US Storm Events 2024: Impact on Population Health and Seasonal Paterns

Chrinovic Mukulu

2025-12-03

```
knitr::opts_chunk$set(
  echo = TRUE,
  message = FALSE,
  warning = FALSE
)

library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.5.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
library(stringr)

folder_path <- ("C:/Users/mukuluc/Desktop/final")
list.files(folder_path)
```

```
## [1] "Final.Rmd"
## [2] "Storm-Data-Bulk-csv-Format.pdf"
## [3] "Storm-Data-Export-Format.pdf"
## [4] "StormEvents_details-ftp_v1.0_d2025_c20251118.csv"
## [5] "StormEvents_fatalities-ftp_v1.0_d2025_c20251118.csv"
## [6] "StormEvents_joined_data.csv"
## [7] "StormEvents_locations-ftp_v1.0_d2025_c20251118.csv"
## [8] "zones.xls"
```

```
details_file    <- file.path(folder_path,  "StormEvents_details-ftp_v1.0_d2025_c20251118.csv")
fatalities_file <- file.path(folder_path, "StormEvents_fatalities-ftp_v1.0_d2025_c20251118.csv")
locations_file  <- file.path(folder_path, "StormEvents_locations-ftp_v1.0_d2025_c20251118.csv")

details    <- read_csv(details_file)
```

```
## Rows: 44721 Columns: 51

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (25): STATE, MONTH_NAME, EVENT_TYPE, CZ_TYPE, CZ_NAME, WFO, BEGIN_DATE_T...
## dbl (25): BEGIN_YEARMONTH, BEGIN_DAY, BEGIN_TIME, END_YEARMONTH, END_DAY, EN...
## lgl  (1): CATEGORY
##
## 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.
```

```
fatalities <- read_csv(fatalities_file)
```

```
## Rows: 423 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): FATALITY_TYPE, FATALITY_DATE, FATALITY_SEX, FATALITY_LOCATION
## dbl (7): FAT_YEARMONTH, FAT_DAY, FAT_TIME, FATALITY_ID, EVENT_ID, FATALITY_A...
##
## 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.
```

```
locations  <- read_csv(locations_file)
```

```
## Rows: 31130 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): AZIMUTH, LOCATION
## dbl (9): YEARMONTH, EPISODE_ID, EVENT_ID, LOCATION_INDEX, RANGE, LATITUDE, L...
##
## 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.
```

```
dim(details)
```

```
## [1] 44721    51
```

```r
dim(fatalities)
```

```
## [1] 423  11
```

```r
dim(locations)
```

```
## [1] 31130    11
```

```r
head(details)
```

