

CANCER Heart Disease

**By Yasmin Badri,
Kennedy Bagnol and
Cindy Davis**

Problem Statement

How accurately can Cancer and Heart Disease Mortality rates be predicted given the data below for each state in the US?



Environment

- Air Quality
- Water Quality
- Sunshine
- Temperature



Socio-Economic

- Income
- Education
- Government Assistance
- Poverty



Health

- Obesity
- Physical Activity
- Smoking
- Excess Alcohol



Access

- Stores
- Restaurants
- Local Farmers
- Fitness Centers

Data Collection

Initial Data

We started by collecting health data we thought could contribute to Cancer and Heart disease. This included air quality, lead, general health data such as obesity and diabetes.



Later Data

As we modeled, we continually searched for more data that we thought might contribute to our model. We added healthcare coverage, socioeconomic data, and energy consumption.

Datasets Used

AQI

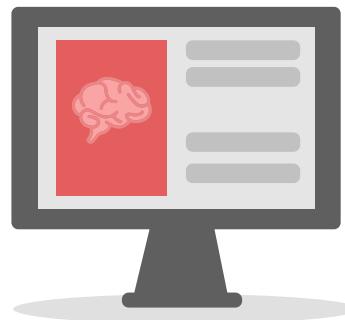
The overall air quality for each state.

Health Care Coverage

Tracked the uninsured, insured through employer, privately insured, government insurance

Health Data

Obesity rate, diabetes rate, nutrition scores, activity scores, etc.



And Much More!

Smokers

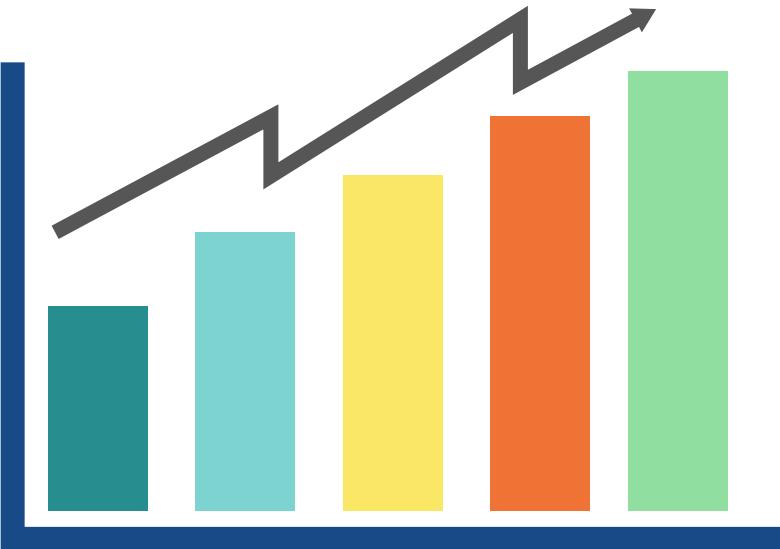
The percent of smokers for each state

Socioeconomic Data

Percent on WIC, percent educational attainment, poverty rate, median income

General US Data

Sunshine, happiness, college graduation rate



Interesting Note

- One aspect about our modeling is that our dataset was quite different from what we have worked with previously.
- Rather than using our entire dataset to predict one target, we instead were predicting a value in a row based on the features in the row.
- This was because we were trying to predict a death rate per state, not an overall death rate given a series of variables.
- This could be a possible explanation for some of the results.

MODELING PROCESS

HIGHEST CORRELATING FEATURES

Used a function to extract the highest correlated features with our target.



POLYNOMIAL FEATURES & PCA

Engineered new features using polynomial features and/or principal component analysis



MACHINE LEARNING

Instantiated several different models to find our best scores including Linear Regression, LassoCV, Bagging and Random Forest Regressor

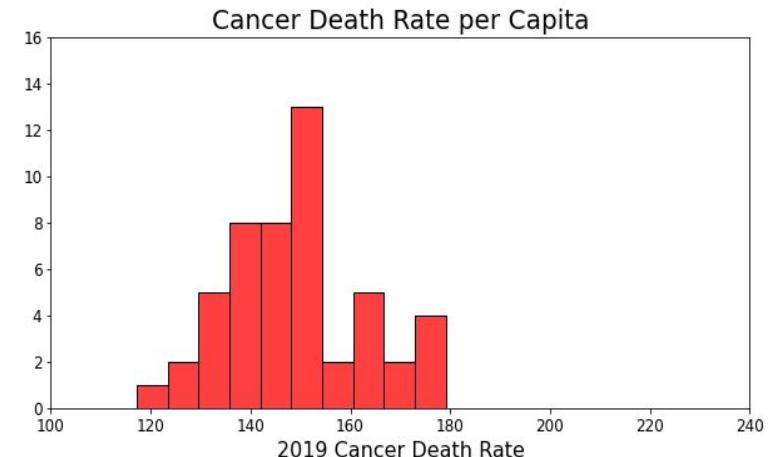
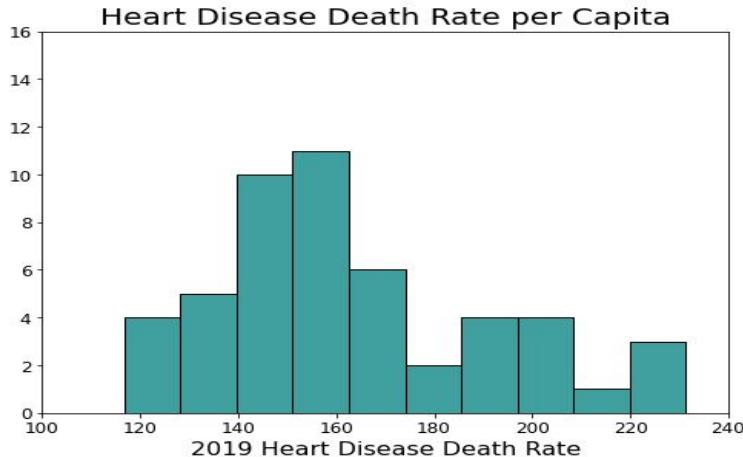


PREDICTIONS

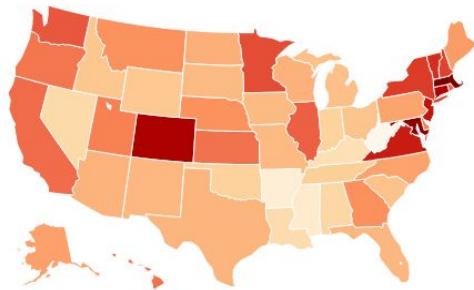
Once we had our best model, we used it to make our target predictions



PREDICTION TARGETS



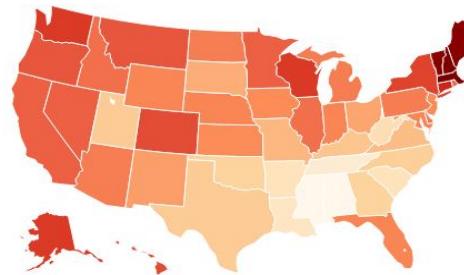
Percent Education Attainment by State



Mean

35
30
25
20

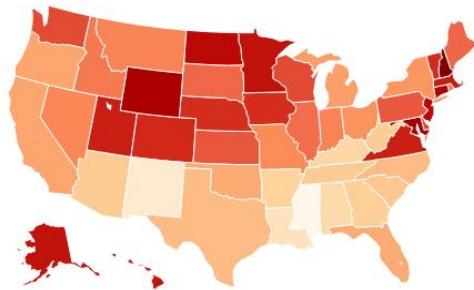
Percent Non-religious by State



Mean

60
50
40
30

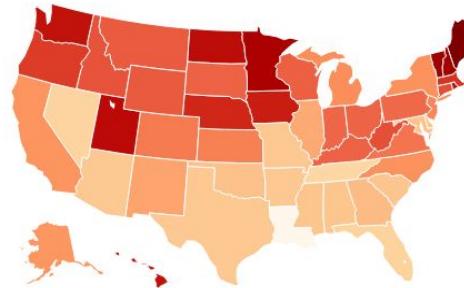
Percent Above Poverty Rate Mean by State



Mean

90
88
86
84
82
80

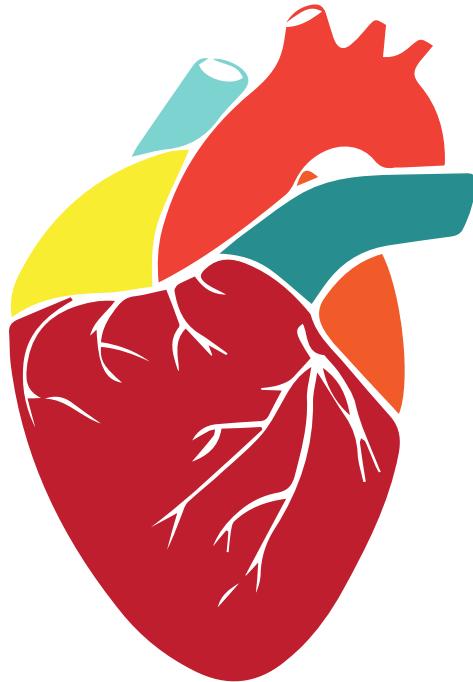
Percent Peace Index Mean by State



Mean

70
60
50
40
30
20

Heart Disease Facts



Leading cause of death in the US among adult men and women.



659,000 people in the US die from heart disease each year, 1 in every 4 deaths.

