Climate Change:

Have Climate Variables Changed Over Time?

By: Brianna Noel

Contents

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.
    a. 1963
    b. 1975
    c. 1987
    d. 1999
    e. 2011
    f. 2023
7. **Season:** The season:
    a. Winter- December, January, February
    b. Spring- March, April, May
    c. Summer- June, July, August
    d. 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
    a. Low Pressure
    b. Normal Pressure
    c. High Pressure
9. **Mean Visibility Range:** Visibility ranges based on mean visibility for the day in miles to tenths
    a. Low Visibility
    b. Moderate Visibility
    c. 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.

a. 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.
b. 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.
c. Frequency tables were then created to observe the remaining variables. Any unnecessary variables identified from the frequency tables were removed.
d. Among the remaining variables, there was no missing data, so further data cleaning was unnecessary at this step.
e. 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.
f. 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.

g. 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

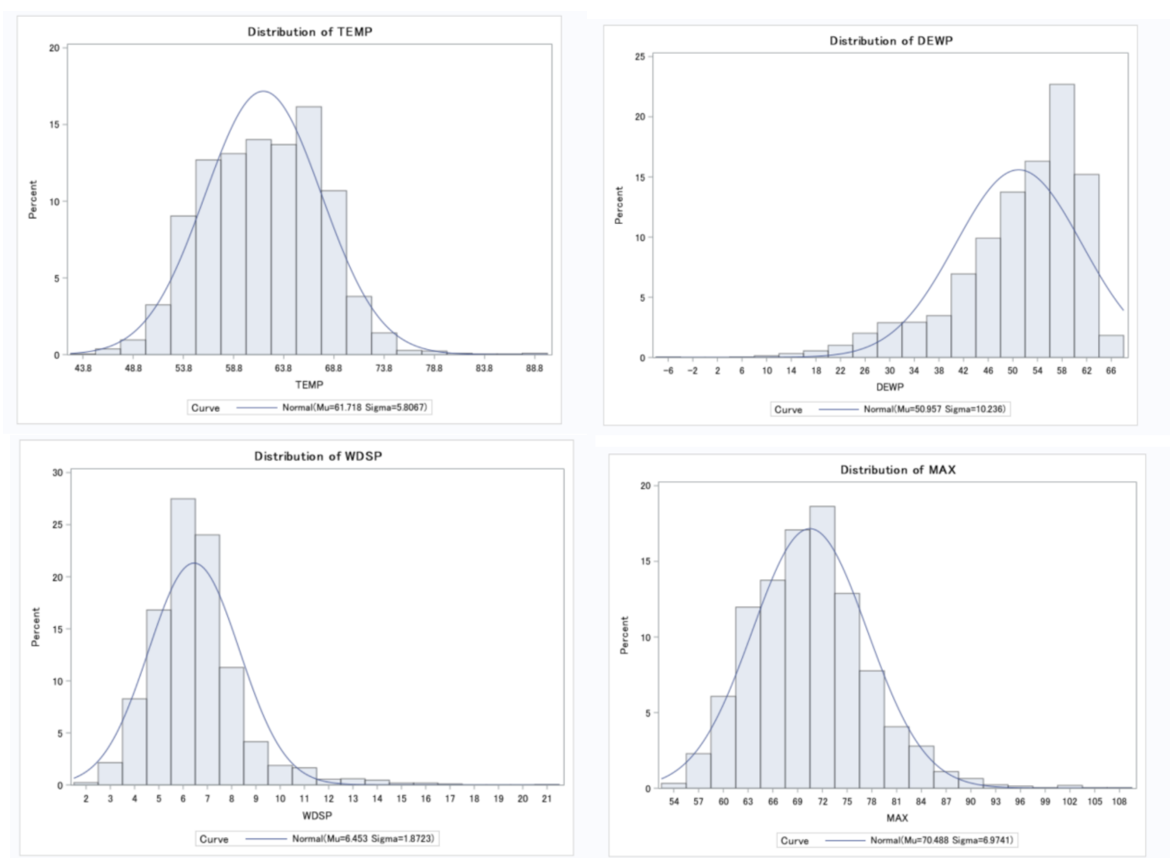
Descriptive Statistics

Descriptive Statistics

The MEANS Procedure

Variable	N	Mean	Median	Std Dev	Range	Lower Quartile	Upper Quartile
TEMP	2190	61.7180365	61.9000000	5.8067452	45.0000000	57.2000000	66.1000000
DEWP	2190	50.9570320	53.5000000	10.2355115	73.6000000	46.1000000	58.5000000
WDSP	2190	6.4530137	6.3000000	1.8722804	19.5000000	5.3000000	7.3000000
MAX	2190	70.4884475	70.0000000	6.9741282	55.0000000	66.0000000	75.0000000

