Regression_Analysis_Housing_Electricity

2024-01-18

```
### import libraries

library(car)
```

```
## Loading required package: carData
```

```
library(MASS)
library(dplyr)
```

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

## The following object is masked from 'package:MASS':
##
##     select

## The following object is masked from 'package:car':
##
##     recode

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

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

```
library(tidyr)
library(fastDummies)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

```
library(coefplot)
```

```
## Loading required package: ggplot2
```

```
library(ggplot2)
library(leaps)
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

Loading the data

```
df = read.csv("data_cleaned_R_final.csv", head = TRUE)

head(df, 10)
```

```
##     X age income      political_party
## 1  25  65   3000              CDU/CSU
## 2  26  59    800         Keine Angabe
## 3  27  60   1750         Keine Angabe
## 4  28  73   2500                  SPD
## 5  30  43   2500 Einer anderen Partei
## 6  31  49   2300              CDU/CSU
## 7  32  57    600              CDU/CSU
## 8  33  39   5000                  SPD
## 9  34  62      0         Keine Angabe
## 10 36  45   2600         Keine Angabe
##                                                                       education
## 1  (Fach-) Hochschulabschluss (Bachelor, Master, Magister, Diplom, Staatsexamen)
## 2       Allgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)
## 3                     Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 4               Realschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
## 5                     Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 6                     Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 7               Realschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
## 8  (Fach-) Hochschulabschluss (Bachelor, Master, Magister, Diplom, Staatsexamen)
## 9  (Fach-) Hochschulabschluss (Bachelor, Master, Magister, Diplom, Staatsexamen)
## 10                    Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
##    EUROSTAT      RLK2022                                  KTU2022
## 1        PU      zentral                        Städtischer Kreis
## 2        PU sehr zentral                     kreisfreie Großstadt
## 3        IN     peripher Ländlicher Kreis mit Verdichtungsansätzen
```

```
## 4        IN sehr zentral                        Städtischer Kreis
## 5        PU sehr zentral                     kreisfreie Großstadt
## 6        IN      zentral                     kreisfreie Großstadt
## 7        IN      zentral                        Städtischer Kreis
## 8        PU sehr zentral                     kreisfreie Großstadt
## 9        PU sehr zentral                     kreisfreie Großstadt
## 10       PU sehr zentral                     kreisfreie Großstadt
##          federal_state CO2_housing CO2_electricity CO2_housing_electricity
## 1             Saarland   5038.2000        1053.000               6091.2000
## 2               Hessen   1785.0000         487.500               2272.5000
## 3               Bayern    200.1024         663.000                863.1024
## 4               Bayern    648.4800         975.000               1623.4800
## 5               Berlin   1923.4862         390.000               2313.4862
## 6       Sachsen-Anhalt   2793.0960         663.000               3456.0960
## 7    Baden-Württemberg   1620.0000         112.000               1732.0000
## 8               Berlin    902.6745          26.320                928.9945
## 9  Nordrhein-Westfalen   2340.0000         825.825               3165.8250
## 10              Hessen    868.1526          47.600                915.7526
##    CO2_cruise CO2_flight CO2_public_transport CO2_car1 CO2_car2 CO2_car3
## 1           0     2440.0                  0.0 1432.728    0.000        0
## 2        2710     5985.0                107.8 1944.608 1037.124        0
## 3           0      598.5                107.8    0.000    0.000        0
## 4           0     2287.6                  0.0 1432.728    0.000        0
## 5           0        0.0                107.8    0.000    0.000        0
## 6           0      532.0                107.8 3581.820    0.000        0
## 7           0        0.0                  0.0    0.000    0.000        0
## 8        4878     2074.8                107.8 5185.620 5185.620        0
## 9           0        0.0                107.8 2226.012 2782.515        0
## 10          0     3894.0                107.8    0.000    0.000        0
##    CO2_car4 CO2_car5 CO2_car_total CO2_mobility CO2_food CO2_other_consumption
## 1         0        0      1432.728     3872.728 1494.628              3766.100
## 2         0        0      2981.731    11784.531 1731.025              1444.879
## 3         0        0         0.000      706.300 1180.241              2433.480
## 4         0        0      1432.728     3720.328 1709.007              4152.125
## 5         0        0         0.000      107.800 1735.132              3766.100
## 6         0        0      3581.820     4221.620 1033.474              2317.600
## 7         0        0         0.000        0.000 1295.785              1520.925
## 8         0        0     10371.240    17431.840 2384.497              1216.740
## 9         0        0      5008.527     5116.327 1790.341              1376.075
## 10        0        0         0.000     4001.800 1407.010              3398.905
##    public_emission CO2_total belief_diff_housing_electricity
## 1             1152 16376.656                             -31
## 2             1152 18384.935                             -38
## 3             1152  6335.123                              40
## 4             1152 12356.940                              -2
## 5             1152  9074.518                             -43
## 6             1152 12180.790                              -6
## 7             1152  5700.710                              -1
## 8             1152 23114.072                               5
## 9             1152 12600.568                             -48
## 10            1152 10875.468                              -1
##    belief_diff_mobility belief_diff_food belief_diff_other_consumption
## 1                   -14                5                           -68
## 2                   -42              -26                            23
```

```
## 3                     11                49                             9
## 4                    -31                -9                           -36
## 5                     -2               -26                           -53
## 6                     22                93                            24
## 7                     72                60                            37
## 8                    -67               -61                            12
## 9                    -34                -5                            18
## 10                   -48                11                           -64
##    belief_diff_total
## 1                -15
## 2                -76
## 3                 57
## 4                 -8
## 5                 -1
## 6                 13
## 7                 68
## 8                -66
## 9                -16
## 10                -2
```

Hypotheses for the regression model

1. The first dependent variable: actual CO2 emission H1a: age makes differences in the actual CO2 emission from everyday activity.
H1b: income makes differences in the actual CO2 emission from everyday activity.
H1c: education level makes differences in the actual CO2 emission from everyday activity.
H1d: the place of residence (city or countryside) in the actual CO2 emission from every day activity. H1e: the region (the federal state) makes differences in the actual CO2 emission from everyday activity.
H1f: the political party that the respondent supports makes differences in the actual CO2 emission from everyday activity.

2. The second dependent variable: cons H2a: age makes differences in the consumers' belief about CO2 emission from everyday activity.
H2b: income makes differences in the consumers' belief about CO2 emission from everyday activity.
H2c: education level makes differences in the consumers' belief about CO2 emission from everyday activity.
H2d: the place of residence (city or countryside) makes differences in the consumers' belief about CO2 emission from everyday activity.
H2e: the region (the federal state) makes differences in the consumers' belief about CO2 emission from everyday activity.
H2f: the political party that the respondent supports makes differences in the consumers' belief about CO2 emission from everyday activity.

Independent variables in the dataset

1. age: age, numerical variable
2. income: monthly net income in Euro, numerical variable, less than 10,000 EUR only (outlier removed)
3. education: categorical variable
4. urban_rural_class: categorical variable
5. federal_state: federal state, categorical variable
6. political_party: political_party, categorical variable

Dependent variables in the dataset

1. Actual CO2 from housing, electricity, mobility, food, other consumption

1) CO2_housing_electricity
2) CO2_mobility
3) CO2_food
4) CO2_other_consumption
5) CO2_total

2. Belief about CO2

1) belief_diff_housing_electricity
2) belief_diff_mobility
3) belief_diff_food
4) belief_diff_other_consumption
5) belief_diff_total

Data preparation

```
# change into categorical variable

df$education <-as.factor(df$education)
df$EUROSTAT <-as.factor(df$EUROSTAT)
df$RLK2022 <-as.factor(df$RLK2022)
df$KTU2022 <-as.factor(df$KTU2022)
df$political_party <-as.factor(df$political_party)
df$federal_state <-as.factor(df$federal_state)
```

```
## Select the classification for the urban_rural

#df1_1<-  subset(df, select = -c(KTU2022, RLK2022) #EUROSTATS

df1_1<-  subset(df, select = -c(KTU2022, EUROSTAT)) #RLK2022

#df1_1<-  subset(df, select = -c(RLK2022, EUROSTAT)) #KTU2022

names(df1_1)[names(df1_1) == 'RLK2022'] <- 'urban_rural_class'  #change the variable name!!

head(df1_1)
```

```
##    X age income      political_party
## 1 25  65   3000              CDU/CSU
## 2 26  59    800         Keine Angabe
## 3 27  60   1750         Keine Angabe
## 4 28  73   2500                  SPD
## 5 30  43   2500 Einer anderen Partei
## 6 31  49   2300              CDU/CSU
##                                                                      education
## 1 (Fach-) Hochschulabschluss (Bachelor, Master, Magister, Diplom, Staatsexamen)
## 2      Allgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)
```

```
## 3                  Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 4            Realschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
## 5                  Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 6                  Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
##   urban_rural_class  federal_state CO2_housing CO2_electricity
## 1           zentral       Saarland   5038.2000          1053.0
## 2      sehr zentral         Hessen   1785.0000           487.5
## 3          peripher         Bayern    200.1024           663.0
## 4      sehr zentral         Bayern    648.4800           975.0
## 5      sehr zentral         Berlin   1923.4862           390.0
## 6           zentral Sachsen-Anhalt   2793.0960           663.0
##   CO2_housing_electricity CO2_cruise CO2_flight CO2_public_transport CO2_car1
## 1               6091.2000          0     2440.0                  0.0 1432.728
## 2               2272.5000       2710     5985.0                107.8 1944.608
## 3                863.1024          0      598.5                107.8    0.000
## 4               1623.4800          0     2287.6                  0.0 1432.728
## 5               2313.4862          0        0.0                107.8    0.000
## 6               3456.0960          0      532.0                107.8 3581.820
##   CO2_car2 CO2_car3 CO2_car4 CO2_car5 CO2_car_total CO2_mobility CO2_food
## 1    0.000        0        0        0      1432.728     3872.728 1494.628
## 2 1037.124        0        0        0      2981.731    11784.531 1731.025
## 3    0.000        0        0        0         0.000      706.300 1180.241
## 4    0.000        0        0        0      1432.728     3720.328 1709.007
## 5    0.000        0        0        0         0.000      107.800 1735.132
## 6    0.000        0        0        0      3581.820     4221.620 1033.474
##   CO2_other_consumption public_emission CO2_total
## 1              3766.100            1152 16376.656
## 2              1444.879            1152 18384.935
## 3              2433.480            1152  6335.123
## 4              4152.125            1152 12356.940
## 5              3766.100            1152  9074.518
## 6              2317.600            1152 12180.790
##   belief_diff_housing_electricity belief_diff_mobility belief_diff_food
## 1                             -31                  -14                5
## 2                             -38                  -42              -26
## 3                              40                   11               49
## 4                              -2                  -31               -9
## 5                             -43                   -2              -26
## 6                              -6                   22               93
##   belief_diff_other_consumption belief_diff_total
## 1                           -68               -15
## 2                            23               -76
## 3                             9                57
## 4                           -36                -8
## 5                           -53                -1
## 6                            24                13
```

```
## Creating a demo-dataset for a quick regression model building

# Independent variables: age, income, political_party, education, urban_rural, federal_state
# Dependent variables: CO2_housing_electricity


df1 <- as_tibble(df1_1)
```

```
head(df1)
```

```
## # A tibble: 6 x 29
##       X   age income political~1 educa~2 urban~3 feder~4 CO2_h~5 CO2_e~6 CO2_h~7
##   <int> <int>  <dbl> <fct>       <fct>   <fct>   <fct>     <dbl>   <dbl>   <dbl>
## 1    25    65   3000 CDU/CSU     (Fach-~ zentral Saarla~   5038.    1053    6091.
## 2    26    59    800 Keine Anga~ Allgem~ sehr z~ Hessen    1785      488.   2272.
## 3    27    60   1750 Keine Anga~ Berufs~ periph~ Bayern     200.     663     863.
## 4    28    73   2500 SPD         Realsc~ sehr z~ Bayern     648.     975    1623.
## 5    30    43   2500 Einer ande~ Berufs~ sehr z~ Berlin    1923.     390    2313.
## 6    31    49   2300 CDU/CSU     Berufs~ zentral Sachse~   2793.     663    3456.
## # ... with 19 more variables: CO2_cruise <dbl>, CO2_flight <dbl>,
## #   CO2_public_transport <dbl>, CO2_car1 <dbl>, CO2_car2 <dbl>, CO2_car3 <dbl>,
## #   CO2_car4 <dbl>, CO2_car5 <dbl>, CO2_car_total <dbl>, CO2_mobility <dbl>,
## #   CO2_food <dbl>, CO2_other_consumption <dbl>, public_emission <dbl>,
## #   CO2_total <dbl>, belief_diff_housing_electricity <dbl>,
## #   belief_diff_mobility <dbl>, belief_diff_food <dbl>,
## #   belief_diff_other_consumption <dbl>, belief_diff_total <dbl>, and ...
```

```
df1 <- df1 %>% select(2, 3, 4, 5, 6, 7, 10) #10, 20, 21, 22, 24

df1
```

```
## # A tibble: 588 x 7
##      age income political_party      education           urban~1 feder~2 CO2_h~3
##    <int>  <dbl> <fct>                <fct>               <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU              (Fach-) Hochschula~ zentral Saarla~   6091.
##  2    59    800 Keine Angabe         Allgemeine oder fa~ sehr z~ Hessen    2272.
##  3    60   1750 Keine Angabe         Berufsausbildung, ~ periph~ Bayern     863.
##  4    73   2500 SPD                  Realschulabschluss~ sehr z~ Bayern    1623.
##  5    43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin    2313.
##  6    49   2300 CDU/CSU              Berufsausbildung, ~ zentral Sachse~   3456.
##  7    57    600 CDU/CSU              Realschulabschluss~ zentral Baden-~   1732
##  8    39   5000 SPD                  (Fach-) Hochschula~ sehr z~ Berlin     929.
##  9    62      0 Keine Angabe         (Fach-) Hochschula~ sehr z~ Nordrh~   3166.
## 10    45   2600 Keine Angabe         Berufsausbildung, ~ sehr z~ Hessen     916.
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: CO2_housing_electricity
```

```
## Creating a demo-dataset for a quick regression model building

# Independent variables: age, income, political_party, education, urban_rural, federal_state
# Dependent variables: belief_diff_housing_electricity


df2 <- as_tibble(df1_1)

head(df1_1)
```

```
##    X age income      political_party
## 1 25  65   3000              CDU/CSU
## 2 26  59    800         Keine Angabe
```

```
## 3 27  60   1750         Keine Angabe
## 4 28  73   2500                  SPD
## 5 30  43   2500 Einer anderen Partei
## 6 31  49   2300              CDU/CSU
##                                                                         education
## 1 (Fach-) Hochschulabschluss (Bachelor, Master, Magister, Diplom, Staatsexamen)
## 2     Allgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)
## 3                  Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 4             Realschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
## 5                  Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 6                  Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
##   urban_rural_class  federal_state CO2_housing CO2_electricity
## 1           zentral       Saarland   5038.2000          1053.0
## 2      sehr zentral         Hessen   1785.0000           487.5
## 3          peripher         Bayern    200.1024           663.0
## 4      sehr zentral         Bayern    648.4800           975.0
## 5      sehr zentral         Berlin   1923.4862           390.0
## 6           zentral Sachsen-Anhalt   2793.0960           663.0
##   CO2_housing_electricity CO2_cruise CO2_flight CO2_public_transport CO2_car1
## 1               6091.2000          0     2440.0                  0.0 1432.728
## 2               2272.5000       2710     5985.0                107.8 1944.608
## 3                863.1024          0      598.5                107.8    0.000
## 4               1623.4800          0     2287.6                  0.0 1432.728
## 5               2313.4862          0        0.0                107.8    0.000
## 6               3456.0960          0      532.0                107.8 3581.820
##   CO2_car2 CO2_car3 CO2_car4 CO2_car5 CO2_car_total CO2_mobility CO2_food
## 1    0.000        0        0        0      1432.728     3872.728 1494.628
## 2 1037.124        0        0        0      2981.731    11784.531 1731.025
## 3    0.000        0        0        0         0.000      706.300 1180.241
## 4    0.000        0        0        0      1432.728     3720.328 1709.007
## 5    0.000        0        0        0         0.000      107.800 1735.132
## 6    0.000        0        0        0      3581.820     4221.620 1033.474
##   CO2_other_consumption public_emission CO2_total
## 1              3766.100            1152 16376.656
## 2              1444.879            1152 18384.935
## 3              2433.480            1152  6335.123
## 4              4152.125            1152 12356.940
## 5              3766.100            1152  9074.518
## 6              2317.600            1152 12180.790
##   belief_diff_housing_electricity belief_diff_mobility belief_diff_food
## 1                             -31                  -14                5
## 2                             -38                  -42              -26
## 3                              40                   11               49
## 4                              -2                  -31               -9
## 5                             -43                   -2              -26
## 6                              -6                   22               93
##   belief_diff_other_consumption belief_diff_total
## 1                           -68               -15
## 2                            23               -76
## 3                             9                57
## 4                           -36                -8
## 5                           -53                -1
## 6                            24                13
```

```r
df2 <- df2 %>% select(2, 3, 4, 5, 6, 7, 25) #25, 26, 27, 28, 29

df2
```

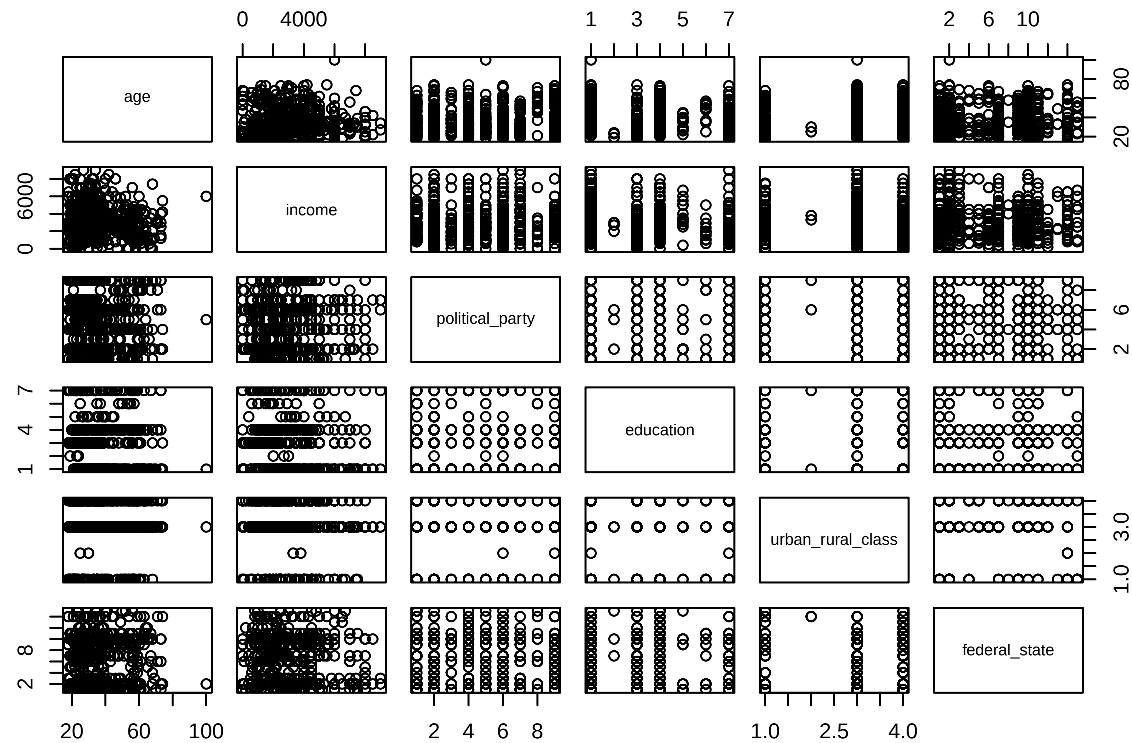
```
## # A tibble: 588 x 7
##      age income political_party      education            urban~1 feder~2 belie~3
##    <int>  <dbl> <fct>                <fct>                <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU              (Fach-) Hochschula~ zentral Saarla~     -31
##  2    59    800 Keine Angabe         Allgemeine oder fa~ sehr z~ Hessen      -38
##  3    60   1750 Keine Angabe         Berufsausbildung, ~ periph~ Bayern       40
##  4    73   2500 SPD                  Realschulabschluss~ sehr z~ Bayern       -2
##  5    43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin      -43
##  6    49   2300 CDU/CSU              Berufsausbildung, ~ zentral Sachse~      -6
##  7    57    600 CDU/CSU              Realschulabschluss~ zentral Baden-~      -1
##  8    39   5000 SPD                  (Fach-) Hochschula~ sehr z~ Berlin        5
##  9    62      0 Keine Angabe         (Fach-) Hochschula~ sehr z~ Nordrh~     -48
## 10    45   2600 Keine Angabe         Berufsausbildung, ~ sehr z~ Hessen       -1
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: belief_diff_housing_electricity
```

I. Exploratory Data Analysis

Check the Jupytor notebook: EDA_scatter_plot_actual_belief

II. Multivariate Regression: CO2 housing electricity

```r
# Checking the possible correlation in the data

plot(df1[1:6])
```



