Understanding the Impact of Socioeconomic and Regional Factors on Health and Educational Outcomes

Rishika Randev, Jenny Wu, Uzoma Uwazurike Jr., Shiyue Zhou

Abstract

This study examines the relationships between socioeconomic, health, and education indicators in U.S. counties using data from the American Community Survey, the USDA Food Environment Atlas, small area income and poverty estimates, and other sources. We will explore regional influences and predictors of life expectancy, and also identify socioeconomic factors associated with school funding adequacy. By analyzing differences in life expectancy and education funding across socioeconomic and regional contexts, this study aims to gain insight into public well-being and economic inequality across the U.S. Through exploratory data analysis and regression modeling with interaction terms, we identify significant regional disparities in both life expectancy and school funding adequacy. The findings highlight that regions, particularly the West and Northeast, tend to have better health and education outcomes, emphasizing the critical role of location and economic conditions in shaping societal well-being.

Introduction

Life expectancy and school funding are important indicators of public well-being because they capture fundamental aspects of health, education, and social equity that are essential for the development of individuals and communities. Life expectancy is an important health indicator that reflects not only the overall health of a society, but also its stability, economic resilience, and levels of inequality. According to "The Root Causes of Health Inequity - Communities in Action" report from the National Institutes of Health, we can infer that differences in life expectancy between regions and socioeconomic groups often reveal systemic inequalities, with marginalized communities experiencing shorter life spans due to limited access to health care, healthy living environments, and economic opportunities.[1] Similarly, adequately funded schools ensure that students have equal access to learning opportunities, serving as a crucial investment in social mobility. The Stanford Center for Opportunity Policy in Education highlights that "the unequal allocation and inadequate levels of resources in schools and communities [are] at the heart of many gaps in student opportunity."[2] This emphasizes the importance of equitable and adequate school funding in closing opportunity gaps and empowering students from disadvantaged backgrounds to achieve academic and economic success. Consequently, it can be inferred that underfunded schools, often situated in economically disadvantaged or segregated areas, exacerbate educational disparities and limit students' ability to break the cycle of poverty. These gaps are rooted in systemic inequalities, such as residential segregation, which concentrates poverty, limits access to quality education and healthcare, and perpetuates income and health disparities across generations. By addressing these inequalities, policymakers can foster fairer, more prosperous communities and ensure that everyone benefits from improved health and education systems.

For this purpose, we focus our analysis on two key research questions. First, we investigate how socioeconomic factors—residential segregation, access to healthy foods, and health insurance coverage—impact average life expectancy. **We seek to answer the question, "What is the impact of residential segregation, access to healthy foods, and health insurance coverage on a county's average life expectancy?".** This analysis includes a comparison across different regions to identify variations in the relationship between residential segregation and life expectancy. Second, we evaluate how residential segregation and region influence school funding adequacy. **We seek the answer the question, "Does residential segregation and region impact the likelihood of a school being adequately funded?"**. This question allows us to assess how the racial makeup and geographic location—given that each region of the U.S. is characterized by differing political and cultural norms—affect a county's educational investment. The observational unit for this analysis is the U.S. county.

To obtain data on the aforementioned variables, we used three primary sources.

The first source was the County Health Rankings & Roadmaps program,[3] associated with the University of Wisconsin Population Health Institute, which provided us with 2024 datasets for all U.S. states. This program consolidates the latest county-level measurements of various population health, economic, demographic, and social factors from the American Community Survey, the USDA Food Environment Atlas, Small Area Income and Poverty Estimates, and other reliable government sources into publicly available datasets every year. Our specific variables of interest were originally collected as part of:

- National Center for Health Statistics - Natality & Mortality Files, 2019-2021 (average life expectancy)
- American Community Survey 5-year estimates, 2018-2022 (residential segregation index & median household income)
- USDA Food Environment Atlas, 2019 (percentage of population with limited access to healthy foods)
- Small Area Health Insurance Estimates, 2021 (percentage of adults uninsured)

- School Finance Indicators Database, 2021 (school funding adequacy)

The second source was IPUMS, which provided us with median poverty tax data for every county from 2018-2022.[4]

The third source was the U.S. Census Bureau's Regions and Divisions of the United States, which maps every U.S. to one of four geographic regions, and also a division within the regions. We specifically included regions as a variable in this analysis. This data was loaded into R using a publicly available GitHub repository titled *Census Regions*.[5]

Methods

Variable Selection

Variables of interest were selected *a priori* based on their perceived relevance to the two outcomes we wanted to better understand: average life expectancy and school funding adequacy. While nutrition and access to quality healthcare are inherently linked to general wellbeing, it is worthwhile to assess how influential the environmental and financial aspects of these factors are on our tangible outcome. Therefore, we selected two key predictors for our life expectancy model: the percentage of the population with limited access to healthy foods and the percentage of uninsured adults in a county. The former is defined by the US Department of Agriculture as "the percentage of the [county] population that is low-income and does not live close to a grocery store", where "close" is defined differently for rural and nonrural areas.[6] The latter variable represents the percentage of adults in a county younger than 65 who lack health insurance. We also were interested in seeing how residential segregation would impact life expectancy among these predictors, given existing studies on the negative health effects of segregation, and whether this relationship would exhibit any variations across different regions. This is why we chose to include residential segregation index values. The residential segregation index, used by the County Health Rankings & Roadmaps program, measures how evenly two groups, such as Black and white residents, are distributed across geographic areas, with values ranging from 0 (complete integration) to 100 (complete segregation).[7]

Income is known to have a strong relationship with.[8]

School funding adequacy is defined by the County Health Rankings program as "the average dollar gap between actual per-pupil spending and the amount needed for students to achieve national average test scores, considering districts' varying equity-based needs."[9] To explore how racial diversity interface with variations in school funding, we selected residential segregation to be a part of this regression. In addition, given that state policy has a large influence on school administration and the determination of school needs, and that states in the same region tend to have similar political leanings and cultural beliefs, we wanted to analyze how much of a relationship exists between region and actual school funding levels.

