**Storm Events & Health Ailments Exploratory Data Science Report**

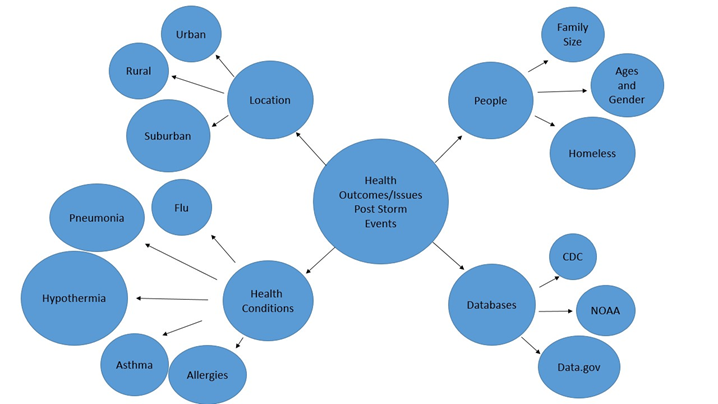
Patrick Nelson

March 22, 2020

**Introduction**

The U.S Department of Commerce’s National Oceanic and Atmospheric Administration’s National Centers for Environmental Information Storm Events Database(NOAA) contains information on a wide variety of storm events that have affected the United States over several decades. With the use of this information, a task was assigned to develop a data analytics project with the objective of addressing business needs and providing useful insights using the findings. Along with the NOAA database, there was a requirement to make use of at least one other database to obtain useful datasets. In order to complete this task, the Data Science Pipeline had to be used. This pipeline is used to describe and organize the 5 phases of the project which include Data Ingestion & Storage, Data Wrangling, Exploratory Data Analysis(EDA), Visualization, and Machine Learning.

Storm events can be extremely destructive in a variety of ways. Other than destroying homes and communities, storm events have been linked to increased occurrences of health ailments. For example, researchers have discovered that flood waters from hurricanes can transmit diseases such as typhoid, hepatitis A and E, and E. coli. (Meyers, 2019). Severe thunderstorms can trigger deadly asthma epidemics(Kuhne, 2019). In fact, cases of asthmatic occurrences post-thunderstorms have been recorded since the 1980s. Records show that there were “several outbreaks reported in the United Kingdom, Canada, Italy, Australia and the United States.” (Kuhne, 2019). Kuhne (2019) also describes incidents of increased pollen disbursement after thunderstorms. Storm events can cause widespread long-term effects on health, which can be disastrous to communities. Carbon monoxide poisoning is also a big concern as a result from damages from storm events like hurricanes. Researchers have agreed that “storms and generators were a deadly combination for carbon monoxide poisoning”(Rapaport, 2018). Hurricanes cause major power outages, which leads to the use of “portable generators, camp stoves, and charcoal camp grills” which “can all emit enough carbon monoxide to cause severe injury and death when they’re used indoors or on porches where there isn’t enough ventilation”(Rapaport, 2018) . Tornadoes and snow storms are also examples of storm events that can produce similar outcomes.

The data analytics project's focus was on how people are affected by health ailments post-storm events. Examples of health ailments include asthma, hypothermia, flu, and carbon monoxide poisoning. Project objectives and questions were outlined. The project objectives included determining which specific health ailments are related to the weather, determine if frequent storm events in specific locations are related to health ailments, and determine if there are relationships between storm events, health ailments and specific demographics such as age, family size, income, race and gender. The concept map above gives a visual of what was researched throughout the duration of the project. Highlighted in the concept map are examples of some health ailments related to storm events and some demographics to be researched like urban, rural, or suburban areas affected. 

