Climate Change:

Have Climate Variables Changed Over Time?

**By: Brianna Noel**

Contents

[Abstract: 3](#_Toc184987509)

[Introduction: 4](#_Toc184987510)

[Problem Definition 4](#_Toc184987511)

[Objective 4](#_Toc184987512)

[Importance 4](#_Toc184987513)

[Data Selection and Preparation: 5](#_Toc184987514)

[Data Overview 5](#_Toc184987515)

[Variable Selection 5](#_Toc184987516)

[Data Cleaning 6](#_Toc184987517)

[Data Analysis: 8](#_Toc184987518)

[Exploratory Data Analysis 8](#_Toc184987519)

[Statistical Methods 12](#_Toc184987520)

[ANOVA and Interpretation of Results (Decision and Conclusion) 12](#_Toc184987521)

[T-Tests and Interpretation of Results (Decision and Conclusion) 17](#_Toc184987522)

[Simple Linear Regression Analysis and Interpretation of Results (Decision and Conclusion) 23](#_Toc184987523)

[Further Interpretation of Results 24](#_Toc184987524)

[Decision Overview 24](#_Toc184987525)

[Implications 24](#_Toc184987526)

[Limitations 26](#_Toc184987527)

[Conclusion: 27](#_Toc184987528)

[References: 28](#_Toc184987529)

[Appendix: 29](#_Toc184987530)

# Abstract:

According to NOAA climate.gov, “Earth’s temperature has risen by an average of 0.11° Fahrenheit (0.06° Celsius) per decade since 1850, or about 2° F in total.” This raises questions, such as if climate is rising over decades on a global scale, can we observe changes in temperature trends from local data that also reflect this trend? The purpose of this report is to analyze six datasets consisting of daily weather observations from six different years (1963, 1975, 1987, 1999, 2011, and 2023) from a randomly selected airport (Los Angeles International Airport, CA, US). These weather factors, such as mean temperature, wind speed, and dew point interact with categorical variables like year, season, and visibility range. Exploratory data analysis, ANOVA, t-tests, and simple regression will be used to analyst and interpret the relationships among the datasets and explore trends across the years, seasons, and other weather factors.

# Introduction:

## Problem Definition

Having a deeper understanding of the relationships between weather variables like average temperature, wind speed, and seasonal impacts can provide meaningful insights into global climate trends. Exploring how temperature interacts with other weather-related variables, such as wind speed, dew point, and sea level pressure ranges is crucial for environmental research on climate change. The question of how closely local climate data reflects global climate trends is fundamental for further exploration of weather patterns.

## Objective

The purpose of this report is to investigate if climate variables like daily mean temperature and wind speed have changed over the past six decades (1963 to 2023) and across different factors, such as seasons and visibility ranges. The goal of this analysis is also to explore the relationships among these weather-related variables and to investigate how they interact with daily average temperatures.

## Importance

Through the exploration of these relationships, the aim of this analysis is to provide valuable insights that could be used in meteorological forecasting, climate changes studies, and other environmental research. By understanding how these weather-related factors interact with each other and change over variations like time from a local standpoint, we can add value to the understanding of the complex nature of climate change on a global scale.

# Data Selection and Preparation:

## Data Overview

The source of the datasets is the NCEI (National Centers Environmental Information)- NOAA (National Oceanic and Atmospheric Administration)- Global Surface Summary of the Day. This analysis used data from the Los Angeles International Airport, CA, US. This analysis consists of six datasets, each containing the daily weather data from the specified year, including 365 days per year. The six different years are as follows: 1963, 1975, 1987, 1999, 2011, 2023.

