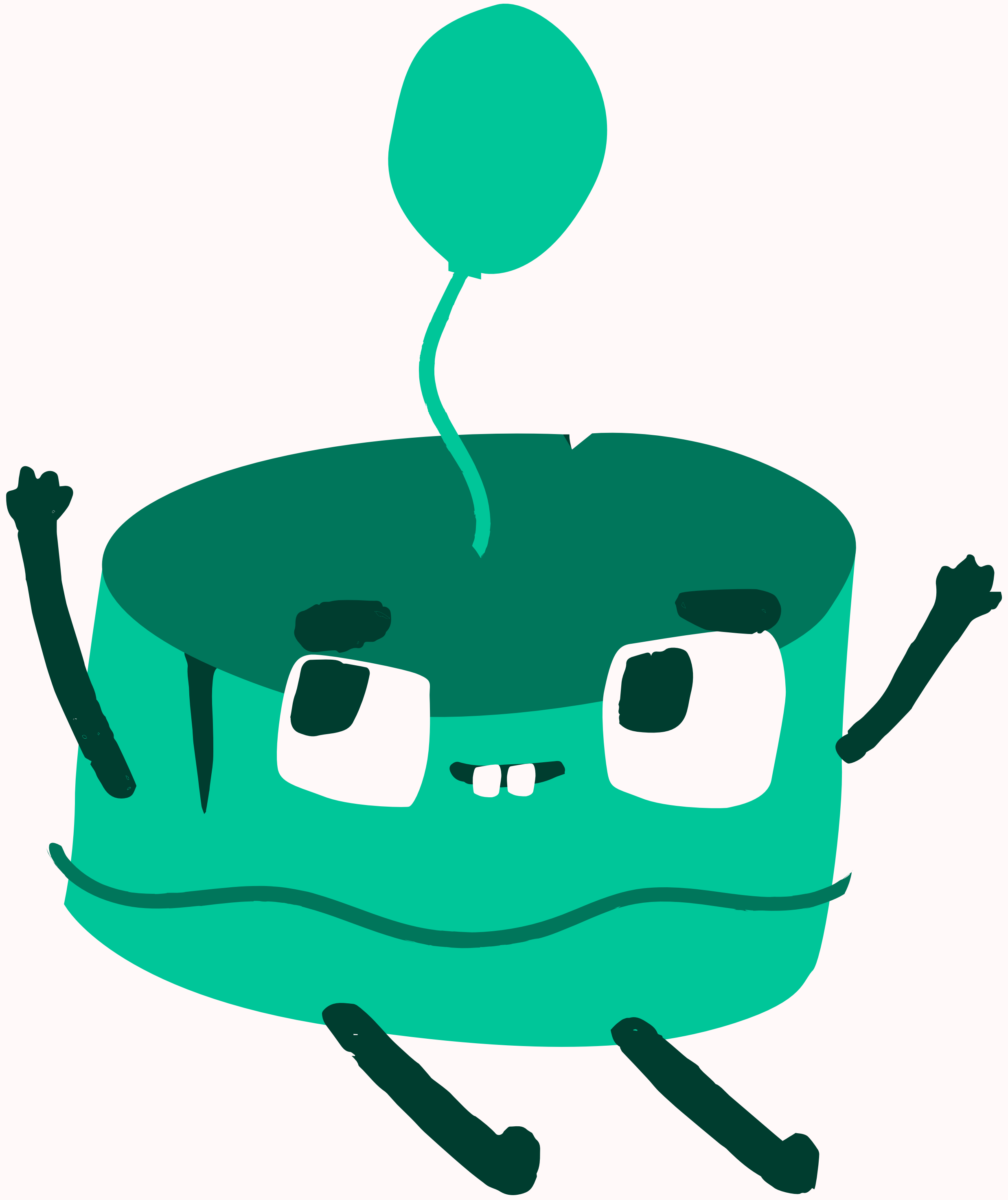
DECISION RULES

- Daniel Alejandro Morales Castillo

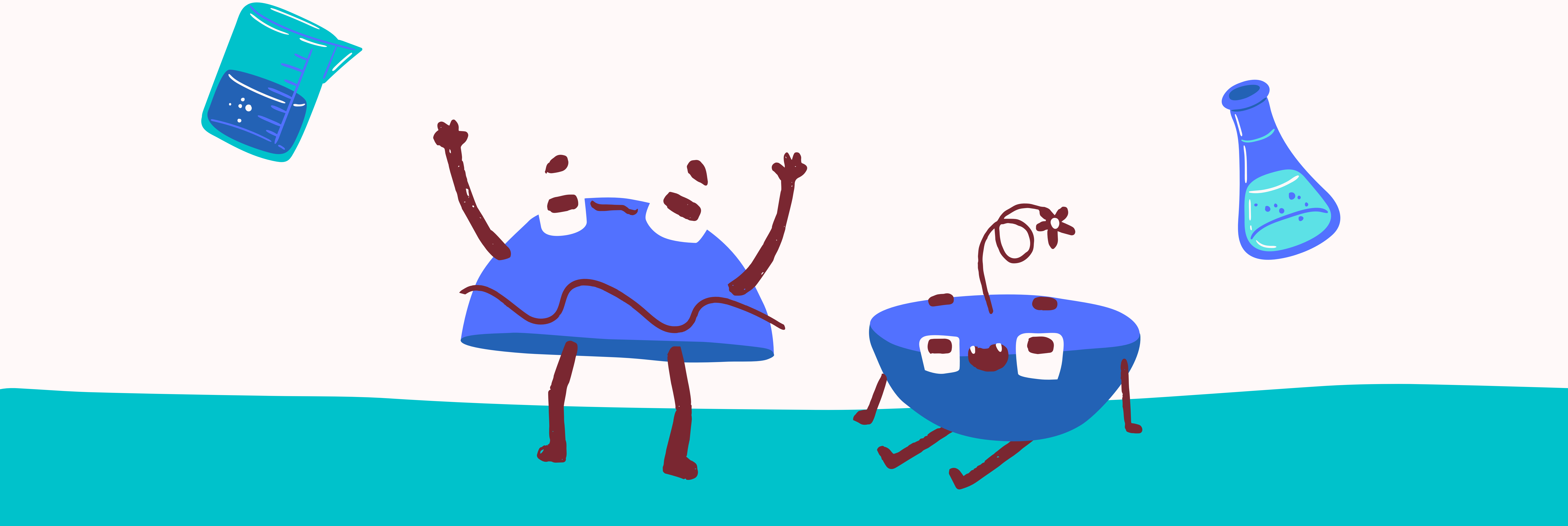


ABSTRACT

With the algorithm "decision rules (1r)" and the columns from our dataset, precipitation, maximum temperature, minimum temperature and wind, it is planned to predict the type of weather of a specific day; drizzle, rain, sun, snow, fog.

INDEX

BASIC DATA ANALYSIS



DATA DOMAIN

weather

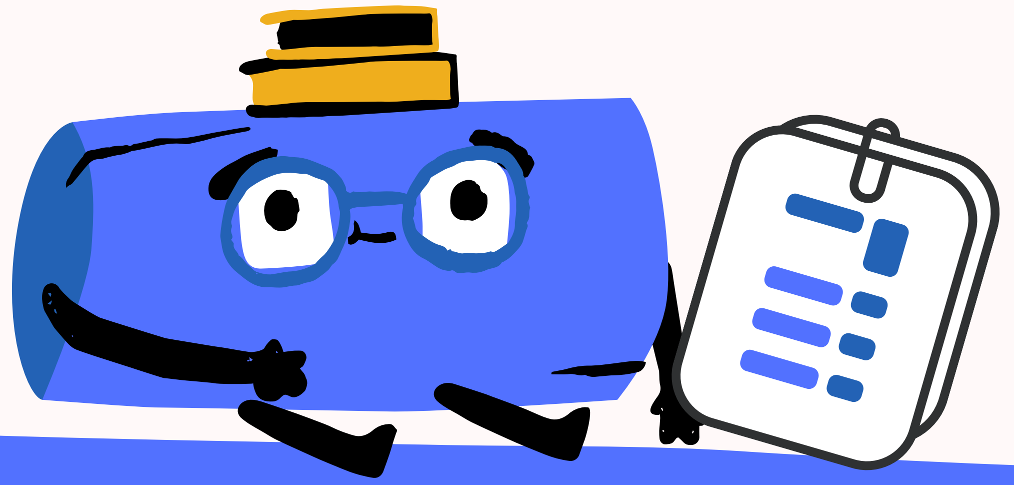
The conditions in the air above the earth such as wind, rain, or temperature, especially at a particular time over a particular area.

precipitation

Water that falls from the clouds towards the ground, especially as rain or snow.

temperature

the measured amount of heat in a place or in the body.



VARIABLES

date

When was taken the data

precipitation

water level

temp_max

the max temperature

temp_min

the min temperature

wind

wind speed

weather

The one that we will predict

How the data was recollected?