**Data Ingestion**

The first step of the data science pipeline involves the ingestion of the data sets that are to be used for the project. The databases that were used are the U.S Department of Commerce’s National Oceanic and Atmospheric Administration’s National Centers for Environmental Information Storm Events Database(NOAA) and the Centers for Disease Control and Prevention’s(CDC) National Environmental Public Health Tracking Network. The NOAA database contains data dating back to January 1950 and up to December 2019; collected by NOAA’s National Weather Service(NWS) and some other outside sources. The database is updated every 75-90 days after the end of each data month. This data was not available online until 1999, when the web application was published. The instance of all of the datasets is storm events. Storm events are defined by the database as the occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption of commerce. The years in the database that will be focused on are from 2013 to 2018.

The CDC launched the Environmental Public Health Tracking Network in July of 2009. As of today, the network consists of over 20 data topics. The tracking network was created to bring together health and environmental data from all over the country to a single source with the goal of providing information in order to help improve the environment of citizens. Data is available for use from 2000 to 2017. The Health Tracking Network contains several data sets on a variety of topics. In this project there were 2 instances focused on from the CDC datasets; asthma and carbon monoxide(CO) poisoning. Unfortunately, there is a big fault in the Tracking Network’s database. With some of the datasets it is stated that some data may be missing due to it not being provided to the CDC or the data was not collected. Therefore, some states are not represented in some of the datasets either due to one of the reasons listed above.

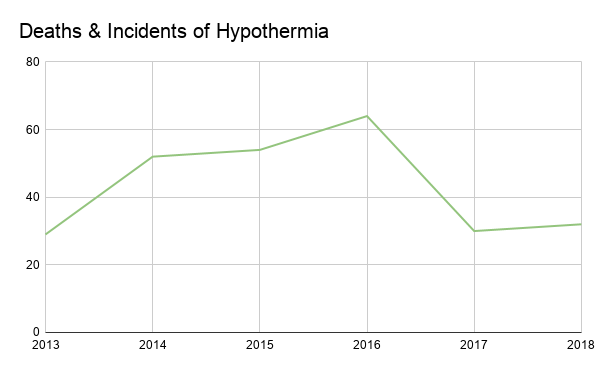
**Data Wrangling**

Each data set was broken down to discuss what features were used, what portion of the data was used, the number of instances and features, the data engineering done, and the data errors/issues encountered and how they were resolved. The features that were used from the NOAA database include State, Year, Month, Event type, Event narrative, and Episode narrative. The State, Year, and Month describe exactly as labeled. Event type refers to the type of storm event that occured; giving the name of the event. The Event Narrative and Episode Narrative provide descriptions of the storm events that occured. These are descriptions of what happened during the storm and afterwards, like the damages done as a result of the weather phenomenon. For the project, all of the United States was looked at to gather information on health ailments. Health ailments are not region specific, so it was best to gather national data to get the most accurate information on storm events’ relation to health ailments. It was also necessary to give the best data insights on the topic. The total number of instances that were used is approximately 50,000 per year, with 5 years being monitored. Out of 51 total features in the NOAA datasets, only 6 of the features were necessary for research purposes on the topic.

