// Module 2 Final Project

Using multivariate linear regression (MLR)
to predict life expectancy.

Jonathan Vasquez

Contents

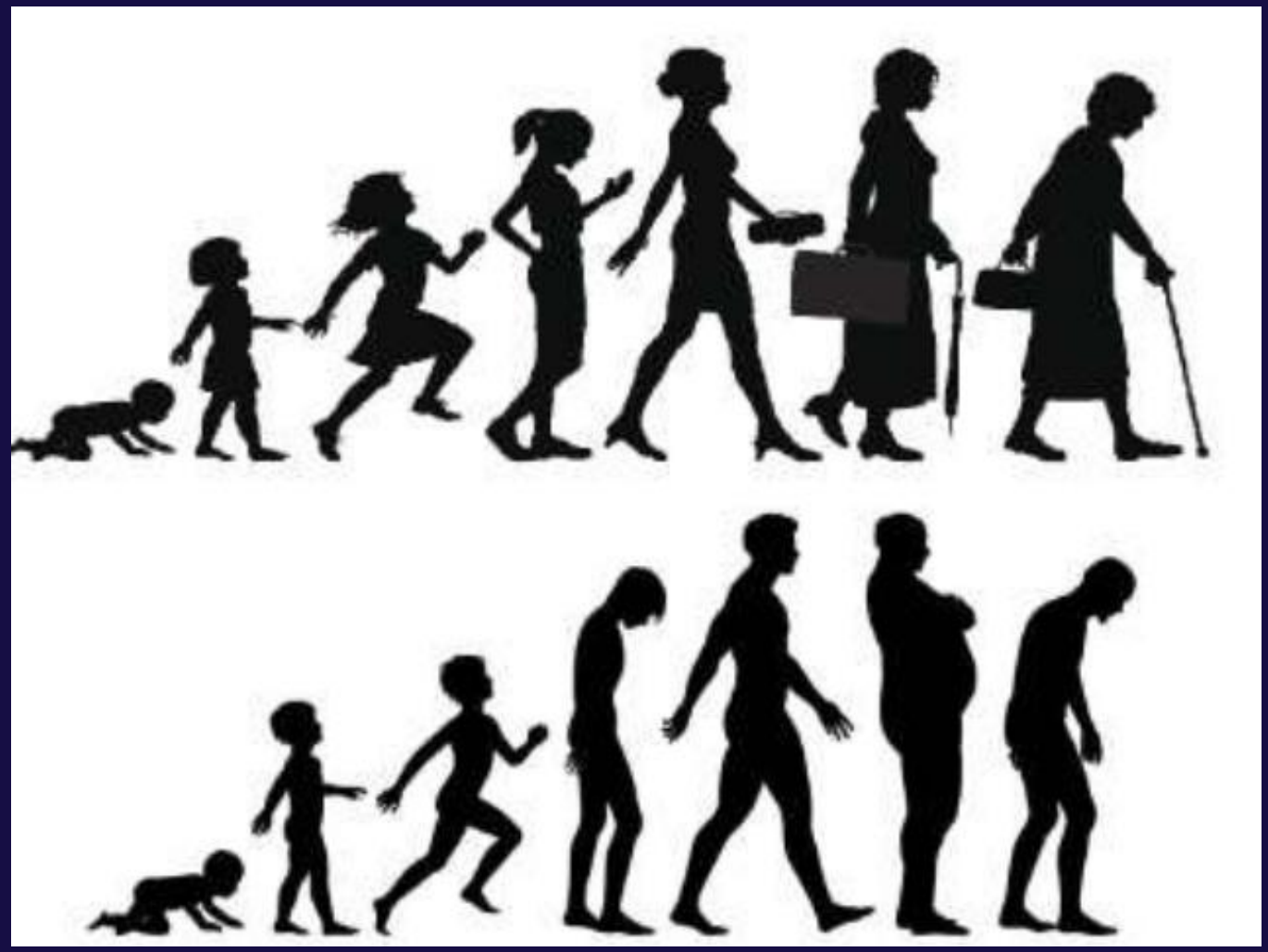
// Overall Goal

Predict life expectancy via:

