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

Candela Cerpa & Samantha Pace & Vinny Whatley

2023-12-13

Contents

1. Rationale and Research Questions **4**

2. Dataset Information **4**

3. Data Wrangling **5**

Heating and Cooling Degree Days Calculations 5

4. Exploratory Analysis **6**

5. Analysis **9**

Question 1: Has there been an increase in the number of heating and cooling days in the U.S. Southwest from 1980 to 2022? 9

Question 2: Is there a difference in heating and cooling days in urban and rural spaces in the U.S. Southwest over time? 12

Question 3: How have minimum and maximum temperatures changed over time? 14

6. Summary and Conclusions **18**

7. References **19**

List of Tables

1 Results for the Seasonal Mann Kendall test for Heating Degree Days 10
2 Results for the Seasonal Mann Kendall test for Cooling Degree Days 10
3 ANOVA Summary Table for Nevada Heat_Degrees 12
4 ANOVA Summary Table for Nevada Cool_Degrees 12
5 ANOVA Summary Table for Arizona Heat_Degrees 12
6 ANOVA Summary Table for Arizona Cool_Degrees 13
7 ANOVA Summary Table for California Heat_Degrees 13
8 ANOVA Summary Table for California Cool_Degrees 13
9 Results for the Seasonal Mann Kendall test on minimum and maximum temperatures 16

List of Figures

1 Average Tempearture for Phoenix, Las Vegas, San Diego 6
2 Phoenix Daily Maiximum and Minimum Temperatures 7
3 Number of Phoenix's Heating and Cooling Degree Days, 1980 7
4 Number of Phoenix's Heating and Cooling Degree Days, 2022 8
5 Comparison of Heat Degree Days in Phoenix in 1980 and 2022 8
6 Comparison of Cool Degree Days in Phoenix in 1980 and 2022 9

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 has 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 increase, so do the number of heat-related illnesses.

The U.S. has not escaped these impact. The southwest region of the US which we defined as Arizona, California, and Nevada has witnessed an uptick in droughts. Nevada which utilizes water from Lake Me has seen record lows in recent years cautioning them to adopt water conservation strategies while California has been no stranger to forest fires. We plan to do this by analyzing data from the National Centers for Environmental Information, using max temperature and minimum temperature. Thus, 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

Our datasets were retrieved from NOAA's National Centers for Environmental Information (NCEI), specifically from the GHCN (Global Historical Climatology Network). 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. 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.

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, PRCP, PSUN, TMAX, TMIN, TAVG, TSUN, and created columns for Month, Year, and DayofYear 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:

Heating degree days = 65 - daily average temperature

Cooling degree days = daily average temperature - 65

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.

Describe how you wrangled your dataset in a format similar to a methods section of a journal article.

To wrangle our data, we: * Read in the .csv's * Filtered by location and created new files for each location * Kept certain columns (name, date, precipitation, sun coverage %, max temp, min temp, average tem, total sun) * Format the date column into a date object * Changed the station name to the city name * Created a column for month, year, and day of year by extracting the information from the existing date column * Updated the column names to the full title ** Check the NAs with the summary function * Saw that some cities have NAs in max, min, and average temperature * Interpolate gaps in min and max using linear interpolation in Fountain Hills, Pahrump, and Ramona * Calculate the average daily temperature and fill in the existing column * Created separate columns for heating and cooling degrees and calculated the values ** As we created the two columns, we filled them with 0s ** We calculate heating degrees by indexing the mean temperatures that are below 65F then subtracting the mean temp from 65 ** We calculate cooling degrees by indexing the mean temperatures that are above 65F then subtracting 65 from the mean temp

Add a table that summarizes your data structure (variables, units, ranges and/or central tendencies, data source if multiple are used, etc.). This table can be made in markdown text or inserted as a kable function in an R chunk. If the latter, do not include the code used to generate your table.

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 to be horizontal with no major changes in slope.

