Documentation : output of console window

Note: The script will set the working directory to the R file which is in the zip file. The script searches for the folder data where the R file is extracted. (script and folder with name data which contains all csv inside to be in the same folder).

I have not used relative path as the number of lines of code increases. And also, I have used the code to read all csv files from the "data" folder at one go. (which appends data at one go to a single tibble).

Preparing R environment: setting working directory and removing existing environment variables etc.

```
> rm(list = ls())
> #setting the working directory for the current script to execute
> #getting the directory of the current script
> #this following has to be executed twice if the script has not invoked R studi
o.
> #hence executing twice to set the working directory to the script path
> for(i in 1:2){
+   Wrkng_Dir <- setwd(dirname(rstudioapi::getActiveDocumentContext()$path))}
> Wrkng_Dir
[1] "D:/DataScience/IntroDsc/Assignment"
> #Appending data files folder to the path
> dataFile_Path <- file.path(Wrkng_Dir, "data")
> dataFile_Path
[1] "D:/DataScience/IntroDsc/Assignment/data"
> #getting the file names under InputData folder of captured data
> filenames <- list.files(dataFile_Path)
> filenames
 [1] "201808.csv" "201809.csv" "201810.csv" "201811.csv" "201812.csv"
 [6] "201901.csv" "201902.csv" "201903.csv" "201904.csv" "201905.csv"
[11] "201906.csv" "201907.csv" "201908.csv" "201909.csv" "201910.csv"
[16] "201911.csv" "201912.csv" "202001.csv" "202002.csv"
> #calculating the number of files in InputData folder
> numfiles <- length(filenames)
> numfiles
[1] 19
>
```

PART-A:

1. Loading all csv files.
2. Concatenating to one data frame.

```
> tblACT_wthr
# A tibble: 578 x 21
   Date  `Minimum temper~ `Maximum temper~ `Rainfall (mm)` `Evaporation (m~
   <chr> <chr>            <chr>            <chr>           <chr>
 1 1/08~ 7.6              15.4             0               NA
 2 2/08~ -3.8             14.3             0               NA
 3 3/08~ -3.6             19.5             0               NA
 4 4/08~ 3.7              12.8             13.8            NA
 5 5/08~ -1               15               0               NA
 6 6/08~ 1.2              13.7             0               NA
 7 7/08~ 2.4              9.7              6.6             NA
 8 8/08~ 2.6              12.1             0               NA
 9 9/08~ 1.6              13.7             0               NA
10 10/0~ -2.5             15.6             0.2             NA
# ... with 568 more rows, and 16 more variables: `Sunshine (hours)` <chr>,
#   `Direction of maximum wind gust` <chr>, `Speed of maximum wind gust
#   (km/h)` <chr>, `Time of maximum wind gust` <chr>, `9am
#   Temperature` <chr>, `9am relative humidity (%)` <chr>, `9am cloud amount
#   (oktas)` <chr>, `9am wind direction` <chr>, `9am wind speed
#   (km/h)` <chr>, `9am MSL pressure (hPa)` <chr>, `3pm Temperature` <chr>,
#   `3pm relative humidity (%)` <chr>, `3pm cloud amount (oktas)` <chr>, `3pm
#   wind direction` <chr>, `3pm wind speed (km/h)` <chr>, `3pm MSL pressure
#   (hPa)` <chr>
>
```

##

3. Check for problems while loading.
   "calm" is replaced in 2 columns(9am and 3pm wind speed) with value 0.
   And class of columns are changed to numeric and double accordingly.

```
> #3. check for problems while loading
> #"calm" value in 9am and 3pm wind speed replaced with 0 km/hr
> tblACT_wthr$`9am wind speed (km/h)` <- lapply(tblACT_wthr$`9am wind speed (km/
h)`, gsub, pattern = "Calm", replacement = 0, fixed = TRUE)
> tblACT_wthr$`3pm wind speed (km/h)` <- lapply(tblACT_wthr$`3pm wind speed (km/
h)`, gsub, pattern = "Calm", replacement = 0, fixed = TRUE)
> |
```

```
> sapply(tblACT_wthr, class)
                                Date                 Minimum temperature
                         "character"                           "numeric"
                Maximum temperature                       Rainfall (mm)
                           "numeric"                           "numeric"
                   Evaporation (mm)                     Sunshine (hours)
                         "character"                         "character"
   Direction of maximum wind gust Speed of maximum wind gust (km/h)
                         "character"                           "numeric"
         Time of maximum wind gust                    9am Temperature
                         "character"                           "numeric"
         9am relative humidity (%)          9am cloud amount (oktas)
                           "numeric"                           "numeric"
                  9am wind direction             9am wind speed (km/h)
                         "character"                           "numeric"
             9am MSL pressure (hPa)                    3pm Temperature
                           "numeric"                           "numeric"
         3pm relative humidity (%)          3pm cloud amount (oktas)
                           "numeric"                           "numeric"
                  3pm wind direction             3pm wind speed (km/h)
                         "character"                           "numeric"
             3pm MSL pressure (hPa)
                           "numeric"
```

##

PART-B:

1. Removing variables with all NA. 2 columns with all NA removed out of 21 columns.

```
> tblACT_wthr <- tblACT_wthr %>% select_if(all_na)
> tblACT_wthr
# A tibble: 578 x 19
   Date   `Minimum temper~ `Maximum temper~ `Rainfall (mm)` `Direction of m~
   <chr>             <dbl>            <dbl>           <dbl> <chr>
 1 1/08~               7.6             15.4               0 NW
 2 2/08~              -3.8             14.3               0 NNW
 3 3/08~              -3.6             19.5               0 NW
 4 4/08~               3.7             12.8              13.8 NNW
 5 5/08~              -1               15                 0 NW
 6 6/08~               1.2             13.7               0 NW
 7 7/08~               2.4              9.7               6.6 WNW
 8 8/08~               2.6             12.1               0 WNW
 9 9/08~               1.6             13.7               0 NNW
10 10/0~              -2.5             15.6               0.2 NNW
# ... with 568 more rows, and 14 more variables: `Speed of maximum wind gust
#   (km/h)` <dbl>, `Time of maximum wind gust` <chr>, `9am
#   Temperature` <dbl>, `9am relative humidity (%)` <dbl>, `9am cloud amount
#   (oktas)` <dbl>, `9am wind direction` <chr>, `9am wind speed
#   (km/h)` <dbl>, `9am MSL pressure (hPa)` <dbl>, `3pm Temperature` <dbl>,
#   `3pm relative humidity (%)` <dbl>, `3pm cloud amount (oktas)` <dbl>, `3pm
#   wind direction` <chr>, `3pm wind speed (km/h)` <dbl>, `3pm MSL pressure
#   (hPa)` <dbl>
> |
```

