# Are we doing better at reducing the total loss in large earthquakes?

*Chen Chen*

*12/09/2017*

## *Introduction*

Earthquakes, one of the most damaging natural disasters, take away hundreds to thousands of lives and houses worldwide each year. While it is difficult to predict when the next earthquake will hit, the earthquake prone areas can take actions to be more prepared if such disasters happen. Through this project, I want to analyze earthquakes for the past 50 years to investigate (1) the factors that contributed to high fatalities/damages in large earthquakes and (2) the correlation between large earthquake fatality/damages and a country's economic situation. The questions that I attempt to answer through this project are:

1) overall, are we doing better at reducing damages and fatality over the past 50 years;
2) are the total deaths/damages correlated with the magnitude/depth/location of the earthquakes?
3) is there a correlation between a country's GDP/population and the total deaths/damages;

## *Explanation of the Dataset*

The dataset used in this study is downloaded from NOAA significant earthquake database, which contains damaging earthquakes from 2150 B.C. to the present. Earthquakes between 1967 and 2017 were downloaded for this analysis. This dataset contains important information of the earthquake data, location, depth, magnitude, total deaths, and total damages. A country's economic status for the past 57 years, represented by GDP and population were downloaded from the World Band database.

The difficulty of analyzing the dataset is how to handle the missing values. For instance, 1222 out of 2012 earthquakes do not have the total deaths count, and 1691 out of 2012 earthquakes do not list the total damage values. It is unclear if the empty field represents unavailable data or there were no damage or deaths in these earthquakes. However, considering that the downloaded earthquake data were from the past 50 years when global earthquake recording networks were in place and documentations of these earthquakes were relatively complete , it is very likely that these missing fields represent no deaths or damages.

One limitation of this dataset is the absence the local population density information at each earthquake location. While a country's population is an indication of how dense the population is, the distribution of population varies from cities to cities. In addition, large countries will have great values of population, but the earthquakes may occur along faults that are far away from the residential areas. Another limiation is the lack of information on building code/style at each earthquake location. One would image that areas with stronger buildings will have fewer deaths/damages. After all, earthquakes don't kill people, but buildings do. Because of the lack of information on the local population density and building code, we cannot quantify the correlation between these factors and the total deaths/damages.

