Slide 1 Life Expectancy vs. Education

* Hello, my name is Colin O’Neill and I was tasked with investigating the relationship between Life Expectancy and Schooling. Schooling, as defined by the WHO, is the average # of years of formal education per capita for each country in the dataset.

Slide 2 Data Cleaning

* Obviously, to begin my analysis, I had to clean my data to ensure every value in the education column was valid, as well as those values in the Life expectancy column. To do this, I first looked at the number of items in each column.

Slide 3 First Data Cleaning Image

* I ran a value count and saw mismatched counts.

Slide 4 Second Data Cleaning Image

* Ran a second value count and still saw mismatched counts.

Slide 5 Third Data Cleaning Image

* I was having trouble with my regression analyses because there were too many zero values in the schooling column so I filtered the dataset for non-zero values as those are not relevant data points as it is most likely indicative of lack of information rather than legitimate information. At the end I ended with a clean dataset of 2742 rows.

Slide 6 Life Expectancy vs. Education (Consolidated)

* To perform the meat of my analysis I decided to perform 3 regression analyses. The first regression analysis is displayed on this slide showing Life Expectancy vs. Education at the combined consolidated level (i.e. developing and developed countries included in the dataset). There is very strong correlation here as indicated by the strong r value of .78.

Slide 7 Life Expectancy vs. Education (Developing)

* To dig deeper into the legitimacy of the previous regression analysis, I decided to bifurcate the data between developed and developing nations. As see in this regression analysis of just developing nations, there is still a strong correlation of r value .73. Perhaps this means that as a country increases education requirements, this directly impacts life expectancy, or perhaps there is something else at play (i.e. a confounding variable).

Slide 8 Life Expectancy vs. Education (Developed)

* The last regression analysis that I performed supports my hypothesis that there is correlation but not causation between life expectancy and education, as there is a weak correlation between Life Expectancy and Education for developed countries. As a country becomes developed, there are other factors at play increasing Life Expectancy outside of a country’s schooling/education.

Slide 9 Conclusion: Life Expectancy vs. Schooling

* Per my data analysis, there is a strong correlation between Life Expectancy and Schooling, specifically as it pertains to developing countries. However, the weak correlation observed when isolating the data to just developed countries indicates that there are confounding variables at play that are causing the correlation between Life Expectancy and Schooling in developing countries. As such, we can conclude correlation but NOT causation.