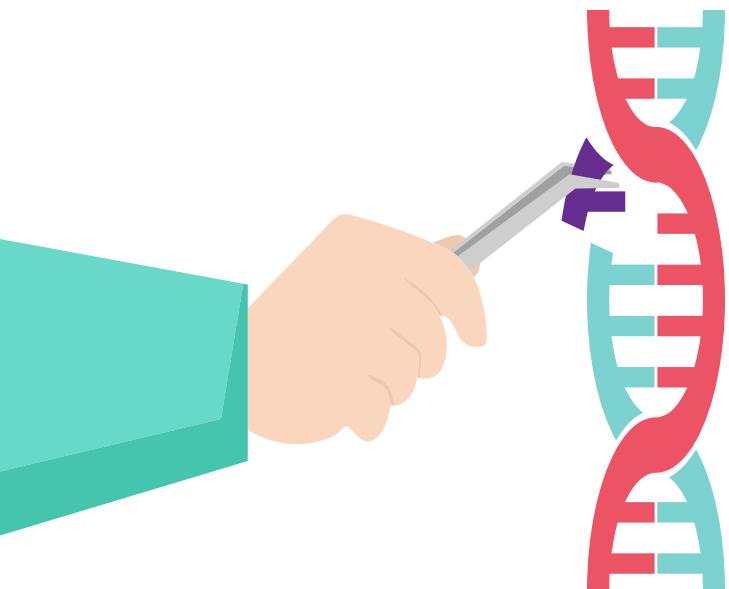


Costs the US over \$360 billion per year.



Thankfully year-over-year rates are trending downwards

Cancer Facts



Leading cause of death in the US among adult men and women after heart disease.



599,000 people in the US die from cancer each year.



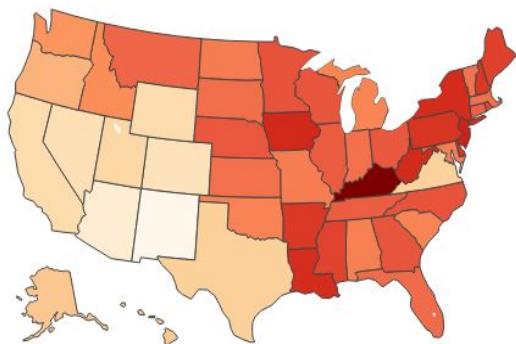
Costs the US over \$208 billion per year.



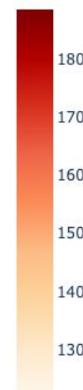
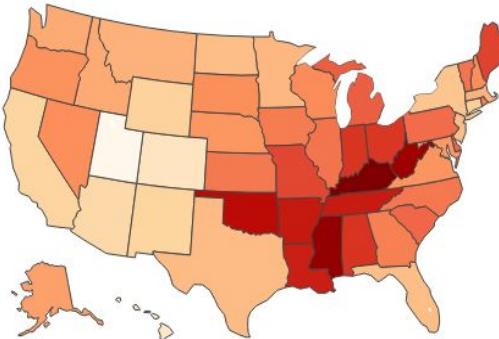
Thankfully year-over-year rates are trending downwards

Cancer Rates

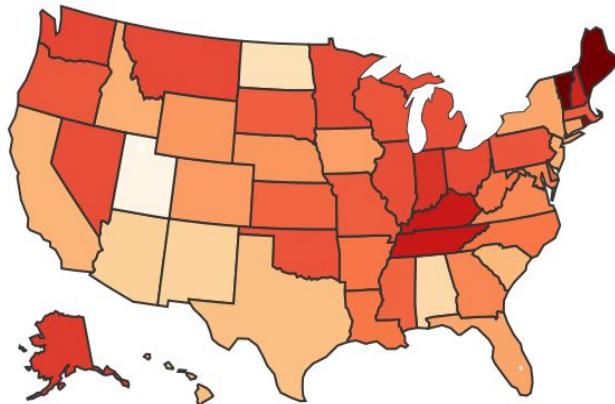
Cancer Incidence Rate by State



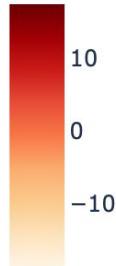
Cancer Death Rate by State



Cancer Results



Heatmap of Diff



We found that predicting cancer was easier than predicting heart disease.



Our best performing model was the LassoCV model with a training score of .72 and a testing score of .75.

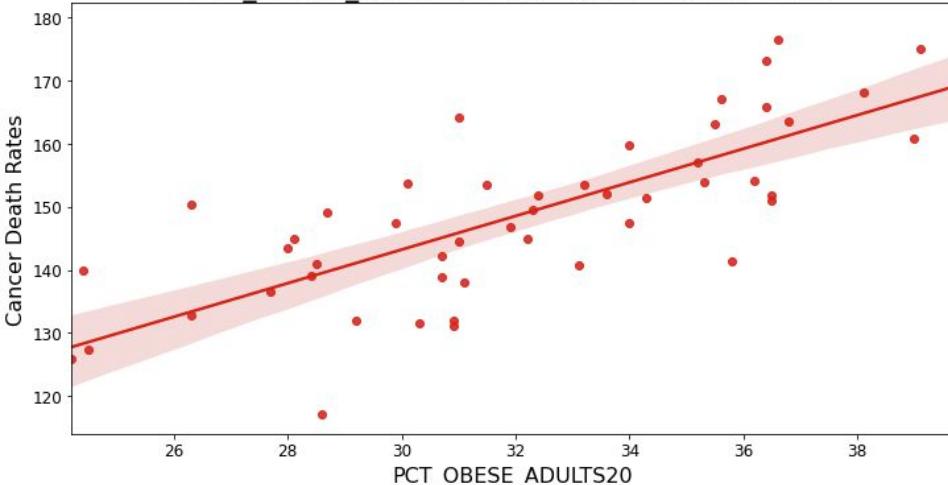


PCA analysis actually did not help our LassoCV model much, only raising the testing score by .03. However, it helped other models greatly.

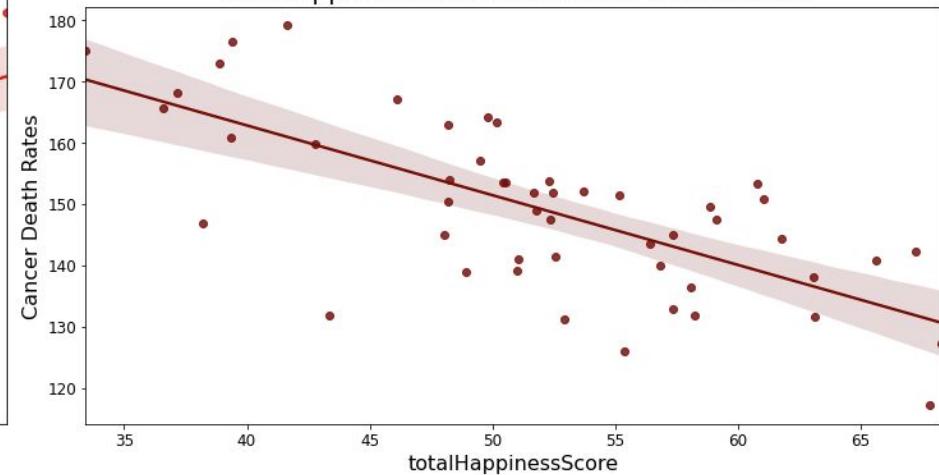


The highest correlating features were percent obesity and diabetes, as well as the happiness score and nutrition and physical activity score.

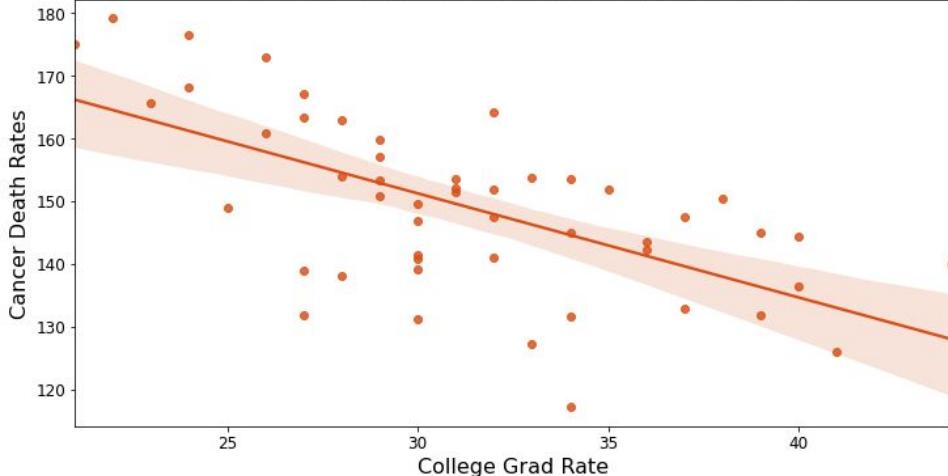
PCT_OBESE_ADULTS20 with Cancer Death Rates



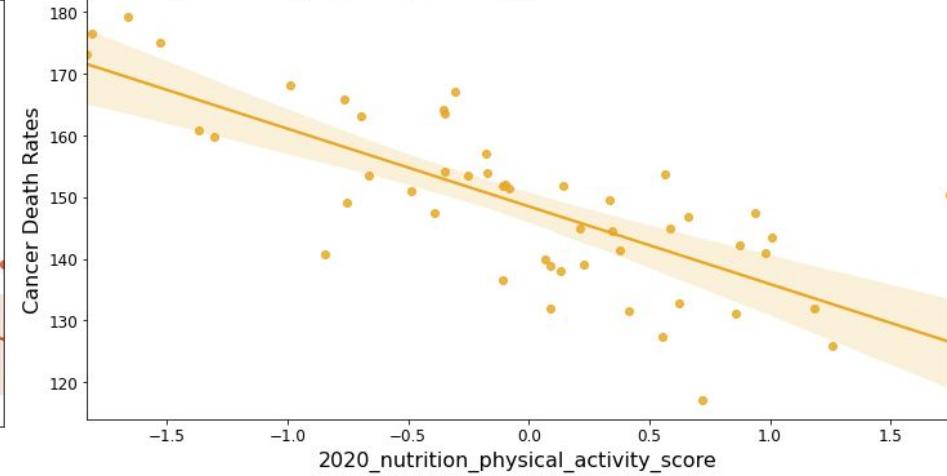
totalHappinessScore with Cancer Death Rates

