

# NOAA Storm Events in Literature

```
library(noaastormevents)
library(tidyverse)
library(viridis)
library(lubridate)
events_2019 <- create_storm_data(date_range = c("2019-01-01", "2019-12-31"))
```

## What is the Storm Events Database

The **National Oceanic & Atmospheric Administration (NOAA) Storm Events database** collects information on storms, significant weather phenomena, rare or unusual weather, and other significant meteorological events across the United States. NOAA stores this information in csv files for each year. It has been recording weather information since 1950. However, the database has undergone several changes since then. In 1950, NOAA only recorded tornado events and now it records 48 different event types. Following these changes, the NOAA Storm Events website states that they have reformatted and standardized events types without changing any specific values or details (“Storm Events Database,” n.d.).

The **database** assigns each weather phenomenon with a location, date, event type, event ID, episode ID, property damage estimate, crop damage estimate, county name, state, event narrative, and episode narrative. The database categorizes large storms as episodes which contain several individual events. For example, a hurricane will be given one episode ID, and the rain, wind, floods, etc associated with that hurricane will be given event IDs that fall under that episode.

The US National Weather Service (NWS) collects **weather data** from storm trackers, federal agencies, the media, the public, and several other sources. The NWS then uses this data to create the NOAA Storm Events database that is released as a monthly publication.

[BA: I think that one of the things it’s important to understand about this database is how they record with a hierarchy, so there will be different listings for one storm both across several types of hazards and several geographic locations. For example, a big hurricane is one storm system but would result in listings for lots of counties and for multiple hazards (storm surge, wind, extreme rain, etc.) within some of those counties. They’ve got ID numbers that help in joining together different listings for the same storm. It might be helpful for us to add a figure somewhere early in the draft to explain this aspect of how listings are included in the database. You could take a look through the noaastormevents to see if there’s a figure or illustration there that might make a good starting point for helping to illustrate that.]

## Use of Storm Events Database in research on societal impacts

[BA: Let’s add some examples of papers in which this database was used to identify storm exposures, and then the paper was looking at how they impacted humans.]

## Concerns with Storm Events Database

[BA: Could we start on a paragraph here that gives an introduction to this section about potential biases?

One thing here would be to talk about the idea of the probability of an event being recorded given that it happens.]

Though the NOAA Storm Events database is quite extensive, **several biases affect its data**. These biases fall under five major categories which are common in weather data. These include hazard bias, temporal bias, threshold bias, accounting bias, and geographic bias. These biases result from various factors including structural changes over time, reporting errors, inherent bias, and others.

The Storm Events website gives a disclaimer that “an effort is made to use the best available information but because of time and resource constraints, **information from [the] sources may be unverified by the NWS**” (“Storm Data Faq Page,” n.d.).

- extensive database
- several changes over time
- not all events may be reported, some may be double counted
- lots of sources
- inherent bias in data collection in this specific field

[BA: Maybe it would be helpful for us to start a small table here that lists each type of bias that we cover and a short description of each? We can start that as an itemized list, because that will be easier to format right now, and then convert it to a table later.]

- **hazard bias:** [Add a short definition / description]
- **temporal bias:** [Add a short definition / description]
- **threshold bias:** [Add a short definition / description]
- **accounting bias:** [Add a short definition / description]
- **geographic bias:** [Add a short definition / description]

## Hazard Bias

[BA: I think there might be a few different things going on with this one. Let’s talk about how we can make sure we bring these different elements together in describing this type of bias. Here are some of the elements that seemed to me to be coming up here: (1) some types of events might be more likely to be recorded than others; (2) some types of hazards within an event might be more likely to be recorded than others; (3) for some types of events, maybe there’s some encouragement to list it as a single event; (4) for the events that are listed, they can differ a lot in severity, and there’s not much information in the database to help figure out how severe they are; (5) some of the listings are coming from different sources (e.g., media rather than the NWS)]

Within the Storm Events database, there may be inaccurate recording in the number of certain hazard types [BA: Could we be more precise with this statement? For example, is this likely to always (or almost always) be an underreporting rather than an overreporting? Also, for this type of bias, maybe we want to clarify that, even when storms are entered into the database, all of their associated hazards might not be entered. That’s what’s going on with this type of bias, right? If I’ve understood this correctly, maybe move towards something like, “The Storm Events database aims to record listings for all hazards brought by a particular storm event. However, in some cases, when a storm event is recorded certain types of hazards within the storm can tend to go unreported.” Have I gotten the idea right here?].

## Introduction

Something like: The Storm Events database aims to record listings for all hazards brought by a particular storm event. However, in some cases, when a storm event is recorded **certain types of hazards within the storm can tend to go unreported**. [BA: We may want to add something like, “and, in general, the probability

that an event is recorded here might differ by type of event.” I’m thinking that this type of hazard would especially come into play for a study where you’re trying to *compare* the frequency of two or more hazards with each other, or even trying to see how each is related with an outcome. We might want to see if we can find some examples of studies that have used this database while looking at two or more types of hazards, to help us in thinking through how a reporting bias that’s different by type of hazard might bias the results. Many of our example studies focus on one type of hazard (e.g., tornadoes, rip tides), but maybe we can find some that consider two or more in the same study?]

## Major Points:

1. some types of events might be more likely to be recorded than others
  - **overrepresentation of flood events** in Storm Data because of how they get information [BA: Is this because the flood insurance program (or at least some part of it) is through the federal government? If so, we might need to add that information to clarify this statement.]
  - “NWS is obliged to provide monetary loss estimates for any flood event even if the damage assessment is a ‘guesstimate’” (Gall, Borden, and Cutter 2009)
  - drought hazards are apparently notoriously underreported because there is a lack of physical damage and it is **hard to quantify spatial and monetary losses** (Gall, Borden, and Cutter 2009) [BA: This a *great* example of this point]
  - P(event recorded | event happened) may be different by events [BA: Here, do we mean that it might be different for tornado events versus avalanche events, for example? Or do we mean that, for any type of event, it might be different depending on some characteristics of the event (for example, how severe a specific event was)? Or maybe both?] [BA: I’ve added a bit on this point through my comment in the Introduction. This seems to be the heart of this type of bias, right?]
    - I think maybe both. As we have been starting to discuss, events that happen closer to larger populations of people may have a higher likelihood of being reported than an event that is in a more secluded area. (So this could apply to tornadoes vs avalanches) Additionally, events that are more severe may get more attention than events of lower severity. But, I’m not sure if it is worth it to explain this here or just explain it in a different section where it is more clear. [BA: I see. I think we’ve got the geographic one well-covered in that section, although we could mention here that it comes in to play if a certain *type* of event tends to happen in well-populated versus less-populated areas (or even if it’s large enough that its coverage includes at least some well-populated areas, like a hurricane often does). The idea that more severe events are more likely to be included is a very interesting point here, and one that I think we could distinguish from the idea of different *types* of hazards having different probabilities of being reported. Any thoughts on any examples from the vignettes that we could use as a starting point for illustrating and exploring these points?]
    - “fatal weather events are more likely to be reported due to enhanced media attention” (Gensini and Ashley 2010) [BA: This would play in with both the event severity question (more severe events are more likely to be deadly) and where the event occurs (so somewhat in the geographic), as higher population density would contribute to risk of deaths. I think that one of the tornado papers might cover this idea somewhat?]
2. even when storms are entered into the database, **all of their associated hazards might not be entered or vice versa** (if there is a large event that causes other events, it is as if the events are getting double counted)
  - example from Gall losses: hurricane wind, storm surge, and a tornado that were all caused by a hurricane get counted as three separate events even though they all stemmed from one hurricane
  - one episode can correlate to multiple events (Konisky, Hughes, and Kaylor 2016)

- script for how noaa deals with this? or examples of this? Probably explained in narrative sections of events [BA: This comes in with the episode ID and the event ID. I think there's an example in the vignette where we check this out. One is used to help collect all the types of hazards from one overall event using a common ID. We might want to bring that analysis in here.]

