Course Intro

Elements of Data Visualization (CS 329E)

Jan 12, 2026

“Visualization gives you answers to questions you didn’t know you had.”

– Ben Shneiderman

“There is no such thing as information overload. There is only bad design.”

– Edward Tufte

Class page & syllabus walkthrough:

https://www.cs.utexas.edu/~scohen/cs329e.html

Project 1 overview

Data foundations: Locate and gather one primary dataset that you'd like to study and one or more supplemental datasets. Brainstorm research questions to guide your exploration study.

Data load: Import the datasets, examine some sample data from each source dataset, inspect each dataset for missing values, and plot feature distributions.

Data preparation: apply data cleaning and wrangling techniques, and understand the trade-offs of different options, potential pitfalls, and their downstream impact on the analysis workflow.

Data visualization: Map each of your questions about the data to one or more charts, create and compare your charts, use the resulting charts to derive more insights from the data.

Data analysis: compute meaningful statistics for the data, identify the different components of the data, and evaluate the similarity of certain groups of data.

Project 1 timeline

Milestone 1: Data Foundations, 1 week
Milestone 2: Data Load, 1 week
Milestone 3: Data Preparation, 1 week
Milestone 4: Data Visualization, 2 weeks
Milestone 5: Data Analysis, 1 week

Project 2 overview

Problem definition: review the Request For Proposal (RFP) and conduct some research on the company, their leadership, products and services. Explore the provided sample dataset.

Visual narrative: Create visualizations that deep-dive into the business problem, show the correlations, and prove that X is more valuable than Y.

Stakeholder briefing: Translate your visuals into a 5-slide "Executive Story." Focus on the implications of the data, not the mechanics of the Python code.

Pitch: Present to the Executive Board. You have 5-7 minutes. If the board asks, "How does this improve our ROI?" you must have a data-backed answer ready.

Dashboard: Develop an interactive dashboard using Streamlit. This will serve as an evidence repository and allow executives to "drill down" into specific details.

Project 2 timeline

Milestone 1: Problem definition, 1 week
Milestone 2: Visual narrative, 2 weeks
Milestone 3: Stakeholder briefing, 2 weeks
Milestone 4: Pitch, 1 week
Milestone 5: Dashboard, 2 weeks

Project group and environment setup:

Project 1 Milestone 0

Data formats

Time series data: flu activity in the US, what does it look like year-over-year and by state?

Network data: compare airlines and routes to each other, looking at relationships, flows, hierarchies, communities, and spatial networks.

Geospatial data: air pollution data, what are the geographical distribution, density, clustering and auto-correlation?

Text data: study Shakespeare's works, explore the word frequency, lexical dispersion, and dependency graphs.

Image data: landmark images, what are the categories of those images and distributions? What happens when you reduce the colors? Looking at images taken of the landmark and from the landmark.

Video data: explore national park videos, what are their source, size, compression, and aspect ratios?

**We will study each of these six formats over the coming weeks.*