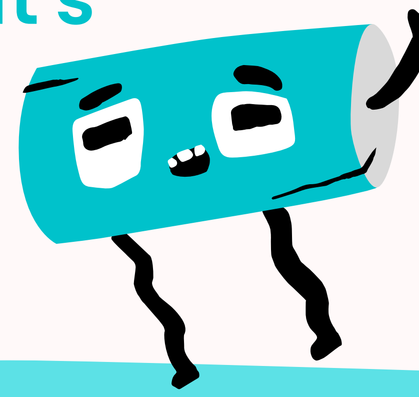
Was collected through websites that show the weather, like weather.com, wunderground.com, etc. This dataset was obtained from the Kaggle website. It is a compilation of different dates and their climatic elements.

Limitations of study

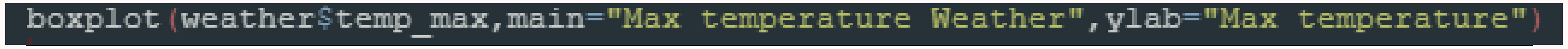
It is limited to only Seattle weather and 2012 to 2015 years, with 1462 observations.

Disadvantages

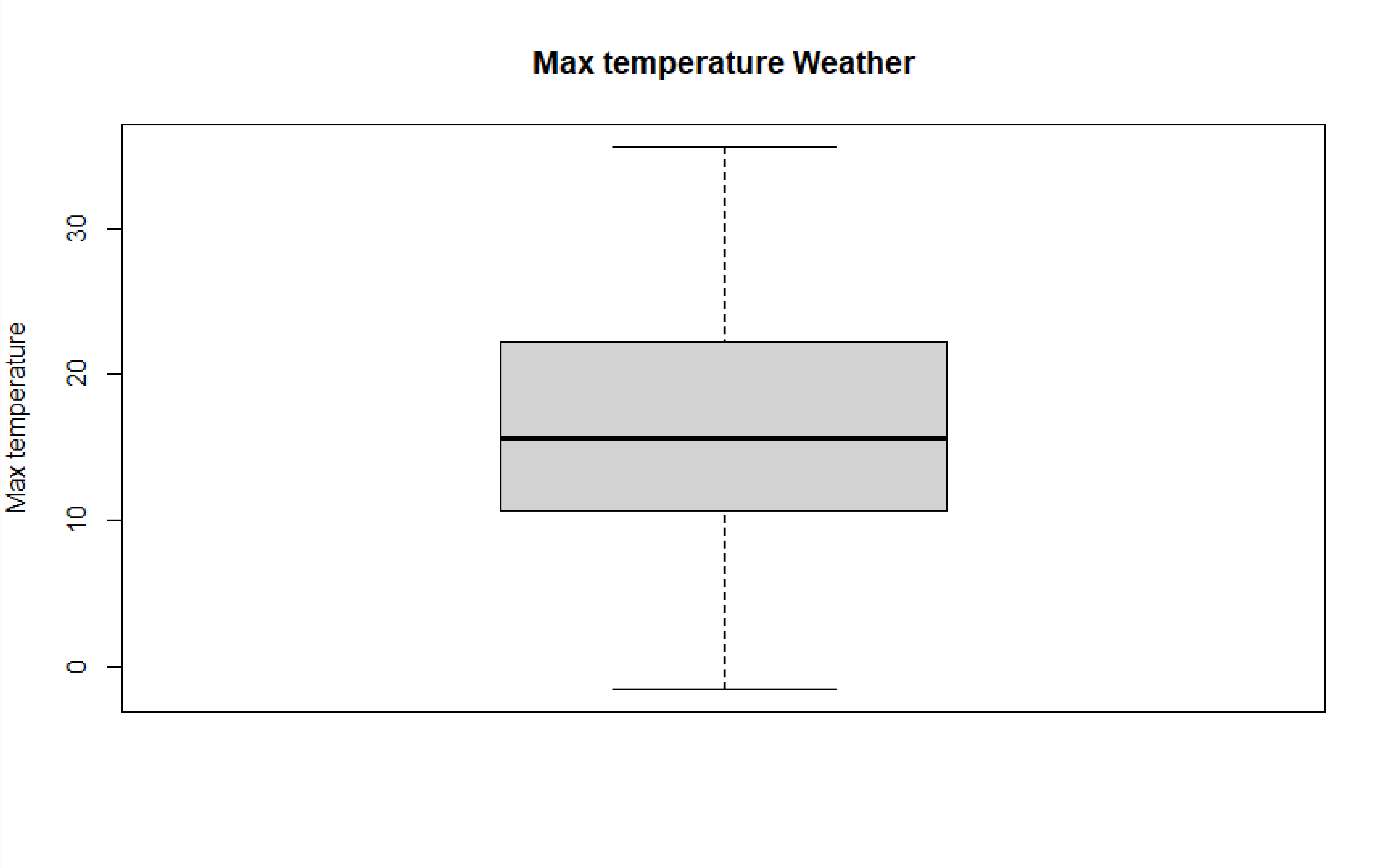
Although weather predictions are mostly accurate we know that sometimes weather changes pretty fast, and if we want to use those 2012 data in 2022, the weather might be quite a bit different. There are some rows were the precipitation is 0 but the weather it's rainy.



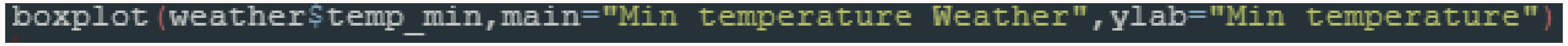
INTERESTING PLOTS :]



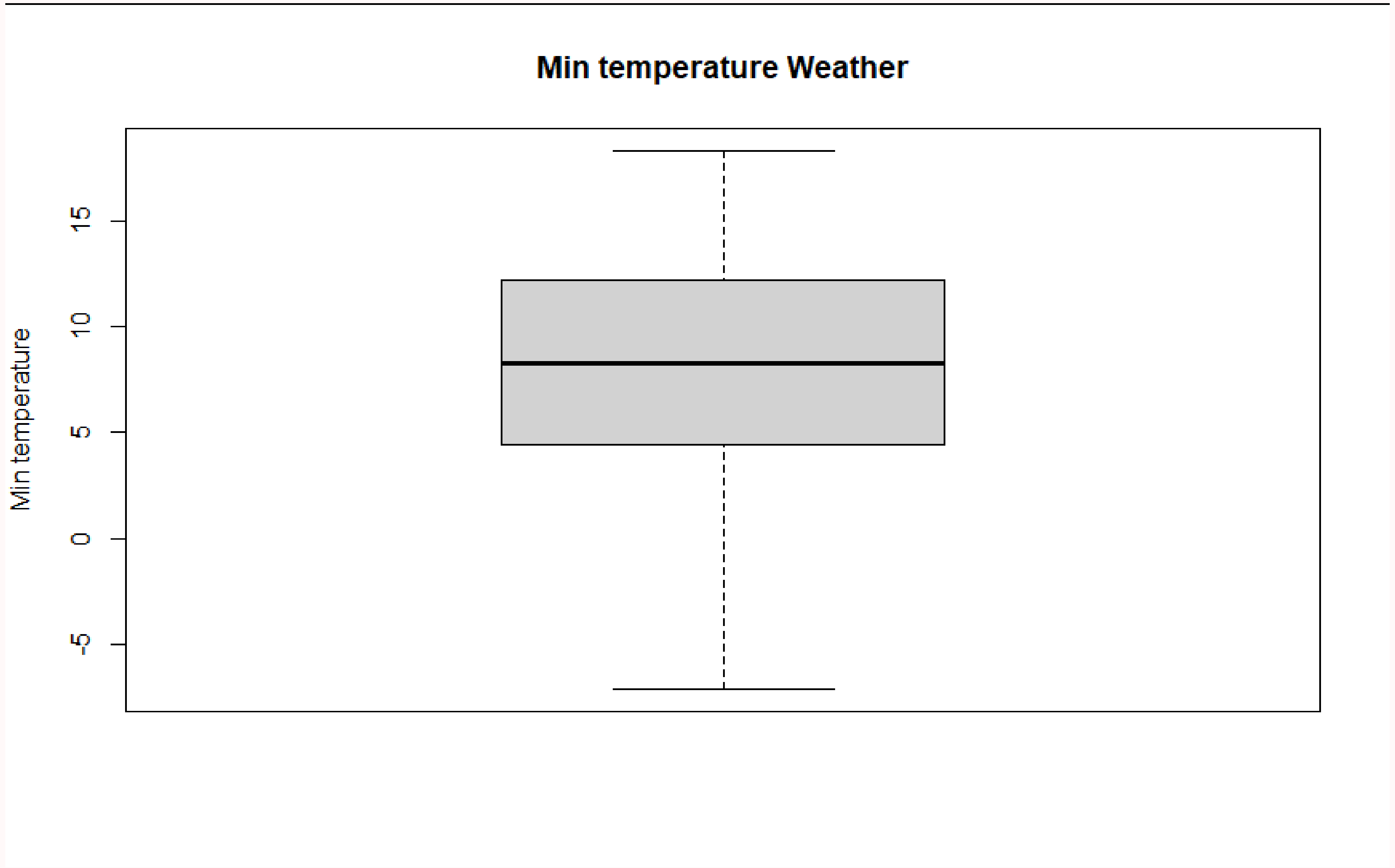
```
boxplot(weather$temp_max,main="Max temperature Weather",ylab="Max temperature")
```



