Data 607_Final Project

Jiaxin Zheng

2024-12-10

Research question:"Do higher temperatures result in more severe hurricanes or floods?"

In this Projet I will be exporting with two data sets.

- The NOAA Storm Events Dataset: it will provide detailed information about the natural disaster, including date, event type and location.
- The NASA Power API: Will supply climate temperature and precipitation to the locations and time periods of the disasters. (only can get one location at a time, So I will compare Temperature vs. Locations (Florida and New York))

By analysis these datasets will highlight the influence of climate change on Hurricanes or floods.

First: read the two cvs files from NOAA Storm Events Dataset

- 1.NOAA storm location dataset with Event_ID connect to the second dataset

```r
# Read NOAA storm location CSV file
df1 <- read.csv("https://raw.githubusercontent.com/Jennyjjxxzz/Data607_Fianl_Project/refs/heads/main/St

head(df1)
```

```
##   YEARMONTH EPISODE_ID EVENT_ID LOCATION_INDEX RANGE AZIMUTH       LOCATION
## 1    202303     176742  1074014              1  2.07       N ABERNATHY ARPT
## 2    202303     176742  1074015              1  1.49     SSW       PROSPECT
## 3    202303     176742  1074016              1  1.67       W      NASHVILLE
## 4    202303     176742  1074017              1  0.45     ENE GOODLETTSVILLE
## 5    202303     176742  1074018              1  2.19     SSW   MILLERSVILLE
## 6    202303     176742  1074019              1  0.00       N GOODLETTSVILLE
##   LATITUDE LONGITUDE    LAT2    LON2
## 1  35.1800  -87.0500 3510800  873000
## 2  35.0100  -87.0100   35600   87600
## 3  36.1700  -86.8100 3610200 8648600
## 4  36.3220  -86.7123 3619320 8642738
## 5  36.3394  -86.7100 3620364 8642600
## 6  36.3200  -86.7200 3619200 8643200
```

- 2. NOAA's detail CSV file with date, event type, source, state, begin latitude and longitude.

```r
# Read NOAA storm detail CSV file
df2 <- read.csv("https://raw.githubusercontent.com/Jennyjjxxzz/Data607_Fianl_Project/refs/heads/main/St

head(df2)
```

```
##   BEGIN_YEARMONTH BEGIN_DAY BEGIN_TIME END_YEARMONTH END_DAY END_TIME
## 1          202310        25        230        202310      27      551
## 2          202310        25        230        202310      27     1437
## 3          202310        25        230        202310      27     1126
## 4          202310        25        230        202310      27     1301
## 5          202310        25        230        202310      27      600
## 6          202310        25        230        202310      27      800
##   EPISODE_ID EVENT_ID        STATE YEAR MONTH_NAME EVENT_TYPE            SOURCE
## 1     186682  1145781 NORTH DAKOTA 2023    October Heavy Snow            Public
## 2     186682  1145783 NORTH DAKOTA 2023    October Heavy Snow   Law Enforcement
## 3     186682  1145784 NORTH DAKOTA 2023    October Heavy Snow            Public
## 4     186682  1145796 NORTH DAKOTA 2023    October Heavy Snow Emergency Manager
## 5     186682  1145884 NORTH DAKOTA 2023    October Heavy Snow          CoCoRaHS
## 6     186682  1145780 NORTH DAKOTA 2023    October Heavy Snow     COOP Observer
##   BEGIN_LAT BEGIN_LON END_LAT END_LON
## 1        NA        NA      NA      NA
## 2        NA        NA      NA      NA
## 3        NA        NA      NA      NA
## 4        NA        NA      NA      NA
## 5        NA        NA      NA      NA
## 6        NA        NA      NA      NA
```