### Number of events reported per episode ID:

For these episodes with the most events in 2019, the following graph shows the number of events reported for the episode. This figure demonstrates how **one large weather 'episode' can include several other events types**. For example, Episode 133801 in this figure indicates several reports of flood, debris flow, strong winds, high wind, and hail. All of these different event types occurred as a result of one large storm labelled episode 133801. This unique episode ID prevents related events from getting double counted as separate entities.

[Can we add anything from the database's documentation in terms of how they decide how to define an episode and how to link events with that episode? Or really, anything that have in their documentation on this episode / event piece might be interesting to try to work in here.]

Copied from NWS Instruction Documentation:

> 2.9.1 Episode Narrative. Generate an episode with a narrative; otherwise individual events cannot be entered into the Storm Data software. An episode narrative describes the entire episode in a general fashion, and briefly describes the synoptic meteorology associated with the episode. Information in the episode narrative can be very useful for researchers and other users of Storm Data. This narrative does not need to be long or elaborate, rather make it brief and informative. An example would be "A strong cold front passed through the Washington, D.C.area, triggering several instances of damaging thunderstorm winds and large, baseball-size hail."

To ensure events being logged in a single episode are part of the same synoptic meteorological system, events within the same episode may begin no more than five (5) calendar days apart. This will enable the Storm Data preparer to properly document events that double back into a specific region or events that are very slow moving. Examples include Hurricane and Winter Storm events.

The episode narrative will appear in the Storm Data publication after all events contained within the episode. The episode narrative does not appear in the examples shown in Appendix A, which is reserved for only event narratives. Additionally, a brief summary of fatalities and injuries should be part of the episode narrative for zone-based events.

2.9.2 Event Narrative. Detailed information pertaining specifically to the event and not the overall episode will appear in an event narrative. The event narrative describes the significance or impact of an event within an episode. An event narrative is required for all Tornado events, all Thunderstorm Wind events, and all Lightning events, whether over land or marine zones.

This narrative will appear immediately below the header-strip in the publication and should contain descriptive information about the times, locations, and severity of destruction of property, trees, crops, power lines, roads, bridges, etc. Additionally, a brief summary of fatalities and injuries should be part of the event narrative for county-based and marine zone-based events. For Thunderstorm Wind events with estimated gusts, use sentences such as "Several trees were toppled by powerful downburst gusts."

```
## Warning: `select()` is deprecated as of dplyr 0.7.0.
## Please use `select()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.

## Warning: `filter()` is deprecated as of dplyr 0.7.0.
## Please use `filter()` instead.
## See vignette('programming') for more help
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

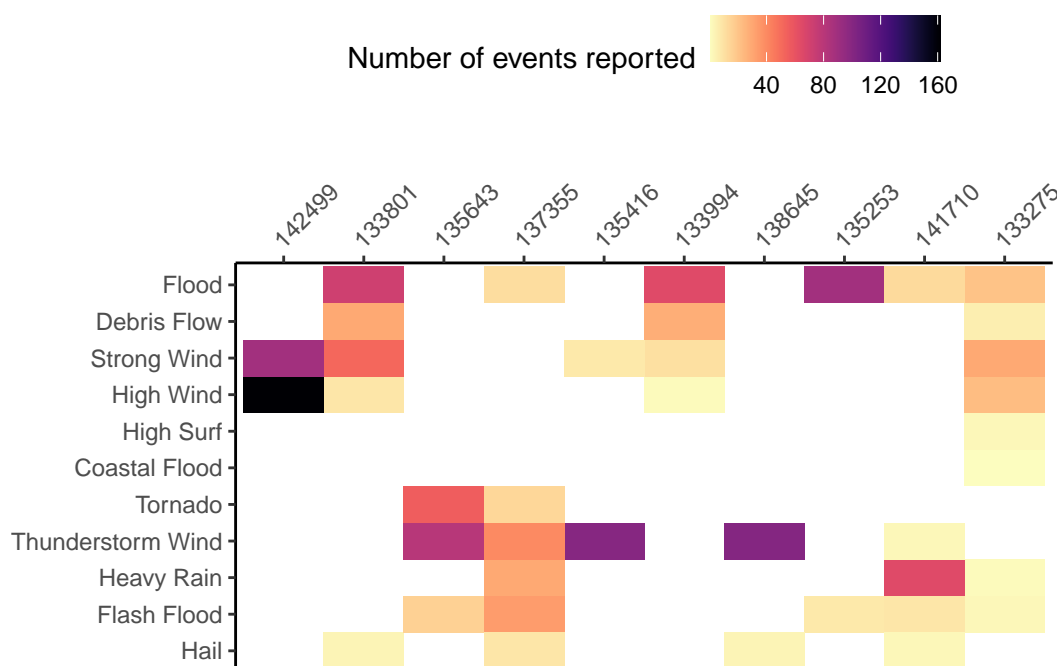
```
## Warning: `mutate_()` is deprecated as of dplyr 0.7.0.
## Please use `mutate()` instead.
## See vignette('programming') for more help
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

Table 1: Table continues below

event_type	damage_property	damage_crops	cz_name
------------	-----------------	--------------	---------

state	event_narrative	episode_narrative
-------	-----------------	-------------------

