ECNS 460: Project Stage 2 - Visualization

Tim Cuddy & Tommy Mellott

Our project uses weather data to measure temperature change across the United States since 1900. We are using this data to help demonstrate the effects of climate change in the United States. Beyond the figures provided in this report, we will create a spatial visualization for the 100 randomly sampled weather stations we use in our project. Most of our data is taken from the Daily Global Historical Climatology Network (GHCN-Daily), which is a public dataset that reports raw temperature and climatological data from weather stations across the World, over the last 150 years.

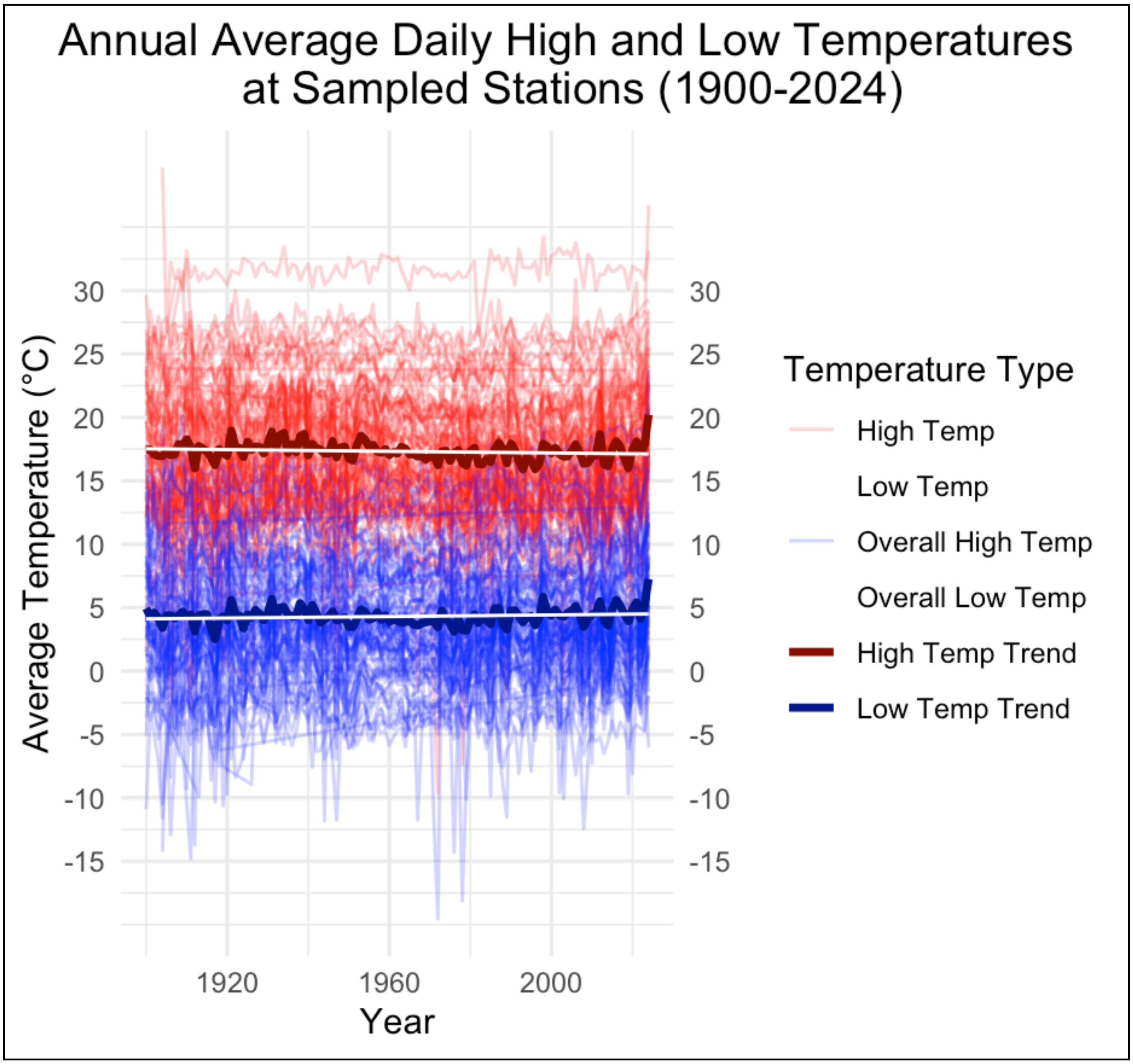
This is an extremely large, unwieldy dataset, with millions of observations and Tens of thousands of variables. The reason for that bulk is because we have date specific temperature information for every single temperature station from when they first started recording data, in some cases as early as the 1870s, until the present day. Our first objective, therefore, was to make this data more manageable and presentable. For that reason, we chose to take a random sample of 100 weather stations spread across the US. This decision will likely reduce some of our variation, although it hopefully won't enable any outliers. We accept that consequence because sampling enables us to present information on individual stations in a detailed manner, without having to sort through 80,000 variables.

