Projet Température : JUNG Thierry

Téléchargement

On télécharge les données d'entrainement et on les parcoure. On s'apercoit qu'il y a presque une cinquantaine de variables. Celles-ci sont presque exclusivement numérique.

```
rm(list=ls())

MeteoTrain <- read.csv("meteo.train.csv", header = T, quote="")
summary(MeteoTrain)
```

```
##        X.           X..Year..      X..Month..        X..Day..
X..Hour..
##  Min.   :   2.0   Min.   :2010   Min.   : 1.000   Min.   : 1.0   Min.   :0
##  1st Qu.: 721.5   1st Qu.:2012   1st Qu.: 3.000   1st Qu.: 8.0   1st Qu.:0
##  Median :1451.0   Median :2014   Median : 6.000   Median :16.0   Median :0
##  Mean   :1459.8   Mean   :2014   Mean   : 6.436   Mean   :15.8   Mean   :0
##  3rd Qu.:2189.0   3rd Qu.:2016   3rd Qu.: 9.000   3rd Qu.:23.0   3rd Qu.:0
##  Max.   :2940.0   Max.   :2018   Max.   :12.000   Max.   :31.0   Max.   :0
##   X..Minute.. X..Temperature.daily.mean..2.m.above.gnd...
##  Min.   :0    Min.   :-7.63
##  1st Qu.:0    1st Qu.: 6.71
##  Median :0    Median :12.08
##  Mean   :0    Mean   :12.23
##  3rd Qu.:0    3rd Qu.:17.54
##  Max.   :0    Max.   :29.45
##  X..Relative.Humidity.daily.mean..2.m.above.gnd...
##  Min.   :38.33
##  1st Qu.:64.82
##  Median :72.21
##  Mean   :71.40
##  3rd Qu.:78.63
##  Max.   :95.54
##  X..Mean.Sea.Level.Pressure.daily.mean..MSL...
##  Min.   : 978.9
##  1st Qu.:1012.4
##  Median :1017.0
##  Mean   :1017.0
##  3rd Qu.:1022.0
##  Max.   :1042.4
##  X..Total.Precipitation.daily.sum..sfc...
##  Min.   : 0.000
##  1st Qu.: 0.000
##  Median : 0.100
##  Mean   : 2.085
##  3rd Qu.: 2.300
##  Max.   :31.500
```

```
##  X..Snowfall.amount.raw.daily.sum..sfc...
##  Min.   :0.00000
##  1st Qu.:0.00000
##  Median :0.00000
##  Mean   :0.04965
##  3rd Qu.:0.00000
##  Max.   :8.61000
##  X..Total.Cloud.Cover.daily.mean..sfc...
##  Min.   :  0.00
##  1st Qu.: 23.80
##  Median : 51.67
##  Mean   : 50.76
##  3rd Qu.: 78.53
##  Max.   :100.00
##  X..High.Cloud.Cover.daily.mean..high.cld.lay...
##  Min.   :  0.000
##  1st Qu.:  1.657
##  Median : 11.880
##  Mean   : 20.284
##  3rd Qu.: 33.260
##  Max.   :100.000
##  X..Medium.Cloud.Cover.daily.mean..mid.cld.lay...
##  Min.   :  0.00
##  1st Qu.:  1.83
##  Median : 24.98
##  Mean   : 31.50
##  3rd Qu.: 54.21
##  Max.   :100.00
##  X..Low.Cloud.Cover.daily.mean..low.cld.lay...
##  Min.   :  0.00
##  1st Qu.:  9.42
##  Median : 36.35
##  Mean   : 39.34
##  3rd Qu.: 65.76
##  Max.   :100.00
##  X..Sunshine.Duration.daily.sum..sfc...
##  Min.   :   0.0
##  1st Qu.: 114.3
##  Median : 366.8
##  Mean   : 373.1
##  3rd Qu.: 587.7
##  Max.   :1015.8
##  X..Shortwave.Radiation.daily.sum..sfc...
##  Min.   : 265.2
##  1st Qu.:2096.2
##  Median :3675.3
##  Mean   :3984.6
##  3rd Qu.:5723.6
##  Max.   :8363.3
##  X..Wind.Speed.daily.mean..10.m.above.gnd...
```

```
##  Min.   : 1.260
##  1st Qu.: 6.428
##  Median : 9.195
##  Mean   :10.707
##  3rd Qu.:12.977
##  Max.   :42.210
##  X..Wind.Direction.daily.mean..10.m.above.gnd...
##  Min.   : 11.19
##  1st Qu.:152.40
##  Median :206.36
##  Mean   :201.82
##  3rd Qu.:254.19
##  Max.   :331.67
##  X..Wind.Speed.daily.mean..80.m.above.gnd...
##  Min.   : 1.34
##  1st Qu.: 8.68
##  Median :12.41
##  Mean   :14.28
##  3rd Qu.:17.61
##  Max.   :54.03
##  X..Wind.Direction.daily.mean..80.m.above.gnd...
##  Min.   : 12.18
##  1st Qu.:157.42
##  Median :213.78
##  Mean   :206.23
##  3rd Qu.:259.06
##  Max.   :333.43
##  X..Wind.Speed.daily.mean..900.mb...
X..Wind.Direction.daily.mean..900.mb...
##  Min.   : 2.25                      Min.   : 17.37
##  1st Qu.:13.02                      1st Qu.:144.02
##  Median :19.57                      Median :233.47
##  Mean   :24.57                      Mean   :206.22
##  3rd Qu.:32.10                      3rd Qu.:265.93
##  Max.   :97.06                      Max.   :344.82
##  X..Wind.Gust.daily.mean..sfc...
X..Temperature.daily.max..2.m.above.gnd...
