Sta 478 Project 1

Matthew Flanders

2/6/2022

Introduction

Data and Orginization

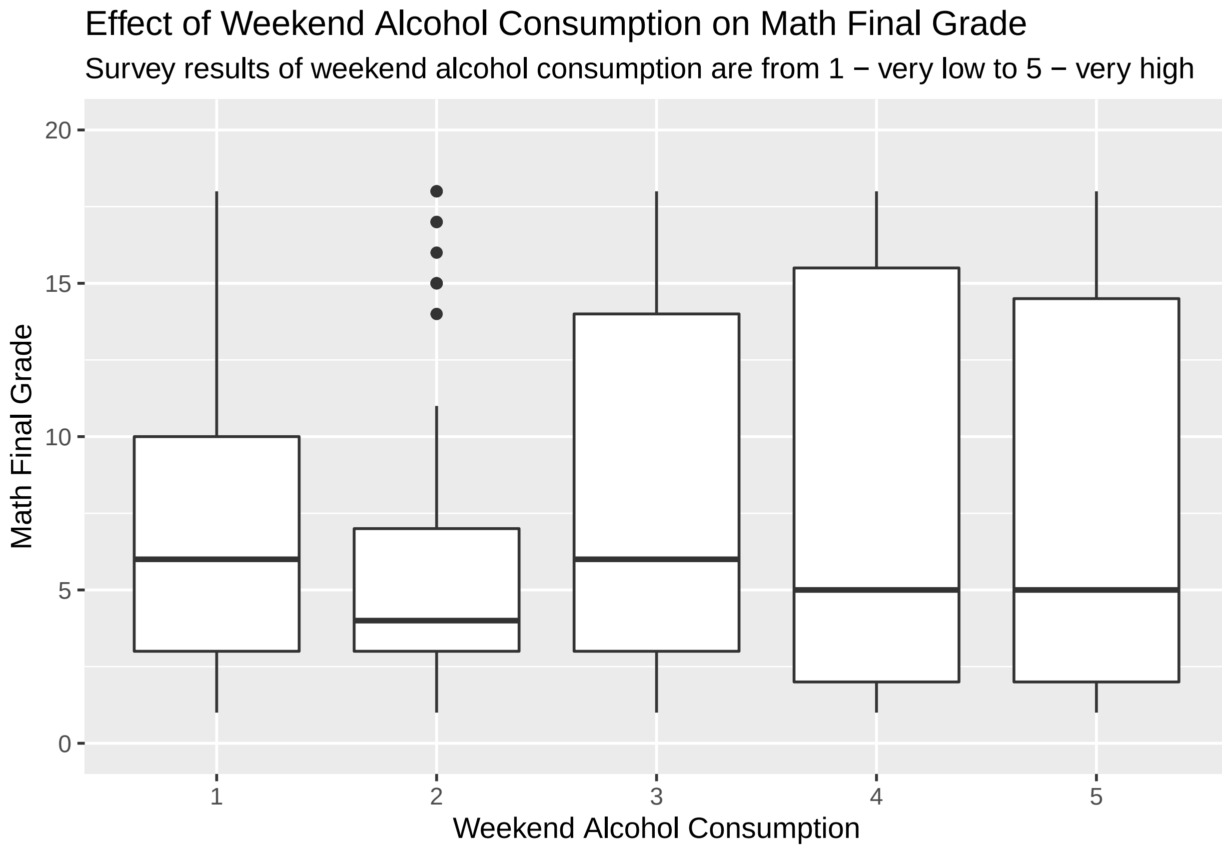
Models and Analysis

Discussions of Major Results

Conclusions

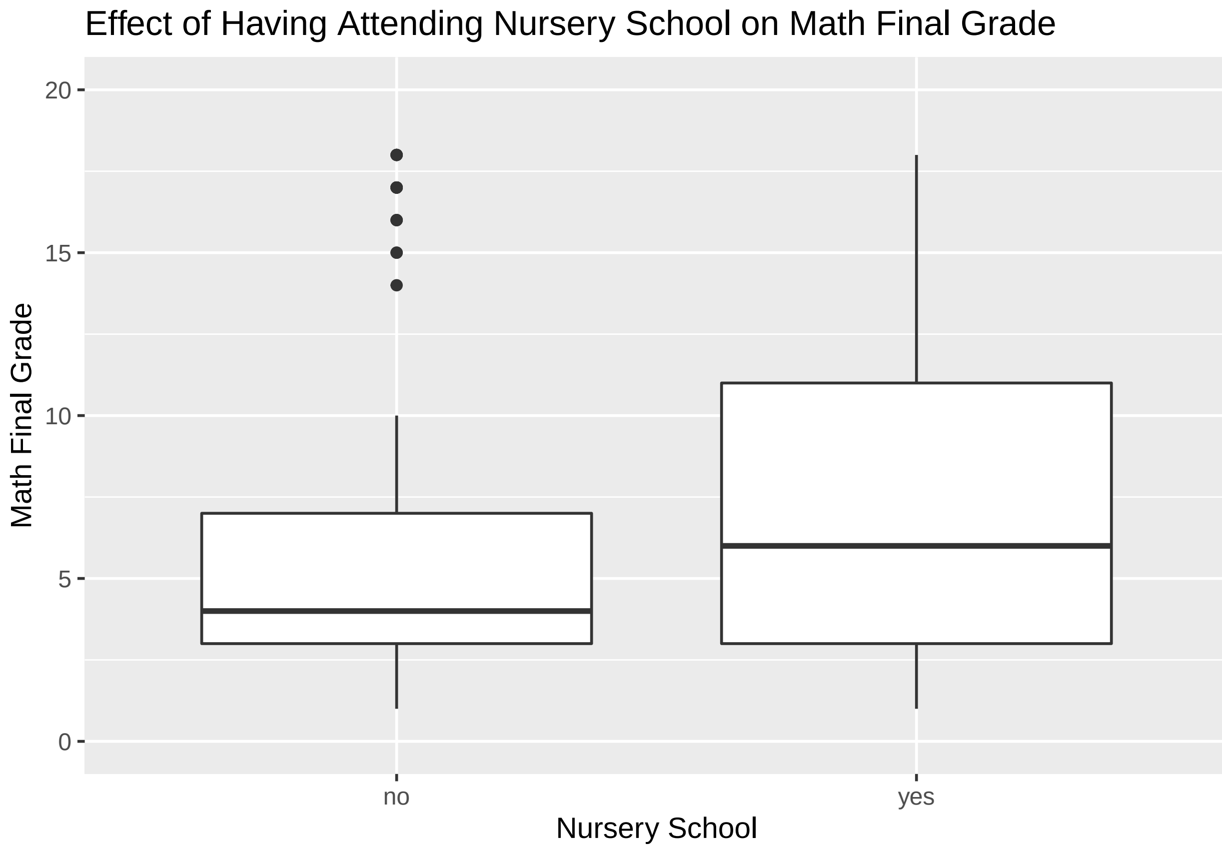
References

Variable visual and statistical analysis of starting variables using student survey data of Portuguese students for final math and Portuguese grades. The data set contains 649 observations for the Portuguese data and 395 observations for the math data set. Of the 33 variables 30 variables relate to a student's demographics such as what the parents do for work, what level of education the student's parents have attained, previous academic history, as well as other questions that relate to how a student spends time and what their social life is like.



