GEOG573 DATA ANALYSIS PROJECT

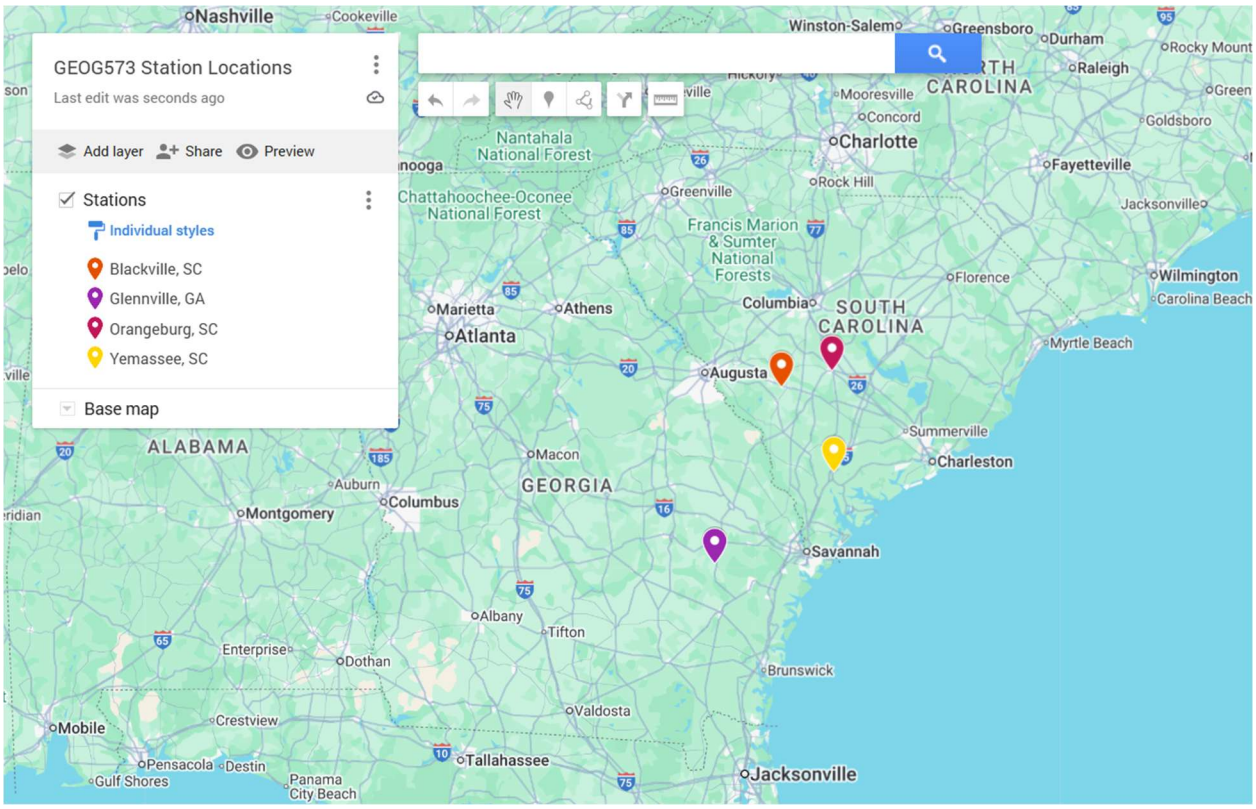
Comparing Climate Indices Across the Southeast

Matthew Botteon

December 3, 2025

Introduction

The goal of our class's collective project is to explore trends related to climatic change and variability in and around Allendale County, South Carolina. I am specifically investigating the ETCCDI Climate Change Indices, statistics that measure climate extremes related to temperature and precipitation and how they have changed over time. I have calculated the indices for four stations in the Global Historical Climate Network: Blackville, SC, Glennville, GA, Orangeburg, SC, and Yemassee, SC. These four locations are depicted below:



Interactive link:
https://www.google.com/maps/d/edit?mid=1ZUSH9Jsw0ao6LbwdBVBsse7KuJHN-NQ&usp=sharing

Each station was selected for its long record of observational data. Blackville goes back to 1895, Yemassee to 1899, Glennville to 1904, and Orangeburg to 1919. The ETCCDI climate indices were computed using the RClimDex software, an R package that provides a GUI for file upload and performs both quality control and calculates the indices. Before uploading a file, it must be formatted into a .csv file with six columns: year, month, day, prcp, tmax, and tmin. The quality control replaces any missing or unreasonable values and removes any impossible dates. The quality control returns a indcal.csv file to input to the index calculation. The calculation requires a base period, for which I used 1947 to 2024. I start at 1947 despite datasets going back

to the 1800s due to gaps in the data from the Orangeburg station. RClimDex calculates monthly indices if no more than 3 days are missing in a month, and annual indices if no more than 15 days are missing in a year. Most years have indices available, but the observations for some stations have significant gaps. This complicates long term analysis, but over the full century I was able to calculate statistics with some validity.

