Weather Analysis of Akron, OH, and Asheville, NC

# Introduction

The purpose of this study is to analyze the historical weather data from Akron, OH, and Asheville, NC, and determine any differences between the two. Specifically, this study will look at differences in average temperature.

This study will look at the overall average temperature and test whether Asheville’s average temperature is greater than 5 degrees warmer than Akron’s. Since warmer weather is normally desirable in the winter, but not during the summer, tests specifically for winter and summer will also be performed. A test will be performed to determine if Asheville’s average winter temperature is greater than 8 degrees warm that Akron, and a test will be performed to determine if Asheville’s average summer temperature is less than 2 degrees warmer then Akron.

# Overview of the Data

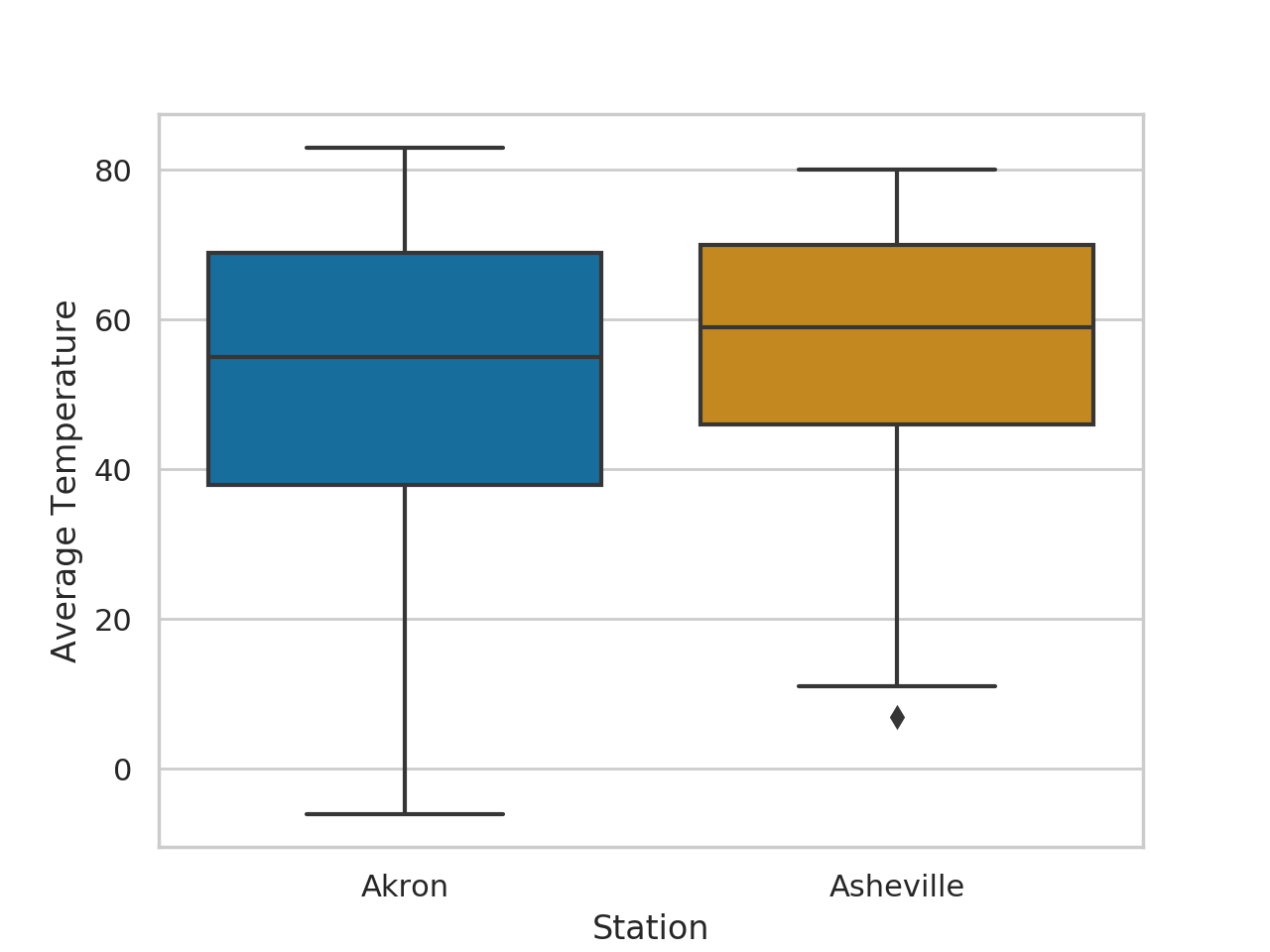
That data for this study was obtained from the National Oceanic and Atmospheric Administration (NOAA). The NOAA dataset that was used is the Global Historical Climate Network Daily dataset. Daily weather data was obtained from each city starting from 01/01/2014 and going to 12/31/2017. The data for Akron, OH is from NOAA station GHCND:USW00014895 at the Akron Canton Airport in Akron, OH. The data for Asheville, NC is from NOAA station GHCND:US1NCBC0110 at the Asheville Airport in Asheville, NC.