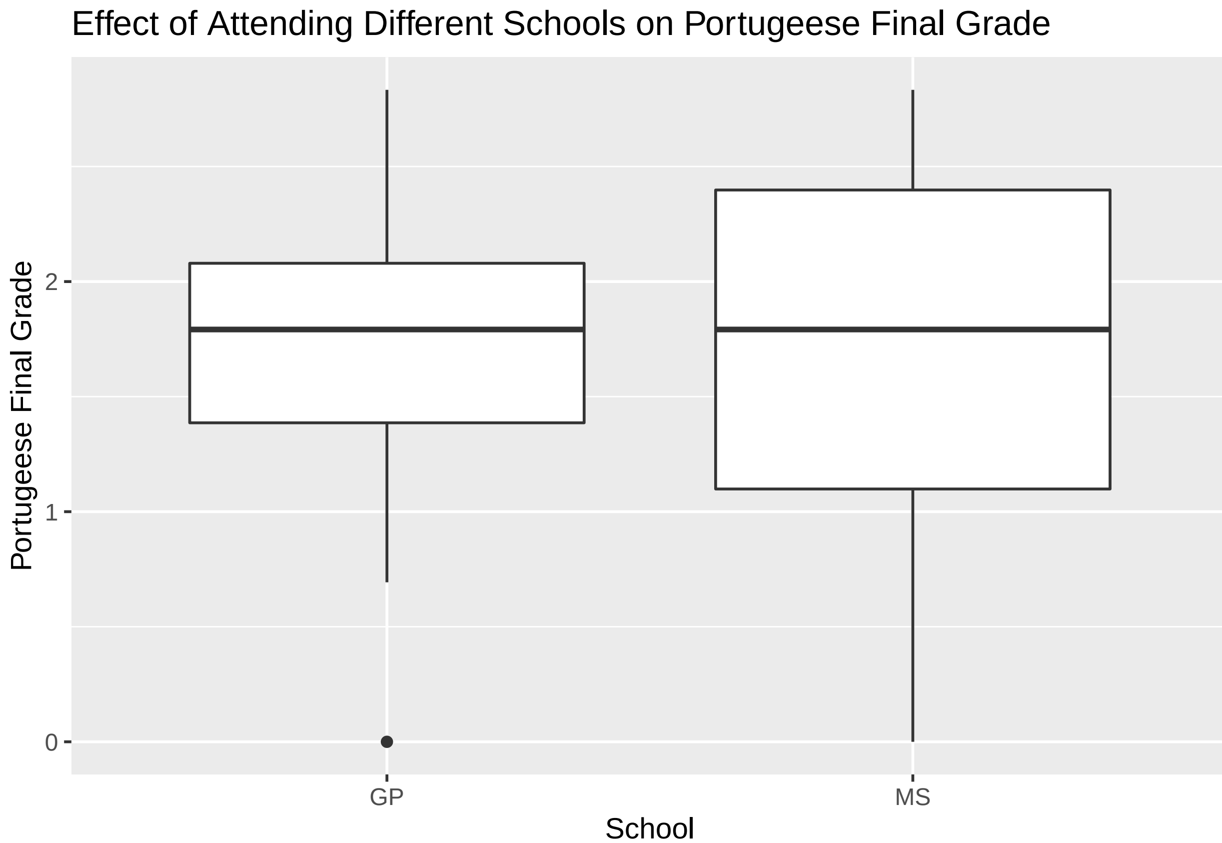
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Walc	4	190.5869	47.64672	1.469466	0.2107179
Residuals	390	12645.5650	32.42453	NA	NA

Figure 1. Box plot results of weekend alcohol use and the effect is has on final math grades. Here the medians are fairly close and near 5. Performing a 1-way ANOVA test yields a P-value of 0.2107179, with the P-value being greater than 0.05 we do not have enough evidence to say that weekend alcohol use alone has an effect on final math grades.

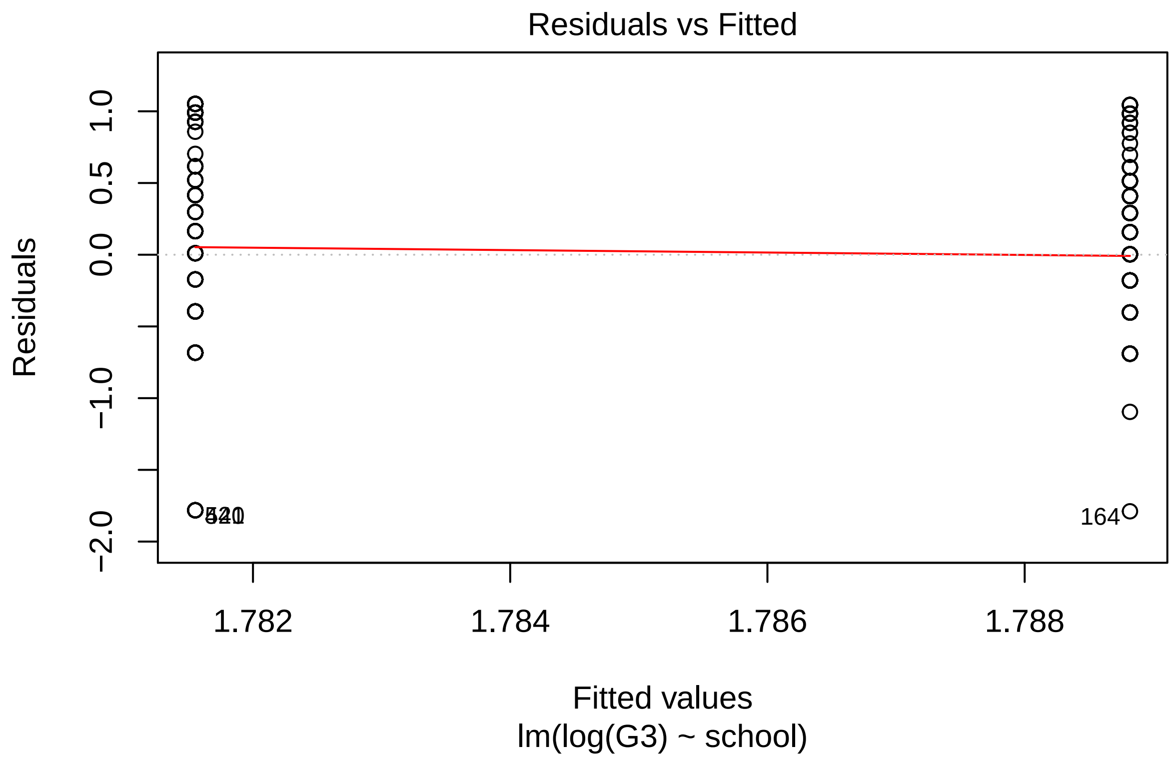


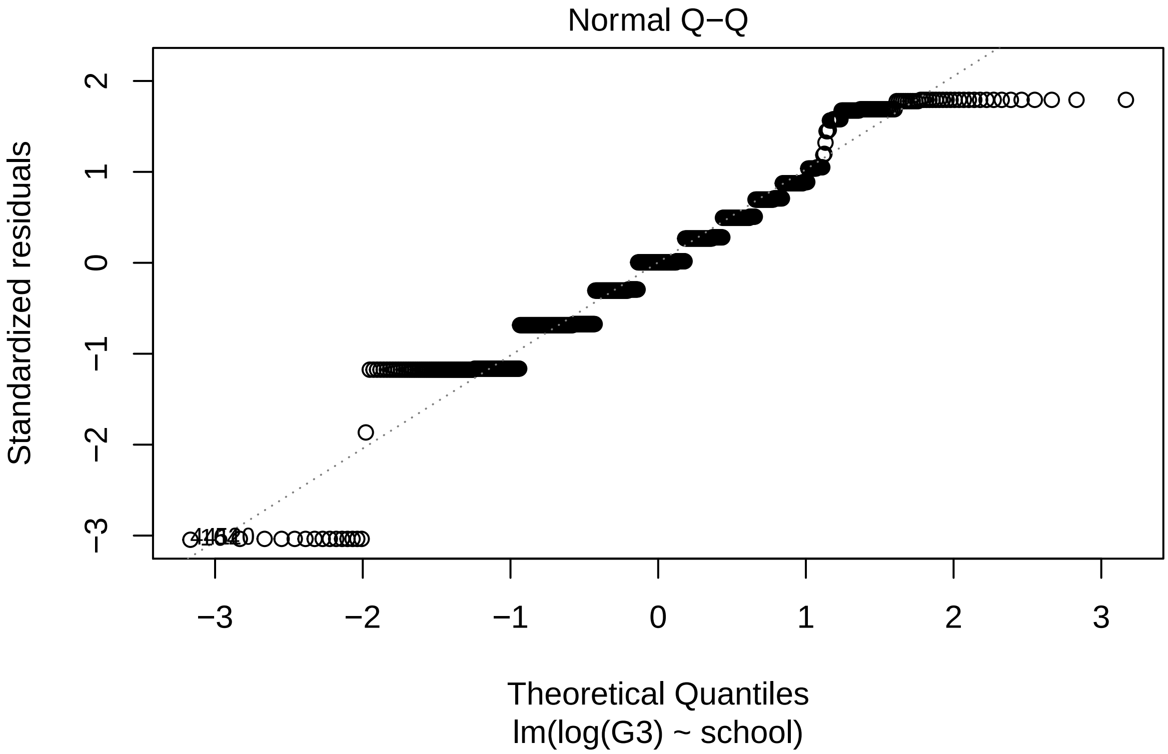
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
nursery	1	68.64006	68.64006	2.112827	0.1468673
Residuals	393	12767.51183	32.48731	NA	NA

