Exploring the Socio-Economic factors that affect a Country's Life Expectancy

Group 6
Principal Investigators:
Muzi Li (mli737@wisc.edu), Evelyn Tan (ztan75@wisc.edu),
Kshitij Galav (galav@wisc.edu)

Introduction
This project examines the complex relationship between important socio-economic factors and life expectancy. We find that economic prosperity is often accompanied by higher life expectancy, implying an increase in socio-economic status and population health as a result of countries becoming richer. To delve deeper into how the economy affects life expectancy, we turn our focus to health expenditure and primary education attainment. The study uses data visualization and regression models to quantify and elucidate the impact of these variables on life expectancy, enhancing our understanding of the intricate interactions between economic growth, education, healthcare and social well-being.

Description of the data
Our data comes from the World Bank's DataBank, which is a reliable platform. Considering that some data before 2000 are missing, we decided to analyze the data from 2000 to 2019 and excluded any cases of missing values when processing the data. Additionally, one parameter, "Primary_Education_Attainment_Percent" has partial data missing. We have chosen to utilize only the 1200 entries with numerical values to ensure that our analysis is based on reliable and complete information. Next, we grouped the data by year and calculated the average value of each parameter for each year to obtain a more detailed time trend. Below, are the definitions of some of the variables we use for our analysis:
- "Life_Expectancy": Total life expectancy at birth
- "Health_Expend_Per_Capita": Health expenditure per capita in current US $
- "GDP_Per_Capita": GDP per capita in current US $
- "Primary_Education_Attainment_Percent": Percent of population 25 and above with at least completed primary education (referenced as 'Education Attainment' in the report)

Methodology
In terms of data visualization, we mainly use three methods: scatter plots, bar plots, and line plots. In addition, in order to better demonstrate the relationship between parameters, we take the logarithm of GDP per capita and health expenditure to emphasize the trend between data by avoiding clustering of data due to the difference in size of the data. In order to further analyze the relevance of these relationships, we used a linear regression model to analyze the relationship between parameters and show the degree and direction of the influence of the independent variables on the dependent variables.

