weather

Gözde Nur Özdemir

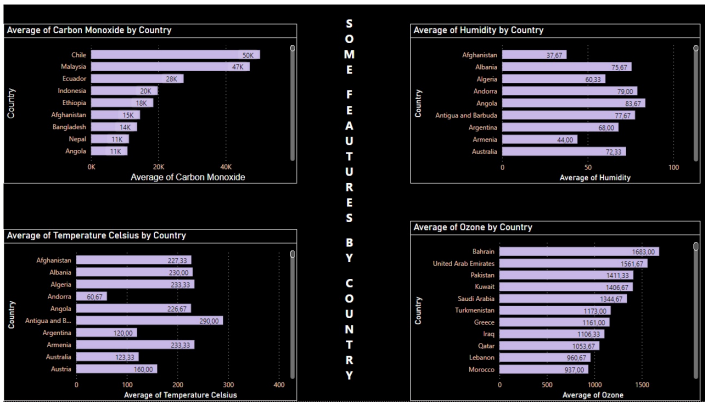
2023-09-01

WEATHER DATA SET

(https://www.yandex.com.tr/gorsel/search?from=tabbar&img_url=https%3A%2F%2Fcatherineasquithgallery.com%2Fuploads%2Fposts%2F2021-02%2F1613686171_41-p-fon-dlya-prezentatsii-pogoda-43.png&lr=104782&pos=18&rpt=simage&text=weather)

This data set was taken from Kaggle (https://www.kaggle.com/datasets/nelgiriyewithana/global-weather-repository?select=GlobalWeatherRepository.csv)

Power BI



This graphs are made by Power BI

Preperation

```
# Please first download required library
library(tidyverse)
library(ggplot2)
library(dplyr)
library(rvest)
library(stringr)
library(corrplot)
```

General Information About Data Set

```
dt<-read.csv("C:/Users/gozde/Desktop/wheather/GlobalWeatherRepository.csv")
set.seed(292)
rand_number_1 <- sample(x = 1:585, size = 200)
dt<-dt[rand_number_1,1:41]
summary(dt)
```

```
##    country          location_name         latitude       longitude
##  Length:200         Length:200         Min.   :-41.30   Min.   :-120.49
##  Class :character   Class :character   1st Qu.:  6.15   1st Qu.: -18.18
##  Mode  :character   Mode  :character   Median : 15.59   Median :  18.83
##                                        Mean   : 18.89   Mean   :  16.40
##                                        3rd Qu.: 40.63   3rd Qu.:  44.79
##                                        Max.   : 63.83   Max.   : 179.22
##    timezone         last_updated_epoch  last_updated       temperature_celsius
##  Length:200         Min.   :1.693e+09   Length:200         Min.   : 5.00
##  Class :character   1st Qu.:1.693e+09   Class :character   1st Qu.:18.77
##  Mode  :character   Median :1.693e+09   Mode  :character   Median :24.70
##                     Mean   :1.693e+09                      Mean   :23.27
##                     3rd Qu.:1.693e+09                      3rd Qu.:28.55
##                     Max.   :1.693e+09                      Max.   :42.00
##  temperature_fahrenheit condition_text        wind_mph         wind_kph
##  Min.   : 41.00         Length:200         Min.   : 2.200   Min.   : 3.60
##  1st Qu.: 65.78         Class :character   1st Qu.: 3.800   1st Qu.: 6.10
##  Median : 76.45         Mode  :character   Median : 6.500   Median :10.45
##  Mean   : 73.89                            Mean   : 7.444   Mean   :11.97
##  3rd Qu.: 83.40                            3rd Qu.: 9.400   3rd Qu.:15.10
##  Max.   :107.60                            Max.   :30.000   Max.   :48.20
##   wind_degree    wind_direction      pressure_mb    pressure_in
##  Min.   : 10.0   Length:200         Min.   :1000   Min.   :29.53
##  1st Qu.: 90.0   Class :character   1st Qu.:1007   1st Qu.:29.74
##  Median :176.5   Mode  :character   Median :1011   Median :29.85
##  Mean   :172.2                      Mean   :1011   Mean   :29.86
##  3rd Qu.:240.0                      3rd Qu.:1014   3rd Qu.:29.94
##  Max.   :360.0                      Max.   :1034   Max.   :30.53
##    precip_mm         precip_in         humidity         cloud
##  Min.   : 0.0000   Min.   :0.0000   Min.   : 12.0   Min.   :  0.00
##  1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.: 56.0   1st Qu.:  6.00
##  Median : 0.0000   Median :0.0000   Median : 78.5   Median : 32.00
##  Mean   : 0.4585   Mean   :0.0177   Mean   : 71.7   Mean   : 39.55
##  3rd Qu.: 0.1000   3rd Qu.:0.0000   3rd Qu.: 89.0   3rd Qu.: 75.00
##  Max.   :28.7000   Max.   :1.1300   Max.   :100.0   Max.   :100.00
##  feels_like_celsius feels_like_fahrenheit visibility_km   visibility_miles
##  Min.   : 3.60      Min.   : 38.40        Min.   : 2.00   Min.   : 1.000
##  1st Qu.:18.77      1st Qu.: 65.78        1st Qu.:10.00   1st Qu.: 6.000
##  Median :26.55      Median : 79.80        Median :10.00   Median : 6.000
##  Mean   :25.21      Mean   : 77.38        Mean   :10.21   Mean   : 6.055
##  3rd Qu.:31.35      3rd Qu.: 88.47        3rd Qu.:10.00   3rd Qu.: 6.000
##  Max.   :59.10      Max.   :138.30        Max.   :24.00   Max.   :14.000
##     uv_index        gust_mph        gust_kph     air_quality_Carbon_Monoxide
##  Min.   :1.000   Min.   : 1.10   Min.   : 1.80   Min.   : 140.2
##  1st Qu.:1.000   1st Qu.: 6.00   1st Qu.: 9.70   1st Qu.: 219.5
##  Median :1.000   Median : 9.95   Median :16.00   Median : 270.4
##  Mean   :2.965   Mean   :11.22   Mean   :18.06   Mean   : 441.7
##  3rd Qu.:6.000   3rd Qu.:14.50   3rd Qu.:23.40   3rd Qu.: 415.6
##  Max.   :9.000   Max.   :49.40   Max.   :79.60   Max.   :7370.0
##  air_quality_Ozone air_quality_Nitrogen_dioxide air_quality_Sulphur_dioxide
##  Min.   :  0.00    Min.   : 0.000               Min.   : 0.000
##  1st Qu.: 18.95    1st Qu.: 0.775               1st Qu.: 0.300
##  Median : 36.30    Median : 3.300               Median : 1.200
##  Mean   : 43.70    Mean   : 8.845               Mean   : 5.965
##  3rd Qu.: 59.05    3rd Qu.: 8.825               3rd Qu.: 5.225
##  Max.   :277.50    Max.   :98.700               Max.   :76.300
##  air_quality_PM2.5 air_quality_PM10 air_quality_us.epa.index
##  Min.   :  0.500   Min.   :  0.50   Min.   :1.00
##  1st Qu.:  2.975   1st Qu.:  4.70   1st Qu.:1.00
##  Median :  6.450   Median :  9.75   Median :1.00
##  Mean   : 16.716   Mean   : 26.00   Mean   :1.37
##  3rd Qu.: 14.025   3rd Qu.: 22.30   3rd Qu.:1.00
##  Max.   :496.200   Max.   :596.90   Max.   :6.00
##  air_quality_gb.defra.index   sunrise            sunset
##  Min.   : 1.00              Length:200         Length:200
##  1st Qu.: 1.00              Class :character   Class :character
##  Median : 1.00              Mode  :character   Mode  :character
##  Mean   : 1.82
##  3rd Qu.: 2.00
##  Max.   :10.00
##    moonrise           moonset           moon_phase        moon_illumination
##  Length:200         Length:200         Length:200         Min.   :93.00
##  Class :character   Class :character   Class :character   1st Qu.:93.00
##  Mode  :character   Mode  :character   Mode  :character   Median :98.00
##                                                           Mean   :96.25
##                                                           3rd Qu.:98.00
##                                                           Max.   :98.00
```

