Assignment 4

aditya venugopalan a1899824

2023-10-26

Excecutive summary

#1.1 Project Outline #In this universe one of the basic things that is required for living is water. MWC is given the responsibilty make sure water is availbe in the city of Melbourne, Australia. The main motivation behind this project is to understand the relationship between evaporation rate in a day , amount of sunlight hours and the maximum speed of wind(gust of wind) in kmph, the minimum temperature as well as the maximum temperature and the relative humidity at 9 am. This report analysis will help the Melbourne Water Corporation(MWC) to have some insights on estimating the rate of evaporation at their reservoirs and help them understand is there any factors affecting the evaporation rate at any given day.

1.2 Approach

We have done the analysis on the obtained dataset , which conisisted some data on random samples of Melbourne's weather observations and evaporation for the financial year of 2018-2019. This report also includes the the relation which will answer the queries asked by the MWC. Multiple statistical modles were developed in order to understand whether the rate of evaporation is affected by some factors or not.

1.3 Findings

Month: After evaluation we were able to find that the aamount of evaporation which we have termed as evaporation rate is comparetively higher in the months of January, March, April , November and December . And because of this we were able to conclude that there is a significant relationship observed between the evaporation rate and the months.

Day Of Week: We were able to observe the fact that the day of the week didn't have much influence on the rate of evaporation significantly and hence no relationship can be seen .

Number of hours of bright sunlight:

#Even in the case of number of hours of sunlight we are unable any effects caused by the number of hours of bright sunlight as they form no relationship.

Speed of the maximum wind gust in kilometer per hour. We can see the speed of gust of wind is a significant factor while predicting the rate of evaporation. We can see a moderate positive linear relationship. For every 1km/h increase in speed of maximum wind gust will increase in 0.522 mm increase in rate of evaporation.

Minimum temperature in degree Celsius : It was observed that minimum temperature is significantly important for the purpose of prediction of amount of evaporation. The coreresponding figure also displays postive linear relationship.For every 1 C increase in the minimum temperature will cause a 0.329 mm increase in the evaporation amount.

Relative humidity (measured at 9am): This is a unique case as we can see a significant negative relationship is developed between humidity and rate of evaporation . for every increase in humidity will cause decrease in 0.084 mm in amount of evaporation

Recommendations

Based on my analyis , we can say with 95% confidence that on december 26th 2024 and January 11 2024 , Cardinia Reservoir will have more than 9mm of evaporation , which should motivate the MWC to make temproray measures to ensure continuous supply of water which also means they should consider transferring the water from the Silvan Reservoir upstream.

METHODS

2.1 Software used

In this section, I'll be analyzing a random sampled dataset of Melbourne's weather observations devleoped by the Melbourne Water Corporation("MWC") in order to understand the evaporation rate from the reservoir called Cardina Reservoir . The analysis was successfull with the help of Rstudio.

2.2 Dataset used

In this particular section we will be able to observe and understand how varaibles such as (a) Month, (b) Day of the week, (c)Number of hours of bright sunlight, (d) Speed of the maximum wind gust

#2 .3 Bivarite Summary

Month : In the figure 1: which tell us relation of rate of evaporation against month , we can see spread of few months March , April , September and November to be wider and the location higher than 9. In the case of January,February,March April , and November and December we can see few outliers, which tells us there is a significant relationship

Day of the week: In the figure 2 the boxplot tells us the spread with is quite same with all the days of the week and location not showing varaince among the days and there is no significant relationship observed.

Sunlight hours: Figure 3 is a scatterplot which shows us no significant relationship due to random scattering of points

Maximum speed of wind gust(kmph): Even though the name of this varaible was slighlty changed for better presentation , we can still observe that there is a moderate linear positive relationship betwwen rate or amount of evaporation against the maximum speed of wind.

Minimum Temperature (in c): The figure 5 which is a scatterplot which tells us a postive relationship betwwen the rate of evaporation agaoinst minimum temperature.

Relative humidity: The figure number 6 displays a sytong positive relationship between evaporation and humidity when measuread at 9 am

Model selection

We will begin with a model which has all the possible predictor for the response variable . We also used the Anova() function from the "car" package and calculated the p-value for predictor and found out that the varaible sunlight hours didnt had enough significance was removed.

After exculding sunlight hours , second model was created and we realise that the Sunlight hours didnt have enough significance and hence it was removed

