Regression_Analysis_Food

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_food


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

df1
```

```
## # A tibble: 588 x 7
##      age income political_party      education           urban~1 feder~2 CO2_f~3
##    <int>  <dbl> <fct>                <fct>               <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU              (Fach-) Hochschula~ zentral Saarla~   1495.
##  2    59    800 Keine Angabe         Allgemeine oder fa~ sehr z~ Hessen    1731.
##  3    60   1750 Keine Angabe         Berufsausbildung, ~ periph~ Bayern    1180.
##  4    73   2500 SPD                  Realschulabschluss~ sehr z~ Bayern    1709.
##  5    43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin    1735.
##  6    49   2300 CDU/CSU              Berufsausbildung, ~ zentral Sachse~   1033.
##  7    57    600 CDU/CSU              Realschulabschluss~ zentral Baden-~   1296.
##  8    39   5000 SPD                  (Fach-) Hochschula~ sehr z~ Berlin    2384.
##  9    62      0 Keine Angabe         (Fach-) Hochschula~ sehr z~ Nordrh~   1790.
## 10    45   2600 Keine Angabe         Berufsausbildung, ~ sehr z~ Hessen    1407.
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: CO2_food
```

```
## 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_food


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

df2
```

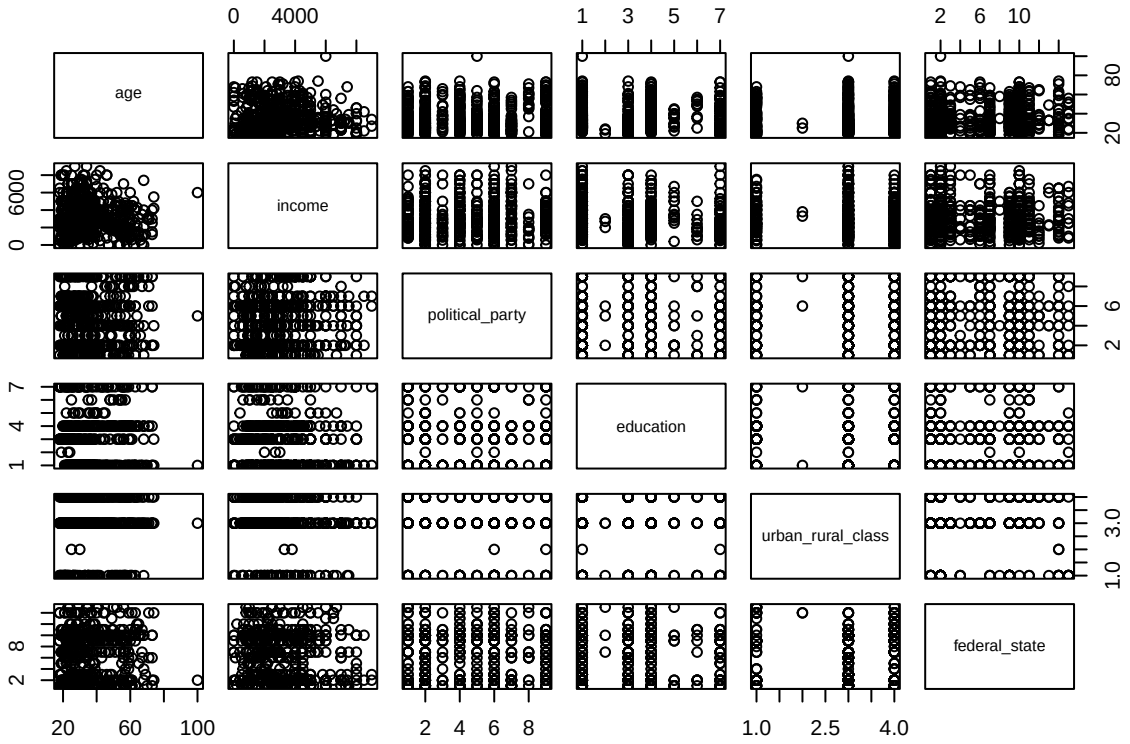
```
## # A tibble: 588 x 7
##      age income political_party      education           urban~1 feder~2 belie~3
##    <int>  <dbl> <fct>                <fct>               <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU              (Fach-) Hochschula~ zentral Saarla~       5
##  2    59    800 Keine Angabe         Allgemeine oder fa~ sehr z~ Hessen      -26
##  3    60   1750 Keine Angabe         Berufsausbildung, ~ periph~ Bayern       49
##  4    73   2500 SPD                  Realschulabschluss~ sehr z~ Bayern       -9
##  5    43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin      -26
##  6    49   2300 CDU/CSU              Berufsausbildung, ~ zentral Sachse~      93
##  7    57    600 CDU/CSU              Realschulabschluss~ zentral Baden-~      60
##  8    39   5000 SPD                  (Fach-) Hochschula~ sehr z~ Berlin      -61
##  9    62      0 Keine Angabe         (Fach-) Hochschula~ sehr z~ Nordrh~      -5
## 10    45   2600 Keine Angabe         Berufsausbildung, ~ sehr z~ Hessen       11
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: belief_diff_food
```

I. Exploratory Data Analysis

Check the Jupytor notebook: EDA_scatter_plot_actual_belief

II. Multivariate Regression: CO2 food

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

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



1. Modeling

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

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

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

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

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

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

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

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

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

```
## defining a reference level

df1$political_party  <- relevel(df1$political_party, ref='Bündnis 90/Die Grünen')
df1$education  <- relevel(df1$education, ref='(Fach-) Hochschulabschluss (Bachelor, Master, Magister, 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_food ~ age + income + political_party + education +  urban_rural_class + federal_state

summary(model1)
```

```
##
## Call:
## lm(formula = CO2_food ~ age + income + political_party + education +
##     urban_rural_class + federal_state, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1228.42  -351.29   -80.26   261.35  3024.52
##
## Coefficients:
##                                                                  Estimate
## (Intercept)                                                   1649.95700
## age                                                             -4.75323
## income                                                           0.01770
## political_partyAfD                                             494.41873
## political_partyBündnis Sarah Wagenknecht                       237.97895
```

```
## political_partyCDU/CSU                                                          401.30696
## political_partyDie Linke                                                         60.97770
## political_partyEiner anderen Partei                                             269.78750
## political_partyFDP                                                              338.27776
## political_partyKeine Angabe                                                      52.06262
## political_partySPD                                                              390.52818
## education(Noch) kein Abschluss                                                 1036.26216
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   80.48354
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule             87.69220
## educationDoktorgrad oder Habilitation                                            10.56860
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  156.98337
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss       102.58551
## urban_rural_classperipher                                                       -182.03001
## urban_rural_classsehr peripher                                                  -112.32325
## urban_rural_classzentral                                                         -57.63702
## federal_stateBaden-Württemberg                                                   -44.78441
## federal_stateBayern                                                              81.71124
## federal_stateBerlin                                                              -23.55651
## federal_stateBrandenburg                                                         79.06846
## federal_stateBremen                                                             -114.86223
## federal_stateHamburg                                                            -152.23293
## federal_stateHessen                                                              -63.76098
## federal_stateMecklenburg-Vorpommern                                             -164.39711
## federal_stateNiedersachsen                                                       226.47128
## federal_stateRheinland-Pfalz                                                      -8.27212
## federal_stateSaarland                                                            -43.16881
## federal_stateSachsen-Anhalt                                                     -512.64010
## federal_stateSchleswig-Holstein                                                  120.69573
## federal_stateThüringen                                                           122.87544
##                                                                                Std. Error
## (Intercept)                                                                      102.10932
## age                                                                                1.80773
## income                                                                             0.01208
## political_partyAfD                                                               86.49161
## political_partyBündnis Sarah Wagenknecht                                        121.22905
## political_partyCDU/CSU                                                           77.68269
## political_partyDie Linke                                                         93.63026
## political_partyEiner anderen Partei                                              69.49607
## political_partyFDP                                                               89.96842
## political_partyKeine Angabe                                                     155.83677
## political_partySPD                                                               79.10575
## education(Noch) kein Abschluss                                                  316.48391
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   61.51900
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule             62.75856
## educationDoktorgrad oder Habilitation                                           153.41144
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  175.11760
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss       79.81045
## urban_rural_classperipher                                                        80.42637
## urban_rural_classsehr peripher                                                  398.15281
## urban_rural_classzentral                                                         58.96386
## federal_stateBaden-Württemberg                                                   76.59576
## federal_stateBayern                                                              78.64226
## federal_stateBerlin                                                              95.53706
## federal_stateBrandenburg                                                        200.66496
```

```
## federal_stateBremen                                                                  146.36473
## federal_stateHamburg                                                                 119.11131
## federal_stateHessen                                                                   91.38668
## federal_stateMecklenburg-Vorpommern                                                  383.60044
## federal_stateNiedersachsen                                                            92.97044
## federal_stateRheinland-Pfalz                                                         114.66837
## federal_stateSaarland                                                                180.05848
## federal_stateSachsen-Anhalt                                                          277.55695
## federal_stateSchleswig-Holstein                                                      134.12031
## federal_stateThüringen                                                               203.29023
##                                                                                     t value
## (Intercept)                                                                          16.159
## age                                                                                  -2.629
## income                                                                                1.465
## political_partyAfD                                                                    5.716
## political_partyBündnis Sarah Wagenknecht                                              1.963
## political_partyCDU/CSU                                                                5.166
## political_partyDie Linke                                                              0.651
## political_partyEiner anderen Partei                                                   3.882
## political_partyFDP                                                                    3.760
## political_partyKeine Angabe                                                           0.334
## political_partySPD                                                                    4.937
## education(Noch) kein Abschluss                                                        3.274
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)     1.308
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                  1.397
## educationDoktorgrad oder Habilitation                                                 0.069
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss      0.896
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss            1.285
## urban_rural_classperipher                                                            -2.263
## urban_rural_classsehr peripher                                                       -0.282
## urban_rural_classzentral                                                             -0.977
## federal_stateBaden-Württemberg                                                       -0.585
## federal_stateBayern                                                                   1.039
## federal_stateBerlin                                                                  -0.247
## federal_stateBrandenburg                                                              0.394
## federal_stateBremen                                                                  -0.785
## federal_stateHamburg                                                                 -1.278
## federal_stateHessen                                                                  -0.698
## federal_stateMecklenburg-Vorpommern                                                  -0.429
## federal_stateNiedersachsen                                                            2.436
## federal_stateRheinland-Pfalz                                                         -0.072
## federal_stateSaarland                                                                -0.240
## federal_stateSachsen-Anhalt                                                          -1.847
## federal_stateSchleswig-Holstein                                                       0.900
## federal_stateThüringen                                                                0.604
##                                                                                    Pr(>|t|)
## (Intercept)                                                                         < 2e-16
## age                                                                                0.008791
## income                                                                             0.143391
## political_partyAfD                                                                 1.78e-08
## political_partyBündnis Sarah Wagenknecht                                           0.050140
## political_partyCDU/CSU                                                             3.34e-07
## political_partyDie Linke                                                           0.515148
## political_partyEiner anderen Partei                                                0.000116
```

```
## political_partyFDP                                                                0.000188
## political_partyKeine Angabe                                                       0.738442
## political_partySPD                                                                1.05e-06
## education(Noch) kein Abschluss                                                    0.001125
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) 0.191324
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule              0.162884
## educationDoktorgrad oder Habilitation                                             0.945102
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  0.370404
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss        0.199202
## urban_rural_classperipher                                                         0.024003
## urban_rural_classsehr peripher                                                    0.777964
## urban_rural_classzentral                                                          0.328750
## federal_stateBaden-Württemberg                                                    0.558997
## federal_stateBayern                                                               0.299247
## federal_stateBerlin                                                               0.805333
## federal_stateBrandenburg                                                          0.693709
## federal_stateBremen                                                               0.432926
## federal_stateHamburg                                                              0.201759
## federal_stateHessen                                                               0.485654
## federal_stateMecklenburg-Vorpommern                                               0.668408
## federal_stateNiedersachsen                                                        0.015167
## federal_stateRheinland-Pfalz                                                      0.942517
## federal_stateSaarland                                                             0.810614
## federal_stateSachsen-Anhalt                                                       0.065284
## federal_stateSchleswig-Holstein                                                   0.368561
## federal_stateThüringen                                                            0.545803
##
## (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: 528.4 on 554 degrees of freedom
## Multiple R-squared:  0.1539, Adjusted R-squared:  0.1035
## F-statistic: 3.053 on 33 and 554 DF,  p-value: 6.253e-08
```

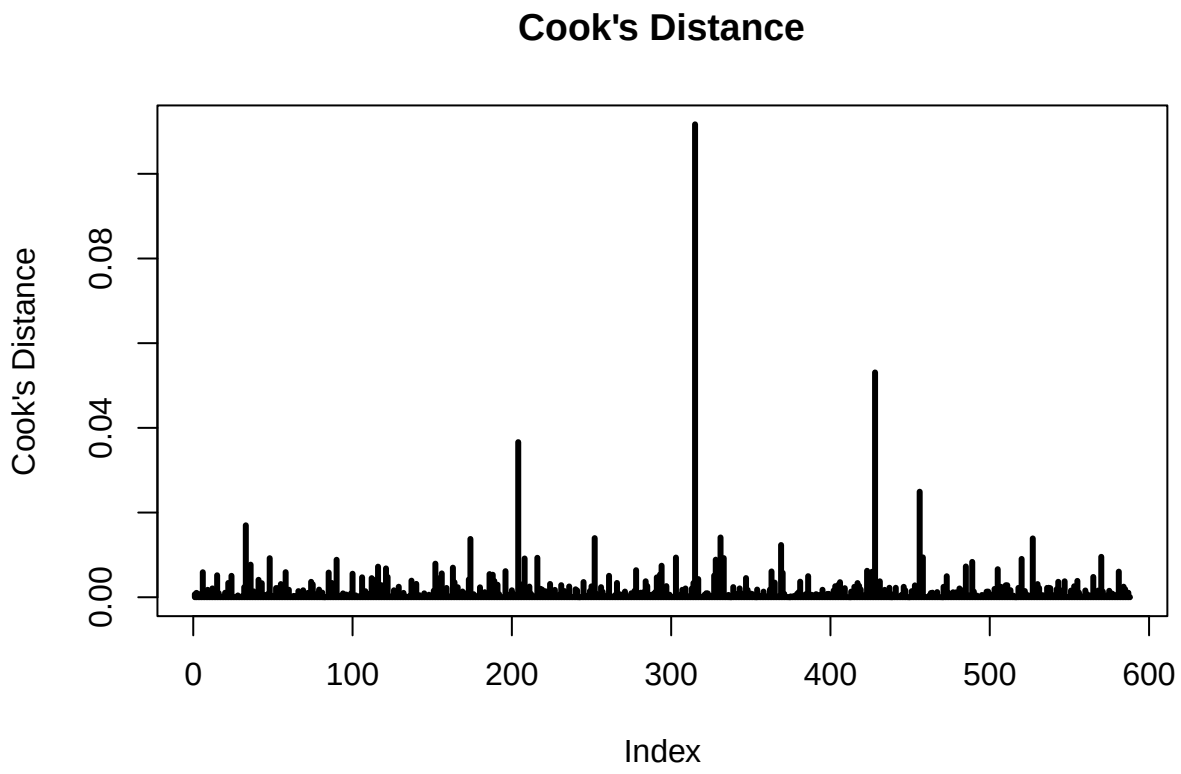
```
# Checking the VIFs for multicollinearity

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
```

```
# threshold for multicollinearity
# Calculating the threshold

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)
```



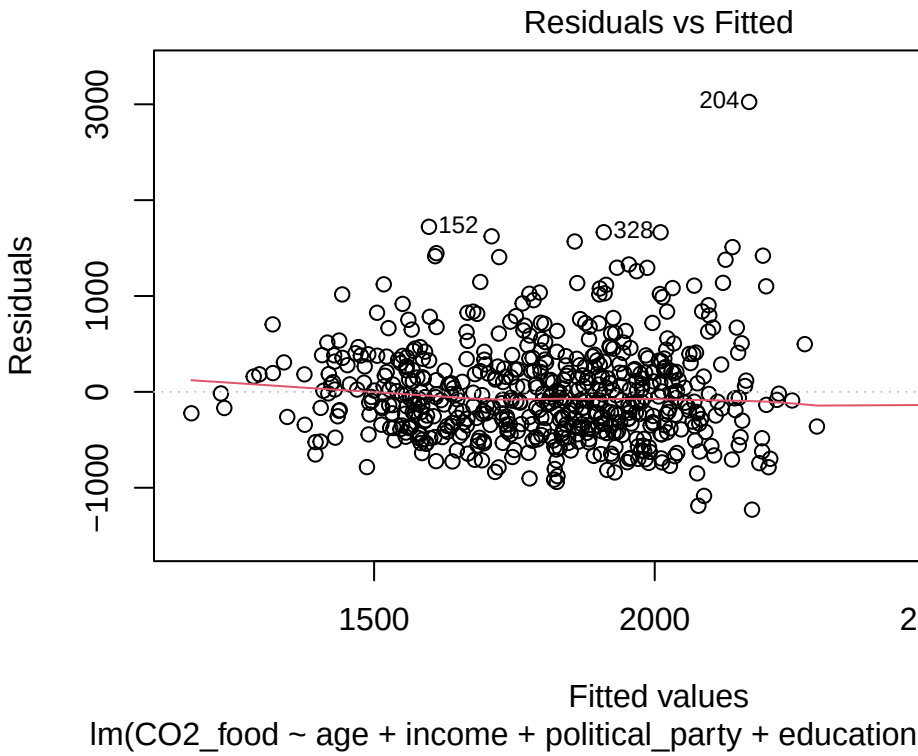

```
influential = cooks.distance(model1)[which(cook > 3*mean(cook, na.rm=TRUE))]
influential
```

```
##           6          33          36          48          58          85
## 0.005905500 0.016987374 0.007701172 0.009222091 0.005922695 0.005829832
##          90         100         116         121         152         156
## 0.008849586 0.005559505 0.007228999 0.006809638 0.007934072 0.005632054
##         163         174         186         196         204         208
## 0.006989085 0.013764261 0.005469170 0.006165627 0.036633543 0.009152248
##         216         252         278         294         303         315
## 0.009319069 0.013966868 0.006389329 0.007449710 0.009375852 0.111679185
##         328         330         331         333         363         369
## 0.008878170 0.005651801 0.014119699 0.009246118 0.006092358 0.012333071
##         370         423         426         428         456         458
## 0.005769210 0.006248660 0.006002794 0.053089194 0.024925336 0.009419272
##         485         489         505         520         527         570
## 0.007270443 0.008350133 0.006622905 0.009055882 0.013890920 0.009547564
##         581
## 0.006051068
```

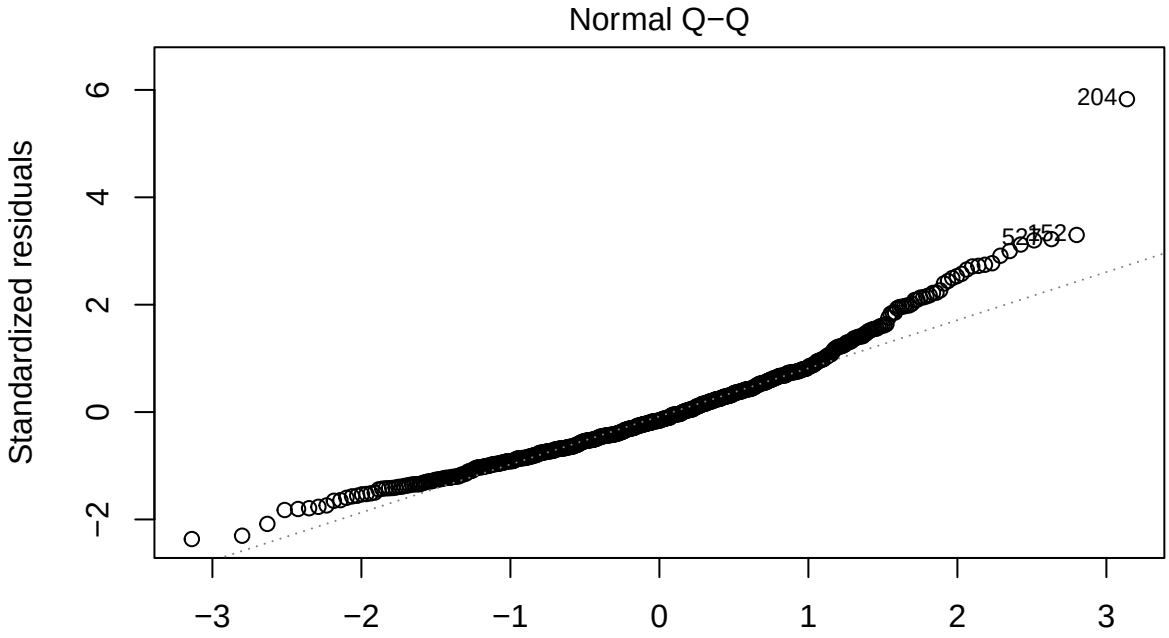
```
influential = influential[!is.na(influential)]
influential_vector = c(as.numeric(rownames(data.frame(influential))))


