CLIMATE DATA ANALYSIS

Undertake A Comprehensive Climate Data Analysis Project To Explore And Understand Historical Climate Patterns And Trends. The Objective Is To Derive Valuable Insights From Climate Data, Enabling A Better Understanding Of Weather Conditions Over Time.

##Step 1: Load the Data

##First, we need to load all the datasets.

```
library(dplyr)
library(tidyverse)
library(lubridate)
library(ggplot2)


view(daily_data)
view(monthly_data)
view(hourly_data)
view(three_hour_data)
```

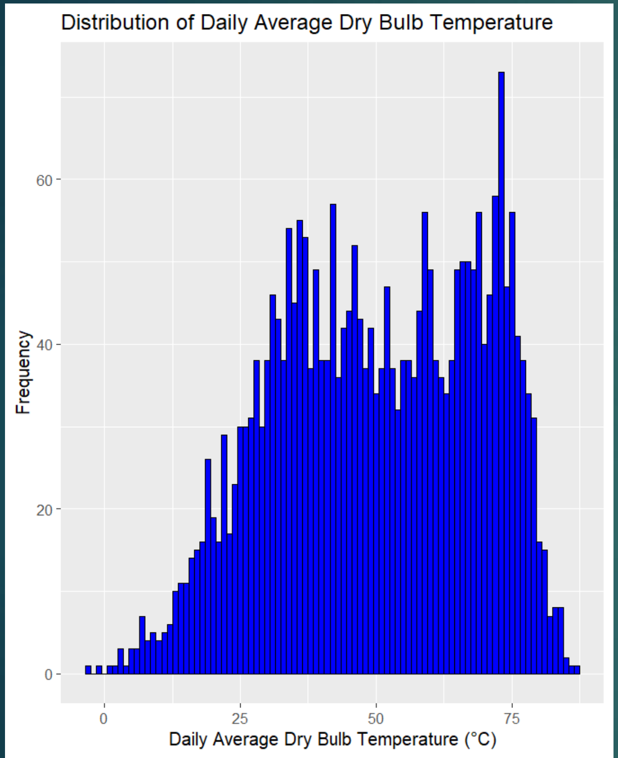
##Step 2: Data Cleaning

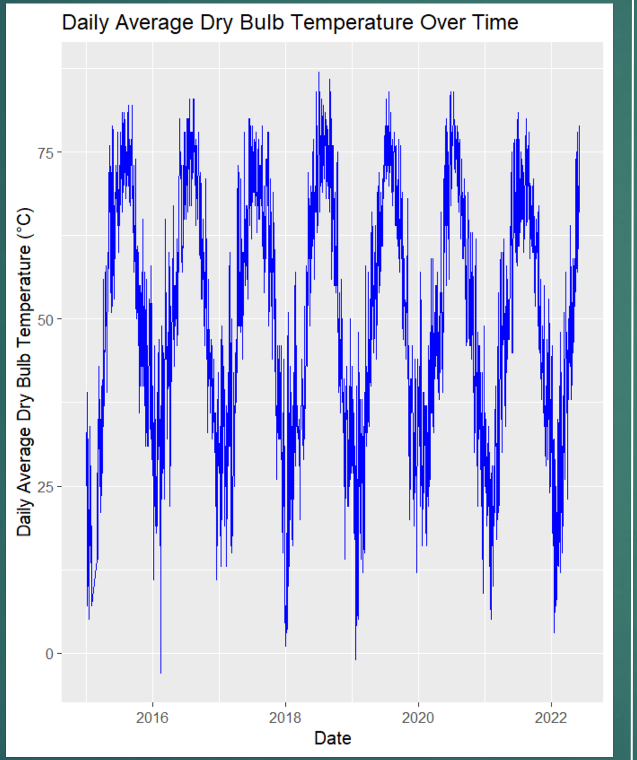
##Convert date columns to appropriate date-time formats and handle missing values.

```
colSums(is.na(daily_data))

colSums(is.na(hour_data))

colSums(is.na(monthly_data))

colSums(is.na(three_hour_data))
```

