Project 1 Report: Global Maternal and Newborn Health

Contents

1. Introduction

a. Set up

Load packages.

b. Description of the dataset(s)

`mmr_ihme` is a dataset from the Institute for Health Metrics and Evaluation (gapminder.org). It contains the maternal mortality ratios (MMR), numeric, for each country (1 row/country) from 1989 to 2014. MMR is the number of maternal deaths per 100,000 live births.

`newborn_mortality_rate_per_1000`is a dataset from the UN Inter-agency Group for Child Mortality Estimation (childmortality.org, gapminder.org). Each row is a country, and the newborn mortality rate (NMR), a numeric variable, is given for the years 1950-2014. Newborn mortality rate is measured as the number of children dying within the first 28 days of life per 1000 live births.

`birth_attendants` is a dataset from World Bank (worldbank.org, gapminder.org). Each row is a country, and the percent of births attended by skilled health staff is given for each country for each year from 1979-2019. This numeric variable is measured as the percentage of deliveries attended by personnel trained to give the necessary supervision, care, and advice to women during pregnancy, labor, and the postpartum period.

`medical_doctors_per_1000_people` is from World Bank (worldbank.org, gapminder.org). Each row represents a country, and the number of medical doctors per 1000 people, numeric, is given for each country from 1959-2018. "Medical doctors" includes general and specialized medical practitioners.

These four datasets can be joined together with the "country" ID.

```
# Upload data
mmr_ihme <- read_csv("mmr_ihme.csv")
newborn_mortality_rate_per_1000 <- read_csv("newborn_mortality_rate_per_1000.csv")
birth_attendants <- read_csv("birth_attendants.csv")
medical_doctors_per_1000_people <- read_csv("medical_doctors_per_1000_people.csv")
```

c. Introducing the Topic

While there has been progress in reducing the maternal mortality ratio and newborn mortality rate, there are still preventable deaths. Some countries still have high incidences of maternal and newborn deaths. Ensuring that mothers can have healthy pregnancies and deliveries and that newborns can have a healthy start should be a global health priority. According to the Gates Foundation, maternal health is closely linked to a newborn's survival. Factors such as the availability and access of skilled birth attendants and medical doctors mothers may correlate to maternal mortality ratios. These health factors may also influence newborn health. There should be a positive relationship between NMR and MMR, based on the Gates Foundation's statement. It is also likely that there will be a positive relationship between MMR and the percentage of births attended by skilled personnel.

d. Research Question

What relationships exist between MMR, NMR, the percent of births attended by skilled personnel, and the number of medical doctors per 1000 people?

Before beginning, pairwise comparisons between each variable (MMR, NMR, percentage of births attended by skilled personnel, and medical doctors per 1000 people) were made. From this, the most interesting relationships of these comparisons will be explored in this report.

2. Tidying

```
#Select only 2014 column from Maternal Mortality Ratio and rename column
mmr2014 <- mmr_ihme %>% select(country, "2014") %>% rename("MMR" = "2014")


#Select only 2014 column from % Skilled Birth Attendants and rename column
attendants2014 <- birth_attendants %>% select (country, "2014") %>%
                                       rename("Attendants" = "2014")


#Select only 2014 column from Newborn Mortality Rate per 1000 and rename column
nmr2014 <- newborn_mortality_rate_per_1000 %>% select (country, "2014") %>%
                                           rename ("NMR" = "2014")

#Select only 2014 column from Medical Doctors per 1000 people and rename column
doctors2014 <- medical_doctors_per_1000_people %>% select (country, "2014") %>%
                                                   rename ("Doctors" = "2014")

#Determine number of observations in each dataset using count()
count(mmr2014)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   194
```

```
count(attendants2014)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   204
```

```
count(nmr2014)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   195
```

```
count(doctors2014)
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1   207
```

Each dataset was cleaned before joining. This was done by selecting only the reported data from 2014 for each country. Individually, each dataset contains a "country" variable and the statistic variable (MMR, NMR, percentage of births attended by skilled personnel, or medical doctors per 1000 people)

Before joining the datasets, there were:

- 194 observations in the maternal mortality ratio dataset,
- 204 observations in the % Births Attended by Skilled Professionals dataset,
- 195 observations in the Newborn Deaths per 1000 dataset, and
- 207 observations in the Medical Doctors per 1000 people dataset.

The "country" ID appears in all datasets. The datasets have no other variables/IDs in common. No IDs will be left out after joining, since each dataset has only the country variable and the variable related to that dataset.

3. Joining/ Merging

```
# Put all data frames into list
df_list <- list(mmr2014, attendants2014, nmr2014, doctors2014)

# Missing values
df_list %>% reduce(anti_join, by = 'country')
```

```
## # A tibble: 1 x 2
##   country   MMR
##   <chr>   <dbl>
## 1 Taiwan   12.6
```

```
# Merge all data frames together
finaldata <- df_list %>% reduce(full_join, by='country')

# Number of observations in final dataset
nrow(finaldata)
```

```
## [1] 214
```

After joining the 4 datasets, there were a total of 214 observations. This could be due to how the data was collected and what the original data collectors classified as a country. For example, Hong Kong and Taiwan were considered countries in some of the datasets, but if they are officially recognized as countries depends. An anti_join revealed Taiwan's MMR value was dropped.

