Appendix

Tiffany Tran

2023-12-08

```
# taking 80% subset of data
library(readr)
countries <- read_csv("C:/Users/Tiffany/Desktop/school/sta 108/countries.csv")
```

```
## Rows: 186 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (2): Country, Code
## dbl (11): LandArea, Population, Rural, Health, Internet, BirthRate, ElderlyP...
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
n <- nrow(countries)
set.seed(138)
subset_id <- sample(n, 0.8*n)
countries_subset <- countries[subset_id, ]
countries <- countries_subset
n <- nrow(countries)
```

```
# loop to assign variable names
predictors <- names(countries[3:13])
for (x in predictors){
  assign(paste(x), countries[[x]])
}
```

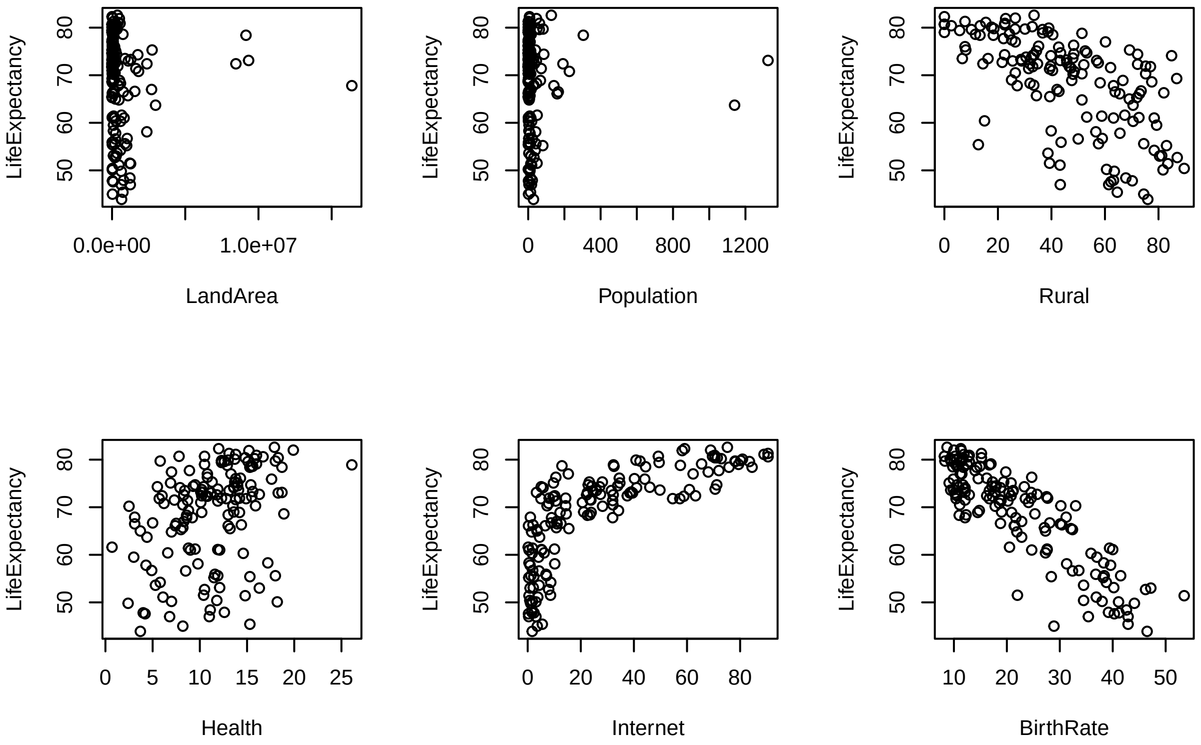
```
# descriptives of all variables
library(psych)
```

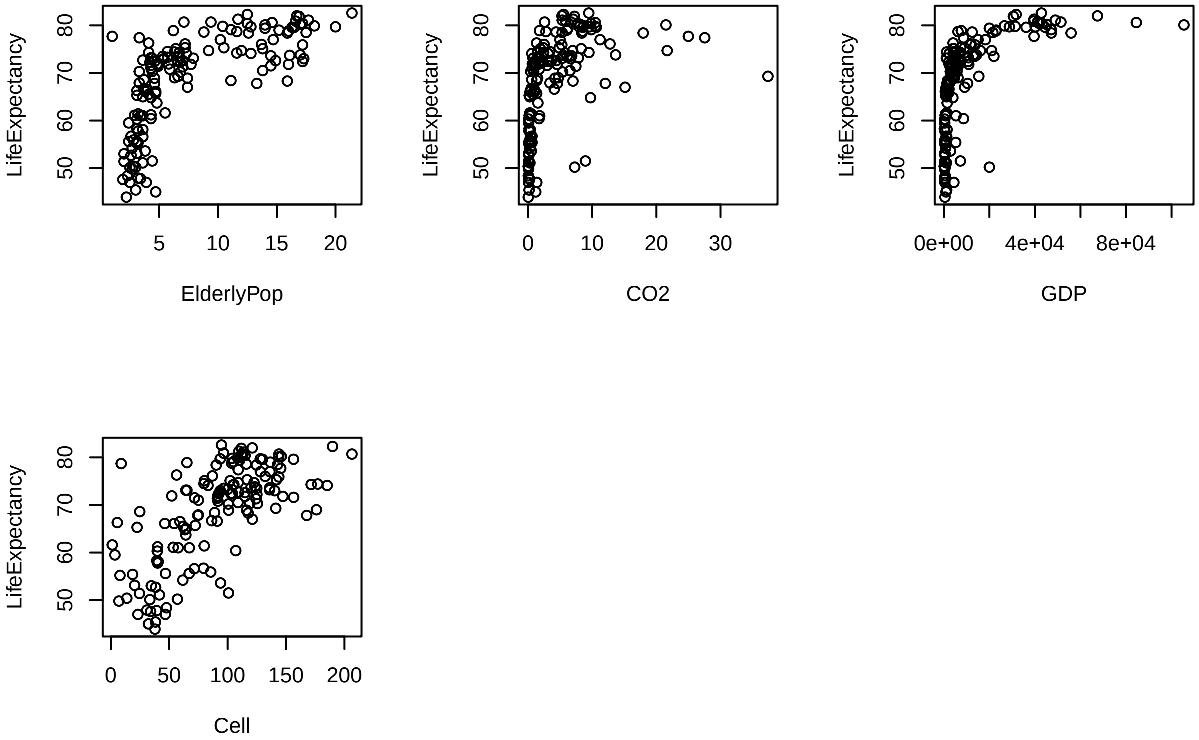
```
## Warning: package 'psych' was built under R version 4.3.2
```

```
variables = cbind(LifeExpectancy, LandArea, Population, Rural, Health, Internet,
                  BirthRate, ElderlyPop, CO2, GDP, Cell)
describe(variables, skew=F)
```