The data files used contained one record for each day from 01/01/2014 to 12/31/2017. The fields that were used from the dataset are:

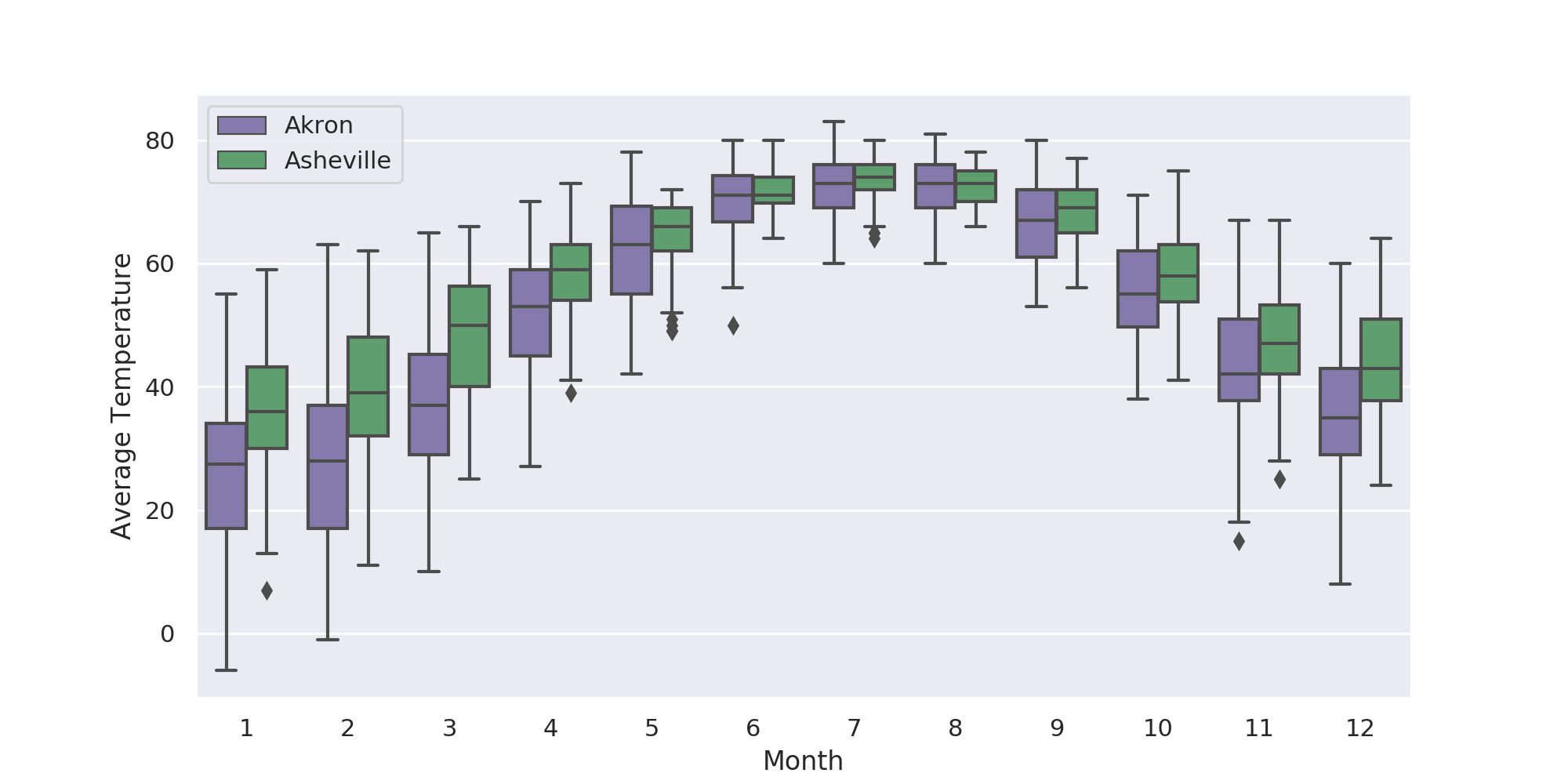
|  |  |  |
| --- | --- | --- |
| **Field Name** | **Data Type** | **Description** |
| STATION | String | Station ID |
| Date | Date | Date of the weather record |
| TAGV | Integer | Average temperature in Fahrenheit (average of hourly values) |