Merge the two CSV files

```r
# Merge NOAA Locations and Details datasets
noaa_merged <- df1 %>%
  inner_join(df2, by = c("EPISODE_ID" = "EPISODE_ID", "EVENT_ID" = "EVENT_ID"))

head(noaa_merged)
```

```
##   YEARMONTH EPISODE_ID EVENT_ID LOCATION_INDEX RANGE AZIMUTH       LOCATION
## 1    202303     176742  1074014              1  2.07       N ABERNATHY ARPT
## 2    202303     176742  1074015              1  1.49     SSW       PROSPECT
## 3    202303     176742  1074016              1  1.67       W      NASHVILLE
## 4    202303     176742  1074017              1  0.45     ENE GOODLETTSVILLE
## 5    202303     176742  1074018              1  2.19     SSW   MILLERSVILLE
## 6    202303     176742  1074019              1  0.00       N GOODLETTSVILLE
##   LATITUDE LONGITUDE    LAT2    LON2 BEGIN_YEARMONTH BEGIN_DAY BEGIN_TIME
## 1  35.1800  -87.0500 3510800  873000          202303         3       1102
## 2  35.0100  -87.0100   35600   87600          202303         3       1112
## 3  36.1700  -86.8100 3610200 8648600          202303         3       1124
## 4  36.3220  -86.7123 3619320 8642738          202303         3       1137
## 5  36.3394  -86.7100 3620364 8642600          202303         3       1137
## 6  36.3200  -86.7200 3619200 8643200          202303         3       1137
##   END_YEARMONTH END_DAY END_TIME     STATE YEAR MONTH_NAME        EVENT_TYPE
## 1        202303       3     1102 TENNESSEE 2023      March Thunderstorm Wind
```

```
## 2        202303       3     1112 TENNESSEE 2023      March Thunderstorm Wind
## 3        202303       3     1124 TENNESSEE 2023      March Thunderstorm Wind
## 4        202303       3     1137 TENNESSEE 2023      March Thunderstorm Wind
## 5        202303       3     1137 TENNESSEE 2023      March Thunderstorm Wind
## 6        202303       3     1137 TENNESSEE 2023      March Thunderstorm Wind
##              SOURCE BEGIN_LAT BEGIN_LON END_LAT  END_LON
## 1 Emergency Manager   35.1800  -87.0500 35.1800 -87.0500
## 2            Public   35.0100  -87.0100 35.0100 -87.0100
## 3 Emergency Manager   36.1700  -86.8100 36.1700 -86.8100
## 4            Public   36.3220  -86.7123 36.3220 -86.7123
## 5   Trained Spotter   36.3394  -86.7100 36.3394 -86.7100
## 6            Public   36.3200  -86.7200 36.3200 -86.7200
```

```
#Create the date column (there are repeat date columns)
noaa_clean <- noaa_merged %>%
  mutate(
    BEGIN_DATE = as.Date(paste0(BEGIN_YEARMONTH, sprintf("%02d", BEGIN_DAY)), format = "%Y%m%d")
  )
```

```
#Select columns for final clean dataset
noaa_clean <- noaa_clean %>%
  select(
    BEGIN_DATE, STATE, EVENT_TYPE, YEAR, MONTH_NAME, BEGIN_LAT, BEGIN_LON
  )

head(noaa_clean)
```

```
##   BEGIN_DATE     STATE        EVENT_TYPE YEAR MONTH_NAME BEGIN_LAT BEGIN_LON
## 1 2023-03-03 TENNESSEE Thunderstorm Wind 2023      March   35.1800  -87.0500
## 2 2023-03-03 TENNESSEE Thunderstorm Wind 2023      March   35.0100  -87.0100
## 3 2023-03-03 TENNESSEE Thunderstorm Wind 2023      March   36.1700  -86.8100
## 4 2023-03-03 TENNESSEE Thunderstorm Wind 2023      March   36.3220  -86.7123
## 5 2023-03-03 TENNESSEE Thunderstorm Wind 2023      March   36.3394  -86.7100
## 6 2023-03-03 TENNESSEE Thunderstorm Wind 2023      March   36.3200  -86.7200
```

Second: pull out the data from NASA Power API (Only can pull out one location at a time). (This is for Florida)

```
# Florida NASA Power API URL
fl_url <- paste0("https://power.larc.nasa.gov/api/temporal/hourly/point?",
                 "start=20230101&end=20231231&latitude=27.9944&longitude=-81.7603",
                 "&community=ag&parameters=T2M&format=json&header=true&time-standard=lst")
```

```
# Fetch Florida climate data
fl_response <- httr::GET(fl_url)
```

```
# Parse JSON response
fl_climate_data <- fromJSON(content(fl_response, "text"), flatten = TRUE)
```

```
## No encoding supplied: defaulting to UTF-8.
```

```r
# Extract temperature data (T2M) for Florida
fl_t2m_data <- fl_climate_data$properties$parameter$T2M
fl_climate_df <- tibble(
  Date = names(fl_t2m_data),
  T2M = unlist(fl_t2m_data)
) %>%
  mutate(
    Date = as.Date(substr(Date, 1, 8), format = "%Y%m%d"),
    State = "FLORIDA"
  )

head(fl_climate_df)
```

```
## # A tibble: 6 x 3
##   Date         T2M State
##   <date>     <dbl> <chr>
## 1 2023-01-01  19.1 FLORIDA
## 2 2023-01-01  18.9 FLORIDA
## 3 2023-01-01  18.8 FLORIDA
## 4 2023-01-01  18.6 FLORIDA
## 5 2023-01-01  18.5 FLORIDA
## 6 2023-01-01  18.3 FLORIDA
```

