Chris Hoffman

Data Science

Spring 2022

Final Project Report

Introduction

The warming of the earth and climate change has been studied and reported extensively on a global scale, but far less so on the regional level. This project explores one aspect of climate change and global warming more locally through the study of tornadoes and their association with winter temperatures in the counties of Wisconsin. Climate data shows that Wisconsin has become warmer and wetter over the last 60 years, and that tornadoes have not necessarily become more frequent over the years but have increased in density (Wisconsin DHS, 2019).

Exactly what causes tornadoes to form is still unknown and studying the impact of global warming on tornadoes can be difficult, as they are localized and short lived; however, increased temperature increases the moisture capacity of the atmosphere, priming areas for the supercell thunderstorms that can lead to tornadoes (National Geographic, 2019). This project will not search for a causal link but will look at how good of an indicator winter temperature in a location is of whether or not tornadoes will occur in the following tornado season.

Question

The question being pursued in this project is “Does the temperature over the course of winter in individual counties of Wisconsin indicate whether or not a tornado will occur in that county?”

Data Collection

All tornado and temperature data collected was downloaded as .csv files from NOAA. Initial temperature data collected consisted of average, maximum, and minimum monthly temperature from 1950-2020 throughout Wisconsin as a whole. After initial analysis, additional data was collected. This included average, maximum, and minimum monthly temperature and total monthly precipitation for each Wisconsin county from 1950-2020. The tornado data used consisted of a record of all tornadoes from 1950-2020 with the date, start time, severity, injuries, fatalities, start coordinates, end coordinates, state, and other descriptors for each tornado. Data for geospatial visualization was collected from the US Census Bureau in the form of an .shp file.

Data Formatting

After initially extracting Wisconsin tornadoes from the tornado data, the start county of the tornadoes was calculated. This was done with a library called geopy, which took the start coordinates of the tornado and extracted the county from them. Some counties ended up not being from Wisconsin due to tornadoes starting out of state before moving into Wisconsin. These tornadoes were removed from the data set. Additionally, Menominee County was created in 1959, which is 9 years after our dataset begins. This resulted in Menominee County not being included in the tornado dataset. Menominee County was removed from all climate data used because of this and was not used for modeling or prediction. Tornado data was then grouped by year and county with the average magnitude of tornadoes, total fatalities, total injuries, total number of tornadoes, and maximum magnitude of tornadoes each year calculated.

County level monthly climate data was taken and cut down to winter months. This included January, February, and March. December was excluded as winter officially starts towards the end of the month, and including it would make grouping by year difficult with a new year starting January 1. The data was then grouped by year and county and average winter temperature, average maximum winter temperature, average minimum winter temperature, average winter precipitation, total winter precipitation, maximum winter temperature, and minimum winter temperature each year were calculated in the process.

The formatted tornado and climate data were then merged into one table. This table showed all climate and tornado data for each county each year. From this, 5 year rolling averages were calculated for all climate data each year. All data from years 1950-1953 was NaN for these values since the data was not there to compute those averages. A 5 year rolling average was also calculated for tornado counts, but the current year was not included in the calculation as that would give away tornado occurrence for that year. Therefore, all data from 1950-1954 was NaN for this rolling average.

Tornado occurrence was added as a column by checking to see if the count for a given county each year was greater than 0. If it was, it was indicated as 1, meaning a tornado occurred that year, and if it wasn't then it was indicated with a 0. Whether or not a tornado occurred in previous years was also calculated. This was done for whether a tornado occurred last year, within the last two years, and within the last three years for each county each year.

Preliminary Analysis

Tornado data was initially analyzed statewide to see general trends and behaviors. This began with tornadoes per year being plotted over time. Figure 1 shows a slight trend upward in

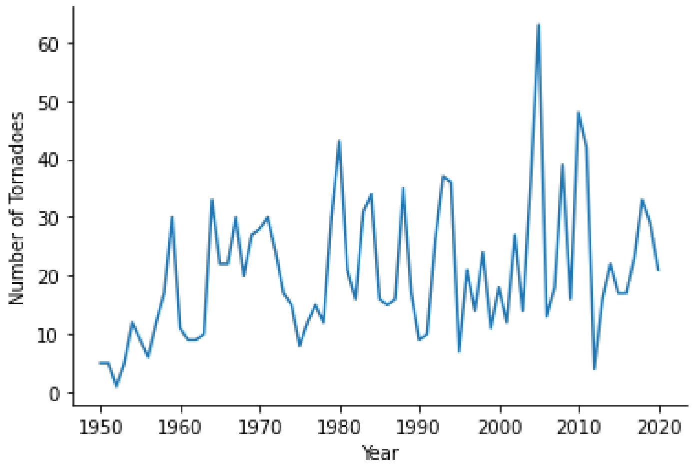

Figure 1. Number of tornadoes in Wisconsin per year from 1950-2020.

