

Natural Disasters

Jaslyn Miura

2026-02-04

Goal: The goal with this data is to explore how the types and amount of natural disasters that have occurred around the world have varied since 1960-2018.

Questions of Interest: What type natural disasters occur the most around the world? Which countries experience the most natural disasters per year around the world? How have the distribution of natural disasters changed over the years (type and count)? Could the changing trends identify major shifts, such as the rise of climate change?

```
# load necessary libraries  
library(tidyverse)
```

```
Warning: package 'ggplot2' was built under R version 4.5.2
```

```
Warning: package 'tibble' was built under R version 4.5.2
```

```
Warning: package 'tidyverse' was built under R version 4.5.2
```

```
Warning: package 'readr' was built under R version 4.5.2
```

```
Warning: package 'purrr' was built under R version 4.5.2
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
v dplyr     1.1.4     v readr     2.1.6  
v forcats   1.0.0     v stringr   1.6.0  
v ggplot2   4.0.1     v tibble    3.3.1  
v lubridate  1.9.4     v tidyverse  1.3.2  
v purrr     1.2.1  
-- 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 becom
```

```

library(paletteer)
library(lubridate)

# read in natural disaster world data
nat_dis <- read_csv(here:::here("data", "pend-gdis-1960-2018-disasterlocations-csv", "pend-gd...
  mutate(year = ymd(year, truncated = 2))

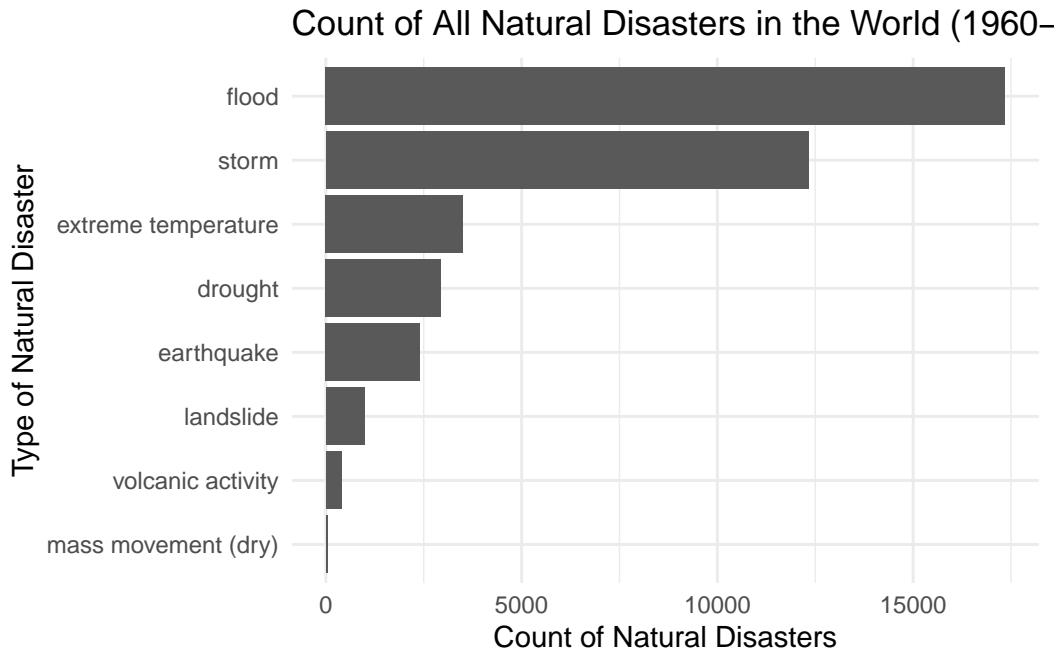
Rows: 39953 Columns: 18
-- Column specification -----
Delimiter: ","
chr (11): id, country, iso3, geolocation, adm1, adm2, adm3, location, hist_c...
dbl (7): gwno, year, geo_id, level, historical, latitude, longitude

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.

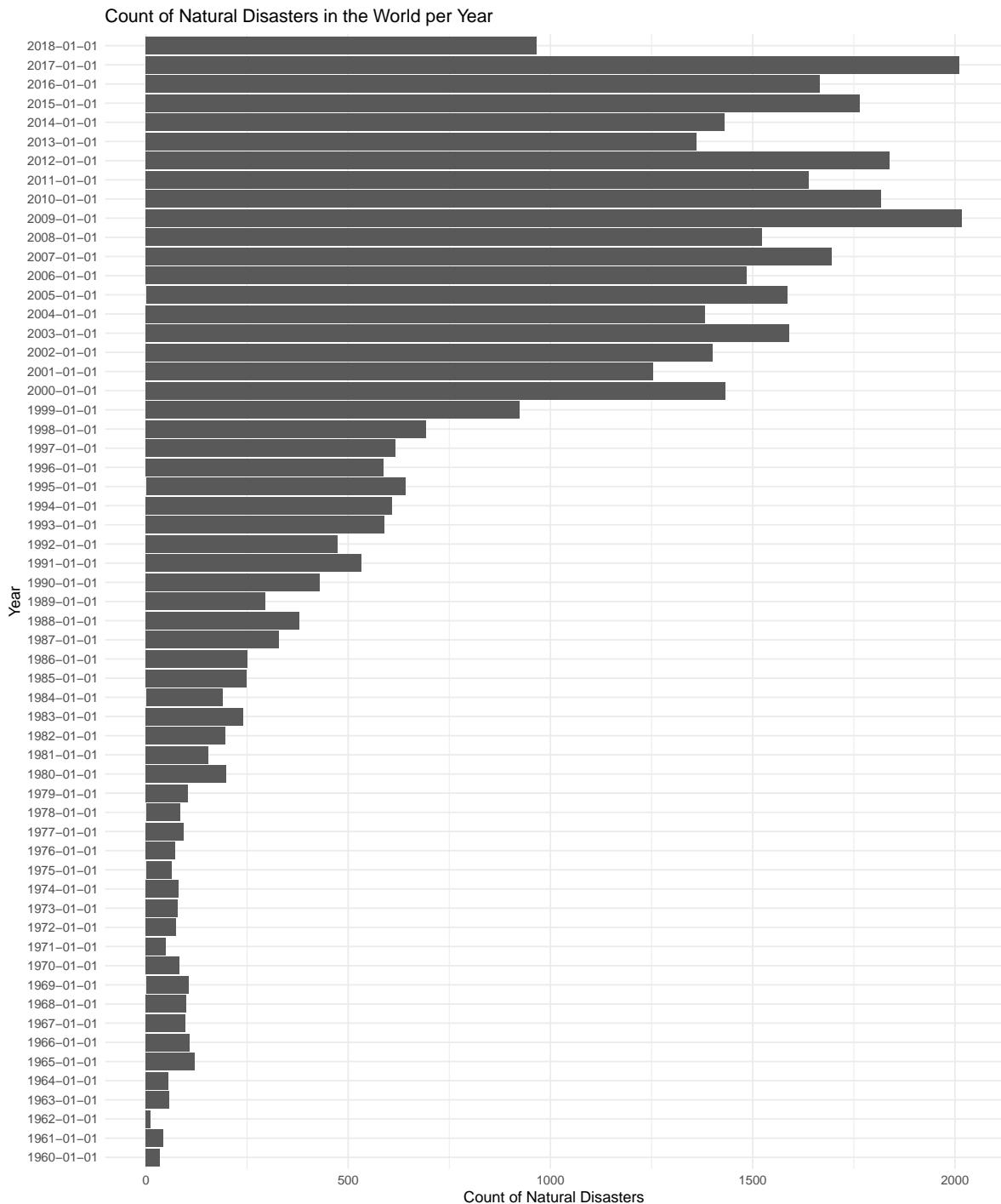
# filter world data to the U.S.
us <- nat_dis %>%
  filter(country == "United States")

# summarize data by type
nat_dis %>%
  group_by(disastertype) %>%
  summarize(disastertype_count = n()) %>%
  mutate(disastertype = fct_reorder(.f = disastertype, .x = disastertype_count)) %>%
# create a bar plot
  ggplot(aes(x = disastertype_count, y = disastertype)) +
  geom_col() +
  labs(x = "Count of Natural Disasters",
       y = "Type of Natural Disaster",
       title = "Count of All Natural Disasters in the World (1960-2018)") +
  theme_minimal()