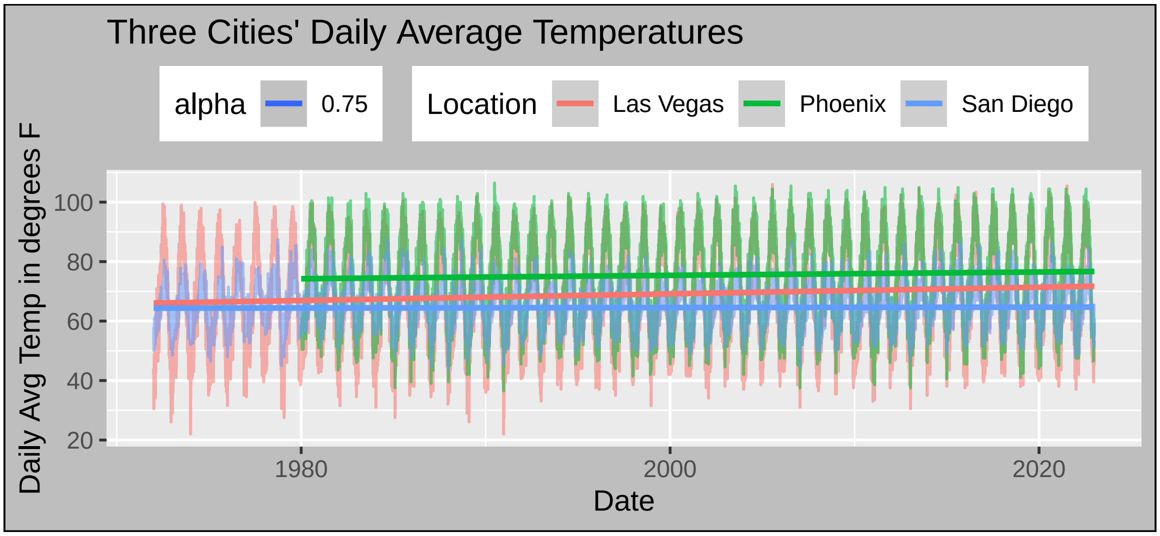


Figure 1: Average Tempearture for Phoenix, Las Vegas, San Diego

For the remainder of the exploratory analysis section, we focused on a single city - Phoenix, Arizona - in order 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 show that there has been a modest increase in both over time. The black horizontal line on 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 degree days than heating.

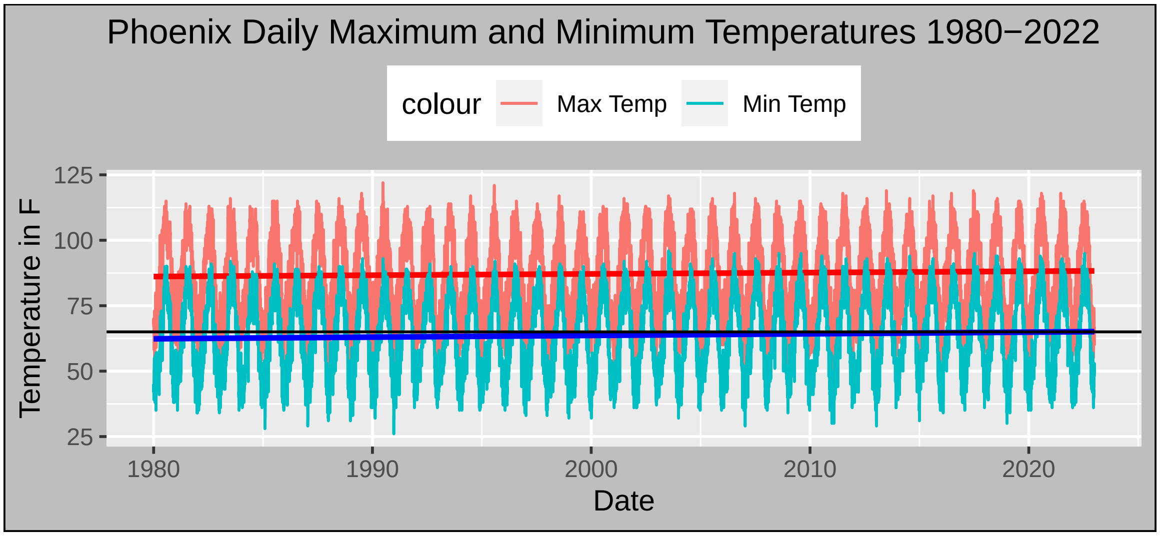


Figure 2: Phoenix Daily Maiximum and Minimum Temperatures

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 days in March, April, and October and a decrease in the number of heating days in February and March.