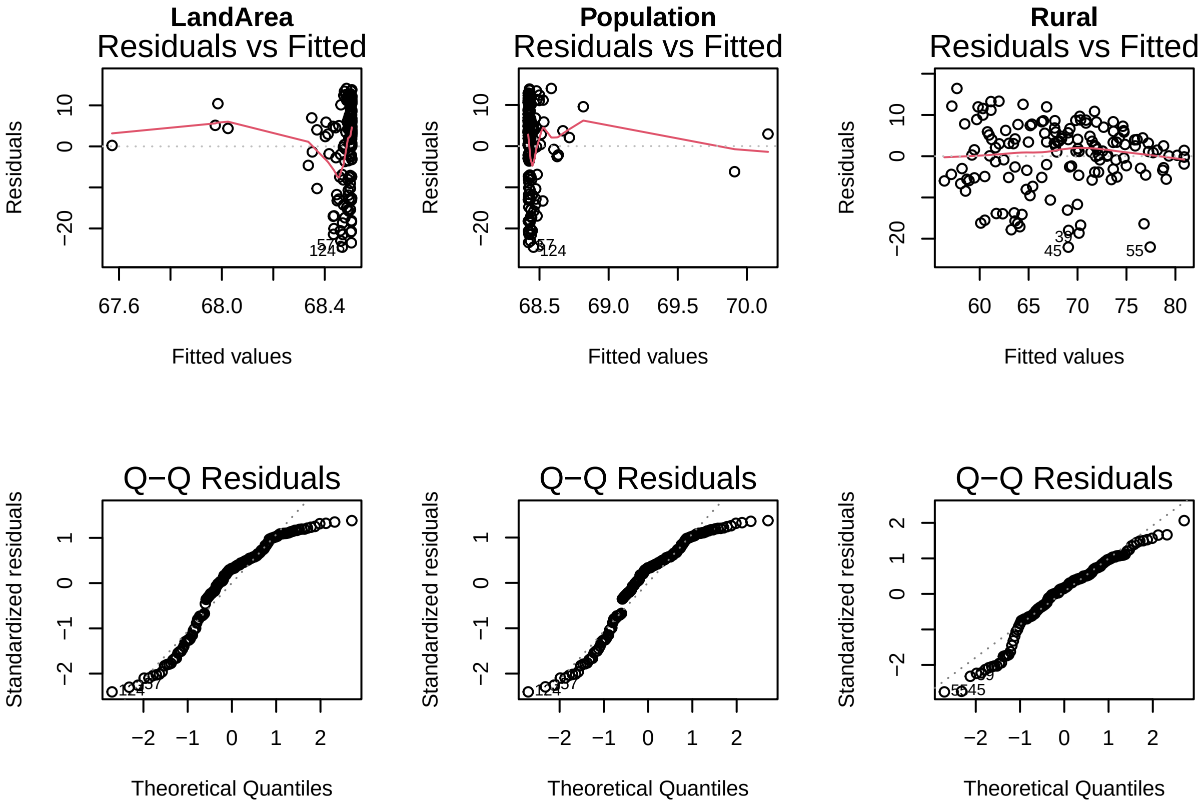
```
##                vars   n      mean         sd   min         max       range
## LifeExpectancy    1 148     68.47      10.22 43.90       82.60       38.70
## LandArea          2 148 653690.14 1874764.70 28.00 16376870.00 16376842.00
## Population        3 148     38.12     147.14  0.06     1324.65     1324.59
## Rural             4 148     45.44      23.06   0.00       89.60       89.60
## Health            5 148     11.28       4.37   0.70       26.10       25.40
## Internet          6 148     27.71      26.81   0.20       90.50       90.30
## BirthRate         7 148     22.30      11.12   8.20       53.50       45.30
## ElderlyPop        8 148      7.76       5.23   1.00       21.40       20.40
## CO2               9 148      4.53       5.68   0.02       37.39       37.37
## GDP              10 148  12025.88   17479.17 192.12   105437.67   105245.55
## Cell             11 148     90.29      43.55   1.24      206.43      205.19
##                         se
## LifeExpectancy      0.84
## LandArea       154104.71
## Population         12.09
## Rural               1.90
## Health              0.36
## Internet            2.20
## BirthRate           0.91
## ElderlyPop          0.43
## CO2                 0.47
## GDP              1436.78
## Cell                3.58
```

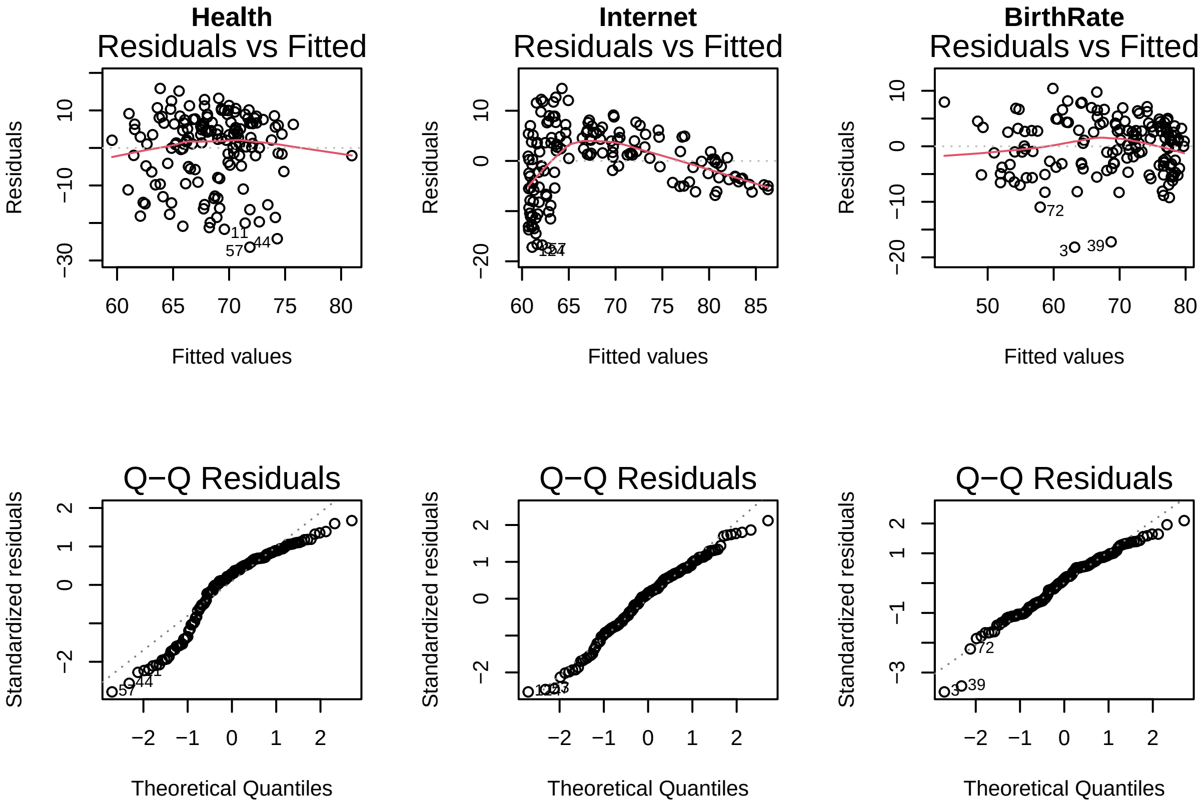
```
# simple linear regression for each variable
  # start with scatterplots
par(mfrow=c(2,3))
for (x in predictors){
  if (x == "LifeExpectancy"){
    next
  }
  plot(countries[[x]], LifeExpectancy, xlab = x)
}
```

