Final Report
due November 16, 2021 by 11:59 PM

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

11/16/2021

Background

Research Question

Data Overview

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.    :   16.0
##  1st Qu.:  177.0
##  Median :  602.5
##  Mean    : 1202.1
##  3rd Qu.: 1477.8
##  Max.    :11027.0
```

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:

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

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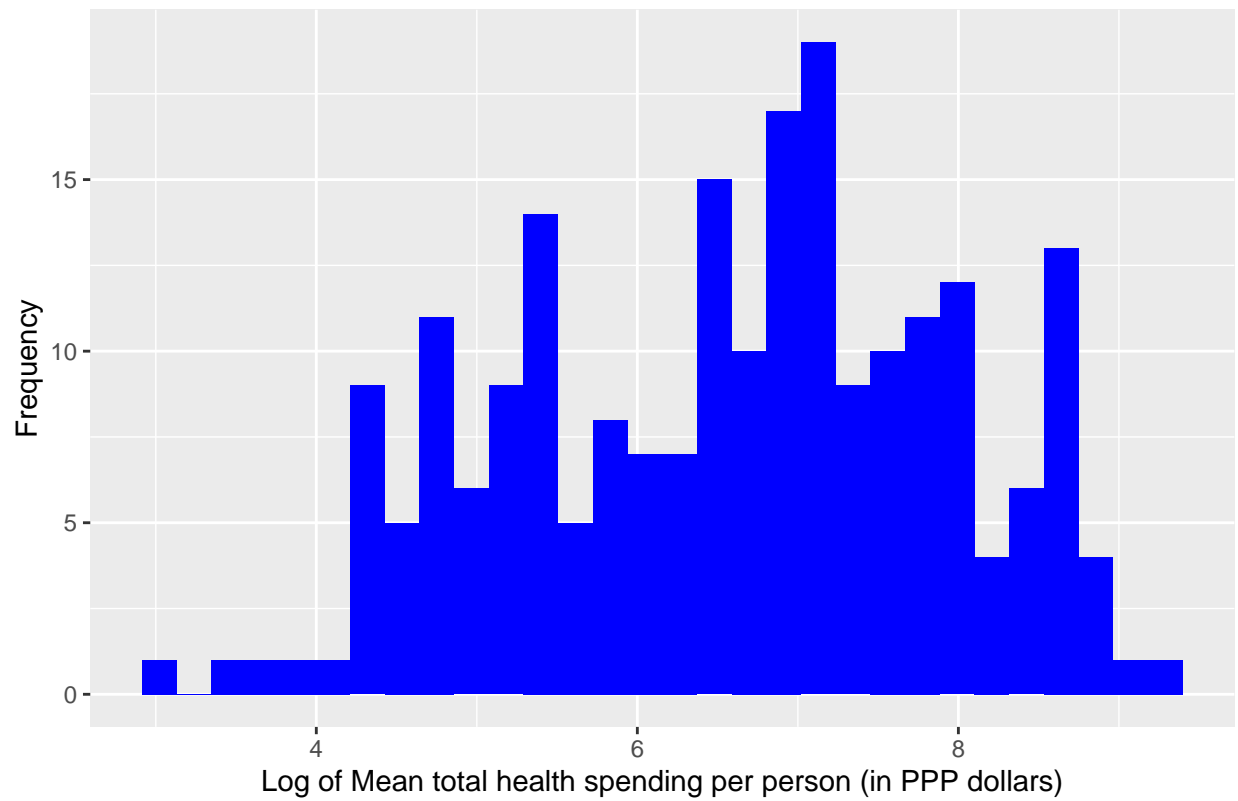
Then, we looked at healthcare spending over time in a few selected countries to see any trends in the data:

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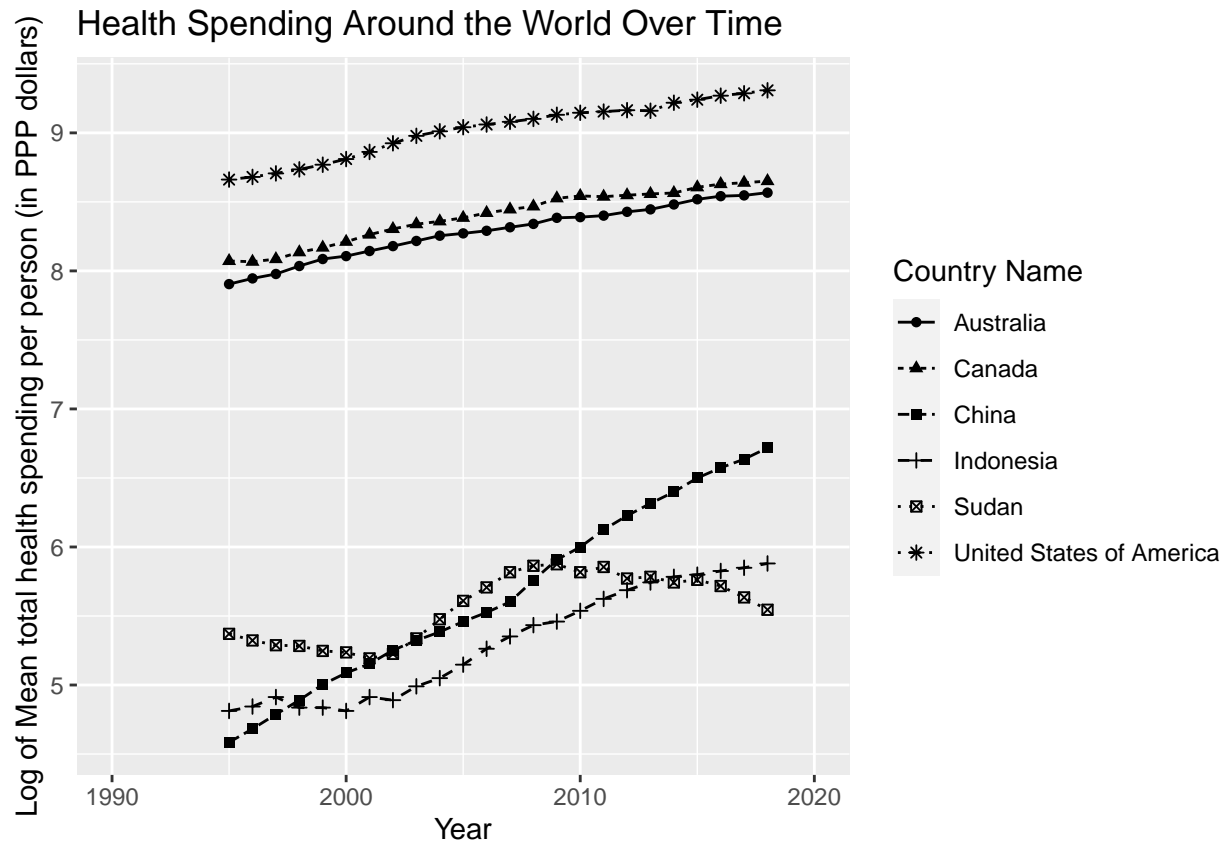
We found that healthcare spending generally increased over time in developed countries and was also significantly larger in Western than Eastern countries. This would