

# 65 Years Weather Data of Bangladesh

Jawwad Ahmed Ornob

2024-08-29

```
## Here I'm downloading all the packages
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
library(lubridate)
```

```
##
```

```
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## date, intersect, setdiff, union
```

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
```

```
## method from
```

```
## as.zoo.data.frame zoo
```

```
library(tidyr)
```

```
## Loading Dataset
```

```
weather_data <- read.csv("D:/R Programming/bd_weather.csv")
```

```
## Exploring the Data
```

```
head(weather_data)
```

```
##   X Station.Names YEAR Month Max.Temp Min.Temp Rainfall Relative.Humidity
## 1 0      Barisal 1949   1    29.4    12.3      0          68
## 2 1      Barisal 1950   1    30.0    14.1      0          77
## 3 2      Barisal 1951   1    28.2    12.3      0          77
## 4 3      Barisal 1952   1    26.6    12.3      2          77
## 5 4      Barisal 1953   1    30.0    13.3     10          75
## 6 5      Barisal 1954   1    27.8    12.7      0          72
##   Wind.Speed Cloud.Coverage Bright.Sunshine Station.Number   X_COR   Y_COR
## 1 0.4537037          0.6      7.831915      41950 536809.8 510151.9
## 2 0.4537037          0.8      7.831915      41950 536809.8 510151.9
## 3 0.4537037          0.6      7.831915      41950 536809.8 510151.9
## 4 0.4537037          1.0      7.831915      41950 536809.8 510151.9
## 5 0.4537037          1.6      7.831915      41950 536809.8 510151.9
## 6 0.4537037          0.5      7.831915      41950 536809.8 510151.9
##   LATITUDE LONGITUDE ALT   Period
## 1    22.7     90.36   4 1949.01
## 2    22.7     90.36   4 1950.01
## 3    22.7     90.36   4 1951.01
## 4    22.7     90.36   4 1952.01
## 5    22.7     90.36   4 1953.01
## 6    22.7     90.36   4 1954.01
```

```
str(weather_data)
```

```
## 'data.frame':    21120 obs. of  18 variables:
##  $ X                : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ Station.Names     : chr   "Barisal" "Barisal" "Barisal" "Barisal" ...
##  $ YEAR              : int  1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 ...
##  $ Month             : int   1 1 1 1 1 1 1 1 1 1 ...
##  $ Max.Temp          : num  29.4 30 28.2 26.6 30 27.8 26.6 29.4 30.1 31.1 ...
##  $ Min.Temp          : num  12.3 14.1 12.3 12.3 13.3 12.7 12.3 14.3 15.1 15.5 ...
##  $ Rainfall          : num   0 0 0 2 10 0 2 17 104 0 ...
##  $ Relative.Humidity: num   68 77 77 77 75 72 77 74 80 77 ...
##  $ Wind.Speed        : num   0.454 0.454 0.454 0.454 0.454 ...
##  $ Cloud.Coverage    : num   0.6 0.8 0.6 1 1.6 0.5 1 0.4 1 1.7 ...
##  $ Bright.Sunshine   : num   7.83 7.83 7.83 7.83 7.83 ...
##  $ Station.Number    : int  41950 41950 41950 41950 41950 41950 41950 41950 41950 41950 ...
##  $ X_COR             : num  536810 536810 536810 536810 536810 ...
##  $ Y_COR             : num  510152 510152 510152 510152 510152 ...
##  $ LATITUDE          : num  22.7 22.7 22.7 22.7 22.7 22.7 22.7 22.7 22.7 22.7 ...
##  $ LONGITUDE         : num  90.4 90.4 90.4 90.4 90.4 ...
##  $ ALT              : int   4 4 4 4 4 4 4 4 4 4 ...
##  $ Period            : num  1949 1950 1951 1952 1953 ...
```

```
summary(weather_data)
```

```
##           X           Station.Names           YEAR           Month
```

```
## Min. : 0 Length:21120 Min. :1948 Min. : 1.00
## 1st Qu.: 5280 Class :character 1st Qu.:1972 1st Qu.: 3.75
## Median :10560 Mode :character Median :1988 Median : 6.50
## Mean :10560 Mean :1986 Mean : 6.50
## 3rd Qu.:15839 3rd Qu.:2001 3rd Qu.: 9.25
## Max. :21119 Max. :2013 Max. :12.00
## Max.Temp Min.Temp Rainfall Relative.Humidity
## Min. :21.60 Min. : 6.20 Min. : 0.0 Min. :34.00
## 1st Qu.:31.70 1st Qu.:16.90 1st Qu.: 8.0 1st Qu.:75.00
## Median :33.90 Median :23.40 Median :110.0 Median :81.00
## Mean :33.48 Mean :21.15 Mean :197.5 Mean :79.46
## 3rd Qu.:35.50 3rd Qu.:25.40 3rd Qu.:310.0 3rd Qu.:85.00
## Max. :44.00 Max. :28.10 Max. :2072.0 Max. :97.00
## Wind.Speed Cloud.Coverage Bright.Sunshine Station.Number
## Min. : 0.000 Min. :0.000 Min. : 0.000 Min. :41858
## 1st Qu.: 0.700 1st Qu.:1.500 1st Qu.: 5.000 1st Qu.:41907
## Median : 1.200 Median :3.300 Median : 6.783 Median :41941
## Mean : 1.415 Mean :3.479 Mean : 6.411 Mean :41934
## 3rd Qu.: 1.900 3rd Qu.:5.500 3rd Qu.: 7.800 3rd Qu.:41960
## Max. :11.200 Max. :7.900 Max. :11.000 Max. :41998
## X_COR Y_COR LATITUDE LONGITUDE
## Min. : 0 Min. : 0 Min. :20.87 Min. :88.56
## 1st Qu.:435304 1st Qu.:499111 1st Qu.:22.64 1st Qu.:89.36
## Median :540099 Median :544955 Median :23.17 Median :90.39
## Mean :534711 Mean :563482 Mean :23.36 Mean :90.45
## 3rd Qu.:650012 3rd Qu.:683166 3rd Qu.:24.29 3rd Qu.:91.46
## Max. :734765 Max. :844822 Max. :25.75 Max. :92.26
## ALT Period
## Min. : 0.00 Min. :1948
## 1st Qu.: 4.00 1st Qu.:1972
## Median : 7.00 Median :1988
## Mean :12.99 Mean :1986
## 3rd Qu.:19.00 3rd Qu.:2001
## Max. :63.00 Max. :2013
```

```
colnames(weather_data)
```

```
## [1] "X" "Station.Names" "YEAR"
## [4] "Month" "Max.Temp" "Min.Temp"
## [7] "Rainfall" "Relative.Humidity" "Wind.Speed"
## [10] "Cloud.Coverage" "Bright.Sunshine" "Station.Number"
## [13] "X_COR" "Y_COR" "LATITUDE"
## [16] "LONGITUDE" "ALT" "Period"
```

```
## Checking for Missing Values
```

```
sum(is.na(weather_data))
```

```
## [1] 0
```

```
## Basic Statistics for Rainfall
```

```
rainfall_data <- summary(weather_data$Rainfall)
mean_rainfall <- mean(rainfall_data)
median_rainfall <- median(rainfall_data)
standard_deviation_of_rainfall <- sd(rainfall_data)
total_rainfall <- sum(rainfall_data)

print(mean_rainfall)
```

```
## [1] 449.5883
```

```
print(median_rainfall)
```

```
## [1] 153.7648
```

```
print(standard_deviation_of_rainfall)
```

```
## [1] 803.4536
```

```
print(total_rainfall)
```

```
## [1] 2697.53
```

```
## Convert month numbers to month names
```

```
weather_data$Month <- month.name[weather_data$Month]
head(weather_data)
```

```
##   X Station.Names YEAR   Month Max.Temp Min.Temp Rainfall Relative.Humidity
## 1 0      Barisal 1949 January   29.4    12.3      0             68
## 2 1      Barisal 1950 January   30.0    14.1      0             77
## 3 2      Barisal 1951 January   28.2    12.3      0             77
## 4 3      Barisal 1952 January   26.6    12.3      2             77
## 5 4      Barisal 1953 January   30.0    13.3     10             75
## 6 5      Barisal 1954 January   27.8    12.7      0             72
##   Wind.Speed Cloud.Coverage Bright.Sunshine Station.Number   X_COR   Y_COR
## 1 0.4537037           0.6      7.831915      41950 536809.8 510151.9
## 2 0.4537037           0.8      7.831915      41950 536809.8 510151.9
## 3 0.4537037           0.6      7.831915      41950 536809.8 510151.9
## 4 0.4537037           1.0      7.831915      41950 536809.8 510151.9
## 5 0.4537037           1.6      7.831915      41950 536809.8 510151.9
## 6 0.4537037           0.5      7.831915      41950 536809.8 510151.9
##   LATITUDE LONGITUDE ALT   Period
## 1    22.7     90.36   4 1949.01
## 2    22.7     90.36   4 1950.01
## 3    22.7     90.36   4 1951.01
## 4    22.7     90.36   4 1952.01
## 5    22.7     90.36   4 1953.01
## 6    22.7     90.36   4 1954.01
```

```
## Here Starts Anaysis
```

```
## Average Rainfall by Year
```

```
avg_rainfall_by_year <- weather_data %>%  
  group_by(YEAR) %>%  
  summarize(avg_rainfall = mean(Rainfall, na.rm = TRUE)) %>%  
  arrange(desc(avg_rainfall))  
print(avg_rainfall_by_year)
```

```
## # A tibble: 66 x 2  
##   YEAR avg_rainfall  
##   <int>      <dbl>  
## 1  1983        236.  
## 2  1959        231.  
## 3  2007        230.  
## 4  1991        229.  
## 5  1998        227.  
## 6  1984        227.  
## 7  2004        226.  
## 8  1988        224.  
## 9  1993        222.  
## 10 1987        221.  
## # i 56 more rows
```

```
# Finding out the highest average rainy year
```

```
highest_average_rainy_year <- head(avg_rainfall_by_year, n=1)  
highest_average_rainy_year
```

```
## # A tibble: 1 x 2  
##   YEAR avg_rainfall  
##   <int>      <dbl>  
## 1  1983        236.
```

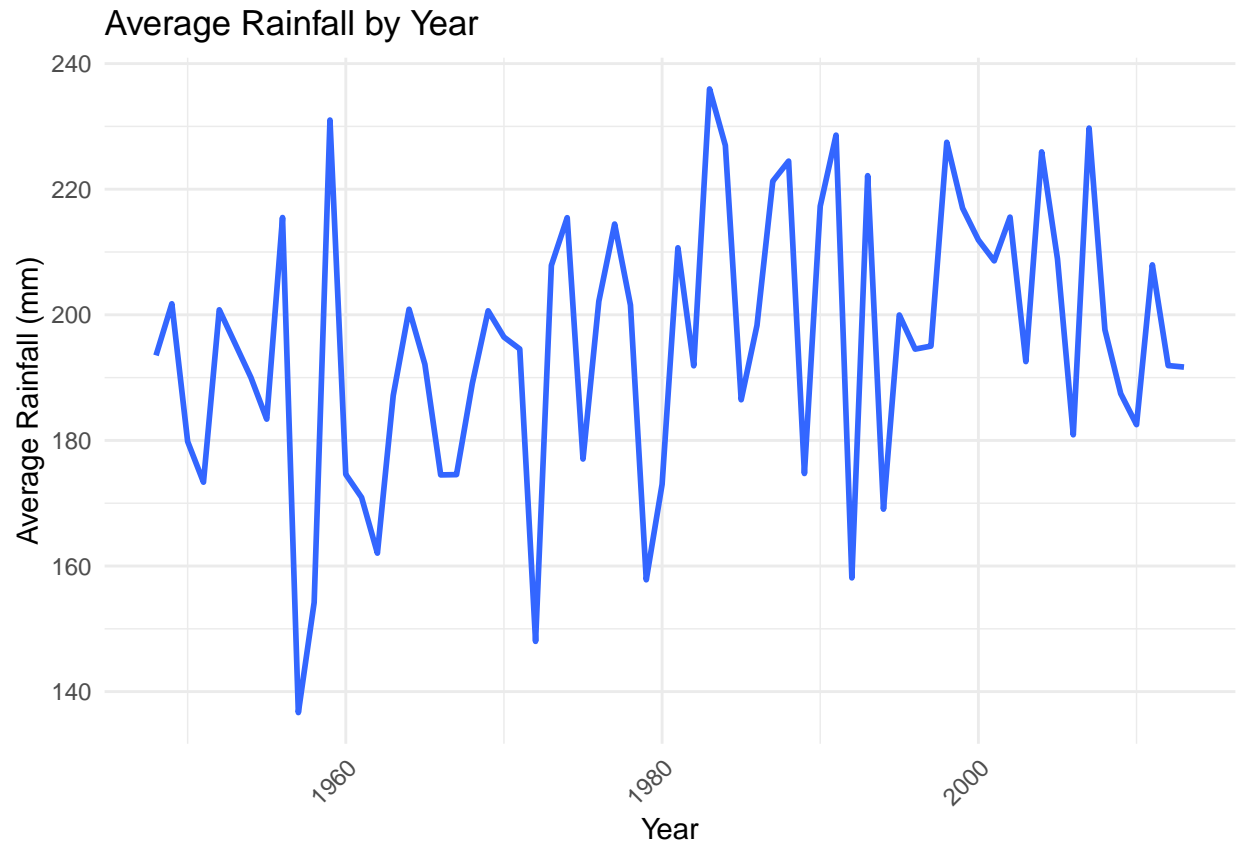
```
# Finding out the lowest average rainy year
```

```
lowest_average_rainy_year <- tail(avg_rainfall_by_year, n=1)  
lowest_average_rainy_year
```

