Regression_Analysis_Total

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)
```

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
```

```
# The total number of data points in the dataset

nrow(df)
```

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

```
unique(df$RLK2022)
```

```
## [1] "zentral"       "sehr zentral"  "peripher"      "sehr peripher"
```

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 (outliers removed)
3. education: categorical variable
4. urban_rural_class: categorical variable, based on RLK 2022 classification
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 is selected

#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
```

```
# Independent variables: age, income, political_party, education, urban_rural, federal_state
# Dependent variables: CO2_total


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

df1
```

```
## # A tibble: 588 x 7
##      age income political_party      education           urban~1 feder~2 CO2_t~3
##    <int>  <dbl> <fct>                <fct>               <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU              (Fach-) Hochschula~ zentral Saarla~  16377.
##  2    59    800 Keine Angabe         Allgemeine oder fa~ sehr z~ Hessen   18385.
##  3    60   1750 Keine Angabe         Berufsausbildung, ~ periph~ Bayern    6335.
##  4    73   2500 SPD                  Realschulabschluss~ sehr z~ Bayern   12357.
##  5    43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin    9075.
##  6    49   2300 CDU/CSU              Berufsausbildung, ~ zentral Sachse~  12181.
##  7    57    600 CDU/CSU              Realschulabschluss~ zentral Baden-~   5701.
##  8    39   5000 SPD                  (Fach-) Hochschula~ sehr z~ Berlin   23114.
##  9    62      0 Keine Angabe         (Fach-) Hochschula~ sehr z~ Nordrh~  12601.
## 10    45   2600 Keine Angabe         Berufsausbildung, ~ sehr z~ Hessen   10875.
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: CO2_total
```

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


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
```

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

df2
```

```
## # A tibble: 588 x 7
```

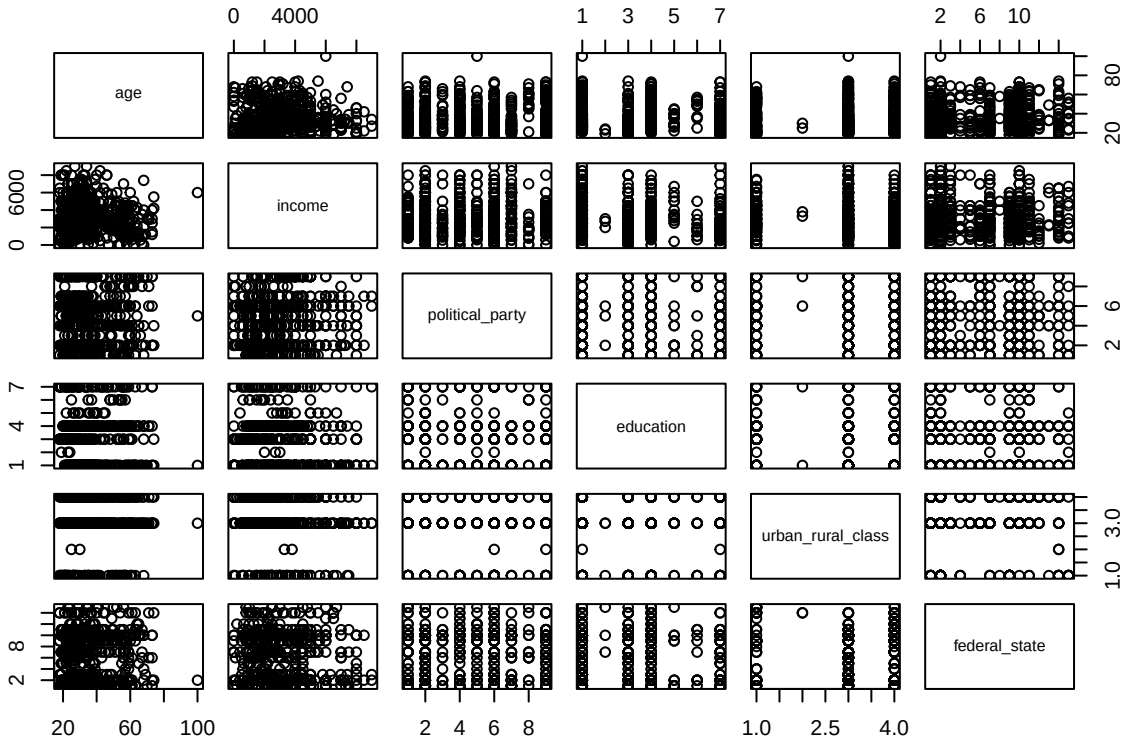
```
##      age income political_party      education           urban~1 feder~2 belie~3
##    <int>  <dbl> <fct>                <fct>               <fct>   <fct>     <dbl>
##  1    65   3000 CDU/CSU              (Fach-) Hochschula~ zentral Saarla~     -15
##  2    59    800 Keine Angabe         Allgemeine oder fa~ sehr z~ Hessen      -76
##  3    60   1750 Keine Angabe         Berufsausbildung, ~ periph~ Bayern       57
##  4    73   2500 SPD                  Realschulabschluss~ sehr z~ Bayern       -8
##  5    43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin       -1
##  6    49   2300 CDU/CSU              Berufsausbildung, ~ zentral Sachse~      13
##  7    57    600 CDU/CSU              Realschulabschluss~ zentral Baden-~      68
##  8    39   5000 SPD                  (Fach-) Hochschula~ sehr z~ Berlin      -66
##  9    62      0 Keine Angabe         (Fach-) Hochschula~ sehr z~ Nordrh~     -16
## 10    45   2600 Keine Angabe         Berufsausbildung, ~ sehr z~ Hessen       -2
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: belief_diff_total
```

I. Exploratory Data Analysis

Check the Jupytor notebook: EDA_scatter_plot_actual_belief

II. Multivariate Regression: CO2 total

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

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



1. Modeling

We can see that there is some level of positive correlation between age and income variables.

```
# Checking the number of variables to find the ones with the highest number

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 reference levels according to the values with the highest frequency

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, non-scaled dataset

model1 <- lm(CO2_total ~ age + income + political_party + education +  urban_rural_class + federal_stat

summary(model1)
```

```
##
## Call:
## lm(formula = CO2_total ~ age + income + political_party + education +
##     urban_rural_class + federal_state, data = df1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -19734  -5327  -1957   2038 160042
##
## Coefficients:
##                                                                  Estimate
## (Intercept)                                                    8145.3827
```

```
## age                                                                          -41.7031
## income                                                                         1.5334
## political_partyAfD                                                          2200.7114
## political_partyBündnis Sarah Wagenknecht                                    2606.0394
## political_partyCDU/CSU                                                      9305.5682
## political_partyDie Linke                                                    1766.9693
## political_partyEiner anderen Partei                                          264.3604
## political_partyFDP                                                          2335.7555
## political_partyKeine Angabe                                                 2781.9705
## political_partySPD                                                          3905.7658
## education(Noch) kein Abschluss                                             -4154.0124
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -1992.9472
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule       -2847.6093
## educationDoktorgrad oder Habilitation                                      -4134.1188
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  -3880.7668
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss   -597.9738
## urban_rural_classperipher                                                  -1785.3020
## urban_rural_classsehr peripher                                             -2220.2331
## urban_rural_classzentral                                                   -1174.2346
## federal_stateBaden-Württemberg                                              -287.5748
## federal_stateBayern                                                         2788.0825
## federal_stateBerlin                                                         3551.3652
## federal_stateBrandenburg                                                   -2379.6996
## federal_stateBremen                                                        -1962.8886
## federal_stateHamburg                                                       -1006.8613
## federal_stateHessen                                                         2330.9722
## federal_stateMecklenburg-Vorpommern                                        -8763.3458
## federal_stateNiedersachsen                                                   574.0948
## federal_stateRheinland-Pfalz                                                5150.2829
## federal_stateSaarland                                                        818.4358
## federal_stateSachsen-Anhalt                                                -2952.0208
## federal_stateSchleswig-Holstein                                             2521.9645
## federal_stateThüringen                                                      3084.7399
##                                                                           Std. Error
## (Intercept)                                                                 2881.1435
## age                                                                           51.0074
## income                                                                         0.3409
## political_partyAfD                                                          2440.4702
## political_partyBündnis Sarah Wagenknecht                                    3420.6309
## political_partyCDU/CSU                                                      2191.9153
## political_partyDie Linke                                                    2641.8962
## political_partyEiner anderen Partei                                         1960.9194
## political_partyFDP                                                          2538.5727
## political_partyKeine Angabe                                                 4397.1314
## political_partySPD                                                          2232.0688
## education(Noch) kein Abschluss                                              8929.9934
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  1735.8363
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule        1770.8122
## educationDoktorgrad oder Habilitation                                       4328.6978
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   4941.1645
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss  2251.9528
## urban_rural_classperipher                                                   2269.3316
## urban_rural_classsehr peripher                                             11234.3848
## urban_rural_classzentral                                                    1663.7398
```

```
## federal_stateBaden-Württemberg                                                  2161.2461
## federal_stateBayern                                                             2218.9909
## federal_stateBerlin                                                             2695.6990
## federal_stateBrandenburg                                                        5662.0156
## federal_stateBremen                                                             4129.8658
## federal_stateHamburg                                                            3360.8760
## federal_stateHessen                                                             2578.5907
## federal_stateMecklenburg-Vorpommern                                            10823.7714
## federal_stateNiedersachsen                                                      2623.2786
## federal_stateRheinland-Pfalz                                                    3235.5130
## federal_stateSaarland                                                           5080.5777
## federal_stateSachsen-Anhalt                                                     7831.6200
## federal_stateSchleswig-Holstein                                                 3784.3740
## federal_stateThüringen                                                          5736.0907
##                                                                                t value
## (Intercept)                                                                      2.827
## age                                                                             -0.818
## income                                                                           4.499
## political_partyAfD                                                               0.902
## political_partyBündnis Sarah Wagenknecht                                         0.762
## political_partyCDU/CSU                                                           4.245
## political_partyDie Linke                                                         0.669
## political_partyEiner anderen Partei                                              0.135
## political_partyFDP                                                               0.920
## political_partyKeine Angabe                                                      0.633
## political_partySPD                                                               1.750
## education(Noch) kein Abschluss                                                  -0.465
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  -1.148
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule            -1.608
## educationDoktorgrad oder Habilitation                                           -0.955
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  -0.785
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss      -0.266
## urban_rural_classperipher                                                       -0.787
## urban_rural_classsehr peripher                                                  -0.198
## urban_rural_classzentral                                                        -0.706
## federal_stateBaden-Württemberg                                                  -0.133
## federal_stateBayern                                                              1.256
## federal_stateBerlin                                                              1.317
## federal_stateBrandenburg                                                        -0.420
## federal_stateBremen                                                             -0.475
## federal_stateHamburg                                                            -0.300
## federal_stateHessen                                                              0.904
## federal_stateMecklenburg-Vorpommern                                             -0.810
## federal_stateNiedersachsen                                                       0.219
## federal_stateRheinland-Pfalz                                                     1.592
## federal_stateSaarland                                                            0.161
## federal_stateSachsen-Anhalt                                                     -0.377
## federal_stateSchleswig-Holstein                                                  0.666
## federal_stateThüringen                                                           0.538
##                                                                                Pr(>|t|)
## (Intercept)                                                                     0.00487
## age                                                                             0.41394
## income                                                                         8.34e-06
## political_partyAfD                                                              0.36758
```

```
## political_partyBündnis Sarah Wagenknecht                                        0.44647
## political_partyCDU/CSU                                                          2.56e-05
## political_partyDie Linke                                                         0.50388
## political_partyEiner anderen Partei                                              0.89281
## political_partyFDP                                                               0.35792
## political_partyKeine Angabe                                                      0.52720
## political_partySPD                                                               0.08070
## education(Noch) kein Abschluss                                                   0.64199
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  0.25141
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule             0.10839
## educationDoktorgrad oder Habilitation                                            0.33997
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss  0.43256
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss       0.79070
## urban_rural_classperipher                                                        0.43179
## urban_rural_classsehr peripher                                                   0.84341
## urban_rural_classzentral                                                         0.48062
## federal_stateBaden-Württemberg                                                   0.89419
## federal_stateBayern                                                              0.20948
## federal_stateBerlin                                                              0.18824
## federal_stateBrandenburg                                                         0.67444
## federal_stateBremen                                                              0.63477
## federal_stateHamburg                                                             0.76461
## federal_stateHessen                                                              0.36640
## federal_stateMecklenburg-Vorpommern                                              0.41850
## federal_stateNiedersachsen                                                       0.82685
## federal_stateRheinland-Pfalz                                                     0.11200
## federal_stateSaarland                                                            0.87208
## federal_stateSachsen-Anhalt                                                      0.70637
## federal_stateSchleswig-Holstein                                                  0.50542
## federal_stateThüringen                                                           0.59095
##
## (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: 14910 on 554 degrees of freedom
## Multiple R-squared:  0.1083, Adjusted R-squared:  0.05516
## F-statistic: 2.038 on 33 and 554 DF,  p-value: 0.0006897
```

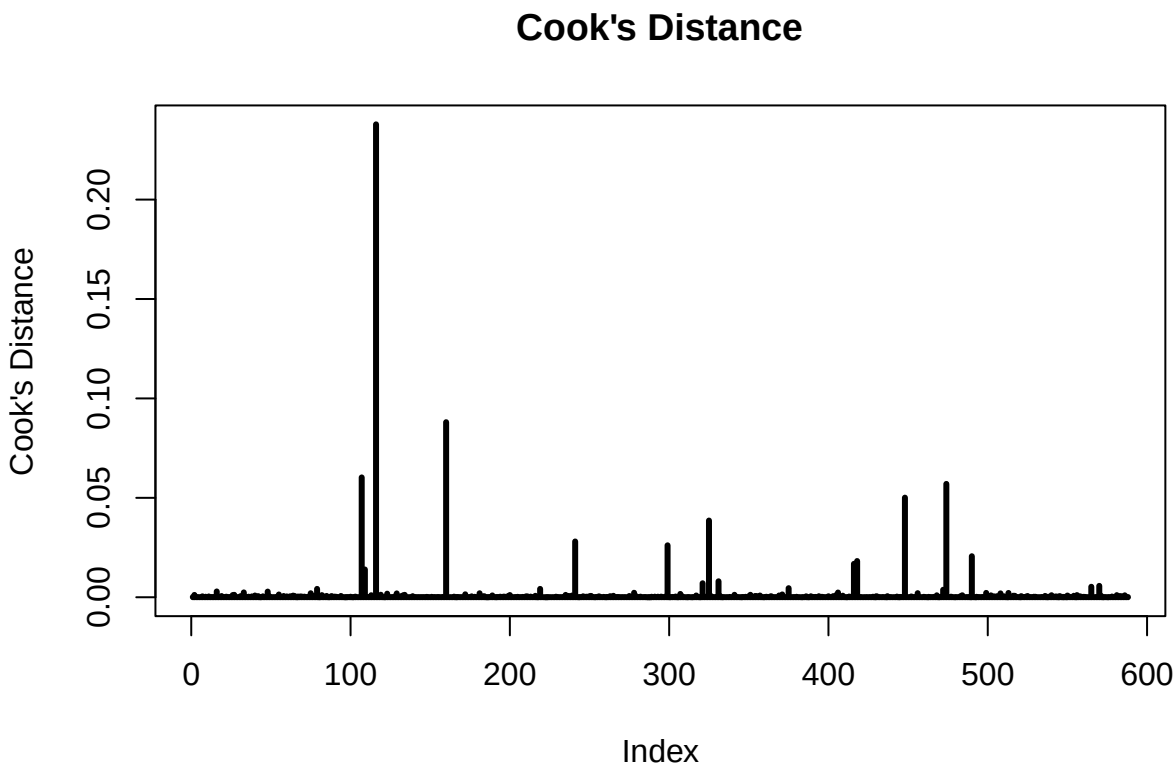
```
# Checking the VIFs for multicollinearity

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 the outliers: estimate of the influence of data point; summary of how much a regression mode

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
```

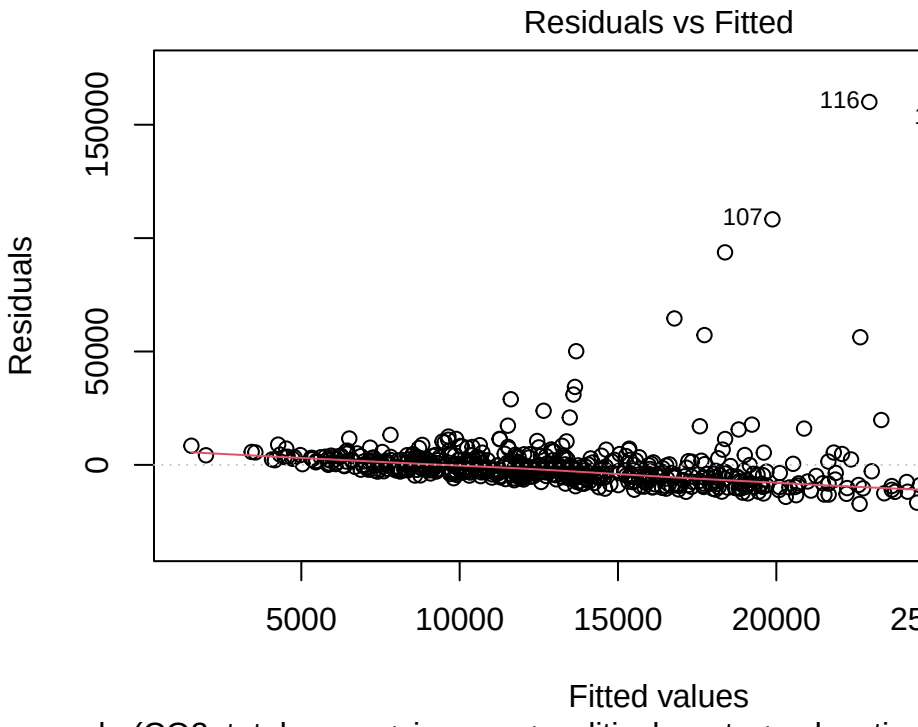
```
##         107         109         116         160         241         299
## 0.060274420 0.013990806 0.237846858 0.088047668 0.028049196 0.026121855
##         321         325         331         375         416         418
## 0.007034491 0.038578640 0.008024303 0.004568137 0.016814168 0.018226988
##         448         474         490         565         570
## 0.050100158 0.057018122 0.020630810 0.005273394 0.005737906
```

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


df1[influential_vector, ]
```

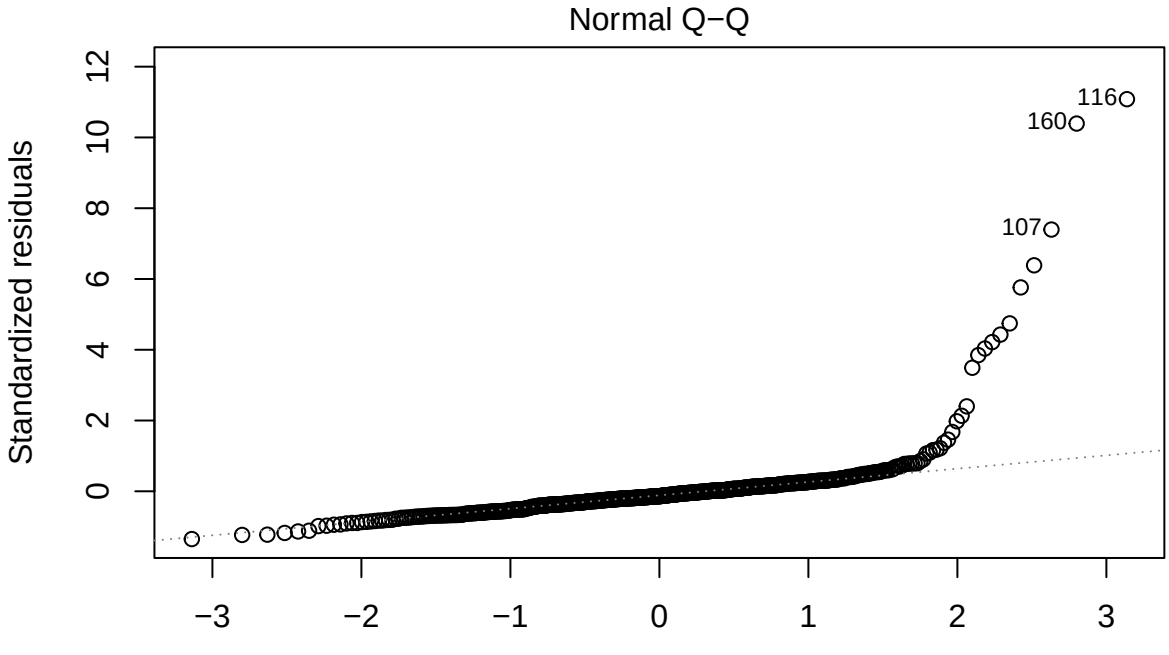
```
## # A tibble: 17 x 7
##      age income political_party           education      urban~1 feder~2 CO2_t~3
##    <int>  <dbl> <fct>                     <fct>          <fct>   <fct>     <dbl>
##  1    32   7000 Bündnis 90/Die Grünen     (Fach-) Hochs~ sehr z~ Hessen  128151.
##  2    22    600 FDP                       Allgemeine od~ sehr z~ Rheinl~  48023.
##  3    23   2000 CDU/CSU                   Realschulabsc~ zentral Rheinl~ 182979.
##  4    29   4500 CDU/CSU                   (Fach-) Hochs~ sehr z~ Bayern  178779.
##  5    21   5000 Bündnis 90/Die Grünen     Allgemeine od~ periph~ Schles~  63804.
##  6    43   3500 SPD                       (Fach-) Hochs~ zentral Hessen   81349.
```