```
## # A tibble: 6 x 51
##   BEGIN_YEARMONTH BEGIN_DAY BEGIN_TIME END_YEARMONTH END_DAY END_TIME EPISODE_ID
##             <dbl>     <dbl>      <dbl>         <dbl>   <dbl>    <dbl>      <dbl>
## 1          202503        31       1104        202503      31     1106     201366
## 2          202503        30       1552        202503      30     1555     200337
## 3          202501         5       1800        202501       6     2227     197733
## 4          202501         3       1300        202501       3     1900     197761
## 5          202501         3       1300        202501       3     1900     197761
## 6          202501         3       1300        202501       3     1900     197761
## # i 44 more variables: EVENT_ID <dbl>, STATE <chr>, STATE_FIPS <dbl>,
## #   YEAR <dbl>, MONTH_NAME <chr>, EVENT_TYPE <chr>, CZ_TYPE <chr>,
## #   CZ_FIPS <dbl>, CZ_NAME <chr>, WFO <chr>, BEGIN_DATE_TIME <chr>,
## #   CZ_TIMEZONE <chr>, END_DATE_TIME <chr>, INJURIES_DIRECT <dbl>,
## #   INJURIES_INDIRECT <dbl>, DEATHS_DIRECT <dbl>, DEATHS_INDIRECT <dbl>,
## #   DAMAGE_PROPERTY <chr>, DAMAGE_CROPS <chr>, SOURCE <chr>, MAGNITUDE <dbl>,
## #   MAGNITUDE_TYPE <chr>, FLOOD_CAUSE <chr>, CATEGORY <lgl>, ...
```

```r
StormEvents_joined_data <- details %>%
  left_join(locations,  by = "EVENT_ID") %>%
  left_join(fatalities, by = "EVENT_ID")
```

```
## Warning in left_join(., fatalities, by = "EVENT_ID"): Detected an unexpected many-to-many relationshi
## i Row 896 of `x` matches multiple rows in `y`.
## i Row 325 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
dim(StormEvents_joined_data)
```

```
## [1] 56092    71
```

```r
head(StormEvents_joined_data)
```

```
## # A tibble: 6 x 71
##   BEGIN_YEARMONTH BEGIN_DAY BEGIN_TIME END_YEARMONTH END_DAY END_TIME
##             <dbl>     <dbl>      <dbl>         <dbl>   <dbl>    <dbl>
## 1          202503        31       1104        202503      31     1106
## 2          202503        30       1552        202503      30     1555
## 3          202501         5       1800        202501       6     2227
```

```
## 4          202501         3       1300        202501       3     1900
## 5          202501         3       1300        202501       3     1900
## 6          202501         3       1300        202501       3     1900
## # i 65 more variables: EPISODE_ID.x <dbl>, EVENT_ID <dbl>, STATE <chr>,
## #   STATE_FIPS <dbl>, YEAR <dbl>, MONTH_NAME <chr>, EVENT_TYPE <chr>,
## #   CZ_TYPE <chr>, CZ_FIPS <dbl>, CZ_NAME <chr>, WFO <chr>,
## #   BEGIN_DATE_TIME <chr>, CZ_TIMEZONE <chr>, END_DATE_TIME <chr>,
## #   INJURIES_DIRECT <dbl>, INJURIES_INDIRECT <dbl>, DEATHS_DIRECT <dbl>,
## #   DEATHS_INDIRECT <dbl>, DAMAGE_PROPERTY <chr>, DAMAGE_CROPS <chr>,
## #   SOURCE <chr>, MAGNITUDE <dbl>, MAGNITUDE_TYPE <chr>, FLOOD_CAUSE <chr>, ...
```

```
write_csv(StormEvents_joined_data,
          file.path(folder_path, "StormEvents_joined_data.csv"))
StormEvents_clean <- StormEvents_joined_data %>%
  mutate(
    BEGIN_DATE = mdy_hms(BEGIN_DATE_TIME, quiet = TRUE),


    MONTH = month(BEGIN_DATE, label = TRUE, abbr = TRUE),
    EVENT_TYPE = str_trim(EVENT_TYPE),
    STATE = str_trim(STATE),


    TOTAL_INJURIES = INJURIES_DIRECT + INJURIES_INDIRECT,
    TOTAL_DEATHS = DEATHS_DIRECT + DEATHS_INDIRECT,


    TOTAL_HEALTH_HARM = TOTAL_INJURIES + TOTAL_DEATHS
  )

head(StormEvents_clean)
```