df1[influential_vector, ]
```

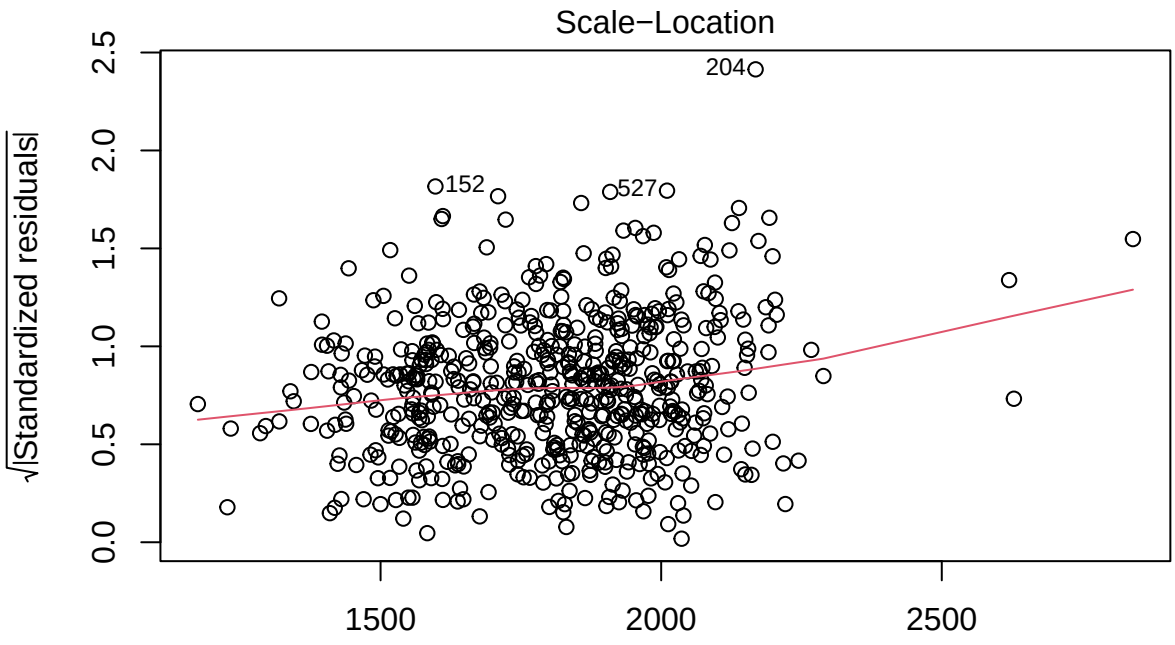
```
## # A tibble: 43 x 7
##      age income political_party education                urban~1 feder~2 CO2_f~3
##    <int>  <dbl> <fct>           <fct>                    <fct>   <fct>     <dbl>
##  1    49   2300 CDU/CSU         Berufsausbildung, Lehre~ zentral Sachse~   1033.
##  2    37   3500 Keine Angabe    Hauptschulabschluss (Vo~ sehr z~ Bayern     948.
##  3    19   1000 FDP             Allgemeine oder fachgeb~ sehr z~ Bayern     890.
##  4    19   5000 CDU/CSU         Allgemeine oder fachgeb~ sehr z~ Rheinl~   3259.
##  5    53   1500 AfD             Hauptschulabschluss (Vo~ periph~ Bayern    1363.
##  6    57   1000 AfD             Hauptschulabschluss (Vo~ periph~ Baden-~   1248.
##  7    19   3000 FDP             Allgemeine oder fachgeb~ sehr z~ Hessen    3226.
##  8    50   2500 AfD             Berufsausbildung, Lehre~ periph~ Thürin~   1418.
##  9    23   2000 CDU/CSU         Realschulabschluss (Mit~ zentral Rheinl~   3003.
## 10    58      0 CDU/CSU         Realschulabschluss (Mit~ zentral Bayern    2981.
## # ... with 33 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: CO2_food
```

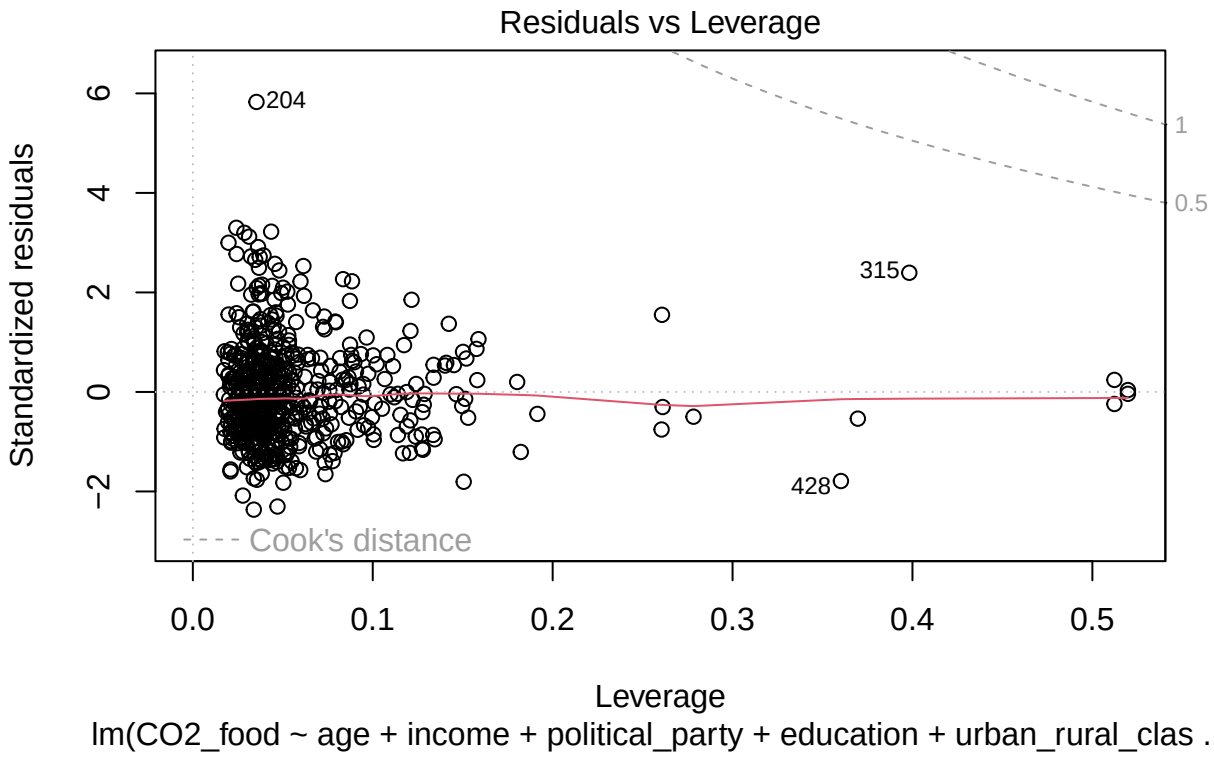
```
plot(model1)
```



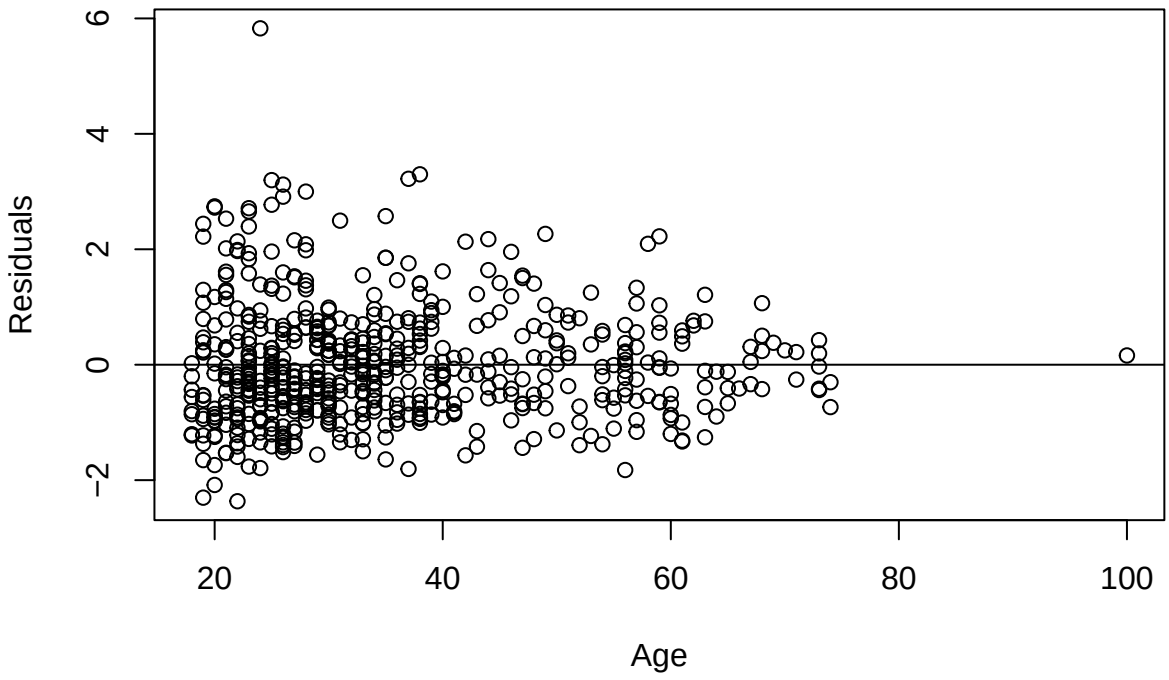
2. Assumptions check in the residuals



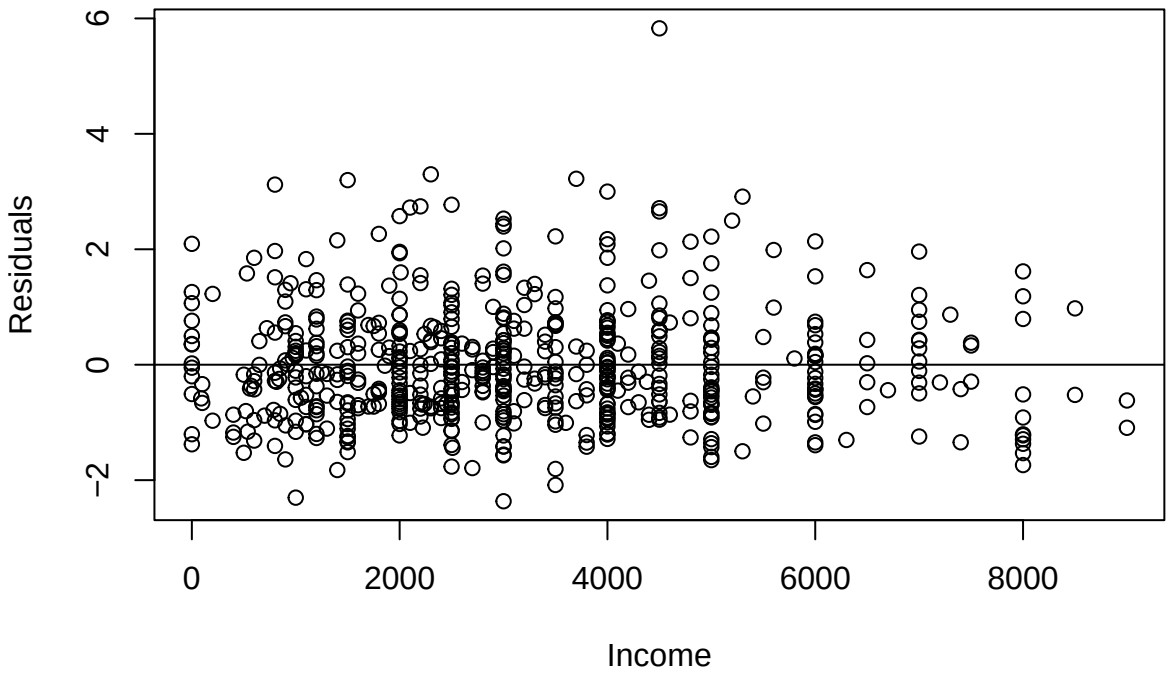

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
204
152
527
1500
2000
2500
Fitted values
lm(CO2_food ~ age + income + political_party + education + urban_rural_clas ...



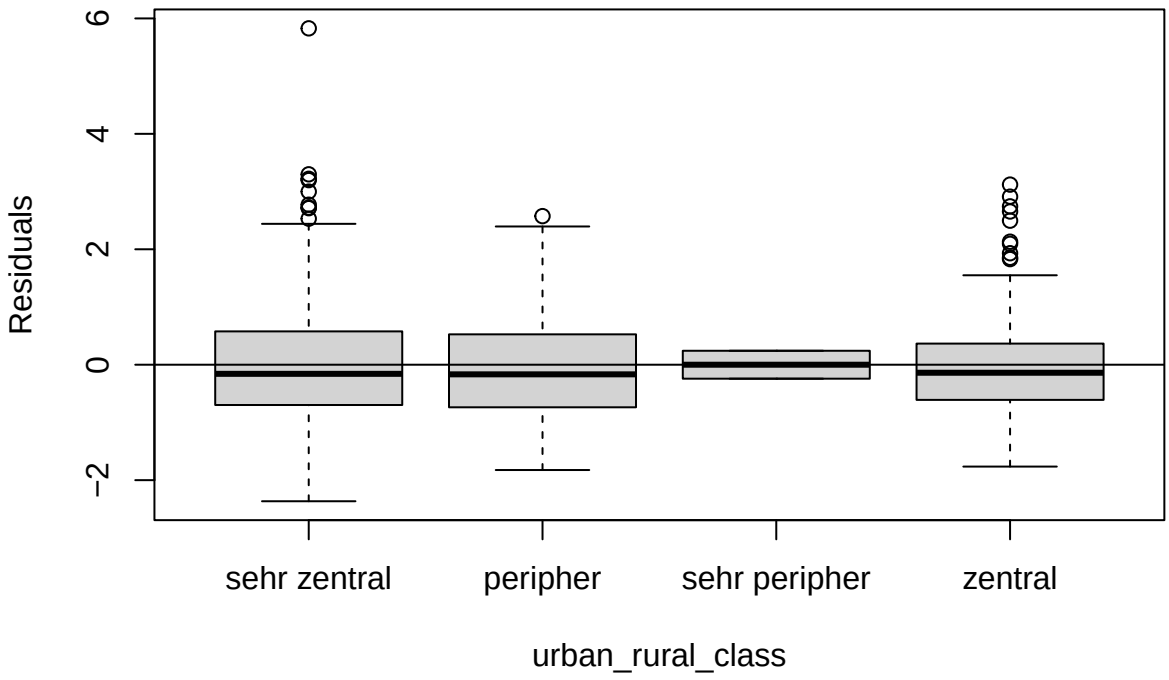

```
res1 = stdres(model1) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

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



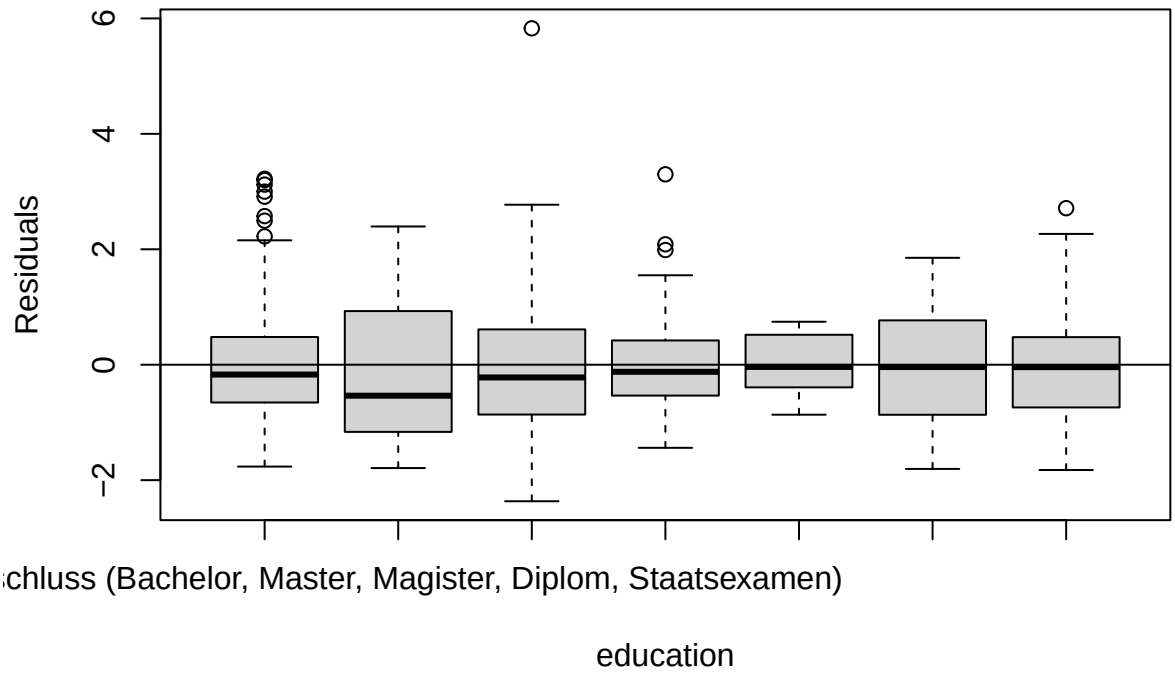

```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



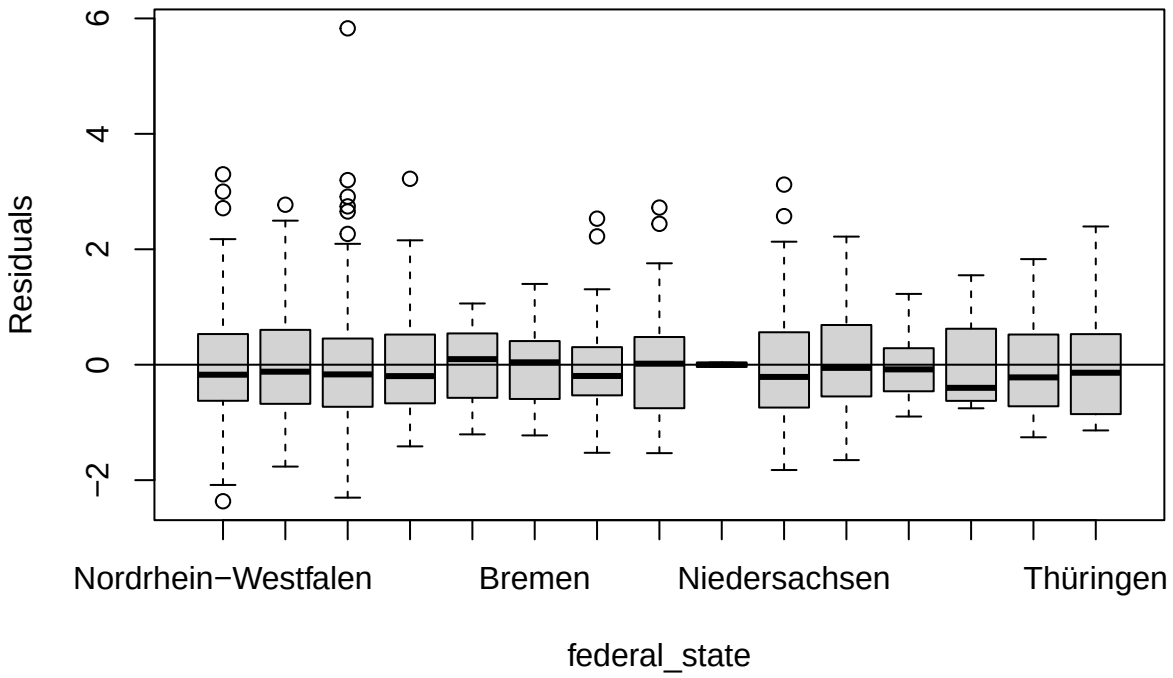

```
plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
abline(h = 0)
```



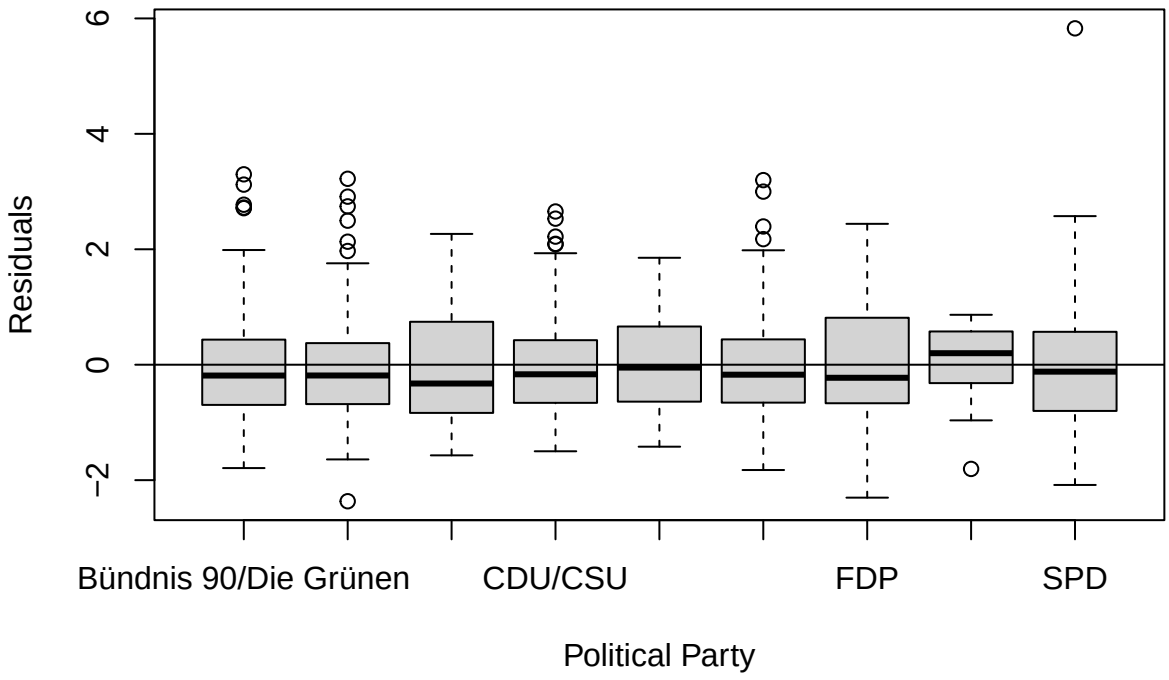

```
plot(df1$education, res1, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