Because a major source of public school funding is tax revenue, median property tax and median household income were also included in the school funding model as potential confounders. This allowed us to assess whether the effects of segregation and region on funding existed outside of the effects of known economic factors.

Exploratory Data Analysis & Data Preprocessing

2024 datasets for all U.S. states were combined and merged with median property tax and region mapping data, resulting in a final dataset with 3144 county observations and 9 different variables (state, average life expectancy, residential segregation index, % limited access to healthy foods, % uninsured adults, school funding adequacy, median household income, median property tax, and region). Because school funding adequacy was given as a numerical value, and we wanted to focus on better understanding the dichotomy between school funding adequacy and inadequacy, this variable was converted into a binary categorical variable (where the value was set to 1 if the numerical value was positive, indicating the counties' schools were adequately funded, and set to 0 if the numerical value was negative, indicating underfunding).

Exploratory data analysis for the two outcome variables was conducted by 1) calculating descriptive statistics, 2) creating scatterplots of each individual predictor against the outcome (for life expectancy), and 3) creating pairwise plots between variables (for school funding). During the initial stages of analysis, we found that of our observations were missing residential segregation index data, and that certain states were missing nearly all of their segregation values. Because of this, we decided to drop all states missing over 50% of their values for any variable. This led to our analysis being limited to 33 states, and excluding the following: Alaska, Colorado, Idaho, Iowa, Kansas, Minnesota, Montana, Nebraska, North Dakota, South Dakota, Utah, Wyoming, and Vermont. We were left with 2379 total observations, and the missing percentage for residential segregation dropped to 22%.

After this, remaining missing values in all variables were imputed using the mice package and predictive mean matching to generate one complete dataset. The original school funding adequacy numerical variable was first imputed, and then converted into binary values for subsequent analysis.

Model Fitting & Assessment

For our first research question, a multiple linear regression model was fit, regressing average life expectancy on residential segregation index, percentage with limited access to healthy foods, percentage uninsured adults, region, and median household income. Linear regression assumptions were assessed using diagnostic plots, especially residual vs. fitted and quantile-quantile plots. The adjusted R-squared value was used to evaluate the fit of the model, and the effect of region as an interaction term with residential segregation was evaluated using nested F tests. We also calculated VIF to check for multicollinearity and Cook's distance to identify influential points.

For our second research question, a binary logistic regression model was fit with school funding adequacy (categorical) as the outcome and residential segregation index, region, median property tax, and median household income as predictors. The model was assessed using a confusion matrix, an ROC curve, and a comparison of deviance between the full model and a reduced model that excluded region as a predictor.

