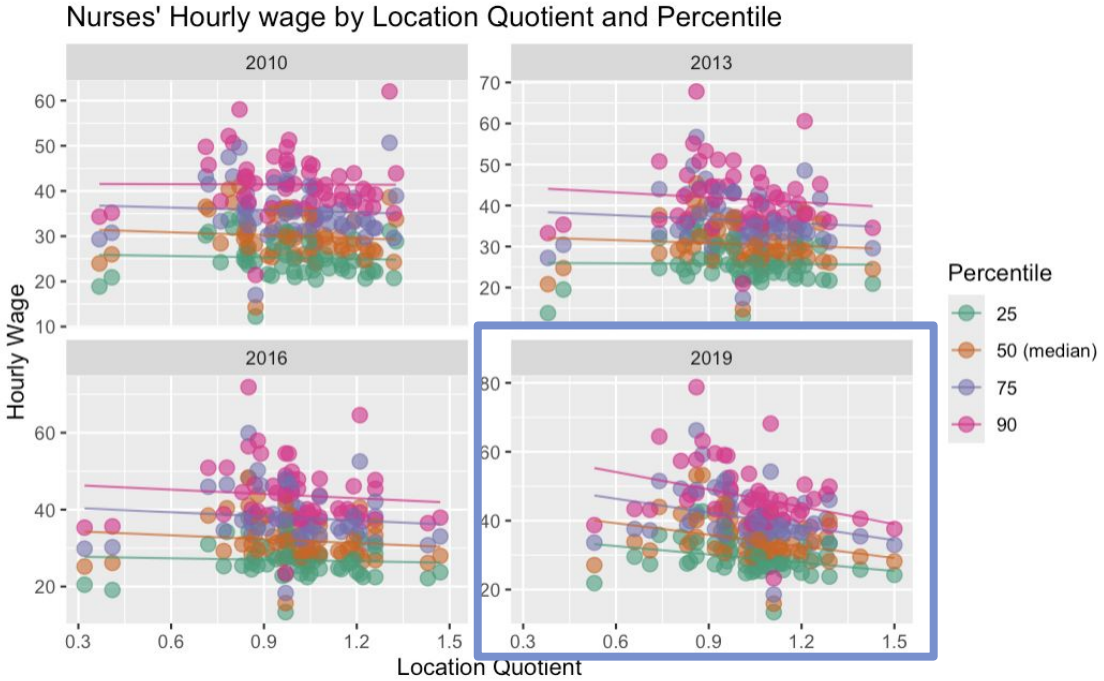
Nurse Hourly EDA

Question 2: Is there a relationship between location quotient (LQ) and hourly wage?

Hypothesis: Places with smaller LQ have higher wages

In 2019, places with smaller location quotients (<1) have higher wages



- General decreasing trend in 2019
- Hourly wage remained constant in previous years.

