

DATA 101

Introduction to Data Science

Professor:

Dr. Joanna Bieri
joanna_bieri@redlands.edu

Class:

Fall 2024
T,Th - 1:15-2:30 and 2:40-3:55

Office Hours:

Posted on the class website

Welcome to Data Science

The goal of this course is to introduce the process and best practices of Data Science. Classes, lectures, and group interactions should be a safe learning space where each member of the community feels valued, listened to, and respected. It is the responsibility of each of us to ensure that we are lifting up our peers, breaking free from old biases, reaching across cultural or socioeconomic boundaries, and supporting each other. If at any time in the class you feel that there is something that could be done to improve your learning or your experience please let me know. I am here to lift you up and help you learn!

This course is a introduction to data science using Python with a focus on question formulation, data collection and cleaning, exploratory data analysis, data visualization, prediction and communication of results. This will include data wrangling, modeling, visualization and ethics. Emphasis on reproducible results and being a good data custodian, eg. CARE and FAIR data principles. Attendance is expected at all class meetings unless you have notified me ahead of time.

Course Learning Objectives

By the end of the semester your work must reflect your ability to

1. Use computers to explore and wrangle data.
2. Write basic code, or update given code, to do data science projects and modeling.
3. Summarize, visualize, and analyze given data.

4. Present reproducible results of a data science project in written, oral, and graphical form.
5. Understand and implement the principles behind being a good data custodian.
6. Engage with moral issues surrounding the collection, use, and storage of data.
7. Formulate a data focused research question or research area.
8. Work in a team using good communication, version control, and code sharing techniques.
9. Have some fun exploring the data!

What are your goals for the class?

Technology

You absolutely need access to a computer that can run Anaconda Python and JupyterLab. We will install all the packages you need on the first day of class. We will use JupyterLab to do our data science programming and GitLab for version control. If all of this sounds like a foreign language - have no fear! It will all be explained in class.

Homework and announcements will be posted on the class website. Homework will be submitted on Canvas.

Textbooks

This class does not have a required textbook, optional resources and interesting reads:

- Data Science for Beginners - Free Online Resources - Cost: \$0
- Data Feminism, Catherine D'Ignazio (978-0262044004) - Cost: \$25

- Weapons of Math Destruction, Cathy O'Neil (978-0553418811) - Cost: \$33
- Invisible Women, Caroline Criado Perez (978-1419735219) - Cost \$14
- The Theory that Would Not Die, Sharon Bertsch McGrayne (978-0300188226) - Cost \$14
- Freakonomics, Levitt and Dubner (978-0063032378) - Cost \$12

Classwork

Consistent participation and engagement are crucial to your success in the class. Please read the following carefully:

- i. **Pre-Class Preparation:** You should come to class prepared to learn. This includes completing the homework assignments and reading from the previous class, taking the pre-class quiz, and watching any pre-class lectures available on the course website.
- ii **In-Class Participation:** In class you should be actively working on and discussing Data Science. Bring your computer and try out all the things we are learning about! We will talk about more advanced ideas in class. Missing more than 2 days of class (unexcused) will result in a 50% reduction in your In-Class Participation grade and each additional missed class will decrease your grade further
- iii **Weekly Homework:** Homework will be assigned daily and due each week on Friday evening. Each assignment should include both working

Python code where appropriate for the assignment along with a written discussions of the data and reflection of your progress. Homework is graded as pass-fail. If you fail you can resubmit the assignment once.

- iv **Exams:** There will be two midterm exams - both of which you work on in groups.
- v **Final Project:** Working in teams of 2-4 students you will complete a final data science project inspired by an area of interest to the group.

Evaluation Procedures

There are four elements of the course which contribute to your overall grade, as follows:

Pre-Class Preparation	10%
In-Class Participation	10%
Homework	30%
Midterm Exams	30%
Final Project	20%

How to Succeed in this Class

All work for the course will be graded according following criteria.

