San José State University
Department of Computer Science
CS 185C Section 01 Spring 2022
Introduction to Data Analysis and Visualization

Course and Contact Information

Instructor: Wendy Lee Ph.D.
Office Location: MacQuarrie Hall MQH413
Telephone: (408) 924-5161
Email: wendy.lee@sjsu.edu

Class Days/Time: Tuesday/Thursday 12:00 PM - 1:15 PM
Office Hours: Monday/Friday 12:00-1:00 PM by appointment only:
https://calendly.com/wendy-lee-sjsu/fall-2022-office-hours
Classroom: **Science Building SCI 311**
Prerequisites: CS22B and graduate standing, or CS146 with a grade of C- or better
Zoom link will be posted in the Canvas course shell.

Course Description

Topics in data wrangling, analysis and visualization. This course will cover tools and techniques to efficiently work with 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

- This course will be conducted in-person after February 14, 2022. Between January 27 and February 11, 2022, this course will be conducted online via Zoom.
- Course materials such as syllabus, handouts, notes, hands-on exercises, projects, quizzes, exams, etc. can be found on Canvas Learning Management System. You are responsible for regularly checking with the Canvas messaging system to learn of any updates.
- Students are required to have bring their laptop computers to attend in-person or online lectures. SJSU has a free equipment loan program available for students (https://sjsuequipment.getconnect2.com). Students are responsible for ensuring that they have access to reliable Wi-Fi during tests. If students are unable to have reliable Wi-Fi, they must inform the instructor, as soon as possible or at the latest one week before the test date to determine an alternative. See Learn Anywhere website (https://www.sjsu.edu/learnanywhere/equipment/index.php) for current Wi-Fi options on campus.

Course Learning Outcomes (CLO)

Upon successful completion of this course, students will be familiar with the following concepts and will be able to apply them in appropriate situations:

1. Manipulate large datasets and handle missing or inconsistent values in datasets.
2. Perform data analysis with scientific computing libraries such as numpy and scipy.

3. Discover and visualize datasets using data visualization libraries such as seaborn and matplotlib.
4. Perform machine learning using Scikit-learn.

Required Texts/Readings

Textbook

- *Biological data exploration with Python, pandas and seaborn* by Martin Jones, 2020.

Other Readings

Additional course readings, examples, exercises, etc. will be assigned and provided by the instructor.

Python 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

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/studying, or course related activities.

- **Quizzes (10%)**: Quizzes will take place on Tuesday at the beginning of class to assess students' knowledge on the course materials from the week before. A unique password will be provided for each quiz during lecture. ***No make up quizzes will be given.***
- **Hands-on Assignments (40%)**: Hands-on assignments will be posted and must be submitted on Canvas. All assignments must be submitted by the posted due date to receive full credit. All work submitted on individual assignments must be your own. You may not share or copy code from fellow students or from the web. Infractions will be detected and will lead to an automatic 0. If someone else copies your work, with or without your permission, you will be held responsible.
- **Midterms (MT) (20%)**: MT1 (10%): March 22, 2022, MT2 (10%) April 19, 2022. No make-up exams will be given if a student misses the midterm exam submission deadline (unless you have a legitimate excuse or other personal emergencies and can provide documented evidence).
- **Final Project & Presentation (30%)**: The final project is a group project. Each group consists of 2 students. Group partners will be assigned by the instructor. Final project (25%) and presentation (5%) will be used to assess student's understanding of the course materials at the end of the semester instead of a final exam. The Final Project report and all the associated files must be submitted in Canvas are due on May 10, 2022 at 11:59 am. Presentation: Each group gives a 10-minute, in-class presentation May 10 or May 12, 2022, during class time.

Grading Information

- 10% Quizzes
- 40% Hands-on Assignments
- 20% Midterm I & II
- 30% Final Project

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

Late assignments: No late homework will be accepted. However, under exceptional circumstances, one problem set per student might be accepted late. It will need to be handed in prior to the following class meeting and will be graded with 30% off. Such an extension should be requested from the instructor.

Academic Honesty: You may only submit your own work for all quizzes, assignments, exams, and projects. ***Copying and any other form of cheating will not be tolerated and will result in a failing grade (F) for the course, as well as disciplinary consequences from the university.***

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

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*

Classroom Protocol

In-person classroom protocol

- All students registered for a College of Science (CoS) class with an in-person component must view the CoS COVID-19 Training slides and the SJSU Phased Adapt Plan website and acknowledge reading them according to their instructor's directions. By working together to follow these county and SJSU safety practices, we can keep our college safer. Students who

do not follow COVID-19 Safety practice(s) outlined in the training, the SJSU Phased Adapt Plan, or instructions from their instructors, TAs or CoS Safety Staff may be dismissed from CoS buildings, facilities or field sites. Please review this training as needed throughout the semester, as updates will be implemented as changes occur (and posted to the same links).

- **Dual Role of SCI 311**: Lecture/Lab SCI 311 will be used as a dual-purpose room. It can be a regular lecture room, or it can be a computer laboratory for hands-on exercises.
    - **Lecture Mode:** This is when SCI 311 is used as a regular lecture room. Students are expected to listen and follow the lecture. Be considerate to your classmates and follow the lecture.
    - **Lab Mode:** This is when SCI 311 is used as a computer lab. Students must bring their laptop to every class. Students are expected to work collaboratively on problems of the Hands-On and share your ideas and solutions with your classmates.