1. Modeling

```
table(df1$political_party)
```

```
##
##                       AfD     Bündnis 90/Die Grünen Bündnis Sarah Wagenknecht
##                        58                       143                        23
##                   CDU/CSU                 Die Linke      Einer anderen Partei
##                        75                        44                       111
##                       FDP             Keine Angabe                       SPD
##                        48                        15                        71
```

```
table(df1$education)
```

```
##
## (Fach-) Hochschulabschluss (Bachelor, Master, Magister, Diplom, Staatsexamen)
##                                                                           253
##                                                          (Noch) kein Abschluss
##                                                                             3
##      Allgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)
##                                                                           131
##                   Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
##                                                                           118
##                                                  Doktorgrad oder Habilitation
##                                                                            13
##       Hauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss
##                                                                            11
```

```
##            Realschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
##                                                                         59
```

```
table(df1$urban_rural_class)
```

```
##
##      peripher sehr peripher  sehr zentral       zentral
##            79             2           350           157
```

```
table(df1$federal_state)
```

```
##
##      Baden-Württemberg                 Bayern                 Berlin
##                     94                    100                     44
##            Brandenburg                 Bremen                Hamburg
##                      8                     15                     25
##                 Hessen Mecklenburg-Vorpommern          Niedersachsen
##                     50                      2                     58
##    Nordrhein-Westfalen        Rheinland-Pfalz               Saarland
##                    117                     30                     10
##         Sachsen-Anhalt     Schleswig-Holstein              Thüringen
##                      4                     22                      9
```

```
## defining a reference level

df1$political_party  <- relevel(df1$political_party, ref='Bündnis 90/Die Grünen')
df1$education  <- relevel(df1$education, ref='(Fach-) Hochschulabschluss (Bachelor, Master, Magister, Di
df1$urban_rural_class  <- relevel(df1$urban_rural_class, ref='sehr zentral')
df1$federal_state  <- relevel(df1$federal_state, ref='Nordrhein-Westfalen')
```

```
# regression model with all variables

model1 <- lm(CO2_housing_electricity  ~ age + income + political_party + education +  urban_rural_class

summary(model1)
```

```
##
## Call:
## lm(formula = CO2_housing_electricity ~ age + income + political_party +
##     education + urban_rural_class + federal_state, data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2671.4  -762.0  -235.4   437.8 13841.0
##
## Coefficients:
##                                                                                          Estimate
## (Intercept)                                                                            1444.32318
## age                                                                                      12.94510
## income                                                                                   -0.05536
## political_partyAfD                                                                      325.16106
## political_partyBündnis Sarah Wagenknecht                                                166.97042
```

```
## political_partyCDU/CSU                                                               17.13978
## political_partyDie Linke                                                           -142.15689
## political_partyEiner anderen Partei                                                  33.17008
## political_partyFDP                                                                  504.59733
## political_partyKeine Angabe                                                         119.44216
## political_partySPD                                                                  118.45958
## education(Noch) kein Abschluss                                                     -608.98577
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   83.56266
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                -91.19049
## educationDoktorgrad oder Habilitation                                                 9.27019
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  -494.68473
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss          -19.29133
## urban_rural_classperipher                                                           300.42781
## urban_rural_classsehr peripher                                                     -876.89481
## urban_rural_classzentral                                                           -237.95096
## federal_stateBaden-Württemberg                                                     -330.23463
## federal_stateBayern                                                                -203.80324
## federal_stateBerlin                                                                -144.95052
## federal_stateBrandenburg                                                           -279.64348
## federal_stateBremen                                                                 391.75815
## federal_stateHamburg                                                               -361.35209
## federal_stateHessen                                                                 338.73305
## federal_stateMecklenburg-Vorpommern                                                -581.97535
## federal_stateNiedersachsen                                                          141.91428
## federal_stateRheinland-Pfalz                                                        696.83833
## federal_stateSaarland                                                              1485.10203
## federal_stateSachsen-Anhalt                                                        1259.44507
## federal_stateSchleswig-Holstein                                                     283.79099
## federal_stateThüringen                                                              660.92693
##                                                                                    Std. Error
## (Intercept)                                                                         289.24263
## age                                                                                   5.12072
## income                                                                                0.03422
## political_partyAfD                                                                  245.00273
## political_partyBündnis Sarah Wagenknecht                                            343.40265
## political_partyCDU/CSU                                                              220.04990
## political_partyDie Linke                                                            265.22422
## political_partyEiner anderen Partei                                                 196.85985
## political_partyFDP                                                                  254.85140
## political_partyKeine Angabe                                                         441.43510
## political_partySPD                                                                  224.08098
## education(Noch) kein Abschluss                                                      896.49641
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  174.26340
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                177.77469
## educationDoktorgrad oder Habilitation                                               434.56494
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   496.05145
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss          226.07717
## urban_rural_classperipher                                                           227.82185
## urban_rural_classsehr peripher                                                     1127.83798
## urban_rural_classzentral                                                            167.02552
## federal_stateBaden-Württemberg                                                      216.97097
## federal_stateBayern                                                                 222.76807
## federal_stateBerlin                                                                 270.62556
## federal_stateBrandenburg                                                            568.41886
```

```
## federal_stateBremen                                                            414.60387
## federal_stateHamburg                                                           337.40376
## federal_stateHessen                                                            258.86887
## federal_stateMecklenburg-Vorpommern                                           1086.61583
## federal_stateNiedersachsen                                                     263.35516
## federal_stateRheinland-Pfalz                                                   324.81836
## federal_stateSaarland                                                          510.04738
## federal_stateSachsen-Anhalt                                                    786.22894
## federal_stateSchleswig-Holstein                                                379.91939
## federal_stateThüringen                                                         575.85538
##                                                                              t value
## (Intercept)                                                                    4.993
## age                                                                            2.528
## income                                                                        -1.618
## political_partyAfD                                                             1.327
## political_partyBündnis Sarah Wagenknecht                                       0.486
## political_partyCDU/CSU                                                         0.078
## political_partyDie Linke                                                      -0.536
## political_partyEiner anderen Partei                                            0.168
## political_partyFDP                                                             1.980
## political_partyKeine Angabe                                                    0.271
## political_partySPD                                                             0.529
## education(Noch) kein Abschluss                                                -0.679
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  0.480
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule          -0.513
## educationDoktorgrad oder Habilitation                                          0.021
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  -0.997
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss     -0.085
## urban_rural_classperipher                                                      1.319
## urban_rural_classsehr peripher                                                -0.778
## urban_rural_classzentral                                                      -1.425
## federal_stateBaden-Württemberg                                                -1.522
## federal_stateBayern                                                           -0.915
## federal_stateBerlin                                                           -0.536
## federal_stateBrandenburg                                                      -0.492
## federal_stateBremen                                                            0.945
## federal_stateHamburg                                                          -1.071
## federal_stateHessen                                                            1.309
## federal_stateMecklenburg-Vorpommern                                           -0.536
## federal_stateNiedersachsen                                                     0.539
## federal_stateRheinland-Pfalz                                                   2.145
## federal_stateSaarland                                                          2.912
## federal_stateSachsen-Anhalt                                                    1.602
## federal_stateSchleswig-Holstein                                                0.747
## federal_stateThüringen                                                         1.148
##                                                                              Pr(>|t|)
## (Intercept)                                                                  7.96e-07
## age                                                                           0.01175
## income                                                                        0.10630
## political_partyAfD                                                            0.18500
## political_partyBündnis Sarah Wagenknecht                                      0.62700
## political_partyCDU/CSU                                                        0.93794
## political_partyDie Linke                                                      0.59218
## political_partyEiner anderen Partei                                           0.86625
```

```
## political_partyFDP                                                               0.04820
## political_partyKeine Angabe                                                      0.78682
## political_partySPD                                                               0.59726
## education(Noch) kein Abschluss                                                   0.49723
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  0.63176
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule             0.60819
## educationDoktorgrad oder Habilitation                                            0.98299
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   0.31908
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss        0.93203
## urban_rural_classperipher                                                        0.18782
## urban_rural_classsehr peripher                                                   0.43720
## urban_rural_classzentral                                                         0.15482
## federal_stateBaden-Württemberg                                                   0.12857
## federal_stateBayern                                                              0.36066
## federal_stateBerlin                                                              0.59244
## federal_stateBrandenburg                                                         0.62294
## federal_stateBremen                                                              0.34512
## federal_stateHamburg                                                             0.28465
## federal_stateHessen                                                              0.19124
## federal_stateMecklenburg-Vorpommern                                              0.59246
## federal_stateNiedersachsen                                                       0.59019
## federal_stateRheinland-Pfalz                                                     0.03236
## federal_stateSaarland                                                            0.00374
## federal_stateSachsen-Anhalt                                                      0.10975
## federal_stateSchleswig-Holstein                                                  0.45539
## federal_stateThüringen                                                           0.25158
##
## (Intercept)                                                                     ***
## age                                                                             *
## income
## political_partyAfD
## political_partyBündnis Sarah Wagenknecht
## political_partyCDU/CSU
## political_partyDie Linke
## political_partyEiner anderen Partei
## political_partyFDP                                                              *
## political_partyKeine Angabe
## political_partySPD
## education(Noch) kein Abschluss
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule
## educationDoktorgrad oder Habilitation
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
## urban_rural_classperipher
## urban_rural_classsehr peripher
## urban_rural_classzentral
## federal_stateBaden-Württemberg
## federal_stateBayern
## federal_stateBerlin
## federal_stateBrandenburg
## federal_stateBremen
## federal_stateHamburg
## federal_stateHessen
```

```
## federal_stateMecklenburg-Vorpommern
## federal_stateNiedersachsen
## federal_stateRheinland-Pfalz                                               *
## federal_stateSaarland                                                      **
## federal_stateSachsen-Anhalt
## federal_stateSchleswig-Holstein
## federal_stateThüringen
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1497 on 554 degrees of freedom
## Multiple R-squared:  0.08865,    Adjusted R-squared:  0.03436
## F-statistic: 1.633 on 33 and 554 DF,  p-value: 0.01562
```

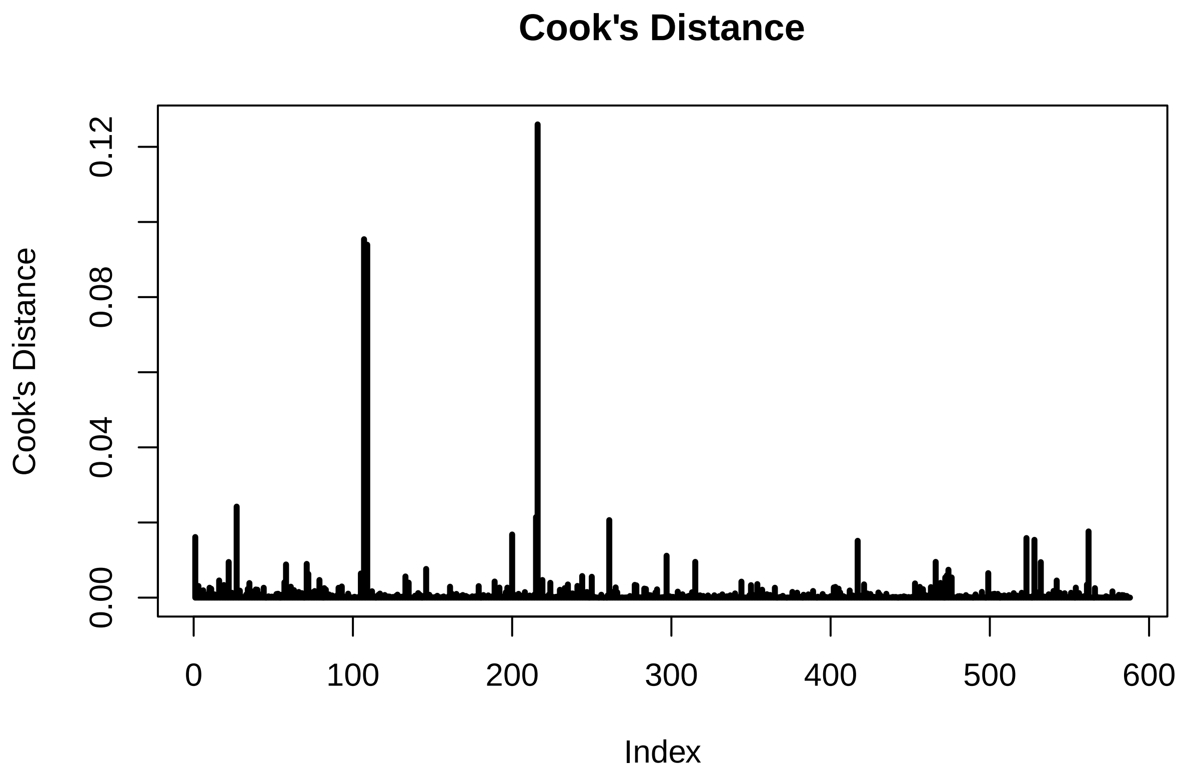
```r
# Checking the VIFs for multicollinearity

vif(model1)
```

```
##                         GVIF Df GVIF^(1/(2*Df))
## age                 1.313360  1        1.146019
## income              1.099357  1        1.048502
## political_party     1.794759  8        1.037231
## education           1.848270  6        1.052520
## urban_rural_class 2.066166  3        1.128568
## federal_state       3.002832 14        1.040051
```

```r
# threshold for multicollinearity
# Calculating the threshold

max(10, 1/(1-summary(model1)$r.square))
```

```
## [1] 10
```

```r
# Checking outliers: estimate of the influence of data point; summary of how much a regression model ch

cook = cooks.distance(model1)
plot(cook,
     type="h",
     lwd=3,
     ylab = "Cook's Distance",
     main="Cook's Distance")
abline(h = 1)
```




```
influential = cooks.distance(model1)[which(cook > 3*mean(cook, na.rm=TRUE))]
influential
```

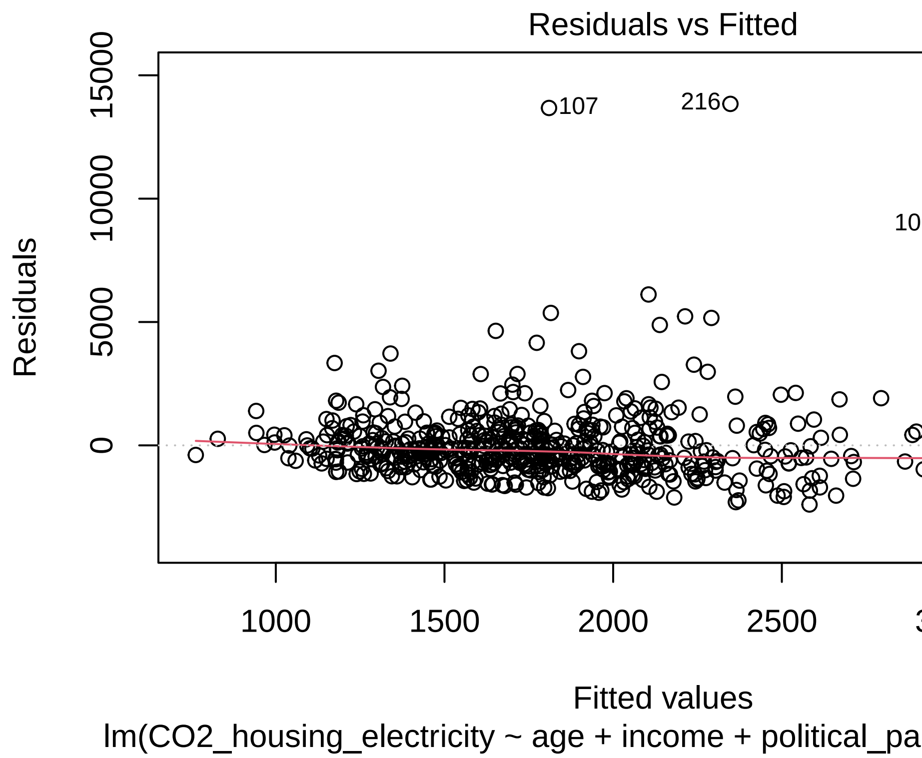
```
##           1          22          27          58          71          72
## 0.016127624 0.009480507 0.024224697 0.008826271 0.008973082 0.006299245
##         105         107         109         133         146         200
## 0.006429258 0.095412252 0.093940930 0.005639933 0.007636482 0.016803997
##         215         216         244         250         261         297
## 0.021428641 0.125965406 0.005741790 0.005549451 0.020633259 0.011149565
##         315         417         466         472         473         474
## 0.009513089 0.015125598 0.009513787 0.005364578 0.005906450 0.007422763
##         476         499         523         528         532         562
## 0.005368812 0.006519123 0.015858419 0.015381180 0.009444404 0.017618512
```

```
influential = influential[!is.na(influential)]
influential_vector = c(as.numeric(rownames(data.frame(influential))))


df1[influential_vector, ]
```

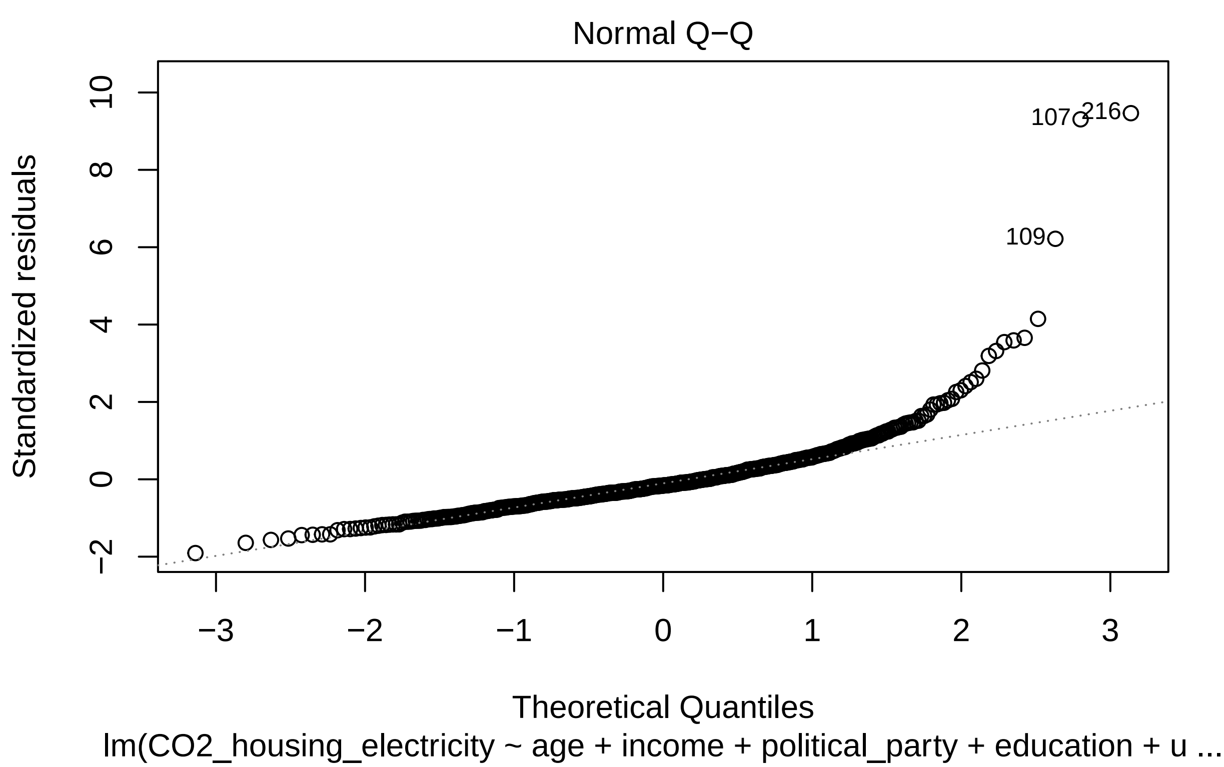
```
## # A tibble: 30 x 7
##      age income political_party           education      urban~1 feder~2 CO2_h~3
##    <int>  <dbl> <fct>                     <fct>          <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU                   (Fach-) Hochs~ zentral Saarla~   6091.
##  2    52   4800 Die Linke                 (Fach-) Hochs~ periph~ Thürin~   4534.
```

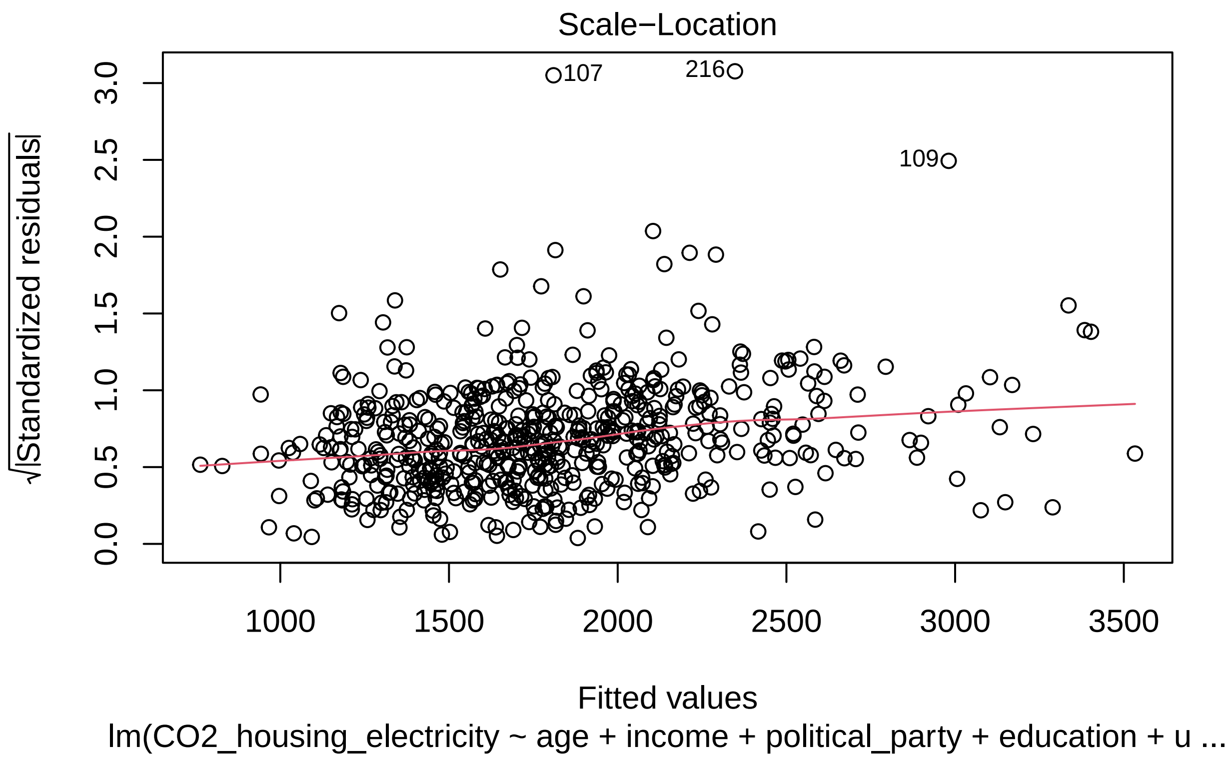
```
## 3    36   1000 AfD                        Berufsausbild~ zentral Saarla~   6713.
## 4    53   1500 AfD                        Hauptschulabs~ periph~ Bayern    4093.
## 5    56   1000 Keine Angabe               Berufsausbild~ periph~ Thürin~   4700.
## 6    49   2000 Keine Angabe               Berufsausbild~ sehr z~ Baden-~   3771.
## 7    49   3000 Bündnis 90/Die Grünen      Berufsausbild~ zentral Rheinl~   5261.
## 8    32   7000 Bündnis 90/Die Grünen      (Fach-) Hochs~ sehr z~ Hessen   15486.
## 9    22    600 FDP                        Allgemeine od~ sehr z~ Rheinl~  11925
## 10   29   1900 Bündnis Sarah Wagenknecht Berufsausbild~ sehr z~ Rheinl~    451.
## # ... with 20 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: CO2_housing_electricity
```

```
plot(model1)
```