- Students can demonstrate pre-class preparation by coming to class prepared. This means watching all the videos, completing all required reading, and starting on the assignments or worksheets. Additionally, daily quizzes will demonstrate your preparation for class.
- Students can demonstrate in-class participation by attending every class. In addition, students should be actively involved in the discussion during class time. For some this will be active speaking, for others it will be taking notes and posting them for the class, for others it will be talking in small groups or helping with collaboration. There are lots of ways to participate, let me know how you best like to contribute!

- Each weekly homework will be submitted on Canvas and graded for 10 points. See the homework prompts on the class website for specific details on each assignment. You will submit your own individual homework, but working in groups is encouraged.
- Exams will have three parts. First you will work individually on the Exam (take home - no outside help) and submit your work before the start of the exam period. Then working in groups you will compare your work with your peers in the class, helping everyone in your group to get the best possible grade. After the exam period, you will work individually on the exam (take home - no outside help) and submit it for grading on Canvas.
- The final project is your chance to show mastery of the concepts for the class. You will propose your final project during the semester and work with your peers and professor to clearly define your goals.

Grading Scales:

4.0	95-100%
3.7	90-94%
3.3	87-89%
3.0	83-86%
2.7	80-82%
2.3	77-79%
2.0	73-76%
1.7	70-72%
1.3	67-69%
1.0	63-66%
0.7	60-62%

Please note: According to the University Course Catalog (p. 13) 3.7 and 4.0 are reserved for “outstanding” work; 3.3 and 3.0 are both defined as “excellent,” not mediocre.

Communication

- The most reliable way to reach me is by email. Please note that my normal working hours are 9 a.m. to 5 p.m., Monday to Friday. I do not respond to emails after 5 p.m. or on weekends, except in an emergency.
- You can make appointments with me via email. Appointments can happen in Duke 209 or on Teams.
- It is important that you communicate throughout the semester. Let me know if there are ways I can improve your learning in the class. If you are going to miss class or need an extension on the homework, the earlier you tell me the better!

Health Protocols

- In an effort to keep the classroom community safe and healthy, please follow the guidelines outlined here:
 - Wearing a mask is voluntary.
 - Do not come to class if you feel ill or have been exposed to someone who is ill.
- In any case of the above, e-mail me directly to reconcile any class work and/or attendance issues. Please contact me if you have any concerns as to your health needs and goals for the semester.

Academic Honesty

The University of Redlands enforces strict standards as regards academic honesty, and students may be dismissed for breaches of these standards.

In light of this, please note that:

- intentional plagiarism—i.e. piecemeal or wholesale appropriation of text from one or more printed or internet source—will result in a fail grade for the course.
- plagiarism by default—i.e. uncredited adoption of ideas from source texts due to carelessness in citation—will result in a fail grade for the project.

Artificial Intelligence: There is no tolerance for the use of generative artificial intelligence in place of individual work and thinking in this course. All work is to be considered a student's own. Any violation of that will result in a failing grade for the work.

If you are still in any doubt about what constitutes plagiarism, please ask me before you hand in your work!

Office of Equity and Title IX

In order to provide a safe and equitable learning environment for all students, faculty, and staff, discrimination, harassment, retaliation, sexual misconduct, and sexual harassment (including sexual assault, dating or domestic violence, and stalking) are not tolerated at the University of Redlands. The University prohibits unlawful discrimination or harassment (as defined in the Policy Prohibiting Discrimination, Harassment, Sexual Misconduct, and Retaliation) on the basis of age, color, race, ethnicity, national origin, ancestry, sex, marital status, pregnancy, status as a complaining party of domestic violence, sexual orientation, gender, gender identity or expression, physical or mental disability, genetic information, religion/creed, citizenship status (except to comply with legal requirements for employment), military/veteran status, or any other characteristic protected by law. If you or someone you know has experienced or experiences any of these behaviors, know that you are not alone. You can contact the Office of Equity and Title IX for reporting options, supportive measures, and resources to support you.

