SOCIAL DETERMINANTS OF HEALTH AND CARDIOVASCULAR DISEASES

Max Shalom, Yeshivah of Flatbush High School | **Mentors:** Yuan Zhao & Vishwali Mhasawade | **Professor:** Rumi Chunara, Chunara Lab

The Pinkerton Foundation

Center for K12 STEM Education

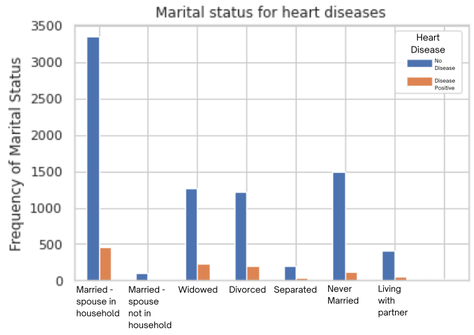
NYU TANDON SCHOOL OF ENGINEERING

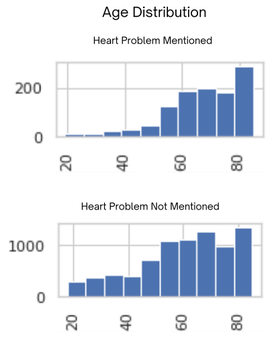
Introduction

- Populations in low socioeconomic communities and geographies usually have a higher burden of cardiovascular diseases (CVD).
- Using data from the National Health Interview Survey, we can pinpoint at which factors such as education level, poverty, and malnutrition, are most relevant in contracting a CVD.

Descriptive Statistics

Using pandas and Matplotlib in Python, I created charts to help portray the results of my research. Above are bar graphs and histograms of age and marital status and to the right is an except from a large table I created with over 10 variables and the percentage of how many people within these variables answered these questions in the data survey. Using these, we can hope to figure out the most relevant variables that affect one's chances of contracted a CVD such as heart disease.





Data Table

After analyzing many different types of social health factors, I found that some factors were stagnant across disease responses while others like age and highest education level were most important.

	Mentioned Heart Problem (LAHCA7==1) (N=)	Did Not Mention Heart Problem (LAHCA7==2) (N=)	Not Clear/Don't Know (N=)
Age (AGE_P) (mean, median)	67.62, 69	60.27, 62	62.06, 65
Marital Status (R_MARITL)(n,%): Married, Single, Other	51%, 26%	54%, 16%	66%, 9%
Alcohol Abuse (LAHCA29_)(n,%): Mentioned, Not mentioned	0%, 100%	0.2%, 99.8%	0%
Health Insurance (COVER)(n,%) Under 65 years old: Private, Medicaid, Other, Uninsured, Don't know	30.2%, 39.1%, 21.5%, 8.9%, 0.2%	39.7%, 33.9%, 20%, 9.7%, 0.8%	30.2%, 39.1%, 21.5%, 8.9%, 0.2%
Health Insurance (COVER65)(n,%) 65 years old and older: Private, Dual eligible, Medicare Advantage, Medicare Only, Other, Uninsured	36%, 14.2%, 18.6%, 17.6%, 13.6%, 0.3%	35.2%, 12.7%, 22.6%, 16.5%, 12.2%, 0.6%, 0.2%	40.7%, 37%, 14.8%, 18.5%, 18.5%, 0%, 3.7%
Private Health Insurance (HIKINDNA) (n,%): Yes - information, Yes - but no information, Refused, Don't know	32.8%, 67.1%, 0.1%	37.5%, 62%, 0.1%, 0.4%	26.4%, 68%, 5.7%
Income (ERNYR_P)(median)	$25,000–34,999	$25,000–34,999	$25,000–34,999
Born in the US (PLBORN)(n,%)): Yes, No	89%, 11%	88%, 12%	77%, 23%
How well is English (ENGLANG)(median)	Very Well	Very Well	Very Well
Work full time (WRKFTALL)(n,%): Yes, No, Don't Know	11.6%, 88.4%,	13.1%, 86.7%, 0.1%	22.2%, 77.7%
Highest Education (EDUC1)(n,%): High school and under College, Graduate, and Above	26.7%, 39.9%, 44.9%	23.4%, 41.6%, 48.7%	53%, 39.9%, 10.5%

Logistical Regression

We can assess the association between individual and community-level social determinants and CVD, including heart disease, and use regression and machine learning methods to predict CVD using social determinants.

```
------------------------------------------------------------------------
              Coef.    Std.Err.      z      P>|z|     [0.025    0.975]
------------------------------------------------------------------------
ENGLANG      0.0202      0.0474    0.4267    0.6696   -0.0727    0.1132
PSSRR       -0.0723      0.0499   -1.4491    0.1473   -0.1701    0.0255
EDUC1       -0.0090      0.0040   -2.2592    0.0239   -0.0168   -0.0012
CITIZENP    -0.0913      0.1415   -0.6456    0.5185   -0.3687    0.1860
PLBORN      -0.2753      0.1290   -2.1336    0.0329   -0.5281   -0.0224
HIKINDNA     0.0629      0.0490    1.2850    0.1988   -0.0330    0.1589
LAHCA29_    -1.3984      0.1428   -9.7949    0.0000   -1.6782   -1.1185
R_MARITL     0.0184      0.0143    1.2878    0.1978   -0.0096    0.0464
AGE_P        0.0290      0.0024   11.9857    0.0000    0.0243    0.0338
MRACRPI2     0.0306      0.0135    2.2579    0.0240    0.0040    0.0572
SEX         -0.3917      0.0658   -5.9491    0.0000   -0.5207   -0.2627
WTFA        -0.0000      0.0000   -0.8577    0.3911   -0.0001    0.0000

Factors with a P>|z| value of <0.05 can significantly impact one's
chances of contracting a CVD, since negative coefficients decrease the
risk and positive ones raise.
========================================================================
Accuracy of logistic regression classifier on test set: 0.89
```

Future Research

Using the newfound research results, the next step is to apply deep data analysis and machine learning techniques to estimate and predict efficient ways to prevent the contraction of these CVDs, using the social data of who has already contracted them.