

## Assignment 4 ETC5513

Team name

Yan Ma, Yiwen Zhang, Yuqi Wang, Putu Wahyu Saputra

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Data</b>	<b>3</b>
2.1	Data source . . . . .	3
2.2	Data limitation . . . . .	3
<b>3</b>	<b>Methodology</b>	<b>3</b>
<b>4</b>	<b>Section 1: Locations</b>	<b>5</b>
4.1	Most frequently happened storms in US . . . . .	5
4.2	Thunderstorm wind in US in 2019 . . . . .	7
<b>5</b>	<b>Section 2: Injuries and Deaths</b>	<b>9</b>
5.1	Direct injuries and deaths by month and year . . . . .	9
5.2	Direct injuries and deaths by state . . . . .	11
5.3	Direct injuries and deaths by major event type . . . . .	13
<b>6</b>	<b>Section 3 : Trend of events</b>	<b>15</b>

6.1	The trend of occurrence of the top five most frequent events changing by year . . . . .	15
<b>7</b>	<b>Section 4: Tornado</b>	<b>18</b>
<b>8</b>	<b>Conclusion</b>	<b>21</b>

## 1 Introduction

Storm can disturb the surface of earth by thunderstorm, snowstorm, rainstorm, sandstorm, flood and so on to cause the road impassibility, the damage of properties and even the injuries and death. Current research Christopherson and Birkeland [2015] shows that hurricanes in the eastern North Pacific can provide water to parts of the southwestern United States and parts of Mexico. More than half of Japan's rainfall comes from typhoons, but at the same time strong convective weather such as hail can also damage buildings and cars parked on the road. , And seriously affected the output of crops such as corn and wheat. This report will focus on the storms happened in US from 2017 to 2019 to analyse it from locations, damages and frequency. This report will answer the research questions as following :

1. Section 1 :
  - What is the most frequently happened storm event in US from 2017-2019?
  - Where is the locations and time features of thunderstorm wind?
2. Section 2 :
  - How many direct injuries and deaths caused by storm events and other unusual weather phenomena in each month from 2017 to 2019 in the US?
  - How many direct injuries and deaths in each state from 2017 to 2019 in the US?
  - What's the major events caused direct injuries and deaths over years and what's the injuries and deaths by each type?
3. Section 3 :
  - What is the trend of occurrence of the top five most frequent events in US
  - How the trend changing from 2017 to 2019?
4. Section 4 :
  - Which state that hit by tornado the most and when the event is likely to happen?

## 2 Data

### 2.1 Data source

The dataset that we use in this report is Storm Events Database from 2017-2019. The data contain a chronological listing, by state, of weather phenomena (thunderstorm, hail, tornado, etc). NOAA [2019] We obtain the dataset from NOAA website that contains 31 variables and 187139 observations. NOAA [2019] The format from original source is “csv.gz”, and we extract them into .csv files.

### 2.2 Data limitation

One of the defects of this dataset is that there is no further division of the scale of disasters, because the losses, casualties and even distribution areas caused by disasters of different scales are different, for example, it is quite difficult to distinguish between small versus large hail through this dataset. So, if there is a further division of the scale of disasters, it will be more conducive to analysis.

## 3 Methodology

- This project is a descriptive research aiming at visualising and analysing the storms and other significant weather phenomena.
- Use Gromelund and Wickham [2011] package to convert the date and time columns to proper format.

- Use select function to choose variables for later analysis.
- Use rbind function to bind the data of different years.
- Use Wickham [2016] package to create plots.

The packages used are as follows :

- ggplot2 [Wickham, 2016]
- lubridate [Grolemund and Wickham, 2011]
- readr [Wickham et al., 2018]
- tidyverse [Wickham et al., 2019]
- usmap [Di Lorenzo, 2019]
- kableExtra [Zhu, 2019]
- gridExtra [Auguie, 2017]
- bookdown [Xie, 2020a]
- plot\_usmap [Di Lorenzo, 2019]
- ggthemes[Arnold, 2019]
- ggmap [Kahle and Wickham, 2013]

## 4 Section 1: Locations

This section will focus on the most frequently happened storms and their locations in US from 2017 to 2019.

### 4.1 Most frequently happened storms in US

From the Table 1 the top 10 numbers of storm events in US are all concentrate in 2017 and 2019 in the past three years. The thunderstorm wind happened most frequently, which is 18617 cases in 2019 and 16472 cases in 2017, after that is hail followed by 10398 in 2017 and 9013 in 2019.

Table 1: Top10 storm events in 3 years

EVENT_TYPE	YEAR	count
Thunderstorm Wind	2019	18617
Thunderstorm Wind	2017	16472
Hail	2017	10398
Hail	2019	9013
Flood	2019	4949
Flash Flood	2019	4072
Winter Weather	2019	3803
High Wind	2019	3772
Flash Flood	2017	3662
High Wind	2017	3536

Table 2 shows the top 10 numbers of storm events happened in US in 2019. First of all is thunderstorm wind happened 18617 times, after that is hail with 9013 cases and flood with 4949 cases and flash flood followed by 4072 cases. The other six are winter weather, high wind, winter storm, heavy snow, marine thunderstorm wind and tornado in order.

Table 2: Top5 storm event frequency in 2019

EVENT_TYPE	count
Thunderstorm Wind	18617
Hail	9013
Flood	4949
Flash Flood	4072
Winter Weather	3803

Figure 1 shows the distributions of the most frequently happened five storm events' locations in 2019. It is clear to see that these storm events happened more in eastern of US and some in the middle of US.

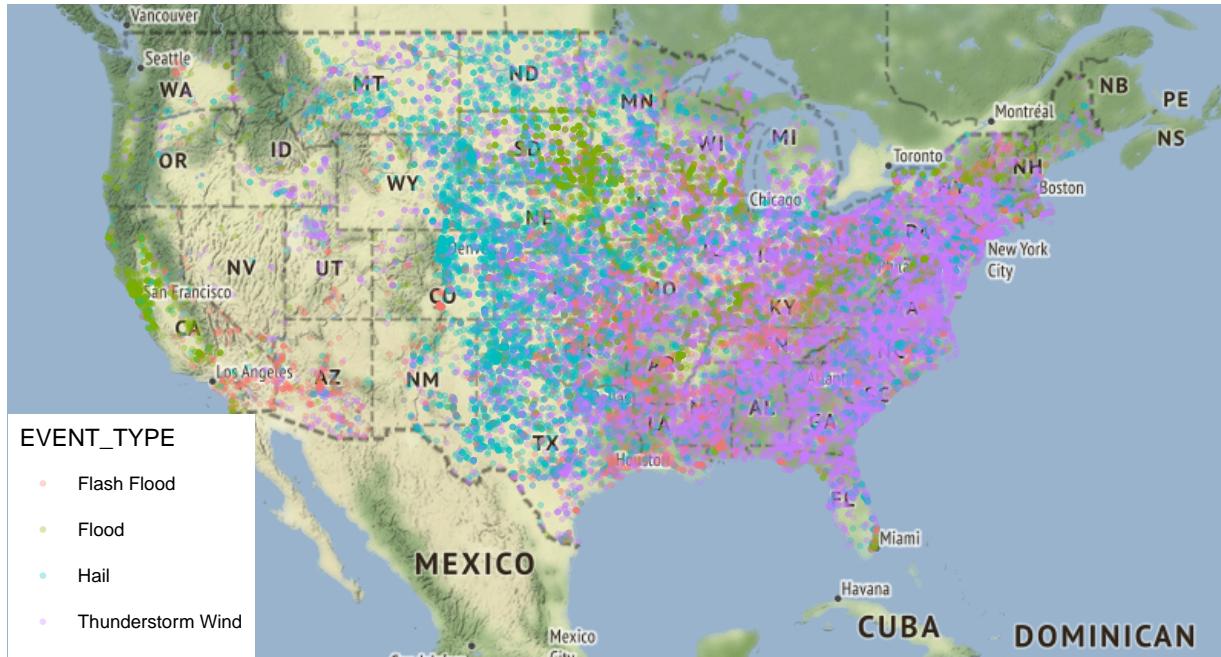


Figure 1: Top5 storm event locations in 2019

## 4.2 Thunderstorm wind in US in 2019

Figure 2 shows the thunderstorm wind locations in US in 2019 of different months. It is clear to see that the most active months of thunderstorm wind were from May to August [Christopherson and Birkeland, 2015]. Furthermore, the locations of these events happened from middle and east of US slowly transformed to the east from July to October.