Many faculty and staff at the University of Redlands are considered "Responsible Employees," which means that if you tell me about a situation involving any of the above, I must report the matter to the Office of Equity and Title IX. Although I make that report, you are in control of how you would

like to proceed, including whether or not you wish to pursue a formal complaint. Our goal is to make sure you are aware of the range of reporting options available to you and have access to the support and resources you need.

To report an incident directly, you can:

- Contact the Interim Director of Equity & Title IX, Christopher Jones, at 909-748-8289 or titleix@redlands.edu
- Report online at: www.redlands.edu/titleixandequity

You can also report to local law enforcement at 909-798-7681, ext. 1. If you are ever in immediate danger, please call 911 or email/text 911@redlandspolice.org if you cannot call. To reach Public Safety on campus, call 909-748-8888 or use the Rave Gaudian App

If you wish to speak to someone confidentially (meaning not connecting with the Office of Equity and Title IX Office), you can contact the following resources:

- Campus:
 - Counseling Center: 909-748-8108 or 24-Hour Crisis Line: 909-748-8960
 - TimelyCare, 24/7 emotional support; 12 free telehealth counseling sessions
- Community:
 - Partners Against Violence, 24-hour sexual assault crisis line: 909-885-8884
 - Option House, 24-hour dating/domestic violence crisis line: 909-381-3471 Online: <https://www.rainn.org/> (sexual assault); <https://www.loveisrespect.org/>

You can also report to local law enforcement at (909) 798-7681, ext. 1. If you are ever in immediate

danger, please call 911 or email/text 911@redlandspolice.org if you cannot call.

If you wish to speak to someone confidentially (meaning not connecting with the Office of Equity and Title IX Office), you can contact the following resources:

- Campus: Counseling Service: 909-748-8108 or 24-Hour Crisis Line: 909-748-8960 or Chaplain's Office: 909-748-8368
- Community: Partners Against Violence, 24-hour sexual assault crisis line: 909-885-8884
- Option House. 24-hour dating/domestic violence crisis line: 909-381-3471
- Online chat: <https://www.loveisrespect.org/>

For more information, visit www.redlands.edu/titleixandequity

Accommodations

If you are a student with a disability requesting reasonable academic accommodations in this course, please contact Academic Support and Accessibility (ASA). ASA is located in the Student Success Center on the ground floor of the Armacost Library. You can reach the office at 909-748-8069 or asa@redlands.edu. All requests for reasonable accommodations require registration with ASA in advance of need. Faculty, students, and ASA will work together regarding classroom accommodations. You are encouraged to discuss approved accommodations with your faculty. More information is available on the ASA webpage.

If there are ways that simple changes to the class could improve your learning please feel free to reach out to me directly.

Counseling Center

The Counseling Center provides free and confidential mental health services, including short-term

individual therapy, group therapy, single-session therapy, consultations, and urgent appointments to all students with in-person or virtual options. Our Counseling Center is committed to inclusivity and to providing a supportive space for everyone. Please call 909-748-8108 to schedule an appointment or email counseling_center@redlands.edu. If a student is in crisis, please call 909-748-8960 for the 24/7 mental health crisis line. For more information on our resources, go here. Another option for individual therapy for all students is TimelyCare, which provides virtual therapy immediately (Talk Now) or up to 12 scheduled virtual therapy sessions per year. Students can choose their therapist from a list of providers for the scheduled therapy option.

Conflict Resolution Center

Experiencing a conflict? Whether it's with a friend, roommate, another member of a student organization, or faculty or staff member, conflicts happen. Learning to navigate conflicts is important to success in virtually any field, and a vital step in being a part of a community and having healthy, meaningful relationships with others. See <https://sites.redlands.edu/conflict-resolution-center/student-resources/> for more information.