In our third model after excluding sunlight hours temperature we

```r
#Loading the required libraries
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(tidyr)
library(dplyr)
library(ggplot2)
library(inspectdf)
library(stringr)
library(readr)
library(forcats)
library(ggthemes)
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(lubridate)
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```r
#importing the dataset on which we will perform analysis
melbourne <- read.csv("melbourne.csv")
head(melbourne)
```

```
##        Date Minimum.temperature..Deg.C. Maximum.Temperature..Deg.C.
## 1 2019-01-1                       15.5                       26.2
## 2 2019-01-2                       18.4                       22.2
## 3 2019-01-3                       15.9                       29.5
## 4 2019-01-4                       18.0                       42.6
## 5 2019-01-5                       17.4                       21.2
## 6 2019-01-6                       14.6                       22.1
##   Rainfall..mm. Evaporation..mm. Sunshine..hours.
## 1           0.0              7.0             11.0
## 2           0.0              7.0              7.5
## 3           0.0              6.6              9.3
## 4           0.0              7.8             12.2
## 5           0.4             15.4              5.8
## 6           1.4              6.4             13.3
##   Direction.of.maximum.wind.gust Speed.of.maximum.wind.gust..km.h.
## 1                              S                                35
## 2                            SSW                                39
## 3                            SSW                                26
## 4                             NW                                54
## 5                            SSW                                39
## 6                            SSW                                33
##   Time.of.maximum.wind.gust X9am.Temperature..Deg.C. X9am.relative.humidity....
## 1                  17:44:00                      19.8                         74
## 2                  15:23:00                      19.5                         64
## 3                  14:53:00                      18.1                         75
## 4                  12:03:00                      29.5                         31
## 5                  08:24:00                      18.0                         63
## 6                  11:12:00                      17.7                         55
##   X9am.cloud.amount..oktas. X9am.wind.direction X9am.wind.speed..km.h.
## 1                         7                   S                      6
## 2                         8                 SSE                      7
## 3                         8                   S                      2
## 4                         0                 NNE                      9
## 5                         7                   S                     13
## 6                         1                  SW                      9
##   X9am.MSL.pressure..hPa. X3pm.Temperature..Deg.C. X3pm.relative.humidity....
## 1                  1013.0                      24.4                        45
## 2                  1013.9                      21.4                        62
## 3                  1012.6                      24.6                        60
## 4                  1005.5                      42.0                        16
## 5                  1013.5                      19.1                        58
## 6                  1020.4                      20.6                        48
##   X3pm.cloud.amount..oktas. X3pm.wind.direction X3pm.wind.speed..km.h.
## 1                         1                 SSW                     11
## 2                         1                 SSW                     19
## 3                         0                 SSW                     13
## 4                         1                  NW                     15
## 5                         7                   S                     11
## 6                         1                 SSW                     13
##   X3pm.MSL.pressure..hPa.
## 1                  1011.5
## 2                  1012.9
## 3                  1009.9
## 4                  1001.0
```

```
## 5                  1013.4
## 6                  1019.5
```

```
# Here i will try to eliminate or remove some data which are irrelevant and not required for anlayis
melbourne <- melbourne %>%
  mutate(Date = ymd(Date)) %>%
  mutate(month = as_factor(month(Date)),
         `X9am.wind.speed..km.h.` = as.integer(str_replace(`X9am.wind.speed..km.h.`, "Calm", "0")),
         weekday = as_factor(wday(Date)))
melbourne <- melbourne %>%
rename(minimum_temperature = 'Minimum.temperature..Deg.C.',
maximum_temperature = 'Maximum.Temperature..Deg.C.',
humidity_9am = 'X9am.relative.humidity....',
evaporation_rate = 'Evaporation..mm.',
Sunlight_hours = "Sunshine..hours.",
Maximum_speed_gust_of_wind= "Speed.of.maximum.wind.gust..km.h.")

head(melbourne)
```

```
##         Date minimum_temperature maximum_temperature Rainfall..mm.
## 1 2019-01-01                15.5                26.2           0.0
## 2 2019-01-02                18.4                22.2           0.0
## 3 2019-01-03                15.9                29.5           0.0
## 4 2019-01-04                18.0                42.6           0.0
## 5 2019-01-05                17.4                21.2           0.4
## 6 2019-01-06                14.6                22.1           1.4
##   evaporation_rate Sunlight_hours Direction.of.maximum.wind.gust
## 1              7.0           11.0                              S
## 2              7.0            7.5                            SSW
## 3              6.6            9.3                            SSW
## 4              7.8           12.2                             NW
## 5             15.4            5.8                            SSW
## 6              6.4           13.3                            SSW
##   Maximum_speed_gust_of_wind Time.of.maximum.wind.gust X9am.Temperature..Deg.C.
## 1                         35                 17:44:00                      19.8
## 2                         39                 15:23:00                      19.5
## 3                         26                 14:53:00                      18.1
## 4                         54                 12:03:00                      29.5
## 5                         39                 08:24:00                      18.0
## 6                         33                 11:12:00                      17.7
##   humidity_9am X9am.cloud.amount..oktas. X9am.wind.direction
## 1           74                         7                  S
## 2           64                         8                SSE
## 3           75                         8                  S
## 4           31                         0                NNE
## 5           63                         7                  S
## 6           55                         1                 SW
##   X9am.wind.speed..km.h. X9am.MSL.pressure..hPa. X3pm.Temperature..Deg.C.
## 1                      6                 1013.0                      24.4
## 2                      7                 1013.9                      21.4
## 3                      2                 1012.6                      24.6
## 4                      9                 1005.5                      42.0
## 5                     13                 1013.5                      19.1
## 6                      9                 1020.4                      20.6
```

```
##   X3pm.relative.humidity.... X3pm.cloud.amount..oktas. X3pm.wind.direction
## 1                         45                         1                 SSW
## 2                         62                         1                 SSW
## 3                         60                         0                 SSW
## 4                         16                         1                  NW
## 5                         58                         7                   S
## 6                         48                         1                 SSW
##   X3pm.wind.speed..km.h. X3pm.MSL.pressure..hPa. month weekday
## 1                     11                 1011.5     1       3
## 2                     19                 1012.9     1       4
## 3                     13                 1009.9     1       5
## 4                     15                 1001.0     1       6
## 5                     11                 1013.4     1       7
## 6                     13                 1019.5     1       1
```

```
#continuing the cleaning of data. Now i'll remove the missing data but first i would love to show how m
inspect_na(melbourne)
```

```
## # A tibble: 23 x 3
##    col_name                        cnt  pcnt
##    <chr>                         <int> <dbl>
##  1 X9am.wind.direction              17 4.66
##  2 evaporation_rate                  8 2.19
##  3 Rainfall..mm.                     2 0.548
##  4 Date                              0 0
##  5 minimum_temperature               0 0
##  6 maximum_temperature               0 0
##  7 Sunlight_hours                    0 0
##  8 Direction.of.maximum.wind.gust    0 0
##  9 Maximum_speed_gust_of_wind        0 0
## 10 Time.of.maximum.wind.gust         0 0
## # i 13 more rows
```

```
#As mentioned earlier i will not hesitate to remove the values or data which will might cause some trou
melbourne <- na.omit(melbourne)
head(melbourne)
```

```
##         Date minimum_temperature maximum_temperature Rainfall..mm.
## 1 2019-01-01                15.5                26.2           0.0
## 2 2019-01-02                18.4                22.2           0.0
## 3 2019-01-03                15.9                29.5           0.0
## 4 2019-01-04                18.0                42.6           0.0
## 5 2019-01-05                17.4                21.2           0.4
## 6 2019-01-06                14.6                22.1           1.4
##   evaporation_rate Sunlight_hours Direction.of.maximum.wind.gust
## 1              7.0           11.0                              S
## 2              7.0            7.5                            SSW
## 3              6.6            9.3                            SSW
## 4              7.8           12.2                             NW
## 5             15.4            5.8                            SSW
## 6              6.4           13.3                            SSW
##   Maximum_speed_gust_of_wind Time.of.maximum.wind.gust X9am.Temperature..Deg.C.
## 1                         35                 17:44:00                     19.8
```

```
## 2                        39                 15:23:00                     19.5
## 3                        26                 14:53:00                     18.1
## 4                        54                 12:03:00                     29.5
## 5                        39                 08:24:00                     18.0
## 6                        33                 11:12:00                     17.7
##   humidity_9am X9am.cloud.amount..oktas. X9am.wind.direction
## 1           74                        7                   S
## 2           64                        8                 SSE
## 3           75                        8                   S
## 4           31                        0                 NNE
## 5           63                        7                   S
## 6           55                        1                  SW
##   X9am.wind.speed..km.h. X9am.MSL.pressure..hPa. X3pm.Temperature..Deg.C.
## 1                      6                 1013.0                     24.4
## 2                      7                 1013.9                     21.4
## 3                      2                 1012.6                     24.6
## 4                      9                 1005.5                     42.0
## 5                     13                 1013.5                     19.1
## 6                      9                 1020.4                     20.6
##   X3pm.relative.humidity.... X3pm.cloud.amount..oktas. X3pm.wind.direction
## 1                         45                         1                 SSW
## 2                         62                         1                 SSW
## 3                         60                         0                 SSW
## 4                         16                         1                  NW
## 5                         58                         7                   S
## 6                         48                         1                 SSW
##   X3pm.wind.speed..km.h. X3pm.MSL.pressure..hPa. month weekday
## 1                     11                 1011.5     1       3
## 2                     19                 1012.9     1       4
## 3                     13                 1009.9     1       5
## 4                     15                 1001.0     1       6
## 5                     11                 1013.4     1       7
## 6                     13                 1019.5     1       1
```

```
#Since we have extracted variables with their proper names . Its time to start our first analysis , tha
#FIGURE 1
ggplot(melbourne, aes(x= month, y = evaporation_rate)) +
  geom_boxplot(aes(fill = month))+
  theme_bw() +
  labs(caption = "Figure 1: Boxplot showing evaporation rate in relation against month")+
  theme(plot.caption = element_text(hjust = 0.5))
```