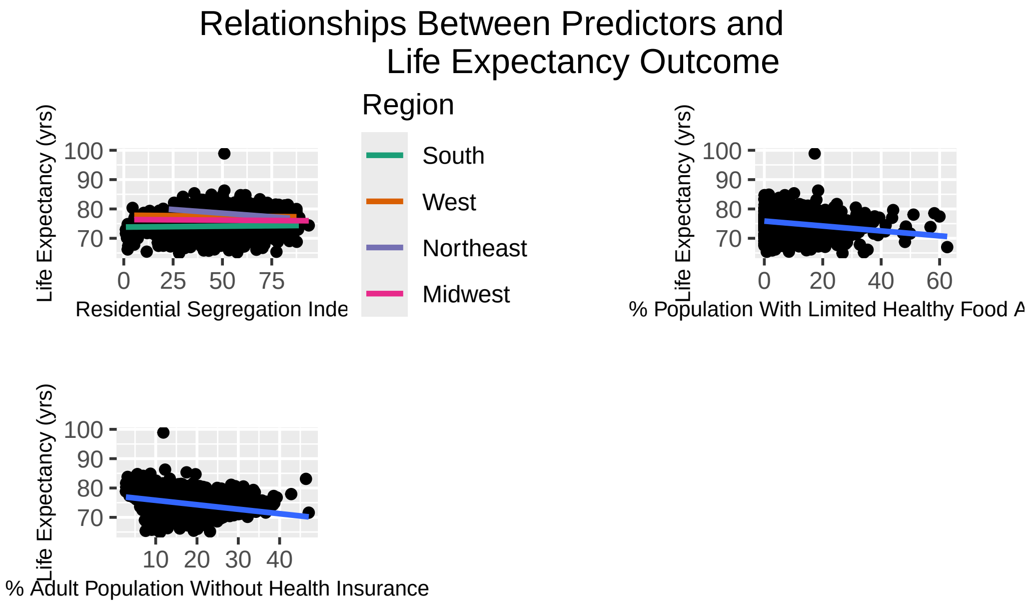
Results

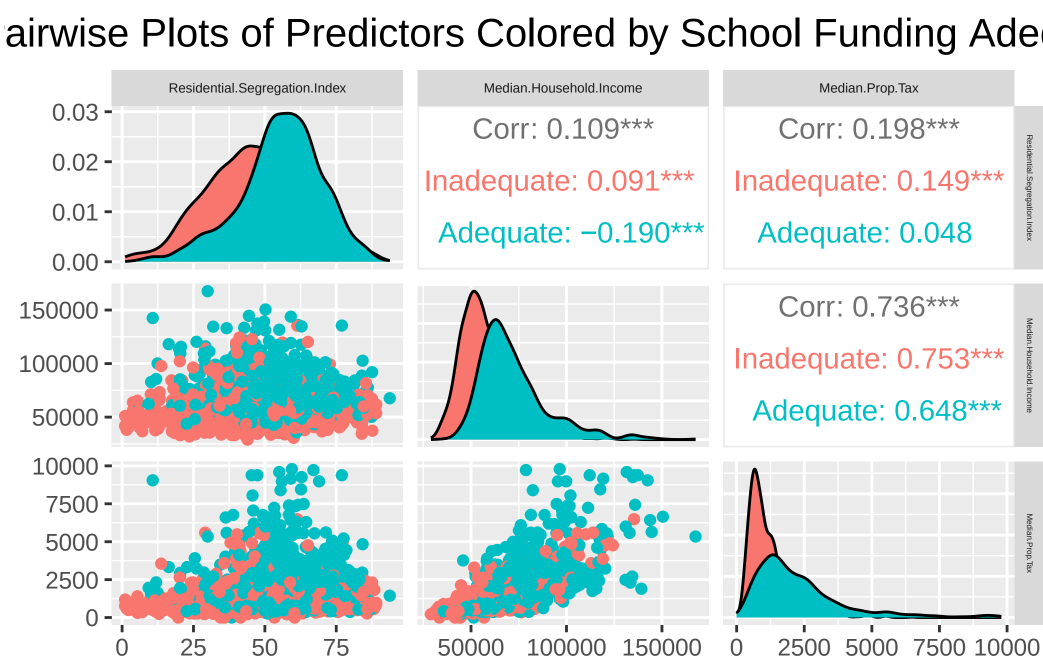
Data Overview

Table 1 displays summary statistics for the predictors in the second model, stratified by school funding adequacy. School funding adequacy counts show that around 66% of the counties in our analysis had under funded schools. Additionally, the number of counties per region indicate an imbalance between categories, with the South region being over represented (holding 60% of all counties), and the East and Northeast holding around 9% of the counties each. Table A1 in the Appendix provides summary statistics for all variables, and shows that both the mean and median for life expectancy are approximately 75 years.

Table 1: School Funding Adequacy Summary Statistics

Variable	**Inadequate School Funding**	**Adequate School Funding**
Residential Segregation Index: Mean (SD)	46 (17)	56 (14)
Region-West: Count (%)	114 (7.3%)	89 (10.8%)
Region-Northeast: Count (%)	3 (0.2%)	200 (24.3%)
Region-Midwest: Count (%)	267 (17.1%)	285 (34.7%)
Region-South: Count (%)	1173 (75.3%)	248 (30.2%)
Median Property Tax, USD: Mean (SD)	1119 (763)	2282 (1648)
Median Household Income, USD: Mean (SD)	56513 (13200)	72218 (18255)





