# Analyzing Natural Disasters

In this exercise, we will analyze the relationship between various factors related to natural disasters and their economic impact. The dataset `natural_disasters_2024.csv` contains information about different natural disasters recorded globally, including their type, severity, and the economic damage they caused. Understanding these relationships can provide insights into which factors are most predictive of economic losses, aiding in disaster preparedness and response planning.

### Dataset Overview

The dataset `natural_disasters_2024.csv` contains the following variables:

| Name | Description |
|---|---|
| `Disaster_Type` | The type of disaster (e.g., hurricane, earthquake, flood). |
| `Magnitude` | A severity score (1-10) indicating the intensity of the disaster. |
| `Location` | The area affected by the disaster. |
| `Economic_Loss` | The estimated economic damage caused by the disaster (in USD). |
| `Fatalities` | The number of casualties resulting from the disaster. |
| `Date` | The date when the disaster occurred. |

```r
natural_disasters <- read.csv("natural_disasters_2024.csv")
options(scipen = 999)
```

1. Calculate the total economic loss for each disaster type across all locations. Identify the disaster type with the highest total economic loss, then provide the amount of the loss.

```r
total_loss_per_type <- tapply(natural_disasters$Economic_Loss,
    natural_disasters$Disaster_Type, sum, na.rm = TRUE)

max(total_loss_per_type)
```

```
[1] 1022281208596
```

Answer: 1022281208594

2. For each location, calculate the number of disasters recorded and the average fatalities. Identify the location with the highest number of disasters and the one with the highest average fatalities.

```r
# Calculate the number of disasters for each location
disasters_per_location <- tapply(natural_disasters$Disaster_ID,
    natural_disasters$Location, length)

# Calculate the average fatalities for each location
average_fatalities_per_location <- tapply(natural_disasters$Fatalities,
    natural_disasters$Location, mean, na.rm = TRUE)

location_most_disasters <- names(disasters_per_location)[which.max(disasters_per_location)]

location_highest_avg_fatalities <- names(average_fatalities_per_location)[which.max(average_

location_most_disasters
```

```
[1] "Brazil"
```

```r
location_highest_avg_fatalities
```

```
[1] "Brazil"
```

Answer (number of disasters): Brazil Answer (number of fatalities): Brazil

3. Fit a multiple linear regression model with Economic_Loss as the dependent variable and Fatalities as the independent variable. Report the coefficient for Fatalities.

```r
# Fit a multiple linear regression model
model_multiple <- lm("Economic_Loss ~ Fatalities", data = natural_disasters)

summary(model_multiple)
```

```
Call:
lm(formula = "Economic_Loss ~ Fatalities", data = natural_disasters)

Residuals:
       Min         1Q     Median         3Q        Max
-507446782 -241648110    2831941  247885441  498254628

Coefficients:
               Estimate  Std. Error t value            Pr(>|t|)
(Intercept) 500731642.8   5690312.5  87.997 <0.0000000000000002 ***
Fatalities        851.3       986.7   0.863               0.388
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 286100000 on 9998 degrees of freedom
Multiple R-squared:  7.445e-05, Adjusted R-squared:  -2.557e-05
F-statistic: 0.7444 on 1 and 9998 DF,  p-value: 0.3883
```

Answer: 851.3

4. For each disaster type, calculate the total economic loss in the location with the highest overall fatalities. Which disaster type contributed the most to the economic loss in that location?

```r
# Find the location with the highest total fatalities
total_fatalities_per_location <- tapply(natural_disasters$Fatalities,
    natural_disasters$Location, sum, na.rm = TRUE)

location_highest_fatalities <- names(total_fatalities_per_location)[which.max(total_fataliti

# Subset the data for this location
subset_data <- natural_disasters[natural_disasters$Location ==
    location_highest_fatalities, ]

# Calculate the total economic loss for each disaster type
# in this location
total_loss_per_type_in_location <- tapply(subset_data$Economic_Loss,
    subset_data$Disaster_Type, sum, na.rm = TRUE)

# Find the disaster type contributing most to the economic
# loss
```

```r
disaster_highest_loss <- names(total_loss_per_type_in_location)[which.max(total_loss_per_type

total_loss_per_type_in_location
```

```
  Earthquake        Flood    Hurricane      Tornado     Wildfire 
179370658645 176094507457 162255649961 194012746732 163522268892 
```

```r
disaster_highest_loss
```

```
[1] "Tornado"
```