Project 4: Drug overdose and alcohol related deaths

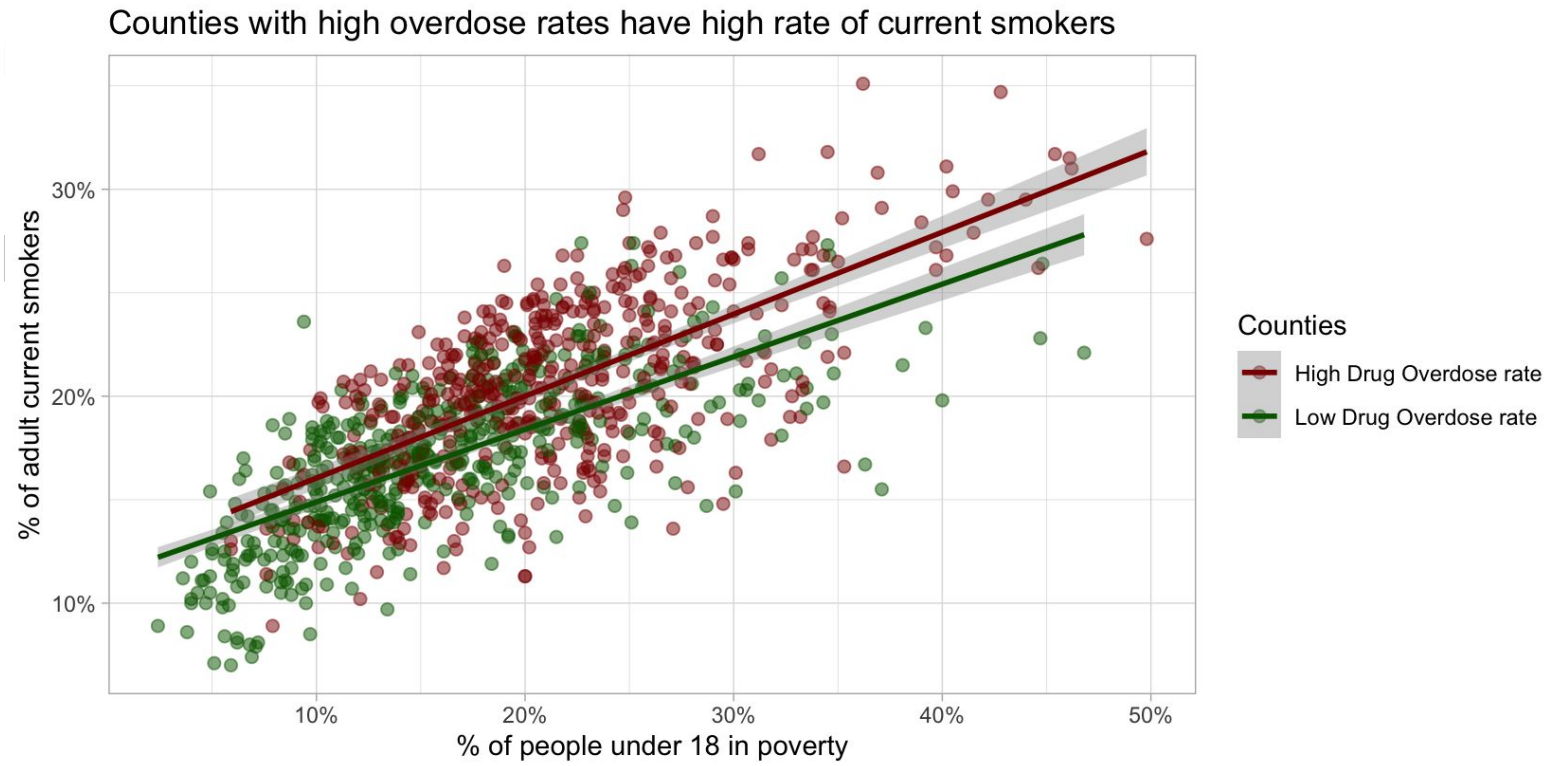
Dennis Campoverde-Lema, Samantha Parras, and Jodie Chen

Are there demographic and social factors that are predictors of drug overdose deaths and alcohol-related incidents deaths(e.g., driving accidents)?