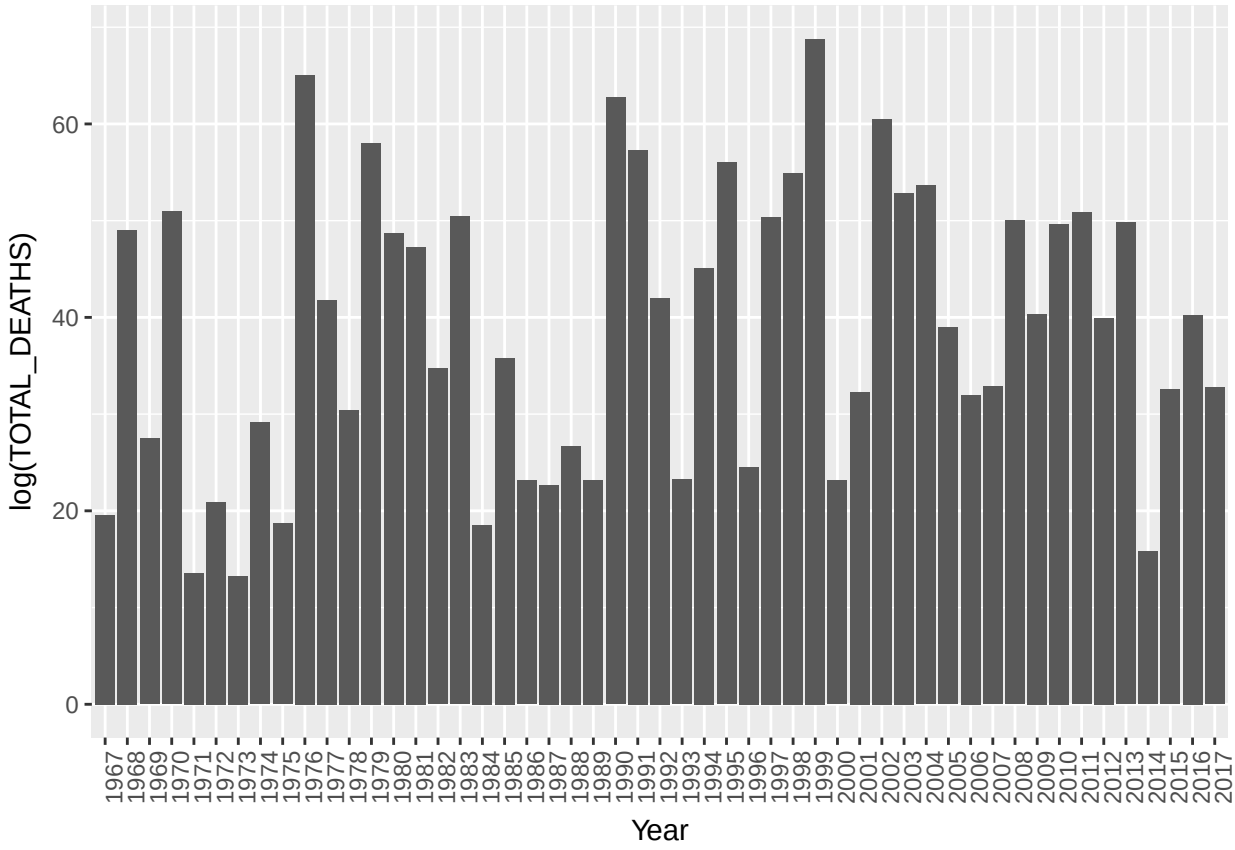
## *Data Cleaning and Wrangling*

For the earthquake data, the date and time need to be converted from characters to date in R. The earthquake data was also re-ordered by countries and variables that were used in the analysis.

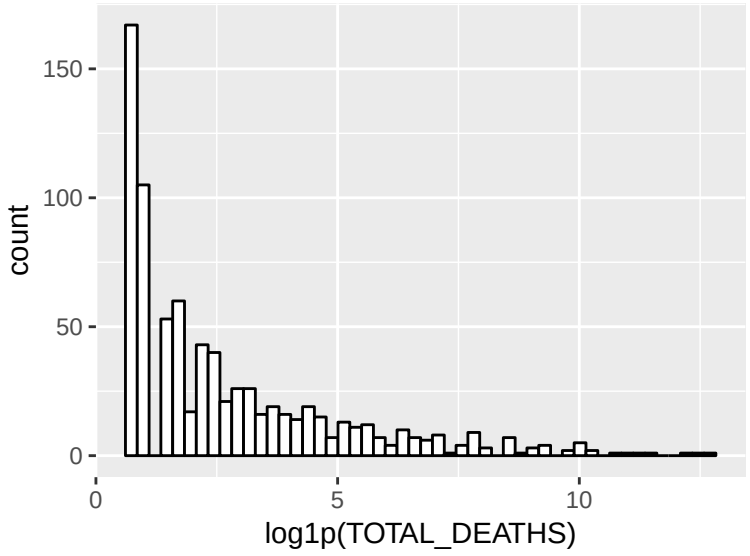
The GDP and population data were collapsed into Year-Value pairs, in order to be joined with with the earthquake data. There are ".." values in the GDP and population data, which were replaced with NA values.

## *Preliminary Results*

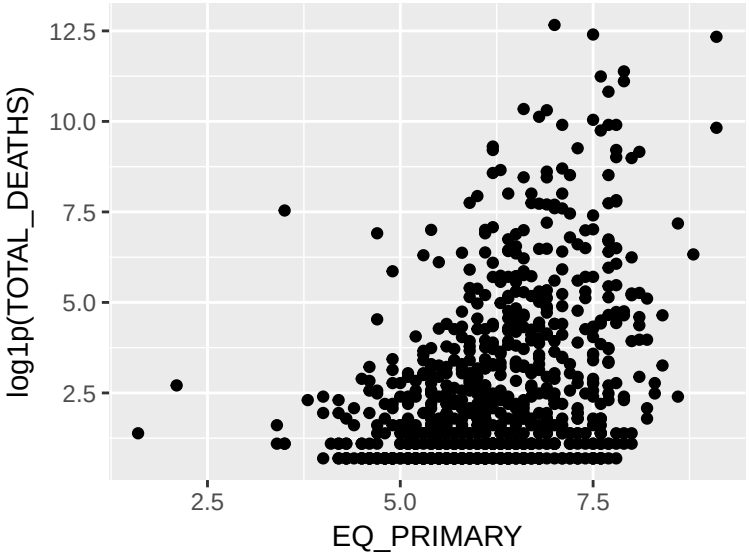
1. Plotting the total deaths in large earthquakes over the past 50 years reveals no decrease in the total deaths in large earthquakes.



