Regression_Analysis_Other_Consumption

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_other_consumption


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, 22) #10, 20, 21, 22, 24

df1
```

```
## # A tibble: 588 x 7
##      age income political_party      education           urban~1 feder~2 CO2_o~3
##    <int>  <dbl> <fct>                <fct>               <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU              (Fach-) Hochschula~ zentral Saarla~   3766.
##  2    59    800 Keine Angabe         Allgemeine oder fa~ sehr z~ Hessen    1445.
##  3    60   1750 Keine Angabe         Berufsausbildung, ~ periph~ Bayern    2433.
##  4    73   2500 SPD                  Realschulabschluss~ sehr z~ Bayern    4152.
##  5    43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin    3766.
##  6    49   2300 CDU/CSU              Berufsausbildung, ~ zentral Sachse~   2318.
##  7    57    600 CDU/CSU              Realschulabschluss~ zentral Baden-~   1521.
##  8    39   5000 SPD                  (Fach-) Hochschula~ sehr z~ Berlin    1217.
##  9    62      0 Keine Angabe         (Fach-) Hochschula~ sehr z~ Nordrh~   1376.
## 10    45   2600 Keine Angabe         Berufsausbildung, ~ sehr z~ Hessen    3399.
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: CO2_other_consumption
```

```
## 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_other_consumption


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, 28) #25, 26, 27, 28, 29

df2
```

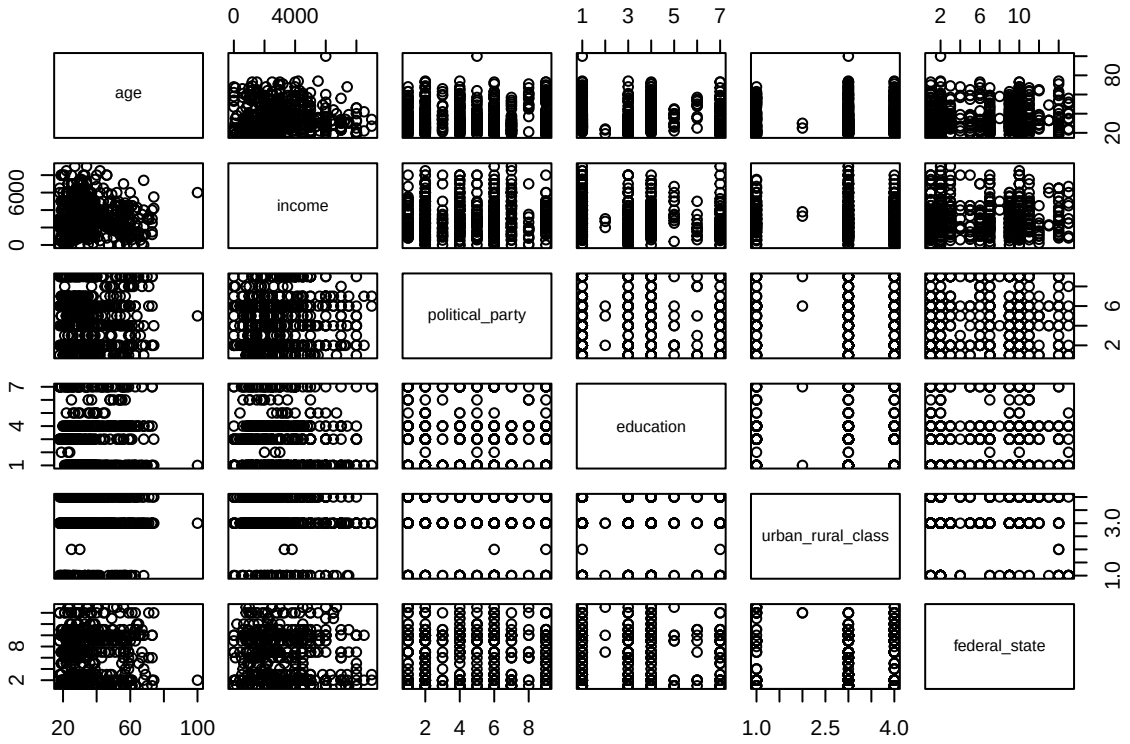
```
## # A tibble: 588 x 7
##      age income political_party      education           urban~1 feder~2 belie~3
##    <int>  <dbl> <fct>                <fct>               <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU              (Fach-) Hochschula~ zentral Saarla~     -68
##  2    59    800 Keine Angabe         Allgemeine oder fa~ sehr z~ Hessen       23
##  3    60   1750 Keine Angabe         Berufsausbildung, ~ periph~ Bayern        9
##  4    73   2500 SPD                  Realschulabschluss~ sehr z~ Bayern      -36
##  5    43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin      -53
##  6    49   2300 CDU/CSU              Berufsausbildung, ~ zentral Sachse~      24
##  7    57    600 CDU/CSU              Realschulabschluss~ zentral Baden-~      37
##  8    39   5000 SPD                  (Fach-) Hochschula~ sehr z~ Berlin       12
##  9    62      0 Keine Angabe         (Fach-) Hochschula~ sehr z~ Nordrh~      18
## 10    45   2600 Keine Angabe         Berufsausbildung, ~ sehr z~ Hessen      -64
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: belief_diff_other_consumption
```

I. Exploratory Data Analysis

Check the Jupytor notebook: EDA_scatter_plot_actual_belief

II. Multivariate Regression: CO2 other consumption

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

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



1. Modeling

```
# finding the most frequent values for the categorical variables

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, D
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_other_consumption ~ age + income + political_party + education +  urban_rural_class + f

summary(model1)
```

```
##
## Call:
## lm(formula = CO2_other_consumption ~ age + income + political_party +
##     education + urban_rural_class + federal_state, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1676.54  -597.49   -17.44   508.54  2780.07
##
## Coefficients:
##                                                                    Estimate
## (Intercept)                                                     1483.18792
## age                                                                9.54800
## income                                                             0.20676
```

```
## political_partyAfD                                                                  -90.96748
## political_partyBündnis Sarah Wagenknecht                                            -43.11592
## political_partyCDU/CSU                                                               23.91325
## political_partyDie Linke                                                           -169.78353
## political_partyEiner anderen Partei                                                 122.30447
## political_partyFDP                                                                  156.86712
## political_partyKeine Angabe                                                        -314.16071
## political_partySPD                                                                 -132.15993
## education(Noch) kein Abschluss                                                      -98.14205
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -279.70479
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule               -205.74263
## educationDoktorgrad oder Habilitation                                              -206.07402
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  -250.70736
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss        -225.79068
## urban_rural_classperipher                                                          -181.51265
## urban_rural_classsehr peripher                                                     -378.47522
## urban_rural_classzentral                                                           -118.72940
## federal_stateBaden-Württemberg                                                     -150.61586
## federal_stateBayern                                                                 178.88642
## federal_stateBerlin                                                                 -45.93766
## federal_stateBrandenburg                                                            199.22633
## federal_stateBremen                                                                   9.65408
## federal_stateHamburg                                                                148.45155
## federal_stateHessen                                                                 -49.98087
## federal_stateMecklenburg-Vorpommern                                                 -40.93265
## federal_stateNiedersachsen                                                          -71.52746
## federal_stateRheinland-Pfalz                                                       -186.05848
## federal_stateSaarland                                                               260.69301
## federal_stateSachsen-Anhalt                                                         436.37233
## federal_stateSchleswig-Holstein                                                     111.62866
## federal_stateThüringen                                                              166.40373
##                                                                                  Std. Error
## (Intercept)                                                                         145.47028
## age                                                                                   2.57539
## income                                                                                0.01721
## political_partyAfD                                                                  123.22048
## political_partyBündnis Sarah Wagenknecht                                            172.70925
## political_partyCDU/CSU                                                              110.67082
## political_partyDie Linke                                                            133.39057
## political_partyEiner anderen Partei                                                  99.00773
## political_partyFDP                                                                  128.17372
## political_partyKeine Angabe                                                         222.01322
## political_partySPD                                                                  112.69819
## education(Noch) kein Abschluss                                                      450.87953
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   87.64318
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                89.40913
## educationDoktorgrad oder Habilitation                                               218.55797
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   249.48170
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         113.70215
## urban_rural_classperipher                                                           114.57961
## urban_rural_classsehr peripher                                                      567.22933
## urban_rural_classzentral                                                             84.00300
## federal_stateBaden-Württemberg                                                      109.12232
## federal_stateBayern                                                                 112.03788
```

```
## federal_stateBerlin                                                                 136.10710
## federal_stateBrandenburg                                                            285.87781
## federal_stateBremen                                                                 208.51885
## federal_stateHamburg                                                                169.69220
## federal_stateHessen                                                                 130.19425
## federal_stateMecklenburg-Vorpommern                                                 546.49727
## federal_stateNiedersachsen                                                          132.45056
## federal_stateRheinland-Pfalz                                                        163.36256
## federal_stateSaarland                                                               256.52074
## federal_stateSachsen-Anhalt                                                         395.42215
## federal_stateSchleswig-Holstein                                                     191.07481
## federal_stateThüringen                                                              289.61790
##                                                                                    t value
## (Intercept)                                                                         10.196
## age                                                                                  3.707
## income                                                                              12.014
## political_partyAfD                                                                  -0.738
## political_partyBündnis Sarah Wagenknecht                                            -0.250
## political_partyCDU/CSU                                                               0.216
## political_partyDie Linke                                                            -1.273
## political_partyEiner anderen Partei                                                  1.235
## political_partyFDP                                                                   1.224
## political_partyKeine Angabe                                                         -1.415
## political_partySPD                                                                  -1.173
## education(Noch) kein Abschluss                                                      -0.218
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   -3.191
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                -2.301
## educationDoktorgrad oder Habilitation                                               -0.943
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss    -1.005
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss          -1.986
## urban_rural_classperipher                                                           -1.584
## urban_rural_classsehr peripher                                                      -0.667
## urban_rural_classzentral                                                            -1.413
## federal_stateBaden-Württemberg                                                      -1.380
## federal_stateBayern                                                                  1.597
## federal_stateBerlin                                                                 -0.338
## federal_stateBrandenburg                                                             0.697
## federal_stateBremen                                                                  0.046
## federal_stateHamburg                                                                 0.875
## federal_stateHessen                                                                 -0.384
## federal_stateMecklenburg-Vorpommern                                                 -0.075
## federal_stateNiedersachsen                                                          -0.540
## federal_stateRheinland-Pfalz                                                        -1.139
## federal_stateSaarland                                                                1.016
## federal_stateSachsen-Anhalt                                                          1.104
## federal_stateSchleswig-Holstein                                                      0.584
## federal_stateThüringen                                                               0.575
##                                                                                   Pr(>|t|)
## (Intercept)                                                                        < 2e-16
## age                                                                                0.00023
## income                                                                             < 2e-16
## political_partyAfD                                                                 0.46068
## political_partyBündnis Sarah Wagenknecht                                           0.80295
## political_partyCDU/CSU                                                             0.82901
```

```
## political_partyDie Linke                                                          0.20361
## political_partyEiner anderen Partei                                               0.21724
## political_partyFDP                                                                0.22152
## political_partyKeine Angabe                                                       0.15761
## political_partySPD                                                                0.24142
## education(Noch) kein Abschluss                                                    0.82777
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) 0.00150
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule              0.02175
## educationDoktorgrad oder Habilitation                                             0.34615
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  0.31538
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss        0.04755
## urban_rural_classperipher                                                         0.11373
## urban_rural_classsehr peripher                                                    0.50490
## urban_rural_classzentral                                                          0.15810
## federal_stateBaden-Württemberg                                                    0.16807
## federal_stateBayern                                                               0.11091
## federal_stateBerlin                                                               0.73586
## federal_stateBrandenburg                                                          0.48616
## federal_stateBremen                                                               0.96309
## federal_stateHamburg                                                              0.38205
## federal_stateHessen                                                               0.70120
## federal_stateMecklenburg-Vorpommern                                               0.94032
## federal_stateNiedersachsen                                                        0.58939
## federal_stateRheinland-Pfalz                                                      0.25522
## federal_stateSaarland                                                             0.30995
## federal_stateSachsen-Anhalt                                                       0.27026
## federal_stateSchleswig-Holstein                                                   0.55931
## federal_stateThüringen                                                            0.56582
##
## (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: 752.8 on 554 degrees of freedom
## Multiple R-squared:  0.3142, Adjusted R-squared:  0.2733
## F-statistic:  7.69 on 33 and 554 DF,  p-value: < 2.2e-16
```

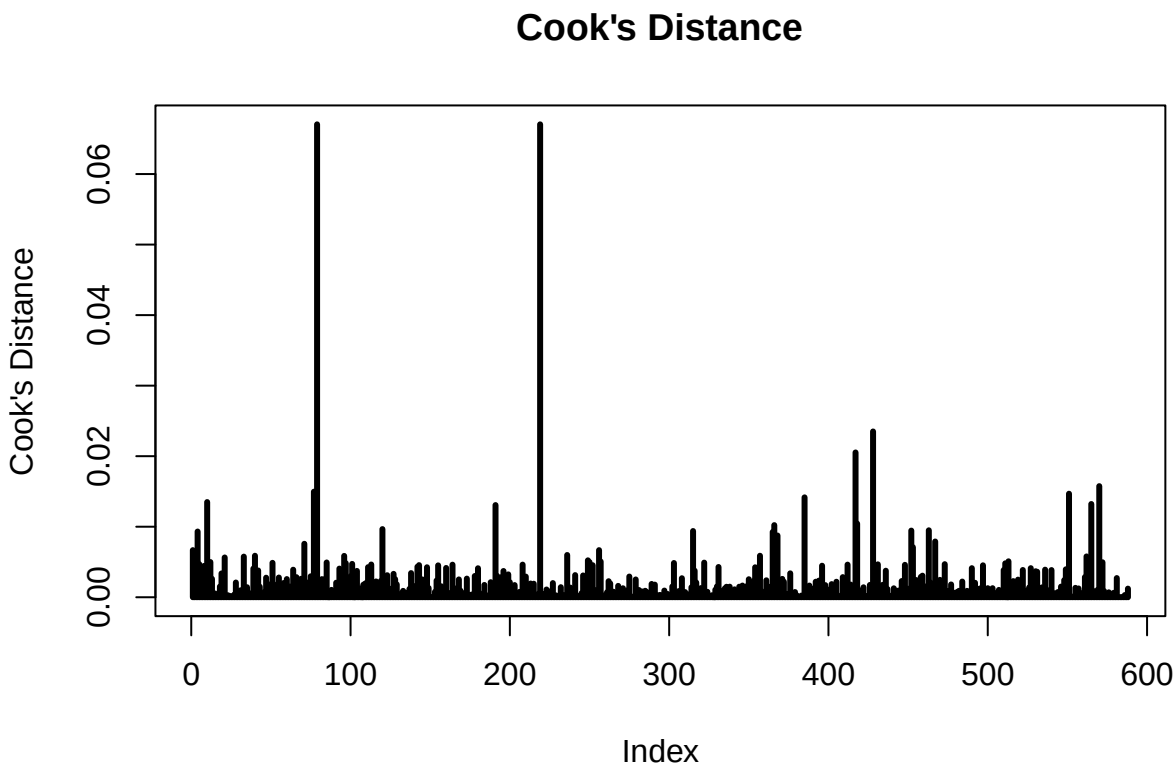
```
# Checking the VIFs for multicollinearity

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
```

```
# Calculating the threshold for multicollinearity

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

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

```
# 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
```

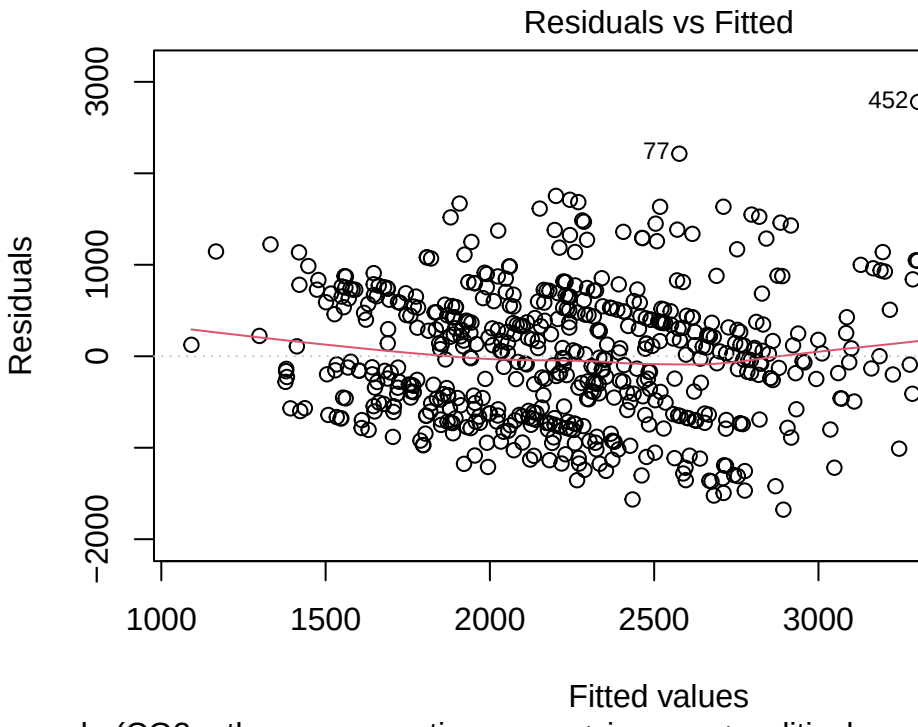
```
##           1           4          10          21          33          40
## 0.006678063 0.009318080 0.013496170 0.005641452 0.005759147 0.005884612
##          71          77          79          96         120         191
## 0.007578456 0.014965915 0.067047005 0.005864675 0.009661257 0.013069687
##         219         236         256         315         357         365
## 0.067047005 0.005991985 0.006680510 0.009390571 0.005884612 0.009260705
##         366         368         385         417         418         428
## 0.010226910 0.008756813 0.014160513 0.020527714 0.010442998 0.023502630
##         452         453         463         467         551         562
## 0.009450234 0.007095319 0.009494270 0.007904132 0.014680721 0.005788038
##         565         570
## 0.013219117 0.015751975
```

