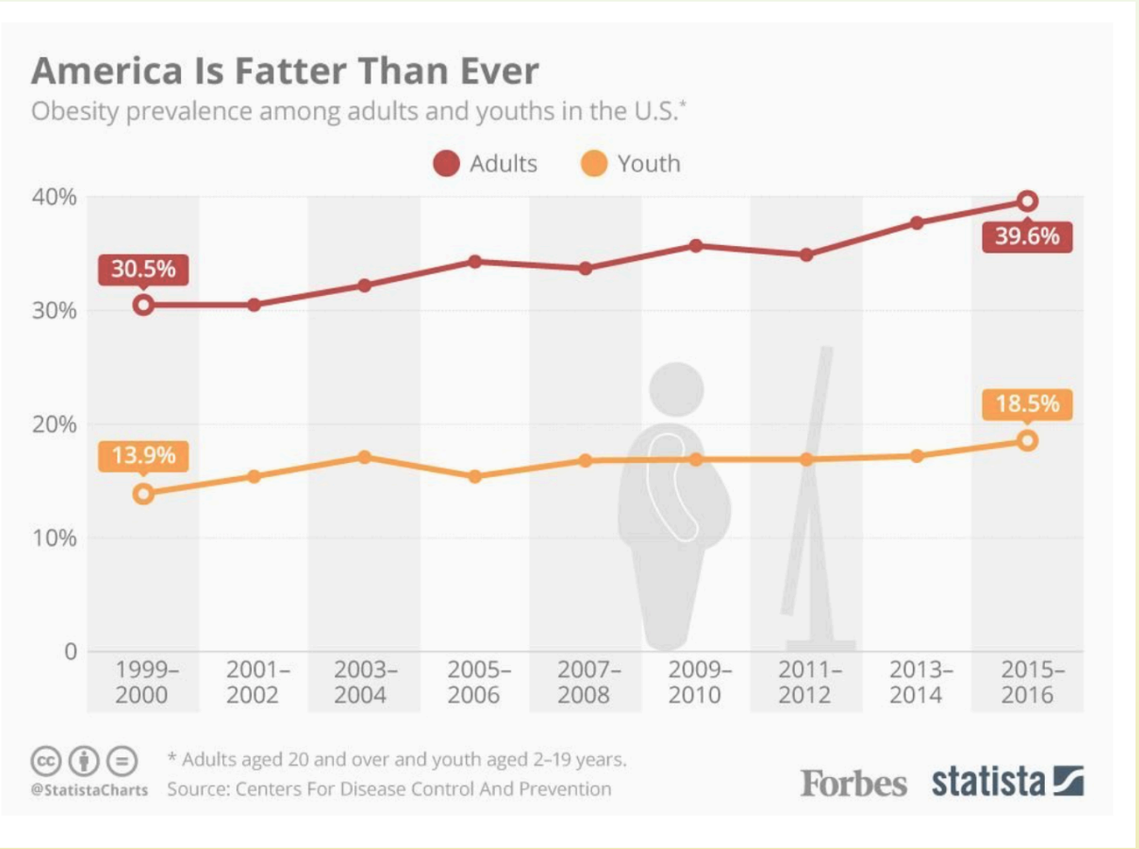
Obesity in the United States

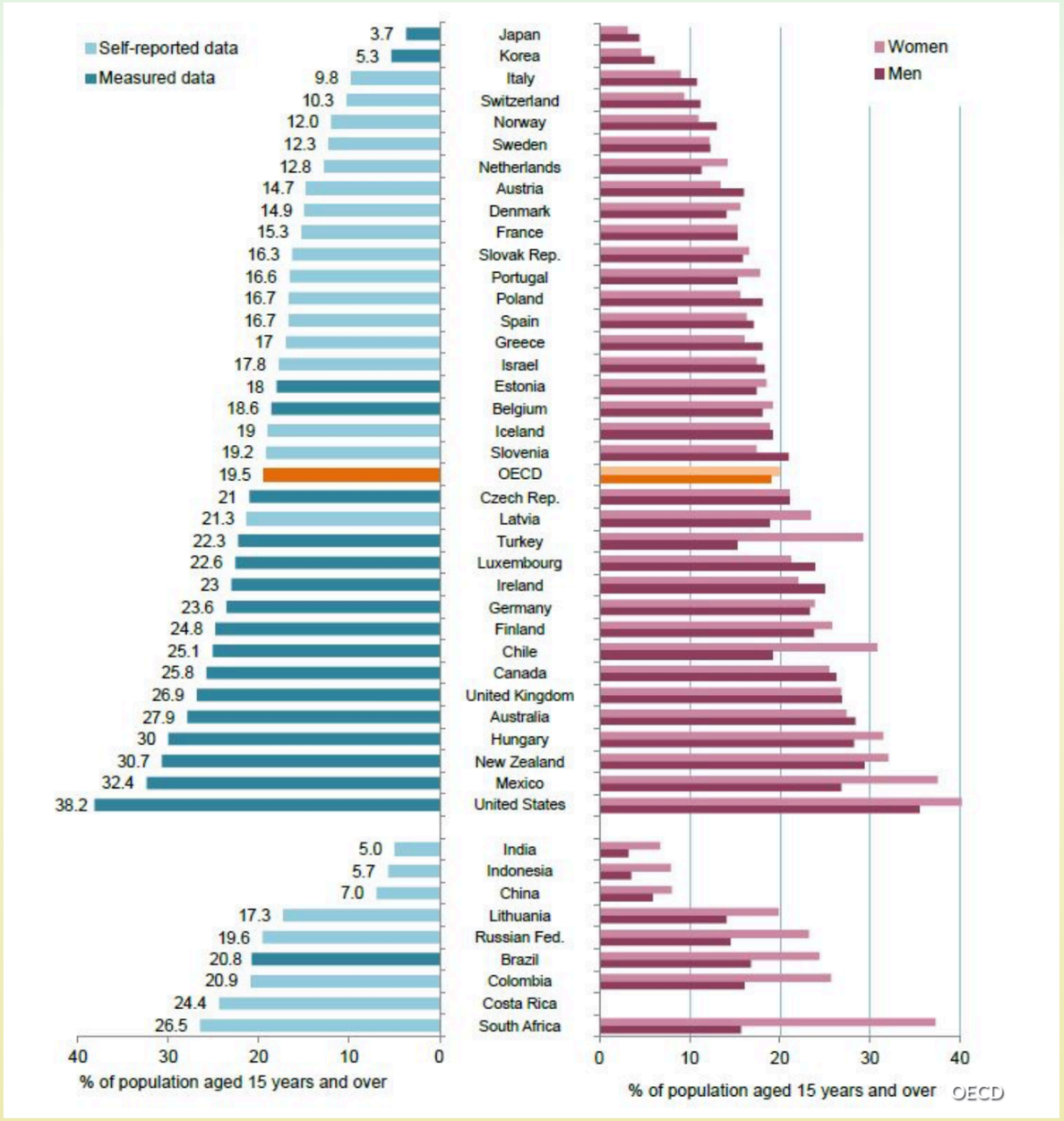
Kalina & Joe

Facts:

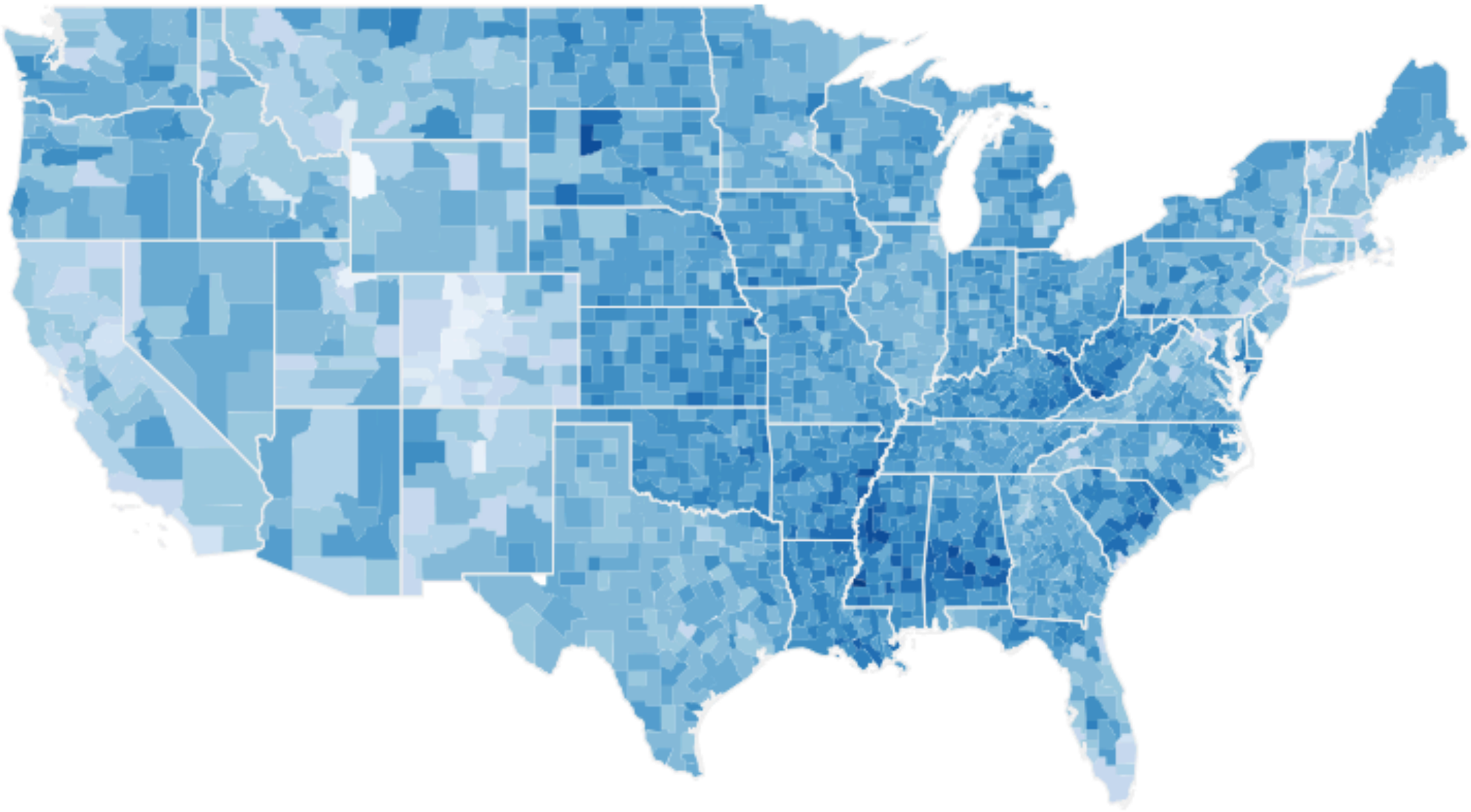
Still increasing number of obese adults



The United States at the top of the ranking-



Percentage of Obese Adults per County



How can we help society?

Can we predict overweight and obesity using available data?