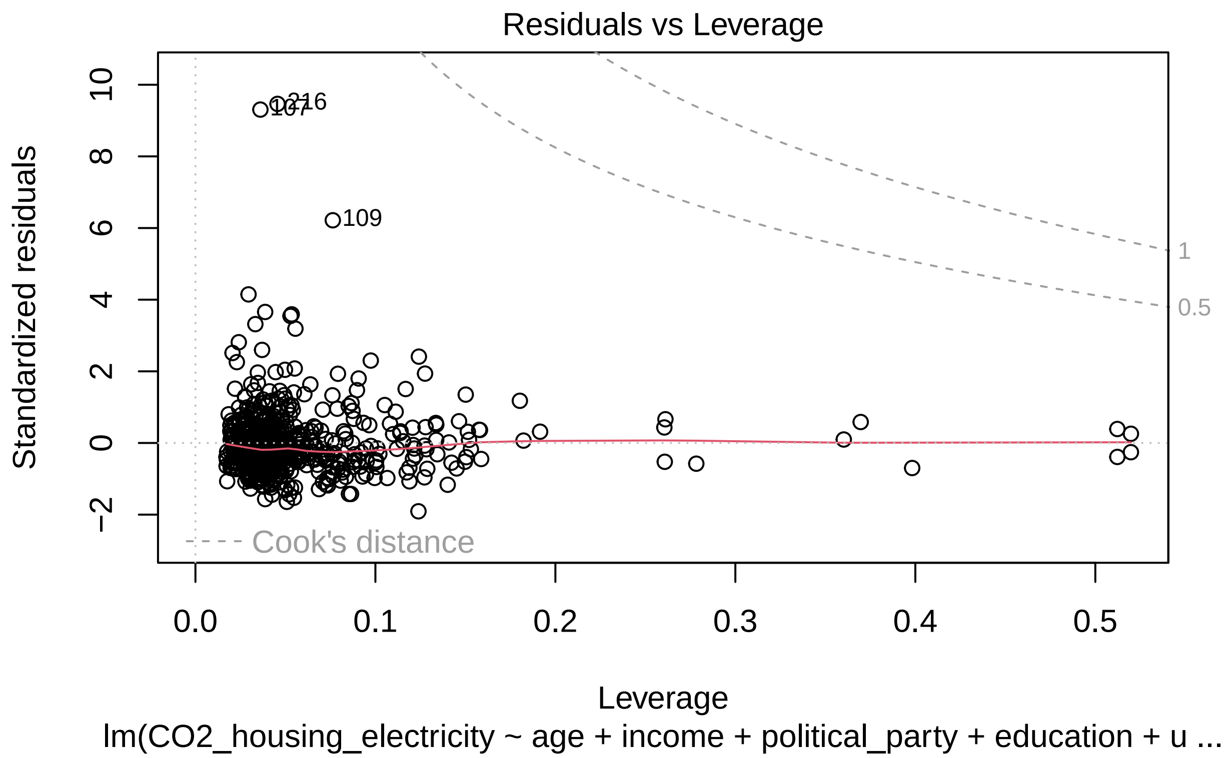


2. **Assumptions check in the residuals**

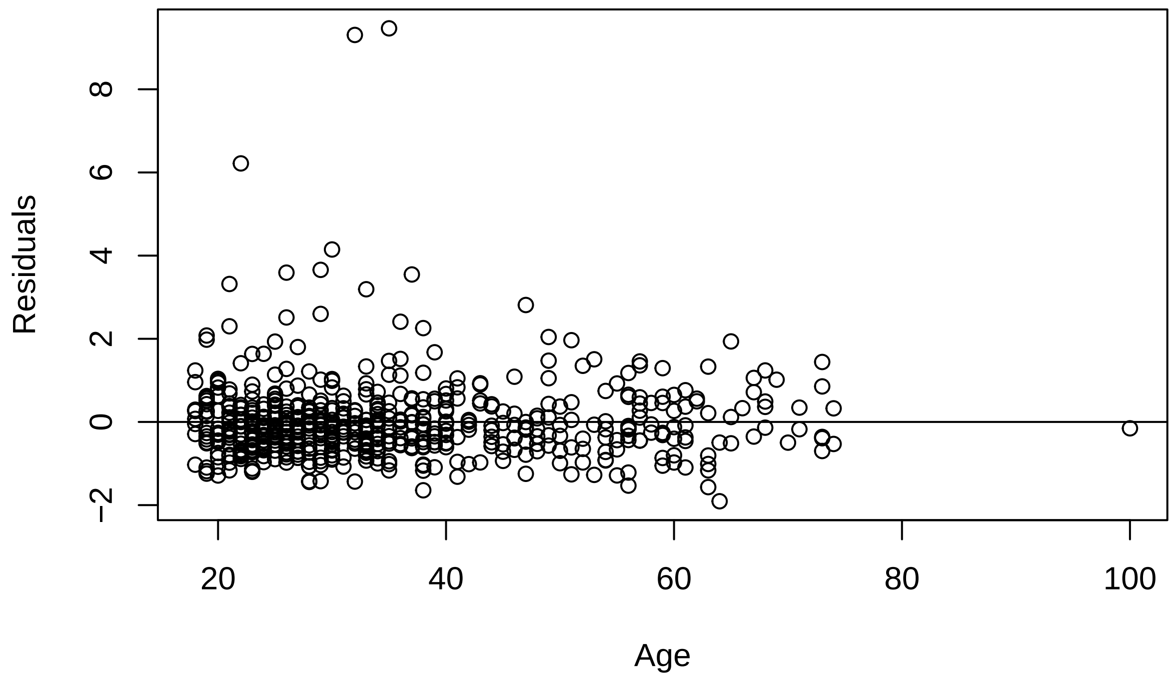




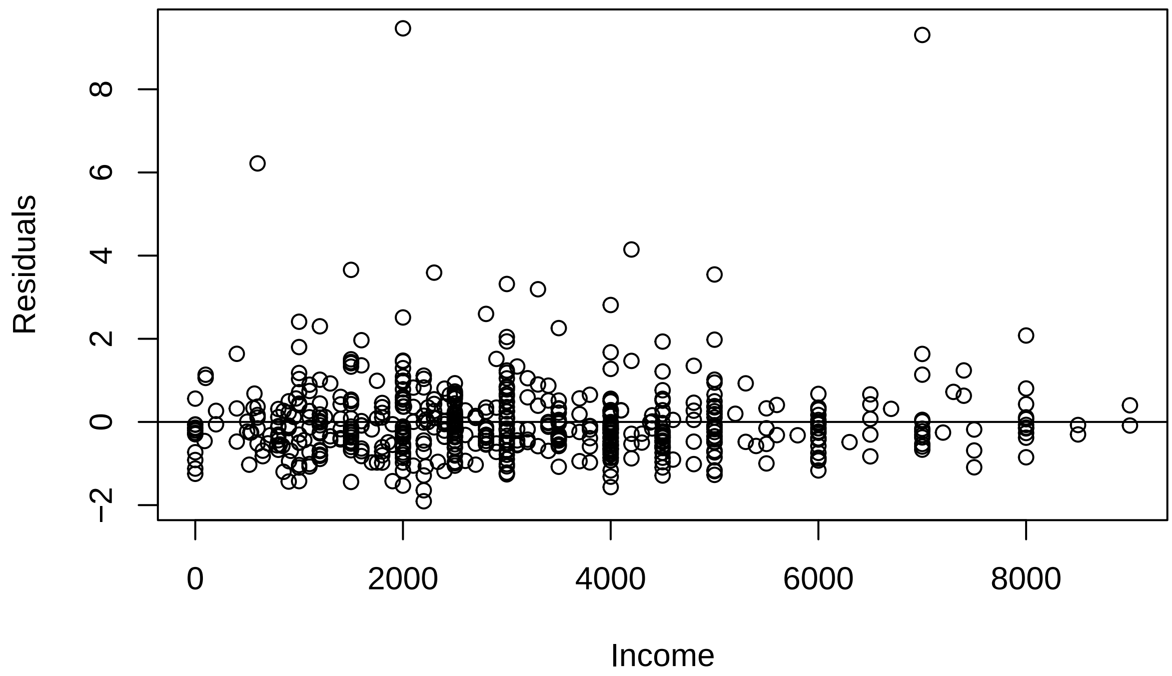
Scale−Location
107
216
109
0.0
0.5
1.0
1.5
2.0
2.5
3.0
√|Standardized residuals|
1000
1500
2000
2500
3000
3500
Fitted values
lm(CO2_housing_electricity ~ age + income + political_party + education + u ...



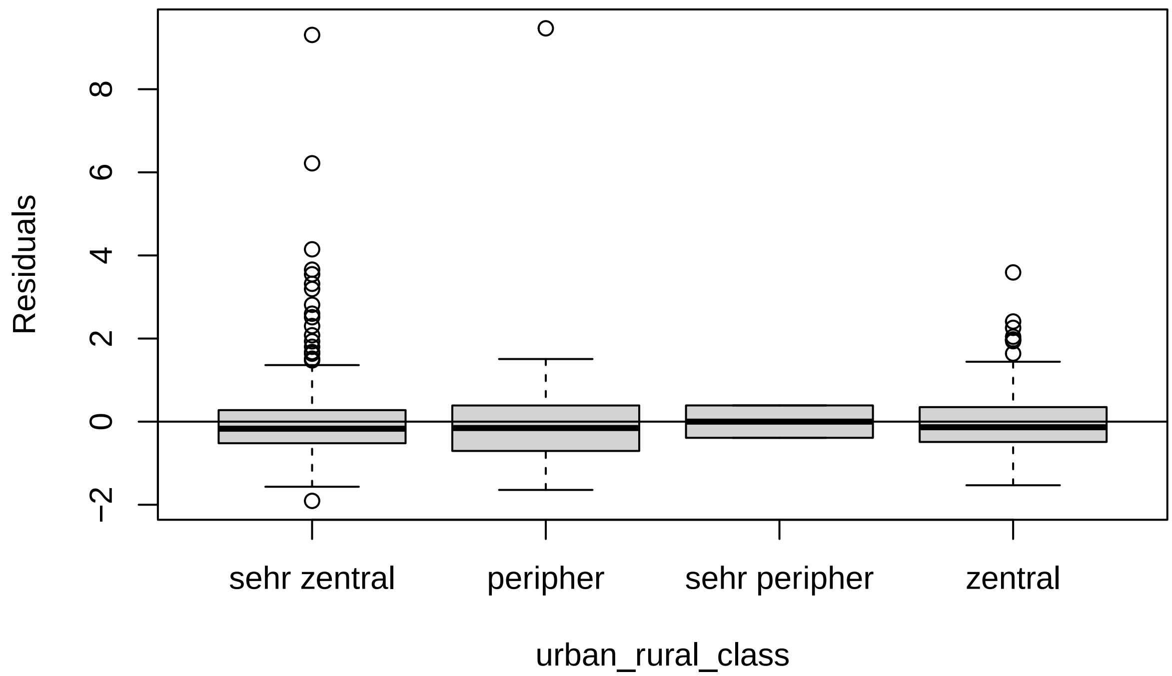

```
res1 = stdres(model1) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

plot(df1$age, res1, xlab = "Age", ylab = "Residuals")
abline(h = 0)
```



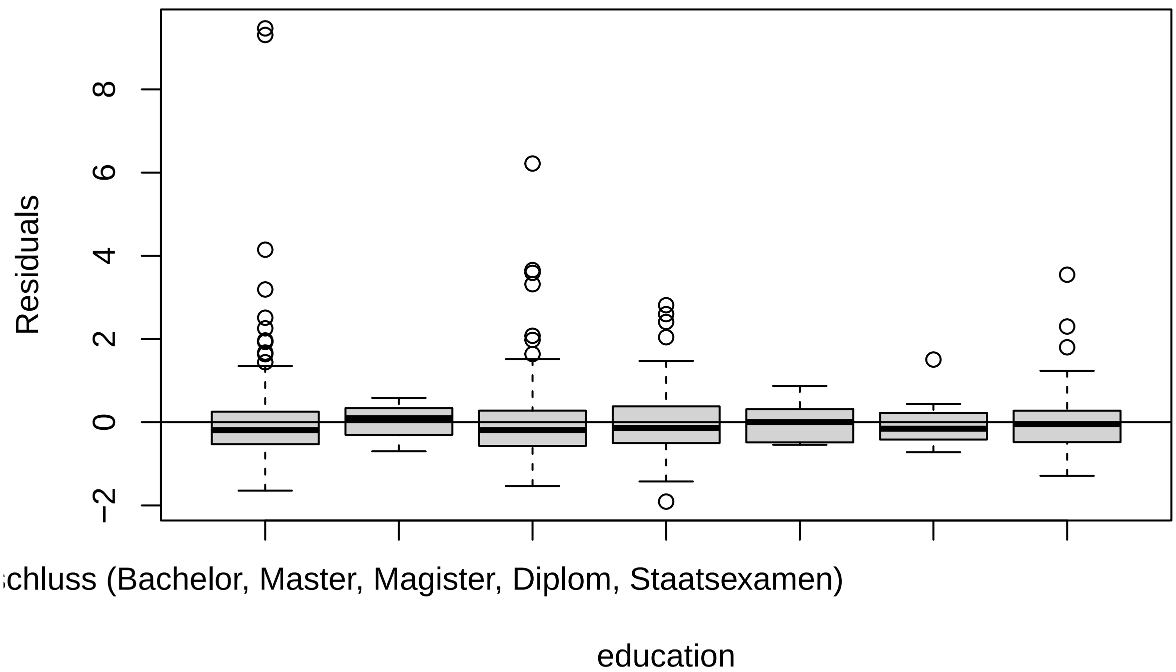

```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



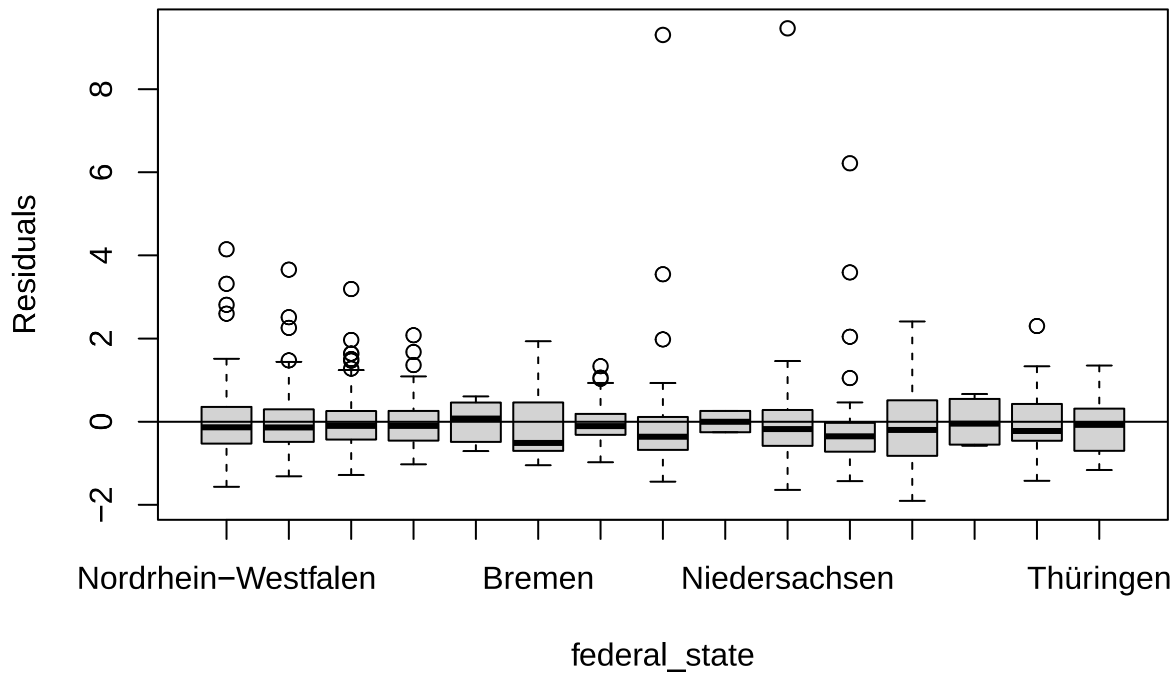

```
plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
abline(h = 0)
```



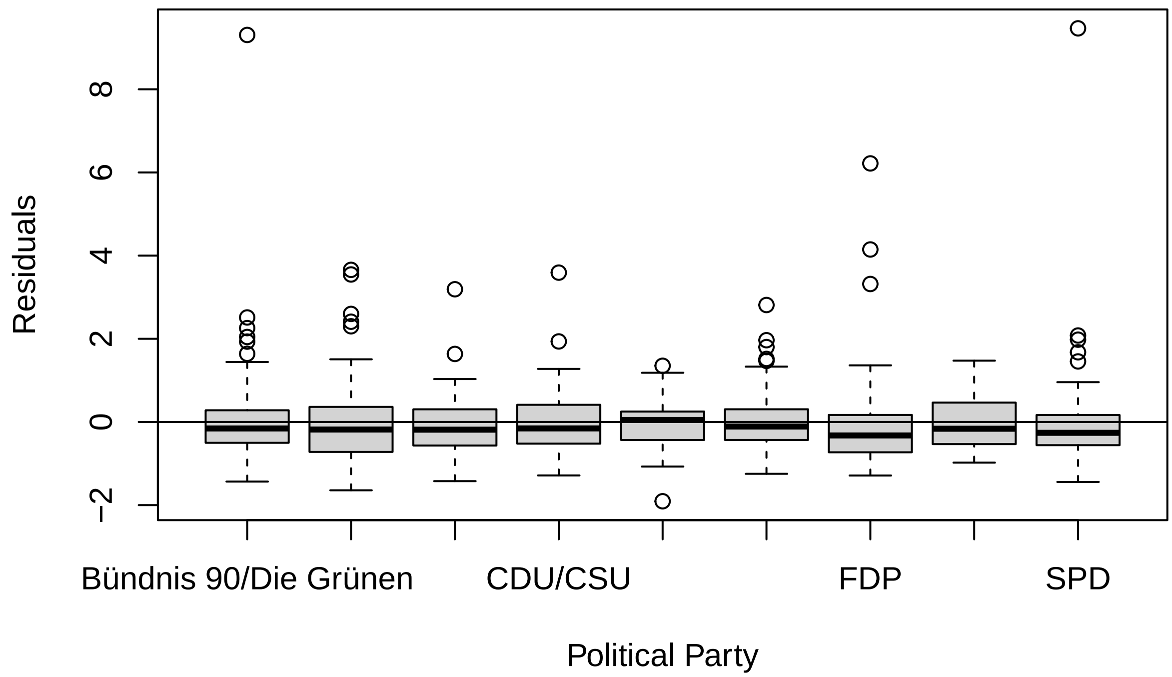

```
plot(df1$education, res1, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



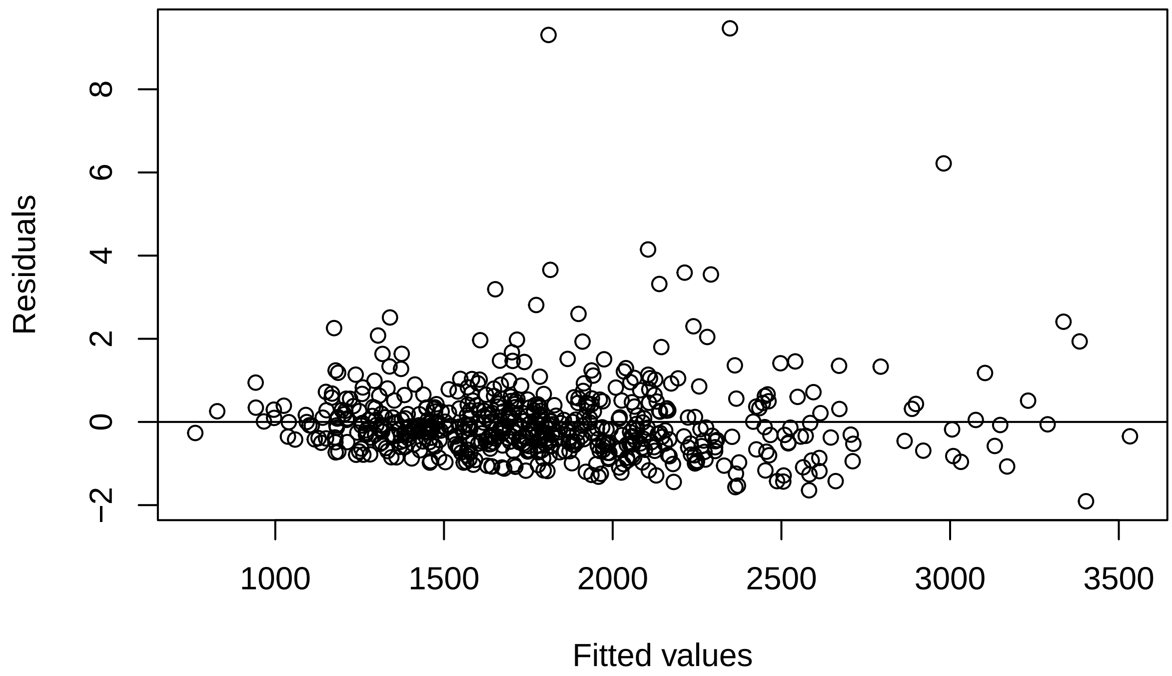

```
plot(df1$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```



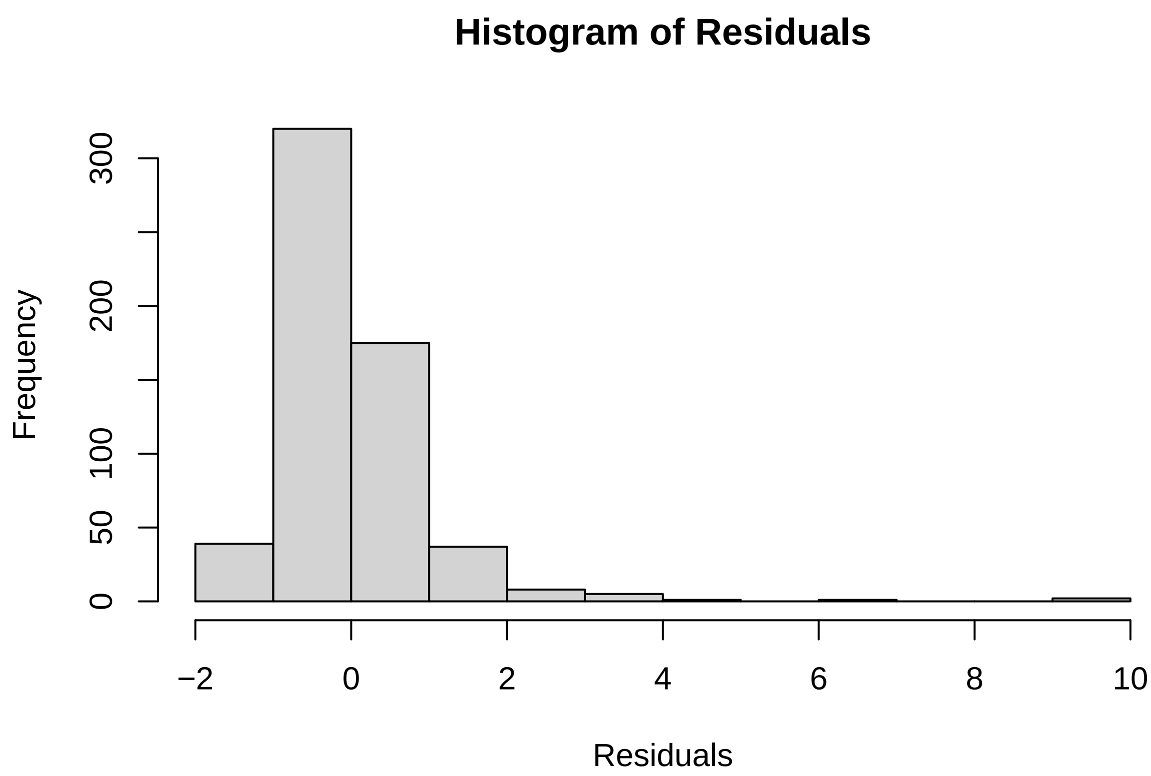