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


df1[influential_vector, ]
```

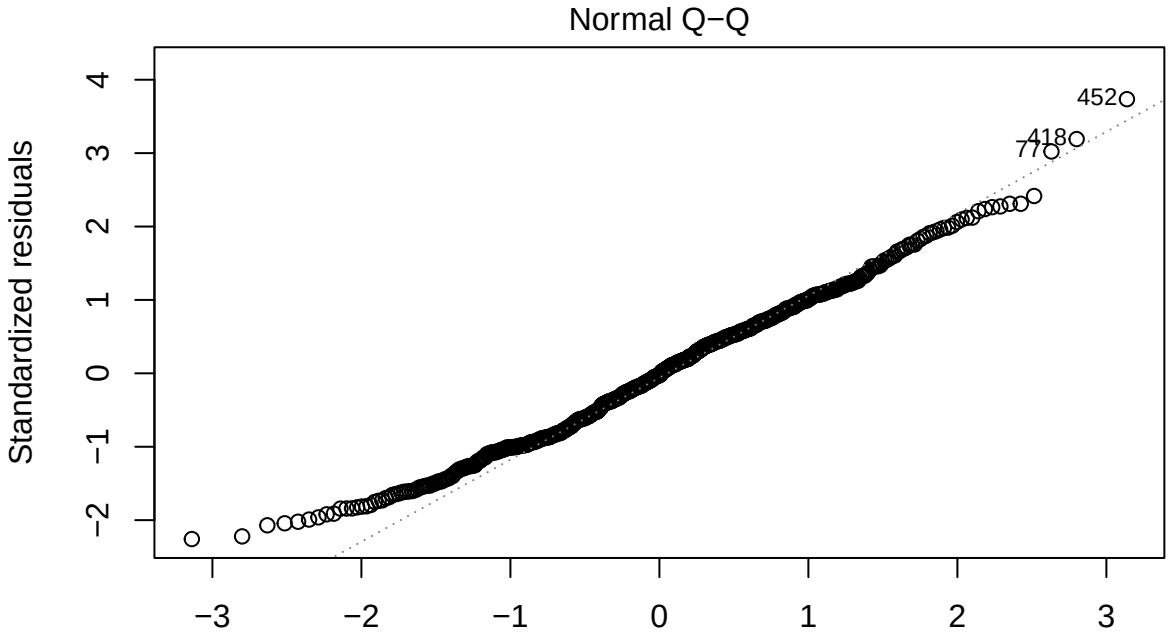
```
## # A tibble: 32 x 7
##      age income political_party education               urban~1 feder~2 CO2_o~3
##    <int>  <dbl> <fct>           <fct>                     <fct>   <fct>     <dbl>
```

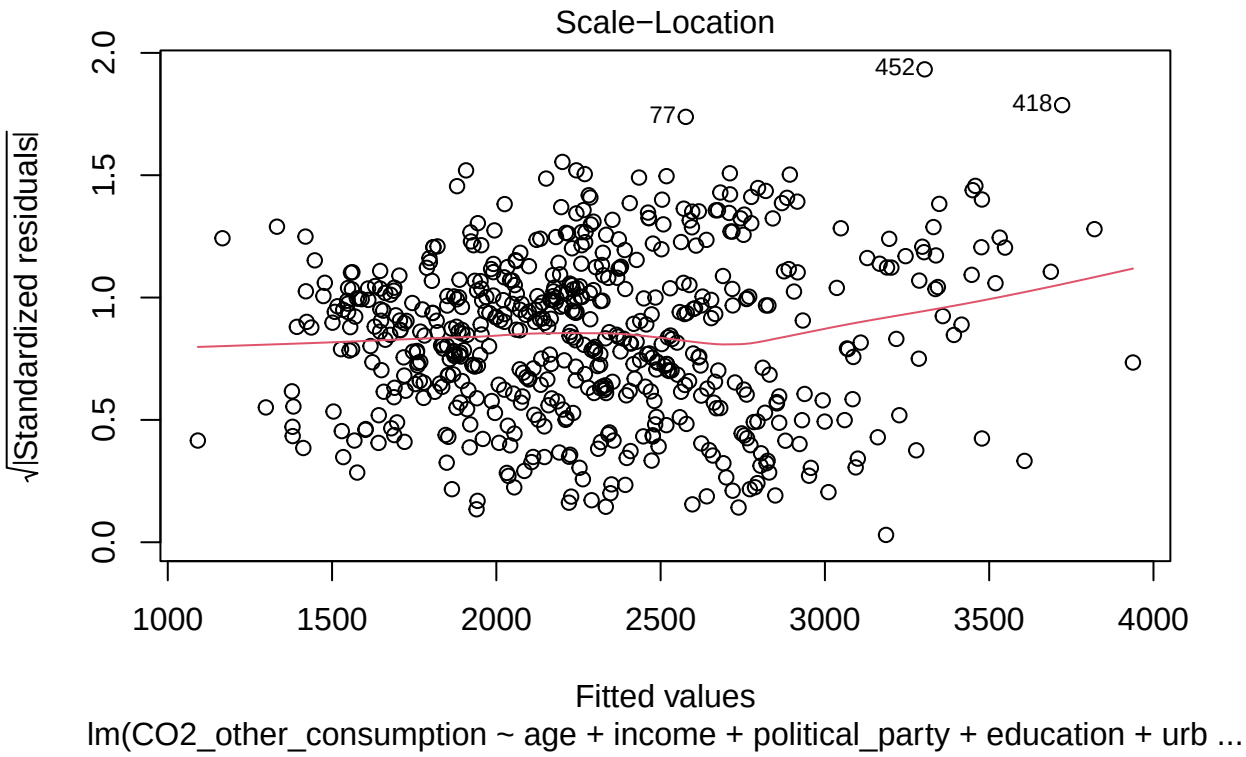
```
##  1    65   3000 CDU/CSU         (Fach-) Hochschulabschl~ zentral Saarla~   3766.
##  2    73   2500 SPD             Realschulabschluss (Mit~ sehr z~ Bayern    4152.
##  3    45   2600 Keine Angabe    Berufsausbildung, Lehre~ sehr z~ Hessen    3399.
##  4    54   2900 AfD             Hauptschulabschluss (Vo~ zentral Rheinl~   2752.
##  5    37   3500 Keine Angabe    Hauptschulabschluss (Vo~ sehr z~ Bayern    1445.
##  6    58   4000 CDU/CSU         (Fach-) Hochschulabschl~ periph~ Meckle~   2890.
##  7    56   1000 Keine Angabe    Berufsausbildung, Lehre~ periph~ Thürin~   2427.
##  8    27   5400 CDU/CSU         (Fach-) Hochschulabschl~ zentral Rheinl~   4791.
##  9    30   3800 SPD             Realschulabschluss (Mit~ sehr p~ Schles~   1156.
## 10    34   4500 SPD             (Fach-) Hochschulabschl~ sehr z~ Hamburg   1445.
## # ... with 22 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: CO2_other_consumption
```

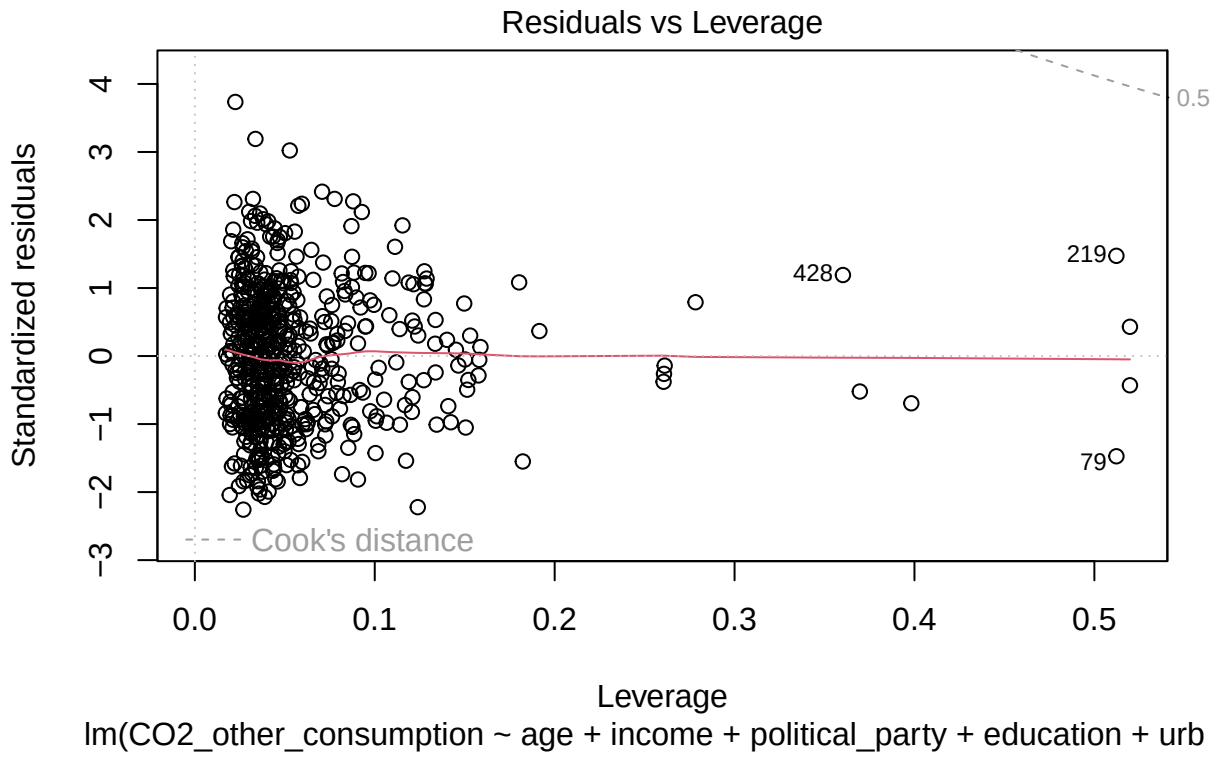
```
plot(model1)
```



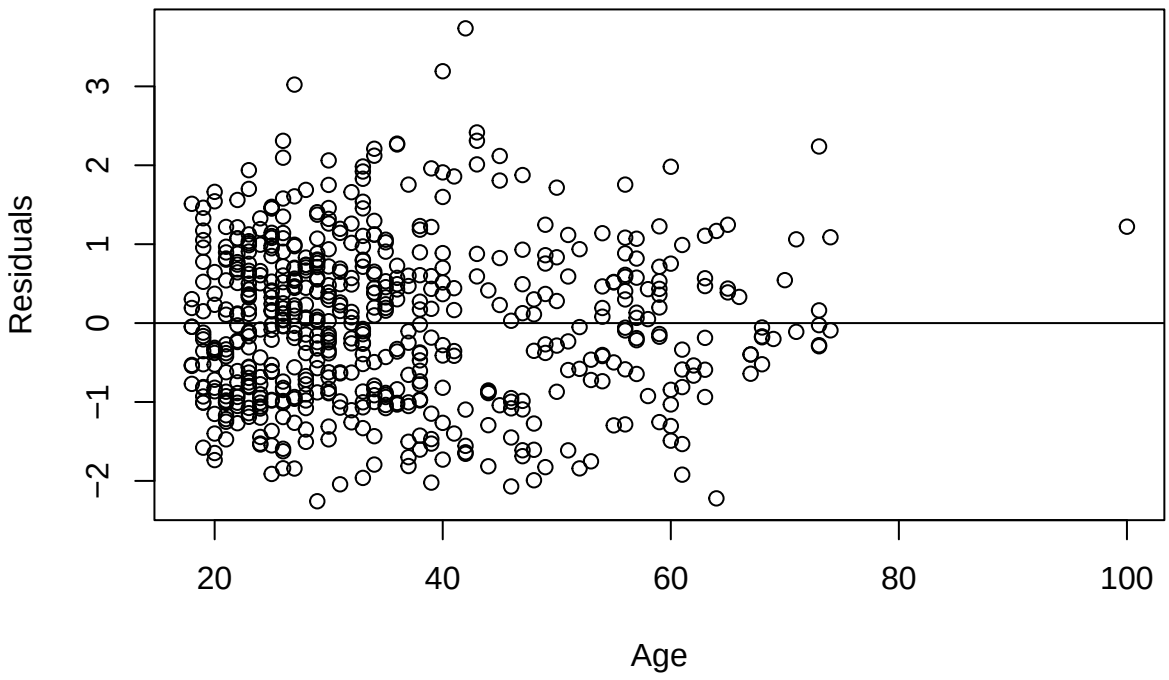
2. Assumptions check in the residuals



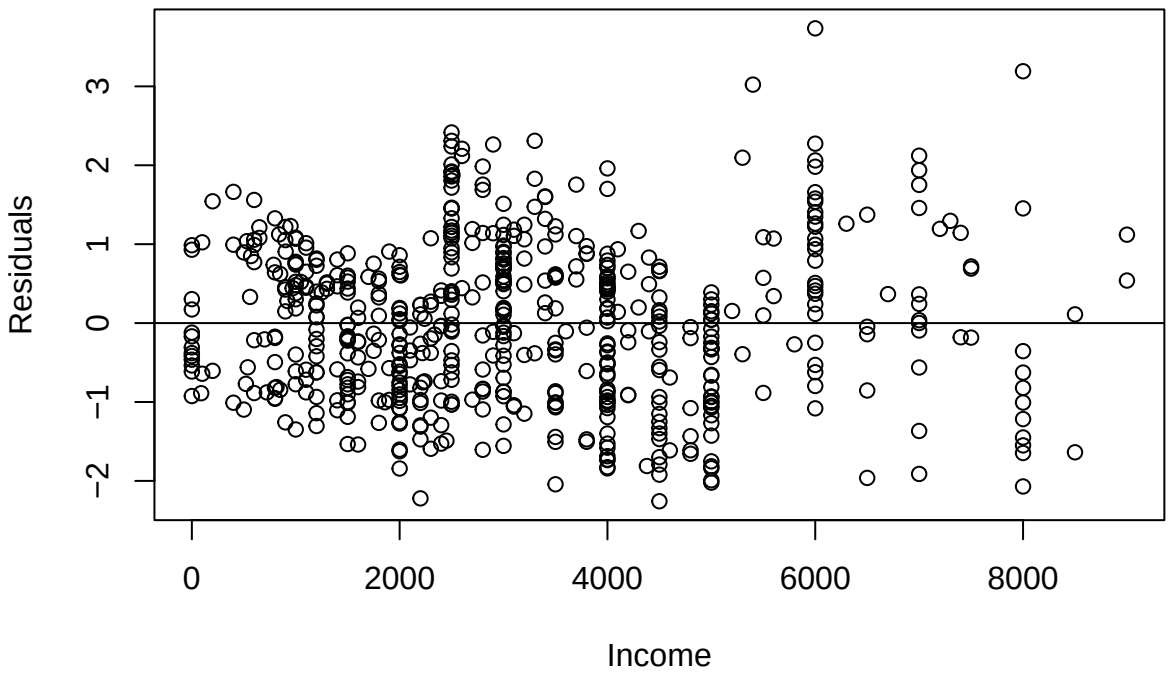

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
452
418
77
1000
1500
2000
2500
3000
3500
4000
Fitted values
lm(CO2_other_consumption ~ age + income + political_party + education + urb ...



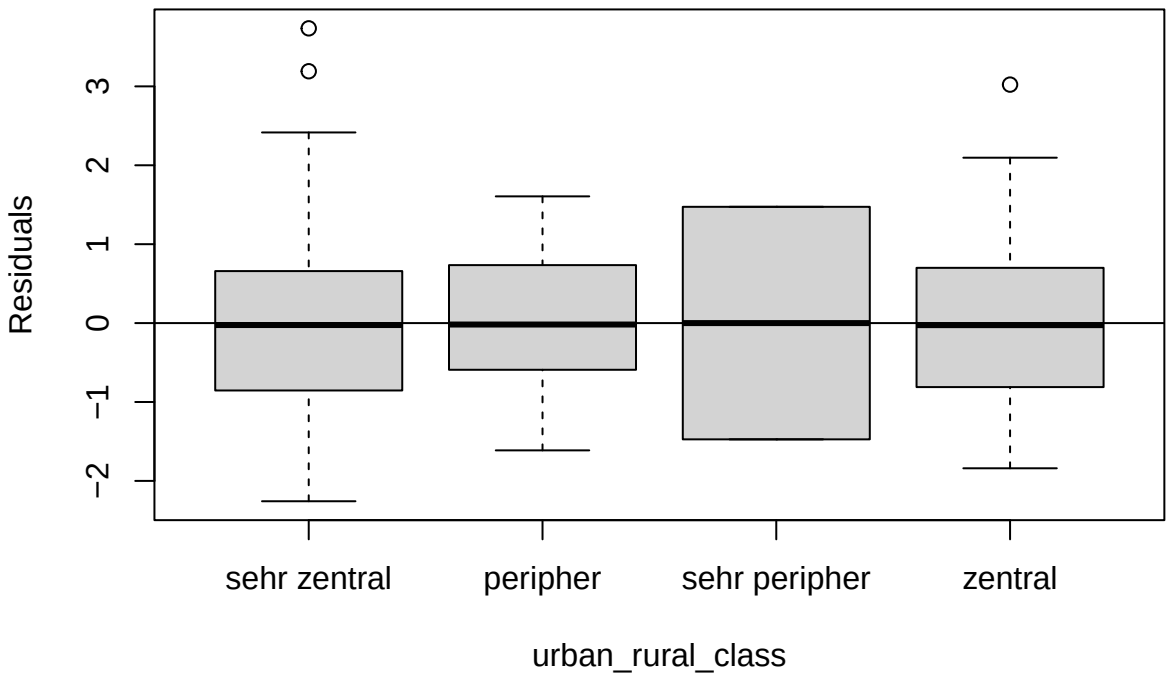

```
res1 = stdres(model1) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

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



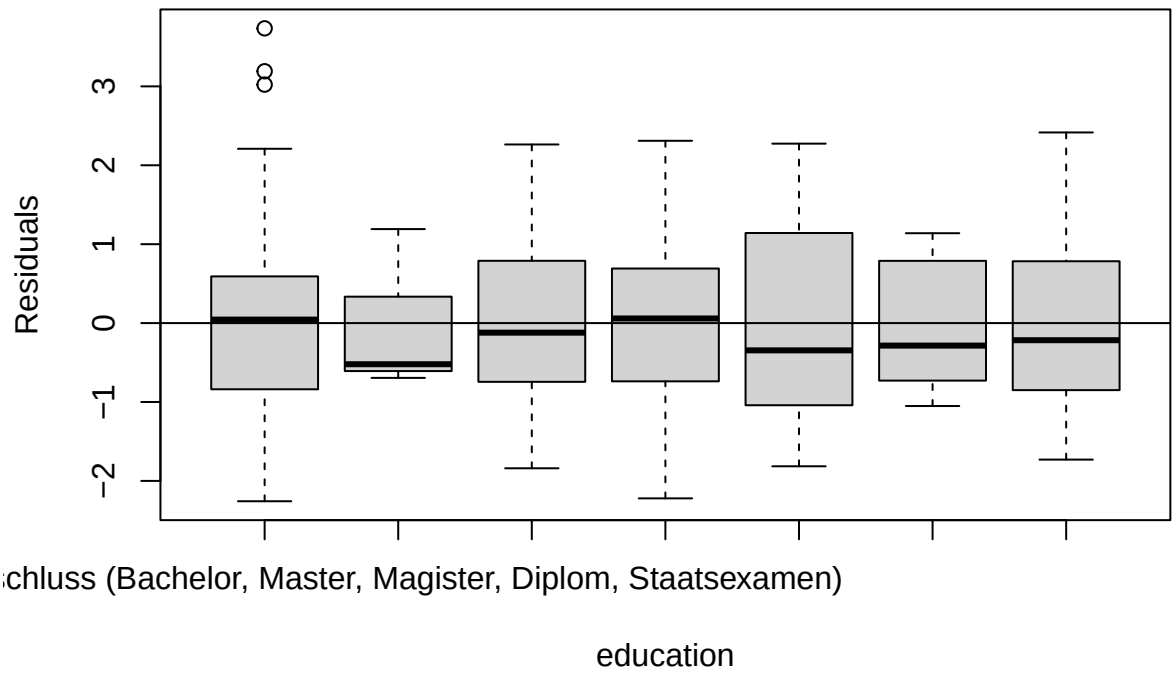

```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



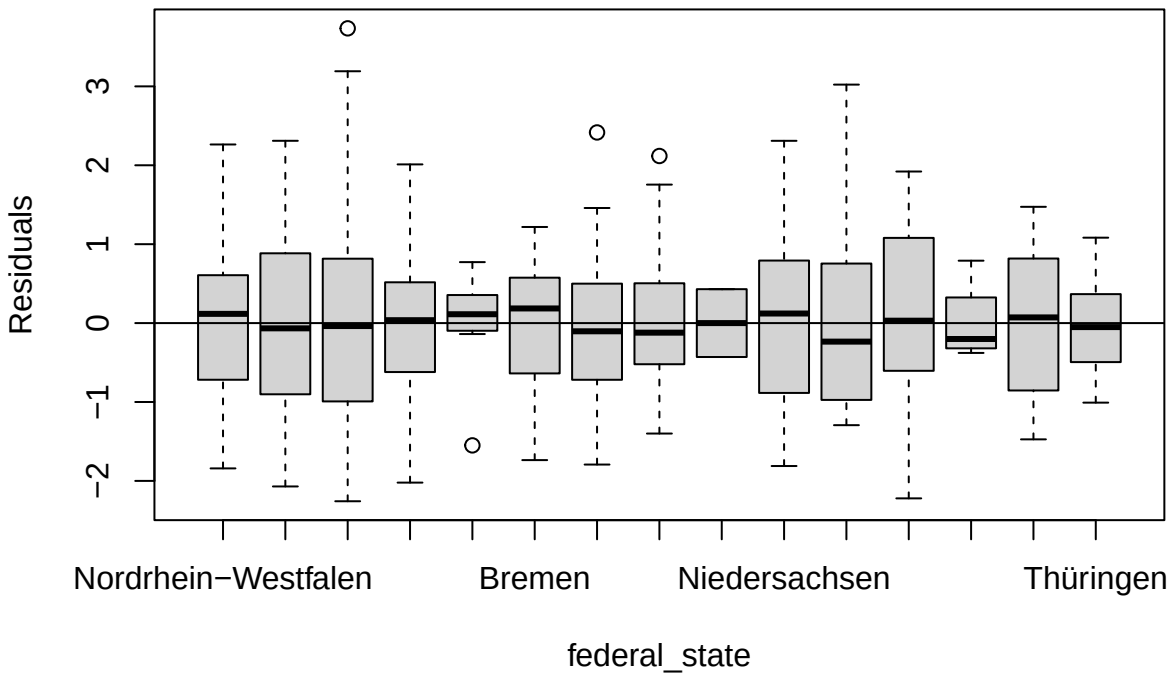

