Multiple linear regression

Grading the professor

Many college courses conclude by giving students the opportunity to evaluate the course and the instructor anonymously. However, the use of these student evaluations as an indicator of course quality and teaching effectiveness is often criticized because these measures may reflect the influence of non-teaching related characteristics, such as the physical appearance of the instructor. The article titled, "Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity" (Hamermesh and Parker, 2005) found that instructors who are viewed to be better looking receive higher instructional ratings. (Daniel S. Hamermesh, Amy Parker, Beauty in the classroom: instructors pulchritude and putative pedagogical productivity, *Economics of Education Review*, Volume 24, Issue 4, August 2005, Pages 369-376, ISSN 0272-7757, 10.1016/j.econedurev.2004.07.013. http://www.sciencedirect.com/science/article/pii/S0272775704001165.)

In this lab we will analyze the data from this study in order to learn what goes into a positive professor evaluation.

The data

The data were gathered from end of semester student evaluations for a large sample of professors from the University of Texas at Austin. In addition, six students rated the professors' physical appearance. (This is aslightly modified version of the original data set that was released as part of the replication data for *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Gelman and Hill, 2007).) The result is a data frame where each row contains a different course and columns represent variables about the courses and professors.

```
load("more/evals.RData")
```

variable	description
`score`	average professor evaluation score: (1) very unsatisfactory - (5) excellent.
`rank`	rank of professor: teaching, tenure track, tenured.
`ethnicity`	ethnicity of professor: not minority, minority.
`gender`	gender of professor: female, male.
`language`	language of school where professor received education: english or non-english.
`age`	age of professor.
`cls_perc_eval`	percent of students in class who completed evaluation.
`cls_did_eval`	number of students in class who completed evaluation.
`cls_students`	total number of students in class.
`cls_level`	class level: lower, upper.
`cls_profs`	number of professors teaching sections in course in sample: single, multiple.
`cls_credits`	number of credits of class: one credit (lab, PE, etc.), multi credit.

variable	description
bty_f1lower	beauty rating of professor from lower level female: (1) lowest - (10) highest.
bty_f1upper	beauty rating of professor from upper level female: (1) lowest - (10) highest.
bty_f2upper	beauty rating of professor from second upper level female: (1) lowest - (10) highest.
bty_m1lower	beauty rating of professor from lower level male: (1) lowest - (10) highest.
bty_m1upper	beauty rating of professor from upper level male: (1) lowest - (10) highest.
bty_m2upper	beauty rating of professor from second upper level male: (1) lowest - (10) highest.
bty_avg	average beauty rating of professor.
pic_outfit	outfit of professor in picture: not formal, formal.
pic_color	color of professor's picture: color, black & white.

Loading Packages

```
library(ggplot2)
library(DATA606)
```

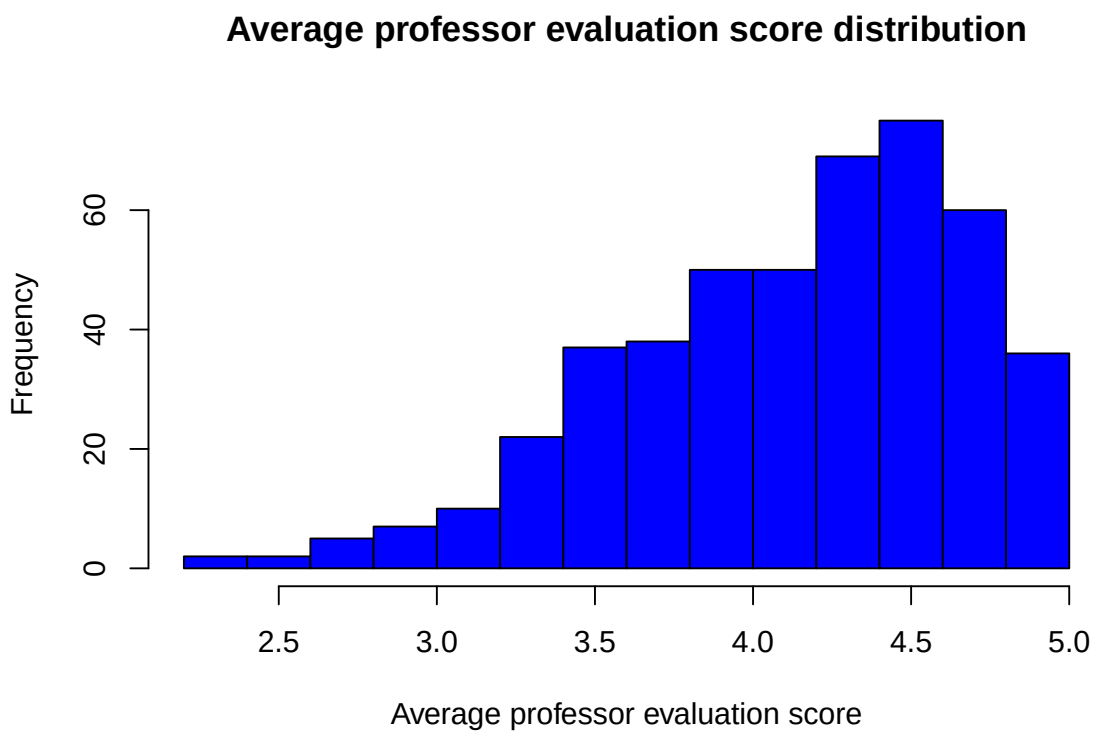
```
##
## Welcome to CUNY DATA606 Statistics and Probability for Data Analytics
## This package is designed to support this course. The text book used
## is OpenIntro Statistics, 3rd Edition. You can read this by typing
## vignette('os3') or visit www.OpenIntro.org.
##
## The getLabs() function will return a list of the labs available.
##
## The demo(package='DATA606') will list the demos that are available.
```

Exploring the data

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.
   Answer: This is an observational study as there are no control and experimental groups. Since, this is only an observational study, there cannot be causation between the explanatory and response variables instead there can only be a correlation. We can rephrase the question as Whether beauty has a positive or negative correlation to student course evaluation.

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?
   Answer:

```
hist(evals$score, main = "Average professor evaluation score distribution", xlab = "Average profess
col = "blue")
```



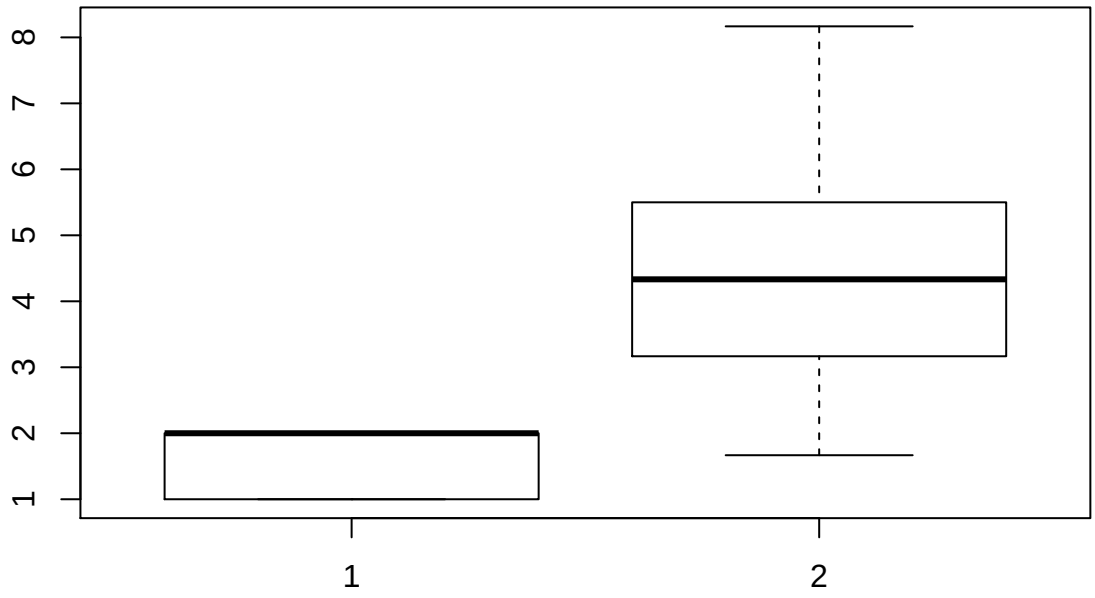

```
summary(evals$score)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.300   3.800   4.300   4.175   4.600   5.000
```

The distribution is uni-modal and skewed left with majority of the observation with score of between 4 and 5. This is not what i expected. Student have far more positive rating than negative. I expected a normal distribution where most teachers would be rated as average and fewer teachers will be evaluated in the extremes excellent or unsatisfactory.

