Trends in Heating and Cooling Days in Suburban and Urban Settlements in the U.S. Southwest

GitHub Repository: https://github.com/vinny-chasing-sunset/CerpaPaceWhatley___ENV872_EDA_FinalProject

Candela Cerpa & Samantha Pace & Vinny Whatley

2023-12-13

Contents

List of Tables

List of Figures

1. Rationale and Research Questions

Climate change, long-term shifts in temperatures and weather patterns, is a naturally occurring process that has been accelerated by human actions, primarily the burning of fossil fuels. One of the impacts of climate change is rising temperatures, which have negative impacts on Earth's ecosystems, such as melting glaciers, intensifying storms, and worsening drought conditions. Humans are directly impacted by rising temperatures: as the number of hot days and heat waves increases, so does the number of heat-related illnesses.

The U.S. has not escaped these impacts. The U.S. Southwest, which we define as Arizona, California, and Nevada, has witnessed an uptick in droughts. Lake Me, the main source of drinking water for Nevada, has seen record lows in recent years, cautioning the state to adopt water conservation strategies. News about California's wildfire seasons seem to increase in intensity and length every year. Arizona has reported some of the highest temperatures in the U.S. Lastly, the impacts are not spread evenly across human settlements. The Urban Heat Island phenomenon describes the situation where cities, with their heat-trapping cement and lower tree canopy cover, tend to be hotter than nearby rural areas.

In light of these issues, this project aims to understand the intensity of rising temperatures in the U.S. Southwest. Our research questions are: 1. Has there been an increase in the number of Heating and Cooling Days in the U.S. Southwest from 1980 to 2022?

2. Is there a difference in Heating and Cooling Days in urban and rural spaces in the U.S. Southwest over time?

3. How have minimum and maximum temperatures changed over time?

2. Dataset Information

We retrieved historical temperature data from NOAA's National Centers for Environmental Information (NCEI), specifically from the GHCN (Global Historical Climatology Network). We chose this dataset because NCEI is a reputable source of weather-related historic data. We selected a total of six stations, encompassing three major cities and each with an associated nearby suburban location. We selected the nearby suburban locations to be within 50 miles of the major city and with a population of less than 50,000 in 2022. The selected locations are as follows:

- Phoenix and Fountain Hills, Arizona
- Las Vegas and Pahrump, Nevada
- San Diego and Ramona, California

We primarily examined the daily data for TMAX = Maximum temperature (Fahrenheit) and TMIN = Minimum temperature (Fahrenheit) for each of the six locations. As Table 1 shows, our datasets had varying levels of coverage and start dates for each location. All the datasets had high coverage from 1998-2022, and all but Ramona, CA have data from 1980-1998. More information about this data's documentation can be found at https://www.ncei.noaa.gov/pub/data/cdo/documentation/GHCND_documentation.pdf and in the Metadata folder of this project's repository.

Table 1: Data Structure Summary

Location	Date	Daily Temp Max (°F)	Daily Temp Min (°F)	Temp Avg (°F)	Month	Year	Day of Year	Heating Degrees (°F)	Cooling Degrees (°F)
Phoenix	1980-01-01 to 2022-12-31	43 to 122	26 to 96	36.5 to 106.5	1 to 12	1980 to 2022	1 to 366	0 to 28.5	0 to 41.5
Fountain Hills	1980-01-01 to 2022-12-31	25 to 125	23 to 97	33.5 to 107.5	1 to 12	1980 to 2022	1 to 366	0 to 31.5	0 to 42.5
Las Vegas	1972-01-01 to 2022-12-31	31 to 117	11 to 95	22 to 106	1 to 12	1972 to 2022	1 to 366	0 to 43	0 to 41
Pahrump	1972-01-01 to 2022-12-31	27 to 115	-2 to 86	15 to 98	1 to 12	1972 to 2022	1 to 366	0 to 50	0 to 33
San Diego	1972-01-01 to 2023-01-01	50 to 107	34 to 78	43.5 to 87.5	1 to 12	1972 to 2023	1 to 366	0 to 21.5	0 to 22.5
Ramona	1998-04-16 to 2023-01-01	47 to 117	0 to 76	35 to 92	1 to 12	1998 to 2023	1 to 366	0 to 30	0 to 27

3. Data Wrangling

We took several steps in our workflow to clean our datasets. The raw data included both the city and suburb in the same CSV, so we first filtered the data to separate out the locations. For each of the 6 locations, we selected the columns we were interested in exploring, which included NAME, DATE, TMAX, TMIN, TAVG, and created columns for Month, Year, and Day of Year to enable different years to be overlaid on a graph.

The Fountain Hills, Pahrump, and Ramona datasets had some missing data points in the maximum and minimum daily temperatures, so we used linear interpolation to fill in the missing data. The raw data for the 'Average Temperature' column contained many missing values, so we inputted the calculated mean of the daily maximum and minimum for this column.

Heating and Cooling Degree Days Calculations

Having the daily average temperature calculated for each location, we then calculated the Heating Degree Days and Cooling Degree Days. Heating Degree Days are when people turn their heater on because the average temperature is low, and Cooling Degree Days is when people turn their air conditioner on because the average temperature is warm. The industry standard threshold is 65 degrees F, meaning that anything higher than 65 F would constitute a Cooling Degree Day, and anything lower than 65 would mean it is a Heating Degree Day (US EPA, 2016). The equations to calculate the Heating and Cooling Degree Days are as follows:

- If mean temperature $>$ 65°F then
- mean temperature - 65°F = Cooling Degree Days
- If mean temperature $<$ 65°F then
- 65°F - mean temperature = Heating Degree Days

From these calculations we are able to look at the number of Heating and Cooling Degree Days, as well as the intensity, which are two main components examined in the exploratory analysis section.

4. Exploratory Analysis

For the purposes of exploring our data, we first plotted the three major cities together to get a sense of how their average daily temperatures compared over time, shown in Figure 1. Based on the figure, there are clear seasonal trends throughout the year, and Phoenix has the highest daily average temperature over time, followed by Las Vegas and then San Diego. The trend for Las Vegas has a noticeable positive slope over time, while the trend for Phoenix demonstrates only a slight increase and San Diego's trend remains horizontal with no major changes in slope.

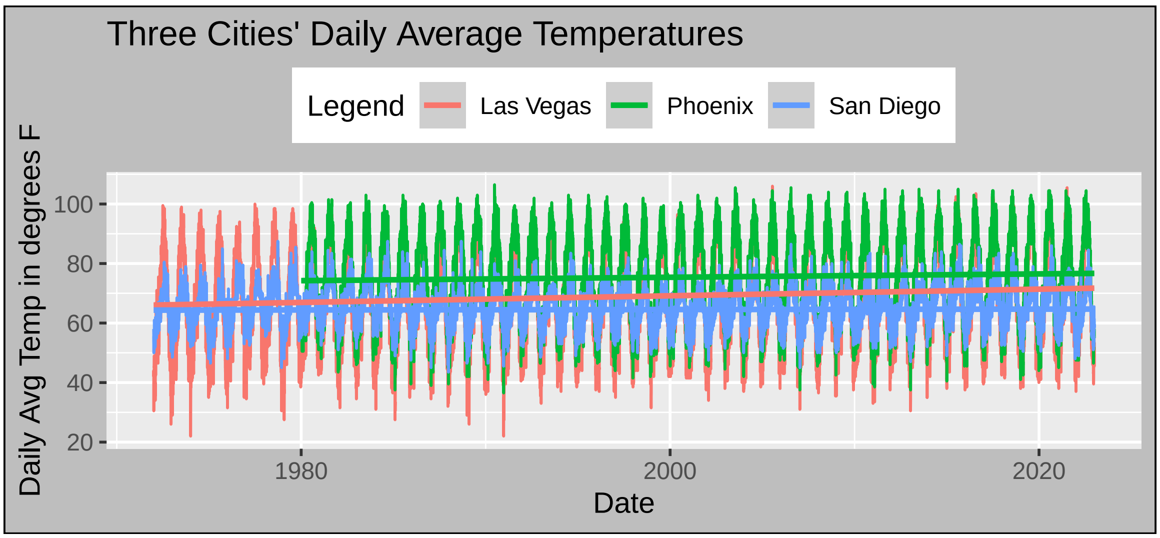
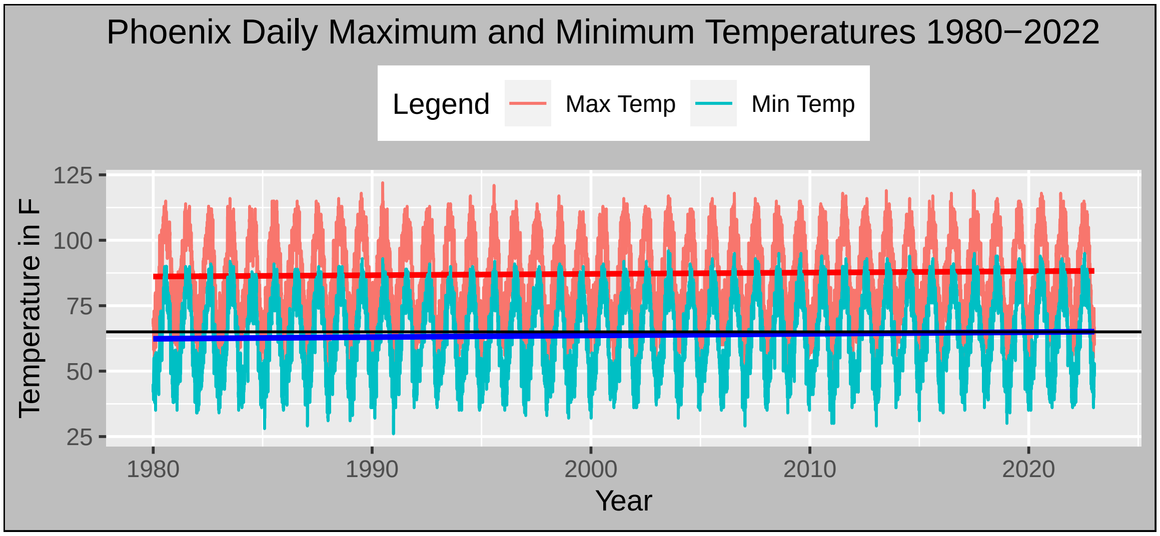


