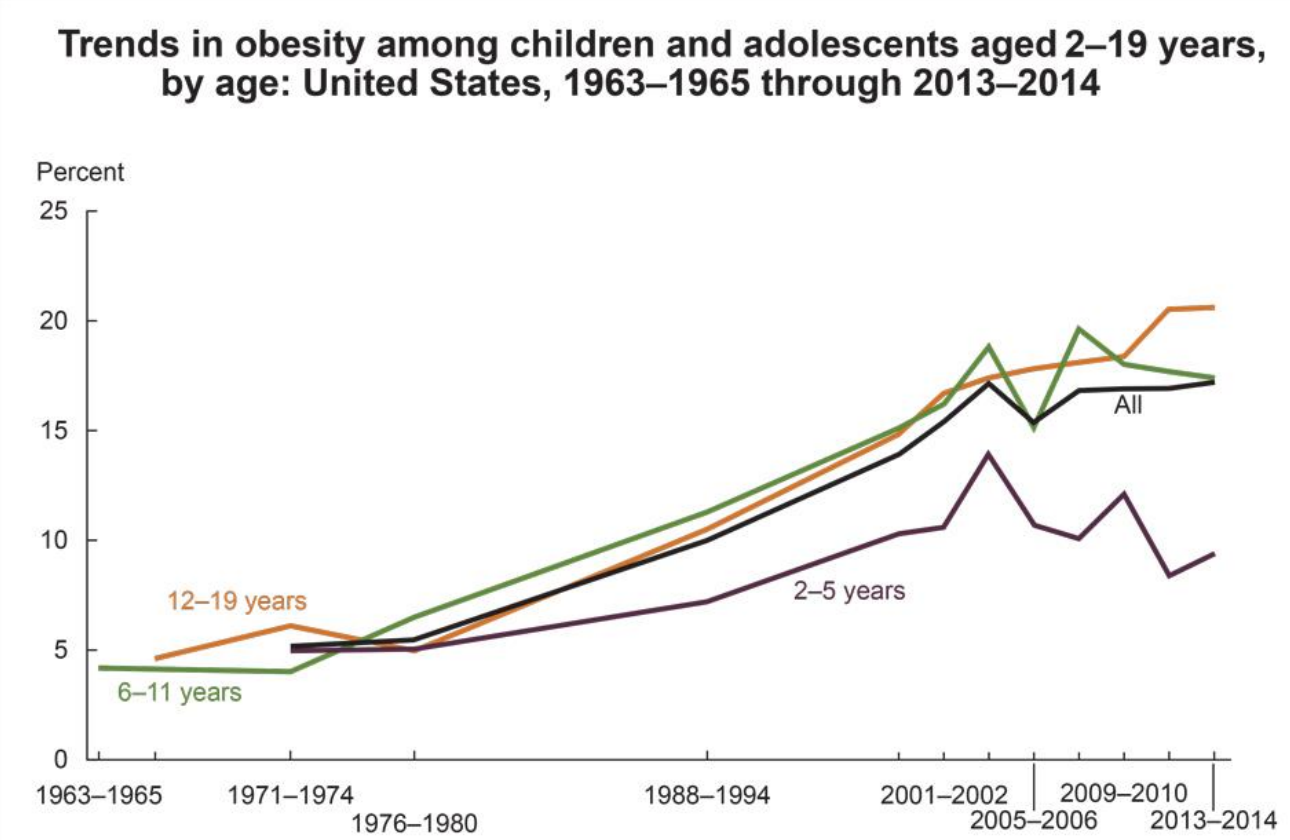
Food Insecurity

The Social Problem Causing Obesity

Fiona Baenziger

The Problem

Past 40 years, obesity increase of 8.5% in women & 7.6% in men



Graph from the National Institute of Diabetes and Digestive and Kidney Diseases

Economic Situation

In short: The rich are getting richer, the poor are getting poorer.

As economic inequalities have increased, so have inequalities in weight.

<u>Who is the most affected?</u> The middle class.

Hunger

a **personal, physical sensation** of discomfort

VS

Food Insecurity

a lack of available financial resources for food at the level of the household

Eating & Health Module Dataset

American Time Use Survey (ATUS) Eating & Health Module Files from 2014

From the US Bureau of Labor Statistics *on Kaggle*

- 37 variables
- 11.2K entries

KNOWN: Income, BMI, Changes in Income, Time Spent Eating, Height, Weight, Dietary Choices, Exercise Levels, etc…