Histograms



"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

Frequency Tables

The FREQ Procedure

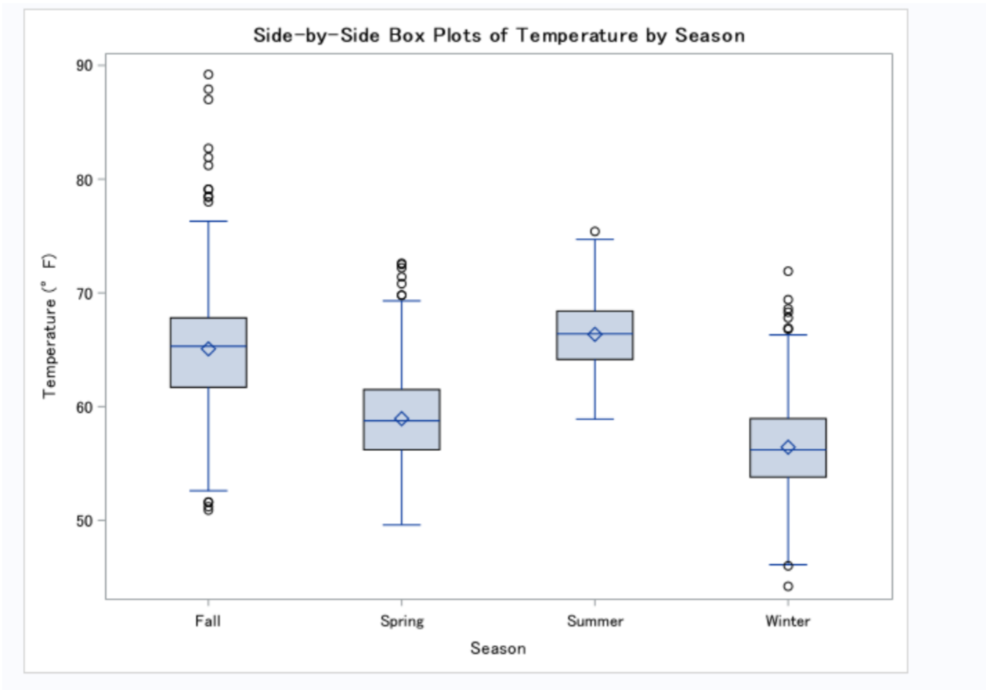
YEAR	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1963	365	16.67	365	16.67
1975	365	16.67	730	33.33
1987	365	16.67	1095	50.00
1999	365	16.67	1460	66.67
2011	365	16.67	1825	83.33
2023	365	16.67	2190	100.00

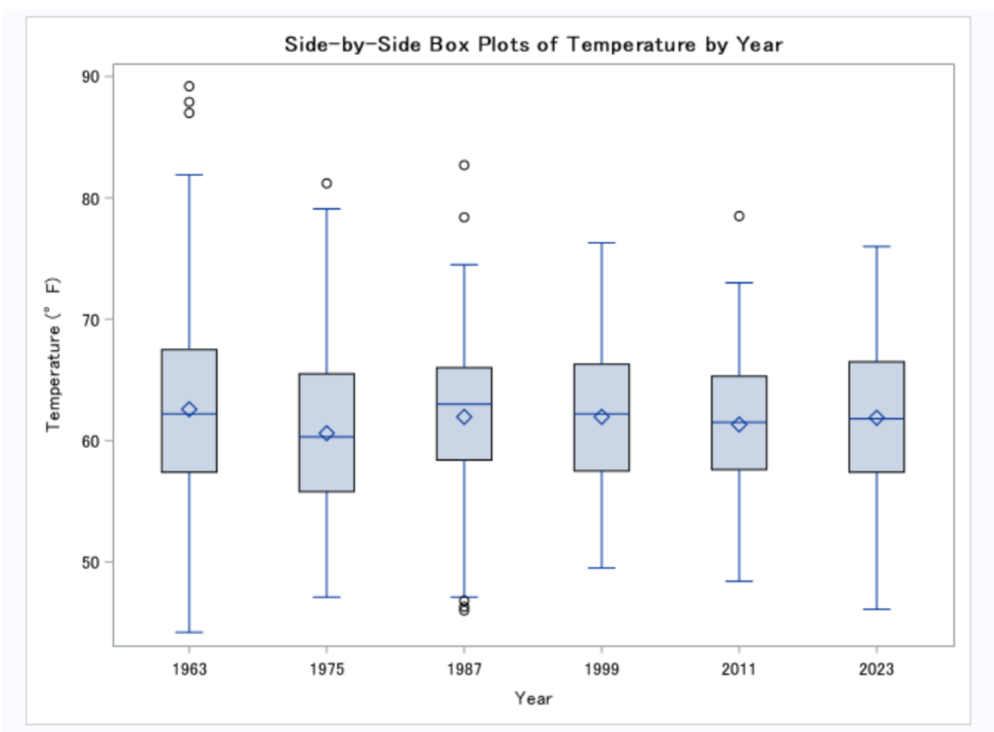
SEASON	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Fall	546	24.93	546	24.93
Spring	552	25.21	1098	50.14
Summer	552	25.21	1650	75.34
Winter	540	24.66	2190	100.00

