

EDA and Visualization of Earthquake Occurrence in Indonesia over the last 20 years using R



(<https://www.mapsofworld.com>)

Introduction

Indonesia is a country with a high possibilities of natural disasters such as volcanic eruptions, forest fires, floods, and, most notably, earthquakes. Every day, I received an earthquake notification from my phone's Meteorological, Climatological, and Geophysical agency application. Not just one, but sometimes two or three notifications appeared in one day, informing me that there has been an earthquake in one of Indonesia's regions. This inspired me to create an EDA (Exploratory Data Analysis) of how many earthquakes occurred in the Indonesia region over the last 20 years.

Earthquakes are known as one of the most destructive natural disasters, capable of destroying nearby towns in a split second with no warning. A Tectonic earthquake is one of the natural phenomenon caused by the movement

of the earth's crust. A Tectonic earthquake is the most frequent type of earthquake^[1]. Earthquake power can be recorded and measured as the earthquake magnitude. The magnitude can be recorded using seismograph with the output of earthquake power in Richter scale^[2].

The Indonesian archipelago is located on the tectonic zone where the Pacific, Eurasian, and Indo-Australian plates collide. Because Indonesia is at the center of a complex tectonic zone, earthquakes occur almost every day more than three times^[3]. According to data from the USGS (United States Geological Survey) website, 90 % of earthquakes, including the largest, occurred along the Ring of Fire region^[4]. Surprisingly, Indonesia is located above the ring of fire and the complex tectonic zone.

The EDA and data visualization can be used to visualize the number of earthquake occurrences. In this article, I will use the R programming language to perform EDA and visualize data to obtain specific information. The dataset was gathered from USGS website, which collects data on earthquakes that occur around the world^[5]. We will be using earthquake data from the years 2000 to 2020.

EDA and Visualization

Before we begin the coding section, we must import specific library that we will use later on.

```
#Importing specific library
```

```
library(dplyr)
```

```
library(tidyr)
```

```
library(lubridate)
```

```
library(ggplot2)
```

```
library(plotly)
```

```
#Load the dataset
```

```
df_eq1 <- read.csv("~/Rstudio/Projects/EDA Earthquake/Dataset/Raw/EQ 2000-2004.csv")
```

```
df_eq2 <- read.csv("~/Rstudio/Projects/EDA Earthquake/Dataset/Raw/EQ 2005-2010.csv")
```

```
df_eq3 <- read.csv("~/Rstudio/Projects/EDA Earthquake/Dataset/Raw/EQ 2011-2015.csv")
```

```
df_eq4 <- read.csv("~/Rstudio/Projects/EDA Earthquake/Dataset/Raw/EQ 2016-2020.csv")
```

```
#Combine the dataset
df_eq <- rbind(df_eq4,df_eq3,df_eq2,df_eq1)
df_eq
```

The majority of the data in the dataset comes from earthquakes recorded in ASEAN countries. We must specify the coordinates of the Indonesian territory so that the data can only be filled with earthquakes occurred in the Indonesian region.

```
#Specified earthquake's coordinate happened in Indonesia territory
eq_INA = df_eq %>%
  filter(longitude>=93 & longitude<=141.25,latitude>=-15 & latitude<=9)
eq_INA
```

After we specify the data, we must classify the data of earthquakes based on their magnitude.

Based on its magnitude richter scale, the earthquake is classified into eight categories[6].

1. Magnitude less than 2: Micro Earthquake
2. Magnitude 2 to 3.9: Minor Earthquake
3. Magnitude 4 to 4.9: Light Earthquake
4. Magnitude 5 to 5.9: Moderate Earthquake
5. Magnitude 6 to 6.9: Strong Earthquake
6. Magnitude 7 to 7.9: Major Earthquake
7. Magnitude 8 to 9.9: Great Earthquake
8. Magnitude more than 10: Epic Earthquake

This analysis only includes earthquakes with magnitudes greater than 2 on the Richter scale.

```
#Classifying earthquake based on magnitude richter scale
eq_INA = eq_INA %>%
  mutate(mag_class = factor(ifelse(mag>=2 & mag<=4,"minor",ifelse(mag>=4 &
mag<=5,"light",ifelse(mag>=5 & mag<=6,"moderate",ifelse(mag>=6 &
mag<=7,"strong",ifelse(mag>=7 & mag<=8,"major","great"))))))))
```