Get Url_2 temperature data for New York.

```r
# New York NASA Power API URL
ny_url <- paste0("https://power.larc.nasa.gov/api/temporal/hourly/point?",
                 "start=20230101&end=20231231&latitude=40.7128&longitude=-74.0060",
                 "&community=ag&parameters=T2M&format=json&header=true&time-standard=lst")
```

```r
# Fetch New York climate data
ny_response <- httr::GET(ny_url)
```

```r
# Parse JSON response
ny_climate_data <- fromJSON(content(ny_response, "text"), flatten = TRUE)
```

```
## No encoding supplied: defaulting to UTF-8.
```

```r
# Extract temperature data (T2M) for New York
ny_t2m_data <- ny_climate_data$properties$parameter$T2M
ny_climate_df <- tibble(
  Date = names(ny_t2m_data),
  T2M = unlist(ny_t2m_data)
) %>%
  mutate(
    Date = as.Date(substr(Date, 1, 8), format = "%Y%m%d"),
    State = "NEW YORK"
  )

head(ny_climate_df)
```

```
## # A tibble: 6 x 3
##   Date         T2M State
##   <date>     <dbl> <chr>
## 1 2023-01-01  8.04 NEW YORK
## 2 2023-01-01  7.69 NEW YORK
## 3 2023-01-01  7.43 NEW YORK
## 4 2023-01-01  7.04 NEW YORK
## 5 2023-01-01  6.87 NEW YORK
## 6 2023-01-01  6.61 NEW YORK
```

```r
# Combine Florida and New York climate datasets as one
climate_df <- bind_rows(fl_climate_df, ny_climate_df)

head(climate_df)
```

```
## # A tibble: 6 x 3
##   Date         T2M State
##   <date>     <dbl> <chr>
## 1 2023-01-01  19.1 FLORIDA
## 2 2023-01-01  18.9 FLORIDA
## 3 2023-01-01  18.8 FLORIDA
## 4 2023-01-01  18.6 FLORIDA
## 5 2023-01-01  18.5 FLORIDA
## 6 2023-01-01  18.3 FLORIDA
```

```r
# Check for any na data values
colSums(is.na(noaa_clean))
```

```
## BEGIN_DATE      STATE EVENT_TYPE       YEAR MONTH_NAME  BEGIN_LAT  BEGIN_LON
##          0          0          0          0          0          0          0
```

```r
colSums(is.na(climate_df))
```

```
##  Date   T2M State
##     0     0     0
```