Next, we took our random sample and filtered the weather data to include only those 100 stations. We cleaned this data and developed 3 core dataframes: TMAX, TMIN, and Station. In the process of developing these dataframes, we only made a few significant decisions. First, we chose 1900 as the lower bound for our data, removing any data points from the 19th century, which ensured our stations had uniform start and end points. Second, we chose to reformat Julian days (jday) into normal dates, which made the whole thing much easier to understand, and allowed us to break data down by month and year, which is critical going forward. Finally, temperature was previously formatted without decimal points, meaning 22.4 degrees was represented by the number 224, we divide by 10 to bring values back to normal.

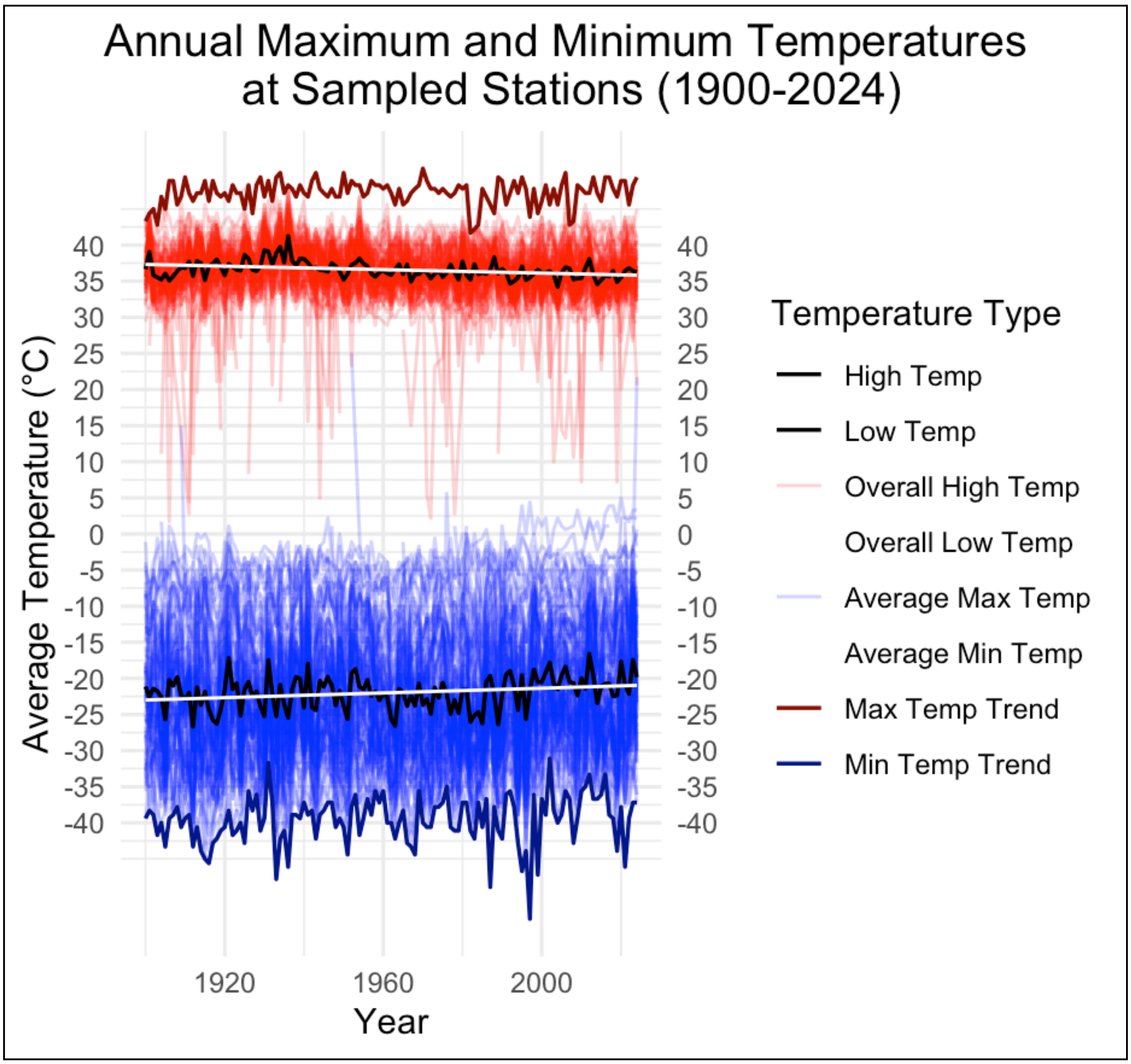
We also had to make a few decisions about extreme and missing values. First, we excluded anything without a jday throughout, as we have no way to analyze without date information. Second, some days have maximum temperature values but not minimum temperature values, we exclude these as necessary, such as for the creation of specific figures, but don't remove them entirely. There are a variety of extreme values as well, showing temperature values well beyond the point of realism, breaking world temperature records. To avoid this, we excluded any temperature value greater than positive or negative 60 degree celsius.

