

Regression_Analysis_Mobility

2024-01-18

```
### import libraries
```

```
library(car)
```

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

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

```
##
```

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

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

```
##
```

```
##      select
```

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

```
##
```

```
##      recode
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

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

```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      date, intersect, setdiff, union
```

```
library(coefplot)
```

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

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

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

```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

Loading the data

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

```
head(df, 10)
```

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

## 4	IN sehr zentral		Städtischer Kreis			
## 5	PU sehr zentral		kreisfreie Großstadt			
## 6	IN zentral		kreisfreie Großstadt			
## 7	IN zentral		Städtischer Kreis			
## 8	PU sehr zentral		kreisfreie Großstadt			
## 9	PU sehr zentral		kreisfreie Großstadt			
## 10	PU sehr zentral		kreisfreie Großstadt			
##	federal_state	C02_housing	C02_electricity	C02_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		
##	C02_cruise	C02_flight	C02_public_transport	C02_car1	C02_car2	C02_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
##	C02_car4	C02_car5	C02_car_total	C02_mobility	C02_food	C02_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	C02_total	belief_diff_housing_electricity			
## 1		1152	16376.656			-31
## 2		1152	18384.935			-38
## 3		1152	6335.123			40
## 4		1152	12356.940			-2
## 5		1152	9074.518			-43
## 6		1152	12180.790			-6
## 7		1152	5700.710			-1
## 8		1152	23114.072			5
## 9		1152	12600.568			-48
## 10		1152	10875.468			-1
##	belief_diff_mobility	belief_diff_food	belief_diff_other_consumption			
## 1		-14	5			-68
## 2		-42	-26			23

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

Hypotheses for the regression model

1. The first dependent variable: actual CO2 emission H1a: age makes differences in the actual CO2 emission from everyday activity.

H1b: income makes differences in the actual CO2 emission from everyday activity.

H1c: education level makes differences in the actual CO2 emission from everyday activity.

H1d: the place of residence (city or countryside) in the actual CO2 emission from every day activity. H1e: the region (the federal state) makes differences in the actual CO2 emission from everyday activity.

H1f: the political party that the respondent supports makes differences in the actual CO2 emission from everyday activity.

2. The second dependent variable: cons H2a: age makes differences in the consumers' belief about CO2 emission from everyday activity.

H2b: income makes differences in the consumers' belief about CO2 emission from everyday activity.

H2c: education level makes differences in the consumers' belief about CO2 emission from everyday activity.

H2d: the place of residence (city or countryside) makes differences in the consumers' belief about CO2 emission from everyday activity.

H2e: the region (the federal state) makes differences in the consumers' belief about CO2 emission from everyday activity.

H2f: the political party that the respondent supports makes differences in the consumers' belief about CO2 emission from everyday activity.

Independent variables in the dataset

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

Dependent variables in the dataset

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

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

2. Belief about CO2

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

Data preparation

```
# change into categorical variable
```

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

```
## Select the classification for the urban_rural
```

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

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

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

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

```
head(df1_1)
```

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

```

## 3          Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 4          Realschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
## 5          Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
## 6          Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
##  urban_rural_class  federal_state C02_housing C02_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
##  C02_housing_electricity C02_cruise C02_flight C02_public_transport C02_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
##  C02_car2 C02_car3 C02_car4 C02_car5 C02_car_total C02_mobility C02_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
##  C02_other_consumption public_emission C02_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: C02_mobility

```

```

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

```
df1
```

```
## # A tibble: 588 x 7
##       age income political_party      education      urban~1 feder~2 CO2_m~3
##     <int> <dbl> <fct>          <fct>          <fct>    <fct>    <dbl>
## 1     65   3000 CDU/CSU      (Fach-) Hochschula~ zentral Saarla~   3873.
## 2     59    800 Keine Angabe Allgemeine oder fa~ sehr z~ Hessen  11785.
## 3     60   1750 Keine Angabe Berufsausbildung, ~ periph~ Bayern    706.
## 4     73   2500 SPD          Realschulabschluss~ sehr z~ Bayern  3720.
## 5     43   2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin   108.
## 6     49   2300 CDU/CSU      Berufsausbildung, ~ zentral Sachse~  4222.
## 7     57    600 CDU/CSU      Realschulabschluss~ zentral Baden~     0
## 8     39   5000 SPD          (Fach-) Hochschula~ sehr z~ Berlin 17432.
## 9     62     0 Keine Angabe (Fach-) Hochschula~ sehr z~ Nordrh~   5116.
## 10    45   2600 Keine Angabe Berufsausbildung, ~ sehr z~ Hessen   4002.
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #   2: federal_state, 3: CO2_mobility
```

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

```
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, 26) #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~   -14
##  2      59     800 Keine Angabe    Allgemeine oder fa~ sehr z~ Hessen    -42
##  3      60    1750 Keine Angabe    Berufsausbildung, ~ periph~ Bayern     11
##  4      73    2500 SPD              Realschulabschluss~ sehr z~ Bayern   -31
##  5      43    2500 Einer anderen Partei Berufsausbildung, ~ sehr z~ Berlin    -2
##  6      49    2300 CDU/CSU          Berufsausbildung, ~ zentral Sachse~    22
##  7      57     600 CDU/CSU          Realschulabschluss~ zentral Baden~    72
##  8      39    5000 SPD              (Fach-) Hochschula~ sehr z~ Berlin   -67
##  9      62       0 Keine Angabe    (Fach-) Hochschula~ sehr z~ Nordrh~   -34
## 10     45    2600 Keine Angabe    Berufsausbildung, ~ sehr z~ Hessen   -48
## # ... with 578 more rows, and abbreviated variable names 1: urban_rural_class,
## #      2: federal_state, 3: belief_diff_mobility
```

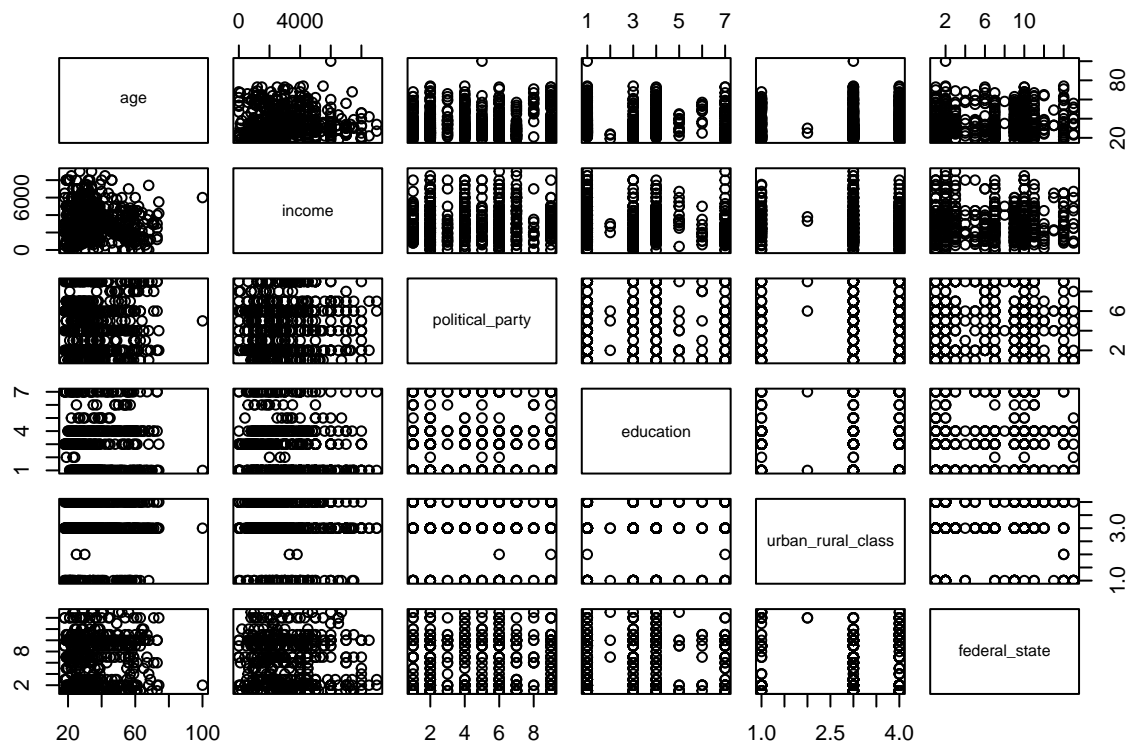
I. Exploratory Data Analysis

Check the Jupyter notebook: EDA_scatter_plot_actual_belief

II. Multivariate Regression: CO2 mobility

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

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



1. Modeling

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

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

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

```
##
## (Fach-) Hochschulabschluss (Bachelor, Master, Magister, Diplom, Staatsexamen)
##                                     253
##                                     (Noch) kein Abschluss
##                                     3
##      Allgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)
##                                     131
##      Berufsausbildung, Lehre oder Ausbildung an einer Fachschule
##                                     118
##      Doktorgrad oder Habilitation
##                                     13
##      Hauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss
##                                     11
##      Realschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss
##                                     59
```

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

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

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

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

```
## defining a reference level
```

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

```
# regression model with all variables
```

```
model1 <- lm(CO2_mobility ~ age + income + political_party + education + urban_rural_class + federal_s
summary(model1)
```

```
##
## Call:
## lm(formula = CO2_mobility ~ age + income + political_party +
##     education + urban_rural_class + federal_state, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18265   -5114   -1762    1723   160734
##
## Coefficients:
##                                     Estimate
## (Intercept)                      2415.9146
## age                             -59.4430
## income                           1.3643
## political_partyAfD                1472.0991
## political_partyBündnis Sarah Wagenknecht 2244.2059
## political_partyCDU/CSU             8863.2082
## political_partyDie Linke           2017.9320
## political_partyEiner anderen Partei   -160.9017
## political_partyFDP                 1336.0133
## political_partyKeine Angabe         2924.6264
## political_partySPD                 3528.9380
## education(Noch) kein Abschluss      -4483.1467
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -1877.2886
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule -2638.3684
## educationDoktorgrad oder Habilitation -3947.8836
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss -3292.3581
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss -455.4774
## urban_rural_classperipher          -1722.1871
## urban_rural_classsehr peripher     -852.5398
## urban_rural_classzentral           -759.9172
## federal_stateBaden-Württemberg      238.0601
## federal_stateBayern                 2731.2880
## federal_stateBerlin                 3765.8099
## federal_stateBrandenburg           -2378.3509
## federal_stateBremen                -2249.4386
## federal_stateHamburg                -641.7278
## federal_stateHessen                 2105.9810
## federal_stateMecklenburg-Vorpommern -7976.0407
## federal_stateNiedersachsen           277.2367
## federal_stateRheinland-Pfalz       4647.7752
## federal_stateSaarland              -884.1904
```

## federal_stateSachsen-Anhalt	-4135.1981
## federal_stateSchleswig-Holstein	2005.8491
## federal_stateThüringen	2134.5338
##	Std. Error
## (Intercept)	2818.3587
## age	49.8959
## income	0.3334
## political_partyAfD	2387.2884
## political_partyBündnis Sarah Wagenknecht	3346.0898
## political_partyCDU/CSU	2144.1498
## political_partyDie Linke	2584.3250
## political_partyEiner anderen Partei	1918.1878
## political_partyFDP	2483.2531
## political_partyKeine Angabe	4301.3108
## political_partySPD	2183.4284
## education(Noch) kein Abschluss	8735.3943
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	1698.0096
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	1732.2233
## educationDoktorgrad oder Habilitation	4234.3684
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	4833.4884
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	2202.8791
## urban_rural_classperipher	2219.8792
## urban_rural_classsehr peripher	10989.5692
## urban_rural_classzentral	1627.4842
## federal_stateBaden-Württemberg	2114.1490
## federal_stateBayern	2170.6354
## federal_stateBerlin	2636.9553
## federal_stateBrandenburg	5538.6310
## federal_stateBremen	4039.8693
## federal_stateHamburg	3287.6371
## federal_stateHessen	2522.3990
## federal_stateMecklenburg-Vorpommern	10587.9037
## federal_stateNiedersachsen	2566.1131
## federal_stateRheinland-Pfalz	3165.0059
## federal_stateSaarland	4969.8636
## federal_stateSachsen-Anhalt	7660.9562
## federal_stateSchleswig-Holstein	3701.9063
## federal_stateThüringen	5611.0920
##	t value
## (Intercept)	0.857
## age	-1.191
## income	4.092
## political_partyAfD	0.617
## political_partyBündnis Sarah Wagenknecht	0.671
## political_partyCDU/CSU	4.134
## political_partyDie Linke	0.781
## political_partyEiner anderen Partei	-0.084
## political_partyFDP	0.538
## political_partyKeine Angabe	0.680
## political_partySPD	1.616
## education(Noch) kein Abschluss	-0.513
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	-1.106
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	-1.523
## educationDoktorgrad oder Habilitation	-0.932

## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	-0.681
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	-0.207
## urban_rural_classperipher	-0.776
## urban_rural_classsehr peripher	-0.078
## urban_rural_classzentral	-0.467
## federal_stateBaden-Württemberg	0.113
## federal_stateBayern	1.258
## federal_stateBerlin	1.428
## federal_stateBrandenburg	-0.429
## federal_stateBremen	-0.557
## federal_stateHamburg	-0.195
## federal_stateHessen	0.835
## federal_stateMecklenburg-Vorpommern	-0.753
## federal_stateNiedersachsen	0.108
## federal_stateRheinland-Pfalz	1.468
## federal_stateSaarland	-0.178
## federal_stateSachsen-Anhalt	-0.540
## federal_stateSchleswig-Holstein	0.542
## federal_stateThüringen	0.380
##	Pr(> t)
## (Intercept)	0.392
## age	0.234
## income	4.92e-05
## political_partyAfD	0.538
## political_partyBündnis Sarah Wagenknecht	0.503
## political_partyCDU/CSU	4.12e-05
## political_partyDie Linke	0.435
## political_partyEiner anderen Partei	0.933
## political_partyFDP	0.591
## political_partyKeine Angabe	0.497
## political_partySPD	0.107
## education(Noch) kein Abschluss	0.608
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	0.269
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	0.128
## educationDoktorgrad oder Habilitation	0.352
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	0.496
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	0.836
## urban_rural_classperipher	0.438
## urban_rural_classsehr peripher	0.938
## urban_rural_classzentral	0.641
## federal_stateBaden-Württemberg	0.910
## federal_stateBayern	0.209
## federal_stateBerlin	0.154
## federal_stateBrandenburg	0.668
## federal_stateBremen	0.578
## federal_stateHamburg	0.845
## federal_stateHessen	0.404
## federal_stateMecklenburg-Vorpommern	0.452
## federal_stateNiedersachsen	0.914
## federal_stateRheinland-Pfalz	0.143
## federal_stateSaarland	0.859
## federal_stateSachsen-Anhalt	0.590
## federal_stateSchleswig-Holstein	0.588
## federal_stateThüringen	0.704

```
##
## (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: 14590 on 554 degrees of freedom
## Multiple R-squared:  0.1012, Adjusted R-squared:  0.04768
## F-statistic: 1.891 on 33 and 554 DF,  p-value: 0.002282
```

```
# Checking the VIFs for multicollinearity
```

```
vif(model1)
```

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

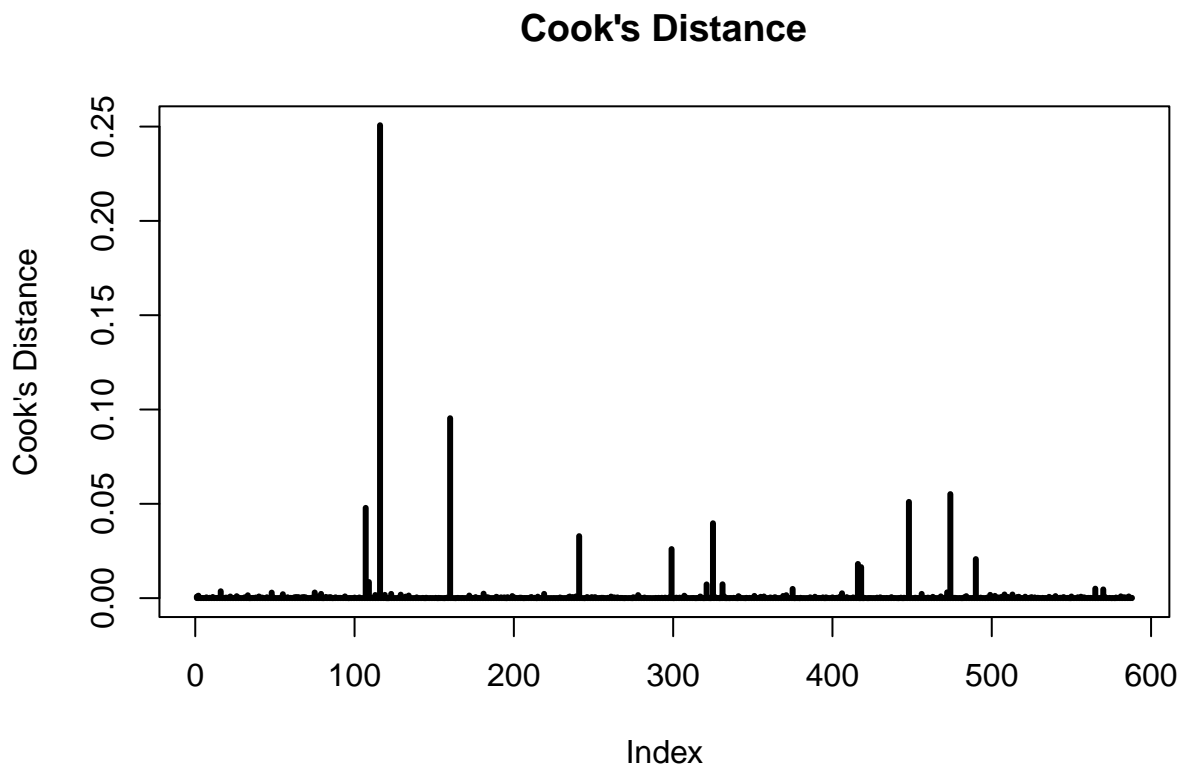
```
# threshold for multicollinearity
# Calculating the threshold

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

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

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

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



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

```
##          107          109          116          160          241          299
## 0.047854479 0.008552693 0.250716388 0.095387356 0.032863228 0.026013831
##          321          325          331          375          416          418
## 0.007283266 0.039692369 0.007329553 0.004883986 0.018121180 0.016440236
```

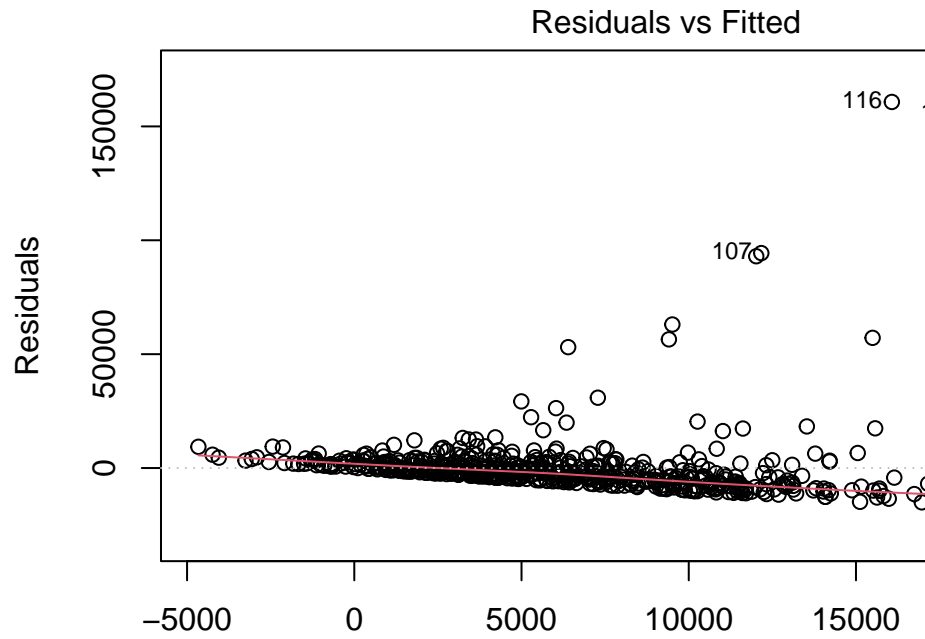
```
##          448          474          490          565          570
## 0.051016645 0.055146060 0.020689502 0.005023236 0.004619679
```

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

df1[influential_vector, ]
```

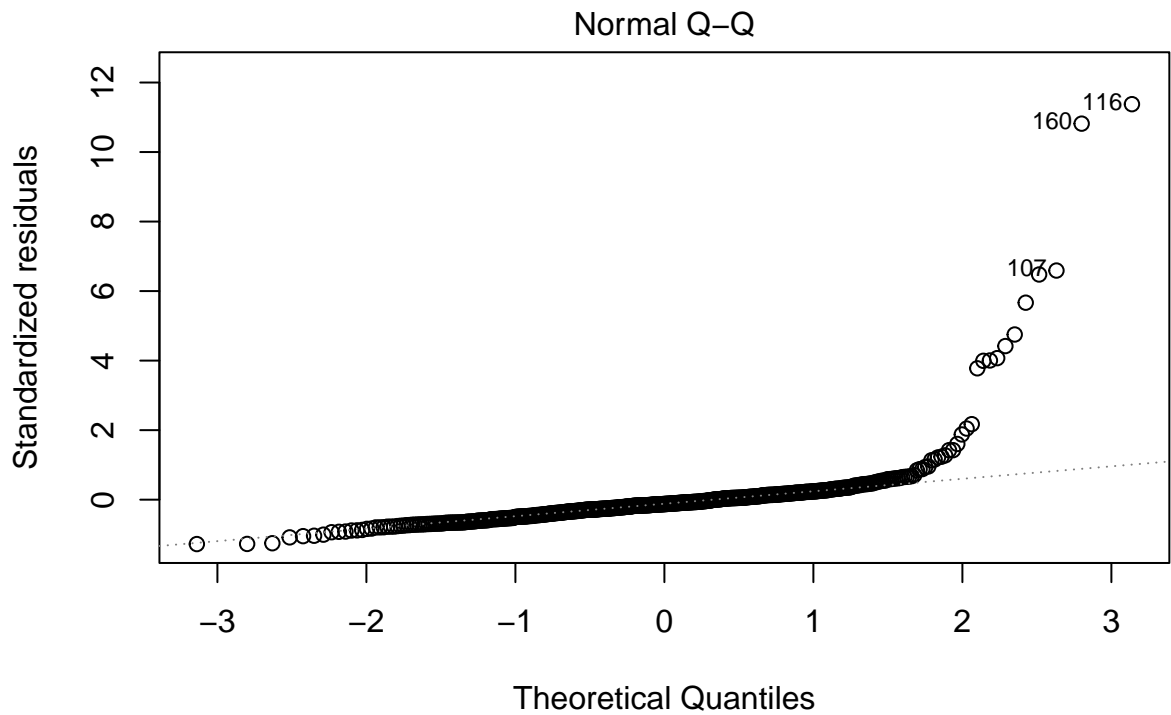
```
## # A tibble: 17 x 7
##   age income political_party education urban~1 feder~2 CO2_m~3
##   <int> <dbl> <fct> <fct> <fct> <fct> <dbl>
## 1 32 7000 Bündnis 90/Die Grünen (Fach-) Hochs~ sehr z~ Hessen 106545.
## 2 22 600 FDP Allgemeine od~ sehr z~ Rheinl~ 32330.
## 3 23 2000 CDU/CSU Realschulabsc~ zentral Rheinl~ 176807.
## 4 29 4500 CDU/CSU (Fach-) Hochs~ sehr z~ Bayern 174051.
## 5 21 5000 Bündnis 90/Die Grünen Allgemeine od~ periph~ Schles~ 59468.
## 6 43 3500 SPD (Fach-) Hochs~ zentral Hessen 72539.
## 7 24 1200 SPD Berufsausbild~ sehr z~ Berlin 38159.
## 8 26 1500 CDU/CSU (Fach-) Hochs~ sehr z~ Baden-- 105013.
## 9 59 3500 Bündnis Sarah Wagenknecht (Fach-) Hochs~ sehr z~ Hamburg 27598.
## 10 38 4000 Die Linke Berufsausbild~ sehr z~ Nordrh~ 34256.
## 11 31 5000 CDU/CSU (Fach-) Hochs~ zentral Nordrh~ 72664.
## 12 40 8000 CDU/CSU (Fach-) Hochs~ sehr z~ Bayern 79958.
## 13 100 6000 Die Linke (Fach-) Hochs~ sehr z~ Bayern 65904.
## 14 19 8000 SPD Allgemeine od~ sehr z~ Berlin 97951.
## 15 34 7000 CDU/CSU (Fach-) Hochs~ sehr z~ Bayern 89783.
## 16 26 2500 Bündnis Sarah Wagenknecht Realschulabsc~ sehr z~ Nieder~ 26291.
## 17 25 8000 Bündnis Sarah Wagenknecht (Fach-) Hochs~ sehr z~ Brande~ 640.
## # ... with abbreviated variable names 1: urban_rural_class, 2: federal_state,
## # 3: CO2_mobility
```