Merge NOAA and NASA datasets as final dataset

```r
# Merge NOAA and NASA datasets
final_merged <- noaa_clean %>%
  inner_join(climate_df, by = c("BEGIN_DATE" = "Date", "STATE" = "State"))
```

```
## Warning in inner_join(., climate_df, by = c(BEGIN_DATE = "Date", STATE = "State")): Detected an unexp
## i Row 151 of ‘x‘ matches multiple rows in ‘y‘.
## i Row 3697 of ‘y‘ matches multiple rows in ‘x‘.
## i If a many-to-many relationship is expected, set ‘relationship =
##   "many-to-many"‘ to silence this warning.
```

```r
head(final_merged)
```

```
##   BEGIN_DATE   STATE EVENT_TYPE YEAR MONTH_NAME BEGIN_LAT BEGIN_LON   T2M
## 1 2023-06-04 FLORIDA       Hail 2023       June     28.14    -82.31 22.32
## 2 2023-06-04 FLORIDA       Hail 2023       June     28.14    -82.31 22.05
## 3 2023-06-04 FLORIDA       Hail 2023       June     28.14    -82.31 21.79
## 4 2023-06-04 FLORIDA       Hail 2023       June     28.14    -82.31 21.55
## 5 2023-06-04 FLORIDA       Hail 2023       June     28.14    -82.31 21.40
## 6 2023-06-04 FLORIDA       Hail 2023       June     28.14    -82.31 21.34
```

```r
df_final <- final_merged %>%
  select(-c(YEAR, MONTH_NAME))

head(df_final)
```

```
##   BEGIN_DATE   STATE EVENT_TYPE BEGIN_LAT BEGIN_LON   T2M
## 1 2023-06-04 FLORIDA       Hail     28.14    -82.31 22.32
## 2 2023-06-04 FLORIDA       Hail     28.14    -82.31 22.05
## 3 2023-06-04 FLORIDA       Hail     28.14    -82.31 21.79
## 4 2023-06-04 FLORIDA       Hail     28.14    -82.31 21.55
## 5 2023-06-04 FLORIDA       Hail     28.14    -82.31 21.40
## 6 2023-06-04 FLORIDA       Hail     28.14    -82.31 21.34
```

```r
colnames(df_final) <- c("Date", "State", "EventType", "Begin_Lat", "Begin_Lon", "Temperature")

head(df_final)
```

```
##         Date   State EventType Begin_Lat Begin_Lon Temperature
## 1 2023-06-04 FLORIDA      Hail     28.14    -82.31       22.32
## 2 2023-06-04 FLORIDA      Hail     28.14    -82.31       22.05
## 3 2023-06-04 FLORIDA      Hail     28.14    -82.31       21.79
## 4 2023-06-04 FLORIDA      Hail     28.14    -82.31       21.55
## 5 2023-06-04 FLORIDA      Hail     28.14    -82.31       21.40
## 6 2023-06-04 FLORIDA      Hail     28.14    -82.31       21.34
```

In this data, we see the avg_temperature and total_events happened in state Florida and New York

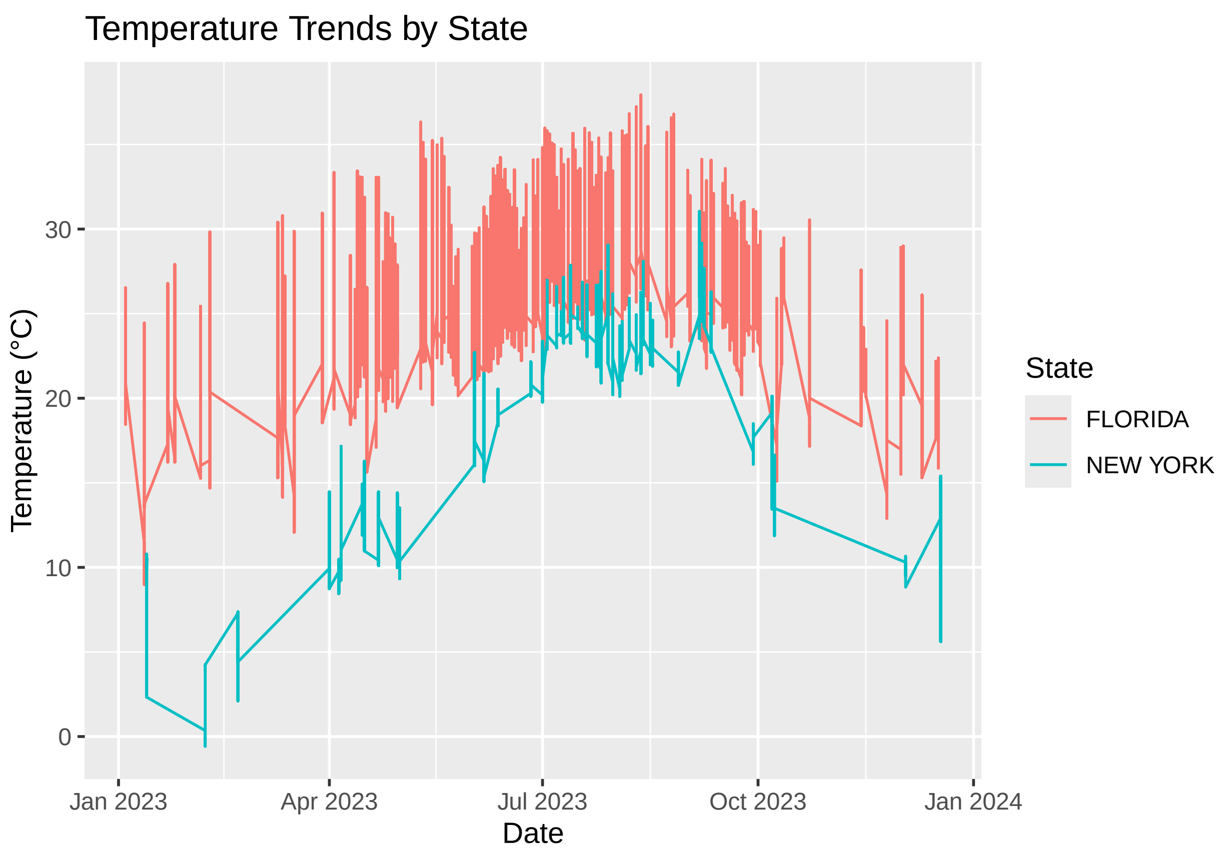
```r
# Count events by state and event type
df_final %>%
  group_by(State, EventType) %>%
  summarize(
    Total_Events = n(),
    Avg_Temperature = mean(Temperature, na.rm = TRUE)
  )
```

```
## `summarise()` has grouped output by 'State'. You can override using the
## `.groups` argument.
```