```
## # A tibble: 6 x 76
##   BEGIN_YEARMONTH BEGIN_DAY BEGIN_TIME END_YEARMONTH END_DAY END_TIME
##             <dbl>     <dbl>      <dbl>         <dbl>   <dbl>    <dbl>
## 1          202503        31       1104        202503      31     1106
## 2          202503        30       1552        202503      30     1555
## 3          202501         5       1800        202501       6     2227
## 4          202501         3       1300        202501       3     1900
## 5          202501         3       1300        202501       3     1900
## 6          202501         3       1300        202501       3     1900
## # i 70 more variables: EPISODE_ID.x <dbl>, EVENT_ID <dbl>, STATE <chr>,
## #   STATE_FIPS <dbl>, YEAR <dbl>, MONTH_NAME <chr>, EVENT_TYPE <chr>,
## #   CZ_TYPE <chr>, CZ_FIPS <dbl>, CZ_NAME <chr>, WFO <chr>,
## #   BEGIN_DATE_TIME <chr>, CZ_TIMEZONE <chr>, END_DATE_TIME <chr>,
## #   INJURIES_DIRECT <dbl>, INJURIES_INDIRECT <dbl>, DEATHS_DIRECT <dbl>,
## #   DEATHS_INDIRECT <dbl>, DAMAGE_PROPERTY <chr>, DAMAGE_CROPS <chr>,
## #   SOURCE <chr>, MAGNITUDE <dbl>, MAGNITUDE_TYPE <chr>, FLOOD_CAUSE <chr>, ...
```

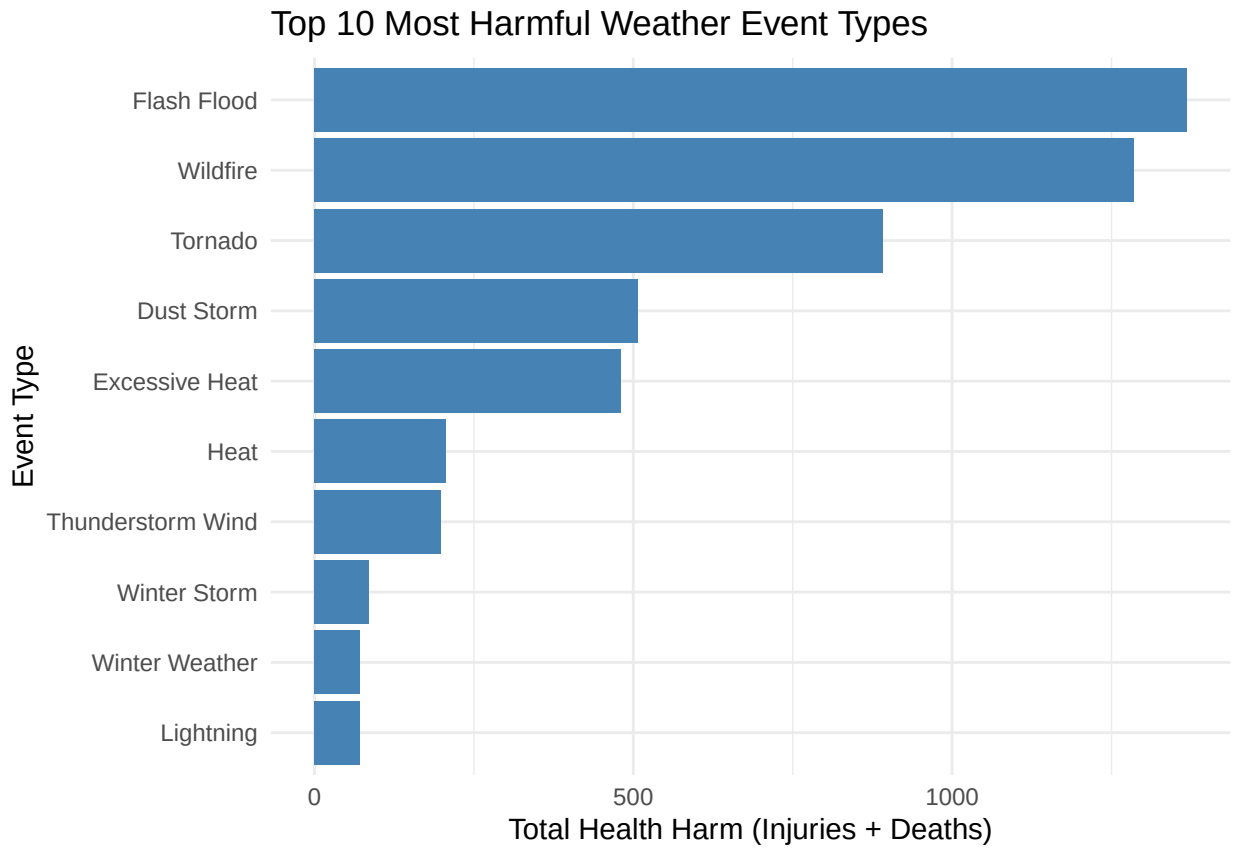
```
Q1_summary <- StormEvents_clean %>%
  group_by(EVENT_TYPE) %>%
  summarise(
    Total_Injuries = sum(TOTAL_INJURIES, na.rm = TRUE),
```

```
    Total_Deaths   = sum(TOTAL_DEATHS, na.rm = TRUE),
    Total_Health_Harm = sum(TOTAL_HEALTH_HARM, na.rm = TRUE)
  ) %>%
  arrange(desc(Total_Health_Harm))

head(Q1_summary, 10)
```

```
