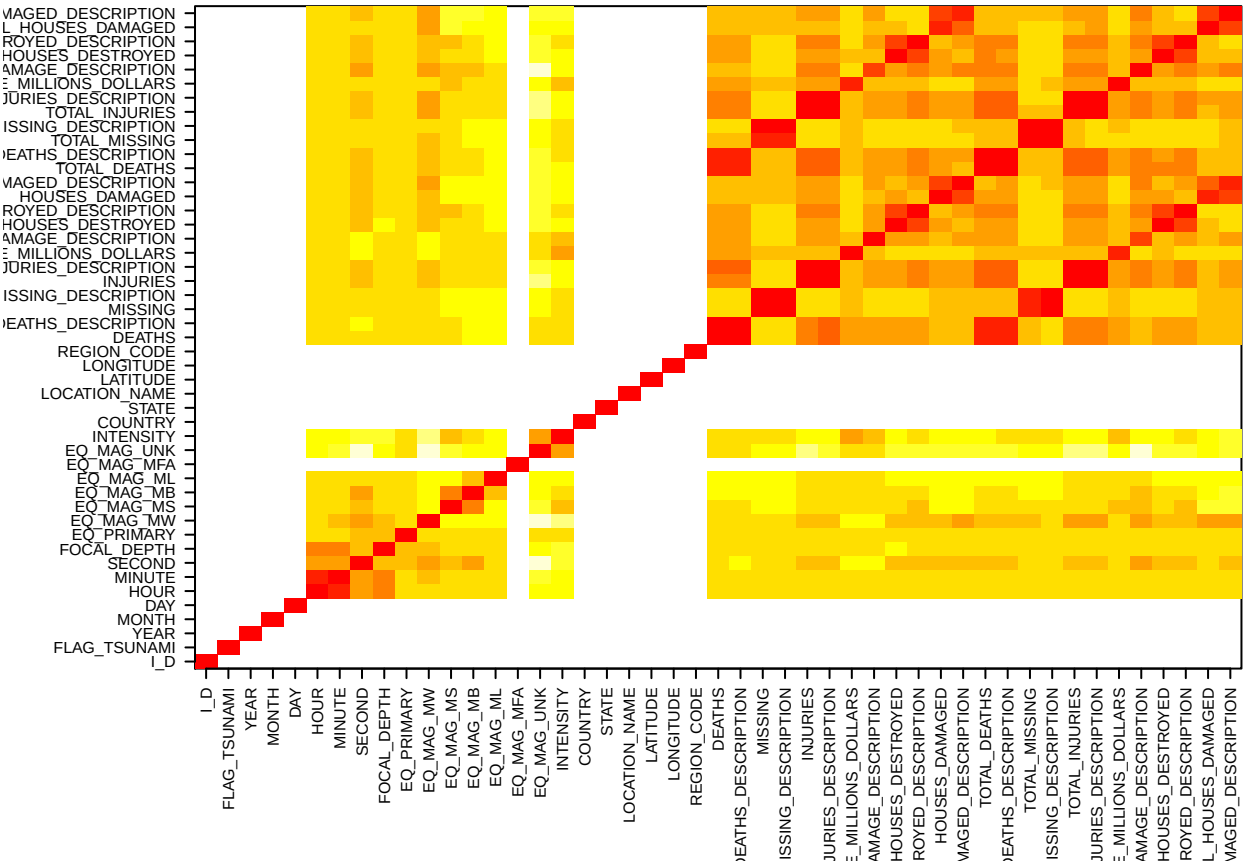
Data Analysis

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



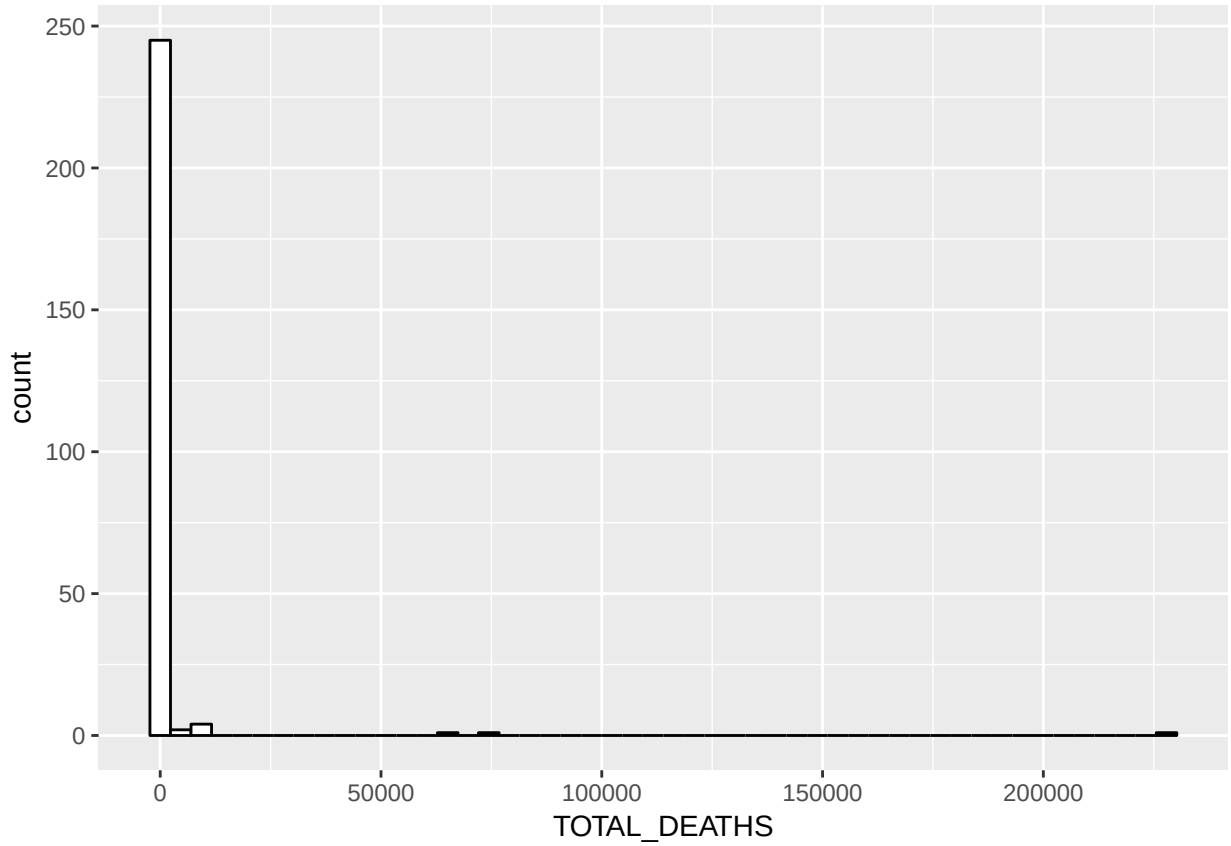

```
# Perform data analysis
gdp_population[gdp_population == ".."] = NA

