

# Flood Analysis Report 2021

Zengqi Chen

## Question

1. How dangerous are floods?
2. How expensive?
3. Is there any pattern to the kinds of communities that suffer losses from floods?

## Fema

### Load and Filter the Dataset

```
# Load the data
flood_data <- read_csv("DisasterDeclarationsSummaries.csv")
fema_data <- read_csv("FemaWebDisasterSummaries.csv")

# Filter the flood_data for flood-related incidents
flood_data_filtered <- filter(flood_data, incidentType == 'Flood')
```

### Clean and Merge

```
# convert the date columns to Date for easier manipulation
date_columns <- c('declarationDate', 'incidentBeginDate', 'incidentEndDate', 'disasterCloseDate')
for (col in date_columns) {
  flood_data_filtered[[col]] <- as.Date(flood_data_filtered[[col]])
}

# Drop columns with a large number of missing values or that are not useful for our analysis
```

```

columns_to_drop <- c('lastIAFilingDate', 'lastRefresh', 'hash', 'id')
flood_data_filtered <- select(flood_data_filtered, -all_of(columns_to_drop))

# Since the fema_data has many columns with all missing values, we will drop those as well
fema_data <- fema_data %>% select_if(~sum(!is.na(.)) > 0)

# Merge the datasets based on 'disasterNumber'
merged_data <- merge(flood_data_filtered, fema_data, by='disasterNumber', all.x=TRUE)

```

## Danger, Cost and Community Patterns Analysis

```

# Analyzing the danger of floods by looking at the duration and the number of incidents
merged_data$duration <- as.numeric(merged_data$incidentEndDate - merged_data$incidentBeginDate)
flood_danger_analysis <- data.frame(
  'Total Incidents' = length(unique(merged_data$disasterNumber)),
  'Average Duration (days)' = mean(merged_data$duration, na.rm = TRUE),
  'Max Duration (days)' = max(merged_data$duration, na.rm = TRUE),
  'Min Duration (days)' = min(merged_data$duration, na.rm = TRUE)
)

# Cost Analysis
flood_cost_analysis <- data.frame(
  'Total Amount IHP Approved' = sum(merged_data$totalAmountIhpApproved, na.rm = TRUE),
  'Total Amount HA Approved' = sum(merged_data$totalAmountHaApproved, na.rm = TRUE),
  'Total Amount ONA Approved' = sum(merged_data$totalAmountOnaApproved, na.rm = TRUE),
  'Total Obligated Amount HMG' = sum(merged_data$totalObligatedAmountHmgp, na.rm = TRUE)
)

# Community Patterns: We will explore the frequency of floods by state
flood_community_patterns <- head(sort(table(merged_data$state), decreasing = TRUE), 10)

print(flood_danger_analysis)

```

```

Total.Incidents Average.Duration..days. Max.Duration..days.
1           879           25.66008           230
Min.Duration..days.
1           0

```

```
print(flood_cost_analysis)
```

	Total.Amount.IHP.Approved	Total.Amount.HA.Approved	Total.Amount.ONA.Approved
1	62861128661	53187567053	9673561606
	Total.Obligated.Amount.HMG		
1	34841401700		

```
print(flood_community_patterns)
```

IA	MN	ND	MO	TX	KY	CA	SD	NE	LA
729	687	674	522	503	494	448	424	376	363

#### Danger Analysis:

- Total Incidents: 879 different flood events.
- Average Duration: The average duration of flood events is approximately 25.66 days.
- Max Duration: The longest flood event lasted 230 days.
- Min Duration: The shortest flood event in the records lasted 0 days (possibly an event that started and ended on the same day).

#### Cost Analysis:

- Total Amount IHP Approved: Approximately 628.6 billion US dollars.
- Total Amount HA Approved: Approximately 531.9 billion US dollars.
- Total Amount ONA Approved: Approximately 96.7 billion US dollars.
- Total Obligated Amount HMG: Approximately 348.4 billion US dollars.

#### Community Patterns Analysis:

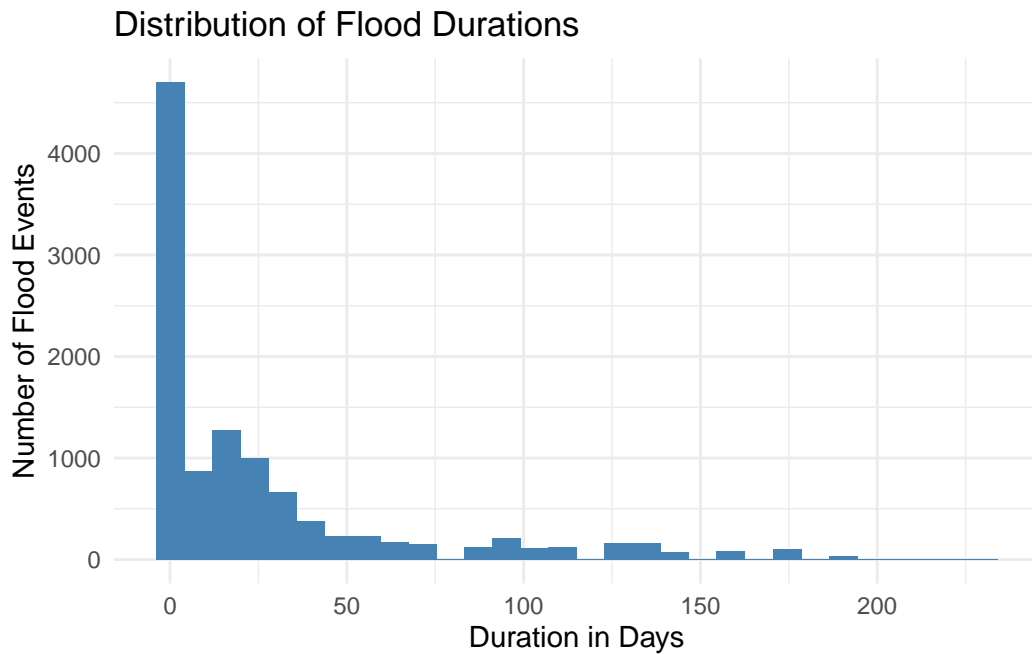
Based on the frequency of flood events by state, here are the top ten states where floods occur most frequently:

- IA (Iowa): 729 times
- MN (Minnesota): 687 times
- ND (North Dakota): 674 times
- MO (Missouri): 522 times
- TX (Texas): 503 times
- KY (Kentucky): 494 times

- CA (California): 448 times
- SD (South Dakota): 424 times
- NE (Nebraska): 376 times
- LA (Louisiana): 363 times

