Syllabus
DATA 271 Data Wrangling and Visualization
23685
Spring 2024

Course Meeting Information

- **This course meets:**
  - **Lecture:** Monday, Wednesday, Friday 11AM – 11:50 AM SCIA 564
  - **Lab activity:** Thursday 11AM – 12:50PM, BSS 317
- **Final exam:** 5/8/2024, Wednesday 10:20AM – 12:10PM, SCIA 564

Please be sure to regularly check Canvas and your Cal Poly Humboldt email for course information and updates.

Instructor Information

- **Name: Dr. Bethany Johnson**
- **Email:** bethany.johnson@humboldt.edu (please include "*DATA 271"* in subject when you email me about this course)
- **Office Hours and location:** We 9-10AM & 12-2PM, Thurs 9-10AM & 1-2 PM in BSS 332 (or by appointment)

Course Description

Catalog description: This course provides an intensive, hands-on introduction to data wrangling, cleaning and visualization using a contemporary programming language. You will learn the fundamental skills required to acquire, transform, manipulate, and visualize data in a computing environment that fosters reproducibility. The overall goal is to create actionable data from raw sources and then perform exploratory analysis. These steps will include importing data, viewing data diagnostically, identifying outliers, imputing data, cleaning data, calculating basic statistics and creating informative plots.

A minimum grade of a C- is required to count towards the Data Science major.

Prerequisites: Data 111 and CS 111.

Course Learning Outcomes. Upon completion of this course, you will be able to:

- Perform your data analysis in a literate programming environment with a variety of data structures.
- Import and manage structured and unstructured data and adapt data to a desirable format for analysis and visualization

- Manage different data structures
- Import, scrape, impute, and export data
- Index, subset, slice, reshape, manipulate, transform, and summarize your data
- Join disparate data sources
- Methodically explore and visualize data
- Create effective graphs and plots using "the grammar of graphics"
- Use current analysis, presentation, and collaboration tools in the Data Science field to communicate with data scientists and other parties (e.g. GitHub, Jupyter Notebooks)
- Describe where data wrangling and visualization resides within the Data Science profession

Course Readings, Materials & Fees

Textbook:

- *Python Data Cleaning Cookbook* by Michael Walker and Data

This book is available for free through the library here.

- *Hands on Data Analysis with Pandas* by Stefanie Molin

This book is **not required, but it may be a helpful resource.** It is available for free through the library here.

Additional Materials: Readings and other materials such as assignments, labs, and slides will be provided on Canvas.

Computer and Internet: You will need a computer and internet access. You may use campus computer labs and, if needed, you may borrow laptops. Please contact me if you need a laptop.

Software needed for this course: Access to Python. Python is free, and is available on computers in our computer labs, or can be downloaded for free to your individual computer. We will be using a cloud-based JupyterHub for this class, however installation of Anaconda is highly encouraged.

Course Topics & Schedule

The schedule below is preliminary and is subject to change with fair notice. A more detailed schedule with updated information can be found on Canvas.

Week	Day	Main Topics
1	Wed Jan 17	Welcome! Data Types in Python
	Fri Jan 19	Data Structures
2	Mon Jan 22	Github

	Wed Jan 24	List and Dictionary Comprehension
	Fri Jan 26	Loops/Conditioning using Comprehension
3	Mon Jan 29	Syntax for Regular Expression
	Wed Jan 31	Python's re module
	Fri Feb 2	Markdown in Python and Reproducible Research
4	Mon Feb 5	NumPy
	Wed Feb 7	Applications of NumPy
	Fri Feb 9	Method Chaining and Intro to Pandas
5	Mon Feb 12	Using Series and DataFrames
	Wed Feb 14	Importing Data from CSV or Excel
	Fri Feb 16	Subsetting/Filtering Data with Pandas
6	Mon Feb 19	Visualization: Matplotlib Basics
	Wed Feb 21	Review for Midterm
	Thurs Feb 22	**MIDTERM 1**
	Fri Feb 23	More on Matplotlib & AutoViz

7	Mon Feb 26	Grammar of Graphics
	Wed Feb 28	Grammar of Graphics
	Fri March 1	Creating Meaningful Subplots
8	Mon March 4	Bar charts, Histograms, Scatter Plots
	Wed March 6	Summary Statistics & Detecting Outliers
	Fri March 8	Seaborn & Other Data Exploation
9	Mon March 18	Advanced Plotting: Box/Violin Plots
	Wed March 20	Advanced Plotting: Swarm Plots and Subplotting
	Fri March 22	Creating New Variables
10	Mon March 25	Importing Data with APIs
	Wed March 27	Cleaning Data: indexing, sorting, subsetting
	Fri March 29	More on Cleaning Data
11	Mon April 1	*Campus Holiday: NO CLASS*
	Wed April 3	Review for Midterm
	Thurs April 4	**Midterm 2**