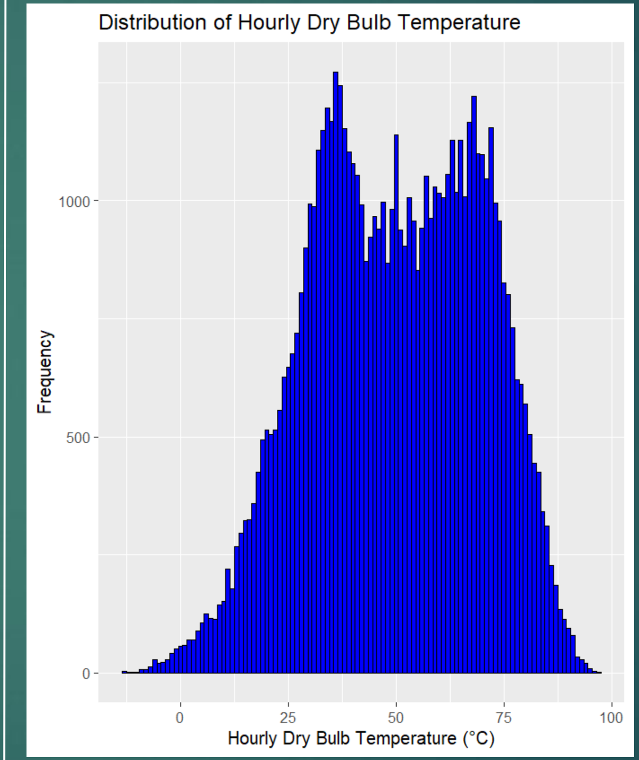
#Step 3: Exploratory Data Analysis (EDA)