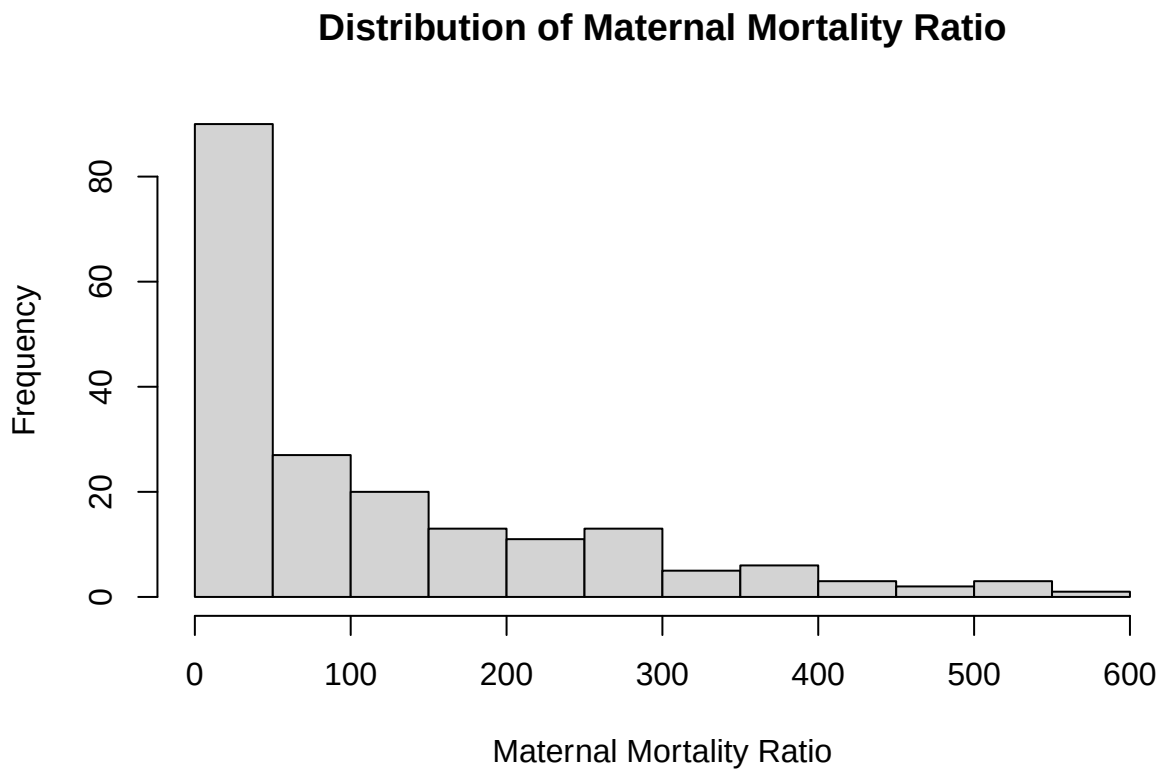
Potential issues are related to some countries not having data for one or all of the variables. Since this is data from one year, some countries may not have had data collected that year, so there are NA values for some or all of the four variables.

4. Wrangling and Analyzing the Data

```
#Look at summary statistics for maternal mortality ratio in 2014
finaldata %>% select(MMR) %>% drop_na() %>% summarize(mean_MMR = mean(MMR),
                                        sd_MMR = sd(MMR),
                                        median_MMR = median(MMR),
                                        IQR_MMR = IQR(MMR)) %>%
                  pivot_longer(cols =
                                  c('mean_MMR', 'sd_MMR', 'median_MMR','IQR_MMR'),
                               names_to = "Statistic",
                               values_to ="Measurement")
```

```
## # A tibble: 4 x 2
##   Statistic  Measurement
##   <chr>            <dbl>
## 1 mean_MMR         116.
## 2 sd_MMR           131.
## 3 median_MMR        60.4
## 4 IQR_MMR          166.
```

```
hist(finaldata$MMR,
     xlab = "Maternal Mortality Ratio",
     main = "Distribution of Maternal Mortality Ratio")
```




```
#Look at summary statistics for newborn mortality rate in 2014
finaldata %>% select(NMR) %>% drop_na() %>% summarize(mean_NMR = mean(NMR),
                                        sd_NMR = sd(NMR),
```

```
                                    median_NMR = median(NMR),
                                    IQR_NMR = IQR(NMR)) %>%
                pivot_longer(cols =
                               c('mean_NMR', 'sd_NMR', 'median_NMR','IQR_NMR'),
                             names_to = "Statistic",
                             values_to ="Measurement")
```

```
## # A tibble: 4 x 2
##   Statistic  Measurement
##   <chr>            <dbl>
## 1 mean_NMR          13.6
## 2 sd_NMR            11.3
## 3 median_NMR         9.8
## 4 IQR_NMR           17.5
```