## Variable Selection

Variables:

1. **Date ID:** The date used as an identifier variable
2. **Mean Temperature:** Mean temperature for the day in degrees Fahrenheit to tenths
3. **Mean Dew Point:** Mean dew point for the day in degrees Fahrenheit to tenths
4. **Mean Wind Speed:** Mean wind speed for the day in knots to tenths
5. **Max Temperature:** Maximum temperature reported for the day in degrees Fahrenheit to tenths
6. **Year:** The year.
   1. 1963
   2. 1975
   3. 1987
   4. 1999
   5. 2011
   6. 2023
7. **Season:** The season:
   1. Winter- December, January, February
   2. Spring- March, April, May
   3. Summer- June, July, August
   4. Fall- September, October, November
8. **Mean Sea Level Pressure Range:** Sea level pressure ranges based on mean sea level pressure for the day in millibars to tenths
   1. Low Pressure
   2. Normal Pressure
   3. High Pressure
9. **Mean Visibility Range:**  Visibility ranges based on mean visibility for the day in miles to tenths
   1. Low Visibility
   2. Moderate Visibility
   3. High Visibility

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Description Table** | | | |
| **Variable Title** | **Variable Name** | **Measurement Units** | **General Type** |
| DATE\_ID | Date ID | N/A | Identifier |
| TEMP | Mean Temperature | Fahrenheit | Quantitative |
| DEWP | Mean Dew Point | Fahrenheit | Quantitative |
| WDSP | Mean Wind Speed | Knots | Quantitative |
| MAX | Max Temperature | Fahrenheit | Quantitative |
| YEAR | Year | N/A | Categorical |
| SEASON | Season | N/A | Categorical |
| SLP\_CAT | Mean Sea Level Pressure Range | N/A (millibars) | Categorical |
| VISIB\_CAT | Mean Visibility Range | N/A (miles) | Categorical |

## Data Cleaning

After each CSV file corresponding to each of the six years was imported into SAS and combined into one dataset (with a new variable year created to differentiate each year), several variables were removed, and others were created.

1. The unnecessary variables that included repeating information were removed first. These variables, such as station, latitude, longitude, elevation, and name, all contained the same data for each observation. This is because each observation is from the Los Angeles International Airport, CA, US, which has the same information for all these variables across all observations. Hence, these variables were removed.
2. Next, the indicator/ attribute variables were removed, as they are not necessary for this analysis. Similarly to the previous step, these variables also contained redundant information, so they were removed, as well.
3. Frequency tables were then created to observe the remaining variables. Any unnecessary variables identified from the frequency tables were removed.
4. Among the remaining variables, there was no missing data, so further data cleaning was unnecessary at this step.
5. Then, new categorical variables were created. A variable indicating the month of each observation was created to determine the season, a new categorical variable created for the analysis. The month variable used to create the season variable was removed afterwards, as well. The season variable remained in the dataset.
6. Descriptive statistics on the mean sea level pressure and mean visibility quantitative variables were conducted to explore the possible ranges for these variables to be converted into categorical variables. Hence, the mean sea level pressure range and the mean visibility range variables were created.
7. Lastly, the date variable name was changed to date ID, as it serves the purpose of an identifier variable in this analysis. The order of the variables was modified to have the ID listed to the right, followed by the quantitative and then the categorical variables.

# Data Analysis:

## Exploratory Data Analysis

Descriptive Statistics

A screenshot of a statistics report

Description automatically generated

Histograms

A graph of a distribution of a function

Description automatically generatedA graph with a blue line

Description automatically generatedA graph with a blue line

Description automatically generatedA graph of a distribution of a number

Description automatically generated

*"Top Left: Mean Temperature | Top Right: Mean Dew Point | Bottom Left: Mean Wind Speed 3 | Bottom Right: Maximum Temperature"*

It is noted that the distribution of mean dew point is skewed left, and the distribution of mean wind speed is slightly skewed right. The distribution of mean temperature and max temperature are more normally distributed.

Frequency Tables

A screenshot of a screenshot of a computer

Description automatically generated

Side-By-Side Box Plots

A diagram of a graph

Description automatically generated with medium confidence

A graph of a graph showing different types of data

Description automatically generated with medium confidence

Correlation Matrix

A table with numbers and letters

Description automatically generated

It is noted that the highest correlation coefficient between two variables is 0.59408 between mean temperature and mean dew point (other than max temperature and mean temperature, which was not wished to be explored in this analysis).