```
plot(model1)
```

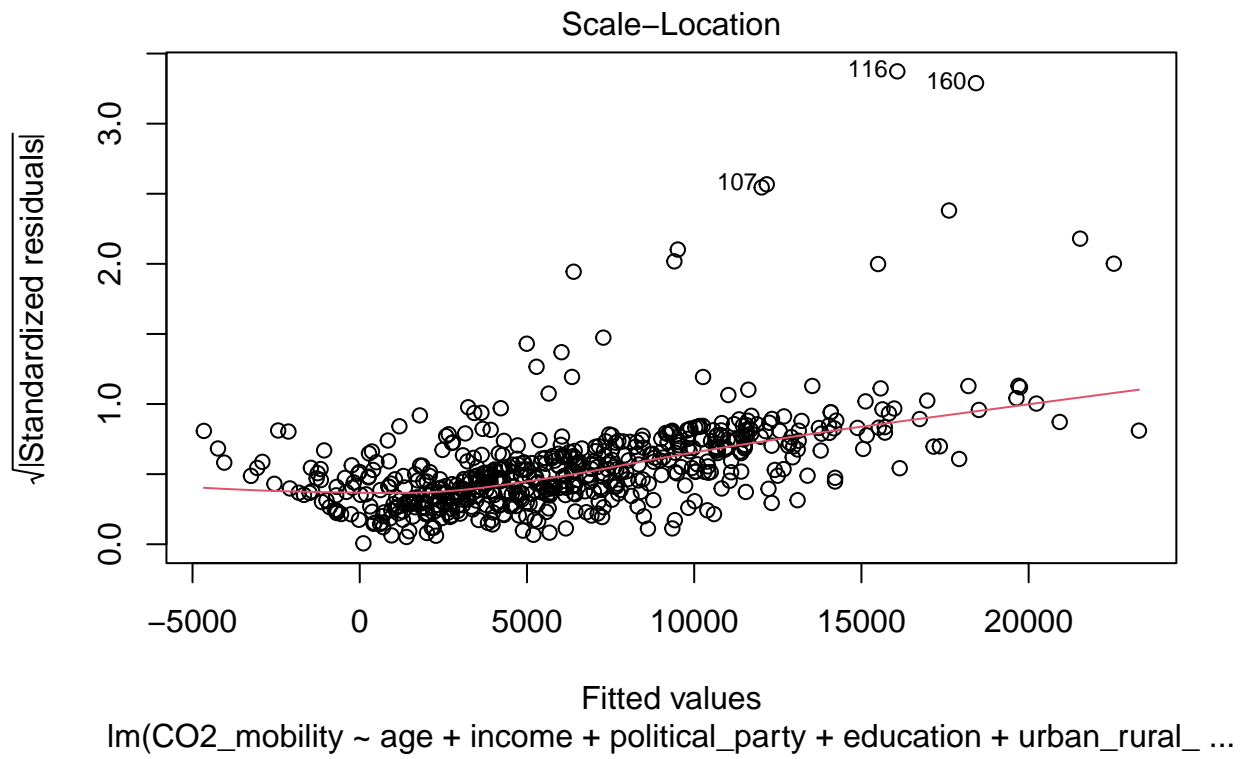



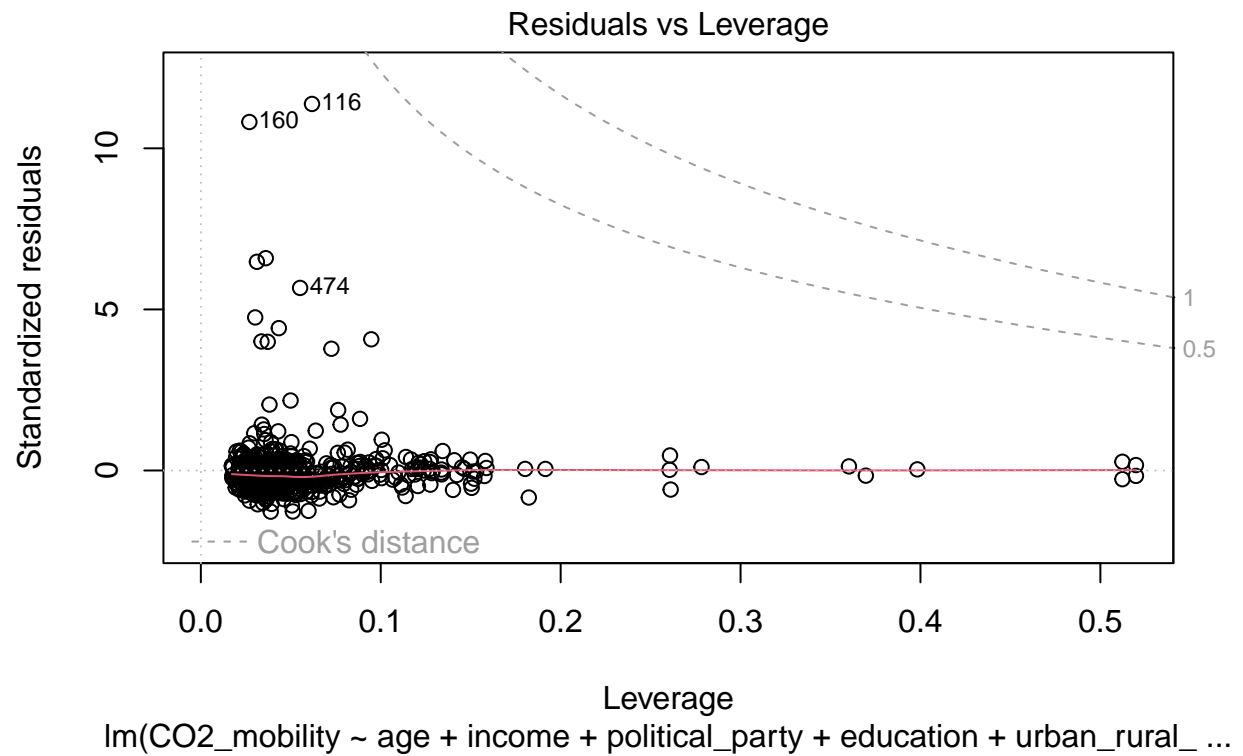
2. Assumptions check in the residuals

$\ln(\text{CO2_mobility} \sim \text{age} + \text{income} + \text{political_party} + \text{education} + \text{urban_rural_} \dots)$



$\ln(\text{CO2_mobility} \sim \text{age} + \text{income} + \text{political_party} + \text{education} + \text{urban_rural_} \dots)$

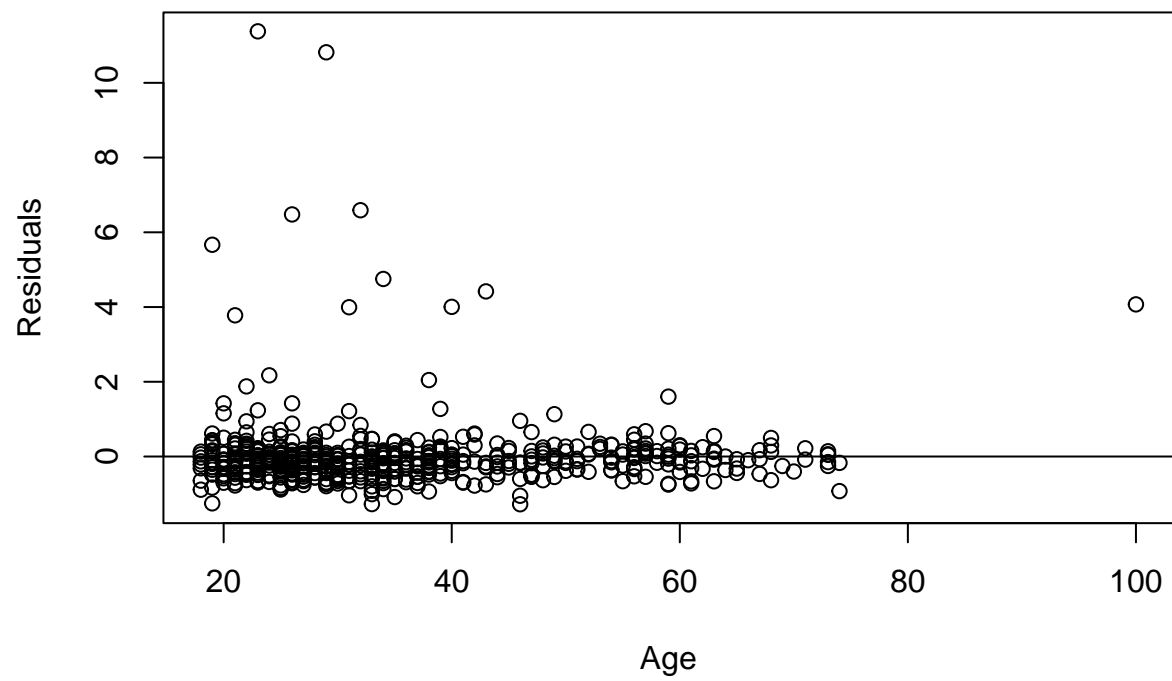




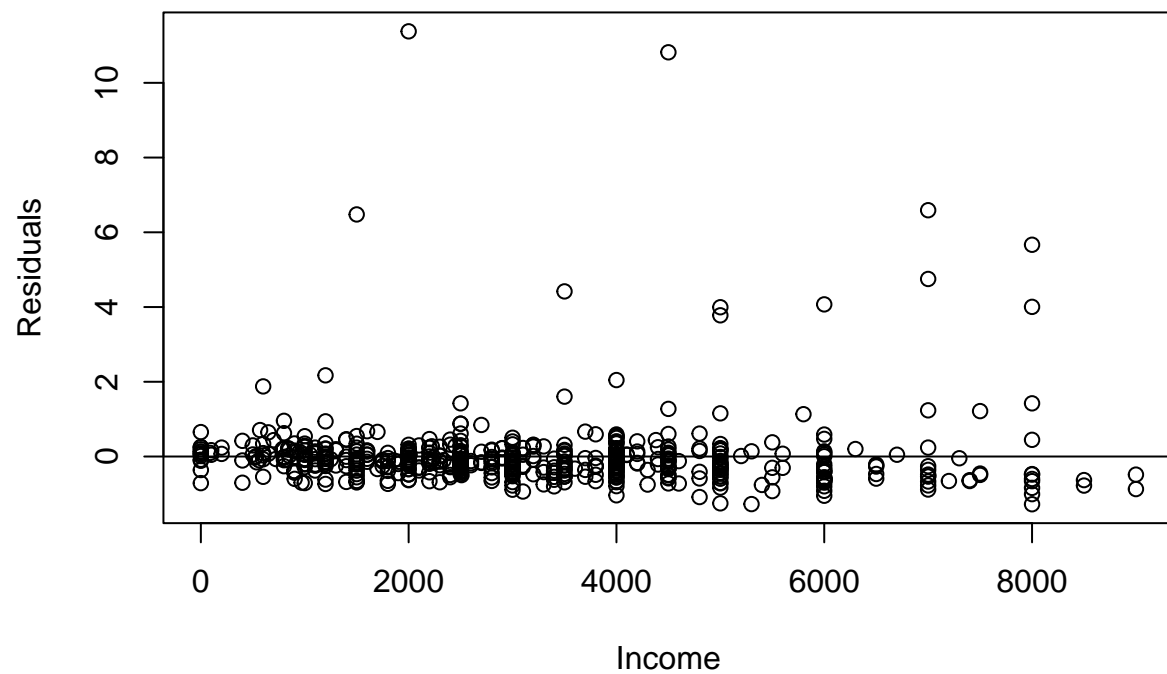
```
res1 = stdres(model1) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

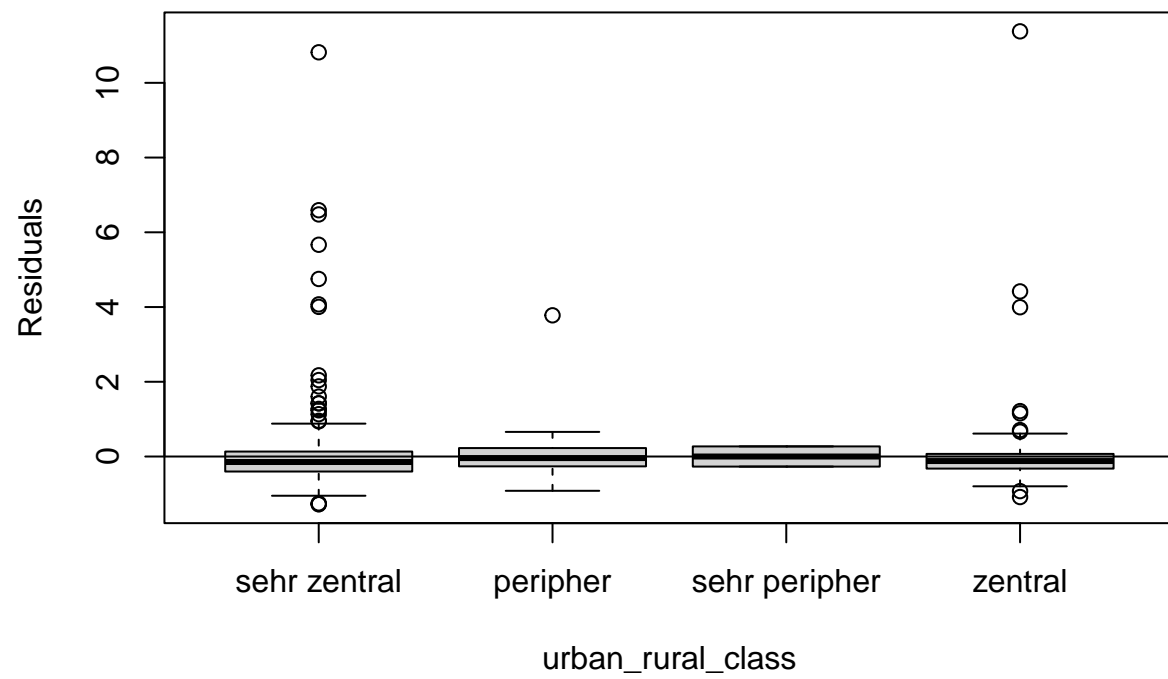
plot(df1$age, res1, xlab = "Age", ylab = "Residuals")
abline(h = 0)
```



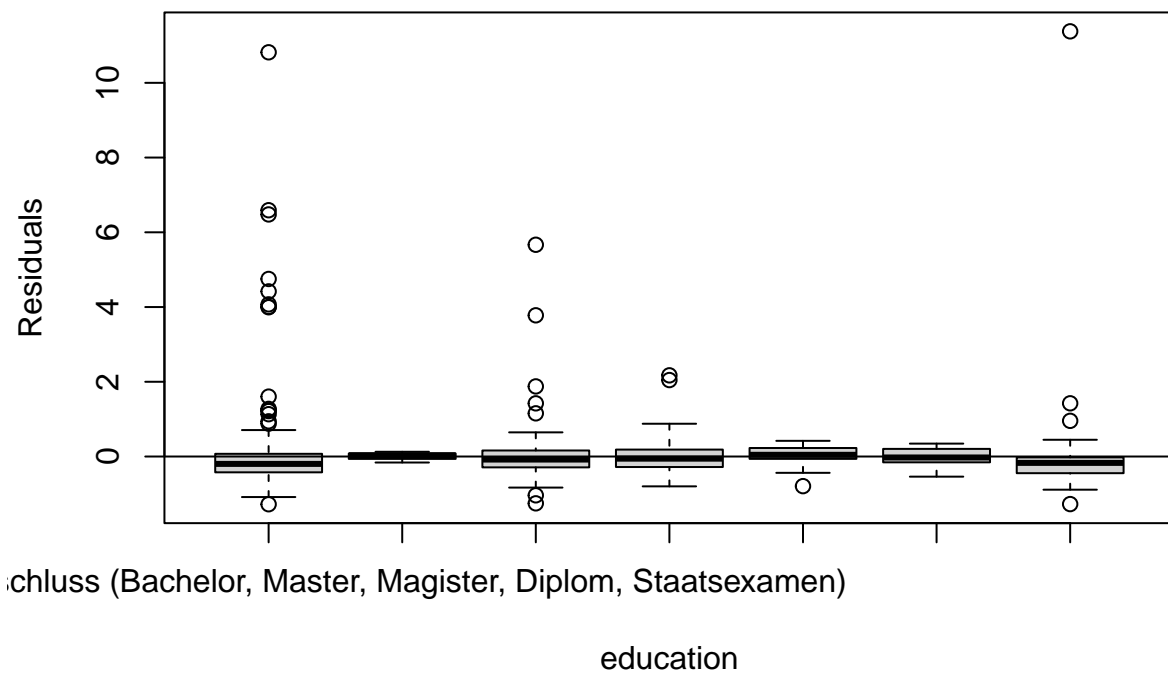
```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")  
abline(h = 0)
```



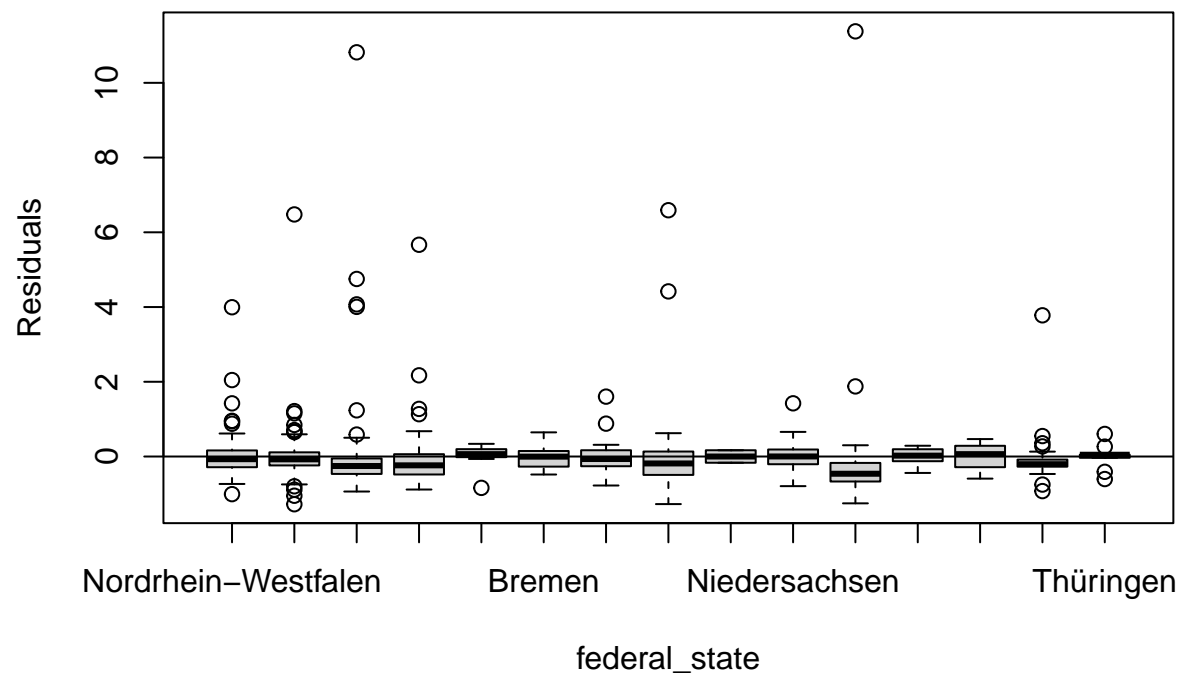
```
plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")  
abline(h = 0)
```



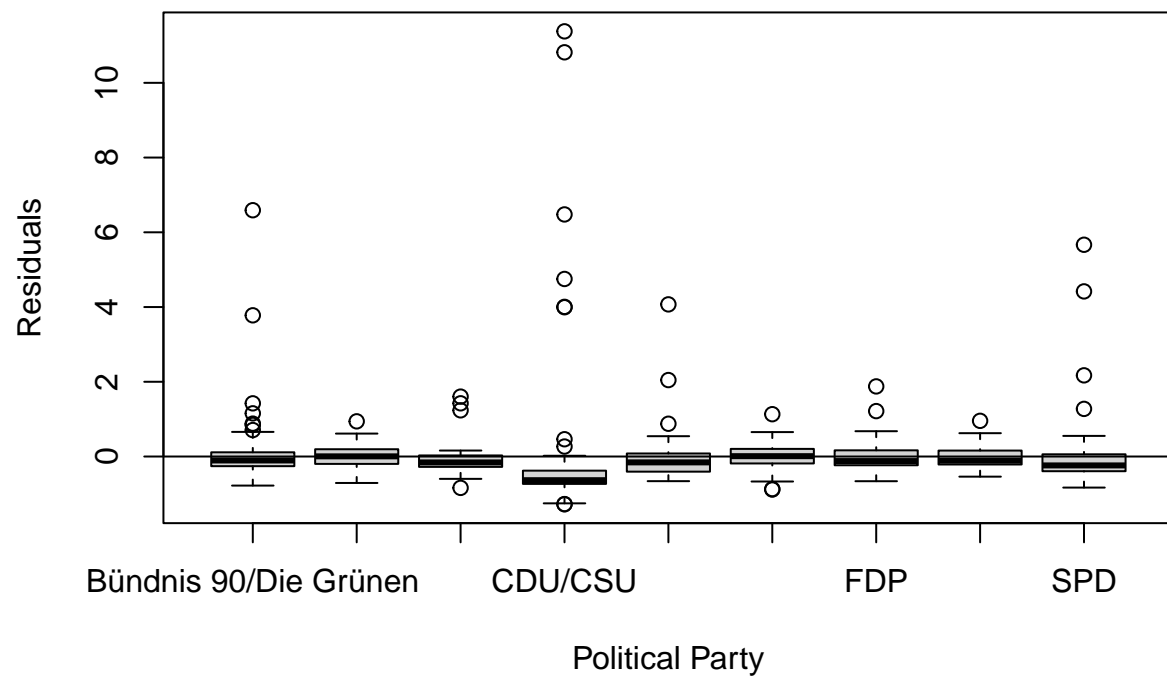
```
plot(df1$education, res1, xlab = "education", ylab = "Residuals")
abline(h = 0)
```



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

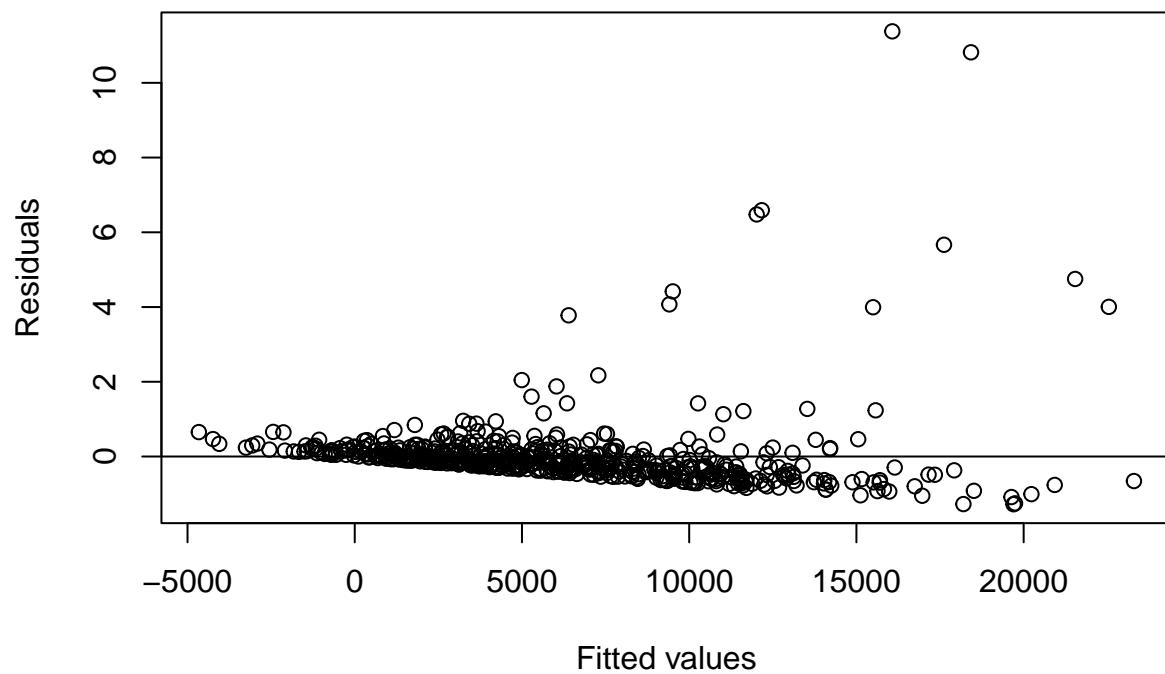


```
plot(df1$political_party, res1, xlab = "Political Party", ylab = "Residuals")
abline(h = 0)
```

Constant variance and independent error term assumption

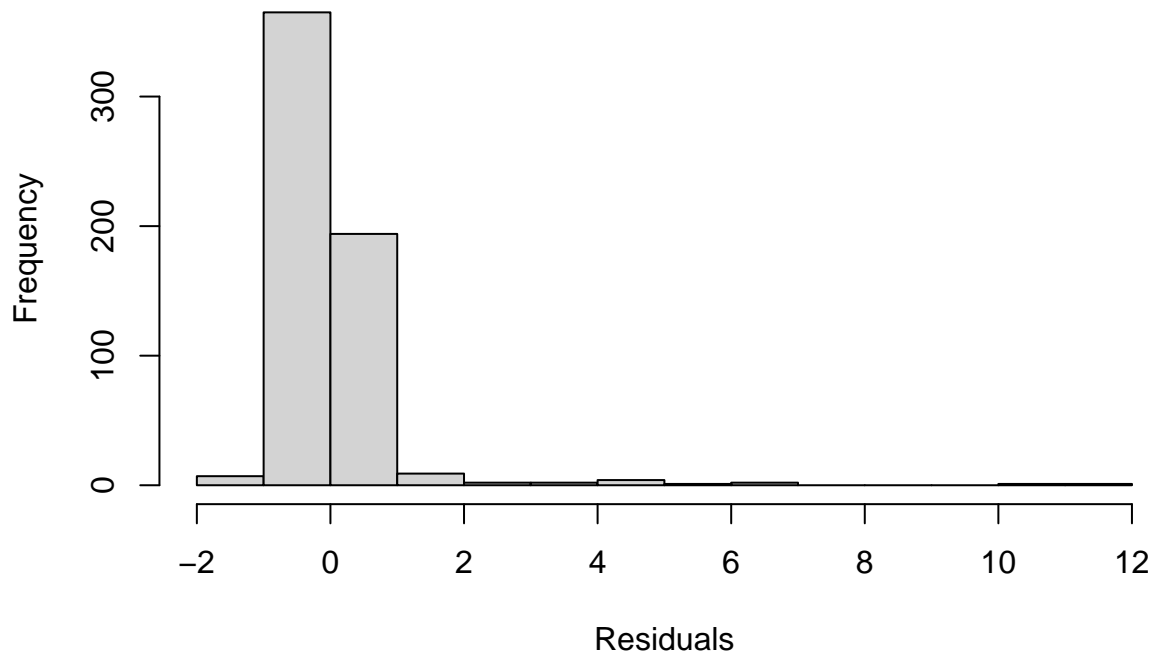
```
plot(fitted(model1), res1, xlab = "Fitted values", ylab = "Residuals")  
abline(h = 0)
```



```
# Normality assumption
```

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

Histogram of Residuals



```
## remove any 0 values in order to run a regression for Box-cox transformation
```

```
df1_new <- df1 %>% filter(df1$CO2_mobility != 0)
nrow(df1_new)
```

3. Improving the regression fit

```
## [1] 584
```

```
# Check the rows where the values are 0
# There are four data points where the CO2 mobility values are 0
```

```
df1 %>% filter_all(any_vars(. %in% c(0)))
```

```
## # A tibble: 17 x 7
##   age income political_party education urban~1 feder~2 CO2_m~3
##   <int> <dbl> <fct> <fct> <fct> <fct> <dbl>
## 1 57 600 CDU/CSU Realschulabsc~ zentral Baden~~ 0
## 2 62 0 Keine Angabe (Fach-) Hochs~ sehr z~ Nordrh~ 5116.
## 3 53 0 Bündnis 90/Die Grünen (Fach-) Hochs~ zentral Baden~~ 2431.
## 4 23 0 FDP (Fach-) Hochs~ zentral Baden~~ 640.
## 5 47 0 Einer anderen Partei Berufsausbild~ periph~ Baden~~ 4667.
```