```
## # A tibble: 1 x 2  
##   YEAR avg_rainfall  
##   <int>      <dbl>  
## 1  1957        137.
```

```
# Displaying Average Rainfall in each Year
```

```
ggplot(data = avg_rainfall_by_year, aes(x = YEAR, y = avg_rainfall)) +  
  geom_smooth(stat = "identity", fill = "skyblue") +  
  labs(title = "Average Rainfall by Year", x = "Year", y = "Average Rainfall (mm)") +  
  theme_minimal() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
## Average rainfall by month
avg_rainfall_by_month <- weather_data %>%
  group_by(Month) %>%
  summarize(avg_rainfall = mean(Rainfall, na.rm = TRUE)) %>%
  arrange(desc(avg_rainfall))
print(avg_rainfall_by_month)
```

```
## # A tibble: 12 x 2
##   Month      avg_rainfall
##   <chr>         <dbl>
## 1 July          515.
## 2 June          459.
## 3 August        411.
## 4 September    314.
## 5 May           262.
## 6 October      184.
## 7 April         111.
## 8 March          45.1
## 9 November      34.3
## 10 February     18.9
## 11 December       7.65
## 12 January       7.61
```

```
# Highest Average Rainfall By Month
```

```
highest_average_rainy_month <- head(avg_rainfall_by_month, n = 1)
highest_average_rainy_month
```

```
## # A tibble: 1 x 2
##   Month avg_rainfall
##   <chr>      <dbl>
## 1 July        515.
```

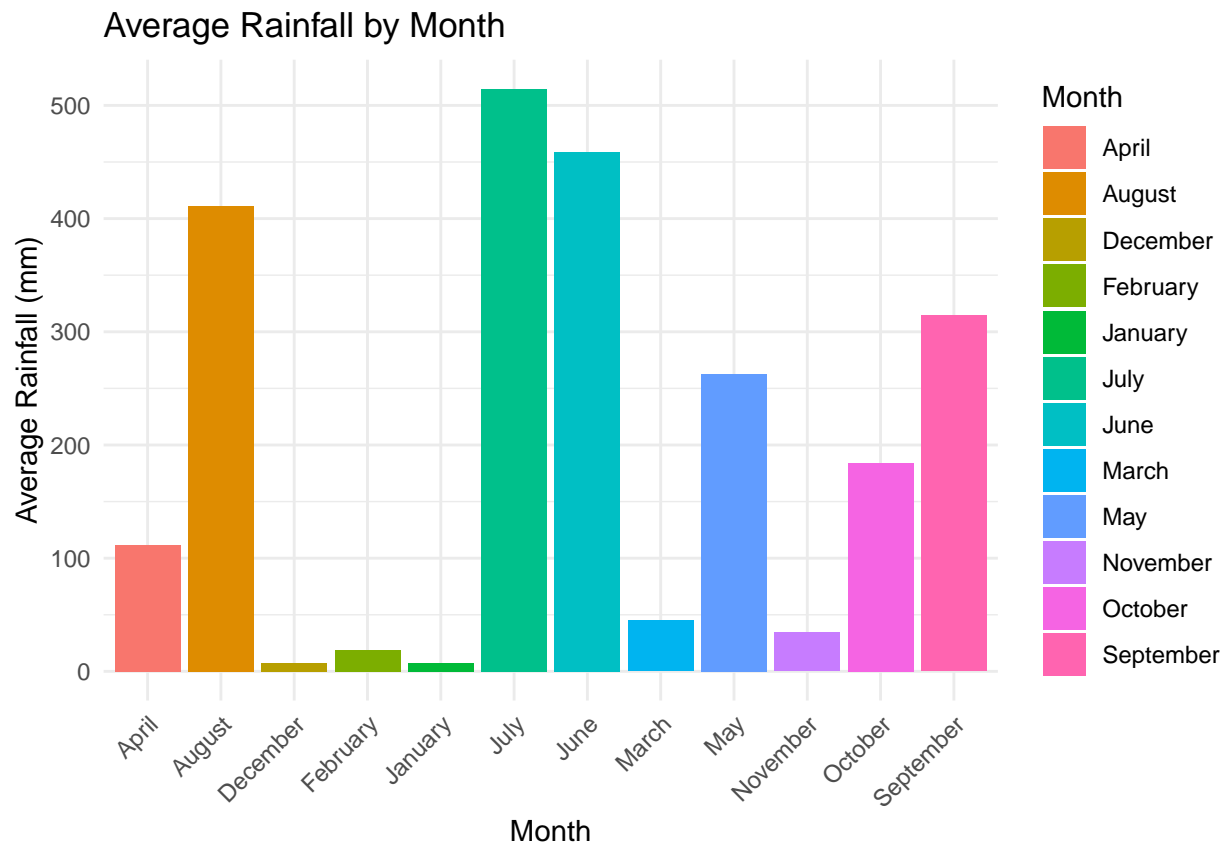
```
# Lowest Average Rainfall By Month
```

```
lowest_average_rainy_month<- tail(avg_rainfall_by_month, n=1)
lowest_average_rainy_month
```

```
## # A tibble: 1 x 2
##   Month  avg_rainfall
##   <chr>      <dbl>
## 1 January      7.61
```

```
# Displaying the average rainfall by months
```

```
ggplot(avg_rainfall_by_month, aes(x = Month, y = avg_rainfall, fill = Month)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Rainfall by Month", x = "Month", y = "Average Rainfall (mm)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
## Average Rainfall According to Station Names (Location)
```

```
avg_rainfall_by_stations <- weather_data %>%  
  group_by(Station.Names) %>%  
  summarise(avg_rainfall=mean(Rainfall, na.rm =TRUE)) %>%  
  arrange(desc(avg_rainfall))  
  
print(avg_rainfall_by_stations)
```

```
## # A tibble: 35 x 2  
##   Station.Names      avg_rainfall  
##   <chr>            <dbl>  
## 1 Teknaf          342.  
## 2 Sylhet          334.  
## 3 Cox's Bazar     303.  
## 4 Sandwip         296.  
## 5 Hatiya          261.  
## 6 Maijdee Court   258.  
## 7 Feni            248.  
## 8 Sitakunda       248.  
## 9 Chittagong (City-Ambagan) 244.  
## 10 Chittagong (IAP-Patenga) 241.  
## # i 25 more rows
```

```
# Highest Rainfall by Station
```

```
highest_average_rainfall_by_station <- head(avg_rainfall_by_stations, n=1)  
highest_average_rainfall_by_station
```

```
## # A tibble: 1 x 2  
##   Station.Names avg_rainfall  
##   <chr>         <dbl>  
## 1 Teknaf       342.
```

```
#Lowest Rainfall By Station
```

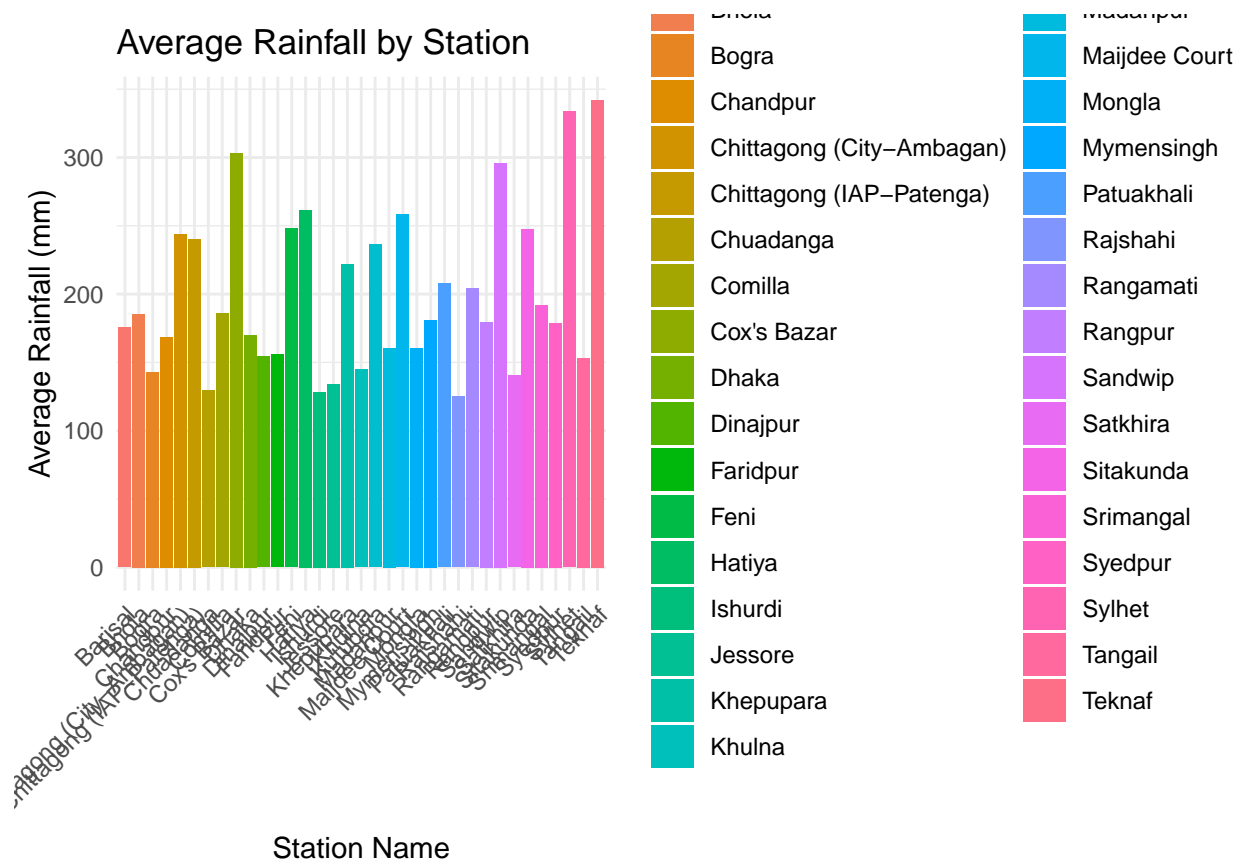
```
lowest_average_rainfall_by_station <- tail(avg_rainfall_by_stations, n=1)  
lowest_average_rainfall_by_station
```

```
## # A tibble: 1 x 2  
##   Station.Names avg_rainfall  
##   <chr>         <dbl>  
## 1 Rajshahi     125.
```

```
#Displaying Rainfall by Stations (Locations) in Bar Chart
```

```
ggplot(data = avg_rainfall_by_stations) +  
  geom_bar(mapping = aes(x = Station.Names, y = avg_rainfall, fill =Station.Names), stat = "identity") +  
  labs(title = "Average Rainfall by Station",  
        x = "Station Name",  
        y = "Average Rainfall (mm)") +  
  theme_minimal() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

