5-2023

Hawk Mountain Raptor Migration Phenology's Relation to Weather

Eric Burgos
Kutztown University of Pennsylvania, eburg118@live.kutztown.edu

Hawk Mountain Raptor Migration Phenology’s Relation to Weather

A Graduate Thesis

Presented to the Faculty of

The Department of Computer Science and Information Technology

Kutztown University of Pennsylvania

Kutztown, Pennsylvania

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

By

Eric Burgos

May 2023

Approved:

July 24, 2023 Dale E Parson

(Date) (Adviser)

___July 25, 2023______ ______________________

(Date) (Second Reader)

___July 25, 2023______ ______________________

(Date) (Chair, Department of Computer Science and Information Technology)

Acknowledgements

I would like to thank everyone who has helped me in big and small ways to make this thesis possible. Thank you to my thesis advisor, Dr. Dale Parson, who shares the love of the outdoors and ultimately suggested this research topic during my graduate studies. In the same sentiment, thank you Dr. Laurie Goodrich who allowed me to utilize Hawk Mountain Sanctuary's data and facilitated trips to Hawk Mountain to witness the beautiful migrating raptors.

Thank you as well to Dr. Schwesinger who has taken the time and helped proofread this thesis as well as suggesting changes and updates. I would also like to thank Kutztown University of Pennsylvania alumni Bryan McNally and Tyler Blankenbiller for laying down the initial groundwork in terms of data analysis.

To my friends and family, who have motivated me and pushed me to see this research through until its completion; thank you.

Lastly, to the hawks that soar and migrate via Hawk Mountain's ridge year after year; thank you for letting me observe and study your beautiful fall migration.

Abstract

We have been studying year-round raptor migration phenology across the United States and North America for multiple decades now. Hawk Mountain Sanctuary’s Autumn migration hawk count began in 1934 and is the longest running raptor migration count in the world [2]. A decline in total raptor counts passing through Hawk Mountain’s North Lookout is well documented and much research has already been done in what could be the main causes for this decrease in counts year-over-year. We know that cold front passages have long been associated with autumnal migration in northeastern North America. Using updated analysis techniques, we examined 60 years’ worth of Hawk Mountain migration counts in relation to local climate variables.

The data was aggregated on an autumnal basis and the climate variables of interest were pulled, cleaned and sorted along with our target variable:

the total raptor counts. For numeric non-target attributes, we recorded and visualized many scalar statistical values. Hawk Mountain's temperature data has not been consistently recorded until around 1980, so, we merged NOAA Allentown weather station data for the days we see in the original dataset. Using this data, we were able to get a good understanding of initial correlations between weather attributes. Linear regression model evaluation using the Pearson Correlation Coefficient was run in order to try to find the best combination of predictors in order to predict the movement of the total raptors migrating through Hawk Mountain's north lookout.

Table of Contents

List of Tables and Figures

Chapter 1

1. Introduction

Climate is the long-term pattern of weather in a particular area. In the last century, the climate around the globe is being affected and climate change is an inevitable occurrence for all life on earth. Climate change in the northeast United States means that there has been and there will be an increase in extreme precipitation, higher temperatures, sea level rise and an increase of heat waves [6]. The increase in temperature helps to contribute to the variability and intensity of other weather patterns such as wind speed, cloud coverage and atmospheric pressure. We as humans are doing many things in order to adapt, such as creating provisions to protect infrastructure, emergency preparation, response and recovery [18]. Plants and animals have also been exhibiting phenological changes to adapt to the changing climate.

In the northeast temperate region of the United States birds of prey, specifically raptors, migrate seasonally between wintering and breeding grounds. Hawk Mountain Sanctuary is located on the southernmost ridge of the Appalachian Mountains along the Kittatinny Ridge in eastern Pennsylvania. Here, Hawk Mountain Sanctuary's conservation scientists and volunteers have been counting the passage of raptors on a daily/hourly basis since 1934. This sanctuary is regarded as the oldest detailed archive on the timing and magnitude of migratory raptors in the world [16]. Not only is the count of migrating raptors recorded but also certain weather variables such as temperature, wind speed/direction, visibility and cloud coverage are also recorded.

The main goal and focus here is to provide accurate and intelligible analysis about the correlations and effects of climate change as it related to Hawk Mountain's mission of raptor preservation and research. We will be using 57 years of data provided by Hawk Mountain and also augmenting National Oceanic Atmospheric Administration's (NOAA) weather data from Allentown Airport. This weather station is the closest to Hawk Mountain and this data is necessary due to missing weather data from the observation site.

Firstly, we will begin by aggregating the data on an autumnal basis. We will be following the standard range of dates that Hawk Mountain uses for their fall migration counts which are the dates ranging from 15 August to 15 December [16]. For non-target climate numeric attributes, we will record scalar statistical measures for each autumnal aggregated year. For the same attributes we will get the absolute value of the daily delta change year-over-year to find the variability of the attribute year to year. As recommended by prior research [4], date-adjusted index, regression-based analysis will be done on the data to find trends and correlations in the data. Most of the analysis within the scope of this project will be run using the Python programming language and sourcing from relevant library packages such as pandas, numpy, matplotlib, scikit-learn, seaborn and other helper libraries. Python was used due to its extensive data-analysis-specific libraries which are used in every step of data analysis, from cleaning, processing, modeling and visualizing data.

Chapter 2

2. Background

2.1 What Is Hawk Mountain Sanctuary?

There are thousands of hawk watch sites around the world that count and keep track of migrating hawks. None of these sites have been tracking migrating raptor populations for as long as Hawk Mountain Sanctuary. Hawk Mountain has the oldest archive of migratory raptor timings and counts in the world. Daily counts are recorded as far back as 1934 and hourly counts can be found starting in 1966. Hawk Mountain sits next to the Kittatinny Ridge (see Figure 1) which is the southernmost ridge of the Appalachian Mountains in eastern Pennsylvania [16].

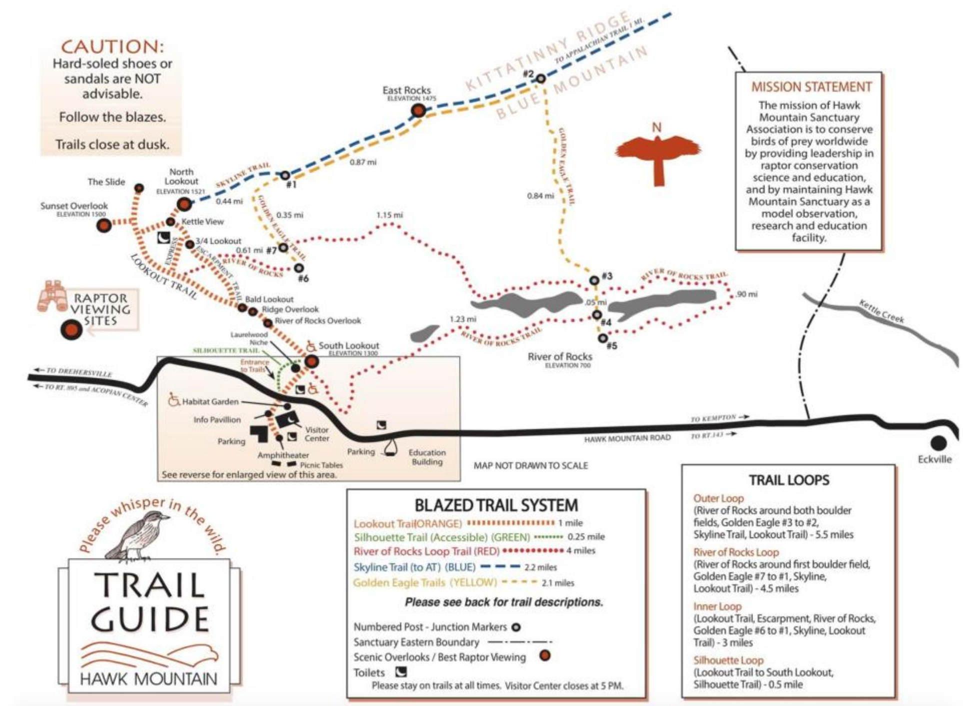

Figure 1 – Hawk Mountain Trail Guide [5]

2.2 What is Autumn Raptor Migration in Northeast United States?

Autumn season in the northeast part of the United States is characterized by colder temperatures, a decrease in sunlight, more low-pressure systems and gusting winds rolling through. This environment signals the start of even colder temperatures come wintertime. This also signals raptors to start their southward migration towards wintering grounds that will be favorable for nesting and a better chance at finding sources of food. This autumnal migration begins in late August and usually ends in early December, which is exactly the time when Hawk Mountain does their yearly autumn migration count and data gathering.

Raptors will fly south utilizing different landmarks which vary depending on the raptor species but the landmark that is used by every group is the Appalachian Mountain chain. Raptors will use this chain of mountain ridges in many ways. First, they act as a visual guide to follow, but most importantly is the air current systems that are generated during the autumn time.

Two methods of travel are generally used by these migrating raptors. One method entails using thermal updrafts that are generated by solar heating of the ridge slopes and valley. A thermal updraft is this rising hot air that raptors will use to migrate around. Raptors can be seen circling, moving up in elevation and then moving on to the next thermal updraft (see Figure 2). The second method that migrating raptors will use is the updraft created along the ridges when horizontal northwest winds strike the north slope of the mountain. Raptors will fly south and will concentrate closer to the treetops if the wind is strong and favorable (see Figure 2).

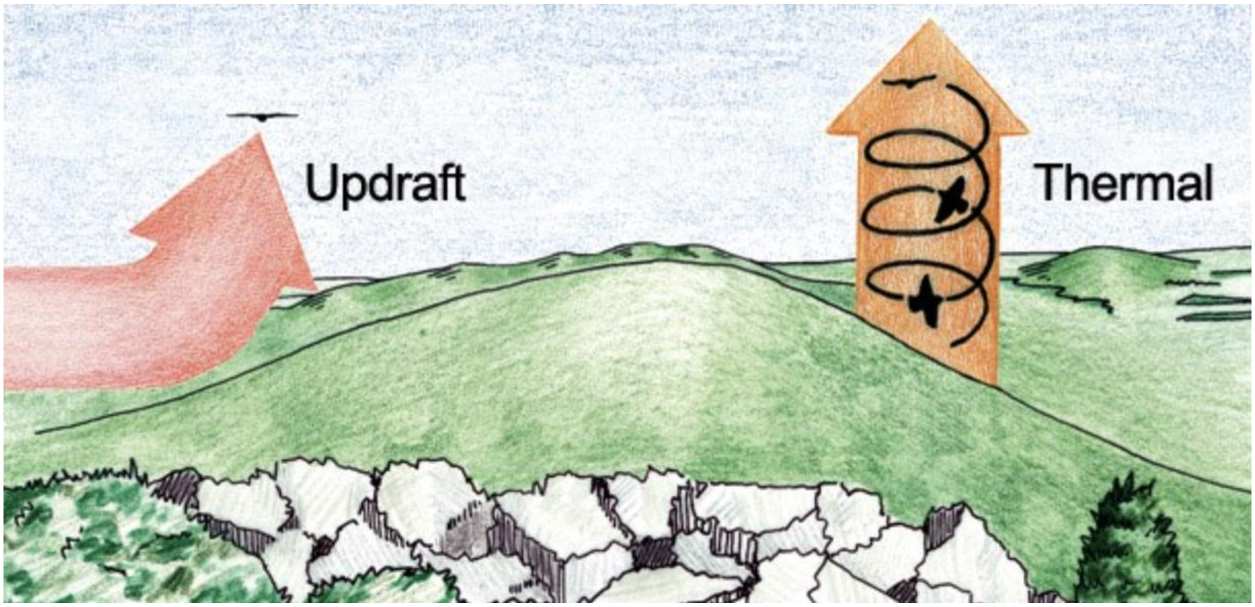

Figure 2 – Thermal Updraft and Wind-Created Updraft

2.3 What Is Time Series Based Weather and Raptor Count Data?

Hawk Mountain stores its raptor count and weather data in a time-series manner. Time series data is a sequence of data points indexed in time order. This data consists of successive measurements made from the same source over a time interval and is used to track change over time [17]. See Figure 3 for a sample of the total raptor count summed up on a yearly basis for all raptor species.

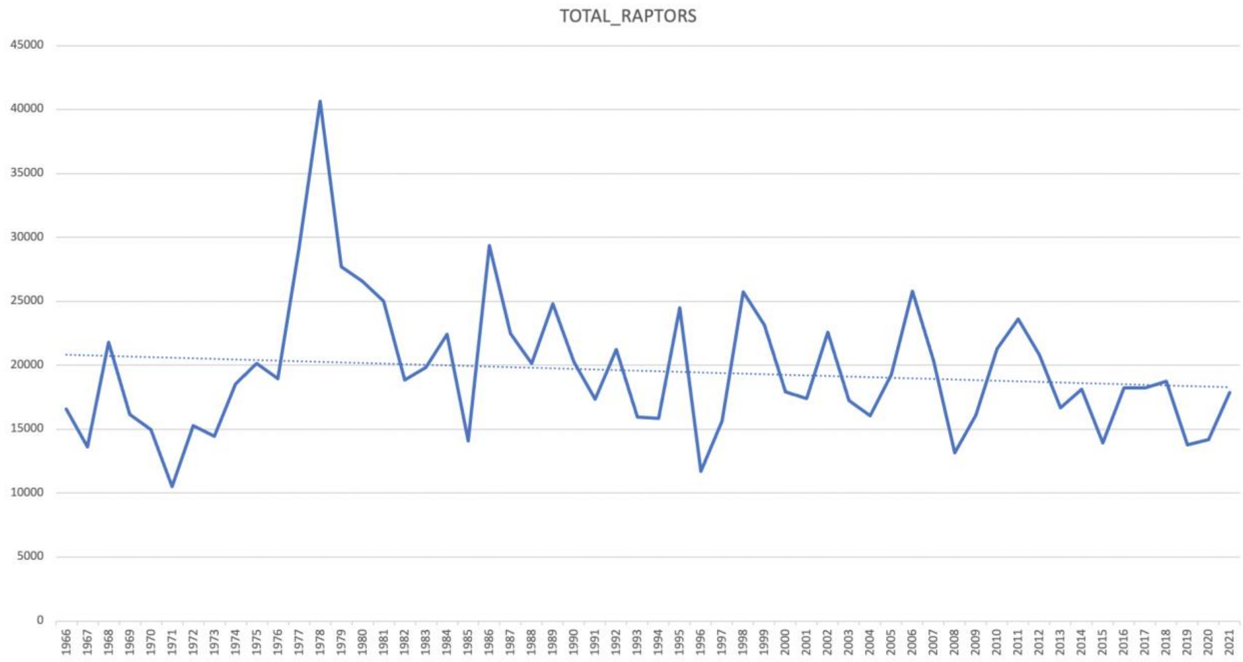

Figure 3 – Hawk Mountain Time-Series Data 1966-2021 Total Raptors Counted per Year

2.4 Python Libraries Used?

In order to clean, parse, process, analyze and run our regression tests on the data we used a multitude of Python open-source libraries available. The following is not a list of all the available data science and analytics Python libraries; we will list only the ones that were sourced for the purposes of this research.

2.4.1 Pandas

The Pandas library for Python offers a fast, flexible and practical package to deal with our data. This library is great for working with relational and labeled data. The data that was provided by Hawk Mountain was in comma-separated value (CSV) format and Pandas is able to import this data very efficiently with just one line of code. Pandas has two primary data structures: Series and DataFrames. A Series is a container for one-dimensional scalar data whereas a DataFrame is a container for multiple Series data.

Pandas comes with many functions that are specifically catered to data wrangling. With Pandas, we can easily handle missing data, data grouping, object type conversion, reshaping data sets, joining data sets, and much more. One of the most important functionalities that this research uses Pandas for is its time series-specific functionality [12].

2.4.2 NumPy (Numerical Python)

The NumPy library was used in tangent with Pandas and other scientific Python packages in order to leverage the fast mathematical functions on arrays and matrices. What it means in terms of this research is that we can take our data, split it into DataFrames via Pandas and then run calculations and mathematical functions extremely efficiently using NumPy [10].

2.4.3 Seaborn and Matplotlib

Seaborn and Matplotlib are both data visualization Python libraries that are used to create informative statistical graphs. The Seaborn library is based on matplotlib and is built on top of it in order to create even more attractive and informative graphics. Seaborn is a bit more comfortable when working with Pandas DataFrames but during this research we are using both libraries almost interchangeably. Most of the graphs in this research are generated using Seaborn and Matplotlib [15] [13].

2.4.4 Scikit-learn

Scikit-learn is an open-source machine learning library for Python that has many uses. This library is built upon other python libraries such as NumPy, Pandas and Matplotlib. Some of the functionality of scikit-learn includes [7]:

a. **Regression:** Predicting a continuous-valued attribute associated with an object.

b. **Classification:** Identifying which category an object belongs to.

c. **Clustering:** Automatic grouping of similar objects into sets.

d. **Preprocessing:** Feature extraction and data normalization.

e. **Model Evaluation:** Comparing, validating, and choosing parameters and models

2.5 What Is Regression Analysis?

Regression analysis is widely seen as the 'go-to" form of analysis for trying to make sense of real-world data. Regression analysis is a way of mathematically sorting out which variables have an impact on the variable in question aka the dependent variable. This type of analysis can quantify and give us an understanding of which factors matter most, which factors we can ignore and how all of these factors interact with each other. For this study, we are laying down the skeletal building blocks for future analysis that will be done on this data. So, it is very crucial that we understand our data and the underlying relationships before we can continue to find correlations and provide accurate and intelligible analysis about the effects of the changing climate as they relate to Hawk Mountain's on-going research.

