Hunger in America

Projecting Food Insecurity Rates in the US

By Khyatee Desai
Khyatee.d@gmail.com

Background

Food insecurity is the inability to **reliably** obtain enough food, due to a lack of resources

Business Problem

Currently, FI rates are calculated retroactively

Business Goal

This project aims to predict
current and future FI rates
to allocate resources to communities
before they suffer

The Data

54 CSV files spanning 2009-2020

950,000 rows total

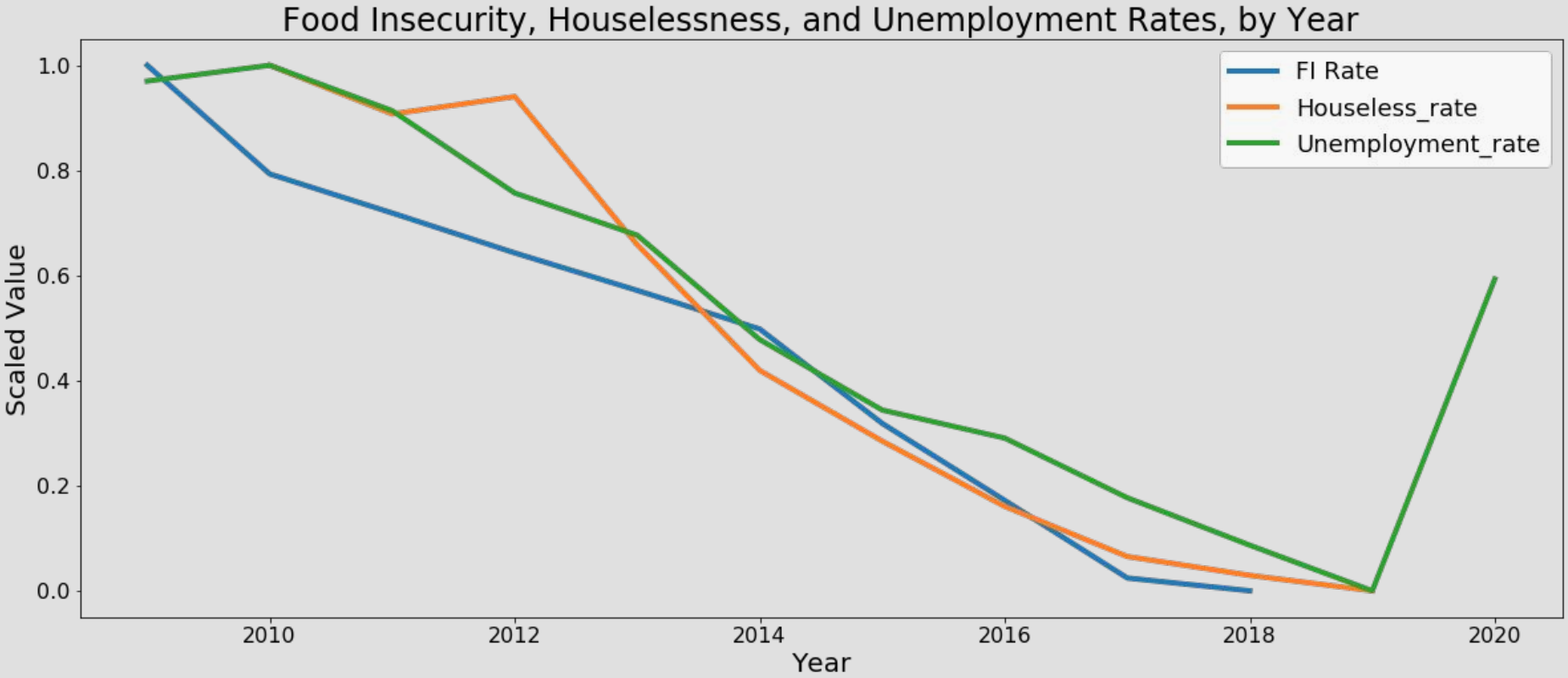
FI rate, rent prices, meal cost, houselessness, unemployment, race, food retail, all at the county level