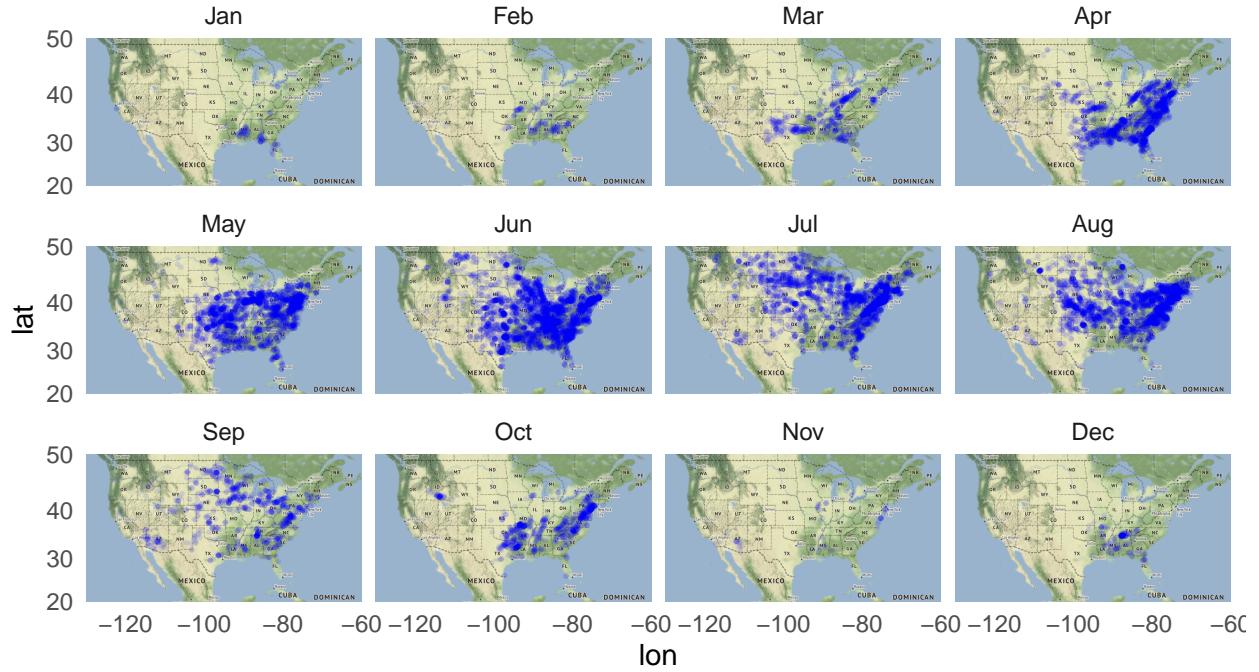


Figure 2: Thunderstorm locations by month in 2019

Table 3 displayed the number of thunderstorm wind happed in different states of US in 2019. As we can see from the table, except for Kansas, Texas and Missouri, the top ten states are all located in the east of US.

Table 3: Top10 states by number of thunderstorm in 2019

STATE	count
PENNSYLVANIA	1245
VIRGINIA	1199
TEXAS	1047
NEW YORK	903
OHIO	891
KANSAS	848
NORTH CAROLINA	791
GEORGIA	673
MISSOURI	648
SOUTH CAROLINA	619

Figure 3 shows that the number of thunderstorm wind occurred in different states by month. The thunderstorm wind most frequently happened from May to August. However the peak of the number of such cases were different from some states, August is the peak month of some eastern US states like New York, Pennsylvania and Virginia, for the middle of US states like Texas and Missouri, the peak months were May and June.

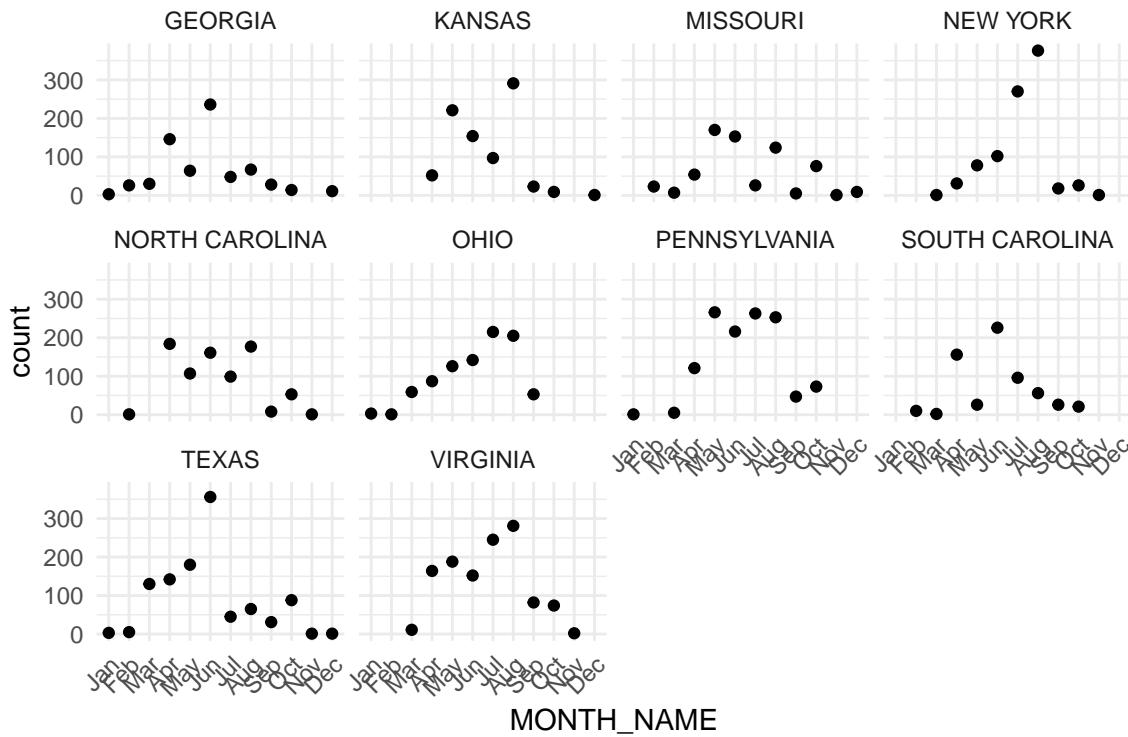


Figure 3: The nuber of thunderstorm events of top10 states in differenrent month

## 5 Section 2: Injuries and Deaths

This section will focus on the injuries and deaths caused by the storm events and other unusual weather phenomena across the US from 2017 to 2019.

### 5.1 Direct injuries and deaths by month and year

Figure 4 and 5 shows the total direct injuries and deaths of the storm events by month.

Over all the years, there are less direct injuries of storm events from September to December, while more direct injuries happened on January June and July. Most direct deaths happened on January, June to September and November.