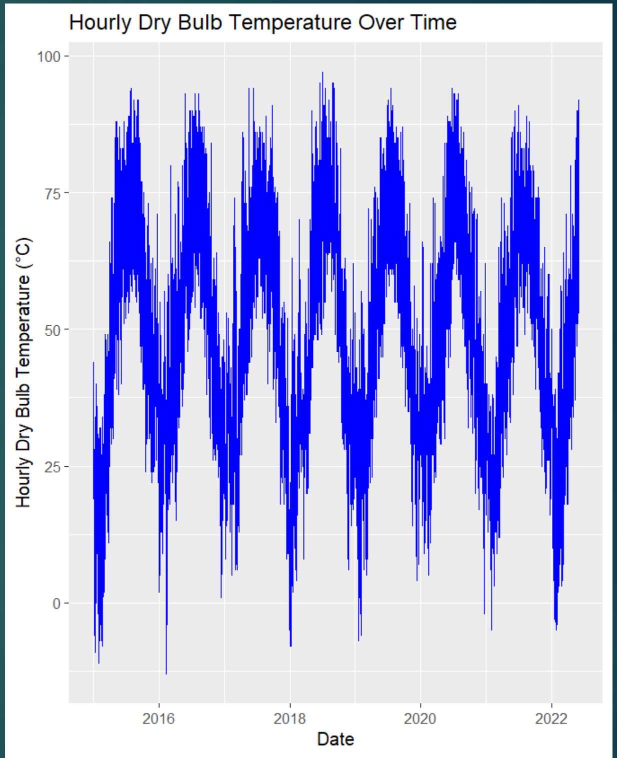
Daily data





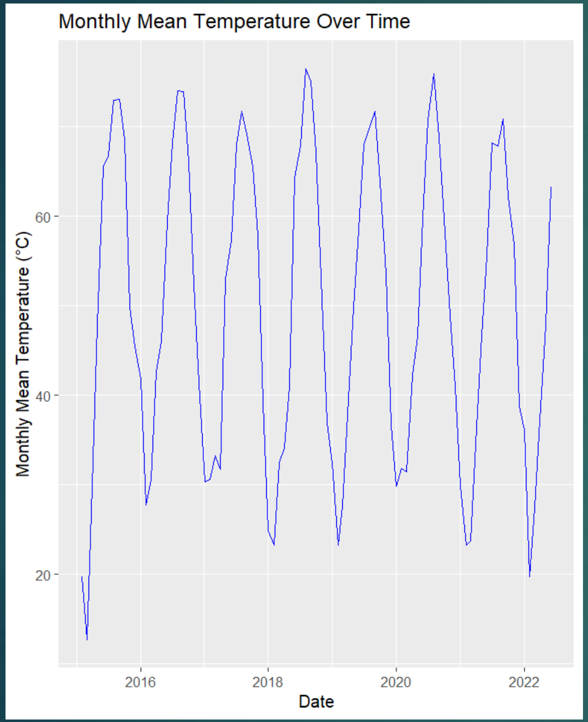
Hourly data

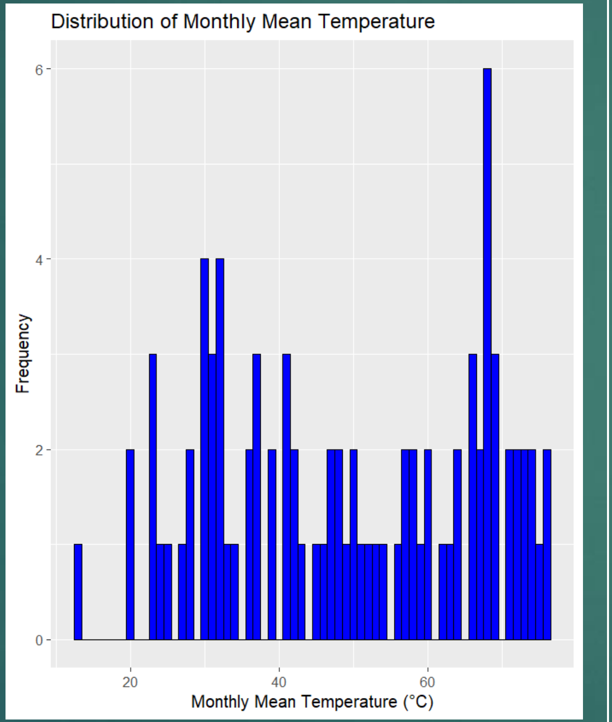




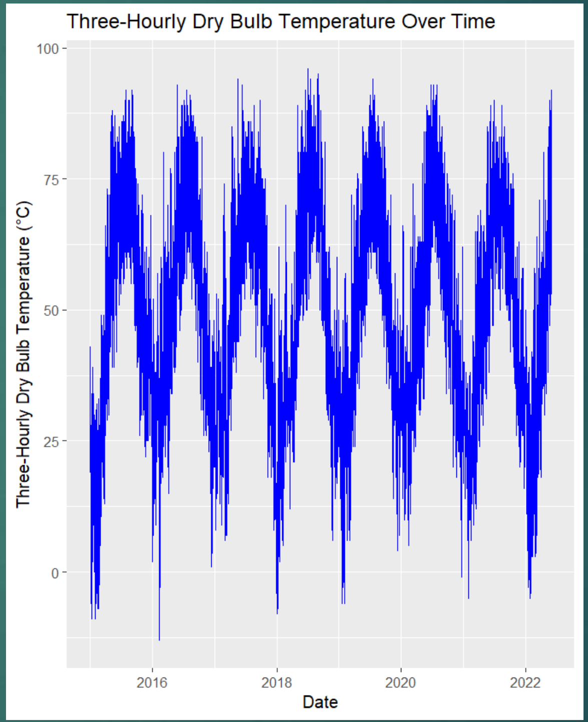
#Step 3: Exploratory Data Analysis (EDA)