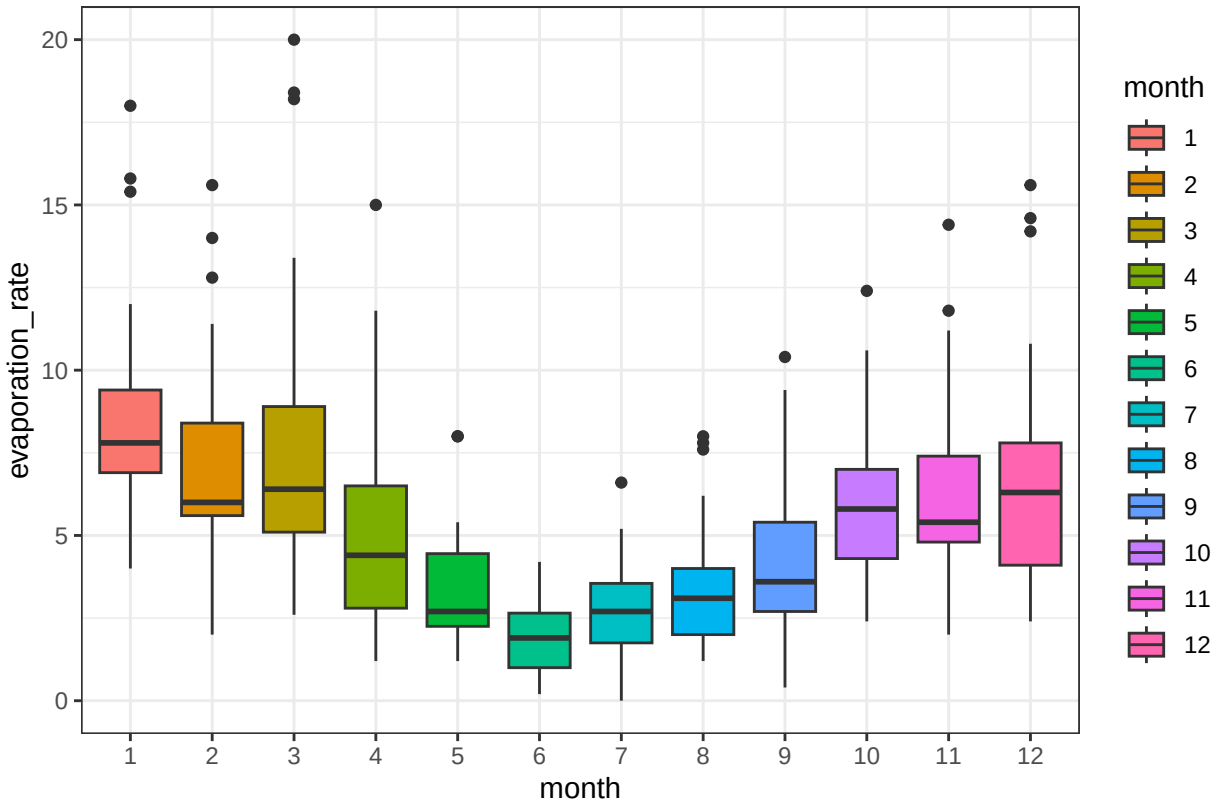


Figure 1: Boxplot showing evaporation rate in relation against month

```
# Figure 2.Now we will have a look how evaporation looks against weekday.
ggplot(melbourne, aes(x= weekday, y = evaporation_rate)) +
  geom_boxplot(aes(fill = weekday))+
  theme_bw() +
  labs(caption = " Figure 2: Boxplot showing evaporation rate in relation against weekday  ")
```