Question 1

Model Results

Based on the multiple linear regression results, percentage of uninsured adults, percentage of population with limited access to healthy foods, and region all emerged with statistically significant coefficient values. The interaction term between region and residential segregation, as well as the individual term of residential segregation, were not statistically significant. While percentage of uninsured adults and percentage with limited access to healthy foods were statistically significant, the coefficients themselves indicate very minor impacts on life expectancy (a 1 unit increase in either of these percentages leads to average life expectancy decreasing by around 0.05 years). The impact of region on life expectancy, however, was notable: life expectancy is on average 4.17 years higher in the West region compared to the South, 6.7 years higher in the Northeast compared to the South, and 2.41 years higher in the Midwest compared to the South.

```
Call:
lm(formula = Life.Expectancy ~ Residential.Segregation.Index *
    Region_fac + X..Limited.Access.to.Healthy.Foods + X..Uninsured.Adults +
    Median.Household.Income, data = combined)

Residuals:
   Min     1Q Median     3Q    Max
-8.119 -1.097  0.005  1.074  7.899

Coefficients:
                                                   Estimate Std. Error t value
(Intercept)                                       6.638e+01  2.935e-01 226.178
Residential.Segregation.Index                     5.641e-03  3.367e-03   1.675
Region_facWest                                    2.996e-01  7.194e-01   0.416
Region_facNortheast                               2.280e+00  8.292e-01   2.749
Region_facMidwest                                 1.221e+00  5.071e-01   2.409
X..Limited.Access.to.Healthy.Foods               -2.276e-02  7.793e-03  -2.921
X..Uninsured.Adults                               1.589e-02  8.500e-03   1.870
Median.Household.Income                           1.249e-04  2.746e-06  45.486
Residential.Segregation.Index:Region_facWest      2.159e-02  1.200e-02   1.799
Residential.Segregation.Index:Region_facNortheast -5.906e-03  1.367e-02  -0.432
Residential.Segregation.Index:Region_facMidwest    4.021e-04  8.644e-03   0.047
                                                 Pr(>|t|)
(Intercept)                                       < 2e-16 ***
Residential.Segregation.Index                     0.09404 .
Region_facWest                                    0.67711
Region_facNortheast                               0.00603 **
Region_facMidwest                                 0.01610 *
X..Limited.Access.to.Healthy.Foods                0.00353 **
X..Uninsured.Adults                               0.06170 .
Median.Household.Income                           < 2e-16 ***
Residential.Segregation.Index:Region_facWest      0.07222 .
Residential.Segregation.Index:Region_facNortheast  0.66573
Residential.Segregation.Index:Region_facMidwest    0.96291
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.803 on 1819 degrees of freedom
  (549 observations deleted due to missingness)
Multiple R-squared:  0.674, Adjusted R-squared:  0.6722
F-statistic: 376.1 on 10 and 1819 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Life.Expectancy ~ Residential.Segregation.Index *
    Region_fac + X..Limited.Access.to.Healthy.Foods + X..Uninsured.Adults +
    Median.Household.Income, data = completed_data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.4382 -1.1413 -0.0355  1.1304 20.3333

Coefficients:
                                               Estimate Std. Error t value
(Intercept)                                   6.588e+01  2.742e-01 240.246
Residential.Segregation.Index                 7.154e-03  3.319e-03   2.156
Region_facWest                                1.961e+00  6.926e-01   2.832
Region_facNortheast                           2.917e+00  8.426e-01   3.463
Region_facMidwest                             1.915e+00  4.353e-01   4.398
X..Limited.Access.to.Healthy.Foods            4.038e-03  6.220e-03   0.649
X..Uninsured.Adults                           3.285e-02  7.503e-03   4.378
Median.Household.Income                       1.254e-04  2.666e-06  47.046
Residential.Segregation.Index:Region_facWest      -3.164e-04  1.164e-02  -0.027
Residential.Segregation.Index:Region_facNortheast -1.264e-02  1.391e-02  -0.909
Residential.Segregation.Index:Region_facMidwest   -7.524e-03  7.501e-03  -1.003
                                                  Pr(>|t|)
(Intercept)                                        < 2e-16 ***
Residential.Segregation.Index                     0.031217 *
Region_facWest                                    0.004667 **
Region_facNortheast                               0.000545 ***
Region_facMidwest                                 1.14e-05 ***
X..Limited.Access.to.Healthy.Foods                0.516251
X..Uninsured.Adults                               1.25e-05 ***
Median.Household.Income                            < 2e-16 ***
Residential.Segregation.Index:Region_facWest      0.978312
Residential.Segregation.Index:Region_facNortheast 0.363617
Residential.Segregation.Index:Region_facMidwest   0.315926
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.964 on 2368 degrees of freedom
Multiple R-squared:  0.6081,    Adjusted R-squared:  0.6065
F-statistic: 367.4 on 10 and 2368 DF,  p-value: < 2.2e-16

Analysis of Variance Table
```

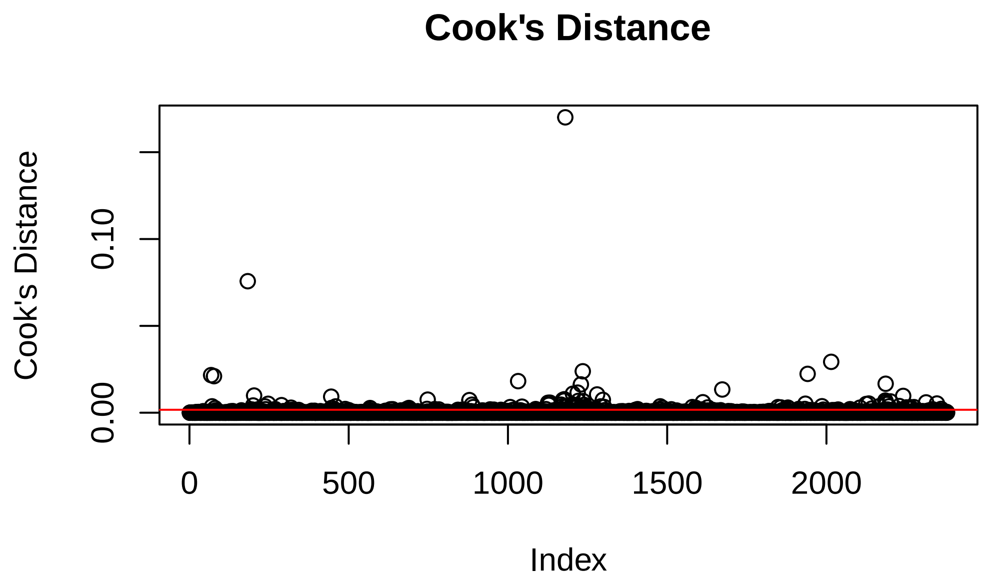
```
Model 1: Life.Expectancy ~ Residential.Segregation.Index * Region_fac +
    X..Limited.Access.to.Healthy.Foods + X..Uninsured.Adults +
    Median.Household.Income
Model 2: Life.Expectancy ~ Residential.Segregation.Index + X..Limited.Access.to.Healthy.Food
    X..Uninsured.Adults + Median.Household.Income + Region_fac
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1   2368 9135.7
2   2371 9142.1 -3   -6.4651 0.5586 0.6424
```

Model Assessment

The adjusted R-squared value for the model was around 0.2, indicating that only a small amount of the variability in life expectancy was explained by the predictors in the model. This points to the fact that are a plethora of other socioeconomic and environmental factors that have an effect on life expectancy, and expanding our model to include more of these would have likely yielded a better fit.