Figure 1: Average Temperature for Phoenix, Las Vegas, San Diego

For the remainder of the exploratory analysis section, we focused on a single city (Phoenix, Arizona) to better understand the data for the statistical analyses. Figure 2 shows Phoenix's maximum and minimum daily temperature from 1980 until 2022. The trend for the minimum and maximum temperatures shows that there has been a modest increase in both over time. The black horizontal line in Figure 2 marks the 65 F threshold for Heating and Cooling Degrees. Based on the trends in this figure, we would expect there to be more Cooling than Heating Degree Days.



Phoenix Daily Maximum and Minimum Temperatures 1980−2022
Legend
Max Temp
Min Temp
Temperature in F
125
100
75
50
25
1980
1990
2000
2010
2020
Year

Figures 3 and 4 compare the number of Heating and Cooling Degree Days throughout the year for Phoenix in 1980 and 2022. The differences in the two figures demonstrate an increase in the number of Cooling Degree Days in March, April, and October and a decrease in the number of Heating Degree Days in February and March.

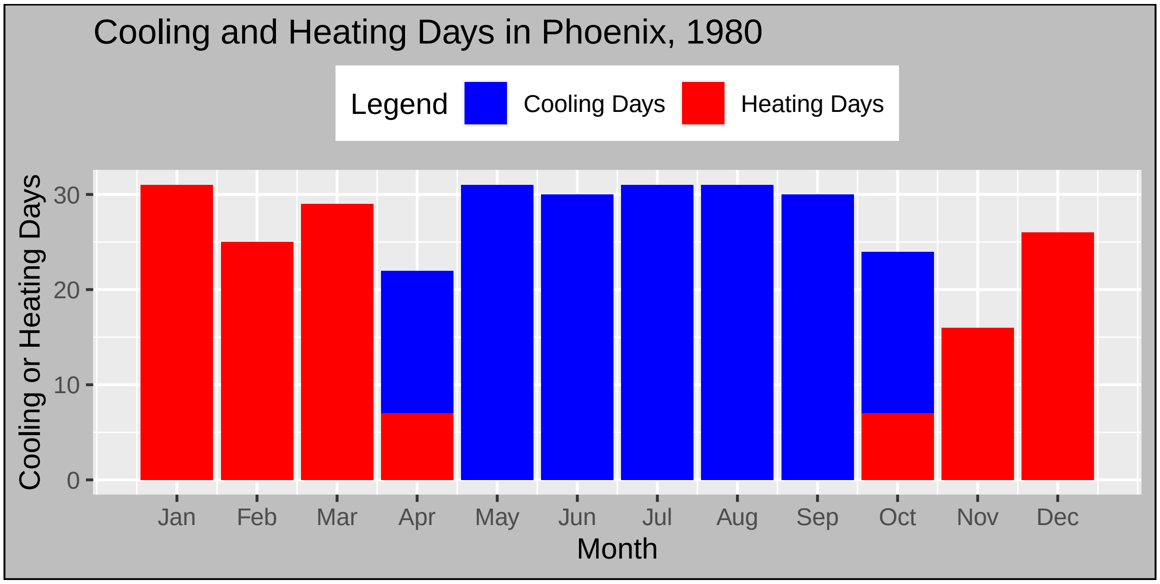


Figure 2: Number of Phoenix's Heating and Cooling Degree Days, 1980

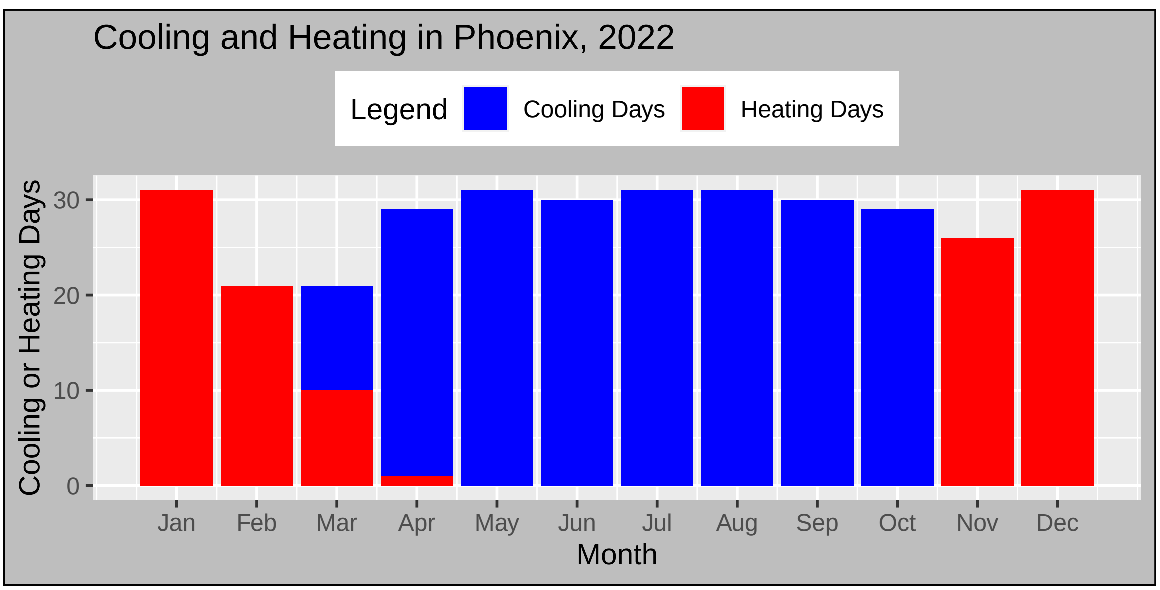


Figure 3: Number of Phoenix's Heating and Cooling Degree Days, 2022

Figures 5 and 6 show the intensity of Heating and Cooling Degree Days in Phoenix in 1980 and 2022. In this instance, intensity is the number of degrees over or under 65 F per day. The peaks of the Heating Degree Day intensity are approximately half the intensity of the Cooling Degree Day intensity. Figure 5 shows that there were some higher peaks in 2022 of Heating Degree Days, and March of 2022 saw a much lower intensity of Heating Degree Days than March of 1980. Figure 6 shows the intensity of Cooling Degree Days peaking consistently in the month of July. July 2022 reached 39 degrees F and July 1980 reached 36 degrees F. In 2022 compared to 1980, Cooling Degree Days started earlier in the year, and consistently reached a slightly higher intensity.