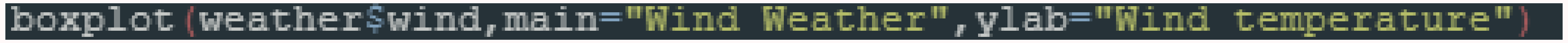
INTERESTING PLOTS :]



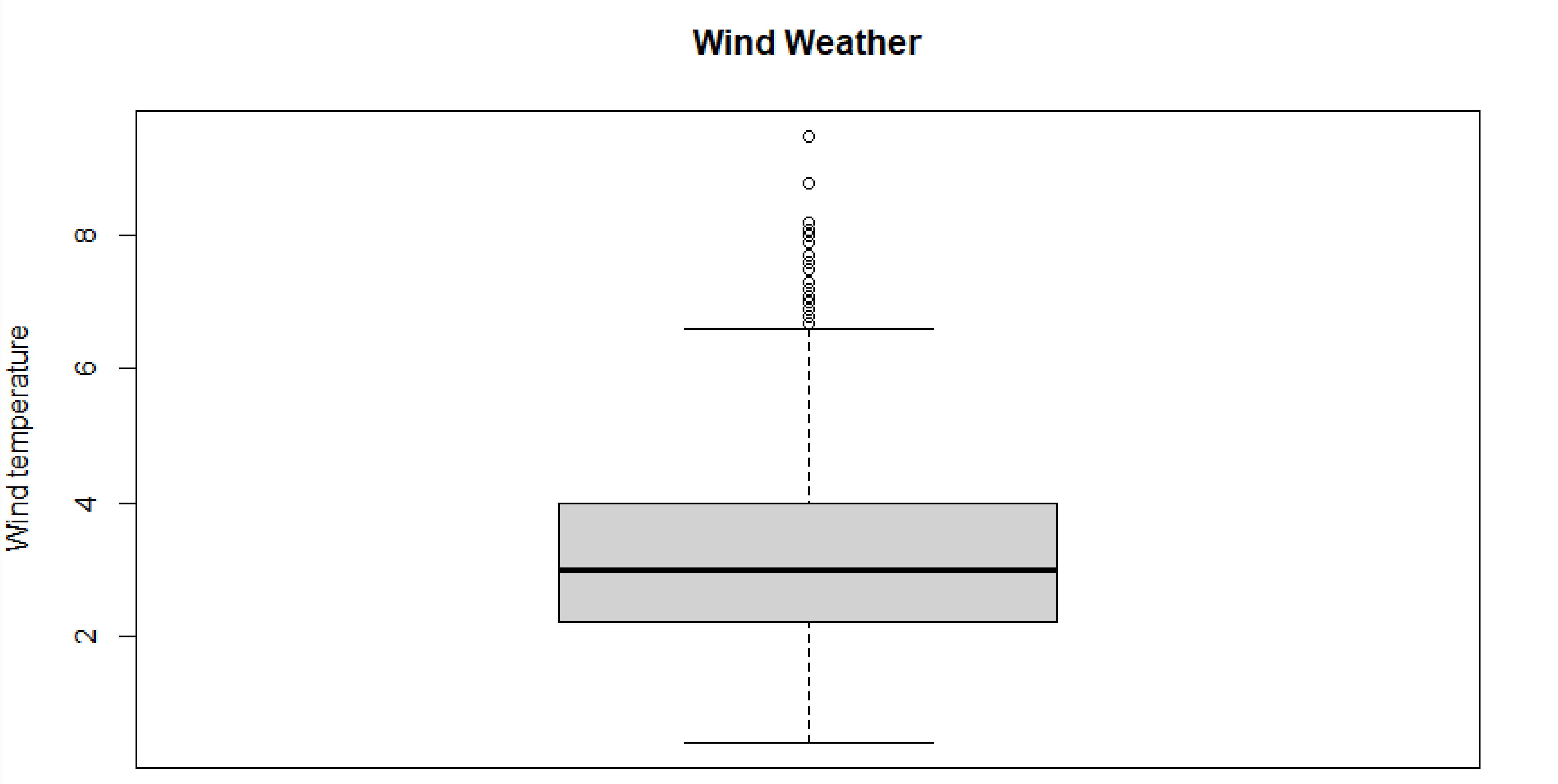
```
boxplot(weather$temp_min,main="Min temperature Weather",ylab="Min temperature")
```



INTERESTING PLOTS :]

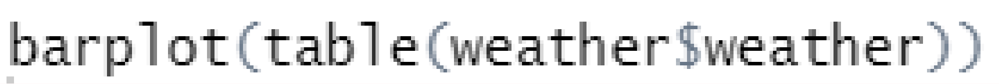


```
boxplot(weather$wind,main="Wind Weather",ylab="Wind temperature")
```

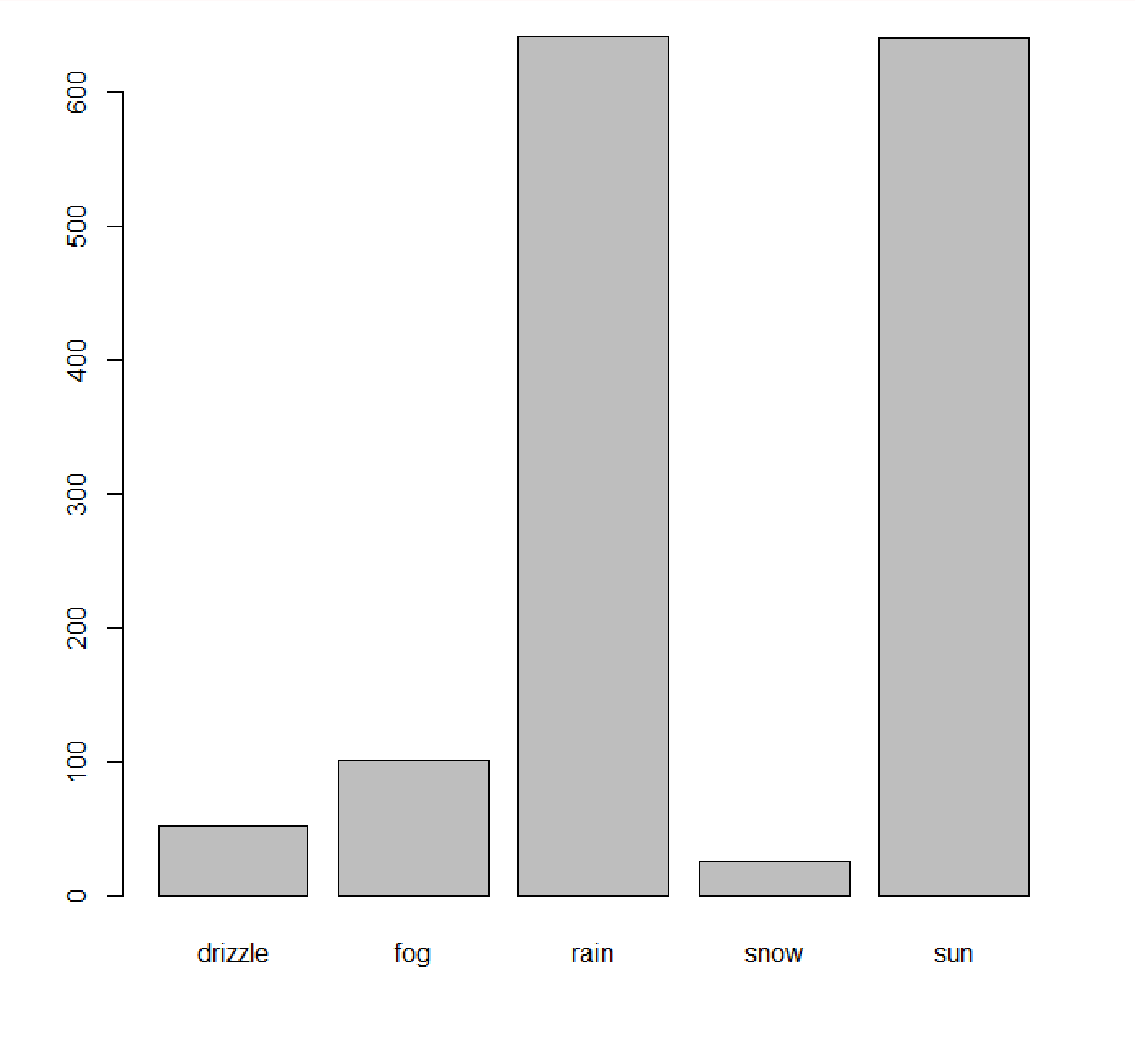


INTERESTING PLOTS :]