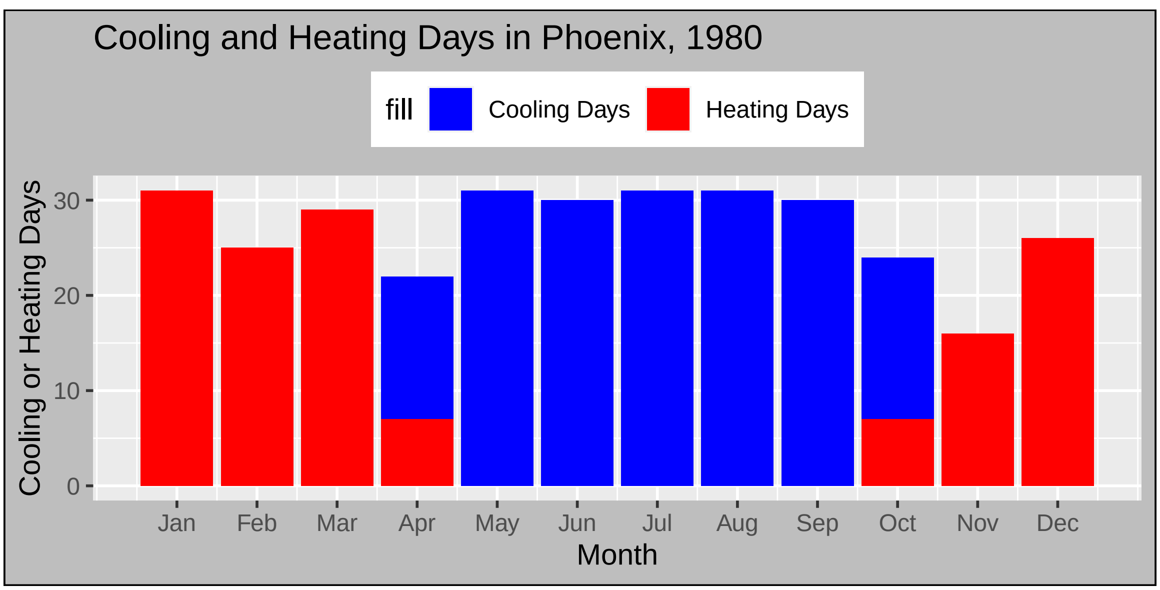


Figure 3: Number of Phoenix's Heating and Cooling Degree Days, 1980

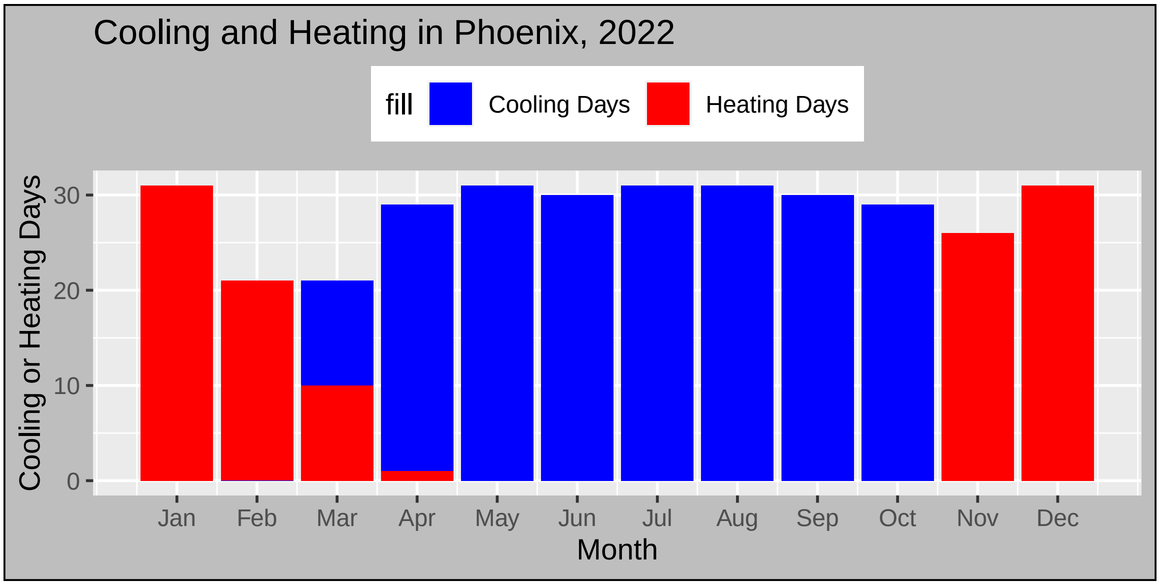


Figure 4: Number of Phoenix's Heating and Cooling Degree Days, 2022

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 is 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 heat 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.

