Data Analysis Practice

Michael V Cumbo

February 12, 2024

1 About this Data

Disaster Declarations Summaries is a summarized dataset describing all federally declared disasters. This dataset lists all official FEMA Disaster Declarations, beginning with the first disaster declaration in 1953 and features all three disaster declaration types: major disaster, emergency, and fire management assistance. The dataset includes declared recovery programs and geographic areas (county not available before 1964; Fire Management records are considered partial due to historical nature of the dataset).

```
library(tidyverse)
library(lubridate)
library(RSQLite)
library(DBI)
library(ggplot2)
library(dplyr)
library(forcats)
library(GGally)
library(stringr)
library(magrittr)

setwd("~/workbook")
con <- dbConnect(RSQLite::SQLite(), "Disaster_Data.db")
dbListTables(con)
```

```
## [1] "US_Declarations_2023"
## [2] "sqlean_define"
## [3] "us_disaster_declarations"
```

```
declarations_2023 <- as_tibble(dbGetQuery(
  con,
  "SELECT
  disasterNumber,
  state,
  declarationType,
  incidentType,
  declarationDate
  FROM US_Declarations_2023
ORDER BY declarationDate;"
))
dbDisconnect(con)
```

```r
# build an arrangment paired with mutaute in
# DPLYR library than pass it to ggplot and build the graph
# this reorders the factors in the
# graph so its easier to break down and look at
v1 <- declarations_2023 %>%
  select(state, DisasterType = incidentType, Year) %>%
  group_by(DisasterType) %>%
  summarise(NumberOfDisasters = n()) %>%
  arrange(NumberOfDisasters) %>%
  mutate(DisasterType = factor(DisasterType, levels = unique(DisasterType))) %>%
  ggplot(aes(x = DisasterType, y = NumberOfDisasters)) +
  geom_col() +
  theme_gray() +
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5, hjust = .4)) +
  labs(y = "Number of Disasters", x = "Disaster Type")

print(v1)
```