Regression is represented mathematically in the following formula;

$$Y = a + bX + e$$

Where:

Y: is the dependent variable that we are trying to predict

a: is the Y-intercept aka "bias"

b: the slope coefficient

X: is the independent variable

e: is the error

2.6 What Is Pearson Correlation Coefficient

To find the initial correlation of the data, a scikit-learn library function [7] was used to normalize attribute data and a pandas library function was used to find the Pearson Correlation Coefficient [11]. A Pearson Correlation Coefficient is just a measure of linear correlation between two sets of data. The ratio of this covariance is expressed as a value between 0.0 and 1.0. This is a great way to get the simplest correlation between attributes that appear to be linear in nature and is generally not a strong correlator with nonlinear data. Formula for Pearson Correlation Coefficient can be seen below:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Figure 4 – Pearson Correlation Coefficient Formula

Where:

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

2.7 Why Pearson Correlation Coefficient

The reason that using the Pearson correlation coefficient analysis is appropriate here, is because there is a need to find these positive and negative linear correlations as this will outline the most important climate attributes as it relates to the TOTAL_mean raptor count. TOTAL_mean is a derived attribute which represents the average of the TOTAL_RAPTORS attribute. It's important to note that at this stage in the analysis, we are not looking to extract predictor coefficients and that these are just correlations. Another important note is that when we have a correlation coefficient of, let's say .95, this does not mean that the two attributes' records are tightly, numerically close in value. The high correlation means that the trend of the two attributes, as explained by the Pearson correlation coefficient formula, is highly correlated.

The last important note here is to explain that if we were to calculate the error coefficient for comparison that the Pearson correlation coefficient is making; it would be 0 although the actual values are not the same. Although, if we calculate the mean absolute error and the mean squared error, they will show many discrepancies and high error values. This just means that our analysis here is not predictive and it's only correlative. Again, the analysis in this paper is for finding the highest correlating attributes and the matrix can be seen in sections 4.4 and 4.5.

Chapter 3

3. Scope of Data

3.1 Hawk Mountain Sanctuary Data

The main dataset that we are working with here is coming from Hawk Mountain Sanctuary's database. It is given to us in CSV format which will be very easy to pull and manipulate using the mentioned Python tools and libraries. The dataset starts from 1966 and is captured on an hourly basis. For the sake of this project, we will be aggregating the data on a year-over-year autumnal basis. The Hawk Mountain Autumn counting season is from August 15 to December 15. The data provided has many columns but for this research we will be only taking a few columns into account [See Appendix A].

We will be taking the following non-target attributes into account from the original Hawk Mountain dataset: 'YEAR', 'MONTH', 'DAY', 'MAX_VISIBILITY', 'FLIGHT_ALT', 'CLOUD_COVER', 'TEMP', 'WIND_SPEED', 'WIND_DIR', 'SKY_CODE', 'NUMBER_OF_OBSERVERS' [See Appendix A]. The target attribute from this dataset is the TOTAL_RAPTORS; this is the attribute that we believe to be dependent on the changing weather attributes previously mentioned. TOTAL_RAPTORS is an aggregate count of all raptor species that Hawk Mountain keeps counts of. For this portion of research, we are only looking at the total raptor count as opposed to looking at each individual species.

3.2 National Oceanic and Atmospheric Administration Data

NOAA [8] data was retrieved from the NOAA website from the Allentown Lehigh Valley International Airport weather station, which as previously mentioned, is the closest weather station to Hawk Mountain [9]. The data was packaged up in CSV format and holds data weather starting from 1948 but we will be only using data starting from 1966 - 2021. This data is needed to account for missing data in the Hawk Mountain's old data that has many missing records. The missing records were mostly weather attributes such as cloud cover and temperature for early years. We will use other attributes from this dataset such as dry bulb temperature, wind speed, daily precipitation [See Appendix A for all attributes pulled]. We will join this data to the original Hawk Mountain data. This will allow us to have more attributes to run analysis on; with hopes of finding correlations and better predictors. The NOAA data came in monthly, daily, and hourly averages. We will be using the hourly data and aggregating it on a daily basis and then on a yearly basis [See Appendix A].

3.3 Data Aggregation

Hawk Mountain data needs to be aggregated so we can run our analysis. For the scope of this analysis, the data will be aggregated on a year-to-year autumnal basis. This means that we will be filtering out months that are not included in Hawk Mountain Autumn season counts. The months that will be kept are months 08 – 12 and we will filter out months 01 – 07. As previously mentioned, August through December is the time when Hawk Mountain staff conducts their annual autumn raptor counts so this is the timeframe that we will be focused on, at least for this portion of the research [See Appendix A].

The NOAA data will also be aggregated on a year-to-year autumnal basis. The same filtering methods that were used to aggregate the Hawk Mountain data will be used for this dataset. This dataset contains many columns, so we will be filtering and aggregating hourly and daily attributes separately [See Appendix A]. Once this data is aggregated, we will calculate some scalar statistical data and merge the data with Hawk Mountains data. This data aggregation and merging, in its final form, will be used not just for this portion of research but also will be used for continued future research.

3.4 Data Merge

The NOAA and Hawk Mountain data were merged and aggregated on a daily and year-to-year autumnal basis. The hourly weather records from the NOAA data were used and averaged on a daily and yearly basis [See Appendix A]. This is the data that is used for much of the analysis here. We are using 43 years' worth of data 1979-2021 and after the data merge, there are 799 attributes that we will filter through and use.

3.5 Data Cleaning and Organization

Data cleaning is arguably one of the most time-consuming parts of data analysis. To be able to get the most accurate results when wrangling and analyzing the data; we need to be able to get rid of records that are missing, do not make sense, are clear outliers, etc. This is where Dr. Parson's Python magic came in very handy. We will not get into the technical details about the data cleaning process but there is a need to include and outline what data was cleaned and why. The final output data from the cleaning is used for the analysis here.

1. We cleaned data that was not within Hawk Mountain's regular observation period of August – December on both the Hawk Mountain data and the NOAA data.
2. Data entry errors were cleaned aka numeric attributes with trailing or leading non-numeric data.

3. Setting certain data such as number of observers and duration that should not be 0 were set to unknown.
4. Clear out blank rows.

Aside from the domain-specific cleaning; the cleaning of the data was standard and was done so that the machine learning tools do not choke up [1]. A “standard” cleaning entails renaming columns to a more recognizable set of labels, skipping unnecessary rows or columns, modifying the DataFrames index, cleaning leading and trailing blank spaces from attribute elements.

Chapter 4

4. Data Analysis

4.1 Hawk Mountain Data Scalar Statistics

The records count trend for Hawk Mountain's numeric attributes of interest can be seen below, aggregated on a year-over-year autumnal basis. The reason for showing this trend is to clearly show that there are many numeric attributes that are missing or that did not start until later years. In a perfect data set, we would not see the decoupling of trends for all these attributes. There are also some discrepancies in the trend that can only be explained by the Hawk Mountain team. For example, the reason why we see increasing counts starting at 1980 is because, as Dr. Laurie Goodrich explained, the counting season was extended by a month. The counting season before 1980 was from September 1st to December 1st and now it's from August 15th to December 15th [3].

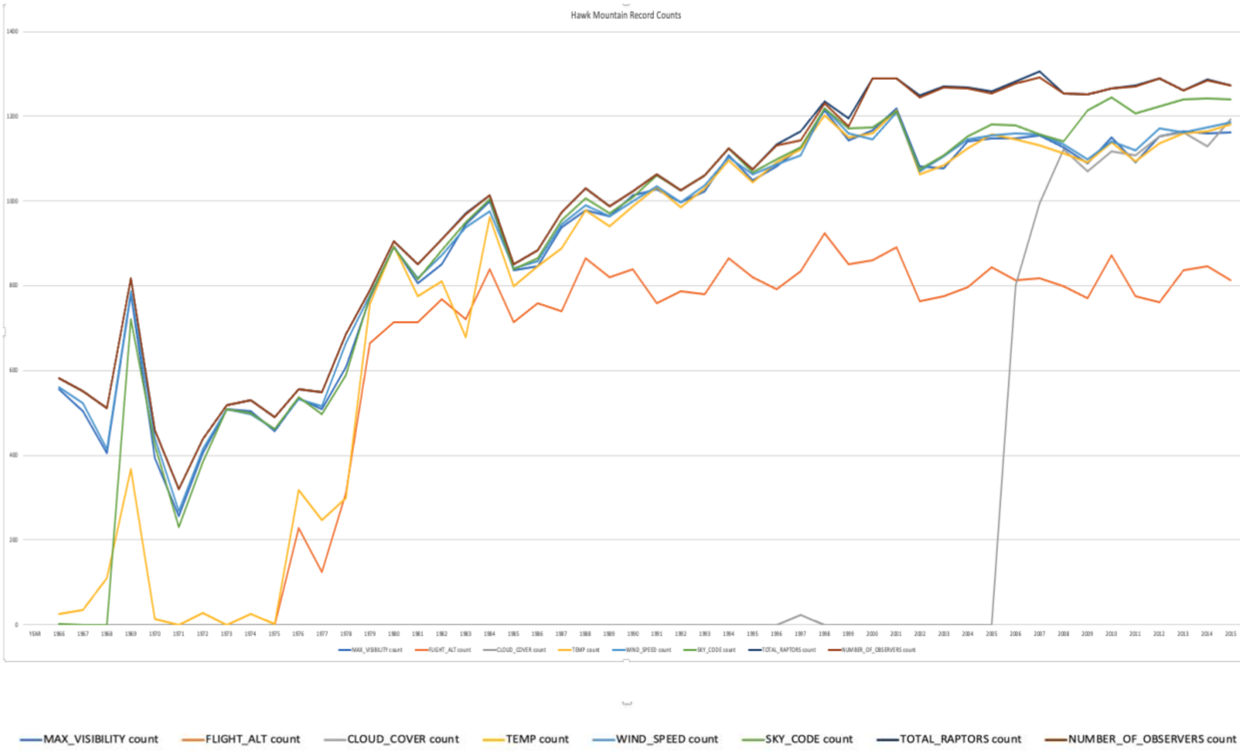

Figure 5 – Hawk Mountain Total Attribute Count Trend per Year 1966-2015

Inconsistent and missing weather data is what led us to pull and merge the NOAA data with the Hawk Mountain data. We filtered out some data from the final merged data as well in order to zoom in a bit and stay within the scope of this research. It should be noted that the final data will be used for future analysis efforts by Kutztown University faculty, students and researchers. The scalar statistics were derived from the original data; they include min, max, mean, population standard deviation, median. We derived this using Python and its data analysis libraries and first derived the attributes for the yearly, autumnal based aggregated Hawk Mountain data [Table 1-1.2]. These are the non-target numeric attributes included in the below tables: temperature (Celsius), wind speed (km/hr), cloud cover, sky code, flight height/altitude, and visibility (See Appendix B for more units of measurement). The attributes such as weather and temperature are self-explanatory but there are a few that require some explanation such as wind speed, flight height/altitude and sky code [Appendix B].

year	WindSpd_mean	HMtempC_mean	CloudCover_mean	SkyCode_mean	FlightHT_mean	Visibility_mean
1979	9.48	12.43		2.31		34.63
1980	9.72	12.96		2.13		33.54
1981	11.06	11.61		2.37		37.65
1982	7.78	14.02		2.02		27.80
1983	8.71	15.85		2.04		15.78
1984	9.76	14.73		2.00		25.48
1985	11.07	14.77		2.19		37.15
1986	9.16	12.06		2.10		39.83
1987	10.06	12.70		2.02		40.38
1988	10.02	11.73		2.04		39.81
1989	10.53	13.22		2.10		34.75
1990	11.98	15.19		2.01		41.11
1991	9.48	13.61		2.18		33.49
1992	12.20	11.74		2.37		40.46
1993	10.02	12.88		2.29		43.52
1994	9.34	13.58		1.92		45.33
1995	11.59	12.26		1.91	1.68	45.76
1996	10.47	14.12		2.40	1.87	36.39
1997	11.56	12.58		2.19	1.52	40.60
1998	10.03	13.83		2.18	1.51	40.89
1999	9.95	14.61		2.23	2.10	37.87
2000	11.15	13.47		2.25	1.42	40.95
2001	10.51	13.12		2.26	1.50	35.43
2002	10.10	13.79		2.54	1.40	34.18
2003	11.27	13.06		2.64	1.43	32.93
2004	9.35	13.05		2.58	1.62	32.07
2005	10.08	12.51		2.31		35.67
2006	10.34	12.23	39.30	2.43		38.03
2007	9.38	12.87	40.49	2.50		37.26
2008	8.14	11.89	47.23	2.20	1.64	40.92
2009	9.23	13.03	55.26	2.93	1.46	40.97
2010	10.65	12.37	41.87	2.45	1.43	41.37
2011	8.05	12.82	45.31	2.61	1.72	32.90
2012	8.09	12.12	47.57	2.40	1.83	36.13
2013	7.83	12.11	47.51	2.36	1.87	35.77
2014	9.07	11.87	54.75	2.45	1.40	32.73
2015	8.61	14.40	44.54	2.13	1.86	36.87
2016	8.55	13.47	50.30	2.03	1.71	35.53
2017	6.51	12.94	48.49	2.02	1.70	31.51
2018	8.46	11.84	65.13	2.69	1.67	27.40
2019	7.71	11.74	55.27	2.13	1.73	20.05
2020	6.37	13.33	56.18	2.40	1.86	21.14
2021	6.34	12.87	57.77	2.36	1.58	26.26

Table 1 – Hawk Mountain Scalar Statistics – Mean per Year 1979-2021

year	WindSpd_median	WindSpd_pstdv	WindSpd_min	WindSpd_max	HMtempC_median	HMtempC_pstdv	HMtempC_min	HMtempC_max	CloudCover_median	CloudCover_pstdv
1979	8.50	7.75	0.00	55.50	12.11	6.99	-16.00	26.00		
1980	8.44	7.91	0.00	44.00	12.69	8.60	-2.00	32.60		
1981	8.17	9.54	0.00	44.00	10.73	7.10	-2.00	26.00		
1982	5.96	6.94	0.00	33.50	15.82	7.23	-7.00	26.67		
1983	7.89	6.58	0.00	33.50	16.56	9.04	-1.00	31.09		
1984	7.44	9.05	0.00	68.00	16.78	7.60	-4.00	29.00		
1985	8.50	9.65	0.00	55.50	13.88	7.84	-1.88	28.33		
1986	7.44	7.60	0.00	34.45	14.25	8.75	-9.00	27.67		
1987	8.50	7.87	0.00	41.38	11.61	7.91	-6.30	32.00		
1988	7.28	8.68	0.00	44.29	11.00	8.89	-13.00	36.44		
1989	7.78	10.09	0.00	68.00	14.00	9.48	-9.50	28.89		
1990	7.85	11.58	0.00	45.79	16.23	8.18	-2.33	30.33		
1991	7.39	7.35	0.00	33.50	13.00	8.81	-5.33	31.38		
1992	8.80	10.03	0.00	51.67	10.88	8.79	-4.67	29.88		
1993	8.25	7.85	0.00	42.83	12.20	8.34	-6.33	32.78		
1994	6.95	7.67	0.30	35.83	14.05	8.47	-6.00	31.50		
1995	8.50	9.73	0.00	45.94	13.00	11.38	-8.75	34.13		
1996	8.50	7.83	0.00	46.10	14.18	8.55	1.00	29.36		
1997	8.52	8.98	0.43	44.00	11.91	7.85	1.00	30.00		
1998	7.83	8.41	0.00	55.50	13.21	7.15	1.71	28.25		
1999	8.00	6.94	0.00	32.65	14.67	7.59	1.60	36.53		
2000	8.89	7.88	0.33	50.07	13.66	7.74	1.00	31.00		
2001	8.42	7.19	0.00	38.75	12.40	6.98	1.00	27.67		
2002	7.74	7.11	0.00	33.50	13.78	8.56	1.00	32.13		
2003	8.50	9.91	0.00	51.19	14.00	7.14	1.00	29.00		
2004	7.41	6.99	0.00	33.70	11.70	6.98	1.00	27.00		
2005	8.25	6.88	0.82	31.13	13.00	9.66	-11.57	27.27		
2006	8.50	6.61	1.36	36.50	12.11	7.34	-4.11	31.56	32.25	35.12
2007	7.90	7.05	1.20	33.88	14.34	9.28	-7.14	27.91	30.00	35.65
2008	7.15	5.70	0.75	34.89	13.50	8.99	-6.75	26.73	47.48	34.00
2009	7.20	6.73	0.00	29.72	13.06	7.37	-6.25	30.56	57.42	33.73
2010	8.41	7.71	0.00	34.20	12.64	8.70	-8.25	29.45	43.64	32.01
2011	6.92	6.41	0.00	47.83	12.33	7.47	-2.29	27.00	44.08	33.46
2012	5.75	6.64	0.00	33.50	11.99	8.00	-3.29	27.70	48.89	32.29
2013	5.82	6.11	0.90	35.80	13.19	9.04	-6.63	30.30	50.00	30.58
2014	7.27	6.90	0.00	44.13	13.00	8.32	-6.56	28.10	58.50	32.08
2015	5.75	7.32	0.25	55.50	13.50	7.72	0.25	30.33	38.75	32.77
2016	6.75	6.30	0.00	32.08	13.30	8.98	-10.00	29.78	52.05	33.13
2017	5.44	4.88	0.38	31.13	14.50	8.62	-7.25	29.89	47.00	33.95
2018	7.36	5.57	0.38	26.11	9.60	9.81	-6.25	30.80	72.55	32.15
2019	5.86	6.89	0.00	49.67	12.35	8.84	-4.44	31.43	62.97	33.07
2020	4.88	4.82	0.00	28.00	13.82	7.63	-2.00	27.00	61.39	32.05
2021	5.19	4.66	0.00	26.22	15.00	8.43	-3.67	29.10	62.31	33.18