UNKNOWN: Age, Sex, Role in Household

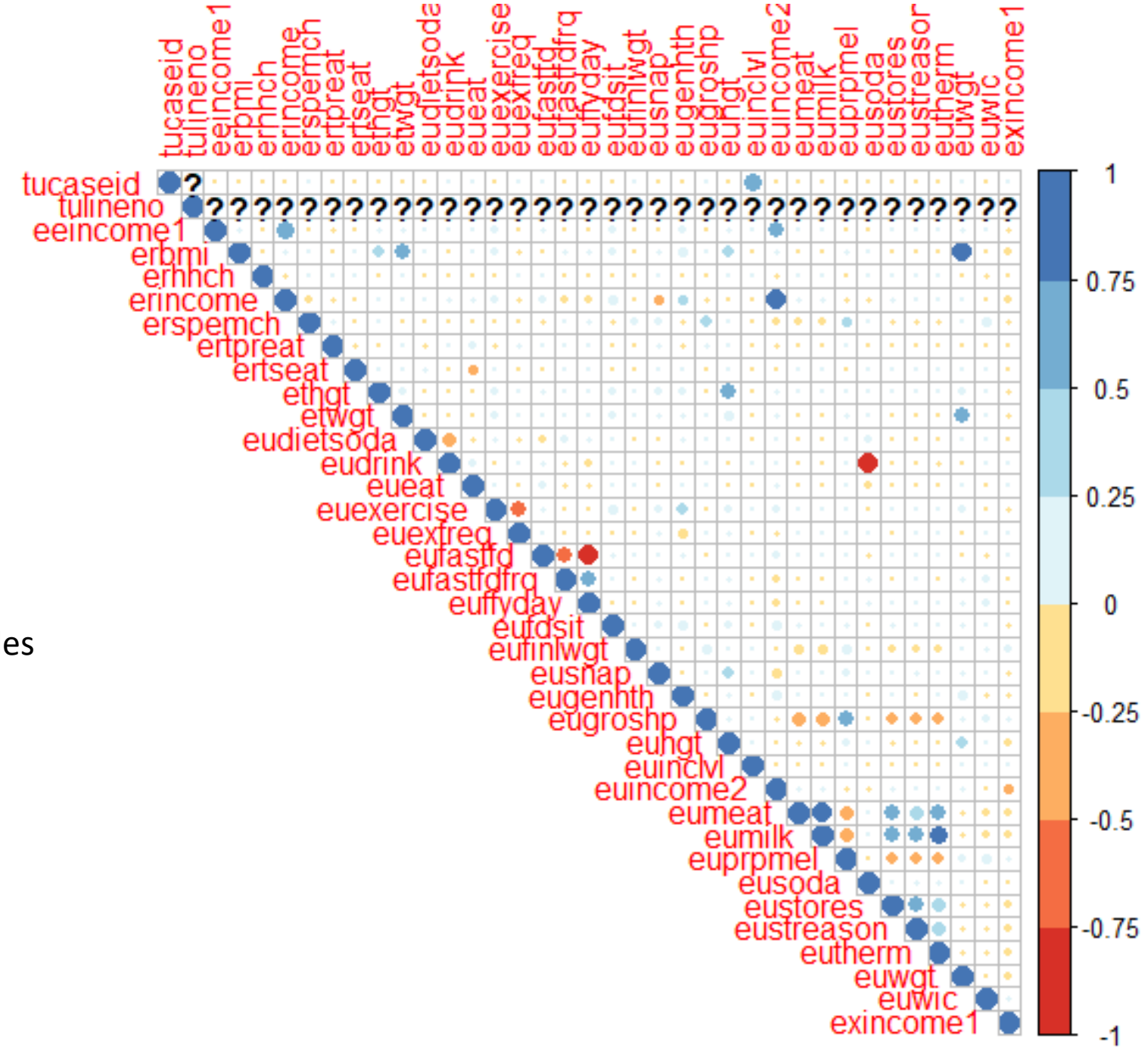
Correlation Plot

- Removed variables with |correlations| > 0.5
  - eudrink to ~~eusoda~~
  - erbmi to ~~euwgt~~