In conducting diagnostics on the multilinear regression model, we first checked to see if the model meets our assumptions of linearity, independence, normality, and homoscedasticity. The residuals vs. fitted plot showed relatively equal variance and no discernible pattern, meaning that the linearity and homoscedasticity assumptions were reasonably met. In addition, because each observation in the dataset represented a unique county in the U.S., independence can be reasonably assumed. Assessing the Q-Q Plot for normality, the residual distribution was approximately normal, despite a slight divergence on both tails of the plot. Finally, through observing the Residuals vs. Leverage plot (not pictured), we observed only a few points falling near or beyond the Cook's distance threshold, suggesting that there may be a few influential points that can be seen as significantly impacting the results of the regression. These points could affect the validity of the model.

With variance inflation factors of roughly less than 2 for all predictors, there was little to no multicollinearity among the predictors. Through a nested-f test comparing a full model with a reduced model that excludes the interaction term between region and residential segregation, we concluded that a full model fit the data substantially better. The reduction in RSS in the full model compared to the reduced model was highly significant, indicating that the interaction between region and residential segregation added meaningful explanatory power to the model.




```
  68   71   77   80  183  193  195  200  203  237  239  247  269  289  319  445
  68   71   77   80  183  193  195  200  203  237  239  247  269  289  319  445
 446  456  459  489  567  631  638  683  689  746  748  770  783  859  879  887
 446  456  459  489  567  631  638  683  689  746  748  770  783  859  879  887
 891  944  956 1007 1022 1032 1044 1087 1119 1126 1128 1132 1167 1170 1173 1175
 891  944  956 1007 1022 1032 1044 1087 1119 1126 1128 1132 1167 1170 1173 1175
1176 1177 1179 1180 1183 1202 1204 1206 1209 1210 1213 1214 1215 1217 1219 1221
1176 1177 1179 1180 1183 1202 1204 1206 1209 1210 1213 1214 1215 1217 1219 1221
1224 1226 1229 1231 1235 1236 1239 1241 1242 1245 1253 1271 1274 1280 1281 1291
1224 1226 1229 1231 1235 1236 1239 1241 1242 1245 1253 1271 1274 1280 1281 1291
1298 1302 1407 1473 1478 1480 1485 1513 1579 1586 1589 1609 1612 1627 1673 1848
1298 1302 1407 1473 1478 1480 1485 1513 1579 1586 1589 1609 1612 1627 1673 1848
1857 1867 1874 1879 1882 1913 1932 1934 1941 1986 1990 2015 2037 2074 2105 2125
1857 1867 1874 1879 1882 1913 1932 1934 1941 1986 1990 2015 2037 2074 2105 2125
2133 2142 2166 2184 2185 2186 2187 2189 2192 2196 2201 2229 2241 2243 2251 2261
2133 2142 2166 2184 2185 2186 2187 2189 2192 2196 2201 2229 2241 2243 2251 2261
2263 2275 2313 2333 2347 2360
2263 2275 2313 2333 2347 2360
```

```
Call:
lm(formula = Life.Expectancy ~ Residential.Segregation.Index *
```