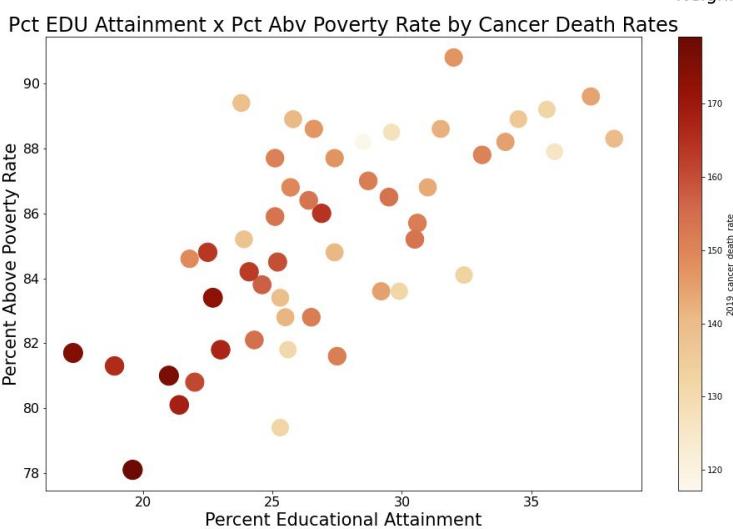
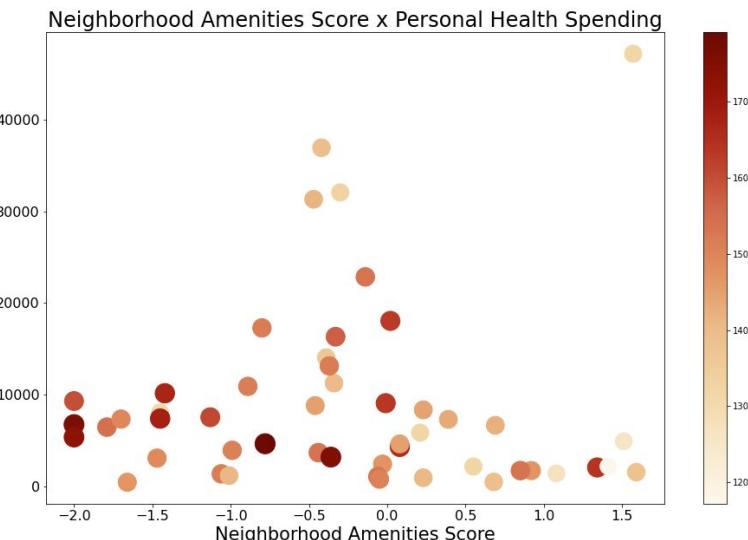
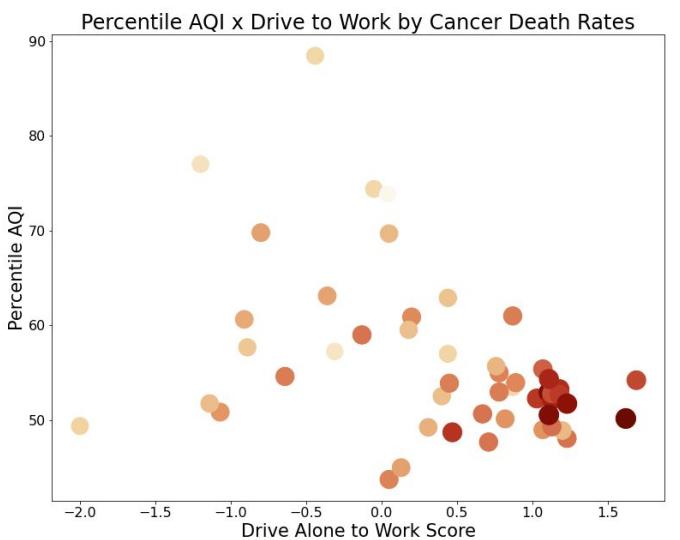


College Grad Rate with Cancer Death Rates

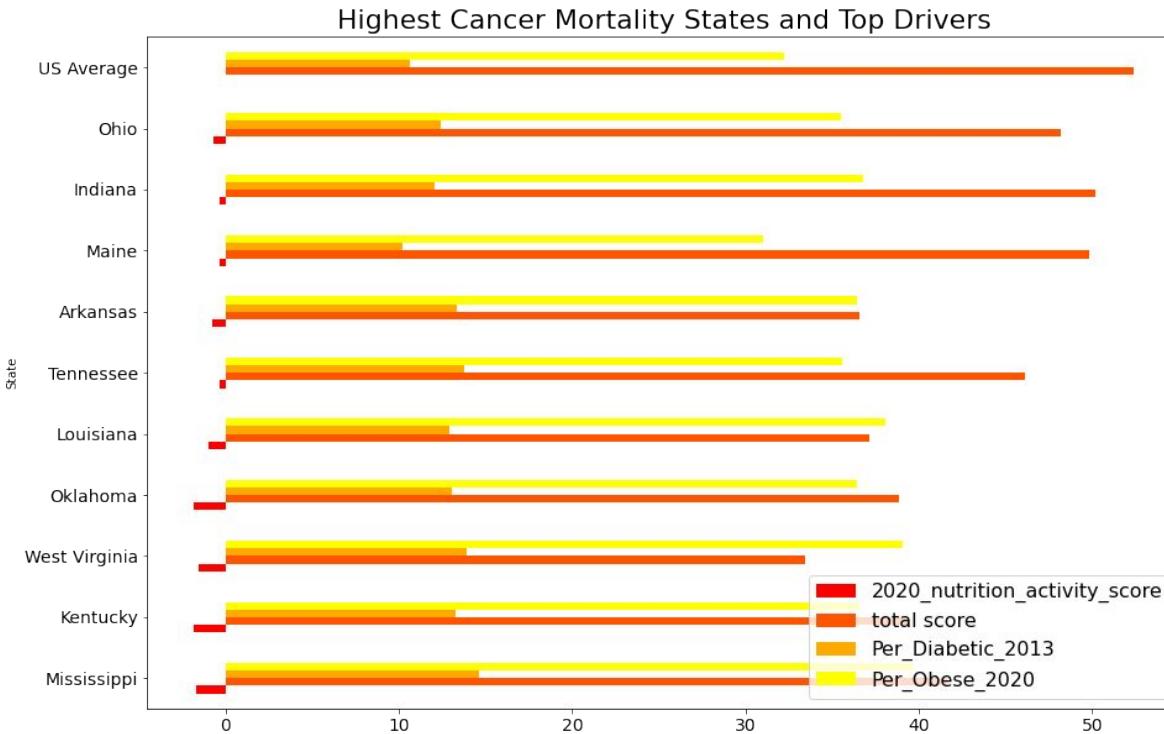


2020_nutrition_physical_activity_score with Cancer Death Rates





States with highest Cancer Mortality Rates



Nutrition and physical activity score lower than the US average



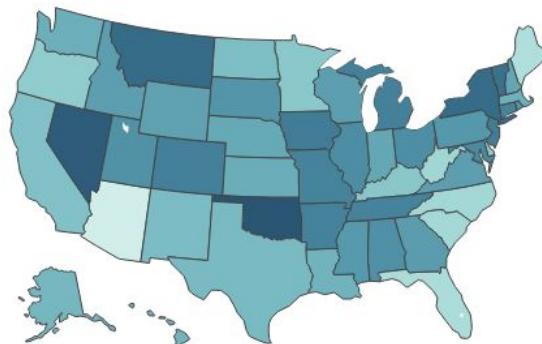
Obesity and diabetes rates higher than the US average



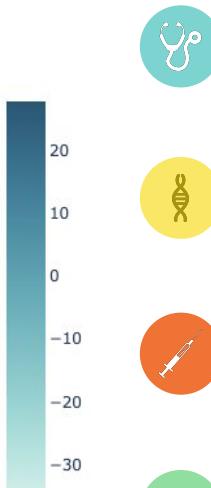
Lower total happiness scores than the US average

Heart Disease Results

Prediction Errors for Heart Disease by State



Train rmse: 14.51671577346816
Test rmse: 12.506243762207557



Heart Disease was difficult to predict.



