Lab8-Lin

Bin Lin

2016-12-4

```
#install.packages("StMoSim")
library(StMoSim)
```

```
## Warning: package 'StMoSim' was built under R version 3.3.2
```

```
## Loading required package: RcppParallel
```

```
## Warning: package 'RcppParallel' was built under R version 3.3.2
```

```
## Loading required package: Rcpp
```

```
##
## Attaching package: 'Rcpp'
```

```
## The following object is masked from 'package:RcppParallel':
##
##     LdFlags
```

```
library(IS606)
```

```
##
## Welcome to CUNY IS606 Statistics and Probability for Data Analytics
## This package is designed to support this course. The text book used
## is OpenIntro Statistics, 3rd Edition. You can read this by typing
## vignette('os3') or visit www.OpenIntro.org.
##
## The getLabs() function will return a list of the labs available.
##
## The demo(package='IS606') will list the demos that are available.
```

```
##
## Attaching package: 'IS606'
```

```
## The following object is masked from 'package:utils':
##
##     demo
```

```
#startLab('Lab8')
setwd('C:/Users/blin261/Documents/Lab8')
load("more/evals.RData")
```

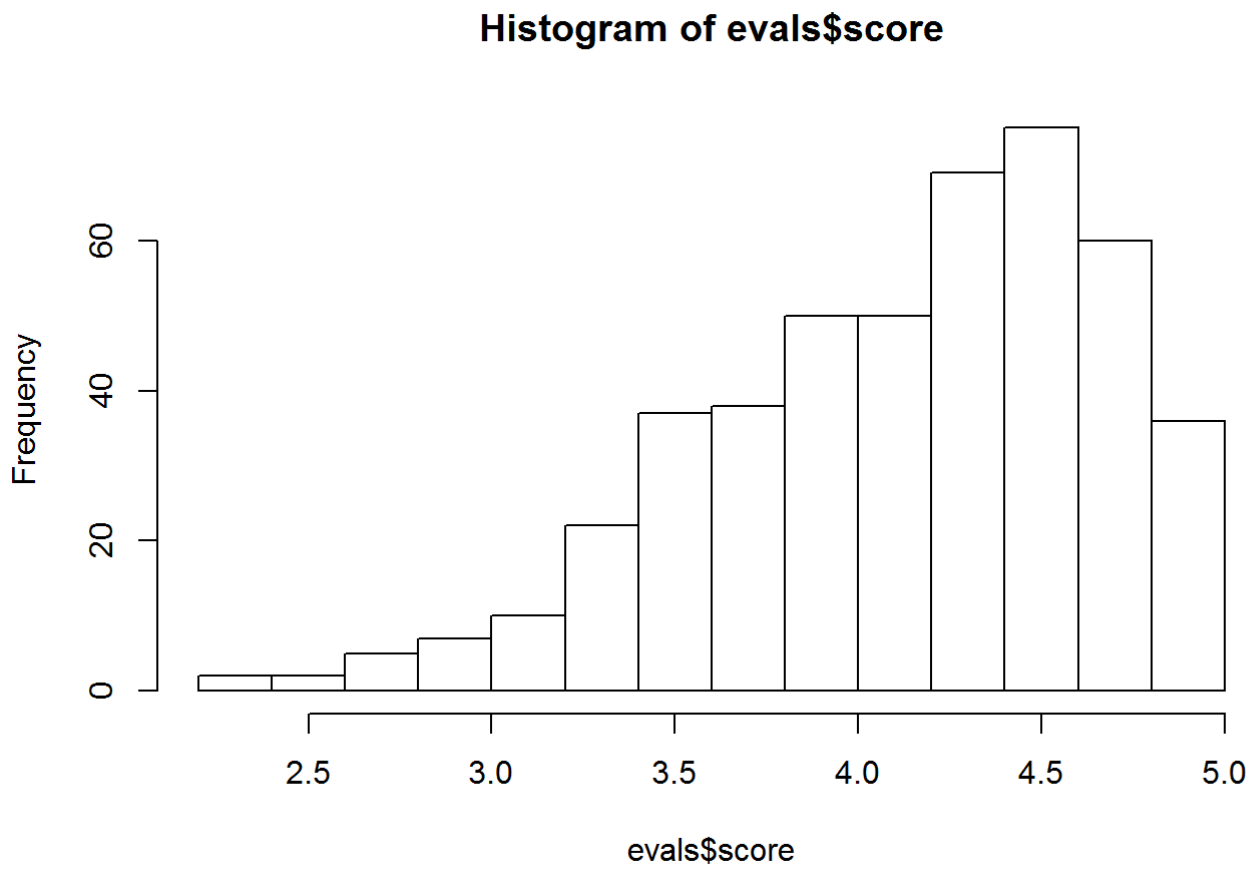
Exercise 1: Is this an observational study or an experiment? The original research question posed in the paper is whether beauty leads directly to the differences in course evaluations. Given the study design, is it possible to answer this question as it is phrased? If not, rephrase the question.

This is an observational study. Since we are not introducing intervention to this study. There is no control or experimental group. All it is doing is to evaluate existing data. It is not possible to answer this question as it is phrased. The question can be is the instructor's beauty positively or negatively correlated with course evaluation.

Exercise 2: Describe the distribution of score. Is the distribution skewed? What does that tell you about how students rate courses? Is this what you expected to see? Why, or why not?

According to the histogram of score, this distribution is skewed to the left. There are more students give instructor higher scores than the lower scores. This is not what I expected to see. I expected the distribution will closed to normal as it is unimodal, bell-shaped and symmetric.

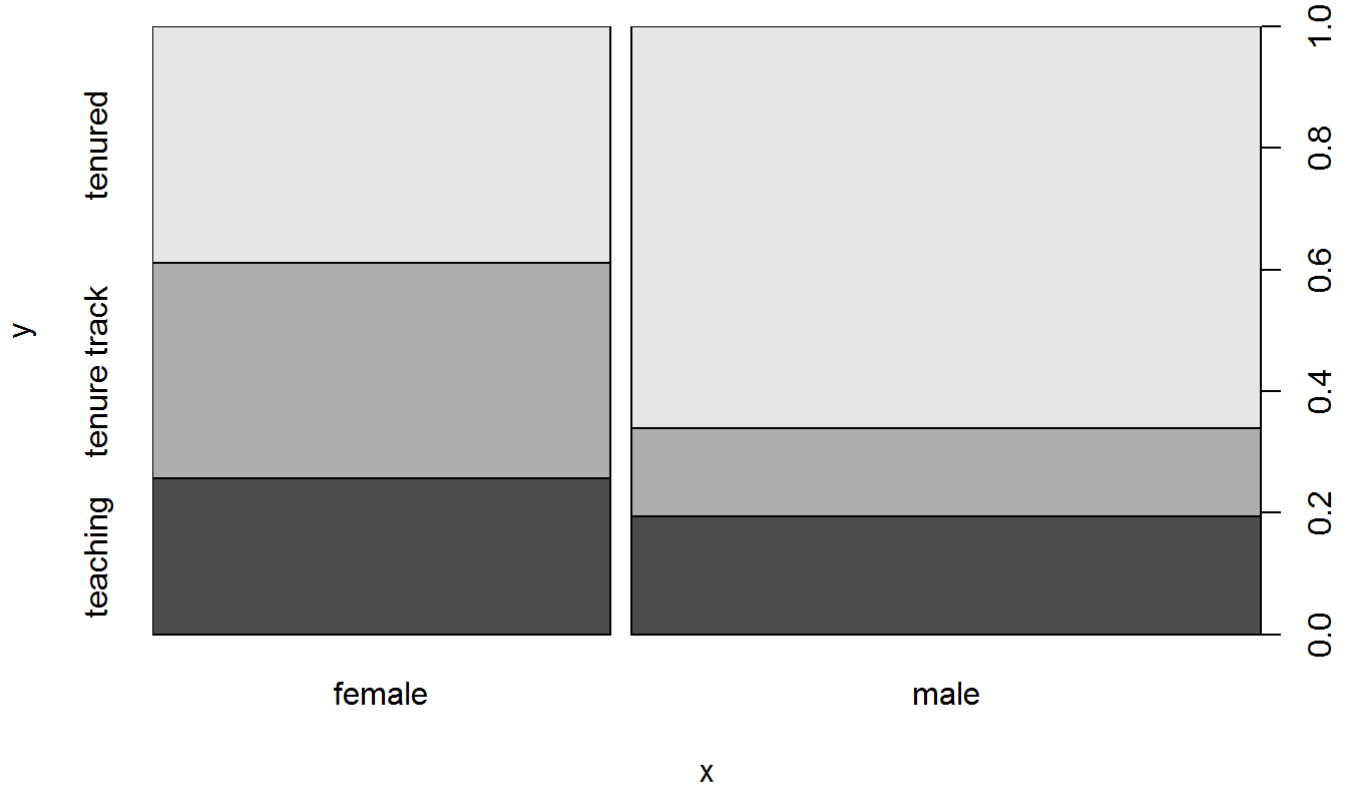
```
hist(evals$score)
```



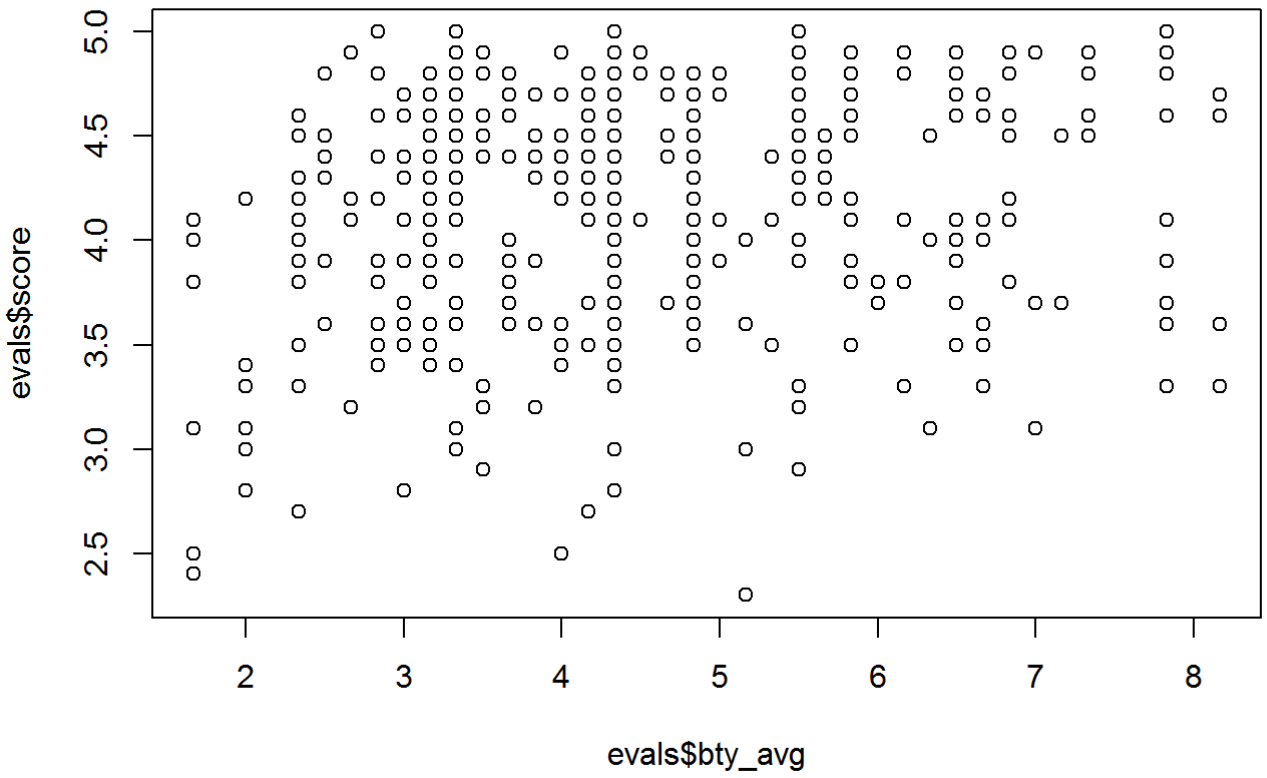
Exercise 3: Excluding score, select two other variables and describe their relationship using an appropriate visualization (scatterplot, side-by-side boxplots, or mosaic plot).

As shown in the boxplot, male instructors are much more likely than female instrctors to be tenured.

```
plot(evals$gender, evals$rank)
```



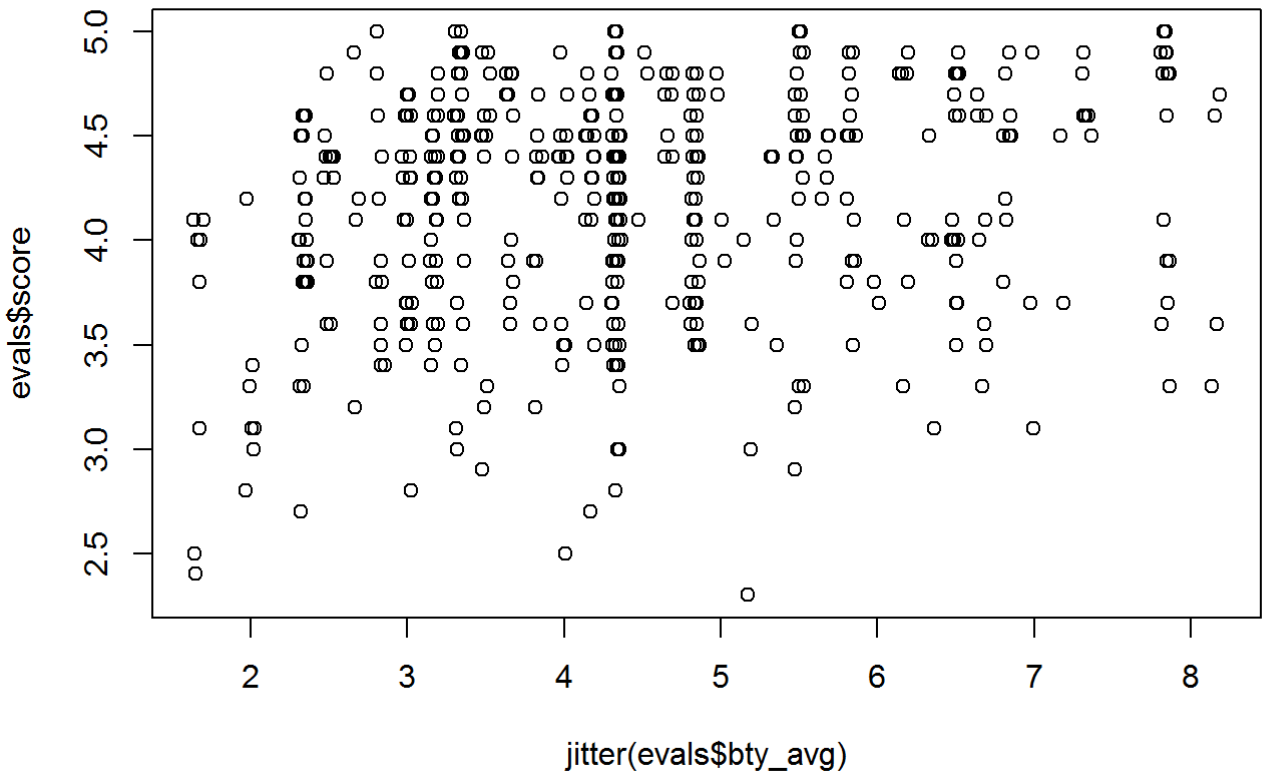

```
plot(evals$score ~ evals$bty_avg)
```



Exercise 4: Replot the scatterplot, but this time use the function jitter() on the yy- or the xx-coordinate. (Use ? jitter to learn more.) What was misleading about the initial scatterplot?