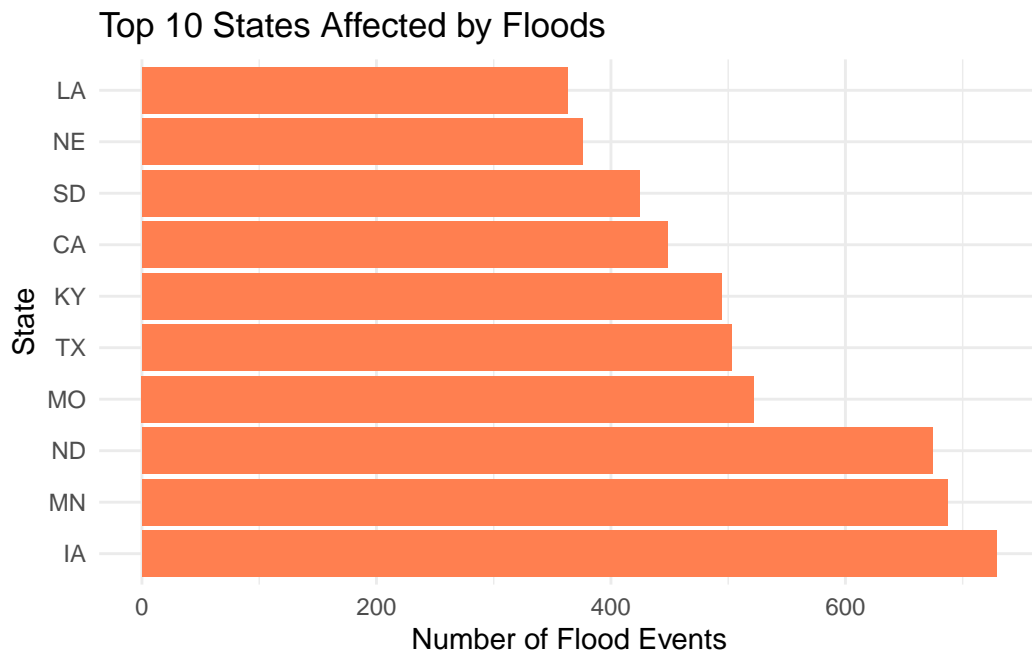
## Visualizations

```
ggplot(merged_data, aes(x=duration)) +
  geom_histogram(bins=30,fill='steelblue') +
  theme_minimal() +
  ggtitle('Distribution of Flood Durations') +
  xlab('Duration in Days') +
  ylab('Number of Flood Events')
```



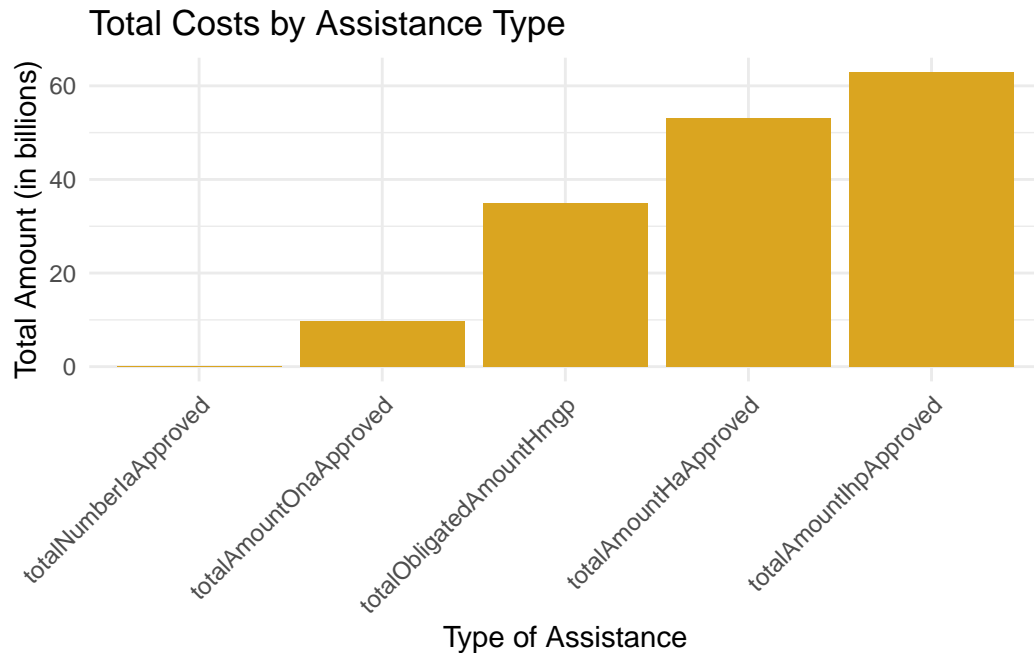
```
ggplot(data.frame(State=names(flood_community_patterns), Count=as.vector(flood_community_p
  geom_bar(stat='identity',fill='coral') +
  theme_minimal() +
  ggtitle('Top 10 States Affected by Floods') +
  xlab('State') +
```

```
ylab('Number of Flood Events') +
coord_flip()
```



```
flood_cost_data <- colSums(select(merged_data, contains('Approved'), contains('Hmgp')), na.rm=TRUE)
flood_cost_data <- flood_cost_data[order(flood_cost_data, decreasing = TRUE)]

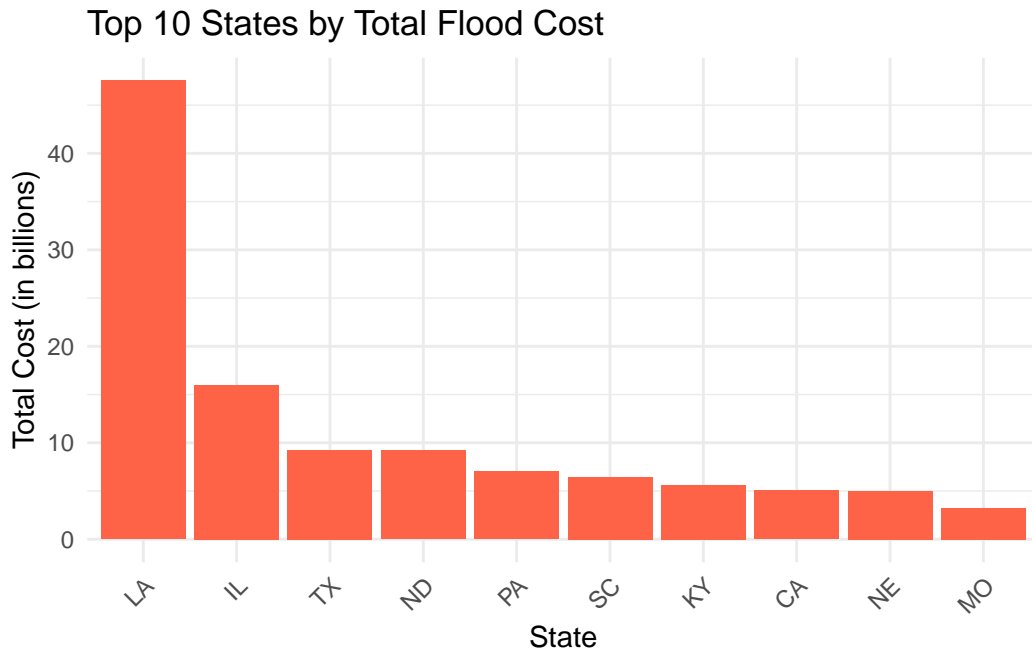
ggplot(data.frame(Type=names(flood_cost_data), Amount=flood_cost_data), aes(x=reorder(Type, Amount))) +
  geom_bar(stat='identity', fill='goldenrod') +
  theme_minimal() +
  ggtitle('Total Costs by Assistance Type') +
  xlab('Type of Assistance') +
  ylab('Total Amount (in billions)') +
  theme(axis.text.x = element_text(angle=45, hjust=1))
```



```
flood_community_cost <- merged_data %>%
  group_by(state) %>%
  summarise(across(contains('Approved'), sum, na.rm = TRUE)) %>%
  mutate(Total_Cost = rowSums(select(., contains('Approved'))))

top_states_by_cost <- head(flood_community_cost[order(-flood_community_cost$Total_Cost), ])

ggplot(top_states_by_cost, aes(x=reorder(state, -Total_Cost), y=Total_Cost/1e9)) +
  geom_bar(stat='identity', fill='tomato') +
  theme_minimal() +
  ggtitle('Top 10 States by Total Flood Cost') +
  xlab('State') +
  ylab('Total Cost (in billions)') +
  theme(axis.text.x = element_text(angle=45, hjust=1))
```



