

2015 Multinational Healthcare Analysis

Mitch Strahlman

January 19th, 2021

The Data

The data utilized in this analysis was downloaded from a World Health Statistics Dataset. More specifically, the following datasets were downloaded and imported:

- lifeExpectancyAtBirth.csv
- infantMortalityRate.csv
- crudeSuicideRates.csv
- uhcCoverage.csv
- population10%SDG3.8.2.csv

These datasets covered the most recent health statistics in countries recognized by the WHO. I imported and cleaned these datasets for readability, while removing NA values and replacing the NA values with 0 in the *population10%SDG3.8.2.csv* dataset. I then merged these datasets into a singular, meta-dataset. Lastly, I wrangled this meta-data to represent the following variables per country in 2015 in order for consistent analyses: healthcare expenses, infant mortality rates, life expectancy rates, suicide rates, and a Universal Healthcare Index. The first five lines of the data can be seen below:

```
head(all_data)
```

```
## # A tibble: 6 x 6
##   Location      pop_high_costs infant_mortality life_expectancy suicide_rate   uhc
##   <chr>          <dbl>          <dbl>          <dbl>          <dbl> <dbl>
## 1 Belarus          7.78            3.04            73.8            26.3   74
## 2 Bolivia (P~      6.02            25.1            71.5            12.4   64
## 3 Bosnia and~      8.18            5.42            76.3            8.9    60
## 4 Ethiopia         4.91            43.1            66.8            7.2    39
## 5 Gambia           0.2             39.8            64.1            5.2    45
## 6 Indonesia        3.61            23.3            70.6            3.4    53
```

The distribution of the variables can be seen below:

```
summary(all_data)
```

```
##   Location      pop_high_costs   infant_mortality life_expectancy
## Length:162      Min.   : 0.0000   Min.   : 1.760   Min.   :47.67
## Class :character 1st Qu.: 0.0000   1st Qu.: 6.535   1st Qu.:65.09
## Mode  :character Median : 0.0000   Median :14.525   Median :73.03
##              Mean   : 0.9686   Mean   :23.541   Mean   :71.65
```

```

##          3rd Qu.: 0.0000  3rd Qu.:35.665  3rd Qu.:77.38
##          Max.    :21.7800  Max.    :94.170  Max.    :83.62
## suicide_rate      uhc
## Min.    : 0.700  Min.    :22.00
## 1st Qu.: 4.800  1st Qu.:49.00
## Median : 8.100  Median :66.50
## Mean    : 9.418  Mean    :62.63
## 3rd Qu.:12.475  3rd Qu.:75.00
## Max.    :34.900  Max.    :88.00

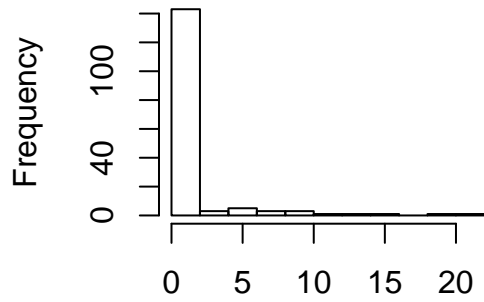
```

Variable Explanation

- pop_high_costs: Percent of country's population with healthcare costs > 10% of household expenses or income.
- infant_mortality: Probability of dying between birth and age 1 per 1000 live births for a single country.
- life_expectancy: A country's life expectancy at birth.
- suicide_rate: Suicides per 100k population for a single country.
- uhc: A country's Universal Healthcare Index is an indicator reported on a 0 to 100 scale that represents a country's ability to achieve universal health coverage including financial risk protection and access to quality healthcare services and essential, affordable medicines.