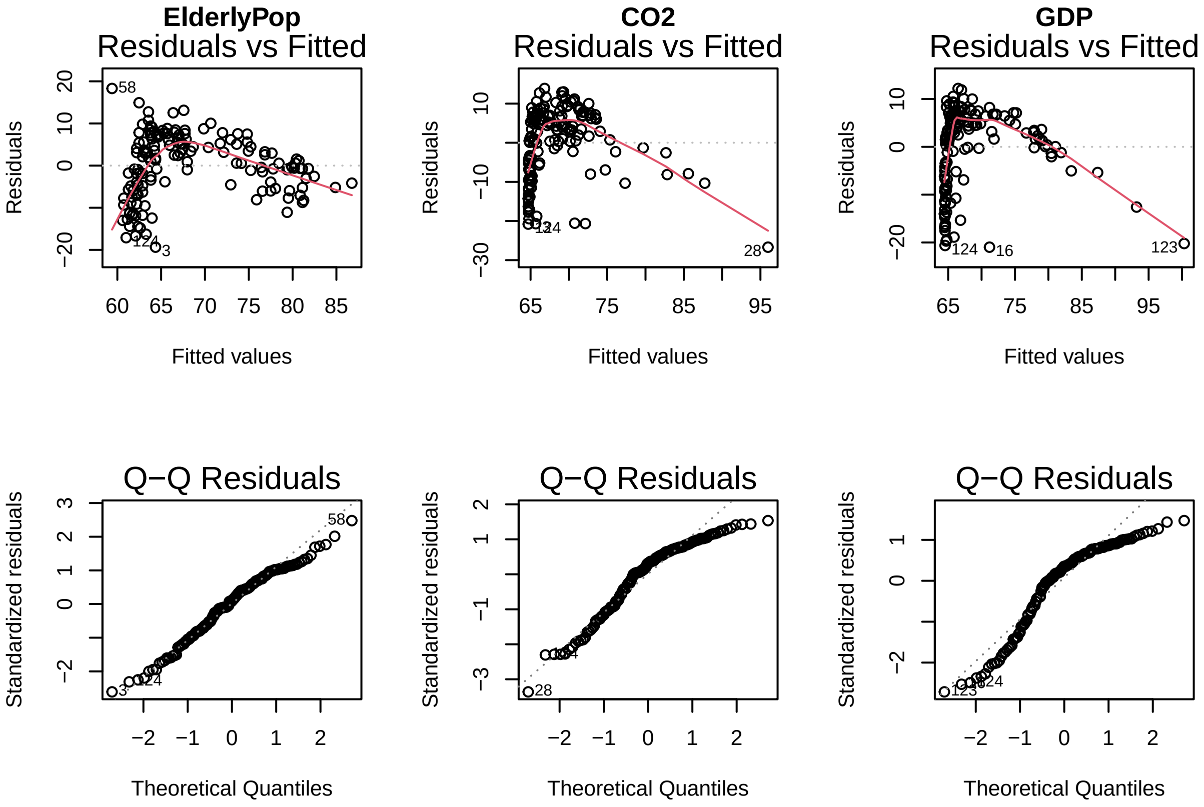

LifeExpectancy
50
60
70
80
0.0e+00
1.0e+07
LandArea
LifeExpectancy
50
60
70
80
0
400
800
1200
Population
LifeExpectancy
50
60
70
80
0
20
40
60
80
Rural
LifeExpectancy
50
60
70
80
0
5
10
15
20
25
Health
LifeExpectancy
50
60
70
80
0
20
40
60
80
Internet
LifeExpectancy
50
60
70
80
10
20
30
40
50
BirthRate

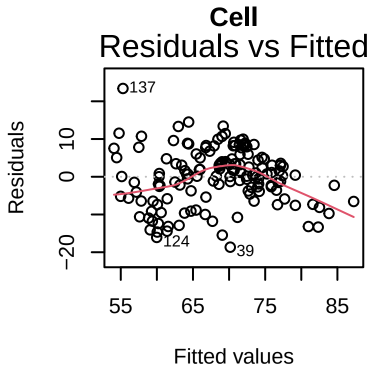


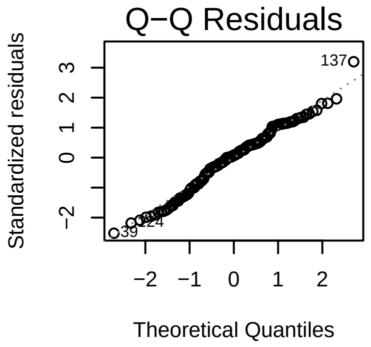
```r
# look at residual vs predictor and q-q plots to assess assumptions
par(mfcol=c(2,3))

for (x in predictors){
  if (x == "LifeExpectancy"){
    next
  }
  plot(lm(LifeExpectancy ~ countries[[x]]), which = 1, main = x)
  plot(lm(LifeExpectancy ~ countries[[x]]), which = 2)
}
```


LandArea
Residuals vs Fitted
Residuals
Fitted values
67.6
68.0
68.4
10
0
−20
57
124
Population
Residuals vs Fitted
Residuals
Fitted values
68.5
69.0
69.5
70.0
10
0
−20
57
124
Rural
Residuals vs Fitted
Residuals
Fitted values
60
65
70
75
80
10
0
−20
39
45
55
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
−2
−1
0
1
2
124
57
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
−2
−1
0
1
2
124
57
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
−2
−1
0
1
2
55
45
39


Health
Residuals vs Fitted
Residuals
Fitted values
10
−10
−30
60
65
70
75
80
11
57
44
Internet
Residuals vs Fitted
Residuals
Fitted values
10
0
−20
60
65
70
75
80
85
BirthRate
Residuals vs Fitted
Residuals
Fitted values
10
0
−10
−20
50
60
70
80
72
3
39
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
2
1
0
−2
−2
−1
0
1
2
57
44
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
2
1
0
−2
−2
−1
0
1
2
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
2
1
−1
−3
−2
−1
0
1
2
72
3
39


