



Fad Diet Tweets

Analyzing Sentiment using Natural Language Processing
and Predicting Obesity with Machine Learning Techniques



Vinh Tran - Final Capstone, 7/29/2019
https://github.com/vinhttran/fad_diets_prediction

Background

- \$72 Billion diet and weight loss market
- 40% of US adults are obese (93.3 million people)
- Obesity shown to have a network effect

BMI (kg/m^2)		Classification ^[18]
from	up to	
	18.5	underweight
18.5	25.0	normal weight
25.0	30.0	overweight
30.0	35.0	class I obesity
35.0	40.0	class II obesity
40.0		class III obesity

In this project, tweets about fad diets are used to study the relationship between fad diet twitter sentiment and obesity prevalence rates in US cities

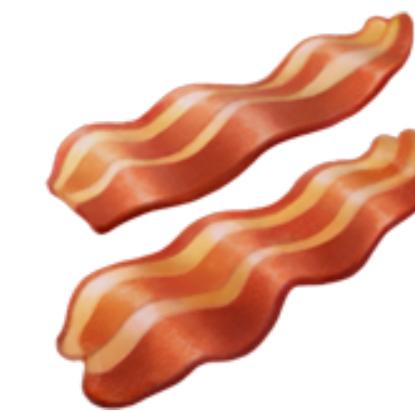
Fad Diets



Keto



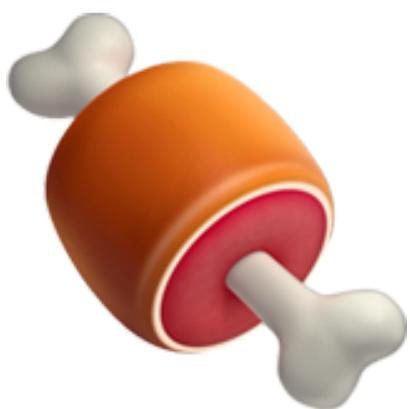
Whole 30



Atkins



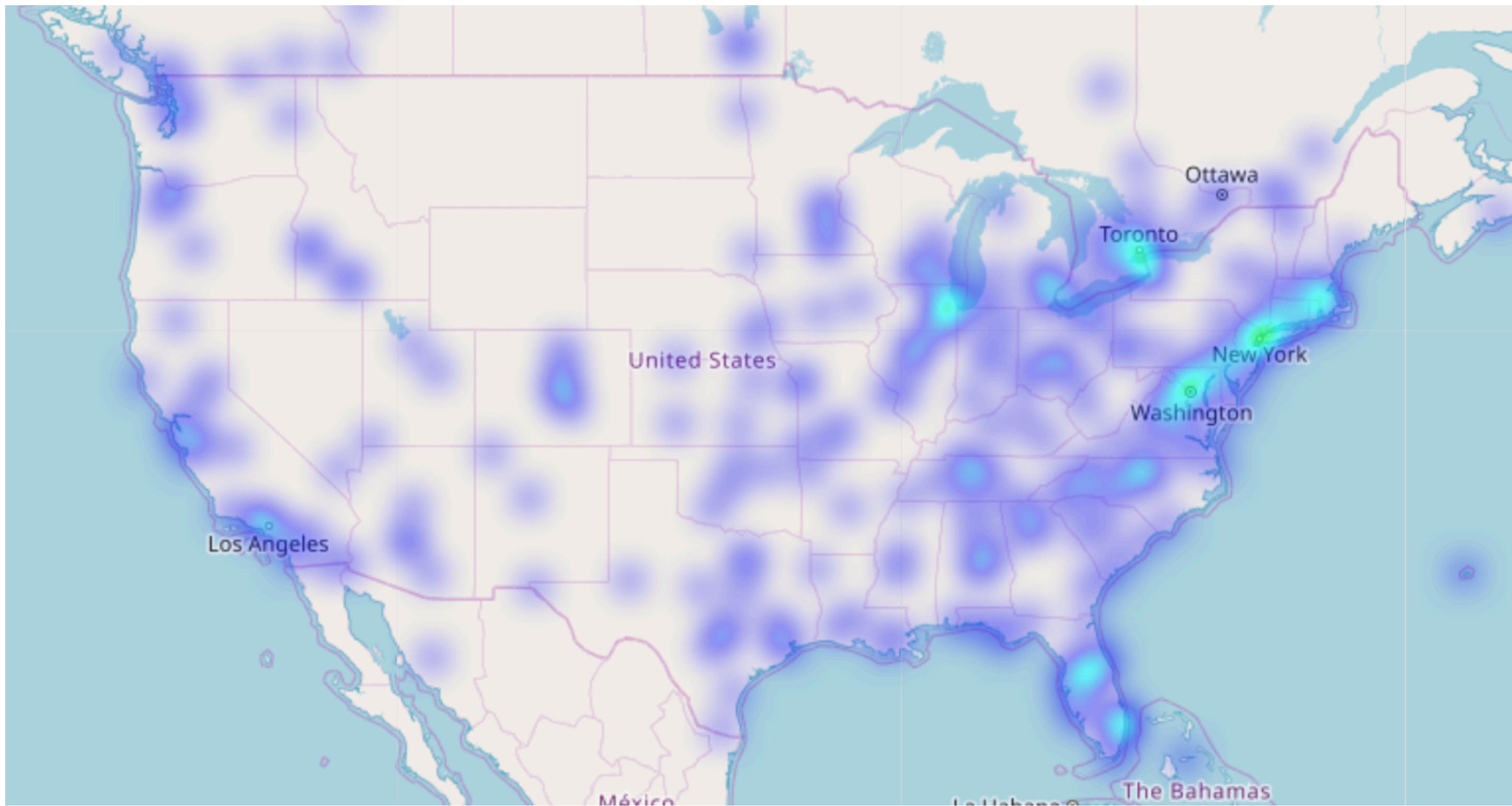
Gluten Free



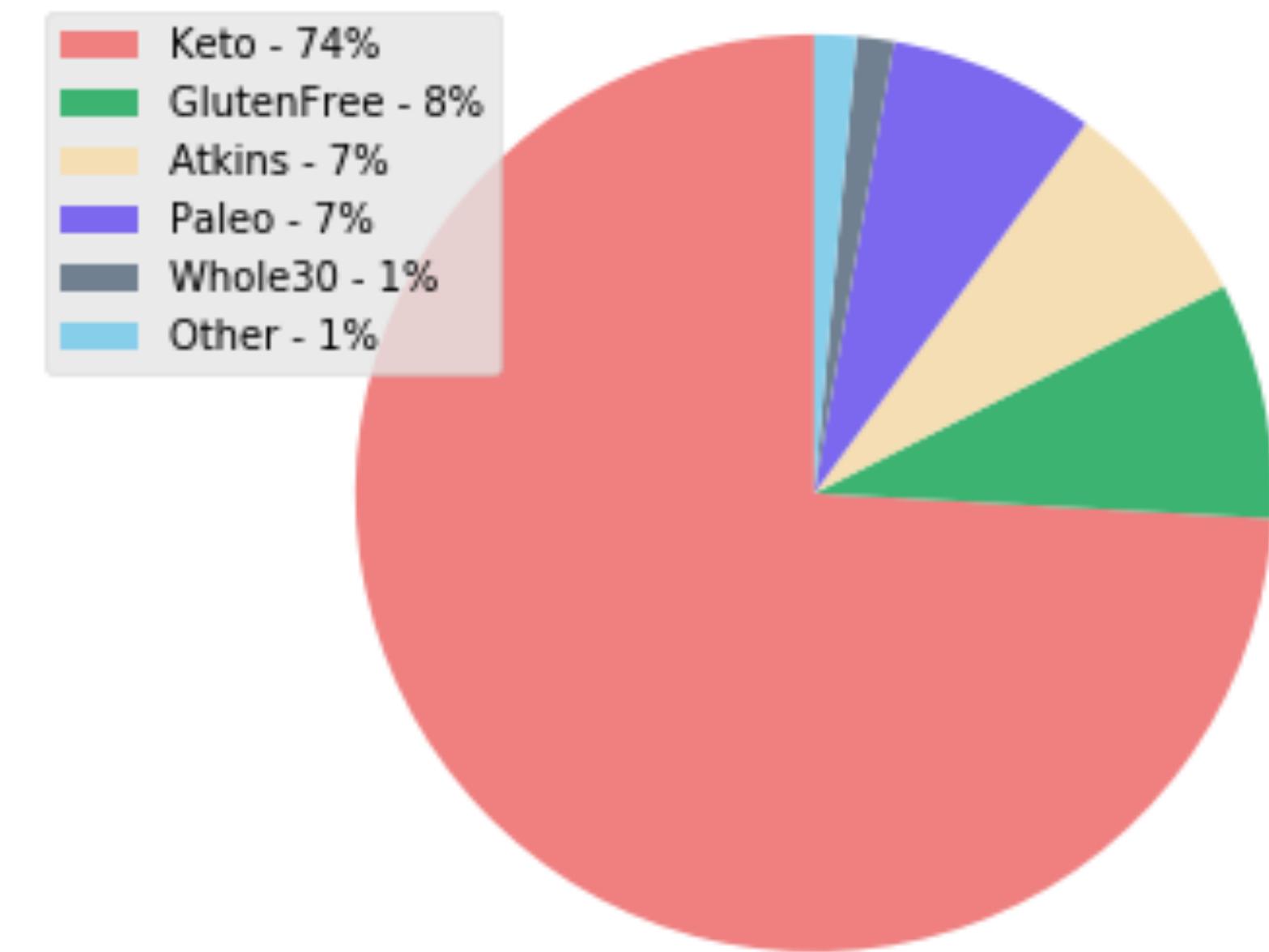
Paleo

Twitter Data

