

# Storm Events Data: Analysis

Emmanuel Guizar Rosales

last rendered on: Aug 16, 2024

```
# install package librarian if needed
if (!("librarian" %in% rownames(installed.packages())))) {
  install.packages("librarian")
}

# load required packages
librarian::shelf(
  tidyverse,
  fs,
  usmap,
  ggpubr,
  sessioninfo
)

# Source required functions
myFunctions <- c(
  "FUNStormEventsData_filterData"
)

for (f in myFunctions) {
  source(paste0("../functions/", f, ".R"))
}

# Preperations to show states boundaries
poly_states <- plot_usmap(regions = "states")

# Read in data_details_fips
fileName <- "data_details_fips.RDS"
pathName <- "../data/stormData"
filePath <- dir_ls(path = pathName, regexp = paste0(fileName, "$")) %>% last()
data_details_fips <- readRDS(filePath)
```

## Purpose & Rationale

As outlined in the Registered Report, we will assess the number of extreme weather episodes recorded in each participant's county of residence within the 30 days prior to study completion. Regarding the time window during which we plan to conduct the study, we aim for maximizing the likelihood of capturing suitable variability in the exposure to extreme weather episodes with notable geographic variability. To this end, we analyzed records of extreme weather episodes over the last ten years.

## Filter Data

We filter the storm events data for the specific years, months, and extreme weather event types we are interested in. We filter for all years from 2014 to 2023 (as data are not complete for the year 2024 yet), we highlight the month of July, and we focus on those types of extreme weather events that are predicted to increase in frequency and severity due to climate change (IPCC 2023): Excessive Heat, Drought, Wildfire, Flash Flood, Coastal Flood, Strong Wind, Hail, and Tornado.

```
# Define variables of interest
myYears <- seq(2014, 2023)
myMonths <- c("July")
myEventTypes <- c(
  "Excessive Heat",
  "Drought",
  "Wildfire",
  "Flash Flood",
  "Coastal Flood",
  "Strong Wind",
  "Hail",
  "Tornado"
)

# Call function
out <- FUNStormEventsData_filterData(
  myData = data_details_fips,
  myYears = myYears,
  myMonths = myMonths,
  myEventTypes = myEventTypes
)
```

## Analysis

```
p.hist <- out$dataForHist %>%
  group_by(year) %>%
  mutate(
    max_nEpisodes = max(nEpisodes),
    yearlyMean_nEpisodes = mean(nEpisodes)
  ) %>%
  ungroup() %>%
  mutate(max_month = ifelse(nEpisodes == max_nEpisodes, TRUE, FALSE)) %>%
  ggplot(aes(
    x = month_name, y = nEpisodes,
    linewidth = max_month,
    fill = month_name %in% myMonths
  )) +
  geom_hline(
    mapping = aes(yintercept = yearlyMean_nEpisodes,
      linetype = "dashed",
      color = "black"
    ) +
    geom_bar(
      stat = "identity",
      color = "black",
      alpha = .7,
      show.legend = FALSE
    ) +
    scale_linewidth_manual(values = c(0.5, 2)) +
    scale_x_discrete(labels = month.abb) +
    scale_fill_manual(
      values = c("darkgrey", "orange"),
    ) +
    labs(
      title = "Number of Extreme Weather Episodes by Month over the Years 2014 to 2023",
      x = "Month",
      y = "Number of Episodes"
    ) +
    theme_bw() +
    theme(
      text = element_text(size = 15),
      plot.title = element_text(hjust = .5),
      axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5)
```

```

) +
facet_wrap(~year, ncol = 5)

jpeg(
  file = "../images/histogramSeasonalDistribution.jpeg",
  width = 14, height = 7.5, units = "in", res = 600
)
print(p.hist)
invisible(dev.off())

