A Study of Factors Affecting Heart Disease Mortality Rate in the United States

Ashraful Islam, Nicholas Luczak, Saswata Paul, Yiyun su
{islama6, luczan, pauls4, syu4}@rpi.edu
Rensselaer Polytechnic Institute, Tetherless World Constellation, Troy, NY, United States

Abstract

In this poster, we present our investigation of the factors that may be responsible for coronary heart disease in the United States. We perform two types of analysis - an analysis of coronary heart disease and median household income for New York State, and an analysis of coronary heart disease and social determinants for the entire United States. We obtain public domain data from www.cdc.gov and www.data.gov to perform our analysis. Our preliminary analysis shows interesting patterns between coronary heart disease mortality and social factors. Moreover, we train a machine learning model to see if it is possible to correctly predict coronary heart disease from the various factors, which shows around 67% accuracy in predicting heart disease mortality from social vulnerability factors.

Workflow

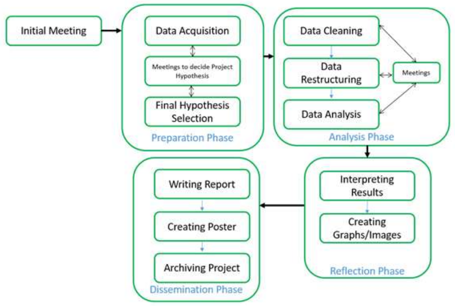

Figure 1. Workflow of our Project

Research Questions

1. How is heart disease mortality rate connected to the median income in New York State?
2. How is heart disease mortality rate connected to social vulnerability in the United States?
3. How is heart disease mortality rate connected to ethnicity in New York?

Data Format

NYSERDA Low-to-Moderate-Income New York State Census Population Analysis Dataset
1. Collected from: catalog.data.gov
2. Type of file: CSV
3. Publisher: data.ny.gov
4. Maintainer: NY Open data
5. Maintainer email: openny@nyserda.ny.gov
6. Time: 2013-2015

Heart Disease Mortality Data Among US Adults (35+) by State/Territory and County
1. Collected from: catalog.data.gov
2. Type of file: CSV
3. Publisher: Centers for Disease Control and Prevention
4. Maintainer: NY Open data
5. Maintainer email: openny@nyserda.ny.gov
6. Time: 2013-2015

Social Vulnerability Index 2012 - 2014
Collected from: svi.cdc.gov
1. Type of file: CSV
2. Publisher: cdc.gov
3. Maintainer: Centers for Disease Control and Prevention
4. Maintainer email: dhdsprequests@cdc.gov

Sponsors:

Data Analysis

Relationship Between Heart Disease Rate and Median Income in New York State

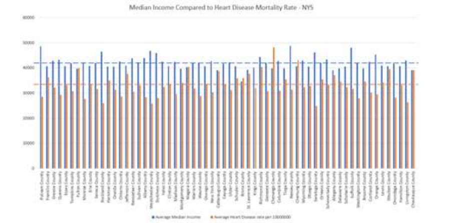

Figure 2. Median income vs Heart Disease Mortality Rate

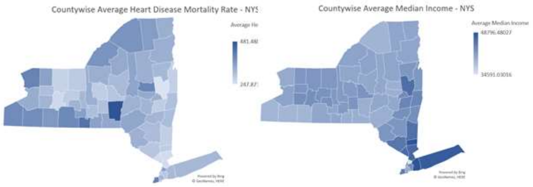
Countywise Average Heart Disease Mortality Rate - NYS
Countywise Average Median Income - NYS

Figure 3. Median Income and Heart Disease Mortality Rate by County

- Heart disease mortality rate is usually inversely related to median income in New York.

Relationship Between Heart Disease Rate and Social Vulnerability

- 15 US census variables to determine the Social vulnerability of each county
- Social vulnerability is a good indicator of Heart disease
- The higher the overall vulnerability of a community the less likely they are to live a healthy lifestyle; therefore, their likelihood of coronary-related mortalities sees an increase as a result.