Scatterplot

A screen shot of a graph

Description automatically generated

## Statistical Methods

### ANOVA and Interpretation of Results (Decision and Conclusion)

ANOVA for Mean Temperatures Across Years

A screenshot of a calculator

Description automatically generated

Since the p-value (0.0002) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean temperatures across the six years are the same. This means that there is a statistically significant difference between the mean temperatures across the six years.

A graph with numbers and a red line

Description automatically generated

Dynamic Visualization of Temperature Changes Across Years

ANOVA for Mean Summer Temperatures Across Years

A screenshot of a computer screen

Description automatically generated

Since the p-value (<0.0001) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean summer temperatures across the six years are the same. This means that there is a statistically significant difference between the mean summer temperatures across the six years.

ANOVA for Mean Winter Temperatures Across Years

A screenshot of a data

Description automatically generated

Since the p-value (0.0010) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean winter temperatures across the six years are the same. This means that there is a statistically significant difference between the mean winter temperatures across the six years.

ANOVA for Mean Temperatures Across Mean Sea Level Pressure Ranges

A screenshot of a data

Description automatically generated

Since the p-value (<0.0001) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean temperatures across the mean sea level pressure ranges are the same. This means that there is a statistically significant difference between the mean temperatures across the mean sea level pressure ranges.

ANOVA for Mean Temperatures Across Mean Visibility Ranges

A screenshot of a data

Description automatically generated

Since the p-value (<0.0001) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean temperatures across mean visibility are the same. This means that there is a statistically significant difference between the mean temperatures across the mean visibility ranges.

ANOVA for Mean Wind Speeds Across Years

A screenshot of a computer screen

Description automatically generated

Since the p-value (<0.0001) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean wind speeds across years are the same. This means that there is a statistically significant difference between the mean wind speeds across the years.

### T-Tests and Interpretation of Results (Decision and Conclusion)

T-Tests for Mean Wind Speed for 1963 vs. 2023

A screenshot of a test

Description automatically generated

Since the p-value (0.0424) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean wind speeds in 1963 and 2023 are the same. This means that there is a statistically significant difference between the mean wind speeds in 1963 vs. 2023.

T-Tests for Mean Temperature for 1963 vs. 2023

A screenshot of a test

Description automatically generated

Since the p-value (0.1196) > 0.05, the null hypothesis is failed to be rejected. There is not sufficient evidence to conclude that the mean temperatures in 1963 and 2023 are statistically significantly different.

T-Tests for Mean Temperature for 1975 vs. 2023

A screenshot of a test

Description automatically generated

Since the p-value (0.0032) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean temperature in 1975 and 2023 are the same. This means that there is a statistically significant difference between the mean temperature in 1975 vs. 2023.

T-Tests for Max Temperature for 1975 vs. 2023

A screenshot of a test results

Description automatically generated

Since the p-value (0.7057) > 0.05, the null hypothesis is failed to be rejected. There is not sufficient evidence to conclude that the max temperature in 1975 and 2023 are statistically significantly different.

T-Tests for Mean Summer Temperature for 1975 vs. 2023

A screenshot of a test

Description automatically generated

Since the p-value (0.0826) > 0.05, the null hypothesis is failed to be rejected. There is not sufficient evidence to conclude that the max temperature in 1975 and 2023 are statistically significantly different.

T-Tests for Mean Winter Temperature for 1975 vs. 2023

A screenshot of a test

Description automatically generated

Since the p-value (0.0242) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the mean winter temperature in 1975 and 2023 are the same. This means that there is a statistically significant difference between the mean winter temperature in 1975 vs. 2023.

### Simple Linear Regression Analysis and Interpretation of Results (Decision and Conclusion)

Simple Linear Regression Mean Dew Point vs. Mean Temperature

A screenshot of a computer

Description automatically generated

Since the p-value (<0.0001) < 0.05, the null hypothesis is rejected. There is not sufficient evidence to conclude that the slope of the line is 0. In other words, there is not sufficient evidence to conclude that there is not a linear relationship between mean dew point and mean temperature.