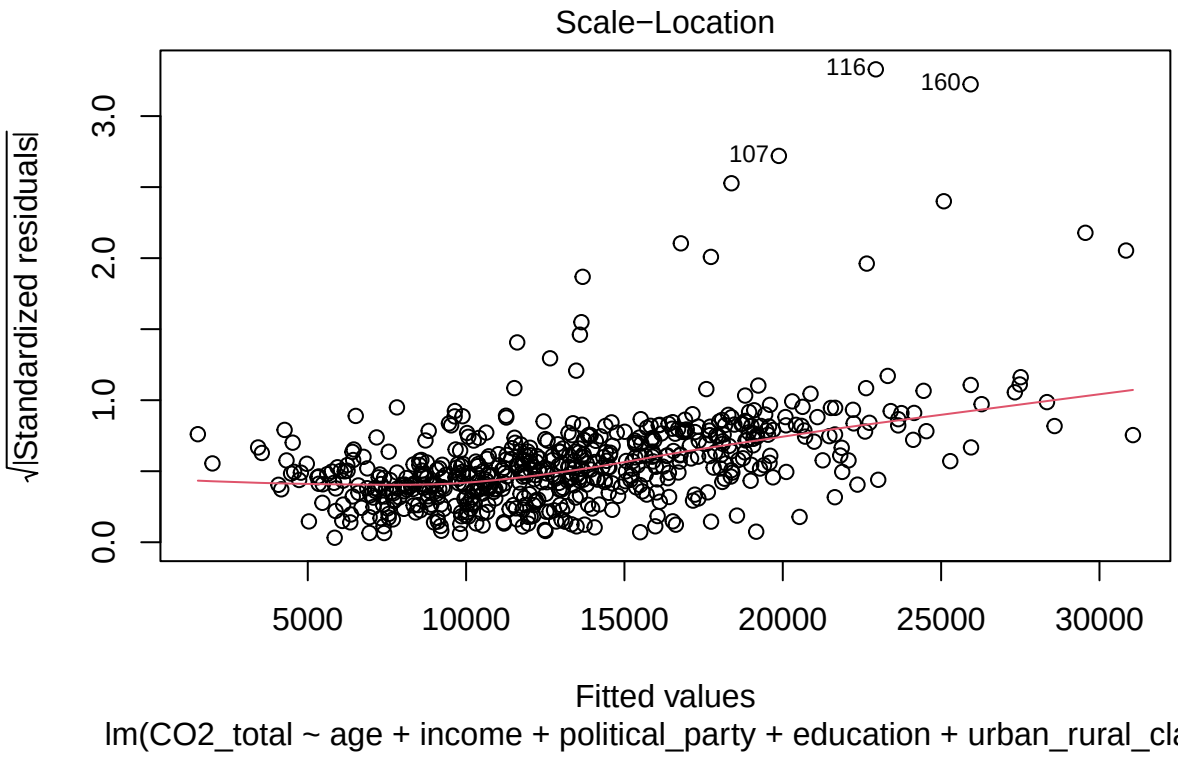
```
##  7    24   1200 SPD                       Berufsausbild~ sehr z~ Berlin   44615.
##  8    26   1500 CDU/CSU                   (Fach-) Hochs~ sehr z~ Baden-~ 112102.
##  9    59   3500 Bündnis Sarah Wagenknecht (Fach-) Hochs~ sehr z~ Hamburg  36516.
## 10    38   4000 Die Linke                 Berufsausbild~ sehr z~ Nordrh~  40544.
## 11    31   5000 CDU/CSU                   (Fach-) Hochs~ zentral Nordrh~  78943.
## 12    40   8000 CDU/CSU                   (Fach-) Hochs~ sehr z~ Bayern   92635.
## 13   100   6000 Die Linke                 (Fach-) Hochs~ sehr z~ Bayern   74965.
## 14    19   8000 SPD                       Allgemeine od~ sehr z~ Berlin  108589.
## 15    34   7000 CDU/CSU                   (Fach-) Hochs~ sehr z~ Bayern   99220.
## 16    26   2500 Bündnis Sarah Wagenknecht Realschulabsc~ sehr z~ Nieder~  34366.
## 17    25   8000 Bündnis Sarah Wagenknecht (Fach-) Hochs~ sehr z~ Brande~   6984.
## # ... with abbreviated variable names 1: urban_rural_class, 2: federal_state,
## #   3: CO2_total
```

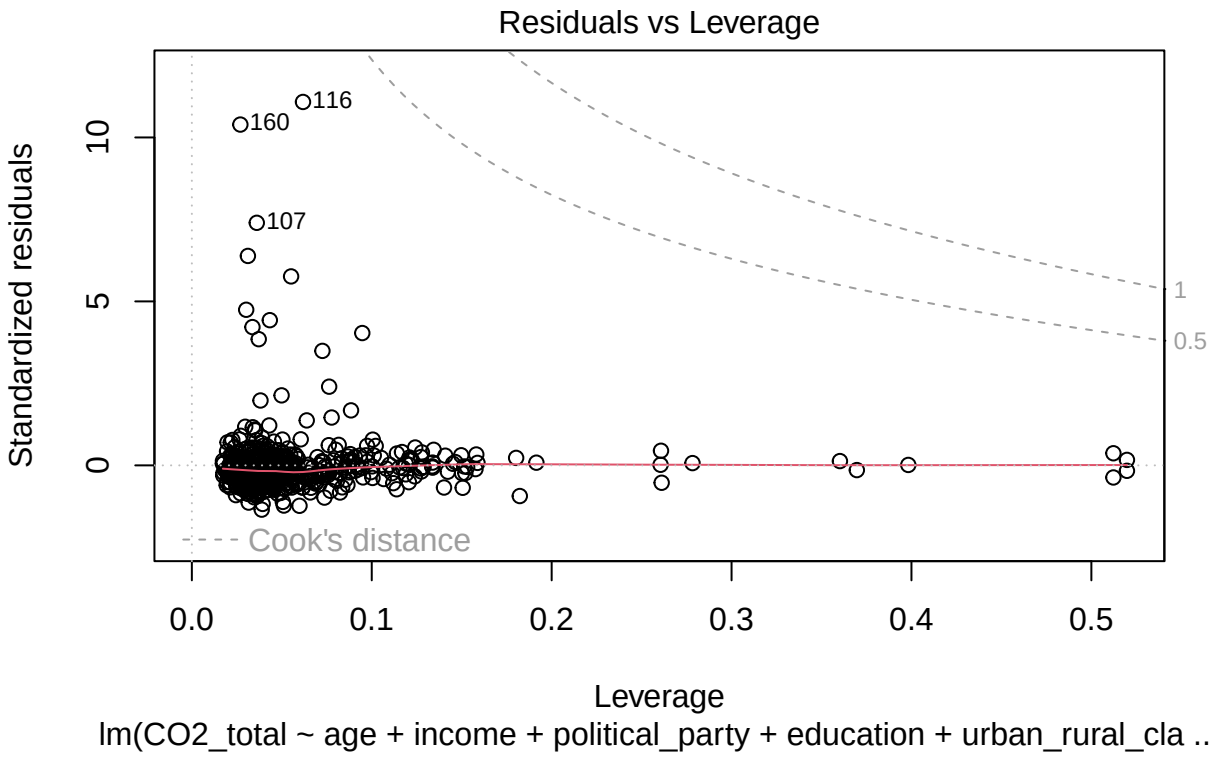
```
plot(model1)
```



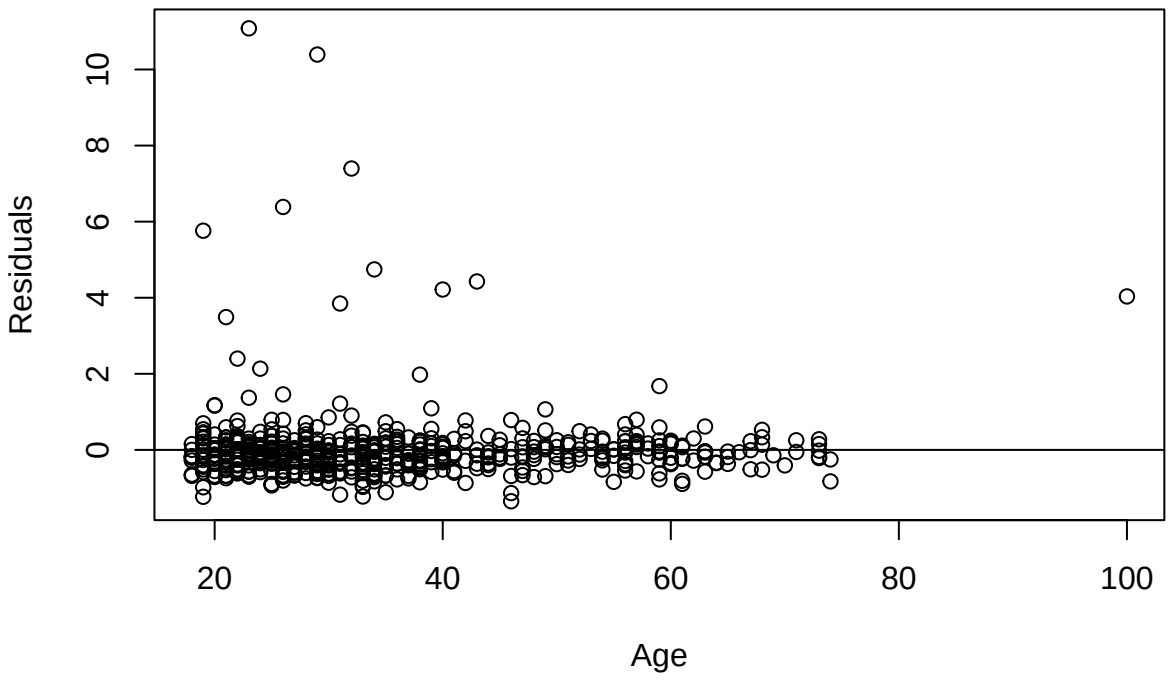
2. Assumptions check in the residuals



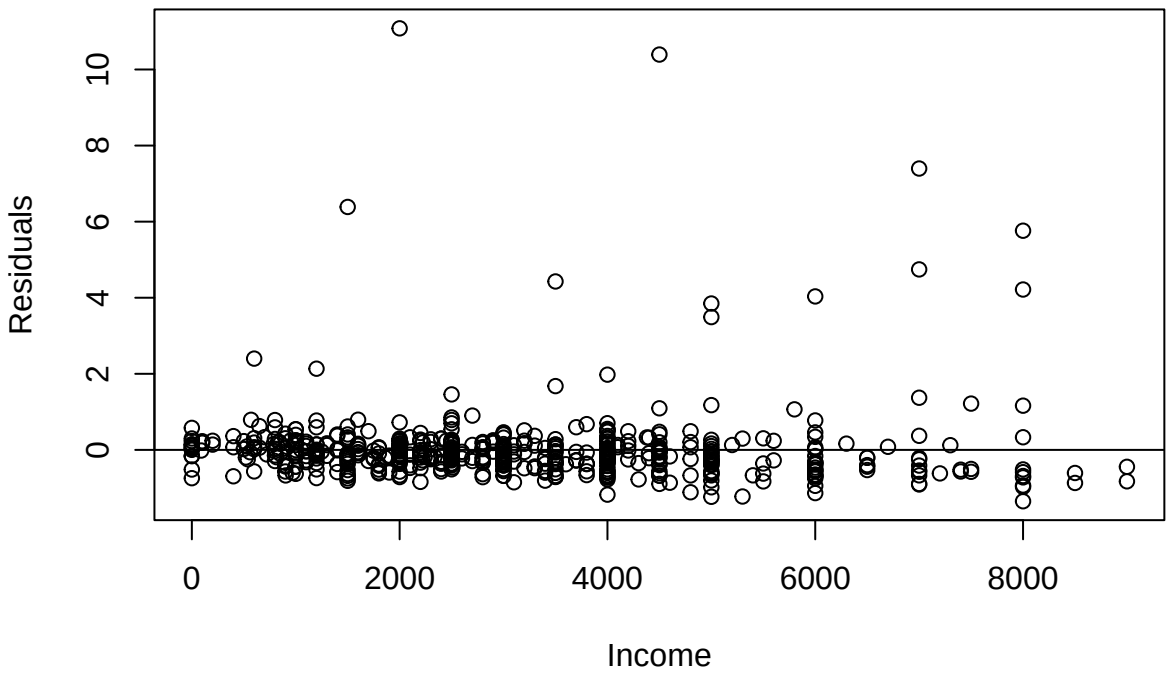

Scale−Location
116
160
107
√|Standardized residuals|
0.0
1.0
2.0
3.0
5000
10000
15000
20000
25000
30000
Fitted values
lm(CO2_total ~ age + income + political_party + education + urban_rural_cla ...



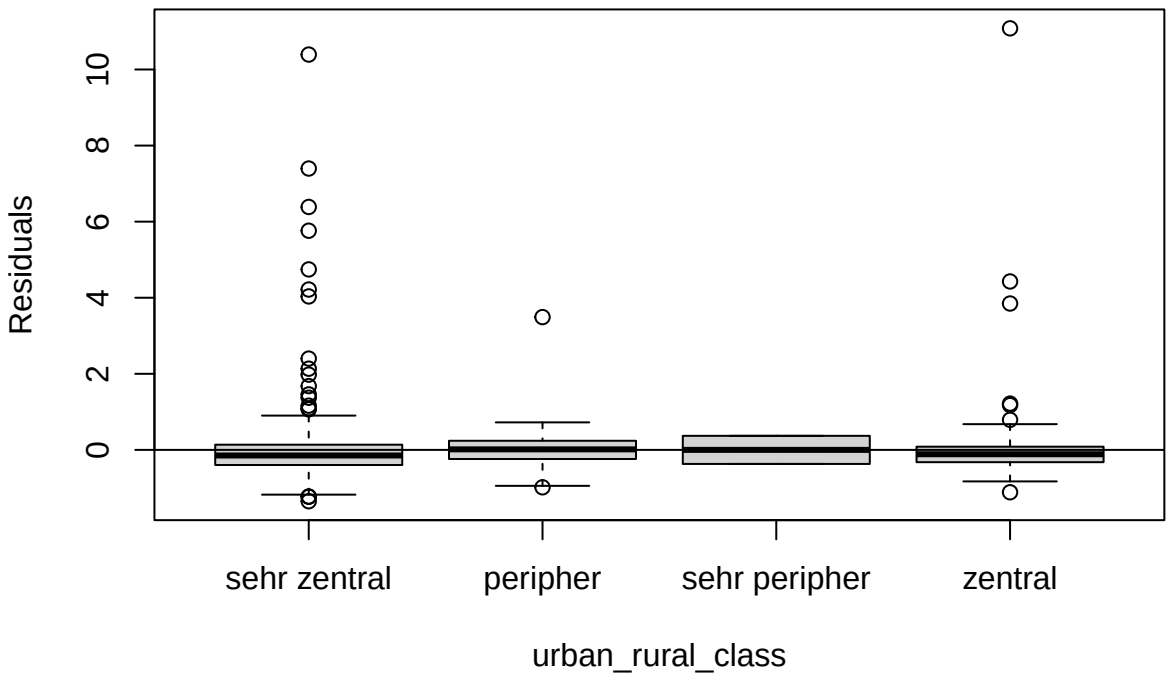

```
res1 = stdres(model1) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption: violated

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



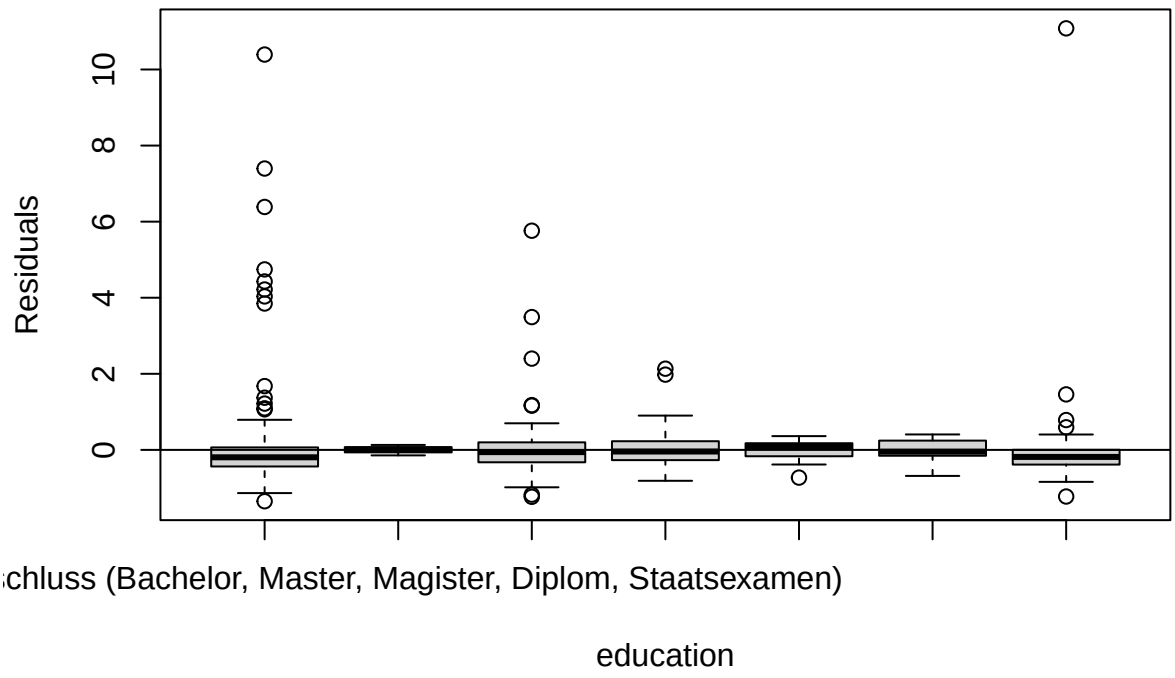

```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



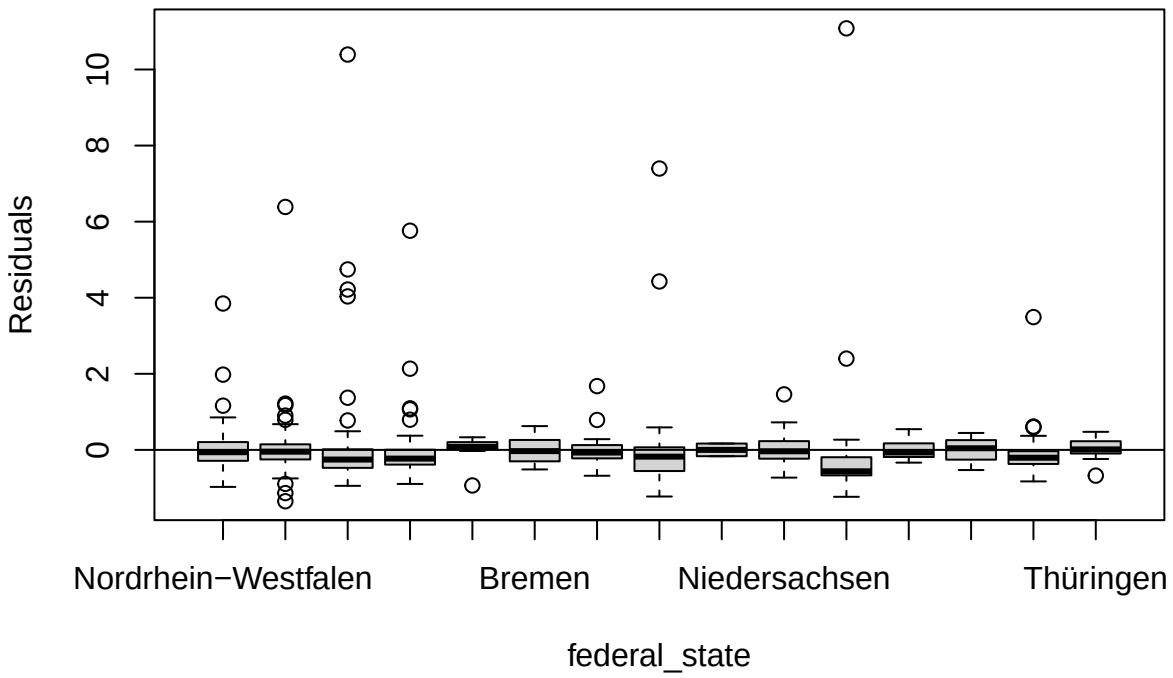

```
plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
abline(h = 0)
```



