

Final Report

due November 16, 2021 by 11:59 PM

The Gre8est Team: Roni Ochakovski, Matthew Wang, Judy Zhong

11/16/2021

Background

Numerous studies have analyzed the correlation between health spending and health outcomes, consistently finding a positive relationship between the two [1]. Other studies have investigated how GDP and educational attainment are associated with health outcomes and found that both are positive predictors of health [2]. Our study looks to build on past findings and conduct a multivariate analysis to better inform global leaders on focus areas for improving global health.

Like similar studies, we have chosen to analyze life expectancy and under-5 mortality rate as measures of health [1]. Period life expectancy at birth is often used as a measure of the overall health of a population. It is derived from the probabilities of people of certain age groups dying given the mortality rates of those age groups over a specific time frame. The probabilities are then used in a survival function to project the average age of death of a newborn of that time period [3]. Meanwhile, under-5 mortality rate reflects the probability of a child born in the year in question dying before the age of 5. It is represented as the number of predicted deaths per 1,000 live births [4].

Research Question

This analysis aims to determine the significance of relationships between a set of World Development Indicators and health outcomes, measured by life expectancy and under-5 mortality rate.

Data

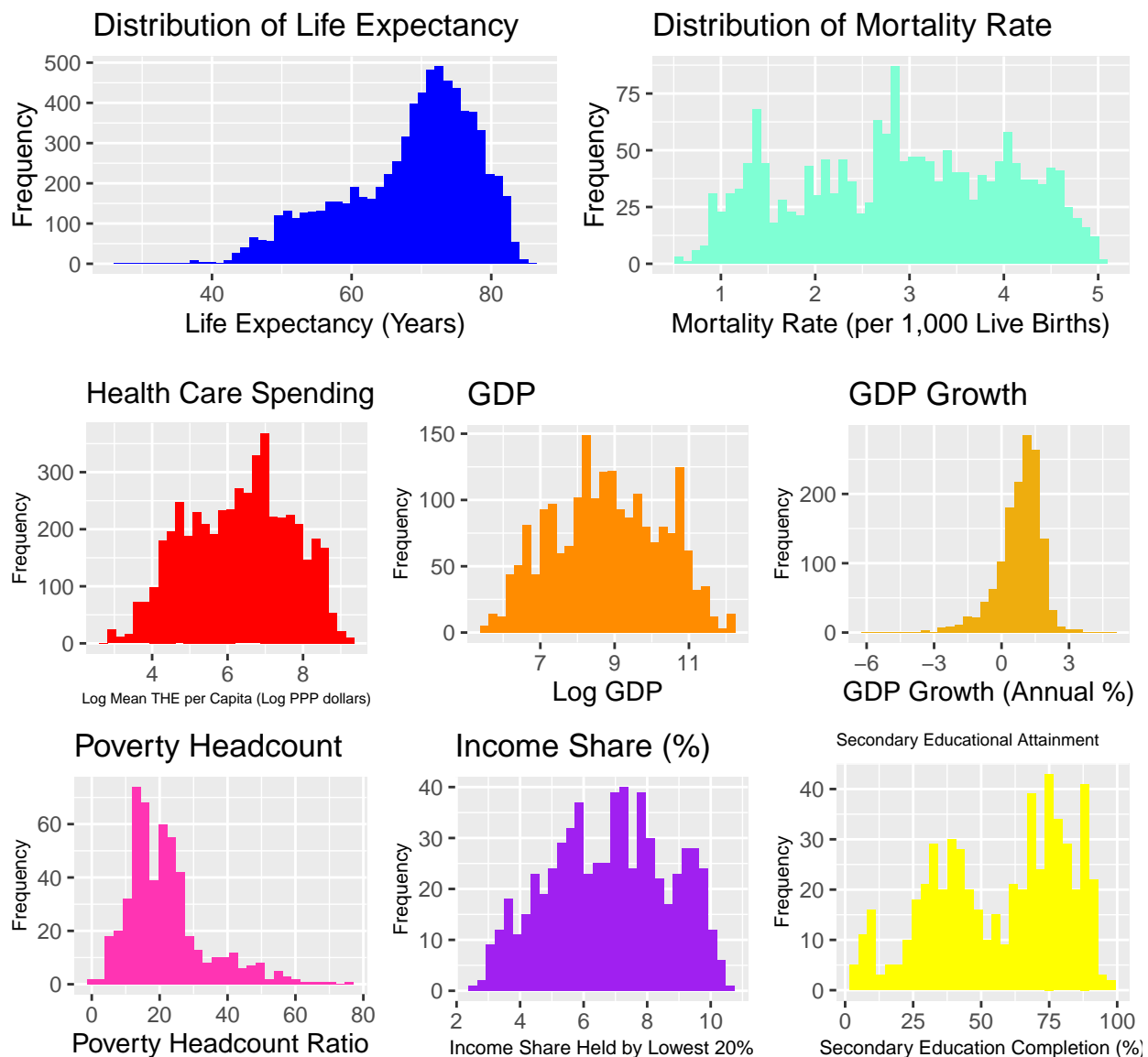
We are combining health spending data from the Global Health Data Exchange [5] and world economic/health-related data from the World Bank [6], ranging from 1990-2020. For these data sets, we will be analyzing data from the 204 countries and territories that are included in both databases. The health spending data was collected from a wide variety of sources that included program reports, budget data, national estimates, and National Health Accounts (NHAs). The variables that we are most concerned with are location ID, location name, year, and total health expenditure (THE) per capita in purchasing power parity (PPP) dollars (2020 USD). Purchasing power parity accounts for the differences in economic and standards of living between countries. The World Bank collects data through various sources, mostly through national statistical systems of member countries. The variables that we are most concerned with are country or area, total life expectancy at birth (years), male life expectancy at birth (years), female life expectancy at birth (years), GDP (current US dollars), GDP growth (annual %), income share held by lowest 20%, inflation/GDP deflation (annual %), mortality rate under 5 (per 1,000 live births), population total, population growth (annual %), and poverty headcount ratio at national poverty lines (% of population) [6].

