

## Final project 242

```
setwd("~/Desktop/Math 242(applied stats)")
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.4.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(leaps)
library(car)

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##   recode

students=read.csv("student-mat.csv",sep=";",header=TRUE)

students$age=as.numeric(students$age)
students$traveltime=as.numeric(students$traveltime)
students$studytime=as.numeric(students$studytime)
students$freetime=as.numeric(students$freetime)
students$goout=as.numeric(students$goout)
students$Dalc=as.numeric(students$Dalc)
students$Walc=as.numeric(students$Walc)
students$health=as.numeric(students$health)
students$absences=as.numeric(students$absences)
students$G1=as.numeric(students$G1)
students$G2=as.numeric(students$G2)
students$G3=as.numeric(students$G3)
students$famrel=as.numeric(students$famrel)
students$failures=as.numeric(students$failures)
#str(students)
```

```
best.subset=regsubsets(G3~.,students,nvmax =10)
```

```
sum=summary(best.subset)
```

```
sum$outmat
```

```
##          schoolMS sexM age addressU famsizeLE3 PstatusT Medu Fedu
## 1 ( 1 ) " "      " "  " " " "      " "      " "      " "
## 2 ( 1 ) " "      " "  " " " "      " "      " "      " "
## 3 ( 1 ) " "      " "  " " " "      " "      " "      " "
## 4 ( 1 ) " "      " "  " " " "      " "      " "      " "
## 5 ( 1 ) " "      " "  "*" " "      " "      " "      " "
## 6 ( 1 ) " "      " "  "*" " "      " "      " "      " "
## 7 ( 1 ) " "      " "  "*" " "      " "      " "      " "
## 8 ( 1 ) "*"      " "  "*" " "      " "      " "      " "
## 9 ( 1 ) "*"      " "  "*" " "      " "      " "      " "
## 10 ( 1 ) "*"      " "  "*" " "      " "      " "      " "
##          Mjobhealth Mjobother Mjobservices Mjobteacher Fjobhealth
## 1 ( 1 ) " "          " "          " "          " "          " "
## 2 ( 1 ) " "          " "          " "          " "          " "
## 3 ( 1 ) " "          " "          " "          " "          " "
## 4 ( 1 ) " "          " "          " "          " "          " "
## 5 ( 1 ) " "          " "          " "          " "          " "
## 6 ( 1 ) " "          " "          " "          " "          " "
## 7 ( 1 ) " "          " "          " "          " "          " "
## 8 ( 1 ) " "          " "          " "          " "          " "
## 9 ( 1 ) " "          " "          " "          " "          " "
## 10 ( 1 ) " "          " "          " "          " "          " "
##          Fjobother Fjobservices Fjobteacher reasonhome reasonother
## 1 ( 1 ) " "          " "          " "          " "          " "
## 2 ( 1 ) " "          " "          " "          " "          " "
## 3 ( 1 ) " "          " "          " "          " "          " "
## 4 ( 1 ) " "          " "          " "          " "          " "
## 5 ( 1 ) " "          " "          " "          " "          " "
## 6 ( 1 ) " "          "*"          " "          " "          " "
## 7 ( 1 ) " "          "*"          " "          " "          " "
## 8 ( 1 ) " "          "*"          " "          " "          " "
## 9 ( 1 ) " "          "*"          " "          " "          " "
## 10 ( 1 ) " "          "*"          " "          " "          " "
##          reasonreputation guardianmother guardianother traveltime
## 1 ( 1 ) " "                  " "          " "          " "
## 2 ( 1 ) " "                  " "          " "          " "
## 3 ( 1 ) " "                  " "          " "          " "
## 4 ( 1 ) " "                  " "          " "          " "
## 5 ( 1 ) " "                  " "          " "          " "
## 6 ( 1 ) " "                  " "          " "          " "
## 7 ( 1 ) " "                  " "          " "          " "
## 8 ( 1 ) " "                  " "          " "          " "
## 9 ( 1 ) " "                  " "          " "          " "
## 10 ( 1 ) " "                  " "          " "          " "
##          studytime failures schoolsupyes famsupyes paidyes activitiesyes
## 1 ( 1 ) " "          " "          " "          " "          " "          " "
```

```

## 2 ( 1 ) " " " " " " " " " "
## 3 ( 1 ) " " " " " " " " " "
## 4 ( 1 ) " " " " " " " " " "
## 5 ( 1 ) " " " " " " " " " "
## 6 ( 1 ) " " " " " " " " " "
## 7 ( 1 ) " " " " " " " " " "
## 8 ( 1 ) " " " " " " " " " "
## 9 ( 1 ) " " " " " " " " " "
## 10 ( 1 ) " " " " " " " " "*"
##
nurseryyes higheryes internetyes romanticyes famrel freetime
## 1 ( 1 ) " " " " " " " " " "
## 2 ( 1 ) " " " " " " " " "*" " "
## 3 ( 1 ) " " " " " " " " "*" " "
## 4 ( 1 ) " " " " " " " " "*" " "
## 5 ( 1 ) " " " " " " " " "*" " "
## 6 ( 1 ) " " " " " " " " "*" " "
## 7 ( 1 ) " " " " " " " " "*" " "
## 8 ( 1 ) " " " " " " " " "*" " "
## 9 ( 1 ) " " " " " " " " "*" "*" " "
## 10 ( 1 ) " " " " " " " " "*" "*" " "
##
goout Dalc Walc health absences G1 G2
## 1 ( 1 ) " " " " " " " " " " "*"
## 2 ( 1 ) " " " " " " " " " " "*"
## 3 ( 1 ) " " " " " " " " "*" " " "*"
## 4 ( 1 ) " " " " " " " " "*" "*" "*"
## 5 ( 1 ) " " " " " " " " "*" "*" "*"
## 6 ( 1 ) " " " " " " " " "*" "*" "*"
## 7 ( 1 ) " " " " "*" " " " "*" "*" "*"
## 8 ( 1 ) " " " " "*" " " " "*" "*" "*"
## 9 ( 1 ) " " " " "*" " " " "*" "*" "*"
## 10 ( 1 ) " " " " "*" " " " "*" "*" "*"

```

```

null = lm(G3~1, data = students)
full = lm(G3~., data = students)

```

```

step(null, scope=list(lower=null, upper=full),direction="forward")

```

```

## Start: AIC=1203.39

```

```

## G3 ~ 1

```

```

##
##      Df Sum of Sq  RSS   AIC
## + G2    1   6771.3 1498.6  530.70
## + G1    1   5312.2 2957.7  799.25
## + failures 1   1074.3 7195.7 1150.43
## + Medu    1    390.0 7880.0 1186.31
## + higher  1    275.3 7994.6 1192.01
## + age     1    215.9 8054.0 1194.94
## + Fedu    1    192.2 8077.7 1196.10
## + Mjob    4     306.6 7963.3 1196.46
## + goout   1     145.8 8124.1 1198.36

```

```

## + romantic      1      139.7 8130.2 1198.66
## + traveltime    1      113.5 8156.4 1199.93
## + address       1       92.5 8177.4 1200.95
## + sex           1       88.5 8181.4 1201.14
## + paid          1       86.0 8183.9 1201.26
## + internet      1       80.2 8189.7 1201.54
## + studytime     1       79.1 8190.8 1201.59
## + schoolsup     1       56.7 8213.2 1202.67
## + famsize       1       54.8 8215.1 1202.76
## + reason        3      129.9 8140.0 1203.14
## <none>          8269.9 1203.39
## + health        1       31.1 8238.8 1203.90
## + Pstatus       1       27.8 8242.1 1204.06
## + guardian      2       66.6 8203.3 1204.20
## + Dalc          1       24.7 8245.2 1204.21
## + Walc          1       22.3 8247.6 1204.32
## + nursery       1       22.0 8247.9 1204.34
## + famrel        1       21.8 8248.1 1204.35
## + school        1       16.8 8253.1 1204.59
## + famsup        1       12.7 8257.2 1204.78
## + absences      1        9.7 8260.2 1204.93
## + activities    1        2.1 8267.8 1205.29
## + freetime      1        1.1 8268.9 1205.34
## + Fjob          4      109.1 8160.9 1206.15
##
## Step:  AIC=530.7
## G3 ~ G2
##
##           Df Sum of Sq   RSS   AIC
## + famrel    1    38.147 1460.5 522.51
## + absences  1    32.858 1465.8 523.94
## + G1        1    27.929 1470.7 525.27
## + failures  1    13.946 1484.7 529.01
## + age       1     8.514 1490.1 530.45
## <none>      1498.6 530.70
## + activities 1     7.285 1491.3 530.77
## + romantic   1     6.960 1491.7 530.86
## + health     1     6.127 1492.5 531.08
## + studytime  1     5.322 1493.3 531.29
## + Walc       1     5.168 1493.5 531.33
## + freetime   1     4.675 1494.0 531.47
## + schoolsup  1     4.603 1494.0 531.48
## + Medu       1     4.245 1494.4 531.58
## + traveltime 1     3.908 1494.7 531.67
## + sex        1     3.685 1494.9 531.73
## + higher     1     3.547 1495.1 531.76
## + Pstatus    1     3.503 1495.1 531.77
## + guardian   2    10.662 1488.0 531.88
## + famsup     1     1.716 1496.9 532.25
## + goout      1     1.670 1497.0 532.26

```

```
## + nursery      1      0.847 1497.8 532.48
## + internet     1      0.772 1497.8 532.50
## + address      1      0.578 1498.0 532.55
## + famsize      1      0.521 1498.1 532.56
## + paid         1      0.387 1498.2 532.60
## + Dalc         1      0.094 1498.5 532.67
## + Fedu         1      0.090 1498.5 532.68
## + school       1      0.001 1498.6 532.70
## + Fjob         4     17.626 1481.0 534.03
## + reason       3      7.038 1491.6 534.84
## + Mjob         4      1.679 1496.9 538.26
##
```

```
## Step: AIC=522.51
```

```
## G3 ~ G2 + famrel
```

```
##
```

	Df	Sum of Sq	RSS	AIC
## + absences	1	36.192	1424.3	514.60
## + G1	1	23.541	1436.9	518.10
## + failures	1	11.583	1448.9	521.37
## + age	1	10.514	1450.0	521.66
## + Walc	1	9.037	1451.4	522.06
## + activities	1	8.756	1451.7	522.14
## <none>			1460.5	522.51
## + studytime	1	6.618	1453.9	522.72
## + romantic	1	4.990	1455.5	523.16
## + schoolsup	1	4.697	1455.8	523.24
## + traveltime	1	4.409	1456.1	523.32
## + Medu	1	4.244	1456.2	523.36
## + guardian	2	11.533	1448.9	523.38
## + Pstatus	1	4.093	1456.4	523.41
## + health	1	3.651	1456.8	523.53
## + higher	1	2.927	1457.5	523.72
## + sex	1	2.392	1458.1	523.87
## + famsup	1	2.084	1458.4	523.95
## + freetime	1	1.555	1458.9	524.09
## + internet	1	1.203	1459.3	524.19
## + goout	1	0.825	1459.7	524.29
## + nursery	1	0.820	1459.7	524.29
## + address	1	0.745	1459.7	524.31
## + famsize	1	0.729	1459.8	524.32
## + Dalc	1	0.634	1459.8	524.34
## + paid	1	0.369	1460.1	524.41
## + school	1	0.109	1460.4	524.49
## + Fedu	1	0.084	1460.4	524.49
## + Fjob	4	20.027	1440.5	525.06
## + reason	3	7.056	1453.4	526.60
## + Mjob	4	1.126	1459.3	530.21

```
##
```

```
## Step: AIC=514.6
```

```
## G3 ~ G2 + famrel + absences
```

```
##
##          Df Sum of Sq    RSS    AIC
## + G1      1   23.7914 1400.5 509.95
## + age      1   19.0504 1405.2 511.28
## + failures  1   13.9314 1410.3 512.72
## + romantic  1    9.9930 1414.3 513.82
## + activities 1    8.3988 1415.9 514.27
## <none>                1424.3 514.60
## + guardian  2   13.4693 1410.8 514.85
## + Walc      1    5.0326 1419.2 515.20
## + studytime  1    4.9837 1419.3 515.22
## + traveltime 1    4.9030 1419.4 515.24
## + health     1    4.3572 1419.9 515.39
## + schoolsup   1    4.2210 1420.1 515.43
## + higher     1    4.0584 1420.2 515.48
## + sex        1    3.6941 1420.6 515.58
## + internet    1    3.0879 1421.2 515.75
## + freetime    1    2.4476 1421.8 515.92
## + Medu       1    1.9815 1422.3 516.05
## + famsup     1    1.7265 1422.6 516.12
## + Pstatus    1    1.4865 1422.8 516.19
## + nursery    1    1.0693 1423.2 516.31
## + school     1    0.7944 1423.5 516.38
## + address    1    0.5228 1423.8 516.46
## + goout      1    0.4260 1423.9 516.48
## + famsize    1    0.3947 1423.9 516.49
## + paid       1    0.2934 1424.0 516.52
## + Dalc       1    0.0235 1424.3 516.60
## + Fedu       1    0.0115 1424.3 516.60
## + Fjob       4   20.7772 1403.5 516.80
## + reason     3    9.3340 1415.0 518.01
## + Mjob       4    0.7815 1423.5 522.39
##
```