- **country**: Country of the weather data
- **location_name**: Name of the location (city)
- **latitude**: Latitude coordinate of the location
- **longitude**: Longitude coordinate of the location
- **timezone**: Timezone of the location
- **last_updated_epoch**: Unix timestamp of the last data update
- **last_updated**: Local time of the last data update
- **temperature_celsius**: Temperature in degrees Celsius
- **temperature_fahrenheit**: Temperature in degrees Fahrenheit
- **condition_text**: Weather condition description
- **wind_mph**: Wind speed in miles per hour
- **wind_kph**: Wind speed in kilometers per hour
- **wind_degree**: Wind direction in degrees
- **wind_direction**: Wind direction as a 16-point compass
- **pressure_mb**: Pressure in millibars
- **pressure_in**: Pressure in inches
- **precip_mm**: Precipitation amount in millimeters
- **precip_in**: Precipitation amount in inches
- **humidity**: Humidity as a percentage
- **cloud**: Cloud cover as a percentage
- **feels_like_celsius**: Feels-like temperature in Celsius
- **feels_like_fahrenheit**: Feels-like temperature in Fahrenheit
- **visibility_km**: Visibility in kilometers
- **visibility_miles**: Visibility in miles
- **uv_index**: UV Index
- **gust_mph**: Wind gust in miles per hour
- **gust_kph**: Wind gust in kilometers per hour
- **air_quality_Carbon_Monoxide**: Air quality measurement: Carbon Monoxide
- **air_quality_Ozone**: Air quality measurement: Ozone

- **air_quality_Nitrogen_dioxide**: Air quality measurement: Nitrogen Dioxide
- **air_quality_Sulphur_dioxide**: Air quality measurement: Sulphur Dioxide
- **air_quality_PM2.5**: Air quality measurement: PM2.5
- **air_quality_PM10**: Air quality measurement: PM10
- **air_quality_us-epa-index**: Air quality measurement: US EPA Index
- **air_quality_gb-defra-index**: Air quality measurement: GB DEFRA Index
- **sunrise**: Local time of sunrise
- **sunset**: Local time of sunset
- **moonrise**: Local time of moonrise
- **moonset**: Local time of moonset
- **moon_phase**: Current moon phase
- **moon_illumination**: Moon illumination percentage

Above information was taken from Kaggle
(country:%20Country%20of%20the%20weather%20data%20location_name:%20Name%20of%20the%20location%20(city)%20latitude:%20Latitude%20coordinate%20of%20the%20location%20longitude
point%20compass%20pressure_mb:%20Pressure%20in%20millibars%20pressure_in:%20Pressure%20in%20inches%20precip_mm:%20Precipitation%20amount%20in%20millimeters%20precip_in:%2(
like%20temperature%20in%20Celsius%20feels_like_fahrenheit:%20Feels-
like%20temperature%20in%20Fahrenheit%20visibility_km:%20Visibility%20in%20kilometers%20visibility_miles:%20Visibility%20in%20miles%20uv_index:%20UV%20Index%20gust_mph:%20Wind%20(
epa-index:%20Air%20quality%20measurement:%20US%20EPA%20Index%20air_quality_gb-defra-
index:%20Air%20quality%20measurement:%20GB%20DEFRA%20Index%20sunrise:%20Local%20time%20of%20sunrise%20sunset:%20Local%20time%20of%20sunset%20moonrise:%20Local%20tim

Plot of Weather Data

Since this data set so big to see any pattern we use small sample.

set.seed(291)