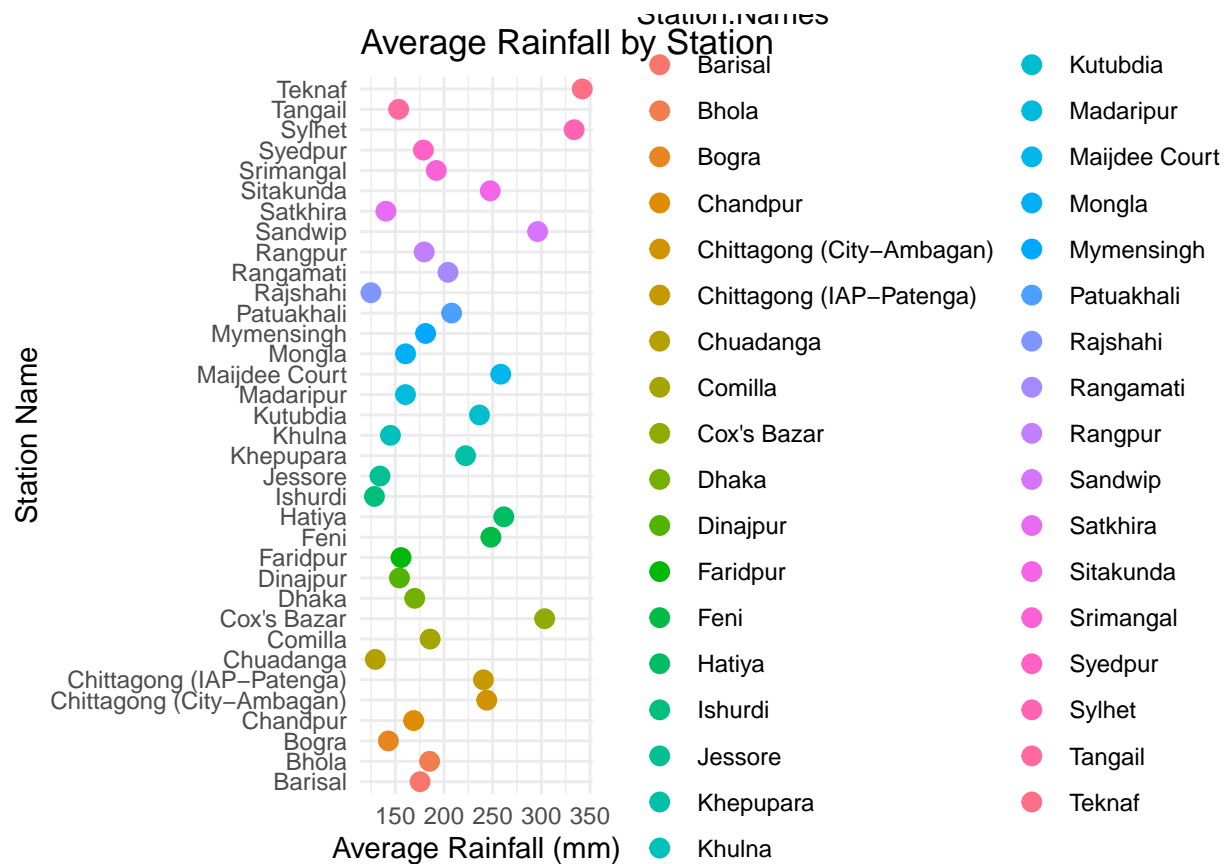




*#Displaying Rainfall by Stations (Locations) in Dot Plot*

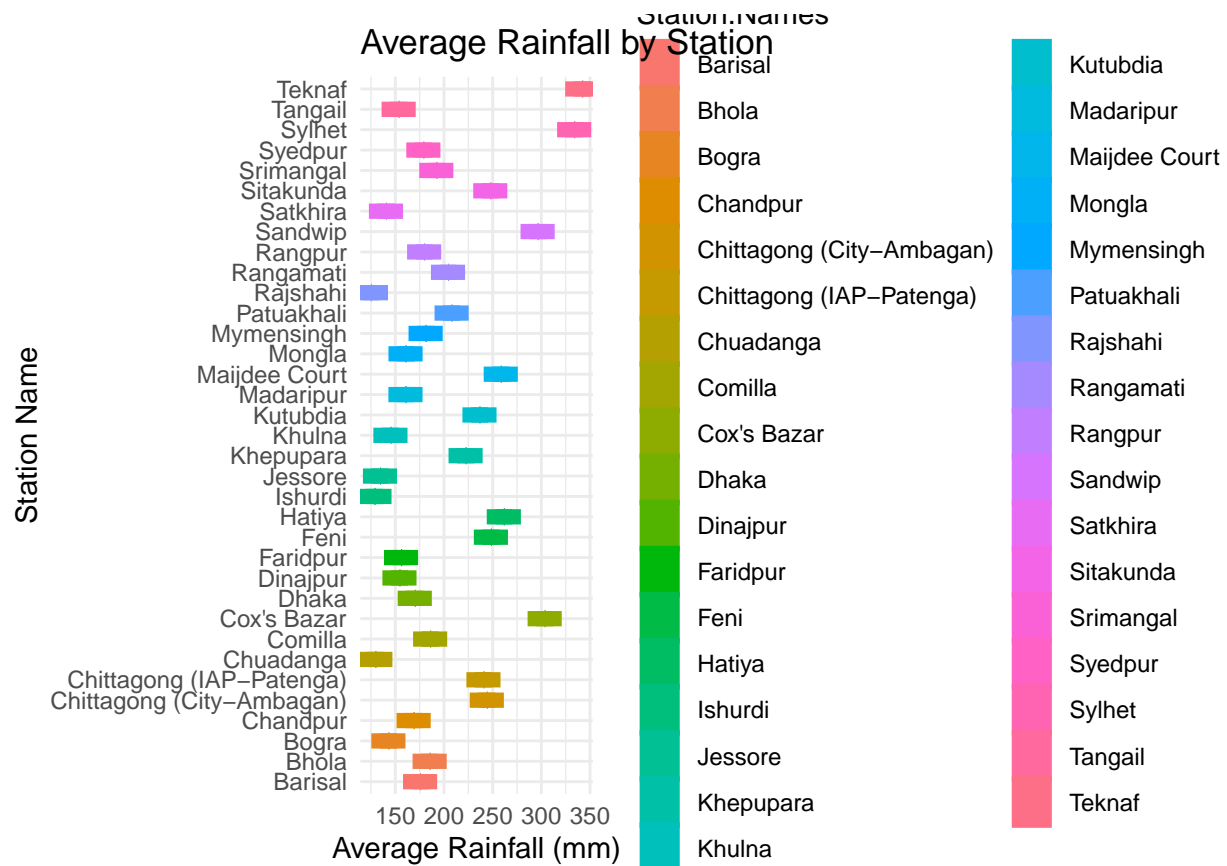
*# Create a dot plot of total rainfall by station*

```
ggplot(data = avg_rainfall_by_stations, aes(x = avg_rainfall, y = Station.Names)) +
  geom_point(aes(color = Station.Names), size = 3) +
  labs(title = "Average Rainfall by Station",
       x = "Average Rainfall (mm)",
       y = "Station Name") +
  theme_minimal()
```



*#Displaying Rainfall by Stations (Locations) in Box Plot*

```
ggplot(data = avg_rainfall_by_stations, aes(x = avg_rainfall, y = Station.Names)) +
  geom_boxplot(aes(color = Station.Names), size = 3) +
  labs(title = "Average Rainfall by Station",
       x = "Average Rainfall (mm)",
       y = "Station Name") +
  theme_minimal()
```



*#Plotting Maximum & Minimum Temperature Trends over the Years*

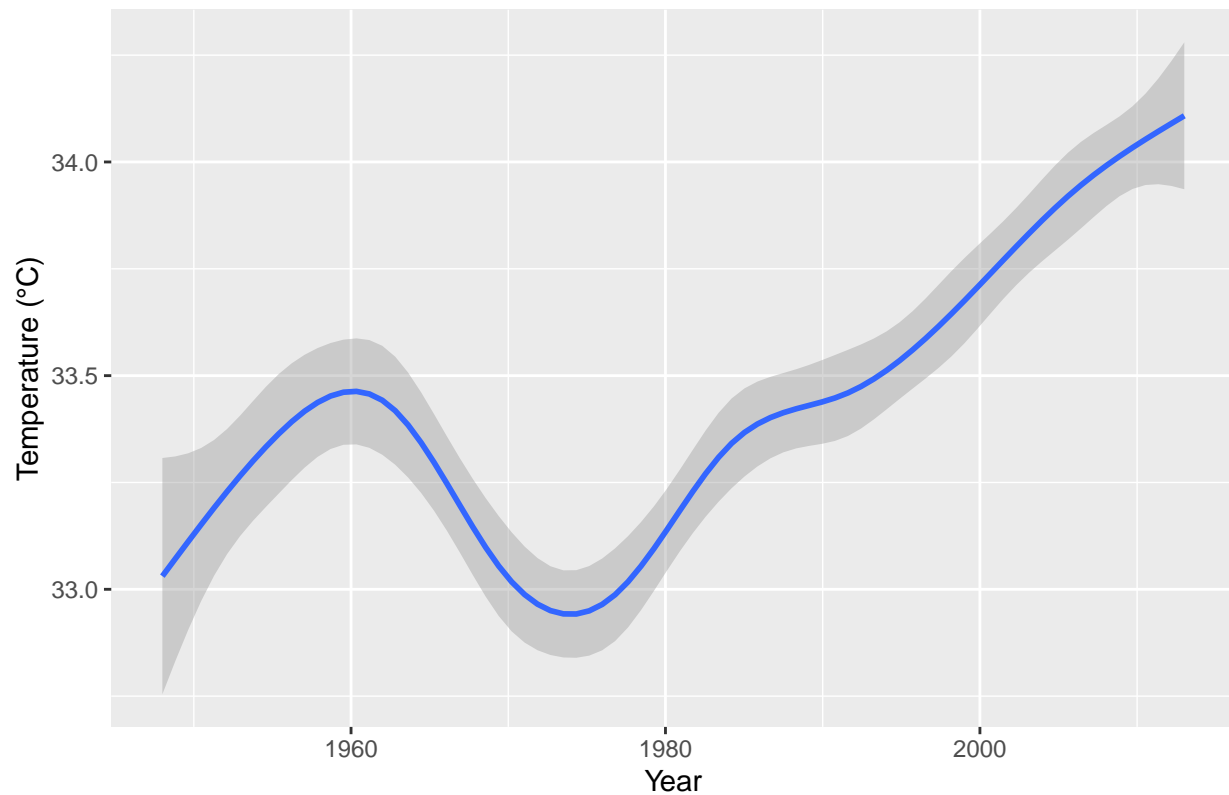
*#Maximum Temperature*

```
ggplot(weather_data, aes(x = YEAR, y = Max.Temp, fill = YEAR)) +
  geom_smooth() +
  labs(title = "Maximum Temperature Trends Over the Years", x = "Year", y = "Temperature (°C)")
```

```
## 'geom_smooth()' using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: The following aesthetics were dropped during statistical transformation: fill.
## i This can happen when ggplot fails to infer the correct grouping structure in
## the data.
## i Did you forget to specify a 'group' aesthetic or to convert a numerical
## variable into a factor?
```

## Maximum Temperature Trends Over the Years



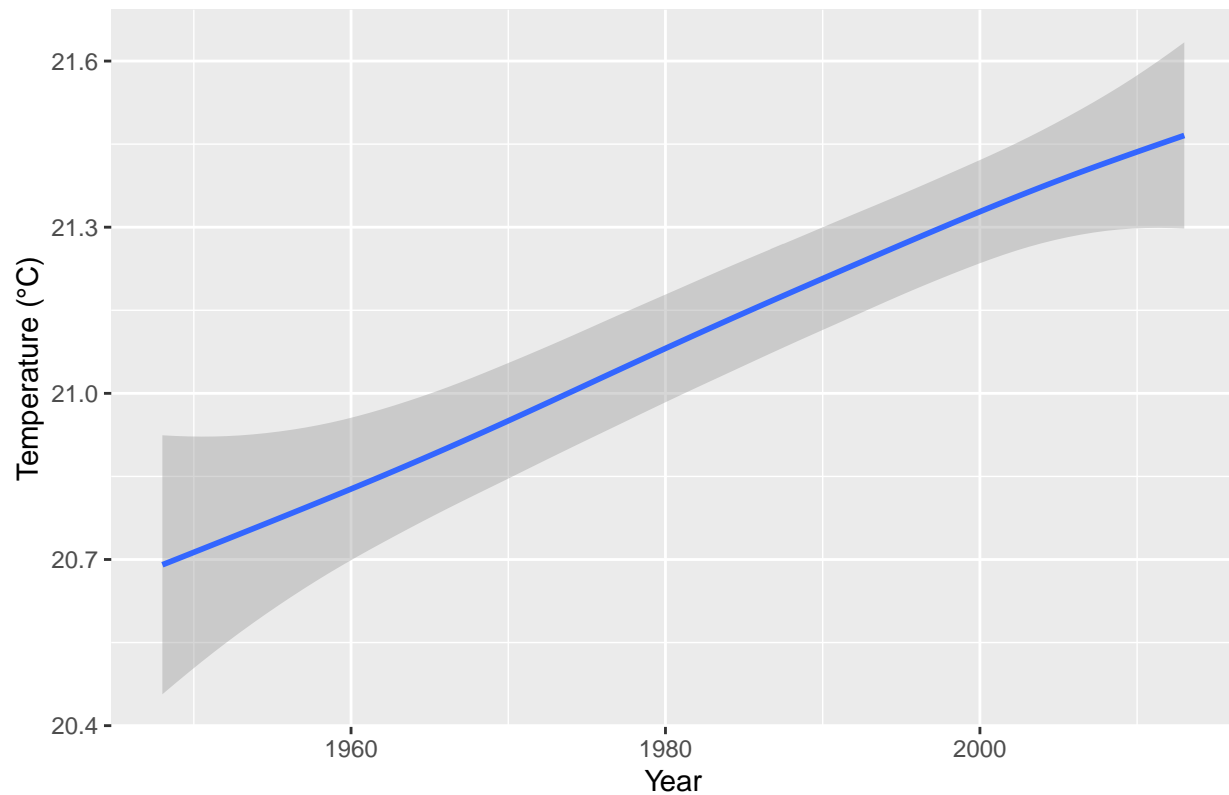
*#Minimum Temperature*

```
ggplot(weather_data, aes(x = YEAR, y = Min.Temp, fill = YEAR)) +  
  geom_smooth() +  
  labs(title = "Minimum Temperature Trends Over the Years", x = "Year", y = "Temperature (°C)")
```

```
## 'geom_smooth()' using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

```
## Warning: The following aesthetics were dropped during statistical transformation: fill.  
## i This can happen when ggplot fails to infer the correct grouping structure in  
##   the data.  
## i Did you forget to specify a 'group' aesthetic or to convert a numerical  
##   variable into a factor?
```

## Minimum Temperature Trends Over the Years



```
#Adding Average Temperature Column in Dataset
weather_data$Avg.Temp <- (weather_data$Max.Temp + weather_data$Min.Temp)/2
```

```
# Average Temperature by Year
avg_temp_by_year <- weather_data %>%
  group_by(YEAR) %>%
  summarise(Avg.Temp= mean(Avg.Temp, na.rm = TRUE)) %>%
  arrange(desc(Avg.Temp))
print(avg_temp_by_year, n=)
```

```
## # A tibble: 66 x 2
##   YEAR Avg.Temp
##   <int>   <dbl>
## 1  2010    28.1
## 2  2006    28.0
## 3  2009    27.9
## 4  1999    27.9
## 5  2005    27.8
## 6  1998    27.7
## 7  2012    27.7
## 8  1987    27.6
## 9  1958    27.6
## 10 1995    27.6
## # i 56 more rows
```

```
# Highest Average Temperature by Year
highest_Average_Temperature_by_year <- head(avg_temp_by_year, n=1)
print(highest_Average_Temperature_by_year)
```