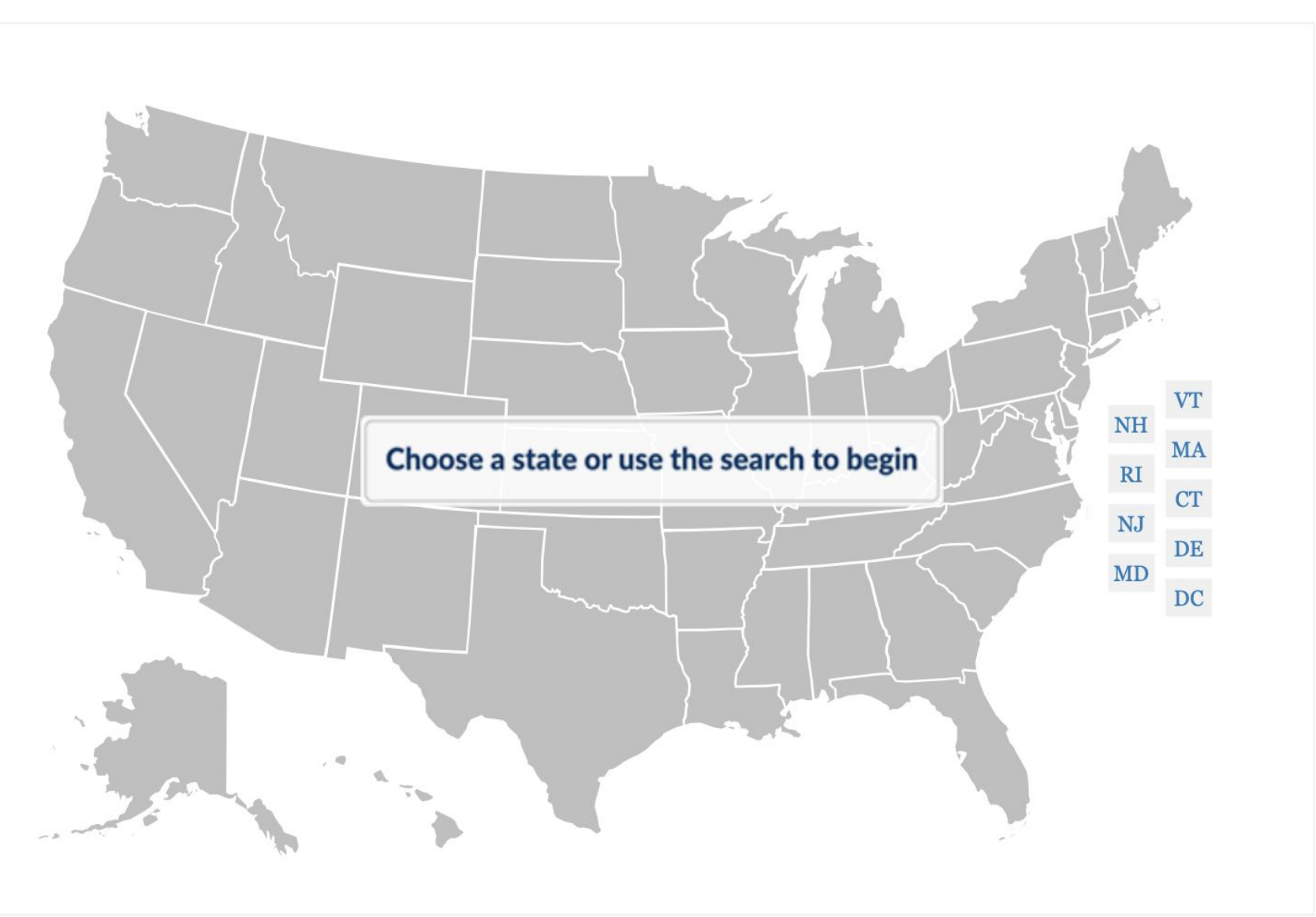
- Health Behaviors
- Clinical Care Access
- Socioeconomic Factors
- Physical Environment

Lasso model's prediction within 0.5587 SD's.