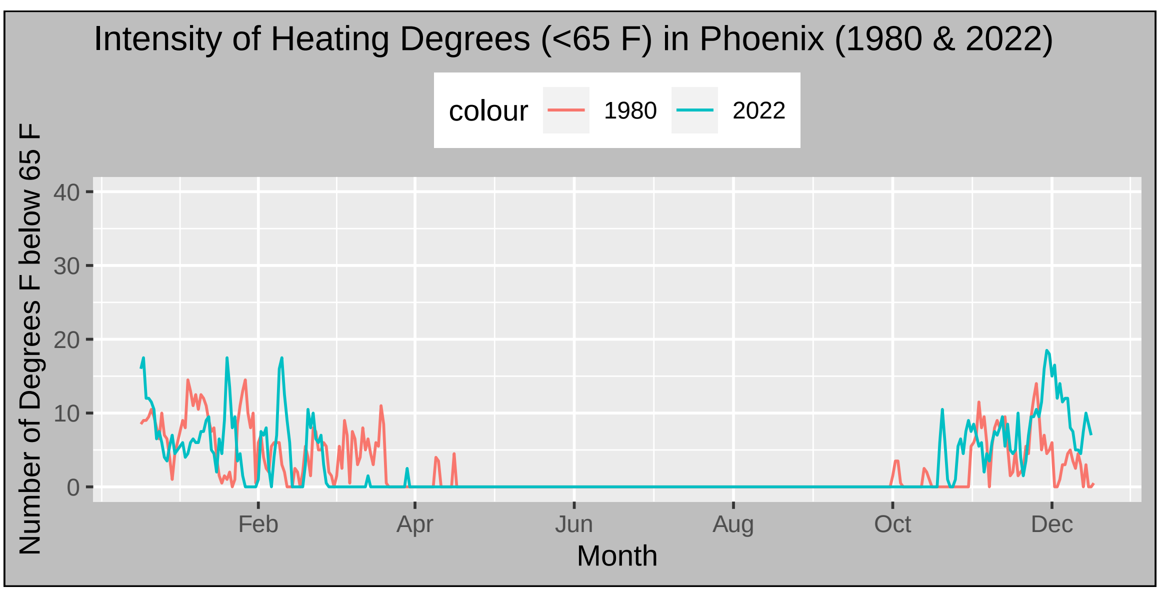


Figure 5: Comparison of Heat Degree Days in Phoenix in 1980 and 2022

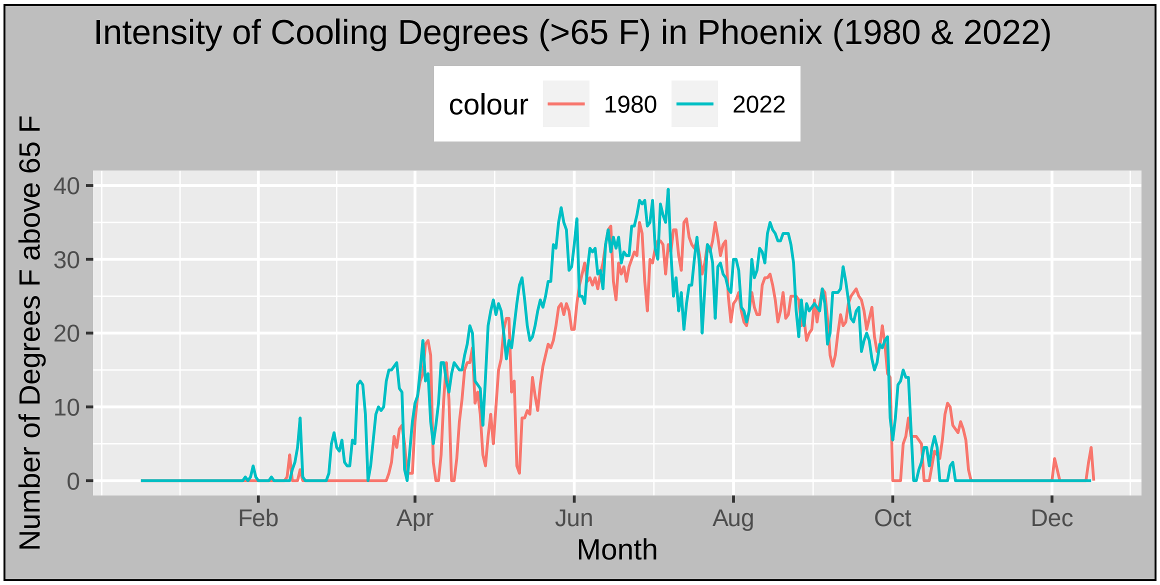
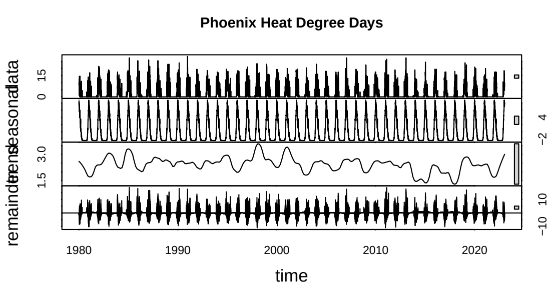


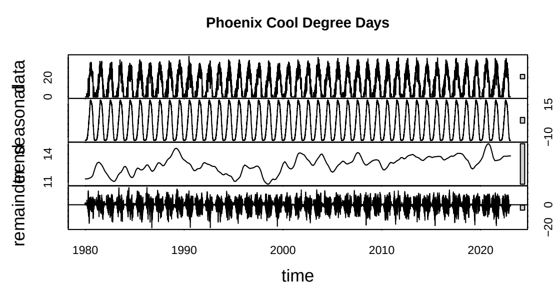
Figure 6: Comparison of Cool Degree Days in Phoenix in 1980 and 2022

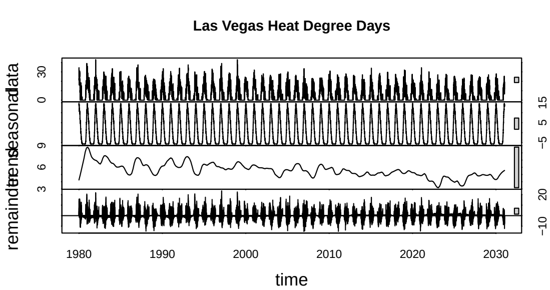
5. Analysis

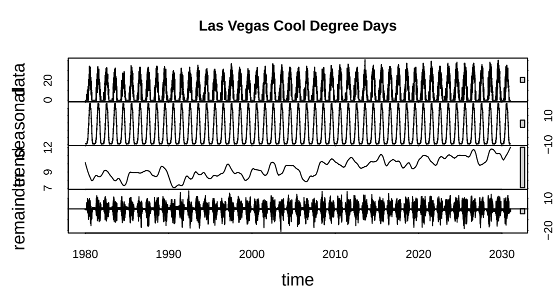
Question 1: Has there been an increase in the number of heating and cooling days in the U.S. Southwest from 1980 to 2022?

