Exploring Opioid-Related Mortality: Visualization & Prediction

Adam Awad Arthur Wang Evan Marcantonio Jay Kinzer

Summary

We visualize **opioid-related mortality** by state and county over time to show how different places are impacted by the opioid epidemic. We use predictive modeling to show **what mortality could look like in the near future**. The application allows users to **explore** how demographic and economic characteristics may impact mortality at the county-level.

Mortality Rates

We utilize two primary measures of mortality: **mean deaths per 100,000 people** and **age-adjusted mortality rate**. These are standard measures provided by the CDC and used in existing research.

Related Variables

Existing research shows relationships between mortality and demographic and economic characteristics. Our data allows users to explore how mortality varies by a county's makeup of **age, sex, race, poverty, income, and education**.

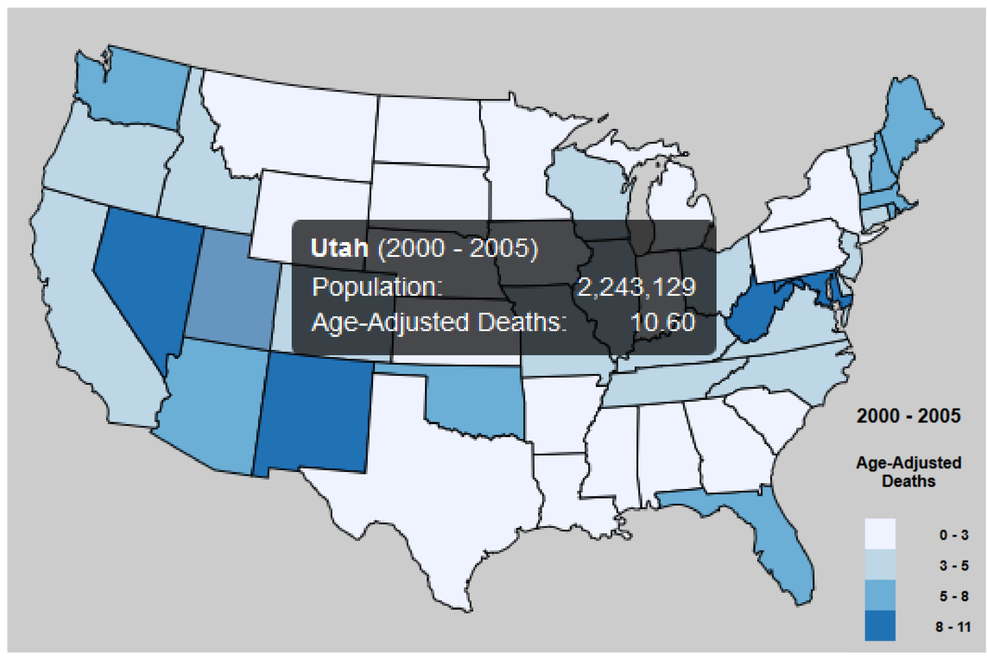
Datasets

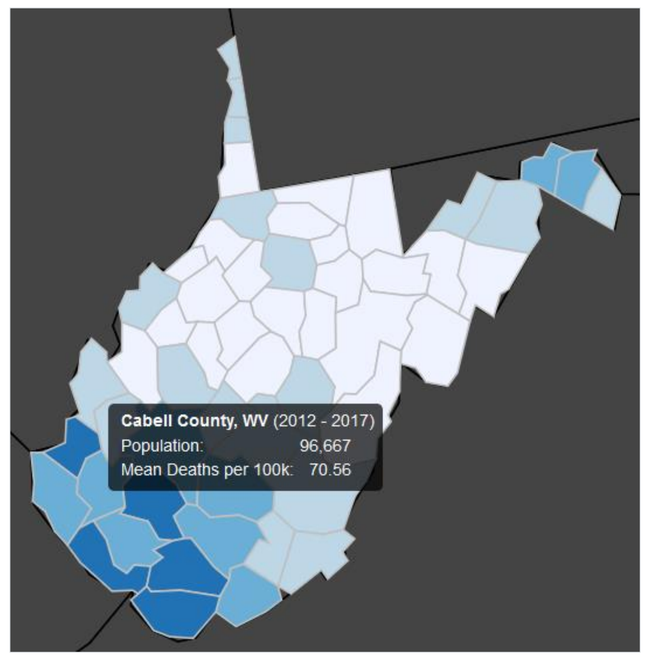
Our primary dataset collects mortality rates by county for three six-year time periods from **2000-2017**. We scraped this data from the CDC's WONDER system using the *Selenium* and *requests* Python libraries to automate hundreds of queries.