ElderlyPop
Residuals vs Fitted
Residuals
Fitted values
−20
0
10
20
60
65
70
75
80
85
58
124
3
CO2
Residuals vs Fitted
Residuals
Fitted values
−30
−10
10
65
75
85
95
124
28
GDP
Residuals vs Fitted
Residuals
Fitted values
−20
0
10
65
75
85
95
124
16
123
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
−2
0
1
2
3
−2
−1
0
1
2
58
3
124
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
−3
−1
1
2
−2
−1
0
1
2
124
28
Q−Q Residuals
Standardized residuals
Theoretical Quantiles
−2
0
1
−2
−1
0
1
2
123
16
124





```
# hypothesis test for slope (t-test)
design = cbind(LandArea, Population, Rural, Health, Internet, BirthRate,
               ElderlyPop, CO2, GDP, Cell)
for (x in predictors){
  if (x == "LifeExpectancy"){
    next
  }
  # print(x)
  x <- countries[[x]]
  model <- lm(LifeExpectancy ~ x)
  # print(summary(model))
  # to shorten appendix, not all summaries will be printed
}

# two variables with not significant p-values
summary(lm(LifeExpectancy ~ LandArea))
```

```
##
## Call:
## lm(formula = LifeExpectancy ~ LandArea)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.568  -7.329   3.295   7.424  14.115
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.851e+01  8.929e-01  76.718   <2e-16 ***
## LandArea    -5.695e-08  4.511e-07  -0.126      0.9
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 10.25 on 146 degrees of freedom
## Multiple R-squared:  0.0001092,  Adjusted R-squared:  -0.006739
## F-statistic: 0.01594 on 1 and 146 DF,  p-value: 0.8997
```

```
summary(lm(LifeExpectancy ~ Population))
```

```
##
## Call:
## lm(formula = LifeExpectancy ~ Population)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.556  -7.253   3.279   7.501  14.014
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 68.418315   0.870761  78.573   <2e-16 ***
## Population   0.001310   0.005747   0.228     0.82
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 10.25 on 146 degrees of freedom
## Multiple R-squared:  0.0003556,  Adjusted R-squared:  -0.006491
## F-statistic: 0.05193 on 1 and 146 DF,  p-value: 0.8201
```

```
# initial full model summary with all variables to determine most significant
model <- lm(LifeExpectancy ~ LandArea+Population+Rural+Health+Internet+
              BirthRate+ElderlyPop+CO2+GDP+Cell)
summary(model)
```

```
##
## Call:
## lm(formula = LifeExpectancy ~ LandArea + Population + Rural +
##     Health + Internet + BirthRate + ElderlyPop + CO2 + GDP +
##     Cell)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.548  -2.521   0.259   2.502  10.058
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.098e+01  3.183e+00  25.442  < 2e-16 ***
## LandArea    -3.294e-07  2.372e-07  -1.389  0.16722
## Population   1.808e-03  3.042e-03   0.594  0.55326
## Rural       -2.131e-02  2.334e-02  -0.913  0.36275
## Health       3.119e-01  9.325e-02   3.345  0.00106 **
## Internet     6.661e-02  3.293e-02   2.023  0.04507 *
## BirthRate   -7.083e-01  6.812e-02 -10.398  < 2e-16 ***
## ElderlyPop  -4.240e-01  1.386e-01  -3.060  0.00266 **
## CO2         -1.157e-01  8.805e-02  -1.314  0.19090
## GDP          4.847e-05  3.658e-05   1.325  0.18744
## Cell         2.514e-02  1.282e-02   1.961  0.05189 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.431 on 137 degrees of freedom
## Multiple R-squared:  0.8248, Adjusted R-squared:  0.812
## F-statistic: 64.47 on 10 and 137 DF,  p-value: < 2.2e-16
```

```
# use leaps to generate order of forward selection
library(leaps)
```

```
## Warning: package 'leaps' was built under R version 4.3.2
```

```
modelforward <- regsubsets(LifeExpectancy ~ LandArea+Population+Rural+Health+
              Internet+BirthRate+ElderlyPop+CO2+GDP+Cell, method ="forward",
              data=countries, nvmax = 9)
summary(modelforward)
```

```
## Subset selection object
## Call: regsubsets.formula(LifeExpectancy ~ LandArea + Population + Rural +
##     Health + Internet + BirthRate + ElderlyPop + CO2 + GDP +
##     Cell, method = "forward", data = countries, nvmax = 9)
## 10 Variables  (and intercept)
##            Forced in Forced out
## LandArea       FALSE      FALSE
## Population     FALSE      FALSE
## Rural          FALSE      FALSE
## Health         FALSE      FALSE
## Internet       FALSE      FALSE
## BirthRate      FALSE      FALSE
## ElderlyPop     FALSE      FALSE
## CO2            FALSE      FALSE
## GDP            FALSE      FALSE
## Cell           FALSE      FALSE
## 1 subsets of each size up to 9
## Selection Algorithm: forward
##          LandArea Population Rural Health Internet BirthRate ElderlyPop CO2 GDP
## 1  ( 1 ) " "      " "        " "   " "    " "      "*"       " "        " " " "
## 2  ( 1 ) " "      " "        " "   "*"    " "      "*"       " "        " " " "
## 3  ( 1 ) " "      " "        " "   "*"    "*"      "*"       " "        " " " "
## 4  ( 1 ) " "      " "        " "   "*"    "*"      "*"       "*"        " " " "
## 5  ( 1 ) " "      " "        " "   "*"    "*"      "*"       "*"        " " " "
## 6  ( 1 ) " "      " "        " "   "*"    "*"      "*"       "*"        "*" " "
## 7  ( 1 ) " "      " "        " "   "*"    "*"      "*"       "*"        "*" "*"
## 8  ( 1 ) "*"      " "        " "   "*"    "*"      "*"       "*"        "*" "*"
## 9  ( 1 ) "*"      " "        "*"   "*"    "*"      "*"       "*"        "*" "*"
##          Cell
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) "*"
## 6  ( 1 ) "*"
## 7  ( 1 ) "*"
## 8  ( 1 ) "*"
## 9  ( 1 ) "*"
```

```
# assess forward selection
summary(lm(LifeExpectancy ~ BirthRate + Health + Internet + ElderlyPop + Cell))
```

```
##
## Call:
## lm(formula = LifeExpectancy ~ BirthRate + Health + Internet +
##     ElderlyPop + Cell)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.7525  -2.8776   0.0746   2.8033  10.5680
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 78.39354    2.78182  28.181  < 2e-16 ***
## BirthRate   -0.69085    0.06380 -10.827  < 2e-16 ***
## Health       0.34149    0.09202   3.711 0.000296 ***
## Internet     0.08849    0.02460   3.596 0.000445 ***
## ElderlyPop  -0.39051    0.13518  -2.889 0.004475 **
## Cell         0.02445    0.01225   1.996 0.047872 *
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 4.46 on 142 degrees of freedom
## Multiple R-squared:  0.816,  Adjusted R-squared:  0.8095
## F-statistic:   126 on 5 and 142 DF,  p-value: < 2.2e-16
```

```
summary(lm(LifeExpectancy ~ BirthRate + Health + Internet + ElderlyPop +
             Cell + CO2))
```

```
##
## Call:
## lm(formula = LifeExpectancy ~ BirthRate + Health + Internet +
##     ElderlyPop + Cell + CO2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.1119  -2.8417   0.0953   2.6737  10.3925
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 79.21076    2.82994  27.990  < 2e-16 ***
## BirthRate   -0.70681    0.06454 -10.951  < 2e-16 ***
## Health       0.32311    0.09258   3.490 0.000645 ***
## Internet     0.10357    0.02669   3.881 0.000159 ***
## ElderlyPop  -0.43028    0.13753  -3.129 0.002134 **
## Cell         0.02645    0.01228   2.153 0.033034 *
## CO2         -0.11997    0.08389  -1.430 0.154915
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.443 on 141 degrees of freedom
## Multiple R-squared:  0.8187, Adjusted R-squared:  0.8109
## F-statistic: 106.1 on 6 and 141 DF,  p-value: < 2.2e-16
```

```
# adding more variables does not seem significant! stop at Cell.
```