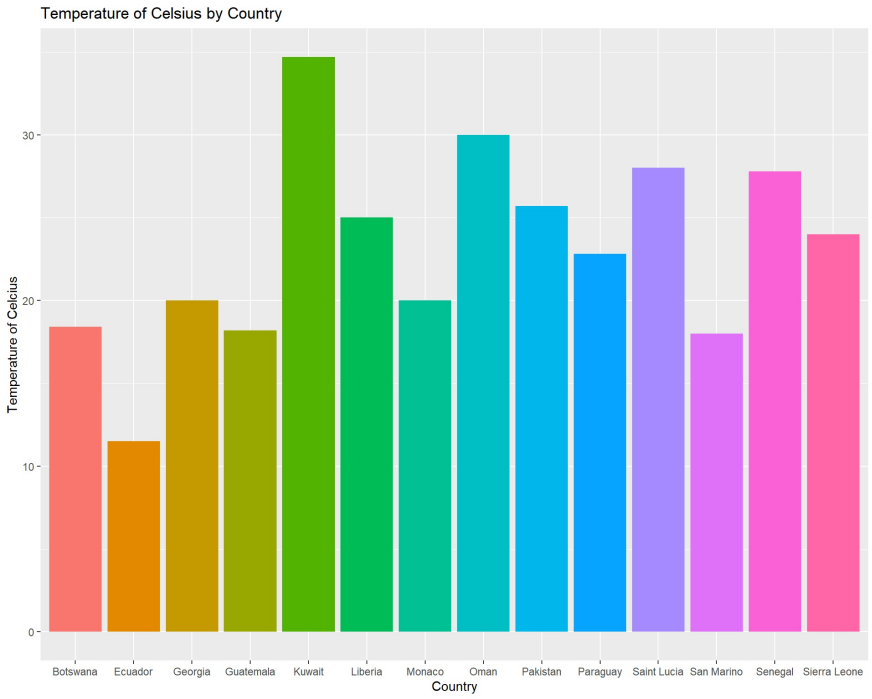
rand_number_2 <- sample(x = 1:200, size = 15)

dt_plot<-dt[rand_number_2,1:41]

```
set.seed(291)
rand_number_2 <- sample(x = 1:200, size = 15)
dt_plot<-dt[rand_number_2,1:41]
```

Temperature of Celsius by Country

```
ggplot(data = dt_plot, aes(x = country, y = temperature_celsius, fill = country)) +
  geom_bar(stat = "identity", position = "dodge",show.legend = F)+
  labs(title="Temperature of Celsius by Country")+
  xlab("Country")+ylab("Temperature of Celcius")
```

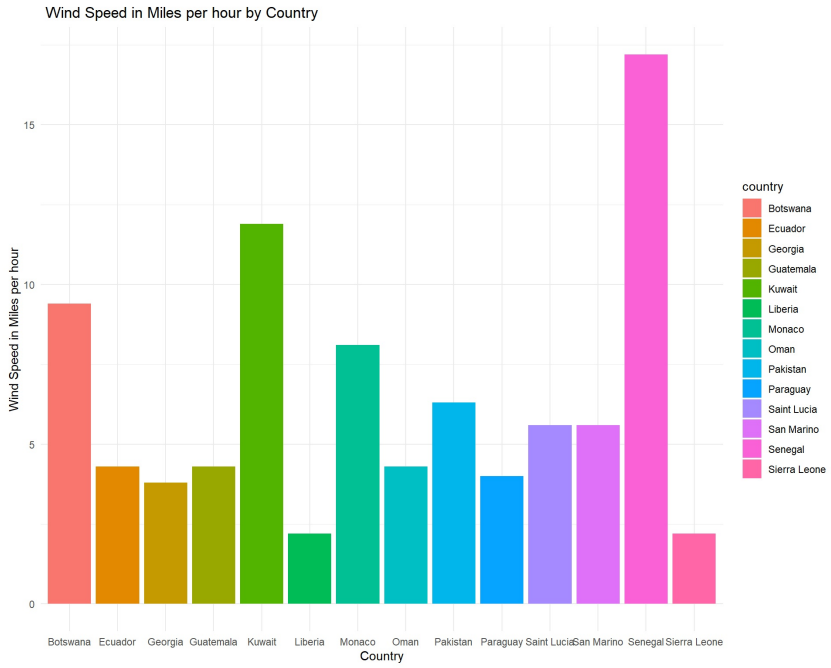


As we can understand from this graph between these 15 countries the hottest country is Kuwait. The coldest country is Ecuador.

- According to NBC News (https://www.nbcnews.com/science/environment/kuwait-worlds-hottest-places-lags-climate-action-rcna20830) : Kuwait reached a scorching temperature of 53.2 degrees Celsius (127.7 degrees Fahrenheit), making it among the hottest places on earth.

Wind Speed in Miles per hour by Country

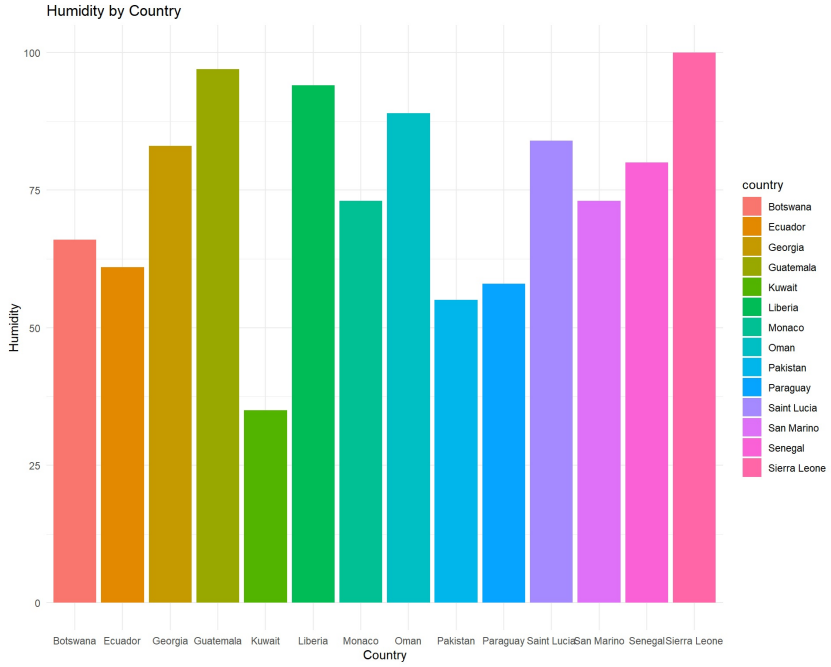
```
ggplot(data = dt_plot, aes(x = country, y = wind_mph, fill = country)) +
  geom_bar(stat = "identity", position = "dodge")+
  labs(title=" Wind Speed in Miles per hour by Country")+
  xlab("Country")+ylab("Wind Speed in Miles per hour")+theme_minimal()
```