```
## 6 58 0 CDU/CSU Realschulabsc~ zentral Bayern 9528.
## 7 68 0 SPD Allgemeine od~ sehr z~ Nordrh~ 1791.
## 8 54 0 CDU/CSU (Fach-) Hochs~ zentral Baden~ 2367.
## 9 18 0 Einer anderen Partei Realschulabsc~ periph~ Nieder~ 119.
## 10 23 0 Bündnis 90/Die Grünen (Fach-) Hochs~ zentral Hessen 4346.
## 11 23 0 Die Linke (Fach-) Hochs~ sehr z~ Baden~ 6077.
## 12 34 1000 CDU/CSU Realschulabsc~ sehr z~ Nordrh~ 0
## 13 21 0 CDU/CSU Allgemeine od~ sehr z~ Hessen 108.
## 14 38 200 Einer anderen Partei Allgemeine od~ zentral Saarla~ 0
## 15 25 0 Bündnis 90/Die Grünen (Fach-) Hochs~ zentral Hessen 640.
## 16 21 0 Bündnis Sarah Wagenknecht Allgemeine od~ zentral Baden~ 2165.
## 17 27 1000 Einer anderen Partei Realschulabsc~ sehr z~ Bremen 0
## # ... with abbreviated variable names 1: urban_rural_class, 2: federal_state,
## # 3: CO2_mobility
```

```
# Regression with all variables after dropping data point having 0 values in CO2 mobility
```

```
modell1_1 = lm(CO2_mobility ~ age+ income + political_party + education + urban_rural_class + federal_state, data = df1_new)
summary(modell1_1)
```

```
##
## Call:
## lm(formula = CO2_mobility ~ age + income + political_party +
##     education + urban_rural_class + federal_state, data = df1_new)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18563   -5126   -1739    1705   160249
##
## Coefficients:
##                                     Estimate
## (Intercept)                        2516.9978
## age                               -58.8292
## income                             1.3474
## political_partyAfD                  1422.3814
## political_partyBündnis Sarah Wagenknecht 2195.7572
## political_partyCDU/CSU               9100.3126
## political_partyDie Linke             2019.1050
## political_partyEiner anderen Partei   -229.5201
## political_partyFDP                   1332.0179
## political_partyKeine Angabe          2881.3815
## political_partySPD                   3518.8186
## education(Noch) kein Abschluss       -4481.4048
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) -1912.4756
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule -2644.1833
## educationDoktorgrad oder Habilitation -3988.5052
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss -3249.8567
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss -148.7563
## urban_rural_classperipher            -1730.7286
## urban_rural_classsehr peripher        -952.8969
## urban_rural_classzentral              -755.7879
## federal_stateBaden-Württemberg        227.5019
## federal_stateBayern                   2623.9383
```

## federal_stateBerlin	3702.6327
## federal_stateBrandenburg	-2421.7616
## federal_stateBremen	-2396.2889
## federal_stateHamburg	-734.1947
## federal_stateHessen	2002.6802
## federal_stateMecklenburg-Vorpommern	-8138.0089
## federal_stateNiedersachsen	163.1776
## federal_stateRheinland-Pfalz	4503.6841
## federal_stateSaarland	-1342.3336
## federal_stateSachsen-Anhalt	-4268.0242
## federal_stateSchleswig-Holstein	1934.2351
## federal_stateThüringen	2094.9403
##	Std. Error
## (Intercept)	2837.0611
## age	50.1876
## income	0.3356
## political_partyAfD	2396.5888
## political_partyBündnis Sarah Wagenknecht	3356.1186
## political_partyCDU/CSU	2166.1266
## political_partyDie Linke	2594.6478
## political_partyEiner anderen Partei	1933.7181
## political_partyFDP	2490.4198
## political_partyKeine Angabe	4313.8494
## political_partySPD	2190.1739
## education(Noch) kein Abschluss	8760.7023
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	1706.7855
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	1737.6938
## educationDoktorgrad oder Habilitation	4247.0040
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	4848.1509
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	2248.9215
## urban_rural_classperipher	2226.7759
## urban_rural_classsehr peripher	11023.1126
## urban_rural_classzentral	1636.7055
## federal_stateBaden-Württemberg	2128.3649
## federal_stateBayern	2180.8096
## federal_stateBerlin	2646.9930
## federal_stateBrandenburg	5555.0918
## federal_stateBremen	4181.0736
## federal_stateHamburg	3299.4909
## federal_stateHessen	2533.0135
## federal_stateMecklenburg-Vorpommern	10619.9705
## federal_stateNiedersachsen	2576.8127
## federal_stateRheinland-Pfalz	3178.3780
## federal_stateSaarland	5228.3645
## federal_stateSachsen-Anhalt	7684.3851
## federal_stateSchleswig-Holstein	3713.9156
## federal_stateThüringen	5628.7113
##	t value
## (Intercept)	0.887
## age	-1.172
## income	4.014
## political_partyAfD	0.594
## political_partyBündnis Sarah Wagenknecht	0.654
## political_partyCDU/CSU	4.201

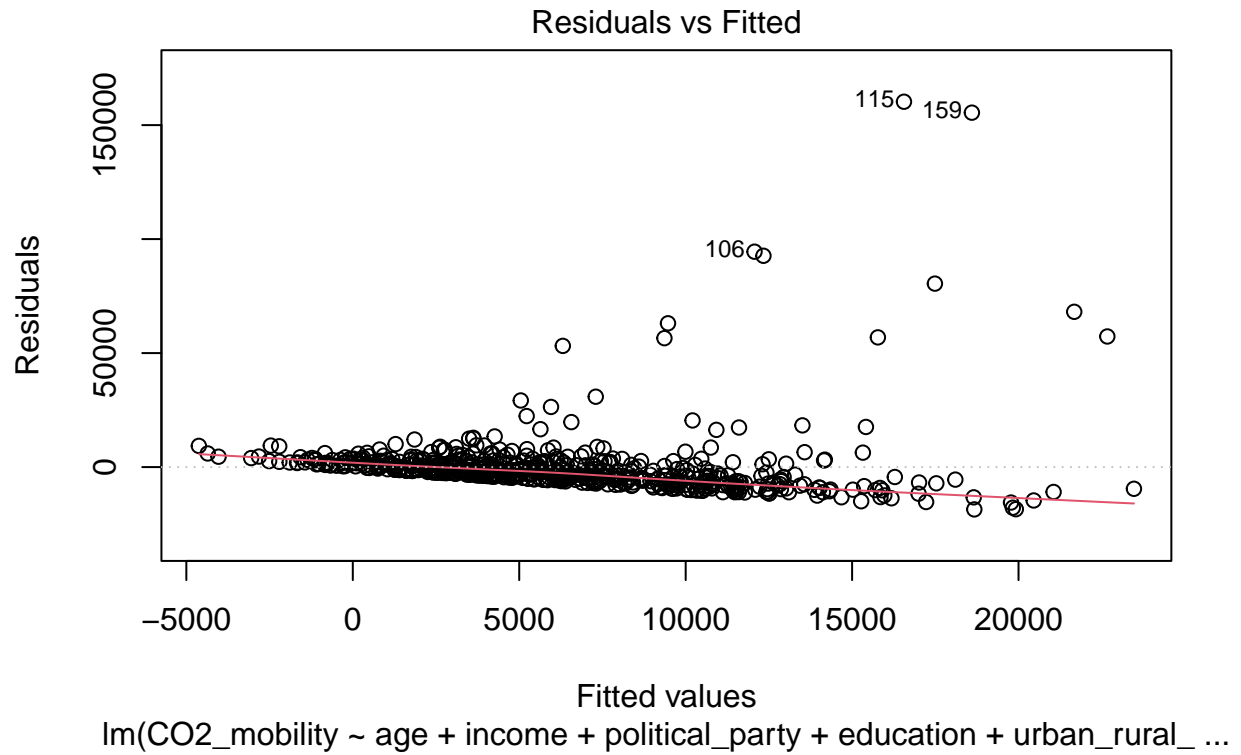
## political_partyDie Linke	0.778
## political_partyEiner anderen Partei	-0.119
## political_partyFDP	0.535
## political_partyKeine Angabe	0.668
## political_partySPD	1.607
## education(Noch) kein Abschluss	-0.512
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	-1.121
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	-1.522
## educationDoktorgrad oder Habilitation	-0.939
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	-0.670
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	-0.066
## urban_rural_classperipher	-0.777
## urban_rural_classsehr peripher	-0.086
## urban_rural_classzentral	-0.462
## federal_stateBaden-Württemberg	0.107
## federal_stateBayern	1.203
## federal_stateBerlin	1.399
## federal_stateBrandenburg	-0.436
## federal_stateBremen	-0.573
## federal_stateHamburg	-0.223
## federal_stateHessen	0.791
## federal_stateMecklenburg-Vorpommern	-0.766
## federal_stateNiedersachsen	0.063
## federal_stateRheinland-Pfalz	1.417
## federal_stateSaarland	-0.257
## federal_stateSachsen-Anhalt	-0.555
## federal_stateSchleswig-Holstein	0.521
## federal_stateThüringen	0.372
##	Pr(> t)
## (Intercept)	0.375
## age	0.242
## income	6.78e-05
## political_partyAfD	0.553
## political_partyBündnis Sarah Wagenknecht	0.513
## political_partyCDU/CSU	3.10e-05
## political_partyDie Linke	0.437
## political_partyEiner anderen Partei	0.906
## political_partyFDP	0.593
## political_partyKeine Angabe	0.504
## political_partySPD	0.109
## education(Noch) kein Abschluss	0.609
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	0.263
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	0.129
## educationDoktorgrad oder Habilitation	0.348
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	0.503
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	0.947
## urban_rural_classperipher	0.437
## urban_rural_classsehr peripher	0.931
## urban_rural_classzentral	0.644
## federal_stateBaden-Württemberg	0.915
## federal_stateBayern	0.229
## federal_stateBerlin	0.162
## federal_stateBrandenburg	0.663
## federal_stateBremen	0.567

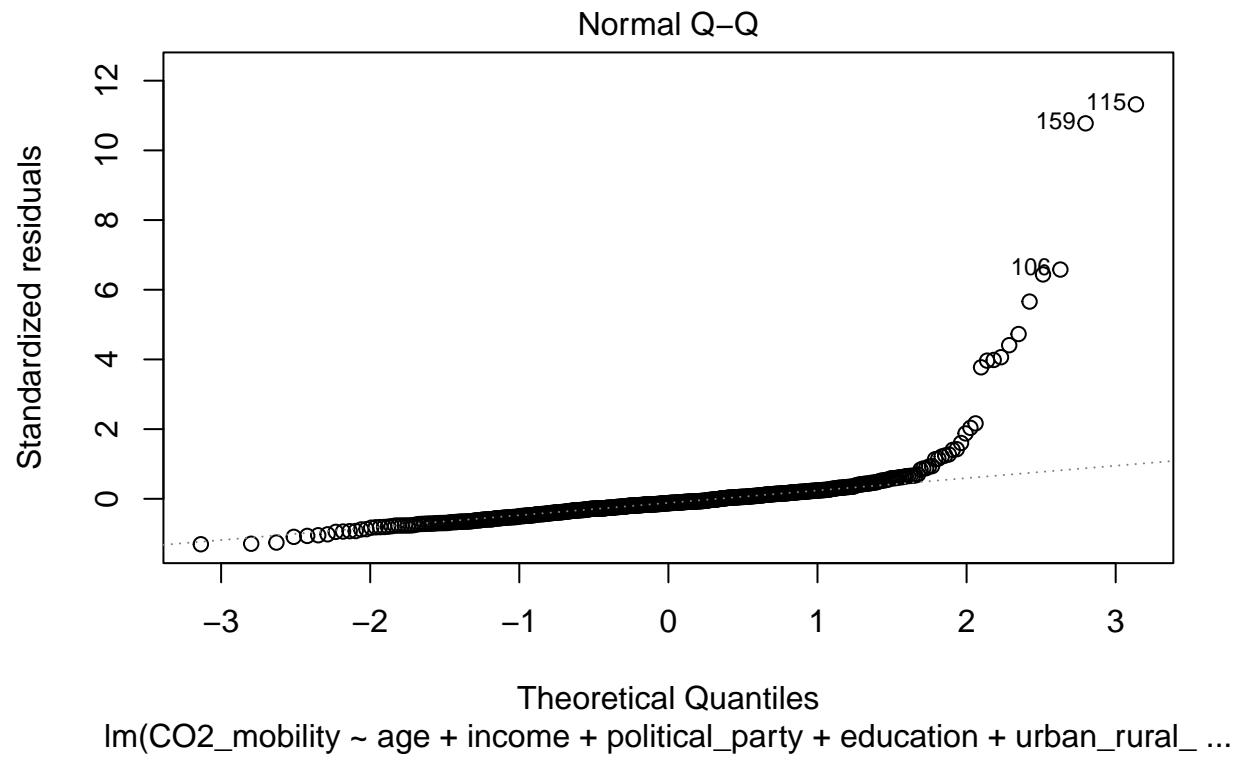
```

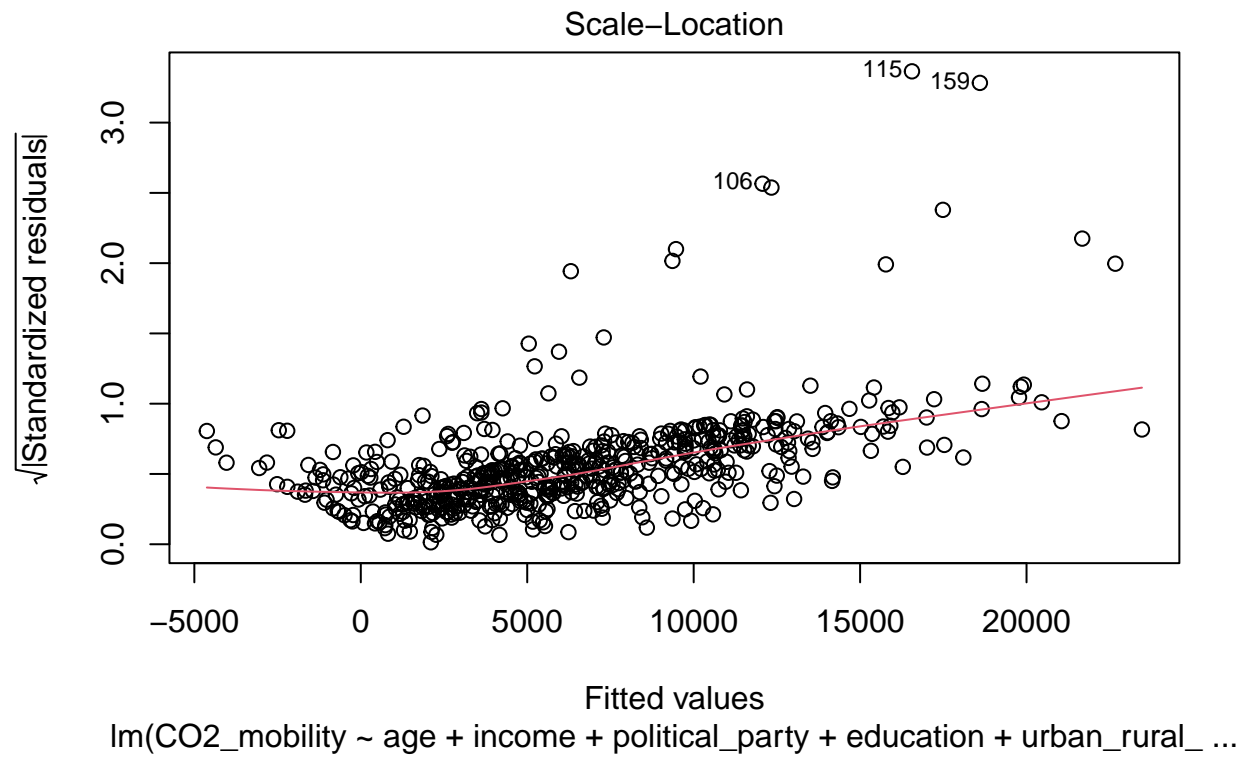
## federal_stateHamburg 0.824
## federal_stateHessen 0.429
## federal_stateMecklenburg-Vorpommern 0.444
## federal_stateNiedersachsen 0.950
## federal_stateRheinland-Pfalz 0.157
## federal_stateSaarland 0.797
## federal_stateSachsen-Anhalt 0.579
## federal_stateSchleswig-Holstein 0.603
## federal_stateThüringen 0.710
##
## (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: 14630 on 550 degrees of freedom
## Multiple R-squared:  0.1016, Adjusted R-squared:  0.04765
## F-statistic: 1.884 on 33 and 550 DF, p-value: 0.002411

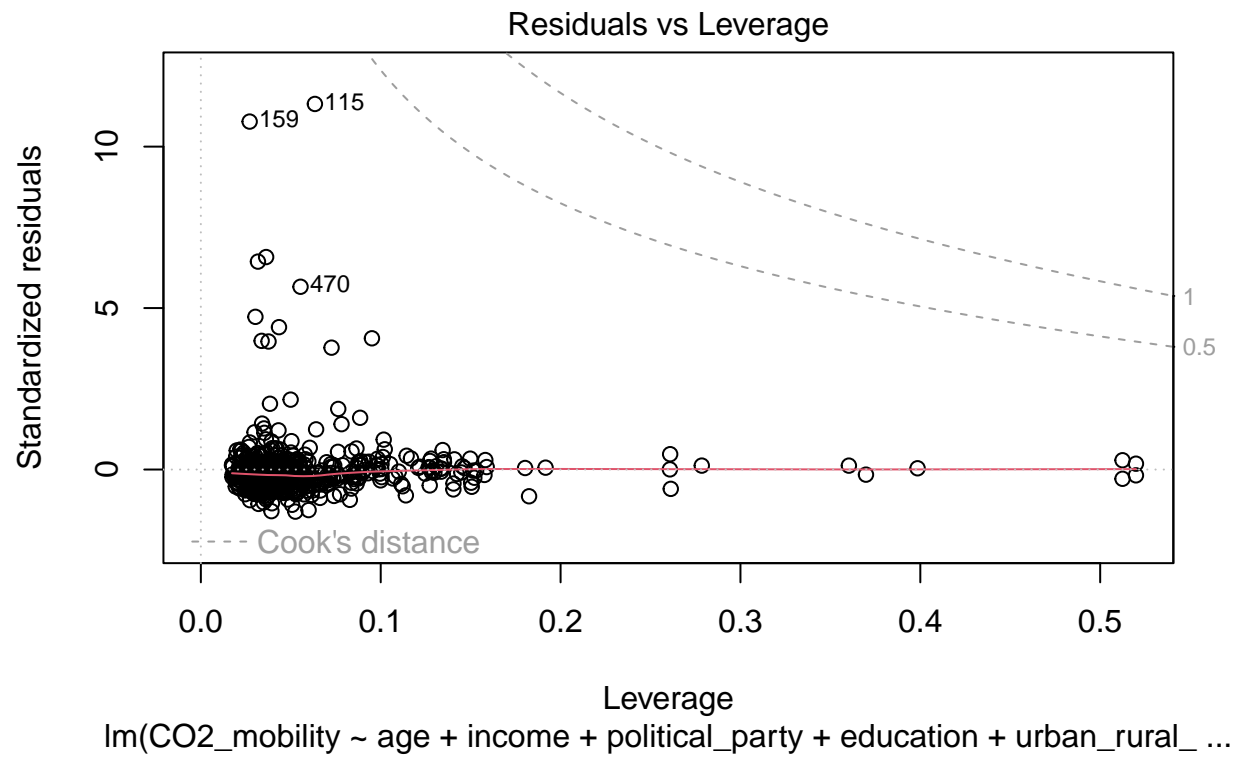
```

```
plot(model1_1)
```



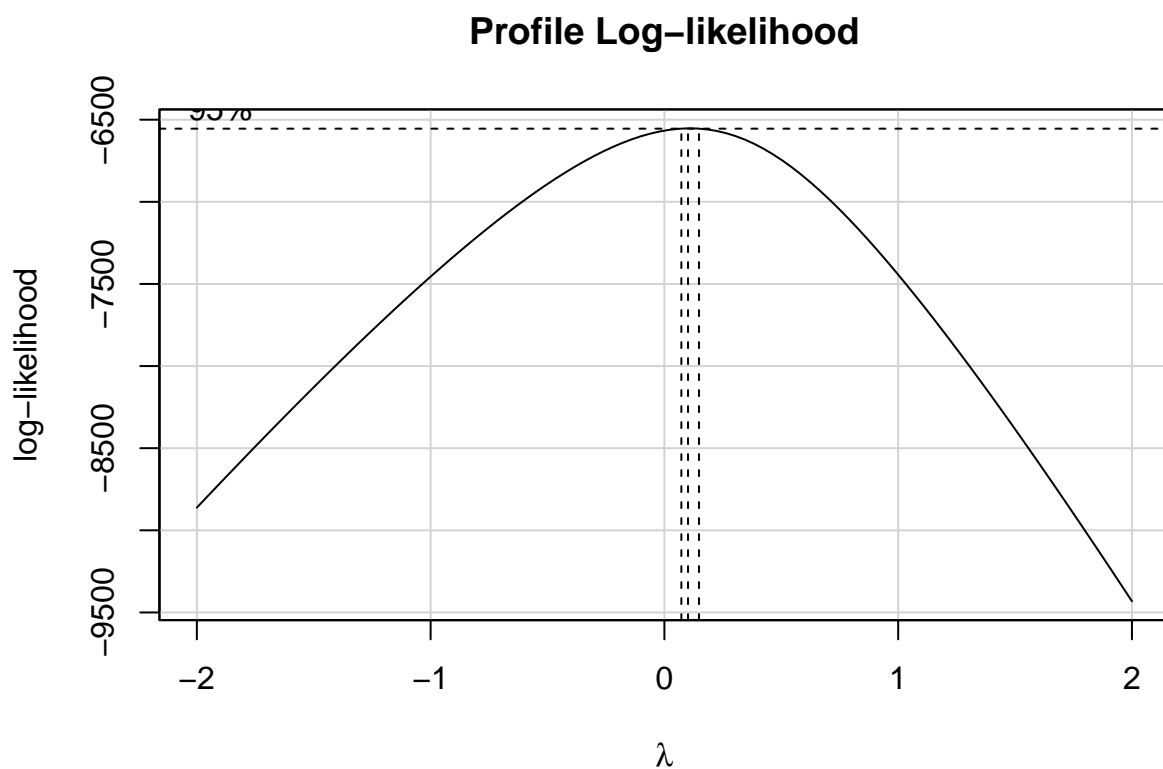






```
# box-cox transformation
```

```
bc = boxCox(model1_1)
```



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

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

```
# since the lambda value from the box-cox test is 0,  $y = \log(y)$ 
```

```
model1_2 = lm(log(CO2_mobility + 1) ~ age + income + political_party + education + urban_rural_class
summary(model1_2)
```

```
##
## Call:
## lm(formula = log(CO2_mobility + 1) ~ age + income + political_party +
##     education + urban_rural_class + federal_state, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5536 -0.7787  0.1783  0.9266  4.9405
##
## Coefficients:
##
## (Intercept)
## age
```

```
Estimate
6.6719661
-0.0044184
```

## income	0.0003073
## political_partyAfD	0.7308849
## political_partyBündnis Sarah Wagenknecht	0.4756050
## political_partyCDU/CSU	0.4907804
## political_partyDie Linke	0.5000582
## political_partyEiner anderen Partei	0.2203809
## political_partyFDP	0.7836987
## political_partyKeine Angabe	0.7288442
## political_partySPD	0.8060326
## education(Noch) kein Abschluss	-0.3624659
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	-0.3669161
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	-0.3309114
## educationDoktorgrad oder Habilitation	-0.0942231
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	-0.6507820
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	-0.7662780
## urban_rural_classperipher	0.5536884
## urban_rural_classsehr peripher	1.5074934
## urban_rural_classzentral	0.2344592
## federal_stateBaden-Württemberg	-0.2029594
## federal_stateBayern	-0.0022639
## federal_stateBerlin	0.3897379
## federal_stateBrandenburg	-1.0307906
## federal_stateBremen	-1.1441133
## federal_stateHamburg	-0.1682531
## federal_stateHessen	-0.0634752
## federal_stateMecklenburg-Vorpommern	-1.4958557
## federal_stateNiedersachsen	-0.2635256
## federal_stateRheinland-Pfalz	-0.0016887
## federal_stateSaarland	-0.8071818
## federal_stateSachsen-Anhalt	-0.2127387
## federal_stateSchleswig-Holstein	-0.2695219
## federal_stateThüringen	-0.1347985
##	Std. Error
## (Intercept)	0.2924563
## age	0.0051776
## income	0.0000346
## political_partyAfD	0.2477249
## political_partyBündnis Sarah Wagenknecht	0.3472181
## political_partyCDU/CSU	0.2224948
## political_partyDie Linke	0.2681710
## political_partyEiner anderen Partei	0.1990471
## political_partyFDP	0.2576830
## political_partyKeine Angabe	0.4463397
## political_partySPD	0.2265707
## education(Noch) kein Abschluss	0.9064570
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	0.1761996
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	0.1797499
## educationDoktorgrad oder Habilitation	0.4393932
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	0.5015629
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	0.2285890
## urban_rural_classperipher	0.2303531
## urban_rural_classsehr peripher	1.1403689
## urban_rural_classzentral	0.1688813
## federal_stateBaden-Württemberg	0.2193817

## federal_stateBayern	0.2252432
## federal_stateBerlin	0.2736324
## federal_stateBrandenburg	0.5747343
## federal_stateBremen	0.4192104
## federal_stateHamburg	0.3411525
## federal_stateHessen	0.2617451
## federal_stateMecklenburg-Vorpommern	1.0986888
## federal_stateNiedersachsen	0.2662812
## federal_stateRheinland-Pfalz	0.3284273
## federal_stateSaarland	0.5157143
## federal_stateSachsen-Anhalt	0.7949644
## federal_stateSchleswig-Holstein	0.3841405
## federal_stateThüringen	0.5822535
##	t value
## (Intercept)	22.814
## age	-0.853
## income	8.883
## political_partyAfD	2.950
## political_partyBündnis Sarah Wagenknecht	1.370
## political_partyCDU/CSU	2.206
## political_partyDie Linke	1.865
## political_partyEiner anderen Partei	1.107
## political_partyFDP	3.041
## political_partyKeine Angabe	1.633
## political_partySPD	3.558
## education(Noch) kein Abschluss	-0.400
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	-2.082
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	-1.841
## educationDoktorgrad oder Habilitation	-0.214
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	-1.298
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	-3.352
## urban_rural_classperipher	2.404
## urban_rural_classsehr peripher	1.322
## urban_rural_classzentral	1.388
## federal_stateBaden-Württemberg	-0.925
## federal_stateBayern	-0.010
## federal_stateBerlin	1.424
## federal_stateBrandenburg	-1.794
## federal_stateBremen	-2.729
## federal_stateHamburg	-0.493
## federal_stateHessen	-0.243
## federal_stateMecklenburg-Vorpommern	-1.361
## federal_stateNiedersachsen	-0.990
## federal_stateRheinland-Pfalz	-0.005
## federal_stateSaarland	-1.565
## federal_stateSachsen-Anhalt	-0.268
## federal_stateSchleswig-Holstein	-0.702
## federal_stateThüringen	-0.232
##	Pr(> t)
## (Intercept)	< 2e-16
## age	0.393823
## income	< 2e-16
## political_partyAfD	0.003308
## political_partyBündnis Sarah Wagenknecht	0.171317

## political_partyCDU/CSU	0.027807
## political_partyDie Linke	0.062752
## political_partyEiner anderen Partei	0.268697
## political_partyFDP	0.002467
## political_partyKeine Angabe	0.103051
## political_partySPD	0.000406
## education(Noch) kein Abschluss	0.689406
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	0.037765
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	0.066163
## educationDoktorgrad oder Habilitation	0.830284
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	0.194996
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	0.000856
## urban_rural_classperipher	0.016560
## urban_rural_classsehr peripher	0.186735
## urban_rural_classzentral	0.165601
## federal_stateBaden-Württemberg	0.355294
## federal_stateBayern	0.991984
## federal_stateBerlin	0.154919
## federal_stateBrandenburg	0.073437
## federal_stateBremen	0.006550
## federal_stateHamburg	0.622074
## federal_stateHessen	0.808476
## federal_stateMecklenburg-Vorpommern	0.173912
## federal_stateNiedersachsen	0.322777
## federal_stateRheinland-Pfalz	0.995899
## federal_stateSaarland	0.118113
## federal_stateSachsen-Anhalt	0.789101
## federal_stateSchleswig-Holstein	0.483209
## federal_stateThüringen	0.817003
##	
## (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: 1.513 on 554 degrees of freedom
## Multiple R-squared:  0.2285, Adjusted R-squared:  0.1826
## F-statistic: 4.973 on 33 and 554 DF,  p-value: < 2.2e-16
```

```
# Checking the VIFs for multicollinearity
```

```
vif(model1_2)
```

```
##              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
```

