clinical_app_starter

June 27, 2020

0.1 Part 2: Clinical Application

0.1.1 Contents

Fill out this notebook as part 2 of your final project submission.

You will have to complete the Code (Load Metadata & Compute Resting Heart Rate) and Project Write-up sections.

- Section 0.1.3 is where you will implement some parts of the **pulse rate algorithm** you created and tested in Part 1 and already includes the starter code.
- Section 0.1.3 - These are the imports needed for Part 2 of the final project.
  - glob
  - os
  - numpy
  - pandas
- Section 0.1.3
- Section 0.1.3
- Section 0.1.3
- Section **??**
- Section **??** to describe the clinical significance you observe from the **pulse rate algorithm** applied to this dataset, what ways/information that could improve your results, and if we validated a trend known in the science community.

0.1.2 Dataset (CAST)

The data from this project comes from the Cardiac Arrythmia Suppression Trial (CAST), which was sponsored by the National Heart, Lung, and Blood Institute (NHLBI). CAST collected 24 hours of heart rate data from ECGs from people who have had a myocardial infarction (MI) within the past two years.[1] This data has been smoothed and resampled to more closely resemble PPG-derived pulse rate data from a wrist wearable.[2]

1. **CAST RR Interval Sub-Study Database Citation** - Stein PK, Domitrovich PP, Kleiger RE, Schechtman KB, Rottman JN. Clinical and demographic determinants of heart rate variability in patients post myocardial infarction: insights from the Cardiac Arrhythmia Suppression Trial (CAST). Clin Cardiol 23(3):187-94; 2000 (Mar)

2. **Physionet Citation** - Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals (2003). Circulation. 101(23):e215-e220.

0.1.3 Code

Imports When you implement the functions, you'll only need to you use the packages you've used in the classroom, like Pandas and Numpy. These packages are imported for you here. We recommend you don't import other packages outside of the Standard Library , otherwise the grader might not be able to run your code.

```
In [1]: import glob
        import os

        import numpy as np
        import pandas as pd
```

Load the dataset The dataset is stored as .npz files. Each file contains roughly 24 hours of heart rate data in the 'hr' array sampled at 1Hz. The subject ID is the name of the file. You will use these files to compute resting heart rate.

Demographics metadata is stored in a file called 'metadata.csv'. This CSV has three columns, one for subject ID, age group, and sex. You will use this file to make the association between resting heart rate and age group for each gender.

Find the dataset in `../datasets/crisdb/`

```
In [2]: hr_filenames = glob.glob('/data/crisdb/*.npz')
```

Load Metadata Load the metadata file into a datastructure that allows for easy lookups from subject ID to age group and sex.

```
In [3]: metadata_filename = '/data/crisdb/metadata.csv'
        metadata = None
        # Load the metadata file into this variable.
        with open(metadata_filename, 'r') as f:
            metadata = pd.read_csv(f)
```

Compute Resting Heart Rate For each subject we want to compute the resting heart rate while keeping track of which age group this subject belongs to. An easy, robust way to compute the resting heart rate is to use the lowest 5th percentile value in the heart rate timeseries.