```
plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
abline(h = 0)
```



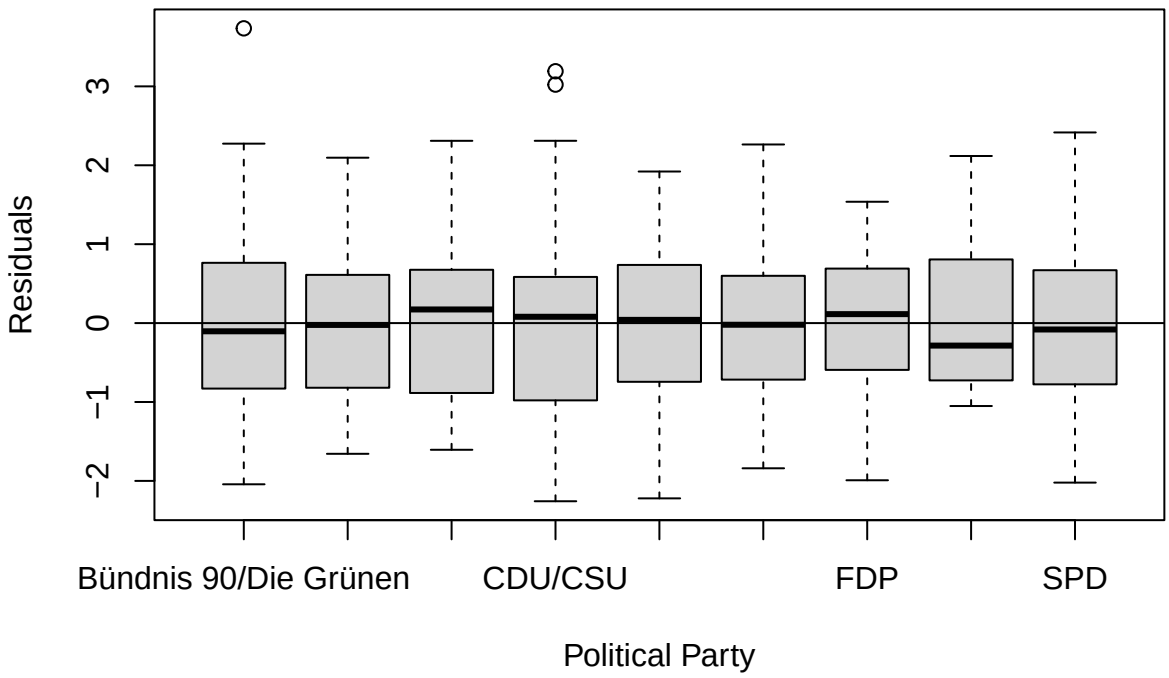

```
plot(df1$education, res1, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



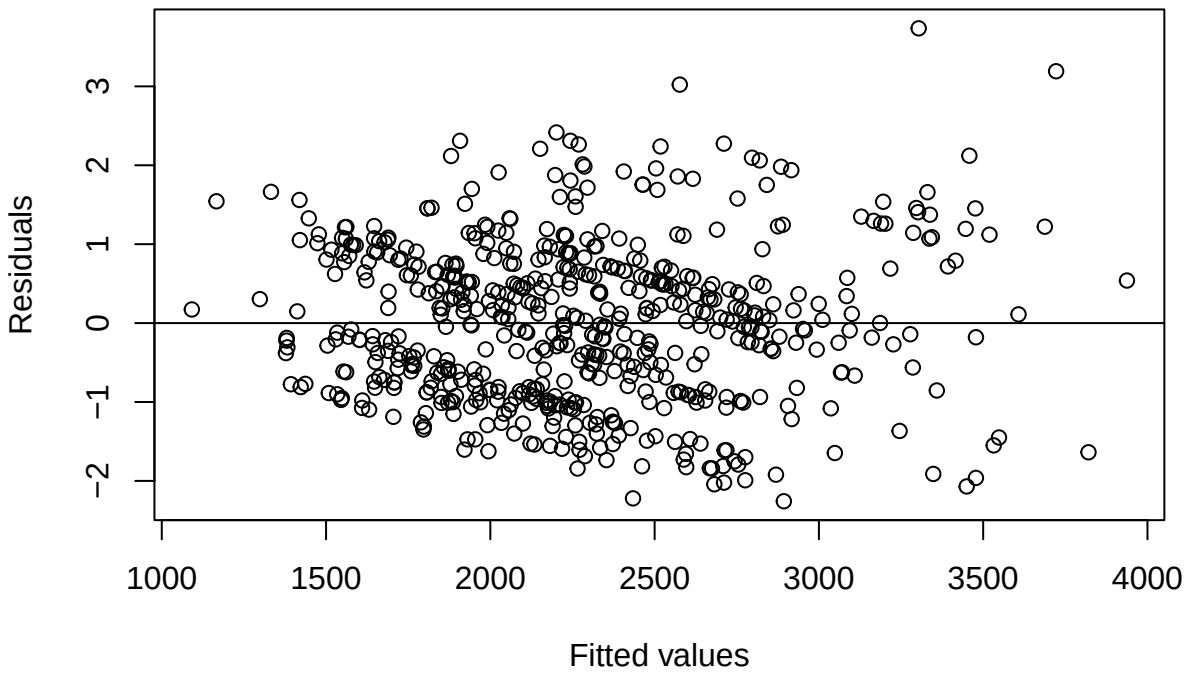

```
plot(df1$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```




```
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)
```



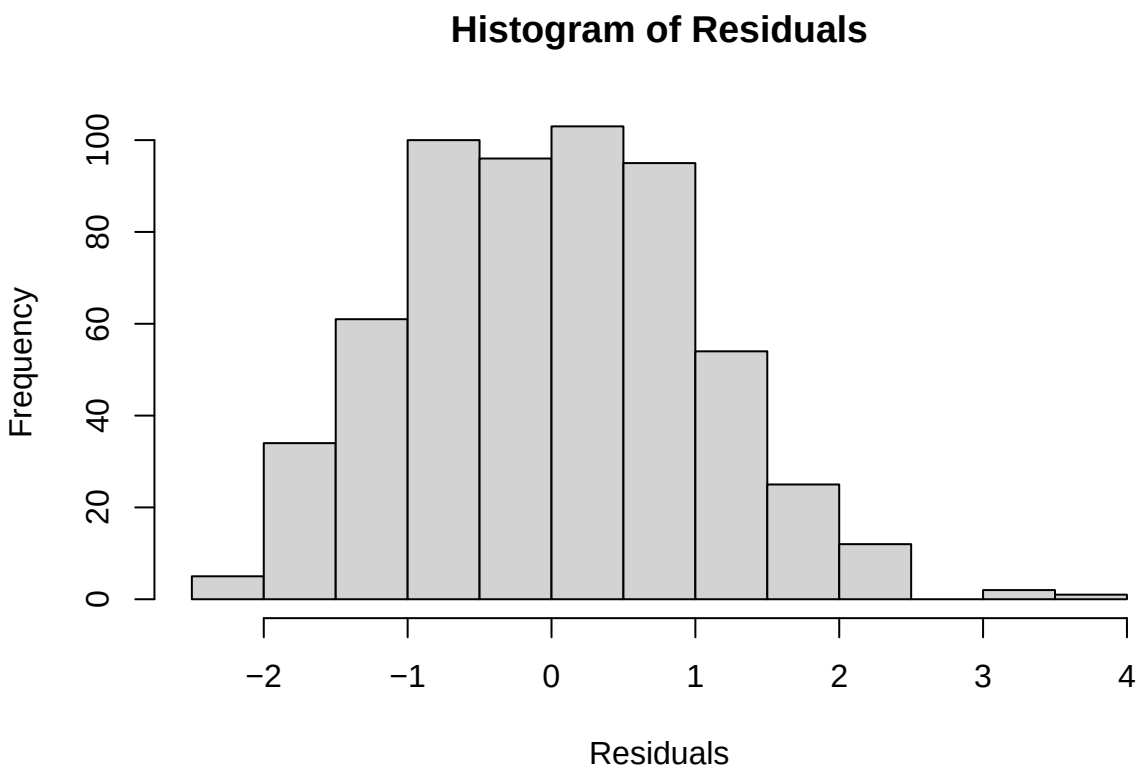

```
# Durbin-Watson Test:  Independence of the error terms
# HO (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(model1)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.02147313      1.952612   0.556
##  Alternative hypothesis: rho != 0
```

```
# Breusch-Pagan Test: Heteroscedasticity
# HO: Homoscedasticity is present

library(lmtest)
bptest(model1)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1
## BP = 86.394, df = 33, p-value = 1.146e-06
```

```
# 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 = 0.99105, p-value = 0.001251
```

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

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

3. Variable Selection, model outcome and assumption check

```
## Start:  AIC=7822.58
## CO2_other_consumption ~ age + income + political_party + education +
##     urban_rural_class + federal_state
##
```

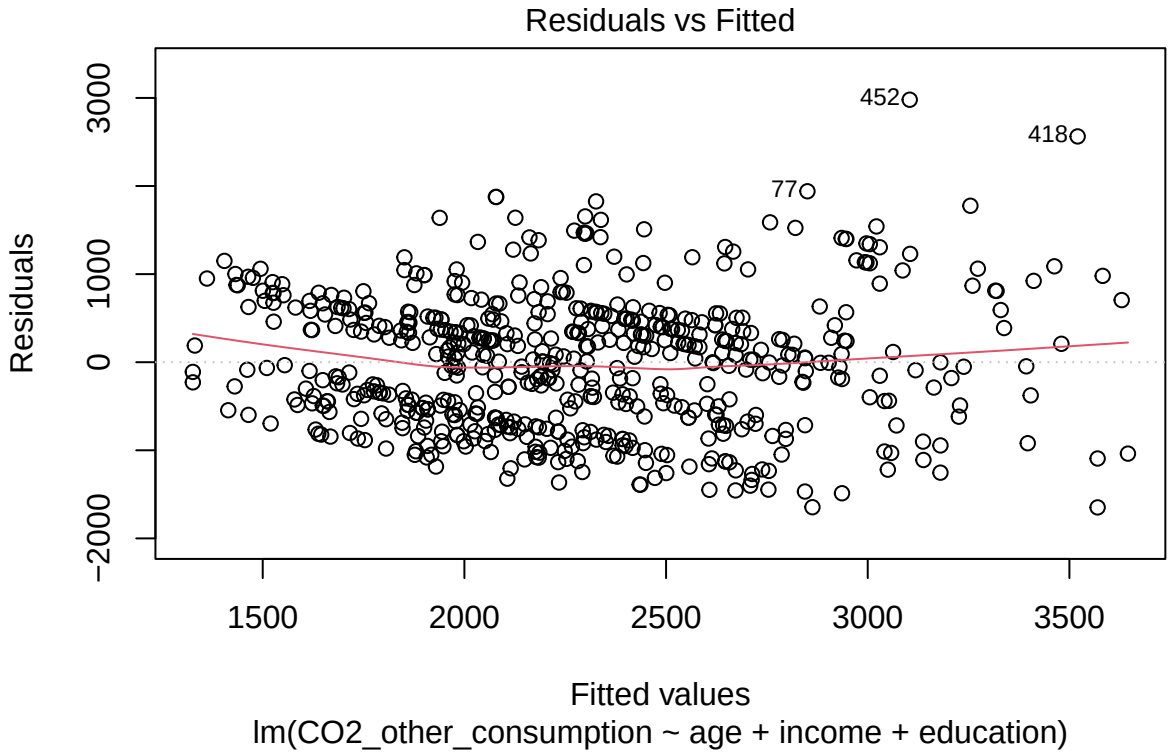
```
##                     Df Sum of Sq       RSS    AIC
## - federal_state     14   9112753 323076551 7811.4
## - political_party    8   6540053 320503852 7818.7
## - urban_rural_class  3   2002238 315966037 7820.3
## <none>                           313963798 7822.6
## - education          6   7632753 321596552 7824.7
## - age                1   7789496 321753294 7835.0
## - income             1  81794002 395757801 7956.7
##
## Step:  AIC=7811.4
## CO2_other_consumption ~ age + income + political_party + education +
##     urban_rural_class
##
##                     Df Sum of Sq       RSS    AIC
## - political_party    8   5712587 328789139 7805.7
## - urban_rural_class  3   1923297 324999848 7808.9
## <none>                           323076551 7811.4
## - education          6   9010132 332086683 7815.6
## - age                1   8332876 331409428 7824.4
## - income             1  88452397 411528949 7951.7
##
## Step:  AIC=7805.71
## CO2_other_consumption ~ age + income + education + urban_rural_class
##
##                     Df Sum of Sq       RSS    AIC
## - urban_rural_class  3   1858654 330647793 7803.0
## <none>                           328789139 7805.7
## - education          6  10307799 339096938 7811.9
## - age                1   6621009 335410148 7815.4
## - income             1  94038551 422827690 7951.6
##
## Step:  AIC=7803.02
## CO2_other_consumption ~ age + income + education
##
##             Df Sum of Sq       RSS    AIC
## <none>                   330647793 7803.0
## - education  6  11468398 342116191 7811.1
## - age        1   6393972 337041765 7812.3
## - income     1  95229567 425877360 7949.8
```

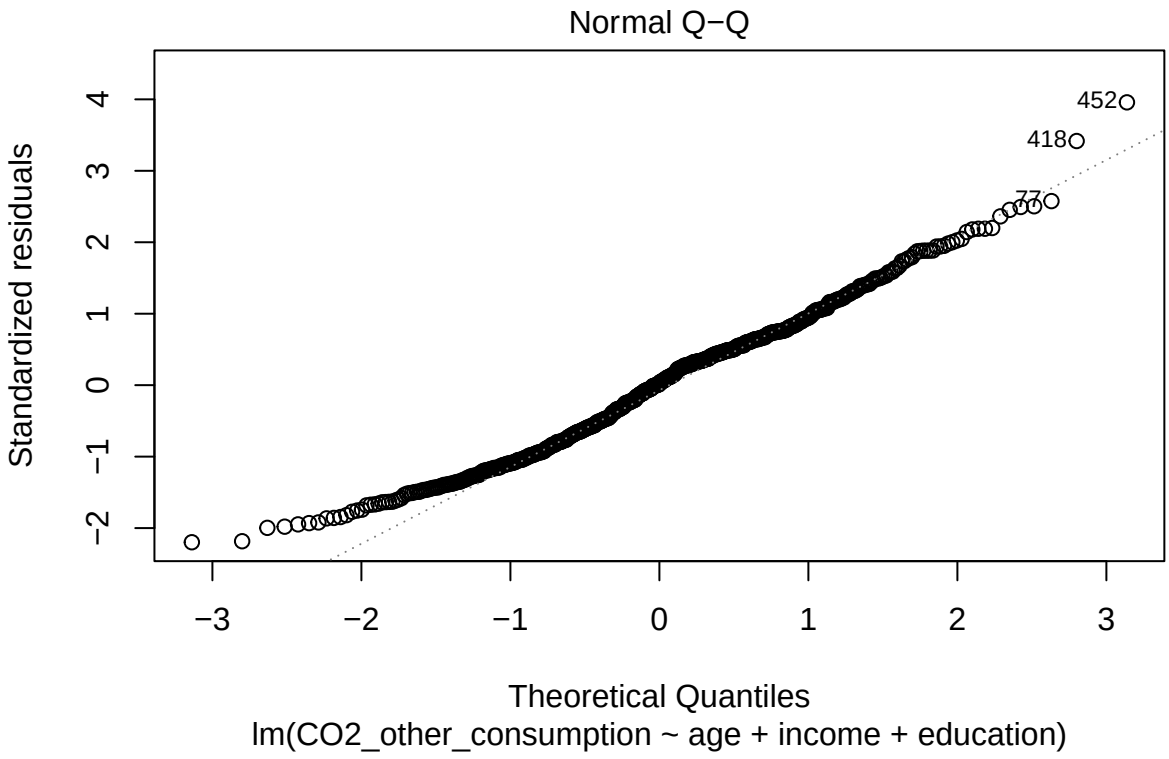
```
summary(step_model1)
```

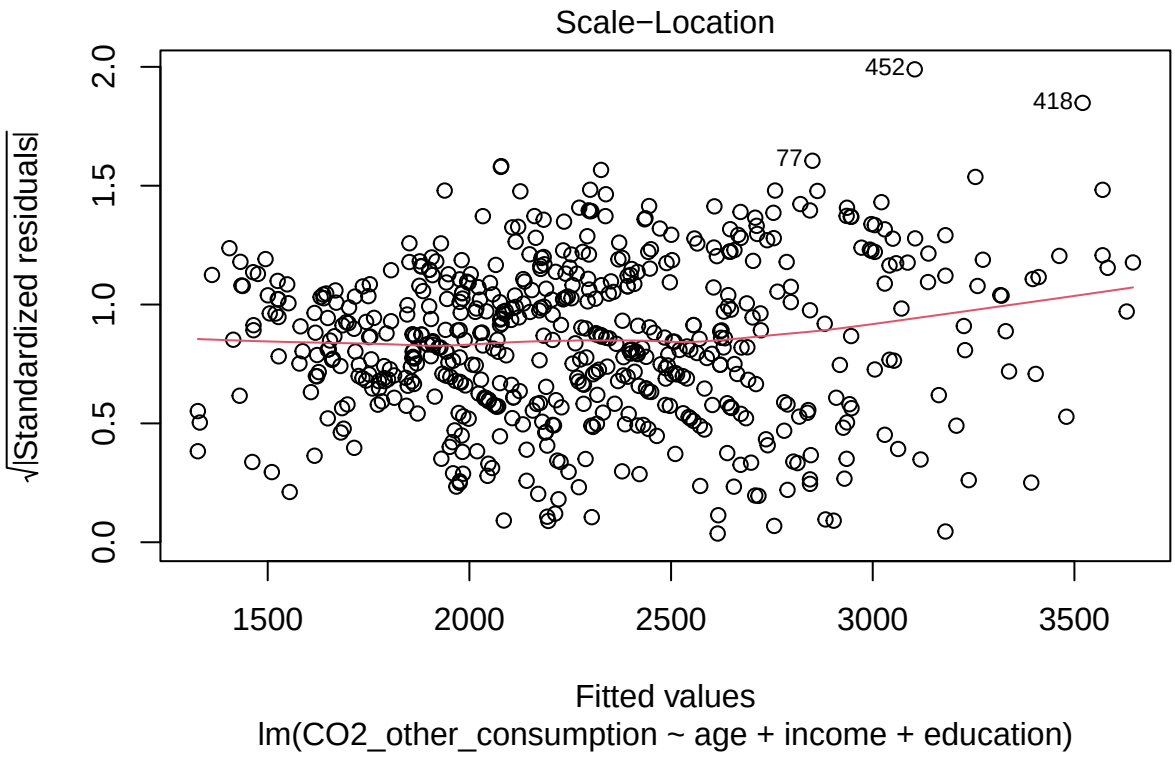
```
##
## Call:
## lm(formula = CO2_other_consumption ~ age + income + education,
##     data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1648.2  -597.4    18.7   493.1  2979.6
##
## Coefficients:
##                                                                    Estimate
## (Intercept)                                                      1459.16394
```

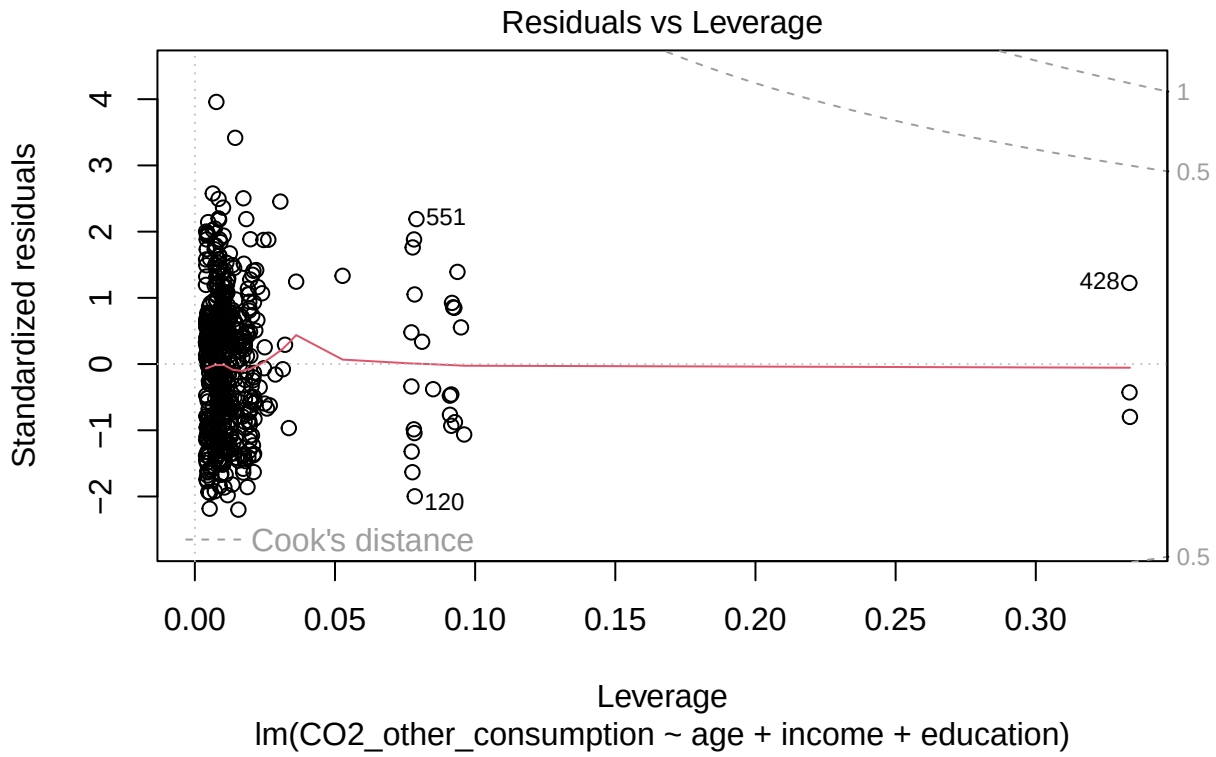
```
## age                                                                          8.24859
## income                                                                       0.21641
## education(Noch) kein Abschluss                                            -107.83899
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -305.61520
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule      -228.53941
## educationDoktorgrad oder Habilitation                                     -296.71997
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  -445.39275
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss -275.92855
##                                                                           Std. Error
## (Intercept)                                                                116.37738
## age                                                                          2.46512
## income                                                                       0.01676
## education(Noch) kein Abschluss                                             440.38123
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   84.41276
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule        86.67182
## educationDoktorgrad oder Habilitation                                      215.04527
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   235.49506
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss  110.02315
##                                                                           t value
## (Intercept)                                                                12.538
## age                                                                         3.346
## income                                                                     12.913
## education(Noch) kein Abschluss                                             -0.245
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  -3.620
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule       -2.637
## educationDoktorgrad oder Habilitation                                      -1.380
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   -1.891
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss  -2.508
##                                                                           Pr(>|t|)
## (Intercept)                                                                < 2e-16
## age                                                                       0.000873
## income                                                                     < 2e-16
## education(Noch) kein Abschluss                                            0.806639
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) 0.000320
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule      0.008593
## educationDoktorgrad oder Habilitation                                     0.168180
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  0.059083
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss 0.012417
##
## (Intercept)                                                               ***
## age                                                                       ***
## income                                                                    ***
## 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 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 755.7 on 579 degrees of freedom
## Multiple R-squared:  0.2777, Adjusted R-squared:  0.2677
## F-statistic: 27.83 on 8 and 579 DF,  p-value: < 2.2e-16
```