The following estimated simple linear regression equation can be concluded from the output:

## Further Interpretation of Results

### Decision Overview

|  |  |
| --- | --- |
| **Statistical Method Test Decisions** | |
| **Test** | **Decision** |
| ANOVA for Mean Temperatures Across Years | Reject the null |
| ANOVA for Mean Summer Temperatures Across Years | Reject the null |
| ANOVA for Mean Winter Temperatures Across Years | Reject the null |
| ANOVA for Mean Temperatures Across Mean Sea Level Pressure Ranges | Reject the null |
| ANOVA for Mean Temperatures Across Mean Visibility Ranges | Reject the null |
| ANOVA for Mean Wind Speeds Across Years | Reject the null |
| T-Tests for Mean Wind Speed for 1963 vs. 2023 | Reject the null |
| T-Tests for Mean Temperature for 1963 vs. 2023 | Fail to reject the null |
| T-Tests for Mean Temperature for 1975 vs. 2023 | Reject the null |
| T-Tests for Max Temperature for 1975 vs. 2023 | Fail to reject the null |
| T-Tests for Mean Summer Temperature for 1975 vs. 2023 | Fail to reject the null |
| T-Tests for Mean Winter Temperature for 1975 vs. 2023 | Reject the null |

### Implications

There are many practical implications of the findings in this analysis. From the ANOVA tests, we made the decision to reject the null hypothesis for all six tests conducted. This means it was concluded that…

1. There was not sufficient evidence to conclude that the mean temperatures across the years were statistically the same
2. There was not sufficient evidence to conclude that the mean summer temperatures across the years were statistically the same
3. There was not sufficient evidence to conclude that the mean winter temperatures across the years were statistically the same
4. There was not sufficient evidence to conclude that the mean temperatures across mean sea level pressure ranges were statistically the same
5. There was not sufficient evidence to conclude that the mean temperatures across mean visibility ranges were statistically the same
6. There was not sufficient evidence to conclude that the mean wind speeds across the years were statistically the same

This information gave further insight into how these specific weather-related variables have statistical differences across various factors. For example, it highlighted how the mean temperatures across the years are statistically not the same, providing more potential insight into how local climate data may reflect global climate data. Additionally, it potentially provides insight into exploring how these other weather-related variables, such as mean wind speeds, sea level pressures, and visibility, are changing over time on a local scale. This allowed for further t-tests in this study to be conducted to explore where the statistical differences occurred across these variables.

The t-Tests showed that…

1. There was not sufficient evidence to conclude that the mean wind speeds in 1963 and 2023 are statistically the same
2. There was not sufficient evidence to conclude that the mean temperatures in 1963 and 2023 are statistically different
3. There was not sufficient evidence to conclude that the mean temperatures in 1975 and 2023 are statistically the same
4. There was not sufficient evidence to conclude that the max temperatures in 1975 and 2023 are statistically different
5. There was not sufficient evidence to conclude that the mean summer temperatures in 1975 and 2023 are statistically different
6. There was not sufficient evidence to conclude that the mean winter temperatures in 1975 and 2023 are statistically the same

From the t-Tests, the first t-Test adds value that weather-related variables other than mean temperature, such as wind speeds, can be explored, as it can be seen how they are statistically different from 1963 vs. 2023. From the second t-Test, a statistical difference between the mean temperatures in 1963 vs. 2023 cannot be found, but from the third t-Test, a statistical difference between the mean temperatures in 1975 vs. 2023 can be found. Even though the reason for this remains unconclusive, this may still suggest there was a change in average climate across decades on this local scale (between 1975 and 2023). From the fourth t-Test, there was not a statistical difference between the maximum temperatures in 1975 vs. 2023, even though there was a statistical difference between the mean temperatures in those same years from the third t-test. This may suggest further research into the differences in changes among mean temperatures over time versus maximum temperatures over time, which may support future environmental research. From the fifth t-Test, even though the mean summer temperatures in 1975 and 2023 were not found to be statistically different, the mean winter temperatures in 1975 and 2023 were found to be statistically different. This may suggest further research into the differences in changes among mean temperatures over time amongst different seasons, which may aid in future climate change research, as well.