SLP_CAT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
High Pressure	549	25.07	549	25.07
Low Pressure	540	24.66	1089	49.73
Normal Pressure	1101	50.27	2190	100.00

VISIB_CAT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
High Visibility	554	25.30	554	25.30
Low Visibility	541	24.70	1095	50.00
Moderate Visibility	1095	50.00	2190	100.00

Side-By-Side Box Plots



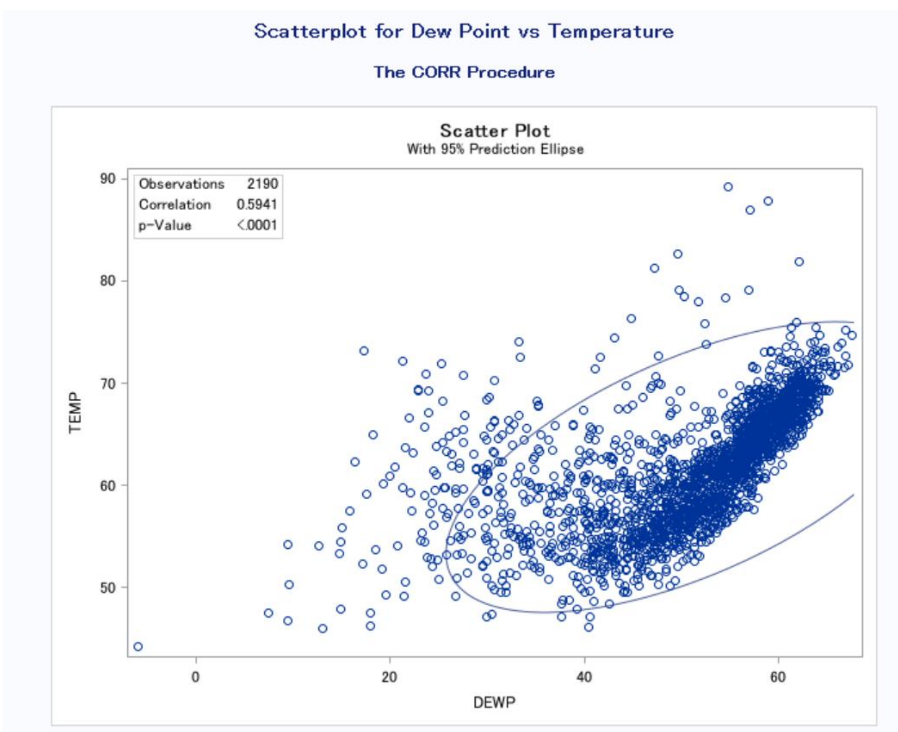


Correlation Matrix

Pearson Correlation Coefficients, N = 2190 Prob > |r| under H0: Rho=0				
	TEMP	**DEWP**	**WDSP**	**MAX**
TEMP	1.00000	0.59408 <.0001	−0.05383 0.0118	0.84209 <.0001
DEWP	0.59408 <.0001	1.00000	0.00926 0.6650	0.28268 <.0001
WDSP	−0.05383 0.0118	0.00926 0.6650	1.00000	−0.24322 <.0001
MAX	0.84209 <.0001	0.28268 <.0001	−0.24322 <.0001	1.00000

