dicussion12

Keith Colella

2023-11-16

```
library(tidyverse)
library(cowplot)
library(car)
```

Assignment

Using R, build a multiple regression model for data that interests you. Include in this model at least one quadratic term, one dichotomous term, and one dichotomous vs. quantitative interaction term. Interpret all coefficients. Conduct residual analysis. Was the linear model appropriate? Why or why not?

Response

Data

https://www.kaggle.com/datasets/mirichoi0218/insurance
(https://www.kaggle.com/datasets/mirichoi0218/insurance)

```
data <- read_csv('data/insurance.csv')
```

```
## Rows: 1338 Columns: 7
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (3): sex, smoker, region
## dbl (4): age, bmi, children, charges
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
glimpse(data)
```

```
## Rows: 1,338
## Columns: 7
## $ age      <dbl> 19, 18, 28, 33, 32, 31, 46, 37, 37, 60, 25, 62, 23, 56, 27, 1…
## $ sex      <chr> "female", "male", "male", "male", "male", "female", "female",…
## $ bmi      <dbl> 27.900, 33.770, 33.000, 22.705, 28.880, 25.740, 33.440, 27.74…
## $ children <dbl> 0, 1, 3, 0, 0, 0, 1, 3, 2, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0…
## $ smoker   <chr> "yes", "no", "no", "no", "no", "no", "no", "no", "no", "no", …
## $ region   <chr> "southwest", "southeast", "southeast", "northwest", "northwes…
## $ charges  <dbl> 16884.924, 1725.552, 4449.462, 21984.471, 3866.855, 3756.622,…
```

Cleaning

```
for (col in colnames(data)) {
  data[col] %>%
    filter(is.na(!!sym(col))) %>%
    nrow() %>% print()
}
```

```
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
## [1] 0
```

```
data$sex %>% unique()
```

```
## [1] "female" "male"
```

```
data$smoker %>% unique()
```

```
## [1] "yes" "no"
```

```
data$region %>% unique()
```

```
## [1] "southwest" "southeast" "northwest" "northeast"
```

```
data <- data %>%
  mutate(
    sex_female = if_else(sex == 'female', 1, 0),
    smoker = if_else(smoker == 'yes', 1, 0),
    north = if_else(region == 'northwest' | region == 'northeast', 1, 0),
    east = if_else(region == 'northeast' | region == 'southeast', 1, 0)
  ) %>%
  select(-sex, -region)

head(data)
```

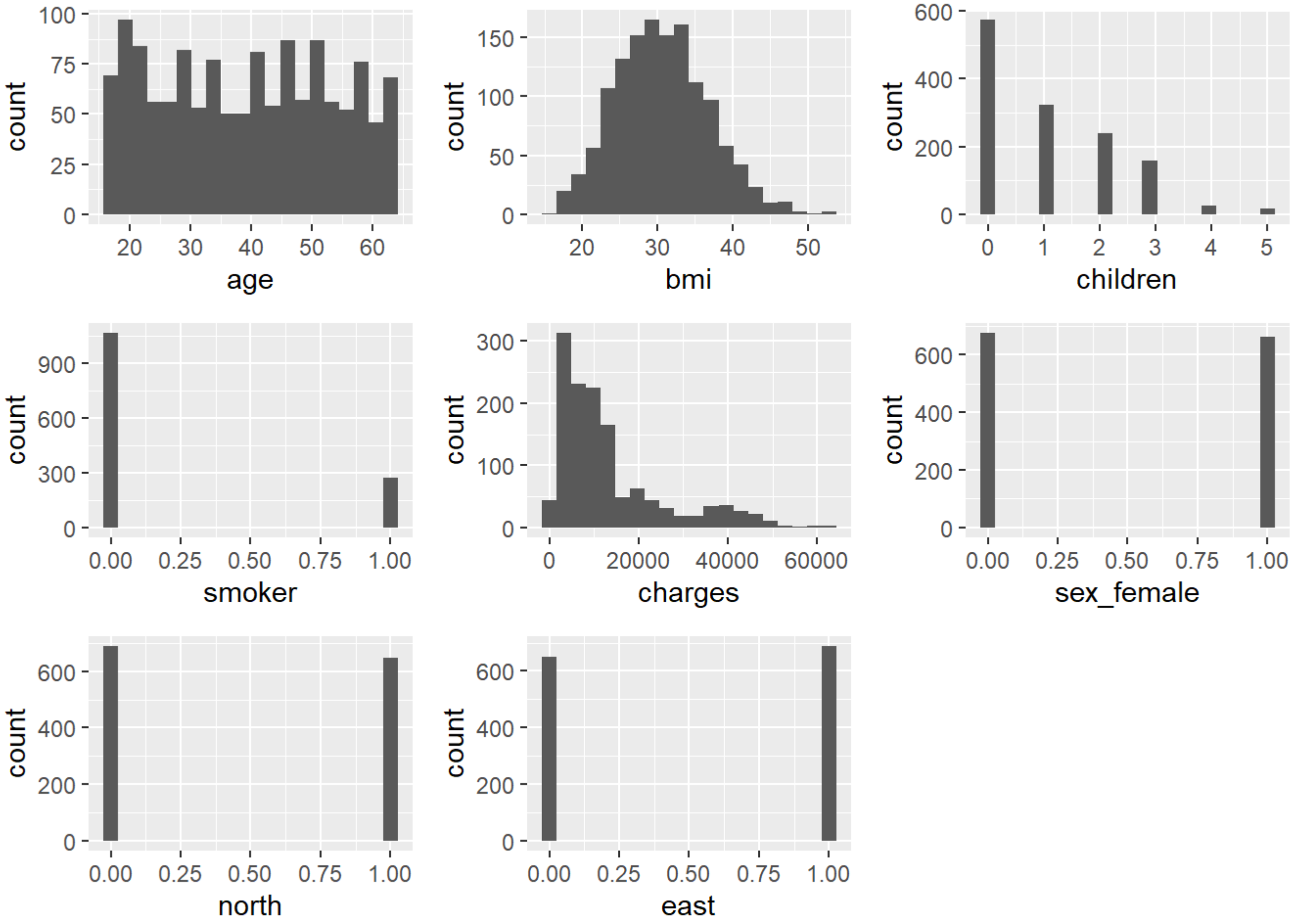
```
## # A tibble: 6 × 8
##     age   bmi children smoker charges sex_female north  east
##   <dbl> <dbl>    <dbl>  <dbl>   <dbl>      <dbl> <dbl> <dbl>
## 1    19  27.9        0      1  16885.          1     0     0
## 2    18  33.8        1      0   1726.          0     0     1
## 3    28  33          3      0   4449.          0     0     1
## 4    33  22.7        0      0  21984.          0     1     0
## 5    32  28.9        0      0   3867.          0     1     0
## 6    31  25.7        0      0   3757.          1     0     1
```

```
summary(data)
```

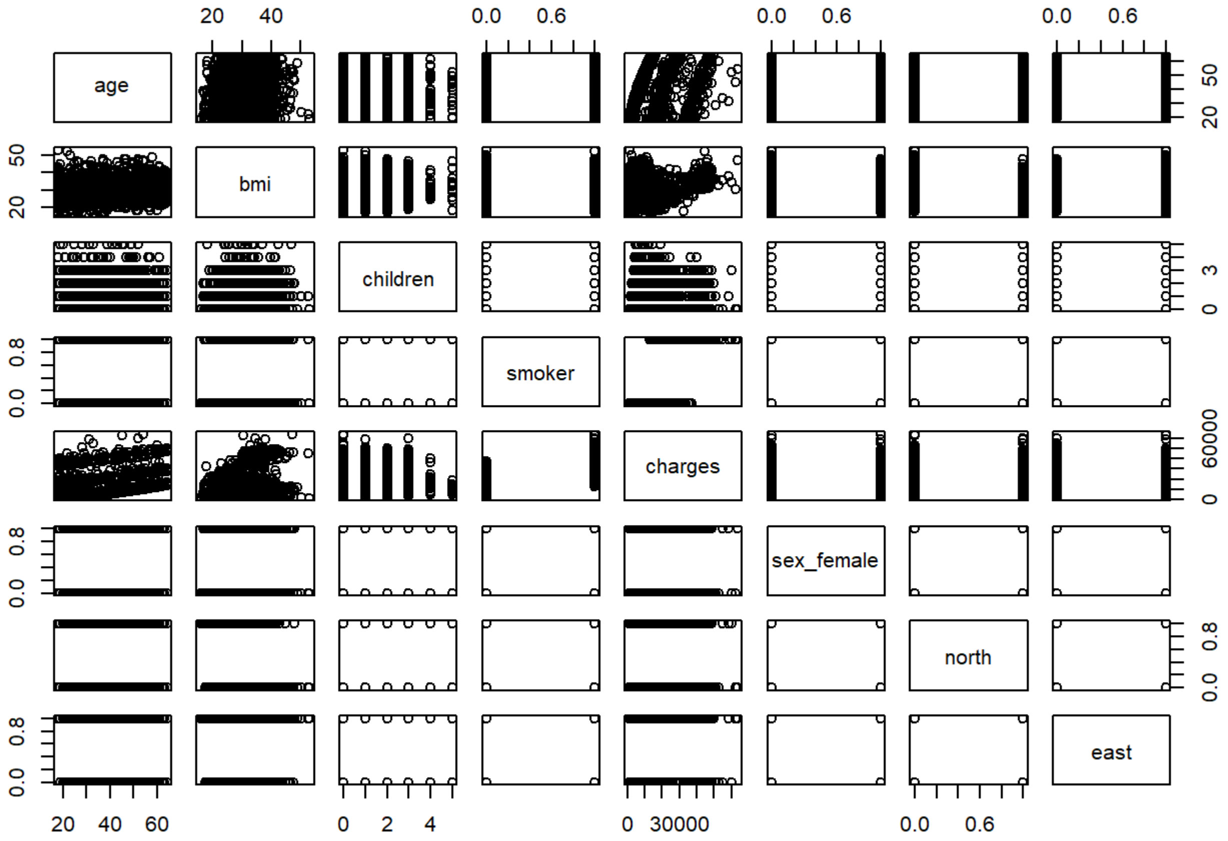
```
##       age             bmi           children         smoker      
##  Min.   :18.00   Min.   :15.96   Min.   :0.000   Min.   :0.0000  
##  1st Qu.:27.00   1st Qu.:26.30   1st Qu.:0.000   1st Qu.:0.0000  
##  Median :39.00   Median :30.40   Median :1.000   Median :0.0000  
##  Mean   :39.21   Mean   :30.66   Mean   :1.095   Mean   :0.2048  
##  3rd Qu.:51.00   3rd Qu.:34.69   3rd Qu.:2.000   3rd Qu.:0.0000  
##  Max.   :64.00   Max.   :53.13   Max.   :5.000   Max.   :1.0000  
##     charges        sex_female         north             east       
##  Min.   : 1122   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.: 4740   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000  
##  Median : 9382   Median :0.0000   Median :0.0000   Median :1.0000  
##  Mean   :13270   Mean   :0.4948   Mean   :0.4851   Mean   :0.5142  
##  3rd Qu.:16640   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000  
##  Max.   :63770   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000  
```

Plots

```
histograms <- function(df) {
  plots <- list()

  for (i in 1:ncol(df)) {
    col <- colnames(df)[i]
    p <- df %>%
      ggplot(aes(!!sym(col))) +
      geom_histogram(bins = 20)
    plots[[i]] <- p
  }

  return(plot_grid(plotlist = plots, nrow = 3))
}

histograms(data)
```


count
100
75
50
25
0
20
30
40
50
60
age
count
150
100
50
0
20
30
40
50
bmi
count
600
400
200
0
0
1
2
3
4
5
children
count
900
600
300
0
0.00
0.25
0.50
0.75
1.00
smoker
count
300
200
100
0
0
20000
40000
60000
charges
count
600
400
200
0
0.00
0.25
0.50
0.75
1.00
sex_female
count
600
400
200
0
0.00
0.25
0.50
0.75
1.00
north
count
600
400
200
0
0.00
0.25
0.50
0.75
1.00
east

