NOAA Storm Events in Literature

```
library(noaastormevents)
library(tidyverse)
```

What is the Storm Events Database

[Add a brief description of this database, including: who collects it; what data is collected; how it's collected; where the database is available; any major changes over time to what data have been collected and how they've been collected; information about what's included in each "entry" in the database.]

From NOAA Storm Events Website: "The Storm Events Database contains the records used to create the official NOAA Storm Data publication, documenting:

The occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce; Rare, unusual, weather phenomena that generate media attention, such as snow flurries in South Florida or the San Diego coastal area; and Other significant meteorological events, such as record maximum or minimum temperatures or precipitation that occur in connection with another event.

The database currently contains data from January 1950 to February 2020, as entered by NOAA's National Weather Service (NWS). Due to changes in the data collection and processing procedures over time, there are unique periods of record available depending on the event type. NCEI has performed data reformatting and standardization of event types but has not changed any data values for locations, fatalities, injuries, damage, narratives and any other event specific information."

[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.]

- how data is collected and aggregated for the databbase: "An effort is made to use the best available information but because of time and resource constraints, information from these sources may be unverified by the NWS" ("Storm Data Faq Page," n.d.).

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.]

[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 beause of how they get information

    -"NWS is obliged to provide moentary 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)
    - 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?]

  - 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 seculded 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.
  - "fatal weather events are more likely to be reported due to enhanced media attention" (Gensini and Ashley 2010)

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 correlates to mulitple events (Konisky, Hughes, and Kaylor 2016)
     - script for how noaa deals with this? or examples of this? Probably explained in narrative sections of events
   - 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
   - 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.
   - we could list event_type from database? [BA: I think this is a great idea, either here or earlier when we talk about the database. I think there may already be some code in one of the vignettes for doing that. Take a look and see if you see that—if not, I can help find it if it's there or give some suggestions for writing some code for this if we don't already have an example.]
   - 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.

- maybe something like this where we group by event_type (this is from overview vignette sonwe would have to modify to make it work here):

```r
library(dplyr)
floyd_events %>%
  group_by(event_type) %>%
  summarize(n = n()) %>%
  arrange(desc(n)) %>%
  knitr::kable(col.names = c("Event type", "Number of events"),
               caption = "NOAA Storm Events within 200 km and within a 5-day window of Hurricane Floyd.
```

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, forecastors 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 numbber of sources including trained spotters, the public, law enforcement, broadcast media, social media, local and county officials, etc.

- include graph from details vignette listing who reported events [Dr. Anderson do you have the code for this graph?]
- 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?]

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 inconsistensies 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 probablity of a rip current occuring 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 envrionmental 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?

[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.]

Threshold Bias

The varying severity of an event creates threshold bias in the storm events database. Events of larger maginutude and/or damage to human health are better documented, while events of smaller magnitude are less reported becasue 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, ff 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 becasue 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

There are discrepencies in the type of loss information collected in the storm events database leading to accounting bias. 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. These estimates are comprised

of direct and indirect monetary losses. Direct monetary losses from damage to infrastructure, buildings, crops, etc. are easier to quanitfy than indirect losses like lost revenue, business closures, societal losses, environmental damage (Gall, Borden, and Cutter 2009). These monetary losses can also be separated into community, state, refional, and global levels (Gall, Borden, and Cutter 2009). 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).

- how does noaa report different types of losses?
- what level does noaa report economic losses at?
- How can nonmonetary losses be quanitifed?
- provide example/tabble of how NOAA records financial losses
  - can this be found in NOAA disclaimer?

[BA: We probably want to clarify somewhere here that there's also, for damages / money, the fact that 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. That's a separate idea, though, from this idea that we might be *better* at counting all the damage from a storm with time.]

- New process for loss estimation developed by NWS in 2007
- Switch between categorical estimates to whole dollar figure estimates: uses logarithmic at some point?
  - pull up loss data for events pre and post 1995
  - check for changes in reporting for NOAA under Database Details

Geographic Bias

Introduction

Another area of concern in the NOAA storm events database arises with geographic bias. This bias 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.]

Additionally, the supply of information is greater in areas closer to the weather event (Konisky, Hughes, and Kaylor 2016).

- 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

There can also be changes at country or state level over time that lead to excluding or double counting events or loss data (Gall, Borden, and Cutter 2009). The NWS changed its reporting strategy from loss estimates by climate region to loss estimates in specific counties where event occurred around 1995 (Gall, Borden, and Cutter 2009). The NOAA website 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.

Currently, the smallest unit of aggregation to use all parts of database are Weather Forecasting Offices and there are about 122 nationwide (Konisky, Hughes, and Kaylor 2016). There may be inconsistencies between county regions and weather forecasting offices. NOAA Storm Events 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."

Currently, the smallest unit of aggregation to use all parts of the database are Weather Forecasting Offices. There are currently about 122 nationwide (Konisky, Hughes, and Kaylor 2016). There may be inconsistencies between where county regions are defined and where weather forecasting offices are located. NOAA Storm Events 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.). Problems may arise here if the exact location of an event is not already documented in the Storm Data software location database. [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 occurrance 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.]

- the above quote is from the NWS documentation pdf that I used a misc. bibtex to cite
- include zone and fips script here?

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): 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): 799–810.

Gensini, Victor A, and Walker S Ashley. 2010. "An Examination of Rip Current Fatalities in the United States." *Natural Hazards* 54 (1): 159–75.

Konisky, David M, Llewelyn Hughes, and Charles H Kaylor. 2016. "Extreme Weather Events and Climate Change Concern." *Climatic Change* 134 (4): 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): 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.