Weather Prediction

2022-12-12

```
#Read the data
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 4.2.2
```

```
## --------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## --------------------------------------------------------------------------------
```

```
##
## Attaching package: 'plyr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
data<-read.csv("weather.csv")
head(data)
```

```
##   MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir WindGustSpeed
## 1     8.0    24.3      0.0         3.4      6.3          NW            30
## 2    14.0    26.9      3.6         4.4      9.7         ENE            39
```

```
## 3    13.7    23.4      3.6        5.8       3.3        NW            85
## 4    13.3    15.5     39.8        7.2       9.1        NW            54
## 5     7.6    16.1      2.8        5.6      10.6       SSE            50
## 6     6.2    16.9      0.0        5.8       8.2        SE            44
##    WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm
## 1         SW         NW            6           20          68          29
## 2          E          W            4           17          80          36
## 3          N        NNE            6            6          82          69
## 4        WNW          W           30           24          62          56
## 5        SSE        ESE           20           28          68          49
## 6         SE          E           20           24          70          57
##    Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday
## 1      1019.7      1015.0        7        7    14.4    23.6        No
## 2      1012.4      1008.4        5        3    17.5    25.7       Yes
## 3      1009.5      1007.2        8        7    15.4    20.2       Yes
## 4      1005.5      1007.0        2        7    13.5    14.1       Yes
## 5      1018.3      1018.5        7        7    11.1    15.4       Yes
## 6      1023.8      1021.7        7        5    10.9    14.8        No
##    RainTomorrow
## 1          Yes
## 2          Yes
## 3          Yes
## 4          Yes
## 5           No
## 6           No
```

summary(data)

```
##     MinTemp          MaxTemp         Rainfall        Evaporation
## Min.   :-5.300   Min.   : 7.60   Min.   : 0.000   Min.   : 0.200
## 1st Qu.: 2.300   1st Qu.:15.03   1st Qu.: 0.000   1st Qu.: 2.200
## Median : 7.450   Median :19.65   Median : 0.000   Median : 4.200
## Mean   : 7.266   Mean   :20.55   Mean   : 1.428   Mean   : 4.522
## 3rd Qu.:12.500   3rd Qu.:25.50   3rd Qu.: 0.200   3rd Qu.: 6.400
## Max.   :20.900   Max.   :35.80   Max.   :39.800   Max.   :13.800
##
##    Sunshine       WindGustDir        WindGustSpeed    WindDir9am
## Min.   : 0.000   Length:366        Min.   :13.00   Length:366
## 1st Qu.: 5.950   Class :character  1st Qu.:31.00   Class :character
## Median : 8.600   Mode  :character  Median :39.00   Mode  :character
## Mean   : 7.909                     Mean   :39.84
## 3rd Qu.:10.500                     3rd Qu.:46.00
## Max.   :13.600                     Max.   :98.00
## NA's   :3                          NA's   :2
##   WindDir3pm        WindSpeed9am     WindSpeed3pm     Humidity9am
## Length:366        Min.   : 0.000   Min.   : 0.00   Min.   :36.00
## Class :character  1st Qu.: 6.000   1st Qu.:11.00   1st Qu.:64.00
## Mode  :character  Median : 7.000   Median :17.00   Median :72.00
##                   Mean   : 9.652   Mean   :17.99   Mean   :72.04
##                   3rd Qu.:13.000   3rd Qu.:24.00   3rd Qu.:81.00
##                   Max.   :41.000   Max.   :52.00   Max.   :99.00
##                   NA's   :7
##   Humidity3pm      Pressure9am      Pressure3pm       Cloud9am
## Min.   :13.00   Min.   : 996.5   Min.   : 996.8   Min.   :0.000
```

```
##    1st Qu.:32.25     1st Qu.:1015.4     1st Qu.:1012.8     1st Qu.:1.000
##    Median :43.00     Median :1020.1     Median :1017.4     Median :3.500
##    Mean   :44.52     Mean   :1019.7     Mean   :1016.8     Mean   :3.891
##    3rd Qu.:55.00     3rd Qu.:1024.5     3rd Qu.:1021.5     3rd Qu.:7.000
##    Max.   :96.00     Max.   :1035.7     Max.   :1033.2     Max.   :8.000
##
##       Cloud3pm          Temp9am            Temp3pm          RainToday
##    Min.   :0.000    Min.   : 0.100    Min.   : 5.10    Length:366
##    1st Qu.:1.000    1st Qu.: 7.625    1st Qu.:14.15    Class :character
##    Median :4.000    Median :12.550    Median :18.55    Mode  :character
##    Mean   :4.025    Mean   :12.358    Mean   :19.23
##    3rd Qu.:7.000    3rd Qu.:17.000    3rd Qu.:24.00
##    Max.   :8.000    Max.   :24.700    Max.   :34.50
##
##    RainTomorrow
##    Length:366
##    Class :character
##    Mode  :character
##
##
##
##
```