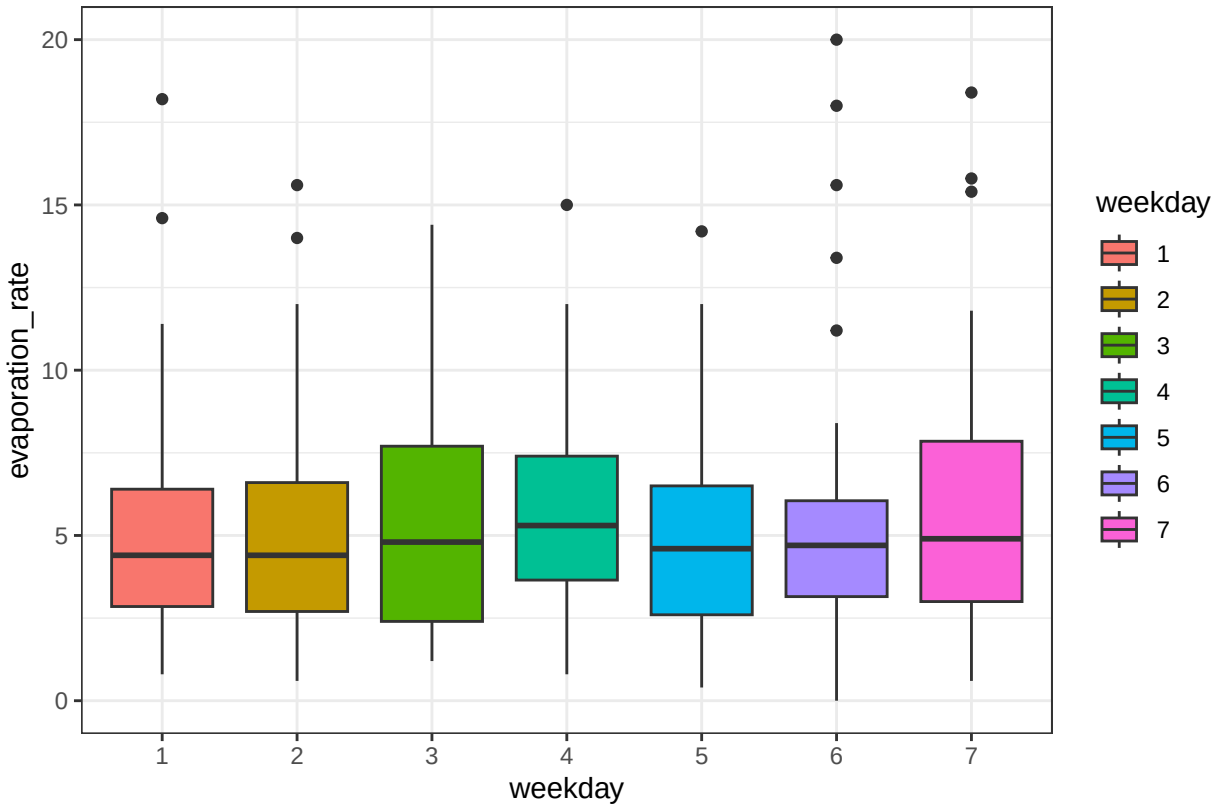


Figure 2: Boxplot showing evaporation rate in relation against weekday

```
#Figure 3. A boxplot which will help us to understand the relation of evaporation rate against Number o
ggplot(melbourne, aes(x= Sunlight_hours, y= evaporation_rate)) +
  geom_point() +
  theme_bw() +
  labs(caption = " Figure 3: A scatterplot showing relation of evaporation rate against Number of hours
  theme(plot.caption = element_text(hjust = 0.5))
```

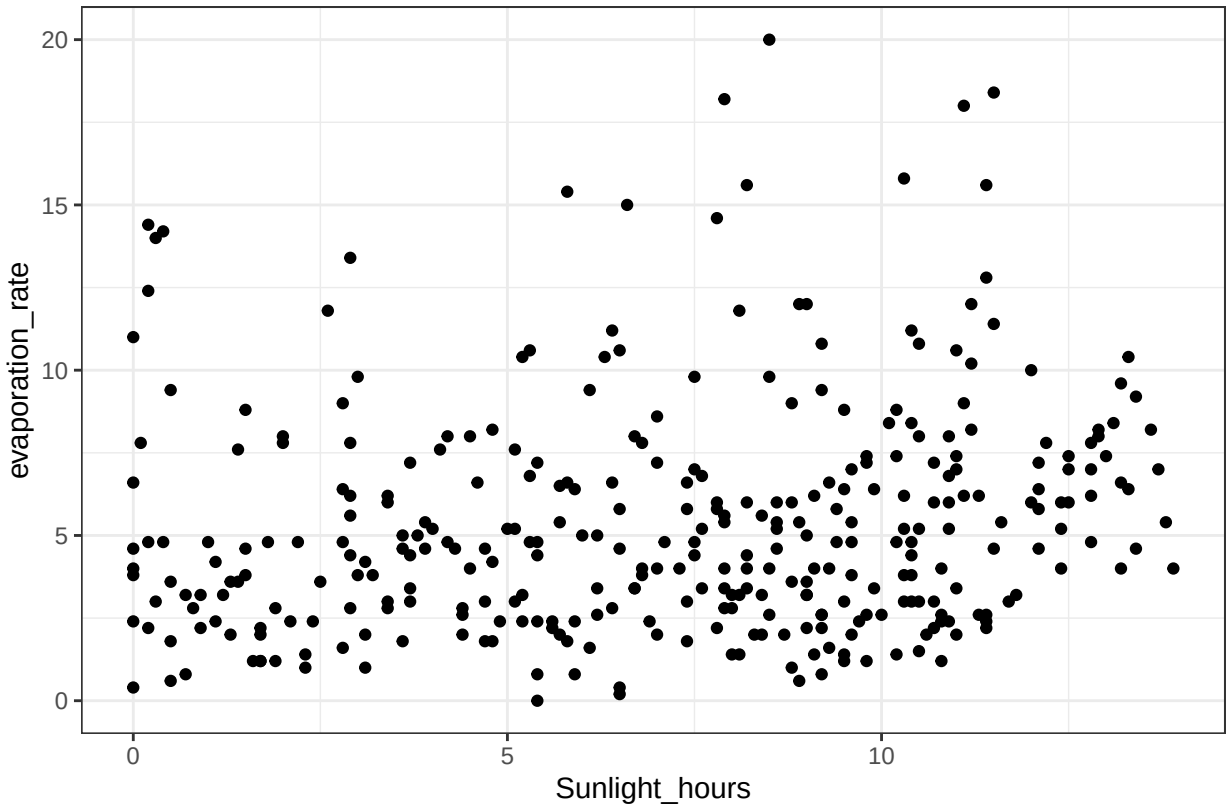


Figure 3: A scatterplot showing relation of evaporation rate against Number of hours of bright sunlight

```
# Figure 4
ggplot(melbourne, aes(x = Maximum_speed_gust_of_wind, y = evaporation_rate)) +
  geom_point() +
  theme_bw() +
  labs(caption = " Figure 4: Scatterplot showing evaporation rate in relation against Speed of the maxi
  theme(plot.caption = element_text(hjust = 0.5))
```