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



Statistical Methods

ANOVA and Interpretation of Results (Decision and Conclusion)

ANOVA for Mean Temperatures Across Years

$$H_0:\ \mu_{1963} = \mu_{1975} = \mu_{1987} = \mu_{1999} = \mu_{2011} = \mu_{2023}$$

$$H_A: At\ least\ one\ \mu_i = \mu_j\ for\ some\ i\ \neq j$$

One-Way ANOVA for Temperatures Across Years

The ANOVA Procedure

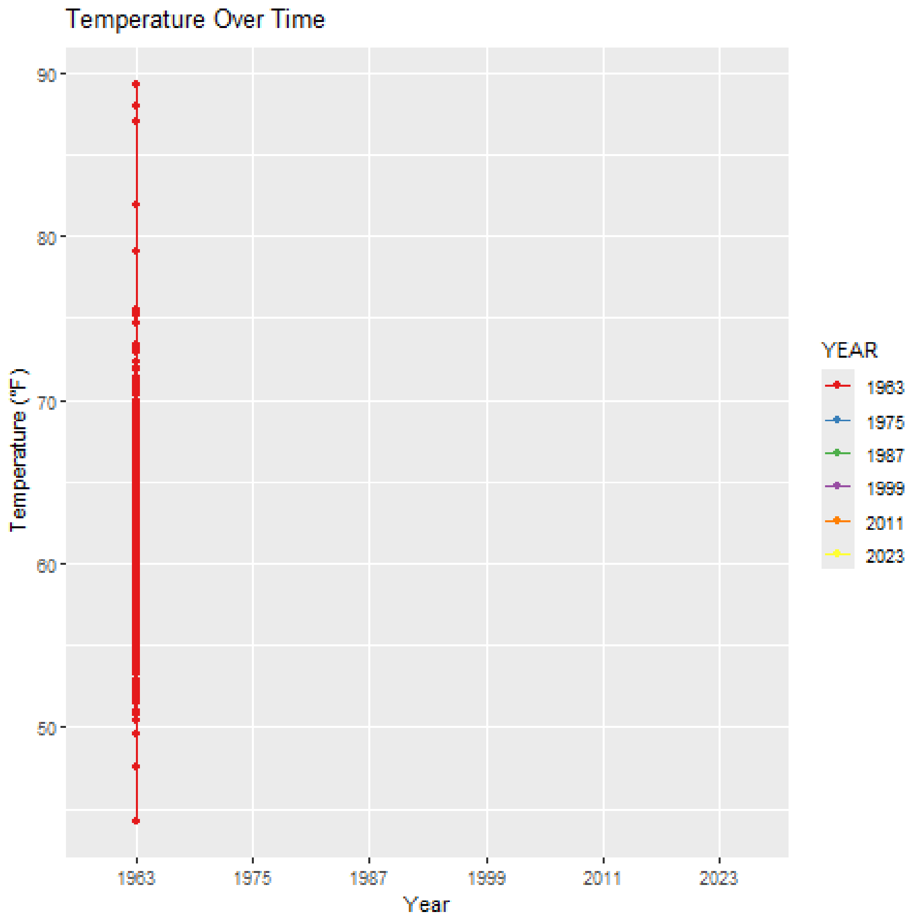
Dependent Variable: TEMP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	829.46435	165.89287	4.96	0.0002
Error	2184	72979.87321	33.41569		
Corrected Total	2189	73809.33756			

R-Square	Coeff Var	Root MSE	TEMP Mean
0.011238	9.366194	5.780631	61.71804

Source	DF	Anova SS	Mean Square	F Value	Pr > F
YEAR	5	829.4643516	165.8928703	4.96	0.0002

Since the p-value (0.0002) < $\alpha = 0.05$, 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.