Of the 27 indices, I selected a few to examine in detail. Those are:

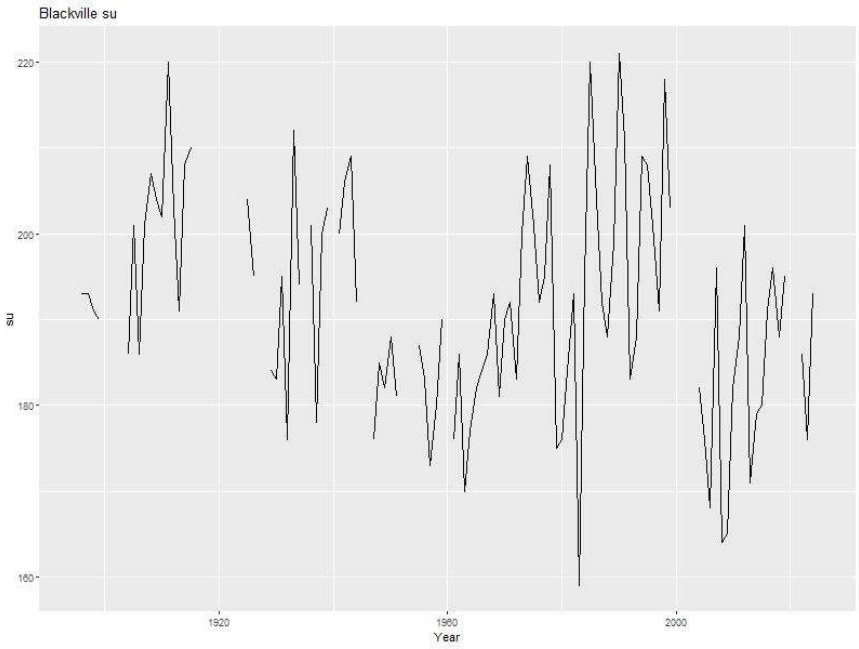
- SU, Summer Days: Annual count of days when Tmax > 25 °C
- FD, Frost Days: Annual count of days when Tmin < 0° C
- WSDI, Warm Spell Duration Index: Annual count of days with at least 6 consecutive days when Tmax > the 90^{th} percentile temperature
- TX_X, Monthly Maximum Value of daily maximum temperature

When calculating these indices, RClimDex returns a .csv file with the count for each month or year. I then used R to visualize the counts over the period of record for each station to determine if there was a long term trend. The graphs contain a value for each valid year, and a gap in the line appears for the years below the threshold that were eliminated by quality control. To help determine the trend, I also utilzed the Mann-Kendall function in the Kendall R library. This function returns a tau value and p-value for each index. The Mann-Kendall tau value measures the strength and direction of a trend on a scale of -1 to +1. This statistic is useful for climate data because it is non-parametric, meaning it does not assume a normal distribution, and it is robust to outliers. I wanted to also use Sen's slope, which has these same strengths, but it is highly affected by missing data. Sen's slope computes pairwise slopes between all available points, so missing larger amounts of data in certain timeframes will skew the calculation toward years where more data is present. In the data from these four stations, missing data tends to be clustered rather than randomly distributed, so the Sen's slope value would have been skewed. The Mann-Kendall test is more resistant to this clustering since it does not consider the magnitude of the difference between two values; it instead uses the direction of the difference of ranked values to determine a trend. Below are the graphs and Mann-Kendall tau values for each of the four stations across the previously mentioned indices, with p-value significance analysis to follow.

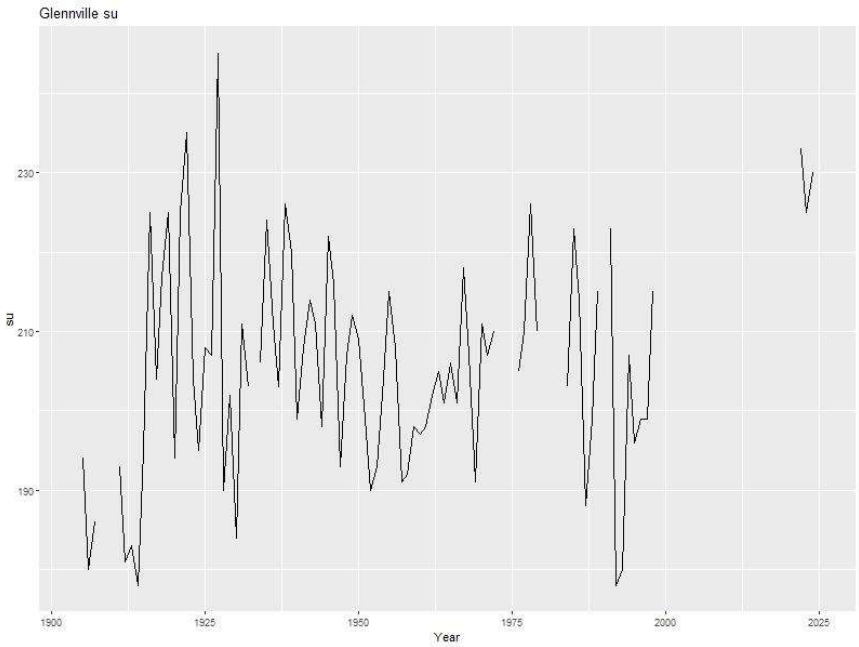
Summer Days

Summer Days, SU, tallies the number of days each year where the maximum temperature exceeds 25 °C.