```
plot(step_model1)
```

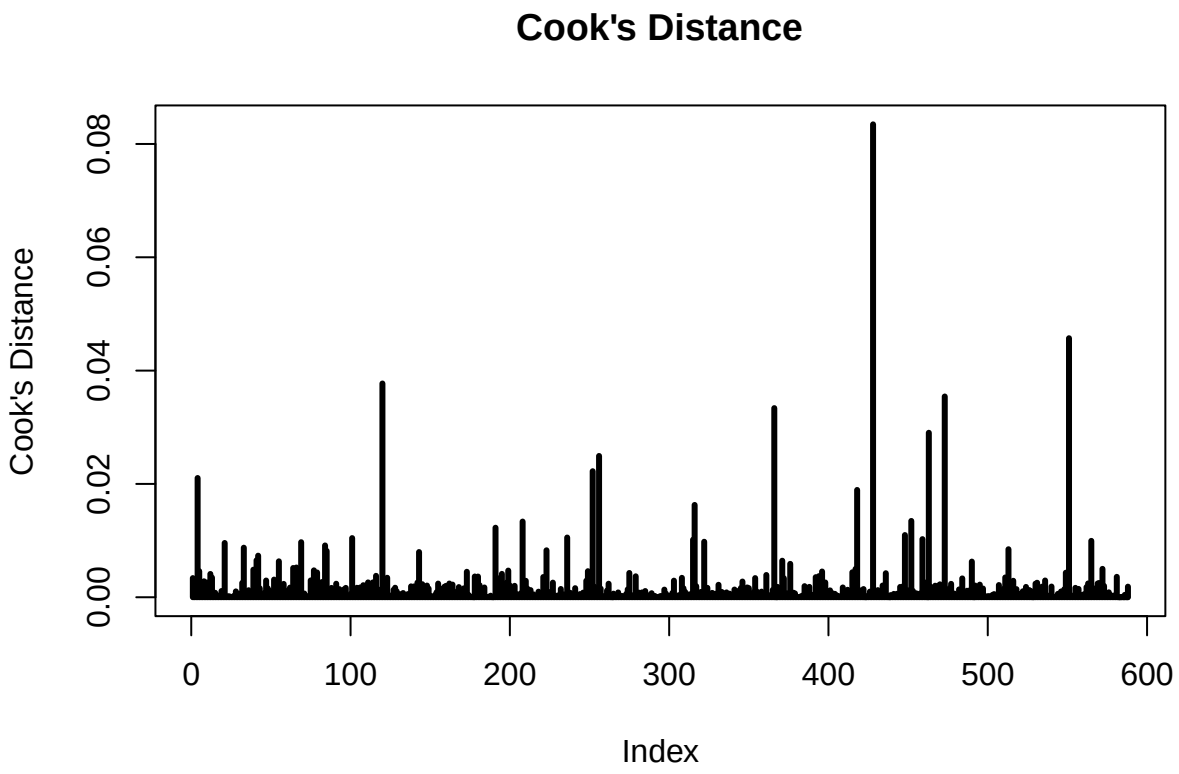



Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(CO2_other_consumption ~ age + income + education)
452
418
77
−3
−2
−1
0
1
2
3
−2
−1
0
1
2
3
4


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
452
418
77
1500
2000
2500
3000
3500
Fitted values
lm(CO2_other_consumption ~ age + income + education)



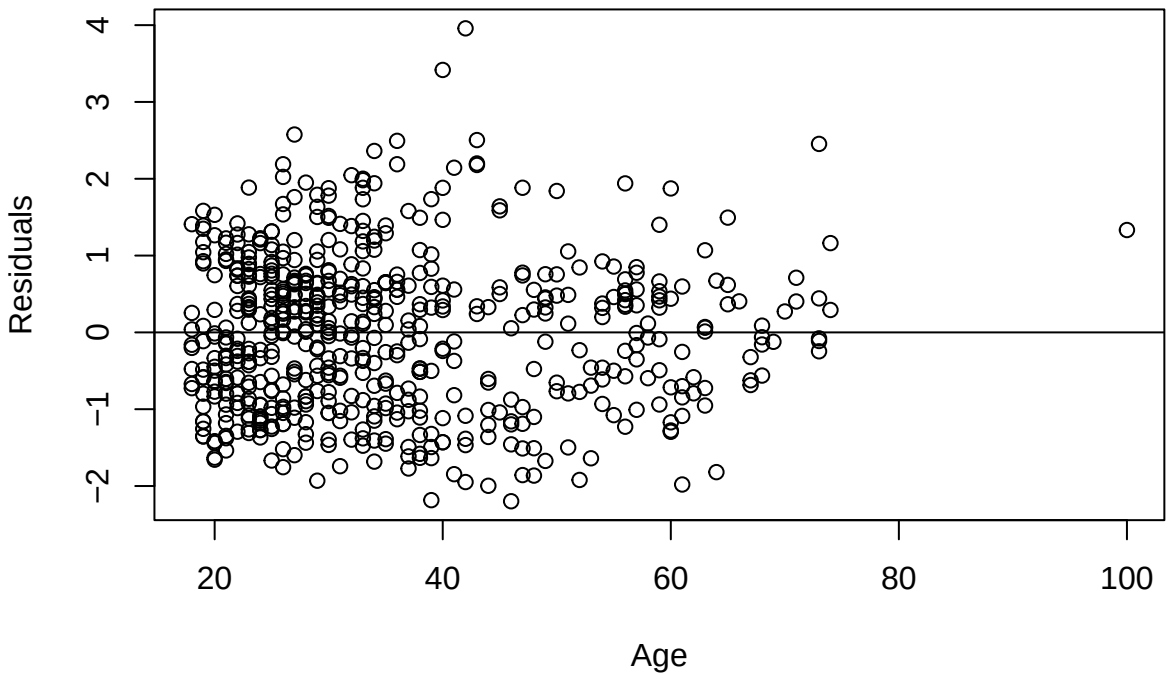

```
# Checking outliers: estimate of the influence of data point; summary of how much a regression model ch

cook = cooks.distance(step_model1)
plot(cook,
     type="h",
     lwd=3,
     ylab = "Cook's Distance",
     main="Cook's Distance")
abline(h = 1)
```



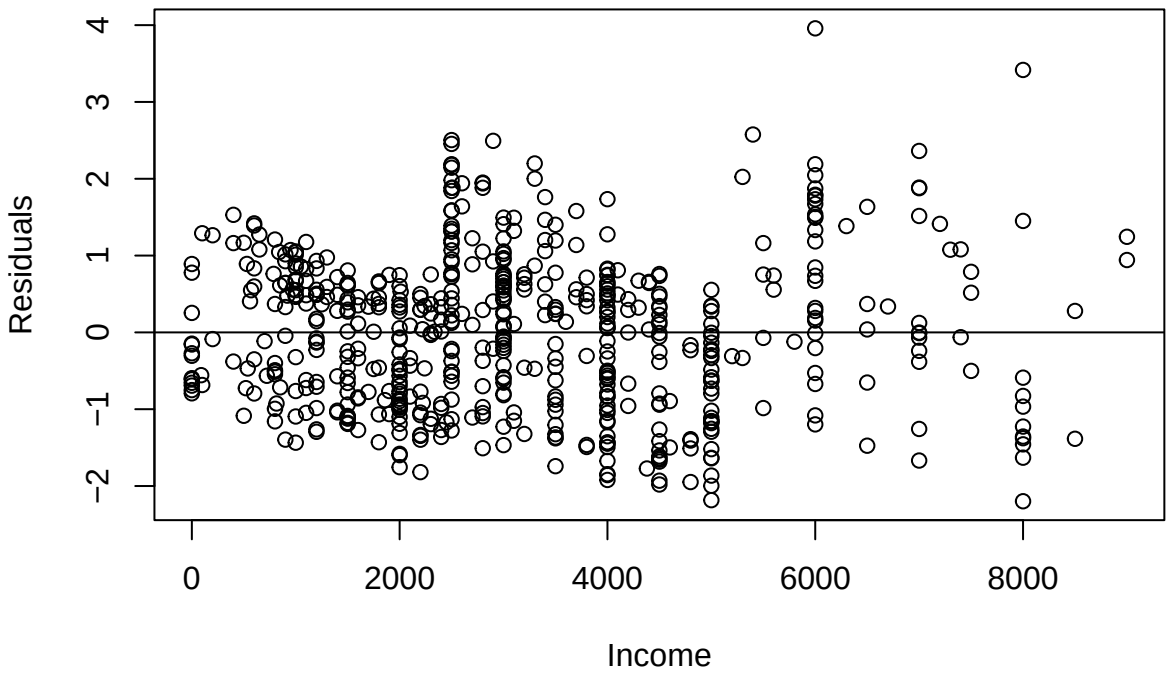

```
influential = cooks.distance(step_model1)[which(cook >1)]

influential
```

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

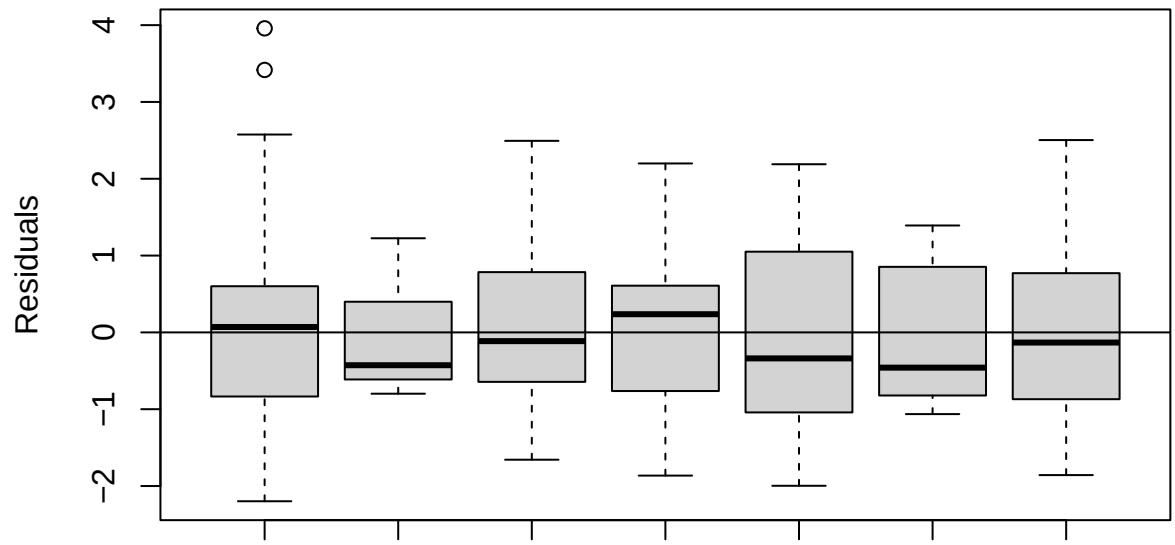
```
res1 = stdres(step_model1) ## (Standardized) Residuals

# 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)
```



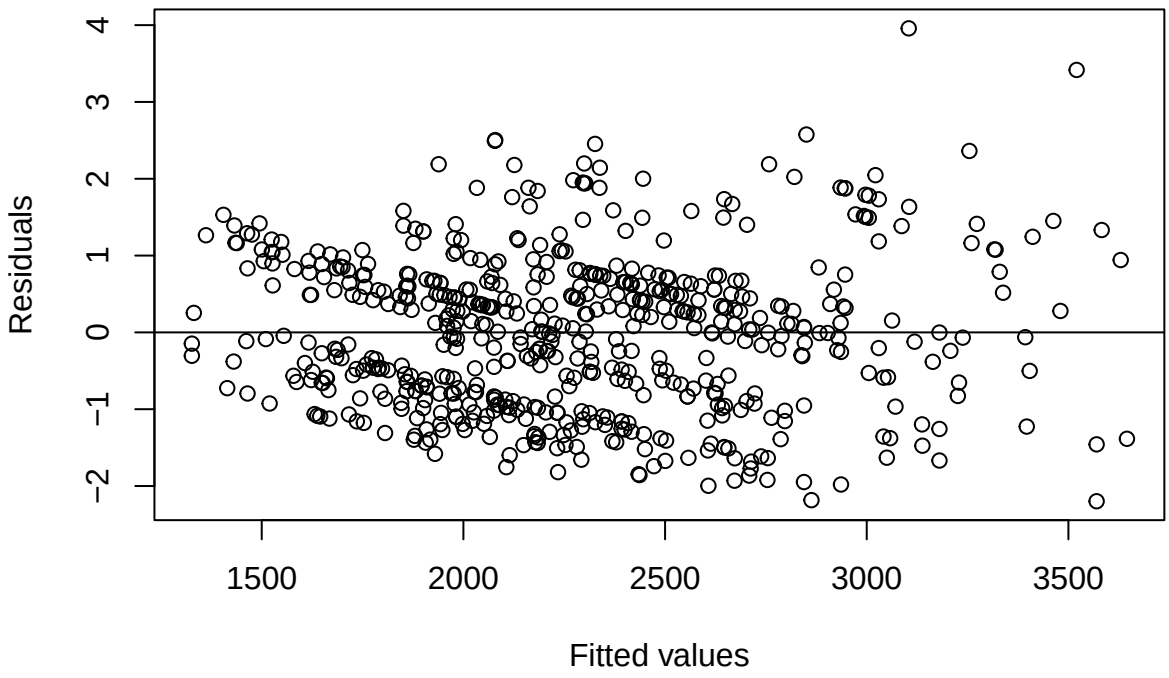

```
plot(df1$education, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



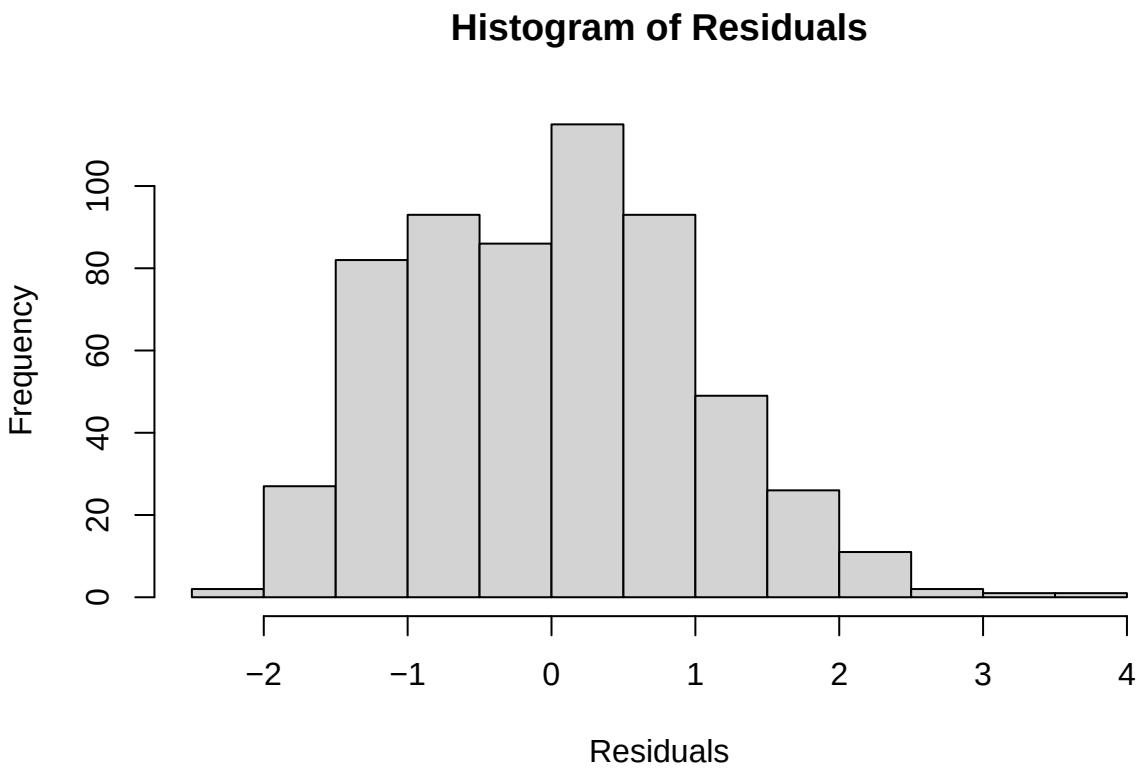
schluss (Bachelor, Master, Magister, Diplom, Staatsexamen)

Income

```
# Constant variance and independent error term assumption

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



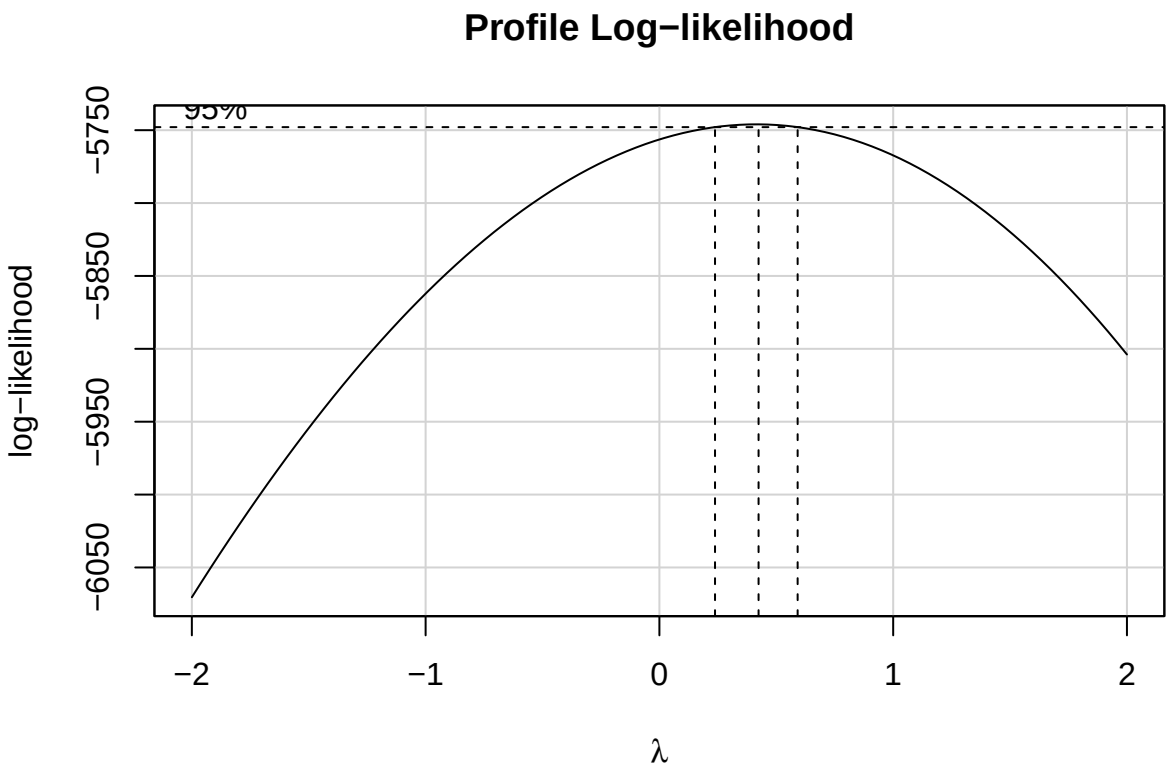