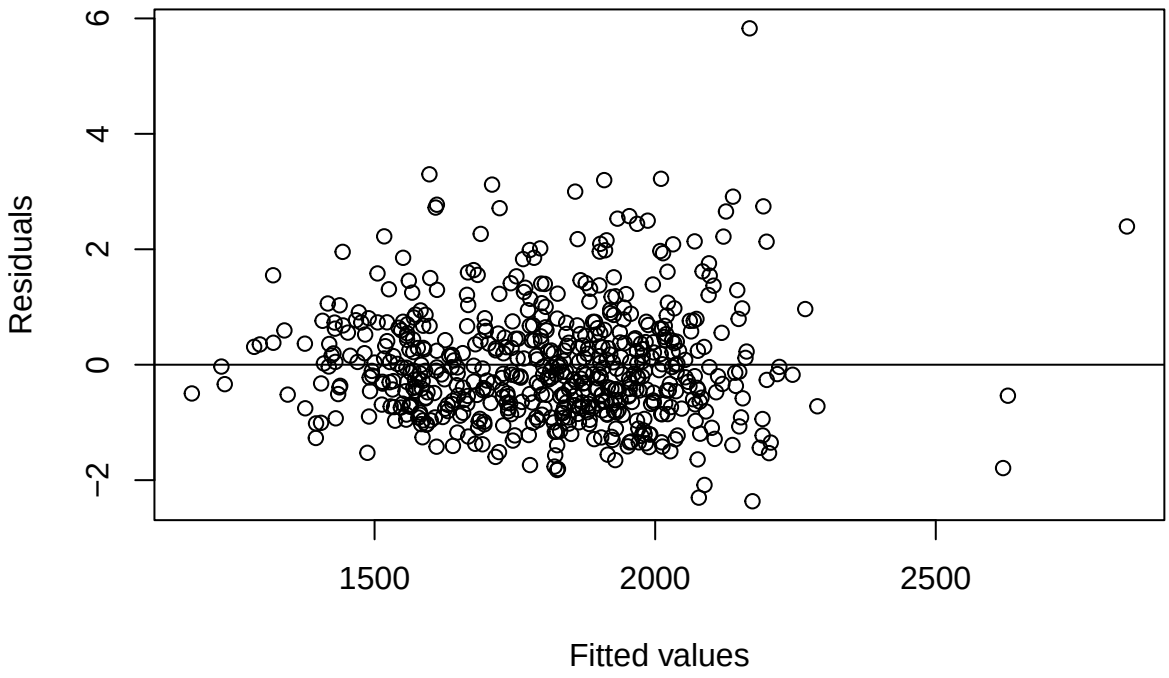

```
plot(df1$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```



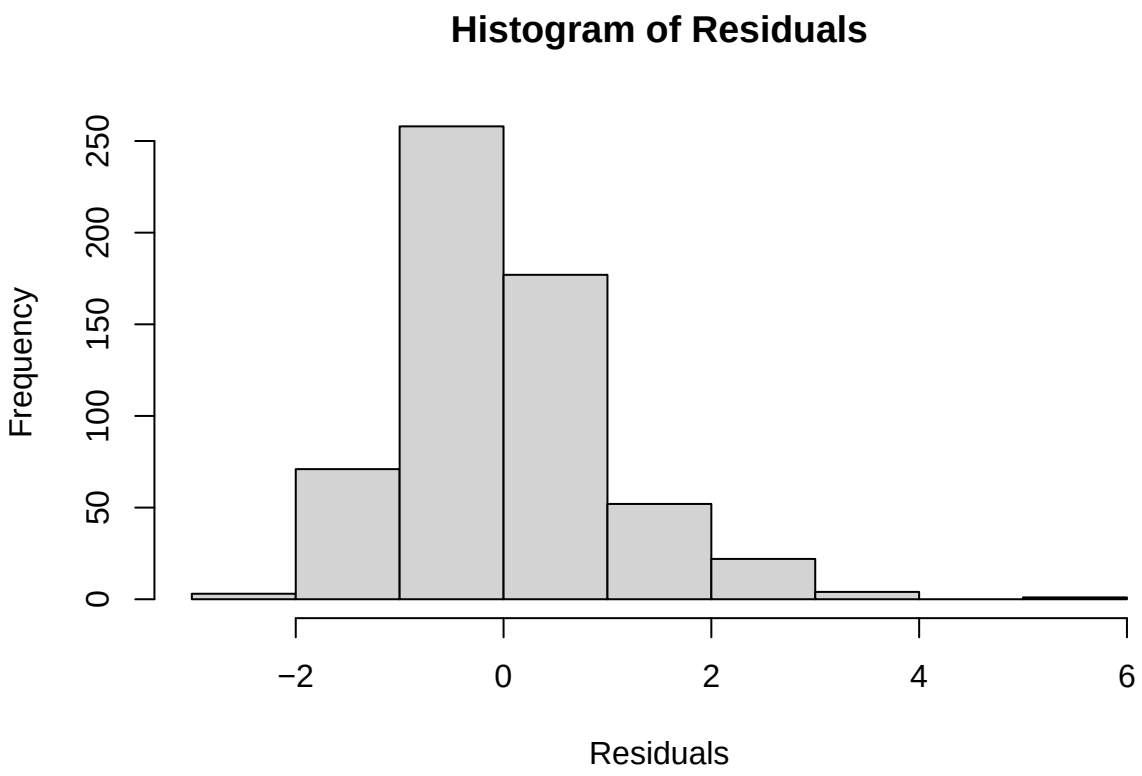

```
plot(df1$political_party, res1, xlab = "Political Party", ylab = "Residuals")
abline(h = 0)
```




```
# Constant variance and independent error term assumption

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




```
# Normality assumption

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




```
## 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 = 0.94981, p-value = 3.066e-13
```

```
### 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=7406.36
## CO2_food ~ age + income + political_party + education + urban_rural_class +
##     federal_state
##
```

```
##                      Df Sum of Sq       RSS    AIC
## - federal_state      14   5461417 160151258 7398.8
## - urban_rural_class   3   1438123 156127963 7405.8
## <none>                             154689841 7406.4
## - income              1    599569 155289409 7406.6
## - education           6   3798811 158488651 7408.6
## - age                 1   1930466 156620307 7411.7
## - political_party     8  16816209 171506050 7451.0
##
## Step:  AIC=7398.76
## CO2_food ~ age + income + political_party + education + urban_rural_class
##
##                      Df Sum of Sq       RSS    AIC
## - urban_rural_class   3    361311 160512568 7394.1
## <none>                             160151258 7398.8
## - income              1    627914 160779172 7399.1
## - education           6   3576894 163728151 7399.8
## - age                 1   1757161 161908418 7403.2
## - political_party     8  15973782 176125039 7438.7
##
## Step:  AIC=7394.09
## CO2_food ~ age + income + political_party + education
##
##                    Df Sum of Sq       RSS    AIC
## <none>                           160512568 7394.1
## - income            1    641618 161154186 7394.4
## - education         6   3420504 163933072 7394.5
## - age               1   1828133 162340701 7398.7
## - political_party   8  15721609 176234177 7433.0
```

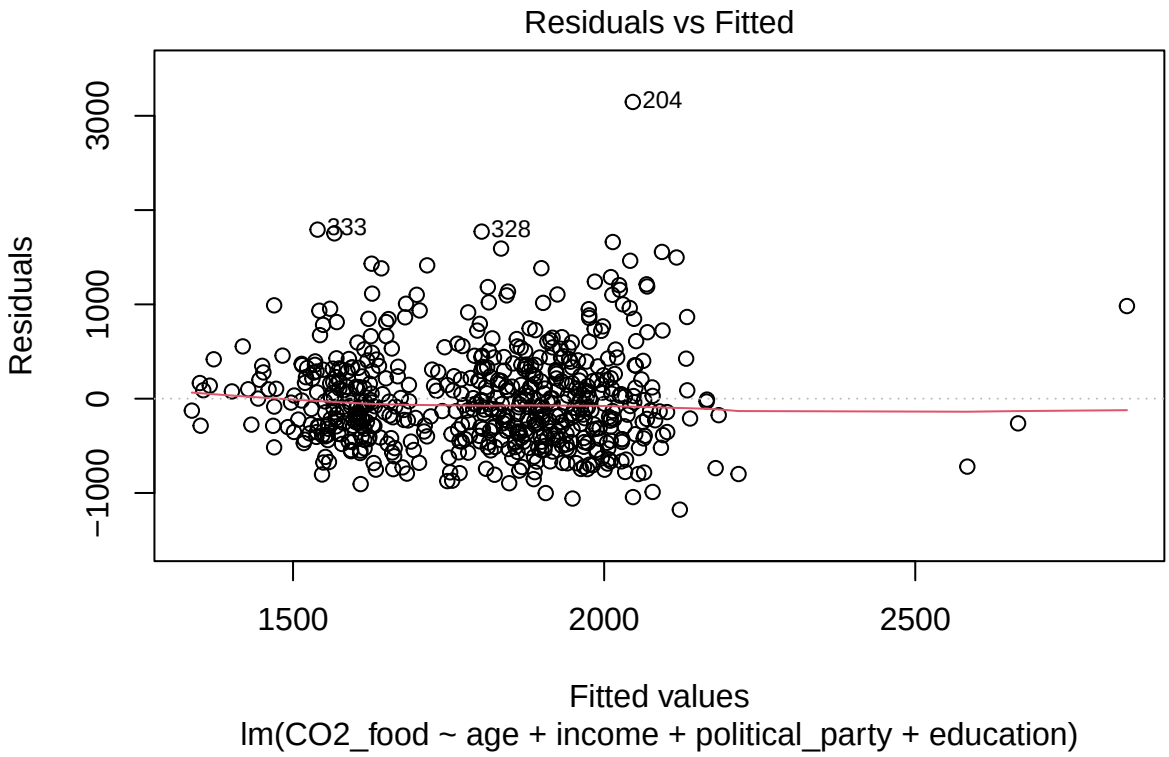
```
summary(step_model1)
```

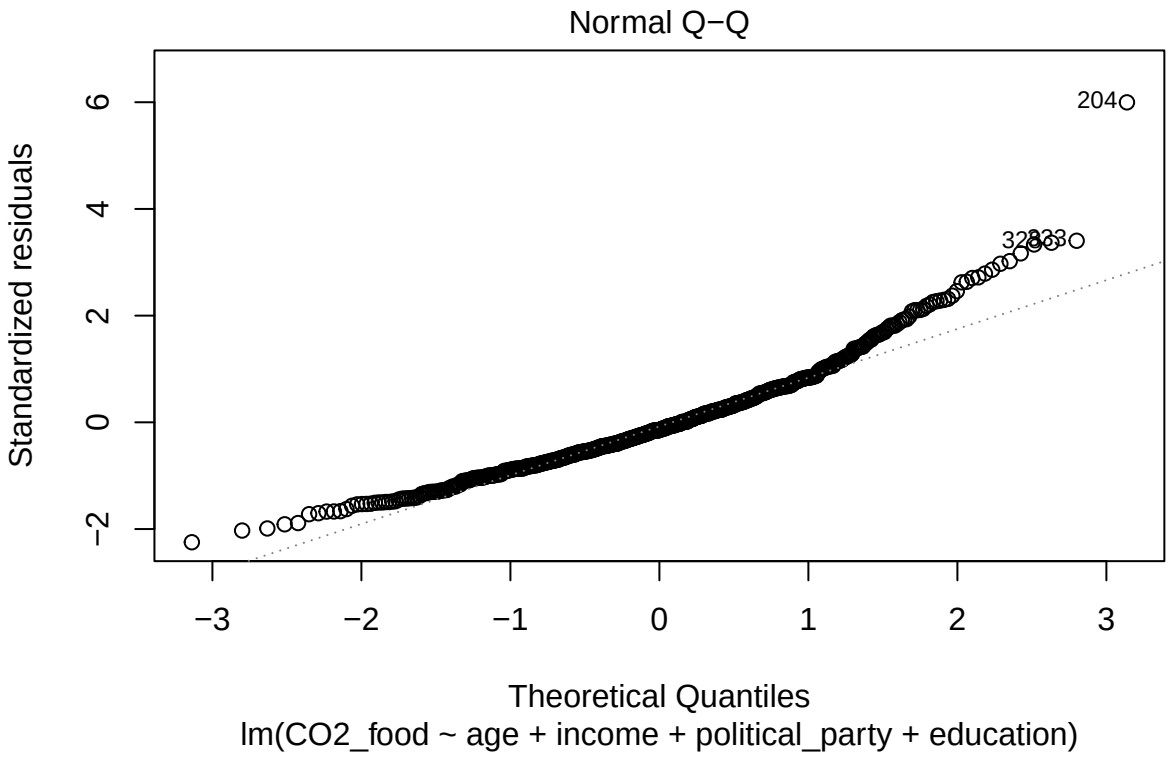
```
##
## Call:
## lm(formula = CO2_food ~ age + income + political_party + education,
##     data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1176.66  -364.06   -76.22   279.85  3146.72
##
## Coefficients:
##                                                                    Estimate
## (Intercept)                                                     1643.75236
## age                                                               -4.56653
## income                                                             0.01796
## political_partyAfD                                               472.66377
## political_partyBündnis Sarah Wagenknecht                         266.31954
## political_partyCDU/CSU                                           370.52532
## political_partyDie Linke                                          71.31061
## political_partyEiner anderen Partei                              246.60463
## political_partyFDP                                               322.43817
## political_partyKeine Angabe                                       11.89768
## political_partySPD                                               379.28628
```

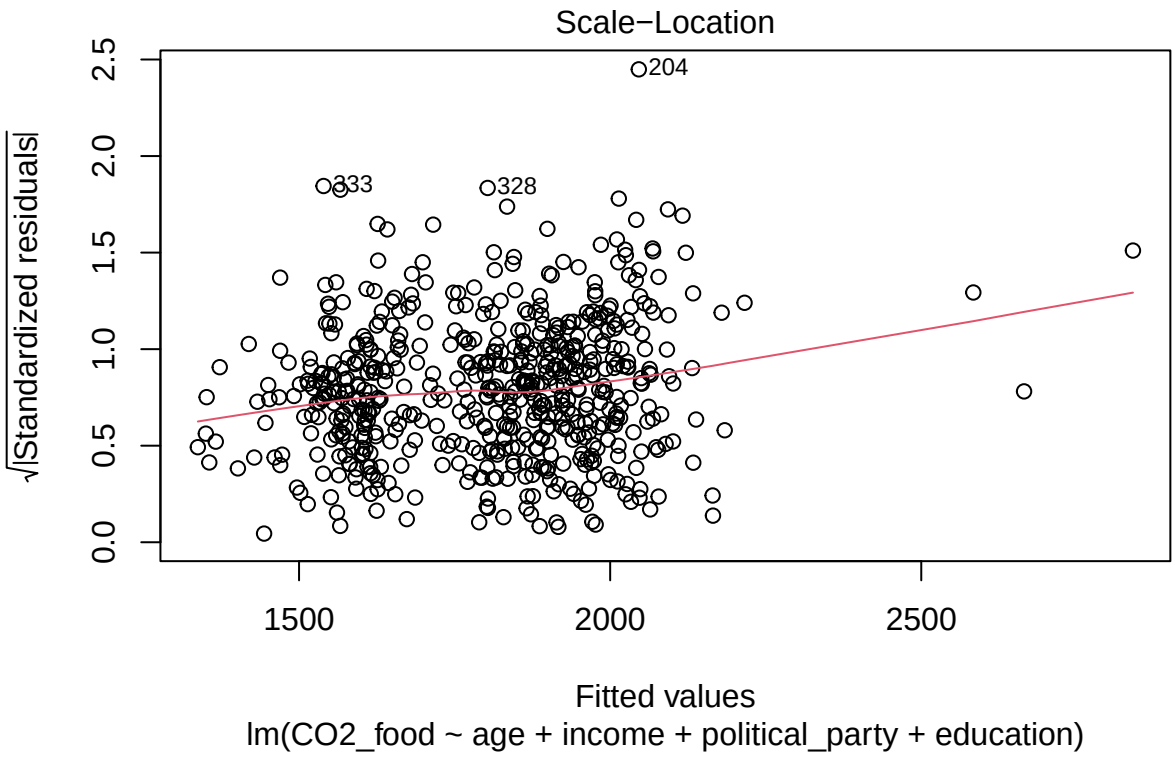
```
## education(Noch) kein Abschluss                                                1001.01093
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   51.79976
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule               54.90458
## educationDoktorgrad oder Habilitation                                               8.00367
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  126.51540
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         95.99489
##                                                                               Std. Error
## (Intercept)                                                                     89.01839
## age                                                                              1.79068
## income                                                                           0.01189
## political_partyAfD                                                              84.86621
## political_partyBündnis Sarah Wagenknecht                                       120.42916
## political_partyCDU/CSU                                                          76.92301
## political_partyDie Linke                                                        92.15183
## political_partyEiner anderen Partei                                             68.07532
## political_partyFDP                                                              89.55848
## political_partyKeine Angabe                                                    152.28022
## political_partySPD                                                              78.14095
## education(Noch) kein Abschluss                                                 309.93720
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  59.83958
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule            61.48889
## educationDoktorgrad oder Habilitation                                          151.68538
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  170.72390
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss      78.53385
##                                                                               t value
## (Intercept)                                                                    18.465
## age                                                                            -2.550
## income                                                                          1.511
## political_partyAfD                                                              5.570
## political_partyBündnis Sarah Wagenknecht                                        2.211
## political_partyCDU/CSU                                                          4.817
## political_partyDie Linke                                                        0.774
## political_partyEiner anderen Partei                                             3.623
## political_partyFDP                                                              3.600
## political_partyKeine Angabe                                                     0.078
## political_partySPD                                                              4.854
## education(Noch) kein Abschluss                                                  3.230
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  0.866
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule            0.893
## educationDoktorgrad oder Habilitation                                           0.053
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  0.741
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss      1.222
##                                                                               Pr(>|t|)
## (Intercept)                                                                    < 2e-16
## age                                                                           0.011027
## income                                                                        0.131397
## political_partyAfD                                                            3.94e-08
## political_partyBündnis Sarah Wagenknecht                                      0.027402
## political_partyCDU/CSU                                                        1.87e-06
## political_partyDie Linke                                                      0.439347
## political_partyEiner anderen Partei                                           0.000318
## political_partyFDP                                                            0.000345
## political_partyKeine Angabe                                                   0.937752
## political_partySPD                                                            1.56e-06
```

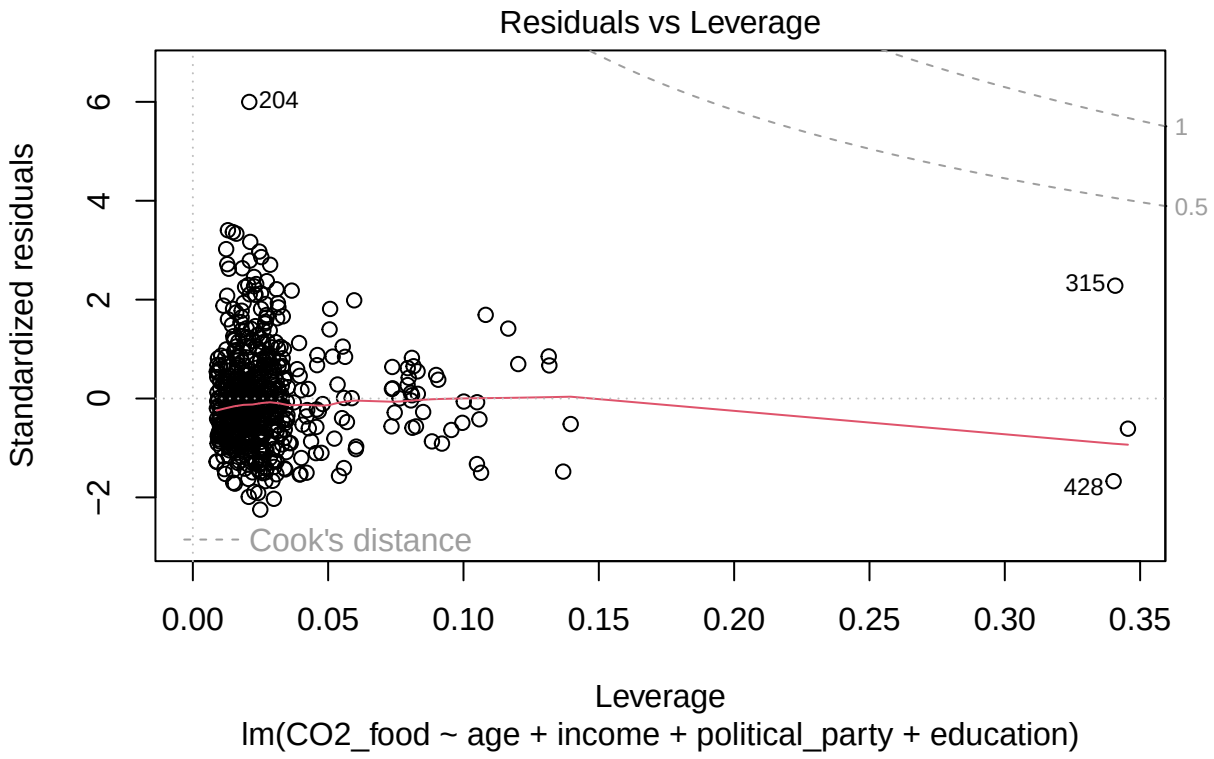
```
## education(Noch) kein Abschluss                                                    0.001310
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) 0.387049
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule              0.372277
## educationDoktorgrad oder Habilitation                                             0.957938
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  0.458966
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss        0.222084
##
## (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
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 530.2 on 571 degrees of freedom
## Multiple R-squared:  0.122,  Adjusted R-squared:  0.0974
## F-statistic: 4.959 on 16 and 571 DF,  p-value: 1.207e-09
```