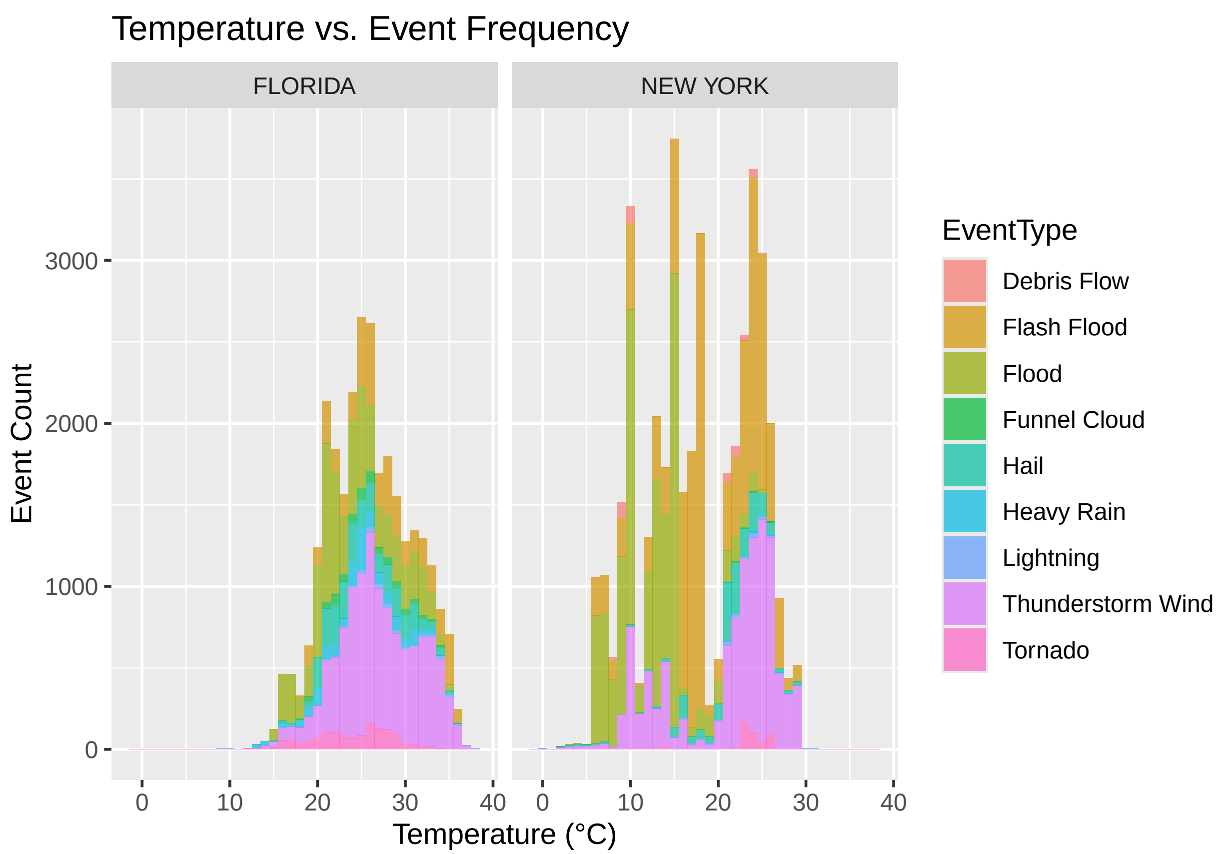
```
## # A tibble: 17 x 4
## # Groups:   State [2]
##    State    EventType         Total_Events Avg_Temperature
##    <chr>    <chr>                    <int>           <dbl>
##  1 FLORIDA  Flash Flood               3840            27.4
##  2 FLORIDA  Flood                     7152            23.9
##  3 FLORIDA  Funnel Cloud               648            25.9
##  4 FLORIDA  Hail                      2472            25.4
##  5 FLORIDA  Heavy Rain                1440            24.3
##  6 FLORIDA  Lightning                  264            28.8
##  7 FLORIDA  Thunderstorm Wind        11064            27.2
##  8 FLORIDA  Tornado                   1392            23.9
##  9 NEW YORK Debris Flow                432            16.1
## 10 NEW YORK Flash Flood              15504            19.1
## 11 NEW YORK Flood                    11808            12.2
## 12 NEW YORK Funnel Cloud                48            22.9
## 13 NEW YORK Hail                      1944            21.0
## 14 NEW YORK Heavy Rain                  24            24.3
## 15 NEW YORK Lightning                  168            19.7
## 16 NEW YORK Thunderstorm Wind        10464            21.0
## 17 NEW YORK Tornado                    480            22.9
```