## Step: AIC=509.95

## G3 ~ G2 + famrel + absences + G1

```
##
##          Df Sum of Sq    RSS    AIC
## + age      1   24.3963 1376.1 505.01
## + romantic  1   14.1977 1386.3 507.92
## + failures  1   10.5866 1389.9 508.95
## + schoolsup  1   10.1652 1390.3 509.07
## + activities 1    9.0870 1391.4 509.38
## + guardian  2   14.7640 1385.7 509.76
## + Walc      1    7.4216 1393.1 509.85
## <none>                1400.5 509.95
## + studytime  1    7.0187 1393.5 509.96
## + health     1    4.1037 1396.4 510.79
## + traveltime 1    3.4521 1397.0 510.97
## + sex        1    3.2814 1397.2 511.02
## + higher     1    3.1817 1397.3 511.05
```

```
## + famsup      1      2.6539 1397.8 511.20
## + internet    1      2.1433 1398.3 511.34
## + freetime    1      1.9314 1398.6 511.40
## + Pstatus     1      1.9102 1398.6 511.41
## + Medu        1      1.4345 1399.1 511.54
## + nursery     1      1.3007 1399.2 511.58
## + paid        1      1.0450 1399.5 511.65
## + goout       1      0.6040 1399.9 511.78
## + school      1      0.5139 1400.0 511.80
## + famsize     1      0.3588 1400.1 511.85
## + Dalc        1      0.2474 1400.2 511.88
## + Fedu        1      0.1339 1400.4 511.91
## + address     1      0.1314 1400.4 511.91
## + Fjob        4     18.3616 1382.1 512.74
## + reason      3      9.1130 1391.4 513.37
## + Mjob        4      0.8742 1399.6 517.70
```

```
##
```

```
## Step: AIC=505.01
```

```
## G3 ~ G2 + famrel + absences + G1 + age
```

```
##
```

	Df	Sum of Sq	RSS	AIC
## + activities	1	12.5228	1363.6	503.40
## + Walc	1	10.6246	1365.5	503.95
## + romantic	1	10.2015	1365.9	504.07
## + school	1	8.6788	1367.4	504.51
## <none>			1376.1	505.01
## + studytime	1	6.6399	1369.5	505.10
## + failures	1	4.9442	1371.2	505.59
## + traveltime	1	4.3828	1371.7	505.75
## + internet	1	4.1699	1371.9	505.81
## + schoolsup	1	3.7993	1372.3	505.92
## + sex	1	3.0896	1373.0	506.12
## + health	1	2.6649	1373.4	506.24
## + nursery	1	2.4657	1373.6	506.30
## + freetime	1	2.1061	1374.0	506.40
## + goout	1	1.6437	1374.5	506.54
## + Fedu	1	1.4182	1374.7	506.60
## + Pstatus	1	1.3569	1374.7	506.62
## + Dalc	1	1.2126	1374.9	506.66
## + address	1	0.9541	1375.1	506.73
## + paid	1	0.9191	1375.2	506.74
## + famsup	1	0.8036	1375.3	506.78
## + higher	1	0.7315	1375.4	506.80
## + famsize	1	0.6769	1375.4	506.81
## + Medu	1	0.1458	1376.0	506.97
## + guardian	2	6.4951	1369.6	507.14
## + reason	3	9.5860	1366.5	508.25
## + Fjob	4	14.7272	1361.4	508.76
## + Mjob	4	0.3768	1375.7	512.90

```
##
```

```

## Step: AIC=503.4
## G3 ~ G2 + famrel + absences + G1 + age + activities
##
##           Df Sum of Sq    RSS    AIC
## + Walc      1  10.2847 1353.3 502.41
## + romantic  1   9.3225 1354.2 502.69
## + school    1   7.1107 1356.5 503.33
## <none>                1363.6 503.40
## + failures  1   5.4096 1358.2 503.83
## + studytime 1   5.2626 1358.3 503.87
## + traveltime 1   4.4381 1359.1 504.11
## + sex       1   4.3756 1359.2 504.13
## + schoolsup  1   4.2818 1359.3 504.15
## + internet  1   3.7086 1359.9 504.32
## + freetime  1   3.1033 1360.5 504.50
## + health    1   2.8611 1360.7 504.57
## + nursery   1   2.5760 1361.0 504.65
## + goout     1   2.2934 1361.3 504.73
## + address   1   1.5175 1362.1 504.96
## + higher    1   1.2111 1362.4 505.05
## + Dalc      1   0.8797 1362.7 505.14
## + Fedu      1   0.7709 1362.8 505.17
## + paid      1   0.7506 1362.8 505.18
## + famsup    1   0.7327 1362.8 505.18
## + famsize   1   0.6858 1362.9 505.20
## + Pstatus   1   0.6489 1362.9 505.21
## + Medu      1   0.4713 1363.1 505.26
## + guardian  2   6.6611 1356.9 505.46
## + reason    3  10.4696 1353.1 506.35
## + Fjob      4  12.8252 1350.8 507.66
## + Mjob      4   0.5571 1363.0 511.24
##
## Step: AIC=502.41
## G3 ~ G2 + famrel + absences + G1 + age + activities + Walc
##
##           Df Sum of Sq    RSS    AIC
## + romantic  1   8.3953 1344.9 501.95
## <none>                1353.3 502.41
## + failures  1   6.6930 1346.6 502.45
## + school    1   6.5721 1346.7 502.48
## + schoolsup  1   5.8202 1347.5 502.70
## + internet  1   3.9534 1349.3 503.25
## + traveltime 1   2.9130 1350.4 503.55
## + studytime 1   2.5246 1350.8 503.67
## + Dalc      1   1.9632 1351.3 503.83
## + health    1   1.7913 1351.5 503.88
## + nursery   1   1.7535 1351.5 503.89
## + higher    1   1.6330 1351.7 503.93
## + freetime  1   1.4335 1351.9 503.99
## + famsup    1   1.3350 1352.0 504.02

```



```

## + sex          1      1.3044 1352.0 504.03
## + address      1      0.9254 1352.4 504.14
## + Fedu         1      0.9253 1352.4 504.14
## + Pstatus      1      0.7789 1352.5 504.18
## + Medu         1      0.5767 1352.7 504.24
## + paid         1      0.4453 1352.8 504.28
## + famsize      1      0.2520 1353.0 504.33
## + goout        1      0.0364 1353.2 504.40
## + guardian     2      6.0500 1347.2 504.64
## + reason       3      9.7896 1343.5 505.54
## + Fjob         4     14.6237 1338.7 506.11
## + Mjob         4      0.6410 1352.7 510.22
##
## Step: AIC=501.95
## G3 ~ G2 + famrel + absences + G1 + age + activities + Walc +
##      romantic
##
##              Df Sum of Sq    RSS    AIC
## + school      1      6.8978 1338.0 501.92
## <none>                1344.9 501.95
## + failures    1      6.1267 1338.8 502.14
## + schoolsup    1      5.1200 1339.8 502.44
## + traveltime  1      2.8946 1342.0 503.10
## + internet    1      2.7909 1342.1 503.13
## + health      1      2.1235 1342.8 503.32
## + studytime   1      2.0196 1342.9 503.35
## + Dalc        1      1.8262 1343.1 503.41
## + famsup      1      1.4993 1343.4 503.51
## + nursery     1      1.4816 1343.4 503.51
## + freetime    1      1.4673 1343.4 503.52
## + higher      1      1.1595 1343.7 503.61
## + Pstatus     1      0.9829 1343.9 503.66
## + Medu        1      0.8907 1344.0 503.69
## + sex         1      0.8542 1344.0 503.70
## + Fedu        1      0.7263 1344.2 503.73
## + address     1      0.6772 1344.2 503.75
## + paid        1      0.5884 1344.3 503.78
## + famsize     1      0.3720 1344.5 503.84
## + goout       1      0.0313 1344.9 503.94
## + guardian    2      5.6138 1339.3 504.30
## + reason      3     10.1369 1334.8 504.96
## + Fjob        4     14.0043 1330.9 505.81
## + Mjob        4      0.5643 1344.3 509.78
##
## Step: AIC=501.92
## G3 ~ G2 + famrel + absences + G1 + age + activities + Walc +
##      romantic + school
##
##              Df Sum of Sq    RSS    AIC
## <none>                1338.0 501.92

```

```

## + failures      1      5.5442 1332.5 502.28
## + schoolsup     1      5.5306 1332.5 502.28
## + Dalc          1      2.3951 1335.6 503.21
## + famsup        1      2.3777 1335.6 503.21
## + health        1      2.2750 1335.7 503.24
## + internet      1      2.1553 1335.8 503.28
## + studytime     1      1.3703 1336.6 503.51
## + freetime      1      1.2613 1336.7 503.54
## + traveltime    1      1.2457 1336.8 503.55
## + Medu          1      1.1679 1336.8 503.57
## + nursery       1      1.1527 1336.8 503.58
## + Pstatus       1      1.1134 1336.9 503.59
## + sex           1      0.9004 1337.1 503.65
## + higher        1      0.7921 1337.2 503.68
## + Fedu          1      0.7102 1337.3 503.71
## + paid          1      0.6183 1337.4 503.73
## + famsize       1      0.2163 1337.8 503.85
## + goout         1      0.1263 1337.9 503.88
## + address       1      0.0228 1338.0 503.91
## + guardian      2      5.6634 1332.3 504.24
## + Fjob          4     16.5231 1321.5 505.01
## + reason        3      9.3473 1328.7 505.15
## + Mjob          4      0.5407 1337.5 509.76

##
## Call:
## lm(formula = G3 ~ G2 + famrel + absences + G1 + age + activities +
##      Walc + romantic + school, data = students)
##
## Coefficients:
##      (Intercept)          G2          famrel          absences          G1
##      0.65012      0.96153      0.38361      0.04714      0.18058
##      age  activitiesyes          Walc  romanticyes  schoolMS
##     -0.26017     -0.32049      0.11979     -0.32668      0.45495

reg=lm(formula = G3 ~ G2 + famrel + absences + G1 + age + activities + Walc +
romantic + school, data = students)
summary(reg)

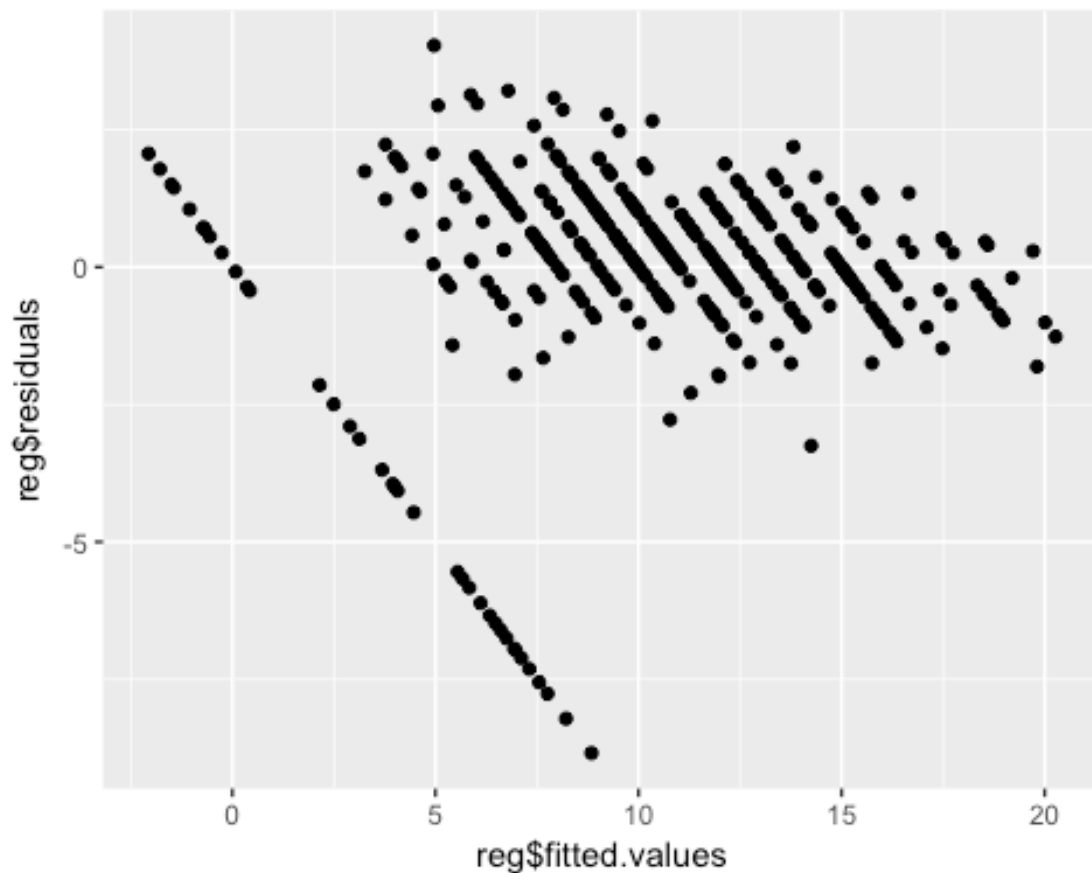
##
## Call:
## lm(formula = G3 ~ G2 + famrel + absences + G1 + age + activities +
##      Walc + romantic + school, data = students)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.8416 -0.4534  0.2645  1.0247  4.0315
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)

