College of Science · Computer Science

Introduction to Data VisualizationCS 133

- Fall 2026
- Section 02
- In Person
- 3 Unit(s)
- 08/19/2026 to 12/07/2026
- Modified 08/13/2026

Contact Information

Instructor(s):	Frank Luo
Email:	zhiqiang.luo@sjsu.edu
Class Days/Time:	TuTh 3:00PM - 4:15PM
Classroom:	Boccardo Business Center 125
Office Hours:	Tu/Th 4:15 – 5:15pm at DH282
TA:	

Prerequisites:	Prerequisite(s): CS 146 with a grade of "C-" or better, or CS 22B and graduate standing. Computer Science orSoftware Engineering majors only.
Class Meeting Dates:	Aug 19, 2026 – Dec 7, 2026

Course Description and Requisites

Topics in data analysis and visualization. Covers tools and techniques to efficiently analyze and visualize large volumes of data in meaningful ways to help solve complex problems in fields such as life sciences, business, and social sciences.

Prerequisite(s): CS 146 with a grade of "C-" or better, or CS 22B and graduate standing. Computer Science or Software Engineering majors only.

Letter Graded

Classroom Protocols

Classroom Learning Environment:

This course will follow a hands-on learning approach where we will work through coding exercises together in class. Please come to class with a charged laptop ready to dive into some code!

Code of conduct:
Short version: In this course, I aim to foster a positive learning environment - no form of harassment will be tolerated, including verbal comments and images that exclude people based on gender, socio-economic status, or appearance.

The full code of conduct is provided on the Canvas course space for this course.

Plagiarism and cheating

Just like a written essay, using somebody's computer code without proper acknowledgement is considered plagiarism. Homework problems should be based entirely on students' own work. Students can (and are encouraged to) discuss general coding techniques and problem solving strategies for homework problems but this should never include copying (whether by typing, file transfer or cutting and pasting), looking at somebody else's code on their computer to get help, or allowing copying to occur. Students found violating this policy once will receive zero credit for those problems. Continued violations will result in disciplinary action. If you have any questions about this policy, please don't hesitate to ask for clarification.

Program Information

Diversity Statement - At SJSU, it is important to create a safe learning environment where we can explore, learn, and grow together. We strive to build a diverse, equitable, inclusive culture that values, encourages, and supports students from all backgrounds and experiences.

Course Learning Outcomes (CLOs)

Course Learning Outcomes (CLO)

Upon successful completion of this course, students will be able to:

- CLO 1 Manipulate large datasets and handle missing or inconsistent values in datasets.
- CLO 2 Perform statistical analysis using packages such as Numpy and Scipy
- CLO 3 Analyze and visualize datasets using packages such as seaborn and matplotlib
- CLO 4 Develop interactive visualization using packages such as Plotly
- CLO 5 Recognize and reduce data and spatial biases

Course Materials

Required Materials

There is **no required textbook**. Course content will be delivered through instructor-provided Google Colab notebooks, lecture notes, and publicly available Python documentation.

Recommended References

1. McKinney, W. (2022). *Python for Data Analysis* (3rd ed.). O'Reilly Media.
2. Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow* (3rd ed.). O'Reilly Media.
3. Wilke, C. O. (2019). *Fundamentals of Data Visualization*. Available online.

Grading Information

Category	Percent of Total Grade
Quizzes/Exercise	20
Homework	20
Midterm1	20
Midterm2	20

Project	20

Wk	Date	CLO	Topic	Notes
W1 (Th)	Aug 20	—	Course introduction: the modern data-analysis pipeline	Syllabus and course expectations. Assign take-home Python/Jupyter/Colab/GitHub 25.
W2 (T)	Aug 25	CLO 1	Pandas I — Foundations & Data Access	Series/DataFrame, loading data, indexing, selecting, filtering, sorting, calculated col
W2 (Th)	Aug 27	CLO 1	Pandas II — Integration, Reshaping & Scale	Merge/join, groupby, long vs. wide data, dtypes, memory, vectorization, chunked re
W3 (T)	Sep 1	CLO 1	Pandas III — Data Quality & Cleaning	Missing-data detection; MCAR/MAR/MNAR; imputation; duplicates; inconsistent outliers; validation.
W3 (Th)	Sep 3	CLO 1	Applied Data Wrangling & EDA Lab	Milestone 0 due: project dataset selected, source documented, initial schema/data di reproducible load, and preliminary quality audit. Lab integrates raw → analysis-read
W4 (T)	Sep 8	CLO 2	NumPy for Data Analysis	Arrays, vectorization, broadcasting, mathematical operations, and relationship to pa
W4 (Th)	Sep 10	CLO 2	Descriptive Statistics & Probability Distributions	Summary statistics, variability, probability distributions, sampling concepts; NumPy
W5 (T)	Sep 15	CLO 2	Statistical Inference I	Sampling distributions, confidence intervals, one-/two-sample t-tests, assumptions,
W5 (Th)	Sep 17	CLO 2	Statistical Inference II	Chi-square, ANOVA, post-hoc thinking, practical significance, and effect size.
W6 (T)	Sep 22	CLO 2	Correlation & Regression	Correlation, simple linear regression, residual thinking, uncertainty, and interpretati
W6 (Th)	Sep 24	CLO 2	Non-parametric Methods & Applied Statistics Lab	Milestone 1 due: cleaned analysis dataset + documented cleaning decisions + descri + at least one justified inferential analysis.
W7 (T)	Sep 29	CLO 1–2	Integrated Data Analysis Case Study + Exam Review	End-to-end pandas → statistics case study; interpretation and exam preparation.
W7 (Th)	Oct 1	CLO 1–2	Exam #1	Covers CLO 1 & 2: data preparation, NumPy/SciPy, descriptive and inferential anal
			Project Overview	Introduce the project scope, expectations, milestones, deliverables, and connection t data-analysis workflow.