Analysis
Plotting
1. Overall trend analysis
Initially, our goal was to conduct a comprehensive analysis of four key variables: mean life expectancy, mean GDP per capita, mean health expenditure per capita, and mean primary education attainment. To facilitate observation and comparison of their trends, we organized subplots in a single column. Examining the subplots, we observed a consistent increase in mean life expectancy. Additionally, there was an apparent positive trend in mean GDP per capita and mean health expenditure per capita, with minor fluctuations. Notably, these two variables demonstrated a strikingly similar pattern. In terms of mean primary education attainment, despite occasional fluctuations, the values stabilized at a certain level post-2007.
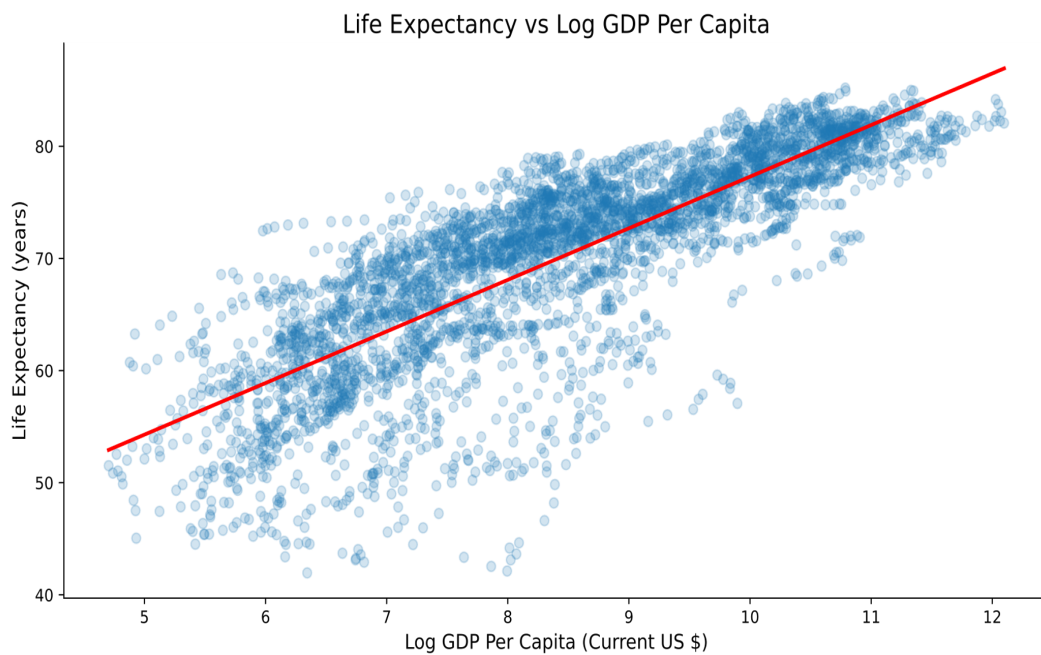
Figure 1
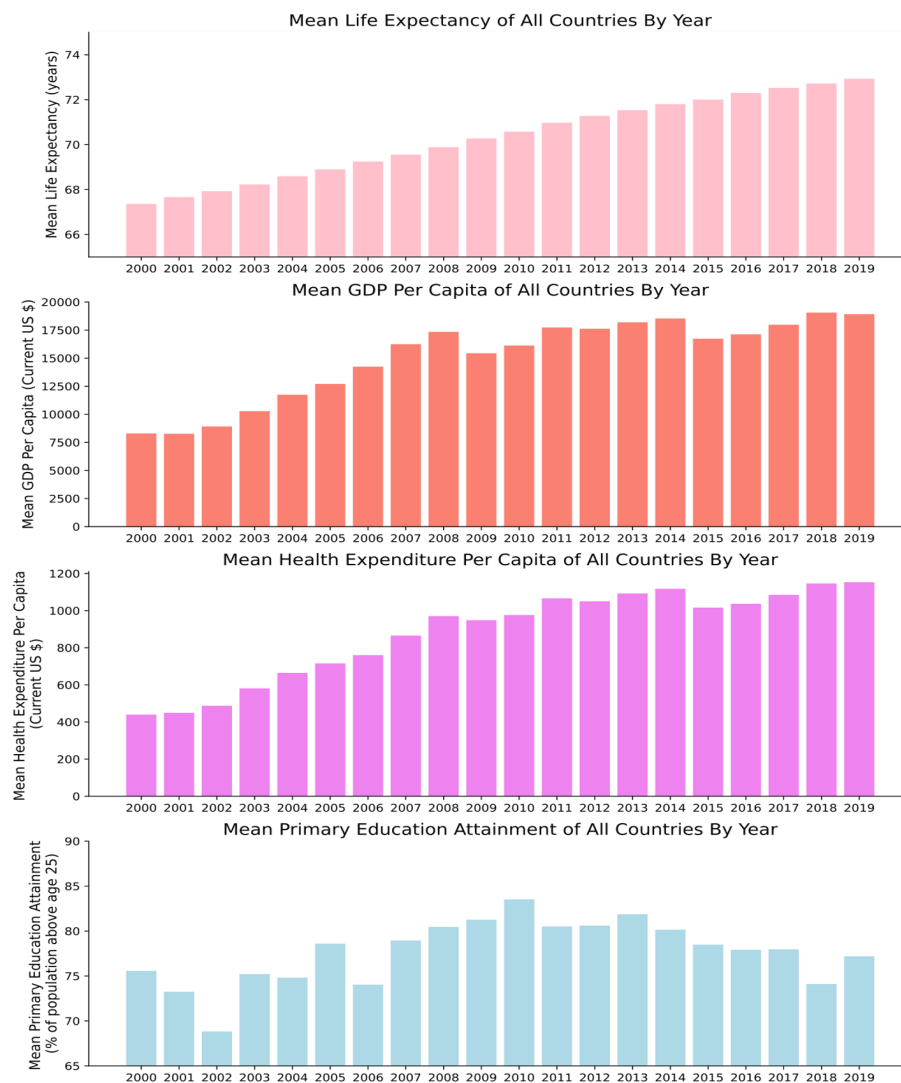
2. Log GDP Per Capita and Life Expectancy

Next, we compared the log of GDP per capita with life expectancy. As shown in Figure 2, There is a positive correlation between these two variables, indicating that the improvement in economic levels leads to an extension of life expectancy. In order to delve deeper into how the growth of per capita GDP affects life expectancy, we chose variables that affect life expectancy and are affected by GDP per capita as well, such as log health expenditure per capita and education attainment. In the following figure, we show the relationship between these factors and life expectancy.