Exploratory Analysis

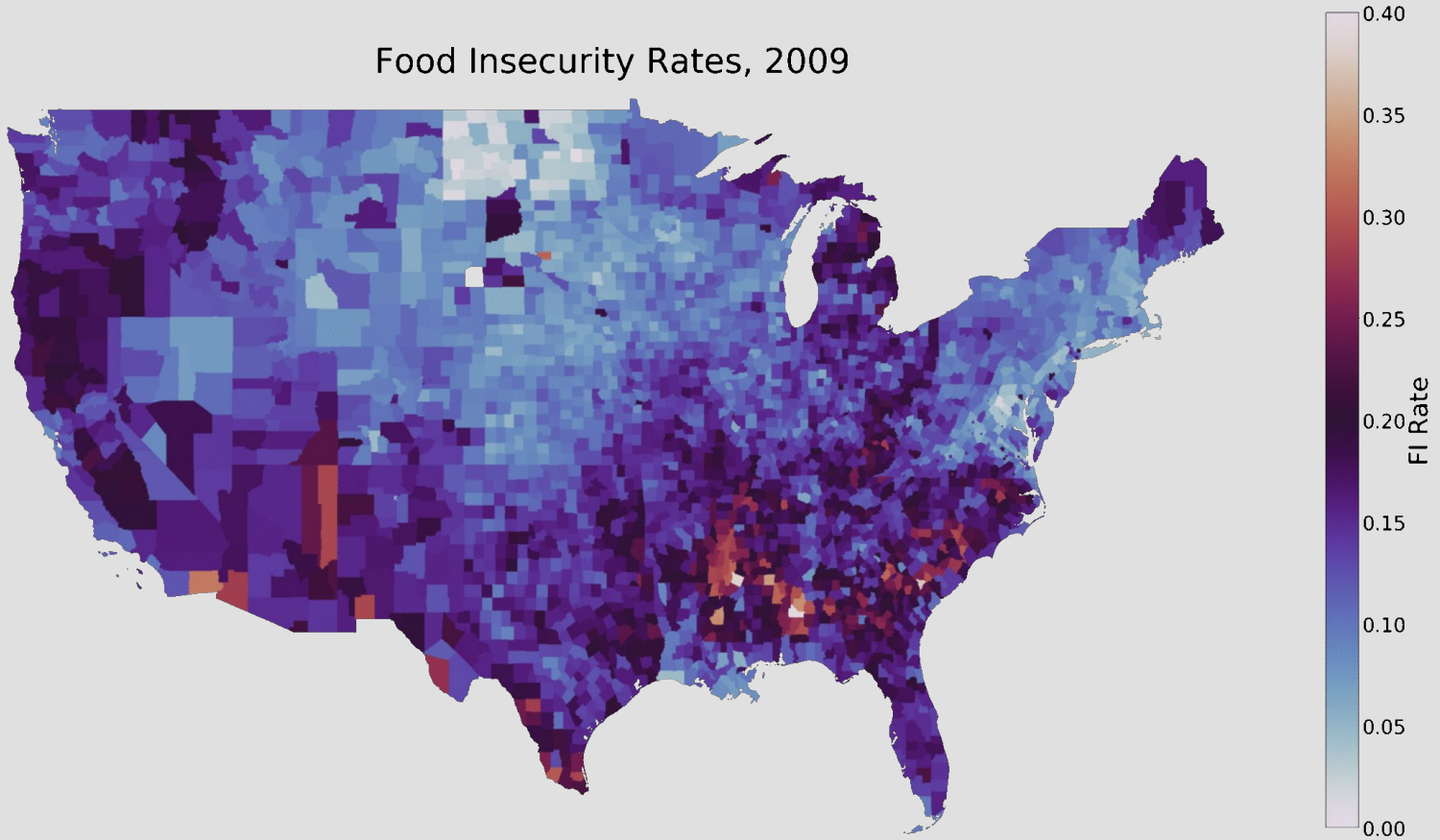
How has FI rate changed over time?

Which areas experience the highest rates of FI?