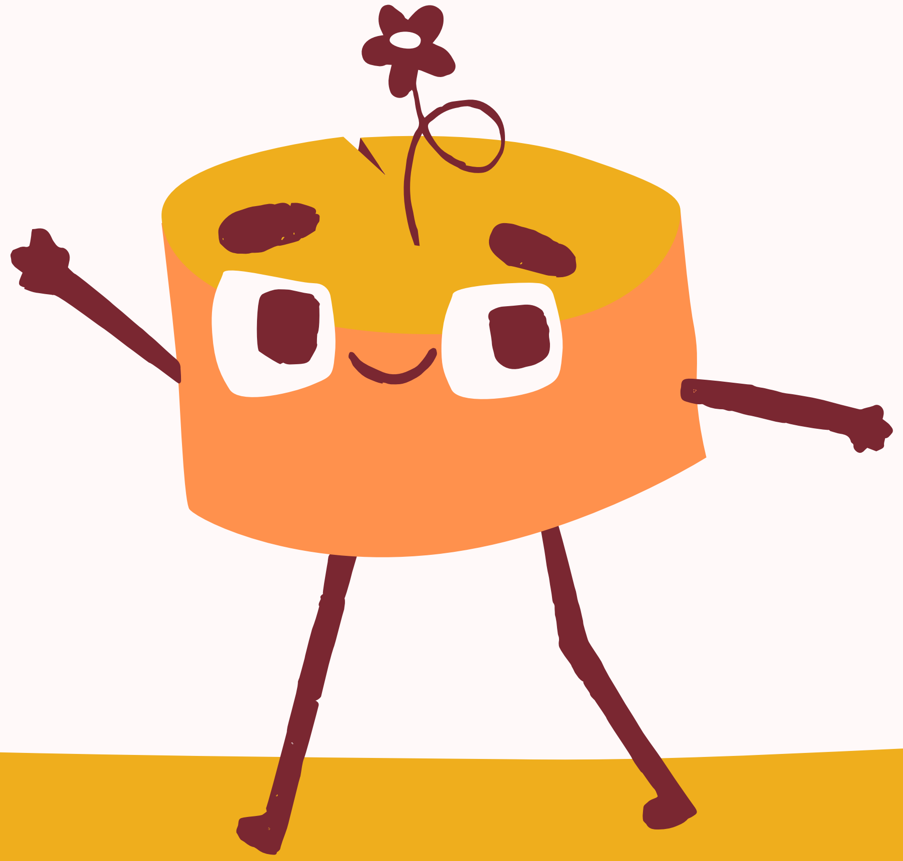
Most common climates in Seattle, according to the dataset.



```
barplot(table(weather$weather))
```



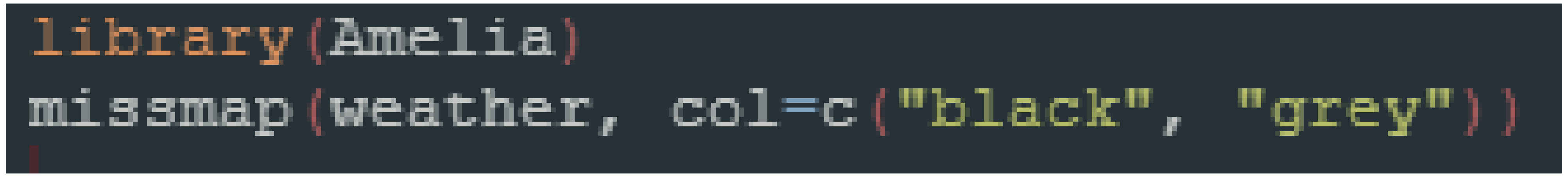
PREPROCESSING



LET'S TRANSFORM THE DATA INTO A SUITABLE DATASET

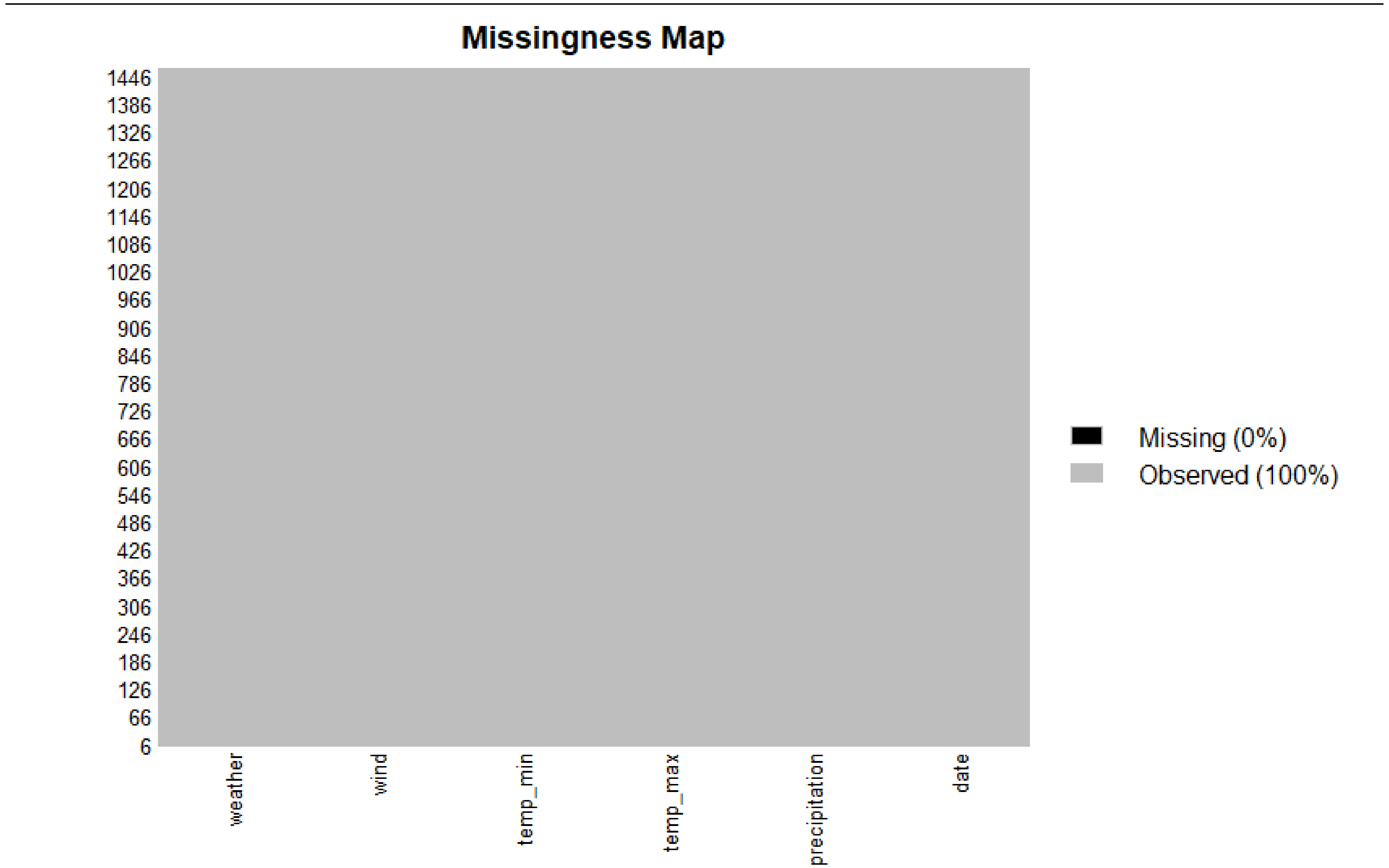
MISSING VALUES

IN R:

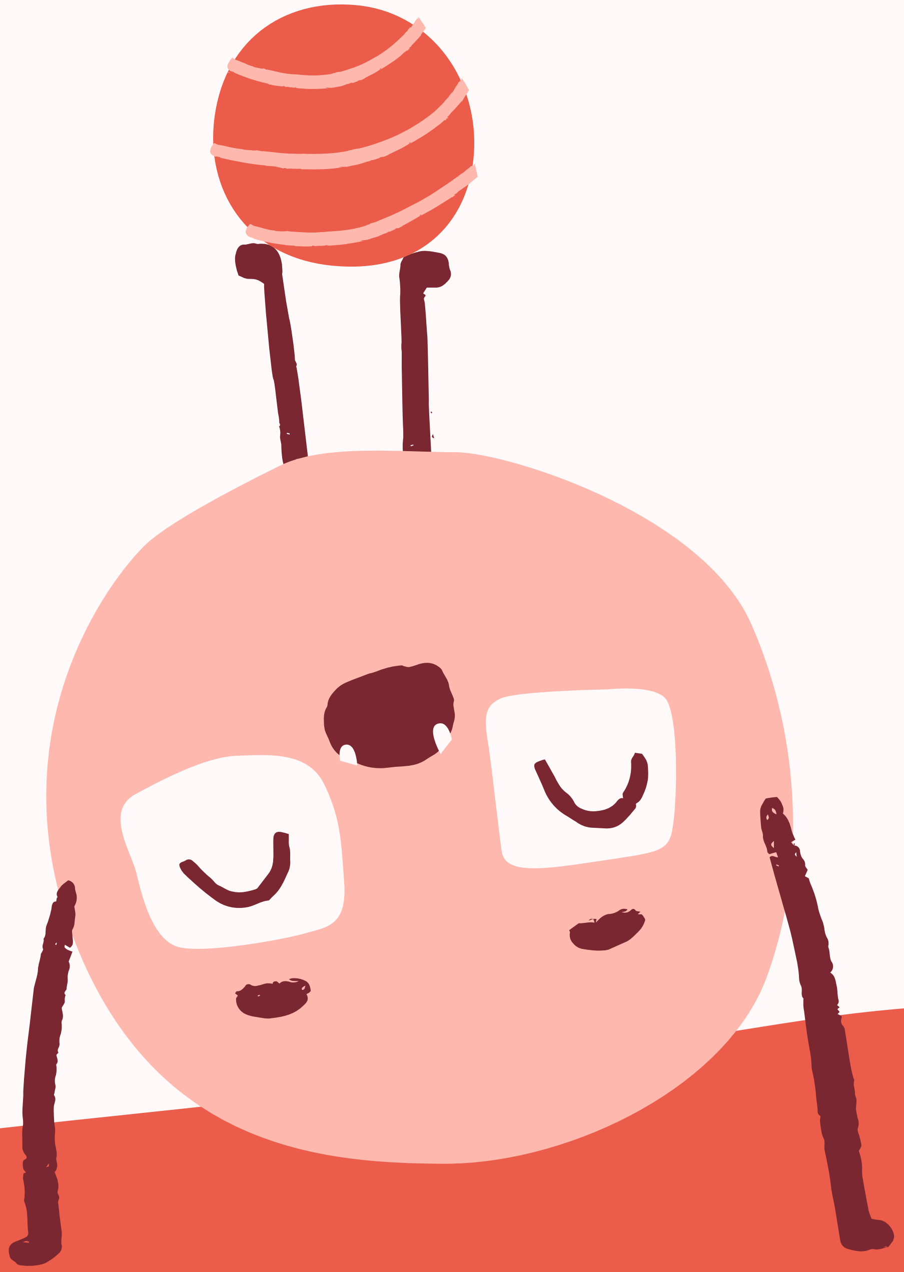


```r
library(Amelia)
missmap(weather, col=c("black", "grey"))
```

OUTPUT:



PROCESSING AND RESULTS

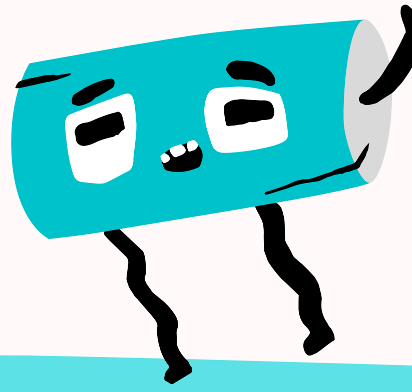


OneR

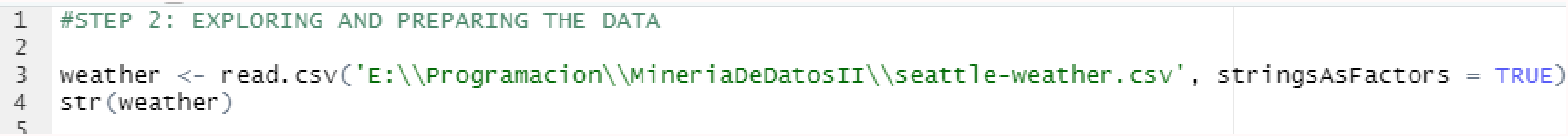
- OneR is a simple and effective classification algorithm, often used in machine learning applications. It can be difficult to improve, due to its simplicity,

- The algorithm creates a rule for each attribute in the training data, then chooses the rule with the smallest error rate, the ”one rule”. .

- To create a rule for each attribute, the most frequent class for each attribute value must be determined. With good data work good results can be obtained

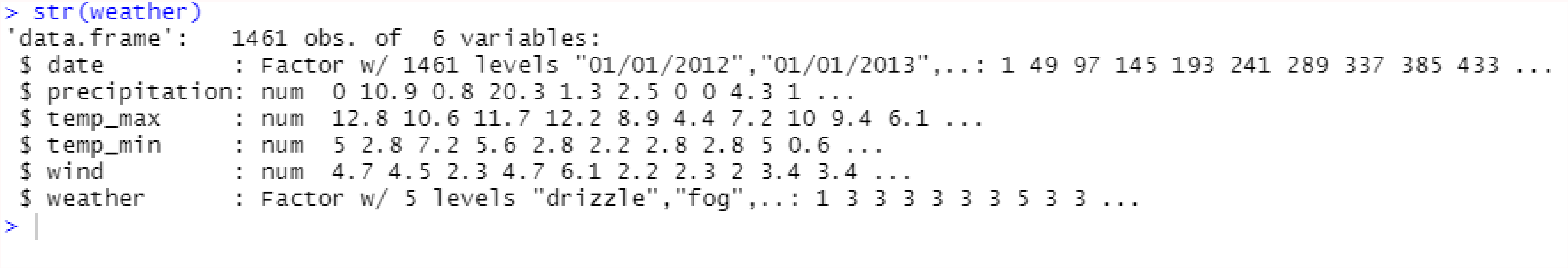


IN R:



```
#STEP 2: EXPLORING AND PREPARING THE DATA

weather <- read.csv('E:\\Programacion\\MineriaDeDatosII\\seattle-weather.csv', stringsAsFactors = TRUE)
str(weather)
```

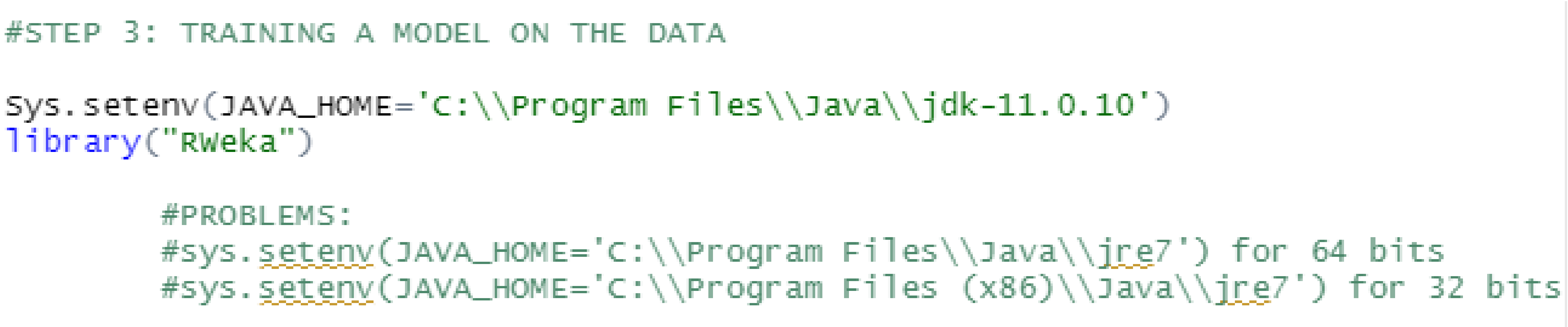
OUTPUT:



```
> str(weather)
'data.frame':	1461 obs. of  6 variables:
 $ date         : Factor w/ 1461 levels "01/01/2012","01/01/2013",..: 1 49 97 145 193 241 289 337 385 433 ...
 $ precipitation: num  0 10.9 0.8 20.3 1.3 2.5 0 0 4.3 1 ...
 $ temp_max     : num  12.8 10.6 11.7 12.2 8.9 4.4 7.2 10 9.4 6.1 ...
 $ temp_min     : num  5 2.8 7.2 5.6 2.8 2.2 2.8 2.8 5 0.6 ...
 $ wind         : num  4.7 4.5 2.3 4.7 6.1 2.2 2.3 2 3.4 3.4 ...
 $ weather      : Factor w/ 5 levels "drizzle","fog",..: 1 3 3 3 3 3 3 5 3 3 ...
> 
```

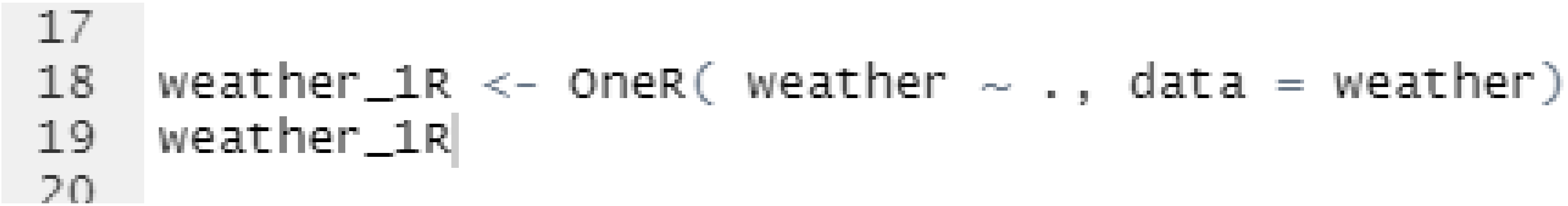
Training a model data

We use the library Rweka



```
#STEP 3: TRAINING A MODEL ON THE DATA

Sys.setenv(JAVA_HOME='C:\\Program Files\\Java\\jdk-11.0.10')
library("RWeka")

        #PROBLEMS:
        #sys.setenv(JAVA_HOME='C:\\Program Files\\Java\\jre7') for 64 bits
        #sys.setenv(JAVA_HOME='C:\\Program Files (x86)\\Java\\jre7') for 32 bits
```

