

Activity1

Monte Davityan

2022-09-07

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.7      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(ggplot2)
library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

fire <- read.csv("forestfires.csv")

head(fire)

##   X Y month day FFMC  DMC   DC  ISI temp RH wind rain area
## 1 7 5  mar fri 86.2 26.2 94.3  5.1  8.2 51  6.7  0.0   0
## 2 7 4  oct tue 90.6 35.4 669.1  6.7 18.0 33  0.9  0.0   0
## 3 7 4  oct sat 90.6 43.7 686.9  6.7 14.6 33  1.3  0.0   0
## 4 8 6  mar fri 91.7 33.3  77.5  9.0  8.3 97  4.0  0.2   0
## 5 8 6  mar sun 89.3 51.3 102.2  9.6 11.4 99  1.8  0.0   0
## 6 8 6  aug sun 92.3 85.3 488.0 14.7 22.2 29  5.4  0.0   0

str(fire)

## 'data.frame':   517 obs. of  13 variables:
##  $ X      : int  7 7 7 8 8 8 8 8 8 7 ...
##  $ Y      : int  5 4 4 6 6 6 6 6 6 5 ...
##  $ month: chr  "mar" "oct" "oct" "mar" ...
##  $ day   : chr  "fri" "tue" "sat" "fri" ...
##  $ FFMC  : num  86.2 90.6 90.6 91.7 89.3 92.3 92.3 91.5 91 92.5 ...
##  $ DMC   : num  26.2 35.4 43.7 33.3 51.3 ...
##  $ DC    : num  94.3 669.1 686.9 77.5 102.2 ...
##  $ ISI   : num  5.1 6.7 6.7 9 9.6 14.7 8.5 10.7 7 7.1 ...
##  $ temp  : num  8.2 18 14.6 8.3 11.4 22.2 24.1 8 13.1 22.8 ...
##  $ RH    : int  51 33 33 97 99 29 27 86 63 40 ...
##  $ wind  : num  6.7 0.9 1.3 4 1.8 5.4 3.1 2.2 5.4 4 ...
##  $ rain  : num  0 0 0 0.2 0 0 0 0 0 0 ...
```

```
## $ area : num 0 0 0 0 0 0 0 0 0 0 ...
```

```
summary(fire)
```

```
##           X           Y           month           day
## Min.      :1.000   Min.      :2.0   Length:517      Length:517
## 1st Qu.:3.000   1st Qu.:4.0   Class :character Class :character
## Median :4.000   Median :4.0   Mode  :character Mode  :character
## Mean      :4.669   Mean      :4.3
## 3rd Qu.:7.000   3rd Qu.:5.0
## Max.      :9.000   Max.      :9.0
##           FPMC           DMC           DC           ISI
## Min.      :18.70   Min.      : 1.1   Min.      : 7.9   Min.      : 0.000
## 1st Qu.:90.20   1st Qu.: 68.6   1st Qu.:437.7   1st Qu.: 6.500
## Median :91.60   Median :108.3   Median :664.2   Median : 8.400
## Mean      :90.64   Mean      :110.9   Mean      :547.9   Mean      : 9.022
## 3rd Qu.:92.90   3rd Qu.:142.4   3rd Qu.:713.9   3rd Qu.:10.800
## Max.      :96.20   Max.      :291.3   Max.      :860.6   Max.      :56.100
##           temp           RH           wind           rain
## Min.      : 2.20   Min.      : 15.00   Min.      :0.400   Min.      :0.00000
## 1st Qu.:15.50   1st Qu.: 33.00   1st Qu.:2.700   1st Qu.:0.00000
## Median :19.30   Median : 42.00   Median :4.000   Median :0.00000
## Mean      :18.89   Mean      : 44.29   Mean      :4.018   Mean      :0.02166
## 3rd Qu.:22.80   3rd Qu.: 53.00   3rd Qu.:4.900   3rd Qu.:0.00000
## Max.      :33.30   Max.      :100.00   Max.      :9.400   Max.      :6.40000
##           area
## Min.      : 0.00
## 1st Qu.: 0.00
## Median : 0.52
## Mean      : 12.85
## 3rd Qu.: 6.57
## Max.      :1090.84
```

