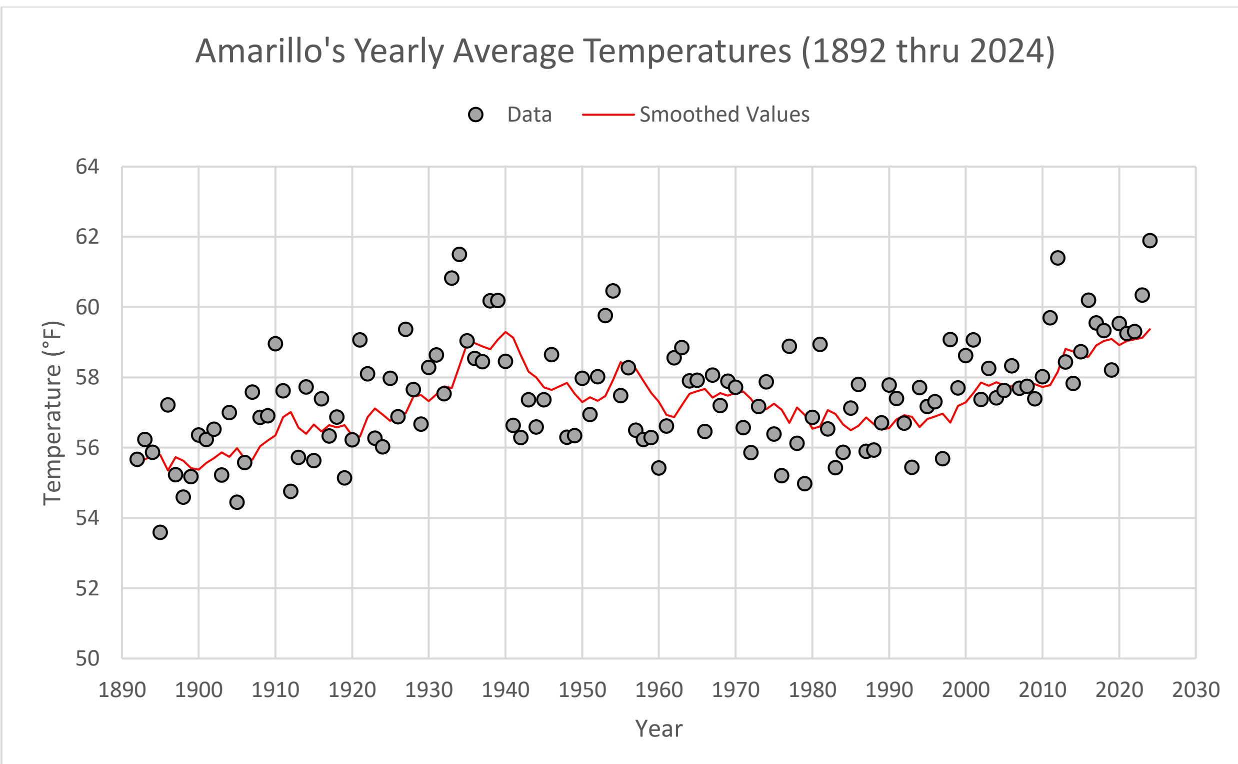
Amarillo's Yearly Average Temperatures (1892 – 2024)

By Robert Ashcraft

I recently downloaded Amarillo's yearly average temperatures from NOWData (a database which is maintained by the National Weather Service) and imported the data into an Excel worksheet. When plotted, the value for 1947 appeared to be an outlier. After downloading the daily temperature data for 1947, I found that roughly 1/3 of the data for that year were missing. Fortunately, 1947 was the only year with missing temperature data, and it is not included in the analysis that follows.

Step 1 – Time Series Plot

The first step in the analysis was a time series plot of the yearly average temperatures, with exponentially smoothed values shown in red.