##

2. Removing columns with more than 90% NA: There are no such columns found.

```
> tblActwthr_tidied
# A tibble: 578 x 19
   Date   `Minimum temper~ `Maximum temper~ `Rainfall (mm)` `Direction of m~
   <chr>             <dbl>            <dbl>           <dbl> <chr>
 1 1/08~               7.6             15.4               0 NW
 2 2/08~              -3.8             14.3               0 NNW
 3 3/08~              -3.6             19.5               0 NW
 4 4/08~               3.7             12.8              13.8 NNW
 5 5/08~              -1               15                 0 NW
 6 6/08~               1.2             13.7               0 NW
 7 7/08~               2.4              9.7               6.6 WNW
 8 8/08~               2.6             12.1               0 WNW
 9 9/08~               1.6             13.7               0 NNW
10 10/0~              -2.5             15.6               0.2 NNW
# ... with 568 more rows, and 14 more variables: `Speed of maximum wind gust
#   (km/h)` <dbl>, `Time of maximum wind gust` <chr>, `9am
#   Temperature` <dbl>, `9am relative humidity (%)` <dbl>, `9am cloud amount
#   (oktas)` <dbl>, `9am wind direction` <chr>, `9am wind speed
#   (km/h)` <dbl>, `9am MSL pressure (hPa)` <dbl>, `3pm Temperature` <dbl>,
#   `3pm relative humidity (%)` <dbl>, `3pm cloud amount (oktas)` <dbl>, `3pm
#   wind direction` <chr>, `3pm wind speed (km/h)` <dbl>, `3pm MSL pressure
#   (hPa)` <dbl>
> |
```

3. Replacing column names with space to underscore.

```
> #3. replacing column names with space to underscore
> names(tblActwthr_tidied) <- gsub(" ", "_", names(tblActwthr_tidied))
> colnames(tblActwthr_tidied)
 [1] "Date"                                "Minimum_temperature"
 [3] "Maximum_temperature"                 "Rainfall_(mm)"
 [5] "Direction_of_maximum_wind_gust"      "Speed_of_maximum_wind_gust_(km/h)"
 [7] "Time_of_maximum_wind_gust"           "9am_Temperature"
 [9] "9am_relative_humidity_(%)"           "9am_cloud_amount_(oktas)"
[11] "9am_wind_direction"                  "9am_wind_speed_(km/h)"
[13] "9am_MSL_pressure_(hPa)"              "3pm_Temperature"
[15] "3pm_relative_humidity_(%)"           "3pm_cloud_amount_(oktas)"
[17] "3pm_wind_direction"                  "3pm_wind_speed_(km/h)"
[19] "3pm_MSL_pressure_(hPa)"
>
```

##

4. Changing date from char type to "date" data type.

```
> #4.Changing date from char type to date data type
> tblActwthr_tidied$Date <- as.Date(tblActwthr_tidied$Date,format="%d/%m/%Y")
> class(tblActwthr_tidied$Date)
[1] "Date"
>
```

##

5. Creating Month and Year column out of Date column. And relocating month and year next to Date column.

```
> #5. extracting date column and adding month and year column
> tblActwthr_tidied$Month <- as.numeric(format(as.Date(tblActwthr_tidied$Date), "%m"))
> tblActwthr_tidied$Year <- as.numeric(format(as.Date(tblActwthr_tidied$Date), "%Y"))
> tblActwthr_tidied <- tblActwthr_tidied[c(1,20,21,2,3,4,5,6,7,8,9,10,11,12,
+                                          13,14,15,16,17,18,19)]
> head(tblActwthr_tidied)
# A tibble: 6 x 21
  Date       Month  Year Minimum_tempera~ Maximum_tempera~ `Rainfall_(mm)`
  <date>     <db7> <db7>           <db7>           <db7>            <db7>
1 2018-08-01     8  2018             7.6            15.4                0
2 2018-08-02     8  2018            -3.8            14.3                0
3 2018-08-03     8  2018            -3.6            19.5                0
4 2018-08-04     8  2018             3.7            12.8               13.8
5 2018-08-05     8  2018            -1              15                  0
6 2018-08-06     8  2018             1.2            13.7                0
# ... with 15 more variables: Direction_of_maximum_wind_gust <chr>,
#   `Speed_of_maximum_wind_gust_(km/h)` <db7>, Time_of_maximum_wind_gust <chr>,
#   `9am_Temperature` <db7>, `9am_relative_humidity_(%)` <db7>,
#   `9am_cloud_amount_(oktas)` <db7>, `9am_wind_direction` <chr>,
#   `9am_wind_speed_(km/h)` <db7>, `9am_MSL_pressure_(hPa)` <db7>,
#   `3pm_Temperature` <db7>, `3pm_relative_humidity_(%)` <db7>,
#   `3pm_cloud_amount_(oktas)` <db7>, `3pm_wind_direction` <chr>,
#   `3pm_wind_speed_(km/h)` <db7>, `3pm_MSL_pressure_(hPa)` <db7>
>
```

##

6. Month and Year column changed to ordinals.

```
> #6. Month and Year column changed to ordinals
> tblActwthr_tidied$Month <- factor(tblActwthr_tidied$Month, levels = c(1:12))
> tblActwthr_tidied$Year <- factor(tblActwthr_tidied$Year, levels = c(2018, 2019 , 2020), or
dered = FALSE)
> class(tblActwthr_tidied$Month)
[1] "factor"
> class(tblActwthr_tidied$Year)
[1] "factor"
>
```

```
> head(tblActwthr_tidied$Month)
[1] 8 8 8 8 8 8
Levels: 1 2 3 4 5 6 7 8 9 10 11 12
> head(tblActwthr_tidied$Year)
[1] 2018 2018 2018 2018 2018 2018
Levels: 2018 2019 2020
>
```