# Need to make sure the names are consistent
colnames(gdp_population)[1] <- "COUNTRY"
gdp_population$COUNTRY <- toupper(gdp_population$COUNTRY)
country_eq$Year <- format(country_eq$DATE,"%Y")
df_eq_gdp <- left_join(country_eq,gdp_population)

colnames(df_eq_gdp)[48] <- "GDP_constant2010USD"
colnames(df_eq_gdp)[49] <- "GDP_currentUSD"
colnames(df_eq_gdp)[50] <- "GDP_per_capita_2010USD"
colnames(df_eq_gdp)[51] <- "GDP_per_capita_currentUSD"
```

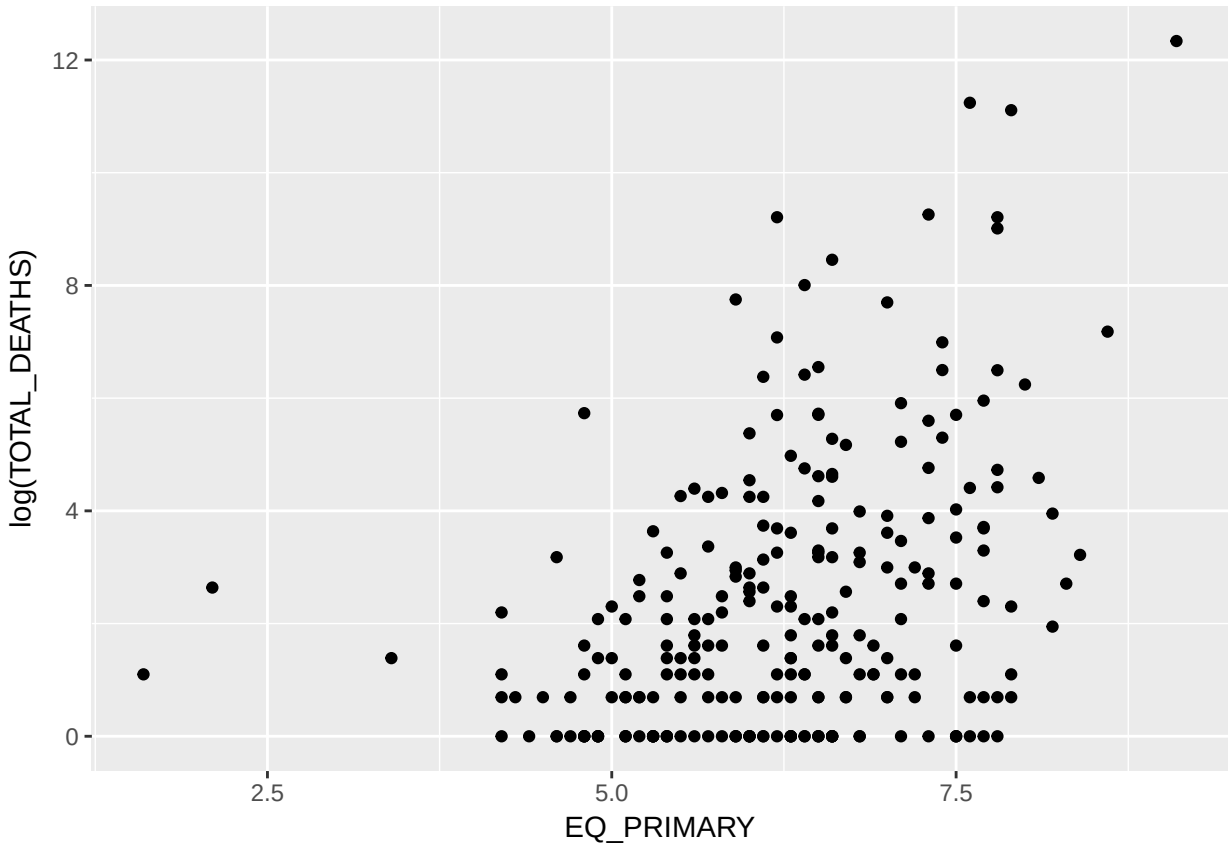
Distrubtion of total deaths

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



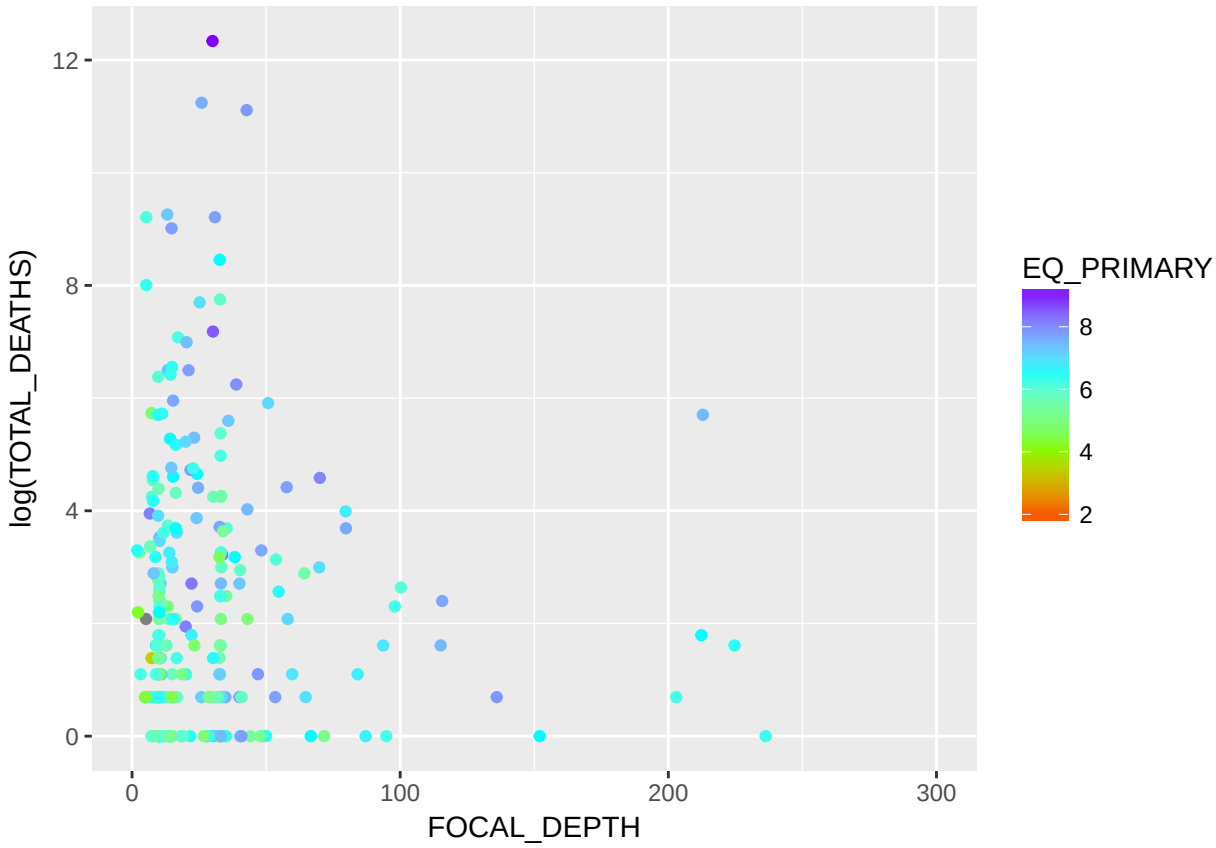
