Abstract

Climate change poses significant challenges globally, including rising temperatures with far-reaching consequences. This study examines temperature trends in Oklahoma, focusing on summers since 2000 using data from the ACME Mesonet station in Grady County. Analysis of daily temperature averages, maximums, and minimums for July reveals a consistent, albeit slight, increase in temperatures over the 23-year period. While yearly increases are minimal, comparing temperatures in 2000 and 2023 shows a significant rise. These findings underscore the importance of long-term trend analysis for understanding climate change impacts. The study highlights the need for continued monitoring and analysis of temperature trends, suggesting a potential shift towards hotter summers in Oklahoma. Future research could expand the dataset to include a longer timeframe or analyze other months and incorporate precipitation data for a more comprehensive understanding of climate trends. Understanding these trends is crucial for informing climate change mitigation and adaptation strategies in the region.

Introduction

Climate change is a pressing global issue with far-reaching impacts, including shifts in temperature patterns that can have profound consequences. One area of concern is the increasing temperatures experienced during summer months, which can have significant implications for various aspects of human life, including water resources, natural systems, and human health. In this study, we aim to investigate whether summers in Oklahoma have experienced a noticeable increase in temperature since 2000, focusing specifically on data from the ACME Mesonet station in Grady County. For this study, we are only looking at data from the month of July.

The state of Oklahoma, located in the central United States, is no stranger to the effects of climate change. Over the past century, Oklahoma has experienced a noticeable increase in temperature trends, with some of the warmest years on record occurring in recent decades. The past few decades have seen a notable rise in global temperatures, where ""the five warmest years over the last century have likely been: 2005, 1998, 2002, 2003, and 2006. The top 10 warmest years have all occurred since 1990" (EPA, 2008). This trend is expected to continue, leading to longer and hotter summers in many regions, including Oklahoma.

Climate change will have numerous and diverse impacts, including impacts on human health, natural systems, and the built environment. Many of the consequences of climate change relate to water resources, including: warming air and water; change in the location and amount of rain and snow; increased storm intensity; sea level rise; and changes in ocean characteristics (EPA, 2008). Through the future decades in general, Oklahoma will become hotter, and droughts and floods have the potential to rise in severity and occurrence. Under a changing climate, Oklahoma's water sector will deal with longer periods of drought, flooding events, severe weather events, fire weather, heat waves, and more, as each incident has the potential to stress the water industry (Warner, 2023).

One of the critical consequences of increasing temperatures in Oklahoma is the impact on water resources. As temperatures continue to rise, "surface water evaporates at a quicker rate, depleting the amount of water available" leading to water scarcity and stress on water infrastructure (Warner 2023). The water sector in Oklahoma is particularly vulnerable to these changes, as longer periods of drought, flooding events, and severe weather are projected to become more frequent and severe.

In addition to water resources, increasing temperatures also pose risks to other aspects of the environment, such as the increased potential for forest fires. With higher temperatures and drier surface conditions predicted for the southeastern US, including Oklahoma, we may see an "increase low-level lapse rates and near-surface soil drying in the south-eastern US, leading to increases in future KBDI and HI, and associated potential for fires respectively to ignite and spread" (Bedel 2015).

Given the significant impacts of increasing temperatures on various aspects of life in Oklahoma, it is crucial to understand the trends in temperature change over time. This project aims to analyze temperature data from the ACME Mesonet station in Oklahoma to determine if summers in Oklahoma have indeed become hotter since 2000. By focusing on the month of July, which represents the middle of the summer season, this study will provide valuable insights into the changing climate of Oklahoma and its potential implications for the future.

Data

In order to obtain my dataset I used the past data request form from the Oklahoma Mesonet website to download daily temperature maximum, minimum, and averages (TMAX, TMIN, TAVG) in Fahrenheit from the ACME station in Grady County for the period from July 2000 to July 2023. Although the dataset contained data for every day, I only required daily data for July of each year. Therefore, I filtered the dataset to include only the daily TAVG, TMAX, and TMIN for every July from 2000 to 2023.

Throughout the years, there were a few instances where temperature readings were missing, possibly due to faulty or broken equipment. These missing values were indicated by -999,000, which is a significant and extreme outlier. To prevent these values from affecting statistical analysis, particularly those involving the mean, I filtered the data to exclude any values below 0 or above 150, as these would be errors in data collection.

The original dataset did not include qualitative data, which is necessary for the Chi-squared test. To address this, I categorized the TMAX, TMIN, and TAVG values into the categories 'High,' 'Medium,' and 'Average.' Values within one standard deviation of the mean were categorized as 'Average,' while those more than one standard deviation above or below the mean were categorized as 'High' and 'Low,' respectively. For specific categorization values, refer to Table 14 in the results/analysis section.