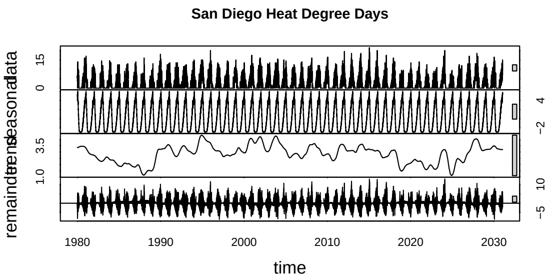
We used a time series analysis for the heating and cooling degree days for each of the three major cities to address this research question. We made univariate time series objects for the heating and cooling degree days for each city, and then decomposed each time series object to separate out the seasonality,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

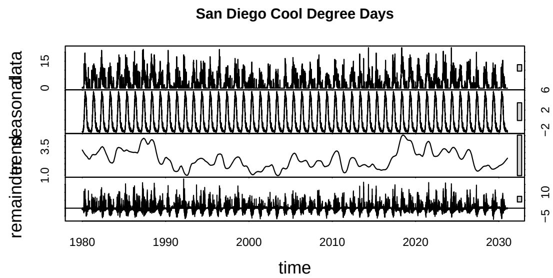












The Seasonal Mann Kendall test was performed on each of the city's heating and cooling degree days data to analyze if the data is stationary. The null hypothesis for the Seasonal Mann Kendall test is that the data is stationary, meaning there is no significant trend over time. The alternative hypothesis 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 1. 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 only Phoenix and Las Vegas. Las Vegas has the steepest slope.

Table 1: 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 2 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 higher the alpha level of 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. The tau values are the slope of the trend line, and indicate if the trend is positive or negative. For both Phoenix and Las Vegas, the slope is positive, meaning there is a positive trend over time in cooling degree days in these locations.

Table 2: 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 question 2, we thought it was best to perform ANOVA testing. This would allow us the check the means of the "Heat Degrees" and "Cool 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 "Heat Degrees" and "Cool Degrees" . But to get an accurate picture we thought it would be best to take the means of the Heat 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, Heat Degrees, and Cool Degrees. Then running 2 anova test for each of those 3 states, the first test having heat degrees as the dependent variable and the second test having cool degrees as dependent variable.

After that using the anova data for heat and cool 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 heat temperature. While the same story was told when looking at the average cool 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 heat degree's location was the statically significant contributor but for cool 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 3: ANOVA Summary Table for Nevada Heat_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 4: ANOVA Summary Table for Nevada Cool_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 5: ANOVA Summary Table for Arizona Heat_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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	30890	789816.57781	25.56868	NA	NA

Table 6: ANOVA Summary Table for Arizona Cool_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

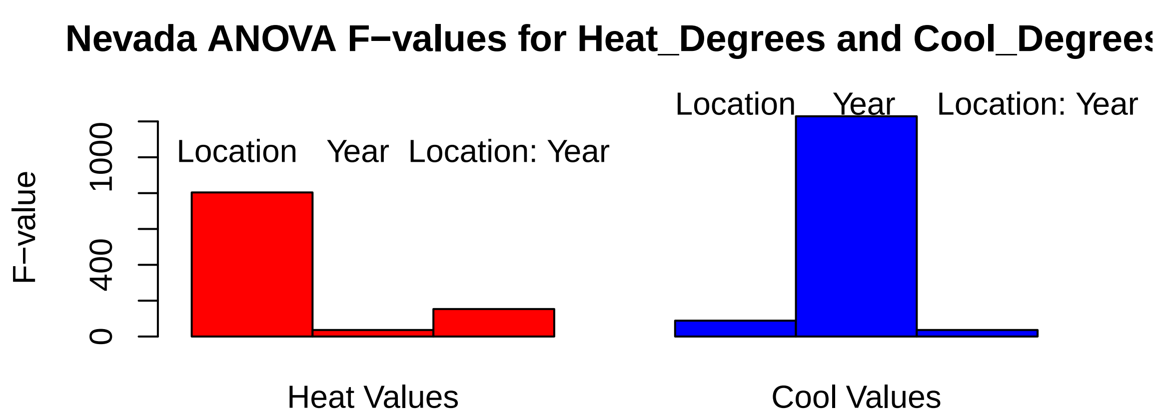
Table 7: ANOVA Summary Table for California Heat_Degrees