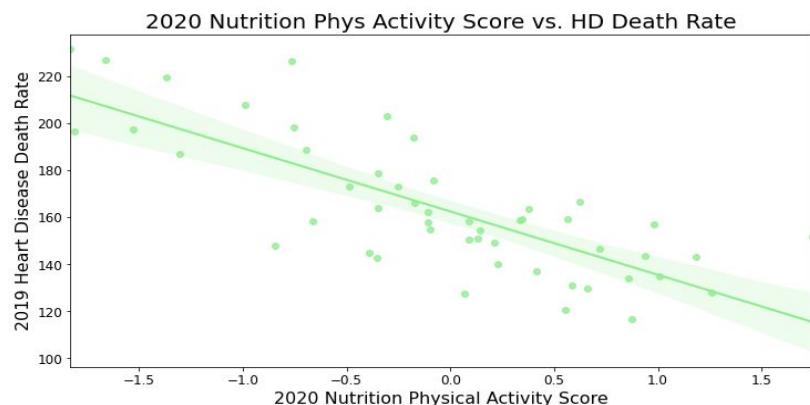
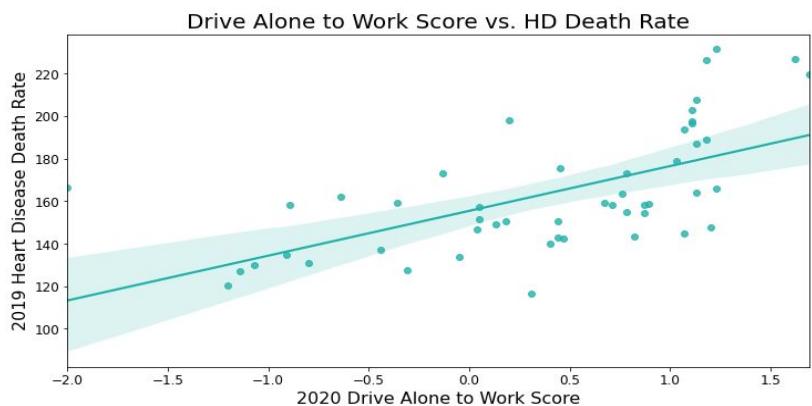
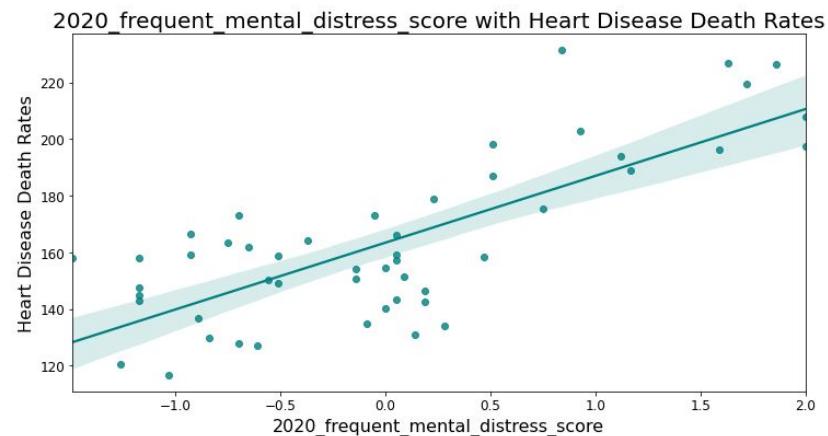
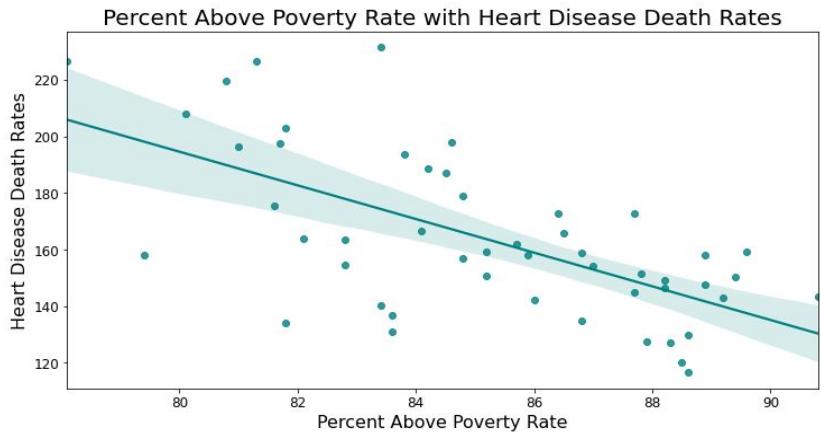
Best performing model was Linear Regression with principal component analysis (PCA) with a training score of 0.78 and testing score 0.46 with CV of 0.66,

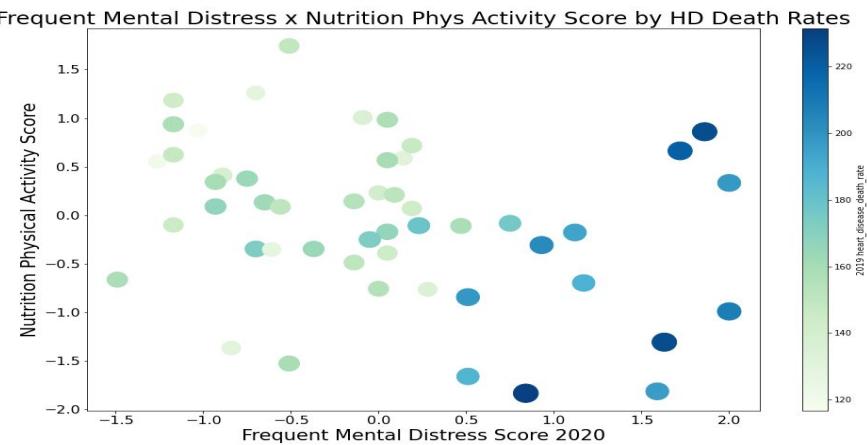
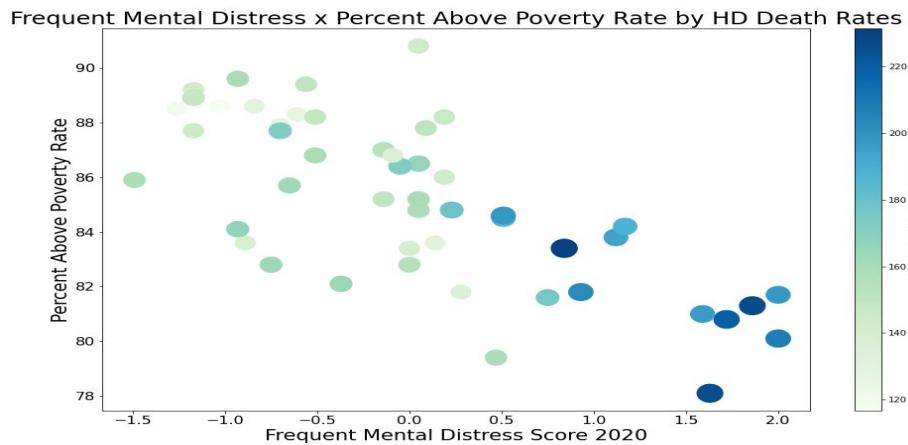
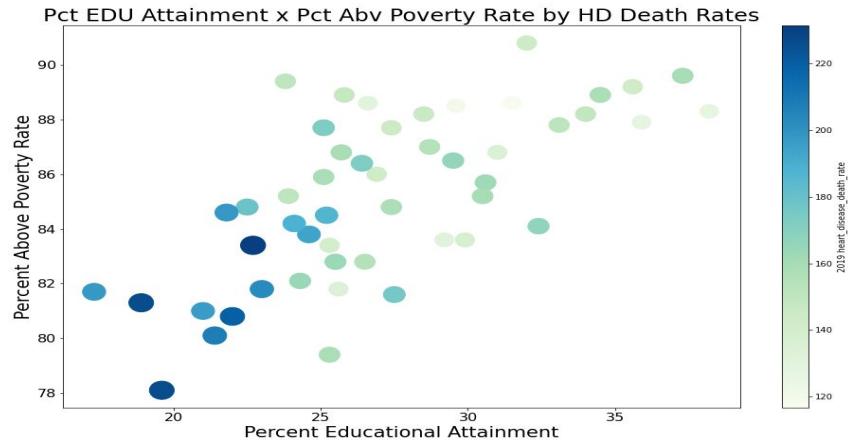
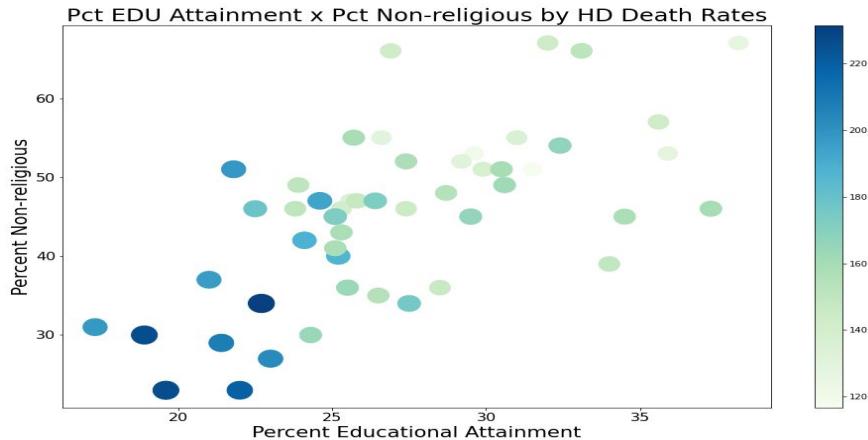