Senegal is the windiest place between these 15 countries.

Humidity by Country

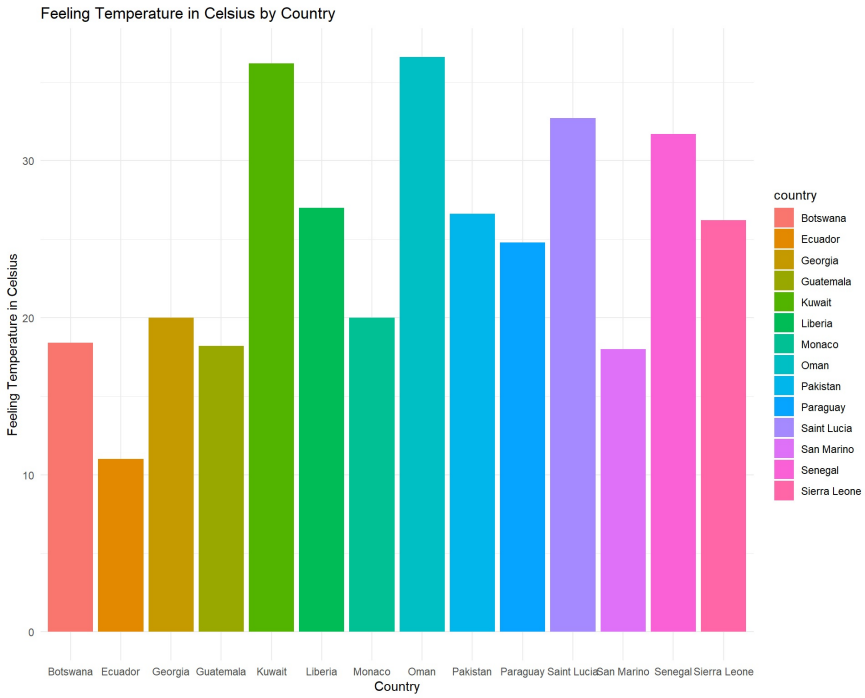
```
ggplot(data = dt_plot, aes(x = country, y = humidity, fill = country)) +
  geom_bar(stat = "identity", position = "dodge")+
  labs(title="Humidity by Country")+
  xlab("Country")+ylab("Humidity")+theme_minimal()
```



All countries are mostly high except Kuwait. Also, as we discussed, Kuwait is a hot country, so this affects the humidity.

Feels-like temperature in Celsius by Country

```
ggplot(data = dt_plot, aes(x = country, y =feels_like_celsius , fill = country)) +
  geom_bar(stat = "identity", position = "dodge")+
  labs(title="Feeling Temperature in Celsius by Country")+
  xlab("Country")+ylab("Feeling Temperature in Celsius")+theme_minimal()
```



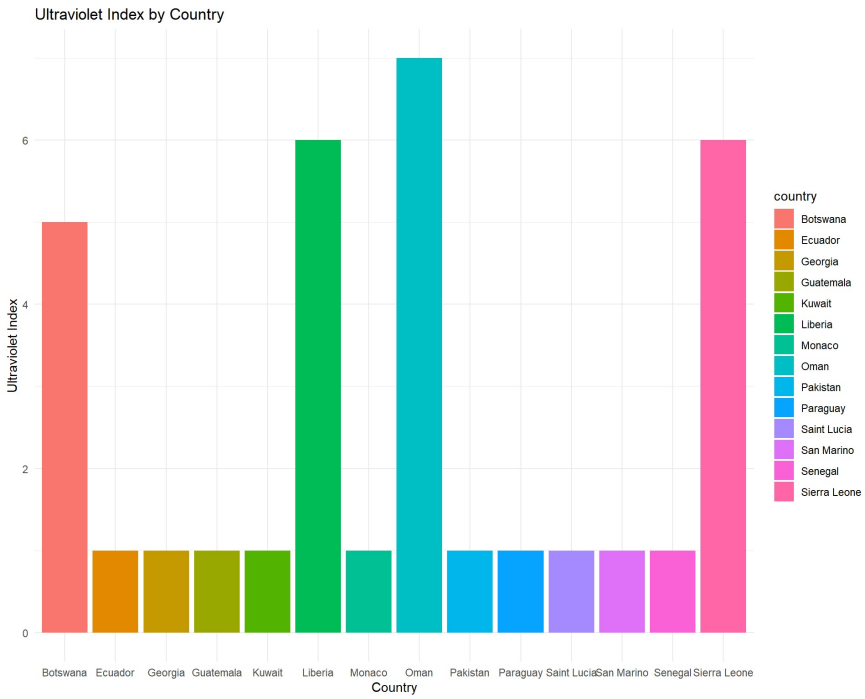
As we excepted Kuwait's feels-like temperature is high since it's humidity and normal temperature is high.