The initial scatterplot have multiple points that have same value. Because they overlap each other, on the scatterplot, they appear to be just one point.

```
plot(evals$score ~ jitter(evals$bty_avg))
```



Exercise 5: Let's see if the apparent trend in the plot is something more than natural variation. Fit a linear model called m_bty to predict average professor score by average beauty rating and add the line to your plot using abline(m_bty). Write out the equation for the linear model and interpret the slope. Is average beauty score a statistically significant predictor? Does it appear to be a practically significant predictor?

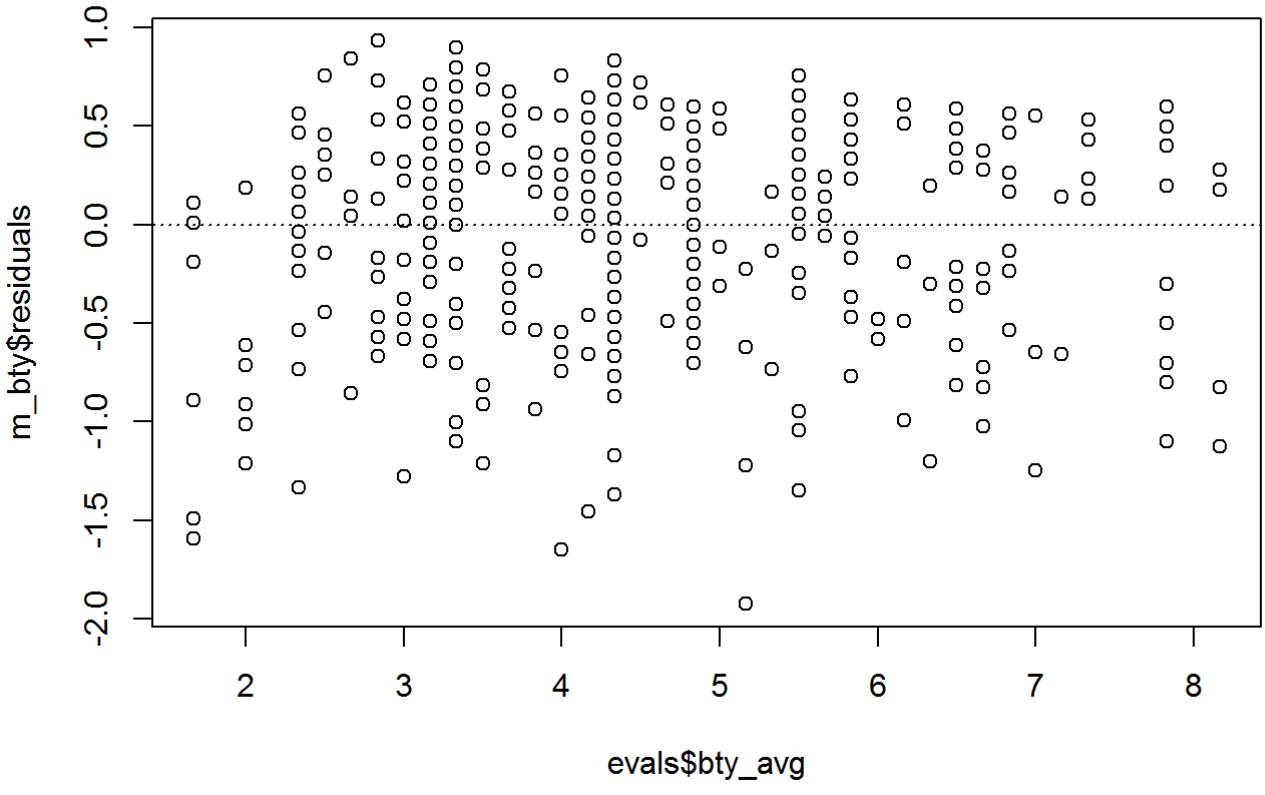
The equaltion for the linear model is: y = 3.88034 + 0.06664 * x. It is statistically significant because the p-value is 5.08e-05 which is very closed to 0. However, it may not be practically significant, because the slope of this line is only 0.06664, which mean for any 1 point increase of the beauty score, the instructors' evaluation score only increase by barely 0.0664.

```
m_bty <- lm(evals$score ~ evals$bty_avg)
summary(m_bty)
```

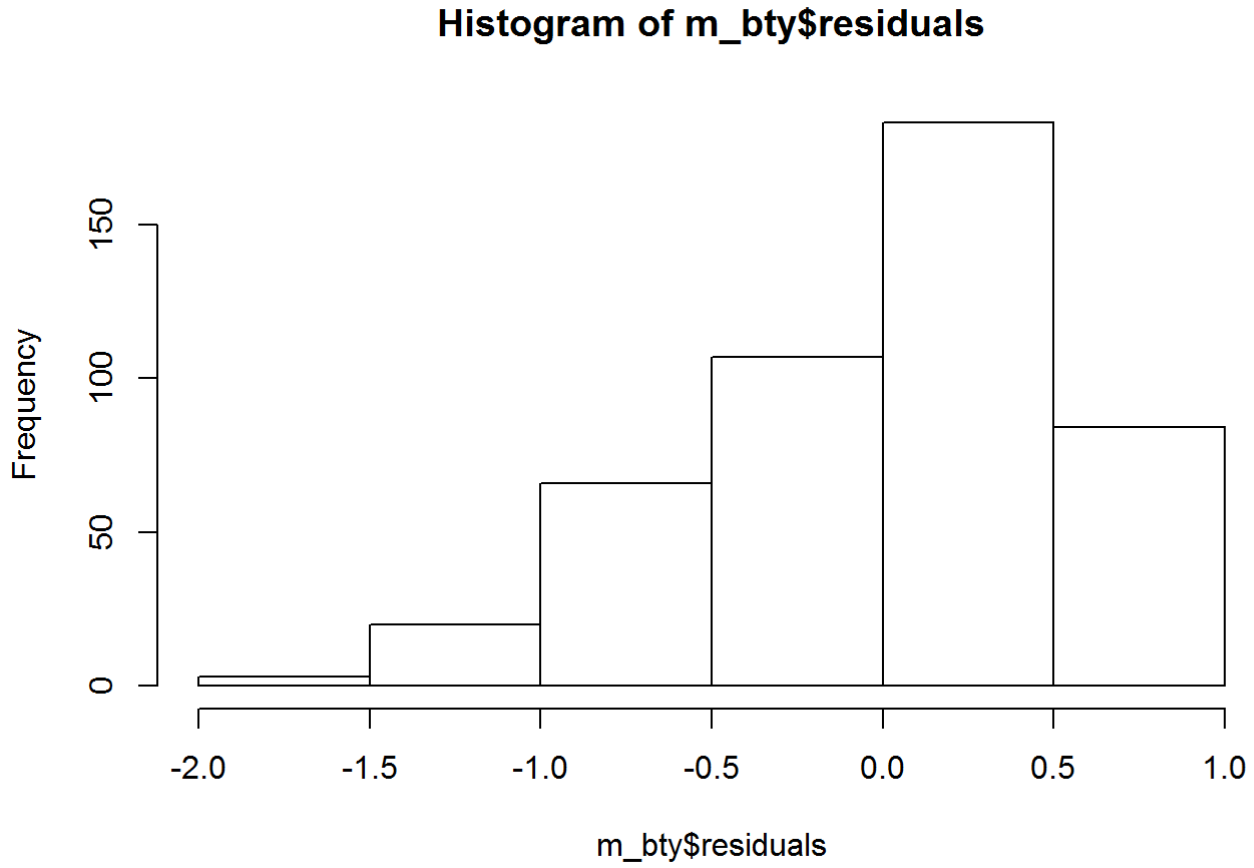
```
##
## Call:
## lm(formula = evals$score ~ evals$bty_avg)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9246 -0.3690  0.1420  0.3977  0.9309
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.88034    0.07614   50.96  < 2e-16 ***
## evals$bty_avg  0.06664    0.01629    4.09 5.08e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5348 on 461 degrees of freedom
## Multiple R-squared:  0.03502,    Adjusted R-squared:  0.03293
## F-statistic: 16.73 on 1 and 461 DF,  p-value: 5.083e-05
```

Exercise 6: Use residual plots to evaluate whether the conditions of least squares regression are reasonable. Provide plots and comments for each one (see the Simple Regression Lab for a reminder of how to make these).

```
plot(m_bty$residuals ~ evals$bty_avg)
abline(h = 0, lty = 3)
```



```
hist(m_bty$residuals)
```

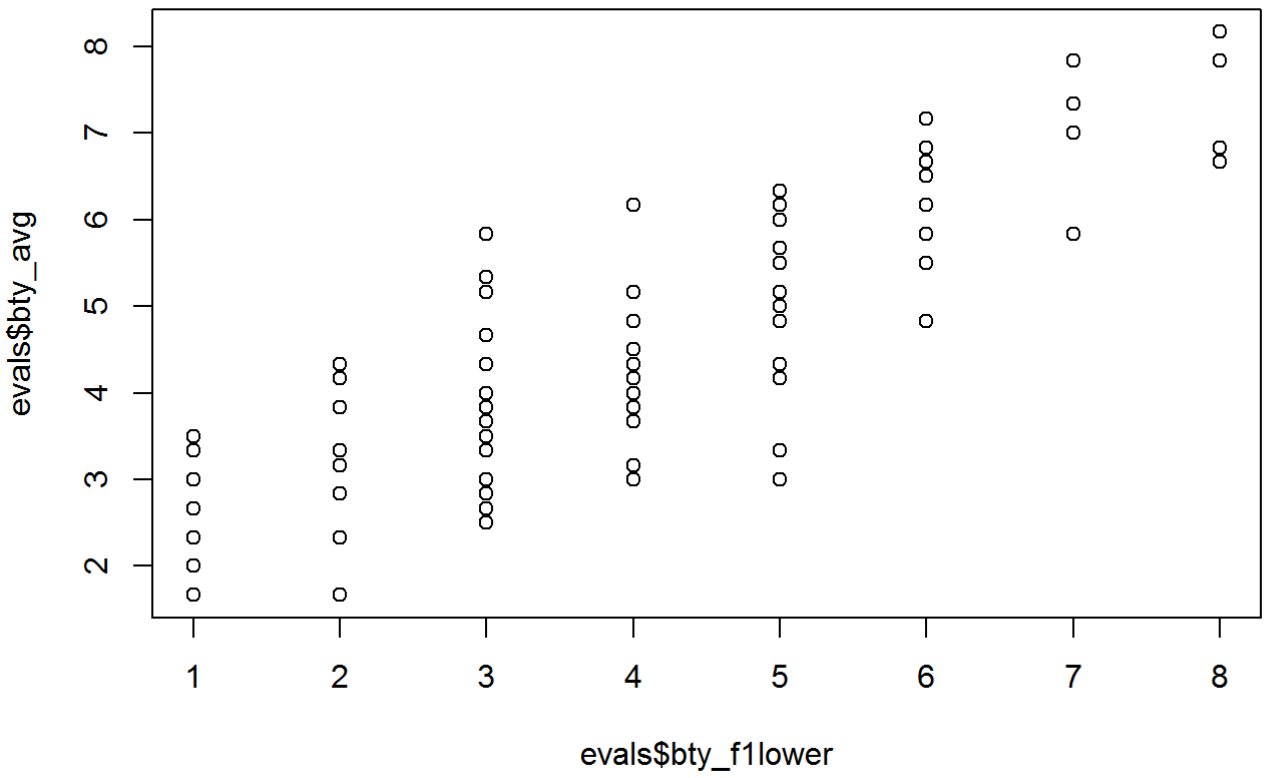


Linearity: The scatterplot shows linear relationship, and the residual plot shows no patten, and all the data points are evenly distributed on both side of the 0 line.

Nearly normal residuals: The histogram of the residual shows unimodel and bell shaped distribution that is quite symmetric. The normal probability plot also indicate normal distribution of residuals.

Constant variability: The variability of residuals around the 0 line is roughly constant. No pattern or fan shape observed.