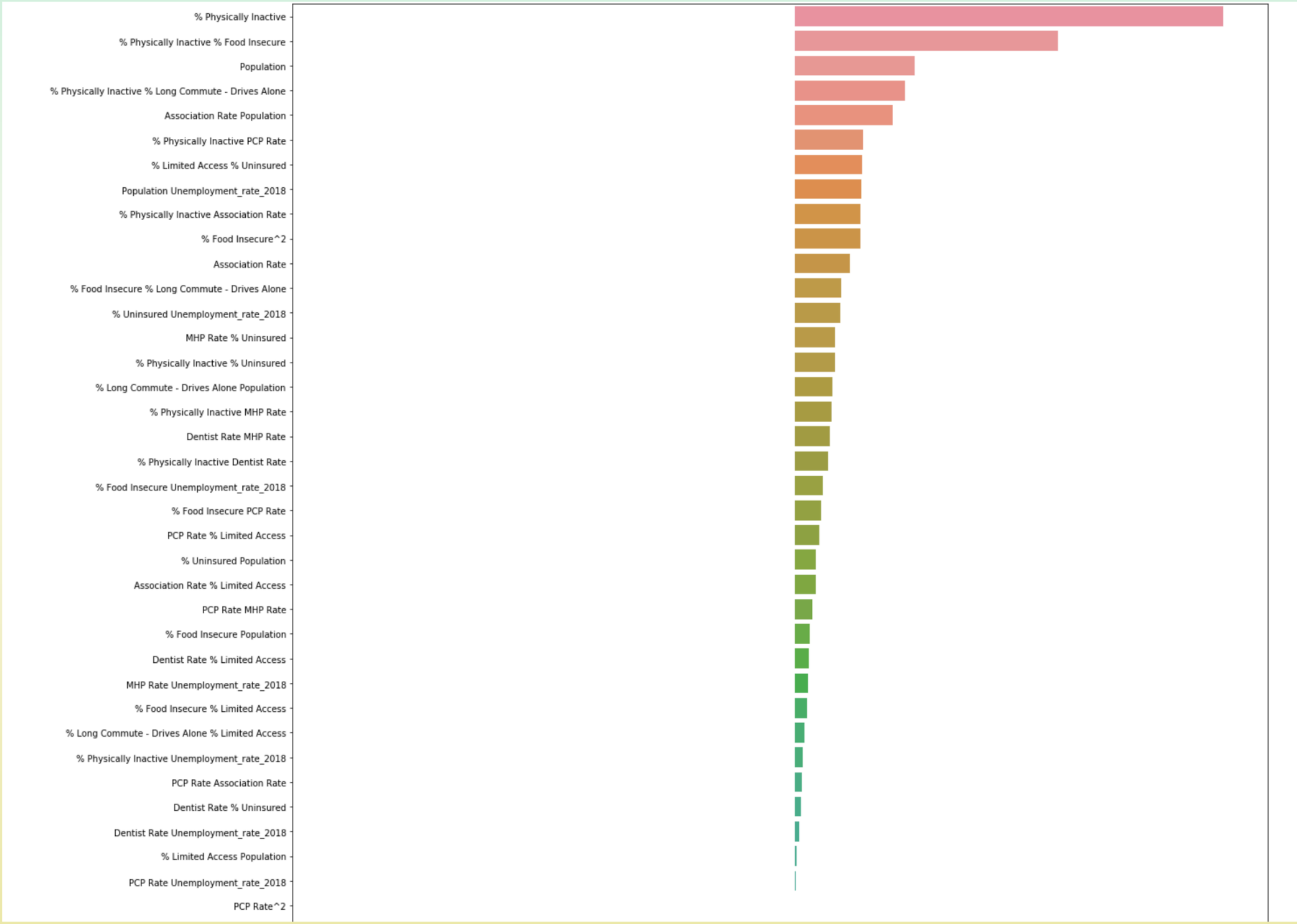
How can we prevent obesity?

Which factors can help with stopping increasing numbers of obese adults?