In the end…

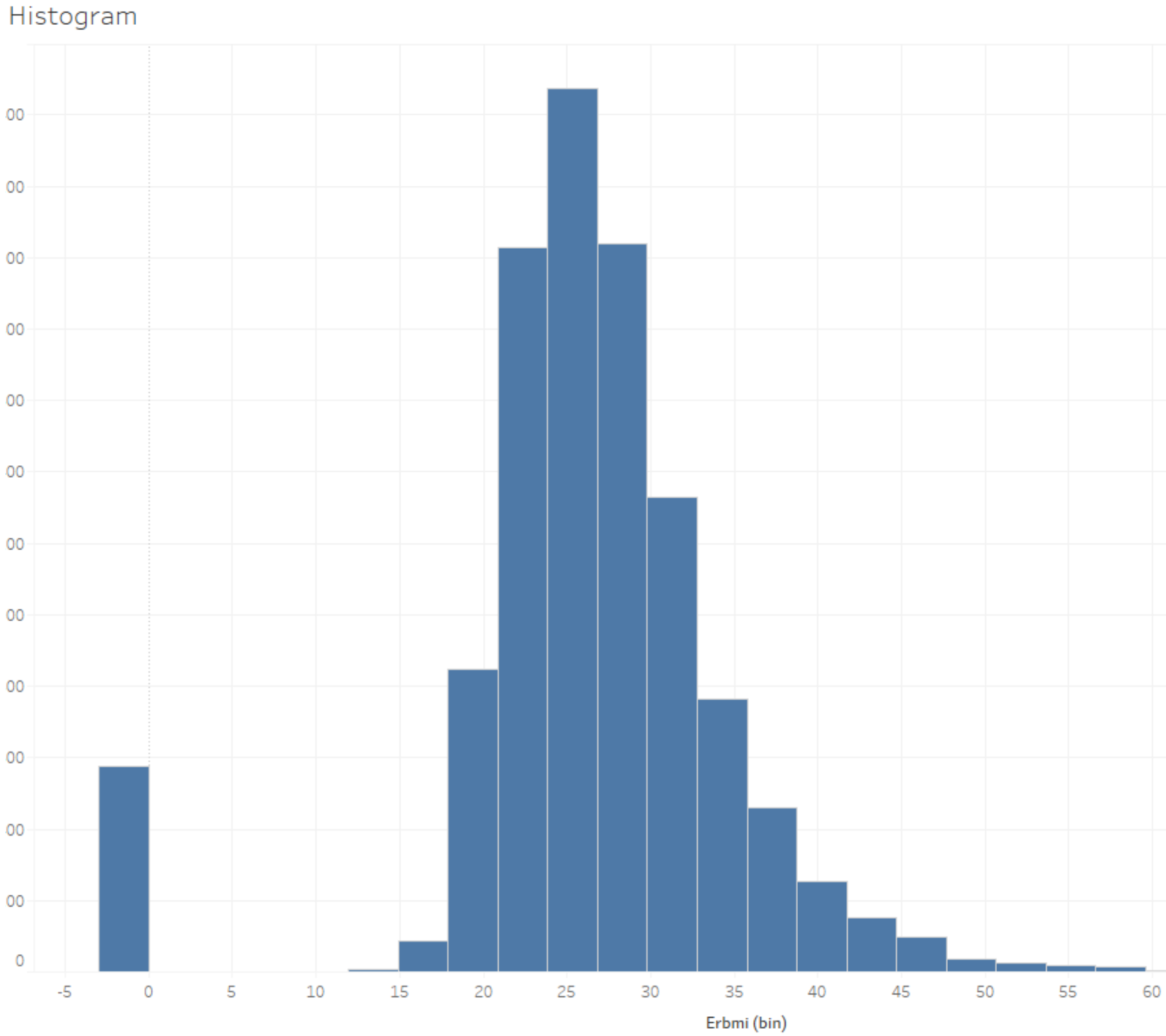
- Manually removed 15 variables
  - Repetitive information
  - Correlated
  - Irrelevant

- So… 22 variables.