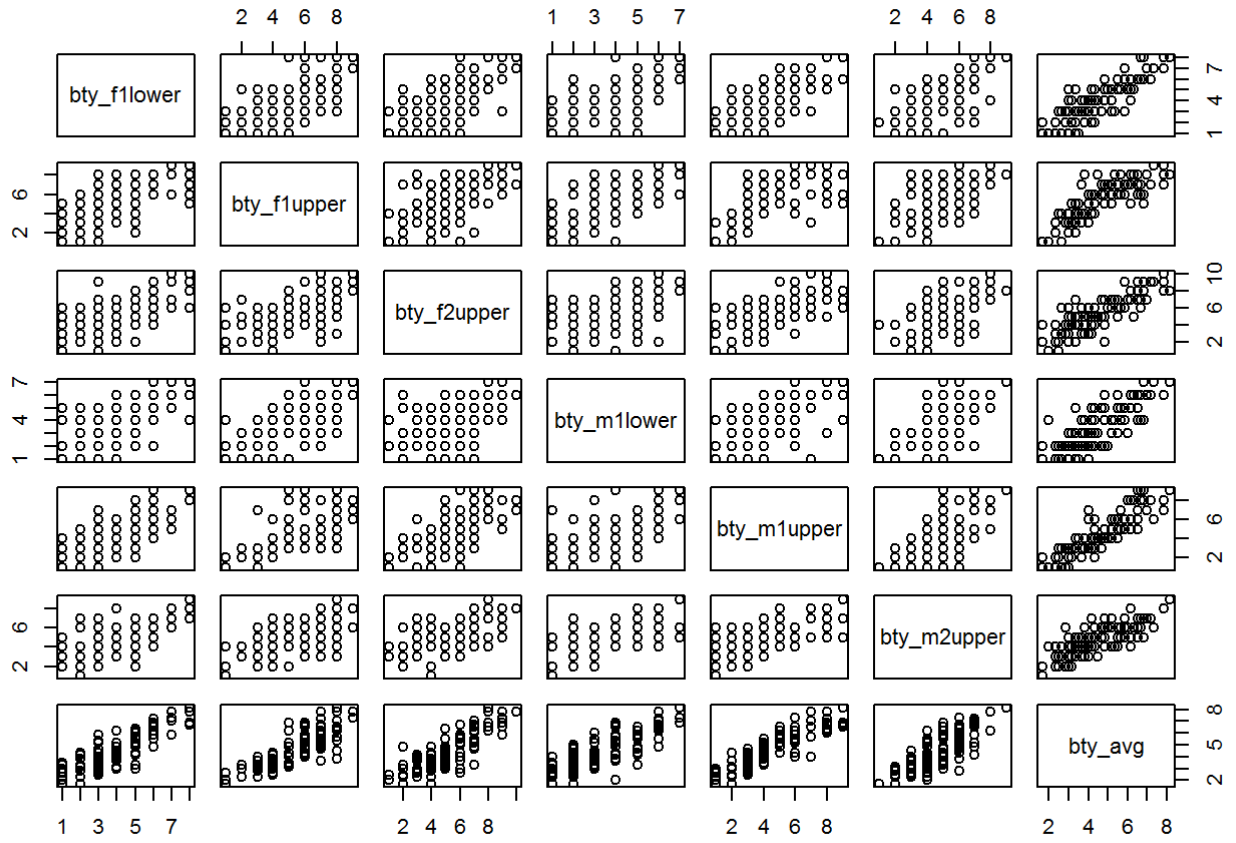
```
plot(evals$bty_avg ~ evals$bty_f1lower)
```




```
cor(evals$bty_avg, evals$bty_f1lower)
```

```
## [1] 0.8439112
```

```
plot(evals[,13:19])
```

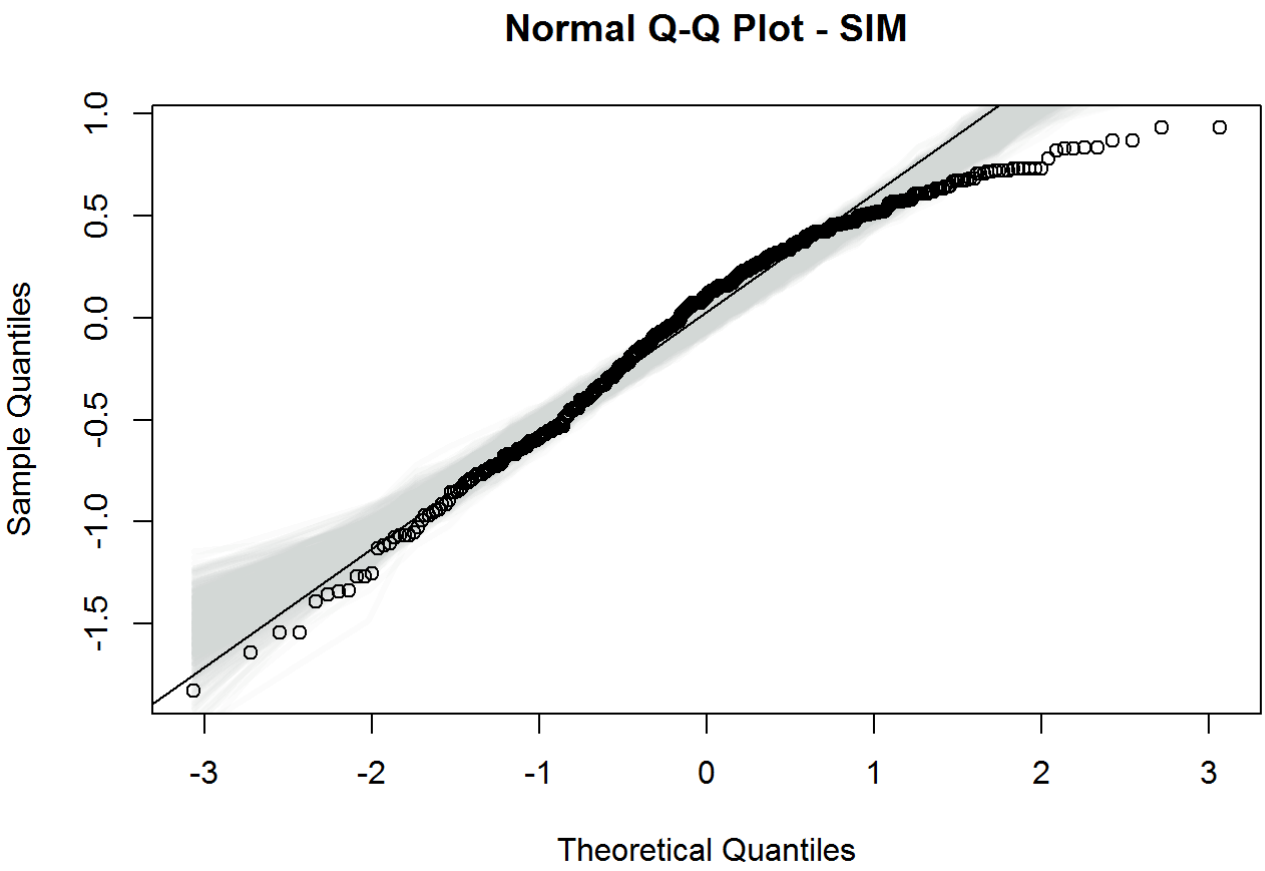



```
m_bty_gen <- lm(evals$score ~ evals$bty_avg + evals$gender, data = evals)
summary(m_bty_gen)
```

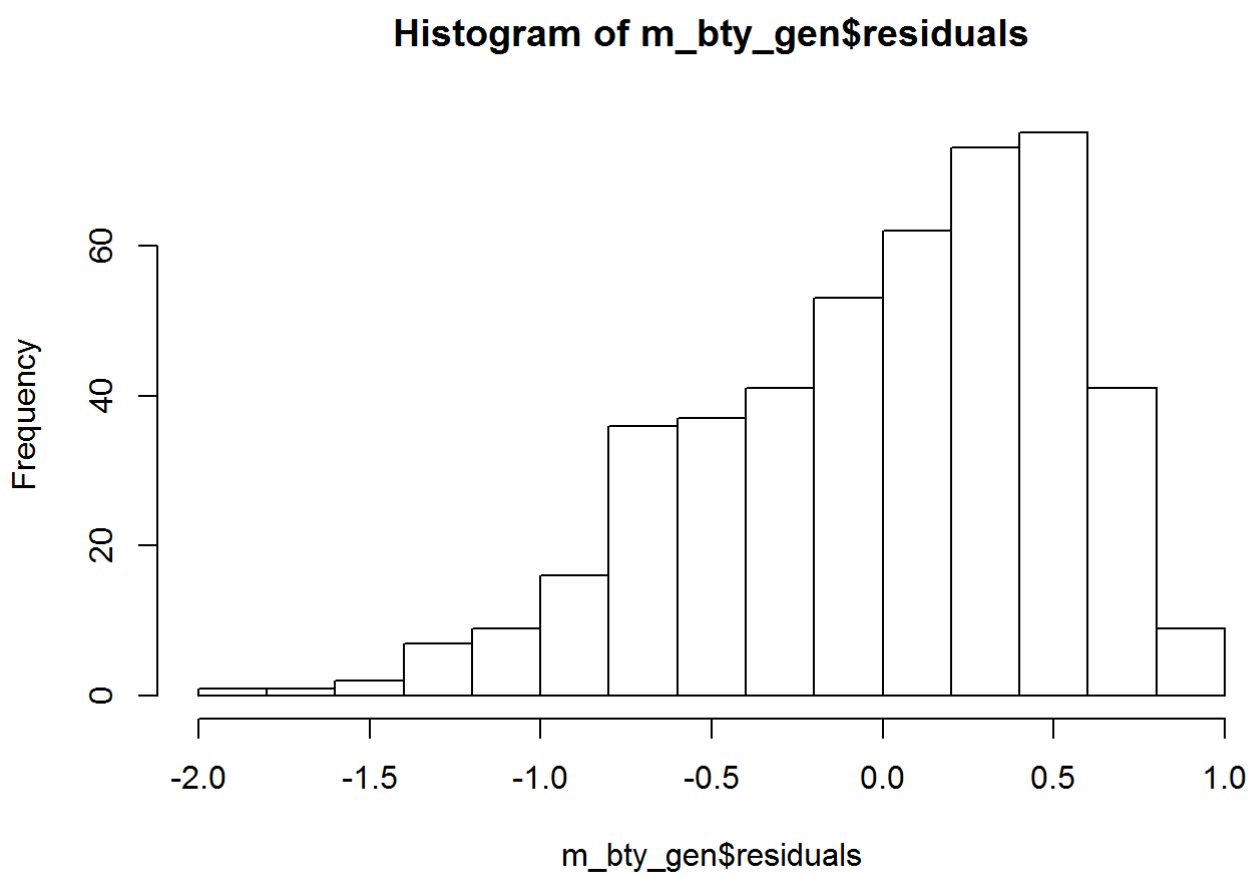
```
##
## Call:
## lm(formula = evals$score ~ evals$bty_avg + evals$gender, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8305 -0.3625  0.1055  0.4213  0.9314
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.74734    0.08466  44.266  < 2e-16 ***
## evals$bty_avg     0.07416    0.01625   4.563 6.48e-06 ***
## evals$gendermale  0.17239    0.05022   3.433 0.000652 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5287 on 460 degrees of freedom
## Multiple R-squared:  0.05912,    Adjusted R-squared:  0.05503
## F-statistic: 14.45 on 2 and 460 DF,  p-value: 8.177e-07
```

Exercise 7: P-values and parameter estimates should only be trusted if the conditions for the regression are reasonable. Verify that the conditions for this model are reasonable using diagnostic plots.

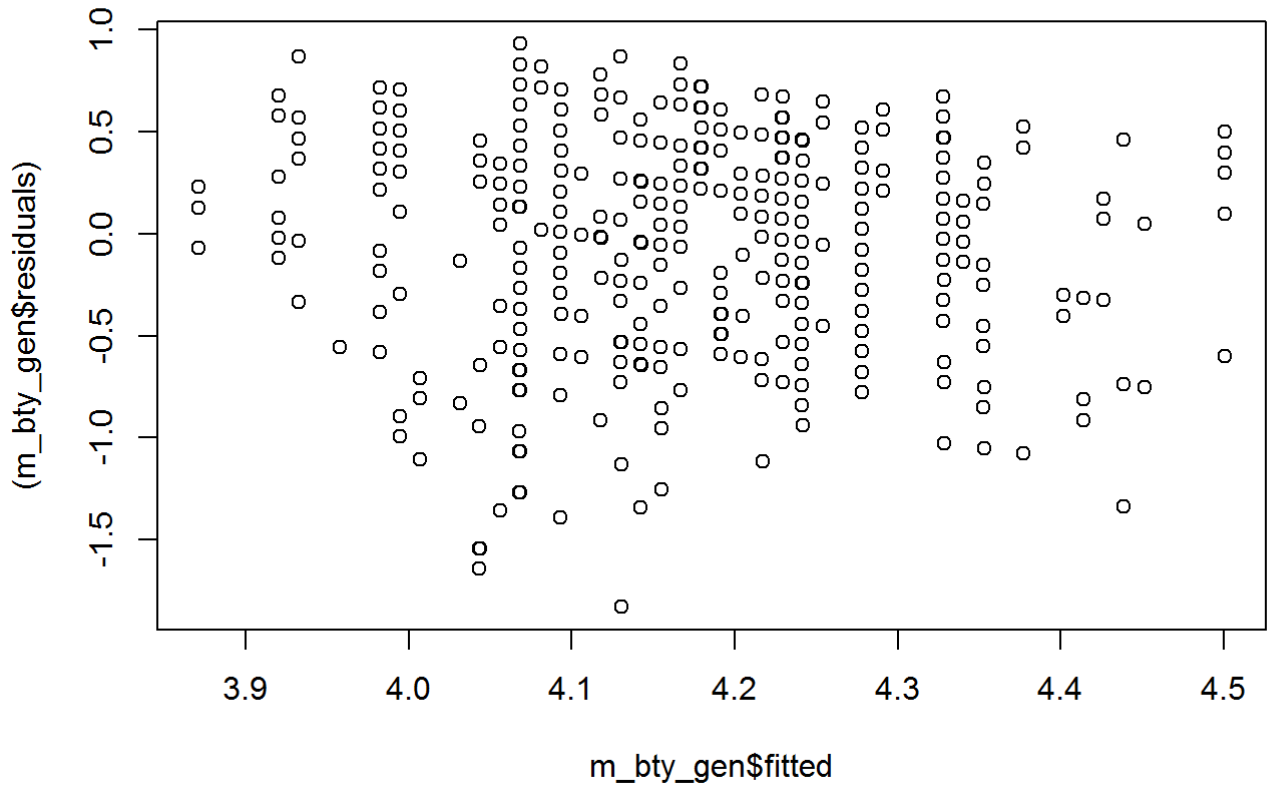
```
#1. residuals are nearly normal (primary concern relates to residuals that are outliers)
qqnormSim(m_bty_gen$residuals)
qqline(m_bty_gen$residuals)
```



