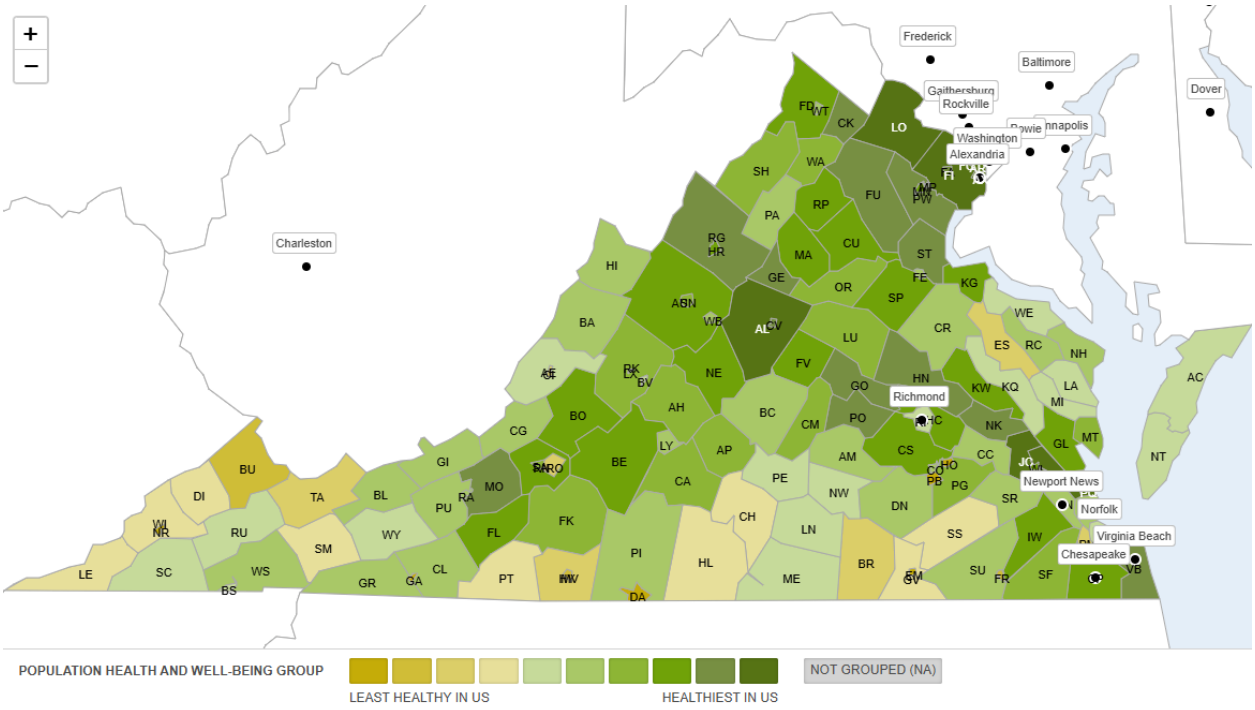
Insert heading

Group Name



Introduction

Although life expectancy is often regarded as a concise indicator of a community's overall well-being, notable disparities still exist even within a single state. Virginia's 133 counties and independent cities have an average life expectancy of just over seventy-seven years, according to the 2025 County Health Rankings report, highlighting disparities in the state's progress toward public health objectives (County Health Rankings & Roadmaps, 2025). Three recurrent factors that influence longevity are household economic resources, the burden of chronic diseases (obesity), and the characteristics of local environments (such as rural versus urban areas), according to decades of epidemiological research, including the Centers for Disease Control and Prevention's findings on social determinants of health (Hacker, 2022) and the World Health Organization's analyses of obesity-related mortality (WHO, 2025). Understanding how these factors manifest at the county level in Virginia can guide targeted interventions and equitable allocation of resources.

The overarching question guiding this study is:

Which socioeconomic, health-behavior, and environmental factors best explain the variation in 2025 life expectancy across Virginia's counties and independent cities?

Drawing on publicly available data from the County Health Rankings portal, we assembled a dataset in which each row represents a single jurisdiction, and the outcome of interest is the average life expectancy in years.

Three specific research questions are presented to focus the general investigation. First, *is the average life expectancy longer in jurisdictions with greater median household incomes?* This tackles the quantifiable relationship that previous national studies have proposed between longevity and economic prosperity. Second, *is life expectancy lower in counties with the highest tertile of adult obesity rates than in those with the lowest tertile?* This investigates the relationship between mortality risk and a quantifiably modifiable health behavior variable. Third, *is there a difference in mean life expectancy between Virginia's primarily rural and urban jurisdictions?* This question investigates whether geographic context alone confers a longevity advantage or disadvantage by contrasting a qualitative classification of place.

Data Summary

Data Sources

This analysis is based on the 2025 Virginia extract from the *County Health Rankings & Roadmaps program*, a collaborative effort between the University of Wisconsin Population Health Institute and the Robert Wood Johnson Foundation that gathers county-level health indicators nationwide each year. Before harmonizing the series to a single 2025 reference year, the institute gathers each metric directly from approved federal sources, including the Bureau of Labor Statistics, the American Community Survey, and the CDC WONDER database.

For this study, the population is the complete set of 133 Virginia counties and independent cities. Two companion tables provided by the Rankings, Select Measure Data and Additional Measure Data, contain (i) core health outcomes and (ii) socioeconomic and health-behavior covariates. These tables were combined on each county's five-digit FIPS code to create a single cross-sectional data frame. A continuous response variable—life expectancy at birth (years)—was retained exactly as reported, ensuring comparability with CDC methodology.

Several small, logically motivated changes were necessary. The ratio of primary care physicians was provided as a text string, such as "2,210:1." To ensure that lower figures indicate increased provider availability, it was transformed to a straightforward numerical count of residents per physician and then split into 5 bins "<=1k", "1-2k", "2-3k", "3-5k", ">=5k". To draw attention to non-linear gradients in built-environment resources, the percentage of people with access to exercise opportunities was recoded into an ordered three-level factor (Low, Mid, and High). Ultimately, the rural population percentage was divided into four equal-width groups, referred to as rural bands: 0–25%, 26–50%, 51–75%, and 76–100%. This discretization facilitates the understanding of regional differences while maintaining a monotonic ordering. After removing occasional entries with missing outcome values and one jurisdiction without a county name, 132 observations were included in the study sample.

The County Health Rankings have a high level of credibility, as they are frequently referenced in peer-reviewed health services research and utilize open documentation of data provenance and imputation procedures. However, three cautions are worth mentioning. First, particular data show suppressed numbers for a limited number of counties, which can contribute to the uncertainty. Second, there is no set time for retrieval because several focus areas of data are gathered in different years. Lastly, measurements are obtained from a wide range of reliable sources, each with its own unique data collection techniques, and then compiled into sheets. These restrictions will be reviewed when evaluating model findings; however, they do not

compromise the dataset's overall integrity.

Exploratory Data Analysis

EDA Summary

Methods and Analysis

Results

Conclusions

Appendix A: Data Dictionary

Variable Name	Abbreviated Name	Description

Appendix B: Data Rows

Appendix C: Final Model Output and Plots

Appendix D: References

Background

Data Sources

Additional Help