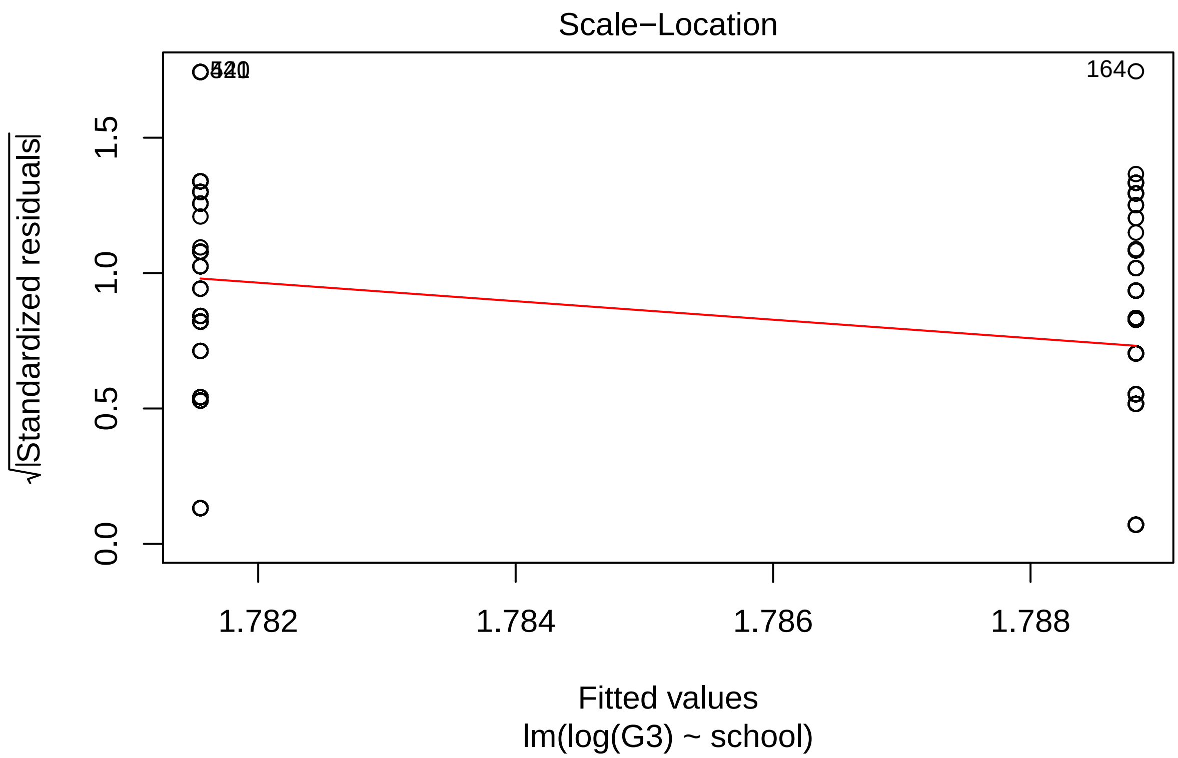
Figure 2. Box plot results of a student having attended nursery school and the effect is has on final math grades. Here the median for the students that did not attend nursery school is 4, while those that did have a median final math grade of 6. 1-way ANOVA test yields a P-value of 0.1468673, with the P-value being greater than 0.05 we do not have enough evidence to say that attending nursery school alone has an effect on final math grades.

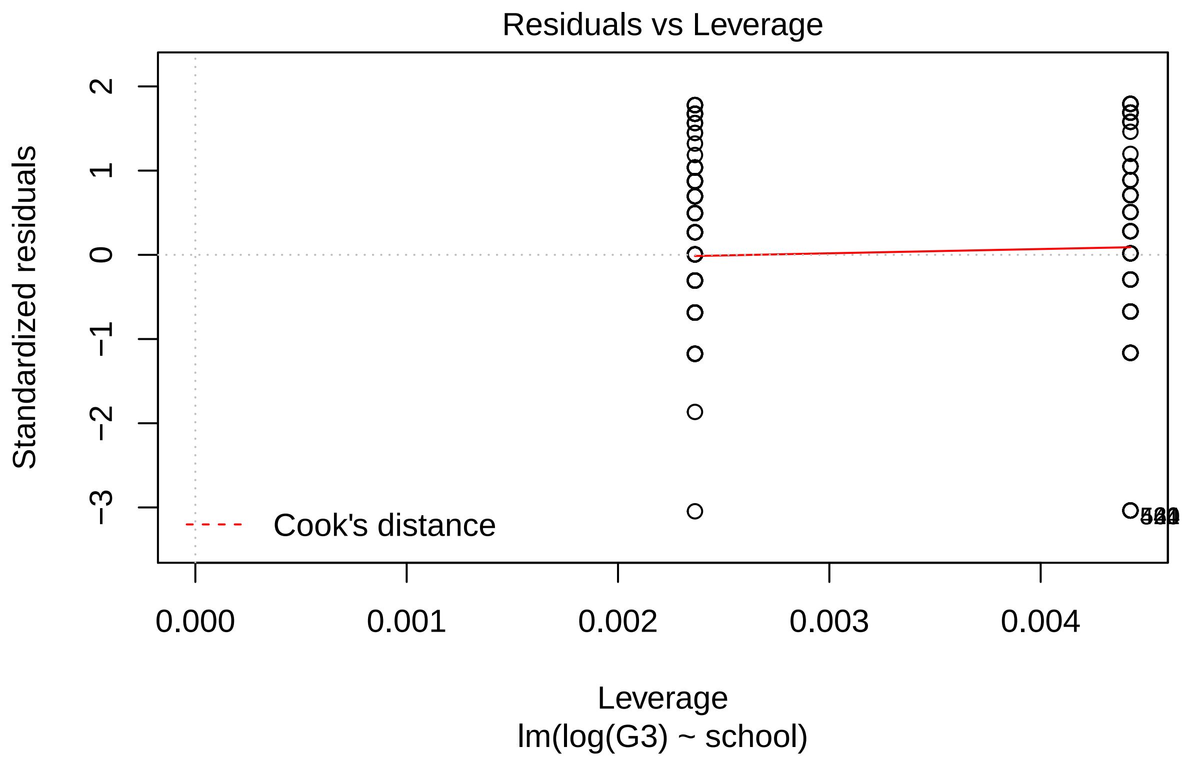


	Df	Sum Sq	Mean Sq	F value	Pr(>F)
school	1	162.8945	162.89453	9.443111	0.0022085
Residuals	647	11160.8096	17.25009	NA	NA

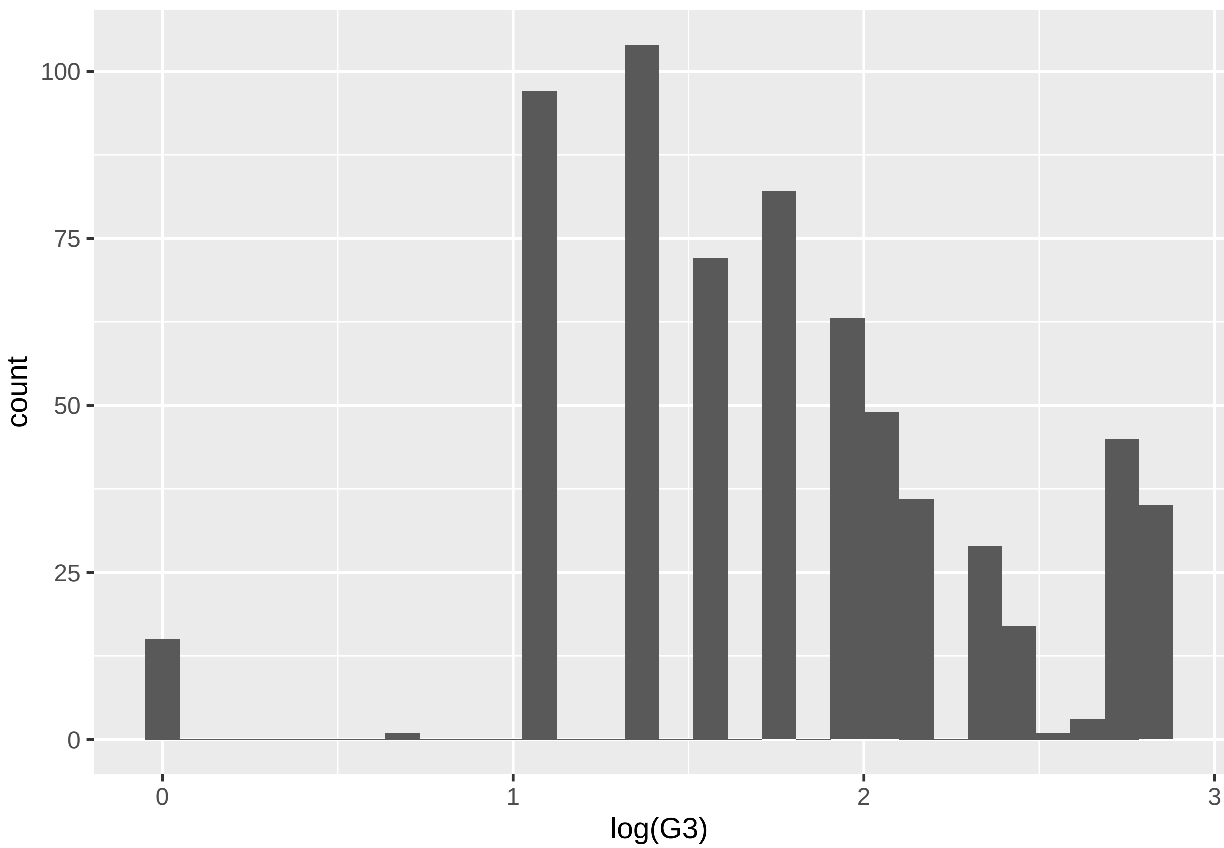

Residuals vs Fitted
Residuals
1.0
0.5
0.0
−1.0
−2.0
521
541
164
1.782
1.784
1.786
1.788
Fitted values
lm(log(G3) ~ school)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(G3) ~ school)
−3
−2
−1
0
1
2
3