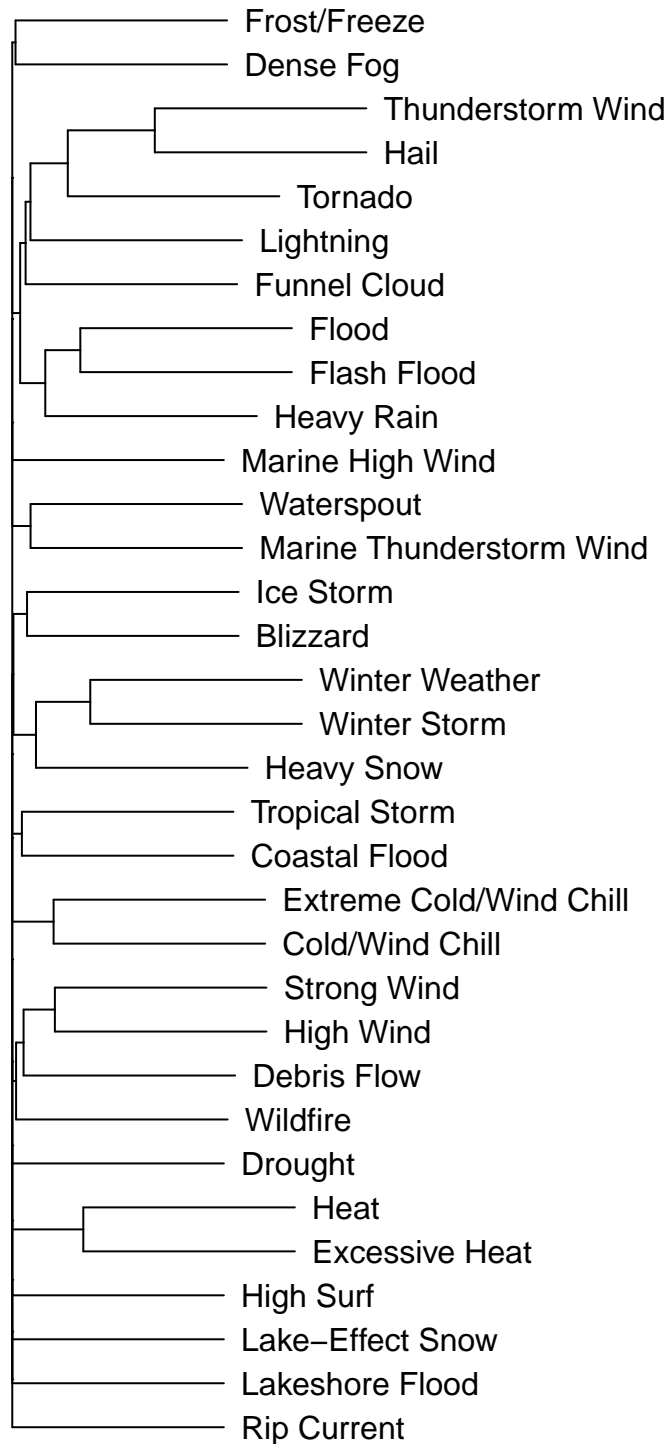
```
## `summarise()` ungrouping output (override with `.groups` argument)
## `summarise()` regrouping output by 'EPISODE_ID' (override with `.groups` argument)
## `summarise()` regrouping output by 'EPISODE_ID' (override with `.groups` argument)
```



## Cluster Analysis

Once we removed event types with less than 50 listings in 2019, we did a cluster analysis of event types, to **group events that are more likely to occur together within an episode**. The following plot shows the resulting cluster structure of these event types. The figure shows that there are certain event types which are very commonly reported together such as high wind and strong wind or heat and excessive heat. It is important to keep these trends in mind as there may be overlap or uncertainty in how to categorize what happened during an episode.

```
## `summarise()` regrouping output by 'EPISODE_ID' (override with `.groups` argument)
```



- it appears that NOAA does try to account for this: “In the event it is obvious that a continuous or nearly continuous swath of thunderstorm wind or hail damage occurred, a single event should be entered into Storm Data. This single event would be described as occurring from Point A to Point B, during Time C to Time D. The related event narrative could describe the width and length of the damage swath. Scientifically, a swath is more accurate and reduces the chance of a researcher interpreting a single event as a series of events occurring across multiple points.”

### 3. difficulty of sorting events

- this could include the ideas of severity and source [BA: For the source part, we could look at how different types of events vary in their probability of being reported by different sources (I think we'd talked about that idea for another section, too). There's some code in the Details vignette we can start from for that.]
- there is also limited information on event severity or distinction between events within the database (Luh et al. 2015). The NOAA Storm Data Disclaimer states that "Some information appearing in Storm Data may be provided by or gathered from sources outside the National Weather Service (NWS), such as the media, law enforcement and/or other government agencies, private companies, individuals, etc. ("Storm Data Faq Page," n.d.)
- Definitional differences (Gall, Borden, and Cutter 2009) [BA: Could you add some more details or explanation on this point? I'm not sure I get it yet.]
  - this seems to be more of problem when comparing events across databases. The Gall Losses article gives the example of hurricane-induced storm surges. In SHELDUS, a storm surge gets categorized under "coastal hazards" and Storm Events categorizes them under "ocean and lake surf". So if someone was trying to detail a large events like a massive hurricane using multiple databases this could create a problem. [BA: Maybe this also comes into play for the different source? Would the media use the same criteria for identifying a disaster as a National Weather Service office? If not, and if the media tend to report certain types of disasters more often than others, than this could maybe create some bias.]
- NOAA Storm Events defines the types of events that are allowed in the data set. This can be found under Table 1 of Section 2.1.1 of NWS Directive 10-1605 at <http://www.nws.noaa.gov/directives/sym/pd01016005curr.pdf>.

## Event Type:

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

Event type	Number of events in 2019
Thunderstorm Wind	18,617
Hail	9,013
Flood	4,949
Flash Flood	4,072
Winter Weather	3,808
High Wind	3,784
Winter Storm	3,336
Heavy Snow	2,872
Marine Thunderstorm Wind	2,502
Tornado	1,728
Strong Wind	1,591
Heavy Rain	1,441
Heat	1,294
Extreme Cold/Wind Chill	1,065
Drought	1,007
Blizzard	852
Excessive Heat	852
Frost/Freeze	654
Dense Fog	652
Cold/Wind Chill	470
High Surf	468
Funnel Cloud	348
Lightning	343

Event type	Number of events in 2019
Coastal Flood	240
Ice Storm	200
Debris Flow	185
Waterspout	183
Wildfire	180
Lakeshore Flood	146
Tropical Storm	143
Lake-Effect Snow	78
Marine High Wind	77
Rip Current	74
Avalanche	44
Dust Storm	40
Marine Tropical Storm	37
Marine Hail	32
Tropical Depression	26
Storm Surge/Tide	25
Marine Strong Wind	17
Astronomical Low Tide	16
Hurricane	10
Dust Devil	8
Freezing Fog	6
Marine Hurricane/Typhoon	6
Marine Tropical Depression	4
Seiche	4
Sneakerwave	3
Dense Smoke	2
Sleet	1
Volcanic Ashfall	1

## Temporal Bias

[BA: We might want to start with this type of bias, and then geographic bias. I think those two are very easy for people to grasp. Then, when we talk about other types of biases, we might want to talk about them just in terms of things that might cause differences within a certain period of time and location, so we can separate them from geographic and temporal issues. There are a few examples that could really fall into several categories (for example, the flood insurance question, which has a time component to it since policies changed over time), so we might want to discuss any time-related things in this section and start with it.]