##

7. Imputation done for all numeric columns. Replaced NA with median values.

```
> for (cols in which(sapply(tblActwthr_tidied, is.numeric))) {
+   for (row in which(is.na(tblActwthr_tidied[, cols]))) {
+     tblActwthr_tidied[row, cols] <- median(tblActwthr_tidied[[cols]],  na.rm = TRUE)
+   }
+ }
> head(tblActwthr_tidied)
# A tibble: 6 x 21
  Date       Month Year  Minimum_tempera~ Maximum_tempera~ `Rainfall_(mm)`
  <date>     <fct> <fct>            <dbl>            <dbl>            <dbl>
1 2018-08-01 8     2018               7.6             15.4                0
2 2018-08-02 8     2018              -3.8             14.3                0
3 2018-08-03 8     2018              -3.6             19.5                0
4 2018-08-04 8     2018               3.7             12.8               13.8
5 2018-08-05 8     2018              -1               15                  0
6 2018-08-06 8     2018               1.2             13.7                0
# ... with 15 more variables: Direction_of_maximum_wind_gust <chr>,
#   `Speed_of_maximum_wind_gust_(km/h)` <dbl>, Time_of_maximum_wind_gust <chr>,
#   `9am_Temperature` <dbl>, `9am_relative_humidity_(%)` <dbl>,
#   `9am_cloud_amount_(oktas)` <dbl>, `9am_wind_direction` <chr>,
#   `9am_wind_speed_(km/h)` <dbl>, `9am_MSL_pressure_(hPa)` <dbl>,
#   `3pm_Temperature` <dbl>, `3pm_relative_humidity_(%)` <dbl>,
#   `3pm_cloud_amount_(oktas)` <dbl>, `3pm_wind_direction` <chr>,
#   `3pm_wind_speed_(km/h)` <dbl>, `3pm_MSL_pressure_(hPa)` <dbl>
> |
```

##

PART-C

1. Min median mean max for specified columns. Min,median,max and average for specified variables at one go.

```
> stats_tblActWhthr <- tblActwthr_tidied %>%
+   summarise(min_mntemp=min(Minimum_temperature),
+             med_mntemp=median(Minimum_temperature),
+             mean_mntemp=mean(Minimum_temperature),
+             max_mntemp=max(Minimum_temperature),
+             min_mxtemp=min(Maximum_temperature),
+             med_mxtemp=median(Maximum_temperature),
+             mean_mxtemp=mean(Maximum_temperature),
+             max_mxtemp=max(Maximum_temperature),
+             min_9amtemp=min(`9am_Temperature`),
+             med_9amtemp=median(`9am_Temperature`),
+             mean_9amtemp=mean(`9am_Temperature`),
+             max_9amtemp=max(`9am_Temperature`),
+             min_3pmtemp=min(`3pm_Temperature`),
+             med_3pmtemp=median(`3pm_Temperature`),
+             mean_3pmtemp=mean(`3pm_Temperature`),
+             max_3pmtemp=max(`3pm_Temperature`),
+             min_mxgstspd=min(`Speed_of_maximum_wind_gust_(km/h)`),
+             med_mxgstspd=median(`Speed_of_maximum_wind_gust_(km/h)`),
+             mean_mxgstspd=mean(`Speed_of_maximum_wind_gust_(km/h)`),
+             max_spdofmxwndgust=max(`Speed_of_maximum_wind_gust_(km/h)`))
> stats_tblActWhthr
# A tibble: 1 x 20
  min_mntemp med_mntemp mean_mntemp max_mntemp min_mxtemp med_mxtemp mean_mxtemp
       <dbl>      <dbl>       <dbl>      <dbl>      <dbl>      <dbl>       <dbl>
1       -6.4          8        7.84       26.7        8.6       23.0        23.1
# ... with 13 more variables: max_mxtemp <dbl>, min_9amtemp <dbl>,
#   med_9amtemp <dbl>, mean_9amtemp <dbl>, max_9amtemp <dbl>, min_3pmtemp <dbl>,
#   med_3pmtemp <dbl>, mean_3pmtemp <dbl>, max_3pmtemp <dbl>, min_mxgstspd <dbl>,
#   med_mxgstspd <dbl>, mean_mxgstspd <dbl>, max_spdofmxwndgust <dbl>
> |
```

##

2. average min temp per month and year(Minimium_Teperature) and
3. average max temp per month and year(Maximum_Temperature).Calculated together for mintemp and maxtemp variable per month and per year.(Two column done together)

```
> minmax_temp_month <- tblActwthr_tidied %>% group_by(Year, Month) %>%
+   summarise(mean_mintemp_prmn = mean(Minimum_temperature),
+             mean_maxtemp_prmn = mean(Maximum_temperature))
`summarise()` has grouped output by 'Year'. You can override using the `.groups` argum
ent.
> minmax_temp_year <- tblActwthr_tidied %>% group_by(Year) %>%
+   summarise(mean_mintemp_pryr = mean(Minimum_temperature),
+             mean_maxtemp_pryr = mean(Maximum_temperature))
```

```
> minmax_temp_month
# A tibble: 19 x 4
# Groups:   Year [3]
   Year  Month mean_mintemp_prmn mean_maxtemp_prmn
   <fct> <fct>             <dbl>              <dbl>
 1 2018  8                0.765               13.9
 2 2018  9                1.82                18.0
 3 2018  10               7.30                22.8
 4 2018  11              10.5                 24.5
 5 2018  12              13.7                 29.3
 6 2019  1               17.7                 34.5
 7 2019  2               12.9                 29.1
 8 2019  3               12                   26.0
 9 2019  4                7.47                22.6
10 2019  5                3.34                16.6
11 2019  6               -0.0833              14.1
12 2019  7                0.668               13.7
13 2019  8                0.110               14.2
14 2019  9                2.12                18.8
15 2019  10               6.21                23.6
16 2019  11               9.49                26.8
17 2019  12              13.1                 31.7
18 2020  1               15.2                 31.7
19 2020  2               15.1                 27.7
> minmax_temp_year
# A tibble: 3 x 3
  Year  mean_mintemp_pryr mean_maxtemp_pryr
* <fct>             <dbl>              <dbl>
1 2018               6.83               21.7
2 2019               7.06               22.6
3 2020              15.2                29.8
```

##

4. Wind gust speed average grouped by direction of wind gust.

```
> windgst_avg
# A tibble: 17 x 2
   Direction_of_maximum_wind_gust mean_gst
 * <chr>                            <dbl>
 1 E                                 41.9
 2 ENE                               40.1
 3 ESE                               42.5
 4 N                                 40.0
 5 NE                                34
 6 NNE                               32.1
 7 NNW                               44.4
 8 NW                                50.0
 9 S                                 44.7
10 SE                                39.3
11 SSE                               38.3
12 SSW                               40.7
13 SW                                49.5
14 W                                 46.6
15 WNW                               50.2
16 WSW                               41.2
17 NA                                43
> |
```