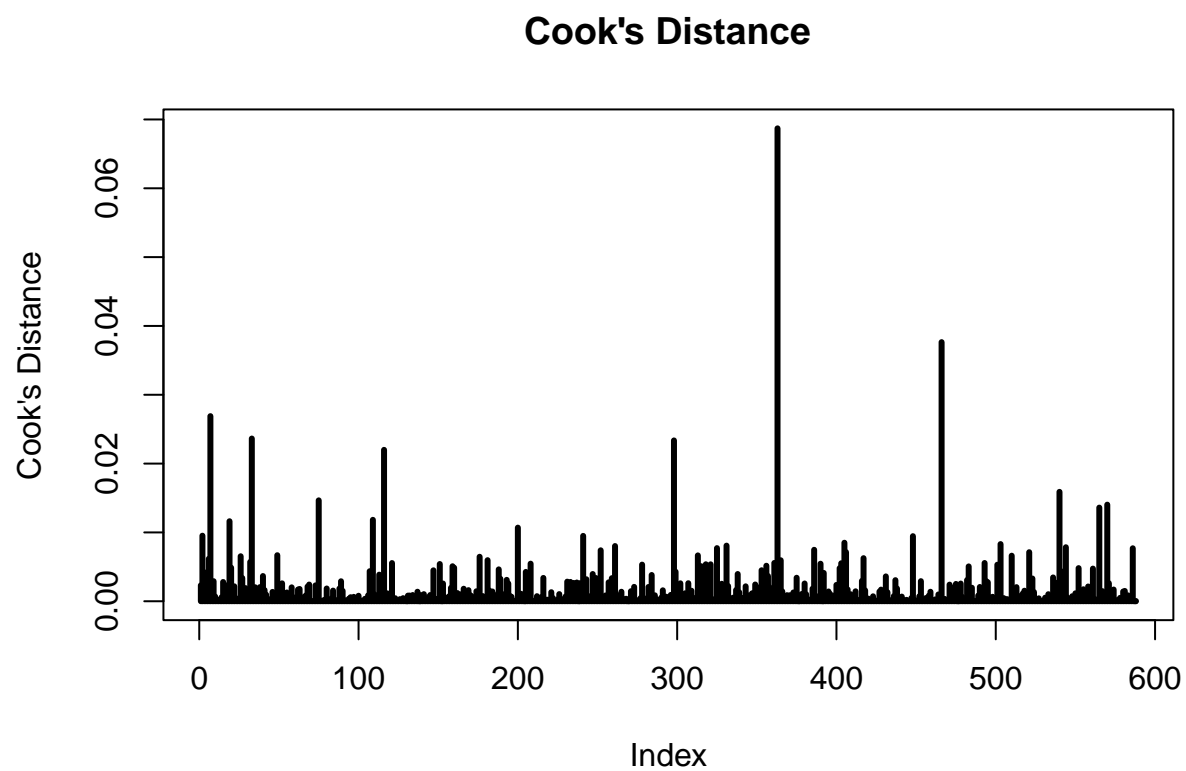
```
# threshold for multicollinearity
# Calculating the threshold
```

```
max(10, 1/(1-summary(model1_2)$r.square))
```

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

```
# Checking outliers: estimate of the influence of data point; summary of how much a regression model changes
# no outliers detected
```

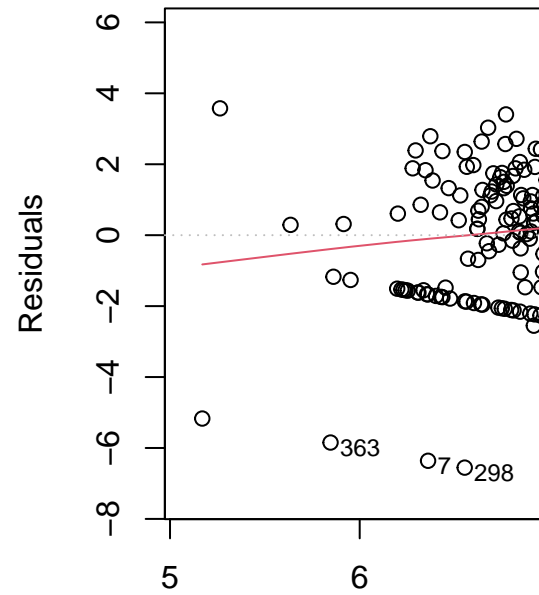
```
cook = cooks.distance(model1_2)
plot(cook,
      type="h",
      lwd=3,
      ylab = "Cook's Distance",
      main="Cook's Distance")
abline(h = 1)
```

```
influential = cooks.distance(model1_2)[which(cook > 1)]  
influential
```

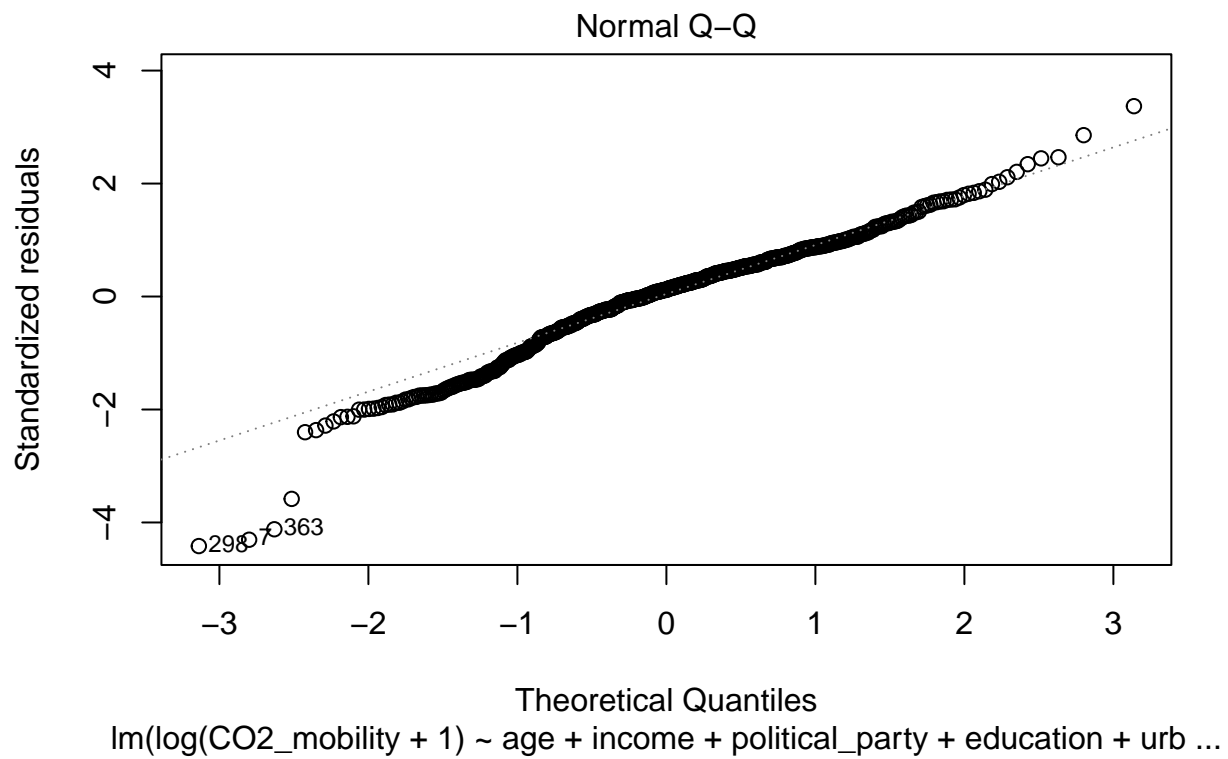
```
## named numeric(0)
```

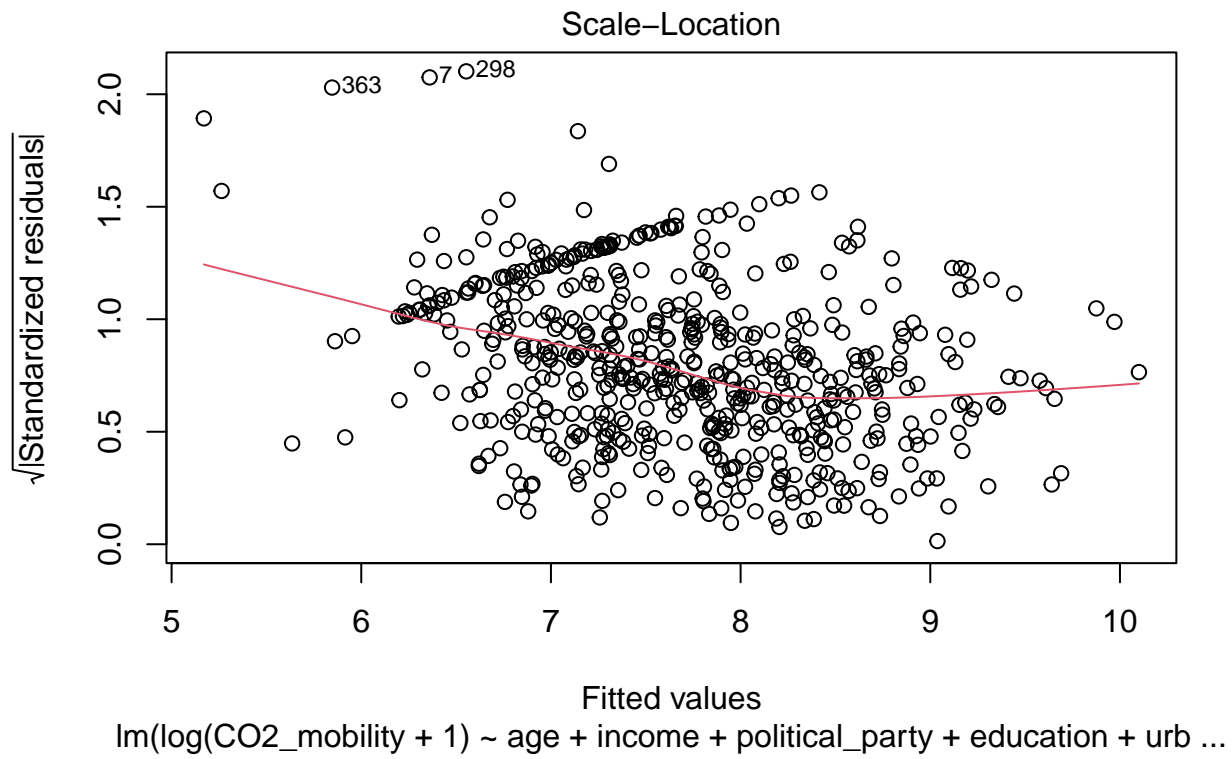
```
plot(model1_2)
```

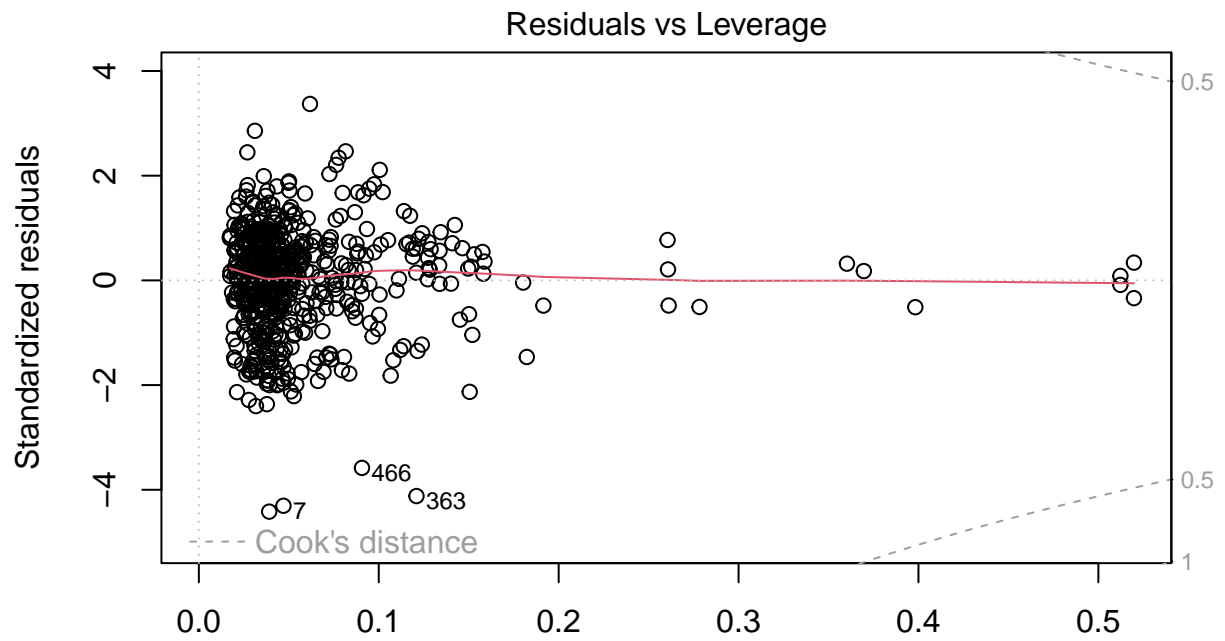


4. Assumptions check in the residuals of the transformed regression

$\ln(\log(\text{CO2_mobility} + 1)) \sim \text{age} + \text{income} + \text{political_party} + \text{education} + \text{urb} \dots$





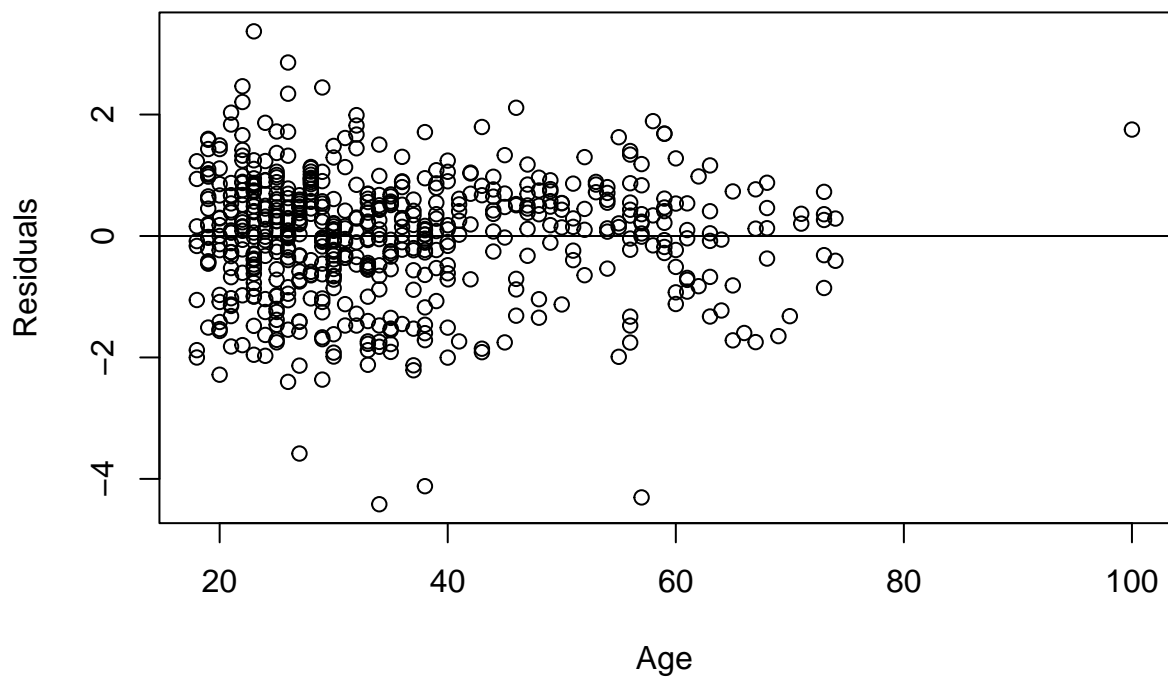


Im(log(CO2_mobility + 1) ~ age + income + political_party + education + urb ...

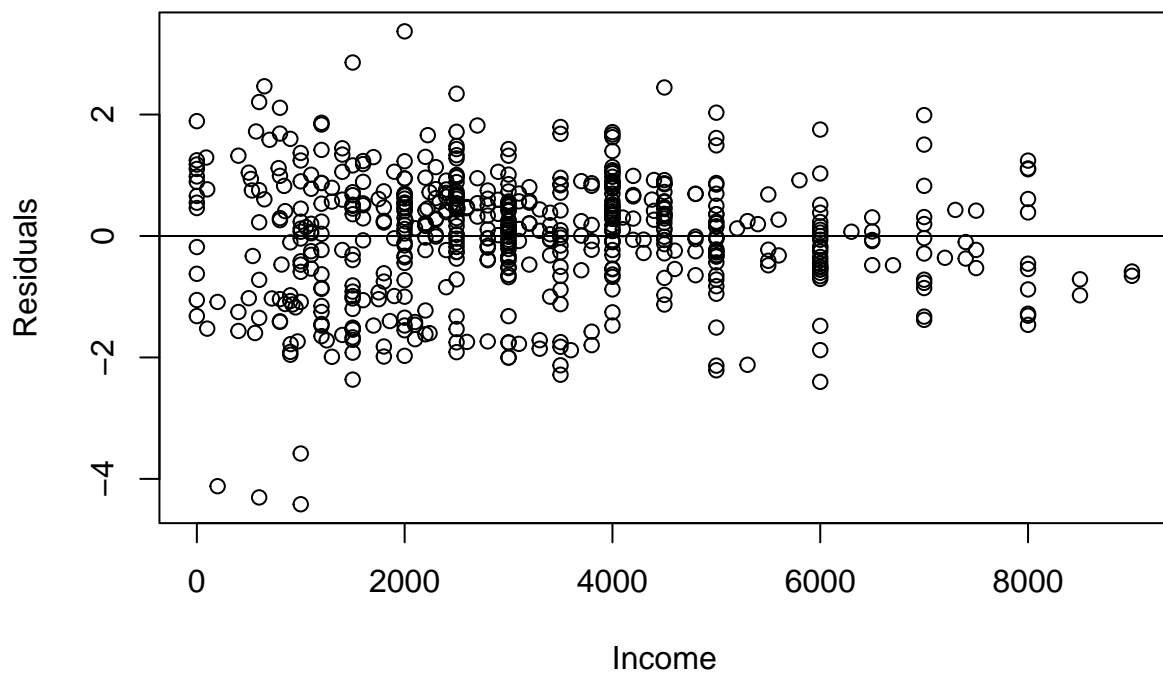
```
res1 = stdres(model1_2) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption: still violated

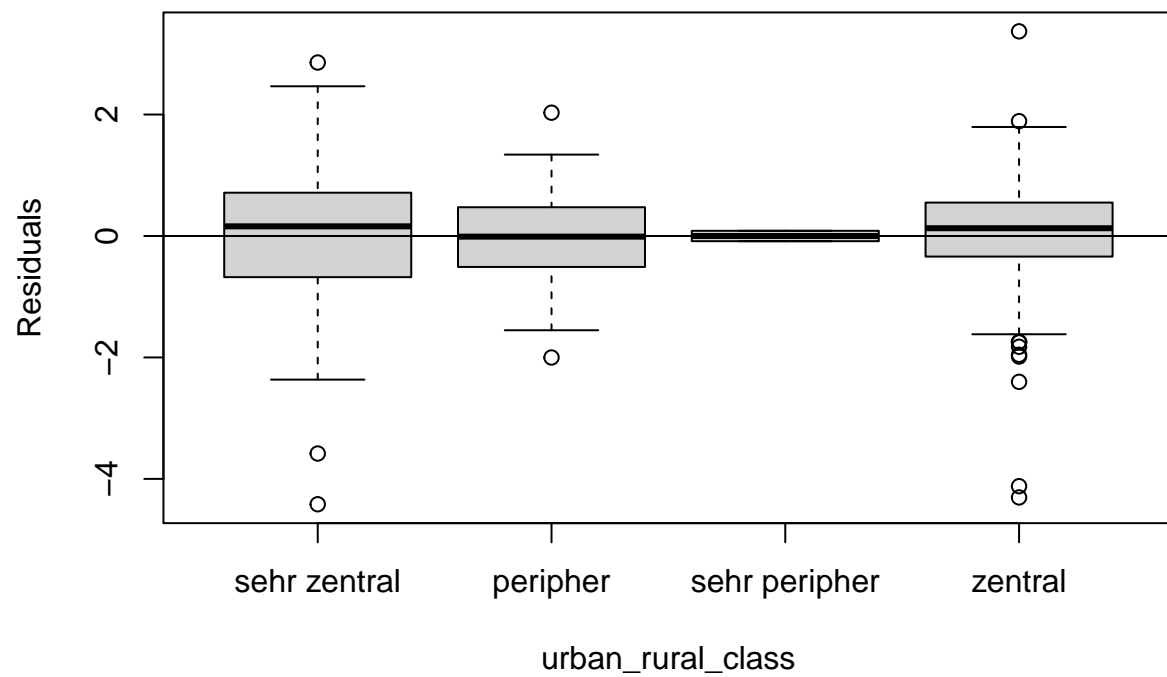
plot(df1$age, res1, xlab = "Age", ylab = "Residuals")
abline(h = 0)
```



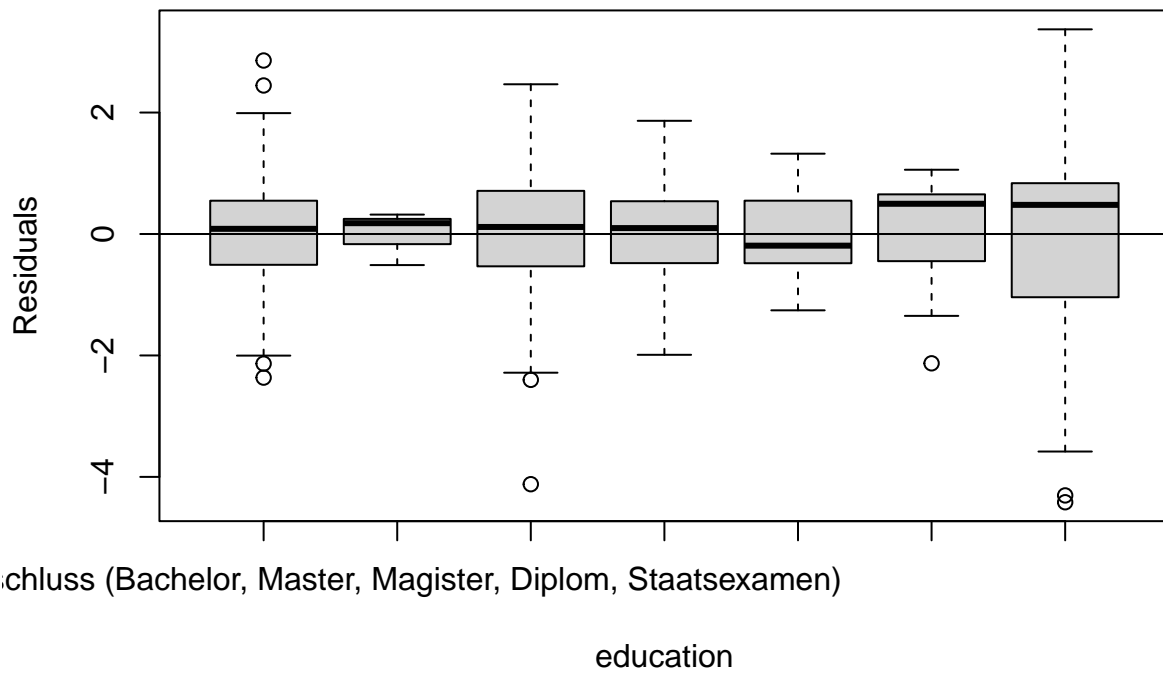
```
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")  
abline(h = 0)
```