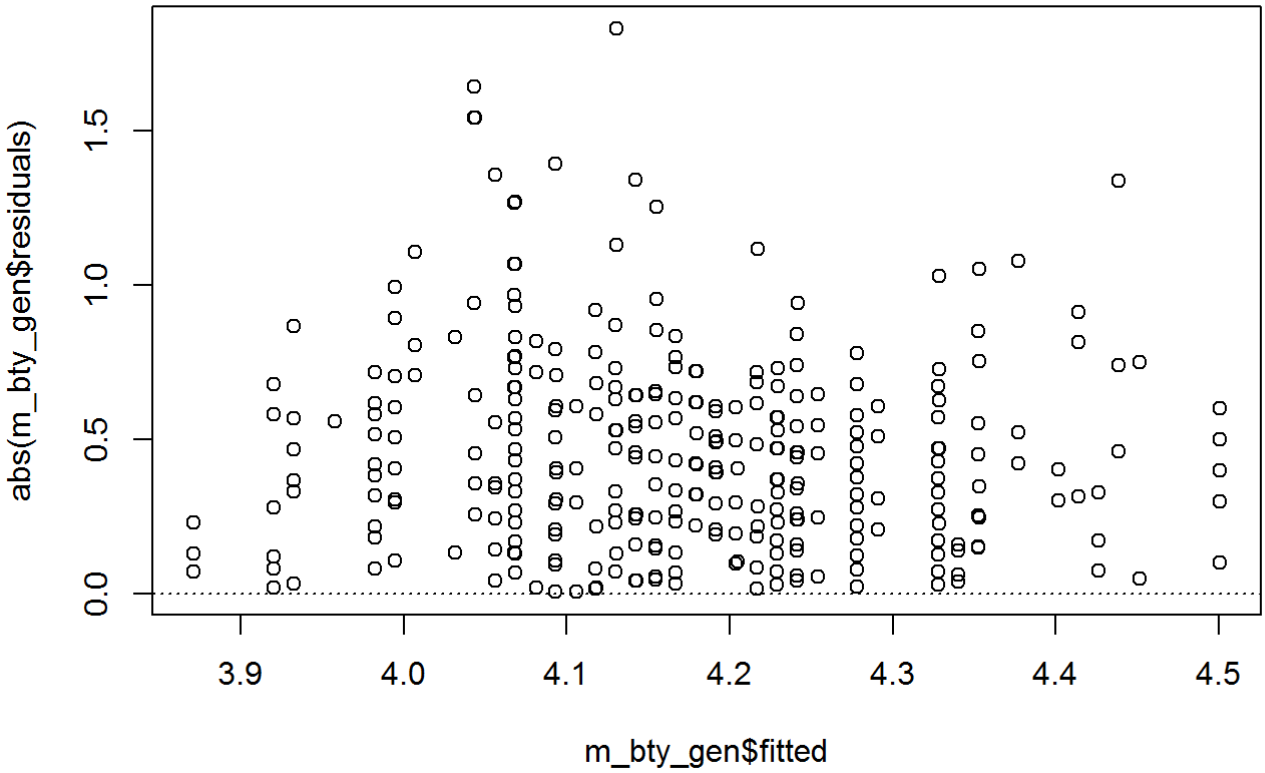

```
hist(m_bty_gen$residuals)
```



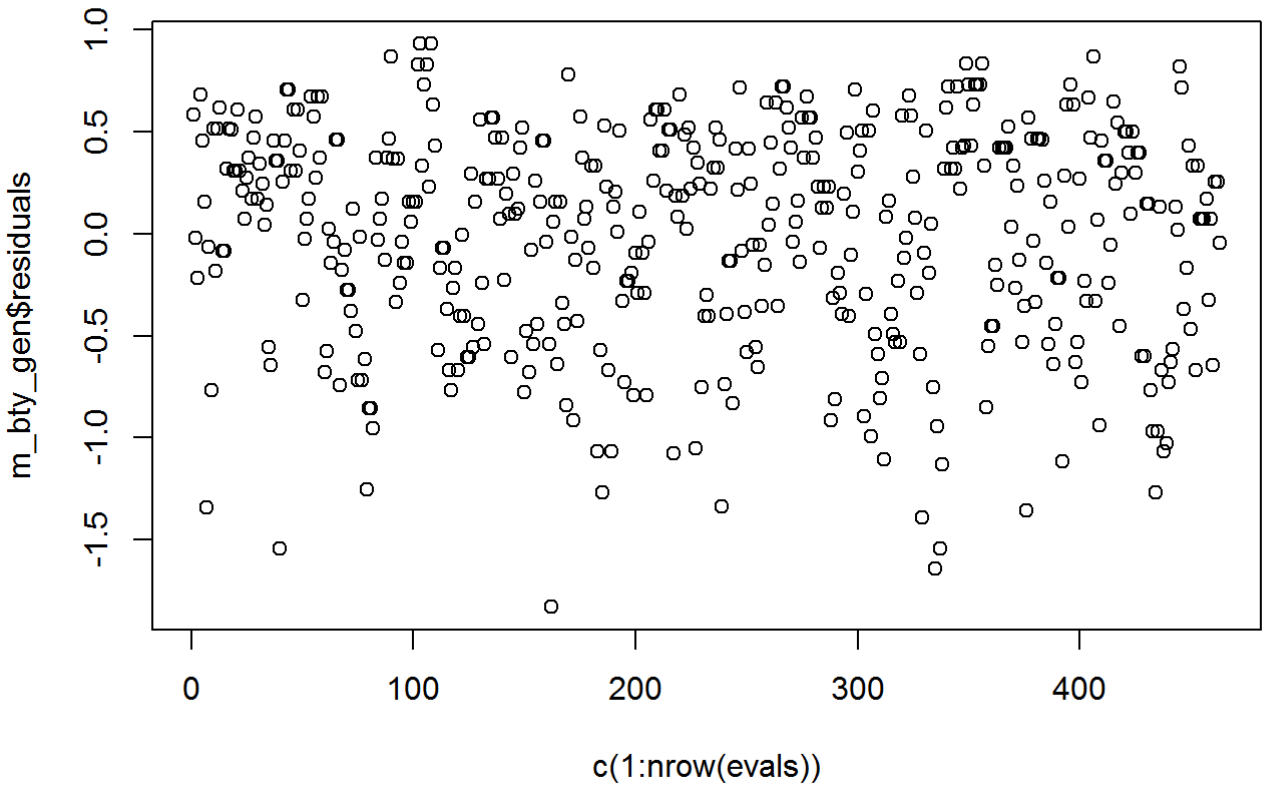

```
# The qqplot is bended which means the distribution of residual is not very normal. Histogram is
 skewed which tells us the same thing.

#2. residuals have constant variability
plot((m_bty_gen$residuals) ~ m_bty_gen$fitted)
```



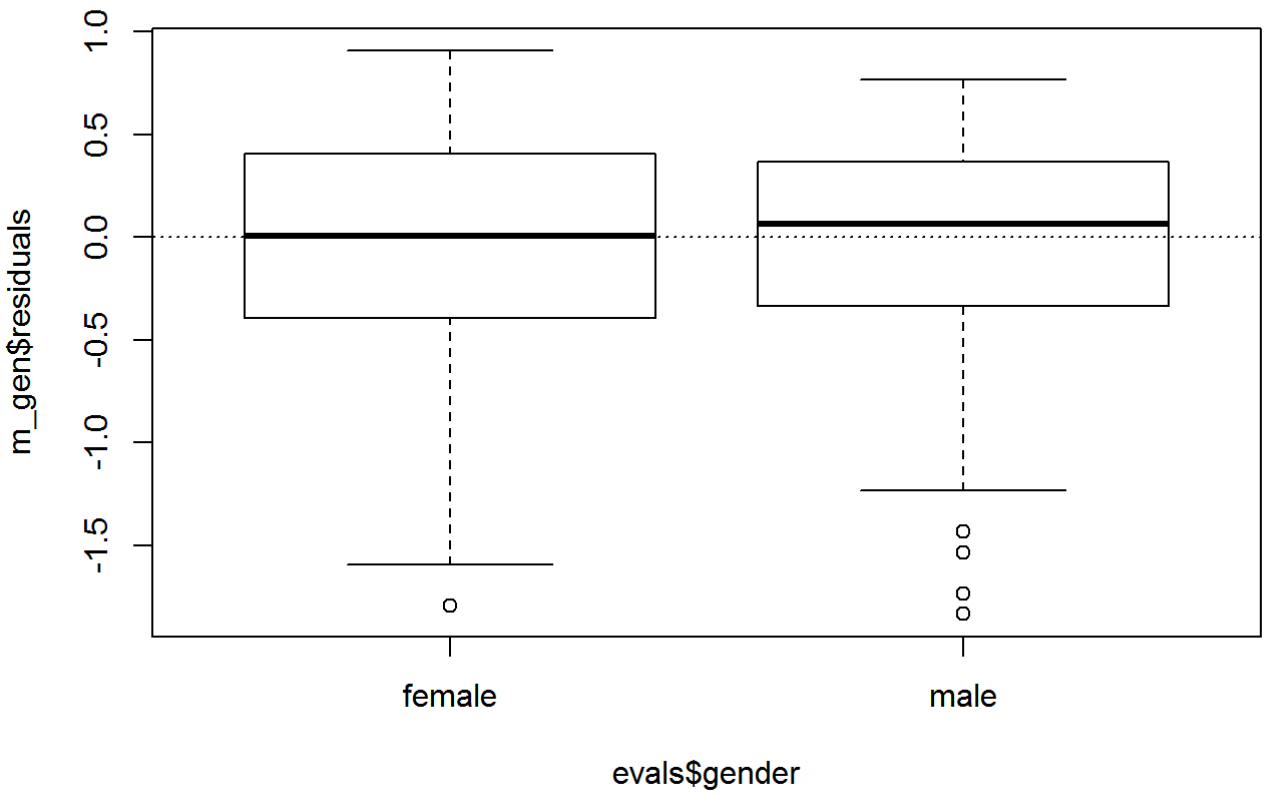

```
plot(abs(m_bty_gen$residuals) ~ m_bty_gen$fitted)
abline(h = 0, lty = 3)
```



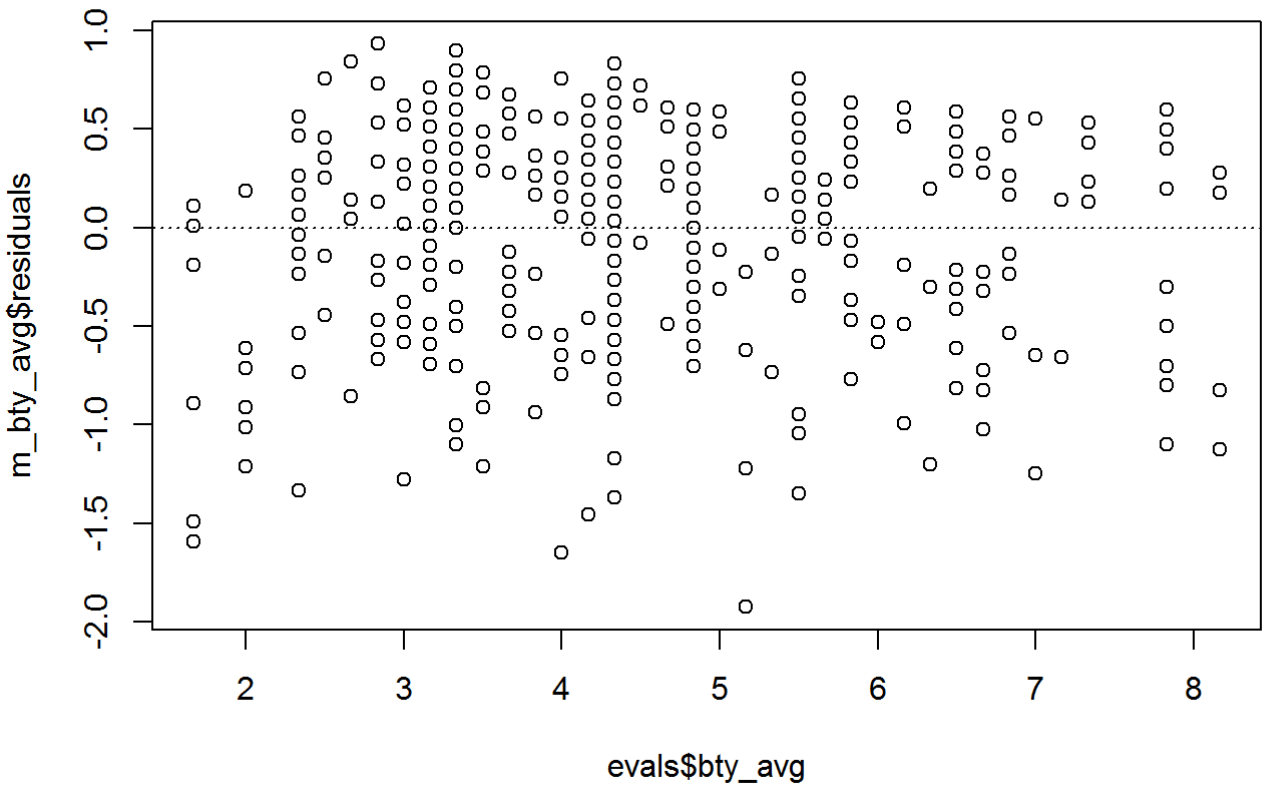

```
#The scatter plot shows no patterns of the residuals, most of residuals are closed to 0, which m
eans they are quite constant.


#3. residuals are independent
plot(m_bty_gen$residuals ~ c(1:nrow(evals)))
```




```
# The values of residuals does not appear to affect each other. They are pretty random through o
ut.

#4. each variable is linearly related to the outcome
m_gen <- lm(evals$score ~ evals$gender)
plot(m_gen$residuals ~ evals$gender)
abline(h = 0, lty = 3)
```




```
m_bty_avg <- lm(evals$score ~ evals$bty_avg)
plot(m_bty_avg$residuals ~ evals$bty_avg)
abline(h = 0, lty = 3)
```




```
#It seems like gender dose affects the evaluation score based on the boxplot. In addition, there
 is no pattern observed in the residual scatter plot of bty_avg, and all the data points are eve
nly distributed on both side of the 0 line. So the relationship is linear
```

Exercise 8: Is bty_avg still a significant predictor of score? Has the addition of gender to the model changed the parameter estimate for bty_avg?