[BA: Potential reasons for temporal bias: (1) We’ve gotten better at “seeing” extreme weather as it happens—radar, larger networks of monitors, etc. There may be some good examples from measuring tornadoes of this idea, where the improvement in weather monitoring has helped increase the number of reported events; (2) Some changes have increased the chance that we notice / report events when they happen. For example, population growth over time in some areas might mean there’s a higher chance that an event that happens in that area gets reported. (3) Some temporal changes are structural, in terms of the database changing it’s “rules” for what it recorded. The expansion from just tornadoes to other types of events is an example here. This is one where we can help control for any bias, since there’s clear documentation on some of those changes. However, I wonder if there’s still some kind of lead in period, where it takes them a little while once they start adding an event to succeed in recording most or all of them? We could maybe check how the rate of recording changes with time for tornadoes versus some of the “new” events in the database in the late 1990s.]

[BA: I like that we distinguish between the idea of long-term trends in the probability of an event being



recorded in the database (e.g., from one decade to the next) versus seasonal time trends (e.g., winter versus summer). We probably want to explicitly write a few sentences introducing that idea that temporal bias can operate at different time scales.]

### **Introduction to Temporal Bias:**

Within the NOAA Storm Events database, there are **comparable differences in event and loss recording over time**. Temporal bias is helpful in describing and understanding these changes within the database. There are long term patterns of temporal bias that may be caused by changes in measuring and recording events and losses. There are also seasonal patterns within the database that relate to unique attributes and temporality of certain weather events.

### **Long Term Trends:**

The long term patterns of temporal bias are **related to advancements in monitoring and detecting storm events**, changes in population size and locations, and structural changes in recording strategies and monetary loss accounting.

Technological advancements within storm recording such as the development of radar and larger networks of monitors have increased our ability to detect extreme weather events as they happen. For example, over time, forecasters and storm spotters have learned how to recognize key weather patterns and structures that make it more likely for a tornado to form. The development and advancement of doppler radar and dual-polarization radar technology has also increased our ability to collect and use data to detect tornado events (“SEVERE Weather 101: Tornado Detection,” n.d.). One study states that “the number of tornadoes reported in the United States has doubled from about 600 per year in the 1950s to around 1200 in the 2000s”(Verbout et al. 2006). This may be influenced by factors other than meteorological changes.

[BA: Great! This is a great explanation of this idea that we may be able to “see” more events as we go along. There may be some papers in the climate change field on this topic, too. It turns out that this change in technology makes it hard to determine if there has been clear evidence of a trend in frequency of some types of extreme events. This comes up a lot with hurricane research, because we missed a lot of storms that didn’t make landfall early in the twentieth century and we don’t miss them now. I can see if I can find a couple of papers along those lines that we could add in here.]

Increases in population size throughout the US and further land development may also contribute to increased event reporting. For example, population growth over time in some areas might mean there’s a higher chance that an event that happens in that area gets reported. This is because events are reported to the NWS by a number of sources including trained spotters, the public, law enforcement, broadcast media, social media, local and county officials, etc.

### **How events are reported**

Here is a link to the National Weather Service’s information on how to become a storm spotter: Storm Spotter

Acronyms: ASOS= “Automated Surface Observing System (ASOS) units are automated sensor suites that are designed to serve meteorological and aviation observing needs”

Mesonet= “portmanteau of mesoscale network, a network of (typically) automated weather and environmental monitoring stations designed to observe mesoscale meteorological phenomena”

COOP Observer= The National Weather Service (NWS) Cooperative Observer Program (Coop) is truly the Nation’s weather and climate observing network of, by and for the people. More than 8,700 volunteers take observations on farms, in urban and suburban areas, National Parks, seashores, and mountaintops. The data are truly representative of where people live, work and play.

AWOS= Automated Weather Observing System (AWOS) is a fully configurable airport weather system that provides continuous, real time information and reports on airport weather conditions. AWOS stations are mostly operated, maintained and controlled by aviation service providers.

CoCoRaHS= The Community Collaborative Rain, Hail and Snow Network, or CoCoRaHS, is a network of volunteer weather observers in the United States, Canada, and the Bahamas that take daily readings of precipitation and report them to a central data store over the internet.

SNOTEL= snow telemetry

RAWS= The Remote Automatic Weather Stations system is a network of automated weather stations run by the U.S. Forest Service and Bureau of Land Management and monitored by the National Interagency Fire Center, mainly to observe potential wildfire conditions.

CMAN Station= The Coastal-Marine Automated Network (C-MAN) is a meteorological observation network along the coastal United States.

AWSS= automated weather sensor system

WLON= a radio station

SHAVE project= Severe Hazards Analysis & Verification Experiment

I did not find much information about these sources in the NWS NOAA documentation. However, from my research it appears that several of these sources seem to be communities of people that are reporting events. This leads to more questions about how objective the reporting is from these groups. There may be differences in how certain people classify certain events.

## `summarise()` ungrouping output (override with `.groups` argument)

Source of event report	Number of events reported in 2019
Trained Spotter	9,740
Public	8,371
Emergency Manager	6,054
Mesonet	5,710
Law Enforcement	4,053
ASOS	3,169
COOP Observer	2,746
911 Call Center	2,615
Broadcast Media	2,357
Social Media	2,054
AWOS	2,030
Department of Highways	1,915
River/Stream Gage	1,760
State Official	1,748
NWS Storm Survey	1,745
Official NWS Observations	1,424
CoCoRaHS	1,324
Amateur Radio	1,176
Fire Department/Rescue	1,169
Drought Monitor	958
NWS Employee	900
Buoy	590
Other Federal Agency	479
SNOTEL	470
Storm Chaser	446
RAWS	402
County Official	398

Source of event report	Number of events reported in 2019
C-MAN Station	396
Utility Company	393
Newspaper	274
Park/Forest Service	244
Unknown	194
Local Official	71
AWSS	29
Mariner	22
Lifeguard	21
Post Office	20
Tribal Official	14
Coast Guard	9
Airplane Pilot	7
Insurance Company	5
WLON	3
SHAVE Project	1

### More Points for Long Term Trends

- include information from the US census
- Here is link to the population data: [US census data](#)
- Here is a link to the locations of National Weather Service offices across the US: [NWS locations](#)

[BA: We may want to remind the readers here of the role that storm trackers / reporters play in reporting some of these events. I think that for some types of events, people can volunteer as “spotters” and get some training and then be able to report events they see. I think in one vignette, there’s some code that helps in seeing how many of the event reports come from those types of inputs. The population size question could be really important there. Also, I wonder if there’s any link between how close National Weather Service offices are and the population density? If so, maybe that is another link with population size / density?]