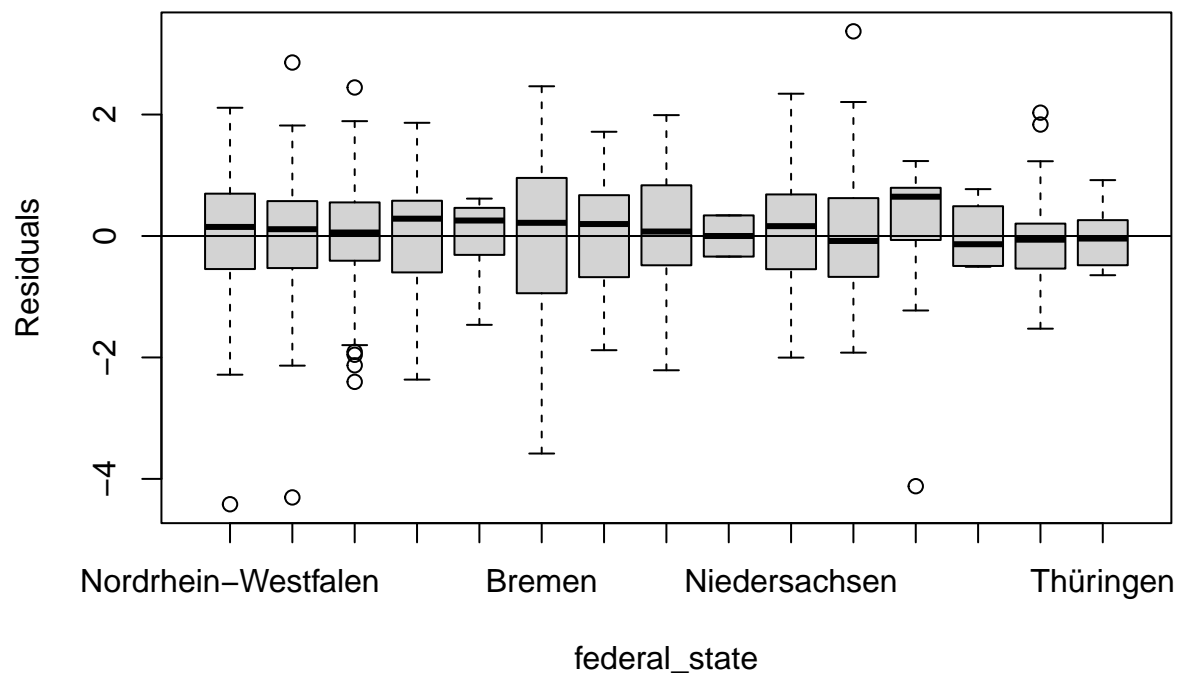
```
plot(df1$urban_rural_class, res1, xlab = "urban_rural_class", ylab = "Residuals")  
abline(h = 0)
```



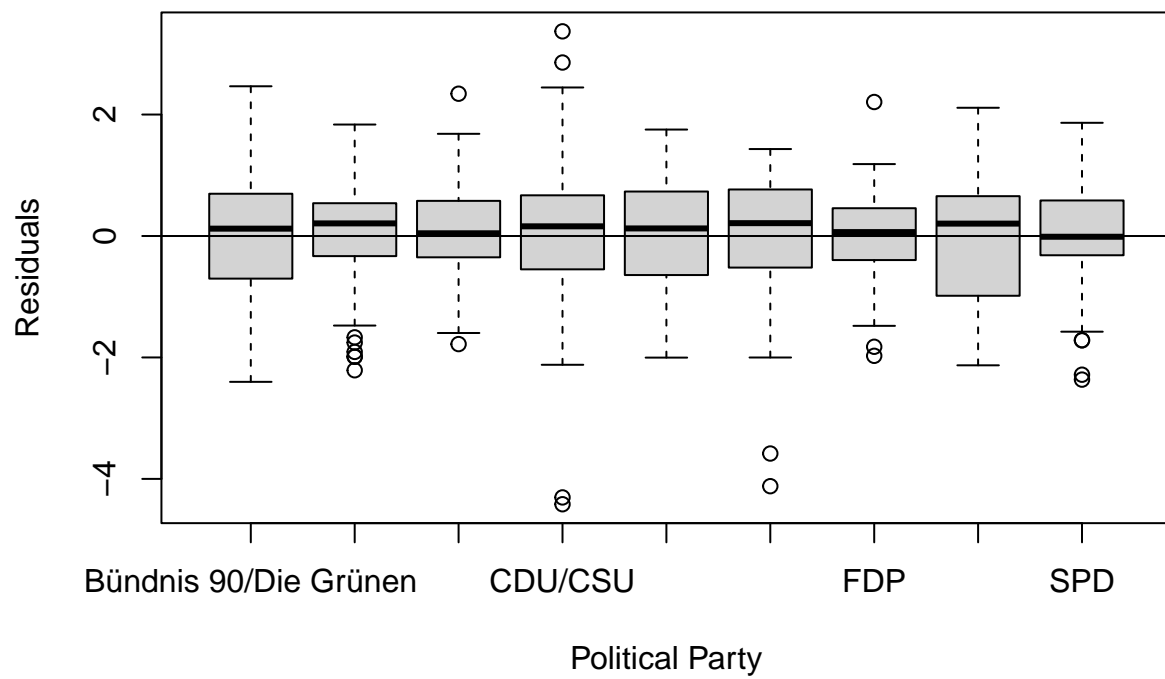
```
plot(df1$education, res1, xlab = "education", ylab = "Residuals")  
abline(h = 0)
```



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

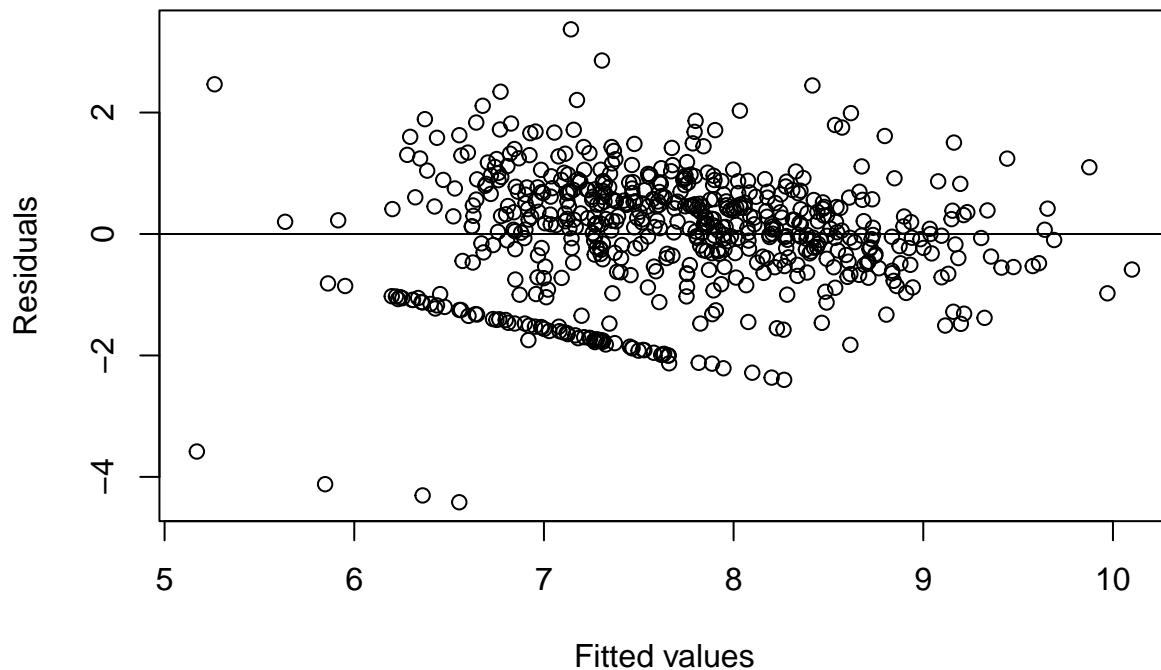



```
plot(df1$political_party, res1, xlab = "Political Party", ylab = "Residuals")
abline(h = 0)
```



Constant variance and independent error term assumption

```
plot(fitted(model1_2), res1, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```



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

```
durbinWatsonTest(model1_2)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 -0.02375885 2.046619 0.602
## Alternative hypothesis: rho != 0
```

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

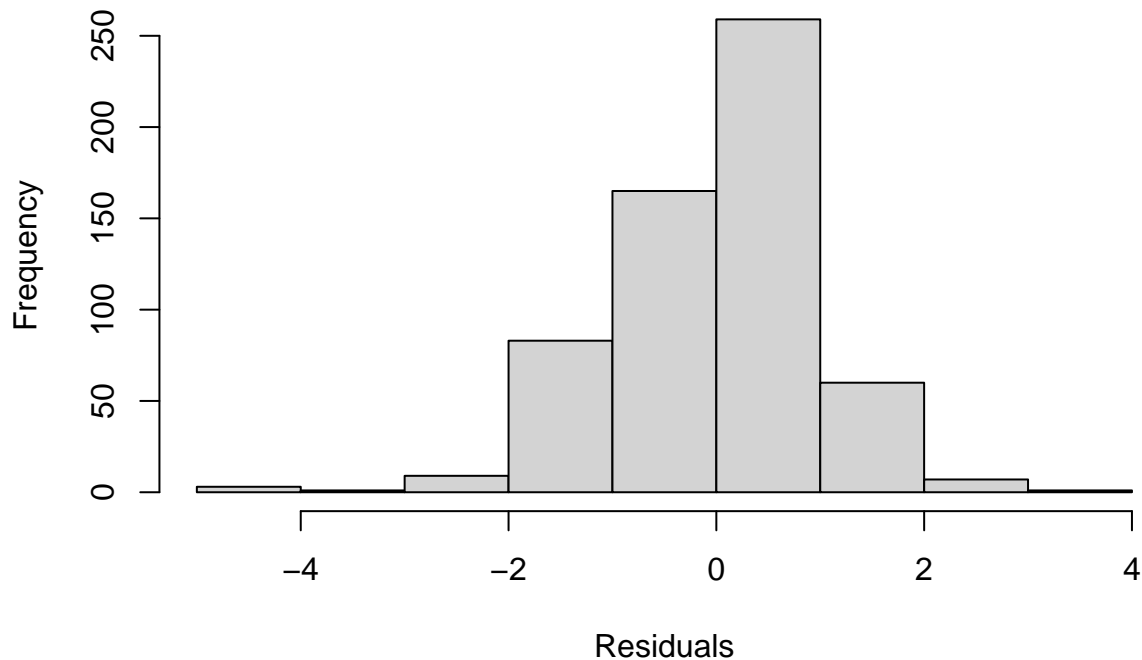
```
library(lmtest)
bptest(model1_2)
```

```
##
## studentized Breusch-Pagan test
##
## data: model1_2
## BP = 85.255, df = 33, p-value = 1.66e-06
```

```
# Normality assumption
```

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

Histogram of Residuals



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

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

```
shapiro.test(res1_num)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: res1_num  
## W = 0.97153, p-value = 2.857e-09
```

5. Variable Selection, model outcome and assumption check

```
### Backward regression using AIC: starting with all of the variables  
# r-squared not good  
options(scipen = -2)  
  
step_model1 <- stepAIC(model1_2, trace=TRUE, direction= "backward")
```

FINAL MODEL

```
## Start: AIC=520.31
## log(CO2_mobility + 1) ~ age + income + political_party + education +
##   urban_rural_class + federal_state
##
##           Df Sum of Sq  RSS   AIC
## - federal_state    14    44.984 1314.0 512.79
## - age                1     1.668 1270.6 519.08
## <none>                                1269.0 520.31
## - urban_rural_class    3    16.553 1285.5 521.93
## - education            6    31.654 1300.6 522.80
## - political_party      8    50.993 1320.0 527.48
## - income              1   180.740 1449.7 596.61
##
## Step: AIC=512.79
## log(CO2_mobility + 1) ~ age + income + political_party + education +
##   urban_rural_class
##
##           Df Sum of Sq  RSS   AIC
## - age                1     3.269 1317.2 512.26
## <none>                                1314.0 512.79
## - urban_rural_class    3    14.564 1328.5 513.28
## - education            6    34.819 1348.8 516.17
## - political_party      8    51.257 1365.2 519.30
## - income              1   195.447 1509.4 592.33
##
## Step: AIC=512.26
## log(CO2_mobility + 1) ~ income + political_party + education +
##   urban_rural_class
##
##           Df Sum of Sq  RSS   AIC
## <none>                                1317.2 512.26
## - urban_rural_class    3    14.458 1331.7 512.68
## - education            6    34.629 1351.9 515.51
## - political_party      8    49.569 1366.8 517.98
## - income              1   196.458 1513.7 592.00
```

```
summary(step_model1)
```

```
##
## Call:
## lm(formula = log(CO2_mobility + 1) ~ income + political_party +
##   education + urban_rural_class, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5082 -0.7462  0.1629  0.9648  5.1984
##
## Coefficients:
##              Estimate
## (Intercept)  6.439e+00
## income       3.148e-04
## political_partyAfD  7.066e-01
## political_partyBündnis Sarah Wagenknecht  4.166e-01
## political_partyCDU/CSU  4.964e-01
```

## political_partyDie Linke	4.349e-01
## political_partyEiner anderen Partei	2.592e-01
## political_partyFDP	8.455e-01
## political_partyKeine Angabe	6.620e-01
## political_partySPD	7.689e-01
## education(Noch) kein Abschluss	-2.629e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	-3.540e-01
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	-3.932e-01
## educationDoktorgrad oder Habilitation	-1.233e-01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	-5.684e-01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	-7.816e-01
## urban_rural_classperipher	4.319e-01
## urban_rural_classsehr peripher	1.329e+00
## urban_rural_classzentral	1.006e-01
##	Std. Error
## (Intercept)	1.905e-01
## income	3.417e-05
## political_partyAfD	2.445e-01
## political_partyBündnis Sarah Wagenknecht	3.461e-01
## political_partyCDU/CSU	2.204e-01
## political_partyDie Linke	2.648e-01
## political_partyEiner anderen Partei	1.977e-01
## political_partyFDP	2.561e-01
## political_partyKeine Angabe	4.315e-01
## political_partySPD	2.233e-01
## education(Noch) kein Abschluss	8.886e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	1.666e-01
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	1.747e-01
## educationDoktorgrad oder Habilitation	4.356e-01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	4.921e-01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	2.262e-01
## urban_rural_classperipher	1.951e-01
## urban_rural_classsehr peripher	1.087e+00
## urban_rural_classzentral	1.499e-01
##	t value
## (Intercept)	33.799
## income	9.212
## political_partyAfD	2.890
## political_partyBündnis Sarah Wagenknecht	1.204
## political_partyCDU/CSU	2.252
## political_partyDie Linke	1.642
## political_partyEiner anderen Partei	1.311
## political_partyFDP	3.302
## political_partyKeine Angabe	1.534
## political_partySPD	3.444
## education(Noch) kein Abschluss	-0.296
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	-2.125
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	-2.251
## educationDoktorgrad oder Habilitation	-0.283
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	-1.155
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	-3.455
## urban_rural_classperipher	2.213
## urban_rural_classsehr peripher	1.223
## urban_rural_classzentral	0.671

```

##                                                    Pr(>|t|)
## (Intercept)                                     < 2e-16
## income                                           < 2e-16
## political_partyAfD                             4.00e-03
## political_partyBündnis Sarah Wagenknecht       2.29e-01
## political_partyCDU/CSU                         2.47e-02
## political_partyDie Linke                      1.01e-01
## political_partyEiner anderen Partei            1.90e-01
## political_partyFDP                             1.02e-03
## political_partyKeine Angabe                   1.26e-01
## political_partySPD                             6.16e-04
## education(Noch) kein Abschluss                 7.67e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS) 3.40e-02
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule 2.48e-02
## educationDoktorgrad oder Habilitation          7.77e-01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss 2.49e-01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss 5.91e-04
## urban_rural_classperipher                      2.73e-02
## urban_rural_classsehr peripher                 2.22e-01
## urban_rural_classzentral                      5.03e-01
##
## (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                             ***
## 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
## ---
## Signif. codes:  0 '***' 1e-03 '**' 1e-02 '*' 5e-02 '.' 0.1 ' ' 1
##
## Residual standard error: 1.522 on 569 degrees of freedom
## Multiple R-squared:  0.1992, Adjusted R-squared:  0.1738
## F-statistic: 7.862 on 18 and 569 DF,  p-value: < 2.2e-16

```

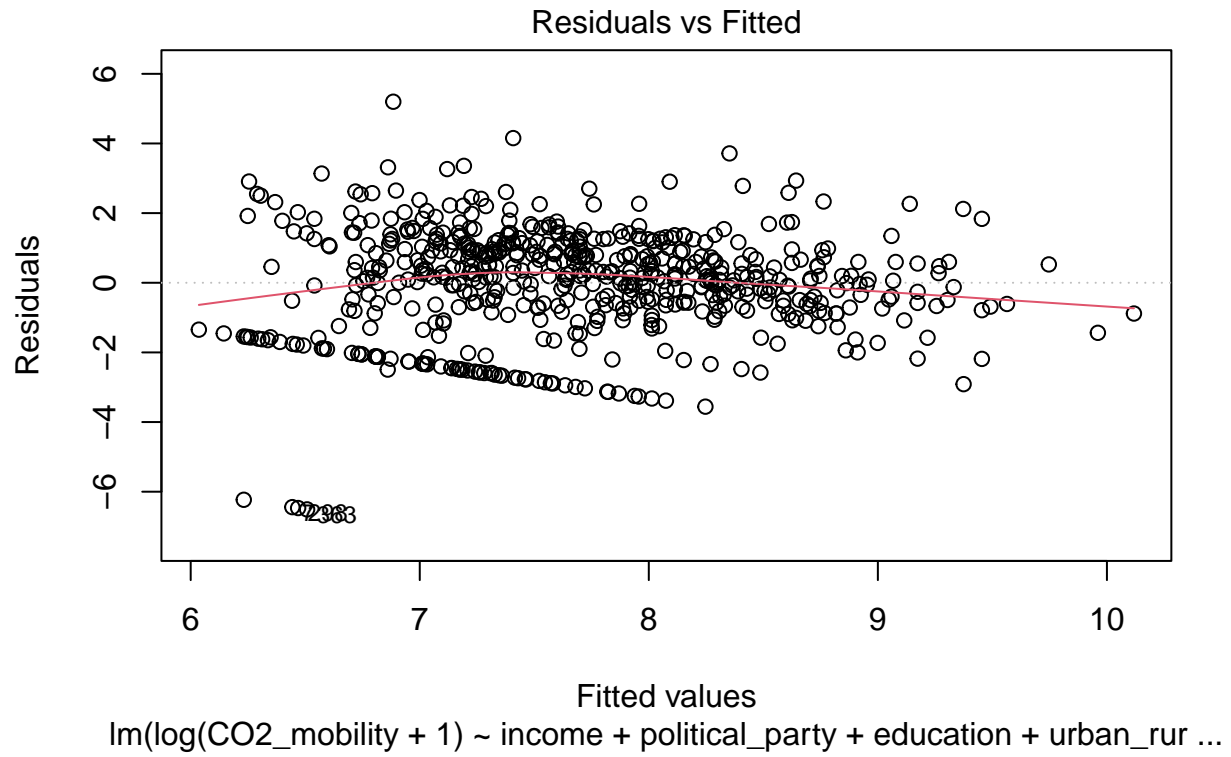
```
vif(step_model1)
```

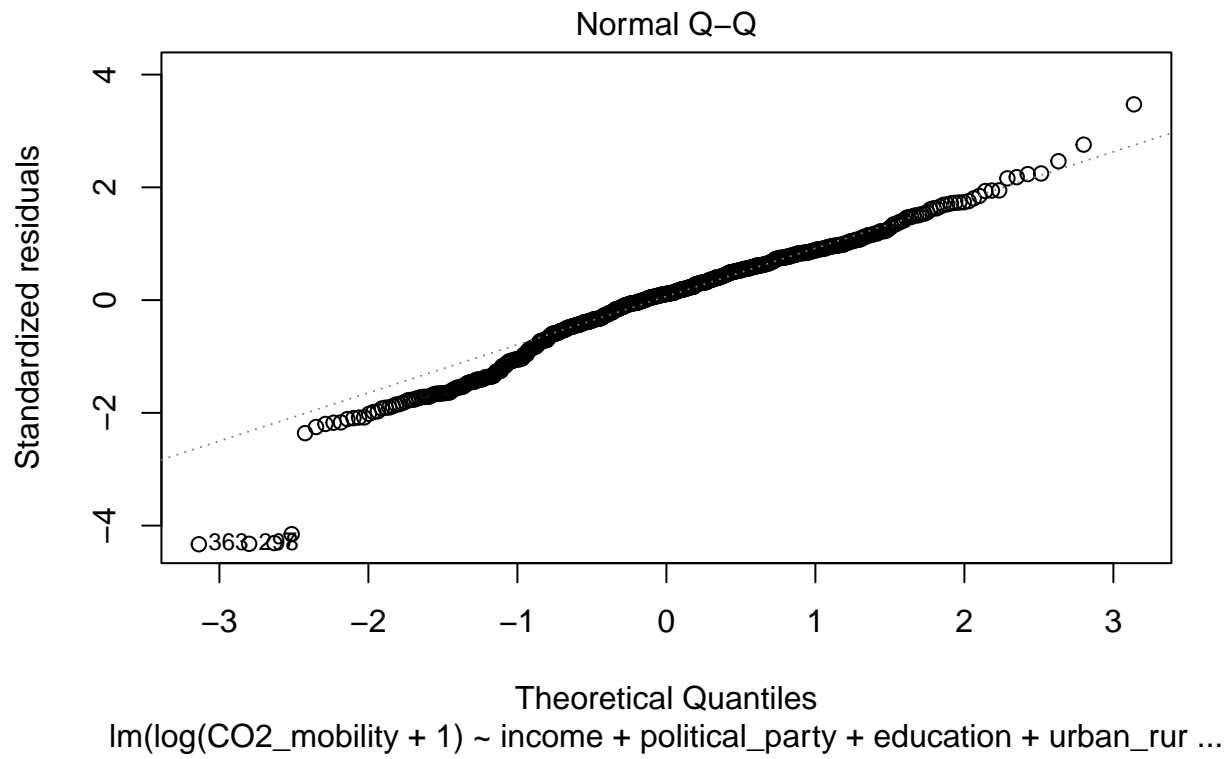
```