Figures

The first figure we made presents daily high and low temperatures for each station, averaged across the whole year. This is an important figure to see up front, since our data is broken down into daily high and low temperatures, we need to ask ourselves how that data behaves. This graph is not, however, a good metric to demonstrate climate change, as is clear by the almost completely flat or slightly downward sloping trend lines. There are two main reasons we think that is the case. First, climate change often presents itself as a story of extremes, as maximum temperatures get higher and lower across the board. When taken as an annual average, however, this impact is not visible. This graph helped to guide the choices we made regarding what other graphs to make.



The second figure is similar, but instead of averages, this shows the highest and lowest overall temperature a station measured in each year. This presents a much wider range of temperatures and provides us with a slightly more interesting story. The purpose of this figure is, first, to contrast with the first one by showing whether extreme temperatures and average temperatures are changing differently. Second, this figure helps us see if the range of temperatures seen throughout the year is changing. To both of those questions we do get answers, although they are somewhat different from what intuition may have suggested. The contrast is clear between this graph and the first we looked at, while averages see nearly imperceptible changes (on the scale of these graphs), the extremes presented in graph 2 demonstrate clear shifts, showing a decrease in maximums and an increase in minimum temperatures. This demonstrates that, among our sample, there is a decrease in annual temperature range, which counters traditional expectations.



The third and fourth figures we developed are paired, just as the first and second were. Both of these seek to answer the same questions that were asked by the first two: what change in temperature patterns have we seen? The key unique element that figures 3 and 4 share is that they narrow the sample to a single month and, therefore, a single season. This allows us to avoid counterbalancing, like we encountered in table 1, and doesn't provide us only with a single moment's worth of temperature, like we encountered in table 2.

In figure 3, we look only at the average high and low temperatures for the month of August. This graph shows decreasing average high temperatures and increasing average low temperatures since 1900, this is in line with the results of Figure 2 and supports a conclusion of decreasing annual temperature ranges.