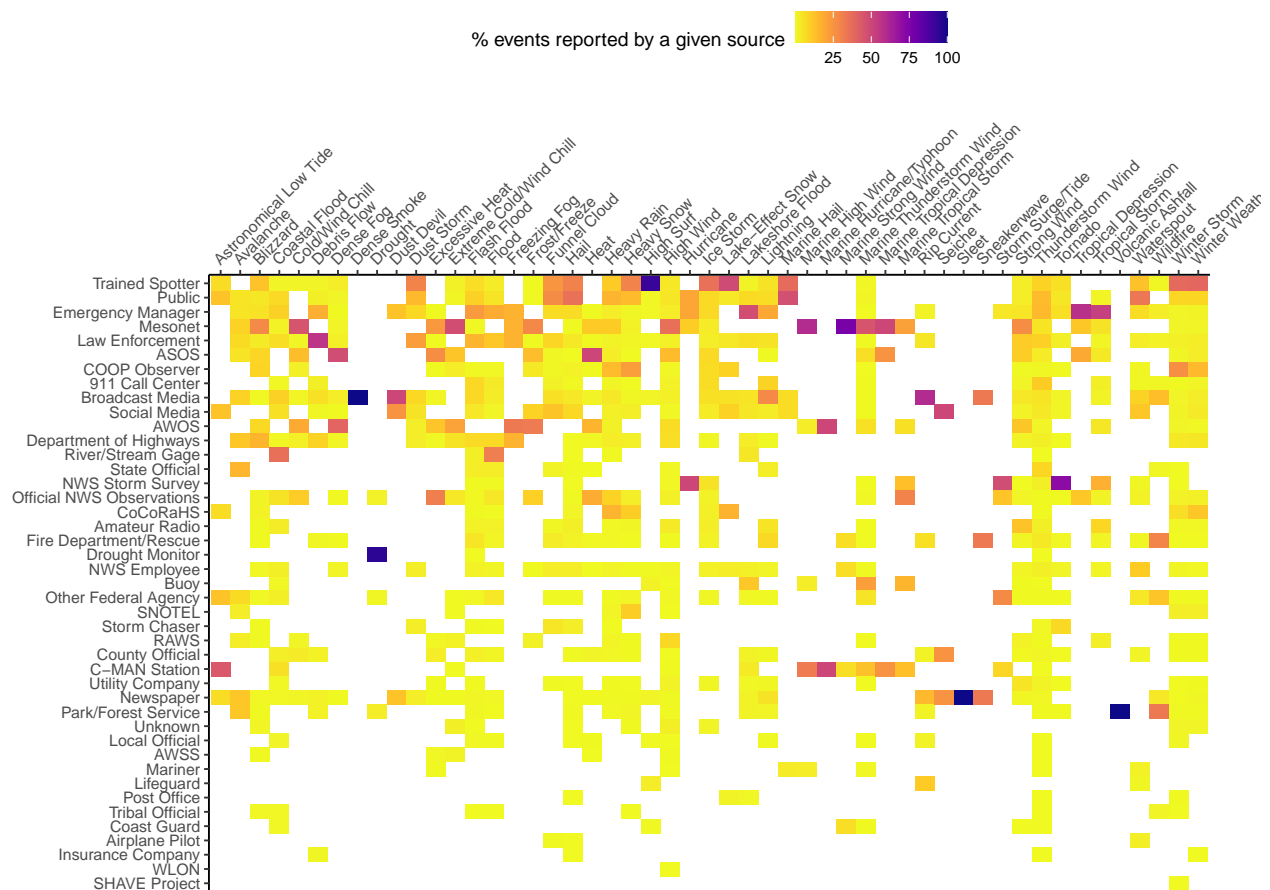
### Breakdown of percentage of events reported by a given source:

The following graph shows, for each type of event in 2019, [the percent reported by each source](#). Both axes of the plot are ordered by overall frequency (i.e., overall number of each type of event and overall number of reports from each source).

For some types of events, [reporting is dominated by a specific source](#). For example, most high surf reports come from trained spotters, while most drought reports come from drought monitors and most tornado reports come from the NWS Storm Survey. Another example is that rip currents are mostly reported by broadcast media and newspapers. This may be because rip currents can pose a serious danger to humans and cause death in certain instances. These instances are more likely to make the news than other types of general events.

For other types of events, reporting sources are more diversified.

Additionally, [certain sources are more influential in reporting certain events](#). For example, the public often reports hail and marine hail events. This may be because the public is usually highly affected by this type of weather event.



Some temporal changes are structural, in terms of the database changing what types of events are recorded. In the NOAA database, from 1950 to 1954, only tornado events were recorded (“Storm Events Database,” n.d.). Then, from 1955 to 1996, only tornado, thunderstorm wind and hail events were recorded as digital data (“Storm Events Database,” n.d.). Starting in 1996, the database began including all 48 event types that are recorded today (“Storm Events Database,” n.d.). Additionally, the NWS reports that there was reduced funding from 1976 to 1979 that led to a decrease in flood damage data collection (Downton, Miller, and Pielke Jr 2005). Then, after 1983, Congress began to require The US Army Corps of Engineers to provide yearly flood damage reports in contract with the NWS (Downton, Miller, and Pielke Jr 2005). This provides a more consistent report of flood damage that may not have otherwise been reported. These changes pose inconsistencies in the number of certain storm events over large periods of time.

[BA: We might want to add some examples, here or in one of the more general sections, about how these issues are common across a lot of disaster databases. I’ve added a few papers that talk about common problems for disaster database, rather than just NOAA Storm Events, and those might help in rounding out our discussion.]

## Seasonal Trends:

More temporal bias may be showcased on a short term scale by seasonal differences in number of events reported for certain storm types. For example, NOAA exhibits higher rates of rip currents in the summer versus the winter although the probability of a rip current occurring during summer may not be inherently higher. It may be that more people attend the beach and swim in the ocean during the summer and are thus more likely to experience and/or report a rip current. A study completed by Gensini and Ashley found that “summer season weekends are shown to have the more [rip current] fatalities than any other time of the year” (Gensini and Ashley 2010).

- Seasonal Differences in number of events reported
  - Example of rip currents in summer vs winter
- Direct quote: The strength of rip currents can be seasonal. During hurricane season (from June to November) there is a greater chance for rip currents to develop. (“Rip Current,” n.d.)
- **Rip currents** are related to several environmental factors including waves (surf heights, period, direction), beach (slope, orientation, material), water levels (tidal cycle, tide ranges), winds (affect wave breaking) and wind-driven currents alongshore, others like local coastal configuration and beach and promontories by natural or human made. (“Rip Currents: A Natural Killer on Beaches,” n.d.)
- “The most likely scenario for rip hazards is not high surf but high exposure of beachgoers in the warm water of the summer-fall period. When low-energy, longer-period waves (significant wave heights of 0.5 -1.5 meters in 10-15 second sequences) lead to the highest number of rip incidents. During spring/neap tides or very low daily tidal cycles, a mass rescue event can occur, with hundreds of rescues in several locations on a beach, or at several beaches under the same conditions.” (“Rip Currents: A Natural Killer on Beaches,” n.d.)
- “societal factors (e.g., weekends and holidays) that could change the risk of a rip current fatality. For example, low and high pressure systems off the east coast of the United States can produce onshore flow to many surf zones. Both systems may invoke a high rip current formation risk on the LURCS, but swimmers will be more inclined to enter the water on days when a high pressure system is offshore. The clear skies generally associated with large-scale subsidence of a high pressure system would provide beachgoers with favorable weather for beach activities as opposed to a day amid a low pressure system with stratus clouds and precipitation occurring” (Gensini and Ashley 2010)
- create table of number reported by lifeguards or media vs month of the year
  - other events that are more or less reported? [BA: We might want to talk about the idea of some events that truly are seasonal in their occurrence, like avalanches and hurricanes, versus some that happen year-round but have varying chance of being reported. Also, you can look through the package vignette and see if you see any other examples of things that might not be reported evenly throughout the year.]
  - coastal flood? high surf? waterspout?

