Melbourne Water Corporation - Heaps moist

Satyaveer Pattanaik

2023-02-20

Executive summary

The Melbourne Water Corporation (MWC) manages the supply of water in Melbourne, Australia. In light of recent changes in Melbourne's climate, MWC want to build a new model assessing the effects of Melbourne's day-to-day weather's on evaporation, to aid in their management of their Cardinia Reservoir, in the city's South East. Predicting the rate of evaporation is crucial for the stability of the city's water supply.

This report outlines the temporal and meteorological factors that affect evaporation on a given day and the way these factors affect evaporation. I build a model that incorporates some of these factors, and use it to predict evaporation rates for individual days at the Cardinia reservoir.

The model identifies January 13, 2020 to be a day when evaporation exceeds 10 mm, which could be an issue. For this day, the MWC should take temporary measures to ensure a continuous supply of water, which may include transferring water from its Silvan Reservoir upstream.

Methods

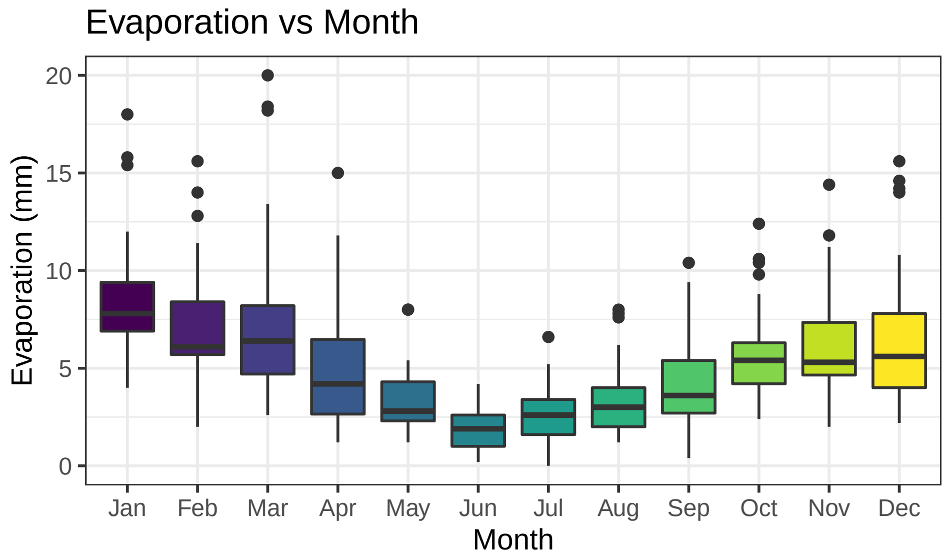
For our analysis we use a dataset that contains Melbourne's weather observations, including evaporation, for the previous financial year. We are interested in the following potential influences on amount of evaporation in a day:

- Month
- Day of the week
- Maximum temperature (in degrees Celsius)
- Minimum temperature (in degrees Celsius)
- Relative humidity % (as measured at 9am)

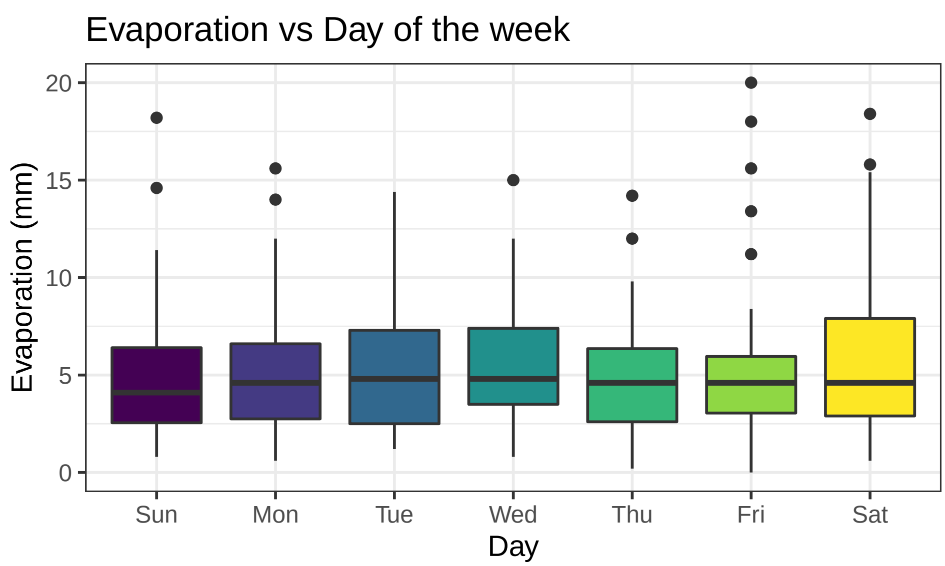
This section describes how each factor affects evaporation. We then go on to build our model, incorporating some of these factors.