Nevada and Oklahoma had the largest error margin in predictions

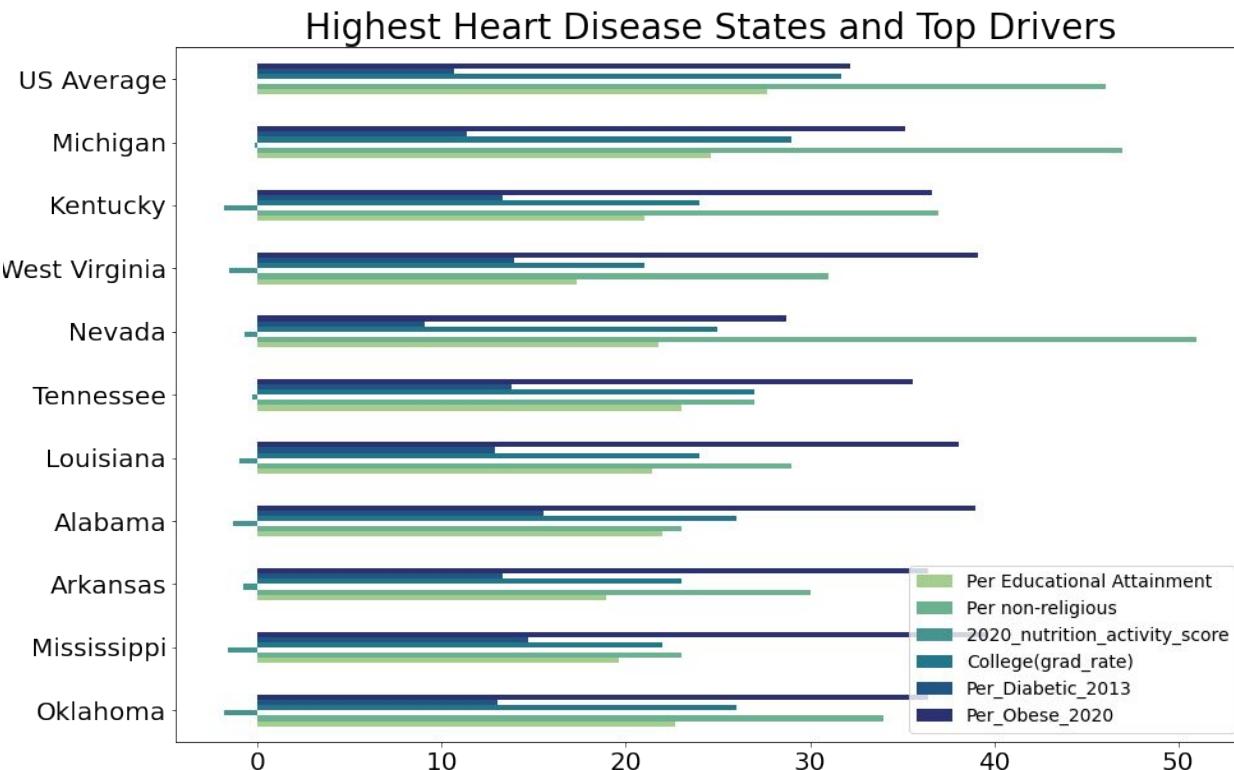


The highest correlating features were percent cholesterol, percent obesity, percent diabetes, frequent mental distress, nutrition and physical activity score, and WIC participation rates





States with Highest Heart Disease Mortality Rates



Nutrition and physical activity scores lower than the US average



Higher adult obesity and diabetic rates than the US average

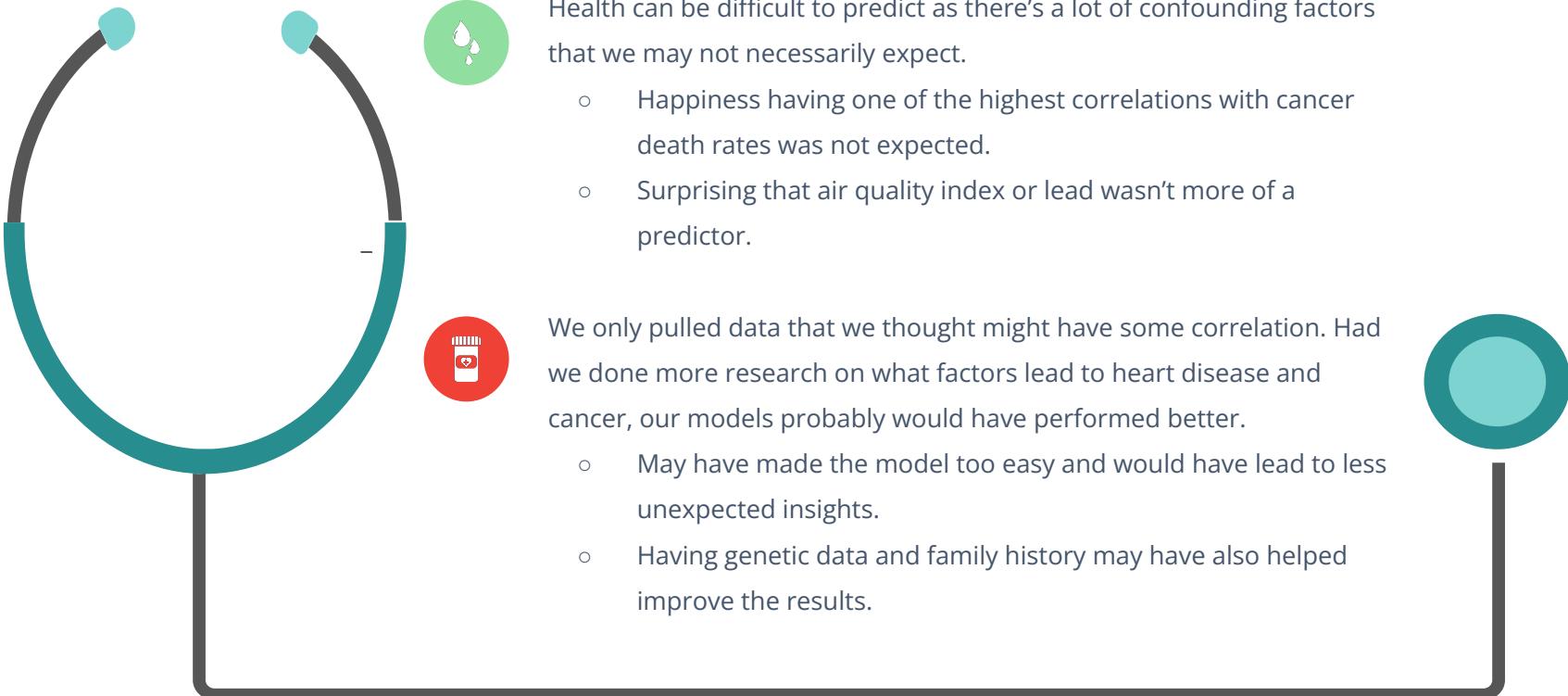


Non-religious rates lower than the US average

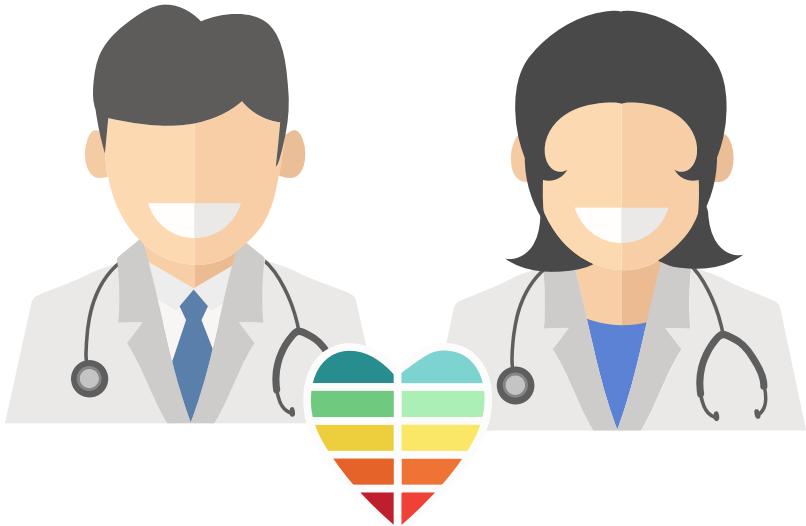


Education rates lower than the US average

Why the low model scores?



Recommendations



Exercise on a regular basis and maintain a healthy weight



Maintain a healthy diet and avoid excessive sugar



Continue to learn



Avoid exposure to industrial and environmental toxins



Know your risk, family history for heart disease and cancer



Backup

Data Sources

HD Stats

<https://www.cdc.gov/heartdisease/facts.htm>



Cancer Stats

[Stats of the States - Cancer Mortality \(cdc.gov\)](#)



Smoking Data

[Current Cigarette Smoking Among Adults in the United States | CDC](#)



State Data

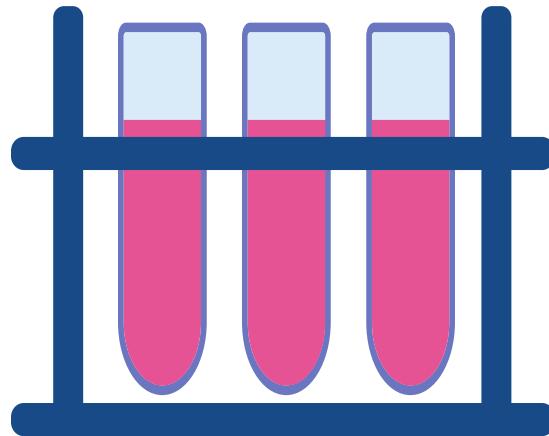
[https://data.world/johnsnowlabs/us-healthcare-spending-by-state](#)

[Health Insurance Coverage of the Total Population \(CPS\) | KFF](#)

[USDA ERS - Data Access and Documentation Downloads](#)

[https://worldpopulationreview.com/state-rankings/happiest-states](#)

[https://data.nrel.gov/submissions/50](#)



