

Analyzing Social Health Factors

Team EcoStats
Karan Sheth, Liting Chiang, Wing Au,

Introduction

- The goal of this project was to investigate the various relationships between the multiple factors that make up social health.
- This project explores various data sets related to economic well-being including poverty rates, mortality rates, educational attainment, and social factors (e.g., trust in religion, faith in politics, etc.).
- Since social health is such a broad topic, big data frameworks are useful for aggregating data from multiple sources and performing fine-grained analysis

Background

- Existing research has shown that increased life satisfaction and overall happiness have lead to longer life expectancy. [1], [2]
- Higher education levels have been linked to increased life satisfaction [3].
- Increased income has been linked to higher life satisfaction [4].
- **Understanding the socio-economic conditions that can contribute to higher life satisfaction and happiness can inform better public policy decisions**

Data

- U.S. Census Bureau Data (6.5 GB)
 - Educational Attainment, Poverty, Income, Senior Citizen Living Conditions, Age by Zip Code
- CDC U.S. Mortality Data (5.5 GB)
 - Manner of Death, gender, education, location of death
- World Values Survey Data (2 GB)
 - Questions about their family, neighborhood, values, beliefs in political systems, confidence and faith in various procedures, and so on. (396041 rows and

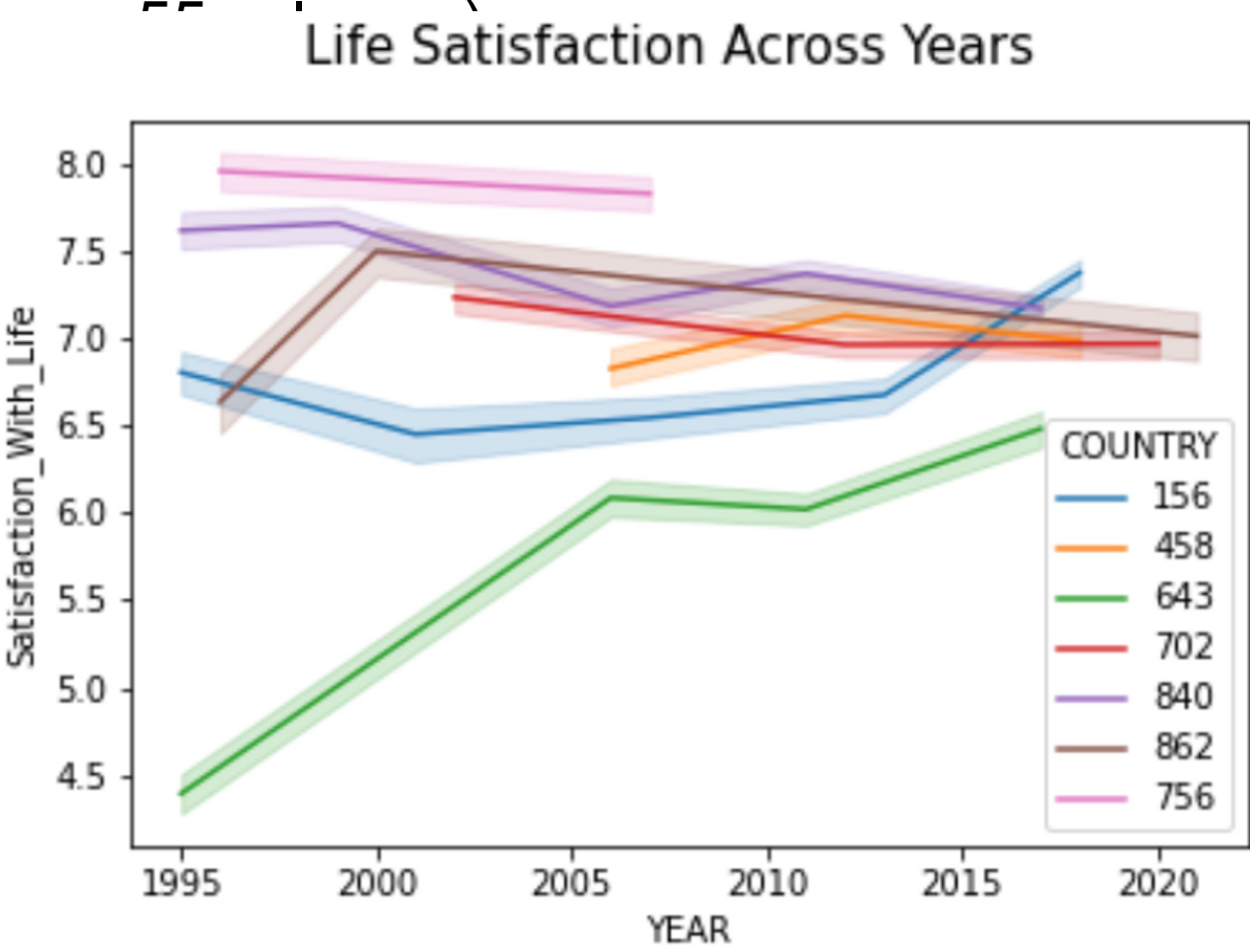


Figure 3: Life Satisfsaction Across Years

Methodologies

Clustering Analysis based on U.S. Census Data

- Data Preprocessing and Feature Engineering (Spark)
- Clustering by Similarity Search and Euclidean distance

Understanding Cause of Death in United States

- Extracted key features from the data to understand how a person's social background influences the causes of death, particularly accidents and suicide.

Predicting Life Satisfaction of country based on Social Factors

- Employing large-scale machine learning to forecast people's life satisfaction depend-ing on their socio-economic circumstances.