Total deaths VS Earthquake Magnitude

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



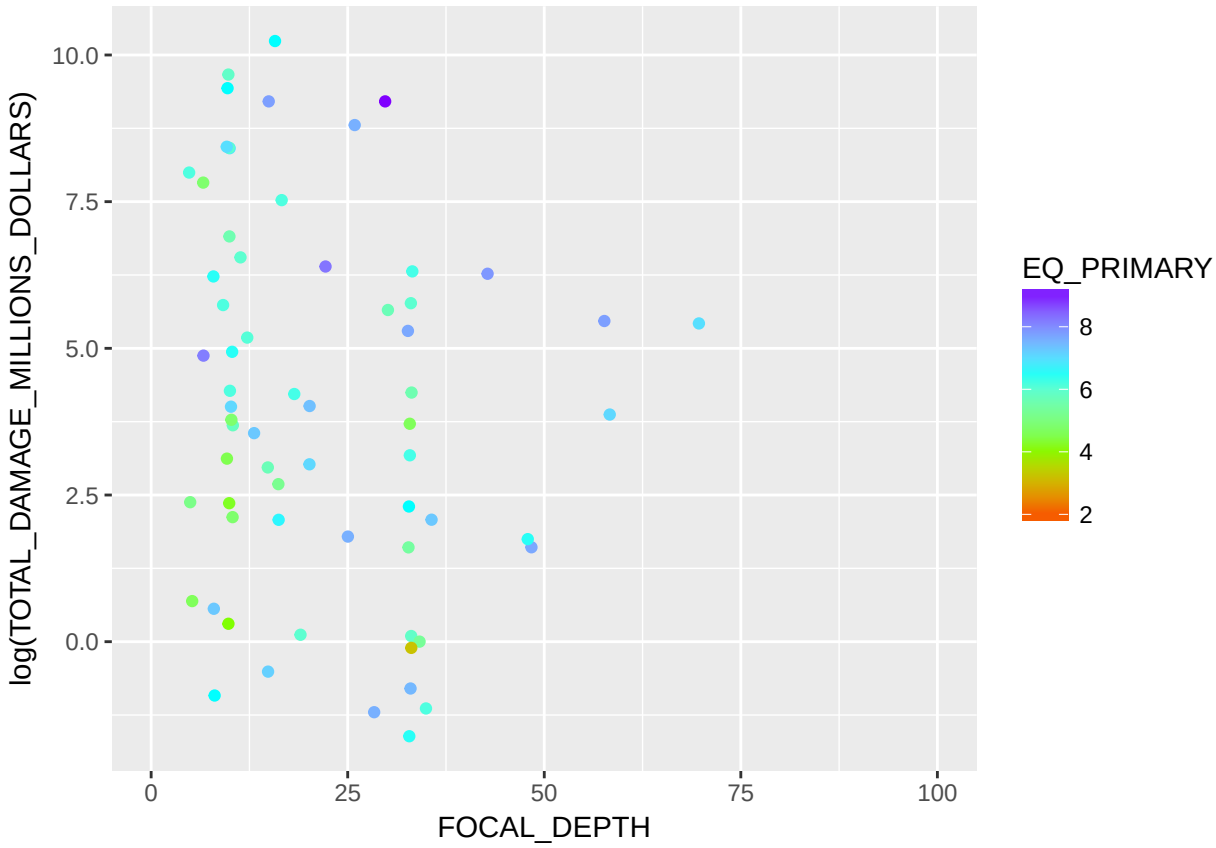
Total deaths VS Focal Depth of earthquakes

```
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))