##  Min.   : 2.25                   Min.   :-3.84
##  1st Qu.: 9.48                   1st Qu.:10.58
##  Median :14.06                   Median :16.54
##  Mean   :16.69                   Mean   :16.54
##  3rd Qu.:21.15                   3rd Qu.:22.36
##  Max.   :79.38                   Max.   :35.77
##  X..Temperature.daily.min..2.m.above.gnd...
##  Min.   :-12.520
##  1st Qu.:  3.350
##  Median :  8.005
##  Mean   :  8.062
##  3rd Qu.: 13.092
##  Max.   : 23.940
```

```
##  X..Relative.Humidity.daily.max..2.m.above.gnd...
##  Min.   : 59.00
##  1st Qu.: 83.00
##  Median : 89.00
##  Mean   : 87.69
##  3rd Qu.: 94.00
##  Max.   :100.00
##  X..Relative.Humidity.daily.min..2.m.above.gnd...
##  Min.   :19.00
##  1st Qu.:45.00
##  Median :54.00
##  Mean   :54.04
##  3rd Qu.:63.00
##  Max.   :92.00
##  X..Mean.Sea.Level.Pressure.daily.max..MSL...
##  Min.   : 981.9
##  1st Qu.:1015.4
##  Median :1019.5
##  Mean   :1019.9
##  3rd Qu.:1024.7
##  Max.   :1045.4
##  X..Mean.Sea.Level.Pressure.daily.min..MSL...
##  Min.   : 977
##  1st Qu.:1009
##  Median :1015
##  Mean   :1014
##  3rd Qu.:1019
##  Max.   :1039
##  X..Total.Cloud.Cover.daily.max..sfc...
X..Total.Cloud.Cover.daily.min..sfc...
##  Min.   :  0.00                        Min.   :  0.000
##  1st Qu.:100.00                        1st Qu.:  0.000
##  Median :100.00                        Median :  0.000
##  Mean   : 88.23                        Mean   :  8.692
##  3rd Qu.:100.00                        3rd Qu.:  2.400
##  Max.   :100.00                        Max.   :100.000
##  X..High.Cloud.Cover.daily.max..high.cld.lay...
##  Min.   :  0.00
##  1st Qu.: 15.00
##  Median : 97.00
##  Mean   : 60.17
##  3rd Qu.:100.00
##  Max.   :100.00
##  X..High.Cloud.Cover.daily.min..high.cld.lay...
##  Min.   :  0.0000
##  1st Qu.:  0.0000
##  Median :  0.0000
##  Mean   :  0.9432
##  3rd Qu.:  0.0000
##  Max.   :100.0000
```

```
##  X..Medium.Cloud.Cover.daily.max..mid.cld.lay...
##  Min.   :  0.00
##  1st Qu.: 22.75
##  Median :100.00
##  Mean   : 70.94
##  3rd Qu.:100.00
##  Max.   :100.00
##  X..Medium.Cloud.Cover.daily.min..mid.cld.lay...
##  Min.   :  0.000
##  1st Qu.:  0.000
##  Median :  0.000
##  Mean   :  2.097
##  3rd Qu.:  0.000
##  Max.   :100.000
##  X..Low.Cloud.Cover.daily.max..low.cld.lay...
##  Min.   :  0
##  1st Qu.:100
##  Median :100
##  Mean   : 80
##  3rd Qu.:100
##  Max.   :100
##  X..Low.Cloud.Cover.daily.min..low.cld.lay...
##  Min.   :  0.000
##  1st Qu.:  0.000
##  Median :  0.000
##  Mean   :  3.879
##  3rd Qu.:  0.000
##  Max.   :100.000
##  X..Wind.Speed.daily.max..10.m.above.gnd...
##  Min.   : 2.52
##  1st Qu.:12.32
##  Median :17.36
##  Mean   :19.06
##  3rd Qu.:23.44
##  Max.   :79.99
##  X..Wind.Speed.daily.min..10.m.above.gnd...
##  Min.   : 0.00
##  1st Qu.: 1.14
##  Median : 2.41
##  Mean   : 3.57
##  3rd Qu.: 4.45
##  Max.   :27.73
##  X..Wind.Speed.daily.max..80.m.above.gnd...
##  Min.   : 3.98
##  1st Qu.:18.27
##  Median :23.85
##  Mean   :25.35
##  3rd Qu.:29.92
##  Max.   :93.84
##  X..Wind.Speed.daily.min..80.m.above.gnd...
```

```
X..Wind.Speed.daily.max..900.mb...
##  Min.   : 0.000                              Min.   :  4.02
##  1st Qu.: 1.140                              1st Qu.: 24.54
##  Median : 2.600                              Median : 37.12
##  Mean   : 4.727                              Mean   : 41.82
##  3rd Qu.: 5.830                              3rd Qu.: 54.37
##  Max.   :37.700                              Max.   :136.25
##  X..Wind.Speed.daily.min..900.mb... X..Wind.Gust.daily.max..sfc...
##  Min.   : 0.00                      Min.   : 4.32
##  1st Qu.: 3.05                      1st Qu.:19.08
##  Median : 6.73                      Median :26.10
##  Mean   :11.09                      Mean   :29.31
##  3rd Qu.:15.31                      3rd Qu.:37.08
##  Max.   :76.13                      Max.   :95.04
##  X..Wind.Gust.daily.min..sfc... X..pluie.demain.. X..temp.demain...
##  Min.   : 0.000                 Mode :logical     Min.   :-7.10
##  1st Qu.: 2.160                 FALSE:579         1st Qu.: 6.77
##  Median : 3.960                 TRUE :601         Median :12.37
##  Mean   : 6.502                                   Mean   :12.19
##  3rd Qu.: 8.280                                   3rd Qu.:17.34
##  Max.   :57.960                                   Max.   :29.96
```