For the simple linear regression, the two variables with the highest correlation coefficients were selected for the estimated simple linear regression equation. The variables average dew point and temperature have a relatively moderate positive linear relationship with a correlation coefficient of 0.59408. The coefficient of determination (r-squared), 0.3529, indicates that 35.29% of the variation in temperature can be explained by average dew point. This analysis suggests that there is a linear relationship between mean dew point and mean temperature. This provides insight into how mean dew point, and other weather-related factors, can potentially be used to predict mean temperature (may give insight into potential future multiple linear regression analysis). This could add value into how environmental researchers can attempt to predict temperature based on other weather-related variables, which could give meaningful insights for global climate change research.

### Limitations

There are many limitations to this analysis, and some of these limitations are listed as follows:

1. **Limited Years:** The years used in this analysis are very limited. These specific years used (1963, 1975, 1987, 1999, 2011, and 2023) are a very broad selection of years that can be studied for changes in temperature and other weather-related variables,
2. **Limited Variables:** The variables used in this analysis are also very limited. More weather-related variables could have added more insights into studying climate change trends on a local scale.
3. **Limited Location:** The location used to reflect climate change on a local scale is only limited to one specific airport (only LA International Airport, CA, US). More locations, specifically more in other regions of the country and/or world, could add more valuable insights into investigating climate change trends.

# Conclusion:

The findings of this analysis provide valuable insights into the statistical relationships between several weather-related variables like mean temperature, wind speed, and dew point, over many decades at the Los Angeles International Airport, CA, US. Through the application of statistical methods including ANOVA, t-Tests, and simple linear regression analysis, notable trends across the years, seasons, and other weather-related factors were observed.

The ANOVA tests showed that there are statistically significant differences in the mean temperatures across the years, and all the null hypotheses for all conducted ANOVA tests were rejected. This implies that there are statistical differences in temperature and other weather-related factors like mean sea level pressure ranges, visibility ranges, and wind speed across the selected years (1963, 1975, 1987, 1999, 2011, and 2023). These findings suggest that local climate data, such as mean temperatures and wind speed, at not static and may reflect wider trends in global climate change.

The t-Tests revealed statistical changes in mean temperatures over time, specifically in 1975 vs 2023. They also revealed changes in wind speeds over time like 1963 vs. 2023. Moreover, among the same years 1975 vs. 2023 where mean temperatures across all seasons over time were statistically varied, further t-Tests among seasons revealed more results. As some differences in summer mean temperatures and winter mean temperatures were found, this suggest that there is a need for future research to understand the underlying causes in these variations among changes in mean temperatures among different seasons over time.

The simple linear regression analysis among mean dew point and temperature showcased a moderate positive correlation, with about 35.29% of the variation in temperature explained by variations in the mean dew point. These results imply that mean dew point may be a meaningful predictor of temperature, having potential value for future research in climate modeling and forecasting.

However, several limitations were present, such as the selection of only six years, the limited range of weather-related variables, and the restriction of only one location. Hence, it would be beneficial to broaden further analysis to include more years, additional variables, and several locations for more comprehensive insights.

# References:

Dahlman, R. L. A. L. (2024, January 18). Climate change: Global temperature. NOAA Climate.gov. https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

Data Access. National Centers for Environmental Information (NCEI). (n.d.). https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-day

# Appendix:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Author: Bri Noel \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Create new library \*/

libname report "C:\Users\Brino\OneDrive\Desktop\Fall 2024\Statistical Methods\Week 14\Report";

/\* Import CSV files for each year \*/

**proc** **import** out = report.climate\_1963

datafile ="C:\Users\Brino\OneDrive\Desktop\Fall 2024\Statistical Methods\Week 14\Report\LA\_climate\_1963.csv"

dbms=csv replace;

getnames=yes;

**run**;

**proc** **import** out = report.climate\_1975

datafile ="C:\Users\Brino\OneDrive\Desktop\Fall 2024\Statistical Methods\Week 14\Report\LA\_climate\_1975.csv"

dbms=csv replace;

getnames=yes;