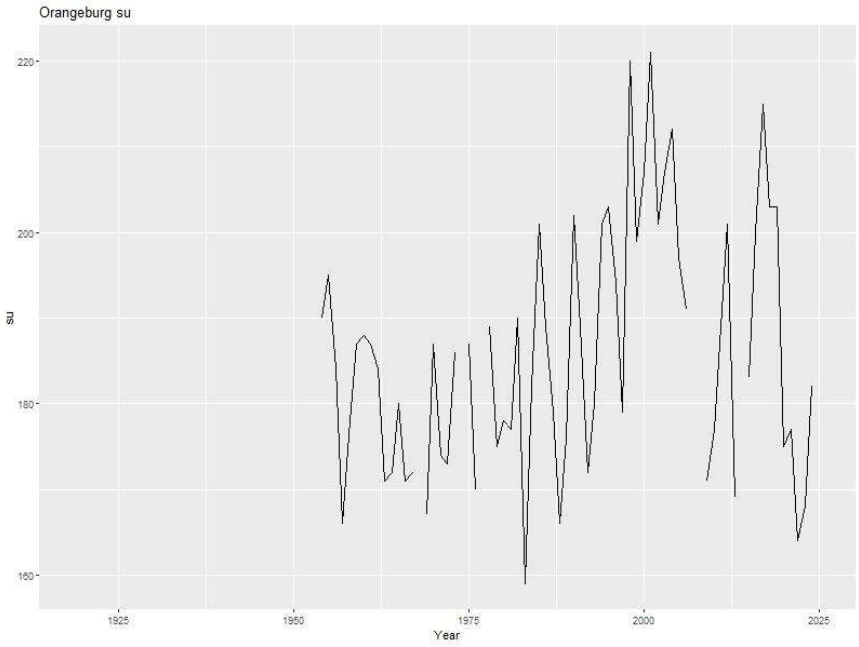
Blackville: $\tau = -0.131$



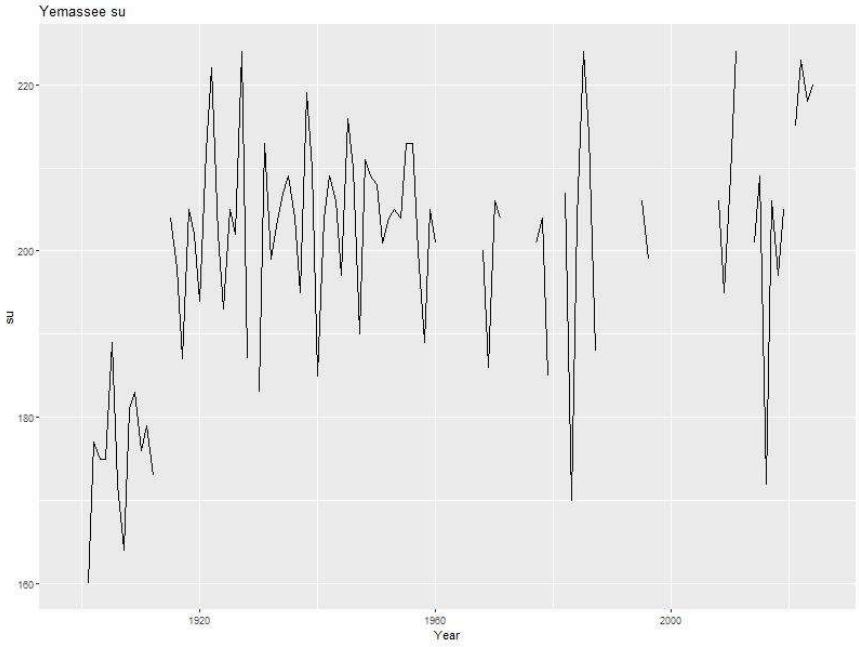
Glennville: $\tau = 0.150$



Orangeburg: $\tau = 0.176$



Yemassee: $\tau = 0.290$



Three of these stations, Glennville, Orangeburg, and Yemassee, have slight to moderate positive trends, while one, Blackville has a slight negative trend in the annual count of days over 25 °C. Using the Yemassee station as an example, since it has the most extreme trend, the Mann-Kendall tau value can be interpreted as a probability. 0.290 correlates to the difference between the probability of concordant data minus the probability of discordant data, meaning there are 29% more increasing pairs over the time series than decreasing pairs. To determine the statistical significance of these values, the Mann-Kendall function also provides a p-value. A p-value less than 0.05 means there is a statistically significant trend with 95% confidence. The p-values for the stations are:

Blackville: 0.0525, Glennville: 0.0401, Orangeburg: 0.0403, Yemassee: 0.000056

All three of the stations with a positive tau value, meaning a positive trend of increasing summer days, fall below a 0.05 p-value. This means the results are significant enough to have confidence there is a trend. Yemassee has such a small p-value that we can derive a very strong trend from the data. Blackville is the only station with a p-value above 0.05, at 0.0525. Since it is so close, this suggests there is a trend, but it is weak evidence, and we cannot be certain there is a trend of declining summer days over the past century in Blackville. It is odd that one station is an outlier when compared to the other three, especially considering their close geographic proximity, but the lack of certainty toward the opposite trend gives me some solace the data somewhat correlates. An increase in summer days per year would make sense because of the global rise in temperature that has been recorded since the beginning of the industrial era- and that is the more dominant trend here, although it is only moderate and is barely statistically significant.