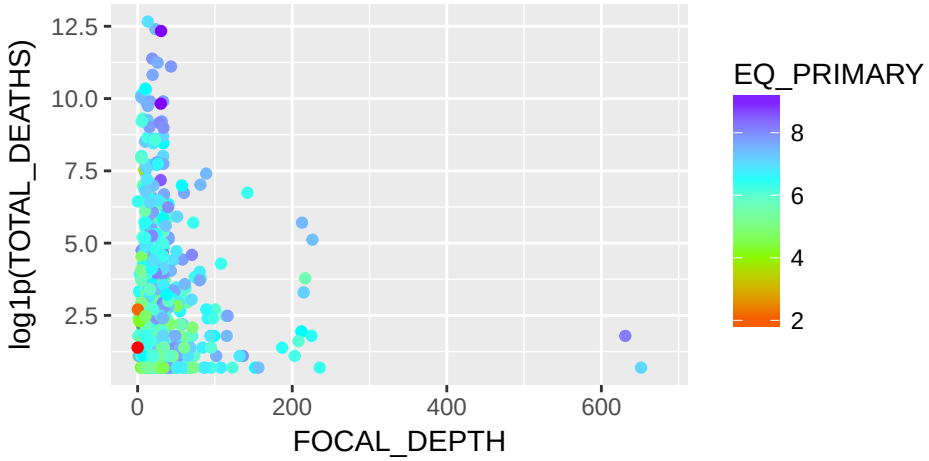
2. Distribution of total deaths exhibits an exponentially decreasing distribution. Only a few earthquakes have caused significant deaths count, while majority of the earthquakes took away fewer lives.



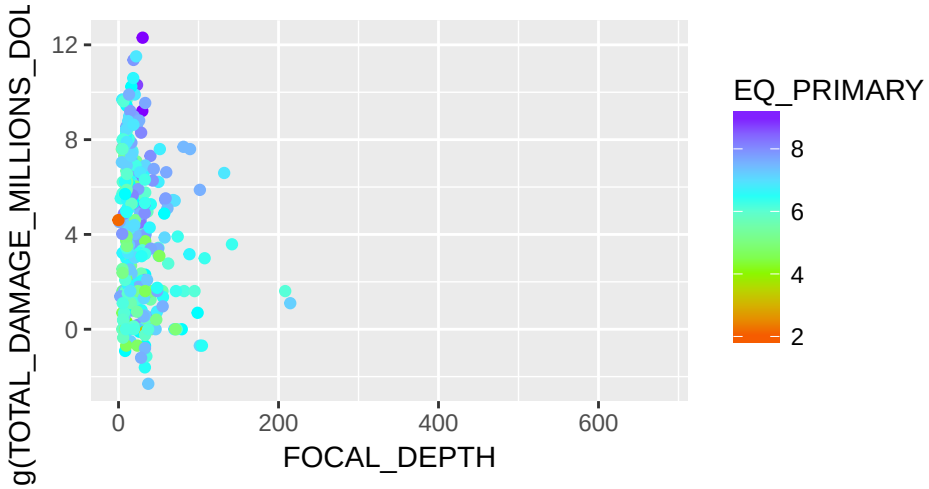

3. There is a positive correlation of earthquake magnitude and the total death in an earthquake. As expected, more people died in large earthquakes.



4. There is a weak correlation between focal depth and total deaths. More deaths are correlated with shallower earthquakes. There is usually no death in very deep earthquakes (> 300 km), except for one earthquake in Peru that occurred in 1970 had 1 death on record, and the 1994 M8.2 earthquake in Bolivia that killed 5 people.




5. There is no obvious correlation between focal depth and total damages. But it is interesting to see that only earthquakes shallow than 100 km depths have caused damages.




Based on these results, I want to perform regression analysis on the total deaths and variables of earthquakes to explore what factors have higher influence on the total deaths. Similarly, regression analysis on the total damages and variables of earthquakes will also be done.

The plots in the preliminary results also show weak correlation between earthquake depth/magnitude with the total death/damages, so the regression analysis may not achieve a good fit for the data. As suggested by my mentor, I also plan to run random forest algorithm to aggregate the results.