Changing boolean values

```
#data$RainToday <- revalue(data$RainToday, c("Yes"=1))
#data$RainTomorrow <- revalue(data$RainTomorrow, c("Yes"=1))
#data$RainToday <- revalue(data$RainToday, c("No"=0))
#data$RainTomorrow <- revalue(data$RainTomorrow, c("No"=0))
data$RainToday=as.factor(data$RainToday)
data$RainTomorrow=as.factor(data$RainTomorrow)
head(data)
```

```
##   MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir WindGustSpeed
## 1     8.0    24.3      0.0         3.4      6.3          NW            30
## 2    14.0    26.9      3.6         4.4      9.7         ENE            39
## 3    13.7    23.4      3.6         5.8      3.3          NW            85
## 4    13.3    15.5     39.8         7.2      9.1          NW            54
## 5     7.6    16.1      2.8         5.6     10.6         SSE            50
## 6     6.2    16.9      0.0         5.8      8.2          SE            44
##   WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm
## 1         SW         NW            6           20          68          29
## 2          E          W            4           17          80          36
## 3          N        NNE            6            6          82          69
## 4        WNW          W           30           24          62          56
## 5        SSE        ESE           20           28          68          49
## 6         SE          E           20           24          70          57
##   Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday
## 1      1019.7      1015.0        7        7    14.4    23.6        No
## 2      1012.4      1008.4        5        3    17.5    25.7       Yes
## 3      1009.5      1007.2        8        7    15.4    20.2       Yes
## 4      1005.5      1007.0        2        7    13.5    14.1       Yes
## 5      1018.3      1018.5        7        7    11.1    15.4       Yes
```

```
## 6      1023.8      1021.7      7      5    10.9    14.8          No
##    RainTomorrow
## 1          Yes
## 2          Yes
## 3          Yes
## 4          Yes
## 5           No
## 6           No
```

Check for any missing values and remove them

```
sum(is.na(data))
```

```
## [1] 47
```

```
data=na.omit(data)
head(data)
```

```
##    MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir WindGustSpeed
## 1     8.0    24.3      0.0         3.4      6.3          NW            30
## 2    14.0    26.9      3.6         4.4      9.7         ENE            39
## 3    13.7    23.4      3.6         5.8      3.3          NW            85
## 4    13.3    15.5     39.8         7.2      9.1          NW            54
## 5     7.6    16.1      2.8         5.6     10.6         SSE            50
## 6     6.2    16.9      0.0         5.8      8.2          SE            44
##    WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm
## 1          SW         NW            6           20          68          29
## 2           E          W            4           17          80          36
## 3           N        NNE            6            6          82          69
## 4         WNW          W           30           24          62          56
## 5         SSE        ESE           20           28          68          49
## 6          SE          E           20           24          70          57
##    Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday
## 1       1019.7      1015.0        7        7    14.4    23.6        No
## 2       1012.4      1008.4        5        3    17.5    25.7       Yes
## 3       1009.5      1007.2        8        7    15.4    20.2       Yes
## 4       1005.5      1007.0        2        7    13.5    14.1       Yes
## 5       1018.3      1018.5        7        7    11.1    15.4       Yes
## 6       1023.8      1021.7        7        5    10.9    14.8        No
##    RainTomorrow
## 1          Yes
## 2          Yes
## 3          Yes
## 4          Yes
## 5           No
## 6           No
```