Table 1.1 – Hawk Mountain Scalar Statistics Min, Median, Max, Std Deviation per Year 1979-2021

year	CloudCover_min	CloudCover_max	SkyCode_median	SkyCode_pstdv	SkyCode_min	SkyCode_max	FlightHT_median	FlightHT_pstdv	FlightHT_min	FlightHT_max	Visibility_median	Visibility_pstdv	Visibility_min	Visibility_max
1979			2.38	1.57	0.00	5.50					33.14	21.97	0.00	81.00
1980			1.90	1.52	0.00	6.00					33.20	16.98	4.00	81.00
1981			2.50	1.50	0.00	5.50					33.57	23.77	5.00	81.00
1982			2.13	1.40	0.00	5.00					21.67	20.81	1.00	81.00
1983			2.11	1.54	0.00	5.67					12.58	11.41	0.00	50.00
1984			1.95	1.57	0.00	7.00					15.00	24.85	1.00	81.00
1985			2.13	1.48	0.00	5.00					32.89	24.56	0.00	77.00
1986			2.14	1.54	0.00	6.20					40.00	21.77	1.40	80.00
1987			2.11	1.43	0.00	8.00					40.00	22.52	0.00	79.67
1988			2.00	1.49	0.00	5.00					40.00	21.84	5.10	77.00
1989			2.13	1.40	0.00	7.00					32.77	23.07	0.00	77.00
1990			2.00	1.56	0.00	6.25					37.31	24.66	4.00	78.50
1991			2.17	1.44	0.00	5.25					27.00	22.75	2.63	77.00
1992			2.22	1.71	0.00	8.00					35.59	23.94	3.60	77.00
1993			2.38	1.58	0.00	6.00					42.42	24.99	0.00	91.00
1994			1.70	1.56	0.00	7.00					45.00	24.83	2.00	85.00
1995			1.75	1.43	0.00	5.78	2.00	1.06	0.00	4.00	48.75	22.73	0.00	80.00
1996			2.50	1.67	0.00	8.00	2.00	0.89	0.00	4.00	36.05	21.94	1.00	76.22
1997			1.89	1.51	0.00	6.33	1.00	0.91	0.00	4.00	39.04	24.35	0.01	80.00
1998			1.90	1.75	0.00	7.00	1.50	0.99	0.00	5.00	38.60	25.16	2.09	82.50
1999			2.10	1.67	0.00	6.25	2.00	0.96	0.00	4.00	33.61	25.53	0.00	80.00
2000			2.25	1.73	0.00	6.29	1.00	0.86	0.00	4.00	40.00	25.59	1.44	80.00
2001			2.17	1.61	0.00	5.29	1.50	1.02	0.00	5.00	29.75	23.03	1.13	77.00
2002			2.50	1.62	0.00	6.00	1.00	0.87	0.00	4.00	30.22	22.04	0.00	77.00
2003			2.44	1.70	0.00	5.63	1.00	1.01	0.00	5.00	29.08	22.49	0.00	77.00
2004			2.55	1.70	0.00	7.00	1.50	0.91	0.00	5.00	30.00	23.43	0.10	77.00
2005			2.20	1.83	0.00	7.00					31.31	23.94	0.00	77.00
2006	0.00	100.00	2.10	1.67	0.00	7.00					34.67	24.87	0.00	77.00
2007	0.00	100.00	2.55	1.72	0.00	7.00					32.80	25.52	0.07	77.00
2008	0.00	100.00	2.19	1.58	0.00	6.78	1.50	0.95	0.00	6.00	39.05	24.65	0.00	77.00
2009	0.00	100.00	2.83	1.81	0.00	8.00	1.00	0.79	0.00	4.00	37.00	25.71	0.00	77.00
2010	0.00	98.00	2.29	1.94	0.00	7.00	1.00	0.87	0.00	5.00	36.45	24.21	2.50	77.00
2011	0.00	100.00	2.33	1.91	0.00	7.00	1.50	1.17	0.00	5.00	24.65	25.18	0.00	77.00
2012	0.00	100.00	2.18	1.70	0.00	7.00	2.00	1.02	0.00	5.00	32.67	24.74	0.00	77.00
2013	0.00	100.00	2.00	1.83	0.00	8.00	2.00	1.11	0.00	4.00	31.95	24.63	0.00	77.00
2014	0.00	100.00	2.38	1.62	0.00	8.00	1.00	0.49	1.00	2.00	30.67	20.76	0.33	77.00
2015	0.00	100.00	1.89	1.77	0.00	7.00	2.00	0.98	0.00	5.00	33.00	23.01	3.56	77.00
2016	0.00	100.00	2.00	1.63	0.00	8.00	1.50	1.01	0.00	4.00	32.43	22.70	0.00	77.00
2017	0.00	100.00	2.00	1.48	0.00	6.33	2.00	0.94	0.00	4.00	28.72	20.68	1.14	77.00
2018	0.00	100.00	2.75	1.48	0.00	6.56	2.00	0.87	0.00	4.00	23.73	20.15	0.00	77.00
2019	0.00	100.00	2.33	1.53	0.00	6.00	2.00	0.95	0.00	4.00	14.63	14.24	3.00	77.00
2020	0.00	100.00	2.35	1.54	0.00	6.00	2.00	0.88	0.00	4.00	15.70	15.99	1.13	77.00
2021	0.00	100.00	2.36	1.57	0.00	5.44	1.50	1.03	0.00	5.00	20.25	18.27	0.22	77.00

Table 1.2 – Hawk Mountain Scalar Statistics Mean, Median, Max, Std Deviation per Year 1979-2021

4.2 NOAA Data Scalar Statistics

Table 2-2.3 shows the NOAA attributes that are of interest to us. I redacted the actual attribute names here in order to be able to fit the table on this document. The original NOAA data attributes are prefixed by the word "Hourly" (See Appendix A). To recap, the hourly data is taken from the NOAA database and the data was aggregated on a daily and then yearly basis. The data was merged with the Hawk Mountain data, but we did not completely replace the Hawk Mountain weather attributes. This is just a way to use weather data that is arguably more accurately recorded by an actual weather station. We will run some correlation analysis using both sets of data. Figure 6 shows a comparison of the temperature attributes between Hawk Mountain and NOAA's data, showing a linear trendline fit and a simple moving average.

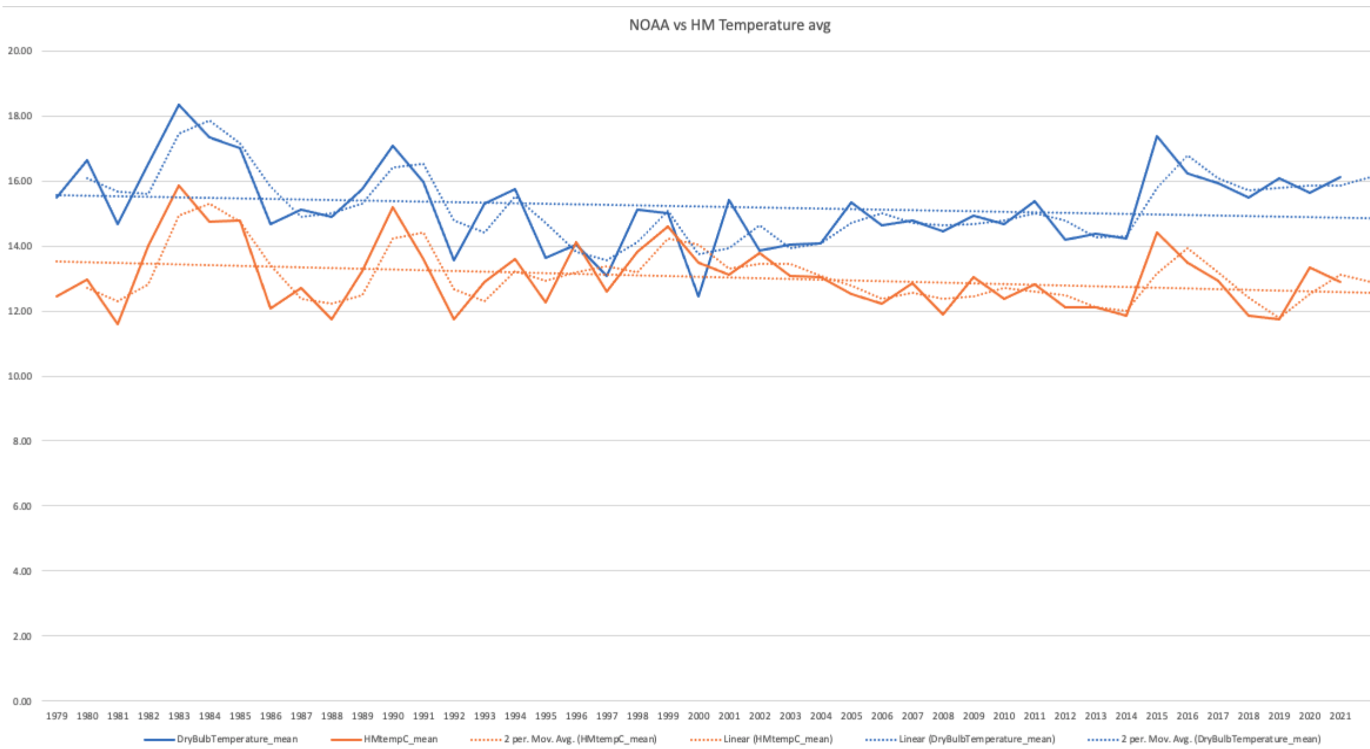

Figure 6 – Hawk Mountain vs NOAA Average Temperature Data in Celsius per Year. Also includes Slope and Simple Moving Average 1979-2021

year	DryBulbTemperature_mean	WetBulbTemperature_mean	DewPointTemperature_mean	WindSpeed_mean	Precipitation_mean	StationPressure_mean	RelativeHumidity_mean	Visibility_mean
1979	15.49	11.45	7.55	10.00	0.01	29.65	61.06	14.77
1980	16.65	11.03	5.60	11.01	0.01	29.64	50.57	17.03
1981	14.68	10.37	5.92	10.22	0.01	29.62	58.09	16.73
1982	16.53	11.71	7.16	9.19	0.01	29.71	56.11	15.98
1983	18.34	12.73	8.05	10.20	0.01	29.64	53.58	16.31
1984	17.33	12.58	8.07	9.30	0.00	29.69	56.14	16.02
1985	17.00	12.25	7.89	9.91	0.01	29.73	57.39	15.79
1986	14.67	10.28	5.80	10.09	0.01	29.73	57.31	16.88
1987	15.10	10.71	6.32	9.76	0.01	29.63	57.99	16.18
1988	14.88	10.48	5.84	10.37	0.01	29.60	56.85	18.09
1989	15.74	11.52	7.41	10.45	0.01	29.61	59.78	15.80
1990	17.09	12.54	8.26	11.06	0.00	29.62	57.88	17.34
1991	15.97	11.16	6.41	9.68	0.00	29.69	55.65	15.66
1992	13.56	9.59	5.36	9.48	0.01	29.70	59.82	16.40
1993	15.28	11.39	7.66	9.01	0.01	29.66	62.54	15.95
1994	15.74	11.36	6.96	9.99	0.01	29.69	58.28	17.48
1995	13.64	9.22	4.53	9.33	0.00	29.63	57.28	13.65
1996	14.03	10.16	6.30	7.44	0.02	29.65	62.20	9.39
1997	13.09	9.35	5.42	7.78	0.01	29.57	62.22	9.49
1998	15.11	11.16	7.28	7.45	0.02	29.64	61.98	9.09
1999	15.01	1.90	6.51	7.98	0.03	29.64	59.66	9.20
2000	12.46	1.47	4.70	7.72	0.01	29.61	61.38	9.53
2001	15.42	6.28	6.98	7.82	0.01	29.76	59.40	9.59
2002	13.84	5.07	5.92	7.82	0.01	29.56	61.65	9.60
2003	14.04	10.85	7.53	8.12	0.02	29.64	67.09	9.13
2004	14.09	0.19	6.47	7.57	0.02	29.74	62.22	9.34
2005	15.35	10.59	5.98	7.16	0.02	29.62	56.38	9.21
2006	14.63	10.49	6.21	7.39	0.02	29.63	59.27	9.22
2007	14.78	10.71	6.91	7.00	0.02	29.66	61.89	8.91
2008	14.45	10.21	6.08	6.98	0.05	29.66	59.88	9.34
2009	14.91	11.45	8.26	7.03	0.02	29.64	66.51	9.33
2010	14.65	10.16	5.76	8.02	0.03	29.55	58.23	9.55
2011	15.37	11.69	8.16	6.47	0.08	29.62	64.45	9.46
2012	14.18	10.42	6.85	6.54	0.04	29.61	63.75	9.01
2013	14.37	10.19	5.87	6.85	0.03	29.67	58.88	9.21
2014	14.22	9.80	5.32	7.38	0.02	29.62	57.58	9.37
2015	17.36	12.55	8.32	6.58	0.04	29.68	57.65	9.45
2016	16.23	11.23	6.48	7.51	0.01	29.67	54.74	9.60
2017	15.94	11.43	7.08	7.05	0.03	29.65	57.83	9.53
2018	15.50	11.74	8.11	7.60	0.04	29.66	63.00	9.35
2019	16.07	11.25	6.56	7.54	0.03	29.65	55.46	9.50
2020	15.61	11.48	7.60	7.39	0.02	29.65	61.04	9.55
2021	16.12	12.12	8.43	7.39	0.02	29.61	61.97	9.84

Table 2 – NOAA Scalar Statistics – Mean per Year 1979-2021

year	DryBulbTemperature_pstdv	WetBulbTemperature_pstdv	DewPointTemperature_pstdv	WindSpeed_pstdv	Precipitation_pstdv	StationPressure_pstdv	RelativeHumidity_pstdv	Visibility_pstdv
1979	7.10	6.49	7.72	4.90	0.06	0.19	14.06	8.11
1980	8.72	7.10	8.27	5.22	0.05	0.19	14.39	8.47
1981	7.07	6.46	7.90	4.99	0.07	0.23	15.22	8.65
1982	7.41	6.49	7.87	4.20	0.06	0.19	14.67	9.02
1983	9.20	7.22	7.78	4.16	0.04	0.22	14.41	7.95
1984	7.74	7.41	9.24	3.65	0.01	0.19	12.69	8.20
1985	7.86	7.05	8.41	4.18	0.08	0.18	15.24	8.89
1986	8.47	7.55	8.76	3.93	0.07	0.19	13.72	8.39
1987	7.83	7.07	8.43	4.34	0.04	0.21	14.84	7.52
1988	8.59	7.92	9.74	4.38	0.03	0.21	15.16	8.33
1989	9.49	8.73	10.29	4.79	0.04	0.20	15.61	8.34
1990	7.80	7.33	9.00	4.89	0.01	0.17	13.94	7.87
1991	9.12	7.96	9.52	3.93	0.01	0.19	16.03	7.51
1992	8.74	7.85	9.24	4.16	0.03	0.20	15.06	8.20
1993	8.84	7.99	9.27	3.94	0.05	0.22	15.22	8.12
1994	7.31	6.93	8.60	3.98	0.03	0.19	15.06	8.38
1995	10.93	8.94	9.58	3.78	0.02	0.19	16.67	6.89
1996	9.00	7.98	8.80	3.60	0.10	0.23	13.21	1.22
1997	8.31	7.48	8.59	3.54	0.04	0.19	13.94	1.36
1998	8.08	7.55	9.06	3.13	0.08	0.17	15.17	1.66
1999	8.20	9.02	8.61	3.31	0.14	0.30	15.47	1.57
2000	9.94	9.96	10.76	3.41	0.05	0.18	14.01	1.24
2001	7.66	6.03	8.13	3.41	0.05	0.16	12.97	1.35
2002	10.07	7.41	9.32	3.50	0.04	0.23	14.74	1.22
2003	8.62	8.36	9.79	4.45	0.09	0.20	15.20	1.68
2004	8.51	8.51	9.19	3.54	0.06	0.18	13.90	1.41
2005	9.87	8.48	9.45	3.83	0.08	0.22	14.92	1.64
2006	7.25	6.69	8.22	3.59	0.09	0.22	14.25	1.68
2007	9.78	8.41	9.11	3.11	0.07	0.21	14.25	2.16
2008	9.01	7.70	8.55	3.57	0.19	0.23	14.73	1.51
2009	7.45	6.68	7.45	3.89	0.08	0.20	13.91	1.73
2010	9.29	7.90	8.84	3.80	0.10	0.21	15.86	1.13
2011	7.62	7.19	8.50	3.63	0.30	0.24	14.94	1.36
2012	8.67	7.56	8.31	4.18	0.17	0.24	14.52	2.07
2013	9.10	8.30	9.66	3.34	0.10	0.20	13.74	1.74
2014	8.98	7.80	8.90	3.61	0.07	0.19	14.07	1.65
2015	7.67	6.51	7.55	3.24	0.15	0.21	13.47	1.39
2016	9.26	7.97	9.24	3.87	0.07	0.18	13.12	1.33
2017	8.99	8.15	9.60	3.71	0.11	0.20	13.60	1.35
2018	9.72	9.02	10.53	3.40	0.14	0.20	13.01	1.46
2019	8.82	7.78	9.19	3.47	0.11	0.20	13.48	1.46
2020	8.03	7.25	8.26	3.52	0.08	0.20	13.11	1.37
2021	8.08	7.55	8.81	3.24	0.07	0.19	11.87	0.56