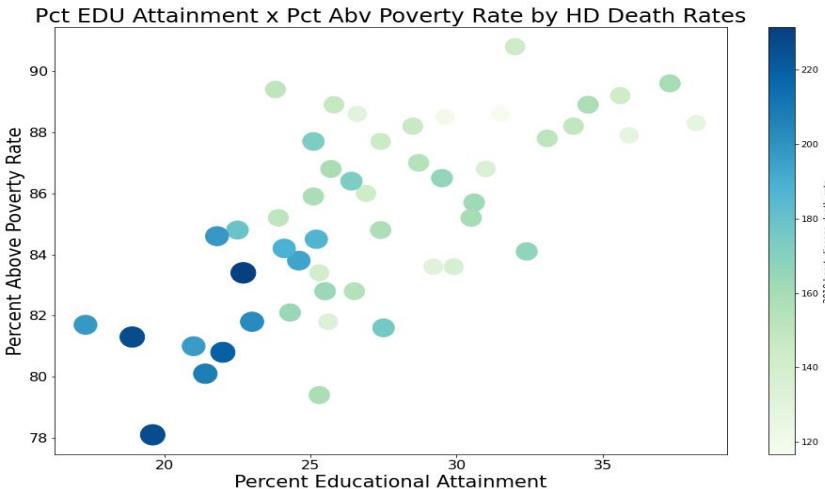
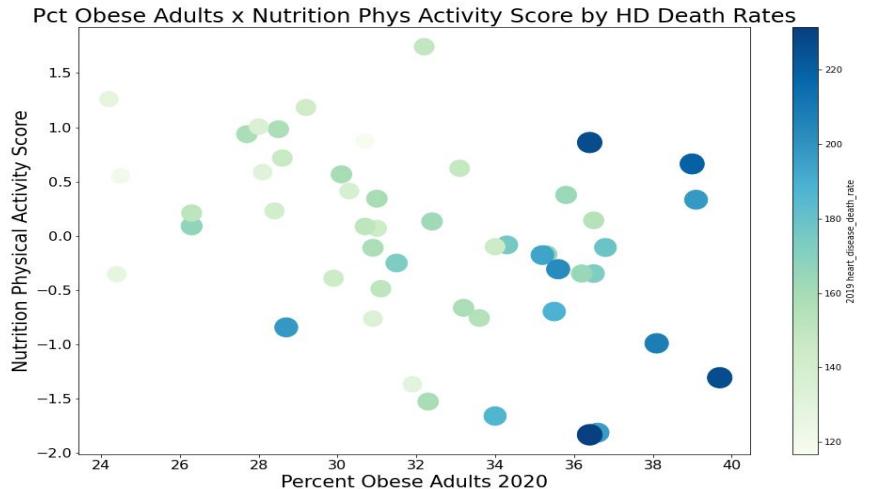
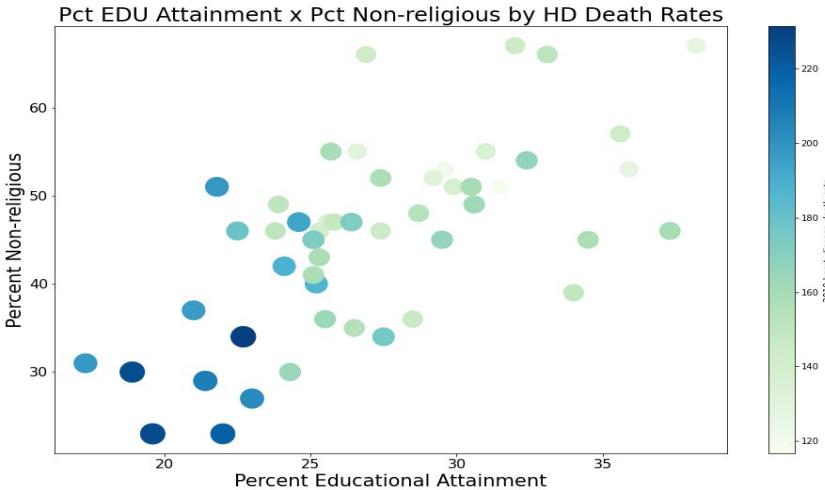
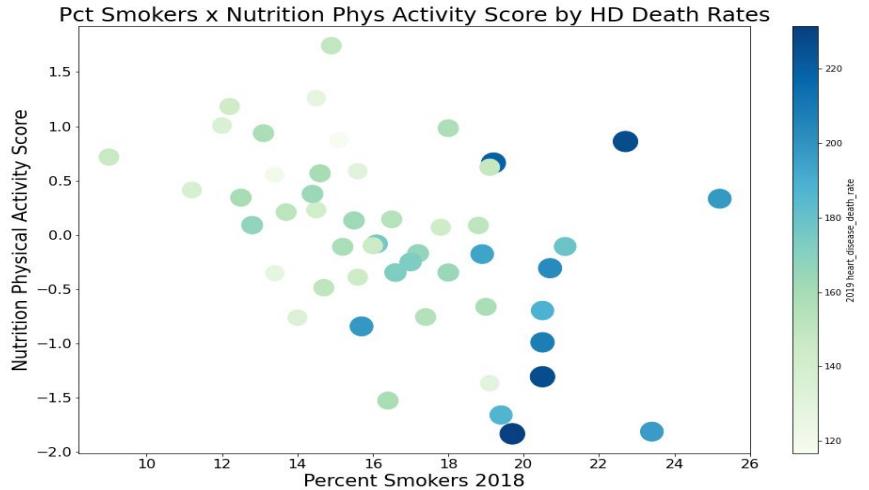
[Data, Trends and Maps | Overweight & Obesity | CDC](#)

[Adult Obesity Prevalence Maps | Overweight & Obesity | CDC](#)

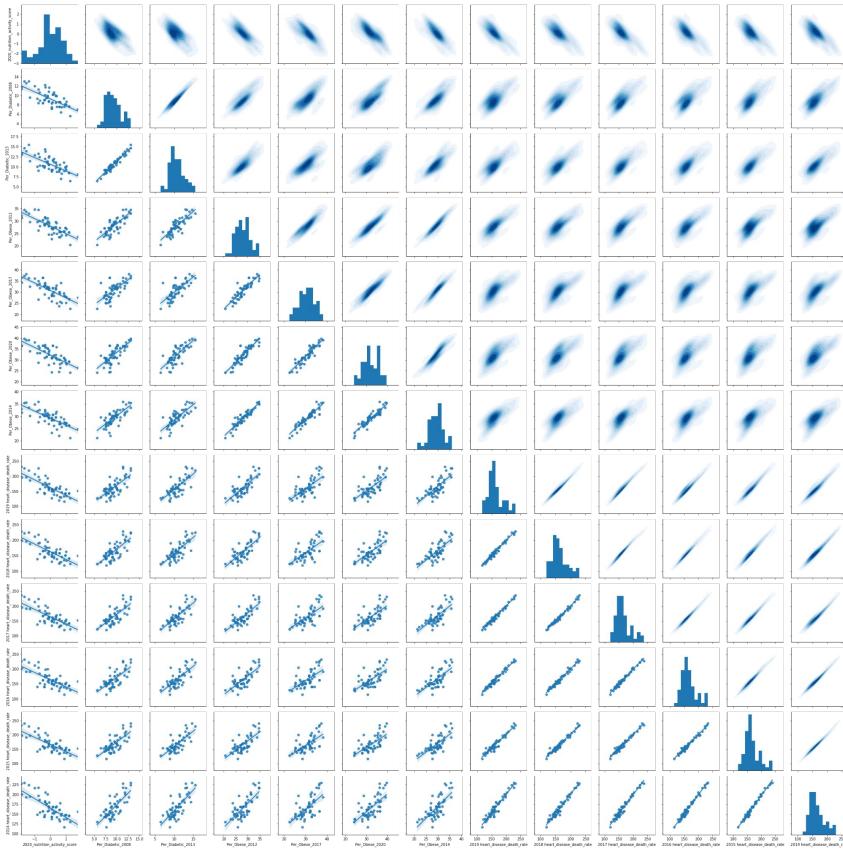
[The Fast Food Capitals of America | NiceRx](#)

[https://www.waterqualitydata.us/portal/](#)

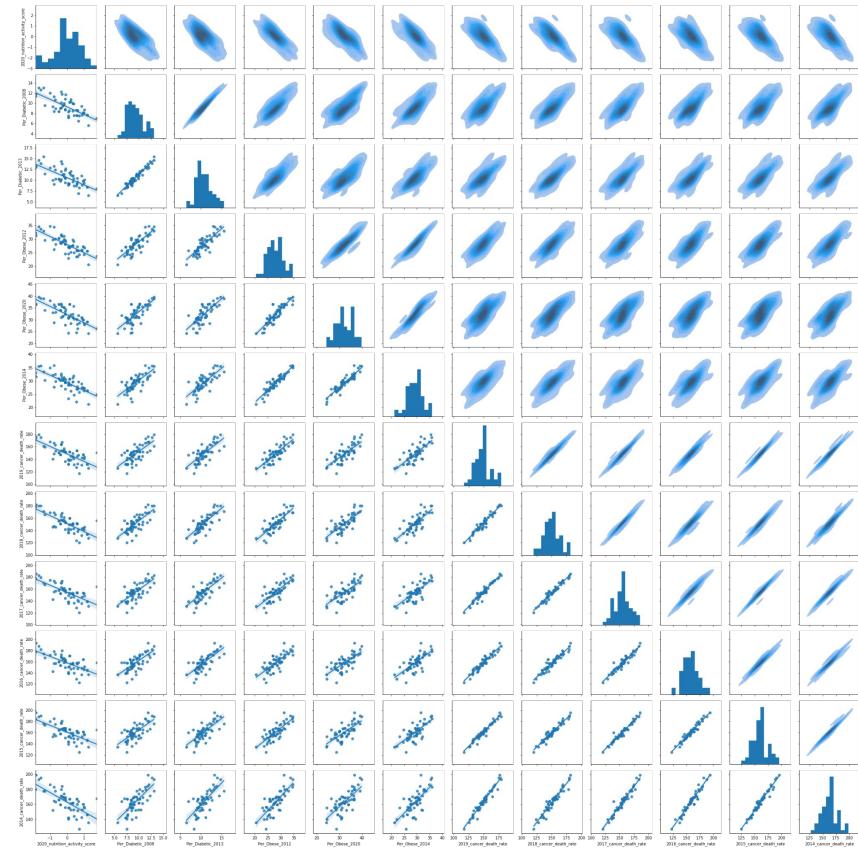
[https://www.epa.gov/outdoor-air-quality-data](#)