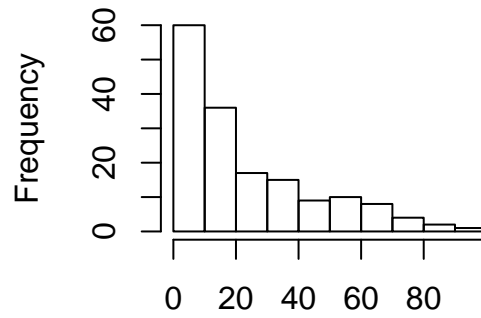
Histograms of Each Variable

HC Healthcare Distribution



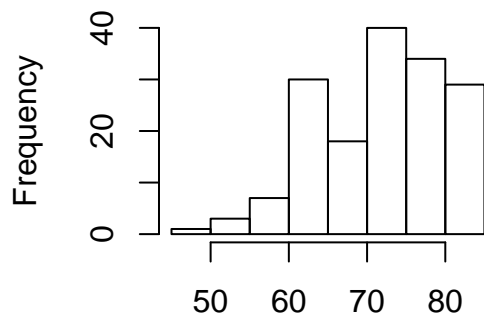
High Cost Healthcare Rate

Infant Mortality Distribution



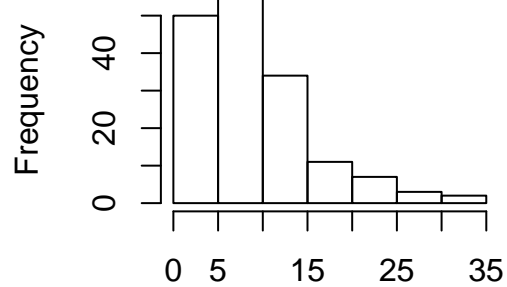
Infant Mortality Rate

Life Expectancy Distribution



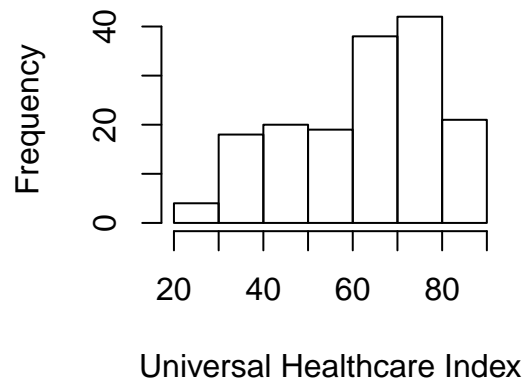
Life Expectancy at Birth

Suicide Rate Distribution



Suicides per 100K Population

UHC Index Distribution



Data Takeaways

- 162 countries were included in this meta-dataset. Some were removed due to missing data.
- Only 20 countries had non-zero values for population with healthcare costs > 10% of household expenses or income, which caused the variable to be positively skewed.
- The average number of suicides per 100k people across all countries included in the data is 9.4.
- The average infant mortality and life expectancy rates seem to be standard.
- Infant mortality and suicide rate are positively skewed, while life expectancy and Universal Healthcare Index are negatively skewed.

Top and Bottom 5 Countries for Each Variable

5 Countries with the Highest Percent of Households with High Healthcare Costs

```
## # A tibble: 5 x 2
##   Location          pop_high_costs
##   <chr>              <dbl>
## 1 Republic of Korea    21.8
## 2 Republic of Moldova  18.5
## 3 Myanmar             14.4
## 4 Poland              13.7
## 5 Montenegro          10.3
```

Top and Bottom 5 Countries with the Highest and Lowest Infant Mortality Rates

```
## # A tibble: 5 x 2
##   Location          infant_mortality
##   <chr>              <dbl>
## 1 Sierra Leone      94.2
## 2 Central African Republic 89.7
## 3 Somalia           82.8
## 4 Nigeria            79.5
## 5 Chad              76.1
```

```
## # A tibble: 5 x 2
##   Location infant_mortality
##   <chr>      <dbl>
## 1 Iceland      1.76
## 2 Japan        2.01
## 3 Slovenia     2.01
## 4 Finland      2.03
## 5 Singapore    2.18
```

Top and Bottom 5 Countries with Highest and Lowest Life Expectancy

```
## # A tibble: 5 x 2
##   Location    life_expectancy
##   <chr>      <dbl>
## 1 Japan      83.6
## 2 Singapore  82.8
## 3 Switzerland 82.6
## 4 Spain      82.3
## 5 Luxembourg 82.3
```

```
## # A tibble: 5 x 2
##   Location                life_expectancy
##   <chr>                  <dbl>
## 1 Lesotho                47.7
## 2 Central African Republic 50.5
## 3 Eswatini               54.0
## 4 Somalia                54.6
## 5 Mozambique             56.0
```

Top and Bottom 5 Countries with Highest and Lowest Suicide Rates

```
## # A tibble: 5 x 2
##   Location          suicide_rate
##   <chr>            <dbl>
## 1 Lithuania        34.9
## 2 Russian Federation 32.1
## 3 Guyana            29.1
## 4 Republic of Korea 28.3
## 5 Belarus           26.3
```

```
## # A tibble: 5 x 2
##   Location          suicide_rate
##   <chr>            <dbl>
## 1 Barbados          0.7
## 2 Antigua and Barbuda 0.8
## 3 Grenada           1.1
## 4 Bahamas           1.8
## 5 Syrian Arab Republic 1.9
```