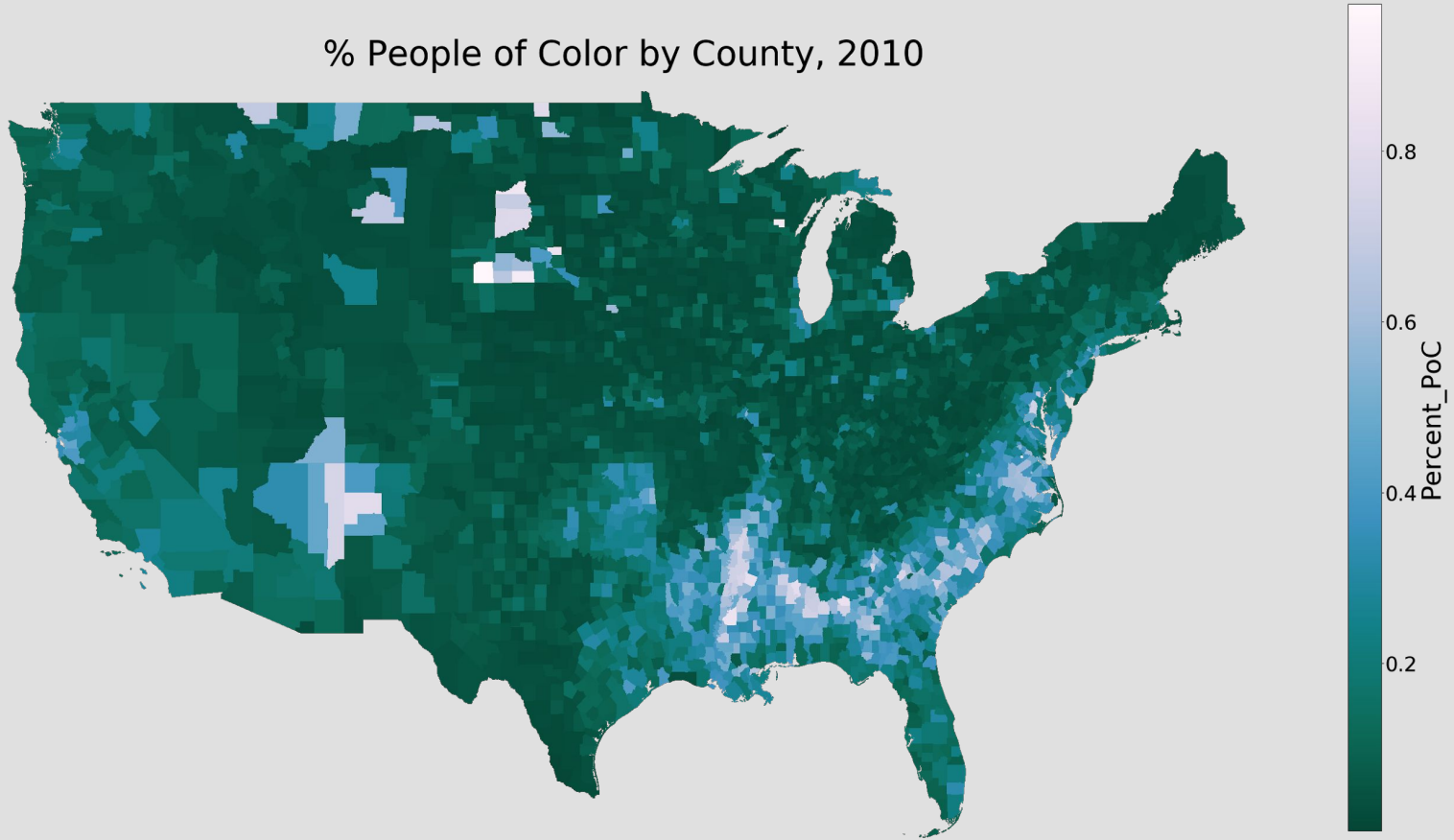
How does FI rate relate to other features in the data?