**run**;

**proc** **import** out = report.climate\_1987

datafile ="C:\Users\Brino\OneDrive\Desktop\Fall 2024\Statistical Methods\Week 14\Report\LA\_climate\_1987.csv"

dbms=csv replace;

getnames=yes;

**run**;

**proc** **import** out = report.climate\_1999

datafile ="C:\Users\Brino\OneDrive\Desktop\Fall 2024\Statistical Methods\Week 14\Report\LA\_climate\_1999.csv"

dbms=csv replace;

getnames=yes;

**run**;

**proc** **import** out = report.climate\_2011

datafile ="C:\Users\Brino\OneDrive\Desktop\Fall 2024\Statistical Methods\Week 14\Report\LA\_climate\_2011.csv"

dbms=csv replace;

getnames=yes;

**run**;

**proc** **import** out = report.climate\_2023

datafile ="C:\Users\Brino\OneDrive\Desktop\Fall 2024\Statistical Methods\Week 14\Report\LA\_climate\_2023.csv"

dbms=csv replace;

getnames=yes;

**run**;

/\*\*\*\*\* Data Cleaning and New Variable Creation \*\*\*\*\*/

/\* Combine all six datasets into one dataset \*/

**data** report.climate\_data;

set report.climate\_1963 (in=a)

report.climate\_1975 (in=b)

report.climate\_1987 (in=c)

report.climate\_1999 (in=d)

report.climate\_2011 (in=e)

report.climate\_2023 (in=f);

/\* Create a YEAR variable to identify the data \*/

if a then YEAR = "1963";

else if b then YEAR = "1975";

else if c then YEAR = "1987";

else if d then YEAR = "1999";

else if e then YEAR = "2011";

else if f then YEAR = "2023";

**run**;

/\* Remove unneccessary variables with repeated information not needed for analysis \*/

**data** report.climate\_data (drop = STATION LATITUDE LONGITUDE ELEVATION NAME);

set report.climate\_data;

**run**;

/\* Remove indicator or \_attribute variables \*/

**data** report.climate\_data (drop = TEMP\_ATTRIBUTES DEWP\_ATTRIBUTES SLP\_ATTRIBUTES STP\_ATTRIBUTES VISIB\_ATTRIBUTES WDSP\_ATTRIBUTES MAX\_ATTRIBUTES MIN\_ATTRIBUTES PRCP\_ATTRIBUTES);

set report.climate\_data;

**run**;

/\* Create frequency tables for remaining variables \*/

**proc** **freq** data = report.climate\_data;

table TEMP;

**run**;

**proc** **freq** data = report.climate\_data;

table DEWP;

**run**;

**proc** **freq** data = report.climate\_data;

table SLP;

**run**;

**proc** **freq** data = report.climate\_data;

table STP;

**run**;

**proc** **freq** data = report.climate\_data;

table VISIB;

**run**;

**proc** **freq** data = report.climate\_data;

table WDSP;

**run**;

**proc** **freq** data = report.climate\_data;

table MXSPD;

**run**;

**proc** **freq** data = report.climate\_data;

table GUST;

**run**;

**proc** **freq** data = report.climate\_data;

table MAX;

**run**;

**proc** **freq** data = report.climate\_data;

table MIN;

**run**;

**proc** **freq** data = report.climate\_data;

table PRCP;

**run**;

**proc** **freq** data = report.climate\_data;

table SNDP;

**run**;

**proc** **freq** data = report.climate\_data;

table FRSHTT;

**run**;

/\* Remove unnecessary variables identified from previous frequency tables \*/

**data** report.climate\_data (drop = STP MXSPD GUST PRCP SNDP FRSHTT);

set report.climate\_data;

**run**;

/\* Create new categorical variable SEASON \*/

**data** report.new\_climate\_data (drop = MONTH\_NUM);

set report.climate\_data;