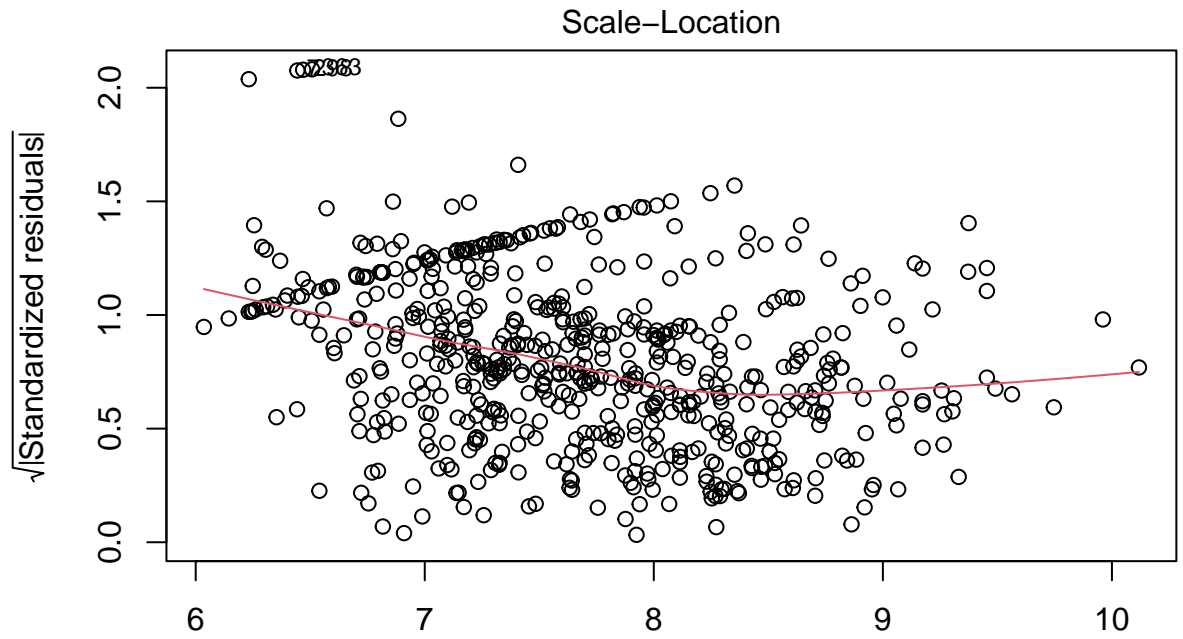
##              GVIF Df GVIF^(1/(2*Df))
## income          1.060993  1      1.030045
## political_party  1.302468  8      1.016653
## education       1.271724  6      1.020233
## urban_rural_class 1.129197  3      1.020458

```

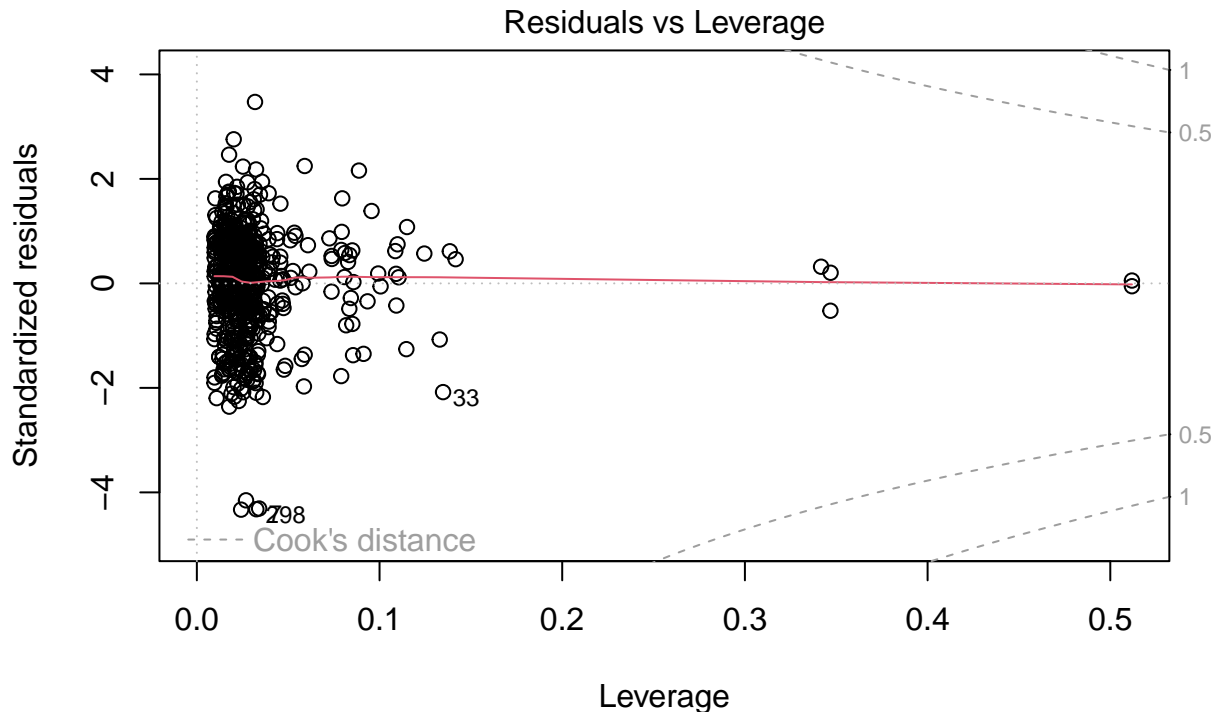
```
plot(step_model1)
```







Fitted values
 $\text{lm}(\log(\text{CO2_mobility} + 1) \sim \text{income} + \text{political_party} + \text{education} + \text{urban_rur} \dots$



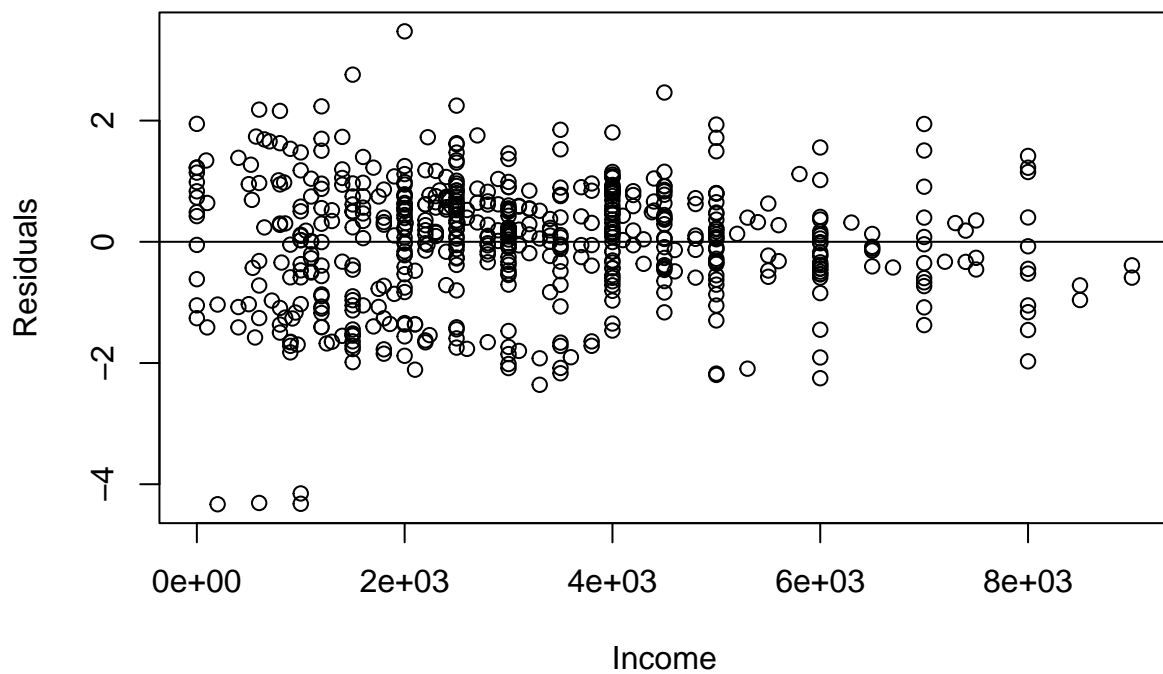
$\text{lm}(\log(\text{CO2_mobility} + 1) \sim \text{income} + \text{political_party} + \text{education} + \text{urban_rur} \dots)$

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

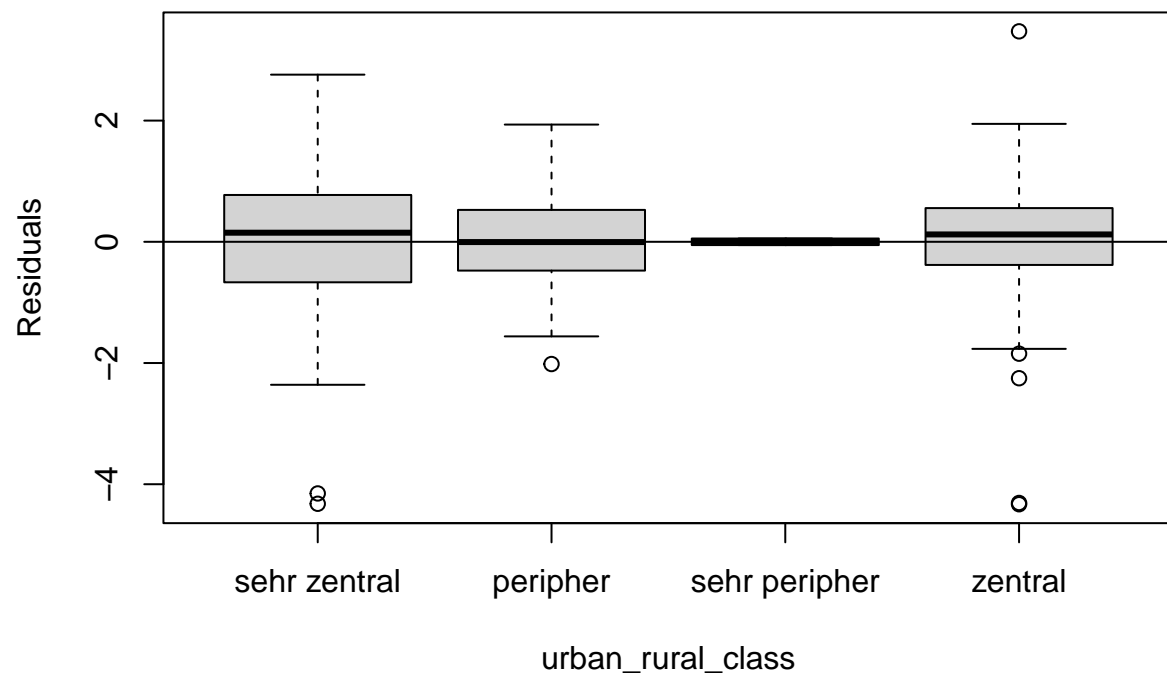
# Linearity assumption/Mean zero assumption

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

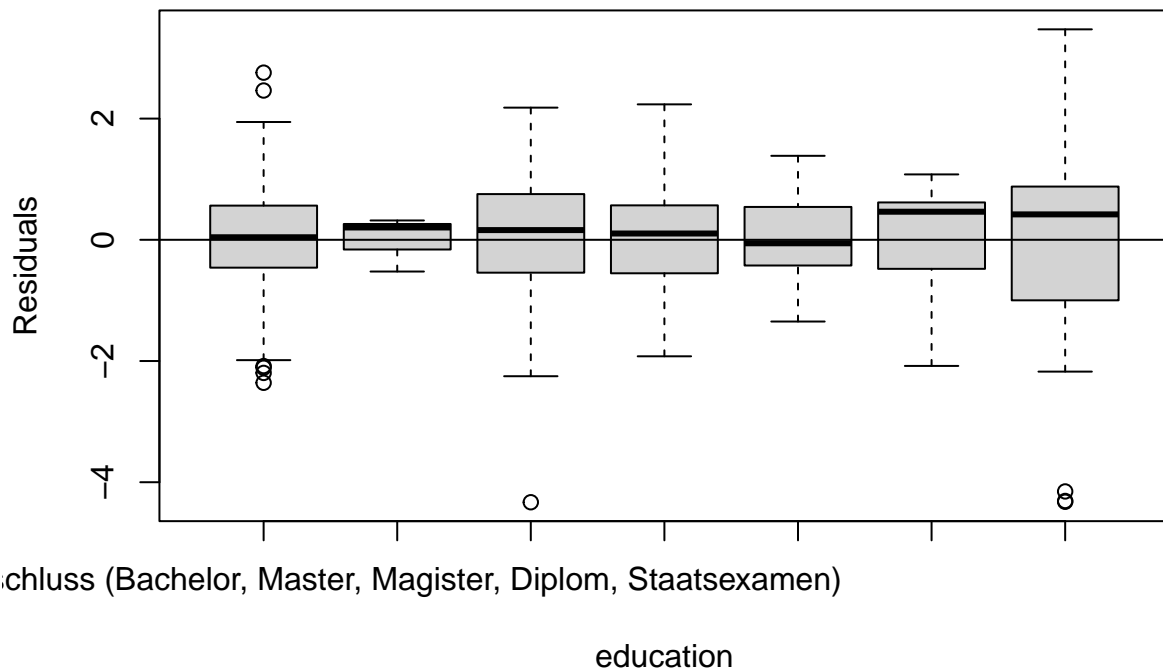
plot(df1$income, res1, xlab = "Income", ylab = "Residuals")
abline(h = 0)
```



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

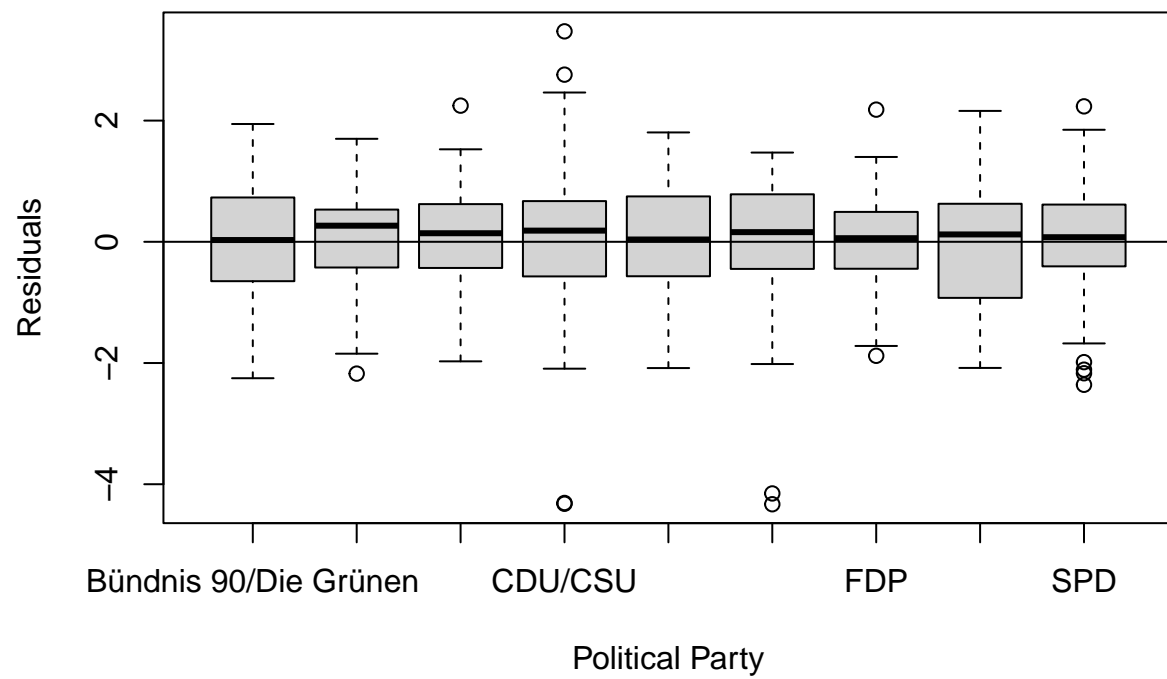


```
plot(df1$education, res1, xlab = "education", ylab = "Residuals")  
abline(h = 0)
```



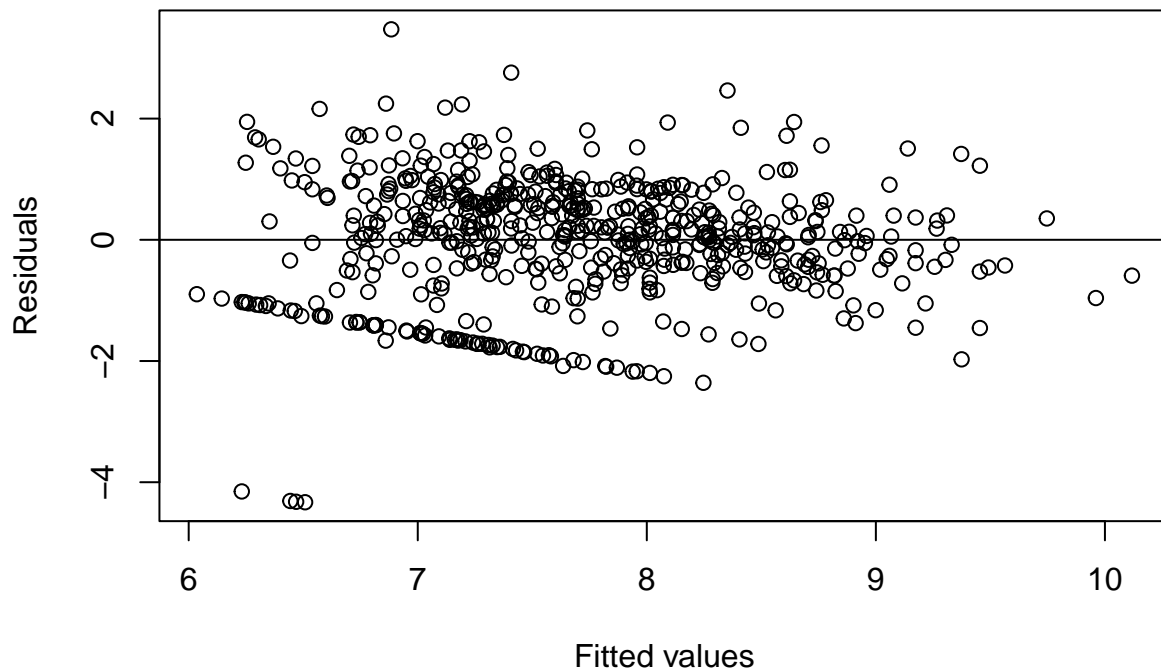
```
#plot(df1_new$federal_state, res1, xlab = "federal_state", ylab = "Residuals")
#abline(h = 0)

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



Constant variance and independent error term assumption

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



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

```
durbinWatsonTest(step_model1)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 -0.02565914 2.051207 0.536
## Alternative hypothesis: rho != 0
```

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

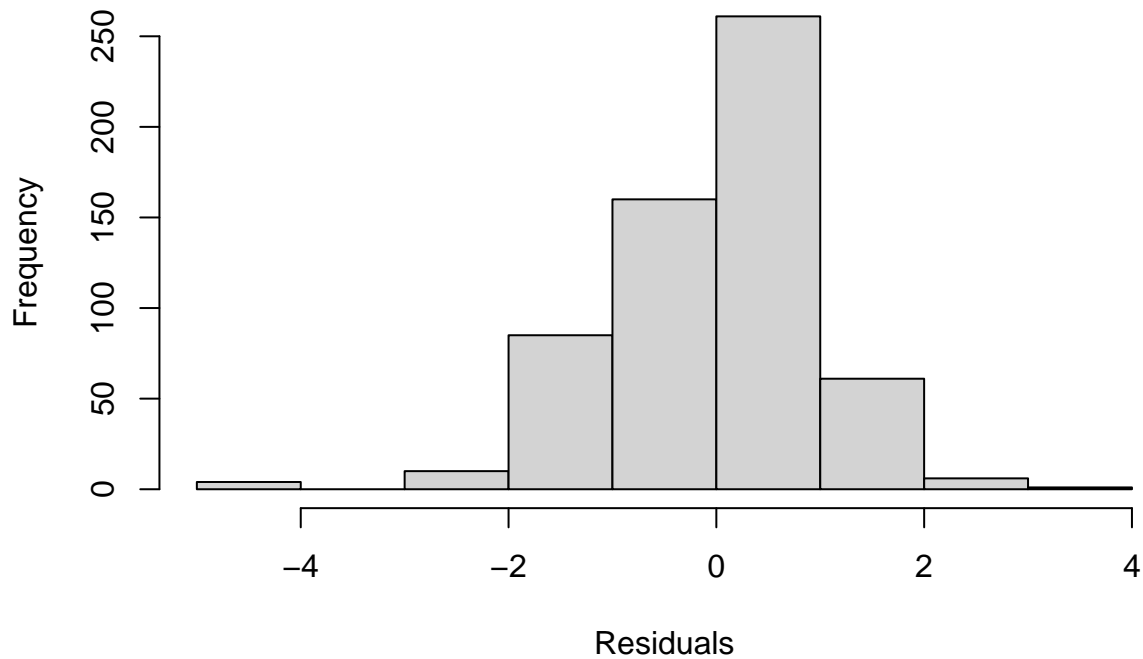
```
bptest(step_model1)
```

```
##
## studentized Breusch-Pagan test
##
## data: step_model1
## BP = 69.042, df = 18, p-value = 6.56e-08
```

```
# Normality assumption
```

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


Histogram of Residuals



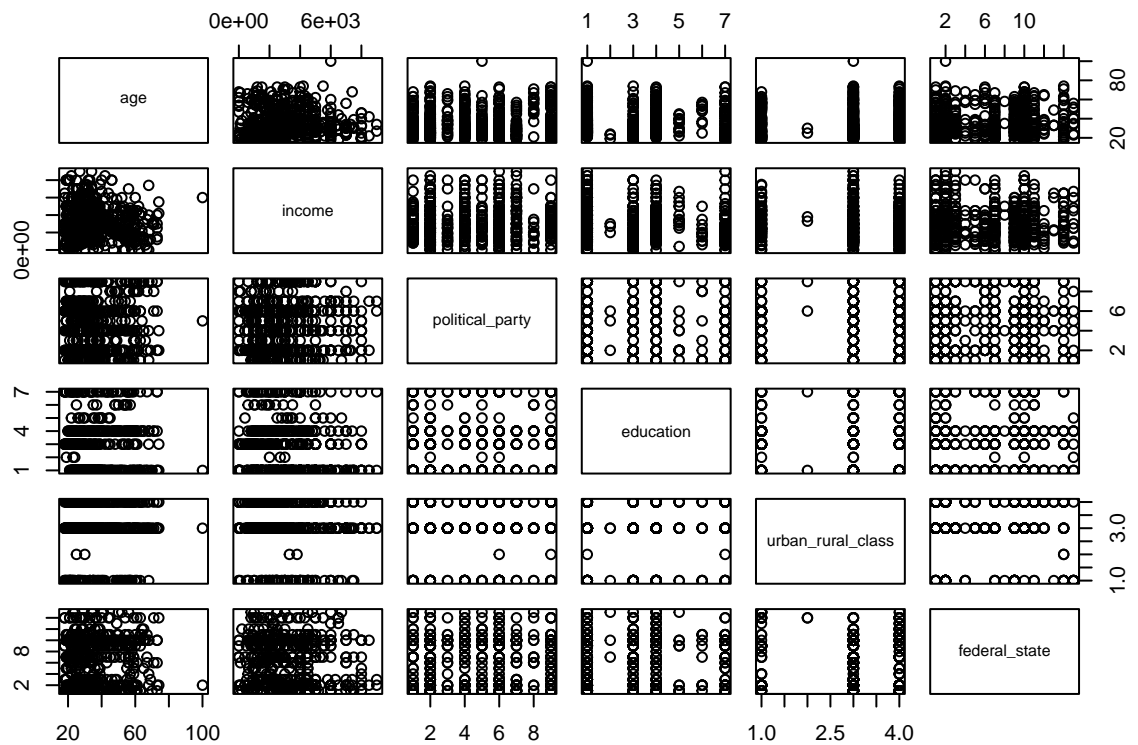
```
## normality test using shapiro-test: reject the H0  
#H0: the sample comes from a normal distribution
```

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

```
##  
## Shapiro-Wilk normality test  
##  
## data: res1_num  
## W = 0.96882, p-value = 7.478e-10
```

III. Multivariate Regression: belief diff mobility

```
# Checking the possible correlation in the data  
plot(df2[1:6])
```



1. Modeling

```
## defining a reference level: check chunk 11
```

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

```
# regression model with all variables
```

```
model2 <- lm(belief_diff_mobility ~ age + income + political_party + education + urban_rural_class + fe
summary(model2)
```

```
##
## Call:
## lm(formula = belief_diff_mobility ~ age + income + political_party +
##     education + urban_rural_class + federal_state, data = df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -76.826 -17.387   2.023  14.714  97.893
##
## Coefficients:
##
##              Estimate
## (Intercept) -1.285e+01
## age         2.806e-01
```

## income	-2.079e-03
## political_partyAfD	-3.835e+00
## political_partyBündnis Sarah Wagenknecht	5.104e-01
## political_partyCDU/CSU	-6.525e+00
## political_partyDie Linke	-3.387e+00
## political_partyEiner anderen Partei	-3.069e+00
## political_partyFDP	-7.047e+00
## political_partyKeine Angabe	-4.019e+00
## political_partySPD	-1.526e+01
## education(Noch) kein Abschluss	2.336e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	4.696e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	6.097e+00
## educationDoktorgrad oder Habilitation	1.020e+01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	1.492e+01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	1.099e+01
## urban_rural_classperipher	-3.225e+00
## urban_rural_classsehr peripher	6.923e+00
## urban_rural_classzentral	-9.624e-02
## federal_stateBaden-Württemberg	3.229e+00
## federal_stateBayern	1.020e+00
## federal_stateBerlin	-7.788e-01
## federal_stateBrandenburg	2.963e+01
## federal_stateBremen	1.729e+01
## federal_stateHamburg	4.102e+00
## federal_stateHessen	-1.215e+00
## federal_stateMecklenburg-Vorpommern	1.017e+01
## federal_stateNiedersachsen	1.722e-01
## federal_stateRheinland-Pfalz	-3.663e+00
## federal_stateSaarland	9.757e+00
## federal_stateSachsen-Anhalt	1.341e+01
## federal_stateSchleswig-Holstein	1.763e+00
## federal_stateThüringen	-6.133e+00
##	Std. Error
## (Intercept)	5.428e+00
## age	9.609e-02
## income	6.421e-04
## political_partyAfD	4.598e+00
## political_partyBündnis Sarah Wagenknecht	6.444e+00
## political_partyCDU/CSU	4.129e+00
## political_partyDie Linke	4.977e+00
## political_partyEiner anderen Partei	3.694e+00
## political_partyFDP	4.782e+00
## political_partyKeine Angabe	8.284e+00
## political_partySPD	4.205e+00
## education(Noch) kein Abschluss	1.682e+01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	3.270e+00
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	3.336e+00
## educationDoktorgrad oder Habilitation	8.155e+00
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	9.309e+00
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	4.242e+00
## urban_rural_classperipher	4.275e+00
## urban_rural_classsehr peripher	2.116e+01
## urban_rural_classzentral	3.134e+00
## federal_stateBaden-Württemberg	4.072e+00

## federal_stateBayern	4.180e+00
## federal_stateBerlin	5.078e+00
## federal_stateBrandenburg	1.067e+01
## federal_stateBremen	7.780e+00
## federal_stateHamburg	6.332e+00
## federal_stateHessen	4.858e+00
## federal_stateMecklenburg-Vorpommern	2.039e+01
## federal_stateNiedersachsen	4.942e+00
## federal_stateRheinland-Pfalz	6.095e+00
## federal_stateSaarland	9.571e+00
## federal_stateSachsen-Anhalt	1.475e+01
## federal_stateSchleswig-Holstein	7.129e+00
## federal_stateThüringen	1.081e+01
##	t value
## (Intercept)	-2.368
## age	2.920
## income	-3.238
## political_partyAfD	-0.834
## political_partyBündnis Sarah Wagenknecht	0.079
## political_partyCDU/CSU	-1.580
## political_partyDie Linke	-0.681
## political_partyEiner anderen Partei	-0.831
## political_partyFDP	-1.474
## political_partyKeine Angabe	-0.485
## political_partySPD	-3.628
## education(Noch) kein Abschluss	1.389
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	1.436
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	1.828
## educationDoktorgrad oder Habilitation	1.251
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	1.603
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	2.591
## urban_rural_classperipher	-0.754
## urban_rural_classsehr peripher	0.327
## urban_rural_classzentral	-0.031
## federal_stateBaden-Württemberg	0.793
## federal_stateBayern	0.244
## federal_stateBerlin	-0.153
## federal_stateBrandenburg	2.778
## federal_stateBremen	2.223
## federal_stateHamburg	0.648
## federal_stateHessen	-0.250
## federal_stateMecklenburg-Vorpommern	0.499
## federal_stateNiedersachsen	0.035
## federal_stateRheinland-Pfalz	-0.601
## federal_stateSaarland	1.019
## federal_stateSachsen-Anhalt	0.909
## federal_stateSchleswig-Holstein	0.247
## federal_stateThüringen	-0.568
##	Pr(> t)
## (Intercept)	1.82e-02
## age	3.64e-03
## income	1.28e-03
## political_partyAfD	4.05e-01
## political_partyBündnis Sarah Wagenknecht	9.37e-01

## political_partyCDU/CSU	1.15e-01
## political_partyDie Linke	4.96e-01
## political_partyEiner anderen Partei	4.06e-01
## political_partyFDP	1.41e-01
## political_partyKeine Angabe	6.28e-01
## political_partySPD	3.12e-04
## education(Noch) kein Abschluss	1.65e-01
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	1.52e-01
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	6.81e-02
## educationDoktorgrad oder Habilitation	2.11e-01
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	1.10e-01
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	9.81e-03
## urban_rural_classperipher	4.51e-01
## urban_rural_classsehr peripher	7.44e-01
## urban_rural_classzentral	9.76e-01
## federal_stateBaden-Württemberg	4.28e-01
## federal_stateBayern	8.07e-01
## federal_stateBerlin	8.78e-01
## federal_stateBrandenburg	5.66e-03
## federal_stateBremen	2.66e-02
## federal_stateHamburg	5.17e-01
## federal_stateHessen	8.03e-01
## federal_stateMecklenburg-Vorpommern	6.18e-01
## federal_stateNiedersachsen	9.72e-01
## federal_stateRheinland-Pfalz	5.48e-01
## federal_stateSaarland	3.08e-01
## federal_stateSachsen-Anhalt	3.64e-01
## federal_stateSchleswig-Holstein	8.05e-01
## federal_stateThüringen	5.71e-01
##	
## (Intercept)	*
## age	**
## income	**
## political_partyAfD	
## political_partyBündnis Sarah Wagenknecht	
## political_partyCDU/CSU	
## political_partyDie Linke	
## political_partyEiner anderen Partei	
## political_partyFDP	
## political_partyKeine Angabe	
## political_partySPD	***
## education(Noch) kein Abschluss	
## educationAllgemeine oder fachgebundene Hochschulreife/Abitur (Gymnasium bzw. EOS)	
## educationBerufsausbildung, Lehre oder Ausbildung an einer Fachschule	.
## educationDoktorgrad oder Habilitation	
## educationHauptschulabschluss (Volksschulabschluss) oder gleichwertiger Abschluss	
## educationRealschulabschluss (Mittlere Reife) oder gleichwertiger Abschluss	**
## urban_rural_classperipher	
## urban_rural_classsehr peripher	
## urban_rural_classzentral	
## federal_stateBaden-Württemberg	
## federal_stateBayern	
## federal_stateBerlin	
## federal_stateBrandenburg	**

