Data Analysis

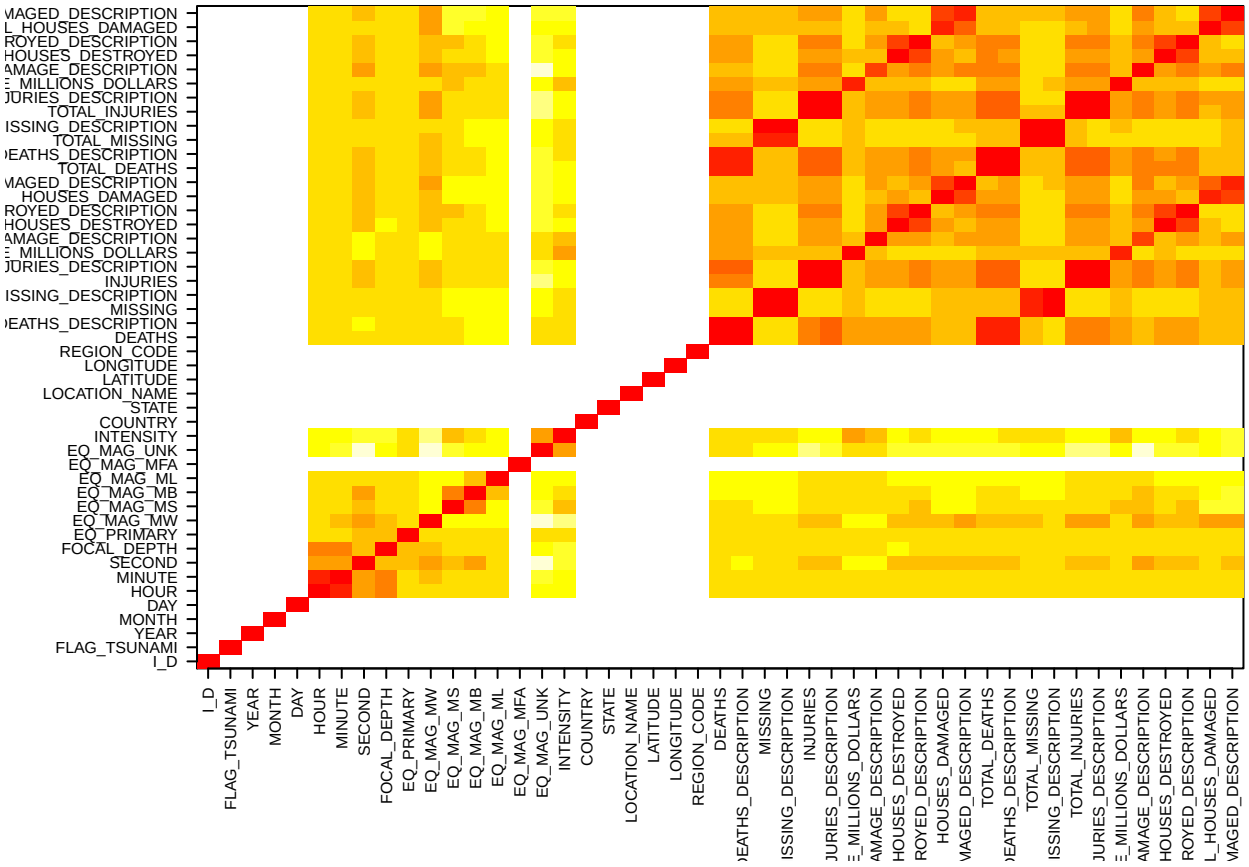
Chen Chen

11/29/2017

After loading the earthquake data, the population data, and the GDP data into R studio, I plotted the following plots:

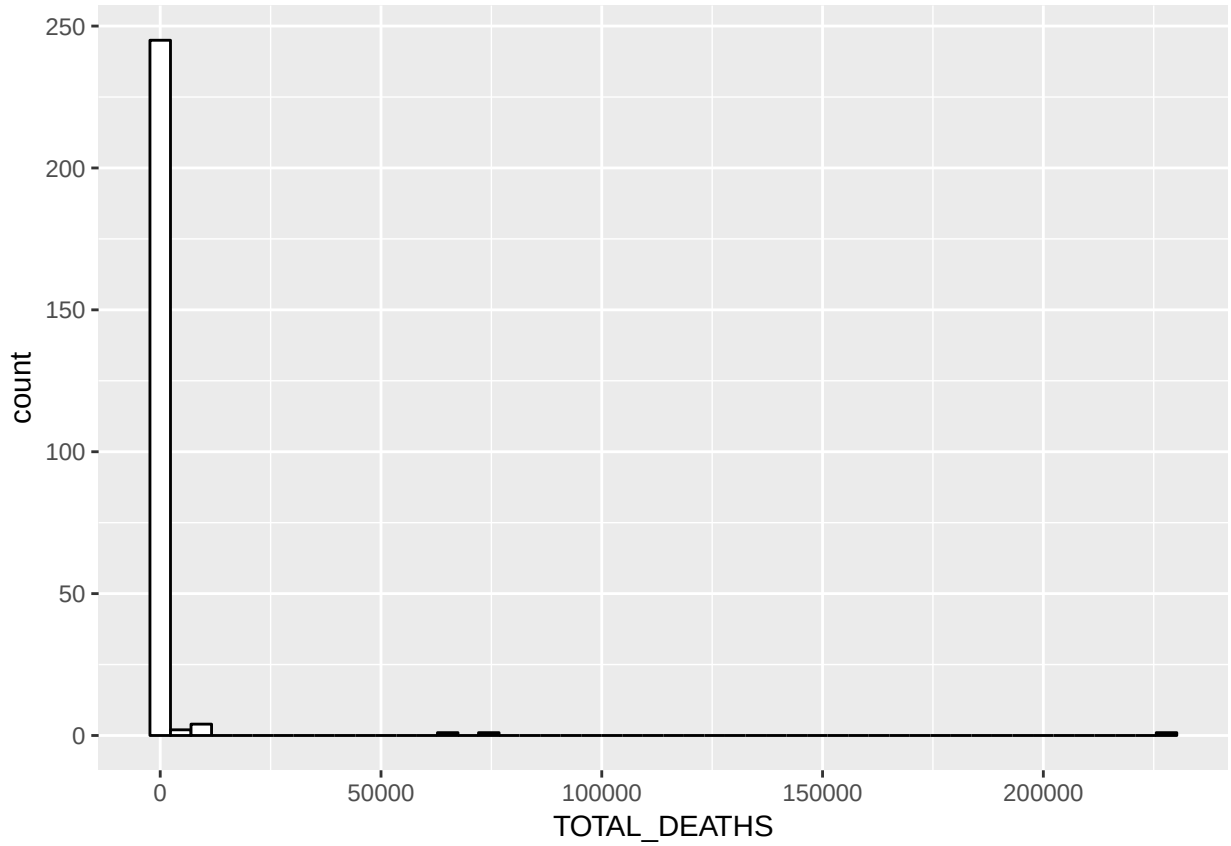
Correlation of the earthquake data

Total deaths and deaths are strongly correlated. The death count and the death description are strongly correlated as well.