```
## # A tibble: 1 x 2
##   YEAR Avg.Temp
##   <int>   <dbl>
## 1  2010    28.1
```

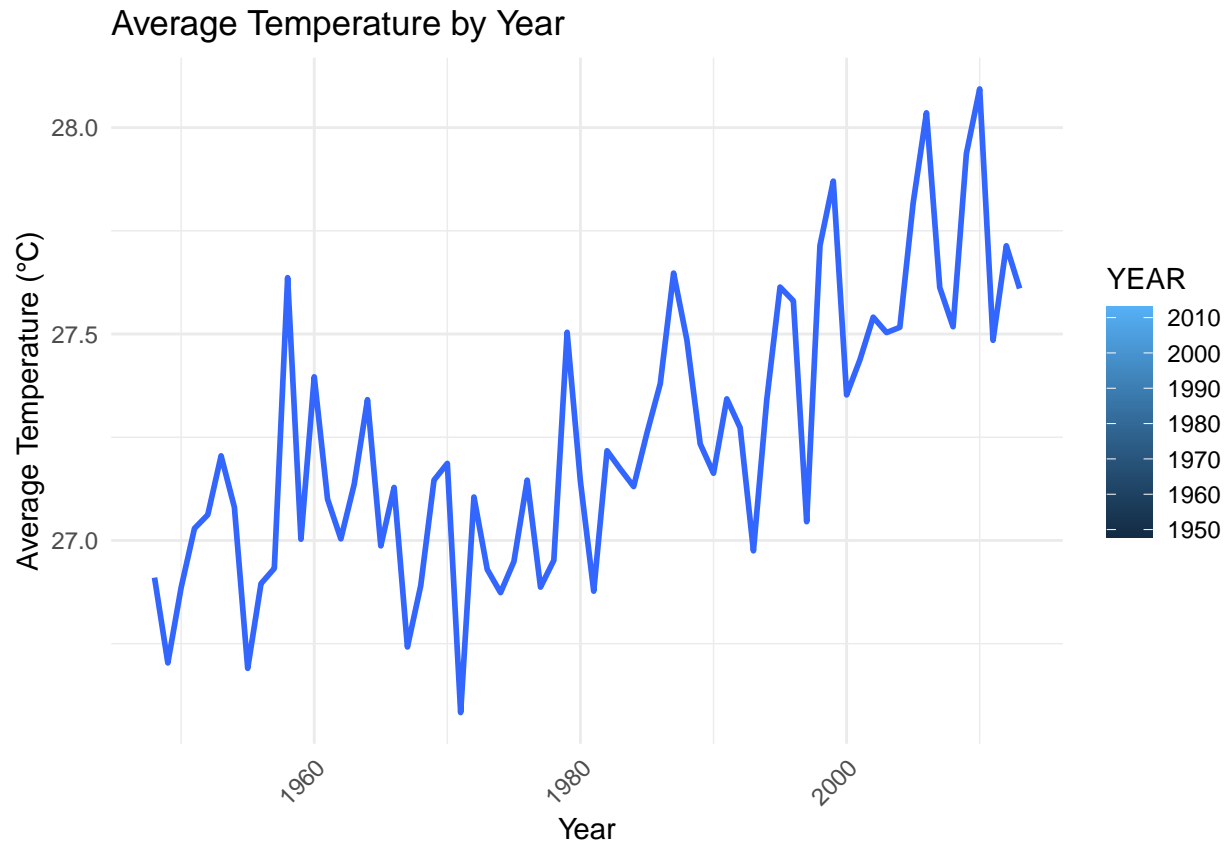
```
# Lowest Average Temperature by Year
lowest_Average_temperature_by_year <- tail(avg_temp_by_year, n=1)
print(lowest_Average_temperature_by_year)
```

```
## # A tibble: 1 x 2
##   YEAR Avg.Temp
##   <int>   <dbl>
## 1  1971    26.6
```

```
#Displaying Average Temperature in each Year
```

```
#Line Chart
```

```
ggplot(data = avg_temp_by_year) +
  geom_smooth(mapping = aes(x = YEAR, y = Avg.Temp, fill= YEAR), stat = "identity") +
  labs(title = "Average Temperature by Year",
       x = "Year",
       y = "Average Temperature (°C)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
# Average Temperature by Month
```

```
avg_temp_by_month <- weather_data %>%
  group_by(Month) %>%
  summarise(Avg.Temp= mean(Avg.Temp, na.rm = TRUE)) %>%
  arrange(desc(Avg.Temp))
print(avg_temp_by_month)
```

```
## # A tibble: 12 x 2
##   Month      Avg.Temp
##   <chr>      <dbl>
## 1 June       30.5
## 2 May        30.4
## 3 August     30.0
## 4 September  30.0
## 5 July       29.8
## 6 April      29.8
## 7 October    28.8
## 8 March      27.5
## 9 November   25.5
## 10 February  23.4
## 11 December  21.6
## 12 January   20.4
```

```
#Highest Average Temperature by Month
```

```
highest_Average_Temperature_by_month <- head(avg_temp_by_month, n=1)  
highest_Average_Temperature_by_month
```

```
## # A tibble: 1 x 2  
##   Month Avg.Temp  
##   <chr>   <dbl>  
## 1 June    30.5
```

```
# Lowest Average Temperature by Month
```

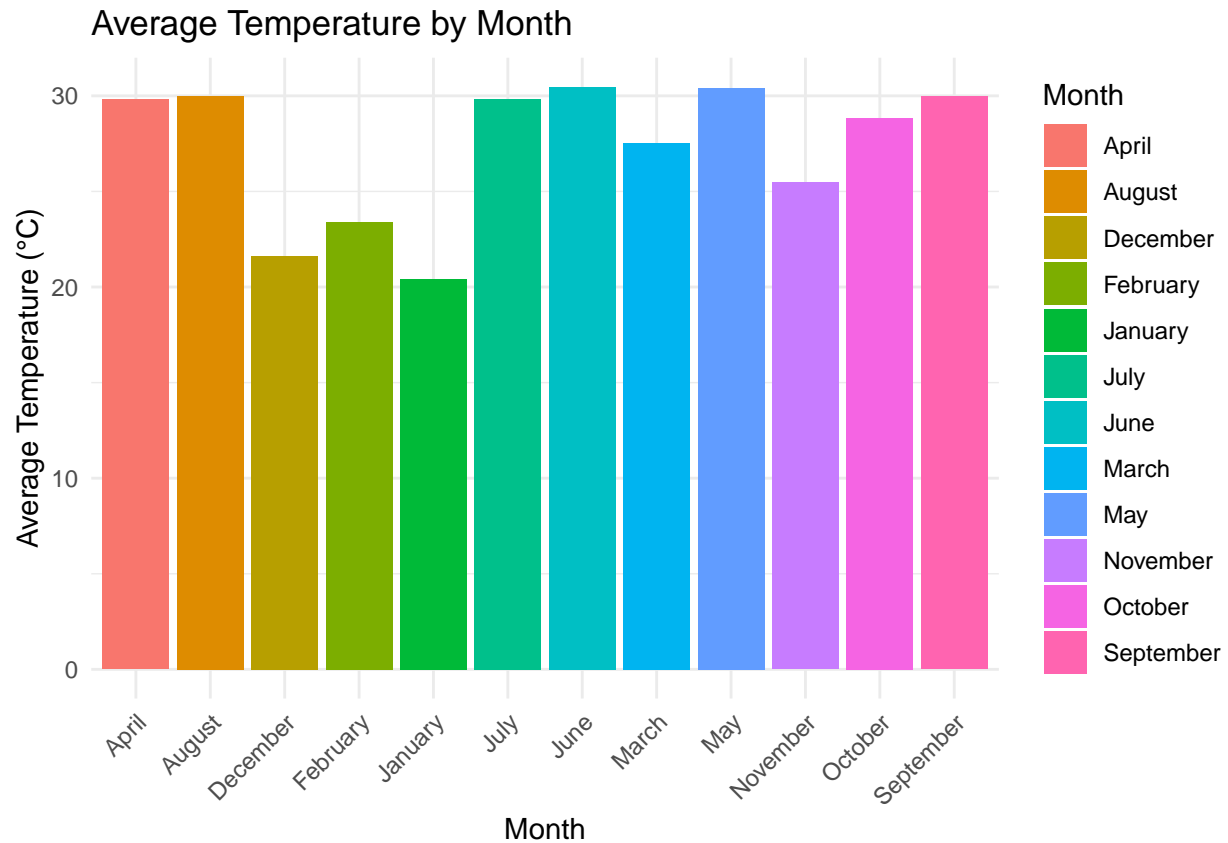
```
lowest_Average_temperature_by_month <- tail(avg_temp_by_month, n=1)  
lowest_Average_temperature_by_month
```

```
## # A tibble: 1 x 2  
##   Month   Avg.Temp  
##   <chr>    <dbl>  
## 1 January  20.4
```

```
#Displaying Average Temperature by Month
```

```
ggplot(data = avg_temp_by_month) +  
  geom_bar(mapping = aes(x = Month, y = Avg.Temp, fill = Month), stat = "identity") +  
  labs(title = "Average Temperature by Month",  
        x = "Month",  
        y = "Average Temperature (°C)") +  
  theme_minimal() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```





*#Average Temperature by Stations (Location)*

```
avg_temp_by_station <- weather_data %>%
  group_by(Station.Names) %>%
  summarise(Avg.Temp= mean(Avg.Temp, na.rm = TRUE)) %>%
  arrange(desc(Avg.Temp))
print(avg_temp_by_station)
```

```
## # A tibble: 35 x 2
##   Station.Names      Avg.Temp
##   <chr>             <dbl>
## 1 Mongla            28.4
## 2 Khulna            28.0
## 3 Chuadanga         28.0
## 4 Chittagong (City-Ambagan) 28.0
## 5 Satkhira          27.9
## 6 Khepupara         27.7
## 7 Patuakhali        27.7
## 8 Jessore           27.7
## 9 Madaripur         27.6
## 10 Rangamati         27.6
## # i 25 more rows
```

```
# Highest Average Temperature By Station
```

```
highest_Average_Temperature_by_station <- head(avg_temp_by_station, n=1)
print(highest_Average_Temperature_by_station)
```

```
## # A tibble: 1 x 2
##   Station.Names Avg.Temp
##   <chr>         <dbl>
## 1 Mongla       28.4
```

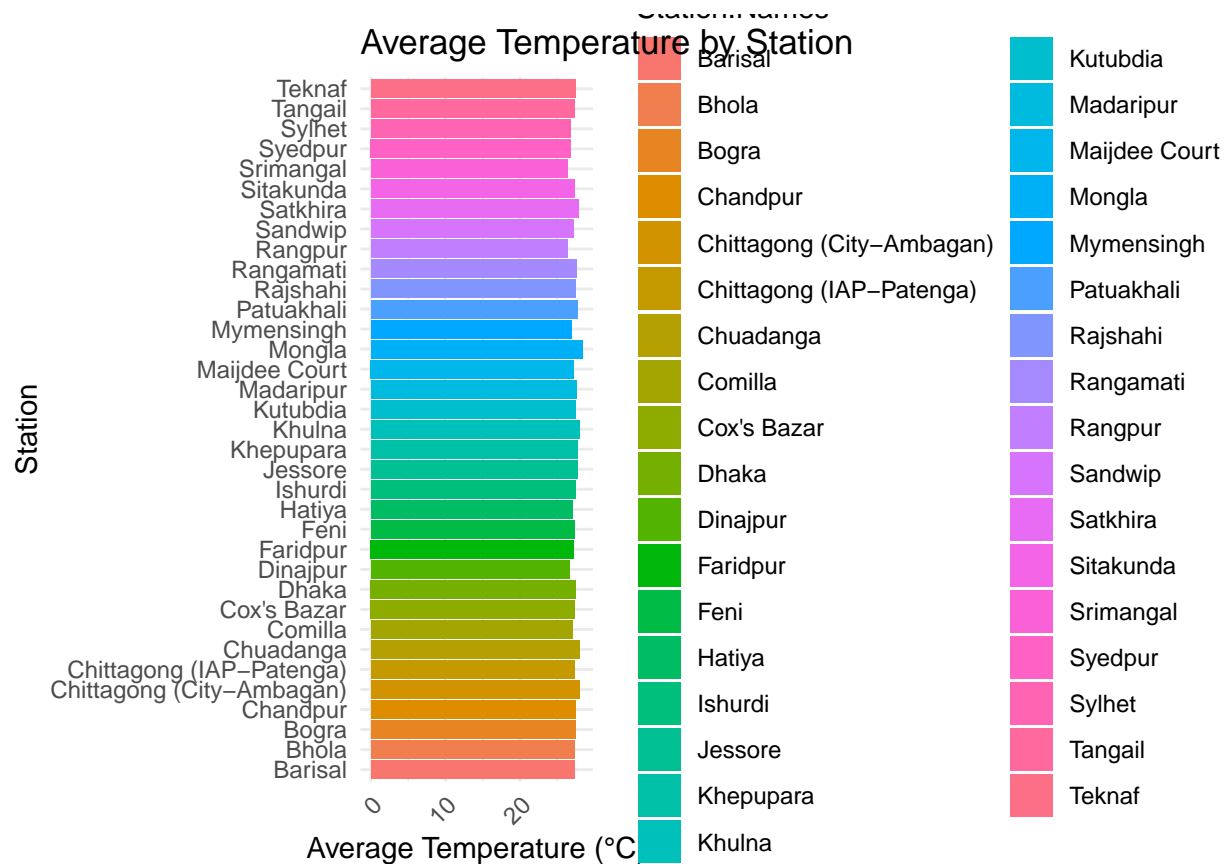
```
#Lowest Average Temperature by Station
```

```
lowest_Average_temperature_by_station <- tail(avg_temp_by_station, n=1)
print(lowest_Average_temperature_by_station)
```

```
## # A tibble: 1 x 2
##   Station.Names Avg.Temp
##   <chr>         <dbl>
## 1 Srimangal     26.4
```

```
# Dsiplaying Average Temperature by Station
```

```
ggplot(data = avg_temp_by_station) +
  geom_bar(mapping = aes(x = Station.Names, y = Avg.Temp, fill = Station.Names), stat = "identity") +
  labs(title = "Average Temperature by Station",
       x = "Station",
       y = "Average Temperature (°C)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip()
```