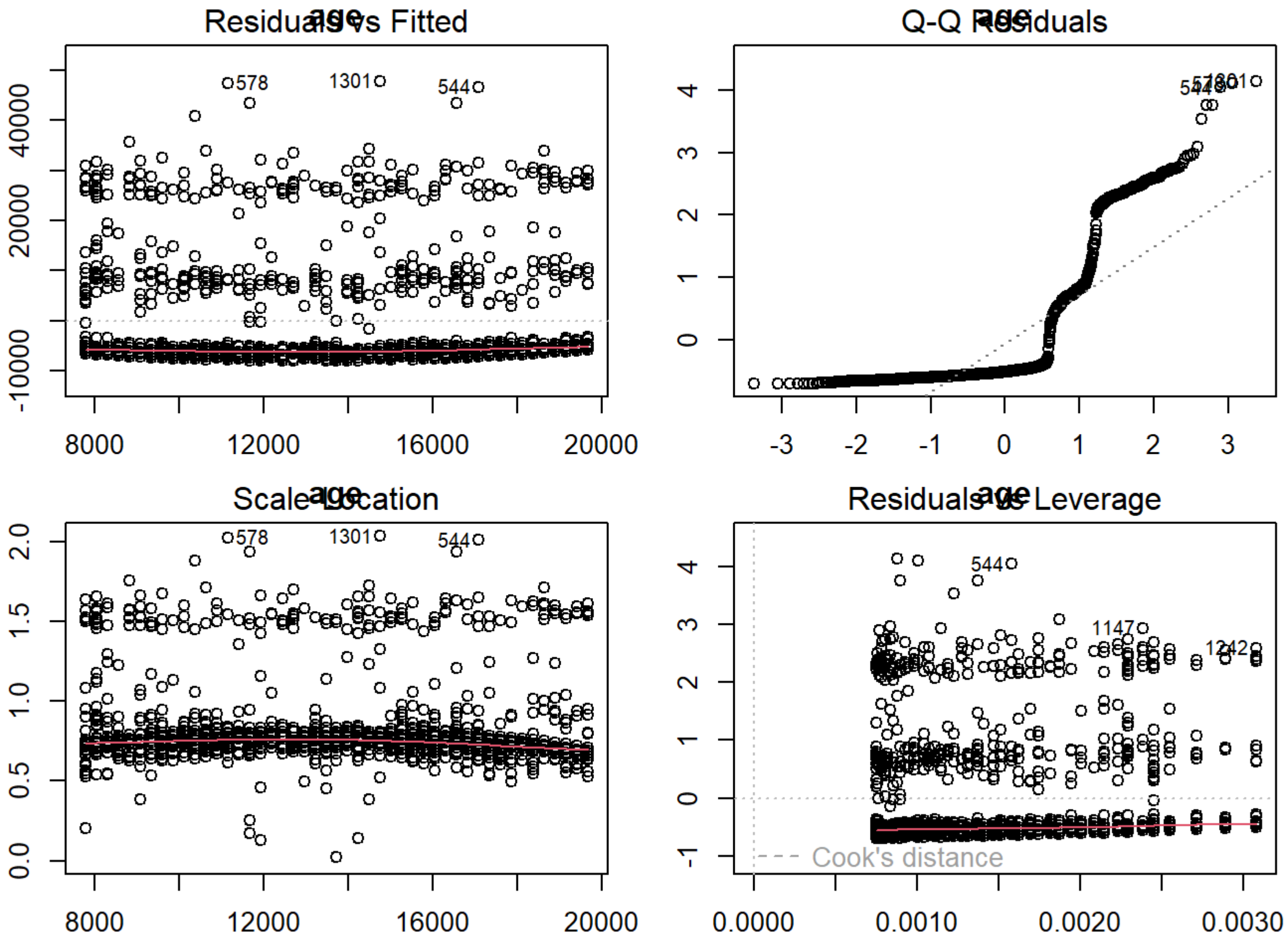
```
pairs(data)
```



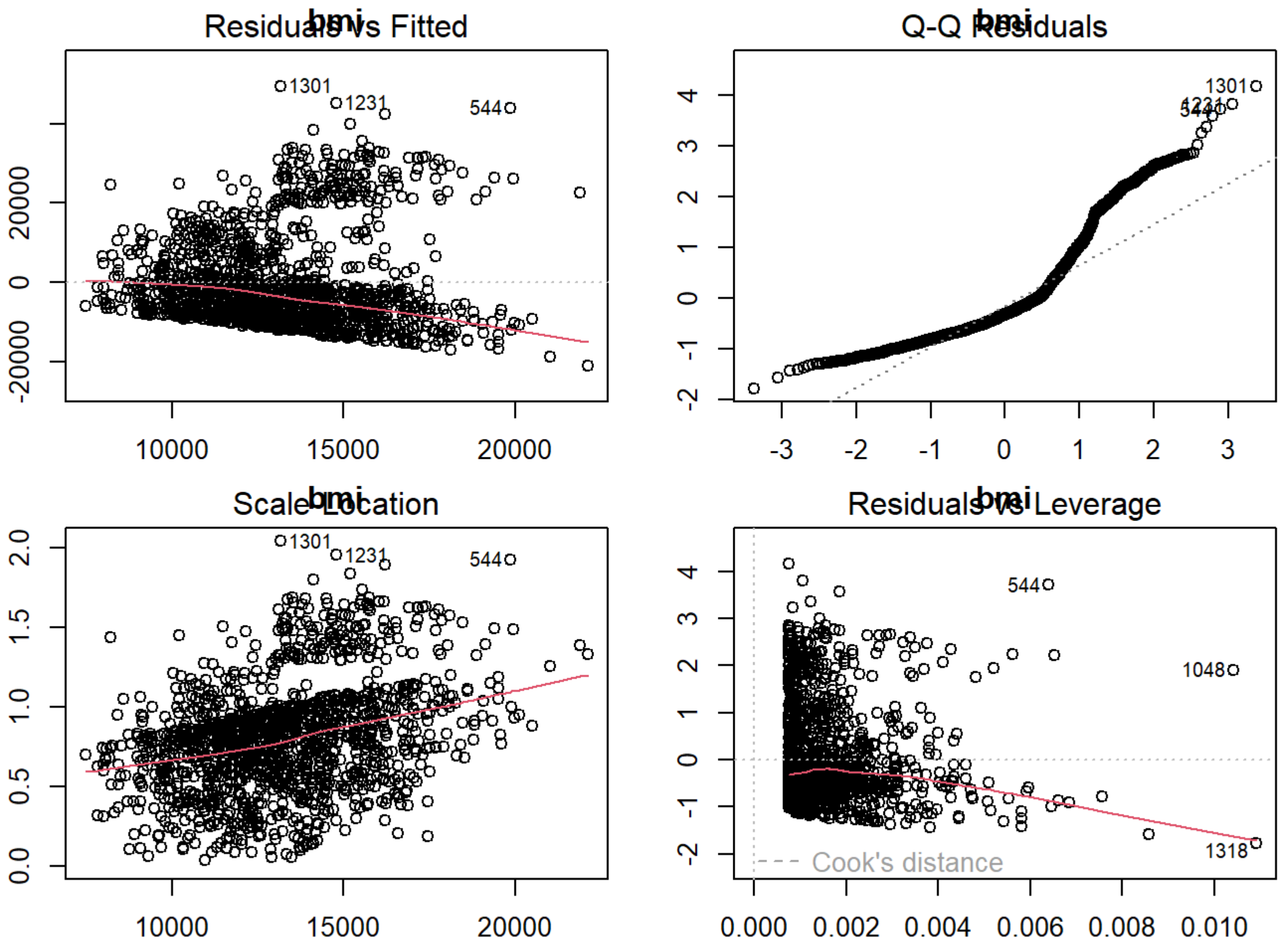
one by one

```
for (predictor in colnames(data)) {
  formula <- as.formula(paste('charges ~',predictor))
  model <- lm(formula, data = data)
  print(summary(model))
  par(mfrow = c(2, 2), mar = c(2,2,2,2))
  plot(model, main = predictor)
}
```

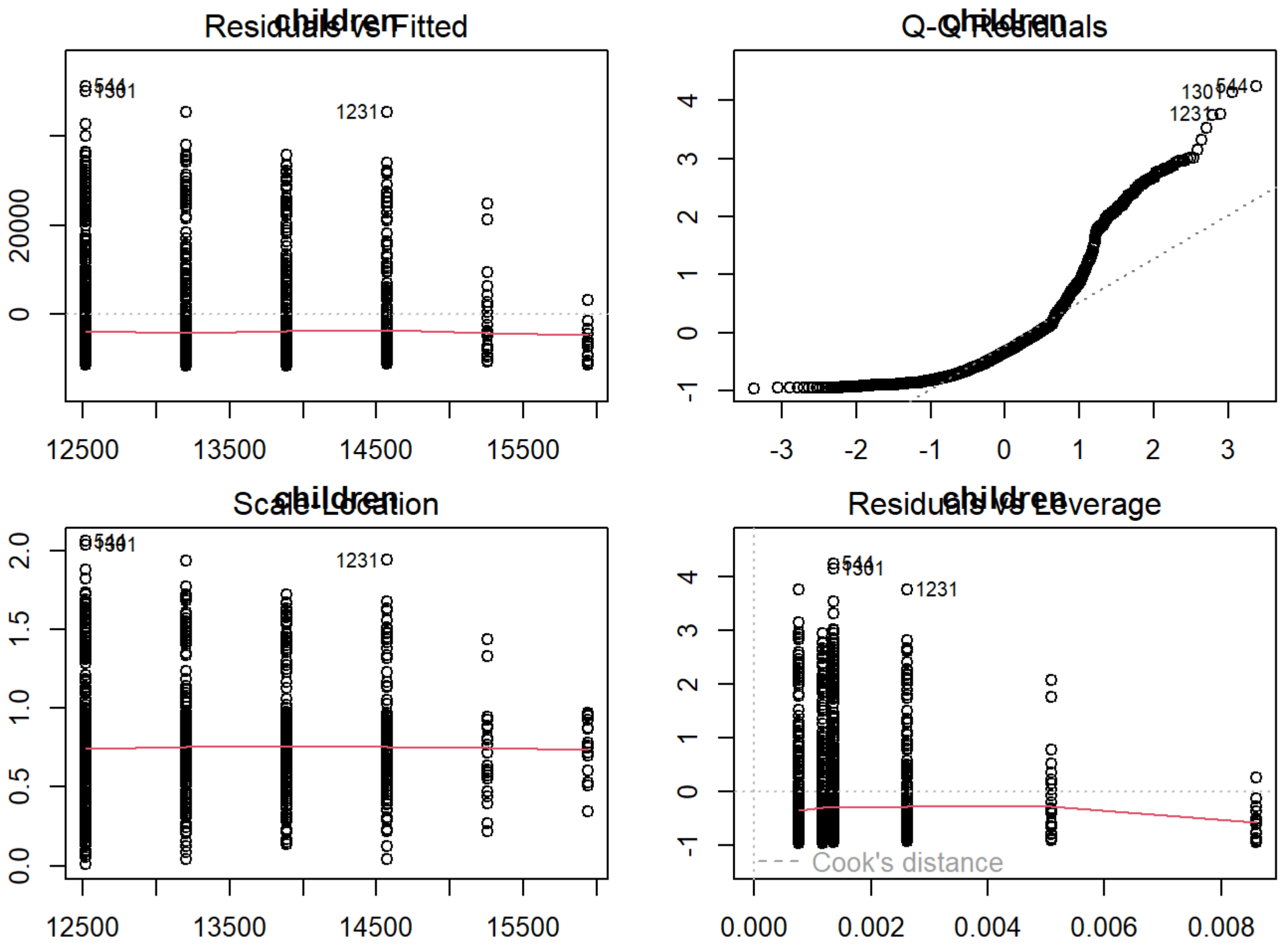
```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
##  -8059  -6671  -5939   5440  47829 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   3165.9      937.1   3.378 0.000751 ***
## age            257.7       22.5  11.453  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11560 on 1336 degrees of freedom
## Multiple R-squared:  0.08941,    Adjusted R-squared:  0.08872 
## F-statistic: 131.2 on 1 and 1336 DF,  p-value: < 2.2e-16
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -20956  -8118  -3757   4722  49442
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1192.94    1664.80   0.717    0.474
## bmi           393.87      53.25   7.397 2.46e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11870 on 1336 degrees of freedom
## Multiple R-squared:  0.03934,    Adjusted R-squared:  0.03862
## F-statistic: 54.71 on 1 and 1336 DF,  p-value: 2.459e-13
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -11585  -8759  -4071   3468  51248
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12522.5      446.5  28.049   <2e-16 ***
## children       683.1      274.2   2.491   0.0129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12090 on 1336 degrees of freedom
## Multiple R-squared:  0.004624,   Adjusted R-squared:  0.003879
## F-statistic: 6.206 on 1 and 1336 DF,  p-value: 0.01285
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -19221  -5042   -919   3705  31720
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8434.3      229.0   36.83   <2e-16 ***
## smoker       23616.0      506.1   46.66   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7470 on 1336 degrees of freedom
## Multiple R-squared:  0.6198, Adjusted R-squared:  0.6195
## F-statistic:  2178 on 1 and 1336 DF,  p-value: < 2.2e-16
```

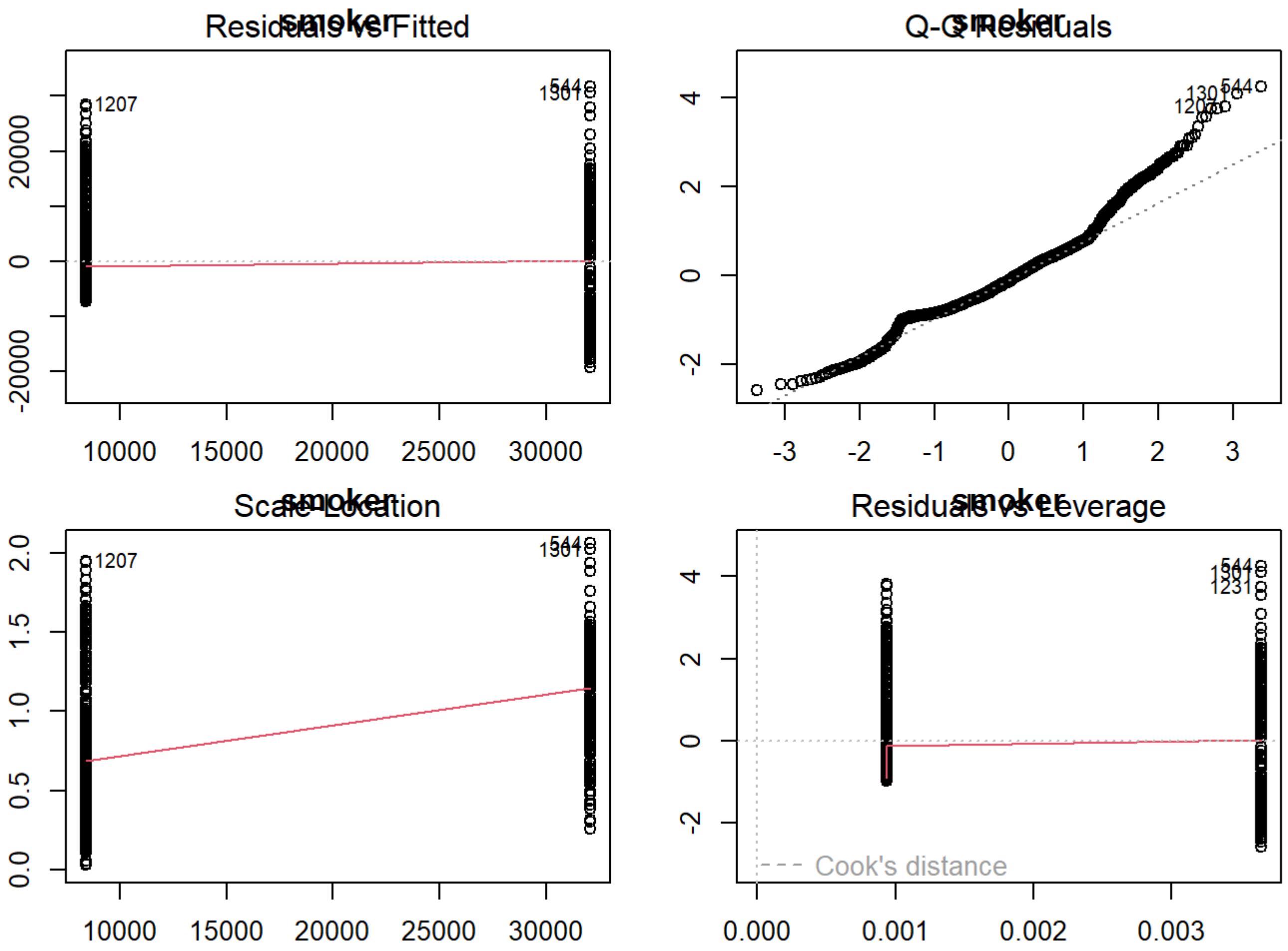
```
## Warning in model.matrix.default(mt, mf, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(mt, mf, contrasts): problem with term 1 in
## model.matrix: no columns are assigned
```