Analyse et préparation

On ne garde que les variables sur les moyennes pour éviter les effets de correlations ainsi que les différentes directions du vent a différentes altitudes. On retire également les données de nébulosité ainsi que de vitesse du vent qui semblent respectivement corrélées avec l'ensoleillement et les rafales de vent.

```
MeteoMeanTrain <- subset(MeteoTrain, select = c(7,8,9,10,11,17,19,24,48))
head (MeteoMeanTrain)
```

```
##   X..Temperature.daily.mean..2.m.above.gnd...
## 1                                       14.99
## 2                                       17.31
## 3                                       21.62
## 4                                       20.22
## 5                                       22.64
## 6                                       18.44
##   X..Relative.Humidity.daily.mean..2.m.above.gnd...
## 1                                             76.46
## 2                                             77.62
## 3                                             69.50
## 4                                             75.08
## 5                                             73.46
## 6                                             76.83
##   X..Mean.Sea.Level.Pressure.daily.mean..MSL...
## 1                                       1014.99
## 2                                       1017.26
## 3                                       1014.59
```

```
## 4                                        1007.74
## 5                                        1003.81
## 6                                        1012.00
##   X..Total.Precipitation.daily.sum..sfc...
## 1                                     1.0
## 2                                     0.0
## 3                                     3.7
## 4                                     0.2
## 5                                     0.0
## 6                                     2.2
##   X..Snowfall.amount.raw.daily.sum..sfc...
## 1                                        0
## 2                                        0
## 3                                        0
## 4                                        0
## 5                                        0
## 6                                        0
##   X..Shortwave.Radiation.daily.sum..sfc...
## 1                                  6709.71
## 2                                  7974.40
## 3                                  4833.59
## 4                                  5389.84
## 5                                  7216.12
## 6                                  3590.26
##   X..Wind.Direction.daily.mean..10.m.above.gnd...
## 1                                          274.72
## 2                                          229.69
## 3                                          214.62
## 4                                          204.97
## 5                                          179.23
## 6                                          219.80
##   X..Wind.Gust.daily.mean..sfc... X..temp.demain...
## 1                           14.88             14.27
## 2                            9.48             19.88
## 3                           13.50             19.31
## 4                            5.31             22.06
## 5                           12.21             20.85
## 6                           13.69             16.13
```

```
Cormean = cor(MeteoMeanTrain)
Cormean
```

```
##
X..Temperature.daily.mean..2.m.above.gnd...
## X..Temperature.daily.mean..2.m.above.gnd...
1.00000000
## X..Relative.Humidity.daily.mean..2.m.above.gnd...
-0.41567501
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...
-0.13506874
```

```
## X..Total.Precipitation.daily.sum..sfc...
-0.01476159
## X..Snowfall.amount.raw.daily.sum..sfc...
-0.19906884
## X..Shortwave.Radiation.daily.sum..sfc...
0.70156783
## X..Wind.Direction.daily.mean..10.m.above.gnd...
0.02681726
## X..Wind.Gust.daily.mean..sfc...
-0.27913573
## X..temp.demain...
0.95470419
##
X..Relative.Humidity.daily.mean..2.m.above.gnd...
## X..Temperature.daily.mean..2.m.above.gnd...
-0.415675014
## X..Relative.Humidity.daily.mean..2.m.above.gnd...
1.000000000
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...
-0.005151056
## X..Total.Precipitation.daily.sum..sfc...
0.361476224
## X..Snowfall.amount.raw.daily.sum..sfc...
0.163114074
## X..Shortwave.Radiation.daily.sum..sfc...
-0.624825488