```

```
## (Intercept)    0.65012    1.45806    0.446 0.655933
## G2             0.96153    0.04904   19.607 < 2e-16 ***
## famrel        0.38361    0.10668    3.596 0.000365 ***
## absences      0.04714    0.01232    3.826 0.000152 ***
## G1            0.18058    0.05517    3.273 0.001159 **
## age          -0.26017    0.08405   -3.095 0.002109 **
## activitiesyes -0.32049    0.18989   -1.688 0.092278 .
## Walc          0.11979    0.07521    1.593 0.112050
## romanticyes   -0.32668    0.20622   -1.584 0.113989
## schoolMS      0.45495    0.32293    1.409 0.159693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.864 on 385 degrees of freedom
## Multiple R-squared:  0.8382, Adjusted R-squared:  0.8344
## F-statistic: 221.6 on 9 and 385 DF,  p-value: < 2.2e-16

students2=cbind(students,reg$residuals,reg$fitted.values)

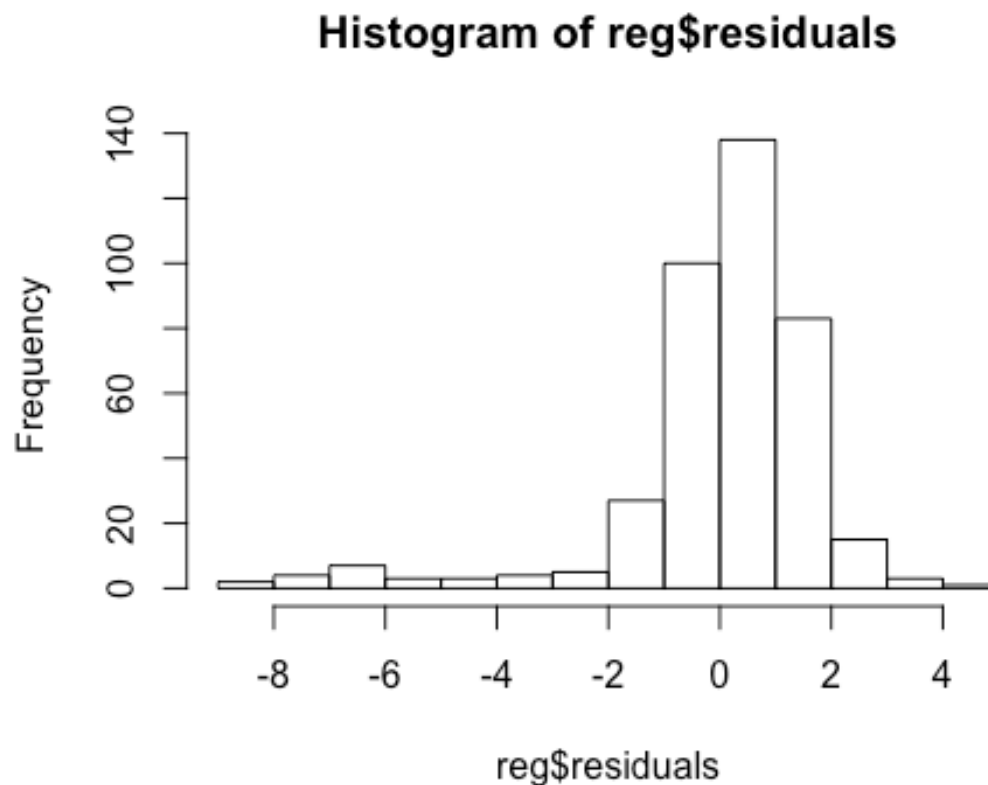
students2%>%
  ggplot(aes(x=reg$fitted.values, y=reg$residuals))+
  geom_point()
```



```
shapiro.test(reg$residuals) #test if normal distribution (if under 100 data points) You want a high p value, you want to fail to reject the null
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: reg$residuals  
## W = 0.79675, p-value < 2.2e-16
```

```
hist(reg$residuals)
```

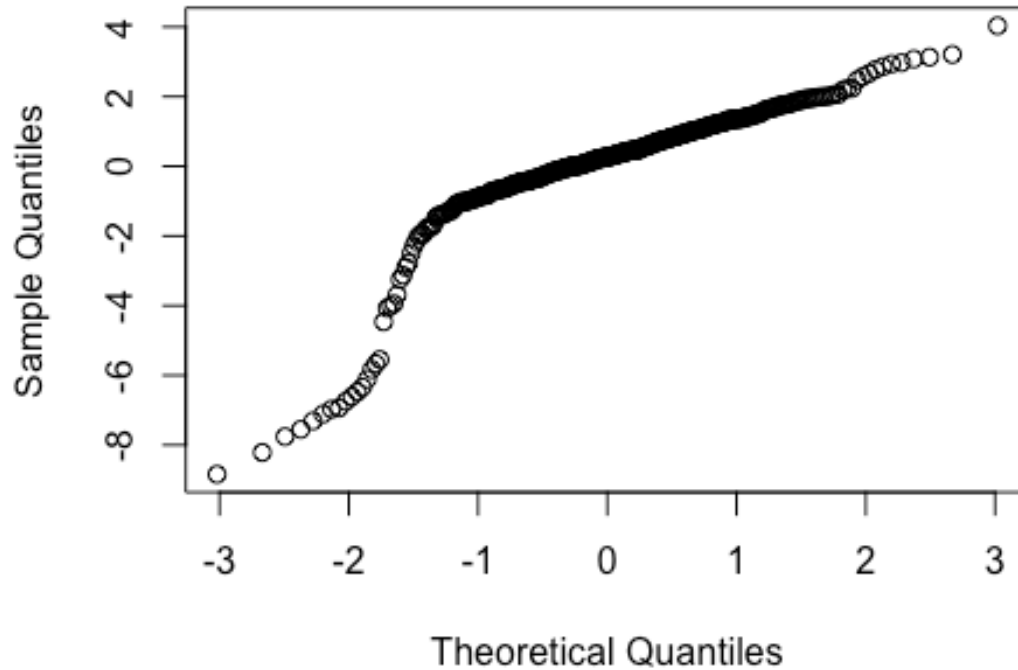


```
ncvTest(reg)
```

```
## Non-constant Variance Score Test  
## Variance formula: ~ fitted.values  
## Chisquare = 91.92734    Df = 1    p = 8.991798e-22
```

```
qqnorm(reg$residuals)
```