```
plot(step_model1)
```

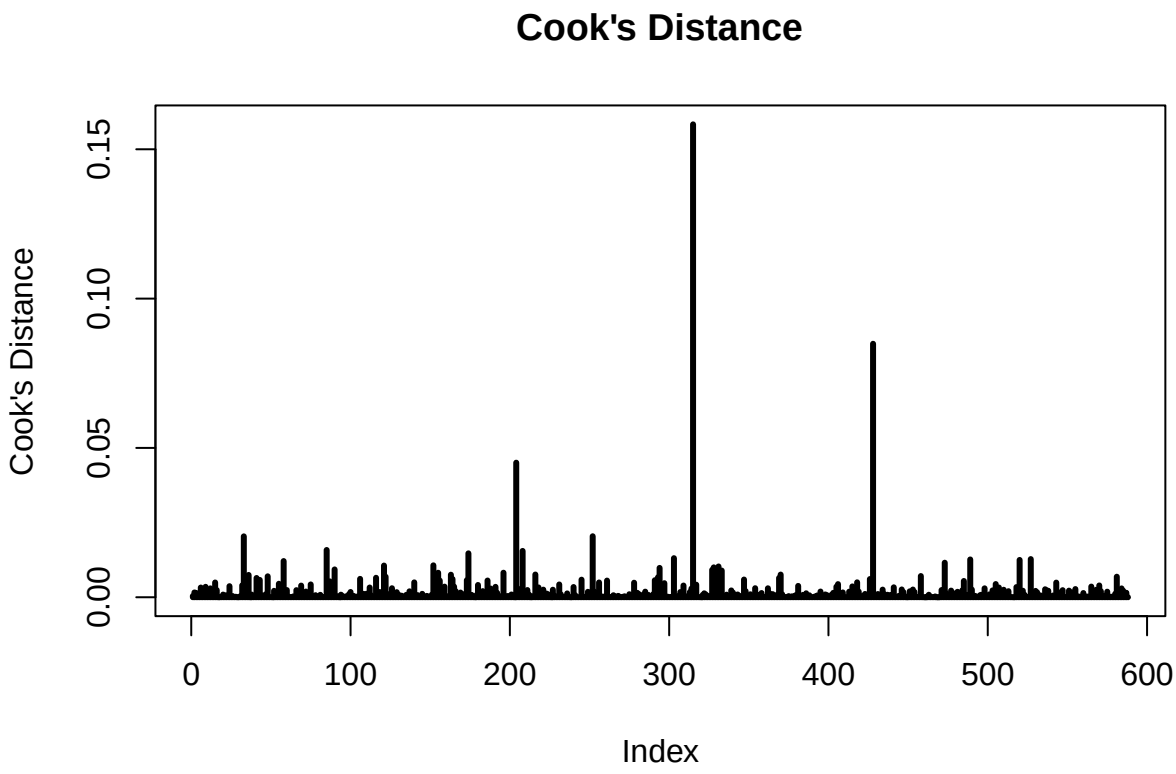

Residuals vs Fitted
Residuals
3000
1000
0
−1000
204
333
328
1500
2000
2500
Fitted values
lm(CO2_food ~ age + income + political_party + education)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(CO2_food ~ age + income + political_party + education)
204
−3
−2
−1
0
1
2
3
−2
0
2
4
6


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
204
333
328
1500
2000
2500
Fitted values
lm(CO2_food ~ age + income + political_party + 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)
```



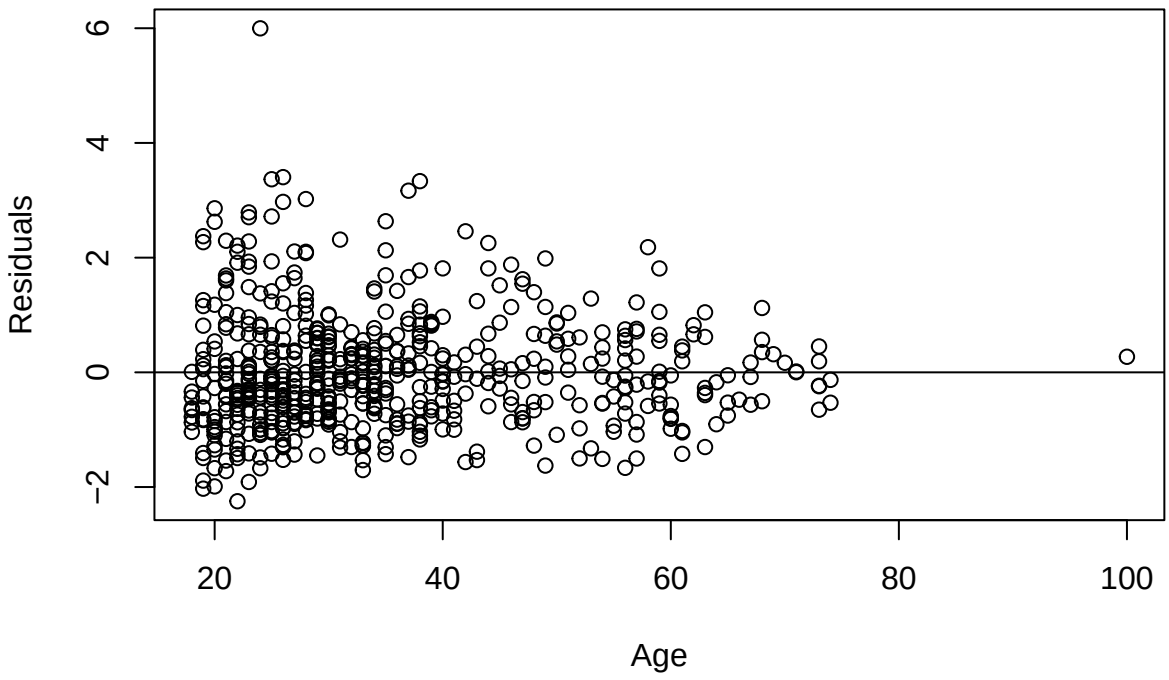

```
influential = cooks.distance(step_model1)[which(cook >1)]
influential
```

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

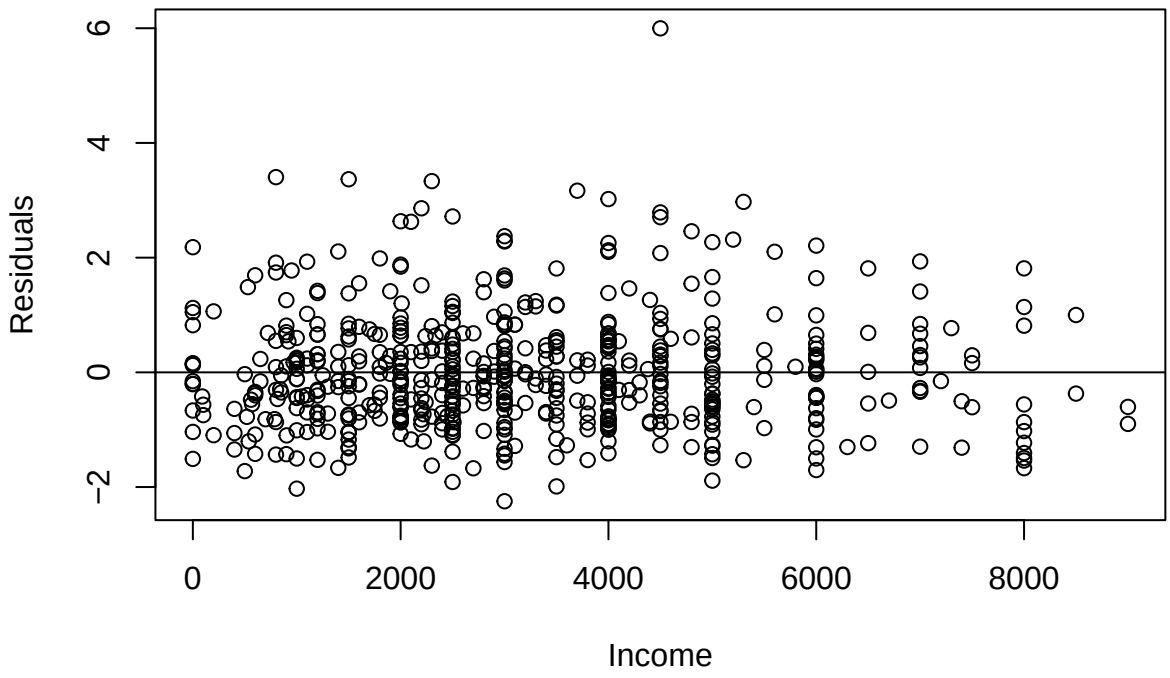
```
influential = influential[!is.na(influential)]
influential_vector = c(as.numeric(rownames(data.frame(influential))))


df1[influential_vector, ]
```

```
## # A tibble: 0 x 7
## # ... with 7 variables: age <int>, income <dbl>, political_party <fct>,
## #   education <fct>, urban_rural_class <fct>, federal_state <fct>,
## #   CO2_food <dbl>
```

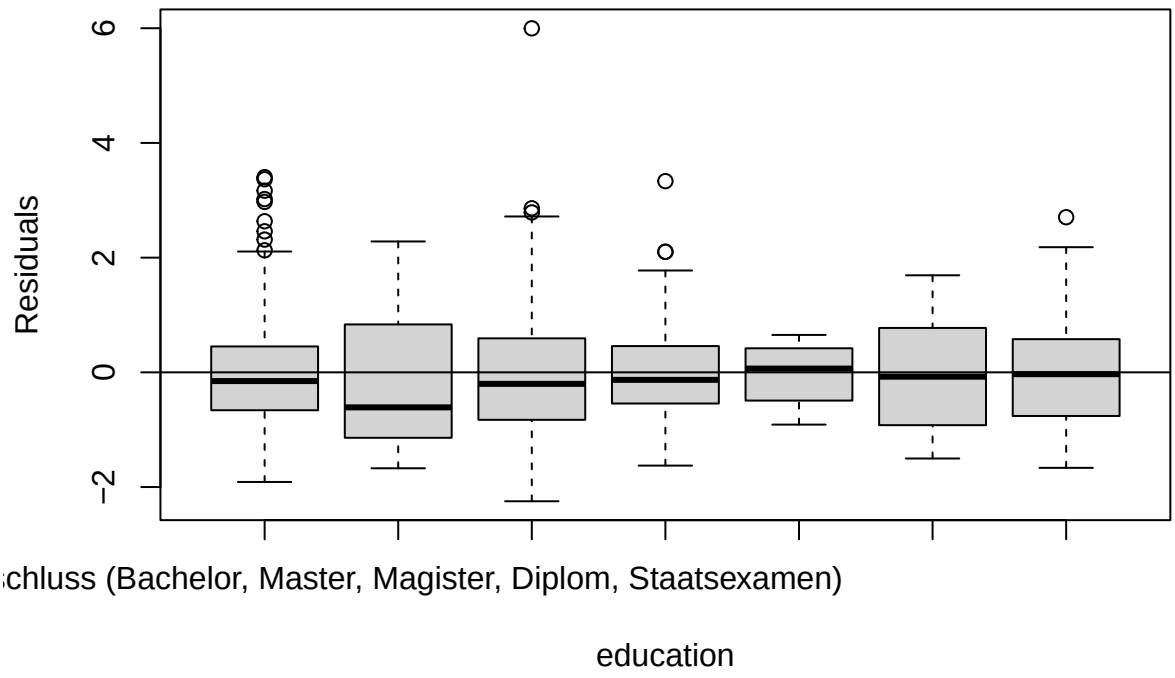
```
res1 = stdres(step_model1) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

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



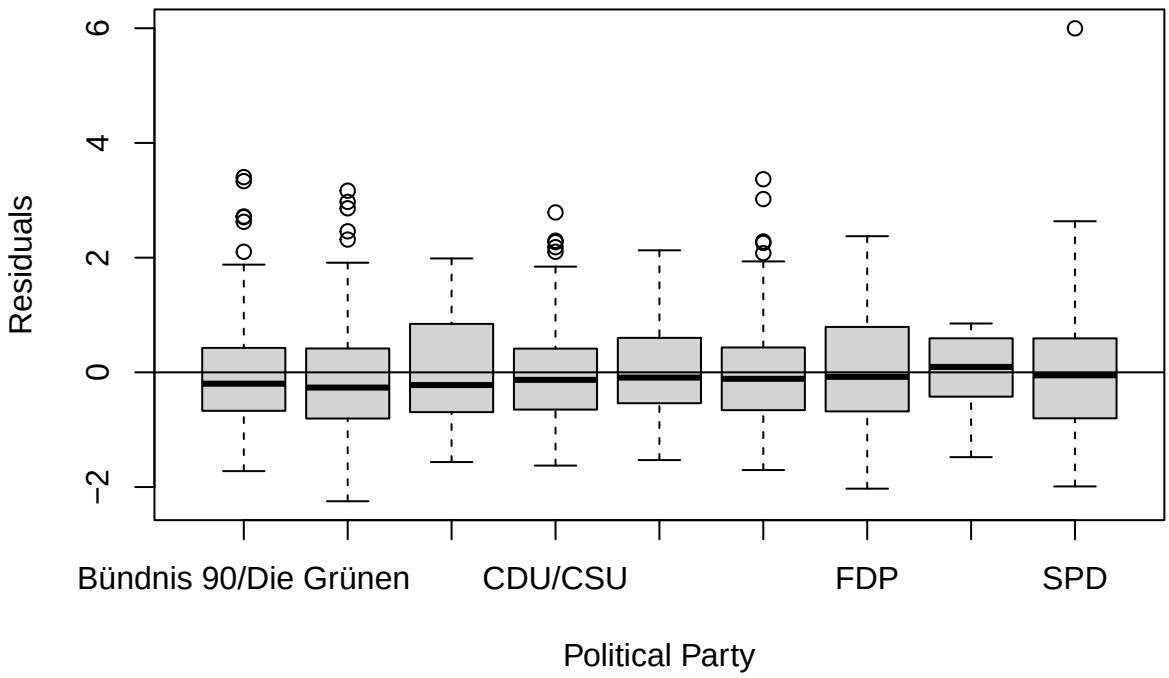

```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



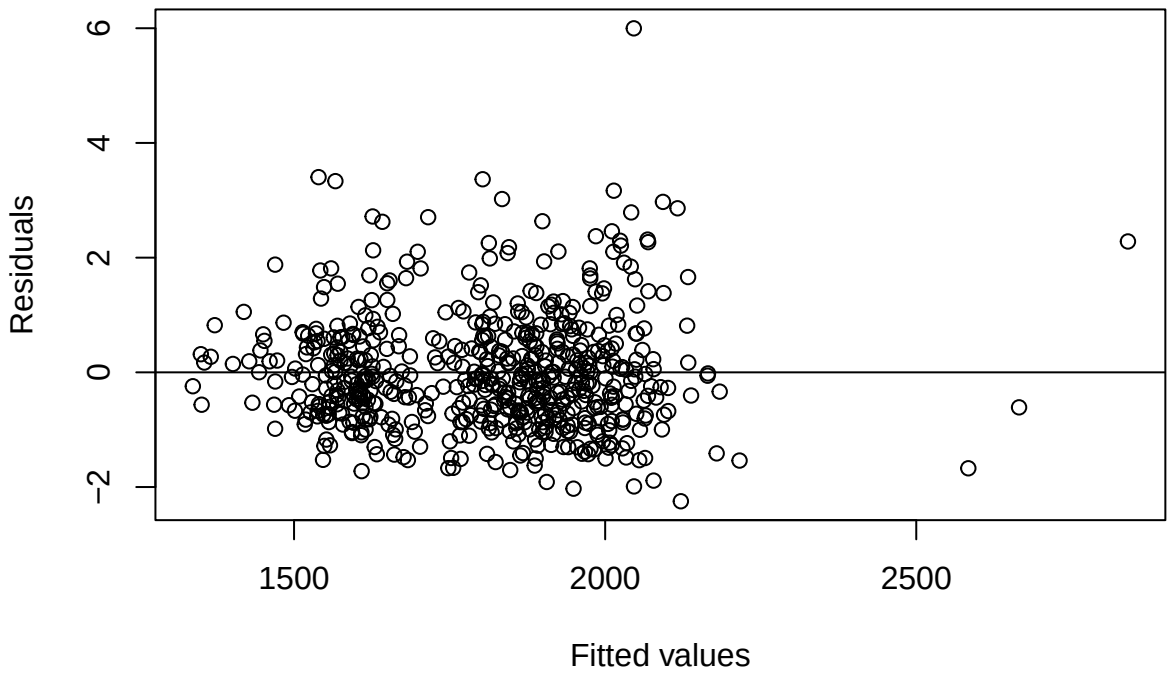

```
#plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
#abline(h = 0)

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



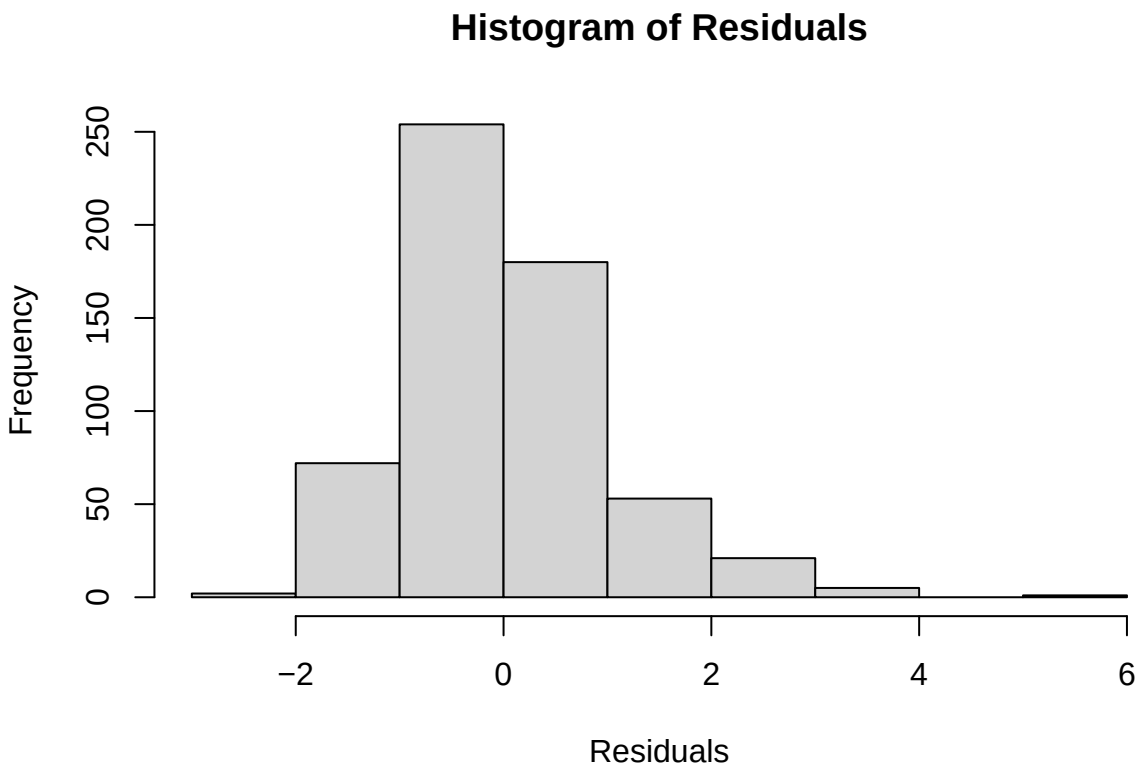

```
#plot(df1$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
#abline(h = 0)

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




```
# Constant variance and independent error term assumption: clustering observed

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



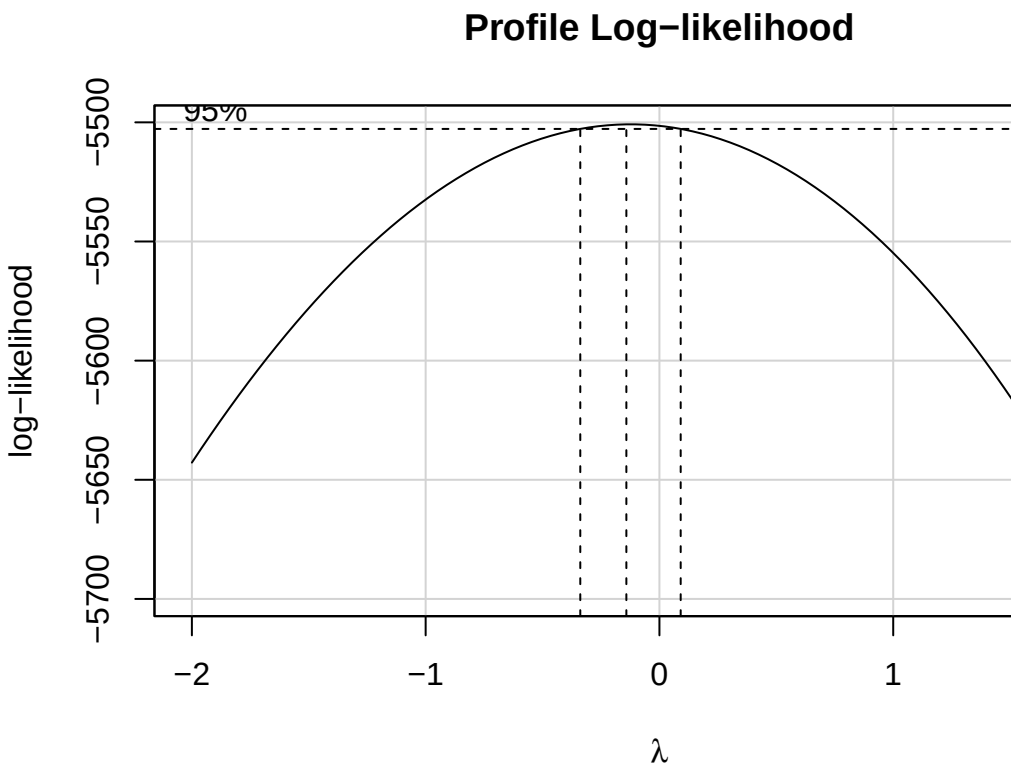

```
# Normality assumption

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 = 0.94742, p-value = 1.335e-13
```

```
# box-cox transformation
bc = boxCox(step_model1)
```



4. Improving the regression fit

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

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