## X..Wind.Direction.daily.mean..10.m.above.gnd...
0.235941167
## X..Wind.Gust.daily.mean..sfc...
0.063222256
## X..temp.demain...
-0.426221689
##
X..Mean.Sea.Level.Pressure.daily.mean..MSL...
## X..Temperature.daily.mean..2.m.above.gnd...
-0.135068739
## X..Relative.Humidity.daily.mean..2.m.above.gnd...
-0.005151056
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...
1.000000000
## X..Total.Precipitation.daily.sum..sfc...
-0.308874351
## X..Snowfall.amount.raw.daily.sum..sfc...
-0.098376778
## X..Shortwave.Radiation.daily.sum..sfc...
0.026171330
## X..Wind.Direction.daily.mean..10.m.above.gnd...
-0.106642884
## X..Wind.Gust.daily.mean..sfc...
-0.294691686
```

```
## X..temp.demain...
-0.066533919
##
X..Total.Precipitation.daily.sum..sfc...
## X..Temperature.daily.mean..2.m.above.gnd...
-0.01476159
## X..Relative.Humidity.daily.mean..2.m.above.gnd...
0.36147622
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...
-0.30887435
## X..Total.Precipitation.daily.sum..sfc...
1.00000000
## X..Snowfall.amount.raw.daily.sum..sfc...
0.16813648
## X..Shortwave.Radiation.daily.sum..sfc...
-0.32195279
## X..Wind.Direction.daily.mean..10.m.above.gnd...
0.26063649
## X..Wind.Gust.daily.mean..sfc...
0.31996822
## X..temp.demain...
-0.05564471
##
X..Snowfall.amount.raw.daily.sum..sfc...
## X..Temperature.daily.mean..2.m.above.gnd...
-0.19906884
## X..Relative.Humidity.daily.mean..2.m.above.gnd...
0.16311407
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...
-0.09837678
## X..Total.Precipitation.daily.sum..sfc...
0.16813648
## X..Snowfall.amount.raw.daily.sum..sfc...
1.00000000
## X..Shortwave.Radiation.daily.sum..sfc...
-0.15008514
## X..Wind.Direction.daily.mean..10.m.above.gnd...
0.06063235
## X..Wind.Gust.daily.mean..sfc...
0.12702759
## X..temp.demain...
-0.20957717
##
X..Shortwave.Radiation.daily.sum..sfc...
## X..Temperature.daily.mean..2.m.above.gnd...
0.70156783
## X..Relative.Humidity.daily.mean..2.m.above.gnd...
-0.62482549
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...
0.02617133
```

```
## X..Total.Precipitation.daily.sum..sfc...
-0.32195279
## X..Snowfall.amount.raw.daily.sum..sfc...
-0.15008514
## X..Shortwave.Radiation.daily.sum..sfc...
1.00000000
## X..Wind.Direction.daily.mean..10.m.above.gnd...
-0.08747224
## X..Wind.Gust.daily.mean..sfc...
-0.35797063
## X..temp.demain...
0.73074372
##
X..Wind.Direction.daily.mean..10.m.above.gnd...
## X..Temperature.daily.mean..2.m.above.gnd...
0.02681726
## X..Relative.Humidity.daily.mean..2.m.above.gnd...
0.23594117
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...
-0.10664288
## X..Total.Precipitation.daily.sum..sfc...
0.26063649
## X..Snowfall.amount.raw.daily.sum..sfc...
0.06063235
## X..Shortwave.Radiation.daily.sum..sfc...
-0.08747224
## X..Wind.Direction.daily.mean..10.m.above.gnd...
1.00000000
## X..Wind.Gust.daily.mean..sfc...
0.23062499
## X..temp.demain...
-0.01174729
##
X..Wind.Gust.daily.mean..sfc...
## X..Temperature.daily.mean..2.m.above.gnd...                          -
0.27913573
## X..Relative.Humidity.daily.mean..2.m.above.gnd...
0.06322226
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...                        -
0.29469169
## X..Total.Precipitation.daily.sum..sfc...
0.31996822
## X..Snowfall.amount.raw.daily.sum..sfc...
0.12702759
## X..Shortwave.Radiation.daily.sum..sfc...                             -
0.35797063
## X..Wind.Direction.daily.mean..10.m.above.gnd...
0.23062499
## X..Wind.Gust.daily.mean..sfc...
1.00000000
```

```
## X..temp.demain...                                                    -
0.32269739
##                                                  X..temp.demain...
## X..Temperature.daily.mean..2.m.above.gnd...           0.95470419
## X..Relative.Humidity.daily.mean..2.m.above.gnd...    -0.42622169
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...        -0.06653392
## X..Total.Precipitation.daily.sum..sfc...             -0.05564471
## X..Snowfall.amount.raw.daily.sum..sfc...             -0.20957717
## X..Shortwave.Radiation.daily.sum..sfc...              0.73074372
## X..Wind.Direction.daily.mean..10.m.above.gnd...      -0.01174729
## X..Wind.Gust.daily.mean..sfc...                      -0.32269739
## X..temp.demain...                                     1.00000000
```