tornadoes yearly since 1950, but more notably it shows a cyclic up and down pattern between years. There are usually years of fewer tornadoes followed by years of many tornadoes. This showed that past tornado occurrence may be able to be used for modeling future tornadoes.

The average yearly winter temperature in Wisconsin overall was then plotted over time. This showed a cyclic pattern similar to that of the tornado data. Years of cold winters were

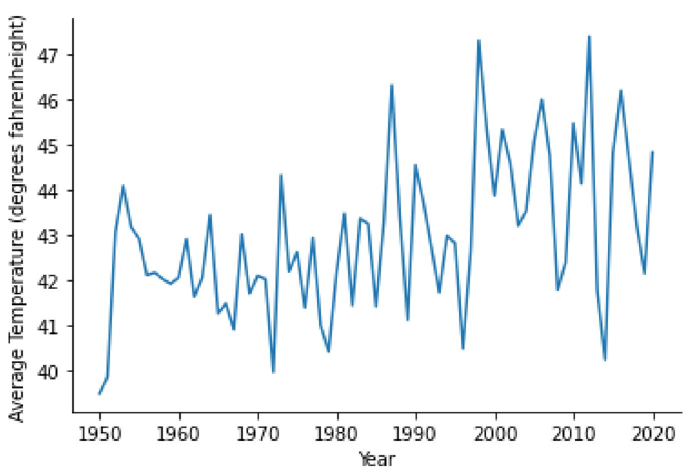

Figure 2. Average Winter Temperature in Wisconsin from 1950-2020.

followed by years of warm winters (Figure 2). The volitility in winter temperature each year showed that in building a model it would be beneficial to look at averages across multiple years rather than individual yearly anomaolies. Looking at a rolling average gives a better idea of the

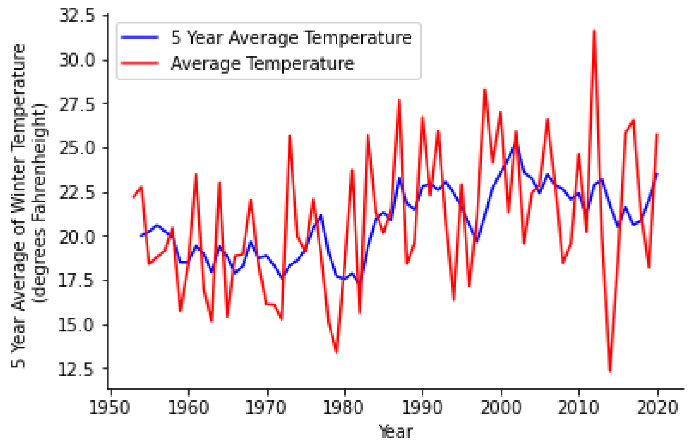

Figure 3. Rolling 5-year average winter temperature in Wisconsin from 1953-2020.

climate in an area compared to the temperature of a single year. Rolling average can also be used for looking into the future. The slope of the average line is trending upwards before getting to 2020 in Figure 3, so it could be predicted that temperature will continue to increase.

To get an idea of how climate data and tornado location was distributed throughout the

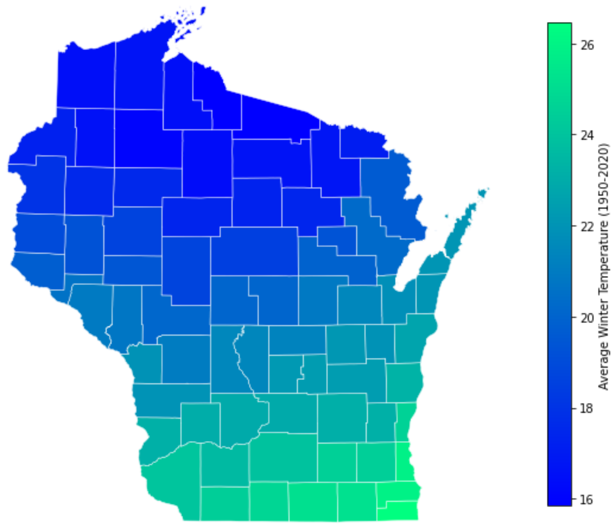

Figure 4. Average winter temperature across Wisconsin counties from 1950-2020.