Dynamic Visualization of Temperature Changes Across Years

ANOVA for Mean Summer Temperatures Across Years

$$H_0:\ \mu_{Summer\ of\ 1963} = \mu_{Summer\ of\ 1975} = \mu_{Summer\ of\ 1987} = \mu_{Summer\ of\ 1999} = \mu_{Summer\ of\ 2011} = \mu_{Summer\ of\ 2023}$$

$$H_A: At\ least\ one\ \mu_i = \mu_j\ for\ some\ i\ \neq j$$

One-Way ANOVA for Summer Temperatures Across Years

The ANOVA Procedure

Dependent Variable: TEMP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	315.740308	63.148062	7.72	<.0001
Error	546	4466.918804	8.181170		
Corrected Total	551	4782.659112			

R-Square	Coeff Var	Root MSE	TEMP Mean
0.066018	4.310806	2.860274	66.35127

Source	DF	Anova SS	Mean Square	F Value	Pr > F
YEAR	5	315.7403080	63.1480616	7.72	<.0001

Since the p-value (<0.0001) < $\alpha = 0.05$, 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

$$H_0:\ \mu_{Winter\ of\ 1963} = \mu_{Winter\ of\ 1975} = \mu_{Winter\ of\ 1987} = \mu_{Winter\ of\ 1999} = \mu_{Winter\ of\ 2011} = \mu_{Winter\ of\ 2023}$$

$$H_A\text{: } At\ least\ one\ \mu_i = \mu_j\ for\ some\ i\ \neq j$$

One-Way ANOVA for Winter Temperatures Across Years

The ANOVA Procedure

Dependent Variable: TEMP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	344.744611	68.948922	4.19	0.0010
Error	534	8792.503889	16.465363		
Corrected Total	539	9137.248500			

R-Square	Coeff Var	Root MSE	TEMP Mean
0.037730	7.190561	4.057753	56.43167

Source	DF	Anova SS	Mean Square	F Value	Pr > F
YEAR	5	344.7446111	68.9489222	4.19	0.0010

Since the p-value (0.0010) < $\alpha = 0.05$, 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

---

$$H_0\text{: } \mu_{low} = \mu_{normal} = \mu_{high}$$

$$H_A\text{: } At\ least\ one\ \mu_i = \mu_j\ for\ some\ i\ \neq j$$

One-Way ANOVA for Temperatures Across Sea Level Pressure Ranges

The ANOVA Procedure

Dependent Variable: TEMP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	14747.98890	7373.99445	273.05	<.0001
Error	2187	59061.34866	27.00565		
Corrected Total	2189	73809.33756			

R-Square	Coeff Var	Root MSE	TEMP Mean
0.199812	8.420060	5.196696	61.71804

Source	DF	Anova SS	Mean Square	F Value	Pr > F
SLP_CAT	2	14747.98890	7373.99445	273.05	<.0001

Since the p-value (<0.0001) < $\alpha = 0.05$, 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

$$H_0\text{: } \mu_{low} = \mu_{moderate} = \mu_{high}$$

$$H_A: At\ least\ one\ \mu_i = \mu_j\ for\ some\ i\ \neq j$$

One-Way ANOVA for Temperatures Across Visibilty Ranges

The ANOVA Procedure

Dependent Variable: TEMP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	966.71346	483.35673	14.51	<.0001
Error	2187	72842.62410	33.30710		
Corrected Total	2189	73809.33756			

R-Square	Coeff Var	Root MSE	TEMP Mean
0.013097	9.350962	5.771230	61.71804

Source	DF	Anova SS	Mean Square	F Value	Pr > F
VISIB_CAT	2	966.7134599	483.3567299	14.51	<.0001

Since the p-value (<0.0001) < $\alpha = 0.05$, 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

$$H_0:\ \mu_{1963} = \mu_{1975} = \mu_{1987} = \mu_{1999} = \mu_{2011} = \mu_{2023}$$

$$H_A: At\ least\ one\ \mu_i = \mu_j\ for\ some\ i\ \neq j$$

One-Way ANOVA for Wind Speeds Across Years

The ANOVA Procedure

Dependent Variable: WDSP

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	326.359384	65.271877	19.40	<.0001
Error	2184	7347.035726	3.364027		
Corrected Total	2189	7673.395110			

R-Square	Coeff Var	Root MSE	WDSP Mean
0.042531	28.42282	1.834128	6.453014

Source	DF	Anova SS	Mean Square	F Value	Pr > F
YEAR	5	326.3593836	65.2718767	19.40	<.0001