```
plot(df1$political_party, res1, xlab = "Political Party", ylab = "Residuals")
abline(h = 0)
```




```
# Constant variance and independent error term assumption

plot(fitted(model1), res1, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```




```
# Normality assumption

hist(res1, xlab="Residuals", main= "Histogram of Residuals")
```




```r
### Backward regression using AIC: starting with all of the variables

step_model1 <- stepAIC(model1, trace=TRUE, direction= "backward")
```

3. Variable Selection, model outcome and assumption check

```
## Start:  AIC=8630.84
## CO2_housing_electricity ~ age + income + political_party + education +
##     urban_rural_class + federal_state
##
##                     Df Sum of Sq        RSS    AIC
## - education          6   4723490 1245963158 8621.1
## - political_party    8  15175336 1256415003 8622.0
## <none>                           1241239667 8630.8
## - income             1   5863257 1247102924 8631.6
## - urban_rural_class  3  14793184 1256032851 8631.8
## - age                1  14318420 1255558087 8635.6
## - federal_state     14  71384514 1312624181 8635.7
##
## Step:  AIC=8621.07
## CO2_housing_electricity ~ age + income + political_party + urban_rural_class +
##     federal_state
##
```

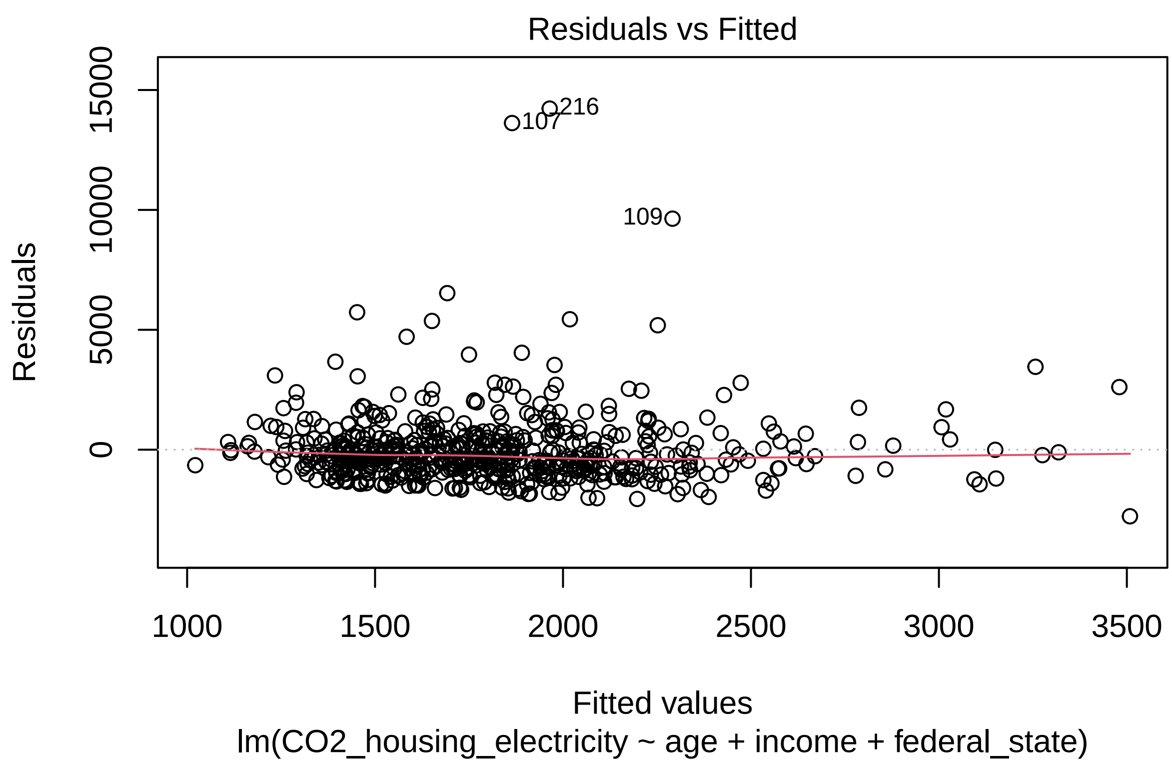
```
##                     Df Sum of Sq        RSS    AIC
## - political_party    8  14763817 1260726974 8612.0
## <none>                           1245963158 8621.1
## - urban_rural_class  3  13808719 1259771877 8621.6
## - income             1   5688390 1251651548 8621.7
## - age                1  12719286 1258682444 8625.0
## - federal_state     14  70692591 1316655748 8625.5
##
## Step:  AIC=8612
## CO2_housing_electricity ~ age + income + urban_rural_class +
##     federal_state
##
##                     Df Sum of Sq        RSS    AIC
## - urban_rural_class  3  12154533 1272881508 8611.6
## <none>                           1260726974 8612.0
## - income             1   4948920 1265675894 8612.3
## - federal_state     14  67372513 1328099487 8614.6
## - age                1  13065133 1273792107 8616.1
##
## Step:  AIC=8611.64
## CO2_housing_electricity ~ age + income + federal_state
##
##                 Df Sum of Sq        RSS    AIC
## <none>                       1272881508 8611.6
## - income         1   4916171 1277797679 8611.9
## - federal_state 14  65378580 1338260087 8613.1
## - age            1  13498518 1286380025 8615.8
```

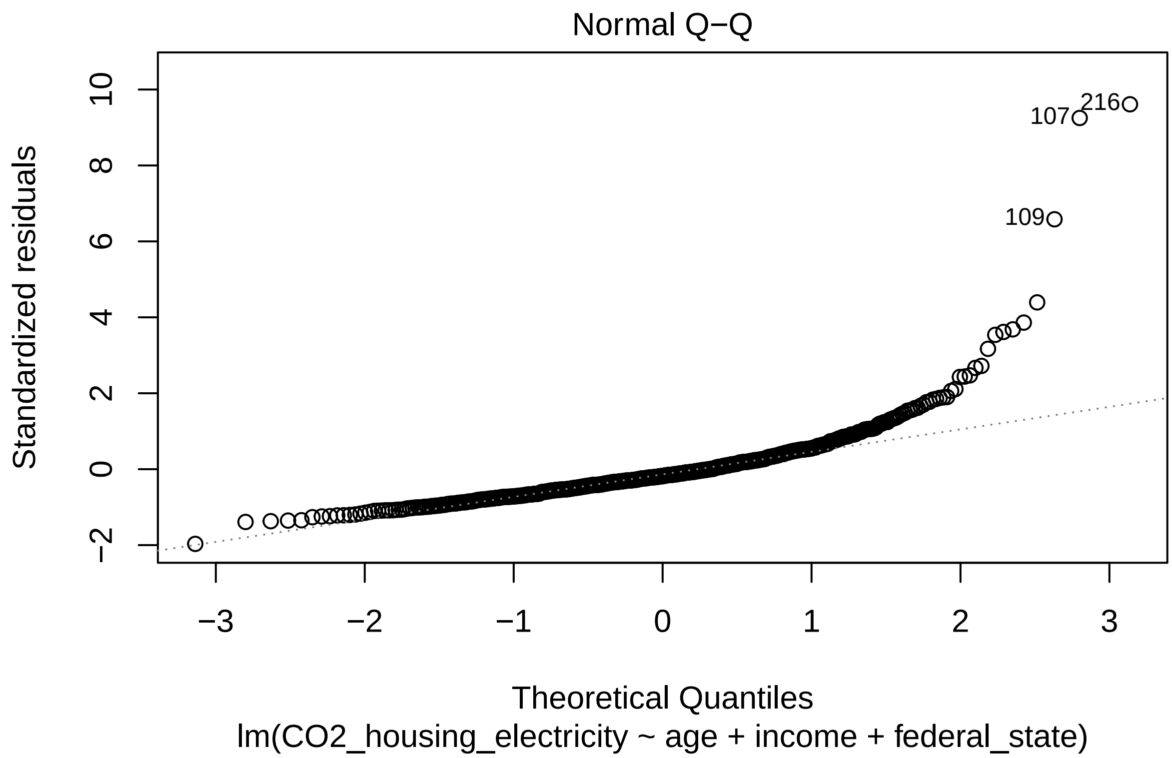
```
summary(step_model1)
```

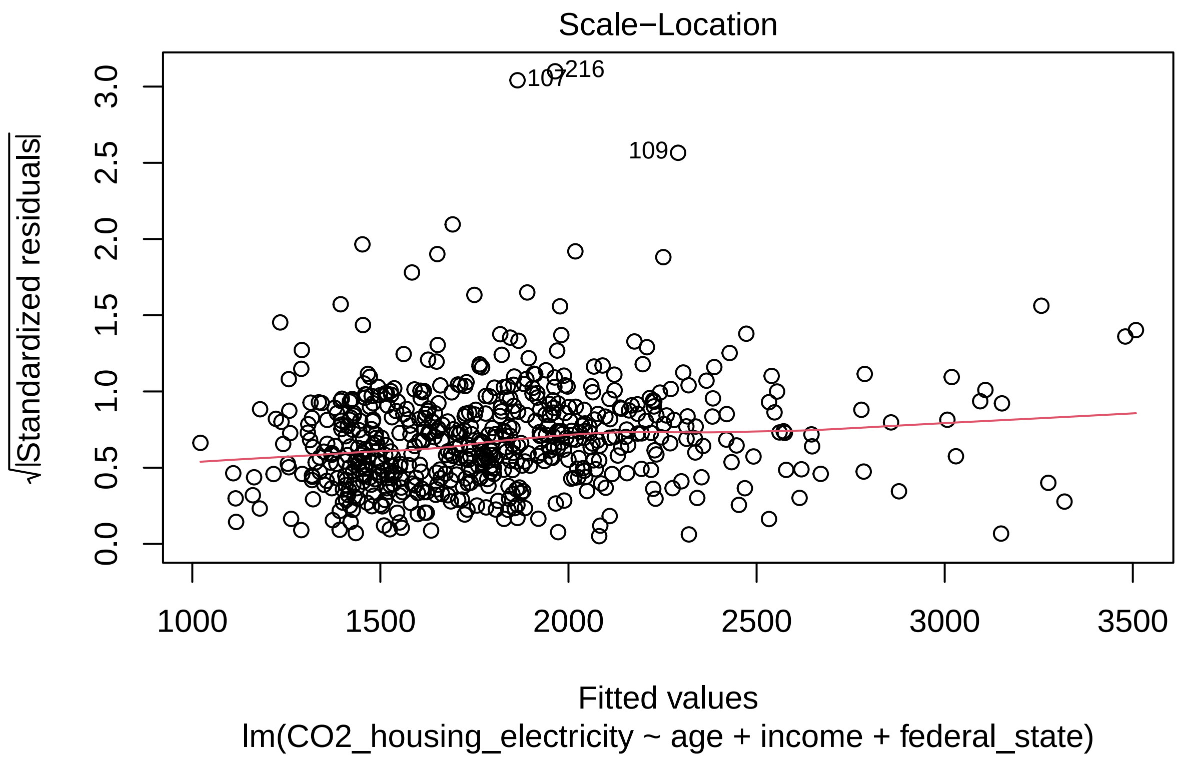
```
##
## Call:
## lm(formula = CO2_housing_electricity ~ age + income + federal_state,
##     data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2777.0  -790.1  -270.1   388.5 14224.0
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       1566.08603  235.44154   6.652 6.79e-11 ***
## age                                 11.08721    4.50563   2.461  0.01416 *
## income                              -0.04919    0.03312  -1.485  0.13809
## federal_stateBaden-Württemberg    -361.59221  206.92024  -1.747  0.08109 .
## federal_stateBayern               -185.65602  204.01227  -0.910  0.36319
## federal_stateBerlin               -149.43881  264.20237  -0.566  0.57187
## federal_stateBrandenburg          -288.91151  546.44541  -0.529  0.59721
## federal_stateBremen                358.99164  409.57688   0.876  0.38113
## federal_stateHamburg              -312.66275  329.59330  -0.949  0.34321
## federal_stateHessen                287.88789  252.48782   1.140  0.25468
## federal_stateMecklenburg-Vorpommern -591.65061 1067.01177  -0.554  0.57946
## federal_stateNiedersachsen         108.76419  240.33453   0.453  0.65104
## federal_stateRheinland-Pfalz       510.89852  305.78884   1.671  0.09532 .
```

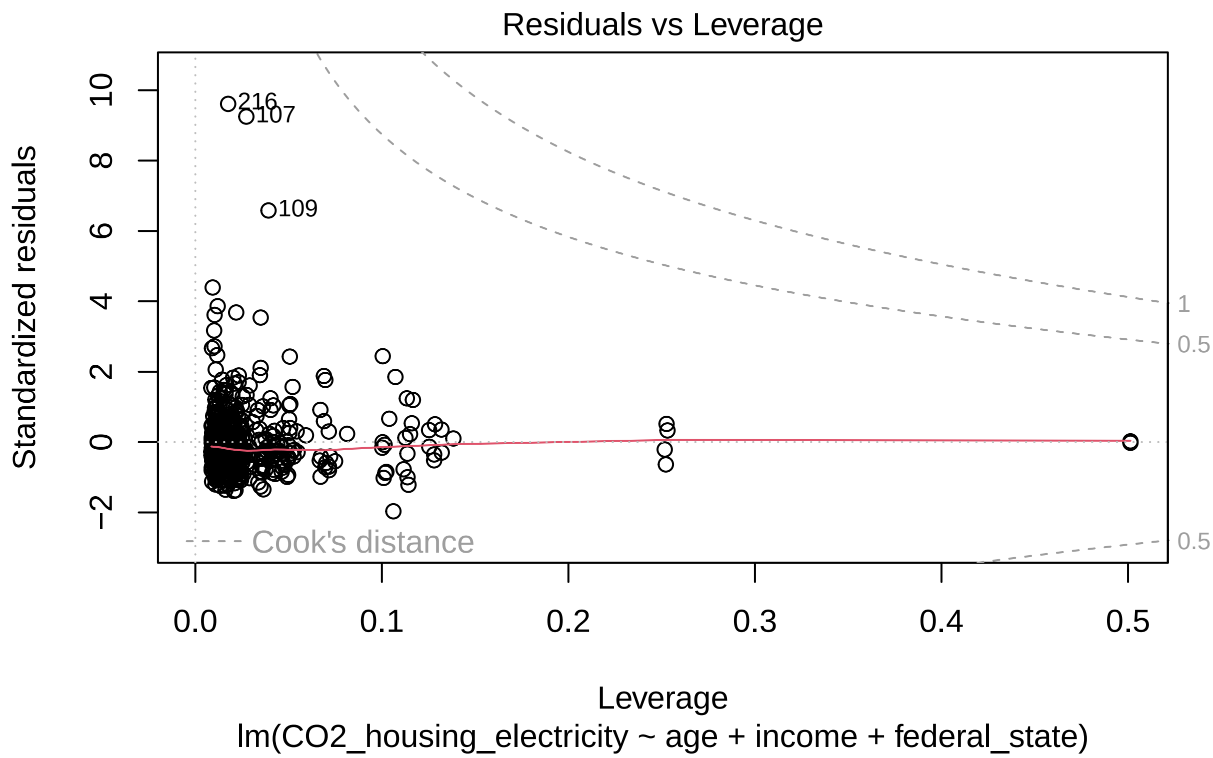
```
## federal_stateSaarland               1340.90572  493.76640   2.716  0.00681 **
## federal_stateSachsen-Anhalt         1033.62431  760.13352   1.360  0.17443
## federal_stateSchleswig-Holstein      237.61095  347.47597   0.684  0.49437
## federal_stateThüringen               880.97153  517.04822   1.704  0.08895 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1493 on 571 degrees of freedom
## Multiple R-squared:  0.06541,    Adjusted R-squared:  0.03923
## F-statistic: 2.498 on 16 and 571 DF,  p-value: 0.001053
```

```
plot(step_model1)
```

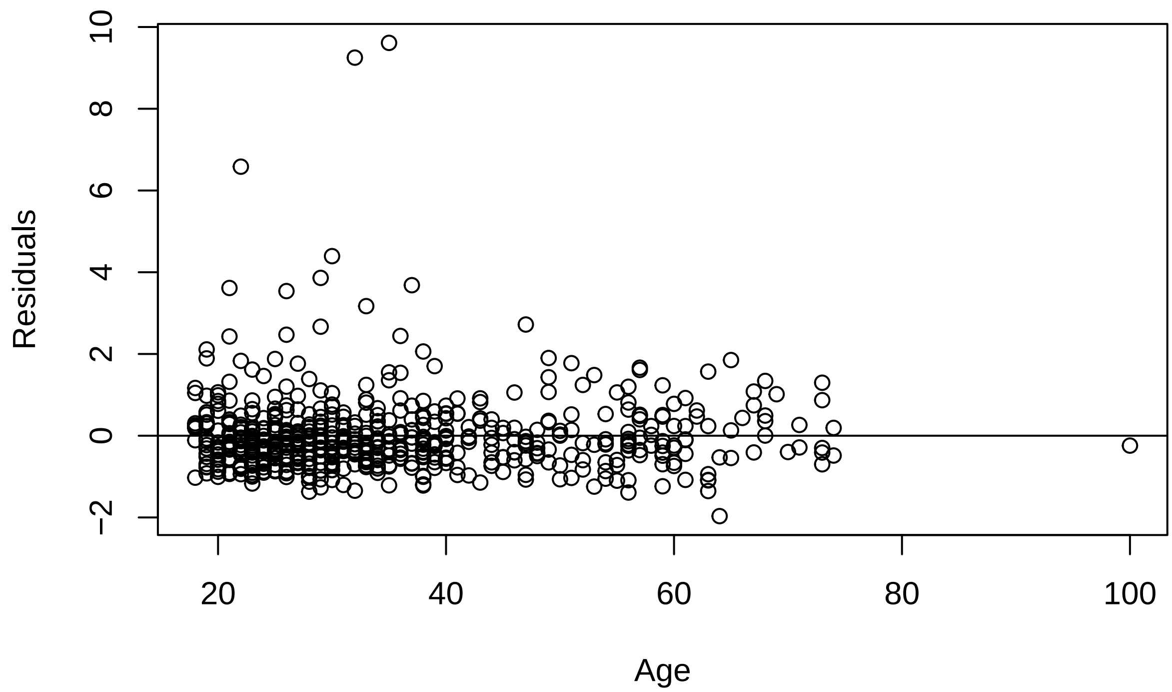



Normal Q−Q
Standardized residuals
10
8
6
4
2
0
−2
216
107
109
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(CO2_housing_electricity ~ age + income + federal_state)

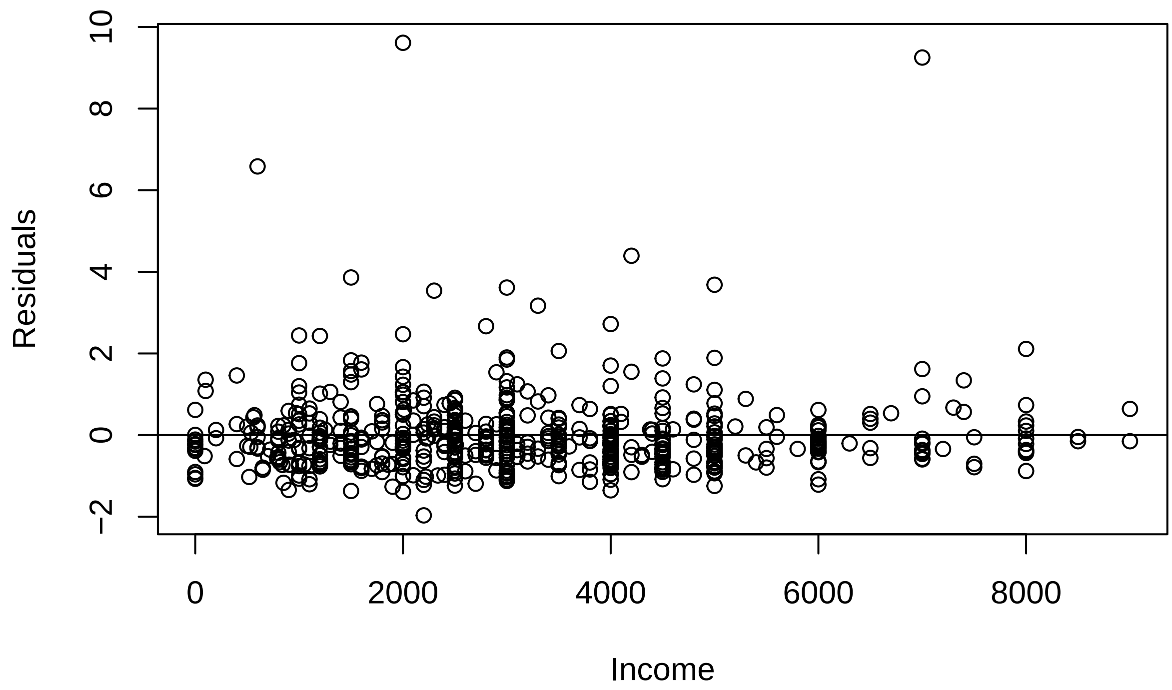

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
107
216
109
1000
1500
2000
2500
3000
3500
Fitted values
lm(CO2_housing_electricity ~ age + income + federal_state)



