Data Visualization
CS 450
Mondays and Thursdays 1pm to 2:15pm in FMH 319

Instructor: Aritra Dasgupta (aritra.dasgupta@njit.edu)
Assistant Professor, Data Science
niiv.njitvis.com
Office hours: 12pm to 1pm on Mondays in GITC 5502
(Questions/Appointments via email are always welcome)

1. **Introduction**
   - Real-world use cases from public health (e.g., COVID, sciences (e..g., charts used on in science domains) and journalism (e.g., charts used for communicating real-time events)
   - Automation Vs Human-Data Interaction
   - Interpreting Basic Visualizations: why and how misinterpretation can happen

2. **Role of Visualization in the Data Science Pipeline**
   - Communication and decision-making
   - Importance of Perception and Cognition
   - Introduction to Tableau
   - **Assignment**: setting up Tableau workbook and generating a "Hello World" chart

3. **Data Abstraction and Organization**
   - Data Types
   - Attribute Types
   - Sorting, Grouping, Filtering, Aggregation
   - **Tableau Assignment**: Given a number of sample data tables, identify the items, attribute types, and data organization needed for extracting specific pieces of information (e.g., identify the student who got the highest number of points, or the course in which students got the highest average points). Implement them in Tableau.

4. **Visual Encoding**
   - Analytical Tasks
   - Design principles
   - Design pitfalls
   - Hands-on demonstrations with Tableau
   - How visualization can mislead
   - Common errors and critiquing

- Optimizing visualization and the role of computation
- Introduction to Python Seaborn library
- **Assignment** on Visualization Critiquing (detect design pitfalls and suggest alternative design)

Tableau Assignment: Generate charts in Tableau along with the implementation and reasoning behind the design optimization

5. **Visualizing Univariate, Bivariate, and Multivariate Data**
   - Histograms
   - Dot plots
   - Box plots
   - Bar charts
   - Scatter plots
   - Heatmaps
   - Slope plots
   - Small multiples
   - Parallel coordinates
   - Optimizing visual search
   - Dimensionality reduction
   - Communicating uncertainty in the data

   Assignment in Python on developing visualizations suited for understanding univariate, bivariate and multivariate distributions given a data set

6. **Comparative Visualization Techniques**
   - Role of visualization in model comparison
   - Visual comparison techniques
   - Optimizing visual comparison
   - **Assignment in Python on implementing a multivariate data visualization technique that enables quick comparison of specific patterns (e.g., correlation, similarity, outlier, etc.)**

7. **Midterm Exam**

8. **Visualizing Spatio-temporal Data**
   - Basics of map-based visualization
   - Temporal data visualization techniques -beyond line charts
   - Visualizing events

9. **Interactive Data Visualization (basic)**
   - Goals of exploratory data analysis

- Design principles for interactive visualization techniques
- Faceted search and exploration in human-data interfaces

10. **Interactive Data Visualization (advanced)**
    - Interaction modalities: direct manipulation vs indirect manipulation
    - Demonstration and examples using Tableau dashboards
    - Demonstration and examples using Dash
    - **Mini-project using real-world data sets** for developing a simple interactive user interface (using Dash) that takes user-specified parameters and then generates an interactive visualization, where both direct (e.g., select a group of data points) and indirect manipulation (e.g., filtering outliers) are enabled

11. **Color in Visualization**
    - Color Perception
    - Color Maps
    - **Assignment:** Describe how color can be used as part of the mini-project, giving examples of tasks and color maps.

12. **Visually Communicating Data Insights-1**
    - Model Explanation techniques
    - Visualizing Machine-learning model explanations

13. **Visually Communicating Data Insights-2**
    - Inferential reasoning using rules and decision-tree visualizations
    - Visualization in scientific applications

14. **Project Presentations and Critique**

15. **Final Exam**

--

Textbook:
-Visualization Analysis and Design, Tamara Munzner, 2014 (A K Peters)

Grading Policy: (attendance is strongly encouraged as interaction with me and the TAs will help you learn the subject better)

Quizzes and Class Performance: 20%
Midterm: 25%
Course Project: 25%
Final Exam: 30%

Tentative Dates for Quizzes and Midterm:

Quiz 1: 2/23

Midterm: 3/30

Quiz 2: 4/20

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

Grading:

Class Attendance: 10%

Assignments: 20%

Course Project: 20%

Midterm Exam: 20%

Final Exam: 30%

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 users first and thinking about the solution from the users' 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 self-disciplined to uphold expected standards of professional integrity. Academic dishonesty will not be tolerated at this institution.

Attendance Policy:

Attendance is required with the following exception: you can have 2 unexcused absences in total throughout the semester. Beyond that, you have to send documentation to the Dean of Students as described here: https://www.njit.edu/dos/student-excusals Failure to do so will result in a letter grade penalty for each time you fail to adhere to the rules. Essentially, this means that failing to provide documentation 4 times can result in you getting an F.