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.
`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.

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.

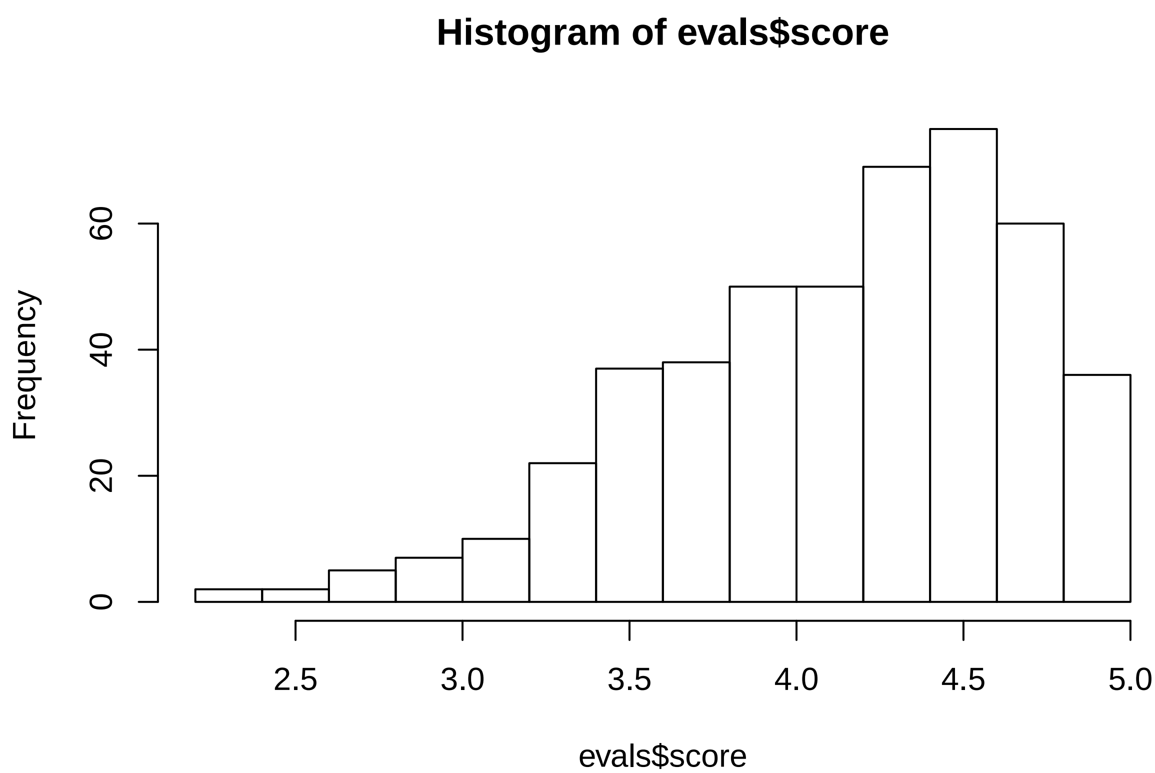
I believe this is an observational study.I don't think that the original research question can be answered with the study design. Rather, a better question is whether beauty has any correlation to the scores.

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?