```
# Normality assumption

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




```
## normality test using shapiro-test: reject the H0
#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 = 0.98689, p-value = 3.945e-05
```

```
bc = boxCox(step_model1)
```




```
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
```

```
#Non-linear transformation to the selected variables

options(scipen = -2)
model_extra <- lm(sqrt(CO2_other_consumption) ~ age + income + education, data = df1)

summary(model_extra)
```

FINAL MODEL

```
##
## Call:
## lm(formula = sqrt(CO2_other_consumption) ~ age + income + education,
##     data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.9019  -6.4128   0.9995   5.6666  22.7669
```

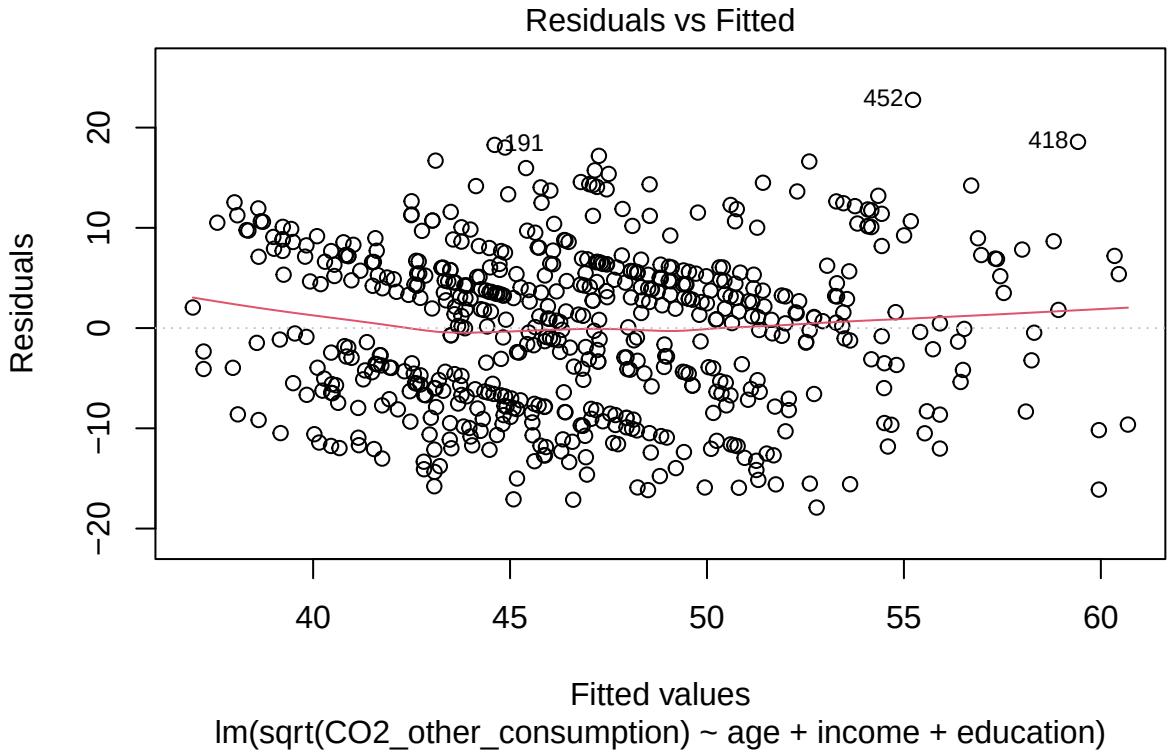
```
##
## Coefficients:
##                                                                                  Estimate
## (Intercept)                                                                     3.843e+01
## age                                                                             8.823e-02
## income                                                                          2.183e-03
## education(Noch) kein Abschluss                                                 -7.364e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -3.063e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule           -2.275e+00
## educationDoktorgrad oder Habilitation                                          -3.281e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss -4.751e+00
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss     -3.071e+00
##                                                                                Std. Error
## (Intercept)                                                                     1.222e+00
## age                                                                             2.589e-02
## income                                                                          1.760e-04
## education(Noch) kein Abschluss                                                  4.626e+00
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  8.867e-01
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule            9.104e-01
## educationDoktorgrad oder Habilitation                                           2.259e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   2.474e+00
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss      1.156e+00
##                                                                                t value
## (Intercept)                                                                     31.435
## age                                                                              3.407
## income                                                                          12.399
## education(Noch) kein Abschluss                                                  -0.159
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  -3.454
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule            -2.498
## educationDoktorgrad oder Habilitation                                           -1.452
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   -1.920
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss      -2.657
##                                                                                Pr(>|t|)
## (Intercept)                                                                     < 2e-16
## age                                                                            7.01e-04
## income                                                                          < 2e-16
## education(Noch) kein Abschluss                                                 8.74e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) 5.92e-04
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule           1.28e-02
## educationDoktorgrad oder Habilitation                                          1.47e-01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  5.53e-02
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss     8.10e-03
##
## (Intercept)                                                                    ***
## age                                                                            ***
## income                                                                         ***
## 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     **
## ---
## Signif. codes:  0 '***' 1e-03 '**' 1e-02 '*' 5e-02 '.' 0.1 ' ' 1
```

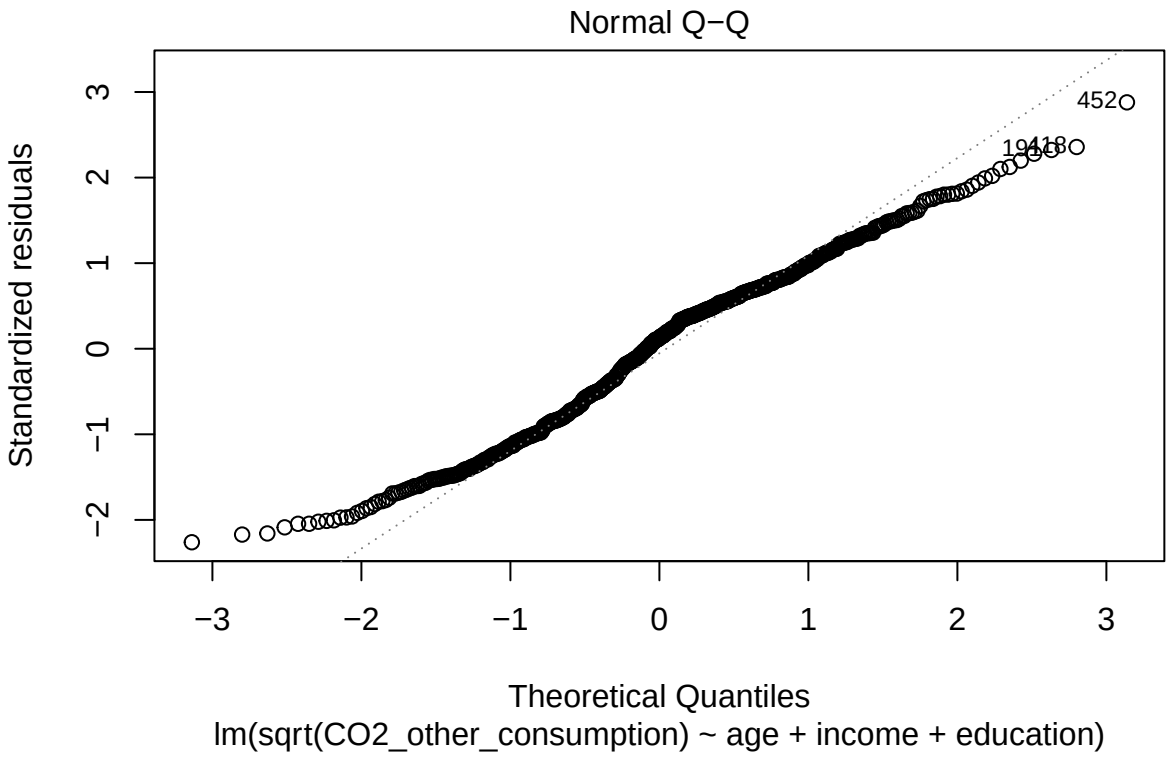
```
##
## Residual standard error: 7.938 on 579 degrees of freedom
## Multiple R-squared:  0.2642, Adjusted R-squared:  0.254
## F-statistic: 25.99 on 8 and 579 DF,  p-value: < 2.2e-16
```

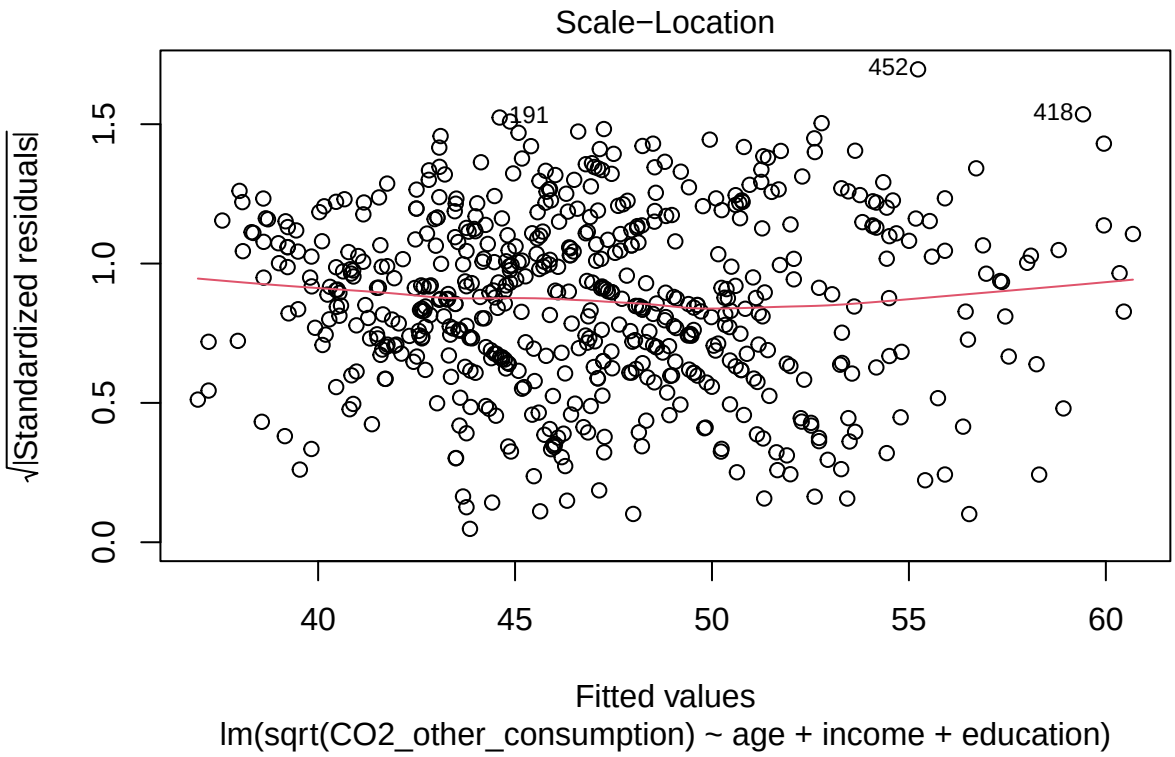
```
vif(model_extra)
```

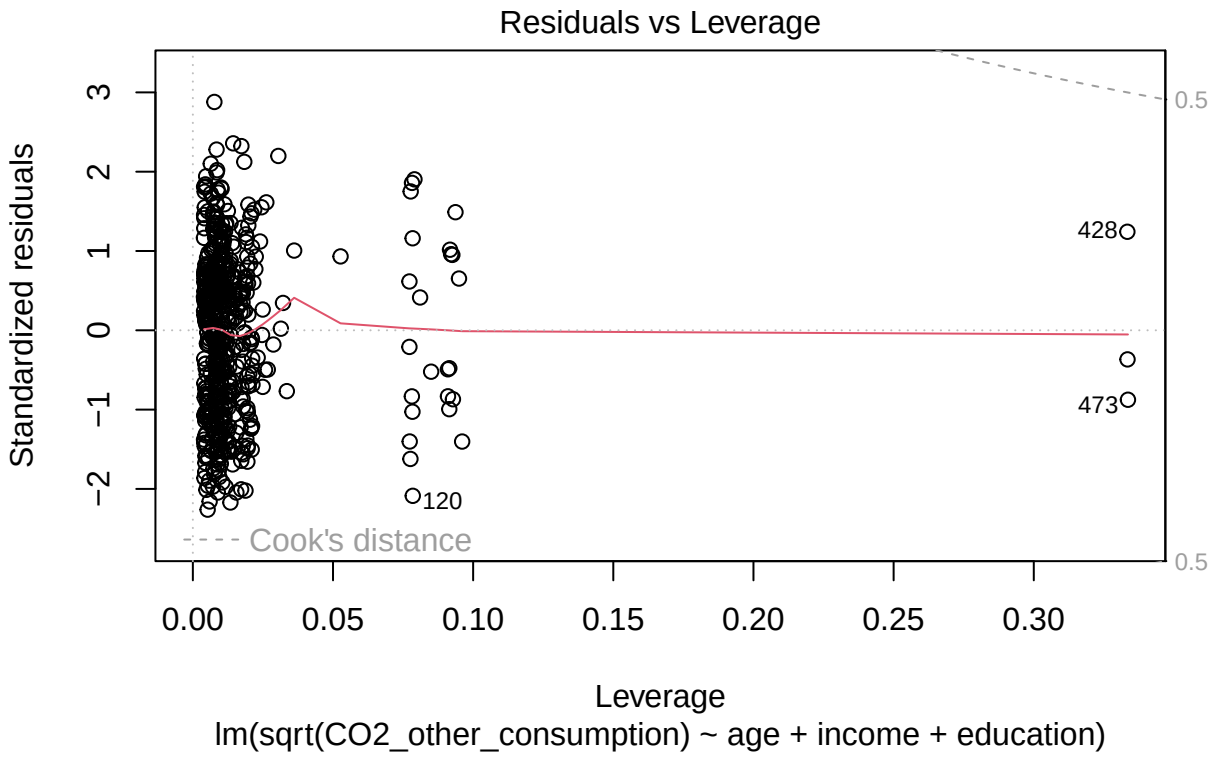
```
##                GVIF Df GVIF^(1/(2*Df))
## age       1.194146  1        1.092770
## income    1.034515  1        1.017111
## education 1.232726  6        1.017589
```

```
plot(model_extra)
```

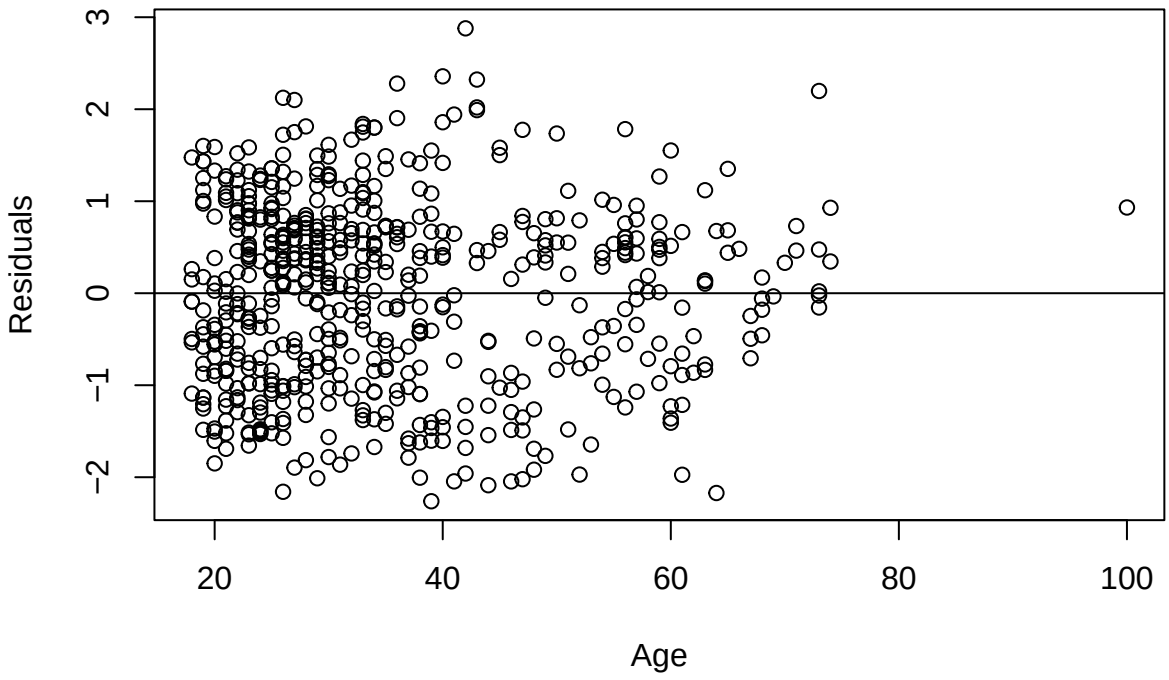



Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(sqrt(CO2_other_consumption) ~ age + income + education)
452
1941
418
−3
−2
−1
0
1
2
3

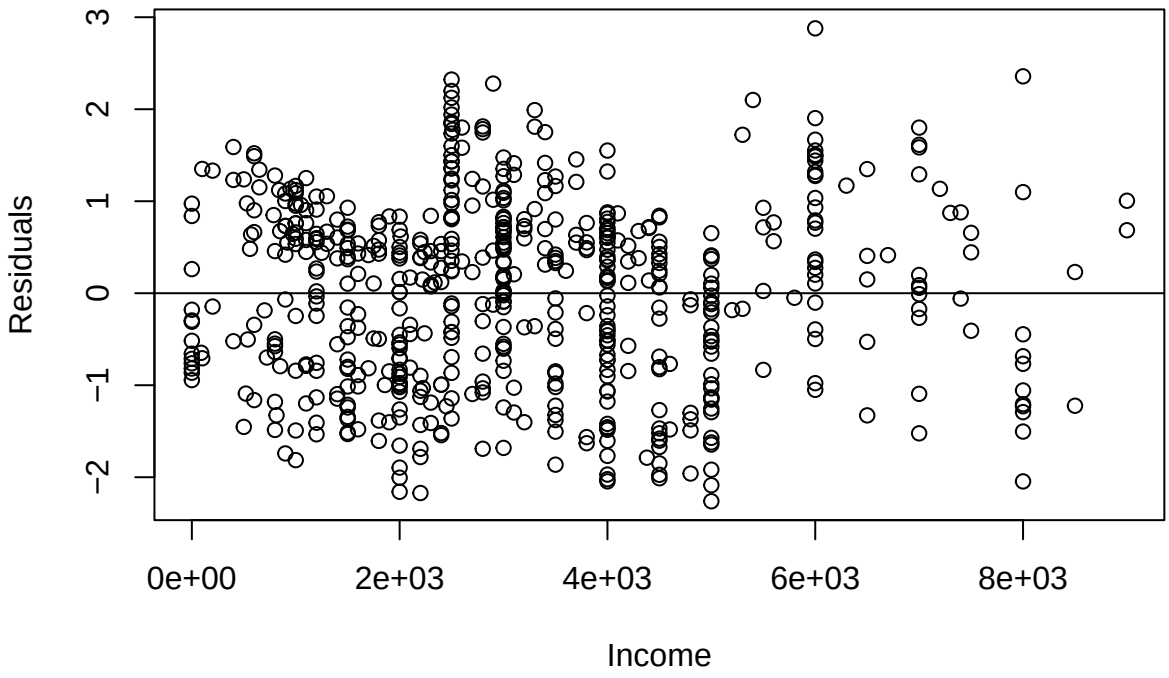

Scale−Location
452
191
418
√|Standardized residuals|
0.0
0.5
1.0
1.5
40
45
50
55
60
Fitted values
lm(sqrt(CO2_other_consumption) ~ age + income + education)