```
#ggpairs(fire)
```

```
fire <- fire %>% mutate(month = as.factor(month),
                        day = as.factor(day))
```

```
# Reponse area (multiple regression)
```

```
lm.fit <- lm(area ~ ., data = fire )
```

```
summary(lm.fit)
```

```
##
## Call:
## lm(formula = area ~ ., data = fire)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -55.32  -17.84   -6.82    4.99  1039.28
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -15.16402   76.56086  -0.198   0.8431
## X              2.25583    1.49786   1.506   0.1327
## Y             -0.14765    2.81881  -0.052   0.9582
## monthaug      46.88205   38.08792   1.231   0.2190
```

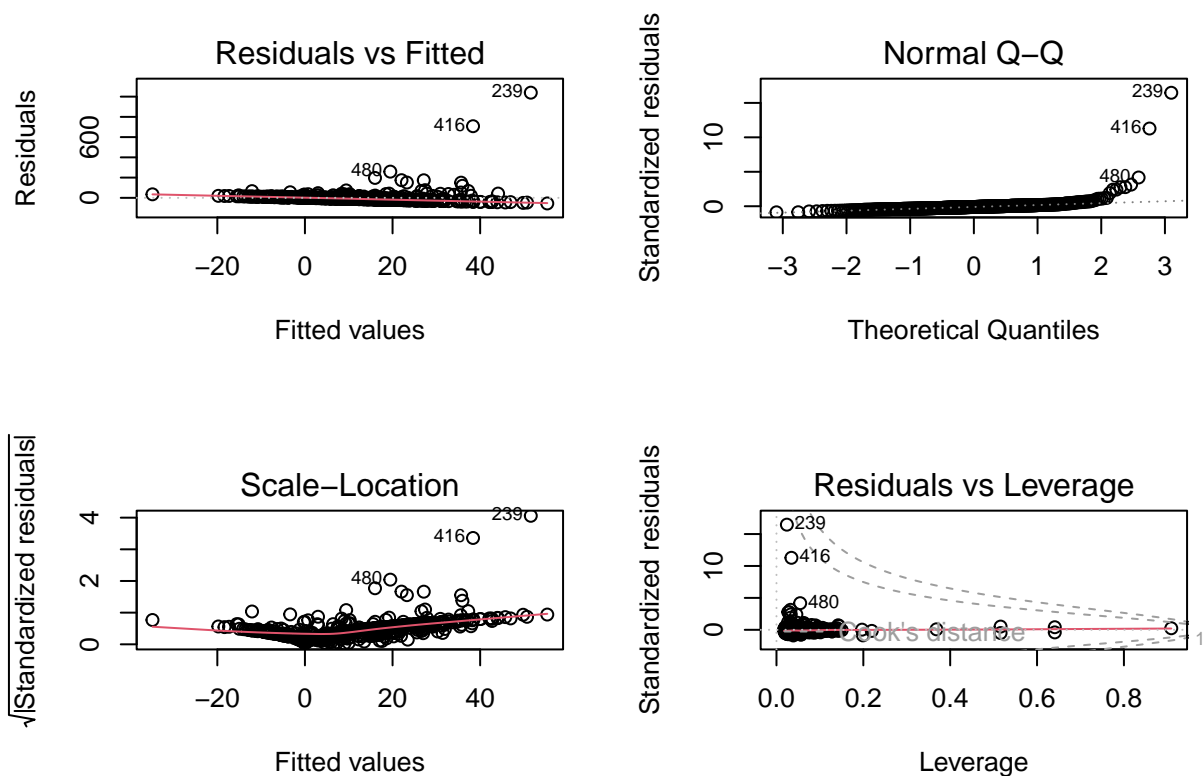
```