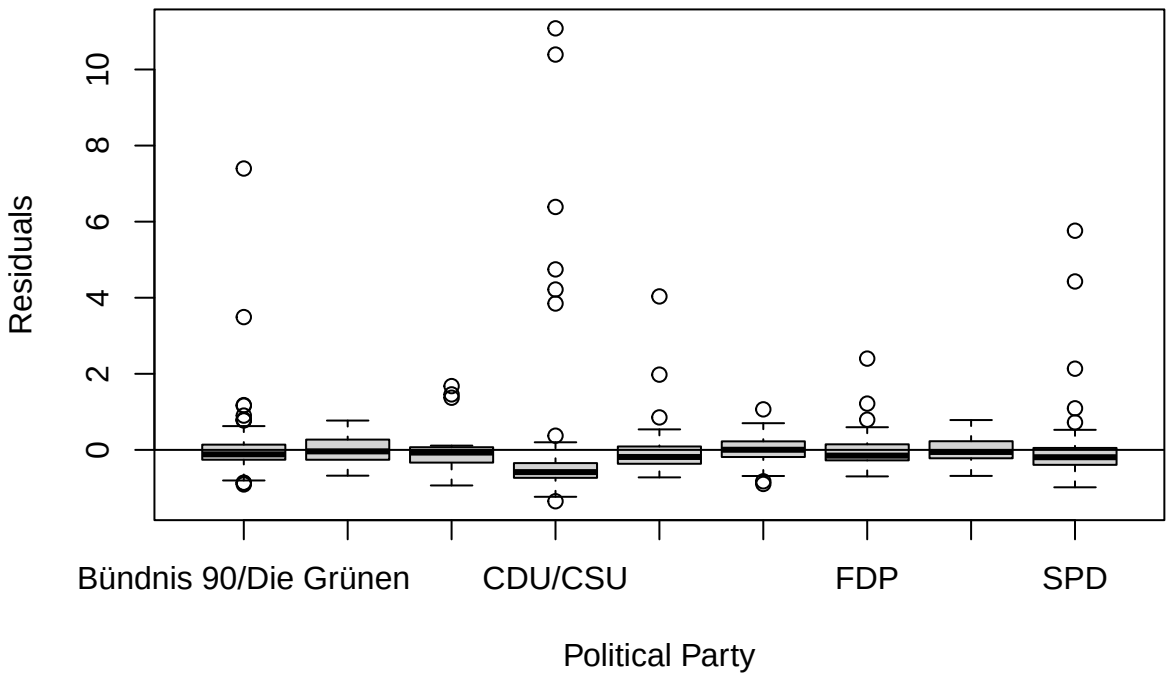

```
plot(df1$education, res1, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



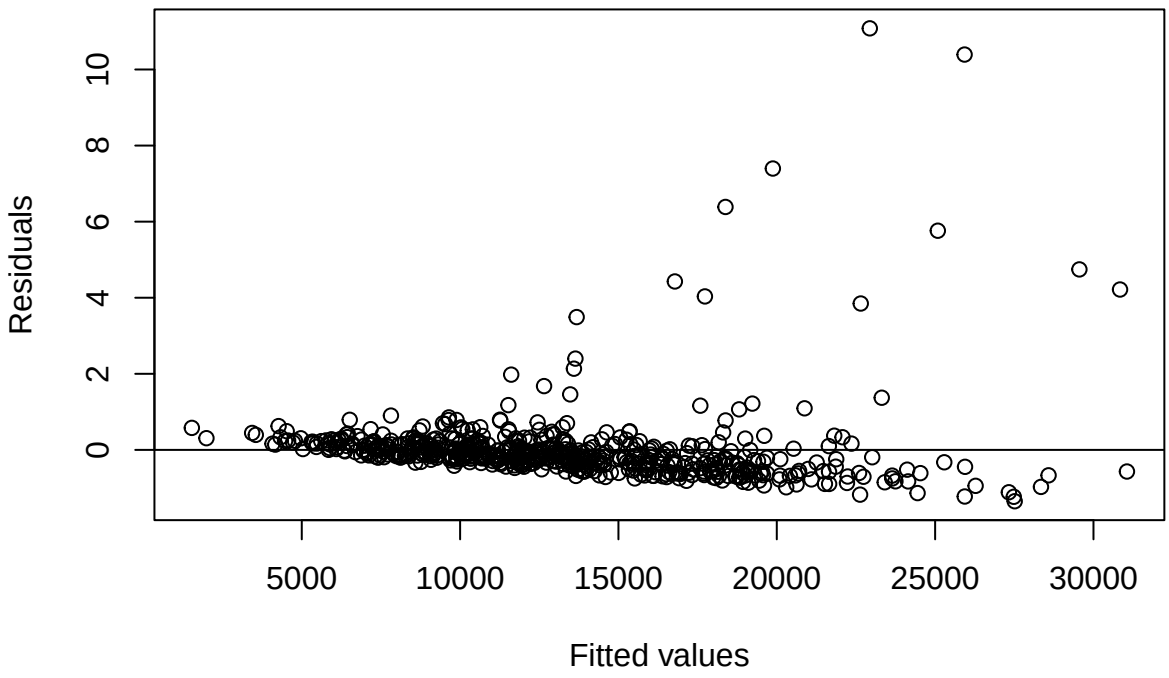

```
plot(df1$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```



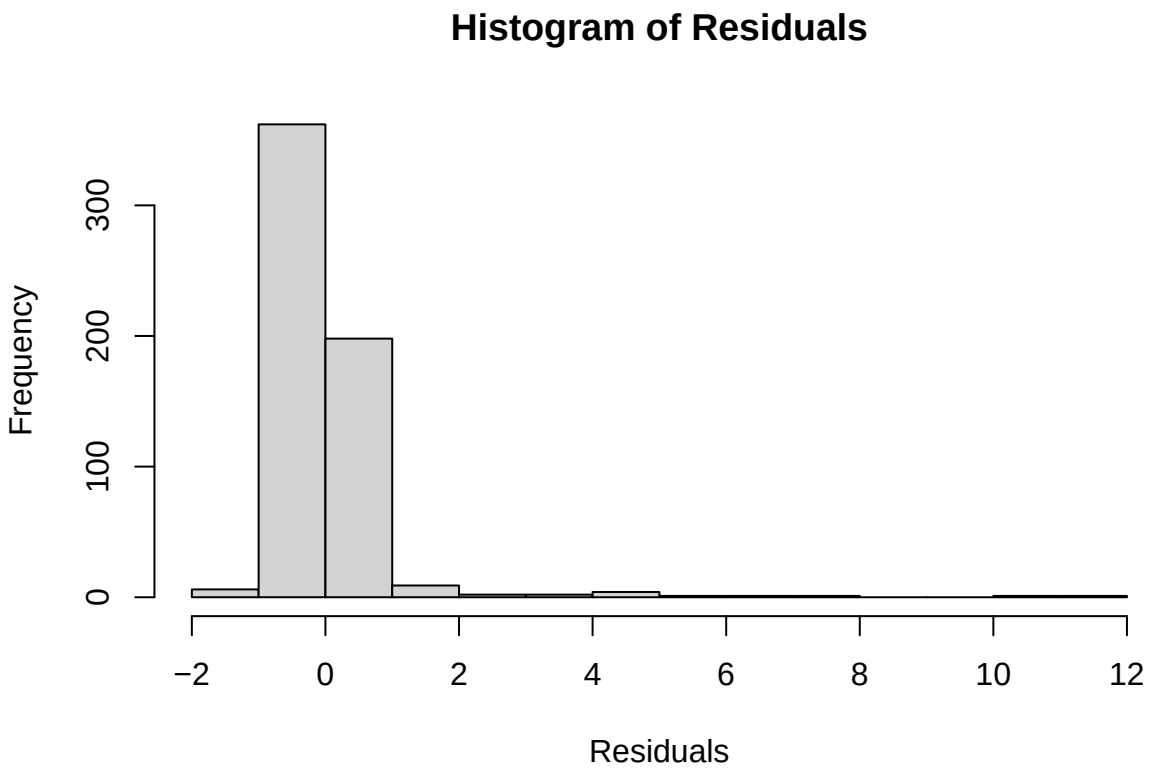

```
plot(df1$political_party, res1, xlab = "Political Party", ylab = "Residuals")
abline(h = 0)
```




```
# Constant variance and independent error term assumption: violated

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




```
# Normality assumption : violated

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




```
### Backward regression using AIC: starting with all of the variables
### Final model: CO2_total ~ income + political_party

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

3. Variable Selection

```
## Start:  AIC=11334.08
## CO2_total ~ age + income + political_party + education + urban_rural_class +
##     federal_state
##
##                     Df  Sum of Sq        RSS   AIC
## - federal_state     14 2050088689 1.2521e+11 11316
## - education          6  907418903 1.2406e+11 11326
## - urban_rural_class  3  184396781 1.2334e+11 11329
## - age                1  148600836 1.2331e+11 11333
## <none>                             1.2316e+11 11334
## - political_party    8 4827118892 1.2798e+11 11341
## - income             1 4499013255 1.2766e+11 11353
##
## Step:  AIC=11315.79
## CO2_total ~ age + income + political_party + education + urban_rural_class
##
```

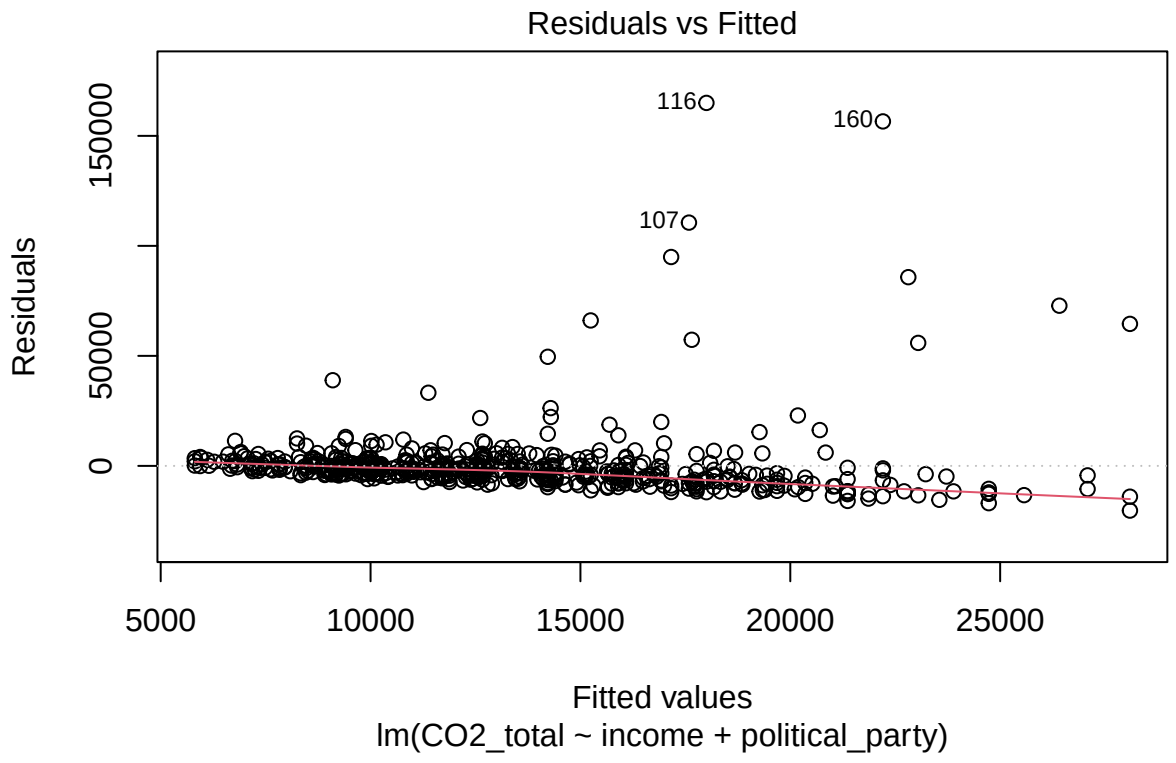
```
##                       Df  Sum of Sq        RSS   AIC
## - education            6 1103377391 1.2631e+11 11309
## - urban_rural_class    3  135423195 1.2534e+11 11310
## - age                  1  227009624 1.2543e+11 11315
## <none>                              1.2521e+11 11316
## - political_party      8 5027931674 1.3024e+11 11323
## - income               1 5186882912 1.3039e+11 11338
##
## Step:  AIC=11308.95
## CO2_total ~ age + income + political_party + urban_rural_class
##
##                       Df  Sum of Sq        RSS   AIC
## - urban_rural_class    3  256526676 1.2657e+11 11304
## - age                  1  273276668 1.2658e+11 11308
## <none>                              1.2631e+11 11309
## - political_party      8 4953451234 1.3126e+11 11316
## - income               1 5659279571 1.3197e+11 11333
##
## Step:  AIC=11304.14
## CO2_total ~ age + income + political_party
##
##                     Df  Sum of Sq        RSS   AIC
## - age                1  311187508 1.2688e+11 11304
## <none>                             1.2657e+11 11304
## - political_party    8 4854420395 1.3142e+11 11310
## - income             1 5795022031 1.3236e+11 11328
##
## Step:  AIC=11303.58
## CO2_total ~ income + political_party
##
##                     Df  Sum of Sq        RSS   AIC
## <none>                             1.2688e+11 11304
## - political_party    8 4626486717 1.3151e+11 11309
## - income             1 5791449738 1.3267e+11 11328
```

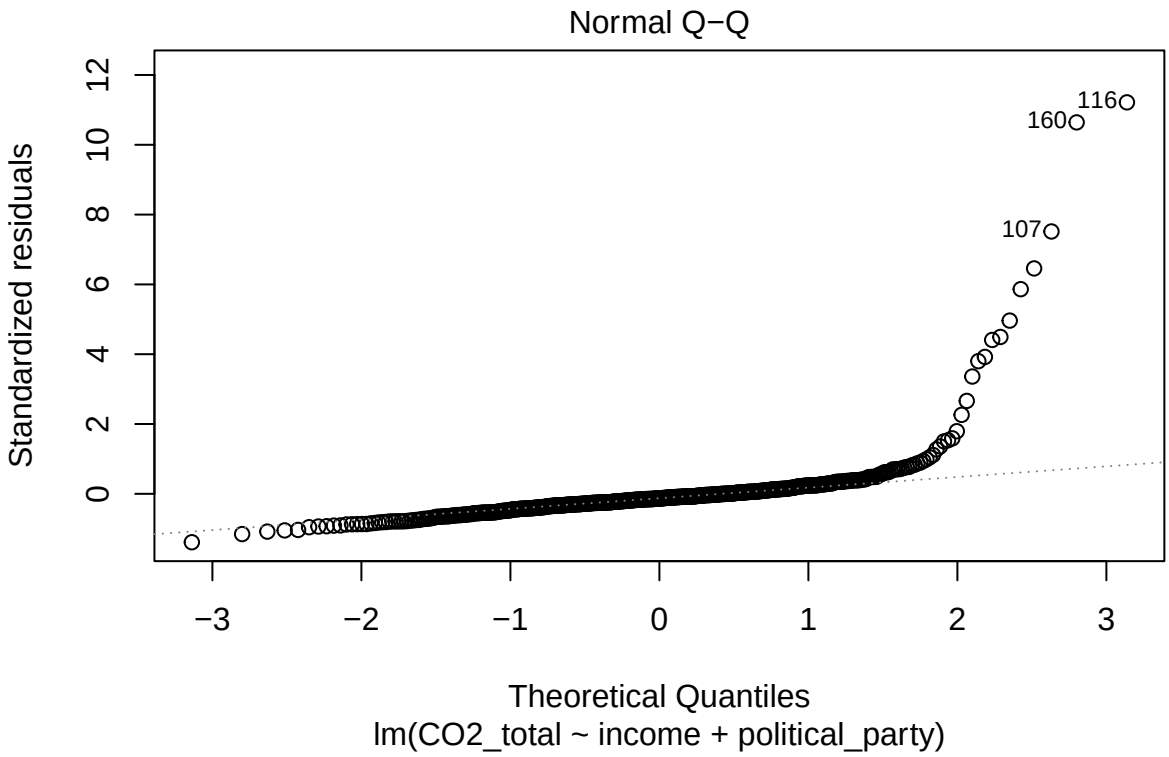
```
summary(step_model1)
```

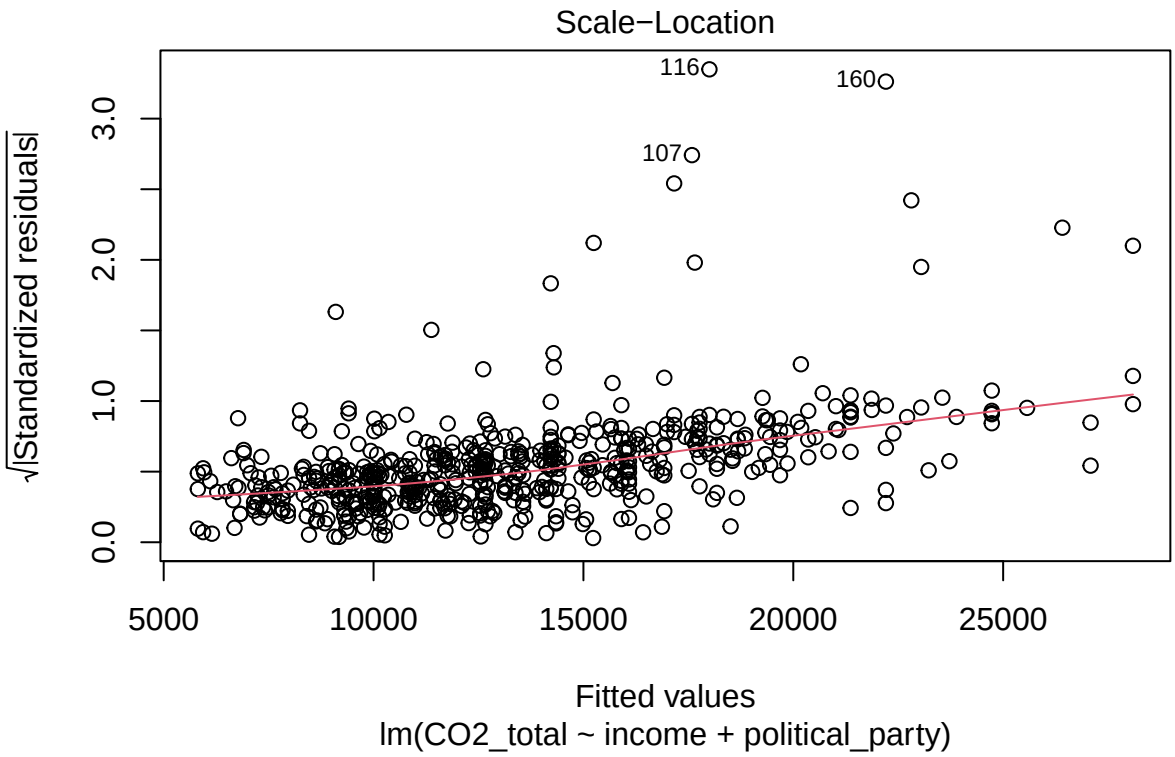
```
##
## Call:
## lm(formula = CO2_total ~ income + political_party, data = df1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -20314  -4874  -1969   1160 164977
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              5813.5677  1616.5023   3.596  0.00035
## income                                      1.6817     0.3274   5.136 3.83e-07
## political_partyAfD                       1575.3245  2308.5965   0.682  0.49528
## political_partyBündnis Sarah Wagenknecht 2595.5486  3333.1883   0.779  0.43648
## political_partyCDU/CSU                   8825.2105  2114.4715   4.174 3.46e-05
## political_partyDie Linke                 1749.2279  2555.2437   0.685  0.49389
## political_partyEiner anderen Partei       126.0188  1874.2080   0.067  0.94642
```

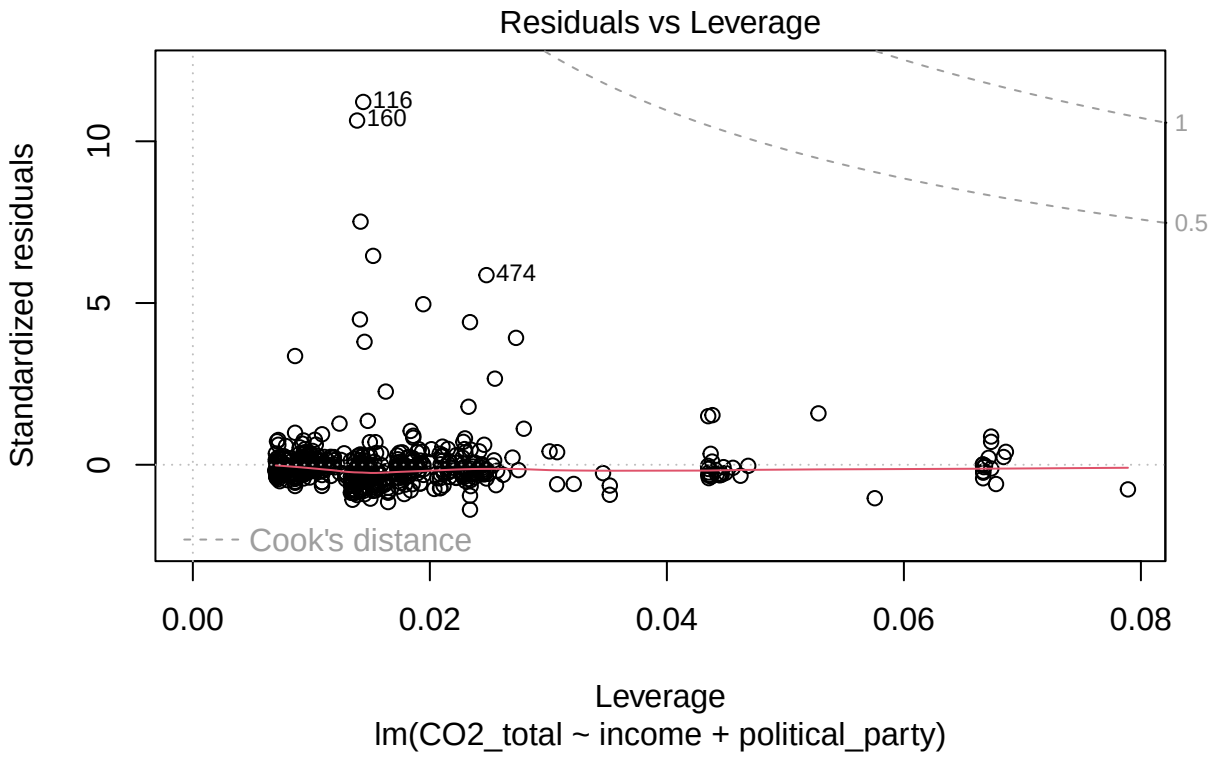
```
## political_partyFDP                          2276.4377  2477.2698   0.919  0.35852
## political_partyKeine Angabe                 1089.5258  4039.5542   0.270  0.78748
## political_partySPD                          3544.9519  2151.5530   1.648  0.09997
##
## (Intercept)                                 ***
## income                                      ***
## 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 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14820 on 578 degrees of freedom
## Multiple R-squared:  0.08133,    Adjusted R-squared:  0.06703
## F-statistic: 5.686 on 9 and 578 DF,  p-value: 1.442e-07
```