Lasso Training Error: 1.0133
Lasso Testing Error: 1.6517
SD: 2.956



// Data Gathering / Data Cleaning



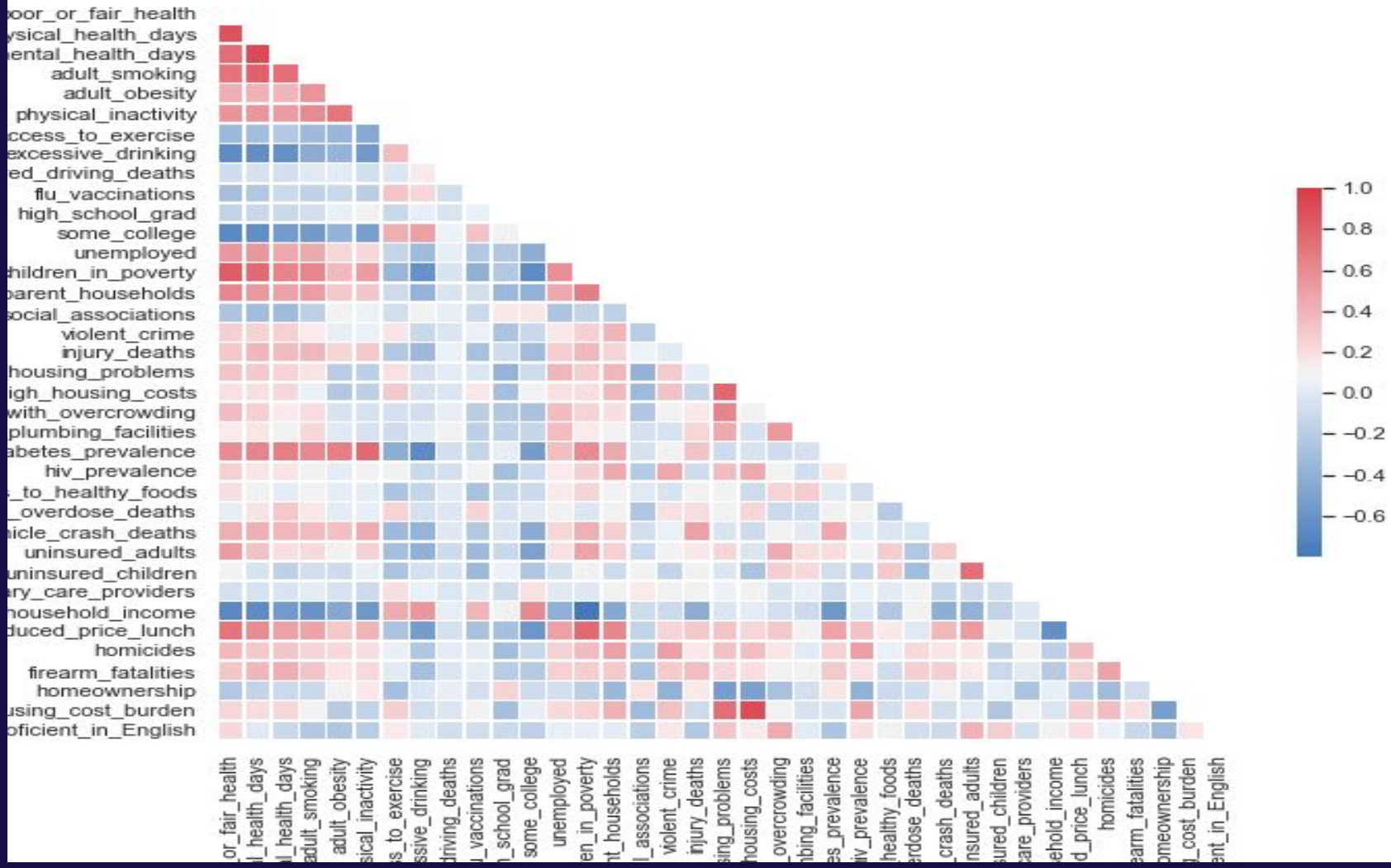
Data Gathering:

Source: County Health Rankings
Collaboration: RWJ & UoW
Break-down: 3,193 counties/50 states
Year: 2019

Data Cleaning:

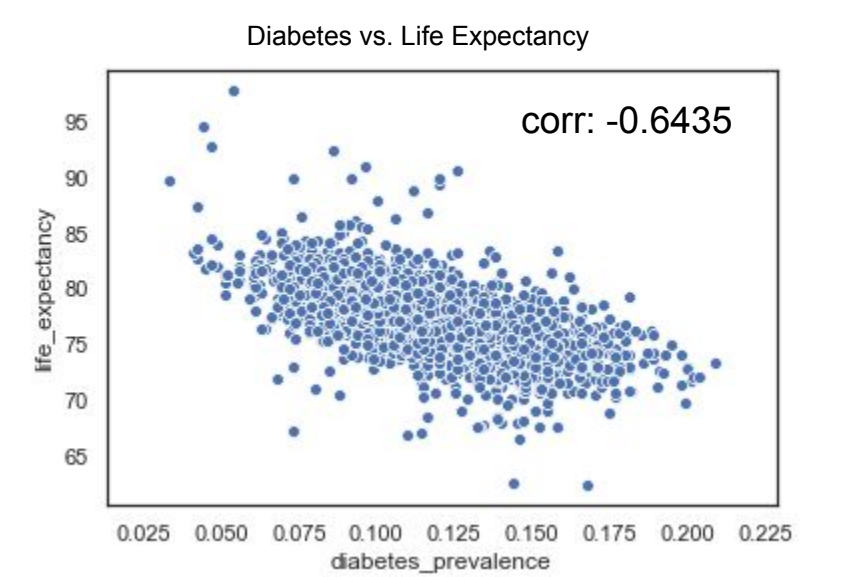
- Dropped 488 columns
- Replaced ‘ ’ for ‘_’
- Converted objects to floats
- Filled in NaN’s