```



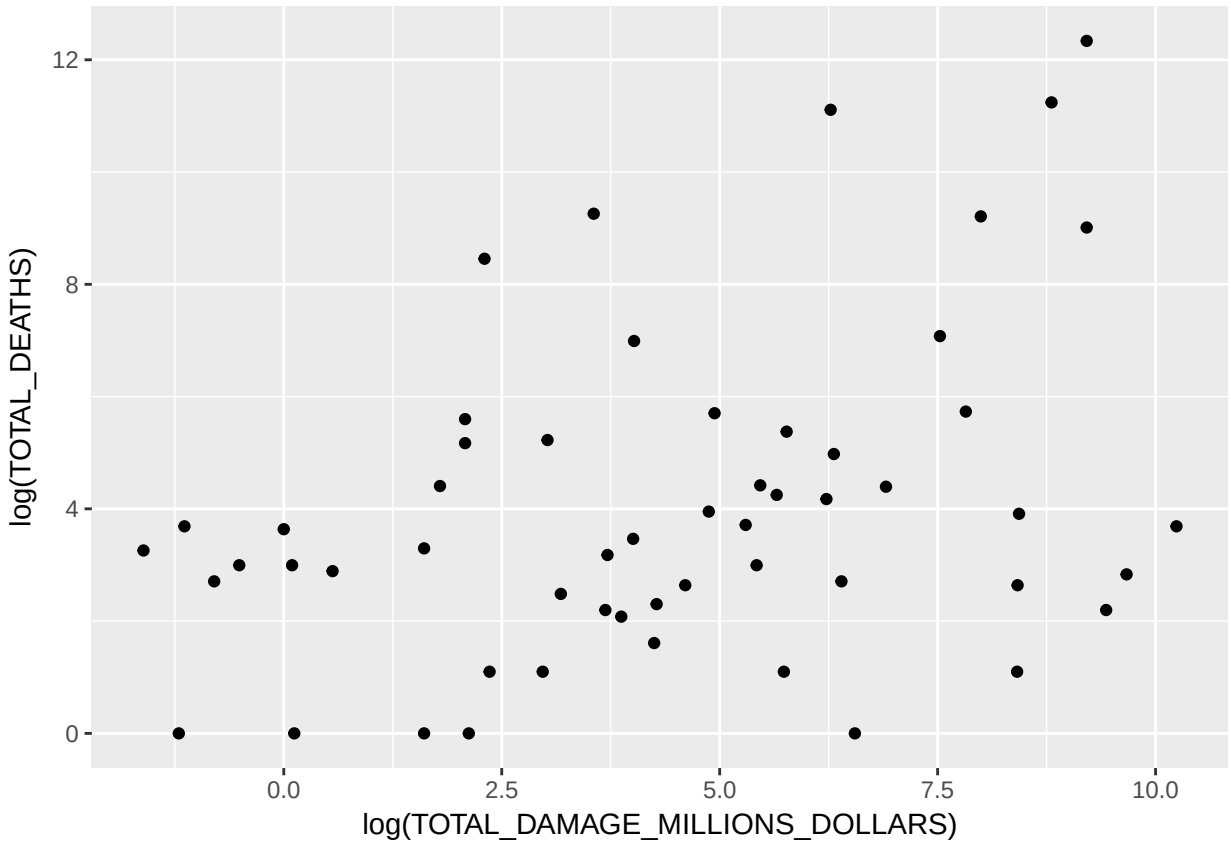
Total damage VS Focal Depth of earthquakes

```
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))
```



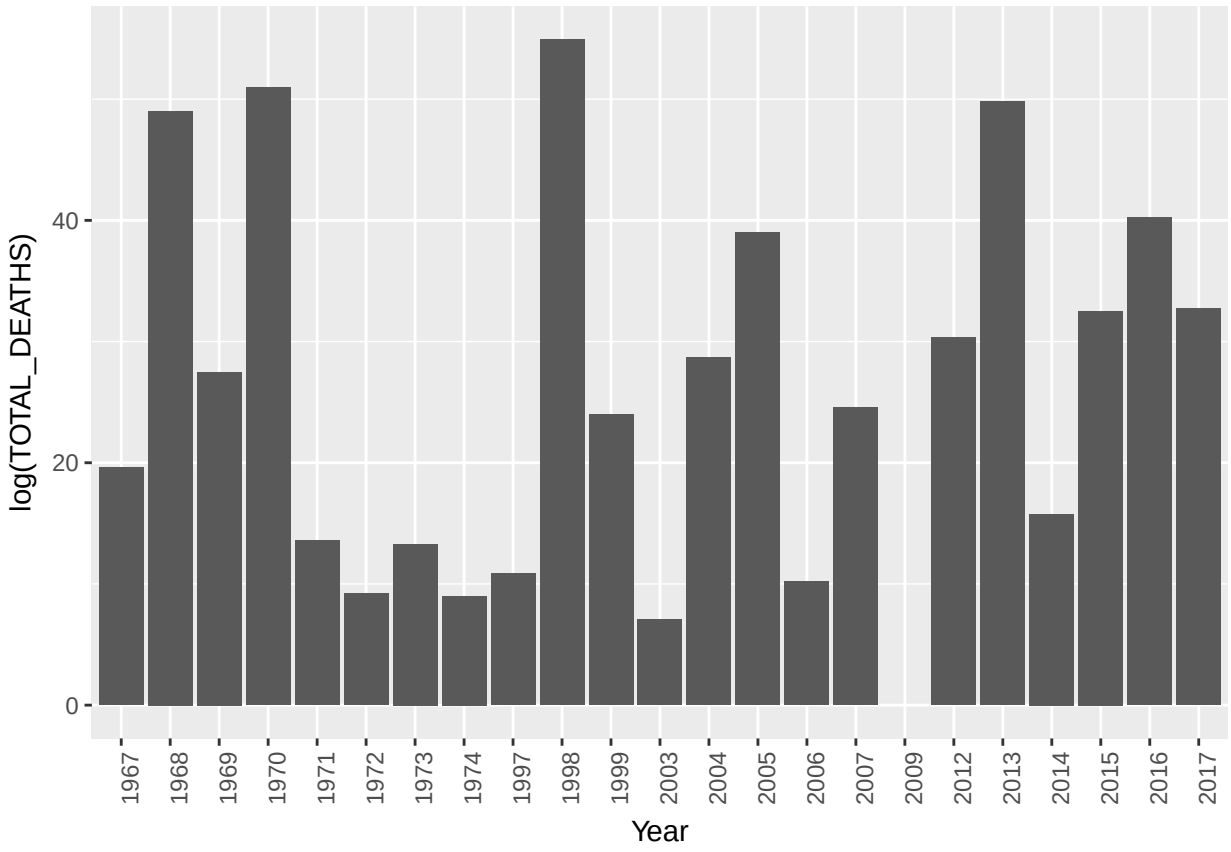