Florida's temperature is warmer than New York, but New York's natural disasters are more frequent.

```r
ggplot(df_final, aes(x = Date, y = Temperature, color = State)) +
  geom_line() +
  labs(title = "Temperature Trends by State", x = "Date", y = "Temperature (°C)")
```




```r
ggplot(df_final, aes(x = Temperature, fill = EventType)) +
  geom_histogram(binwidth = 1, alpha = 0.7) +
  facet_wrap(~State) +
  labs(title = "Temperature vs. Event Frequency", x = "Temperature (°C)", y = "Event Count")
```



Now, we focus on hurricanes (typhoons) and floods in the year 2023 to examine whether higher temperatures influence the frequency of these events.

- There is no hurricanes (typhoons) were recorded in the New York and Florida areas during this period

```
# filter the hurricanes (typhoons) and flood

filtered_data <- df_final %>%
  filter(EventType %in% c("Hurricane(Typhoon)", "Flood"))

head(filtered_data)
```

```
##         Date   State EventType Begin_Lat Begin_Lon Temperature
## 1 2023-11-15 FLORIDA     Flood   25.9265  -80.3477       21.40
## 2 2023-11-15 FLORIDA     Flood   25.9265  -80.3477       21.28
## 3 2023-11-15 FLORIDA     Flood   25.9265  -80.3477       21.13
## 4 2023-11-15 FLORIDA     Flood   25.9265  -80.3477       21.04
## 5 2023-11-15 FLORIDA     Flood   25.9265  -80.3477       20.99
## 6 2023-11-15 FLORIDA     Flood   25.9265  -80.3477       20.91
```