## Normal Q-Q Plot



```
reg2=lm(formula = G3 ~ G2 + famrel + absences + G1 + age, data=students)
summary(reg2)
```

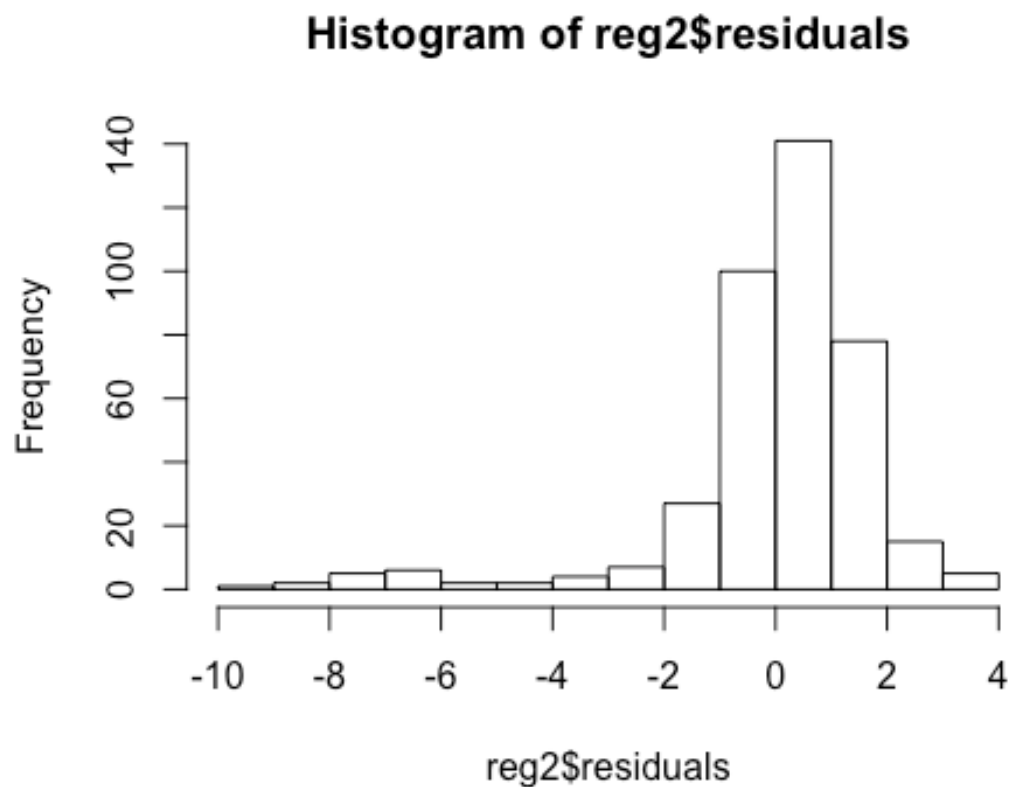
```
##
## Call:
## lm(formula = G3 ~ G2 + famrel + absences + G1 + age, data = students)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.0811 -0.4081  0.2733  0.9927  3.7111
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.07765    1.37626  -0.056  0.955033
## G2           0.97804    0.04895  19.981  < 2e-16 ***
## famrel       0.35725    0.10622   3.363  0.000847 ***
## absences     0.04365    0.01205   3.623  0.000329 ***
## G1           0.15794    0.05503   2.870  0.004329 **
## age         -0.20167    0.07679  -2.626  0.008978 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.881 on 389 degrees of freedom
```

```
## Multiple R-squared:  0.8336, Adjusted R-squared:  0.8315
## F-statistic: 389.8 on 5 and 389 DF,  p-value: < 2.2e-16

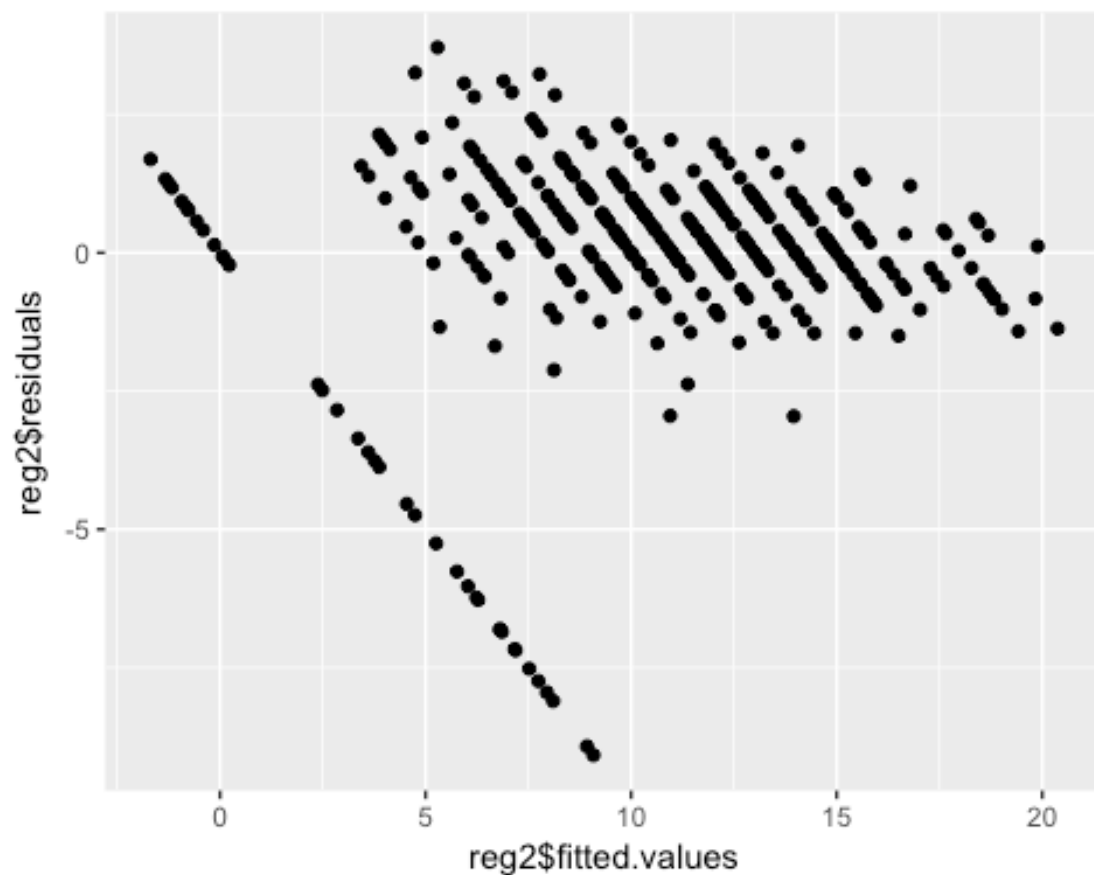
students2=students%>%
  select(G3, G2 , famrel ,absences , G1 , age)

students2=cbind(students2,reg2$residuals,reg2$fitted.values)

hist(reg2$residuals)
```



```
students2%>%
  ggplot(aes(x=reg2$fitted.values, y=reg2$residuals))+
  geom_point()
```



```
shapiro.test(reg2$residuals) #test if normal distribution (if under 100 data
points) You want a high p value, you want to fail to reject the null
```

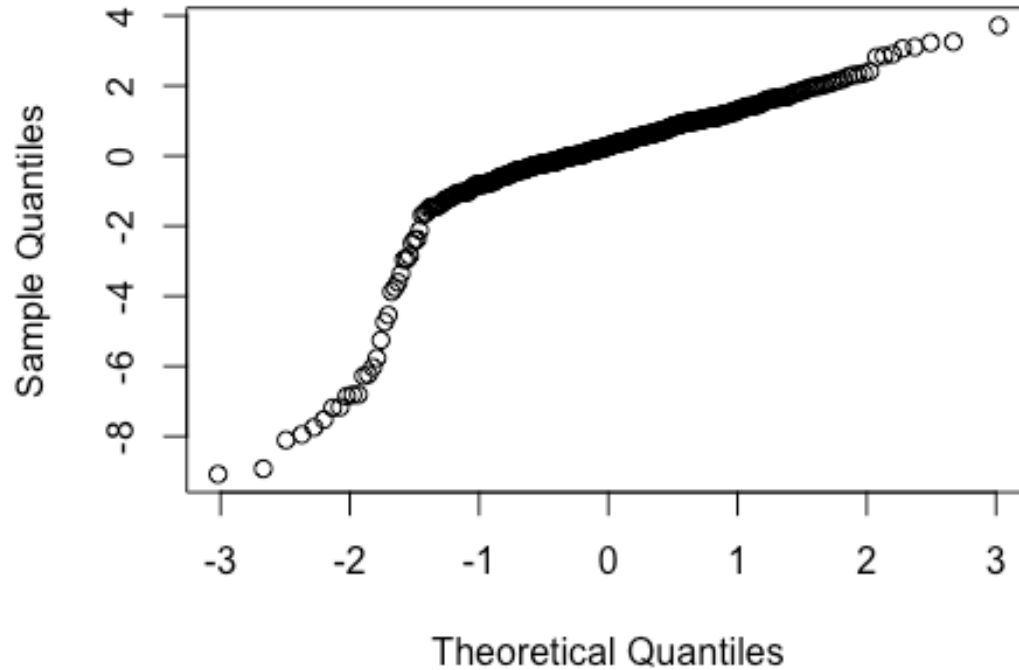
```
##
##  Shapiro-Wilk normality test
##
## data:  reg2$residuals
## W = 0.77152, p-value < 2.2e-16
```

```
ncvTest(reg2)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 87.37829    Df = 1    p = 8.962888e-21
```

```
qqnorm(reg2$residuals)
```

## Normal Q-Q Plot



```
students3=students%>%
  mutate(logfinal=log(G3+1))

students3=students3%>%
  select(school:G2, logfinal)

best.subset=regsubsets(logfinal~.,students3,nvmax =10)
sum=summary(best.subset)
sum$outmat
```

##		schoolMS	sexM	age	addressU	famsizeLE3	PstatusT	Medu	Fedu
## 1	( 1 )	" "	" "	" "	" "	" "	" "	" "	" "
## 2	( 1 )	" "	" "	" "	" "	" "	" "	" "	" "
## 3	( 1 )	" "	" "	" "	" "	" "	" "	" "	" "
## 4	( 1 )	" "	" "	" "	" "	" "	" "	" "	" "
## 5	( 1 )	" "	" "	" "	" "	" "	" "	" "	" "
## 6	( 1 )	" "	" "	" "	" "	" "	" "	" "	" "
## 7	( 1 )	" "	" "	" "	" "	" "	" "	" "	" "
## 8	( 1 )	"*"	" "	" "	" "	" "	" "	" "	" "
## 9	( 1 )	"*"	" "	" "	" "	" "	" "	" "	" "
## 10	( 1 )	"*"	" "	" "	" "	" "	" "	" "	" "
##		Mjobhealth	Mjobother	Mjobservices	Mjobteacher	Fjobhealth			



```

## 1 ( 1 ) " " " " " " " "
## 2 ( 1 ) " " " " " " " "
## 3 ( 1 ) " " " " " " " "
## 4 ( 1 ) " " " " " " " "
## 5 ( 1 ) " " " " " " " "
## 6 ( 1 ) " " " " " " " "
## 7 ( 1 ) " " " " " " " "
## 8 ( 1 ) " " " " " " " "
## 9 ( 1 ) " " " " " " " "
## 10 ( 1 ) " " " " " " " "
##
## Fjobother Fjobservices Fjobteacher reasonhome reasonother
## 1 ( 1 ) " " " " " " " "
## 2 ( 1 ) " " " " " " " "
## 3 ( 1 ) " " " " " " " "
## 4 ( 1 ) " " " " " " " "
## 5 ( 1 ) " " " " " " " "
## 6 ( 1 ) " " " " " " " "
## 7 ( 1 ) " " " " " " " "
## 8 ( 1 ) " " " " " " " "
## 9 ( 1 ) " " " " " " " "
## 10 ( 1 ) " " " " " " " "
##
## reasonreputation guardianmother guardianother traveltime
## 1 ( 1 ) " " " " " " " "
## 2 ( 1 ) " " " " " " " "
## 3 ( 1 ) " " " " " " " "
## 4 ( 1 ) " " " " " " " "
## 5 ( 1 ) " " " " " " " "
## 6 ( 1 ) " " " " " " " "
## 7 ( 1 ) " " " " " " " "
## 8 ( 1 ) " " " " " " " "
## 9 ( 1 ) " " " " " " " "
## 10 ( 1 ) " " " " " " " "
##
## studytime failures schoolsupyes famsupyes paidyes activitiesyes
## 1 ( 1 ) " " " " " " " "
## 2 ( 1 ) " " " " " " " "
## 3 ( 1 ) " " " " "*" " " " " "
## 4 ( 1 ) " " "*" " " " " " " "
## 5 ( 1 ) " " "*" "*" " " " " " "
## 6 ( 1 ) " " "*" "*" " " " " " "
## 7 ( 1 ) " " "*" "*" " " " " " "
## 8 ( 1 ) " " "*" "*" " " " " " "
## 9 ( 1 ) " " "*" "*" " " " " " "
## 10 ( 1 ) " " "*" "*" " " " " " "
##
## nurseryyes higheryes internetyes romanticyes famrel freetime
## 1 ( 1 ) " " " " " " " "
## 2 ( 1 ) " " " " " " " "
## 3 ( 1 ) " " " " " " " "
## 4 ( 1 ) " " " " " " " "
## 5 ( 1 ) " " " " " " " "
## 6 ( 1 ) " " " " " " " "

```

```
## 7 ( 1 ) " " " " " " " " "*" " "
## 8 ( 1 ) " " " " " " " " "*" " "
## 9 ( 1 ) " " " " " " "*" "*" " "
## 10 ( 1 ) " " " " " " "*" "*" " "
##
##          goout Dalc Walc health absences G1 G2
## 1 ( 1 ) " " " " " " " " " " "*"
## 2 ( 1 ) " " " " " " " " "*" " " "*"
## 3 ( 1 ) " " " " " " " " "*" " " "*"
## 4 ( 1 ) " " " " " " " " "*" "*" "*"
## 5 ( 1 ) " " " " "*" " " " "*" " " "*"
## 6 ( 1 ) " " " " "*" " " " "*" "*" "*"
## 7 ( 1 ) " " " " "*" " " " "*" "*" "*"
## 8 ( 1 ) " " " " "*" " " " "*" "*" "*"
## 9 ( 1 ) " " " " "*" " " " "*" "*" "*"
## 10 ( 1 ) " " " " "*" "*" "*" "*" "*"

```

```
null = lm(logfinal~1, data = students3)
full = lm(logfinal~., data = students3)

```

```
step(null, scope=list(lower=null, upper=full),direction="forward")

```

```
## Start: AIC=-195.74

```

```
## logfinal ~ 1

```

```
##
##          Df Sum of Sq    RSS    AIC
## + G2      1  132.908 106.52 -513.65
## + G1      1   80.968 158.46 -356.78
## + failures 1   25.997 213.44 -239.14
## + higher   1    7.655 231.78 -206.58
## + Medu     1    7.202 232.23 -205.81
## + absences 1    5.331 234.10 -202.64
## + paid     1    5.296 234.14 -202.58
## + romantic 1    4.465 234.97 -201.18
## + age      1    4.244 235.19 -200.81
## + Fedu     1    3.042 236.39 -198.79
## + traveltime 1    2.186 237.25 -197.37
## + goout    1    1.835 237.60 -196.78
## + reason   3    4.100 235.33 -196.57
## + famsize  1    1.694 237.74 -196.55
## + sex      1    1.550 237.88 -196.31
## + address  1    1.226 238.21 -195.77
## <none>          239.43 -195.74
## + Mjob      4    4.766 234.67 -195.69
## + internet  1    1.134 238.30 -195.62
## + studytime 1    0.942 238.49 -195.30
## + Pstatus   1    0.818 238.61 -195.10
## + nursery   1    0.537 238.90 -194.63
## + famrel    1    0.514 238.92 -194.59
## + health    1    0.260 239.17 -194.17
## + Walc      1    0.253 239.18 -194.16