Calculating descriptive statistics for the two data sets we obtain the following:

|  |  |
| --- | --- |
| **Descriptive Statistics for Akron Data** | |
|  | TAVG |
| Observations | 1461 |
| Mean | 51.930869 |
| Median | 55 |
| Standard Deviation | 19.175898 |
| Minimum | -6 |
| Maximum | 83 |



|  |  |
| --- | --- |
| **Descriptive Statistics For Asheville Data** | |
|  | TAVG |
| Observations | 1,461 |
| Mean | 56.988364 |
| Median | 59 |
| Standard Deviation | 14.726939 |
| Minimum | 7 |
| Maximum | 80 |



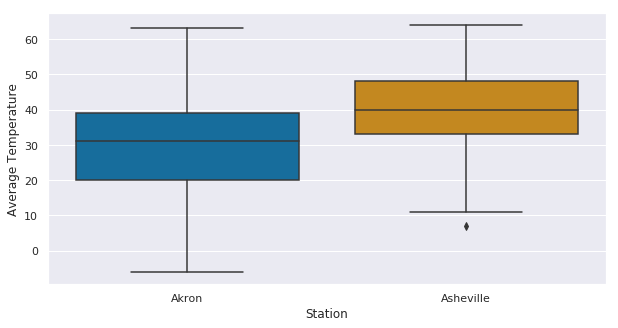
After looking at the descriptive statistics, and the box plots we can make the following observations:

* The Asheville sample mean is higher than the Akron sample mean.
* The Asheville sample standard deviation is smaller than the Akron standard deviation.
* Asheville appears to be warmer in the winter, but close to the same temperature in the summer.

Normally warmer temperatures are desirable in the winter, but not desirable in the summer. Because of that, descriptive statistics for just the winter months (December, January, February) and just summer months (June, July, August) where also calculated.

## Winter

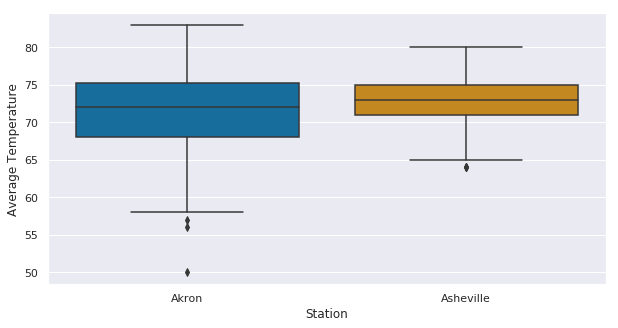
|  |  |
| --- | --- |
| **Descriptive Statistics for Akron Winter** | |
|  | TAVG |
| Observations | 361 |
| Mean | 29.700831 |
| Median | 31 |
| Standard Deviation | 13.16368 |
| Minimum | -6 |
| Maximum | 63 |



|  |  |
| --- | --- |
| **Descriptive Statistics For Asheville Winter** | |
|  | TAVG |
| Observations | 361 |
| Mean | 39.936288 |
| Median | 40 |
| Standard Deviation | 10.420004 |
| Minimum | 7 |
| Maximum | 64 |
|  |  |