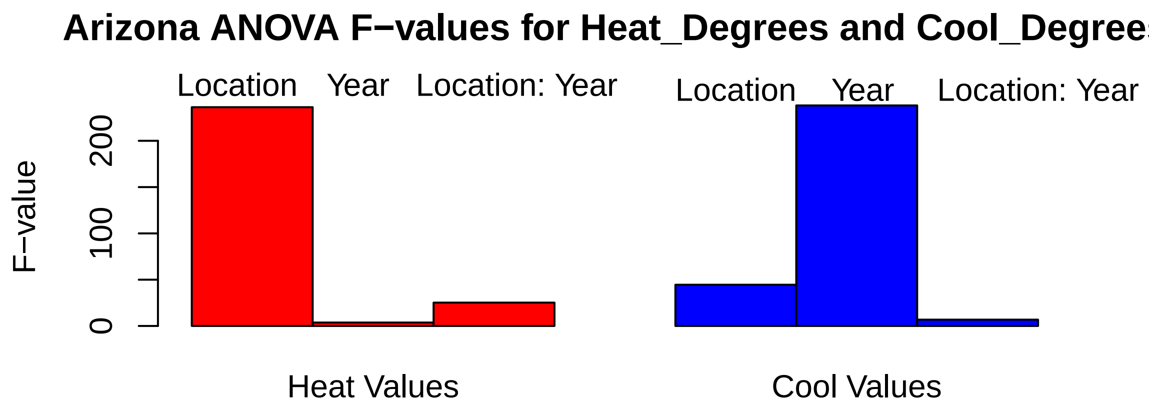
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Location	1	5.854260e+04	58542.604743	1870.433248	0.0000000
Year	1	3.572846e+03	3572.845863	114.152244	0.0000000
Location:Year	1	6.610809e+00	6.610809	0.211215	0.6458223
Residuals	18105	5.666676e+05	31.298954	NA	NA

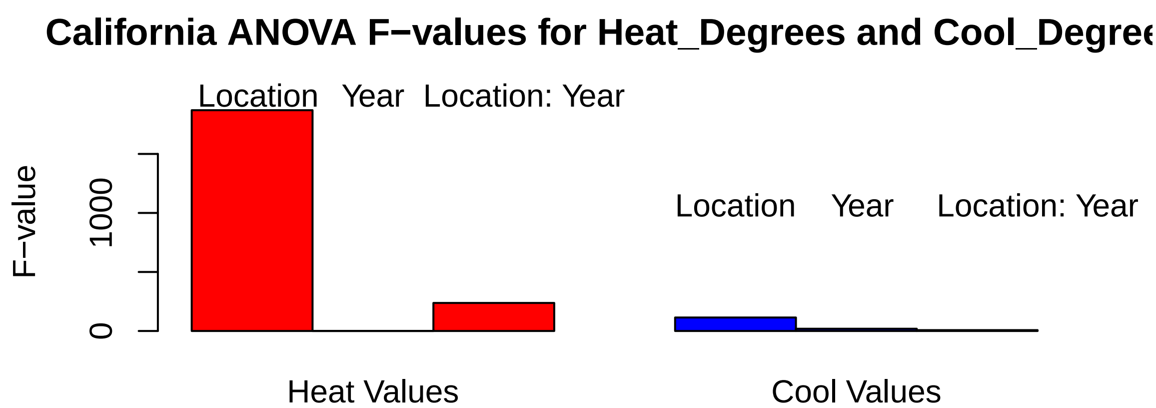
Table 8: ANOVA Summary Table for California Cool_Degrees

	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.



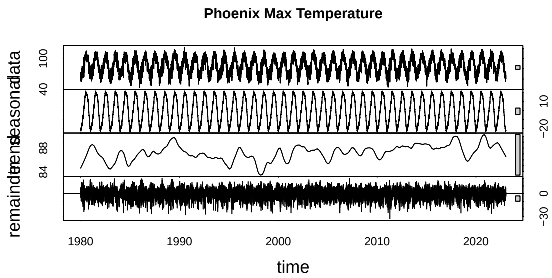


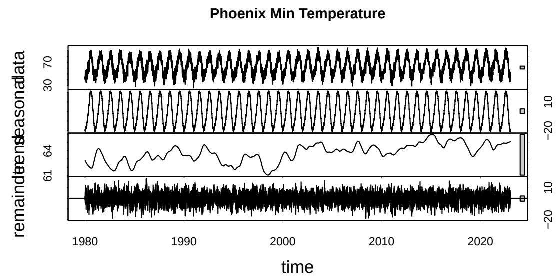


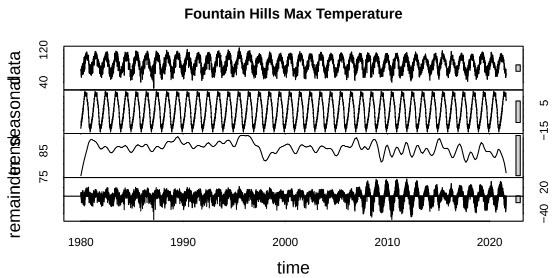
In summary, with anova testing we learned that biggest contributor to average heat and cooling temperatures for each can be attributed to solely location. Requiring us to do further analysis this time instead with time-series.

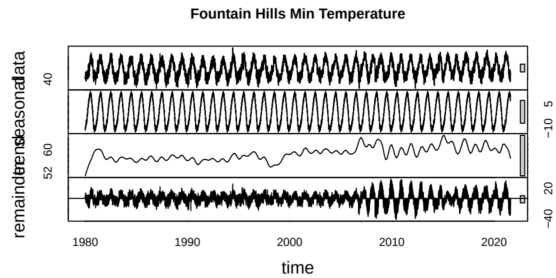
Question 3: How have minimum and maximum temperatures changed over time?