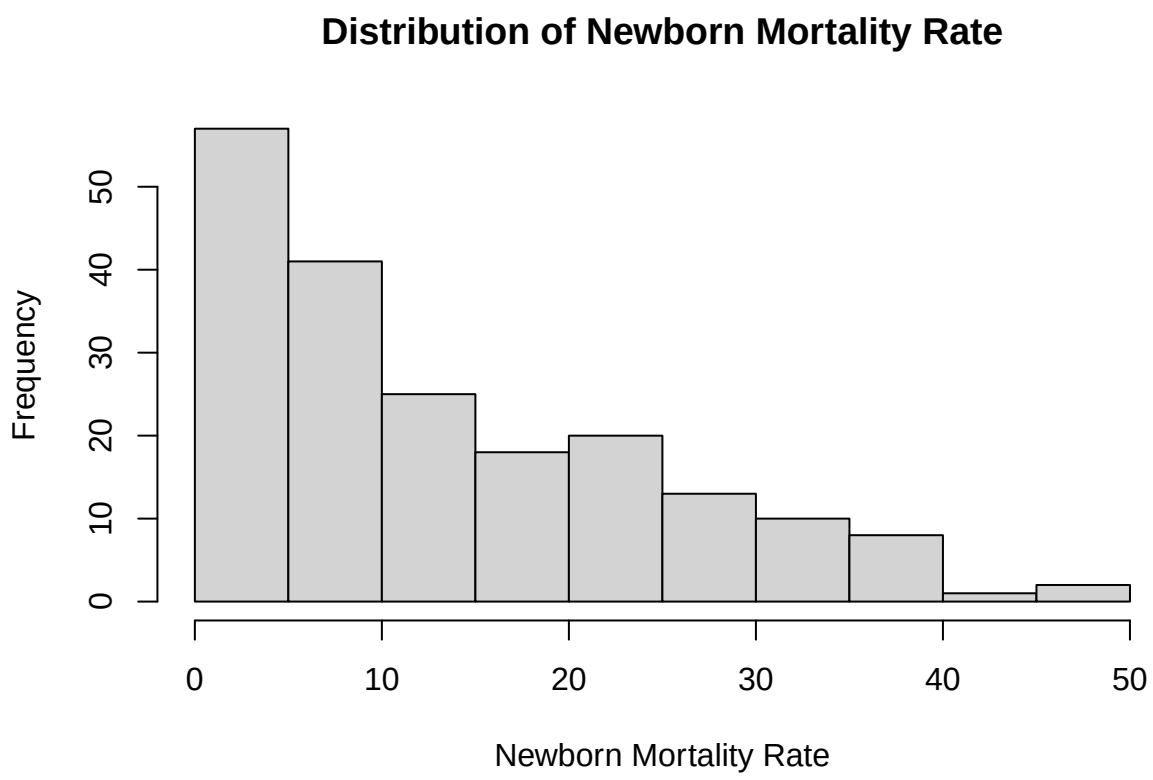
```
summary(finaldata$NMR)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.70    4.25    9.80   13.61   21.75   48.70      19
```

```
IQR(finaldata$NMR, na.rm = TRUE)
```

```
## [1] 17.5
```

```
hist(finaldata$NMR,
     xlab = "Newborn Mortality Rate",
     main = "Distribution of Newborn Mortality Rate")
```



```r
#Look at summary statistics for birth attendants in 2014
finaldata %>% select(Attendants) %>% drop_na() %>%
                                      summarize(mean_Attendants = mean(Attendants),
                                      sd_Attendants = sd(Attendants),
                                        median_Attendants = median(Attendants),
                                        IQR_Attendants = IQR(Attendants)) %>%
                  pivot_longer(cols =
                                 c('mean_Attendants', 'sd_Attendants',
                               names_to = "Statistic",
                               values_to ="Measurement")
```

```
## # A tibble: 4 x 2
##   Statistic         Measurement
##   <chr>                   <dbl>
## 1 mean_Attendants         93.2
## 2 sd_Attendants           14.2
## 3 median_Attendants       99.1
## 4 IQR_Attendants           4.90
```

