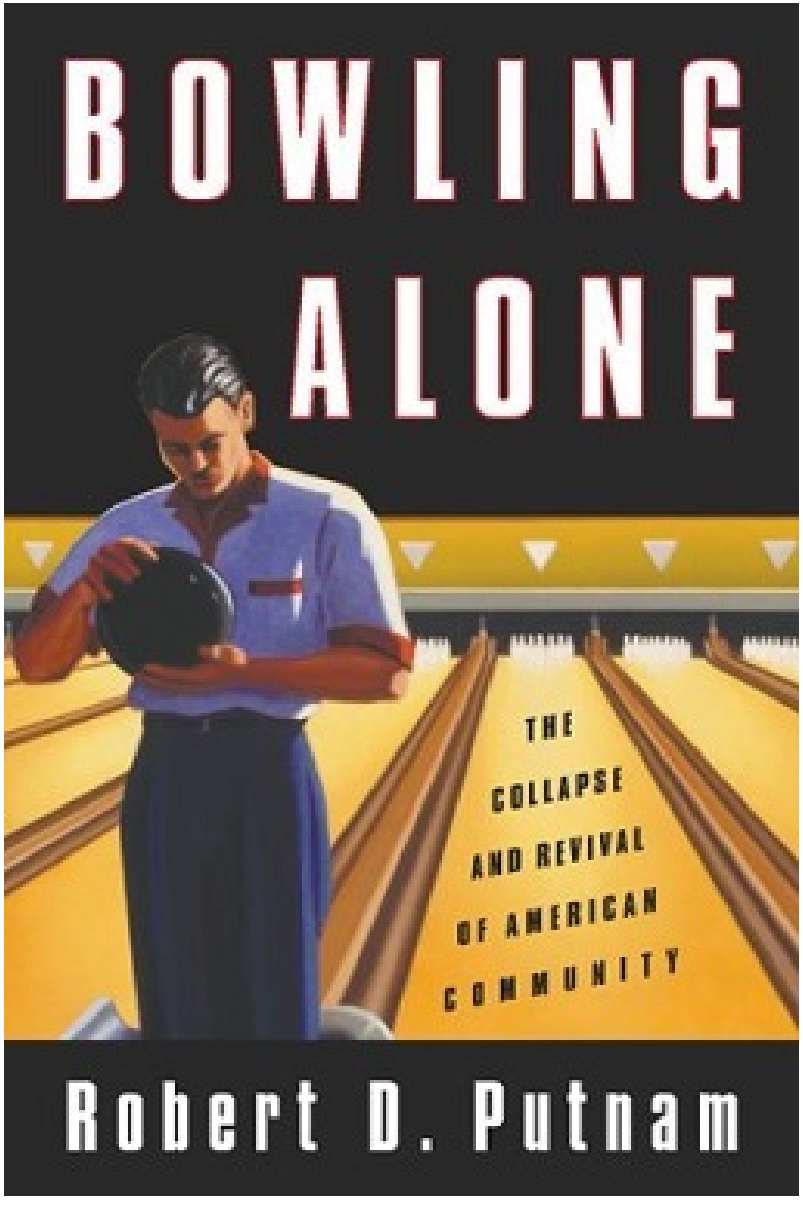
Deaths of Despair: An Analysis of Mortality in the American Rust Belt

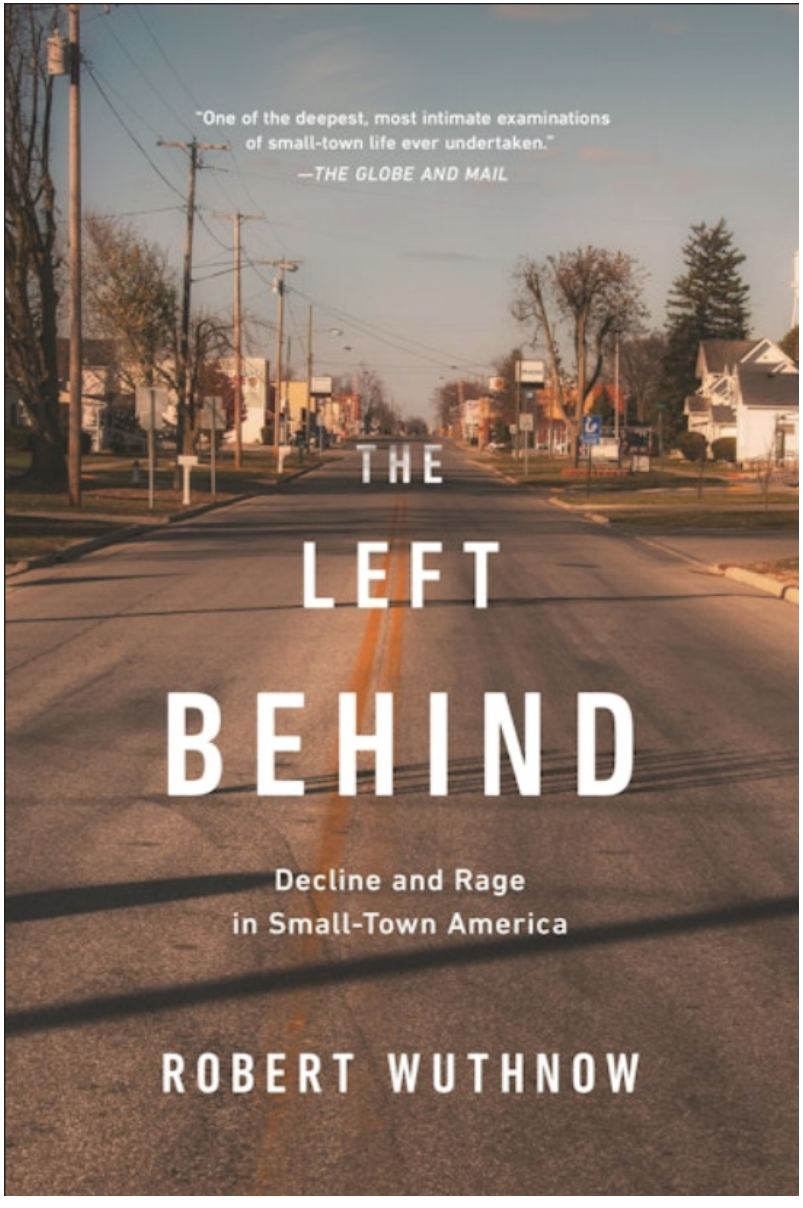
Samuel Lowe, Jacob Zelko

Northeastern University

April 16, 2025

Introduction & Background





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 Varibles

-

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

-

Feature Selection: Singular Value Decomposition

- Even though SVD models are harder to interpret, they may be substantially more accurate
- This helps in targeting which states are most likely to have mortality from certain causes, which can help respond to public health crises in the short term
- More interpretable work is needed to establish preventable measures and reduce mortality in the long term

Feature Selection: Results from SVD

-

Future Work