Answer: Tornado

5. Calculate the range of economic losses (maximum - minimum) for each disaster type and location combination. Which combination has the greatest range?

```r
# Calculate the maximum economic loss for each disaster
# type and location combination
max_loss_per_type_location <- tapply(natural_disasters$Economic_Loss,
    list(natural_disasters$Disaster_Type, natural_disasters$Location),
    max, na.rm = TRUE)

# Calculate the minimum economic loss for each disaster
# type and location combination
min_loss_per_type_location <- tapply(natural_disasters$Economic_Loss,
    list(natural_disasters$Disaster_Type, natural_disasters$Location),
    min, na.rm = TRUE)

# Calculate the range as the difference between max and min
range_loss_per_type_location <- max_loss_per_type_location -
    min_loss_per_type_location

# Find the combination with the greatest range
max_range_value <- max(range_loss_per_type_location, na.rm = TRUE)
max_range_combination <- which(range_loss_per_type_location ==
    max_range_value, arr.ind = TRUE)

# Print results
range_loss_per_type_location
```

```
             Brazil     China     India Indonesia     Japan       USA
Earthquake 996707003 997286257 995831648 997151590 983893935 990821664
Flood      984039581 997407126 997993517 994481143 997013083 997559987
Hurricane  994937902 996799596 994541024 987167417 997083787 995781615
Tornado    996321580 985560422 995345211 997876944 994215491 993513435
Wildfire   981128562 988517188 992353776 991374497 998816222 997326273
```

```r
rownames(range_loss_per_type_location)[max_range_combination[1]]
```

```
[1] "Wildfire"
```

```r
colnames(range_loss_per_type_location)[max_range_combination[2]]
```

```
[1] "Japan"
```

6. For each disaster type, calculate the total number of disasters recorded in the dataset and the average fatalities per event. Report the average fatalities, for the disaster type with the highest average fatalities.

```r
# Calculate the total number of disasters for each type
# using tapply
total_disasters_per_type <- tapply(natural_disasters$Disaster_ID,
    natural_disasters$Disaster_Type, length)

# Calculate the average fatalities for each type using
# tapply
max(tapply(natural_disasters$Fatalities, natural_disasters$Disaster_Type,
    mean, na.rm = TRUE))
```

```
[1] 5063.269
```

Answer: 5063.27

7. For each location, calculate the number of “High Impact” disasters (economic loss above the median). Identify the location with the most “High Impact” disasters.

```r
median_loss <- median(natural_disasters$Economic_Loss, na.rm = TRUE)
natural_disasters$High_Impact <- natural_disasters$Economic_Loss >
    median_loss

# Calculate the number of 'High Impact' disasters per
# location using tapply
high_impact_per_location <- tapply(natural_disasters$High_Impact,
    natural_disasters$Location, sum, na.rm = TRUE)

# Find the location with the most 'High Impact' disasters
location_most_high_impact <- names(high_impact_per_location)[which.max(high_impact_per_locati

# Print results
location_most_high_impact
```

```
[1] "Brazil"
```

Answer: Brazil

8. Fit a linear regression model with Fatalities as the dependent variable and Magnitude as the single predictor. Is the relationship statistically significant at the 0.05 level? Provide the p-value for the coefficient.

```r
# Fit a linear regression model with Magnitude as the
# single predictor for Fatalities
model_magnitude_fatalities <- lm(Fatalities ~ Magnitude, data = natural_disasters)

summary(model_magnitude_fatalities)
```

```

Call:
lm(formula = Fatalities ~ Magnitude, data = natural_disasters)

Residuals:
    Min      1Q  Median      3Q     Max 
-5121.7 -2504.0   -10.7  2505.6  5132.3 

Coefficients:
            Estimate Std. Error t value            Pr(>|t|)    
(Intercept)  5170.24      68.02  76.014 <0.0000000000000002 ***
Magnitude     -33.33      11.07  -3.012              0.0026 ** 
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2899 on 9998 degrees of freedom
Multiple R-squared:  0.0009064, Adjusted R-squared:  0.0008065
F-statistic: 9.071 on 1 and 9998 DF,  p-value: 0.002604
```

Answer: 0

9. Which month had the most fatalities?

```r
# Extract the month from the Date column
natural_disasters$Month <- format(as.Date(natural_disasters$Date,
    format = "%m/%d/%Y"), "%m")

total_fatalities_per_month <- tapply(natural_disasters$Fatalities,
    natural_disasters$Month, sum, na.rm = TRUE)

# Find the month with the highest total fatalities
names(total_fatalities_per_month)[which.max(total_fatalities_per_month)]
```

```
[1] "01"
```

Answer: January

10. Find the day of the month with the highest total fatalities. What was the total number of fatalities on that day?

```r
# Extract the day of the month from the Date column
natural_disasters$Day <- format(as.Date(natural_disasters$Date,
    format = "%m/%d/%Y"), "%d")

# Calculate the total number of fatalities for each day
# using tapply
total_fatalities_per_day <- tapply(natural_disasters$Fatalities,
    natural_disasters$Day, sum, na.rm = TRUE)

# Find the day with the highest total fatalities
day_most_fatalities <- names(total_fatalities_per_day)[which.max(total_fatalities_per_day)]

# Print results
day_most_fatalities
```

```
[1] "03"
```

Answer: 3