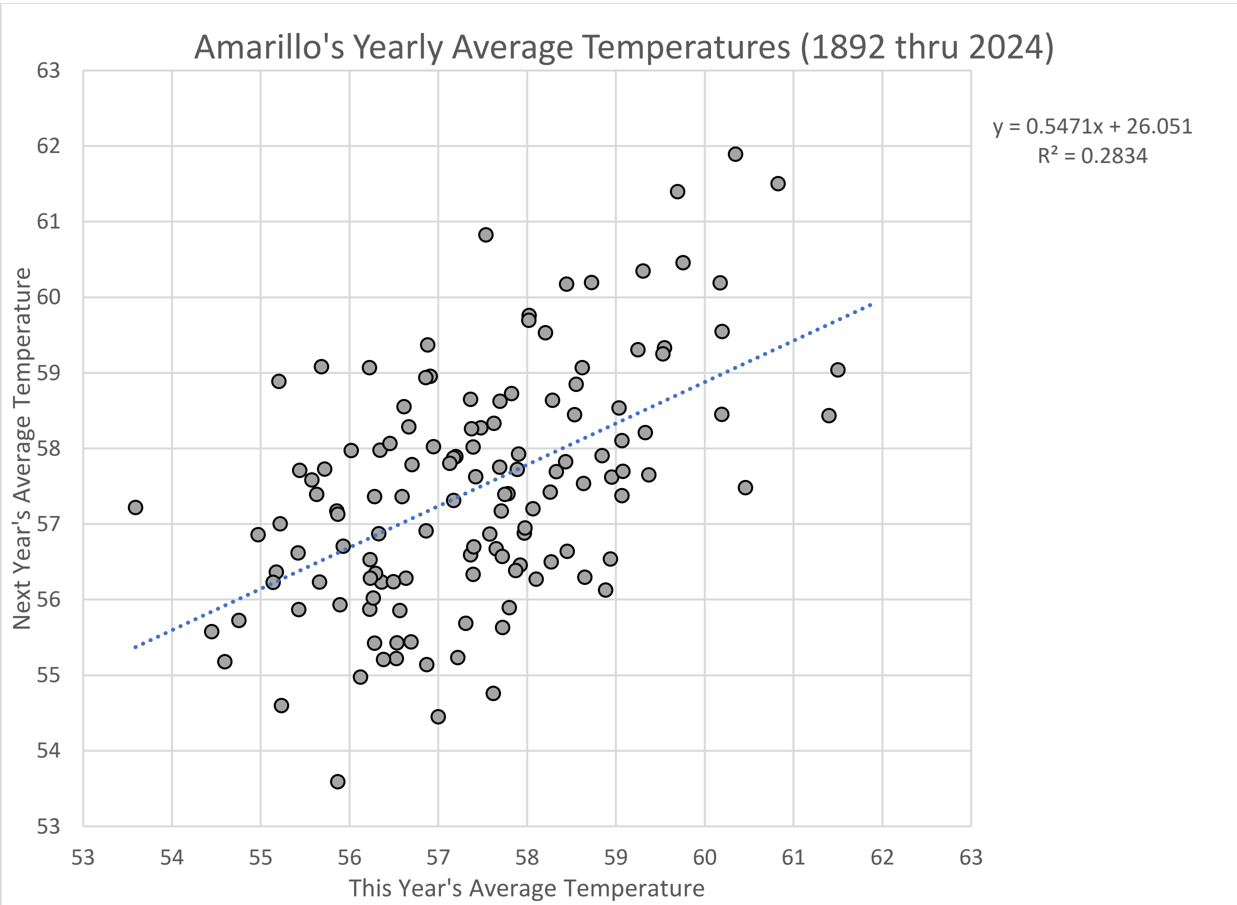


As wild as the weather can be here in Amarillo, I was surprised to see that the yearly average temperatures have a range of only 8°F. Note that 2024 is now the warmest year on record, pushing 1934 into second place.

There was a noticeable warming trend from 1892 until the mid 1930's, followed by a cooling trend until 1980, followed by a warming trend through the present. The first warming trend lasted about 45 years. The cooling trend also lasted about 45 years. If our current warming trend lasts 45 years, it will end in 2025.