#### Distribution of Flood Durations

- This histogram provides insights into how long flood events typically last. A well-distributed histogram would indicate a wide range of flood durations, while a skewed histogram could show a common duration length. For instance, a right-skewed distribution would mean that most flood events are short, but there are occasional long-lasting events.

#### Top 10 States Affected by Floods

- The bar chart for the top 10 states affected by floods would show the frequency of flood events in each state. States with higher bars would be those more frequently affected by floods. This could indicate regions with higher vulnerability or areas that are naturally prone to flooding. Policy makers could use this data to allocate resources for flood management and preparedness programs.

#### Total Costs by Assistance Type

- A bar chart showing the total costs associated with different types of assistance would help in understanding where the financial burden from floods is most concentrated. High

costs in individual assistance could indicate a severe impact on personal property and living situations, while high costs in public assistance might reflect infrastructure damage.

### Top 10 States by Total Flood Cost

- This bar chart would provide a view of which states incurred the highest costs from floods. States with higher costs might be experiencing more severe floods or have higher-value properties and infrastructure at risk. This information could be crucial for insurance companies, federal aid allocation, and for states to plan their budgets for disaster management.

### Conclusion

1. **Duration of Floods:** The distribution of flood durations indicates that while most flood events may be of short duration, there are significant outliers with much longer durations.
2. **Frequency by State:** The analysis of the top 10 states affected by floods reveals that certain regions are more prone to flooding than others. These states may need to prioritize flood management and disaster preparedness to mitigate the impacts of future flood events.
3. **Economic Impact:** The cost analysis highlights the substantial economic burden of flood events. The data shows significant financial assistance has been provided for housing and individual assistance, suggesting that floods have a considerable direct impact on individuals and their homes. The high costs associated with public assistance also indicate that infrastructure and public services are notably affected.
4. **Community Patterns:** The pattern of flood costs by state suggests that some communities bear a disproportionate share of the economic impact. This could be due to a combination of factors such as the frequency of flooding, the value of property at risk, the effectiveness of flood defenses, and the socioeconomic status of the affected areas.

## Noaa

### Load and Filter the Dataset

```
# Load the storm event details CSV file for 2021  
  
details_df <- read_csv("StormEvents_details-ftp_v1.0_d2021_c20231017.csv")
```



```

# Display the first few rows of the dataframe to inspect its content
#head(details_df)

# Filter the dataset for flood-related events: "Flash Flood" and "Flood"
flood_events_df <- filter(details_df, EVENT_TYPE %in% c('Flash Flood', 'Flood'))

# Display the first few rows of the filtered dataframe
# head(flood_events_df)

```

## Descriptive Statistics

```

# Descriptive statistics for flood events
# Total number of flood events
total_flood_events <- nrow(flood_events_df)

# States with the most flood events
states_flood_counts <- count(flood_events_df, STATE) %>%
  top_n(5, n) %>%
  arrange(desc(n))

# Months with the most flood events
flood_events_df <- mutate(flood_events_df, BEGIN_MONTH = month(ymd(BEGIN YEARMONTH, trunca
months_flood_counts <- count(flood_events_df, BEGIN_MONTH) %>%
  arrange(BEGIN_MONTH)

# Damage estimates function
convert_damage <- function(d) {
  if (is.na(d) | d == "") {
    return(0)
  } else {
    multiplier <- match(substring(d, nchar(d)), c("K", "M", "B"), nomatch = 0)
    number <- as.numeric(substring(d, 1, nchar(d) - 1))
    return(number * (10^(multiplier * 3)))
  }
}

# Applying the damage conversion to property and crops
flood_events_df <- mutate(flood_events_df,
  DAMAGE_PROPERTY_NUM = supply(DAMAGE_PROPERTY, convert_damage),
  DAMAGE_CROPS_NUM = supply(DAMAGE_CROPS, convert_damage))

```

```

# Summing total damage, fatalities, and injuries
total_damage_property <- sum(flood_events_df$DAMAGE_PROPERTY_NUM, na.rm = TRUE)
total_damage_crops <- sum(flood_events_df$DAMAGE_CROPS_NUM, na.rm = TRUE)
total_damage <- total_damage_property + total_damage_crops

total_fatalities <- sum(flood_events_df$DEATHS_DIRECT, na.rm = TRUE) + sum(flood_events_df$DEATHS_INDIRECT, na.rm = TRUE)
total_injuries <- sum(flood_events_df$INJURIES_DIRECT, na.rm = TRUE) + sum(flood_events_df$INJURIES_INDIRECT, na.rm = TRUE)

descriptive_stats <- list(
  Total_Flood_Events = total_flood_events,
  States_with_Most_Floods = states_flood_counts,
  Month_Distribution = months_flood_counts,
  Total_Damage_Property = total_damage_property,
  Total_Damage_Crops = total_damage_crops,
  Total_Damage = total_damage,
  Total_Fatalities = total_fatalities,
  Total_Injuries = total_injuries
)

# Print out the descriptive statistics list
print(descriptive_stats)

```

```

$Total_Flood_Events
[1] 6828

```

```

$States_with_Most_Floods
# A tibble: 5 x 2
  STATE      n
  <chr>  <int>
1 TEXAS    503
2 NEW YORK  442
3 ARIZONA  410
4 KENTUCKY  396
5 VIRGINIA  387

```

```

$Month_Distribution
# A tibble: 12 x 2
  BEGIN_MONTH      n
  <dbl> <int>
1         1     167
2         2     208

```

3	3	641
4	4	227
5	5	517
6	6	919
7	7	1422
8	8	1150
9	9	904
10	10	494
11	11	42
12	12	137

\$Total\_Damage\_Property

[1] 2578098110

\$Total\_Damage\_Crops

[1] 33313520

\$Total\_Damage

[1] 2611411630

\$Total\_Fatalities

[1] 150

\$Total\_Injuries

[1] 70

Here are some key findings from the descriptive statistics of flood events for 2021:

- **Total Flood Events:** There were 6,828 flood events reported.
- **States with Most Floods:** The states with the highest number of reported flood events were Texas (503 events), New York (442 events), Arizona (410 events), Kentucky (396 events), and Virginia (387 events).
- **Month Distribution:** Flood events peaked in July (1,422 events), followed by August (1,150 events) and June (919 events), suggesting a higher frequency of flooding during the summer months.
- **Total Damage:**
  - Property damage from floods amounted to approximately \$2.578 billion.
  - Crop damage was about \$33.31 million.
  - The combined total estimated damage was approximately \$2.611 billion.