There are several noticeable time periods. First, there are over 500 injuries and over 200 deaths on July 2018. Second, over 100 deaths happened on November 2018.

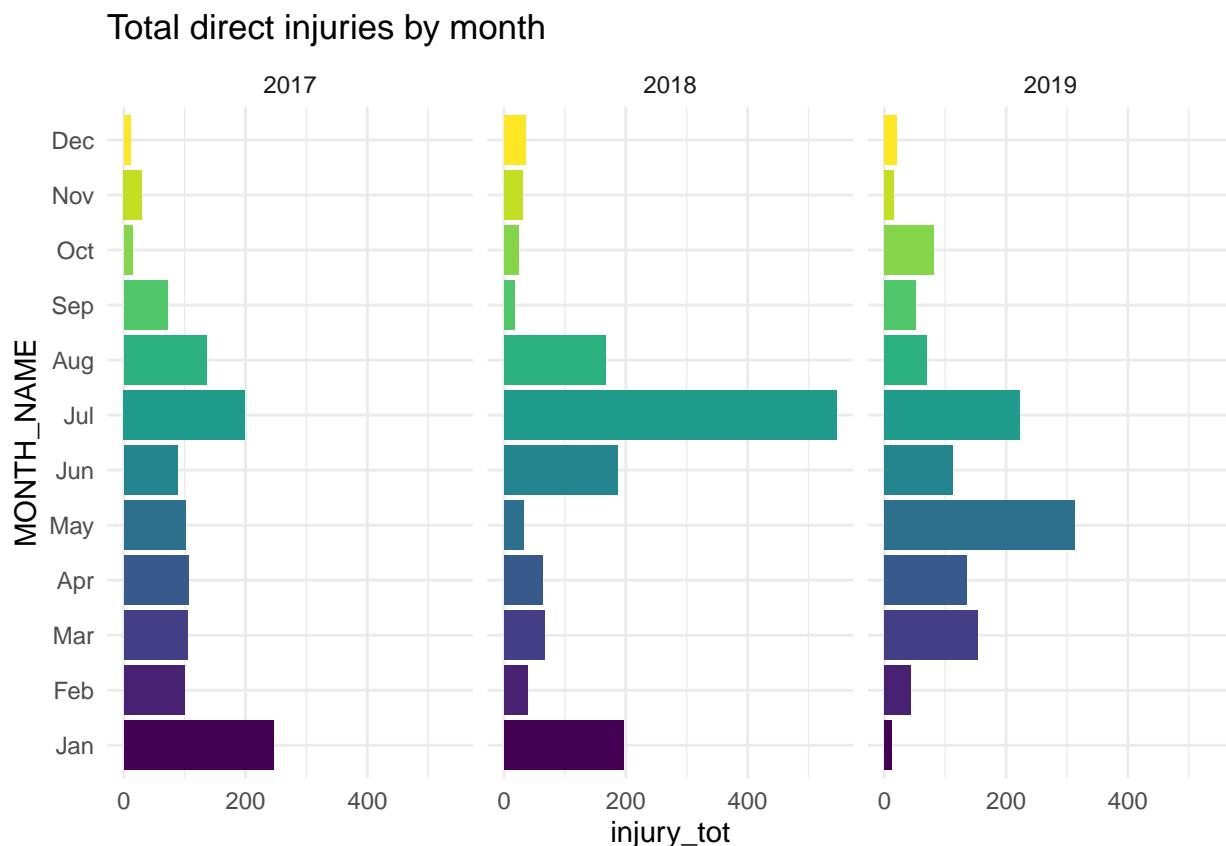


Figure 4: Total direct injuries by month

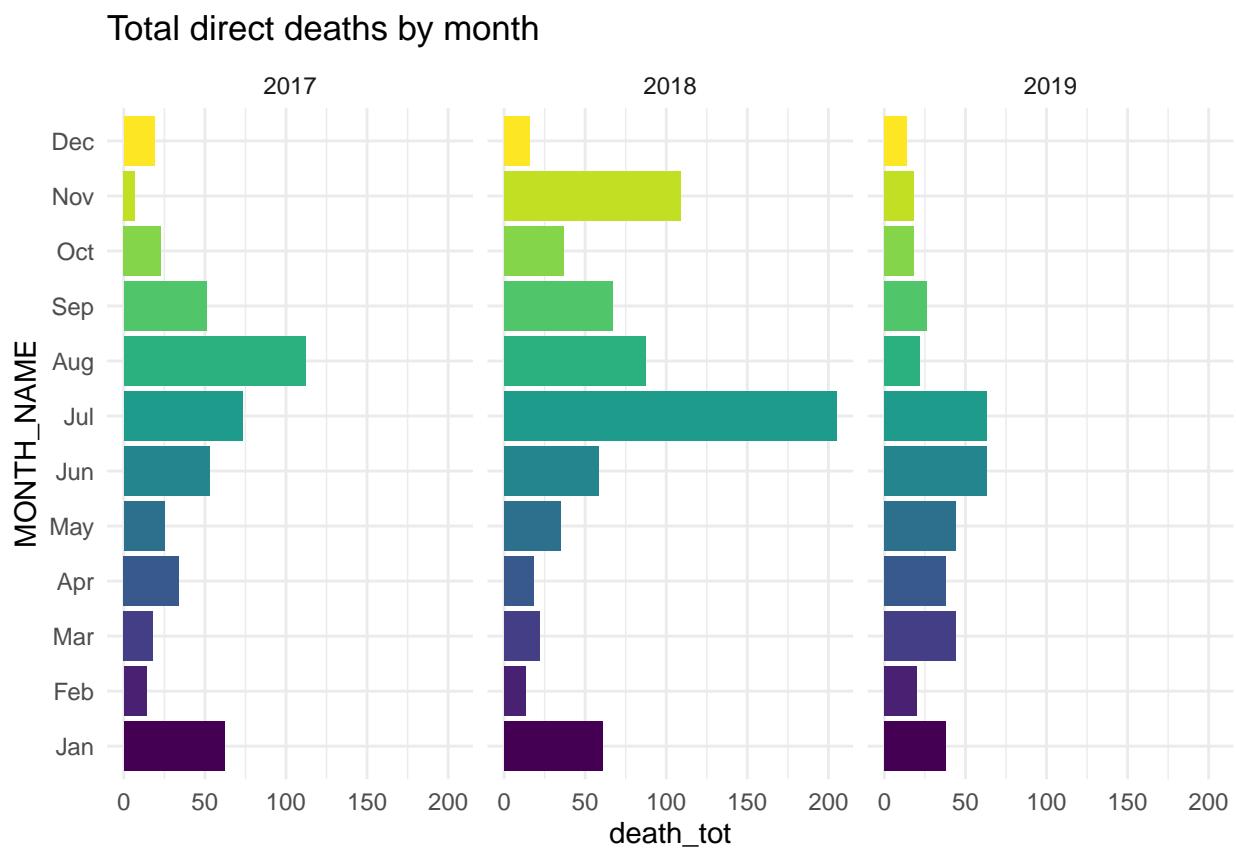


Figure 5: Total direct deaths by month

## 5.2 Direct injuries and deaths by state

Figure 6 shows the direct injuries of each state through 2017 to 2019. In 2017, Georgia had over 150 direct injuries which is the most among all the states. In 2018, California had over 350 direct injuries which is the most among all the states due to a historical largest wildfire(Wikipedia contributors [2020]). Meanwhile, Oklahoma and Missouri also had over 200 direct injuries. The most injuries in 2019 happened in Ohio, with over 200 injuries.