```
plot(step_model1)
```

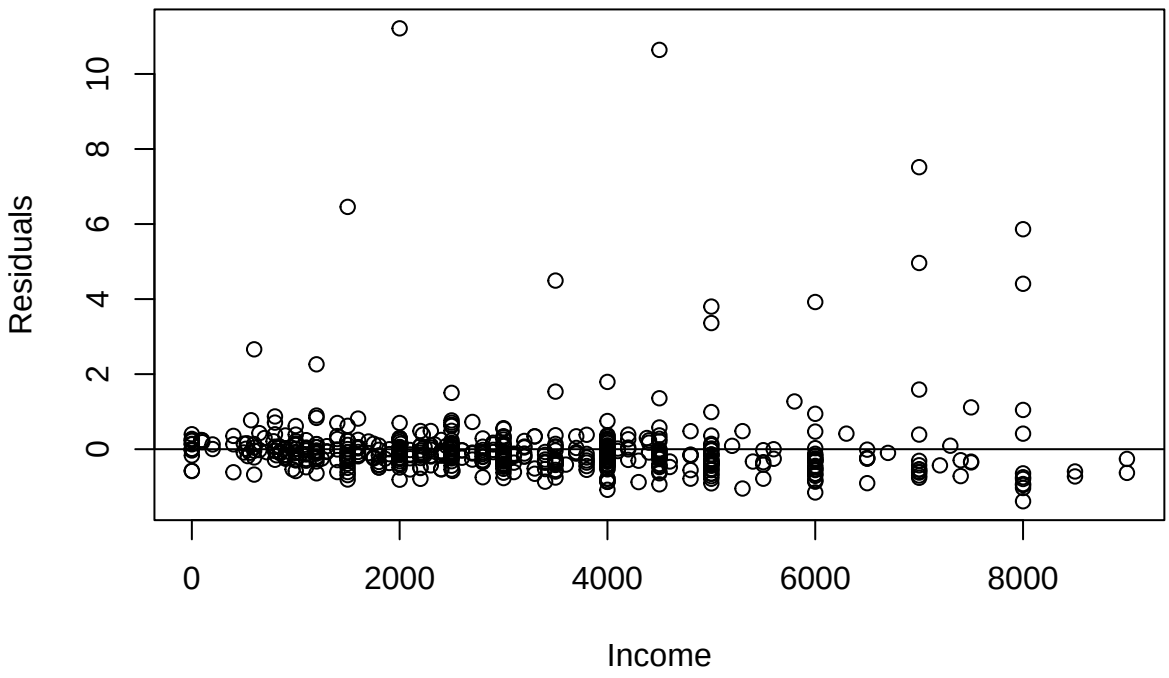



Normal Q−Q
Standardized residuals
0
2
4
6
8
10
12
116
160
107
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(CO2_total ~ income + political_party)

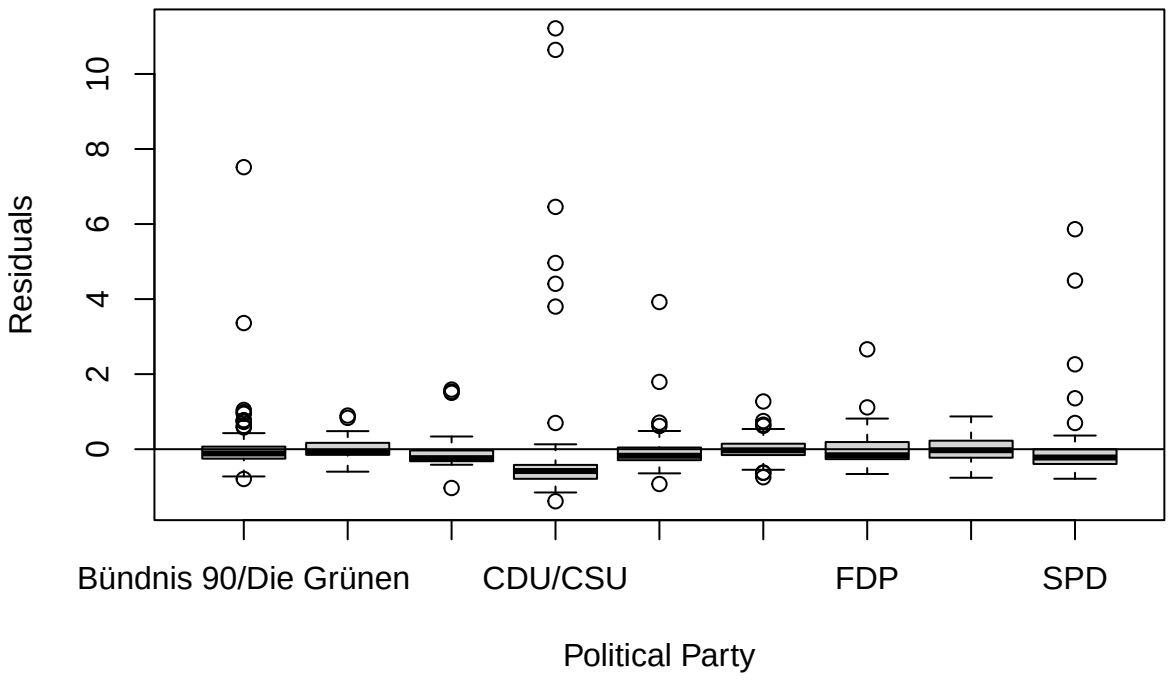

Scale−Location
116
160
107
√|Standardized residuals|
0.0
1.0
2.0
3.0
5000
10000
15000
20000
25000
Fitted values
lm(CO2_total ~ income + political_party)



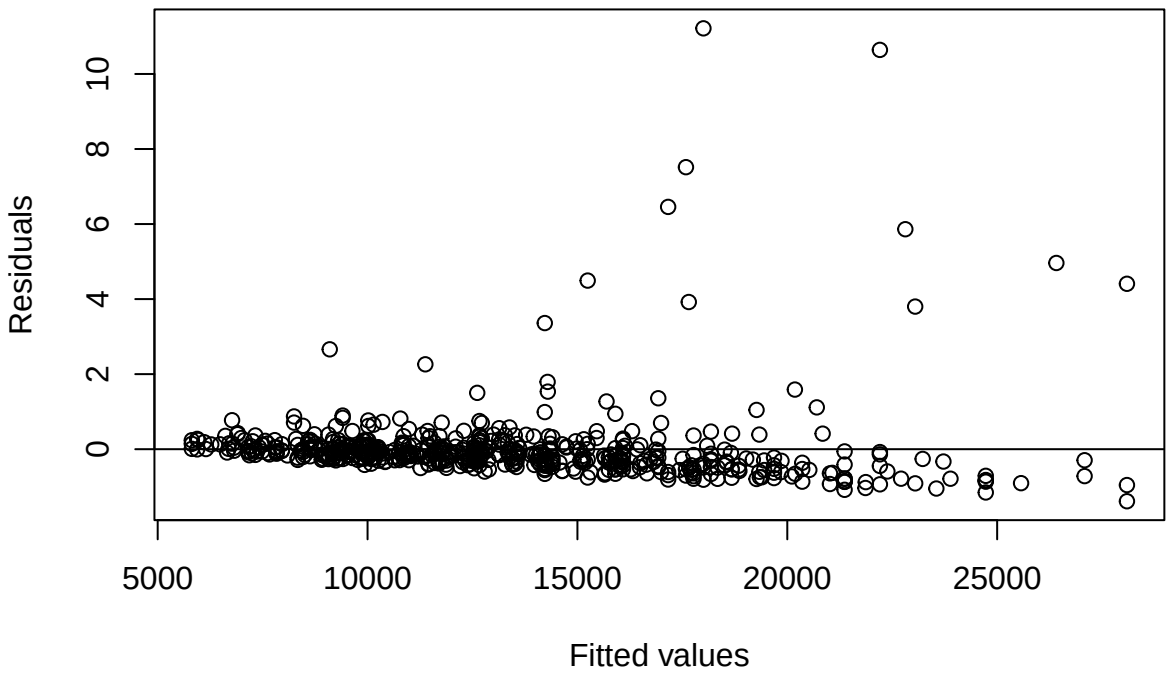

```r
res1 = stdres(step_model1) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption: violated

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

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



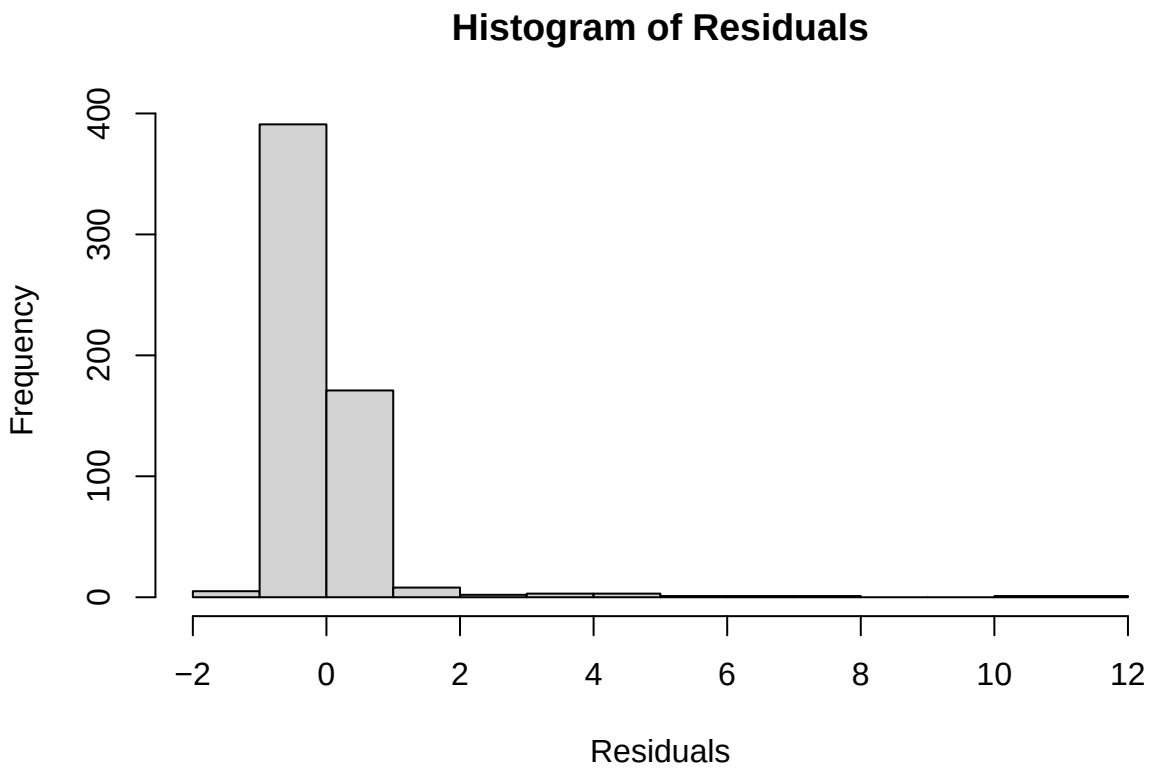

```
#plot(df1_scaled$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
#abline(h = 0)

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

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

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



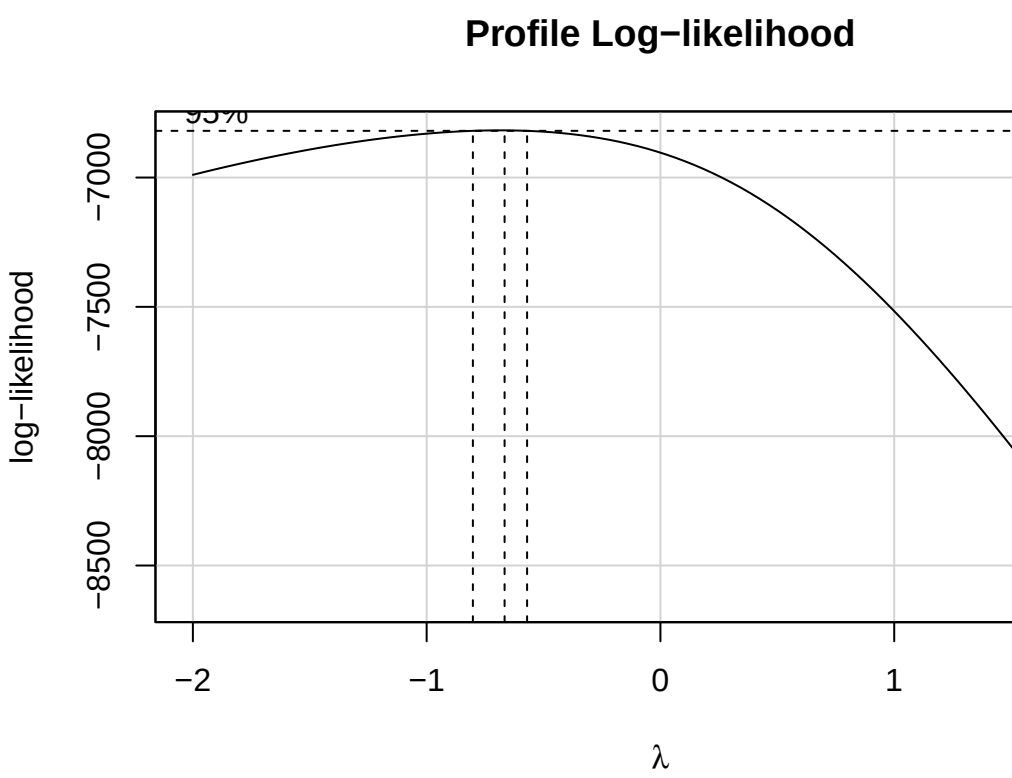

```
# Constant variance and independent error term assumption: violated

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




```
# Normality assumption: violated

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




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



4. Improving the regression fit

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

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

FINAL MODEL Here is the final model for CO2 total and the selected independent variables:

```
# transform the y variables y = y^(-0.5)

options(scipen = 0, digits=2)

model1_trans <- lm(1/sqrt(CO2_total) ~ income + political_party, data = df1)

summary(model1_trans)
```

```
##
## Call:
## lm(formula = 1/sqrt(CO2_total) ~ income + political_party, data = df1)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.007540 -0.001150  0.000046  0.001213  0.006271
##
## Coefficients:
```

