

TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Data Overview

This report utilizes a comprehensive dataset spanning nearly five years (January 2020 – May 2024), aimed at analyzing global climate dynamics. Unlike standard meteorological datasets, this collection integrates traditional weather metrics (such as temperature, precipitation, and humidity) with critical environmental and geographical indicators. By combining atmospheric data with urbanization indices, CO2 concentrations, and particulate matter levels, the dataset provides a holistic view of the interaction between human activity and climate patterns across various global coordinates and altitudes.



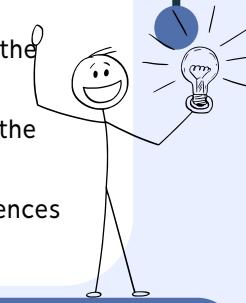
Month	Date	Avg_Temp	Max_Temp	Min_Temp	Precipitation	Humidity	Wind_Speed	Solar_Irradiance	Cloud_Cover	CO2_Concentration	Latitude	Longitude	Altitude	Proximity_to_Water	Urbanization_Index	Vegetation_Index	ENSO_Index	Particulate_Matter	Sea_Surface_Temp
2020	2 - 2020-02-01	7.69964234	25.901915	3.1730745	2957344	95.17102	10.5480480	252.31094	30.8566952	419.559	40.7120	-74.006	10	15	9.29990e+04	0.61304003	-4.200594e-01	24.30470	15.50959
2020	3 - 2020-03-01	7.69964234	18.63421	10.4304821	64.32324	94.95948	11.0002736	247.39140	20.3002236	416.6505	40.7120	-74.006	10	15	3.47715e-01	0.46651930	-4.200594e-01	32.30270	21.28100
2020	5 - 2020-05-01	19.8909048	36.082159	10.076997	15.9720223	73.99665	13.180301	169.54644	2.1937105	401.0646	40.7120	-74.006	10	15	7.617320e-02	0.25119651	-2.061405e-01	22.64605	15.50267
2020	6 - 2020-06-01	2.9500832	18.819905	21.162107	26.294740	35.52644	27.989051	158.75344	96.0547075	400.3515	40.7120	-74.006	10	15	7.319470e-01	0.04055950	2.350706e-01	28.25000	15.50051
2020	8 - 2020-08-01	15.241457	23.019366	12.824100	24.283125	75.3805	14.2432400	89.70139	92.0199697	423.20705	40.7120	-74.006	10	15	4.620463e-01	0.3570404	5.235904e-01	40.53486	15.70015
2020	9 - 2020-09-01	34.3622057	47.777639	39.16120	106.43042	94.61957	8.0009585	227.88279	90.8030524	412.891	40.7120	-74.006	10	15	3.66741e-01	0.35640303	-1.147100e-01	34.35790	14.49461
2020	10 - 2020-10-01	25.0791364	7.6482655	15.0791109	157.0021	66.02121	9.121695	235.22905	51.9337020	444.605	40.7120	-74.006	10	15	2.51307e-01	0.12970201	-7.330704e-01	15.12072	15.13093
2020	11 - 2020-11-01	11.64340211	16.300470	15.541059	15.300470	15.300470	0.000581	106.18524	106.18524	417.1719	40.7120	-74.006	10	15	1.377564e-01	0.22564805	9.000000e-01	37.23240	24.24007
2021	1 - 2021-01-01	7.58786661	15.279342	6.0404057	105.01403	67.46966	8.061072	139.63255	78.7567901	425.652	9999999999999999	-74.006	10	15	9.370211e-01	0.33019552	7.091903e-01	10.00000	10.00000
2021	3 - 2021-03-01	17.05919304	22.484153	25.511360	17.114277	49.27952	20.030398	106.42712	24.803971	442.5642	40.7120	-74.006	10	15	7.112595e-01	0.0555031	-3.77965e-01	22.30793	26.33234
2021	5 - 2021-05-01	4.4620074	36.307219	34.8452359	30.17100	47.731169	26.446039	74.0652571	0.00040403	40.7120	-74.006	10	15	2.737616e-01	0.3642646	-6.207624e-01	30.77415	17.99666	
2021	8 - 2021-08-01	5.7306712	12.875159	12.057101	147.05711	94.34709	11.0022747	245.41191	90.985260	413.087	40.7120	-74.006	10	15	7.003950e-01	0.1519705	-7.101221e-01	40.64695	11.21057
2021	9 - 2021-09-01	10.70791205	37.966851	7.3795005	50.03432	60.37200	7.0171761	244.37701	74.8007211	403.072	40.7120	-74.006	10	15	2.530005e-01	0.5955350	7.080200e-01	43.35958	11.03082
2021	10 - 2021-10-01	9.44961516	37.221183	0.0028939	105.655330	67.43030	14.0570368	279.23491	72.3571703	406.7727	40.7120	-74.006	10	15	7.301920e-01	0.25350870	6.952319e-01	42.341807	12.44467
2021	11 - 2021-11-01	26.049007	4.05030	3.2749682	12.221952	41.03249	8.141670	23.22039	13.1872191	403.493	40.7120	-74.006	10	15	3.167320e-01	0.57070612	7.938203e-01	41.91930	13.05613
2021	12 - 2021-12-01	30.75134221	37.385212	12.221967	93.048105	40.33000	27.789085	126.034042	35.4709000	403.7540	40.7120	-74.006	10	15	7.164842e-01	0.43030743	2.059130e-01	35.94205	10.01747
2021	1 - 2021-12-01	29.75134229	22.484153	12.221967	22.484153	27.789085	12.3620714	248.94905	90.9709225	431.0114	40.7120	-74.006	10	15	9.859150e-02	0.15127041	10.016030	16.20209	16.20209
2021	12 - 2021-12-01	10.02058766	35.200070	9.5234691	64.07013	94.7013	64.04808	0.0471939	7.0147939	415.0515	40.7120	-74.006	10	15	7.403550e-01	0.14680230	-6.895700e-01	35.95977	21.30097
2021	12 - 2021-12-01	13.38691351	4.915621	12.875159	165.219410	55.33300	3.1777006	52.23391	0.021954932	435.070	40.7120	-74.006	10	15	6.923200e-01	0.15197050	-8.632400e-02	19.04071	26.41705
2022	1 - 2022-01-01	17.11413036	20.373155	12.057101	62.056718	54.917160	11.0373030	174.30791	10.0731030	413.3991	40.7120	-74.006	10	15	5.718055e-01	0.05053958	-2.737110e-01	44.25915	11.24404
2022	2 - 2022-02-01	30.6423816	25.799054	15.0534166	27.82015	58.73007	9.1216975	32.0004021	12.0200021	427.8327	40.7120	-74.006	10	15	9.441110e-01	0.16248441	-4.059986e-01	32.59002	26.51432
2022	3 - 2022-03-01	38.840716	24.28398	27.251104	162.163897	94.612487	9.5558793	145.5001	4.4717944	435.7910	40.7120	-74.006	10	15	2.235040e-01	0.1679461	-6.945320e-01	11.20200	15.30201
2022	4 - 2022-04-01	4.70518116	22.889123	27.945369	102.04640	47.78023	0.0987036	255.80414	30.3006407	432.5210	40.7120	-74.006	10	15	2.960050e-01	0.70834839	5.000000e-01	41.42027	16.09633
2022	5 - 2022-05-01	32.27192059	21.134659	9.920027	17.2232903	95.30000	12.0385111	265.09510	0.05556209	405.944	40.7120	-74.006	10	15	2.915160e-01	0.26300020	4.247456e-01	13.97571	26.50703
2022	6 - 2022-06-01	3.38119058	24.905132	3.6611920	111.052052	36.06641	10.0305721	23.02502	9.34403074	412.6130	40.7120	-74.006	10	15	2.911140e-01	0.19553538	1.300195e-01	42.04211	21.21302
2022	7 - 2022-07-01	1.84107067	20.675000	6.0205533	51.020002	36.09810	15.95995	204.19004	95.00046574	412.801	40.7120	-74.006	10	15	5.050050e-01	0.34710747	9.4337040e-01	14.44239	15.54405
2022	8 - 2022-08-01	-0.11320019	17.445677	16.019951	16.019951	40.09000	4.4940951	65.30753	63.1735379	415.0400	40.7120	-74.006	10	15	7.305775e-02	0.37482681	-7.132240e-02	20.47443	25.90063
2022	9 - 2022-09-01	2.0890125	11.284682	9.233654	41.07190	14.071932	107.50011	42.1766071	427.8601	40.7120	-74.006	10	15	6.170705e-01	0.16705707	4.574950e-01	21.85724	10.50151	
2022	10 - 2022-10-01	-4.0230011	17.074009	13.974009	12.02254	6.6591041	16.001441	41.1215307	431.155	40.7120	-74.006	10	15	1.876574e-01	0.13207370	3.100591e-01	32.86040	20.77290	
2022	12 - 2022-12-01	3.31119009	30.910475	17.045366	17.045366	79.37937	4.0000004	97.7314	93.5150623	423.079	40.7120	-74.006	10	15	9.493020e-01	0.16570327	1.621950e-01	30.09653	26.30379
2022	12 - 2022-12-01	22.74900005	35.4000132	7.703567	17.0110707	32.04008	30.1795453	72.23506	67.9513070	414.0001	40.7120	-74.006	10	15	7.444310e-01	0.0880370	1.527070e-01	30.6600000	26.45235
2023	1 - 2023-01-01	4.6700000	-28.050000	-28.050000	12.03692	42.03691	13.0000509	42.0369000	405.61304	40.7120	-74.006	10	15	5.255710e-01	0.74000007	9.965545e-01	30.60749	26.96005	
2023	2 - 2023-02-01	28.27773641	26.900002	19.641100	19.641100	52.02736	12.7912002	205.93542	47.9660002	440.195	40.7120	-74.006	10	15	9.234040e-01	0.08223754	-3.103000e-01	26.95777	12.59500
2023	3 - 2023-03-01	23.0719015	17.795015	17.795015	50.0200054	79.07002	12.7912004	19.10794	43.6001	41.7102	-74.006	10	15	3.884460e-01	9.496500e-02	1.045000e-01	14.21005	25.24028	
2023	5 - 2023-05-01	4.49547252	13.810005	6.3640523	67.934012	65.73002	4.0400954	225.18006	91.9515059	428.930	40.7120	-74.006	10	15	9.929520e-02	0.08004502	-8.016150e-02	31.03000	27.03006
2023	6 - 2023-06-01	7.01647040	26.0712136	27.3104504	17.1470000	9.79000412	292.73056	421.0006	40.7120	-74.006	10	15	9.910404e-01	0.3031952	7.016104e-01	14.60005	23.87		