```

```

## + guardian      2      1.375 238.06 -194.02
## + freetime      1      0.102 239.33 -193.91
## + famsup        1      0.074 239.36 -193.87
## + schoolsup     1      0.071 239.36 -193.86
## + school        1      0.052 239.38 -193.83
## + activities    1      0.010 239.42 -193.76
## + Dalc          1      0.008 239.42 -193.76
## + Fjob          4      1.721 237.71 -190.59
##
## Step:  AIC=-513.65
## logfinal ~ G2
##
##           Df Sum of Sq      RSS      AIC
## + absences    1      7.1646  99.359 -539.16
## + schoolsup    1      2.6588 103.865 -521.64
## + G1          1      2.4879 104.036 -520.99
## + Walc        1      2.2123 104.311 -519.94
## + paid        1      1.1981 105.326 -516.12
## + failures    1      1.1352 105.388 -515.89
## + famrel      1      0.8613 105.662 -514.86
## + romantic    1      0.6882 105.836 -514.21
## + Dalc        1      0.6873 105.836 -514.21
## <none>                106.524 -513.65
## + higher      1      0.5087 106.015 -513.55
## + activities  1      0.4654 106.058 -513.38
## + health      1      0.3835 106.140 -513.08
## + studytime   1      0.3617 106.162 -513.00
## + goout       1      0.2732 106.251 -512.67
## + freetime    1      0.2280 106.296 -512.50
## + Pstatus     1      0.1832 106.341 -512.33
## + age         1      0.1683 106.355 -512.28
## + famsup      1      0.1680 106.356 -512.28
## + famsize     1      0.1343 106.389 -512.15
## + school      1      0.1231 106.401 -512.11
## + address     1      0.1213 106.402 -512.10
## + internet    1      0.0987 106.425 -512.02
## + traveltime  1      0.0848 106.439 -511.97
## + Medu        1      0.0415 106.482 -511.81
## + sex         1      0.0382 106.486 -511.80
## + Fedu        1      0.0253 106.498 -511.75
## + nursery     1      0.0028 106.521 -511.67
## + guardian    2      0.2438 106.280 -510.56
## + Fjob        4      1.1336 105.390 -509.88
## + reason      3      0.5945 105.929 -509.87
## + Mjob        4      0.3021 106.222 -506.78
##
## Step:  AIC=-539.16
## logfinal ~ G2 + absences
##
##           Df Sum of Sq      RSS      AIC

```

```

## + schoolsup      1    2.49701 96.862 -547.21
## + G1             1    2.42510 96.934 -546.92
## + romantic       1    1.55656 97.803 -543.39
## + failures       1    1.48257 97.877 -543.10
## + Walc           1    1.29641 98.063 -542.34
## + paid           1    1.13550 98.224 -541.70
## + famrel         1    1.10136 98.258 -541.56
## + age            1    0.78382 98.575 -540.29
## + higher         1    0.72504 98.634 -540.05
## + health         1    0.50204 98.857 -539.16
## <none>           99.359 -539.16
## + activities     1    0.42257 98.937 -538.84
## + freetime       1    0.40363 98.956 -538.77
## + internet       1    0.36248 98.997 -538.60
## + school         1    0.35378 99.005 -538.57
## + Dalc           1    0.28886 99.070 -538.31
## + studytime      1    0.19749 99.162 -537.94
## + goout          1    0.17373 99.185 -537.85
## + sex            1    0.13587 99.223 -537.70
## + famsup         1    0.12228 99.237 -537.64
## + traveltime     1    0.11527 99.244 -537.62
## + address        1    0.08065 99.278 -537.48
## + famsize        1    0.06946 99.290 -537.43
## + Fedu           1    0.05756 99.302 -537.39
## + nursery        1    0.01219 99.347 -537.21
## + Medu           1    0.00823 99.351 -537.19
## + Pstatus        1    0.00402 99.355 -537.17
## + guardian       2    0.48075 98.878 -537.07
## + reason         3    0.72245 98.637 -536.04
## + Fjob           4    1.03526 98.324 -535.29
## + Mjob           4    0.18428 99.175 -531.89
##
## Step:  AIC=-547.21
## logfinal ~ G2 + absences + schoolsup
##
##           Df Sum of Sq  RSS    AIC
## + Walc      1   1.70551 95.157 -552.23
## + G1        1   1.54918 95.313 -551.58
## + failures   1   1.31141 95.551 -550.60
## + romantic   1   1.20216 95.660 -550.14
## + paid       1   1.16514 95.697 -549.99
## + famrel     1   1.11023 95.752 -549.76
## + school     1   0.69566 96.166 -548.06
## + health     1   0.61045 96.252 -547.71
## + activities 1   0.53912 96.323 -547.42
## + higher     1   0.53274 96.329 -547.39
## + freetime   1   0.50324 96.359 -547.27
## <none>       96.862 -547.21
## + internet   1   0.36706 96.495 -546.71
## + Dalc       1   0.34530 96.517 -546.62

```

```

## + sex      1  0.33145 96.531 -546.56
## + studytime 1  0.28455 96.578 -546.37
## + goout    1  0.26058 96.602 -546.27
## + age      1  0.21161 96.651 -546.07
## + traveltime 1  0.14668 96.715 -545.81
## + address  1  0.12138 96.741 -545.71
## + Fedu     1  0.10982 96.752 -545.66
## + famsize  1  0.08720 96.775 -545.57
## + nursery  1  0.03849 96.824 -545.37
## + famsup   1  0.03820 96.824 -545.37
## + Medu     1  0.00487 96.857 -545.23
## + Pstatus  1  0.00007 96.862 -545.21
## + guardian 2  0.36235 96.500 -544.69
## + reason   3  0.73170 96.130 -544.21
## + Fjob     4  1.01741 95.845 -543.38
## + Mjob     4  0.23652 96.626 -540.18
##
## Step: AIC=-552.23
## logfinal ~ G2 + absences + schoolsup + Walc
##
##           Df Sum of Sq  RSS    AIC
## + failures  1  1.67998 93.477 -557.26
## + famrel    1  1.45574 93.701 -556.32
## + G1        1  1.17821 93.978 -555.15
## + romantic  1  1.06204 94.095 -554.66
## + paid      1  0.98421 94.172 -554.33
## + higher    1  0.68564 94.471 -553.08
## + school    1  0.57249 94.584 -552.61
## + activities 1  0.48921 94.667 -552.26
## <none>                95.157 -552.23
## + health    1  0.45070 94.706 -552.10
## + internet  1  0.38019 94.776 -551.81
## + age       1  0.29034 94.866 -551.43
## + freetime  1  0.26593 94.891 -551.33
## + Fedu      1  0.11261 95.044 -550.70
## + Dalc      1  0.10575 95.051 -550.67
## + famsup    1  0.09626 95.060 -550.63
## + address   1  0.05647 95.100 -550.46
## + studytime 1  0.05373 95.103 -550.45
## + traveltime 1  0.04935 95.107 -550.43
## + sex       1  0.04406 95.113 -550.41
## + famsize   1  0.02509 95.132 -550.33
## + nursery   1  0.00559 95.151 -550.25
## + goout     1  0.00068 95.156 -550.23
## + Pstatus   1  0.00019 95.156 -550.23
## + Medu      1  0.00008 95.157 -550.23
## + guardian  2  0.27733 94.879 -549.38
## + reason    3  0.62416 94.532 -548.83
## + Fjob      4  0.95900 94.198 -548.23
## + Mjob      4  0.27394 94.883 -545.37

```

```

##
## Step: AIC=-557.26
## logfinal ~ G2 + absences + schoolsup + Walc + failures
##
##           Df Sum of Sq    RSS    AIC
## + G1       1   1.51007 91.967 -561.70
## + famrel    1   1.33402 92.143 -560.94
## + romantic  1   0.93185 92.545 -559.22
## + school    1   0.64980 92.827 -558.02
## + paid      1   0.60871 92.868 -557.84
## + activities 1   0.58446 92.892 -557.74
## + health    1   0.49384 92.983 -557.36
## <none>                93.477 -557.26
## + internet  1   0.42960 93.047 -557.08
## + Fedu      1   0.38574 93.091 -556.90
## + freetime  1   0.38523 93.091 -556.90
## + higher    1   0.27514 93.201 -556.43
## + studytime 1   0.13976 93.337 -555.85
## + age       1   0.08737 93.389 -555.63
## + sex       1   0.07736 93.399 -555.59
## + address   1   0.07235 93.404 -555.57
## + traveltime 1   0.06646 93.410 -555.54
## + Dalc      1   0.06246 93.414 -555.53
## + Medu      1   0.06056 93.416 -555.52
## + famsup    1   0.04839 93.428 -555.47
## + nursery   1   0.02823 93.448 -555.38
## + famsize   1   0.02488 93.452 -555.37
## + Pstatus   1   0.00117 93.475 -555.27
## + goout     1   0.00005 93.477 -555.26
## + reason    3   0.54565 92.931 -553.58
## + guardian  2   0.05523 93.421 -553.50
## + Fjob      4   0.86294 92.614 -552.93
## + Mjob      4   0.11902 93.358 -549.77
##
## Step: AIC=-561.7
## logfinal ~ G2 + absences + schoolsup + Walc + failures + G1
##
##           Df Sum of Sq    RSS    AIC
## + famrel    1   1.50128 90.465 -566.20
## + romantic  1   0.72541 91.241 -562.82
## + school    1   0.67437 91.292 -562.60
## + internet  1   0.53616 91.430 -562.01
## + health    1   0.53563 91.431 -562.00
## + activities 1   0.52854 91.438 -561.97
## + freetime  1   0.49621 91.470 -561.83
## <none>                91.967 -561.70
## + paid      1   0.41509 91.551 -561.48
## + higher    1   0.32210 91.644 -561.08
## + Fedu      1   0.25807 91.709 -560.81
## + address   1   0.13632 91.830 -560.28

```

```

## + traveltime 1 0.13582 91.831 -560.28
## + sex 1 0.11038 91.856 -560.17
## + studytime 1 0.08866 91.878 -560.08
## + Dalc 1 0.05997 91.907 -559.95
## + Medu 1 0.05050 91.916 -559.91
## + age 1 0.03499 91.932 -559.85
## + famsize 1 0.03063 91.936 -559.83
## + nursery 1 0.02236 91.944 -559.79
## + famsup 1 0.01970 91.947 -559.78
## + goout 1 0.00124 91.965 -559.70
## + Pstatus 1 0.00000 91.967 -559.70
## + reason 3 0.55402 91.413 -558.08
## + guardian 2 0.05527 91.911 -557.93
## + Fjob 4 0.70274 91.264 -556.73
## + Mjob 4 0.04848 91.918 -553.91
##
## Step: AIC=-566.2
## logfinal ~ G2 + absences + schoolsup + Walc + failures + G1 +
## famrel
##
## Df Sum of Sq RSS AIC
## + school 1 0.77437 89.691 -567.59
## + internet 1 0.61892 89.846 -566.91
## + activities 1 0.59359 89.872 -566.80
## + romantic 1 0.58799 89.877 -566.77
## <none> 90.465 -566.20
## + paid 1 0.39934 90.066 -565.95
## + health 1 0.37232 90.093 -565.83
## + higher 1 0.31321 90.152 -565.57
## + freetime 1 0.25444 90.211 -565.31
## + Fedu 1 0.24540 90.220 -565.27
## + traveltime 1 0.14587 90.319 -564.84
## + address 1 0.14541 90.320 -564.83
## + studytime 1 0.09367 90.372 -564.61
## + age 1 0.07777 90.388 -564.54
## + Dalc 1 0.05867 90.407 -564.45
## + sex 1 0.04704 90.418 -564.40
## + Medu 1 0.04586 90.419 -564.40
## + famsize 1 0.03457 90.431 -564.35
## + famsup 1 0.03132 90.434 -564.34
## + nursery 1 0.01703 90.448 -564.27
## + goout 1 0.01304 90.452 -564.26
## + Pstatus 1 0.00060 90.465 -564.20
## + reason 3 0.55163 89.914 -562.61
## + guardian 2 0.05551 90.410 -562.44
## + Fjob 4 0.74257 89.723 -561.45
## + Mjob 4 0.05478 90.411 -558.44
##
## Step: AIC=-567.59
## logfinal ~ G2 + absences + schoolsup + Walc + failures + G1 +