```
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              1.18e-02   2.18e-04   54.46  < 2e-16
## income                                  -4.13e-07   4.41e-08   -9.36  < 2e-16
## political_partyAfD                      -1.13e-03   3.11e-04   -3.65  0.00029
## political_partyBündnis Sarah Wagenknecht -6.18e-04   4.48e-04   -1.38  0.16851
## political_partyCDU/CSU                  -1.14e-03   2.85e-04   -4.02  6.6e-05
## political_partyDie Linke                -4.52e-04   3.44e-04   -1.31  0.18944
## political_partyEiner anderen Partei     -6.87e-04   2.52e-04   -2.72  0.00663
## political_partyFDP                      -1.18e-03   3.33e-04   -3.53  0.00045
## political_partyKeine Angabe             -4.09e-04   5.44e-04   -0.75  0.45240
## political_partySPD                      -1.10e-03   2.89e-04   -3.79  0.00016
##
## (Intercept)                             ***
## income                                  ***
## 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 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002 on 578 degrees of freedom
## Multiple R-squared:  0.179,  Adjusted R-squared:  0.166
## F-statistic:   14 on 9 and 578 DF,  p-value: <2e-16
```

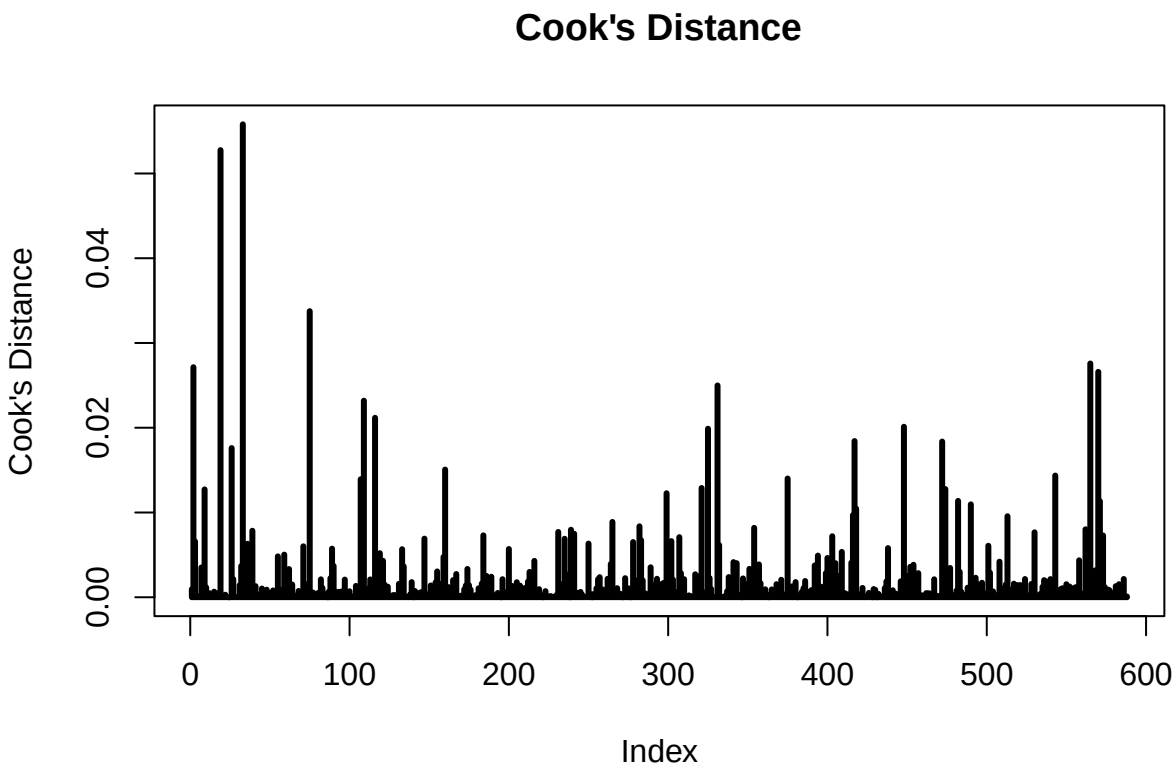
```
# Checking the VIFs for multicollinearity

vif(model1_trans)
```

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

```
# threshold for multicollinearity
# Calculating the threshold

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

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

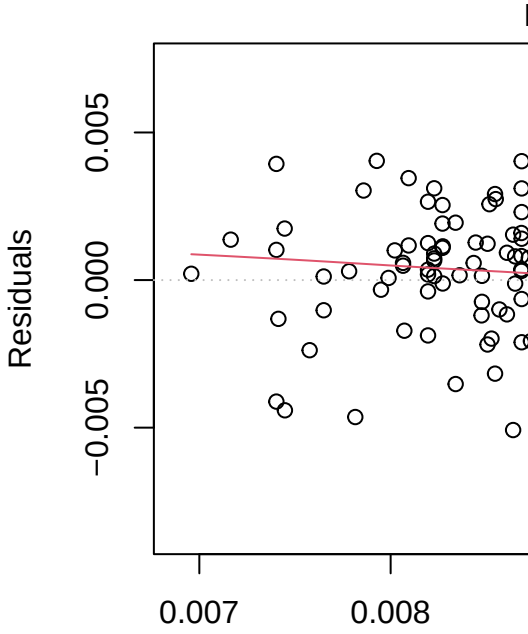
```
# Checking outliers: estimate of the influence of data point; summary of how much a regression model ch

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



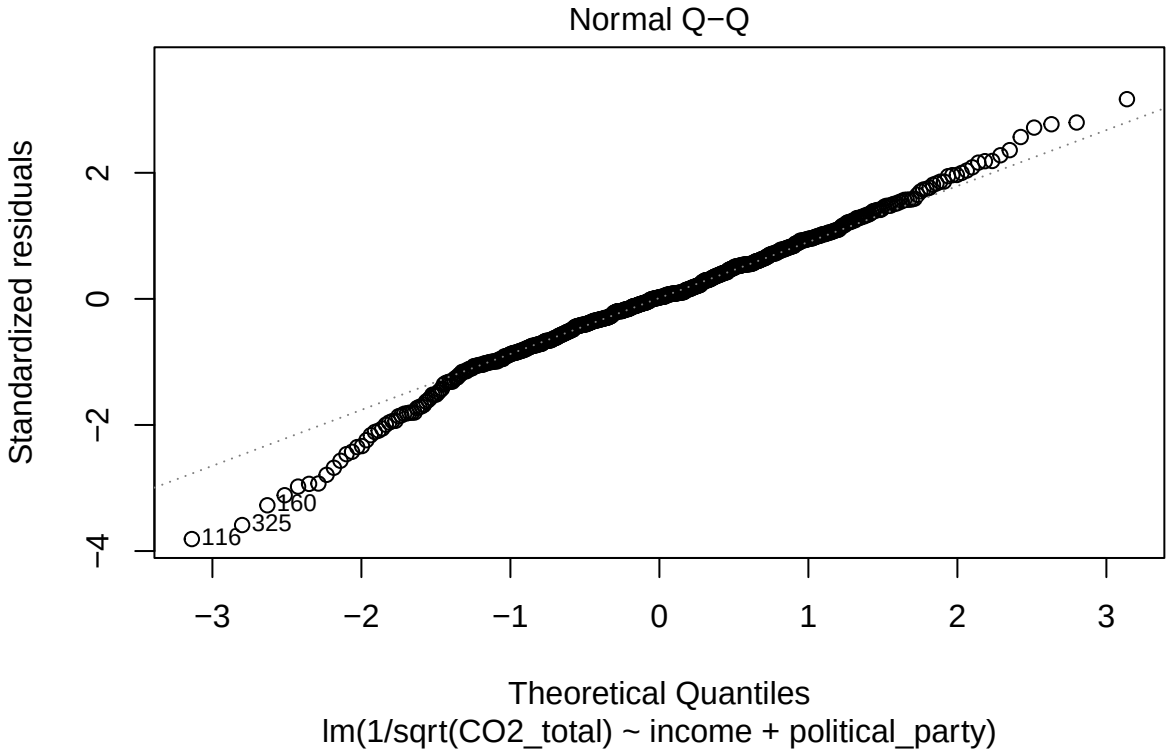

```
influential = cooks.distance(model1_trans)[which(cook >1)]

influential
```

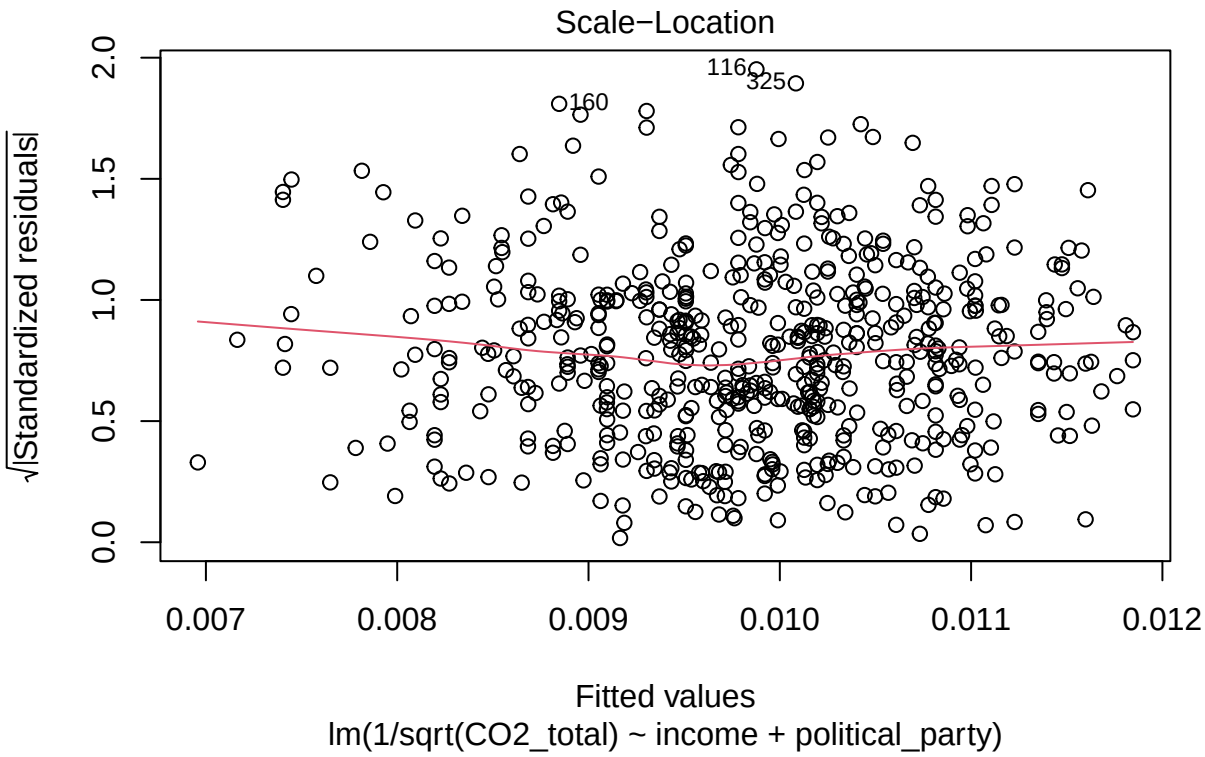
```
## named numeric(0)
```

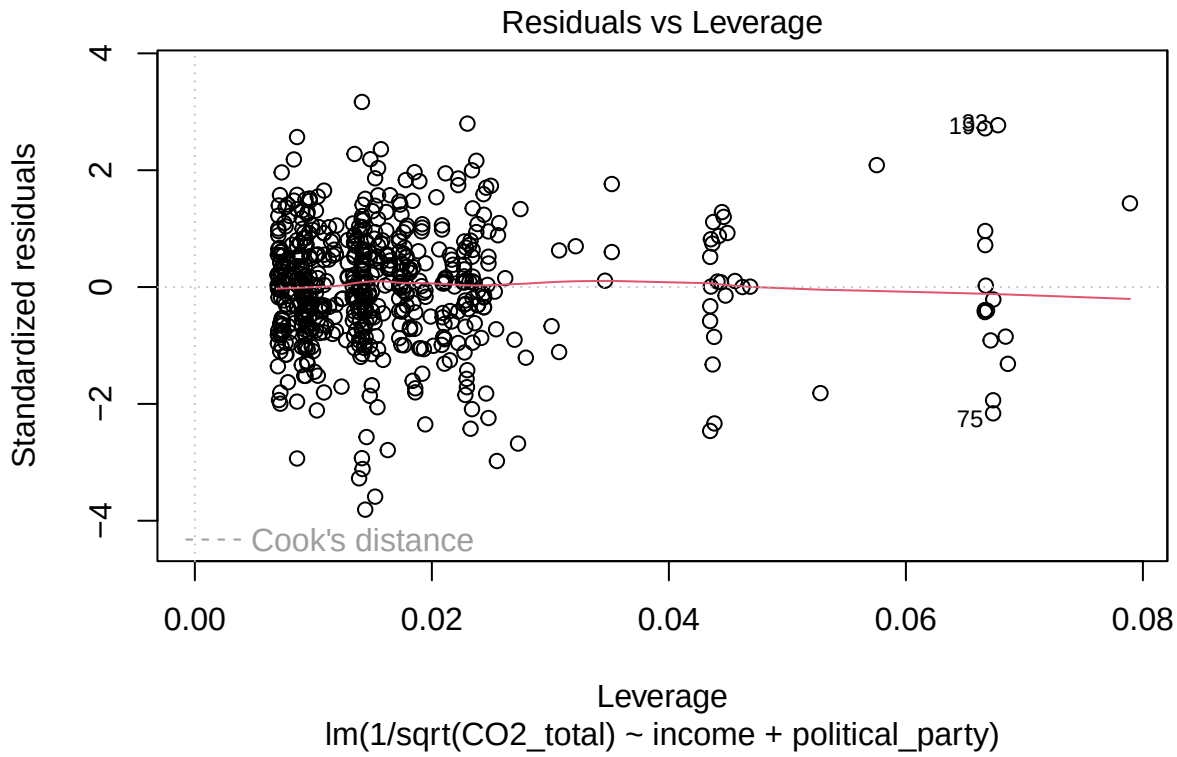
```

plot(model1_trans)
```



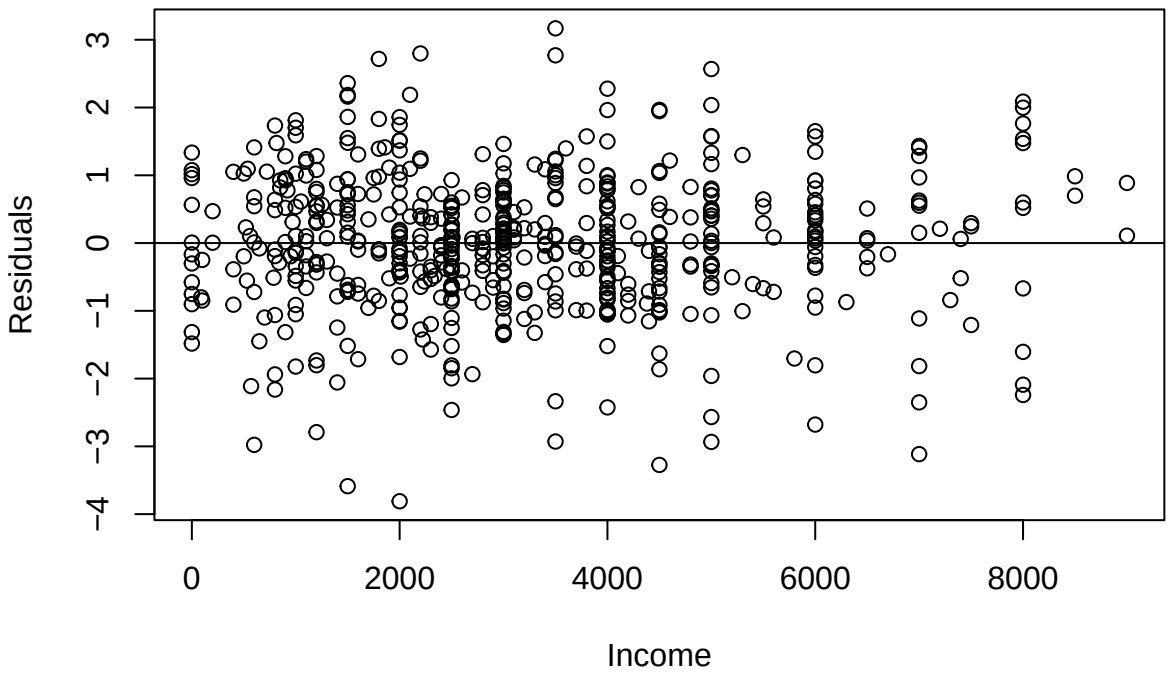
5. Assumptions check in the residuals of the transformed regression



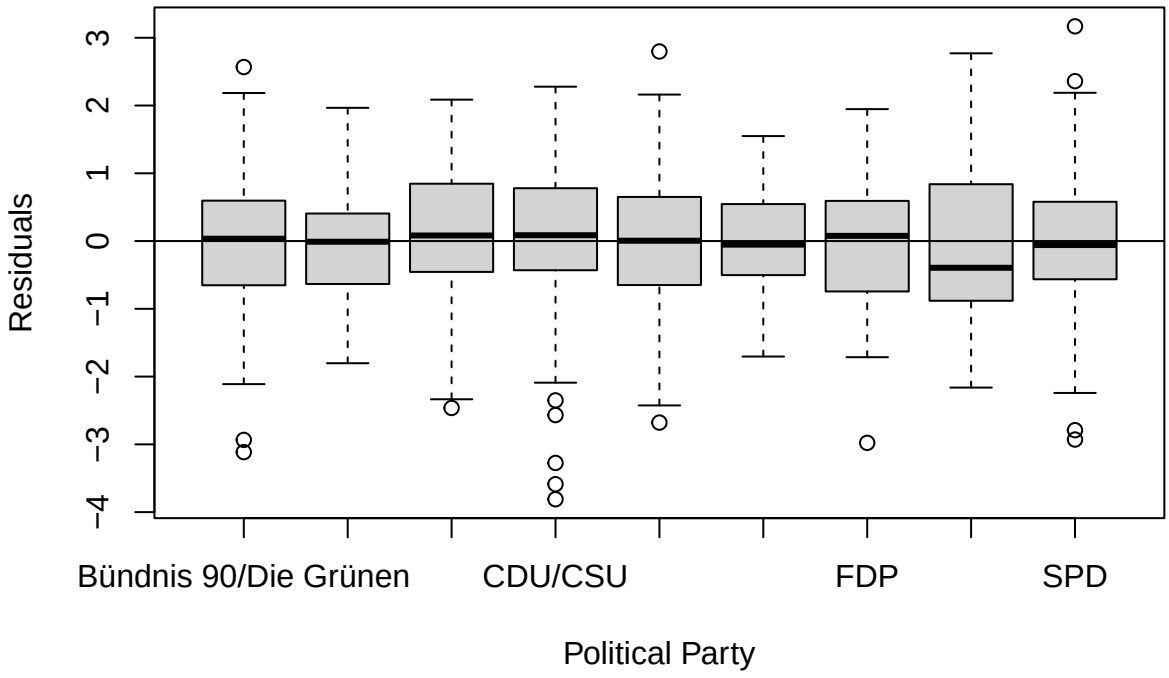

Scale−Location
116
325
160
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
0.007
0.008
0.009
0.010
0.011
0.012
Fitted values
lm(1/sqrt(CO2_total) ~ income + political_party)




```
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)
```



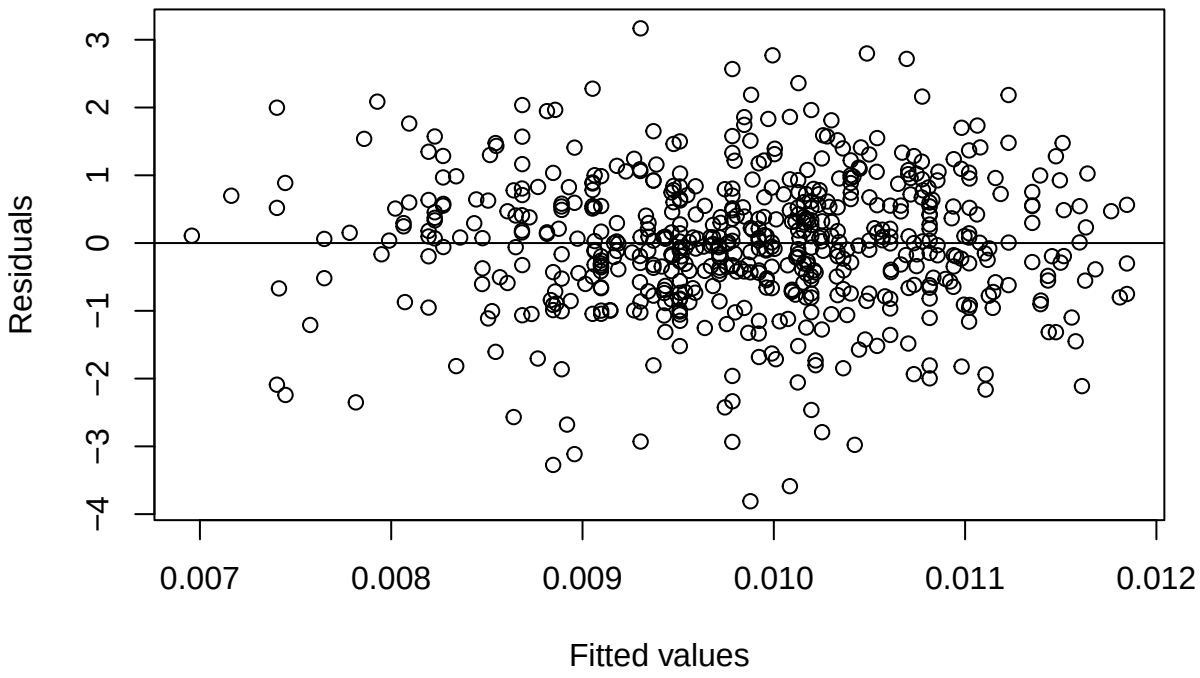

```
#plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")
#abline(h = 0)

#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)
```




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

durbinWatsonTest(model1_trans)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1           0.019             2    0.63
##  Alternative hypothesis: rho != 0
```

```
# Constant variance and independent error term assumption

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