Total Damage VS Total Deaths

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



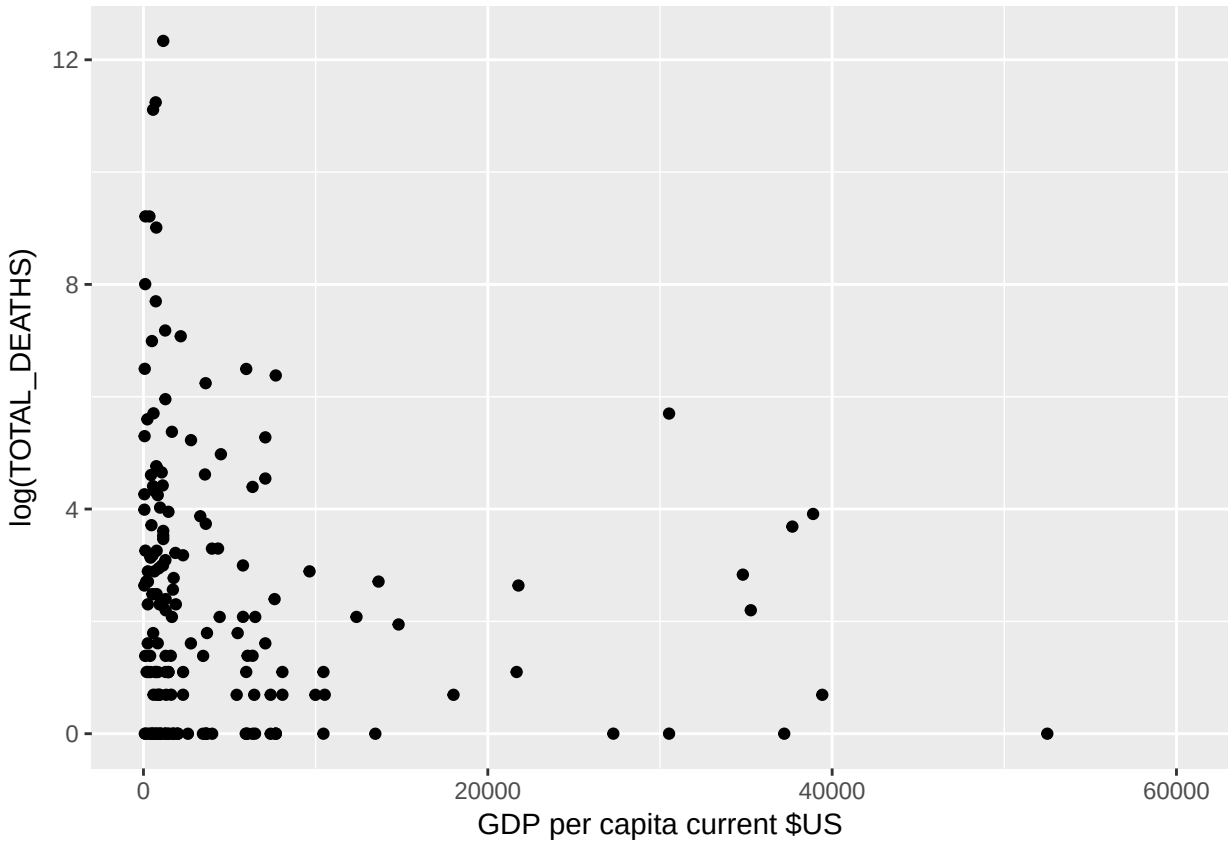
Total deaths over the years

```
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")