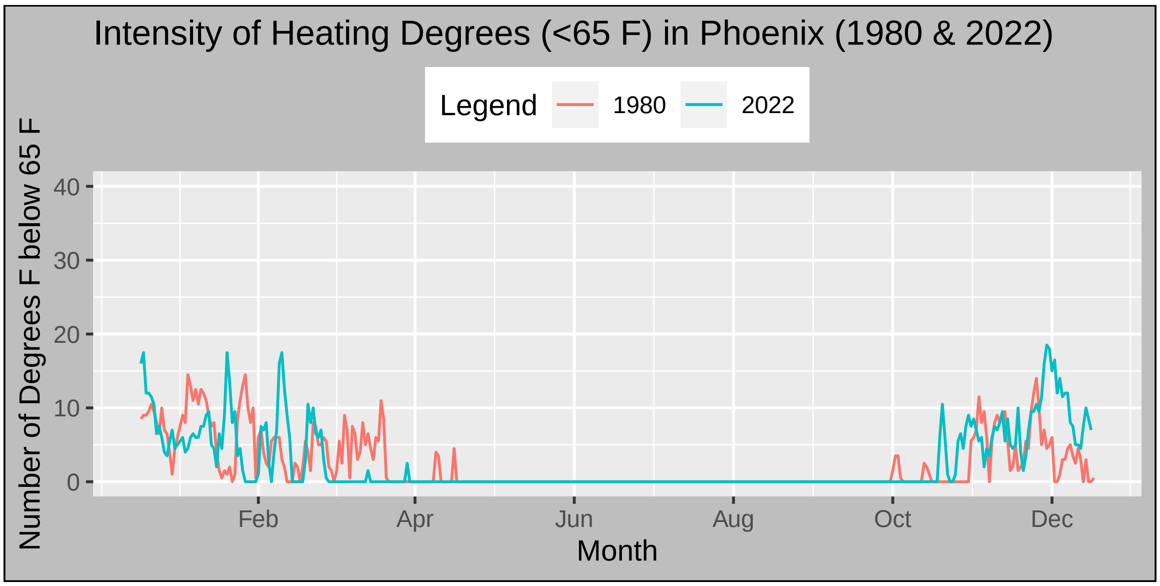
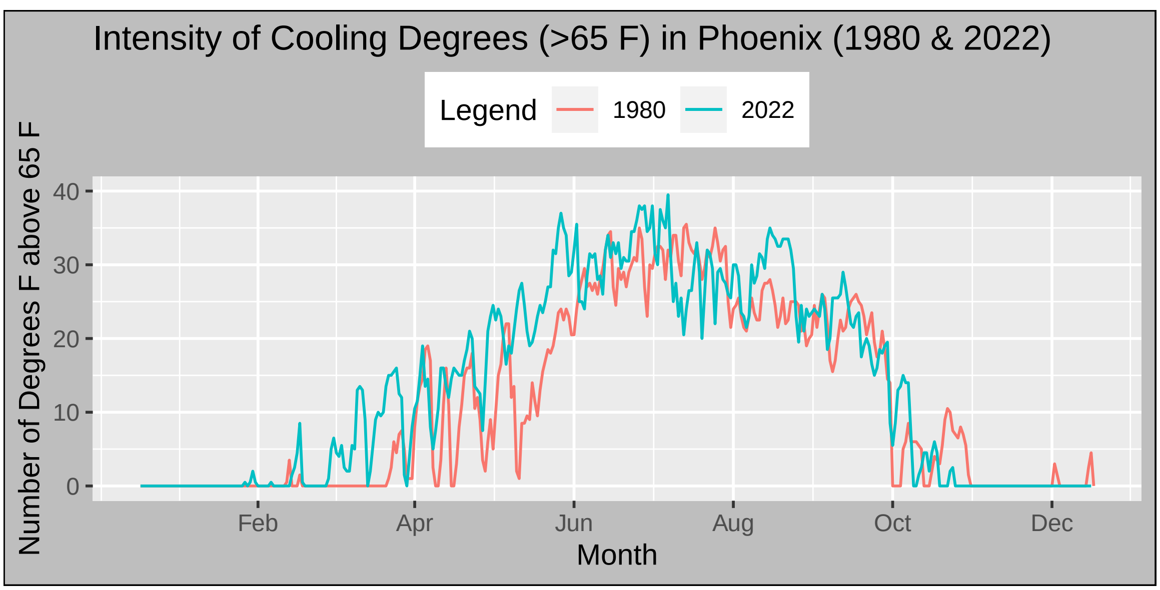


Figure 4: Comparison of Heating Degree Days in Phoenix in 1980 and 2022



Comparing Figures 3 and 4 to 5 and 6 reflects the importance of understanding the intensity over quantity of Heating and Cooling Days. While the initial comparison shows a slight increase in the number of days, the intensity comparison denotes a far steeper increase. The intensity also impacts energy usage, because people usually have a buffer of temperatures at which they will turn on the heat or AC that is not just 65 degrees F. More days that are closer to 100 degrees F have a worse impact on human health and energy usage than more days that are closer to 70 degrees F. In our analysis, we thus chose to look more closely at intensity over quantity of Heating and Cooling Degree Days.

5. Analysis

Question 1: Has there been an increase in the number of Heating and Cooling Days in the U.S. Southwest from 1980 to 2022?

To address this research question, we used a time series analysis for the Heating and Cooling Degree Days for each of the three major cities. We made univariate time series objects for the Heating and Cooling Degree Days for each city, and then decomposed each time series object to separate out the seasonality, trend, and remainder from the data. The graphs below showcase the time series plots.