Now that we get the proper data frame, we started with the idea to count the number of earthquakes based on its level of magnitude and year.

```
eq_INA %>%
  group_by(mag_class) %>%
  summarise(number_of_earthquakes = n()) %>%
  ggplot(aes(x = mag_class, y = number_of_earthquakes)) +
  geom_bar(stat = 'identity', fill = "red") + geom_label(aes(label = number_of_earthquakes)) + labs(title = 'Earthquake distribution based on magnit
```

```

ude classes',
  subtitle = 'Huge increased on the number of earthquake occurred wit
hin the smaller magnitude classes.',
  caption = "The dataset contains list of recorded earthquake in Indo
nesia from year 2000-2020.",
  x = 'magnitude classes',
  y = 'Number of earthquakes')

```

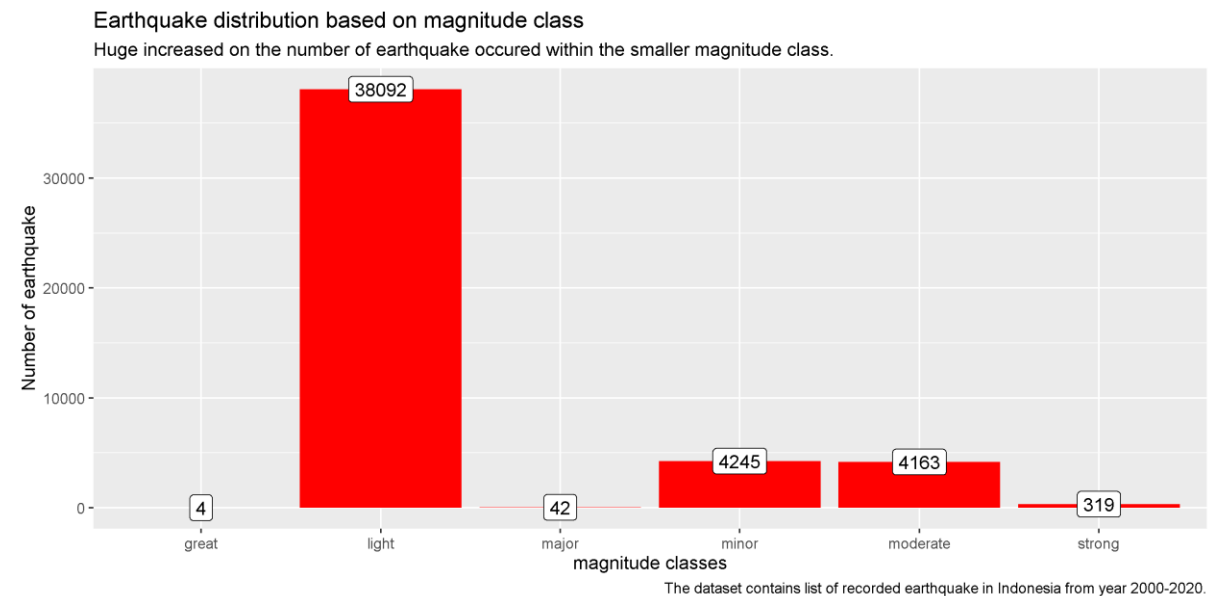


Fig 1. Earthquake distribution based on magnitude classes.

According to the bar chart of Fig 1. above, there are over 4245 minor earthquakes, 38092 light earthquakes, and 4163 moderate earthquakes occurred on Indonesian territory between 2000 and 2020. The most concerning thing is that there are 319 strong earthquake, 42 major earthquake, and 4 great earthquake. The magnitude classes above strong will almost certainly cause significant damage to a nearby town. To get more specific data about when this earthquake occurred, we need to collect more information about earthquakes that occur every year.

```

eq_INA %>%
  group_by(year) %>%
  summarise(number_of_earthquakes = n()) %>%
  ggplot(aes(x = year, y = number_of_earthquakes)) +
  geom_bar(stat = 'identity', fill = "blue") + geom_label(aes(label = numbe
r_of_earthquakes)) +
  labs(title = 'Earthquake distribution based on number of earthquakes eve
ry year',
  subtitle = 'Huge number of earthquakes has been spotted within time
period of year 2005-2007.',
  caption = "The dataset contains list of recorded earthquake in Indo
nesia from year 2000-2020.",