Figure 2

3.(a) Log GDP Per Capita and Education Attainment

Firstly, we analyzed the relationship between education attainment and GDP per capita. As illustrated in Figure 3, with the growth of GDP per capita, education attainment is also on the rise. Education can be considered an investment in human capital. With the increase in GDP per capita, countries are likely to lean towards investing in their education systems to enhance the overall education level of the population, adapting to the demands of a more complex and high-skilled labor market.

Figure 3

Primary Education Attainment vs Log GDP Per Capita
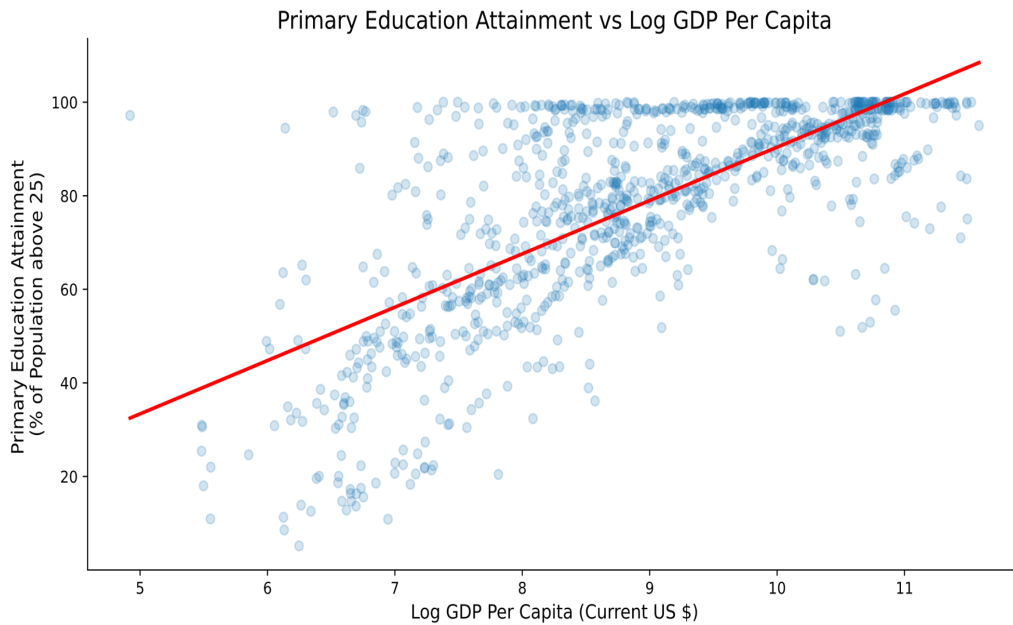
3.(b) Education Attainment and Life Expectancy

To further investigate the relationship between GDP per capita and life expectancy, we will conduct a detailed analysis of the connection between education attainment and life expectancy. In Figure 4, it can be observed that as education attainment increases, life expectancy also rises. This may reflect the positive impact of education in promoting health awareness, improving the quality of life, and enhancing the overall health levels of society.

Figure 4

Life Expectancy vs Primary Education Attainment
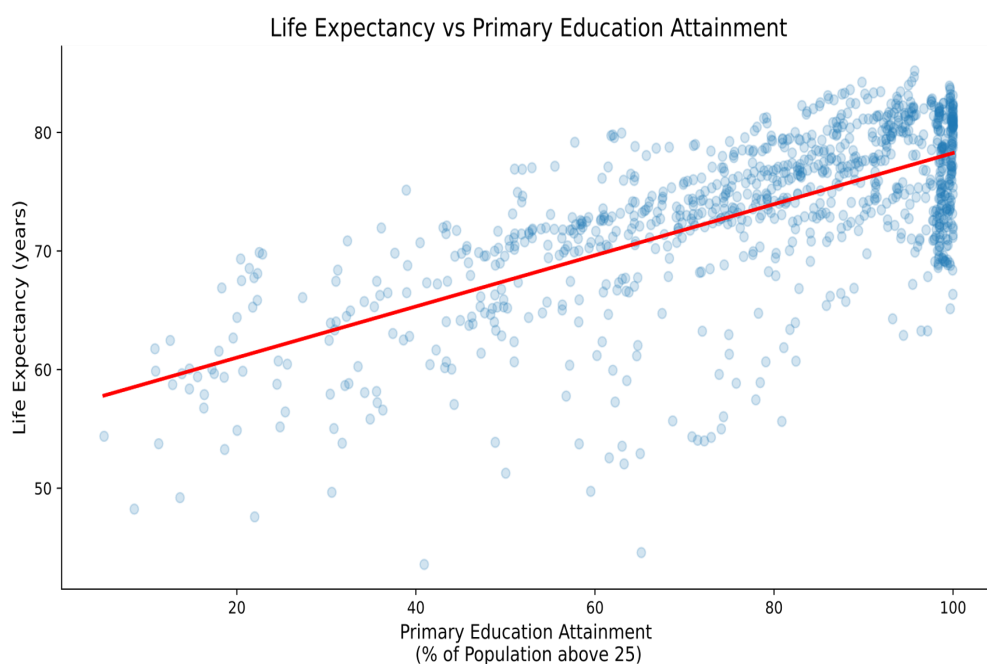
4.(a) Log GDP Per Capita and Health Expenditure Per Capita