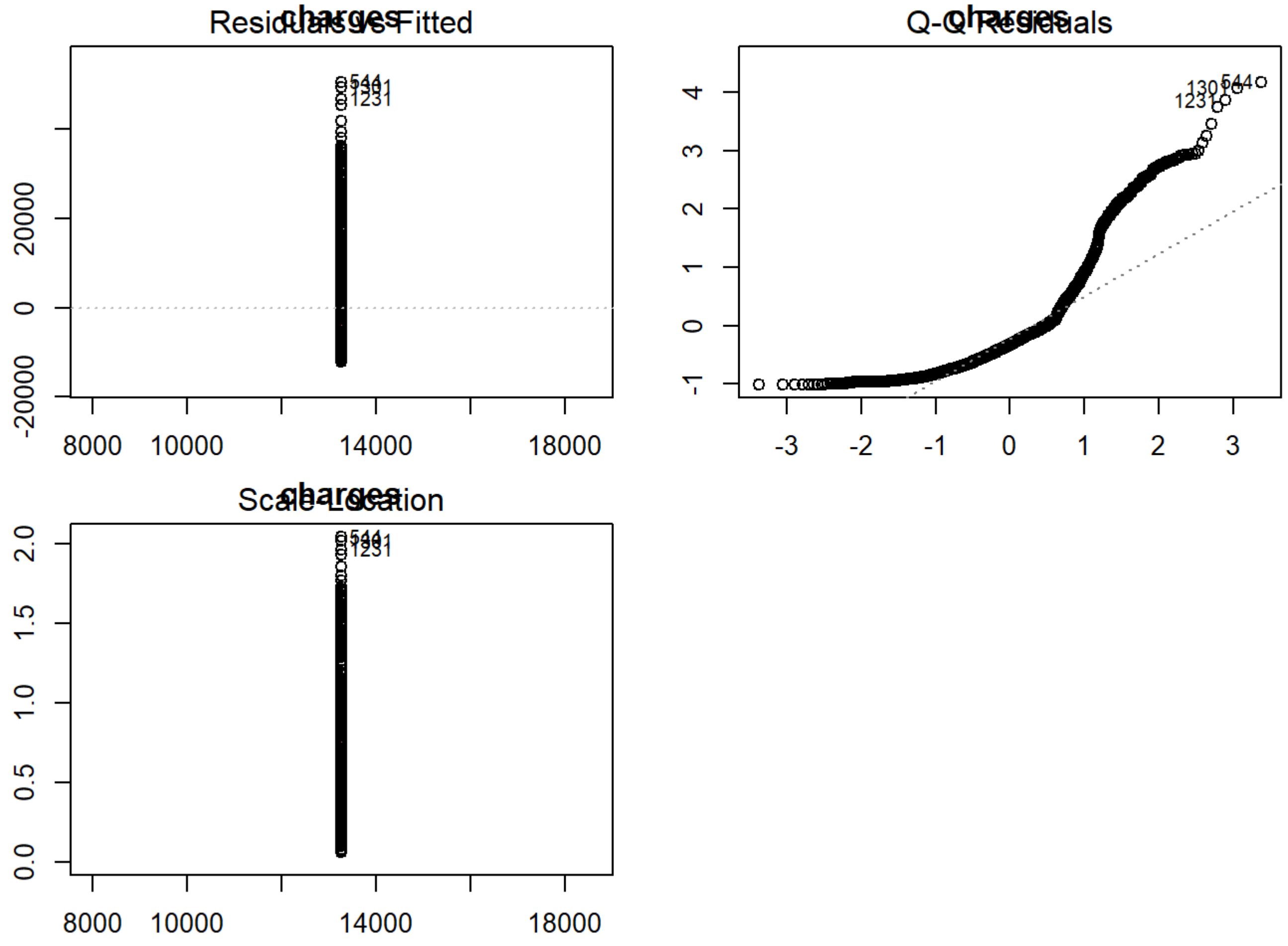
```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -12149  -8530  -3888   3369  50500
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13270.4      331.1   40.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12110 on 1337 degrees of freedom
```

```
## Warning in model.matrix.default(object, data = structure(list(charges =
## c(16884.924, : the response appeared on the right-hand side and was dropped
```

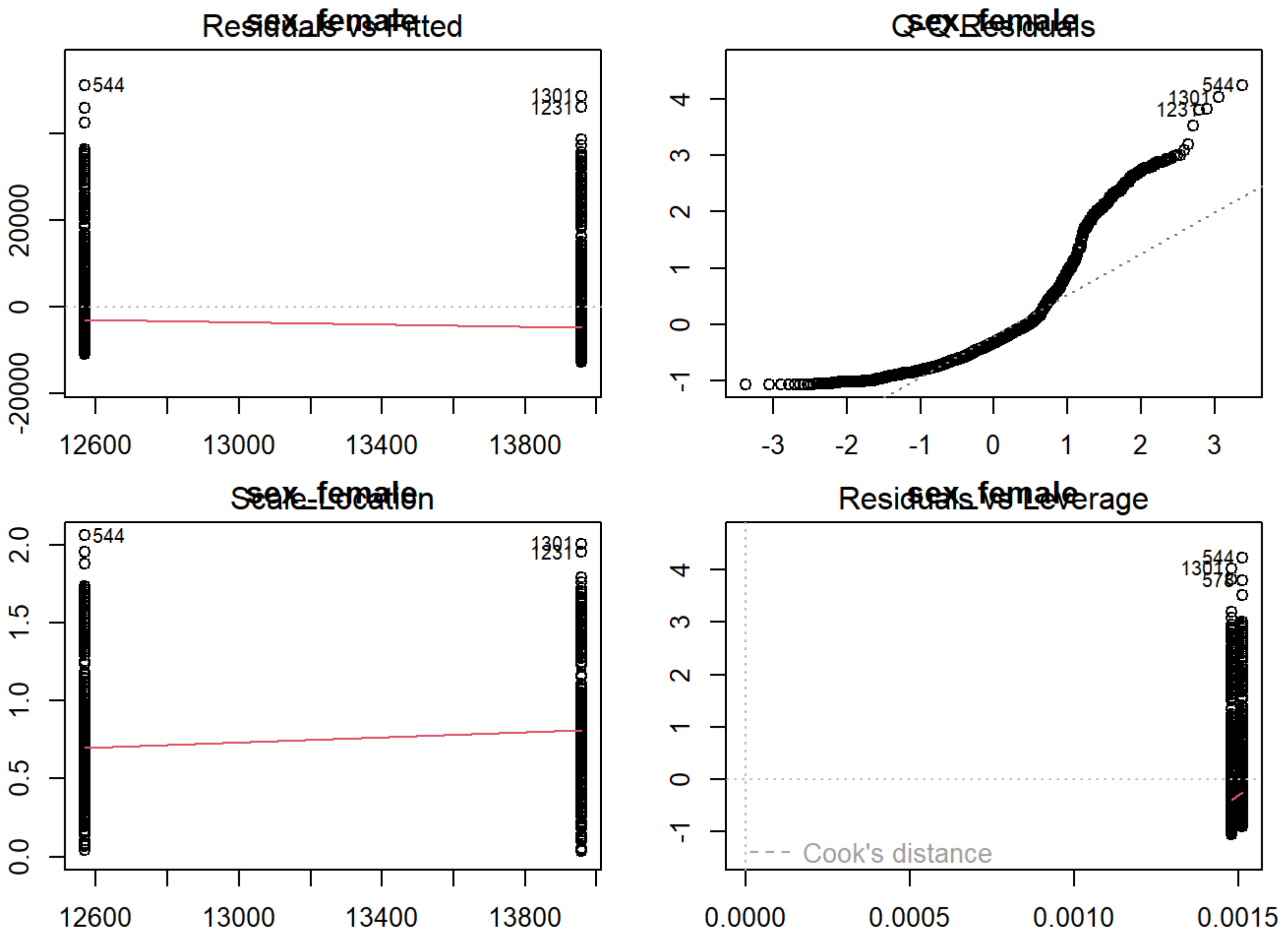
```
## Warning in model.matrix.default(object, data = structure(list(charges =
## c(16884.924, : problem with term 1 in model.matrix: no columns are assigned
```



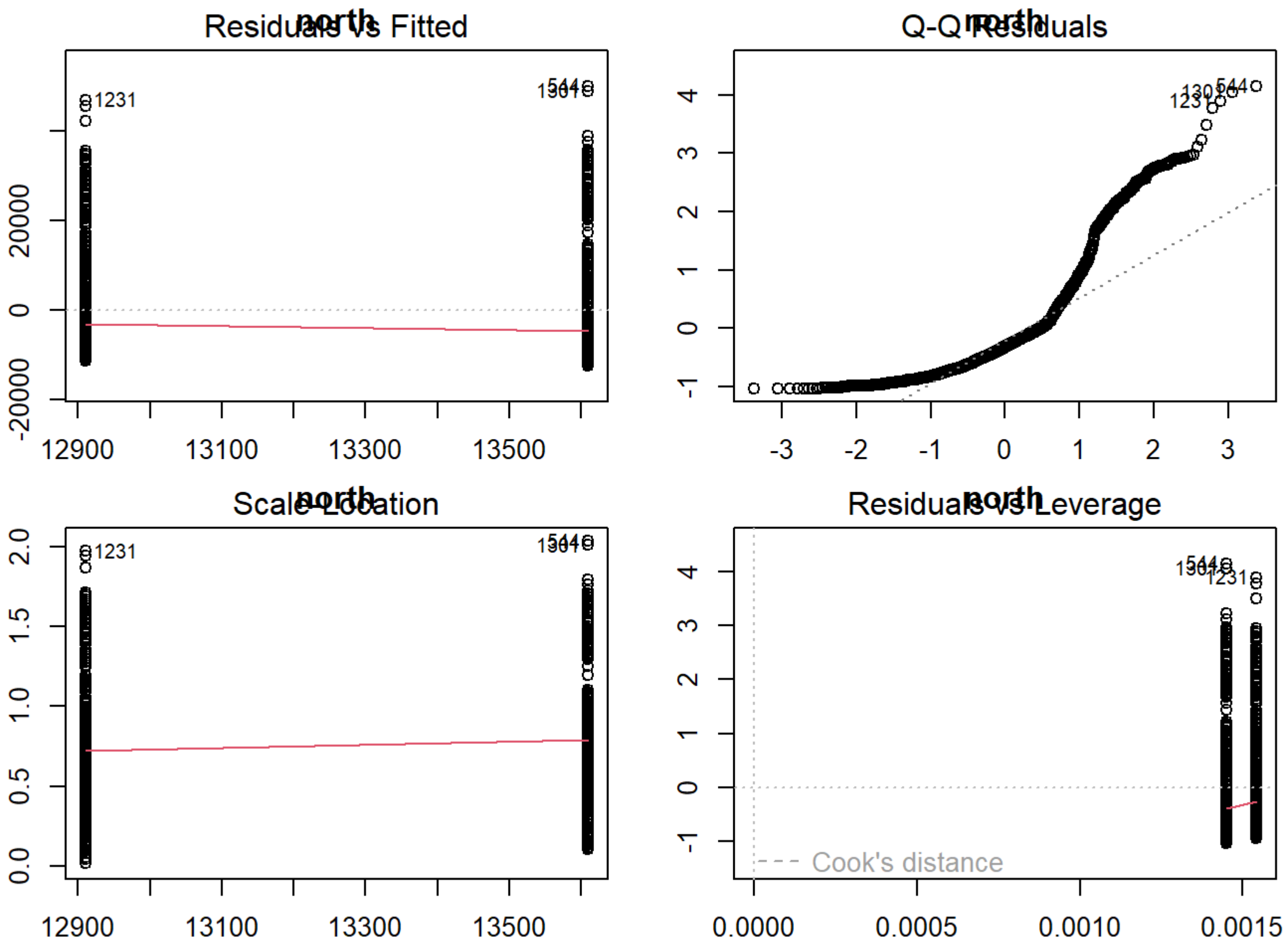
```
## hat values (leverages) are all = 0.0007473842
##  and there are no factor predictors; no plot no. 5
```