Data Dictionary

1. Temporal & Spatial Metadata

Variables identifying the time and location of observations, essential for time-series and geospatial analysis.

- **Year (Integer):** Represents the year of data collection, ranging from 2020 to 2024. This variable is essential for performing annual trend analysis.
- **Month (Integer):** Represents the month of the year (1–12). It is useful for identifying seasonality and monthly weather patterns.
- **Date (Date, YYYY-MM-DD):** Represents the full date, specifically the first day of each month. It serves as the main axis for time-series analysis.
- **Latitude (Float, Degrees):** The geographic coordinate specifying the north-south position of the location.
- **Longitude (Float, Degrees):** The geographic coordinate specifying the east-west position of the location.
- **Altitude (Float, Meters):** The height of the location above sea level, which significantly influences local temperature and atmospheric pressure.



TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

HELP

2. Temperature Variables

Avg_Temp (Float, °C): The average air temperature measured at a height of 2 meters during the month. Expected range: -50°C to +60°C.

Max_Temp (Float, °C): The highest temperature recorded during the specific month.

Min_Temp (Float, °C): The lowest temperature recorded during the specific month.

3. Precipitation & Humidity

Precipitation (Float, mm): The total accumulated rainfall or snowfall in millimeters for the month.

Humidity (Float, %): The relative humidity percentage, indicating the amount of water vapor in the air relative to the maximum possible saturation.

4. Wind & Solar Variables

Wind_Speed (Float, m/s): The average speed of near-surface air movement measured in meters per second.

Solar_Irradiance (Float, W/m²): The power of solar energy reaching the Earth's surface, measured in watts per square meter.

Cloud_Cover (Float, %): The percentage of the sky obscured by clouds.

5. Greenhouse Gases

CO2_Concentration (Float, ppm): The concentration of Carbon Dioxide in the atmosphere, measured in parts per million. This is a key indicator for studying global warming.

6. Environmental & Urban Indicators

Proximity_to_Water (Float, km): The distance in kilometers to the nearest major body of water (ocean, sea, or large lake).

Urbanization_Index (Float, 0.0–1.0): A normalized index representing the level of urban development, where higher values indicate more urbanized areas.

Vegetation_Index (Float, -1.0 to +1.0): An index representing the density of green vegetation (e.g., NDVI). Positive values indicate healthier/denser vegetation.

7. Advanced Climate Indicators

ENSO_Index (Float, -2.0 to +2.0): The El Niño Southern Oscillation index, used to monitor ocean-climate interactions (El Niño and La Niña phases).

Particulate_Matter (Float, µg/m³): The concentration of air pollutants (PM10 / PM2.5) measured in micrograms per cubic meter.

Sea_Surface_Temp (Float, °C): The temperature of the water at the ocean's surface.



Data Cleaning & Preprocessing

To ensure high data quality and analytical accuracy, a comprehensive preprocessing pipeline was implemented using the R programming language (specifically tidyverse and zoo packages). The process addressed structural inconsistencies, missing values, and logical anomalies through the following stages:



TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT



3.1. Standardizing Data Structure (Renaming & Typing)

Renaming: Original column names were translated and standardized into concise English identifiers. This ensures compatibility with visualization libraries and mathematical functions.

Type Casting: All numerical features (Temperature, Wind Speed, etc.) were explicitly converted to numeric types, while temporal features (Year, Month) were cast to integer. This step prevents execution errors during statistical aggregation.

