Capstone Project Proposal

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

Chen Chen

11/2/2017

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 the correlation between large earthquake fatality/damages and a country's economic situation, specifically answering the followng questions:

1) overall, are we doing better at reducing damages and fatality;
2) which countries have managed to decrease damages/fatality over the years;
3) is there a correlation between a country's GDP and the total deaths/damages;
4) is there a correlation between a country's population and the total deaths/damages;
5) what can we learn from the countries that are able to decrease the total loss at large earthquakes.

Learning from a devastating earthquake is by no means limited to the above aspects. Earthquake engineers and scientists can gather first-hand information on building structures and urban planning flaws by visiting those earthquakes affected areas, which will be very useful for future prevention purposes. This study aims to look into the correlation between several variables and the earthquake damages through data analysis, and the results will be useful in several aspects. For example, insurance companies can use this information to help determine premium, and a country can also refer to this type of information when budgeting natural disaster funding.

Data Collection

Many resources are available online for earthquake data. For instance, the United State Geological Survey updates their earthquake statistics each year, from which we can get an overview of the estimated deaths and the number of large earthquakes for each year (USGS earthquake statistics).

NOAA has a significant earthquake database, which contains earthquakes from 2150 B.C. to the present that meet at least one of the following criteria: Moderate damage (approximately $1 million or more), 10 or more deaths, Magnitude 7.5 or greater, Modified Mercalli Intensity X or greater, or the earthquake generated a tsunami.

I downloaded the earthquakes between 1967 to 2017 from this database, and the data format is .tsv file. Take a quick glance of the data in R:

```
eq_data <- read.table(file = 'results.tsv', sep = '\t', header = TRUE, stringsAsFactors = F)
head(eq_data,2)
```

```
##    I_D FLAG_TSUNAMI YEAR MONTH DAY HOUR MINUTE SECOND FOCAL_DEPTH
## 1 4387              1967     1   4    5     58     NA          10
```

```
## 2 4388              1967     1   5    0     14     NA          35
##   EQ_PRIMARY EQ_MAG_MW EQ_MAG_MS EQ_MAG_MB EQ_MAG_ML EQ_MAG_MFA EQ_MAG_UNK
## 1        6.0        NA        NA        NA        NA         NA        6.0
## 2        7.5        NA        NA        NA        NA         NA        7.5
##   INTENSITY  COUNTRY STATE LOCATION_NAME LATITUDE LONGITUDE REGION_CODE
## 1        NA   GREECE              GREECE     38.4      22.0         130
## 2        10 MONGOLIA            MONGOLIA     48.1     102.8          40
##   DEATHS DEATHS_DESCRIPTION MISSING MISSING_DESCRIPTION INJURIES
## 1     NA                 NA      NA                  NA       NA
## 2     NA                 NA      NA                  NA       NA
##   INJURIES_DESCRIPTION DAMAGE_MILLIONS_DOLLARS DAMAGE_DESCRIPTION
## 1                   NA                      NA                  2
## 2                   NA                      NA                  1
##   HOUSES_DESTROYED HOUSES_DESTROYED_DESCRIPTION HOUSES_DAMAGED
## 1               NA                           NA             NA
## 2               NA                           NA             NA
##   HOUSES_DAMAGED_DESCRIPTION TOTAL_DEATHS TOTAL_DEATHS_DESCRIPTION
## 1                         NA           NA                       NA
## 2                         NA           NA                       NA
##   TOTAL_MISSING TOTAL_MISSING_DESCRIPTION TOTAL_INJURIES
## 1            NA                        NA             NA
## 2            NA                        NA             NA
##   TOTAL_INJURIES_DESCRIPTION TOTAL_DAMAGE_MILLIONS_DOLLARS
## 1                         NA                            NA
## 2                         NA                            NA
##   TOTAL_DAMAGE_DESCRIPTION TOTAL_HOUSES_DESTROYED
## 1                       NA                    NA
## 2                       NA                    NA
##   TOTAL_HOUSES_DESTROYED_DESCRIPTION TOTAL_HOUSES_DAMAGED
## 1                                 NA                   NA
## 2                                 NA                   NA
##   TOTAL_HOUSES_DAMAGED_DESCRIPTION
## 1                               NA
## 2                               NA
```

We only need a few columns for this analysis, so the data frame needs to be cleaned up in the analysis.

We also need data about a country's economic status. I found GDP and population data of all the countries for the past 57 years from the World Bank database.