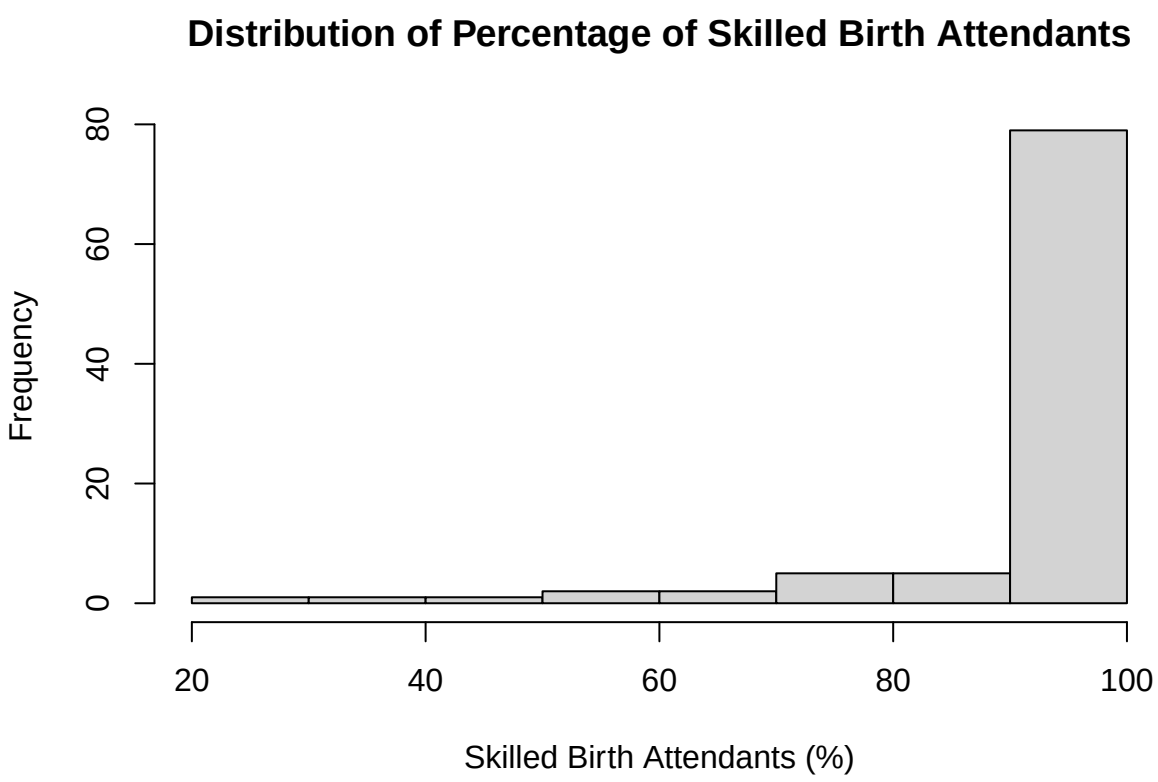
```r
summary(finaldata$Attendants)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   24.30   95.00   99.10   93.22   99.90  100.00     118
```

```r
IQR(finaldata$Attendants, na.rm = TRUE)
```

```
## [1] 4.9
```

```r
hist(finaldata$Attendants,
     xlab = "Skilled Birth Attendants (%)",
     main= "Distribution of Percentage of Skilled Birth Attendants")
```



The distribution of maternal mortality ratio is right skewed. The median maternal mortality ratio is 60.35 deaths per 100,000 live births. The IQR is 166.4.

The distribution of newborn mortality rate is also right skewed. The median newborn mortality rate is 9.80 deaths per 1,000 live births, and the IQR is 17.5.

The distribution of the percentage of births attended by skilled personnel is left skewed. The median is 99.10 %, and the IQR is 4.9.

5. Visualizing the Data

```
# 1 variable density plot
finaldata %>% select(MMR) %>% drop_na(MMR) %>% ggplot(aes(x = MMR)) +
  geom_density(fill = "lightblue") +
  scale_x_continuous(breaks = seq(0,600,50)) +
  scale_y_continuous(breaks = seq(0, 0.006, .001)) +
  labs(title = "Distribution of Maternal Mortality Ratio",
       x = "Maternal Mortality Ratio (per 100,000 live births)",
       y = "Density") +
  theme_bw()
```

From this density plot, it can be seen that maternal mortality rate is right skewed. The most frequent maternal mortality ratio is around 50 deaths per 100,000 live births, with a density of 0.005. There are outliers for maternal mortality ratio, which are as high as about 550 deaths per 100,000 live births.

```
#Creating a new variable to describe maternal mortality ratio
#Remove any countries that are missing values for any of the four variabes
cleaned <- finaldata %>% drop_na()
```