Results

(A)Clustering Analysis on U.S. County Economic Data

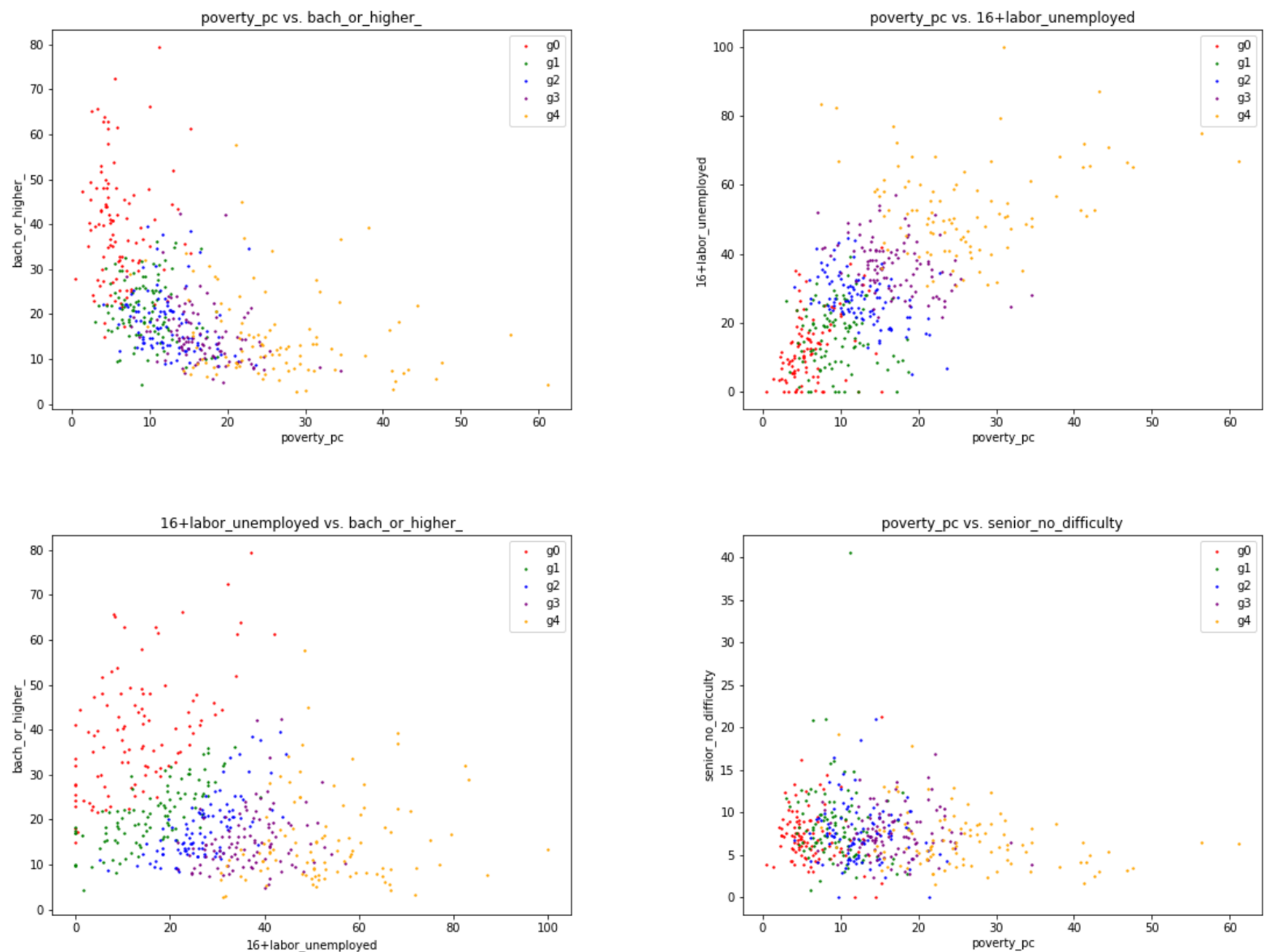


Figure 1: Poverty vs. Educational Attainment; Poverty vs. Unemployment Rate for 16+; Unemployment Rate vs. Educational Attainment; Poverty vs % Senior Citizens Living with No Difficulty