Since the p-value (<0.0001) < $\alpha = 0.05$, 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

$$H_0:\ \mu_{1963} = \mu_{2023}$$

$$H_A:\ \mu_{1963} \neq \mu_{2023}$$

t-Test for Wind Speeds between 1963 and 2023

The TTEST Procedure

Variable: WDSP

YEAR	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1963		365	6.8205	1.7008	0.0890	3.1000	15.8000
2023		365	6.5416	1.9942	0.1044	3.3000	21.2000
Diff (1-2)	Pooled		0.2789	1.8534	0.1372		
Diff (1-2)	Satterthwaite		0.2789		0.1372		

YEAR	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1963		6.8205	6.6455	6.9956	1.7008	1.5858	1.8341
2023		6.5416	6.3364	6.7469	1.9942	1.8593	2.1505
Diff (1-2)	Pooled	0.2789	0.00957	0.5482	1.8534	1.7629	1.9537
Diff (1-2)	Satterthwaite	0.2789	0.00955	0.5483			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	728	2.03	0.0424
Satterthwaite	Unequal	710.31	2.03	0.0424

Since the p-value (0.0424) < $\alpha = 0.05$, 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

$$H_0:\ \mu_{1963} = \mu_{2023}$$

$$H_A:\ \mu_{1963} \neq \mu_{2023}$$

t-Test for Temperatures between 1963 and 2023

The TTEST Procedure

Variable: TEMP

YEAR	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1963		365	62.5822	6.4080	0.3354	44.2000	89.2000
2023		365	61.8800	5.7486	0.3009	46.1000	76.0000
Diff (1-2)	Pooled		0.7022	6.0873	0.4506		
Diff (1-2)	Satterthwaite		0.7022		0.4506		

YEAR	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1963		62.5822	61.9226	63.2418	6.4080	5.9744	6.9100
2023		61.8800	61.2883	62.4717	5.7486	5.3597	6.1989
Diff (1-2)	Pooled	0.7022	-0.1824	1.5868	6.0873	5.7900	6.4169
Diff (1-2)	Satterthwaite	0.7022	-0.1825	1.5868			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	728	1.56	0.1196
Satterthwaite	Unequal	719.58	1.56	0.1196

Since the p-value (0.1196) > $\alpha = 0.05$, 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

$$H_0:\ \mu_{1975} = \mu_{2023}$$

$$H_A:\ \mu_{1975} \neq \mu_{2023}$$

t-Test for Temperatures between 1975 and 2023

The TTEST Procedure

Variable: TEMP

YEAR	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1975		365	60.6066	5.8712	0.3073	47.1000	81.2000
2023		365	61.8800	5.7486	0.3009	46.1000	76.0000
Diff (1-2)	Pooled		-1.2734	5.8102	0.4301		
Diff (1-2)	Satterthwaite		-1.2734		0.4301		

YEAR	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1975		60.6066	60.0022	61.2109	5.8712	5.4739	6.3311
2023		61.8800	61.2883	62.4717	5.7486	5.3597	6.1989
Diff (1-2)	Pooled	-1.2734	-2.1178	-0.4291	5.8102	5.5265	6.1249
Diff (1-2)	Satterthwaite	-1.2734	-2.1178	-0.4291			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	728	-2.96	0.0032
Satterthwaite	Unequal	727.68	-2.96	0.0032

Since the p-value (0.0032) < $\alpha = 0.05$, 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

$$H_0\colon\ \mu_{1975} = \mu_{2023}$$

$$H_A\colon\ \mu_{1975} \neq \mu_{2023}$$

t-Test for Max Temperatures between 1975 and 2023

The TTEST Procedure

Variable: MAX

YEAR	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1975		365	70.3660	7.0652	0.3698	57.0000	100.9
2023		365	70.5636	7.0635	0.3697	54.0000	91.0000
Diff (1-2)	Pooled		-0.1975	7.0644	0.5229		
Diff (1-2)	Satterthwaite		-0.1975		0.5229		

YEAR	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1975		70.3660	69.6388	71.0933	7.0652	6.5871	7.6186
2023		70.5636	69.8365	71.2906	7.0635	6.5856	7.6169
Diff (1-2)	Pooled	-0.1975	-1.2242	0.8291	7.0644	6.7194	7.4469
Diff (1-2)	Satterthwaite	-0.1975	-1.2242	0.8291			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	728	-0.38	0.7057
Satterthwaite	Unequal	728	-0.38	0.7057