```

```
x = 'Year',
y = 'Number of earthquakes')
```

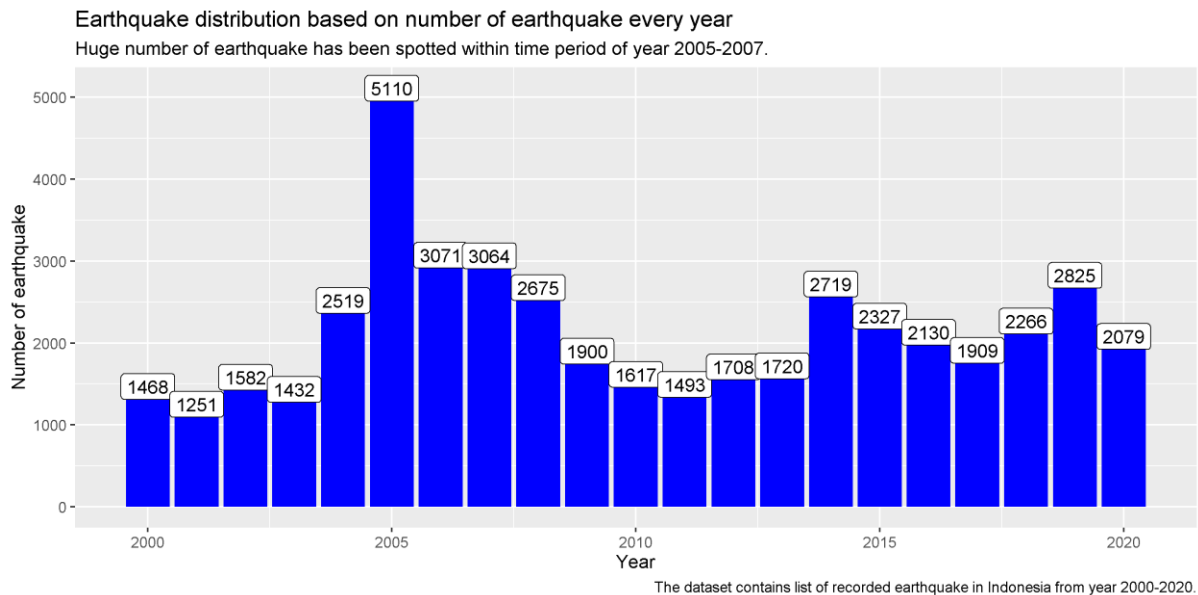


Fig 2. Earthquake distribution based on sequence of year.

The bar chart of Fig 2. above clearly shows that there are more than 1000 earthquakes occurred in Indonesia every year. There was also a significant increase in the number of earthquakes in 2005, with over 5110 earthquakes occurring on that year alone. The 2005 earthquake doubled the number of earthquakes recorded from previous year in 2004.

This is interesting.

One of the hypotheses for why such a phenomenon could occur is that there was an increase in the number of minor earthquakes preceding the aftershock of the major earthquake. We will use a scatter plot to gather more information about the earthquake distribution based on the magnitude scale and time period.

```
eq_INA %>%
  ggplot(aes(x = date, y = mag)) +
  geom_point() + labs(title = 'Scatter plot of earthquake distribution based on magnitude scale from 2000-2020',
    subtitle = 'There has been spotted 4 great earthquake within time period of year 2004, 2005, 2007, and 2012.',
    caption = "The dataset contains list of recorded earthquake in Indonesia from year 2000-2020.",
    x = 'Year',
    y = 'Magnitude scale')
```