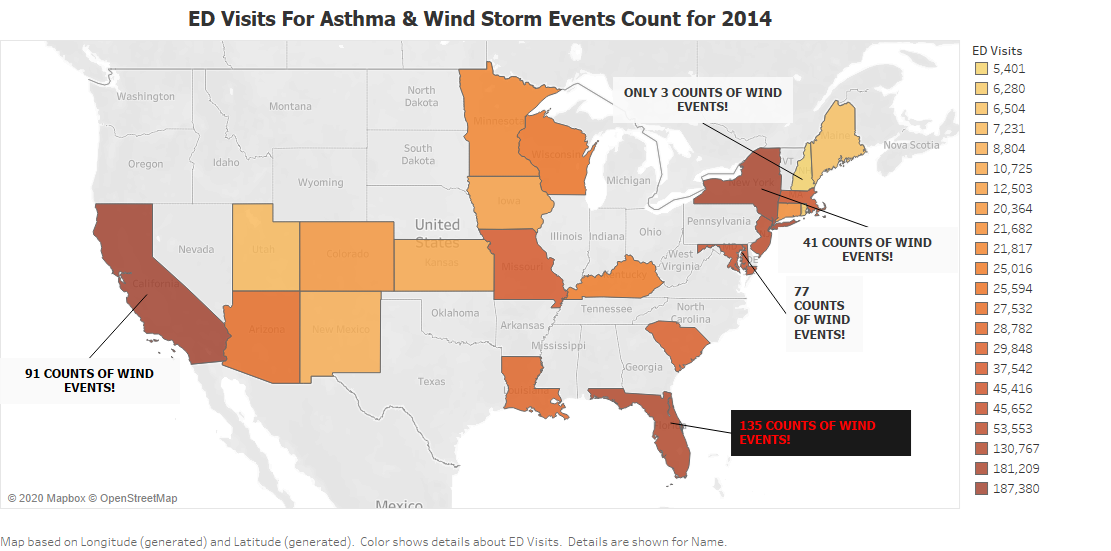
The features that were used for the CDC datasets include the State, Year, and Value features. The State and Year features contain data as labeled. The Value feature contains the count of the instance of the dataset. In this case it contains the counts for the number of emergency department visits for asthma cases and carbon monoxide poisoning cases in their respective datasets. Like with the NOAA Storm Events database, the location focus of the CDC datasets were every state within the United States. The total number of instances used for research was 230 accompanied by a total of 3 different features used for each dataset.

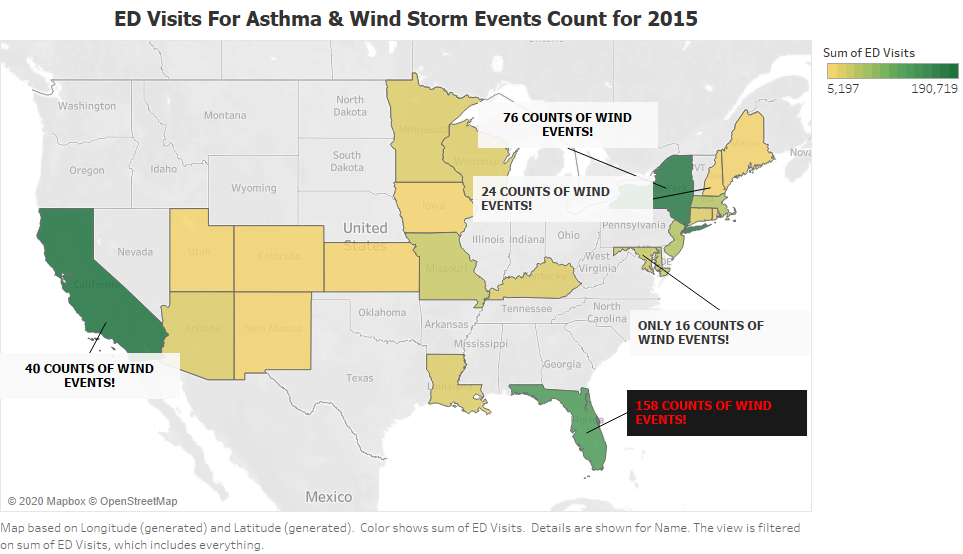
Within the datasets, there were a lot of features that were not necessary to use. The NOAA datasets had the majority of its features removed for the final analytic datasets, which left only 6 of them actually used. Within those 6 datasets, filtering was done to the Month and Event Type features in order to focus on specific months and storm events. This was done because the spring months had to be used specifically to focus on relating storm events and asthma cases during a season with high pollen counts. Other times the Event Type was filtered in order to create separate datasets containing only specific storm events related to specific health ailments. To create the separate datasets, the filtered data had to be copied and pasted into another new spreadsheet and saved. The find and select option in Excel was also useful in filtering out data on ailments in the Event Narrative and Episode Narrative features. These finds were then able to be counted as totals for data. For the CDC datasets, there actually was not any engineering necessary to be done to the data. Browsing through the data returned a favorable result of all of the datasets that were viewed not having any missing data or unusual inputs. The only concern was the limited amount of data there was due to not having data from all of the states in the United States. All of the data present though for the CDC datasets were pretty neat and clean.