Regression

En faisant une regression linéaire sur l'ensemble de celles-ci, 3 variables ressortent comme très pertinentes : "La température", "La pression" et "Le rayonnement solaire". 2 variables ressortent comme pertinentes : "La direction du vent" et "les rafales de vent"

```
mBig = lm(formula = X..temp.demain... ~
X..Temperature.daily.mean..2.m.above.gnd... +
X..Relative.Humidity.daily.mean..2.m.above.gnd... +
X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Total.Precipitation.daily.sum..sfc... +
X..Snowfall.amount.raw.daily.sum..sfc... +
X..Shortwave.Radiation.daily.sum..sfc... +
X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc... , data=MeteoMeanTrain)
summary(mBig)
```

```
##
## Call:
## lm(formula = X..temp.demain... ~
X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Relative.Humidity.daily.mean..2.m.above.gnd... +
X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
##     X..Total.Precipitation.daily.sum..sfc... +
X..Snowfall.amount.raw.daily.sum..sfc... +
##     X..Shortwave.Radiation.daily.sum..sfc... +
X..Wind.Direction.daily.mean..10.m.above.gnd... +
##     X..Wind.Gust.daily.mean..sfc..., data = MeteoMeanTrain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3893 -1.2533  0.0285  1.2228  6.7876
##
## Coefficients:
##                                                  Estimate Std. Error t
value
## (Intercept)                                    -3.949e+01  8.164e+00  -
```

```
4.837
## X..Temperature.daily.mean..2.m.above.gnd...        8.793e-01  1.264e-02
69.558
## X..Relative.Humidity.daily.mean..2.m.above.gnd...  9.791e-03  8.503e-03
1.151
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...      3.876e-02  7.945e-03
4.878
## X..Total.Precipitation.daily.sum..sfc...           3.408e-02  1.739e-02
1.960
## X..Snowfall.amount.raw.daily.sum..sfc...          -2.575e-01  1.501e-01  -
1.715
## X..Shortwave.Radiation.daily.sum..sfc...           3.830e-04  4.689e-05
8.169
## X..Wind.Direction.daily.mean..10.m.above.gnd...   -2.516e-03  1.005e-03  -
2.503
## X..Wind.Gust.daily.mean..sfc...                   -1.599e-02  6.847e-03  -
2.335
##                                                  Pr(>|t|)
## (Intercept)                                      1.50e-06 ***
## X..Temperature.daily.mean..2.m.above.gnd...       < 2e-16 ***
## X..Relative.Humidity.daily.mean..2.m.above.gnd...   0.2498
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...    1.22e-06 ***
## X..Total.Precipitation.daily.sum..sfc...            0.0502 .
## X..Snowfall.amount.raw.daily.sum..sfc...            0.0866 .
## X..Shortwave.Radiation.daily.sum..sfc...         7.99e-16 ***
## X..Wind.Direction.daily.mean..10.m.above.gnd...     0.0124 *
## X..Wind.Gust.daily.mean..sfc...                     0.0197 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.951 on 1171 degrees of freedom
## Multiple R-squared:  0.9226, Adjusted R-squared:  0.9221
## F-statistic:  1745 on 8 and 1171 DF,  p-value: < 2.2e-16
```

Regression

On effectue une regression sur les 5 variables puis sur les 3 variables.

```
m5var = lm(formula = X..temp.demain... ~
X..Temperature.daily.mean..2.m.above.gnd... +
X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Shortwave.Radiation.daily.sum..sfc... +
X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc... , data=MeteoMeanTrain)
summary(m5var)
```

```
##
## Call:
## lm(formula = X..temp.demain... ~
X..Temperature.daily.mean..2.m.above.gnd... +
```

```
##     X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Shortwave.Radiation.daily.sum..sfc... +
##     X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc...,
##     data = MeteoMeanTrain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3159 -1.2571  0.0346  1.2400  6.9283
##
## Coefficients:
##                                              Estimate Std. Error t
value
## (Intercept)                                -3.640e+01  7.960e+00  -
4.572
## X..Temperature.daily.mean..2.m.above.gnd...  8.891e-01  1.195e-02
74.401
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...  3.648e-02  7.769e-03
4.695
## X..Shortwave.Radiation.daily.sum..sfc...  3.211e-04  3.856e-05
8.326
## X..Wind.Direction.daily.mean..10.m.above.gnd... -1.905e-03  9.585e-04  -
1.987
## X..Wind.Gust.daily.mean..sfc...            -1.653e-02  6.417e-03  -
2.577
##                                             Pr(>|t|)
## (Intercept)                                5.34e-06 ***
## X..Temperature.daily.mean..2.m.above.gnd...  < 2e-16 ***
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...  2.98e-06 ***
## X..Shortwave.Radiation.daily.sum..sfc...  2.30e-16 ***
## X..Wind.Direction.daily.mean..10.m.above.gnd...   0.0471 *
## X..Wind.Gust.daily.mean..sfc...            0.0101 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.955 on 1174 degrees of freedom
## Multiple R-squared:  0.9221, Adjusted R-squared:  0.9217
## F-statistic:  2778 on 5 and 1174 DF,  p-value: < 2.2e-16
```

```
m3var = lm(formula = X..temp.demain... ~
X..Temperature.daily.mean..2.m.above.gnd... +
X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Shortwave.Radiation.daily.sum..sfc..., data=MeteoMeanTrain)
summary(m3var)
```

```
##
## Call:
## lm(formula = X..temp.demain... ~
X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
```

```
X..Shortwave.Radiation.daily.sum..sfc...,
##     data = MeteoMeanTrain)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3409 -1.3312  0.0325  1.2697  6.9819
##
## Coefficients:
##                                              Estimate Std. Error t
value
## (Intercept)                                -4.513e+01  7.520e+00  -
6.001
## X..Temperature.daily.mean..2.m.above.gnd...    8.902e-01  1.182e-02
75.285
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...  4.428e-02  7.388e-03
5.994
## X..Shortwave.Radiation.daily.sum..sfc...       3.508e-04  3.785e-05
9.267
##                                               Pr(>|t|)
## (Intercept)                                 2.61e-09 ***
## X..Temperature.daily.mean..2.m.above.gnd...  < 2e-16 ***
## X..Mean.Sea.Level.Pressure.daily.mean..MSL... 2.72e-09 ***
## X..Shortwave.Radiation.daily.sum..sfc...     < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.965 on 1176 degrees of freedom
## Multiple R-squared:  0.9212, Adjusted R-squared:  0.921
## F-statistic:  4582 on 3 and 1176 DF,  p-value: < 2.2e-16
```