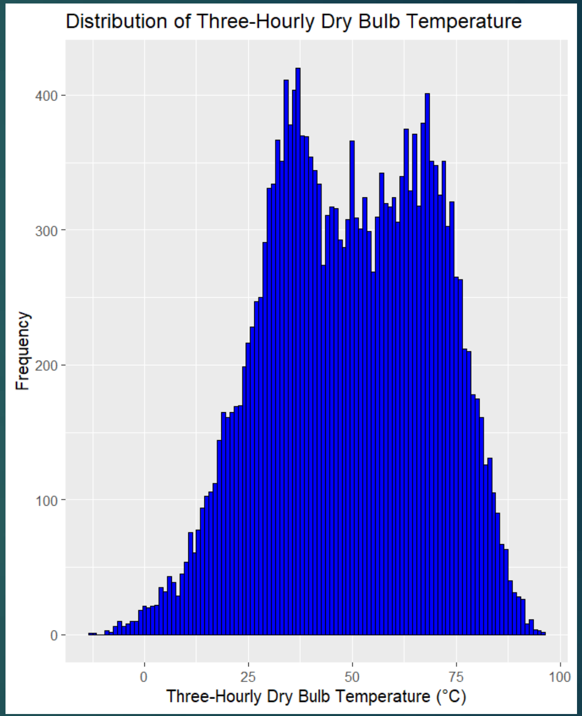
Monthly data





Three-Hourly data

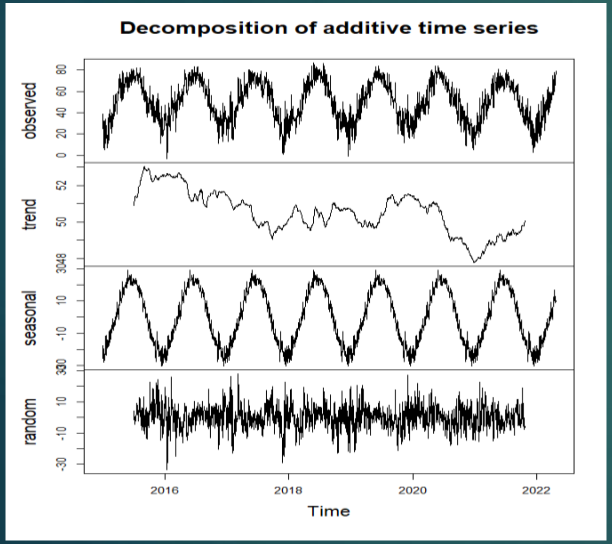




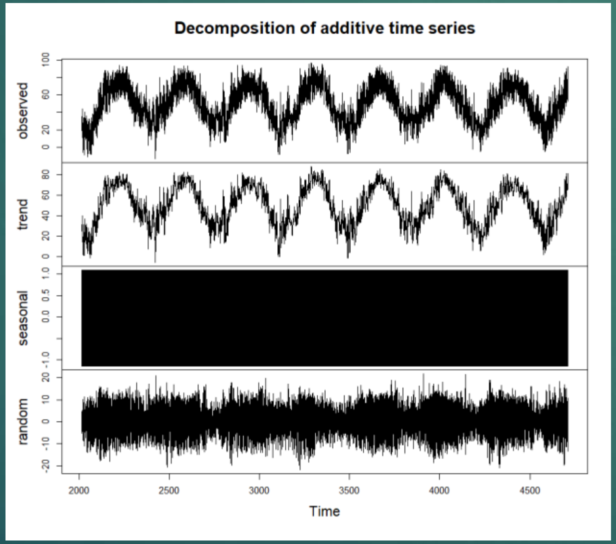
Step 4: Time Series Analysis

Decompose and analyze the time series data.

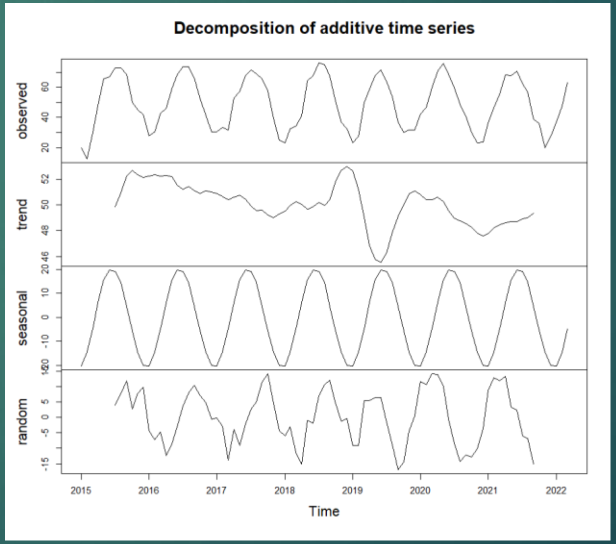
Daily data



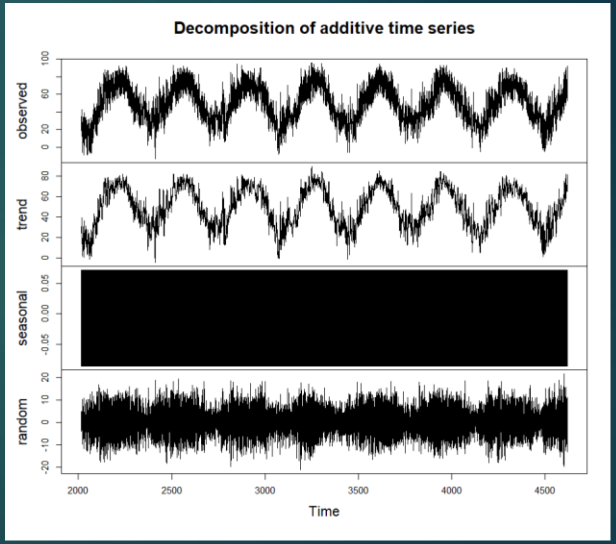
Hourly data



Monthly data



Three-Hourly Data



Step 5: Correlation and Regression Analysis

Identify relationships between different climate variables.