	Fri April 5	Advanced Pandas
12	Mon April 8	Web Scraping
	Wed April 10	More Web Scraping
	Fri April 12	Ethics of Web Scraping
13	Mon April 15	Peer Review of Projects
	Wed April 17	Peer Review of Projects
	Fri April 19	Importing Data from SQL
14	Mon April 22	Tidying Data
	Wed April 24	Reshaping Data
	Fri April 26	Tidying for Machine Learning
15	Mon April 29	Types of Missing Data & Why It Occurs
	Wed May 1	Handling Missing Data With Pandas
	Fri May 3	Prepare for Final Presentations
16	**Final**: Wed May 8	Final Project Presentations

Course Structure & Modality

- This course includes in-person meetings on-campus. Class time will include lectures, discussions, and activities. This course will also include work on Canvas, where you will find a detailed schedule, links to course materials, assignment instructions, and so on. Each week, be sure to review Canvas before class so that you are prepared for in-class work and upcoming assignments.
- This is a 4-unit course, and thus you should expect at least 180 hours of work during the semester, which equates to an average of at least 12 hours per week between in-class/instructional activities, review of materials, and completion of assignments.

NOTE: Class meets 3 hours per week for lecture and 2 hours per week for activity, so expect at least 7 hours per week of work outside of class.

Assignments:

Attendance and Participation: Regular attendance is mandatory for this course and is vital to your succuss. Attending classes offers several benefits including:

1. Active Learning: In-class engagement promotes active learning through discussions and activities.
2. Immediate Feedback: Attending class allows you to seek clarification, ask questions, and get feedback.
3. Collaborative Environment: Interacting with your peers fosters shared knowledge and diverse perspectives.
4. Access to Important Information: Attending class allows you to stay up-to-date with announcements and additional resources.
5. Building a learning community: Attendance helps you make connections with your fellow students and build networks beyond this course.

Remember that attendance is not just about fulfilling a requirement but about taking an active role in your learning journey. Your presence and participation contribute to a rich and engaging learning environment that benefit both you and your peers. We all want you to be here!

Exceptions and Accommodations: If unforeseen circumstances prevent attendance, please communicate with me *as soon as possible* to discuss options.

Exams: There will be two in-class exams, the last being the final exam. A study guide and details for each exam will be provided in class and on Canvas. They will be held during lab sessions in Week 6 and Week 11.

Homework: Bi-weekly problem sets will be assigned and collected through a file upload on Canvas. Your lowest homework score will be dropped.

Lab Activities: Weekly lab reports will be assigned and collected through a file upload on Canvas. Your lowest score will be dropped.

Projects: There will be 1 large project for you to work on throughout the semester. Your project idea is due in week 5, your project checkpoint is due week 10, and your final project is due Friday, May 3. More information is in the project description and specific dates are available on Canvas.

Notes:

- If you have problems submitting any assignments on Canvas, contact Canvas 24/7 Support (help button in Canvas) or the Cal Poly Humboldt Help Desk (help@humboldt.edu or 707-826-4357).
- This class includes assignments that will be submitted online outside of Canvas; however, faculty review of your work will not be done publicly. If assignments ask you to submit work in a public space that is visible to people outside of the course, please contact the instructor if you would like to discuss alternative options.

Grading & Evaluation:

Assignment Category	% Course Grade
Homework	25%
Lab activities	25%
Midterms (2)	10% each
Project Idea & Checkpoint	5% each
Final Project	20%

You will earn a grade in this course based on the following scale:

- **A** = 93% or above; **A-** = 92.9-90%
- **B+** = 89.9-87%; **B** = 86.9-83%; **B-** = 82.9-80%
- **C+** = 79.9-77%; **C** = 76.9-73%; **C-** = 72.9-70%
- **D+** = 69.9-67%; **D** = 66.9-60%; **F** = <60%.
- For **CR/NC**, 70% or above is needed to pass

Late / Make-Up Policy

If you have an emergency, please notify me immediately if this affects your coursework or attendance. Late work will be accepted only with a serious and compelling reason and documentation. I will ask you to work with the CARE office in this situation. Their information is here: https://deanofstudents.humboldt.edu/CARE.