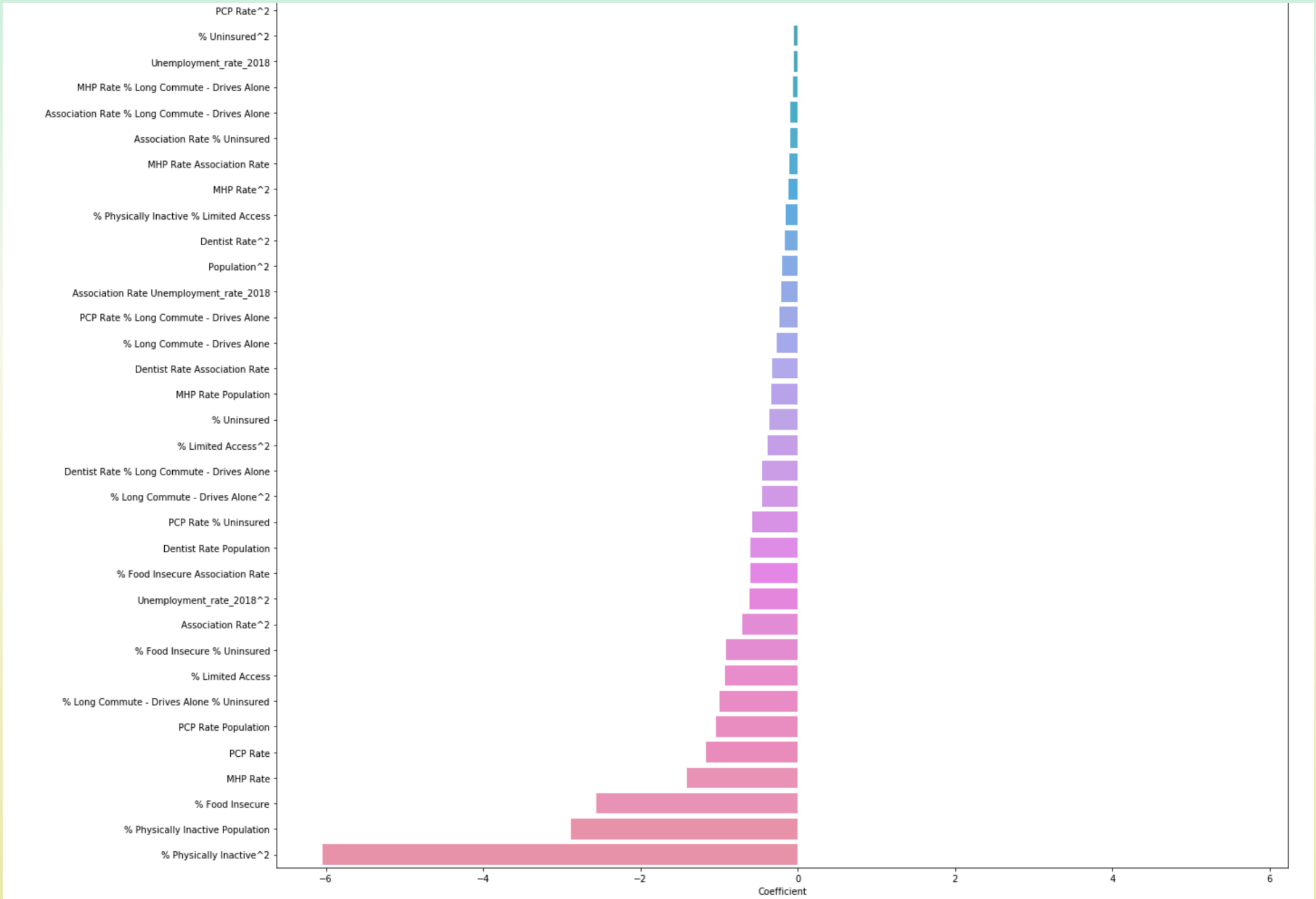
Our factors:

- Percentage of adults that report no leisure-time physical activity
- Food insecurity
- Primary Care Physicians per 100,000 population
- Dentists per 100,000 population
- Mental Health Providers per 100,000 population
- Social Associations per 10,000 population
- Among workers who commute in their car alone, the percentage that commute more than 30 minutes
- Percentage of Limited access to healthy foods
- Percentage of Uninsured adults
- Population
- Unemployment rate in 2018