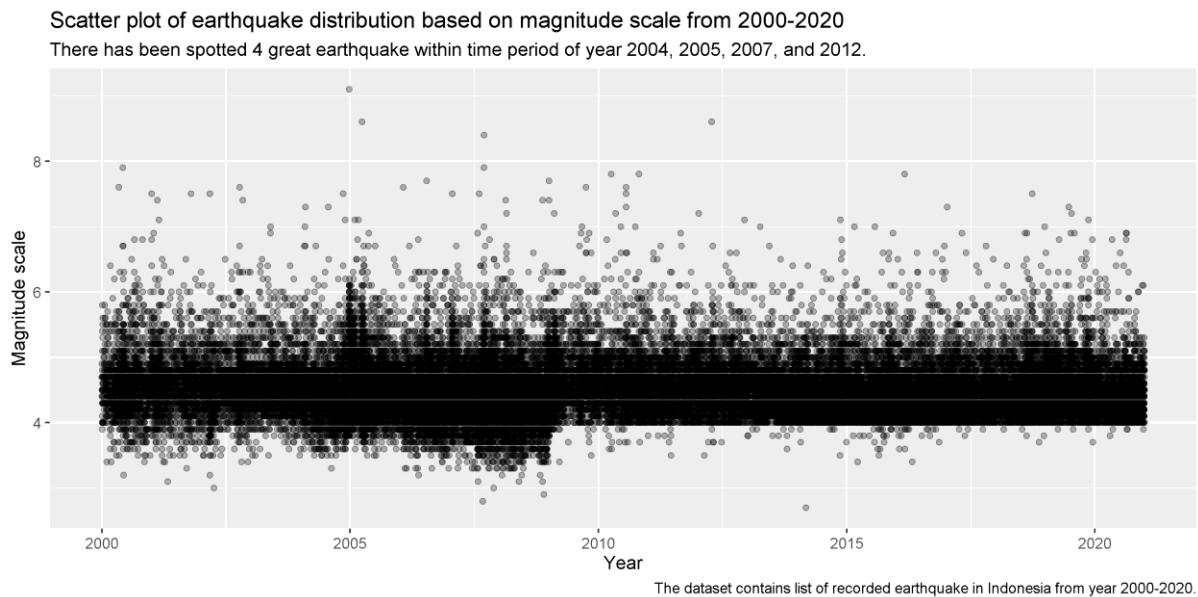


Fig 3. Scatter plot of earthquake distribution based on magnitude scale from 2000-2020.

The scatter plot from Fig 3. above shows that there were three major earthquakes between 2004 and 2007. We will mainly focus our attention to this phenomenon.

According to media news, the Sumatra-Andaman earthquake with a magnitude of 9.1 on the Richter scale occurred in December 2004. This earthquake occurred on a tectonic subduction zone where the India plate, as part of the Sunda plate, is subducted beneath the Burma microplate^[7].

Following the December 2004 great earthquake, the next great earthquake occurred four months later in March 2005. This earthquake occurred prior to the effect of the December 2004 great earthquake with a magnitude of 8.6^[8].

The third great earthquake occurred on September 2007 in which there are one great earthquake with magnitude of 8.4 and one major earthquake with magnitude of 7.9 occurring on the same day^[9].

We believe that this massive earthquake is the primary cause of the great increase in the number of earthquakes. Furthermore, we will focus our investigation to the years 2004 to 2007 where mainly great earthquakes occurred.

```
eq_INA %>%
  filter(year==2004 | year==2005 | year==2006 | year==2007) %>%
  group_by(mag_class, year) %>%
  summarise(number_of_earthquakes = n()) %>%
  ggplot(aes(x = mag_class, y = number_of_earthquakes)) +
  geom_bar(stat = 'identity', fill = "forest green") + geom_label(aes(label = number_of_earthquakes)) + facet_wrap(~year, ncol=1, strip.position =
"left") +
  labs(title = 'Earthquake distribution based on magnitude class in time
period of 2004-2007',
       subtitle = 'Huge increased on the number of minor earthquake occurred in 2005, doubled the last year.',
       caption = "The dataset contains list of recorded earthquake in Indonesia from year 2004-2007.",
       x = 'magnitude classes',
       y = 'Number of earthquakes')
```

