

COMP 4433—Data Visualization

Course Overview

This course explores visualization techniques and theory. The course covers how to use visualization tools to effectively present data as part of quantitative statements within a publication/report and as an interactive system. Both design principles (color, layout, scale, and psychology of vision) as well as technical visualization tools/languages will be covered.

Objectives

Students will understand and be able to apply visualization techniques for both exploratory data analysis and formal reporting. Students will write Python scripts controlling packages: numpy, pandas, matplotlib, seaborn, plotnine, and plotly.

Textbooks and Materials

Suggested textbooks for COMP 4433 Data Visualization

- Dobler, M., & Grossmann, T. (2019). *Data visualization with Python*. Packt Publishing. ISBN 978-1-78995-646-7.
- McKinney, W. *Python for data analysis* (2nd ed.). O'Reilly. ISBN 2018, 978-1-491-95766-0.

Additional suggested textbooks

- Poladi, S. (2018). *Matplotlib 3.0 cookbook*. Packt Publishing. ISBN 978-1-78913-571-8.
- Lutz, M. (2009). *Learning Python* (4th ed.). O'Reilly. ISBN 978-0-596-15806-4.

Grading

Assignment/Assessment	Points	Weight on Final Grade
Assignment 1	100	20%
Assignment 2	100	20%
Assignment 3	100	20%
Assignment 4	100	20%
Assignment 5	100	20%

Grading Scale

- A is 93-100
- A- 90-92.99
- B+ 86-89.99
- B 83-85.99
- B- 80-82.99
- C+ 76-79.99
- C 73-75.99
- C- 70-72.99
- D+ 66-69.99
- D 63-65.99
- D- 60-62.99
- F < 60

Assignment and Assessment Information

Assignment 1 (due by midnight MST the day prior to Live Session 2)

- Assignment 1, Part 1: Construct a CSV file with the first eight elements of the periodic table. Include columns for name, symbol, and atomic number. Read that into a pandas DataFrame. Inside the program, add a ninth and 10th element, and then add a column with the atomic weights rounded to the nearest integer.
- Assignment 1, Part 2: Make a list of strings for nine Greek letters, for example 'alpha'. Make that list such that the strings are not in alphabetic order. Make two 9-element numpy arrays of random floating-point numbers with an estimated mean of 10 and a standard deviation of 1.5. Make another array of nine elements ranging from zero to two times pi. Name it 'angle'. Make another array holding the cosine of that 'angle' array. Construct a dictionary from all of the above. Form a DataFrame from that dictionary, and print it out. Sort the DataFrame ascending on the Greek letters, drop two columns of your choice, drop one of the rows, and print that out.
- Assignment 1, Part 3: Write a program in Python to create and print out the first 12 Fibonacci numbers. Then iterate over the last five numbers to build another list with the ratio of each number to its predecessor. What do you observe about this latter list?
- Assignment 1, Part 4: Provide a function that converts temperature in Kelvin to Rankine. Make a list of five Kelvin temperatures, and print out their values in Rankine. Repeat using a lambda function.

Assignment 2 (due by midnight MST the day prior to Live Session 4)

- Assignment 2, Part 1: Create a list of the atomic weights of the first six elements of the periodic table, each rounded to the nearest integer. Provide two pie charts as follows: (1) each slice annotated with a percentage of the whole and (2) each slice annotated with its atomic weight. Explode a different element with each chart.
- Assignment 2, Part 2: Read into a DataFrame the file `py_ide2.csv`, and provide both a horizontal bar chart and a vertical bar chart, complete with all labels. Be sure to rotate the IDE names so that they are readable.
- Assignment 2, Part 3: Construct a list of eight strings that represent days evenly spread out. Drawing from the random uniform distribution, make an array of eight floats ranging from 100 to 200 in value. Establish a DataFrame from that list and that array, convert the dates to pandas datetime objects, and set them to the index. Make two charts in the same window or canvas as follows: (1) a line plot of the values vs. dates and (2) a bar chart of the same.
- Assignment 2, Part 4: Pull from Yahoo! Finance the closing prices and volumes of the stock of your choice over the trading days of one month, and plot the prices and volumes on a canvas in two separate panels, one above the other, with the dates aligned.

Assignment 3 (due by midnight MST the day prior to Live Session 6)

- Assignment 3, Part 1: Using the built-in Seaborn dataset `mpg`, provide a heatmap of the correlation of all the numeric columns and provide a pairplot of the same.
- Assignment 3, Part 2: Using the built-in Seaborn dataset `diamonds`, establish a FacetGrid based on 'cut' and 'color'. Eliminate colors 'D' and 'E' as well as the cut 'Fair'. Within that grid, plot the scatterplot for 'price' vs. 'carat'.
- Assignment 3, Part 3: Using the built-in Seaborn dataset `car_crashes`, prepare plots with a scattergram with the linear model for both the total vs. speeding and the total vs. alcohol.
- Assignment 4, Part 4: Using the built-in Seaborn dataset `iris`, provide a plot with four subplots wherein the distribution of each of the numeric columns is presented as a set of boxplots, one for each 'species'.