The Target Variable

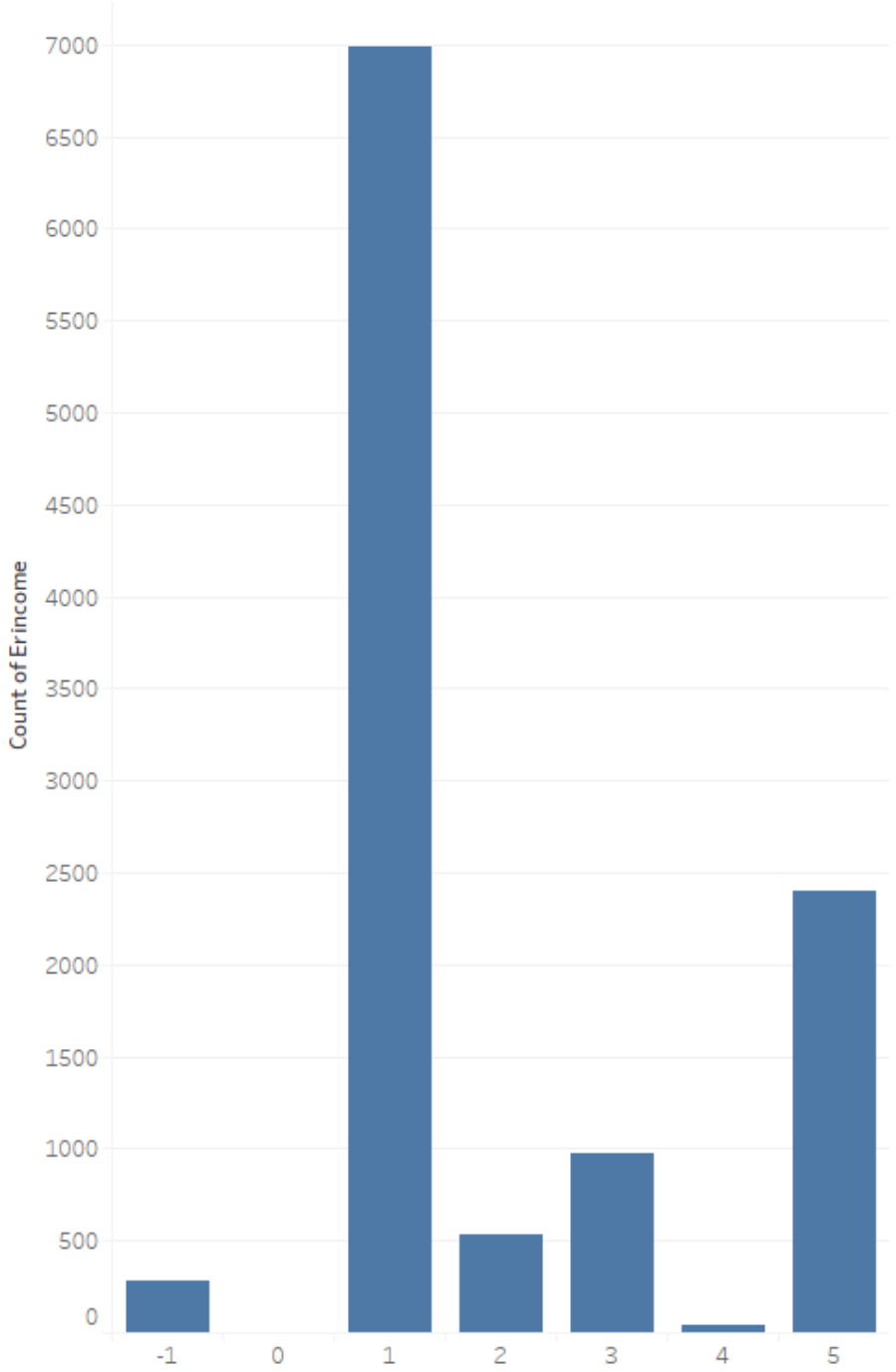
- Body Mass Index (BMI)
  - A weight-to-height ratio
- BMI has 575 missing values
- Attempted Binning BMI:
  - Underweight = < 18.5
  - Normal Weight = 18.5 – 24.9
  - Overweight = 25 – 29.9
  - Obesity = > 30
- Used Continuous BMI value



Income - Variable of Interest

- Based on number of individuals in the household
- Classified as percentage above or below poverty line
- Very imbalanced