```

```

##      famrel + school
##
##              Df Sum of Sq    RSS      AIC
## + romantic    1   0.66519 89.026 -568.53
## + internet    1   0.47224 89.219 -567.68
## + activities  1   0.46394 89.227 -567.64
## <none>                89.691 -567.59
## + health      1   0.44211 89.249 -567.55
## + age         1   0.43802 89.253 -567.53
## + paid        1   0.40880 89.282 -567.40
## + higher      1   0.30843 89.382 -566.95
## + freetime    1   0.24042 89.451 -566.65
## + Fedu        1   0.19790 89.493 -566.47
## + Dalc        1   0.11212 89.579 -566.09
## + famsup      1   0.09755 89.593 -566.02
## + sex         1   0.06872 89.622 -565.90
## + studytime   1   0.06072 89.630 -565.86
## + traveltime  1   0.03428 89.657 -565.74
## + address     1   0.02172 89.669 -565.69
## + famsize     1   0.01673 89.674 -565.67
## + Medu        1   0.01321 89.678 -565.65
## + goout       1   0.00545 89.685 -565.62
## + nursery     1   0.00465 89.686 -565.61
## + Pstatus     1   0.00225 89.689 -565.60
## + guardian    2   0.08339 89.608 -563.96
## + reason      3   0.50837 89.183 -563.84
## + Fjob        4   0.91126 88.780 -563.63
## + Mjob        4   0.03897 89.652 -559.77
##
## Step:  AIC=-568.53
## logfinal ~ G2 + absences + schoolsup + Walc + failures + G1 +
##      famrel + school + romantic
##
##              Df Sum of Sq    RSS      AIC
## + health      1   0.47384 88.552 -568.64
## + paid        1   0.45327 88.572 -568.55
## <none>                89.026 -568.53
## + activities  1   0.42107 88.605 -568.41
## + internet    1   0.36366 88.662 -568.15
## + age         1   0.35362 88.672 -568.11
## + higher      1   0.24863 88.777 -567.64
## + freetime    1   0.24022 88.786 -567.60
## + Fedu        1   0.16848 88.857 -567.28
## + famsup      1   0.11036 88.915 -567.02
## + Dalc        1   0.10216 88.924 -566.99
## + sex         1   0.03707 88.989 -566.70
## + studytime   1   0.03274 88.993 -566.68
## + traveltime  1   0.03080 88.995 -566.67
## + famsize     1   0.02562 89.000 -566.65
## + address     1   0.01137 89.014 -566.58

```



```

## + Pstatus      1    0.00534 89.020 -566.56
## + goout        1    0.00477 89.021 -566.56
## + Medu         1    0.00420 89.022 -566.55
## + nursery      1    0.00132 89.024 -566.54
## + guardian     2    0.06979 88.956 -564.84
## + reason       3    0.51360 88.512 -564.82
## + Fjob         4    0.88013 88.146 -564.46
## + Mjob         4    0.03786 88.988 -560.70
##
## Step:  AIC=-568.64
## logfinal ~ G2 + absences + schoolsup + Walc + failures + G1 +
##      famrel + school + romantic + health
##
##           Df Sum of Sq    RSS    AIC
## + paid      1    0.52640 88.026 -569.00
## <none>                        88.552 -568.64
## + activities 1    0.44245 88.109 -568.62
## + internet   1    0.30112 88.251 -567.99
## + age        1    0.27956 88.272 -567.89
## + higher     1    0.23692 88.315 -567.70
## + freetime   1    0.21421 88.338 -567.60
## + Fedu       1    0.18948 88.362 -567.49
## + Dalc       1    0.11578 88.436 -567.16
## + famsup     1    0.09536 88.457 -567.07
## + traveltime 1    0.03284 88.519 -566.79
## + famsize    1    0.03198 88.520 -566.78
## + studytime  1    0.02132 88.531 -566.74
## + sex        1    0.01248 88.539 -566.70
## + Pstatus    1    0.00650 88.545 -566.67
## + address    1    0.00644 88.545 -566.67
## + Medu       1    0.00232 88.550 -566.65
## + nursery    1    0.00116 88.551 -566.65
## + goout      1    0.00009 88.552 -566.64
## + reason     3    0.55868 87.993 -565.14
## + guardian   2    0.08613 88.466 -565.03
## + Fjob       4    0.78823 87.764 -564.17
## + Mjob       4    0.06299 88.489 -560.92
##
## Step:  AIC=-569
## logfinal ~ G2 + absences + schoolsup + Walc + failures + G1 +
##      famrel + school + romantic + health + paid
##
##           Df Sum of Sq    RSS    AIC
## <none>                        88.026 -569.00
## + internet   1    0.41416 87.611 -568.86
## + activities 1    0.41264 87.613 -568.85
## + age        1    0.28364 87.742 -568.27
## + freetime   1    0.25323 87.772 -568.14
## + Fedu       1    0.22187 87.804 -567.99
## + higher     1    0.14283 87.883 -567.64

```

```

## + Dalc      1  0.14021 87.885 -567.63
## + studytime 1  0.07299 87.953 -567.33
## + sex       1  0.05020 87.975 -567.22
## + traveltime 1 0.04627 87.979 -567.21
## + famsize   1  0.04115 87.984 -567.18
## + Medu      1  0.01775 88.008 -567.08
## + Pstatus   1  0.01465 88.011 -567.06
## + nursery   1  0.01044 88.015 -567.04
## + address   1  0.01023 88.015 -567.04
## + famsup    1  0.00837 88.017 -567.03
## + goout     1  0.00006 88.025 -567.00
## + reason    3  0.59325 87.432 -565.67
## + guardian  2  0.06380 87.962 -565.28
## + Fjob      4  0.70873 87.317 -564.19
## + Mjob      4  0.09135 87.934 -561.41

##
## Call:
## lm(formula = logfinal ~ G2 + absences + schoolsup + Walc + failures +
##      G1 + famrel + school + romantic + health + paid, data = students3)
##
## Coefficients:
## (Intercept)          G2      absences  schoolsupyes          Walc
##      0.13859      0.17858      0.01788      0.22449      0.04778
##      failures          G1      famrel      schoolMS  romanticyes
##     -0.09173     -0.03477      0.06391      0.15651     -0.09561
##      health      paidyes
##      0.02736      0.07575

reg=lm(logfinal ~ G2+ absences + paid + school +
      schoolsup + Walc + failures + internet + age + higher + Medu,
      data = students3)
summary(reg)

##
## Call:
## lm(formula = logfinal ~ G2 + absences + paid + school + schoolsup +
##      Walc + failures + internet + age + higher + Medu, data = students3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.01655 -0.11627  0.07416  0.27095  0.92652
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.888959   0.435998   2.039  0.04215 *
## G2           0.153385   0.007183  21.353 < 2e-16 ***
## absences     0.018003   0.003240   5.556 5.17e-08 ***
## paidyes      0.090053   0.051693   1.742  0.08230 .
## schoolMS     0.156076   0.084589   1.845  0.06579 .

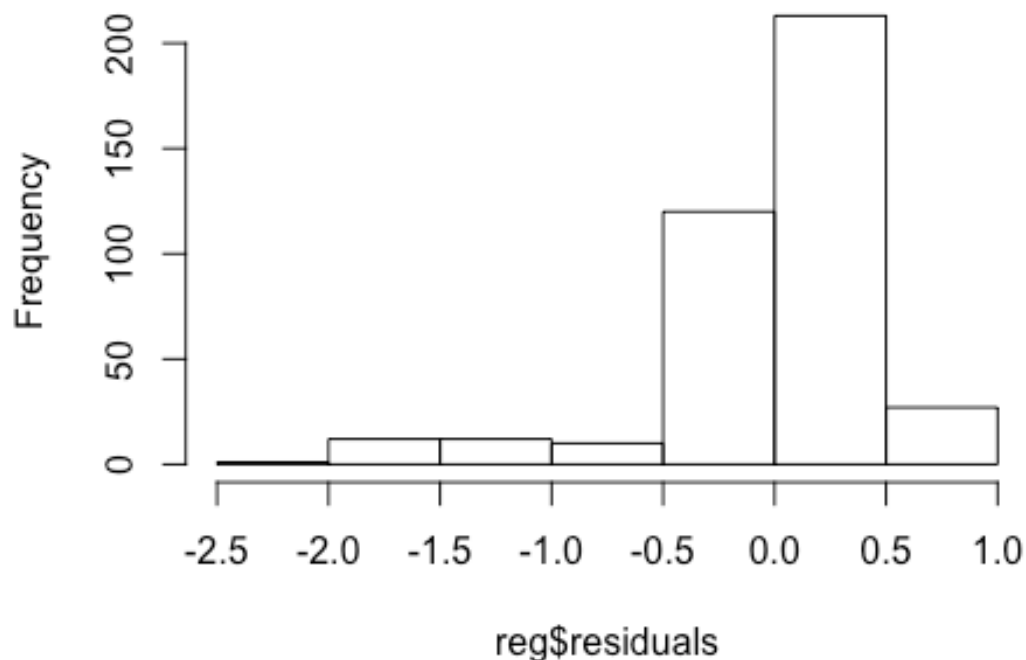
```

```
## schoolsupyes  0.237848  0.077339  3.075  0.00225 **
## Walc         0.054240  0.019693  2.754  0.00616 **
## failures     -0.074596  0.037908  -1.968  0.04981 *
## internetyes  -0.096020  0.068754  -1.397  0.16335
## age          -0.031918  0.023029  -1.386  0.16655
## higheryes    0.078839  0.121097   0.651  0.51541
## Medu         -0.012255  0.024287  -0.505  0.61415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4877 on 383 degrees of freedom
## Multiple R-squared:  0.6195, Adjusted R-squared:  0.6086
## F-statistic:  56.7 on 11 and 383 DF,  p-value: < 2.2e-16

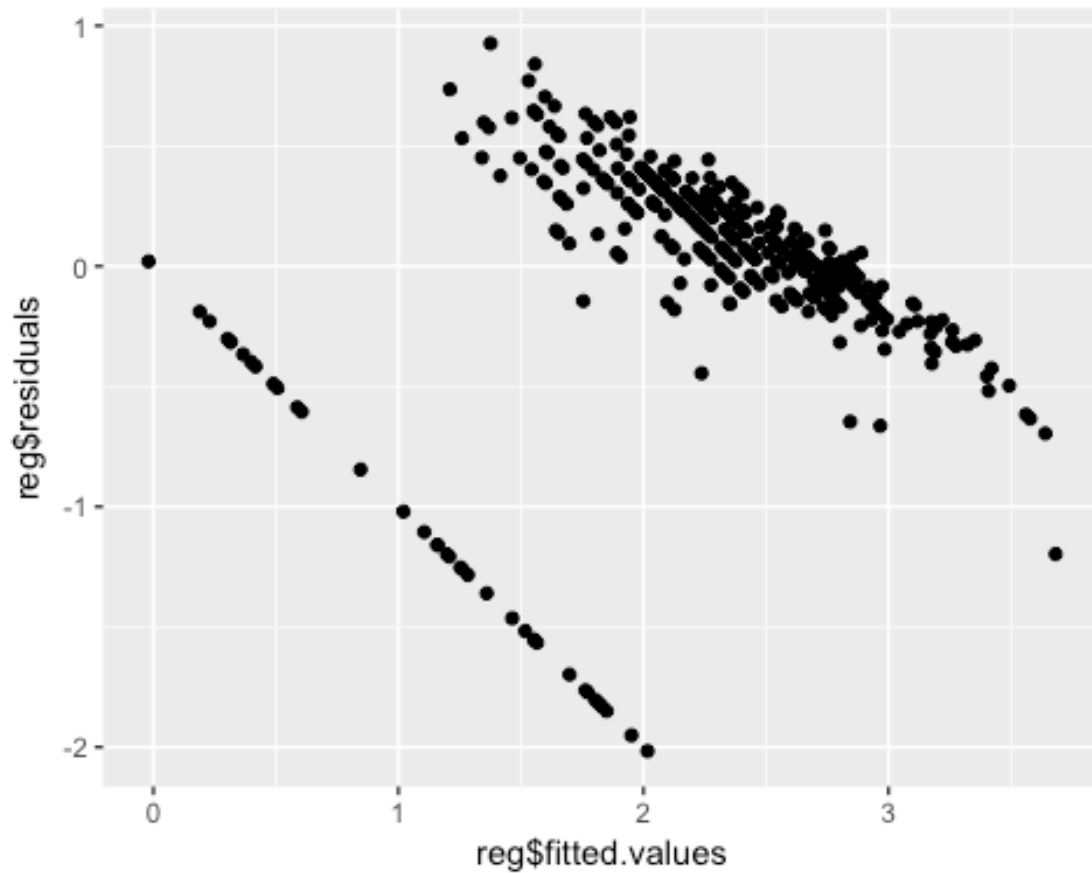
students3=cbind(students3,reg$residuals,reg$fitted.values)

hist(reg$residuals)
```