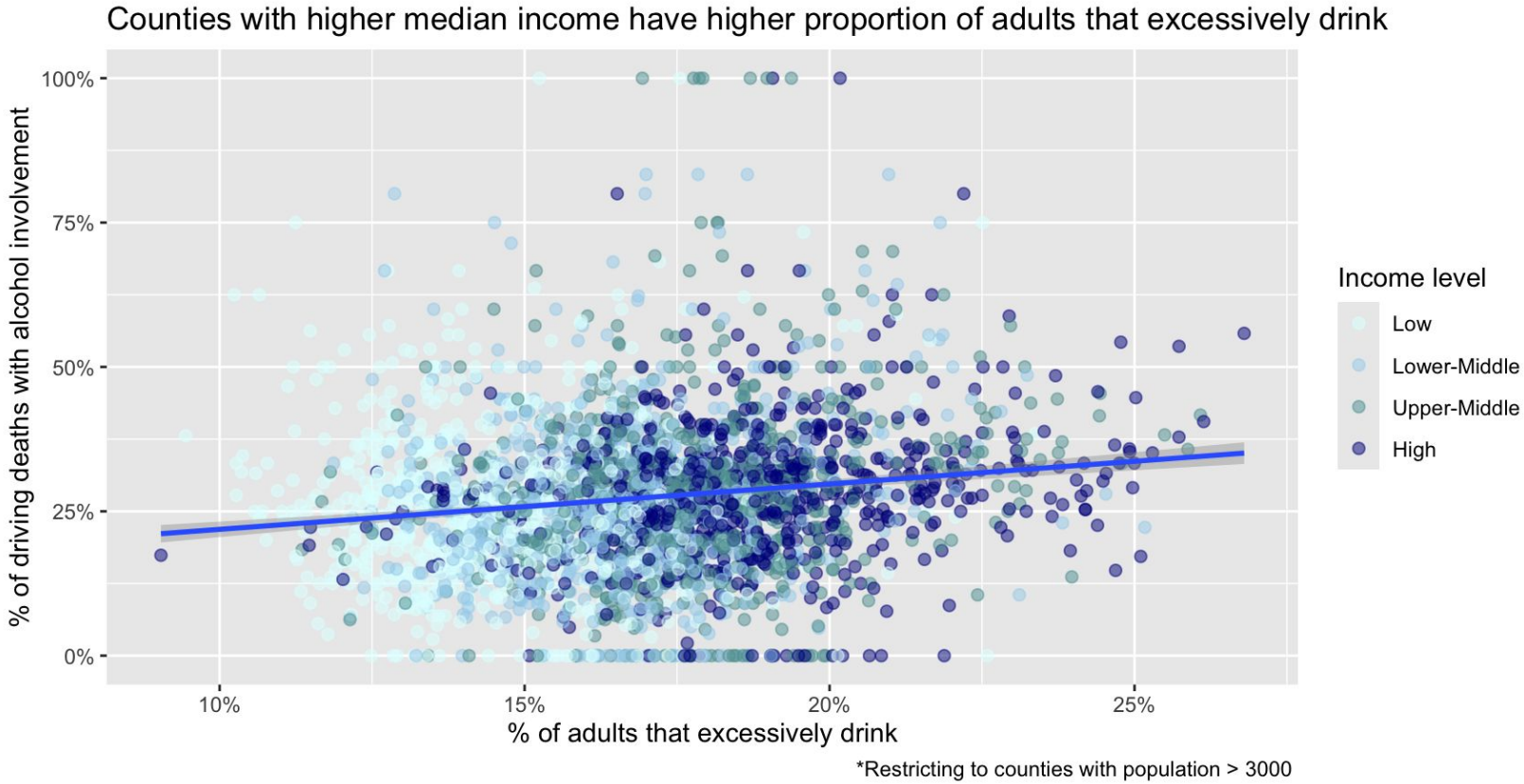
Data Description

- **Source:** 2024 County Health Rankings
- Key variables were renamed for clarity and ease for interpretation. For example:
  - Drug_overdose_death_raw_value to Drug_Overdose_Deaths
  - Alcohol_imparied_driving_death_raw_value to Alcohol_imparied_Driving_Deaths
- **Demographics:** Population by age, race, and gender
- **Socioeconomic factors:** Unemployment Rate, Income Inequality, Education Rates, etc.
- **Data Processing:**
  - Converted all variables to numeric.
  - missing or zero values