```
# Breusch-Pagan TEST
# Heteroscedasticity, constant error terms
# H0: Homoscedasticity is present
# Reject H0

library(lmtest)
```

```
## Loading required package: zoo
```

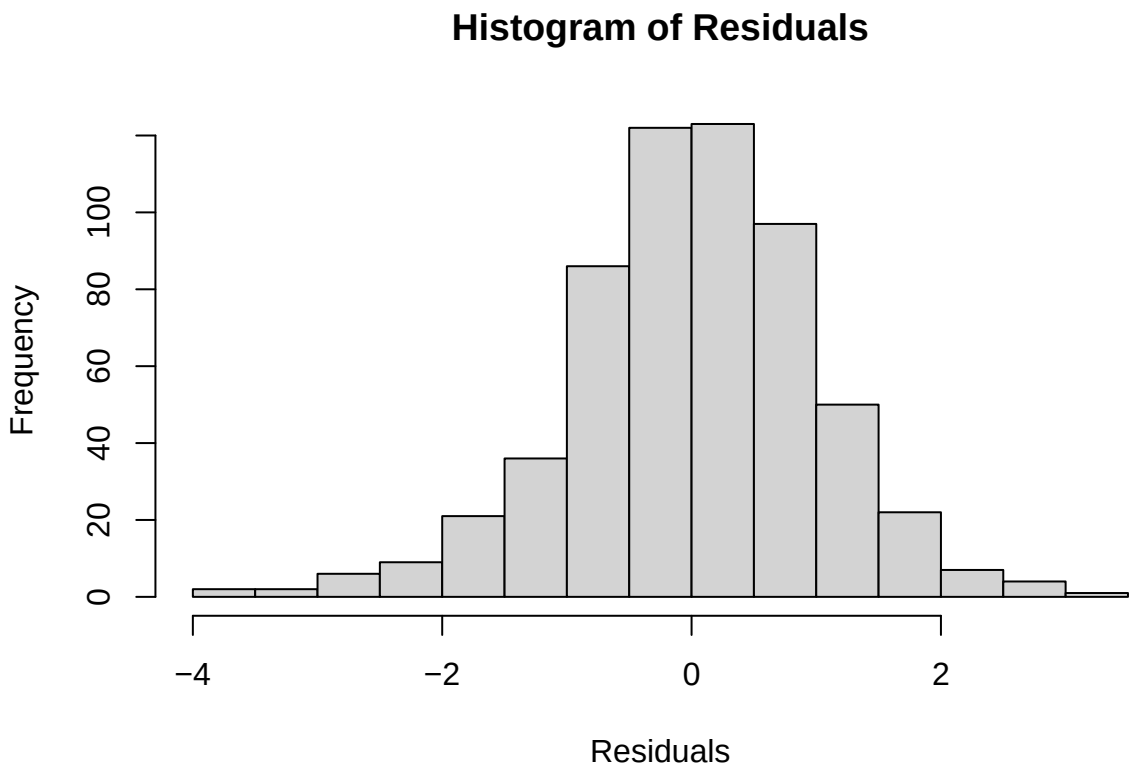
```
##
## Attaching package: 'zoo'
```

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

```
bptest(model1_trans)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model1_trans
## BP = 22, df = 9, p-value = 0.01
```

```r
# Normality assumption

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

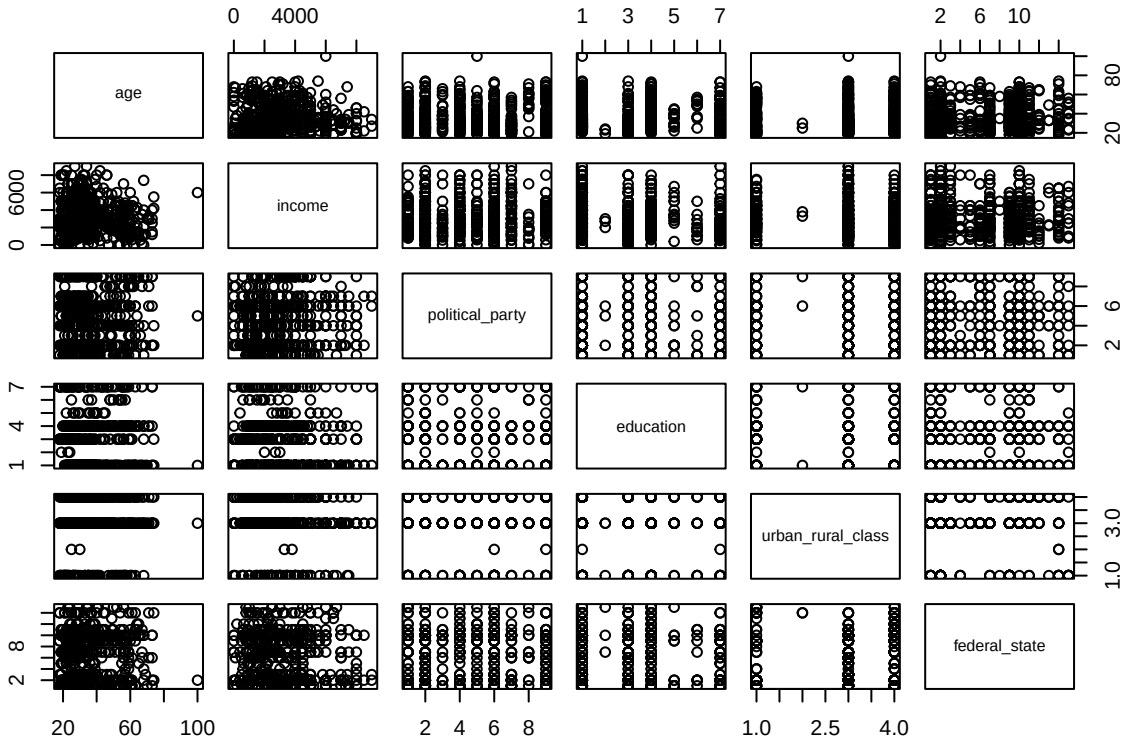



```r
# Normality test using Shapiro-test: reject the H0
# H0: the sample comes from a normal distribution
# Reject H0

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

```
##
##  Shapiro-Wilk normality test
##
## data:  res1_num
## W = 1, p-value = 6e-05
```

III. Multivariate Regression: belief diff total

```r

# Checking the possible correlation in the data

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



1. Modeling

```
## defining a reference level

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

```
# regression model

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

```
##
## Call:
## lm(formula = belief_diff_total ~ age + income + political_party +
##     education + urban_rural_class + federal_state, data = df2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -78.31 -19.19   0.79  19.22  81.63
##
## Coefficients:
##                                                                    Estimate
## (Intercept)                                                        8.72e+00
## age                                                               -5.61e-05
```

```
## income                                                                    -3.42e-03
## political_partyAfD                                                        -1.37e+01
## political_partyBündnis Sarah Wagenknecht                                  -5.60e+00
## political_partyCDU/CSU                                                    -7.66e+00
## political_partyDie Linke                                                  -1.58e-01
## political_partyEiner anderen Partei                                       -1.88e+00
## political_partyFDP                                                        -1.22e+01
## political_partyKeine Angabe                                               -6.16e+00
## political_partySPD                                                        -9.51e+00
## education(Noch) kein Abschluss                                             1.19e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)  2.03e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule       5.77e+00
## educationDoktorgrad oder Habilitation                                      3.79e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   6.44e+00
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         5.91e+00
## urban_rural_classperipher                                                 -7.01e-01
## urban_rural_classsehr peripher                                            -2.19e+01
## urban_rural_classzentral                                                   3.42e+00
## federal_stateBaden-Württemberg                                             4.63e+00
## federal_stateBayern                                                        1.34e-01
## federal_stateBerlin                                                       -9.83e+00
## federal_stateBrandenburg                                                   1.90e+01
## federal_stateBremen                                                        4.72e+00
## federal_stateHamburg                                                       2.28e+00
## federal_stateHessen                                                       -8.32e-01
## federal_stateMecklenburg-Vorpommern                                        2.75e+01
## federal_stateNiedersachsen                                                -7.29e+00
## federal_stateRheinland-Pfalz                                              -6.95e+00
## federal_stateSaarland                                                     -1.07e+00
## federal_stateSachsen-Anhalt                                                6.13e+00
## federal_stateSchleswig-Holstein                                           -1.93e+00
## federal_stateThüringen                                                    -5.83e+00
##                                                                          Std. Error
## (Intercept)                                                                5.49e+00
## age                                                                        9.72e-02
## income                                                                     6.50e-04
## political_partyAfD                                                         4.65e+00
## political_partyBündnis Sarah Wagenknecht                                   6.52e+00
## political_partyCDU/CSU                                                     4.18e+00
## political_partyDie Linke                                                   5.03e+00
## political_partyEiner anderen Partei                                        3.74e+00
## political_partyFDP                                                         4.84e+00
## political_partyKeine Angabe                                                8.38e+00
## political_partySPD                                                         4.25e+00
## education(Noch) kein Abschluss                                             1.70e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   3.31e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule       3.37e+00
## educationDoktorgrad oder Habilitation                                      8.25e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss   9.42e+00
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss         4.29e+00
## urban_rural_classperipher                                                  4.32e+00
## urban_rural_classsehr peripher                                             2.14e+01
## urban_rural_classzentral                                                   3.17e+00
## federal_stateBaden-Württemberg                                             4.12e+00
```

```
## federal_stateBayern                                                                  4.23e+00
## federal_stateBerlin                                                                  5.14e+00
## federal_stateBrandenburg                                                             1.08e+01
## federal_stateBremen                                                                  7.87e+00
## federal_stateHamburg                                                                 6.40e+00
## federal_stateHessen                                                                  4.91e+00
## federal_stateMecklenburg-Vorpommern                                                  2.06e+01
## federal_stateNiedersachsen                                                           5.00e+00
## federal_stateRheinland-Pfalz                                                         6.17e+00
## federal_stateSaarland                                                                9.68e+00
## federal_stateSachsen-Anhalt                                                          1.49e+01
## federal_stateSchleswig-Holstein                                                      7.21e+00
## federal_stateThüringen                                                               1.09e+01
##                                                                                    t value
## (Intercept)                                                                         1.59
## age                                                                                 0.00
## income                                                                             -5.27
## political_partyAfD                                                                 -2.95
## political_partyBündnis Sarah Wagenknecht                                           -0.86
## political_partyCDU/CSU                                                             -1.83
## political_partyDie Linke                                                           -0.03
## political_partyEiner anderen Partei                                                -0.50
## political_partyFDP                                                                 -2.53
## political_partyKeine Angabe                                                        -0.74
## political_partySPD                                                                 -2.24
## education(Noch) kein Abschluss                                                      0.70
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)   0.61
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                1.71
## educationDoktorgrad oder Habilitation                                               0.46
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss    0.68
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss          1.38
## urban_rural_classperipher                                                          -0.16
## urban_rural_classsehr peripher                                                     -1.02
## urban_rural_classzentral                                                            1.08
## federal_stateBaden-Württemberg                                                      1.12
## federal_stateBayern                                                                 0.03
## federal_stateBerlin                                                                -1.91
## federal_stateBrandenburg                                                            1.76
## federal_stateBremen                                                                 0.60
## federal_stateHamburg                                                                0.36
## federal_stateHessen                                                                -0.17
## federal_stateMecklenburg-Vorpommern                                                 1.33
## federal_stateNiedersachsen                                                         -1.46
## federal_stateRheinland-Pfalz                                                       -1.13
## federal_stateSaarland                                                              -0.11
## federal_stateSachsen-Anhalt                                                         0.41
## federal_stateSchleswig-Holstein                                                    -0.27
## federal_stateThüringen                                                             -0.53
##                                                                                  Pr(>|t|)
## (Intercept)                                                                       0.1128
## age                                                                               0.9995
## income                                                                             2e-07
## political_partyAfD                                                                0.0033
## political_partyBündnis Sarah Wagenknecht                                          0.3909
```

```
## political_partyCDU/CSU                                                                0.0673
## political_partyDie Linke                                                               0.9750
## political_partyEiner anderen Partei                                                    0.6143
## political_partyFDP                                                                     0.0118
## political_partyKeine Angabe                                                            0.4626
## political_partySPD                                                                     0.0257
## education(Noch) kein Abschluss                                                         0.4859
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)      0.5401
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule                   0.0877
## educationDoktorgrad oder Habilitation                                                  0.6465
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss       0.4943
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss             0.1689
## urban_rural_classperipher                                                              0.8713
## urban_rural_classsehr peripher                                                         0.3069
## urban_rural_classzentral                                                               0.2815
## federal_stateBaden-Württemberg                                                         0.2615
## federal_stateBayern                                                                    0.9748
## federal_stateBerlin                                                                    0.0561
## federal_stateBrandenburg                                                               0.0793
## federal_stateBremen                                                                    0.5488
## federal_stateHamburg                                                                   0.7217
## federal_stateHessen                                                                    0.8657
## federal_stateMecklenburg-Vorpommern                                                    0.1826
## federal_stateNiedersachsen                                                             0.1455
## federal_stateRheinland-Pfalz                                                           0.2602
## federal_stateSaarland                                                                  0.9124
## federal_stateSachsen-Anhalt                                                            0.6815
## federal_stateSchleswig-Holstein                                                        0.7886
## federal_stateThüringen                                                                 0.5938
##
## (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: 28 on 554 degrees of freedom
## Multiple R-squared:  0.122,  Adjusted R-squared:  0.07
## F-statistic: 2.34 on 33 and 554 DF,  p-value: 5.17e-05
```

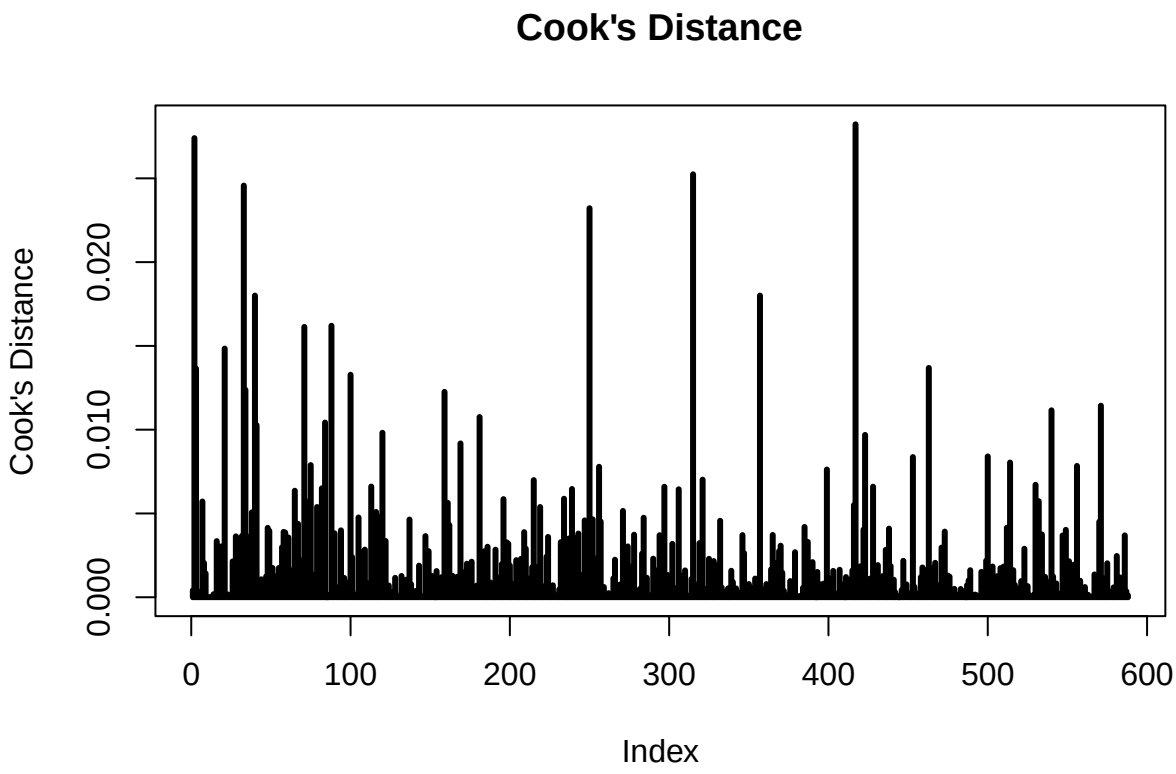
```
# Checking the VIFs for multicollinearity: federal_state variable should be removed

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 the outliers

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




```
influential = cooks.distance(model2)[which(cook >3*mean(cook, na.rm=TRUE))]
influential
```

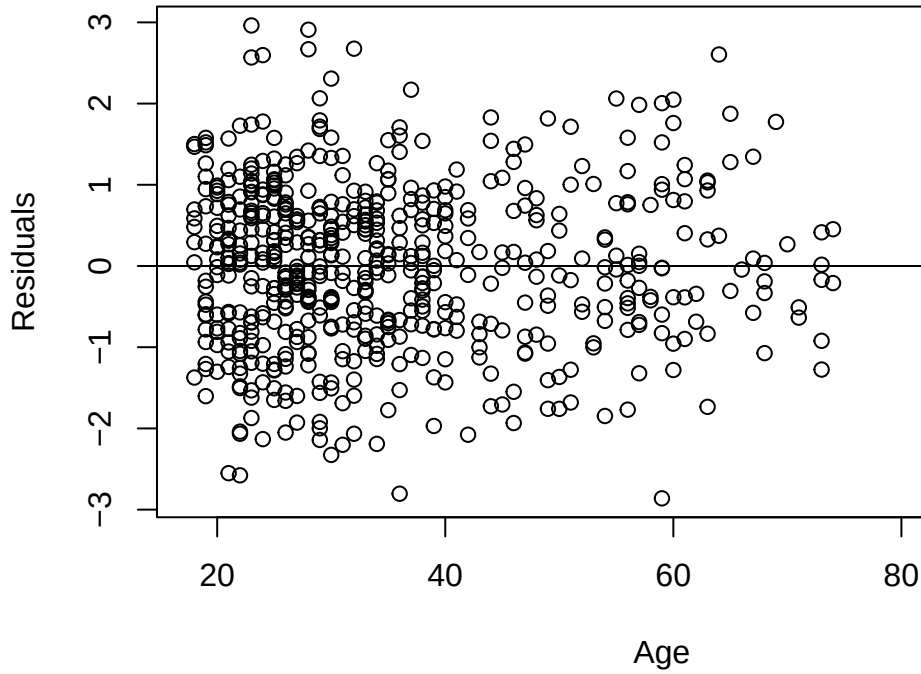
```
##      2      3      7     21     33     34     40     41     65     71     72
## 0.0274 0.0136 0.0057 0.0148 0.0246 0.0124 0.0180 0.0103 0.0064 0.0161 0.0057
##     75     82     84     88    100    113    120    159    161    169    181
## 0.0079 0.0065 0.0104 0.0162 0.0133 0.0066 0.0098 0.0123 0.0056 0.0092 0.0108
##    196    215    234    239    250    256    297    306    315    321    357
## 0.0059 0.0070 0.0059 0.0065 0.0232 0.0078 0.0066 0.0064 0.0252 0.0070 0.0180
##    399    417    423    428    453    463    500    514    530    532    540
## 0.0076 0.0282 0.0097 0.0066 0.0084 0.0137 0.0084 0.0080 0.0067 0.0057 0.0112
##    556    571
## 0.0078 0.0114
```

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


#df2_no_outliers = df2_scaled[-influential_vector, ]

df2[influential_vector, ]
```

```
## # A tibble: 46 x 7
##      age income political_party education               urban~1 feder~2 belie~3
##    <int>  <dbl> <fct>           <fct>                   <fct>   <fct>     <dbl>
```

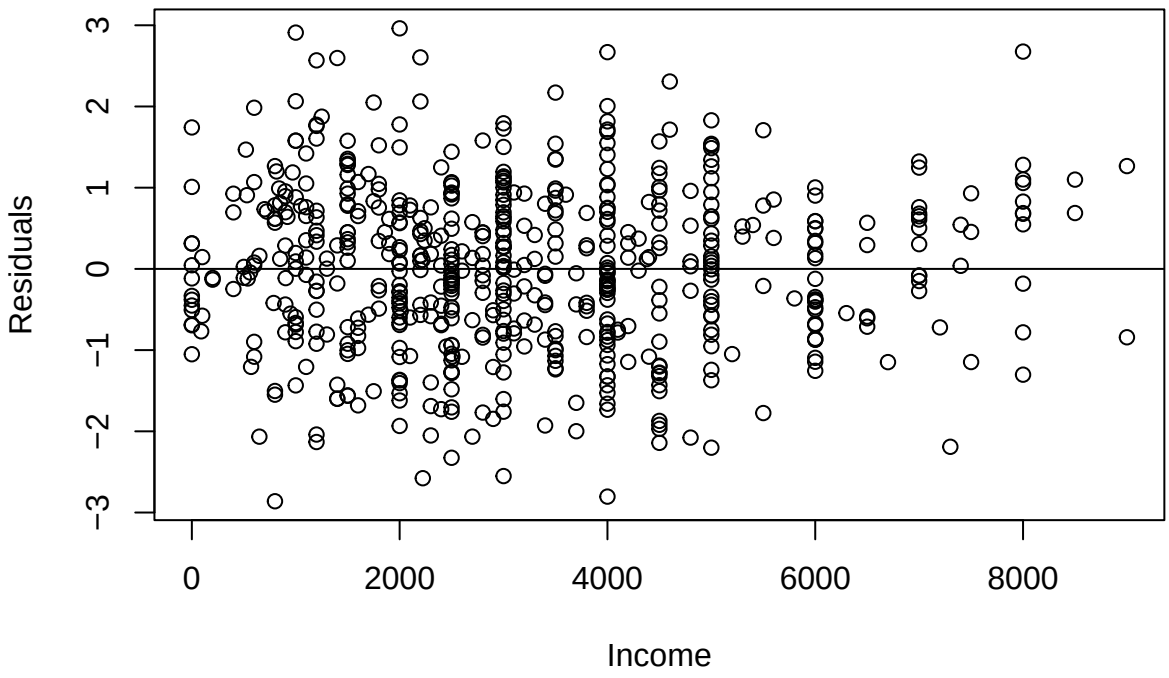
```
##  1    59    800 Keine Angabe    Allgemeine oder fachgeb~ sehr z~ Hessen      -76
##  2    60   1750 Keine Angabe    Berufsausbildung, Lehre~ periph~ Bayern       57
##  3    57    600 CDU/CSU         Realschulabschluss (Mit~ zentral Baden-~      68
##  4    54   2900 AfD             Hauptschulabschluss (Vo~ zentral Rheinl~     -61
##  5    37   3500 Keine Angabe    Hauptschulabschluss (Vo~ sehr z~ Bayern       54
##  6    59   4000 AfD             Berufsausbildung, Lehre~ sehr z~ Bremen       46
##  7    58   4000 CDU/CSU         (Fach-) Hochschulabschl~ periph~ Meckle~      29
##  8    50   2000 Keine Angabe    Hauptschulabschluss (Vo~ zentral Rheinl~     -37
##  9    68   2100 AfD             Berufsausbildung, Lehre~ zentral Brande~     -12
## 10    56   1000 Keine Angabe    Berufsausbildung, Lehre~ periph~ Thürin~      39
## # ... with 36 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: belief_diff_total
```