```



```
# summarize data by year
nat_dis %>%
  group_by(year) %>%
  summarize(year_count = n()) %>%
# create a bar plot
  ggplot(aes(x = year_count, y = factor(year))) +
  geom_col() +
  labs(x = "Count of Natural Disasters",
       y = "Year",
       title = "Count of Natural Disasters in the World per Year") +
  theme_minimal()
```



```
# summarize data by counts
nat_dis %>%
```

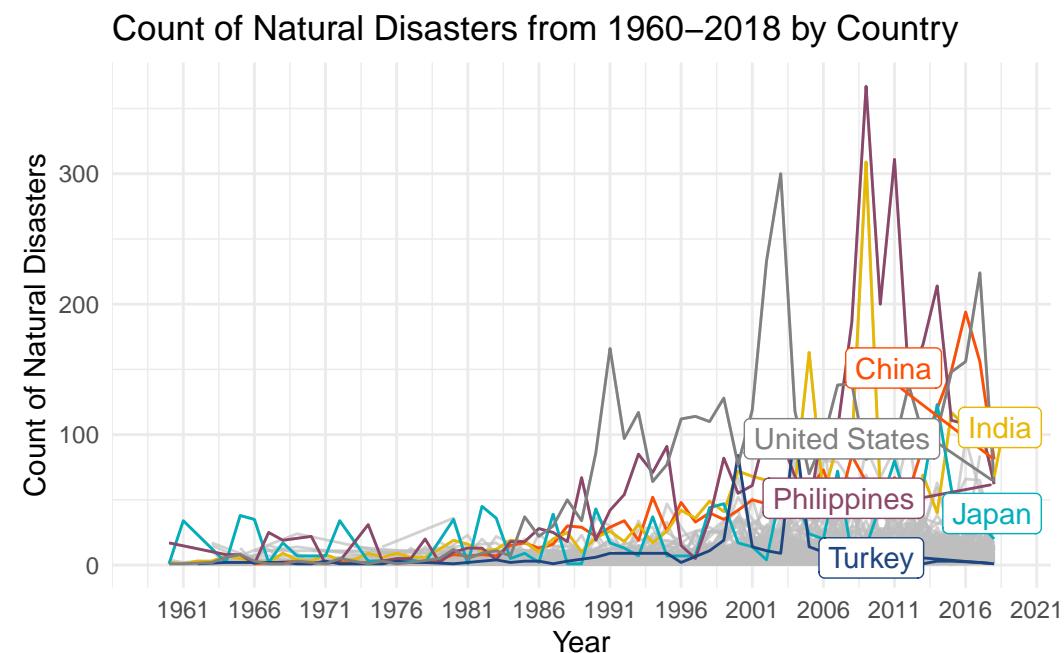
```

group_by(country, year) %>%
  summarise(disaster_count = n(),
            .groups = "drop") %>%
  group_by(country) %>%
  mutate(country_max = max(disaster_count), .groups = "drop") %>%
# create a line plot
ggplot(aes(x = year, y = disaster_count, group = country, color = country)) +
  geom_line() +
  gghighlight::gghighlight(country_max > 100) +
  scale_color_paletteer_d("ltc::expevo") +
  scale_x_date(limits = as.Date(c("1959-01-01", "2018-12-31")),
               date_breaks = "5 year",
               date_labels = "%Y") +
  labs(x = "Year",
       y = "Count of Natural Disasters",
       title = "Count of Natural Disasters from 1960–2018 by Country") +
  theme_minimal()

```

Warning: Tried to calculate with `group_by()`, but the calculation failed.
 Falling back to ungrouped filter operation...

