Fall 2019

IS 698-001: Data Visualization and Interpretation

Aritra Dasgupta

Data Visualization & Interpretation

IS 698 Fri: 230-5pm CKB 212

Instructor: Aritra Dasgupta (aritra.dasgupta@njit.edu)
aedeegee.github.io
Office hours: GITC 5116, Friday 1pm -2pm
Appointments via email are also welcome

TA: Akm Islam (azi3@njit.edu)
Office hours: GITC 5119, Monday 1pm -2pm

Course Description:

Data is only as good as the insights derived from it. With increasing adoption of data-driven analysis and decision-making in diverse disciplines, we need to provide people with tools that help them understand data and derive key insights. Visualization is a powerful medium for augmenting human cognitive abilities and the value of data visualization is recognized widely across scientific and business domains. A Department of Energy report[i] says: *"To eliminate as many manual tasks as possible, it is imperative to develop intelligent data analysis methods coupled with interactive visualization through easy-to-use user interfaces"*. A Mckinsey report[ii] says: *"As data grows more complex, distilling it and bringing it to life through visualization is becoming critical to help make the results of data analyses digestible for decision makers. We estimate that demand for visualization grew roughly 50 percent annually from 2010 to 2015*".

To meet this real-world need and demand for a workforce trained in visualization techniques, this course will train students to design, develop, and make sense of visual representations of data. This is the first graduate course in YWCC exclusively focusing on data visualization techniques and is an elective in the M.S. in Data Science program.

Students will learn:

- **Why** and **when** visualization can be applied in the data science pipeline
- What NOT to do: **pitfalls** of visualization design, importance of **perception** and **cognition**

- **Interactive** visualization strategies: tackle **high-dimensional data** and handle **uncertainty**
- **Optimization** techniques for effective **comparison** and **interpretability** of patterns
- **Evaluation** of visualization techniques with respect to their quality and the insights
- **Implementation** using Python, JavaScript, Tableau

Grading Policy:

Group (<=3 people) project:	35%
Final exam:	30%
Midterm:	25%
Class Participation:	10%

Attendance Policy:

My expectation is that you well let me know in advance about the reason for your absence. Class participation is part of your grade, so you run the risk of impacting your grade if you miss classes without letting me know about a valid reason.

Textbooks (recommended but not required):

1. Visualization Analysis and Design (AK Peters Visualization Series), Tamara Munzner

2. Interactive Data Visualization for the Web: An Introduction to Designing with D3 2nd Edition, Scott Murray

3. The Visual Display of Quantitative Information, Edward Tufte

Web and Software resources:

Course portal: canvas.njit.edu

Tableau:
https://www.tableau.com/academic/students

D3:
https://d3js.org/
https://square.github.io/intro-to-d3/

Python libraries:
http://seaborn.pydata.org/index.html
https://dash.plot.ly/?_ga=2.11438772.1202715669.1567783402-402630880.1564242752

Syllabus:

1. Introduction
    a. How visualization affects data interpretation
    b. Role of visualization in data science
    c. Two flavors of data visualization: exploratory and communicative
    d. Course projects overview

2. Visualization design principles I
    a. Data and task abstraction
    b. Best practices for encoding
    c. Marks and channels
    d. Tableau assignment
    e. Course project details

3. Visualization design principles II
    a. Effectiveness and expressiveness
    b. How to critique visualizations
    c. Design problems and consequences
    d. How *not to* cause misinterpretation

4. Exploratory Visualization techniques
    a. Handling high-dimensional data
    b. Comparison techniques
    c. Small multiples
    d. Handling uncertainty
    e. Depicting time

5. Interactive visualization
    a. Why interactivity is needed
    b. Handling multiple views
    c. Brushing and Linking
    d. D3 instructions and assignments

6. Data Interpretability
    a. Causes of misinterpretation
    b. Role of communicative visualization in sciences

c. Graphical interpretability: metrics and approaches.

7. Visualization in data science: methods and examples
    a. Explaining machine learning models
    b. Interpretability challenges and solutions
    c. Transparency and human-machine trust
    d. Impactful case studies:
    biology, healthcare, cyber security, climate science, social science

How you can do well in the course:

In data visualization, the process is more important than the outcome, meaning that in an academic course, the focus is more on how you can reason about the need for and the effectiveness of visualization techniques, and less on whether you got the correct answer. Often, there is no single correct answer, but multiple viable solutions. As long as you are able to reason about your solution by applying the principles learnt in the class, you will be doing well in the course. This will need a combination of computational thinking as well as design thinking (putting the user first and thinking about the solution from a user's perspective). I hope you all will learn and develop these skills as part of this course, which will be the biggest take-aways and which you can apply in any real-life, data-driven problem-solving scenario.

Plagiarism and Academic Integrity

The approved "University Code on Academic Integrity" is currently in effect for all courses. Should a student fail a course due to a violation of academic integrity, they will be assigned the grade of "XF" rather than the "F" and this designation will remain permanently on their transcript.
All students are encouraged to look over the University Code on Academic Integrity and understand this document. Students are expected to uphold the integrity of this institution by reporting any violation of academic integrity to the Office of the Dean of Students. The identity of the student filing the report will be kept anonymous.
NJIT will continue to educate top tier students that are academically sound and are self-disciplined to uphold expected standards of professional integrity. Academic dishonesty will not be tolerated at this institution.

---

[i] DOE ASCR, 2013 http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/ASCR_DataCrosscutting2_8_28_13.pdf

[ii] Mckinsey, 2016
https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20analytics/our%20insights/the%20age%20of%20analytics%20competing%20in%20a%20data%20driven%20world/mgi-the-age-of-analytics-full-report.ashx