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

```
boxplot(evals$gender, evals$bty_avg)
```



The boxplot shows that average beauty rating of female professor is higher than men. This might end up resluting higher course evaluation rating for female professor compared to men.

Simple linear regression

The fundamental phenomenon suggested by the study is that better looking teachers are evaluated more favorably. Let's create a scatterplot to see if this appears to be the case:

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

Before we draw conclusions about the trend, compare the number of observations in the data frame with the approximate number of points on the scatterplot. Is anything awry?

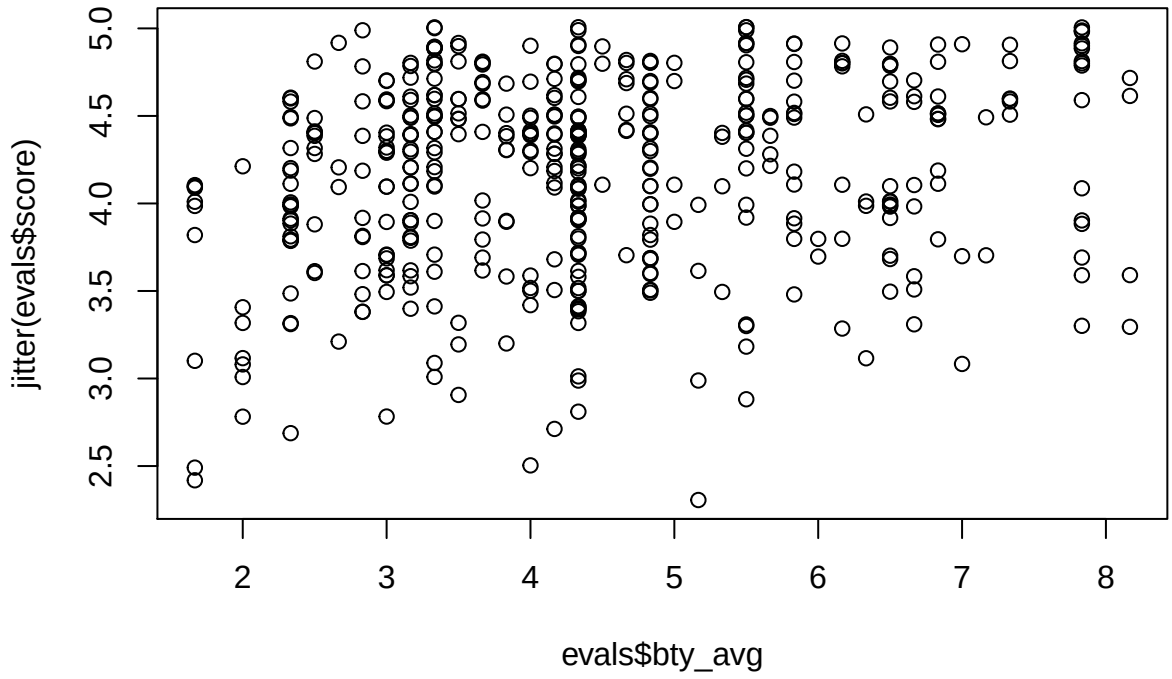
```
nrow(evals)
```

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

There seem to be less observations ploted in the scatterplot than actual observations(463).

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

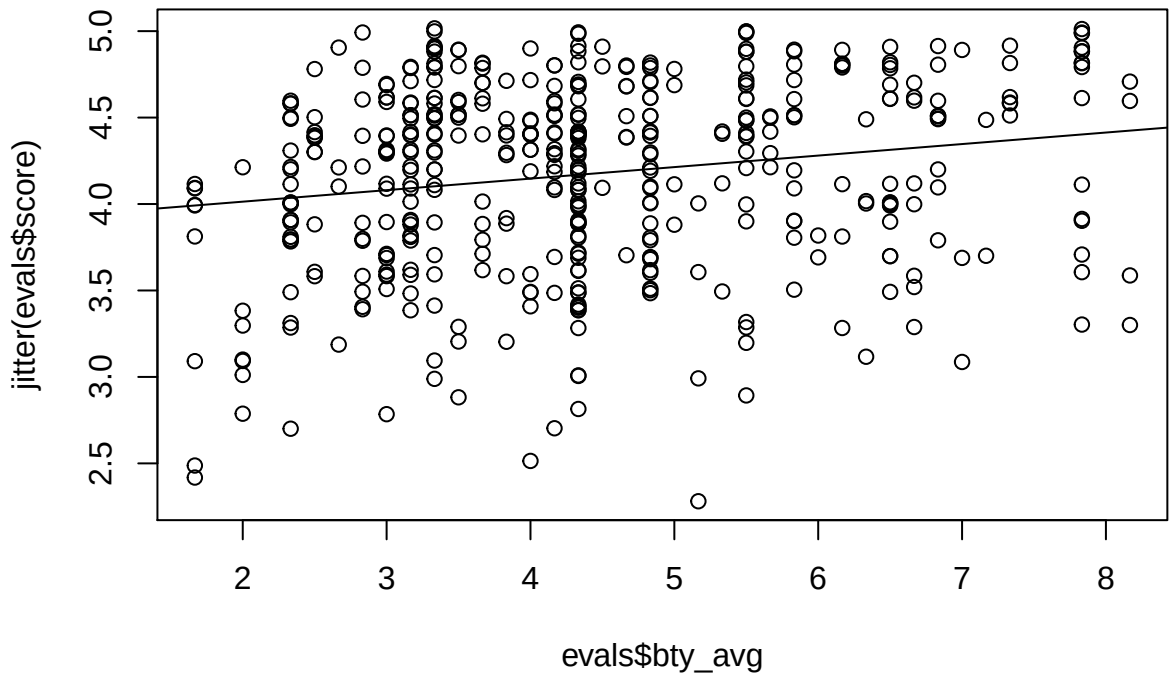
```
plot(jitter(evals$score) ~ evals$bty_avg)
```



It appears that quite a large number of points had the same values for (x,y), hence, they could not be differenciated on the scatter plot. By adding a small amount of noise on the score variable (y), we can now differentiate the points with the help of `jitter()` function.

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?
   Answer:

```
m_bty<- lm(score ~ bty_avg, data = evals)
plot(jitter(evals$score) ~ evals$bty_avg)
abline(m_bty)
```




```
summary(m_bty)
```

```
##
## Call:
## lm(formula = score ~ bty_avg, data = evals)
##
## 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 ***
## 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
```

The equation of the regression line: score = 3.88 + .06664 * bty_avg

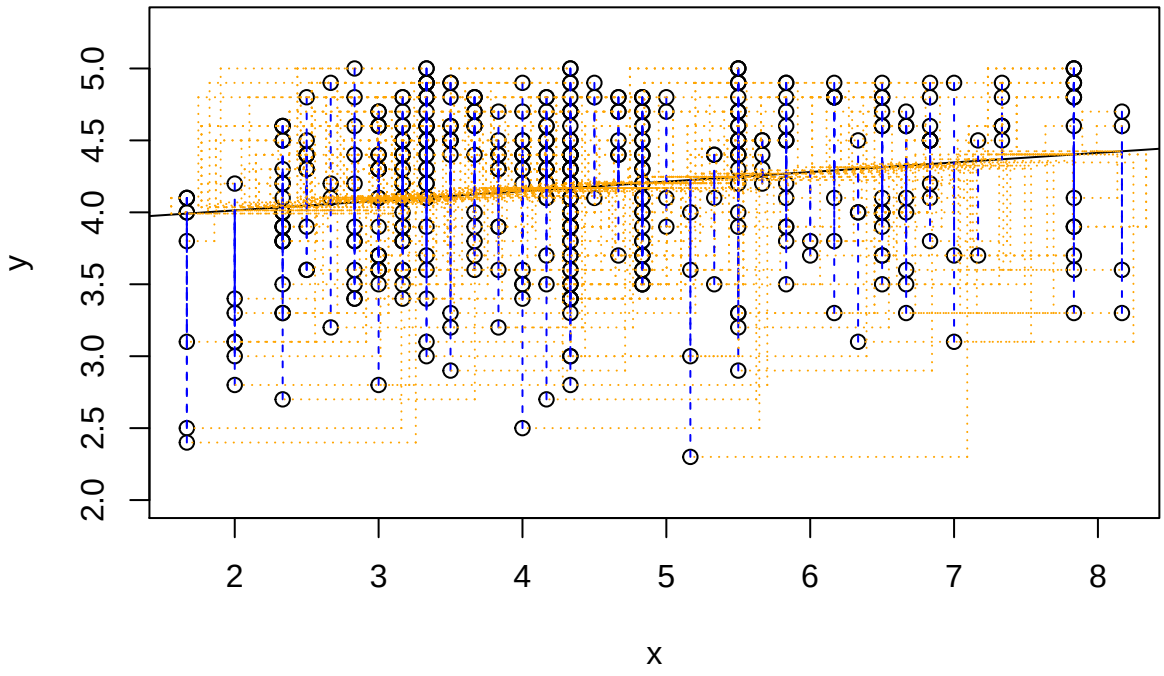
Slope interpretation: When the average beauty score of the professor goes up by 1, we would expect the rating evaluation to go up by .06664.