## monthdec      47.37821    36.94830    1.282    0.2004
## monthfeb       5.58985    25.94816    0.215    0.8295
## monthjan      14.76909    56.40617    0.262    0.7936
## monthjul      28.87889    33.05232    0.874    0.3827
## monthjun       6.71548    30.33765    0.221    0.8249
## monthmar      -4.22256    23.41447   -0.180    0.8570
## monthmay      12.79646    50.91572    0.251    0.8017
## monthnov      -4.41010    68.37767   -0.064    0.9486
## monthoct      68.97536    45.42009    1.519    0.1295
## monthsep      73.73192    42.67672    1.728    0.0847 .
## daymon        5.96928    10.48154    0.570    0.5693
## daysat       19.40993    10.06218    1.929    0.0543 .
## daysun        5.14460     9.78870    0.526    0.5994
## daythu        9.67192    11.10696    0.871    0.3843
## daytue        7.79282    10.88291    0.716    0.4743
## daywed        5.47914    11.40526    0.480    0.6312
## FFMC         -0.09527     0.76985   -0.124    0.9016
## DMC           0.20106     0.08681    2.316    0.0210 *
## DC           -0.12880     0.05872   -2.194    0.0287 *
## ISI          -0.54416     0.83105   -0.655    0.5129
## temp          1.29620     1.03082    1.257    0.2092
## RH           -0.13476     0.28845   -0.467    0.6406
## wind          1.97427     1.77824    1.110    0.2674
## rain         -2.81545     9.92647   -0.284    0.7768
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.88 on 489 degrees of freedom
## Multiple R-squared:  0.04578,    Adjusted R-squared:  -0.006905
## F-statistic: 0.8689 on 27 and 489 DF,  p-value: 0.6581

par(mfrow=c(2,2))
plot(lm.fit)

## Warning: not plotting observations with leverage one:
## 517

```



```
# Binary area
fire <- fire %>%
  mutate(binary_area = if_else(area != 0, "Not zero", "Zero")) %>%
  mutate(binary_area = as.factor(binary_area))

glm.fit <- glm(binary_area ~ ., family = "binomial", data = fire %>% select(-area))
summary(glm.fit)
```

```
##
## Call:
## glm(formula = binary_area ~ ., family = "binomial", data = fire %>%
##   select(-area))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5873  -1.0993  -0.8112   1.1860   1.6016
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.644e+00  2.888e+00   1.608   0.108
## X            -5.838e-02  4.806e-02  -1.215   0.225
## Y            -4.134e-02  9.078e-02  -0.455   0.649
## monthaug     2.074e-01  1.214e+00   0.171   0.864
## monthdec    -1.682e+01  7.894e+02  -0.021   0.983
## monthfeb    -4.220e-01  8.237e-01  -0.512   0.608
## monthjan     1.505e+01  1.556e+03   0.010   0.992
```

```

## monthjul      1.292e-01  1.054e+00  0.123  0.902
## monthjun      3.762e-01  9.718e-01  0.387  0.699
## monthmar      4.897e-01  7.494e-01  0.653  0.513
## monthmay     -8.583e-03  1.603e+00 -0.005  0.996
## monthnov      1.631e+01  2.400e+03  0.007  0.995
## monthoct      1.005e+00  1.456e+00  0.691  0.490
## monthsep     -5.052e-03  1.360e+00 -0.004  0.997
## daymon       -1.331e-01  3.400e-01 -0.391  0.695
## daysat       -6.636e-02  3.229e-01 -0.206  0.837
## daysun        1.264e-02  3.146e-01  0.040  0.968
## daythu        3.645e-03  3.569e-01  0.010  0.992
## daytue       -2.725e-01  3.504e-01 -0.778  0.437
## daywed       -3.474e-01  3.696e-01 -0.940  0.347
## FFMC         -3.146e-02  3.039e-02 -1.035  0.301
## DMC          1.138e-03  2.769e-03  0.411  0.681
## DC          -4.078e-04  1.871e-03 -0.218  0.827
## ISI          1.591e-02  2.803e-02  0.568  0.570
## temp        -4.861e-02  3.352e-02 -1.450  0.147
## RH          -5.851e-03  9.514e-03 -0.615  0.539
## wind        -8.036e-02  5.786e-02 -1.389  0.165
## rain        -6.886e-03  3.492e-01 -0.020  0.984
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 715.69  on 516  degrees of freedom
## Residual deviance: 678.24  on 489  degrees of freedom
## AIC: 734.24
##
## Number of Fisher Scoring iterations: 15

```