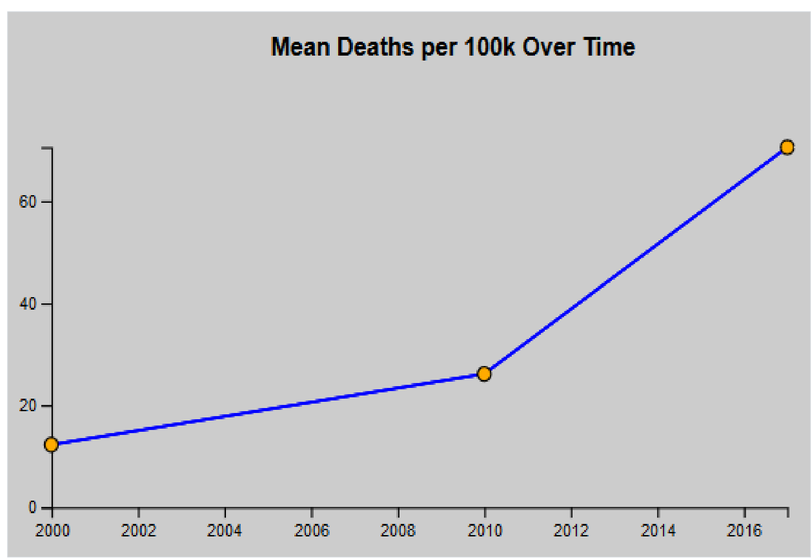
This dataset comprises **~10,000 records** (~1 MB). We supplemented this data with demographic/economic datasets downloaded manually from government archives, comprising **~6.9 million records** (~526 MB).

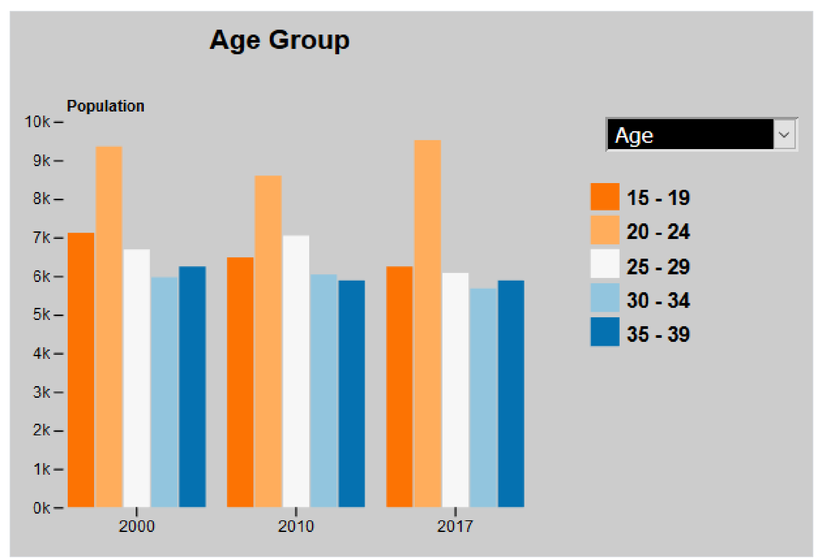
For visualization and predictive modeling, we cleaned these datasets using pandas. This reduced the file size to ~2.5 MB.