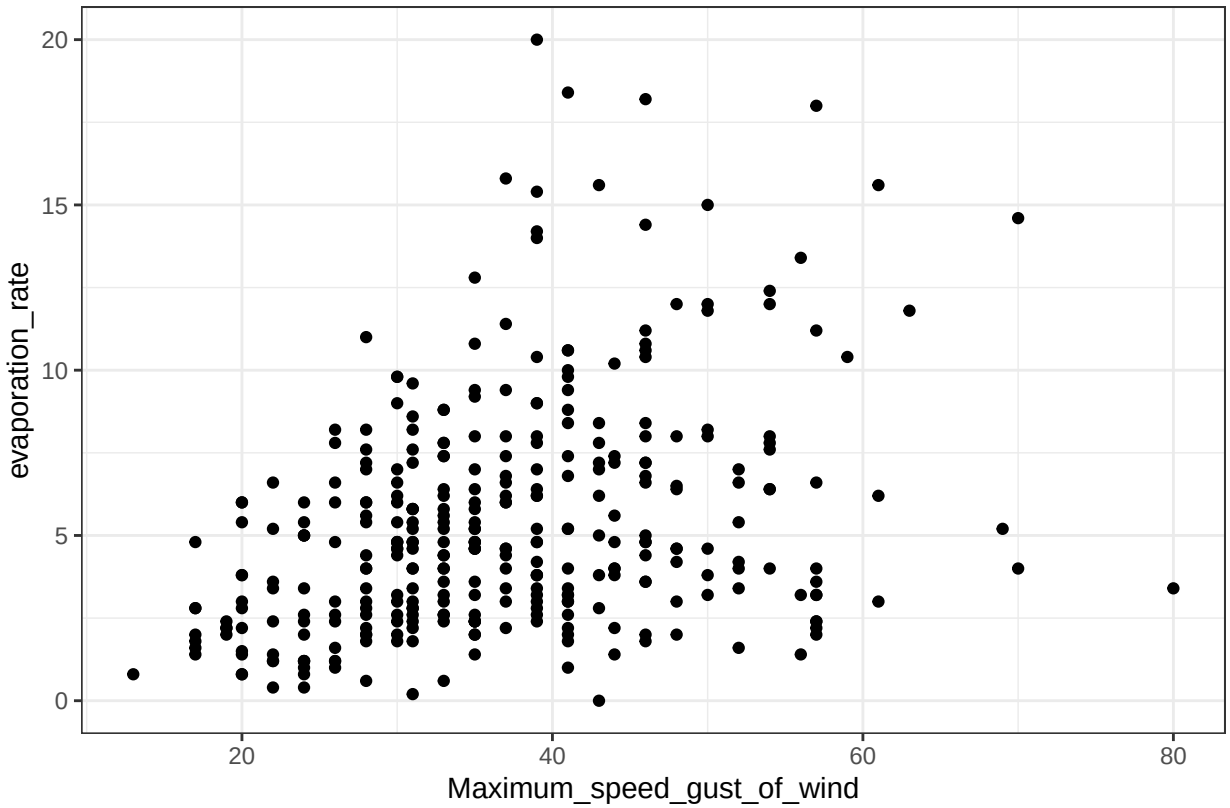


Figure 4: Scatterplot showing evaporation rate in relation against Speed of the maximum wind gust in kilometers per

```
# Figure 5
ggplot(melbourne, aes(x = minimum_temperature, y = evaporation_rate)) +
  geom_point() +
  theme_bw() +
  labs(caption = " Figure 5: Scatterplot showing evaporation rate in relation against Minimum Temperatu
  theme(plot.caption = element_text(hjust = 0.5))
```

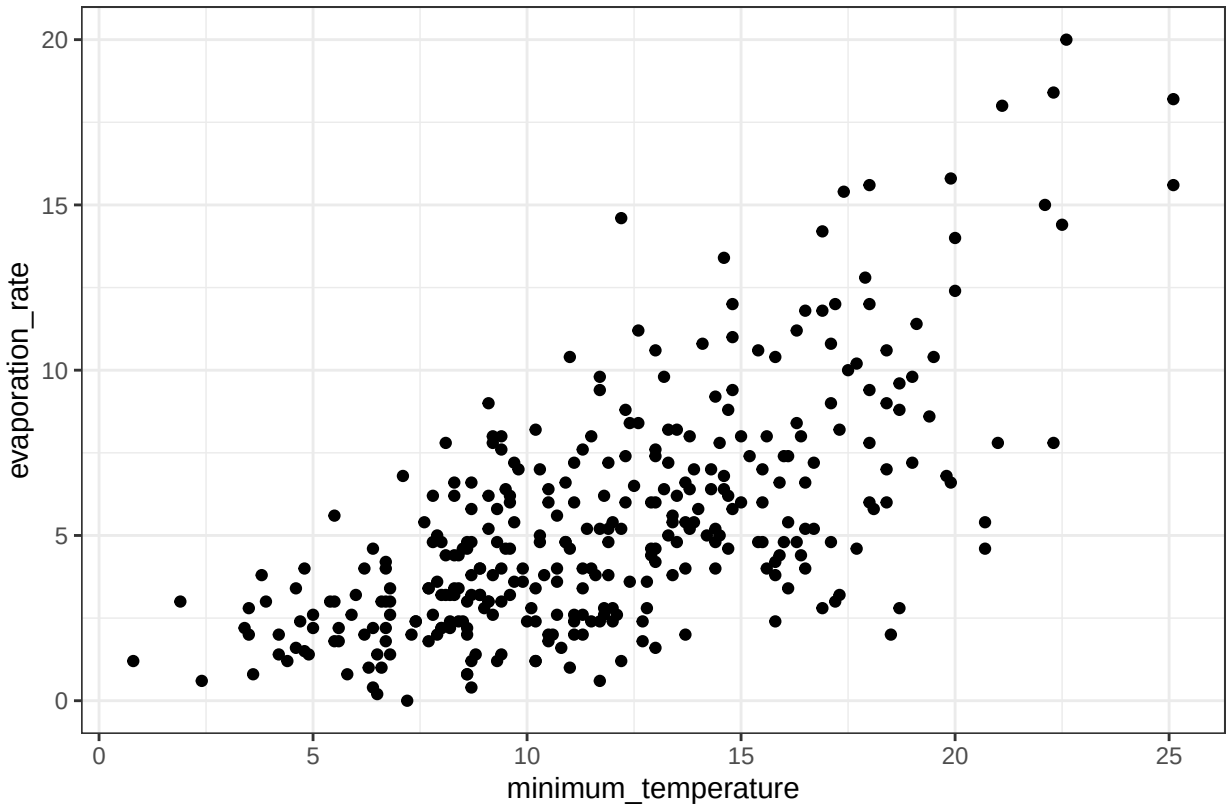


Figure 5: Scatterplot showing evaporation rate in relation against Minimum Temperataure

```
# Figure 6. A scatterplot showing relation of evaporation rate against Relative humidity, as measured a
ggplot(melbourne, aes(x= humidity_9am, y= evaporation_rate)) +
  geom_point() +
  theme_bw() +
  labs(caption = " Figure 6: A scatterplot showing relation of evaporation rate against Relative humidi
  theme(plot.caption = element_text(hjust = 0.5))
```