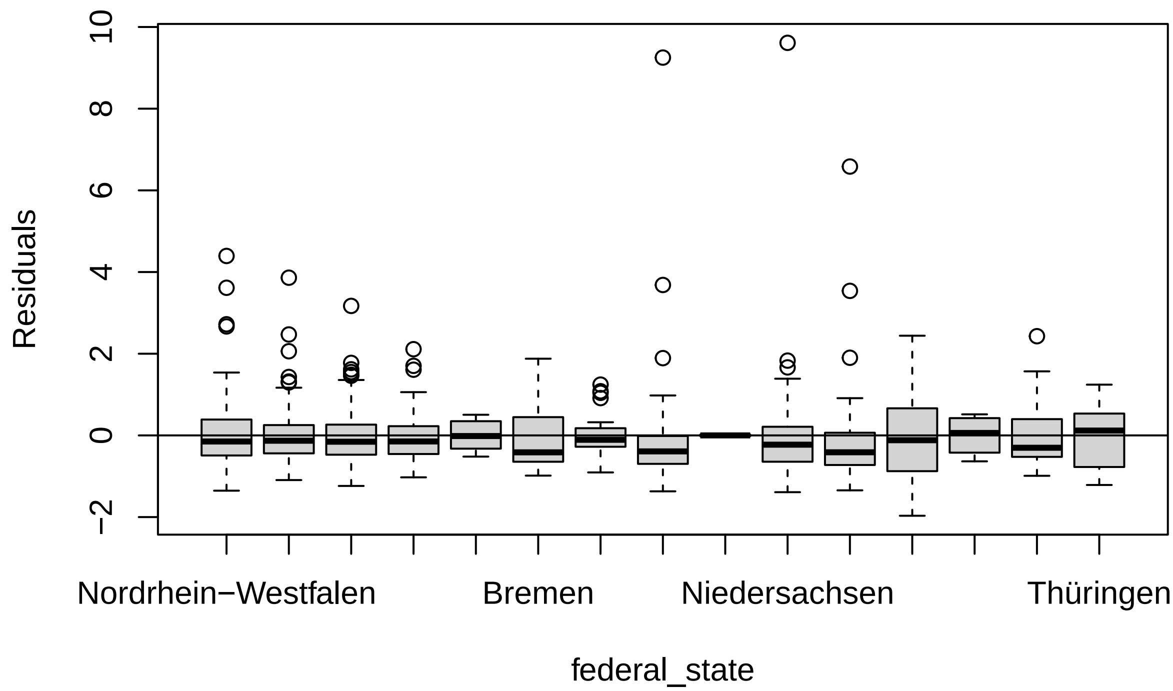

```r
res1 = stdres(step_model1) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

plot(df1$age, res1, xlab = "Age", ylab = "Residuals")
abline(h = 0)
```




```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```

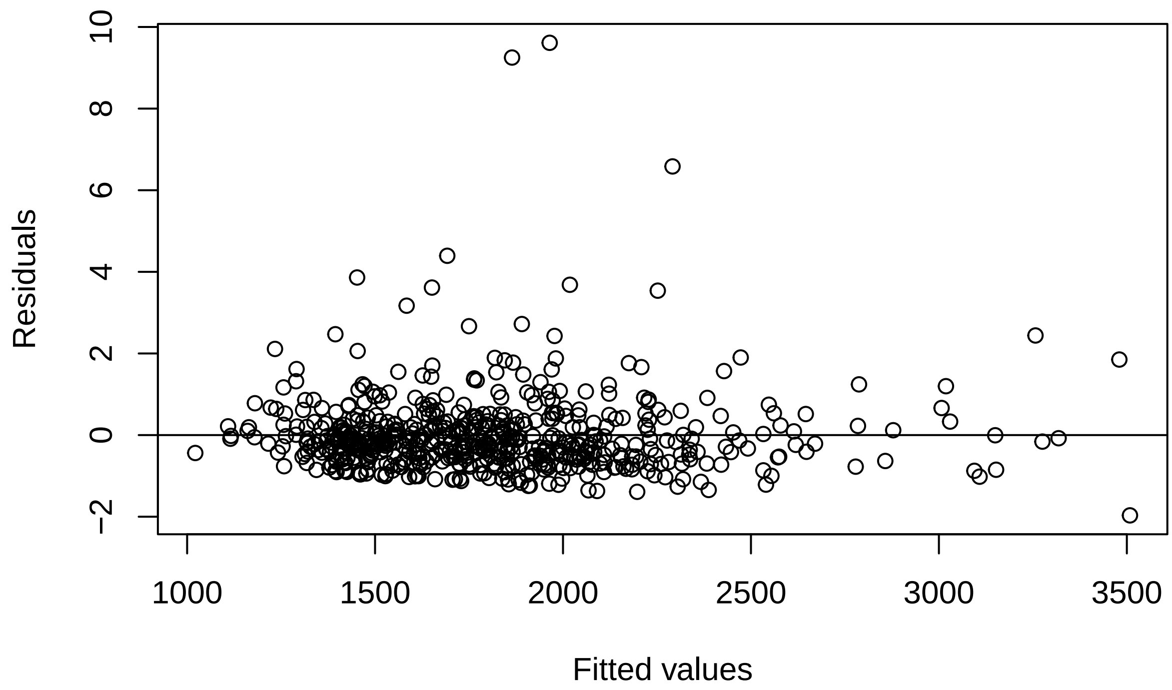



```
#plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
#abline(h = 0)

#plot(df1_scaled$education, res1, xlab = "education", ylab = "Residuals")
#abline(h = 0)

plot(df1$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```

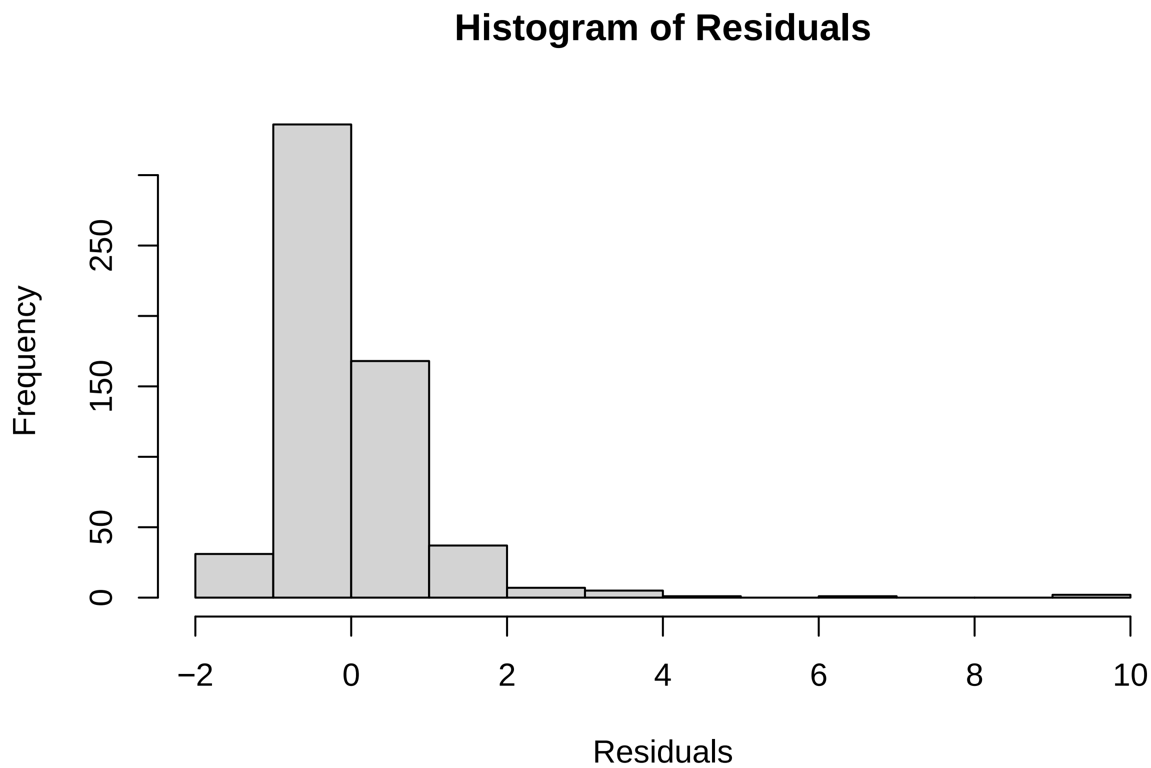



```
#plot(df1_scaled$political_party, res1, xlab = "Political Party", ylab = "Residuals")
#abline(h = 0)
```

```
# Constant variance and independent error term assumption

plot(fitted(step_model1), res1, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```



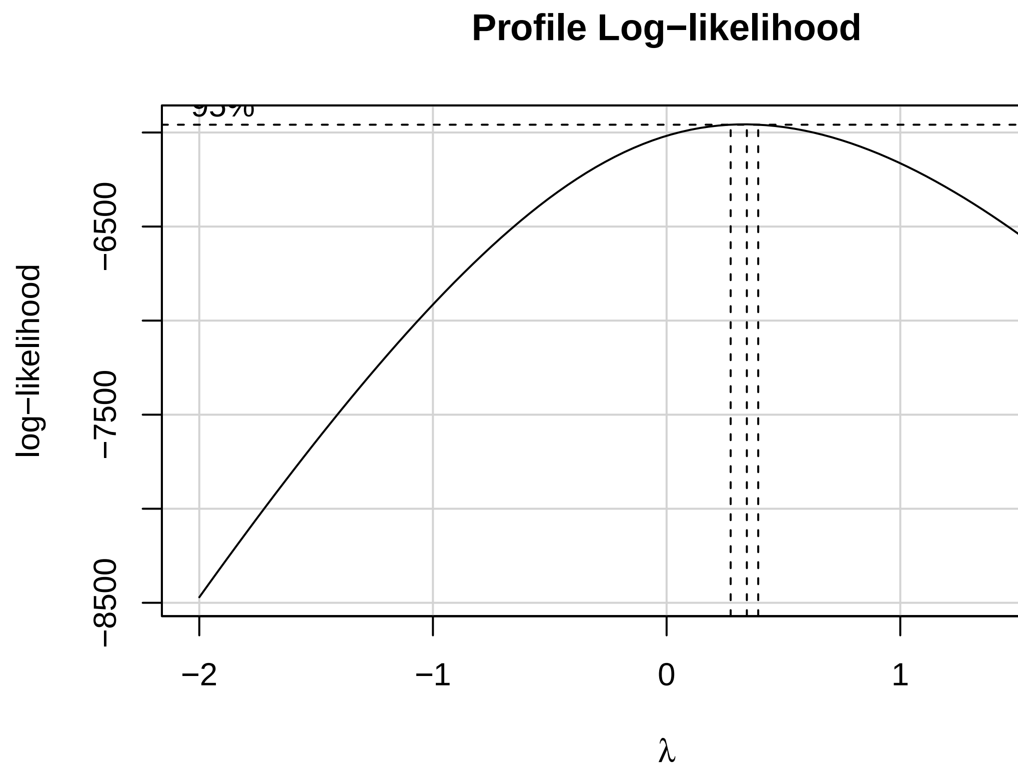

```
# Normality assumption

hist(res1, xlab="Residuals", main= "Histogram of Residuals")
```




```
## normality test using shapiro-test: reject the H0, not normally distributed
#H0:  the sample comes from a normal distribution

res1_num = res1[is.finite(res1)]

shapiro.test(res1_num)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res1_num
## W = 0.71431, p-value < 2.2e-16
```

```
# Box-cox transformation

bc = boxCox(step_model1)
```



4. Improving the regression fit

```
opt.lambda = bc$x[which.max(bc$y)]
round(opt.lambda/0.5)*0.5 # round it to the nearest 0.5
```

```
## [1] 0.5
```

FINAL MODEL

```
# Non-linear transformation with the lambda 0.5

options(scipen = -2)

model1_trans = lm(sqrt(CO2_housing_electricity)  ~ age + income + federal_state, data = df1)

summary(model1_trans)
```

```
##
## Call:
## lm(formula = sqrt(CO2_housing_electricity) ~ age + income + federal_state,
##     data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.436  -8.123  -0.712   6.929  86.427
```

```
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         3.715e+01  2.331e+00  15.938  < 2e-16 ***
## age                                 1.379e-01  4.460e-02   3.093 2.08e-03 **
## income                             -6.080e-04  3.279e-04  -1.854 6.42e-02 .
## federal_stateBaden-Württemberg     -5.137e+00  2.048e+00  -2.508 1.24e-02 *
## federal_stateBayern                -2.337e+00  2.019e+00  -1.157 2.48e-01
## federal_stateBerlin                -2.027e+00  2.615e+00  -0.775 4.39e-01
## federal_stateBrandenburg           -1.917e+00  5.409e+00  -0.354 7.23e-01
## federal_stateBremen                 4.646e+00  4.054e+00   1.146 2.52e-01
## federal_stateHamburg               -3.296e+00  3.263e+00  -1.010 3.13e-01
## federal_stateHessen                 2.198e+00  2.499e+00   0.880 3.79e-01
## federal_stateMecklenburg-Vorpommern -5.289e+00  1.056e+01  -0.501 6.17e-01
## federal_stateNiedersachsen          5.000e-02  2.379e+00   0.021 9.83e-01
## federal_stateRheinland-Pfalz        4.405e+00  3.027e+00   1.455 1.46e-01
## federal_stateSaarland               1.334e+01  4.888e+00   2.730 6.53e-03 **
## federal_stateSachsen-Anhalt         1.295e+01  7.524e+00   1.721 8.59e-02 .
## federal_stateSchleswig-Holstein     3.160e+00  3.440e+00   0.919 3.59e-01
## federal_stateThüringen              9.924e+00  5.118e+00   1.939 5.30e-02 .
## ---
## Signif. codes:  0 '***' 1e-03 '**' 1e-02 '*' 5e-02 '.' 0.1 ' ' 1
##
## Residual standard error: 14.78 on 571 degrees of freedom
## Multiple R-squared:  0.08588,    Adjusted R-squared:  0.06026
## F-statistic: 3.353 on 16 and 571 DF,  p-value: 1.139e-05
```

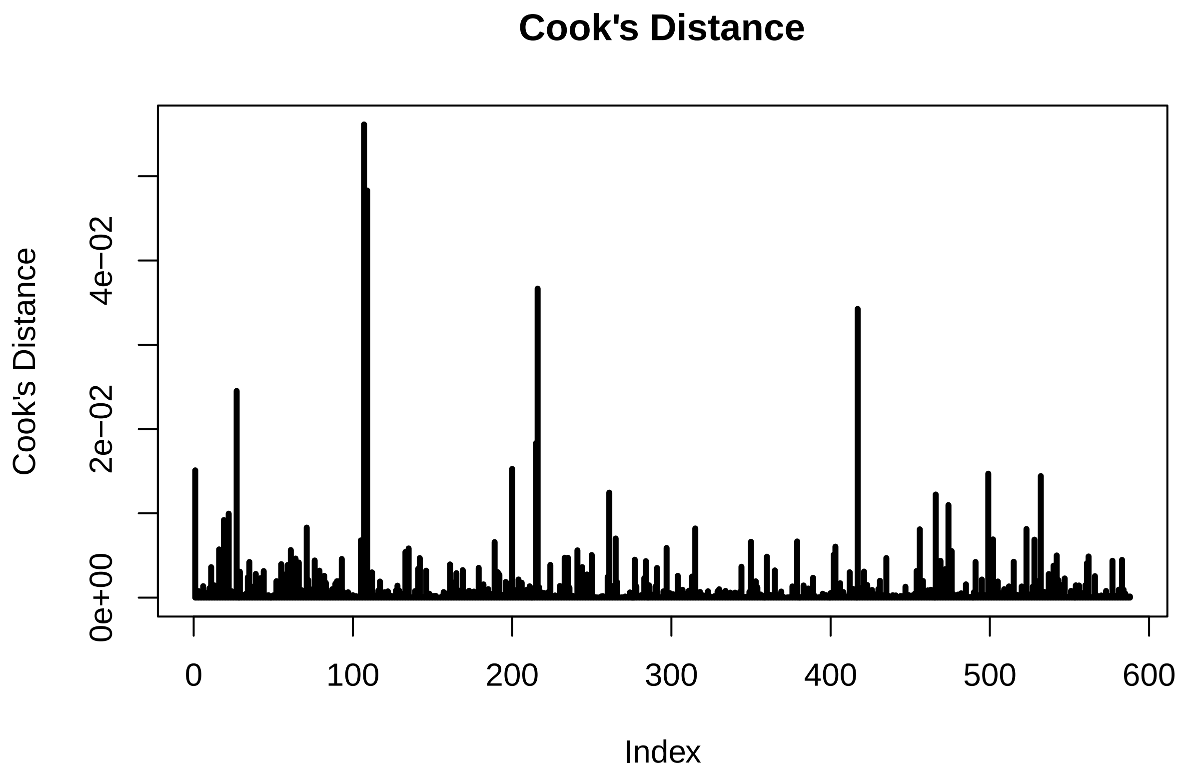
```r
# Checking the VIFs for multicollinearity

vif(model1_trans)
```

```
##                   GVIF Df GVIF^(1/(2*Df))
## age           1.021943  1        1.010912
## income        1.035322  1        1.017508
## federal_state 1.056661 14        1.001970
```

```r
# threshold for multicollinearity
# Calculating the threshold

max(10, 1/(1-summary(model1_trans)$r.square))
```

```
## [1] 10
```

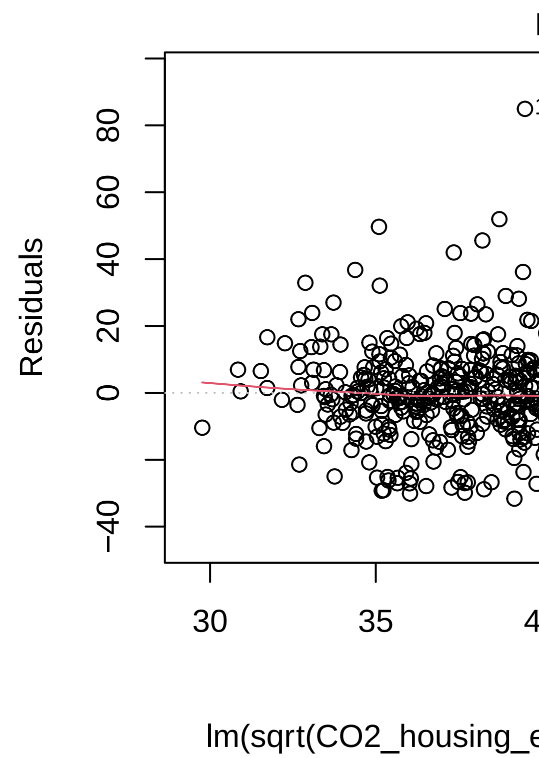
```r
# Checking outliers: estimate of the influence of data point; summary of how much a regression model ch

cook = cooks.distance(model1_trans)
plot(cook,
     type="h",
     lwd=3,
     ylab = "Cook's Distance",
     main="Cook's Distance")
abline(h = 1)
```



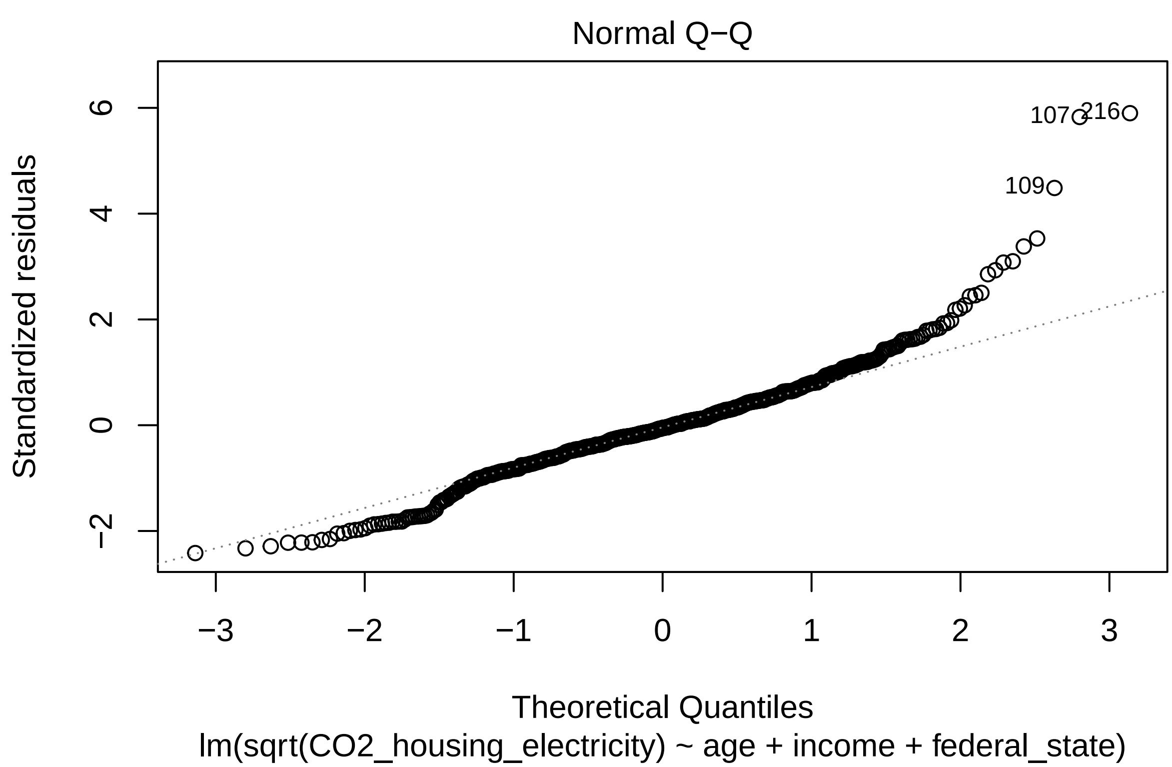