Table 2.1 – NOAA Scalar Statistics – Std Deviation per Year 1979-2021

year	DryBulbTemperature_median	WetBulbTemperature_median	DewPointTemperature_median	WindSpeed_median	Precipitation_median	StationPressure_median	RelativeHumidity_median	Visibility_median	DryBulbTemperature_min	WetBulbTemperature_min	DewPointTemperature_min	WindSpeed_min
1979	15.38	10.71	7.93	9.65	0.00	29.67	60.29	13.39	1.39	-1.91	-8.40	0.00
1980	15.75	10.73	6.02	9.91	0.00	29.65	47.41	17.67	1.56	-2.89	-15.39	2.50
1981	13.66	9.06	4.91	9.21	0.00	29.63	55.35	16.45	1.39	-2.22	-11.78	0.63
1982	18.27	13.48	9.94	8.56	0.00	29.71	53.75	13.83	-2.78	-6.11	-15.56	1.50
1983	17.89	12.59	9.06	9.33	0.00	29.62	48.50	17.34	0.83	-2.59	-9.91	1.56
1984	18.33	13.19	9.20	8.82	0.00	29.69	55.75	14.91	-3.06	-5.56	-11.39	3.00
1985	17.59	12.18	8.40	9.17	0.00	29.73	56.27	14.91	-2.78	-5.00	-12.99	2.56
1986	16.09	10.94	5.52	9.34	0.00	29.74	54.30	16.29	-4.29	-7.38	-16.35	4.20
1987	14.69	10.50	6.31	8.89	0.00	29.63	55.55	16.33	-7.02	-9.65	-19.60	2.20
1988	15.06	10.61	5.78	10.22	0.00	29.60	55.30	18.74	-7.96	-10.74	-23.70	2.22
1989	18.69	13.74	8.97	9.66	0.00	29.63	58.00	15.70	-6.11	-7.64	-16.25	1.89
1990	17.74	13.22	9.32	10.15	0.00	29.65	54.76	18.19	0.83	-2.50	-10.00	2.71
1991	16.08	11.20	7.53	9.20	0.00	29.66	53.22	14.82	-3.75	-6.81	-16.25	1.56
1992	12.78	8.70	5.95	8.57	0.00	29.71	56.00	17.12	-5.56	-8.70	-20.00	0.00
1993	14.67	10.28	6.78	8.28	0.00	29.67	59.57	14.94	-8.28	-10.17	-17.78	2.80
1994	16.80	12.06	8.12	9.38	0.00	29.68	53.42	17.74	-1.35	-4.84	-17.31	2.00
1995	15.44	11.62	6.88	8.53	0.00	29.65	53.61	9.97	-8.82	-10.76	-18.26	0.00
1996	14.77	11.62	7.57	6.79	0.00	29.63	60.47	9.94	-6.11	-7.50	-14.17	0.00
1997	13.17	8.59	5.21	7.43	0.00	29.59	59.37	9.94	-4.72	-7.29	-14.51	0.00
1998	15.39	11.21	8.19	7.09	0.00	29.63	60.33	9.94	-5.56	-8.67	-19.22	1.27
1999	15.74	-1.67	7.89	7.83	0.00	29.81	57.42	9.94	-4.60	-6.39	-15.32	1.08
2000	13.62	-5.00	5.45	7.37	0.00	29.67	59.56	9.95	-8.22	-6.39	-19.44	0.43
2001	15.66	9.44	7.12	7.53	0.00	29.78	56.68	9.95	-5.93	-3.06	-13.98	2.10
2002	13.15	0.97	6.78	8.00	0.00	29.57	59.17	9.95	-7.30	-4.17	-17.54	0.00
2003	14.44	11.26	8.52	7.14	0.00	29.67	63.73	9.95	-2.87	-5.35	-13.61	0.00
2004	13.84	-3.33	6.81	6.92	0.00	29.75	59.20	9.95	-11.11	-6.67	-22.56	2.27
2005	16.94	12.67	8.50	7.00	0.00	29.64	51.13	10.00	-7.99	-9.86	-16.74	0.00
2006	14.06	11.32	6.94	6.80	0.00	29.63	56.60	10.00	-2.01	-5.21	-13.83	0.00
2007	16.02	11.82	8.13	6.86	0.00	29.67	57.75	10.00	-2.78	-4.14	-13.02	0.00
2008	16.04	12.32	8.36	6.65	0.00	29.65	56.33	10.00	-3.47	-6.25	-14.03	0.00
2009	15.11	11.58	8.06	6.11	0.00	29.66	63.00	10.00	-3.26	-6.04	-14.10	1.40
2010	14.11	10.76	6.57	7.80	0.00	29.56	54.50	10.00	-6.04	-7.92	-13.54	0.00
2011	15.25	11.41	8.19	5.75	0.00	29.63	62.00	10.00	-1.02	-3.61	-9.07	0.60
2012	14.58	11.17	7.35	5.86	0.00	29.64	60.82	10.00	-0.56	-1.98	-8.02	0.00
2013	16.44	11.77	7.17	7.10	0.00	29.66	56.85	10.00	-5.65	-8.10	-17.83	0.33
2014	15.17	10.50	6.16	6.90	0.00	29.61	54.91	10.00	-3.61	-6.67	-16.42	0.00
2015	17.11	12.90	8.89	6.27	0.00	29.68	54.09	10.00	1.39	-2.00	-8.83	0.00
2016	15.90	12.19	7.32	6.96	0.00	29.67	52.35	10.00	-6.73	-9.57	-20.19	0.27
2017	18.06	13.22	8.24	6.70	0.00	29.65	54.33	10.00	-11.11	-13.03	-20.86	0.33
2018	16.37	11.42	8.16	7.39	0.00	29.67	61.15	10.00	-3.83	-7.04	-18.64	0.00
2019	17.08	12.31	7.61	6.80	0.00	29.65	53.17	10.00	-0.80	-4.38	-16.83	1.10
2020	16.92	12.01	7.62	6.96	0.00	29.63	58.61	10.00	-5.40	-6.37	-9.44	0.00
2021	17.41	12.54	9.35	7.22	0.00	29.62	60.44	10.00	1.43	-1.98	-8.58	0.00

Table 2.2 – NOAA Scalar Statistics – Min, Median per Year 1979-2021

Precipitation_min	StationPressure_min	RelativeHumidity_min	Visibility_min	DryBulbTemperature_max	WetBulbTemperature_max	DewPointTemperature_max	WindSpeed_max	Precipitation_max	StationPressure_max	RelativeHumidity_max	Visibility_max	WindDirection
0.00	29.08	35.50	1.82	28.25	23.11	22.33	23.00	0.57	30.04	95.00	29.83	
0.00	28.82	26.30	1.16	34.44	23.98	20.56	30.00	0.34	30.12	90.50	28.41	270.00
0.00	29.08	26.56	1.49	28.40	22.04	20.93	26.20	0.73	30.19	97.00	29.83	292.50
0.00	29.19	32.20	0.62	28.40	22.22	19.78	24.00	0.51	30.15	99.00	29.83	270.00
0.00	28.88	32.42	1.57	33.89	23.89	20.56	23.50	0.38	30.15	96.00	29.83	247.50
0.00	29.20	25.90	3.48	30.78	24.44	22.22	19.00	0.09	30.13	92.00	29.83	180.00
0.00	29.18	27.83	0.75	30.93	23.95	21.43	29.00	0.60	30.15	96.00	29.83	270.00
0.00	29.27	35.00	0.25	29.44	23.89	21.67	21.33	0.56	30.27	96.00	29.83	270.00
0.00	29.09	31.75	0.71	30.93	24.63	22.04	29.73	0.41	30.11	100.00	29.83	270.00
0.00	28.95	27.33	2.98	34.26	26.67	25.00	23.13	0.24	30.01	97.07	29.83	270.00
0.00	28.95	34.44	2.90	32.31	24.86	23.89	27.50	0.35	30.01	100.00	29.83	270.00
0.00	29.13	35.44	2.09	30.69	23.61	21.67	27.00	0.06	29.97	96.50	29.83	247.50
0.00	29.24	18.70	2.87	33.19	25.37	22.35	24.00	0.11	30.30	96.00	29.83	225.00
0.00	29.31	31.33	0.78	30.07	23.70	21.17	23.88	0.24	30.31	100.00	29.83	
0.00	29.18	39.89	2.41	32.16	25.42	23.06	24.75	0.34	30.32	100.00	29.83	180.00
0.00	29.00	21.67	1.90	28.39	23.78	22.08	23.80	0.22	30.07	100.00	29.83	270.00
0.00	29.13	29.90	0.53	32.29	24.56	23.78	21.88	0.17	30.08	100.00	29.83	270.00
0.00	29.15	28.73	3.98	28.56	22.96	20.37	20.71	1.03	30.37	99.00	9.95	0.00
0.00	28.97	32.30	0.98	28.22	23.22	20.83	18.56	0.23	30.03	100.00	9.95	0.00
0.00	29.25	34.00	2.16	28.40	23.94	22.82	19.83	0.85	30.06	99.00	9.96	0.00
0.00	29.20	26.75	2.25	34.86	19.44	21.19	16.44	1.21	29.90	96.25	11.30	0.00
0.00	29.25	34.60	1.69	29.72	21.39	22.78	17.22	0.42	29.79	100.00	9.99	270.00
0.00	29.57	30.36	1.13	27.78	13.33	20.86	20.27	0.47	29.98	99.27	9.97	0.00
0.00	29.14	23.36	3.50	32.22	17.78	21.85	19.67	0.26	29.88	95.91	9.97	0.00
0.00	29.03	40.73	1.49	29.50	25.20	23.59	26.38	0.62	30.22	100.00	9.97	0.00
0.00	29.42	35.10	2.21	28.67	17.50	22.53	21.50	0.37	30.00	94.73	9.97	0.00
0.00	28.83	32.91	3.49	29.70	23.50	21.67	18.63	0.72	30.16	92.66	10.00	0.00
0.00	28.87	32.22	2.40	29.11	21.67	19.56	17.07	0.67	30.09	89.93	10.00	0.00
0.00	28.83	32.50	0.87	31.01	25.30	22.98	17.43	0.72	30.08	90.88	10.00	0.00
0.00	28.99	36.18	3.50	29.89	24.24	21.92	19.38	1.60	30.16	94.75	10.00	0.00
0.00	28.90	41.90	1.38	30.31	24.14	22.58	19.70	0.80	30.08	98.00	10.00	0.00
0.00	28.90	26.89	4.59	32.37	24.83	23.44	17.64	0.51	30.05	96.38	10.00	0.00
0.00	29.02	37.20	2.59	29.31	23.50	22.22	18.50	2.80	30.18	94.25	10.00	0.00
0.00	28.34	38.40	0.69	30.19	25.00	22.96	30.13	1.42	30.17	99.13	10.00	0.00
0.00	29.12	30.30	0.78	31.39	25.11	22.50	14.89	0.81	30.22	94.13	10.00	0.00
0.00	29.15	28.00	1.00	30.30	23.54	21.78	18.11	0.45	30.10	91.50	10.00	0.00
0.00	29.16	32.89	2.28	32.33	23.46	21.77	16.22	1.02	30.31	93.50	10.00	0.00
0.00	29.15	30.78	2.38	32.67	24.50	22.10	19.70	0.66	30.25	93.00	10.00	0.00
0.00	28.93	31.91	1.06	30.61	24.49	22.92	17.63	1.08	30.13	89.85	10.00	0.00
0.00	29.11	26.00	0.50	32.28	25.10	22.27	16.63	0.96	30.07	90.67	10.00	0.00
0.00	29.09	23.70	2.16	30.91	23.61	21.06	21.56	0.84	30.23	88.88	10.00	0.00
0.00	28.95	35.45	0.45	31.30	23.27	22.56	19.82	0.55	30.11	96.00	10.00	0.00
0.00	29.05	38.50	6.22	30.67	23.89	21.54	21.38	0.65	30.12	88.67	10.00	0.00

Table 2.3 – Continued NOAA Scalar Statistics – Min, Max and per Year 1979-2021

4.3 Hawk Mountain and NOAA Daily Fluctuations Year-Over-Year

Table 3, figures 7 and 8 are showing the derived attributes that quantify the 24-hour absolute value of each day-to-day change from previous day to the current day for: HMtempC_24_mean (light blue), Visibility_24_mean (orange), CloudCover_24_mean (gray), HourlyDryBulbTemperature_24_mean (yellow), HourlyWetBulbTemperature_24_mean (lighter blue), HourlyDewPointTemperature_24_mean (green), HourlyStationPressure_24_mean (dark blue), HourlyRelativeHumidity_24_mean (brown). The attributes prefixed with "Hourly" are NOAA data and those without it are from Hawk Mountain. This shows the variability of certain important weather attributes. We want to get a good idea if certain years are showing abnormal, increasing, or decreasing fluctuations. There are not many hard fluctuations over the last 43 years. There are a few attributes that had some missing data (see station pressure and wet bulb temperature on figure 8) but we will be disregarding this as we couldn't find a reason for the missing data.