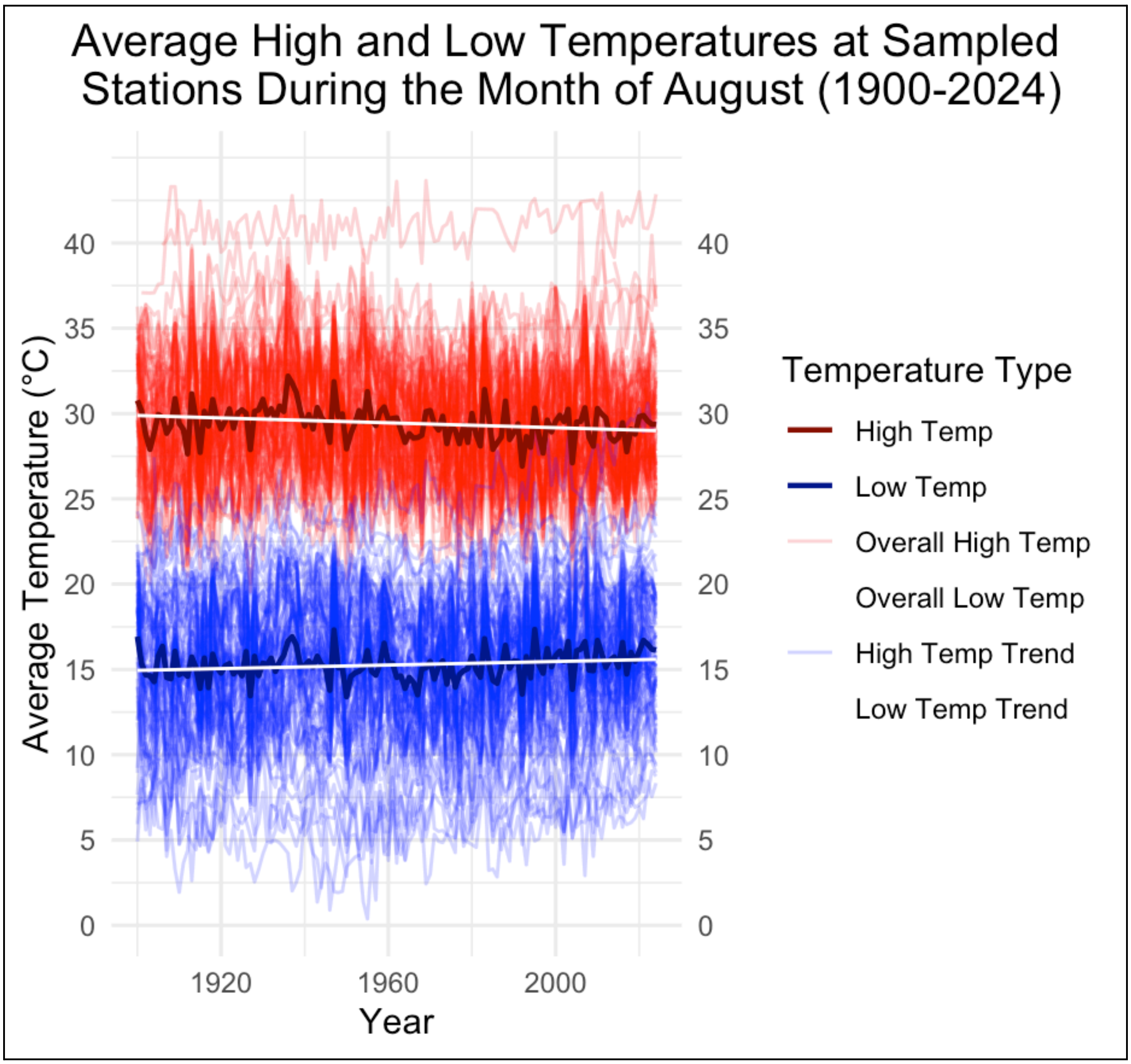


Figure 4 tells a slightly different story, showing average temperatures only for the month of February. In this graph, for the first time, we see evidence of an increase across all measures, with the average high and low temperatures increasing slightly since 1900. This supports the idea of milder winters as a result of climate change, and also suggests that the somewhat unexpected results we saw in figures 1 and 2 may be a result of summer variation, while winter sees the increasing temperatures that are described by climate change scientists.

