

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

Chen Chen

January 10, 2018

Questions

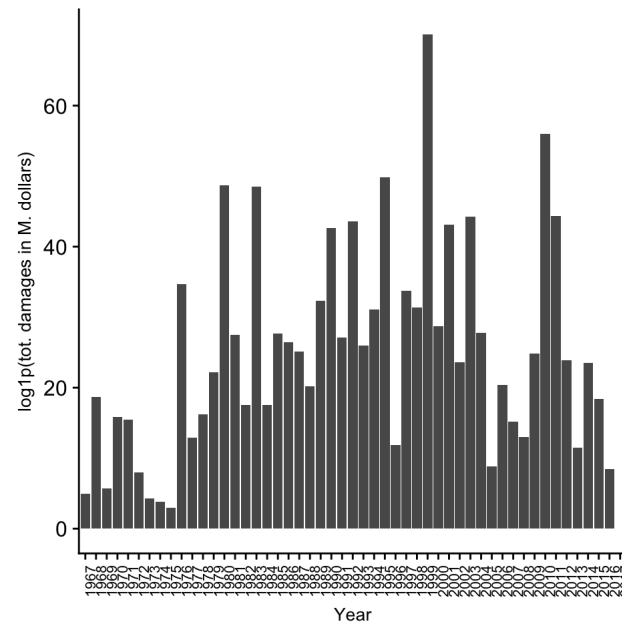
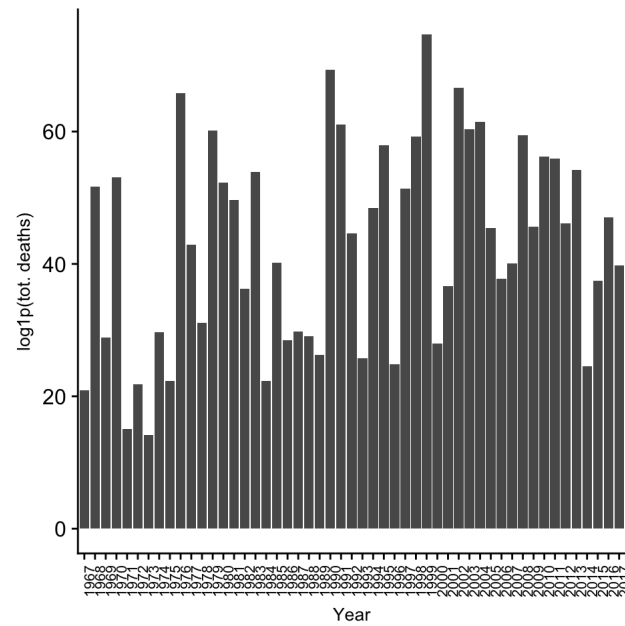
1. Overall, are we doing better at reducing damages and fatality over the past 50 years?
2. Are the total deaths/damages correlated with the magnitude/depth/location of the earthquakes?
3. Is there a correlation between a country's GDP/population and the total deaths/damages?

Dataset

- [Significant earthquake database](#) from NOAA: damaging earthquakes from 2150 B.C. to present.
- Earthquakes between 1967 and 2017 were downloaded, containing information on the earthquake data, location, depth, magnitude, total deaths, total damages, etc.
- World Bank Database on Country's [GDP](#) and [population](#)