- **Total Fatalities and Injuries:**

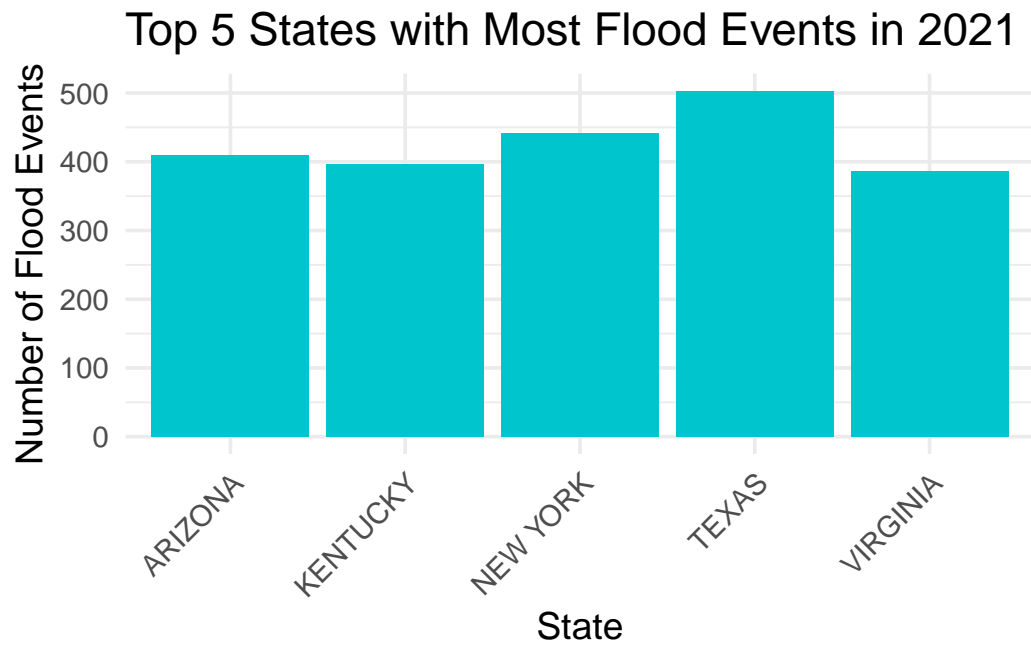
- There were 150 reported fatalities related to flood events.
- There were 70 reported injuries due to flood events.

## Visualizations

```
# Function to plot data
plot_data <- function(data, x, y, title, xlabel, ylabel, kind = 'bar') {
  p <- ggplot(data, aes_string(x = x, y = y)) +
    theme_minimal(base_size = 14) +
    labs(title = title, x = xlabel, y = ylabel) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))

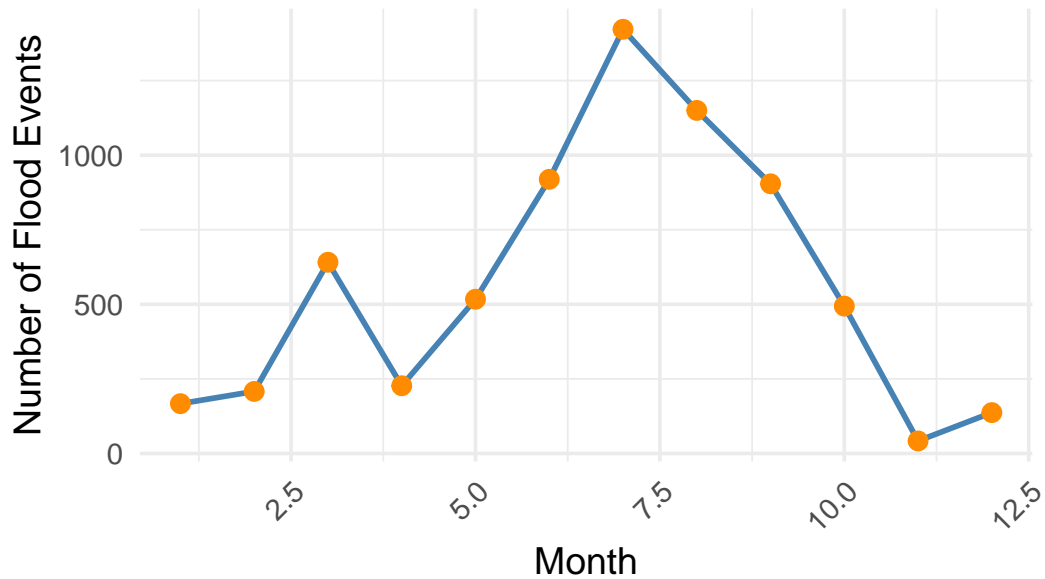
  if (kind == 'bar') {
    p + geom_bar(stat = 'identity', fill = 'turquoise3')
  } else if (kind == 'line') {
    p + geom_line(color = 'steelblue', size = 1) +
      geom_point(color = 'darkorange', size = 3)
  }
}

# States with the most flood events plot
states_flood_counts_df <- as.data.frame(states_flood_counts)
plot_data(states_flood_counts_df, 'STATE', 'n', 'Top 5 States with Most Flood Events in 20
```



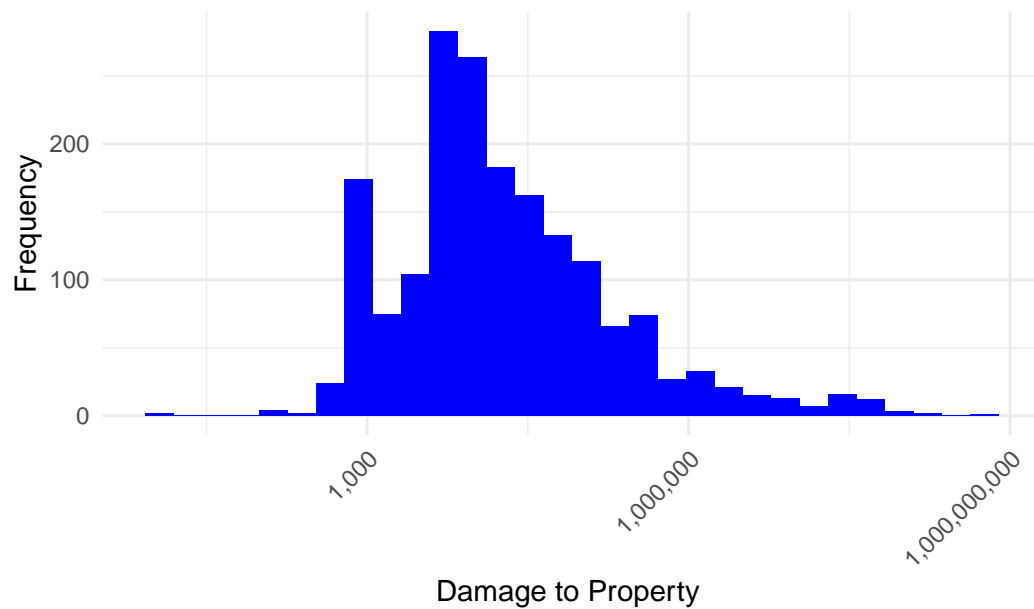
```
# Month Distribution of Flood Events plot
months_flood_counts_df <- as.data.frame(months_flood_counts)
plot_data(months_flood_counts_df, 'BEGIN_MONTH', 'n', 'Month Distribution of Flood Events
```