Heatmap of Tweets

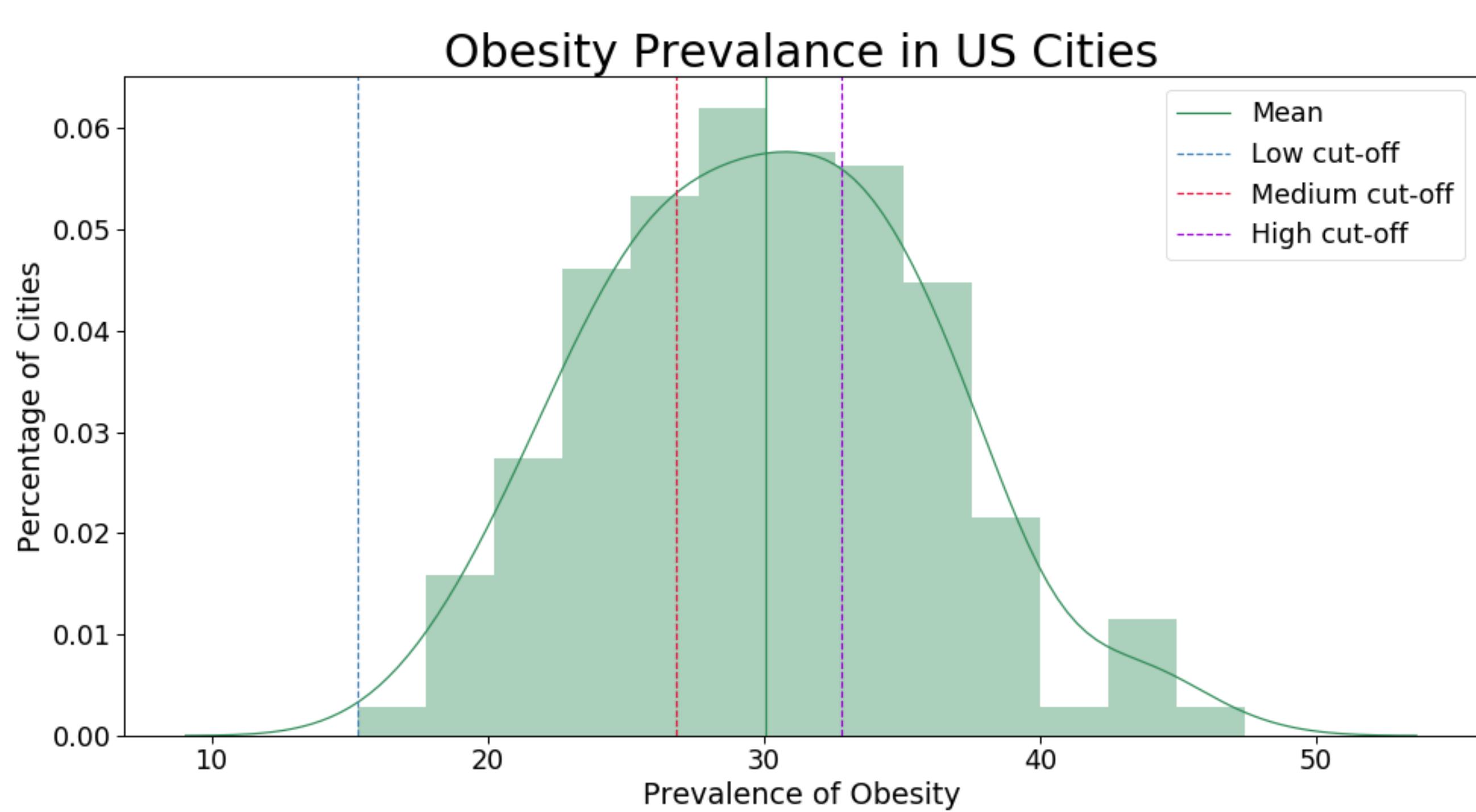


Popularity of Diets

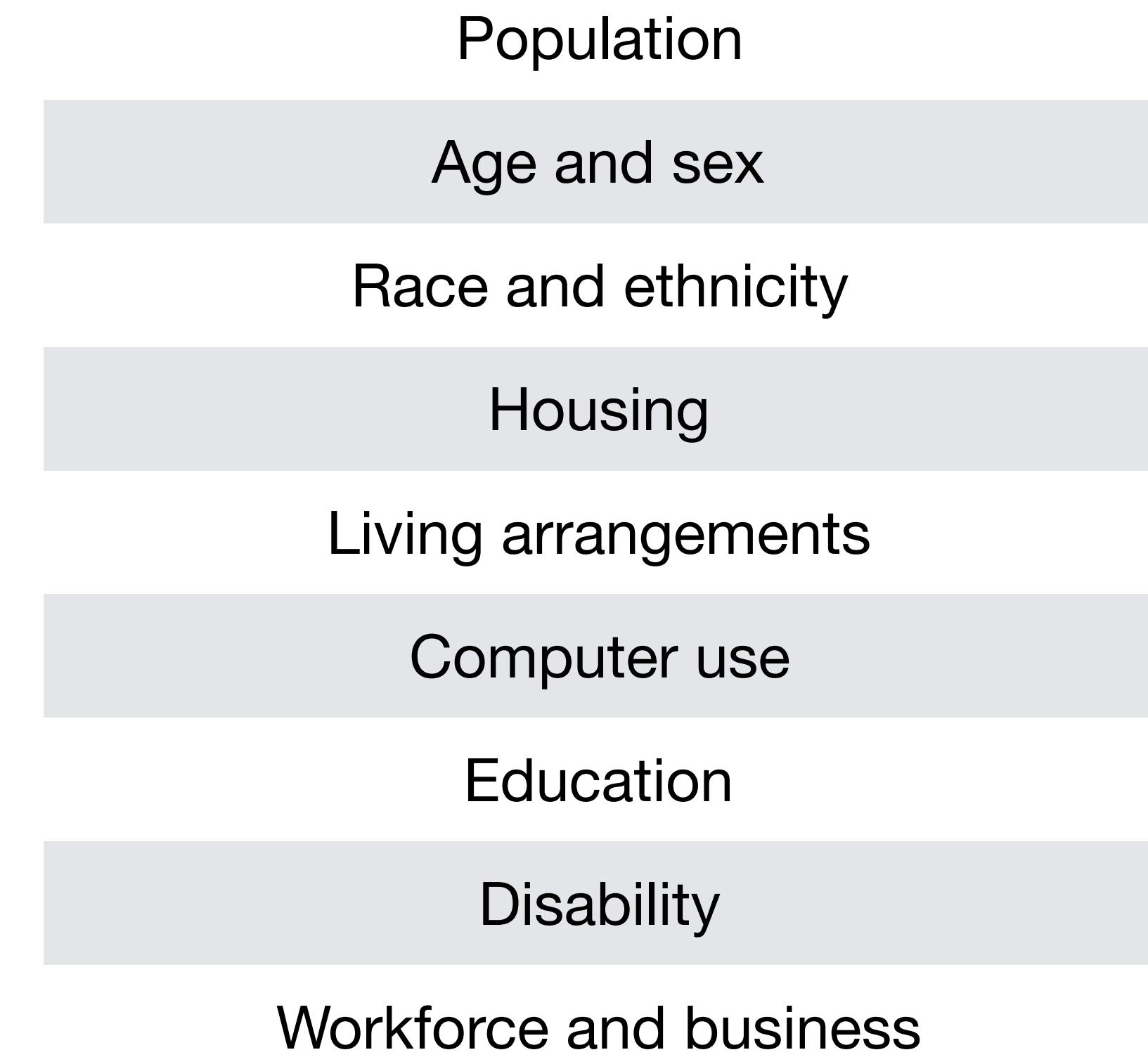


CDC and Census Data

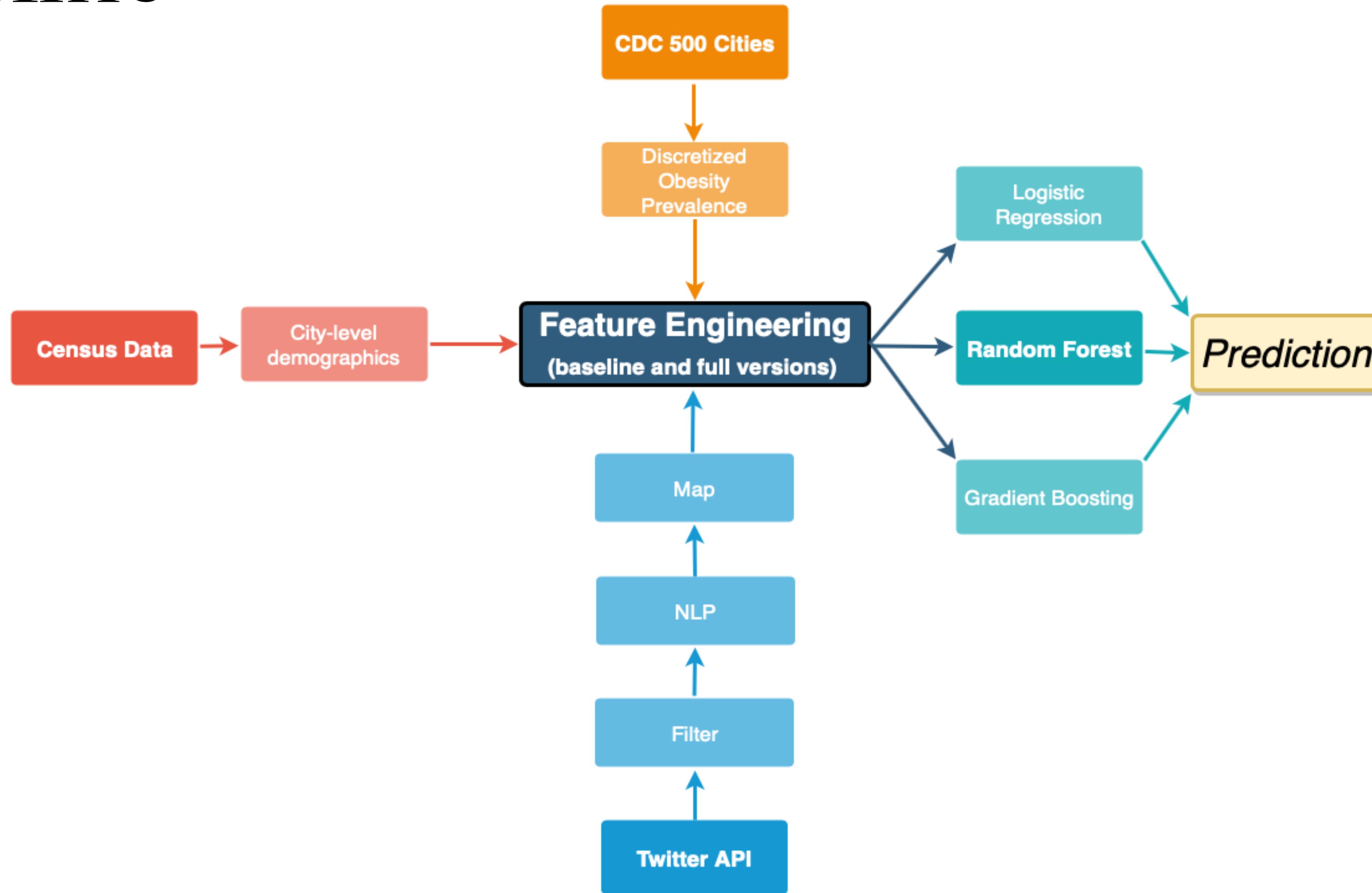
CDC 500 Cities



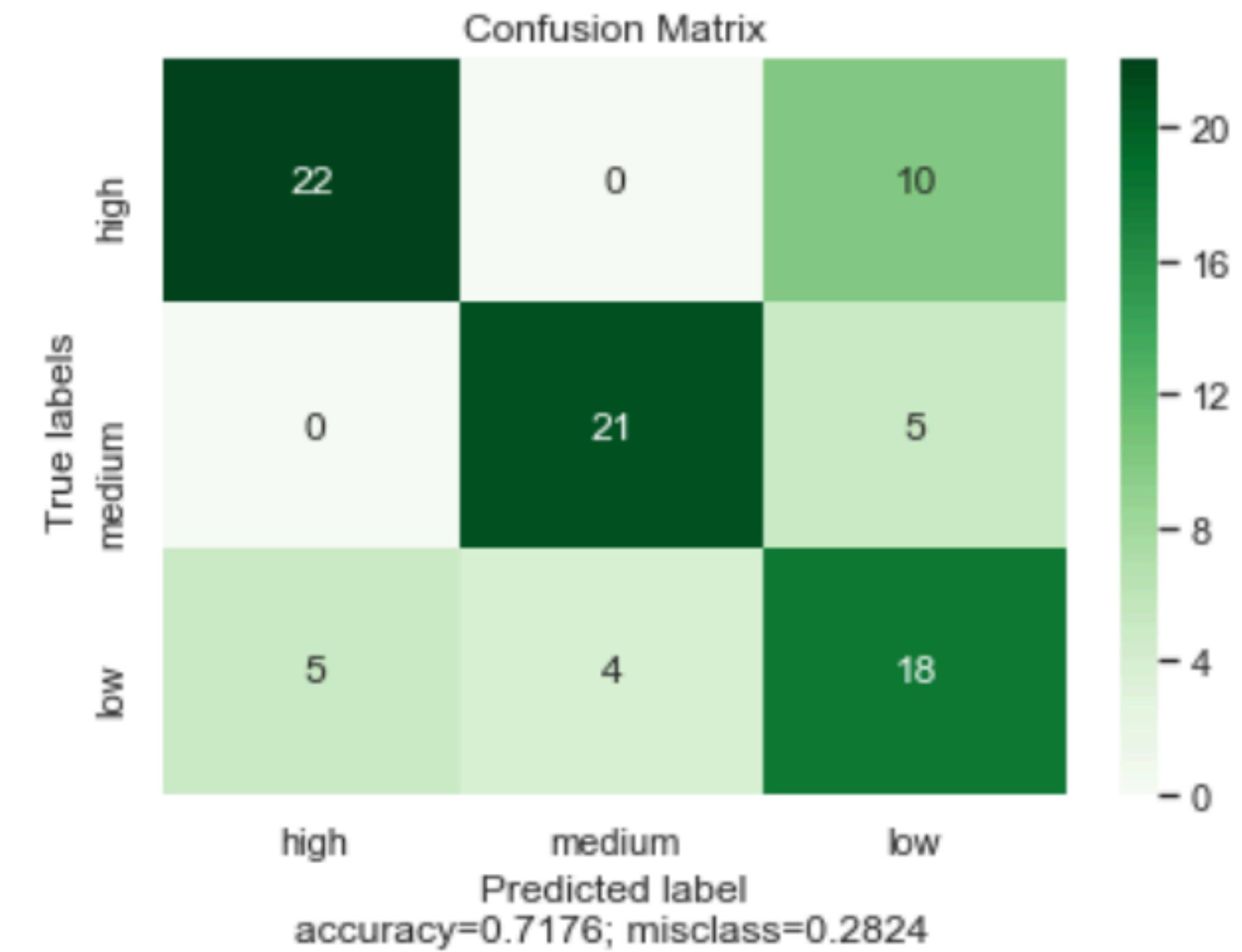
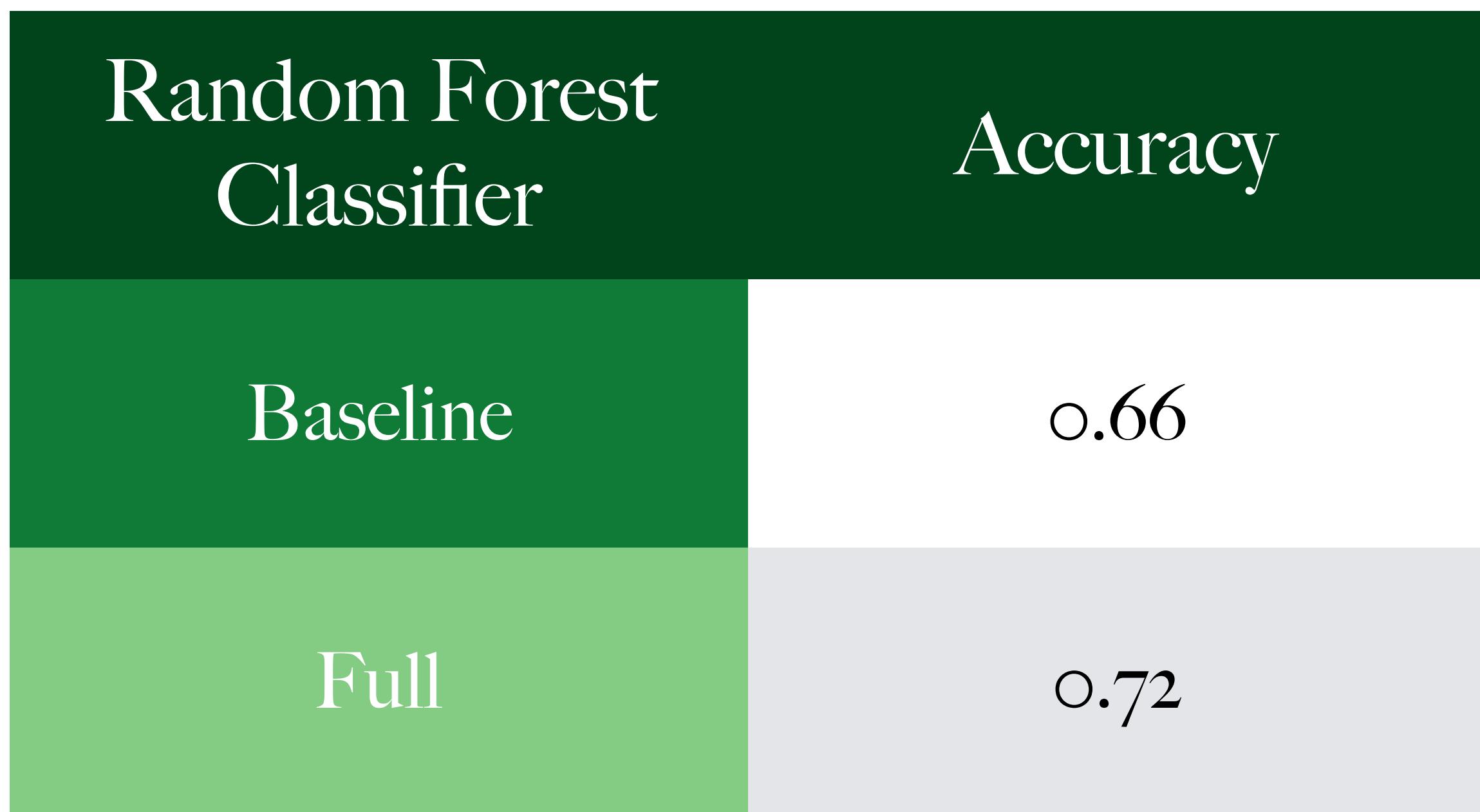
American Community
Surveys 2017