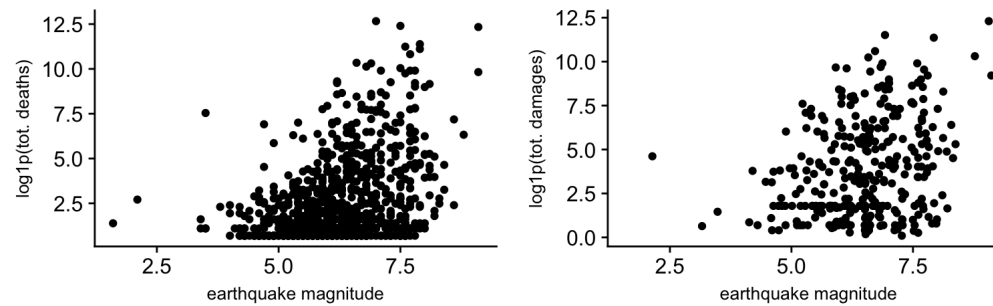
Visualization of the Data

- No decrease in the total deaths/damages in large earthquakes

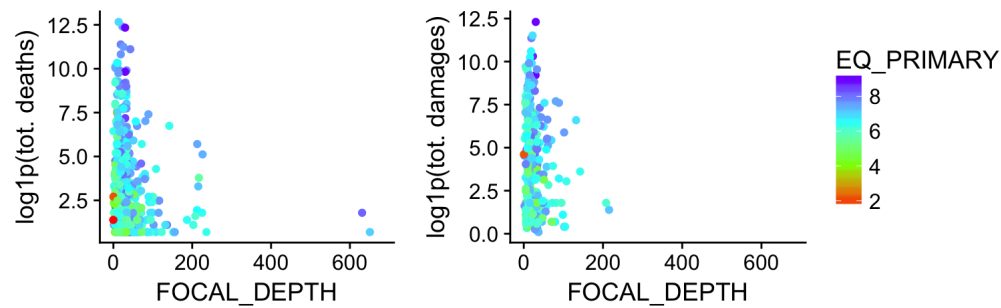


Visualization of the Data

- Positive correlation of earthquake magnitude and the total deaths/damages in an earthquake.

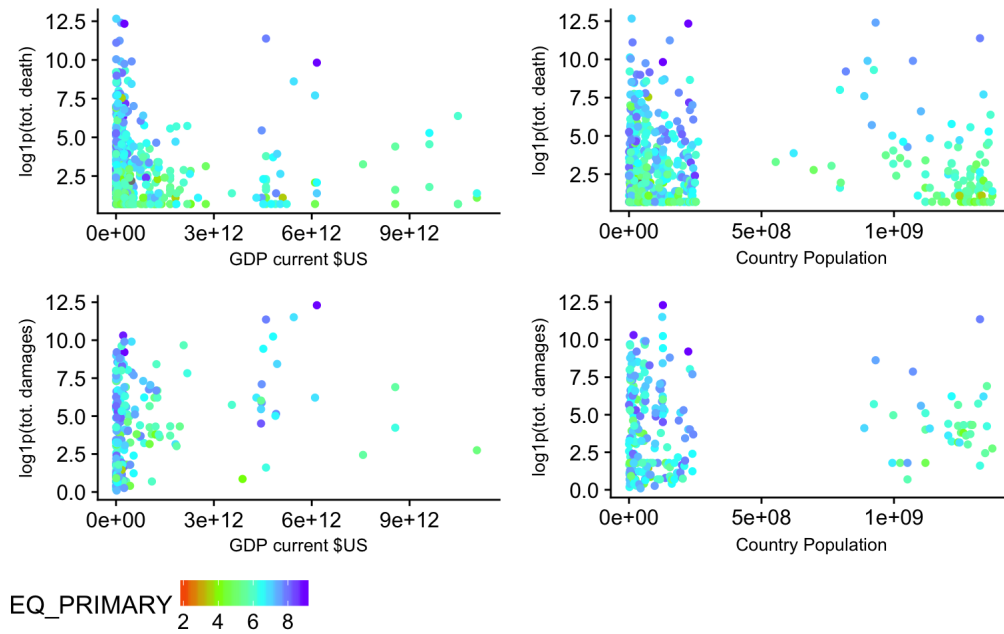


- More deaths/damages are correlated with shallower earthquakes.



Visualization of the Data

- No clear correlation between the total loss and a country's population.
- It appears that countries with higher GDP values suffered fewer deaths and less damages.