be confirmed when we mapped health care spending. While mapping, we also mapped what would later be one of our response variables: life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

Health Spending Around the World
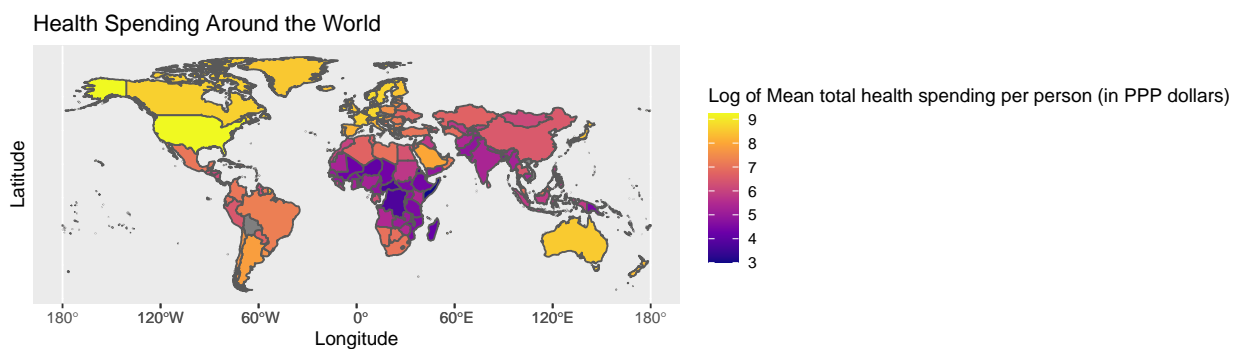
Mortality Rates Around the World
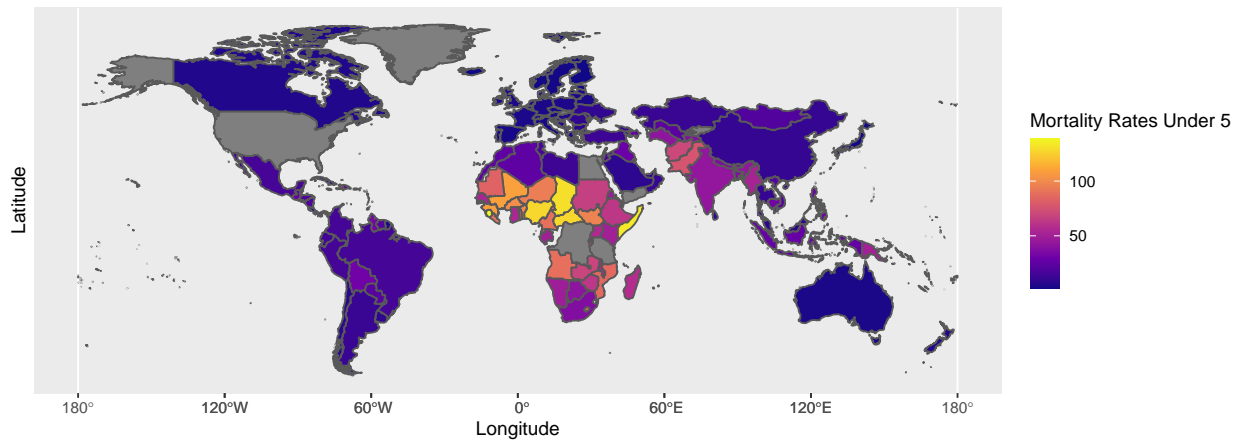
Life Expectancies Around the World
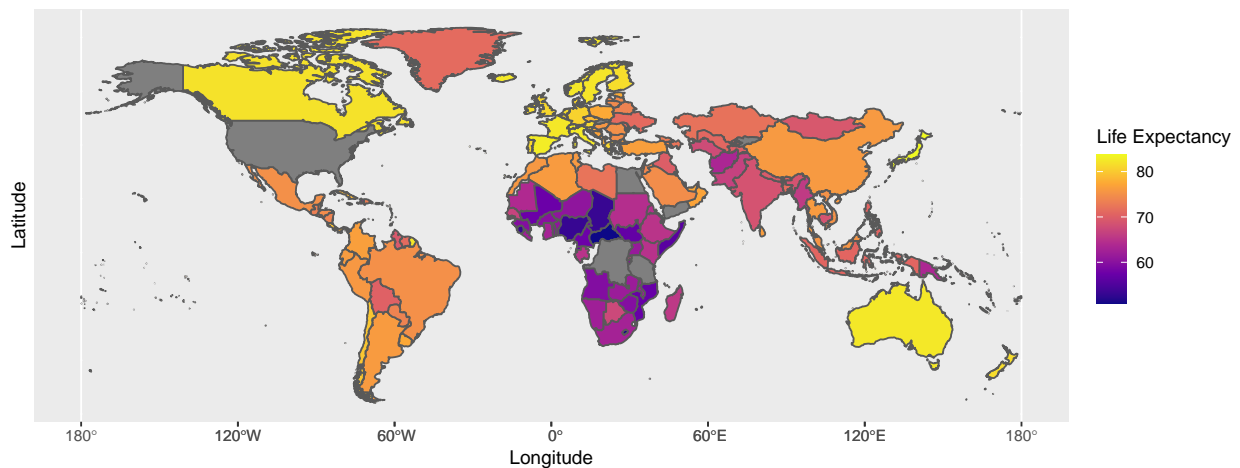