Pipeline



Modeling

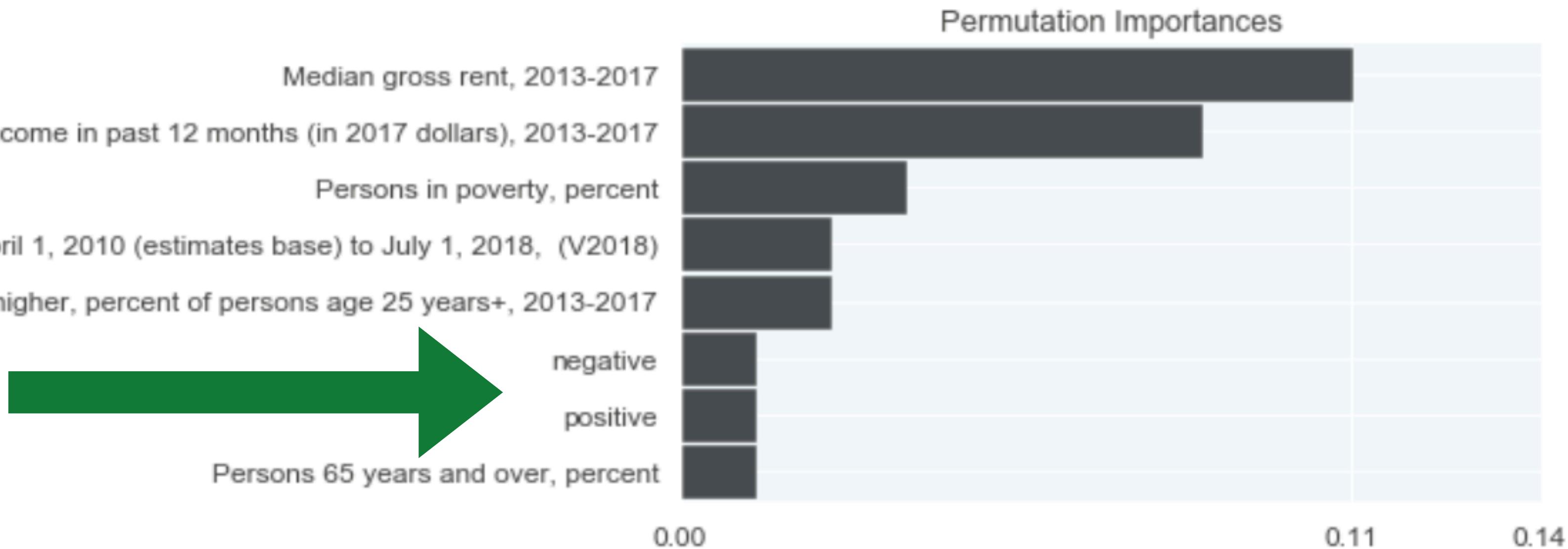


Feature Importance

Baseline



Full



Conclusions

- Adding positive and negative sentiment scores improved random forest model accuracy by 6%
- Cities where tweets are more highly positive and negative are important features in predicting high obesity prevalence rates
- More evidence of a network effect
- Likely persons with obesity have more extreme emotions related to fad diets, leading to more positive and negative tweets



Q & A



References

1. <https://www.businesswire.com/news/home/20190225005455/en/72-Billion-Weight-Loss-Diet-Control-Market>
2. <https://www.hsph.harvard.edu/nutritionsource/an-epidemic-of-obesity/>
3. <https://www.cdc.gov/obesity/data/adult.html>
4. <https://www.nejm.org/doi/full/10.1056/NEJMsa066082>
5. <https://www.mayoclinic.org>
6. <https://www.cdc.gov/500cities/index.htm>)
7. <https://www.census.gov/quickfacts/fact/table/mobilecityalabama,US/PST045218>
8. https://eli5.readthedocs.io/en/latest/blackbox/permuation_importance.html
9. <https://www.niemanlab.org/2019/06/twitter-is-turning-off-location-data-on-tweets-a-small-win-for-privacy-but-a-small-loss-for-journalists-and-researchers/>)

Future

Limitations

- Select fad diets analyzed
- Fads are fleeting
- Twitter recently removed their precise location tagging
- Limited to cities, no rural representation

Future

- Link BMI scores to individual tweets
- See if findings hold outside US
- Dashboard to capture tweets in real-time as part of a model that continually predicts obesity
- Examine other social media

Tools Used

- Twitter API
- Census and CDC data
- Natural Language Processing (TextBlob)
- Logistic Regression, Random Forest, Gradient Boosting
- Python Libraries: Pandas, NumPy, Sklearn, Matplotlib, Seaborn, Folium, Tweepy, GridSearchCV

Fad Diet Details

Popular diets often promoting/restricting certain foods or food groups, promising various improved health-outcomes (quickly), hard to sustain, and only anecdotally effective

Diet	Description
Keto	Low-carb, high-fat. The idea is to put your body into a metabolic state called ketosis.
Gluten Free	Exclude the protein gluten found in many grains. In non-celiac or wheat-sensitive persons its purported benefits are improved health, weight loss and increased energy.
Atkins	Low-carb. Promoted by the physician Dr. Robert Atkins, who wrote a best-selling book about it in 1972.
Paleo	Also known as the caveman diet. Based on foods similar to what might have been eaten during the Paleolithic era with a diet high in lean meats, fish, fruits, vegetables, nuts and seeds. Limits dairy products, legumes and grains.
Whole 30	For 30 days, no sugar, alcohol, grains, legumes, dairy or treats in general are allowed. Moderate amounts of meat, seafood and eggs; vegetables aplenty; some fruit; and natural fats such as nuts and avocado.