3.2. Temporal Feature Engineering

Since climate data depends heavily on time progression, a unified Date object was constructed by combining the Year and Month columns (assigning the 1st day of each month).

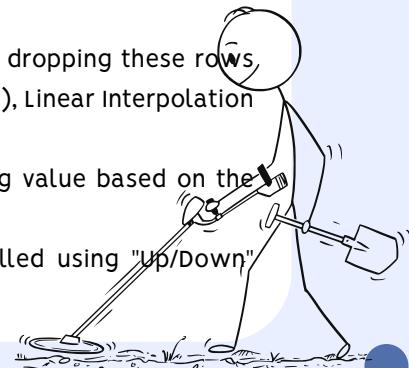
Purpose: This created a continuous time-series index, enabling accurate plotting of trends over the 53-month period.

3.3. Handling Missing Values (Imputation Strategy)

An analysis of the dataset revealed missing values across several columns. Instead of dropping these rows, (which would lead to data loss) or filling them with a static mean (which distorts trends), Linear Interpolation was applied using the na.approx function from the zoo package.

Justification: Climate data is continuous by nature. Interpolation estimates the missing value based on the point before and after it, preserving the natural fluctuation of weather patterns.

Edge Cases: Any remaining gaps at the very beginning or end of the series were filled using "Up/Down" propagation.



3.4. Deduplication

The dataset was scanned for identical duplicate rows.

Outcome: Duplicate records were identified and removed to ensure that each data point represents a unique observation, preventing bias in statistical calculations.

3.5. Logical Consistency & Outlier Treatment

To maintain scientific validity, the dataset was examined for physically implausible measurements. Several domain-based correction rules were applied:

Precipitation and Wind Speed:

Negative values are not physically meaningful and were therefore corrected to zero. In rare instances where negative wind speed appeared (likely due to sensor polarity issues), the absolute value was taken to preserve magnitude information.

Humidity and Cloud Cover:

Since these variables represent percentages, all values were constrained within the 0–100% range. Any measurement outside this domain was clipped to the nearest valid boundary.

Outlier Detection:

Extreme values were identified using the Interquartile Range (IQR) method. Outliers were replaced with the median to reduce distortion while maintaining the central tendency of the dataset.



TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Data Visualization & Exploratory Analysis

Exploratory Data Analysis (EDA)

Range: -5°C to 35°C (40°C difference)

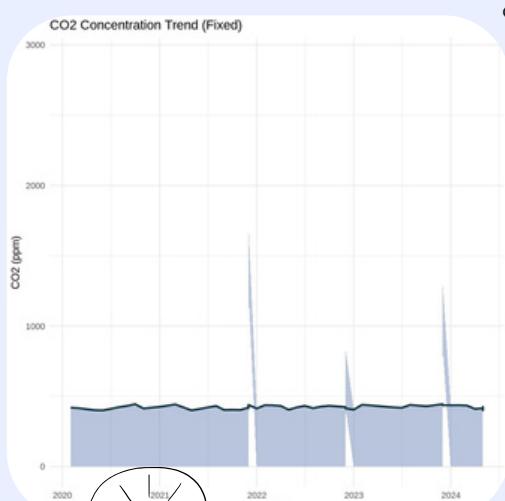
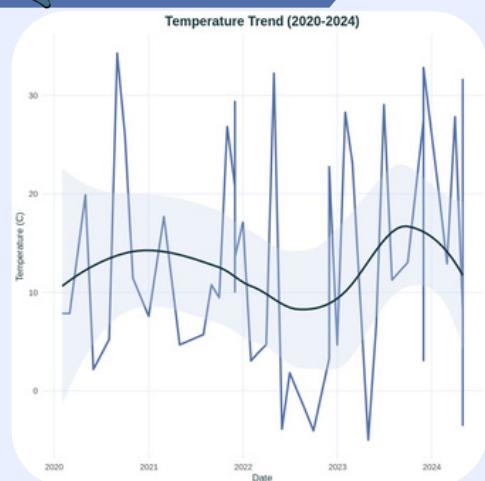
Overall Average: 13°C

Highest Average: 16°C (2023)

Lowest Average: 9°C (End of 2022)

Seasonal Fluctuations: ±15°C annually

Trend: Fluctuating with no clear upward/downward pattern



⚠ Warning: Data contains 3 outliers that were ignored

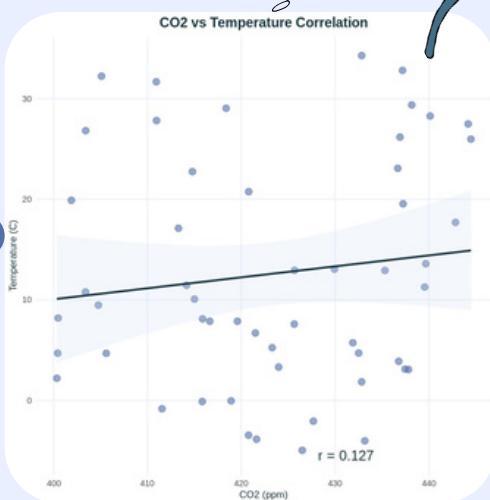
Normal Range: 408–425 ppm

Average: ~ 420 ppm

Fluctuations: ±5 ppm seasonally

Trend: Slight increase (+2–3 ppm from 2020 to 2024)

Conclusion: CO₂ concentration is relatively stable with a very slight increase – consistent with the trend.



Bivariate Correlation Analysis

Correlation Coefficient: $r = 0.127$

⚠ Interpretation: Very weak positive correlation (almost non-existent)

Data:

CO₂: 400–445 ppm

Temperature: -5 to 34°C

Number of Observations: 41 points

Result: Only 12.7% of the change in temperature is explained by CO₂ change. The remaining change (87.3%) is due to other factors:

✓ Seasonal fluctuations

✓ Humidity and Wind

✓ Geographical location

✓ Solar Irradiance

Conclusion: In this short-term monthly data, The effect of CO₂ on temperature is unclear. The real effect of CO₂ appears over the long term (decades).

TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Distribution Analysis

Distribution Type: Bimodal (Two peaks)

Data Distribution:

Cold (-5 to 5°C): 22% of data

Moderate (5–15°C): 43% (The most)

Warm (15–25°C): 17%

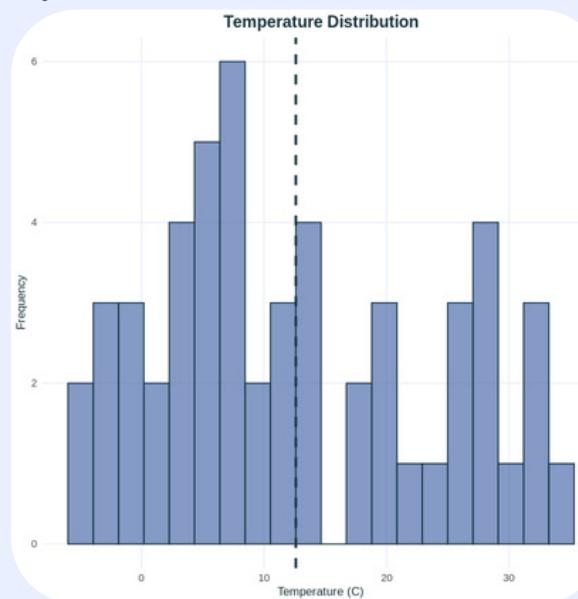
Hot (25–35°C): 18%

📌 The Two Peaks:

Winter Peak: ~ 7.5°C

Summer Peak: ~ 27.5°C

Conclusion: The distribution shows two clear seasons (Winter and Summer) With a greater concentration in the moderate region (5–15°C).



