By Maidul Islam (mi252), Rishab Kaul (rk493)
Published 5/14/19
Introduction to Data Science

How do rising temperatures affect the number and severity of storms in the United States?

With the increase of news related to the devastating effects of storms and their said correlation with rising temperatures, we were curious to know how these two topics relate. According to scientists, climate change is a key player in more extreme weather [1]. To gain further understanding between storms and rising temperatures, we collected data from a number of different sources to give us a plethora of data on storms and temperature data in the US, sorted by state and year. State-level data points were necessary for a more accurate picture. We then measured human and property damage by state to find a better relationship between climate change and storms. We found that there was a positive correlation between the number of storm events and anomalies in temperature. Positive increases in temperature for a year correlated with a higher amount of storms for that year.

The significance of finding relationships between these two data sets will emphasize the United States' need to be more cognizant of the pollutants that may contribute to rising temperatures. While our findings will not be conclusive about whether rising temperatures will directly affect the severity of storms, we hope that our findings will point us to a possible area of concern relating extremity of weather patterns and its relation to rising temperatures.

Storm Attributes and Temperature Anomalies

Our first data set was compiled from the National Oceanic and Atmospheric Administration's (NOAA) and it involved data regarding storms that occurred in the US. Before exploring how rising temperature correlates with the number and severity of storms, we first explored the storms dataset alone to gauge preliminary trends. To create our models using the storms data set, we used quantitative attributes to measure the severity of weather events over time. For instance, we took into account the amount of damage in dollars caused to property, the state the event took place in, the deaths and injuries that were caused by the event, and finally the damage to crops in dollars. Using this data set we are able to gauge the change in these different attributes over time. Our findings indicate that crop damage, property damage, number of deaths, and number of storms each have had subtle increases over the period of 1950-2018.

Our second dataset was also compiled from the NOAA. It contains anomalies and average temperature for a year for each state in the United States. Anomalies describe the deviation of that year's average temperature from the known average temperatures of all years for that state. For example, if the anomaly was +2.0 for a particular year, this would describe how the average temperature for that year is 2 more than the average temperature for all years. Using years, we modeled the climate data along with storm data. First, we wanted to see how the number of storms per year correlates with the average shift in temperature. A higher storm count and positive change in temperatures over time would signal that higher temperature shifts correlate with more storms per year. After modeling these data points into a scatter plot, we discovered that both the average temperature change and storm counts per year positivity correlate with each other. Next, we modeled storm data attributes total crop damage, total deaths per year, and total property damage per year with temperature anomalies over the time period of 1950-2018. All of these attributes correlate positively with higher temperature anomalies.

For each model of the two datasets, points cluster together around an average change in temperature of zero. This is because, since 1950, most anomaly averages are situated around the zero marks, rather than away from it. As the years increased, anomalies increased as well as evidenced by global data on climate change. For each model, the severity of weather events was at a low around zero. However, as the changes in average temperature began to rise, changes in severity (damage to crops, number of storms, and number of deaths) began to slowly gain upwards momentum.

Distance between States and Temperature Anomalies

After having modeled both datasets individually and datasets together with the average change in temperature, we wanted to quantify how close the relationship between the states that had the highest change in temperature and those that had the most property damage. Our plan was to create two data sets, one derived from the temperature dataset and the other from the storms data set. Both of these data sets would be ordered in ascending order by the same range of dates and they would each contain states. Then we took the first state from the derived temperature data set and the first state from the derived property damage data set and we calculated the straight line path between the two states in kilometers. Our hypothesis was that states with the highest property damage should be close to the states with the most change in temperature. With this thought process, a smaller distance between the states from both of these data sets would mean that a state with higher temperatures would be more likely to have more property damage overall.