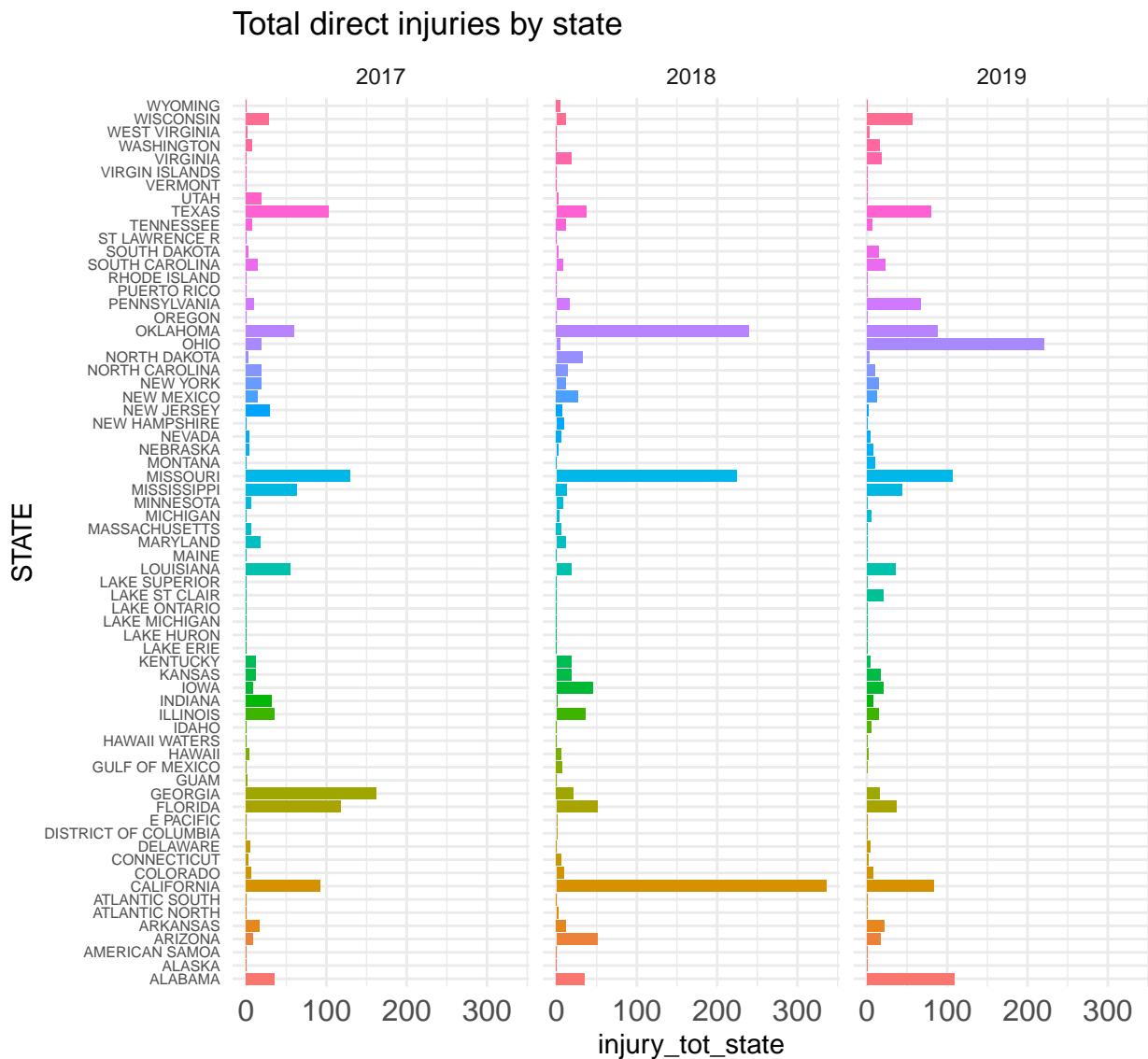


Figure 6: Total direct injuries by state

Figure 7 shows the direct deaths of each state through 2017 to 2019. In 2017, Texas had about 100 direct deaths, which is the most among all the states. In 2018, Arizona had over 150 direct deaths and California had about 150 direct deaths, which are more than other states. In 2019, no state had more than 50 direct deaths, and Texas is slightly higher than others.

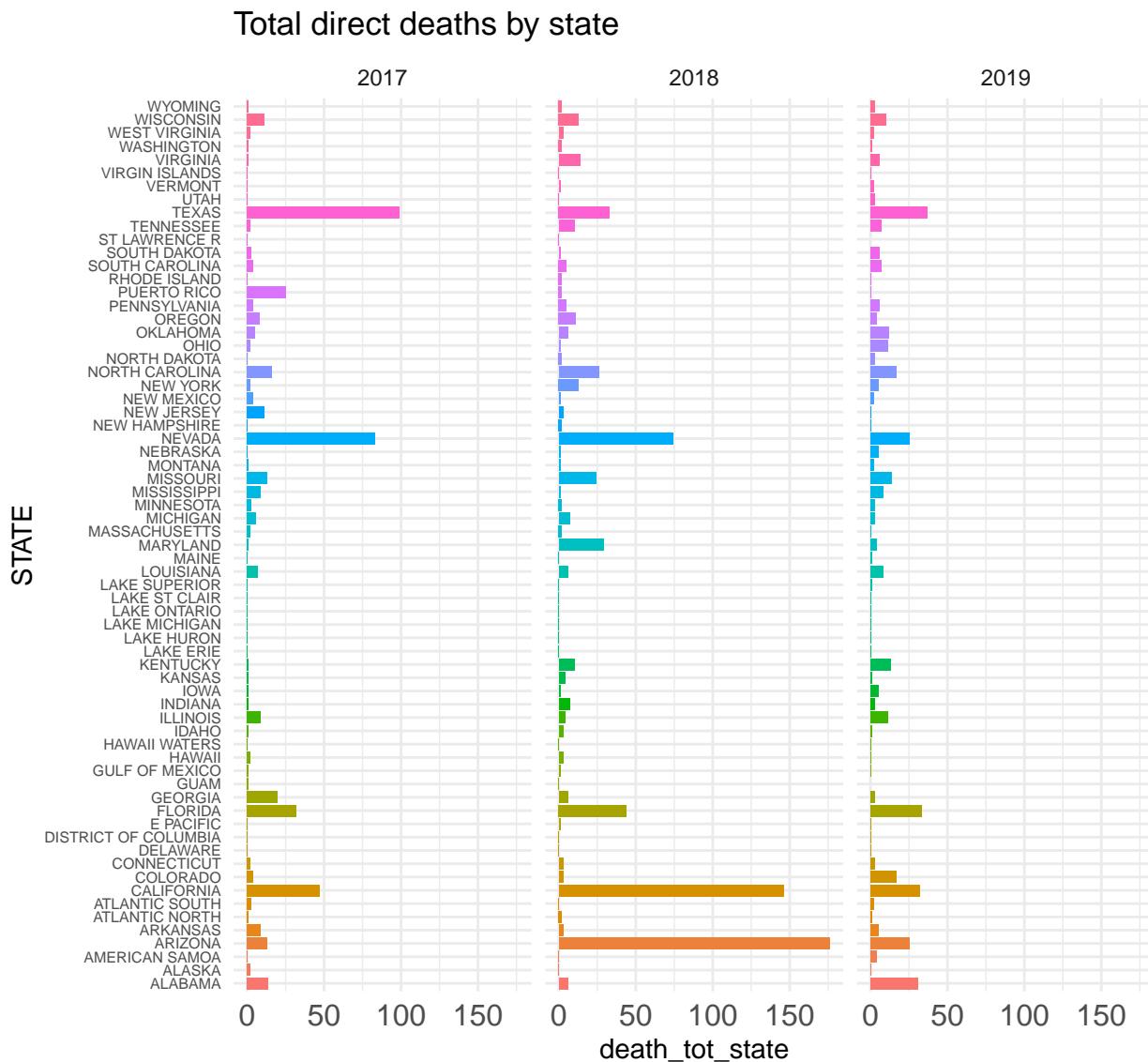


Figure 7: Total direct deaths by state

### 5.3 Direct injuries and deaths by major event type

Table 4 shows the top 5 types of events with most direct injuries from 2017 to 2019. And figure 8 shows the injuries they caused every year. Tornado caused most direct injuries in 2017 and 2019, while heat caused more direct injuries than other types of events in 2018.

