SJSU SAN JOSÉ STATE UNIVERSITY

College of Science · Computer Science

Introduction to Data Visualization Section 01

CS 133

Fall 2024 3 Unit(s) 08/21/2024 to 12/09/2024 Modified 08/21/2024

Contact Information

Instructor: Wendy Lee Ph.D.

Office Location: MacQuarrie Hall 413

Email: wendy.lee@sjsu.edu

Office Hours: Wednesday/Thursday 3:00-4:00 PM by appointment only:

https://calendly.com/wendy-lee-sjsu/fall-2024-office-hours (https://calendly.com/wendy-lee-sjsu/fall-2024-office-hours)

CS133 Class Schedule: Tuesday/Thursday 12:00 PM - 1:15 PM @ **DH 233**

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

Students are expected to adhere to the Student Conduct Code found at http://www.sjsu.edu/studentconduct/ students/. Additionally, students should regularly attend lectures and labs (if applicable), treat instructors and peers with respect, and **refrain from the use of cell phones during any classroom activities.**

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 Goals

Understand:

- How to access data
- How to format data for visualizations
- Different data types and how to visualize them
- Data visualization as a tool for problem solving

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 and Panel.

CLO 5 Recognize and reduce data and spatial biases.

Student Learning Outcomes (SLO)

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

SLO 3 Describe and implement, at an introductory level, data analysis concepts, models, and algorithms in machine learning and artificial intelligence.

SLO 4 Explain and summarize, at a developed level, results and report findings in oral and written forms.

Program Learning Outcomes (PLO)

Upon successful completion of this course, students will meet the student learning outcomes that support the following program learning outcomes of the BS Data Science Program:

PLO 2 Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements.

PLO 3 Communicate effectively in a variety of professional contexts.

PLO 6 Apply theory, techniques, and tools throughout the data lifecycle and employ the resulting knowledge to satisfy stakeholders' needs.

Course Materials

Required Texts/Readings Textbooks:

- *Biological data exploration with Python, pandas and seaborn* by Martin Jones. June, 2020. (https://pythonforbiologists.com/biological-data-exploration-book) ISBN-13: 979-8612757238 Additional course readings, examples, exercises, etc. will be assigned and provided by the instructor.

Other Readings:

- *Hands-On Data Visualization: Interactive Storytelling From Spreadsheets to Code* by Jack Dougherty and Ilya Ilyankou, 2021. ISBN-13: 978-1492086000. Free open-access web edition at https://HandsOnDataViz.org
- Additional course readings, examples, exercises, etc. will be assigned and provided by the instructor.

Other technology requirements / equipment / material

Students will need to have either a personal laptop/desktop with Internet service or access to an on campus computer lab.

Programming environment:

- Python 3.7 or 3.8 available at https://www.python.org/downloads/
- Google Colab (https://colab.research.google.com/) with Chrome or any supported web browser

Anaconda (optional) for local installation of Jupyter notebook.
https://www.anaconda.com/products/individual

Course Requirements and Assignments

The course will consist of quizzes, hands-on lab reports, two midterm exams and a final exam.

1. **Quizzes (10%):** Quizzes will take place once a week at the end of class to assess students' knowledge of the course materials from the week prior. A unique password will be provided for each quiz during lecture.
2. **Hands-on Lab Report (20%):** The purpose of the hands-on lab is to develop students' understanding of the material and the skills in problem-solving. Students will work on the hands-on exercise with a group partner assigned by the instructor. Each student must write and submit independent lab reports. Hands-on lab reports are only accepted in Canvas. Students must submit lab reports on time to receive full credit (Unless they use a late pass. 2 late passes total). All work submitted on individual assignments **must be your own**. You may not share or copy code from fellow students or from the web/chatGPT. Infractions will be detected and will lead to an automatic failing grade for the course. If someone else copies your work, with or without your permission, you will be held responsible.
3. **Midterms (30%):**
   - Midterm 1 (15%): October 8th, 2024
   - Midterm 2 (15%): November 12th, 2024

   No make-up exams will be given if a student misses the midterm exam submission deadline (except for a legitimate excuse or other personal emergencies and student can provide documented evidence).
4. **Final Project & Presentation (20%):** Final project and presentation will be used to assess student's understanding of the course materials at the end of the semester instead of a final exam. Each team will be given a unique problem to solve for the final project. Each team consists of 2 students. Here are the key deliverables and due dates:

   - **Team Formation:** September 26, 2024
   - **Project Proposal:** October 12, 2024
   - **Deliverables and Timeline:** October 22, 2024
   - **Progress Report:** November 12, 2024
   - **Final Project Due:** December 3, 2024
   - **Presentation:** Each group gives a 10-minute, in-class presentation on December 3 or December 5, 2024, during class time.
5. **Final Exam (20%):** A cumulative Final Exam will be given on **Tuesday, December 17, 9:45 AM-12:00 PM**. If there is a time conflict, please inform the instructor at least two weeks in advance for rescheduling.

✔ Grading Information

- 10% Quizzes
- 20% Hands-on Assignments
- 30% Midterm I (15%) & Midterm II (15%)
- 20% Term Project
- 20% Final Exam

Late Policy - Hands-Ons ONLY

Life happens - You can submit **two** hands-ons late, no explanation why necessary. Please just add in the comment box of your submission "USING LATE PASS"

Late Passes: You can submit the assignment up to 3 days after the deadline.

Quiz Drop

A student's 2 lowest quiz scores will be dropped.

Incomplete work

Points will be deducted for incomplete question responses and solutions that are partially functional. Consult individual assignments for details of point allocation for each problem.

Academic Honesty

All assignments submitted, including quizzes, hands-on activities, exams, and projects, are expected to be the student's own original work. The instructor may, at any time, ask a student to explain the meaning of any part of an answer they have submitted. If the student cannot adequately explain their answer, the penalty for the first incident will be the loss of all points on that question. The penalty for the second and subsequent incidents will be the loss of all points on the assignment, along with a report to the Office of Student and Ethical Conduct.

Makeup Exams

Makeup exams will only be given in cases of illness (documented by a physician) or in documentable, extreme emergency cases.

Grading Scale

Grade	Percentage
A plus	97.0 to 100%
A	93.0 to 96.99%
A minus	90.0 to 92.99%
B plus	87.0 to 89.99%
B	82.0 to 86.99%
B minus	80.0 to 81.99
C plus	77.0 to 79.99%
C	72.0 to 76.99%

C minus	70.0 to 71.99%
D plus	67.0 to 69.99%
D	62.0 to 66.99
D minus	60.0 to 61.99%
F	<60.0

University Policies

Per University Policy S16-9 (PDF) (http://www.sjsu.edu/senate/docs/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 (https://www.sjsu.edu/curriculum/courses/syllabus-info.php) web page. Make sure to visit this page to review and be aware of these university policies and resources.

Course Schedule

The course schedule is subject to change with fair notice. Changes will be announced on Canvas.

Week	Date	Topics
1	8/22	Syllabus. Introductions. Course Expectations.
2	8/27	Introduction to Pandas
2	8/29	Pandas, Series, and DataFrame Objects
3	9/3	Data Exploration Using Pandas
3	9/5	Data Exploration Using Pandas
4	9/10	Advanced Features in Pandas
4	9/12	Intro to Seaborn

5	9/17	Representing Categorical Data
5	9/19	Reshaping Data
6	9/24	Reshaping Data
6	9/26	Handling Complicated Data Files
7	10/1	Case Study: Applying Pandas and Seaborn for Problem Solving
7	10/3	Midterm Review
8	10/8	Midterm #1
8	10/10	Matrix Charts and Heatmaps
9	10/15	Introduction to Interactive Plots with Plotly
9	10/17	Creating Maps with Geopy and Folium
10	10/22	Creating Interactive Dashboards with Panel
10	10/24	Handling Large Datasets with Polars
11	*10/29*	Visualizing High-dimensional Data in a Low-dimensional Space
12	11/5	Relational Databases - SQLite, Accessing data in Google Drive and Google Sheets
12	11/7	Midterm #2 Review
13	11/12	Midterm #2
13	11/14	Integrating Data Visualization in Web App

14	11/19	Integrating Data Visualization in Web App
14	11/21	Deploying Data Visualization Web App to the Cloud
15	11/26	Deploying Data Visualization Web App to the Cloud
15	11/28	*Thanksgiving Holiday - no class*
16	12/3	Project Due. Project Presentations
16	12/5	Project Presentations
18	12/17	Final Exam: 9:45 AM - 12:00 PM