After graphing the distances between the states, our findings indicate that the relationship between states with high temperatures and property damages caused by extreme weather events were unclear. If the relationship between the state with the most positive change in temperature and the state with the most property damage were the same, we would see a decline in distance between states which would indicate that as time increases, states with higher temperatures would correlate with higher numbers of property damage. However, our model from plotting distances between states over time describes an unclear trend. The absence of a clear pattern can be attributed to numerous reasons. One is that we did not take into account that all states are not equally prone to storms. Some states may have more storms in general because of their geography. If we could somehow differentiate states based on how prone they are to storms and quantify their geography, we may have a clear pattern in this model.

Detailed analysis

Attributes used

We used the following data attributes from the storms data set:

- Damage_property - Gives us detail on the estimated amount of damage to property by the specific weather event. For example,10.00K = $10,000; 10.00M = $10,000,000
- Deaths_direct - Gave us the number of deaths directly related to the said weather event.
- Damage_crops - The estimated amount of damage to crops by the said weather event. For instance, 10.00K = $10,000; 10.00M = $10,000,000)
- State - in which state the storm event took place

For the temperature data by states, we used the following attributes:

- Date - the year that the data point was taken
- State - the state that the data was collected from
- Anomaly - the change in deviation from the base year

Summary Statistics

Damage to Property

- Mean: $476,245.96
- Median: $0.0
- Standard Deviation: 39102867.19

Direct Deaths

- Mean: 0.0117 deaths
- Median: 0 deaths
- Standard Deviation: 0.637

Damage to Crops

- Mean: $68,903.80
- Median: $0.0
- Standard Deviation: 3560839.84

Number of Storm Events

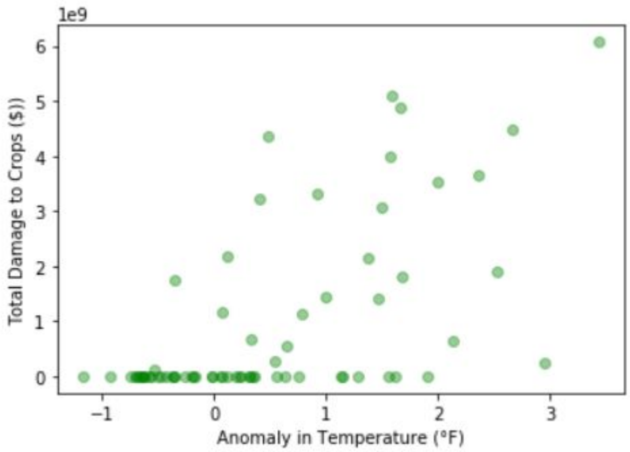
- Mean: 22,233.90 events
- Median: 7335.0 events
- Standard Deviation: 25162.48

Highest Storm Damage ($) vs Year

slope=81916331.75, intercept=-161134864511.02, rvalue=0.4983, coefficient of determination= 0.2483

In the above scatter plot, we took each unique year in the storms data set and set its value to the storm that caused the highest property damage that year. Our scatter plot shows that for every year increase, there will be $76,782,131 in approximate damage done to properties through the states. The R from this graph is 0.4766, which implies a positive correlation. The purpose of this graph is to show the most severe storm for each year. The general pattern shows an upwards trajectory. The coefficient of determination is 0.248, which means that only a meager 24.8% of the data can be explained by the linear relationship between the two variables.

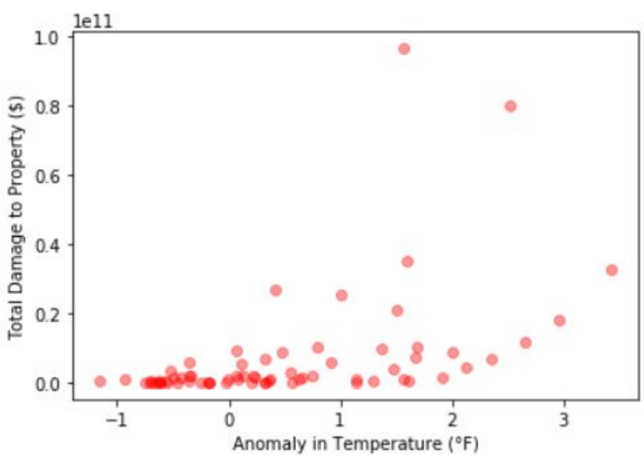
Total Damage to Crops vs Anomaly in Temperature



slope=918327738.85, intercept=456660373.169, rvalue=0.6107, coefficient of determination= 0.3730