##

5. Highest Rainfall month and year.

```
> Month <- tbl_maxrain[which.max(tbl_maxrain$rain),"Month"]
> Year <- tbl_maxrain[which.max(tbl_maxrain$rain),"Year"]
> rain <- round(tbl_maxrain[which.max(tbl_maxrain$rain),"rain"],2)
> print(paste0("Highest rainfall happens in month-",Month,
+              "  and in year-",Year[[1]],
+              " and is=",rain,"mm"))
[1] "Highest rainfall happens in month-2  and in year-2020 and is=60.4mm"
> |
```

##

6. No rainfall month and year.

```
> #6. no rainfall month and year.##################################################
> #grouping by month and year summing up all rainfall in mm
> #checking which month has 0mm out of all.
> tbl_norain <- tblActwthr_tidied %>%
+       group_by(Year,Month) %>%
+       summarise(rain = sum(`Rainfall_(mm)`))
`summarise()` has grouped output by 'Year'. You can override using the `.groups` argument.
> yrnorain <- tbl_norain[which(tbl_norain$rain == 0),"Year"]
> mntnorain <- tbl_norain[which(tbl_norain$rain == 0),"Month"]
> if(any(tbl_norain$rain==0)){
+   print(paste("There is no rainfall in the month",mntnorain,"and the year-",
+               yrnorain,))
+ }else{print("There is rainfall in one or the other months in every year,so there was no dry month")}
[1] "There is rainfall in one or the other months in every year,so there was no dry month"
> |
```

##

7. Highest humid level month in 2019.

```
> tbl_meanhumid <- tblActwthr_tidied %>%
+   filter(Year==2019) %>%
+   gather(key= "humid_key",
+          value= "humid_val",
+          c(`9am_relative_humidity_(%)`,`3pm_relative_humidity_(%)`)) %>%
+   group_by(Month) %>%
+   summarise(humid = mean(humid_val))
> Month <- tbl_meanhumid[which.max(tbl_meanhumid$humid),"Month"]
> humid <- round(tbl_meanhumid[which.max(tbl_meanhumid$humid),"humid"],2)
> print(paste0("Highest humidity is in month-",Month,
+              " and is=",humid,"%"," in the year-2019"))
[1] "Highest humidity is in month-6 and is=72.47% in the year-2019"
>
```

##

8. min max and avg temp , wind_speed and humidity per month and per quarter 2019.

```
> Mnthlydata_2019 <- tblActwthr_tidied %>%
+   filter(Year==2019) %>%
+   gather(key= "temp_key",
+          value= "temp_val",
+          c(`9am_Temperature`, `3pm_Temperature`)) %>%
+   gather(key= "wndspd_key",
+          value= "wndspd_val",
+          c(`9am_wind_speed_(km/h)`, `3pm_wind_speed_(km/h)`)) %>%
+   gather(key= "humid_key",
+          value= "humid_val",
+          c(`9am_relative_humidity_(%)`, `3pm_relative_humidity_(%)`)) %>%
+   group_by(Month) %>%
+   summarize(min_temp=min(temp_val),
+             max_temp=max(temp_val),
+             avg_temp=mean(temp_val),
+             wndspd_min=min(wndspd_val),
+             wndspd_max=max(wndspd_val),
+             wndspd_avg=mean(wndspd_val),
+             humid_min=min(humid_val),
+             humid_max=max(humid_val),
+             humid_avg=mean(humid_val))
> Mnthlydata_2019
# A tibble: 12 x 10
   Month min_temp max_temp avg_temp wndspd_min wndspd_max wndspd_avg humid_min humid_max humid_avg
 * <fct>     <dbl>    <dbl>    <dbl>      <dbl>      <dbl>      <dbl>     <dbl>     <dbl>     <dbl>
 1 1          17       40.7     27.9          2         35       14.8        10        86      47.9
 2 2          13.4     35.2     23.0          0         44       15.6         8        79      47.5
 3 3           9.8     33.3     20.6          0         39       15.4        19        98      56.5
 4 4           8.4     25.7     17.5          0         33       11.5        26        99      58.5
 5 5           3.5     22.4     12.1          0         46       15.1        28        99      68.6
 6 6          -0.5     16.5      9.05         0         39       11.9        40       100      72.5
 7 7           0.6     16.6      9.19         0         48       17.5        33       100      67.7
 8 8           0.9     18.4      9.52         0         37       17.0        27        99      61.1
 9 9           5.6     22.9     14.0          0         46       18.0        17        87      48.2
10 10          7.6     30.3     18.1          0         61       18.4        11        87      42.2
11 11          9       38.2     21.1          0         46       21.6         7        85      35.2
12 12         10.4     40.1     25.1          0         41       18.0         6        69      33.3
>
```