Results and Analysis

Data Visualization

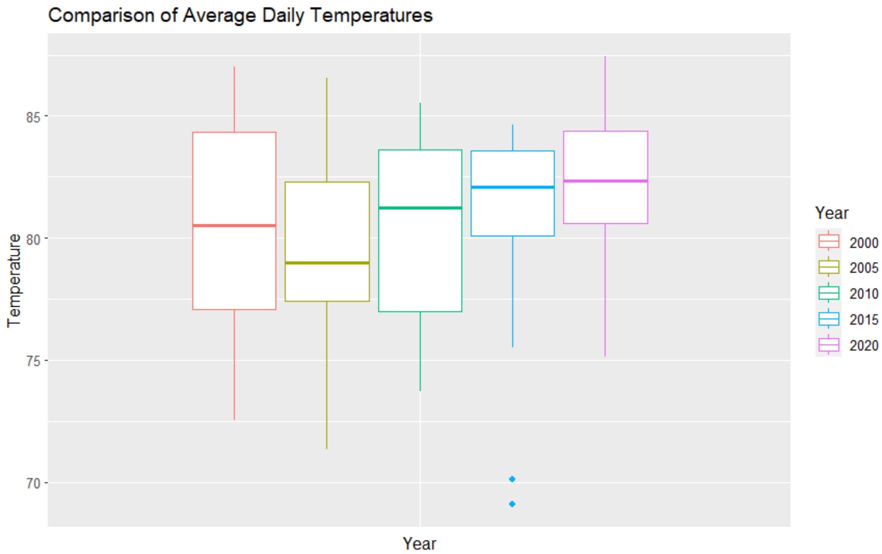

Figure 1: Average Daily Temperatures Every 5 Years

Figures 1, 2, and 3 display boxplots for daily temperature averages, maximums, and minimums for July every five years, beginning with the year 2000. Figures 1, 2, and 3 show temperature data every five years rather than every other year for the sake of simplicity and to show the general trend of averages over the years.Years 2000 and 2020 are fairly symmetrical, but year 2005 has a median closer to the minimum and years 2010 and 2015 have medians closer to the maximum, despite 2015 having two outliers below the minimum.

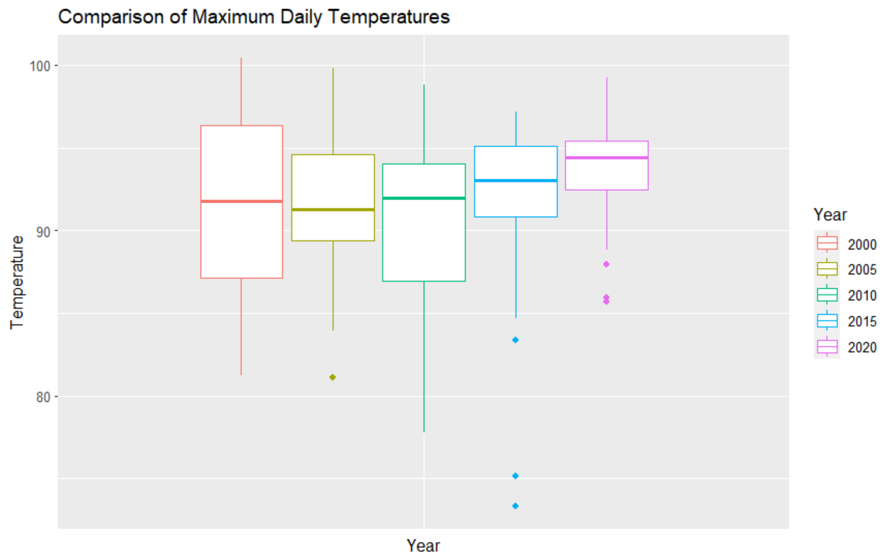

Figure 2: Maximum Daily Temperatures Every 5 Years

The year 2000 has a symmetrical distribution for maximum temperature as well. Year 2005 has a median closer to the minimum, as well as a low outlier. Years 2010, 2015, and 2020 contain medians closer to the maximum, and 2015 and 2020 both have three low outliers.

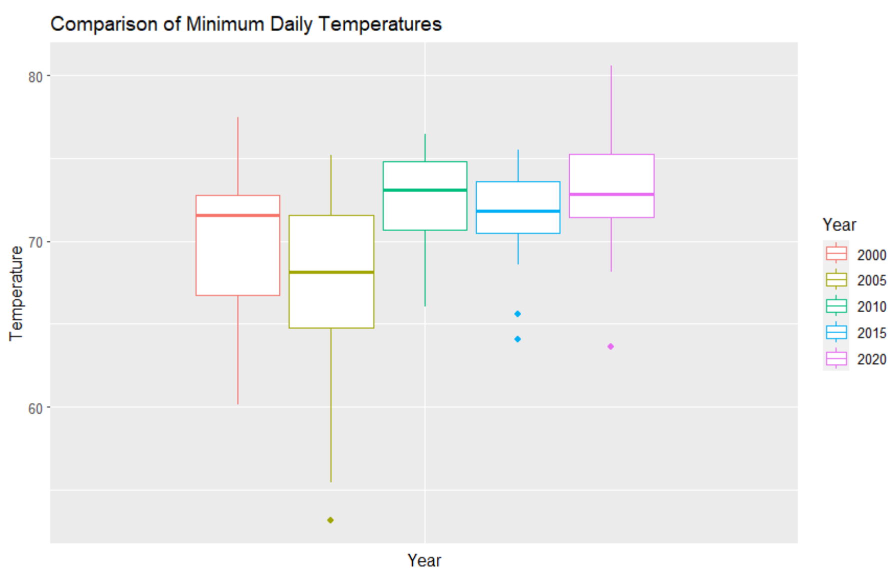

Figure 3: Minimum Daily Temperature Every 5 Years

All years displayed in Figure 3 show asymmetrical distributions. Years 2000, 2010, and 2015 contain medians closer to the maximum, while 2005 and 2020 contain means closer to the minimum. 2005 and 2020 have one low outlier, and 2015 has two low outliers.