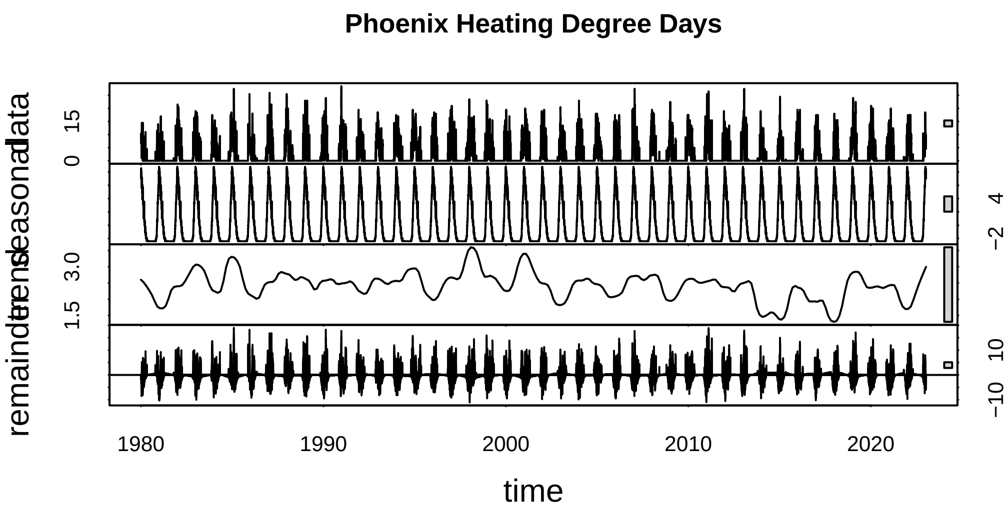


Figure 5: Phoenix Decomposed Time Series for Heating Degree Days

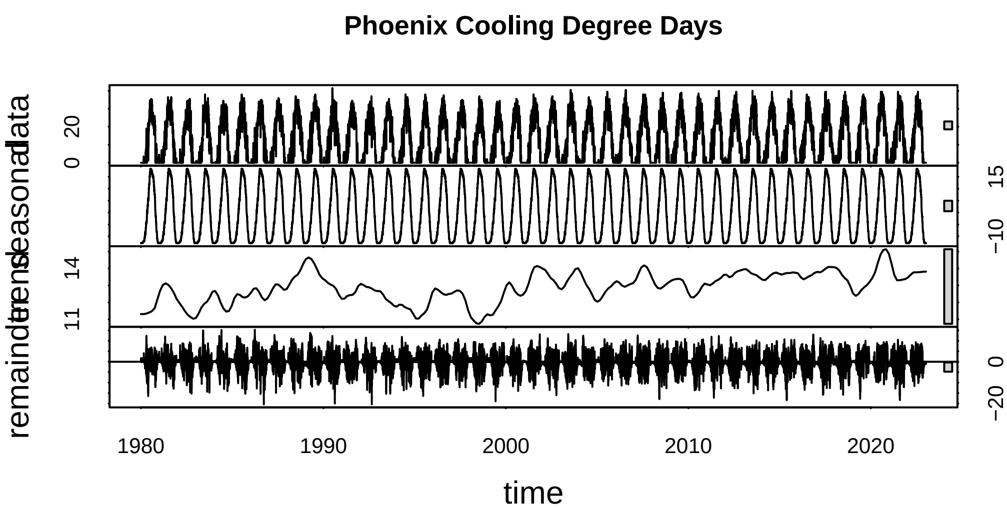


Figure 6: Phoenix Decomposed Time Series for Cooling Degree Days

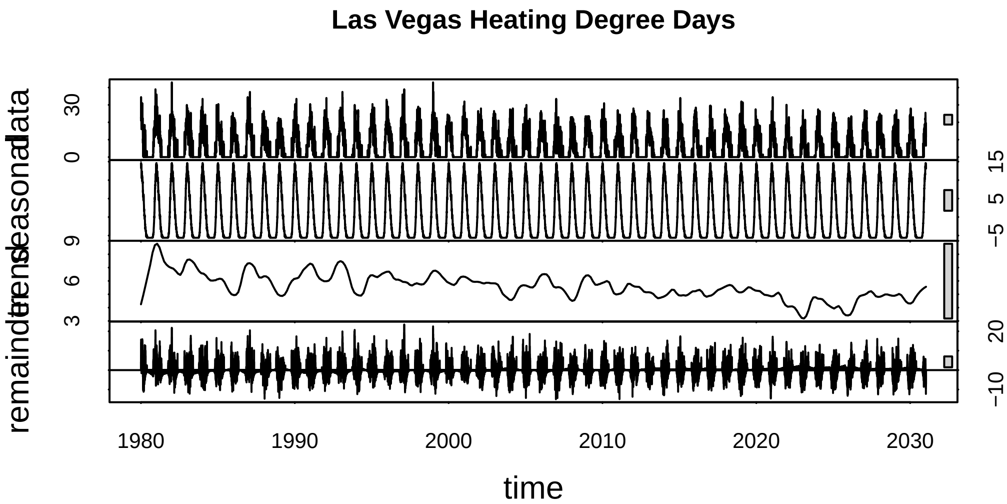


Figure 7: Las Vegas Decomposed Time Series for Heating Degree Days

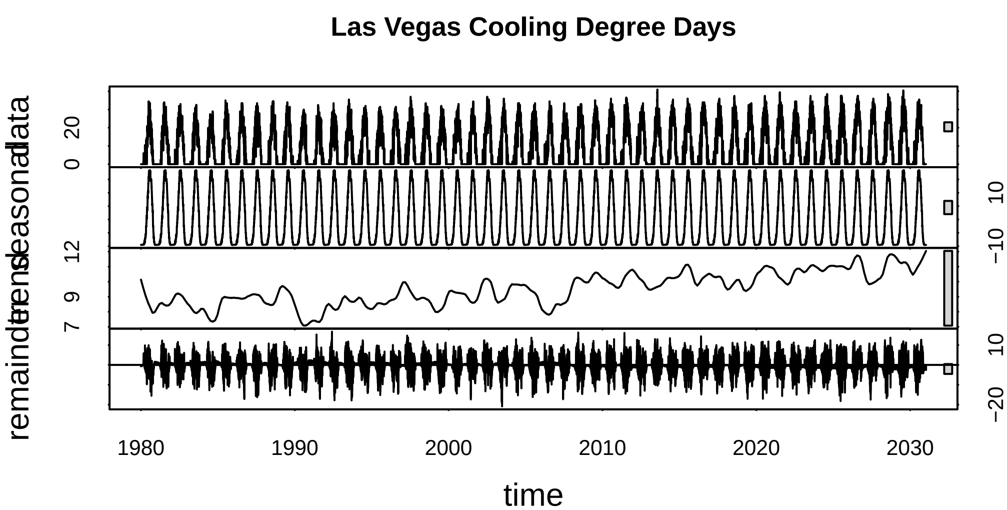


Figure 8: Las Vegas Decomposed Time Series for Cooling Degree Days

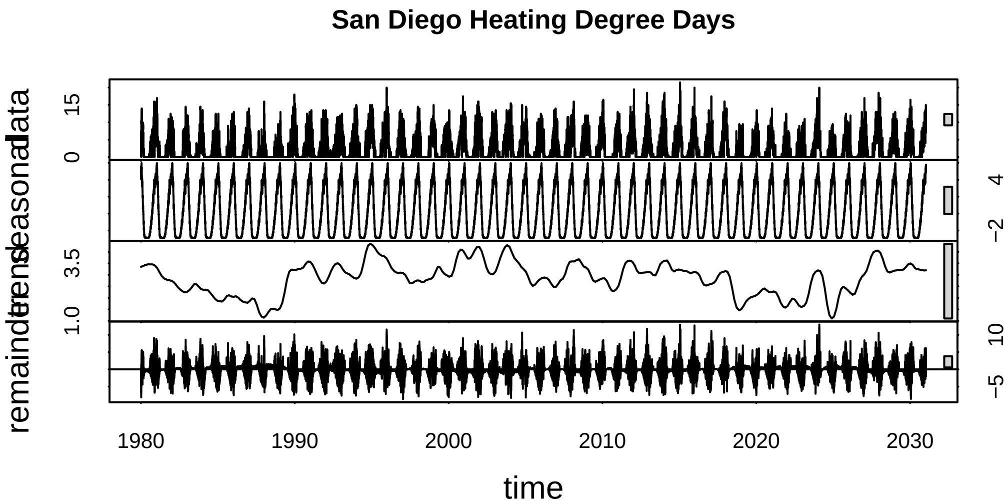


Figure 9: San Diego Decomposed Time Series for Heating Degree Days

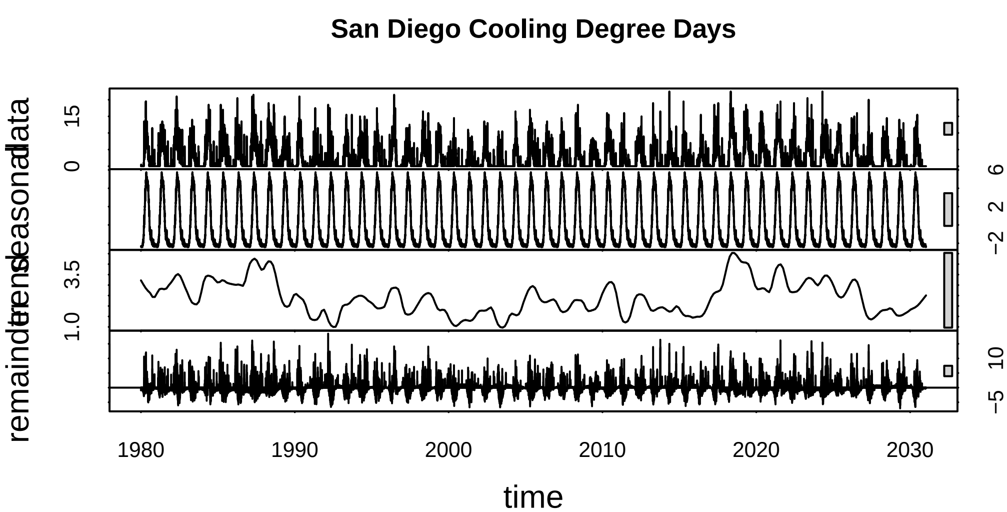


Figure 10: San Diego Decomposed Time Series for Cooling Degree Days

The Seasonal Mann-Kendall test was performed on each of the city's Heating and Cooling Degree Days data to analyze if the data is stationary. The null hypothesis for the Seasonal Mann-Kendall test is that the data is stationary, meaning there is no significant trend over time. The alternative hypothesis is that there is a significant trend in the data over time.

The results of the Seasonal Mann-Kendall test for the Heating Degree Days in Phoenix, Las Vegas, and San Diego are found in Table 2. The results indicate that Phoenix and Las Vegas have a p-value of less than 0.05 (3.0198e-13 and 2.22e-16, respectively), so the null hypothesis can be rejected. The p-value for San Diego is greater than 0.5 (having a value of 0.49529), so we cannot reject the null hypothesis. The tau value indicates the slope of the trend, and all are negative, but this is only significant for Phoenix and Las Vegas. Las Vegas has the steepest slope.

Table 2: Results for the Seasonal Mann-Kendall test for Heating Degree Days

City	p-value	tau value
Phoenix	3.0198e-13	-0.0592
Las Vegas	2.22e-16	-0.16
San Diego	0.49529	-0.00414

Table 3 shows the results of the Seasonal Mann-Kendall Test for the Cooling Degree Days over time in Phoenix, Las Vegas, and San Diego. Phoenix and Las Vegas have a p-value of less than 0.05 (p-value = 2.22e-16 for each city), so the null hypothesis can be rejected. The Heating Degree Days for Phoenix and Las Vegas have a significant trend over time. The Cooling Degree Days for San Diego produced a p-value of 0.36317, which is greater than 0.05, so we cannot reject the null hypothesis for this location. There is not a significant trend over time in the Cooling Degree Days for San Diego. For both Phoenix and Las Vegas, the slope of the trend line is positive, meaning there is a positive trend over time in Cooling Degree Days in these locations.