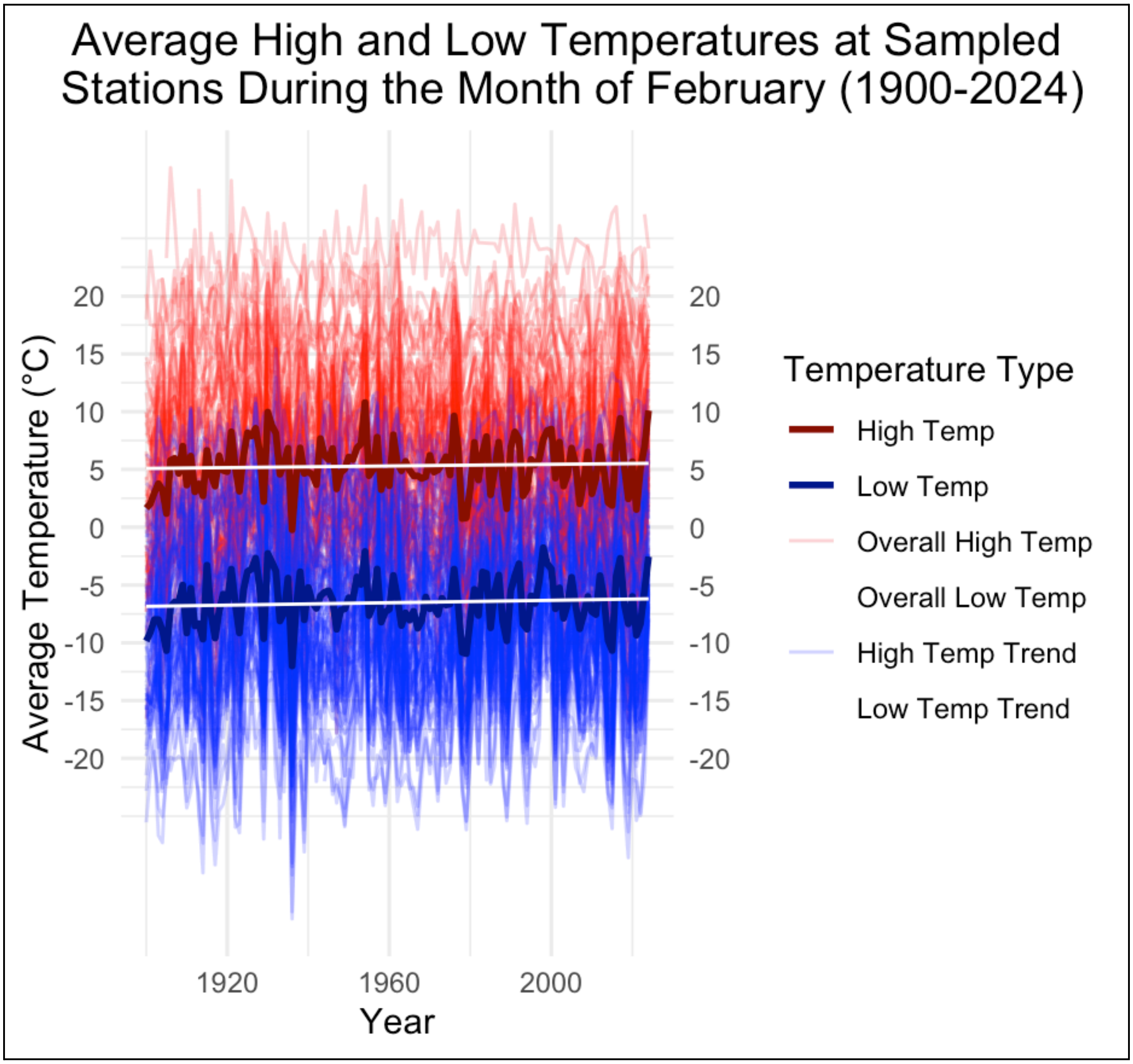


Figure 5 looks at Elevation, which is a variable which will presumably have a meaningful impact on temperature and, potentially, on climate change's impact. In this graph we didn't see a significant impact, while average daily high temperature is higher near sea level, the variation from 166 to 2400 feet in elevation doesn't appear to make a meaningful difference in general temperatures. Nor does elevation appear to have any significant impact on the impacts of climate change, none of these groups show significant temperature change of the sample period, which also means none show any unique impact of elevation.