Take a look at the GDP data and we notice that we will need to clean up the header and replace empty cells with NA:

```
world_gdp <- read.csv(file='world_gdp_Data.csv',sep=",",header=T,stringsAsFactors = F)
head(world_gdp,2)
```

```
##         Series.Name    Series.Code Country.Name Country.Code
## 1 GDP (current US$) NY.GDP.MKTP.CD  Afghanistan          AFG
## 2 GDP (current US$) NY.GDP.MKTP.CD      Albania          ALB
##     X1960..YR1960.   X1961..YR1961.   X1962..YR1962.   X1963..YR1963.
## 1 537777811.111111 548888895.555556 546666677.777778 751111191.111111
## 2               ..               ..               ..               ..
##     X1964..YR1964.   X1965..YR1965.   X1966..YR1966.   X1967..YR1967.
## 1 800000044.444444 1006666637.77778 1399999966.66667 1673333417.77778
## 2               ..               ..               ..               ..
##     X1968..YR1968.   X1969..YR1969.   X1970..YR1970.   X1971..YR1971.
```

```
## 1 1373333366.66667 1408888922.22222 1748886595.55556 1831108971.11111
## 2               ..                ..                ..               ..
##    X1972..YR1972.   X1973..YR1973.   X1974..YR1974.   X1975..YR1975.
## 1 1595555475.55556 1733333264.44444 2155555497.77778 2366666615.55556
## 2               ..                ..                ..               ..
##    X1976..YR1976.   X1977..YR1977.   X1978..YR1978.   X1979..YR1979.
## 1 2555555566.66667 2953333417.77778 3300000108.88889 3697940409.61098
## 2               ..                ..                ..               ..
##    X1980..YR1980.   X1981..YR1981. X1982..YR1982. X1983..YR1983.
## 1 3641723321.99546 3478787909.09091             ..             ..
## 2               ..                ..                ..             ..
##    X1984..YR1984.  X1985..YR1985. X1986..YR1986. X1987..YR1987.
## 1               ..              ..              ..             ..
## 2 1924242453.00793 1965384586.2409   2173750012.5     2156624900
##   X1988..YR1988. X1989..YR1989. X1990..YR1990.   X1991..YR1991.
## 1             ..             ..             ..               ..
## 2     2126000000   2335124987.5   2101624962.5 1139166645.83333
##    X1992..YR1992.   X1993..YR1993.   X1994..YR1994.   X1995..YR1995.
## 1               ..                ..                ..               ..
## 2 709452583.880319 1228071037.84446 1985673798.10258 2424499009.14264
##    X1996..YR1996.   X1997..YR1997.   X1998..YR1998.   X1999..YR1999.
## 1               ..                ..                ..               ..
## 2 3314898291.75235 2359903108.38446 2707123772.16195 3414760915.27878
##    X2000..YR2000.   X2001..YR2001.   X2002..YR2002.   X2003..YR2003.
## 1               .. 2461665937.89386 4128820723.04713 4583644246.48061
## 2 3632043907.97733 4060758804.12084 4435078647.74817 5746945912.58082
##    X2004..YR2004.   X2005..YR2005.   X2006..YR2006.   X2007..YR2007.
## 1 5285465685.86423 6275073571.54659 7057598406.61553 9843842455.48323
## 2  7314865175.6199 8158548716.68554  8992642348.7871 10701011896.7708
##    X2008..YR2008.   X2009..YR2009.   X2010..YR2010.   X2011..YR2011.
## 1 10190529882.4878 12486943505.7381 15936800636.2487 17930239399.8149
## 2 12881352687.7773 12044212903.8168  11926953258.916 12890867538.5302
##    X2012..YR2012.   X2013..YR2013.   X2014..YR2014.   X2015..YR2015.
## 1 20536542736.7297 20046334303.9661 20050189881.6659 19702986340.5494
## 2 12319784787.2987 12781029643.5936 13219857459.1009 11390365293.8057
##    X2016..YR2016.
## 1 19469022207.6357
## 2 11926892452.8499
```

I will also need to clean up the population data in a similar fashion.

Proposed Analyses

Data Wrangling using dplyr package in R

As mentioned in the data in the above section, I will create a subset data frame from the significant earthquakes data frame with only the columns needed in the analysis. I also need to replace empty cells in the GDP and population data with NA and clean up the header information.

After these three data frames are cleaned up, I plan to add GDP and population data as columns of variables to the earthquake data frame, so all the data are kept in one place.

Preliminary inspection of the data

I plan to make the following plots:

- total deaths by year for the past 50 years;
- total deaths by country;
- total deaths by earthquake magnitude;
- total deaths VS GDP;
- total deaths VS population.

A series of plots with total damages will also be made.

Futher analysis based on the results of the preliminary findings

Depending on what correlation I can find from the previous step, I will design additional analyses or plots.

Recommendations based on my findings

If I find certain countries are doing better than the others, I can do some literature research to find what actions these countries have taken. And these actions can be applied in other countries too.

Additional aspects to think from the findings include: how they can be utilized by insurance companies, urban planning department, or the UN country assistance development frameworks (UNDAFs).