Table 3: Results for the Seasonal Mann-Kendall test for Cooling Degree Days

City	p-value	tau value
Phoenix	2.22e-16	0.0983
Las Vegas	2.22e-16	0.189
San Diego	0.36317	-0.00542

In summary, we can conclude there has been a significant increase in the trend of Cooling Degree Days over time in some cities in the U.S. Southwest from 1980 to 2022, but not uniformly. There has been some significant decrease in the trend of Heating Degree Days over time in some U.S. Southwest cities from 1980 to 2022, but not in all locations. The increase in Cooling Degree Days and decrease in Heating Degree Days is evidence of the overall warming temperatures in Las Vegas and Phoenix. Tests performed on the San Diego dataset indicated that this data may be stationary, meaning there is no significant trend upward or downward in the Heating and Cooling Degree Days over time.

Question 2: Is there a difference in Heating and Cooling Days in urban and rural spaces in the U.S. Southwest over time?

To address this research question, we applied ANOVA tests to find whether there was a significant difference in the means of Heating and Cooling degree days between urban and rural spaces. In particular, we compared the Location, Year, and Location and Year as potential significant factors that contributed to the potential difference in Degree Days means. This required creating data frames for each sate to apply the test where Heating or Cooling Degree Days were the dependent variable and Location and Year as the independent variable.

To address question 2, we thought it was best to perform ANOVA testing. This would allow us the check the means of the "Heating Degrees" and "Cooling Degrees" for the each state against Location, Year, and Location:Year finding out what was statically significant to the contribution of the average mean of "Heating Degrees" and "Cooling Degrees" . But to get an accurate picture we thought it would be best to take the means of the Heating and Cooling variables of each state and then compare it to each other state in the dataset. This required us to create a data frame for each state where we combined the metropolitan city and suburban city data by Location, Year, Heating Degrees, and Cooling Degrees. Then running 2 ANOVA test for each of those 3 states, the first test having Heating degrees as the dependent variable and the second test having Cooling degrees as dependent variable.