## # A tibble: 10 x 4
##    EVENT_TYPE        Total_Injuries Total_Deaths Total_Health_Harm
##    <chr>                      <dbl>        <dbl>              <dbl>
##  1 Flash Flood                   12         1356               1368
##  2 Wildfire                     413          872               1285
##  3 Tornado                      730          162                892
##  4 Dust Storm                   417           91                508
##  5 Excessive Heat               172          309                481
##  6 Heat                          33          173                206
##  7 Thunderstorm Wind            107           91                198
##  8 Winter Storm                  23           63                 86
##  9 Lightning                     57           14                 71
## 10 Winter Weather                44           27                 71
```

```
Q1_summary %>%
  slice_max(Total_Health_Harm, n = 10) %>%
  ggplot(aes(x = reorder(EVENT_TYPE, Total_Health_Harm),
             y = Total_Health_Harm)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  labs(
    title = "Top 10 Most Harmful Weather Event Types",
    x = "Event Type",
    y = "Total Health Harm (Injuries + Deaths)"
  ) +
  theme_minimal()
```




```
# Question 1
# The analysis shows that certain severe weather events have a disproportionately large impact on popul
```

```
Q2_counts <- StormEvents_clean %>%
  group_by(STATE, EVENT_TYPE) %>%
  summarise(Event_Count = n(), .groups = "drop")

head(Q2_counts)
```

```
## # A tibble: 6 x 3
##   STATE   EVENT_TYPE              Event_Count
##   <chr>   <chr>                         <int>
## 1 ALABAMA Cold/Wind Chill                  26
## 2 ALABAMA Drought                          23
## 3 ALABAMA Extreme Cold/Wind Chill           6
## 4 ALABAMA Flash Flood                     124
## 5 ALABAMA Flood                            13
## 6 ALABAMA Funnel Cloud                      1
```