Yes, because the p-value is 6.48e-06 which is closed to 0, so it is still a significant predictor of score. Yes, since the the slope and standard error for bty_avg are both changed.

Exercise 9: What is the equation of the line corresponding to males? (Hint: For males, the parameter estimate is multiplied by 1.) For two professors who received the same beauty rating, which gender tends to have the higher course evaluation score?

y = 3.74734 + 0.07416 * X1 + 0.17239 * 1 For two professors who received the same beauty rating, male gender tende to have the higher course evaluation score.

Exercise 10: Create a new model called m_bty_rank with gender removed and rank added in. How does R appear to handle categorical variables that have more than two levels? Note that the rank variable has three levels: teaching, tenure track, tenured.

R will use teaching as a reference. Only two predictors shows up which are tenure track and tenured

```
m_bty_rank <- lm(evals$score ~ evals$bty_avg + evals$rank)
summary(m_bty_rank)
```

```
##
## Call:
## lm(formula = evals$score ~ evals$bty_avg + evals$rank)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8713 -0.3642  0.1489  0.4103  0.9525
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              3.98155    0.09078  43.860  < 2e-16 ***
## evals$bty_avg            0.06783    0.01655   4.098 4.92e-05 ***
## evals$ranktenure track -0.16070    0.07395  -2.173   0.0303 *
## evals$ranktenured      -0.12623    0.06266  -2.014   0.0445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5328 on 459 degrees of freedom
## Multiple R-squared:  0.04652,    Adjusted R-squared:  0.04029
## F-statistic: 7.465 on 3 and 459 DF,  p-value: 6.88e-05
```

Exercise 11: Which variable would you expect to have the highest p-value in this model? Why? Hint: Think about which variable would you expect to not have any association with the professor score.

“cls_profs” (number of professors teaching sections in course in sample: single, multiple) has the highest p-value in this model. The p-value is 0.77806.

```
m_full <- lm(score ~ rank + ethnicity + gender + language + age + cls_perc_eval
             + cls_students + cls_level + cls_profs + cls_credits + bty_avg
             + pic_outfit + pic_color, data = evals)
summary(m_full)
```

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_students + cls_level + cls_profs + cls_credits +
##     bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77397 -0.32432  0.09067  0.35183  0.95036
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.0952141  0.2905277  14.096  < 2e-16 ***
## ranktenure track       -0.1475932  0.0820671  -1.798  0.07278 .
## ranktenured            -0.0973378  0.0663296  -1.467  0.14295
## ethnicitynot minority   0.1234929  0.0786273   1.571  0.11698
## gendermale              0.2109481  0.0518230   4.071 5.54e-05 ***
## languagenon-english    -0.2298112  0.1113754  -2.063  0.03965 *
## age                    -0.0090072  0.0031359  -2.872  0.00427 **
## cls_perc_eval           0.0053272  0.0015393   3.461  0.00059 ***
## cls_students            0.0004546  0.0003774   1.205  0.22896
## cls_levelupper          0.0605140  0.0575617   1.051  0.29369
## cls_profssingle        -0.0146619  0.0519885  -0.282  0.77806
## cls_creditsone credit   0.5020432  0.1159388   4.330 1.84e-05 ***
## bty_avg                 0.0400333  0.0175064   2.287  0.02267 *
## pic_outfitnot formal   -0.1126817  0.0738800  -1.525  0.12792
## pic_colorcolor         -0.2172630  0.0715021  -3.039  0.00252 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.498 on 448 degrees of freedom
## Multiple R-squared:  0.1871, Adjusted R-squared:  0.1617
## F-statistic: 7.366 on 14 and 448 DF,  p-value: 6.552e-14
```

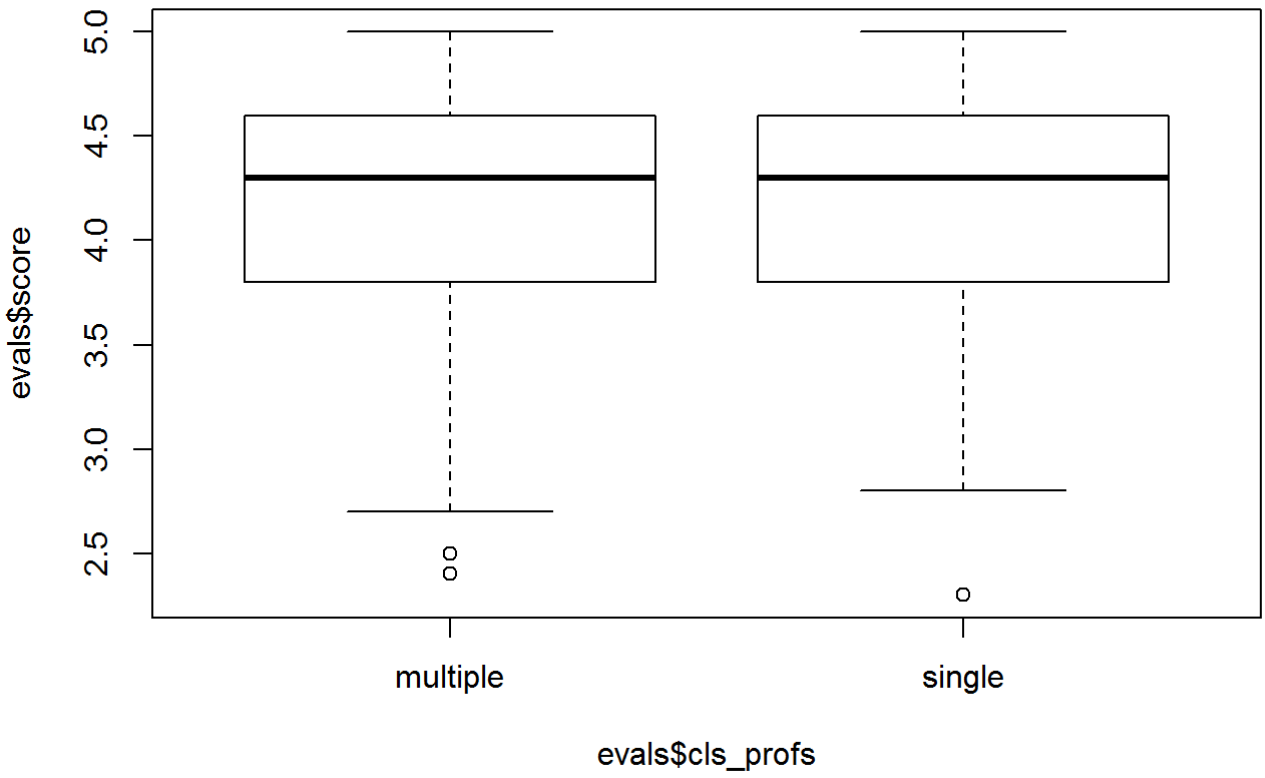
Exercise 12: Check your suspicions from the previous exercise. Include the model output in your response.

If we only build the linear model between evaluation score and cls_profs, the p-value (0.585) still be over 0.05. The boxplot shows the distribution between single and multiple professors teaching in the course are pretty much the same.

```
m_bty_cls_profs <- lm(evals$score ~ evals$cls_profs)
summary(m_bty_cls_profs)
```

```
##
## Call:
## lm(formula = evals$score ~ evals$cls_profs)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8554 -0.3846  0.1154  0.4154  0.8446
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.18464    0.03111 134.493   <2e-16 ***
## evals$cls_profssingle -0.02923    0.05343  -0.547    0.585
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5443 on 461 degrees of freedom
## Multiple R-squared:  0.0006486,  Adjusted R-squared:  -0.001519
## F-statistic: 0.2992 on 1 and 461 DF,  p-value: 0.5847
```

```
plot(evals$score ~ evals$cls_profs)
```




```
evals$ethnicity
```

```
##   [1] minority     minority     minority     minority     not minority
##   [6] not minority not minority not minority not minority not minority
##  [11] not minority not minority not minority not minority not minority
##  [16] not minority not minority not minority not minority not minority
##  [21] not minority not minority not minority not minority not minority
##  [26] not minority not minority not minority not minority not minority
##  [31] not minority not minority not minority not minority not minority
##  [36] not minority not minority not minority not minority not minority
##  [41] not minority not minority not minority not minority not minority
##  [46] not minority not minority not minority not minority not minority
##  [51] not minority not minority not minority not minority not minority
##  [56] not minority not minority not minority not minority minority
##  [61] minority     minority     not minority not minority not minority
##  [66] not minority not minority not minority not minority not minority
##  [71] not minority not minority not minority not minority not minority
##  [76] not minority not minority not minority not minority not minority
##  [81] not minority not minority not minority not minority not minority
##  [86] not minority not minority not minority not minority not minority
##  [91] not minority not minority not minority not minority not minority
##  [96] not minority not minority not minority not minority not minority
## [101] not minority not minority not minority not minority not minority
## [106] not minority not minority not minority not minority not minority
## [111] not minority not minority not minority not minority not minority
## [116] not minority not minority not minority not minority not minority
## [121] not minority not minority not minority not minority not minority
## [126] not minority minority     not minority not minority not minority
## [131] not minority not minority not minority not minority not minority
## [136] not minority not minority not minority not minority not minority
## [141] not minority not minority not minority not minority not minority
## [146] not minority minority     minority     minority     minority
## [151] minority     minority     minority     not minority not minority
## [156] not minority not minority not minority not minority not minority
## [161] not minority not minority not minority not minority not minority
## [166] not minority not minority not minority not minority not minority
## [171] not minority not minority not minority not minority not minority
## [176] not minority not minority minority     minority     minority
## [181] minority     minority     minority     minority     minority
## [186] minority     minority     minority     minority     minority
## [191] not minority not minority not minority minority     minority
## [196] minority     minority     minority     minority     minority
## [201] minority     minority     minority     minority     minority
## [206] not minority not minority not minority not minority not minority
## [211] not minority not minority not minority not minority not minority
## [216] not minority not minority not minority not minority not minority
## [221] not minority not minority not minority not minority not minority
## [226] not minority not minority not minority not minority not minority
## [231] not minority not minority not minority not minority not minority
## [236] not minority not minority not minority not minority not minority
## [241] not minority not minority not minority not minority not minority
## [246] not minority not minority not minority not minority not minority
## [251] not minority not minority not minority not minority not minority
## [256] not minority not minority not minority not minority not minority
## [261] not minority not minority not minority not minority not minority
```

```
## [266] not minority not minority not minority not minority not minority
## [271] not minority not minority not minority not minority not minority
## [276] not minority not minority not minority not minority not minority
## [281] not minority not minority not minority not minority not minority
## [286] not minority not minority not minority not minority not minority
## [291] not minority not minority not minority not minority not minority
## [296] not minority not minority not minority not minority not minority
## [301] not minority not minority not minority not minority not minority
## [306] not minority not minority not minority not minority not minority
## [311] not minority not minority not minority not minority not minority
## [316] not minority not minority not minority not minority not minority
## [321] not minority not minority not minority not minority not minority
## [326] not minority not minority not minority not minority not minority
## [331] not minority not minority not minority not minority not minority
## [336] not minority not minority not minority not minority not minority
## [341] not minority not minority not minority not minority not minority
## [346] not minority not minority minority     minority     minority
## [351] minority     minority     minority     minority     minority
## [356] minority     minority     not minority not minority not minority
## [361] not minority not minority not minority not minority not minority
## [366] not minority not minority not minority not minority not minority
## [371] not minority not minority not minority not minority minority
## [376] minority     not minority not minority not minority not minority
## [381] not minority not minority not minority not minority not minority
## [386] not minority not minority not minority not minority not minority
## [391] not minority not minority not minority not minority not minority
## [396] not minority not minority not minority not minority not minority
## [401] not minority not minority not minority not minority not minority
## [406] not minority not minority not minority not minority not minority
## [411] not minority not minority not minority minority     minority
## [416] minority     minority     minority     not minority not minority
## [421] not minority not minority not minority not minority not minority
## [426] not minority not minority not minority not minority not minority
## [431] not minority not minority not minority not minority not minority
## [436] not minority not minority not minority minority     minority
## [441] minority     not minority not minority not minority not minority
## [446] not minority not minority not minority not minority not minority
## [451] not minority not minority not minority not minority not minority
## [456] not minority not minority not minority not minority minority
## [461] minority     minority     minority
## Levels: minority not minority
```

Exercise 13: Interpret the coefficient associated with the ethnicity variable.

The coefficient associated with ethnicity is 0.1235, which means when all the other variables are the same, the prefessot who is not a minority will have average increase of 0.1235 of the evaluation score.

Exercise 14: Drop the variable with the highest p-value and re-fit the model. Did the coefficients and significance of the other explanatory variables change? (One of the things that makes multiple regression interesting is that coefficient estimates depend on the other variables that are included in the model.) If not, what does this say about whether or not the dropped variable was collinear with the other explanatory variables?

Yes, both the coefficients and significance have changed. If not, it tell us the variable that is removed is not collinear with other explanatory variables.