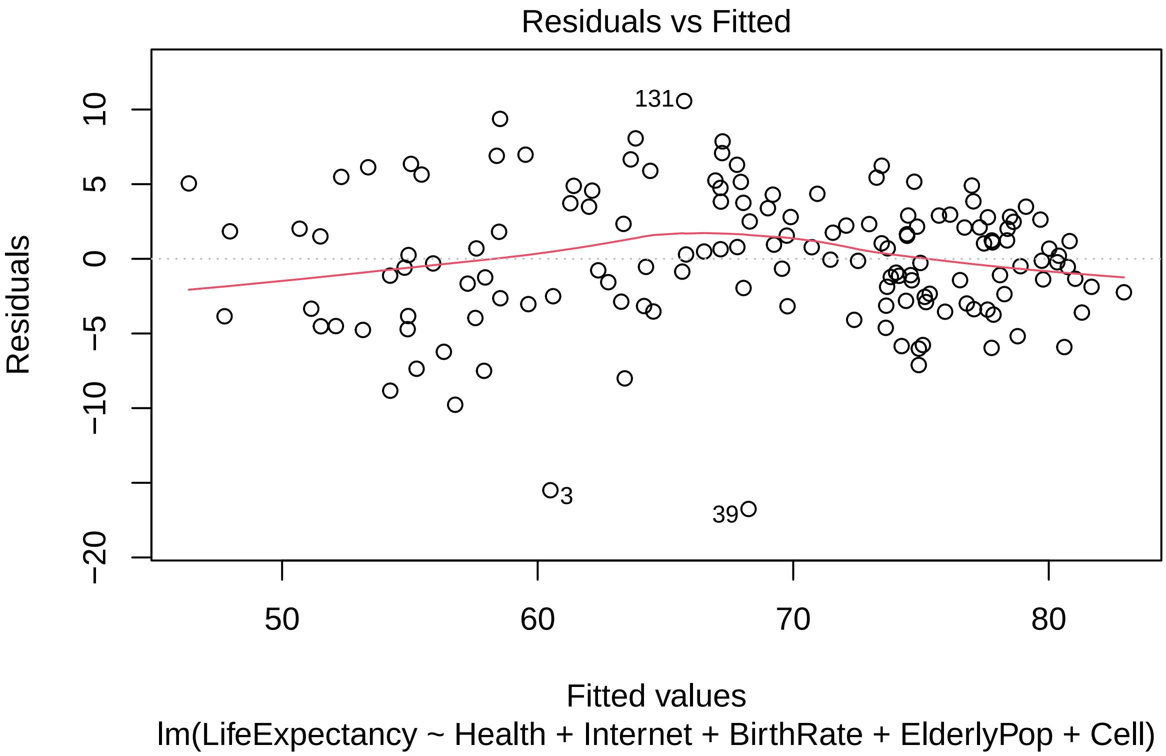
```
# perform backwards selection with leaps to check model
modelbackward <- regsubsets(LifeExpectancy ~ LandArea+Population+Rural+Health+
              Internet+BirthRate+ElderlyPop+CO2+GDP+Cell, method ="backward",
              data=countries, nvmax = 9)
summary(modelbackward)
```

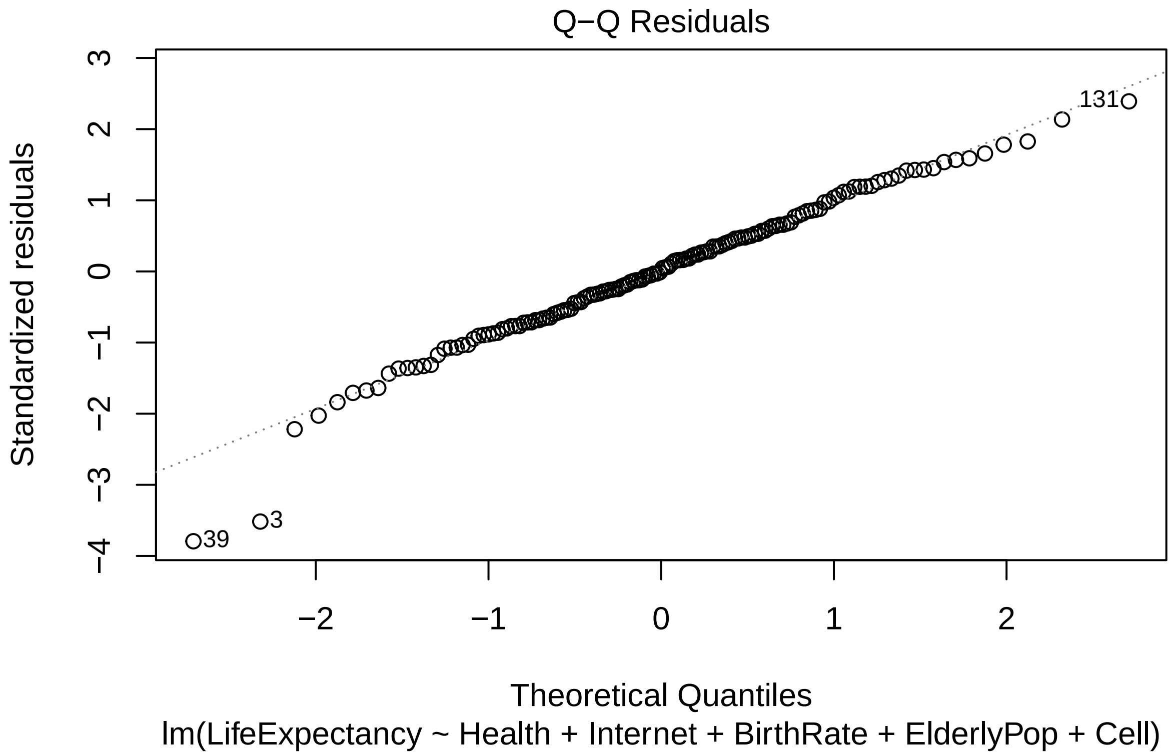
```
## Subset selection object
## Call: regsubsets.formula(LifeExpectancy ~ LandArea + Population + Rural +
##     Health + Internet + BirthRate + ElderlyPop + CO2 + GDP +
##     Cell, method = "backward", data = countries, nvmax = 9)
## 10 Variables  (and intercept)
##            Forced in Forced out
## LandArea       FALSE      FALSE
## Population     FALSE      FALSE
## Rural          FALSE      FALSE
## Health         FALSE      FALSE
## Internet       FALSE      FALSE
## BirthRate      FALSE      FALSE
## ElderlyPop     FALSE      FALSE
## CO2            FALSE      FALSE
## GDP            FALSE      FALSE
## Cell           FALSE      FALSE
## 1 subsets of each size up to 9
## Selection Algorithm: backward
##          LandArea Population Rural Health Internet BirthRate ElderlyPop CO2 GDP
## 1  ( 1 ) " "      " "        " "   " "    " "      "*"       " "        " " " "
## 2  ( 1 ) " "      " "        " "   "*"    " "      "*"       " "        " " " "
## 3  ( 1 ) " "      " "        " "   "*"    "*"      "*"       " "        " " " "
## 4  ( 1 ) " "      " "        " "   "*"    "*"      "*"       "*"        " " " "
## 5  ( 1 ) " "      " "        " "   "*"    "*"      "*"       "*"        " " " "
## 6  ( 1 ) " "      " "        " "   "*"    "*"      "*"       "*"        "*" " "
## 7  ( 1 ) " "      " "        " "   "*"    "*"      "*"       "*"        "*" "*"
## 8  ( 1 ) "*"      " "        " "   "*"    "*"      "*"       "*"        "*" "*"
## 9  ( 1 ) "*"      " "        "*"   "*"    "*"      "*"       "*"        "*" "*"
##          Cell
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) "*"
## 6  ( 1 ) "*"
## 7  ( 1 ) "*"
## 8  ( 1 ) "*"
## 9  ( 1 ) "*"
```

```
# since Population, Rural, and LandArea do not seem significant from our
#   observations, remove for criterion testing
reddesign = cbind(Health, Internet, BirthRate, ElderlyPop, CO2, GDP, Cell)
leaps(x=reddesign, y=LifeExpectancy, names=c("Health", "Internet", "BirthRate",
      "ElderlyPop", "CO2", "GDP", "Cell"), method = "adjr2", nbest = 3)
```