**Histogram of reg\$residuals**



```
students2%>%
  ggplot(aes(x=reg$fitted.values, y=reg$residuals))+
  geom_point()
```



```
shapiro.test(reg$residuals) #test if normal distribution (if under 100 data points) You want a high p value, you want to fail to reject the null
```

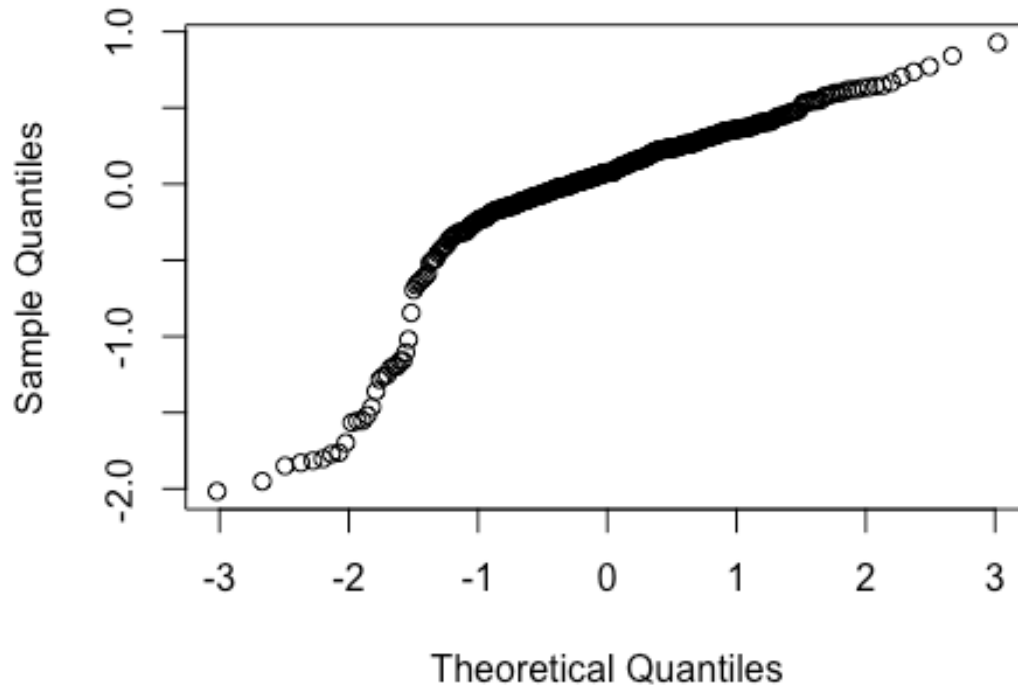
```
##  
##  Shapiro-Wilk normality test  
##  
## data:  reg$residuals  
## W = 0.81526, p-value < 2.2e-16
```

```
ncvTest(reg)
```

```
## Non-constant Variance Score Test  
## Variance formula: ~ fitted.values  
## Chisquare = 118.9213    Df = 1    p = 1.089686e-27
```

```
qqnorm(reg$residuals)
```

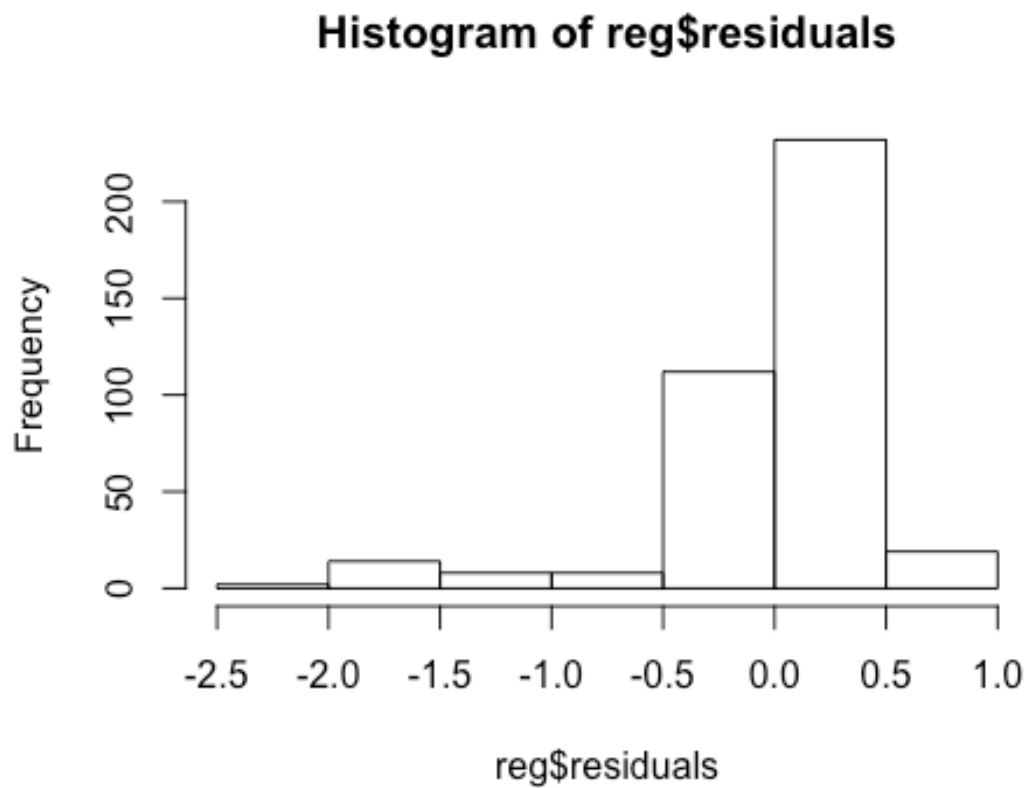
## Normal Q-Q Plot



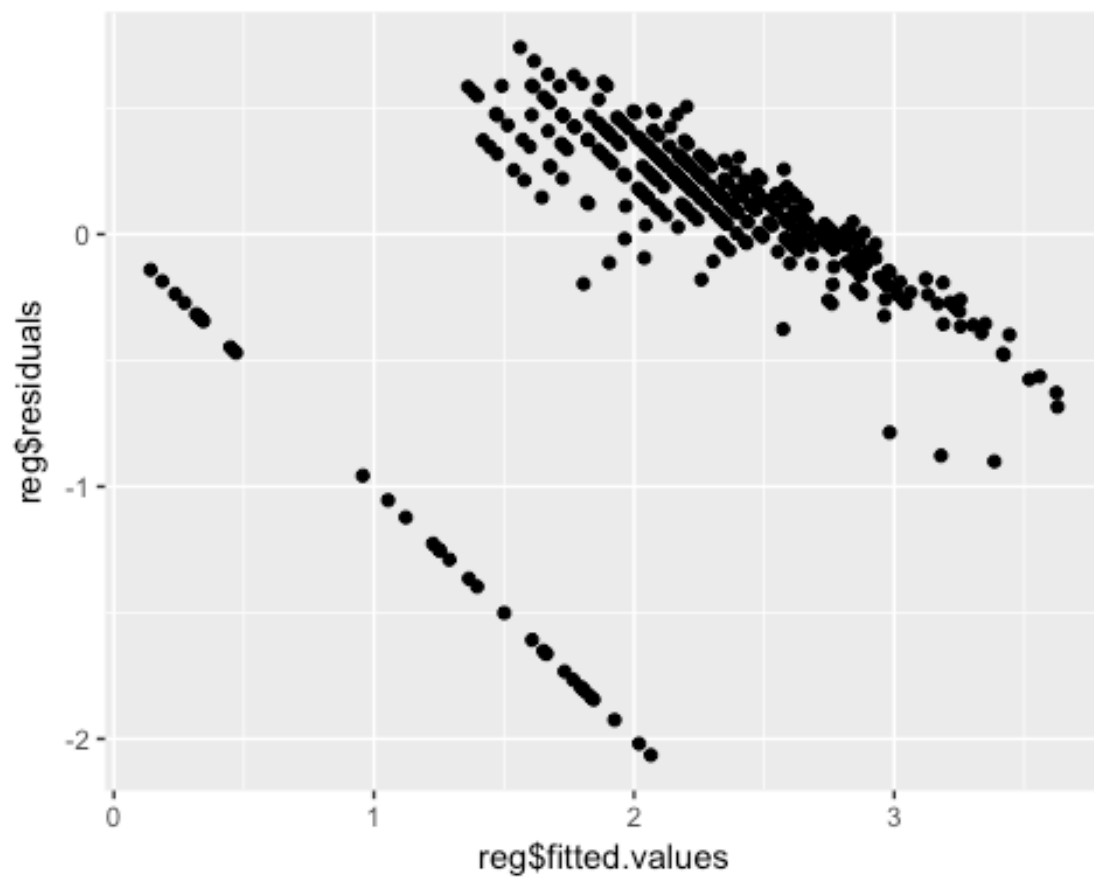
```
reg=lm(formula = logfinal ~ G2 + famrel + absences + G1 + age, data=students3
)
summary(reg)

##
## Call:
## lm(formula = logfinal ~ G2 + famrel + absences + G1 + age, data = students
3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0644 -0.1067  0.1029  0.2864  0.7404
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.883569   0.361572   2.444  0.01498 *
## G2           0.188277   0.012860  14.641 < 2e-16 ***
## famrel       0.067974   0.027907   2.436  0.01531 *
## absences     0.017987   0.003165   5.683  2.6e-08 ***
## G1          -0.045077   0.014458  -3.118  0.00196 **
## age         -0.031561   0.020175  -1.564  0.11855
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##  
## Residual standard error: 0.4941 on 389 degrees of freedom  
## Multiple R-squared:  0.6033, Adjusted R-squared:  0.5982  
## F-statistic: 118.3 on 5 and 389 DF,  p-value: < 2.2e-16  
hist(reg$residuals)
```



```
students3%>%  
  ggplot(aes(x=reg$fitted.values, y=reg$residuals))+  
  geom_point()
```



```
ncvTest(reg)
```