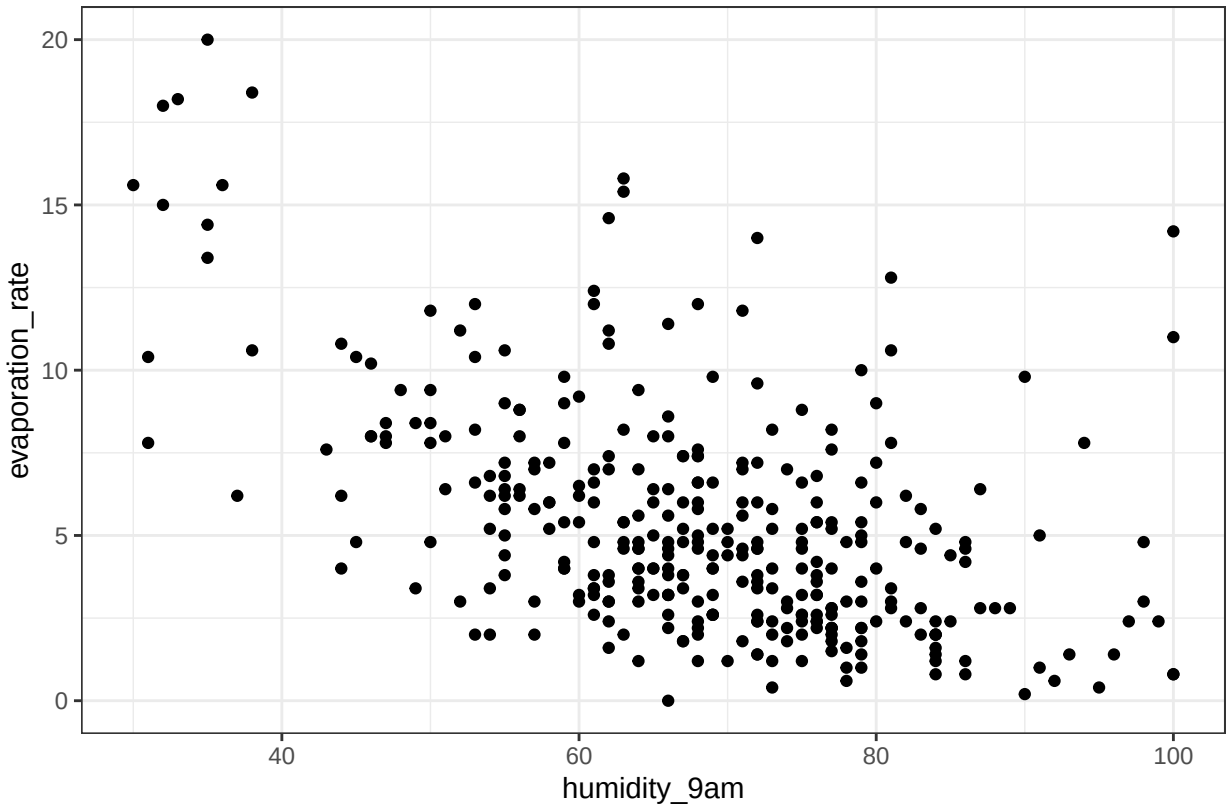


Figure 6: A scatterplot showing relation of evaporation rate against Relative humidity, as measured at 9am

```
# I will begin will fitting the full model with all possible predictors
lm_melbourne <- lm(evaporation_rate ~ month +
                     weekday+
                    Sunlight_hours +
                    Maximum_speed_gust_of_wind+
                    maximum_temperature +
                    minimum_temperature +
                    humidity_9am +
                    month: humidity_9am,
                   data=melbourne )
Anova(lm_melbourne)
```

```
## Anova Table (Type II tests)
##
## Response: evaporation_rate
##                              Sum Sq  Df F value    Pr(>F)
## month                         98.99  11  1.9517 0.0328695 *
## weekday                       36.36   6  1.3144 0.2503076
## Sunlight_hours                 1.68   1  0.3636 0.5469549
## Maximum_speed_gust_of_wind    70.24   1 15.2339 0.0001170 ***
## maximum_temperature            1.74   1  0.3767 0.5398587
## minimum_temperature          146.46   1 31.7633 3.974e-08 ***
## humidity_9am                 231.23   1 50.1463 9.940e-12 ***
## month:humidity_9am           158.47  11  3.1243 0.0005192 ***
## Residuals                   1401.76 304
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# from the above observations, we can conclude that "Sunlight_hours " has the highest p-value , therefo

lm_melbourne2 <- lm(evaporation_rate ~ month +
                     weekday+
                    Maximum_speed_gust_of_wind+
                    maximum_temperature +
                    minimum_temperature +
                    humidity_9am +
                    month: humidity_9am,
                   data=melbourne )
Anova(lm_melbourne2)
```

```
## Anova Table (Type II tests)
##
## Response: evaporation_rate
##                              Sum Sq  Df F value    Pr(>F)
## month                         98.59  11  1.9478 0.0332691 *
## weekday                       35.93   6  1.3013 0.2562310
## Maximum_speed_gust_of_wind    69.90   1 15.1901 0.0001195 ***
## maximum_temperature            0.87   1  0.1886 0.6644135
## minimum_temperature          211.93   1 46.0583 5.984e-11 ***
## humidity_9am                 232.75   1 50.5825 8.167e-12 ***
## month:humidity_9am           158.50  11  3.1315 0.0005047 ***
## Residuals                   1403.43 305
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#from the above observations and calculations we can conclude that "maximum_temperature" has highest p
lm_melbourne3 <- lm(evaporation_rate ~ month +
                     weekday+
                    Maximum_speed_gust_of_wind+
                    minimum_temperature +
                    humidity_9am +
                    month: humidity_9am,
                   data=melbourne )
Anova(lm_melbourne3)
```

```
## Anova Table (Type II tests)
##
## Response: evaporation_rate
##                              Sum Sq  Df F value    Pr(>F)
## month                        118.50  11  2.3474 0.0086669 **
## weekday                       35.17   6  1.2774 0.2673598
## Maximum_speed_gust_of_wind    71.27   1 15.5295 0.0001007 ***
## minimum_temperature          241.61   1 52.6464 3.303e-12 ***
## humidity_9am                 262.74   1 57.2513 4.533e-13 ***
## month:humidity_9am           161.08  11  3.1908 0.0004033 ***
## Residuals                   1404.30 306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#from the above observations and calculations we can conclude that "weekday" has highest p value, there
```

```r
lm_melbourne4 <- lm(evaporation_rate ~ month +
                     Maximum_speed_gust_of_wind+
                     minimum_temperature +
                     humidity_9am +
                     month: humidity_9am,
                    data=melbourne )
Anova(lm_melbourne4)
```

```
## Anova Table (Type II tests)
##
## Response: evaporation_rate
##                             Sum Sq  Df F value    Pr(>F)
## month                       119.06  11  2.3459 0.0086754 **
## Maximum_speed_gust_of_wind   71.05   1 15.3999 0.0001071 ***
## minimum_temperature         245.37   1 53.1819 2.522e-12 ***
## humidity_9am                253.36   1 54.9155 1.189e-12 ***
## month:humidity_9am          167.05  11  3.2917 0.0002730 ***
## Residuals                  1439.47 312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(lm_melbourne4)
```

```
##
## Call:
## lm(formula = evaporation_rate ~ month + Maximum_speed_gust_of_wind +
##     minimum_temperature + humidity_9am + month:humidity_9am,
##     data = melbourne)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8270 -1.1603 -0.0941  1.1352  9.5483
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 6.358170   2.249806   2.826 0.005016 **
## month2                      0.556089   3.259698   0.171 0.864653
## month3                      5.397192   2.651877   2.035 0.042672 *
## month4                      1.534093   3.035119   0.505 0.613601
## month5                     -4.320274   3.410596  -1.267 0.206200
## month6                     -9.303646   3.924989  -2.370 0.018378 *
## month7                     -7.931404   3.518775  -2.254 0.024888 *
## month8                     -9.266054   3.260777  -2.842 0.004783 **
## month9                     -2.484941   3.070197  -0.809 0.418916
## month10                    -6.044544   3.080561  -1.962 0.050633 .
## month11                    -0.648342   2.707044  -0.240 0.810874
## month12                    -0.353236   2.844030  -0.124 0.901235
## Maximum_speed_gust_of_wind  0.052284   0.013323   3.924 0.000107 ***
## minimum_temperature         0.329534   0.045187   7.293 2.52e-12 ***
## humidity_9am               -0.084676   0.031560  -2.683 0.007685 **
```