Wk	Date	CLO	Topic	Notes
W8 (T)	Oct 6	CLO 3	Visualization Foundations + Matplotlib	Visual encodings, choosing chart types, figure/axes model, subplots, styling, annota accessibility.
W8 (Th)	Oct 8	CLO 3	Seaborn: Distributions, Categories & Relationships	Hist/KDE, box/violin, categorical comparisons, scatterplots, regression plots.
W9 (T)	Oct 13	CLO 3	Multivariate Visualization	Hue/size/style, facets, pair plots, multiple variables, comparison across subgroups.
W9 (Th)	Oct 15	CLO 3	Large & High-Dimensional Data Visualization	Heatmaps, correlation matrices, aggregation, sampling, density, overplotting, large-
W10 (T)	Oct 20	CLO 3	Data Storytelling & Visualization Design	Composition, annotation, hierarchy, color, accessibility, explanatory vs. exploratory
W10 (Th)	Oct 22	CLO 3, 5	Visualization Critique & Misleading Charts Lab	Milestone 2 due: static visualization portfolio/story with justified chart choices, acce and short bias/misleading-chart critique.
W11 (T)	Oct 27	CLO 4	Interactive Visualization with Plotly Express	Interactive charts, hover, filtering, animation, and when interaction adds analytical v
W11 (Th)	Oct 29	CLO 4	Plotly Graph Objects & Dashboard Design	Graph Objects customization; linked views, layout, controls/widgets, dashboard des
W12 (T)	Nov 3	CLO 4, 5	Interactive & Geospatial Visualization	Maps, spatial encodings, interactive exploration, geographic aggregation, introducti bias concerns.
W12 (Th)	Nov 5	CLO 3–4	Visualization / Dashboard Project Lab	Milestone 3 due: working interactive visualization/dashboard prototype integrating findings; peer usability review and revision plan.
W13 (T)	Nov 10	CLO 1–3	Machine Learning I — Predictive Modeling Workflow	scikit-learn workflow; features/target; train/test split; preprocessing concept; linear r prediction vs. inference.
W13 (Th)	Nov 12	CLO 1–3	Machine Learning II	Classification framing; logistic regression and decision-tree concepts; predicted clas probabilities.

Wk	Date	CLO	Topic	Notes
			— Classification	
W14 (T)	Nov 17	CLO 3–4	Exam #2	Covers CLO 3 & 4: static, multivariate, large-data, interactive, dashboard, and geos visualization.
W14 (Th)	Nov 19	CLO 2, 3, 5	Machine Learning III — Model Evaluation	Regression metrics; confusion matrix; accuracy, precision, recall, F1; cross-validati subgroup performance.
W15 (T)	Nov 24	CLO 5	Bias, Ethics & Responsible Data/ML	Sampling, measurement, selection, survivorship, spatial/MAUP, visualization and a bias audit of own project.
W15 (Th)	Nov 26	—	No class — Thanksgiving Day	Institutional holiday.
W16 (T)	Dec 1	—	Final Project Presentations — Group A	Final project should demonstrate an end-to-end data workflow; ML is an optional ex required component.
W16 (Th)	Dec 3	—	Final Project Presentations — Group B; Course Synthesis	Course wrap-up: prepare → analyze → visualize → predict → evaluate responsibly

Grading Scale					
A+	97%	A	93%	A-	90%
B+	87%	B	83%	B-	80%
C+	77%	C	73%	C-	70%
D+	67%	D	63%	D-	60%
F	below 60.0%				

"This course must be passed with a C- or better as a CSU graduation requirement."

University Policies

Per University Policy S16-9 (PDF), relevant university policy concerning all courses, such as student responsibilities, academic integrity, accommodations, dropping and adding, consent for recording of class, etc. and available student services (e.g. learning assistance, counseling, and other resources) are listed on the Syllabus Information web page. Make sure to visit this page to review and be aware of these university policies and resources.

Course Schedule