As the wind pressure increases the humidity also shows a rise
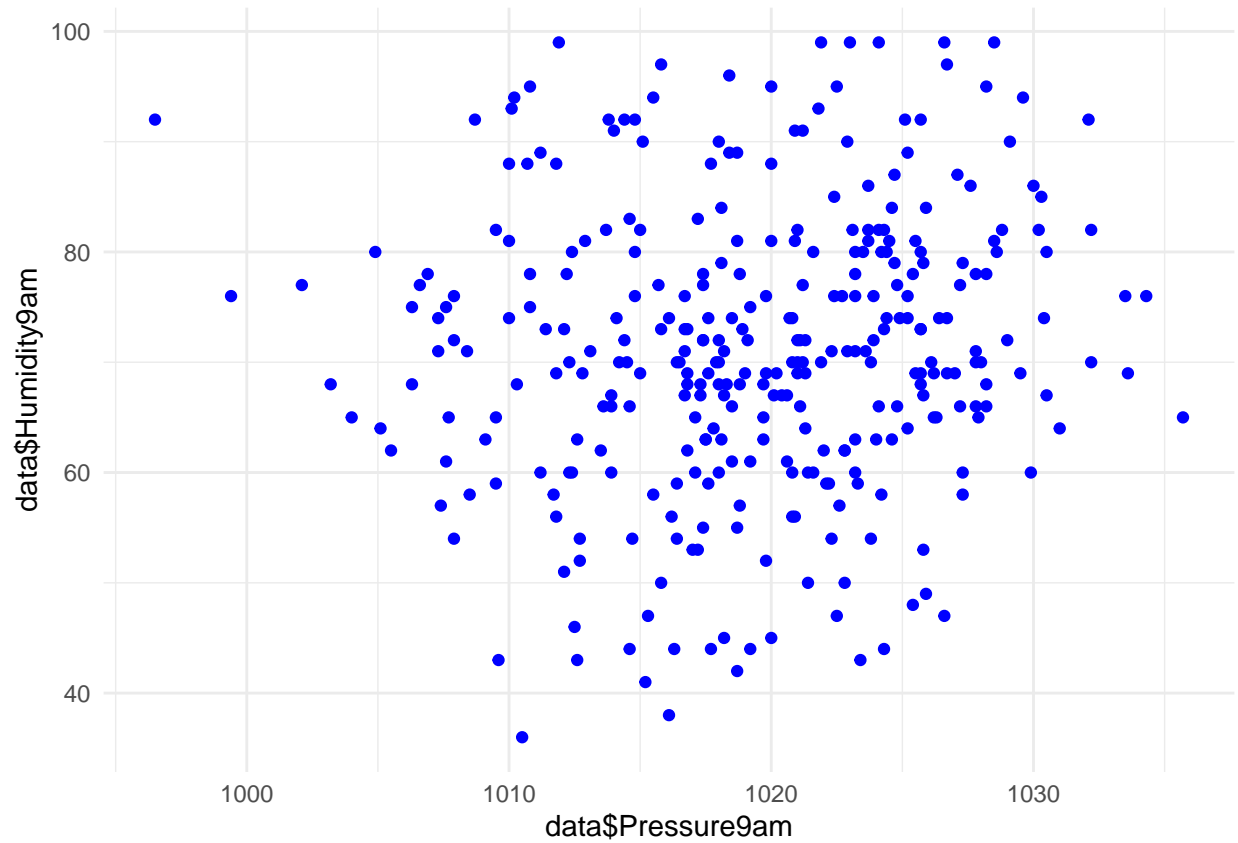
```
library(ggplot2)
```

```
ggplot(data) +
```