On the other hand, we proceed to examine the relationship between GDP per capita and health expenditure per capita. As seen in Figure 5, there is a positive correlation between the two variables, and we posit that with the rise in income levels, there may be a change in people's demands for health and medical services. The increase in income levels might lead individuals to pay more attention to their physical well-being, and the improved financial capacity may empower them to afford the higher costs associated with healthcare, thereby driving the growth of healthcare expenditure.

Figure 5

Health Expenditure Per Capita vs GDP Per Capita
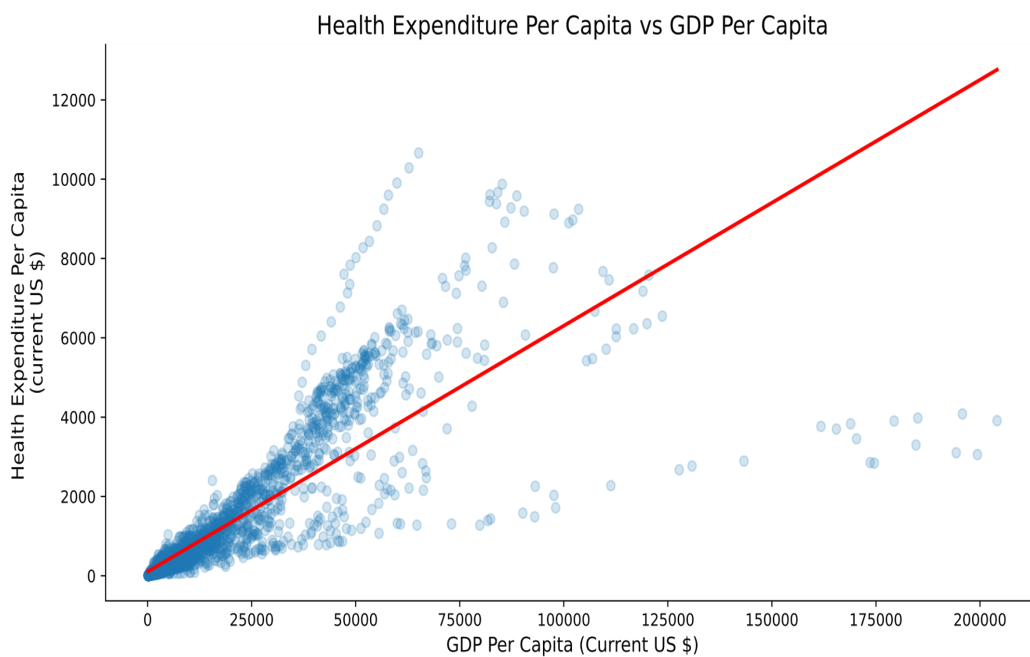
4.(b) Log Health Expenditure Per Capita and Life Expectancy

Furthermore, we delve into the impact of the log health expenditure per capita on life expectancy. As depicted in Figure 6, there is a positive correlation between the log of health expenditure per capita and life expectancy. With the increase in health expenditure per capita, life expectancy also extends. This suggests that investments in healthcare conditions may have a positive influence on life expectancy.