```
## federal_stateBremen
## federal_stateHamburg
## federal_stateHessen
## federal_stateMecklenburg-Vorpommern
## federal_stateNiedersachsen
## federal_stateRheinland-Pfalz
## federal_stateSaarland
## federal_stateSachsen-Anhalt
## federal_stateSchleswig-Holstein
## federal_stateThüringen
## ---
## Signif. codes:  0 '***' 1e-03 '**' 1e-02 '*' 5e-02 '.' 0.1 ' ' 1
##
## Residual standard error: 28.09 on 554 degrees of freedom
## Multiple R-squared:  0.1176, Adjusted R-squared:  0.06499
## F-statistic: 2.236 on 33 and 554 DF,  p-value: 1.278e-04
```

```
# Checking the VIFs for multicollinearity
```

```
vif(model2)
```

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

```
# threshold for multicollinearity
```

```
# Calculating the threshold
```

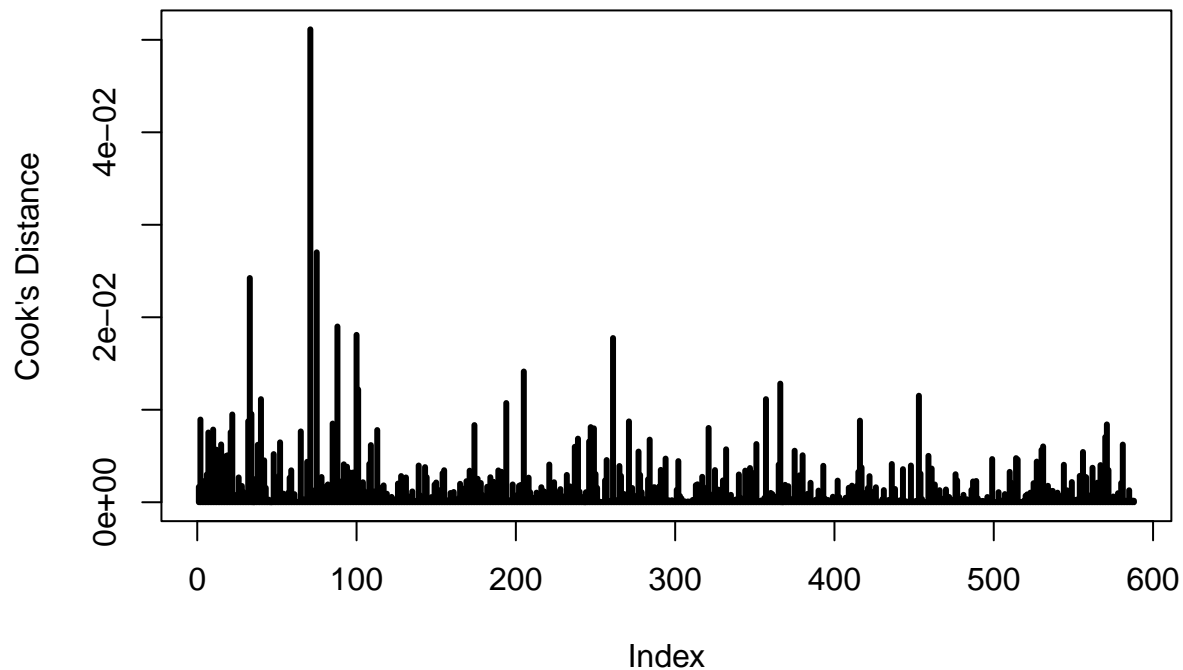
```
max(10, 1/(1-summary(model2)$r.square))
```

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

```
# Checking outliers
```

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

Cook's Distance



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

```
##           2           7           10           12           15           21
## 8.941968e-03 7.541535e-03 7.857655e-03 5.687019e-03 6.250193e-03 7.572721e-03
##          22          32          33          34          38          40
## 9.492056e-03 8.759963e-03 2.424498e-02 9.550146e-03 6.213189e-03 1.114879e-02
##          52          65          71          75          85          88
## 6.478947e-03 7.656839e-03 5.113081e-02 2.702019e-02 8.511976e-03 1.900174e-02
##         100         101         109         113         174         194
## 1.809747e-02 1.219187e-02 6.182040e-03 7.805895e-03 8.343031e-03 1.072930e-02
##         205         237         239         246         247         249
## 1.413706e-02 5.995577e-03 6.877588e-03 6.570804e-03 8.125633e-03 7.989901e-03
##         261         271         277         284         321         332
## 1.775548e-02 8.741018e-03 5.480297e-03 6.772870e-03 8.032307e-03 5.724478e-03
##         351         357         366         375         416         453
## 6.293711e-03 1.114879e-02 1.282733e-02 5.573314e-03 8.824362e-03 1.151811e-02
##         530         531         556         570         571         581
## 5.662137e-03 6.037726e-03 5.434877e-03 7.047557e-03 8.415183e-03 6.234873e-03
```

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

df2[influential_vector, ]
```

```
## # A tibble: 48 x 7
```

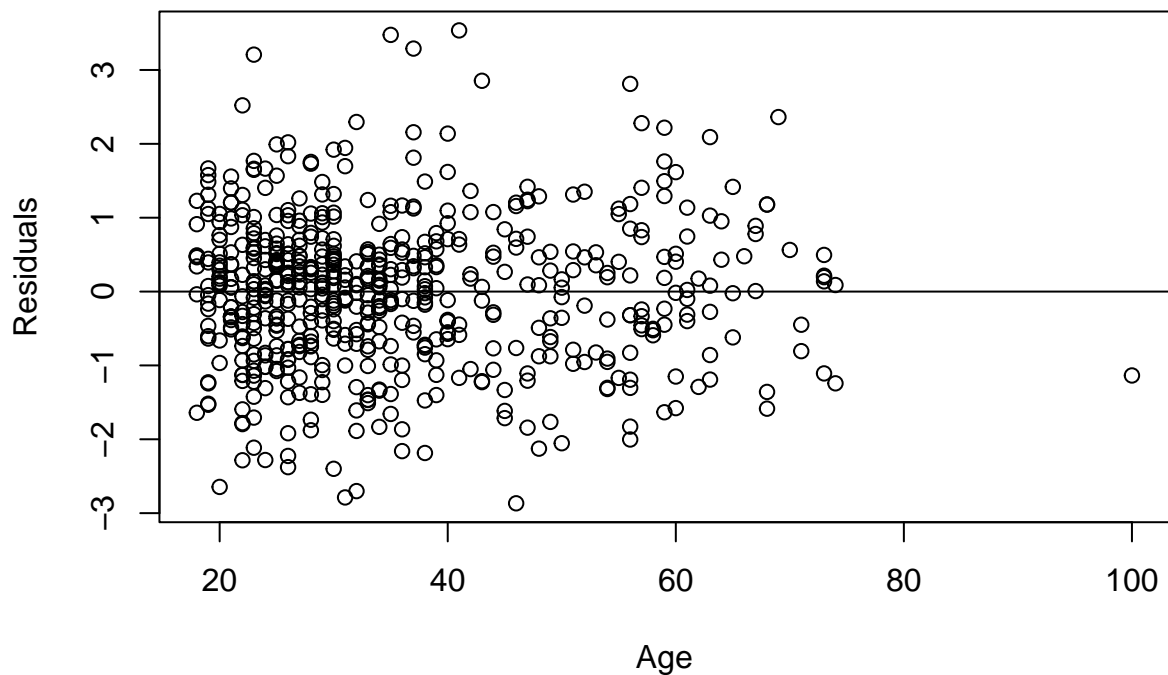
```
##      age income political_party      education      urban-1 feder-2 belie-3
##      <int> <dbl> <fct>          <fct>          <fct>   <fct>   <dbl>
##  1     59    800 Keine Angabe      Allgemeine oder fa~ sehr z~ Hessen    -42
##  2     57    600 CDU/CSU          Realschulabschluss~ zentral Baden~~    72
##  3     45   2600 Keine Angabe      Berufsausbildung, ~ sehr z~ Hessen   -48
##  4     48   5000 FDP              Berufsausbildung, ~ sehr z~ Baden~~   -66
##  5     56   1400 Einer anderen Partei Realschulabschluss~ periph~ Nieder~~   -50
##  6     54   2900 AfD              Hauptschulabschlus~ zentral Rheinl~   -31
##  7     52   4800 Die Linke          (Fach-) Hochschule~ periph~ Thürin~    14
##  8     48   1750 Keine Angabe      Hauptschulabschlus~ sehr z~ Hessen    40
##  9     37   3500 Keine Angabe      Hauptschulabschlus~ sehr z~ Bayern    58
## 10     59   4000 AfD              Berufsausbildung, ~ sehr z~ Bremen    62
## # ... with 38 more rows, and abbreviated variable names 1: urban_rural_class,
## #      2: federal_state, 3: belief_diff_mobility
```

2. Assumptions check in the residuals It does not work with this dataset because of the NaN

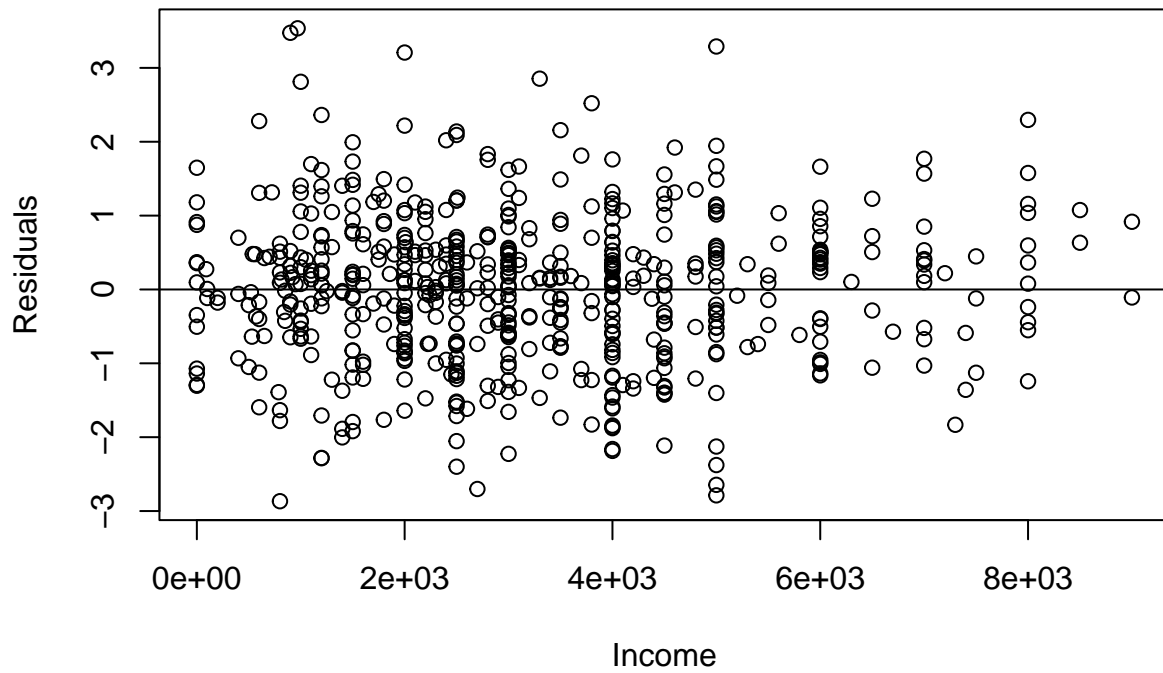
```
res2 = stdres(model2) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

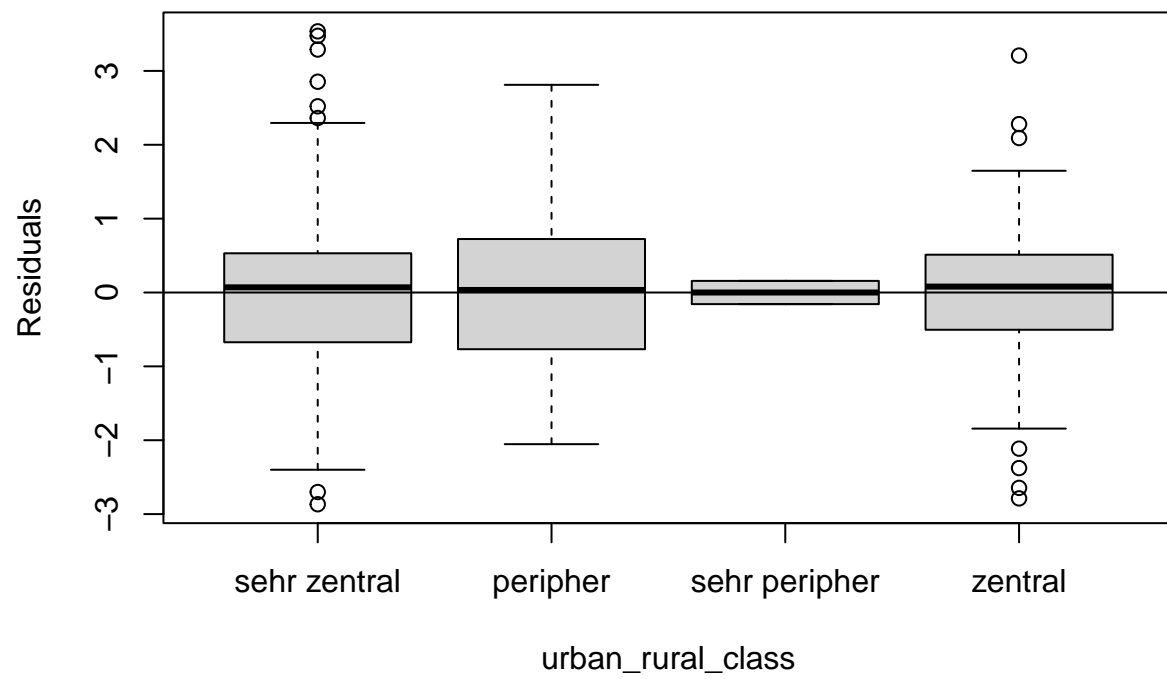
plot(df2$age, res2, xlab = "Age", ylab = "Residuals")
abline(h = 0)
```



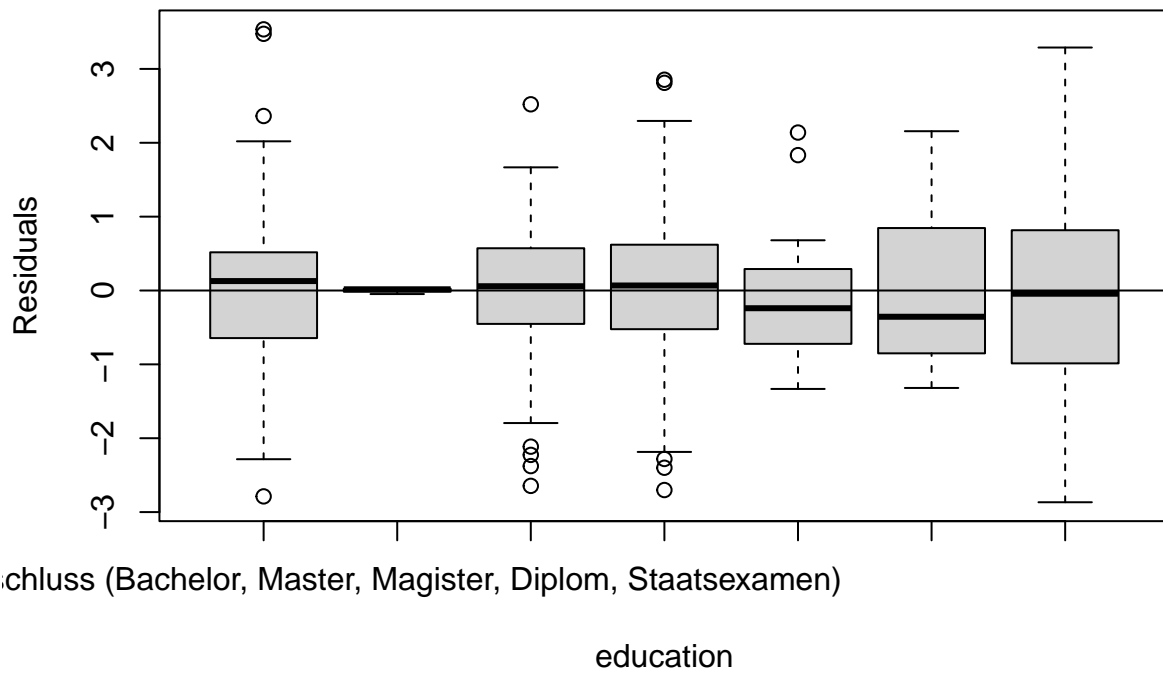

```
plot(df2$income, res2, xlab = "Income", ylab = "Residuals")  
abline(h = 0)
```



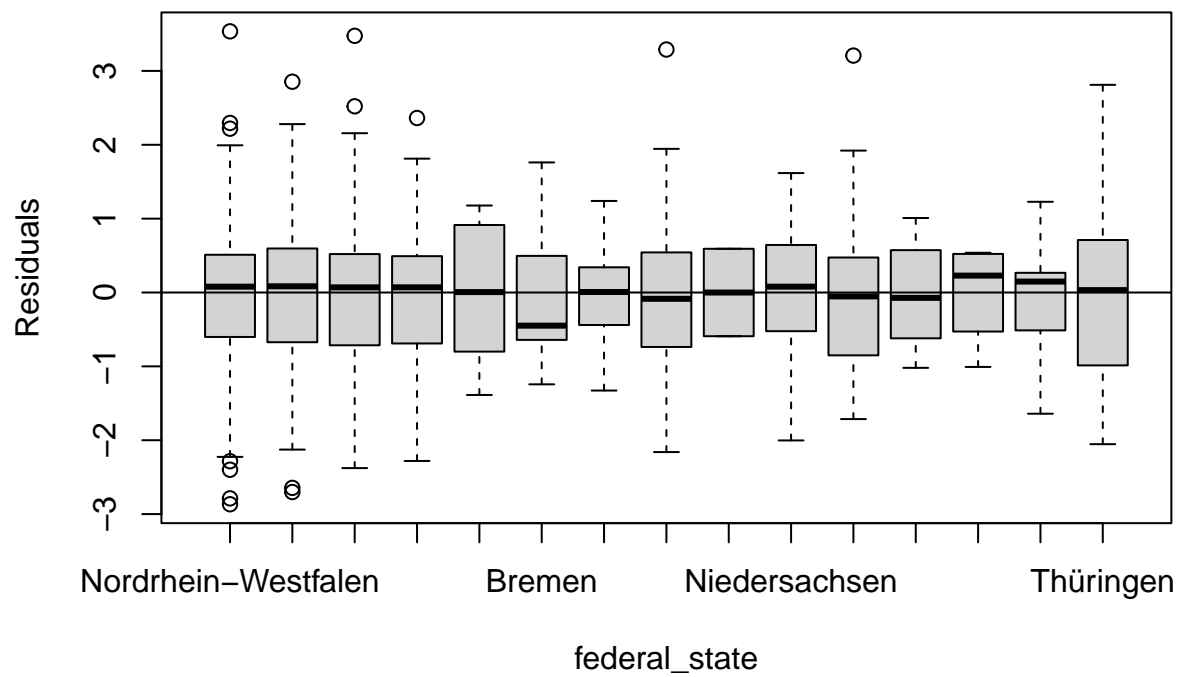
```
plot(df2$urban_rural_class, res2, xlab = "urban_rural_class", ylab = "Residuals")  
abline(h = 0)
```



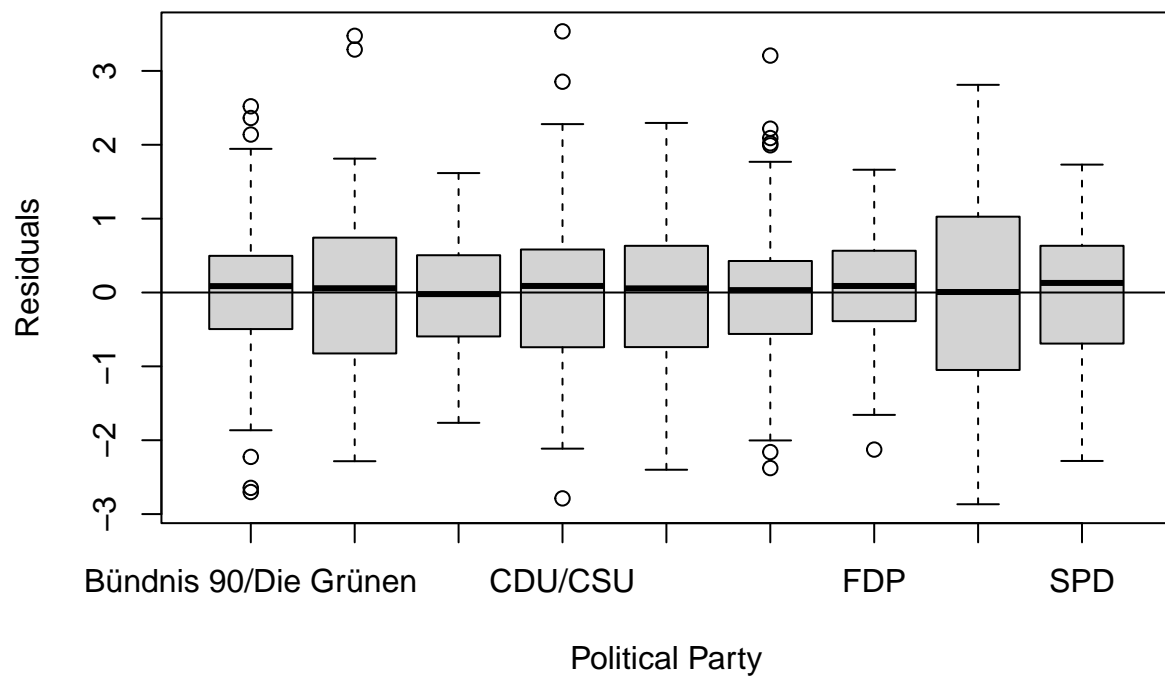
```
plot(df2$education, res2, xlab = "education", ylab = "Residuals")  
abline(h = 0)
```



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

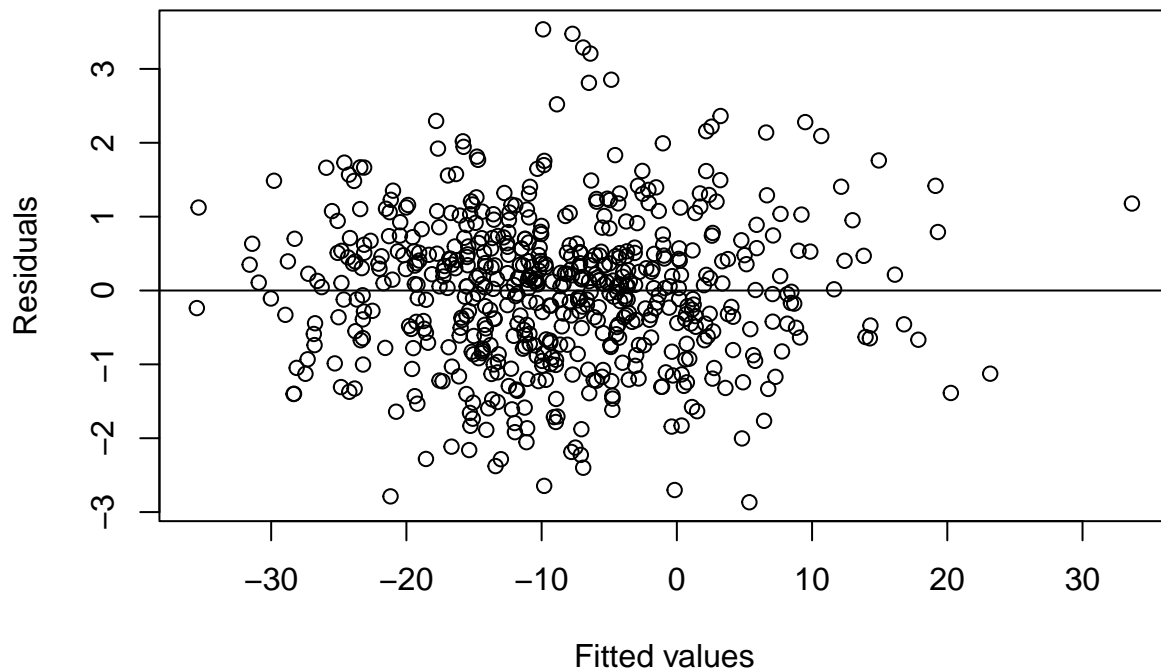


```
plot(df2$political_party, res2, xlab = "Political Party", ylab = "Residuals")
abline(h = 0)
```



Constant variance and independent error term assumption

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



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

```
durbinWatsonTest(model2)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 -9.736497e-03 2.018601 0.862
## Alternative hypothesis: rho != 0
```

