Kevin O’Donnell

Dr. Bernhardt

STAT 4380

3 May 2021

Rising Air Temperatures and Extreme Precipitation in Connecticut

This project explores the relationship between climate change and precipitation. My goal is to see if there is a link between increasing air temperatures and extreme precipitation events (like hurricanes and floods). Climate change is said to exacerbate extreme weather events, so I want to determine if that claim is corroborated through outliers in precipitation readings. To do this, I have compiled daily summary data from stations in each of the 8 counties in Connecticut from 1900 to 2021. The data contains the daily high, low, precipitation for the station, as well as identifying information such as the station name and location. I will explore 3 questions to arrive at my final conclusion:

1. Have air temperatures increased in Connecticut from 1900 to 2020?
2. Are days with very high or low levels of precipitation becoming more common?
3. Is there a link between these two phenomenon?

# Have Air Temperatures Increased in Connecticut from 1900 to 2020?

Chart, line chart, histogram

Description automatically generated

Chart, line chart, histogram

Description automatically generated

According to the plots above, average high and low temperatures are volatile year to year, because weather is extremely variable by nature, but there is a general upward trend for both variables. High temperatures have generally increased by around 4-5 degrees since 1900 and low temperatures have increased more dramatically by about 6-7 degrees. Average yearly low temperatures have undergone a rapid increase in the last 2 decades, which accounts for much of the difference between 1900 and today.

Chart, bar chart

Description automatically generated

Chart, bar chart, histogram

Description automatically generated

To reduce the variability in temperature that is seen in the first 2 graphs, I plotted average high and low temperatures by decade. This shows a similar trend, though the increase looks less significant on this scale. Average high temperatures by decade seem to be slightly higher in the last decade compared to the early 1900s and, again, average low temperatures appear to have increased more rapidly in recent years, creating a noticeable difference between the 2010s and 1900s.

A picture containing application

Description automatically generated

A picture containing application

Description automatically generated

A picture containing application

Description automatically generated

A picture containing application

Description automatically generated

To visualize the changing temperatures in a more appealing way and see which counties have experienced the greatest differences in average temperature, I created maps for the average high and low temperatures by decade in the 1950s and 2010s. In general, average high temperature in the 2010s is higher than in the 1950s, because many counties on the 1950s map are shaded darker than the same counties on the 2010s map. New Haven County in particular appears warmer in the 2010s compared to the 1950s. It had an average high of just below 60 degrees in the 1950s, but its average high in the 2010s was around 63 degrees. With that being said, some counties, like Litchfield and Tolland, have very similar average high temperatures for both decades.

The average low temperature maps follow a similar trend. The 1950s map is darker overall, meaning that average low temperatures were generally lower in that decade compared to the 2010s. New London county experienced the most dramatic difference in average low temperature between these 2 decades. Again, Tolland and Litchfield experienced very small differences in low temperature during the same time period.

Chart, scatter chart, box and whisker chart

Description automatically generated

Chart

Description automatically generated

In an attempt to solidify the results displayed in the previous three visualizations, I created confidence intervals for average high and low temperatures by decade through bootstrapping techniques. I replicated 1000 samples of the length of the data with replacement from each decade for high and low temperatures. From there, I calculated the mean of each sample and found the 95% confidence interval for the mean high and low temperature for each decade. I utilized the bootstrapping method to make my data more robust and to test whether the average temperatures were significantly different today compared to the early 1900s.

The results affirm this because the 95% confidence intervals for average high and low temperatures are much different in the 2010s compared to the 1920s, or any other decade in the data set. There is no overlap between the confidence intervals for the 2010s and those for any other decade. Additionally, each graph shows that the confidence interval is shifting right, in general, as time passes. In conclusion, the mean high and low temperatures appear to be increasing over the last 100 years.

# Are days with very high or low levels of precipitation becoming more common?

Chart, line chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

In the graphs above, it is apparent that average yearly precipitation and the standard deviation of precipitation fluctuate, but have remained around the same level since 1900. There is no significant increase or decrease in the average amount of precipitation, nor is there any evidence that the variation within each year has increased or decreased. Additionally, the proportion of days with extreme precipitation, which also has fluctuated greatly, has generally decreased over time.

A picture containing application

Description automatically generated

Application

Description automatically generated with medium confidence

I created a map for the average proportion of days with extreme precipitation to further explore the trend, or lack thereof, in extreme weather events over time. It is unclear if extreme events were more common in the 1950s or the 2010s because neither map is noticeably darker or lighter overall. The southern counties in Connecticut seem to have lower proportions of days with extreme precipitation events in the 2010s, but the opposite is true for the northern counties. Since these results were inclusive in determining the overall trend, I used the same bootstrapping technique as I used above to produce more robust data.

Table

Description automatically generated

The 95% confidence intervals for the proportion of days with extreme precipitation are almost identical every decade. There is considerable overlap and, unlike the confidence intervals for high and low temperature, there is no evidence for a significant difference between any decades from the 1920s to the 2010s. This is consistent with the previous plots because it shows that extreme precipitation events in Connecticut are not becoming any more or less common as time goes on.