state, the data was plotted on a map of wisconsin. Figure 4 shows average winter temperatures distributed across Wisconsin. The warmer counties tend to be towards the south as expected

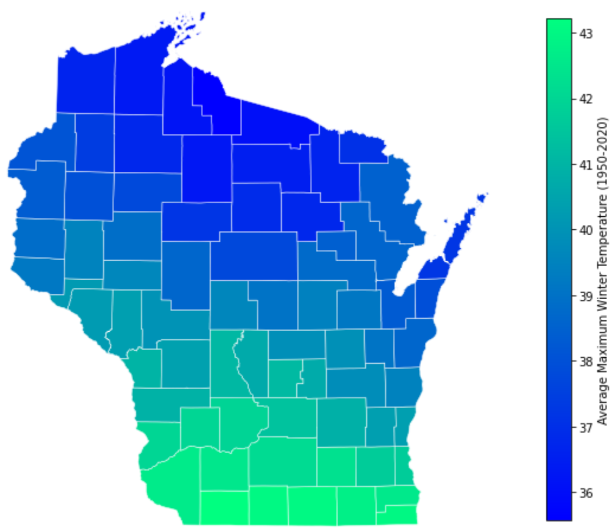

Figure 5. Average maximum winter temperature across Wisconsin counties from 1950-2020.

(Figure 4). The average maximum winter temperatures in Wisconsin are shown in Figure 5. The distribution of the maximum winter temperatures deviate from the simple averages. There is more of a tendency for counties along the southwestern part of the state to have higher maximum temperatures (Figure 5). Figure 6 shows th minimum winter temperatures—the distribution of

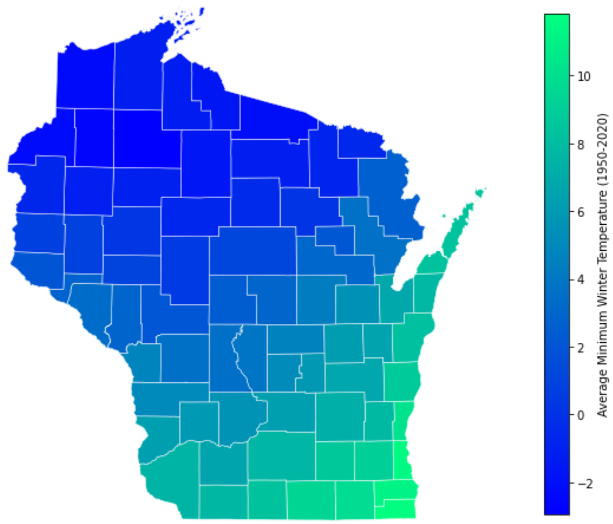

Figure 6. Average minimum winter temperature across Wisconsin counties from 1950-2020.

which more closely matches that of the simple average winter temperatures with the warmer temperatures stretching along the southeast side (Figure 4 & 6). Figure 7 shows the hottest

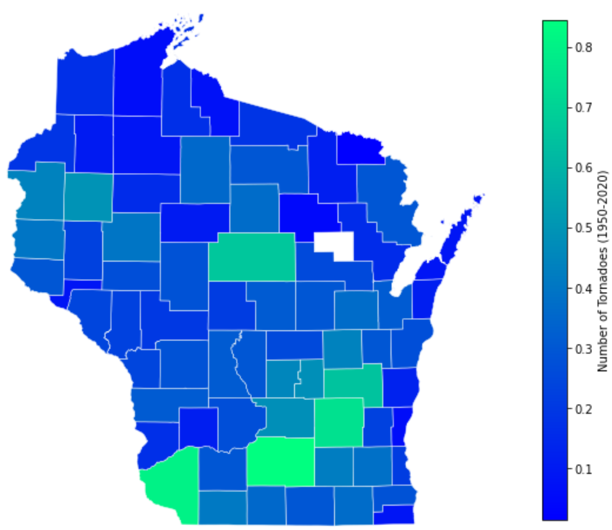

Figure 7. Total number of tornadoes in Wisconsin counties from 1950-2020. Menominee county (white) is excluded due to lack of data.

counties in terms of number of tornadoes from 1950-2020. The distribution of the hardest hit counties does not match that of the temperature data, so it was clear that more than just temperature data alone would be needed for modeling.