Figure 6

Life Expectancy vs Log Health Expenditure Per Capita
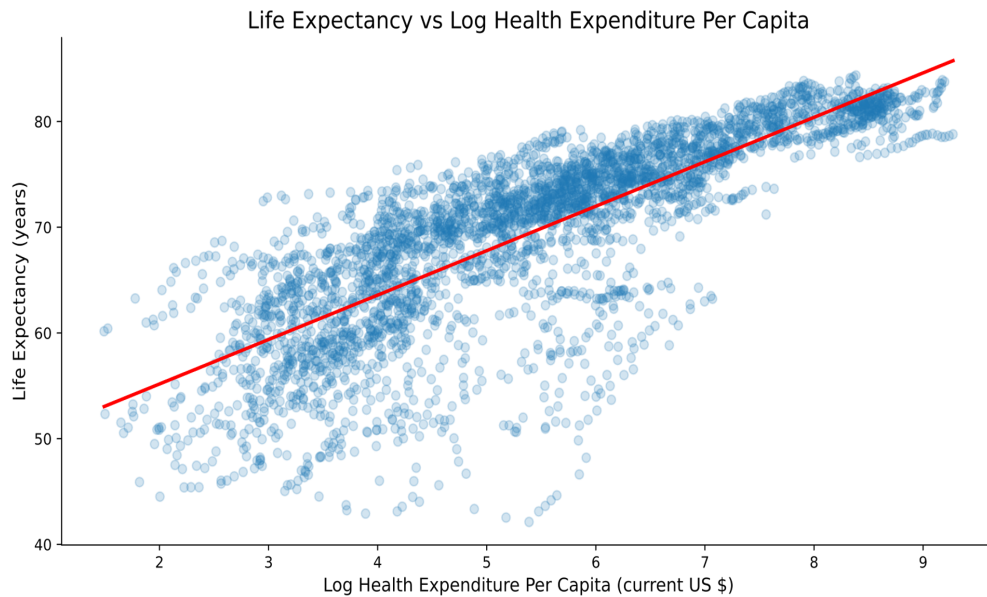
Regression Models

In order to better demonstrate and validate the trends in the image above, we ran three regression models for these four variables, and here are the results.

Life Expectancy vs Log of GDP Per Capita (Table 1)

In order to be able to quantify the relationship between life expectancy and the logarithm of GDP per capita, a regression model was done. According to the regression study, Life Expectancy and Log GDP Per Capita have a significant correlation (Adj. R-squared: 0.667). The positive coefficient for Log GDP Per Capita (4.6011) indicates a positive correlation with life expectancy, meaning that life expectancy tends to climb along with the logarithm of GDP per capita. The model's overall significance is highlighted by the F-statistic of 8062.0 and p-value of 0.00. As p-value is obtained less than 0.05, we can conclude a statistically significant relationship. The results were, as we would expect, positively correlated.

Life Expectancy vs Primary Education Attainment (Table 2)

There is a significant association between primary education attainment and life expectancy, according to the regression analysis. The model shows that elementary education attainment accounts for about 45.4% of the difference in life expectancy. The primary education attainment percent coefficient is 0.2156, meaning that a percentage increase in primary school attainment is predicted to result in a 0.2156-year increase in life expectancy. The dependability of the findings is emphasized by the low p-value ($p < 0.0001$), which supports the statistical significance. This implies that primary education plays a critical role in influencing and raising life expectancy, which in turn improves population health as a whole.

Life Expectancy vs Primary Education Attainment and Log Health Expenditure Per Capita (Table 3)

Primary education attainment and log health expenditure per capita account for approximately 68.4% of the variation in life expectancy. With p-values less than 0.0001, the two predictors—Primary Education Attainment Percent and Log Health Expenditure Per Capita—show statistically significant coefficients that highlight their impact on life expectancy. An increase in education attainment has a positive impact on life expectancy, as seen by the Primary Education Attainment Percent positive coefficient. Meanwhile, a positive coefficient for both Log Health Expenditure Per Capita and Primary Education Attainment indicates that longer life expectancy is correlated with higher health spending and better education. Moreover, the coefficient of Log Health Expenditure Per Capita is significantly higher than the coefficient of Primary Education Attainment ($3.1250 > 0.0526$). This indicates that the impact of Log Health Expenditure Per Capita on life expectancy is greater compared to the impact of Primary Education Attainment on life expectancy. This emphasizes how important investments in healthcare and education are in improving the general health of the people.