[BA: I’m not sure if you saw it yet, but I added in another paper on rip currents that to the “literature” folder that might be another helpful reference for this section.]

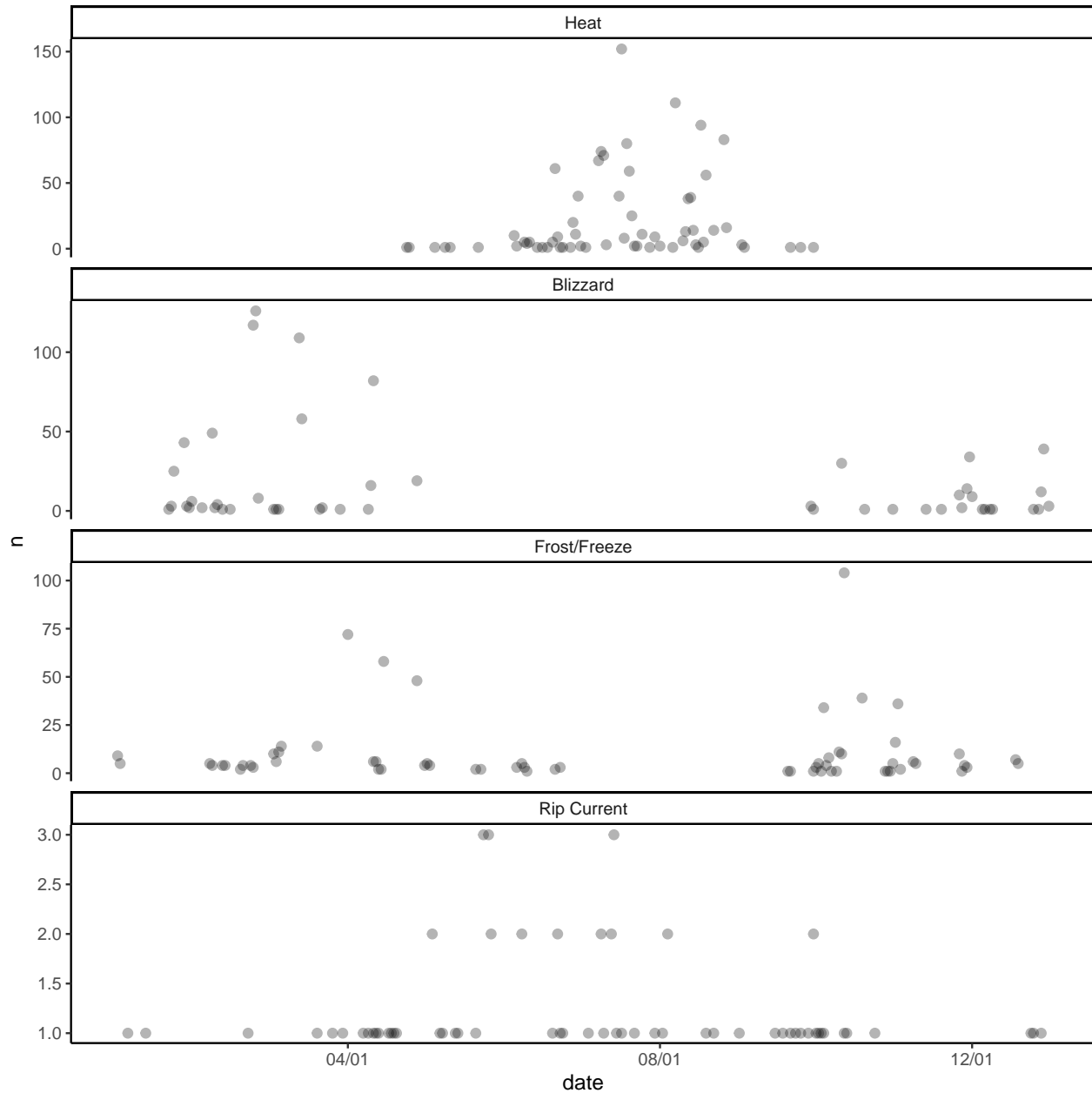
### How start dates for events are distributed over the year

[I don’t know if we’ve explained yet that events are listed with a start and end date. We might want to explain that here.] Here are how the start dates for listings for each event type are distributed over the year (event types are ordered by decreasing total count during the year; note that the y-axes vary depending on the range of events by date for each event type):

Many **event types are clearly seasonal** (e.g., winter weather, winter storms, heavy snow, cold, extreme cold, blizzards, ice storms, lake-effect snow, and avalanches are all much more common during winter months, while heat and extreme heat are typically limited to summer months). However, for some events, reporting seasonal patterns might be based not just on the true pattern of events but also on the **timing of important exposures and impacts of the events**. For example, rip currents have many more listings during the spring and summer, which may be related to events being more likely to be listed when more people are swimming. Frost event listings are particularly high at the start and end of the frost season, rather than in the middle of winter, which may be related to the impacts of frost on crops being higher in spring and fall than during the winter. If working with this data, it important to keep in mind that the data are based on reporting, and there may be related influences on the probability of an event being reported and included in the data that differ from using data from something like a weather station. [Could you check on Google Scholar and see if

we could find out anything from the literature to support our idea here that you're going to get frost through the winter, but that the most economically damaging frost comes near the edges of the frost season.]

```
## `summarise()` regrouping output by 'date' (override with `.groups` argument)
```



## Threshold Bias

The varying severity of an event creates **threshold bias** in the storm events database. Events of larger magnitude and/or damage to human health are better documented, while events of smaller magnitude are less reported because less people are affected (Gall, Borden, and Cutter 2009). Small events may even be excluded due to threshold criteria in the database (Gall, Borden, and Cutter 2009). For example, if a flood has received a presidential disaster declaration, FEMA storm survey teams go to the scene and provide extra damage assessments (Downton, Miller, and Pielke Jr 2005).

- Events of smaller magnitude are less reported because less people are affected (Gall, Borden, and Cutter 2009)
  - may even be excluded due to threshold criteria (Gall, Borden, and Cutter 2009)
  - verify this under NOAA disclaimer

## Accounting Bias

### Introduction/How damages are reported

There are **discrepancies** in how monetary losses and damage information are collected in the NWS storm events database that lead to **accounting bias**.