Ultraviolet Index by Country

According to the Wikipedia (https://en.wikipedia.org/wiki/Ultraviolet_index) :

The **ultraviolet index**, or **UV index**, is an international standard measurement of the strength of the sunburn (https://en.wikipedia.org/wiki/Sunburn)-producing ultraviolet (https://en.wikipedia.org/wiki/Ultraviolet) (UV) radiation (https://en.wikipedia.org/wiki/Radiation) at a particular place and time.

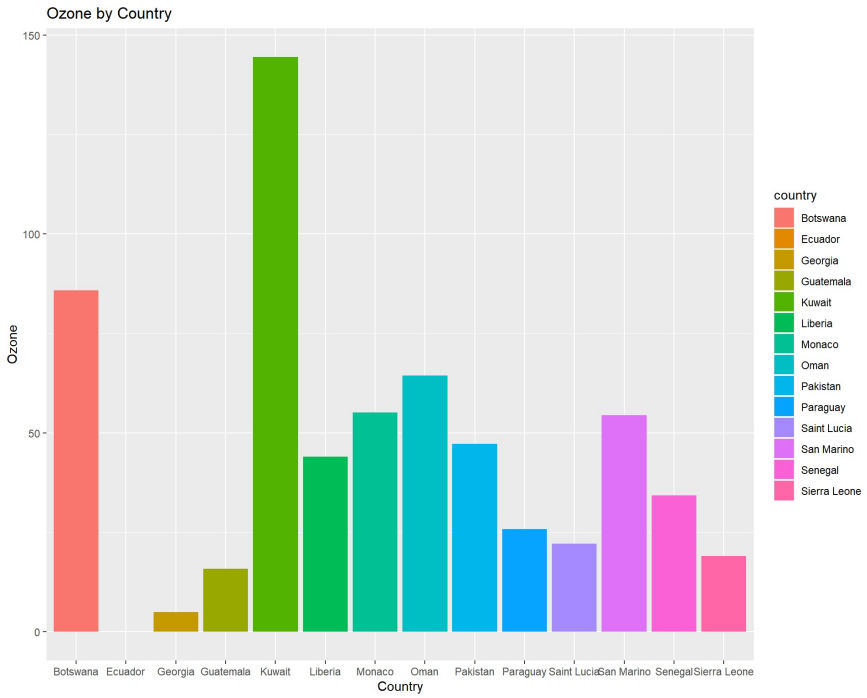
```
ggplot(data = dt_plot, aes(x = country, y =uv_index , fill = country)) +
  geom_bar(stat = "identity", position = "dodge")+
  labs(title="Ultraviolet Index by Country")+
  xlab("Country")+ylab("Ultraviolet Index")+theme_minimal()
```



Oman has the most uv_index.

Ozone by Country

```
ggplot(data = dt_plot, aes(x = country, y =air_quality_Ozone , fill = country)) +
  geom_bar(stat = "identity", position = "dodge")+
  labs(title="Ozone by Country")+
  xlab("Country")+ylab("Ozone")
```



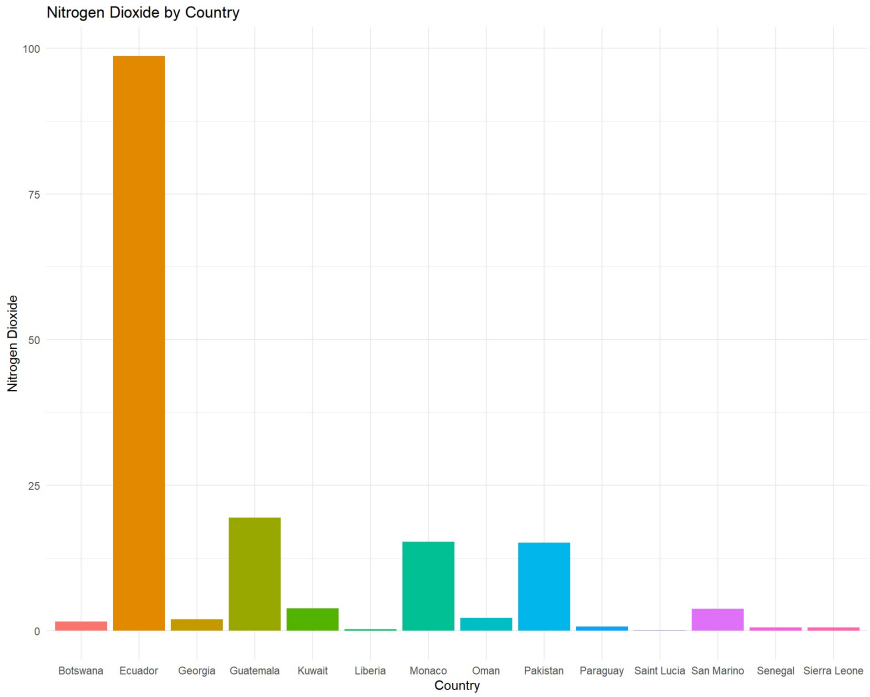
Kuwait's ozone level is really high. This is bad because ozone is really harmful for people's health.

According to the Answer More (https://www.answersmore.com/what-happens-if-there-is-too-much-ozone/) :

Ozone makes people more sensitive to allergens—the most common triggers for asthma attacks. Also, asthmatics may be more severely affected by reduced lung function and airway inflammation.