EDA - drug overdoses



EDA - alcohol impaired deaths



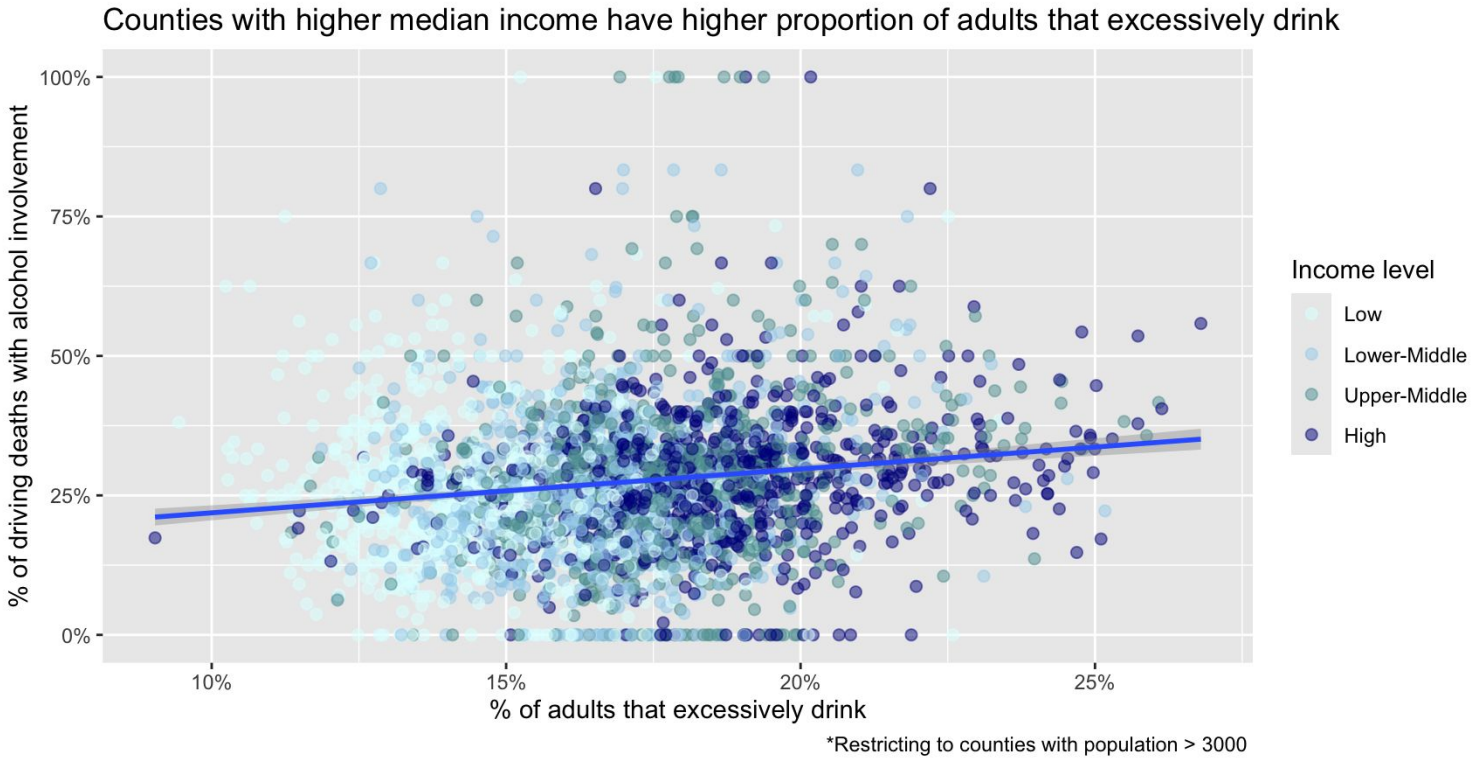
Modeling Strategy

1. **Data:**
    a. 2024 County Health Rankings
2. **Analytical Methods:**
    a. **Correlation Matrix:** Finding variables that correlate with our two response variables
    b. **Linear Models:** Baseline understanding of relationships.
    c. **Ridge Regression:** Addresses multicollinearity, improves model stability.
    d. **Lasso Regression:** Enhances model by selecting significant predictors.
3. **Validation:**
    a. Use cross-validation to test model accuracy.