This is a quote from the NWS documentation on how estimates are recorded. Not sure if we should include it here or above when we describe the database:

“Estimates should be in the form of US Dollar values and rounded to three significant digits, followed by the magnitude of the value (i.e., 1.55 Billion \$USD for \$1,550,000,000). Values used to signify magnitude include: Thousand \$USD, Million \$USD, and Billion \$USD. If additional precision is available, it may be provided in the narrative part of the entry. When damage is due to more than one element of the storm, indicate, when possible, the amount of damage caused by each element. If the dollar amount of damage is unknown, or not available, check the “no information available” box.”

[BA: Overall for this section, I know that there seemed to be some odd things occasionally cropping up in the data from the `noaastormevents` package when you look at damages. You might want to look through the “Details” vignette for anything that specifically talks about damages. I think there were occasionally cases, like Hurricane Floyd in 1999 in North Carolina, where it looked like a lot of the damage was maybe reported for the capital county (the county with Raleigh in it) rather than coastal counties (coastal counties were reported, too, but the numbers were lower, I think). That raised some questions for me about whether sometimes general damages were reported for the state, and just recorded in the capital location, versus in the county where the damage occurred. Also, it may be interesting to see, both from the data and the documentation, (1) whether they split up the total costs across different events in an episode (for example, if there are reports of multiple events during a hurricane for a county, where do they record the damage?) and (2) how different costs are from county to county for the same episode (and particularly for counties that are close to each other but across state lines—do the damages tend to be more similar within a state than between two states?).]

### Inaccuracies in reporting

The National Weather Service Documentation states that “property damage estimates should be entered as actual dollar amounts, if a reasonably accurate estimate from an insurance company or other qualified individual is available” (NWS, n.d.). It also states that “if this estimate is not available, then the preparer has two choices: either check the “no information available” box, or make an estimate” (NWS, n.d.).

The only exception to this rule is that the NWS is required to provide monetary damage amounts for flood events by the U.S. Army Corps of Engineers (NWS, n.d.). Otherwise, the Storm Data preparer is only encouraged to report monetary damage estimates. These estimates can be obtained from emergency managers, U.S. Geological Survey, U.S. Army Corps of Engineers, utility companies, and newspaper articles.

The estimates of damage to insured property are typically obtained from local field office reports that often lack quality and accuracy control (Downton, Miller, and Pielke Jr 2005).

[BA: It would be interesting to see if damages are more likely to be reported based on (1) who reports the event and (2) what type of event it is. We could write some code to check this in our data.]

## Difficulty of quantifying damages

The estimates are comprised of **direct and indirect monetary losses**. Direct monetary losses from damage to infrastructure, buildings, crops, etc. are easier to quantify than indirect losses like lost revenue, business closures, societal losses, environmental damage (Gall, Borden, and Cutter 2009).

“Insurance company records include only insured property, and records of the Federal Emergency Management Agency (FEMA) include only property that qualifies for federal assistance in presidentially declared disasters. Few state and local governments maintain damage records beyond those required by FEMA; only in newspaper archives from cities and towns across the nation might one find more complete historical reporting of local flood damage.” (Downton, Miller, and Pielke Jr 2005)

Monetary losses can also be categorized into community, state, regional, and global levels which can further complicate the precision of the recording (Gall, Borden, and Cutter 2009).

## Structural changes in how NOAA reports damages

Until 1994, damage estimates were recorded on a logarithmic scale (Downton, Miller, and Pielke Jr 2005). NOAA now records damage estimates in thousands of dollars. [BA: Could you clarify this a bit? I’m not sure what is meant here by the use of a logarithmic scale. Is this in terms of how they rounded the estimates?]

- pull up loss data for events pre and post 1995?

“From 1976 through 1979, NWS reports indicate that reduction of funding led to cutbacks in the compilation of flood damage data. Data collection continued as in prior years, but there appears to have been less checking and updating of initial damage information. Further, publication of annual summaries ceased. In 1980, compilation of flood damage estimates was discontinued entirely.” (Downton, Miller, and Pielke Jr 2005)

“NWS policies on what losses to include have changed somewhat over the years. Damage estimates published through 1975 focused primarily on damage to property and crops, but included some indirect losses. Present policy is to focus exclusively on physical damage to property and crops. Until 1992, separate estimates were given for property and agricultural damage, but in 1993 that distinction was eliminated.” (Downton, Miller, and Pielke Jr 2005)

- provide example/table of how NOAA records financial losses

## Change in value of the US dollar over time

The **value of a dollar has changed over time**. This can be addressed by normalizing any dollar measurements to a certain year to prevent bias. [It might be helpful to add some examples of studies that do this. I think there are some from Roger Pielke that look at hurricane damages over time in the US that normalize the damages. We could (briefly) include an explanation of how they do that.]

## Geographic Bias

### Introduction

Another area of concern in the NOAA storm events database arises with **geographic bias** which explains the idea that there may be **inconsistencies in the recording of events due to geographic location**. First, there may be differences in the supply of information from different regions based on population differences or locations of weather stations. Secondly, structural changes in the database create differences in how location data is obtained and recorded.



## Location / Density of people

The reporting of events can be affected by the **geographic location of the event itself** (Luh et al. 2015). In an urban area, more people may be present to witness/experience a storm event than in a rural area.

- probably a tornado example we can include here [Somewhere we think there's a paper on that]

The reporting of events can be affected by geographic location based on **whether or not people are present to record the event**. The supply of information is greater in areas closer to the weather event (Konisky, Hughes, and Kaylor 2016). [BA: You could go more into the rural / urban question here. Would the luh2015vulnerability reference be a good example to include for this idea? The tornado example could also go here probably.]

- this idea is stated in the context of a study about correlations between weather events and public attitudes about climate change. This idea might be helpful in describing geographic bias as more concern is shown towards extreme weather in areas that are more strongly affected?

[More things get reported where there are more people?]

[Any papers on changes in numbers of media reports of disasters?]

## Structural

Geographic bias also results from **structural changes at the county or state level**. In 1995, the NWS changed its reporting strategy from loss estimates by climate region to loss estimates in specific counties where the event occurred (Gall, Borden, and Cutter 2009). This type of change could lead to excluding loss data or double counting events (Gall, Borden, and Cutter 2009). The NOAA website even provides a disclaimer that “the source data ingested into the database are widely varied and leads to many questions about the precision and accuracy of the location data” (“Storm Data Faq Page,” n.d.).

- script for checking event location pre and post 1995?

[BA: If this is a change that would happen over time, do we want to consider moving it to temporal? Or might it happen at different times in different places? For this, maybe we want to include a map of climate regions? I think a lot of readers might not know how large or small those are off the top of their heads. Also, I really like the idea of comparing event reports before and after this change. Can we see that, before the change, they followed the boundaries of the climate region?]