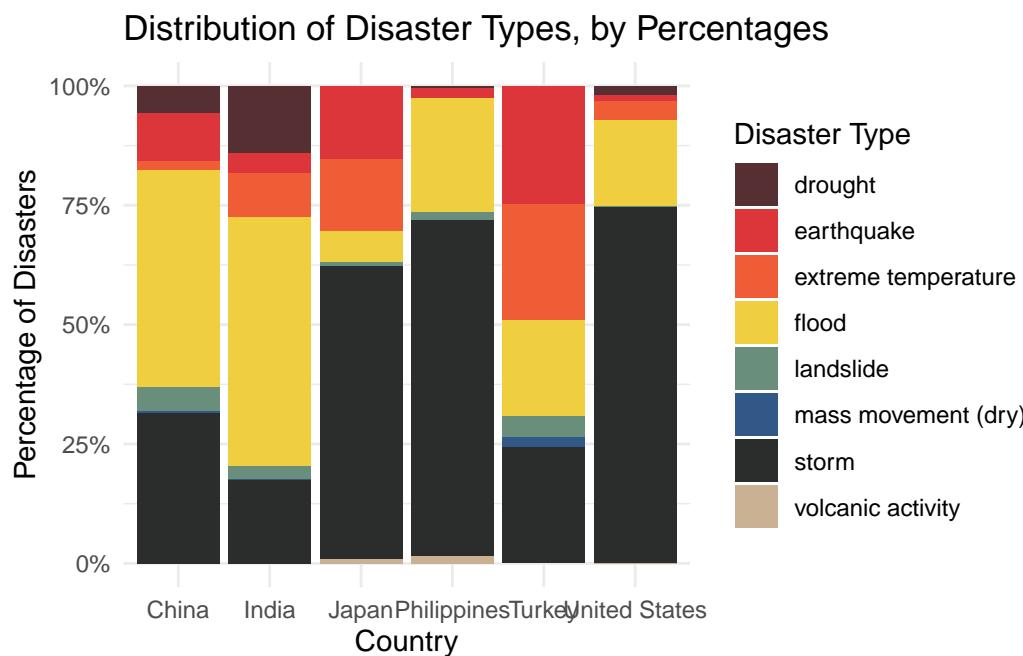
`label_key: country`



```

# summarize data by country and type
nat_dis %>%
  filter(country %in% c("United States", "Turkey", "China", "Philippines", "Japan", "India"))
  group_by(country, disastertype) %>%
  summarise(disastertype_count = n(), .groups = "drop") %>%
# create a stacked bar chart
  ggplot(aes(x = country, y = disastertype_count, fill = disastertype)) +
  geom_col(position = "fill") +
  scale_fill_palatteer_d("futurevisions::pegasi") +
  scale_y_continuous(labels = scales::label_percent(scale = 100)) +
  labs(x = "Country",
       y = "Percentage of Disasters",
       title = "Distribution of Disaster Types, by Percentages",
       fill = "Disaster Type") +
  theme_minimal()

```



1. What have you learned about your data? Have any potentially interesting patterns emerged? Point to specific visualizations that you created as you describe your findings. From this initial data exploration, I found that there are many interesting trends within the different countries and years. For example, seeing the different distributions of disaster types within each country. I also noticed that there are more data in the recent years, however it is interesting to see how the count of natural disasters increases with the years.

I think I want to try to create some kind of figure for my infographic that can highlight how the increase in natural disasters follows along with increasing climate change.

2. In FPM #1, you outlined some questions that you wanted to answer using these data. Have you made any strides towards answering those questions? If yes, how so? If no, what next steps do you need to take (e.g. I need to create X plot type, I still need to track down Y data, I need to restructure existing data so that you can visualize it in Z ways, etc.)? Have any new questions emerged?

Since I want to incorporate the emergence of climate change as an event, I'll have to find a source that identifies a date in time when climate change accelerates. I envision plotting this date as a breakpoint line or highlighting it across a timescale. Some final plots that would be helpful in visualizing the world data would be to create some sort of map, while I have some location data such as long and lat coordinates, it would be helpful to actually have geometries. The map I envision making is a choropleth map.

3. What challenges do you foresee encountering with your data? These can be data wrangling and / or visualization challenges

This dataset contains a lot of information, specifically information for many countries in the world. I think a problem when working with this data is going to be strategically picking which countries to focus on within my infographic. Since I want to be intentional with the countries I share, I want to identify countries that experience the effects of climate change to see if their natural disaster trends reflect climate change. This selection will be done through more data wrangling and calculations of basic statistics.