Statistical Significance: Looking at the p value for the t score, we can say that the average beauty score is statiscally significant p(>|t|) = 5.08 x 10-5.
We would not however find this to be a practically significant predictor. Since this must be evaluated additionally of the professor evaluation. Not all student might want to rate beauty and there is no universal beauty standard. we would imagine that it may difficult to have evaluation across gender line (i.e. male evaluating male, female evaluating female). Moreover, cultural and individual elements will play significant role in rating beauty. Therefore, it would be difficult to accurately value this variable.

R2 is about 3.3%, hence, 3.3% of evaluations can be predicted accurately with the model.

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).

```
library(DATA606)
plot_ss(x = evals$bty_avg, y = evals$score, showSquares = TRUE)
```




```
## Click two points to make a line.

## Call:
## lm(formula = y ~ x, data = pts)
##
## Coefficients:
## (Intercept)            x
##     3.88034      0.06664
##
## Sum of Squares:  131.868
```

Conditions for the least squares line
Linearity: Though it is difficult to tell with so many points, but it appear that the data show a slightly linear positive linearity.
Nearly Normal residuals: From the Histogram, the residuals show a slightly left skewed distribution. The normal probability plot of the residuals shows that the points do not follow the line for upper quadriles.
Constant Variability: From the residual plot, we can observe that there seems to have constant variability.
Independent observations: We do not have much information on how the sample was taken. We can assume indenpendence of the observations.

Multiple linear regression

The data set contains several variables on the beauty score of the professor: individual ratings from each of the six students who were asked to score the physical appearance of the professors and the average of these six scores. Let's take a look at the relationship between one of these scores and the average beauty score.

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

As expected the relationship is quite strong - after all, the average score is calculated using the individual scores. We can actually take a look at the relationships between all beauty variables (columns 13 through 19) using the following command:

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

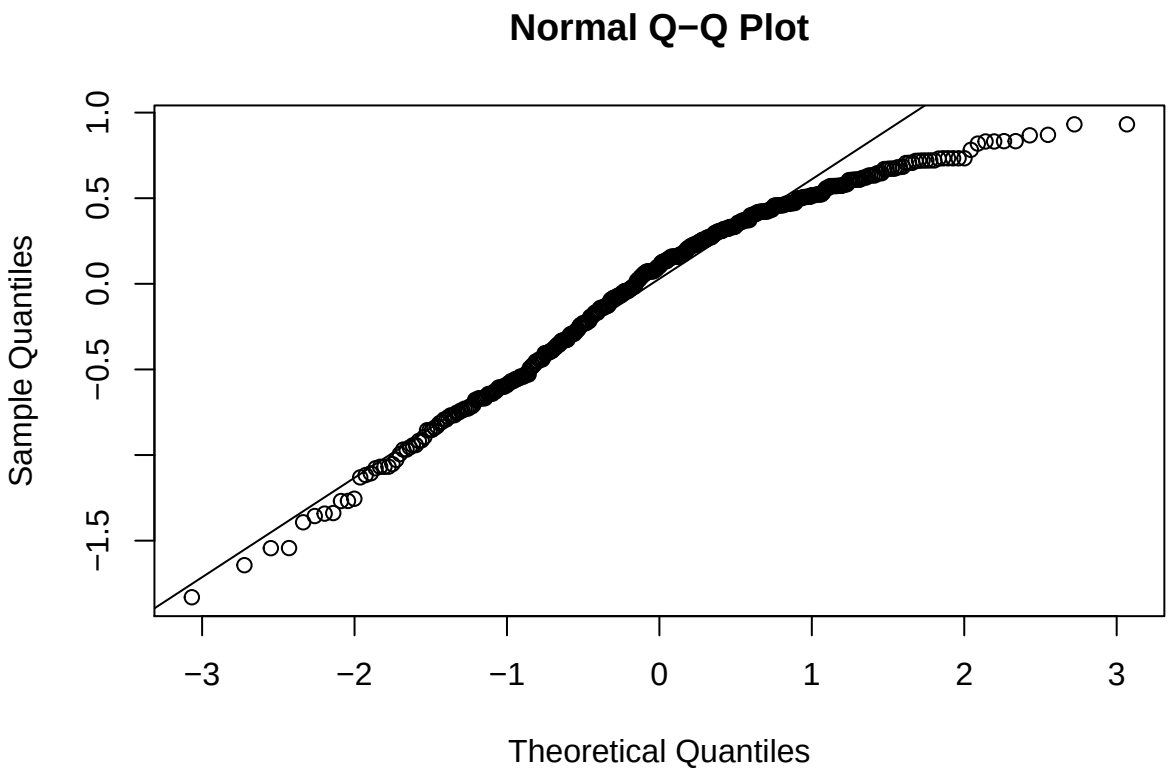
These variables are collinear (correlated), and adding more than one of these variables to the model would not add much value to the model. In this application and with these highly-correlated predictors, it is reasonable to use the average beauty score as the single representative of these variables.

In order to see if beauty is still a significant predictor of professor score after we've accounted for the gender of the professor, we can add the gender term into the model.

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

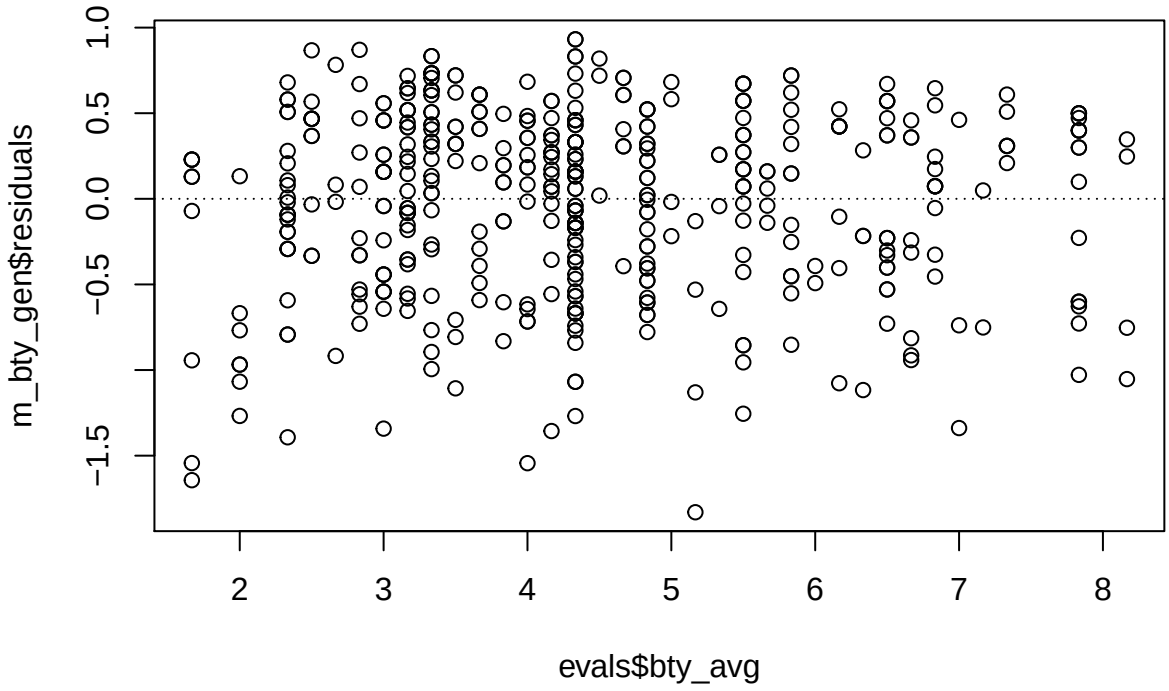
7. P-values and parameter estimates should only be trusted if the conditions for the regression are reasonable. Verify that the conditions for this model are reasonable using diagnostic plots.

```
library(ggplot2)
m_bty_gen <- lm(score ~ bty_avg + gender, data = evals)
qqnorm(m_bty_gen$residuals)
qqline(m_bty_gen$residuals)
```