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
520
541
164
1.782
1.784
1.786
1.788
Fitted values
lm(log(G3) ~ school)

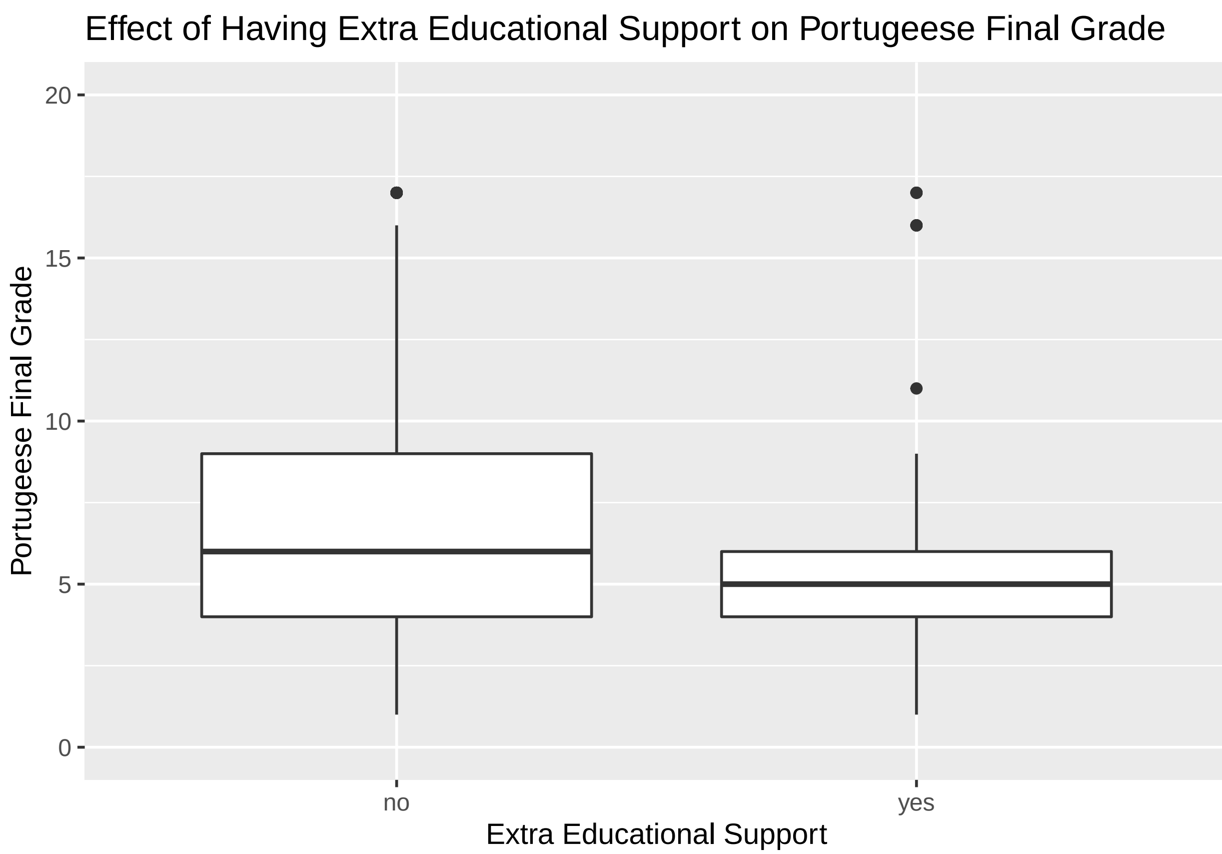



```
##
##  Shapiro-Wilk normality test
##
## data:  mm1$residuals
## W = 0.94803, p-value = 2.524e-14

##
## Call:
## lm(formula = log(G3) ~ school, data = porData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78882 -0.40252  0.00294  0.40841  1.05166
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.788819   0.028591   62.57   <2e-16 ***
## schoolMS    -0.007268   0.048451   -0.15    0.881
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.588 on 647 degrees of freedom
## Multiple R-squared:  3.477e-05,  Adjusted R-squared:  -0.001511
## F-statistic: 0.0225 on 1 and 647 DF,  p-value: 0.8808

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
##   Group.1        x
## 1      GP 6.678487
## 2      MS 7.730088
```

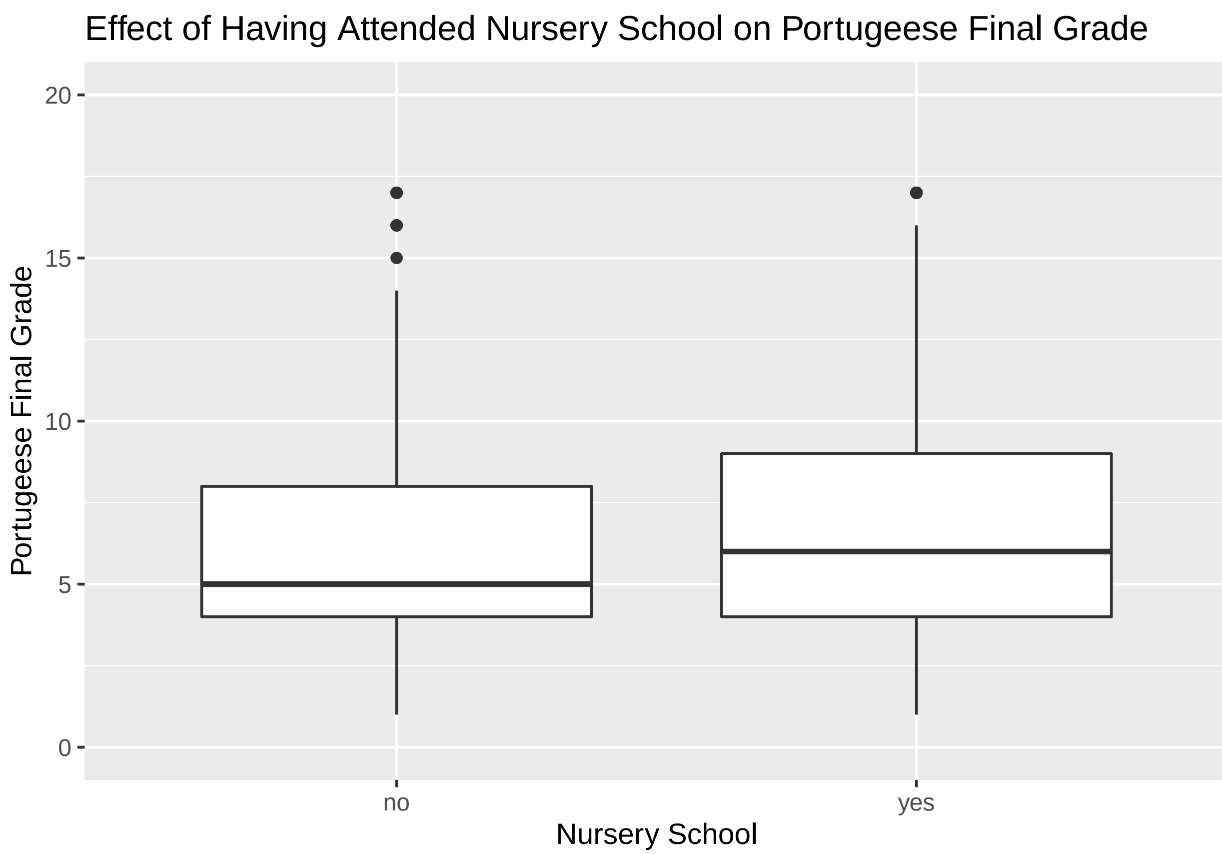
Figure 3. Box plot results of students attending different schools and the effect is has on final Portuguese grades. Here the medians are the same with both schools having a median final Portuguese grade of 6. 1-way ANOVA test yields a P-value of 0.0022085, with the P-value being less than 0.05 we have enough evidence to say that attending a different school alone has an effect on final Portuguese grades. Analyzing the linear model built we can see that by attending Mousinho da Silveira (MS) students can be expected to score 1.052 points higher on final Portuguese grades.