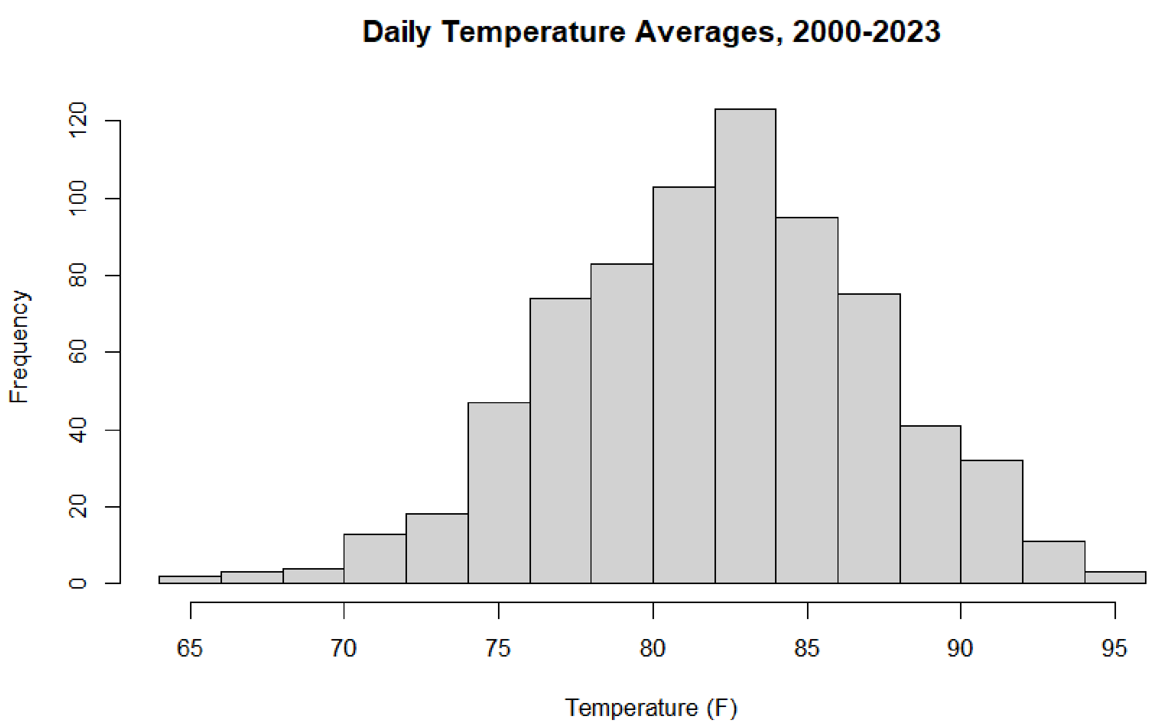

Figure 4: Histogram of Daily Temperature Averages 2000-2023

Figure 4 shows that overall, temperature averages for July over all years have a slight left skew but have a fairly normal distribution. Therefore, we will use mean for central tendency analysis:

N	Mean	SD	Median	Trimmed	Minimum	Maximum	Range	Skew	Kurtosis
727	82.05	5.11	82.18	82.1	65.31	94.45	29.14	-0.18	0.03

Table 1: Central Tendency for Temperature Averages

The mean and median are significantly close in value, further proving that temperature averages over the years have a normal distribution. There is a small negative skew indicating the distribution has a slight left tail, but it is small enough to be considered insignificant. The kurtosis is significantly small as well, indicating a peak and tails comparable to a normal distribution.

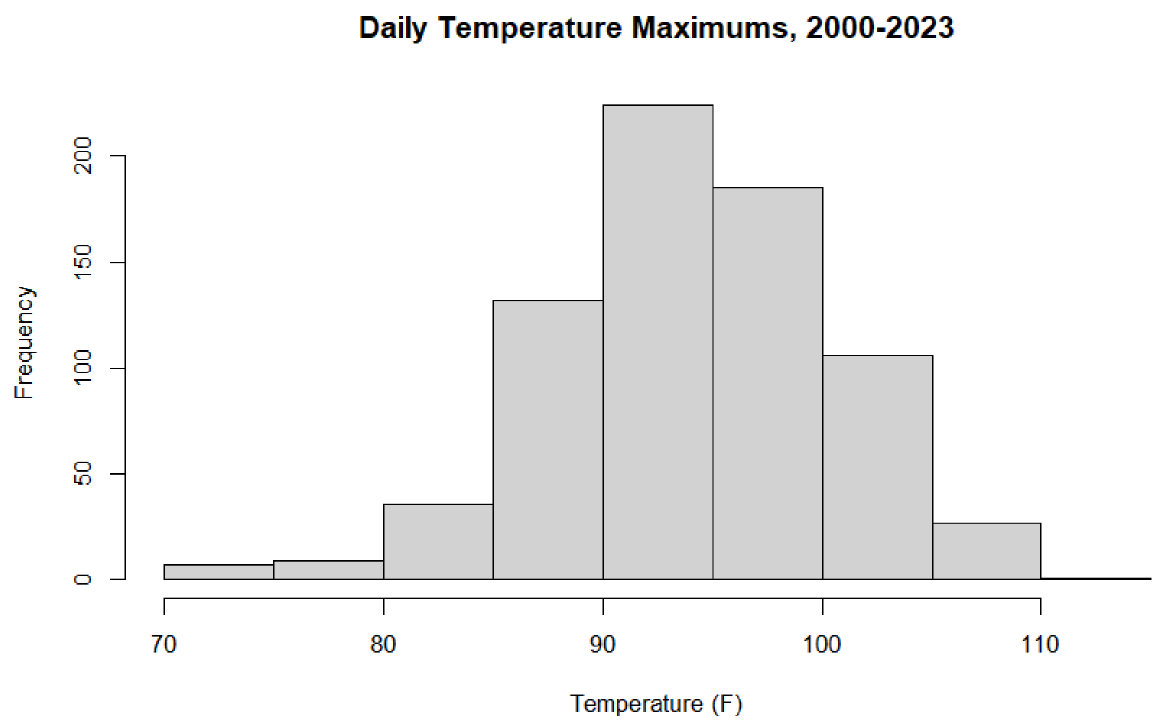

Figure 5: Histogram of Daily Temperature Maximums 2000-2023

Figure 5 shows that overall, temperature maximums for July over all years have a slight left skew but have a fairly normal distribution. Therefore, we will use mean for central tendency analysis:

N	Mean	SD	Median	Trimmed	Minimum	Maximum	Range	Skew	Kurtosis
727	94	6.45	93.99	94.14	70.05	110.3	40.25	-0.33	0.52

Table 2: Central Tendency for Temperature Maximums

Table 2 shows the mean and median are significantly close in value for maximums as well. The skew and kurtosis are slightly larger in magnitude than the skew and kurtosis for temperature averages. The histogram, skew, and kurtosis all indicate a slight left tail, but the skew and kurtosis values are small enough to be insignificant and this can be considered a normal distribution.