Assignment 4 (due by midnight MST the day prior to Live Session 8)

- Using free and open source tools, provide a set of choropleth visualizations for each of the columns containing dates such that the resulting visualizations (48 states only) tell the story by conveying through color, texture, or both the time lines of achievement of each milestone/column in the provided dataset. Missing data are of particular interest in that when a state has never achieved a given milestone, that should be indicated in a standout manner such as cross-hatching. Consider that the publication may be grayscale. Provide a solution for that as well. Provide the titles, labels, and legends necessary for clarification. File support is given as follows: `SturmCodebook` has the explanation. `SturmData` is the data CSV.

Assignment 5 (due by midnight MST the day prior to Live Session 10)

- Assignment 5, Part 1: Using the file `capitals_lat_lon.csv`, form a `DataFrame`, and write out an HTML file that when rendered in a browser displays a marker for each capital city. Hint: Use `folium`.
- Assignment 5, Part 2: Estimate the latitude and longitude for three points of a triangle that would just cover Africa or come close. (Hint: Rabat, Cape Town, and Mogadishu, for example.) Use `folium` to demonstrate where the points are on the map so that a person could see at a glance that they do nearly cover Africa. Use `Shapely` to define a polygon from those three points. Compute the area and perimeter of the triangle assuming flat earth and allowing each degree of latitude and each degree of longitude to be considered one unit of length.
- Assignment 5, Part 3: Estimate the boundaries for Kansas (use just four lat-lon points) and Nebraska (use just six points). Construct a GeoJSON file from that. Then write a Python program to read that file, form a dictionary, and plot the result.
- Assignment 5, Part 4: Generate at random the “happiness index” for each state in the United States. Provide a choropleth of the United States with that data represented thereon.
- Assignment 5, Part 5: Generate 500 points from the random exponential distribution. Choose your own parameters for that. Use `Plotly` to plot a histogram of that distribution such that there are at least 15 bins with counts greater than zero.

Weekly Schedule

There will be a graded assignment assigned each odd week and due the following week by midnight the day prior to the live session. The schedule also includes many asynchronous exercises in addition to the assignments. Please complete each week’s asynchronous exercises 24 hours before each live session. Please note that all readings are suggested readings.

Week 1

Readings:

- Reading 1: McKinney, Chapters 1–3
- Reading 2: Dobler, Chapter 1

Week 2

Readings:

- Reading 1: McKinney, Chapters 4–5
- Reading 2: Dobler, Chapter 2

Complete Assignment 1

Week 3

Readings:

- Reading 1: McKinney, Chapters 7–9

- Reading 2: Dobler, Chapter 3

Week 4

Readings:

- Reading 1: McKinney, Chapters 11 and 12
- Reading 2: Dobler, Chapter 4

Complete Assignment 2

Week 5

Readings:

- Reading 1: McKinney, Chapters 13–14
- Reading 2: Dobler, Chapter 5

Week 6

Readings:

- Reading 1: Online documentation for ggplot2 and plotnine

Complete Assignment 3

Week 7

Readings:

- Reading 1: Dobler, Chapter 6

Week 8

Readings:

- Reading 1: Dobler, Chapter 7
- Review all previous readings

Complete Assignment 4

Week 9

Readings:

- Reading 1: McKinney, Chapters 6 and 10
- Reading 2: Online documentation for plotly and dash

Week 10

Readings:

- Review all reading assignments

Complete Assignment 5

Attendance Policy

Attendance at all live session meetings is mandatory.

Program Mission

Our MS in Data Science provides students with a broad course of study in programming, algorithms, statistics, and data management, as well as a depth of understanding in specific fields such as data mining, machine learning, and parallel systems. Graduates of the data science program go on to work in a wide variety of careers, including business, government, education, and the natural sciences.

Honor Code and Academic Integrity

All students are expected to abide by the [University of Denver Honor Code](#). These expectations include the application of academic integrity and honesty in your class participation and assignments. Violations of these policies include, but are not limited to

- Plagiarism, including any representation of another's work or ideas as one's own in academic and educational submissions
- Cheating, including any actual or attempted use of resources not authorized by the instructor(s) for academic submissions
- Fabrication, including any falsification or creation of data, research, or resources to support academic submissions

Violations of the Honor Code may have serious consequences including, but not limited to, a zero for an assignment or exam, a failing grade in the course, and reporting of violations to the Office of Student Conduct.

Diversity, Inclusiveness, Respect

DU has a core commitment to fostering a diverse learning community that is inclusive and respectful. Our diversity is reflected by differences in race, culture, age, religion, sexual orientation, socioeconomic background, and myriad other social identities and life experiences. The goal of inclusiveness, in a diverse community, encourages and appreciates expressions of different ideas, opinions, and beliefs, so that conversations and interactions that could potentially be divisive turn instead into opportunities for intellectual and personal enrichment.

A dedication to inclusiveness requires respecting what others say, their right to say it, and the thoughtful consideration of others' communication. Both speaking up AND listening are valuable tools for furthering thoughtful, enlightening dialogue. Respecting one another's individual differences is critical in transforming a collection of diverse individuals into an inclusive, collaborative, and excellent learning community. Our core commitment shapes our core expectation for behavior inside and outside of the classroom.