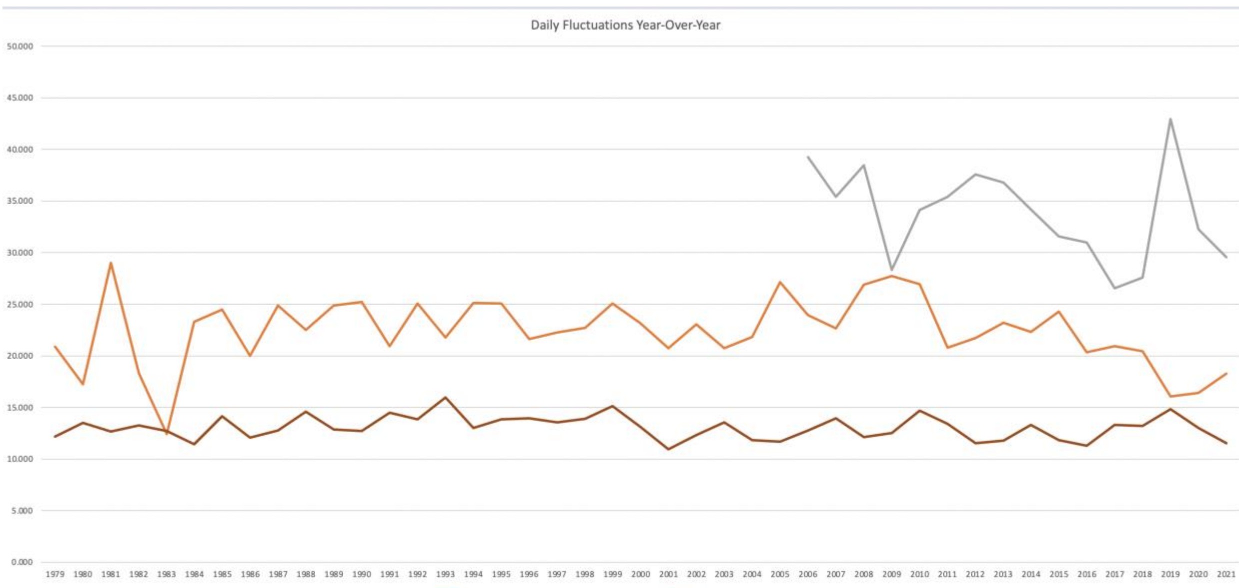

Figure 7 – Daily Fluctuations Averaged Absolute Value per Year 1979-2021

Figure 8 – Daily Fluctuations Averaged Absolute Value per Year 1979-2021

year	HMtempC_24_mean	Visibility_24_mean	CloudCover_24_mean	HourlyDryBulbTemperature_24_mean	HourlyWetBulbTemperature_24_mean	HourlyDewPointTemperature_24_mean	HourlyStationPressure_24_mean	HourlyRelativeHumidity_24_mean	HourlyVisibility_24_mean
1979	3.146	20.867		2.556	2.697	4.000	0.149	12.179	6.417
1980	3.310	17.261		3.333	2.824	4.223	0.158	13.497	8.760
1981	2.764	29.005		2.547	2.598	4.049	0.146	12.659	8.250
1982	2.966	18.340		3.281	2.798	4.204	0.134	13.274	7.554
1983	3.311	12.432		3.068	2.713	4.135	0.141	12.699	7.125
1984	3.139	23.308		2.879	2.831	4.186	0.148	11.426	6.771
1985	3.049	24.505		3.155	3.038	4.630	0.145	14.128	8.024
1986	3.647	20.024		3.493	3.461	4.889	0.168	12.073	7.960
1987	3.279	24.887		3.230	3.028	4.252	0.144	12.770	7.454
1988	3.137	22.525		3.022	3.007	4.670	0.177	14.593	8.316
1989	3.319	24.886		3.521	3.360	4.928	0.153	12.864	7.823
1990	3.552	25.212		2.993	2.889	4.092	0.140	12.727	7.182
1991	3.271	20.933		3.223	3.064	4.817	0.143	14.485	7.356
1992	3.321	25.078		2.933	2.902	4.377	0.150	13.853	7.591
1993	3.686	21.767		3.875	3.590	5.033	0.178	15.992	7.850
1994	3.285	25.133		2.934	2.969	4.474	0.148	13.023	8.144
1995	4.154	25.089		2.904	3.086	4.758	0.176	13.863	4.534
1996	2.923	21.640		3.044	2.680	4.010	0.141	13.927	0.685
1997	3.214	22.264		2.714	2.462	3.974	0.144	13.549	0.758
1998	2.483	22.722		3.005	2.929	4.208	0.130	13.909	1.107
1999	3.314	25.076		3.421	3.611	4.952	0.625	15.144	0.980
2000	2.695	23.168		3.035		4.110		13.115	0.625
2001	2.575	20.727		3.078		3.842		10.937	0.655
2002	2.536	23.042		3.111	0.556	4.006	0.155	12.301	0.677
2003	2.698	20.732		2.963	3.037	4.428	0.166	13.534	1.111
2004	3.169	21.835		2.825		3.551		11.832	0.797
2005	2.472	27.163		2.724	2.636	4.107	0.162	11.700	0.999
2006	2.907	23.933	39.228	3.068	2.725	3.950	0.156	12.760	1.180
2007	2.556	22.651	35.417	2.787	2.613	3.926	0.169	13.936	1.525
2008	2.683	26.887	38.442	2.995	2.937	4.110	0.162	12.106	0.845
2009	2.941	27.750	28.324	2.819	2.698	3.674	0.163	12.530	0.948
2010	2.795	26.967	34.132	2.985	2.650	4.227	0.159	14.685	0.698
2011	2.923	20.778	35.404	3.040	2.883	3.881	0.153	13.400	0.643
2012	2.679	21.716	37.568	2.515	2.347	3.676	0.157	11.522	1.088
2013	2.785	23.193	36.811	3.037	2.788	3.957	0.158	11.768	0.886
2014	2.872	22.300	34.159	3.316	2.967	4.325	0.148	13.297	0.986
2015	2.890	24.280	31.595	3.015	2.854	4.118	0.153	11.844	0.705
2016	3.099	20.328	30.958	2.641	2.579	3.924	0.133	11.287	0.628
2017	3.126	20.944	26.535	3.046	2.761	4.015	0.134	13.305	0.624
2018	2.792	20.428	27.577	3.389	2.905	4.017	0.165	13.236	0.930
2019	2.966	16.091	42.963	2.988	3.050	4.865	0.156	14.818	0.826
2020	2.623	16.387	32.266	3.163	3.066	4.276	0.152	13.019	0.688
2021	3.100	18.301	29.566	2.976	2.738	3.669	0.146	11.558	0.140

Table 3 – NOAA Day-to-Day Averaged Absolute Value per Year 1979-2021

4.4 Hawk Mountain Pearson Correlation Coefficient Matrix

A correlation coefficient matrix was generated, see Table 4. The focus here is the correlation coefficients statistics under the TOTAL_mean columns. The mean total raptor counts year-over-year were taken instead of the sum so there can be consistency. Before generating the correlation coefficient matrix, the attribute data was normalized due to the varying ranges of the data as can be seen on Figure 9. Normalizing the target attribute is not common when running regressors on the original data, but this data is a year-to-year average, so normalizing the target attribute was an okay strategy. We are not using the TOTAL raptor count sum, we are using the TOTAL_mean, again, to keep the matrix uniform. For future regression analysis, the target attribute will most likely not be normalized. Two attributes, CloudCover_mean and FlightHT_mean had some spotty data so the coefficients will be skewed but they are being kept here just for added information and for future use (See Appendix A for code).

To summarize the coefficient matrix; the biggest positive correlations to the total raptor count average are WindSpd_mean (blue), Visibility_mean (green) and last is HMtempC_mean (orange). The biggest negative correlations are FlightHT_mean (light blue), CloudCover_mean (gray), SkyCode_mean (yellow). This correlation distribution seems to fit with our understanding of hawk migration. The understanding here is that if there is higher visibility and have generally more windy days, then that’s when there are higher raptor counts. The small positive correlation coefficient between temperature and the average raptor count was not expected.

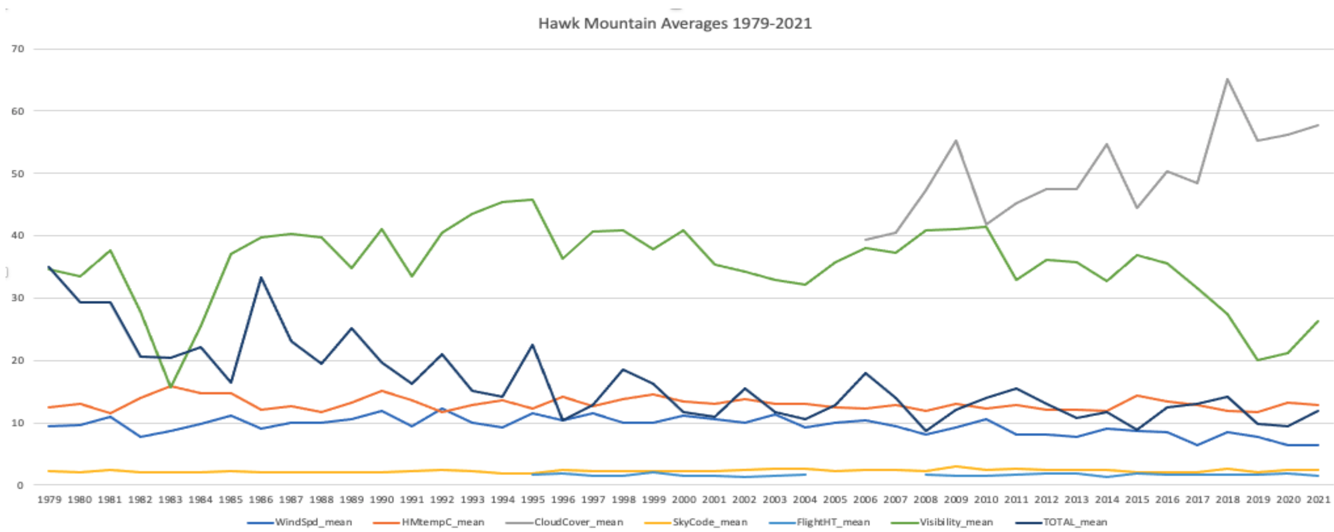

Figure 9 – Hawk Mountain Averages per Year 1979-2021

	WindSpd_mean	HMtempC_mean	CloudCover_mean	SkyCode_mean	FlightHT_mean	Visibility_mean	TOTAL_mean
WindSpd_mean	1	0.124668316	-0.698042532	-0.031864117	-0.366635942	0.589484366	0.327269703
HMtempC_mean	0.124668316	1	-0.398786301	-0.303558972	-0.151344423	-0.195832377	0.014516022
CloudCover_mean	-0.698042532	-0.398786301	1	0.347912496	0.531797176	-0.311294056	-0.516441099
SkyCode_mean	-0.031864117	-0.303558972	0.347912496	1	0.29785919	-0.059678004	-0.284315365
FlightHT_mean	-0.366635942	-0.151344423	0.531797176	0.29785919	1	-0.12085016	-0.660697785
Visibility_mean	0.589484366	-0.195832377	-0.311294056	-0.059678004	-0.12085016	1	0.131788915
TOTAL_mean	0.327269703	0.014516022	-0.516441099	-0.284315365	-0.660697785	0.131788915	1

Table 4 – Hawk Mountain Pearson Correlation Coefficient Matrix

4.5 NOAA Pearson Correlation Coefficient Matrix

Similarly, as done for the Hawk Mountain data, the merged NOAA data was run through, and a correlation coefficient matrix was generated. See Figure 10 for the data visualized and see Table 5 for the matrix (See Appendix A for code). The attribute data was normalized first and then the correlation coefficient matrix was generated.

To summarize the coefficient matrix; the biggest positive correlations to the total raptor count average are HourlyWindSpeed_mean (yellow), HourlyStationPressure_mean (green), HourlyVisibility_mean (brown), HourlyDryBulbTemperature_mean (blue), HourlyWetBulbTemperature_mean (orange). The biggest negative correlations to the total raptor count average are HourlyPrecipitation_mean (light blue), HourlyRelativeHumidity_mean (dark blue), HourlyDewPointTemperature_mean (gray). This correlation distribution, like the Hawk Mountain correlation coefficient matrix, is quantifying our understanding of what attributes matter when we hope to see higher averages of raptors migrating. The correlation coefficients seem to be a lot more pronounced here using the NOAA data as opposed to the Hawk Mountain data. It seems there's a high correlation between wind speed and visibility. Here, we see a slightly higher positive correlation between temperature and raptor counts. We were expecting a negative correlation here due to our original perception that as temperature increases due to localized climate change, we should see a decrease in raptor counts. This is not the case; it seems that there's a general decrease in temperature since 1979 just as there is a decrease in raptor counts since then.

There are two things to add as a side note here. The average station pressure does not seem to fluctuate much at all over the years, which can be the reason why it yielded a high correlation coefficient of .735. So, it seems that it may not be useful but again we will leave the attribute in place for future use. The other thing to note is the sudden drop in average visibility in 1996. There was no real explanation as to why the drop in visibility so this attribute should be either disregarded in future research or further studied.

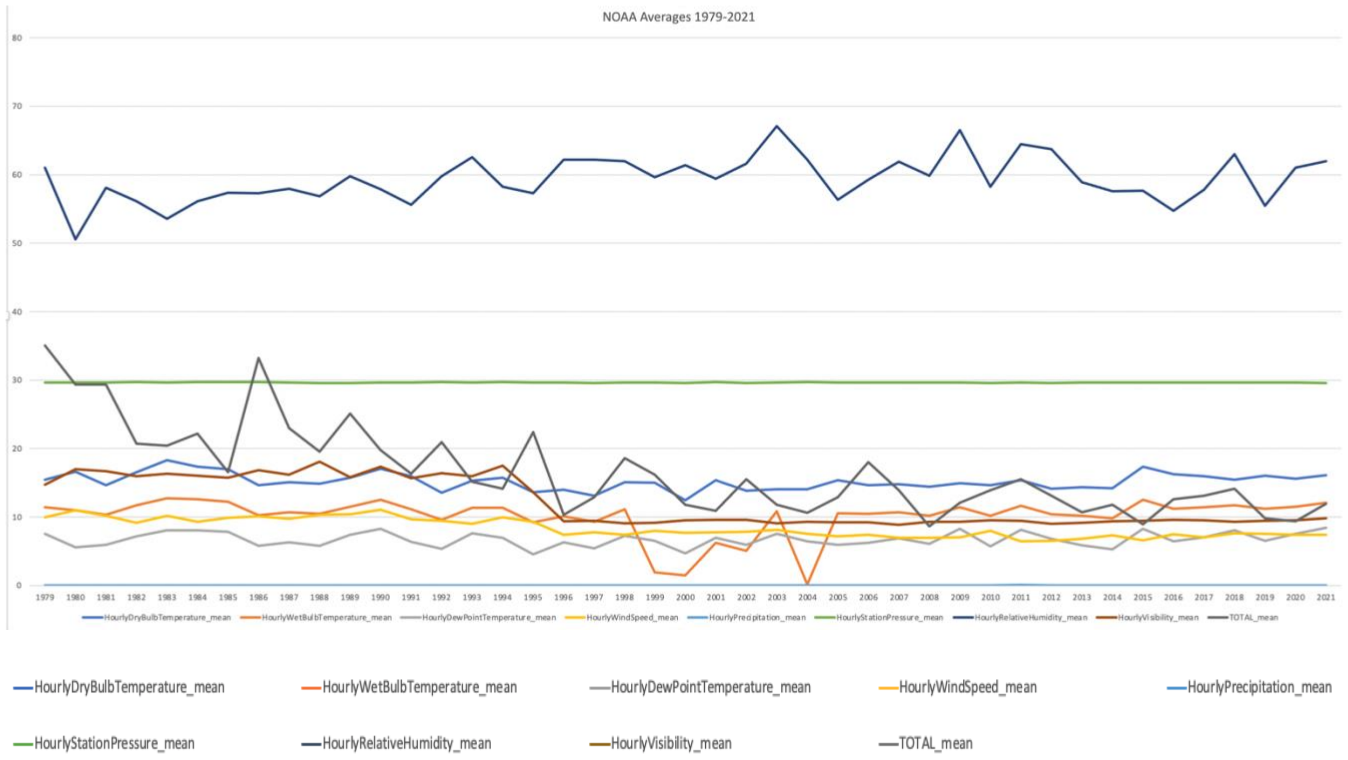

Figure 10 – Hawk Mountain Averages

	DryBulbTemperature_mean	WetBulbTemperature_mean	DewPointTemperature_mean	WindSpeed_mean	Precipitation_mean	StationPressure_mean	RelativeHumidity_mean	Visibility_mean	TOTAL_mean
DryBulbTemperatu	1.000	0.570	0.674	0.381	-0.133	0.301	-0.538	0.434	0.222
WetBulbTemperat	0.570	1.000	0.479	0.213	0.034	0.066	-0.208	0.330	0.206
DewPointTempera	0.674	0.479	1.000	-0.014	0.178	-0.011	0.244	0.053	-0.063
WindSpeed_mean	0.381	0.213	-0.014	1.000	-0.656	0.835	-0.414	0.944	0.770
Precipitation_mea	-0.133	0.034	0.178	-0.656	1.000	-0.594	0.318	-0.579	-0.394
StationPressure_m	0.301	0.066	-0.011	0.835	-0.594	1.000	-0.273	0.829	0.735
RelativeHumidity_	-0.538	-0.208	0.244	-0.414	0.318	-0.273	1.000	-0.420	-0.269
Visibility_mean	0.434	0.330	0.053	0.944	-0.579	0.829	-0.420	1.000	0.727
TOTAL_mean	0.222	0.206	-0.063	0.770	-0.394	0.735	-0.269	0.727	1.000

Table 5 – NOAA Pearson Correlation Coefficient Matrix

4.6 Hawk Mountain Wind Direction Attribute Analysis

The reason for more closely analyzing wind direction attributes is due to the high correlation that were seen in the previous section. It is also common knowledge among hawk enthusiasts and scientists at Hawk Mountain that north and northwest winds aka tail winds are used by hawks of all species to aid in their migration [2]. Another bit of common knowledge is that the northwesterly winds will hit the ridge at Hawk Mountain which will in turn create updrafts which help the hawks migrate (See Figure 2). Table 6 and Figure 12 visualize and quantify the daily tallies of Hawk Mountain's recorded wind direction averaged on a year-to-year basis.

Just as we did with the averaged Hawk Mountain and NOAA weather attributes, we normalized this wind direction data and generated a Pearson correlation coefficient matrix. The correlation matrix can be seen on Table 7. The top two notable wind directions that were seen as having the highest positive correlation with the TOTAL_mean raptor counts are the following: wndN, wndNW. These are winds coming from the north blowing towards the south and northwest blowing towards the southeast. The top 3 notable wind directions that were seen as having the highest negative correlation with TOTAL_mean raptor counts are the following: wndW, wndWNW, wndWSW. These are all winds coming from the west. If we take a look at Figure 2, we can see that the Kittatinny ridge is facing slightly northwest so, these winds that coming from the west seem the be head winds for the migrating raptors.

This falls in line with what prior Hawk Mountain research shows – that when we have higher wind speeds (see Sections 4.4 and 4.5 above) and we have more north and northwesterly winds coming through then we will have a higher chance of seeing migrating raptors flying by. This seems to not only hold true for daily behavior, but it holds true for year-after-year autumnal behavior.