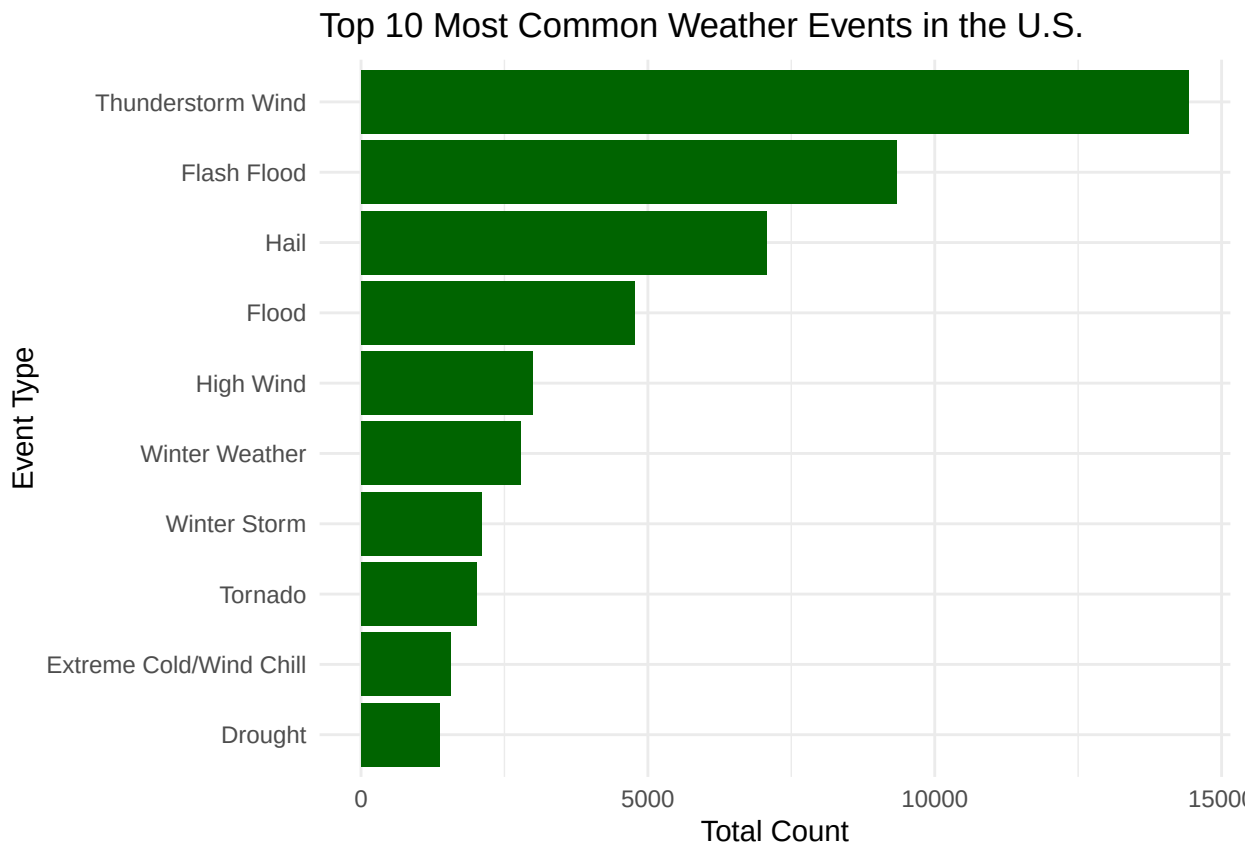
```r
Q2_top_event_by_state <- Q2_counts %>%
  group_by(STATE) %>%
  slice_max(Event_Count, n = 1) %>%
  arrange(STATE)

head(Q2_top_event_by_state, 15)
```

```
## # A tibble: 15 x 3
## # Groups:   STATE [15]
##    STATE                EVENT_TYPE                Event_Count
##    <chr>                <chr>                           <int>
##  1 ALABAMA              Thunderstorm Wind                1091
##  2 ALASKA               Blizzard                           39
##  3 AMERICAN SAMOA       Flash Flood                        41
##  4 ARIZONA              Flash Flood                        73
##  5 ARKANSAS             Thunderstorm Wind                 247
##  6 ATLANTIC NORTH       Marine Thunderstorm Wind          364
##  7 ATLANTIC SOUTH       Marine Thunderstorm Wind          337
##  8 CALIFORNIA           Flood                             601
##  9 COLORADO             Hail                              267
## 10 CONNECTICUT          Winter Weather                     23
## 11 DELAWARE             Thunderstorm Wind                  27
## 12 DISTRICT OF COLUMBIA Thunderstorm Wind                   9
## 13 E PACIFIC            Waterspout                          2
## 14 FLORIDA              Thunderstorm Wind                 197
## 15 GEORGIA              Thunderstorm Wind                 964
```

```r
Q2_counts_overall <- Q2_counts %>%
  group_by(EVENT_TYPE) %>%