## Aggregation of reporting

When a location is entered for an event in the NOAA Storm Events database, **inconsistencies may arise between where the event actually occurred and the location that is recorded in the database**. Currently, the smallest unit of aggregation to use all parts of the database are called Weather Forecasting Offices. Each office covers a geographic area of the US and there are currently about 122 nationwide (Konisky, Hughes, and Kaylor 2016). These geographic areas overlap with county outlines but do not always directly coincide. It is unclear exactly how NOAA decides when to assign an event to a county or a Weather Forecasting Office. In regards to this decision, the NWS documentation states that “a hydrometeorological event will be referenced, minimally, to the nearest hundredth of a mile, to the geographical center (not from the village/city boundaries or limits) of a particular village/city, airport, inland lake, or location providing that the reference point is documented in the Storm Data software location database” (NWS, n.d.).

[BA: In this section, we may also want to talk about how some events might be very localized (for example, a flash flood, or even storm surge that's only near the coast of the county), but the county-level reporting of the event means that all the county gets classified as experiencing or not experiencing the event. I think that sometimes the “Narratives” of the events give details that you could use to find out exactly what parts of the county were affected, and some events like tornadoes give the starting and ending latitude and longitude, so

you can figure out the path and where that was in the county. However, overall, it's much easier from the database to get an idea of an event occurrence for the county as a whole rather than for locations within the county.] [We may want to talk a bit about the idea of ecological bias here.] [Maybe want to talk about spatial alignment.]

- include zone and fips script here?

## Forecast Zone versus County

The NOAA Storm Events database also uses **forecast zones as another unit of aggregation**. Forecast zones are created as subsets of counties for more specific location data. Specific types of events are typically either always reported for a county (CZTYPE of "C") or always reported for a forecast zone (CZTYPE of "Z") (see table below)\*\*. Events typically reported by county include floods, tornado-like events, and a few other events often related to thunderstorms. Events typically reported by forecast zone include severe winter weather, extreme heat, events related to the water or coast, and a few others ("High Wind", "Dense Fog", "Strong Wind", "Wildfire", "Dust Storm", "Dense Smoke").

This demonstrates the **difficulty of narrowing down where exactly certain events occurred**. It is important to keep in mind that some events might be very localized (for example, a flash flood, or even storm surge that's only near the coast of the county), but the county-level reporting of the event means that the whole county gets classified as experiencing or not experiencing the event.

```
## `summarise()` regrouping output by 'CZ_TYPE' (override with `.groups` argument)
```

Event type	County	Forecast Zone	Total	% county
Thunderstorm Wind	18,617	0	18,617	100%
Hail	9,013	0	9,013	100%
Flood	4,949	0	4,949	100%
Flash Flood	4,072	0	4,072	100%
Tornado	1,728	0	1,728	100%
Heavy Rain	1,441	0	1,441	100%
Funnel Cloud	348	0	348	100%
Lightning	343	0	343	100%
Debris Flow	185	0	185	100%
Dust Devil	8	0	8	100%
Heat	1	1,293	1,294	0%
High Wind	1	3,783	3,784	0%
Winter Weather	0	3,808	3,808	0%
Winter Storm	0	3,336	3,336	0%
Heavy Snow	0	2,872	2,872	0%
Marine Thunderstorm Wind	0	2,502	2,502	0%
Strong Wind	0	1,591	1,591	0%
Extreme Cold/Wind Chill	0	1,065	1,065	0%
Drought	0	1,007	1,007	0%
Blizzard	0	852	852	0%
Excessive Heat	0	852	852	0%
Frost/Freeze	0	654	654	0%
Dense Fog	0	652	652	0%
Cold/Wind Chill	0	470	470	0%
High Surf	0	468	468	0%
Coastal Flood	0	240	240	0%
Ice Storm	0	200	200	0%
Waterspout	0	183	183	0%
Wildfire	0	180	180	0%
Lakeshore Flood	0	146	146	0%

Event type	County	Forecast Zone	Total	% county
Tropical Storm	0	143	143	0%
Lake-Effect Snow	0	78	78	0%
Marine High Wind	0	77	77	0%
Rip Current	0	74	74	0%
Avalanche	0	44	44	0%
Dust Storm	0	40	40	0%
Marine Tropical Storm	0	37	37	0%
Marine Hail	0	32	32	0%
Tropical Depression	0	26	26	0%
Storm Surge/Tide	0	25	25	0%
Marine Strong Wind	0	17	17	0%
Astronomical Low Tide	0	16	16	0%
Hurricane	0	10	10	0%
Freezing Fog	0	6	6	0%
Marine Hurricane/Typhoon	0	6	6	0%
Marine Tropical Depression	0	4	4	0%
Seiche	0	4	4	0%
Sneakerwave	0	3	3	0%
Dense Smoke	0	2	2	0%
Sleet	0	1	1	0%
Volcanic Ashfall	0	1	1	0%

## Systemic Bias

- I haven't been able to find a lot of support or evidence for this type of bias. Maybe this could be a good place to describe how these problems span many databases.
- Differences in initial data collection and compilation create difficulties in comparing databases
  - Source and how losses are computed
  - Actual dollar losses vs inflation adjusted losses
  - Whole dollars vs loss categories

## References

- Downton, Mary W, J Zoe Barnard Miller, and Roger A Pielke Jr. 2005. "Reanalysis of Us National Weather Service Flood Loss Database." *Natural Hazards Review* 6 (1). American Society of Civil Engineers: 13–22.
- Gall, Melanie, Kevin A Borden, and Susan L Cutter. 2009. "When Do Losses Count? Six Fallacies of Natural Hazards Loss Data." *Bulletin of the American Meteorological Society* 90 (6). American Meteorological Society: 799–810.
- Gensini, Victor A, and Walker S Ashley. 2010. "An Examination of Rip Current Fatalities in the United States." *Natural Hazards* 54 (1). Springer: 159–75.
- Konisky, David M, Llewelyn Hughes, and Charles H Kaylor. 2016. "Extreme Weather Events and Climate Change Concern." *Climatic Change* 134 (4). Springer: 533–47.
- Luh, Jeanne, Elizabeth C Christenson, Aizhan Toregozhina, David A Holcomb, Tucker Witsil, Laura R Hamrick, Edema Ojomo, and Jamie Bartram. 2015. "Vulnerability Assessment for Loss of Access to Drinking

Water Due to Extreme Weather Events.” *Climatic Change* 133 (4). Springer: 665–79.

NWS, NOAA. n.d. “NATIONAL Weather Service Instruction.”

“Rip Current.” n.d. <https://www.nationalgeographic.org/encyclopedia/rip-current/>.

“Rip Currents: A Natural Killer on Beaches.” n.d. <https://www.weather.gov/mdl/RipCurrent>.

“SEVERE Weather 101: Tornado Detection.” n.d. <https://www.nssl.noaa.gov/education/svrwx101/tornadoes/detection/>.

“Storm Data Faq Page.” n.d. <https://www.ncdc.noaa.gov/stormevents/faq.jsp>.

“Storm Events Database.” n.d. <https://www.ncdc.noaa.gov/stormevents/details.jsp>.

Verbout, Stephanie M, Harold E Brooks, Lance M Leslie, and David M Schultz. 2006. “Evolution of the Us Tornado Database: 1954–2003.” *Weather and Forecasting* 21 (1): 86–93.