Machine Learning Methods - Linear Regression Analysis

- High residual and low R-squared value in the summary of this model suggest a bad fit.

```
##
## Call:
## lm(formula = TOTAL_DEATHS ~ FOCAL_DEPTH + EQ_PRIMARY + INTENSITY +
##     LATITUDE + LONGITUDE, data = df_eq_gdp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12574  -5071  -2347    822 229070
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -29396.051   9095.856  -3.232  0.00139 **
## FOCAL_DEPTH    -4.789     38.116  -0.126  0.90011
## EQ_PRIMARY    2579.113   1381.193   1.867  0.06302 .
## INTENSITY    1872.967    786.637   2.381  0.01801 *
## LATITUDE       50.497     52.875   0.955  0.34049
## LONGITUDE     10.689     13.159   0.812  0.41737
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17410 on 251 degrees of freedom
## (1755 observations deleted due to missingness)
## Multiple R-squared:  0.06531,    Adjusted R-squared:  0.04669
## F-statistic: 3.508 on 5 and 251 DF,  p-value: 0.004401
```

Machine Learning Methods - Linear Regression Analysis

- Including country's GDP and population does not provide a good fit model.

```
##
## Call:
## lm(formula = TOTAL_DEATHS ~ FOCAL_DEPTH + EQ_PRIMARY + INTENSITY +
##     LATITUDE + LONGITUDE + GDP_currentUSD + Population, data = df_eq_gdp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22102  -5239  -1374   1977  219126
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.603e+04  1.157e+04  -3.114  0.00214 **
## FOCAL_DEPTH    2.414e+01  6.273e+01   0.385  0.70083
## EQ_PRIMARY     3.053e+03  1.727e+03   1.768  0.07867 .
## INTENSITY      2.047e+03  1.006e+03   2.035  0.04330 *
## LATITUDE       3.940e+01  6.644e+01   0.593  0.55390
## LONGITUDE     -9.284e+00  1.778e+01  -0.522  0.60217
## GDP_currentUSD  2.611e-10  1.753e-09   0.149  0.88177
## Population     1.418e-05  4.637e-06   3.058  0.00256 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19070 on 186 degrees of freedom
## (1818 observations deleted due to missingness)
## Multiple R-squared:  0.1173, Adjusted R-squared:  0.08412
## F-statistic: 3.532 on 7 and 186 DF, p-value: 0.001382
```


Machine Learning Methods - Random Forest

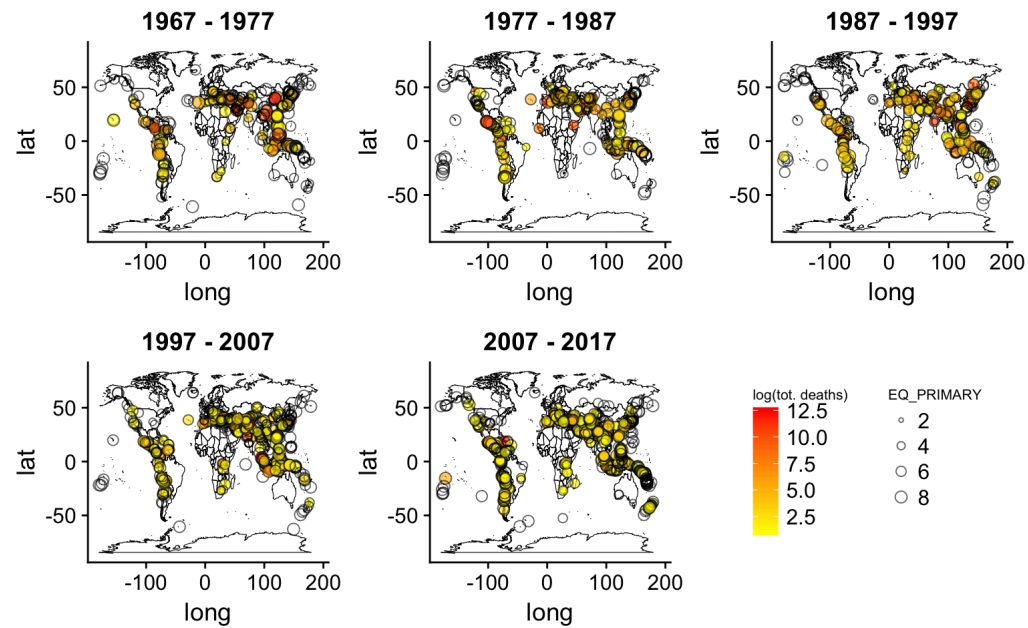
- Large mean of squared residuals and the negative value for the % of variance indicates that random forest method was not adequate to define a good fit model.

```
##          %IncMSE IncNodePurity
## FOCAL_DEPTH -0.215833    2477988429
## EQ_PRIMARY  6.183413    4834659130
## INTENSITY   4.422953    27634649725
## LATITUDE    3.690651    5535607345
## LONGITUDE   3.467354    5643303160
## GDP_currentUSD 4.940315    6390495863
## Population  6.871774    12252170534
```