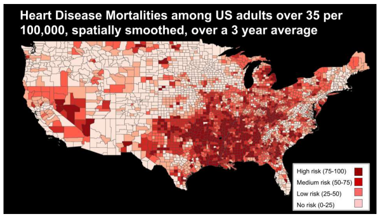
Heart Disease Mortalities among US adults over 35 per 100,000, spatially smoothed, over a 3 year average
High risk (75-100)
Medium risk (50-75)
Low risk (25-50)
No risk (0-25)

Figure 4. Heart Disease Mortality Rate in the US

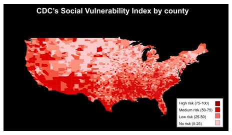

Figure 5. Overall social Vulnerability Index

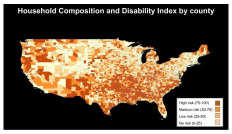

Figure 6. Household Composition and Disability

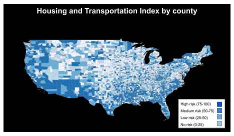

Figure 7. Housing and Transportation

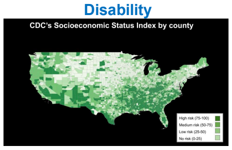

Figure 8. Socioeconomic Status

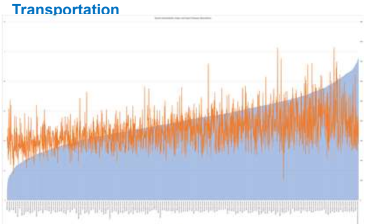

Figure 9. Social Vulnerability and Heart Disease Mortality

Resources:
www.cdc.gov, www.catalog.data.gov, www.data.ny.gov, www.svi.cdc.gov
https://matplotlib.org/3.1.1/index.html, https://seaborn.pydata.org/, https://scikit-learn.org/stable/, https://pandas.pydata.org/

Relationship Between Heart Disease Rate and Ethnicity in New York State

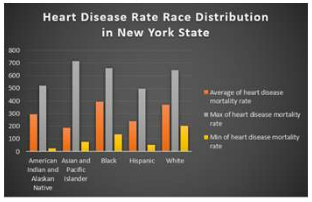

Figure 10. Ethnicity and Heart Disease Mortality

Learning Heart Disease Mortality from Social Vulnerability Factors

12 SVI factors were used

Correlation between heart disease mortality and SVI factors

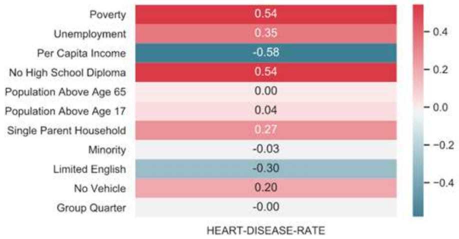

Figure 11. Correlation of Heart Disease Mortality with various SVI

- Heart-disease-death-rate highly correlates with **poverty** and **no-high-school-diploma-household**
- Heart-disease-death-rate inversely correlates with **per-capita-income** and **limited-English-speaking-ability**

Machine Learning

- Convert heart rate mortality values to categorical values:
  - Low: 0 – 320
  - Medium: 320 – 420
  - High: 420 – 800
- Training data: 777 Samples
- Testing data: 333 Samples
- Each input feature has 12 dimension

SVM Classification
- 3 class SVM with RBF kernel
- Test Accuracy: 66.36%

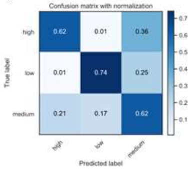

Figure 12. SVM Confusion Matrix

Decision Tree Classifier
- Test Accuracy: 52.60%

K-NN Classifier
- 10 nearest neighbors
- Test Accuracy: 66.36%

SVM and K-NN seem to work best for this problem

Conclusion

1. Heart disease mortality rate is inversely proportional to median income in NY
2. Social vulnerability is a good indicator of heart disease mortality rate in the US
3. In NY, Asian and Pacific Islander has a lower chance of getting heart disease based on their less population in New York state.

Scan the QR code check out our results!