Step 2 – Lag 1 Plot

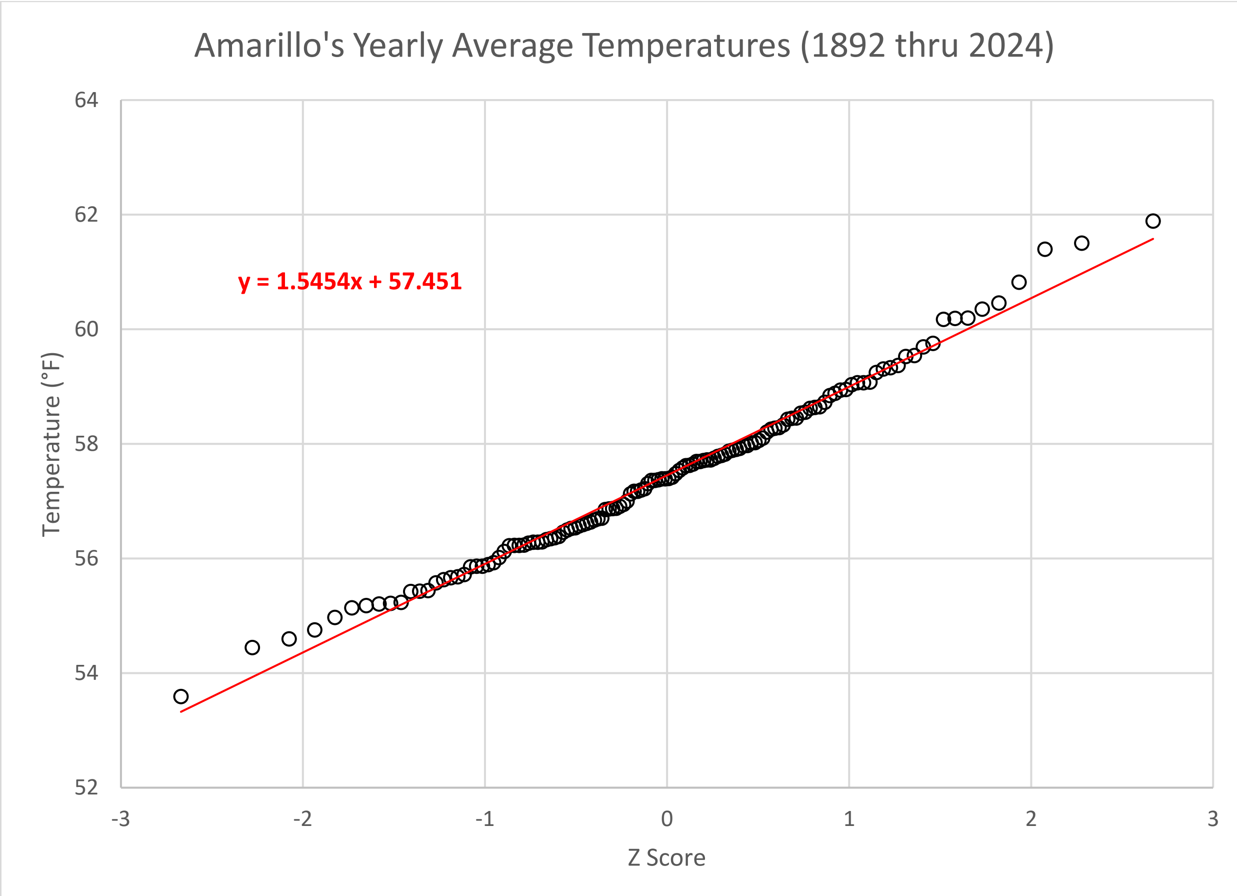
A Lag 1 plot was made to see if any correlation existed between successive yearly average temperatures. The X-axis in this plot is the current year's average temperature, while the Y-axis is the next year's average temperature, i.e., we're plotting the points (T_i, T_{i+1}). If the data points are independent, the plot will appear to be a roughly circular set of random points.