After that using the ANOVA data for Heating and Cooling degrees we created tables for both dependent variables using the Kable function. With this we were able to learn a few things, firstly that for Nevada, location contributed and was the only significant variable to average Heating temperature. While the same story was told when looking at the average Cooling temperature, location being statically significant. Looking at only this and no other analysis would lead you to believe that location is the reason we see such high and low temperatures, especially considering that Nevada is a desert and it typically sees highly polarized temperatures dependent on the time of day.

Secondly, looking at the table for Arizona, we find that this data tells the same story as previously stated for Nevada. That Location is the statically significant factor in average Heating and Cooling temperatures. Lastly, from California's data we learn that again for Heating degree's location was the statically significant contributor but for Cooling degrees Year instead is statically significant. We are thinking this is the case because our California data set (from 1998 to 2022) does not cover the same amount of time as the Nevada dataset (1972 to 2022) and Arizona dataset (1980 to 2022). Leading to slightly different results.

Table 4: ANOVA Summary Table for Nevada Heating Degrees

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Location	1	62571.738	62571.73758	803.92955	0
Year	1	6896.027	6896.02683	88.60102	0
Location:Year	1	2832.745	2832.74535	36.39547	0
Residuals	37026	2881821.104	77.83236	NA	NA

Table 5: ANOVA Summary Table for Nevada Cooling Degrees

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Location	1	117133.636	117133.63640	1228.73828	0e+00
Year	1	14619.575	14619.57493	153.36014	0e+00
Location:Year	1	2439.567	2439.56728	25.59119	4e-07
Residuals	37026	3529628.787	95.32839	NA	NA

Table 6: ANOVA Summary Table for Arizona Heating Degrees

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Location	1	6041.75640	6041.75640	236.29519	0.0000000
Year	1	1136.79017	1136.79017	44.46026	0.0000000
Location:Year	1	94.69209	94.69209	3.70344	0.0543096
Residuals	30890	789816.57781	25.56868	NA	NA

Table 7: ANOVA Summary Table for Arizona Cooling Degrees

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Location	1	32647.6265	32647.6265	238.217915	0.0000000
Year	1	3447.7808	3447.7808	25.157209	0.0000005
Location:Year	1	926.9335	926.9335	6.763498	0.0093086
Residuals	30890	4233456.5144	137.0494	NA	NA

Table 8: ANOVA Summary Table for California Heating Degrees

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Location	1	5.854260e+04	58542.604743	1870.433248	0.0000000
Year	1	3.572846e+03	3572.845863	114.152244	0.0000000
Location:Year	1	6.610809e+00	6.610809	0.211215	0.6458223
Residuals	18105	5.666676e+05	31.298954	NA	NA

Table 9: ANOVA Summary Table for California Cooling Degrees

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Location	1	306.4351	306.43511	18.888099	0.0000139
Year	1	3851.3184	3851.31840	237.388211	0.0000000
Location:Year	1	117.6937	117.69367	7.254422	0.0070792
Residuals	18105	293730.3397	16.22371	NA	NA

We then extracted the f-values from those datasets and created bar charts to visualize the stark difference between statically significant variables compared to the other variables.

In summary, with anova testing we learned that biggest contributor to average Heating and cooling temperatures for each can be attributed to solely location. Requiring us to do further analysis this time instead with time-series.

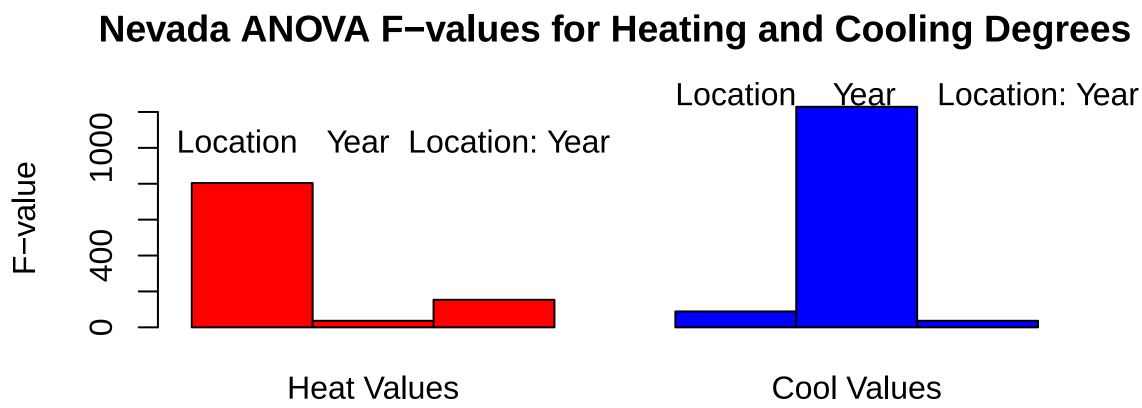


Figure 11: Nevada ANOVA F-values for Heating and Cooling Degrees

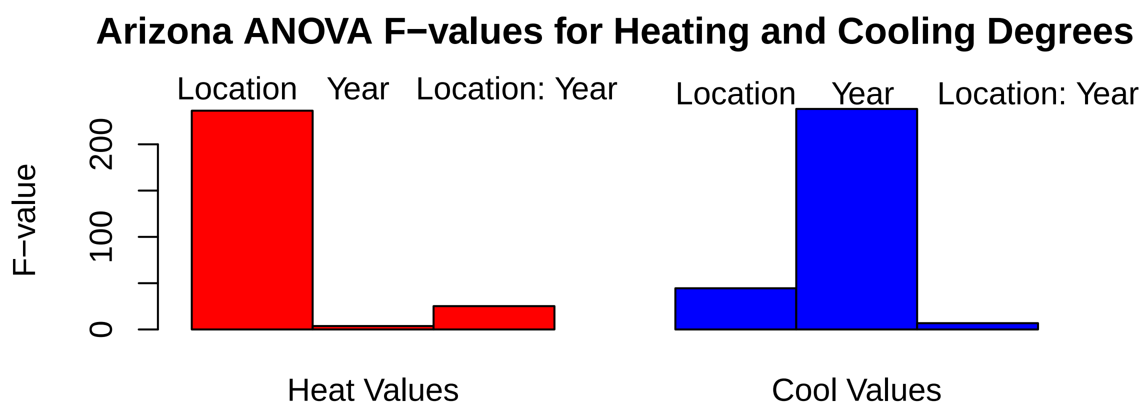


Figure 12: Arizona ANOVA F-values for Heating and Cooling Degrees

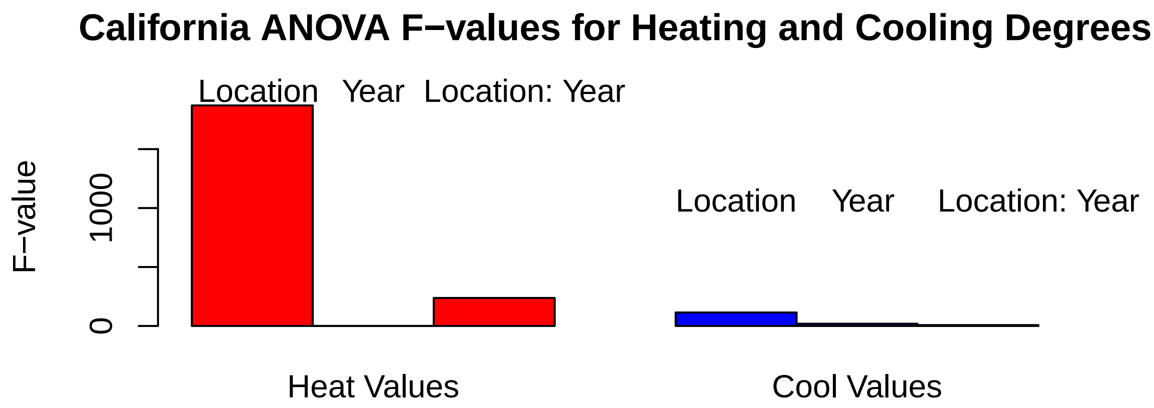


Figure 13: California ANOVA F-values for Heating and Cooling Degrees

Question 3: How have minimum and maximum temperatures changed over time?

To address the question of whether minimum and maximum temperatures have changed over time, we conducted a time series analysis on every location. Similarly to the analysis for Question 1, we created univariate time series objects for each measure for every site and then decomposed each time series to review the overall trend and presence of seasonality. The trend graph, the third row in every decomposition, makes it easier to see what the general trend is. The trend of Las Vegas minimum temperature, for example, has a very clear upward trend. The trends of Ramona's maximum and minimum temperatures are also clearly downward. All other trends, however, are difficult to ascertain just by looking at them, and the degree of the slope is an important value to know.