This analysis (and the rest of the report) is prepared using RStudio.

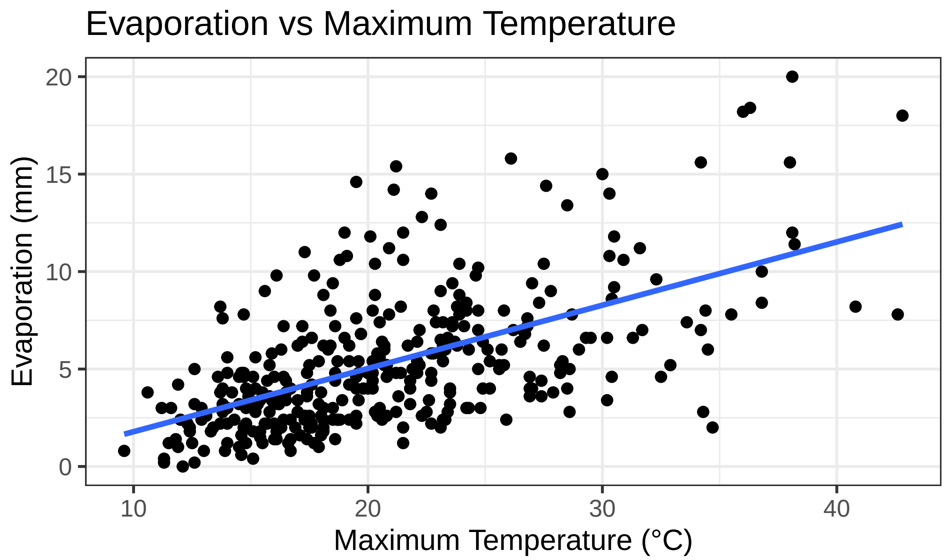
Factors affecting Evaporation



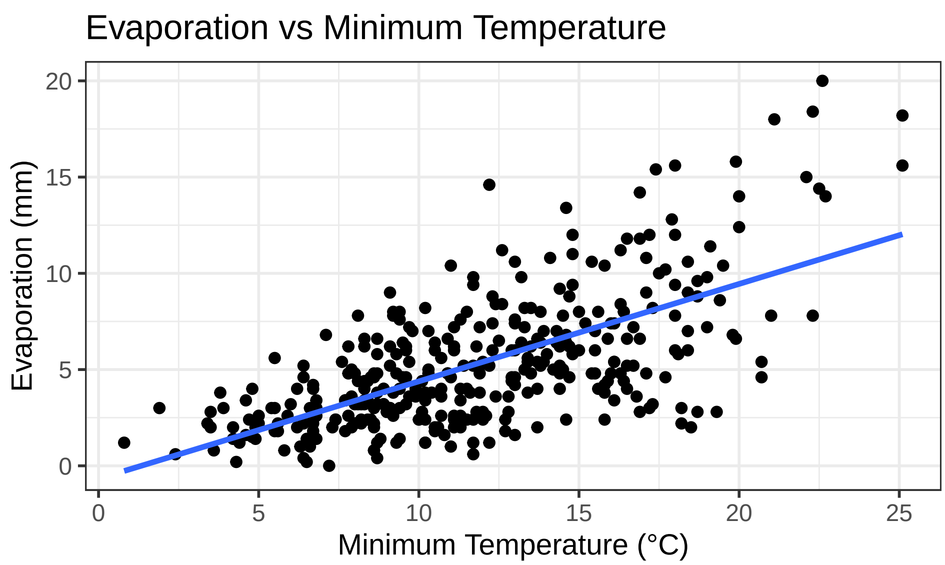
The amount of evaporation does seem to vary with each month (or more logically, with each season). Evaporation is highest during the summer months of December-February and lowest during the winter months of June-August.