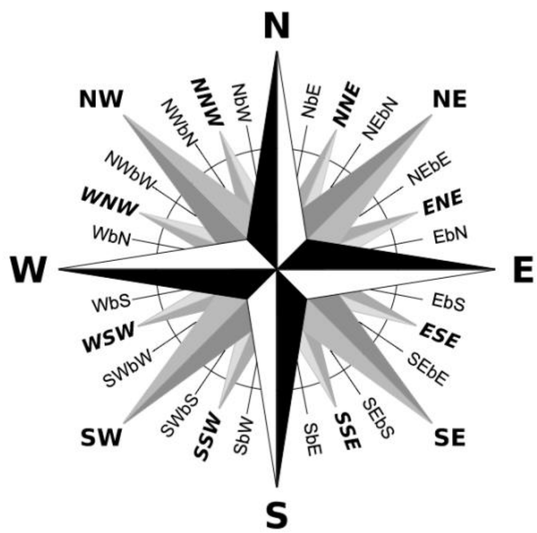

Figure 11 – 32-point compass [*https://en.wikipedia.org/wiki/Points_of_the_compass*](https://en.wikipedia.org/wiki/Points_of_the_compass)

year	wndN	wndNNE	wndNE	wndENE	wndE	wndESE	wndSE	wndSSE	wndS	wndSSW	wndSW	wndWSW	wndW	wndWNW	wndNW	wndNNW	wndUNK	TOTAL_mean
1979	110	0	6	0	20	0	43	0	95	2	105	1	39	0	271	3	95	35.08987342
1980	127	0	2	0	4	0	84	0	50	2	73	0	68	5	428	1	61	29.36022099
1981	129	0	29	0	24	0	68	0	43	0	82	0	36	1	345	1	94	29.35328638
1982	89	3	11	0	11	1	97	1	61	2	53	2	106	4	293	0	176	20.7032967
1983	99	2	41	0	15	0	137	1	35	0	16	0	123	6	312	2	181	20.4556701
1984	116	0	54	0	60	0	157	0	41	0	38	0	51	2	420	0	74	22.15794669
1985	80	0	22	0	33	0	119	0	39	1	13	0	73	0	394	0	78	16.53521127
1986	82	0	0	0	29	0	79	0	28	0	27	0	110	0	425	0	103	33.23329558
1987	29	0	11	0	5	0	58	1	108	0	75	0	105	0	477	0	104	22.99283521
1988	89	0	8	0	4	1	95	0	86	0	65	0	131	1	437	0	114	19.52376334
1989	85	0	12	0	28	0	95	1	130	0	31	0	134	0	356	0	115	25.12056738
1990	100	0	8	0	15	0	101	0	80	0	28	0	95	0	474	0	123	19.76074219
1991	110	0	20	0	3	0	121	0	110	0	125	0	132	0	350	2	89	16.31450094
1992	75	0	11	0	28	0	63	0	133	0	107	0	107	0	387	0	104	20.93103448
1993	98	0	20	0	13	0	102	0	152	0	79	0	112	1	375	2	96	15.16666667
1994	64	0	29	0	23	0	96	0	83	0	90	1	137	1	505	0	89	14.17889088
1995	27	1	19	4	34	32	54	46	84	21	79	16	138	160	252	29	79	22.44505495
1996	18	3	8	4	29	22	109	10	73	29	40	14	172	125	266	22	181	10.37411348
1997	62	1	15	0	19	12	106	27	72	7	74	24	167	132	316	25	90	12.92733278
1998	39	3	9	0	23	9	65	35	78	8	84	34	159	156	307	53	152	18.60621387
1999	53	16	25	5	45	35	165	37	55	6	55	37	109	99	302	30	114	16.21723896
2000	42	10	20	1	14	11	107	45	54	12	25	0	192	146	339	72	112	11.78881579
2001	50	6	13	12	22	6	74	61	102	14	83	14	165	127	245	68	166	10.96219282
2002	52	9	13	15	15	13	90	25	89	12	50	15	108	142	262	66	136	15.52646048
2003	44	3	23	6	60	20	60	59	73	18	55	19	135	129	205	56	151	11.76980874
2004	59	0	27	18	52	21	112	25	49	5	76	21	86	182	260	55	135	10.60555923
2005	25	5	9	10	32	41	108	64	62	15	51	24	115	183	295	39	121	12.93212366
2006	31	3	31	11	20	31	100	61	59	24	42	19	110	226	268	38	123	17.99930168
2007	25	7	15	5	38	10	146	33	75	8	77	22	170	134	257	36	122	13.93013699
2008	78	5	47	4	43	7	50	51	72	11	25	19	178	168	239	23	234	8.679867987
2009	35	15	43	24	37	24	135	24	55	4	34	5	169	92	275	37	82	12.07882883
2010	33	4	37	0	7	10	58	20	67	10	43	8	210	263	259	38	199	13.93967213
2011	19	4	8	7	45	29	78	9	117	8	96	21	179	100	288	12	258	15.56624918
2012	32	4	24	7	56	16	87	10	114	9	89	15	137	98	321	25	245	13.10188679
2013	14	2	40	15	45	12	89	14	96	11	86	28	171	103	289	35	215	10.69467607
2014	50	4	53	12	51	8	113	21	118	8	38	34	150	105	328	12	183	11.79128739
2015	49	8	31	16	22	32	83	43	108	9	97	34	159	115	296	30	141	8.934573445
2016	24	15	33	2	40	18	76	25	125	14	27	6	193	121	375	9	124	12.57408684
2017	89	5	22	16	33	21	103	19	76	17	54	16	120	88	292	10	160	13.10344828
2018	47	7	23	3	36	21	126	51	63	12	39	8	106	120	287	28	137	14.1670446
2019	27	7	47	8	35	16	102	21	107	7	81	22	162	84	234	14	186	9.808844508
2020	37	14	55	4	12	18	34	34	138	42	111	19	195	74	251	15	126	9.414342629
2021	23	4	5	3	54	24	109	46	61	10	112	22	187	114	291	10	150	11.96254181

Table 6 – Hawk Mountain Wind Direction Daily Tallies per Year 1979-2021

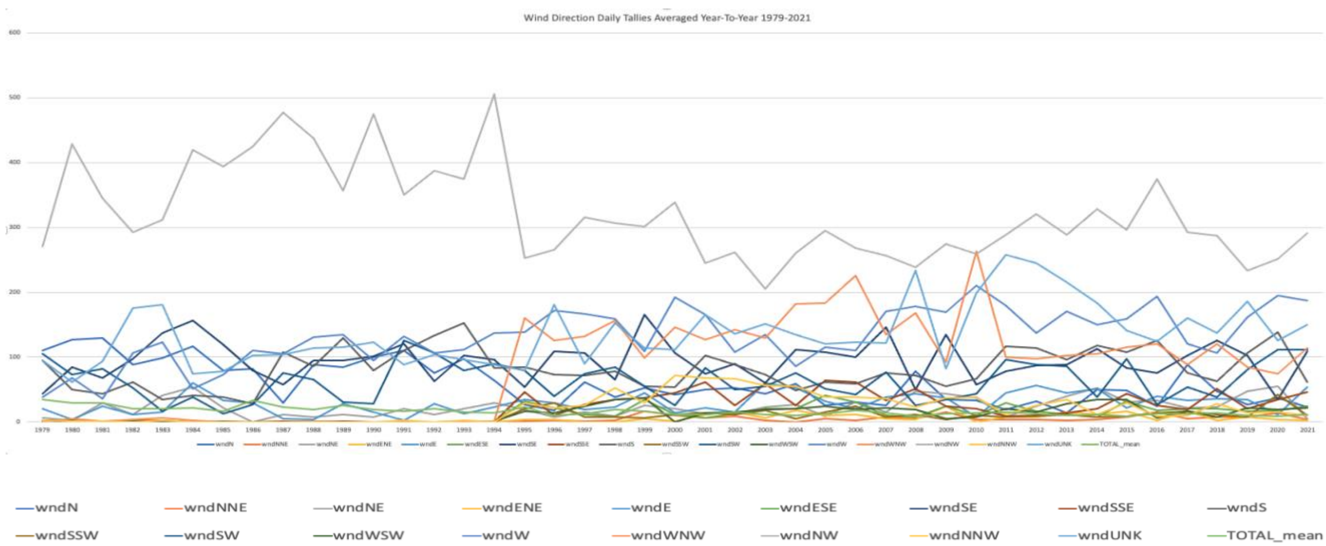

Figure 12 – Hawk Mountain Wind Direction Daily Tallies Year-To-Year 1979-2021

	'wndN'	'wndNNE'	'wndNE'	'wndENE'	'wndE'	'wndESE'	'wndSE'	'wndSSE'	'wndS'	'wndSSW'	'wndSW'	'wndWSW'	'wndW'	'wndWNW'	'wndNW'	'wndNNW'	'wndUNK'	'TOTAL_mean'
'wndN'	1.0000	-0.4156	-0.0979	-0.3292	-0.2785	-0.5805	0.0266	-0.4818	-0.1757	-0.4822	0.0153	-0.5283	-0.6968	-0.6219	0.5253	-0.4405	-0.3472	0.6387
'wndNNE'	-0.4156	1.0000	0.4102	0.4652	0.1896	0.5424	0.2740	0.4794	0.1575	0.4550	-0.0572	0.4096	0.5531	0.3999	-0.4300	0.4278	0.1894	-0.3969
'wndNE'	-0.0979	0.4102	1.0000	0.3539	0.3036	0.1672	0.1501	0.1882	0.0893	0.2780	-0.1065	0.2501	0.2696	0.2101	-0.3812	0.0921	0.2356	-0.3683
'wndENE'	-0.3292	0.4652	0.3539	1.0000	0.3990	0.5835	0.2842	0.4294	0.1018	0.3122	0.0659	0.5074	0.3051	0.5015	-0.4732	0.5617	0.3172	-0.3775
'wndE'	-0.2785	0.1896	0.3036	0.3990	1.0000	0.4655	0.2962	0.3914	-0.0487	0.1880	0.0009	0.4874	0.1577	0.3474	-0.3964	0.2814	0.3652	-0.2048
'wndESE'	-0.5805	0.5424	0.1672	0.5835	0.4655	1.0000	0.2429	0.7243	0.0478	0.6191	0.1224	0.6951	0.4405	0.7223	-0.5741	0.5097	0.2723	-0.3475
'wndSE'	0.0266	0.2740	0.1501	0.2842	0.2962	0.2429	1.0000	0.0800	-0.3820	-0.1632	-0.2716	0.1551	-0.0429	0.0553	-0.0030	0.1196	-0.0686	-0.1564
'wndSSE'	-0.4818	0.4794	0.1882	0.4294	0.3914	0.7243	0.0800	1.0000	0.0319	0.6329	0.0652	0.6333	0.5198	0.8068	-0.6566	0.7493	0.2707	-0.3720
'wndS'	-0.1757	0.1575	0.0893	0.1018	-0.0487	0.0478	-0.3820	0.0319	1.0000	0.2744	0.5163	0.1756	0.3088	-0.0247	-0.2556	-0.0092	0.2015	-0.1445
'wndSSW'	-0.4822	0.4550	0.2780	0.3122	0.1880	0.6191	-0.1632	0.6329	0.2744	1.0000	0.1525	0.4762	0.5672	0.6120	-0.6177	0.4542	0.3270	-0.3337
'wndSW'	0.0153	-0.0572	-0.1065	0.0659	0.0009	0.1224	-0.2716	0.0652	0.5163	0.1525	1.0000	0.2670	0.0769	-0.0175	-0.2002	0.0342	0.0455	0.1232
'wndWSW'	-0.5283	0.4096	0.2501	0.5074	0.4874	0.6951	0.1551	0.6333	0.1756	0.4762	0.2670	1.0000	0.4935	0.6717	-0.6603	0.5233	0.4718	-0.4371
'wndW'	-0.6968	0.5531	0.2696	0.3051	0.1577	0.4405	-0.0429	0.5198	0.3088	0.5672	0.0769	0.4935	1.0000	0.5862	-0.7441	0.4548	0.4815	-0.6367
'wndWNW'	-0.6219	0.3999	0.2101	0.5015	0.3474	0.7223	0.0553	0.8068	-0.0247	0.6120	-0.0175	0.6717	0.5862	1.0000	-0.7998	0.7888	0.4207	-0.4500
'wndNW'	0.5253	-0.4300	-0.3812	-0.4732	-0.3964	-0.5741	-0.0030	-0.6566	-0.2556	-0.6177	-0.2002	-0.6603	-0.7441	-0.7998	1.0000	-0.6274	-0.7642	0.4770
'wndNNW'	-0.4405	0.4278	0.0921	0.5617	0.2814	0.5097	0.1196	0.7493	-0.0092	0.4542	0.0342	0.5233	0.4548	0.7888	-0.6274	1.0000	0.2978	-0.3445
'wndUNK'	-0.3472	0.1894	0.2356	0.3172	0.3652	0.2723	-0.0686	0.2707	0.2015	0.3270	0.0455	0.4718	0.4815	0.4207	-0.7642	0.2978	1.0000	-0.3641
'TOTAL_mea	0.6387	-0.3969	-0.3683	-0.3775	-0.2048	-0.3475	-0.1564	-0.3720	-0.1445	-0.3337	0.1232	-0.4371	-0.6367	-0.4500	0.4770	-0.3445	-0.3641	1.0000

Table 7 – Wind Direction Daily Tallies Year-To-Year Correlation Coefficient Matrix

Chapter 5

5. Conclusion

Running foundational analysis can be hard especially when the subject matter has been and continues to be studied very widely and closely by many scientists and experts. Hawk Mountain Sanctuary has an amazing team of scientists, enthusiasts and volunteers that help uncover new trends and ideas surrounding global raptor conservation.

Using the latest and up-to-date python libraries and code, the initial data cleaning, sorting, aggregating, and wrangling of a merged Hawk Mountain and NOAA data was done. Guided by prior, extensive research the merged data was further sorted and analyzed.

The quantifying of known phenomena such as the positive correlation influence that a clear, windy, and dry day can have on a migrating raptor was done. The negative correlators were found as well. There were, on average, less raptors migrating when conditions were cloudy, less windy, less visible and when there were more headwinds on average for the year. The data of interest were correlated and visualized against the target attribute; the raptor counts.

Chapter 6

6. Future Work

The future for Hawk Mountain Sanctuary data analysis is bright. Researchers are constantly learning and exploring not only phenological events related to hawks but just everything about them ranging from nesting conditions, pollutants, satellite tracking, land conservation and more.

In terms of future work for this specific study and dataset, Kutztown University of Pennsylvania's Dr. Dale E. Parson is incorporating the data and analysis of this project into his courses. The collecting, merging, sorting, aggregating and fundamental analysis will be used going forward in order to try to uncover, correlate and explain the shifts in the Hawk's phenological behavior.

The continuation of analysis needs to be a careful one due to the nature of this merged data. The Hawk Mountain data are collected by hand and technologies used for the collection of this data are constantly evolving. There may be a human error component as well that should be investigated. The NOAA data holds a lot of attributes that could also be studied but, again, due to the difference in distance and topology that the Allentown NOAA weather station is at in relation to Hawk Mountain, there needs to be careful consideration of which attributes to consider and use. The work presented here is a good starting point for the continuation of hawk migration analysis and its relationship to weather attributes.

Bibliography

[1] Analysis of Hawk Mountain Sanctuary Observation Data from 1948 through 2021. https://acad.kutztown.edu/~parson/HawkMtnDaleParson2022/.

[2] *Autumn hawk migration*. Hawk Mountain Sanctuary. (n.d.). Retrieved February 14, 2022, from https://www.hawkmountain.org/visit/events/autumn-hawk-migration

[3] Dr. Laurie Goodrich, email message regarding observation periods, July 12, 2022.

[4] Farmer, C., D. Husseli, and D. Mizrahi. 2007. Detecting Population Trends in Migratory Birds of Prey. The Auk 124:1047-1062.

[5] "Hawk Mountain Trail System." *Hawk Mountain Sanctuary*, https://www.hawkmountain.org/visit/hiking/trail-descriptions.

[6] Horton, Convening Lead Authors Radley. "Menu." *National Climate Assessment*, https://nca2014.globalchange.gov/report/regions/northeast.

[7] "Learn." *Scikit*, https://scikit-learn.org/stable/index.html.

[8] "National Oceanic and Atmospheric Administration." *Homepage | National Oceanic and Atmospheric Administration*, https://www.noaa.gov/.

[9] National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Information, Allentown Lehigh Valley International Airport weather data from 1948 through 2021, https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USW00014737/detail

[10] *NumPy*, https://numpy.org/.

[11] “Pandas.dataframe.corr#.” *Pandas.DataFrame.corr - Pandas 1.5.1 Documentation*, https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html.

[12] “Package Overview#.” *Package Overview - Pandas 1.5.0 Documentation*, https://pandas.pydata.org/docs/getting_started/overview.html.

[13] “Quick Start Guide#.” *Quick Start Guide - Matplotlib 3.6.0 Documentation*, https://matplotlib.org/stable/tutorials/introductory/quick_start.html.

[14] “Raptorpedia.” *Hawk Mountain Sanctuary*, https://www.hawkmountain.org/conservation-science/resources/raptorpedia.

[15] “Statistical Data Visualization#.” *Seaborn*, https://seaborn.pydata.org/.