In this scatter plot, the total damage to property is plotted against anomalies in temperature. The correlation coefficient here is about 0.61, which implies a positive correlation. The coefficient of determination of about 0.37. This means that about 37% of the data plotted can be explained by a linear relationship between total damage to crops and anomalies in temperature. As you can see above, a lot of the points are situated around zero anomalies. This happens because most years did not deviate a lot from average temperatures in the past. As the years increased, anomalies generally tend to be higher, and higher anomalies seem to correlate with higher damage to crops.

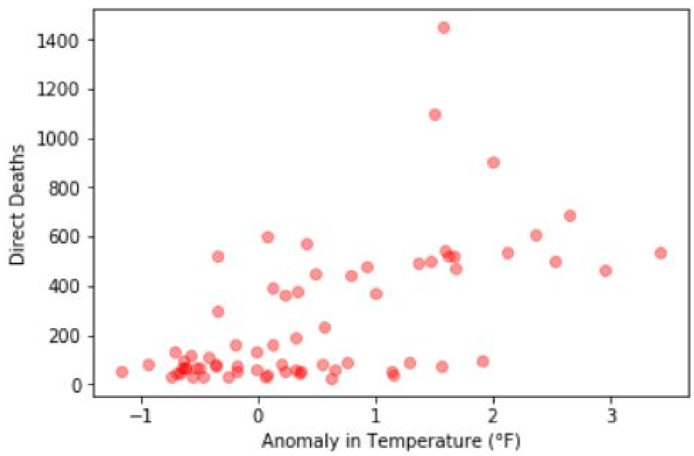
Total Damage to Property vs Anomaly in Temperature



slope=7372473100.03, intercept=3417862427.64, rvalue=0.4768, coefficient of determination= 0.2274

Here, we see a clear linear relationship between total damage to property and anomalies in temperature. The correlation coefficient here is about 0.477, which implies a positive correlation. The coefficient of determination of about 0.227. About 22.7% of the data plotted can be explained by a linear relationship between total damage to property and anomalies in temperature. Similar to the previous graph, a lot of the points are situated around zero anomalies, as during most years in the dataset, there are not too many extreme anomalies. With this data, higher anomalies correlate with higher damage to crops. This makes sense because according to NASA, warmer heat in the atmosphere could provide more fuel to create more disastrous storms[2].

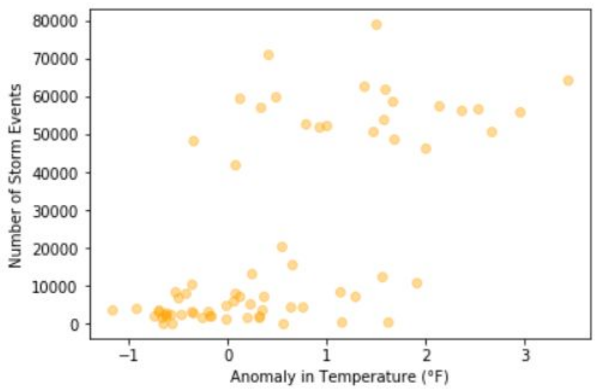
Direct Deaths vs Anomaly in Temperature



slope=163.240, intercept=179.195, rvalue=0.604, coefficient of determination= 0.3651

Here, we explore a linear relationship between direct deaths from storms and the anomaly in temperature for each year in the dataset. The correlation coefficient here is about 0.604. This implies a positive correlation between deaths directly caused by storm events and the average anomaly in temperature during that year. The coefficient of determination is 0.365, which means that about 36.5% of the data can be explained by the linear relationship between the two variables. Higher temperature anomaly here seems to correlate with deaths from storms.