Table 4: Major event types cause direct injuries over years

EVENT_TYPE	injury_event
Tornado	1253
Heat	545
Thunderstorm Wind	480
Excessive Heat	306
Lightning	268

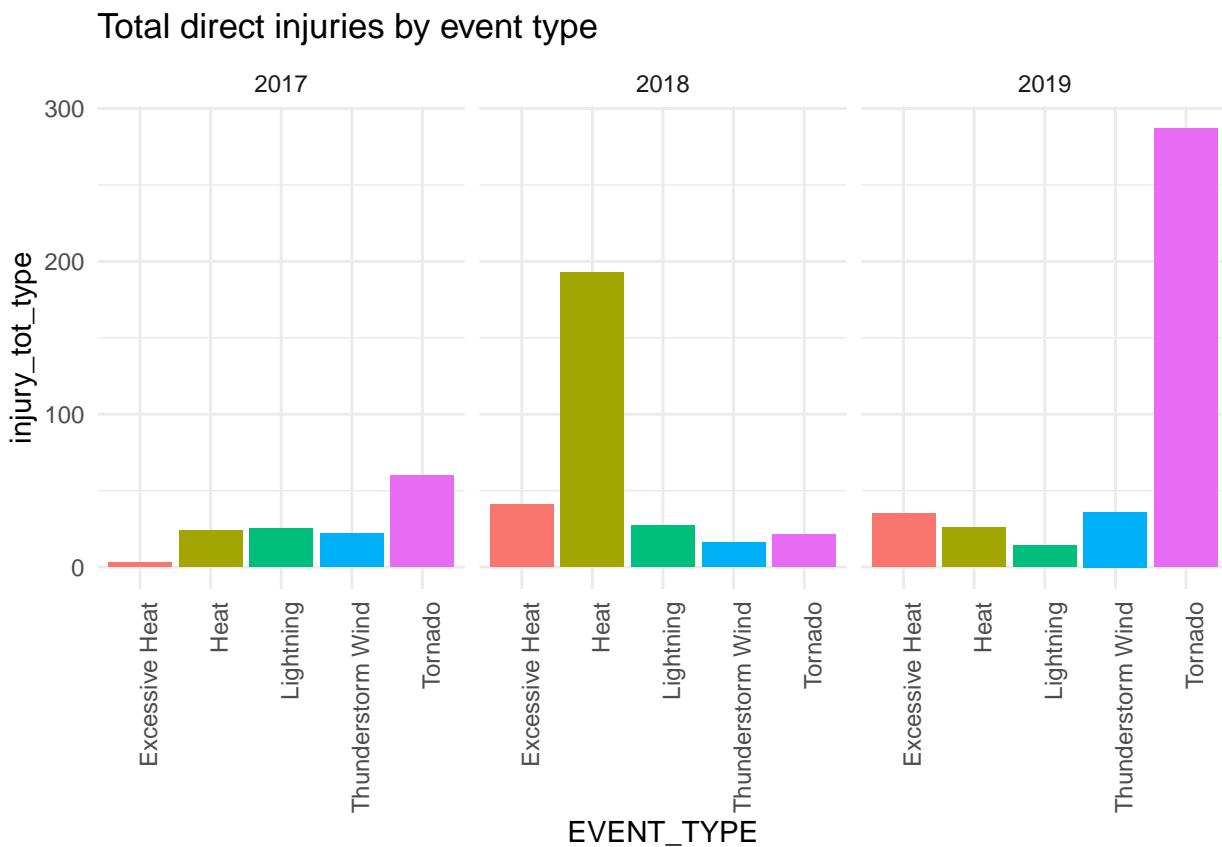


Figure 8: Total direct injuries by event type

Table 5 shows the top 5 types of events with most direct deaths from 2017 to 2019. And figure 9 shows the deaths they caused every year. Excessive Heat caused most direct deaths among the five major event types in 2018, while flash flood caused slightly more direct deaths in 2017 and 2019. In addition, direct deaths caused by wildfire is decreasing in this time period.

Table 5: Major event types cause deaths over years

EVENT_TYPE	death_event
Excessive Heat	238
Flash Flood	219
Heat	213
Rip Current	179
Wildfire	139

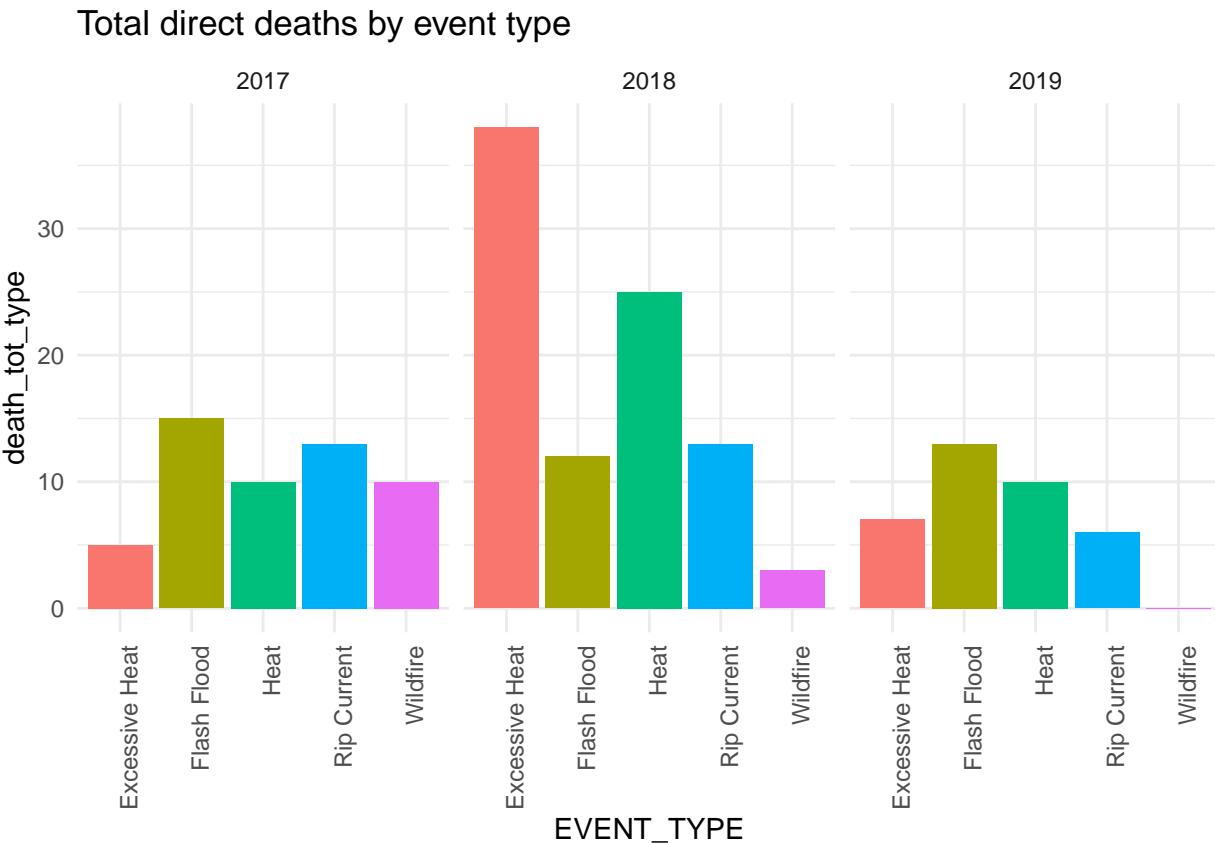


Figure 9: Total direct deaths by event type

## 6 Section 3 : Trend of events

### 6.1 The trend of occurrence of the top five most frequent events changing by year

In this section, the main purpose is to explore the trend of occurrence of the top five most frequent events changing by year both across the whole USA and in each state. First, create a table and group the data by “STATE” and “YEAR”. Then count the times of occurrence for all events and rename the name of this column as “FREQ” to present the frequency. And next, create a new table to arrange the data by descending order with displaying the top 5 events that happen most frequently. Then, use “right\_join” to connect these two tables. Finally, create two figures to display the trend of occurrence of the top 5 event changing by year.