  summarise(Total_Events = sum(Event_Count), .groups = "drop") %>%
  arrange(desc(Total_Events))

Q2_counts_overall %>%
  slice_max(Total_Events, n = 10) %>%
  ggplot(aes(x = reorder(EVENT_TYPE, Total_Events),
             y = Total_Events)) +
  geom_col(fill = "darkgreen") +
  coord_flip() +
  labs(
    title = "Top 10 Most Common Weather Events in the U.S.",
    x = "Event Type",
    y = "Total Count"
  ) +
  theme_minimal()
```




```
#Question 2
# Across the United States, different states experience certain weather events far more frequently than
```

```
Q3_monthly <- StormEvents_clean %>%
  group_by(MONTH, EVENT_TYPE) %>%
  summarise(Event_Count = n(), .groups = "drop")

Q3_top_event_by_month <- Q3_monthly %>%
  group_by(MONTH) %>%
  slice_max(Event_Count, n = 1) %>%
  arrange(MONTH)

Q3_top_event_by_month
```

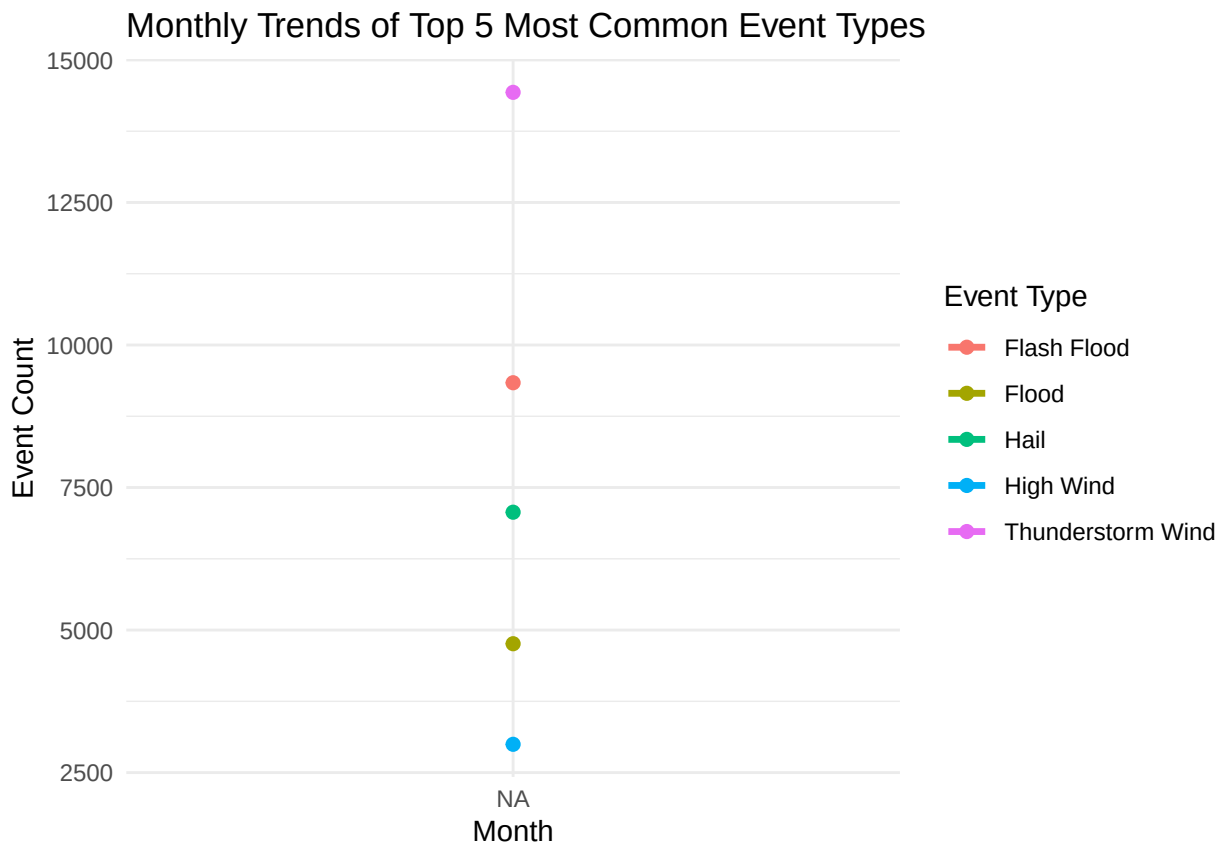
```
## # A tibble: 1 x 3
## # Groups:   MONTH [1]
##   MONTH EVENT_TYPE        Event_Count
##   <ord> <chr>                   <int>
## 1 <NA>  Thunderstorm Wind       14432
```

```r
top5_events <- Q3_monthly %>%
  group_by(EVENT_TYPE) %>%
  summarise(Total = sum(Event_Count)) %>%
  slice_max(Total, n = 5) %>%
  pull(EVENT_TYPE)

Q3_monthly %>%
  filter(EVENT_TYPE %in% top5_events) %>%
  ggplot(aes(x = MONTH, y = Event_Count, color = EVENT_TYPE, group = EVENT_TYPE)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  labs(
    title = "Monthly Trends of Top 5 Most Common Event Types",
    x = "Month",
    y = "Event Count",
    color = "Event Type"
  ) +
  theme_minimal()
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.