```
# regression with non-linear transformation

options(scipen=0, digits=2)

model1_trans = lm(log(CO2_food+1) ~ age + income + political_party + education, data = df1)

summary(model1_trans)
```

FINAL MODEL

```
##
## Call:
## lm(formula = log(CO2_food + 1) ~ age + income + political_party +
##     education, data = df1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -0.7622 -0.1852 -0.0072  0.1840  0.9877
##
## Coefficients:
##                                                                                 Estimate
## (Intercept)                                                                     7.34e+00
## age                                                                            -2.05e-03
## income                                                                          9.34e-06
## political_partyAfD                                                              2.62e-01
## political_partyBündnis Sarah Wagenknecht                                        1.58e-01
## political_partyCDU/CSU                                                          2.18e-01
## political_partyDie Linke                                                        5.51e-02
## political_partyEiner anderen Partei                                             1.57e-01
## political_partyFDP                                                              1.90e-01
## political_partyKeine Angabe                                                     9.90e-03
## political_partySPD                                                              2.18e-01
## education(Noch) kein Abschluss                                                  4.62e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  1.34e-02
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule            3.70e-02
## educationDoktorgrad oder Habilitation                                           3.71e-02
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  7.01e-02
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss      4.99e-02
##                                                                                Std. Error
## (Intercept)                                                                     4.70e-02
## age                                                                             9.45e-04
## income                                                                          6.27e-06
## political_partyAfD                                                              4.48e-02
## political_partyBündnis Sarah Wagenknecht                                        6.36e-02
## political_partyCDU/CSU                                                          4.06e-02
## political_partyDie Linke                                                        4.86e-02
## political_partyEiner anderen Partei                                             3.59e-02
## political_partyFDP                                                              4.73e-02
## political_partyKeine Angabe                                                     8.04e-02
## political_partySPD                                                              4.12e-02
## education(Noch) kein Abschluss                                                  1.64e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  3.16e-02
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule            3.24e-02
## educationDoktorgrad oder Habilitation                                           8.00e-02
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  9.01e-02
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss      4.14e-02
##                                                                                t value
## (Intercept)                                                                     156.33
## age                                                                              -2.17
## income                                                                            1.49
## political_partyAfD                                                                5.84
## political_partyBündnis Sarah Wagenknecht                                          2.49
## political_partyCDU/CSU                                                            5.37
## political_partyDie Linke                                                          1.13
## political_partyEiner anderen Partei                                               4.38
## political_partyFDP                                                                4.01
## political_partyKeine Angabe                                                       0.12
## political_partySPD                                                                5.28
## education(Noch) kein Abschluss                                                    2.82
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  0.42
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule              1.14
```

```
## educationDoktorgrad oder Habilitation                                                    0.46
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss         0.78
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss               1.20
##                                                                                      Pr(>|t|)
## (Intercept)                                                                           < 2e-16
## age                                                                                    0.0303
## income                                                                                 0.1372
## political_partyAfD                                                                    8.8e-09
## political_partyBündnis Sarah Wagenknecht                                               0.0131
## political_partyCDU/CSU                                                                1.1e-07
## political_partyDie Linke                                                               0.2573
## political_partyEiner anderen Partei                                                   1.4e-05
## political_partyFDP                                                                    6.9e-05
## political_partyKeine Angabe                                                            0.9020
## political_partySPD                                                                    1.8e-07
## education(Noch) kein Abschluss                                                         0.0049
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)      0.6724
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                   0.2545
## educationDoktorgrad oder Habilitation                                                  0.6429
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss       0.4367
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss             0.2292
##
## (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
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.28 on 571 degrees of freedom
## Multiple R-squared:  0.128,  Adjusted R-squared:  0.104
## F-statistic: 5.24 on 16 and 571 DF,  p-value: 2.37e-10
```

```
# Checking the VIFs for multicollinearity

vif(model1_trans)
```

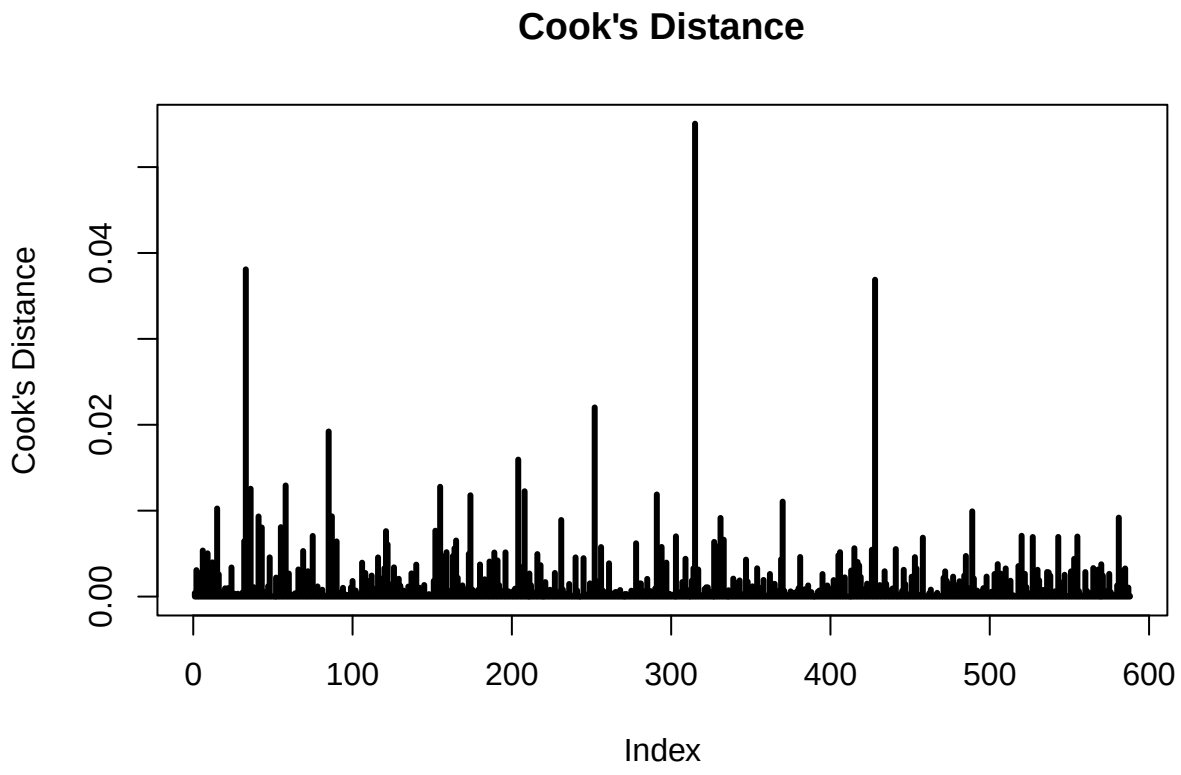
```
##                 GVIF Df GVIF^(1/(2*Df))
## age              1.3  1            1.1
## income           1.1  1            1.0
## political_party  1.3  8            1.0
```

```
## education        1.4  6             1.0
```

```
# threshold for multicollinearity
# Calculating the threshold

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

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

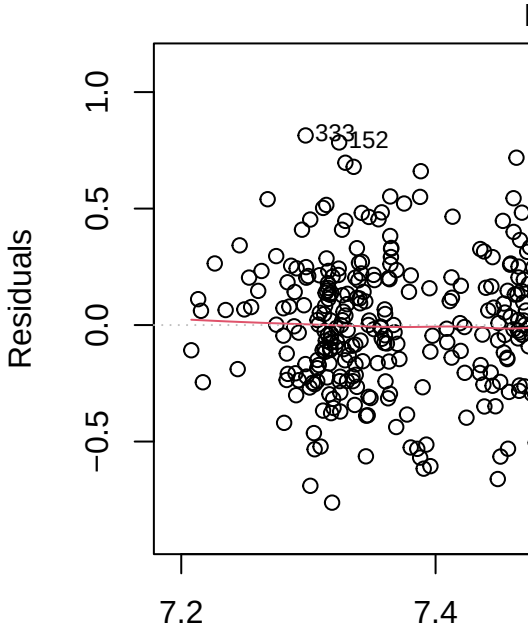
```
# Checking outliers: estimate of the influence of data point; summary of how much a regression model ch

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



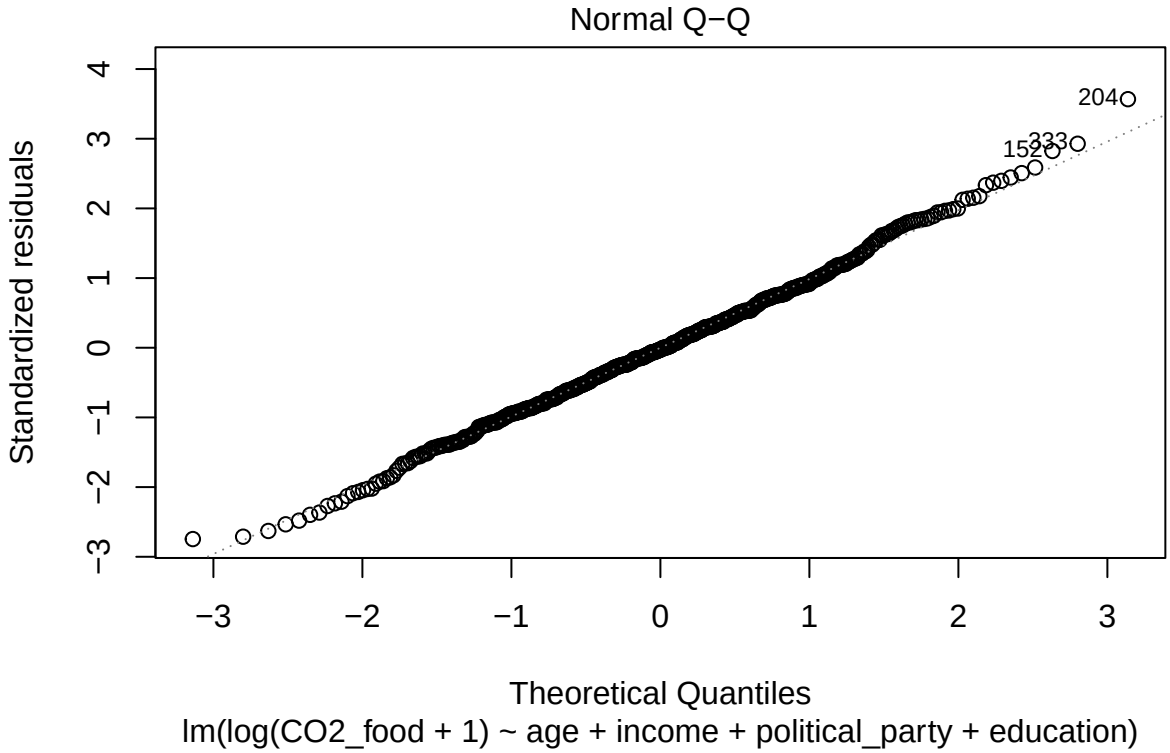

```
influential = cooks.distance(model1_trans)[which(cook >1)]

influential
```

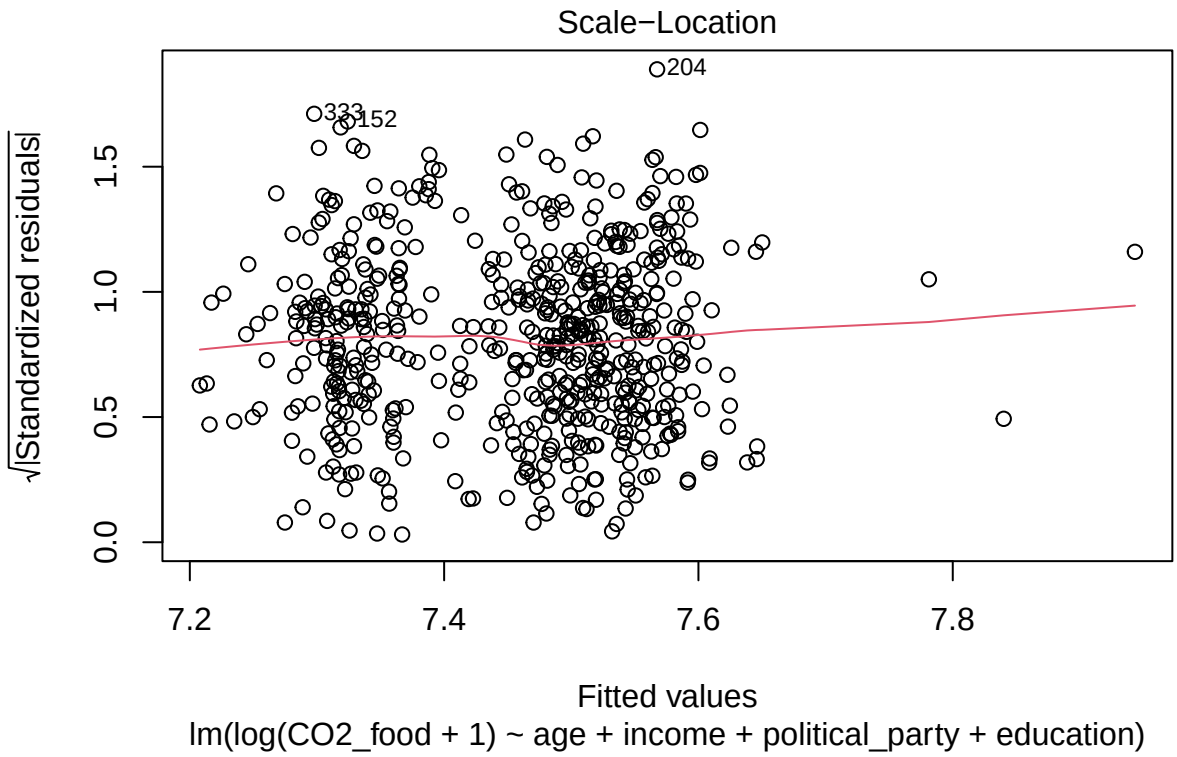
```
## named numeric(0)
```

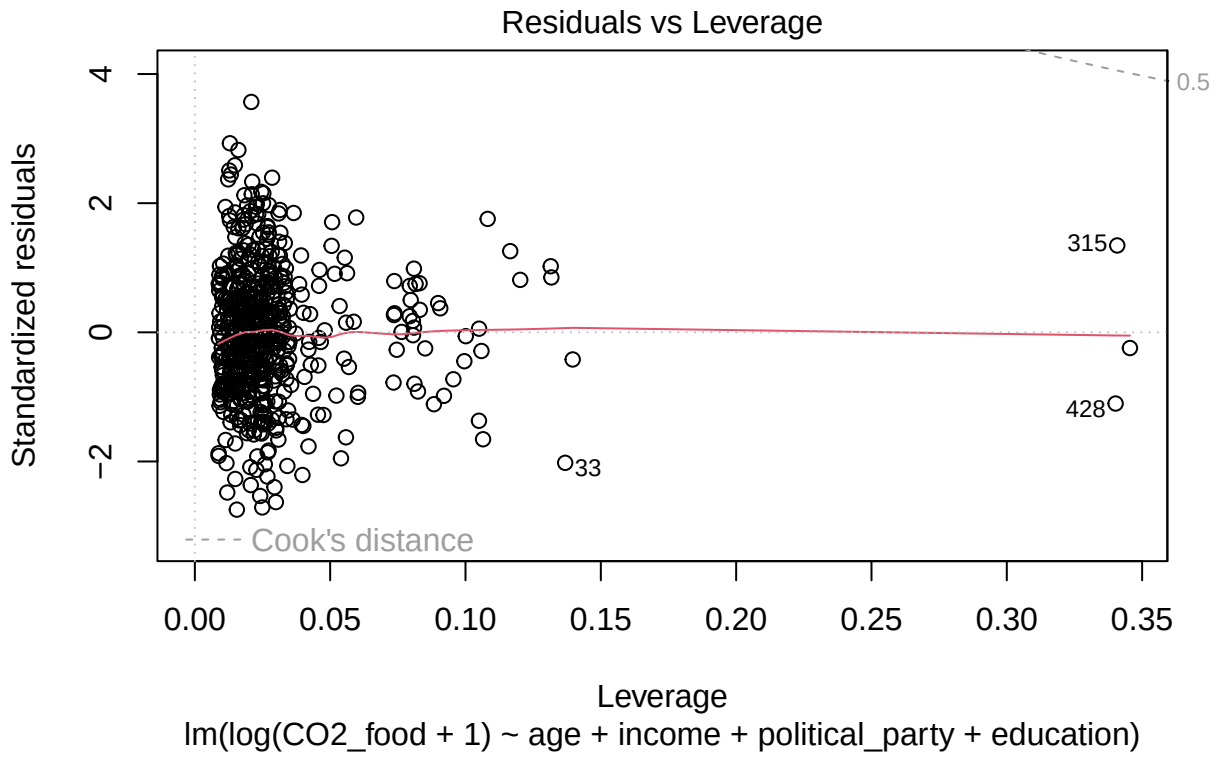
```
plot(model1_trans)
```



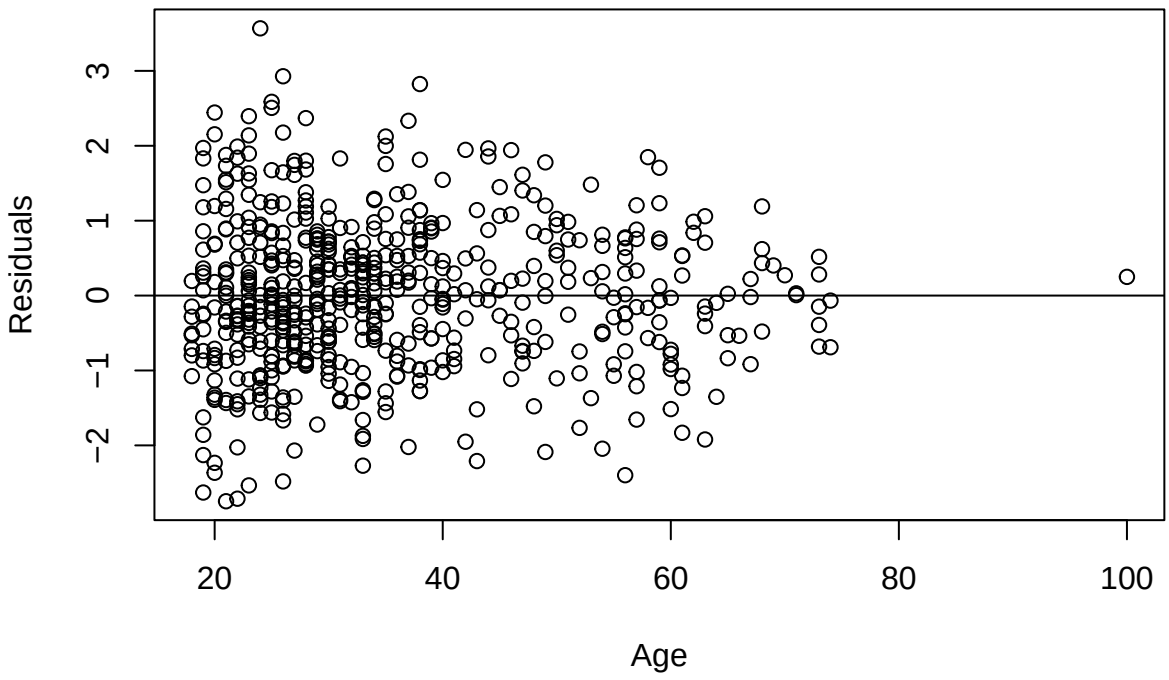
5. Assumptions check in the residuals of the transformed regression



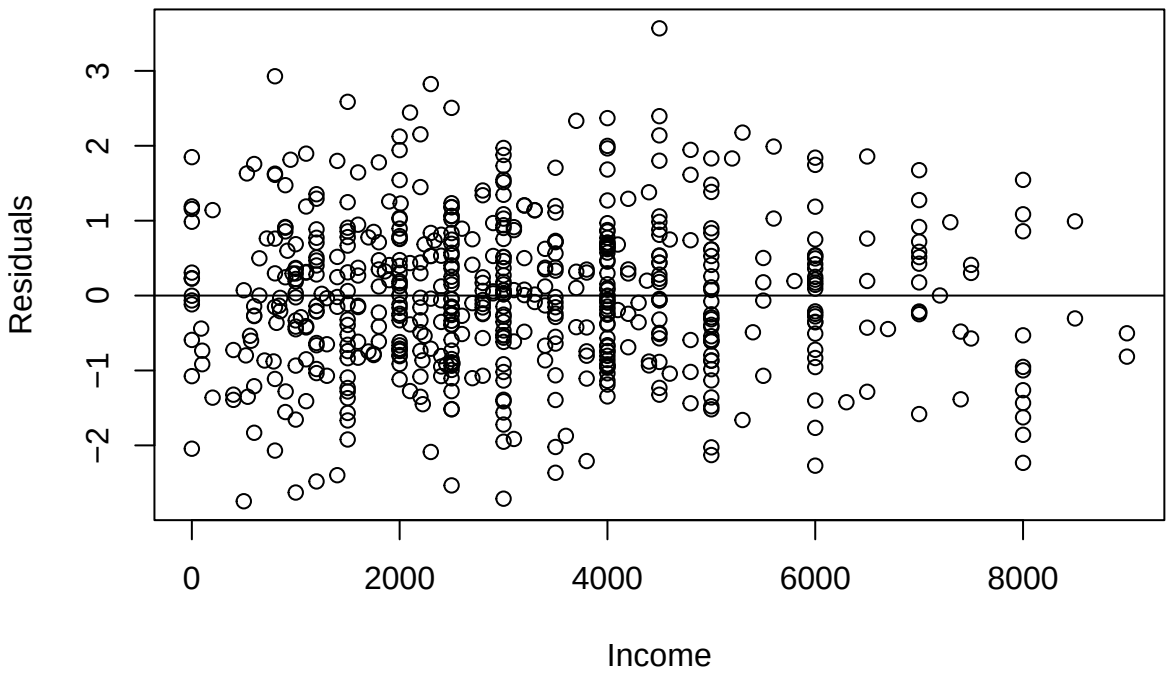

Scale−Location
204
333
152
√|Standardized residuals|
0.0
0.5
1.0
1.5
7.2
7.4
7.6
7.8
Fitted values
lm(log(CO2_food + 1) ~ age + income + political_party + education)



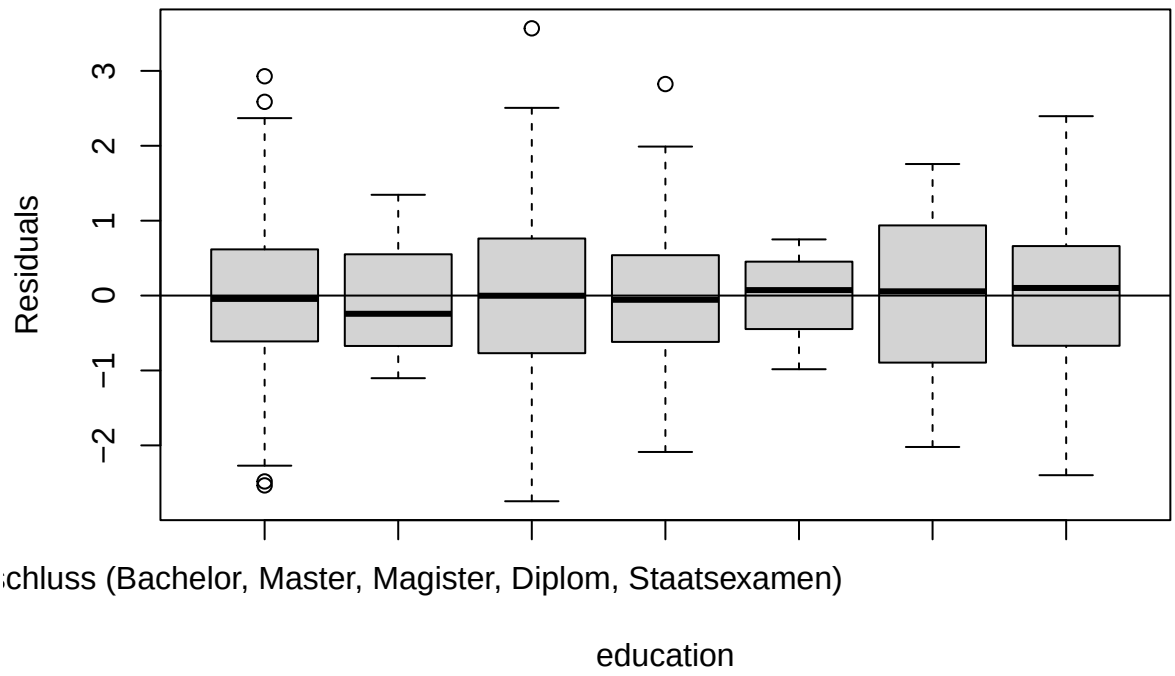