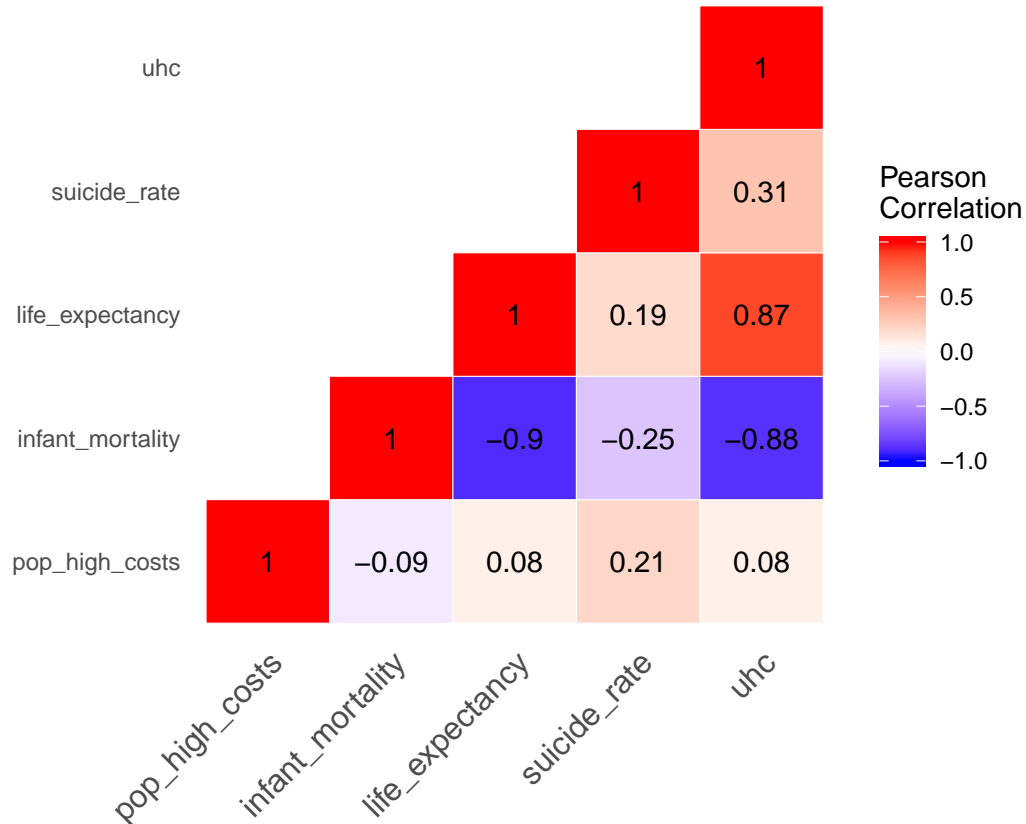
Top and Bottom 5 Countries with Highest and Lowest UHC Index

```
## # A tibble: 9 x 2
##   Location      uhc
##   <chr>      <dbl>
## 1 Canada      88
## 2 Australia   86
## 3 Norway      86
## 4 Republic of Korea 85
## 5 Iceland     85
## 6 Netherlands 85
## 7 New Zealand 85
## 8 Singapore    85
## 9 Sweden      85
```

```
## # A tibble: 5 x 2
##   Location      uhc
##   <chr>      <dbl>
## 1 Somalia      22
## 2 Madagascar   24
## 3 Chad         27
## 4 South Sudan  30
## 5 Central African Republic 32
```

Correlation Matrix

The below matrix delineates the Pearson's r correlation value between each of the continuous variables.



Discussion

We can see that there are high r values for the relationships between *life expectancy* and *infant mortality*, *Universal Healthcare Index* and *infant mortality*, and *Universal Healthcare Index* and *life expectancy*. This is to be expected, as better healthcare coverage will lead to lower infant mortality and higher life expectancy. There are also moderate relationships between *suicide rates* and *Universal Healthcare Index* and *suicide rates* and *infant mortality*. In the next section, we will examine the relationship on a deeper level between some of these variables through the construction of a couple of linear regression models.