(A)Clustering Analysis on U.S. County Economic Data

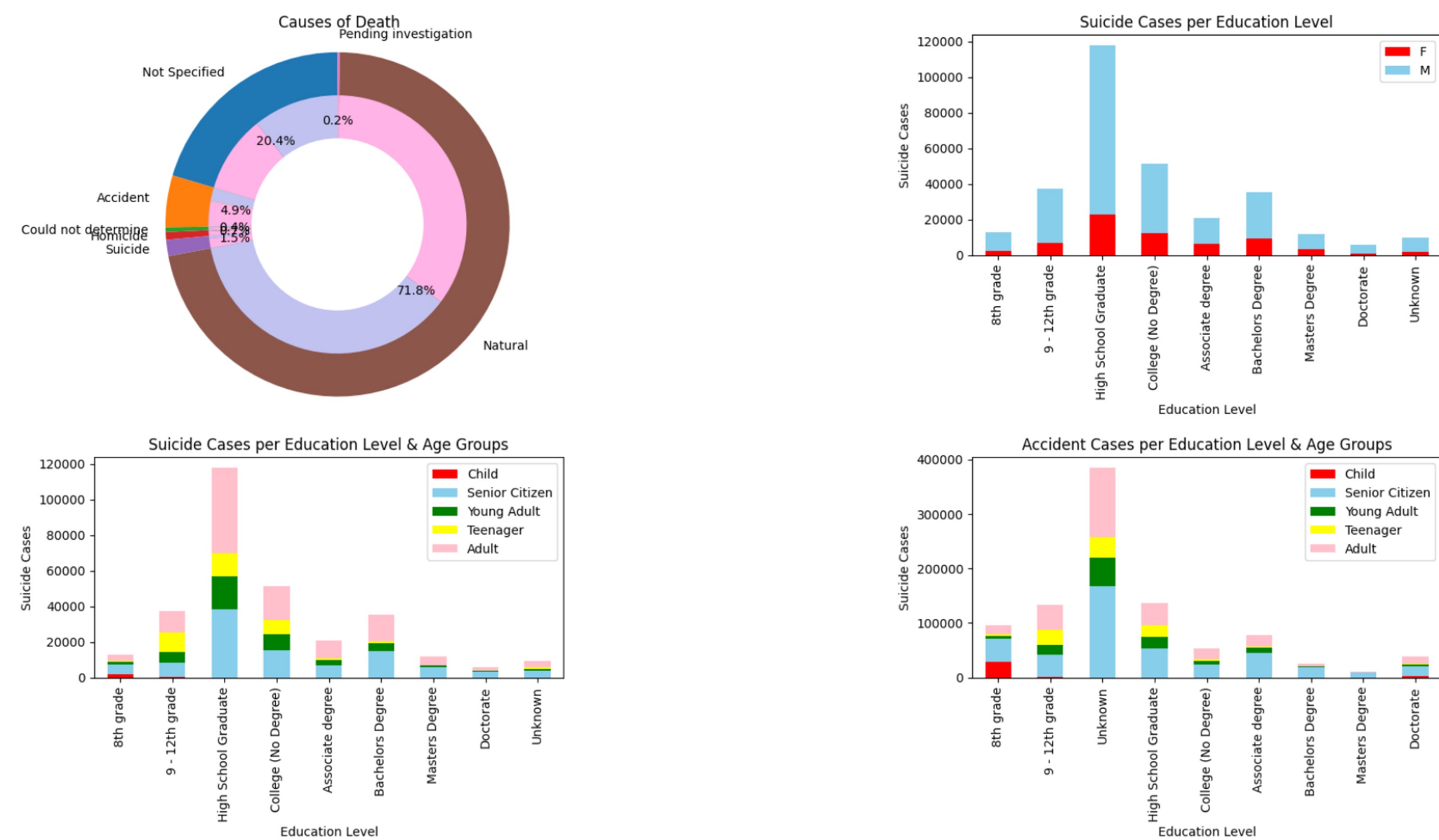


Figure 2: Analyzing the manner of death based on social condition

B) U.S. CDC Mortality Rate Breakdown

Models	Linear Regression	Logistic Regression	Random Forest	XG Boost
Accuracy	0.5854	0.6514	0.6774	0.6834

Table 1: Accuracy of Different Models while predicting Life Satisfaction

C) Predicting Life Satisfaction Via Social Factors

Conclusion

- An important factor to reducing overall poverty is expanding educational opportunities, as higher educational attainment was linked to lower poverty and unemployment rates
- The high proportion of suicide cases among those with only high school degrees warrants suggests greater need for supporting economically weaker areas in the United States, in particular adult men
- Tracking citizens' social features can be a useful tool in predicting life satisfaction.

References

[1] C. Guven and R. Saloumidis, "Life satisfaction and longevity: longitudinal evidence from the german socio-economic panel," German economic review, vol. 15, no. 4, pp. 453–472, 2014.

[2] H. Lee and G. K. Singh, "Inequalities in life expectancy and all-cause mortality in the united states by levels of happiness and life satisfaction: A longitudinal study," International Journal of Maternal and Child Health and AIDS, vol. 9, no. 3, p. 305, 2020.

[3] W. Zhang, K. L. Braun, and Y. Y. Wu, "The educational, racial and gender crossovers in life satisfaction: Findings from the longitudinal health and retirement study," Archives of gerontology and geriatrics, vol. 73, pp. 60–68, 2017.

[4] J. M., "29 eye-opening can money buy happiness statistics in 2021," 2022.

[5] "Us census bureau: <https://www.census.gov/>."

[6] CDC, "Centers for disease control and prevention: <https://www.kaggle.com/datasets/cdc/mortality?dataset> 2016.

[7] I. R., "World values survey: All rounds - country-pooled: <https://www.worldvaluessurvey.org/wvscontents.jsp>," 2021.