## Month Distribution of Flood Events in 2021

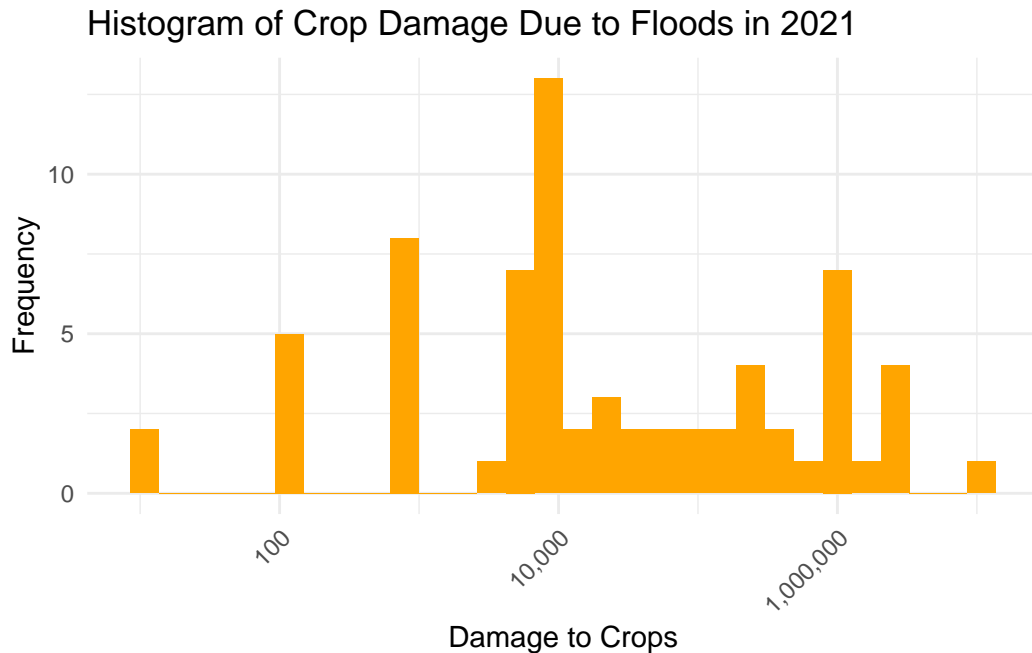


```
# Histogram of Damage to Property
non_zero_property_damage <- filter(flood_events_df, DAMAGE_PROPERTY_NUM > 0)
ggplot(non_zero_property_damage, aes(x = DAMAGE_PROPERTY_NUM)) +
  geom_histogram(bins = 30, fill = 'blue') +
  scale_x_log10(labels = comma) +
  labs(title = 'Histogram of Property Damage Due to Floods in 2021', x = 'Damage to Property') +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Histogram of Property Damage Due to Floods in 2021



```
# Histogram of Damage to Crops
non_zero_crop_damage <- filter(flood_events_df, DAMAGE_CROPS_NUM > 0)
ggplot(non_zero_crop_damage, aes(x = DAMAGE_CROPS_NUM)) +
  geom_histogram(bins = 30, fill = 'orange') +
  scale_x_log10(labels = comma) +
  labs(title = 'Histogram of Crop Damage Due to Floods in 2021', x = 'Damage to Crops', y =
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



**1. Top 5 States with Most Flood Events in 2021:**

- The bar chart shows Texas, New York, Arizona, Kentucky, and Virginia as the top states affected by flood events.

**2. Month Distribution of Flood Events in 2021:**

- The line chart indicates that flood events were more frequent during the summer months, with a peak in July.

**3. Histogram of Property Damage Due to Floods in 2021:**

- The histogram indicates that while many events result in lower damage costs, a few events can cause very high property damage.

**4. Histogram of Crop Damage Due to Floods in 2021:**

- Similarly, the histogram for crop damage demonstrates the frequency of different levels of crop damage due to flood events.



## Merge with fatalities and locations

```
# Read the CSV files into DataFrames
fatalities_df <- read_csv("StormEvents_fatalities-ftp_v1.0_d2021_c20231017.csv")
locations_df <- read_csv("StormEvents_locations-ftp_v1.0_d2021_c20231017.csv")

# Display the first few rows of each dataframe to inspect its content
head(fatalities_df)
```

```
# A tibble: 6 x 11
  FAT_YEARMONTH FAT_DAY FAT_TIME FATALITY_ID EVENT_ID FATALITY_TYPE
      <dbl>    <dbl>    <dbl>    <dbl>    <dbl> <chr>
1      202106         9         0      42960   953511 D
2      202111        25         0      44786   989138 I
3      202111         7         0      45176   992676 I
4      202111        11         0      45190   993049 I
5      202111        11         0      45191   993049 I
6      202111        29         0      45193   993061 I
# i 5 more variables: FATALITY_DATE <chr>, FATALITY_AGE <dbl>,
#   FATALITY_SEX <chr>, FATALITY_LOCATION <chr>, EVENT_YEARMONTH <dbl>
```

```
head(locations_df)
```

```
# A tibble: 6 x 11
  YEARMONTH EPISODE_ID EVENT_ID LOCATION_INDEX RANGE AZIMUTH LOCATION LATITUDE
      <dbl>    <dbl>    <dbl>    <dbl> <dbl> <chr>    <chr>    <dbl>
1      202104      156920   948472         1  36.4 SSE     EDGEWATER    28.5
2      202104      156920   948473         1   3.57 WSW     CAPE CANA~    28.4
3      202104      156920   948475         1    2.8 WNW     MELBOURNE    28.1
4      202104      156920   948476         1  11.2 N       CAPE CANA~    28.5
5      202104      156920   948477         1    1.4 ESE     VALKARIA     28.0
6      202104      156920   948479         1  10.2 E       VERO BEACH    27.7
# i 3 more variables: LONGITUDE <dbl>, LAT2 <dbl>, LON2 <dbl>
```

```
# Merge the details and fatalities datasets on 'EVENT_ID'
merged_df <- left_join(flood_events_df, fatalities_df, by = 'EVENT_ID')

# Merge the resulting dataset with the locations dataset on 'EVENT_ID'
merged_df <- left_join(merged_df, locations_df, by = 'EVENT_ID')
```

```
# Display the first few rows of the merged dataset to inspect its content
#head(merged_df)
```

1. Merge the details and fatalities datasets based on **EVENT\_ID**.
2. Merge the combined dataset with the locations dataset based on **EVENT\_ID**

## Clean

Here we:

- Fill NA values for fatality-related columns with 'None'
- Fill NA values for location-related columns with 'Unknown'

```
# Cleaning the merged dataset
# Remove duplicate columns resulted from the merge (those with suffix _x and _y)
merged_df <- merged_df %>%
  select(-contains("_y")) %>%
  rename_with(~ str_remove(., "_x"))

# Fill NA values for fatality-related columns with 'None'
fatality_columns <- c('FATALITY_ID', 'FATALITY_TYPE', 'FATALITY_DATE', 'FATALITY_AGE', 'FA
merged_df[fatality_columns] <- lapply(merged_df[fatality_columns], function(x) ifelse(is.na(x), 'None', x))

# Fill NA values for location-related columns with 'Unknown'
location_columns <- c('LOCATION_INDEX', 'RANGE', 'AZIMUTH', 'LOCATION', 'LATITUDE', 'LONGIT
merged_df[location_columns] <- lapply(merged_df[location_columns], function(x) ifelse(is.na(x), 'Unknown', x))

# Selecting a subset of relevant columns for analysis
columns_to_keep <- c('EVENT_ID', 'STATE', 'YEAR', 'MONTH_NAME', 'EVENT_TYPE', 'BEGIN_DATE',
  'EPISODE_NARRATIVE', 'EVENT_NARRATIVE', 'DAMAGE_PROPERTY_NUM', 'DAMAG
  'INJURIES_DIRECT', 'INJURIES_INDIRECT', 'FATALITY_TYPE', 'FATALITY_DA
  'LOCATION_INDEX', 'RANGE', 'AZIMUTH', 'LOCATION', 'LATITUDE', 'LONGIT
merged_clean_df <- select(merged_df, all_of(columns_to_keep))

# Look at the first few rows of the cleaned dataset
# head(merged_clean_df)
```

## EDA

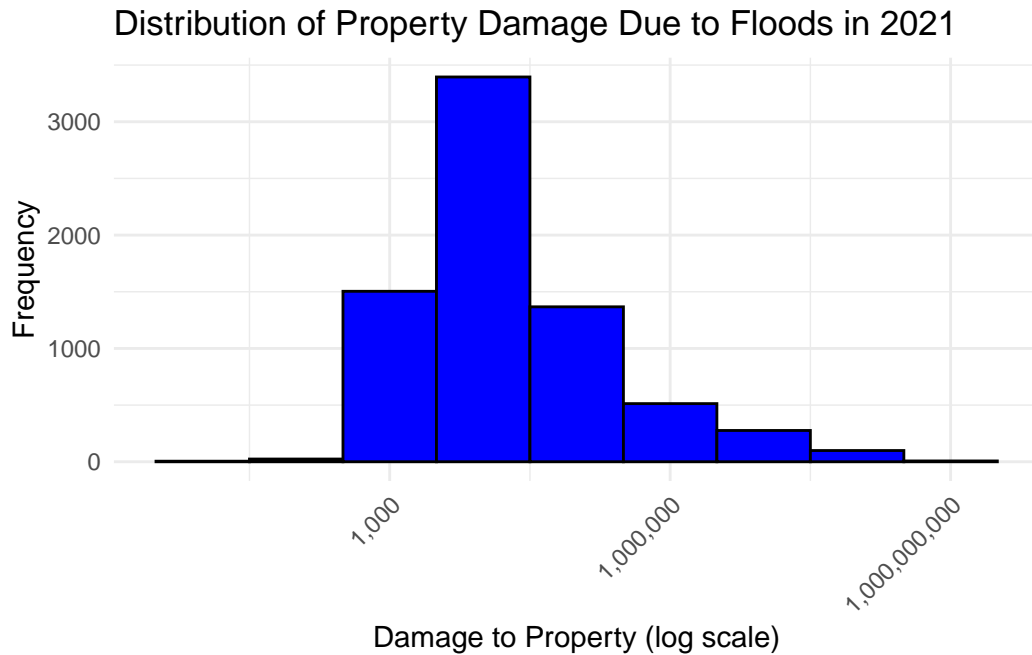
For the EDA report, we can explore several aspects:

1. **Overview of Flood Events:** Summarize the total number of flood and flash flood events, and their geographic and temporal distribution.
2. **Impact Analysis:** Assess the damage caused by floods, including property and crop damage, as well as human casualties and fatalities.
3. **Narrative Insights**
4. **Geospatial Analysis:**

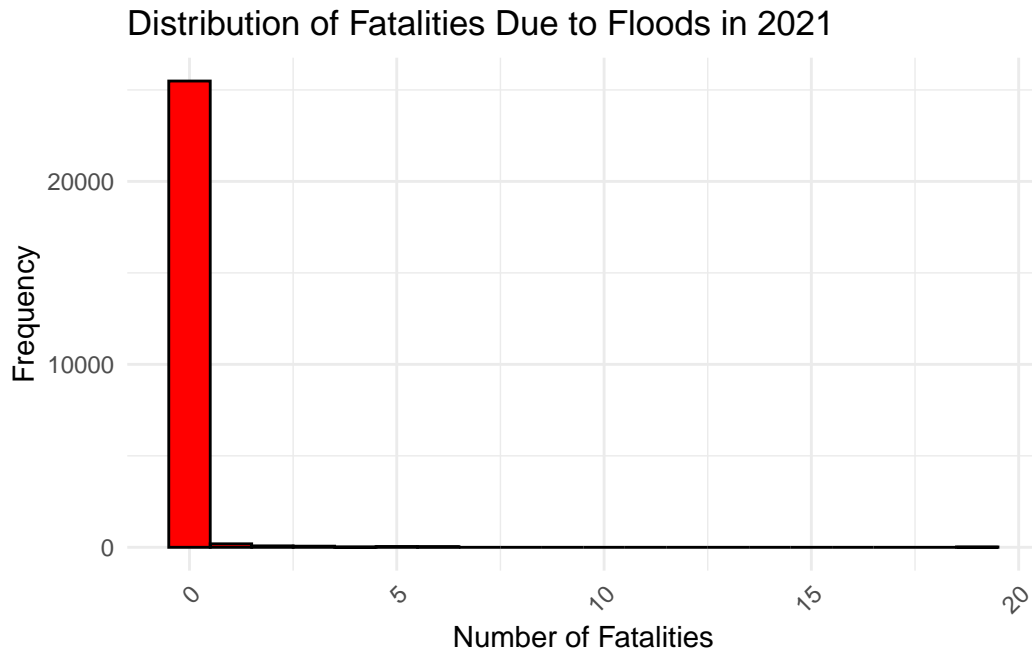
```
# Convert BEGIN_DATE_TIME to Date-Time object and create a new MONTH column
merged_clean_df <- merged_clean_df %>%
  mutate(
    BEGIN_DATE_TIME = ymd_hms(BEGIN_DATE_TIME),
    MONTH = floor_date(BEGIN_DATE_TIME, "month")
  )

# Aggregating property and crop damage by month
monthly_damage <- merged_clean_df %>%
  group_by(MONTH) %>%
  summarise(
    Total_Property_Damage = sum(DAMAGE_PROPERTY_NUM, na.rm = TRUE),
    Total_Crop_Damage = sum(DAMAGE_CROPS_NUM, na.rm = TRUE)
  )

# Histogram for the distribution of damage
ggplot(merged_clean_df, aes(x = DAMAGE_PROPERTY_NUM)) +
  geom_histogram(binwidth = 1, fill = 'blue', colour = 'black') +
  scale_x_continuous(labels = scales::comma, trans = 'log10') +
  labs(title = 'Distribution of Property Damage Due to Floods in 2021', x = 'Damage to Property') +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
# Histogram for the distribution of casualties
ggplot(merged_clean_df, aes(x = DEATHS_DIRECT + DEATHS_INDIRECT)) +
  geom_histogram(binwidth = 1, fill = 'red', colour = 'black') +
  labs(title = 'Distribution of Fatalities Due to Floods in 2021', x = 'Number of Fatalities') +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
# Breakdown of damage and fatalities by state
state_damage_fatalities <- merged_clean_df %>%
  group_by(STATE) %>%
  summarise(
    Total_Property_Damage = sum(DAMAGE_PROPERTY_NUM, na.rm = TRUE),
    Total_Crop_Damage = sum(DAMAGE_CROPS_NUM, na.rm = TRUE),
    Total_Fatalities = sum(DEATHS_DIRECT + DEATHS_INDIRECT, na.rm = TRUE),
    Event_Count = n_distinct(EVENT_ID)
  ) %>%
  mutate(Total_Damage = Total_Property_Damage + Total_Crop_Damage) %>%
  arrange(desc(Total_Damage))
```