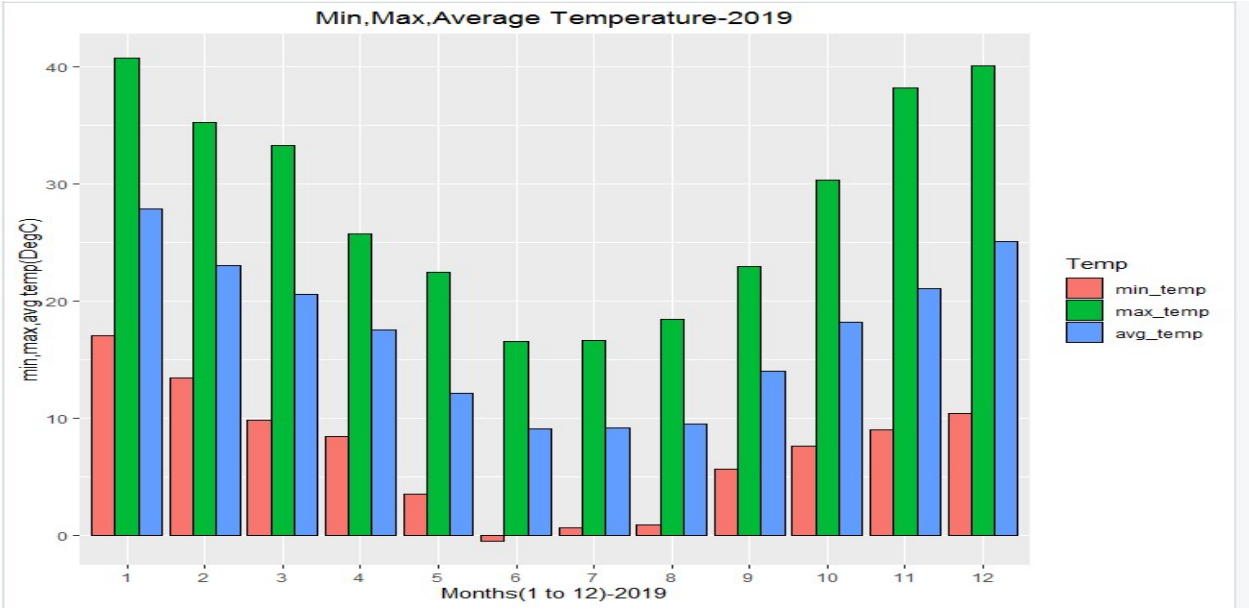
```
> #min max and avg temp , wind_speed and humidity per quarter####################
> if(!("lubridate" %in% rownames(installed.packages()))){
+   install.packages("lubridate")
+ }
> library(lubridate)
> qrterlydata_2019 <- tblActwthr_tidied %>%
+   filter(Year==2019) %>%
+   gather(key= "temp_key",
+          value= "temp_val",
+          c(`9am_Temperature`, `3pm_Temperature`)) %>%
+   gather(key= "wndspd_key",
+          value= "wndspd_val",
+          c(`9am_wind_speed_(km/h)`, `3pm_wind_speed_(km/h)`)) %>%
+   gather(key= "humid_key",
+          value= "humid_val",
+          c(`9am_relative_humidity_(%)`, `3pm_relative_humidity_(%)`)) %>%
+   group_by(Date=quarter(Date, with_year = T)) %>%
+   summarize(min_temp=min(temp_val),
+             max_temp=max(temp_val),
+             avg_temp=mean(temp_val),
+             wndspd_min=min(wndspd_val),
+             wndspd_max=max(wndspd_val),
+             wndspd_avg=mean(wndspd_val),
+             humid_min=min(humid_val),
+             humid_max=max(humid_val),
+             humid_avg=mean(humid_val))
> qrterlydata_2019
# A tibble: 4 x 10
    Date min_temp max_temp avg_temp wndspd_min wndspd_max wndspd_avg humid_min humid_max humid_avg
* <dbl>    <dbl>    <dbl>    <dbl>      <dbl>      <dbl>      <dbl>     <dbl>     <dbl>     <dbl>
1 2019.      9.8     40.7     23.8          0         44       15.3         8        98      50.8
2 2019.     -0.5     25.7     12.9          0         46       12.9        26       100      66.5
3 2019.      0.6     22.9     10.9          0         48       17.5        17       100      59.1
4 2019.      7.6     40.1     21.4          0         61       19.3         6        87      36.9
>
```

##

9. Barplot for specified variables monthly and quarterly-2019.
   Monthy graphs for min,max,average temp, wind speed and humidity.

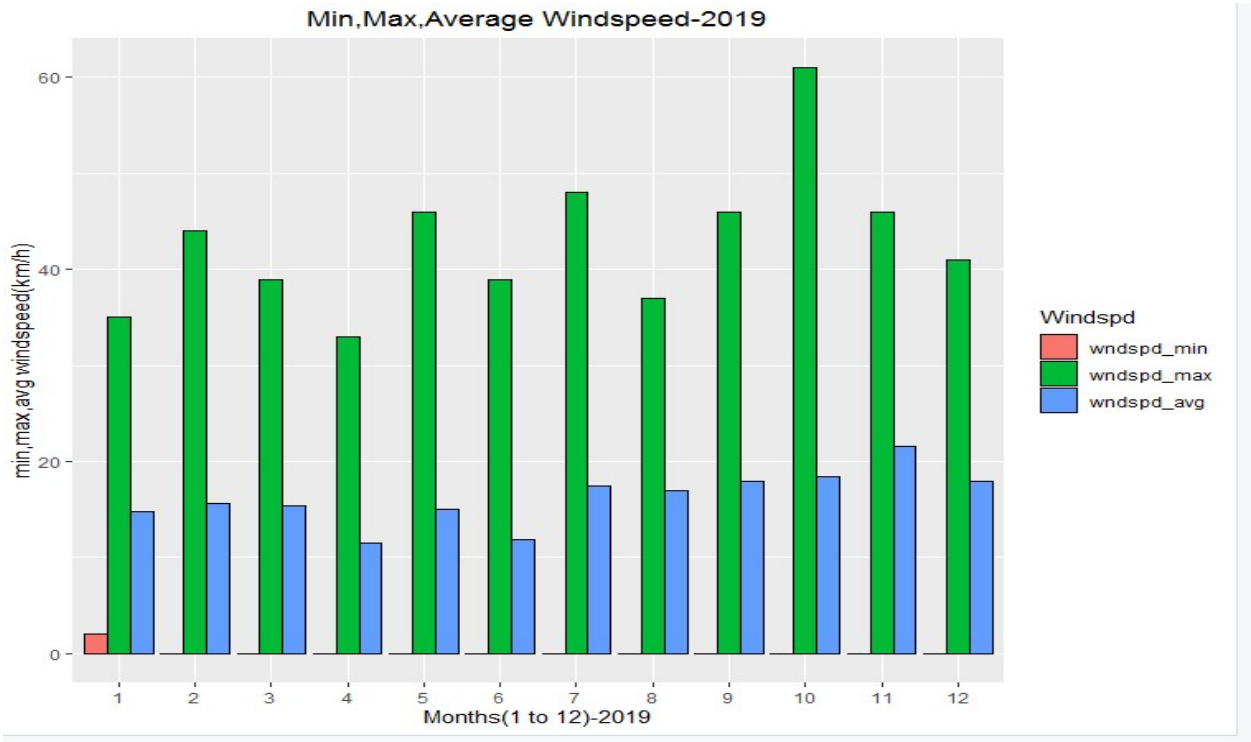
```
> #9. barplots for all the specified variables.
> if(!("ggplot2" %in% rownames(installed.packages()))){
+   install.packages("ggplot2")
+ }
> library(ggplot2)
> if(!("reshape2" %in% rownames(installed.packages()))){
+   install.packages("reshape2")
+ }
> library(reshape2)
> #Selting data to gather all the values to single column across(min,max and avg)
> #renaming value column to "Temp", "WindSpd" and "Humidity" accordingly
> Mnthlydata_temp <- melt(select(Mnthlydata_2019,c(1:4)))
Using Month as id variables
> colnames(Mnthlydata_temp)[2] <- "Temp"
```

```
> Mnthlydata_temp
   Month      Temp      value
1      1  min_temp 17.000000
2      2  min_temp 13.400000
3      3  min_temp  9.800000
4      4  min_temp  8.400000
5      5  min_temp  3.500000
6      6  min_temp -0.500000
7      7  min_temp  0.600000
8      8  min_temp  0.900000
9      9  min_temp  5.600000
10    10  min_temp  7.600000
11    11  min_temp  9.000000
12    12  min_temp 10.400000
13     1  max_temp 40.700000
14     2  max_temp 35.200000
15     3  max_temp 33.300000
16     4  max_temp 25.700000
17     5  max_temp 22.400000
18     6  max_temp 16.500000
19     7  max_temp 16.600000
20     8  max_temp 18.400000
21     9  max_temp 22.900000
22    10  max_temp 30.300000
23    11  max_temp 38.200000
24    12  max_temp 40.100000
25     1  avg_temp 27.882258
26     2  avg_temp 22.991071
27     3  avg_temp 20.566129
28     4  avg_temp 17.511667
29     5  avg_temp 12.137097
30     6  avg_temp  9.050000
31     7  avg_temp  9.190323
32     8  avg_temp  9.519355
33     9  avg_temp 13.986667
34    10  avg_temp 18.141935
35    11  avg_temp 21.083333
36    12  avg_temp 25.082258
```