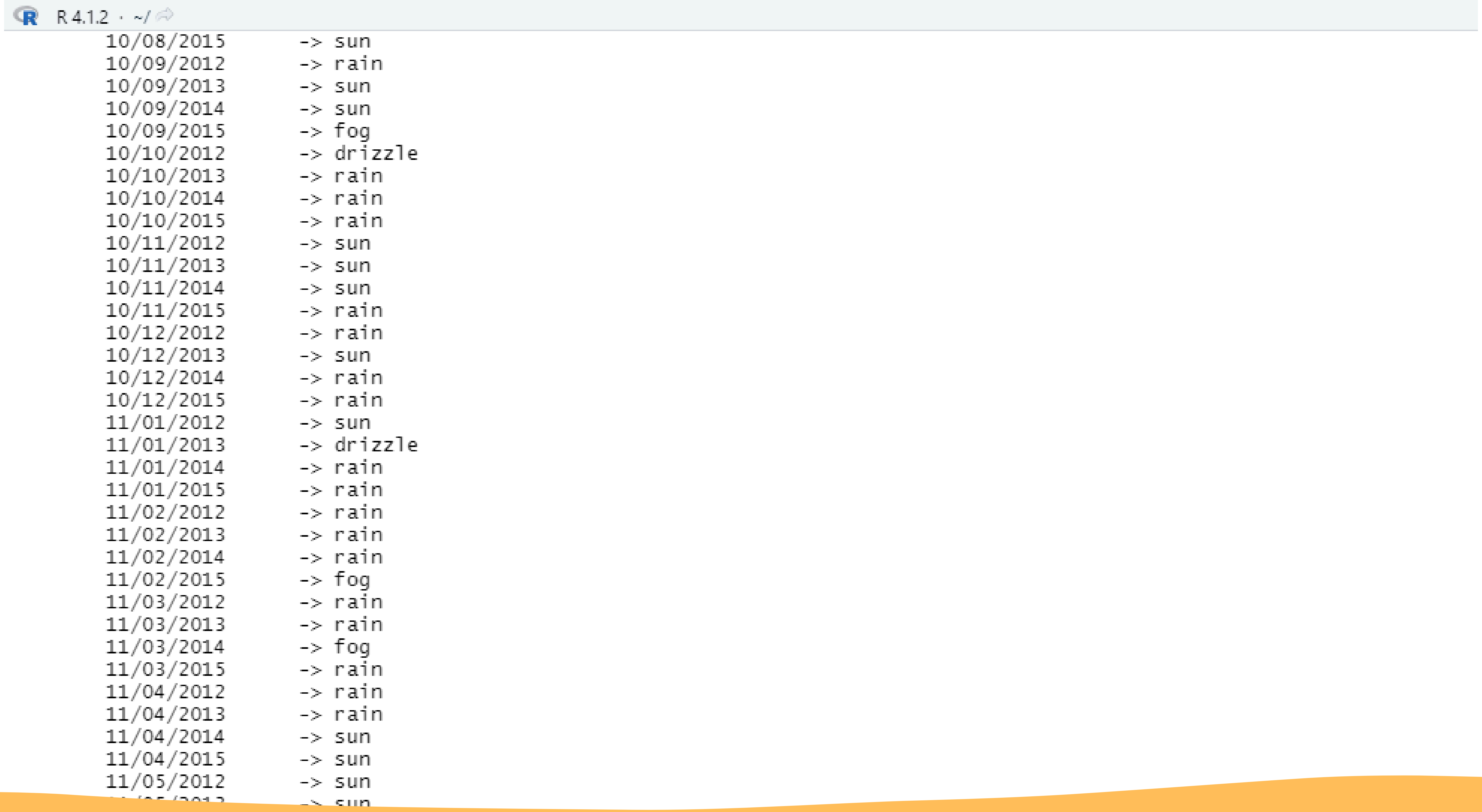
We use the function oneR



```

weather_1R <- OneR( weather ~ ., data = weather)
weather_1R

```

Training a model data

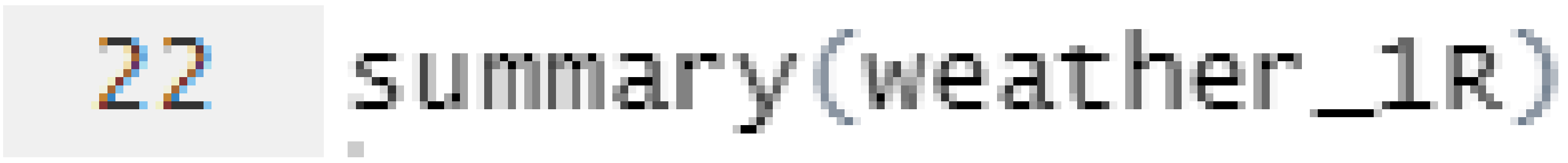
OUTPUT



```
        10/08/2015      -> sun
        10/09/2012      -> rain
        10/09/2013      -> sun
        10/09/2014      -> sun
        10/09/2015      -> fog
        10/10/2012      -> drizzle
        10/10/2013      -> rain
        10/10/2014      -> rain
        10/10/2015      -> rain
        10/11/2012      -> sun
        10/11/2013      -> sun
        10/11/2014      -> sun
        10/11/2015      -> rain
        10/12/2012      -> rain
        10/12/2013      -> sun
        10/12/2014      -> rain
        10/12/2015      -> rain
        11/01/2012      -> sun
        11/01/2013      -> drizzle
        11/01/2014      -> rain
        11/01/2015      -> rain
        11/02/2012      -> rain
        11/02/2013      -> rain
        11/02/2014      -> rain
        11/02/2015      -> fog
        11/03/2012      -> rain
        11/03/2013      -> rain
        11/03/2014      -> fog
        11/03/2015      -> rain
        11/04/2012      -> rain
        11/04/2013      -> rain
        11/04/2014      -> sun
        11/04/2015      -> sun
        11/05/2012      -> sun
        11/05/2013      -> sun
```