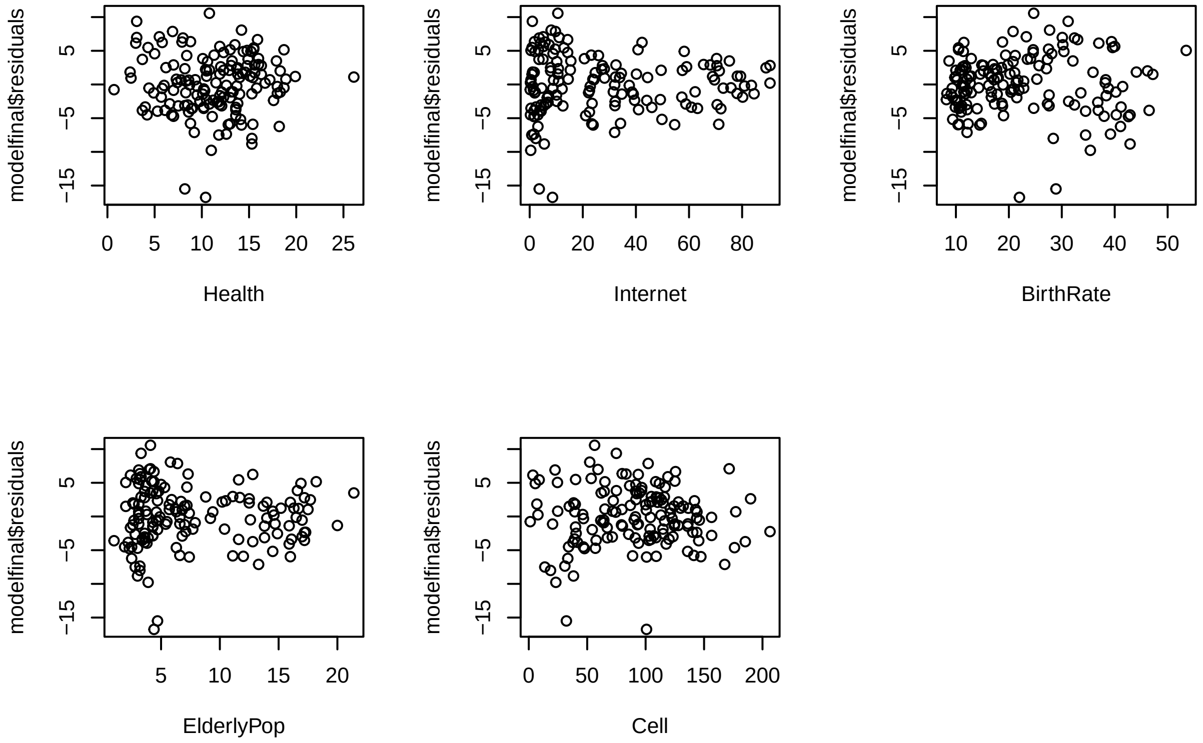
```
## $which
##   Health Internet BirthRate ElderlyPop   CO2   GDP  Cell
## 1  FALSE    FALSE      TRUE      FALSE FALSE FALSE FALSE
## 1  FALSE     TRUE     FALSE      FALSE FALSE FALSE FALSE
## 1  FALSE    FALSE     FALSE      FALSE FALSE FALSE  TRUE
## 2   TRUE    FALSE      TRUE      FALSE FALSE FALSE FALSE
## 2  FALSE     TRUE      TRUE      FALSE FALSE FALSE FALSE
## 2  FALSE    FALSE      TRUE      FALSE FALSE  TRUE FALSE
## 3   TRUE     TRUE      TRUE      FALSE FALSE FALSE FALSE
## 3   TRUE    FALSE      TRUE      FALSE FALSE  TRUE FALSE
## 3   TRUE    FALSE      TRUE      FALSE FALSE FALSE  TRUE
## 4   TRUE     TRUE      TRUE       TRUE FALSE FALSE FALSE
## 4   TRUE    FALSE      TRUE       TRUE FALSE  TRUE FALSE
## 4   TRUE    FALSE      TRUE      FALSE FALSE  TRUE  TRUE
## 5   TRUE     TRUE      TRUE       TRUE FALSE FALSE  TRUE
## 5   TRUE     TRUE      TRUE       TRUE FALSE  TRUE FALSE
## 5   TRUE     TRUE      TRUE       TRUE  TRUE FALSE FALSE
## 6   TRUE     TRUE      TRUE       TRUE  TRUE FALSE  TRUE
## 6   TRUE     TRUE      TRUE       TRUE FALSE  TRUE  TRUE
## 6   TRUE     TRUE      TRUE       TRUE  TRUE  TRUE FALSE
## 7   TRUE     TRUE      TRUE       TRUE  TRUE  TRUE  TRUE
##
## $label
## [1] "(Intercept)" "Health"      "Internet"    "BirthRate"   "ElderlyPop"
## [6] "CO2"         "GDP"         "Cell"
##
## $size
##  [1] 2 2 2 3 3 3 4 4 4 5 5 5 6 6 6 7 7 7 8
##
## $adjr2
##  [1] 0.7602355 0.5517384 0.4716335 0.7826309 0.7821891 0.7796151 0.7939919
##  [8] 0.7931627 0.7926749 0.8055759 0.7999909 0.7998504 0.8095488 0.8063919
## [15] 0.8061010 0.8109402 0.8101789 0.8079946 0.8127930
```

```
# generate residual plots to check assumptions
modelfinal <- lm(LifeExpectancy ~ Health+Internet+BirthRate+ElderlyPop+Cell)
plot(modelfinal, which = c(1, 2))
```


Residuals vs Fitted
Residuals
10
5
0
−5
−10
−20
50
60
70
80
Fitted values
lm(LifeExpectancy ~ Health + Internet + BirthRate + ElderlyPop + Cell)
131
3
39