Correlation Matrix

	Avg_Temp	Precipitation	Humidity	Wind_Speed	CO2_Concentration
Avg_Temp	1	0.23	0.01	-0.06	0.13
Precipitation	0.23	1	0.24	-0.06	0.17
Humidity	0.01	0.24	1	0.24	0.27
Wind_Speed	0.23	-0.06	0.24	1	-0.13
CO2_Concentration	0.13	0.17	0.27	-0.13	1

Correlation Matrix

Strongest Correlations:

Humidity ↔ CO2: 0.27 (Weak positive)

Humidity ↔ Precipitation: 0.24

Humidity ↔ Wind_Speed: 0.24

Avg_Temp ↔ Precipitation: 0.23

Avg_Temp ↔ Wind_Speed: 0.23

⚠️ General Conclusion:

✓ All correlations are very weak (<0.3)

✓ Variables are relatively independent of each other

✓ CO2 has no clear effect on other variables

✓ Reason: Short-term data + strong seasonal fluctuations

Interpretation: In short-term monthly data (4.5 years), Seasonal fluctuations are much stronger than relationships between variables, Therefore, correlations are weak and do not reflect true long-term relationships.



TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Monthly Seasonality Analysis

Monthly Average Temperature (2020-2024):

Hottest Month: December (19.5°C)

Coldest Month: June (1.7°C)

Difference: 17.8°C

Annual Average: ~11.0°C

Month Distribution:

- Cold (< 10°C): 2 months
- Moderate (10-15°C): 8 months
- Warm (> 15°C): 2 months

Seasonal Pattern:

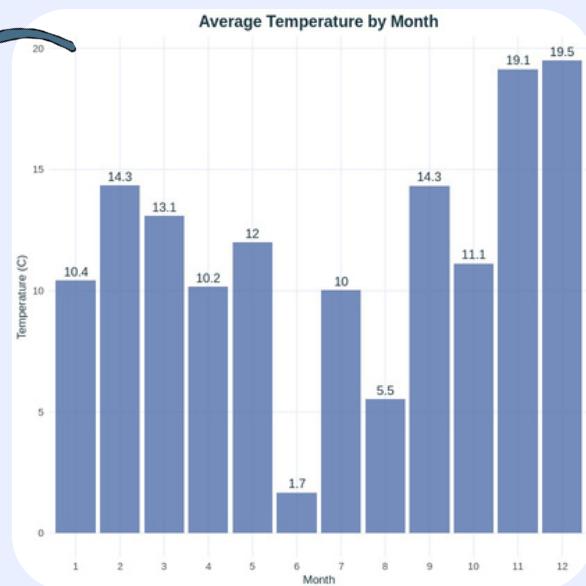
⚠ Pattern is inverted (unconventional):

- Winter (Dec-Feb): 14.7°C (Warm)
- Summer (Jun-Aug): 5.7°C (cold)
- Spring: 11.8°C (Moderate)
- Fall: 14.8°C (Warm Moderate)

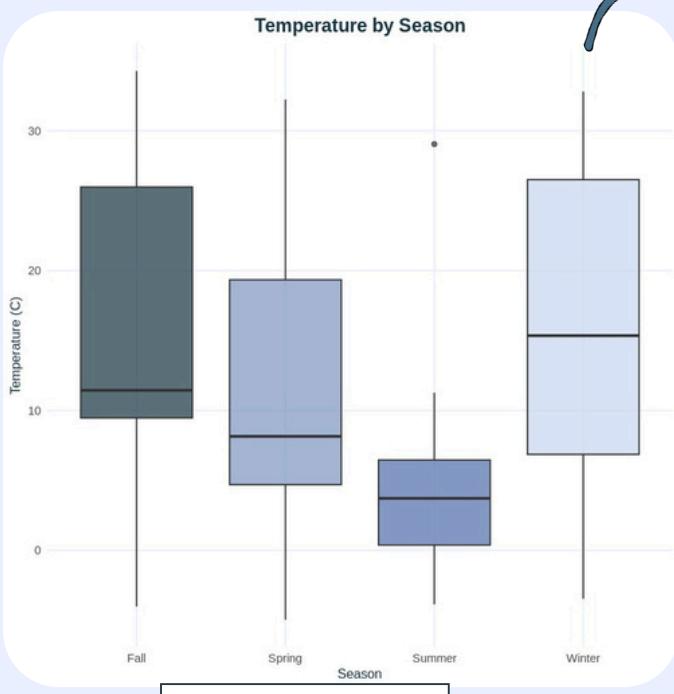
Conclusion: Most months are moderate (10-15°C) With two extremes: June is

very cold, December is warm. The pattern indicates a Southern Hemisphere

region or unconventional data.



Correlation Matrix



Temperature Analysis by Season (Box Plot):

📌 Median for each season:

- Winter: 15.5°C The Warmest!
- Fall: 11.5°C
- Spring: 8.0°C
- Summer: 4.0°C ❄️ The Coldest!

Variability (IQR):

- Winter: 20°C (Most variable)
- Fall: 17°C
- Spring: 14.5°C
- Summer: 6°C (Most stable)

Ranges:

- Winter: -3 to 33°C
- Fall: -3 to 34°C
- Spring: -5 to 32°C
- Summer: -3 to 11°C (+ outlier value 29.5°C)

⚠ Important Notes:

- ✓ Pattern is inverted (Winter is warmer than Summer)
 - ✓ Indicates a Southern Hemisphere location
 - ✓ Very high variability in all seasons
 - ✓ Summer is the most stable despite being the coldest
- Conclusion: Data shows a Southern Hemisphere climate With high fluctuations in all seasons Winter (Dec-Feb) is warm Summer (Jun-Aug) is cold

TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Descriptive Statistics

Average Climate Variables (2020–2024):

🌡️ Temperature:

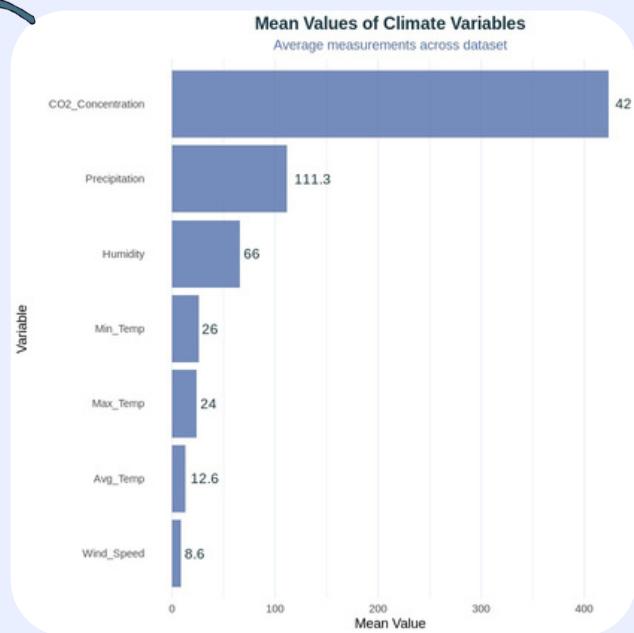
- Average: 12.6°C
- Max: 24°C
- Min: 26°C

Weather: • Humidity: 66% • Precipitation: 111.3 mm/month •

Wind: 8.6 m/s (~31 km/h)

CO2: 423 ppm (Global Level)

All values are logical and reflect a moderate climate.