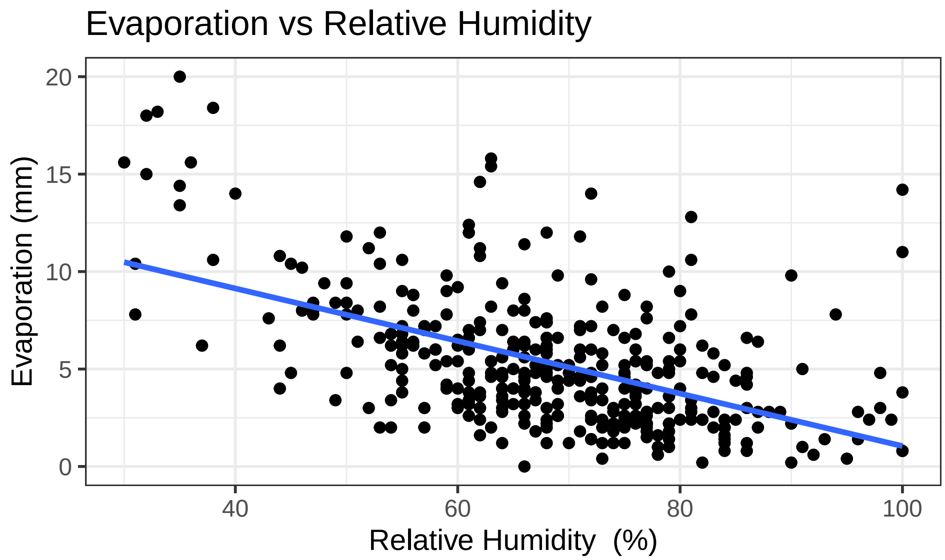
The day of the week does not have any effect on the amount of evaporation.



There exists a positive relationship between Evaporation and Maximum temperature. We can observe that the amount of evaporation on a given day increases as the maximum temperature increases.



Similar to the previous graph, we can observe that the amount of evaporation on a given day also increases as the minimum temperature increases.



We can notice that Evaporation and Relative Humidity has a negative relationship. The amount of evaporation on a given day decreases with an increase in relative humidity.

Model selection

After careful deliberation, the final model is significantly affected by the following factors:

1. Month
2. Minimum Temperature
3. Relative Humidity
4. The effect of relative humidity in different months

It is not surprising to see that the day of the week does not affect evaporation as we had seen earlier (refer to 'Evaporation vs Day of the week' plot in 'Factors affecting Evaporation' section).

It is surprising to see maximum temperature not make the cut into our model, as it did have a significant effect on the amount of evaporation in our analysis earlier (refer to 'Evaporation vs Maximum Temperature' plot in 'Factors affecting Evaporation' section). But this omission is justified as minimum temperature has the same positive effect on evaporation, making the effect of maximum temperature redundant.

Assumptions of the model

Our model satisfies the linearity, homoscedasticity and normality assumptions.

However, we cannot be too sure if our model satisfies the independence assumption. Independence relies on subjects being independent of one another which is not necessarily the case here. The weather on a given day may be affected by the weather on the previous day.

Results

The intercept of our model is 6.533. This means if all the variables (e.g temperature, month) are considered null and void, the amount of evaporation will be 6.53 mm.

The slope of minimum temperature is 0.369. This means that, keeping all other variables constant, if the minimum temperature increases by 1°C, the amount of evaporation increases by 0.369 mm.

The slope of relative humidity is -0.083. This means that, keeping all other variables constant, if the relative humidity increases by 1%, the amount of evaporation decreases by 0.083 mm.

The slope of the Month variable varies with each month, just as we found in our analysis previously (refer to 'Evaporation vs Month' plot in 'Factors affecting Evaporation' section). For example for the month of August, the slope is -4.504. This means that on a given day in August, the evaporation is expected to decrease by 4.504 mm. This makes sense because as seen before, during winter months evaporation is very low.

