Earthquake Death Correlation

Maya Maciel-Seidman

2024-02-04

Set Up

```
# Load tidyverse:
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#Load knitr:
library(knitr)

#Load dplyr:
library(dplyr)
```

Data

```
# Load data downloaded in .csv format:
earthquakes <- read.csv("./earthquakes.csv")

# Find dimensions of dataset:
dim(earthquakes)
```

```
## [1] 5963   16
```

This dataset is NOAA's significant Earthquake Database from 2150 BC to October 16, 2017. It was downloaded as a csv from benjiao's GitHub page at this link: https://github.com/benjiao/significant-earthquakes/blob/master/earthquakes.csv. This dataset records all significant earthquakes around the globe

that occurred in the timeline described above. For each earthquake, the dataset contains characteristics including magnitude, focal depth (depth of the earthquake's hypocenter), death toll, and location, among other variables. The dimensions of this dataset are 5963 x 16. This means that there are 5,963 observations (earthquakes) and 16 variables describing each earthquake.

Research Question

Using this dataset, I aim to investigate if an earthquake's focal depth (depth of hypocenter below Earth's surface) or magnitude is more strongly correlated with the number of deaths resulting from that earthquake. Additionally, I aim to determine the death toll for each earthquake magnitude class (a categorical description of an earthquake's magnitude) to determine if death toll follows the severity of earthquake magnitude.

Variables of Interest

The variables of interest to address my research question are an earthquake's focal depth, magnitude, and death toll.

Data Wrangling

```
# Select only the columns with the variables of interest:
earthquakes <- earthquakes %>% select(deaths, focal_depth, magnitude)

# Look at summary of the data to find if there are NA values:
summary(earthquakes)
```

```
##      deaths        focal_depth       magnitude
##  Min.   :     1   Min.   :  0.00   Min.   :1.600
##  1st Qu.:     4   1st Qu.: 11.00   1st Qu.:5.700
##  Median :    26   Median : 27.00   Median :6.500
##  Mean   :  3988   Mean   : 41.94   Mean   :6.489
##  3rd Qu.:   400   3rd Qu.: 40.00   3rd Qu.:7.300
##  Max.   :830000   Max.   :678.00   Max.   :9.500
##  NA's   :4019     NA's   :2945     NA's   :1789
```

```
# Get rid of NA values:
earthquakes <- earthquakes %>% drop_na()

# Ensure that there are no more NA values:
summary(earthquakes)
```

```
##      deaths          focal_depth       magnitude
##  Min.   :     1.0   Min.   :  0.00   Min.   :1.600
##  1st Qu.:     2.0   1st Qu.: 10.00   1st Qu.:5.700
##  Median :    10.0   Median : 24.00   Median :6.400
##  Mean   :  2257.8   Mean   : 32.45   Mean   :6.404
##  3rd Qu.:    82.5   3rd Qu.: 34.00   3rd Qu.:7.200
##  Max.   :316000.0   Max.   :651.00   Max.   :9.500
```

```r
# Create a new column for magnitude class since the original dataset did not
# have a column for this:
#
earthquakes <- earthquakes %>%
  mutate(magnitude_class =
           case_when(magnitude < 3 ~ "Micro",
                     magnitude >= 3 & magnitude <= 3.9 ~ "Minor",
                     magnitude >= 4 & magnitude <= 4.9 ~ "Light",
                     magnitude >= 5 & magnitude <= 5.9 ~ "Moderate",
                     magnitude >= 6 & magnitude <= 6.9 ~ "Strong",
                     magnitude >= 7 & magnitude <= 7.9 ~ "Major",
                     magnitude >= 8 ~ "Great"))

# Check the first few rows of the data to make sure the new column was added properly:
head(earthquakes)
```

```
##   deaths focal_depth magnitude magnitude_class
## 1      1          18       7.1           Major
## 2  12000          10       5.2        Moderate
## 3  12000          10       5.2        Moderate
## 4  50000          15       7.6           Major
## 5  60000          10       6.5          Strong
## 6   2000          10       6.1          Strong
```

First, I used `select()` to create a subset of the original dataset with only the columns containing variables of interest. I used `summary()` to see a statisitcal summary of the new dataset and found that there were NA values. I then removed all NA values using `drop_na()` and double checked to make sure no NA values were left by using `summary()` again. Then, I created a new column in the dataset for the magnitude class (`magnitude_class`), which is a categorical variable that describes the magnitude of an earthquake in a way that gives more context than just the numerical Richter Scale value of magnitude. Magnitude classes are based on ranges within the Richter Scale magnitude values. I created this new column using `mutate()` and `case_when()`, which allowed for conditional statements to determine the values of the new column. Finally, I used `head()` to check the first few rows of the dataset to ensure that the new column was created correctly.