MONTH\_NUM = month(DATE);

if MONTH\_NUM in (**12**, **1**, **2**) then SEASON = "Winter";

else if MONTH\_NUM in (**3**, **4**, **5**) then SEASON = "Spring";

else if MONTH\_NUM in (**6**, **7**, **8**) then SEASON = "Summer";

else if MONTH\_NUM in (**9**, **10**, **11**) then SEASON = "Fall";

**run**;

/\* Create descriptive stats for two quant variables to make ranges to change to categorical variables \*/

**proc** **means** data = report.climate\_data n mean median stddev range q1 q3;

var SLP VISIB;

**run**;

/\* Create a categorical variable for SLP \*/

**data** report.new\_climate\_data (drop = MIN SLP);

set report.new\_climate\_data;

length SLP\_CAT $**20**;

if SLP < **1013** then SLP\_CAT = "Low Pressure";

else if **1013** <= SLP < **1018** then SLP\_CAT = "Normal Pressure";

else if **1018** <= SLP then SLP\_CAT = "High Pressure";

**run**;

/\* Create a categorical variable for VISIB \*/

**data** report.new\_climate\_data (drop = VISIB);

set report.new\_climate\_data;

length VISIB\_CAT $**20**;

if VISIB < **7.1** then VISIB\_CAT = "Low Visibility";

else if **7.1** <= VISIB < **10.1** then VISIB\_CAT = "Moderate Visibility";

else if **10.1** <= VISIB then VISIB\_CAT = "High Visibility";

**run**;

/\* Create frequency variables for new categorical variables \*/

**proc** **freq** data = report.new\_climate\_data;

table SLP\_CAT VISIB\_CAT;

**run**;

/\* Change DATE variable to a ID variable name \*/

**data** report.new\_climate\_data;

set report.new\_climate\_data;

DATE\_ID = DATE;

format DATE\_ID mmddyy10.

run;

/\* Reorder 1 ID variable, 4 quant variables, 4 categorical variables \*/

**data** report.new\_climate\_data (drop = DATE);

retain DATE\_ID TEMP DEWP WDSP MAX YEAR SEASON SLP\_CAT VISIB\_CAT;

set report.new\_climate\_data;

**run**;

/\*\*\*\*\* EDA \*\*\*\*\*/

/\* Descriptive stats for quant variables \*/

title "Descriptive Statistics";

**proc** **means** data = report.new\_climate\_data n mean median stddev range q1 q3;

var TEMP DEWP WDSP MAX;

**run**;

/\* Histograms for quant variables \*/

title "Histograms";

**proc** **univariate** data=report.new\_climate\_data;

histogram / normal;

var TEMP DEWP WDSP MAX;

**run**;

/\* Frequency tables for categorical variables \*/

title "Frequency Tables";

**proc** **freq** data = report.new\_climate\_data;

tables YEAR SEASON SLP\_CAT VISIB\_CAT;

**run**;

/\* Side-by-side boxplots for categorical variables vs temperature \*/

**proc** **sgplot** data=report.new\_climate\_data;

vbox TEMP / category=SEASON;

title 'Side-by-Side Box Plots of Temperature by Season';

xaxis label='Season';

yaxis label='Temperature (°F)';

**run**;

**proc** **sgplot** data=report.new\_climate\_data;

vbox TEMP / category=YEAR;

title 'Side-by-Side Box Plots of Temperature by Year';

xaxis label='Year';

yaxis label='Temperature (°F)';

**run**;

/\* Correlation matrix for quant variables \*/

title "Correlation Matrix for Quantitative Variables";

**proc** **corr** data = report.new\_climate\_data;

var TEMP DEWP WDSP MAX;

**run**;

/\* Scatter plot for quant variables \*/

title "Scatterplot for Dew Point vs Temperature";

**proc** **corr** data=report.new\_climate\_data plots=scatter(nvar=all);

var DEWP TEMP; \*MAX WDSP;

**run**;

/\*\*\*\*\* Statistical Methods \*\*\*\*\*/

/\* One-Way ANOVA for temperature across years, including post-hoc test \*/

title "One-Way ANOVA for Temperatures Across Years";

**proc** **anova** data=report.new\_climate\_data;

class YEAR;

model TEMP = YEAR;

means YEAR / lsd tukey cldiff;