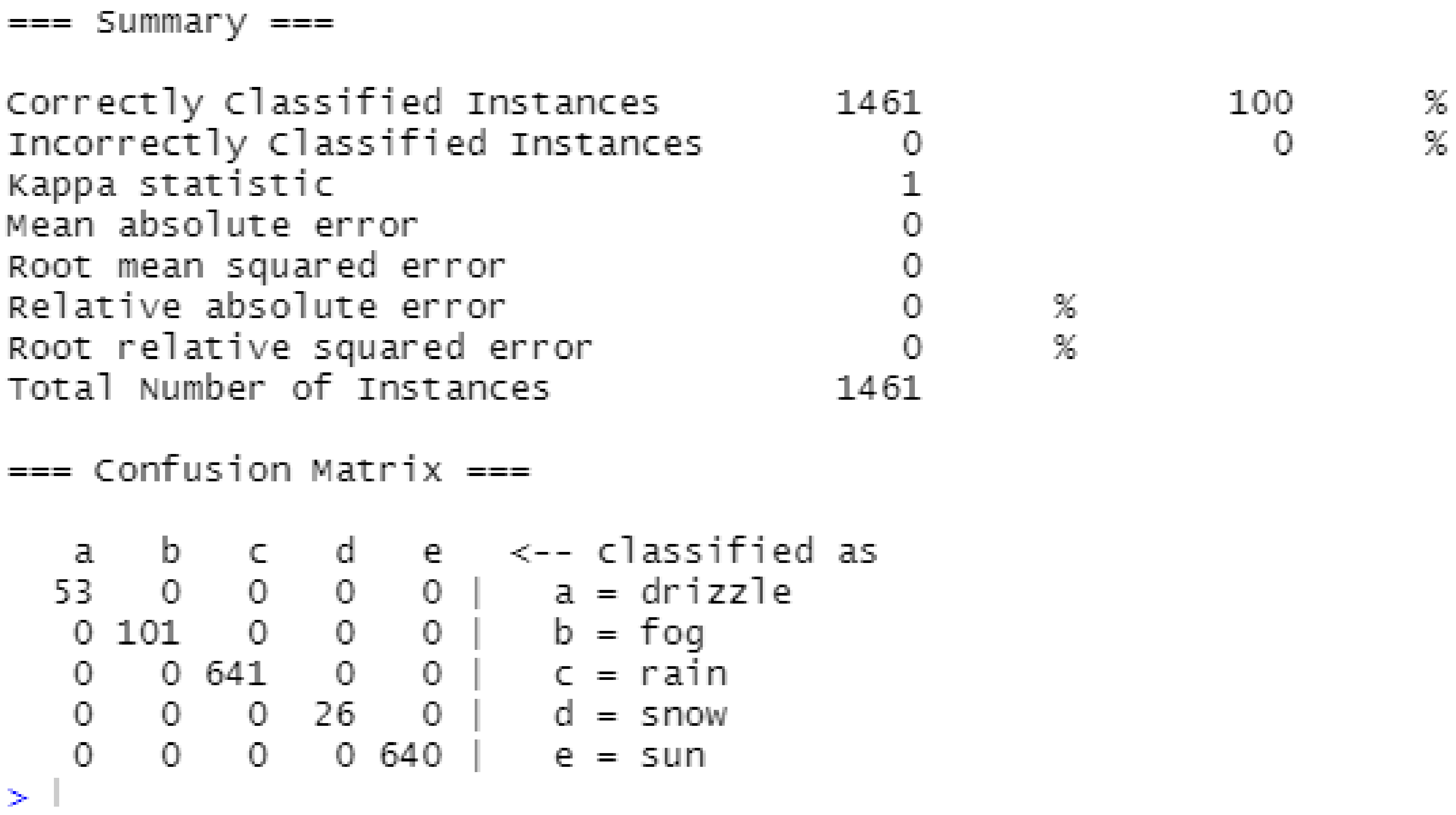
Evaluating the model performance

R fuctions

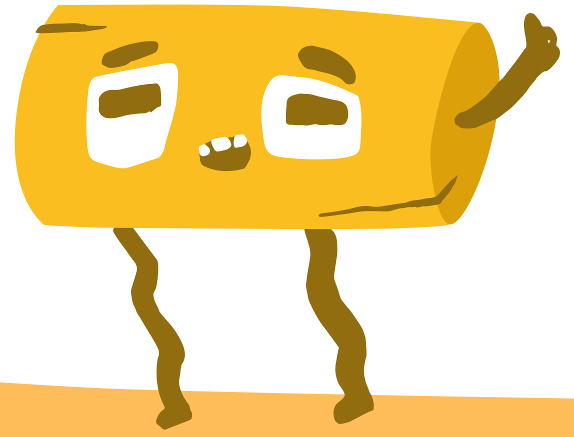


```
summary(weather_1R)
```

Output



```
=== Summary ===

Correctly Classified Instances        1461              100      %
Incorrectly Classified Instances         0                0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances             1461

=== Confusion Matrix ===

   a   b   c   d   e   <-- classified as
  53   0   0   0   0 |   a = drizzle
   0 101   0   0   0 |   b = fog
   0   0 641   0   0 |   c = rain
   0   0   0  26   0 |   d = snow
   0   0   0   0 640 |   e = sun
> |
```

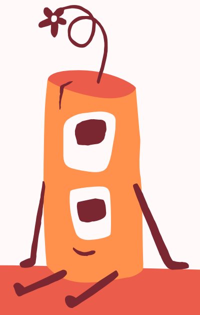


CLASSIFICATION OUTPUTS

Multi-class classification refers to those classification tasks that have more than two class labels.

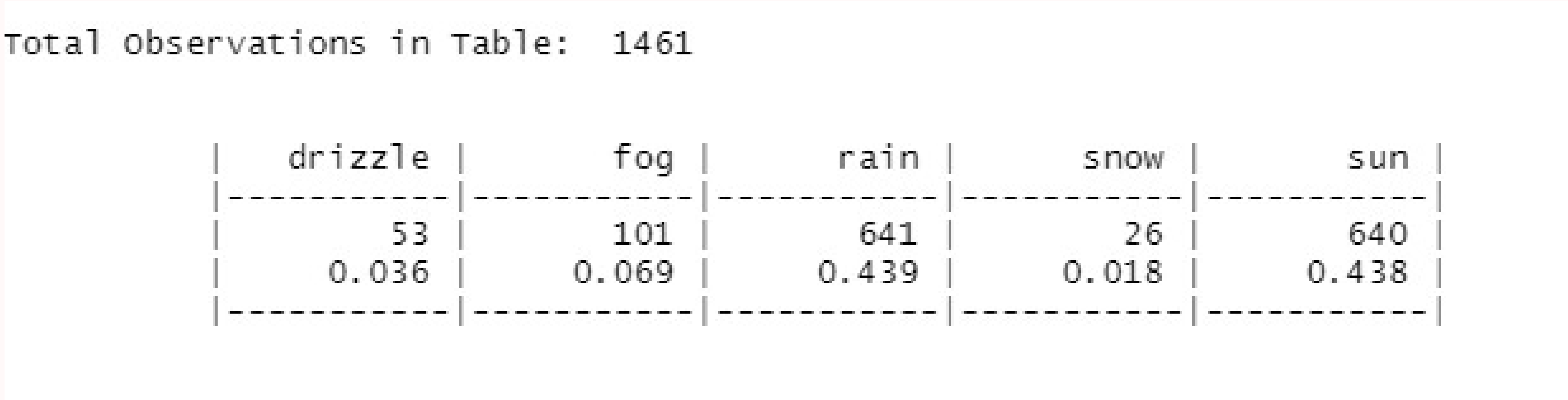
Unlike binary classification, multi-class classification does not have the notion of normal and abnormal outcomes. Instead, examples are classified as belonging to one among a range of known classes.

In this case as we want to predict 5 types of weather, the classification output of our dataset is Multi-Class Classification



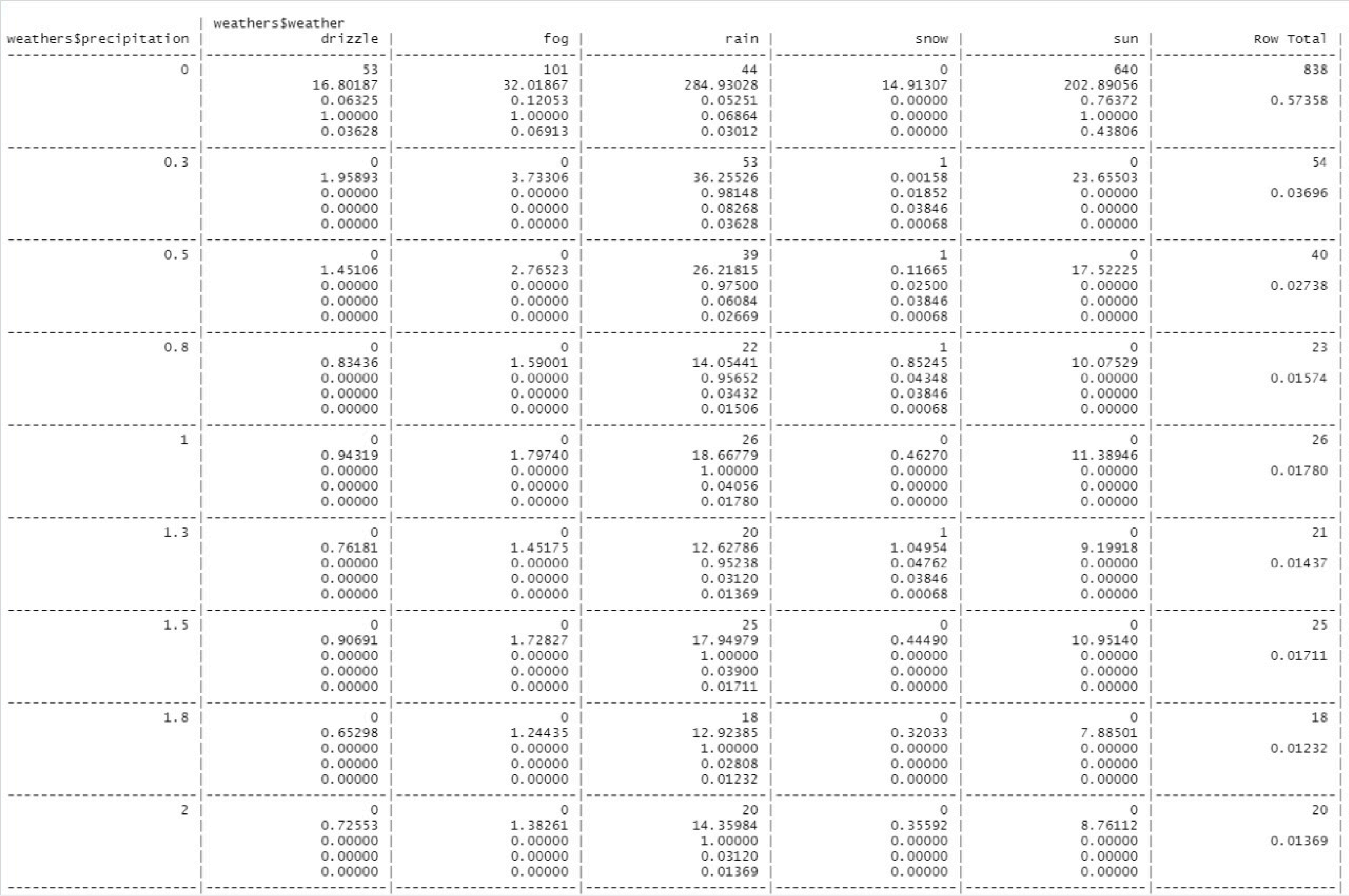
FREQUENCY TABLES, MAE OR ERROR, AND THEIR INTERPRETATION

PERCENTAGE OF CLIMATES IN THE DATASET.