```
influential = cooks.distance(model1_trans)[which(cook >1)]

influential
```

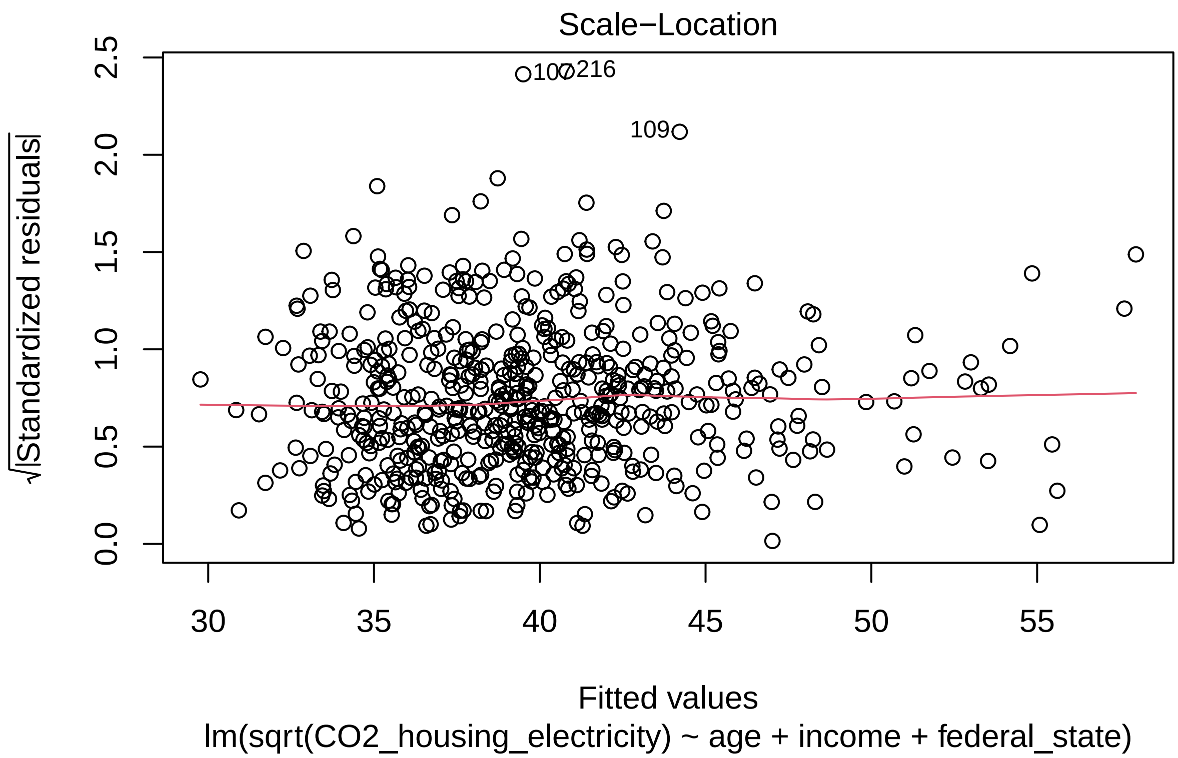
```
## named numeric(0)
```

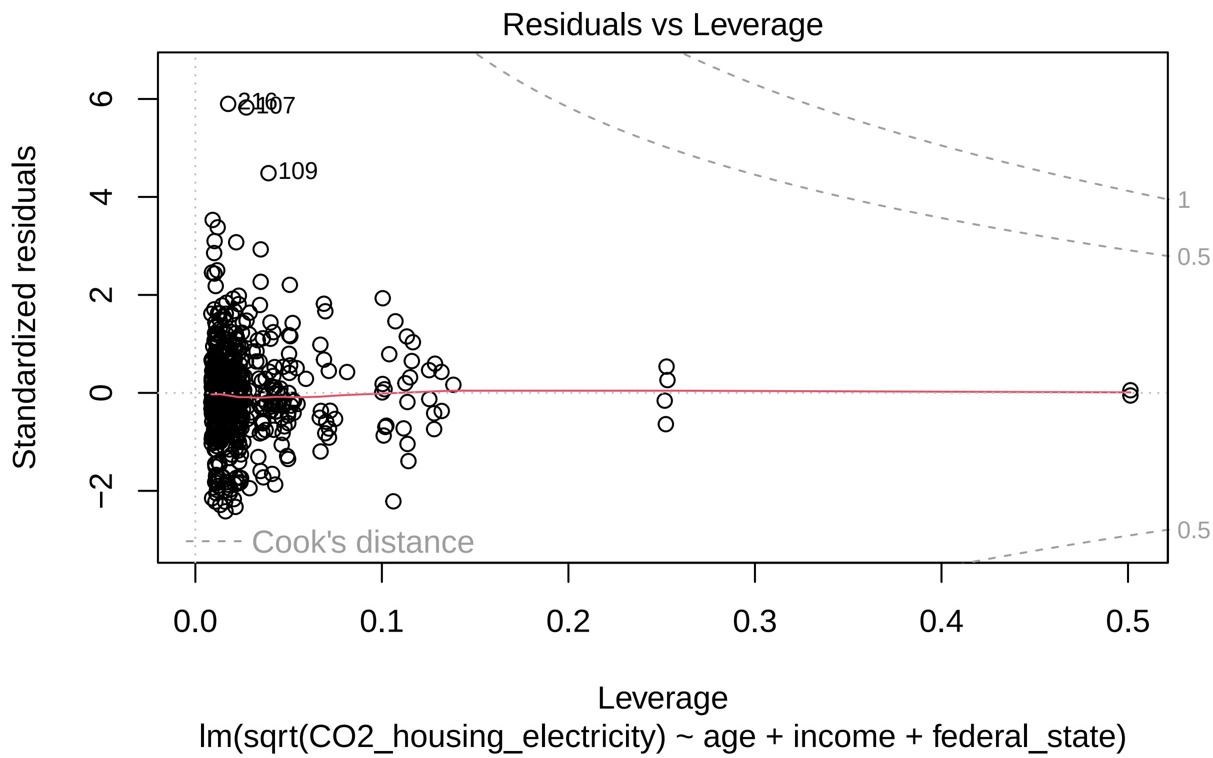
```
plot(model1_trans)
```



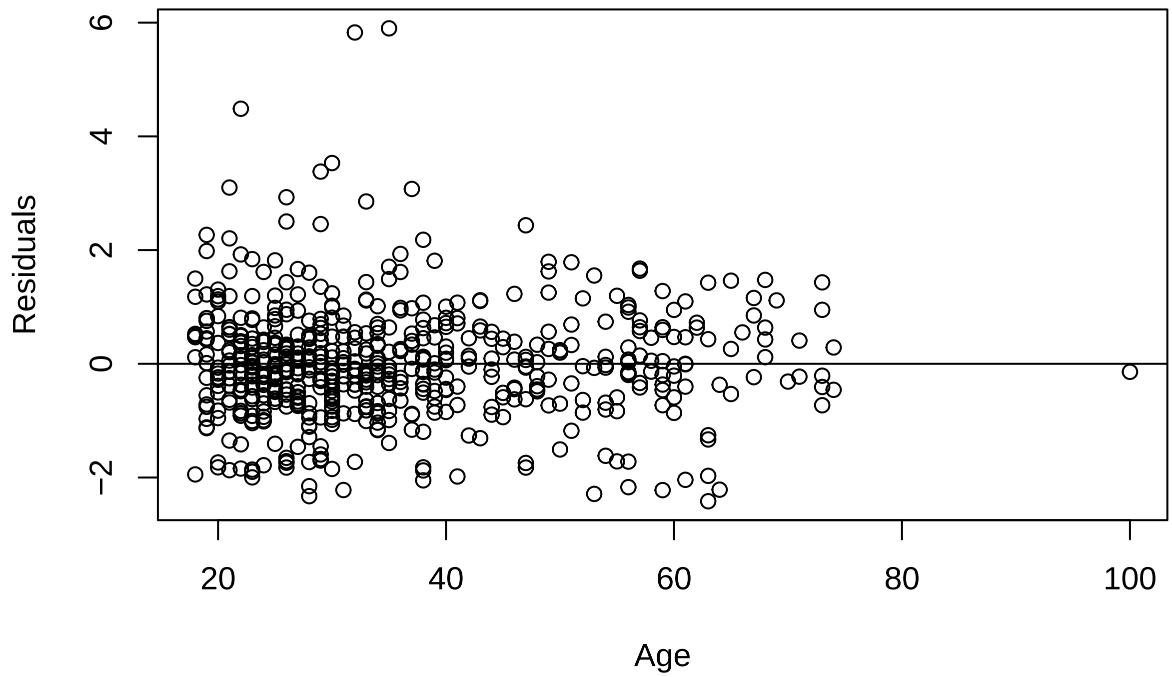
5. Assumptions check in the residuals of the transformed regression



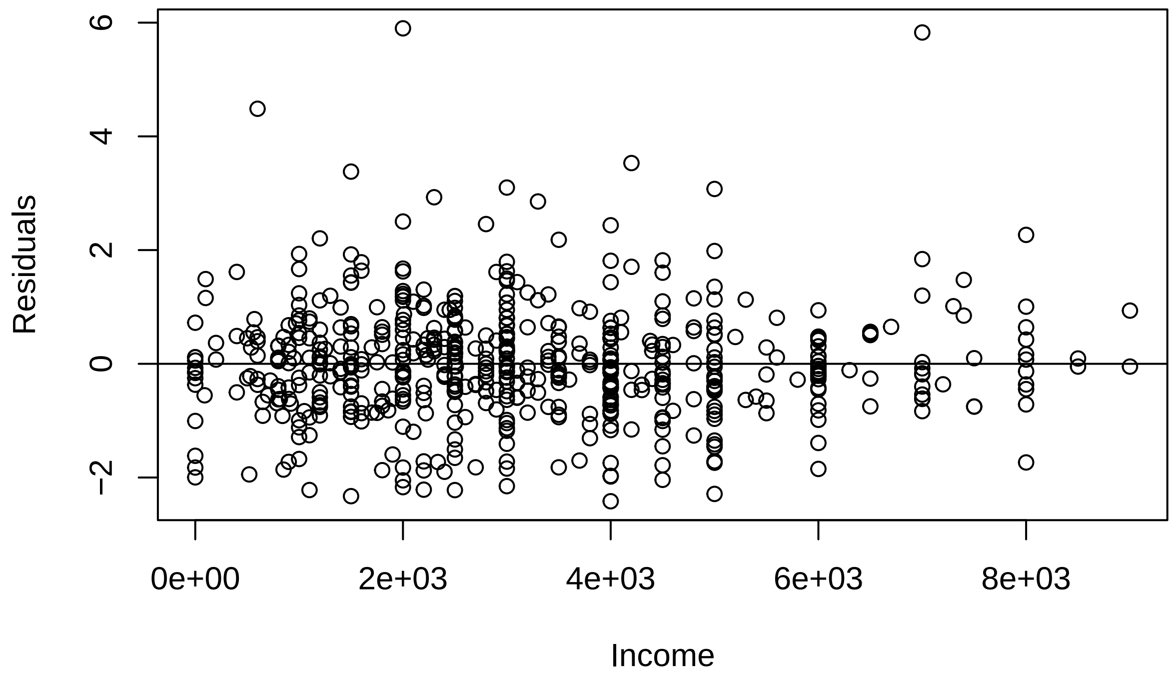

Scale−Location
107
216
109
2.5
2.0
1.5
1.0
0.5
0.0
√|Standardized residuals|
30
35
40
45
50
55
Fitted values
lm(sqrt(CO2_housing_electricity) ~ age + income + federal_state)



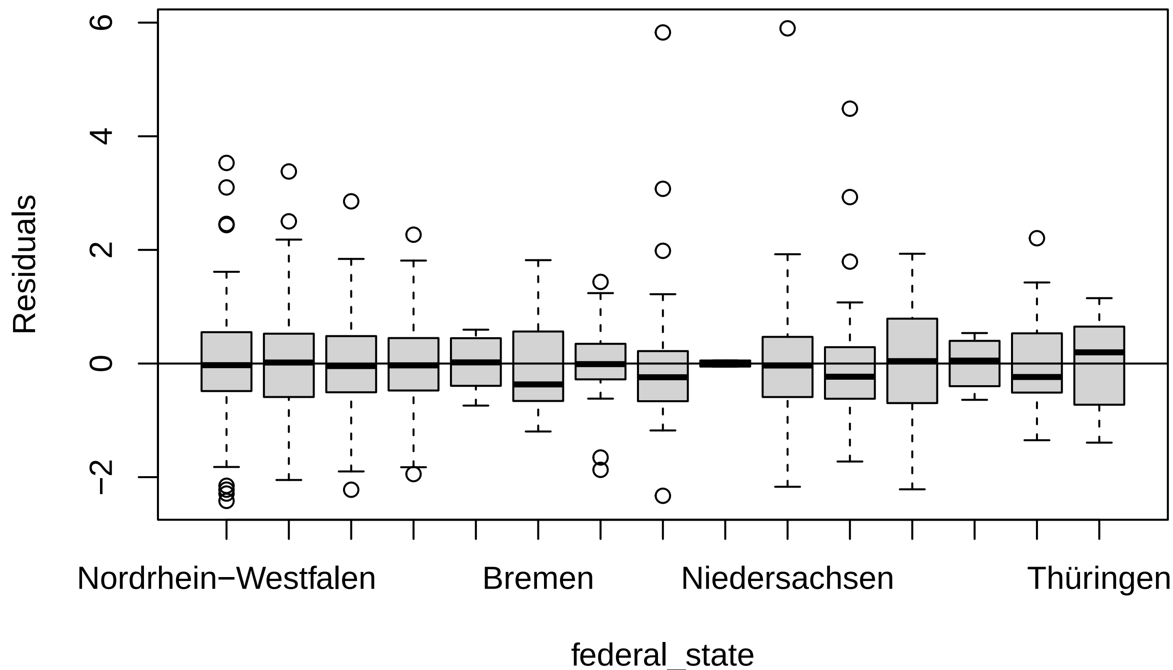

```r
res1 = stdres(model1_trans) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

plot(df1$age, res1, xlab = "Age", ylab = "Residuals")
abline(h = 0)
```




```r
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```




```
#plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
#abline(h = 0)

#plot(df1$education, res1, xlab = "education", ylab = "Residuals")
#abline(h = 0)

plot(df1$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```




```r
#plot(df1$political_party, res1, xlab = "Political Party", ylab = "Residuals")
#abline(h = 0)
```

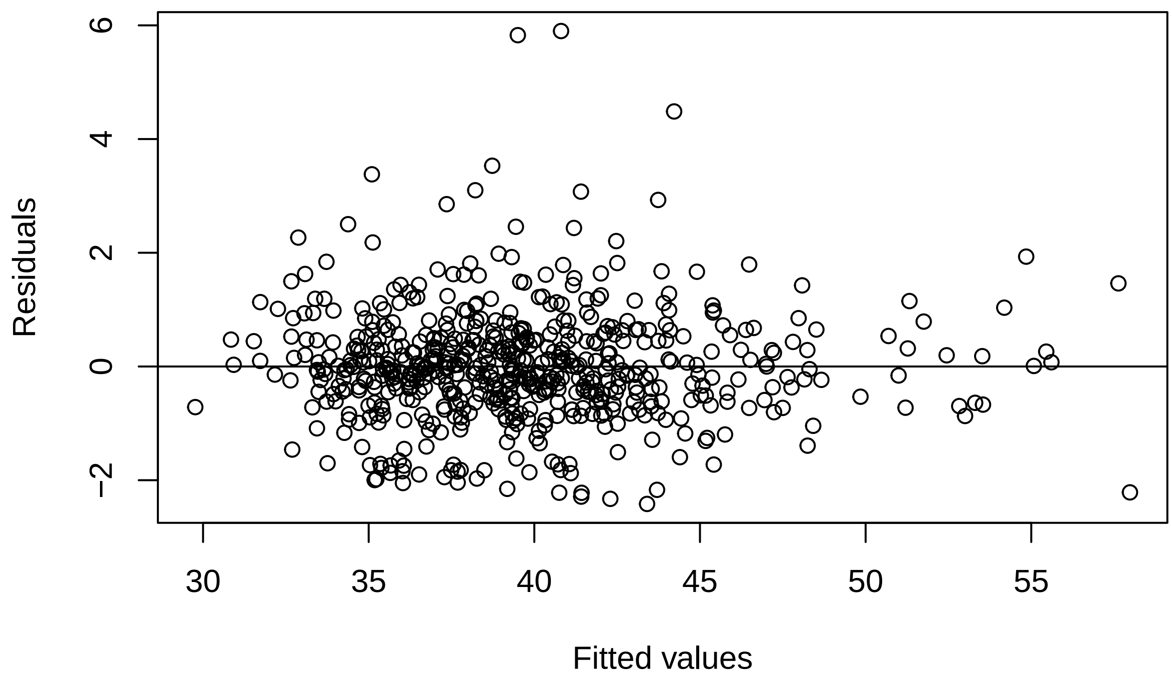
```r
# Durbin-Watson Test: Independence of the error terms
# H0 (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(model1_trans)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.04184228      1.912674     0.3
##  Alternative hypothesis: rho != 0
```

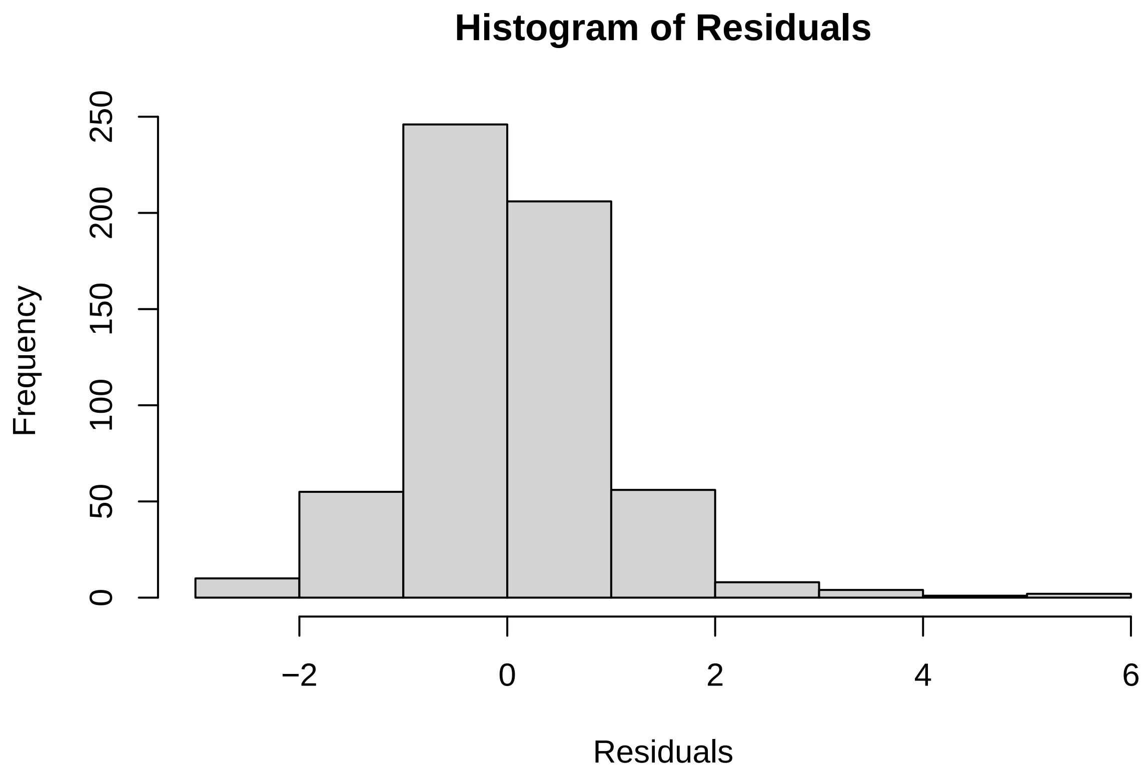
```r
# Breusch-Pagan TEST: Heteroscedasticity
# H0: Homoscedasticity is present

bptest(model1_trans)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1_trans
## BP = 7.5682, df = 16, p-value = 0.9607
```

```
# Constant variance and independent error term assumption

plot(fitted(model1_trans), res1, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```




```
# Normality assumption

hist(res1, xlab="Residuals", main= "Histogram of Residuals")
```

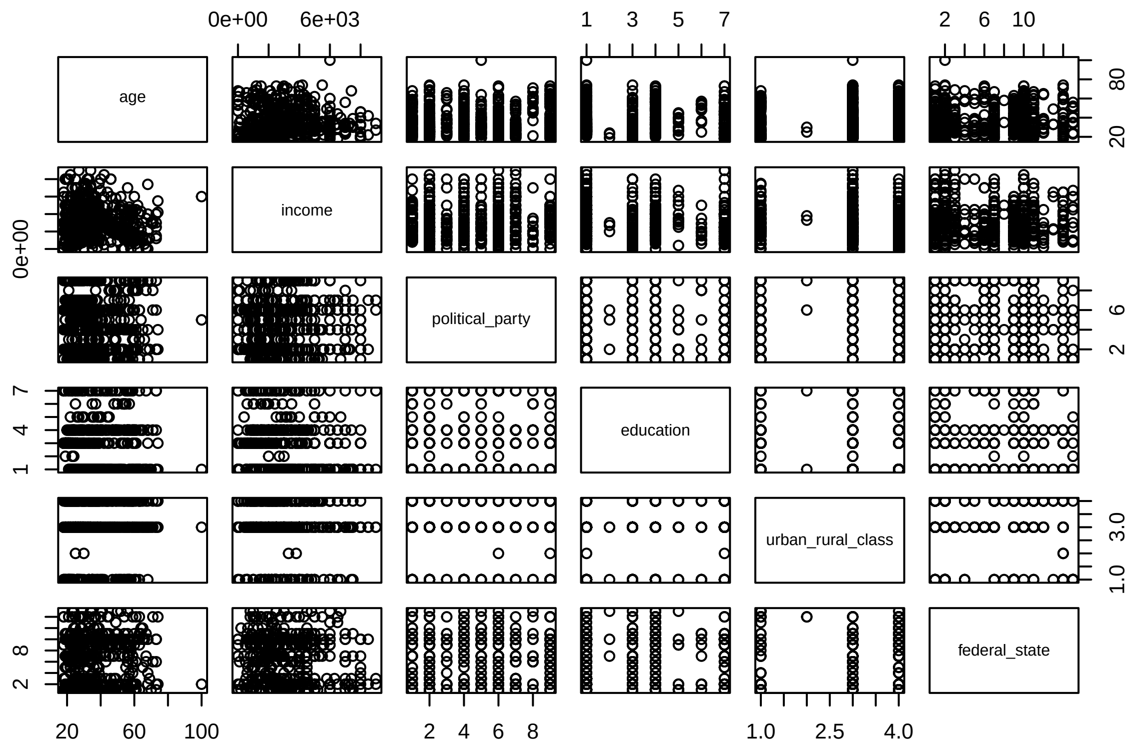



```
## normality test using shapiro-test: reject the H0
#H0:  the sample comes from a normal distribution

res1_num = res1[is.finite(res1)]

shapiro.test(res1_num)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res1_num
## W = 0.94014, p-value = 1.227e-14
```

III. Multivariate Regression: belief diff housing and electricity

```
# Checking the possible correlation in the data

plot(df2[1:6])
```



1. Modeling

```
## defining a reference level

df2$political_party  <- relevel(df2$political_party, ref='Bündnis 90/Die Grünen')
df2$education  <- relevel(df2$education, ref='(Fach-) Hochschulabschluss (Bachelor, Master, Magister, Di
df2$urban_rural_class  <- relevel(df2$urban_rural_class, ref='sehr zentral')
df2$federal_state  <- relevel(df2$federal_state, ref='Nordrhein-Westfalen')
```

```
# regression model

options(scipen=-0, digits=2)

model2 = lm(belief_diff_housing_electricity ~ age + income + political_party + education + urban_rural_

summary(model2)
```