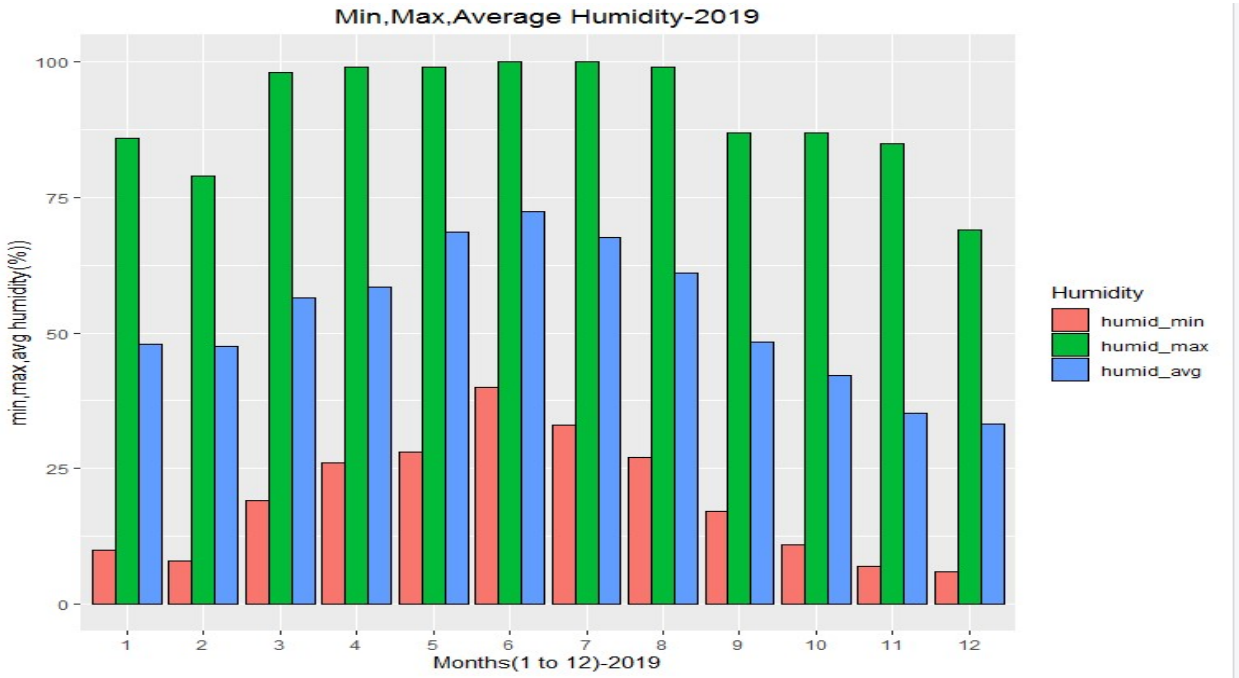
##

```
> Mnthlydata_wndspd
   Month     windspd     value
1      1 wndspd_min   2.00000
2      2 wndspd_min   0.00000
3      3 wndspd_min   0.00000
4      4 wndspd_min   0.00000
5      5 wndspd_min   0.00000
6      6 wndspd_min   0.00000
7      7 wndspd_min   0.00000
8      8 wndspd_min   0.00000
9      9 wndspd_min   0.00000
10    10 wndspd_min   0.00000
11    11 wndspd_min   0.00000
12    12 wndspd_min   0.00000
13     1 wndspd_max  35.00000
14     2 wndspd_max  44.00000
15     3 wndspd_max  39.00000
16     4 wndspd_max  33.00000
17     5 wndspd_max  46.00000
18     6 wndspd_max  39.00000
19     7 wndspd_max  48.00000
20     8 wndspd_max  37.00000
21     9 wndspd_max  46.00000
22    10 wndspd_max  61.00000
23    11 wndspd_max  46.00000
24    12 wndspd_max  41.00000
25     1 wndspd_avg  14.80645
26     2 wndspd_avg  15.64286
27     3 wndspd_avg  15.41935
28     4 wndspd_avg  11.53333
29     5 wndspd_avg  15.08065
30     6 wndspd_avg  11.86667
31     7 wndspd_avg  17.50000
32     8 wndspd_avg  16.96774
33     9 wndspd_avg  17.98333
34    10 wndspd_avg  18.40323
35    11 wndspd_avg  21.56667
36    12 wndspd_avg  17.98387
```



##

```
> colnames(Mnthlydata_humid)[2] <-
> Mnthlydata_humid
   Month  Humidity      value
1      1 humid_min  10.00000
2      2 humid_min   8.00000
3      3 humid_min  19.00000
4      4 humid_min  26.00000
5      5 humid_min  28.00000
6      6 humid_min  40.00000
7      7 humid_min  33.00000
8      8 humid_min  27.00000
9      9 humid_min  17.00000
10    10 humid_min  11.00000
11    11 humid_min   7.00000
12    12 humid_min   6.00000
13     1 humid_max  86.00000
14     2 humid_max  79.00000
15     3 humid_max  98.00000
16     4 humid_max  99.00000
17     5 humid_max  99.00000
18     6 humid_max 100.00000
19     7 humid_max 100.00000
20     8 humid_max  99.00000
21     9 humid_max  87.00000
22    10 humid_max  87.00000
23    11 humid_max  85.00000
24    12 humid_max  69.00000
25     1 humid_avg  47.91935
26     2 humid_avg  47.51786
27     3 humid_avg  56.50000
28     4 humid_avg  58.46667
29     5 humid_avg  68.59677
30     6 humid_avg  72.46667
31     7 humid_avg  67.66129
32     8 humid_avg  61.06452
33     9 humid_avg  48.23333
34    10 humid_avg  42.22581
35    11 humid_avg  35.21667
36    12 humid_avg  33.27419
> |
```