Descriptive Stats

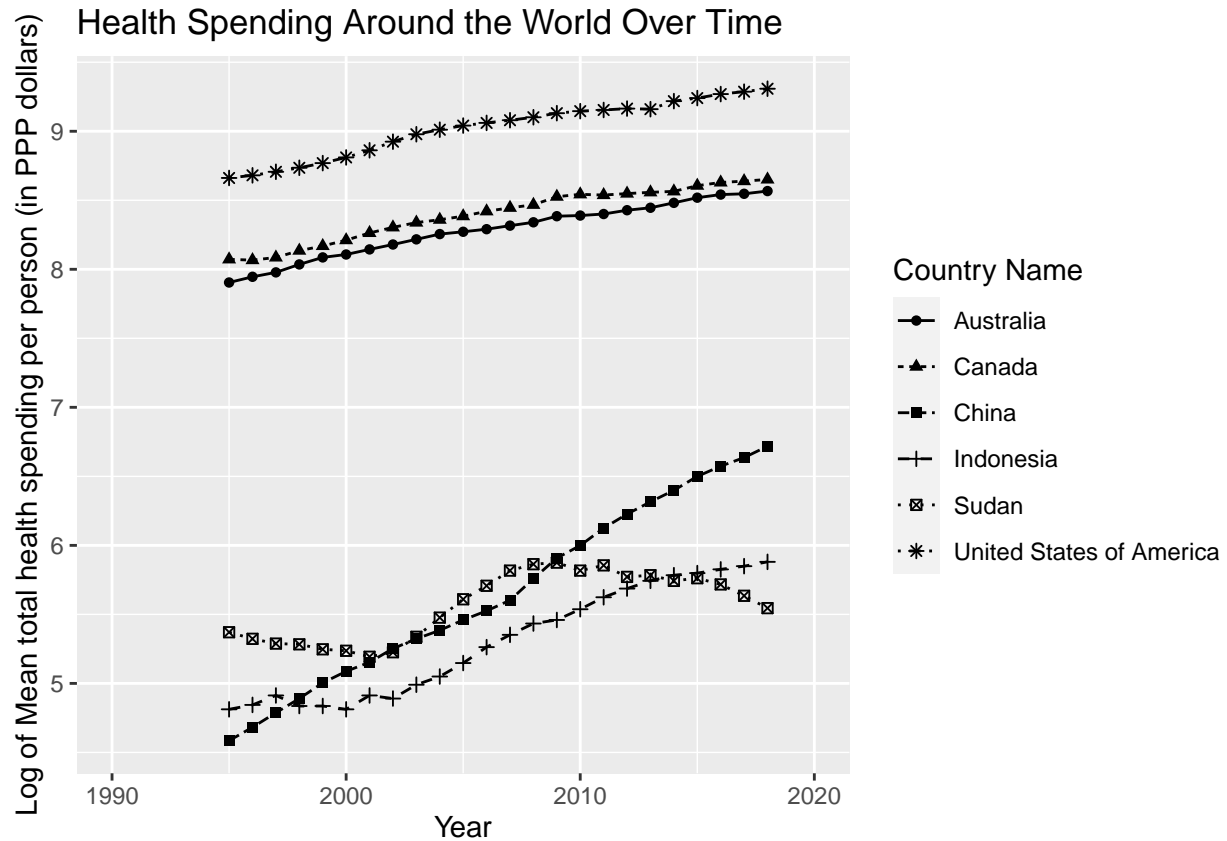
Following the data-cleaning, we ran some descriptive statistics to visualize the data. The first thing we noticed was that healthcare spending variation was extremely large across countries. To standardize the numbers across countries, we attempted to look at just the healthcare spending per person, but even that yielded huge variation across countries:

```
## the_per_cap_ppp_mean
## Min.   : 16
## 1st Qu.: 179
## Median : 603
## Mean   : 1207
## 3rd Qu.: 1534
## Max.   :11027
```

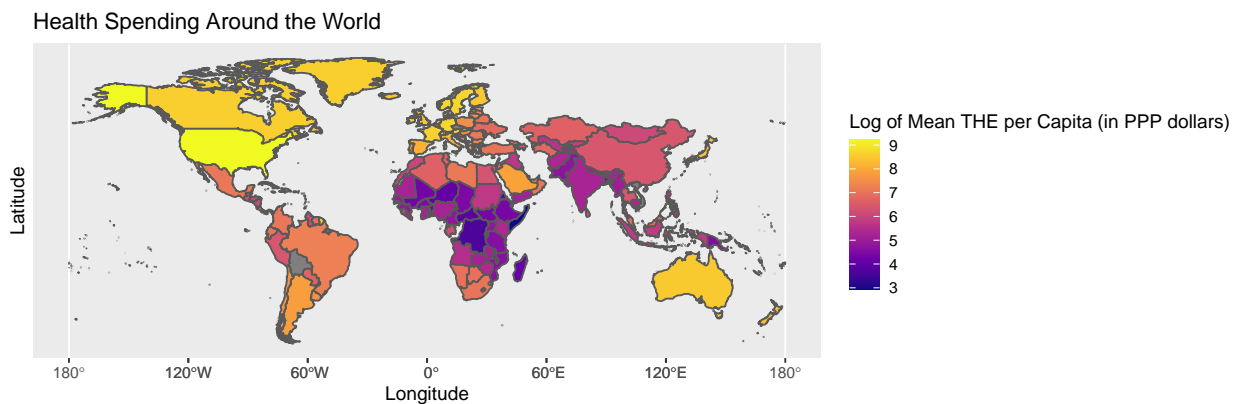
As you can see, the quartiles have enormous range between them. To standardize and reduce variation further, we took the log of the spending per person, and got an approximately normal distribution of healthcare spending across countries:



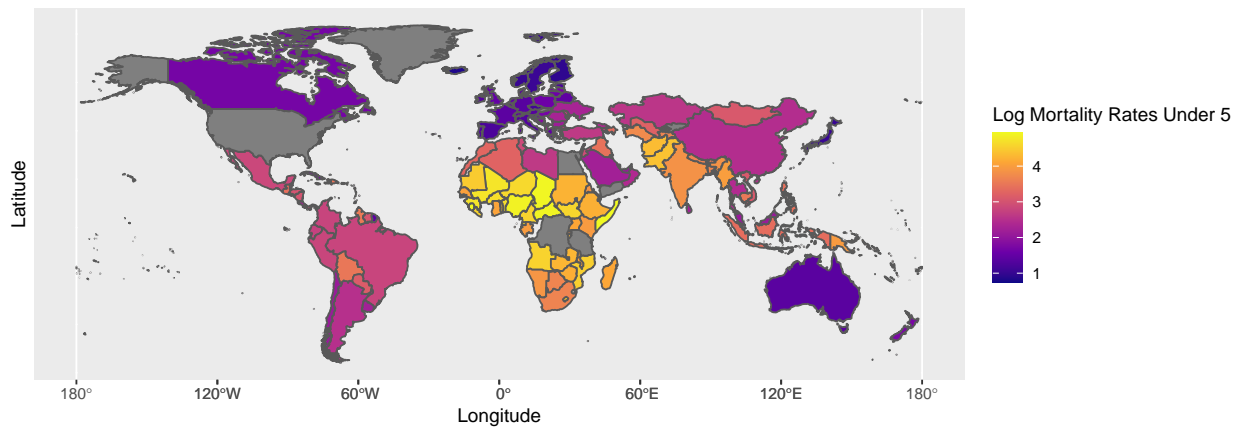
Then, we looked at healthcare spending over time in a few selected countries to see any trends in the data:



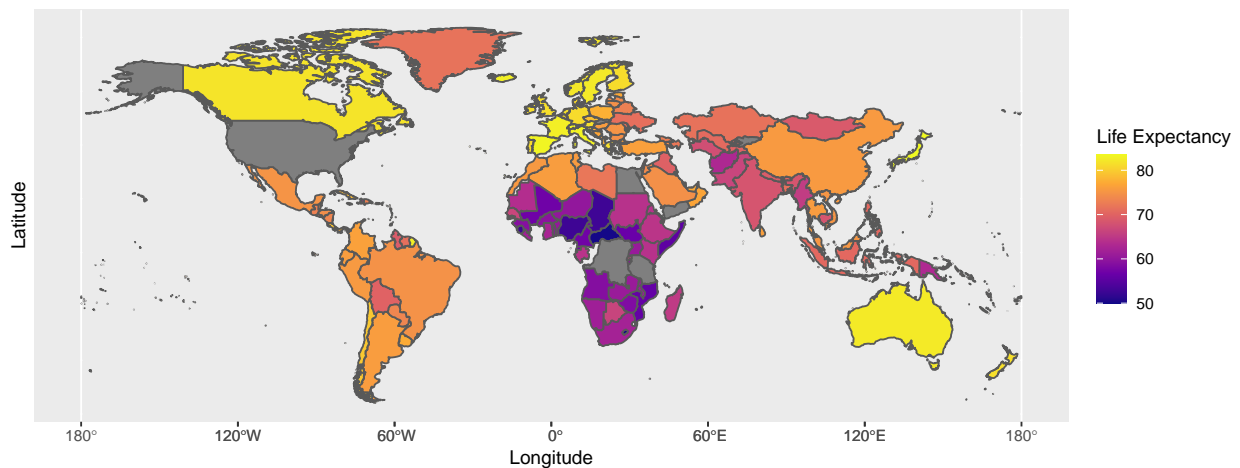
We found that healthcare spending generally increased over time in developed countries and was also significantly larger in Western than Eastern countries. This would be confirmed when we mapped health care spending. While mapping, we also mapped what would later be one of our response variables: life expectancy. We wanted to visualize life expectancy across countries to see any geographic patterns:



Mortality Rates Around the World



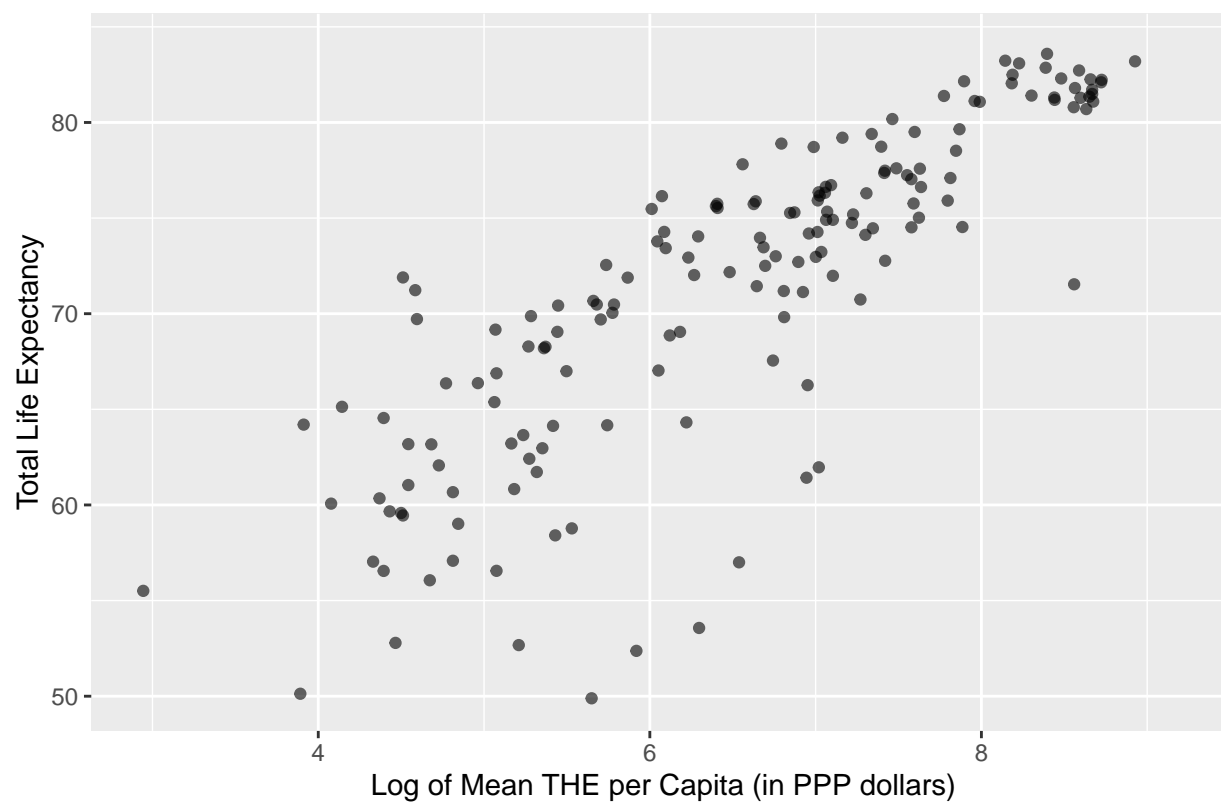
Life Expectancies Around the World

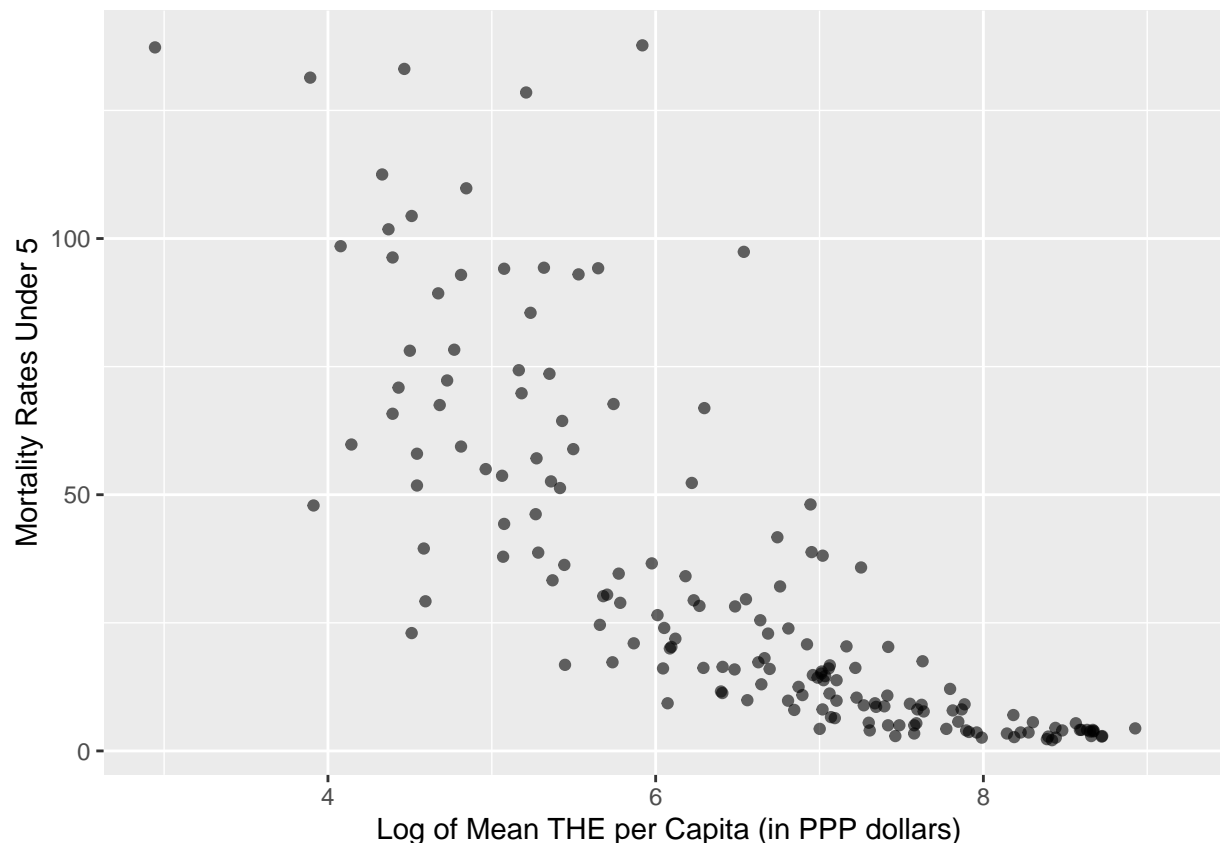


We saw the same pattern as in the line charts where Western countries had higher healthcare spending, higher life expectancy, and lower mortality rates. The same trend was generally true for developed compared to non-developed regions. We also began to notice the problem of missingness in our dataset where mortality rates and life expectancies for some countries (most notably the U.S.) were absent.

Finally, we wanted to visualize the relationship between healthcare spending and our health outcomes (being mortality rates under 5 and life expectancy). We used scatterplots to do so:

Mean THE per Capita versus Life Expectancy





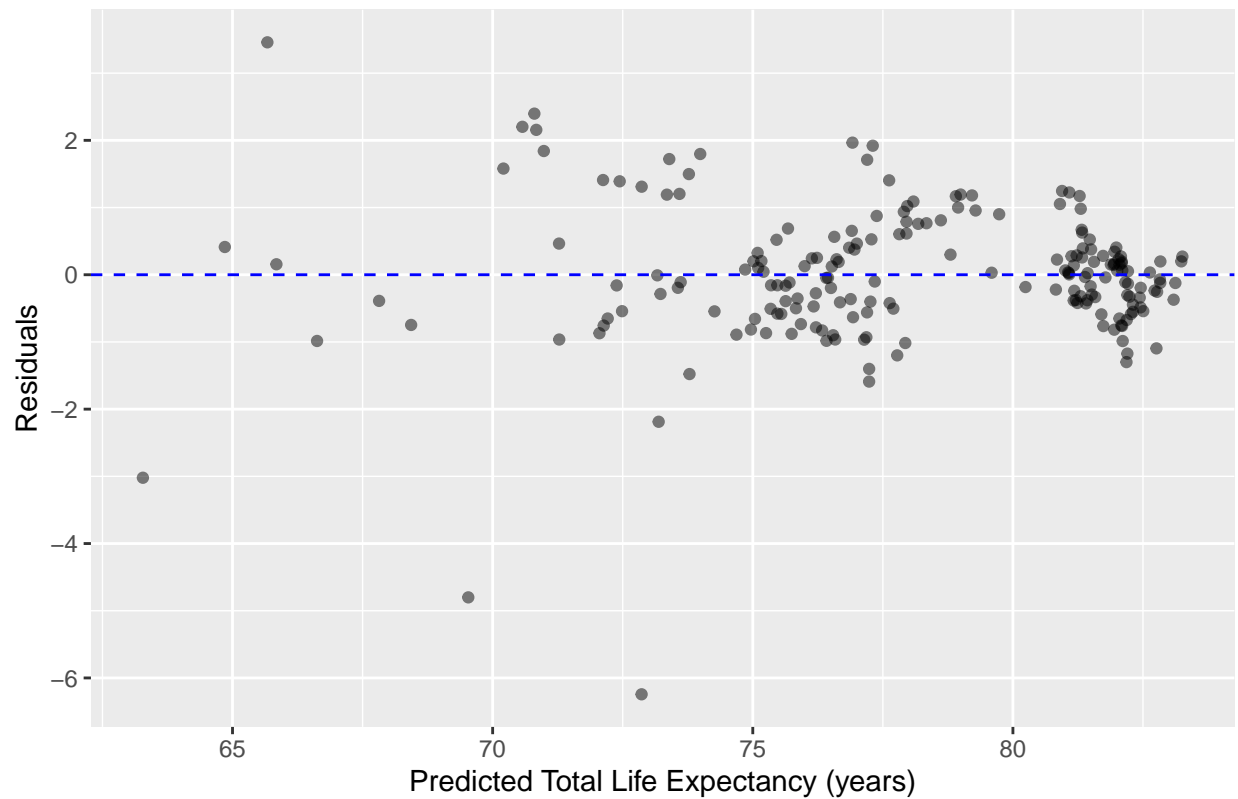
There is a clear positive linear relationship between healthcare spending and life expectancy but a more exponential-decay trend in healthcare spending and mortality rates under 5. Viewing these trends would come in helpful when modelling. Within the descriptive statistics portion of our project, we also explored healthcare spending by GDP, distributions of mortality rates, life expectancies, and fertility rates, male vs. female life expectancy distributions, and line graphs of health outcomes over time. However, the most significant graphs and visualizations are included above as these helped the most in understanding our data set and fitting the proper models. In addition to this, some outcomes such as fertility rates were excluded after data visualization due the missingness in that variable being too great to deal with.