FINAL MODEL

```
##
## Call:
## lm(formula = belief_diff_housing_electricity ~ age + income +
##     political_party + education + urban_rural_class + federal_state,
##     data = df2)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -87.23 -23.68  -0.92  21.89 101.41
##
## Coefficients:
##                                                                                   Estimate
## (Intercept)                                                                       1.15e+00
## age                                                                              -4.97e-01
## income                                                                            1.78e-03
## political_partyAfD                                                                2.24e-01
## political_partyBündnis Sarah Wagenknecht                                         -2.50e+00
## political_partyCDU/CSU                                                            2.48e+00
## political_partyDie Linke                                                         -5.09e-01
## political_partyEiner anderen Partei                                              -1.32e+00
## political_partyFDP                                                               -1.55e+00
## political_partyKeine Angabe                                                       7.48e+00
## political_partySPD                                                                4.39e+00
## education(Noch) kein Abschluss                                                    1.84e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -2.83e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule              2.02e+00
## educationDoktorgrad oder Habilitation                                            -5.53e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  1.35e+01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss       -1.45e+00
## urban_rural_classperipher                                                        -4.17e+00
## urban_rural_classsehr peripher                                                    1.25e+01
## urban_rural_classzentral                                                          4.51e+00
## federal_stateBaden-Württemberg                                                    1.24e+01
## federal_stateBayern                                                               7.10e+00
## federal_stateBerlin                                                               3.67e+00
## federal_stateBrandenburg                                                          1.29e+01
## federal_stateBremen                                                              -4.49e-01
## federal_stateHamburg                                                              6.45e-02
## federal_stateHessen                                                              -6.54e-01
## federal_stateMecklenburg-Vorpommern                                               2.08e+01
## federal_stateNiedersachsen                                                       -7.46e-01
## federal_stateRheinland-Pfalz                                                     -1.91e+00
## federal_stateSaarland                                                            -1.22e+01
## federal_stateSachsen-Anhalt                                                      -1.75e+01
## federal_stateSchleswig-Holstein                                                   3.40e+00
## federal_stateThüringen                                                           -1.40e+01
##                                                                                  Std. Error
## (Intercept)                                                                        6.45e+00
## age                                                                                1.14e-01
## income                                                                             7.64e-04
## political_partyAfD                                                                 5.47e+00
## political_partyBündnis Sarah Wagenknecht                                           7.66e+00
## political_partyCDU/CSU                                                             4.91e+00
## political_partyDie Linke                                                           5.92e+00
## political_partyEiner anderen Partei                                                4.39e+00
## political_partyFDP                                                                 5.69e+00
## political_partyKeine Angabe                                                        9.85e+00
## political_partySPD                                                                 5.00e+00
## education(Noch) kein Abschluss                                                     2.00e+01
```

```
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   3.89e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule               3.97e+00
## educationDoktorgrad oder Habilitation                                             9.70e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss    1.11e+01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         5.05e+00
## urban_rural_classperipher                                                         5.08e+00
## urban_rural_classsehr peripher                                                    2.52e+01
## urban_rural_classzentral                                                          3.73e+00
## federal_stateBaden-Württemberg                                                    4.84e+00
## federal_stateBayern                                                               4.97e+00
## federal_stateBerlin                                                               6.04e+00
## federal_stateBrandenburg                                                          1.27e+01
## federal_stateBremen                                                               9.25e+00
## federal_stateHamburg                                                              7.53e+00
## federal_stateHessen                                                               5.78e+00
## federal_stateMecklenburg-Vorpommern                                               2.42e+01
## federal_stateNiedersachsen                                                        5.88e+00
## federal_stateRheinland-Pfalz                                                      7.25e+00
## federal_stateSaarland                                                             1.14e+01
## federal_stateSachsen-Anhalt                                                       1.75e+01
## federal_stateSchleswig-Holstein                                                   8.48e+00
## federal_stateThüringen                                                            1.29e+01
##                                                                                 t value
## (Intercept)                                                                         0.18
## age                                                                                -4.35
## income                                                                              2.34
## political_partyAfD                                                                  0.04
## political_partyBündnis Sarah Wagenknecht                                           -0.33
## political_partyCDU/CSU                                                              0.50
## political_partyDie Linke                                                           -0.09
## political_partyEiner anderen Partei                                                -0.30
## political_partyFDP                                                                 -0.27
## political_partyKeine Angabe                                                         0.76
## political_partySPD                                                                  0.88
## education(Noch) kein Abschluss                                                      0.92
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  -0.73
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                0.51
## educationDoktorgrad oder Habilitation                                              -0.57
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss     1.22
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         -0.29
## urban_rural_classperipher                                                          -0.82
## urban_rural_classsehr peripher                                                      0.49
## urban_rural_classzentral                                                            1.21
## federal_stateBaden-Württemberg                                                      2.57
## federal_stateBayern                                                                 1.43
## federal_stateBerlin                                                                 0.61
## federal_stateBrandenburg                                                            1.02
## federal_stateBremen                                                                -0.05
## federal_stateHamburg                                                                0.01
## federal_stateHessen                                                                -0.11
## federal_stateMecklenburg-Vorpommern                                                 0.86
## federal_stateNiedersachsen                                                         -0.13
## federal_stateRheinland-Pfalz                                                       -0.26
## federal_stateSaarland                                                              -1.07
```

```
## federal_stateSachsen-Anhalt                                                       -1.00
## federal_stateSchleswig-Holstein                                                    0.40
## federal_stateThüringen                                                            -1.09
##                                                                                Pr(>|t|)
## (Intercept)                                                                       0.859
## age                                                                             1.6e-05
## income                                                                            0.020
## political_partyAfD                                                                0.967
## political_partyBündnis Sarah Wagenknecht                                          0.744
## political_partyCDU/CSU                                                            0.614
## political_partyDie Linke                                                          0.932
## political_partyEiner anderen Partei                                               0.764
## political_partyFDP                                                                0.785
## political_partyKeine Angabe                                                       0.448
## political_partySPD                                                                0.380
## education(Noch) kein Abschluss                                                    0.358
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) 0.467
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule              0.610
## educationDoktorgrad oder Habilitation                                             0.569
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  0.224
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss        0.775
## urban_rural_classperipher                                                         0.413
## urban_rural_classsehr peripher                                                    0.621
## urban_rural_classzentral                                                          0.226
## federal_stateBaden-Württemberg                                                    0.011
## federal_stateBayern                                                               0.154
## federal_stateBerlin                                                               0.544
## federal_stateBrandenburg                                                          0.309
## federal_stateBremen                                                               0.961
## federal_stateHamburg                                                              0.993
## federal_stateHessen                                                               0.910
## federal_stateMecklenburg-Vorpommern                                               0.391
## federal_stateNiedersachsen                                                        0.899
## federal_stateRheinland-Pfalz                                                      0.792
## federal_stateSaarland                                                             0.284
## federal_stateSachsen-Anhalt                                                       0.318
## federal_stateSchleswig-Holstein                                                   0.688
## federal_stateThüringen                                                            0.276
##
## (Intercept)
## age                                                                            ***
## income                                                                         *
## political_partyAfD
## political_partyBündnis Sarah Wagenknecht
## political_partyCDU/CSU
## political_partyDie Linke
## political_partyEiner anderen Partei
## political_partyFDP
## political_partyKeine Angabe
## political_partySPD
## education(Noch) kein Abschluss
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule
## educationDoktorgrad oder Habilitation
```

```
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
## urban_rural_classperipher
## urban_rural_classsehr peripher
## urban_rural_classzentral
## federal_stateBaden-Württemberg                                                    *
## federal_stateBayern
## federal_stateBerlin
## federal_stateBrandenburg
## federal_stateBremen
## federal_stateHamburg
## federal_stateHessen
## federal_stateMecklenburg-Vorpommern
## federal_stateNiedersachsen
## federal_stateRheinland-Pfalz
## federal_stateSaarland
## federal_stateSachsen-Anhalt
## federal_stateSchleswig-Holstein
## federal_stateThüringen
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33 on 554 degrees of freedom
## Multiple R-squared:  0.0898, Adjusted R-squared:  0.0355
## F-statistic: 1.66 on 33 and 554 DF,  p-value: 0.0133
```

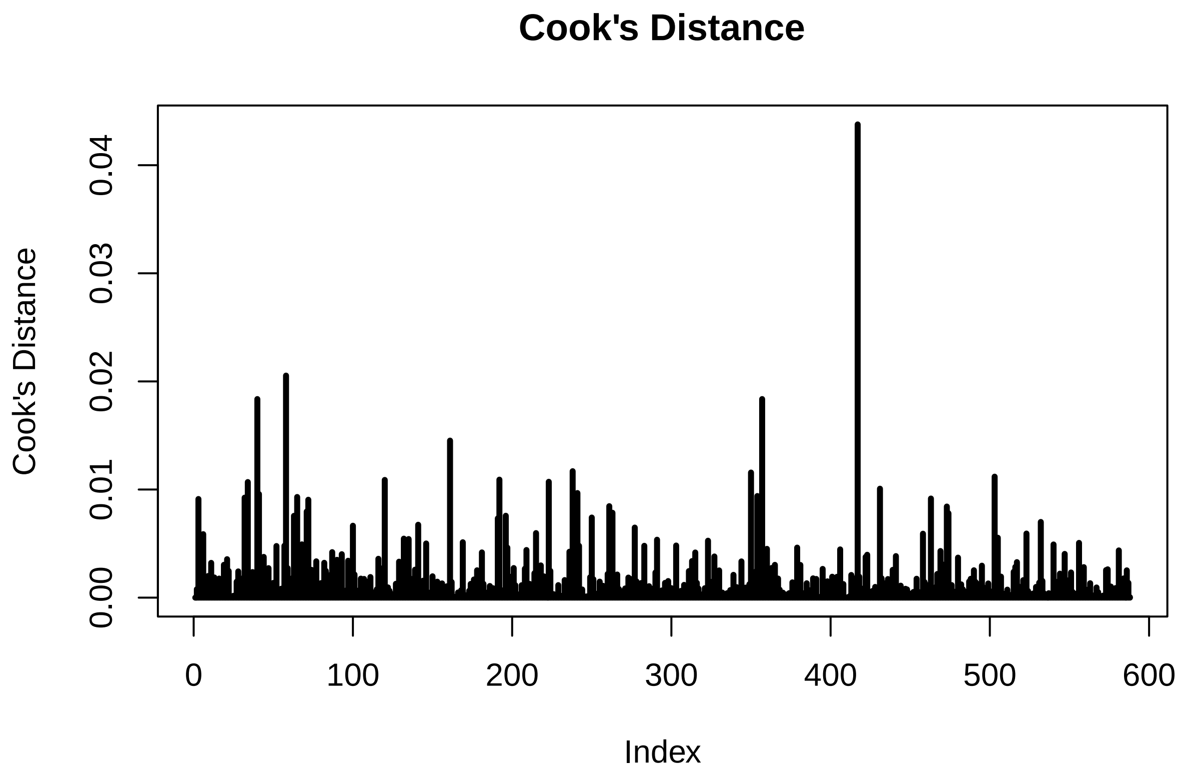
```r
# Checking the VIFs for multicollinearity

vif(model2)
```

```
##                    GVIF Df GVIF^(1/(2*Df))
## age                 1.3  1             1.1
## income              1.1  1             1.0
## political_party     1.8  8             1.0
## education           1.8  6             1.1
## urban_rural_class   2.1  3             1.1
## federal_state       3.0 14             1.0
```

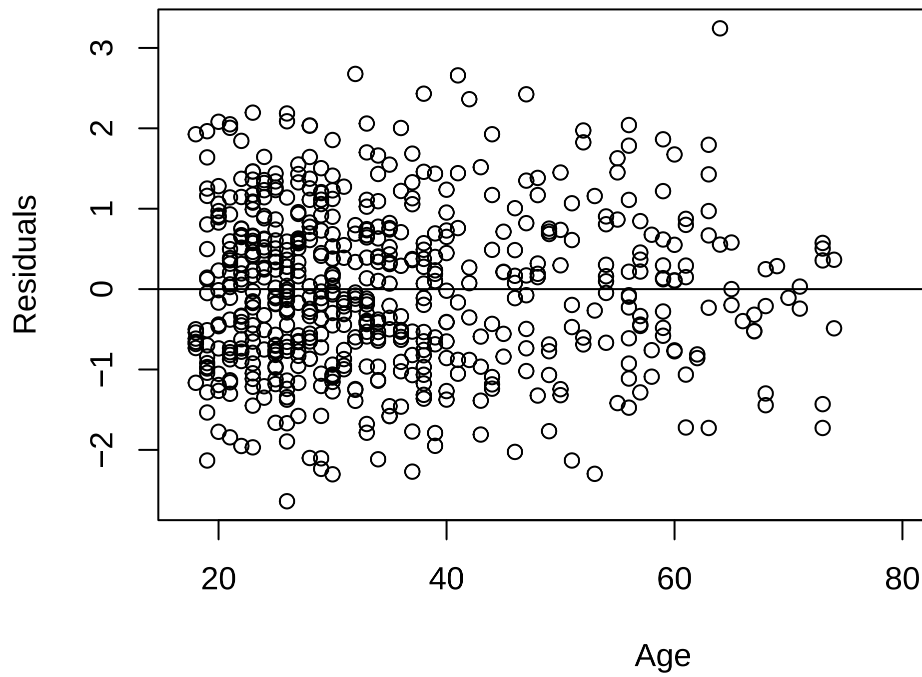
```r
# threshold for multicollinearity
# Calculating the threshold
max(10, 1/(1-summary(model2)$r.square))
```

```
## [1] 10
```

```r
# Checking outliers

cook = cooks.distance(model2)
plot(cook,
     type="h",
     lwd=3,
     ylab = "Cook's Distance",
     main="Cook's Distance")
abline(h = 1)
```

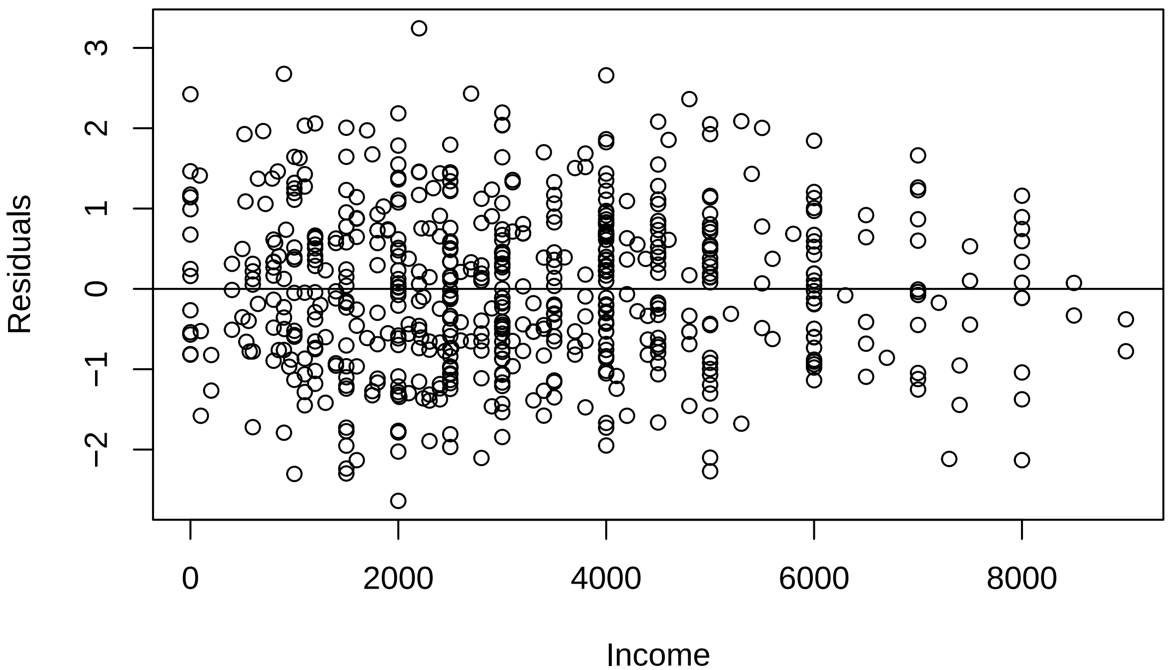



```r
res2 = stdres(model2) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

plot(df2$age, res2, xlab = "Age", ylab = "Residuals")
abline(h = 0)
```

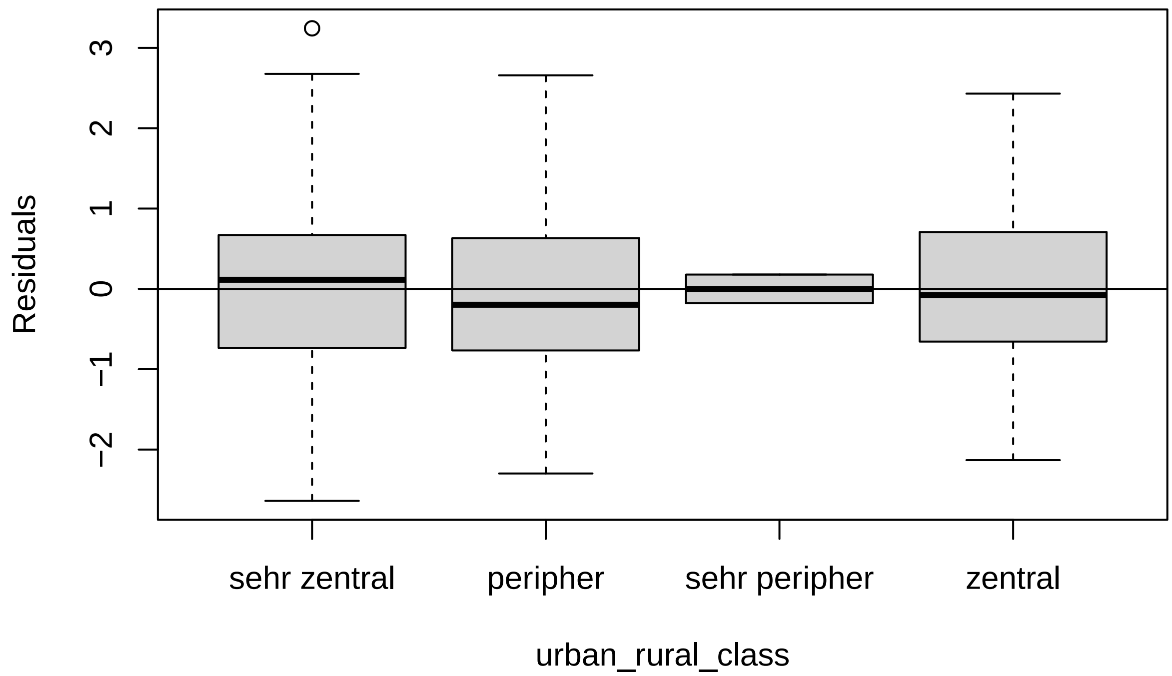


2. Assumptions check in the residuals

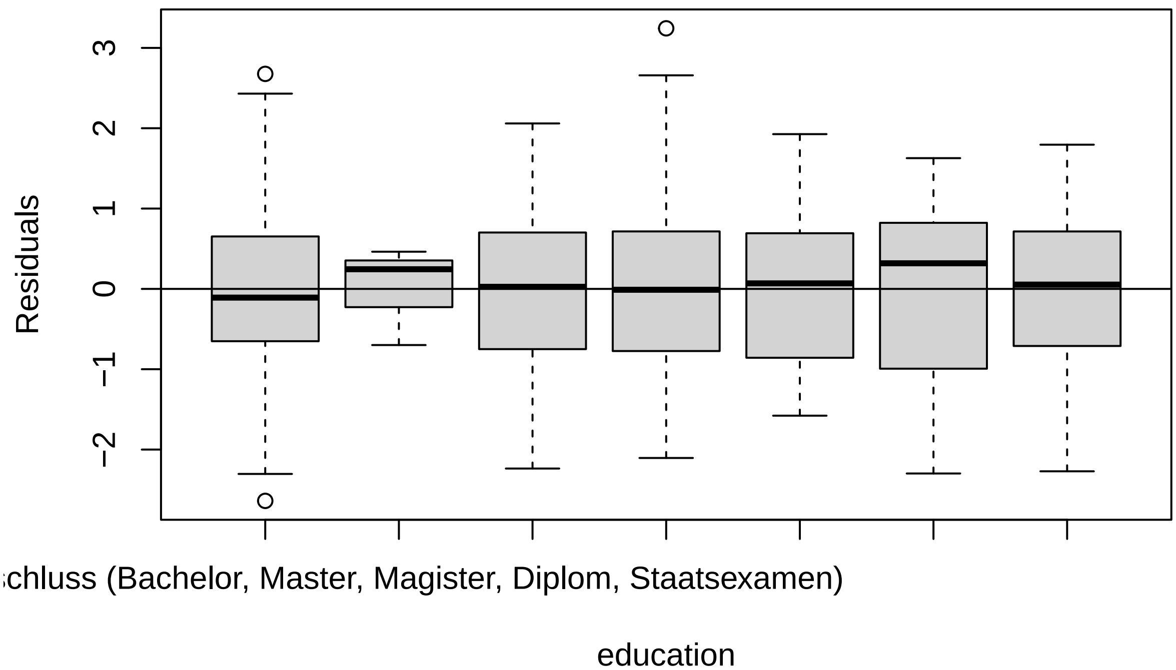
```
plot(df2$income, res2, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



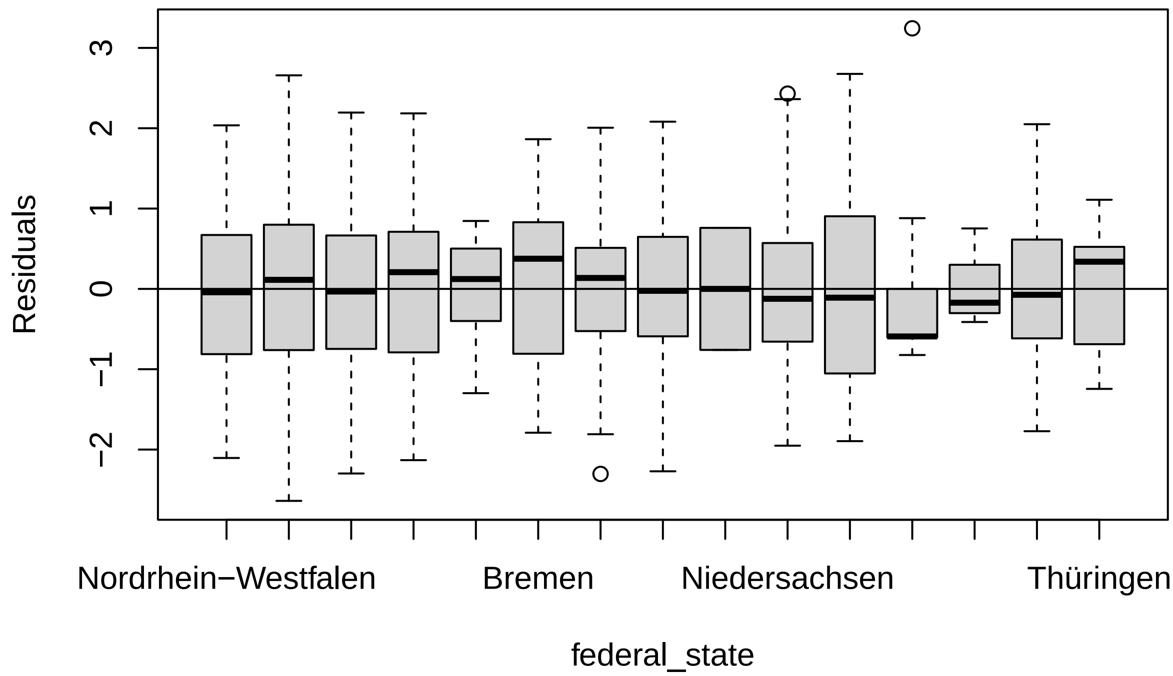