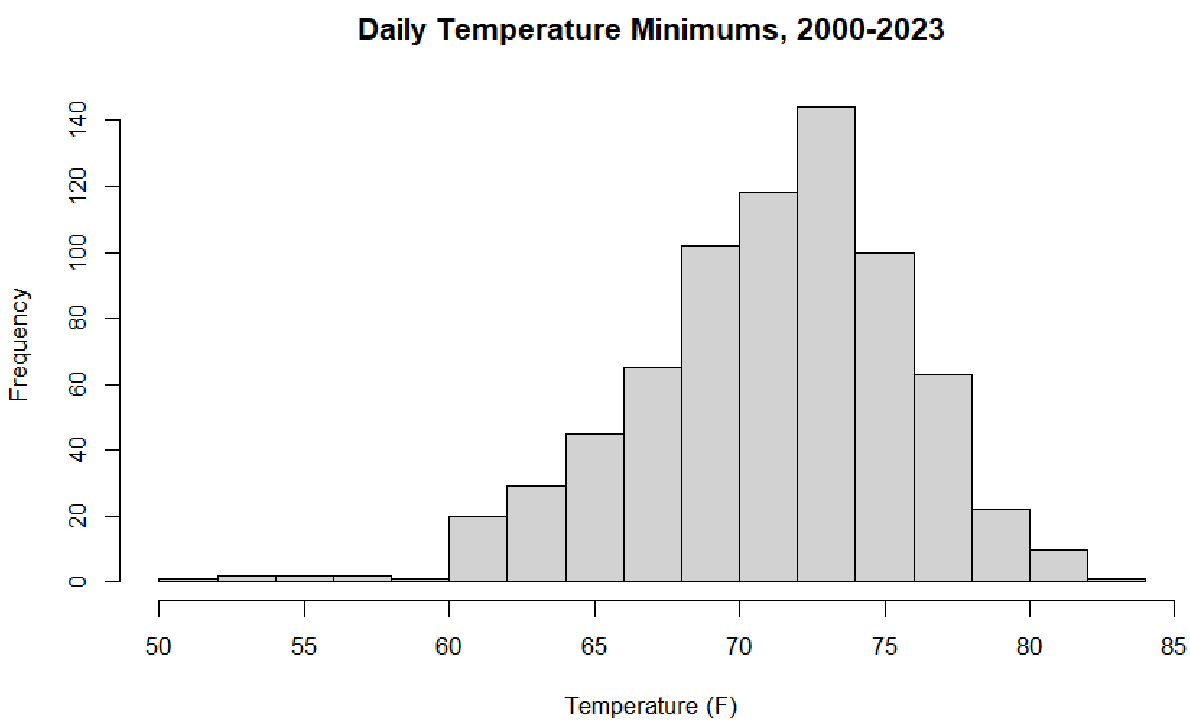

Figure 6: Histogram of Daily Temperature Minimums 2000-2023

Figure 6 shows that overall, temperature maximums for July over all years have a slight left skew but have a fairly normal distribution. Therefore, we will use mean for central tendency analysis:

N	Mean	SD	Median	Trimmed	Minimum	Maximum	Range	Skew	Kurtosis
727	71.07	4.63	71.6	71.3	51.67	82.36	30.69	-0.59	0.77

Table 3: Central Tendency for Temperature Minimums

Similar to temperature maximums, the skew and kurtosis for temperature minimums are slightly larger in magnitude than for temperature averages and indicate a slight left tail. However, because the mean and median are significantly close in value and the skew and kurtosis values are significantly small this is considered a normal distribution.

Correlation and Regression

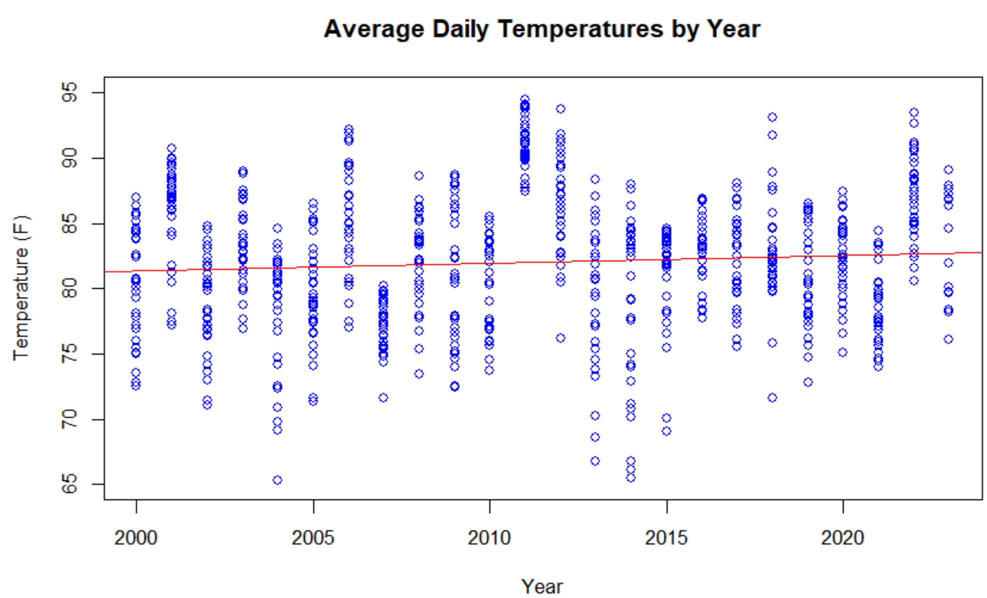

Figure 7: Scatterplot of July Daily Average Temperatures by Year

Table 4 shows that TAVG and Year are only slightly positively correlated, with a correlation coefficient of 0.077. This is very close to 0, and could be considered not correlated. The best-fit line shown in Figure 7 has a y-intercept of -34.1459 and a slope of 0.0578. The linear equation for this is as follows:

$$\boldsymbol{y_{TAVG} = 0.0578 * x_{YEAR} - 34.1459}$$

In other words, every year the temperature averages increase by 0.0578 degrees.

