SJSU SAN JOSÉ STATE UNIVERSITY

College of Science · Computer Science

Introduction to Data Visualization Section 02

CS 133

Fall 2023   3 Unit(s)   08/21/2023 to 12/06/2023   Modified 08/24/2023

👤 Contact Information

Instructor: Hannah DeBaets

Email: hannah.debaets@sjsu.edu

To contact me, please email my SJSU email, not through Canvas mail.

Please keep your emails as brief as possible. (Aim for 5 sentences max)

Office Hours

Wednesday 1PM-2PM, Friday 10AM-11AM
https://sjsu.zoom.us/j/4200295681

Instructor: Hannah DeBaets

Email: hannah.debaets@sjsu.edu

🖥 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.

University Policies

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Programs' Syllabus Information web page at: http://www.sjsu.edu/gup/syllabusinfo/

COVID-19 and Monkeypox

Students registered for a College of Science (CoS) class with an in-person component should view the CoS COVID-19 and Monkeypox Training slides (https://drive.google.com/drive/folders/1Vmp39U9-CNpbwRobtZsGIZPTgRwV_Nh6) for updated CoS, SJSU, county, state and federal information and guidelines, and more information can be found on the SJSU Health Advisories website (https://www.sjsu.edu/healthadvisories/). By working together to follow these safety practices, we can keep our college safer. Failure to follow safety practice(s) outlined in the training, the SJSU Health Advisories website, or instructions from instructors, TAs or CoS Safety Staff may result in dismissal from CoS buildings, facilities or field sites. Updates will be implemented as changes occur (and posted to the same links). Masking guidance: https://www.sjsu.edu/healthadvisories/masks.php

▤ 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:** Quizzes will take place once a week at the beginning 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:** 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.
3. **Midterms I & II:** 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:** 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.
5. **Final Exam:** A cumulative Final Exam will be given on 12/13/2023, from 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

At the end of the semester, after weighting the scores according to the scale below and totaling them, final grades will be assigned as follows:

A plus = 100 to 97.0 points

A = 96.9 to 93 points

A minus = 92.9 to 90.0 points

B plus = 89.9 to 87.0 points

B = 86.9 to 82.0 points

B minus = 81.9 to 80.0 points

C plus = 79.9 to 77.0 points

C = 76.9 to 72.0 points

C minus = 71.9 to 70.0 points

D plus = 69.9 to 67.0 points

D = 66.9 to 62.0 points

D minus = 61.9 to 60.0 points

F = 59.9 points or lower

Breakdown

- **Quizzes (10%)**
- **Hands-on Lab Reports (40%)**
- **Midterm I & II (20%)**
- **Final Project & Presentation (15%)**
- **Final Exam (15%)**

🏛 University Policies

Per [University Policy S16-9 (PDF) (http://www.sjsu.edu/senate/docs/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)](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

Tuesdays and Thursdays

10:30AM - 11:45AM

Classroom: MH422

When	Topic	Notes

When	Topic	Notes
Welcome Week 1 MH 422	Syllabus, Introductions, Course Expectations	• Introduction to Google Colab
Lecture Week 1 MH 422	Introduction to Pandas	
Lecture Week 2 MH 422	Pandas, Series, and DataFrame Objects	
Lecture Week 2 MH 422	Data Exploration Using Pandas	
Lecture Week 3 MH 422	Data Exploration Using Pandas	
Lecture Week 3 MH 422	Advanced Features in Pandas	
Lecture Week 4 MH 422	Intro to Seaborn	
Lecture Week 4 MH 422	Representing Categorical Data	
Lecture Week 5 MH 422	Reshaping Data	
Lecture Week 5 MH 422	Handling Complicated Data Files	
Lecture Week 6 MH 422	Matrix Charts and Heatmaps	
Activity Week 6 MH 422	Case Study: Applying Pandas and Seaborn for Problem Solving	
Lecture Week 7 MH 422	Introduction to Interactive Plots with Plotly	
Review Week 7 MH 422	Midterm 1 Review	• Covers Topics up and including heatmaps
Midterm Week 8 MH 422	Midterm 1	

When	Topic	Notes
Lecture Week 8 MH 422	Relational Databases - SQLite	
Lecture Week 9 MH 422	Accessing data in google sheets	
Lecture Week 9 MH 422	Creating Maps with Geopy and Folium	
Lecture Week 10 MH 422	Creating Interactive Dashboards with Panel	
Lecture Week 10 MH 422	Handling Large Datasets with Vaex	
Activity Week 11 MH 422	Applying Pandas and Seaborn in Data Analysis	
Lecture Week 11 MH 422	Intro to Machine Learning	• Discover and visualize the data to gain insights
Midterm Week 12 MH 422	Midterm 2	
Lecture Week 12 MH 422	Preparing the Data for Machine Learning Algorithms	
Lecture Week 13	Introduction to Scikit-Learn and pipelines	• Feature scaling and Transformation pipelines
Lecture Week 13	Select & Train a Model	
Lecture Week 14	Fine-tune the Model	
No Class Week 14	Thanksgiving Break	
Review Week 15	Review	Review Homeworks, Midterm Answers, Topics for the Final
Presentations Weeks 15	Project Presentations	
Presentations Week 16	Project Presentations	
Final 12/13/2023 9:45 AM - 12:00 PM	Final Exam	