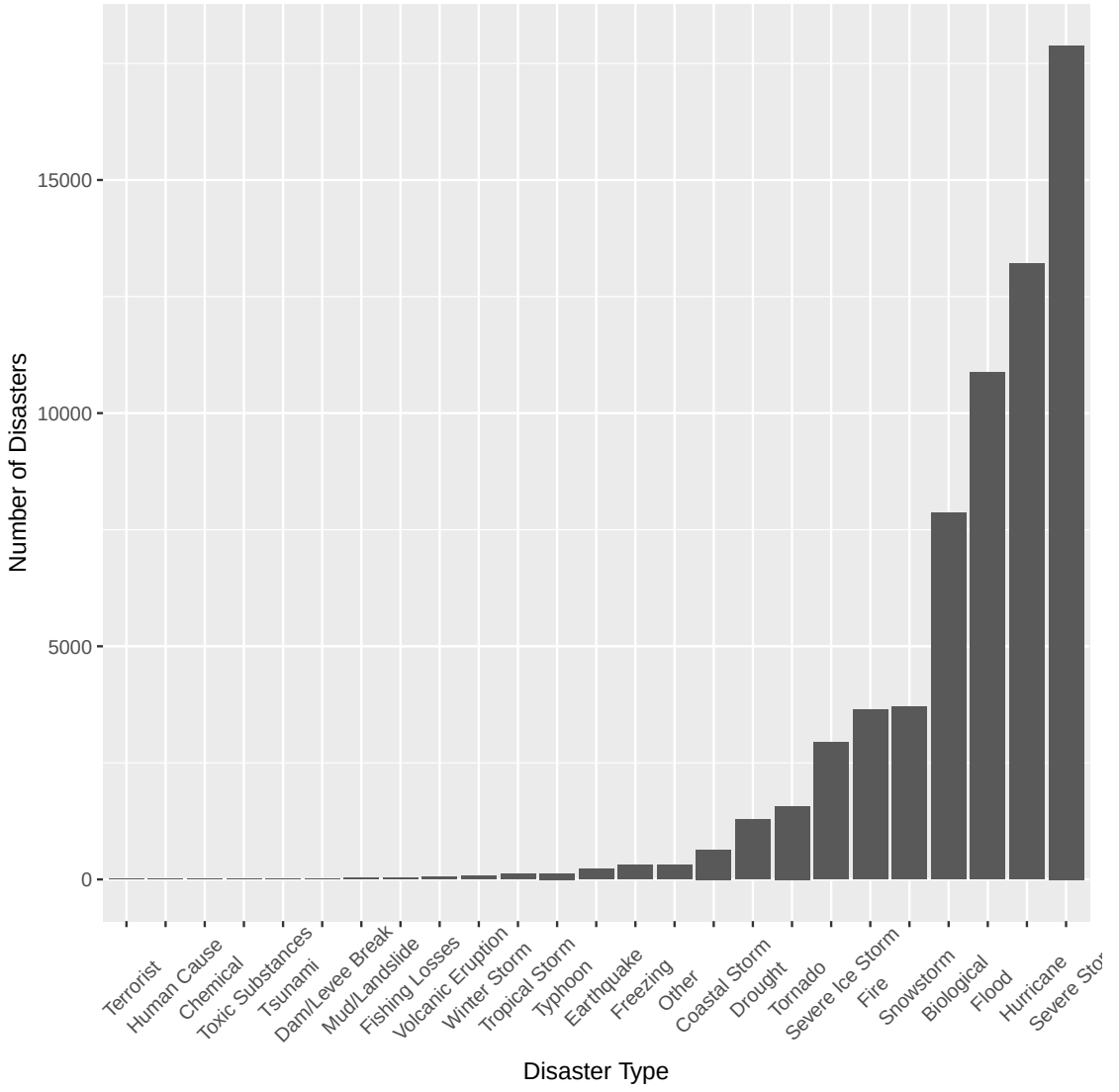


Figure 1: histogram of the total number of disasters declared by the United States

```
##       1       2       3       4
## 1654.88 1680.01 1705.14 1730.27
##       5       6       7       8
## 1755.40 1780.53 1805.66 1830.79
##       9      10      11      12
## 1855.92 1881.05 1906.18 1931.32
##
## Call:
## lm(formula = DisasterCount ~ Year, data = annual_disasters)
##
## Residuals:
##     Min      1Q  Median      3Q
## -714.57 -312.30  -49.99  177.93
##     Max
## 1361.83
##
## Coefficients:
##               Estimate Std. Error
## (Intercept) -49209.962   4821.925
## Year            25.131      2.426
##             t value Pr(>|t|)
## (Intercept)  -10.21 2.80e-15 ***
## Year          10.36 1.51e-15 ***
## ---
## Signif. codes:
##   0 '***' 0.001 '**' 0.01 '*'
##   0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 409.4 on 67 degrees of freedom
## Multiple R-squared:  0.6156,Adjusted R-squared:  0.6098
## F-statistic: 107.3 on 1 and 67 DF,  p-value: 1.514e-15
```

Model Summary

The summary of the linear model offers key insights:

Coefficients:

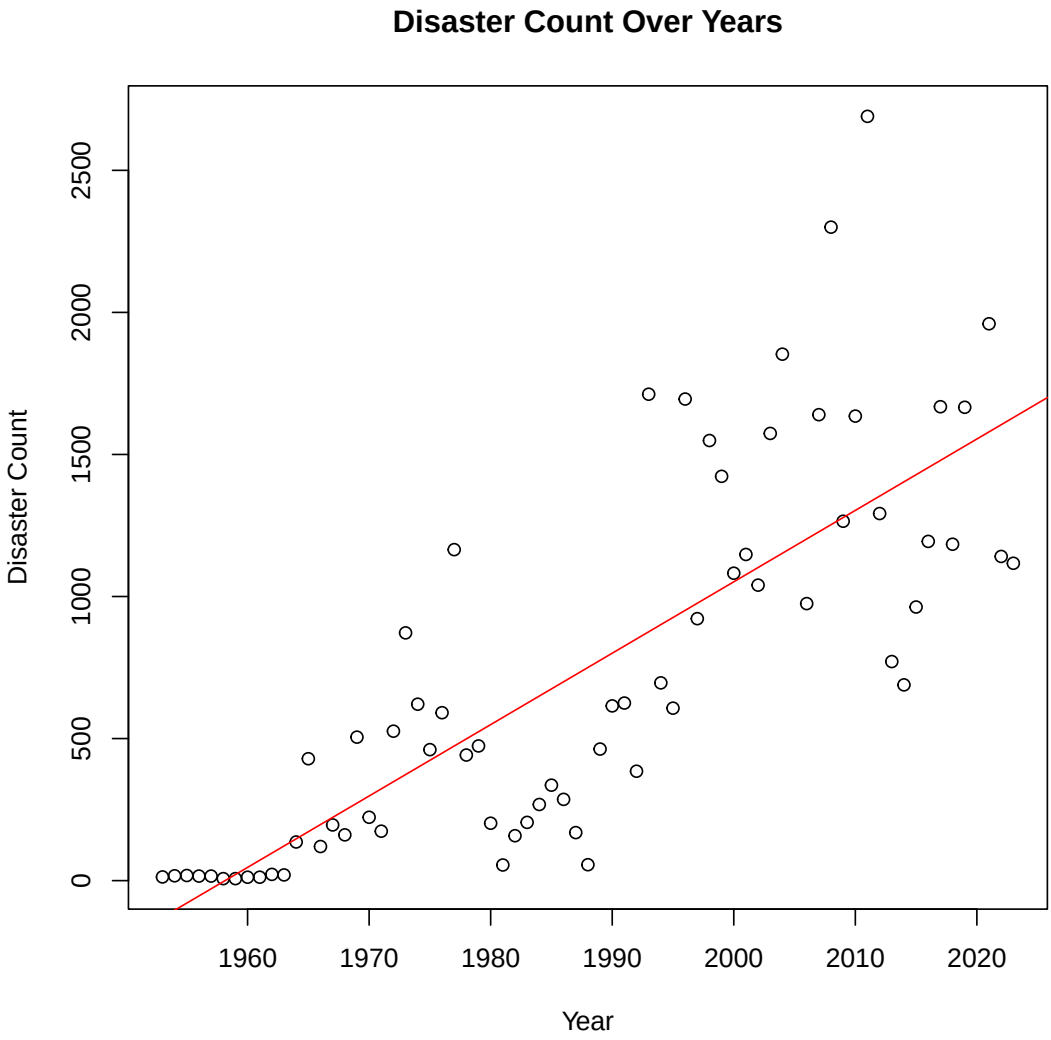
- The intercept is -49209.962, implying the model's prediction for *DisasterCount* when *Year* is 0, which is not applicable in this context.
- The slope coefficient for *Year* is 25.131. This indicates an annual increase of approximately 25.131 in *DisasterCount*, as per the model.
- These predictions indicate an upward trend in *DisasterCount* over the years.

Statistical Significance:

Both the intercept and the slope demonstrate statistical significance with a p-value < 0.001.

Model Fit:

- The R-squared value is 0.6156, signifying that approximately 61.56% of the variability in *DisasterCount* is explained by the year. However, a significant portion of variability remains unexplained.



- The Residual Standard Error (RSE) is 409.4, indicating the average deviation of data points from the fitted line.

Residuals:

The residuals range from -714.57 to 1361.83, suggesting variability around the regression line.

Interpretation and Considerations

- The model indicates a significant upward trend in disaster counts over the years.
- The presence of significant residuals and a R-squared value of 0.6156 implies that, while there is a discernible trend, other unaccounted factors might also be influencing *DisasterCount*.
- Predictions for future years should be approached with caution due to the simplicity of the model and its exclusion of other potential predictive factors.
- When interpreting these results and making future decisions, the limitations of a simple linear regression model and the impact of external factors not included in the model should be considered.