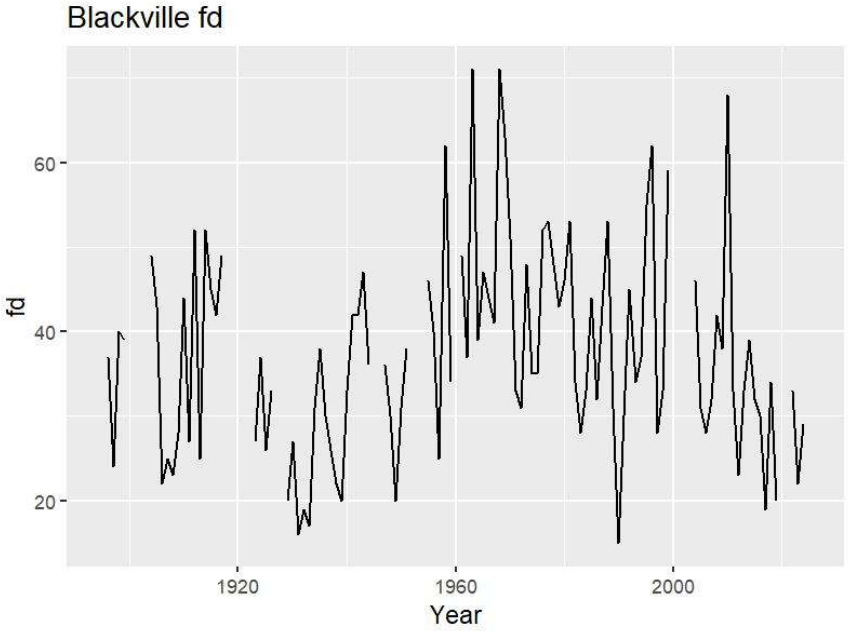
Frost Days

Another index I explored was the opposite- frost days. This measures the number of days each year where the minimum temperature fell below 0 °C, essentially counting the days it fell below freezing. The previous index looked at the peak heat, but this one explores the raising of the floor. Again, I used the RClimDex R package to calculate the index, then wrote some R code to visualize the data.

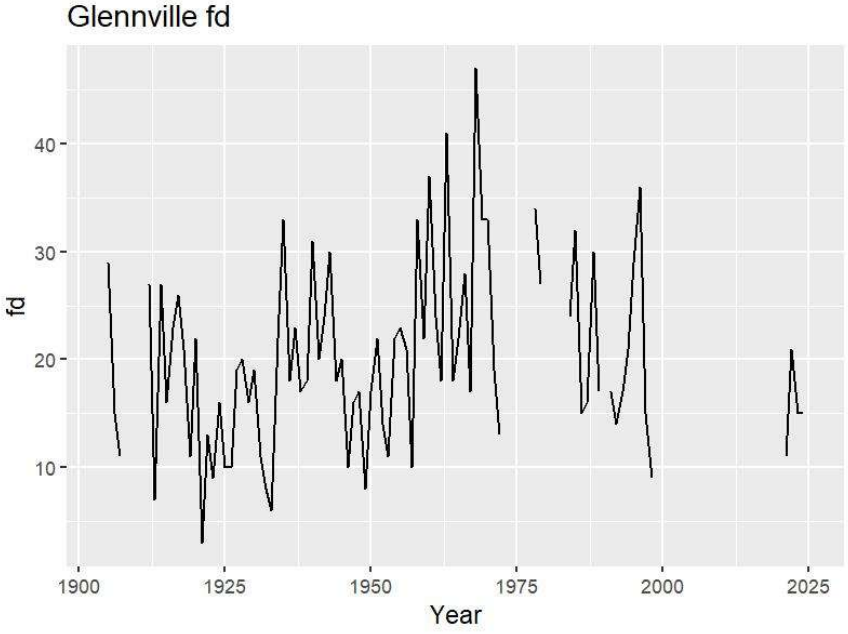
The graphs are listed below, but the first two, Blackville and Glennville, produced p-values of 0.65 and 0.12, neither of which is statistically significant. The third dataset, Orangeburg, has a p-value of 0.00013, very significant, and is the only station to suggest a negative trend in frost days. The magnitude of its tau value, -0.325, is also stronger than any other tau value we have explored, in summer or frost days. The final station, Yemassee, is also statistically significant at a p-value of 0.013 but shows a slight positive increase in frost days.

When examining these graphs, it is important to consider the NA values and where they occur for each station. For the comparison I am about to make, I will assume a compounding warming trend over time, consistent with the direction of the global record, as a hypothesis. Orangeburg shows a strong decreasing trend in frost days, supporting this hypothesis. However, when examining the data, all years from 1919 to 1953 show up as NA. This gives us a much shorter record, diminishing the confidence I have in drawing conclusions from the dataset. Excluding the early period may also leave out a more stable period of time, leaving a more drastic period to contribute to the strong results found in the Mann Kendall test. On the opposite side, Yemassee suggested a positive trend in frost days. When examining this data, 1999 to 2007 and 2012 to 2016, among other scattered recent years, have NA values. Operating under the hypothesis, these years should be warmer. Leaving them out of the calculation may skew the results for this station toward a constant or even a cooling trend. To check the quality control mechanism of RClimDex, I chose 2001 from Yemassee at random, and counted 16 missing days for Tmin, right above the threshold to disqualify that year. Gaps in the data, especially in recent years, skew statistics and make index comparisons between stations difficult.