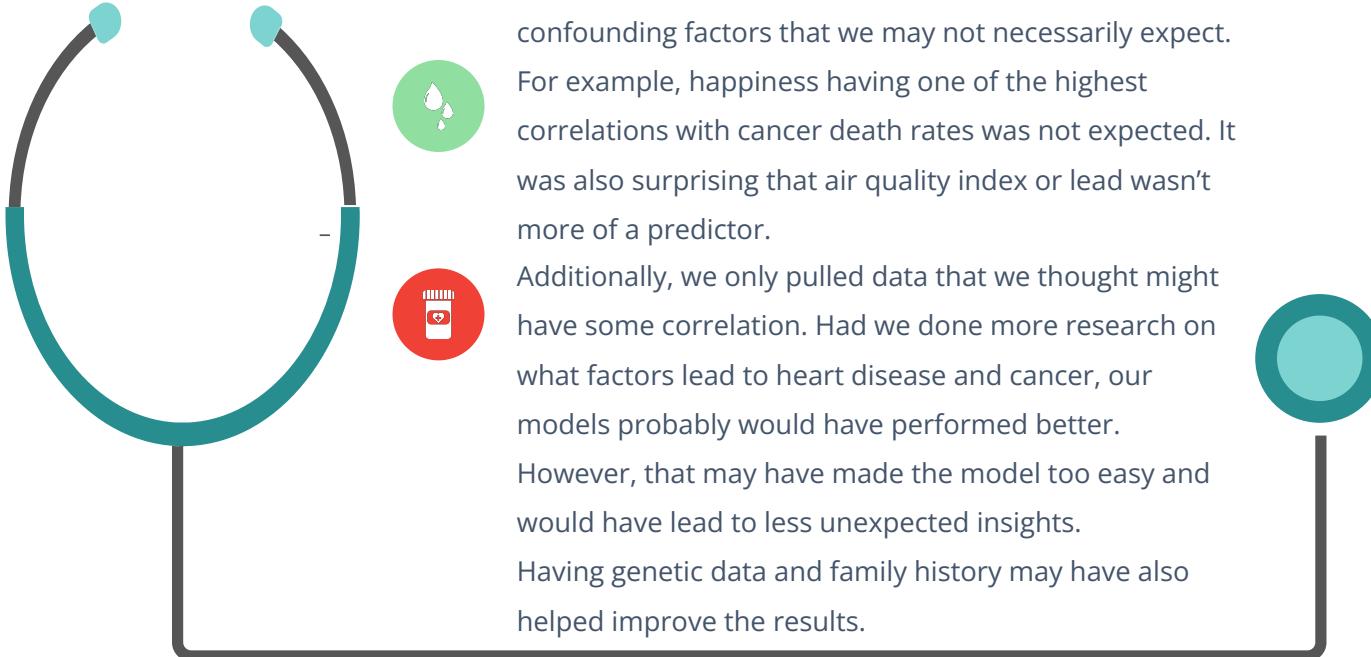
Cancer PairGrid



Heart Disease PairGrid



Possible Explanations



MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy
text of the printing and
typesetting industry.

Subtitle Here

Subtitle Here

Lorem Ipsum is simply dummy
text of the printing and
typesetting industry.

Subtitle Here

Subtitle Here

Lorem Ipsum is simply dummy
text of the printing and
typesetting industry.

Subtitle Here

Subtitle Here

Lorem Ipsum is simply dummy
text of the printing and
typesetting industry.



MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply



Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply



Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply



Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply



Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply

MEDICAL INFOGRAPHIC

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



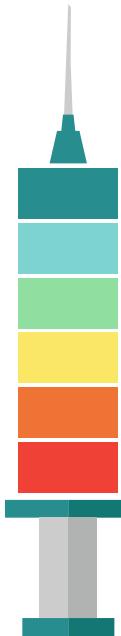
Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



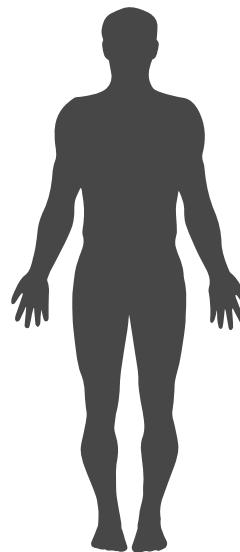
Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply



Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply



Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply



Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply

MEDICAL INFOGRAPHIC

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



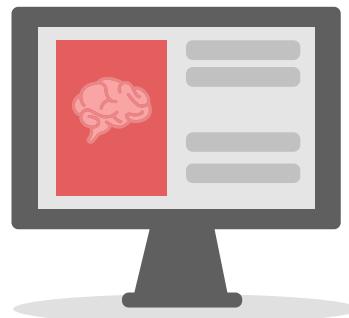
Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



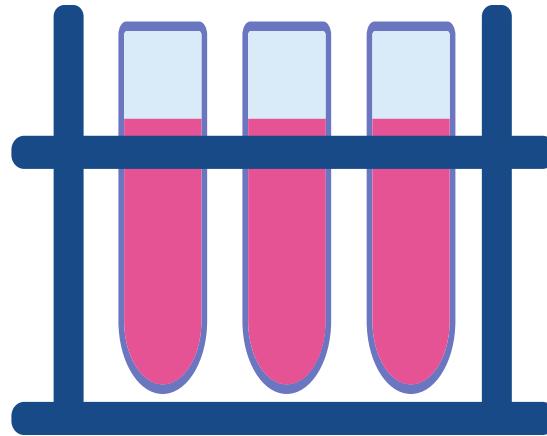
Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text

Subtitle Here

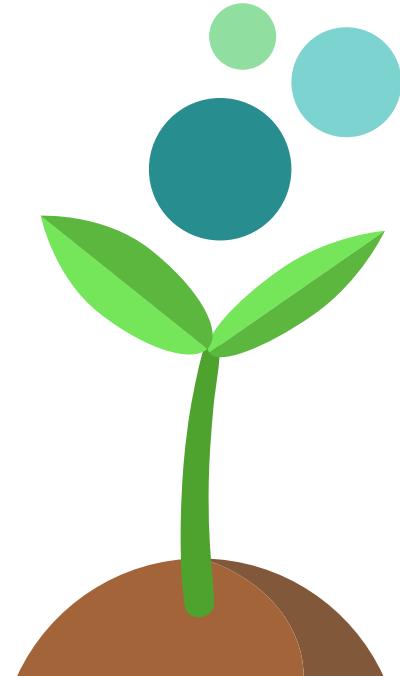
Lorem Ipsum is
simply dummy text
of the printing and
typesetting
industry.

Subtitle Here

Lorem Ipsum is
simply dummy text
of the printing and
typesetting
industry.

Subtitle Here

Lorem Ipsum is
simply dummy text
of the printing and
typesetting
industry.



MEDICAL INFOGRAPHIC

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.

MEDICAL INFOGRAPHIC

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

MEDICAL INFOGRAPHIC

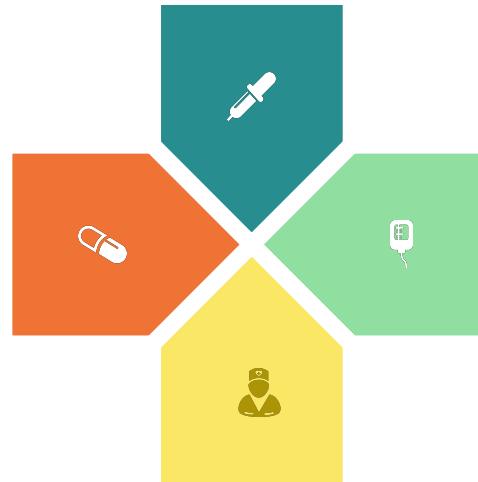
Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

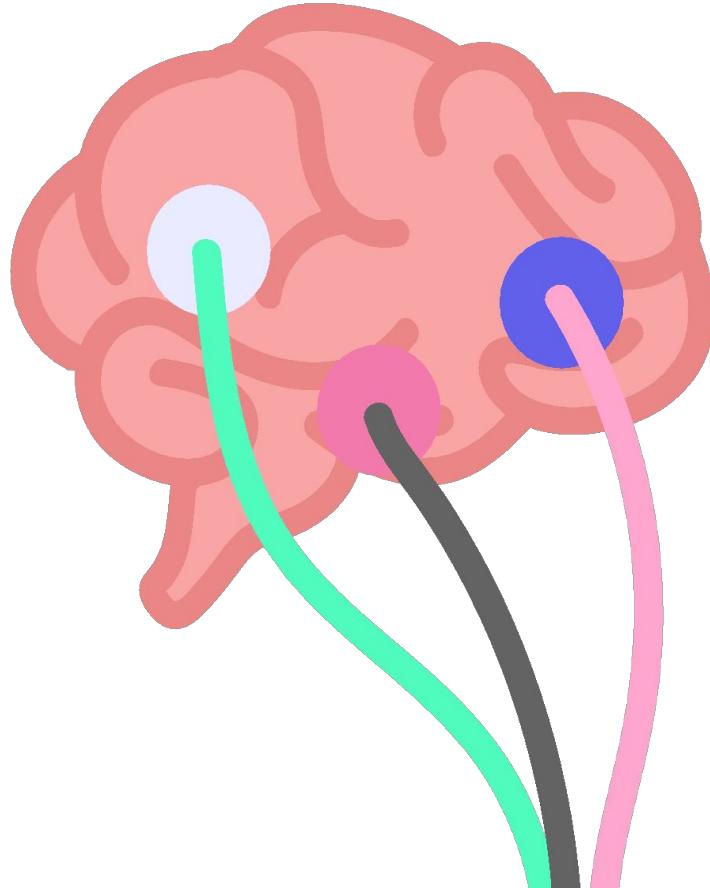


Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



MEDICAL INFOGRAPHIC

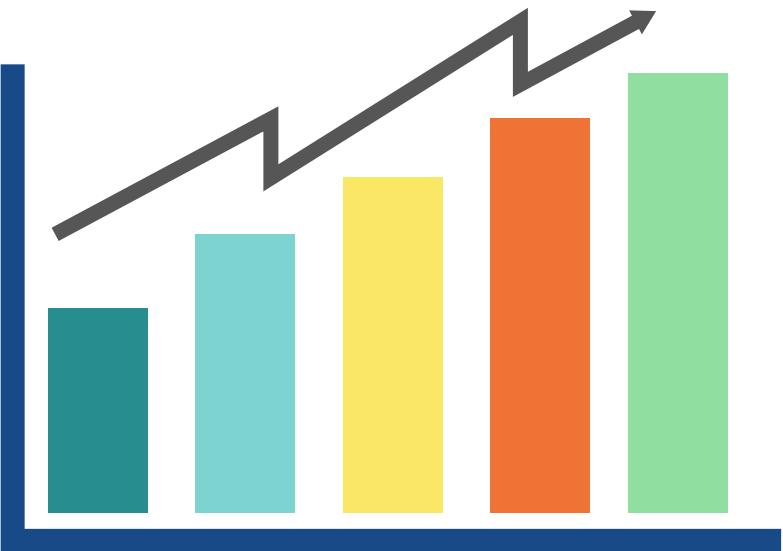
Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.