```
res1 = stdres(model1_trans) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

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



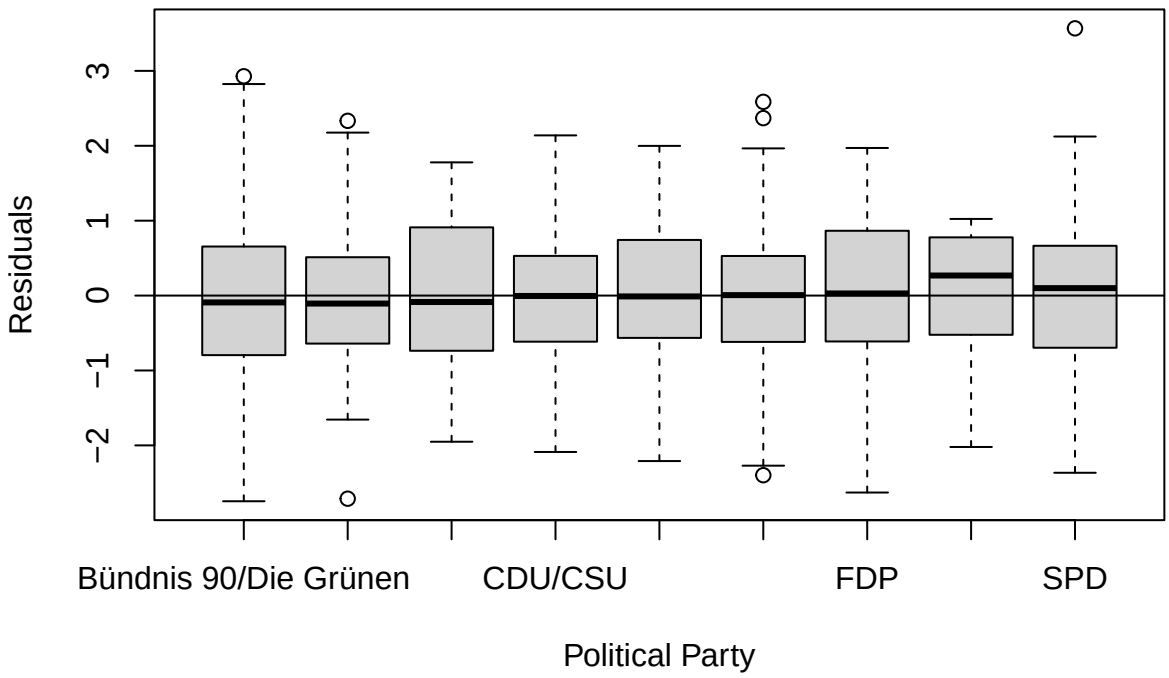

```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



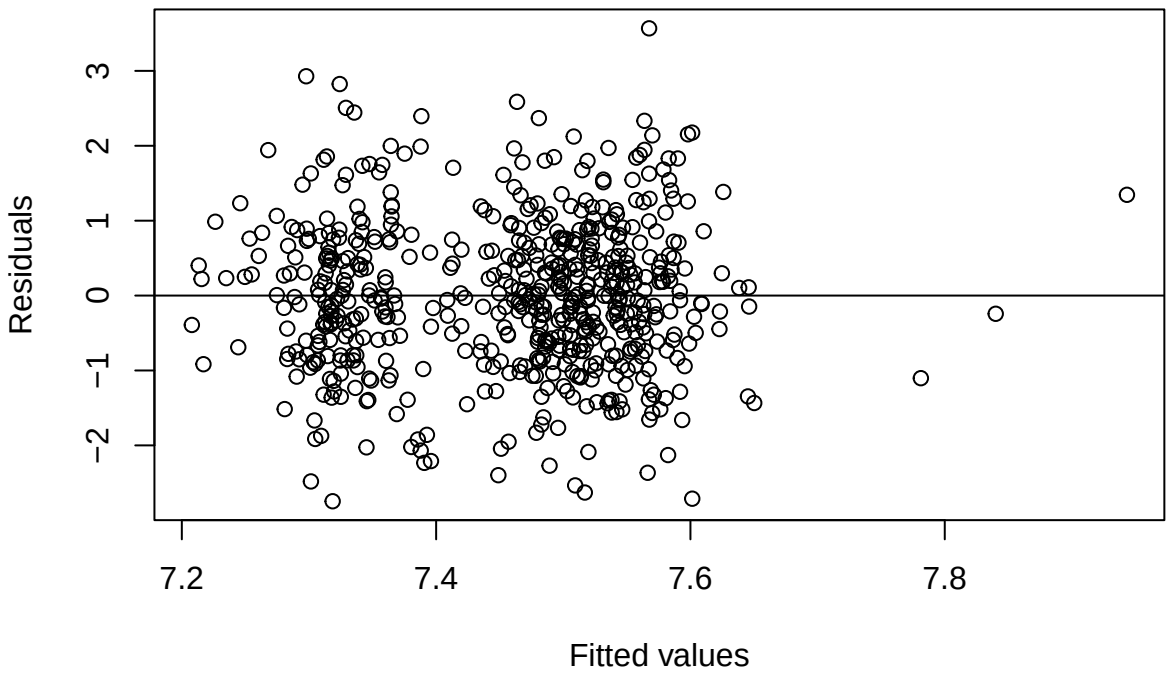

```
#plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
#abline(h = 0)

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_trans), res1, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```



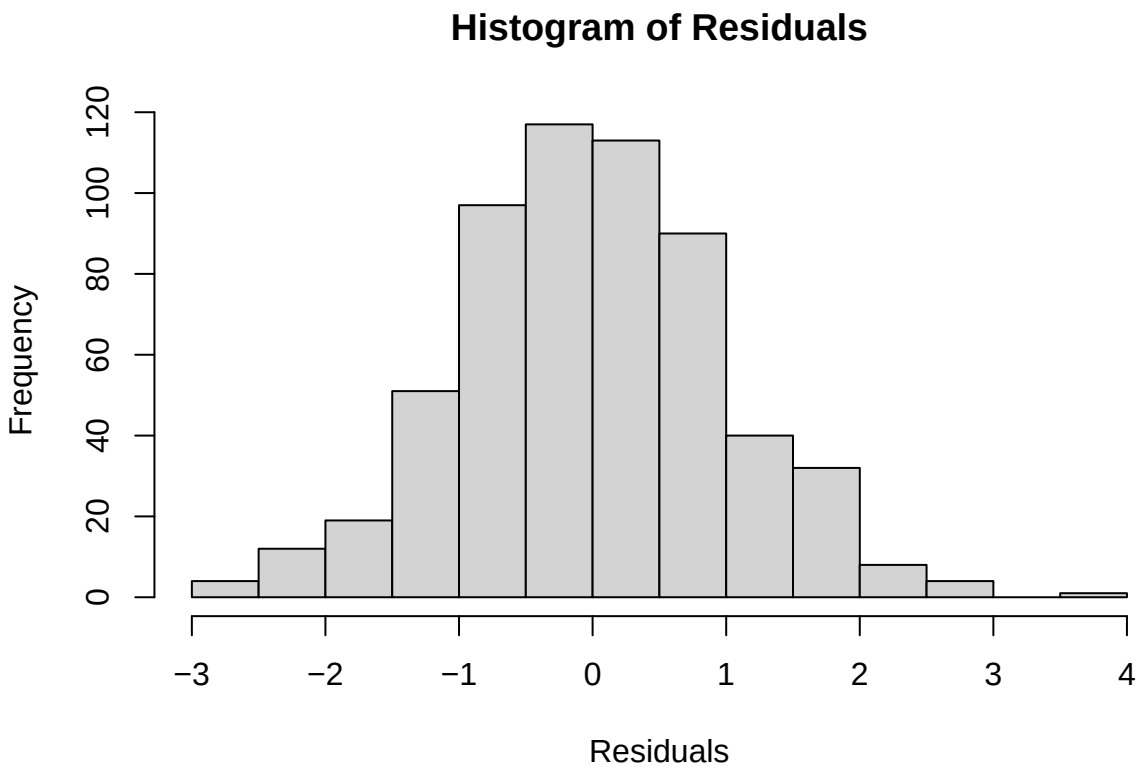

```
# Durbin-Watson Test: Independence of the error terms
# HO (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(model1_trans)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1            0.13           1.7       0
##  Alternative hypothesis: rho != 0
```

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

library(lmtest)
bptest(model1_trans)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1_trans
## BP = 27, df = 16, p-value = 0.04
```

```
# Normality assumption

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

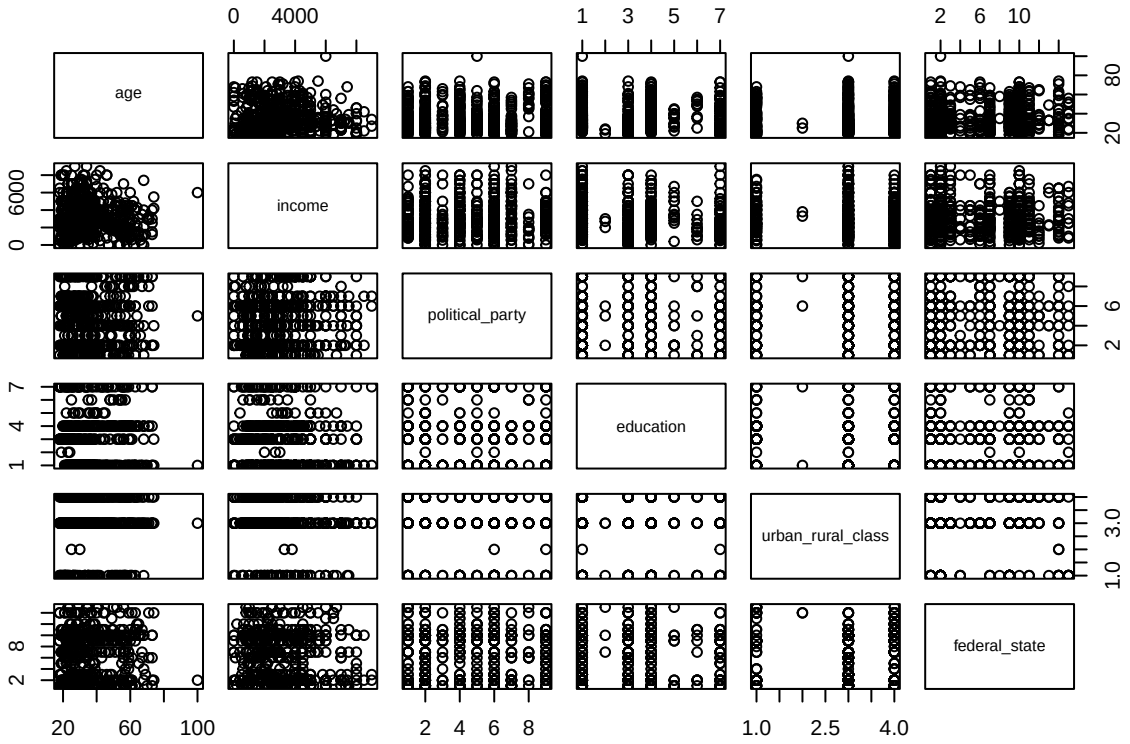



```
## normality test using shapiro-test: reject the H0
#H0:  the sample comes from a normal distribution

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

```
##
##  Shapiro-Wilk normality test
##
## data:  res1_num
## W = 1, p-value = 0.5
```

III. Multivariate Regression: belief diff food

```
# Checking the possible correlation in the data

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



1. Modeling

```
## defining a reference level

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

```
nrow(df2)
```

```
## [1] 588
```

```
# regression model

options(scipen = 0, digits=2)


model2 <- lm(belief_diff_food ~ age + income + political_party + education +  urban_rural_class + federa
summary(model2)
```

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

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -87.63 -19.85   0.21  21.05  80.03
##
## Coefficients:
##                                                                                 Estimate
## (Intercept)                                                                     3.97e+00
## age                                                                             1.27e-01
## income                                                                         -7.52e-04
## political_partyAfD                                                             -1.73e+01
## political_partyBündnis Sarah Wagenknecht                                       -1.12e+01
## political_partyCDU/CSU                                                         -1.14e+01
## political_partyDie Linke                                                       -8.53e-01
## political_partyEiner anderen Partei                                            -9.61e+00
## political_partyFDP                                                             -9.24e+00
## political_partyKeine Angabe                                                     1.15e+01
## political_partySPD                                                             -1.77e+01
## education(Noch) kein Abschluss                                                 -3.21e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -2.87e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule           -1.91e+00
## educationDoktorgrad oder Habilitation                                          -3.82e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss -9.60e+00
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss     -2.88e+00
## urban_rural_classperipher                                                       6.40e+00
## urban_rural_classsehr peripher                                                  1.04e+01
## urban_rural_classzentral                                                        2.99e+00
## federal_stateBaden-Württemberg                                                 -1.50e+00
## federal_stateBayern                                                             1.71e+00
## federal_stateBerlin                                                             4.83e+00
## federal_stateBrandenburg                                                        9.87e+00
## federal_stateBremen                                                             1.16e+01
## federal_stateHamburg                                                            4.50e+00
## federal_stateHessen                                                             4.96e-01
## federal_stateMecklenburg-Vorpommern                                             3.18e+01
## federal_stateNiedersachsen                                                     -8.54e+00
## federal_stateRheinland-Pfalz                                                   -5.09e+00
## federal_stateSaarland                                                           7.10e-01
## federal_stateSachsen-Anhalt                                                     4.06e+01
## federal_stateSchleswig-Holstein                                                -4.57e+00
## federal_stateThüringen                                                          1.45e+00
##                                                                                Std. Error
## (Intercept)                                                                      5.84e+00
## age                                                                              1.03e-01
## income                                                                           6.91e-04
## political_partyAfD                                                               4.95e+00
## political_partyBündnis Sarah Wagenknecht                                         6.93e+00
## political_partyCDU/CSU                                                           4.44e+00
## political_partyDie Linke                                                         5.36e+00
## political_partyEiner anderen Partei                                              3.97e+00
## political_partyFDP                                                               5.15e+00
## political_partyKeine Angabe                                                      8.91e+00
## political_partySPD                                                               4.52e+00
## education(Noch) kein Abschluss                                                   1.81e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  3.52e+00
```

```
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                3.59e+00
## educationDoktorgrad oder Habilitation                                               8.77e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss    1.00e+01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss          4.56e+00
## urban_rural_classperipher                                                           4.60e+00
## urban_rural_classsehr peripher                                                      2.28e+01
## urban_rural_classzentral                                                            3.37e+00
## federal_stateBaden-Württemberg                                                      4.38e+00
## federal_stateBayern                                                                 4.50e+00
## federal_stateBerlin                                                                 5.46e+00
## federal_stateBrandenburg                                                            1.15e+01
## federal_stateBremen                                                                 8.37e+00
## federal_stateHamburg                                                                6.81e+00
## federal_stateHessen                                                                 5.23e+00
## federal_stateMecklenburg-Vorpommern                                                 2.19e+01
## federal_stateNiedersachsen                                                          5.32e+00
## federal_stateRheinland-Pfalz                                                        6.56e+00
## federal_stateSaarland                                                               1.03e+01
## federal_stateSachsen-Anhalt                                                         1.59e+01
## federal_stateSchleswig-Holstein                                                     7.67e+00
## federal_stateThüringen                                                              1.16e+01
##                                                                                   t value
## (Intercept)                                                                        0.68
## age                                                                                1.23
## income                                                                            -1.09
## political_partyAfD                                                                -3.50
## political_partyBündnis Sarah Wagenknecht                                          -1.61
## political_partyCDU/CSU                                                            -2.57
## political_partyDie Linke                                                          -0.16
## political_partyEiner anderen Partei                                               -2.42
## political_partyFDP                                                                -1.80
## political_partyKeine Angabe                                                        1.29
## political_partySPD                                                                -3.91
## education(Noch) kein Abschluss                                                    -1.77
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -0.82
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule              -0.53
## educationDoktorgrad oder Habilitation                                             -0.44
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  -0.96
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss        -0.63
## urban_rural_classperipher                                                          1.39
## urban_rural_classsehr peripher                                                     0.46
## urban_rural_classzentral                                                           0.89
## federal_stateBaden-Württemberg                                                    -0.34
## federal_stateBayern                                                                0.38
## federal_stateBerlin                                                                0.88
## federal_stateBrandenburg                                                           0.86
## federal_stateBremen                                                                1.38
## federal_stateHamburg                                                               0.66
## federal_stateHessen                                                                0.09
## federal_stateMecklenburg-Vorpommern                                                1.45
## federal_stateNiedersachsen                                                        -1.61
## federal_stateRheinland-Pfalz                                                      -0.78
## federal_stateSaarland                                                              0.07
## federal_stateSachsen-Anhalt                                                        2.56
```

```
## federal_stateSchleswig-Holstein                                                  -0.60
## federal_stateThüringen                                                            0.12
##                                                                                Pr(>|t|)
## (Intercept)                                                                      0.4967
## age                                                                              0.2204
## income                                                                           0.2770
## political_partyAfD                                                               0.0005
## political_partyBündnis Sarah Wagenknecht                                         0.1083
## political_partyCDU/CSU                                                           0.0105
## political_partyDie Linke                                                         0.8734
## political_partyEiner anderen Partei                                              0.0159
## political_partyFDP                                                               0.0732
## political_partyKeine Angabe                                                      0.1968
## political_partySPD                                                               0.0001
## education(Noch) kein Abschluss                                                   0.0765
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   0.4151
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule             0.5940
## educationDoktorgrad oder Habilitation                                            0.6635
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   0.3383
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss       0.5283
## urban_rural_classperipher                                                        0.1645
## urban_rural_classsehr peripher                                                   0.6489
## urban_rural_classzentral                                                         0.3763
## federal_stateBaden-Württemberg                                                   0.7319
## federal_stateBayern                                                              0.7040
## federal_stateBerlin                                                              0.3771
## federal_stateBrandenburg                                                         0.3904
## federal_stateBremen                                                              0.1672
## federal_stateHamburg                                                             0.5091
## federal_stateHessen                                                              0.9245
## federal_stateMecklenburg-Vorpommern                                              0.1474
## federal_stateNiedersachsen                                                       0.1086
## federal_stateRheinland-Pfalz                                                     0.4383
## federal_stateSaarland                                                            0.9451
## federal_stateSachsen-Anhalt                                                      0.0107
## federal_stateSchleswig-Holstein                                                  0.5518
## federal_stateThüringen                                                           0.9010
##
## (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: 30 on 554 degrees of freedom
## Multiple R-squared:  0.0945, Adjusted R-squared:  0.0405
## F-statistic: 1.75 on 33 and 554 DF,  p-value: 0.00663
```

