EDA-STAT515-FinalProject.R

smann

2024-05-01

# Load necessary libraries  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(readr)  
library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

library(ggplot2)  
  
# Read the dataset  
earthquakes <- read\_csv("Global\_Seismic\_2023.csv")

## New names:  
## • `` -> `...23`

## Rows: 24682 Columns: 23  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (8): magType, net, id, place, type, status, locationSource, magSource  
## dbl (12): latitude, longitude, depth, mag, nst, gap, dmin, rms, horizontalE...  
## lgl (1): ...23  
## dttm (2): time, updated  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

# Check structure of the dataset  
str(earthquakes)

## spc\_tbl\_ [24,682 × 23] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ time : POSIXct[1:24682], format: "2023-01-01 00:49:25" "2023-01-01 01:41:43" ...  
## $ latitude : num [1:24682] 52.1 7.14 19.16 -4.78 53.4 ...  
## $ longitude : num [1:24682] 178.5 126.7 -66.5 102.8 -166.9 ...  
## $ depth : num [1:24682] 82.8 79.2 24 63.8 10 ...  
## $ mag : num [1:24682] 3.1 4.5 3.93 4.3 3 2.8 4.1 4.1 2.7 4.1 ...  
## $ magType : chr [1:24682] "ml" "mb" "md" "mb" ...  
## $ nst : num [1:24682] 14 32 23 17 19 19 15 40 30 27 ...  
## $ gap : num [1:24682] 139 104 246 187 190 127 87 81 154 119 ...  
## $ dmin : num [1:24682] 0.87 1.152 0.848 0.457 0.4 ...  
## $ rms : num [1:24682] 0.18 0.47 0.22 0.51 0.31 0.18 0.15 0.32 0.14 0.45 ...  
## $ net : chr [1:24682] "us" "us" "pr" "us" ...  
## $ id : chr [1:24682] "us7000j5a1" "us7000j3xk" "pr2023001000" "us7000j3xm" ...  
## $ updated : POSIXct[1:24682], format: "2023-03-11 22:51:52" "2023-03-11 22:51:45" ...  
## $ place : chr [1:24682] "Rat Islands, Aleutian Islands, Alaska" "23 km ESE of Manay, Philippines" "Puerto Rico region" "99 km SSW of Pagar Alam, Indonesia" ...  
## $ type : chr [1:24682] "earthquake" "earthquake" "earthquake" "earthquake" ...  
## $ horizontalError: num [1:24682] 8.46 5.51 0.91 10.25 1.41 ...  
## $ depthError : num [1:24682] 21.21 7.45 15.95 6.58 2 ...  
## $ magError : num [1:24682] 0.097 0.083 0.09 0.238 0.085 0.06 0.213 0.095 0.112 0.123 ...  
## $ magNst : num [1:24682] 14 43 16 5 18 36 6 34 35 18 ...  
## $ status : chr [1:24682] "reviewed" "reviewed" "reviewed" "reviewed" ...  
## $ locationSource : chr [1:24682] "us" "us" "pr" "us" ...  
## $ magSource : chr [1:24682] "us" "us" "pr" "us" ...  
## $ ...23 : logi [1:24682] NA NA NA NA NA NA ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. time = col\_datetime(format = ""),  
## .. latitude = col\_double(),  
## .. longitude = col\_double(),  
## .. depth = col\_double(),  
## .. mag = col\_double(),  
## .. magType = col\_character(),  
## .. nst = col\_double(),  
## .. gap = col\_double(),  
## .. dmin = col\_double(),  
## .. rms = col\_double(),  
## .. net = col\_character(),  
## .. id = col\_character(),  
## .. updated = col\_datetime(format = ""),  
## .. place = col\_character(),  
## .. type = col\_character(),  
## .. horizontalError = col\_double(),  
## .. depthError = col\_double(),  
## .. magError = col\_double(),  
## .. magNst = col\_double(),  
## .. status = col\_character(),  
## .. locationSource = col\_character(),  
## .. magSource = col\_character(),  
## .. ...23 = col\_logical()  
## .. )  
## - attr(\*, "problems")=<externalptr>

# Convert 'time' column to datetime format  
earthquakes$time <- ymd\_hms(earthquakes$time)  
  
# Check for missing values  
colSums(is.na(earthquakes))

## time latitude longitude depth mag   
## 0 0 0 0 0   
## magType nst gap dmin rms   
## 0 0 0 0 0   
## net id updated place type   
## 0 0 0 0 0   
## horizontalError depthError magError magNst status   
## 0 0 0 0 0   
## locationSource magSource ...23   
## 0 0 24682

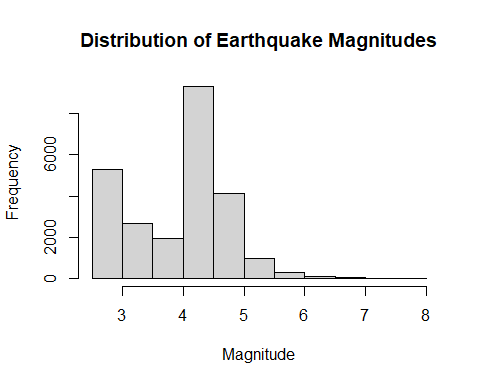
# Summary statistics  
summary(earthquakes$mag)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.600 3.180 4.200 3.968 4.500 7.800