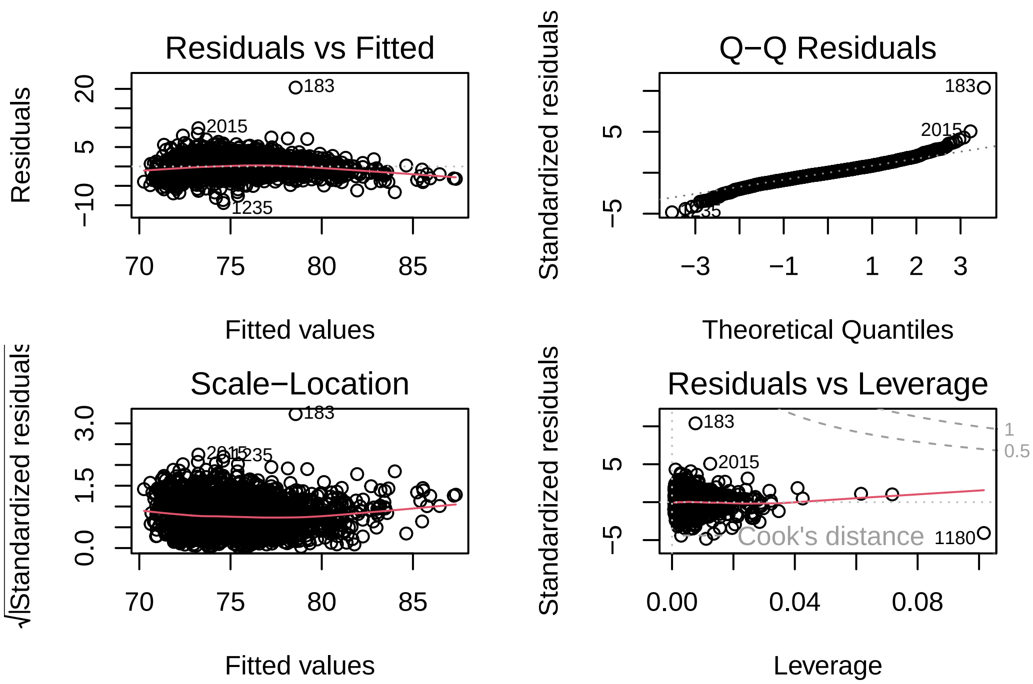
```
    Region_fac + X..Limited.Access.to.Healthy.Foods + X..Uninsured.Adults +
    Median.Household.Income, data = completed_data)

Residuals:
    Min      1Q  Median      3Q     Max
-9.4382 -1.1413 -0.0355  1.1304 20.3333

Coefficients:
                                                  Estimate Std. Error t value
(Intercept)                                      6.588e+01  2.742e-01 240.246
Residential.Segregation.Index                    7.154e-03  3.319e-03   2.156
Region_facWest                                   1.961e+00  6.926e-01   2.832
Region_facNortheast                              2.917e+00  8.426e-01   3.463
Region_facMidwest                                1.915e+00  4.353e-01   4.398
X..Limited.Access.to.Healthy.Foods               4.038e-03  6.220e-03   0.649
X..Uninsured.Adults                              3.285e-02  7.503e-03   4.378
Median.Household.Income                          1.254e-04  2.666e-06  47.046
Residential.Segregation.Index:Region_facWest    -3.164e-04  1.164e-02  -0.027
Residential.Segregation.Index:Region_facNortheast -1.264e-02  1.391e-02  -0.909
Residential.Segregation.Index:Region_facMidwest  -7.524e-03  7.501e-03  -1.003
                                                 Pr(>|t|)
(Intercept)                                       < 2e-16 ***
Residential.Segregation.Index                    0.031217 *
Region_facWest                                   0.004667 **
Region_facNortheast                              0.000545 ***
Region_facMidwest                                1.14e-05 ***
X..Limited.Access.to.Healthy.Foods               0.516251
X..Uninsured.Adults                              1.25e-05 ***
Median.Household.Income                           < 2e-16 ***
Residential.Segregation.Index:Region_facWest     0.978312
Residential.Segregation.Index:Region_facNortheast 0.363617
Residential.Segregation.Index:Region_facMidwest  0.315926
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.964 on 2368 degrees of freedom
Multiple R-squared:  0.6081,    Adjusted R-squared:  0.6065
F-statistic: 367.4 on 10 and 2368 DF,  p-value: < 2.2e-16


Call:
lm(formula = Life.Expectancy ~ Residential.Segregation.Index *
```

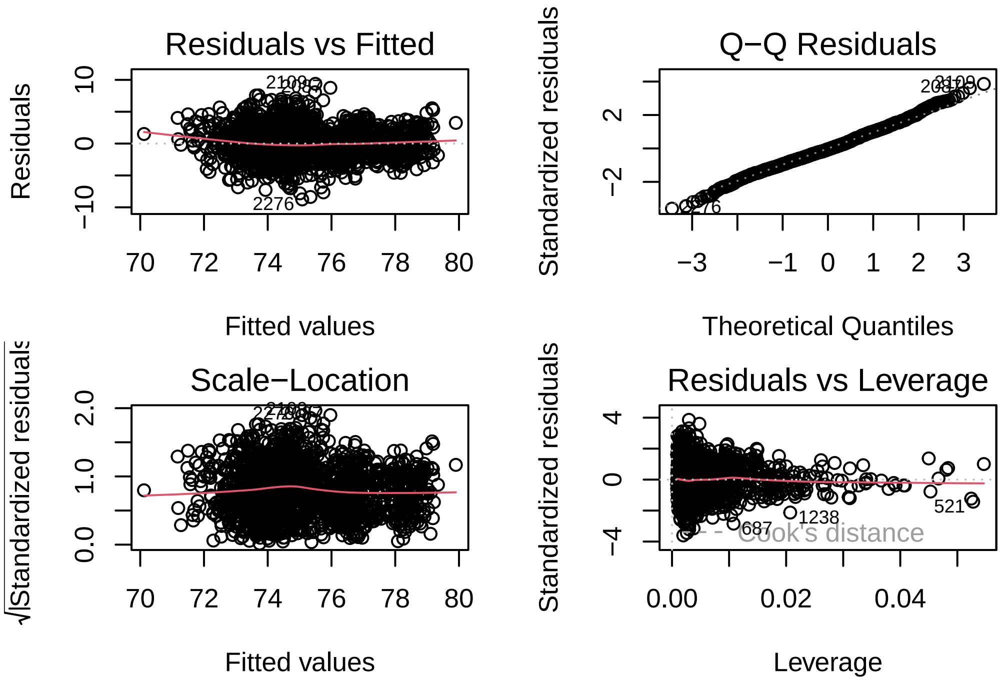
```
    Region_fac + X..Limited.Access.to.Healthy.Foods + X..Uninsured.Adults,
    data = combined_no_influential)

Residuals:
    Min      1Q  Median      3Q     Max
-8.7599 -1.5947 -0.1337  1.5741  9.3751

Coefficients:
                                                   Estimate Std. Error t value
(Intercept)                                       76.102549   0.284684 267.323
Residential.Segregation.Index                      0.007285   0.004708   1.547
Region_facWest                                     3.691181   1.146371   3.220
Region_facNortheast                                7.332445   1.375163   5.332
Region_facMidwest                                  2.520155   0.720384   3.498
X..Limited.Access.to.Healthy.Foods                -0.100848   0.011255  -8.960
X..Uninsured.Adults                               -0.088596   0.011450  -7.738
Residential.Segregation.Index:Region_facWest      -0.007755   0.019175  -0.404
Residential.Segregation.Index:Region_facNortheast -0.079279   0.022453  -3.531
Residential.Segregation.Index:Region_facMidwest   -0.020237   0.012291  -1.646
                                                  Pr(>|t|)
(Intercept)                                        < 2e-16 ***
Residential.Segregation.Index                     0.121969
Region_facWest                                    0.001306 **
Region_facNortheast                                1.1e-07 ***
Region_facMidwest                                 0.000480 ***
X..Limited.Access.to.Healthy.Foods                 < 2e-16 ***
X..Uninsured.Adults                                1.7e-14 ***
Residential.Segregation.Index:Region_facWest      0.685953
Residential.Segregation.Index:Region_facNortheast 0.000425 ***
Residential.Segregation.Index:Region_facMidwest   0.099846 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.434 on 1736 degrees of freedom
  (499 observations deleted due to missingness)
Multiple R-squared:  0.3302,    Adjusted R-squared:  0.3267
F-statistic: 95.08 on 9 and 1736 DF,  p-value: < 2.2e-16


      rstudent unadjusted p-value Bonferroni p
183  10.635104         7.7581e-26   1.8457e-22
2015  5.084328         3.9783e-07   9.4644e-04
1235 -4.855097         1.2817e-06   3.0493e-03
```

```
2347 -4.435539          9.6043e-06   2.2849e-02
1874  4.316914          1.6474e-05   3.9191e-02

numeric(0)
```

```
                                                  GVIF Df GVIF^(1/(2*Df))
Residential.Segregation.Index                 1.855444  1        1.362147
Region_fac                                15046.549757  3        4.968662
X..Limited.Access.to.Healthy.Foods            1.146672  1        1.070828
X..Uninsured.Adults                           1.534443  1        1.238726
Median.Household.Income                       1.248394  1        1.117316
Residential.Segregation.Index:Region_fac  17150.584230  3        5.078239
```

```
                                       GVIF Df GVIF^(1/(2*Df))
Residential.Segregation.Index      1.330770  1        1.153590
X..Limited.Access.to.Healthy.Foods 1.138447  1        1.066980
X..Uninsured.Adults                1.534371  1        1.238697
Median.Household.Income            1.241183  1        1.114084
Region_fac                         1.855213  3        1.108491
```





