



Uncovering Predictors of Substance-Related Death: Social & Demographic Risk

Carnegie Mellon University
Statistics & Data Science

Becky Zhou

Georgina Mbonu

Thomas Kim

University of Pittsburgh

University of Texas Austin

University of Minnesota

Background

Motivation:

- To enhance public health outcomes with substance abuse while simultaneously reducing healthcare expenditures.

Main Question:

- Are there **demographic and social** factors that are predictors of substance abuse outcomes?

Why This Matters:

- "Every day, about **34 people** in the United States **die in drunk-driving crashes** — that's one person every 42 minutes. These deaths were all preventable."("National Highway Traffic Safety Administration")
- "The annual cost of crash deaths involving alcohol-impaired drivers totaled about \$123.3 billion in 2020."("CDC")

Data & Key Variables

Data Source:

- 2025 County Health Rankings Dataset published by the University of Wisconsin Population Health Institute.

Response Variables:

- **Drug Overdose Deaths**
 - The number of people who died from a drug overdose per 100,000.
- **Alcohol Impaired Driving Deaths**
 - Percentage of driving fatalities with alcohol involved.
- **Excessive Drinking**
 - Percentage of adults reporting binge or heavy drinking.

Summary of Model Approach

We applied three modeling approaches to identify key predictors of drug overdose, alcohol-impaired driving deaths, and excessive drinking:

❖ Multiple Linear Regression:

- Chosen for simplicity and interpretability, This model served as our baseline before moving to the Huber.

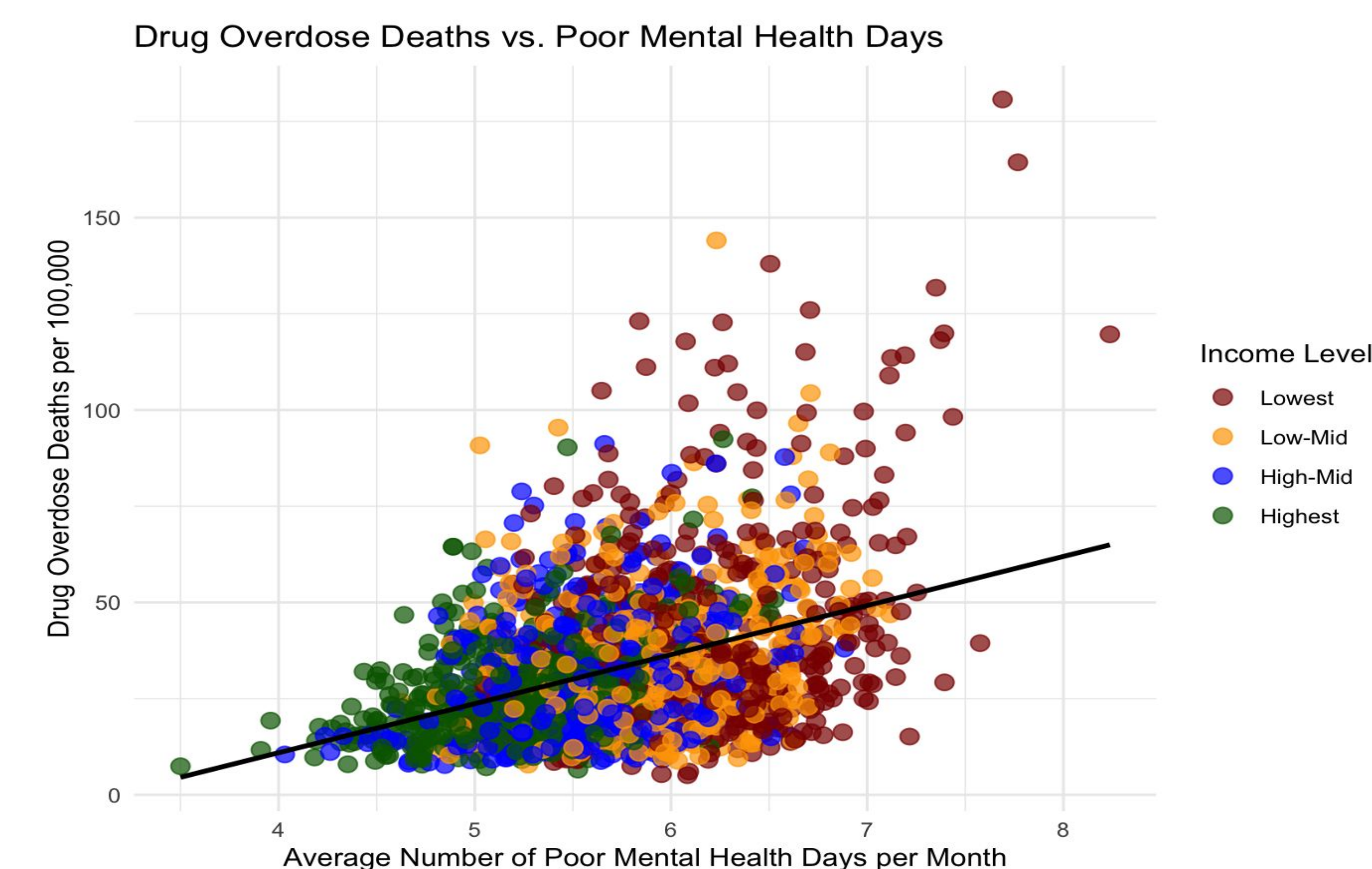
❖ Huber Regression:

- To account for the presence of outliers we chose this model which reduces the effects of extreme values and uncovers relationships we wouldn't see otherwise.