```
  aes(x = data$Pressure9am,y=data$Humidity9am) +
  geom_point(colour = "blue") +
  theme_minimal()
```
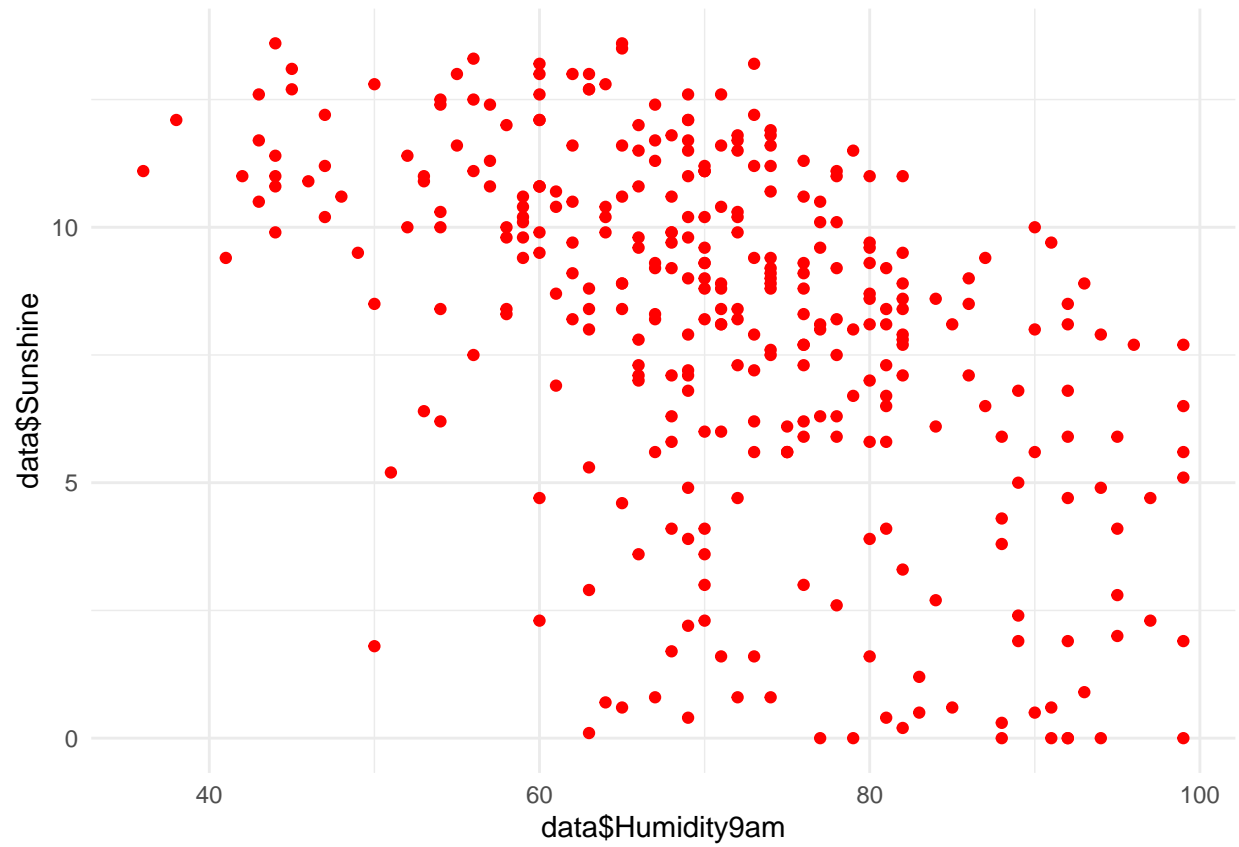