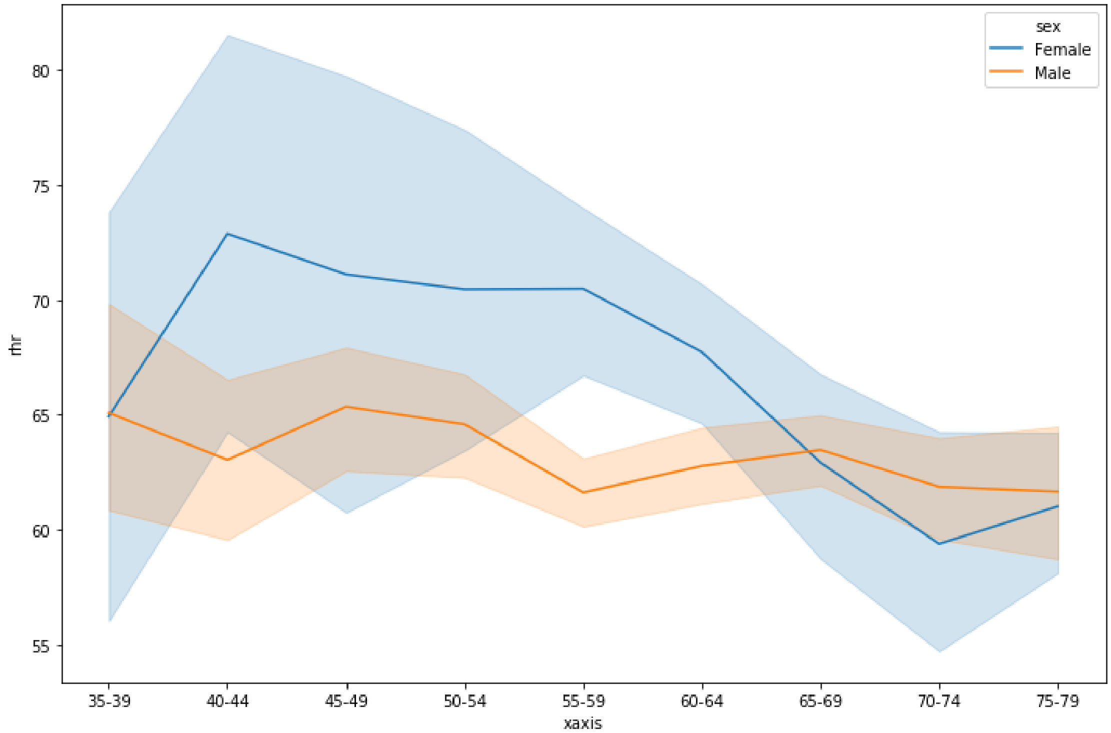
```
In [4]: def AgeAndRHR(metadata, filename):

            # Load the heart rate timeseries
            hr_data = np.load(filename)['hr']
```

```python
    # Compute the resting heart rate from the timeseries by finding the lowest 5th perce
    rhr = np.percentile(hr_data, 5)

    # Find the subject ID from the filename.
    subject = filename[13:-4]

    # Find the age group for this subject in metadata.
    age_group = metadata.loc[metadata['subject'] == subject, 'age'].item()

    # Find the sex for this subject in metadata.
    sex = metadata.loc[metadata['subject'] == subject, 'sex'].item()

    return age_group, sex, rhr

df = pd.DataFrame(data=[AgeAndRHR(metadata, filename) for filename in hr_filenames],
                  columns=['age_group', 'sex', 'rhr'])
```

Plot Resting Heart Rate vs. Age Group We'll use seaborn to plot the relationship. Seaborn is a thin wrapper around matplotlib, which we've used extensively in this class, that enables higher-level statistical plots.

We will use lineplot to plot the mean of the resting heart rates for each age group along with the 95% confidence interval around the mean. Learn more about making plots that show uncertainty here.

```python
In [6]: import seaborn as sns
        from matplotlib import pyplot as plt

        labels = sorted(np.unique(df.age_group))
        df['xaxis'] = df.age_group.map(lambda x: labels.index(x)).astype('float')
        plt.figure(figsize=(12, 8))
        sns.lineplot(x='xaxis', y='rhr', hue='sex', data=df)
        _ = plt.xticks(np.arange(len(labels)), labels)
```



0.1.4 Clinical Conclusion

Answer the following prompts to draw a conclusion about the data. > 1. For women, we see .... > 2. For men, we see ... > 3. In comparison to men, women's heart rate is .... > 4. What are some possible reasons for what we see in our data? > 5. What else can we do or go and find to figure out what is really happening? How would that improve the results? > 6. Did we validate the trend that average resting heart rate increases up until middle age and then decreases into old age? How?

Your write-up will go here...

1) From the above plot for women, we see an increase in resting heart rate from 35-39 year old age group to the 40-44 year old age group, followed by a decrease in resting heart rate as women as they get older, with the execption of the 55-59 year old age group

2) For men, we see the resting heart rate is more consistent around 65 bpm across all age groups with slight increases in the 35-39, 45-49, 65-69 age groups followed by drops in the following age groups

3) In comparison to men, women's heart rate is much higher in the 35-39 through 65-69 age groups and then is lower from the 65-69 age group and beyond.

4) What are some possible reasons for what we see in our data: Some possible reasons could be physical/metabolic differences between the genders, different levels of activity or different physchological factors such as stress levels or different life goals between the genders.

5) What else can we do or go and find to figure out what is really happening? How would that improve the results? Perhaps a subsequent study involving the subjects from this study where the subjects are asked to identify major stress factors in their lives, lifestyle habits or major goals in their lives.

6) Did we validate the trend that average resting heart rate increases up until middle age and then decreases into old age? How? For women the trend depicted in the above plot indicate that heart rate increases up until middle age and then decreases into old age, but for men the plot is a little less conclusive in depicting heart rate increasing up until middle age, because it shows that the heart-rate for the 35-39 year old age group is higher than the heart rate for the 40-44 year old age group. However after the 45-49 year old age group, hear rate does decrease for men.

```
In [ ]:
```