Features

"Population, percent change - April 1, 2010 (estimates base) to July 1, 2018, (V2018)",

"Persons 65 years and over, percent",

"White alone, percent",

"Black or African American alone, percent",

"Asian alone, percent",

"Hispanic or Latino, percent",

"Veterans, 2013-2017",

"Median gross rent, 2013-2017",

"Bachelor's degree or higher, percent of persons age 25 years+, 2013-2017",

"Persons without health insurance, under age 65 years, percent",

"In civilian labor force, total, percent of population age 16 years+, 2013-2017",

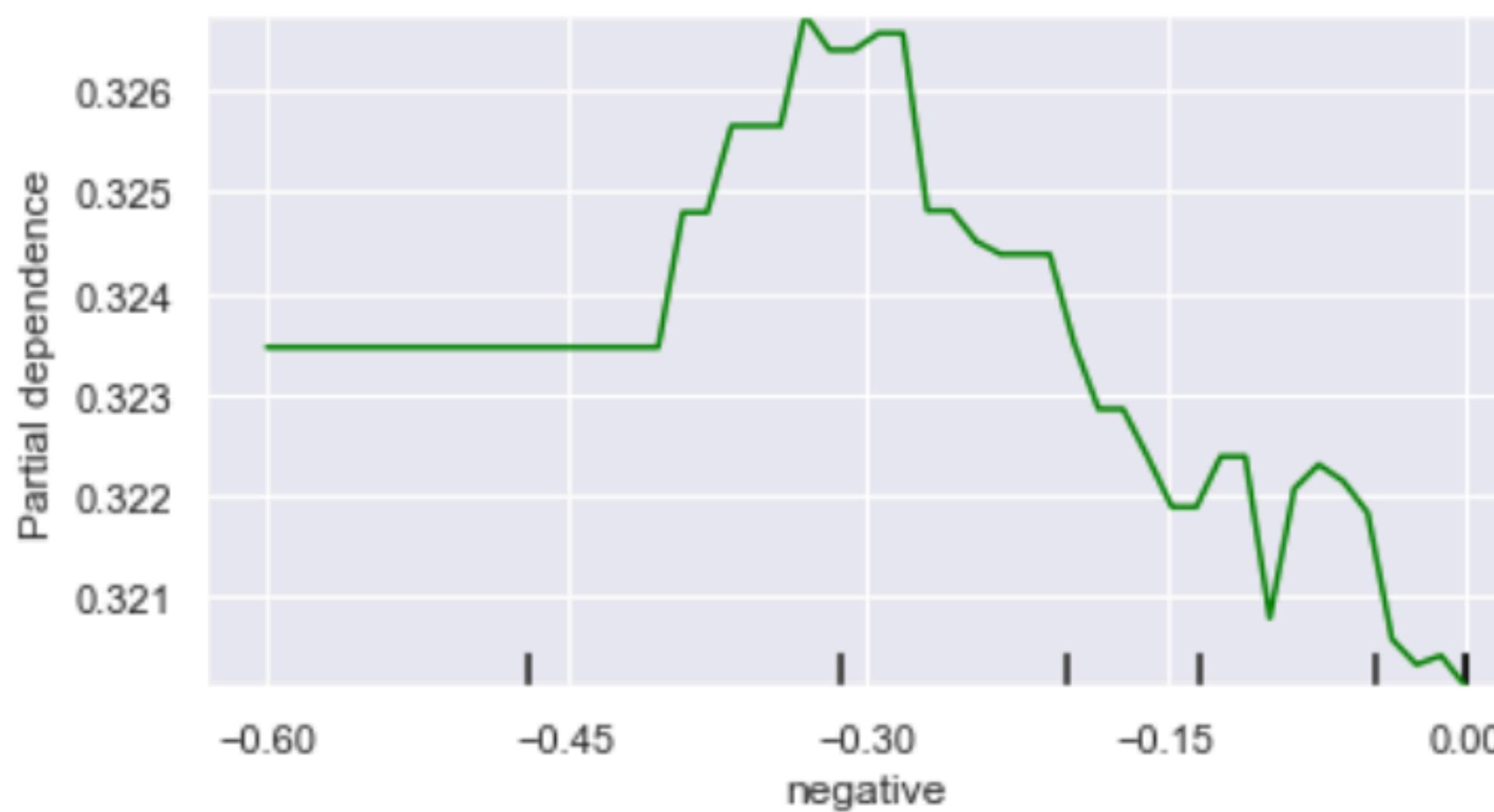
"Mean travel time to work (minutes), workers age 16 years+, 2013-2017",

"Per capita income in past 12 months (in 2017 dollars), 2013-2017",

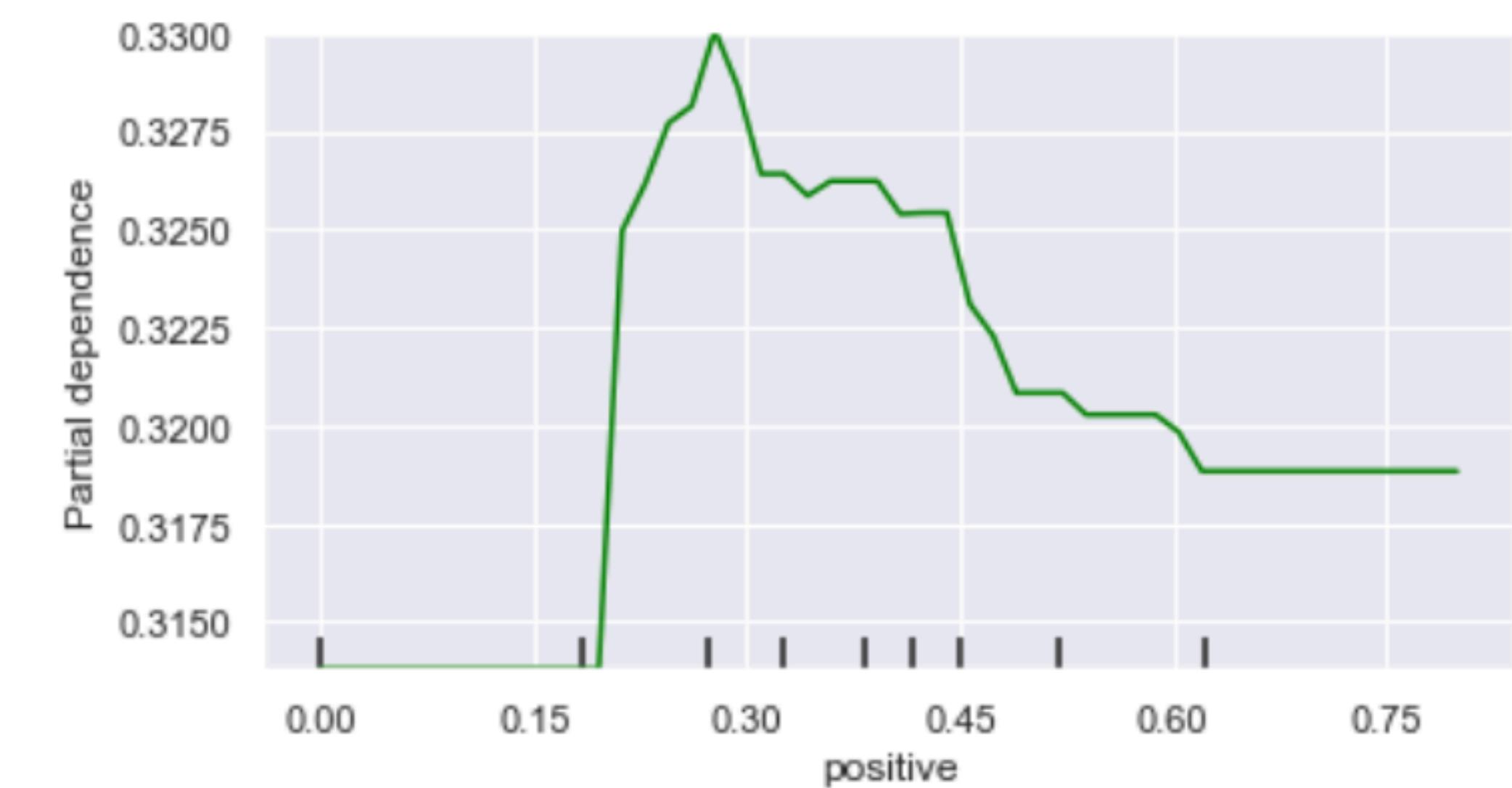
"Persons in poverty, percent"