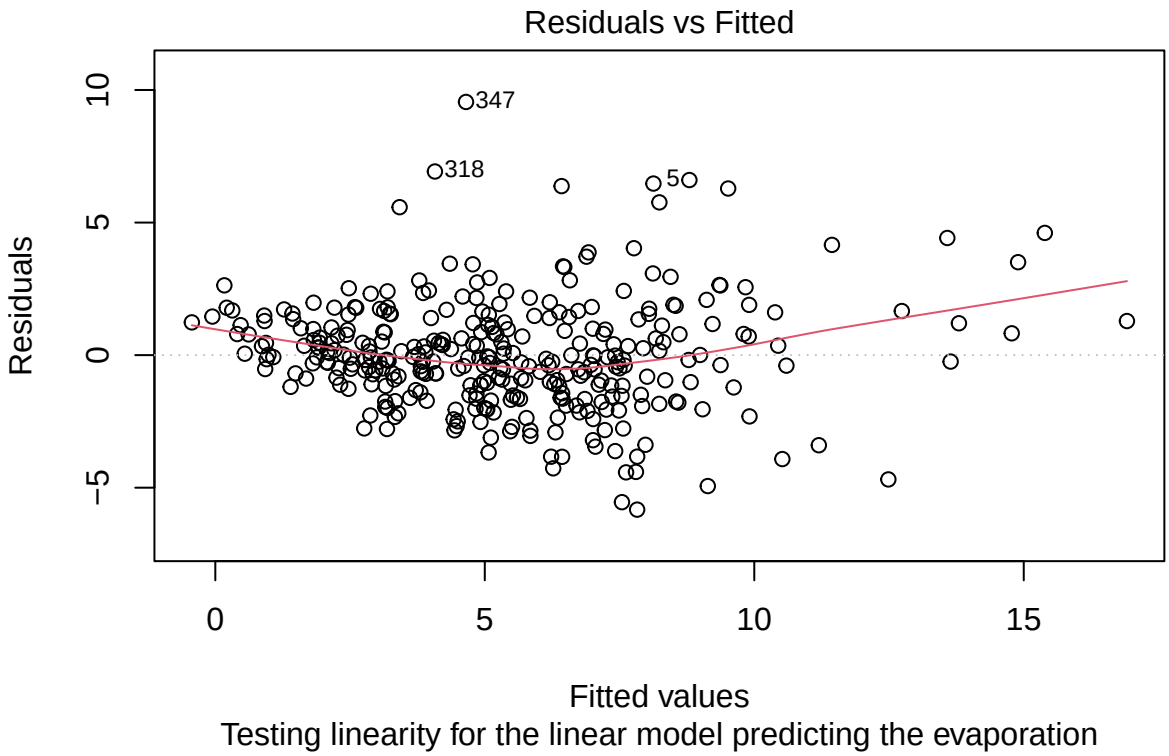
```
## month2:humidity_9am        -0.016776   0.049737  -0.337 0.736121
## month3:humidity_9am        -0.082493   0.040363  -2.044 0.041814 *
## month4:humidity_9am        -0.040056   0.046221  -0.867 0.386815
## month5:humidity_9am         0.028592   0.048794   0.586 0.558321
## month6:humidity_9am         0.091116   0.051724   1.762 0.079119 .
## month7:humidity_9am         0.080370   0.050435   1.594 0.112049
## month8:humidity_9am         0.111816   0.047890   2.335 0.020185 *
## month9:humidity_9am         0.018795   0.047606   0.395 0.693259
## month10:humidity_9am        0.086485   0.047091   1.837 0.067226 .
## month11:humidity_9am        0.004915   0.040430   0.122 0.903316
## month12:humidity_9am       -0.004938   0.041792  -0.118 0.906023
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.148 on 312 degrees of freedom
## Multiple R-squared:  0.6534, Adjusted R-squared:  0.6256
## F-statistic: 23.52 on 25 and 312 DF,  p-value: < 2.2e-16
```

All terms are hence significant.

It can be observed that all varaibles behaved as expected from the relationship observed in Bivariate analysis.

##linearity

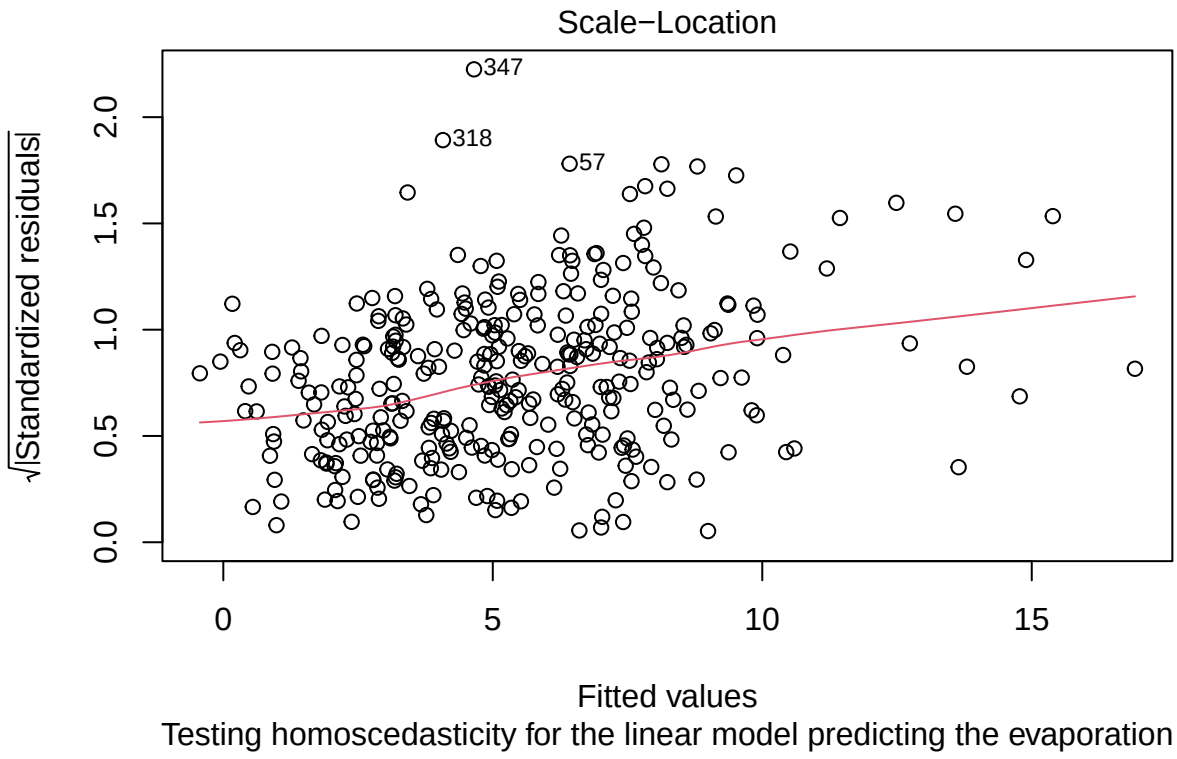
```
plot(lm_melbourne4, which = 1, sub= "Testing linearity for the linear model predicting the evaporation")
```