#### 1. Distribution of Property Damage Due to Floods:

- The histogram, plotted on a logarithmic scale for the x-axis, shows the distribution of property damage amounts. This illustrates that while there are many events with lower levels of damage, there are a few events that result in very high property damage.

#### 2. Distribution of Fatalities Due to Floods:

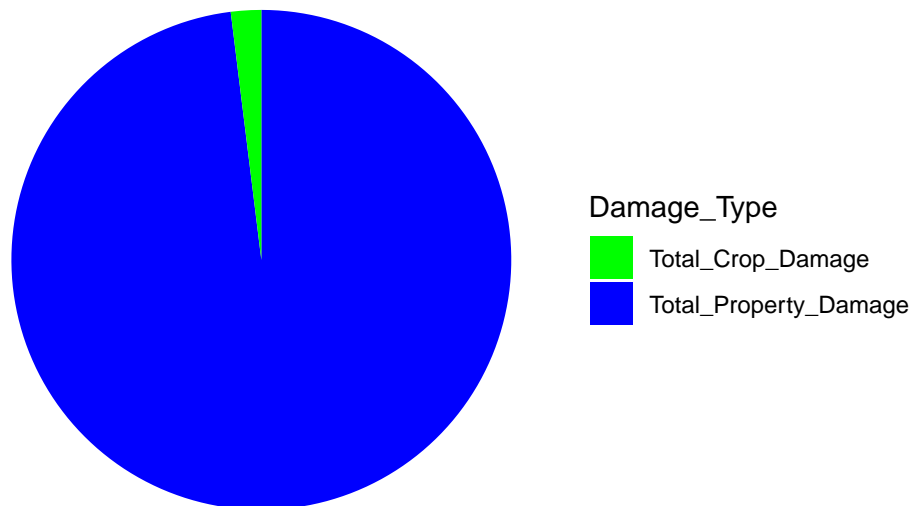
- The histogram for fatalities displays the frequency of different fatality counts due to flood events. This highlights that most flood events do not result in fatalities,

but there are events with significant loss of life.

```
damage_data <- merged_clean_df %>%
  summarise(
    Total_Property_Damage = sum(DAMAGE_PROPERTY_NUM, na.rm = TRUE),
    Total_Crop_Damage = sum(DAMAGE_CROPS_NUM, na.rm = TRUE)
  ) %>%
  gather(key = "Damage_Type", value = "Amount")

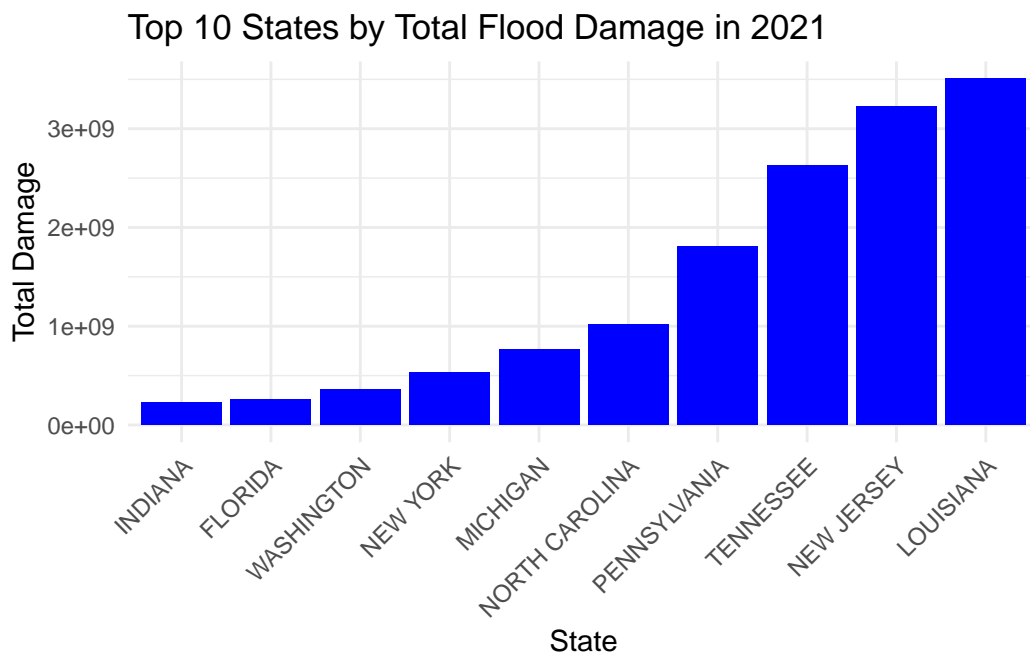
ggplot(damage_data, aes(x = "", y = Amount, fill = Damage_Type)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0) +
  scale_fill_manual(values = c("Total_Property_Damage" = "blue", "Total_Crop_Damage" = "green")) +
  labs(
    title = 'Proportion of Property vs Crop Damage Due to Floods in 2021',
    x = '',
    y = ''
  ) +
  theme_void()
```