	Y-Intercept	Slope	Error	T Value	Pr (>| t |)	Correlation Coefficient
TAVG			0.0279	2.07	0.039	0.077
Best-Fit Line	-34.1459	0.0578	56.1173	-0.61	0.543	
Multiple R-Squared	0.00588					

Table 4: Coefficients for Average Temperatures

The R-squared value is 0.00588, meaning that 0.6% of the variation in temperature averages can be explained by a linear relationship to the year.

Table 5 gives the y-intercept and slope for the best-fit line shown in Figure 8. The linear equation for the best-fit line for TMAX is:

$$\boldsymbol{y_{TMAX} = 0.0628 * x_{YEAR} - 32.3848}$$

Where 0.0628 is the slope and -32.3848 is the y-intercept. In other words, for every year the temperature maximums have gone up 0.0628 degrees.

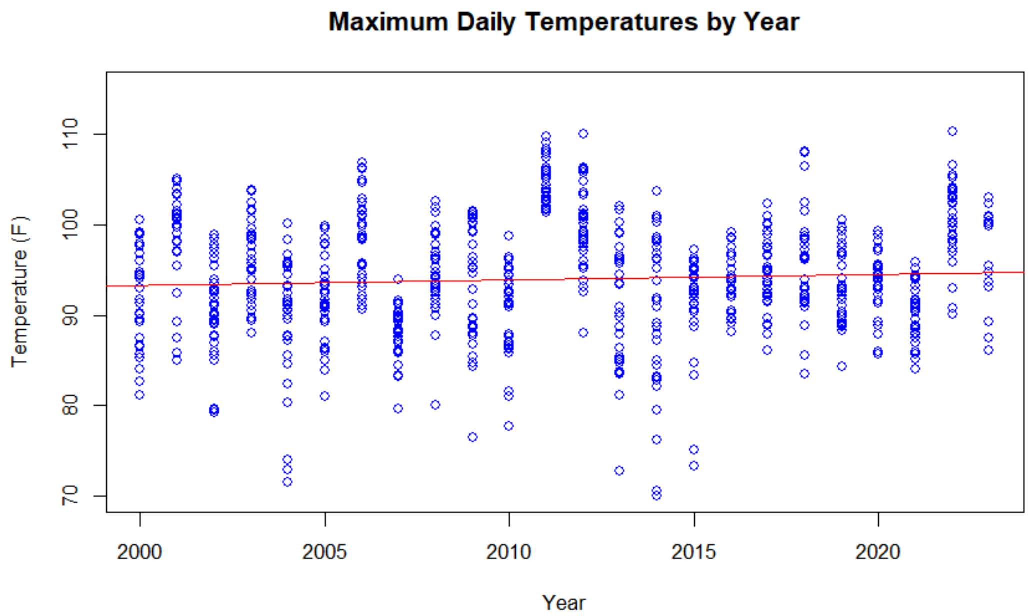

Figure 8: Scatterplot of July Daily Maximum Temperatures by Year

	Y-Intercept	Slope	Error	T Value	Pr (>| t |)	Correlation Coefficient
TMAX			0.0352	1.78	0.075	0.066
Best-Fit Line	-32.3848	0.0628	70.8483	-0.46	0.648	
Multiple R-Squared	0.00437					

Table 5: Coefficients for Maximum Temperatures

The R-squared value for TMAX is 0.00437, meaning that 0.44% of the variation in temperature maximums can be explained by a linear relationship with the year. A correlation coefficient of 0.066 is very small, but shows a slight positive correlation. Being so close to 0 in magnitude, though, means that this correlation may be insignificant.

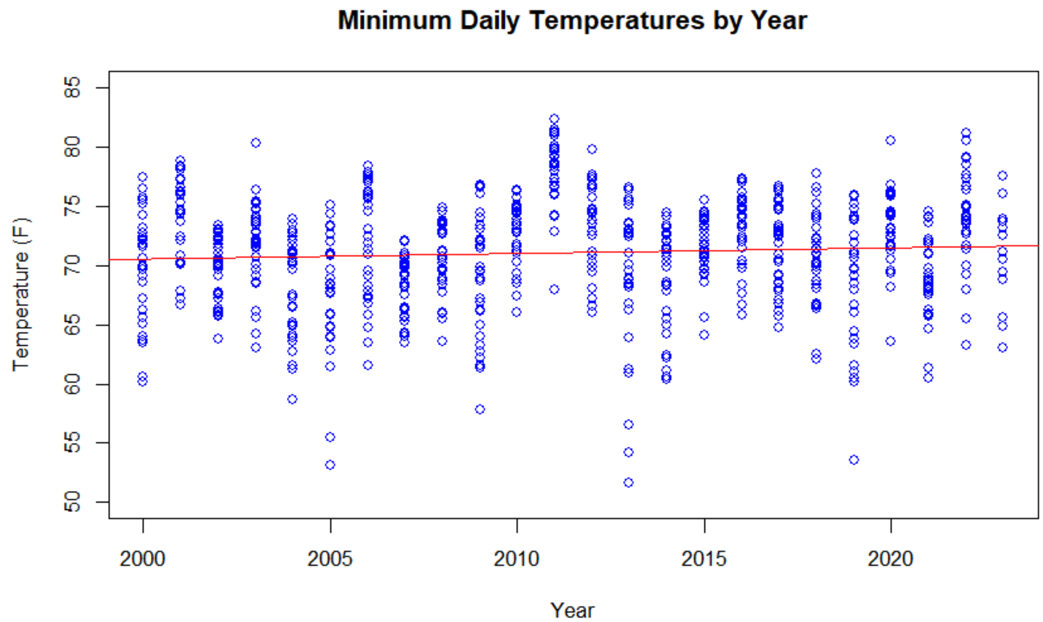

Figure 9: Scatterplot of July Daily Minimums by Year

The line of best fit in Figure 9 is described by the values in Table 6. With a slope of 0.0494 and a y-intercept of -28.3611, the linear equation is:

$$\boldsymbol{y_{TMIN} = 0.0628 * x_{YEAR} - 32.3848}$$

Meaning that for every year, the temperature minimums have gone up 0.0628 degrees.