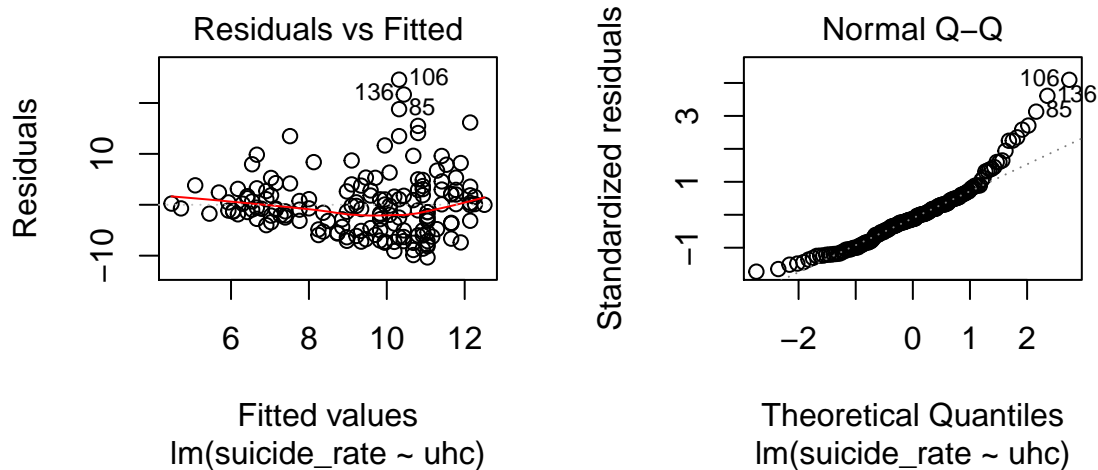
Linear Regression Analysis

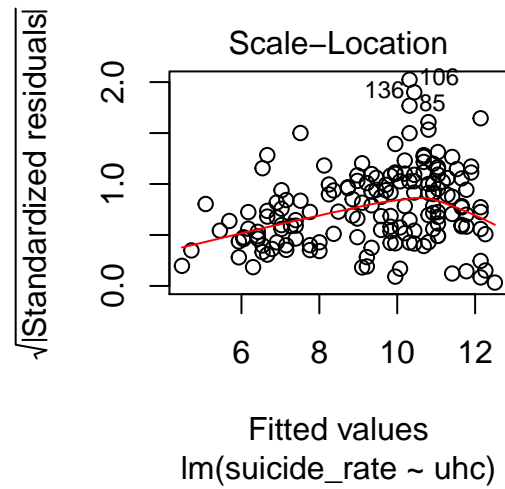
Assessing the Relationship Between Suicide Rate and UHC: Does UHC Index Predict Suicides?

We can see from the below regression model output that UHC Index is a significant predictor of suicide rate.

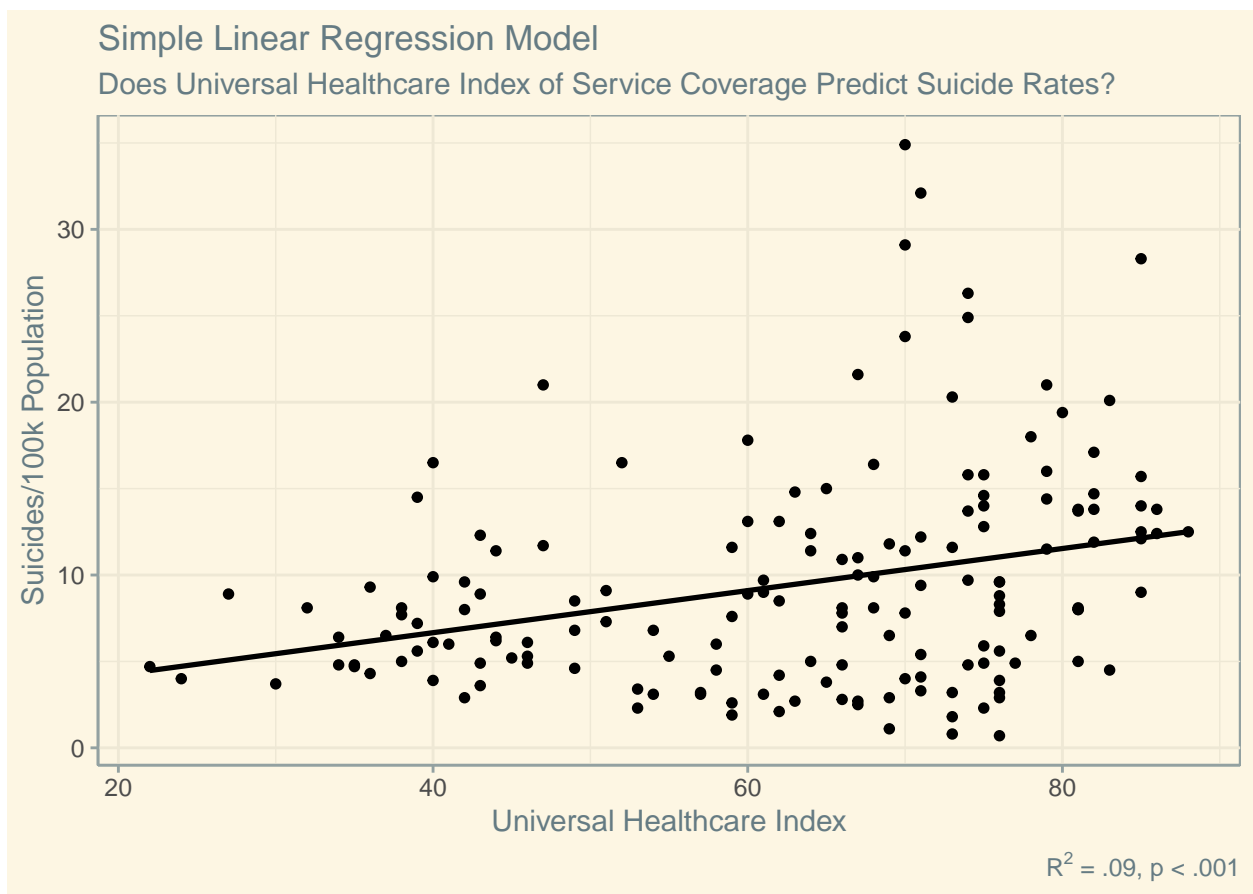
```
##
## Call:
## lm(formula = suicide_rate ~ uhc, data = all_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.3456  -3.9282  -0.7674   2.6761  24.5849
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.79356    1.91909   0.935   0.351
## uhc          0.12174    0.02969   4.100 6.57e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.026 on 160 degrees of freedom
## Multiple R-squared:  0.09506,    Adjusted R-squared:  0.0894
## F-statistic: 16.81 on 1 and 160 DF,  p-value: 6.567e-05
```

The below plots assess the assumptions of the regression model. The first plot assesses the linearity assumption, the second plot assesses the normality assumption, and the third plot assesses the homogeneity of variance assumption. We can see there might be an issue with homogeneity of variance, which will we address further.