```

```

p.map_bin <- plot_usmap(
  data = out$dataForUsPlot,
  values = "episodes_bin",
  regions = "counties",
  exclude = c("AK", "HI"),
  color = "black",
  linewidth = 0.1
) +
  geom_sf(
    data = poly_states[[1]] %>%
      filter(!(abrv %in% c("AK", "HI"))),
    color = "black",
    fill = NA,
    linewidth = .3
) +
  scale_fill_manual(
    name = "Number of Episodes > 0",
    values = c("white", "orange")
) +
  labs(
    title = "Extreme Weather Episodes in July over the Years 2014 to 2023"
) +
  theme_bw() +
  theme(
    text = element_text(size = 15),
    legend.position = "bottom",
    plot.title = element_text(hjust = .5),
    panel.grid = element_blank(),
    axis.ticks = element_blank(),
    axis.text = element_blank()
) +

```

```

facet_wrap(~year, ncol = 5)

jpeg(
  file = "../images/mapGeographicalDistribution_bin.jpeg",
  width = 14, height = 7.5, units = "in", res = 600
)
print(p.map_bin)
invisible(dev.off())

```

Analyzing the seasonal distribution of extreme weather episodes, Figure 1 shows that July consistently shows a high number of extreme weather episodes over the last ten years. Additionally, Figure 2 indicates that within the month of July, these extreme weather episodes also display a high geographical variability.

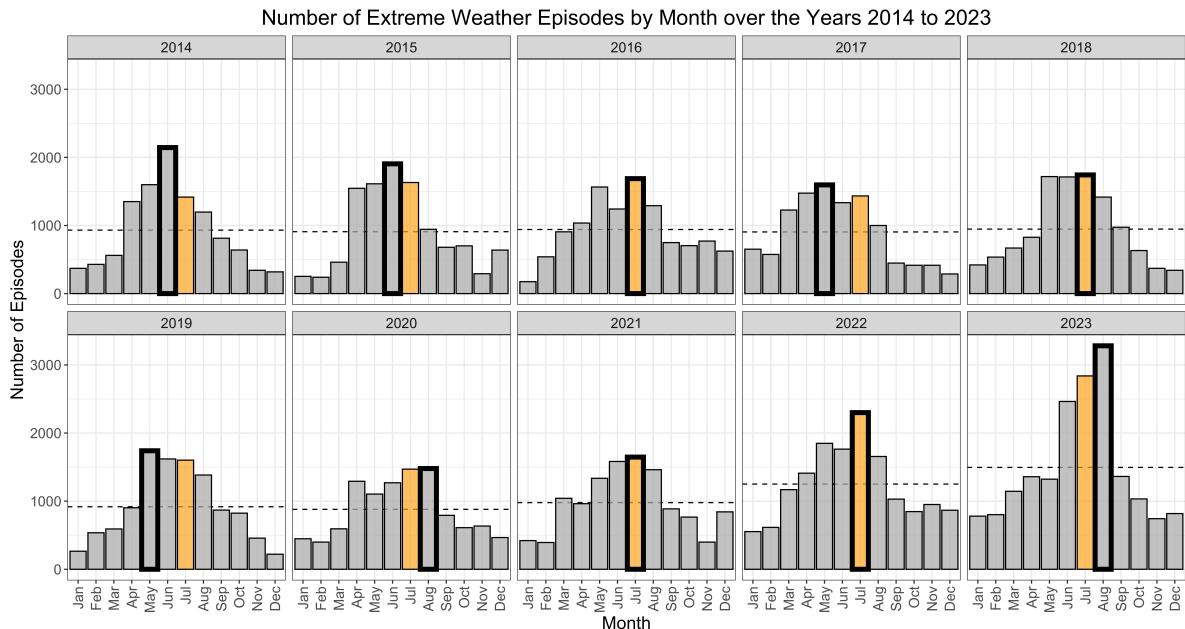


Figure 1: Histograms showing the number of extreme weather episodes by month from 2014 to 2023. The dashed horizontal line indicates the mean number of extreme weather episodes in each year. The thick-bordered bar marks the month with the most extreme weather events each year. The orange bar represents July. July had the most extreme weather events in 4 out of 10 years, and in another 4 years, it was right before or after the peak month. Only episodes that included at least one of the following event types were considered: excessive heat, drought, wildfire, flash flood, coastal flood, strong wind, hail, tornado.