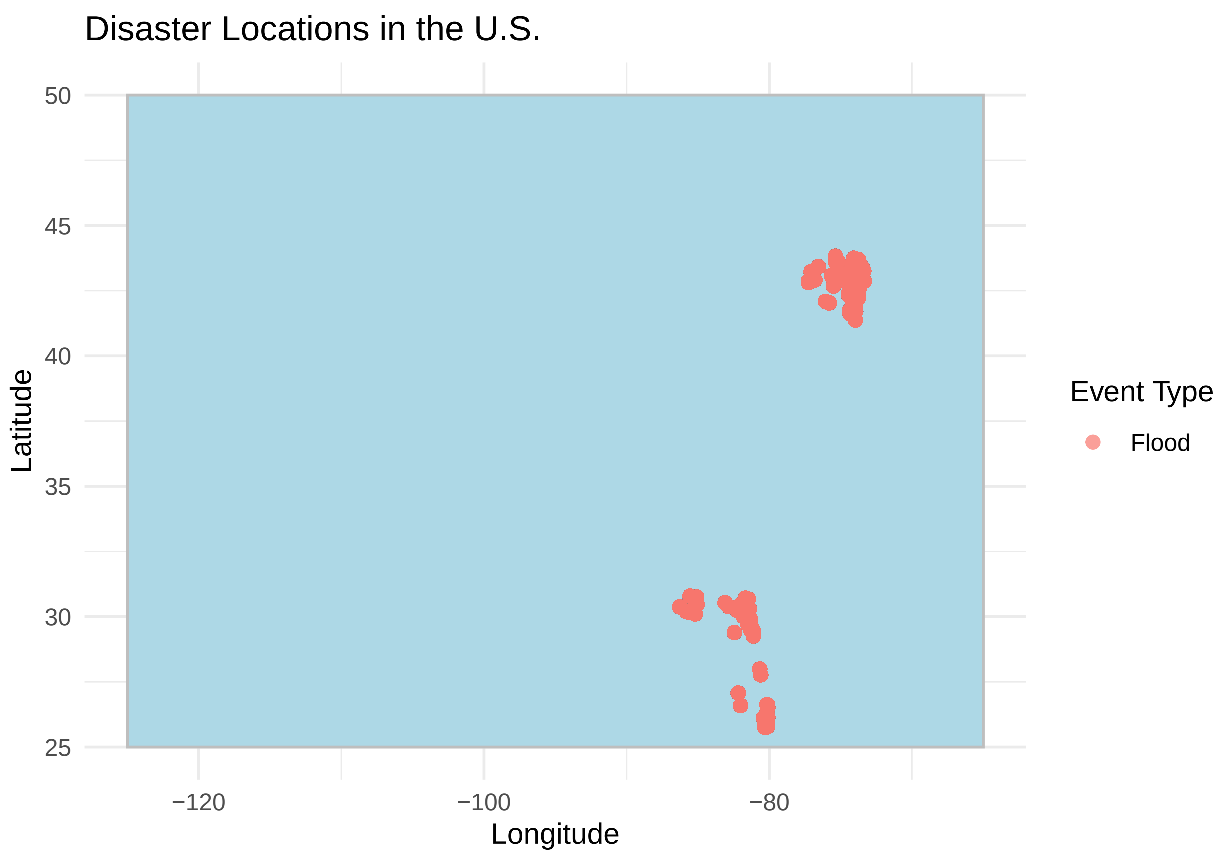
```
# Count floods by state
flood_count_by_state <- filtered_data %>%
  filter(EventType == "Flood") %>%
  group_by(State) %>%
```

```
  summarize(Count = n())

print(flood_count_by_state)
```

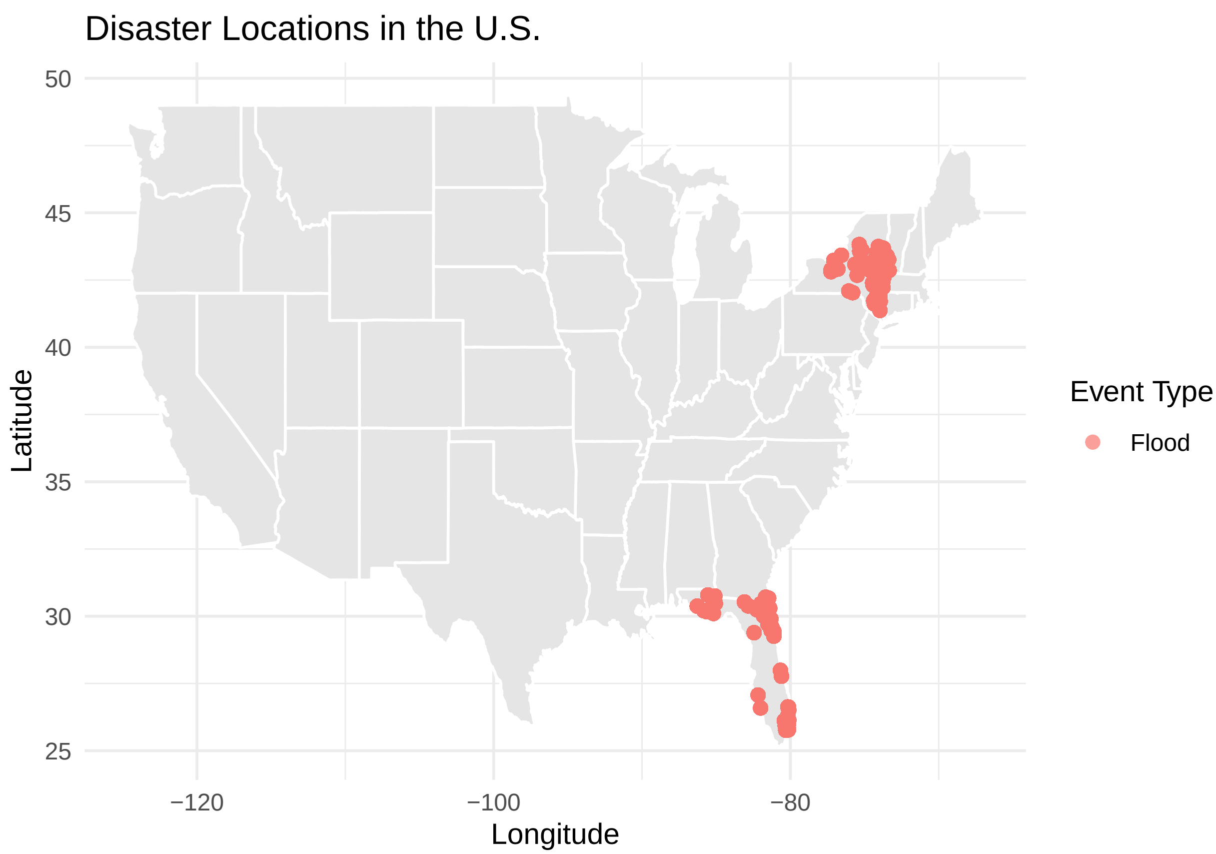
```
## # A tibble: 2 x 2
##   State    Count
##   <chr>    <int>
## 1 FLORIDA   7152
## 2 NEW YORK 11808
```