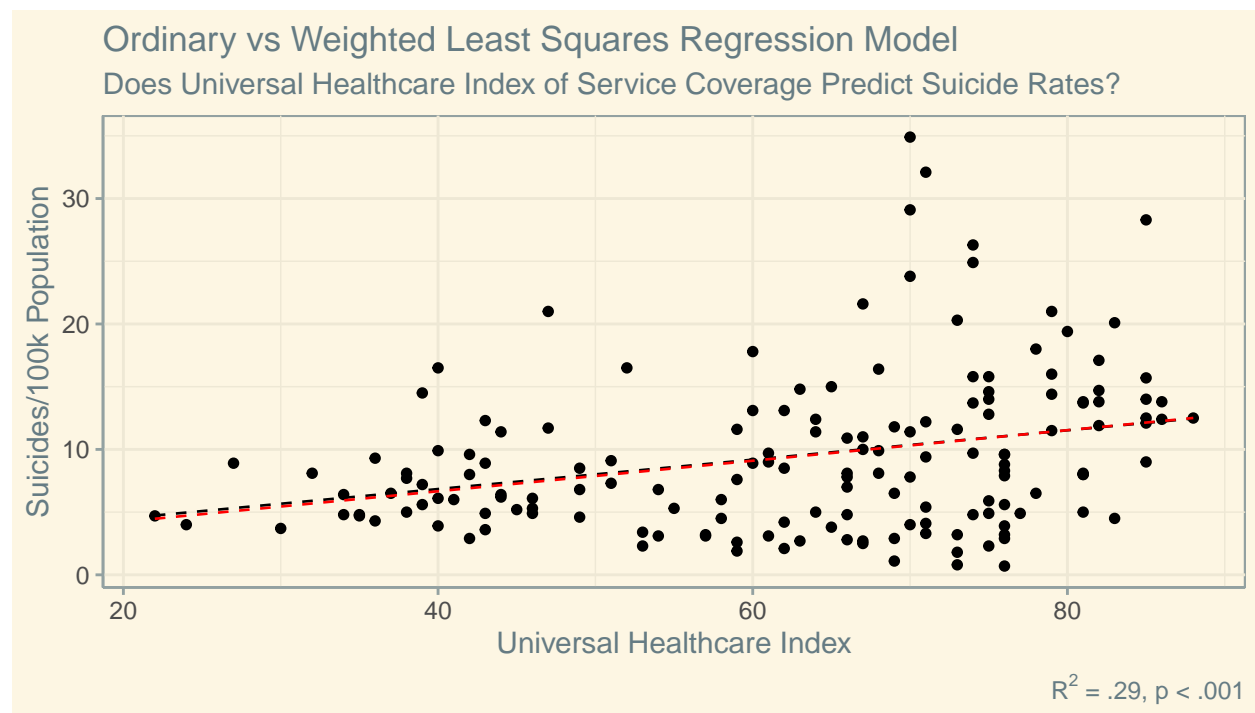
The below plot visualizes the regression model. We can see that, on average, as UHC Index increases, so does suicide rate. The R-squared value is small, but the relationship is significant.



Because there may be an issue with homogeneity of variance, a better approach might be to build a weighted least squares regression model. Weights were created for the residuals, and the output of the model can be seen below.

```
##
## Call:
## lm(formula = suicide_rate ~ uhc, data = all_data, weights = lm_weights)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30426 -0.13009 -0.03088  0.07113  0.62679
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.15580    0.46840   4.602 8.46e-06 ***
## uhc          0.11712    0.01433   8.172 8.69e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1783 on 160 degrees of freedom
## Multiple R-squared:  0.2945, Adjusted R-squared:  0.2901
## F-statistic: 66.79 on 1 and 160 DF,  p-value: 8.686e-14
```

Note that the predictor value is still significant, but the R-squared has increased to .29, reflecting a better fit. To visualize a comparison between the ordinary least squares regression model and the weighted least squares regression model, we can create a plot comparing the respective models. The black dotted line represents the ordinary least squares model, and the red dotted line represents the weighted least squares model. Visually, there does not seem to be a noticeable difference, but the R-squared values represent a difference in fit.