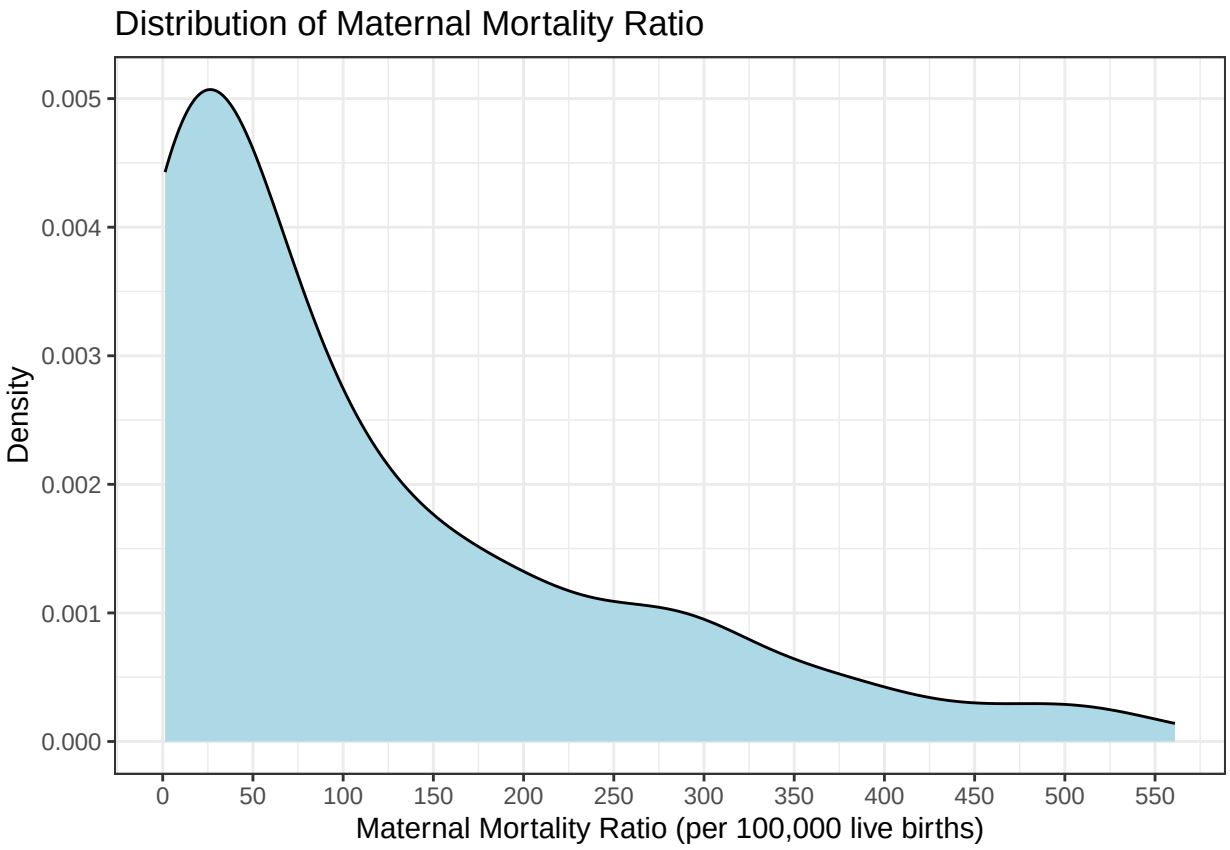


Figure 1: Figure 1

```r
#Find percentile ranges for maternal mortality ratio
quantile(cleaned$MMR, probs = c(0.333,0.667,1))
```

```
##     33.3%     66.7%      100%
##   7.16902  48.80610 489.00000
```

```r
#Create new column for maternal mortality ratio based on the percentile ranges
newvariable <- cleaned %>% mutate(Maternal_Mortality =
                                    case_when(MMR <=7.16902  ~'low',
                                    MMR <= 48.80610 ~ "medium",
                                    MMR <= 489.0000 ~ 'high')) %>%
                           relocate(Maternal_Mortality, .after="MMR") %>%
                           rename("Category" = "Maternal_Mortality")


#Summary statistics for MMR categories
as.data.frame(newvariable) %>% group_by(Category) %>%
  summarize(freq = n()) %>%
  arrange(desc(freq))
```

```
## # A tibble: 3 x 2
##   Category  freq
##   <chr>    <int>
## 1 high        21
## 2 low         21
## 3 medium      20
```

Maternal mortality ratio categories were calculated using percentile groupings of 0.333, 0.667, and 1. Countries within the 33rd percentile had maternal mortality ratios of 7.16902 deaths or less were categorized as "low"; countries with maternal mortality ratios of more than 7.16902 deaths but less than or equal to 48.80610 deaths were considered "medium"; countries with more than 48.80610 deaths were considered "high".

The frequency of countries categorized as "high" for maternal mortality ratios was 21; the frequency of countries categorized as "medium" for maternal mortality ratios was 20; the frequency of countries categorized as "low" was 21. The frequency of high and low was the same, followed by medium.

```r
 # 2 variable plot comparing mean NMR and  MMR categorizations
newvariable %>% select(Category, Attendants) %>% group_by(Category) %>%
  ggplot(aes(x = reorder(Category, -Attendants, na.rm = TRUE), y = Attendants)) +
  geom_bar(aes(y = Attendants),
  stat = "summary", fun = "mean",
  fill = "#4059AD") +
  scale_y_continuous(breaks = seq (0, 100, 10)) +
  labs(title = "Maternal Mortality Ratio & Mean Births Attended by Skilled Personnel",
       x = "Maternal Mortality Ratio Category",
       y = "Mean Percentage of Births Attended by Skilled Personnel") +
   theme(text=element_text(size = 14)) +
  geom_errorbar(stat = "summary", fun.y = "mean_se", width = 0.5) +
  theme_light()
```

This barplot shows the mean number of births attended by skilled personnel for each of the three categories of maternal mortality ratios (high, medium, and low). This visualization indicates that countries categorized as having high maternal mortality ratios have the lowest average number births attended by skilled personnel. Countries categorized as having low or medium maternal mortality ratios have nearly the same average percentages of births attended by skilled professionals. Overall, this plot illustrates that there is a connection between having a high maternal mortality ratio and a lower percentage of births attended by skilled