Use the map to show the temperature vs flood

```
# Define the bounding box for the U.S.
bbox <- c(xmin = -125, xmax = -65, ymin = 25, ymax = 50)

# Create a basic map
ggplot() +
  # Add a rectangular base map
  geom_rect(
    aes(xmin = bbox[1], xmax = bbox[2], ymin = bbox[3], ymax = bbox[4]),
    fill = "lightblue", color = "gray"
  ) +
  # Plot disaster points
  geom_point(
    data = filtered_data,
    aes(x = Begin_Lon, y = Begin_Lat, color = EventType),
    alpha = 0.7, size = 2
  ) +
  labs(
    title = "Disaster Locations in the U.S.",
    x = "Longitude",
    y = "Latitude",
    color = "Event Type"
  ) +
  theme_minimal()
```




```r
# US map
us_map <- map_data("state")

ggplot() +
  # Add state borders
  geom_polygon(
    data = us_map, aes(x = long, y = lat, group = group),
    fill = "gray90", color = "white"
  ) +

  geom_point(
    data = filtered_data,
    aes(x = Begin_Lon, y = Begin_Lat, color = EventType),
    alpha = 0.7, size = 2
  ) +
  labs(
    title = "Disaster Locations in the U.S.",
    x = "Longitude",
    y = "Latitude",
    color = "Event Type"
  ) +
  theme_minimal()
```



Use cor.test function to assess the relationship between two variables, and test the significance

```
frequency_data <- filtered_data %>%
  group_by(Temperature, State, EventType) %>%
  summarize(Event_Count = n(), .groups = "drop")

head(frequency_data)
```

```
## # A tibble: 6 x 4
##   Temperature State    EventType Event_Count
##         <dbl> <chr>    <chr>           <int>
## 1        2.33 NEW YORK Flood               4
## 2        2.58 NEW YORK Flood               4
## 3        3.09 NEW YORK Flood               4
## 4        3.78 NEW YORK Flood               4
## 5        4.45 NEW YORK Flood               4
## 6        4.96 NEW YORK Flood               4
```

The p-value is so small. We reject null hypothesis, there is no correlation between temperature and event frequency.

```r
cor.test(frequency_data$Temperature, frequency_data$Event_Count)
```

```
##
##  Pearson's product-moment correlation
##
## data:  frequency_data$Temperature and frequency_data$Event_Count
## t = -8.4649, df = 805, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3480424 -0.2212493
## sample estimates:
##        cor
## -0.2858967
```

Summary:

The analysis supports that higher temperatures do not result in more frequent natural disaster. Florida has warmer climate with higher average temperatures compared to New York, but in year 2023, New York experienced more floods(11,808 events) than Florida (7,152 events). Temperature may be one influencing factor, but flood frequency appears to be determined by a combination of climatic and environmental factors.

Challenges:

This project was particularly challenging for me because it required merging two separate NOAA CSV files and merging them with the NASA API dataset. Additionally, creating a map plot to visualize the data was a new and complex task that pushed me to learn and apply new skills.