Automatque

On peut aussi essayer de chercher de facon automatique le meilleur modèle avec la fonction step. Puis on détermine le meilleur modèle pour expliquer les données en comparant via l'ANOVA puis via les critères AIC. C'est toujours le modèle Auto qui l'emporte...

```
mAuto = step(lm(formula = X..temp.demain... ~
X..Temperature.daily.mean..2.m.above.gnd... +
X..Relative.Humidity.daily.mean..2.m.above.gnd... +
X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Total.Precipitation.daily.sum..sfc... +
X..Snowfall.amount.raw.daily.sum..sfc... +
X..Shortwave.Radiation.daily.sum..sfc... +
X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc... , data=MeteoMeanTrain))

## Start:  AIC=1586.19
## X..temp.demain... ~ X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Relative.Humidity.daily.mean..2.m.above.gnd... +
```

```
X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
##     X..Total.Precipitation.daily.sum..sfc... +
X..Snowfall.amount.raw.daily.sum..sfc... +
##     X..Shortwave.Radiation.daily.sum..sfc... +
X..Wind.Direction.daily.mean..10.m.above.gnd... +
##     X..Wind.Gust.daily.mean..sfc...
##
##                                                    Df Sum of Sq     RSS
AIC
## - X..Relative.Humidity.daily.mean..2.m.above.gnd...  1       5.0  4462.1
1585.5
## <none>                                                            4457.1
1586.2
## - X..Snowfall.amount.raw.daily.sum..sfc...           1      11.2  4468.3
1587.2
## - X..Total.Precipitation.daily.sum..sfc...           1      14.6  4471.7
1588.1
## - X..Wind.Gust.daily.mean..sfc...                    1      20.8  4477.8
1589.7
## - X..Wind.Direction.daily.mean..10.m.above.gnd...    1      23.9  4480.9
1590.5
## - X..Mean.Sea.Level.Pressure.daily.mean..MSL...      1      90.6  4547.6
1607.9
## - X..Shortwave.Radiation.daily.sum..sfc...           1     254.0  4711.1
1649.6
## - X..Temperature.daily.mean..2.m.above.gnd...        1   18415.7 22872.8
3514.0
##
## Step:  AIC=1585.53
## X..temp.demain... ~ X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Total.Precipitation.daily.sum..sfc... +
##     X..Snowfall.amount.raw.daily.sum..sfc... +
X..Shortwave.Radiation.daily.sum..sfc... +
##     X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc...
##
##                                                   Df Sum of Sq     RSS
AIC
## <none>                                                          4462.1
1585.5
## - X..Snowfall.amount.raw.daily.sum..sfc...         1      10.2  4472.4
1586.2
## - X..Wind.Direction.daily.mean..10.m.above.gnd...  1      19.9  4482.0
1588.8
## - X..Total.Precipitation.daily.sum..sfc...         1      20.1  4482.2
1588.8
## - X..Wind.Gust.daily.mean..sfc...                  1      31.2  4493.3
1591.8
## - X..Mean.Sea.Level.Pressure.daily.mean..MSL...    1      90.9  4553.0
```

```
1607.3
## - X..Shortwave.Radiation.daily.sum..sfc...         1     282.6  4744.7
1656.0
## - X..Temperature.daily.mean..2.m.above.gnd...      1   18470.5 22932.6
3515.1

summary(mAuto)

##
## Call:
## lm(formula = X..temp.demain... ~
X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Total.Precipitation.daily.sum..sfc... +
##     X..Snowfall.amount.raw.daily.sum..sfc... +
X..Shortwave.Radiation.daily.sum..sfc... +
##     X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc...,
##     data = MeteoMeanTrain)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.307 -1.262  0.049  1.242  6.823
##
## Coefficients:
##                                                 Estimate Std. Error t
value
## (Intercept)                                   -3.877e+01  8.141e+00  -
4.762
## X..Temperature.daily.mean..2.m.above.gnd...    8.782e-01  1.261e-02
69.652
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...  3.883e-02  7.946e-03
4.886
## X..Total.Precipitation.daily.sum..sfc...       3.882e-02  1.690e-02
2.297
## X..Snowfall.amount.raw.daily.sum..sfc...      -2.457e-01  1.498e-01  -
1.640
## X..Shortwave.Radiation.daily.sum..sfc...       3.580e-04  4.155e-05
8.616
## X..Wind.Direction.daily.mean..10.m.above.gnd... -2.220e-03  9.717e-04  -
2.284
## X..Wind.Gust.daily.mean..sfc...               -1.855e-02  6.478e-03  -
2.864
##                                                Pr(>|t|)
## (Intercept)                                   2.16e-06 ***
## X..Temperature.daily.mean..2.m.above.gnd...    < 2e-16 ***
## X..Mean.Sea.Level.Pressure.daily.mean..MSL...  1.17e-06 ***
## X..Total.Precipitation.daily.sum..sfc...       0.02179 *
## X..Snowfall.amount.raw.daily.sum..sfc...       0.10127
## X..Shortwave.Radiation.daily.sum..sfc...       < 2e-16 ***
```

```
## X..Wind.Direction.daily.mean..10.m.above.gnd...  0.02253 *
## X..Wind.Gust.daily.mean..sfc...                  0.00426 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.951 on 1172 degrees of freedom
## Multiple R-squared:  0.9225, Adjusted R-squared:  0.9221
## F-statistic:  1994 on 7 and 1172 DF,  p-value: < 2.2e-16
```

```
anova(m3var,m5var)
```

```
## Analysis of Variance Table
##
## Model 1: X..temp.demain... ~ X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Shortwave.Radiation.daily.sum..sfc...
## Model 2: X..temp.demain... ~ X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Shortwave.Radiation.daily.sum..sfc... +
##     X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc...
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1   1176 4539.7
## 2   1174 4488.9  2    50.775 6.6397 0.001357 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(m5var,mAuto)
```

```
## Analysis of Variance Table
##
## Model 1: X..temp.demain... ~ X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Shortwave.Radiation.daily.sum..sfc... +
##     X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc...
## Model 2: X..temp.demain... ~ X..Temperature.daily.mean..2.m.above.gnd... +
##     X..Mean.Sea.Level.Pressure.daily.mean..MSL... +
X..Total.Precipitation.daily.sum..sfc... +
##     X..Snowfall.amount.raw.daily.sum..sfc... +
X..Shortwave.Radiation.daily.sum..sfc... +
##     X..Wind.Direction.daily.mean..10.m.above.gnd... +
X..Wind.Gust.daily.mean..sfc...
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1   1174 4488.9
## 2   1172 4462.1  2    26.793 3.5187 0.02995 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
c(AIC(m3var),AIC(m5var))
```

```
## [1] 4948.559 4939.287
```

```
c(AIC(m5var),AIC(mAuto))
```