Since the p-value (0.7057) > $\alpha = 0.05$, 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

$$H_0:\ \mu_{Summer\ of\ 1975} = \mu_{Summer\ of\ 2023}$$

$$H_A:\ \mu_{Summer\ of\ 1975} \neq \mu_{Summer\ of\ 2023}$$

t-Test for Summer Temperatures between 1975 and 2023

The TTEST Procedure

Variable: TEMP

YEAR	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1975		92	65.8630	2.8970	0.3020	60.3000	73.9000
2023		92	66.7098	3.6416	0.3797	60.8000	74.0000
Diff (1-2)	Pooled		-0.8467	3.2904	0.4851		
Diff (1-2)	Satterthwaite		-0.8467		0.4851		

YEAR	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1975		65.8630	65.2631	66.4630	2.8970	2.5303	3.3889
2023		66.7098	65.9556	67.4639	3.6416	3.1807	4.2598
Diff (1-2)	Pooled	-0.8467	-1.8040	0.1105	3.2904	2.9843	3.6670
Diff (1-2)	Satterthwaite	-0.8467	-1.8043	0.1108			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	182	-1.75	0.0826
Satterthwaite	Unequal	173.24	-1.75	0.0827

Since the p-value (0.0826) > $\alpha = 0.05$, 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

$$H_0:\ \mu_{Winter\ of\ 1975} = \mu_{Winter\ of\ 2023}$$

$$H_A:\ \mu_{Winter\ of\ 1975} \neq \mu_{Winter\ of\ 2023}$$

t-Test for Winter Temperatures between 1975 and 2023

The TTEST Procedure

Variable: TEMP

YEAR	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1975		90	55.4811	3.7501	0.3953	47.1000	66.8000
2023		90	56.7756	3.8854	0.4096	46.1000	65.7000
Diff (1-2)	Pooled		-1.2944	3.8183	0.5692		
Diff (1-2)	Satterthwaite		-1.2944		0.5692		

YEAR	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1975		55.4811	54.6957	56.2666	3.7501	3.2709	4.3951
2023		56.7756	55.9618	57.5893	3.8854	3.3889	4.5537
Diff (1-2)	Pooled	-1.2944	-2.4177	-0.1712	3.8183	3.4595	4.2608
Diff (1-2)	Satterthwaite	-1.2944	-2.4177	-0.1712			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	178	-2.27	0.0242
Satterthwaite	Unequal	177.78	-2.27	0.0242

Since the p-value (0.0242) < $\alpha = 0.05$, 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

$$H_0\colon\ \beta_1(Slope\ of\ the\ Line) = 0$$

$$H_A\colon\ \beta_1(Slope\ of\ the\ Line) \neq 0$$

Simple Linear Regression Dew Point vs. Temperature

The REG Procedure
Model: MODEL1
Dependent Variable: TEMP

Number of Observations Read	2190
Number of Observations Used	2190

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	26050	26050	1193.40	<.0001
Error	2188	47760	21.82806		
Corrected Total	2189	73809			

Root MSE	4.67205	**R-Square**	0.3529
Dependent Mean	61.71804	**Adj R-Sq**	0.3526
Coeff Var	7.56999		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	44.54403	0.50707	87.85	<.0001
DEWP	1	0.33703	0.00976	34.55	<.0001

Since the p-value (<0.0001) < $\alpha = 0.05$, 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:

$$\hat{y} = \mathbf{44.54403 + 0.33703}x$$

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…

a. There was not sufficient evidence to conclude that the mean temperatures across the years were statistically the same
b. There was not sufficient evidence to conclude that the mean summer temperatures across the years were statistically the same
c. There was not sufficient evidence to conclude that the mean winter temperatures across the years were statistically the same
d. There was not sufficient evidence to conclude that the mean temperatures across mean sea level pressure ranges were statistically the same
e. There was not sufficient evidence to conclude that the mean temperatures across mean visibility ranges were statistically the same
f. 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…

a. There was not sufficient evidence to conclude that the mean wind speeds in 1963 and 2023 are statistically the same
b. There was not sufficient evidence to conclude that the mean temperatures in 1963 and 2023 are statistically different
c. There was not sufficient evidence to conclude that the mean temperatures in 1975 and 2023 are statistically the same
d. There was not sufficient evidence to conclude that the max temperatures in 1975 and 2023 are statistically different
e. There was not sufficient evidence to conclude that the mean summer temperatures in 1975 and 2023 are statistically different
f. 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:

a. **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,
b. **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.
c. **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;
```