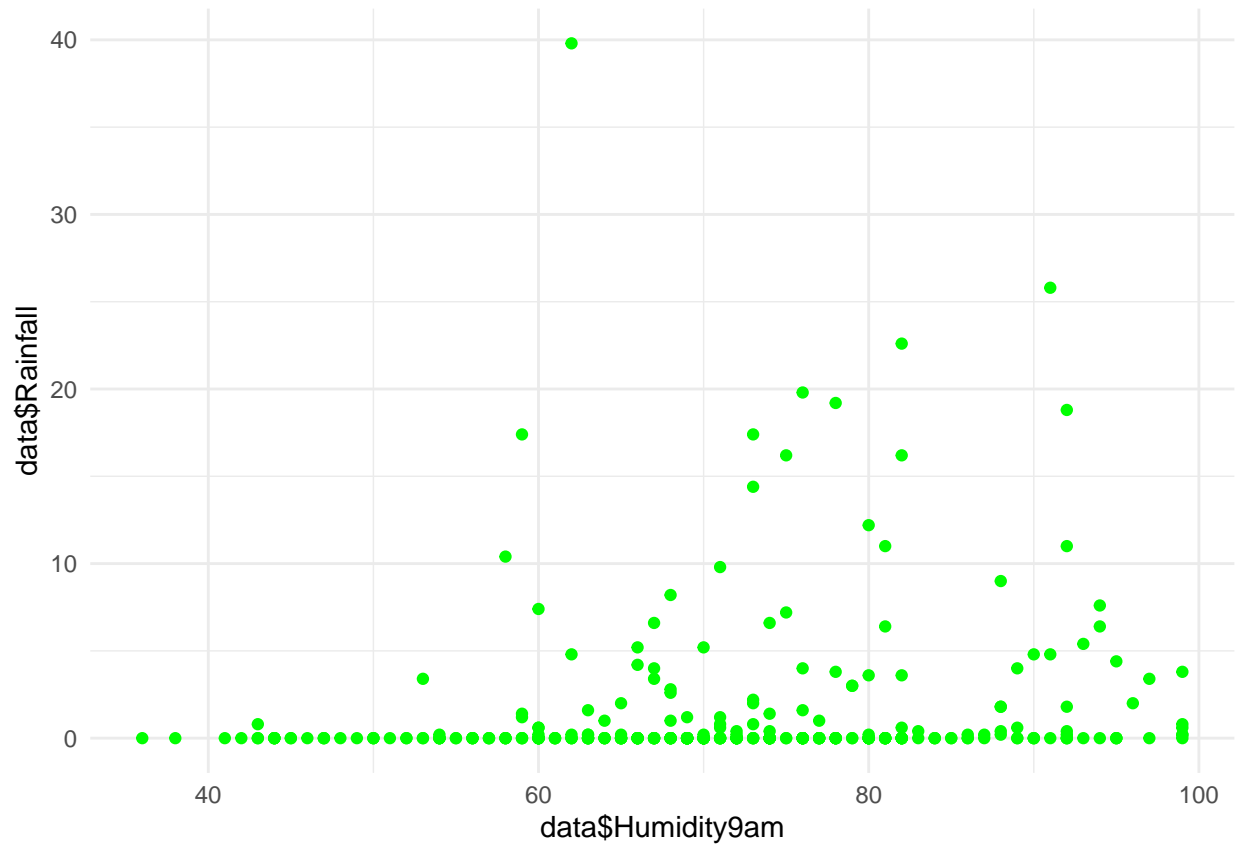The increase in humidity percentage affects the sunshine amount positively.

```
ggplot(data) +
  aes(x = data$Humidity9am,y=data$Sunshine) +
  geom_point(colour = "red") +
  theme_minimal()
```

```
ggplot(data) +
  aes(x = data$Humidity9am,y=data$Rainfall) +
  geom_point(colour = "green") +
  theme_minimal()
```