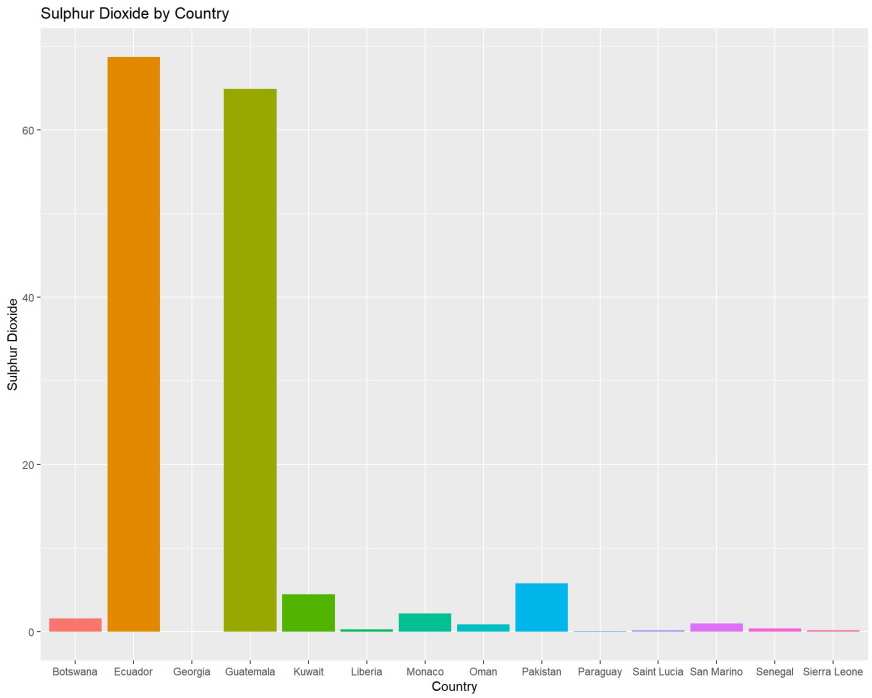
Nitrogen dioxide by Country

```
ggplot(data = dt_plot, aes(x = country, y =air_quality_Nitrogen_dioxide , fill = country)) +
  geom_bar(stat = "identity", position = "dodge",show.legend = F)+
  labs(title="Nitrogen Dioxide by Country")+
  xlab("Country")+ylab("Nitrogen Dioxide")+theme_minimal()
```



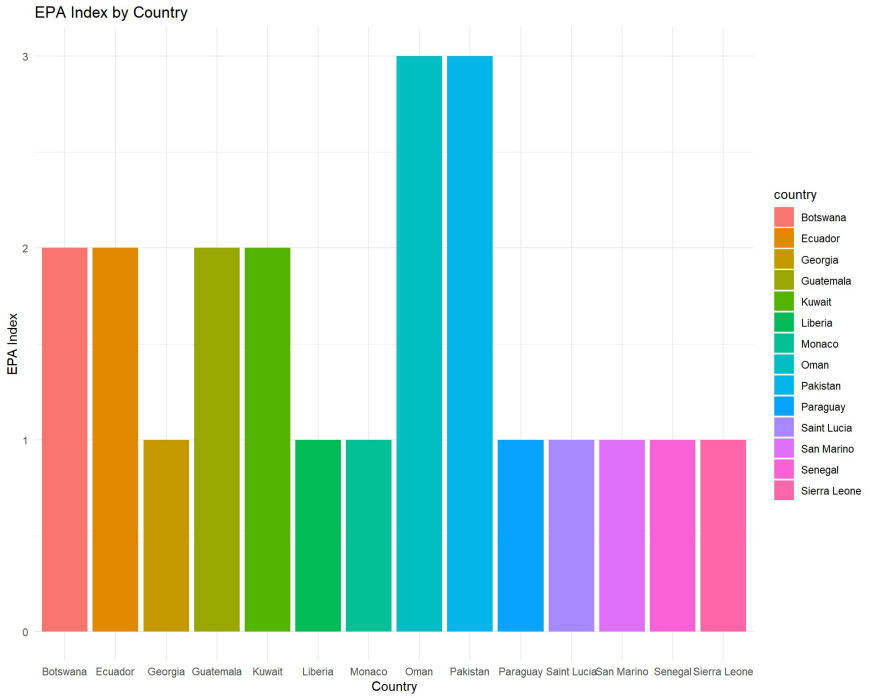
Sulphur Dioxide by Country

```
ggplot(data = dt_plot, aes(x = country, y =air_quality_Sulphur_dioxide , fill = country)) +
  geom_bar(stat = "identity", position = "dodge",show.legend = F)+
  labs(title="Sulphur Dioxide by Country")+
  xlab("Country")+ylab("Sulphur Dioxide")
```

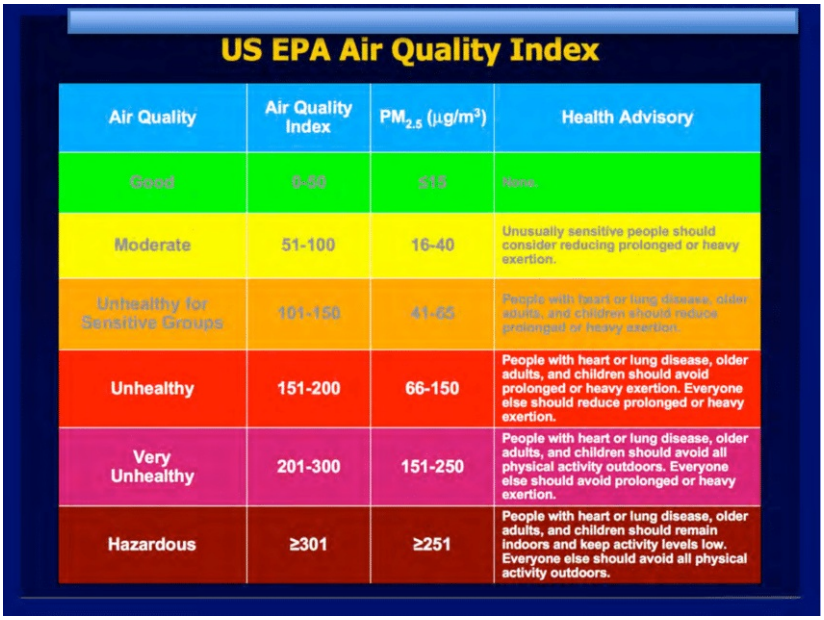


EPA Index by country

```
ggplot(data = dt_plot, aes(x = country, y = air_quality_us.epa.index, fill = country)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "EPA Index by Country") +
  xlab("Country") + ylab("EPA Index") +
  theme_minimal()
```

