Data Science for Public Health 1 and 2

Course location / time: Classes will be done asynchronously. Live lectures are on Tuesday and Thursday at 8:30. Recordings of the lectures will be posted on the Teams site for the class. If you have a conflicting class(es), you should be able to take this class provided you can make some of the live lecture times.

Book

Live version https://smart-stats.github.io/ds4bio_book/book/_build/html/intro.html

Github repo https://github.com/smart-stats/ds4bio_book

Times of live sessions

8:30 AM - 9:50 AM

Live zoom link will be the same for all terms

https://jh.zoom.us/j/4109553504?pwd=amdidU82QTc2QTRmdkpDSkd3RU5pZz09

Instructor office hours

After class, in person first, then online at 10:30 on the Office hours Teams channel.

Instructor office hours canceled because of hard conflicts: 3/9/2023,

TA office hours:

Fridays 11:00 AM – 12:00 PM and Thursdays at 3:00, meet on the Office hours Teams channel

Course instructor: Brian Caffo

TAs:

- **Ziqi Fu**
- **Phyllis Wei**
- **Franklin Fuchs**

Please post all homework related questions on the Homework Teams Channel. For grading questions (such as regrade request and late submission), you can initiate private chats with the TAs. Admin issues should be directed to Dr. Caffo.

I cannot stress enough the need to be active on Teams for the class. Much of the discussion will occur there. New announcements will be made there. For important announcements I will @ the class.

Grading:

- 3rd term There will be once weekly online homework submissions due Mondays at 5:00 PM, unless it is a school holiday. These will be based on 100 points and your percentage of total points will represent 100% of the grade. The specific format of the evaluation will be communicated each week.
- 4th term same as 3rd term except HW and a final project that accounts for 40% of the grade (so 60% HW and 40% final project).
- Please put your name on your submitted materials. In the first text block of an ipynb file, the top lines of a text or MD file or in a top comment of a python code file. Please make sure that your submitted name agrees with your JHED name (or put both so we can properly attribute).
- If you need to submit an assignment late, do the following steps: 1) message Brian (on Teams) to get approval 2) when you submit the assignment to GH classroom message the TA grading that assignment **with the link to the repo**. Also, please follow up with the TA if you have not received a grade for a late submission within one week.
- Grades will be recorded and distributed on Teams. **Please make sure that you have received a grade for assignments submitted when we announce that the assignment has been graded.**
- Any late submission graded without approval will have a 10 point penalty.
- **All grades are final when grades are submitted**, and missing grades will be given a value of 0. The policy on grade changes is *"Once an instructor has submitted a grade to the Homewood Registrar, grade changes can be made only in the case of error in grading, calculation, or transcription."*
- Project due date in the 4th term is firm. Please consider that when working on it.
- Grades due third term 3/31/2023, 6/2/2023
- Note for graduating seniors, fourth term grades are due 5/19/2023. Because of this, the project deadline is firm for graduating seniors.

Homework rules

- To reiterate, put your name in the text of every file. (So, in the top text block of an ipynb or the top line of a md file or the top line of a text file …) Otherwise, we have no way of associating your github username with your actual name. Please make sure the name you give agrees with the one in the JHED system (or put both).
- Every evaluation is open book, open internet ….
- You may consult with fellow students on any assignment. However, please do so on Teams. All work submitted must be your own. Copying or plagiarism will be considered a breach of student conduct.
- If doing a team capstone project, separate github submissions must be made and the role of each team member delineated in the readme file.

- AI based submissions are not allowed. However, feel free to use AI (chat-gpt for example) as a resource.

Fourth term final project There is no project at the end of third term. However, the fourth term includes a final project.

- You should hand in the code for a data web app that you create based on an application of your interest.
- The project includes a 5 minute (maximum) video and one page (maximum) markdown readme file describing your project. Your MD readme file should include a link to the video that I can see or the video included in the repo.
- You name should be on your readme file, in the project video and in all code files as comments.
- This project can be done in a group up to size 5.
- Every project member must submit a repository.
- The project due date is extremely firm for graduating seniors.
- Here's the github classroom assignment link
  https://classroom.github.com/a/V-Ayz7c2
-    Grading Rubric
    - The project must involve data, analysis – it can be prediction, regression, whatever, it has to involve data at some level (project will be rejected if this condition is failed to be met)
    - There must be a dash / shiny or streamlit app (the project will be rejected if this condition is not met)
    - Each team must submit a write up in the form of your readme file, and a 5 minute video (the project will be rejected if this condition is not met).
    - Then the project will be graded on:
        - Is the app documented with instructions on use
        - Effort, points will be deducted for low effort projects (like a simple BMI calculator or something like that)
        - Correctness of results – was the data treated appropriately, was it cleaned appropriately, are there large conceptual or practical errors
        - Communication – was the app well presented, do I have a sense of what it's doing, is the web site and readme well presented?

Development environments
Students tend to like to use their own IDEs. However, we have a class jupyter server

https://jupyter.biostat.jhsph.edu/hub/login?next=%2Fhub%2F

Message Brian for access.

In addition github codespaces are setup as well

Finally, google colab is OK, but you should work towards one of these other solutions as you go

All video passwords are **ds4ph-2023**

Class	Topic	Video	Due Tuesday by 5PM	TA
1/24	Introduction, course structure	Link		
1/26	Git, github, github classroom	Link		
1/31	Basic python, IDEs	Link	HW 1	
2/2	Python programming, functions, python in practice	Link		
2/7	Data cleaning in pandas	Link	HW 2	
2/9	EDA, Interactive graphics, HTML, CSS, javascript	Link		
2/14	SQL, SQL in python	Link	HW 3	
2/16	Base R	Link		
2/21	Tidyverse, virtual environments	Link	HW 4	
2/23	R from python, python from R	Link		
2/28	Web scraping	Link		
3/2	Dash 1	Link		
3/7	DASH 2	Link	HW 5	
3/9	Dash 3	Link		
3/14	Binary classification	Link	HW 6	
3/16	Introduction to regression, **Last class 3rd term CANCELED**			
3/21	**Spring Break**			
3/23	**Spring break**			
3/28	Regression **First class 4th term**	Link		
3/30	Regression	Link		
4/4	Logistic regression	Link	HW 7 Due 4/5	
4/6	Multivariable GLMs			
4/11	Linear models, interpretation	Link	HW 8 Due 4/11	
4/13	Introduction to neural networks	Link		
4/18	Introduction to neural networks	Link	HW 9 Due 4/19	
4/20	Neural networks, fitting	Link		
4/25	Pytorch	Link	HW 10 Due 5/2	
4/27	Neural networks fitting example multivariable prediction	Link		
5/2	Convolutions, convolutional neural networks	Link	HW 11	
5/4	Adversarial neural networks	Link		
5/9	Recurrent neural networks	Link	HW 12	

5/11	Explainable AI	[Link]()		
5/16	AI Fairness, **Project due**		Project	
5/18	Buffer room and AI deployment			
	[Link to github classroom]()			