#### ## Average Humidity According by Year

```
avg_humidity_by_year <- weather_data %>%
  group_by(YEAR) %>%
  summarise(avg_humidity= mean(Relative.Humidity, na.rm = TRUE)) %>%
  arrange(desc(avg_humidity))
print(avg_humidity_by_year)
```

```
## # A tibble: 66 x 2
##   YEAR avg_humidity
##   <int>     <dbl>
## 1 1998      82.1
## 2 1990      81.7
## 3 2003      81.5
## 4 2008      81.4
## 5 1953      81.3
## 6 2002      81.2
## 7 2004      81.2
## 8 2005      81.2
## 9 2000      81.2
## 10 1999      81.1
## # i 56 more rows
```

```
# Highest Average Humidity by Year
highest_average_humidity_by_year <- head(avg_humidity_by_year, n=1)
highest_average_humidity_by_year
```

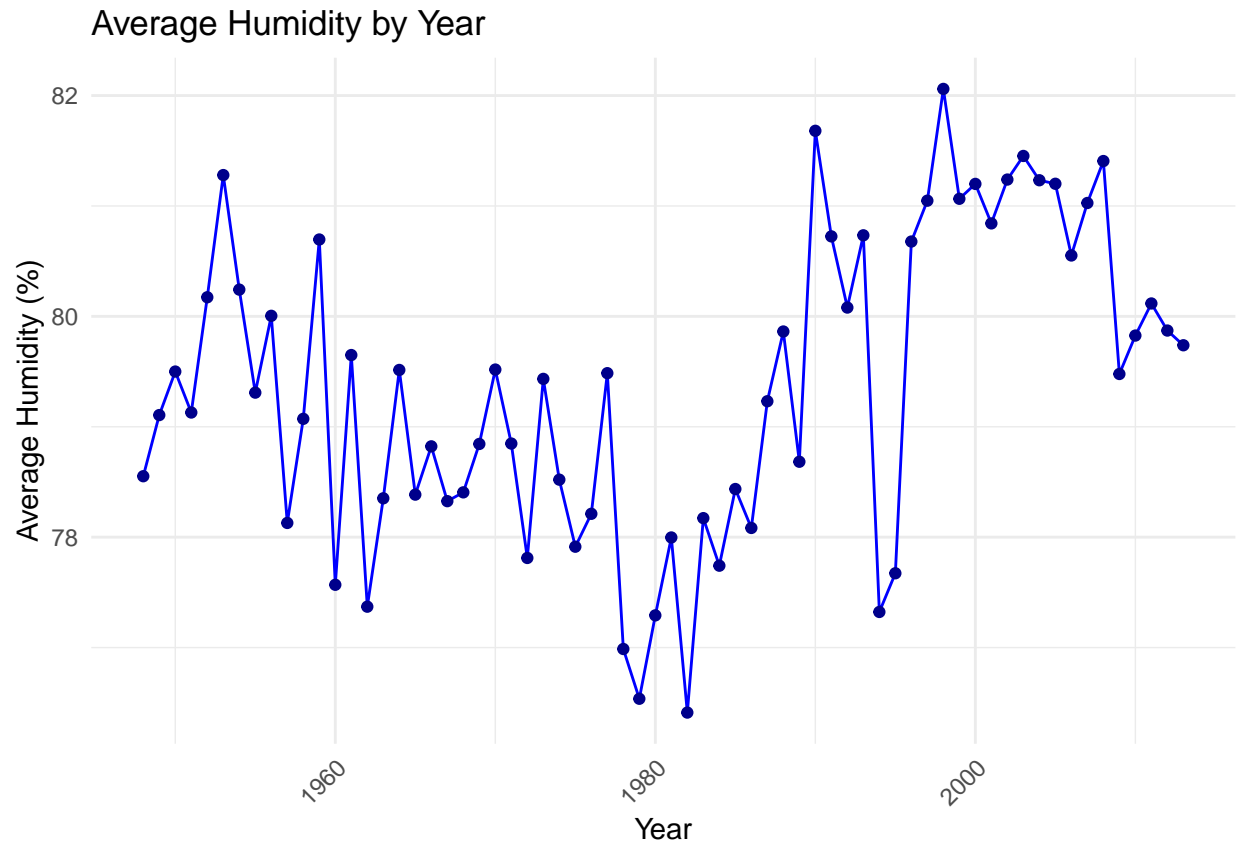
```
## # A tibble: 1 x 2
##   YEAR avg_humidity
##   <int>     <dbl>
## 1  1998         82.1
```

```
# Lowest Average Humidity by Year
lowest_average_humidity_by_year <- tail(avg_humidity_by_year, n=1)
lowest_average_humidity_by_year
```

```
## # A tibble: 1 x 2
##   YEAR avg_humidity
##   <int>     <dbl>
## 1  1982         76.4
```

```
# Displaying Average Humidity by Year

ggplot(data = avg_humidity_by_year, aes(x = YEAR, y = avg_humidity)) +
  geom_line(color = "blue") +
  geom_point(color = "darkblue") +
  labs(title = "Average Humidity by Year",
       x = "Year",
       y = "Average Humidity (%)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



#### ## Average Humidity by Month

```
avg_humidity_by_month <- weather_data %>%
  group_by(Month) %>%
  summarise(avg_humidity= mean(Relative.Humidity, na.rm = TRUE)) %>%
  arrange(desc(avg_humidity))
print(avg_humidity_by_month)
```

```
## # A tibble: 12 x 2
##   Month      avg_humidity
##   <chr>         <dbl>
## 1 July           87.2
## 2 August         86.6
## 3 September      86.0
## 4 June           85.3
## 5 October        83.1
## 6 May            79.1
## 7 November       78.7
## 8 December       77.4
## 9 January        75.6
## 10 April          74.0
## 11 February       71.1
## 12 March          69.6
```

```
# Highest Average Humidity by Month
```

```
highest_average_humidity_by_month <- head(avg_humidity_by_month, n=1)  
highest_average_humidity_by_month
```

```
## # A tibble: 1 x 2  
##   Month avg_humidity  
##   <chr>      <dbl>  
## 1 July      87.2
```

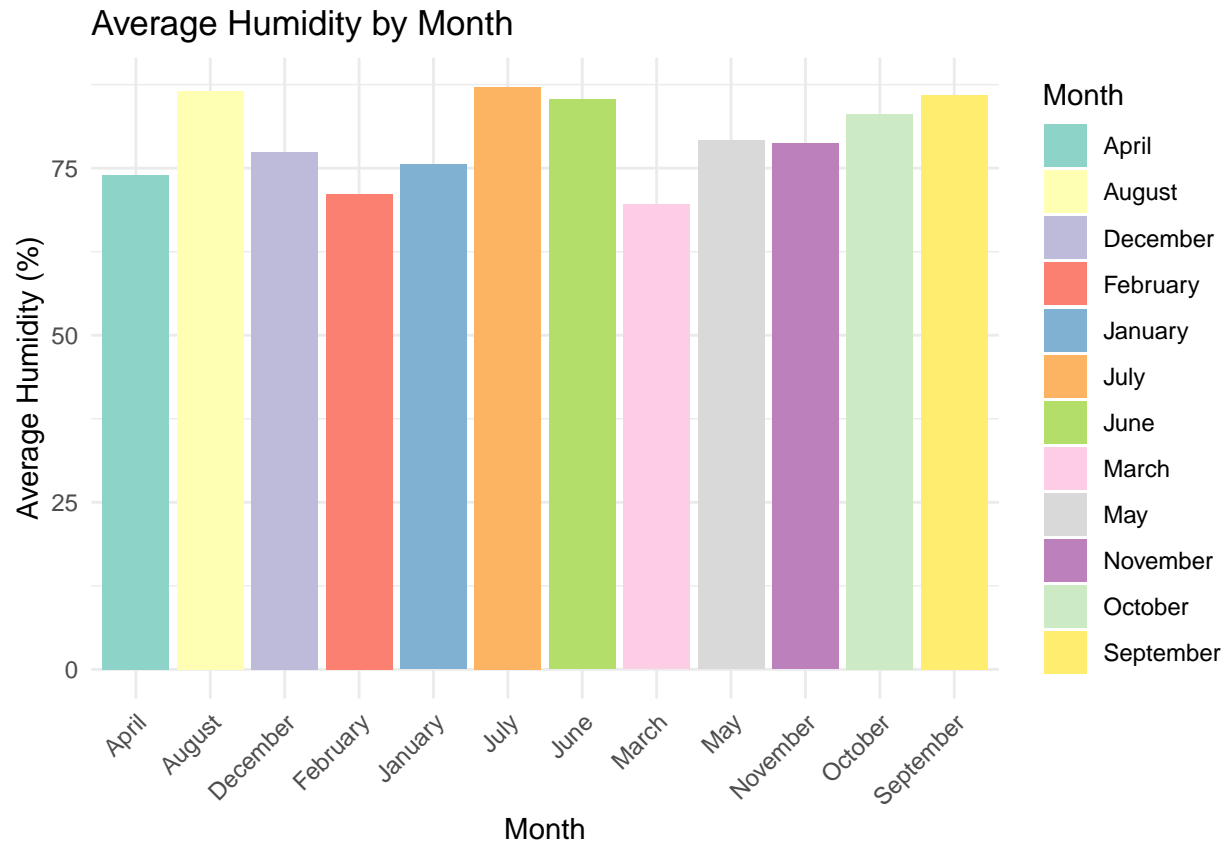
```
# Lowest Average Humidity by Year
```

```
lowest_average_humidity_by_month <- tail(avg_humidity_by_month, n=1)  
lowest_average_humidity_by_month
```

```
## # A tibble: 1 x 2  
##   Month avg_humidity  
##   <chr>      <dbl>  
## 1 March     69.6
```

```
# Displaying Average Humidity by Month
```

```
ggplot(data = avg_humidity_by_month, aes(x = Month, y = avg_humidity, fill = Month)) +  
  geom_bar(stat = "identity") +  
  labs(title = "Average Humidity by Month",  
        x = "Month",  
        y = "Average Humidity (%)") +  
  theme_minimal() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +  
  scale_fill_brewer(palette = "Set3")
```



*# Average Humidity by Stations (Location)*

```
avg_humidity_by_station <- weather_data %>%
  group_by(Station.Names) %>%
  summarise(avg_humidity= mean(Relative.Humidity, na.rm = TRUE)) %>%
  arrange(desc(avg_humidity))
print(avg_humidity_by_station)
```

```
## # A tibble: 35 x 2
##   Station.Names avg_humidity
##   <chr>         <dbl>
## 1 Bhola         83.8
## 2 Sandwip       83.5
## 3 Patuakhali    82.5
## 4 Barisal       82.1
## 5 Srimangal     81.6
## 6 Kutubdia      81.6
## 7 Hatiya        81.4
## 8 Mongla        81.3
## 9 Maijdee Court 81.0
## 10 Khepupara    81.0
## # i 25 more rows
```

```
# Highest Average Humidity by Station
highest_average_humidity_by_station <- head(avg_humidity_by_station, n=1)
highest_average_humidity_by_station
```

```
## # A tibble: 1 x 2
##   Station.Names avg_humidity
##   <chr>         <dbl>
## 1 Bhola         83.8
```

```
# Lowest Average Humidity by Year
lowest_average_humidity_by_station <- tail(avg_humidity_by_station, n=1)
lowest_average_humidity_by_station
```

```
## # A tibble: 1 x 2
##   Station.Names avg_humidity
##   <chr>         <dbl>
## 1 Dinajpur      75.4
```

```
# Displaying Average Humidity by Station
sum(is.na(avg_humidity_by_month$Month))
```

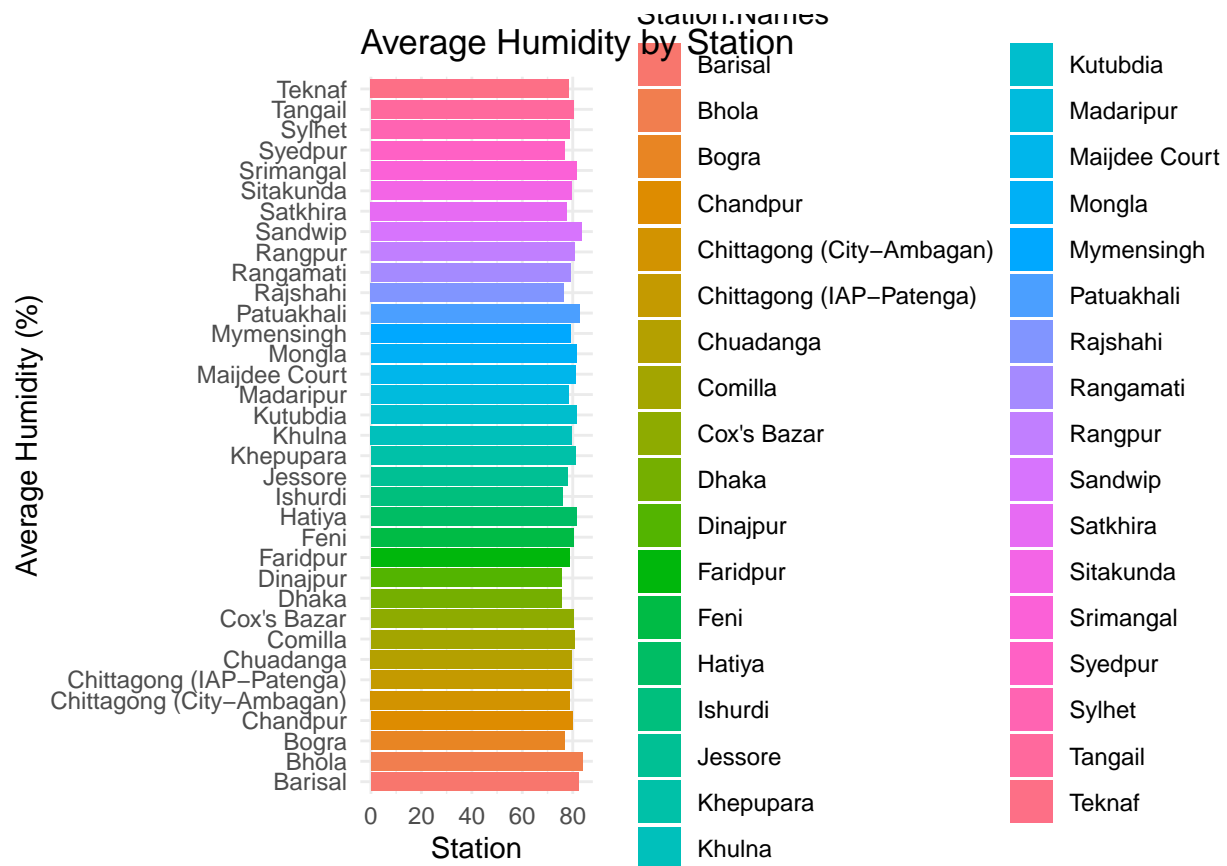
```
## [1] 0
```