  We shall alternate between the two modes. A typical class will begin with a lecture (Lecture Mode) followed by a hands-on (Lab Mode).
- Regular class attendance is strongly encouraged.
- Please arrive to class on-time so that you benefit fully from the course experience and you do not disturb classmates and the instructor while class is in session.
- Students are responsible for knowing all materials covered in class lectures, readings, assignments, and other course-related work.
- Please do not use mobile phones during class time. Laptops, tablets and other devices should only be used for course-related purposes.

Virtual classroom protocol

- **Live Session via Zoom**: Live Zoom meetings will be used as dual-purpose virtual classrooms. A meeting can be a regular lecture room, or it can be a computer laboratory for hands-on team exercises in break-out rooms.
- **Lecture Mode**: This is when Zoom is used as a virtual lecture room. You are expected to listen and follow the lecture. Be considerate to your classmates and follow the lecture. Keep your microphone muted except when speaking to the instructor. You may use the chat in Zoom to post questions during lecture.
- **Lab Mode**: Zoom break-out rooms will be used to group you into teams of three or more to work on hands-on lab exercises. Work collaboratively on the exercises and share your ideas and solutions with your classmates.
- **Attendance**: Live virtual class attendance is strongly encouraged. Follow the rules of netiquette. Be respectful. Dress appropriately if you are going to participate in the virtual classroom with the camera on.
- **Recording of Zoom Classes**: The instructor will record the live virtual classes using Zoom and the recordings will be shared in the Canvas course shell. If you do not wish to be identified in a class recording, please contact the instructor to arrange an “anonymous” option prior to class.
- **Zoom recordings and course materials**: You are allowed view the Zoom recordings for your own study purposes only. You may not record any course materials. You may not share any class recordings or course materials with someone who isn’t enrolled in the without permission from the instructor. The lecture recordings and course materials are protected by copyright.
- **Accessibility**: If you need accommodations or assistive technology you should work with the Accessible Education Center (AEC) and the instructor.

- **Be Punctual**: Please arrive to the live sessions on time so that you benefit fully from the course experience and do not disturb classmates and the instructor while class is in session.
- **Stay on top of coursework**: You are responsible for knowing all material covered in lectures, assignments, quizzes, and course-related work.

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

Per University Policy S14-7, accommodation shall be provided on any graded class work or activities for students wishing to observe religious holidays when such observances require students to be absent from class. It is the responsibility of the student to inform the instructor, in writing, about such holidays before the add deadline at the start of each semester. If such holidays occur before the add deadline, the student must notify the instructor, in writing, at least three days before the date that student will be absent. It is the responsibility of the instructor to make every reasonable effort to honor the student request without penalty, and of the student to make up the work missed.

CS 185C Section 01, Introduction to Data Analysis and Visualization Spring 2022

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	1/27		Syllabus. Introductions. Course Expectations.
2	2/1	BD Ch 2	Introduction to Pandas. Hands-On #2
2	2/3	BD Ch 3,4	Pandas, Series and dataframe objects. Hands-On #3
3	2/8	BD Ch 5	Data exploration using pandas. Hands-On #4
3	2/10	BD Ch 5	Data exploration using pandas. Hands-On #4
4	2/15	BD Ch 12,13	Advanced features in pandas. Hands-On #5
4	2/17	BD Ch 6,7	Intro to seaborn & plotting special types of scatter plots. Hands-On #6
5	2/22	BD Ch 8,9	Using Categorical axes with Seaborn. Hands-On #7
5	2/24	BD Ch 14	Reshaping data. Hands-On #8
6	3/1	BD Ch 16	Handling complicated data files. Hands-On #9
6	3/3		Apply Pandas and Seaborn in Data Analysis. Hands-On #10
7	3/8	BD Ch 15	Matrix charts and heatmaps. Hands-On #11
7	3/10		Midterm 1 Review
8	**3/15**		**Midterm 1**
8	3/17	TBA	Handling very large datasets with Vaex Hands-On #12
9	3/22	TBA	Relational Database - SQLite. Hands-On #13
9	3/24	TBA	Introduction to Interactive Plots with Plotly Hands-On #14
10	*3/29*	***Spring Break - no class***	

Week	Date	Readings	Topics
10	*3/31*		***Spring Break - no class***
11	4/5		Midterm Exam #1 Answers
11	4/7	TBA	Introduction to machine learning.
12	4/12	TBA	Discover and visualize the data to gain insights.
12	4/14	TBA	Prepare the data for Machine Learning algorithms. Hands-On #15
13	**4/19**		**Midterm 2**
13	**4/21**	TBA	Introduction to Scikit-Learn. Hands-On #15
14	4/26	TBA	Feature scaling and Transformation pipelines. Hands-On #16
14	**4/28**	TBA	Select & Train a Model. Hands-On #16
15	5/3	TBA	Fine-tune the Model. Hands-On #16
15	5/5		Review Hands-ons 15/16, Midterm #2 Answers
16	**5/10**		**Project Due. Project Presentations**
16	**5/12**		**Project Presentations**

Important Dates:

Feb 7, 2022: Last Day to Drop Classes without a "W" grade
Feb 14, 2022: Last Day to Add Classes via MySJSU
Apr 22, 2022: Semester Withdrawal Deadline