Features: Partial Dependence

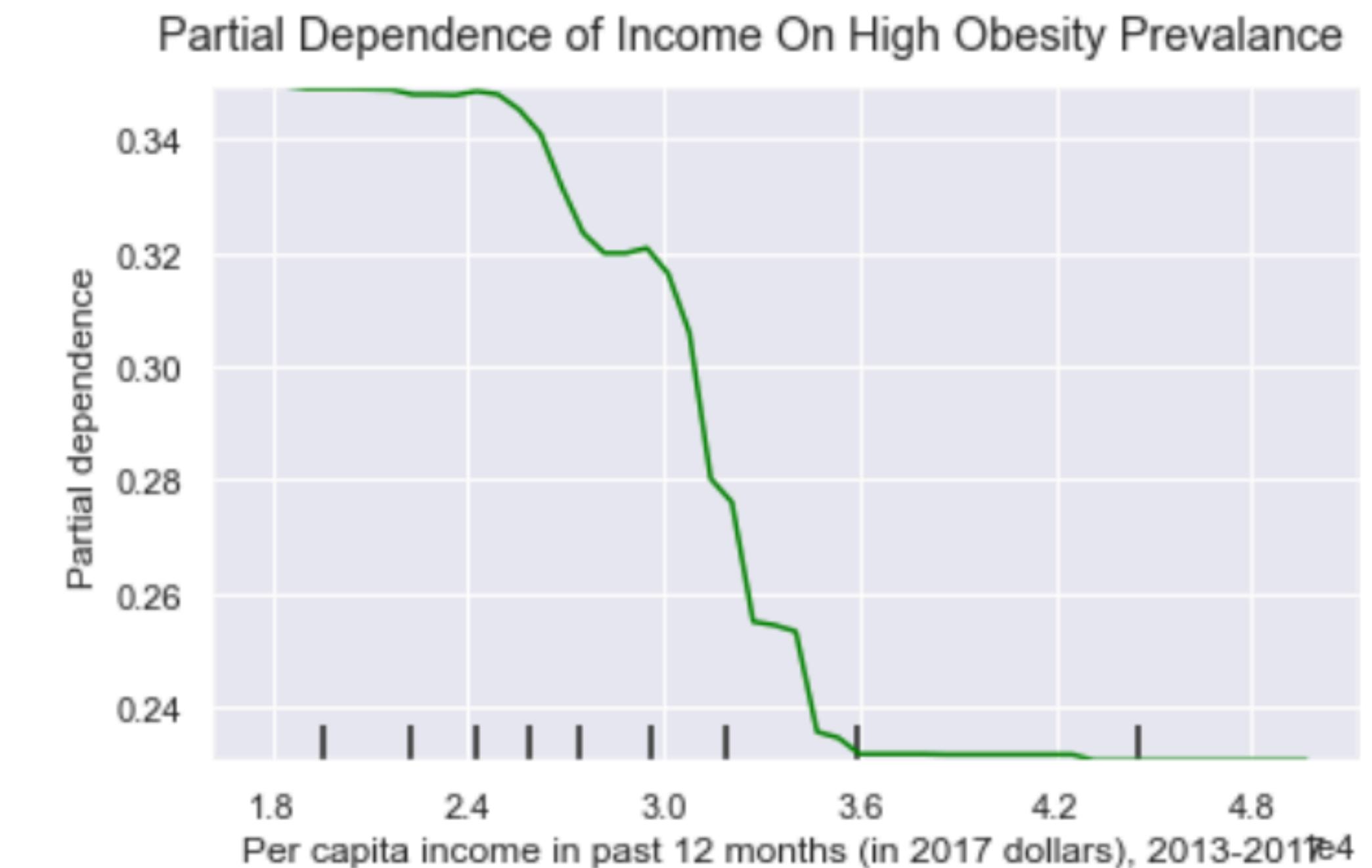
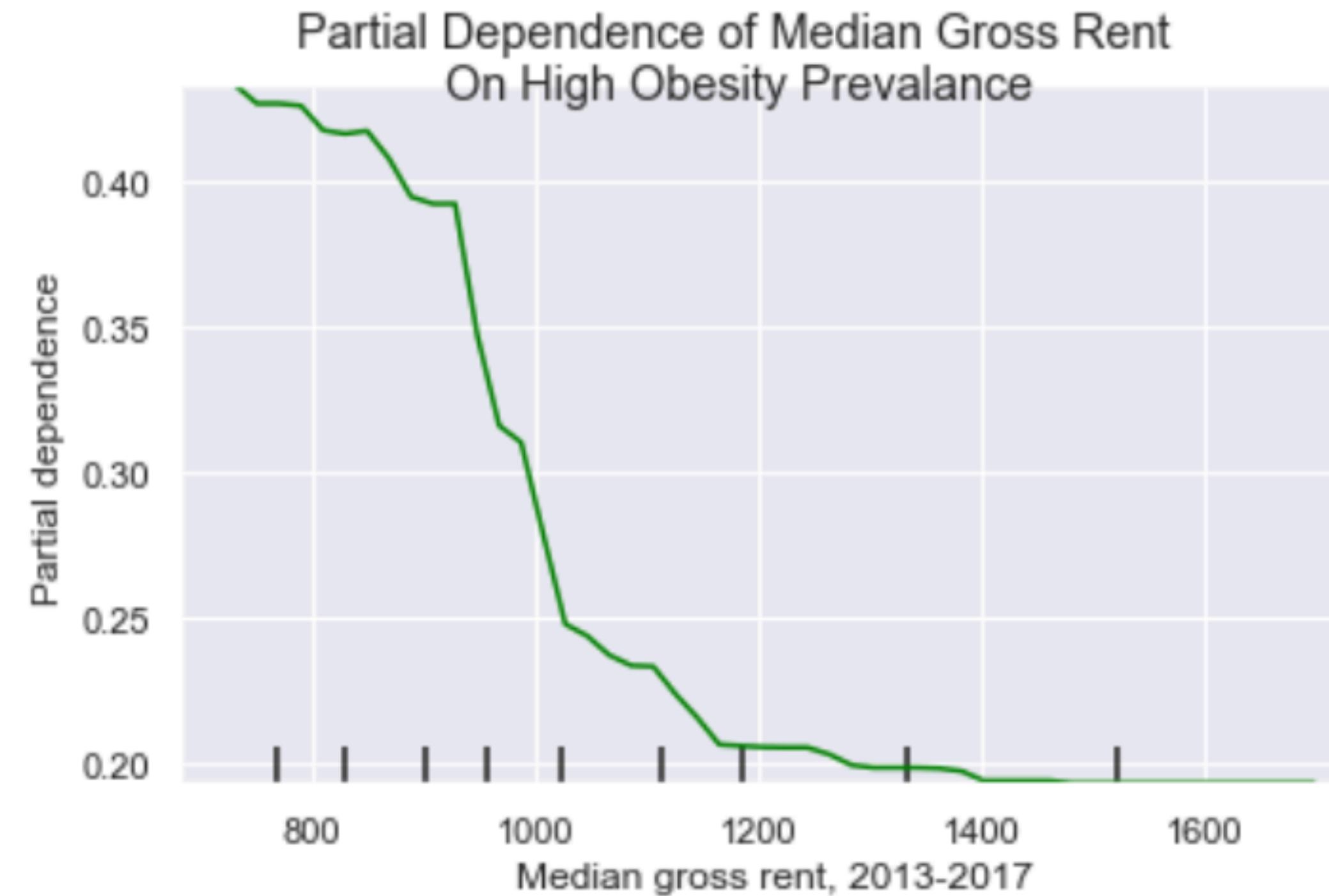
Partial Dependence of Negative Sentiment
On High Obesity Prevalance