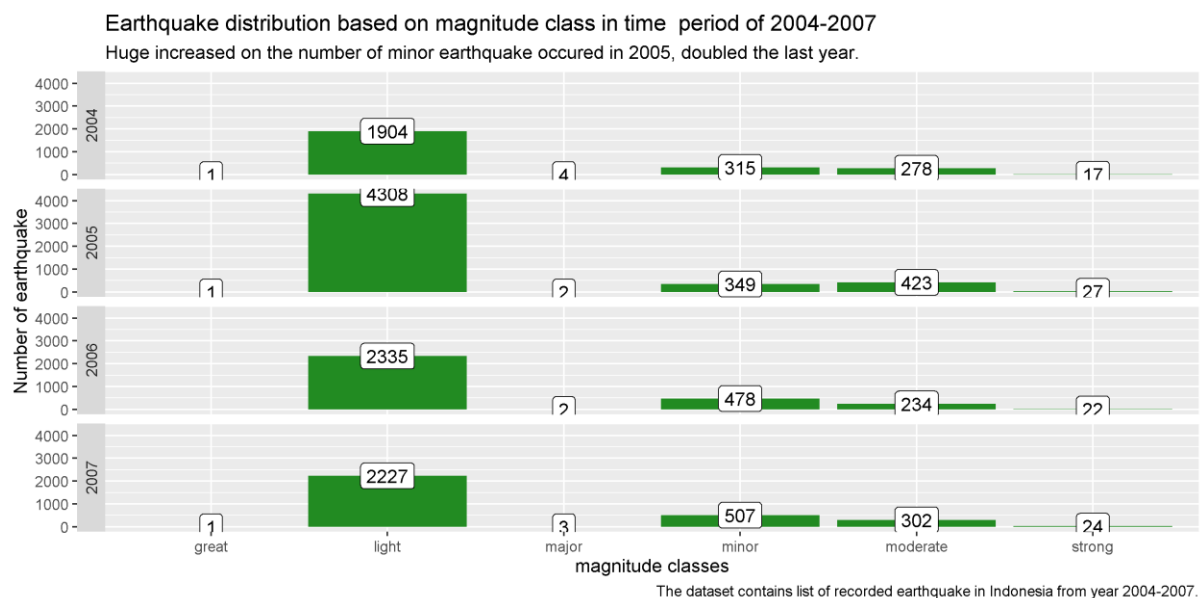


Fig 4. Earthquake distribution based on magnitude class in time period of 2004-2007.

From the bar chart of Fig 4. above, The great earthquake preceding the December 2004 earthquake and the second great earthquake in March 2005 caused a significant increase in light earthquakes in 2005. To get more information about the distribution of earthquakes by month, we need to look at the number of earthquakes that occurred in each sequence of months from 2004 to 2007.

```
eq_INA %>%
  filter(year==2004 | year==2005 | year==2006 | year==2007) %>%
  group_by(month, year) %>%
  summarise(number_of_earthquakes = n()) %>%
  ggplot(aes(x = month, y = number_of_earthquakes)) +
  geom_bar(stat = 'identity', fill = "forest green") + geom_label(aes(label = number_of_earthquakes))
```

```
l = number_of_earthquakes)) + facet_wrap(~year, ncol=1, strip.position =
"left") +
  labs(title = 'Earthquake distribution based on magnitude class each month
in 2004-2007',
        subtitle = 'Huge increase on the number of earthquake occurred prior
to the great earthquake.',
        caption = "The dataset contains list of recorded earthquake in Indo
nesia from year 2004-2007.",
        x = 'Month',
        y = 'Number of earthquakes')
```