Distrubtion of total deaths

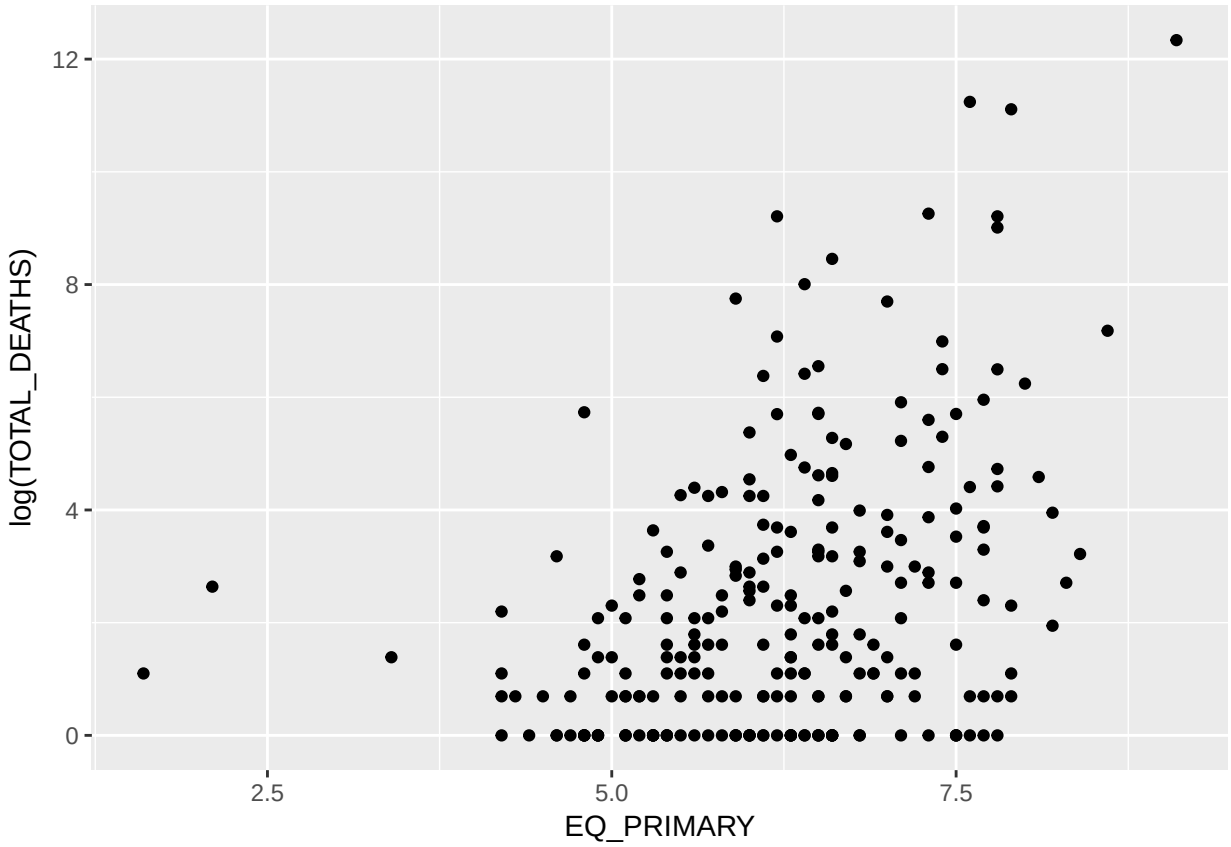
```
ggplot(df_eq_gdp,aes(x=TOTAL_DEATHS)) +
  geom_histogram(bins = 50,color="black",fill = "white")
```



Total deaths VS Earthquake Magnitude

As expected, more people died in large earthquakes.

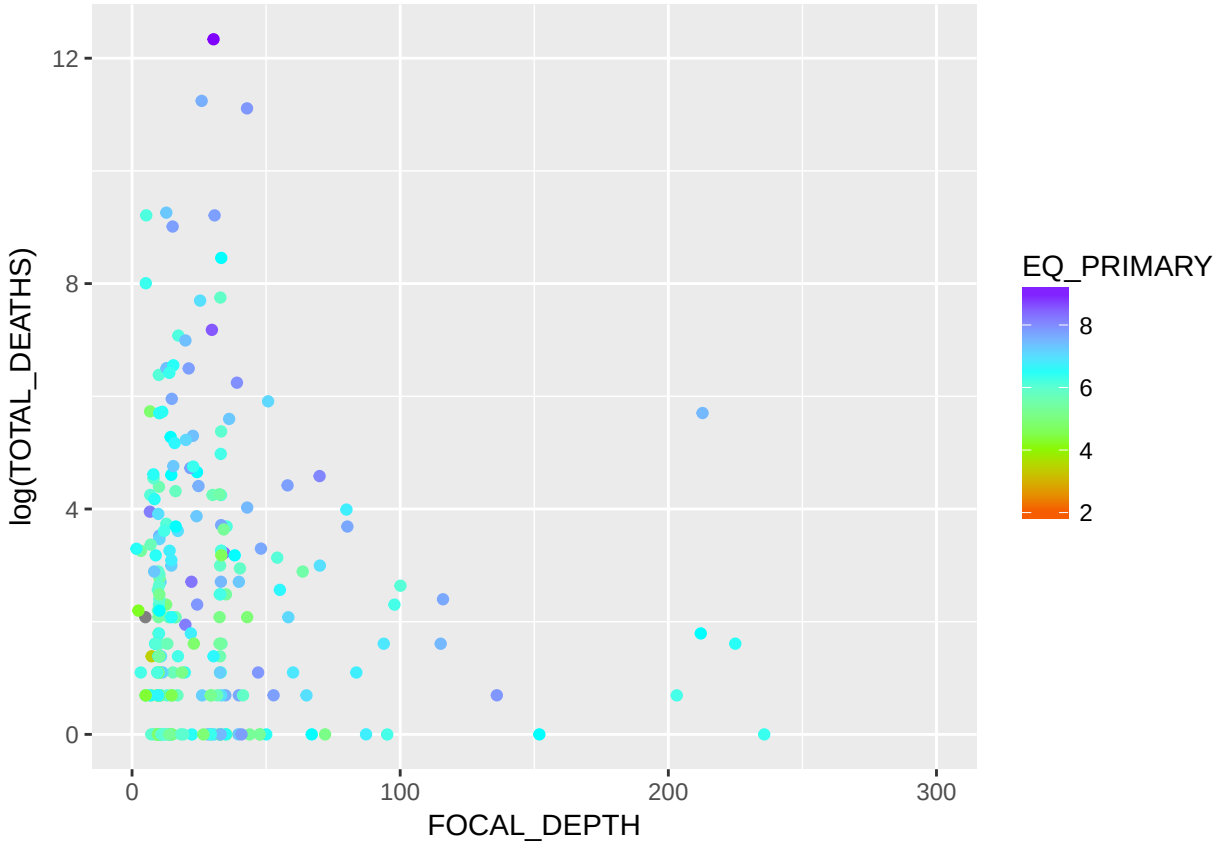
```
ggplot(country_eq,aes(x=EQ_PRIMARY, y = log(TOTAL_DEATHS))) +
  geom_point()
```