```
sum(is.na(avg_humidity_by_month$avg_humidity))
```

```
## [1] 0
```

```
ggplot(data = avg_humidity_by_station, aes(x = avg_humidity, y = Station.Names, fill = Station.Names)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Humidity by Station",
       x = "Station",
       y = "Average Humidity (%)") +
  theme_minimal()
```





```
## Wind speed
```

```
# Average Wind Speed by Year
```

```
avg_windspeed_by_year <- weather_data %>%
  group_by(YEAR) %>%
  summarise(avg_windspeed = mean(Wind.Speed, na.rm = TRUE)) %>%
  arrange(desc(avg_windspeed))
print(avg_windspeed_by_year)
```

```
## # A tibble: 66 x 2
##   YEAR avg_windspeed
##   <int>     <dbl>
## 1 2012         2.50
## 2 2013         2.47
## 3 2011         2.46
## 4 2010         2.39
## 5 2009         2.32
## 6 1984         1.66
## 7 1966         1.64
## 8 1985         1.59
## 9 1967         1.56
## 10 1988         1.52
## # i 56 more rows
```

```
# Highest Average Wind Speed by Year
```

```
highest_average_windspeed_by_year <- head(avg_windspeed_by_year, n=1)  
highest_average_windspeed_by_year
```

```
## # A tibble: 1 x 2  
##   YEAR avg_windspeed  
##   <int>         <dbl>  
## 1  2012           2.50
```

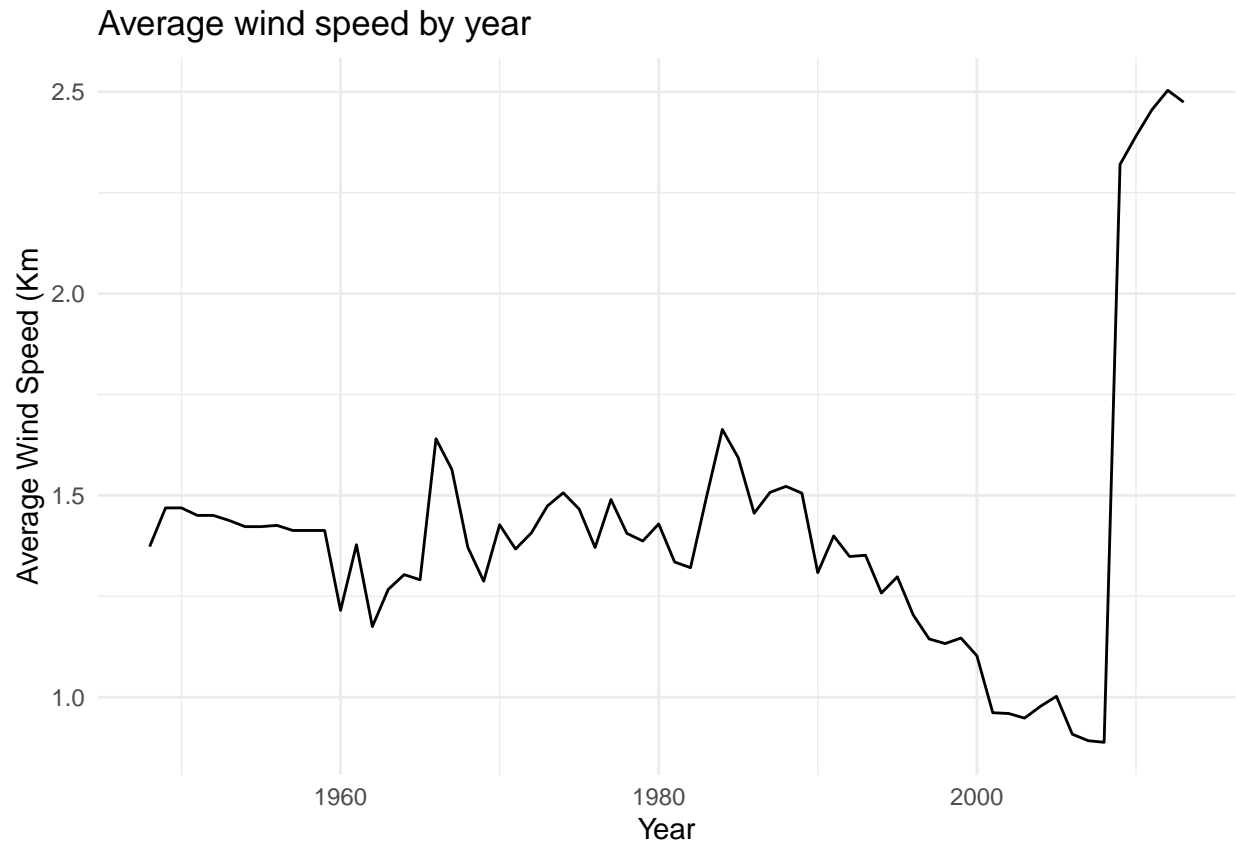
```
# Lowest Average Humidity by Year
```

```
lowest_average_windspeed_by_year <- tail(avg_windspeed_by_year, n=1)  
lowest_average_windspeed_by_year
```

```
## # A tibble: 1 x 2  
##   YEAR avg_windspeed  
##   <int>         <dbl>  
## 1  2008           0.888
```

```
# Displaying average wind speed by year
```

```
ggplot(data = avg_windspeed_by_year, aes(x = YEAR, y = avg_windspeed, fill = avg_windspeed)) +  
  geom_line(stat = "identity") +  
  labs(title = "Average wind speed by year",  
        x = "Year",  
        y = "Average Wind Speed (Km)" ) +  
  theme_minimal()
```



#### ## Average Wind speed by Month

```
avg_windspeed_by_month <- weather_data %>%
  group_by(Month) %>%
  summarise(avg_windspeed = mean(Wind.Speed, na.rm = TRUE)) %>%
  arrange(desc(avg_windspeed))
print(avg_windspeed_by_month)
```

```
## # A tibble: 12 x 2
##   Month      avg_windspeed
##   <chr>          <dbl>
## 1 April          2.06
## 2 June           2.03
## 3 May            1.99
## 4 July           1.97
## 5 August         1.76
## 6 March          1.50
## 7 September     1.36
## 8 February      1.07
## 9 October        0.891
## 10 January       0.890
## 11 December      0.732
## 12 November      0.725
```

```
# Highest Average Wind Speed by Month
highest_average_windspeed_by_month <- head(avg_windspeed_by_month, n=1)
highest_average_windspeed_by_month
```

```
## # A tibble: 1 x 2
##   Month avg_windspeed
##   <chr>         <dbl>
## 1 April          2.06
```

```
# Lowest Average Humidity by Year
lowest_average_windspeed_by_month <- tail(avg_windspeed_by_month, n=1)
lowest_average_windspeed_by_month
```

```
## # A tibble: 1 x 2
##   Month      avg_windspeed
##   <chr>          <dbl>
## 1 November       0.725
```

```
## Determine in which month of 2000 the average wind speed was highest?
```

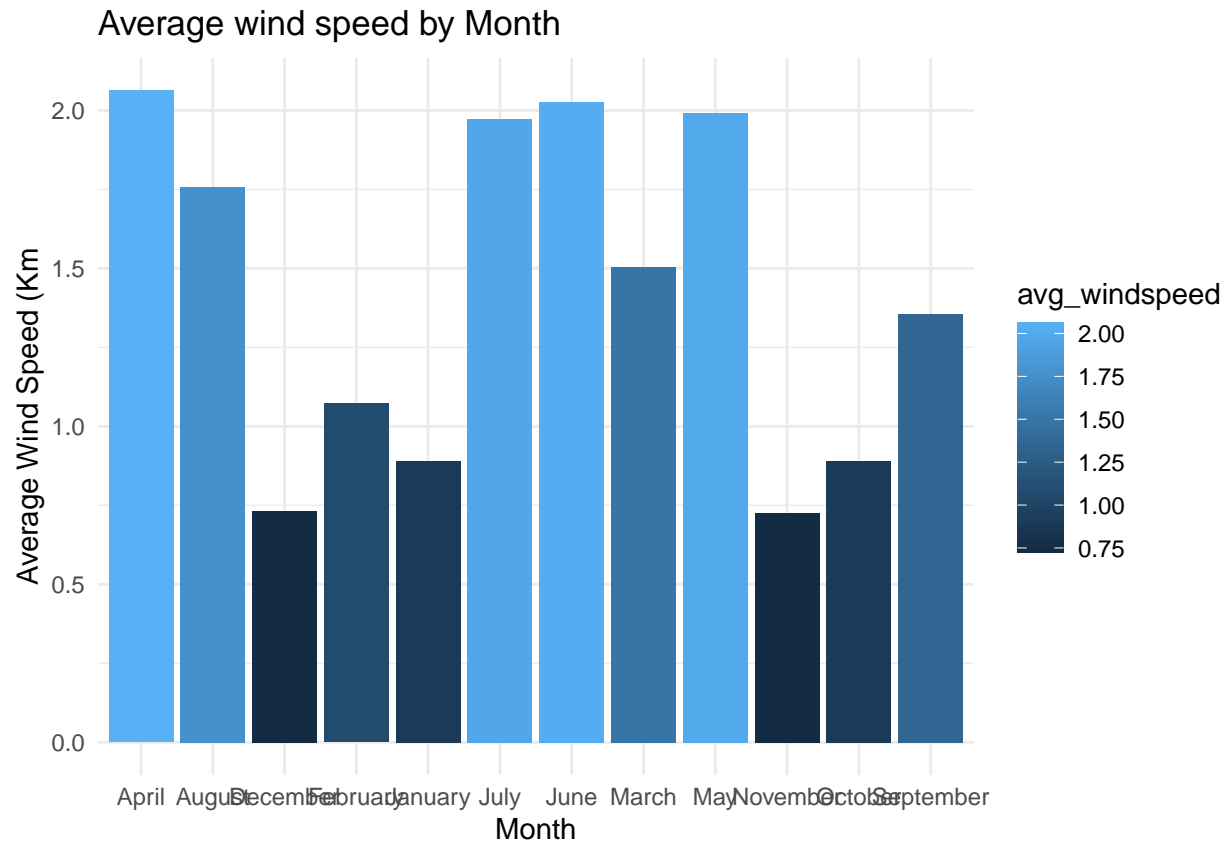
```
avg_windspeed_2000 <- weather_data %>%
  filter(YEAR == 2000) %>%
  group_by(Month) %>%
  summarise(avg_windspeed = mean(Wind.Speed, na.rm = TRUE)) %>%
  arrange(desc(avg_windspeed))
```

```
# Identify the month with the highest average windspeed
highest_windspeed_2000 <- avg_windspeed_2000 %>%
  slice(1) # Get the top row with the highest average windspeed
```

```
# Print the result
print(highest_windspeed_2000)
```

```
## # A tibble: 1 x 2
##   Month avg_windspeed
##   <chr>         <dbl>
## 1 April          2.02
```

```
# Displaying Average Wind Speed by Month
ggplot(data = avg_windspeed_by_month, aes(x = Month, y = avg_windspeed, fill = avg_windspeed)) +
  geom_bar(stat = "identity") +
  labs(title = "Average wind speed by Month",
       x = "Month",
       y = "Average Wind Speed (Km)" ) +
  theme_minimal()
```



*# Average Wind Speed by Latitude & Longitude*

```
avg_windspeed_by_location <- weather_data %>%
  group_by(LATITUDE, LONGITUDE) %>%
  summarise(avg_windspeed = mean(Wind.Speed, na.rm = TRUE)) %>%
  arrange(desc(avg_windspeed))
```

## 'summarise()' has grouped output by 'LATITUDE'. You can override using the  
## '.groups' argument.