	Y-Intercept	**Slope**	**Error**	**T Value**	**Pr (>| t |)**	**Correlation Coefficient**
TMIN			0.0253	1.95	0.051	0.072
Best-Fit Line	-28.3611	0.0494	50.8766	-0.56	0.577	
Multiple R-Squared	0.00524					

Table 6: Coefficients for Minimum Temperatures

The correlation coefficient here is 0.072, which is also close to 0, although it shows a slight positive correlation. However, because of the magnitude of the correlation coefficient, it could be considered insignificant. The multiple R-squared value is 0.00524, meaning that 0.52% of the variance can be explained by a linear relationship to the year.

Confidence Intervals (95%)

The confidence interval for TAVG was calculated using the mean and CI Mean (0.95) from Table 7, where CI Mean (0.95) is the margin of error at 95% confidence.

Median	Mean	SE Mean	CI Mean (0.95)	Var.	SD	Coef. Var.
82.1800	82.0453	0.1896	0.3723	26.1412	5.1128	0.0623

Table 7: Values used to calculate the confidence interval for temperature averages

The confidence interval for TAVG is:

$$\boldsymbol{81.6730 < \mu < 82.4176}$$

Where μ symbolizes the mean. In other words, we are 95% confident that the mean temperature for TAVG from 2000-2023 is between 81.6730 and 82.4176 degrees.

Median	Mean	SE Mean	CI Mean (0.95)	Var.	SD	Coef. Var.
93.9900	94.0033	0.2392	0.4696	41.6037	6.4501	0.0686

Table 8: Values used to calculate the confidence interval for temperature maximums

The confidence interval for TMAX is:

$$\boldsymbol{93.5334 < \mu < 94.4729}$$

This was calculated using the values shown in Table 8. We can be 95% sure that the mean temperature for TMAX from 2000-2023 is between 93.5334 and 94.4729 degrees.

Median	Mean	SE Mean	CI Mean 0.95	Var.	SD	Coef. Var.
71.6000	71.0692	0.1719	0.3374	21.4728	4.6339	0.0652

Table 9: Values used to calculate the confidence interval for temperature minimums

The confidence interval for TMIN is:

$$70.3178 < \mu < 71.4065$$

This was calculated using values shown in Table 9. We can be 95% sure that the mean temperature for TMIN from 2000-2023 is between 70.3178 and 71.4065 degrees.

One-group T-Test

This t-test was conducted using TAVG, from 2000-2023. I wanted to see if averaging the means of TMAX and TMIN would accurately represent daily average temperature- in other words, if we could determine TAVG values based on TMAX and TMIN alone. Using the values for mean from Tables 8 and 9, a μ of 82.5 was used for this t-test. The null hypothesis is that the population mean for TAVG is greater than or equal to 82.5, and the alternative hypothesis is that the population mean is less than 82.5.

t	Degrees of Freedom	p-value	Mean
-2.3979	726	0.00837	82.0453

Table 10: T-test values for TAVG

Table 10 shows a sample mean temperature of 82.0453 degrees, which is 2.3979 standard deviations lower than 82.5. Because the p-value is less than 0.05, we will reject the null hypothesis and conclude that the population mean of TAVG is less than 82.5. The average of the means of TMIN and TMAX are not a good representation of TAVG, likely due to low outliers in TMAX and TMIN.

Two-group t-test

The intention with the two-group t-test was to see if there are significant differences in the means of TAVG, TMAX, and TMIN between the years 2000 and 2023. The null hypothesis for the test with TAVG is that the mean daily average temperature in 2000 is equal to the mean daily average temperature in 2023 and the alternative hypothesis is that the two means are different.

	t	**Degrees of Freedom**	**p-value**	**Mean**
2000	-2.1457	27.844	0.04076	80.43545
2023				83.3720

Table 11: Two-group t-test values for TAVG, years 2000 and 2023

The p-value shown in Table 11 is less than 0.05, which means we reject the null hypothesis and conclude that there is a significant difference in the average temperature means for 2000 and 2023, where 2023 has a higher mean.

The null and alternative hypothesis for TMAX is the same as TAVG- the null hypothesis is that means are the same, and the alternative is that they are different.

	t	**Degrees of Freedom**	**p-value**	**mean**
2000	-2.6452	27.64	0.01331	**91.85871**
2023				**96.466**

Table 12: Two-group t-test values for TMAX, years 2000 and 2023

The p-value for TMAX in Table 12 is significantly less than 0.05, so we can reject the null hypothesis and conclude that there is a significant difference between the TMAX means of 2000 and 2023. The mean for 2023 is higher than the mean for 2000.

The null and alternative hypothesis for TMIN is the same as TAVG and TMAX.

	t	Degrees of Freedom	p-value	mean
2000	-0.70616	30.72	0.4854	70.07968
2023				71.03267

Table 13: Two-group t-test values for TMIN, years 2000-2023

The p-value in Table 13 is also less than 0.05, so we will reject the null hypothesis- there is a significant difference in TMIN means for the years 2000 and 2023, and 2023 has the higher mean here as well.

Overall, the two-group t-tests show that maximum, minimum, and average daily temperatures have all increased significantly between 2000 and 2023.

Chi square

Because the original dataset did not contain qualitative data to conduct a chi-squared test, I added a categorization variable for TAVG, TMAX, and TMIN that arranged them into groups of HIGH, MEDIUM, or LOW. The groups were determined by adding and subtracting standard deviation from the means of TAVG, TMAX, and TMIN- the range of the groups are described in Table 14.