# Is there a connection between increasing air temperatures and extreme precipitation?

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

I have concluded that air temperatures have increased since the early 1900s in Connecticut, but I cannot conclude that extreme precipitation events have become more common. Therefore, I cannot conclude that rising air temperatures are correlated with an increase in severe precipitation. The graphs above corroborate the claim that there is a weak relationship between air temperature and precipitation. Neither the high nor the low temperature have a positive or negative relationship with extreme precipitation. Also, high temperature has a weak correlation with precipitation (-0.15), as does low temperature (-.08). Further analysis with higher quality data is necessary to determine if extreme weather is becoming more common in Connecticut and if it is connected to rising air temperatures.

# Additional Findings

Diagram

Description automatically generated with medium confidence

I created univariate plots for the major variables in the project to examine their distribution. The distributions aren’t crucial for answering the overall research question, but they can inform the data. High and low temperature are approximately bimodal, likely corresponding to the highs and lows during the two longest seasons in Connecticut: summer and winter. Precipitation, however, is right skewed because most days have little to no precipitation. Therefore, extreme precipitation events, as I defined them, will be days with large rain or snowfall totals. The yearly averages for each variable are approximately normally distributed, which is expected for a sampling distribution of this size.

Chart, line chart

Description automatically generated

In addition to observing the univariate distributions, I also looked into a claim made by the EPA that the difference between high and low temperatures, or diurnal temperature variation, is decreasing during the summer. They claim that climate change leads to less “cooling off” during June, July, and August. This plot does show that diurnal temperature variation is lower in the past 2 decades than in the past, but because the smoothed loess model has a parabolic shape, I cannot conclude that it has decreased over time. I would need more data before 1900 and in the future to determine if this trend is cyclical or if diurnal temperature variation truly is decreasing.

# Shortcomings and Challenges

There are multiple shortcomings in this dataset and my analyses that are important to address. At the beginning of the project, I planned to use a much more comprehensive dataset that aggregated station data from across the United States from 1900 to 2020. This data has long term monthly means, as well as normal monthly weather data, for precipitation and air temperature. It covers a wider geographical area and has much more data, which would have allowed me to draw stronger conclusions about climate change as a whole. Unfortunately, it is in an unfamiliar file format and I could not import it into R for my own analysis.

The data that I used was sufficient, but the coverage for most stations was imperfect. Not every station covered all years between 1900 and 2020, so some years, particularly those in the early end of this time frame, did not have as much data. Many of the graphs, therefore, showed much more variation in the early 1900s compared to the late 1900s and 2000s. It would have been ideal if the same amount of data was available for each year so that my conclusions could be stronger. Additionally, within each year, some months were missing at random. For example, the Mount Carmel station was missing many days from the summer months in 1997, which caused its yearly average high temperature to be 5 degrees colder than the next coldest station. Many other gaps in the data are surely present, which could have confounding effects on the visualizations and summary data I produced. For many visualizations, I used decade long periods to try to ensure that these gaps had a lesser effect than they would have for yearly plots. However, I cannot be certain the missing data did not influence the results.

Another issue with my analysis is the definition of extreme precipitation. Since I only had daily precipitation readings and no data that explicitly defined serve weather, it was up to me to arbitrarily define extreme precipitation events. It is possible that this definition was too strict or too loose, which would affect my plots and conclusions. If I were to extend this project, I would try to find a dataset that contained more specific severe weather data and covered a greater geographical area so I could make stronger conclusions about the frequency of severe weather.

Other challenges arose during data wrangling and bootstrapping, beyond issues with missing and limited data. During the data wrangling phase I had to create many different summary variables for temperature and precipitation in order to graph them properly. For example, I had to group by each station and year to calculate the yearly averages for high temperature, low temperature, and precipitation. The process itself was not challenging but organizing every variable within their respective groups and mutate statements was confusing at times. A more daunting challenge was defining extreme precipitation. I decided to define extreme as greater than 2 standard deviations from the mean, since observations with p-values less than the alpha level of 0.05 are usually considered statistically significant. Once I defined these variables, I was able to produce plots without too much extra wrangling.

The bootstrapping phase of this project was the most difficult because I had to learn how to use for loops in R and implement bootstrapping, which was a new skill. I initially had trouble assigning values (like the mean high/low temperature and proportion of days with extreme precipitation) to empty cells in a matrix because I was struggling to implement a counter variable. I solved the problem by using the years I was looping through as the counter variable and omitting all missing rows in the matrix after its creation. Once I created the loop for the high temperature confidence interval, I was able to copy it and make minor adjustments to sample the low temperature and precipitation and record those means.

# Conclusions

In conclusion, the plots I generated during this analysis suggest that air temperatures have increased since 1900 in Connecticut. However, I could not conclude that days with extreme precipitation have become more common during the same period. As such, there is a weak relationship between air temperatures and precipitation. I cannot conclude that rising air temperatures and extreme precipitation are connected, as the hypothesis stated.