Question 2

Model Results

The table below displays the odds ratio results from the logistic regression model for school funding adequacy, with the South as the baseline category for the categorical region variable. While all estimates are statistically significant at the 0.05 significance level, exponentiated regression coefficients from the logistic model demonstrate that the level of residential segregation, the median household income, and median property tax of the county have very minimal impacts on the odds of a school being adequately funded.

Region, however, has a substantial impact on the odds of school funding adequacy. Being located in the Northeast or the Midwest has the greatest impact on how well the school is funded, with the odds of adequate school funding being over 200 times higher in a Northeast county compared to a Southern county, and the odds of adequate school funding being over 3 times higher in a Northeast county compared to a Southern county, holding all other variables constant. Such disparities across the country are caused by a number of factors, including: (1) capacity - how well off a state is based on its economy and resources, and (2) effort - the states willingness to provide funding for education. Wealthier states with a high fiscal capacity, (typically those in the Northeast), have more funding available to spend on education than states with more limited resources (typically those in the South and the West).

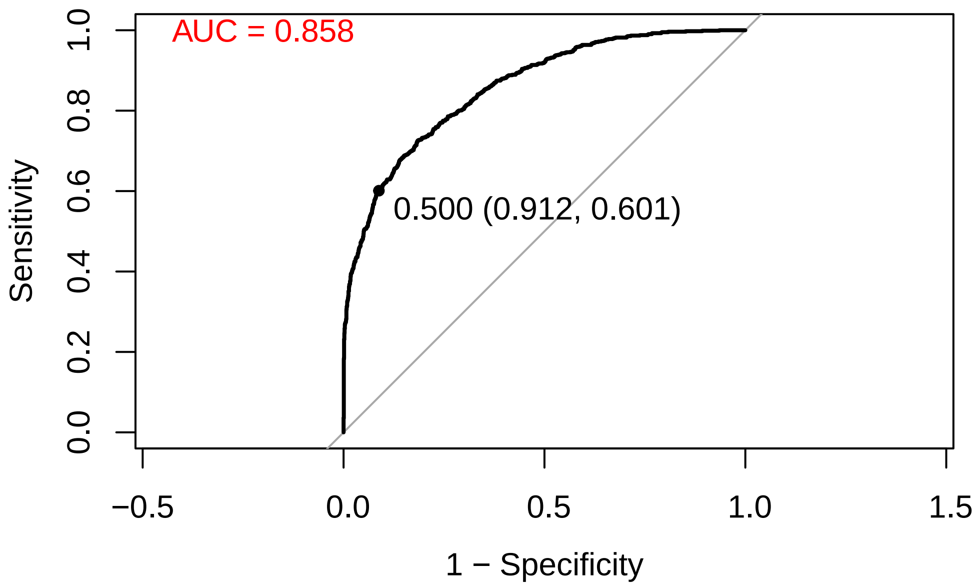
In comparing our full model (with predictor variables for residential segregation, median household income, median property tax, and region) to a model with including all of these variables

	(1)				
	Est.	S.E.	2.5 %	97.5 %	p
(Intercept)	0.001 35	0.000 50	0.000 64	0.002 73	<0.01
Residential Segregation Index (RSI)	1.0196	0.0041	1.0117	1.0277	<0.01
Median Household Income, USD	1.0	5.5×10^{-6}	1.0	1.0	<0.01
Median Property Tax, USD	1.0	8.5×10^{-5}	1.0	1.0	0.01
Region-West	1.83	0.35	1.26	2.66	<0.01
Region-Northeast	232	141	82	974	<0.01
Region-Midwest	3.84	0.52	2.95	5.02	<0.01

except for region, we found that there is statistical significance in the deviance between the residuals of the full model and reduced. As such, we can conclude that including region as a predictor refines and creates a better fit for our model.