Conclusions

From the study above, it is concluded that better economic conditions, in particular higher GDP, increase life expectancy. As a result of countries getting richer, investments in education and health rise, which in turn lead to higher life expectancy. We find that both the log health expenditure per capita and education attainment have a positive effect on life expectancy. The regression model further strengthens these observations by showing a statistically significant association. Combining education attainment and the log of health expenditures provides us with a model for life expectancy with health expenditure having a more significant role. Through this study, governments should encourage sustained investment in education and health care to promote a healthier and longer life for the global population.

Appendix A: Tables

Table 1: Life Expectancy vs Log of GDP Per Capita

```
                        OLS Regression Results
==============================================================================
Dep. Variable:        Life_Expectancy   R-squared:                    0.667
Model:                            OLS   Adj. R-squared:               0.667
Method:                 Least Squares   F-statistic:                  8062.
Date:                Sat, 16 Dec 2023   Prob (F-statistic):            0.00
Time:                        02:57:14   Log-Likelihood:             -12299.
No. Observations:                4029   AIC:                      2.460e+04
Df Residuals:                    4027   BIC:                      2.462e+04
                   Df Model:                            1
                   Covariance Type:              nonrobust
==============================================================================
                        ===
                       coef    std err          t      P>|t|      [0.025
                       0.975]
------------------------------------------------------------------------------
                       ---
    Intercept        31.2799      0.442     70.760      0.000      30.413
                     32.147
    Log_GDP_Per_Capita 4.6011      0.051     89.791      0.000       4.501
                     4.702
==============================================================================
Omnibus:                      819.937   Durbin-Watson:                 2.112
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           1830.061
Skew:                          -1.158   Prob(JB):                       0.00
Kurtosis:                       5.353   Cond. No.                       47.8
==============================================================================

                                Notes:
 [1] Standard Errors assume that the covariance matrix of the errors is correctly
                               specified.
```

Table 2: Life Expectancy vs Primary Education Attainment

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          Life_Expectancy   R-squared:                       0.454
Model:                              OLS   Adj. R-squared:                  0.453
Method:                   Least Squares   F-statistic:                     819.9
Date:                  Sat, 16 Dec 2023   Prob (F-statistic):           8.72e-132
Time:                          02:57:30   Log-Likelihood:                 -3032.4
No. Observations:                   990   AIC:                             6069.
Df Residuals:                       988   BIC:                             6079.
                       Df Model:                              1
                       Covariance Type:                nonrobust
==============================================================================
                       =====================
                                               coef    std err          t      P>|t|
                            [0.025      0.975]
------------------------------------------------------------------------------
                            --------------------
Intercept                                    56.6948      0.610     92.921      0.000
                            55.498      57.892
Primary_Education_Attainment_Percent          0.2156      0.008     28.634      0.000
                            0.201       0.230
==============================================================================
Omnibus:                        207.694   Durbin-Watson:                   2.076
Prob(Omnibus):                    0.000   Jarque-Bera (JB):              411.598
Skew:                            -1.209   Prob(JB):                      4.19e-90
Kurtosis:                         5.033   Cond. No.                        300.
==============================================================================

                                   Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
                                  specified.
```

Table 3: Life Expectancy vs Primary Education Attainment and Log Health Expenditure Per Capita

```
                              OLS Regression Results
==============================================================================
Dep. Variable:          Life_Expectancy   R-squared:                       0.684
Model:                              OLS   Adj. R-squared:                  0.683
Method:                   Least Squares   F-statistic:                     1024.
Date:                Sat, 16 Dec 2023    Prob (F-statistic):           2.14e-237
Time:                        02:57:41    Log-Likelihood:                 -2640.1
No. Observations:                 949    AIC:                             5286.
Df Residuals:                     946    BIC:                             5301.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                                    coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
Intercept                         50.3094      0.528     95.293      0.000
49.273      51.345
Primary_Education_Attainment_Percent  0.0526    0.008      6.199      0.000
0.036       0.069
Log_Health_Expend_Per_Capita       3.1250      0.121     25.760      0.000
2.887       3.363
==============================================================================
Omnibus:                          394.746   Durbin-Watson:                 2.023
Prob(Omnibus):                      0.000   Jarque-Bera (JB):           1953.440
Skew:                              -1.884   Prob(JB):                       0.00
Kurtosis:                           8.933   Cond. No.                       338.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```