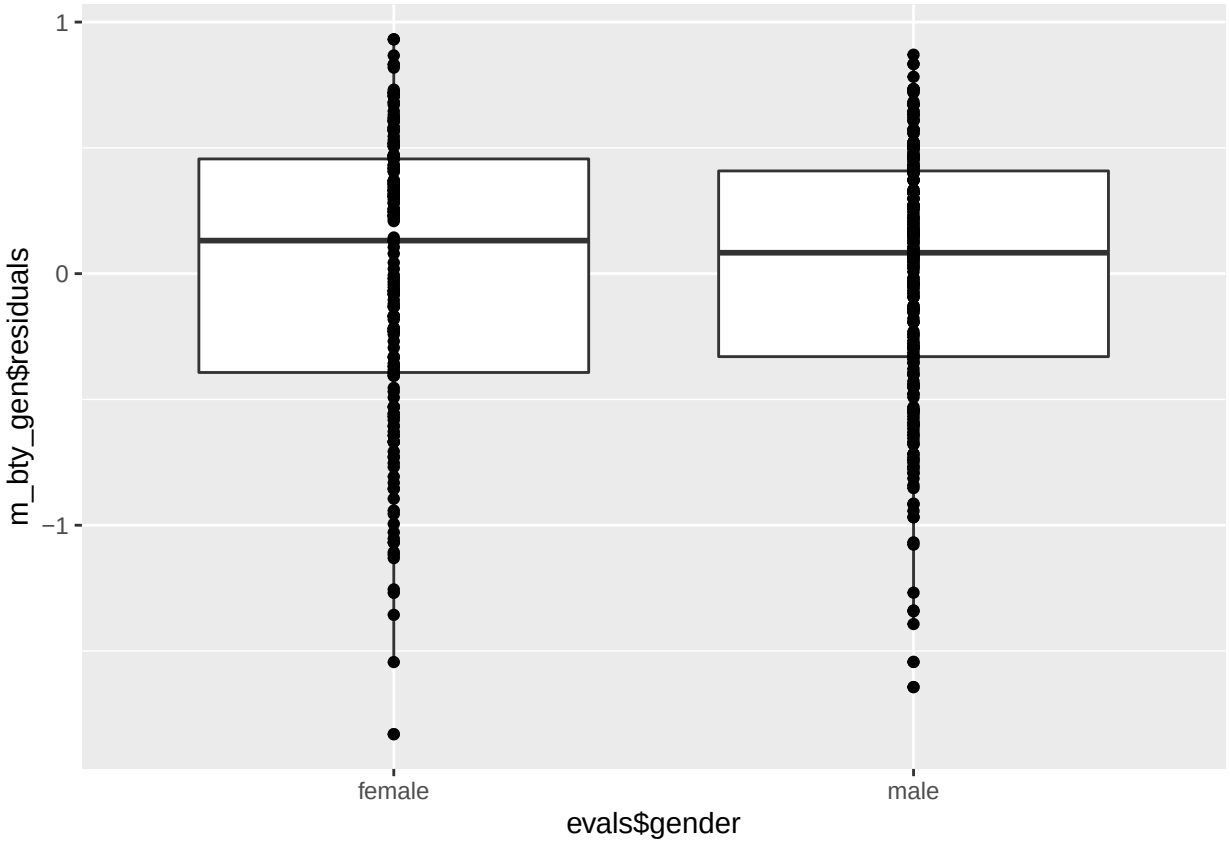

```
#residual are nearly normal

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




```
#variablity is nearly constant

ggplot(evals,aes(y=m_bty_gen$residuals,x=evals$gender))+geom_boxplot()+geom_point()
```



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`?

```
summary(m_bty_gen)
```

```
##
## Call:
## lm(formula = score ~ bty_avg + 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 ***
## bty_avg      0.07416    0.01625   4.563 6.48e-06 ***
## 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
```

Yes it is. In fact, gender made beauty average even more significant as the p-value computed is even smaller now compared to a model where beauty average was the sole variable. Moreover, There was a slight increase in the intercept and increase in the slope

Note that the estimate for `gender` is now called `gendermale`. You'll see this name change whenever you introduce a categorical variable. The reason is that R recodes `gender` from having the values of `female` and `male` to being an indicator variable called `gendermale` that takes a value of 0 for females and a value of 1 for males. (Such variables are often referred to as "dummy" variables.)

As a result, for females, the parameter estimate is multiplied by zero, leaving the intercept and slope form familiar from simple regression.

$$\begin{aligned}\widehat{score} &= \hat{\beta}_0 + \hat{\beta}_1 \times bty\_avg + \hat{\beta}_2 \times (0) \\ &= \hat{\beta}_0 + \hat{\beta}_1 \times bty\_avg\end{aligned}$$

We can plot this line and the line corresponding to males with the following custom function.

```
multiLines(m_bty_gen)
```

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?
   Answer:

$$\begin{aligned}\widehat{score} &= 3.74734 + 0.07416 \text{Ö} bty\_avg + 0.17239 \text{Ö} gender\_male \\ &= 3.74734 + 0.07416 \text{Ö} bty\_avg + 0.1723\end{aligned}$$

Male professor will have a evaluation score higher by 0.17239 all other things being equal.

The decision to call the indicator variable `gendermale` instead of`genderfemale` has no deeper meaning. R simply codes the category that comes first alphabetically as a 0. (You can change the reference level of a categorical variable, which is the level that is coded as a 0, using the`relevel` function. Use `?relevel` to learn more.)

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`. **Answer:**

```
m_bty_rank <- lm(score ~ bty_avg + rank, data = evals)
summary(m_bty_rank)
```

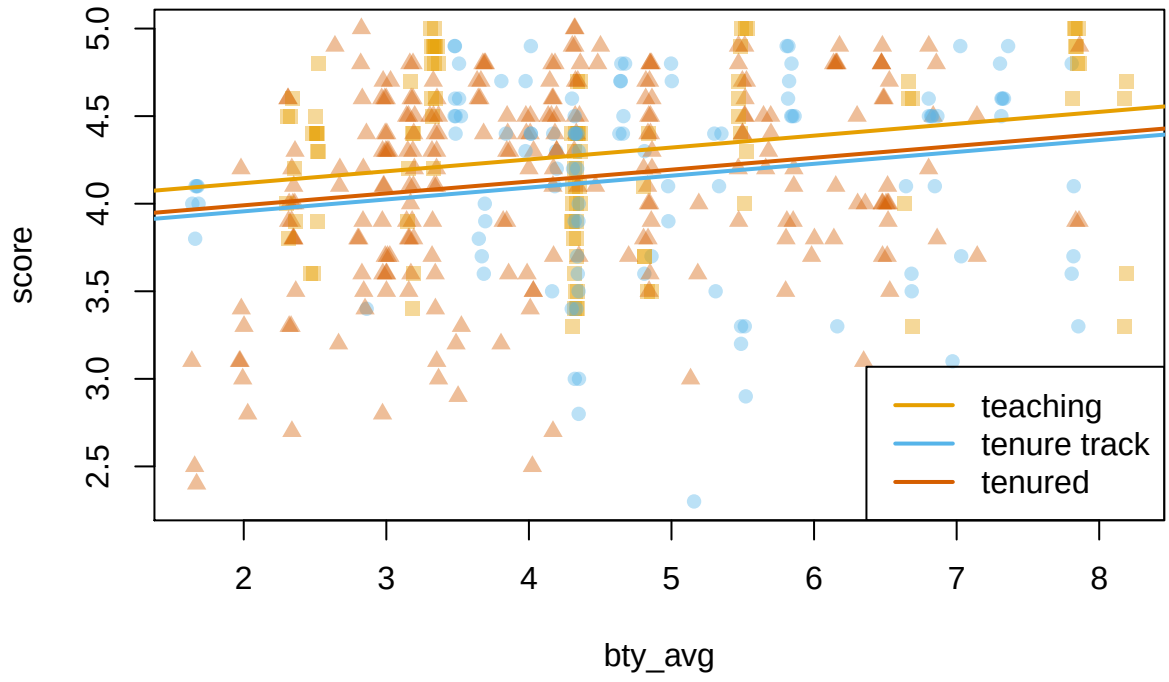
```
##
## Call:
## lm(formula = score ~ bty_avg + rank, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8713 -0.3642  0.1489  0.4103  0.9525
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.98155    0.09078  43.860  < 2e-16 ***
## bty_avg           0.06783    0.01655   4.098 4.92e-05 ***
```