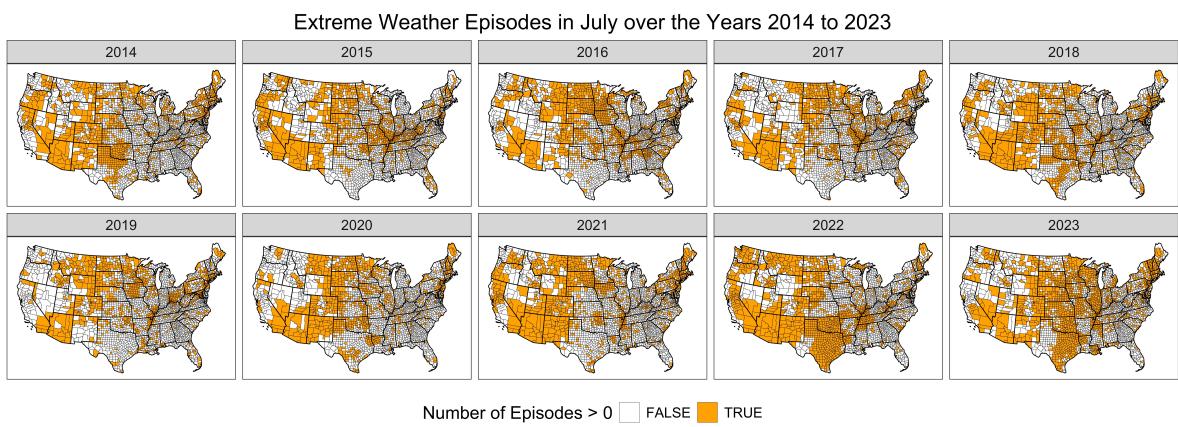


Figure 2: Maps displaying the geographical distribution of the occurrence of at least one extreme weather episode in July over the years 2014 to 2023. Only episodes that included at least one of the following event types were considered: excessive heat, drought, wildfire, flash flood, coastal flood, strong wind, hail, tornado.

```

dataForPlot <- out$dataForUsPlot %>%
  mutate(nEpisodes_withNA = ifelse(nEpisodes == 0, NA_integer_, nEpisodes))

p.map_cont <- plot_usmap(
  data = dataForPlot,
  values = "nEpisodes_withNA",
  regions = "counties",
  exclude = c("AK", "HI"),
  color = "black",
  linewidth = 0.1
) +
  geom_sf(
    data = poly_states[[1]] %>%
      filter(!(abrv %in% c("AK", "HI"))),
    color = "black",
    fill = NA,
    linewidth = .3
) +
  scale_fill_binned(
    name = "Number of Episodes",
    n.breaks = 10,
    type = "viridis",
    na.value = "white"
) +
  labs(
    title = "Extreme Weather Episodes in July over the Years 2014 to 2023"
  ) +
  theme_bw() +
  theme(
    text = element_text(size = 15),
    legend.position = "bottom",
    plot.title = element_text(hjust = .5),
    panel.grid = element_blank(),
    axis.ticks = element_blank(),
    axis.text = element_blank()
  ) +
  facet_wrap(~year, ncol = 5)

jpeg(
  file = "../images/mapGeographicalDistribution_cont.jpeg",
  width = 14, height = 7.5, units = "in", res = 600
)

```

```

print(p.map_cont)
invisible(dev.off())

p.hist_count <- out$dataForUsPlot %>%
  group_by(year, nEpisodes) %>%
  summarise(
    count = n(),
    prcnt = count / n_distinct(out$dataForUsPlot$fips)
  ) %>%
  ggplot(aes(x = nEpisodes, y = prcnt)) +
  geom_bar(stat = "identity", color = "black", fill = "darkgrey") +
  scale_y_continuous(labels = scales::label_percent()) +
  labs(
    x = "Number of Episodes",
    y = "Proportion of Counties"
  ) +
  theme_bw() +
  labs(
    title = "Extreme Weather Episodes in July over the Years 2014 to 2023"
  ) +
  theme(
    text = element_text(size = 15),
    legend.position = "bottom",
    plot.title = element_text(hjust = .5)
  ) +
  facet_wrap(~year, ncol = 5)

jpeg(
  file = "../images/frequencyDistribution_cont.jpeg",
  width = 14, height = 7.5, units = "in", res = 600
)
print(p.hist_count)
invisible(dev.off())

# Calcualte some proportions for display in text
props2023 <- out$dataForUsPlot %>%
  filter(year == 2023) %>%
  count(episodes_bin) %>%
  mutate(
    freq = n/sum(n),
    freq_prcnt = paste0(format(round(freq*100, 2), nsmall = 2), "%")
  )

