

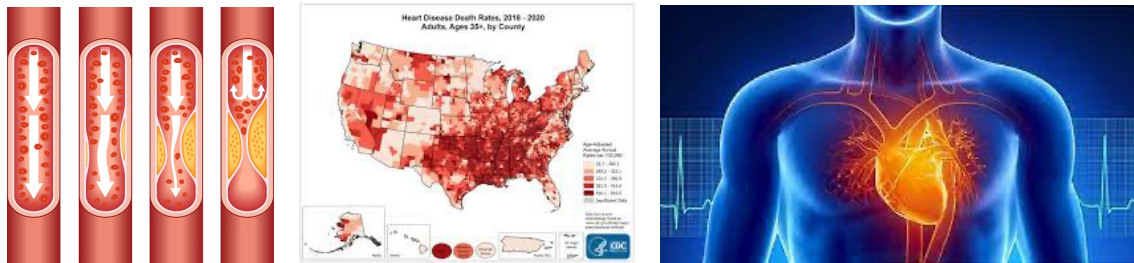
Fall 2022
Homework 6

Released: October 21, 2022

Due: October 28, 2022 @ 9pm ET

Late Due Date (1-48 hours late): October 30, 2022 @ 9pm ET

Homework 6: Exploring Heart Disease Data



Introduction

Heart disease in people of all genders is a leading cause of death in the United States, accounting for approximately 20% of all deaths. Heart disease is characterized by the build-up of plaques in the arteries feeding blood to the heart. Risk factors for heart disease include smoking, lack of exercise, excessive alcohol consumption, and an unhealthy diet. Increasingly, Artificial Intelligence is being incorporated into Electronic Medical Record systems to help doctors identify patients at risk of heart disease. In this assignment you will explore heart disease data to get an idea what working with this sort of data feels like.

Expectations

We expect you to write your code in a *modular style* using a well-defined collection of functions. The functions you implement are totally at your discretion. As always, you will be graded on a) correctness, b) code design and efficiency, and c) [documentation and style](#). You may not use Pandas, numpy, or any machine learning library such as Scikit-Learn that would trivialize some parts of the implementation. (You are allowed to use dictionaries and tuples, but their use is not necessary.)

Requirements

- 1. Read the data:** Read the heart disease data (`heart.csv`) into a list of lists. Convert numeric values to numeric data types as appropriate. Our heart disease dataset consists of 918 patients. Each patient has six attributes (Age, Sex Assigned at Birth, Type of Chest Pain, Resting Blood Pressure (BP), Cholesterol level, Maximum Heart Rate (HR)). The last column indicates whether the patient has heart disease (No=0, Yes=1).
- 2. Write code to answer the following questions and display their answers to the user:**
 - How many patients are there total?
 - How many patients with heart disease?
 - What is the average age of patients?

- What is the average age of patients who have heart disease?
- What is the average resting blood pressure of patients?
- What is the average resting blood pressure of patients who have heart disease?

Make sure to use the `print` function to display your well-labeled answers to the user. For example, your output should look something like:

```
Number of patients: 918
Number of patients with heart disease: 508
Average age: ...
Average age of patients with heart disease: ...
...
```

Hint 1: you may find it helpful in your program to find a filtered list of lists that contains only rows with patients who do have heart disease.

Hint 2: think about how you can design functions so that you can re-use code to answer some of the different questions!

Hint 3: look at the functions that we've written together in class to manipulate data that may be helpful to you here.

3. Generate a scatter plot of Cholesterol (y axis) vs. Maximum Heart Rate (x axis). Color code the markers so that patients with heart disease are red, and patients with no heart disease are blue. Use the true value of heart disease (the last column) from the data set to decide on color.

The plotting command that you'll need is:

```
plt.scatter(maximum_hr, cholesterol, marker = '.', color=colors)
```

where `maximum_hr` is a list of values in the maximum heart rate column, `cholesterol` is the list of values from the cholesterol column, and `colors` is a *list* of color codes corresponding to your data points, for example: `['b', 'r', 'b', 'r', 'b', 'b', ...]`. Note that all three lists (`maximum_hr`, `cholesterol`, and `colors`) must have the same length!

A legend is not required for this graph.

Program Structure

Reminders (we copy + pasted much of this from the style guide!) about what your program should look like:

```
""" Your Name
    course
    assignment
    date

    description (1 - 2 sentences)
```

```
    your copy+pasted program output here
"""

import matplotlib.pyplot as plt
# other imports here

# constants go here

def example_function1(example_parameter):
    """
    Your function comment here. Must include all three of:
    what does it do, what parameters does it take, what value does it return
    """
    # your function code here

def example_function2(example_parameter1, example_parameter2):
    """
    Your function comment here
    """
    # your function code here

def main():
    # Your code goes here!
    # Your code goes here!

if __name__ == "__main__":
    main()
```

What to Submit

Submit your program (`heart_explore.py`), your visualization (`heart.png` or `heart.pdf`). **In your program header (after your 1 - 2 sentence description), include the program text output reporting the answers to all the questions from part 2.**

Please Read: It is your responsibility to verify that your program runs (by examining the autograder output), and that ALL required files have been submitted to gradescope. If you forget to submit your code, your visualization, or your program doesn't run, your grade will be affected. If you *resubmit* your assignment for any reason, your resubmission must include ALL required files. Files from previous submissions are deleted from gradescope and not accessible to the graders! Resubmissions after the due date (but before the late due date) will receive the full late penalty.

No homework files will be accepted for grading or regrading after the late due date!