```
## ranktenure track -0.16070    0.07395  -2.173   0.0303 *
## 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
```

```
names(m_bty_rank)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "contrasts"     "xlevels"       "call"          "terms"
## [13] "model"
```

```
multiLines(m_bty_rank)
```



Since the rank variable has three levels (teaching, tenure track and tenured), R has added another line into the regression summary to account for it. R leaves out one level but mentions the rest as variables.

The interpretation of the coefficients in multiple regression is slightly different from that of simple regression. The estimate for `bty_avg` reflects how much higher a group of professors is expected to score if they have a beauty rating that is one point higher *while holding all other variables constant*. In this case, that translates into considering only professors of the same rank with `bty_avg` scores that are one point apart.

The search for the best model

We will start with a full model that predicts professor score based on rank, ethnicity, gender, language of the university where they got their degree, age, proportion of students that filled out evaluations, class size, course level, number of professors, number of credits, average beauty rating, outfit, and picture color.

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.
Answer: Number of professors teaching sections in course in sample: single, multiple, since the evaluation are done within a class/section. Whether the professor is teaching multiple sections should not have an impact a given class evaluation.

Let's run the model...

```
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)
```

12. Check your suspicions from the previous exercise. Include the model output in your response.
Answer: My suspicion was right. The p value for cls_profssingle variable, is 0.77806 and is the highest in the model

13. Interpret the coefficient associated with the ethnicity variable.
Answer: All other things being equal, Evaluation for professor that not minority tends to be 0.1234929 higher.

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?
Answer:

```
m_full_1 <- 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_1)
```

```
##
## 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
```

Yes. There was a slight change in the coefficients and significance of the other explanatory variables when cls_profs was removed. All the values are now slightly lower, meaning they are more significant now to the level than before.

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.
    Answer:

```
m_back2 <- lm(score ~ ethnicity + gender + language + age + cls_perc_eval + cls_credits + bty_avg + pic
summary(m_back2)
```

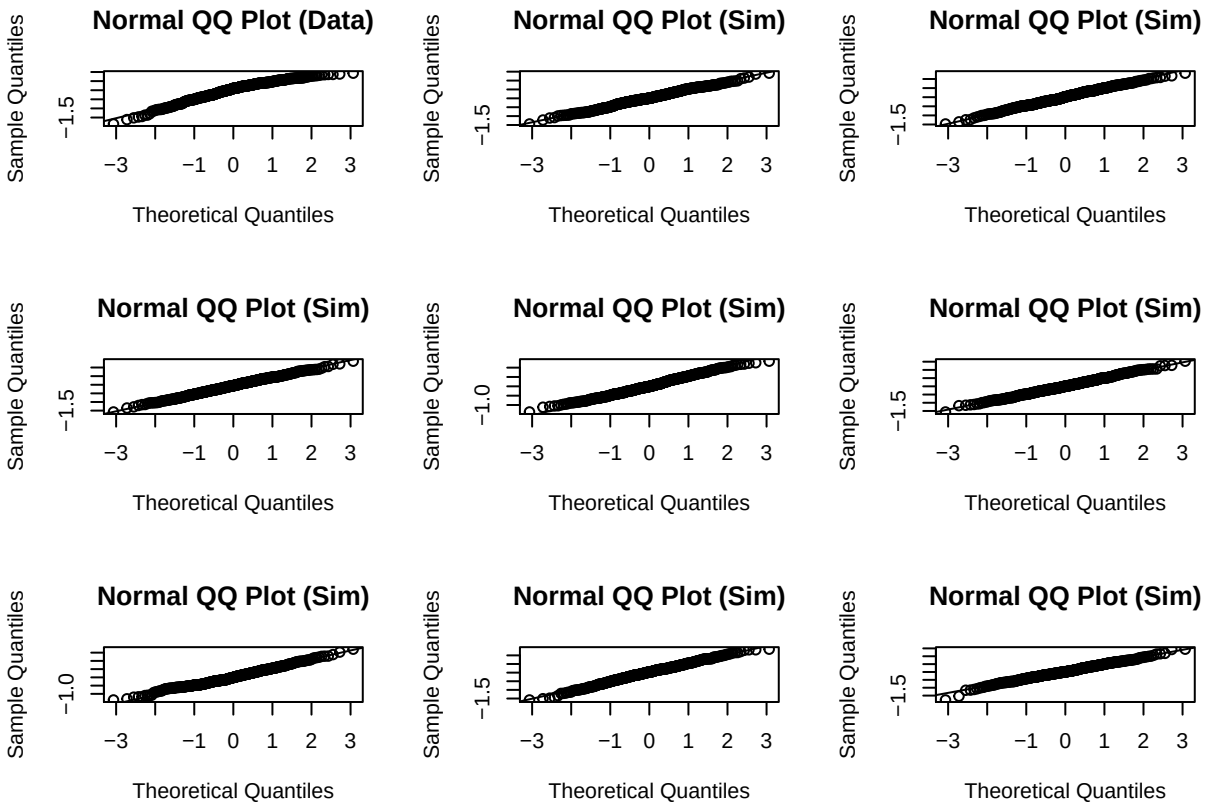
```
##
## Call:
## lm(formula = score ~ ethnicity + gender + language + age + cls_perc_eval +
##     cls_credits + bty_avg + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.85320 -0.32394  0.09984  0.37930  0.93610
##
## Coefficients:
##                        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
```

professor rating = 3.771922 + 0.167872xeth + 0.207112xgender -0.206178xlang -0.006046xage +0.004656xperceval + .505306xcredits + .051069xbeauty - .190579xcolor

16. Verify that the conditions for this model are reasonable using diagnostic plots.
    Answer: The residuals of the model are nearly normal

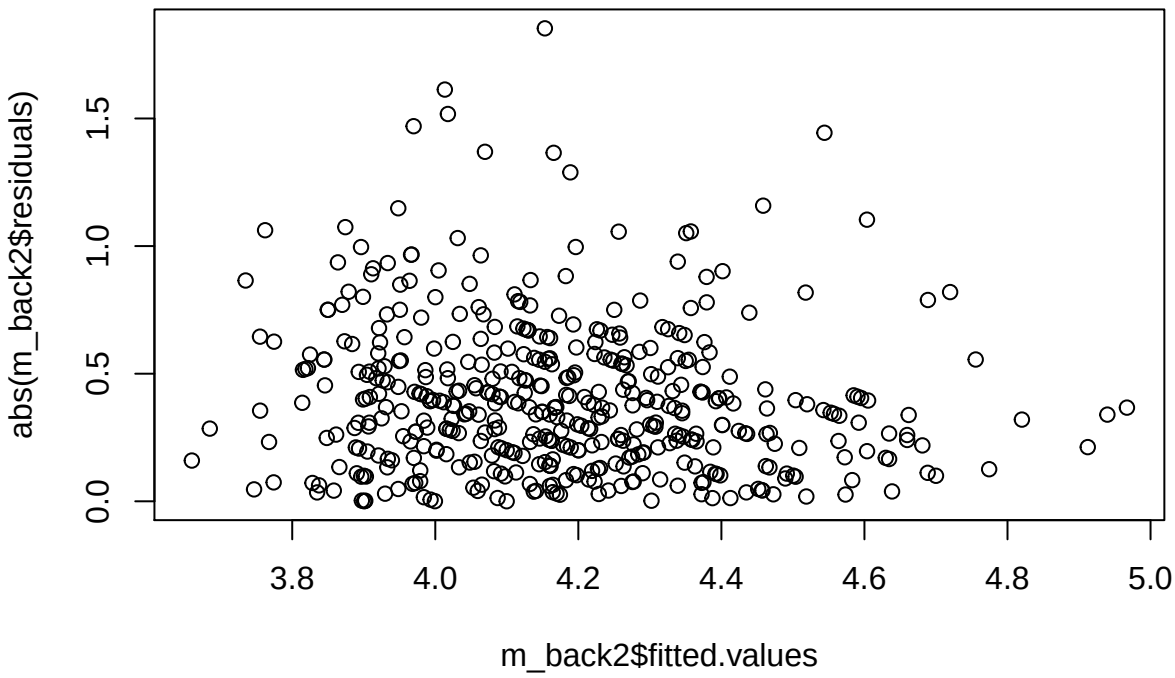
```
qqnormsim(m_back2$residuals)
```



The residuals of the model is not normal as residual values for the the higher and lower quantiles are less than what a normal distribution would predict

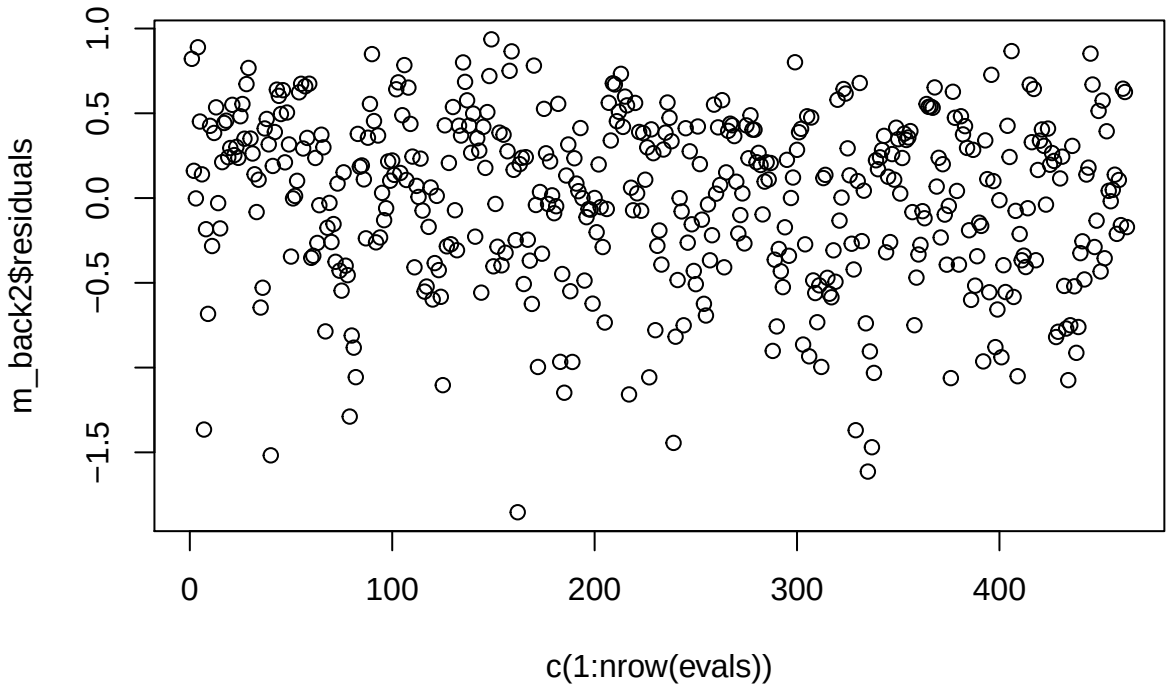
Absolute values of residuals against fitted values. (the variability of the residuals is nearly constant)