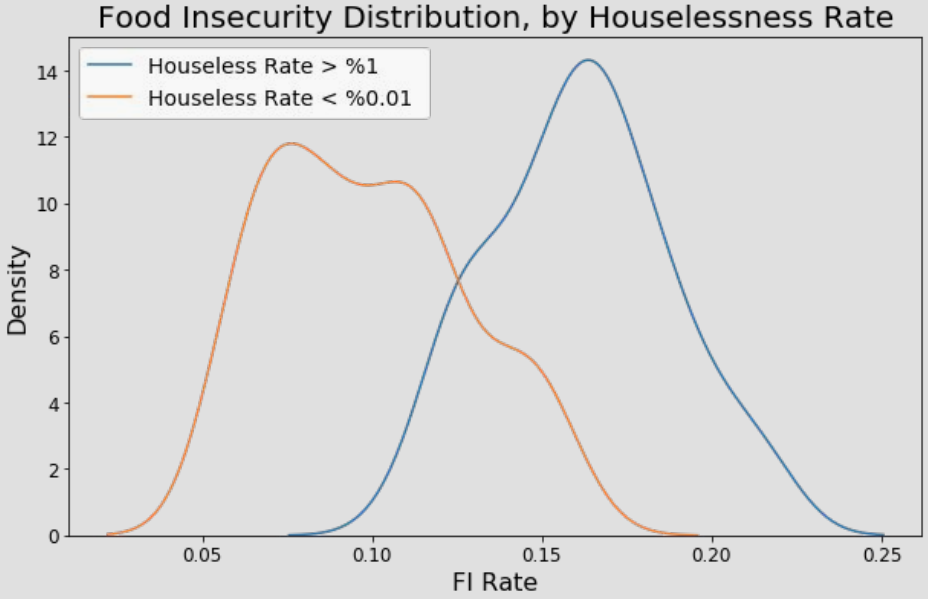
How has FI rate changed over time?

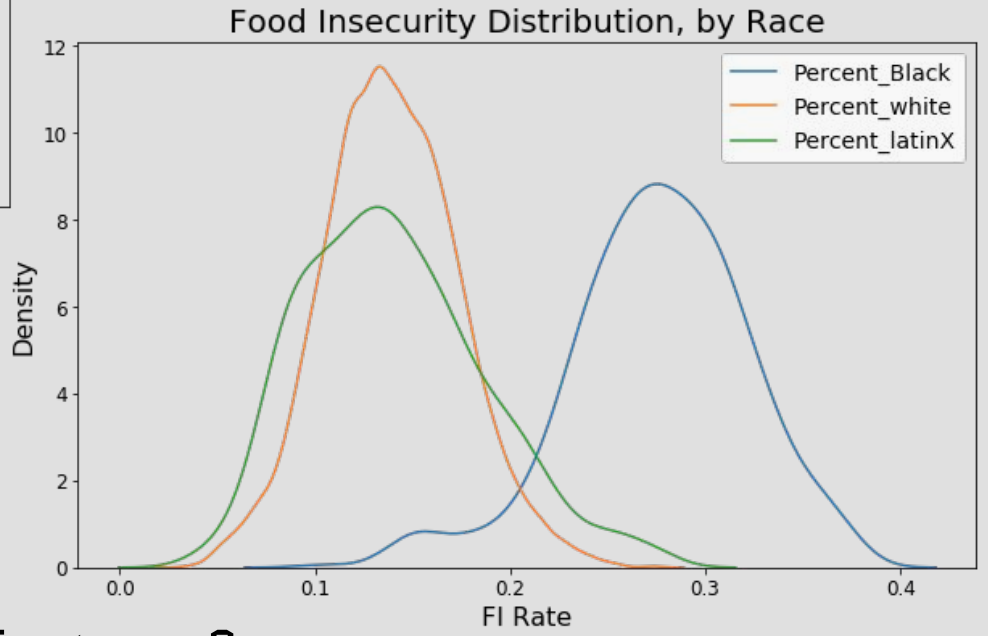


Which areas experience the highest rates of FI?



How does FI rate relate to other features?





How does FI rate relate to other features?

Linear Regression Modeling Process

Data Cleaning

Join datasets by county
Impute missing values
Map coded values to data dictionary