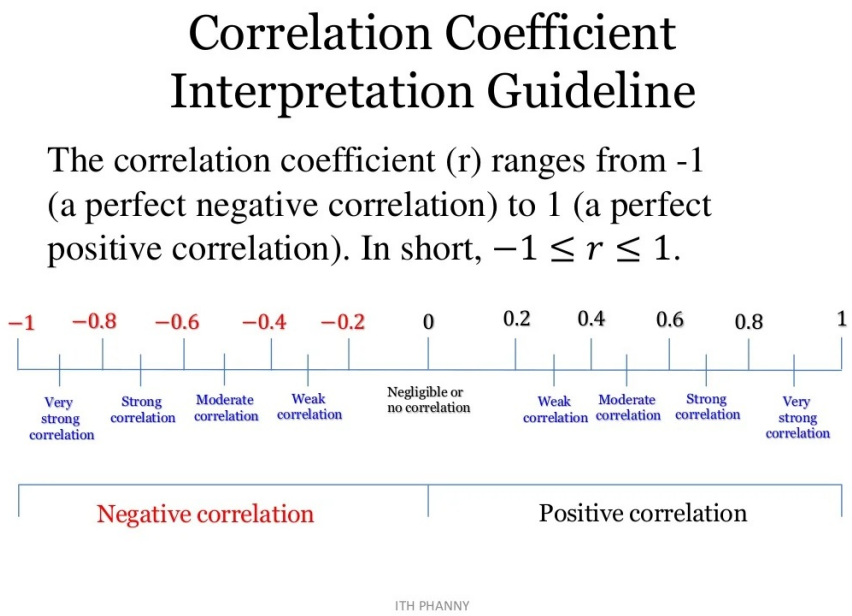


For more information about US EPA Index. Look at the below image.



Correlation matrix: correlations for all variables

If you do not have information about correlation matrix. Here is the general summary information:



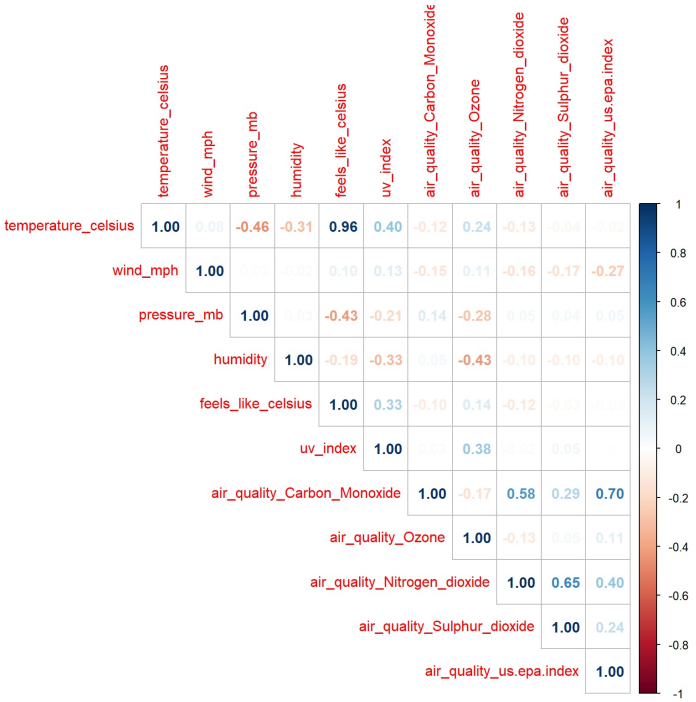
ITH PHANNY

```
dt_cor<-dt[c(8,11,15,19,21,25,28,29,30,31,34)]
round(cor(dt_cor),
  digits = 2 # rounded to 2 decimals
)
```

```
##                             temperature_celsius wind_mph pressure_mb humidity
## temperature_celsius                        1.00     0.08       -0.46    -0.31
## wind_mph                                   0.08     1.00        0.02    -0.02
## pressure_mb                               -0.46     0.02        1.00     0.03
## humidity                                  -0.31    -0.02        0.03     1.00
## feels_like_celsius                         0.96     0.10       -0.43    -0.19
## uv_index                                   0.40     0.13       -0.21    -0.33
## air_quality_Carbon_Monoxide               -0.12    -0.15        0.14     0.05
## air_quality_Ozone                          0.24     0.11       -0.28    -0.43
## air_quality_Nitrogen_dioxide              -0.13    -0.16        0.05    -0.10
## air_quality_Sulphur_dioxide               -0.04    -0.17        0.04    -0.10
## air_quality_us.epa.index                  -0.02    -0.27        0.05    -0.10
##                             feels_like_celsius uv_index
## temperature_celsius                       0.96     0.40
## wind_mph                                  0.10     0.13
## pressure_mb                              -0.43    -0.21
## humidity                                 -0.19    -0.33
## feels_like_celsius                        1.00     0.33
## uv_index                                  0.33     1.00
## air_quality_Carbon_Monoxide              -0.10     0.03
## air_quality_Ozone                         0.14     0.38
## air_quality_Nitrogen_dioxide             -0.12    -0.02
## air_quality_Sulphur_dioxide              -0.03     0.05
## air_quality_us.epa.index                 -0.02     0.01
##                             air_quality_Carbon_Monoxide air_quality_Ozone
## temperature_celsius                               -0.12              0.24
## wind_mph                                          -0.15              0.11
## pressure_mb                                        0.14             -0.28
## humidity                                           0.05             -0.43
## feels_like_celsius                                -0.10              0.14
## uv_index                                           0.03              0.38
## air_quality_Carbon_Monoxide                        1.00             -0.17
## air_quality_Ozone                                 -0.17              1.00
## air_quality_Nitrogen_dioxide                       0.58             -0.13
## air_quality_Sulphur_dioxide                        0.29              0.05
## air_quality_us.epa.index                           0.70              0.11
##                             air_quality_Nitrogen_dioxide
## temperature_celsius                            -0.13
## wind_mph                                       -0.16
## pressure_mb                                     0.05
## humidity                                       -0.10
## feels_like_celsius                             -0.12
## uv_index                                       -0.02
## air_quality_Carbon_Monoxide                     0.58
## air_quality_Ozone                              -0.13
## air_quality_Nitrogen_dioxide                    1.00
## air_quality_Sulphur_dioxide                     0.65
## air_quality_us.epa.index                        0.40
##                             air_quality_Sulphur_dioxide
## temperature_celsius                           -0.04
## wind_mph                                      -0.17
## pressure_mb                                    0.04
## humidity                                      -0.10
## feels_like_celsius                            -0.03
## uv_index                                       0.05
## air_quality_Carbon_Monoxide                    0.29
## air_quality_Ozone                              0.05
## air_quality_Nitrogen_dioxide                   0.65
## air_quality_Sulphur_dioxide                    1.00
## air_quality_us.epa.index                       0.24
##                             air_quality_us.epa.index
## temperature_celsius                        -0.02
## wind_mph                                   -0.27
## pressure_mb                                 0.05
## humidity                                   -0.10
## feels_like_celsius                         -0.02
## uv_index                                    0.01
## air_quality_Carbon_Monoxide                 0.70
## air_quality_Ozone                           0.11
## air_quality_Nitrogen_dioxide                0.40
## air_quality_Sulphur_dioxide                 0.24
## air_quality_us.epa.index                    1.00
```