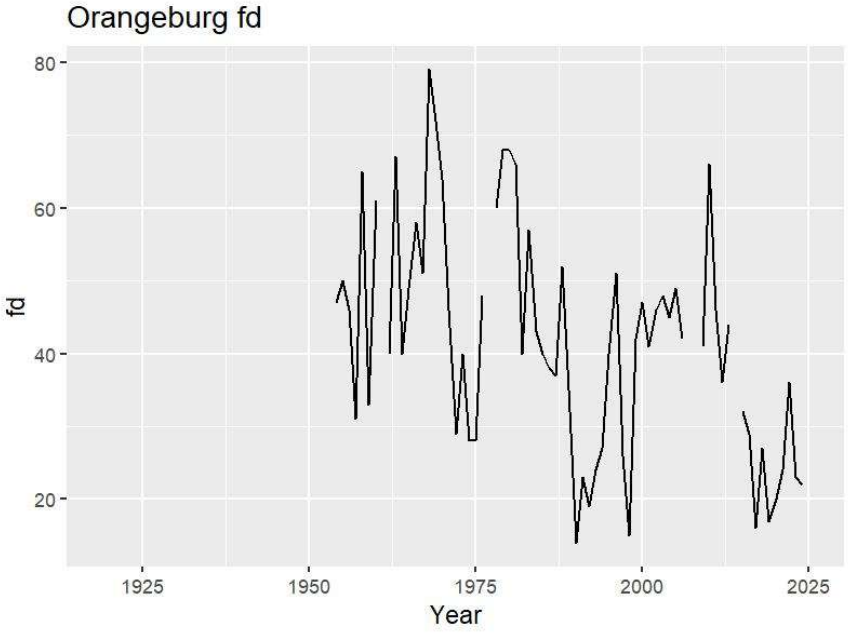
Blackville: $\tau = 0.030$



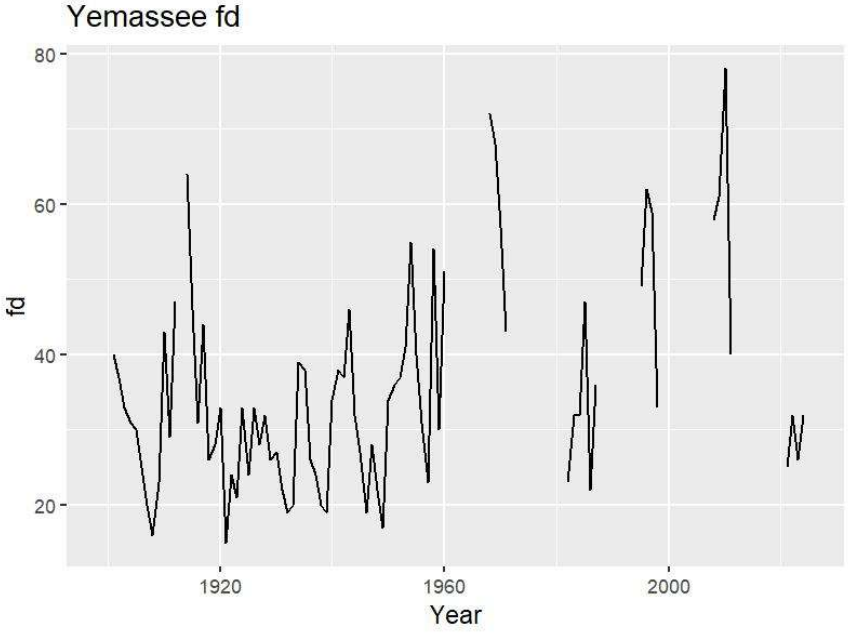
Glennville: $\tau = 0.115$



Orangeburg: $\tau = -0.325$



Yemassee: $\tau = 0.183$

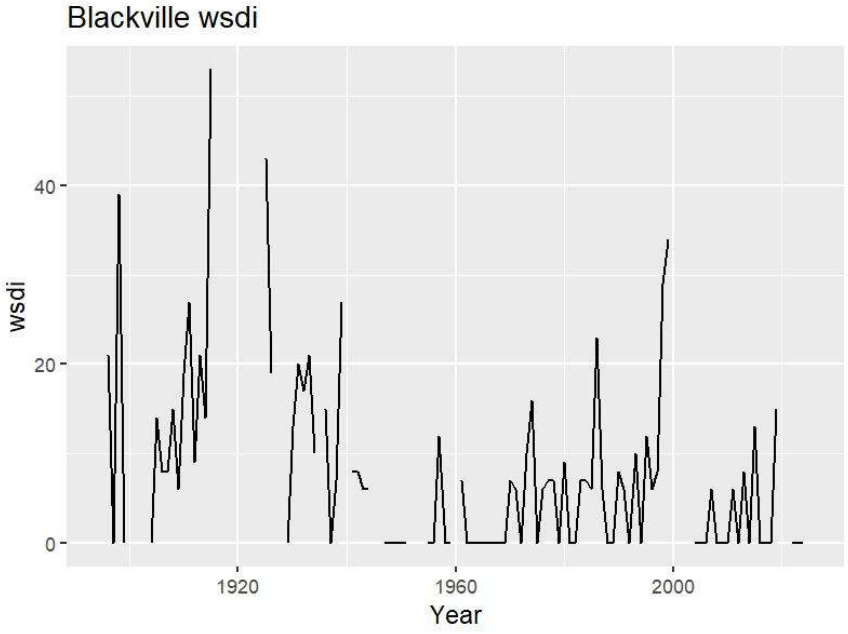


Warm Spell Duration Index

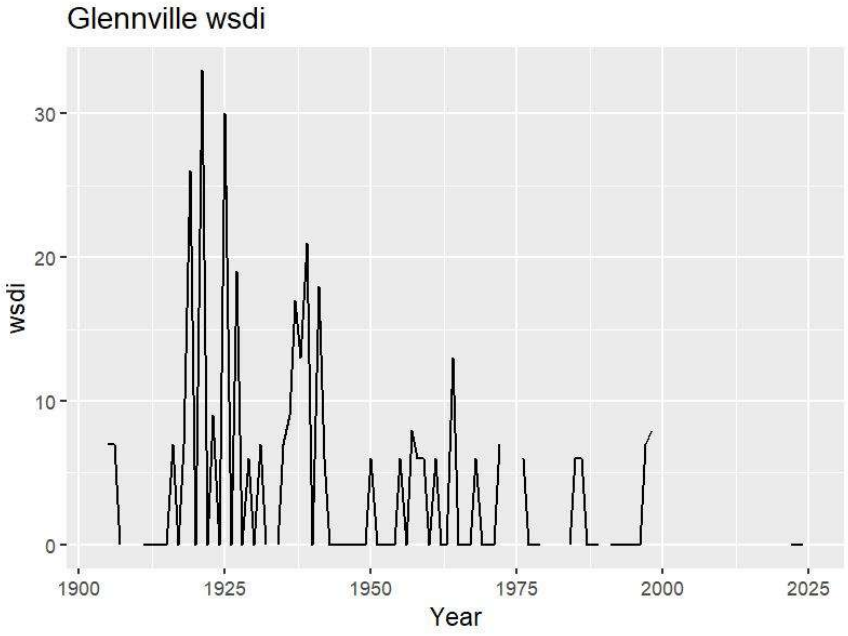
The Warm Spell Duration Index measures “heat waves” by producing an annual count of days with at least 6 consecutive days when Tmax is greater than the 90^{th} percentile temperature. More detail on the calculation of the 90^{th} percentile temperature can be found in the RClimDex manual, which is added to the repository linked at the end of this document. This index attempts to quantify the change in quantity in longer stretches of heat over time.

Below are the graphs and tau values. Only Blackville (0.00012) and Yemassee (0.04) are statistically significant, while Glennville (0.10) and Orangeburg (0.19) are not. Visually, the NA values in these two stations are very apparent- Glennville has almost no data after 2000 and Orangeburg has none before 1950. This gives the two stations only about 50 years of overlap and again makes long term trend comparison difficult and likely contributed to their results not being significant. As far as Blackville and Yemassee, the two significant results, their tau values have opposite directions. Blackville has a moderately negative trend of -0.276 and Yemassee a slightly positive trend of 0.166. An increase in temperature should theoretically lead to an increase in warm spells, since higher heat is more common. However, the Blackville has a much more significant p-value and its tau value has a higher magnitude even though it points to a decrease in warm spells. Yemassee does show an increase even though it too has spotty data after 2000, although the Mann Kendall calculation still has some data towards the present day to use when ranking the data. I selected warm spell thinking it would show the warming effect of temperature in another way besides summer days, but I cannot draw a confident conclusion from this data.