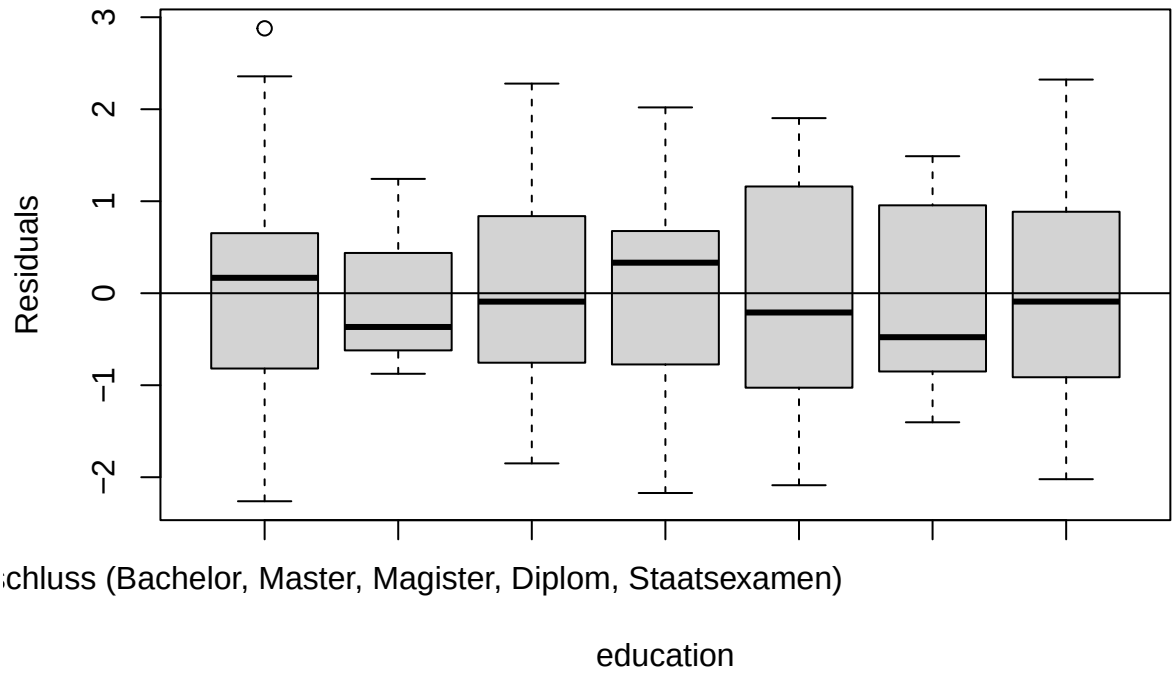

```
res1 = stdres(model_extra) ## (Standardized) Residuals

# 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)
```

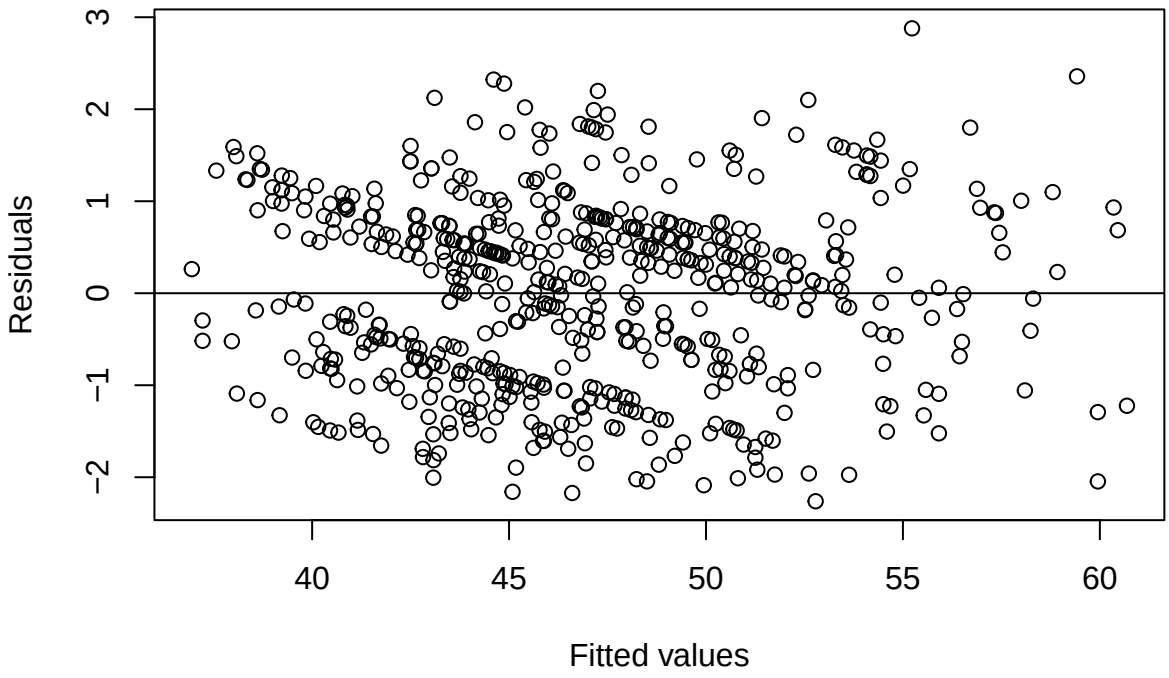



```
#plot(df1_scaled$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
#abline(h = 0)

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




```
#plot(df1_scaled$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(model_extra), res1, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```



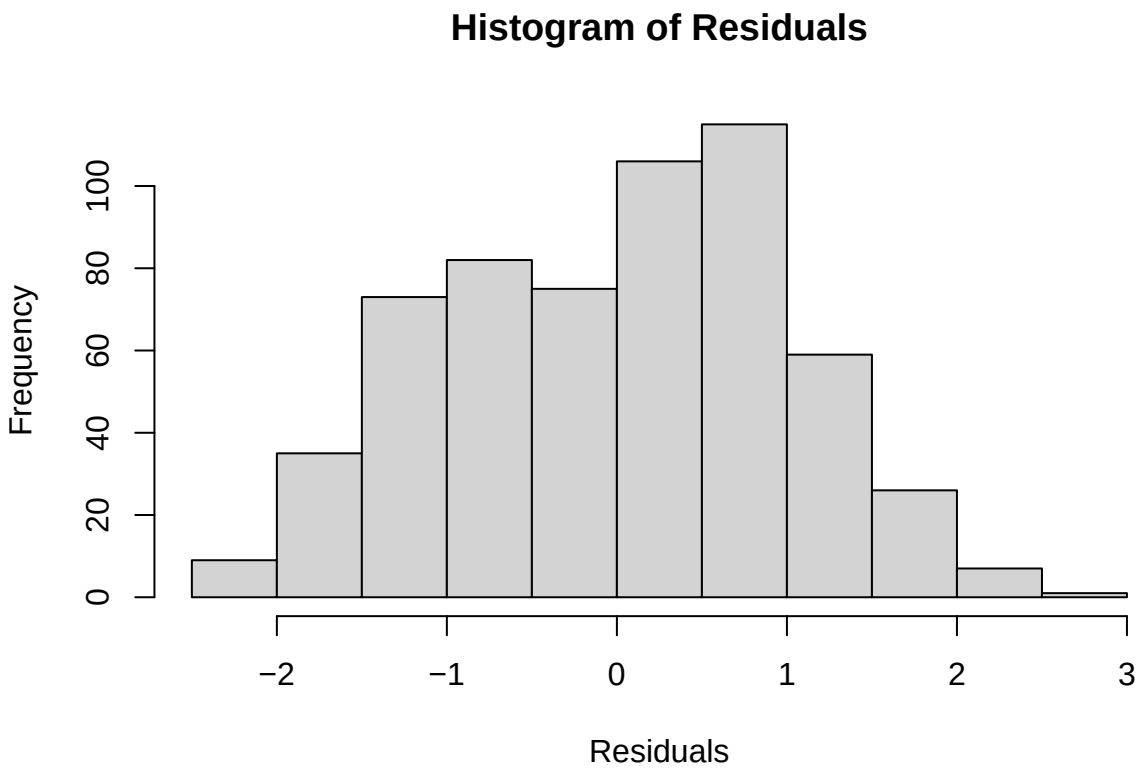

```
# Durbin-Watson Test: Independence of the error terms
# HO (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(model_extra)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1      0.03112004      1.931814   0.344
##  Alternative hypothesis: rho != 0
```

```
# Breusch-Pagan Test: Heteroscedasticity
# HO: Homoscedasticity is present

bptest(model_extra)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model_extra
## BP = 13.357, df = 8, p-value = 0.1001
```

```
# Normality assumption

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

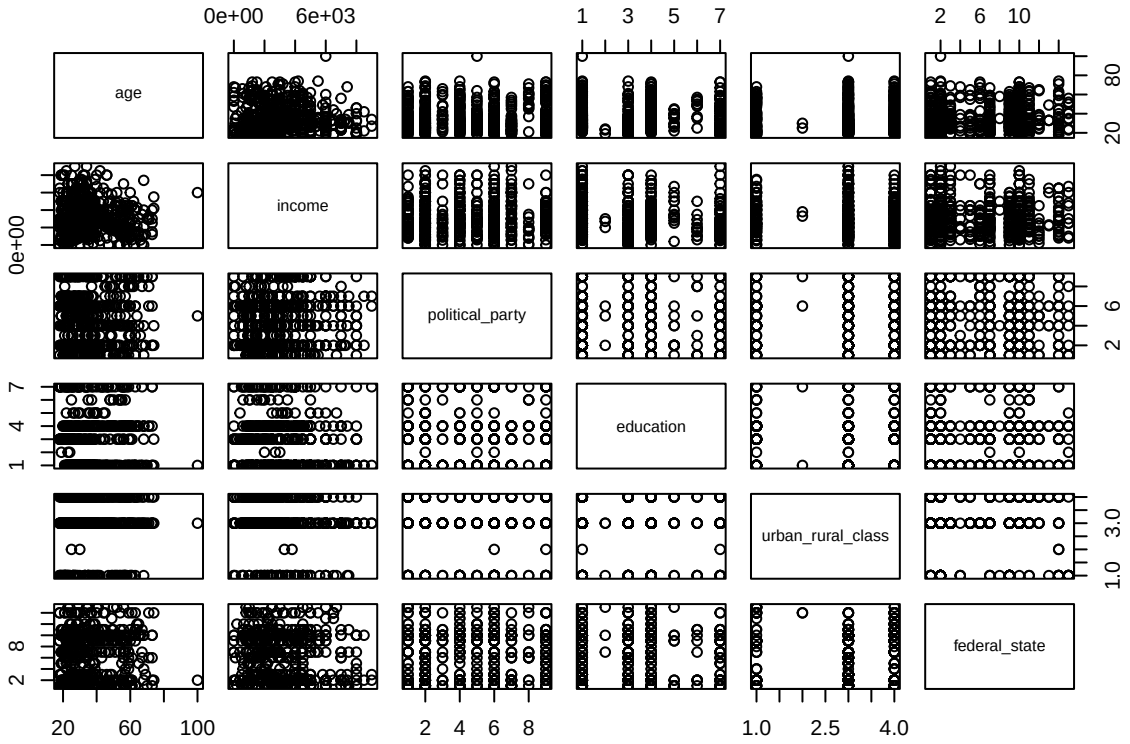



```
## normality test using shapiro-test: reject the H0
#H0:  the sample comes from a normal distribution

res1_num = res1[is.finite(res1)]
shapiro.test(res1_num)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res1_num
## W = 0.98436, p-value = 6.099e-06
```

III. Multivariate Regression: belief diff other consumption

```
# 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, D
df2$urban_rural_class  <- relevel(df2$urban_rural_class, ref='sehr zentral')
df2$federal_state  <- relevel(df2$federal_state, ref='Nordrhein-Westfalen')
```

```
# regression model with all variables

model2 <- lm(belief_diff_other_consumption ~ age + income + political_party + education +  urban_rural_
summary(model2)
```

```
##
## Call:
## lm(formula = belief_diff_other_consumption ~ age + income + political_party +
##     education + urban_rural_class + federal_state, data = df2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -75.558 -24.675   0.221  21.775  99.279
##
## Coefficients:
##                                                                    Estimate
## (Intercept)                                                       4.985e+00
## age                                                              -2.360e-01
```

```
## income                                                                    -5.800e-03
## political_partyAfD                                                         7.001e+00
## political_partyBündnis Sarah Wagenknecht                                   3.395e+00
## political_partyCDU/CSU                                                     6.459e+00
## political_partyDie Linke                                                   1.129e+01
## political_partyEiner anderen Partei                                        3.543e-01
## political_partyFDP                                                         1.563e+00
## political_partyKeine Angabe                                                2.111e+01
## political_partySPD                                                         9.053e+00
## education(Noch) kein Abschluss                                             2.138e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  1.053e+01
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule       9.324e+00
## educationDoktorgrad oder Habilitation                                      1.015e+01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   1.402e+01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         1.722e+01
## urban_rural_classperipher                                                 -7.067e-01
## urban_rural_classsehr peripher                                             3.822e+01
## urban_rural_classzentral                                                   4.793e+00
## federal_stateBaden-Württemberg                                             3.992e+00
## federal_stateBayern                                                        2.191e+00
## federal_stateBerlin                                                        4.771e+00
## federal_stateBrandenburg                                                  -4.672e-01
## federal_stateBremen                                                        3.220e+00
## federal_stateHamburg                                                      -8.957e+00
## federal_stateHessen                                                       -1.093e+00
## federal_stateMecklenburg-Vorpommern                                        1.513e+01
## federal_stateNiedersachsen                                                 3.517e+00
## federal_stateRheinland-Pfalz                                               1.589e+00
## federal_stateSaarland                                                     -8.264e+00
## federal_stateSachsen-Anhalt                                               -1.142e+01
## federal_stateSchleswig-Holstein                                           -6.946e+00
## federal_stateThüringen                                                     1.292e+00
##                                                                          Std. Error
## (Intercept)                                                                6.361e+00
## age                                                                        1.126e-01
## income                                                                     7.526e-04
## political_partyAfD                                                         5.388e+00
## political_partyBündnis Sarah Wagenknecht                                   7.552e+00
## political_partyCDU/CSU                                                     4.840e+00
## political_partyDie Linke                                                   5.833e+00
## political_partyEiner anderen Partei                                        4.330e+00
## political_partyFDP                                                         5.605e+00
## political_partyKeine Angabe                                                9.708e+00
## political_partySPD                                                         4.928e+00
## education(Noch) kein Abschluss                                             1.972e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  3.833e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule       3.910e+00
## educationDoktorgrad oder Habilitation                                      9.557e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   1.091e+01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         4.972e+00
## urban_rural_classperipher                                                  5.010e+00
## urban_rural_classsehr peripher                                             2.480e+01
## urban_rural_classzentral                                                   3.673e+00
## federal_stateBaden-Württemberg                                             4.772e+00
```

```
## federal_stateBayern                                                                 4.899e+00
## federal_stateBerlin                                                                 5.952e+00
## federal_stateBrandenburg                                                            1.250e+01
## federal_stateBremen                                                                 9.118e+00
## federal_stateHamburg                                                                7.420e+00
## federal_stateHessen                                                                 5.693e+00
## federal_stateMecklenburg-Vorpommern                                                 2.390e+01
## federal_stateNiedersachsen                                                          5.792e+00
## federal_stateRheinland-Pfalz                                                        7.144e+00
## federal_stateSaarland                                                               1.122e+01
## federal_stateSachsen-Anhalt                                                         1.729e+01
## federal_stateSchleswig-Holstein                                                     8.356e+00
## federal_stateThüringen                                                              1.266e+01
##                                                                                    t value
## (Intercept)                                                                         0.784
## age                                                                                -2.096
## income                                                                             -7.707
## political_partyAfD                                                                  1.299
## political_partyBündnis Sarah Wagenknecht                                            0.450
## political_partyCDU/CSU                                                              1.335
## political_partyDie Linke                                                            1.936
## political_partyEiner anderen Partei                                                 0.082
## political_partyFDP                                                                  0.279
## political_partyKeine Angabe                                                         2.174
## political_partySPD                                                                  1.837
## education(Noch) kein Abschluss                                                      1.084
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   2.749
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                2.385
## educationDoktorgrad oder Habilitation                                               1.062
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss    1.285
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss          3.464
## urban_rural_classperipher                                                          -0.141
## urban_rural_classsehr peripher                                                      1.541
## urban_rural_classzentral                                                            1.305
## federal_stateBaden-Württemberg                                                      0.837
## federal_stateBayern                                                                 0.447
## federal_stateBerlin                                                                 0.802
## federal_stateBrandenburg                                                           -0.037
## federal_stateBremen                                                                 0.353
## federal_stateHamburg                                                               -1.207
## federal_stateHessen                                                                -0.192
## federal_stateMecklenburg-Vorpommern                                                 0.633
## federal_stateNiedersachsen                                                          0.607
## federal_stateRheinland-Pfalz                                                        0.222
## federal_stateSaarland                                                              -0.737
## federal_stateSachsen-Anhalt                                                        -0.661
## federal_stateSchleswig-Holstein                                                    -0.831
## federal_stateThüringen                                                              0.102
##                                                                                   Pr(>|t|)
## (Intercept)                                                                       4.34e-01
## age                                                                               3.66e-02
## income                                                                            6.01e-14
## political_partyAfD                                                                1.94e-01
## political_partyBündnis Sarah Wagenknecht                                          6.53e-01
```

```
## political_partyCDU/CSU                                                                1.83e-01
## political_partyDie Linke                                                               5.33e-02
## political_partyEiner anderen Partei                                                    9.35e-01
## political_partyFDP                                                                     7.80e-01
## political_partyKeine Angabe                                                            3.01e-02
## political_partySPD                                                                     6.67e-02
## education(Noch) kein Abschluss                                                         2.79e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)      6.18e-03
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                   1.74e-02
## educationDoktorgrad oder Habilitation                                                  2.89e-01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss       1.99e-01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss             5.73e-04
## urban_rural_classperipher                                                              8.88e-01
## urban_rural_classsehr peripher                                                         1.24e-01
## urban_rural_classzentral                                                               1.92e-01
## federal_stateBaden-Württemberg                                                         4.03e-01
## federal_stateBayern                                                                    6.55e-01
## federal_stateBerlin                                                                    4.23e-01
## federal_stateBrandenburg                                                               9.70e-01
## federal_stateBremen                                                                    7.24e-01
## federal_stateHamburg                                                                   2.28e-01
## federal_stateHessen                                                                    8.48e-01
## federal_stateMecklenburg-Vorpommern                                                    5.27e-01
## federal_stateNiedersachsen                                                             5.44e-01
## federal_stateRheinland-Pfalz                                                           8.24e-01
## federal_stateSaarland                                                                  4.62e-01
## federal_stateSachsen-Anhalt                                                            5.09e-01
## federal_stateSchleswig-Holstein                                                        4.06e-01
## federal_stateThüringen                                                                 9.19e-01
##
## (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 '***' 1e-03 '**' 1e-02 '*' 5e-02 '.' 0.1 ' ' 1
##
## Residual standard error: 32.92 on 554 degrees of freedom
## Multiple R-squared:  0.1859, Adjusted R-squared:  0.1374
## F-statistic: 3.834 on 33 and 554 DF,  p-value: 2.443e-11
```

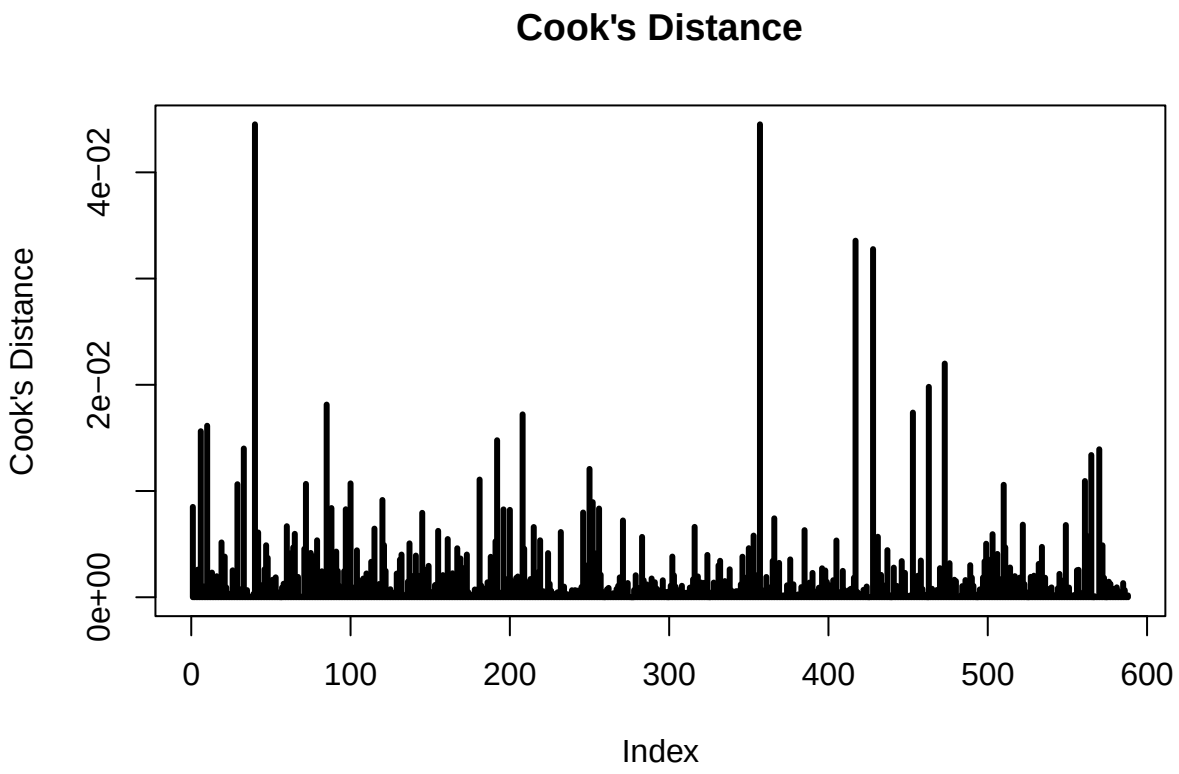
```
# Checking the VIFs for multicollinearity

vif(model2)
```

```
##                       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
```

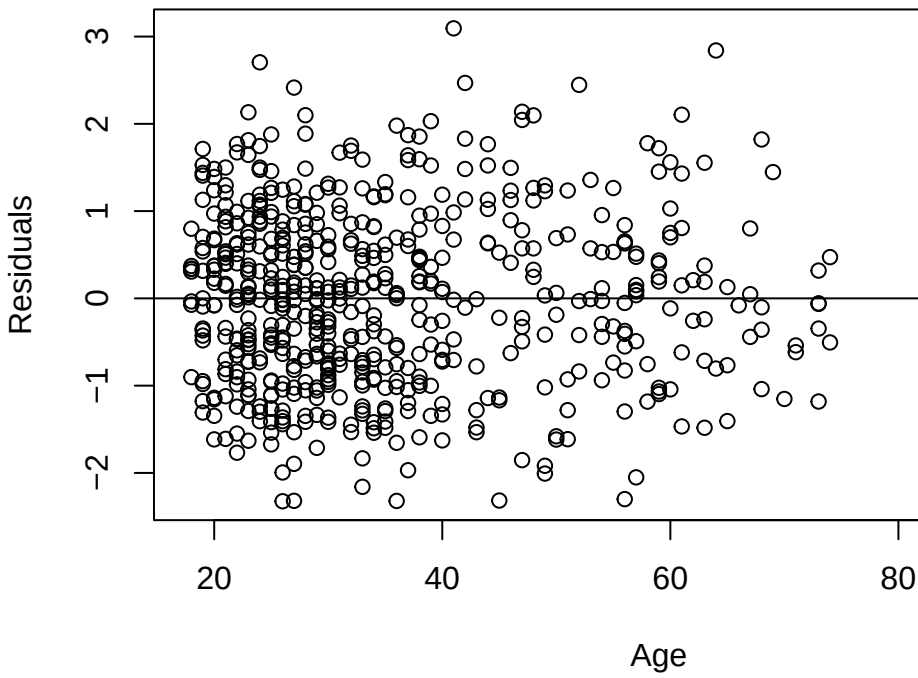
```
# Calculating the threshold for multicollinearity

max(10, 1/(1-summary(model2)$r.square))
```

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