```
plot(df2$urban_rural_class, res2, xlab = "urban_rural_class", ylab = "Residuals")
abline(h = 0)
```



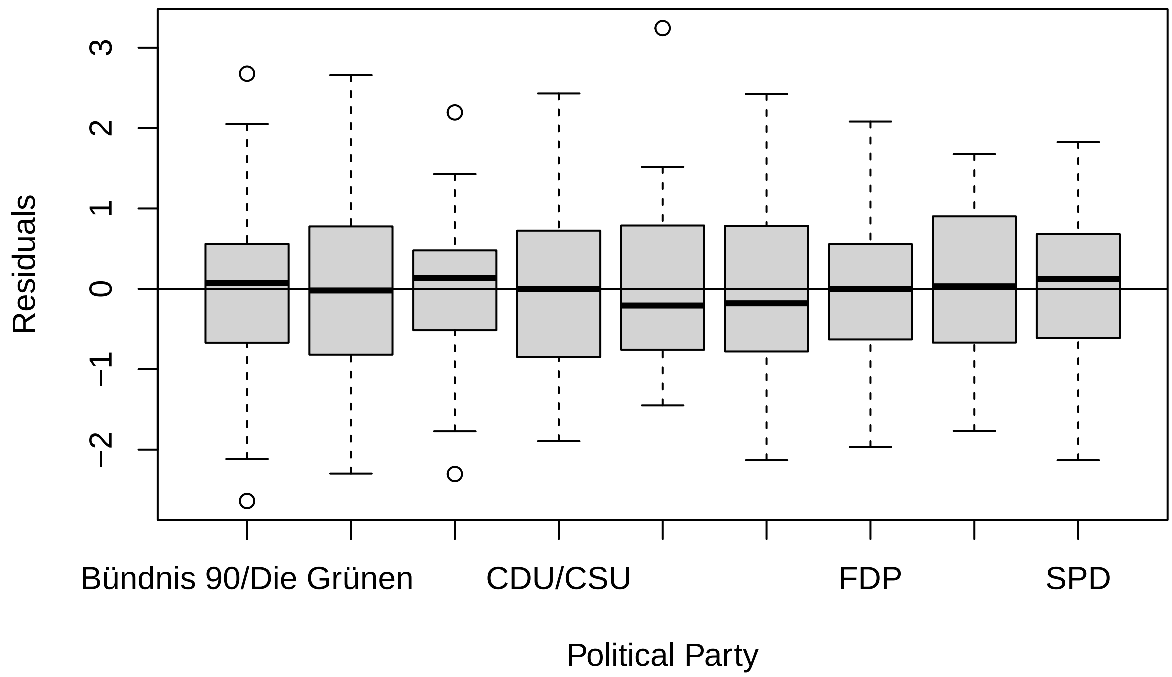

```
plot(df2$education, res2, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



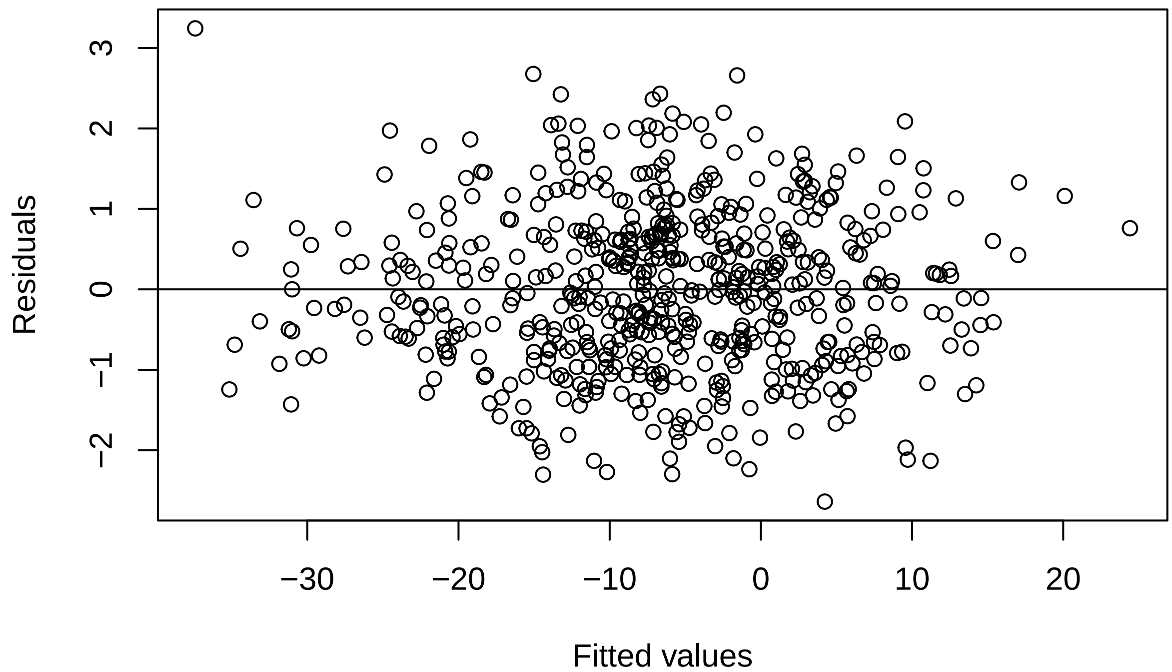

```
plot(df2$federal_state, res2, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```




```
plot(df2$political_party, res2, xlab = "Political Party", ylab = "Residuals")
abline(h = 0)
```




```
# Constant variance and independent error term assumption

plot(fitted(model2), res2, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```



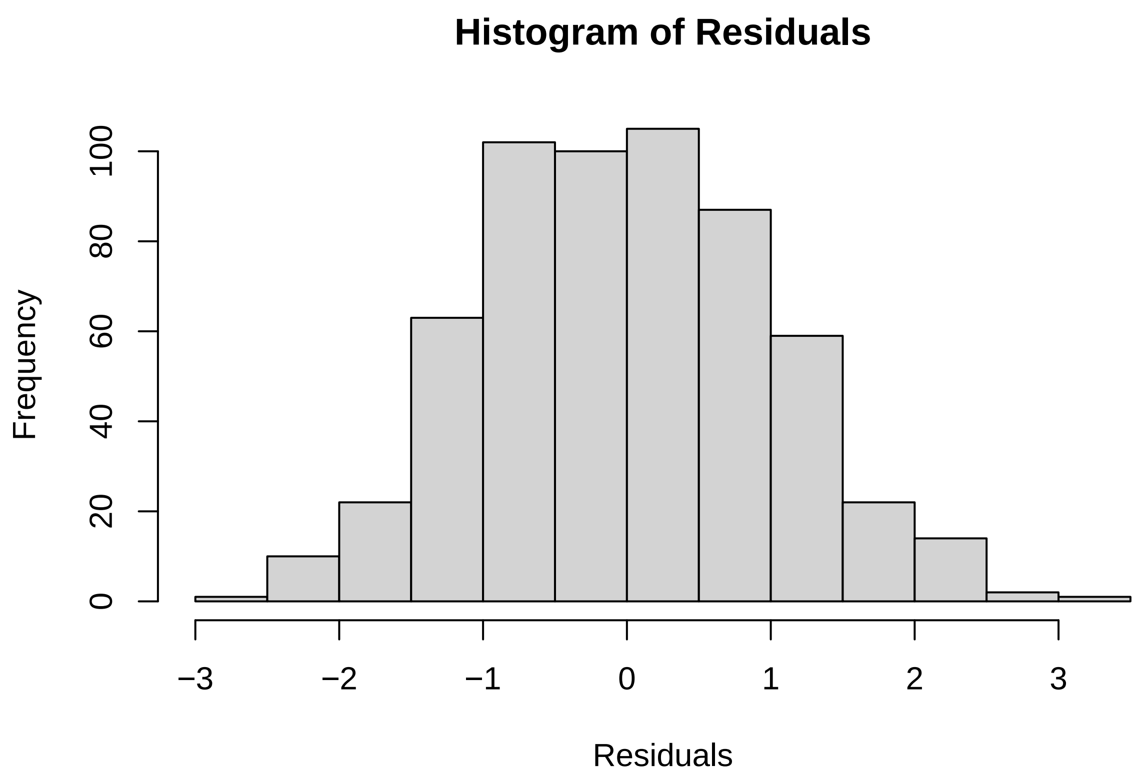

```r
# Durbin-Watson Test: Independence of the error terms
# H0 (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(model2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1          -0.071           2.1   0.096
##  Alternative hypothesis: rho != 0
```

```r
# Breusch-Pagan Test: Heteroscedasticity
# H0: Homoscedasticity is present

bptest(model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model2
## BP = 39, df = 33, p-value = 0.2
```

```r
# Normality assumption

hist(res2, xlab="Residuals", main= "Histogram of Residuals")
```




```r
## normality test using shapiro-test: reject the H0
#H0:  the sample comes from a normal distribution

res2_num = res2[is.finite(res2)]
shapiro.test(res2_num)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res2_num
## W = 1, p-value = 0.2
```

```r
step_model2 <- stepAIC(model2, trace=TRUE, direction= "backward")
```

3. Variable selection

```
## Start:  AIC=4159
## belief_diff_housing_electricity ~ age + income + political_party +
##     education + urban_rural_class + federal_state
##
##                     Df Sum of Sq    RSS  AIC
## - political_party    8      2667 620827 4146
```

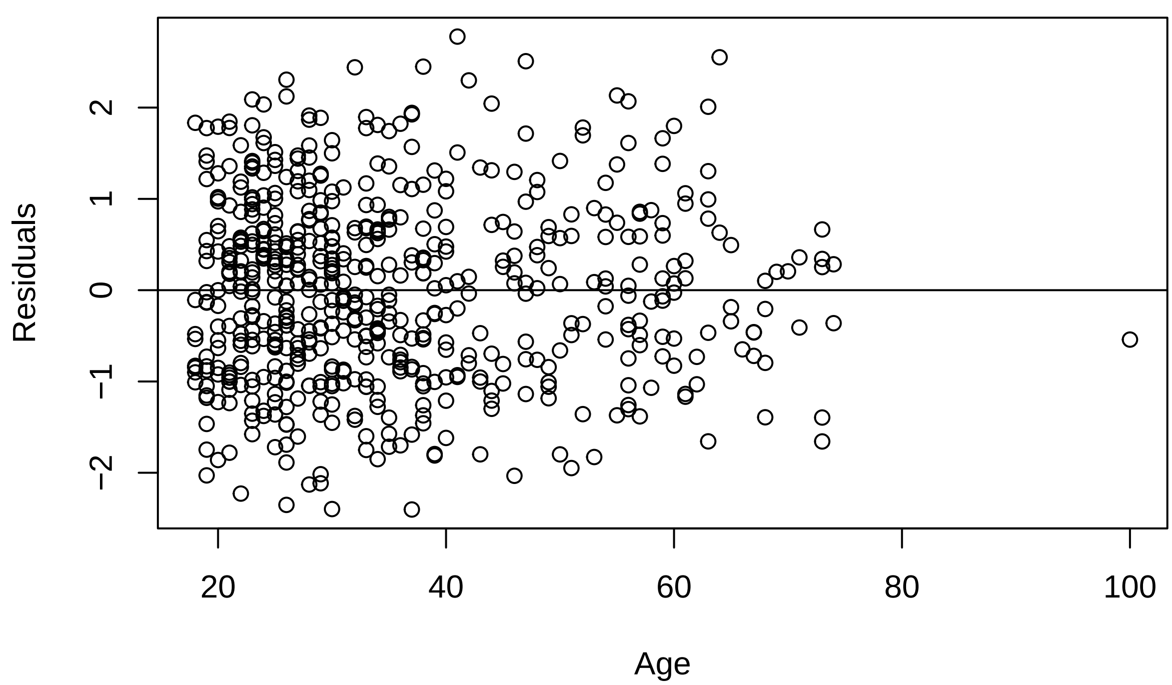
```
## - federal_state     14     20721 638881 4151
## - education          6      4391 622551 4151
## - urban_rural_class  3      3975 622134 4157
## <none>                           618160 4159
## - income             1      6099 624258 4163
## - age                1     21088 639248 4177
##
## Step:  AIC=4146
## belief_diff_housing_electricity ~ age + income + education +
##     urban_rural_class + federal_state
##
##                     Df Sum of Sq    RSS  AIC
## - federal_state     14     20890 641717 4137
## - education          6      4616 625443 4138
## - urban_rural_class  3      4553 625379 4144
## <none>                           620827 4146
## - income             1      6206 627033 4150
## - age                1     19367 640194 4162
##
## Step:  AIC=4137
## belief_diff_housing_electricity ~ age + income + education +
##     urban_rural_class
##
##                     Df Sum of Sq    RSS  AIC
## - education          6      4247 645964 4129
## - urban_rural_class  3      5431 647147 4136
## <none>                           641717 4137
## - income             1      7070 648787 4142
## - age                1     21764 663481 4155
##
## Step:  AIC=4129
## belief_diff_housing_electricity ~ age + income + urban_rural_class
##
##                     Df Sum of Sq    RSS  AIC
## - urban_rural_class  3      4824 650788 4127
## <none>                           645964 4129
## - income             1      7280 653244 4134
## - age                1     20681 666645 4146
##
## Step:  AIC=4127
## belief_diff_housing_electricity ~ age + income
##
##          Df Sum of Sq    RSS  AIC
## <none>                650788 4127
## - income  1      6842 657630 4132
## - age     1     20818 671607 4144
```

```
summary(step_model2)
```

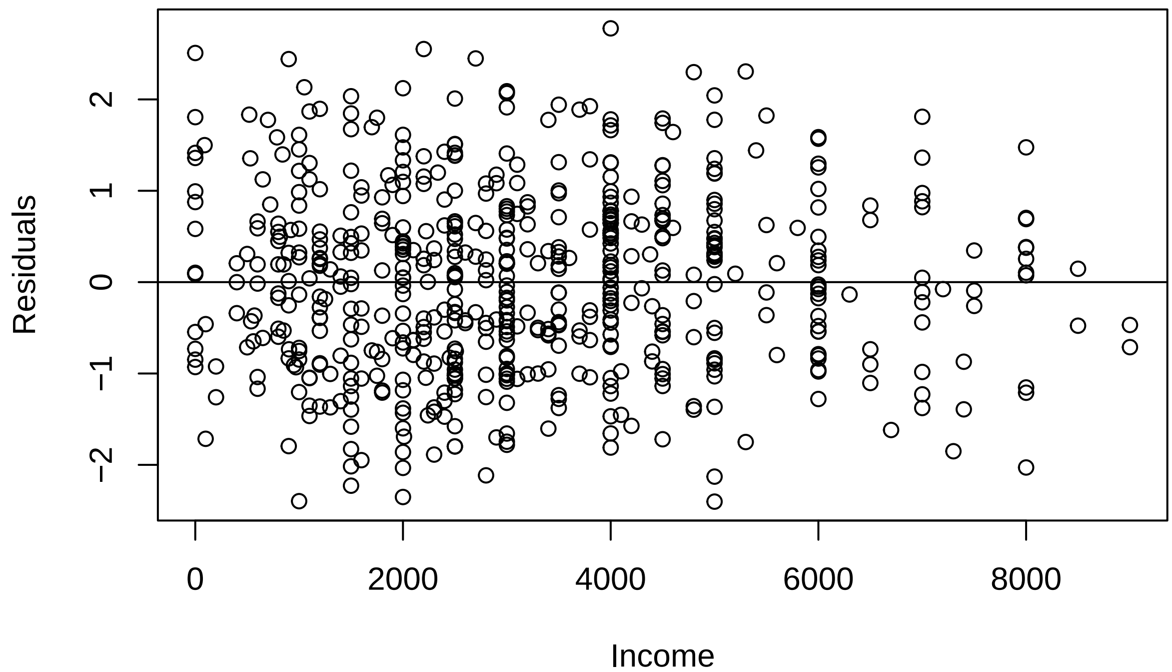
```
##
## Call:
## lm(formula = belief_diff_housing_electricity ~ age + income,
##     data = df2)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -79.99 -25.35   0.52  22.19  92.53
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.915498   4.477743    0.65    0.515
## age         -0.430873   0.099602   -4.33  1.8e-05 ***
## income       0.001804   0.000727    2.48    0.013 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 33 on 585 degrees of freedom
## Multiple R-squared:  0.0417, Adjusted R-squared:  0.0384
## F-statistic: 12.7 on 2 and 585 DF,  p-value: 3.87e-06
```

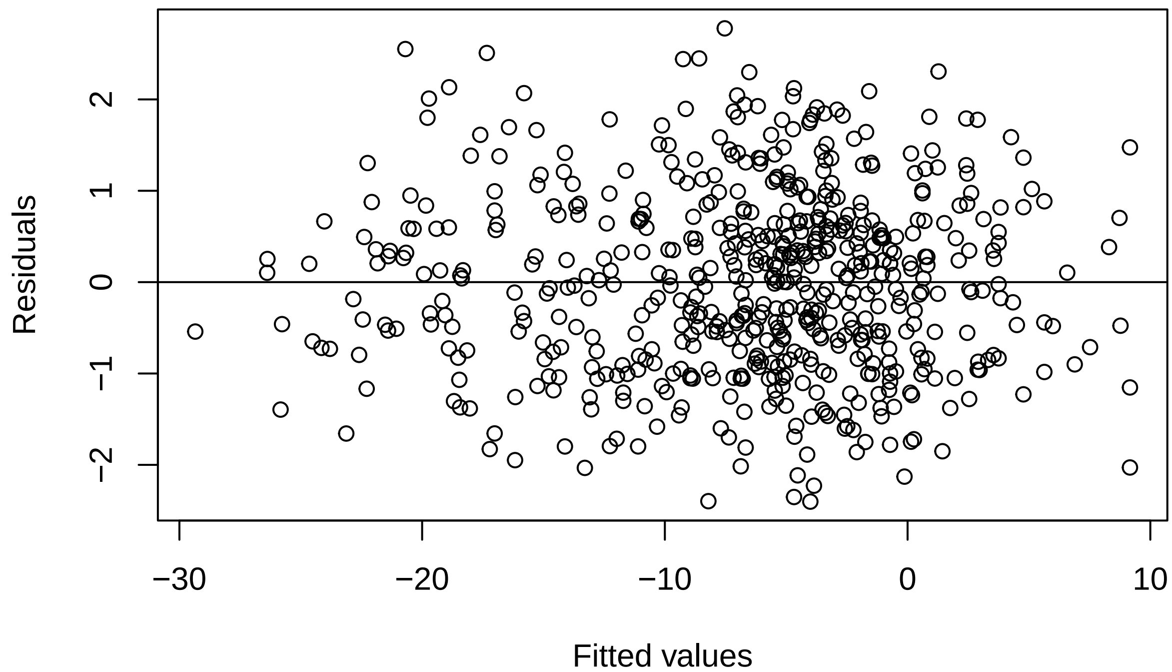
```
res2 = stdres(step_model2) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

plot(df2$age, res2, xlab = "Age", ylab = "Residuals")
abline(h = 0)
```




```
plot(df2$income, res2, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```




```
# Constant variance and independent error term assumption

plot(fitted(step_model2), res2, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```



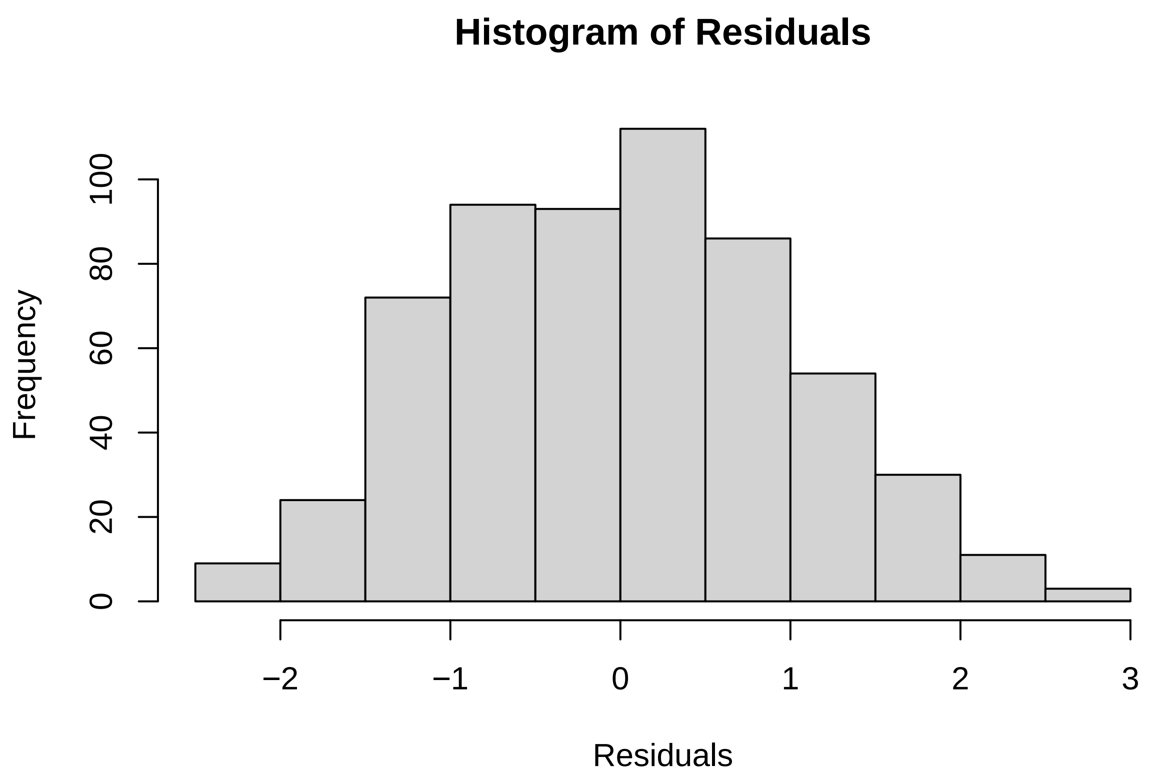

```
# Durbin-Watson Test:  Independence of the error terms
# H0 (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(step_model2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1           -0.04           2.1    0.34
##  Alternative hypothesis: rho != 0
```

```
# Breusch-Pagan TEST: Heteroscedasticity
# H0: Homoscedasticity is present

bptest(step_model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  step_model2
## BP = 2, df = 2, p-value = 0.4
```

```
hist(res2, xlab="Residuals", main= "Histogram of Residuals")
```




```
## normality test using shapiro-test: reject the H0
#H0:  the sample comes from a normal distribution

res2_num = res2[is.finite(res2)]
shapiro.test(res2_num)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res2_num
## W = 1, p-value = 0.01
```