For help documenting or receiving services for serious, confidential, or ongoing circumstances, please contact the Dean of Students office https://deanofstudents.humboldt.edu

Policies, Procedures, and Resources

- **Emergency Procedures:** Review the evacuation plan and emergency procedures for the classroom. During an emergency, information can be found on campus conditions at 707-826-INFO or the HSU Emergency website.

- **Academic Honesty:** Students are responsible for knowing policy regarding academic honesty. For more information, visit: Academic Honesty Policy
- **Students with Disabilities:** Persons who wish to request disability-related accommodations should contact me immediately so I can assist you in a timely manner. If you have not yet done so, please request services with the Student Disability Resource Center in the Learning Commons, Lower Library, 826-4678 (voice) or 826-5392 (TDD). Some accommodations may take up to several weeks to arrange. Student Disability Resource Center.
- **Add/Drop Policy:** Students are responsible for knowing the University policy, procedures, and schedule for dropping or adding classes. The deadline for students to change their course schedule without penalty, known as the Add/Drop date is 11:59 p.m. on the Monday after the second week of classes of the regular semester term. After the deadline, approval to add or drop a class requires approval of a documented serious and compelling reason. See the Resolution on Decoupling Drop/Add Dates from Census (22-14/15-APC), registration help and withdrawal process for additional information.
- **Attendance and Disruptive Behavior:** Students are responsible for knowing policy regarding attendance and disruptive behavior: Class Attendance and Disruptive Behavior.
- **Inclusivity:** Students in this class are encouraged to speak up and participate in class and on assignments outside of class. Each of us must show respect for each other because our class represents a diversity of beliefs, backgrounds, and experiences. I believe that this is what will enrich all of our experiences together. I recognize that our individual differences can deepen our understanding of one another and the world around us, rather than divide us. In this class, people of all ethnicities, genders and gender identities, religions, ages, sexual orientations, disabilities, socioeconomic backgrounds, regions, and nationalities are strongly encouraged to share their rich array of perspectives and experiences. If you feel your differences may in some way isolate you from our classroom community or if you have a specific need, please speak with me early in the semester so that we can work together to help you become an active and engaged member of our class and community.
- **Expectations of the Student**

    1. Log into the Canvas frequently. Attend class regularly. Schedule regular time to work on assignments, read the text, and study.

    2. Prepare to the best of your ability for every aspect of this course.

    3. Take the opportunity to learn how to write your own thoughts and code; don't plagiarize. Be sure to give credit where credit is due and cite your sources or use footnotes or endnotes. In particular, do not post material or take solutions from Chegg or Course Hero or copy code from classmates.

    4. Learning through collaboration (defined as working with or learning from another) is an effective tool used in this class and in your future employment. Collaboration on assignments is encouraged, writing individual code, solutions and lab reports is required.

- **Expectations of the Instructor** If we all live up to our academic responsibilities, this course will be meaningful for all who participate. Please feel free to discuss these points with me at any time during the course this semester.

    1. I will prepare and review course materials to be as current and accurate as possible.

    2. I will be available to answer questions or issues that may arise for you during this course. Expect a 48-hour turnaround time for response to emails during the work week. I do not always read and reply to emails in the evening or on the weekends or holidays.

    3. I will try to the best of my ability to prepare you for the exams and other assessments in this course.

4. I will utilize fair and honest evaluation techniques for each assignment required for this course.

5. To the best of my ability, I will make this a valid and worthwhile learning experience.

6. I will do my best to address the needs of a diverse range of learning styles in this course.

Student support services

Learning Center
The HSU Learning Center has a wide range of academic support services, such as **tutoring**, **supplemental instruction**, study skills, and more. HSU Learning Center

Writing Studio
The Writing Center offers free peer assistance with writing assignments and standardized writing examination preparation. The Writing Studio's web site can be accessed at HSU Writing Studio

Tutoring Services
The Learning Center provides tutorial assistance to students having difficulties in specific Courses. The Tutoring Services web site can be accessed at HSU Tutorial Services

Advising
The Academic and Career Advising Center supports students' pursuit of educational goals. HSU Academic & Career Advising Center

Mentoring
Peer academic support through Retention Through Academic Mentoring Program (RAMP) HSU RAMP Program

Registration
The Office of the Registrar can guide you through all your registration information. HSU Office of the Registrar

Counseling
The Counseling & Psychological Services Center supports the wellbeing of HSU students. **HSU Counseling & Psychological Services**

Information on the syllabus is subject to change with notice. Any significant changes will be marked in the syllabus and communicated to students via email and/or an announcement on Canvas.