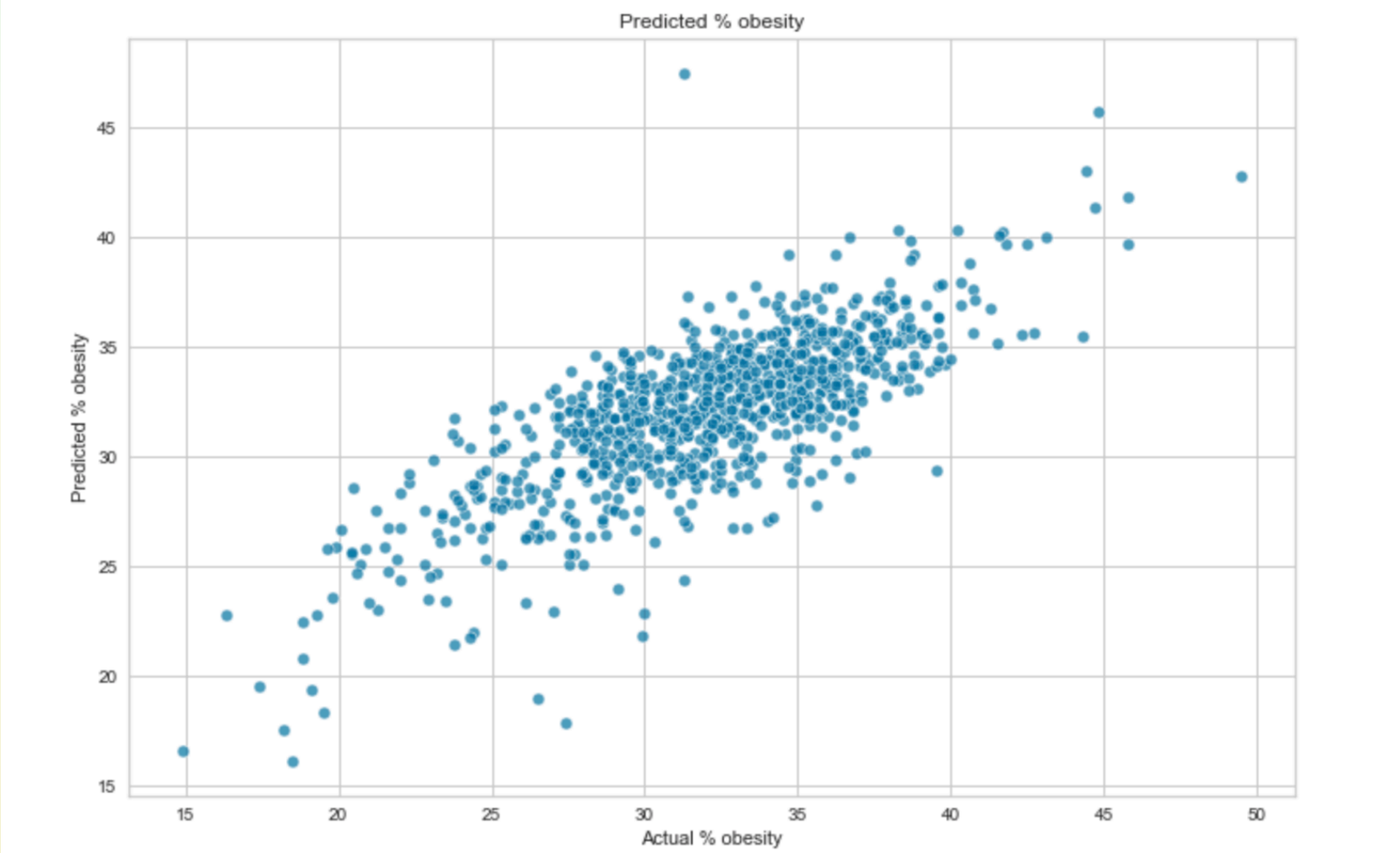
Positive coefficients:

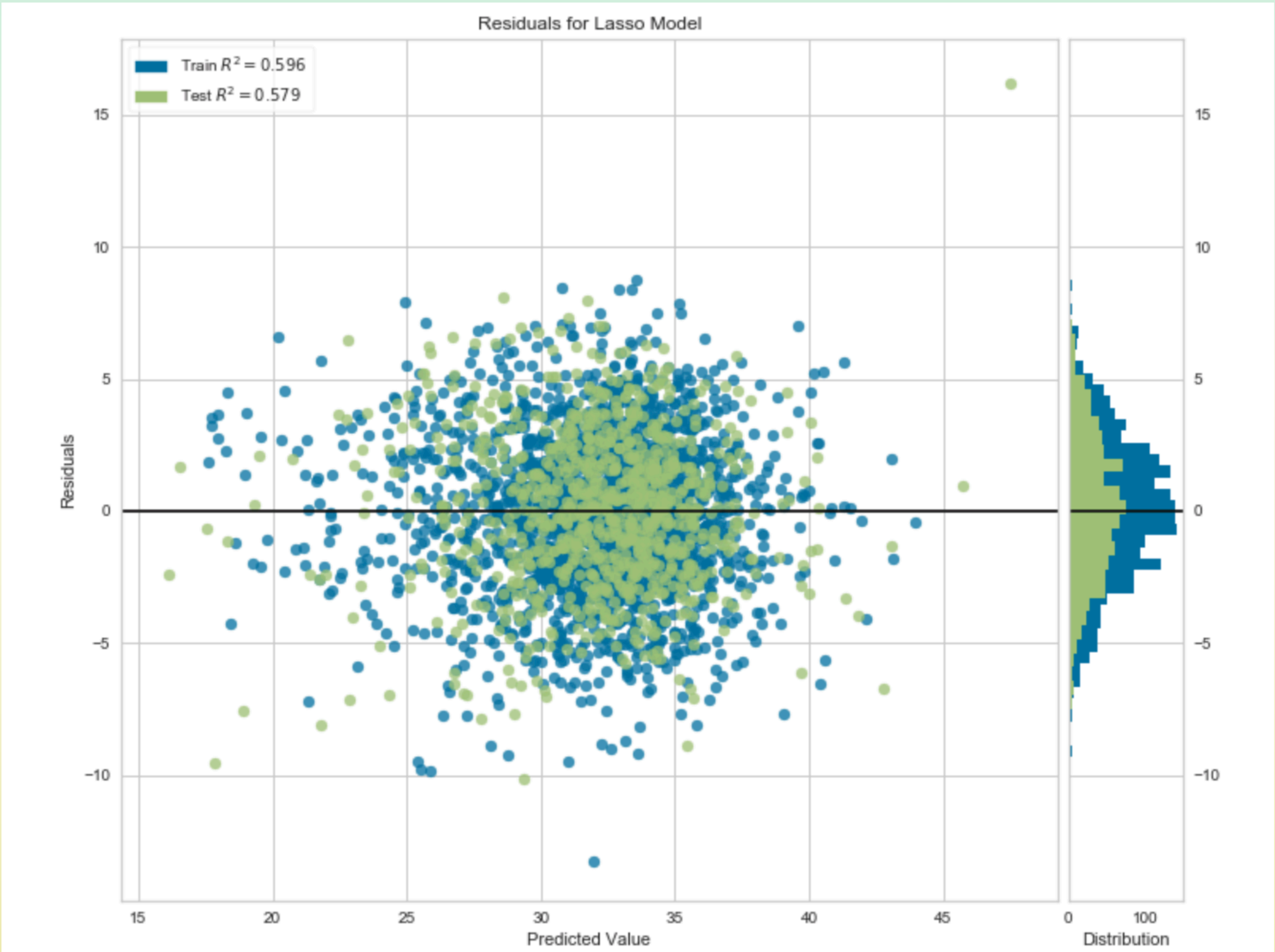


Negative coefficients:



Final predictions:




Residuals for Lasso Model
Train $R^2 = 0.596$
Test $R^2 = 0.579$
Residuals
Predicted Value
Distribution
15
10
5
0
−5
−10
15
20
25
30
35
40
45
0
100

Thank you!

More technical details:

1. Variables to work with
2. Extra data about unemployment
3. Cleaning the data:
   - Replacing null values with state average
   - Dropping small populations with multiple null values
4. Modelling:

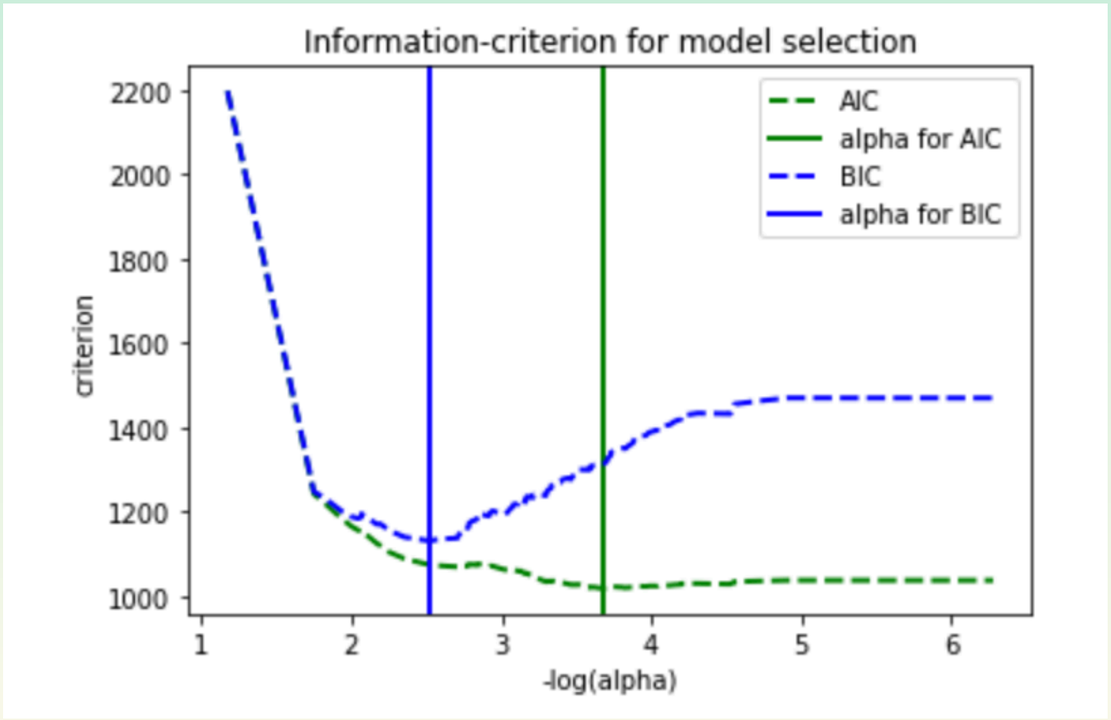
   - Splitting Datasets into Training and Testing
   - Creating baseline model with 5 fold cross validation
     - R2 score ~0.59
     - Not appear to be overfitted
   - Checking for Multicollinearity
     - VIF with a threshold of 10