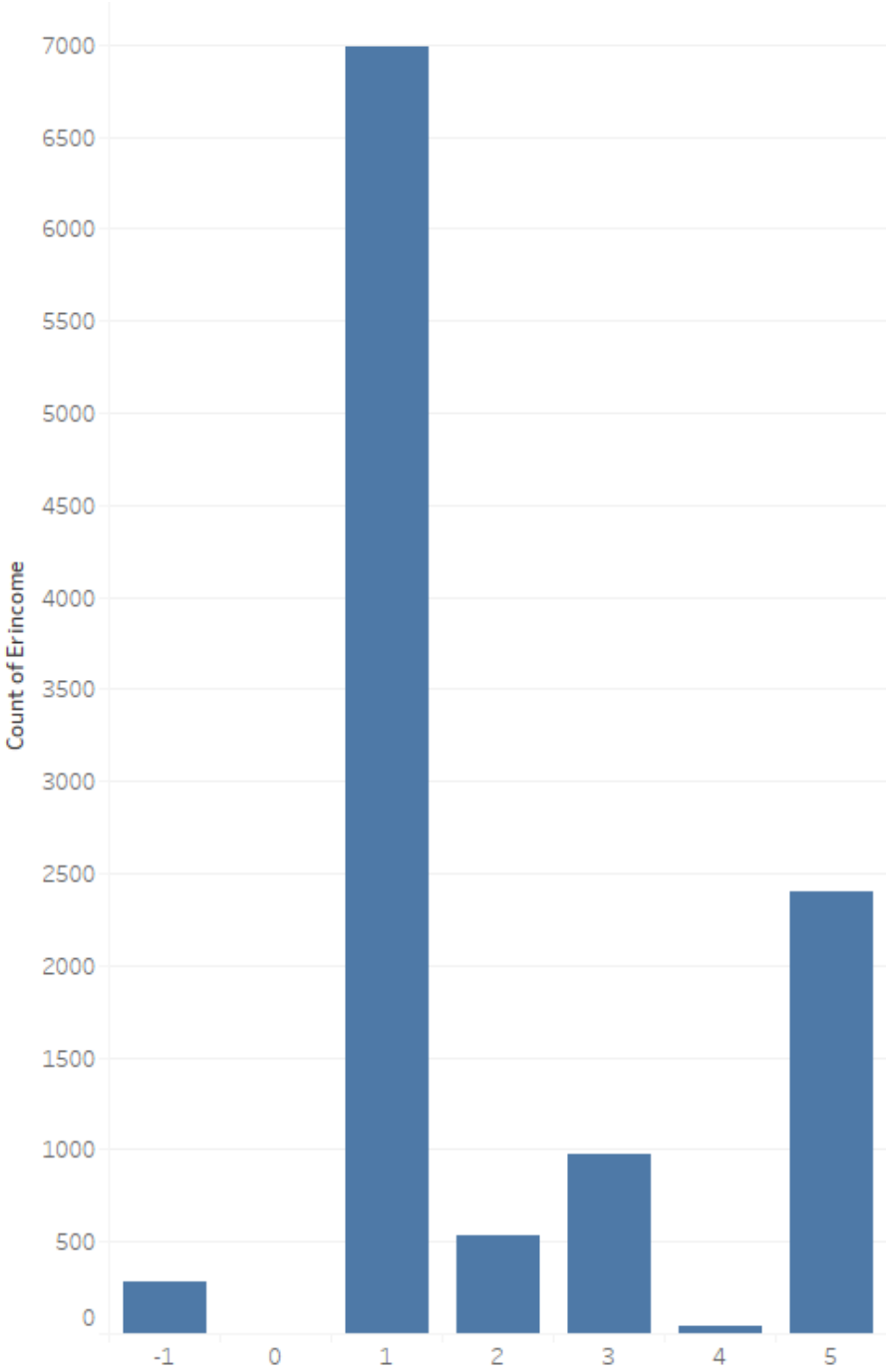
Valid Entries:	1	Income > 185% of poverty threshold
	2	Income < = 185% of poverty threshold
	3	130% of poverty threshold < Income<185% of poverty threshold
	4	Income > 130% of poverty threshold
	5	Income <= 130% of poverty threshold

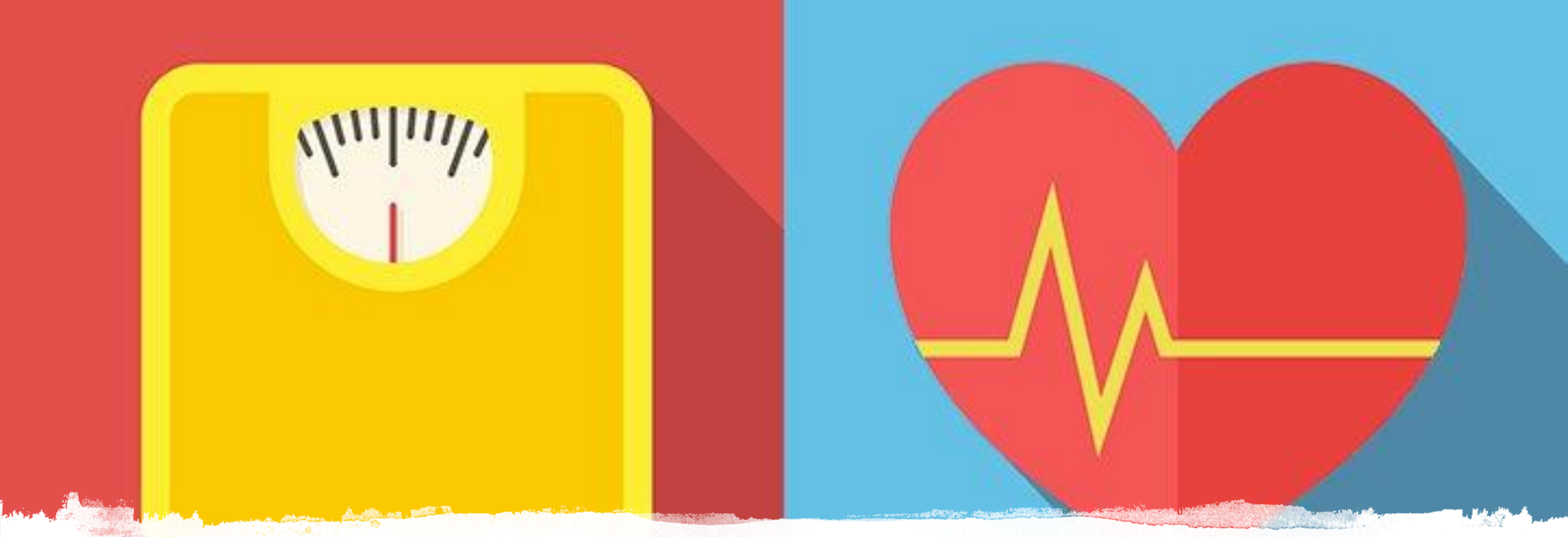


Income - Variable of Interest (cont.)