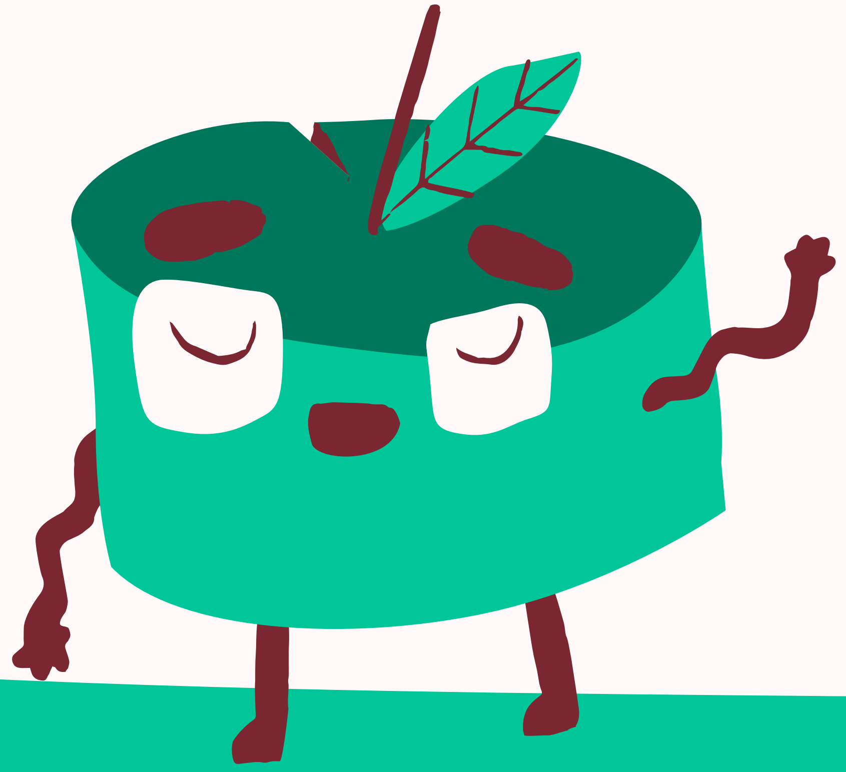
```
Total Observations in Table:  1461


          |   drizzle |       fog |      rain |      snow |       sun |
          |-----------|-----------|-----------|-----------|-----------|
          |        53 |       101 |       641 |        26 |       640 |
          |     0.036 |     0.069 |     0.439 |     0.018 |     0.438 |
          |-----------|-----------|-----------|-----------|-----------|
```

RELATION BETWEEN PRECIPITATION AND WEATHER



```
                       | weathers$weather
weathers$precipitation |               drizzle |                   fog |                  rain |                  snow |                   sun |             Row Total |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                     0 |                    53 |                   101 |                    44 |                     0 |                   640 |                   838 |
                       |              16.80187 |              32.01867 |             284.93028 |              14.91307 |             202.89056 |                       |
                       |               0.06325 |               0.12053 |               0.05251 |               0.00000 |               0.76372 |               0.57358 |
                       |               1.00000 |               1.00000 |               0.06864 |               0.00000 |               1.00000 |                       |
                       |               0.03628 |               0.06913 |               0.03012 |               0.00000 |               0.43806 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                   0.3 |                     0 |                     0 |                    53 |                     1 |                     0 |                    54 |
                       |               1.95893 |               3.73306 |              36.25526 |               0.00158 |              23.65503 |                       |
                       |               0.00000 |               0.00000 |               0.98148 |               0.01852 |               0.00000 |               0.03696 |
                       |               0.00000 |               0.00000 |               0.08268 |               0.03846 |               0.00000 |                       |
                       |               0.00000 |               0.00000 |               0.03628 |               0.00068 |               0.00000 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                   0.5 |                     0 |                     0 |                    39 |                     1 |                     0 |                    40 |
                       |               1.45106 |               2.76523 |              26.21815 |               0.11665 |              17.52225 |                       |
                       |               0.00000 |               0.00000 |               0.97500 |               0.02500 |               0.00000 |               0.02738 |
                       |               0.00000 |               0.00000 |               0.06084 |               0.03846 |               0.00000 |                       |
                       |               0.00000 |               0.00000 |               0.02669 |               0.00068 |               0.00000 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                   0.8 |                     0 |                     0 |                    22 |                     1 |                     0 |                    23 |
                       |               0.83436 |               1.59001 |              14.05441 |               0.85245 |              10.07529 |                       |
                       |               0.00000 |               0.00000 |               0.95652 |               0.04348 |               0.00000 |               0.01574 |
                       |               0.00000 |               0.00000 |               0.03432 |               0.03846 |               0.00000 |                       |
                       |               0.00000 |               0.00000 |               0.01506 |               0.00068 |               0.00000 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                     1 |                     0 |                     0 |                    26 |                     0 |                     0 |                    26 |
                       |               0.94319 |               1.79740 |              18.66779 |               0.46270 |              11.38946 |                       |
                       |               0.00000 |               0.00000 |               1.00000 |               0.00000 |               0.00000 |               0.01780 |
                       |               0.00000 |               0.00000 |               0.04056 |               0.00000 |               0.00000 |                       |
                       |               0.00000 |               0.00000 |               0.01780 |               0.00000 |               0.00000 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                   1.3 |                     0 |                     0 |                    20 |                     1 |                     0 |                    21 |
                       |               0.76181 |               1.45175 |              12.62786 |               1.04954 |               9.19918 |                       |
                       |               0.00000 |               0.00000 |               0.95238 |               0.04762 |               0.00000 |               0.01437 |
                       |               0.00000 |               0.00000 |               0.03120 |               0.03846 |               0.00000 |                       |
                       |               0.00000 |               0.00000 |               0.01369 |               0.00068 |               0.00000 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                   1.5 |                     0 |                     0 |                    25 |                     0 |                     0 |                    25 |
                       |               0.90691 |               1.72827 |              17.94979 |               0.44490 |              10.95140 |                       |
                       |               0.00000 |               0.00000 |               1.00000 |               0.00000 |               0.00000 |               0.01711 |
                       |               0.00000 |               0.00000 |               0.03900 |               0.00000 |               0.00000 |                       |
                       |               0.00000 |               0.00000 |               0.01711 |               0.00000 |               0.00000 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                   1.8 |                     0 |                     0 |                    18 |                     0 |                     0 |                    18 |
                       |               0.65298 |               1.24435 |              12.92385 |               0.32033 |               7.88501 |                       |
                       |               0.00000 |               0.00000 |               1.00000 |               0.00000 |               0.00000 |               0.01232 |
                       |               0.00000 |               0.00000 |               0.02808 |               0.00000 |               0.00000 |                       |
                       |               0.00000 |               0.00000 |               0.01232 |               0.00000 |               0.00000 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
                     2 |                     0 |                     0 |                    20 |                     0 |                     0 |                    20 |
                       |               0.72553 |               1.38261 |              14.35984 |               0.35592 |               8.76112 |                       |
                       |               0.00000 |               0.00000 |               1.00000 |               0.00000 |               0.00000 |               0.01369 |
                       |               0.00000 |               0.00000 |               0.03120 |               0.00000 |               0.00000 |                       |
                       |               0.00000 |               0.00000 |               0.01369 |               0.00000 |               0.00000 |                       |
-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
```

CONCLUSIONS AND LIMITATIONS



Does the study generalize to other domains?

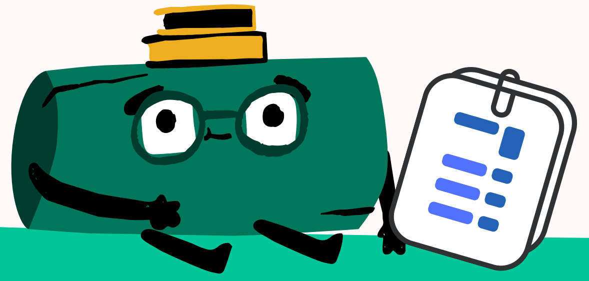
Yes, this is only a basic example but we can classify and predict more things like dog species, if a patient it's healthy of not, etc.

Limitations

The algorithm is overly simple we might not be able to do a lot of things because only uses a single feature.

Advantages

So easy to undestand to people who doesn't have any knowledge in our area, might be used for benchmarking other algorithms and its performance it's really good.



BIBLIOGRAPHY

https://www.kaggle.com/ananthr1/weather-prediction
https://christophm.github.io/interpretable-ml-book/rules.html
http://rasbt.github.io/mlxtend/user_guide/classifier/OneRClassifier/
https://www.youtube.com/watch?v=bAqU3-1FsPA&ab_channel=DaveSullivan
https://machinelearningmastery.com/types-of-classification-in-machine-learning/