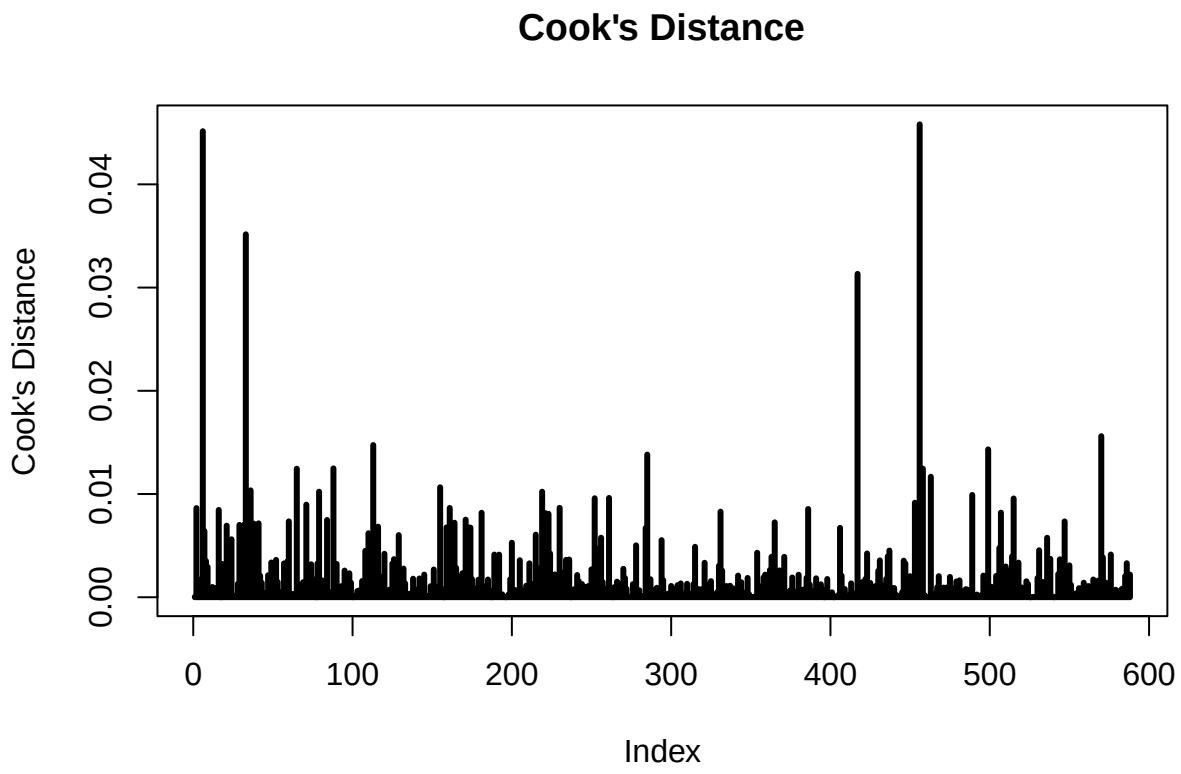
```
# Checking the VIFs for multicollinearity

vif(model2)
```

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

```
# threshold for multicollinearity
# Calculating the threshold

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)
```




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

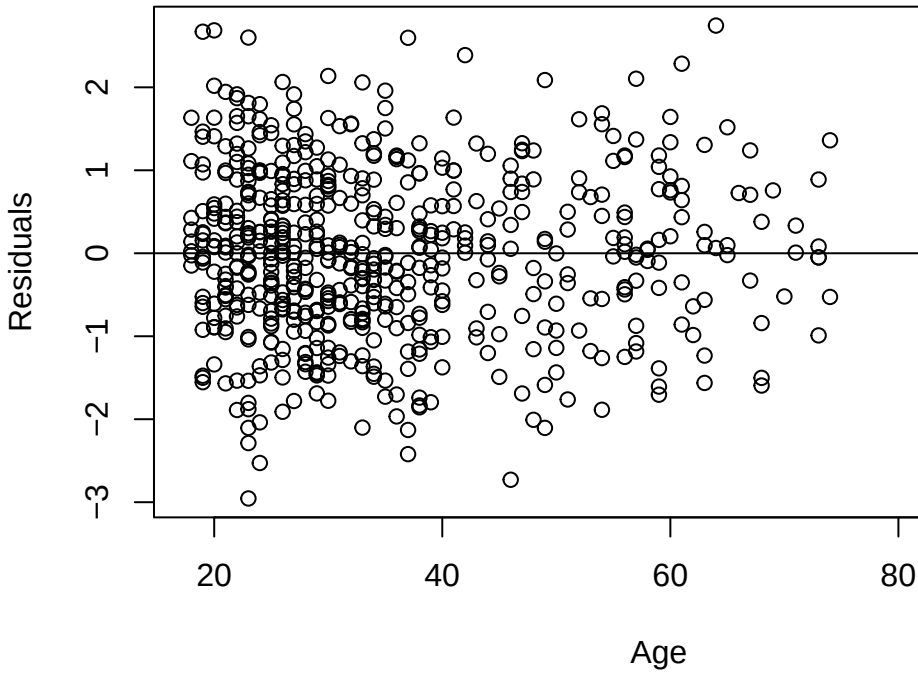
```
##      2      6      7     16     21     29     32     33     36     38     41
## 0.0086 0.0451 0.0064 0.0085 0.0069 0.0070 0.0070 0.0352 0.0104 0.0071 0.0072
##     60     65     71     79     84     88    110    113    116    129    155
## 0.0074 0.0125 0.0090 0.0102 0.0075 0.0125 0.0062 0.0147 0.0068 0.0060 0.0106
##    159    161    164    171    173    174    181    215    219    221    223
## 0.0068 0.0086 0.0072 0.0075 0.0068 0.0067 0.0082 0.0061 0.0102 0.0082 0.0081
##    230    252    261    284    285    331    365    386    406    417    453
## 0.0087 0.0096 0.0096 0.0067 0.0138 0.0083 0.0073 0.0085 0.0067 0.0313 0.0092
##    456    458    463    489    499    507    515    547    570
## 0.0458 0.0125 0.0117 0.0099 0.0143 0.0082 0.0096 0.0073 0.0156
```

```
influential = influential[!is.na(influential)]

influential_vector = c(as.numeric(rownames(data.frame(influential))))


df2[influential_vector, ]
```

```
## # A tibble: 53 x 7
##      age income political_party education                urban~1 feder~2 belie~3
##    <int>  <dbl> <fct>           <fct>                    <fct>   <fct>     <dbl>
##  1    59    800 Keine Angabe    Allgemeine oder fachgeb~ sehr z~ Hessen      -26
```

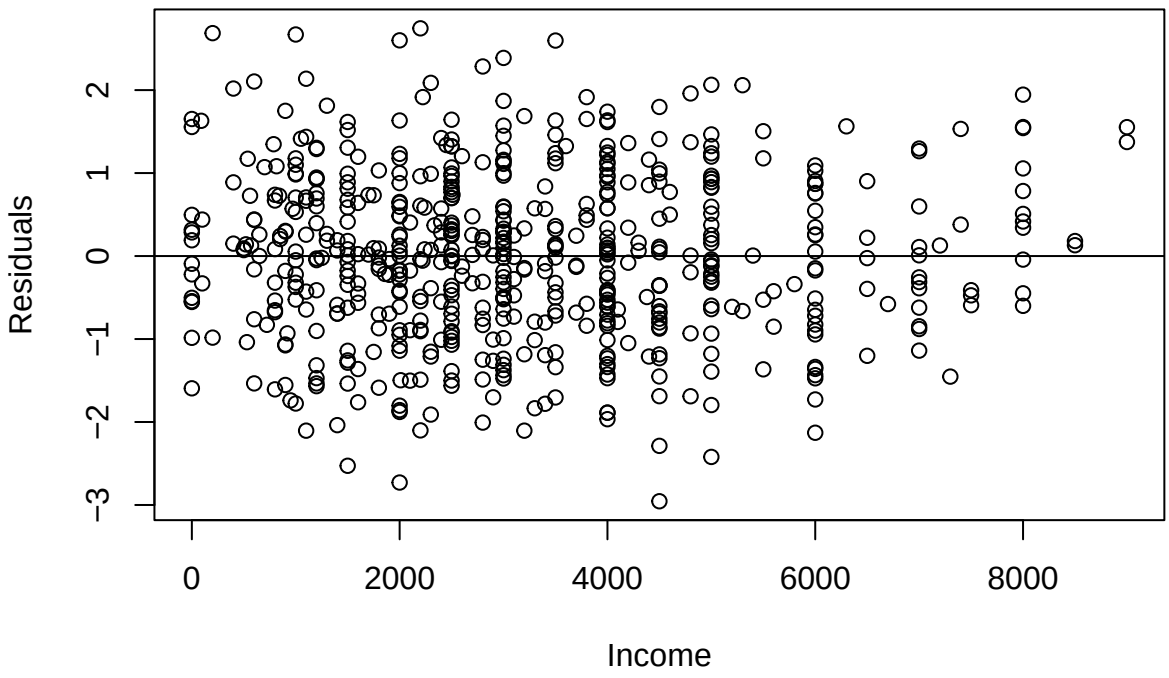
```
##  2    49   2300 CDU/CSU         Berufsausbildung, Lehre~ zentral Sachse~      93
##  3    57    600 CDU/CSU         Realschulabschluss (Mit~ zentral Baden-~      60
##  4    33   6500 CDU/CSU         Berufsausbildung, Lehre~ zentral Sachse~      57
##  5    54   2900 AfD             Hauptschulabschluss (Vo~ zentral Rheinl~     -56
##  6    59   2000 SPD             (Fach-) Hochschulabschl~ zentral Brande~      38
##  7    48   1750 Keine Angabe    Hauptschulabschluss (Vo~ sehr z~ Hessen      -21
##  8    37   3500 Keine Angabe    Hauptschulabschluss (Vo~ sehr z~ Bayern       82
##  9    19   1000 FDP             Allgemeine oder fachgeb~ sehr z~ Bayern       74
## 10    65   1500 CDU/CSU         Berufsausbildung, Lehre~ sehr z~ Bremen       53
## # ... with 43 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: belief_diff_food
```

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

# Linearity assumption/Mean zero assumption

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

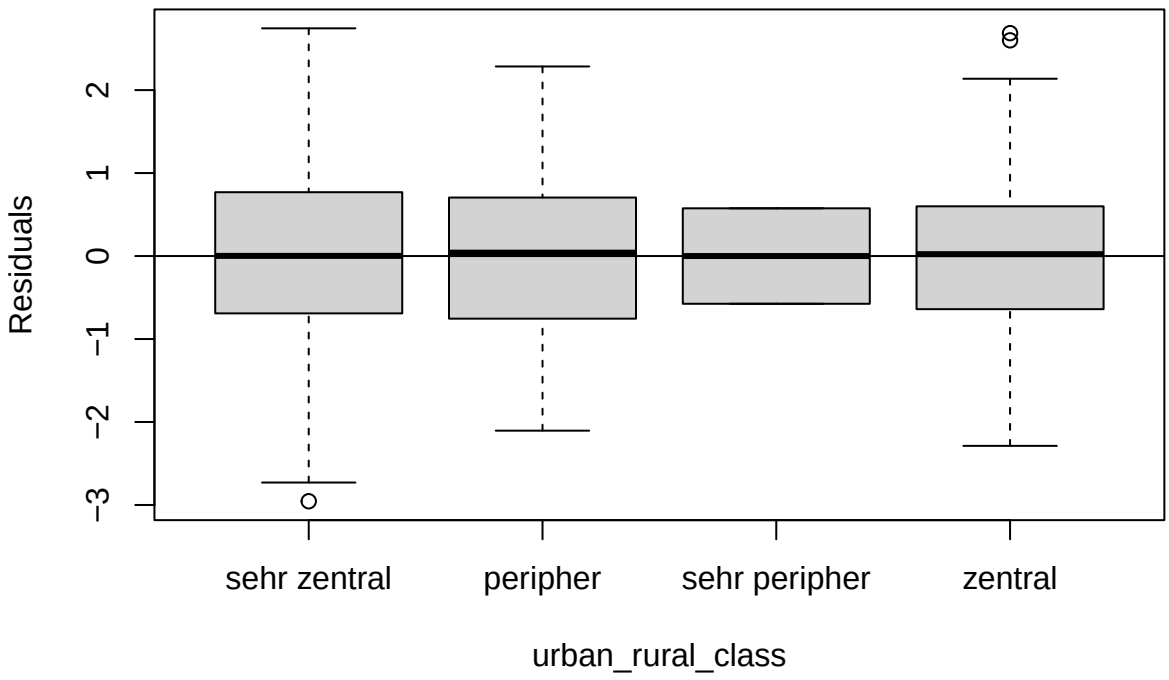


2. Assumptions check in the residuals

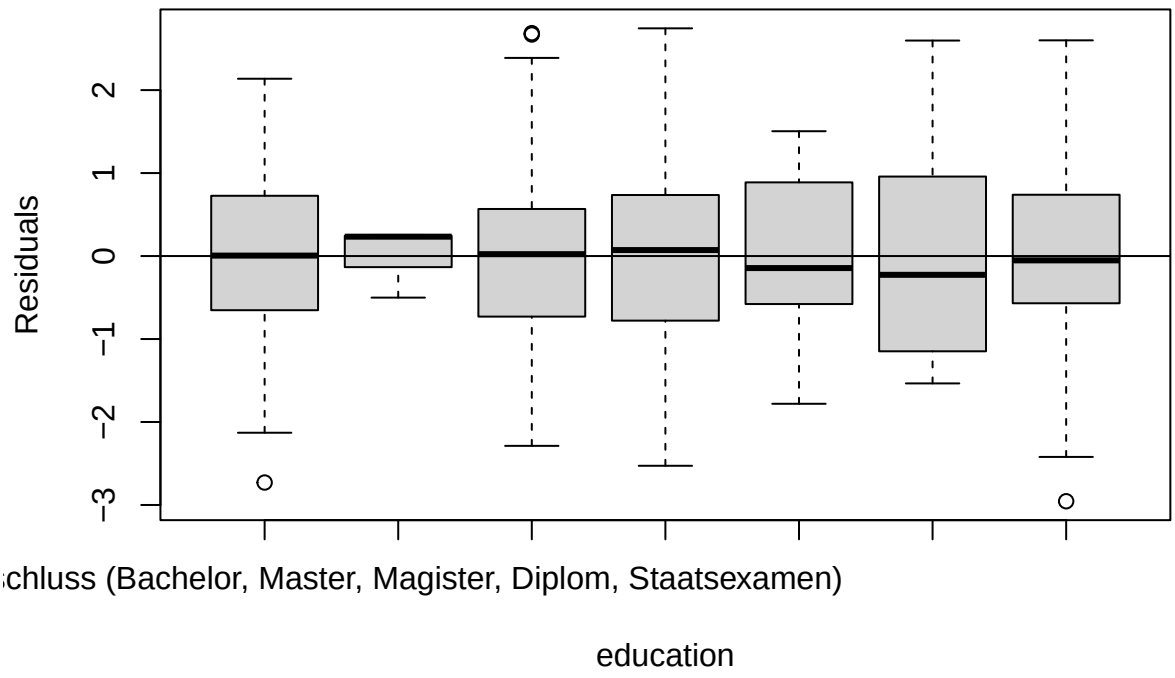
```r
plot(df2$income, res2, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



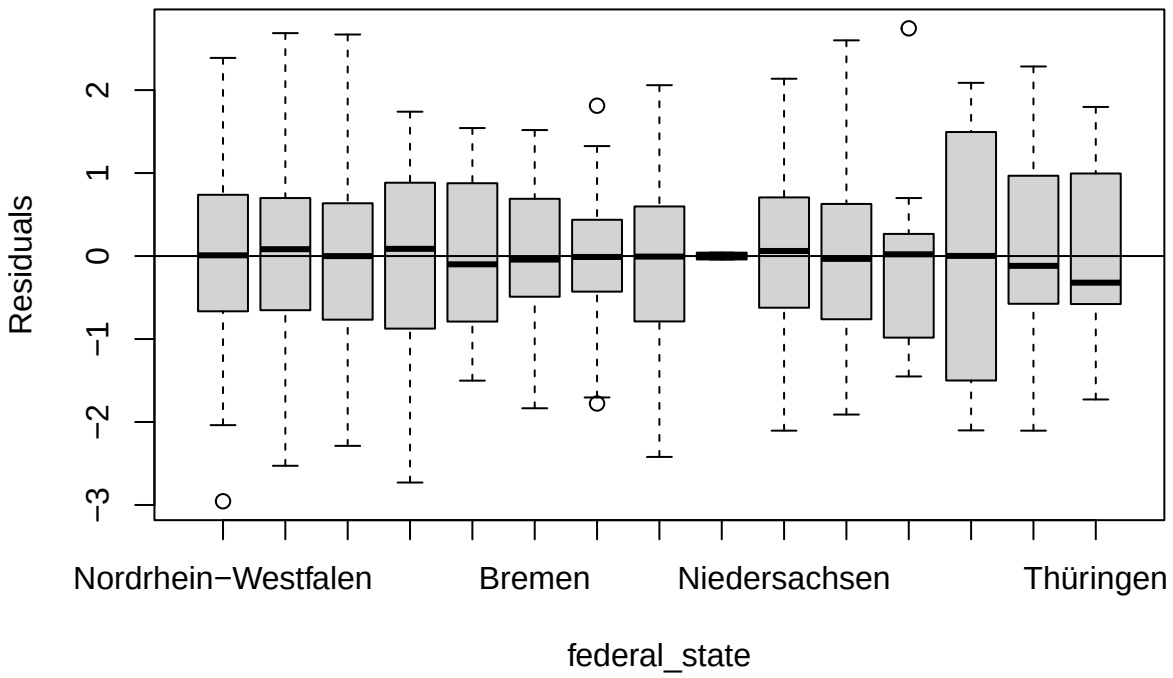

```
plot(df2$urban_rural_class, res2, xlab = "urban_rural_class", ylab = "Residuals")
abline(h = 0)
```



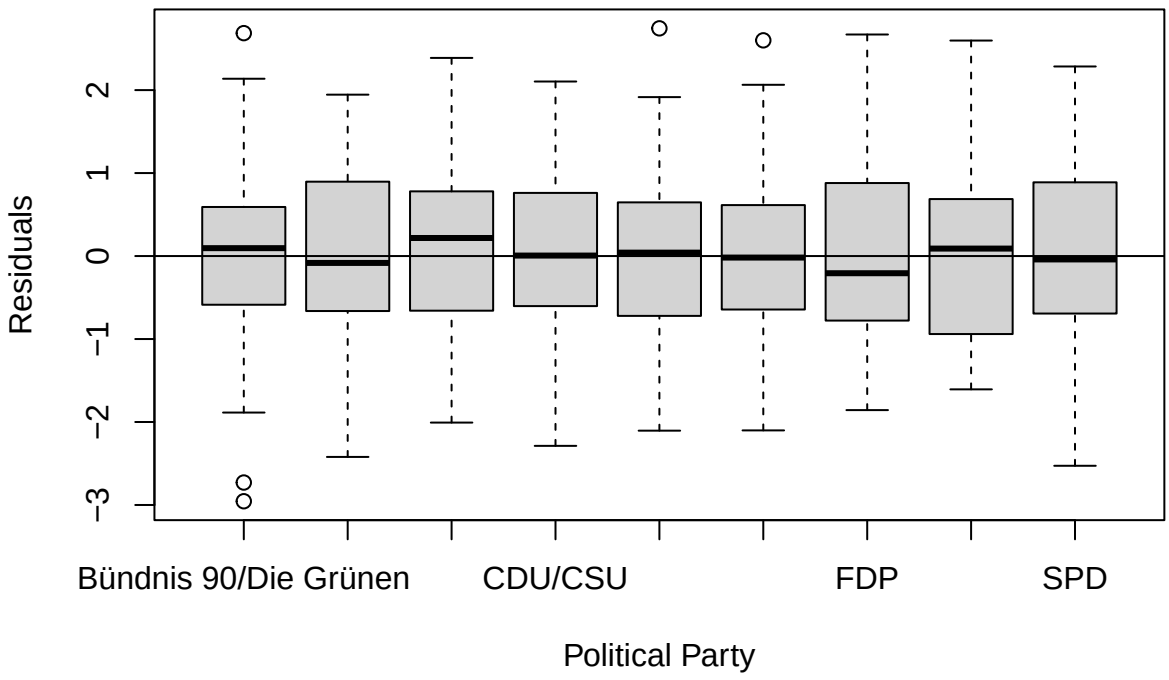

```
plot(df2$education, res2, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



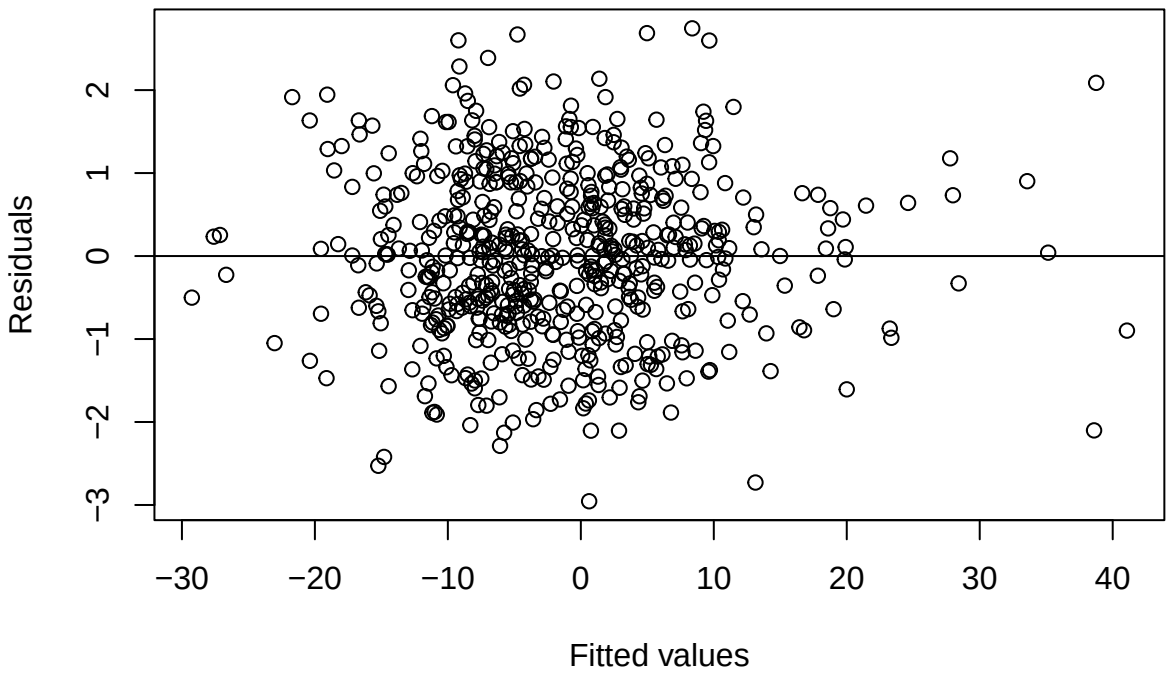

```
plot(df2$federal_state, res2, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```