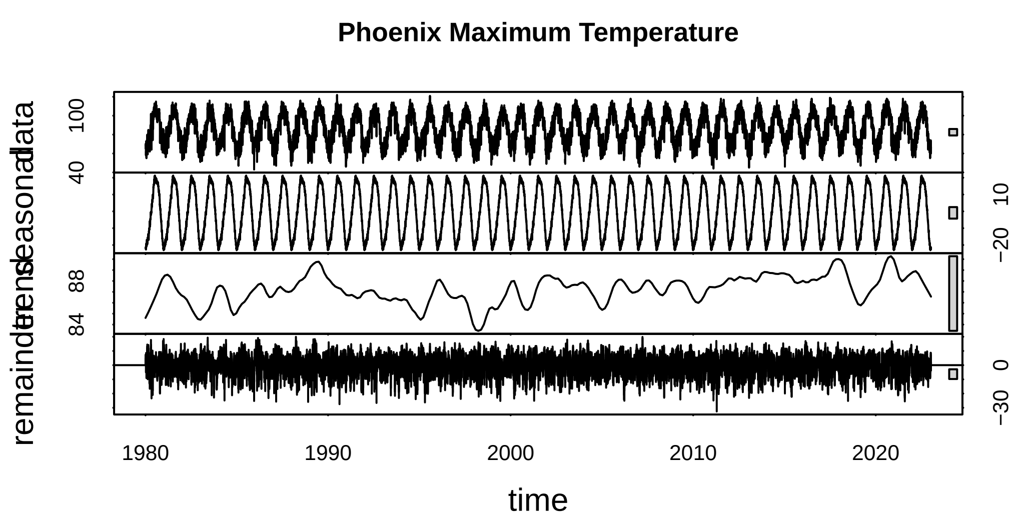


Figure 14: Phoenix Decomposed Time Series for Maximum Temperatures

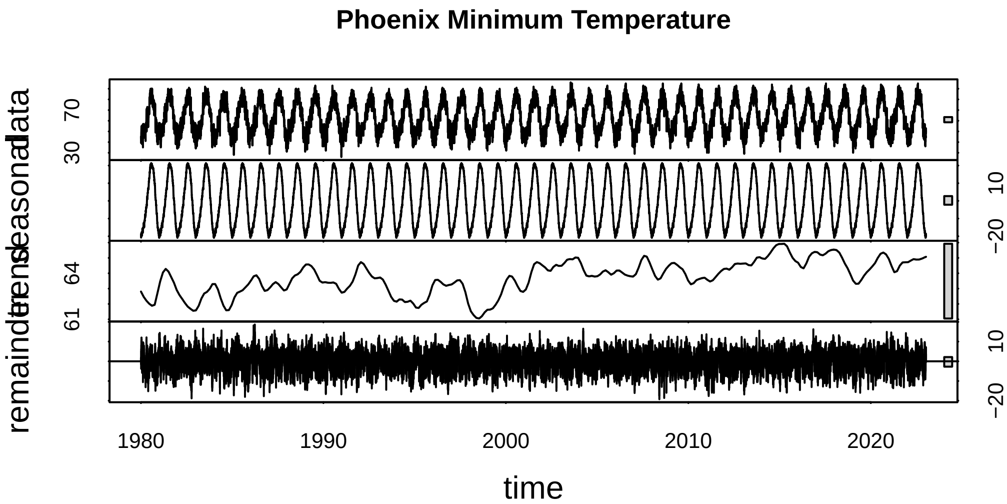


Figure 15: Phoenix Decomposed Time Series for Minimum Temperatures

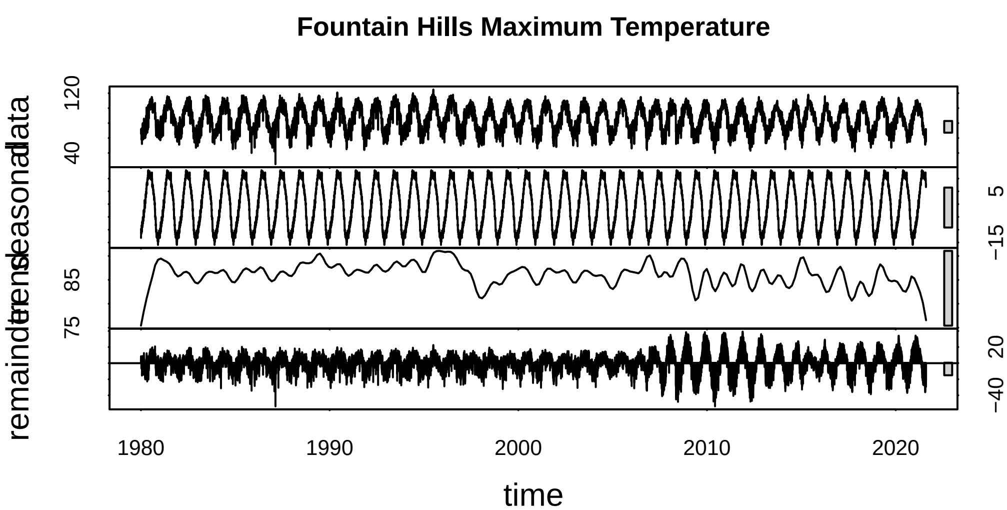


Figure 16: Fountain Hills Decomposed Time Series for Maximum Temperatures

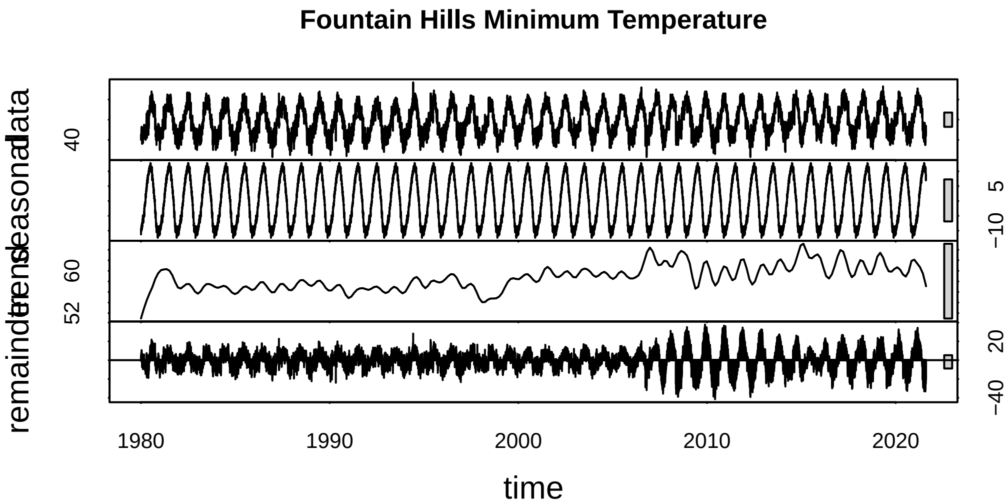


Figure 17: Fountain Hills Decomposed Time Series for Minimum Temperatures

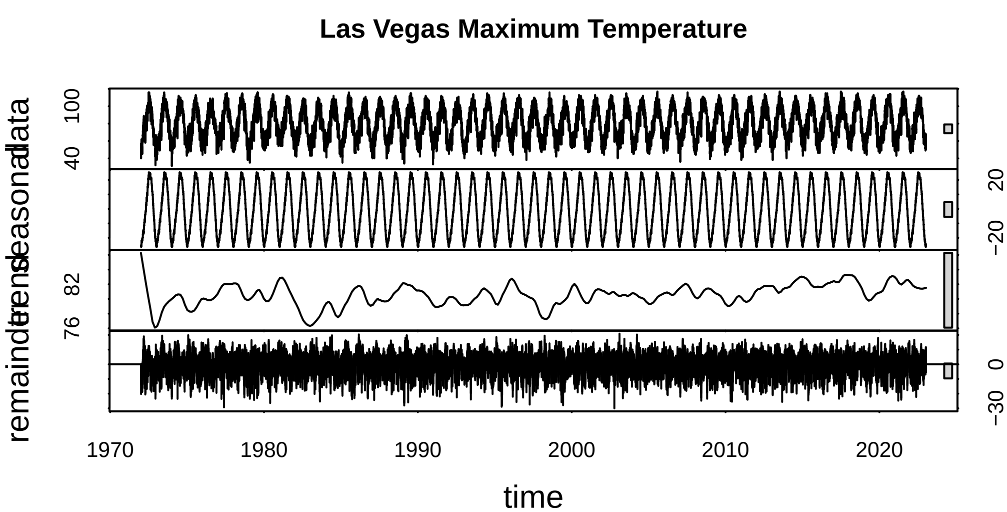


Figure 18: Las Vegas Decomposed Time Series for Maximum Temperatures

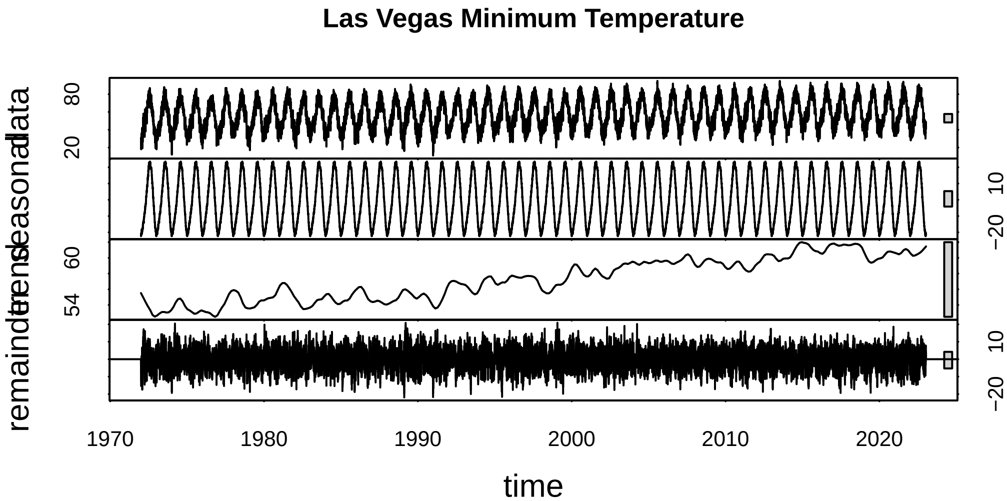
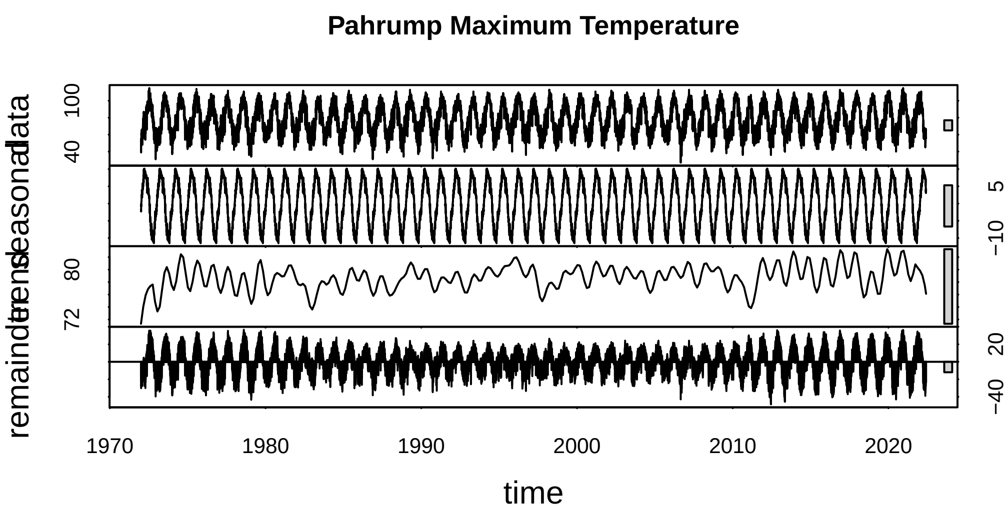
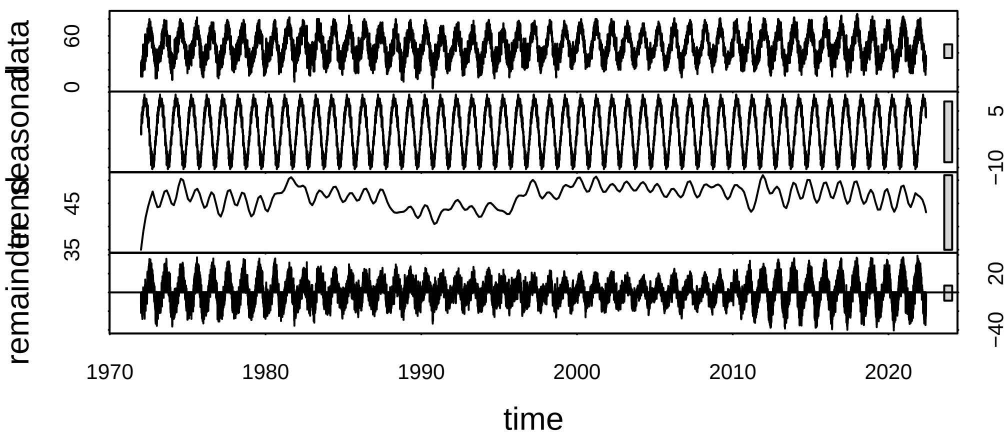