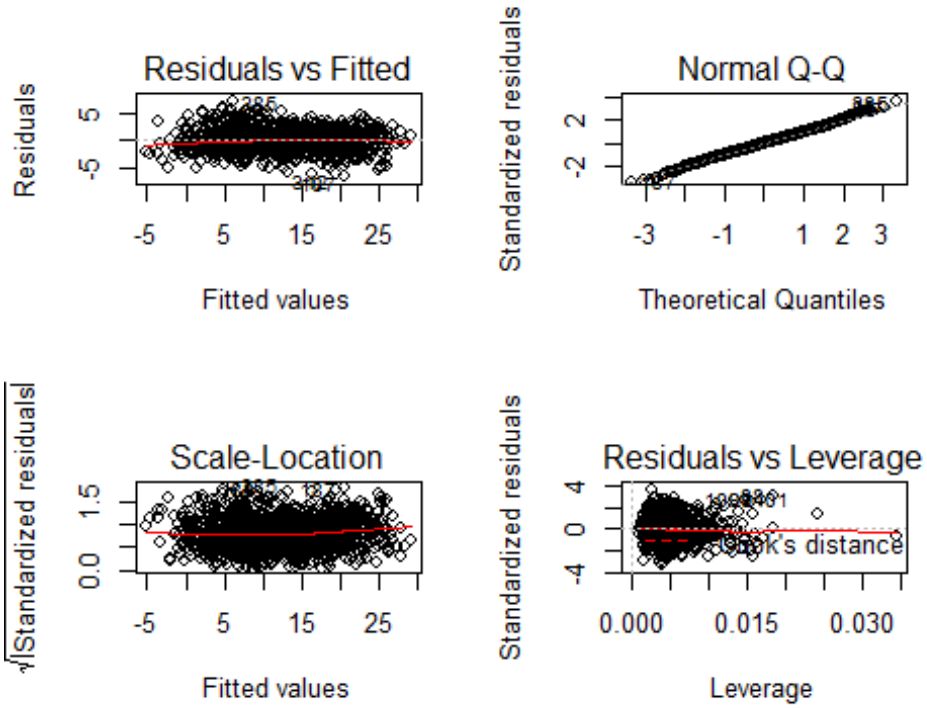
```
## [1] 4939.287 4936.223
```

Validation des hypothèses

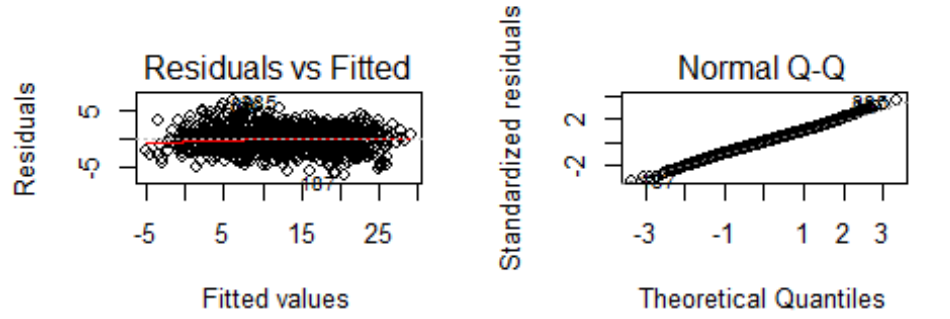
On valide les hypothèses du modèle linéaire en sortant les différents graphes de la régression. 1) On suppose que les données ont été collectées de façon indépendantes. 2) Sur le graphe 1, on valide qu'il n'y a aps de formes particulières et que les résidus sont de moyenne nulle (Par construction) 3) Sur le graphe des résidus standardisés, on vérifie qu'ils ont la même variance. On ne détecte pas de structure particulère. 4) On regarde le QQ-plot des résidus estimés (graphique en haut à droite) est une façon de tester le caractère gaussien des résidus. on voit une droite donc on suppose que les résidus sont gaussiens.