Total deaths VS Focal Depth of earthquakes

There is a weak correlation between focal depth and total deaths. More deaths are correlated with shallower earthquakes.

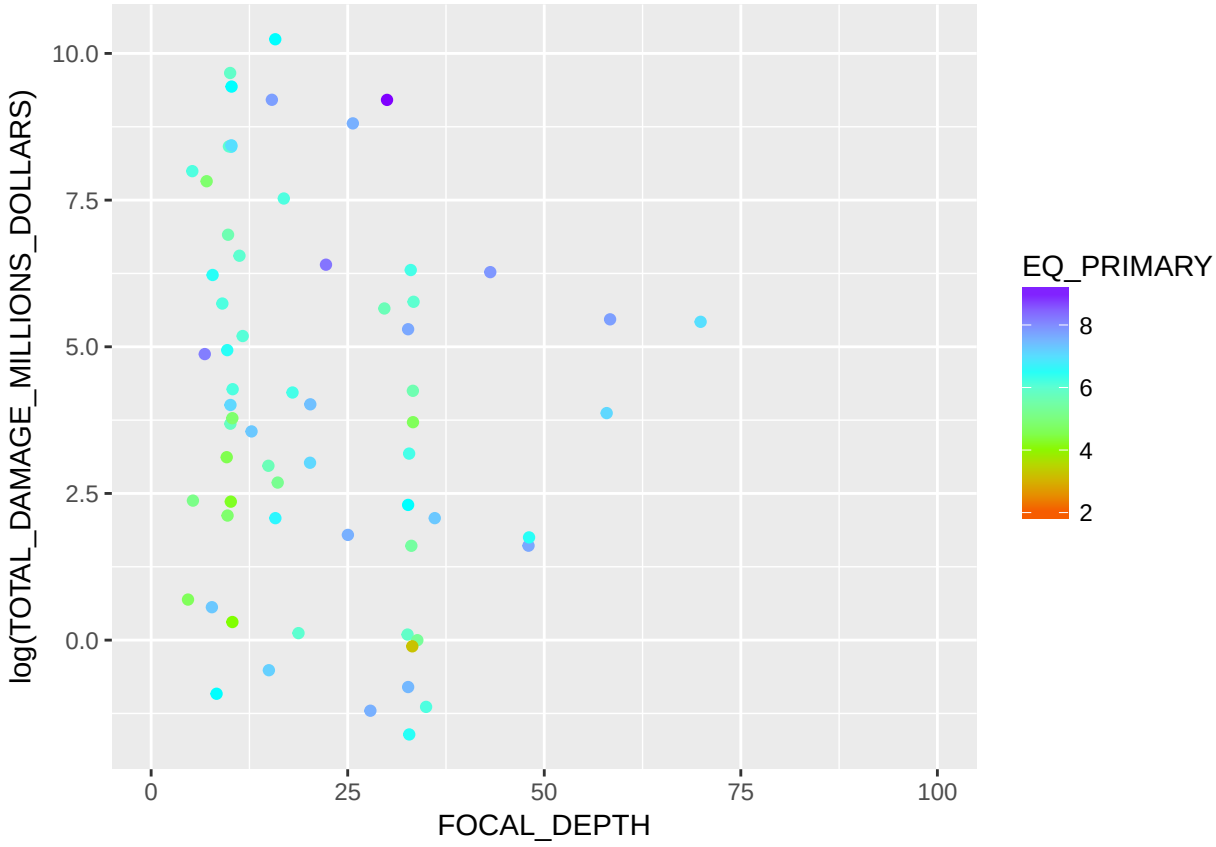
```
ggplot(country_eq,aes(x=FOCAL_DEPTH, y = log(TOTAL_DEATHS),color = EQ_PRIMARY)) +
  geom_jitter() +
  scale_color_gradientn(colors=rainbow(4)) +
  scale_x_continuous(limits = c(0, 300))
```



