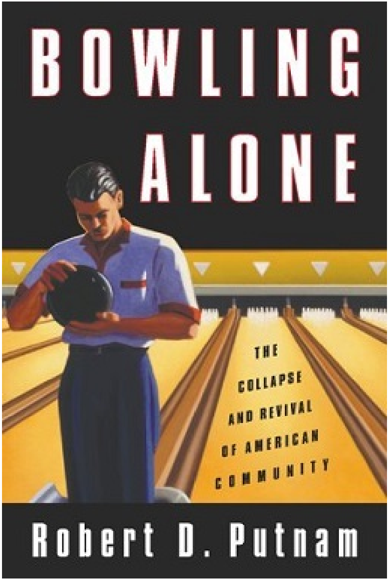
Deaths of Despair: An Analysis of Mortality in the American Rust Belt

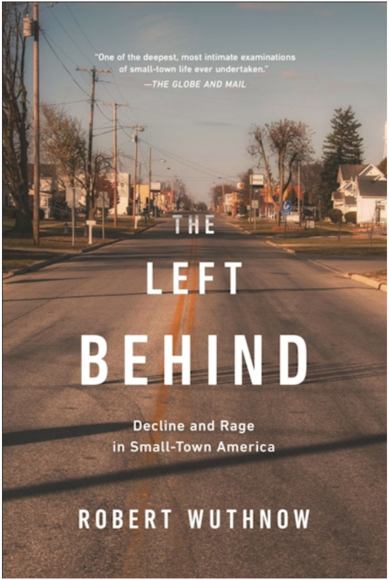
Samuel Lowe, Jacob S. Zelko

Northeastern University

April 16, 2025

Introduction & Background




"One of the deepest, most intimate examinations of small-town life ever undertaken."
—THE GLOBE AND MAIL
THE
LEFT
BEHIND
Decline and Rage
in Small-Town America
ROBERT WUTHNOW

Data Sets: Training Variables

United States Census

- Census conducted once every 10 years, most recently in 2020
- Surveys every household in the United States on socioeconomic and demographic questions

Association of Religion Data Archives (ARDA)

- Religious data, broken down by number of adherents and congregations per state
- Bowling Alone notes that religious involvement is one of the only kinds of social engagement to not fall

IPUMS CPS

- Socioeconomic and health data such as household income, food stamps, smoking frequency, unemployment, etc.
- Huge amount of data spanning numerous surveys across several decades

Challenges: Data Cleaning

- Difficult to make apples to apples comparisons between different data sets
- Even if apples to apples comparisons are possible, it's a lot of work just to clean and prepare data
- Limitation: A lot of ARDA data and IPUMS CPS data was missing in many columns, and sociological data from many surveys is only available in some areas and/or in some years

Data Sets: Target Variables

- Religious congregations in each state
- Separation of rust belt states from rest of US
- Income and tax data
- Alcohol abuse prevalence rates across US

Feature Selection: Lasso Regression

- Idea: a lot of features might not be relevant for predicting certain variables
- Singular value decomposition is great (see below) but it can be very difficult to interpret the resulting features
- Because interpretation is necessary both for sociological research and crafting policy, we started with Lasso to select features

Feature Selection: Results from Lasso

- $\alpha = 1$, $\lambda = 100$
- Wasn't terribly effective with finding out best parameters to use
- May need to experiment further with hyperparameters

Results and Discussion

- Linear regression: 33.4% MSE
- Ridge Regression: 27.8% MSE
- Prevalence was always being overestimated for rust belt states

Future Work

- Experiment further with hyperparameter adjustments or SVD
- Incorporate other exploratory data analysis (would K-means clustering group Rust Belt states together?)
- Add in additional data to experiment with over a several year period