SPLITTING THE DATA INTO TRAIN AND TEST SETS

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.2.2
```

```
## Loading required package: lattice
```

```
index <- sample(c(TRUE, FALSE), nrow(data), replace=TRUE, prob=c(0.80,0.20))
trainData<-data[index,]
testData<-data[!index,]
dim(trainData)
```

```
## [1] 258  21
```
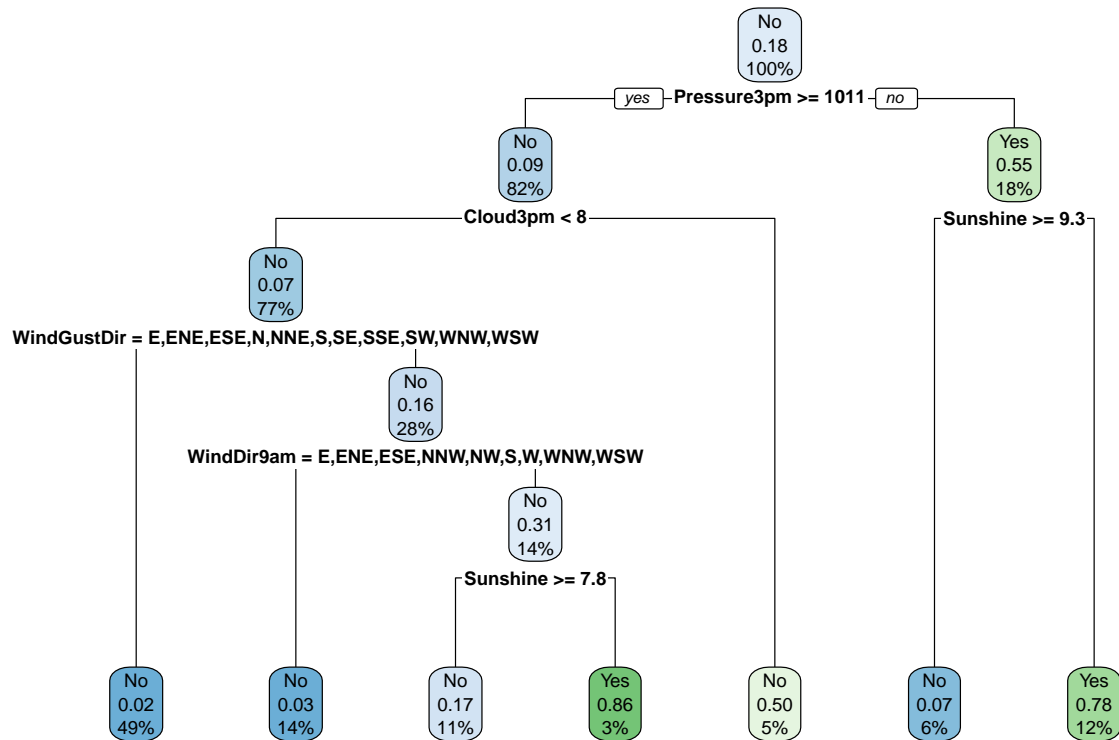
```
dim(testData)
```

```
## [1] 70 21
```

BUILDING THE DECISION TREE MODEL

```
library(rpart)
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 4.2.2
```

```
model=rpart(RainTomorrow~.,trainData)
rpart.plot(model)
```



```
Prediction <- predict(model, testData, type= "class")
confusionMatrix(Prediction, testData$RainTomorrow, positive= "Yes")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction No Yes
##        No  50   6
##        Yes  6   8
##
##               Accuracy : 0.8286
##                 95% CI : (0.7197, 0.9082)
##    No Information Rate : 0.8
##    P-Value [Acc > NIR] : 0.336
##
##                  Kappa : 0.4643
##
##  Mcnemar's Test P-Value : 1.000
##
##            Sensitivity : 0.5714
```

```
##             Specificity : 0.8929
##          Pos Pred Value : 0.5714
##          Neg Pred Value : 0.8929
##              Prevalence : 0.2000
##          Detection Rate : 0.1143
##    Detection Prevalence : 0.2000
##       Balanced Accuracy : 0.7321
##
##         'Positive' Class : Yes
##
```

The Accuracy of the model is 86.57%

```
summary(model)
```

```
## Call:
## rpart(formula = RainTomorrow ~ ., data = trainData)
##   n= 258
##
##            CP nsplit rel error    xerror       xstd
## 1 0.19565217      0 1.0000000 1.0000000 0.1336531
## 2 0.02717391      2 0.6086957 0.6304348 0.1102933
## 3 0.01000000      6 0.5000000 0.9565217 0.1313305
##
## Variable importance
##      Sunshine   Pressure3pm   Pressure9am      Cloud3pm  Humidity3pm
##            17            16            12            12           10
##       Temp3pm       MaxTemp   Humidity9am    WindDir3pm   WindDir9am
##             5             5             4             4            3
##      Cloud9am      Rainfall   WindGustDir WindGustSpeed      Temp9am
##             3             2             2             2            2
##       MinTemp
##             1
##
## Node number 1: 258 observations,    complexity param=0.1956522
##   predicted class=No   expected loss=0.1782946  P(node) =1
##     class counts:   212    46
##    probabilities: 0.822 0.178
##   left son=2 (211 obs) right son=3 (47 obs)
##   Primary splits:
##       Pressure3pm   < 1010.65 to the right, improve=16.154330, (0 missing)
##       Sunshine      < 6.45    to the right, improve=11.077490, (0 missing)
##       Cloud3pm      < 6.5     to the left,  improve=11.022000, (0 missing)
##       Pressure9am   < 1013    to the right, improve= 9.843438, (0 missing)
##       WindGustSpeed < 66      to the left,  improve= 8.505564, (0 missing)
##   Surrogate splits:
##       Pressure9am   < 1013.3  to the right, agree=0.953, adj=0.745, (0 split)
##       Temp9am       < 22.75   to the left,  agree=0.837, adj=0.106, (0 split)
##       WindGustSpeed < 73      to the left,  agree=0.833, adj=0.085, (0 split)
##       Humidity3pm   < 70.5    to the left,  agree=0.829, adj=0.064, (0 split)
##       MinTemp       < 18.95   to the left,  agree=0.826, adj=0.043, (0 split)
##
## Node number 2: 211 observations,    complexity param=0.02717391
```