- Attempted Binning:
  - Class Once: Income > 130 % of Poverty Threshold ~ 7000
  - Class Two: Income <= 130 % of Poverty Threshold ~ 2500+1000+500 = ~4000

- Undersampling of Class One
  - Number of samples *decreased* from 10,637 to 3,728
  - Significantly less samples, but now even sampling of those far above poverty line and those who are not





Feature Selection

- Stepwise Recursion Backwards
- Wrapper Select
- Univariate Feature Selection – Mutual Info
- Full-Blown Wrapper Select

Feature Selection – The Results

Not the Right Fit:

- Stepwise Recursion Backwards
- Full-Blown Wrapper Select

Could be the Right Fit:

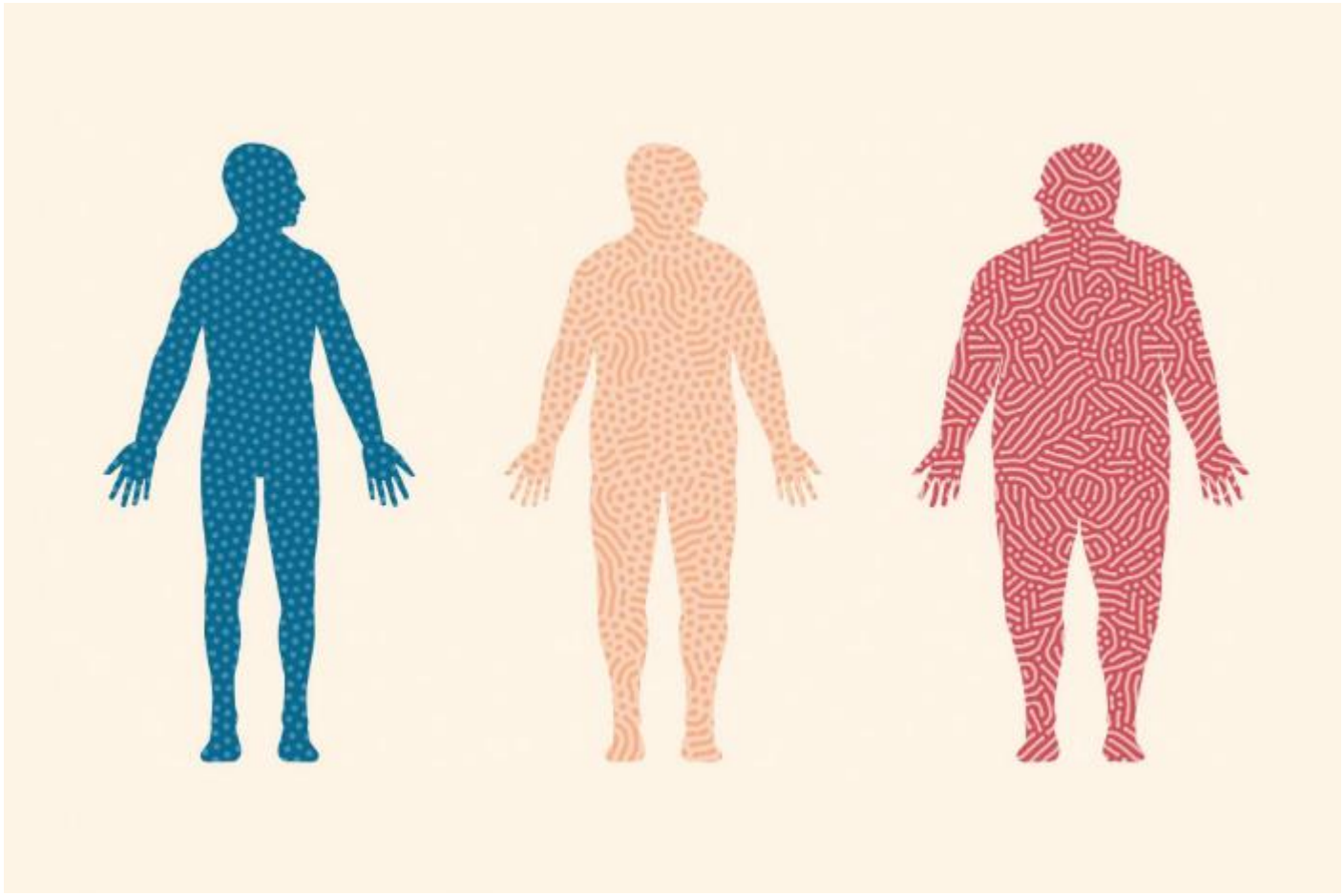
- Univariate Feature Selection – Mutual Info