```



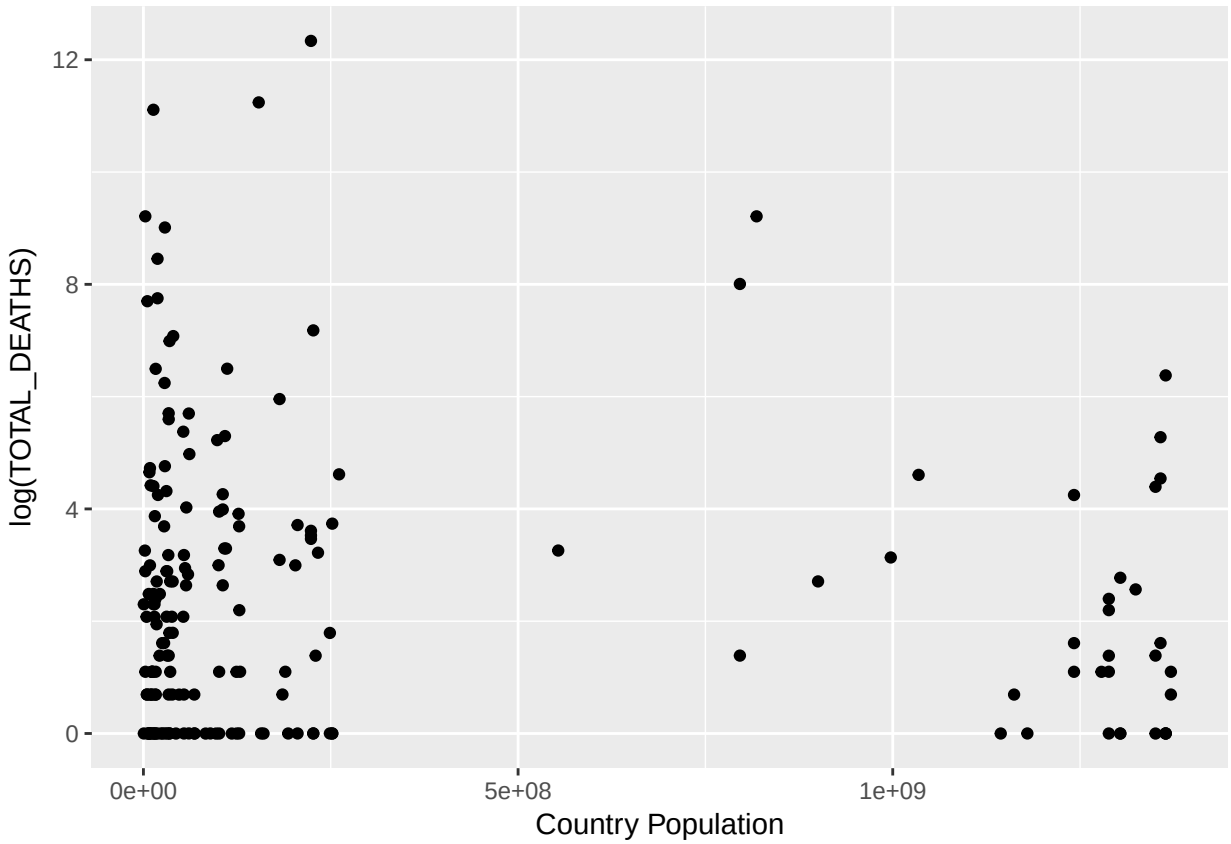
Total deaths VS GDP per capita

```
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")
```


log(TOTAL_DEATHS)
12
8
4
0
0e+00
5e+08
1e+09
Country Population