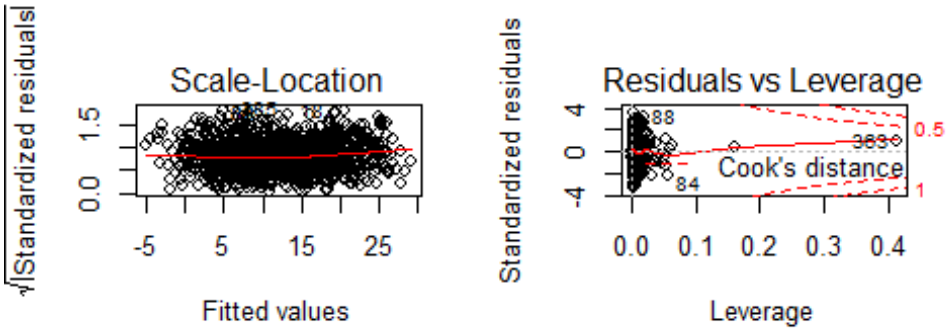
Sur le graphe en bas à droite, on ne détecte pas de points aberrants.

```
par(mfrow=c(2,2))
plot(m5var)
```




```
plot(mAuto)
```





Prédiction

Nous allons faire maintenant une prédiction avec les 2 modèles à 3 et 5 variables. On commence par télécharger la base de test puis on utilise la fonction “predict”.

```
MeteoTest <- read.csv("meteo.test.csv", header = T, quote="")

pred3var = predict(m3var,MeteoTest)
pred5var = predict(m5var,MeteoTest)

head(pred3var)

##        1        2        3        4        5        6
## 19.51080 25.16345 16.15399 17.02050 17.47765 18.24166

head(pred5var)

##        1        2        3        4        5        6
## 19.26881 25.27831 15.87909 17.02819 17.64932 18.32664
```

Validation visuelle

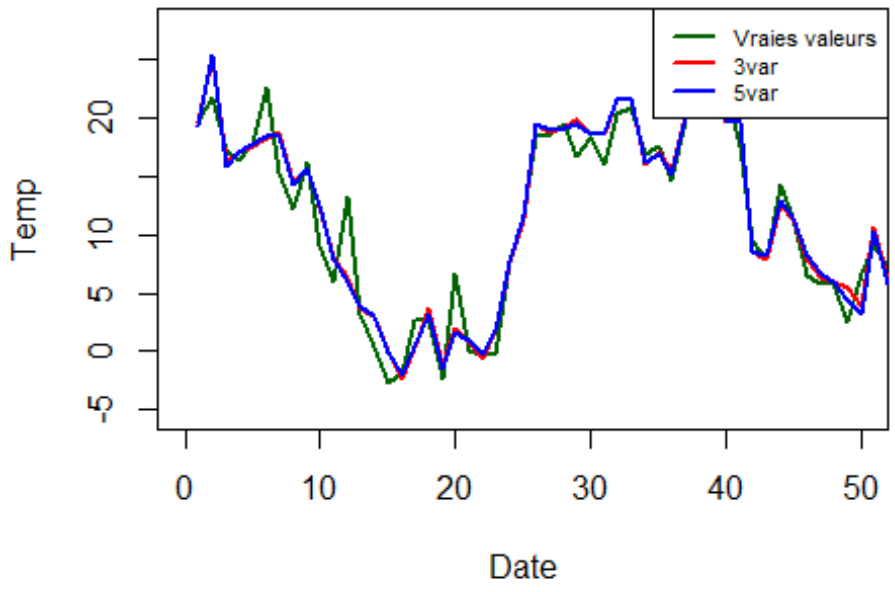
On trace les droites des prévisions en focntions des vraies valeurs pour valider “visuellement” notre modèle.

```
plot(MeteoTest$X..temp.demain...,col="darkgreen",lwd=2,type='l',
ylab="Temp",xlab="Date", xlim=c(0,50))

points(pred3var,col="red",lwd=2,type='l')
```

```
points(pred5var,col="blue",lwd=2,type='l')

legend("topright",c("Vraies valeurs","3var","5var"),
col=c('darkgreen','red','blue'), lty=c(rep(1,3),2),lwd=c(rep(2,3),1),cex=0.7)
```



construction du fichier de prediction

On ajoute les colonnes de prédiction au fichier MeteoTest.

```
MeteoTest$prev3<-pred3var
MeteoTest$prev5<-pred5var
```

construction du fichier de prediction

On “tronque” le fichier pour ne conserver que les jours avec nos prédictions. On exporte le fichier sous format CSV2.

```
Prevision <- subset(MeteoTest, select = c(1,2,3,4,48,49,50) )
head(Prevision)

##   X. X..Year.. X..Month.. X..Day.. X..temp.demain...    prev3    prev5
## 1 36      2010          7        6             19.65 19.51080 19.26881
## 2 44      2010          7       14             21.66 25.16345 25.27831
## 3 54      2010          7       24             17.23 16.15399 15.87909
## 4 58      2010          7       28             16.29 17.02050 17.02819
## 5 74      2010          8       13             17.65 17.47765 17.64932
## 6 86      2010          8       25             22.60 18.24166 18.32664
```

```r
write.csv2(Prevision, file = "PrevisionTemp.csv")
```