Analysis

```r
# Correlation between focal depth and deaths using base R function cor():
focal_depth_correlation <- cor(earthquakes$focal_depth, earthquakes$deaths,
                               method = "pearson")
# Print correlation between focal depth and death toll:
print(paste0("Focal Depth r: ", focal_depth_correlation))
```

```
## [1] "Focal Depth r: -0.0324249632358217"
```

```r
# Correlation between magnitude and deaths using base R function cor():
magnitude_correlation <- cor(earthquakes$magnitude, earthquakes$deaths,
                             method = "pearson")
# Print correlation between magnitude and death toll:
print(paste0("Magnitude r: ", magnitude_correlation))
```

```
## [1] "Magnitude r: 0.140897959639257"
```

```r
# Create vector of the resulting r for each variable:
pearson_correlation_coefficient <- c(focal_depth_correlation,
                                     magnitude_correlation)
# Create vector of variable names:
variable <- c("Focal Depth", "Magnitude")

# Create dataframe of focal depth and magnitude and their correlation
# coefficients:
correlation_df <- data.frame(variable, pearson_correlation_coefficient)

# Magnitude class death toll:
# Create death toll variables for each magnitude class:
micro_deaths <- 0
minor_deaths <- 0
light_deaths <- 0
moderate_deaths <- 0
strong_deaths <- 0
major_deaths <- 0
great_deaths <- 0

# Iterate through dataset and add death counts to each respective magnitude
# class death toll:
for(i in 1:nrow(earthquakes)){
  if(earthquakes[i,4] == "Micro"){
    micro_deaths <- micro_deaths + 1
  }
  else if(earthquakes[i,4] == "Minor"){
    minor_deaths <- minor_deaths + 1
  }
  else if(earthquakes[i,4] == "Light"){
    light_deaths <- light_deaths + 1
  }
  else if(earthquakes[i,4] == "Moderate"){
    moderate_deaths <- moderate_deaths + 1
  }
  else if(earthquakes[i,4] == "Strong"){
    strong_deaths <- strong_deaths + 1
  }
  else if(earthquakes[i,4] == "Major"){
    major_deaths <- major_deaths + 1
  }
  else{
    great_deaths <- great_deaths + 1
  }
}

# Create vector of the resulting death tolls for each magnitude class:
death_toll_vector <- c(micro_deaths,
                       minor_deaths,
                       light_deaths,
                       moderate_deaths,
                       strong_deaths,
```

```
                        major_deaths,
                        great_deaths)

# Create a vector of the different earthquake magnitude classes:
magnitude_class_vector <- c("Micro",
                            "Minor",
                            "Light",
                            "Moderate",
                            "Strong",
                            "Major",
                            "Great")

# Create a dataframe of the earthquake magnitude classes and their respective
# death tolls:
death_toll_df <- data.frame(magnitude_class_vector, death_toll_vector)

# Rename magnitude class column in df:
death_toll_df <- death_toll_df %>% rename(magnitude_class =
                                            magnitude_class_vector)
# Rename death toll column in df:
death_toll_df <- death_toll_df %>% rename(death_toll = death_toll_vector)
```

```
# View the correlation coefficients in table format:
kable(correlation_df)
```

variable	pearson_correlation_coefficient
Focal Depth	-0.032425
Magnitude	0.140898

```
# View the resulting death toll in table format to determine how deaths
# are distributed between magnitude classes:
kable(death_toll_df)
```

magnitude_class	death_toll
Micro	2
Minor	6
Light	78
Moderate	285
Strong	428
Major	290
Great	66

The first step of my analysis was to compute the Pearson correlation coefficients between focal depth and deaths and between magnitude and deaths. I then created a dataframe of the resulting correlations. My next step was to count a death toll for each magnitude class. I did so by iterating through the earthquakes dataset using a for loop and adding the deaths from each earthquake to their respective magnitude class death toll. I then created a dataframe of the magnitude classes and their death tolls. Finally, I used `kable()` to present the two resulting dataframes I created as clean tables in the knitted PDF.

Discussion

To address the first part of my research question, the resulting Pearson correlation coefficients for focal depth and magnitude are -0.032425 and 0.140898, respectively. Based on this, an earthquake's focal depth is negatively correlated to its death toll. This makes sense since the deeper the earthquake is, the less deadly and destructive it would be since it would be farther from Earth's surface. Magnitude was found to be positively correlated with deaths, which also makese sense since this means that more severe earthquakes are correlated with higher death tolls. Both of these correlations are weak, but to answer my research question, magnitude is more strongly correlated to deaths than focal depth is. I was not surprised by this finding, but I was surprised by the weakness of both correlations.

Focusing on the second part of my research question, the table of magnitude class death tolls shows that death toll does not increase directly as earthquake severity increases. The micro (least severe) magnitude class has the smallest death toll (2) and the strong (3rd most severe) magnitude class has the largest death toll (428). The least severe magnitude class having the smallest death toll makes sense. At a first glance, one could assume that the strongest magnitude class would have the largest death toll. However, this is not what my results table suggests. There are many possible reasons for this. For example, major and great level earthquakes are far less common than less severe earthquakes. This would result in less deaths just due to less frequency. Additionally, a lot of NA values for deaths were dropped during data wrangling. Since this dataset goes back to 2150 BC, there could be many deaths that were not recorded at the time of these more severe earthquakes.

Overall, I found answers to my research question. Firstly, earthquake magnitude is more strongly correlated with deaths than earthquake focal depth. Micro magnitude earthquakes have the least number of overall deaths, while strong magnitude earthquakes have the most. My analyis answered my research question, but further analysis would be required to fully investigate relationships between magnitude and death toll to draw more informed conclusions.