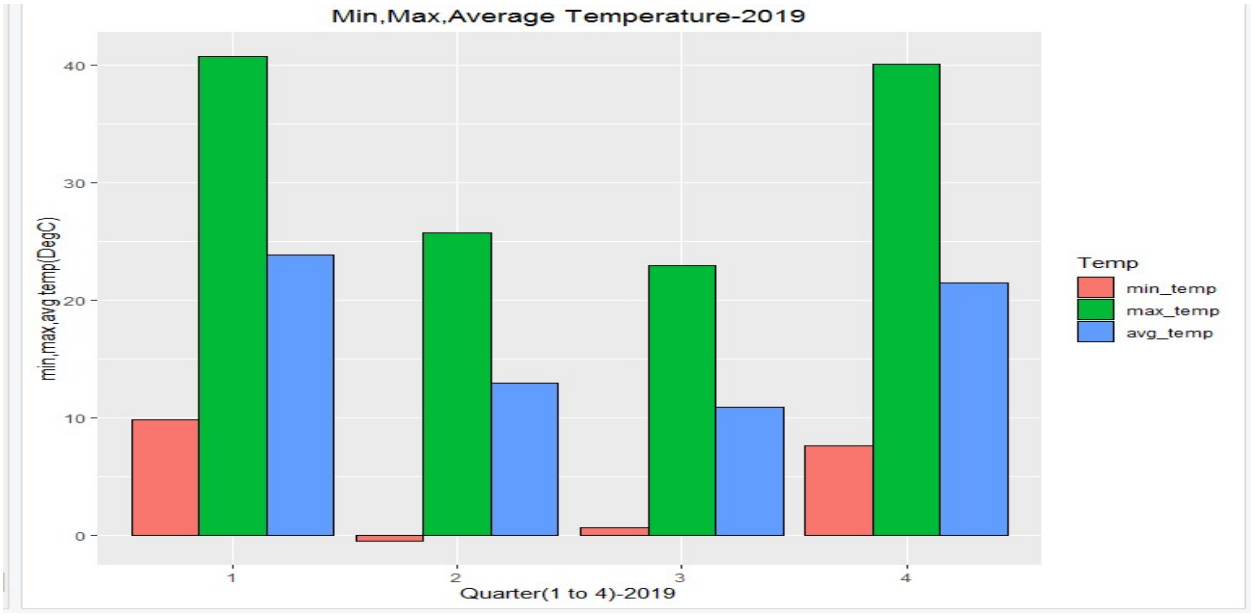


##

Quarterly graphs for min,max,average temp, wind speed and humidity.

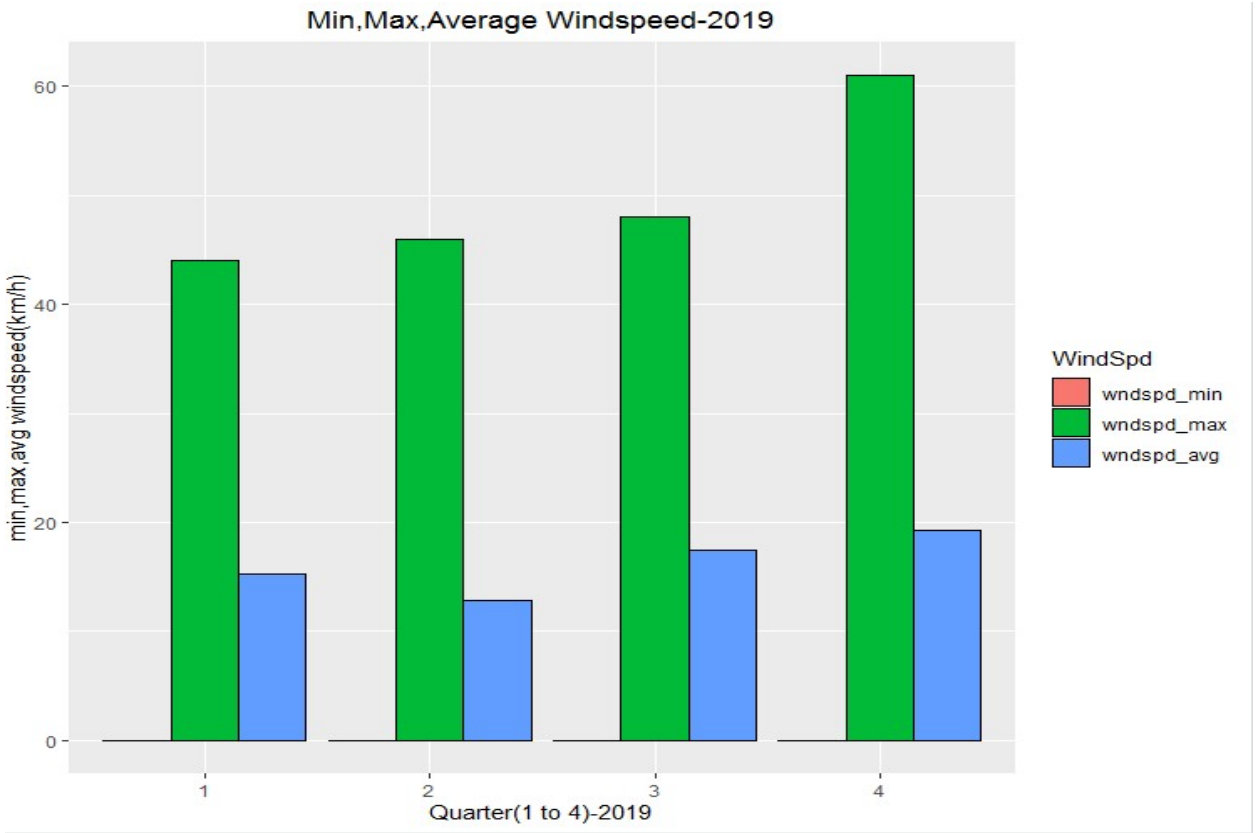
```
> #creating column quatrter to represent each quarter data
> qrterlydata_2019$quarter <- factor(c(1,2,3,4))
> #relocating column quarter as first column
> qrterlydata_2019 <- qrterlydata_2019 %>% relocate(quarter, .before = min_temp)
> #Removing Date column as it is unnecessary.
> qrterlydata_2019 <- select(qrterlydata_2019, -c(Date))
> #Selting data to gather all the values to single column across(min,max and avg)
> #renaming value column to "Temp", "WindSpd" and "Humidity" accordingly
> qrterly_temp <- melt(select(qrterlydata_2019,c(1:4)))
Using quarter as id variables
> colnames(qrterly_temp)[2] <- "Temp"
> qrterly_wndspd <- melt(select(qrterlydata_2019,c(1),c(5:7)))
Using quarter as id variables
> colnames(qrterly_wndspd)[2] <- "WindSpd"
> qrterly_humid <- melt(select(qrterlydata_2019,c(1),c(8:10)))
Using quarter as id variables
> colnames(qrterly_humid)[2] <- "Humidity"
>
```

```
> #bar plot for min max and avg of temperature-2019 each quarter.
> qrterly_temp
   quarter      Temp     value
1        1  min_temp   9.80000
2        2  min_temp  -0.50000
3        3  min_temp   0.60000
4        4  min_temp   7.60000
5        1  max_temp  40.70000
6        2  max_temp  25.70000
7        3  max_temp  22.90000
8        4  max_temp  40.10000
9        1  avg_temp  23.84056
10       2  avg_temp  12.89121
11       3  avg_temp  10.86522
12       4  avg_temp  21.43967
> qrterly_temp %>%
+   ggplot(aes(quarter, value, fill = Temp)) +
+   geom_bar(position="dodge", stat = "identity",color="black") +
+   ggtitle("Min,Max,Average Temperature-2019") +
+   theme(plot.title = element_text(hjust = 0.5)) +
+   xlab("Quarter(1 to 4)-2019")+
+   ylab("min,max,avg temp(DegC)")
>
```