## Summer

|  |  |
| --- | --- |
| **Descriptive Statistics for Akron Summer** | |
|  | TAVG |
| Observations | 368 |
| Mean | 71.774457 |
| Median | 72 |
| Standard Deviation | 5.222118 |
| Minimum | 50 |
| Maximum | 83 |



|  |  |
| --- | --- |
| **Descriptive Statistics For Asheville Summer** | |
|  | TAVG |
| Observations | 368 |
| Mean | 72.633152 |
| Median | 73 |
| Standard Deviation | 3.276319 |
| Minimum | 64 |
| Maximum | 80 |

# Methods Used

In this study the distribution, mean, and variance of the population is unknown. However, our sample size is 1,461 for the test including all months, 361 for the test only including the winter months, and 368 for the test only including summer months. All of these tests have a sample size large enough to assume a normal distribution by the central limit theorem. The two sample Z test was used to to the test hypothesis (Devore 2012). A 95% confidence level was used for all tests.

The datasets used for this study, along with all of the calculations and analysis, and this report are available at <https://github.com/EricMaibach/AkronAshevilleWeatherAnalysis>.

# Analysis

## Test with all months

The first test performed includes all months and tested the following hypothesis:

H0: Akron average temperature – Asheville average temperature = -5

Ha: Akron average temperature – Asheville average temperature < -5

The equation used to calculate the test statistic is:

The calculated test statistic is:

z = -0.090892

This is a lower-tailed test with a significance level of 0.05, so our hypothesis should be rejected if our test statistic is less than -1.645. Our test statistic of -0.090892 is not less than -1.645, so we fail to reject the null hypothesis. The test is inconclusive.

## Test with only winter months (December, January, February)

The second test just looked at the data from the winter months (December, January, February) and tested the following hypothesis:

H0: Akron average winter temperature – Asheville average winter temperature = -8

Ha: Akron average winter temperature – Asheville average winter temperature < -8

The equation used to calculate the test statistic is:

That calculated test statistic is:

Z = -2.530

This is a lower-tailed test with a significance of 0.05, so our hypothesis should be rejected if our test statistic is less than -1.645. Our test statistic of -2.530 is less than -1.645, so we reject the null hypothesis, the Akron average winter temperature – Asheville average winter temperature < -8.

## Test with only summer months (June, July, August)

The third test just looked at the data from the summer months (June, July, August) and tested the following hypothesis:

H0: Akron average summer temperature – Asheville average summer temperature = -2

Ha: Akron average summer temperature – Asheville average summer temperature > -2

The equation used to calculate the test statistic is:

The calculated test statistic is:

Z = 3.551

This is a upper-tailed test with a significance of 0.05, so our hypothesis should be rejected if our test statistic is greater than 1.645. Our test statistic of 3.551 is greater than 1.645, so we reject the null hypothesis, the Akron average summer temperature – Asheville average summer temperature > -2

# Conclusion

The test that was performed with the data from all months of the year was inconclusive. However, when the test was performed with just the winter months the null hypothesis was rejected. Also, when the test was performed with just the summer months, the null hypothesis was rejected We cannot make any conclusions about differences in temperature across the entire year. We can say with 95% confidence that during the winter months the Asheville average temperature is greater than 8 degrees warmer than Akron. We can also say with 95% confidence that during the summer months the Asheville average temperature less than 2 degrees warmer than Akron.

# References

“9.1 z Tests and Confidence Intervals for a Difference Between Two Population Means”

*Probability and Statistics for Engineering and the Sciences*, by Jay L. Devore, Eighth Edition, Brooks/Cole, 2012, pp. 346-354