Modelling

```
## # A tibble: 7 x 7
##   term                estimate std.error statistic  p.value  conf.low conf.high
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)       5.17e+ 1  1.70e+ 0   30.5  1.73e-75  4.84e+ 1  5.51e+ 1
## 2 log(the_per_cap_pp~ 4.32e+ 0  1.96e- 1   22.0  1.74e-54  3.94e+ 0  4.71e+ 0
## 3 `GDP (current US$)` -1.83e-13  1.71e-13   -1.07  2.85e- 1 -5.20e-13  1.54e-13
## 4 `GDP growth (annua~ -7.16e- 2  5.74e- 2   -1.25  2.14e- 1 -1.85e- 1  4.17e- 2
## 5 `Income share held~ -2.23e- 1  1.06e- 1   -2.10  3.75e- 2 -4.33e- 1 -1.31e- 2
## 6 `Poverty headcount~ -9.97e- 2  2.04e- 2   -4.89  2.11e- 6 -1.40e- 1 -5.95e- 2
## 7 `Educational attai~ -4.59e- 2  7.66e- 3   -5.98  1.05e- 8 -6.10e- 2 -3.07e- 2

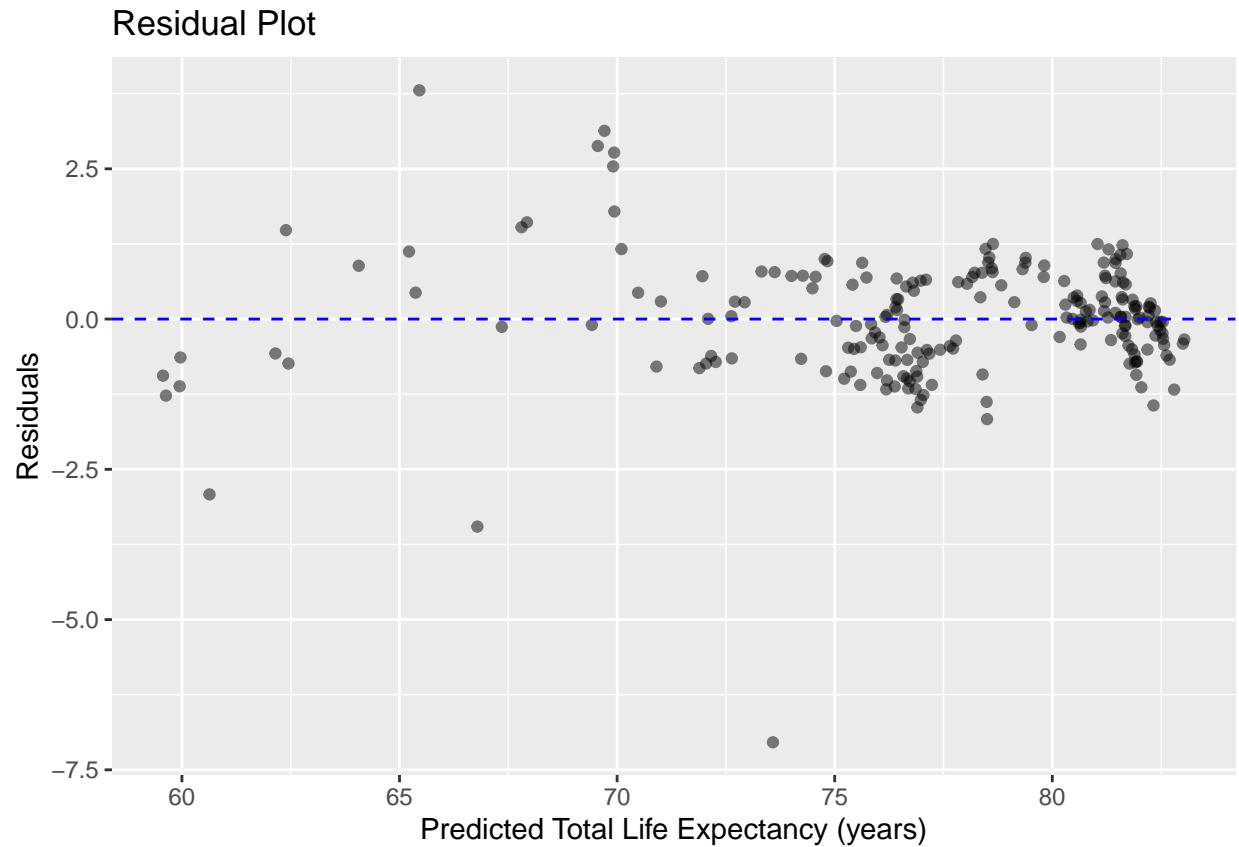
## [1] 0.8388933
## [1] 0.8338587
```