[16] 219. Therrien, J.-F, L. J. Goodrich, D.R. Barber, and K.L. Bildstein. 2012. A long-term database on raptor migration at Hawk Mountain Sanctuary, northeastern United States. Ecology 93(8): 1979

[17] “What Is Time Series Data?: Definition, Examples, Types & Uses.” *InfluxData*, 19 Sept. 2022, https://www.influxdata.com/what-is-time-series-data/.

[18] Wilbanks, Tom, and Steve Fernandez. *CLIMATE CHANGE AND INFRASTRUCTURE, URBAN SYSTEMS, AND VULNERABILITIES*.

Appendix A

Code

Hawk Mountain Data

```
# Imports will be consistent for all code
import pandas as pd
import numpy as np
from numpy.polynomial.polynomial import polyfit
import os
import seaborn as sns
from matplotlib import pyplot as plt

# Load data to work with
os.popen('cd /Users/ericbu/HMThesis ls').read()
path = r'/Users/ericbu/HMThesis/data/FinalHMSHOURLY1966-2015_v1_in_csv.csv'
df_full = pd.read_csv(path)

# Show full dataframe information
df_full.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49304 entries, 0 to 49303
Columns: 152 entries, DAY to birdBin
dtypes: int64(20), object(132)
memory usage: 57.2+ MB


print(df_full.columns.tolist())

['DAY', 'MONTH', 'YEAR', 'JULIAN', 'HOUR_OF_DAY', 'MINUTES', 'WEATHER', 'NUMBER_OF_OBSE
RVERS', 'MAX_VISIBILITY', 'CLOUD_COVER', 'TEMP', 'SKY_CODE', 'WIND_SPEED', 'WIND_DIR', 'F
LIGHT_ALT', 'FLIGHT_DIR', 'TUVU', 'BLVU', 'OSPR', 'BAEA_UNAGED', 'BAEA_IMM', 'BAEA_ADUL
T', 'NOHA', 'SSHA', 'COHA', 'NOGO', 'RSHA', 'BWHA', 'SWHA', 'RTHA', 'RLHA', 'GOEA_UNAGED', 'GO
EA_IMM', 'GOEA_ADULT', 'AMKE', 'MERL', 'PEFA', 'GYRF', 'MIKI', 'STKI', 'UNID_ACCIPITER', 'UNID
_BUTEO', 'UNID_EAGLE', 'UNID_FALCON', 'UNID_RAPTOR', 'OTHER', 'TOTAL_RAPTORS', 'PEOPLE
', 'NAMES', 'windSpeedVal', '"hawkStart date"', 'tempMinutes', 'tempHours', 'tempDay', 'seconds_since_midni
ght', 'seconds_since_newYear', 'temp24Hr', '24HrAgoTemp', '48HrAgoTemp', '72HrAgoTemp', 'WeekAgoTem
p', 'MonthAgoTemp', '24HrAgoVisibility', '48HrAgoVisibility', '72HrAgoVisibility', 'WeekAgoVisibility', 'Mo
nthAgoVisibility', '24HrAgoCLOUDCOVER', '48HrAgoCLOUDCOVER', '72HrAgoCLOUDCOVER', 'Week
AgoCLOUDCOVER', 'MonthAgoCLOUDCOVER', '24HrAgoWINDSPEED', '48HrAgoWINDSPEED', '72Hr
AgoWINDSPEED', 'WeekAgoWINDSPEED', 'MonthAgoWINDSPEED', 'consWind', 'consWindDir', 'consWe
ather', '24HrTempDiff', '48HrTempDiff', '72HrTempDiff', 'WeekTempDiff', 'MonthTempDiff', '24HrVisibility
Diff', '48HrVisibilityDiff', '72HrVisibilityDiff', 'WeekVisibilityDiff', 'MonthVisibilityDiff', '24HrCloudCoverD
iff', '48HrCloudCoverDiff', '72HrCloudCoverDiff', 'WeekCloudCoverDiff', 'MonthCloudCoverDiff', '24HrWin
dSpeedDiff', '48HrWindSpeedDiff', '72HrWindSpeedDiff', 'WeekWindSpeedDiff', 'MonthWindSpeedDiff', 'Pe
opleMinutes', 'BirdsPerPeopleMinute', 'DAILYMINUTES', '"DAILYNUMBER OF OBSERVERS"', '"DAILY
MAX VISIBILITY"', '"DAILYMIN VISIBLITY"', '"DAILYCLOUD COVER"', '"DAILYHIGH TEMP"', '"D
AILYLOW TEMP"', '"DAILYWIND SPEED"', '"DAILYFLIGHT ALT"', 'DAILY24HrAgoVisibility', 'DAIL
Y48HrAgoVisibility', 'DAILY72HrAgoVisibility', 'DAILYWeekAgoVisibility', 'DAILYMonthAgoVisibility',
'DAILY24HrAgoCLOUDCOVER', 'DAILY48HrAgoCLOUDCOVER', 'DAILY72HrAgoCLOUDCOVER', '
```

DAILYWeekAgoCLOUDCOVER', 'DAILYMonthAgoCLOUDCOVER', 'DAILY24HrAgoWINDSPEED', 'DAILY48HrAgoWINDSPEED', 'DAILY72HrAgoWINDSPEED', 'DAILYWeekAgoWINDSPEED', 'DAILYMonthAgoWINDSPEED', 'DAILYconsWind', 'DAILYconsWindDir', 'DAILYconsWeather', 'DAILY24HrTempDiff', 'DAILY48HrTempDiff', 'DAILY72HrTempDiff', 'DAILYWeekTempDiff', 'DAILYMonthTempDiff', 'DAILY24HrVisibilityDiff', 'DAILY48HrVisibilityDiff', 'DAILY72HrVisibilityDiff', 'DAILYWeekVisibilityDiff', 'DAILYMonthVisibilityDiff', 'DAILY24HrCloudCoverDiff', 'DAILY48HrCloudCoverDiff', 'DAILY72HrCloudCoverDiff', 'DAILYWeekCloudCoverDiff', 'DAILYMonthCloudCoverDiff', 'DAILY24HrWindSpeedDiff', 'DAILY48HrWindSpeedDiff', 'DAILY72HrWindSpeedDiff', 'DAILYMonthWindSpeedDiff', 'DAILYPeopleMinutes', 'DAILYyesterdayBirdsPerPeopleMinute', 'DAILYyesterdayBirds', 'birdBin']

File (FinalHMSHOURLY1966-2015_v1_in_csv.csv) available for download: https://drive.google.com/file/d/1OKUcMsSDxOXveqyvbh51wtwWiDUPjq3g/view?usp=sharing

NOAA Data

File weatherData_2021_NOAA.csv' can be downloaded here: https://drive.google.com/file/d/15tCW3mEHn0gSRJvm6VYghJeAGCiH1lUM/view?usp=sharing

NOAA Hourly Data

Load data to work with. Here, we are using original data from NOAA website just like we did for the above daily data.

```
os.popen('cd /Users/ericbu/HMThesis/ ls').read()
path = r'/Users/ericbu/HMThesis/new_yearly_data/weatherData_2021_NOAA.csv'
df_full = pd.read_csv(path)

noaa_weather_attributes= ['DATE', 'HourlyDewPointTemperature', 'HourlyDryBulbTemperature', 'HourlyPrecipitation', 'HourlyRelativeHumidity', 'HourlySeaLevelPressure', 'HourlyStationPressure', 'HourlyVisibility', 'HourlyWetBulbTemperature', 'HourlyWindDirection', 'HourlyWindSpeed', 'YEAR', 'MONTH', 'DAY', 'HOUR', 'MINUTE']
justWeather_reduced = df_full[noaa_weather_attributes].copy()
justWeather_reduced.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 736558 entries, 0 to 736557
Data columns (total 16 columns):
 #   Column                     Non-Null Count   Dtype
---  ------                     --------------   -----
 0   DATE                       736558 non-null  object
 1   HourlyDewPointTemperature  651922 non-null  object
 2   HourlyDryBulbTemperature   196496 non-null  object
 3   HourlyPrecipitation        588679 non-null  object
 4   HourlyRelativeHumidity     651908 non-null  object
 5   HourlySeaLevelPressure     608541 non-null  object
 6   HourlyStationPressure      600453 non-null  object
 7   HourlyVisibility           670179 non-null  object
 8   HourlyWetBulbTemperature   595096 non-null  object
 9   HourlyWindDirection        635986 non-null  object
```

```
 10  HourlyWindSpeed           669388 non-null  object
 11  YEAR                      736558 non-null  int64
 12  MONTH                     736558 non-null  int64
 13  DAY                       736558 non-null  int64
 14  HOUR                      736558 non-null  int64
 15  MINUTE                    736558 non-null  int64
dtypes: int64(5), object(11)
memory usage: 89.9+ MB
```

Hawk Mountain Autumnal-based Data Aggregation

```
# Each individual raptor species count is added and counted under the TOTAL_RAPTORS column.

# Load data to work with
os.popen('cd /Users/ericbu/HMThesis/data ls').read()
path = r'/Users/ericbu/HMThesis/data/FinalHMSHOURLY1966-2015_v1_in_csv.csv'
df_full = pd.read_csv(path)

hm_attributes = ['YEAR', 'MONTH', 'DAY', 'MAX_VISIBILITY', 'FLIGHT_ALT', 'CLOUD_COVER', 'TEMP', 'W
IND_SPEED', 'WIND_DIR', 'SKY_CODE', 'TOTAL_RAPTORS','NUMBER_OF_OBSERVERS']
df_reduced_hm_attributes = df_full[hm_attributes].copy()

df_reduced_hm_attributes.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49304 entries, 0 to 49303
Data columns (total 12 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   YEAR                 49304 non-null  int64
 1   MONTH                49304 non-null  int64
 2   DAY                  49304 non-null  int64
 3   MAX_VISIBILITY       49304 non-null  object
 4   FLIGHT_ALT           49304 non-null  object
 5   CLOUD_COVER          49304 non-null  object
 6   TEMP                 49304 non-null  object
 7   WIND_SPEED           49304 non-null  object
 8   WIND_DIR             49304 non-null  object
 9   SKY_CODE             49304 non-null  object
 10  TOTAL_RAPTORS        49304 non-null  int64
 11  NUMBER_OF_OBSERVERS  49304 non-null  object
dtypes: int64(4), object(8)
memory usage: 4.5+ MB


# Replace ? records with nan value for us to be able to turn these records into float type and run arithmetic oper
ations.

df_reduced_hm_attributes = df_reduced_hm_attributes.replace("?", np.nan)
df_reduced_hm_attributes = df_reduced_hm_attributes[(df_reduced_hm_attributes['MONTH'] != 1) & (df_red
uced_hm_attributes['MONTH'] != 2) & (df_reduced_hm_attributes['MONTH'] != 7)]
```

```
for col in [x for x in df_reduced_hm_attributes.columns if x != 'WIND_DIR' and x != 'YEAR' and x != 'MONT
H' and x != 'DAY'  ]:
   df_reduced_hm_attributes[col] = df_reduced_hm_attributes[col].astype(float)

df_reduced_hm_attributes.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 48648 entries, 0 to 49292
Data columns (total 12 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   YEAR                 48648 non-null  int64
 1   MONTH                48648 non-null  int64
 2   DAY                  48648 non-null  int64
 3   MAX_VISIBILITY       45271 non-null  float64
 4   FLIGHT_ALT           30270 non-null  float64
 5   CLOUD_COVER          10873 non-null  float64
 6   TEMP                 39701 non-null  float64
 7   WIND_SPEED           45605 non-null  float64
 8   WIND_DIR             44289 non-null  object
 9   SKY_CODE             44769 non-null  float64
 10  TOTAL_RAPTORS        48648 non-null  float64
 11  NUMBER_OF_OBSERVERS  48558 non-null  float64
dtypes: float64(8), int64(3), object(1)
memory usage: 4.8+ MB
```

NOAA Autumnal-based Data Aggregation (Daily)

```
# Load data to work with
os.popen('cd /Users/ericbu/HMThesis/ ls').read()
path = r'/Users/ericbu/HMThesis/new_yearly_data/weatherData_2021_NOAA.csv'
df_full = pd.read_csv(path)

noaa_daily_weather_attribute = ['DATE', 'DailyAverageDryBulbTemperature', 'DailyAverageRelativeHumidit
y', 'DailyAverageStationPressure', 'DailyAverageWindSpeed', 'DailyCoolingDegreeDays', 'DailyDepartureFro
mNormalAverageTemperature', 'DailyHeatingDegreeDays', 'DailyMaximumDryBulbTemperature', 'DailyMini
mumDryBulbTemperature', 'DailyPeakWindDirection', 'DailyPeakWindSpeed', 'DailyPrecipitation', 'DailySno
wfall', 'DailySustainedWindDirection', 'DailySustainedWindSpeed', 'DailyWeather', 'Sunrise', 'Sunset', 'YEAR',
'MONTH', 'DAY']
justWeather_reduced = df_full[noaa_daily_weather_attribute].copy()

justWeather_reduced = justWeather_reduced[(justWeather_reduced['YEAR'] > 1965) &
                            (justWeather_reduced['YEAR'] < 2016)]
justWeather_reduced = justWeather_reduced[(justWeather_reduced['MONTH'].astype(int) != 1)
                  & (justWeather_reduced['MONTH'].astype(int) != 2)
                  & (justWeather_reduced['MONTH'].astype(int) != 3)
                  & (justWeather_reduced['MONTH'].astype(int) != 4)
                  & (justWeather_reduced['MONTH'].astype(int) != 5)
                  & (justWeather_reduced['MONTH'].astype(int) != 6)
                  & (justWeather_reduced['MONTH'].astype(int) != 7)]
```

```
# Show metadata of data after keeping only Autumn data.
justWeather_reduced.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 259316 entries, 621649 to 764132
Data columns (total 22 columns):
 #   Column                                  Non-Null Count   Dtype
---  ------                                  --------------   -----
 0   DATE                                    259316 non-null  object
 1   DailyAverageDryBulbTemperature          3867 non-null    object
 2   DailyAverageRelativeHumidity            2331 non-null    float64
 3   DailyAverageStationPressure             3879 non-null    float64
 4   DailyAverageWindSpeed                   3876 non-null    float64
 5   DailyCoolingDegreeDays                  3867 non-null    object
 6   DailyDepartureFromNormalAverageTemperature  3867 non-null    object
 7   DailyHeatingDegreeDays                  3867 non-null    object
 8   DailyMaximumDryBulbTemperature          3867 non-null    object
 9   DailyMinimumDryBulbTemperature          3868 non-null    object
 10  DailyPeakWindDirection                  3843 non-null    object
 11  DailyPeakWindSpeed                      3863 non-null    object
 12  DailyPrecipitation                      3875 non-null    object
 13  DailySnowfall                           2591 non-null    object
 14  DailySustainedWindDirection             3874 non-null    object
 15  DailySustainedWindSpeed                 3882 non-null    object
 16  DailyWeather                            2350 non-null    object
 17  Sunrise                                 3924 non-null    float64
 18  Sunset                                  3924 non-null    float64
 19  YEAR                                    259316 non-null  int64
 20  MONTH                                   259316 non-null  int64
 21  DAY                                     259316 non-null  int64
```

NOAA Autumnal-based Data Aggregation (Hourly)

```
# Load data to work with
os.popen('cd /Users/ericbu/HMThesis/ ls').read()
path = r'/Users/ericbu/HMThesis/new_yearly_data/weatherData_2021_NOAA.csv'
df_full = pd.read_csv(path)

# Grab only hourly attributes
noaa_weather_attributes= ['DATE', 'HourlyDewPointTemperature', 'HourlyDryBulbTemperature', 'HourlyPrec
ipitation', 'HourlyRelativeHumidity', 'HourlySeaLevelPressure', 'HourlyStationPressure',              'HourlyVisib
ility',         'HourlyWetBulbTemperature', 'HourlyWindDirection', 'HourlyWindSpeed', 'YEAR', 'MONTH', 'DA
Y', 'HOUR', 'MINUTE']
justWeather_reduced = df_full[noaa_weather_attributes].copy()
justWeather_reduced = justWeather_reduced[(justWeather_reduced['YEAR'] > 1965) &
                            (justWeather_reduced['YEAR'] < 2016)]
justWeather_reduced = justWeather_reduced[(justWeather_reduced['MONTH'].astype(int) != 1)
                  & (justWeather_reduced['MONTH'].astype(int) != 2)
                  & (justWeather_reduced['MONTH'].astype(int) != 3)
                  & (justWeather_reduced['MONTH'].astype(int) != 4)
```

```
                 & (justWeather_reduced['MONTH'].astype(int) != 5)
                 & (justWeather_reduced['MONTH'].astype(int) != 6)
                 & (justWeather_reduced['MONTH'].astype(int) != 7)]

# Show metadata

justWeather_reduced.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 259316 entries, 622619 to 766243
Data columns (total 16 columns):
 #   Column                    Non-Null Count   Dtype
---  ------                    --------------   -----
 0   DATE                      259316 non-null  object
 1   HourlyDewPointTemperature  222222 non-null  object
 2   HourlyDryBulbTemperature   222399 non-null  object
 3   HourlyPrecipitation        192715 non-null  object
 4   HourlyRelativeHumidity     222218 non-null  object
 5   HourlySeaLevelPressure     201689 non-null  object
 6   HourlyStationPressure      201397 non-null  object
 7   HourlyVisibility           229756 non-null  object
 8   HourlyWetBulbTemperature   198881 non-null  object
 9   HourlyWindDirection        222918 non-null  object
 10  HourlyWindSpeed            229367 non-null  object
 11  YEAR                      259316 non-null  int64
 12  MONTH                     259316 non-null  int64
 13  DAY                       259316 non-null  int64
 14  HOUR                      259316 non-null  int64
 15  MINUTE                    259316 non-null  int64
dtypes: int64(5), object(11)
memory usage: 33.6+ MB
```

