, physical, hearing, vision, neurological, etc.). You must have established your eligibility for support services through Disability Services for Students. Note that services are confidential, may take time to put into place, and are not retroactive. Captions and alternate media for print materials may take three or more weeks to get produced. Please contact Disability Services for Students at `http://disabilityservices.indiana.edu` or 812-855-7578 as soon as possible if accommodations are needed. The office is located on the third floor, west tower, of the Wells Library (Room W302). Walk-ins are welcome 8 AM to 5 PM, Monday through Friday. You can also locate a variety of campus resources for students and visitors who need assistance at `http://www.iu.edu/~ada/index.shtml`.

2. *No electronics—laptops, tablets, and smartphones—may be used in class*, unless the usage is specifically requested by the instructors. It has been shown that using laptops in class hurts learning, *even if* you are using it to take notes. If you must have electronics due to a special reason, please obtain permission beforehand.

3. *Be honest.* Your assignments and papers should be your own work. If you find useful resources for your assignments, share them and cite them. If your friends helped you, acknolwedge them. You should feel free to discuss both online and offline, but do not show your code directly. Any cases of academic misconduct (cheating, fabrication, plagiarism, etc) will be reported to the School and the Dean of Students, following the standard procedure. Cheating is not cool.

4. *You have the responsibility of backing up all your data and code*. Always use at least Box, Dropbox, or Google Drive. Ideally, learn version control systems and use `https://github.iu.edu` or `https://github.com`. Loss of data, code, or papers due to various reasons (e.g. malfunction of your laptop) is not an acceptable excuse for delayed or missing submission.

5. *Inform your excused absences prior to class*. Please contact the instructor until the previous day for an excused absence.
6. If you have any issues, don't hesistate to contact me or IU's Counseling and Psychological Services.

GRADING

- Attendance, Quiz, and Participation: 20%
- Assignments: 40%
- Mid-term and Final project: 40%

COURSE SCHEDULE

The schedule is subject to change. Find the most up-to-date schedule as well as detailed notes at `https://github.com/yy/dviz-course/wiki/Schedule`.

Week 1: Why do we need visualization? / Overview of visualization tools

Week 2: History and integrity / Jupyter notebook / HTML

Week 3: Labor day / Python visualization tools / CSS

Week 4: Perception and Design / Psychophysics experiments with Python / Javascript

Week 5: Data Types and Fundamental Visualizations / 1-D data / Javascript II

Week 6: Fundamental Visualizations II / Histogram and Boxplot / D3.js I

Week 7: Fundamental Visualizations III / Cumulative plots and sampling / D3.js II

Week 8: Fundamental Visualizations IV / Scale and density estimation / D3 scatter plot

Week 9: Midterm presentation / Regression / Axes and interactivity

Week 10 High dimensional data

Week 11 Geospatial data and maps I

Week 12 Geospatial data and maps II

Week 13 Texts and Graphs

Week 14 Thanksgiving

Week 15 Graphs II

Week 16 Final project presentation