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
schoolsup	1	92.49711	92.49711	5.328513	0.0212937
Residuals	647	11231.20705	17.35890	NA	NA

```
##
## Call:
## lm(formula = G3 ~ schoolsup, data = porData)
##
## Coefficients:
##  (Intercept)  schoolsupyes
##        7.174        -1.233
```

Figure 4. Box plot results of students with extra educational support and the effect is has on final Portuguese grades. Here the medians are close, with students not receiving extra support having a meadin final Portuguese grade of 6 and those that did receive extra support a median final Portuguese grade of 5. 1-way ANOVA test yields a P-value of 0.0212937, with the P-value being less than 0.05 we have enough evidence to say that extra educational support alone has an effect on final math grades. Student that recieve extra educational support can see final Portuguese grades 1.233 less than those who did not receive extra support.



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
nursery	1	6.947937	6.947937	0.3972265	0.5287476
Residuals	647	11316.756223	17.491122	NA	NA

Figure 5. Box plot results of students who had attended nursery school and the effect is has on final Portuguese grades. Here the median average final Portuguese grades of students who did not attend nursery school is 5, and the median final Portuguese grade of those that did attend nursery school is 6. 1-way ANOVA test yields a P-value of 0.5287476, with the P-value being greater than 0.05 we not not have enough evidence to say that attending nursery school alone has an effect on final Portuguese grades.

```
##
## Call:
## lm(formula = G3 ~ Walc + school + schoolsup + nursery, data = mathData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.017 -4.608 -1.616  3.001 11.999
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.1484     0.7764   7.919 2.55e-14 ***
## Walc2         -1.1908     0.7806  -1.525   0.1280
## Walc3          0.4161     0.7943   0.524   0.6007
## Walc4          0.6641     0.9226   0.720   0.4721
```

```
## Walc5          0.5840     1.1760   0.497   0.6198
## schoolMS       0.6141     0.9112   0.674   0.5007
## schoolsupyes   1.8250     0.8698   2.098   0.0365 *
## nurseryyes     1.0439     0.7142   1.462   0.1446
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.667 on 387 degrees of freedom
## Multiple R-squared:  0.03181,    Adjusted R-squared:  0.0143
## F-statistic: 1.817 on 7 and 387 DF,  p-value: 0.08268
```

```
## Analysis of Variance Table
##
## Response: G3
##            Df  Sum Sq Mean Sq F value  Pr(>F)
## Walc        4   190.6  47.647  1.4837 0.20636
## school      1     2.6   2.566  0.0799 0.77759
## schoolsup   1   146.6 146.607  4.5653 0.03325 *
## nursery     1    68.6  68.610  2.1365 0.14464
## Residuals 387 12427.8  32.113
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = G3 ~ Walc + schoolsup + nursery, data = mathData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.992 -4.728 -1.732  3.260 11.917
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.2368     0.7647   8.156 4.85e-15 ***
## Walc2         -1.1573     0.7785  -1.487   0.1379
## Walc3          0.4915     0.7858   0.625   0.5321
## Walc4          0.6867     0.9213   0.745   0.4565
## Walc5          0.5733     1.1751   0.488   0.6259
## schoolsupyes   1.7516     0.8624   2.031   0.0429 *
## nurseryyes     1.0036     0.7112   1.411   0.1590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.663 on 388 degrees of freedom
## Multiple R-squared:  0.03068,    Adjusted R-squared:  0.01569
## F-statistic: 2.047 on 6 and 388 DF,  p-value: 0.05868
```

```
## Analysis of Variance Table
##
## Response: G3
##            Df  Sum Sq Mean Sq F value  Pr(>F)
## Walc        4   190.6  47.647  1.4858 0.20572
## schoolsup   1   139.3 139.332  4.3449 0.03777 *
```

```
## nursery     1    63.9  63.864  1.9915 0.15898
## Residuals 388 12442.4  32.068
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

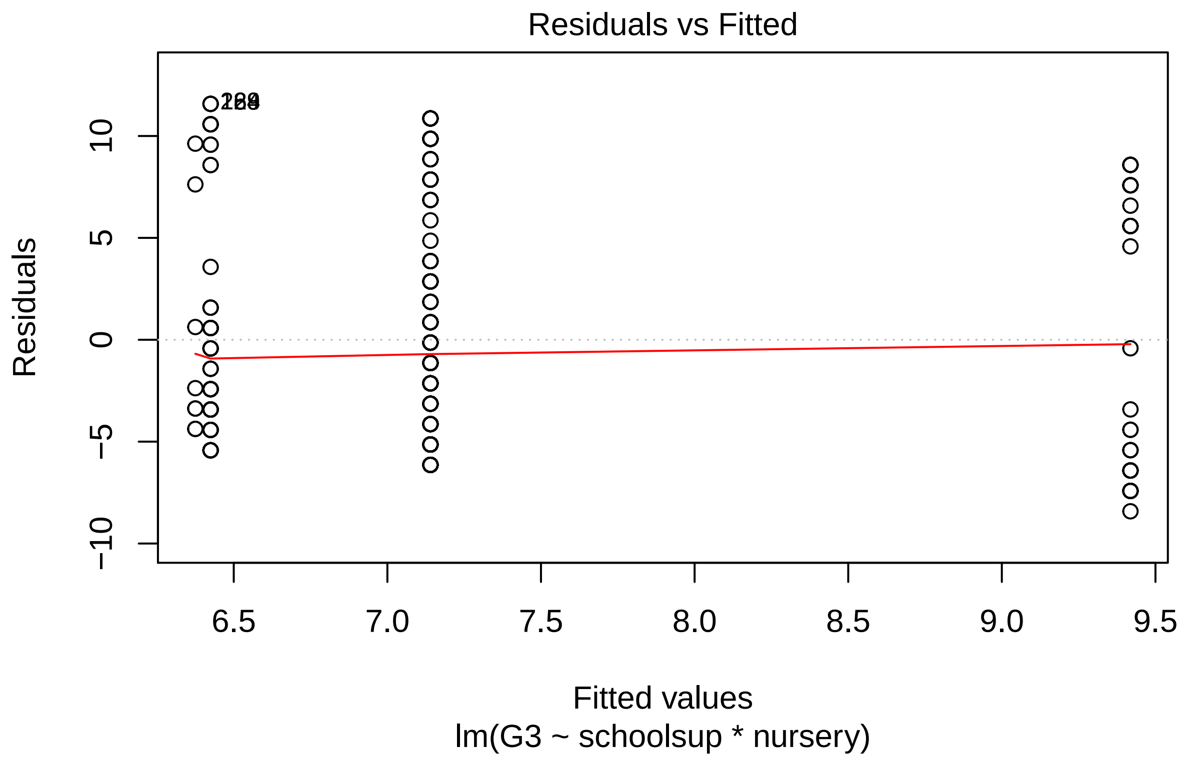
```r
m4 <- lm( G3 ~ schoolsup * nursery, mathData)