Total damage VS Focal Depth of earthquakes

There is no obvious correlation between focal depth and total damages.

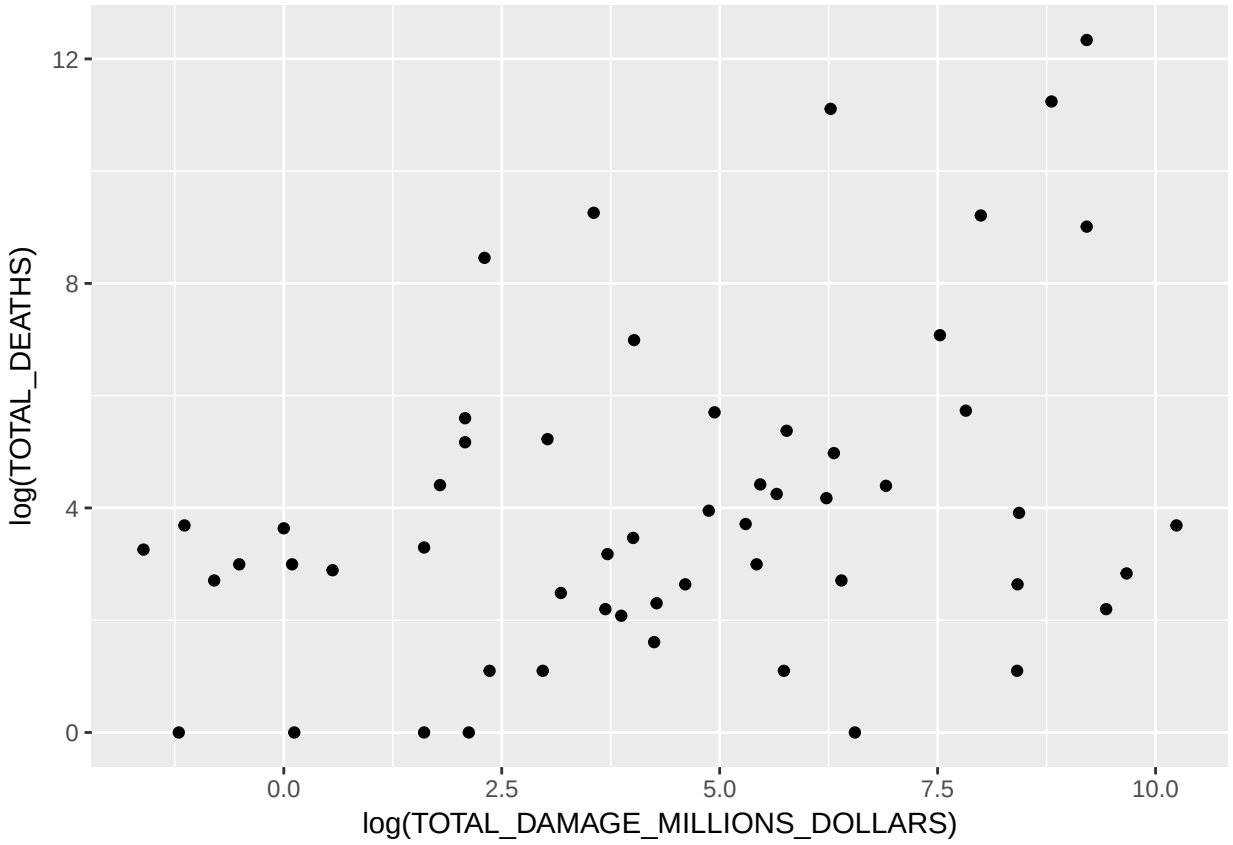
```
ggplot(country_eq,aes(x=FOCAL_DEPTH, y = log(TOTAL_DAMAGE_MILLIONS_DOLLARS),
                      color = EQ_PRIMARY)) +
  geom_jitter() +
  scale_color_gradientn(colors=rainbow(4)) +
  scale_x_continuous(limits = c(0, 100))
```



Total Damage VS Total Deaths

There is a weak correlation of higher damage with higher death rate.