Testing linearity for the linear model predicting the evaporation

#Homoscedasticity

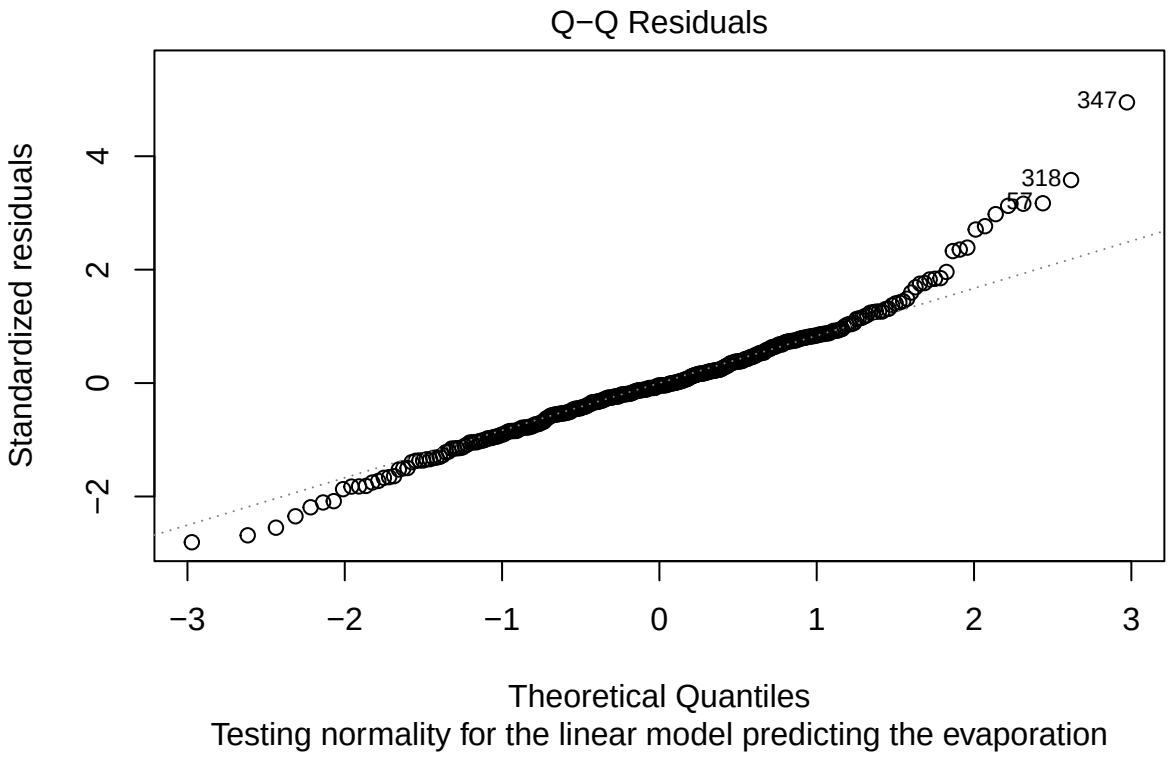
```
plot(lm_melbourne4, which = 3, sub= "Testing homoscedasticity for the linear model predicting the evapo
```



Testing homoscedasticity for the linear model predicting the evaporation

##Normality

```
plot(lm_melbourne4, which=2, sub= "Testing normality for the linear model predicting the evaporation")
```



Testing normality for the linear model predicting the evaporation

Prediction

```
predicted_melbourne = tibble(date=dmy(c("21/2/2024",

"26/12/2024", "11/1/2024","11/7/2024")),

month=as_factor(c(2,12,1,7)),
Maximum_speed_gust_of_wind= c(23.2,76,44.3,10.6),
minimum_temperature= c(13.8, 16.4, 26.5, 6.8),
humidity_9am = c(74,37,35,76),
Sunlight_hours = c(3.4,9,9.6,4),
weekday=wday(dmy(c("21/2/2024", "26/12/2024",
"11/1/2024", "11/7/2024"))))
predicted_melbourne = predicted_melbourne[c(3,1,4,2),]
predicted_melbourne
```

```
## # A tibble: 4 x 7
##   date       month Maximum_speed_gust_of_wind minimum_temperature humidity_9am
##   <date>     <fct>                      <dbl>               <dbl>        <dbl>
## 1 2024-01-11 1                           44.3                26.5           35
## 2 2024-02-21 2                           23.2                13.8           74
## 3 2024-07-11 7                           10.6                 6.8           76
## 4 2024-12-26 12                          76                  16.4           37
```

```
## # i 2 more variables: Sunlight_hours <dbl>, weekday <dbl>
```

```
predicted_interval <- predict(lm_melbourne4, predicted_melbourne, interval = "prediction", level=0.95)
predicted_interval <- tibble(Date = predicted_melbourne$date,
'Lower_Bound' = predicted_interval[,2],
'Expected evaporation Rate' = predicted_interval[,1],
'Upper_Bound' = predicted_interval[,3] )
predicted_interval %>%
knitr::kable(digits = 1,format.args=list(big.mark= ","),
caption = "Table 1: Prediction of intervals of the  evaporation rates in Melbourne on specific dates as
```

Table 1: Table 1: Prediction of intervals of the evaporation rates in Melbourne on specific dates as requested by Melbourne Water Corporation

Date	Lower_Bound	Expected evaporation Rate	Upper_Bound
2024-01-11	9.7	14.4	19.2
2024-02-21	0.8	5.2	9.5
2024-07-11	-3.5	0.9	5.3
2024-12-26	7.4	12.1	16.8