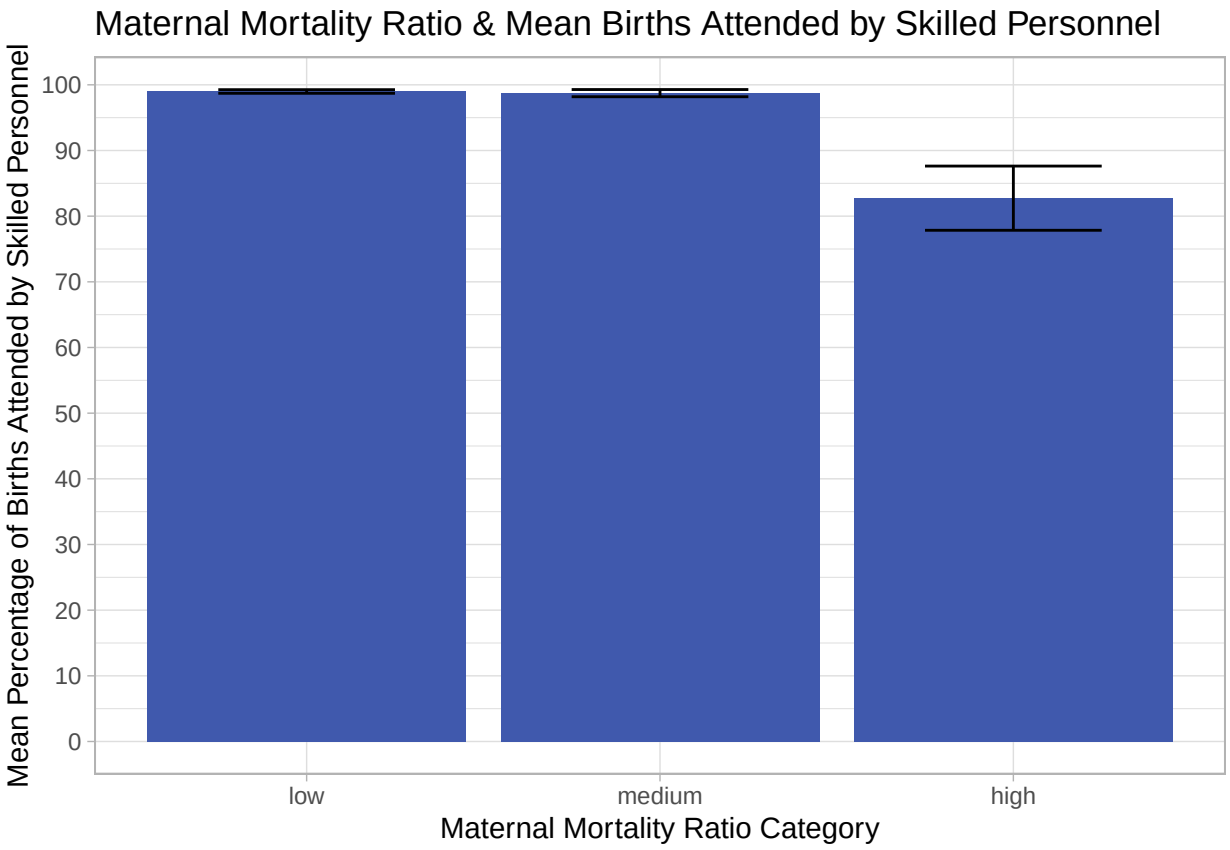


Figure 2: Figure 2

professionals.

```
#plot to compare MMR and NMR
finaldata %>% filter(MMR > 60.35, NMR > 9.80) %>% ggplot(aes(x = MMR, y = NMR)) +
  geom_point(color = "cornflowerblue") +
  geom_smooth(method = "lm", se = FALSE, color ="cornflowerblue") +
  labs(title =
         "Relationship between Maternal Mortality Ratio and Newborn Mortality Rate",
       x = "Maternal Mortality Ratio (per 100,000 live births)",
       y = "Newborn Mortality Rate (per 1,000 live births)")+
  scale_x_continuous(breaks = seq(0, 600, 50)) +
  scale_y_continuous(breaks= seq(0, 50, 5)) +
  theme_minimal()
```

