

# Introduction to Data Visualization Section 03

## CS 133

Spring 2024 3 Unit(s) 01/24/2024 to 05/13/2024 Modified 01/31/2024

### Contact Information

---

**Instructor:** Daniel Quintana

**Email:** daniel.i.quintana@sjsu.edu (Please start the subject line with "CS 133")

**Office Hours:** M/W 5:30-6:30 or by appointment (Please email me if you plan on attending my office hours)

**Class Days/Time:** Tu/Th 10:30AM - 11:15PM

**Classroom:** MacQuarrie Hall 225

### 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

### 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)

---

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.

## 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 of their choice or as 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 material and showcase their ability to implement for practical problems at the end of the semester. Each team will choose or be given a unique problem to solve for the final project.
5. Final Exam: A cumulative Final Exam will be given on May 16th, from 9:45AM to 12:00PM. If there is a time conflict, please inform the instructor at least two weeks in advance for rescheduling.

## Grading Information

---

### Course Percentages

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

## Late Assignments:

Late assignments will receive a 20% reduction for every day late.

## Grading

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

## 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

---

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*)

Week	Date	Readings	Topics
1	1/25	Syllabus	Introductions. Course Expectations. Hands-On #1
2	1/30	BD Ch 2	Introduction to Pandas. Hands-On #2
2	2/1	BD Ch 3, 4	Pandas, Series and dataframe objects. Hands-On #3

3	2/6	BD Ch 5	Data exploration using pandas. Hands-On #4
3	2/8	BD Ch 5	Data exploration using pandas. Hands-On #4
4	2/13	BD Ch 12, 13	Advanced features in pandas. Hands-On #5
4	2/15	BD Ch 6, 7	Intro to seaborn & plotting special types of scatter plots. Hands-On #6
5	2/20	BD Ch 8, 9	Using Categorical axes with Seaborn. Hands-On #7
5	2/22	BD Ch 14	Reshaping data. Hands-On #8
6	2/27	BD Ch 16	Handling complicated data files. Hands-On #9
6	2/29	-	Apply Pandas and Seaborn in Data Analysis. Hands-On #10
7	3/5	BD Ch 15	Matrix charts and heatmaps. Hands-On #11
7	3/7	-	Midterm 1 Review
8	3/12	-	Midterm 1
8	3/14	-	Handling very large datasets with Vaex Hands-On #?
9	3/19	-	Relational Database - SQLite. Hands-On #12
9	3/21	-	Introduction to Interactive Plots with Plotly Hands-On #13
10	3/26	-	Creating Maps with Geopy and Folium Hands-on #14
10	3/28	-	Introduction to machine learning
11	4/2	-	No class - Spring Break
11	4/4	-	No class - Spring Break
12	4/9	-	Discover and visualize the data to gain insights
12	4/11	-	Prepare the data for Machine Learning algorithms. Hands-On #15
13	4/16	-	Midterm 2
13	4/18	-	Introduction to Scikit-Learn. Hands-On #15
14	4/23	-	Feature scaling and Transformation pipelines.
14	4/25	-	Select & Train a Model. Hands-On #16
15	4/30	-	Fine-tune the Model. Hands-On #16

15	5/2	-	Review Hands-ons 15/16, Midterm #2 Answers
16	5/7	-	Project Due. Project Presentations
16	5/9	-	Project Due. Project Presentations
17	5/16	-	(9:45 AM-12:00 PM) Final Exam