Discussion

Date	Mininum Temperature (°C)	Maximum Temperature (°C)	Relative Humidity (%)	Lower bound for mean	Expected evaporation	Upper bound for mean
February 29, 2020	13.8	23.2	74	4.41	5.51	6.60
December 25, 2020	16.4	31.9	57	7.59	8.61	9.62
January 13, 2020	26.5	44.3	35	12.77	14.87	16.98
July 6, 2020	6.8	10.6	76	1.34	2.27	3.19

We can say with 95% confidence that:

- the average evaporation on a February day with a minimum temperature of 13.8°C, a maximum temperature of 23.2°C and relative humidity of 74% in the financial year 2018-19 is between 4.41 and 6.6 mm.
- the average evaporation on a December day with a minimum temperature of 16.4°C, a maximum temperature of 31.9°C and relative humidity of 57% in the financial year 2018-19 is between 7.59 and 9.62 mm.
- the average evaporation on a January day with a minimum temperature of 26.5°C, a maximum temperature of 44.3°C and relative humidity of 35% in the financial year 2018-19 is between 12.77 and 16.98 mm.
- the average evaporation on a July day with a minimum temperature of 6.8°C, a maximum temperature of 10.6°C and relative humidity of 76% in the financial year 2018-19 is between 1.34 and 3.19 mm.

On Jan 13, 2020 we could predict evaporation to be more than 10mm with 95% confidence.

For the remaining days, we could predict that evaporation would be less than 10mm with 95% confidence.

Conclusion

The Melbourne Water Corporation (MWC) manages the supply of water in Melbourne, Australia. In light of recent changes in Melbourne's climate, MWC want to build a new model assessing the effects of Melbourne's day-to-day weather's on evaporation, to aid in their management of their Cardinia Reservoir, in the city's South East. Predicting the rate of evaporation is crucial for the stability of the city's water supply.

This report has outlined the temporal and meteorological factors that affect evaporation on a given day and the way these factors affect evaporation. I built a model that incorporates some of these factors, and have used it to predict evaporation rates for extreme scenarios (as provided by MWC) at the Cardinia reservoir.

After implementation of the model it was determined that:

- On Jan 13, 2020 we could predict evaporation to be more than 10mm with 95% confidence. On this day, the MWC should take temporary measures to ensure a continuous supply of water, which may include transferring water from its Silvan Reservoir upstream.
- For the remaining days, we could predict that evaporation would be less than 10mm with 95% confidence.

Appendix

Taming the data

```
#loading packages and the dataset
pacman::p_load(tidyverse, caret, inspectdf, lubridate)
melbourne <- read_csv("melbourne.csv")
inspect_cat(melbourne)
inspect_types(melbourne)

#converting Date column to date format
melbourne <-
  melbourne %>%
  mutate(Date = ymd(Date))

#making separate columns for new variables: "Month" and "Day"
melbourne <-
  melbourne %>%
  mutate(Month = as.factor(month(Date, label = TRUE)),
         Day = as.factor(wday(Date, label = TRUE)))


#renaming columns for convenience
melbourne <-
  melbourne %>%
  rename(rel_hum = `9am relative humidity (%)`,
         evap = `Evaporation (mm)`,
         max_temp = `Maximum Temperature (Deg C)`,
         min_temp = `Minimum temperature (Deg C)`,
         )
```