```
# Checking outliers

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

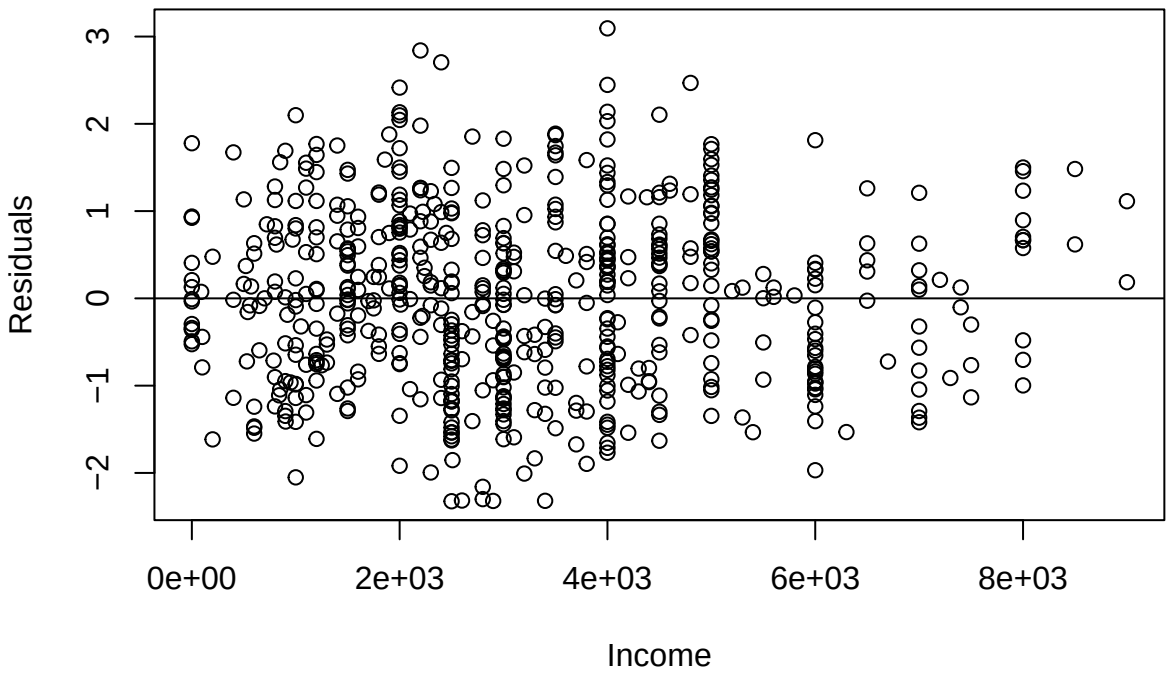



```
res2 = stdres(model2) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

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

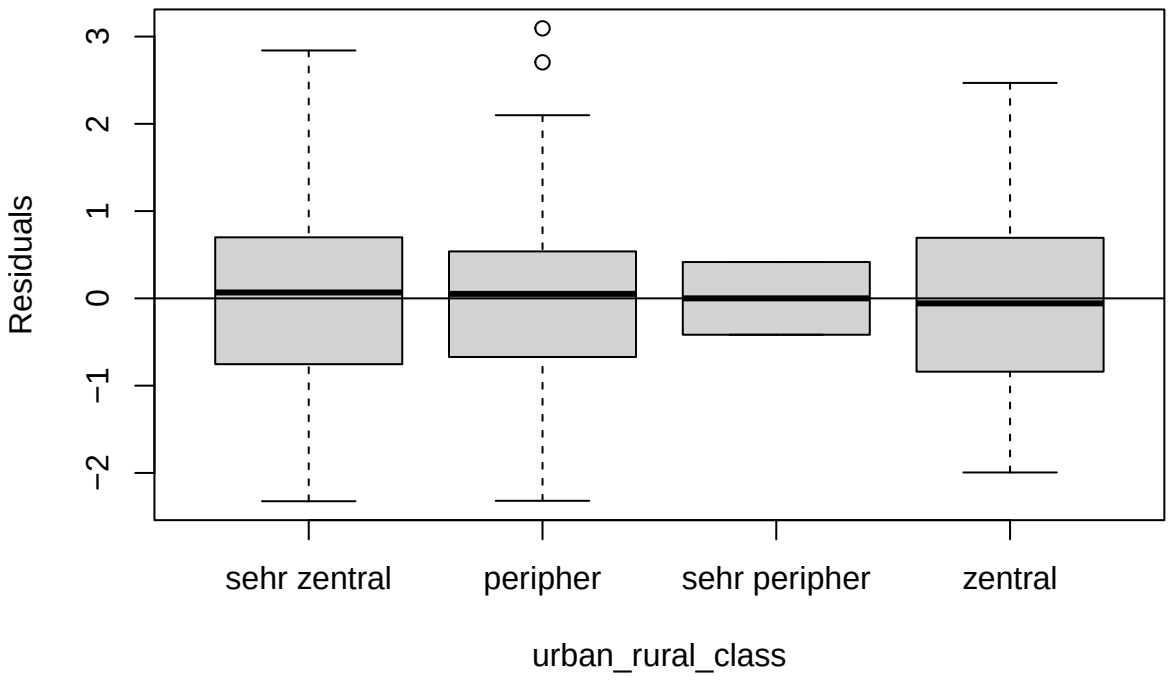


2. Assumptions check in the residuals

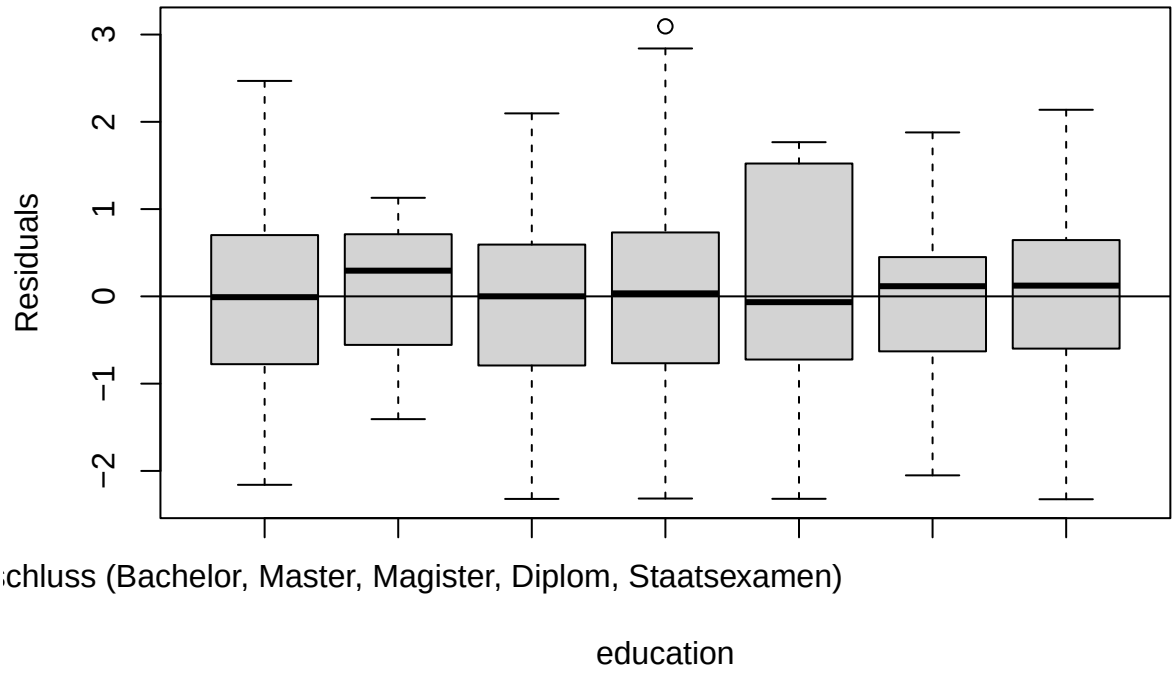
```
plot(df2$income, res2, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



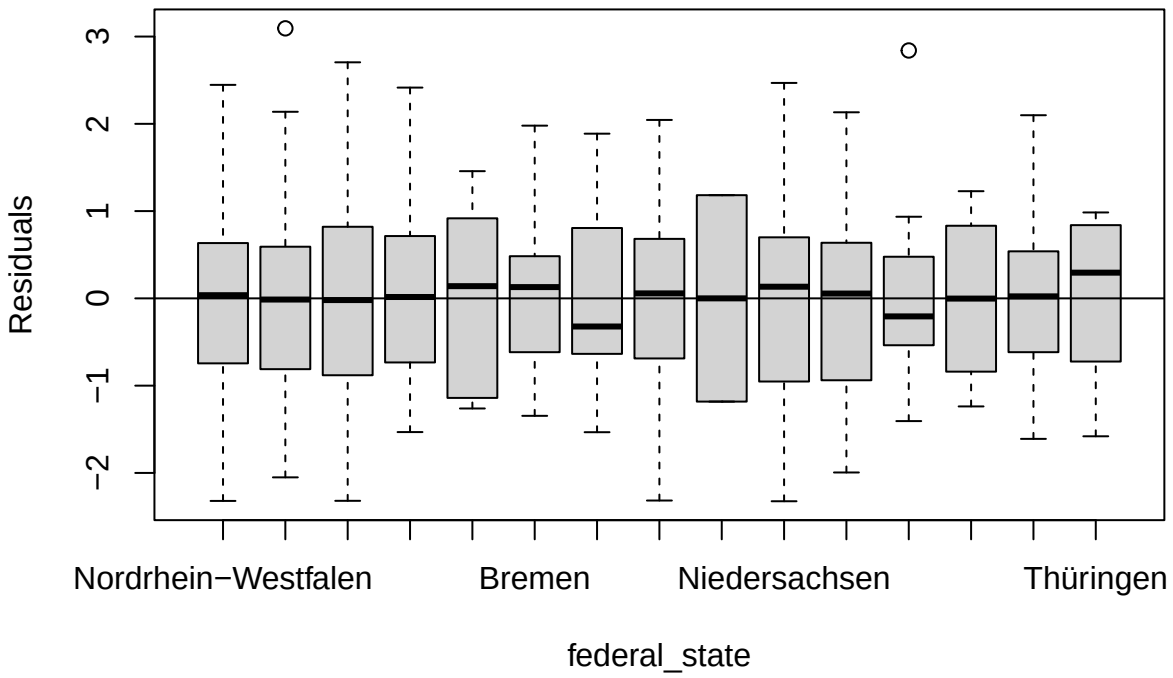

```
plot(df2$urban_rural_class, res2, xlab = "urban_rural_class", ylab = "Residuals")
abline(h = 0)
```



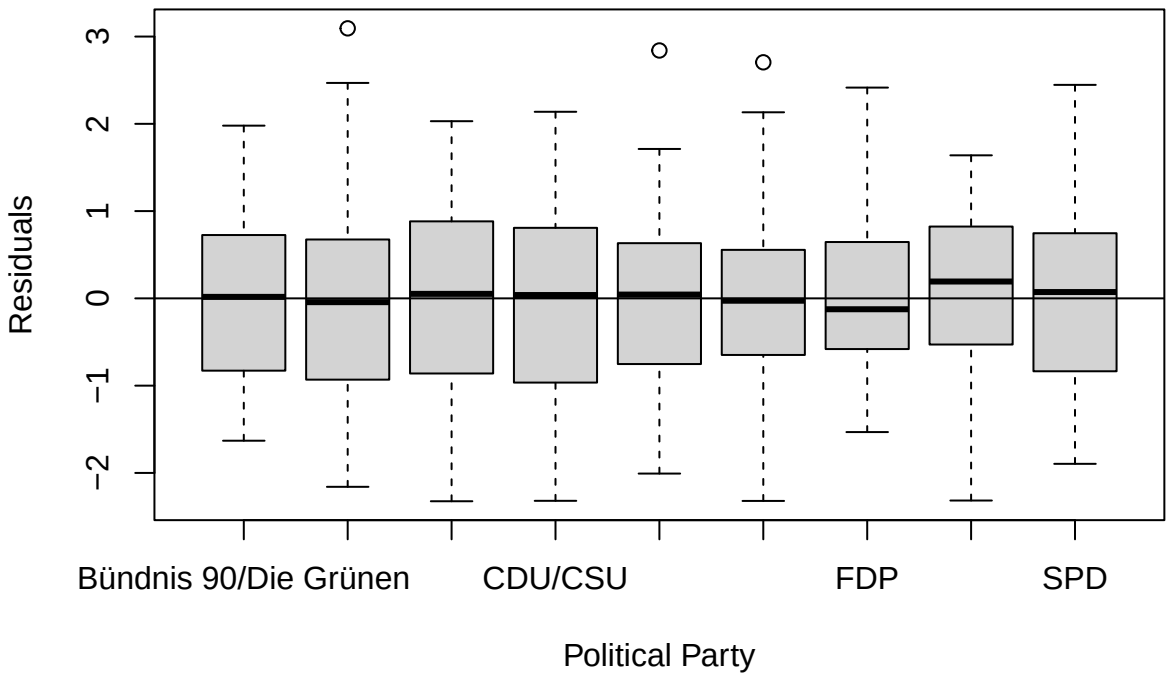

```
plot(df2$education, res2, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



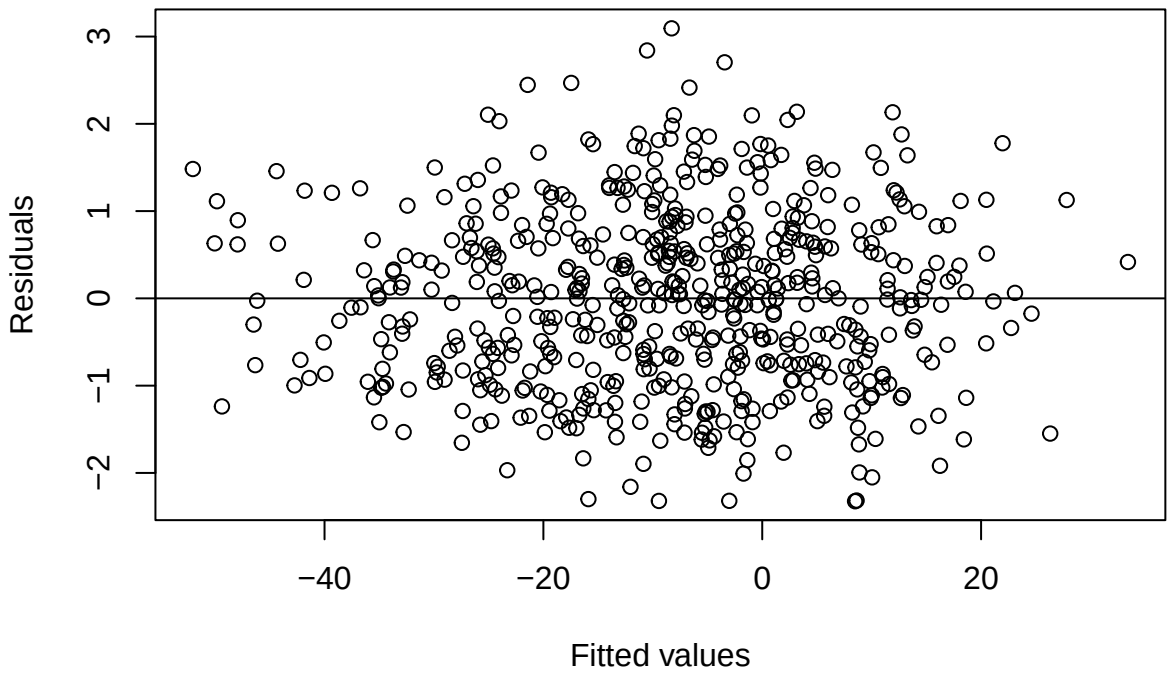

```
plot(df2$federal_state, res2, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```




```
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)
```



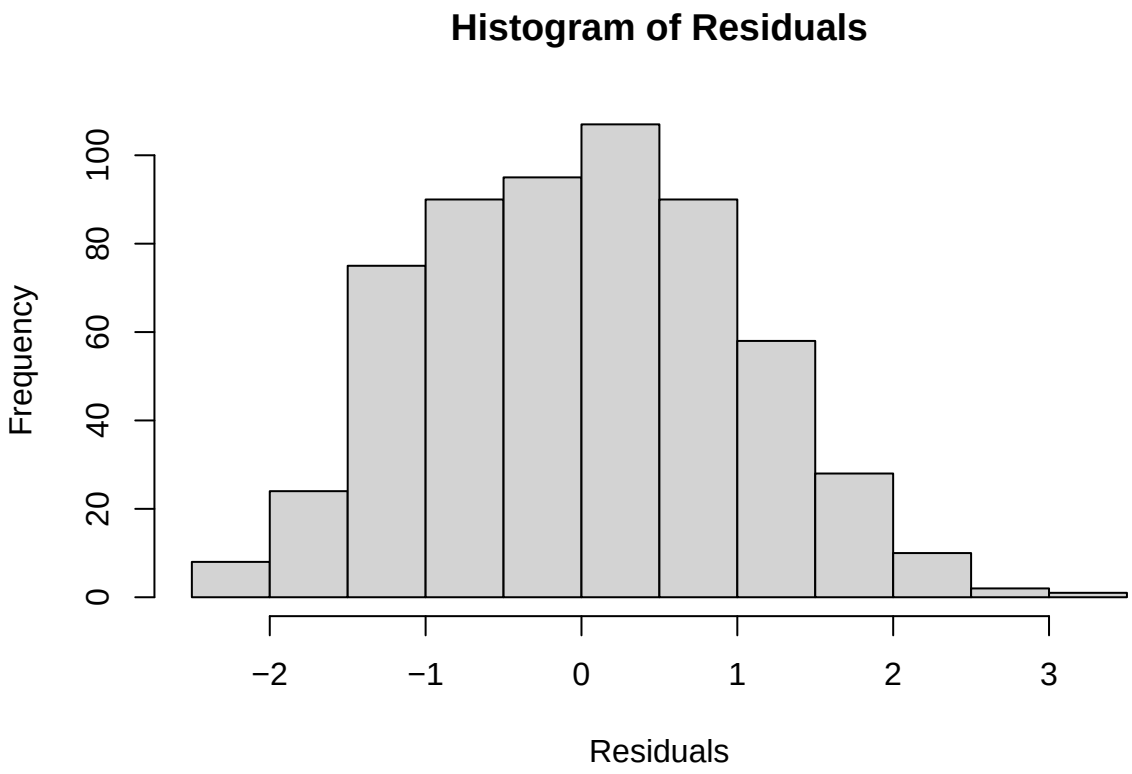

```
# Durbin-Watson Test:  Independence of the error terms
# HO (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(model2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1   -3.216568e-03      2.003034   0.954
##  Alternative hypothesis: rho != 0
```

```
# Breusch-Pagan Test: Heteroscedasticity
# HO: Homoscedasticity is present

bptest(model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model2
## BP = 36.015, df = 33, p-value = 0.3293
```

```
# Normality assumption

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 = 0.99313, p-value = 8.593e-03
```