Feature Engineering

Interaction Features
Log Transformations
Dummy Variables

Feature Selection

Remove multicollinear features
K-best selector
Recursive feature elimination**

Model Evaluation

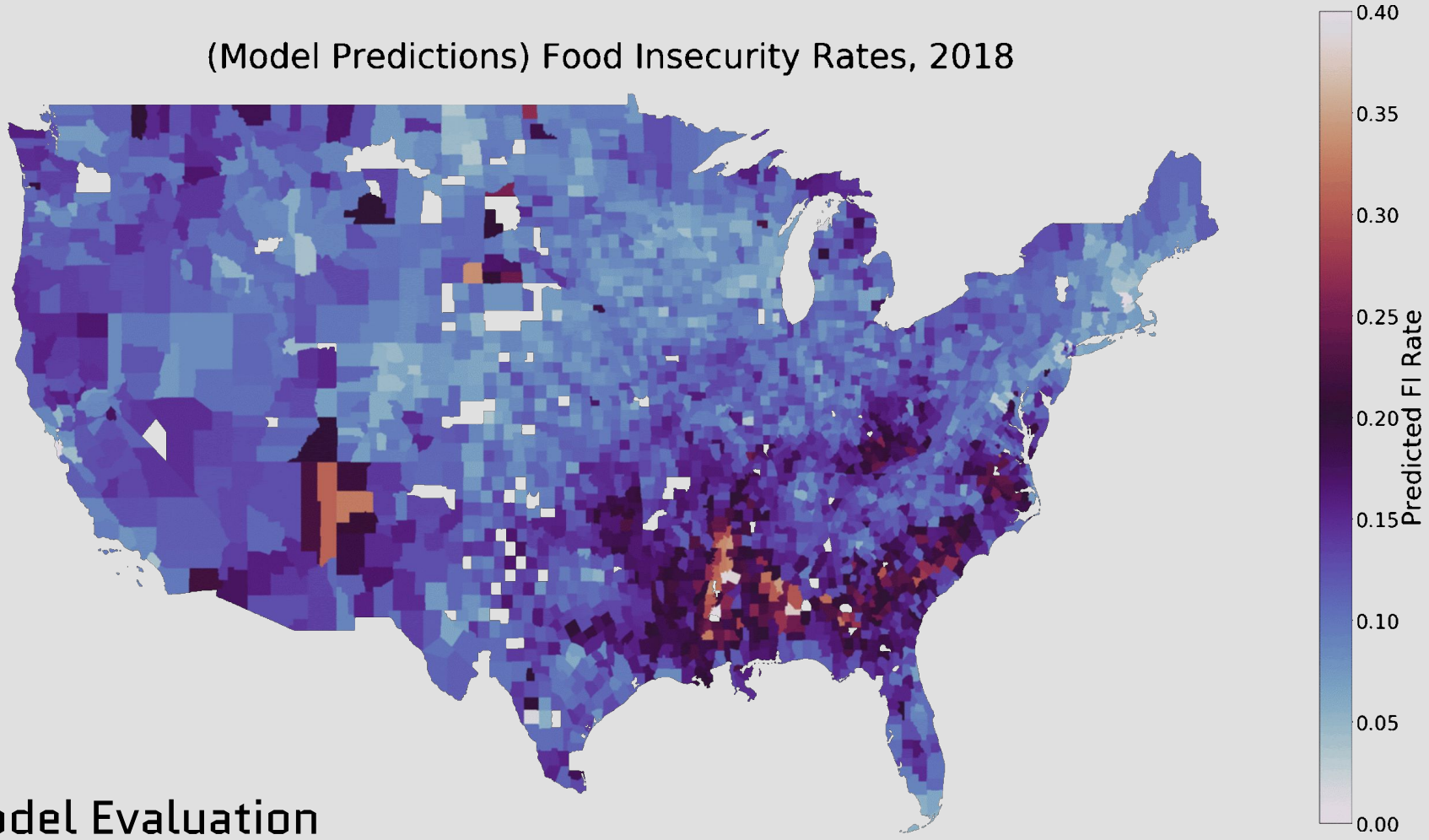
R-Squared
RMSE
K-fold cross validation

R2	0.75	75% of variance explained by model
RMSE	0.0198	Predictions off by 2% on average

Model Evaluation

RFE Features with Highest Coefficients

- Unsheltered_rate
- Houseless_rate_X_Sheltered_rate
- Houseless_rate_X_Percent_male
- Sheltered_rate
- Percent_Black_X_Percent_working
- Sheltered_rate_X_Percent_male
- Unsheltered_rate_X_Percent_male
- Houseless_rate_X_Percent_female



Model Evaluation

Features

Age
Median Income
Mobility
Disability
Food Assistance

Models

Ridge Regression
Lasso Regression

Unsupervised Learning

Predicting 2020 FI Rates

Next Steps