

San José State University
Department of Computer Science
CS 133 Section 01 Data Visualization, Fall 2022

Course and Contact Information

Instructor:	Wendy Lee
Office Location:	MacQuarrie Hall 413
Email:	wendy.lee@sjsu.edu
Office Hours:	Monday 8-9 AM / Friday 8-9 AM (by appointment only)
Class Days/Time:	Tu/Th 12:00PM - 1:15PM
Classroom:	MacQuarrie Hall 222
Prerequisites:	CS22B and graduate standing, or CS146 with a grade of C- or better

Course Description

Topics in data analysis and visualization. This course will cover 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.

Course Format

Class time will be spent in-person either in “lecture” mode or in “lab” mode. Students are required to bring their wireless laptop to each class. Written and oral assessments will be used to measure student learning in this course. In case of a shelter-in-place order in response to a pandemic, this course might be conducted in a hybrid mode where lectures and labs will take place during live Zoom meetings. The Zoom lectures will be recorded and posted in the Canvas Learning Management System at <https://sjsu.instructure.com>.

Course Web Page and MYSJSU Messaging

Course materials such as syllabus, handouts, notes, assignment instructions, etc. can be found on my faculty web page on [Canvas Learning Management System course login website](https://sjsu.instructure.com) at <http://sjsu.instructure.com>. Students are responsible for regularly checking with the messaging system through [MySJSU](http://my.sjsu.edu) at <http://my.sjsu.edu> to learn of any updates.

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 Shiny.
- 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.

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 TBD, from TBD to TBD. If there is a time conflict, please inform the instructor at least two weeks in advance for rescheduling.

University Credit Hour Requirement:

Success in this course is based on the expectation that students will spend, for each unit of credit, a minimum of 45 hours over the length of the course (normally three hours per unit per week) for instruction, preparation and studying. Plan on spending at least 7 hours per week outside of lecture time engaging with the course material.

Grading Information:

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

At the end of the semester, after re-weighting the scores according to the scale above 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

Classroom Protocol

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).

Course Schedule

The course schedule is subject to change with fair notice. Changes will be announced on Canvas. Readings (BD - *Biological data exploration with Python, pandas and seaborn*)

Wk	Date	Readings	Topics
1	8/23		Syllabus. Introductions. Course Expectations.
1	8/25		Introduction to Google Colab. <i>Hands-On #1</i>
2	8/30	BD Ch 2	Introduction to Pandas. <i>Hands-On #2</i>
2	9/1	BD Ch 3,4	Pandas, Series and dataframe objects. <i>Hands-On #3</i>
3	9/6	BD Ch 5	Data exploration using pandas. <i>Hands-On #4</i>
3	9/8	BD Ch 5	Data exploration using pandas. <i>Hands-On #4</i>
4	9/13	BD Ch 12,13	Advanced features in pandas. <i>Hands-On #5</i>
4	9/15	BD Ch 6,7	Intro to seaborn & plotting special types of scatter plots. <i>Hands-On #6</i>
5	9/20	BD Ch 8,9	Using Categorical axes with Seaborn. <i>Hands-On #7</i>
5	9/22	BD Ch 14	Reshaping data. <i>Hands-On #8</i>
6	9/27	BD Ch 16	Handling complicated data files. <i>Hands-On #9</i>
6	9/29		Apply Pandas and Seaborn in Data Analysis. <i>Hands-On #10</i>
7	10/4	BD Ch 15	Matrix charts and heatmaps. <i>Hands-On #11</i>
7	10/6		Midterm 1 Review
8	10/11		Midterm 1
8	10/13		Handling very large datasets with Vaex <i>Hands-On</i>
9	10/18		Relational Database - SQLite. <i>Hands-On #12</i>
9	10/20		Introduction to Interactive Plots with Plotly <i>Hands-On #13</i>
10	10/25		Creating Maps with Geopy and Folium <i>Hands-on #14</i>
10	10/28		Introduction to machine learning
11	11/1		Discover and visualize the data to gain insights
11	11/3		Prepare the data for Machine Learning algorithms. <i>Hands-On #15</i>
12	11/8		Midterm 2
12	11/10		Introduction to Scikit-Learn. <i>Hands-On #15</i>
13	11/15		Feature scaling and Transformation pipelines.
13	11/17		Select & Train a Model. <i>Hands-On #16</i>
14	11/22		Fine-tune the Model. <i>Hands-On #16</i>
14	11/24		No class - Thanksgiving Break
15	11/29		Review <i>Hands-ons 15/16</i> , Midterm #2 Answers
15	12/1		Project Due. Project Presentations
16	12/6		Project Due. Project Presentations
17	12/8 (9:45 AM - 12:00 PM)		Final Exam