Health Data Inequality Prediction & Visualization

Emily Adams (GT)
Monica De Alvaro Mena (GT)
Sriya Surapaneni (GT)
Huy Trinh (GT)
Juan David Perdomo (GT)
Jaxon Turner (GT)

Summary

In the United States, healthcare is **not equal** for all. Health databases are publicly available, but are not being utilized to generate useful insights and solve the problem. This project leverages county-level data to build an interactive tool for **policymakers** to identify and predict the effects of changing economic and health factors on **health outcomes**.

Data

Our data was downloaded from **County Health Rankings & Roadmaps**, specifically CSV files spanning from **2014** to **2023**. Yearly data consists of a row for each county (3,143) and 100+ columns for different health factors and outcomes. We used **R** to aggregate the yearly data and account for missing values.

Model - Random Forest

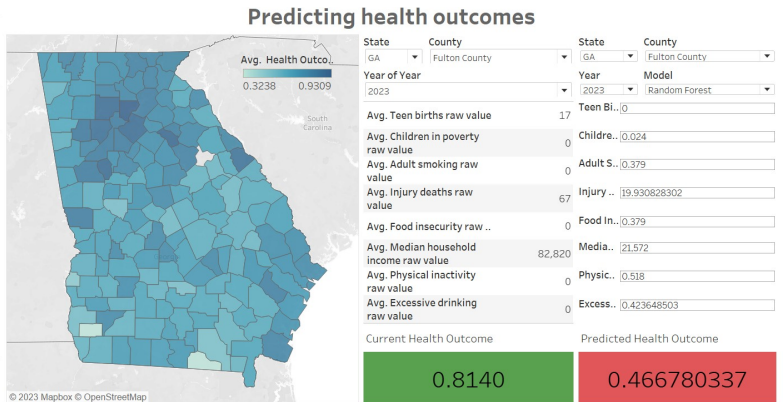
We created a health outcome variable combining the 5 most relevant **outcome columns** from our dataset:

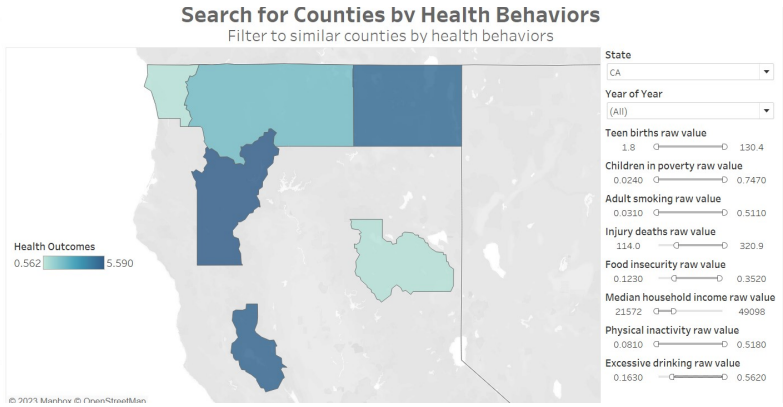
Premature Age-Adjusted Mortality, Poor or Fair Health, Poor Physical Health Days, Low Birthweight, and Poor Mental Health Days.

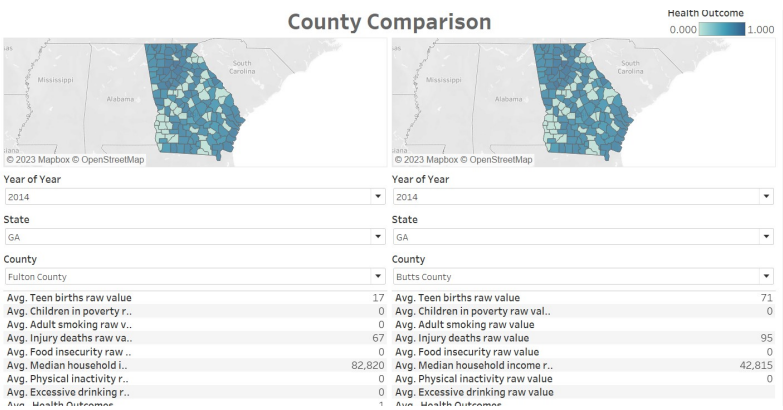
Due to data complexity and heterogeneity, we opted for a **random forest regression model** to get insights into the **most important features** of the data, so the user can see how the **changes alter the prediction** . Our model has a mean squared error of 0.2%.

Interactive Visualization

Using **Tableau,** we created a **choropleth map** of the US where users can see the health risks of their region compared to other regions by **hovering** over them. **Predictive** values can be **input** for select variables to show **updated values** for the new health outcome.







Experiments & Results

We refined an output column from chosen health outcome variables using classification and regression approaches, resulting in accurate predictions with minimal error. The most **important factors** that we found caused the unequal health outcomes were:

- Physical Inactivity
- Children in Poverty
- Adult Smoking
- Median Household Income
- Food Insecurity
- Teen Births
- Injury Deaths
- Excessive Drinking