```
plot(abs(m_back2$residuals) ~ m_back2$fitted.values)
```



There some outliers although overall, most of the residual values are close to the fitted values
The residuals are independent

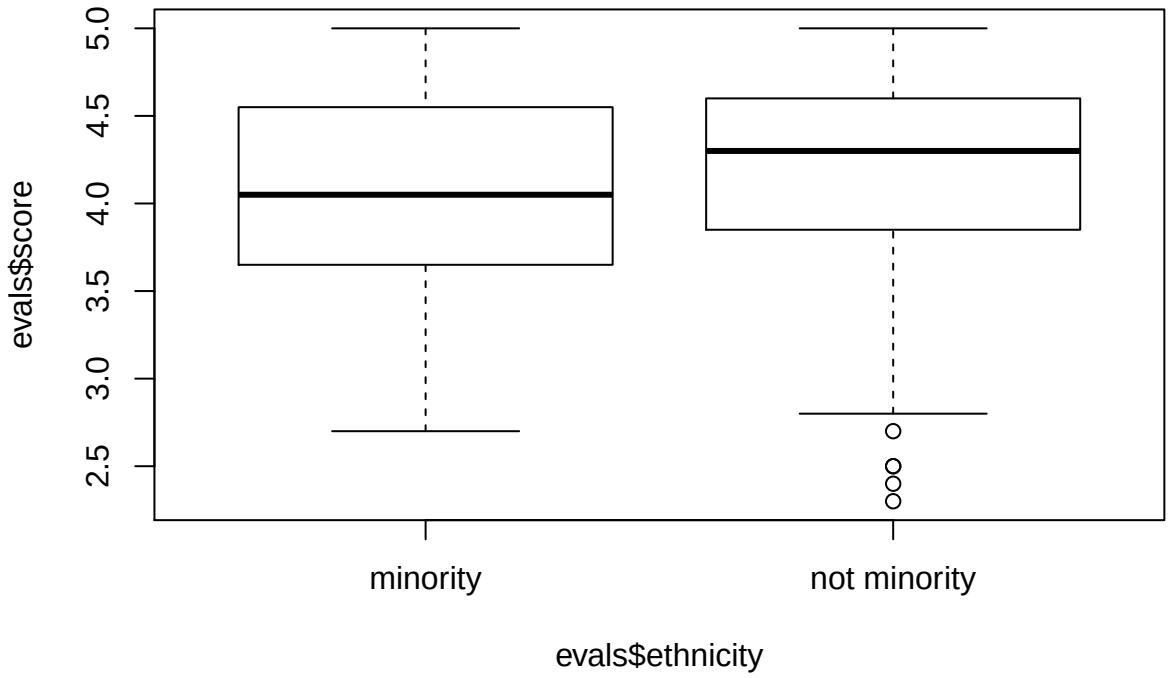
```
plot(m_back2$residuals ~ c(1:nrow(evals)))
```



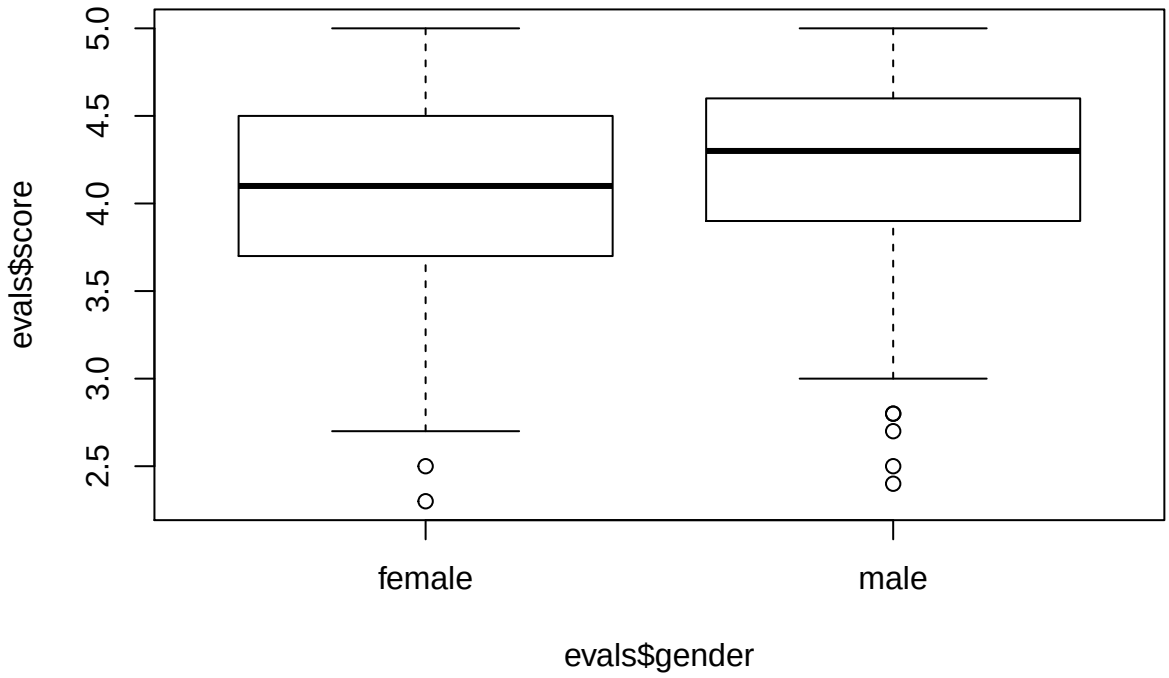
Yes, this condition is met. The residuals based on the sequence when it was gathered shows that they were randomly gathered.

Each variable is linearly related to the outcome.

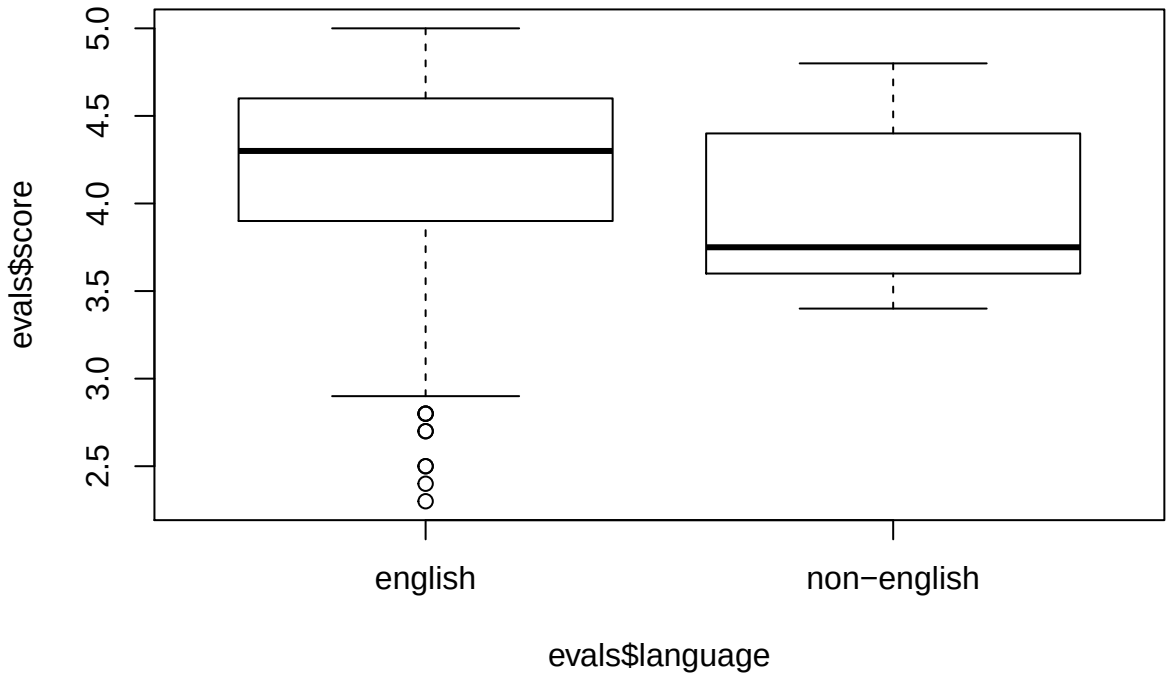
```
plot(evals$score ~ evals$ethnicity)
```



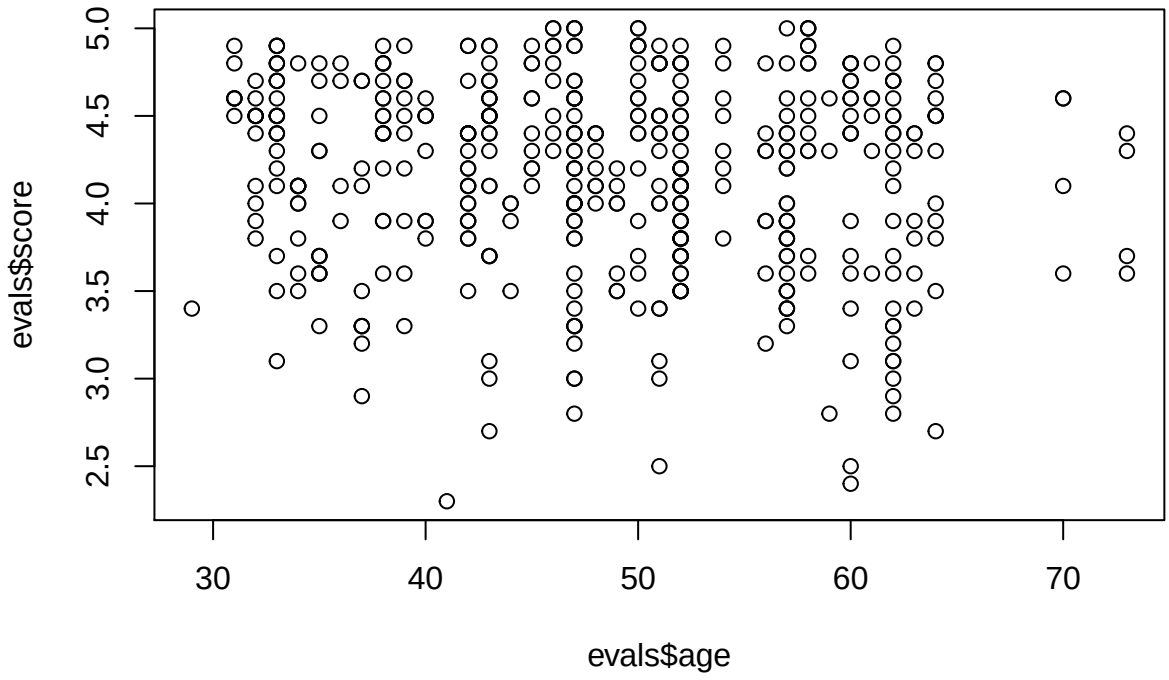