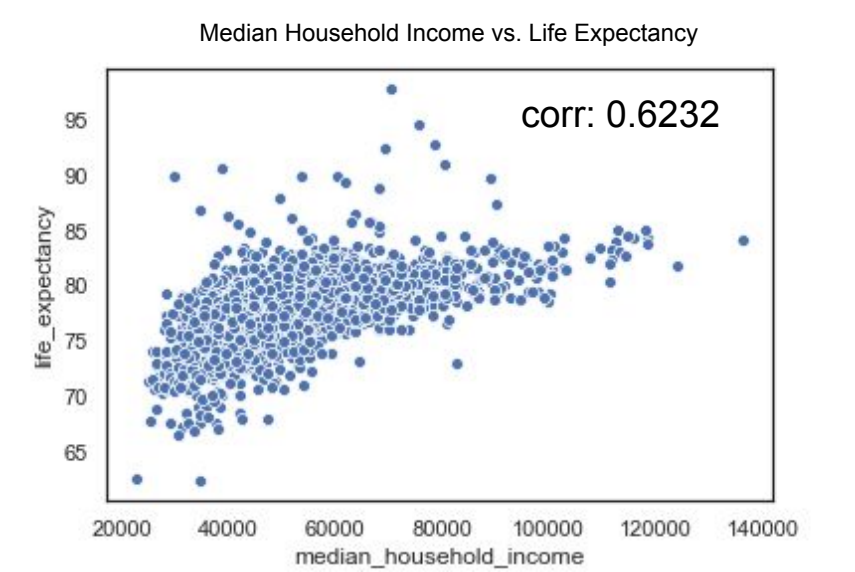
// Exploratory Data Analysis

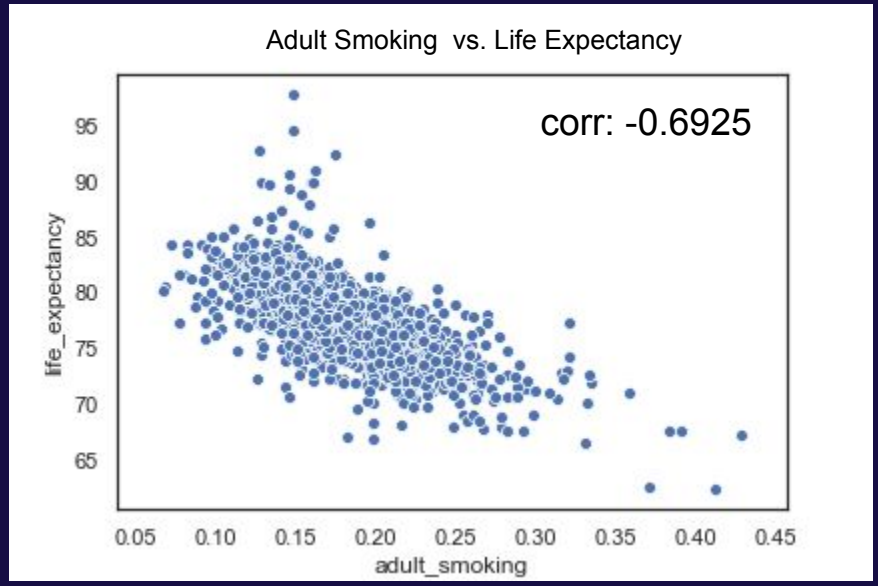


- Tobacco Use
- Diet & Exercise
- Alcohol & Drug Use
- Sexual Activity
- Access to Care
- Quality of Care
- Education
- Employment
- Income
- Family & Social Support
- Community Safety
- Air & Water Quality
- Housing & Transit