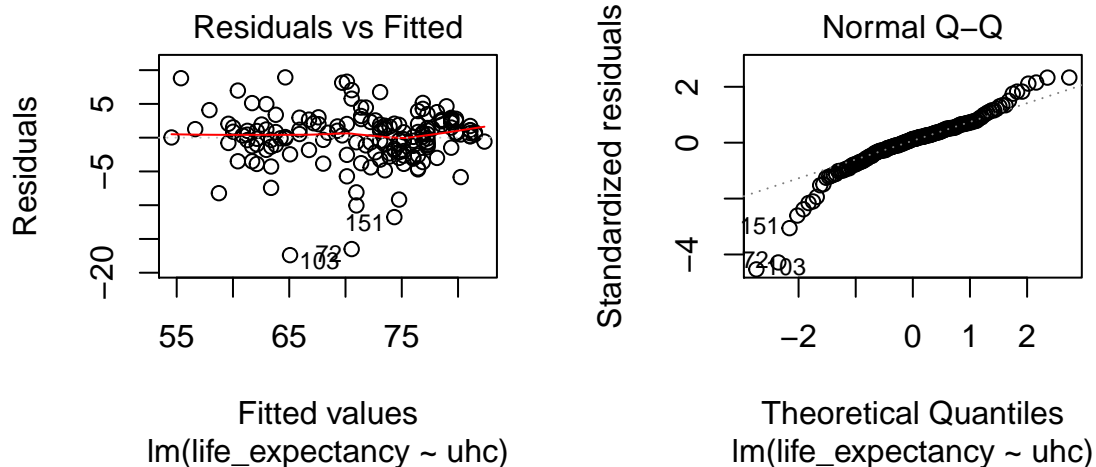


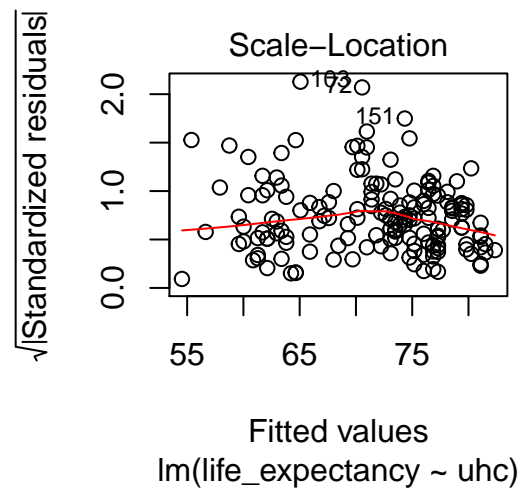
Assessing the Relationship Between UHC Index and Life Expectancy: Does UHC Index Predict Life Expectancy?

We can see from the below regression model output that UHC Index is a significant predictor of life expectancy.

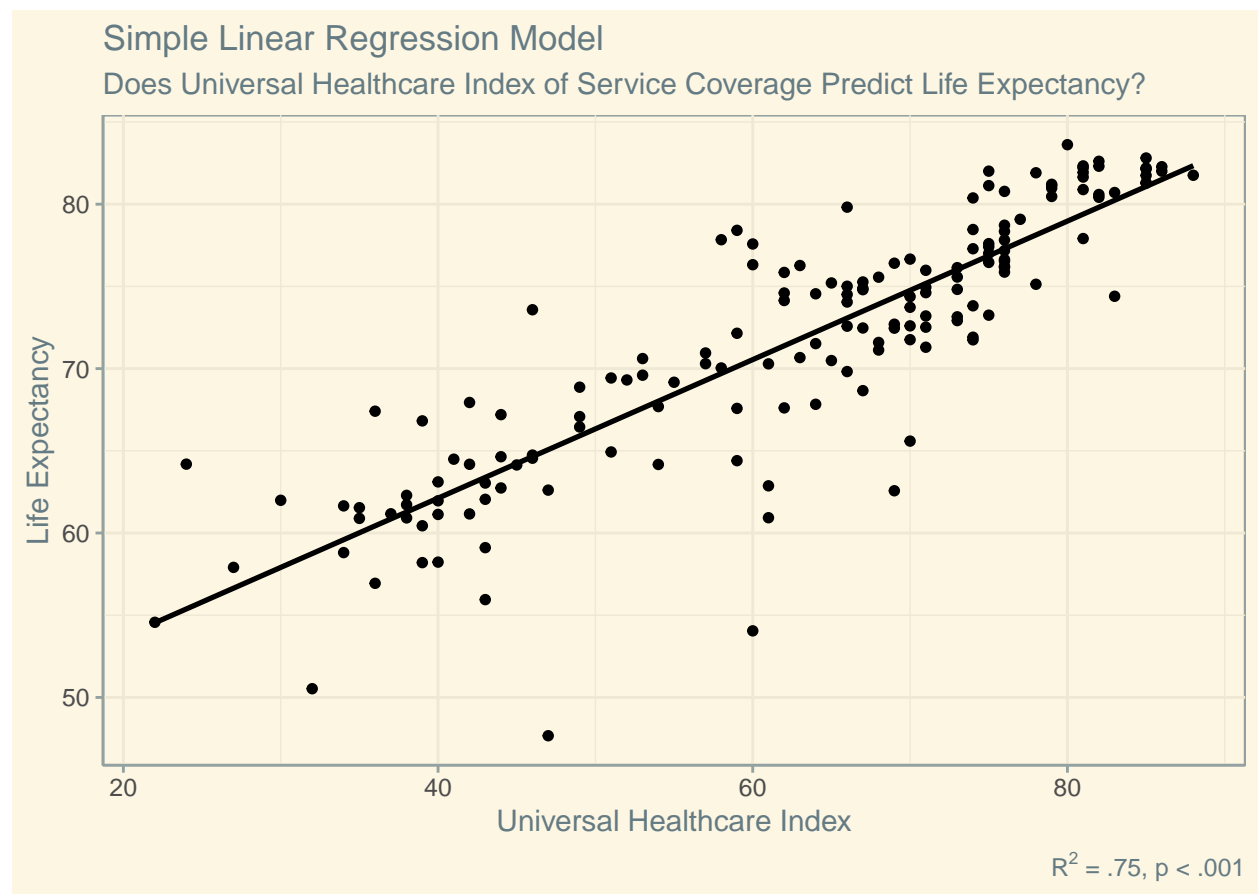
```
##
## Call:
## lm(formula = life_expectancy ~ uhc, data = all_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.3997  -1.4665   0.4924   2.0126   8.9316
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  45.26964    1.22994   36.81  <2e-16 ***
## uhc          0.42128     0.01903   22.14  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.862 on 160 degrees of freedom
## Multiple R-squared:  0.7538, Adjusted R-squared:  0.7523
## F-statistic:  490 on 1 and 160 DF, p-value: < 2.2e-16
```

