**Agenda**

1. Data and Purpose
2. Analysis and Observations
3. Final Takeaways and Future Analysis

**Data and Purpose**

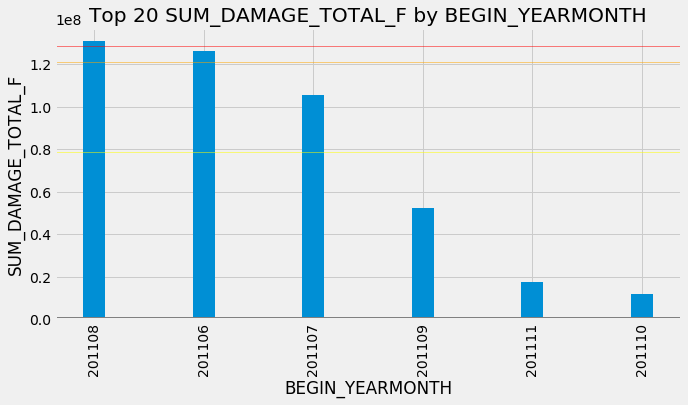
For this assignment, I was given a spreadsheet of over 30,000 records of storm data from the National Oceanic and Atmospheric Administration (NOAA). No direction was given other than to perform exploratory data analysis and uncover trends. All analysis was performed using python and Jupyter Notebooks. I focused on the cost of damages from these events, but the code can quickly be applied for similar analysis of deaths and injuries. I also consolidated the property and crops damages into one field for the analysis.

Due to time constraints, I also decided to focus more on the analysis than the presentation. As a result, elements of the chart like titles and legends are not always fully formed, aesthetically pleasing, and/or easy to read. I will try to make up for this with thoughtful writeups with each graph.

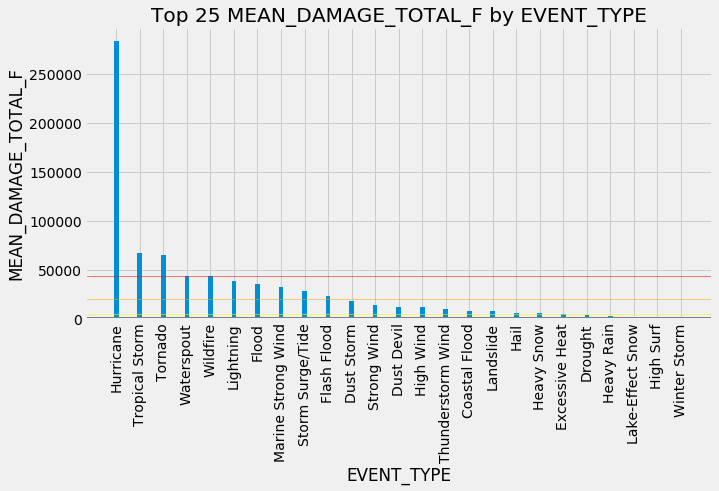
**Analysis and Observations**

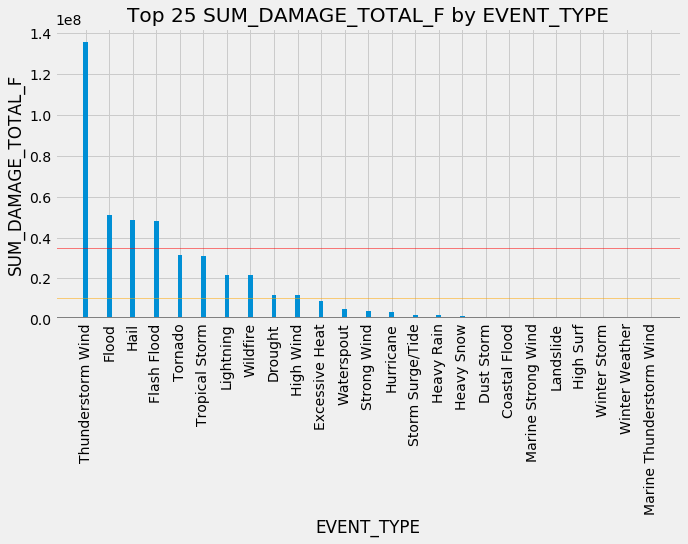
The first thing that stood out during my analysis was episode id 57541. It was classified as a “Watersprout”, but contained only one report of damages for $5,000,000. For context, the 75th percentile for total damages $1,000. For this reason, this event was excluded from some of the charts in the full analysis.

Now, the report will focus on visualization some trends with graphs. All bar graphs in this report contain yellow, orange, and red horizontal lines. The yellow, orange, and red lines are the 50th, 75th, and 90th percentiles respectively for the measurement on the y axis.