The Care Team

The University CARE Team exists to help provide support and resources to students that are overwhelmed, experiencing significant distress, or possibly present some risk to themselves or others. As a faculty member, I may reach out to students about whom I am concerned to talk individually, and/or refer them to the CARE Team. If you have concerns about a fellow student, consider sharing your concern with the CARE Team via their online form. This is part of who we are as a caring,

proactive community where we all look out for one another. Additionally, if you feel that you or someone else needs immediate mental health support, the University has a 24/7 mental health crisis line at 909-748-8960, and the Timely Care app, which offers on-demand emotional care. Both services connect to a live, licensed counselor.

Additional Resources

If you are in need of additional resources, please refer to the available programs below.

Book Lending Program

The Book Lending Program is an initiative to ensure the academic success of First-Generation students (students who are the first to go to college in their families who meet a particular estimated family contribution [EFC] level). Funded through alumni donations, this program provides books and other classroom materials, when needed, for First-Generation students who could not otherwise afford to purchase them. Books are returned at the end of the course, to be used by other First-Generation students the next semester. The program works alongside the Library and faculty members to ensure the availability of books and classroom materials. For more information, click the link above or contact blp@redlands.edu.

Emergency Student Loans

Student Financial Services (SFS) administers a short-term, no-interest loan fund to assist students experiencing an emergency or cash-flow problem. Except in unusual circumstances, these loans do not exceed \$200 and are billed to the student's account. Evidence of repayment ability is a prerequisite for all short-term loans made to students. Students are not eligible for more than one emergency student loan

per term. Contact: SFS@redlands.edu or x8047

Student Lounges

Lounges for all students to sit, work, and eat can be found here on the University website.

Student Affairs Discretionary Fund

These endowed funds in Student Affairs can be used to support student success and remove impediments that otherwise may cause the student to stop or leave school. To utilize this fund, divisional leadership should be made aware of the student in dire need of financial support. This support can be anything from personal expenses, such as utility bills, gas money, emergency trips home due to family tragedy, off-campus counseling, and other medical costs, and occasionally mental health assessment expenses. Students receive grants based on their financial need. Contact: student_affairs@redlands.edu.

Student Food Support Pantry

The Student Food Support Pantry is a resource available to all established full and part-time University of Redlands students facing food insecurities. The Pantry is located on the north side of North Hall. This space is an open, no-questions-asked space with dried and canned goods, and non-perishable items, as well as seasonal fresh produce from our sustainable farm and limited refrigerated goods. Food for this distribution is provided in partnership with Feeding America Riverside and San Bernardino. It is also funded through private donations, ASUR, and the Office of Community Service Learning. For more information, please contact SURF@redlands.edu.

Course Schedule

Schedule is subject to change as we progress through the semester. You will be notified of any changes in class.

Introduction to Data Science – Fall 2024					
DATE	WEEK	DAY	Lecture Topic	Homework	Steps for success – Final Project Info – Writing
09/03/24	1	Tuesday	What is Data Science Install computer packages - GIT - Python or R - Basics of computation	Get your computer set up	CLASS PREP – You should be taking notes when you watch the class videos and following along in the provided python notebook.
09/04/24	1	Wednesday	LAB	Make sure your computer is working!	
09/05/24	1	Thursday	Hello World! Applications	HW1 – Due Friday 11:59pm	Make sure to try the exercises as you prepare for class.
09/10/24	2	Tuesday	Data and visualization		
09/11/24	2	Wednesday	LAB		
09/12/24	2	Thursday	Visualizing Numerical Data Visualizing Categorical Data	HW2 – Due Friday 11:59pm Choose a book or set of articles to explore for reading week. GOAL – Think deeply about what data science means to you and how you might use data science for good.	
09/17/24	3	Tuesday	Reading Day (no class)		
09/18/24	3	Wednesday	NO LAB		
09/19/24	3	Thursday	Reading Day (no class)	HW3 – Due Friday 11:59pm	
09/24/24	4	Tuesday	Data Wrangling		It's okay if this is hard a first, come get help!
09/25/24	4	Wednesday	LAB		
09/26/24	4	Thursday	Working with DataFrames and Tidying Data	HW4 – Due Friday 11:59pm	Start thinking about Final Project Ideas. Look at some of the available data sets on our class webpage.
10/01/24	5	Tuesday	Data Types and Classes		Seriously, it's okay if this is hard, but I really want you to come get some help!
10/02/24	5	Wednesday	LAB		
10/03/24	5	Thursday	Importing and Recoding data	HW5 – Due Friday 11:59pm	Exam 1 handed out.
10/08/24	6	Tuesday	Study Day		
10/09/24	6	Wednesday	NO LAB		
10/10/24	6	Thursday	Work on Exam 1 – Data Basics – in Groups.	Submit individual “draft” work on Exam1 before class! Exam 1 and Self Assessment due – Sunday 10/13/24	Get some initial data for your final project. Start the data preparation process.

