Data Visualization
DS 650 (Fridays CKB 320 1:00 pm-3:40 pm)

Aritra Dasgupta (aritra.dasgupta@njit.edu)
Assistant Professor, Data Science
aedeegee.github.io niiv.njitvis.com
Office hours: 1130am-1230pm Friday, GITC 5502

(Questions/Appointments via email are welcome)

Teaching Assistants:
Vrushali Koli (vk435@njit.edu), PhD student, Data Science
Kaustav Bhattacharjee (kb526@njit.edu)., PhD candidate, Data Science

Important Dates:

Quiz (individual, closed book/laptop, 2 pages of notes):	TBD
Midterm (individual, closed book/laptop, 2 pages of notes):	TBD
Mini-project presentations:	TBD
Final exam:	TBD

Attendance Policy: Attendance is required. Any exceptions must be discussed with me.

Course Description:
Data is only as good as the insights derived from it. With the 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*".

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

Students will learn:

- **Why** and **when** visualization should be used

- What NOT to do: **pitfalls** of visualization design, and the importance of **perception** and **cognition**
- **Visualization techniques** for understanding distributions, proportions, and associations among many variables
- **Optimization** techniques for effective **comparison** and **interpretability** of patterns
- **Evaluation** of visualization techniques with respect to their quality and the insights
- **Implementation** using Tableau or Python

Breakdown of Points:

Assignments & Attendance:	10%
Quiz:	15%
Midterm:	20%
Mini-Projects (2):	30%
Final Exam:	25%

Textbooks:

None required.

References:

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

4. Storytelling with Data: A Data Visualization Guide for Business Professionals, 2015, Cole Nussbaumer Knaflic

Web and Software resources:

Course portal: canvas.njit.edu

Tableau:

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

Python libraries:

http://seaborn.pydata.org/index.html

https://dash.plot.ly/?_ga=2.11438772.1202715669.1567783402-402630880.1564242752

Syllabus:

Introduction
- How visualization affects data interpretation
- Design consequences: examples of good and bad visualization design
- Two flavors of data visualization: exploratory and communicative

Visualization design principles I
- Data and task abstraction
- Best practices for encoding
- Marks and channels

Visualization design principles II
- Effectiveness and expressiveness
- How to critique visualizations
- Design problems and consequences

Visualization techniques
-Handling different attribute types
-Handling high-dimensional data
-Comparison techniques
-Small multiples
-Depicting time

Role of Color in Visualization
- Beyond aesthetics, role of color in data interpretation
- Color perception
- Color maps
-Design Pitfalls

Interactive visualization
- Why interactivity is needed
- Handling multiple views
- Brushing and Linking
- Dashboards

Advanced topics: applying visualization in real world.
- Explaining machine learning models
- Interpretability challenges and opportunities
- Transparency and human-machine trust
- 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 learned 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) skills. I hope you all will learn and develop these skills as part of this course, which will be the biggest takeaways 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