FINAL MODEL

```
### Backward regression using BIC

options(scipen = -2, digits=2)

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

3. Variable selection

```
## Start:  AIC=4142
## belief_diff_other_consumption ~ age + income + political_party +
##     education + urban_rural_class + federal_state
##
##                      Df Sum of Sq    RSS  AIC
## - federal_state     14      7587 607960 4121
## - political_party    8     11921 612294 4138
## - urban_rural_class  3      4730 605103 4141
## <none>                           600373 4142
## - age                1      4759 605132 4145
## - education          6     19430 619803 4149
## - income             1     64363 664736 4200
##
## Step:  AIC=4121
## belief_diff_other_consumption ~ age + income + political_party +
##     education + urban_rural_class
##
##                      Df Sum of Sq    RSS  AIC
## - political_party    8     10668 618627 4116
## - urban_rural_class  3      4357 612317 4120
## <none>                           607960 4121
## - age                1      5580 613540 4125
## - education          6     19510 627470 4128
## - income             1     65790 673750 4180
##
## Step:  AIC=4116
## belief_diff_other_consumption ~ age + income + education + urban_rural_class
##
##                      Df Sum of Sq    RSS  AIC
## - urban_rural_class  3      4698 623325 4114
## <none>                           618627 4116
## - age                1      3357 621985 4117
## - education          6     23987 642615 4126
## - income             1     68964 687591 4176
##
## Step:  AIC=4114
## belief_diff_other_consumption ~ age + income + education
##
##             Df Sum of Sq    RSS  AIC
## <none>                   623325 4114
## - age        1      3333 626658 4115
## - education  6     25461 648786 4126
## - income     1     70312 693637 4175
```

```
summary(step_model2)
```

```
##
## Call:
## lm(formula = belief_diff_other_consumption ~ age + income + education,
##     data = df2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -82.70 -25.64   0.53  22.59 102.53
```

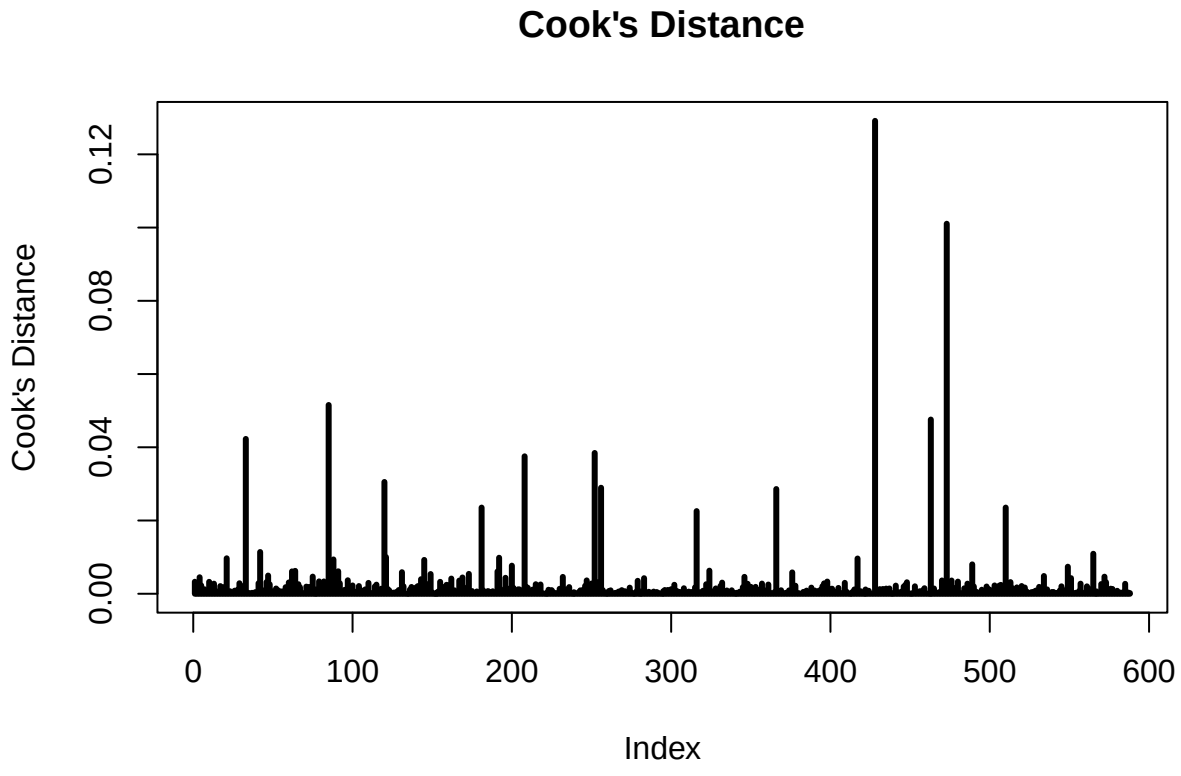
```
##
## Coefficients:
##                                                                                  Estimate
## (Intercept)                                                                      9.85e+00
## age                                                                             -1.88e-01
## income                                                                          -5.88e-03
## education(Noch) kein Abschluss                                                   1.94e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  1.00e+01
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule             9.86e+00
## educationDoktorgrad oder Habilitation                                            1.13e+01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   2.04e+01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss       1.93e+01
##                                                                                 Std. Error
## (Intercept)                                                                       5.05e+00
## age                                                                               1.07e-01
## income                                                                            7.28e-04
## education(Noch) kein Abschluss                                                    1.91e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   3.67e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule              3.76e+00
## educationDoktorgrad oder Habilitation                                             9.34e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss    1.02e+01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss        4.78e+00
##                                                                                 t value
## (Intercept)                                                                        1.95
## age                                                                               -1.76
## income                                                                            -8.08
## education(Noch) kein Abschluss                                                     1.01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)    2.73
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule               2.62
## educationDoktorgrad oder Habilitation                                              1.21
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss    2.00
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         4.04
##                                                                                Pr(>|t|)
## (Intercept)                                                                     5.2e-02
## age                                                                             7.9e-02
## income                                                                          3.7e-15
## education(Noch) kein Abschluss                                                  3.1e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  6.6e-03
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule            9.0e-03
## educationDoktorgrad oder Habilitation                                           2.3e-01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   4.6e-02
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss      6.2e-05
##
## (Intercept)                                                                    .
## age                                                                            .
## income                                                                         ***
## 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     ***
## ---
## Signif. codes:  0 '***' 1e-03 '**' 1e-02 '*' 5e-02 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 33 on 579 degrees of freedom
## Multiple R-squared:  0.155,  Adjusted R-squared:  0.143
## F-statistic: 13.3 on 8 and 579 DF,  p-value: <2e-16
```

```
vif(step_model2)
```

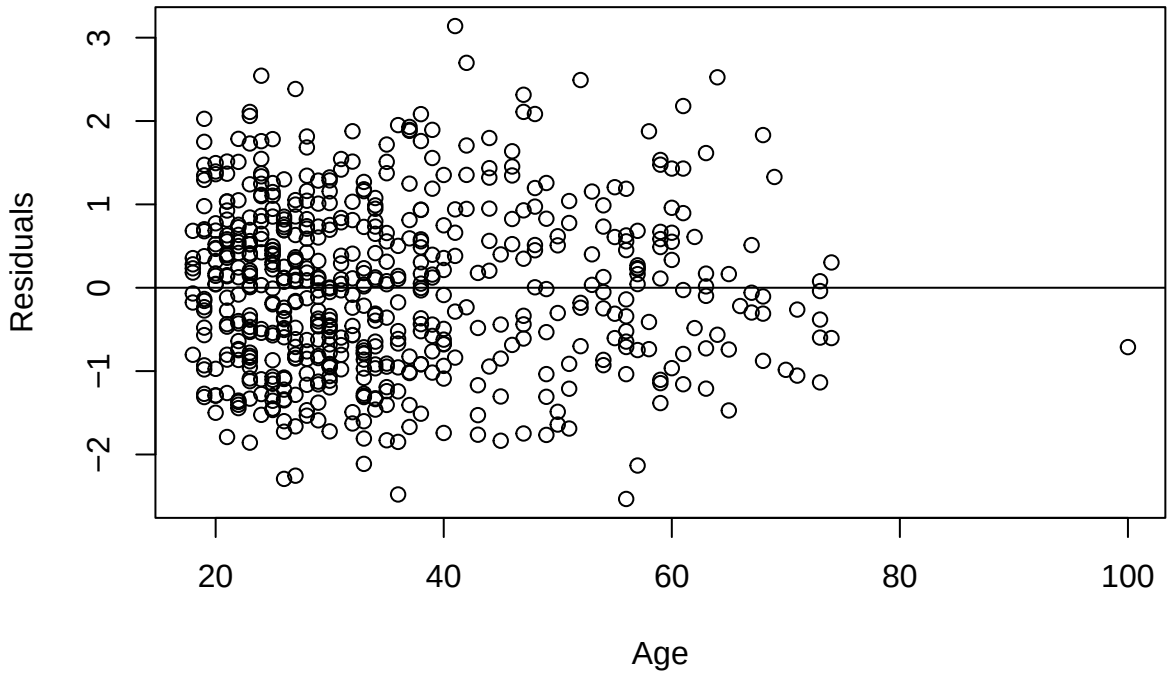
```
##           GVIF Df GVIF^(1/(2*Df))
## age        1.2  1             1.1
## income     1.0  1             1.0
## education  1.2  6             1.0
```

```
# Checking outliers

cook = cooks.distance(step_model2)
plot(cook,
     type="h",
     lwd=3,
     ylab = "Cook's Distance",
     main="Cook's Distance")
abline(h = 1)
```

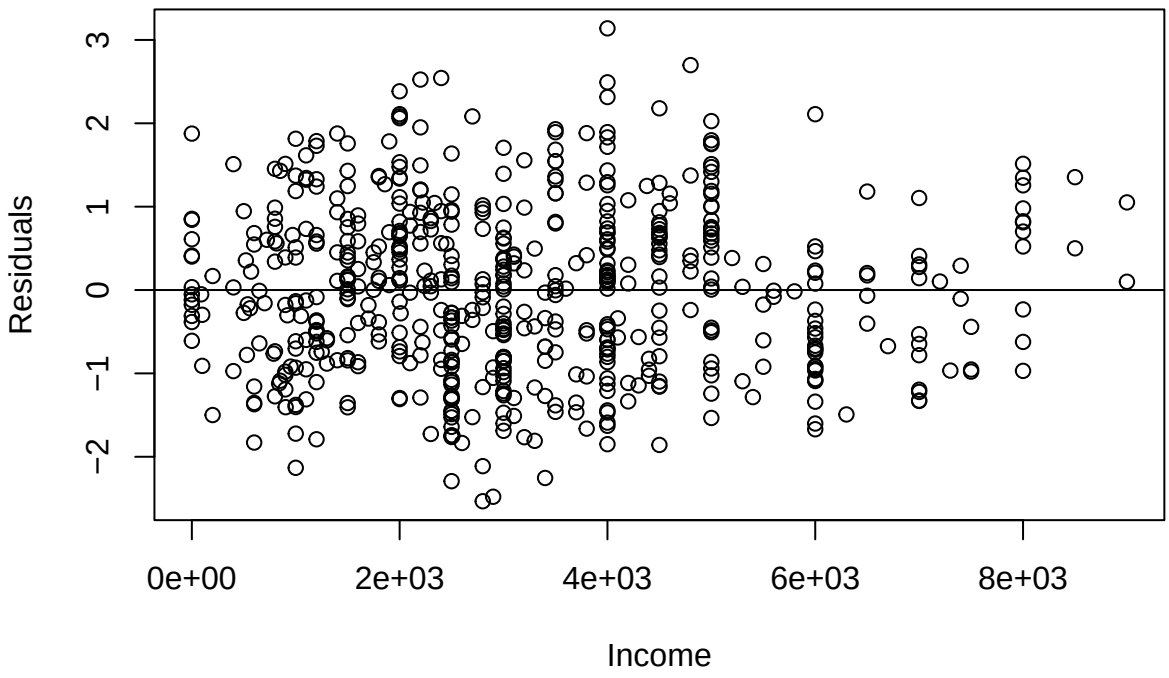



```
res2 = stdres(step_model2) ## (Standardized) Residuals
```

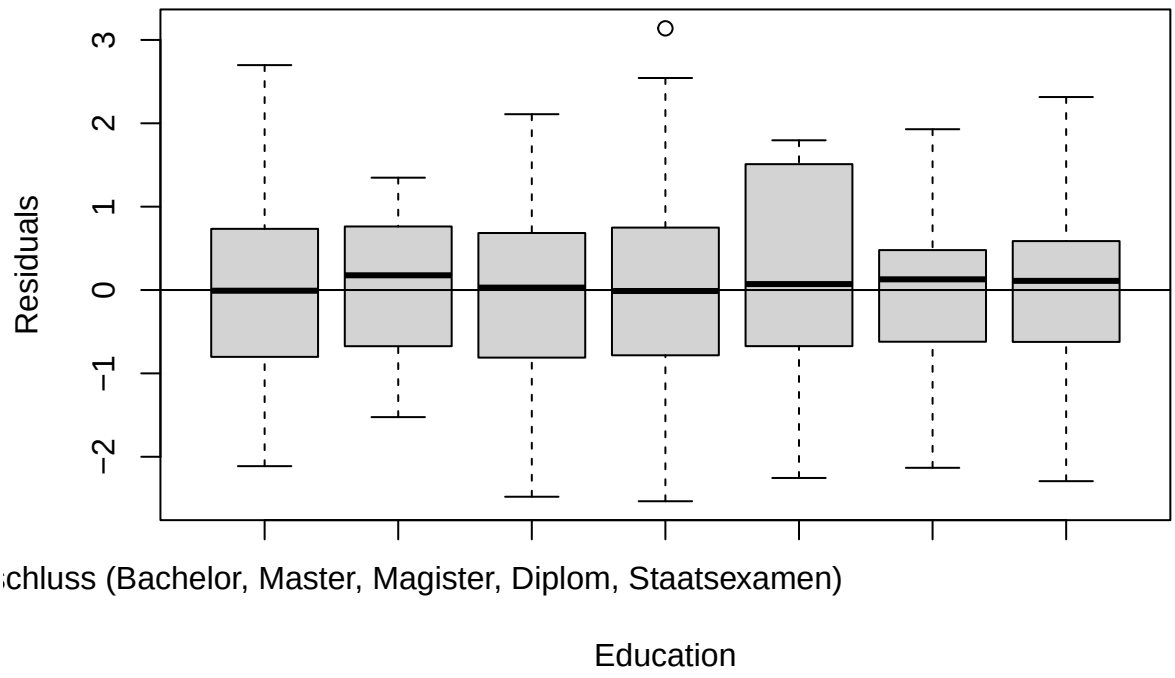
```r
# Linearity assumption/Mean zero assumption

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



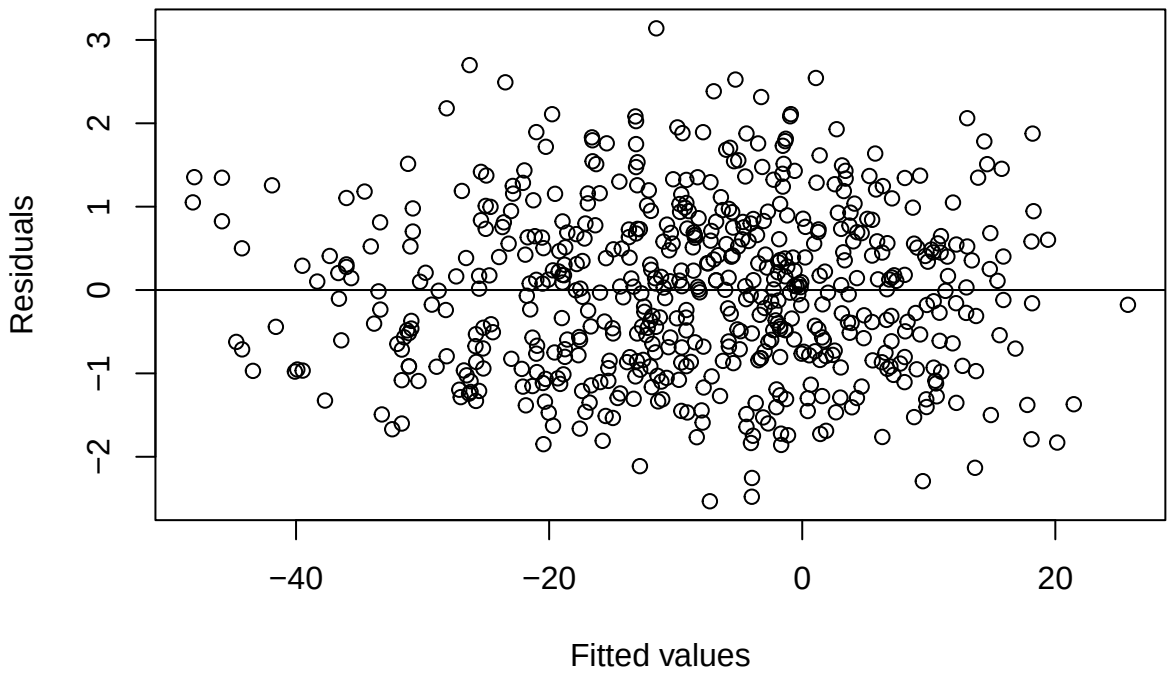

```r
plot(df2$income, res2, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```




```
plot(df2$education, res2, xlab = "Education", 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)
```



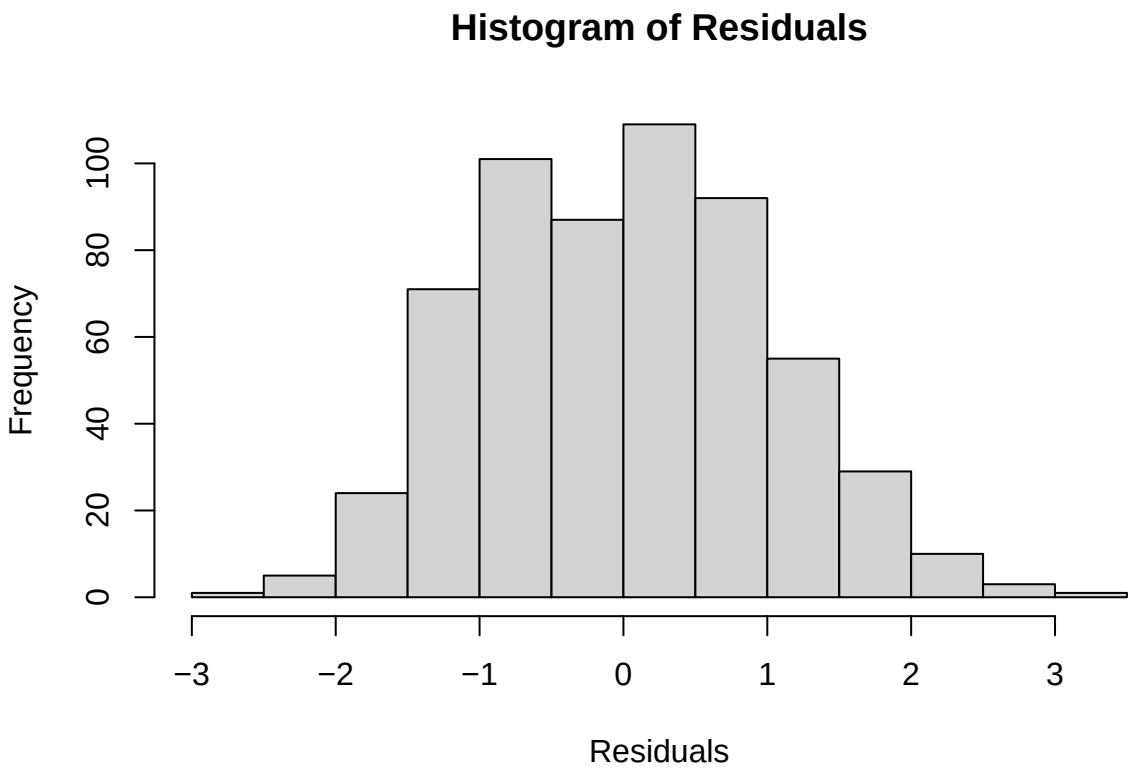

```
# Durbin-Watson Test: Independence of the error terms
# H0 (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(step_model2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1        -4.1e-03             2       1
##  Alternative hypothesis: rho != 0
```

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

bptest(step_model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  step_model2
## BP = 8, df = 8, p-value = 0.4
```

```
# Assumption for the normal distribution

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 = 5e-03
```