Residual Plot



```
## # A tibble: 7 x 7
##   term      estimate std.error statistic    p.value conf.low conf.high
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept) 38.8      6.16      6.30 0.00000000178 26.7     51.0
## 2 spending    7.06     1.54      4.60 0.00000745     4.03    10.1
## 3 spending2  -0.215    0.106     -2.04 0.0429        -0.424   -0.00690
## 4 gdp        -0.131    0.0922     -1.42 0.156         -0.313    0.0505
## 5 poverty    -0.0831   0.0162     -5.14 0.000000627   -0.115   -0.0512
## 6 education   0.191    0.0417      4.58 0.00000818     0.109    0.273
## 7 education2 -0.00203  0.000355    -5.70 0.0000000404  -0.00273 -0.00133

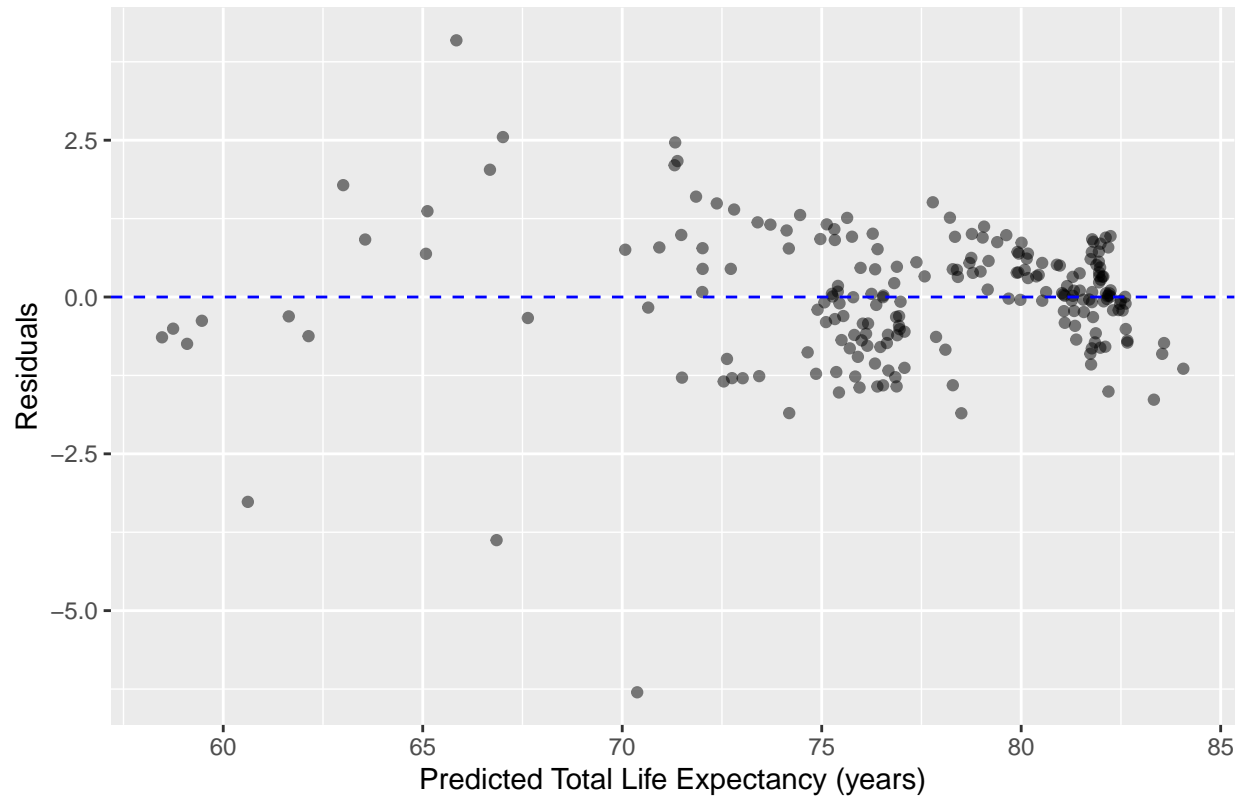
## [1] 0.8997453
## [1] 0.896811
```



```
## # A tibble: 13 x 7
##   term                estimate std.error statistic  p.value conf.low conf.high
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)       59.5      10.9        5.45 1.48e- 7  37.9     81.0
## 2 spending          -2.22      2.64       -0.843 4.00e- 1 -7.42     2.98
## 3 spending2          0.274     0.159        1.72 8.68e- 2 -0.0400  0.589
## 4 gdp                0.451     0.273        1.65 9.97e- 2 -0.0868  0.990
## 5 poverty           -0.182     0.224       -0.813 4.17e- 1 -0.623    0.259
## 6 education          0.466     0.0923        5.05 1.01e- 6  0.284    0.647
## 7 education2        -0.00248 0.000340      -7.29 7.24e-12 -0.00315 -0.00181
## 8 gdpg              -0.313     0.129       -2.42 1.63e- 2 -0.568   -0.0584
## 9 poverty:education -0.00495 0.000833      -5.95 1.19e- 8 -0.00660 -0.00331
##10 education:gdpg     0.00594 0.00226        2.62 9.38e- 3  0.00148  0.0104
##11 gdp:poverty        -0.0222 0.0108       -2.06 4.09e- 2 -0.0434 -0.000921
##12 gdp:education      -0.00598 0.00332       -1.80 7.32e- 2 -0.0125  0.000567
##13 spending:poverty   0.133     0.0267        4.99 1.30e- 6  0.0805   0.186

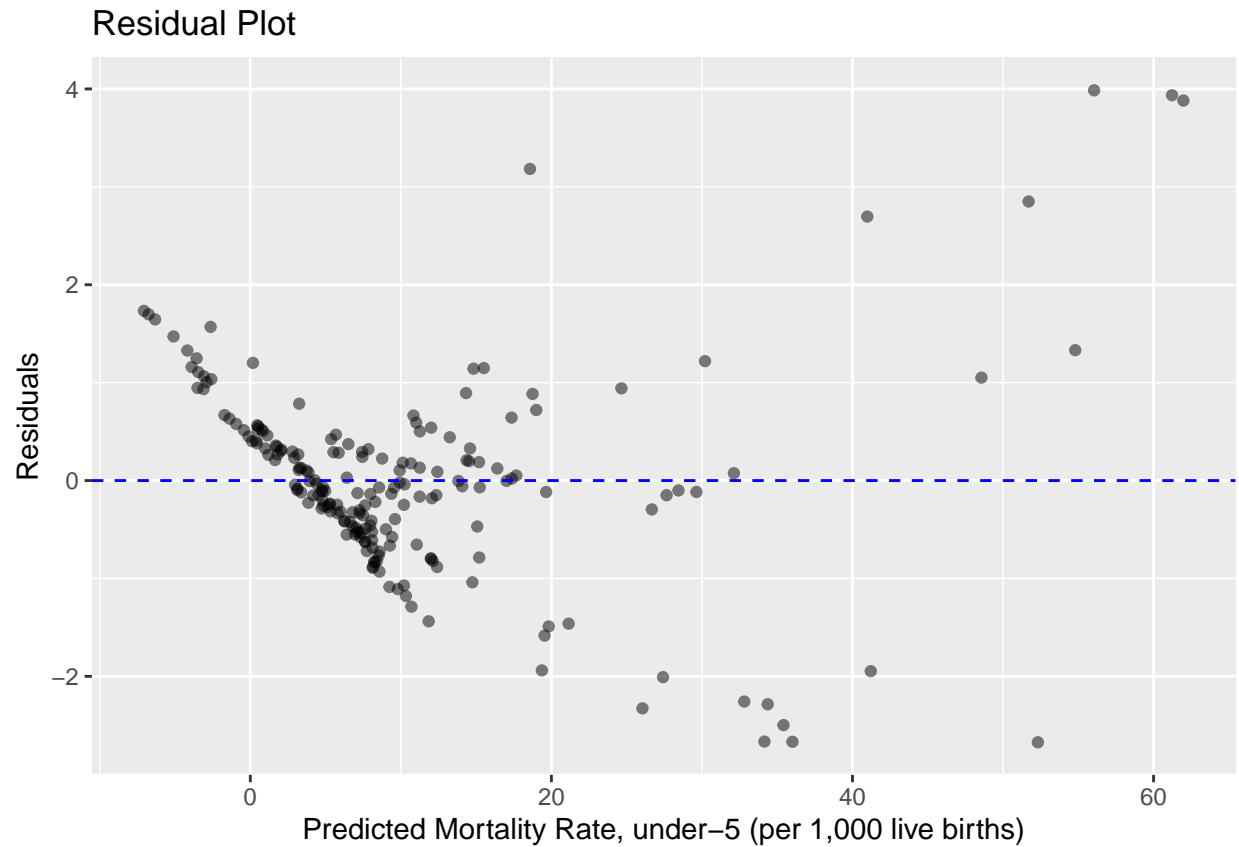
## [1] 0.9203884
## [1] 0.9155877
```