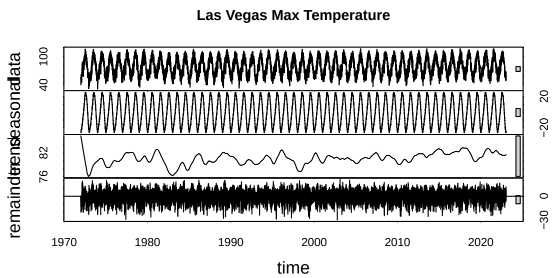
To address the question of whether minimum and maximum temperatures have changed over time, we conducted a time series analysis on every city. We first created univariate time series objects for each measure for each site. We then decomposed each time series to review the trends and seasonality present.

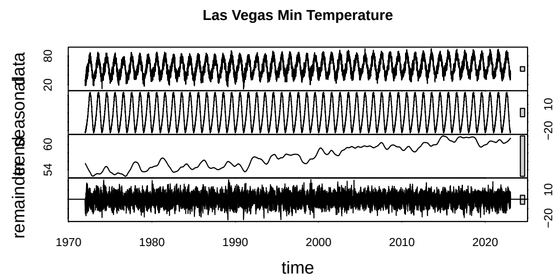

Phoenix Max Temperature
data
seasonal
trend
remainder
100
40
10
−20
88
84
0
−30
1980
1990
2000
2010
2020
time

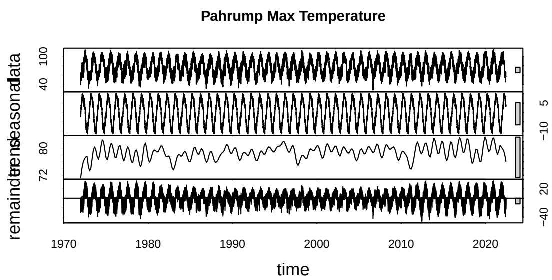

Phoenix Min Temperature
data
seasonal
trend
remainder
70
30
10
−20
64
61
10
−20
1980
1990
2000
2010
2020
time

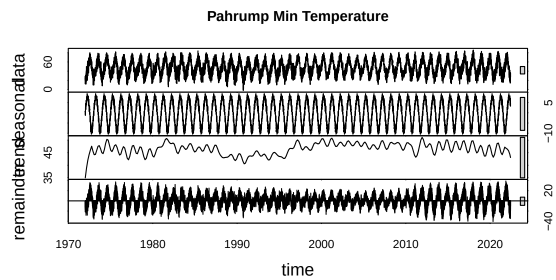

Fountain Hills Max Temperature
data
seasonal
trend
remainder
120
40
5
−15
85
75
20
−40
1980
1990
2000
2010
2020
time

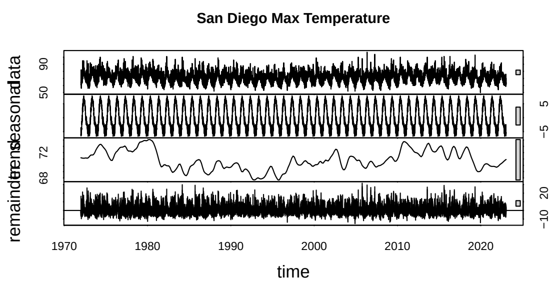

Fountain Hills Min Temperature
data
seasonal
trend
remainder
40
5
−10
60
52
20
−40
1980
1990
2000
2010
2020
time

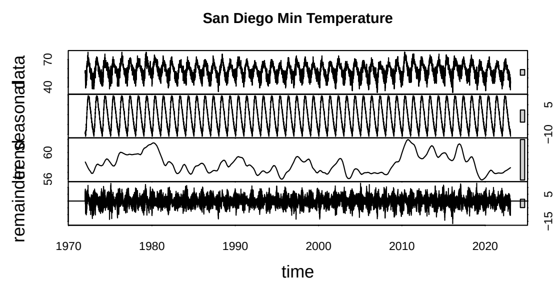

Las Vegas Max Temperature
data
seasonal
trend
remainder
100
40
20
−20
82
76
0
−30
1970
1980
1990
2000
2010
2020
time

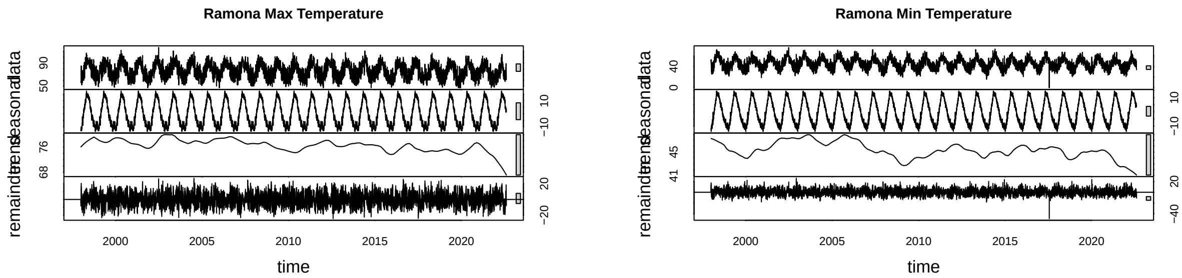

Las Vegas Min Temperature
data
seasonal
trend
remainder
80
20
10
−20
60
54
10
−20
1970
1980
1990
2000
2010
2020
time