Range & Variability Analysis

Climate Variable Ranges:

🌡️ Largest Variance: • Min_Temp: 0–1000 (Outlier – needs review)

• Precipitation: 0–250 mm

Low Variance (Stable):

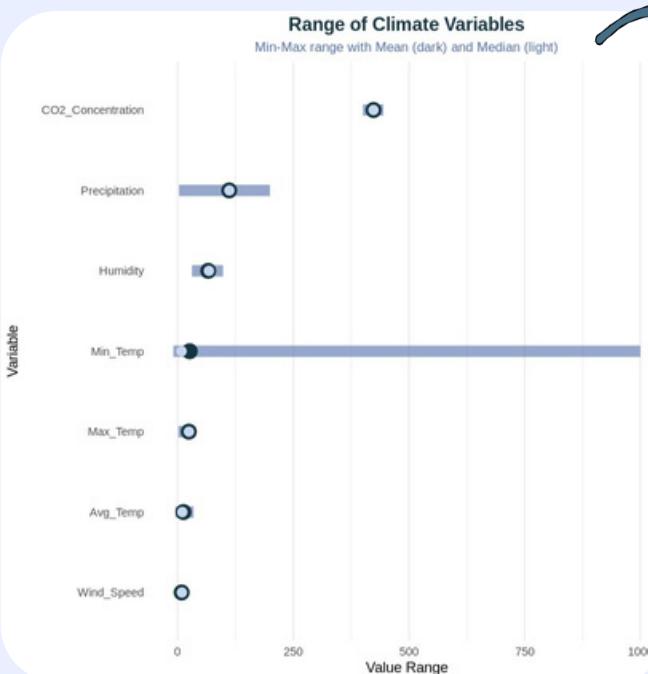
• CO2: 400–450 ppm

• Humidity: 40–90%

• Wind_Speed: 5–15 m/s

• Temperature: Narrow ranges

Conclusion: Most variables are stable with Mean ≈ Median



Descriptive Statistics

Average Climate Variables (2020–2024):

🌡️ Temperature:

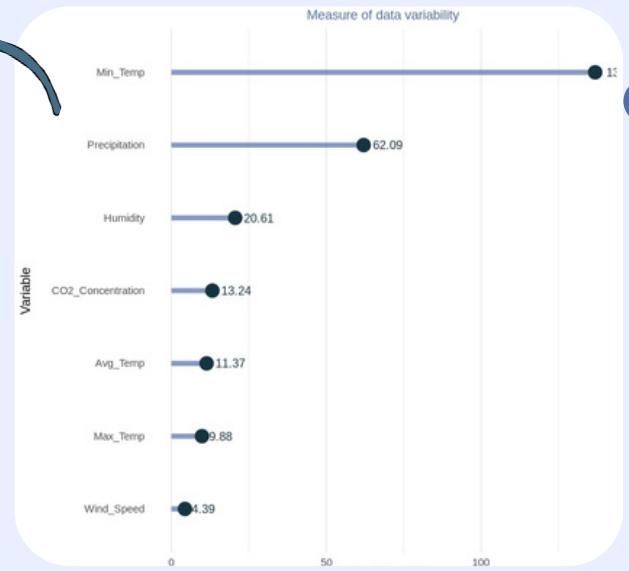
- Average: 12.6°C
- Max: 24°C
- Min: 26°C

Weather: • Humidity: 66% • Precipitation: 111.3 mm/month •

Wind: 8.6 m/s (~31 km/h)

CO2: 423 ppm (Global Level)

All values are logical and reflect a moderate climate.

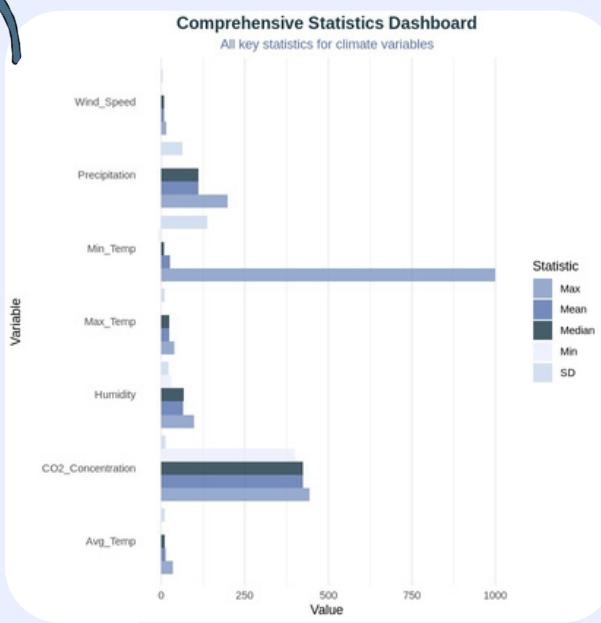
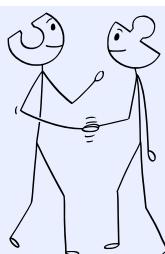


TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Comprehensive Statistical

Highest Numerical Values:

- CO2: 400–450 ppm (Highest)
 - Precipitation: 0–250 mm
 - Humidity: 40–90%
- Most Stable (Low SD):
- Wind_Speed
 - Max_Temp
 - Avg_Temp
- ⚠ Most Volatile:
- Precipitation (Normal)
 - Min_Temp (Outlier – needs correction)
- Mean ≈ Median = Balanced Distribution ✓



Stability Analysis

Coefficient of Variation (CV):

$$\text{Relative Variance} = (\text{SD} \div \text{Mean}) \times 100$$

Most Stable:

- CO2: 3.1% (Almost constant)
- Humidity: 31.2%
- Max_Temp: 41.1%

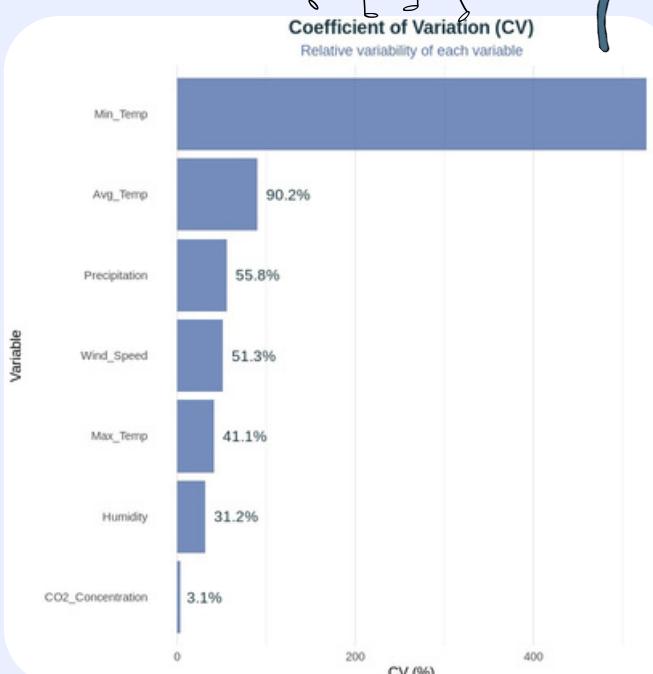
Moderate Variance:

- Wind_Speed: 51.3% • Precipitation: 55.8%

High Variance: • Avg_Temp: 90.2% (Seasonal)

Outlier: • Min_Temp: 520%+ (Needs correction)

The lower the CV = the higher the stability



TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT



One-Factor Analysis

Q-Q Plot for Avg_Temp

Line vs points:

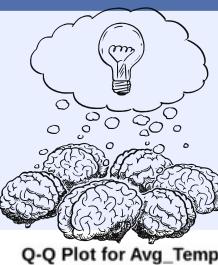
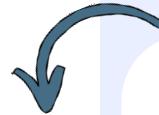
In a perfect normal distribution, all points would lie close to the straight line; here, many points curve away from the line, especially at both tails.

Tails behavior:

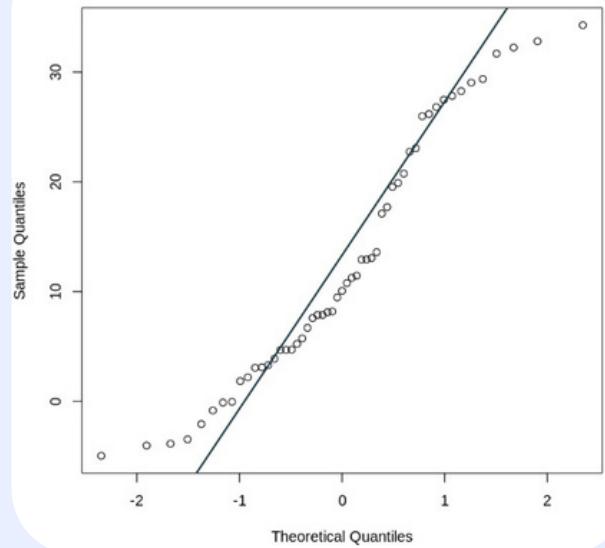
Points at the lower and upper ends are clearly below/above the line, indicating heavier or lighter tails than a normal distribution and possible skewness.

Conclusion:

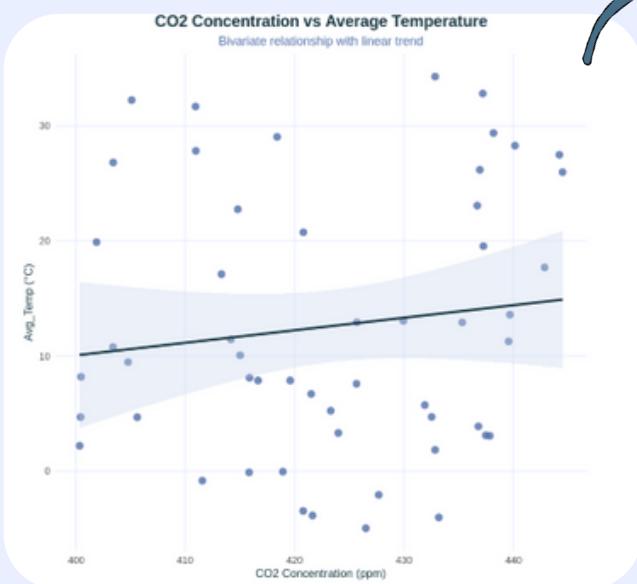
The Q-Q plot visually confirms the Shapiro-Wilk result: Avg_Temp does not follow a perfect normal distribution and shows strong deviations, likely due to seasonal temperature patterns.



Q-Q Plot for Avg_Temp



Two-Factor Analysis



Bivariate: CO2 vs Avg_Temp

Type:

Bivariate relationship between one predictor (CO2 concentration) and one response (average temperature).

Direction (trend):

The regression line is slightly upward, meaning higher CO2 values are associated with slightly higher Avg_Temp on average.

Strength (scatter):

Points are widely scattered around the line, so the linear relationship is weak; CO2 alone explains only a small part of the variation in Avg_Temp.



Practical reading:

CO2 and Avg_Temp move in the same direction overall (positive relationship), but temperature is also strongly affected by other factors (seasonality, location, etc.), so this BI plot shows weak but positive association rather than a strong predictive link.

TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Correlation Heatmap

Climate Variables (Multivariate)

Purpose (Multi):

This heatmap shows all pairwise correlations between several climate variables at the same time (Avg_Temp, Precipitation, Humidity, Wind_Speed, CO2_Concentration).

Darker blue squares = stronger positive relationship, lighter/whitish = weak or near zero, reddish = negative.

Main patterns:

Strong blue on the diagonal is each variable with itself (correlation = 1).

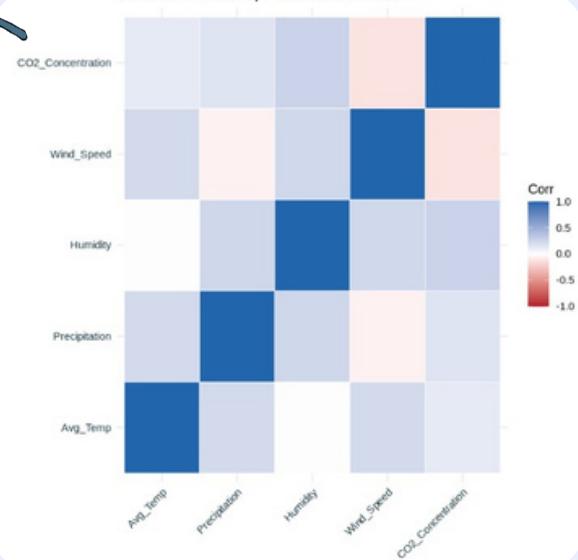
Some pairs like Precipitation–Humidity and Wind_Speed with itself on diagonal appear darker, meaning stronger positive linkage between those variables.

Many off-diagonal cells are pale (close to white), meaning weak correlations, so no single variable fully explains the others.