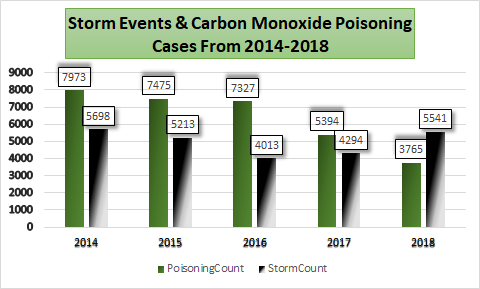
**Exploratory Data Analysis(EDA)**

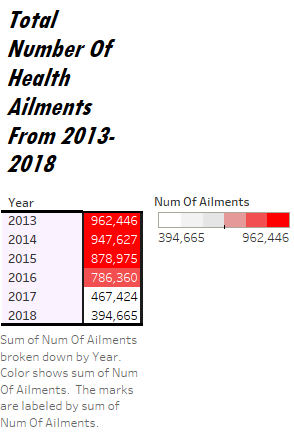
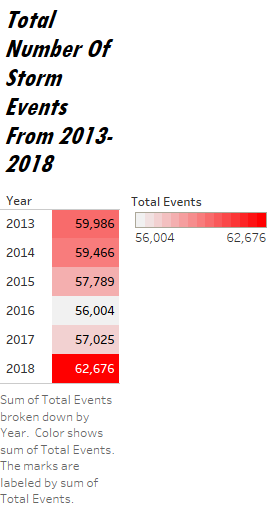
Once all of the datasets went through the data wrangling phase, they were ready to be displayed as visualization for quick and easy understanding of the data. Along with visualization comes statistical analysis of all of the data. Correlations and other trends in the data were analyzed. To the right is a visualization on hypothermia and how it relates to the effect it had on people starting from the year 2013 up to 2018. Each year displays the total number of hypothermia incidents and deaths in the United States due to storm events. Hypothermia deaths were counted using the Event Narrative feature in the NOAA database and the use of some engineering of the dataset. According to the graph, the years 2014, 2015, and 2016 produced the most incidents of hypothermia, with the year 2016 producing the highest total number of hypothermia incidents with a count of 64 incidents. From these findings, conclusions can be drawn that may suggest an increase in the number of winter storms during these 3 years. 2016 may be a year in which there were the most violent snowstorms. 

Next there was a heavy focus on the relationship between asthma and storm events over the 5 year period. 3 visualizations were created to give further insight on the data.

Displayed to the right is a heat map of the number of Emergency Department Visits for cases of Asthma in the United States for the year of 2014. Highlighted, is a count of wind related storm events(Strong winds, high winds, thunderstorm winds, etc) that have occured in 5 specific states in this year.These states were highlighted due to them being the extremes of the data. Florida, California, and New York have the highest number of asthma cases and also have relatively high numbers of wind events. New Hampshire has a low count of asthma cases and has a lower count of wind events. Looking at the visualization, the asthma cases and wind events seem to correlate strongly, the R^2 proved otherwise with a calculation of 0.47. This proves there is actually only a moderate correlation, though approaching the strong correlation range. To compare, another asthma map visualization was created for the following year. 

This data was a lot more spread out. The same states had high counts of asthma again, but only Florida and New York had high counts of wind events in comparison to 2014. California’s wind events count actually went down by more than half even though the count for asthma cases stayed very high. Surprisingly, the wind events count for New Hampshire went up by almost 9x in comparison to the previous year. New Hampshire’s asthma case count stayed low though. When determining the correlation, it was discovered that this data has a very low correlation(R^2=0.16). The next health ailment that will be focused on is carbon monoxide poisoning. 