summary( m4)
```

```
##
## Call:
## lm(formula = G3 ~ schoolsup * nursery, data = mathData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.419 -4.425 -2.140  3.218 11.575
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              6.42466    0.66375   9.679   <2e-16 ***
## schoolsupyes            -0.04966    2.11203  -0.024    0.981
## nurseryyes               0.71556    0.74782   0.957    0.339
## schoolsupyes:nurseryyes  2.32804    2.30808   1.009    0.314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.671 on 391 degrees of freedom
## Multiple R-squared:  0.02036,    Adjusted R-squared:  0.01284
## F-statistic: 2.708 on 3 and 391 DF,  p-value: 0.04495
```

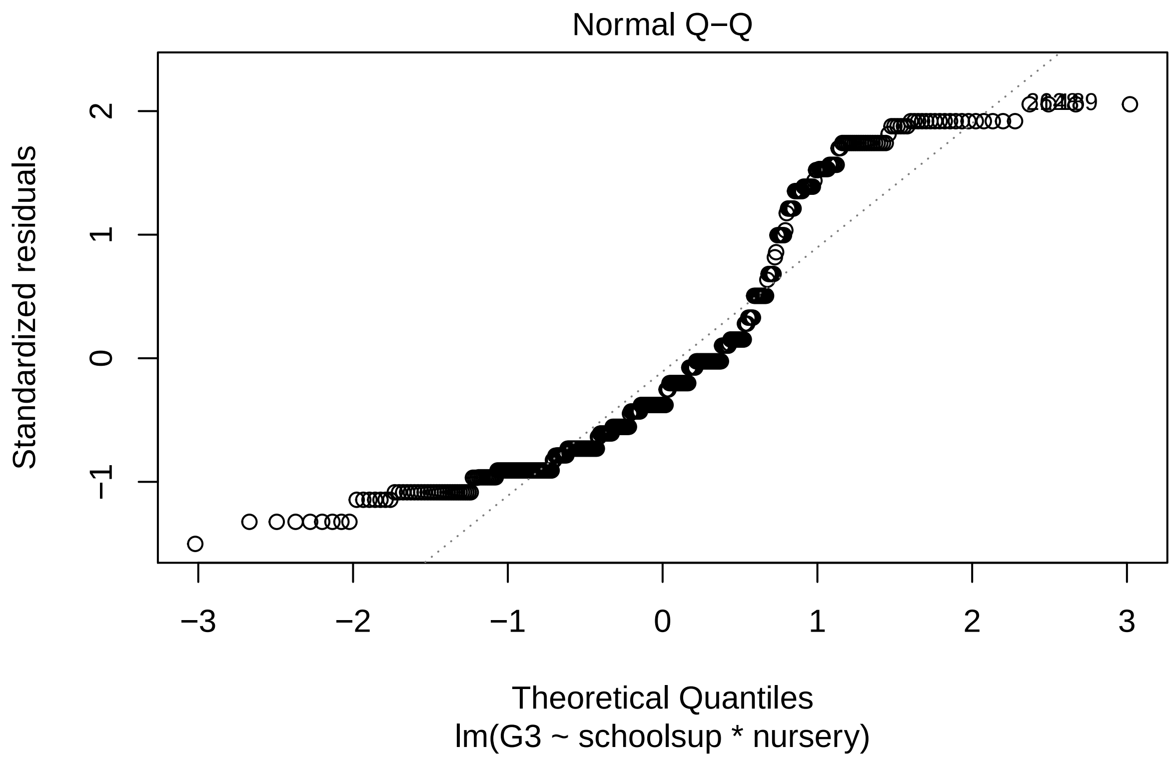
```r
anova(m4)
```

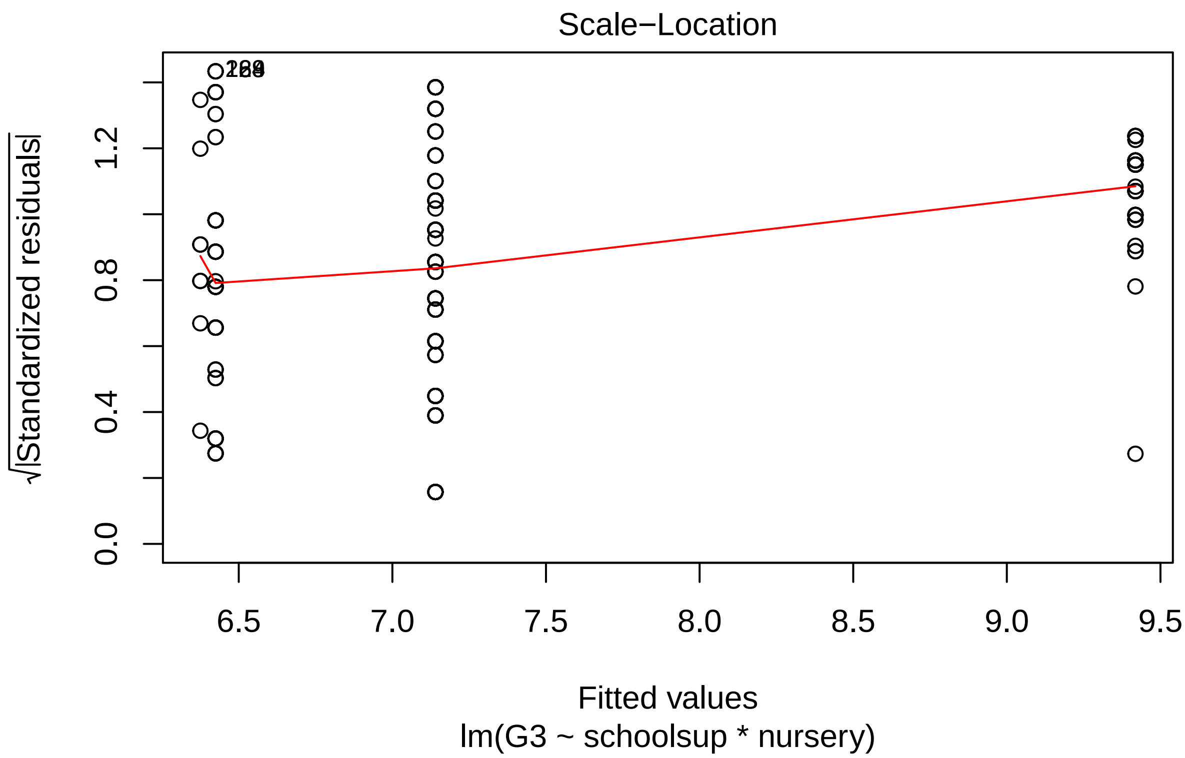
```
## Analysis of Variance Table
##
## Response: G3
##                    Df  Sum Sq Mean Sq F value  Pr(>F)
## schoolsup           1   169.4 169.375  5.2665 0.02227 *
## nursery             1    59.2  59.210  1.8411 0.17561
## schoolsup:nursery   1    32.7  32.719  1.0174 0.31377
## Residuals         391 12574.8  32.161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

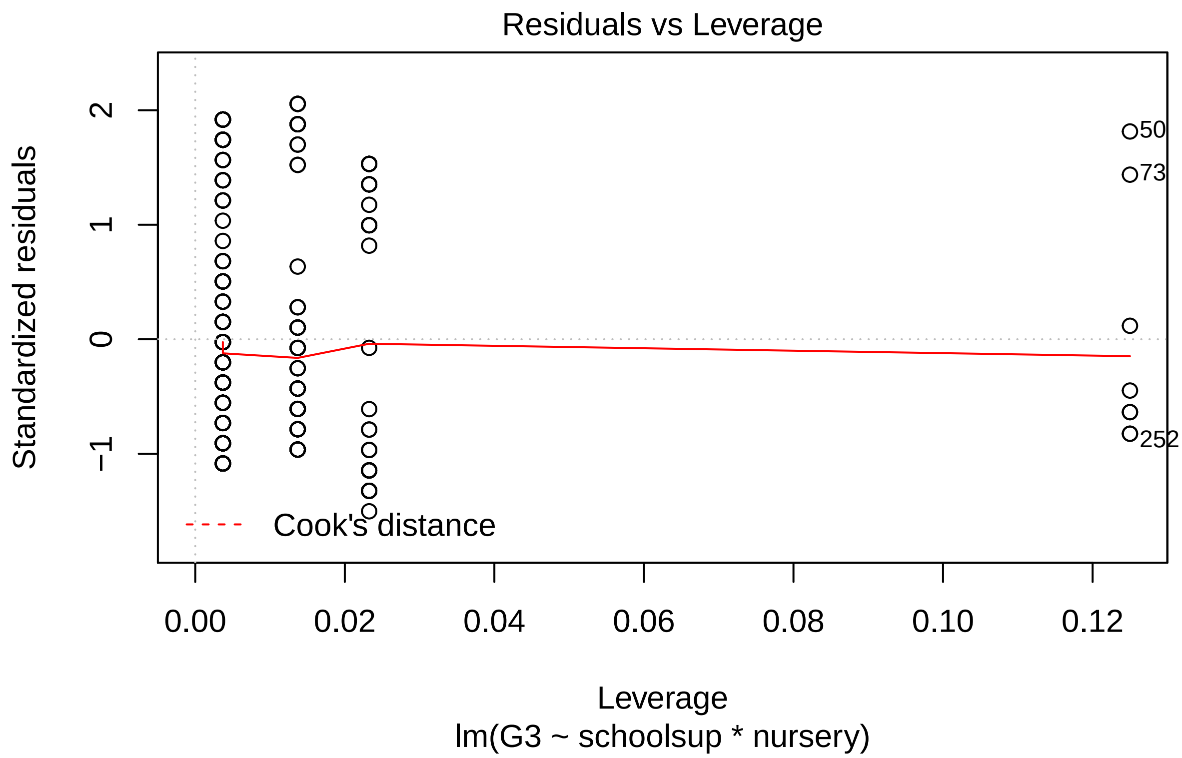
```r
plot(m4)
```



Residuals vs Fitted
Residuals
10
5
0
−5
−10
6.5
7.0
7.5
8.0
8.5
9.0
9.5
Fitted values
lm(G3 ~ schoolsup * nursery)


Normal Q−Q
Standardized residuals
2
1
0
−1
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(G3 ~ schoolsup * nursery)


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
6.5
7.0
7.5
8.0
8.5
9.0
9.5
Fitted values
lm(G3 ~ schoolsup * nursery)