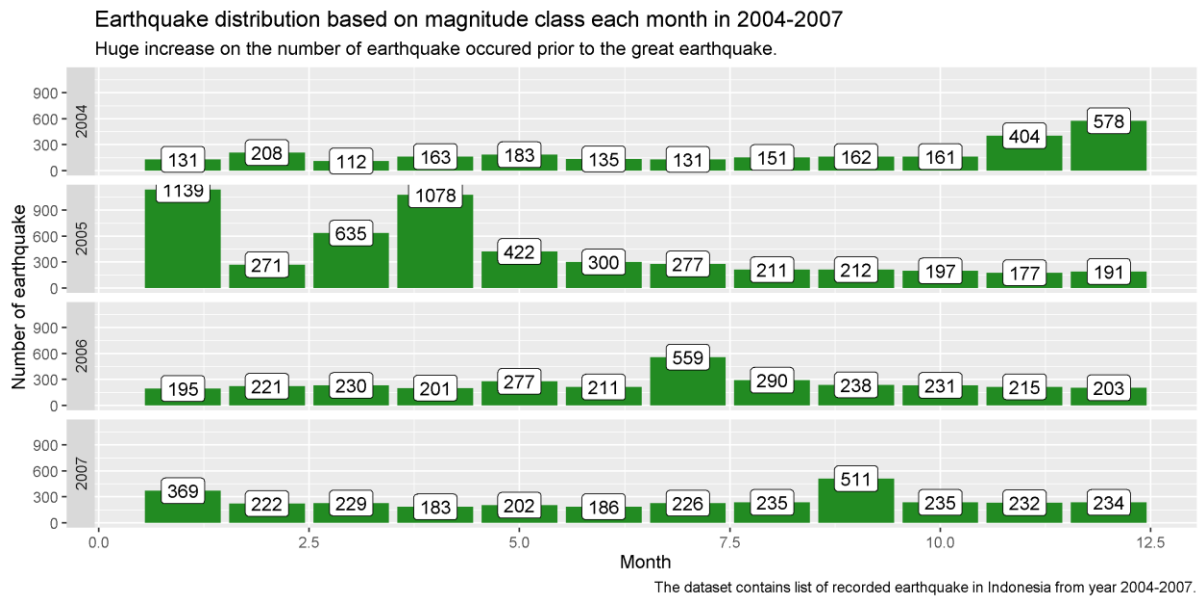


Fig 5. Earthquake distribution based on magnitude class each month in 2004-2007.

We can learn more about the increasing and decreasing number of earthquakes in each month from 2004 to 2007 by looking at the bar chart in Fig. 5. Between December 2004 and April 2005, there were a large number of earthquakes. The earthquakes occurred prior to the great earthquake of 9.1 magnitude in 2004, which resulted in the lesser earthquake on the following month.

There was also an increase in the number of earthquakes prior to the second great earthquake in March 2005. Our previous assumption about the increasing number of minor earthquakes being caused by major earthquakes appears to be correct. According to the earthquake data from December 2004 to April 2005, the great earthquake that occurred a month before caused a significant increase in the number of earthquakes on the following month.

The same phenomenon also happened when the third great earthquake occurred in September 2007. However, the number of earthquakes did not increase as rapidly as it did during the previous great earthquake during this time period.

In relation to the previous number of earthquakes each month every year, we will also exploring the data between magnitude and depth of earthquake. The following scatter plot will show the relation between magnitude and depth regarding the occurred earthquake.

```
#Relationship between mag and depth
eq_INA %>%
  ggplot(aes(x = mag, y = depth)) +
  geom_point() +
  geom_smooth(method = 'lm', se = FALSE) +
  facet_wrap(~year) + labs(title = 'Scatter plot of earthquake distributi
on based on magnitude and depth every year',
  subtitle = 'The blue line indicated that the greater value of magni
tude, the smaller value of the depth.',
  caption = "The dataset contains list of recorded earthquake in Indo
nesia from year 2000-2020.",
  x = 'Magnitude',
  y = 'Depth')
```