Factors affecting Evaporation

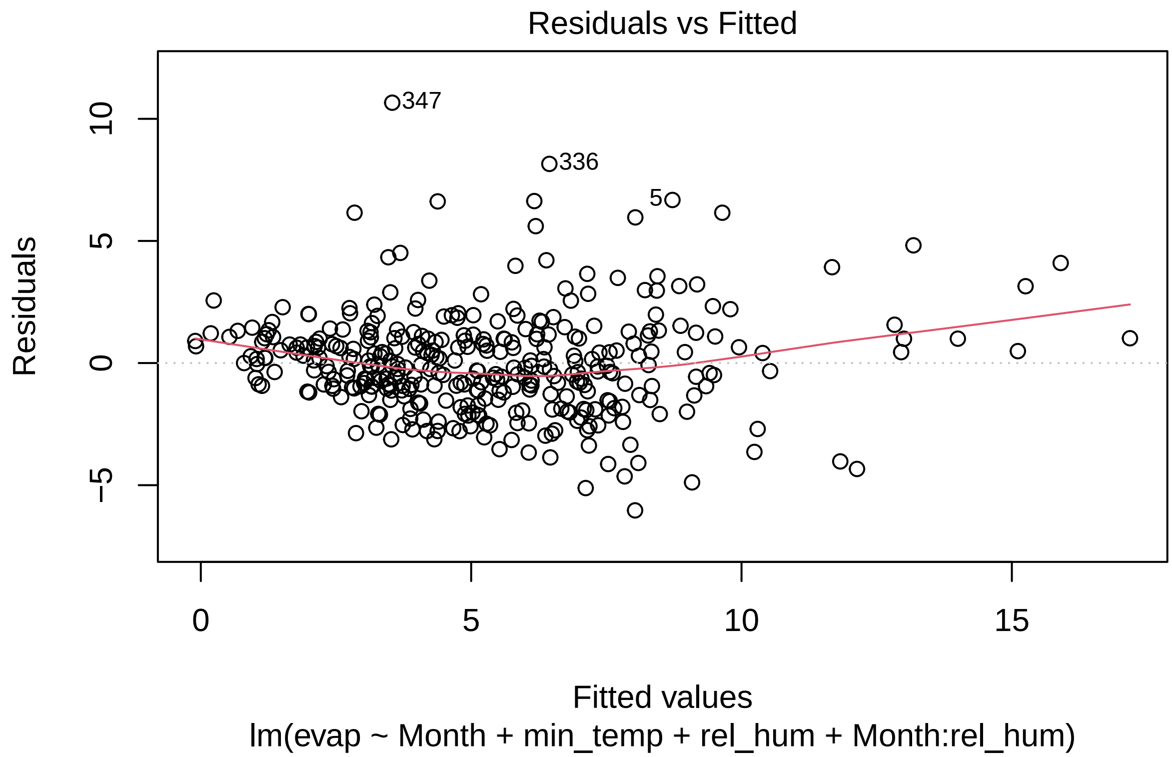
```r
#relationship between Evaporation and Month
melbourne %>%
  ggplot(aes(Month, evap))+
  geom_point()+
  theme_bw()

#relationship between Evaporation and Day of the week
melbourne %>%
  ggplot(aes(Day, evap))+
  geom_point()+
  theme_bw()

#relationship between Evaporation and max temperature
melbourne %>%
  ggplot(aes(max_temp, evap))+
  geom_point()+
  theme_bw()

#relationship between Evaporation and min temperature
melbourne %>%
  ggplot(aes(min_temp, evap))+
  geom_point()+
  theme_bw()

#relationship between Evaporation and relative humidity
melbourne %>%
  ggplot(aes(rel_hum, evap))+
  geom_point()+
  theme_bw()
```