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

# Linearity assumption/Mean zero assumption

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

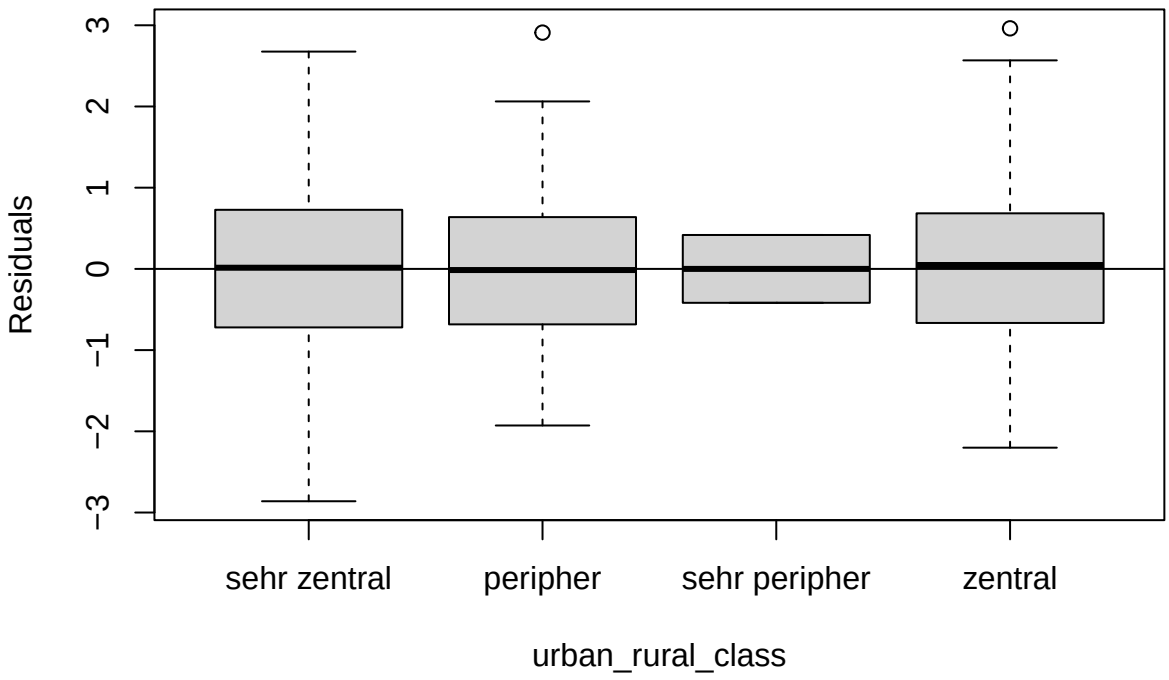


2. Assumptions check in the residuals

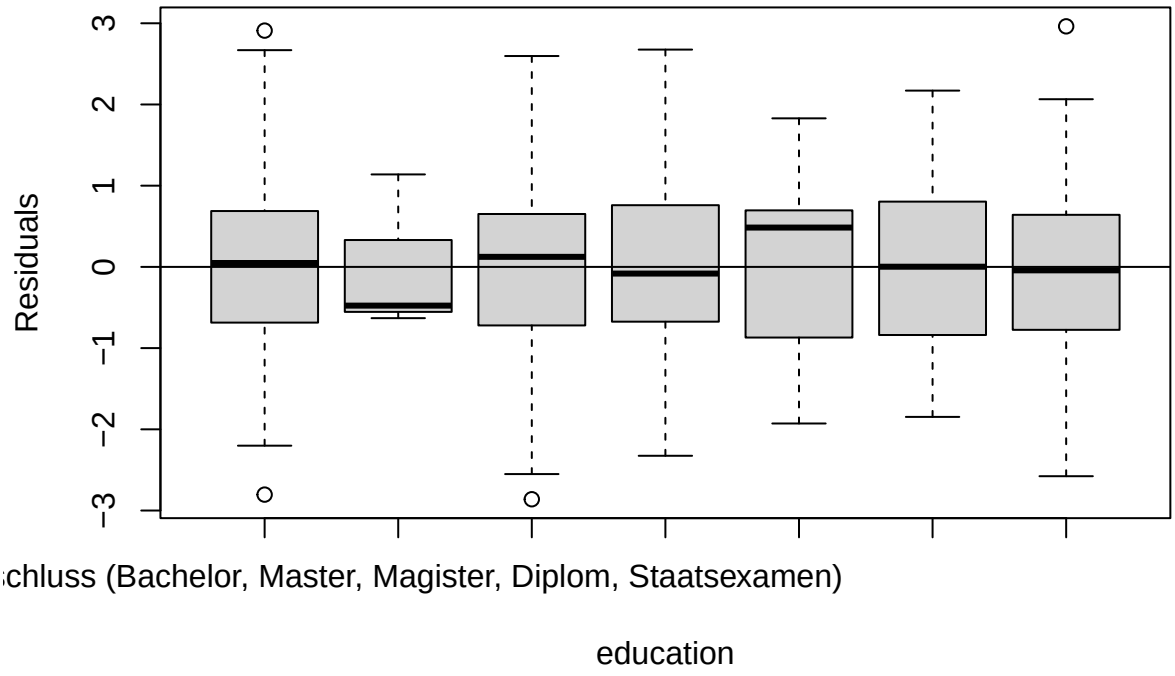
```r
plot(df2$income, res2, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



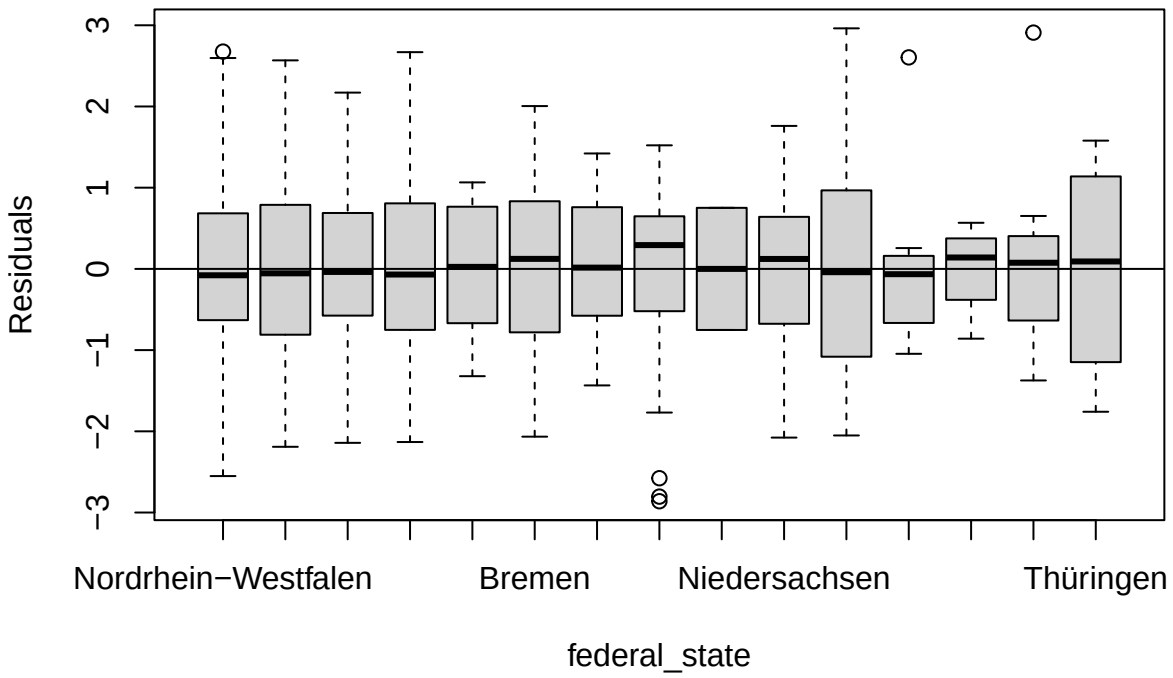

```
plot(df2$urban_rural_class, res2, xlab = "urban_rural_class", ylab = "Residuals")
abline(h = 0)
```



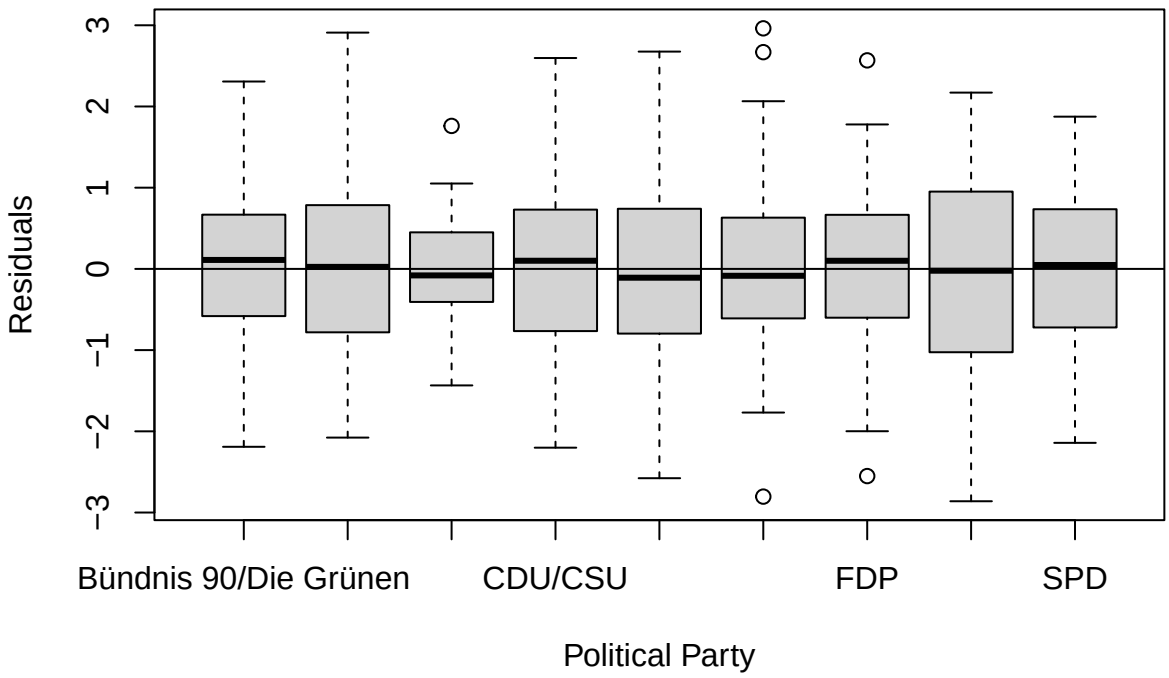

```
plot(df2$education, res2, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



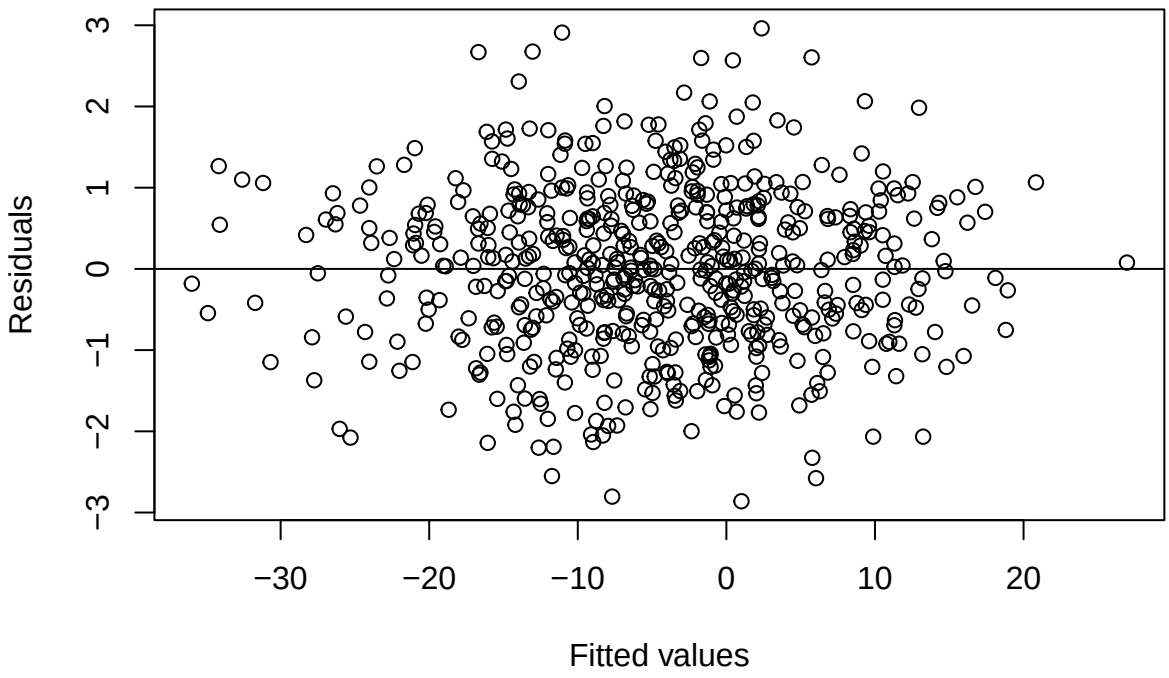

```
plot(df2$federal_state, res2, xlab = "federal_state", ylab = "Residuals")
abline(h = 0)
```




```
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)
```



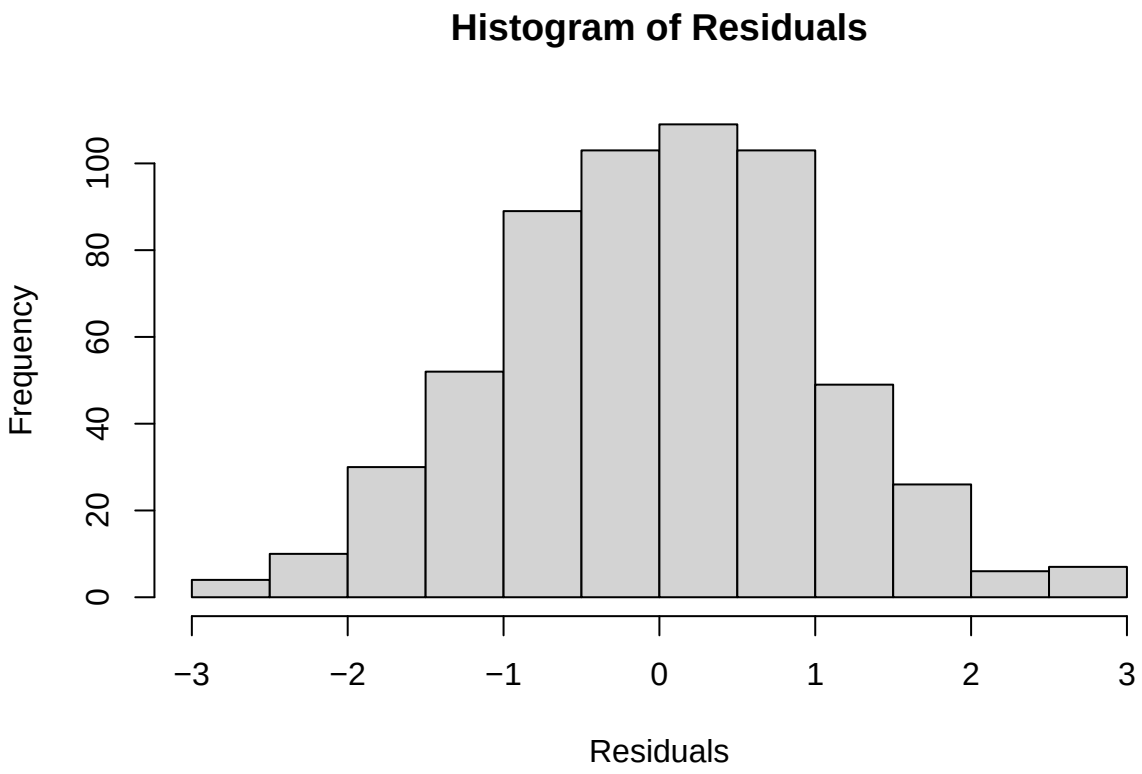

```
# Durbin-Watson Test:  Independence of the error terms
# HO (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(model2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1          -0.023             2    0.59
##  Alternative hypothesis: rho != 0
```

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

bptest(model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model2
## BP = 40, df = 33, p-value = 0.2
```