```
m_full <- lm(score ~ rank + ethnicity + gender + language + age + cls_perc_eval
             + cls_students + cls_level + cls_credits + bty_avg
             + pic_outfit + pic_color, data = evals)
summary(m_full)
```

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_students + cls_level + cls_credits +
##     bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7836 -0.3257  0.0859  0.3513  0.9551
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.0872523  0.2888562  14.150  < 2e-16 ***
## ranktenure track       -0.1476746  0.0819824  -1.801 0.072327 .
## ranktenured            -0.0973829  0.0662614  -1.470 0.142349
## ethnicitynot minority   0.1274458  0.0772887   1.649 0.099856 .
## gendermale              0.2101231  0.0516873   4.065 5.66e-05 ***
## languagenon-english    -0.2282894  0.1111305  -2.054 0.040530 *
## age                    -0.0089992  0.0031326  -2.873 0.004262 **
## cls_perc_eval           0.0052888  0.0015317   3.453 0.000607 ***
## cls_students            0.0004687  0.0003737   1.254 0.210384
## cls_levelupper          0.0606374  0.0575010   1.055 0.292200
## cls_creditsone credit   0.5061196  0.1149163   4.404 1.33e-05 ***
## bty_avg                 0.0398629  0.0174780   2.281 0.023032 *
## pic_outfitnot formal   -0.1083227  0.0721711  -1.501 0.134080
## pic_colorcolor         -0.2190527  0.0711469  -3.079 0.002205 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4974 on 449 degrees of freedom
## Multiple R-squared:  0.187,  Adjusted R-squared:  0.1634
## F-statistic: 7.943 on 13 and 449 DF,  p-value: 2.336e-14
```

Exercise 15: Using backward-selection and p-value as the selection criterion, determine the best model. You do not need to show all steps in your answer, just the output for the final model. Also, write out the linear model for predicting score based on the final model you settle on.

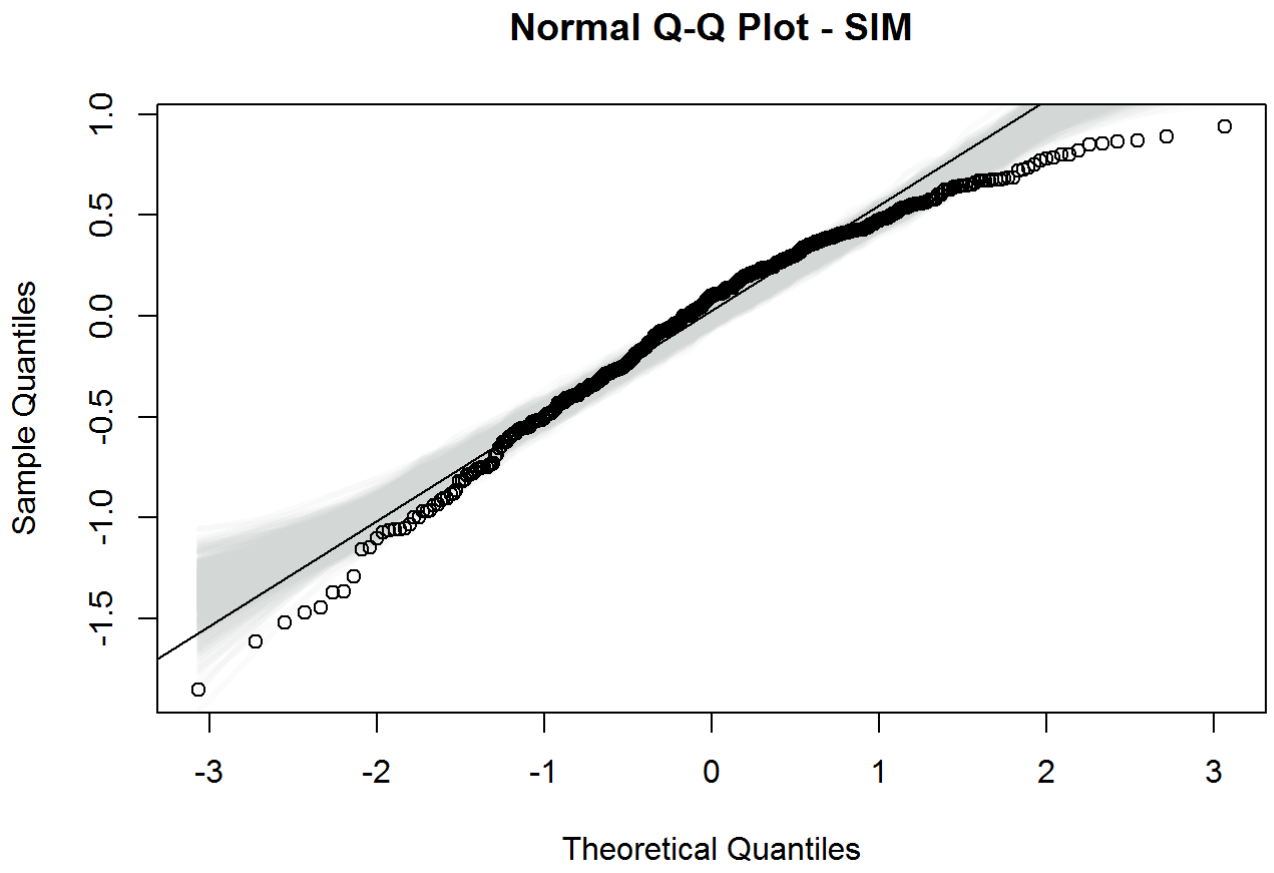
Y = 3.772 + 0.168 * ethnicity + 0.207 * gender -0.206 * languege -0.006 * age + 0.005 * cls_perc_eval + 0.505 * cls_credits + 0.051 * bty_avg - 0.191 * pic_color

```
m_backward <- lm(score ~ ethnicity + gender + language + age + cls_perc_eval + cls_credits + bty
_avg
             + pic_color, data = evals)
summary(m_backward)
```

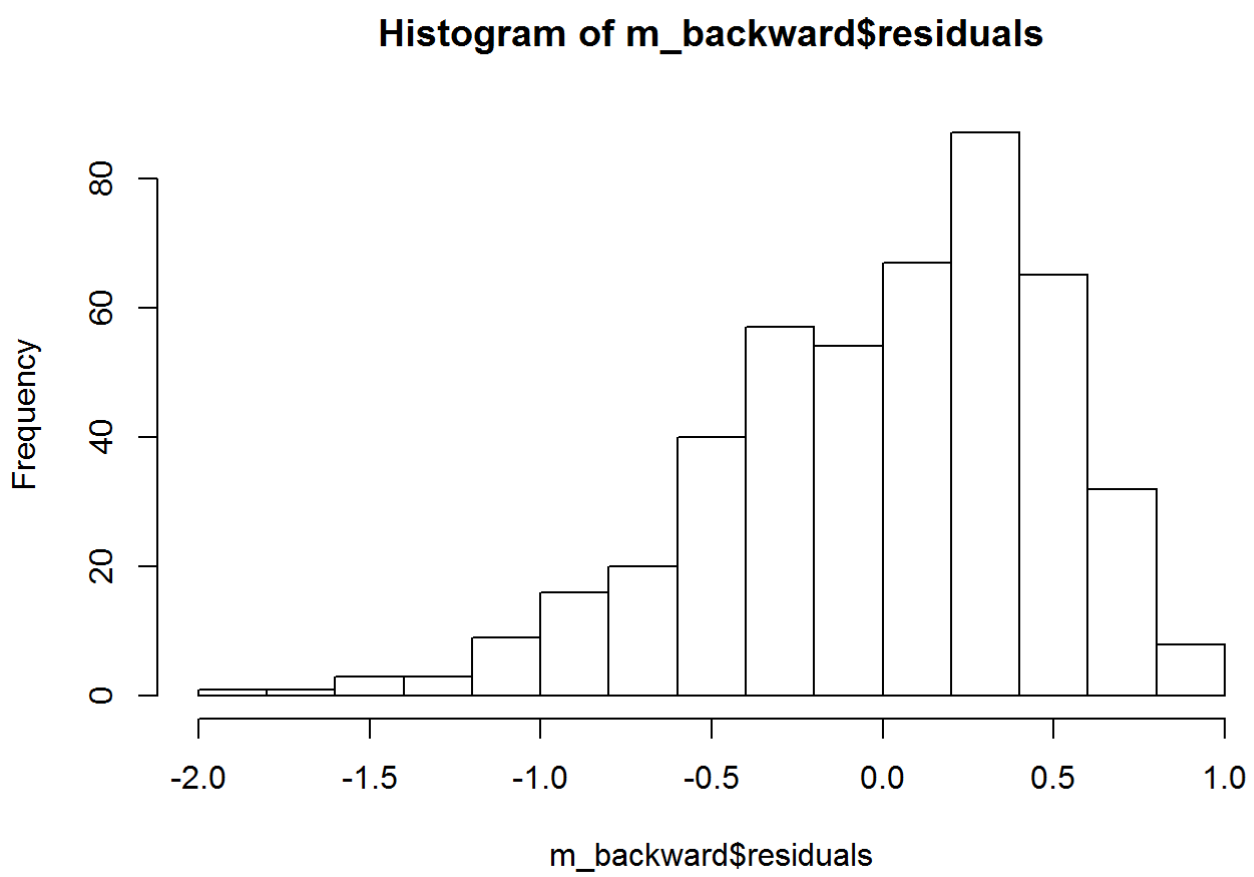
```
##
## Call:
## lm(formula = score ~ ethnicity + gender + language + age + cls_perc_eval +
##     cls_credits + bty_avg + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85320 -0.32394  0.09984  0.37930  0.93610
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.771922   0.232053  16.255  < 2e-16 ***
## ethnicitynot minority   0.167872   0.075275   2.230  0.02623 *
## gendermale              0.207112   0.050135   4.131 4.30e-05 ***
## languagenon-english    -0.206178   0.103639  -1.989  0.04726 *
## age                    -0.006046   0.002612  -2.315  0.02108 *
## cls_perc_eval           0.004656   0.001435   3.244  0.00127 **
## cls_creditsone credit   0.505306   0.104119   4.853 1.67e-06 ***
## bty_avg                 0.051069   0.016934   3.016  0.00271 **
## pic_colorcolor         -0.190579   0.067351  -2.830  0.00487 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4992 on 454 degrees of freedom
## Multiple R-squared:  0.1722, Adjusted R-squared:  0.1576
## F-statistic:  11.8 on 8 and 454 DF,  p-value: 2.58e-15
```

Exercise 16: Verify that the conditions for this model are reasonable using diagnostic plots.

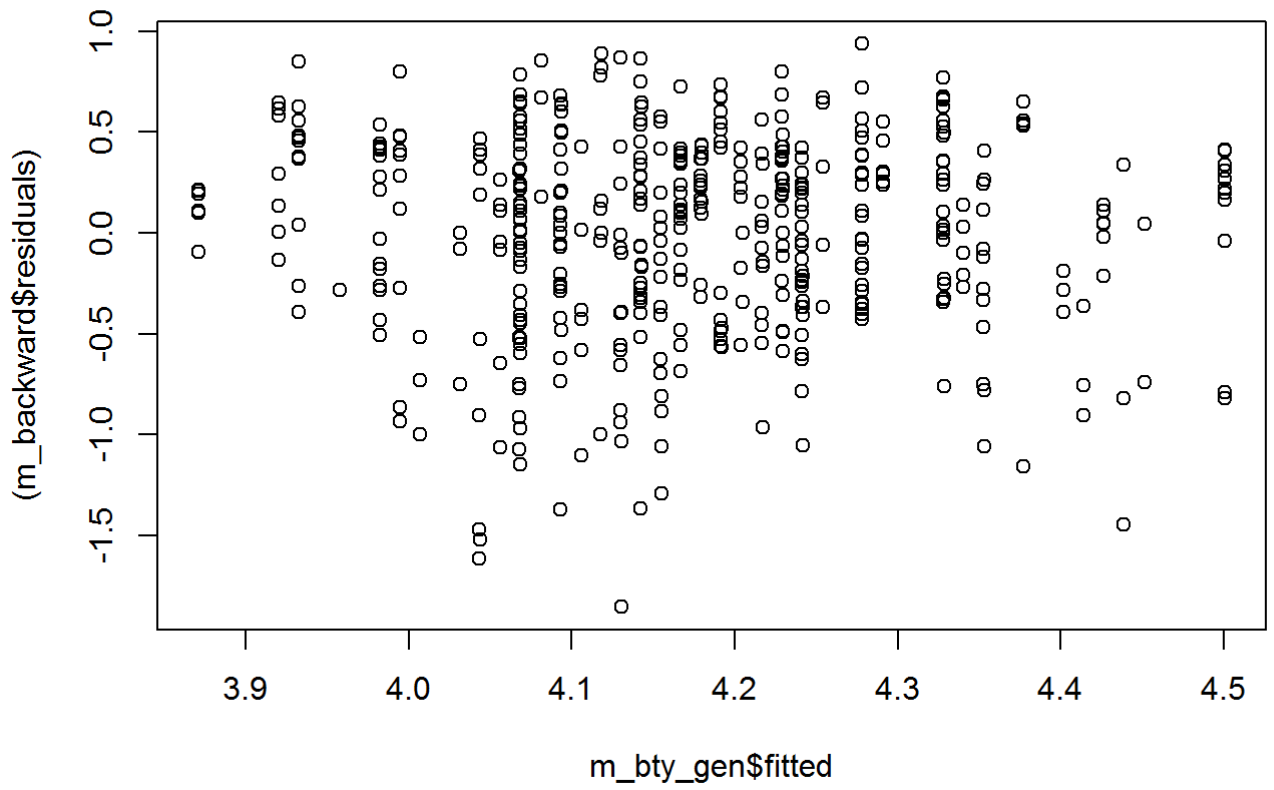
```
#1. residuals are nearly normal (primary concern relates to residuals that are outliers)
qqnormSim(m_backward$residuals)
qqline(m_backward$residuals)
```