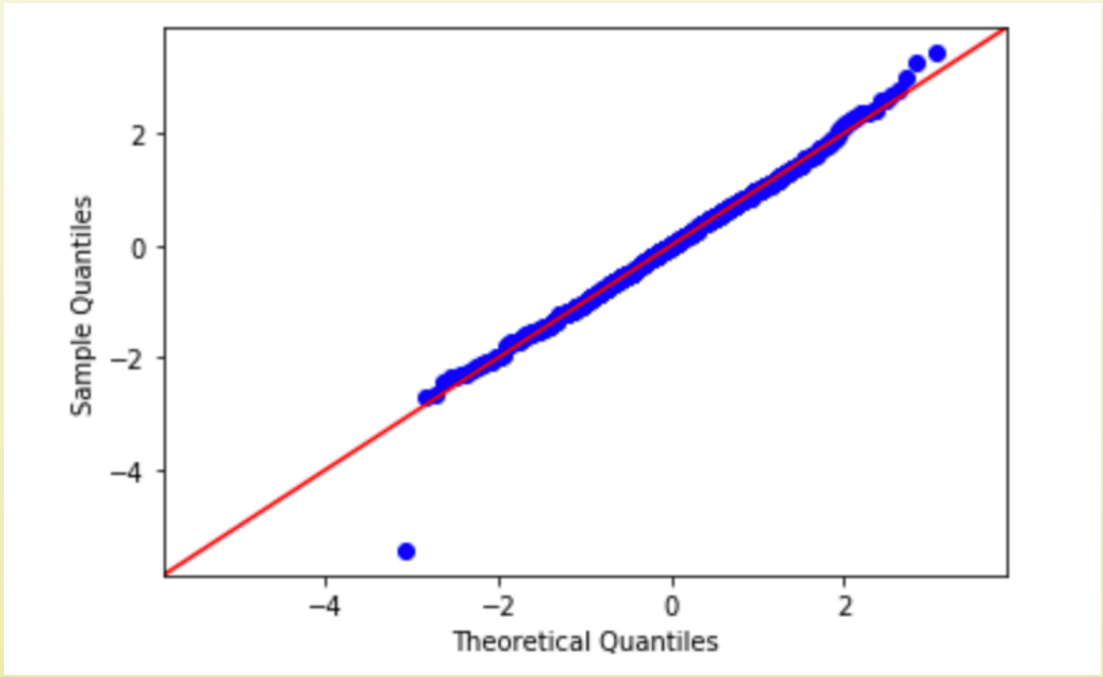
```
Remaining variables
Index(['PCP Rate', 'Dentist Rate', 'MHP Rate', 'Association Rate', '% Long Commute - Drives Alone', '% Limited Acce
ss', '% Uninsured', 'Population', 'Unemployment_rate_2018'], dtype='object')
```

   - VIF returned 9 variables, however will also include % Physically Inactive and % Food Insecure as 'common sense' variables
   - After removing variables, R2 decreases, as expected, and we can see that model is still not overfit
   - Adding second degree polynomial and interaction complexity to model to increase complexity

Evaluation:

- Optimising hyperparameters


Information-criterion for model selection
AIC
alpha for AIC
BIC
alpha for BIC
criterion
2200
2000
1800
1600
1400
1200
1000
1
2
3
4
5
6
-log(alpha)


Sample Quantiles
2
0
−2
−4
−4
−2
0
2
Theoretical Quantiles