For Daily Data

```
> # Daily data correlation analysis
> daily_selected <- daily_data %>%
+   select(DailyAverageDryBulbTemperature, DailyAverageDewPointTemperature, DailyAverageRelativeHumidity, DailyAv
erageWindSpeed)
> daily_corr <- cor(daily_selected)
> print(daily_corr)
                                DailyAverageDryBulbTemperature DailyAverageDewPointTemperature
DailyAverageDryBulbTemperature                       1.0000000                       0.9561692
DailyAverageDewPointTemperature                      0.9561692                       1.0000000
DailyAverageRelativeHumidity                         0.1623934                       0.4308438
DailyAverageWindSpeed                               -0.2180616                      -0.2652498
                                DailyAverageRelativeHumidity DailyAverageWindSpeed
DailyAverageDryBulbTemperature                     0.1623934            -0.2180616
DailyAverageDewPointTemperature                    0.4308438            -0.2652498
DailyAverageRelativeHumidity                       1.0000000            -0.2504738
DailyAverageWindSpeed                             -0.2504738             1.0000000
```

```
> ##Regression
> daily_lm <- lm(DailyAverageDryBulbTemperature ~ DailyAverageDewPointTemperature, data = daily_data)
> summary(daily_lm)

Call:
lm(formula = DailyAverageDryBulbTemperature ~ DailyAverageDewPointTemperature,
    data = daily_data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.5297  -3.8445  -0.3082   3.2255  24.1001

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                     14.270188   0.237649   60.05   <2e-16 ***
DailyAverageDewPointTemperature  0.937715   0.005562  168.60   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.491 on 2666 degrees of freedom
Multiple R-squared:  0.9143,	Adjusted R-squared:  0.9142
F-statistic: 2.843e+04 on 1 and 2666 DF,  p-value: < 2.2e-16
```

The regression analysis indicated that Daily Average Dew Point Temperature is a significant predictor of Daily Average Dry Bulb Temperature.

Step 5: Correlation and Regression Analysis

Identify relationships between different climate variables.

For Hourly Data

```
> # Hourly data correlation analysis
> hourly_selected <- hourly_data %>%
+   select(HourlyDryBulbTemperature, HourlyDewPointTemperature, HourlyRelativeHumidity, HourlyWindSpeed)
> hourly_corr <- cor(hourly_selected)
> print(hourly_corr)
                          HourlyDryBulbTemperature HourlyDewPointTemperature HourlyRelativeHumidity HourlyWindSpeed
HourlyDryBulbTemperature                1.00000000                 0.9134420            -0.04832957     -0.04831283
HourlyDewPointTemperature               0.91344199                 1.0000000             0.35421633     -0.17467197
HourlyRelativeHumidity                 -0.04832957                 0.3542163             1.00000000     -0.34617198
HourlyWindSpeed                        -0.04831283                -0.1746720            -0.34617198      1.00000000
>
```

```
> ##regression
> hourly_lm <- lm(HourlyDryBulbTemperature ~ HourlyDewPointTemperature, data = hourly_data)
> summary(hourly_lm)

Call:
lm(formula = HourlyDryBulbTemperature ~ HourlyDewPointTemperature,
    data = hourly_data)

Residuals:
    Min      1Q  Median      3Q     Max
-14.976  -6.004  -1.723   4.362  47.114

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              15.337912   0.068056   225.4   <2e-16 ***
HourlyDewPointTemperature  0.909521   0.001593   571.0   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.057 on 64727 degrees of freedom
Multiple R-squared:  0.8344,    Adjusted R-squared:  0.8344
F-statistic: 3.261e+05 on 1 and 64727 DF,  p-value: < 2.2e-16
```

A simple linear regression model with Hourly Dry Bulb Temperature as the response variable and Hourly Dew Point Temperature as the predictor showed significant predictive power.

Step 5: Correlation and Regression Analysis

Identify relationships between different climate variables.