Residual Plot



```
## # A tibble: 7 x 7
##   term                estimate std.error statistic  p.value  conf.low conf.high
##   <chr>              <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)      5.50e+ 1  5.97e+ 0     9.23 5.10e-17  4.33e+ 1  6.68e+ 1
## 2 log(the_per_cap_pp~-9.84e+ 0  6.89e- 1    -14.3 5.78e-32 -1.12e+ 1 -8.48e+ 0
## 3 `GDP (current US$)` 2.39e-12  6.01e-13     3.98 9.68e- 5  1.21e-12  3.58e-12
## 4 `GDP growth (annua~ 2.47e- 1  2.02e- 1     1.23 2.22e- 1 -1.51e- 1  6.45e- 1
## 5 `Income share held~ 3.61e+ 0  3.74e- 1     9.66 2.99e-18  2.88e+ 0  4.35e+ 0
## 6 `Poverty headcount~ 5.68e- 1  7.16e- 2     7.92 1.81e-13  4.26e- 1  7.09e- 1
## 7 `Educational attai~-1.39e- 1  2.69e- 2    -5.14 6.66e- 7 -1.92e- 1 -8.54e- 2

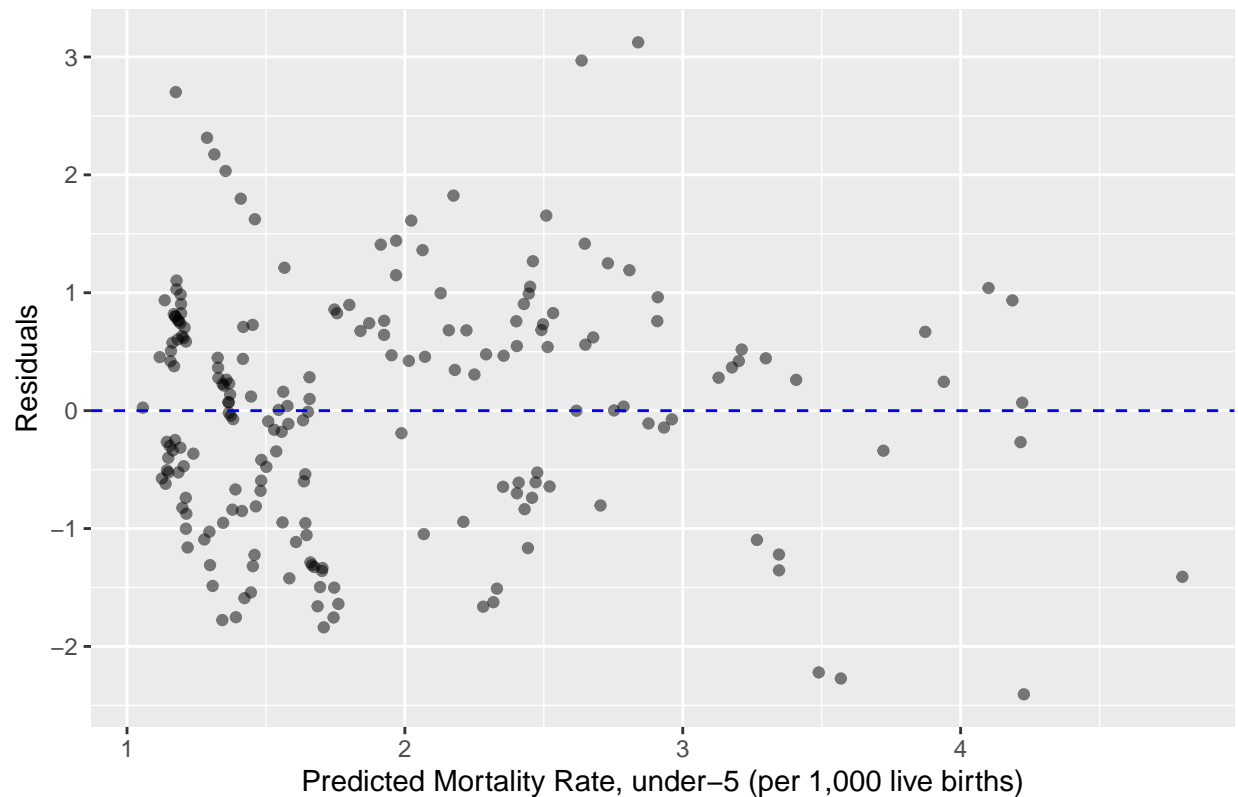
## [1] 0.7867786
## [1] 0.7801155
```



```
## # A tibble: 10 x 7
##   term      estimate std.error statistic  p.value  conf.low  conf.high
##   <chr>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept) 23.4      5.12      4.58 8.40e- 6 13.3     33.5
## 2 spending   -2.54     0.372    -6.81 1.27e-10 -3.27    -1.80
## 3 spending2   0.136    0.0260     5.22 4.59e- 7  0.0847   0.187
## 4 gdp        -0.545    0.379    -1.44 1.52e- 1 -1.29     0.202
## 5 gdp2        0.0115   0.00737     1.56 1.19e- 1 -0.00301  0.0261
## 6 income     -0.772    0.131    -5.91 1.55e- 8 -1.03    -0.515
## 7 income2     0.0459   0.00915     5.02 1.19e- 6  0.0279   0.0640
## 8 poverty2   -0.000212 0.0000674   -3.14 1.94e- 3 -0.000345 -0.0000789
## 9 education  -0.0267   0.00835    -3.20 1.61e- 3 -0.0432  -0.0103
## 10 education2 0.000191 0.0000709     2.69 7.78e- 3  0.0000509 0.000331

## [1] 0.8789936
## [1] 0.8732314
```