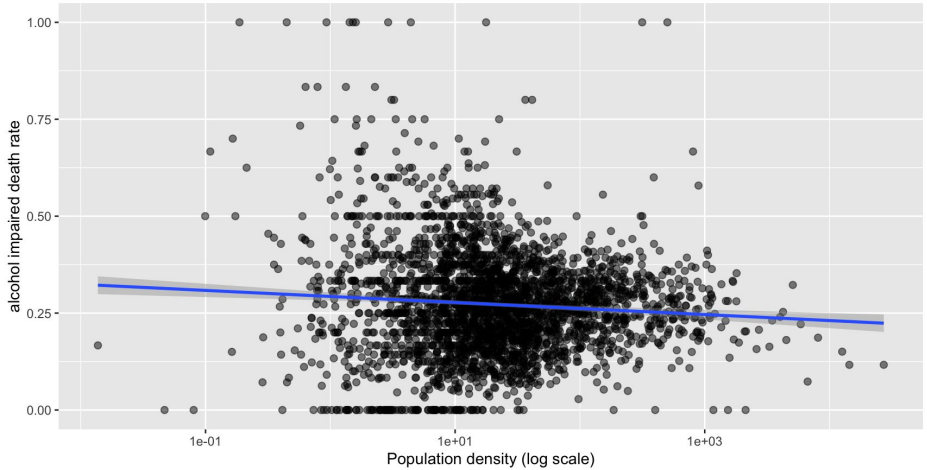
Plan of Action

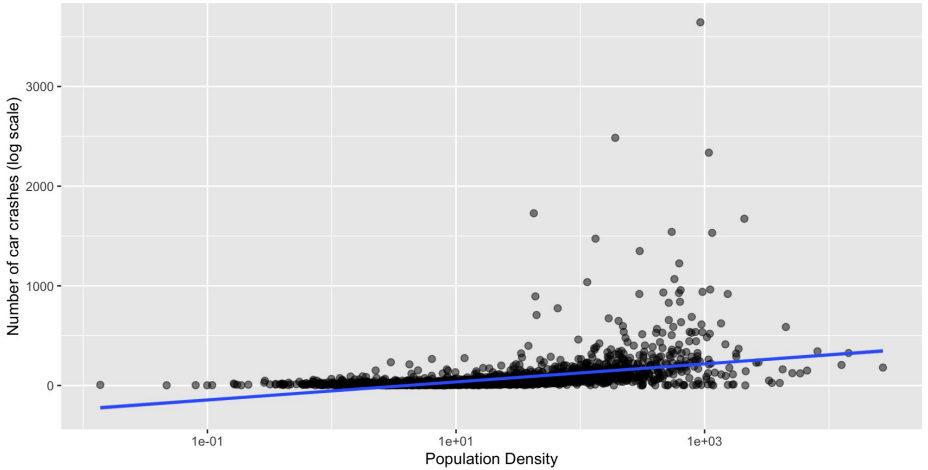
- **Completed Steps:**
  - Data acquisition, Cleaning, and correlation analysis
- **Ongoing Steps:**
  - Refining ridge and lasso models.
- **Next Steps:**
  - Creation of new variables that might improve model performance.
  - Implementing cross-validation techniques to validate the models.
  - Analyzing the output to draw meaningful conclusions.
  - Documenting findings and methodologies.
  - Developing the final presentation slides and rehearsing the presentation.
- **Next Deadline**: July 12, 2024.

Revised EDA - alcohol impaired deaths (no facets–keep)