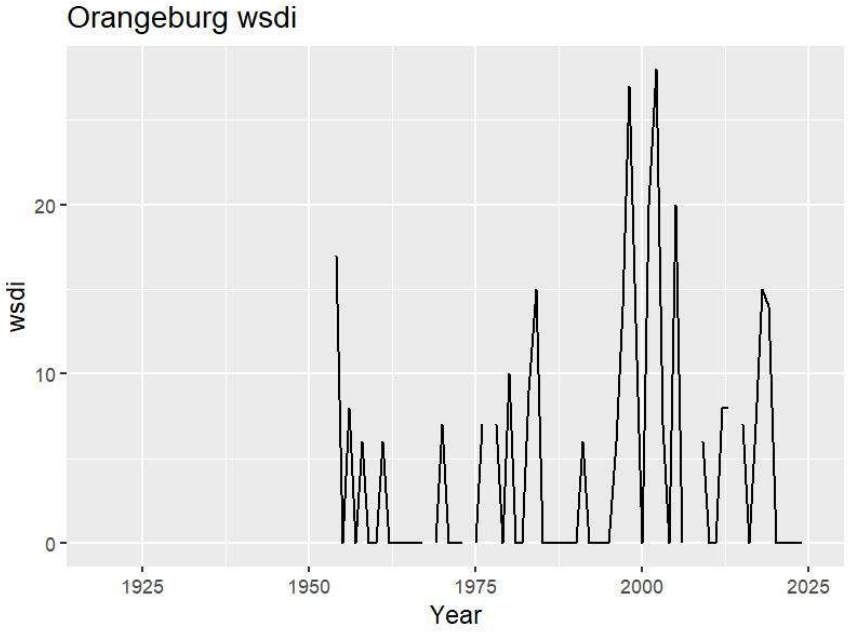
Blackville: $\tau = -0.276$



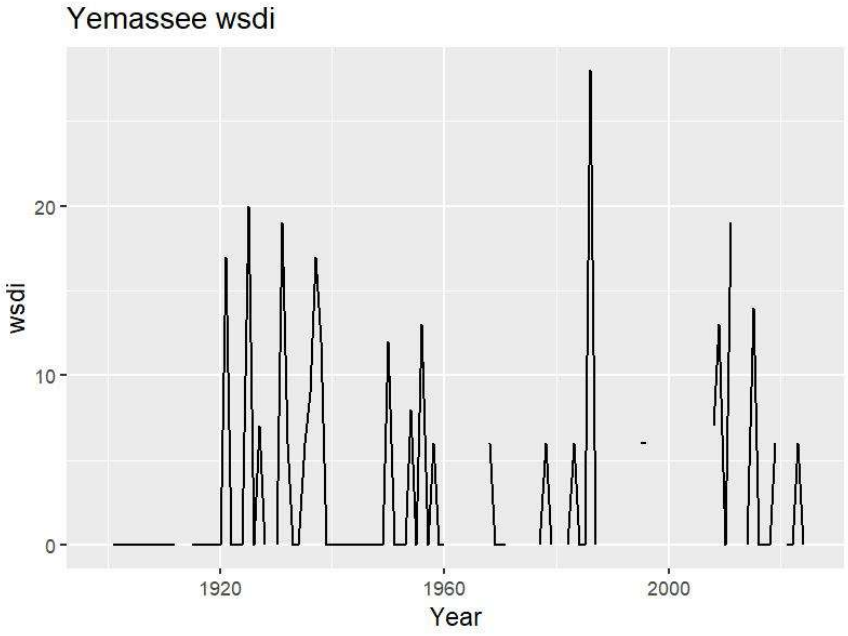
Glennville: $\tau = -0.133$



Orangeburg: $\tau = 0.124$



Yemassee: $\tau = 0.166$

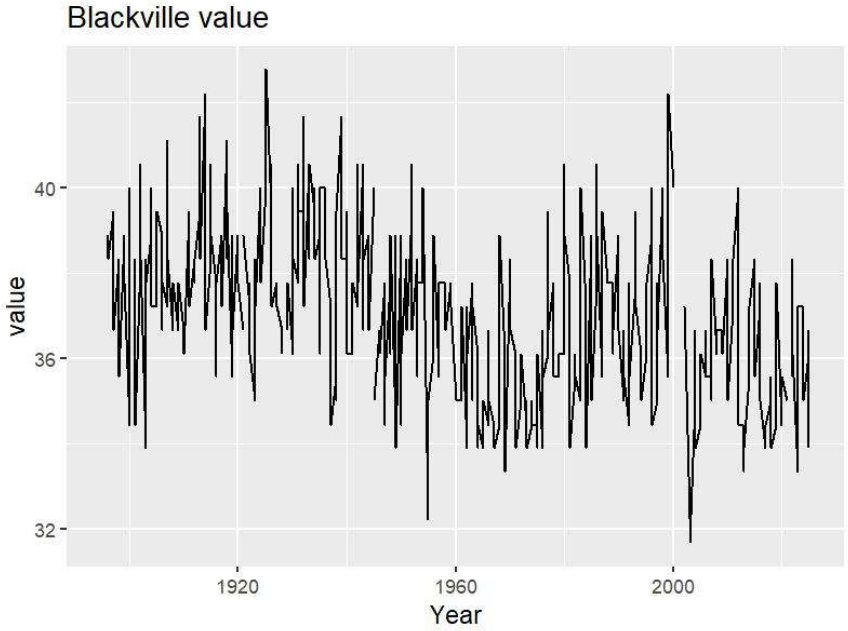


Monthly Maximum Value

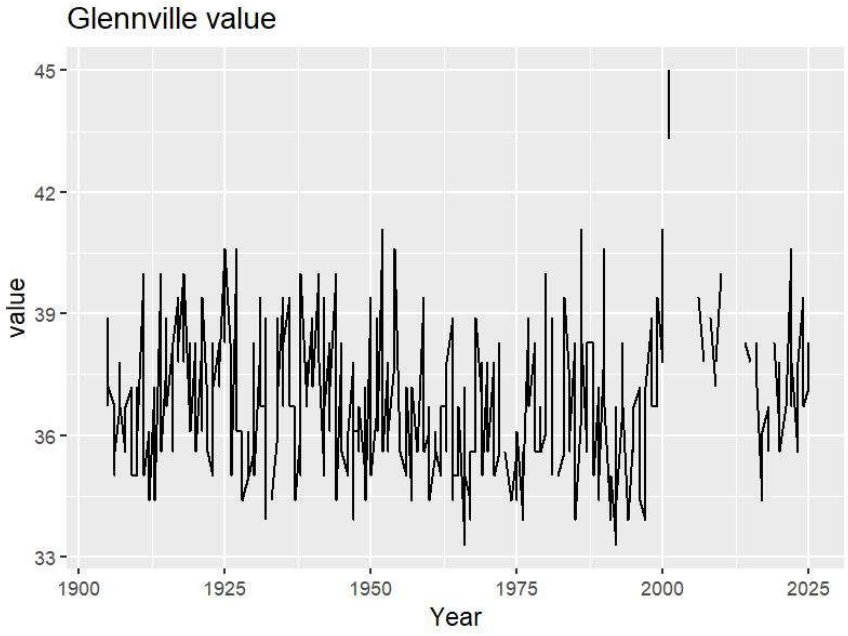
To continue exploring the temperature trend, I decided to look to a monthly measure. Since the threshold is 3 days per month compared to 15 days per year before a unit is thrown out, the more precise monthly data is less prone to exclusions due to NA values. Monthly Maximum Value is the peak Tmax in a given month, aiming to capture the extremes that have been reached rather than counting the number of times a predetermined threshold is surpassed, like in the Summer Days index. For this analysis, I did both a full sequence, to see if there was a slight trend even when going through each season, and then an isolated June, July, and August sequence to eliminate any winter bias. Only the JJA graphs are pictured.