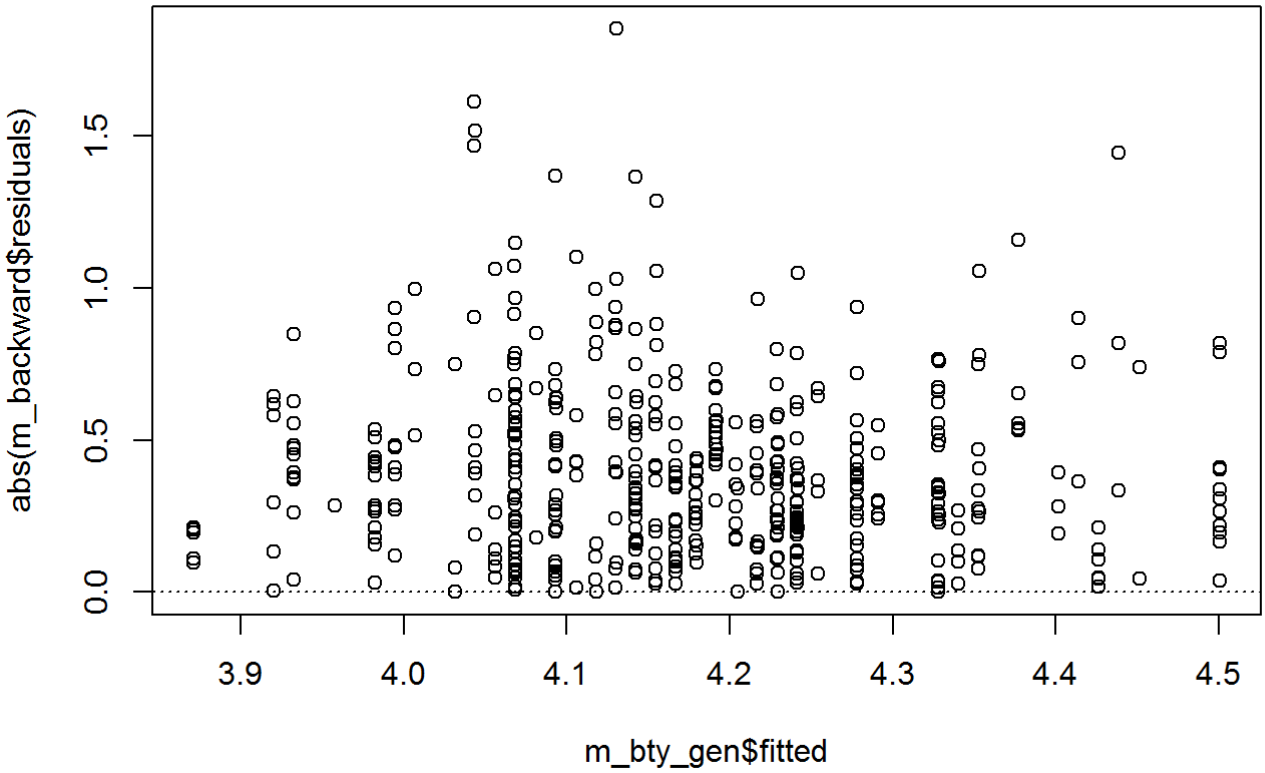

```
hist(m_backward$residuals)
```



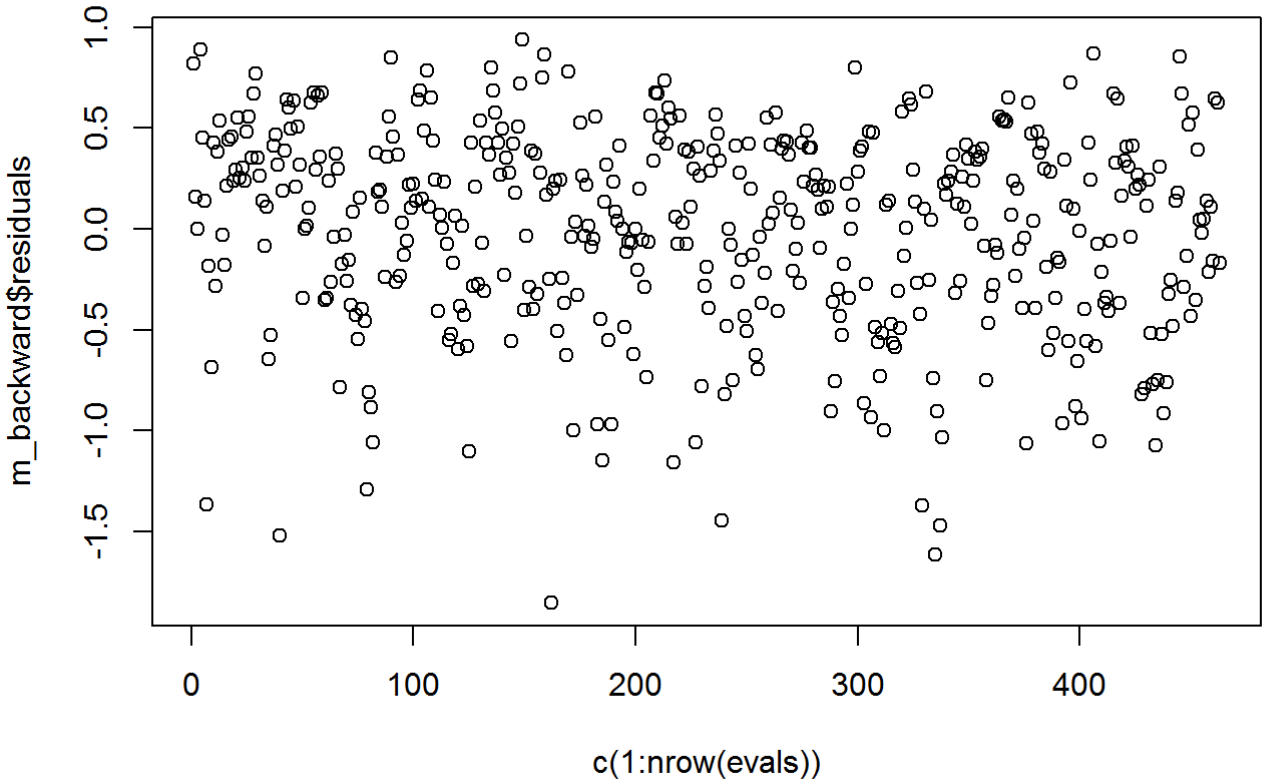

```
# The qqplot is bended which means the distribution of residual is not very normal. Histogram is
 skewed which tells us the same thing.


#2. residuals have constant variability
plot((m_backward$residuals) ~ m_bty_gen$fitted)
```



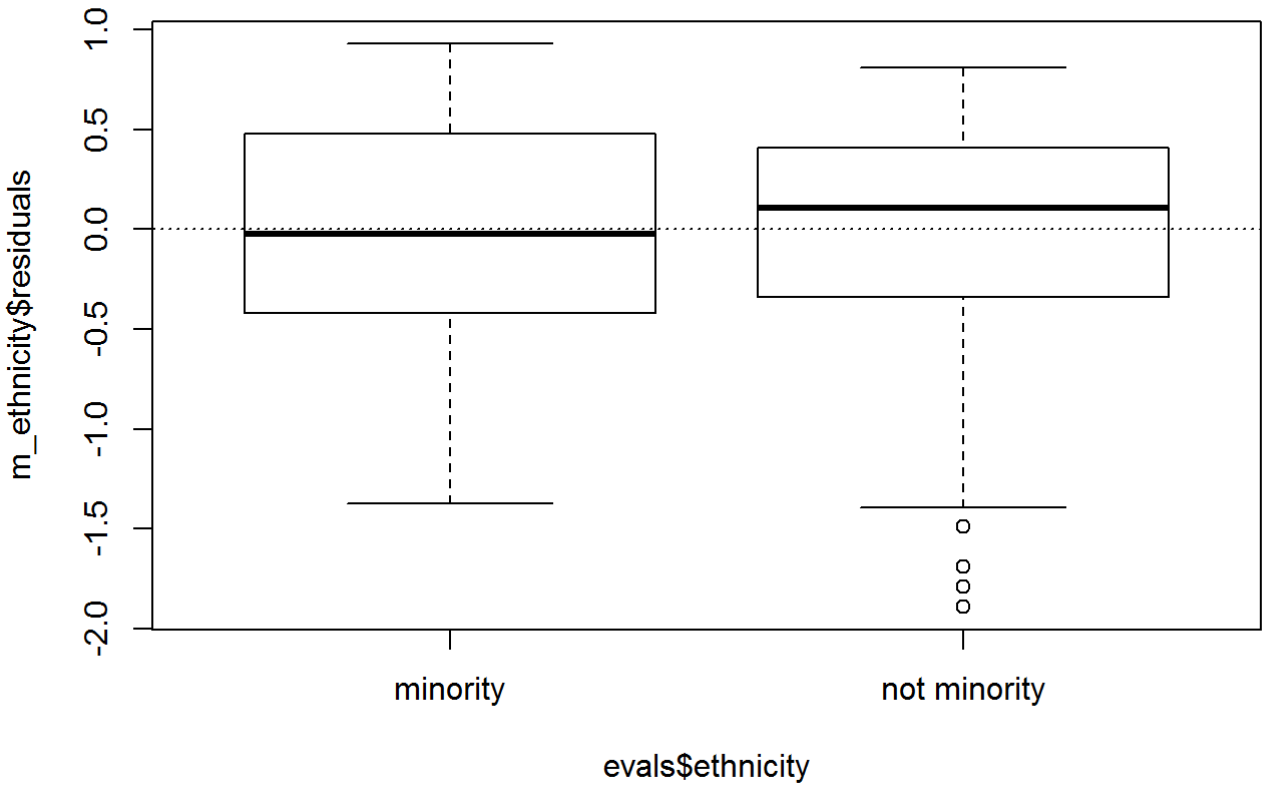

```
plot(abs(m_backward$residuals) ~ m_bty_gen$fitted)
abline(h = 0, lty = 3)
```



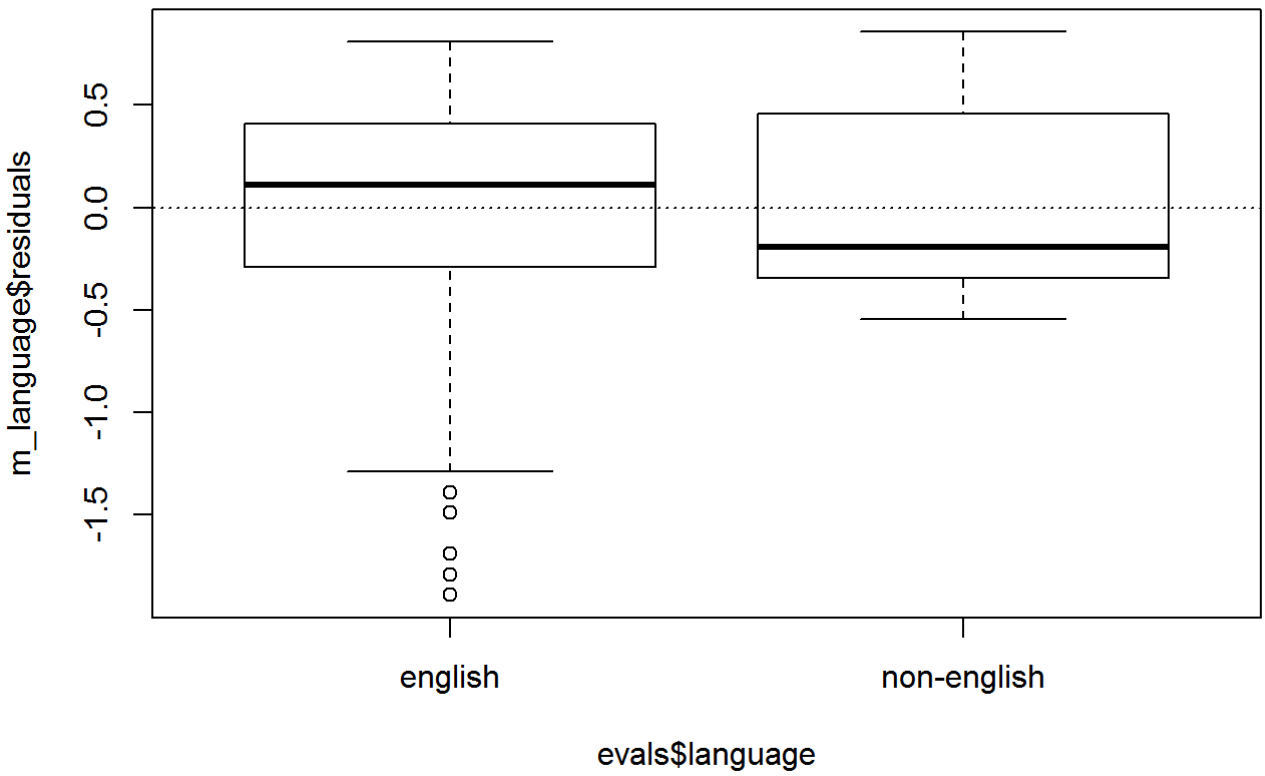

```
#The scatter plot shows no patterns of the residuals, most of residuals are closed to 0, which m
eans they are quite constant.

#3. residuals are independent
plot(m_backward$residuals ~ c(1:nrow(evals)))
```