```
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)
```



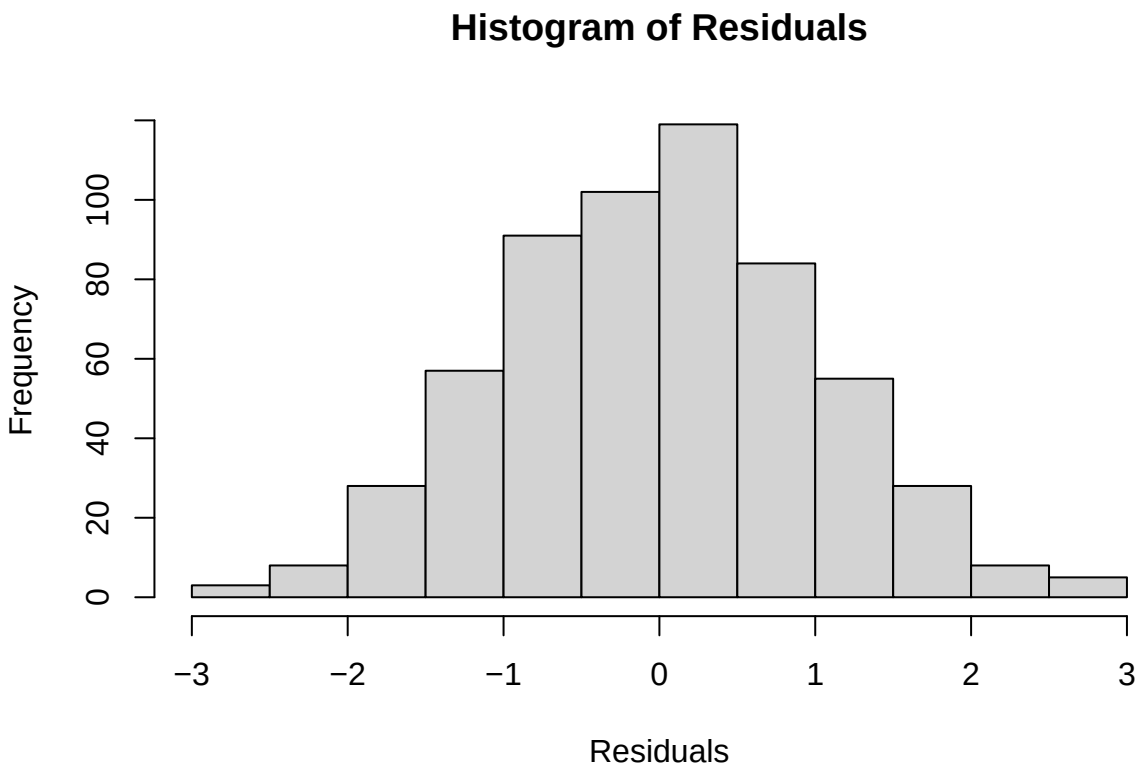

```
# Durbin-Watson Test: Independence of the error terms
# H0 (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(model2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1           0.091           1.8   0.016
##  Alternative hypothesis: rho != 0
```

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

bptest(model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model2
## BP = 25, df = 33, p-value = 0.8
```

```
# 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 = 1, p-value = 0.6
```

3. Variable selection

```
### Backward regression using AIC: starting with all of the variables
options(scipen = -2)

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

FINAL MODEL

```
## Start:  AIC=4e+03
## belief_diff_food ~ age + income + political_party + education +
##     urban_rural_class + federal_state
##
##                      Df Sum of Sq    RSS  AIC
## - federal_state     14     18145 524152 4034
## - education          6      4129 510136 4034
## - urban_rural_class  3      2002 508009 4038
## - income             1      1082 507088 4041
## - age                1      1375 507382 4041
## <none>                           506007 4041
## - political_party    8     28531 534538 4058
##
## Step:  AIC=4e+03
## belief_diff_food ~ age + income + political_party + education +
##     urban_rural_class
##
##                      Df Sum of Sq    RSS  AIC
## - education          6      4837 528988 4028
## - urban_rural_class  3      1548 525700 4030
## - income             1       614 524766 4033
## - age                1      1239 525391 4034
## <none>                           524152 4034
## - political_party    8     26653 550805 4047
##
## Step:  AIC=4e+03
## belief_diff_food ~ age + income + political_party + urban_rural_class
##
##                      Df Sum of Sq    RSS  AIC
## - urban_rural_class  3      1096 530085 4023
## - income             1       353 529342 4026
## <none>                           528988 4028
## - age                1      2038 531027 4028
## - political_party    8     26735 555724 4041
##
## Step:  AIC=4e+03
## belief_diff_food ~ age + income + political_party
##
##                    Df Sum of Sq    RSS  AIC
## - income            1       392 530477 4021
## <none>                          530085 4023
## - age               1      2212 532297 4023
## - political_party   8     26043 556127 4035
##
## Step:  AIC=4e+03
## belief_diff_food ~ age + political_party
##
##                    Df Sum of Sq    RSS  AIC
## <none>                          530477 4021
## - age               1      2210 532687 4022
## - political_party   8     26403 556880 4034
```

```
summary(step_model2)
```

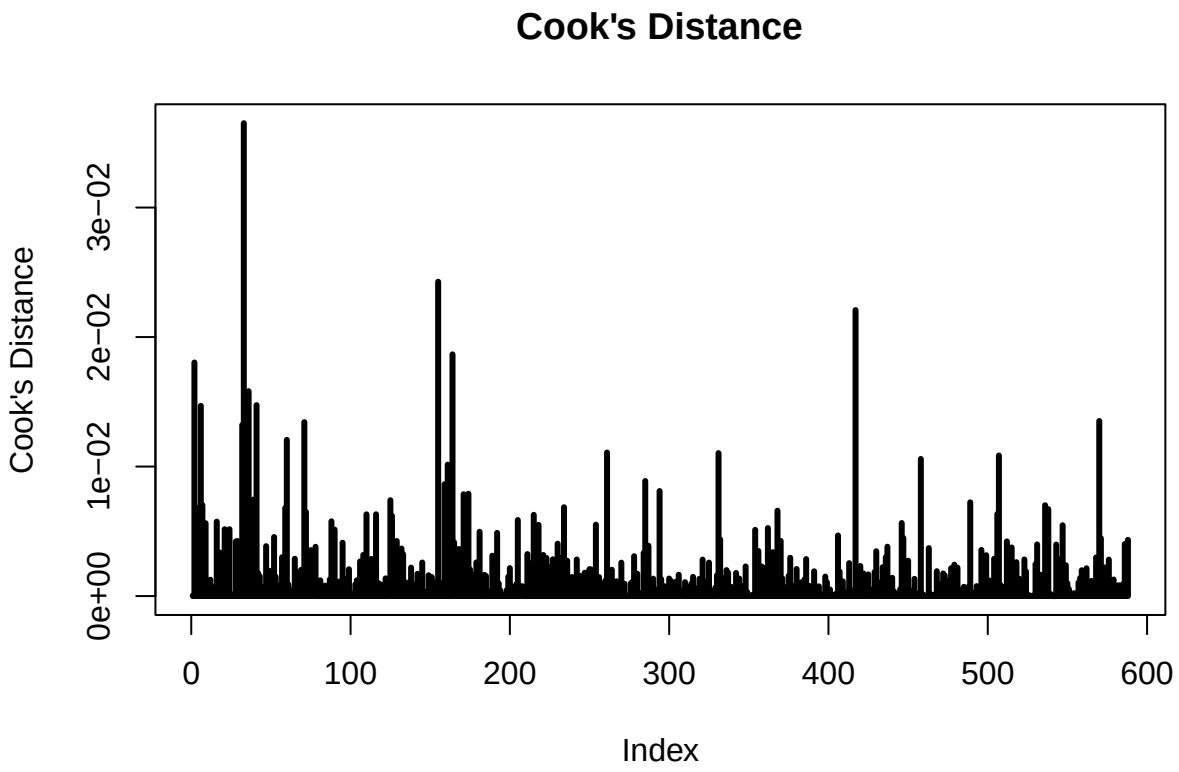
```
##
## Call:
## lm(formula = belief_diff_food ~ age + political_party, data = df2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -90.92 -20.96  -0.04  22.33  95.12
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              5.46e-01   4.00e+00    0.14  8.9e-01
## age                                      1.47e-01   9.45e-02    1.55  1.2e-01
## political_partyAfD                      -1.65e+01   4.74e+00   -3.48  5.4e-04
## political_partyBündnis Sarah Wagenknecht -1.21e+01   6.83e+00   -1.77  7.7e-02
## political_partyCDU/CSU                  -9.85e+00   4.35e+00   -2.26  2.4e-02
## political_partyDie Linke                -1.70e+00   5.22e+00   -0.33  7.5e-01
## political_partyEiner anderen Partei     -8.60e+00   3.83e+00   -2.24  2.5e-02
## political_partyFDP                      -8.97e+00   5.06e+00   -1.77  7.7e-02
## political_partyKeine Angabe              1.12e+01   8.43e+00    1.32  1.9e-01
## political_partySPD                      -1.66e+01   4.43e+00   -3.74  2.0e-04
##
## (Intercept)
## age
## political_partyAfD                      ***
## political_partyBündnis Sarah Wagenknecht .
## political_partyCDU/CSU                  *
## political_partyDie Linke
## political_partyEiner anderen Partei     *
## political_partyFDP                      .
## political_partyKeine Angabe
## political_partySPD                      ***
## ---
## Signif. codes:  0 '***' 1e-03 '**' 1e-02 '*' 5e-02 '.' 0.1 ' ' 1
##
## Residual standard error: 30 on 578 degrees of freedom
## Multiple R-squared:  0.0507, Adjusted R-squared:  0.0359
## F-statistic: 3.43 on 9 and 578 DF,  p-value: 3.98e-04
```

```
vif(step_model2)
```

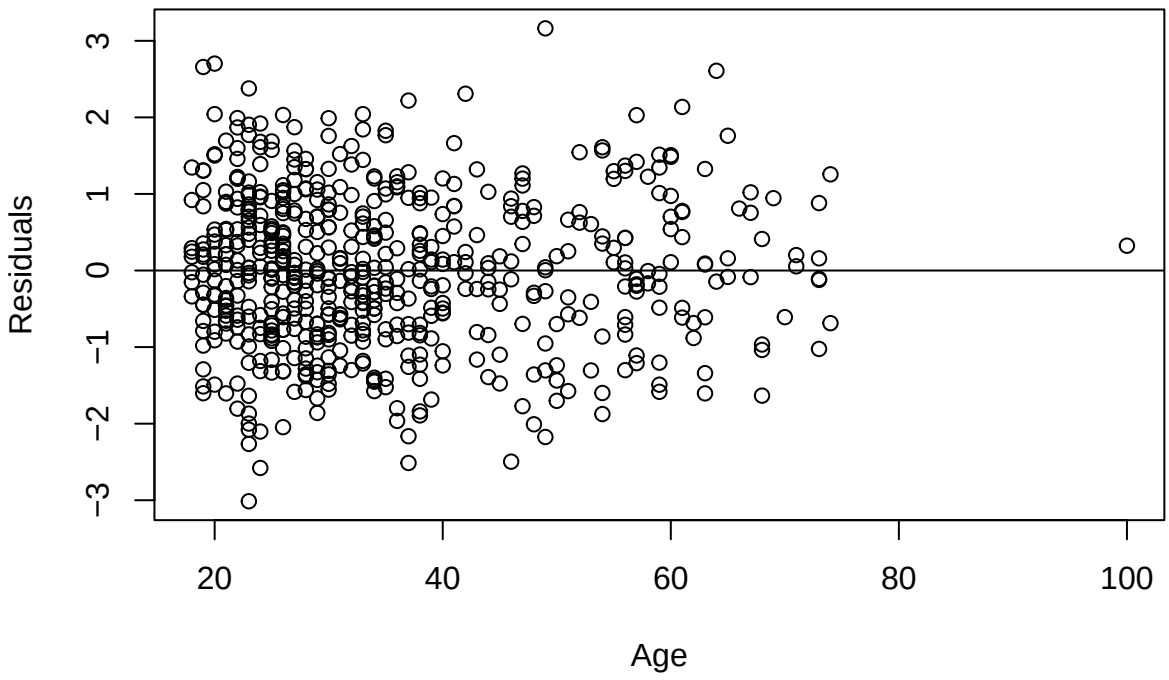
```
##                 GVIF Df GVIF^(1/(2*Df))
## age              1.1  1               1
## political_party  1.1  8               1
```

```
# Checking outliers

cook = cooks.distance(step_model2)
plot(cook,
     type="h",
     lwd=3,
```

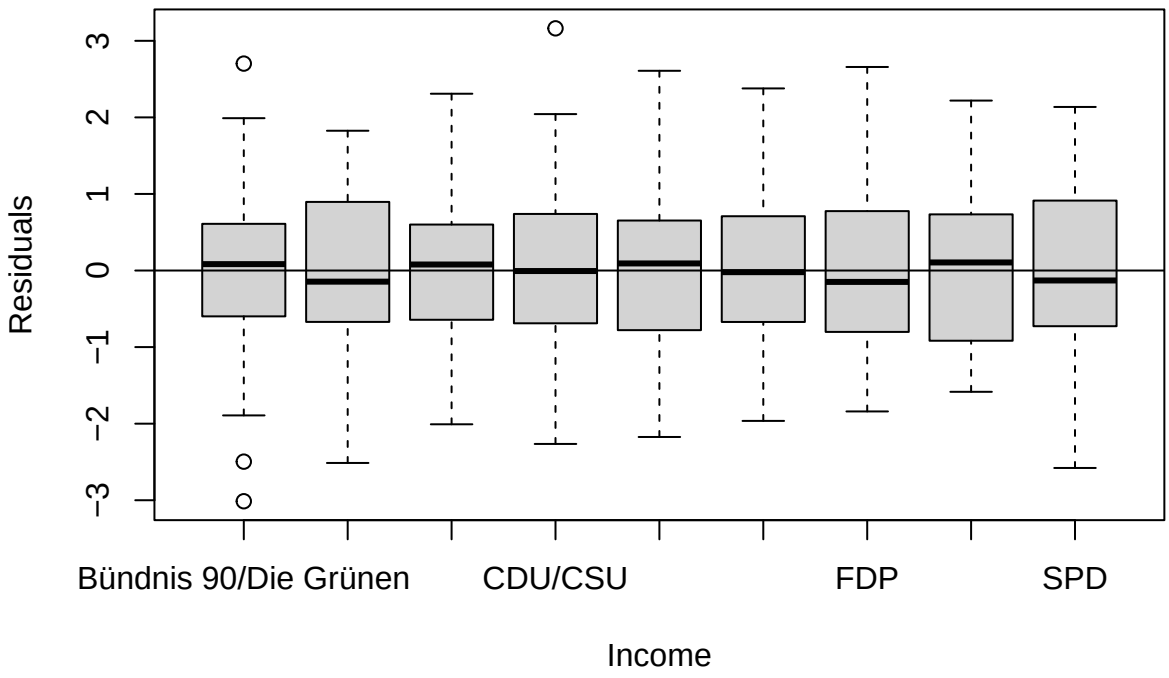
```
     ylab = "Cook's Distance",
     main="Cook's Distance")
abline(h = 1)
```




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

# Linearity assumption/Mean zero assumption

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

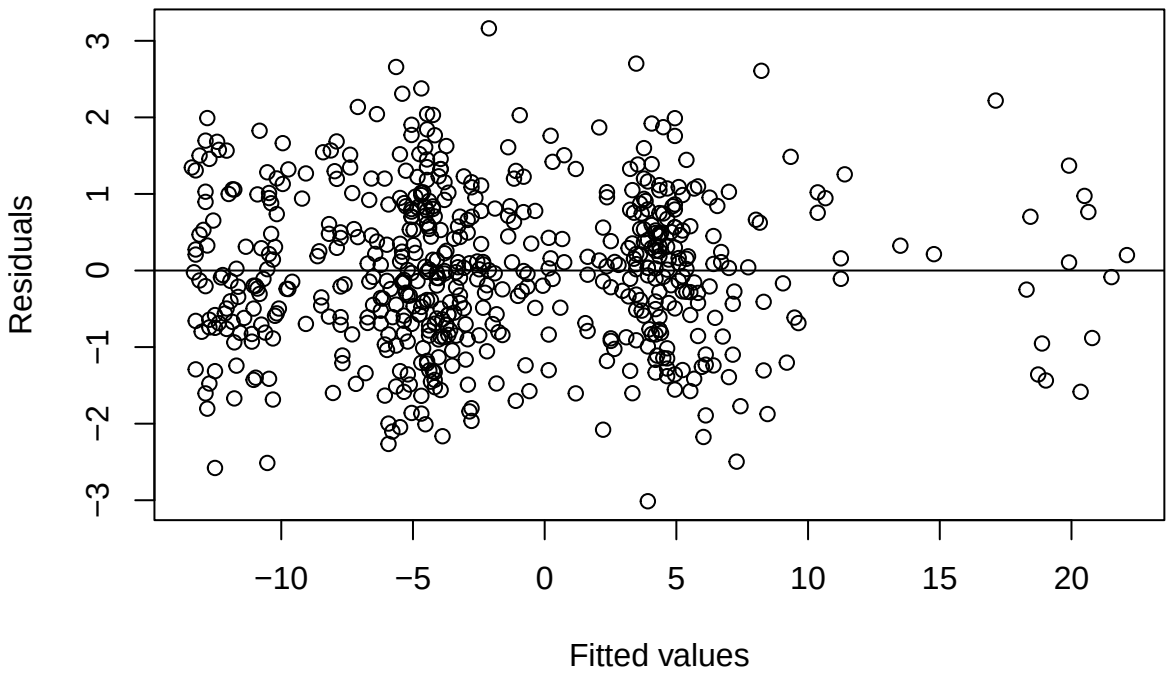



```
plot(df2$political_party, res2, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```




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

```
# Constant variance and independent error term assumption

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



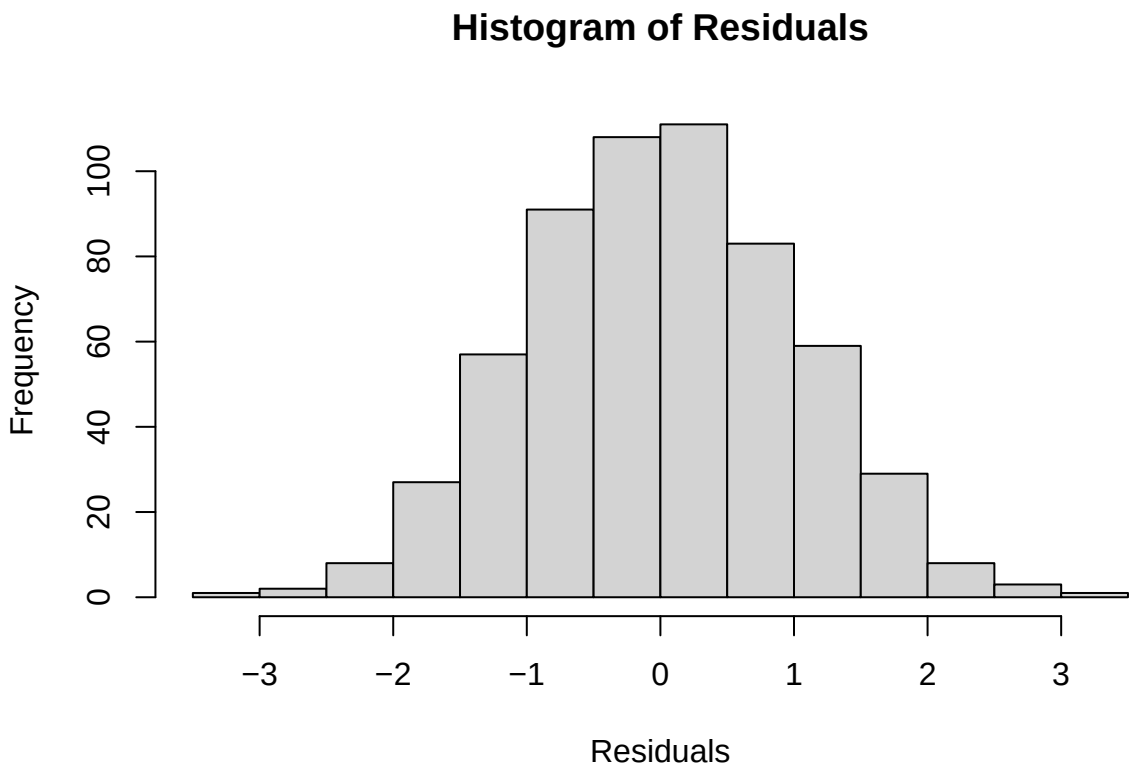

```
# Durbin-Watson Test:  Independence of the error terms
# H0 (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(step_model2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1           0.085           1.8   0.032
##  Alternative hypothesis: rho != 0
```

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

bptest(step_model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  step_model2
## BP = 7, df = 9, p-value = 0.6
```

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




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

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

```
##
##  Shapiro-Wilk normality test
##
## data:  res2_num
## W = 1, p-value = 0.8
```