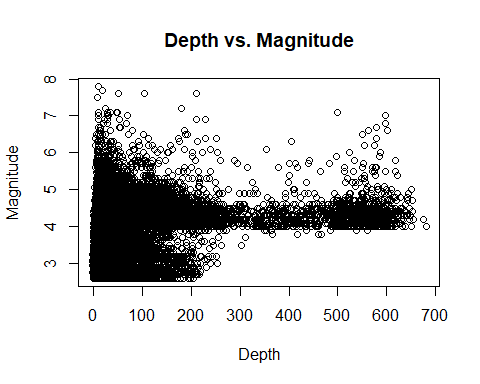
summary(earthquakes$depth)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -3.37 10.00 22.00 66.88 66.34 681.24

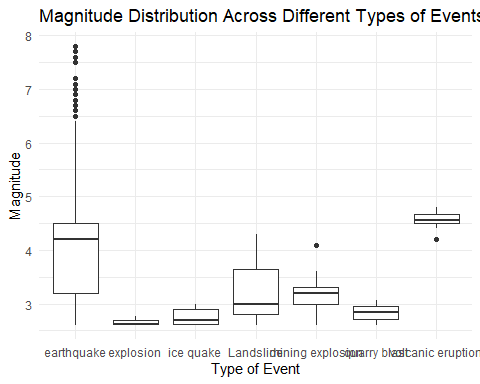
# Histogram of magnitudes  
hist(earthquakes$mag, main = "Distribution of Earthquake Magnitudes", xlab = "Magnitude")



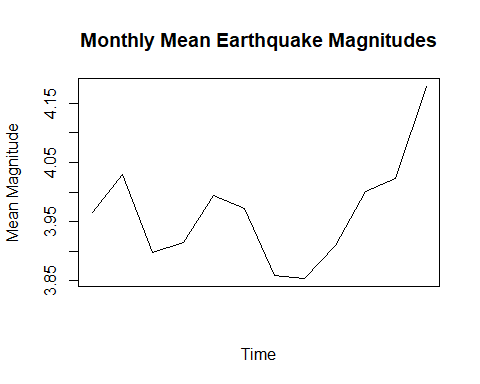
# Scatter plot of depth vs. magnitude  
plot(earthquakes$depth, earthquakes$mag, main = "Depth vs. Magnitude", xlab = "Depth", ylab = "Magnitude")



# Create boxplot for each type of event  
ggplot(earthquakes, aes(x = type, y = mag)) +  
 geom\_boxplot() +  
 labs(title = "Magnitude Distribution Across Different Types of Events",  
 x = "Type of Event", y = "Magnitude") +  
 theme\_minimal()



# Extract year and month from the time column  
earthquakes <- earthquakes %>%  
 mutate(year\_month = format(time, "%Y-%m"))  
  
# Convert the time column to a date format  
earthquakes$time <- as.POSIXct(earthquakes$time)  
  
# Group the data by year and month, and calculate the mean magnitude  
monthly\_magnitudes <- earthquakes %>%  
 group\_by(year\_month) %>%  
 summarize(mean\_magnitude = mean(mag, na.rm = TRUE))  
  
# Convert the data to a time series object  
monthly\_magnitudes\_ts <- ts(monthly\_magnitudes$mean\_magnitude, start = c(year(min(earthquakes$time)), month(min(earthquakes$time))), frequency = 12)  
  
# Plot the time series of monthly mean magnitudes  
plot(monthly\_magnitudes\_ts,  
 main = "Monthly Mean Earthquake Magnitudes",  
 xlab = "Time",  
 ylab = "Mean Magnitude",  
 xaxt = "n")   
  
# Add custom x-axis labels for years  
axis(1, at = seq(1, length(monthly\_magnitudes\_ts), by = 12), labels = unique(substr(monthly\_magnitudes$year\_month, 1, 4)))



# Calculate the correlation coefficient between rms and nst  
correlation\_coefficient <- cor(earthquakes$rms, earthquakes$nst)  
  
# Print the correlation coefficient  
print(paste("Correlation Coefficient between rms and nst:", correlation\_coefficient))

## [1] "Correlation Coefficient between rms and nst: 0.142417511828131"

# Select relevant variables  
relevant\_vars <- c("rms", "mag", "depth", "gap")  
  
# Calculate the correlation coefficients  
correlation\_results <- cor(earthquakes[relevant\_vars])  
  
# Print correlation coefficients  
print(correlation\_results)

## rms mag depth gap  
## rms 1.0000000 0.4614765 0.1082799 -0.2545420  
## mag 0.4614765 1.0000000 0.1559964 -0.3323671  
## depth 0.1082799 0.1559964 1.0000000 -0.1440770  
## gap -0.2545420 -0.3323671 -0.1440770 1.0000000