```
##    predicted class=No    expected loss=0.09478673  P(node) =0.8178295
##      class counts:    191    20
##     probabilities: 0.905 0.095
##    left son=4 (199 obs) right son=5 (12 obs)
##    Primary splits:
##        Cloud3pm    < 7.5      to the left,  improve=4.178380, (0 missing)
##        Sunshine    < 2.1      to the right, improve=3.404271, (0 missing)
##        MinTemp     < 16.6     to the left,  improve=3.289763, (0 missing)
##        WindGustDir splits as  LLRLRLRRLLRRLRLL, improve=2.760603, (0 missing)
##        Pressure3pm < 1016.05  to the right, improve=2.576178, (0 missing)
##    Surrogate splits:
##        Sunshine    < 0.3      to the right, agree=0.962, adj=0.333, (0 split)
##        Humidity3pm < 71.5     to the left,  agree=0.957, adj=0.250, (0 split)
##        Rainfall    < 20       to the left,  agree=0.953, adj=0.167, (0 split)
##
## Node number 3: 47 observations,    complexity param=0.1956522
##    predicted class=Yes  expected loss=0.4468085  P(node) =0.1821705
##      class counts:     21    26
##     probabilities: 0.447 0.553
##    left son=6 (15 obs) right son=7 (32 obs)
##    Primary splits:
##        Sunshine    < 9.3      to the right, improve=10.429880, (0 missing)
##        WindGustDir splits as  RLRRR-RRR-RRRRLL, improve= 6.812990, (0 missing)
##        Humidity3pm < 42       to the left,  improve= 6.506770, (0 missing)
##        Temp9am     < 18.35    to the right, improve= 5.383062, (0 missing)
##        WindDir9am  splits as  LRLR-LLLRRRR-LLL, improve= 5.324268, (0 missing)
##    Surrogate splits:
##        Humidity3pm < 38.5     to the left,  agree=0.851, adj=0.533, (0 split)
##        MaxTemp     < 30.05    to the right, agree=0.830, adj=0.467, (0 split)
##        Cloud3pm    < 1.5      to the left,  agree=0.830, adj=0.467, (0 split)
##        Temp3pm     < 28.4     to the right, agree=0.830, adj=0.467, (0 split)
##        Humidity9am < 60.5     to the left,  agree=0.809, adj=0.400, (0 split)
##
## Node number 4: 199 observations,    complexity param=0.02717391
##    predicted class=No    expected loss=0.07035176  P(node) =0.7713178
##      class counts:    185    14
##     probabilities: 0.930 0.070
##    left son=8 (126 obs) right son=9 (73 obs)
##    Primary splits:
##        WindGustDir splits as  LLLLRLRRLLLRLRLL, improve=2.038848, (0 missing)
##        Temp9am     < 21.6     to the left,  improve=1.861996, (0 missing)
##        Pressure3pm < 1016.05  to the right, improve=1.675859, (0 missing)
##        WindDir9am  splits as  LLLRRRLLLRRRRLLL, improve=1.607276, (0 missing)
##        MinTemp     < 16.4     to the left,  improve=1.547166, (0 missing)
##    Surrogate splits:
##        WindDir3pm  splits as  LLLLLLLRLLLLLLLL, agree=0.724, adj=0.247, (0 split)
##        WindDir9am  splits as  LLLRRRLLLLLLLRRL, agree=0.698, adj=0.178, (0 split)
##        Pressure3pm < 1013.95  to the right, agree=0.683, adj=0.137, (0 split)
##        Humidity9am < 44.5     to the right, agree=0.668, adj=0.096, (0 split)
##        Humidity3pm < 22.5     to the right, agree=0.658, adj=0.068, (0 split)
##
## Node number 5: 12 observations
##    predicted class=No    expected loss=0.5  P(node) =0.04651163
##      class counts:      6     6
```