Final, Merged, and Filtered Yearly Data

```
# Imports
import pandas as pd
import numpy as np
from numpy.polynomial.polynomial import polyfit
import os
import seaborn as sns
from matplotlib import pyplot as plt
from sklearn.preprocessing import Normalizer
```

In [26]:

```
# Load data to work with
os.popen('cd /Users/ericbu/HMThesis/data ls').read()
path = r'/Users/ericbu/HMThesis/parson_july_2022/year_aggregate_HMS_deploy.csv'
df_full = pd.read_csv(path)
```

In [32]:

```
#df_full.reset_index()
df_full.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74 entries, 0 to 73
Columns: 799 entries, year to TOTAL_peak
dtypes: float64(654), int64(145)
memory usage: 462.0 KB
```

In [28]:

```
attributes_of_interest = ['year','duration','HMtempC_mean','Visibility_mean','CloudCover_mean','SkyCode_mean','F
lightHT_mean','WindSpd_mean','HMtempC_median','Visibility_median','CloudCover_median','SkyCode_median','
FlightHT_median','WindSpd_median','HMtempC_pstdv','Visibility_pstdv','CloudCover_pstdv','SkyCode_pstdv','Fli
ghtHT_pstdv','WindSpd_pstdv','HMtempC_min','Visibility_min','CloudCover_min','SkyCode_min','FlightHT_min','
WindSpd_min','HMtempC_max','Visibility_max','CloudCover_max','SkyCode_max','FlightHT_max','WindSpd_ma
x','WindDegrees','wndN','wndNNE','wndNE','wndENE','wndE','wndESE','wndSE','wndSSE','wndS','wndSSW','wndS
W','wndWSW','wndW','wndWNW','wndNW','wndNNW','wndUNK','fltN','fltNNE','fltNE','fltENE','fltE','fltESE','fltS
E','fltSSE','fltS','fltSSW','fltSW','fltWSW','fltW','fltWNW','fltNW','fltNNW','fltUNK','HourlyDryBulbTemperature_
mean','HourlyWetBulbTemperature_mean','HourlyDewPointTemperature_mean','HourlyWindSpeed_mean','Hourly
Precipitation_mean','HourlyStationPressure_mean','HourlyRelativeHumidity_mean','HourlyVisibility_mean','Hourly
DryBulbTemperature_median','HourlyWetBulbTemperature_median','HourlyDewPointTemperature_median','Hourl
yWindSpeed_median','HourlyPrecipitation_median','HourlyStationPressure_median','HourlyRelativeHumidity_med
ian','HourlyVisibility_median','HourlyDryBulbTemperature_pstdv','HourlyWetBulbTemperature_pstdv','HourlyDew
PointTemperature_pstdv','HourlyWindSpeed_pstdv','HourlyPrecipitation_pstdv','HourlyStationPressure_pstdv','Hou
rlyRelativeHumidity_pstdv','HourlyVisibility_pstdv','HourlyDryBulbTemperature_min','HourlyWetBulbTemperatur
e_min','HourlyDewPointTemperature_min','HourlyWindSpeed_min','HourlyPrecipitation_min','HourlyStationPressu
re_min','HourlyRelativeHumidity_min','HourlyVisibility_min','HourlyDryBulbTemperature_max','HourlyWetBulbT
emperature_max','HourlyDewPointTemperature_max','HourlyWindSpeed_max','HourlyPrecipitation_max','HourlyS
tationPressure_max','HourlyRelativeHumidity_max','HourlyVisibility_max','HourlyWindDirection','HMtempC_24_
mean','Visibility_24_mean','CloudCover_24_mean','HourlyDryBulbTemperature_24_mean','HourlyWetBulbTemper
ature_24_mean','HourlyDewPointTemperature_24_mean','HourlyStationPressure_24_mean','HourlyRelativeHumidi
ty_24_mean','HourlyVisibility_24_mean','HMtempC_48_mean','Visibility_48_mean','CloudCover_48_mean','Hourl
yDryBulbTemperature_48_mean','HourlyWetBulbTemperature_48_mean','HourlyDewPointTemperature_48_mean',
'HourlyStationPressure_48_mean','HourlyRelativeHumidity_48_mean','HourlyVisibility_48_mean','HMtempC_72_
mean','Visibility_72_mean','CloudCover_72_mean','HourlyDryBulbTemperature_72_mean','HourlyWetBulbTemper
ature_72_mean','HourlyDewPointTemperature_72_mean','HourlyStationPressure_72_mean','HourlyRelativeHumidi
ty_72_mean','HourlyVisibility_72_mean','HMtempC_24_median','Visibility_24_median','CloudCover_24_median','
HourlyDryBulbTemperature_24_median','HourlyWetBulbTemperature_24_median','HourlyDewPointTemperature_
24_median','HourlyStationPressure_24_median','HourlyRelativeHumidity_24_median','HourlyVisibility_24_median
','HMtempC_48_median','Visibility_48_median','CloudCover_48_median','HourlyDryBulbTemperature_48_median'
,'HourlyWetBulbTemperature_48_median','HourlyDewPointTemperature_48_median','HourlyStationPressure_48_
median','HourlyRelativeHumidity_48_median','HourlyVisibility_48_median','HMtempC_72_median','Visibility_72
_median','CloudCover_72_median','HourlyDryBulbTemperature_72_median','HourlyWetBulbTemperature_72_med
ian','HourlyDewPointTemperature_72_median','HourlyStationPressure_72_median','HourlyRelativeHumidity_72_m
edian','HourlyVisibility_72_median','HMtempC_24_pstdv','Visibility_24_pstdv','CloudCover_24_pstdv','HourlyDry
BulbTemperature_24_pstdv','HourlyWetBulbTemperature_24_pstdv','HourlyDewPointTemperature_24_pstdv','Hou
```

```
rlyStationPressure_24_pstdv','HourlyRelativeHumidity_24_pstdv','HourlyVisibility_24_pstdv','HMtempC_48_pstdv','Visibility_48_pstdv','CloudCover_48_pstdv','HourlyDryBulbTemperature_48_pstdv','HourlyWetBulbTemperature_48_pstdv','HourlyDewPointTemperature_48_pstdv','HourlyStationPressure_48_pstdv','HourlyRelativeHumidity_48_pstdv','HourlyVisibility_48_pstdv','HMtempC_72_pstdv','Visibility_72_pstdv','CloudCover_72_pstdv','HourlyDryBulbTemperature_72_pstdv','HourlyWetBulbTemperature_72_pstdv','HourlyDewPointTemperature_72_pstdv','HourlyStationPressure_72_pstdv','HourlyRelativeHumidity_72_pstdv','HourlyVisibility_72_pstdv','HMtempC_24_min','Visibility_24_min','CloudCover_24_min','HourlyDryBulbTemperature_24_min','HourlyWetBulbTemperature_24_min','HourlyDewPointTemperature_24_min','HourlyStationPressure_24_min','HourlyRelativeHumidity_24_min','HourlyVisibility_24_min','HMtempC_48_min','Visibility_48_min','CloudCover_48_min','HourlyDryBulbTemperature_48_min','HourlyWetBulbTemperature_48_min','HourlyDewPointTemperature_48_min','HourlyStationPressure_48_min','HourlyRelativeHumidity_48_min','HourlyVisibility_48_min','HMtempC_72_min','Visibility_72_min','CloudCover_72_min','HourlyDryBulbTemperature_72_min','HourlyWetBulbTemperature_72_min','HourlyDewPointTemperature_72_min','HourlyStationPressure_72_min','HourlyRelativeHumidity_72_min','HourlyVisibility_72_min','HMtempC_24_max','Visibility_24_max','CloudCover_24_max','HourlyDryBulbTemperature_24_max','HourlyWetBulbTemperature_24_max','HourlyDewPointTemperature_24_max','HourlyStationPressure_24_max','HourlyRelativeHumidity_24_max','HourlyVisibility_24_max','HMtempC_48_max','Visibility_48_max','CloudCover_48_max','HourlyDryBulbTemperature_48_max','HourlyWetBulbTemperature_48_max','HourlyDewPointTemperature_48_max','HourlyStationPressure_48_max','HourlyRelativeHumidity_48_max','HourlyVisibility_48_max','HMtempC_72_max','Visibility_72_max','CloudCover_72_max','HourlyDryBulbTemperature_72_max','HourlyWetBulbTemperature_72_max','HourlyDewPointTemperature_72_max','HourlyStationPressure_72_max','HourlyRelativeHumidity_72_max','HourlyVisibility_72_max','RecordCount','HMtempCCount','TOTAL','TOTAL_1st','TOTAL_25th','TOTAL_50th','TOTAL_75th','TOTAL_last','TOTAL_peak']
```

In [29]:

```
df_reduced = df_full[attributes_of_interest].copy()
```

In [30]:

```
df_reduced = df_reduced[(df_reduced['year'] > 1978)]
```

In [31]:

```
df_reduced.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 43 entries, 31 to 73
Columns: 252 entries, year to TOTAL_peak
dtypes: float64(209), int64(43)
memory usage: 85.0 KB
```

Hawk Mountain Pearson Correlation Coefficient

```
hm_nontarget_scalar_mean = ['year','WindSpd_mean','HMtempC_mean','CloudCover_mean','SkyCode_mean','FlightHT_mean','Visibility_mean','TOTAL_mean']
df_reduced_hm_scalar_zoom_mean = df_reduced[hm_nontarget_scalar_mean].copy()
yearly_df_standardize = df_reduced_hm_scalar_zoom_mean.replace(np.nan, "0")
yearly_df_normalize = Normalizer().fit_transform(yearly_df_standardize)
pearson_correlation_df = pd.DataFrame(yearly_df_normalize).corr(method='pearson')
```

NOAA Pearson Correlation Coefficient

```
noaa_attributes_scalar_zoom_mean = ['year','HourlyDryBulbTemperature_mean','HourlyWetBulbTemperature_mean','HourlyDewPointTemperature_mean','HourlyWindSpeed_mean','HourlyPrecipitation_mean','HourlyStationPressure_mean','HourlyRelativeHumidity_mean','HourlyVisibility_mean','TOTAL_mean']
noaa_attributes_scalar_zoom_mean = df_reduced[noaa_attributes_scalar_zoom_mean].copy()
yearly_df_standardize_noaa = noaa_attributes_scalar_zoom_mean.replace(np.nan, "0")
yearly_df_normalize_noaa = Normalizer().fit_transform(yearly_df_standardize_noaa)
pearson_correlation_df = pd.DataFrame(yearly_df_normalize_noaa).corr(method='pearson')
```

Hawk Mountain Wind Direction Pearson Correlation Coefficient

```
hm_wind_scalar = ['year','wndN','wndNNE','wndNE','wndENE','wndE','wndESE','wndSE','wndSSE','wndS','wndSSW','wndSW','wndWSW','wndW','wndWNW','wndNW','wndNNW','wndUNK','TOTAL_mean']
hm_wind_scalar = df_reduced[hm_wind_scalar].copy()
yearly_df_standardize_hm_winddir = hm_wind_scalar.replace(np.nan, "0")
yearly_df_normalize_hm_winddir = Normalizer().fit_transform(yearly_df_standardize_hm_winddir)
pearson_correlation_df = pd.DataFrame(yearly_df_normalize_hm_winddir).corr(method='pearson')
```

Appendix B

Legend for Hawk Mountain Data

Data from Hawk Mountain that needs to be defined for comprehension

Format

year

Numeric attribute indicating year.

temp

Numeric attribute for temperature. In Celsius.

HMTempC

Numeric attribute from Hawk Mountain's records. In Celsius.

WindSpd

Numeric attribute from Hawk Mountain's records. See below WIND SPEEDS legend.

CloudCover

Numeric attribute from Hawk Mountain's records. Ranging from 0-100, 0 being no clouds, 100 meaning complete cloud cover.

FlightHT

Numeric attribute from Hawk Mountain's records. See below ALTITUDE OF FLIGHT legend.

Visibility

Numeric attribute from Hawk Mountain's records. Ranging from 0 – 100. 0 being low visibility and 100 being high visibility.

TOTAL_RAPTORS

Numeric attribute from Hawk Mountain's records. Total number of hawks counted.

TOTAL_mean

Numeric attribute derived from TOTAL_RAPTORS. It is the average of the TOTAL_RAPTORS attribute.

wnd<wind direction> (i.e wndSE is winds blowing from the Southeast.)

Numeric attribute derived from Hawk Mountain's original dataset. 1 tally is equal to 1 day with wind direction blowing from a certain direction.

Note: For the attributes above (except for TOTAL_mean), there will be an appended suffix showing the minimum, mean, median, maximum, and population standard deviation of each attribute for the specified year. As mentioned in section 3.3, the yearly data is aggregated for the months of August to December as that is when the Hawk Mountain Autumn counting season is. The suffixes are: _min, _mean, _median, _max, _pstdv. These values are found throughout this document and are mainly consolidated in sections 4.1 and 4.2.

Suffix definitions:

_min: The absolute minimum value found for that year.

_mean: The arithmetic mean/average value found for that year.

_median: The simple median value found for that year.

_max: The absolute maximum value found for that year.

_pstdv: The population standard deviation value found for that year. We are using the population standard deviation here. It can be argued that this is a sample standard deviation since we are sampling the yearly data and only taking the autumn months, but we are calculating the standard deviation of that full sample and not sampling further.

Legend for a few Hawk Mountain attributes.

WIND SPEEDS	
0	Less than 1 km/h; calm; smoke rises vertically
1	1-5 km/h; smoke drift shows wind direction
2	6-11 km/h; (4-7 m/h); leaves rustle, wind is felt on face
3	12-19 km/h; (8-12m/h); leaves, small twigs in constant motion; light flag extended
4	20-28 km/h; (13-18 m/h); raises dust, leaves, loose paper; small branches in motion
5	(19-24 m/h); small trees in leaf sway
6	39-49 km/h (25-31 m/h); larger branches in motion; whistling hear in wires
7	50-61 km/h (32-38 m/h); whole trees in motion; resistance felt walking against the wind
8	62-74 km/h (39-46 m/h); twigs, small branches broken off trees; walking generally impeded
9	Greater than 75 km/h (47 m/h)

ALTITUDE OF FLIGHT

0 Below eye level
1 Eye level up t about 30 meters (100 feet) overhead
2 Birds easily seen with unaided eye (eyeglasses no counted as aids)
3 At limit of unaided vision
4 Beyond limit of unaided eye but visible with binocuars – to 10X
5 At limit of bionoculars
6 Beyond limit of binoculars 10 X or less, but can detect with binoculars or telescope of greater power (Mark "1" in COMMENT box and not magnification)
7 No predominant height

SKY CODES

0 Clear; 0-15% cloud cover
1 Partly cloudy; 16-50% cover
2 Mostly cloudy; 51-75% cover
3 Overcast; 76-100% cover
4 Wind-driven sand, dust, snow
5 Fog or haze
6 Drizzle
7 Rain
8 Snow
9 Thunderstorm, with or without precipitation

<u>**Appendix C**</u>

Legend for NOAA data

NOAA data that needs to be defined for comprehension

Format

year

Numeric attribute from NOAA indicating year.

(Hourly)DrybulbTemperature

Numeric attribute from NOAA indicating the dry bulb temperature. Dry bulb temperature is the ambient temperature not affected by moisture in the air. In Celsius.

(Hourly)WetBulbTemperature

Numeric attribute from NOA indicating the web bulb temperature. Wet bulb temperature is a temperature measurement that accounts for the humidity in the air. In Celsius.

(Hourly)DewPointTemperature

Numeric attribute from NOAA indicating the dew point temperature. The dew point temperature is the temperature to which air must be cooled to become saturated with water vapor. Celsius.

(Hourly)WindSpeed

Numeric attribute from NOAA indicating the windspeed. In km/hr.

(Hourly)Precipitation

Numeric attribute from NOAA indicating the precipitation. Inches of rain per hour.

(Hourly)StationPressure

Numeric attribute from NOAA indicating barometric pressure. Barometric pressure at the Allentown NOAA station.

(Hourly)RelativeHumidity

Numeric attribute from NOAA indicating relative humidity. Relative humidity is a percentage that represents the amount of water vapor in the air at a given temperature compared to the max possible water vapor amount at that same temperature.

(Hourly)Visibility

Numeric attribute from NOAA. Ranging from 0-30. 0 being low and 30 being high.

WindDirection

Numeric attribute from NOAA. 0-360 degrees.