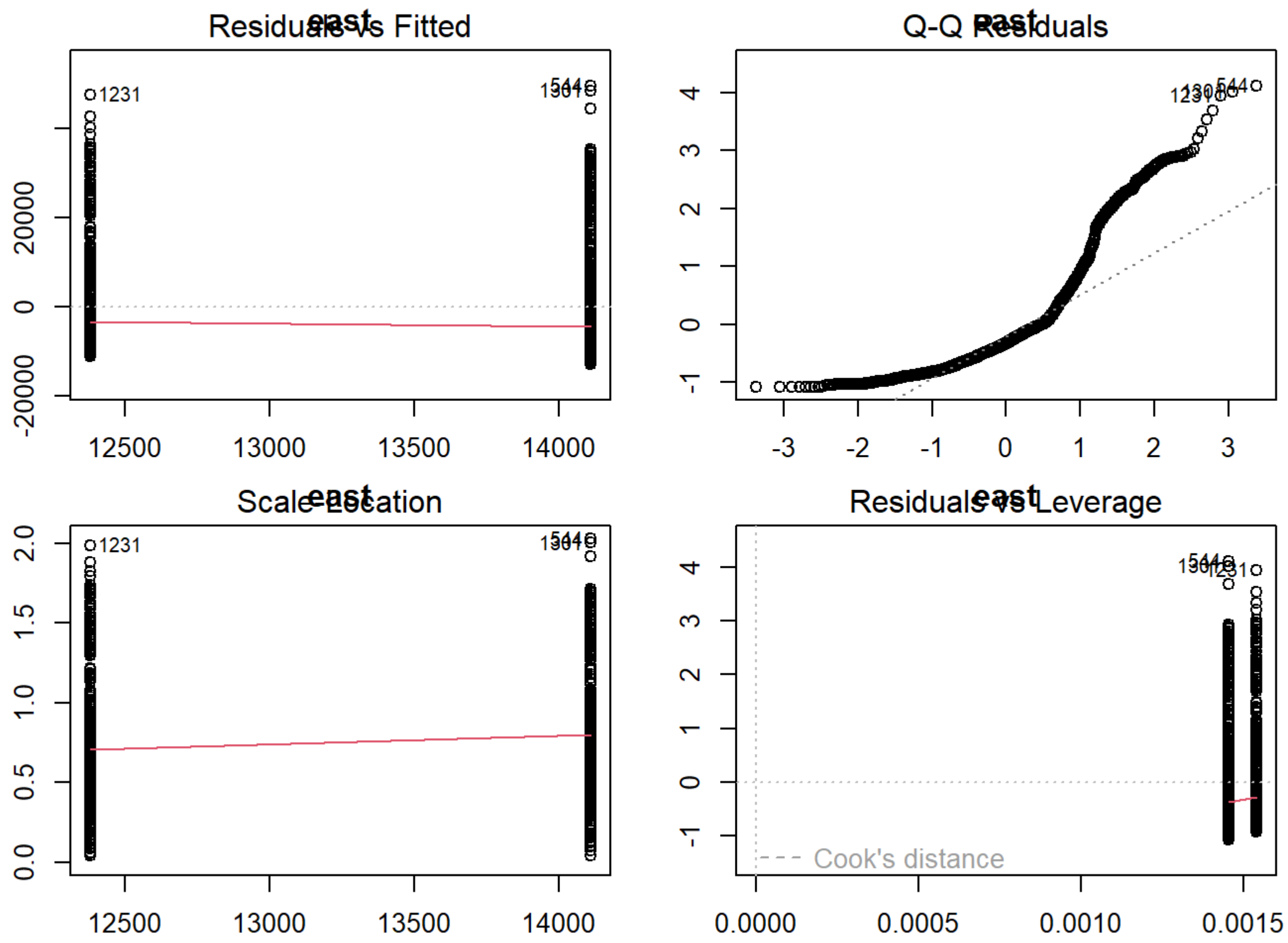
```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -12835  -8435  -3980   3476  51201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13956.8      465.2  30.003   <2e-16 ***
## sex_female   -1387.2      661.3  -2.098   0.0361 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12090 on 1336 degrees of freedom
## Multiple R-squared:  0.003282,	Adjusted R-squared:  0.002536
## F-statistic:   4.4 on 1 and 1336 DF,  p-value: 0.03613
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -12487  -8478  -3872   3475  50162
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13608.8      461.3  29.499   <2e-16 ***
## north         -697.6      662.4  -1.053    0.293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12110 on 1336 degrees of freedom
## Multiple R-squared:  0.0008294,  Adjusted R-squared:  8.148e-05
## F-statistic: 1.109 on 1 and 1336 DF,  p-value: 0.2925
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -12988  -8422  -3918   3384  49661
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    12382        474  26.125  < 2e-16 ***
## east            1727        661   2.613  0.00907 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12080 on 1336 degrees of freedom
## Multiple R-squared:  0.005086,   Adjusted R-squared:  0.004341
## F-statistic: 6.829 on 1 and 1336 DF,  p-value: 0.009069
```

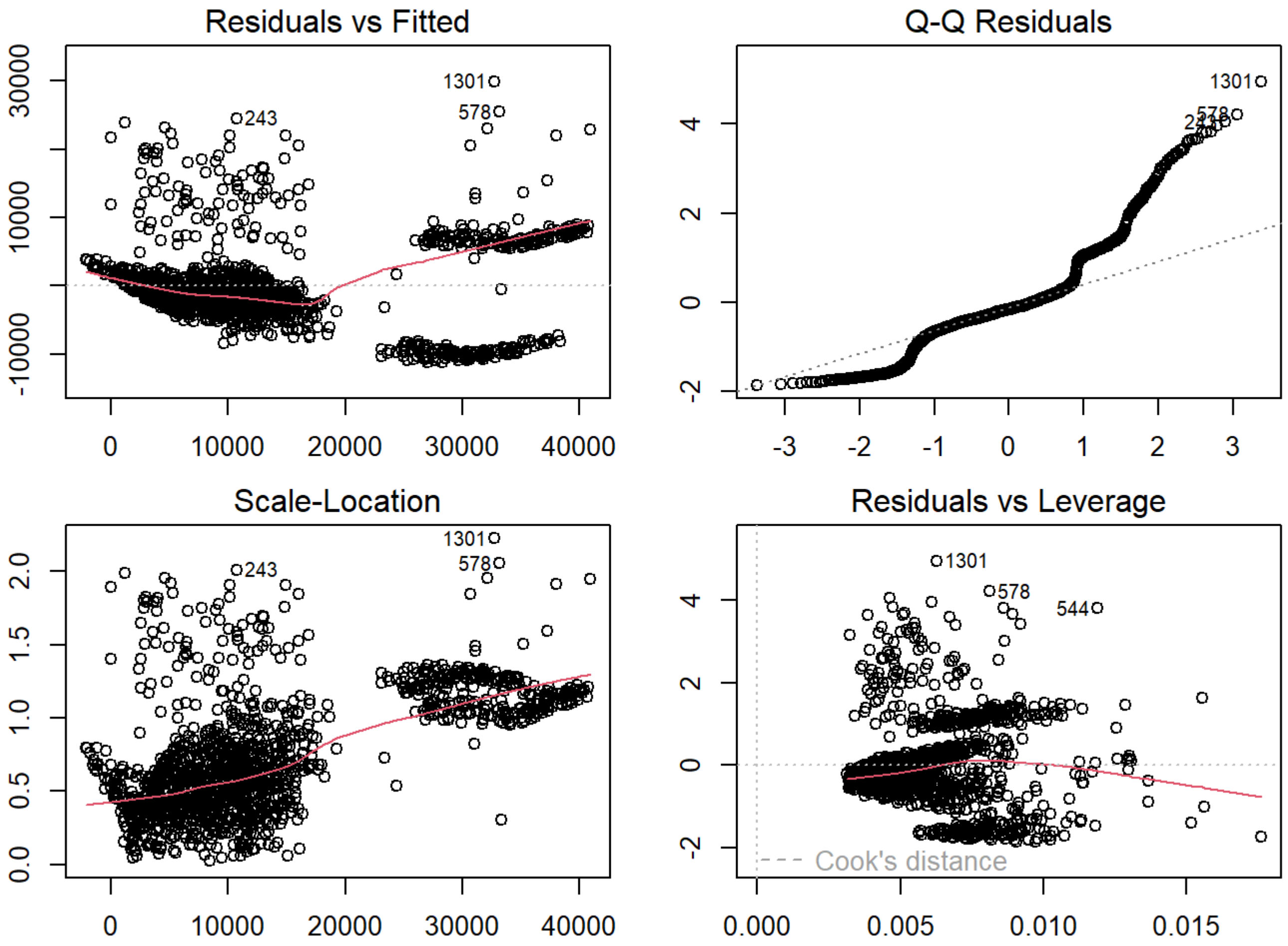


Model

```
model <- lm(charges ~ ., data = data)
summary(model)
```

```
## 
## Call:
## lm(formula = charges ~ ., data = data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -11215.7  -2829.4   -981.2   1382.0  29893.7 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -13076.08    1030.21 -12.693  < 2e-16 ***
## age            257.04      11.89  21.613  < 2e-16 ***
## bmi            336.99      28.38  11.872  < 2e-16 ***
## children       475.46     137.77   3.451 0.000576 ***
## smoker       23841.17     412.90  57.741  < 2e-16 ***
## sex_female     130.69     332.87   0.393 0.694656    
## north          820.37     341.37   2.403 0.016390 *  
## east           136.44     335.05   0.407 0.683900    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 6061 on 1330 degrees of freedom
## Multiple R-squared:  0.7508, Adjusted R-squared:  0.7495 
## F-statistic: 572.6 on 7 and 1330 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2, 2), mar = c(2,2,2,2))
plot(model)
```



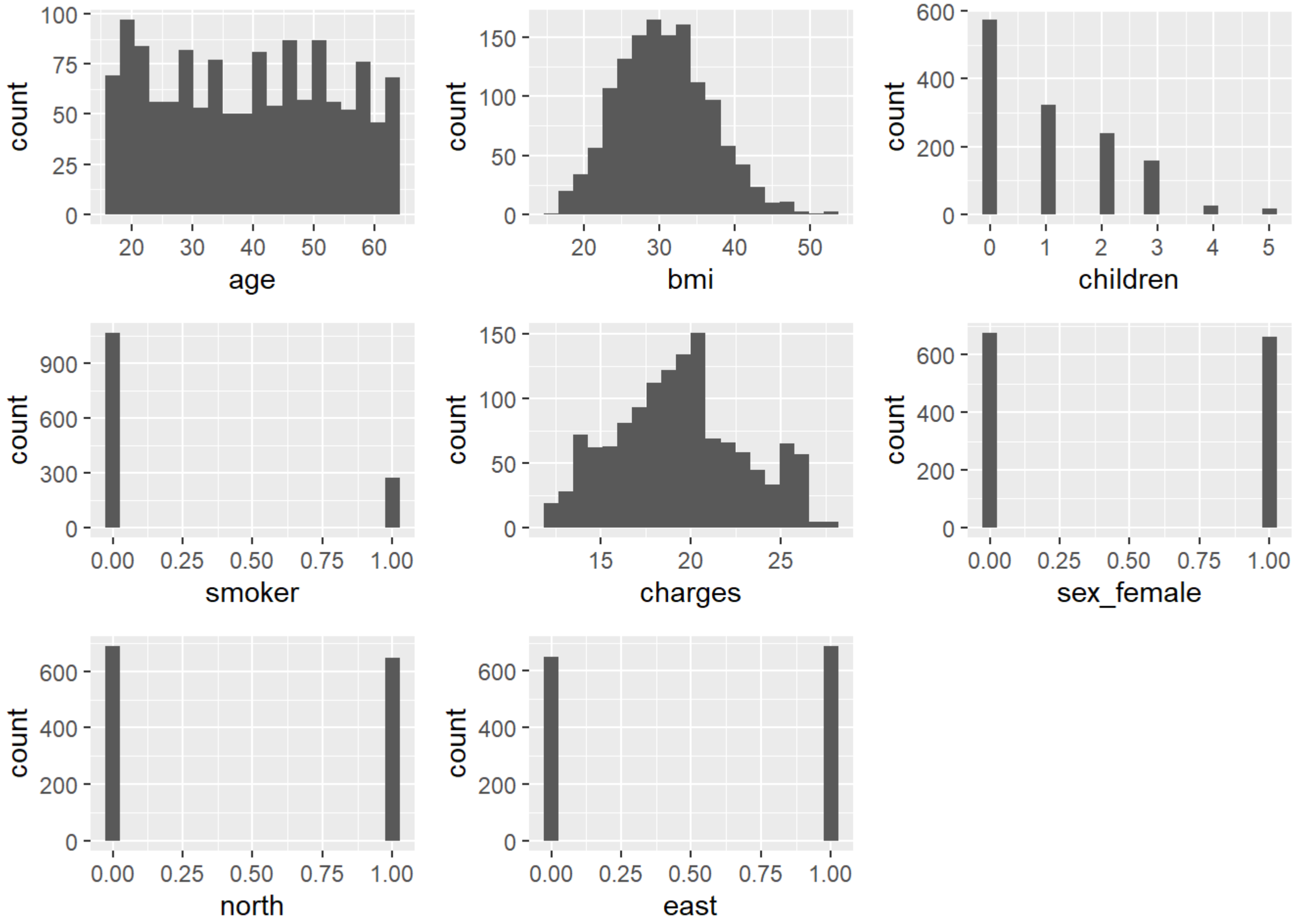
Power Transform

```
transformation <- powerTransform(model)
print(transformation)
```