Proportion of Property vs Crop Damage Due to Floods in 2021



```
# Adjust the dataset to include only the top 10 states by total damage
top_states_damage <- top_n(state_damage_fatalities, 10, Total_Damage)

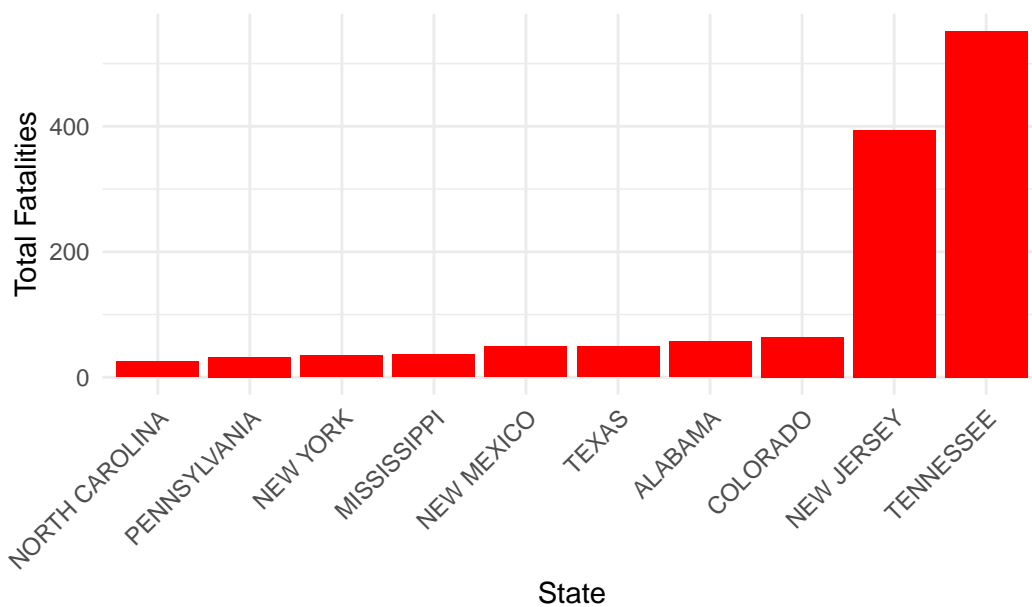
# Bar chart for Total Damage by State (Top 10 States)
ggplot(top_states_damage, aes(x = reorder(STATE, Total_Damage), y = Total_Damage)) +
  geom_bar(stat = "identity", fill = 'blue') +
  labs(title = 'Top 10 States by Total Flood Damage in 2021', x = 'State', y = 'Total Damage') +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
# Adjust the dataset to include only the top 10 states by total fatalities
top_states_fatalities <- top_n(state_damage_fatalities, 10, Total_Fatalities)

# Bar chart for Total Fatalities by State (Top 10 States)
ggplot(top_states_fatalities, aes(x = reorder(STATE, Total_Fatalities), y = Total_Fatalities)) +
  geom_bar(stat = "identity", fill = 'red') +
  labs(title = 'Top 10 States by Total Fatalities Due to Floods in 2021', x = 'State', y = 'Total Fatalities') +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

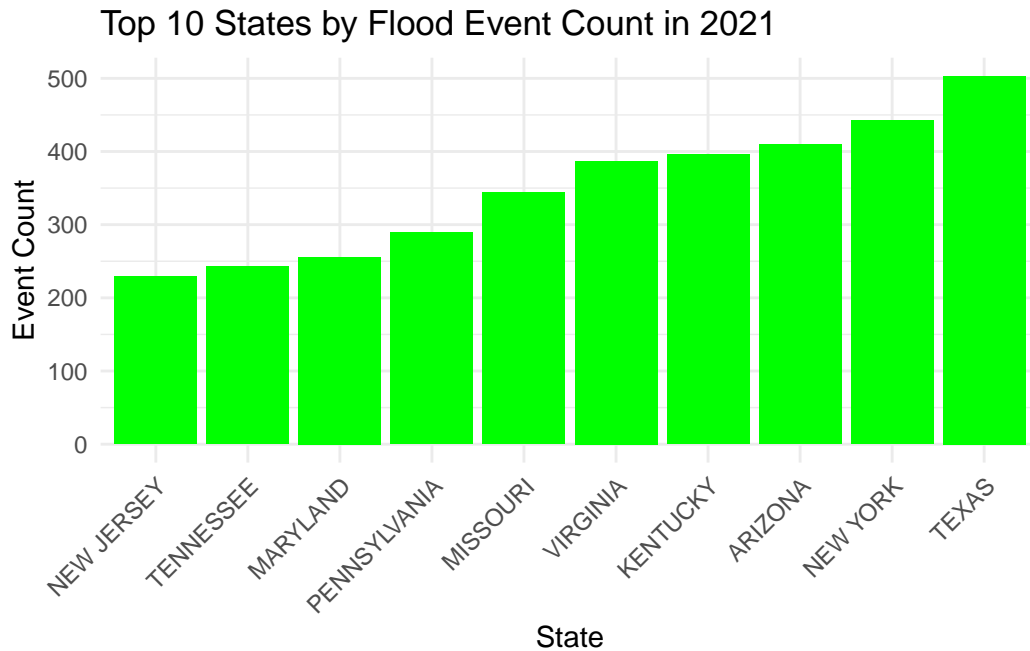
Top 10 States by Total Fatalities Due to Floods in 2021



```
# Adjust the dataset to include only the top 10 states by event count
top_states_event_count <- top_n(state_damage_fatalities, 10, Event_Count)

# Bar chart for Event Count by State (Top 10 States)
ggplot(top_states_event_count, aes(x = reorder(STATE, Event_Count), y = Event_Count)) +
  geom_bar(stat = "identity", fill = 'green') +
  labs(title = 'Top 10 States by Flood Event Count in 2021', x = 'State', y = 'Event Count') +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```





**1. Top 10 States by Total Flood Damage:**

- This bar chart shows the states with the highest total damage from floods. Louisiana and New Jersey stand out with the most significant damage, indicating severe flood impacts in these states.

**2. Top 10 States by Total Fatalities Due to Floods:**

- The chart highlights the states with the most fatalities due to floods. New Jersey has the highest number of fatalities, which could be due to a combination of factors including population density, the severity of flood events, and the vulnerability of the affected communities.

**3. Top 10 States by Flood Event Count:**

- This chart shows the states with the highest number of flood events. It's interesting to note that a high number of events in a state doesn't necessarily correlate with higher damage or fatalities, as seen with New York, which had the most events but less damage and fatalities compared to other states.

## **Conclusion**

**1. Geographical Impact:**

- Floods affected a wide range of states across the United States, with Texas, New York, Arizona, Kentucky, and Virginia being the most frequently hit.
- Louisiana experienced the most significant property damage, indicating the severe impact of flood events there.
- Despite a high number of flood events, some states like New York witnessed lower levels of damage and fatalities, which could suggest better preparedness or less severe flood conditions.

## **2. Temporal Distribution:**

- Flood events were most frequent during the summer months, with July being the peak month. This pattern underscores the influence of seasonal weather conditions on flooding.

## **3. Economic Damage:**

- The total estimated property and crop damage from floods in 2021 amounted to approximately \$2.611 billion, with property damage being the most significant contributor.
- The financial impact of floods is distributed unevenly across states, with some experiencing disproportionately higher damages.

## **4. Human Toll:**

- Floods resulted in 150 fatalities and 70 injuries across the country. New Jersey had the highest number of fatalities, while Tennessee had the most direct fatalities.
- The data suggests that while many flood events do not result in fatalities, certain events can be particularly deadly.

## **5. Community and Preparedness:**

- The data indicates a discrepancy between the number of flood events and their outcomes in terms of damage and fatalities, suggesting varying levels of community resilience and preparedness.
- States with fewer flood events experienced higher damage and fatalities, pointing to the potential severity of individual events and possibly lower preparedness or resilience levels.

## References

- <https://www.ncdc.noaa.gov/stormevents/ftp.jsp>
- <https://www.fema.gov/openfema-data-page/disaster-declarations-summaries-v2>