Model selection

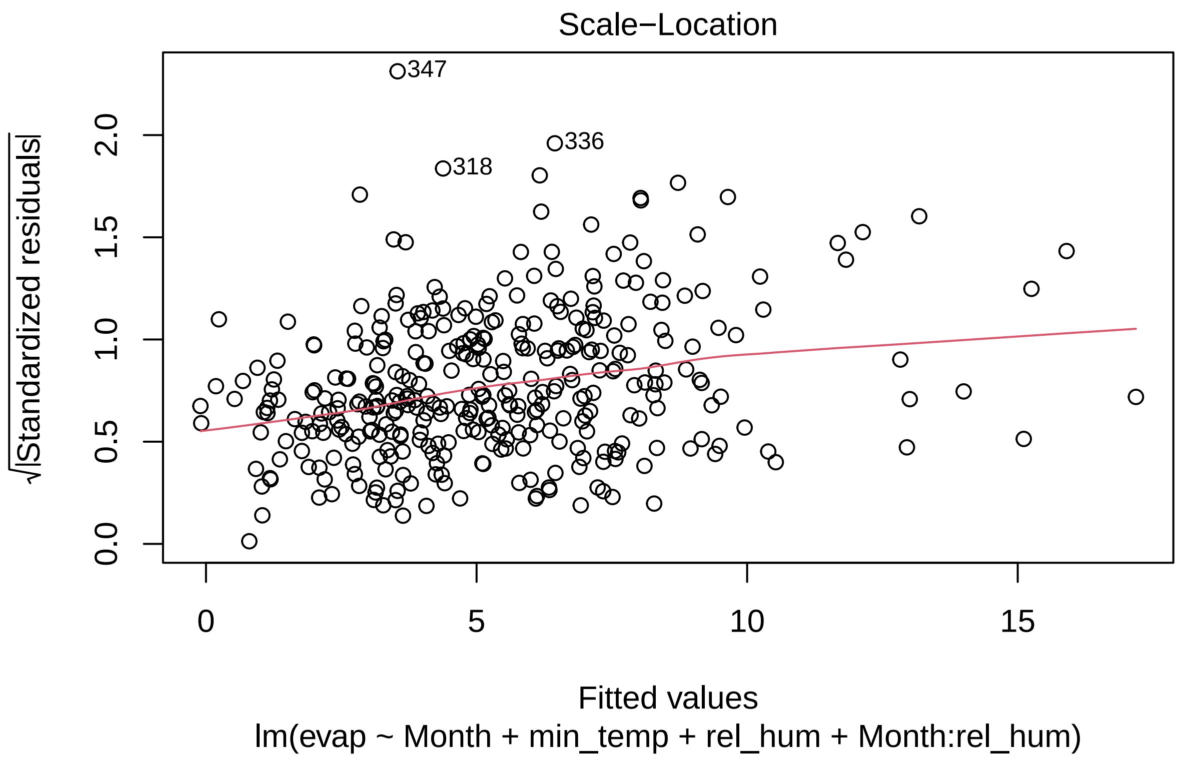
```r
#model with all predictors
melb_all <- lm(evap ~ Month + Day + max_temp + min_temp +
                 rel_hum + Month:rel_hum, data = melbourne)
summary(melb_all)
anova(melb_all)

#model without Day
melb_1 <-  lm(evap ~ Month + max_temp + min_temp +
                rel_hum + Month:rel_hum, data = melbourne)
summary(melb_1)
anova(melb_1)

#model without max_temp
melb_2 <-  lm(evap ~ Month + min_temp +
                rel_hum + Month:rel_hum, data = melbourne)
summary(melb_2)
anova(melb_2)
```