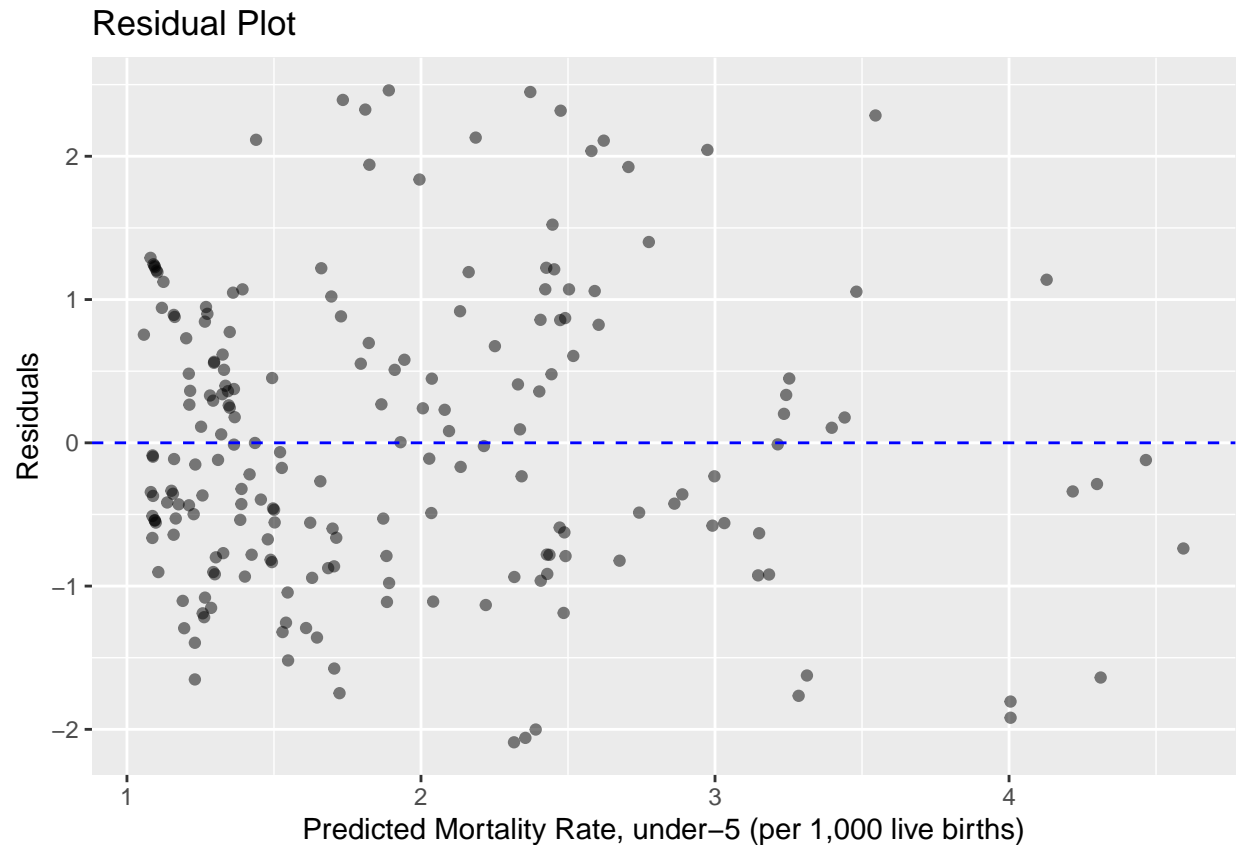
Residual Plot



```
## # A tibble: 19 x 7
```

##	term	estimate	std.error	statistic	p.value	conf.low	conf.high
##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
##	1 (Intercept)	2.42e+ 1	1.16e+ 1	2.08	0.0390	1.24e+ 0	4.72e+ 1
##	2 spending	-1.78e- 1	6.76e- 1	-0.263	0.793	-1.51e+ 0	1.16e+ 0
##	3 spending2	1.20e- 1	3.14e- 2	3.81	0.000191	5.77e- 2	1.82e- 1
##	4 gdp	-1.30e+ 0	9.73e- 1	-1.34	0.183	-3.22e+ 0	6.18e- 1
##	5 gdp2	4.14e- 2	2.09e- 2	1.98	0.0491	1.62e- 4	8.26e- 2
##	6 income	-1.61e- 1	5.23e- 1	-0.309	0.758	-1.19e+ 0	8.71e- 1
##	7 income2	6.93e- 2	1.29e- 2	5.36	0.000000256	4.38e- 2	9.49e- 2
##	8 poverty2	5.22e- 4	2.67e- 4	1.95	0.0522	-5.13e- 6	1.05e- 3
##	9 education	-1.13e- 1	3.16e- 2	-3.58	0.000443	-1.76e- 1	-5.08e- 2
##	10 education2	3.28e- 4	7.01e- 5	4.68	0.00000554	1.90e- 4	4.67e- 4
##	11 gdpg	-1.26e-12	4.78e-13	-2.64	0.00903	-2.20e-12	-3.19e-13
##	12 poverty	-7.96e- 2	3.14e- 2	-2.53	0.0122	-1.42e- 1	-1.76e- 2
##	13 income:education	-1.91e- 3	8.72e- 4	-2.19	0.0297	-3.63e- 3	-1.90e- 4
##	14 spending:gdp	-7.93e- 2	2.23e- 2	-3.55	0.000489	-1.23e- 1	-3.52e- 2
##	15 income:gdpg	1.79e-13	5.94e-14	3.01	0.00302	6.14e-14	2.96e-13
##	16 education:gdpg	-3.55e-15	3.41e-15	-1.04	0.299	-1.03e-14	3.17e-15
##	17 gdp:education	3.43e- 3	1.20e- 3	2.85	0.00487	1.06e- 3	5.81e- 3
##	18 gdp:income	-4.15e- 2	1.95e- 2	-2.13	0.0346	-8.00e- 2	-3.04e- 3
##	19 income:poverty	7.31e- 3	3.09e- 3	2.37	0.0189	1.22e- 3	1.34e- 2

```
## [1] 0.9123402
## [1] 0.9035743
```



References

- [1] Gallet CA, Doucouliagos H. The impact of healthcare spending on health outcomes: A meta-regression analysis. *Soc Sci Med.* 2017 Apr;179:9-17. doi: 10.1016/j.socscimed.2017.02.024. Epub 2017 Feb 20. PMID: 28237460.
- [2] Lutz, Wolfgang and Endale Kebede. "Education and Health: Redrawing the Preston Curve." *Population and Development Review*, vol. 44, no. 2, 2018, pp. 343-61, doi:<https://doi.org/10.1111/padr.12141>.
- [3] Luy, Marc et al. "Life Expectancy: Frequently Used, but Hardly Understood." *Gerontology*, vol. 66, no. 1, 2019, pp. 95-104, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7026938/>.
- [4] "Under-Five Mortality Rate (Probability of Dying by Age 5 Per 1000 Live Births)." Indicator Metadata Registry List. World Health Organization <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/7>.
- [5] [1] Global Burden of Disease Collaborative Network. *Global Health Spending 1995-2018*. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2021.