## ‘geom_line()‘: Each group consists of only one observation.
## i Do you need to adjust the group aesthetic?
```




```
#Question 3
# The monthly analysis reveals clear seasonal patterns in weather events across the United States. Wint
```

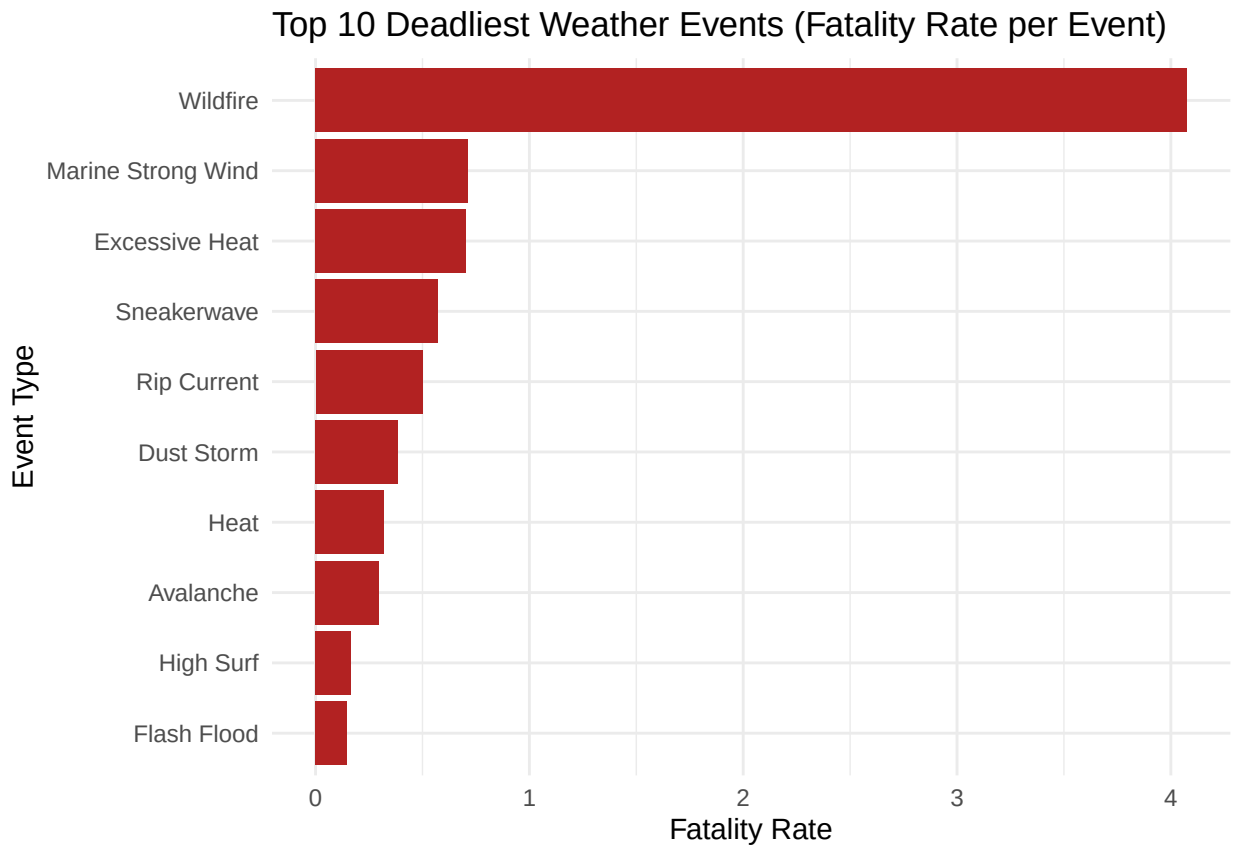
```
Q4A_fatality_rate <- StormEvents_clean %>%
  group_by(EVENT_TYPE) %>%
  summarise(
    Total_Deaths = sum(TOTAL_DEATHS, na.rm = TRUE),
    Event_Count = n(),
    Fatality_Rate = Total_Deaths / Event_Count
  ) %>%
  arrange(desc(Fatality_Rate))

head(Q4A_fatality_rate, 10)
```

```
## # A tibble: 10 x 4
##    EVENT_TYPE         Total_Deaths Event_Count Fatality_Rate
##    <chr>                     <dbl>       <int>         <dbl>
##  1 Wildfire                    872         214         4.07
##  2 Marine Strong Wind            5           7         0.714
##  3 Excessive Heat              309         438         0.705
##  4 Sneakerwave                   4           7         0.571
##  5 Rip Current                  16          32         0.5
```

```
##  6 Dust Storm                  91         237        0.384
##  7 Heat                       173         538        0.322
##  8 Avalanche                    8          27        0.296
##  9 High Surf                    6          36        0.167
## 10 Flash Flood               1356        9338        0.145
```

```
Q4A_fatality_rate %>%
  slice_max(Fatality_Rate, n = 10) %>%
  ggplot(aes(x = reorder(EVENT_TYPE, Fatality_Rate),
             y = Fatality_Rate)) +
  geom_col(fill = "firebrick") +
  coord_flip() +
  labs(
    title = "Top 10 Deadliest Weather Events (Fatality Rate per Event)",
    x = "Event Type",
    y = "Fatality Rate"
  ) +
  theme_minimal()
```