The Right Fit

- ***Wrapper Select via Support Vector Regression (SVR)***

Wrapper Select Feature Selection – The Results

- *eugenhth*
  - In general, would you say that your physical health was excellent, very good, good, fair, or poor?
- *euexercise*
  - In the last 7 days, have you participated in any physical activities?
- *euffyday*
  - Did you purchase any prepared food from a deli, carry-out, delivery food, or fast food yesterday?
- *euwic*
  - In the last 30 days, did you or any member of your household receive benefits from the WIC program, that is, the Women, Infants, and Children program?
- *eudrink*
  - Not including plain water, were there any other times yesterday when you were drinking any beverages?
- *eueat*
  - Were there any times you were eating any meals or snacks yesterday, for example while you were doing something else?



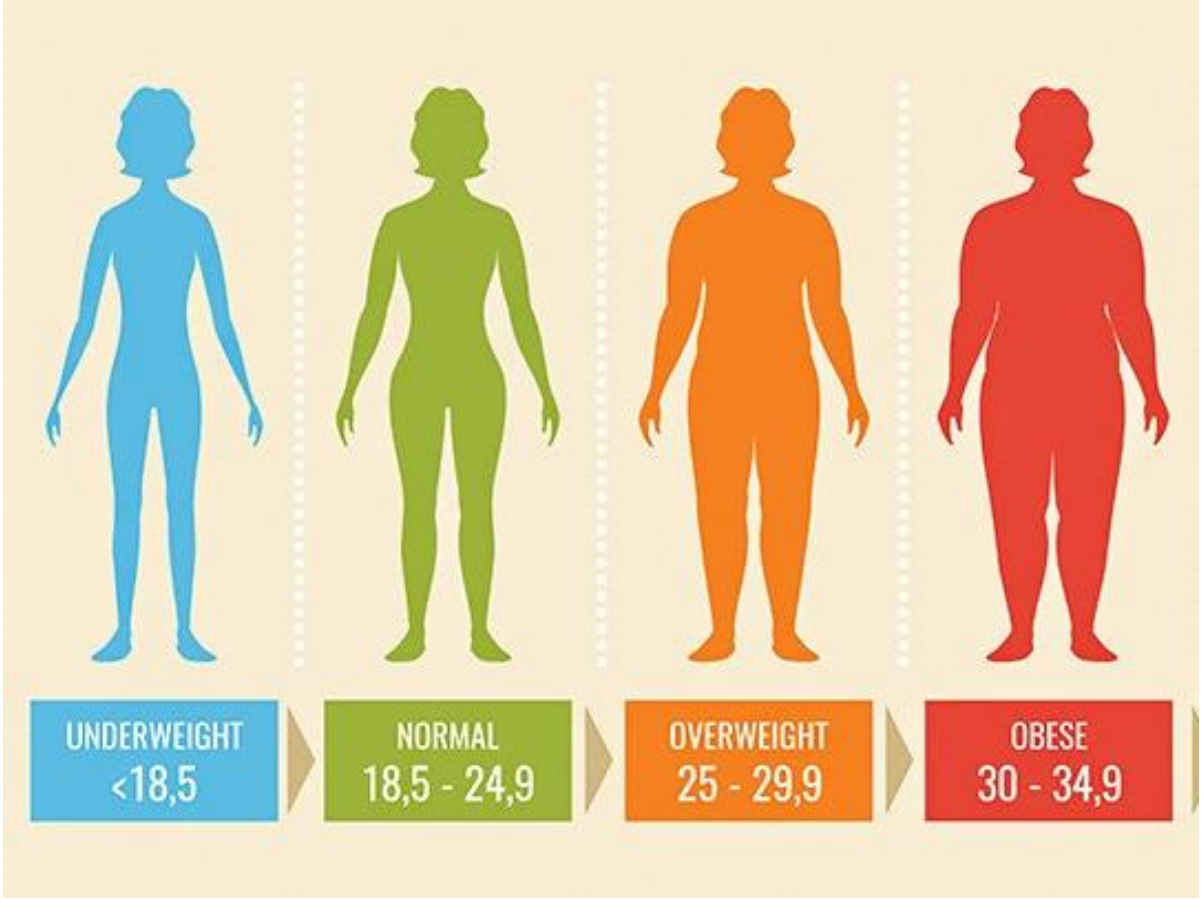
- Random Forests
- Gradient Boosting
- Ada Boost
- Neural Networks
- SVMs