This visualization displays the totals for the number of emergency department visits for Carbon Monoxide(CO) poisoning and the totals for specific storm events related to CO poisoning(tornadoes, hurricanes, and winter storms). These 3 storm events are related to carbon monoxide poisoning due to them causing power outages in affected areas. These are then forced to use technology that, without proper ventilation, can cause CO poisoning. 2014-2016 had approximately 2,000 more cases each year in comparison to 2017 & 2018. Strangely, though 2018 had a significantly lower number of cases of CO poisoning; yet it still managed to produce the second highest amount of storm events that year relating to CO poisoning. Through using the data analysis feature in Microsoft Excel; a very low to no correlation was determined(R^2 = 0.003) actually. Based on this, there are more than likely other factors that are more closely related to CO poisoning totals. Storms contribute to destruction of property, but there are also other ways CO poisoning can occur. 

To sum up the data collected, below are visuals of the total number of specific health ailments and the total number of storm events per year investigated. Due to there being a large variety of different health ailments; the data is only composed of 3 different ailments which include asthma, carbon monoxide poisoning, and hypothermia. From the visual it can be concluded that there does seem to be a trend in the years 2013, 2014, and 2015 having the highest totals in both visualizations, with 2018 being an anomaly. The year 2018 was also an anomaly when it came to the number of carbon monoxide poisonings reported. There was a very high number of storm events and a low number of CO poisonings in the year 2018 in comparison to the other years. The same is also displayed in both of the totals visualizations to the right. The year 2018 has a very high number of storm events with a very low number of ailments. This means the data is consistent with the carbon monoxide and hypothermia reports due to both displaying low incident counts in the year 2018. 

**Data Insights**

Overall there does not seem to be any strong correlations with storm events and the health ailments chosen to be investigated. There is still a possibility of other ailments related to storms to have stronger correlations. The 2014 asthma data contains the strongest correlation and this was only moderate. Looking at the data; a downward trend in the total number of storm events occuring and the total number of health ailments reported is present. The year 2018 does have a pretty great increase in the number of storms occurring, but these are not increasing the number of ailments present. Taking extra precaution against storms is always necessary, but it will not really affect the outcome of people having health ailments post-storm events. As far as dealing with asthma, Florida and New York did have high counts of asthma cases, while also having high counts of wind events. Florida especially had a very high wind event count. These 2 states’ residents should probably be more cautious during the spring season when the pollen count is high. To help spread awareness, governments can find ways to broadcast warning messages to the public.

**Recommendations**

The type of machine learning projects that could be produced from this project would be one that involves health surveillance that would create warnings for when a storm event is predicted to occur. The warnings would be alerts of ailments that are at risk of affecting the area at risk of the storm event to occur. A potential data product that can be created is a public database of health ailments related to specific storm events. Each instance of a specific storm event can have a list of ailments correlating or relating to it in some way. In the future, it would be recommended that other ailments are investigated due to there being such a wide variety. It would probably also be more beneficial to focus on specific storms and their effects in order to get more accurate data and be able to focus actionable insights towards specific storm events.

**References**

Meyers, T. (2019, July 5). How Hurricanes Cause Disease Outbreaks. Retrieved February 7, 2020, from <https://www.directrelief.org/2019/06/how-hurricanes-cause-disease-outbreaks/>

Kuhne, M. (2019, July 8). Thunderstorms can trigger deadly asthma epidemics. Retrieved February 7, 2020, from <https://www.accuweather.com/en/weather-news/thunderstorms-can-trigger-deadly-asthma-epidemics-2/432130>

Rapaport, L. (2018, September 21). Hurricane power outages may increase risk of carbon monoxide poisoning. Retrieved March 23, 2020, from <https://www.reuters.com/article/us-health-storms-carbon-monoxide/hurricane-power-outages-may-increase-risk-of-carbon-monoxide-poisoning-idUSKCN1M12S7>

<https://ephtracking.cdc.gov/DataExplorer/#/>

<https://www.ncdc.noaa.gov/stormevents/>