```

While Figure 2 visualizes the occurrence of at least one extreme weather episode in July for each county and year (binary variable), Figure 4 displays the actual number of such episodes (continuous). The vast majority of counties were exposed to few episodes, indicating that most of the variability is due to whether an extreme weather episode occurred at all or not. This is further supported by Figure 3 showing histograms for the number of extreme weather episodes in July over the past ten years. Most counties reported either zero or one extreme weather episode in July, and the ratio of counties experiencing no episodes to counties experiencing at least one episode seems to gradually approach 1:1. In July 2023, for instance, this ratio reached 1.02, with 50.43% of counties being exposed to zero and 49.57% of counties being exposed to at least one extreme weather episode.

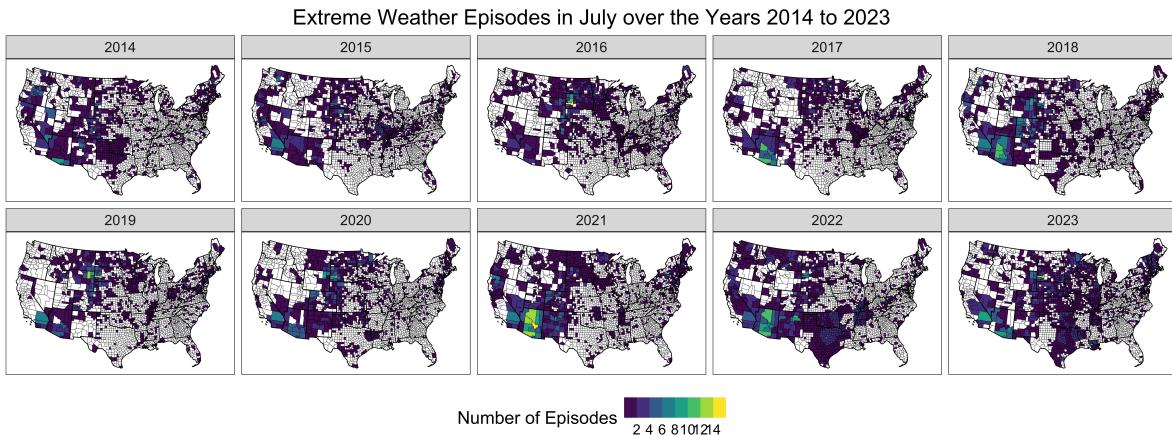


Figure 3: Maps displaying the geographical distribution of the raw number of extreme weather episodes in July over the years 2014 to 2023. The color palette indicates numbers greater than zero, and white represent a count of zero episodes.

## Conclusion

Our analyses indicate that July consistently shows a high number of extreme weather episodes with notable geographic variability (Figure 1 and Figure 2). Therefore, to maximize the likelihood of capturing suitable variability in exposure to extreme weather episodes, we plan to conduct our study at the beginning of August, ensuring that the 30-day period prior to study completion falls within July. Moreover, the main source of variability in exposure to extreme weather episodes in July is due to whether at least one episode occurred or not (Figure 4 and