Pahrump Max Temperature
data
seasonal
trend
remainder
100
40
5
−10
80
72
20
−40
1970
1980
1990
2000
2010
2020
time


Pahrump Min Temperature
data
seasonal
trend
remainder
60
0
5
−10
45
35
20
−40
1970
1980
1990
2000
2010
2020
time


San Diego Max Temperature
data
seasonal
trend
remainder
90
50
5
−5
72
68
20
−10
1970
1980
1990
2000
2010
2020
time


San Diego Min Temperature
data
seasonal
trend
remainder
70
40
5
−10
60
56
5
−15
1970
1980
1990
2000
2010
2020
time

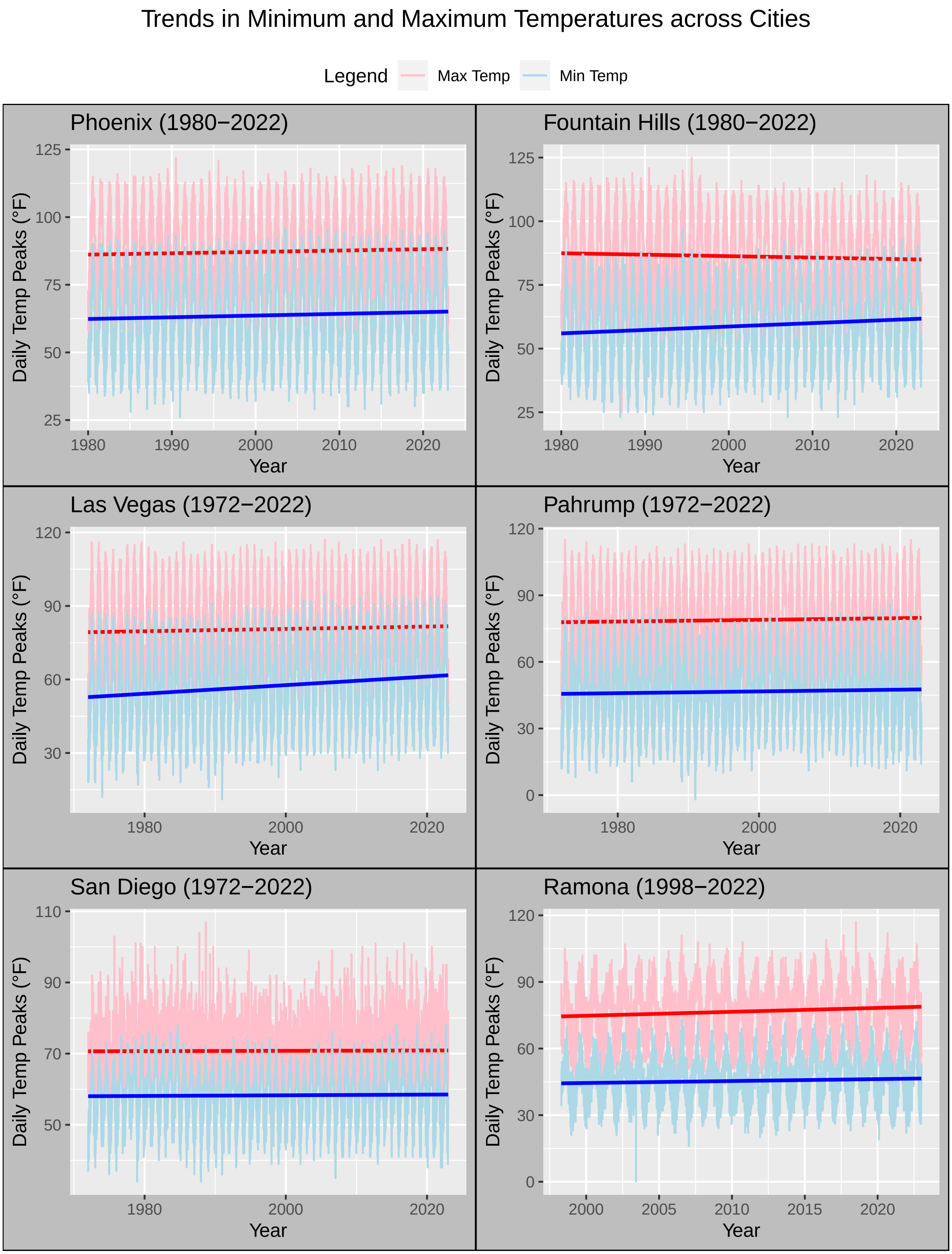


We then conducted a Seasonal Mann Kendall test to test whether the hypothesis that the data is stationary. This would mean that the data exhibit no statistically-significant upward or downward trend over time, meaning that there is no change in the temperature minimum and maximums over time.

The results of the Seasonal Mann Kendall test show us that all trends are statistically significant, so none of the data are stationary. The tau value shows the slope of the trend. 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.

Table 9: 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



The graphs above show the trends increasing or slightly decreasing in the case of .

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)

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