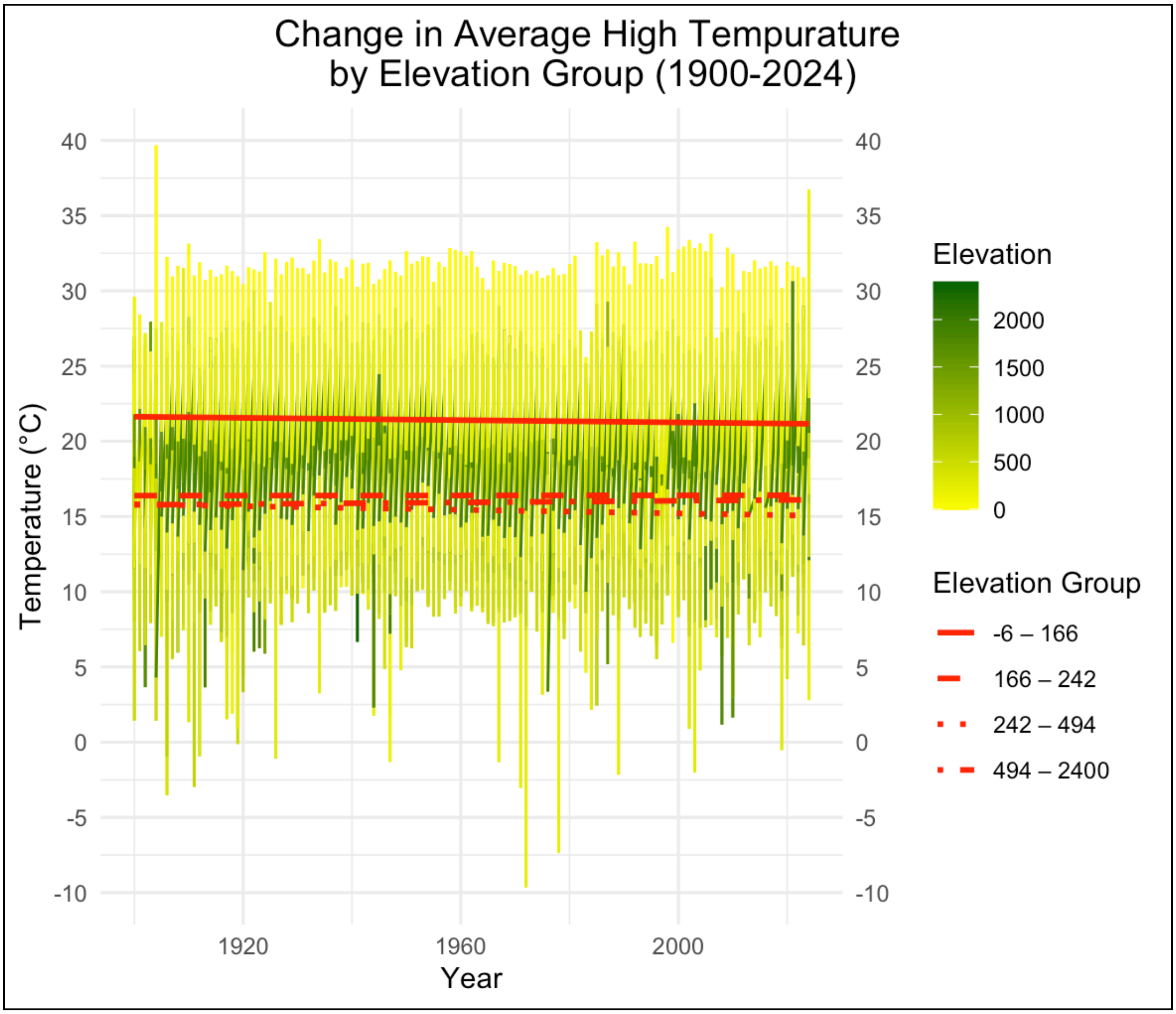


Figure 6 seeks to help show if climate change has impacted areas, at least in the United States, that are closer to the equator differently from their more northern neighbors. This figure looks at the impact that latitude has on overall change in average high temperatures since 1900. We see clear temperature variation based on latitude, as expected, following the model of lower temperatures the further from the equator. There doesn't appear to be a significant difference in overall temperature change as a result of latitude, with the exception of the lowest latitude group. Areas in the most southern part of our sample have seen a decline in average daily high temperatures over the last century.