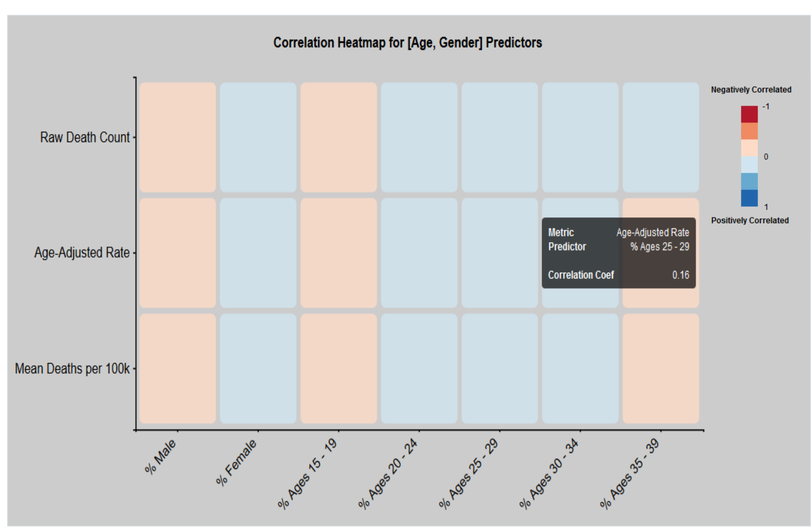
Interactive Choropleth and Detailed Charts

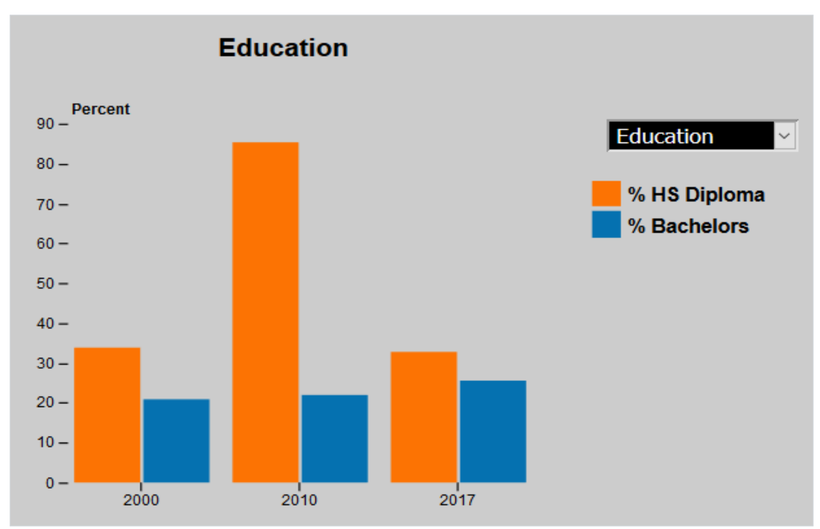












Visualization

We use **D3.js** to generate a choropleth that dynamically updates as users select filters and measures. Users can zoom in on specific states for a "deep dive" of specific counties. The application allows users to switch between historical and predicted data.

Predictive Modeling

Users can visualize mortality rates predicted by a **random forest**, trained using 80% of our input data. Predictions were generated from demographic/economic characteristics projected using **exponential smoothing** to represent what each county might look like in the future.

Evaluation

We asked a sample of **22** users to answer questions about the opioid epidemic from 2012-2017:

1. Which state had the highest mortality rate?
2. In that state, which counties had the highest mortality rates?
3. What makes those counties different from other counties?

We then asked users to rate WONDER and our application on usability, depth, visuals, and how easy it was to answer those questions. We used a scale of 1 (worst) to 5 (best).

Our application had a mean score of **3.9**, compared to **1.4** for WONDER.