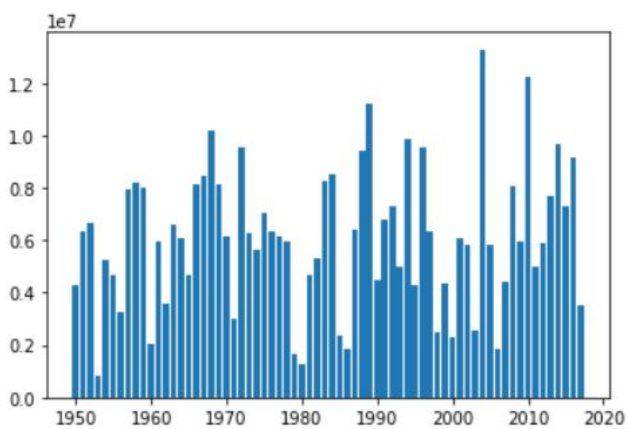
Total Number of Storm Events vs Anomaly in Temperature



slope=15638.03, intercept=14412.509, rvalue=0.650389, Coefficient of Determination: 0.4230

In this graph, we measured the relationship between anomalies in temperature for a year, and the number of storm events that occurred for that year. The correlation coefficient here is about 0.65- Both the correlation coefficient and the plot shows a positive correlation. Increases in temperature anomalies here correlate with increases in storm events, while decreases in temperature anomalies clearly show a lack of storm events. The coefficient of determination is 0.42. About 42% of this data can be explained by the linear relationship between the two variables.

Distance in km between States and Years



slope=76782131.33, intercept=-151006798614.64, rvalue=0.4766841

The above bar chart shows the years on the x axis and the distance in kilometers on the y-axis between the two states highest change in temperature and highest cost in property damage from storms events. The r-value of this graph is 0.47 which means that there is a relation between the two variables. We expected the r-value to be negative because from looking at the scatter plot of temperature and number of storms, there seems to be a high correlation between the two variables implying that higher temperature yields more storms. The states with the most property damage should also have the highest change in temperature. Our model indicates that such a hypothesis is incorrect according to the trend above. We hypothesized that the states with the same ranking on property damage and high temperatures would either be the same state or be close together. However, the data reflects that there is no clear relationship here. We expected a negative relationship since later years should have less distance between the two states. In order to improve this trend, we have to incorporate more variables and study the geography of states more closely.

Conclusion

The data comparisons above show a consistent correlation between rising temperatures in the states and attributes of storms such as damage to property, damage to crops, and deaths. They also show a positive correlation between the number of storms and rising temperatures. While these relationships are positive, they are not indicative of the whole picture between these variables. There are countless other factors that can affect the severity of storms and our data only shows its possible correlation to rise in temperature. Thus, our findings are in no way a conclusive statement on the relationship between these variables. However, our findings do provide an area of concern that needs to be analyzed with the use of more attributes to yield a more accurate picture. Some other areas of concern include outside factors that potentially contribute to the correlation seen, such as geographical regions that are more prone to weather events over others and inconsistencies between the quality of technology used over time. A broad period from 1950-2018 sees changes across the board, from technological advancements to political establishments. As such, more research is needed on this subject, so that these limitations are taken into account when building new models.

References

1. Rosenzweig, Cynthia, et al. “Climate Change and Extreme Weather Events; Implications for Food Production, Plant Diseases, and Pests.” *SpringerLink*, Kluwer Academic Publishers, link.springer.com/article/10.1023/A:1015086831467?LI=true.
2. “The Rising Cost of Natural Hazards.” *NASA*, NASA, earthobservatory.nasa.gov/features/RisingCost/rising_cost5.php.

Data sources:

1. Temperature data was collected from https://www.ncdc.noaa.gov/cag/statewide/time-series/41/tavg/12/12/1895-2019?base_prd=true&firstbaseyear=1896&lastbaseyear=2000
2. The storms data was collected from https://www1.ncdc.noaa.gov/pub/data/swdi/stormevents/csvfiles/
3. Distances between individual states were taken from https://www.mapdevelopers.com/distance_from_to.php