```
# Question 4A
# Weather events with the highest fatality rate per occurrence tend to be rare but extremely dangerous.
```

```r
StormEvents_clean <- StormEvents_clean %>%
  mutate(
    PROP_DAMAGE_NUM = case_when(
      str_detect(DAMAGE_PROPERTY, "K") ~ as.numeric(str_remove(DAMAGE_PROPERTY, "K")) * 1e3,
      str_detect(DAMAGE_PROPERTY, "M") ~ as.numeric(str_remove(DAMAGE_PROPERTY, "M")) * 1e6,
      str_detect(DAMAGE_PROPERTY, "B") ~ as.numeric(str_remove(DAMAGE_PROPERTY, "B")) * 1e9,
      TRUE ~ 0
    )
  )
```

```
## Warning: There were 3 warnings in `mutate()`.
## The first warning was:
## i In argument: `PROP_DAMAGE_NUM = case_when(...)`.
## Caused by warning:
## ! NAs introduced by coercion
## i Run `dplyr::last_dplyr_warnings()` to see the 2 remaining warnings.
```

```r
head(StormEvents_clean$PROP_DAMAGE_NUM)
```

```
## [1] 1e+03 1e+05 0e+00 0e+00 0e+00 0e+00
```

```r
Q4B_damage <- StormEvents_clean %>%
  group_by(EVENT_TYPE) %>%
  summarise(
    Total_Property_Damage = sum(PROP_DAMAGE_NUM, na.rm = TRUE)
  ) %>%
  arrange(desc(Total_Property_Damage))

head(Q4B_damage, 10)
```

```
## # A tibble: 10 x 2
##    EVENT_TYPE        Total_Property_Damage
##    <chr>                             <dbl>
##  1 Tornado                    13293225500
##  2 Flash Flood                 1653107600
##  3 Wildfire                     208200010
##  4 Thunderstorm Wind            127727410
##  5 Hail                          37355850
##  6 Flood                         32033750
##  7 Lightning                     13457200
##  8 High Wind                      5924350
##  9 Ice Storm                      3205000
## 10 Dust Storm                     3192000
```

```r
#Question 4B
# Property damage analysis shows that certain event types cause disproportionately large economic losse
```

```r
Q4C_state_fatal <- StormEvents_clean %>%
  group_by(STATE) %>%
  summarise(
    Total_Deaths = sum(TOTAL_DEATHS, na.rm = TRUE),
    Fatal_Event_Count = sum(TOTAL_DEATHS > 0, na.rm = TRUE)
  ) %>%
  arrange(desc(Total_Deaths))

head(Q4C_state_fatal, 10)
```

```
## # A tibble: 10 x 3
##    STATE         Total_Deaths Fatal_Event_Count
##    <chr>                <dbl>             <int>
##  1 CALIFORNIA             952                80
##  2 WEST VIRGINIA          658                76
##  3 TEXAS                  647                87
##  4 NEVADA                 471                69
##  5 KANSAS                  72                12
##  6 MISSOURI                70                28
##  7 TENNESSEE               67                19
##  8 ARIZONA                 59                27
##  9 OKLAHOMA                52                36
## 10 MONTANA                 39                15
```

```r
#Question 4C
# Several states experience significantly higher numbers of fatal weather events than others. States wi
```