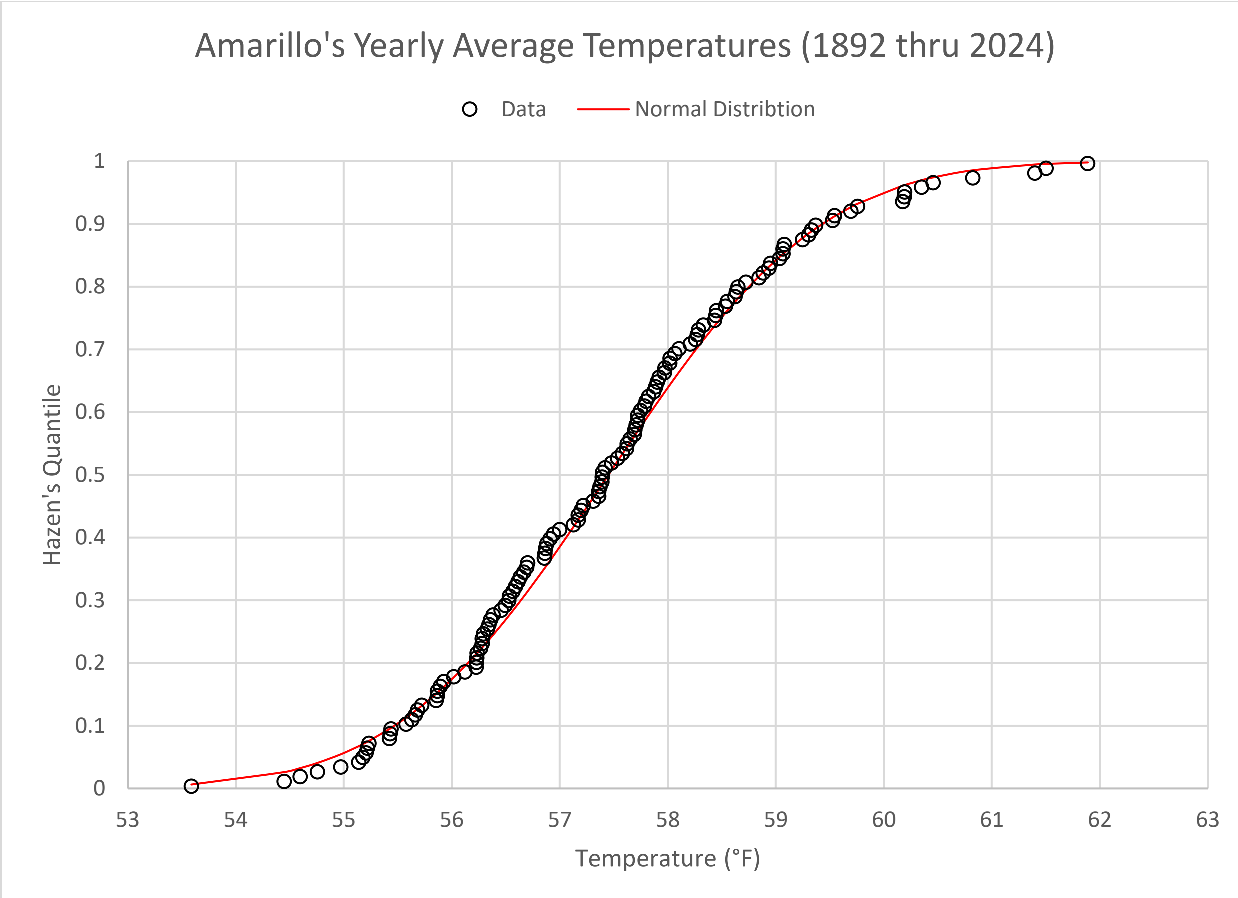
These points are fairly random, but they do have a slight positive slope, indicating a weak correlation between the average temperature for a year and the average temperature for the next year. None of the points appear to be outliers.

Step 3 – Distribution Analysis

The next step in the analysis was to look at the distribution of the yearly average temperatures. Were they approximately normal? Or something else? To check for normality, a plot was made of the sorted temperatures on the Y-axis and the corresponding z scores on the X-axis (see Appendix). The more linear the points in such a graph, the more normally distributed the points are. Also, the slope and intercept of the trendline will be good approximations to the maximum likelihood estimates of the mean and standard deviation of the distribution. Here is the plot, which includes the data points and the trend line.



Except for a few points at the extremes, the data points are very linear. The trend line says the mean of the distribution is approximately 57.451° and the standard deviation is 1.5454°. Actual calculation of the maximum likelihood estimators gave 57.451011° and 1.545979°. The following plot shows the cumulative distribution of the yearly average temperatures with the corresponding normal distribution. The agreement between the data points and the distribution is excellent.



<u>Conclusions</u>

Amarillo's yearly average temperatures show periods of warming and cooling trends. The last cooling period (from 1935 to 1980) lasted about 45 years. If the current warming trend lasts that long, it will end in 2025. The yearly average temperatures have had a range of only 8°F.

The yearly average temperatures are not completely random; there is a weak correlation between successive years.

The yearly average temperatures are very close to being normally distributed, with parameters

$\mu = 57.45101°$

$\sigma = 1.54598°$

APPENDIX

Why Hazen's Quantile?

In the plot of the cumulative distribution for Amarillo's yearly average temperatures, I used Hazen's quantile on the Y-axis. Hazen's quantile was also used to obtain the z scores used in the linearized normal distribution plot.

When constructing a cumulative distribution for N points, the quantile most commonly used is

$$q_i = \frac{i}{N} \qquad i = 1,2, \dots, N$$

A z score does not exist for the quantile q_N = 1. We have a similar problem if we use this quantile

$$q_i = \frac{(i-1)}{N} \quad i = 1,2, \dots, N$$

A z score does not exist for the quantile $q_1 = 0$. The solution is to use Hazen's quantile.

$$q_i = \frac{(i-0.5)}{N} \quad i = 1,2, \dots, N$$

Now there will be a z score for each quantile, and it will be possible to make the linearized normal distribution plot.

As mentioned in the article, when we use Hazen's quantiles, the slope and intercept of the best fit line are good approximations to the maximum likelihood estimates of the normal distribution's parameters.

If we want unbiased estimates of the parameters, the formula to be used for the quantiles is

$$q_i = \frac{(i-0.375)}{N+0.25} \quad i = 1,2, \dots, N$$