**run**;

/\* Subset data for summer and perform One-Way ANOVA for summer temperature across years, including post-hoc test \*/

title "One-Way ANOVA for Summer Temperatures Across Years";

**proc** **anova** data=report.new\_climate\_data;

where SEASON = 'Summer';

class YEAR;

model TEMP = YEAR;

means YEAR / lsd tukey cldiff;

**run**;

/\* Subset data for winter and perform One-Way ANOVA for winter temperature across years, including post-hoc test \*/

title "One-Way ANOVA for Winter Temperatures Across Years";

**proc** **anova** data=report.new\_climate\_data;

where SEASON = 'Winter';

class YEAR;

model TEMP = YEAR;

means YEAR / lsd tukey cldiff;

**run**;

/\* One-Way ANOVA for sea level pressure across years, including post-hoc test \*/

title "One-Way ANOVA for Temperatures Across Sea Level Pressure Ranges";

**proc** **anova** data=report.new\_climate\_data;

class SLP\_CAT;

model TEMP = SLP\_CAT;

means SLP\_CAT / lsd tukey cldiff;

**run**;

/\* One-Way ANOVA for visibility across years, including post-hoc test \*/

title "One-Way ANOVA for Temperatures Across Visibilty Ranges";

**proc** **anova** data=report.new\_climate\_data;

class VISIB\_CAT;

model TEMP = VISIB\_CAT;

means VISIB\_CAT / lsd tukey cldiff;

**run**;

/\* One-Way ANOVA for wind speed across years, including post-hoc test \*/

title "One-Way ANOVA for Wind Speeds Across Years";

**proc** **anova** data=report.new\_climate\_data;

class YEAR;

model WDSP = YEAR;

means YEAR / lsd tukey cldiff;

**run**;

/\* T-test for wind speed between 1963 and 2023 \*/

title "t-Test for Wind Speeds between 1963 and 2023";

**proc** **ttest** data=report.new\_climate\_data;

class YEAR;

var WDSP;

where year in ("1963", "2023");

**run**;

/\* T-test for temperature between 1963 and 2023 \*/

title "t-Test for Temperatures between 1963 and 2023";

**proc** **ttest** data=report.new\_climate\_data;

class YEAR;

var TEMP;

where year in ("1963", "2023");

**run**;

/\* T-test for temperature between 1975 and 2023 \*/

title "t-Test for Temperatures between 1975 and 2023";

**proc** **ttest** data=report.new\_climate\_data;

class YEAR;

var TEMP;

where year in ("1975", "2023");

**run**;

/\* T-test for max temperature between 1975 and 2023 \*/

title "t-Test for Max Temperatures between 1975 and 2023";

**proc** **ttest** data=report.new\_climate\_data;

class YEAR;

var MAX;

where year in ("1975", "2023");

**run**;

/\* T-test for summer temperature between 1975 and 2023 \*/

title "t-Test for Summer Temperatures between 1975 and 2023";

**proc** **ttest** data=report.new\_climate\_data;

where SEASON = 'Summer' and YEAR in ("1975", "2023");

class YEAR;

var TEMP;

**run**;

/\* T-test for winter temperature between 1975 and 2023 \*/

title "t-Test for Winter Temperatures between 1975 and 2023";

**proc** **ttest** data=report.new\_climate\_data;

where SEASON = 'Winter' and YEAR in ("1975", "2023");

class YEAR;

var TEMP;

**run**;

/\* Simple linear regression for dew point vs temperature \*/

title "Simple Linear Regression Dew Point vs. Temperature";

**proc** **reg** data=report.new\_climate\_data;

model TEMP = DEWP;

**run**;

/\*\*\*\*\* Exporting CSV \*\*\*\*\*/

/\* Export new CSV file for cleaned climate data \*/

**proc** **export** data=report.new\_climate\_data

outfile="C:\Users\Brino\OneDrive\Desktop\Fall 2024\Statistical Methods\Week 14\Report\cleaned\_climate\_data.csv"

dbms=csv

replace;

**run**;