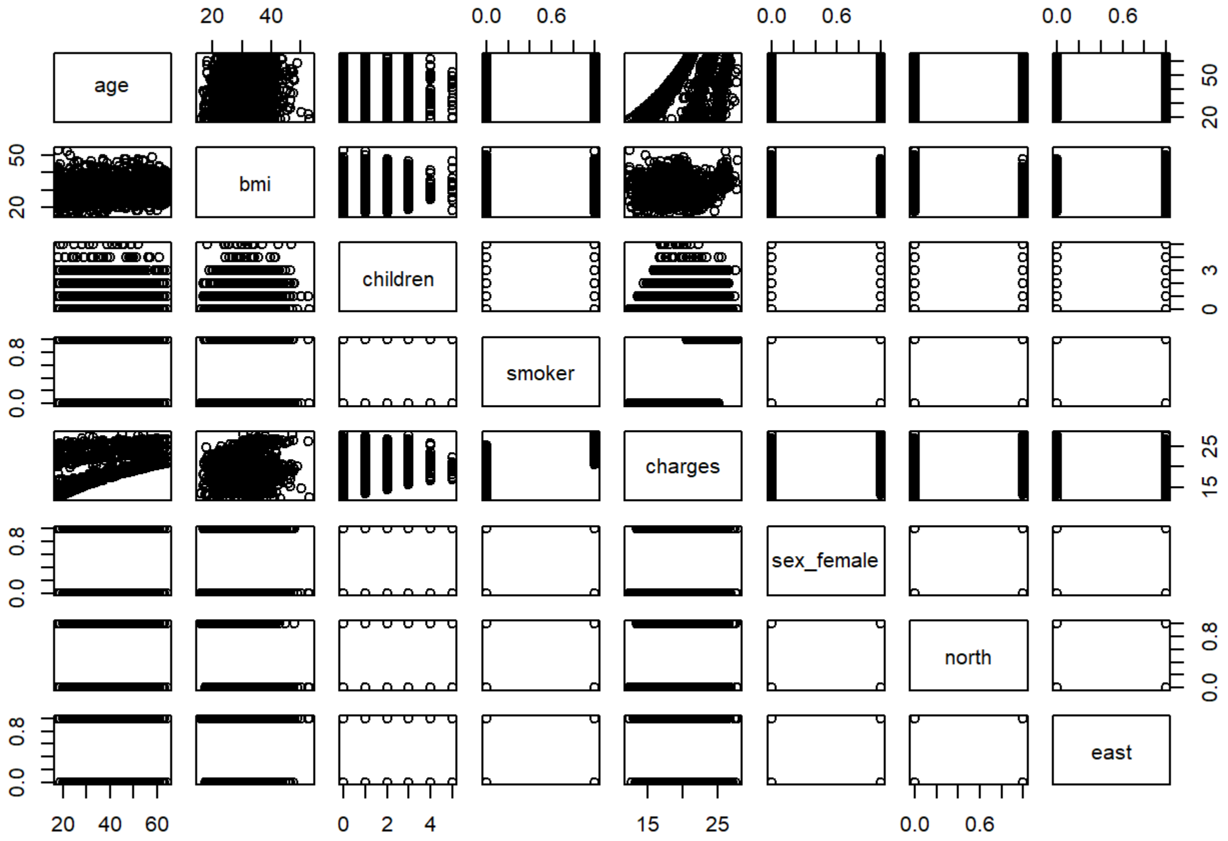
```
## Estimated transformation parameter
##        Y1
## 0.1473381
```

Apply

```
charges_transformed <- bcPower(data$charges, transformation$lambda)

data <- data %>%
  mutate(charges = charges_transformed)

histograms(data)
```


count
age
20 30 40 50 60
0 25 50 75 100
bmi
20 30 40 50
0 50 100 150
children
0 1 2 3 4 5
0 200 400 600
smoker
0.00 0.25 0.50 0.75 1.00
0 300 600 900
charges
15 20 25
0 50 100 150
sex_female
0.00 0.25 0.50 0.75 1.00
0 200 400 600
north
0.00 0.25 0.50 0.75 1.00
0 200 400 600
east
0.00 0.25 0.50 0.75 1.00
0 200 400 600

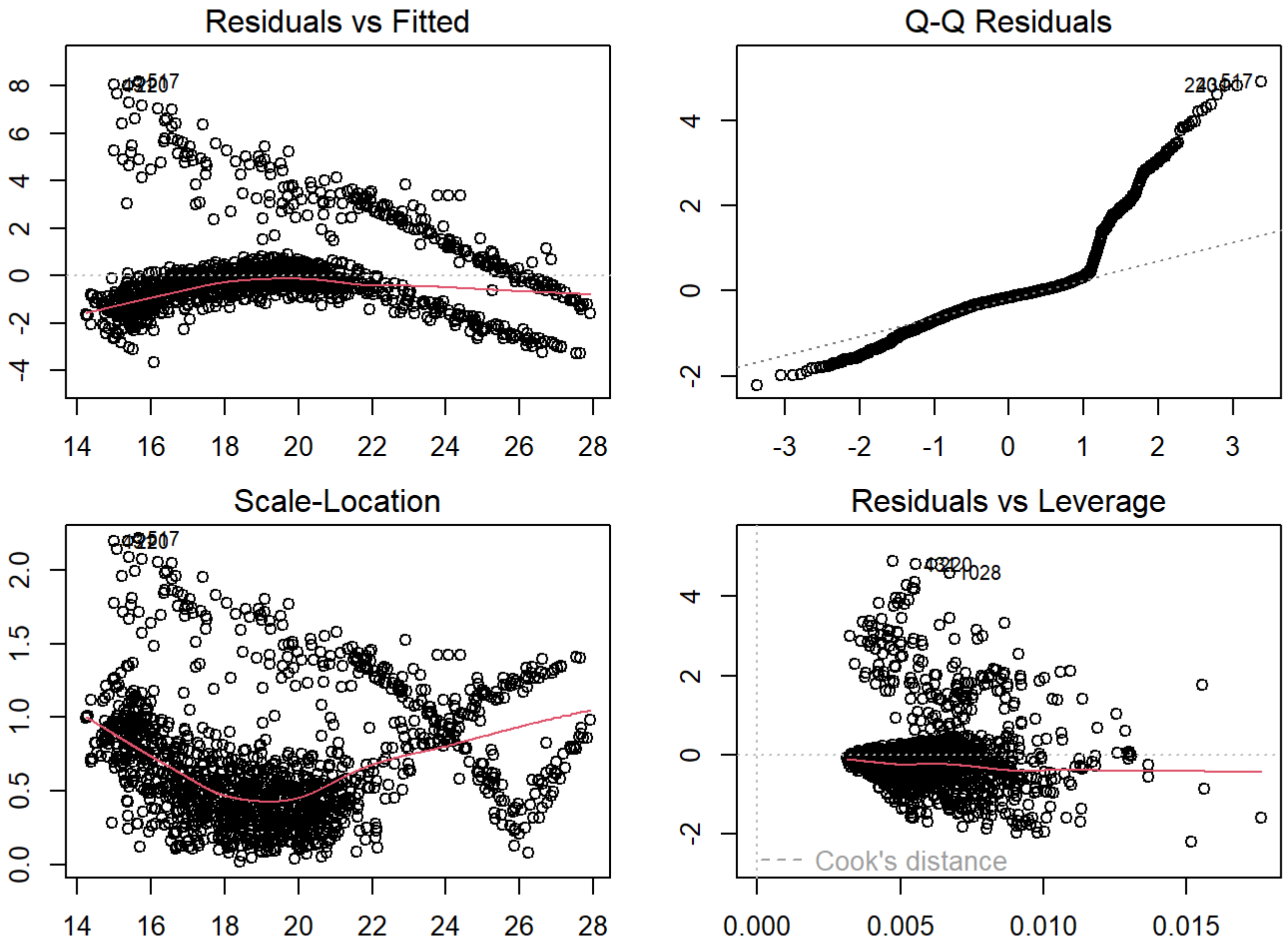
```
pairs(data)
```



```
model <- lm(charges ~ ., data = data)
summary(model)
```

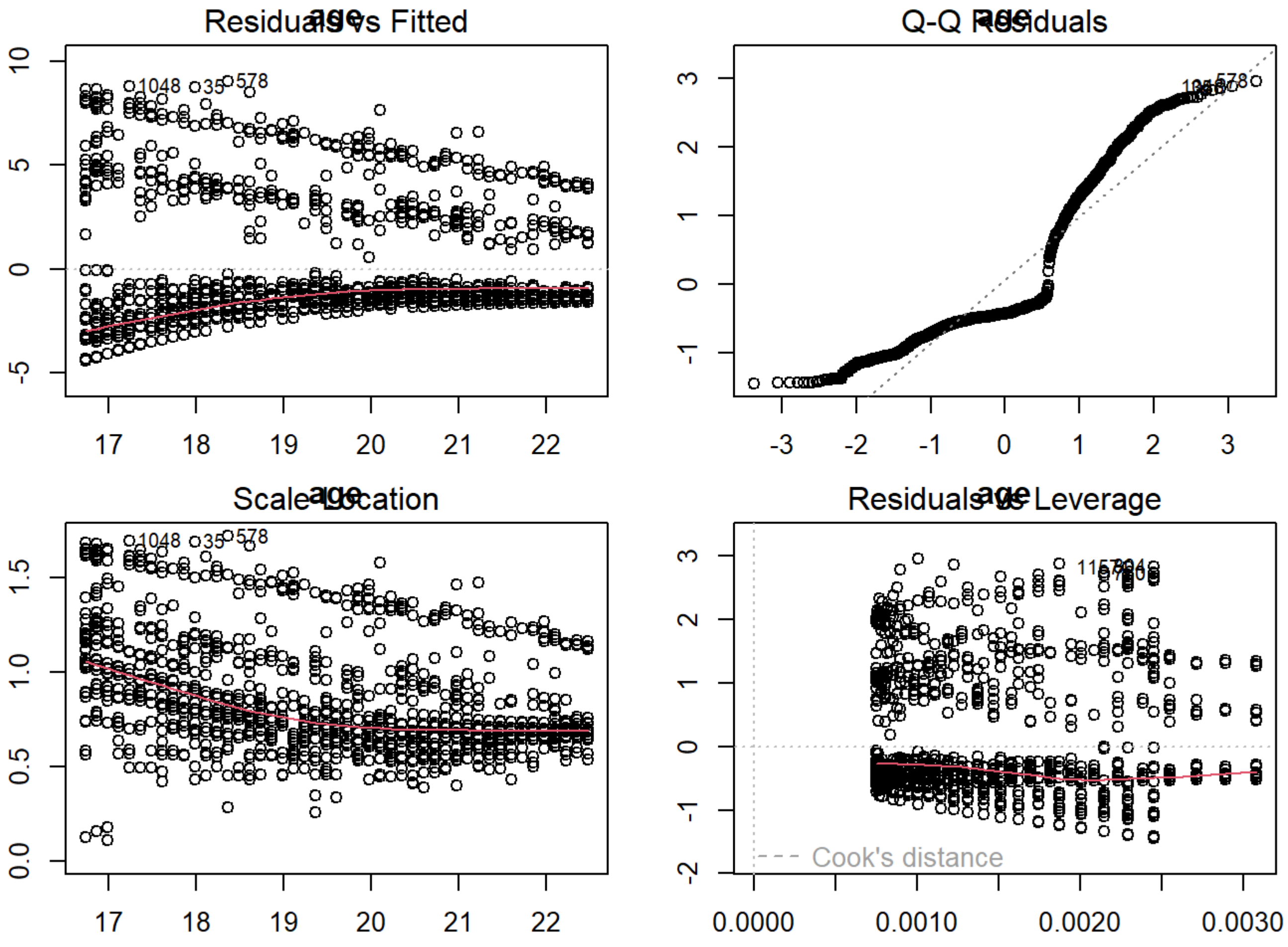
```
##
## Call:
## lm(formula = charges ~ ., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6692 -0.8074 -0.2629  0.1876  8.1868
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.751135   0.284637  37.771  < 2e-16 ***
## age          0.124808   0.003286  37.983  < 2e-16 ***
## bmi          0.056868   0.007842   7.251 6.98e-13 ***
## children     0.343873   0.038065   9.034  < 2e-16 ***
## smoker       6.232202   0.114080  54.630  < 2e-16 ***
## sex_female   0.241518   0.091968   2.626  0.00874 **
## north        0.393816   0.094318   4.175 3.17e-05 ***
## east         0.066824   0.092570   0.722  0.47050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.675 on 1330 degrees of freedom
## Multiple R-squared:  0.7759, Adjusted R-squared:  0.7748
## F-statistic:   658 on 7 and 1330 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2, 2), mar = c(2,2,2,2))
plot(model)
```

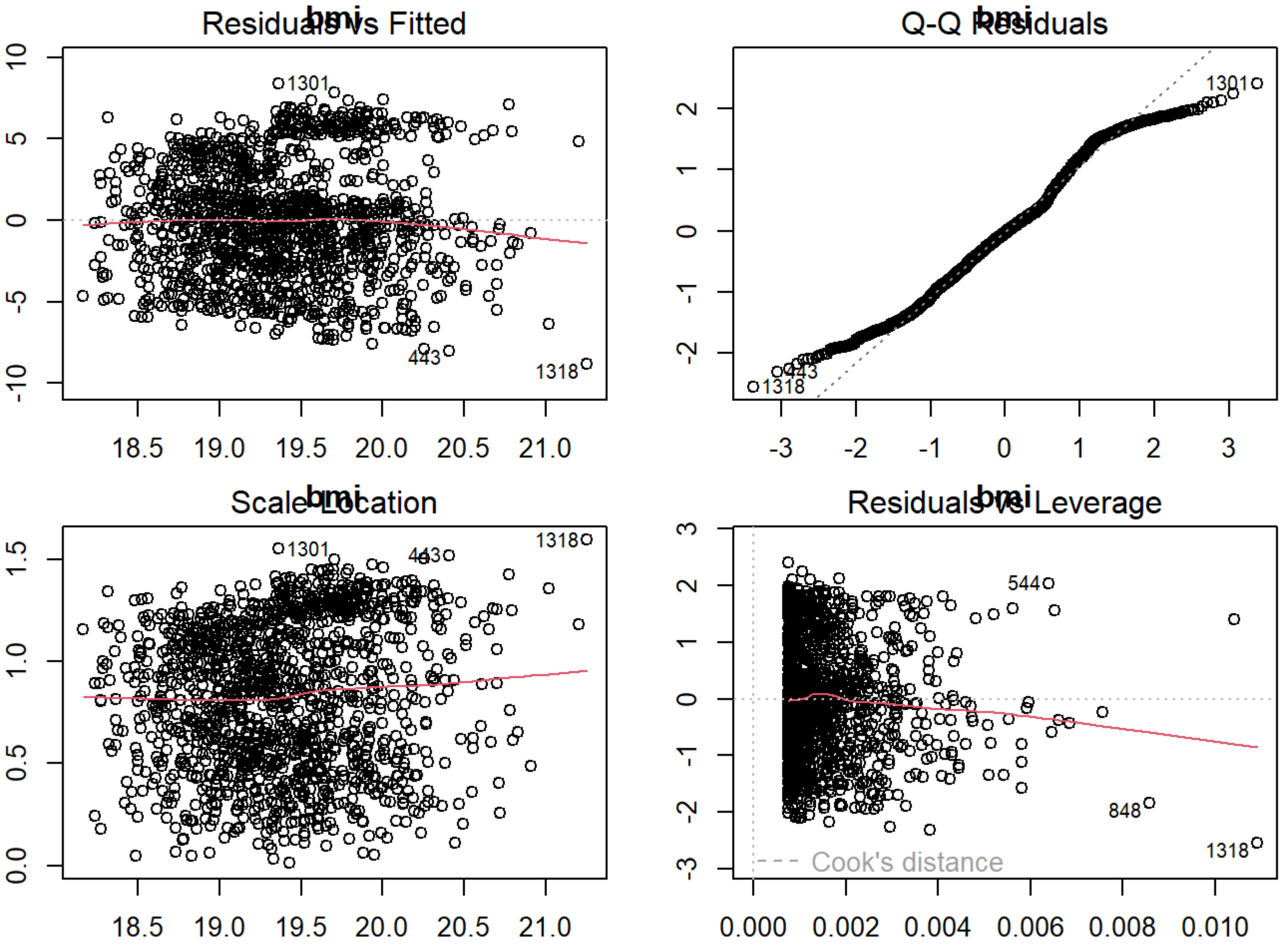


```
for (predictor in colnames(data)) {
  formula <- as.formula(paste('charges ~',predictor))
  model <- lm(formula, data = data)
  print(summary(model))
  par(mfrow = c(2, 2), mar = c(2,2,2,2))
  plot(model, main = predictor)
}
```