The below plots assess the assumptions of the regression model. The first plot assesses the linearity assumption, the second plot assesses the normality assumption, and the third plot assesses the homogeneity of variance assumption.





The below plot visualizes the regression model. We can see that, on average, as UHC Index increases, so does life expectancy. The R-squared is large, and the relationship is significant.

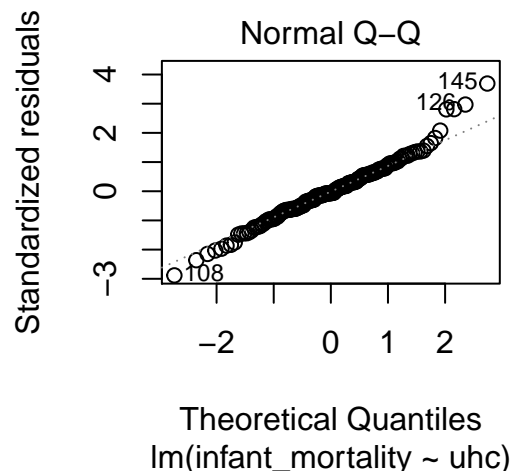
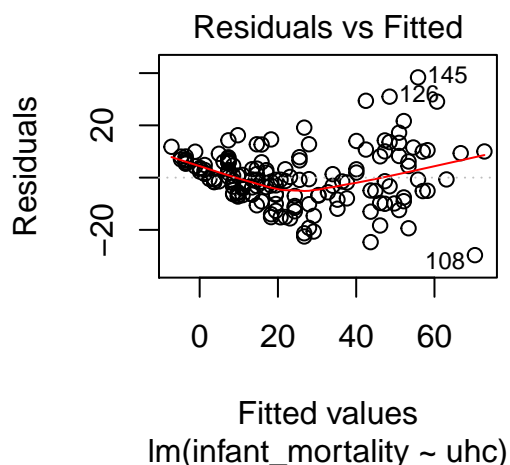


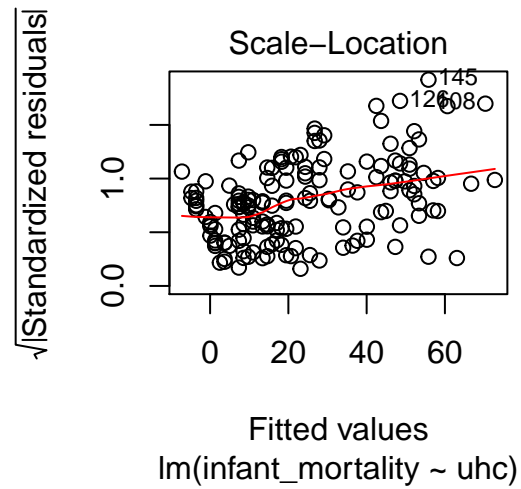
Assessing the Relationship Between UHC Index and Infant Mortality: Does UHC Index Predict Infant Mortality?

We can see from the below regression model output that UHC Index is a significant predictor of infant mortality.