```
# 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 = 0.8
```

FINAL MODEL

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

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

3. Variable Selection, model outcome and assumption check

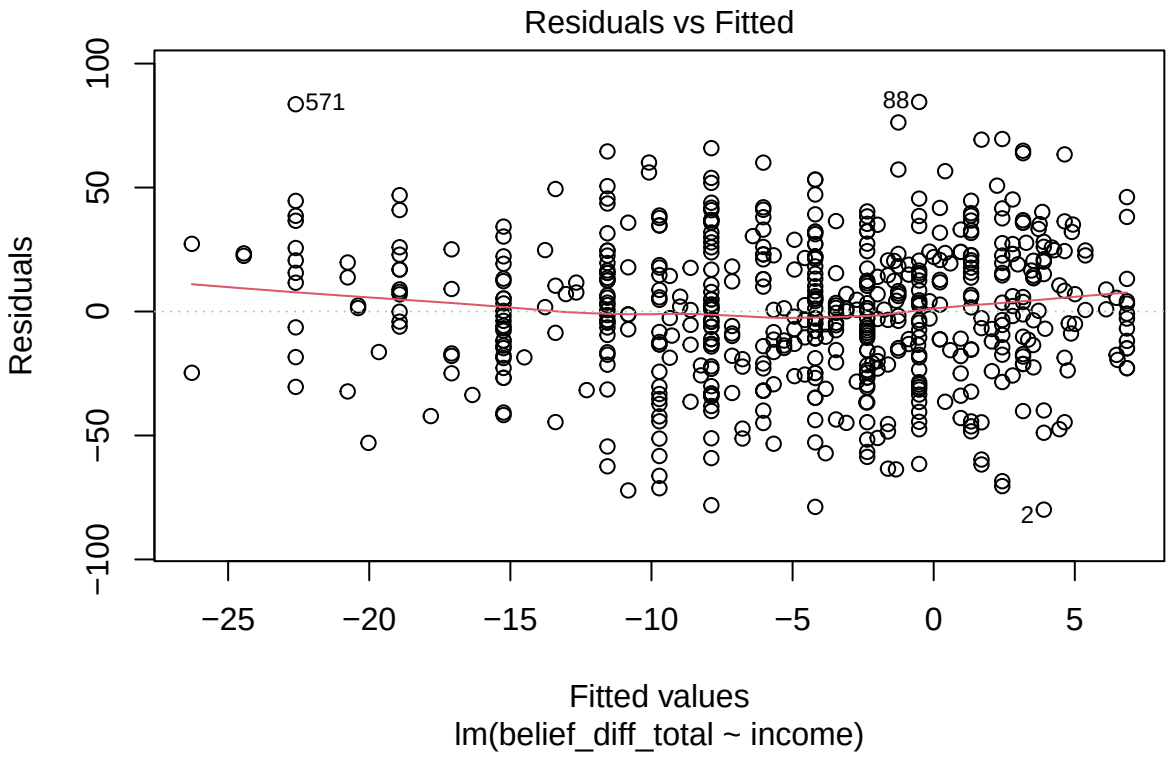
```
## Start:  AIC=3969
## belief_diff_total ~ age + income + political_party + education +
##     urban_rural_class + federal_state
##
##                      Df Sum of Sq    RSS  AIC
## - federal_state     14     15639 462817 3961
## - education          6      3464 450642 3961
## - urban_rural_class  3      2164 449342 3966
## - age                1         0 447178 3967
## <none>                            447178 3969
## - political_party    8     13629 460807 3970
## - income             1     22405 469583 3996
##
## Step:  AIC=3961
## belief_diff_total ~ age + income + political_party + education +
##     urban_rural_class
##
##                      Df Sum of Sq    RSS  AIC
## - education          6      3759 466576 3954
## - urban_rural_class  3      2855 465673 3959
## - age                1         7 462824 3959
## <none>                            462817 3961
## - political_party    8     12812 475629 3961
## - income             1     22837 485655 3987
##
## Step:  AIC=3954
## belief_diff_total ~ age + income + political_party + urban_rural_class
##
##                      Df Sum of Sq    RSS  AIC
## - urban_rural_class  3      2968 469544 3951
## - age                1        98 466674 3952
## - political_party    8     12138 478714 3953
## <none>                            466576 3954
## - income             1     25304 491880 3983
##
## Step:  AIC=3951
## belief_diff_total ~ age + income + political_party
##
##                    Df Sum of Sq    RSS  AIC
## - political_party  8     11412 480956 3950
## - age              1       172 469715 3950
## <none>                         469544 3951
## - income           1     26512 496055 3982
##
## Step:  AIC=3950
## belief_diff_total ~ age + income
##
##          Df Sum of Sq    RSS  AIC
## - age     1        16 480972 3948
## <none>                480956 3950
## - income  1     28452 509408 3981
##
## Step:  AIC=3948
## belief_diff_total ~ income
```

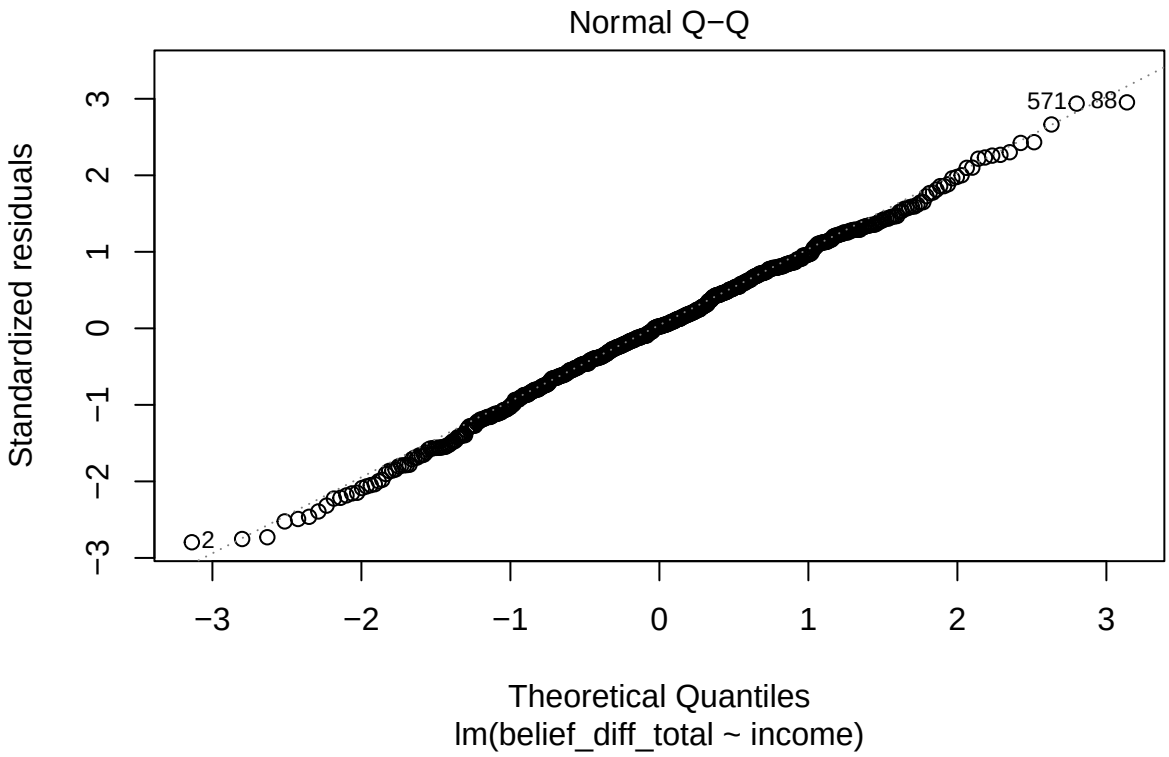
```
##
##          Df Sum of Sq    RSS  AIC
## <none>                480972 3948
## - income  1     28510 509482 3979
```

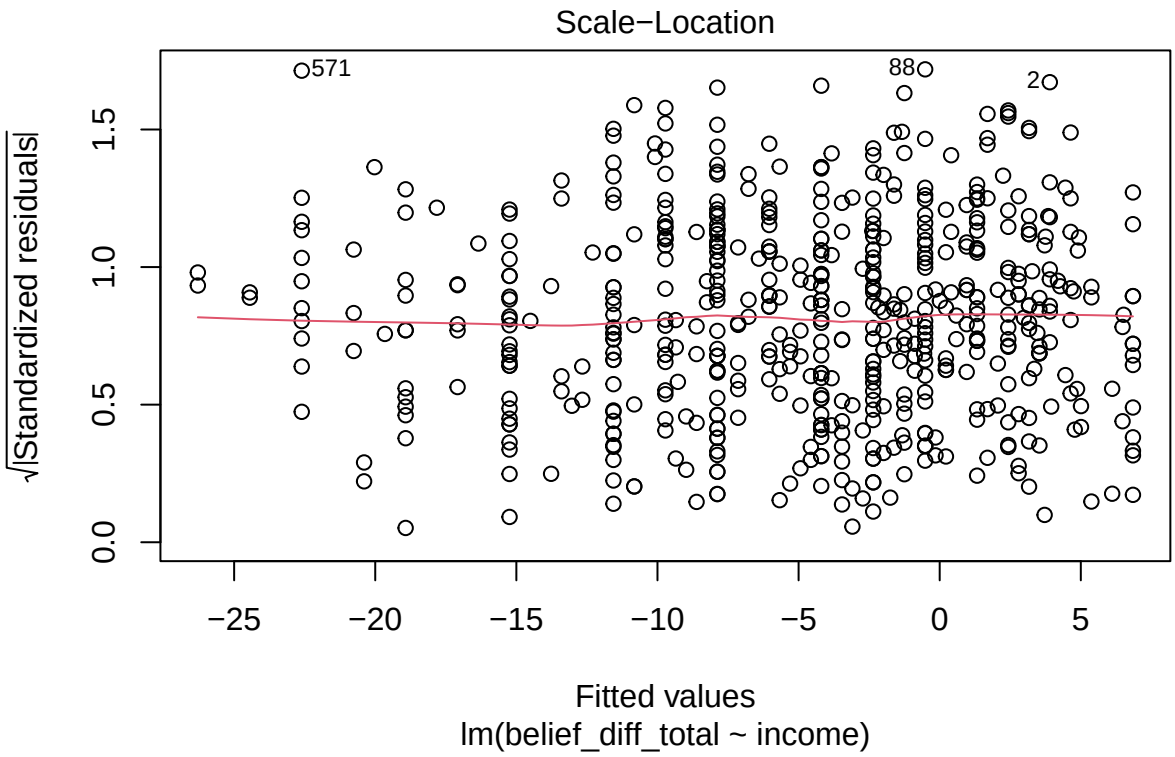
```
summary(step_model2)
```

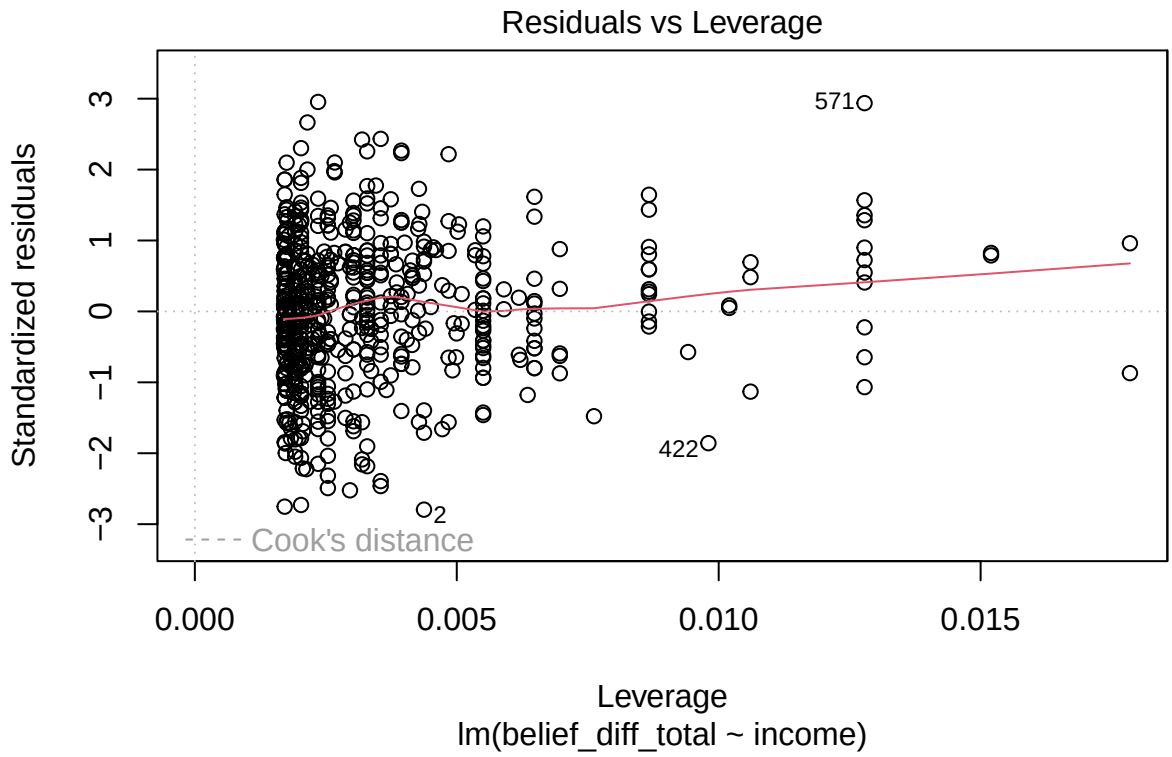
```
##
## Call:
## lm(formula = belief_diff_total ~ income, data = df2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -79.9  -17.9    0.7   20.4   84.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.847630   2.306632    2.97   0.0031 **
## income      -0.003682   0.000625   -5.89  6.4e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29 on 586 degrees of freedom
## Multiple R-squared:  0.056,  Adjusted R-squared:  0.0543
## F-statistic: 34.7 on 1 and 586 DF,  p-value: 6.38e-09
```

```
plot(step_model2)
```

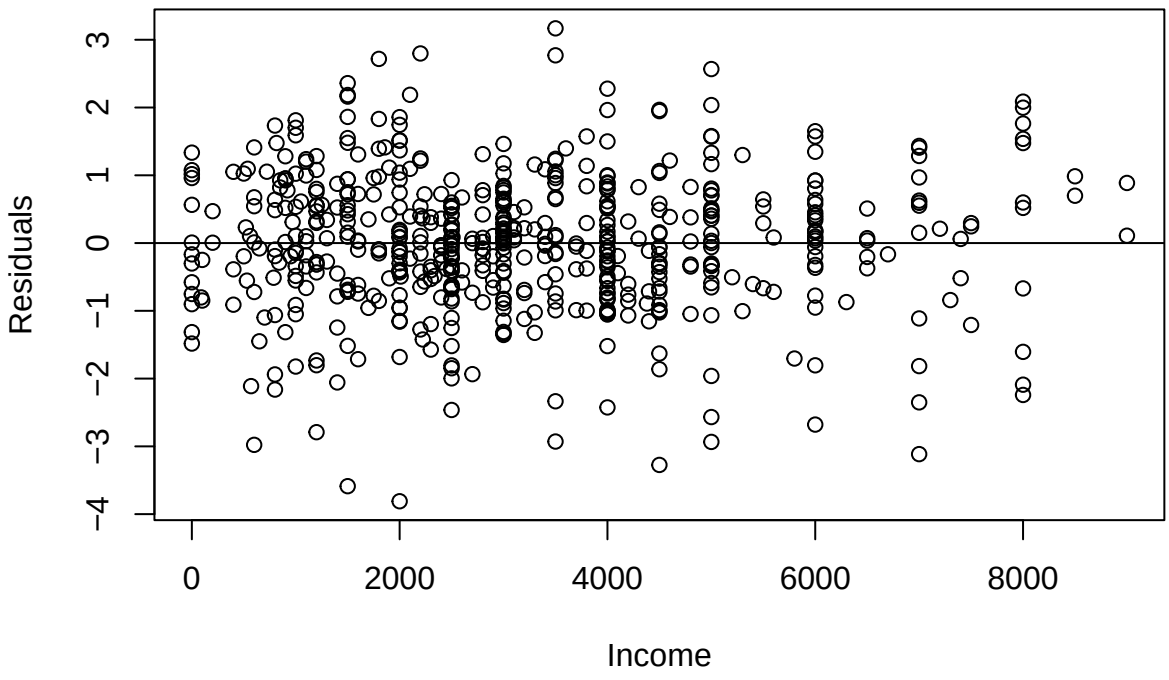

Residuals vs Fitted
Residuals
Fitted values
lm(belief_diff_total ~ income)
571
88
2
100
50
0
−50
−100
−25
−20
−15
−10
−5
0
5


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(belief_diff_total ~ income)
571
88
2
−3
−2
−1
0
1
2
3


Scale−Location
571
88
2
√|Standardized residuals|
0.0
0.5
1.0
1.5
−25
−20
−15
−10
−5
0
5
Fitted values
lm(belief_diff_total ~ income)




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

# Linearity assumption/Mean zero assumption

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

plot(df2$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_new$education, res1, xlab = "education", ylab = "Residuals")
#abline(h = 0)

#plot(df1_new$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
#abline(h = 0)

#plot(df1_new$political_party, res1, xlab = "Political Party", ylab = "Residuals")
#abline(h = 0)
```

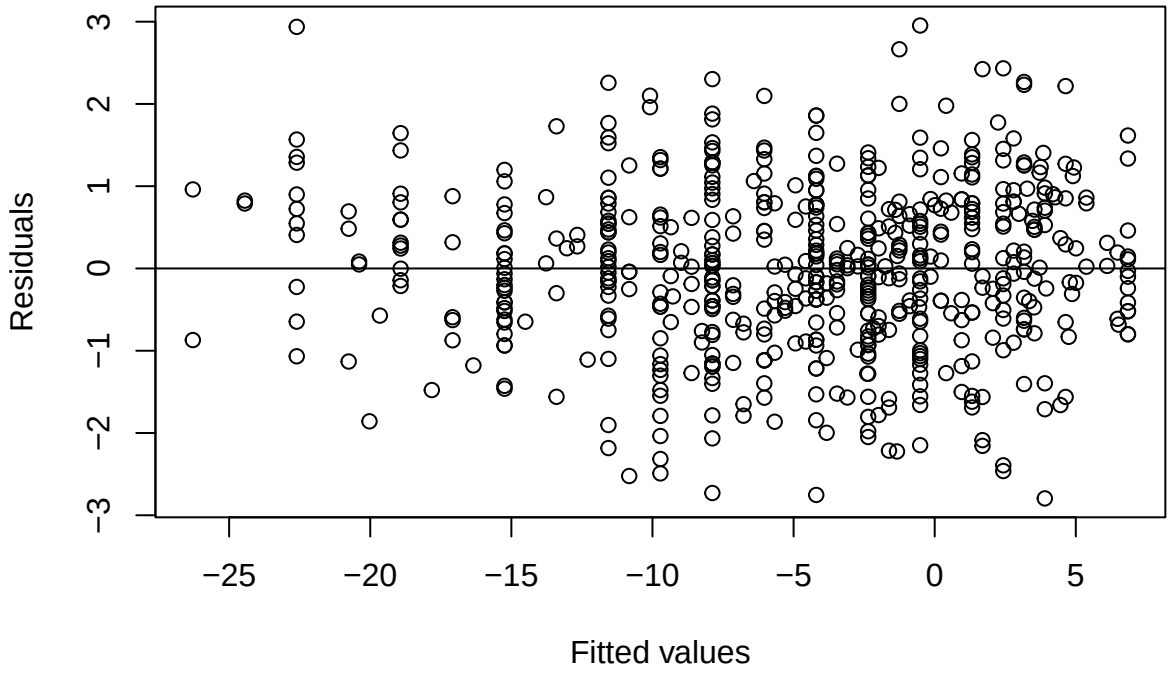
```
# Durbin-Watson Test: independence of the error terms
# H0 (null hypothesis): There is no correlation among the residuals

durbinWatsonTest(step_model2)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1          -0.015             2    0.72
##  Alternative hypothesis: rho != 0
```

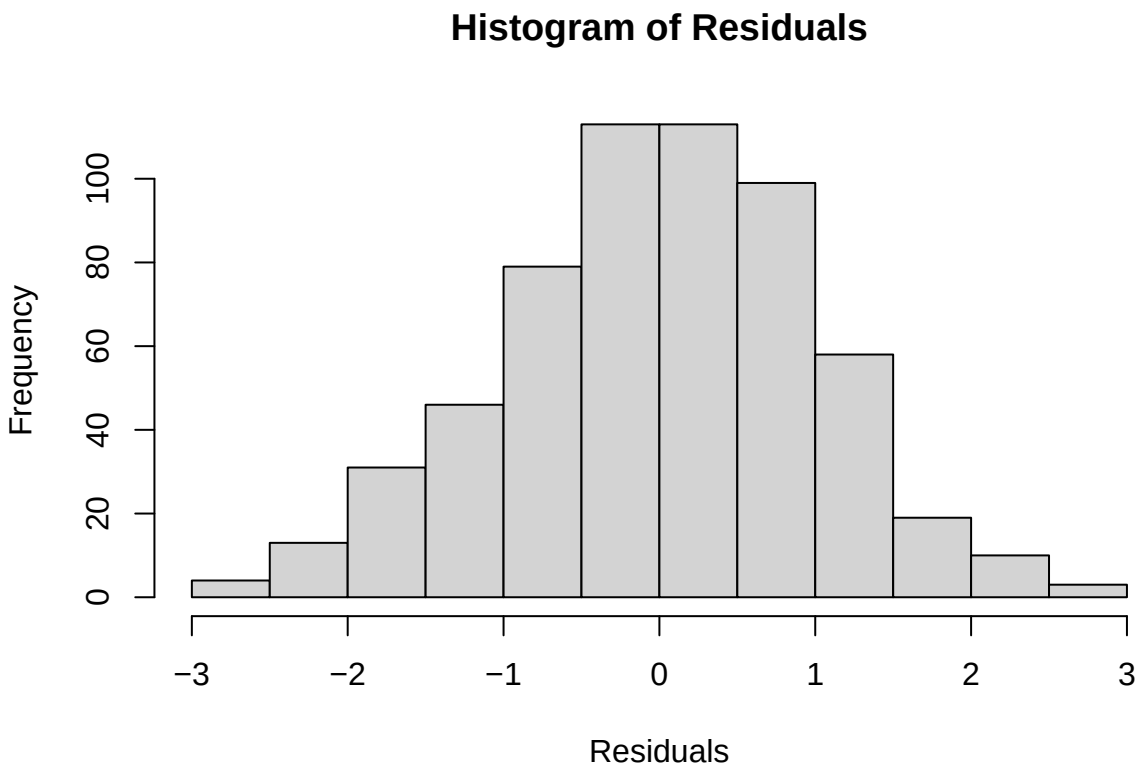
```
# Breusch-Pagan TEST: Heteroscedasticity
# H0: Homoscedasticity is present

bptest(step_model2)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  step_model2
## BP = 0.2, df = 1, p-value = 0.6
```

```r
# Constant variance and independent error term assumption

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




```r
# 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 = 0.5
```