```
plot(evals$score ~ evals$gender)
```



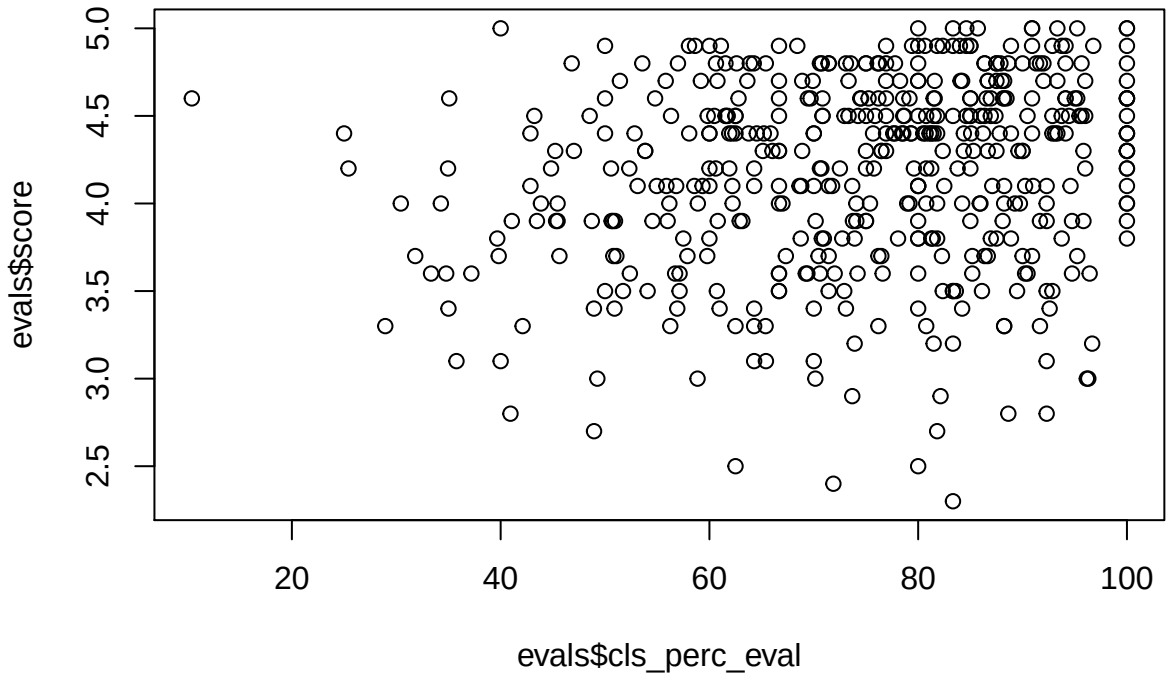