The preceding graph examines total damage by the month the event began. If this is a complete dataset for this timeframe, the graph suggests there may be a seasonal pattern that can be used to predict damages. Intuitively, this conclusion makes sense as certain weather events are more common depending on the season and region.



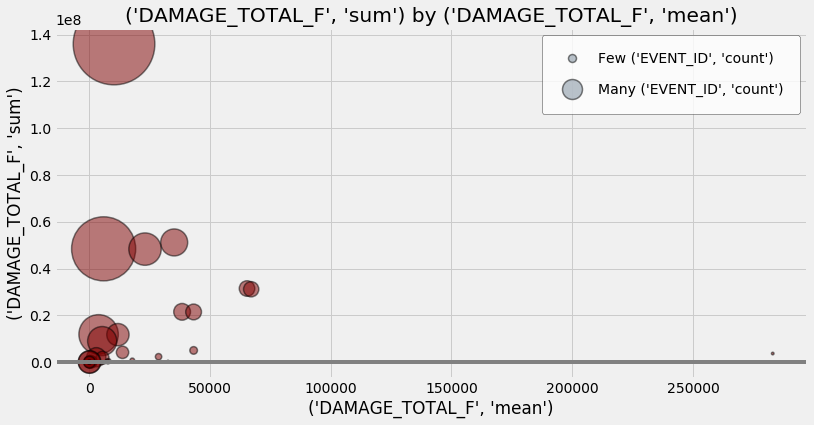


The next set of graphs examines how damage costs change with event types.

As you can see in the first graph, hurricanes are by far the most-costly event per event. The combination force, duration, and size of hurricanes makes them devastating and costly. However, hurricanes don’t occur frequently, so their total impact is not as substantial as other event types. This point is illustrated in the second graph.

Focusing on the second graph it is evident that thunderstorm wind is the leading cause for total damage costs. Unlike hurricanes, their average damages are low, but they occur frequently. The total cost of damages from thunderstorms is about the same as the next three most costly event classes combined (floods, hail, and flash floods).

The impact of tropical storms falls between the hurricanes and thunderstorm winds. Tropical storms are the same atmospheric phenomenon as hurricanes but with lower windspeeds. As a result, they occur more frequently and cause more total damage. Tropical storms and tornados are the only events to land in the top 6 of both total damage and average damage per event.



The final graph displays the relationship between the mean and total damages of each event type. Each circle represents an event type. The mean damage is located along the x axis, and the total damage is located along the y axis. The size of the bubble is correlated to the total number of events for each storm type.

In the bottom right corner of the graph, you can see the small “hurricane” point. The graph visualizes the extreme effects of hurricane, while also conveying the storm types relatively low overall impact.

The large bubble in the upper left represents the “thunderstorm wind” type. The size of the bubble illustrates the volume of thunderstorm wind events. There are 13,323 recorded thunderstorm events which is over 5,000 more than the next most recorded event. As a result, total damages from thunderstorm wind are more than double the next highest storm class.

Next, let’s move our attention to the two overlapping bubbles on the top right of the cluster around $60,000 mean damage. These two bubbles represent tropical storms and tornados. Both of these event types are devastating storms that occur frequently enough to have a relatively high impact on total damage as well.

Finally, the three bubbles grouped near the $50,000,000 mark on the y axis are flood, hail, and flash floods. These types of events have a similar overall effect as tropical storms and tornados. However, their total damage is amassed with more frequent events.

This section is meant to highlight a few of the most interesting aspects of the analysis. The full collection of graphs and stats can be found at [this link](https://github.com/markgroner/charles_schwab/blob/master/groner_analysis.ipynb).

**Final Takeaways and Future Analysis**

After performing the analysis, I am left with more questions than answers. It is important to keep in mind “what problem am I trying to solve?”. The above section highlights some basic takeaways, but the analysis could go much deeper. Depending on the context of the issue, I would like to dig deeper into these questions or areas if given more time:

* Does the legitimacy of the source matter? (Ex: Amateur Radio vs Trained Spotter)
* Since these events are weather driven, it is likely they are seasonal. I would like to see more data to test this hypothesis get more context around events that seem anomalous.
* Taking a further look into specific event types and seeing which events stand out among other storms of the same type.
* Is there a need to consolidate the event types? Is every flooding, tropical storm, hurricane, high wind, its own event or are these occasionally the same underlying event? Looking at location and time could help answer this question. Looking at how certain sources tend to report each event type could also provide insight.
* Creating an interactive map with JavaScript and Leaflet. The map that would allow a user to filter by event types, play a time series, and visualize impact by aggregate or mean damage costs, deaths, or injuiries.
* How do elements like the Fujita scale score and earthquake magnitude impact damages? Are these factors more important than location or does a significantly weaker event have a more devastating effective if it occurs in a more populated area?
* Finally, I believe time-based analysis could be important. How long is each event and how did it affect total damage?