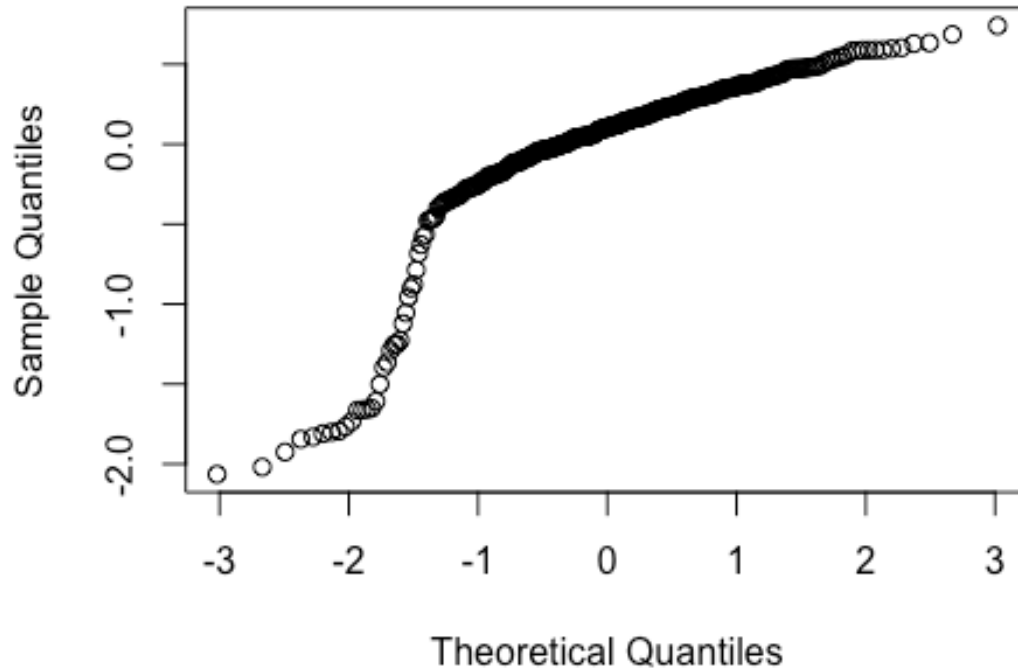
```
## Non-constant Variance Score Test
```

```
## Variance formula: ~ fitted.values
```

```
## Chisquare = 97.68727    Df = 1    p = 4.899617e-23
```

```
qqnorm(reg$residuals)
```

## Normal Q-Q Plot



```
shapiro.test(reg$residuals) #test if normal distribution (if under 100 data points) You want a high p value, you want to fail to reject the null
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  reg$residuals  
## W = 0.77219, p-value < 2.2e-16
```

```
students4=students%>%  
  mutate(G3cube=G3*G3)
```

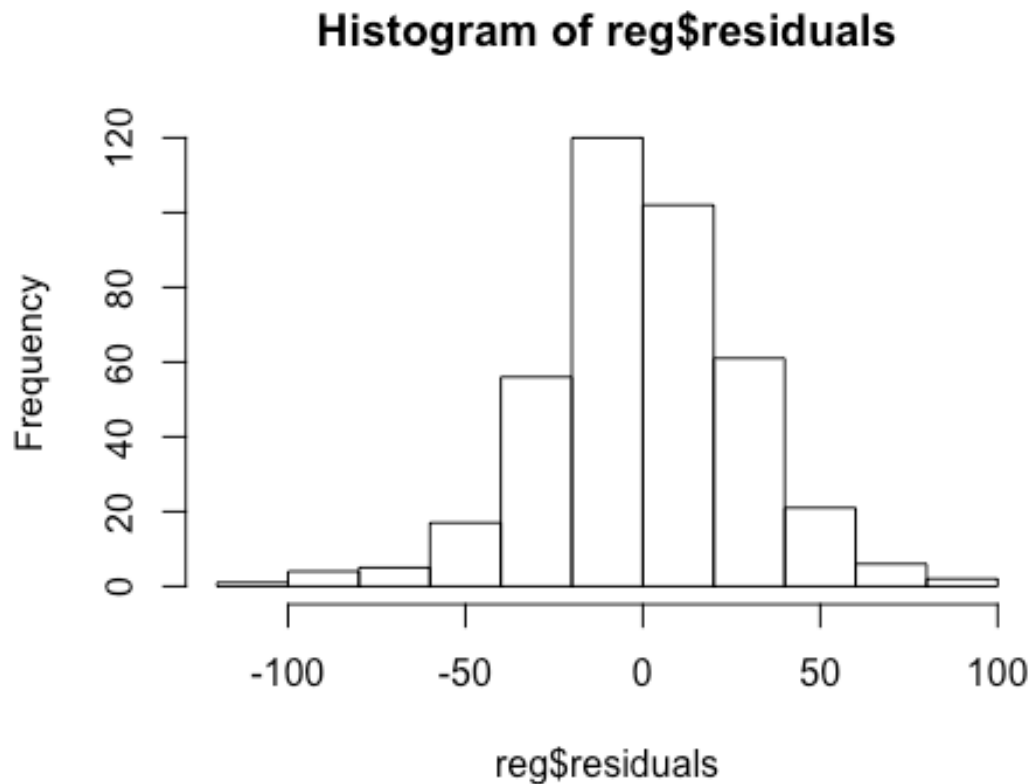
```
reg=lm(formula = G3cube ~ G2 + famrel + absences + G1 + age, data=students4)  
summary(reg)
```

```
##  
## Call:  
## lm(formula = G3cube ~ G2 + famrel + absences + G1 + age, data = students4)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -107.397  -17.689   -1.042   17.460   80.844   
##  
## Coefficients:
```

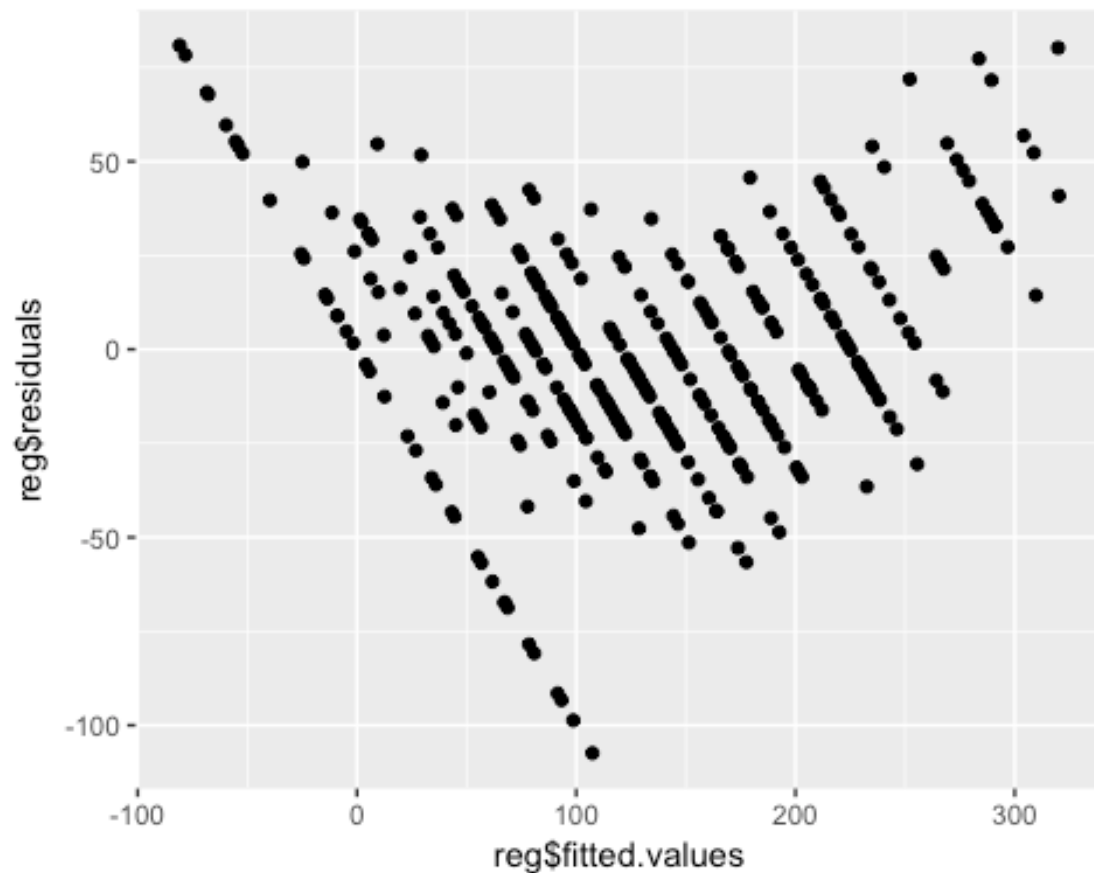


```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -75.2160    21.2868  -3.533 0.000459 ***
## G2           14.0884     0.7571  18.609 < 2e-16 ***
## famrel       5.5300     1.6429   3.366 0.000839 ***
## absences     -0.0974     0.1863  -0.523 0.601492
## G1           8.7203     0.8512  10.245 < 2e-16 ***
## age         -3.7552     1.1878  -3.162 0.001692 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.09 on 389 degrees of freedom
## Multiple R-squared:  0.8845, Adjusted R-squared:  0.883
## F-statistic: 595.8 on 5 and 389 DF,  p-value: < 2.2e-16

hist(reg$residuals)
```



```
students4%>%
  ggplot(aes(x=reg$fitted.values, y=reg$residuals))+
  geom_point()
```



```
shapiro.test(reg$residuals) #test if normal distribution (if under 100 data points) You want a high p value, you want to fail to reject the null
```

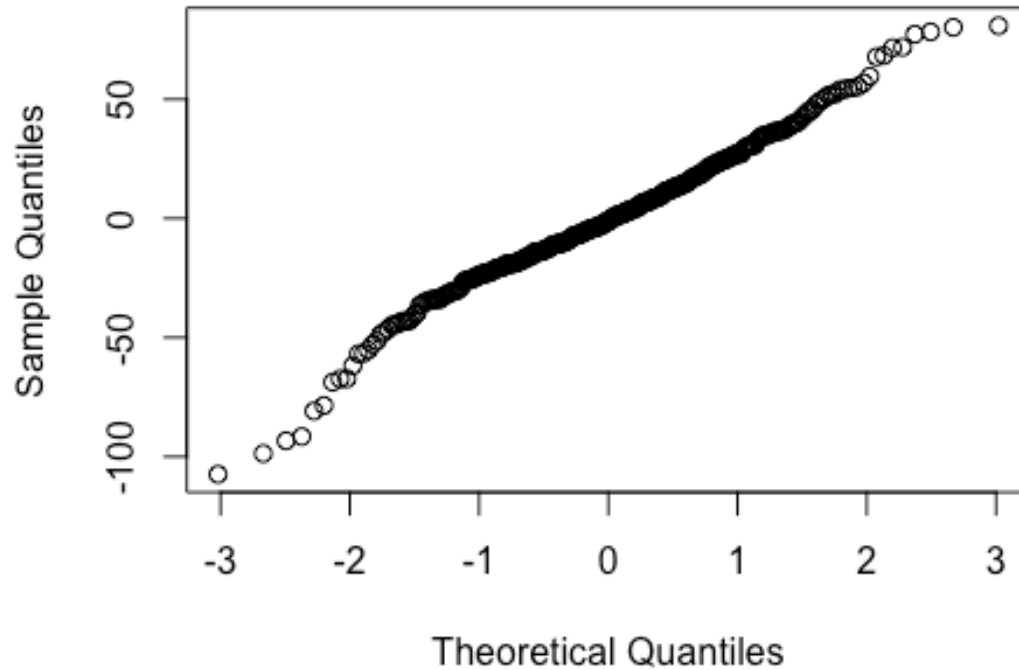
```
##  
##  Shapiro-Wilk normality test  
##  
## data:  reg$residuals  
## W = 0.9854, p-value = 0.0005179
```

```
ncvTest(reg)
```

```
## Non-constant Variance Score Test  
## Variance formula: ~ fitted.values  
## Chisquare = 2.637871    Df = 1    p = 0.1043433
```

```
qqnorm(reg$residuals)
```

## Normal Q-Q Plot



```
mean(students$G3)
## [1] 10.41519
mean(students$G2)
## [1] 10.71392
mean(students$famrel)
## [1] 3.944304
mean(students$absences)
## [1] 5.708861
mean(students$G1)
## [1] 10.90886
mean(students$age)
## [1] 16.6962
-.07765+(.97804*10.71392)+(.35725+3.944304)+(.04365*5.708861)+(.15794*10.90886)+(-.20167*16.6962)
```

```
## [1] 13.30756
```