Testing Assumptions of the model

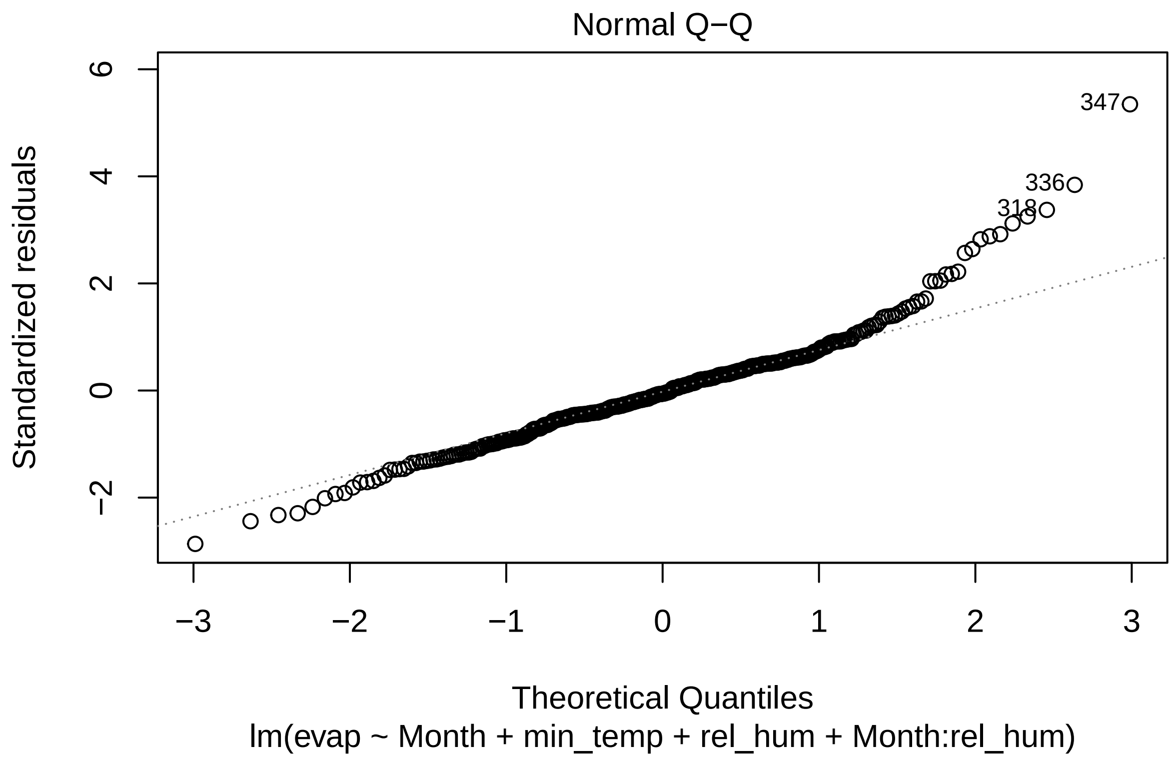
```r
### Linearity

plot(melb_2, which = 1)
```




```r
# There are no trends in the residual versus fitted plot. Hence, our model
# satisfies the linearity assumption.

### Homoscedasticity

plot(melb_2, which = 3)
```




```
# There is a slight bend in the red line due to the effect of outliers
# but there is no noticable trend in the spread of the data. Hence, our model
# satisfies the homoscedasticity assumption.

### Normality

plot(melb_2, which = 2)
```




```
# The points between -1.5 and 1.5 on the x-axis (which is most of the data) lie
# along the dotted line. The residuals are mostly normally distributed, with extreme
# values at both ends. Hence, our model satisfies the normality assumption.

### Independence

# Independence relies on subjects being independent of one another which is not
# necessarily the case here. The weather on a given day is affected by the weather
# on the previous day.
```

```
new_data1 <- tibble(
  min_temp = c(13.8, 16.4, 26.5, 6.8),
  max_temp = c(23.2, 31.9, 44.3, 10.6),
  rel_hum = c(74, 57, 35, 76),
  Month = c("Feb", "Dec", "Jan", "Jul"),
  date = c("February 29, 2020", "December 25, 2020", "January 13, 2020",
           "July 6, 2020")
  )


scenarios <- tibble(
  date = c( "29-02-2020", "25-12-2020" ,"13-01-2020" ,"06-07-2020" ),
 `Minimum temperature (Deg C)` = c(13.8, 16.4, 26.5, 6.8),
 `Maximum Temperature (Deg C)` = c(23.2, 31.9, 44.3, 10.6),
 `9am relative humidity (%)` = c(74, 57, 35, 76),
```

```r
Predicted_value = predict(melb_2, newdata = new_data1, interval = "prediction"))

scenarios %>% knitr::kable(digits = 0, format.args = list(big.mark = ","))
```

date	Minimum temperature (Deg C)	Maximum Temperature (Deg C)	9am relative humidity (%)	Predicted_value
29-02-2020	14	23	74	5.506144
25-12-2020	16	32	57	8.605772
13-01-2020	26	44	35	14.872285
06-07-2020	7	11	76	2.265493

Discussion

```r
new_data = tibble(
  min_temp = c(13.8, 16.4, 26.5, 6.8),
  max_temp = c(23.2, 31.9, 44.3, 10.6),
  rel_hum = c(74, 57, 35, 76),
  Month = c("Feb", "Dec", "Jan", "Jul"),
  date = c("February 29, 2020", "December 25, 2020", "January 13, 2020",
           "July 6, 2020")
  )

preds <- predict(melb_2, newdata = new_data, interval = "confidence", level = 0.95)
```