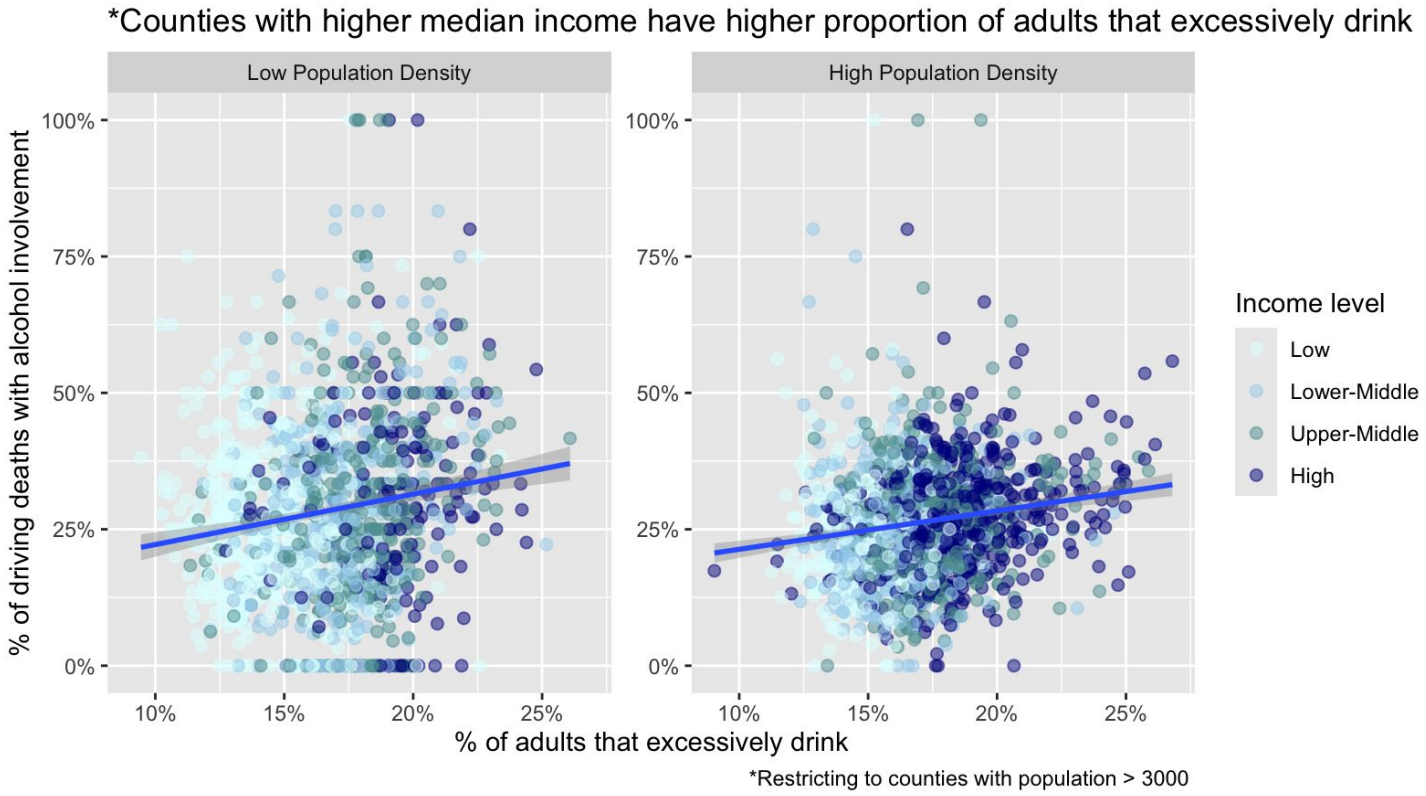


Population Density vs Alcohol Impaired Deaths Rate

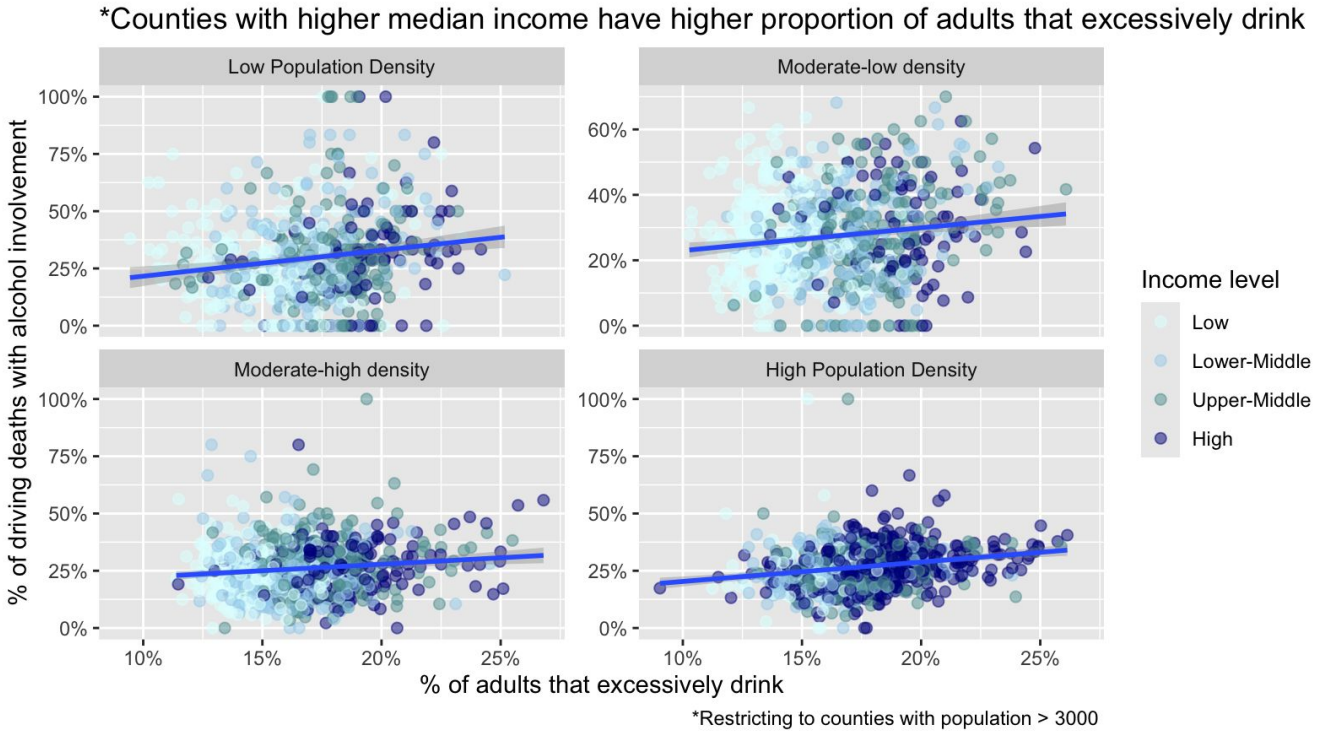




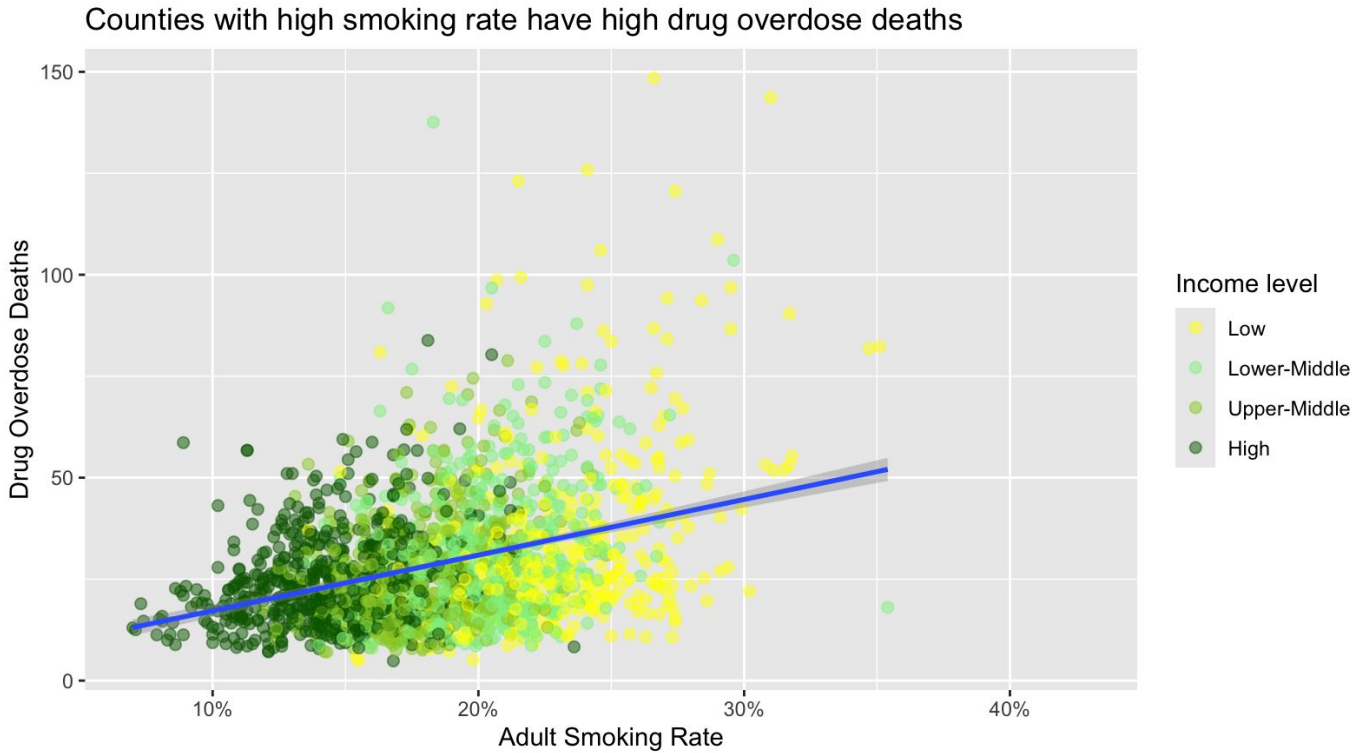
Revised EDA - alcohol impaired deaths (2 facets)



Revised EDA - alcohol impaired deaths (4 facets)



Revised EDA - drug overdoses



Adding more relevant variables

Social–community safety

1. Homicides (1368/1767 NA:not NA)
2. Suicides

Social–education

1. Disconnected youth (%) – 1965 NA : 1170 NA
    a. Numerator is num of people (16-19) neither working or in school
    b. Denominator is total county population, ages 16-19.

Modeling suggestions

- Variable selection
  - Lasso variable selection (built in linear model)
  - Fit model with variables we selected ourselves (using EDA)
  - Bayesian sparse regression using the horseshoe prior (R
- Model selection
  - Multilevel (mixed effects modeling),
  - random forest (be prep to show why chose this model)
  - Maybe a black box to show simpler, interpretable models are better or perform just as well