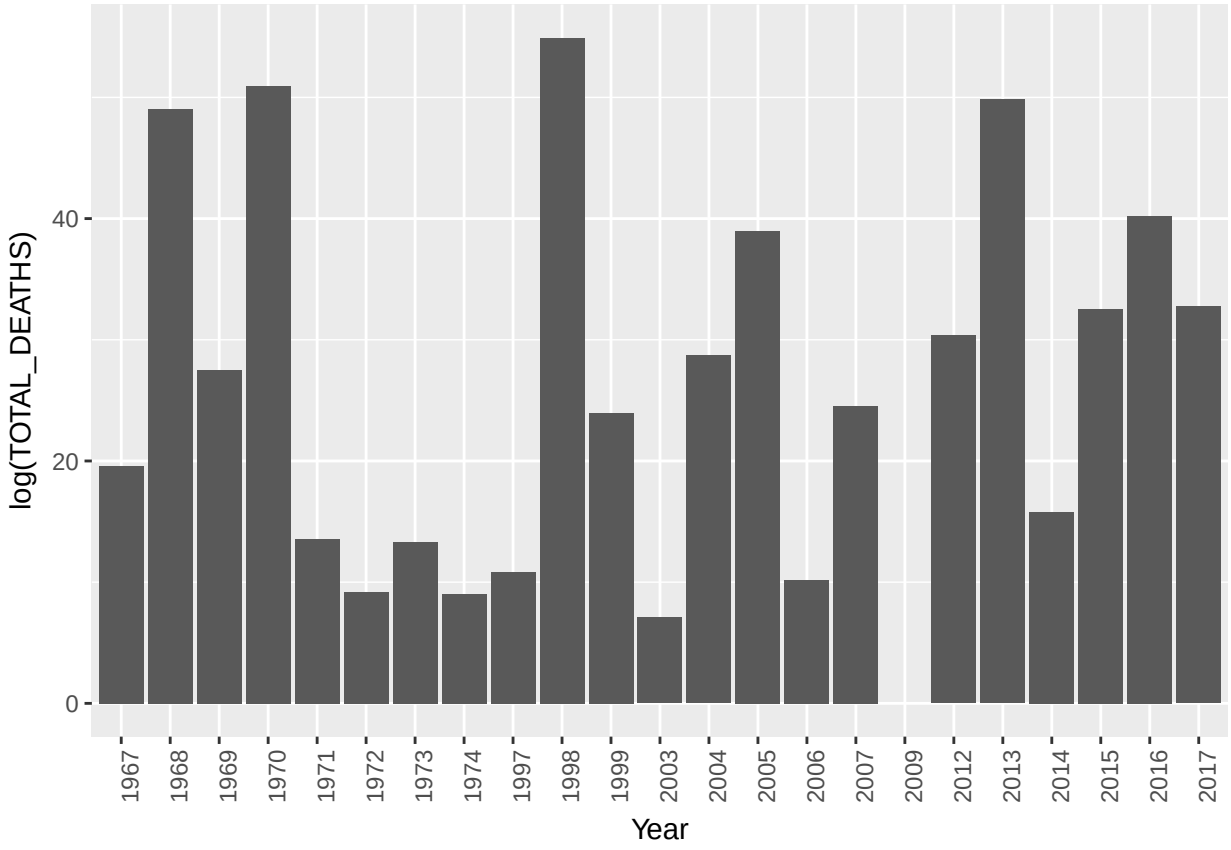
```
ggplot(country_eq,aes(x=log(TOTAL_DAMAGE_MILLIONS_DOLLARS),y=log(TOTAL_DEATHS))) +
  geom_point()
```



Total deaths over the years

There is no decrease of total deaths in large earthquakes over the years. The earthquake data seems to be missing the 1980s data.