```
print(avg_windspeed_by_location)
```

```
## # A tibble: 35 x 3
## # Groups:   LATITUDE [32]
##   LATITUDE LONGITUDE avg_windspeed
##   <dbl>     <dbl>         <dbl>
## 1    22.4     91.8           3.30
## 2    22.3     91.8           2.87
## 3    21.5     92.0           2.34
## 4    23.2     89.2           2.12
## 5    22.0     90.2           2.08
## 6    25.8     88.9           1.98
## 7    22.3     91.1           1.87
## 8    21.8     91.8           1.74
```

```
## 9      24.1      89.0      1.61
## 10     23.5      91.2      1.57
## # i 25 more rows
```

```
# Highest Average Wind Speed by Latitude & Longitude
```

```
highest_average_windspeed_by_location <- head(avg_windspeed_by_location, n=1)
highest_average_windspeed_by_location
```

```
## # A tibble: 1 x 3
## # Groups:   LATITUDE [1]
##   LATITUDE LONGITUDE avg_windspeed
##   <dbl>     <dbl>     <dbl>
## 1     22.4     91.8     3.30
```

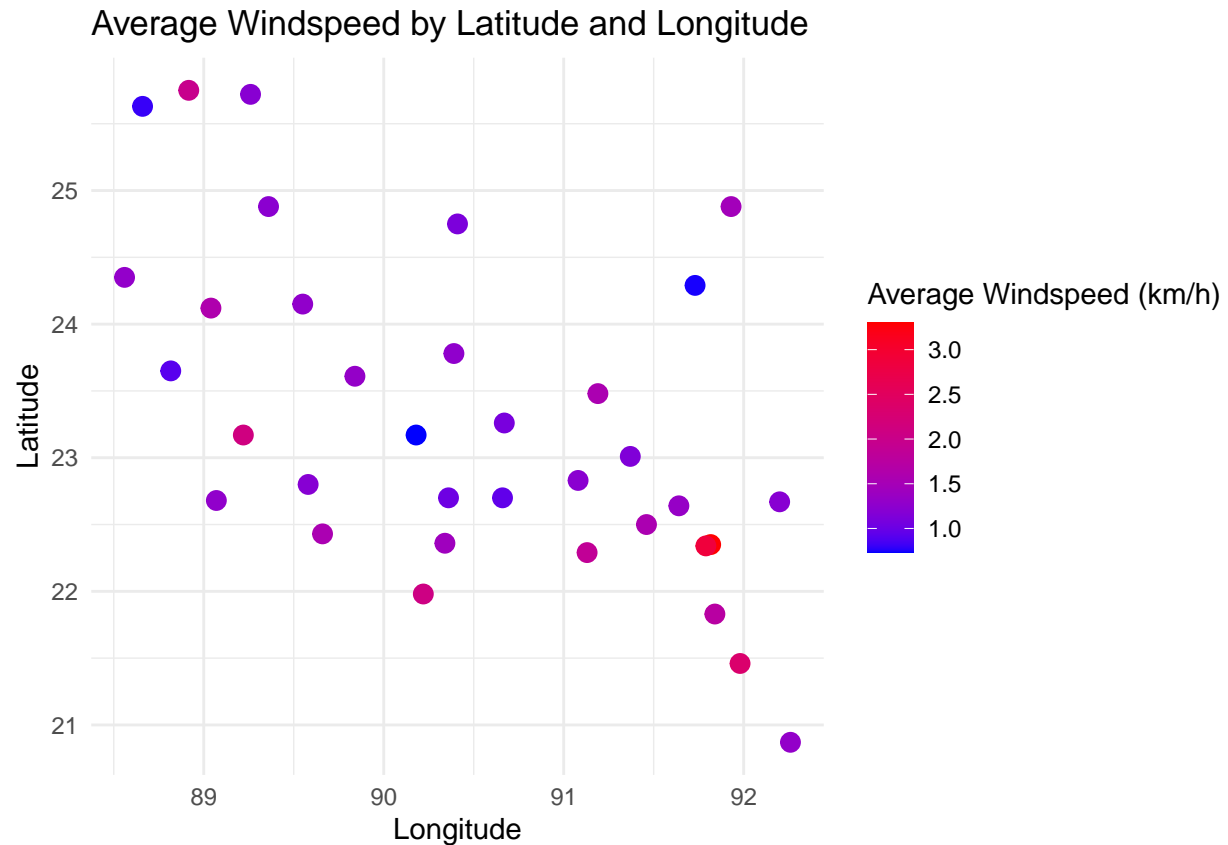
```
# Lowest Average Humidity by Latitude & Longitude
```

```
lowest_average_windspeed_by_location <- tail(avg_windspeed_by_location, n=1)
lowest_average_windspeed_by_location
```

```
## # A tibble: 1 x 3
## # Groups:   LATITUDE [1]
##   LATITUDE LONGITUDE avg_windspeed
##   <dbl>     <dbl>     <dbl>
## 1     23.2     90.2     0.734
```

```
# Displaying Average Wind Speed by Latitude & Longitude
```

```
ggplot(data = avg_windspeed_by_location, aes(x = LONGITUDE, y = LATITUDE, color = avg_windspeed)) +
  geom_point(size = 3) +
  scale_color_gradient(low = "blue", high = "red") + # Color gradient from blue (low windspeed) to red
  labs(title = "Average Windspeed by Latitude and Longitude",
        x = "Longitude",
        y = "Latitude",
        color = "Average Windspeed (km/h)") +
  theme_minimal()
```



```
## Average Temperature by Latitude & Longitude
```

```
avg_temp_by_location <- weather_data %>%
  group_by(LATITUDE, LONGITUDE) %>%
  summarise(avg_temp = mean(Avg.Temp, na.rm = TRUE)) %>%
  arrange(desc(avg_temp))
```

```
## 'summarise()' has grouped output by 'LATITUDE'. You can override using the
## '.groups' argument.
```

```
print(avg_temp_by_location)
```

```
## # A tibble: 35 x 3
## # Groups:   LATITUDE [32]
##   LATITUDE LONGITUDE avg_temp
##   <dbl>     <dbl>   <dbl>
## 1    22.4      89.7    28.4
## 2    22.8      89.6    28.0
## 3    23.6      88.8    28.0
## 4    22.4      91.8    28.0
## 5    22.7      89.1    27.9
## 6    22.0      90.2    27.7
## 7    22.4      90.3    27.7
## 8    23.2      89.2    27.7
## 9    23.2      90.2    27.6
```

```
## 10      22.7      92.2      27.6
## # i 25 more rows
```

```
# Highest Average Temperature by Latitude & Longitude
```

```
highest_average_temp_by_location <- head(avg_temp_by_location, n=1)
highest_average_temp_by_location
```

```
## # A tibble: 1 x 3
## # Groups:   LATITUDE [1]
##   LATITUDE LONGITUDE avg_temp
##   <dbl>     <dbl>     <dbl>
## 1      22.4      89.7      28.4
```

```
# Lowest Average Temperature by Latitude & Longitude
```

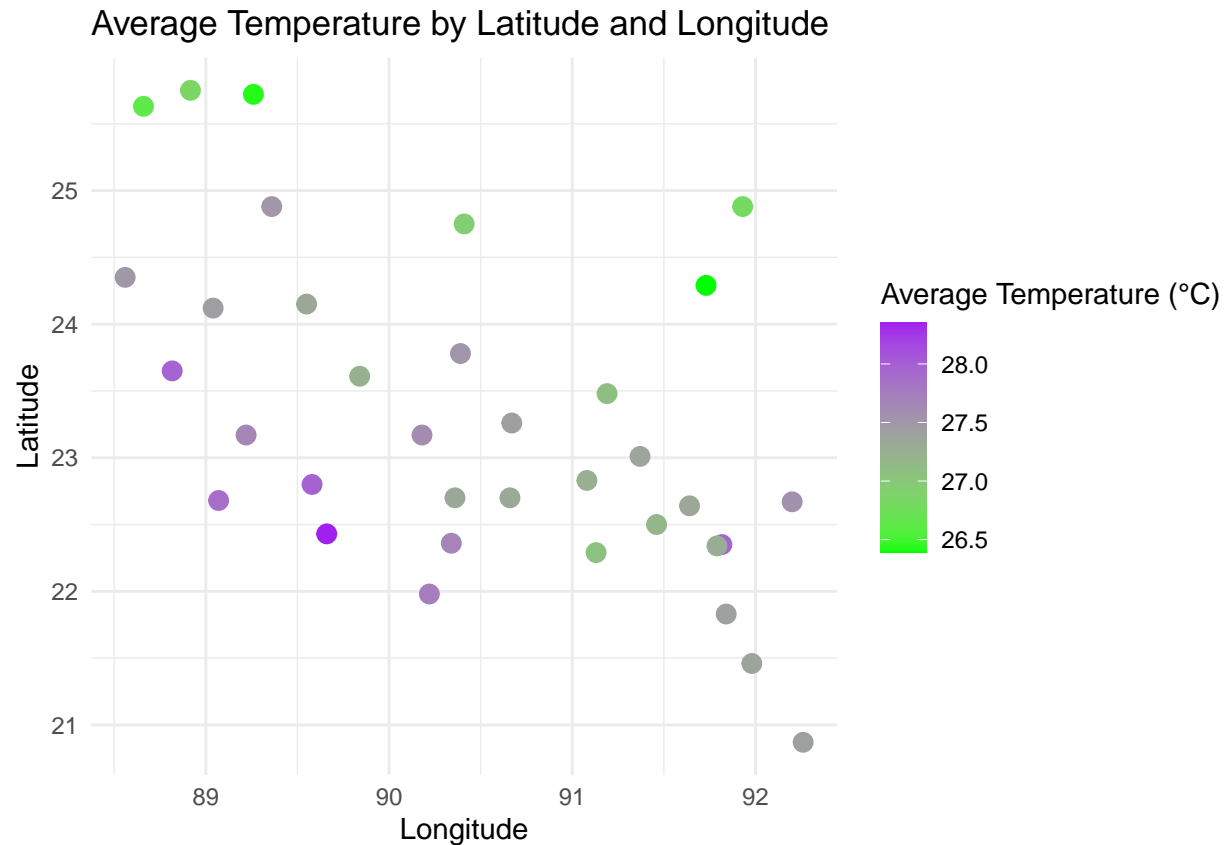
```
lowest_average_temp_by_location <- tail(avg_temp_by_location, n=1)
lowest_average_temp_by_location
```

```
## # A tibble: 1 x 3
## # Groups:   LATITUDE [1]
##   LATITUDE LONGITUDE avg_temp
##   <dbl>     <dbl>     <dbl>
## 1      24.3      91.7      26.4
```

```
## Displaying Average Temperature by Latitude & Longitude
```

```
ggplot(data = avg_temp_by_location, aes(x = LONGITUDE, y = LATITUDE, color = avg_temp)) +
  geom_point(size = 3) +
  scale_color_gradient(low = "green", high = "purple") + # Color gradient from blue (cold) to red (hot)
  labs(title = "Average Temperature by Latitude and Longitude",
       x = "Longitude",
       y = "Latitude",
       color = "Average Temperature (°C)") +
  theme_minimal()
```





```
## Average Rainfall by Latitude & Longitude
avg_rainfall_by_location <- weather_data %>%
  group_by(LATITUDE, LONGITUDE) %>%
  summarise(avg_rainfall = mean(Rainfall, na.rm = TRUE)) %>%
  arrange(desc(avg_rainfall))
```

```
## 'summarise()' has grouped output by 'LATITUDE'. You can override using the
## '.groups' argument.
```

```
print(avg_rainfall_by_location)
```

```
## # A tibble: 35 x 3
## # Groups:   LATITUDE [32]
##   LATITUDE LONGITUDE avg_rainfall
##   <dbl>     <dbl>     <dbl>
## 1    20.9      92.3       342.
## 2    24.9      91.9       334.
## 3    21.5      92.0       303.
## 4    22.5      91.5       296.
## 5    22.3      91.1       261.
## 6    22.8      91.1       258.
## 7    23.0      91.4       248.
## 8    22.6      91.6       248.
## 9    22.4      91.8       244.
## 10   22.3      91.8       241.
```

```
## # i 25 more rows
```

```
# Highest Average Rainfall by Latitude & Longitude
```

```
highest_average_rainfall_by_location <- head(avg_rainfall_by_location, n=1)
```

```
highest_average_rainfall_by_location
```

```
## # A tibble: 1 x 3
```

```
## # Groups:   LATITUDE [1]
```

```
##   LATITUDE LONGITUDE avg_rainfall
```

```
##   <dbl>      <dbl>      <dbl>
```

```
## 1    20.9      92.3      342.
```

```
# Lowest Average Rainfall by Latitude & Longitude
```

```
lowest_average_rainfall_by_location <- tail(avg_rainfall_by_location, n=1)
```