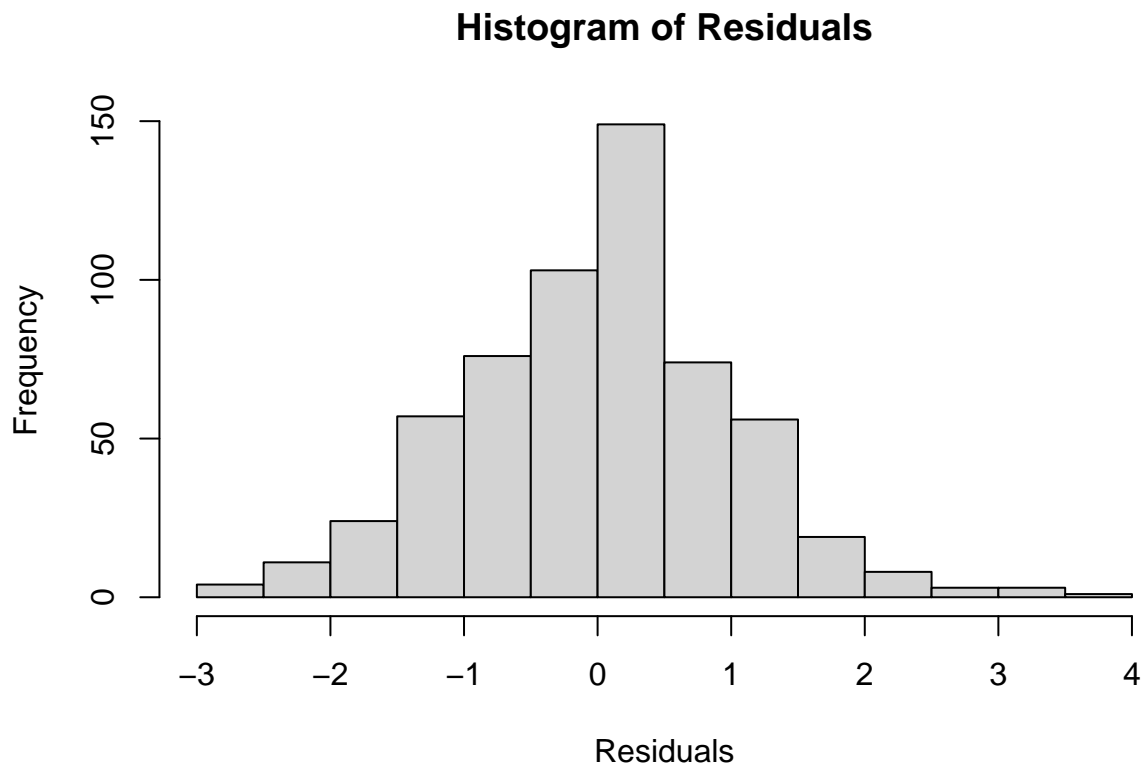
```
# Breusch-Pagan Test: Heteroscedasticity
# H0: Homoscedasticity is present
```

```
bptest(model2)
```

```
##
## studentized Breusch-Pagan test
##
## data: model2
## BP = 48.18, df = 33, p-value = 0.04272
```

```
# Normality assumption
```

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



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

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

```
shapiro.test(res2_num)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: res2_num  
## W = 0.99301, p-value = 7.676e-03
```

FINAL MODEL

```
# After removing the outliers, R-squared improved  
options(scipen=-0, digits=2)  
  
step_model2 <- stepAIC(model2, trace=TRUE, direction= "backward")
```

3. Variable selection

```

## Start: AIC=3955
## belief_diff_mobility ~ age + income + political_party + education +
##   urban_rural_class + federal_state
##
##           Df Sum of Sq    RSS   AIC
## - federal_state 14    13909 450998 3946
## - urban_rural_class 3      623 437712 3950
## - political_party  8    11958 449047 3955
## <none>                                437089 3955
## - education        6      9233 446322 3956
## - age              1      6727 443816 3962
## - income           1      8273 445362 3964
##
## Step: AIC=3946
## belief_diff_mobility ~ age + income + political_party + education +
##   urban_rural_class
##
##           Df Sum of Sq    RSS   AIC
## - urban_rural_class 3      822 451820 3941
## - education        6      8002 459000 3944
## - political_party  8    12320 463318 3946
## <none>                                450998 3946
## - age              1      8050 459048 3954
## - income           1      9234 460232 3956
##
## Step: AIC=3941
## belief_diff_mobility ~ age + income + political_party + education
##
##           Df Sum of Sq    RSS   AIC
## - education        6      7805 459625 3939
## - political_party  8    12141 463961 3940
## <none>                                451820 3941
## - age              1      7982 459801 3949
## - income           1      9397 461216 3951
##
## Step: AIC=3939
## belief_diff_mobility ~ age + income + political_party
##
##           Df Sum of Sq    RSS   AIC
## - political_party  8    11707 471332 3938
## <none>                                459625 3939
## - age              1      9779 469404 3949
## - income           1    11815 471440 3952
##
## Step: AIC=3938
## belief_diff_mobility ~ age + income
##
##           Df Sum of Sq    RSS   AIC
## <none>                                471332 3938
## - age        1      9243 480575 3947
## - income     1    14107 485440 3953

```



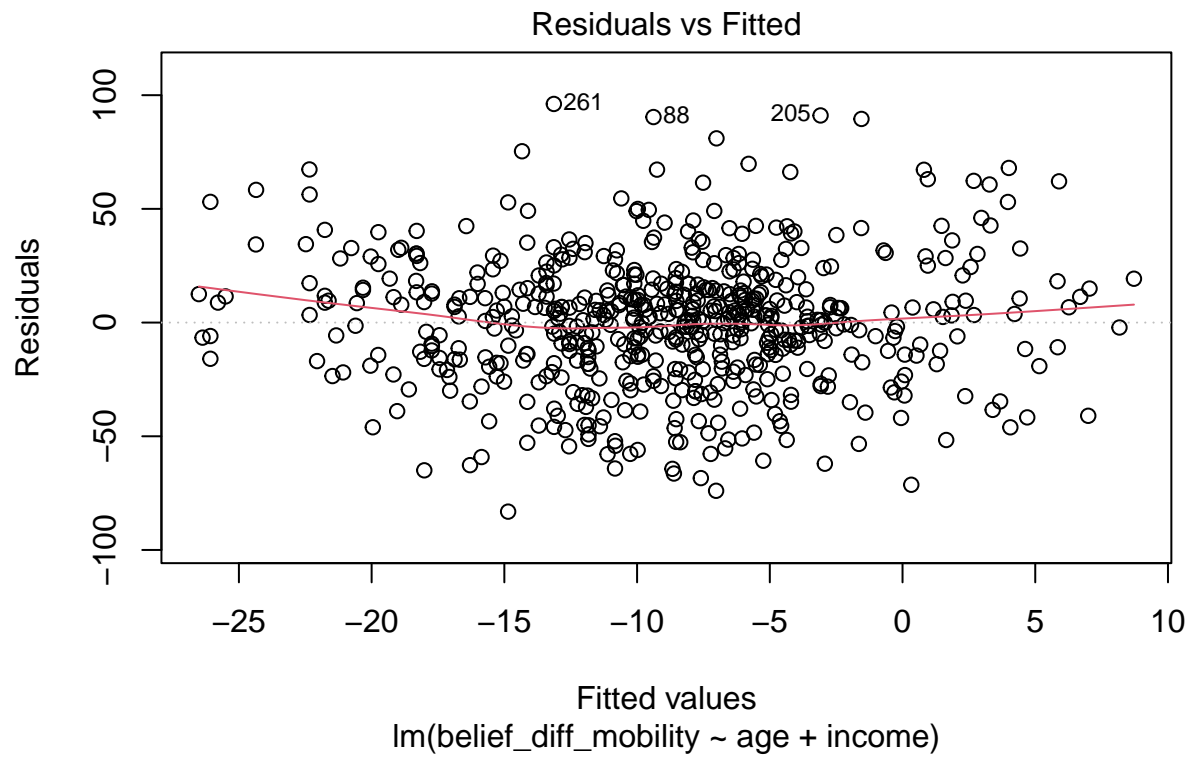
```
summary(step_model2)
```

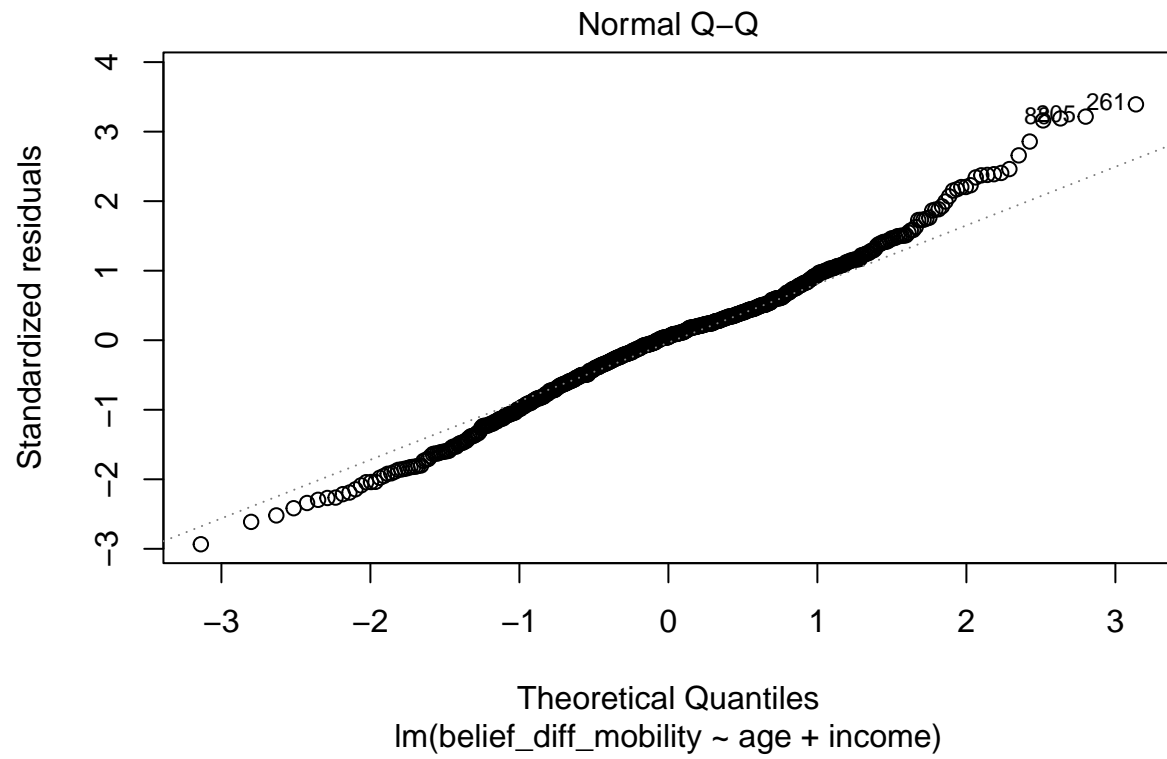
```
##
## Call:
## lm(formula = belief_diff_mobility ~ age + income, data = df2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -83.14 -17.01   1.39  15.14  96.14
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.08e+01   3.81e+00  -2.84  0.00474 **
## age          2.87e-01   8.48e-02   3.39  0.00075 ***
## income       -2.59e-03   6.19e-04  -4.18  3.3e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28 on 585 degrees of freedom
## Multiple R-squared:  0.0484, Adjusted R-squared:  0.0452
## F-statistic: 14.9 on 2 and 585 DF,  p-value: 4.95e-07
```

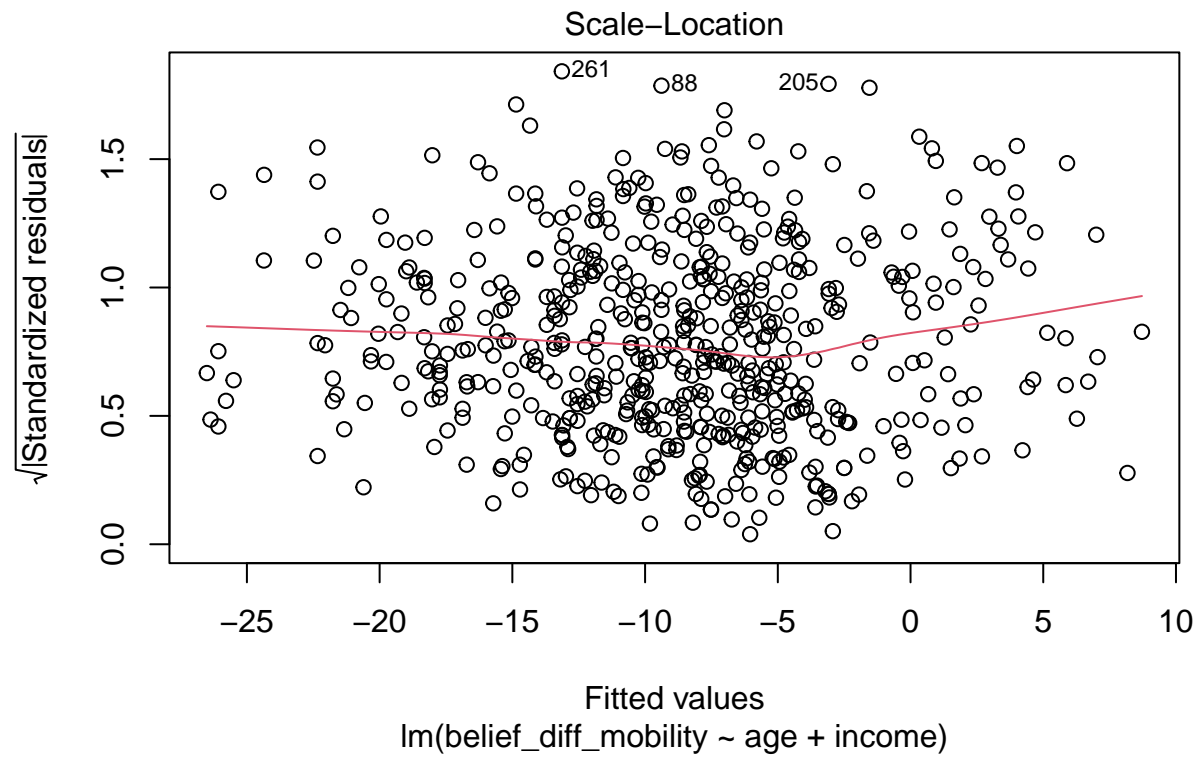
```
vif(step_model2)
```

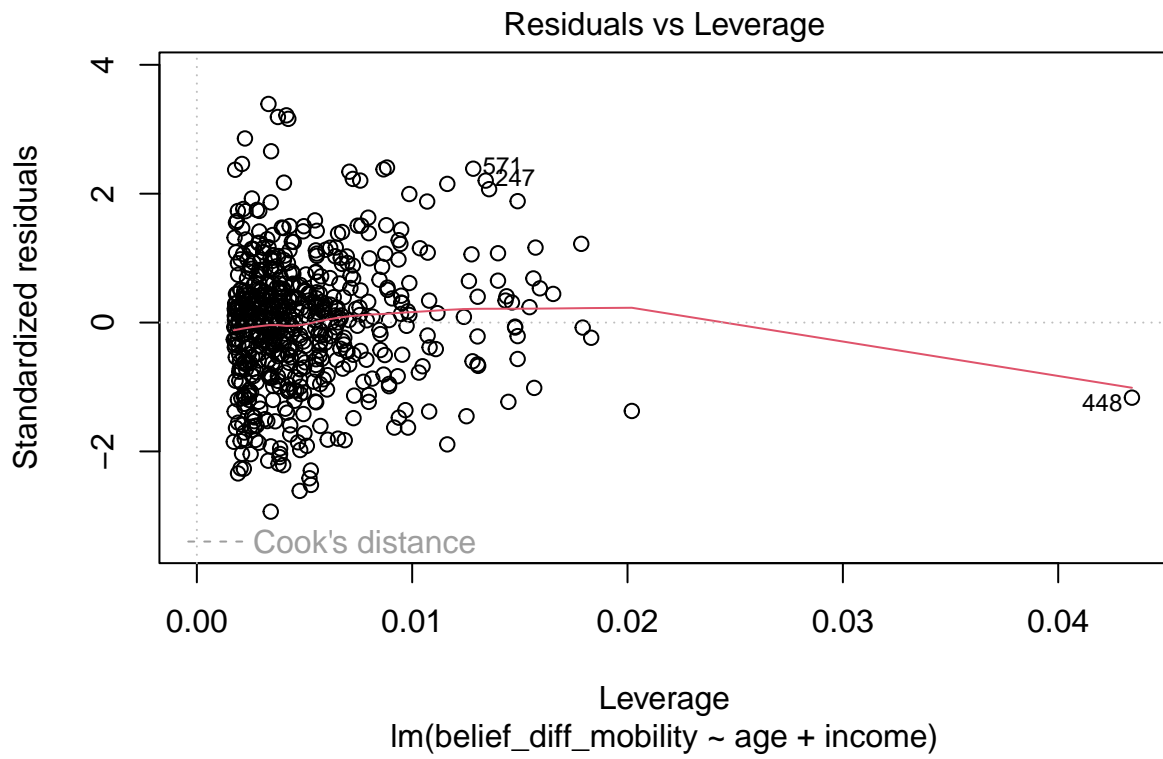
```
##      age income
##      1      1
```

```
plot(step_model2)
```





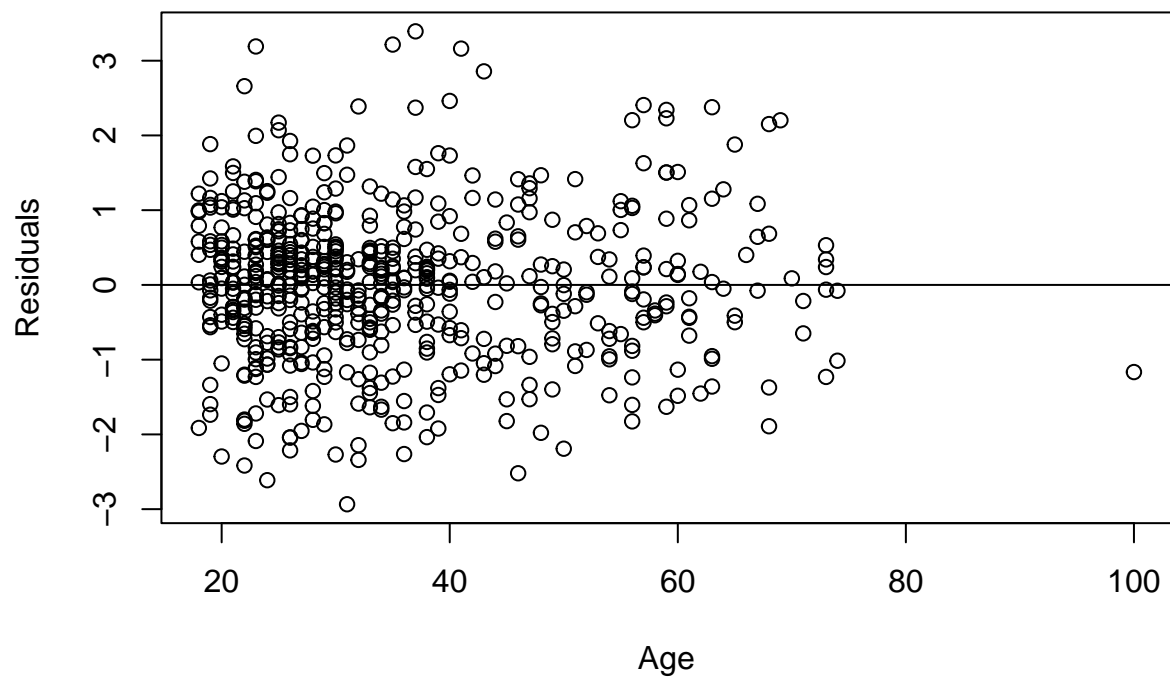




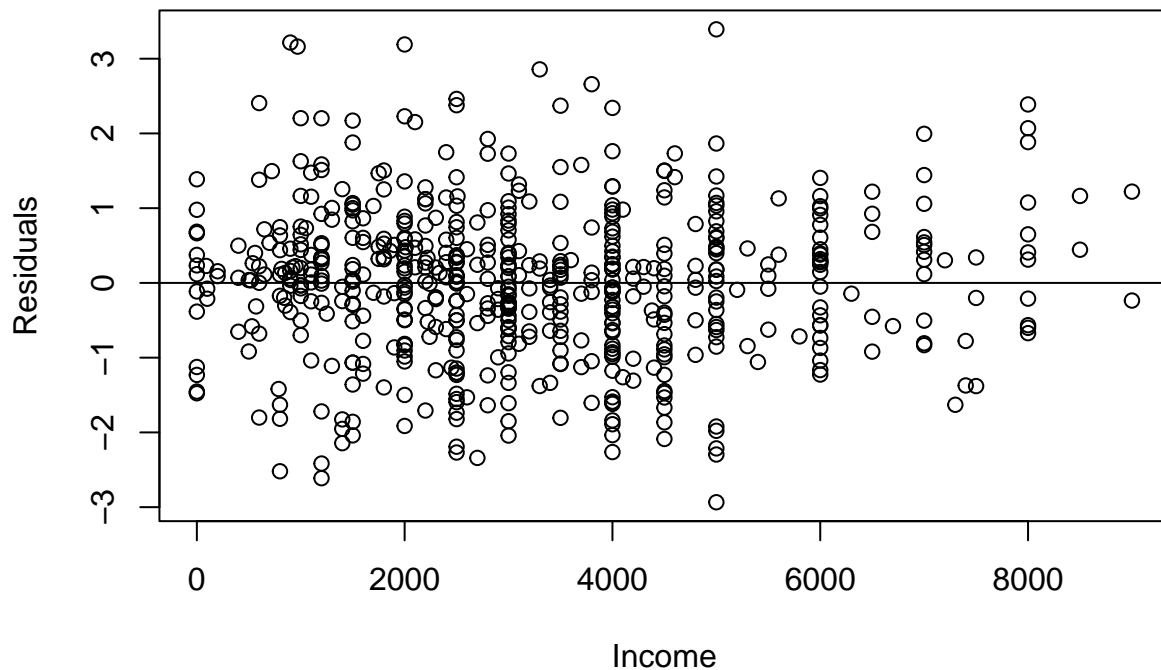
```
res2 = stdres(step_model2) ## (Standardized) Residuals

# Linearity assumption/Mean zero assumption

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



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



```
#plot(df2_scaled_new$urban_rural_class, res2, xlab = "urban_rural_class", ylab = "Residuals")
#abline(h = 0)
```

```
#plot(df2_scaled_new$education, res2, xlab = "education", ylab = "Residuals")
#abline(h = 0)
```

```
#plot(df2_scaled_new$federal_state, res2, xlab = "federal_state", ylab = "Residuals")
#abline(h = 0)
```

```
#plot(df2_scaled_new$political_party, res2, 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 -0.0085 2 0.83
## Alternative hypothesis: rho != 0
```

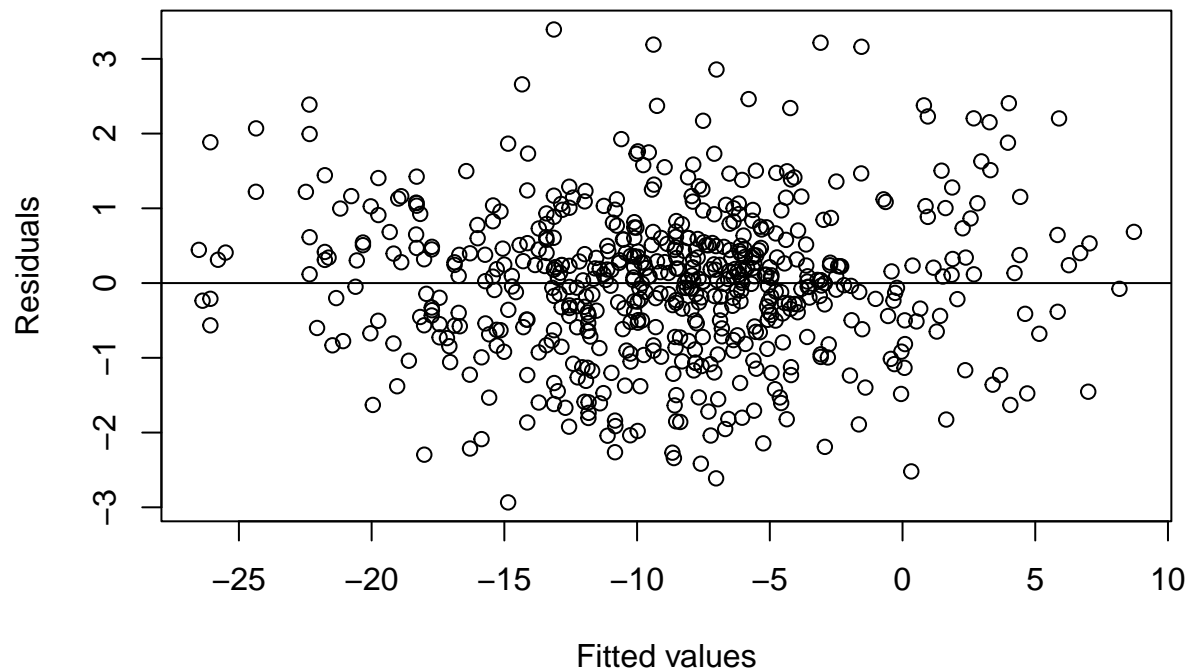
```
# Breusch-Pagan TEST: Heteroscedasticity
# H0: Homoscedasticity is present
```

```
bptest(step_model2)
```

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

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

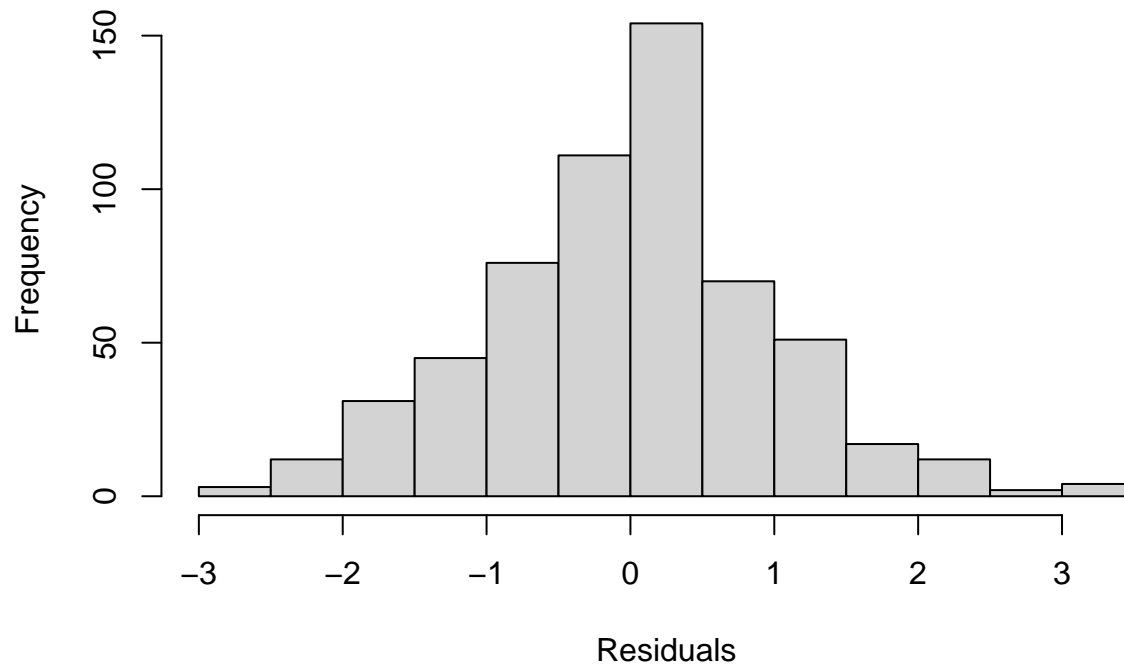
```
plot(fitted(step_model2), res2, xlab = "Fitted values", ylab = "Residuals")
abline(h = 0)
```



```
# Normality assumption
```

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


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 = 0.001
```