We saw the same pattern as in the line charts where Western countries had significantly higher healthcare spending but also higher life expectancy and lower mortality rates. The same trend was generally true for developed compared to non-developed regions. We also began to notice the problem of missingness in our dataset where mortality rates and life epxectancies 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:

Log of Mean total health spending per person (in PPP dollars)
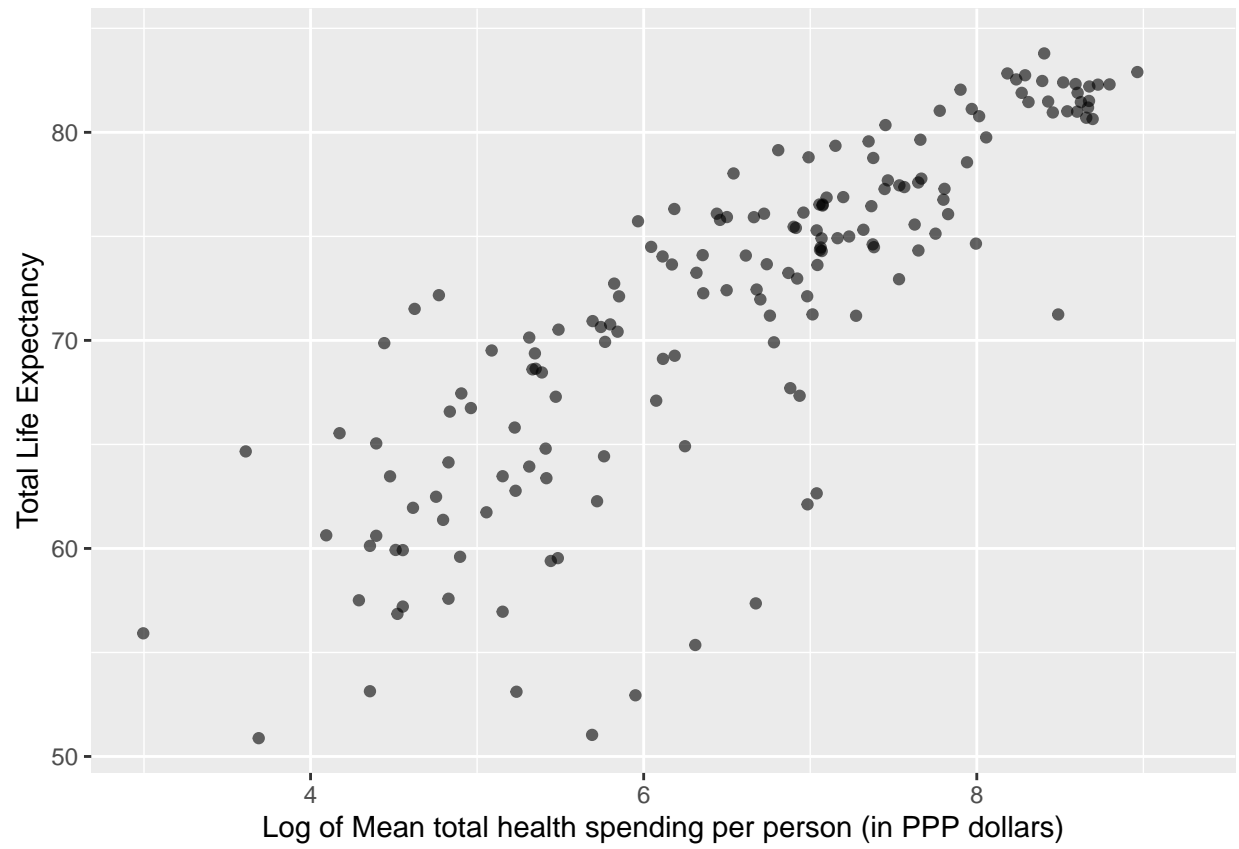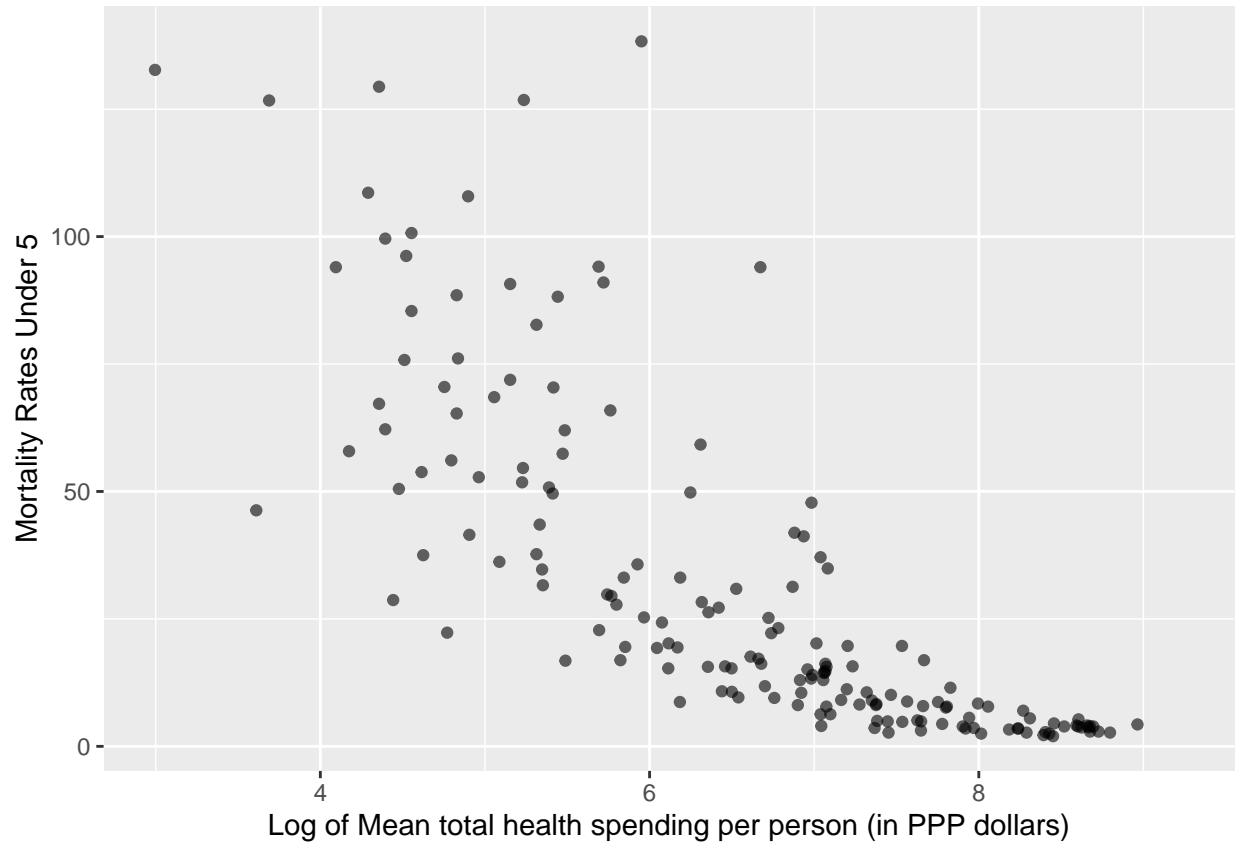
There is a clear positive linear relationship between healthcare spending and life expectancy but a more exponential-decay trend in healthcare spending and mortality rates under 5. Viewing these trends would come in helpful when modelling. Within the descriptive statistics portion of our project, we 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 to great to deal with.

Modelling