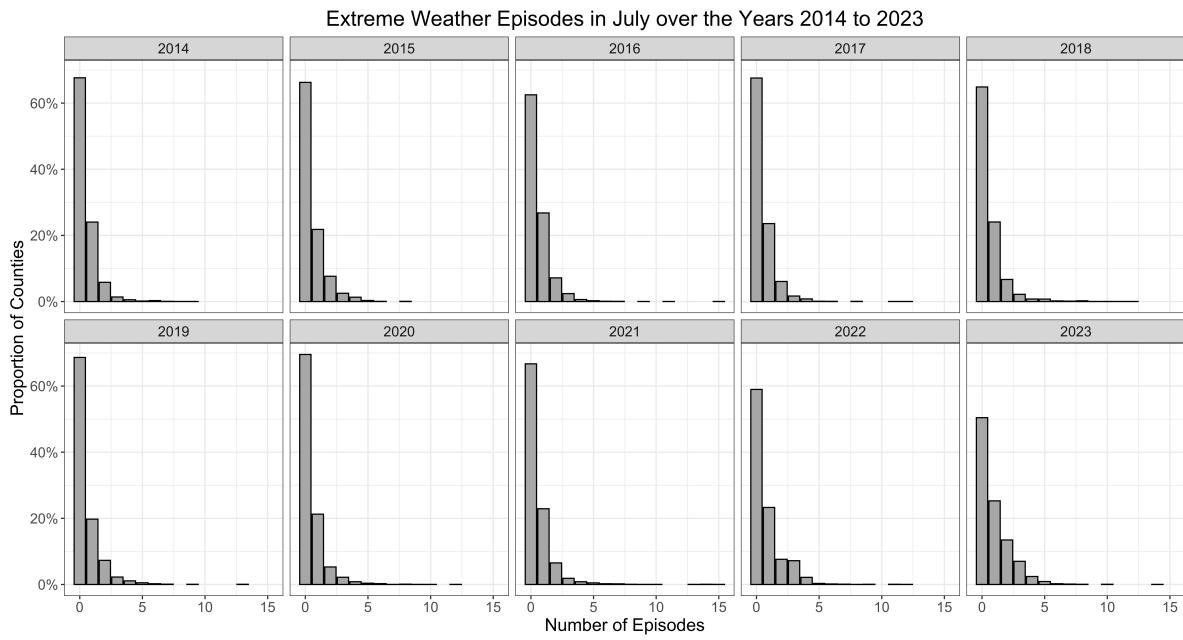


Figure 4: Histograms displaying the distribution of the raw number of extreme weather episodes in July over the years 2014 to 2023. For each number of episodes on the x-axis, the y-axis shows the proportion of counties that recorded this number of episodes.

Figure 4). Thus, our main analyses will focus on whether a participant was exposed to at least one extreme weather episode in the 30 days prior to study completion, treated as a binary variable.

### i Expand for Session Info

```
- Session info -----
  setting  value
  version  R version 4.4.0 (2024-04-24)
  os        macOS Sonoma 14.4.1
  system   aarch64, darwin20
  ui        X11
  language (EN)
  collate  en_US.UTF-8
  ctype    en_US.UTF-8
  tz       Europe/Zurich
  date     2024-08-19
  pandoc   3.1.11 @ /usr/local/bin/ (via rmarkdown)
  quarto    1.5.45 @ /usr/local/bin/quarto

- Packages -----
  package      * version date (UTC) lib source
  dplyr        * 1.1.4   2023-11-17 [1] CRAN (R 4.4.0)
 forcats      * 1.0.0   2023-01-29 [1] CRAN (R 4.4.0)
  fs           * 1.6.4   2024-04-25 [1] CRAN (R 4.4.0)
  ggplot2      * 3.5.1   2024-04-23 [1] CRAN (R 4.4.0)
  ggpubr       * 0.6.0   2023-02-10 [1] CRAN (R 4.4.0)
  lubridate    * 1.9.3   2023-09-27 [1] CRAN (R 4.4.0)
  purrr        * 1.0.2   2023-08-10 [1] CRAN (R 4.4.0)
  readr        * 2.1.5   2024-01-10 [1] CRAN (R 4.4.0)
  sessioninfo  * 1.2.2   2021-12-06 [1] CRAN (R 4.4.0)
  stringr      * 1.5.1   2023-11-14 [1] CRAN (R 4.4.0)
  tibble        * 3.2.1   2023-03-20 [1] CRAN (R 4.4.0)
  tidyverse     * 2.0.0   2023-02-22 [1] CRAN (R 4.4.0)
  usmap        * 0.7.1   2024-03-21 [1] CRAN (R 4.4.0)

[1] /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/library
```

IPCC, ed. 2023. “Weather and Climate Extreme Events in a Changing Climate.” In, 1513–

1766. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009157896.013>.