Data Modeling

Before modeling, the data was split into train and test sets. A logistic regression approach was used for modeling and different combinations of the calculated attributes outlined in data formatting were explored. The final model selected one hot encoded county name and whether a tornado occurred in the previous year in that county and used a standard scaler on the average over 5 years of the average winter temperature in a county. When fitted and tested on the test set there was a final accuracy score of 0.544, a precision score of 0.246, a recall score of 0.593, and an f score of 0.348. Overall, this model is not the best indicator of whether a tornado will occur. This is particularly shown with a low precision score—many counties are incorrectly predicted to have a tornado. However, with an accuracy above 0.544 the model performs better than a simple coin flip, showing that there is something here that may benefit from being explored further.

Deliverable

An interactive map was the chosen deliverable for displaying the results of this project. The map uses blue and green colored counties to show whether a tornado is predicted to occur during 2022. When placing the cursor over a county, the data that was used to make the prediction is displayed as well. A screenshot of this map is shown in Figure 8 (Figure 8).

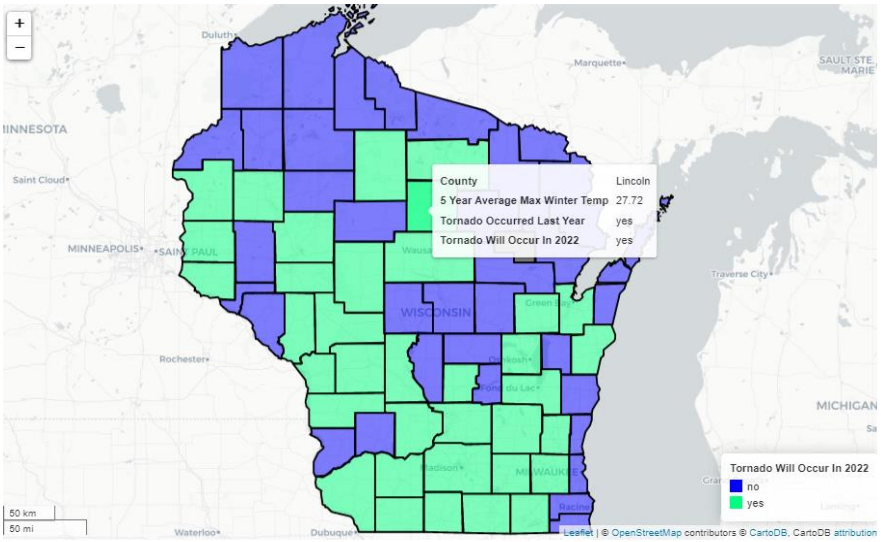

Figure 8. Interactive map showing final model data and predictions.

Reflection

Tornado prediction proved difficult which was to be expected. Professional climatologists work on predicting adverse weather events every day and have still barely scratched the surface on prediction tornadoes ahead of time. A model with around 54% accuracy did show some success being that it was better than a coin flip, but it still is not anything that would or should be used in practice.

If this project were to be continued, additional data would need to be collected. This would include general climate patterns across the country, particularly through tornado alley. In discussing this project with my geography professor, he said that it is generally believed tornadoes are related to world-wide wind patterns, so that would need to be explored as well. While the answer to the initial question proposed in this project would be that winter temperature is not a good indicator of tornado occurrence, the project shows that temperature may play a role, even if that role is small, in future tornado prediction.

References

National Geographic Society. (2019, October 24). *Tornadoes and Climate Change*. National Geographic Society. https://www.nationalgeographic.org/article/tornadoes-and-climate-change/

Wisconsin Department of Health Services. (2019, October 5). Severe Thunderstorms and Tornadoes Toolkit: A planning guide for public health and emergency response professionals. https://www.dhs.wisconsin.gov/publications/p01037.pdf

Data Sources

https://www.weather.gov/mkx/wisconsintornadoes

https://www.spc.noaa.gov/wcm/#data

https://www.ncdc.noaa.gov/cag/county/mapping