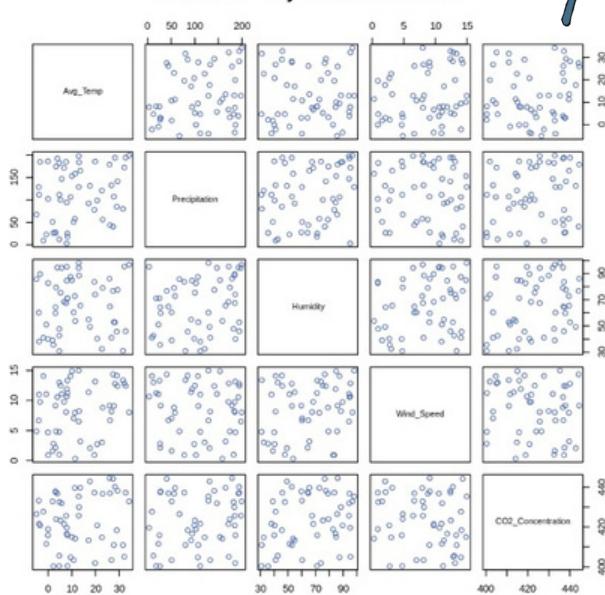
Why this is Multivariate:

Instead of looking at one pair (BI) like CO2 vs Avg_Temp, the heatmap summarizes the structure of relationships among all variables together, giving a global multivariate view of the climate system

Correlation Heatmap - Climate Variables



Pairs Plot - Key Climate Variables



Multi-Factor Analysis

Pairs Plot – Key Climate Variables (Multivariate)

What it shows:

This grid displays a scatter plot for every pair of variables (Avg_Temp, Precipitation, Humidity, Wind_Speed, CO2_Concentration), so each small box is one bivariate relationship, and together they form a multivariate overview.

Main pattern:

Most panels show points widely scattered without a clear line, meaning linear relationships between pairs of variables are generally weak, with no single strong predictor-response pair dominating the system.

Why it is Multivariate:

Instead of focusing on one or two variables, the pairs plot lets you check all pairwise interactions at once, helping to see which relationships might be interesting (more structured clouds) and which look almost random, as part of your multivariate EDA.



TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT



One-Way ANOVA

Objective: To test if Season significantly impacts Avg_Temp.

Results:

- F-statistic: 1.958
- P-value: 0.133 (Greater than 0.05)
- Conclusion: Not Significant. The analysis fails to reject the null hypothesis.

Key Insight: The Box Plot reveals extreme volatility within individual seasons (especially Winter). This high internal variance overshadows the differences between seasonal means, indicating a highly unpredictable climate regardless of the season.

ONE-WAY ANOVA: Effect of Season on Temperature

ANOVA Results:

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Season	3	720	240.1	1.958	0.133
Residuals	49	6007	122.6		

Effect Size:

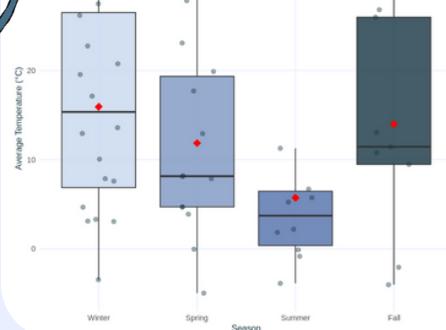
- Eta-squared (η^2): 0.107
- Interpretation: Medium effect

Conclusion:

- p-value = 0.1326 ≥ 0.05
- NOT SIGNIFICANT: No significant seasonal effect

One-Way ANOVA: Temperature by Season

F-test p-value: 0.1326 | $\eta^2 = 0.107$



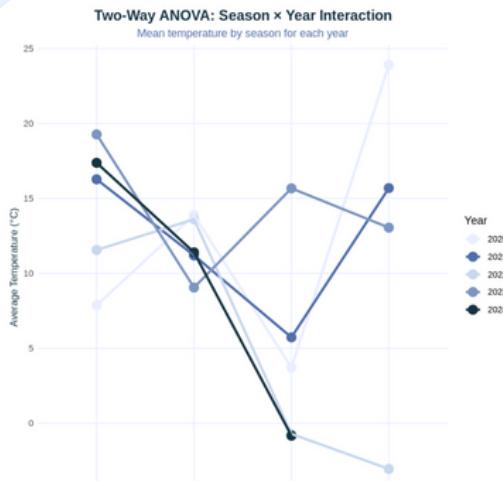
TWO-WAY ANOVA: Season × Year Interaction

Two-Way ANOVA Results:

	DF	Sum Sq	Mean Sq	F value	Pr(>F)
Season	3	720	240.05	1.852	0.156
Year_Factor	4	568	141.97	1.095	0.375
Season:Year_Factor	11	1032	93.84	0.724	0.708
Residuals	34	4407	129.62		

Effect Summary:

- Season effect: F = 1.85 , p = 0.1564 ✗
- Year effect: F = 1.1 , p = 0.3746 ✗
- Season×Year interaction: F = 0.72 , p = 0.708 ✗



Two-Way ANOVA

To ensure the reliability of the ANOVA test, the following assumptions were verified:

- Normality (Shapiro-Wilk Test):
 - $p\text{-value} = 0.1892$ (> 0.05).
- Result: The residuals follow a normal distribution (Assumption Met).
- Homogeneity of Variances (Levene's Test):
 - $p\text{-value} = 0.5252$ (> 0.05).
- Result: The variance across different seasons is equal (Assumption Met).

Conclusion: Since both key assumptions are met, the ANOVA result (indicating no significant seasonal difference) is statistically valid and reliable.



TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Post-hoc Analysis (Tukey's HSD)

Objective: To test if Season significantly impacts Avg_Temp.

Results:

- F-statistic: 1.958
- P-value: 0.133 (Greater than 0.05)
- Conclusion: Not Significant. The analysis fails to reject the null hypothesis.

Key Insight: The Box Plot reveals extreme volatility within individual seasons (especially Winter). This high internal variance overshadows the differences between seasonal means, indicating a highly unpredictable climate regardless of the season.

POST-HOC ANALYSIS: Tukey's HSD

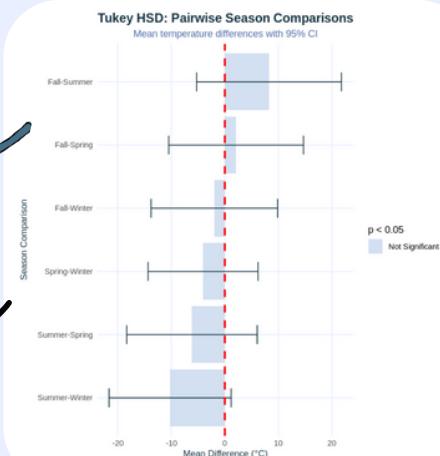
Pairwise Comparisons (Season):
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = Avg_Temp ~ Season, data = climate_final)

\$Season

	diff	lwr	upr	p adj
Spring-Winter	-4.072095	-14.333175	6.188985	0.717846
Summer-Winter	-10.214151	-21.618704	1.190402	0.0940598
Fall-Winter	-1.965826	-13.785225	9.853572	0.9707846
Summer-Spring	-6.142056	-16.334036	6.049923	0.5425811
Fall-Spring	2.106268	-10.474619	14.687156	0.9702333
Fall-Summer	8.248325	-5.281375	21.778025	0.3763808