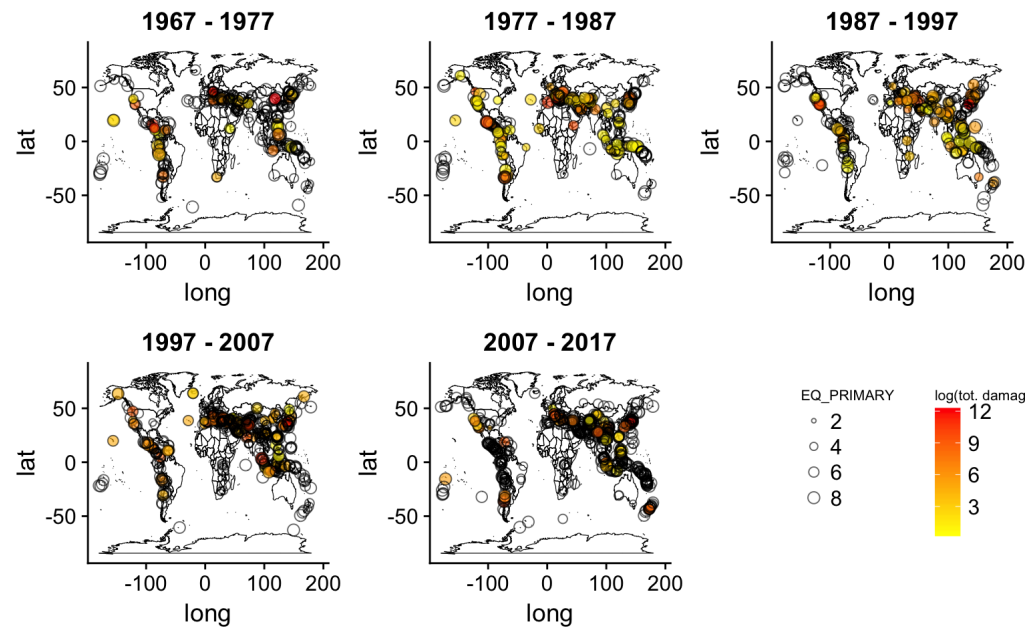
```
##
## Call:
## randomForest(formula = TOTAL_DEATHS ~ FOCAL_DEPTH + EQ_PRIMARY + INTENSITY + LATITUDE + LONGITUDE + GDP_currentUSD + Population, data = df_eq_gdp, imp
##           Type of random forest: regression
##           Number of trees: 2000
## No. of variables tried at each split: 2
##
##           Mean of squared residuals: 408750305
##           % Var explained: -3.45
```

Map the Earthquakes

- Total deaths and total damages was plotted every 10 years, and at certain regions (the Mediterranean and South America), the total loss has decreased over the years.



Map the Earthquakes



Results

- The overall total loss in large earthquakes did not decrease over the years, but the loss seemed to decrease at certain regions.
- Large earthquakes and shallow earthquakes are correlated with more damages and deaths.
- These correlation, however, are not linear, as evidenced by the poor results from the linear regression analysis.
- Preliminary inspection seems to show that rich countries (higher GDP) have less loss in earthquakes.
- The correlation between GDP and loss in earthquakes requires further investigation, because earthquakes only struck certain areas while national GDP evaluates the economic status of an entire country.

Implications

- Results will be useful for the UN to budget emergency fund for earthquake-prone areas.
- E.g., Areas that suffer great loss during the past 50 years and also have high risk for earthquakes in the future, a greater amount of emergency fund should be budgeted.
- The analyses can be furthered to study similar correlations in individual country, state or even city.
- Using these results to identify countries, states and cities that have successfully reduced the total loss over the years and apply their strategies to other regions.