As expected, the full year sequences do not reveal any trend, with no tau value exceeding a magnitude of 0.1 in either direction. For the JJA sequences, there is not much trend either. I figured this could be because of slight seasonal differences, so I isolated August for Blackville since it was the only station with a significant p-value. This produced a tau value of -0.292 with a p-value of 0.000002. This negative trend suggests a decline in temperature over time, inconsistent with the general global trend.

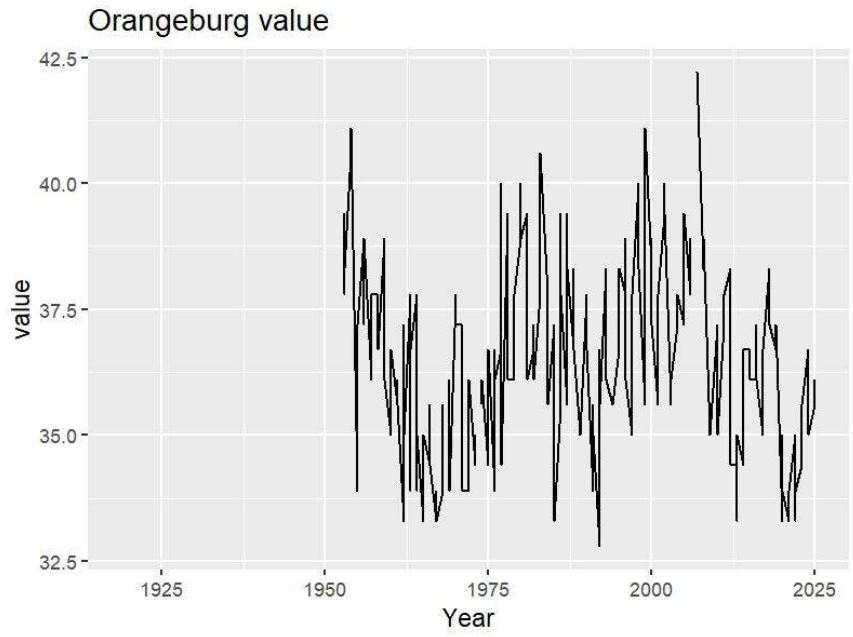
Blackville: All τ = -0.073, JJA τ = -0.277



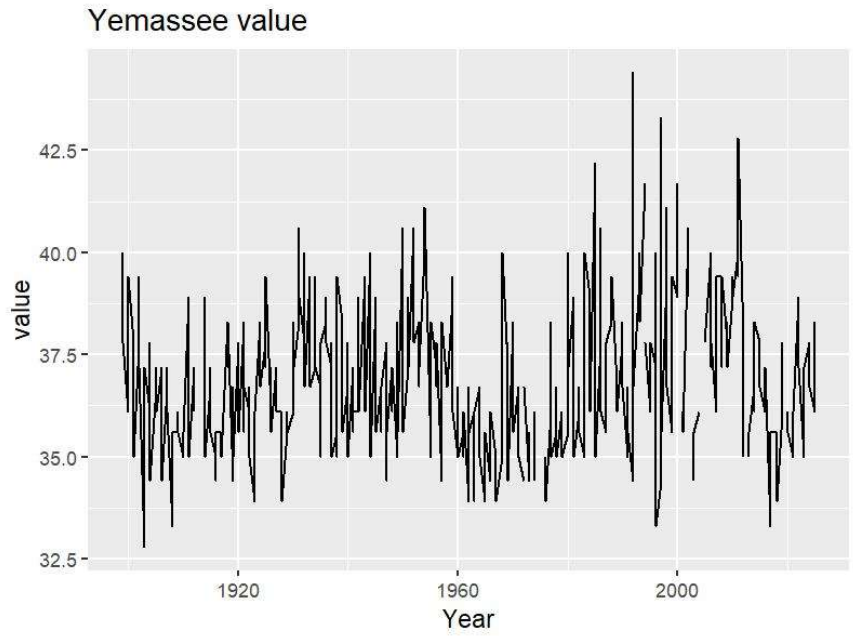
Glennville: All τ = -0.011, JJA τ = -0.007



Orangeburg: All $\tau = 0.016$, JJA $\tau = -0.026$



Yemassee: All $\tau = 0.029$, JJA $\tau = 0.055$



Conclusion

After exploring four climate indices, summer days, frost days, warm spell duration index, and monthly maximum temperature, I am unable to make a confident conclusion on the trend of temperature in this region. I faced difficulty with incomplete data causing insignificant results, and more importantly, this missing data coming during different time periods at each station. I believe these indices hold weight and perhaps a follow-up study could be done with more isolated timeframes and the exploration of new indices. The most complete index in my study was Summer Days, with three of the four stations showing significance and all agreeing on a positive trend. If I were to build a conclusion Yemassee was the most complete station, showing significance in all measures besides the monthly maximum. However, it showed a positive trend in both Summer and Frost Days, two conflicting measures. All of the code and data used in this project are in my repository, linked below, and I look forward to continuing to explore this topic.

Repository: https://github.com/m-botteon/climate_indices