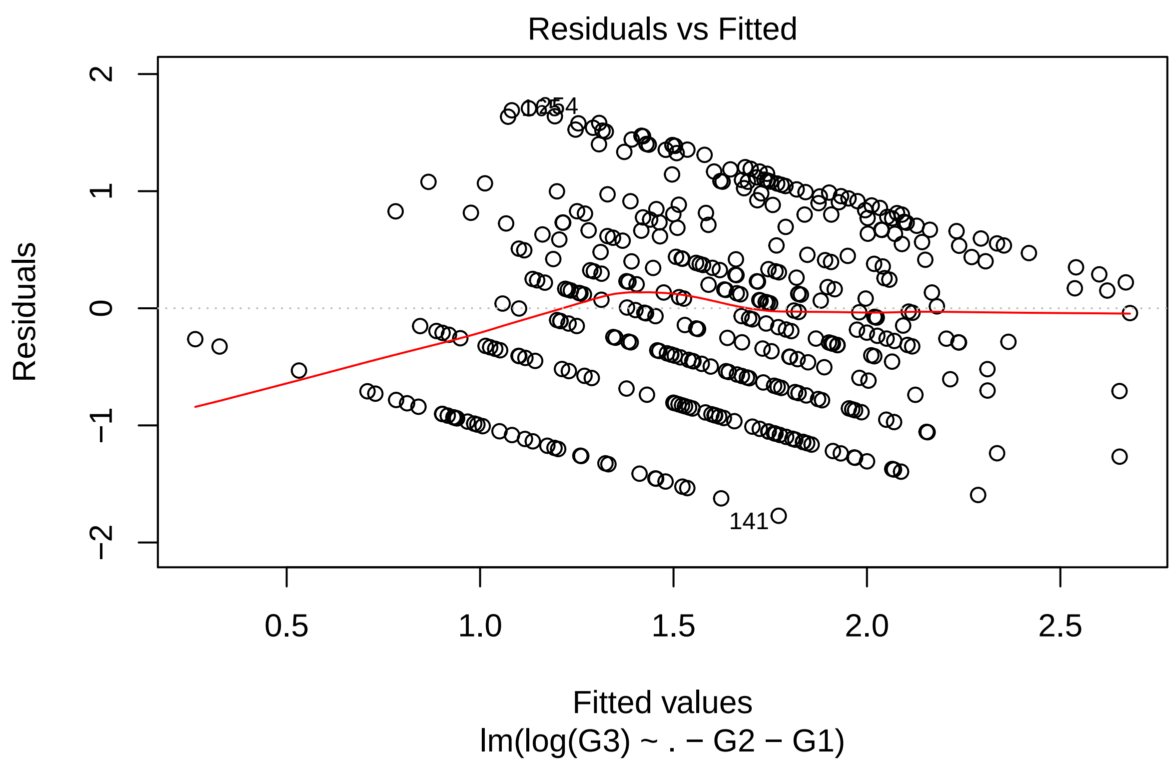

```r
m6 <- lm( log(G3) ~ . - G2 - G1 , mathData)
summary(m6)
```

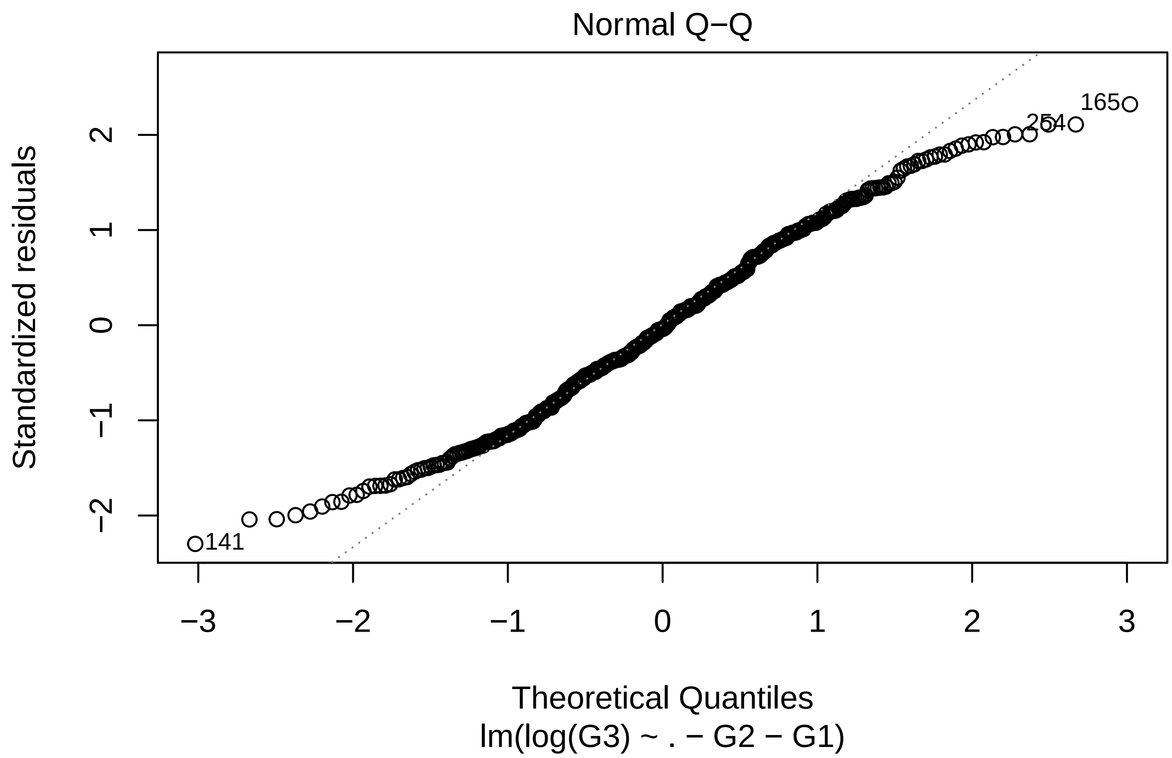
```
##
## Call:
## lm(formula = log(G3) ~ . - G2 - G1, data = mathData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77191 -0.59644 -0.03164  0.63621  1.70693
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.0466045  1.0636694   0.984   0.3259
## schoolMS            0.0067786  0.1756533   0.039   0.9692
## sexM                0.0636875  0.1104666   0.577   0.5647
## age                 0.0687065  0.0479361   1.433   0.1527
## addressU           -0.0169223  0.1314184  -0.129   0.8976
## famsizeLE3          0.1133183  0.1092361   1.037   0.3003
## PstatusT            0.1138492  0.1618019   0.704   0.4822
## Medu1              -1.1775651  0.5574991  -2.112   0.0354 *
## Medu2              -1.0930185  0.5576085  -1.960   0.0508 .
## Medu3              -0.8170385  0.5605136  -1.458   0.1459
## Medu4              -0.6348987  0.5777618  -1.099   0.2726
## Fedu1               0.4000278  0.6648405   0.602   0.5478
```