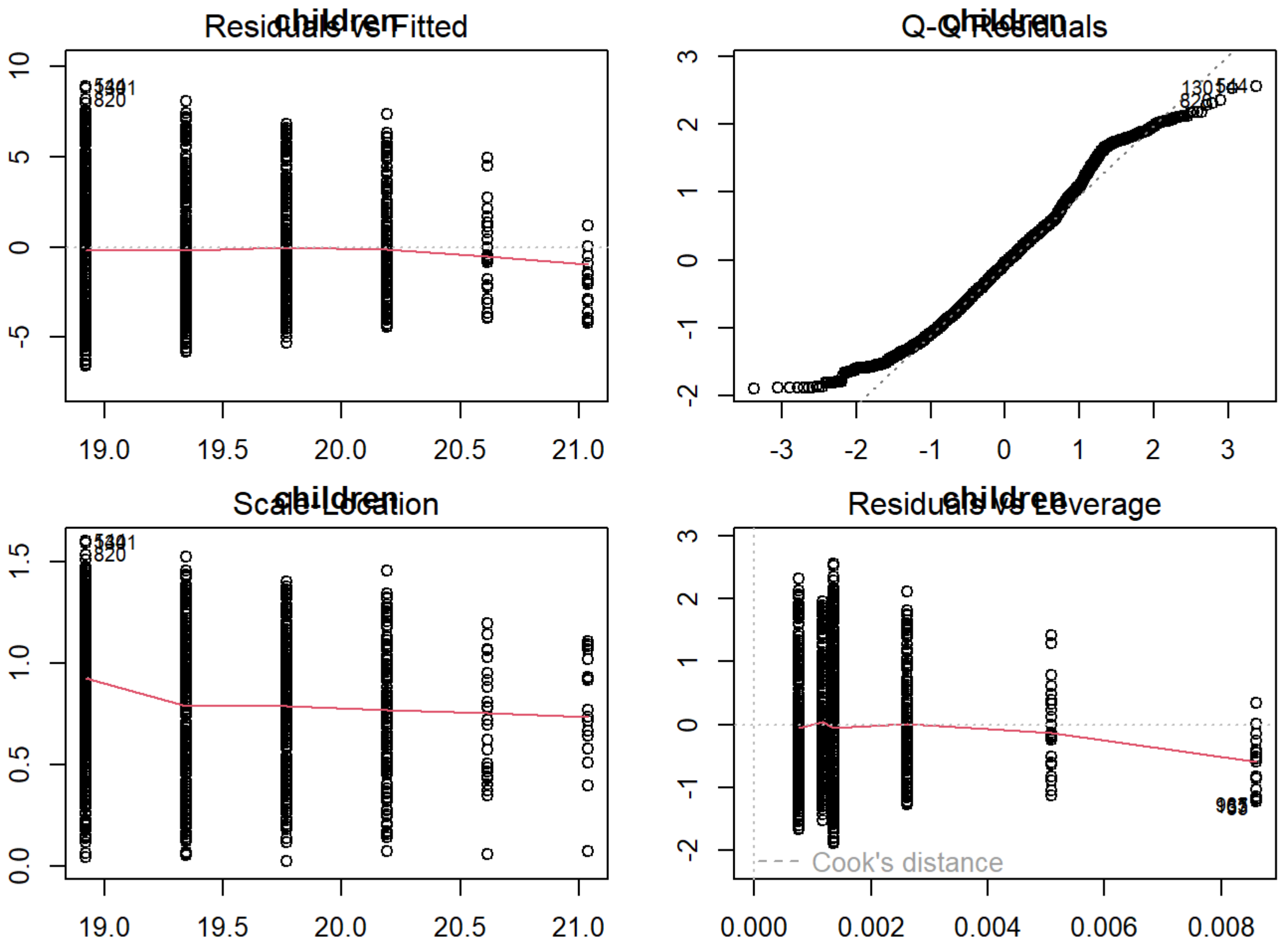
```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.433 -1.739 -1.321  2.088  9.058
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.506092   0.248526   58.37   <2e-16 ***
## age          0.124463   0.005967   20.86   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.066 on 1336 degrees of freedom
## Multiple R-squared:  0.2456, Adjusted R-squared:  0.2451
## F-statistic:   435 on 1 and 1336 DF,  p-value: < 2.2e-16
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.8309 -2.5413 -0.1159  2.5097  8.3980
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.84504    0.48979  34.392  < 2e-16 ***
## bmi          0.08286    0.01567   5.289 1.43e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.493 on 1336 degrees of freedom
## Multiple R-squared:  0.02051,    Adjusted R-squared:  0.01978
## F-statistic: 27.98 on 1 and 1336 DF,  p-value: 1.434e-07
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6090 -2.7294 -0.1486  2.1253  8.9307
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.92299    0.12901 146.675  < 2e-16 ***
## children     0.42279    0.07924   5.336 1.12e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.493 on 1336 degrees of freedom
## Multiple R-squared:  0.02087,    Adjusted R-squared:  0.02013
## F-statistic: 28.47 on 1 and 1336 DF,  p-value: 1.117e-07
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8222 -1.8367  0.3504  1.8138  7.0363
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.13616    0.07746  234.12   <2e-16 ***
## smoker       6.10279    0.17118   35.65   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.527 on 1336 degrees of freedom
## Multiple R-squared:  0.4875, Adjusted R-squared:  0.4872
## F-statistic:  1271 on 1 and 1336 DF,  p-value: < 2.2e-16
```

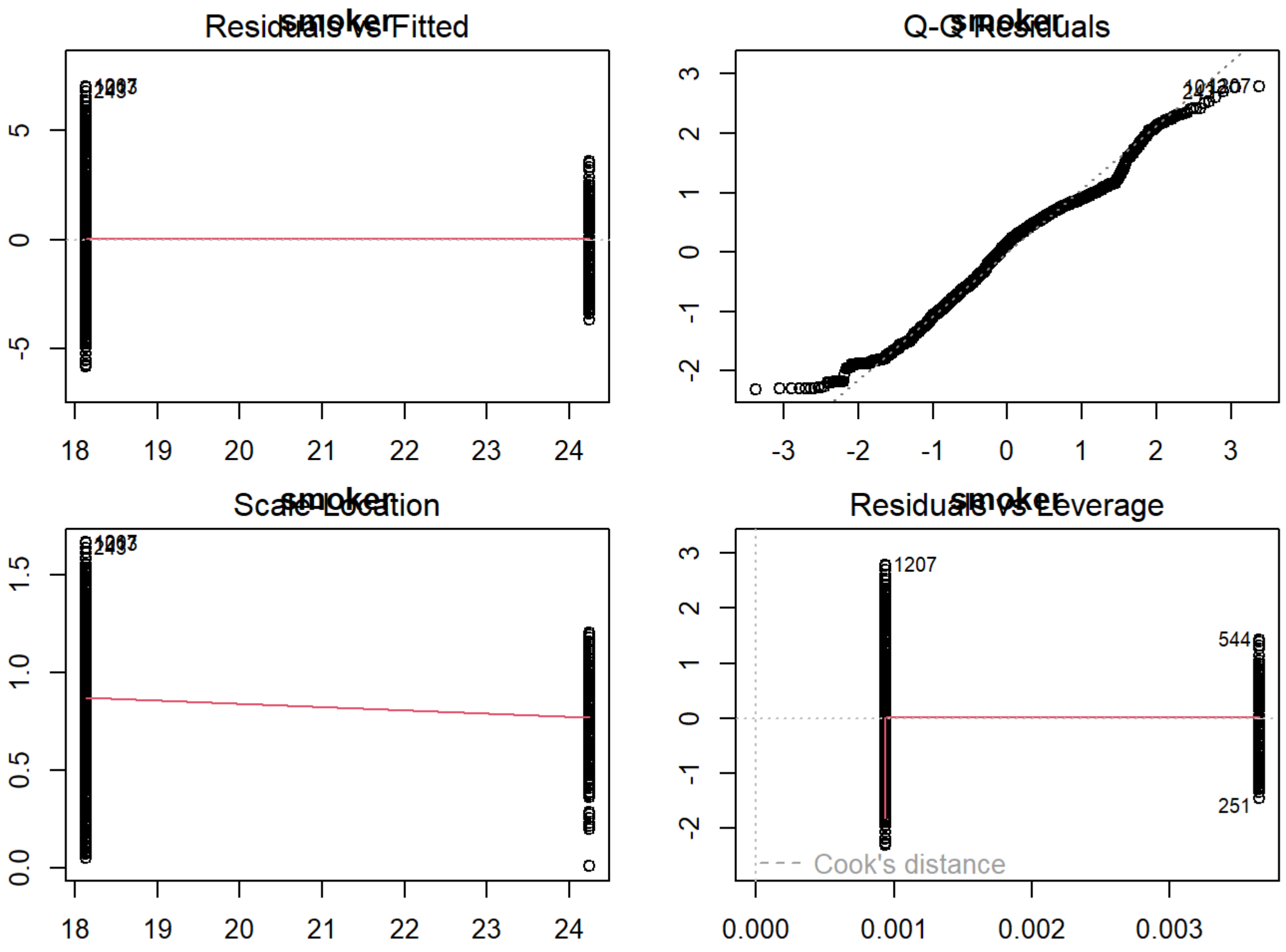
```
## Warning in model.matrix.default(mt, mf, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(mt, mf, contrasts): problem with term 1 in
## model.matrix: no columns are assigned
```

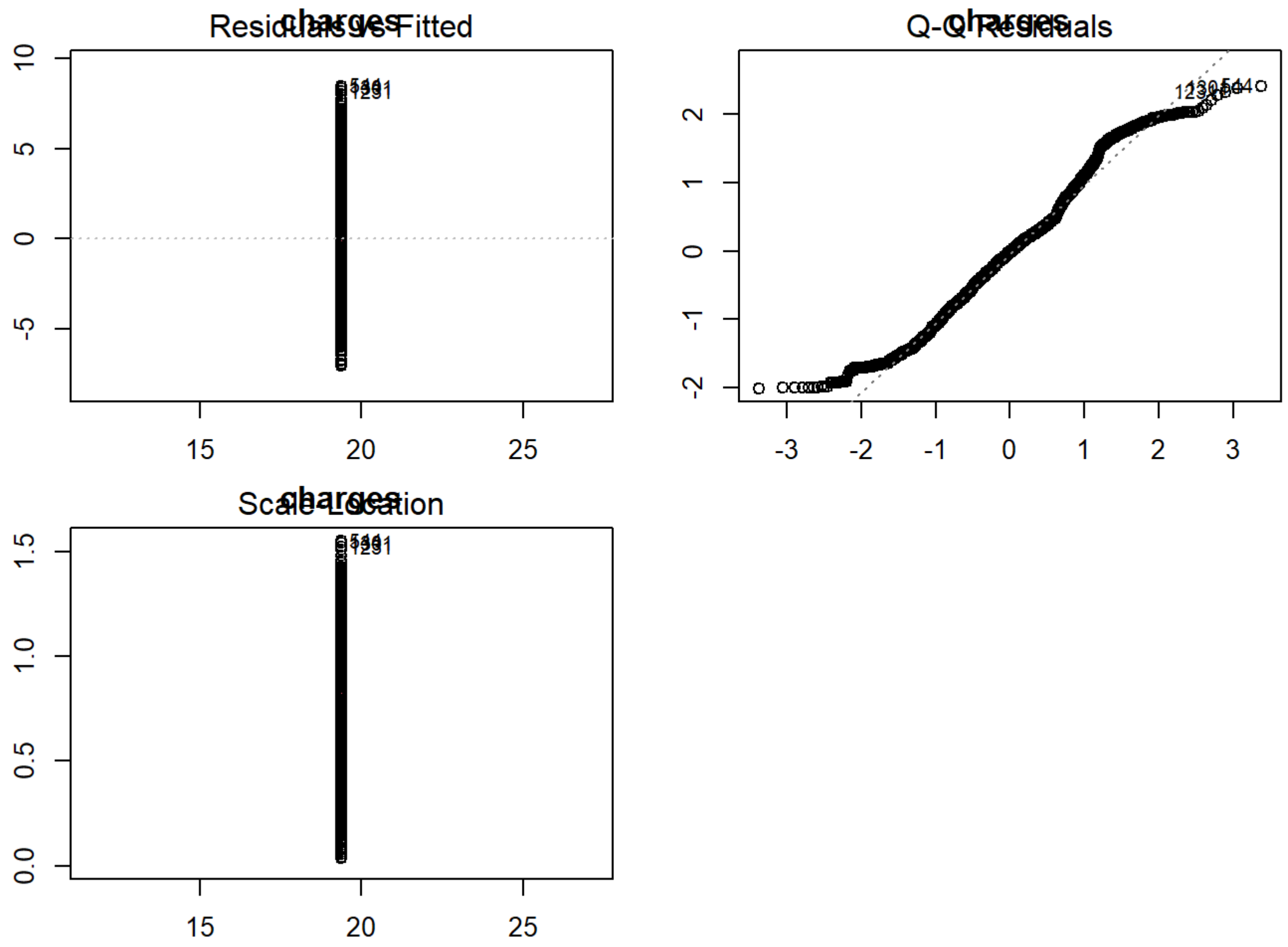
```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0719 -2.5535 -0.0541  2.2468  8.4678
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.38591    0.09646     201   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.528 on 1337 degrees of freedom
```

```
## Warning in model.matrix.default(object, data = structure(list(charges =
## c(21.6939761577649, : the response appeared on the right-hand side and was
## dropped
```