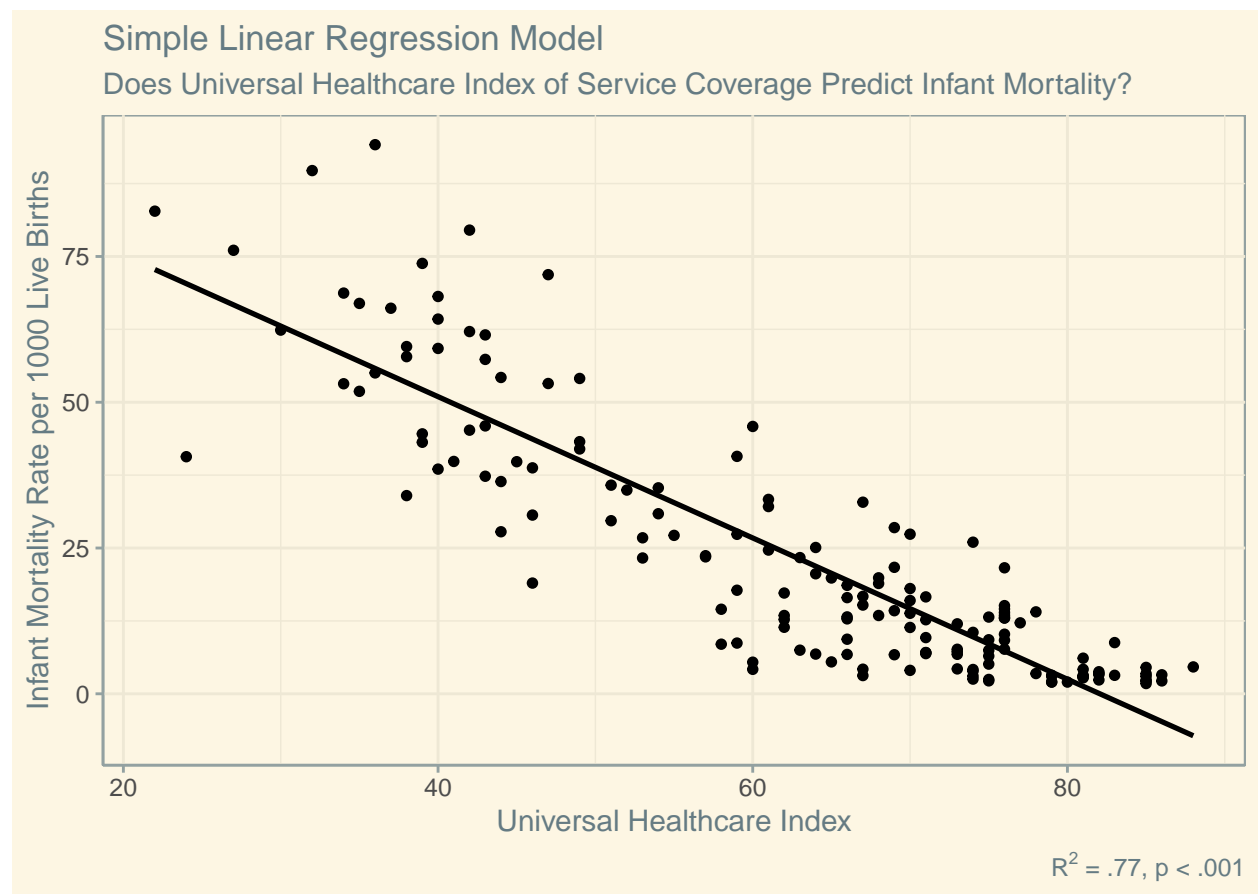
```
##
## Call:
## lm(formula = infant_mortality ~ uhc, data = all_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -29.677  -6.348  -0.680   6.080  38.370
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  99.40956    3.35024   29.67  <2e-16 ***
## uhc          -1.21138    0.05184  -23.37  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.52 on 160 degrees of freedom
## Multiple R-squared:  0.7734, Adjusted R-squared:  0.772
## F-statistic:   546 on 1 and 160 DF,  p-value: < 2.2e-16
```

The below plots assess the assumptions of the regression model. The first plot assesses the linearity assumption, the second plot assesses the normality assumption, and the third plot assesses the homogeneity of variance assumption.





The below plot visualizes the regression model. We can see that, on average, as UHC Index increases, infant mortality decreases. The R-squared is large, and the relationship is significant.



Discussion

Overall, the least squares linear regression model that measures the relationship between suicide rate and UHC index indicates that, when UHC Index increases by 1 point, suicides/100k population increase by about .12. Another way to interpret this is if UHC Index increase by 10 points, suicides/100k population increase by 1.2. Of course, the R-squared value in this model is only .29, so not all of the variation in UHC Index is represented in the suicides/100k population variable. However, a significant relationship still seems to be present. When thinking about why this might be, one possible explanation is that more developed countries with higher UHC indexes come with more middle class, white collar-type jobs. Moreover, a good UHC Index index does not always mean more access to affordable therapy.

The linear regression models that measure the relationships between UHC Index and life expectancy and infant mortality respectively are expected. Countries with more robust healthcare systems would of course be associated with higher life expectancy and lower infant mortality.

Future Research

Overall, I'd like to more closely examine the relationship between suicides and a country's healthcare system by attempting to answer the following research questions:

- Are people more likely to committ suicide when their country does not provide a single-payer healthcare system?
- Are people more likely to committ suicide if their country's single-payer healthcare system does not offer free therapy?
- Is there a relationship between suicide rates and cost of therapy?