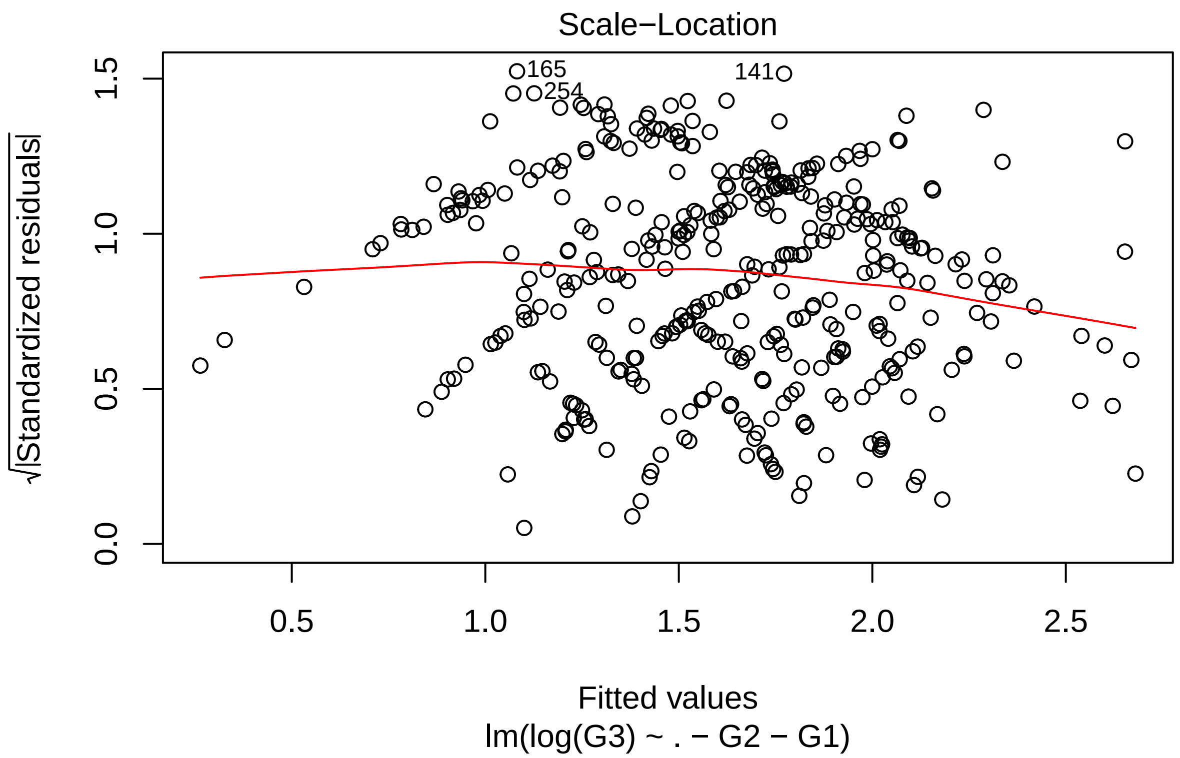
```
## Fedu2             0.2439659  0.6663435   0.366   0.7145
## Fedu3             0.1975929  0.6664146   0.297   0.7670
## Fedu4             0.0728508  0.6778141   0.107   0.9145
## Mjobhealth       -0.1655066  0.2504727  -0.661   0.5092
## Mjobother         0.1708370  0.1565547   1.091   0.2760
## Mjobservices      0.0114974  0.1769826   0.065   0.9482
## Mjobteacher      -0.1223210  0.2349067  -0.521   0.6029
## Fjobhealth        0.3956346  0.3169903   1.248   0.2129
## Fjobother        -0.1690837  0.2271997  -0.744   0.4573
## Fjobservices      0.0773252  0.2339150   0.331   0.7412
## Fjobteacher      -0.0407024  0.2930231  -0.139   0.8896
## reasonhome       -0.0972083  0.1220825  -0.796   0.4265
## reasonother       0.0408137  0.1786692   0.228   0.8195
## reasonreputation  0.0377960  0.1263770   0.299   0.7651
## guardianmother    0.0874250  0.1205720   0.725   0.4689
## guardianother     0.0884170  0.2228676   0.397   0.6918
## traveltime2      -0.0951130  0.1134273  -0.839   0.4023
## traveltime3      -0.1920904  0.2197076  -0.874   0.3826
## traveltime4      -0.3071745  0.3678470  -0.835   0.4043
## studytime2        0.1070648  0.1221751   0.876   0.3815
## studytime3       -0.0428057  0.1685184  -0.254   0.7996
## studytime4       -0.1711339  0.2185073  -0.783   0.4341
## failures1        -0.1796897  0.1559385  -1.152   0.2500
## failures2         0.1270084  0.2501653   0.508   0.6120
## failures3        -0.2618893  0.2681186  -0.977   0.3294
## schoolsupyes      0.2504655  0.1473769   1.699   0.0902 .
## famsupyes         0.0720853  0.1044841   0.690   0.4907
## paidyes          -0.0486792  0.1077453  -0.452   0.6517
## activitiesyes    -0.1167636  0.0984933  -1.185   0.2367
## nurseryyes        0.1803399  0.1220223   1.478   0.1404
## higheryes         0.1406723  0.2416345   0.582   0.5609
## internetyes       0.1327838  0.1353839   0.981   0.3274
## romanticyes      -0.1605450  0.1054043  -1.523   0.1287
## famrel2          -0.1229267  0.4068379  -0.302   0.7627
## famrel3           0.0497885  0.3547289   0.140   0.8885
## famrel4          -0.0860539  0.3442520  -0.250   0.8028
## famrel5          -0.1296383  0.3507216  -0.370   0.7119
## freetime2         0.2072312  0.2497897   0.830   0.4074
## freetime3         0.0006431  0.2378745   0.003   0.9978
## freetime4         0.2039166  0.2457740   0.830   0.4073
## freetime5         0.3243873  0.2812153   1.154   0.2495
## goout2            0.3260881  0.2173838   1.500   0.1346
## goout3            0.4260342  0.2192546   1.943   0.0529 .
## goout4            0.4226273  0.2297344   1.840   0.0667 .
## goout5            0.3496224  0.2474523   1.413   0.1586
## Dalc2            -0.2042117  0.1444206  -1.414   0.1583
## Dalc3            -0.2745863  0.2290987  -1.199   0.2316
## Dalc4             0.3177508  0.3536094   0.899   0.3695
## Dalc5            -0.2932338  0.4076736  -0.719   0.4725
## Walc2            -0.2260255  0.1352233  -1.671   0.0956 .
## Walc3             0.0054306  0.1511698   0.036   0.9714
## Walc4            -0.0438364  0.1947699  -0.225   0.8221
## Walc5             0.1611311  0.2920619   0.552   0.5815
## health2          -0.2013012  0.1992179  -1.010   0.3130
```

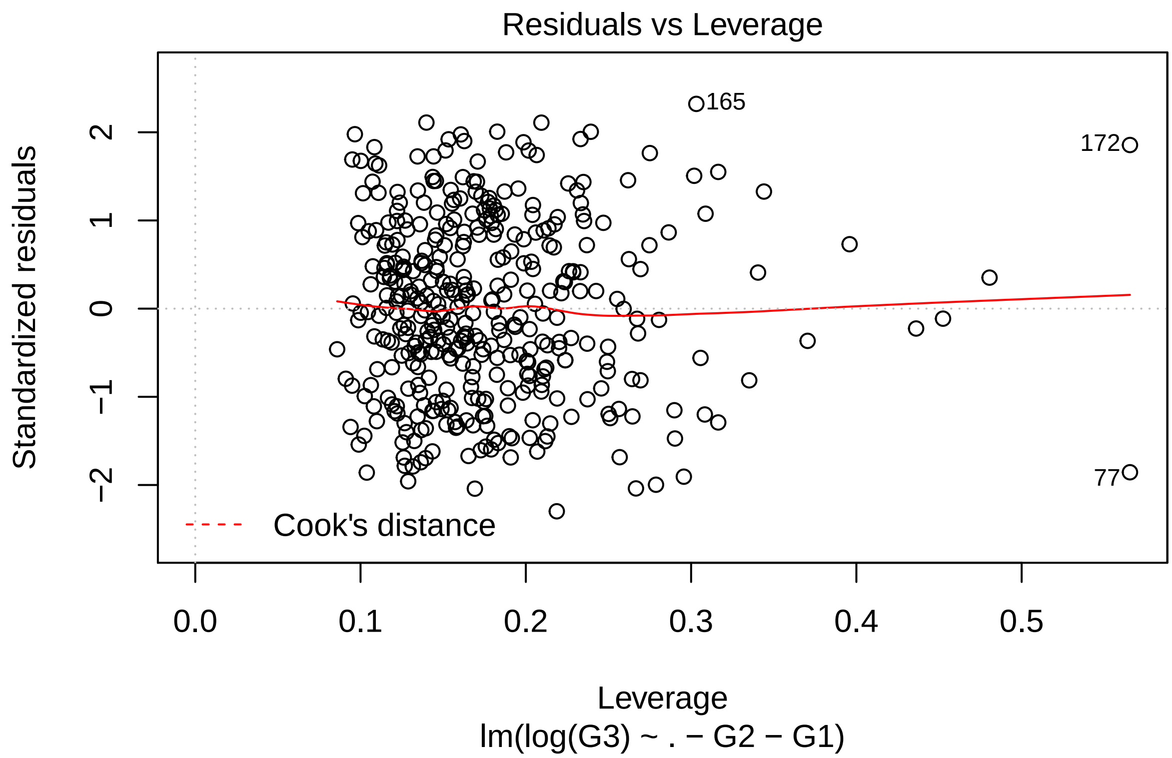
```
## health3            0.0404483  0.1761005   0.230   0.8185
## health4           -0.0974457  0.1843287  -0.529   0.5974
## health5            0.1402555  0.1648774   0.851   0.3956
## absences           0.0173479  0.0041931   4.137 4.48e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8723 on 325 degrees of freedom
## Multiple R-squared:  0.2061, Adjusted R-squared:  0.03751
## F-statistic: 1.223 on 69 and 325 DF,  p-value: 0.1284
```

```
plot(m6)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(G3) ~ . − G2 − G1)
165
254
141
−3
−2
−1
0
1
2
3
−2
−1
0
1
2


Scale−Location
√|Standardized residuals|
Fitted values
lm(log(G3) ~ . − G2 − G1)
165
254
141
0.0
0.5
1.0
1.5
0.5
1.0
1.5
2.0
2.5




```r
shapiro.test(m6$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  m6$residuals
## W = 0.98342, p-value = 0.0001694
```

```r
# cor(mathData %>% select_if(is.numeric()))
# str(mathData)
```