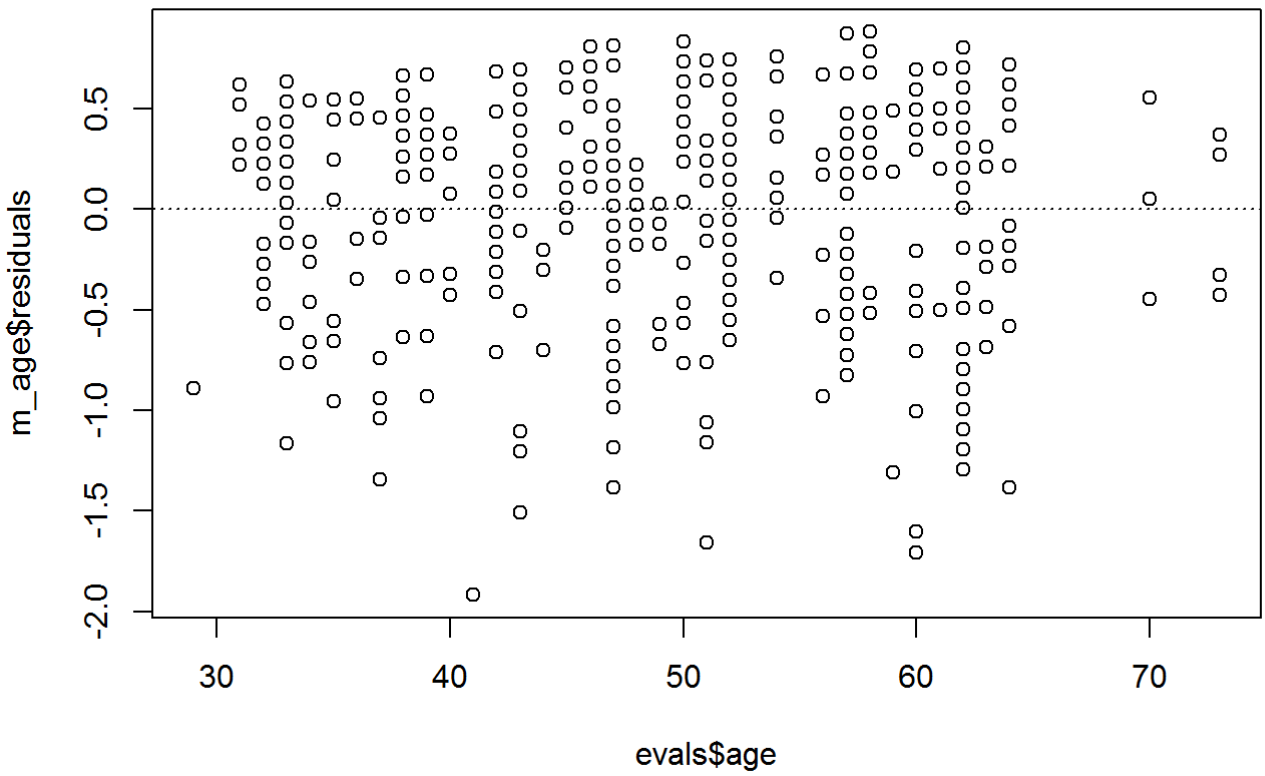

```
# The values of residuals does not appear to affect each other. They are pretty random through o
ut.


#4. each variable is linearly related to the outcome
m_ethnicity <- lm(evals$score ~ evals$ethnicity)
plot(m_ethnicity$residuals ~ evals$ethnicity)
abline(h = 0, lty = 3)
```



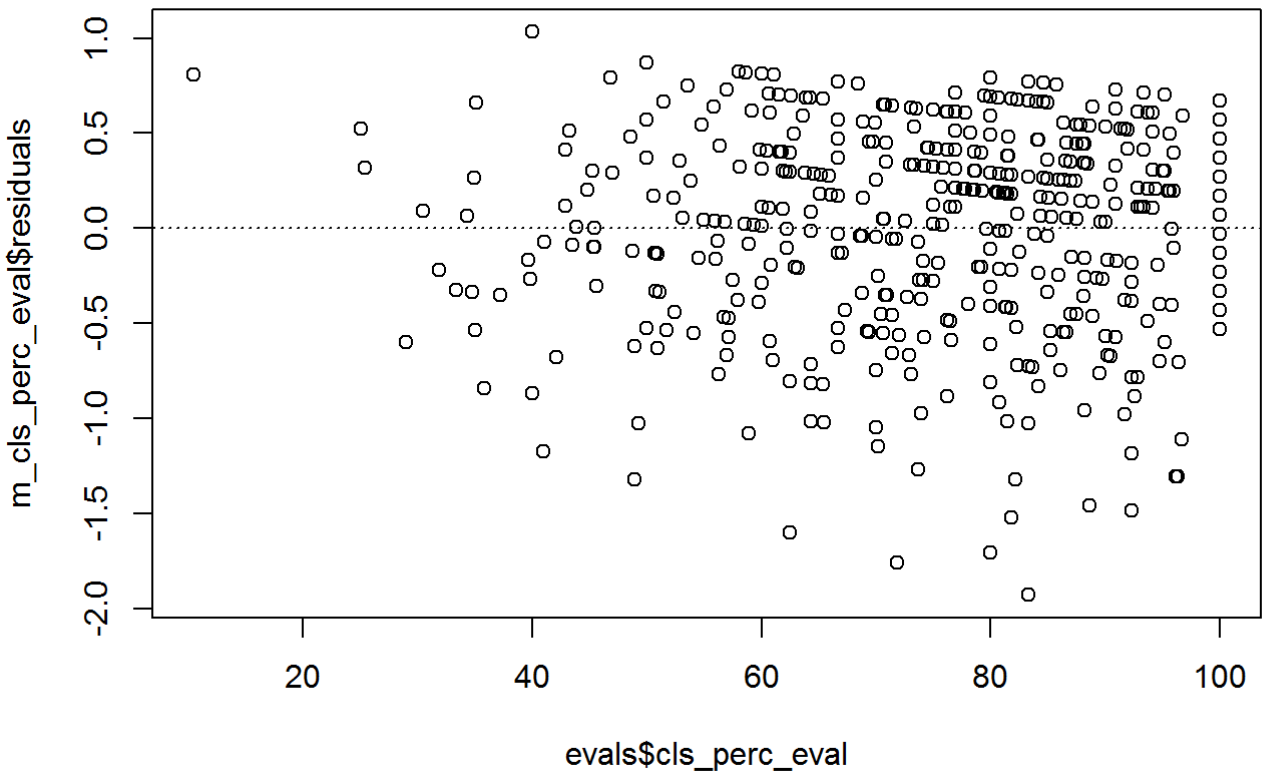

```
m_language <- lm(evals$score ~ evals$language)
plot(m_language$residuals ~ evals$language)
abline(h = 0, lty = 3)
```



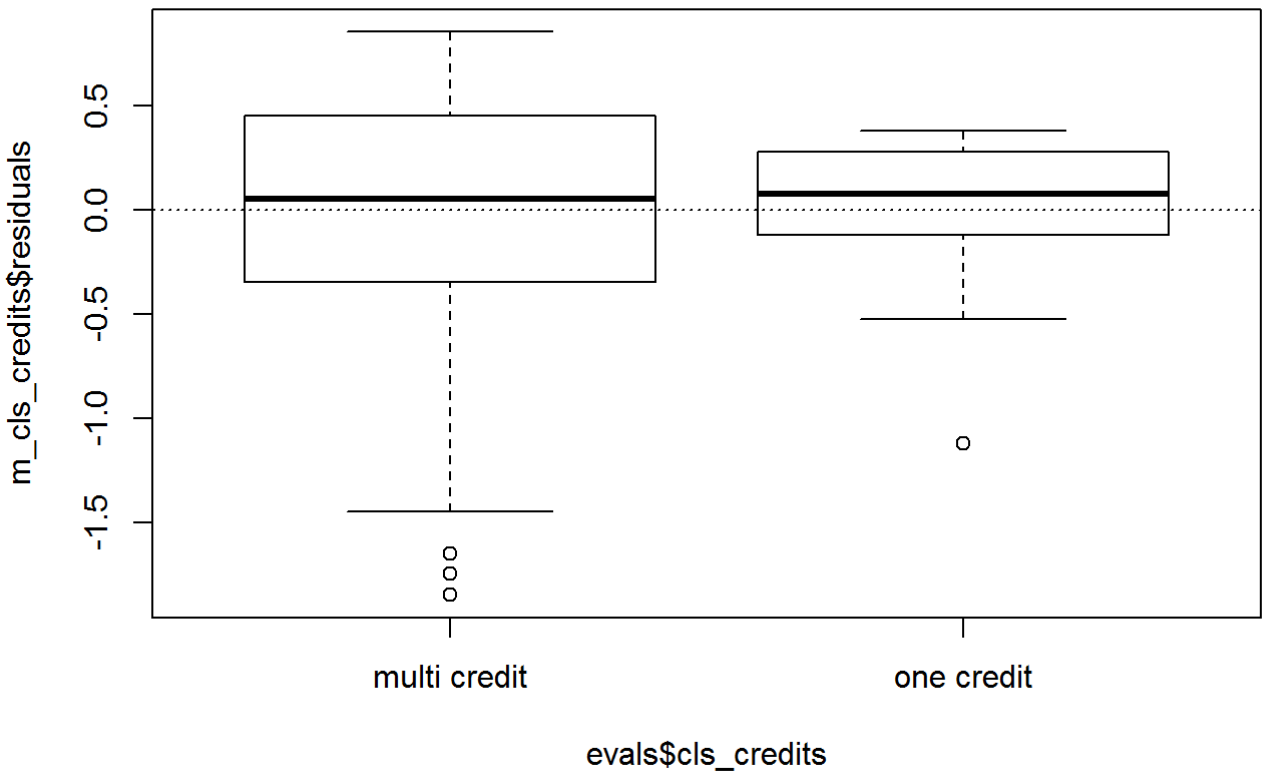

```
m_age <- lm(evals$score ~ evals$age)
plot(m_age$residuals ~ evals$age)
abline(h = 0, lty = 3)
```



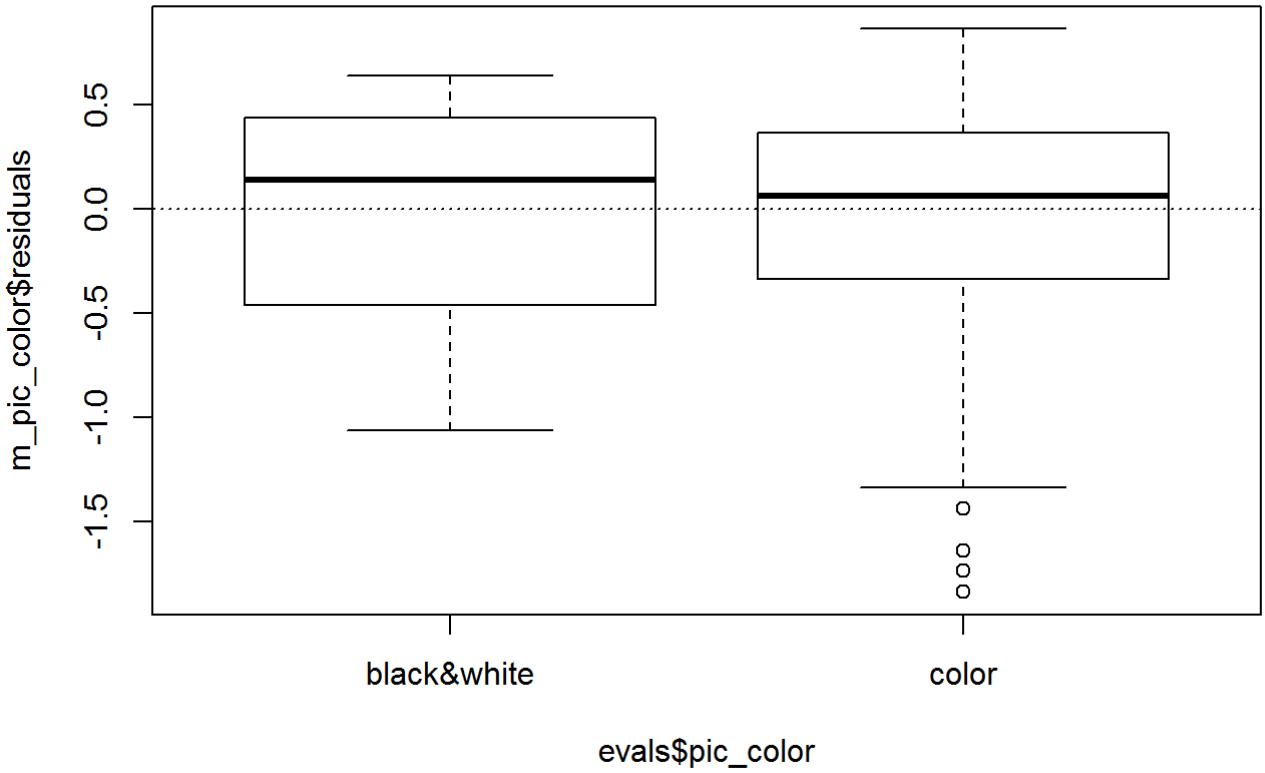

```
m_cls_perc_eval <- lm(evals$score ~ evals$cls_perc_eval)
plot(m_cls_perc_eval$residuals ~ evals$cls_perc_eval)
abline(h = 0, lty = 3)
```




```
m_cls_credits <- lm(evals$score ~ evals$cls_credits)
plot(m_cls_credits$residuals ~ evals$cls_credits)
abline(h = 0, lty = 3)
```




```
m_pic_color <- lm(evals$score ~ evals$pic_color)
plot(m_pic_color$residuals ~ evals$pic_color)
abline(h = 0, lty = 3)
```




```
nrow(evals)
```

```
## [1] 463
```

```
#According to exercise 7, gender and bty_avg are linearly related to outcome. All the graphs sho
wed on top reveals that there are linear relationship bewteen the evalation and these explanator
y variables.
```

Exercise 17: The original paper describes how these data were gathered by taking a sample of professors from the University of Texas at Austin and including all courses that they have taught. Considering that each row represents a course, could this new information have an impact on any of the conditions of linear regression?

Yes, because if some professors are teaching more than one course, the course itself might become an explanatory variable, or it will influence the final evaluation score. (difficult courses maybe associated with lower evaluation score)

Exercise 18: Based on your final model, describe the characteristics of a professor and course at University of Texas at Austin that would be associated with a high evaluation score.

The professors who are non-minority, male, english speaking, young age, teaching one credits course, having black&white color picture, having higher percentage of studetns in class who completed evaluation tend to have higher evaluation score.

Exercise 19: Would you be comfortable generalizing your conclusions to apply to professors generally (at any university)? Why or why not?

No, because this dataset only has 463 samples. And it is only from one university-University of Texas at Austin.