Machine Learning Models

Understanding the RMSE Value

- Target values from 13 – 73.6, a range of **70.6**

- EX. RMSE = 5
  - 7% of target variable
  - Changes a whole CLASS



Machine Learning Model (1) – The Process

		WRAPPER SELECT			
Type	Measure	CROSS VALIDATION (CV = 5)	CROSS VALIDATION (CV = 7)	TEST/TRAIN SPLIT (Test = 30%)	TEST/TRAIN SPLIT (Test = 40%)
Random Forests	RMSE	5.95 (+/- 0.48)	5.51 (+/- 0.63)	5.6768	5.7820
	Expl. Var.	0.08 (+/- 0.05)	0.07 (+/- 0.10)	0.0521	0.0828
Gradient Boosting	**RMSE**	**5.42 (+/- 0.20)**	5.43 (+/- 0.56)	5.6830	5.7403
	Expl. Var.	**0.11 (+/- 0.05)**	0.13 (+/- 0.07)	-0.0364	0.0438
Ada Boost	RMSE	5.80 (+/- 0.35)	5.75 (+/- 0.38)	6.0065	5.6247
	Expl. Var.	0.07 (+/- 0.08)	0.06 (+/- 0.10)	0.0577	0.0866
Neural Networks	RMSE	5.62 (+/- 0.42)	5.46 (+/- 0.56)	5.2468	5.4021
	Expl. Var.	0.09 (+/- 0.04)	0.14 (+/- 0.05)	0.0753	0.1047
SVMs	RMSE	5.52 (+/- 0.65)	5.45 (+/- 0.75)	5.2695	5.7404
	Expl. Var.	0.13 (+/- 0.03)	0.13 (+/- 0.07)	0.1070	0.1467

Machine Learning Model (2) – The Process

		WRAPPER SELECT			
Type	Measure	CROSS VALIDATION (CV = 5)	CROSS VALIDATION (CV = 7)	TEST/TRAIN SPLIT (Test = 30%)	TEST/TRAIN SPLIT (Test = 40%)
Random Forests	RMSE	5.95 (+/- 0.48)	5.51 (+/- 0.63)	5.6768	5.7820
	Expl. Var.	0.08 (+/- 0.05)	0.07 (+/- 0.10)	0.0521	0.0828
Gradient Boosting	RMSE	5.42 (+/- 0.20)	5.43 (+/- 0.56)	5.6830	5.7403
	Expl. Var.	0.11 (+/- 0.05)	0.13 (+/- 0.07)	-0.0364	0.0438
Ada Boost	RMSE	5.80 (+/- 0.35)	5.75 (+/- 0.38)	6.0065	5.6247
	Expl. Var.	0.07 (+/- 0.08)	0.06 (+/- 0.10)	0.0577	0.0866
Neural Networks	**RMSE**	5.62 (+/- 0.42)	5.46 (+/- 0.56)	**5.2468**	5.4021
	Expl. Var.	0.09 (+/- 0.04)	0.14 (+/- 0.05)	**0.0753**	0.1047
SVMs	RMSE	5.52 (+/- 0.65)	5.45 (+/- 0.75)	5.2695	5.7404
	Expl. Var.	0.13 (+/- 0.03)	0.13 (+/- 0.07)	0.1070	0.1467

Machine Learning Model – The Result

- They perform similarly!
- The best models:

		WRAPPER SELECT		
Type	Measure	CROSS VALIDATION (CV = 5)	TEST/TRAIN SPLIT (Test = 30%)	Runtime
Gradient Boosting	RMSE	5.42 (+/- 0.20)	----------	0.5897
	Expl. Var.	0.11 (+/- 0.05)		
Neural Networks	RMSE	----------	5.2468	1.0676
	Expl. Var.		0.0753	

- How can we improve? ***More testing.***

Final Model – Gradient Boosting

- Cross Validation – # of Folds = 5

RMSE	Expl. Var.	Runtime
5.42 (+/- 0.20)	0.11 (+/- 0.05)	0.5897

- Features – 5 of 22 Selected:
  - *erincome* – Relationship between income and poverty level
  - *euexercise* – In the last 7 days, have you participated in physical activity?
  - *eugenhth* – In general, would you say that your physical health was…?
  - *eumeat* – In the last 7 days, did you prepare any meals with meat, poultry, …?
  - *eumilk* – In the last 7 days, did you drink or serve unpasteurized or raw milk?

Conclusions

- Income, as researched, does appear to be a factor in predicting obesity
- General awareness for health is also a factor and are you spending time on your health
- Specific dietary factors contributing to obesity

The Future

- To clean up the model and work on improving the predictive accuracy
- To obtain a more specific income value as opposed to a class
- To incorporate Age, Sex, and Role in Household into the features
- To focus more specifically on the differences in obesity factors between income levels