```
plot(evals$score ~ evals$language)
```



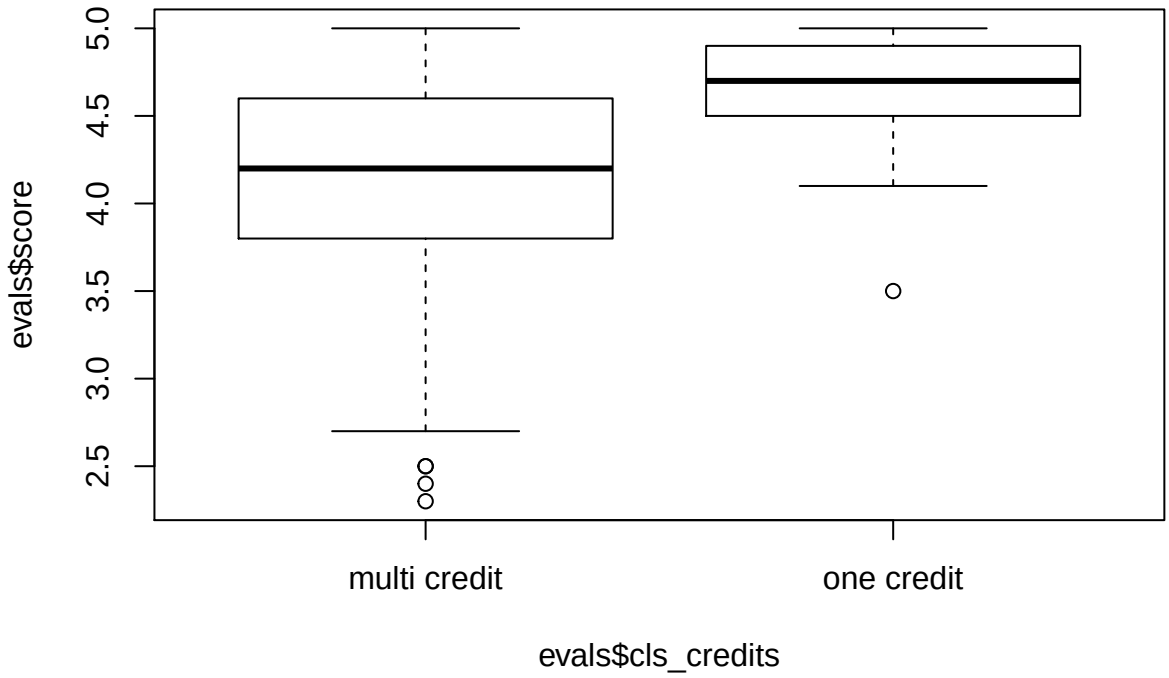

```
plot(evals$score ~ evals$age)
```



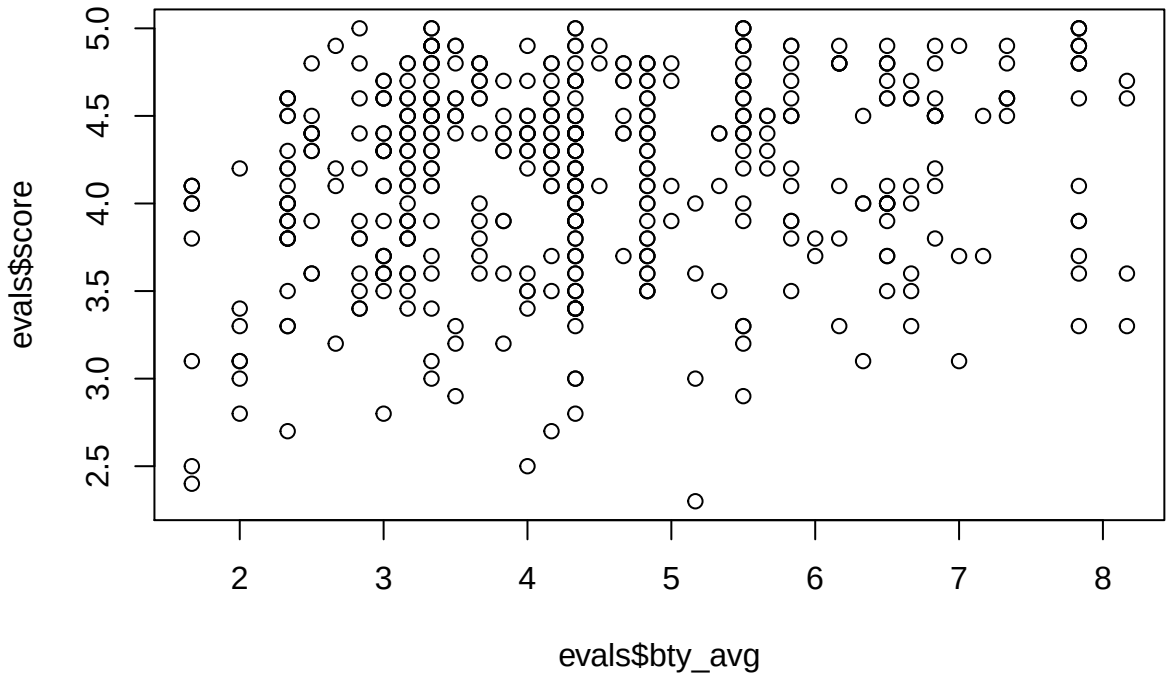

```
plot(evals$score ~ evals$cls_perc_eval)
```



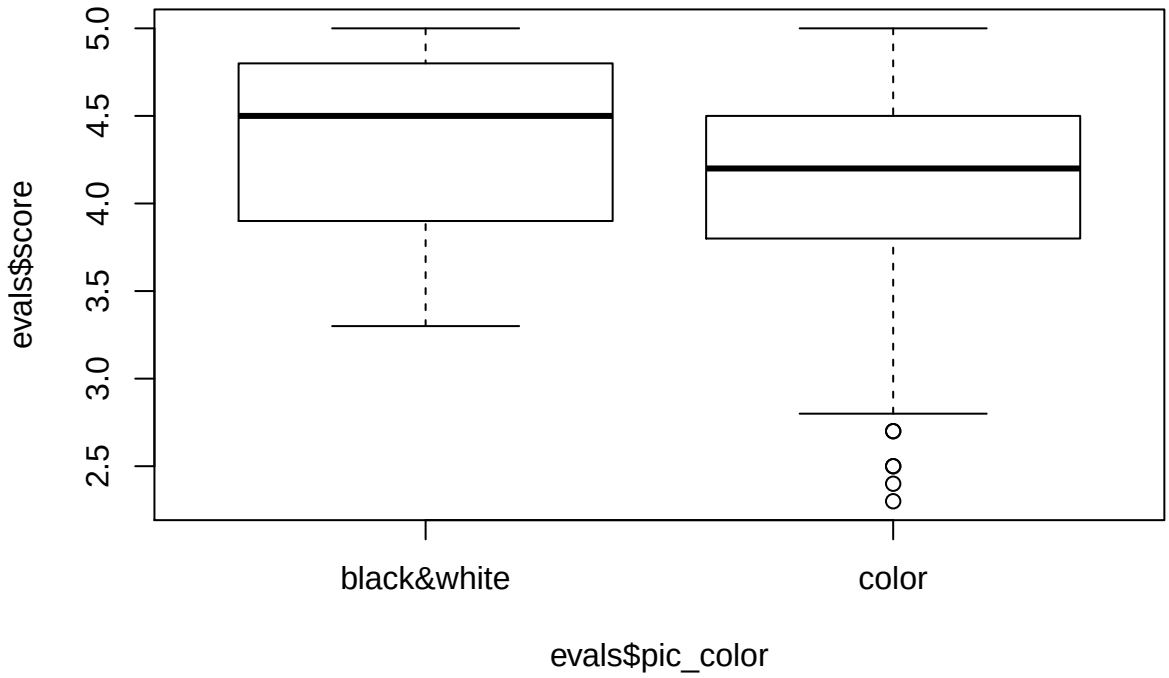

```
plot(evals$score ~ evals$cls_credits)
```




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




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



The variables above are linearly related to the score - some more so than others.

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?
Answer: No. Class courses are independent of each other so evaluation scores from one course is indpendent of the other even if the course is being taught by the same professor.

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.
Answer: The professor is not a minority and male, must have graduated from an American (or English speaking) school and teaches a one credit course. He must also have a high beauty average score from the students and the professor's class photo should be in black and white. He must also be relatively young. And a good percentage of his class must have completed the evaluation

19. Would you be comfortable generalizing your conclusions to apply to professors generally (at any university)? Why or why not?
Answer: No, I would not. We used a samll sample size which doesn't represent all university. Based on the demographic of the school, the importance of each variable may change. For example, if at a campus with a high international student population, have a native English speaker as a professor may not be of importance.