To be more easy to read, let's make as a plot.

```
corrplot(cor(dt_cor),
  method = "number",
  type = "upper" # show only upper side
)
```



Time to interpretation:

As we can guess there is a high positive correlation between temperature Celsius and feels like Celsius.

Also, Carbon monoxide and the US EPA Index are highly positively correlated to each other.

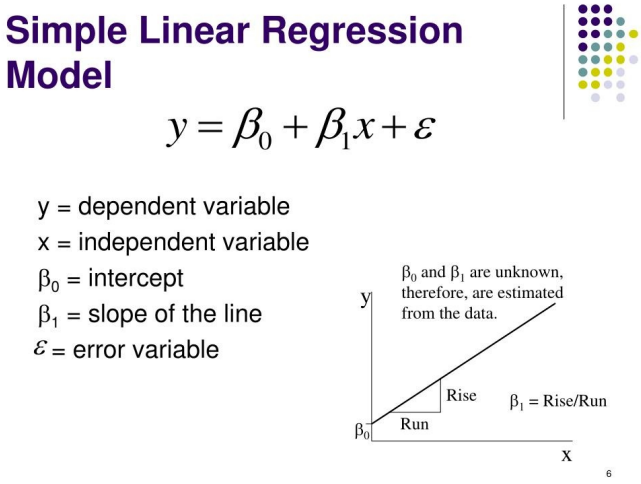
Carbon monoxide and nitrogen dioxide are moderately positively correlated to each other.

Pressure mb and feels like Celsius are moderately negatively correlated to each other.

Also, Ozone and humidity are moderately negatively correlated.

Linear Regression

If you do not have information about linear regression. Here is the general summary information:



Once we have identified two variables that are correlated, we would like to model this relationship. We want to use one variable as a **predictor** or **explanatory** variable to explain the other variable, the **response** or **dependent** variable. In order to do this, we need a good relationship between our two variables. The model can then be used to predict changes in our response variable. A strong relationship between the predictor variable and the response variable leads to a good model.

Ozone and Humidity

```
linear1 <- lm( air_quality_Ozone~ humidity, data=dt_plot)
linear1
```

```
##
## Call:
## lm(formula = air_quality_Ozone ~ humidity, data = dt_plot)
##
## Coefficients:
## (Intercept)     humidity
##     120.740       -1.072
```

Ozone=120.740+(-1.072)*Humidity

Carbon Monoxide and Nitrogen Dioxide

```
linear2 <- lm( air_quality_Carbon_Monoxide~air_quality_Nitrogen_dioxide , data=dt_plot)
linear2
```

```
##
## Call:
## lm(formula = air_quality_Carbon_Monoxide ~ air_quality_Nitrogen_dioxide,
##     data = dt_plot)
##
## Coefficients:
##                   (Intercept)  air_quality_Nitrogen_dioxide
##                        153.10                         40.41
```

Carbon Monoxide=153.10+40.41*Nitrogen Dioxide

Pressure mb and Feels Like Celsius

```
linear3 <- lm( pressure_mb~ feels_like_celsius, data=dt_plot)
linear3
```

```
##
## Call:
## lm(formula = pressure_mb ~ feels_like_celsius, data = dt_plot)
##
## Coefficients:
##         (Intercept)  feels_like_celsius
##           1027.0463             -0.5809
```

Pressure mb=1027.0463-0.5809*(Feels Like Celsius)

Carbon Monoxide and the US EPA Index

```
linear4 <- lm( air_quality_us.epa.index~ air_quality_Carbon_Monoxide, data=dt_plot)
linear4
```

```
##
## Call:
## lm(formula = air_quality_us.epa.index ~ air_quality_Carbon_Monoxide,
##     data = dt_plot)
##
## Coefficients:
##                 (Intercept)  air_quality_Carbon_Monoxide
##                   1.4165158                    0.0001949
```

EPA Index=1.4165+(0.00019)*Carbon Monoxide

Temperature Celsius and Feels Like Celsius

```
linear5 <- lm( feels_like_celsius~temperature_celsius , data=dt_plot)
linear5
```

```
##
## Call:
## lm(formula = feels_like_celsius ~ temperature_celsius, data = dt_plot)
##
## Coefficients:
##         (Intercept)  temperature_celsius
##              -3.469                1.224
```

Feels Like Celsius=(-3.469)+1.224*Temperature Celsius