Figure 19: Las Vegas Decomposed Time Series for Minimum Temperatures



Pahrump Minimum Temperature



San Diego Maximum Temperature

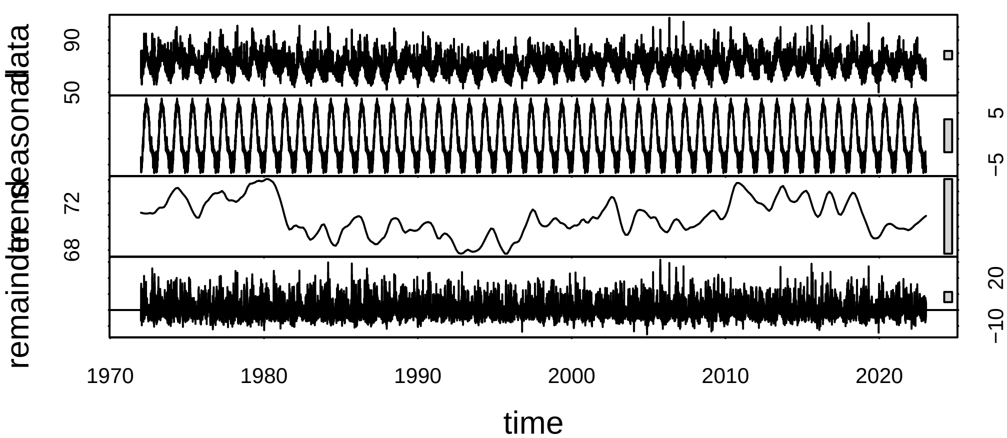


Figure 20: San Diego Decomposed Time Series for Maximum Temperatures

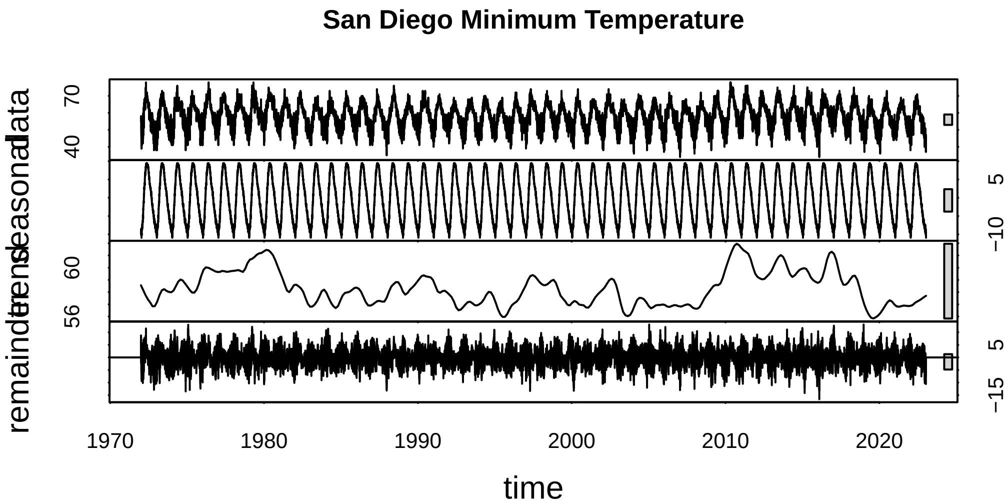


Figure 21: San Diego Decomposed Time Series for Minimum Temperatures

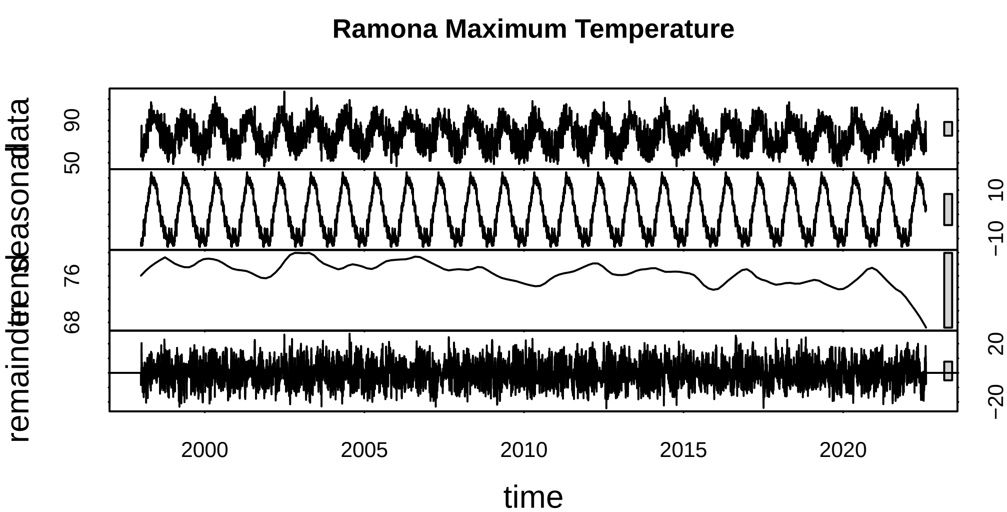


Figure 22: Ramona Decomposed Time Series for Maximum Temperatures

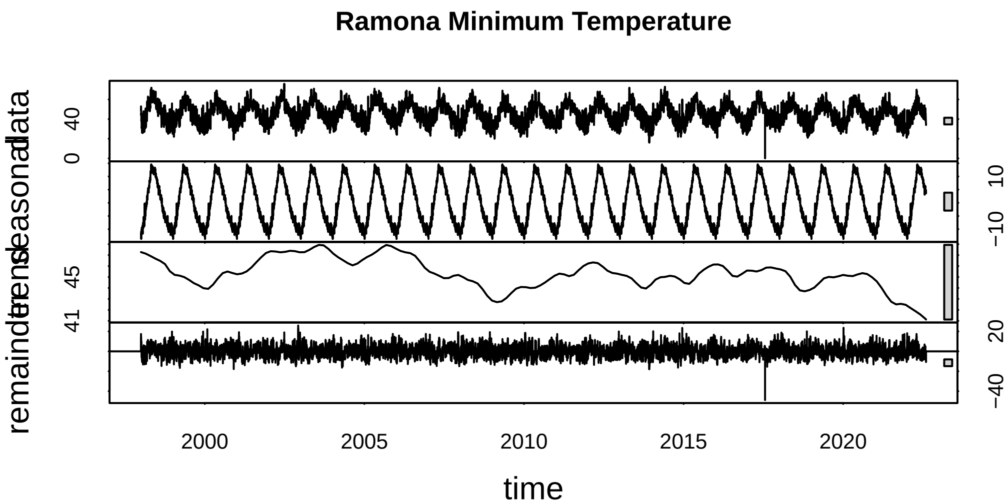


Figure 23: Ramona Decomposed Time Series for Minimum Temperatures

To understand whether the data exhibit a statistically significant upward or downward trend over time, we conducted Seasonal Mann-Kendall tests. As a reminder, the null hypothesis is that the data is stationary and does not exhibit a statistically significant upward or downward trend over time. This would mean that the data exhibit no meaning that there is no change in the temperature minimum and maximum over time. The results of the tests, summarized below, show that all trends are statistically significant, as all p-values are smaller than 0.05, so none of the data are stationary. The tau value shows the slope of the trend line.

Interestingly, Fountain Hill's maximum temperatures, San Diego's minimum temperatures, and Ramona's maximum and minimum temperatures have a downward trend. Every other variable has an upward trend. Furthermore, most maximum temperatures are increasing at a slower rate than the minimum temperatures. The only city where the opposite happens is Ramona, which is already set apart from the rest by being the only city where both temperature variables are decreasing, and the difference is smaller. The steepest slope is in Las Vegas' minimum temperatures.

Table 10: Results for the Seasonal Mann-Kendall test on minimum and maximum temperatures

City	Variable	p-value	tau value
Phoenix	max temperature	<2.22 e-16	0.0617
Phoenix	min temperature	<2.22 e-16	0.0951
Fountain Hills	max temperature	7.66 e-07	-0.0282
Fountain Hills	min temperature	<2.22 e-16	0.0979
Las Vegas	max temperature	<2.22 e-16	0.0671
Las Vegas	min temperature	<2.22 e-16	0.29
Pahrump	max temperature	<2.22 e-16	0.0482
Pahrump	min temperature	<2.22 e-16	0.0678
San Diego	max temperature	8.91 e-4	0.0173
San Diego	min temperature	5.10 e-4	-0.0182
Ramona	max temperature	<2.22 e-16	-0.112
Ramona	min temperature	<2.22 e-16	-0.0912

Though the trends are not linear, as shown in the decomposition graphs, these graphs show the direction of the trends in minimum and maximum temperatures across cities.

In summary, we can conclude that there are significant changes in the trend of minimum and maximum temperatures across all cities in the U.S. Southwest, but these changes are not occurring uniformly Many of the cities, both urban and rural, are seeing their minimums grow at a faster pace than the also-growing maximums. However, exploring why some urban and rural cities have maximum and minimum temperatures with statistically-significant downward trend would be an important next step in research.

6. Summary and Conclusions

(Question 1)… ANOVA testing allowed us to ascertain that Location played the biggest role in determining average means of Heating and Cooling Degrees. This is true for Nevada, Arizona, and California minus California's Cooling degree mean. Which we believe was different due to data set being smaller compared to the other data sets from Nevada and Arizona which have a couple of more decades worth of data. We didn't want to stop here though and decided that further analysis was needed, this time testing the data with time series analysis… (Question 3)

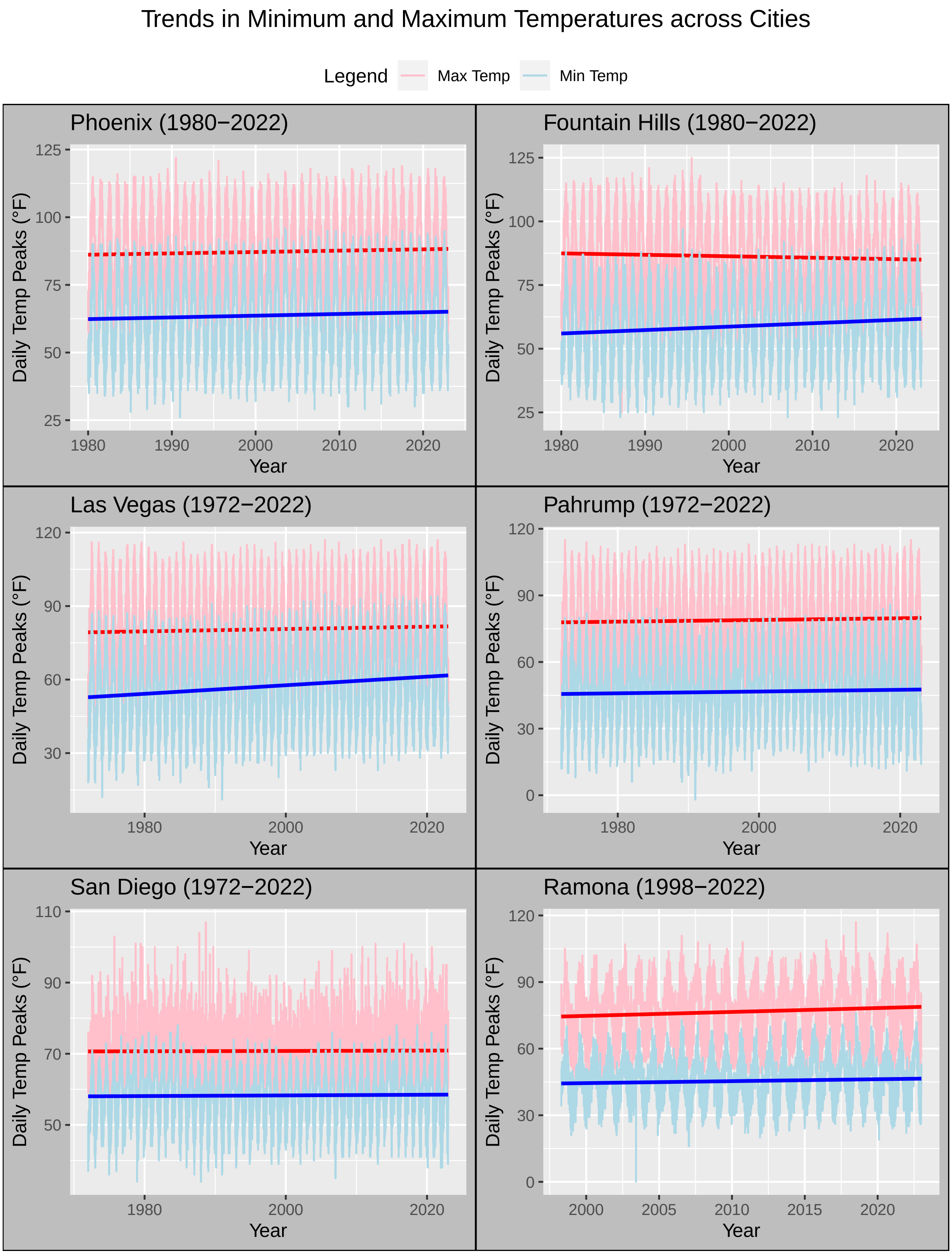


Figure 24: Trends in Minimum and Maximum Temperatures across Cities

7. References

US EPA, O. (2016, July 1). Climate Change Indicators: Heating and Cooling Degree Days [Reports and Assessments]. https://www.epa.gov/climate-indicators/climate-change-indicators-heating-and-cooling-degree-days