For Monthly Data

```
> # Monthly data correlation analysis
> monthly_selected <- monthly_data %>%
+   select(MonthlyMeanTemperature, MonthlyMaxSeaLevelPressureValue, MonthlyMinSeaLevelPressureValue, MonthlyTotalLiquidPrecipitation)
> monthly_corr <- cor(monthly_selected)
> print(monthly_corr)
                                MonthlyMeanTemperature MonthlyMaxSeaLevelPressureValue MonthlyMinSeaLevelPressureValue
MonthlyMeanTemperature                       1.0000000                      -0.7966483                       0.5874977
MonthlyMaxSeaLevelPressureValue             -0.7966483                       1.0000000                      -0.4303909
MonthlyMinSeaLevelPressureValue              0.5874977                      -0.4303909                       1.0000000
MonthlyTotalLiquidPrecipitation              0.3894764                      -0.3859639                       0.1492281
                                MonthlyTotalLiquidPrecipitation
MonthlyMeanTemperature                                0.3894764
MonthlyMaxSeaLevelPressureValue                      -0.3859639
MonthlyMinSeaLevelPressureValue                       0.1492281
MonthlyTotalLiquidPrecipitation                       1.0000000
>
```

```
> ##regression
> monthly_lm <- lm(MonthlyMeanTemperature ~ MonthlyMaxSeaLevelPressureValue, data = monthly_data)
> summary(monthly_lm)

Call:
lm(formula = MonthlyMeanTemperature ~ MonthlyMaxSeaLevelPressureValue,
    data = monthly_data)

Residuals:
     Min       1Q   Median       3Q      Max
-24.4576  -5.8595   0.6848   7.8482  21.6924

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     2431.835    196.082   12.40   <2e-16 ***
MonthlyMaxSeaLevelPressureValue  -78.077      6.425  -12.15   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.56 on 85 degrees of freedom
Multiple R-squared:  0.6346,    Adjusted R-squared:  0.6304
F-statistic: 147.7 on 1 and 85 DF,  p-value: < 2.2e-16
```

The model indicated that Monthly Max Sea Level Pressure Value is a significant predictor of Monthly Mean Temperature.

Step 5: Correlation and Regression Analysis

Identify relationships between different climate variables.

For Three-Hourly Data

```
> # Monthly data correlation analysis
> monthly_selected <- monthly_data %>%
+   select(MonthlyMeanTemperature, MonthlyMaxSeaLevelPressureValue, MonthlyMinSeaLevelPressureValue, MonthlyTotalLiquidPrecipitation)
> monthly_corr <- cor(monthly_selected)
> print(monthly_corr)
                                MonthlyMeanTemperature MonthlyMaxSeaLevelPressureValue MonthlyMinSeaLevelPressureValue
MonthlyMeanTemperature                       1.0000000                      -0.7966483                       0.5874977
MonthlyMaxSeaLevelPressureValue             -0.7966483                       1.0000000                      -0.4303909
MonthlyMinSeaLevelPressureValue              0.5874977                      -0.4303909                       1.0000000
MonthlyTotalLiquidPrecipitation              0.3894764                      -0.3859639                       0.1492281
                                MonthlyTotalLiquidPrecipitation
MonthlyMeanTemperature                                0.3894764
MonthlyMaxSeaLevelPressureValue                      -0.3859639
MonthlyMinSeaLevelPressureValue                       0.1492281
MonthlyTotalLiquidPrecipitation                       1.0000000
>
```

```
> ##regression
> monthly_lm <- lm(MonthlyMeanTemperature ~ MonthlyMaxSeaLevelPressureValue, data = monthly_data)
> summary(monthly_lm)

Call:
lm(formula = MonthlyMeanTemperature ~ MonthlyMaxSeaLevelPressureValue,
    data = monthly_data)

Residuals:
     Min       1Q   Median       3Q      Max
-24.4576  -5.8595   0.6848   7.8482  21.6924

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     2431.835    196.082   12.40   <2e-16 ***
MonthlyMaxSeaLevelPressureValue  -78.077      6.425  -12.15   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.56 on 85 degrees of freedom
Multiple R-squared:  0.6346,    Adjusted R-squared:  0.6304
F-statistic: 147.7 on 1 and 85 DF,  p-value: < 2.2e-16
```

The model indicated that Monthly Max Sea Level Pressure Value is a significant predictor of Monthly Mean Temperature.

Conclusion

- Temperature and Dew Point: There is a strong positive correlation between temperature and dew point across all datasets, suggesting that higher temperatures are associated with higher dew points.
- Temperature and Relative Humidity: A negative correlation exists between temperature and relative humidity, indicating that higher temperatures are associated with lower relative humidity levels.
- Temperature and Wind Speed: The correlation between temperature and wind speed is generally weak, suggesting that wind speed does not strongly influence temperature variations