Model Assessment

Evaluating our model, we found that the logistic regression model demonstrated a strong overall performance with an accuracy rate of 80.33%. The model had a high sensitivity of 91%, indicating that it accurately predicted 91% of actual positive cases of adequate school funding when evaluated on our dataset. However, it had a 60% specificity rate, highlighting it correctly identified cases of adequate school funding only 60% of the time. Moreover, looking at the ROC curve below, the AUC is 0.858, indicating a moderately strong discriminate ability.



Conclusion

By analyzing these two research questions, we observed the significant impact of region on both life expectancy (a health indicator) and school funding (an education indicator), both of which are key components of societal well-being. These indices tend to be higher in the West and Northeast, which are more developed regions. The findings provide empirical evidence supporting the significant disparities in life expectancy across regions with differing economic conditions. They also provide insights into social policy investment decisions, highlighting the importance of fair resource allocation to people's well-being.

Our analysis faced limitations due to missing data, causing our dataset to be restricted to 33 states. Additionally, because The County Health Rankings program consolidates the latest from a variety of governmental sources, most of our variables were pulled from different years between 2018 and 2022; we assumed that trends remained largely consistent between these years, treating time as insignificant, but this may not actually be the case.

Future studies could incorporate more social variables to enhance the analysis and investigate the causal relationships between region and outcomes. Expanding the dataset to include more states and addressing missing data would also provide a more comprehensive understanding.

	Unique	Missing Pct.	Mean	SD	Min	Median	Max
Life.Expectancy	2355	0	75.2	3.1	64.9	75.1	98.9
Residential.Segregation.Index	1850	0	49.3	16.5	1.1	50.3	93.8
X..Limited.Access.to.Healthy.Foods	2345	0	8.2	6.9	0.0	6.5	62.6
X..Uninsured.Adults	2375	0	14.1	6.6	2.8	12.6	47.1
Median.Household.Income	2325	0	61939.5	16877.8	28972.0	58623.0	167605
School.Funding.Adequacy	2345	0	-2993.6	8331.9	-63405.9	-1790.5	29107.
Median.Prop.Tax	1584	0	1520.7	1274.7	0.0	1173.0	9788.0
School.Funding.Cat	2	0	0.3	0.5	0.0	0.0	1.0
Region_fac	N	%					
South	1421	59.7					
West	203	8.5					
Northeast	203	8.5					
Midwest	552	23.2					

Appendix

Table A1: Summary Statistics For Numerical Variables

Statistic	Residential Segregation Index (RSI)	Percentage With Limited Access to Healthy Foods	Percentage of Uninsured Adults	Median Household Income, USD	Median Property Tax, USD	Average Life Expectancy
Min	1.0	0.0	2.8	28972	0	64.9
Median	50.3	6.5	12.6	58623	1173	75.1
Mean	49.3	8.2	14.1	61940	1521	75.2
Max	93.8	62.6	47.1	167605	9788	98.9
Standard Deviation	16.5	6.9	6.6	16877.8	1274.7	3.1

References

1. National Academies of Sciences, Engineering, and Medicine. (2017). *Communities in action: Pathways to health equity.* Washington, DC: The National Academies Press. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK425845/

2. Stanford Center for Opportunity Policy in Education. (2013.). How to close the opportunity gap: Key policy recommendations. Retrieved from https://edpolicy.stanford.edu/sites/default/files/Opp%20Gap%20Policy%20Recommendations.pdf

3. County Health Rankings & Roadmaps. (n.d.). *About us.* Retrieved December 8, 2024, from https://www.countyhealthrankings.org/about-us

4. Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rodgers, and Megan Schouweiler. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS, 2024. Retrieved from https://doi.org/10.18128/D010.V15.0

5. Halpert, C. (2014). Census regions [GitHub repository]. GitHub. https://github.com/cphalpert/census-regions

6. USDA. (2022). *Documentation.* Economic Research Service. https://www.ers.usda.gov/data-products/food-access-research-atlas/documentation/

7. County Health Rankings & Roadmaps. (2024). *Residential segregation - Black/White.* University of Wisconsin Population Health Institute. Retrieved from https://www.countyhealthrankings.org/health-data/health-factors/social-economic-factors/family-and-social-support/residential-segregation-blackwhite?utm_source=chatgpt.com&year=2024

8. Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*, *315*(16), 1750–1766. https://doi.org/10.1001/jama.2016.4226

9. County Health Rankings & Roadmaps. (2024). *School funding adequacy.* University of Wisconsin Population Health Institute. Retrieved from https://www.countyhealthrankings.org/health-data/health-factors/social-economic-factors/education/school-funding-adequacy?utm_source=chatgpt.com&year=2024