Table 6: The frequency of the events

STATE	YEAR	EVENT_TYPE	FREQ
ALABAMA	2017	Coastal Flood	4
ALABAMA	2017	Cold/Wind Chill	1
ALABAMA	2017	Drought	92
ALABAMA	2017	Extreme Cold/Wind Chill	1
ALABAMA	2017	Flash Flood	132
ALABAMA	2017	Flood	24
ALABAMA	2017	Frost/Freeze	11
ALABAMA	2017	Funnel Cloud	2
ALABAMA	2017	Hail	201
ALABAMA	2017	Heavy Rain	6

Table 7: Top five most frequent events

EVENT_TYPE
Thunderstorm Wind
Hail
Flood
Flash Flood
Winter Weather

Table 8: Occurrence of top five most frequent events

STATE	YEAR	EVENT_TYPE	FREQ
ALABAMA	2017	Thunderstorm Wind	630
ALABAMA	2018	Thunderstorm Wind	495
ALABAMA	2019	Thunderstorm Wind	591
ALASKA	2019	Thunderstorm Wind	2
ARIZONA	2017	Thunderstorm Wind	116
ARIZONA	2018	Thunderstorm Wind	185
ARIZONA	2019	Thunderstorm Wind	97
ARKANSAS	2017	Thunderstorm Wind	423
ARKANSAS	2018	Thunderstorm Wind	411
ARKANSAS	2019	Thunderstorm Wind	359

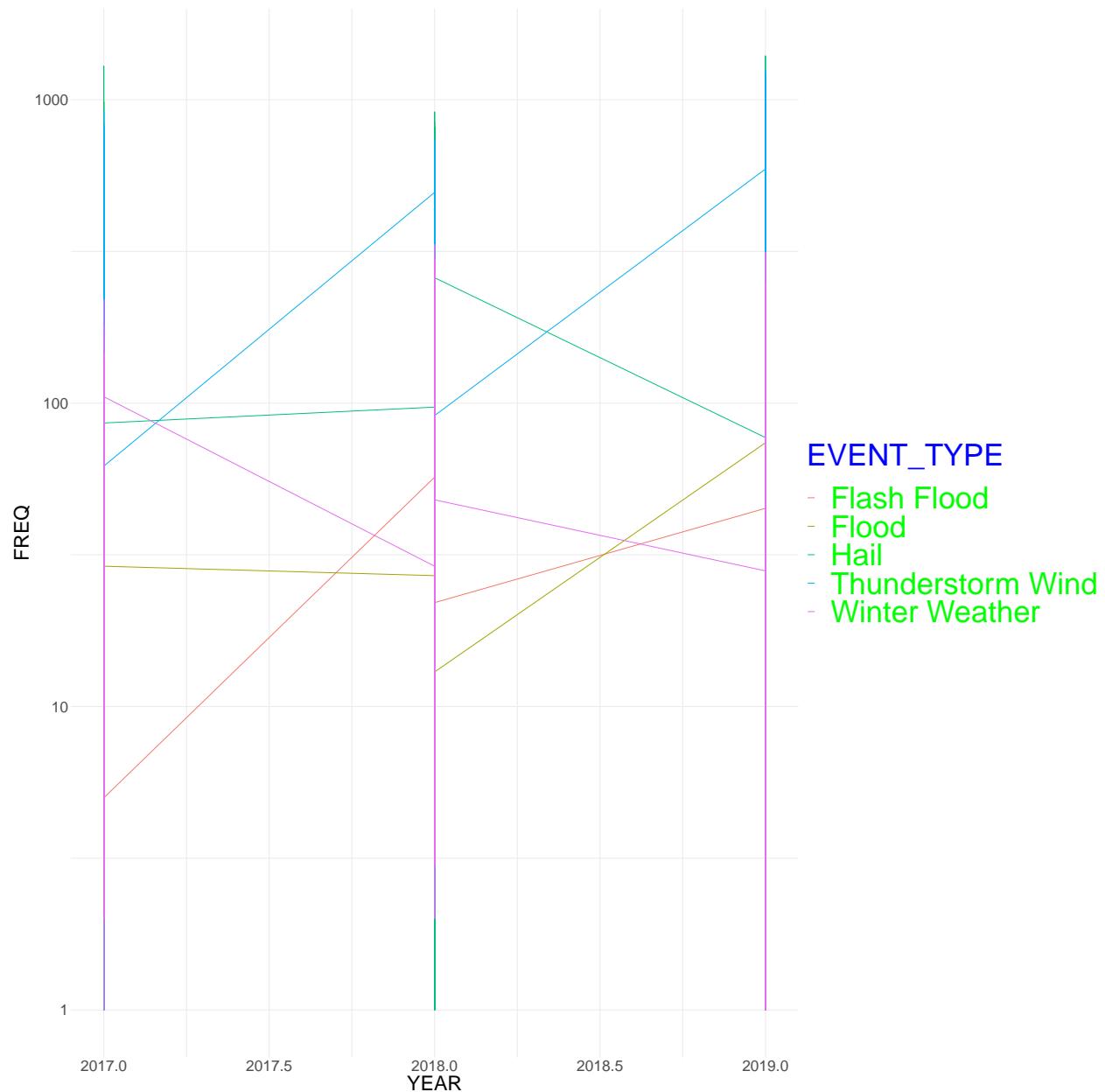


Figure 10: Occurrence of top five most frequent events changing by year

Table 6 displays the first ten lines of the frequency of all events. And Table 7 displays the top five most frequent events. And Table 8 displays the first ten lines of the final results of the occurrence of top five most frequent events. It can be seen that top 5 events have occurred in some regions in three years for example, thunderstorm wind has occurred in Alabama in three years from 495 times to 630 times , while it has only occurred twice in Alaska in 2019. Figure 10 presents the occurrence of top five most frequent events across the whole USA, we could see that thunderstorm wind showed an increasing trend during three years. And hail showed a slow growth in 2017-2018, followed by a sharp decline in 2018-2019. As for the winter weather, it has been on a downward trend for three years, but the slowdown from 2017 to 2018 was larger than that from 2018 to 2019. What's more, flash flood kept growing, but slowed down between 2018 and 2019. At last, flood kept a slow decline at first, then turned into a significant rise. We could see that most events have a similar trend, but hail has decreased in 2018 to 2019 years, which is possibly due to increased melting level heights and greater atmospheric instability.[Dessens et al., 2015] Correspondingly, the number of floods increased from 2018 to 2019, which is increasingly common due to years of relative sea level increases and El Nino.[flo]

## 7 Section 4: Tornado

Based on the calculation of the number of tornadoes that occur each year, the United States becomes a country with more than 1000 events each year. NOAA Table 9 shows the number of tornadoes from 2017 to 2019 in the United States. The table made by `kable` function from `knitr` package. Xie [2020b]

Table 9: Total number of Tornado that occurs from 2017-2019 in United States

YEAR	EVENT_TYPE	n
2017	Tornado	1647
2018	Tornado	1250
2019	Tornado	1728

We use `ggplot2` package to map the distribution overtime. Wickham [2016] To see the number of tornadoes that occur each month during these three years can be seen in Figure ???. From this plot, it can be seen that most tornadoes occur on average in the range of April and May. 2019 recorded a high number of tornadoes in May compared to the previous two years.