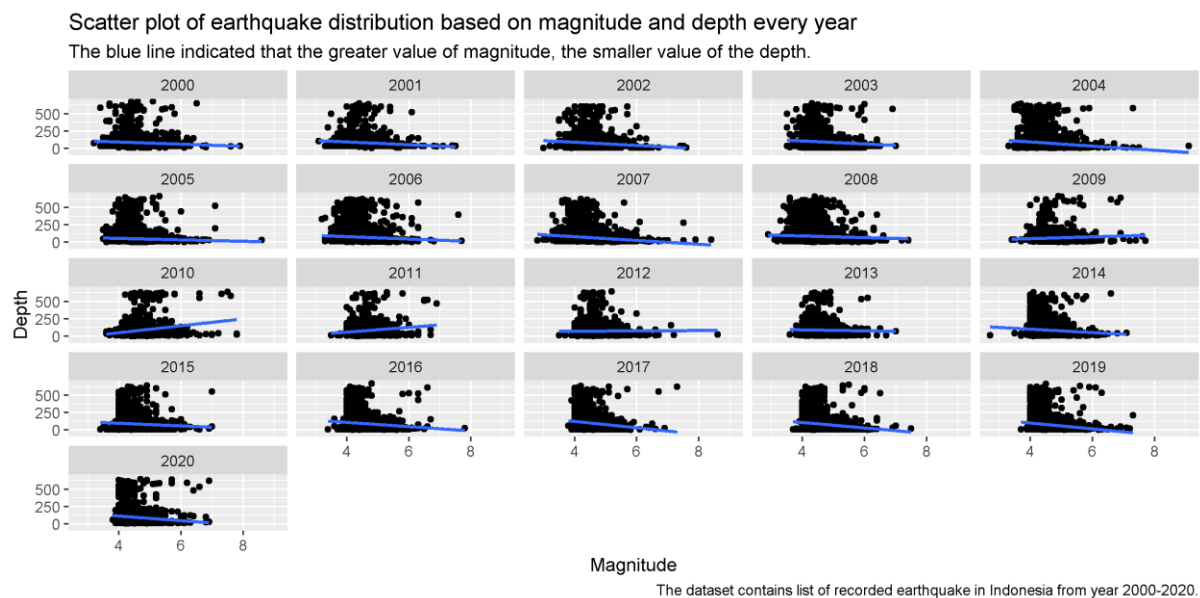


Fig 6. Scatter plot of earthquake distribution based on magnitude and depth every year.

The relationship between mag and depth is inverse, as evidenced by the majority of the scatter plot above from Fig. 6. It means that as the magnitude increases, the depth decreases. The large magnitude earthquake is most likely

to have occurred at a shallow depth. This is correct because the previous great earthquake happened at a depth of less than 50 kilometers beneath the earth's surface.

Conclusion

1. From 2000 to 2020, there were 4245 minor earthquakes, 38092 light earthquakes, and 4163 moderate earthquakes on Indonesian territory. Despite this, the most notable earthquakes are the major and great earthquakes with the major occurring 52 times and the great occurring 4 times.
2. Every year, more than 1000 earthquakes with varying magnitudes occur in Indonesia, ranging from minor to great earthquake.
3. The large number of earthquakes that occurred in 2005 were caused by aftershocks from the previous year's great earthquake in December 2004 and March 2005. Because of the two great earthquakes, the number of earthquakes in 2005 is increasing doubled from the previous year especially for the lesser earthquakes.
4. The greater magnitude Richter scale of earthquake occurred at the shallower depth.

Source

You can get the source code and the earthquake data from [here](#).

Reference

- [1] Adagunodo, A., et all. (2017) Evaluation of $0 < M < 8$ earthquake data sets in African-Asian region during 1966–2015. Data in Brief
- [2] Gongqian, Xu., Hanghang, He., Ying ,Tao., Yang, Wang. (2005) Earthquake. Roskilde Universitetscenter's Digitale Arkiv.
- [3] Pribadi, K., et all. (2021) Learning from past earthquake disasters: The need for knowledge management system to enhance infrastructure resilience in Indonesia. International Journal of Disaster Risk Reduction
- [4] Senduk, R., Indwiarti, Nhita, F. (2019) Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm. Indonesia Journal on Computing.
- [5] <https://earthquake.usgs.gov/earthquakes/search/>

- [6] <https://www.gns.cri.nz/Home/Learning/Science-Topics/Earthquakes/Monitoring-Earthquakes/Other-earthquake-questions/What-is-the-Richter-Magnitude-Scale>
- [7] https://www.usgs.gov/centers/pcmsc/science/tsunami-generation-2004-m91-sumatra-andaman-earthquake?qt-science_center_objects=0#qt-science_center_objects
- [8] <https://reliefweb.int/report/indonesia/indonesia-28-march-earthquake-situation-report-9>
- [9] <https://reliefweb.int/map/indonesia/m85-and-79-southern-sumatra-earthquakes-12-september-2007-and-m70-13-september-2007>