❖ Random Forest:

- This non-parametric model accounts for non-linear relationships and gives insight to variable importance.

EDA / more results



Mental Health & Income Disparities:

- Counties with **more poor mental health days** per month tend to have **higher drug overdose death rates**.
- The association is especially strong in **low-income areas**, where poor mental health and drug overdose are more prevalent.

Discussions

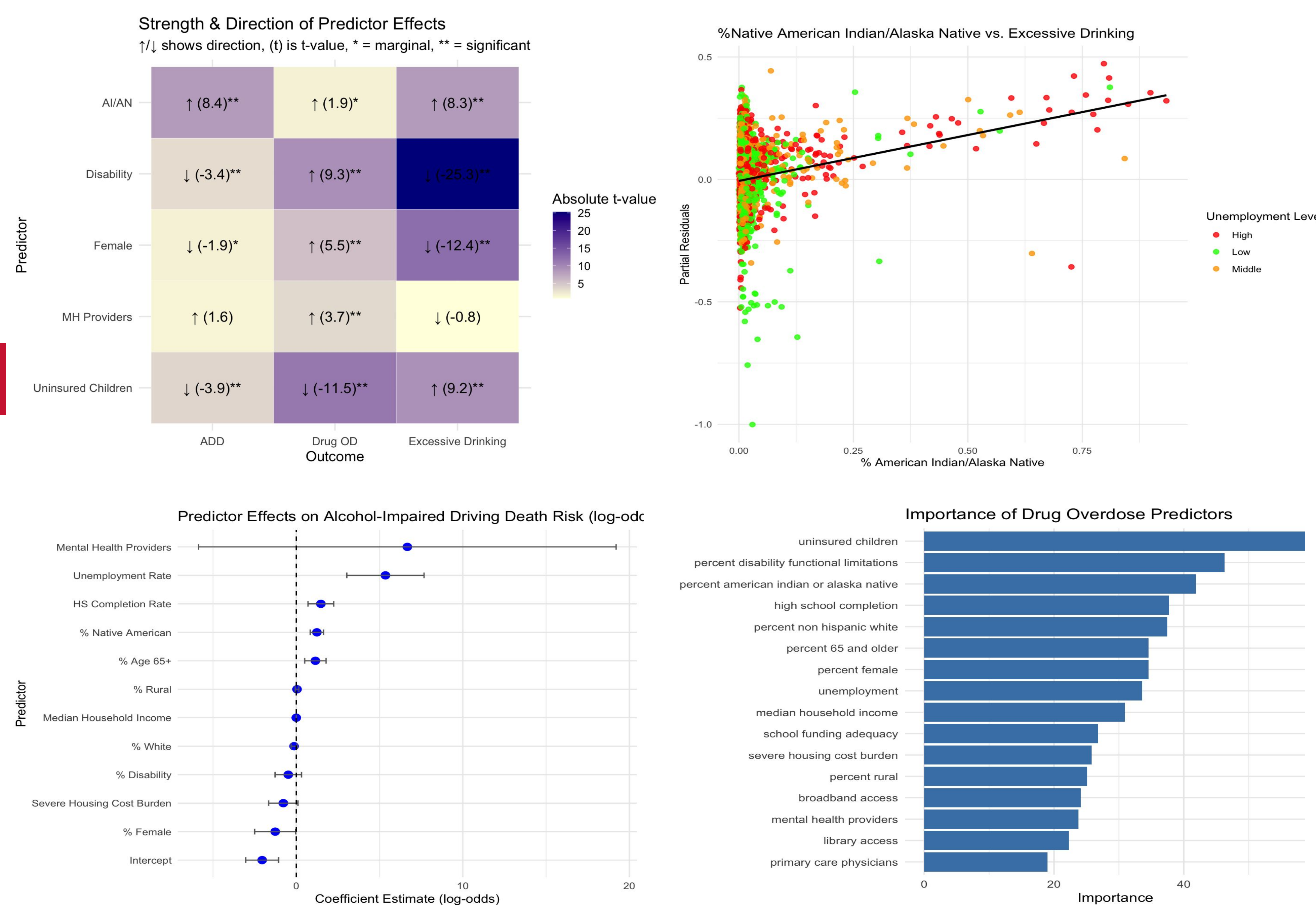
Conclusions/summary:

- Substance abuse outcomes appear to be driven by underlying structural inequalities.
- Counties with higher disability rates, large AI/AN populations, and fewer mental health resources tend to experience worse outcomes related to overdose, impaired driving, and excessive drinking.

Limitations:

- **Missing data** deeply effects variables like race and ethnicity possibly biasing results.
- **Snapshot Information** leads to only capturing inferences about causation.
- **Possible undercounting** because of underreported variables especially in rural and under-resourced areas.

Results



❖ Factors Underlying Rural Risk:

- Rurality by itself is not a good predictor of substance-related outcomes.
- **Structural characteristics often found in rural counties** consistently explain variation in our outcomes (Disability, % AI/AN, Uninsured Children).