```
##     probabilities: 0.500 0.500
##
## Node number 6: 15 observations
##   predicted class=No    expected loss=0.06666667  P(node) =0.05813953
##     class counts:    14      1
##    probabilities: 0.933 0.067
##
## Node number 7: 32 observations
##   predicted class=Yes  expected loss=0.21875  P(node) =0.124031
##     class counts:     7     25
##    probabilities: 0.219 0.781
##
## Node number 8: 126 observations
##   predicted class=No    expected loss=0.01587302  P(node) =0.4883721
##     class counts:   124      2
##    probabilities: 0.984 0.016
##
## Node number 9: 73 observations,    complexity param=0.02717391
##   predicted class=No    expected loss=0.1643836  P(node) =0.2829457
##     class counts:    61     12
##    probabilities: 0.836 0.164
##   left son=18 (37 obs) right son=19 (36 obs)
##   Primary splits:
##       WindDir9am    splits as  LLLRRRLLLRRRRLLL, improve=2.831071, (0 missing)
##       Cloud3pm     < 4.5     to the left,  improve=1.841461, (0 missing)
##       Sunshine     < 10.9    to the right, improve=1.701853, (0 missing)
##       MinTemp      < 14.55   to the left,  improve=1.080769, (0 missing)
##       WindGustSpeed < 54.5   to the left,  improve=1.080769, (0 missing)
##   Surrogate splits:
##       WindDir3pm    splits as  RRRLRRRRRLR--LRL, agree=0.658, adj=0.306, (0 split)
##       Pressure9am  < 1018.3  to the left,  agree=0.644, adj=0.278, (0 split)
##       Temp3pm      < 15.7    to the left,  agree=0.644, adj=0.278, (0 split)
##       MaxTemp      < 16.3    to the left,  agree=0.630, adj=0.250, (0 split)
##       WindGustSpeed < 32     to the right, agree=0.616, adj=0.222, (0 split)
##
## Node number 18: 37 observations
##   predicted class=No    expected loss=0.02702703  P(node) =0.1434109
##     class counts:    36      1
##    probabilities: 0.973 0.027
##
## Node number 19: 36 observations,    complexity param=0.02717391
##   predicted class=No    expected loss=0.3055556  P(node) =0.1395349
##     class counts:    25     11
##    probabilities: 0.694 0.306
##   left son=38 (29 obs) right son=39 (7 obs)
##   Primary splits:
##       Sunshine    < 7.8     to the right, improve=5.287630, (0 missing)
##       Cloud3pm    < 4.5     to the left,  improve=3.906540, (0 missing)
##       Humidity3pm < 28.5    to the left,  improve=2.585470, (0 missing)
##       Cloud9am    < 2.5     to the left,  improve=2.207544, (0 missing)
##       Pressure3pm < 1016.65 to the right, improve=1.754558, (0 missing)
##   Surrogate splits:
##       Humidity3pm < 48.5    to the left,  agree=0.917, adj=0.571, (0 split)
##       Cloud9am    < 6.5     to the left,  agree=0.917, adj=0.571, (0 split)
```

```
##         Cloud3pm    < 6.5      to the left,  agree=0.917, adj=0.571, (0 split)
##         WindDir3pm  splits as  LLRLLLLLR-R--LL-, agree=0.889, adj=0.429, (0 split)
##         Rainfall    < 0.3      to the left,  agree=0.861, adj=0.286, (0 split)
##
## Node number 38: 29 observations
##   predicted class=No   expected loss=0.1724138  P(node) =0.1124031
##     class counts:    24     5
##    probabilities: 0.828 0.172
##
## Node number 39: 7 observations
##   predicted class=Yes  expected loss=0.1428571  P(node) =0.02713178
##     class counts:     1     6
##    probabilities: 0.143 0.857
```