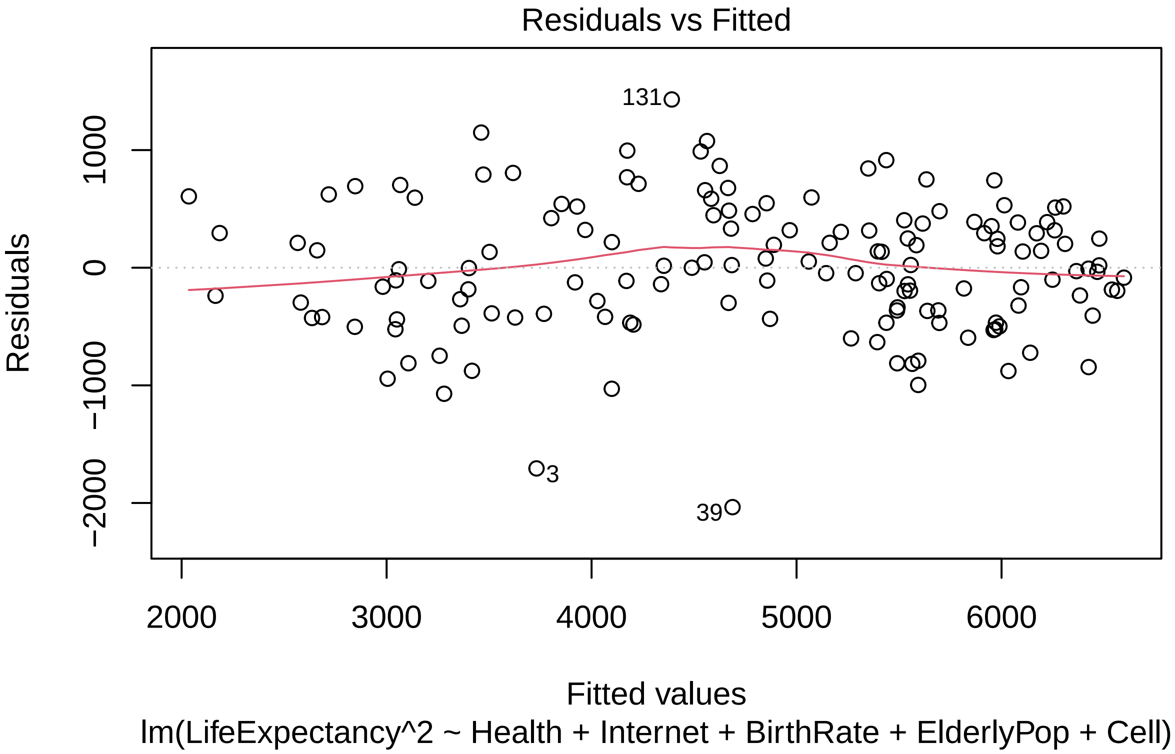


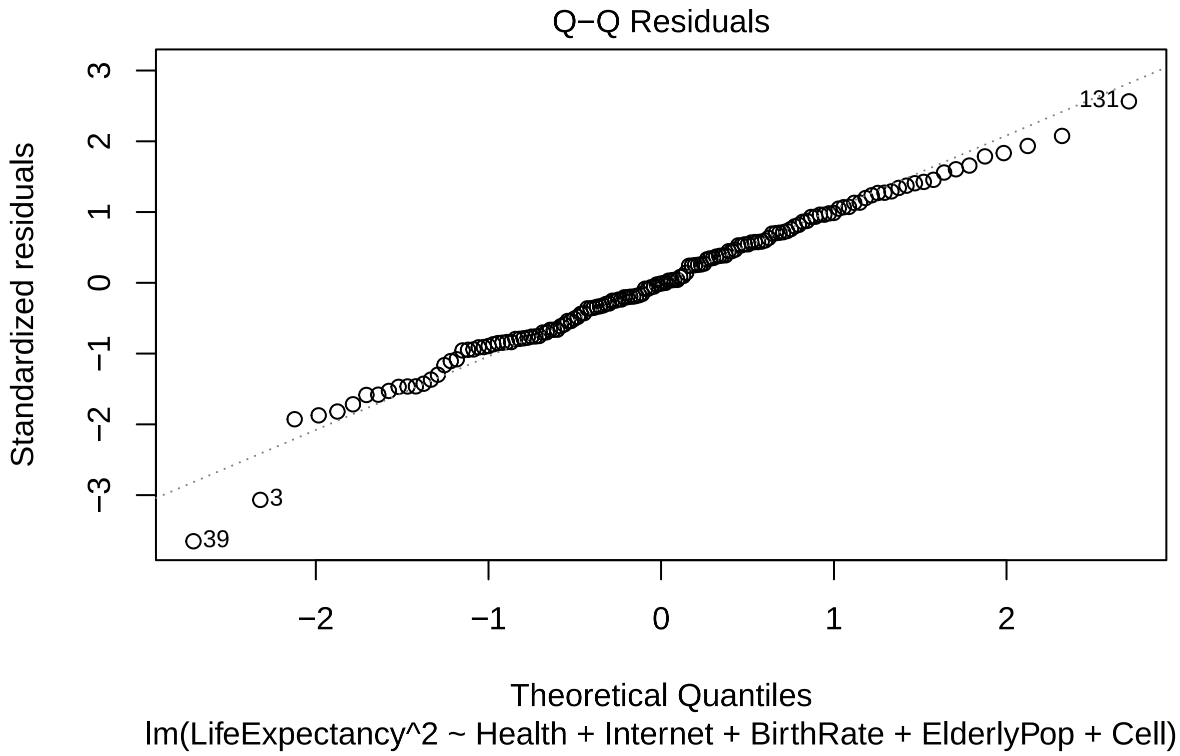
```r
# check to see if nonlinearity is caused by predictors
par(mfrow=c(2,3))
plot(Health, modelfinal$residuals)
plot(Internet, modelfinal$residuals)
plot(BirthRate, modelfinal$residuals)
plot(ElderlyPop, modelfinal$residuals)
plot(Cell, modelfinal$residuals)
```



```
# boxcox transformation
  #library(MASS)
  #boxcox(LifeExpectancy ~ Health+Internet+BirthRate+ElderlyPop+Cell, plotit=T)
# shown in report
```

```
# create new final model
modelnew <- lm(LifeExpectancy^2 ~ Health+Internet+BirthRate+ElderlyPop+Cell)
plot(modelnew, which = c(1,2))
```


Residuals vs Fitted
Residuals
Fitted values
lm(LifeExpectancy^2 ~ Health + Internet + BirthRate + ElderlyPop + Cell)
1000
0
−1000
−2000
2000
3000
4000
5000
6000
131
3
39



```r
# reevaluate partial F tests
summary(lm(LifeExpectancy^2 ~ BirthRate + Health + Internet + ElderlyPop + Cell))
```

```
##
## Call:
## lm(formula = LifeExpectancy^2 ~ BirthRate + Health + Internet +
##     ElderlyPop + Cell)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2035.38  -389.00    -5.52   385.05  1430.43
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5822.218    351.037  16.586  < 2e-16 ***
## BirthRate    -83.070      8.052 -10.317  < 2e-16 ***
## Health        44.904     11.612   3.867 0.000167 ***
## Internet      13.898      3.105   4.476 1.55e-05 ***
## ElderlyPop   -43.246     17.059  -2.535 0.012324 *
## Cell           2.945      1.546   1.905 0.058800 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 562.8 on 142 degrees of freedom
## Multiple R-squared:  0.826,  Adjusted R-squared:  0.8199
## F-statistic: 134.8 on 5 and 142 DF,  p-value: < 2.2e-16
```