MEDICAL INFOGRAPHIC



Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy



Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy



Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy



Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy



Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy

MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

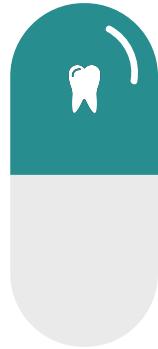
Lorem Ipsum is simply dummy text of the printing and typesetting.



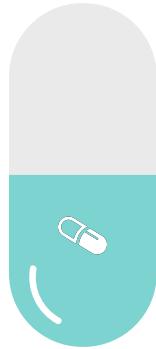
Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

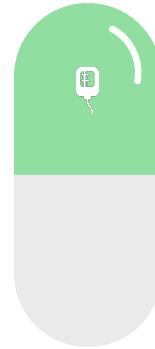
MEDICAL INFOGRAPHIC



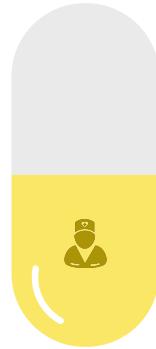
Subtitle Here
Lorem Ipsum is
simply dummy text
of the printing and
typesetting.



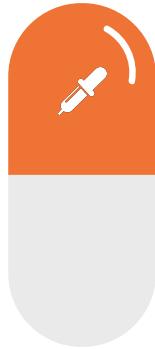
Subtitle Here
Lorem Ipsum is
simply dummy text
of the printing and
typesetting.



Subtitle Here
Lorem Ipsum is
simply dummy text
of the printing and
typesetting.



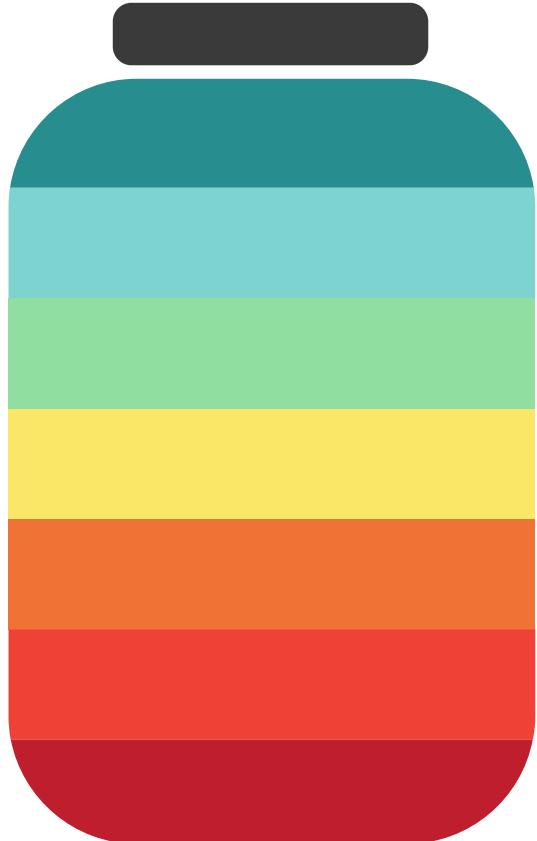
Subtitle Here
Lorem Ipsum is
simply dummy text
of the printing and
typesetting.



Subtitle Here
Lorem Ipsum is
simply dummy text
of the printing and
typesetting.



Subtitle Here
Lorem Ipsum is
simply dummy text
of the printing and
typesetting.



MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

MEDICAL INFOGRAPHIC

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

MEDICAL INFOGRAPHIC

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting. Lorem Ipsum is simply dummy text of the printing and typesetting.

MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

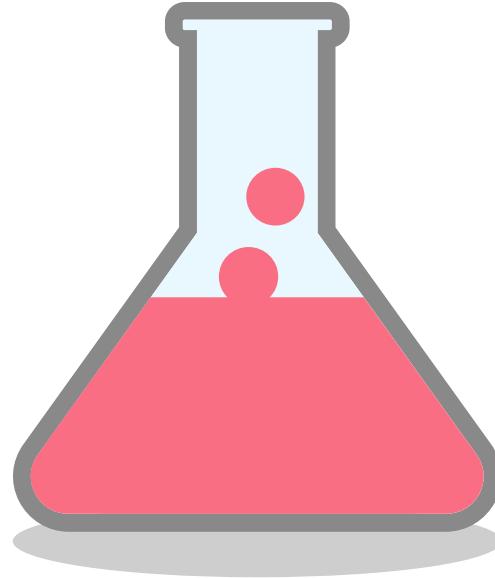
Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.



MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.



Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

MEDICAL INFOGRAPHIC

Subtitle Here

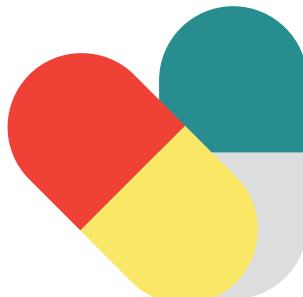
Subtitle Here

Subtitle Here

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.



MEDICAL INFOGRAPHIC

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

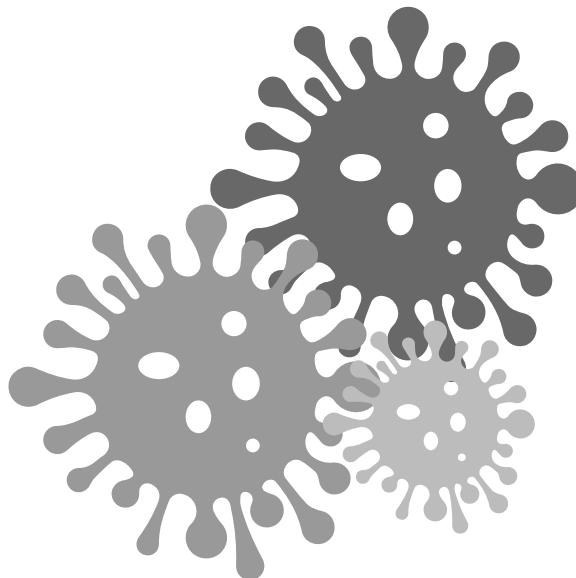
Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting.





MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

MEDICAL INFOGRAPHIC



Lorem Ipsum is simply dummy text
of the printing



Lorem Ipsum is simply dummy text
of the printing



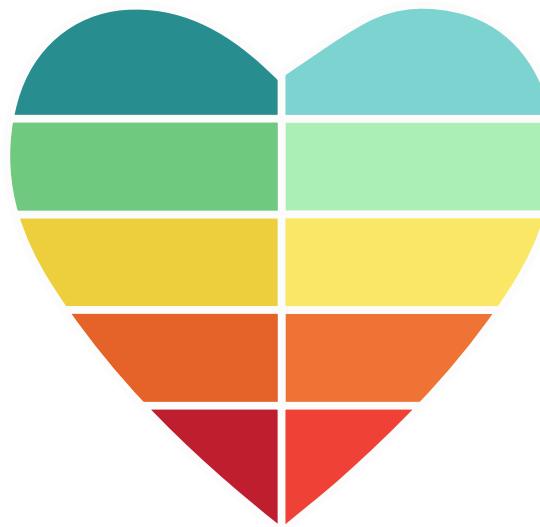
Lorem Ipsum is simply dummy text
of the printing



Lorem Ipsum is simply dummy text
of the printing



Lorem Ipsum is simply dummy text
of the printing



Lorem Ipsum is simply dummy text
of the printing



Lorem Ipsum is simply dummy text
of the printing



Lorem Ipsum is simply dummy text
of the printing



Lorem Ipsum is simply dummy text
of the printing



Lorem Ipsum is simply dummy text
of the printing

MEDICAL

INFOGRAPHIC



Lorem Ipsum is simply dummy text of the printing
Lorem Ipsum is simply dummy text of the printing



Lorem Ipsum is simply dummy text of the printing
Lorem Ipsum is simply dummy text of the printing



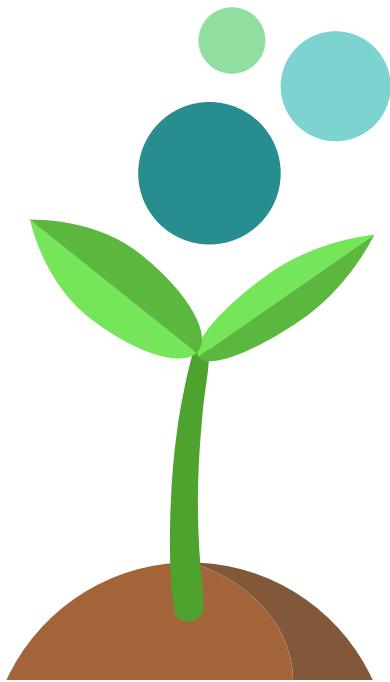
Lorem Ipsum is simply dummy text of the printing
Lorem Ipsum is simply dummy text of the printing



Lorem Ipsum is simply dummy text of the printing
Lorem Ipsum is simply dummy text of the printing



Lorem Ipsum is simply dummy text of the printing
Lorem Ipsum is simply dummy text of the printing



MEDICAL INFOGRAPHIC

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.

Subtitle Here

Lorem Ipsum is simply dummy text of the printing and typesetting industry.