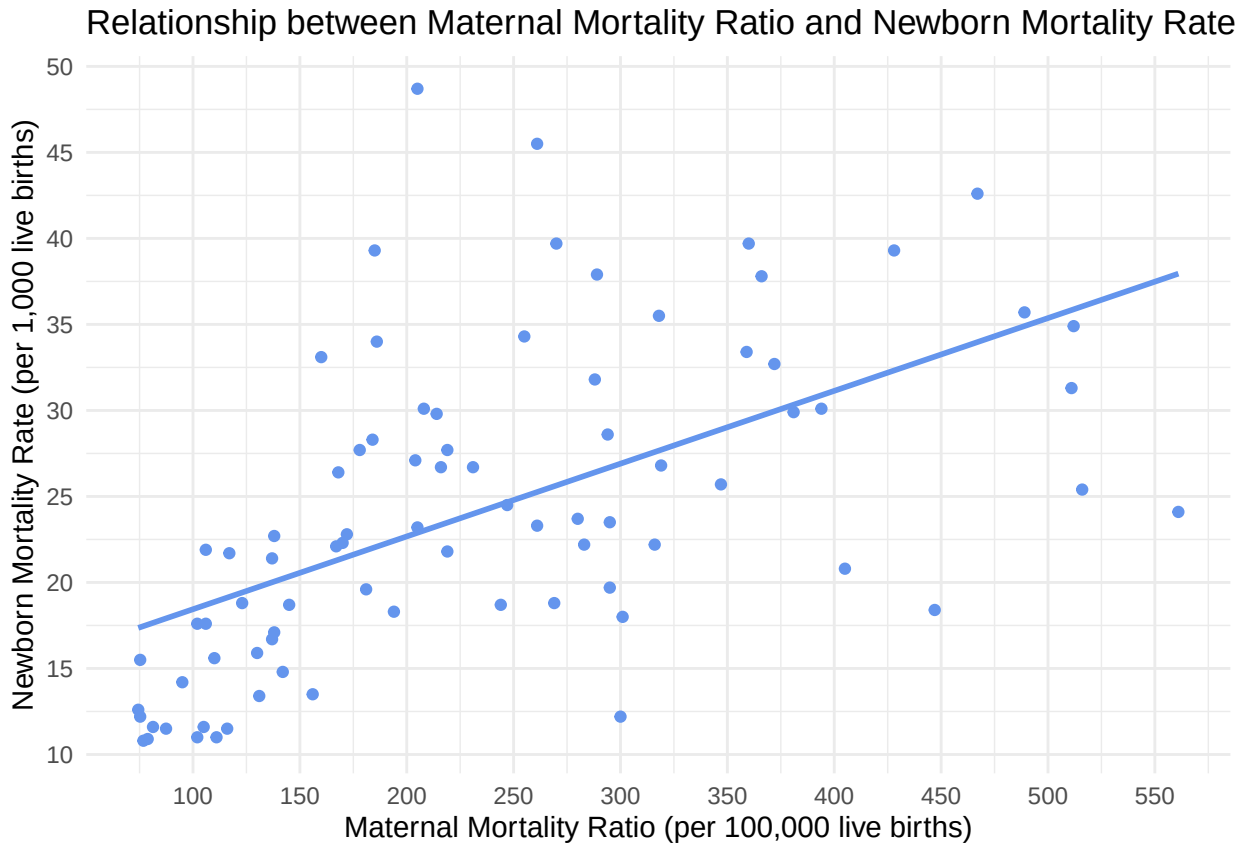


Figure 3: Figure 3

This scatterplot compares the maternal mortality ratio and newborn mortality rate for countries that had MMR and NMR above the median values. 82 countries had both an MMR above the median MMR of 60.35 deaths per 100,000 live births and an NMR above the median 9.80 deaths per 1,000 live births. The positive linear regression line indicates that there is a positive relationship between MMR and NMR for these countries.

```
# Create visualization comparing NMR, MMR, and % Skilled Birth Attendants
dropped <- finaldata %>% drop_na(MMR, NMR, Attendants)

ggplot(data=dropped, aes(x = MMR, y = NMR, color = Attendants)) +
  geom_point(aes(color = Attendants), size = 2) +
```

```
labs(title = "Relationship between MMR, NMR, and Skilled Birth Attendants",
     x = "Maternal Mortality Ratio (per 100,000 live births)",
     y= "Newborn Mortality Rate (per 1000 live births)",
     color = "% Skilled Birth Attendants") +
scale_color_gradient(low = "yellow", high = "darkblue") +
scale_x_continuous(breaks = seq(0, 500, 50)) +
scale_y_continuous(breaks = seq(0, 40, 5)) +
theme_light()
```