	TAVG	TMAX	TMIN
LOW	< 76.94	< 87.56	< 66.43
AVERAGE	76.94 - 87.15	87.56 - 100.45	66.43 - 75.60
HIGH	> 87.15	> 100.45	> 75.7
Mean	82.05	94.00	71.07
Standard Deviation	5.11	6.45	4.63

Table 14: Categorization of LOW, AVERAGE, and HIGH groups

The null hypothesis for the Chi-squared test between TAVG and TMIN is that there is no relationship between the two, and the alternative hypothesis is that there is a relationship.

	AVG	**HIGH**	**LOW**
AVG	412	36	55
HIGH	38	75	0
LOW	46	0	65
Chi-Squared	**Degrees of Freedom**		**p-value**
424.24	4		< 2.2e-16

Table 15: 3x3 Table and Chi-Squared values created using TAVG and TMIN

Table 15 shows a Chi-squared value of 424.24. The p-value is so small it can be considered 0, meaning there is a 0% chance the observed values in the table occurred randomly and there is a very significant relationship between daily average and minimum temperatures.

Because all Ch-squared tests have the same null and alternative hypotheses (that there is not or is a relationship between the variables), the hypotheses for TAVG and TMAX are the same.

	AVG	**HIGH**	**LOW**
AVG	463	18	22
HIGH	11	102	0
LOW	30	0	81
Chi-Squared	**Degrees of Freedom**		**p-value**
886.68	4		< 2.2e-16

Table 16: 3x3 Table and Chi-Squared values created using TAVG and TMAX

Table 16 shows a Chi-squared value of 886.68, with a p-value that is practically 0. There is a 0% chance that TAVG and TMAX are not related, so we will reject the null hypothesis.

The null and alternative hypotheses are the same for the Chi-squared test between TMAX and TMIN.

	AVG	**HIGH**	**LOW**
AVG	389	55	52
HIGH	45	65	1
LOW	70	0	50
Chi-Squared	**Degrees of Freedom**		**p-value**
251.42	4		< 2.2e-16

Table 17: 3x3 Table and Chi-Squared values created using TAVG and TMAX

Table 17 shows a Chi-squared value of 251.42, and a p-value that can be considered 0. We will reject the null hypothesis here as well and conclude that TAVG and TMAX are very closely related. Overall, the temperature categories of the three variables are all very closely related.

ANOVA

For all ANOVA tests, the null hypothesis is that the variable means are different, and the alternative hypothesis is that they are the same. I conducted three ANOVA tests, one each for TAVG, TMAX, and TMIN, all against the year.

	Degrees of Freedom	Sum of Squares	Mean Squared Value	F value	p-value
Year	1	112	111.56	4.287	0.0388
Residuals	725	18867	26.02		

Table 18: ANOVA values for TAVG, with residual standard error 5.1

The p-value shown in Table 18 is less than 0.05, so we will reject the null hypothesis that the yearly means for TAVG are statistically significant.

	Degrees of Freedom	Sum of Squares	Mean Squared Value	F value	p-value
Year	1	132	132	3.182	0.0749
Residuals	725	30072	41.48		

Table 19: ANOVA values for TMAX, with residual standard error 6.4

The p-value in Table 19 is 0.0749, which is greater than 0.05. Therefore we fail to reject the null hypothesis and determine that the temperature maximum means are statistically different per year.

	Degrees of Freedom	Sum of Squares	Mean Squared Value	F value	p-value
Year	1	82	81.70	3.82	0.051
Residuals	725	15508	21.39		

Table 20: ANOVA values for TMIN, with residual standard error 4.6

The p-value in Table 20is 0.051, which is just barely greater than 0.05, and therefore we fail to reject the null hypothesis and can conclude that the means for temperature minimums are statistically different per year.

Overall, TMAX and TMIN show significant statistical differences, but TAVG does not. Because we determined that TAVG values are very significantly related to TMAX and TMIN values in the Chi-squared test, however, I would argue that TMAX and TMIN showing statistically significant differences every year means that TAVG is statistically different as well.

Multiple Regression

The Multiple Regression Equation (Table 21) shows p-values for TMAX and TMIN that are small enough to consider 0, and therefore are very closely related to TAVG.

	Degrees of Freedom	Sum of Squares	Mean Sq Value	F value	p-value	Multiple R-Squared
TMAX	1	17057.0	17057.0	19799.6406	< 2.2e-16	0.9674
TMIN	1	1188.6	1188.6	1379.6618	< 2.2e-16	

Table 21:Multiple Regression Equation

An R-squared value of 0.9674 means that 96.74% of the variance in TAVG can be explained by the variable TMAX and TMIN.

PCA

	PC1	PC2	PC3
Standard Deviation	1.6160	0.6043	0.15254
Proportion of Variance	0.8705	0.1217	0.00776
Cumulative Proportion	0.8705	0.9922	1.0000

Table 22: Importance of Components, eigenvalue = 1.616

Figure 10 and Table 22 show that the only components that should be kept are PC1 and PC2. PC1 by itself explains 87.05% of the variance, but PC1 and PC2 together explain 99.22% of the variance. Even though the weight of PC2 is less than 1, it is still statistically significant.