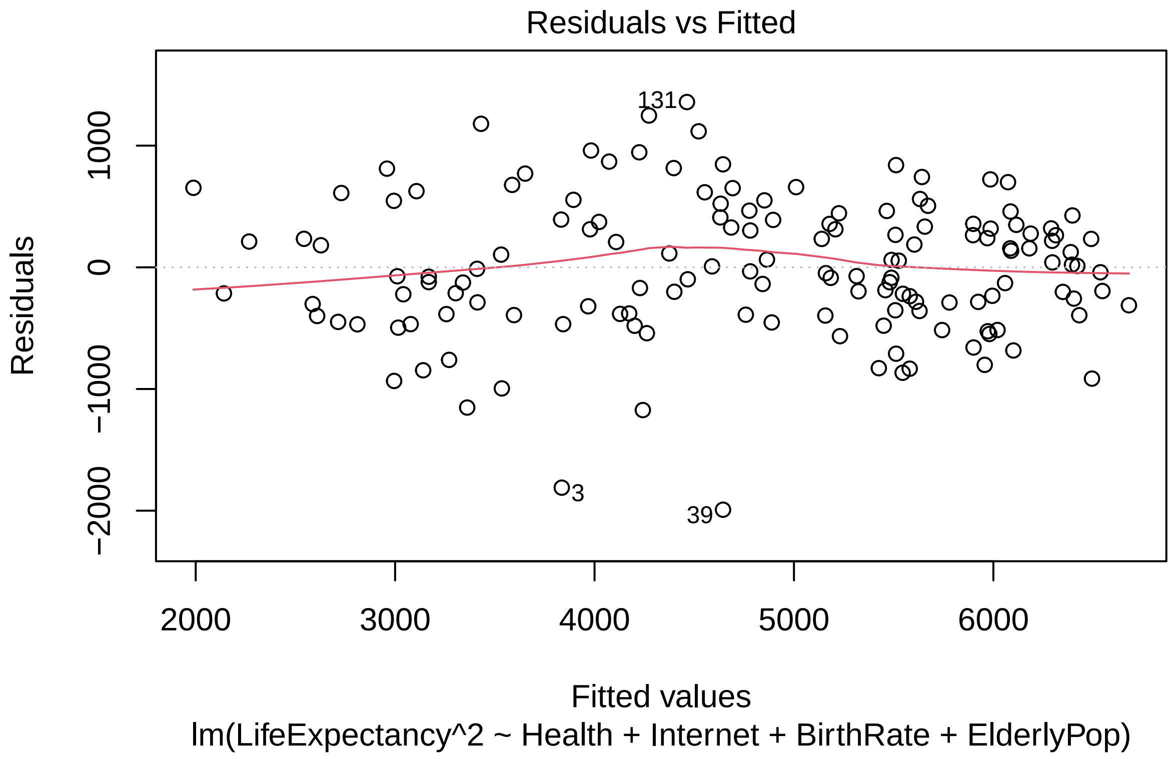
```r
# reevaluate criterion test
reddesign2 = cbind(Health, Internet, BirthRate, ElderlyPop, Cell)
leaps(x=reddesign2, y=LifeExpectancy^2, names=c("Health", "Internet",
      "BirthRate", "ElderlyPop", "Cell"), method = "adjr2", nbest = 3)
```

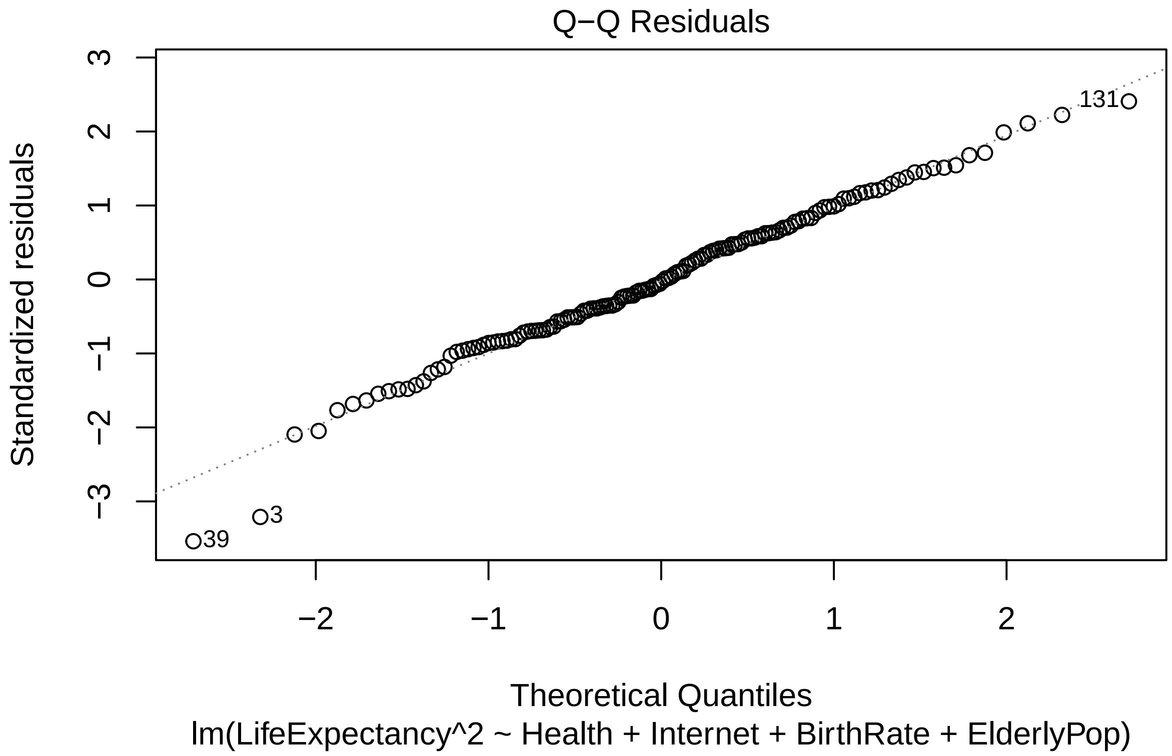
```
## $which
##   Health Internet BirthRate ElderlyPop  Cell
## 1  FALSE    FALSE      TRUE      FALSE FALSE
## 1  FALSE     TRUE     FALSE      FALSE FALSE
## 1  FALSE    FALSE     FALSE       TRUE FALSE
## 2  FALSE     TRUE      TRUE      FALSE FALSE
## 2   TRUE    FALSE      TRUE      FALSE FALSE
## 2  FALSE    FALSE      TRUE      FALSE  TRUE
## 3   TRUE     TRUE      TRUE      FALSE FALSE
## 3  FALSE     TRUE      TRUE       TRUE FALSE
## 3  FALSE     TRUE      TRUE      FALSE  TRUE
## 4   TRUE     TRUE      TRUE       TRUE FALSE
## 4   TRUE     TRUE      TRUE      FALSE  TRUE
## 4  FALSE     TRUE      TRUE       TRUE  TRUE
## 5   TRUE     TRUE      TRUE       TRUE  TRUE
##
## $label
## [1] "(Intercept)" "Health"      "Internet"    "BirthRate"   "ElderlyPop"
## [6] "Cell"
##
## $size
##  [1] 2 2 2 3 3 3 4 4 4 5 5 5 6
##
## $adjr2
##  [1] 0.7587234 0.5939574 0.4975963 0.7953208 0.7865722 0.7689083 0.8082976
##  [8] 0.7999759 0.7987442 0.8165622 0.8130380 0.8022978 0.8198738
```

```r
# remove cell for new final model
modelnew2 <- lm(LifeExpectancy^2 ~ Health+Internet+BirthRate+ElderlyPop)
summary(modelnew2)
```

```
##
## Call:
## lm(formula = LifeExpectancy^2 ~ Health + Internet + BirthRate +
##     ElderlyPop)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1992.02  -380.16   -22.69   361.41  1358.30
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6243.653    275.050  22.700  < 2e-16 ***
## Health        43.826     11.705   3.744 0.000261 ***
## Internet      15.343      3.038   5.050 1.33e-06 ***
## BirthRate    -90.051      7.235 -12.447  < 2e-16 ***
## ElderlyPop   -46.820     17.110  -2.736 0.007000 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 567.9 on 143 degrees of freedom
## Multiple R-squared:  0.8216, Adjusted R-squared:  0.8166
## F-statistic: 164.6 on 4 and 143 DF,  p-value: < 2.2e-16
```

```
plot(modelnew2, which = c(1,2))
```





```r
# multicollinearity
library(faraway)
```

```
## Warning: package 'faraway' was built under R version 4.3.2

##
## Attaching package: 'faraway'

## The following object is masked from 'package:psych':
##
##     logit
```

```r
vif(modelnew2)
```

```
##     Health   Internet  BirthRate ElderlyPop
##   1.190737   3.024301   2.949096   3.652759
```