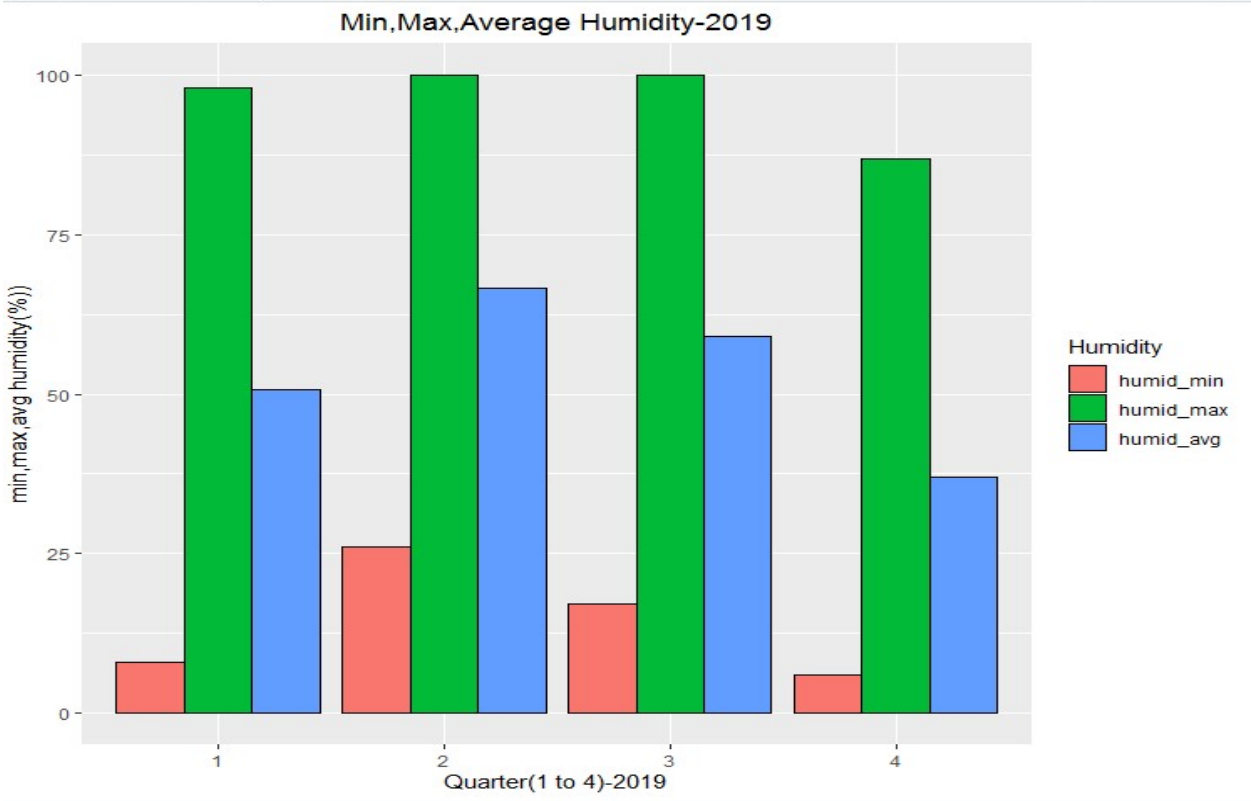
##

```
> qrterly_wndspd
   quarter     WindSpd     value
1        1 wndspd_min   0.00000
2        2 wndspd_min   0.00000
3        3 wndspd_min   0.00000
4        4 wndspd_min   0.00000
5        1 wndspd_max  44.00000
6        2 wndspd_max  46.00000
7        3 wndspd_max  48.00000
8        4 wndspd_max  61.00000
9        1 wndspd_avg  15.27778
10       2 wndspd_avg  12.85165
11       3 wndspd_avg  17.47826
12       4 wndspd_avg  19.29348
> qrterly_wndspd %>%
+   ggplot(aes(quarter, value, fill = WindSpd)) +
+   geom_bar(position="dodge", stat = "identity",color="black") +
+   ggtitle("Min,Max,Average Windspeed-2019") +
+   theme(plot.title = element_text(hjust = 0.5)) +
+   xlab("Quarter(1 to 4)-2019")+
+   ylab("min,max,avg windspeed(km/h)")
>
```



##

```
> qrterly_humid
   quarter  Humidity     value
1        1 humid_min   8.00000
2        2 humid_min  26.00000
3        3 humid_min  17.00000
4        4 humid_min   6.00000
5        1 humid_max  98.00000
6        2 humid_max 100.00000
7        3 humid_max 100.00000
8        4 humid_max  87.00000
9        1 humid_avg  50.75000
10       2 humid_avg  66.53297
11       3 humid_avg  59.10326
12       4 humid_avg  36.92391
> qrterly_humid %>%
+   ggplot(aes(quarter, value, fill = Humidity)) +
+   geom_bar(position="dodge", stat = "identity",color="black") +
+   ggtitle("Min,Max,Average Humidity-2019") +
+   theme(plot.title = element_text(hjust = 0.5)) +
+   xlab("Quarter(1 to 4)-2019")+
+   ylab("min,max,avg humidity(%))")
>
```



##