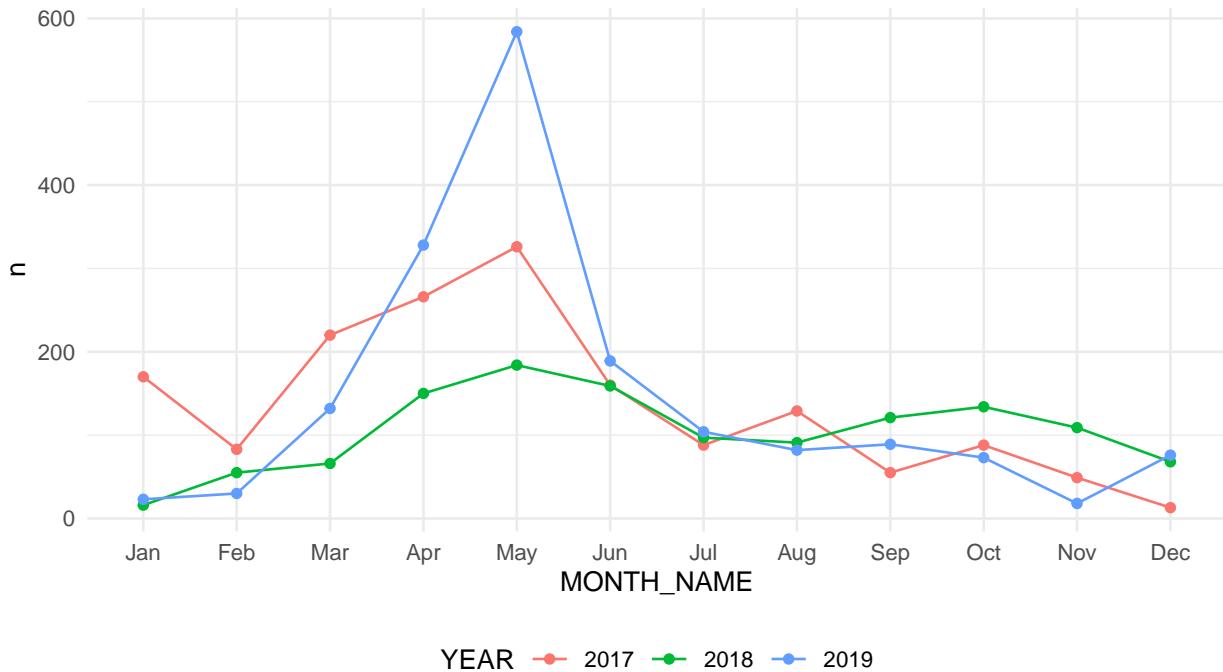


Figure 11: Tornado frequency from 2017-2019 in USA

Figure 12 shows the map of the United States with the highest number of tornadoes during 2017-2019. The map plotted using `plot_usmap` function from `usmap` package. Di Lorenzo [2019]

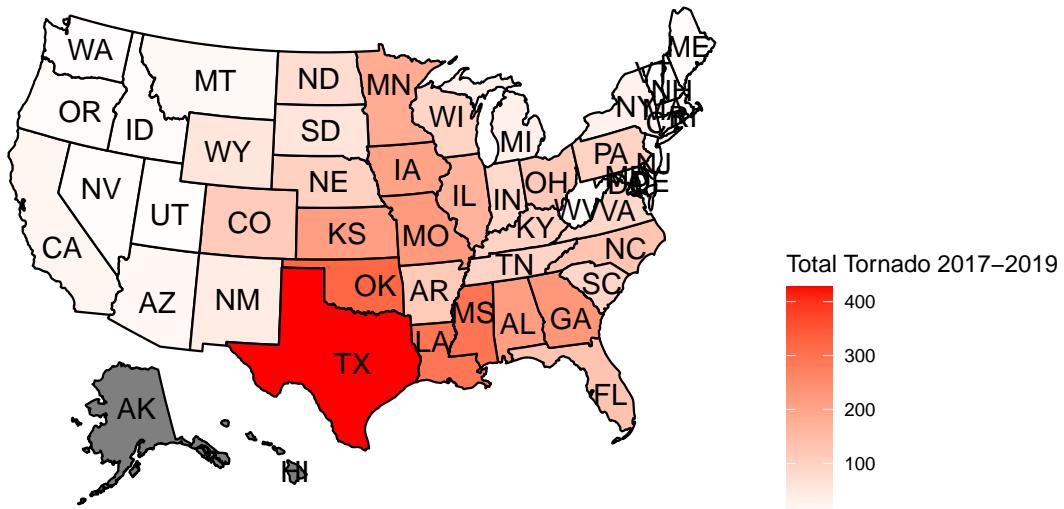


Figure 12: United States map of total tornado in each state

Each tornado that occurs has a scale. Based on dataset documentation, enhanced Fujita Scale describes the strength of the tornado based on the amount and type of damage caused by the tornado. NOAA [2014] The F scale listed as Table 10 :

Table 10: Level of Tornadoes

Scale	Category	Speed
EF0	Light	40 – 72 mph
EF1	Moderate	73 – 112 mph
EF2	Significant	113 – 157 mph
EF3	Severe	158 – 206 mph
EF4	Devastating	207 – 260 mph
EF5	Incredible	261 – 318 mph
EFU	Unknown	Unknown

When there is no tornado damage, but there is evidence a tornado existed (citation)

In Table 11, Texas is the state that has been hit the most by Tornadoes—followed by Kansas, Alabama, Missouri, and Minnesota. The top four states have experienced Tornadoes on an EF4 scale from 2017 to 2019. The table style use `kableExtra` package. Zhu [2019]

Table 11: Tornado frequency by scale for each state in USA from 2017-2019

STATE	EF0	EF1	EF2	EF3	EF4	EFU
TEXAS	201	92	45	13	1	75
KANSAS	116	41	11	6	1	38
ALABAMA	111	86	18	1	1	NA
MISSOURI	102	98	14	5	1	2
MINNESOTA	101	75	7	NA	NA	1

As the state with the highest number of tornadoes, we analyze Texas deeper. Figure 13 shows the number of tornadoes that occur overtime based on scale. In general, the number of tornadoes varies each year. EF0 occurs almost every year, but in 2019, the number of EF0 occurs more than 40 times in May. In 2018, the number of tornadoes on the EF1 and EF2 scales was less than in 2017 and 2019. The EF4 happened once in 2018.