			Tips for effective data visualization.		
10/15/24	7	Tuesday	Doing Data Science		How did the exam go? It's not too late to get some help.
10/16/24	7	Wednesday	LAB		
10/17/24	7	Thursday	Web Scraping	HW6 – Due Friday 11:59pm	
10/22/24	8	Tuesday	Misrepresentation and Data Privacy		
10/23/24	8	Wednesday	LAB		
10/24/24	8	Thursday	Algorithmic Bias	HW7 – Due Friday 11:59pm	
10/29/24	9	Tuesday	Fitting and interpreting Models Modeling nonlinear relationships		Complete some exploratory data analysis on your Final Project data.
10/30/24	9	Wednesday	LAB		
10/31/24	9	Thursday	Models with multiple predictors	HW8 – Due Friday 11:59pm	Exam 2 handed out.
					Don't forget to sign up for Spring Classes:
				Submit individual “draft” work on Exam1 before class!	DATA/GIS 167 – Introduction to Programming in Python
11/05/24	10	Tuesday	Work on Exam 2 – Data Visualizations – in Groups	Exam 2 and Self Assessment due – Wednesday 11/6/24	DATA 100 – Math for Data Science
11/06/24	10	Wednesday	NO LAB		
				Final Project Proposals – due before class.	
11/07/24	10	Thursday	Logistic Regression, Prediction, and Overfitting	HW9 – Due Friday 11:59pm	Final Project Proposals should include a description of your questions, a data set, and some initial Exploratory Data Analysis (EDA).
11/12/24	11	Tuesday	Feature engineering		Final Project Groups Assigned – start discussing you project outlines, structure, and version control plans with your group.
11/13/24	11	Wednesday	LAB	Meet with your Final Project Group	
11/14/24	11	Thursday	Cross Validation	HW10 – Due Friday 11:59pm	
			Quantifying Uncertainty		
11/19/24	12	Tuesday	Bootstrapping		
11/20/24	12	Wednesday	LAB	Meet with your Final Project Group	
11/21/24	12	Thursday	Applications	HW11 – Due Friday 11:59pm	
11/26/24	13	Tuesday	Get help on Final Projects	Work with your group on finishing your final project. Make sure you have addressed each part of the data science life cycle, address ethical issues, and have clear reproducibility in your results.	
11/27/24	13	Wednesday	THANKSGIVING		

11/28/24	13	Thursday	THANKSGIVING	Respond to and review peer work in written form. Practice communicating clearly between data science group members.	
12/03/24	14	Tuesday	Final Projects	Each group will present their results.	
12/04/24	14	Wednesday	NO LAB		
12/05/24	14	Thursday	Final Projects	Each group will present their results.	
12/11/24 – 12/14/24			FINAL EXAMS	Final projects are due 12/14/24	Complete a full cycle data science report that uses referencing to make the results reproducible and more formal writing to describe the process, ethics, results, and conclusions.