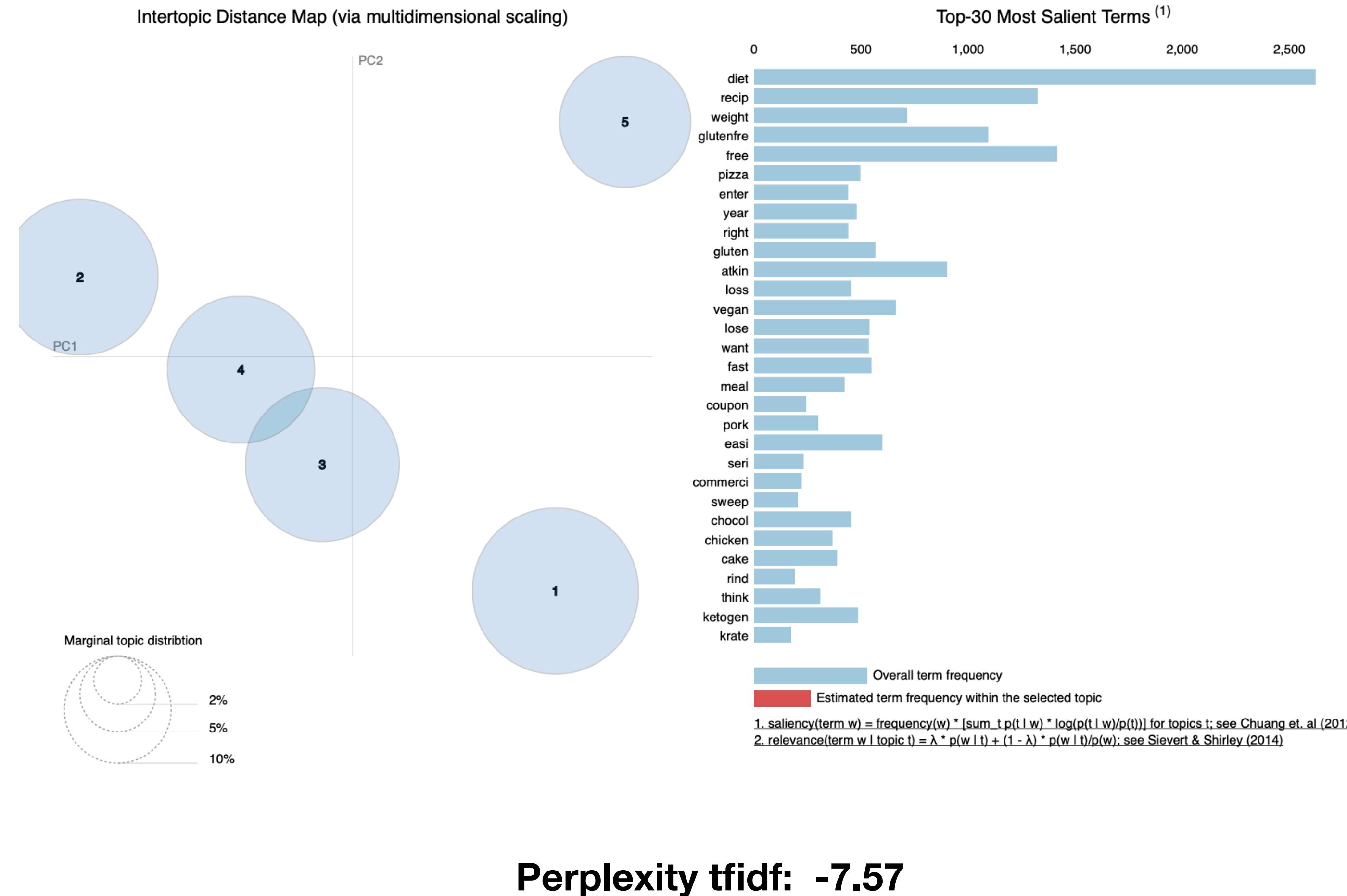
Partial Dependence of Positive Sentiment
On High Obesity Prevalance



Features: Partial Dependence



Topic Modeling using LDA



	Keywords	Topic
1	go, chicken, dinner, fast, think, tri, recip, diet, right, best	Cooking
2	diet, paleo, know, thank, great, day, vegan, time, feel, meal	Feel-good diets
3	good, weight, idea, loss, diet, fee, like, love glutenfre, help	Weight-loss
4	food, come, friendli, happi, health, thing, father, ketodiet, work, ketogen	Health
5	atkin, glutenfre, week, carb, bread, today, coffe, lose, delici, chocol	Indulgences

Sentiment Analysis