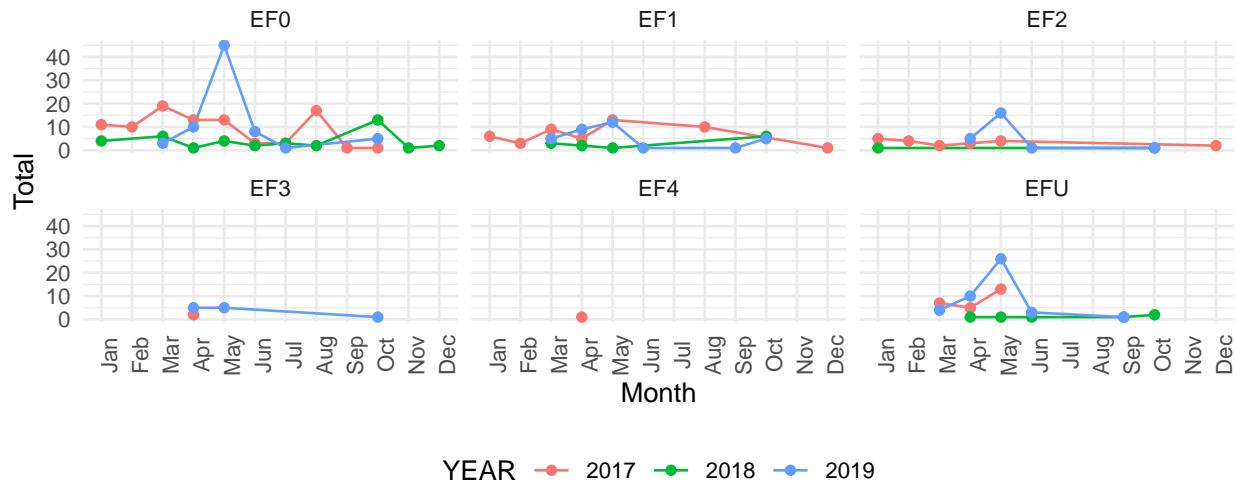


Figure 13: Tornado scale per month in Texas

Figure 14 tells what time of day a tornado with a particular scale occurs in Texas in 2019. Generally, there are differences in each month. May is the most prominent month with almost every day some tornadoes occur.

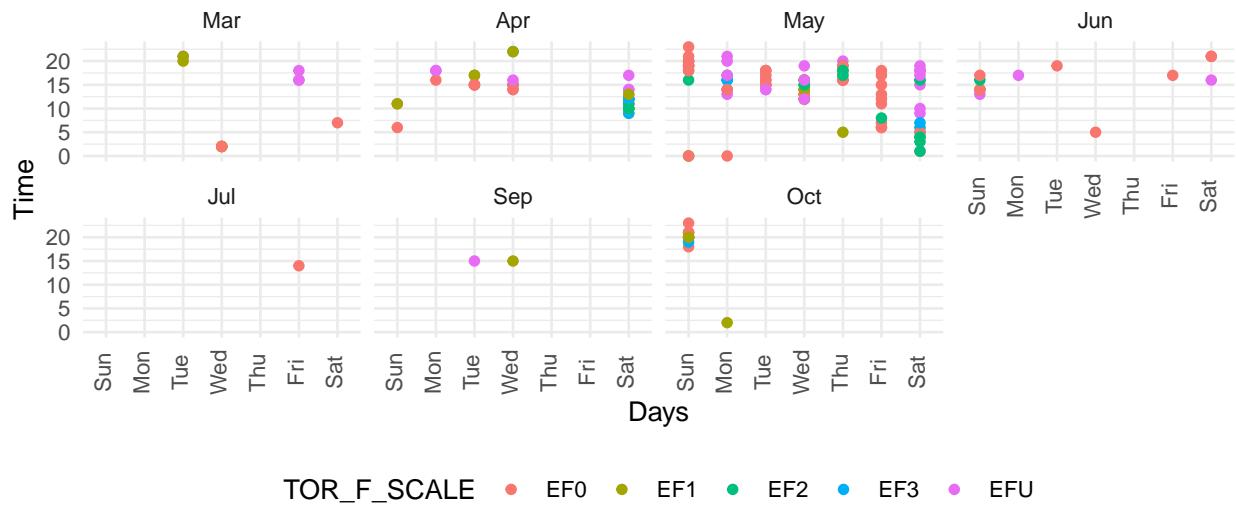


Figure 14: Tornado by hour distribution that happened in Texas in 2019

## 8 Conclusion

- The most active months of thunderstorm wind were from May to August. Furthermore, the locations of these events happened from middle and east of US slowly transformed to the east from July to October.
- Over all the years, there are less direct injuries of storm events from September to December, while more direct injuries happened on January June and July. Most direct deaths happened on January, June to September and November.
- In 2017, Georgia had the most direct injuries and Texas had the most direct deaths among all the states. In 2018, California had the most direct injuries and Arizona had the most direct deaths among all the states. In 2019, no state had more than 50 direct deaths, and Texas is slightly higher than others while Ohio had the most direct injuries.
- Tornado caused most injuries in 2017 and 2019, while heat caused more injuries than other types of events in 2018. Excessive Heat caused most direct deaths among the five major event types in 2018, while flash flood caused slightly more direct deaths in 2017 and 2019. In addition, direct deaths caused by wildfire is decreasing in this time period.
- Seeing from top five most frequent events, most events have a similar trend. However, hail has decreased in 2018 to 2019 years with floods increasing instead.
- Tornado mostly happen in May while Texas is the state with the highest number of tornado in USA.

## References

- U.s. ties record for number of high tide flooding days in 2018. URL <https://www.noaa.gov/media-release/us-ties-record-for-number-of-high-tide-flooding-days-in-2018>.
- Jeffrey B. Arnold. *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*, 2019. URL <https://CRAN.R-project.org/package=ggthemes>. R package version 4.2.0.
- Baptiste Auguie. *gridExtra: Miscellaneous Functions for "Grid" Graphics*, 2017. URL <https://CRAN.R-project.org/package=gridExtra>. R package version 2.3.
- W Robert Christopherson and Ginger Birkeland. *Geosystems: An Introduction to Physical Geography:[Julien Moreau: editing the questions of each chapter of the book]*. Pearson Education Limited, 2015.
- J. Dessens, C. Berthet, and J.l. Sanchez. Change in hailstone size distributions with an increase in the melting level height. *Atmospheric Research*, 158-159:245–253, 2015. doi: 10.1016/j.atmosres.2014.07.004.
- Paolo Di Lorenzo. *usmap: US Maps Including Alaska and Hawaii*, 2019. URL <https://CRAN.R-project.org/package=usmap>. R package version 0.5.0.
- Garrett Grolemund and Hadley Wickham. Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3):1–25, 2011. URL <http://www.jstatsoft.org/v40/i03/>.
- David Kahle and Hadley Wickham. ggmap: Spatial visualization with ggplot2. *The R Journal*, 5(1):144–161, 2013. URL <https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>.
- NOAA. U.s. tornado climatology | national centers for environmental information (ncei) formerly known as national climatic data center (ncdc). URL <https://www.ncdc.noaa.gov/climate-information/extreme-events/us-tornado-climatology>. (Accessed on 06/07/2020).
- NOAA. *Storm Data Export Format*, 2014. URL <ftp://ftp.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/Storm-Data-Export-Format.pdf>.
- NOOA. Storm events database, 2019. URL <https://www.ncdc.noaa.gov/stormevents/>.
- Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL <https://ggplot2.tidyverse.org>.
- Hadley Wickham, Jim Hester, and Romain Francois. *readr: Read Rectangular Text Data*, 2018. URL <https://CRAN.R-project.org/package=readr>. R package version 1.3.1.
- Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686.
- Wikipedia contributors. 2018 california wildfires — Wikipedia, the free encyclopedia, 2020. URL [https://en.wikipedia.org/w/index.php?title=2018\\_California\\_wildfires&oldid=956440002](https://en.wikipedia.org/w/index.php?title=2018_California_wildfires&oldid=956440002). [Online; accessed 9-June-2020].
- Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2020a. URL <https://github.com/rstudio/bookdown>. R package version 0.19.
- Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2020b. URL <https://yihui.org/knitr/>. R package version 1.28.
- Hao Zhu. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*, 2019. URL <https://CRAN.R-project.org/package=kableExtra>. R package version 1.1.0.