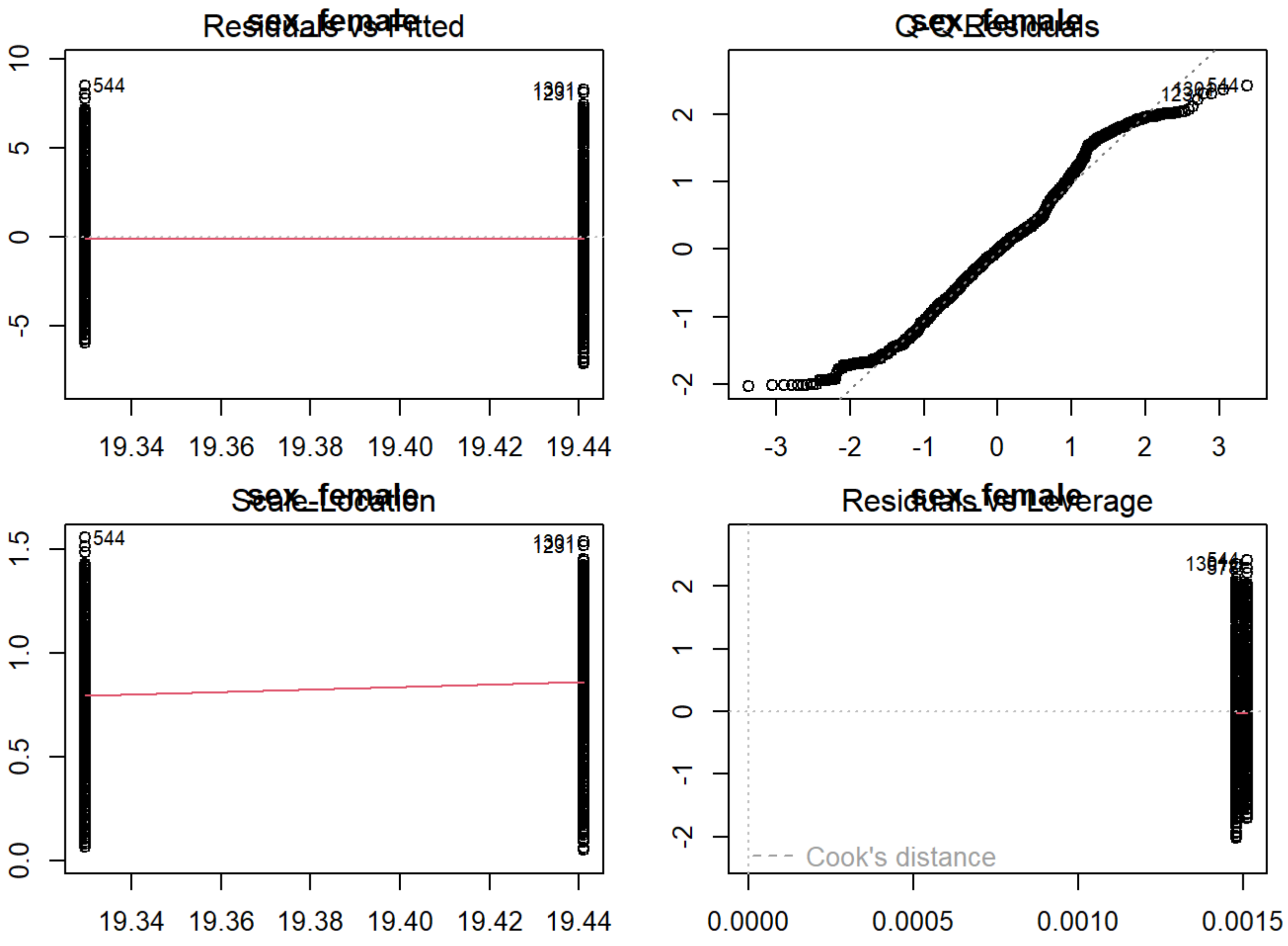
```
## Warning in model.matrix.default(object, data = structure(list(charges =
## c(21.6939761577649, : problem with term 1 in model.matrix: no columns are
## assigned
```



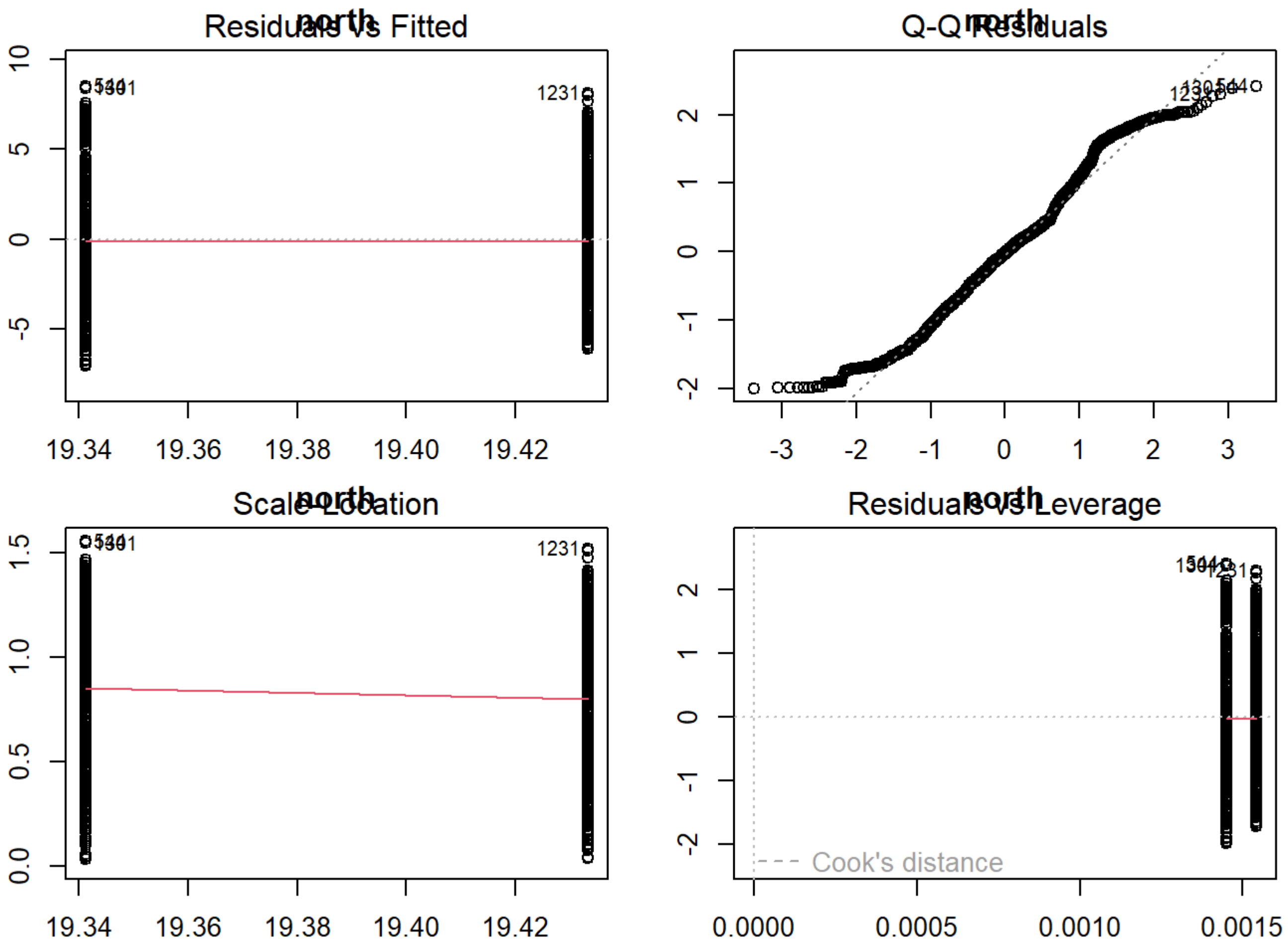
```
## hat values (leverages) are all = 0.0007473842
##  and there are no factor predictors; no plot no. 5
```



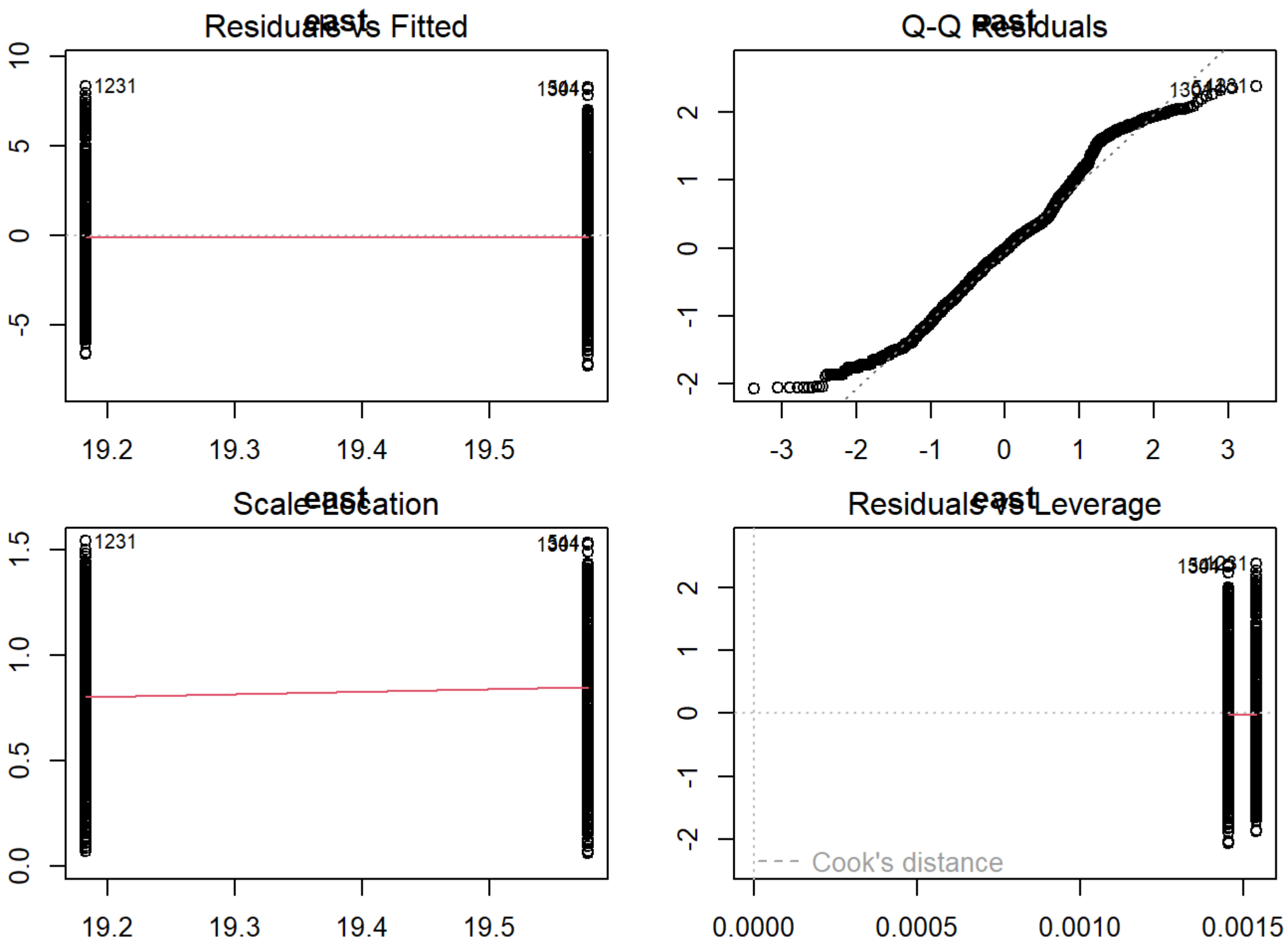
```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1270 -2.5430 -0.0634  2.2910  8.5240
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19.4410     0.1357 143.221   <2e-16 ***
## sex_female   -0.1113     0.1930  -0.577    0.564
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.529 on 1336 degrees of freedom
## Multiple R-squared:  0.0002487,	Adjusted R-squared:  -0.0004996
## F-statistic: 0.3324 on 1 and 1336 DF,  p-value: 0.5644
```



```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.0274 -2.5450 -0.0663  2.2375  8.5124
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.34135    0.13446 143.845   <2e-16 ***
## north        0.09186    0.19306   0.476    0.634
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.529 on 1336 degrees of freedom
## Multiple R-squared:  0.0001694,  Adjusted R-squared:  -0.000579
## F-statistic: 0.2264 on 1 and 1336 DF,  p-value: 0.6343
```