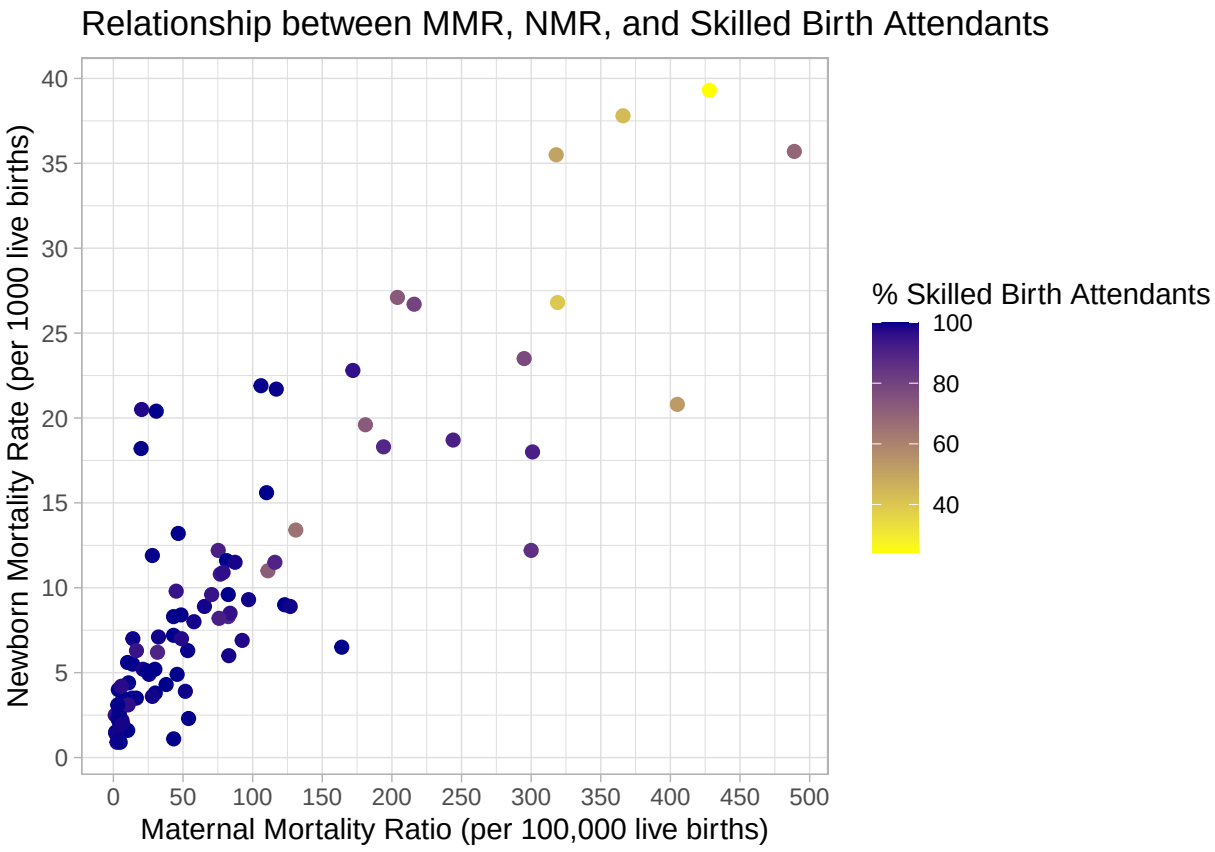


Figure 4: Figure 4

This scatterplot compares maternal mortality ratio, newborn mortality rate, and the percentage of births attended by skilled personnel. Maternal mortality ratio is on the x-axis, newborn mortality rate is on y-axis, and each point is colored by the percentage of births attended by skilled personnel. From the graph, it can be seen that countries with lower maternal mortality ratios and newborn mortality rates have the highest percentage of births attended by skilled personnel (points colored dark purple). Countries with higher maternal mortality ratios and higher newborn mortality rates had a smaller percentage of births attended by skilled personnel.

7. Discussion

Figures 2, 3, and 4 answer the question of how maternal mortality ratio, newborn mortality rate, and the percentage of births attended by skilled personnel are related. Figure 2 illustrates that countries categorized as having "high" maternal mortality ratios, in the 66.7th percentile and higher, have a lower mean percentage of births attended by skilled personnel as compared to countries categorized in either the "low" or "medium"

maternal mortality ratio categories. Across all countries with "high" maternal mortality ratios, the mean percentage of births attended by skilled personnel was around 83%, while for countries in the "low" or "medium" categories, the mean was closer to 98-99%. This would indicate that countries with higher maternal mortality ratios also have lower percentages of births attended by skilled personnel.

Figure 3 illustrates the relationship between maternal mortality ratio and newborn mortality rate. For countries with maternal mortality ratios and newborn mortality rates above the median, there is a positive relationship between the two. It is clear from this visualization that maternal mortality ratio and newborn mortality rate are positively related.

Figure 4 illustrates the relationship between maternal mortality ratio, newborn mortality rate, and the percentage of births attended by skilled personnel. Countries with higher maternal mortality ratios and newborn mortality rates have a lower percentage of births attended by skilled personnel, between 40-60%, as compared to 90-100% in countries with lower maternal mortality ratios and newborn mortality rates. This figure illustrates how the percentage of skilled birth attendants contributes to maternal and newborn mortality.

This project helped me understand how to approach wrangling and visualizing different types of data. It gave me more insight into how to ask questions and how to answer those questions using the skills of a data scientist. The most challenging part of the project was getting started and figuring out the best approach. For future projects, it might be more insightful to look at trends over time for each country, rather than one particular year, and to consider other variables, such as health spending, life expectancy, and number of children. There is still much to be explored and advanced in the areas of maternal and newborn health.

Acknowledgments:

https://www.gatesfoundation.org/our-work/programs/gender-equality/maternal-newborn-and-child-health

To Dr. Guyot for her helpfulness in answering my questions!