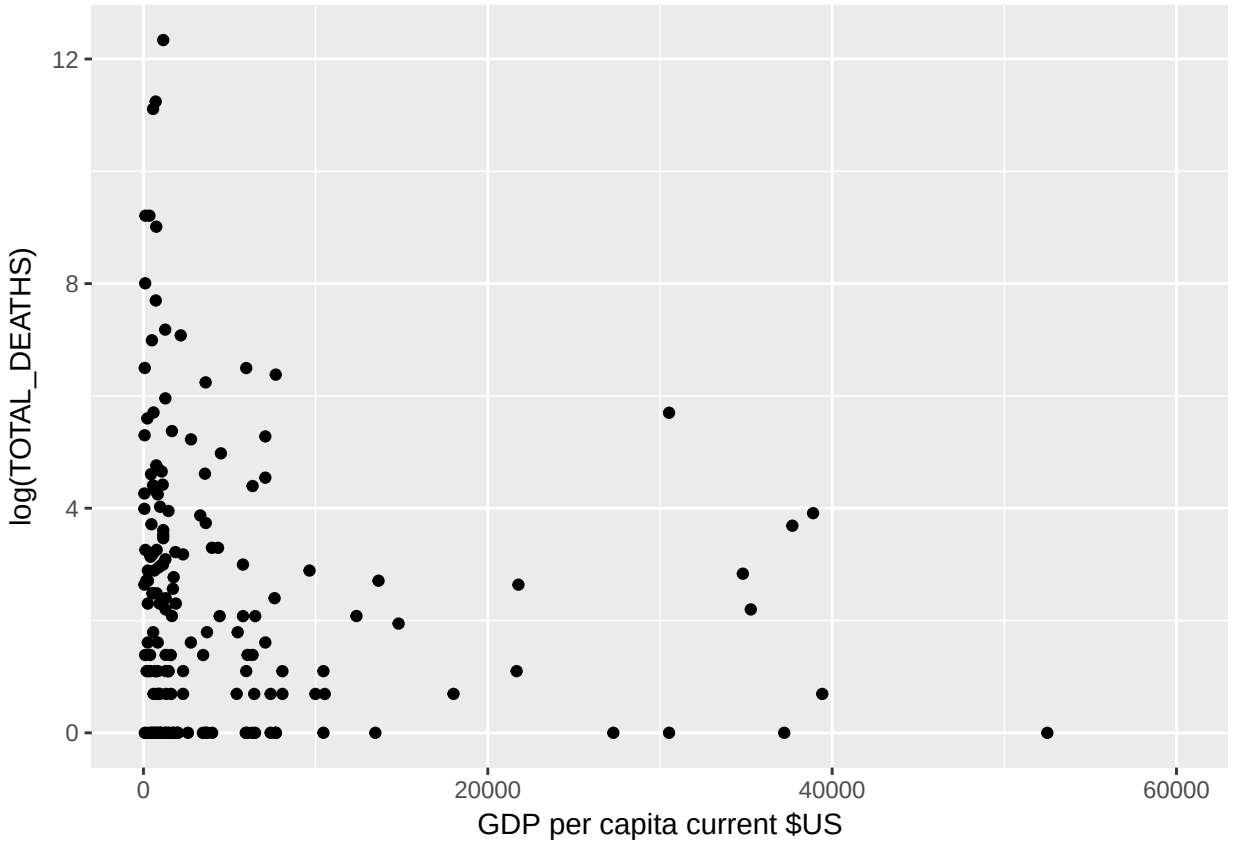
```
ggplot(country_eq, aes(x= format(DATE,"%Y"), y = log(TOTAL_DEATHS))) +
  stat_summary(fun.y=sum,geom="bar") +
  theme(axis.text.x = element_text(angle = 90)) +
  scale_x_discrete("Year")
```



Total deaths VS GDP per capita

The data seems to show that higher death rates are associated with lower GDP per capita.

```
ggplot(df_eq_gdp,aes(x = as.numeric(GDP_per_capita_currentUSD),y=log(TOTAL_DEATHS))) +
  geom_jitter() +
  scale_x_continuous("GDP per capita current $US",limits = c(0,60000))
```



Total deaths VS Country population

```
ggplot(df_eq_gdp,aes(x = as.numeric(Population),y=log(TOTAL_DEATHS))) +
  geom_jitter() +
  scale_x_continuous("Country Population")
```