Significant Differences ($p < 0.05$):
No significant pairwise differences found



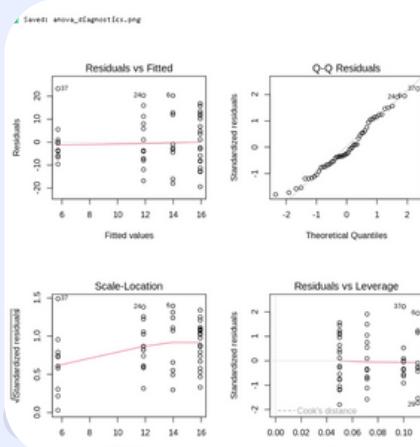
CHECKING ANOVA ASSUMPTIONS

Normality Test (Shapiro-Wilk):
• W = 0.9694
• p-value = 0.1892
• Residuals are normally distributed

Homogeneity of Variances (Levene's Test):
• F = 0.7543
• p-value = 0.5252
• Variances are homogeneous

RECOMMENDATION:
Assumptions met - ANOVA results are valid

Validation of Statistical Assumptions



To ensure the reliability of the ANOVA test, the following assumptions were verified:

- Normality (Shapiro-Wilk Test):
 - $p\text{-value} = 0.1892$ (> 0.05).
 - Result: The residuals follow a normal distribution (Assumption Met).
- Homogeneity of Variances (Levene's Test):
 - $p\text{-value} = 0.5252$ (> 0.05).
 - Result: The variance across different seasons is equal (Assumption Met).

Conclusion: Since both key assumptions are met, the ANOVA result (indicating no significant seasonal difference) is statistically valid and reliable.

TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Multivariate Analysis (MANOVA)

Objective: To determine if Season has a significant effect on the combined atmospheric profile (Temperature + Humidity + Wind Speed) simultaneously.

Multivariate Test Results:

- Pillai's Trace (Most Robust): $p\text{-value} = 0.3123$ (> 0.05) \rightarrow Not Significant.
- Wilks' Lambda: $p\text{-value} = 0.3088$ (> 0.05) \rightarrow Not Significant.
- Roy's Largest Root: $p\text{-value} = 0.0343$ (< 0.05) \rightarrow Significant (Outlier).
- Interpretation: While Roy's test suggests a potential difference, Pillai's Trace is the primary statistic for this data type. Since Pillai's test is not significant, we conclude that the overall climate profile does not fundamentally shift between seasons.

```
MANOVA: Multivariate Analysis of Variance
-----
MANOVA Results (Pillai's Trace):
  DF Pillai approx F num DF den DF Pr(>F)
Season   3 0.20204  1.1794    9   147 0.3123
Residuals 49

All Test Statistics:
  DF Wilks approx F num DF den DF Pr(>F)
Season   3 0.80465  1.1889    9 114.54 0.3088
Residuals 49
  DF Hotelling-Lawley approx F num DF den DF Pr(>F)
Season   3 0.23447  1.1897    9   137 0.3063
Residuals 49
  DF Roy approx F num DF den DF Pr(>F)
Season   3 0.19103  3.1202    3    49 0.0343 *
Residuals 49
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

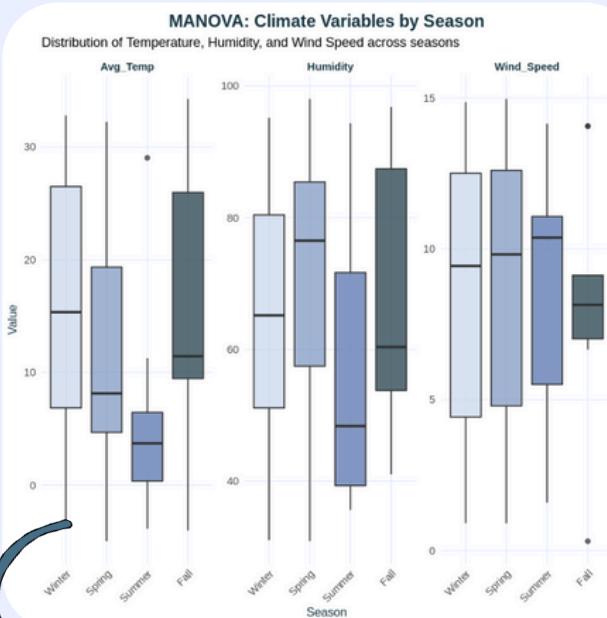
```
MANOVA: Univariate Follow-up ANOVAs
-----
Response Avg_Temp :
  DF Sum Sq Mean Sq F value Pr(>F)
Season   3 728.1 240.05  1.958 0.1326
Residuals 49 6007.3 122.60

Response Humidity :
  DF Sum Sq Mean Sq F value Pr(>F)
Season   3 1677.4 559.15  1.3423 0.2715
Residuals 49 20411.2 416.55

Response Wind_Speed :
  DF Sum Sq Mean Sq F value Pr(>F)
Season   3 0.36  0.120  0.0059 0.9994
Residuals 49 1003.58 20.481

Summary Table: Effect of Season on Each Variable
-----
* Avg_Temp : F = 1.96, p = 0.1326 ns
* Humidity : F = 1.34, p = 0.2715 ns
* Wind_Speed : F = 0.01, p = 0.9994 ns

Significance codes: *** p<0.001, ** p<0.01, * p<0.05, ns = not significant
```



Univariate Follow-up (Individual Variables):

- Avg_Temp: $p = 0.1326$ (NS)
- Humidity: $p = 0.2715$ (NS)
- Wind_Speed: $p = 0.9994$ (NS - Extremely random)
- Conclusion:

The analysis confirms that Seasonality does not significantly drive the climate variables in this dataset, neither individually nor as a combined group. The weather metrics (especially Wind Speed) appear to fluctuate independently of the traditional seasonal cycle.

Visual Analysis (MANOVA)

This visualization visually confirms why the statistical tests found no significant differences:

High Overlap: All three metrics display extensive overlapping ranges between seasons, making them statistically indistinguishable.

Temperature: The extreme vertical spread (variance) in Winter and Fall overshadows any mean differences.

Wind Speed: The medians are nearly identical across all seasons, explaining the perfect p-value of 0.99.

Conclusion: The internal variability within each season is far greater than the differences between the seasons.

TRACKING THE SHIFT: GLOBAL CLIMATE PATTERNS & ENVIRONMENTAL IMPACT

Standardized Seasonal Deviations (Heatmap)

Objective: To compare the relative intensity of climate variables across seasons using Z-scores (deviation from the global average).

- Key Observations:
- Inverted Pattern Confirmed: Summer shows strong negative deviations for Temperature (-1.39) and Humidity (-1.35), confirming it is the coldest and driest season. Conversely, Winter shows a positive Temperature deviation (+0.92).
- Wind Dynamics: Wind speed behaves inversely to other variables; it peaks in Summer (+0.96) and drops significantly in Fall (-1.32).
- Spring Anomaly: Spring is characterized by a distinct spike in Humidity (+1.01).

Conclusion: The standardized data clearly illustrates an "inverted" climate profile (Cold Summer / Warm Winter), strongly suggesting a Southern Hemisphere location.

MANOVA: Standardized Seasonal Means
Z-scores showing seasonal deviations from overall mean

