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

Plan the data exploration: find a dataset you'd like to study and brainstorm some specific questions you have about the data. Supplement your primary dataset with 1-3 additional datasets that will help you answer your questions.

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

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

Visualize data: 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.

Analyze data: 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: Proposal, 1 week
Milestone 2: Load data, 1 week
Milestone 3: Wrangle data, 2 weeks
Milestone 4: Visualize data, 2 weeks
Milestone 5: Analyze data, 1 weeks

Project 2 overview

Plan the data exploration: review the Request For Proposal (RFP) and conduct some basic research on the company, their products and services. Explore the provided sample dataset.

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

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

Pitch to the board: 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.

Build 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: Presentation deck, 2 weeks
Milestone 4: The Pitch, 1 week
Milestone 5: Dashboard, 2 weeks

Project group and environment setup:

Milestone 0

Data types

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 data types over 12 weeks. This will help inform your work for Project 1.*