// Hypothesis Testing







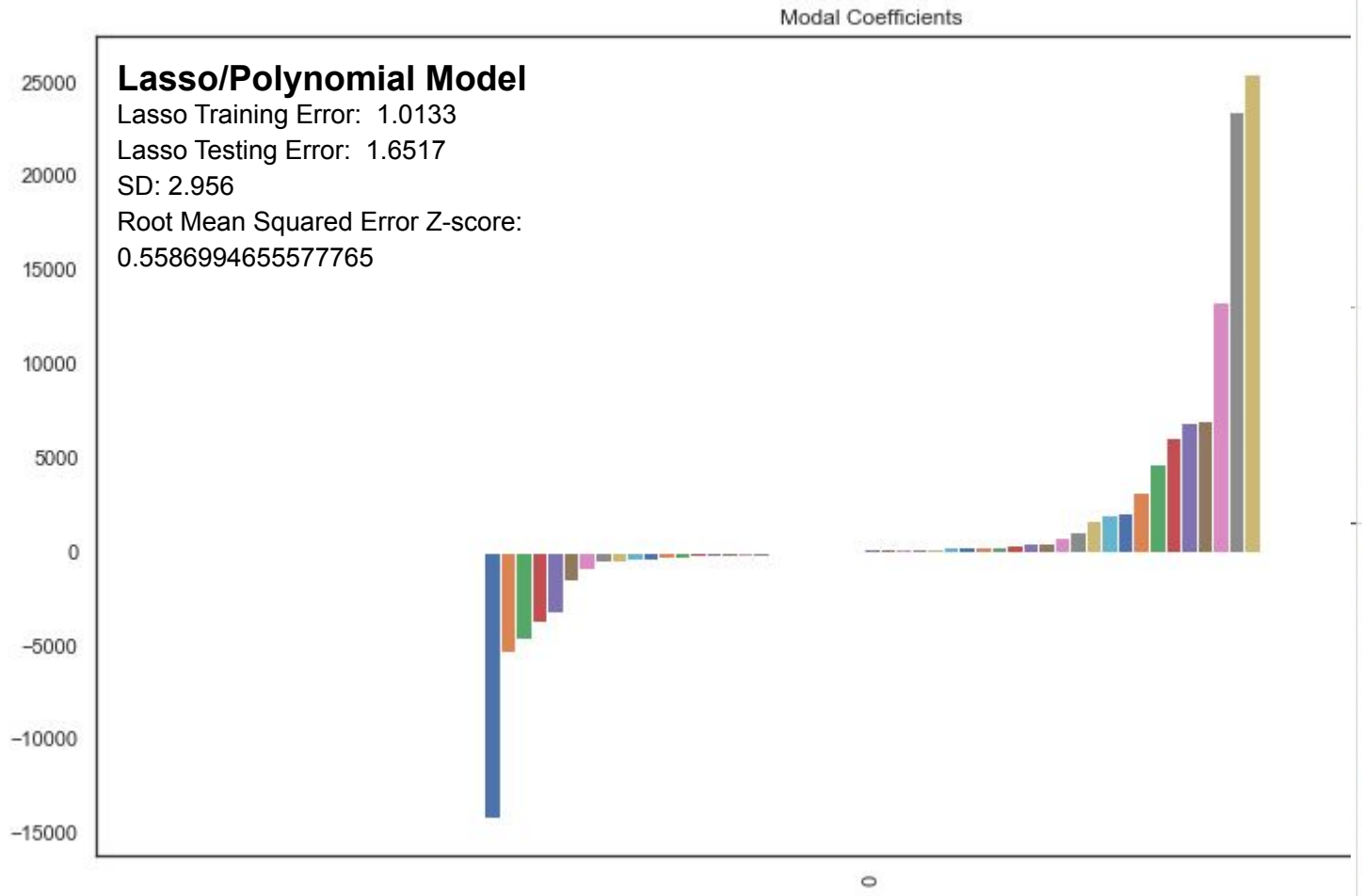
Reject the null hypothesis:

- Diabetes Prevalence & Life Expectancy.
- Median Household Income & Life Expectancy.

Fail to reject the null hypothesis:

- Adult Smoking & Life Expectancy.

// Modeling Data



// Conclusion

- Lasso/Poly model most successful
- RSME Z-score ~0.5587
- Able to predict life expectancy within 1 SD of population mean

// Future work

- Pinpoint specific counties, within states with lowest life expectancy, and create model and compare to original.
  - The counties in Mississippi, state with lowest life expectancy, F-Test model had RMSE Z-score of ~0.6136, similar to overall model.
- Separate out the categories (Health Behaviors, Clinical Care Access, Socioeconomic Factors, Physical Environment) to establish true weight on life expectancy.
- From influx of money from "defund the police" work with state legislators to allocate resources accordingly for health/social reforms.

Questions?

Thank You!

Jonathan Vasquez