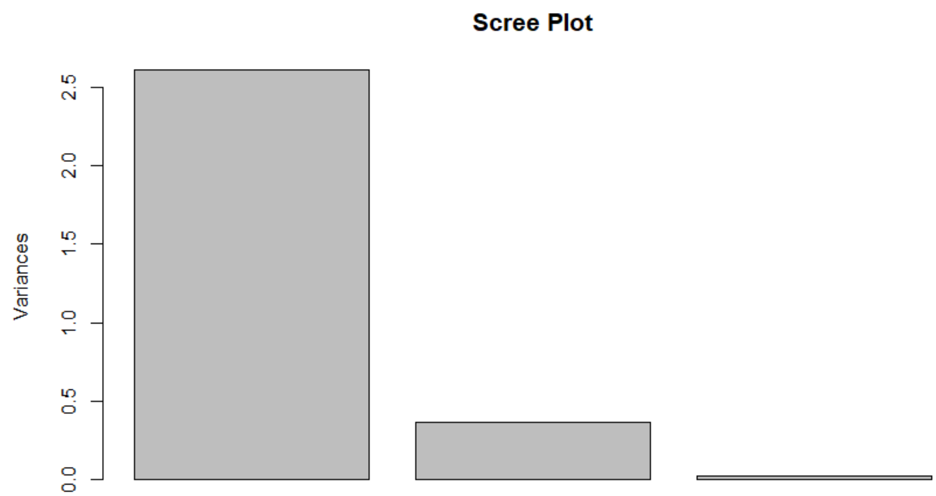

Figure 10: Scree Plot

Cluster Analysis

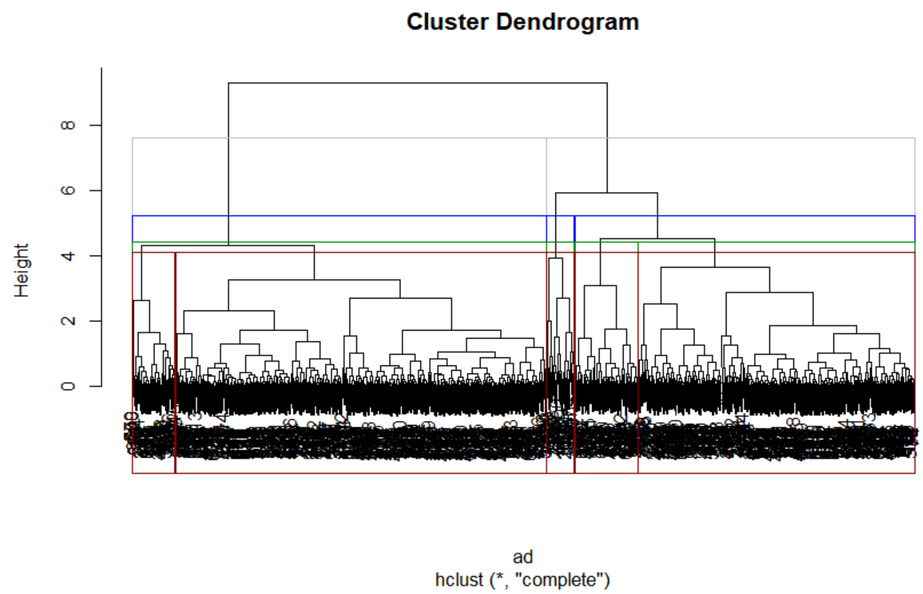

Figure 11: Cluster Dendrogram

Looking at Figure 11, I decided that two clusters (k = 2) is the best representation for this data. Figure 12 shows a large amount of overlap between the two clusters, with both clusters having the most variability along Component 1.

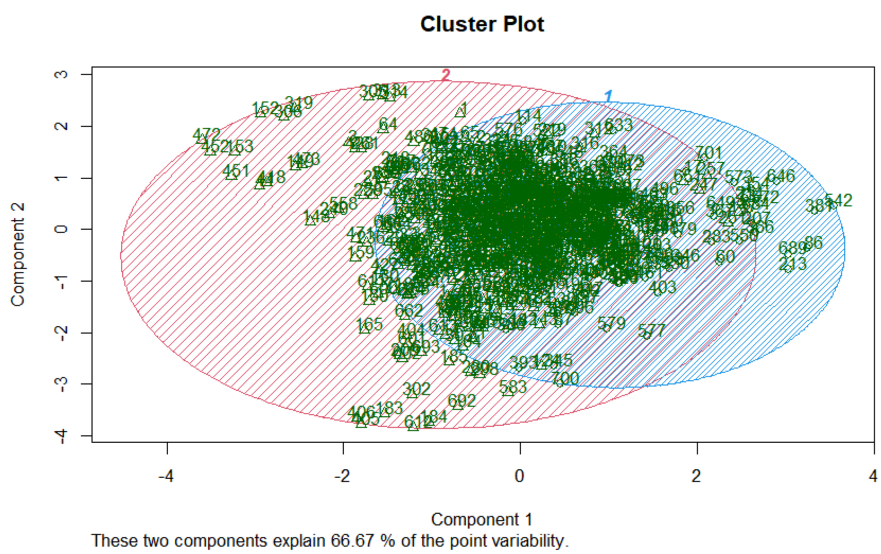

Figure 12: Cluster Plot

Conclusion/Summary

The analysis of temperature data from the ACME Mesonet station in Oklahoma shows the distributions of daily temperature averages, maximums, and minimums for July exhibit normality, indicating consistent patterns in temperature measurements over the years. Moreover, correlation and regression analyses suggest a slight but consistent increase in daily temperature averages, maximums, and minimums over the 23-year period from 2000 to 2023. However, when comparing consecutive years, these increases are minimal. Interestingly, when comparing the first and last years (2000 and 2023), there is a significant rise in temperatures, highlighting the importance of examining long-term trends for a comprehensive understanding of climate change.

In summary, the daily summer (July) temperature averages, maximums, and minimums at the ACME Mesonet station have risen over the past 23 years, with more pronounced differences when comparing the start and end of the period rather than by consecutive year. If I were to repeat this study, I would consider expanding the dataset to include a longer timeframe, such as 50 years, to potentially reveal more substantial trends. Additionally, I would explore analyzing more months or all months to capture a more comprehensive view of temperature changes. Precipitation data could also be incorporated to understand its impact on temperature trends.

Bibliography

Reid, A., Fuhlendorf, S., & Weir, J. (2015, February 10). *Weather variables affecting Oklahoma wildfires*. Rangeland Ecology & Management. https://www.sciencedirect.com/science/article/abs/pii/S1550742410500587

United States. Environmental Protection Agency. Office of Water. (2008). *National Water Program strategy response to climate change.* Washington, D.C.]: U.S. Environmental Protection Agency, Office of Water.

Warner, J. (2023, April 1). *Long-range water municipality planning in the face of climate change in Oklahoma and Texas*. Long-Range Water Municipality Planning in the Face of Climate Change in Oklahoma and Texas. https://hdl.handle.net/11244/337715