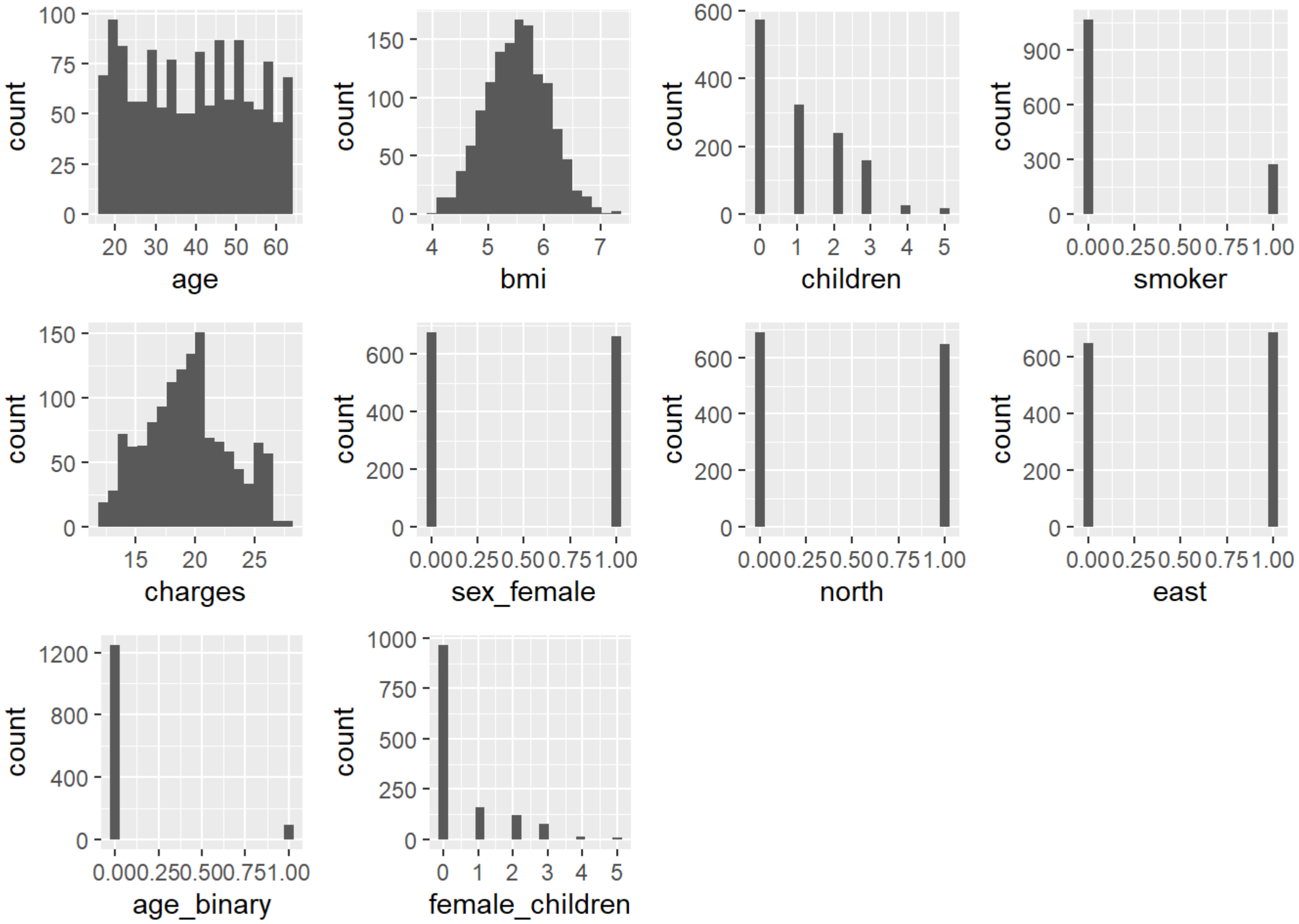


```
##
## Call:
## lm(formula = formula, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2632 -2.5628 -0.0859  2.2258  8.3624
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  19.1835     0.1382 138.778   <2e-16 ***
## east          0.3937     0.1928   2.042   0.0413 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.524 on 1336 degrees of freedom
## Multiple R-squared:  0.003112,	Adjusted R-squared:  0.002366
## F-statistic: 4.171 on 1 and 1336 DF,  p-value: 0.04132
```

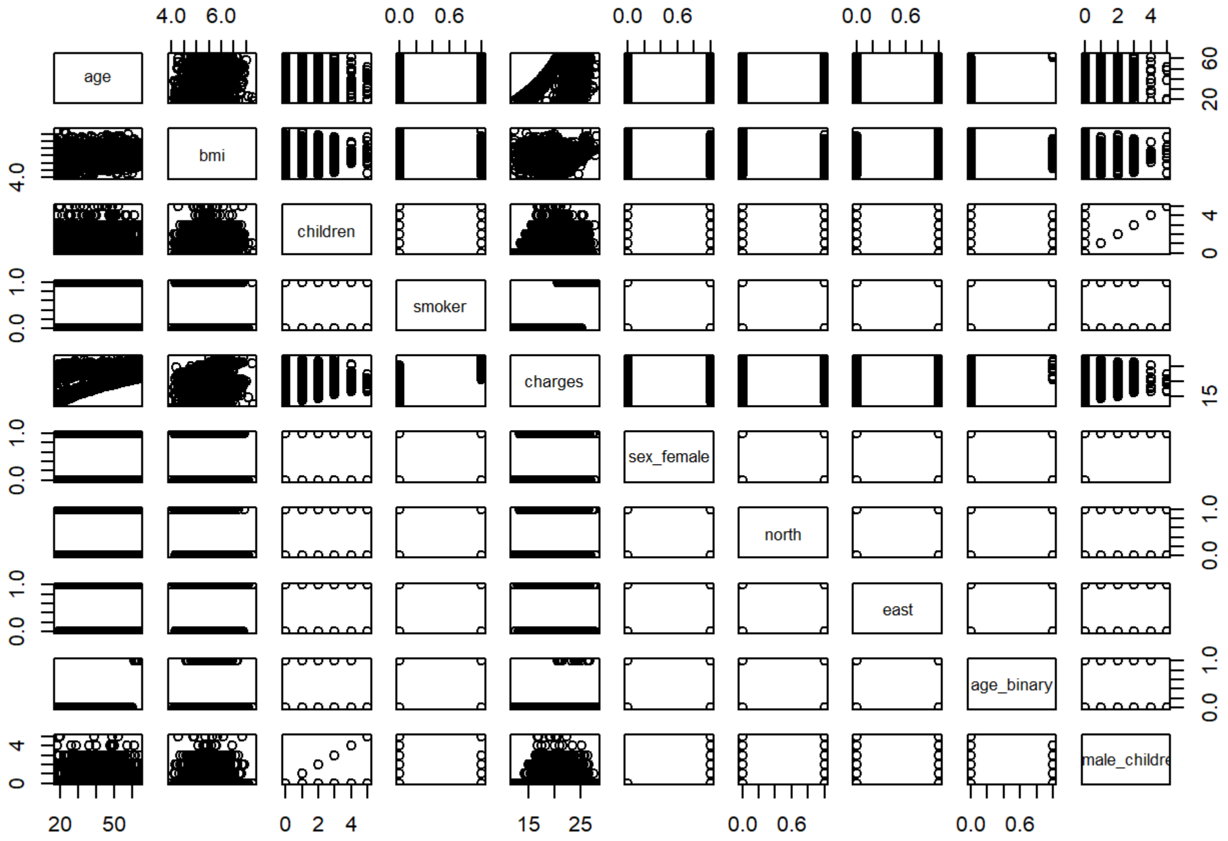


Model 2

```
data2 <- data %>%
  mutate(
    age_binary = if_else(age>60,1,0),
    bmi = sqrt(bmi),
    female_children = sex_female*children
  )

histograms(data2)
```


count
100
75
50
25
0
20 30 40 50 60
age
count
150
100
50
0
4 5 6 7
bmi
count
600
400
200
0
0 1 2 3 4 5
children
count
900
600
300
0
0.00 0.25 0.50 0.75 1.00
smoker
count
150
100
50
0
15 20 25
charges
count
600
400
200
0
0.00 0.25 0.50 0.75 1.00
sex_female
count
600
400
200
0
0.00 0.25 0.50 0.75 1.00
north
count
600
400
200
0
0.00 0.25 0.50 0.75 1.00
east
count
1200
800
400
0
0.00 0.25 0.50 0.75 1.00
age_binary
count
1000
750
500
250
0
0 1 2 3 4 5
female_children

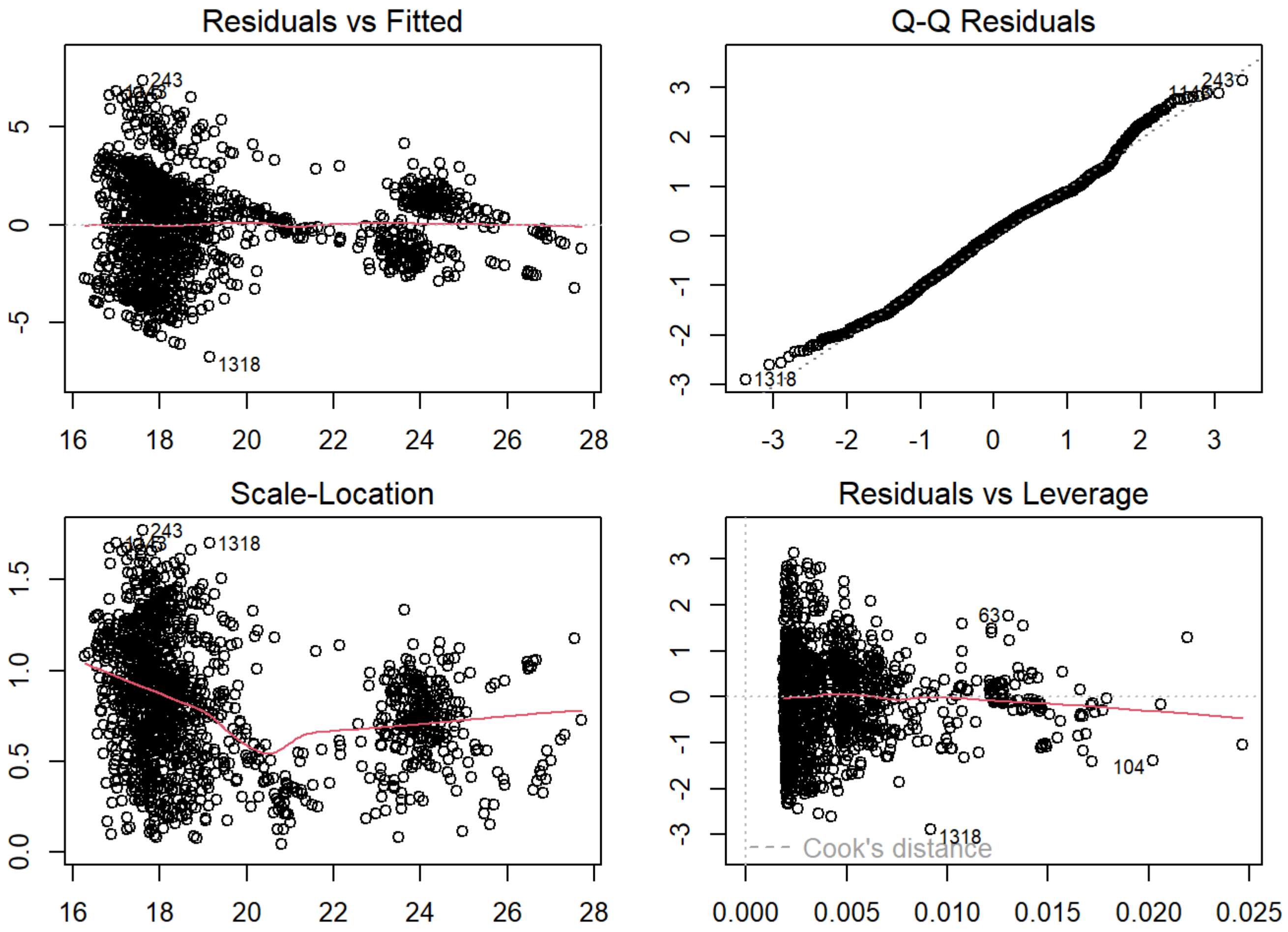
```
pairs(data2)
```



```
model <- lm(charges ~ age_binary + bmi + smoker + female_children + north, data = data2)
summary(model)
```

```
##
## Call:
## lm(formula = charges ~ age_binary + bmi + smoker + female_children +
##     north, data = data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7360 -1.6254  0.1379  1.5302  7.3388
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     12.32270    0.68112  18.092  < 2e-16 ***
## age_binary       2.68051    0.25521  10.503  < 2e-16 ***
## bmi              0.93702    0.11962   7.833 9.66e-15 ***
## smoker           6.13482    0.15882  38.629  < 2e-16 ***
## female_children  0.43206    0.06443   6.706 2.95e-11 ***
## north            0.47847    0.13164   3.635 0.000289 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.339 on 1332 degrees of freedom
## Multiple R-squared:  0.5622, Adjusted R-squared:  0.5605
## F-statistic:   342 on 5 and 1332 DF,  p-value: < 2.2e-16
```

```
par(mfrow = c(2, 2), mar = c(2,2,2,2))
plot(model)
```


Residuals vs Fitted
243
1318
5
0
-5
16
18
20
22
24
26
28
Q-Q Residuals
243
1318
3
2
1
0
-1
-2
-3
-3
-2
-1
0
1
2
3
Scale-Location
243
1318
1.5
1.0
0.5
0.0
16
18
20
22
24
26
28
Residuals vs Leverage
63
104
1318
Cook's distance
3
2
1
0
-1
-2
-3
0.000
0.005
0.010
0.015
0.020
0.025