```
lowest_average_rainfall_by_location
```

```
## # A tibble: 1 x 3
```

```
## # Groups:   LATITUDE [1]
```

```
##   LATITUDE LONGITUDE avg_rainfall
```

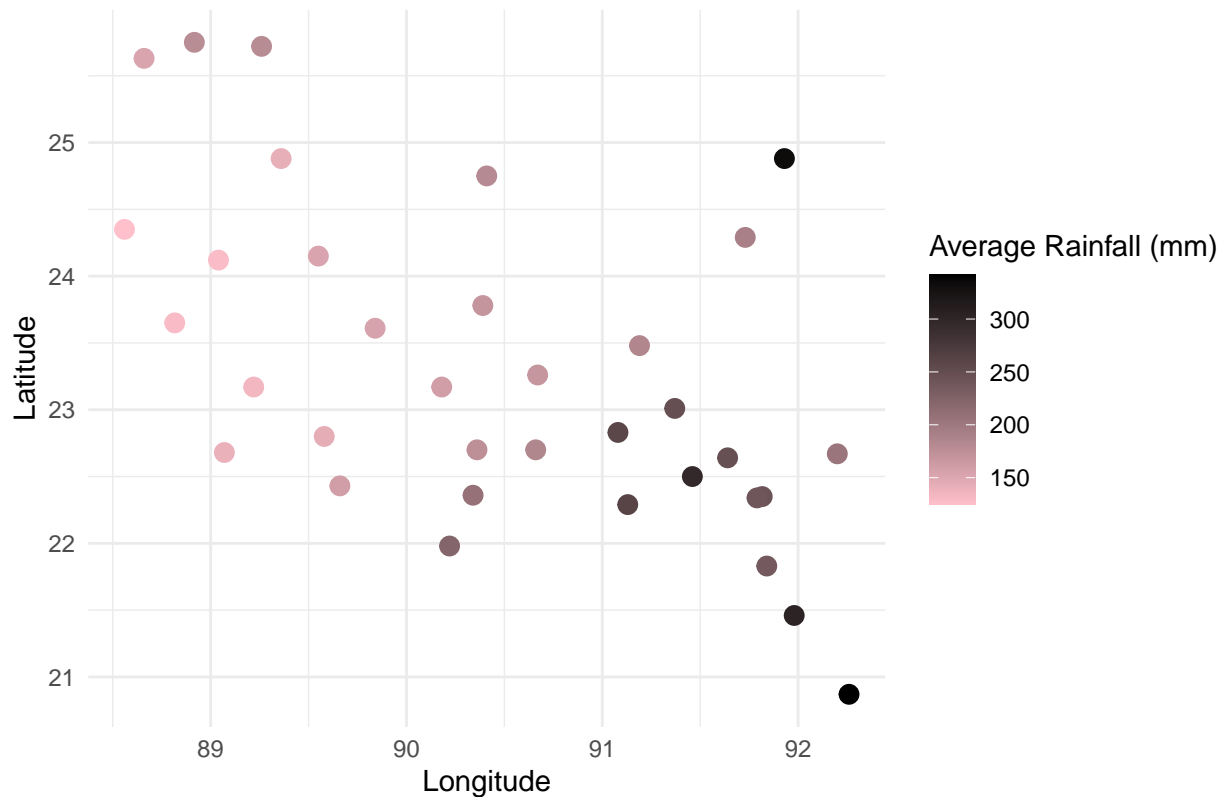
```
##   <dbl>      <dbl>      <dbl>
```

```
## 1    24.4      88.6      125.
```

```
## Displaying Average Rainfall by Latitude & Longitude
```

```
ggplot(data = avg_rainfall_by_location, aes(x = LONGITUDE, y = LATITUDE, color = avg_rainfall)) +  
  geom_point(size = 3) +  
  scale_color_gradient(low = "pink", high = "black") +  
  labs(title = "Average Rainfall by Latitude and Longitude",  
        x = "Longitude",  
        y = "Latitude",  
        color = "Average Rainfall (mm)") +  
  theme_minimal()
```

## Average Rainfall by Latitude and Longitude



### ## Average Humidity by Latitude & Longitude

```
avg_humidity_by_location <- weather_data %>%
  group_by(LATITUDE, LONGITUDE) %>%
  summarise(avg_humidity = mean(Relative.Humidity, na.rm = TRUE)) %>%
  arrange(desc(avg_humidity))
```

## 'summarise()' has grouped output by 'LATITUDE'. You can override using the  
## '.groups' argument.

```
print(avg_humidity_by_location)
```

```
## # A tibble: 35 x 3
## # Groups:   LATITUDE [32]
##   LATITUDE LONGITUDE avg_humidity
##   <dbl>     <dbl>     <dbl>
## 1    22.7     90.7      83.8
## 2    22.5     91.5      83.5
## 3    22.4     90.3      82.5
## 4    22.7     90.4      82.1
## 5    24.3     91.7      81.6
## 6    21.8     91.8      81.6
## 7    22.3     91.1      81.4
## 8    22.4     89.7      81.3
## 9    22.8     91.1      81.0
```

```
## 10      22.0      90.2      81.0
## # i 25 more rows
```

```
# Highest Average Humidity by Latitude & Longitude
```

```
highest_average_humidity_by_location <- head(avg_humidity_by_location, n=1)
highest_average_humidity_by_location
```

```
## # A tibble: 1 x 3
## # Groups:   LATITUDE [1]
##   LATITUDE LONGITUDE avg_humidity
##   <dbl>     <dbl>     <dbl>
## 1      22.7      90.7      83.8
```

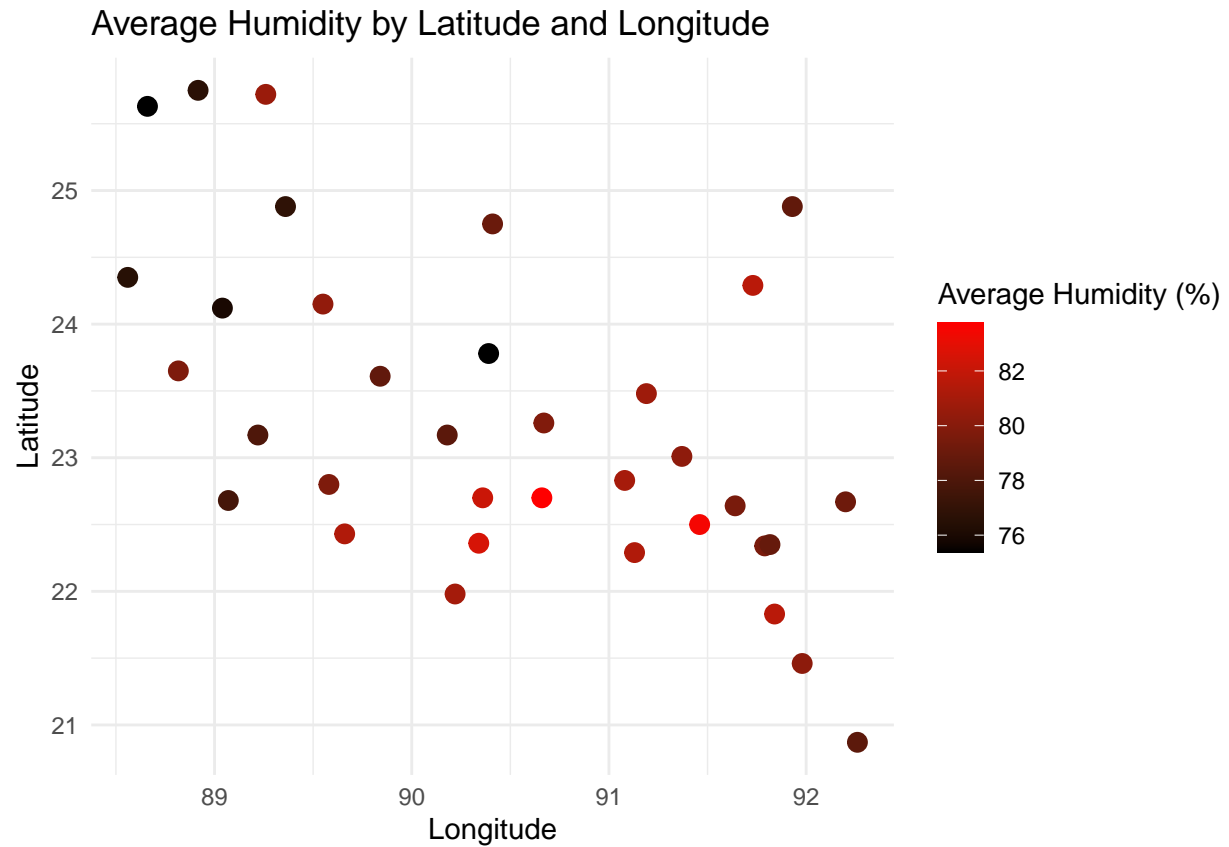
```
# Lowest Average Humidity by Latitude & Longitude
```

```
lowest_average_humidity_by_location <- tail(avg_humidity_by_location, n=1)
lowest_average_humidity_by_location
```

```
## # A tibble: 1 x 3
## # Groups:   LATITUDE [1]
##   LATITUDE LONGITUDE avg_humidity
##   <dbl>     <dbl>     <dbl>
## 1      25.6      88.7      75.4
```

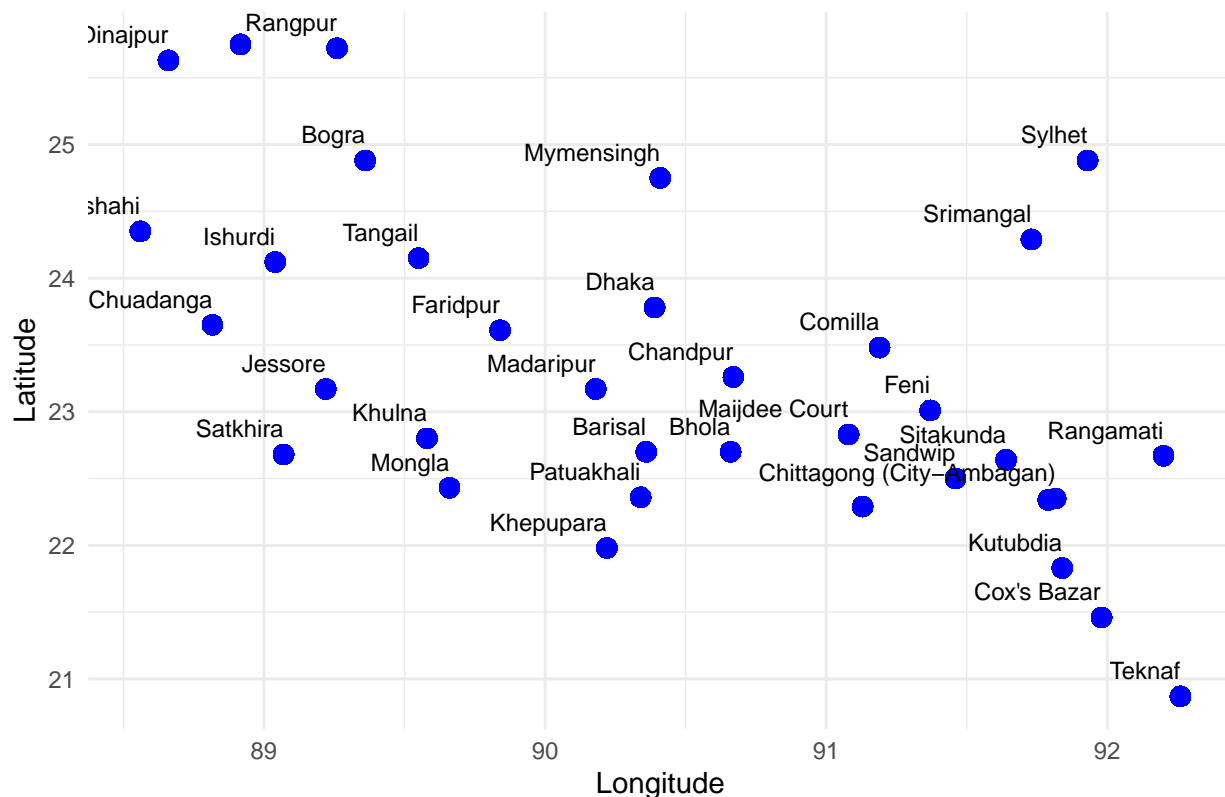
```
## Displaying Average Humidity by Latitude & Longitude
```

```
ggplot(data = avg_humidity_by_location, aes(x = LONGITUDE, y = LATITUDE, color = avg_humidity)) +
  geom_point(size = 3) +
  scale_color_gradient(low = "black", high = "red") +
  labs(title = "Average Humidity by Latitude and Longitude",
       x = "Longitude",
       y = "Latitude",
       color = "Average Humidity (%)") +
  theme_minimal()
```



```
## Displaying Station by Latitude & Longitude
ggplot(data = weather_data, aes(x = LONGITUDE, y = LATITUDE, label = Station.Names)) +
  geom_point(color = "blue", size = 3) +
  geom_text(vjust = -1, hjust = 1, size = 3, check_overlap = TRUE) + # Add station names
  labs(title = "Stations by Latitude and Longitude",
        x = "Longitude",
        y = "Latitude") +
  theme_minimal()
```

## Stations by Latitude and Longitude



### ## Analysis Interpretation

1. The highest Average Rainfall occurred in 1983
2. The Lowest Average Rainfall occurred in 1957
3. July was the most rainy month throughout 65 Years
4. January was the least rainy month throughout 65 Years.
5. The most Rainfall occurred in Teknaf
6. The least Rainfall occurred in Rajshahi
7. From the graph of "Maximum Temperature Trends over the Years" it is noticeable that from 1948 to 1960 maximum temperature was increasing. But after 1960 it declined a little bit. But from 1975 it is increasing Rapidly.
8. From the graph of "Minimum Temperature Trends Over the years" it is noticeable that temperature is increasing gradually.
9. In 2010 , the Average Temperature was highest
10. In 1971, the Average Temperature was lowest.
11. In January, the Average Temperature was lowest throughout 65 Years
12. In June, the Average Temperature was highest throughout 65 Years
13. The Highest Average Temperature was in Mongla
14. The Lowest Average Temperature was in Srimangal
15. The Average Humidity was Highest in 1998
16. The Average Humidity was Lowest in 1982
17. The Average Humidity was highest in July
18. The Average Humidity was lowest in March
19. The Average Humidity was Highest in Bhola
20. The Average Humidity was lowest in Dinajpur
21. The Highest Average Wind Speed was in 2012

22. The Lowest Average Wind Speed was in 2008
23. The Highest Average Wind Speed was in April
24. The Lowest Average Wind Speed was in November
25. The Highest Average Wind Speed was in 22.4° Latitude & 91.8° Longitude
26. The Lowest Average Wind Speed was in 23.2° Latitude & 90.2° Longitude
27. The Highest Average Temperature was in 22.4° Latitude & 89.7° Longitude
28. The Lowest Average Temperature was in 24.3° Latitude & 91.7° Longitude
29. The Highest Average Rainfall was in 20.9° Latitude & 92.3° Longitude
30. The Lowest Average Rainfall was in 24.4° Latitude & 88.6° Longitude
31. The Highest Average Humidity was in 22.7° Latitude & 90.7° Longitude
32. The Lowest Average Humidity was in 25.6° Latitude & 88.7° Longitude