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

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

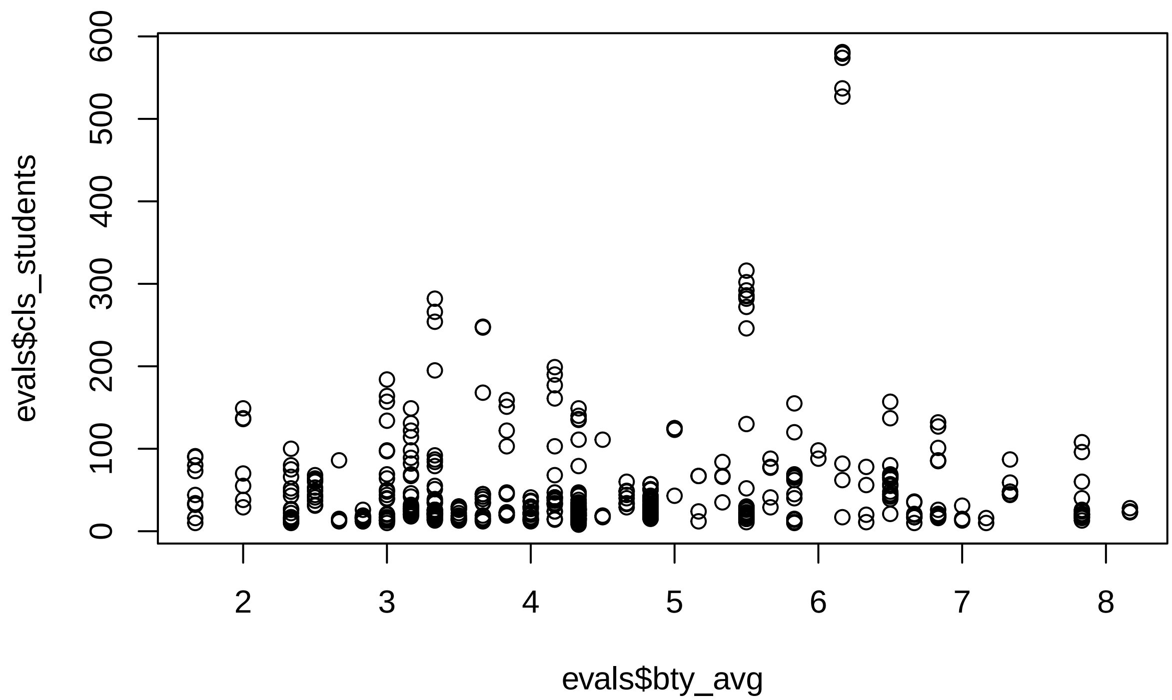
```r
hist(evals$score)
```



The distribution of `score` is skewed to the left. Students in the study have generally scored on the higher end with no scores that are below 2.3. I expected more of a normal distribution of scores as not all professors are good and you generally have a mix.

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

```r
plot(evals$bty_avg, evals$cls_students)
```

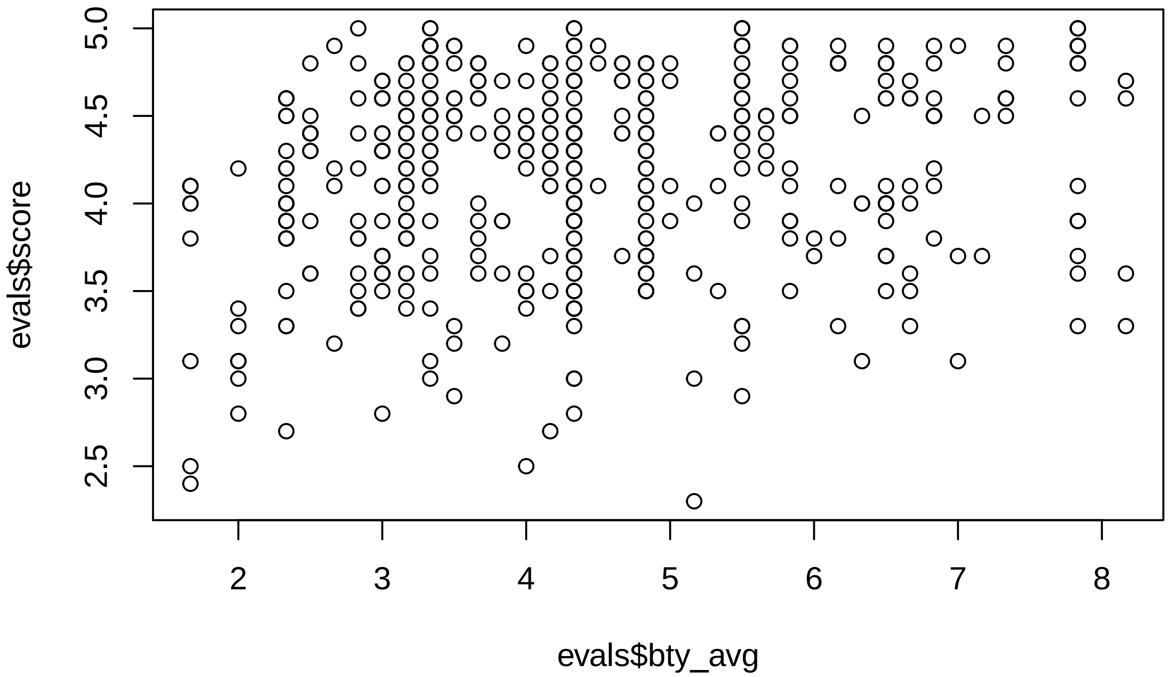


The above scatterplot of Beauty average vs the number of students in a class does not seem to indicate any trend. There is no apparent relationship based on the visualization above.

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:

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

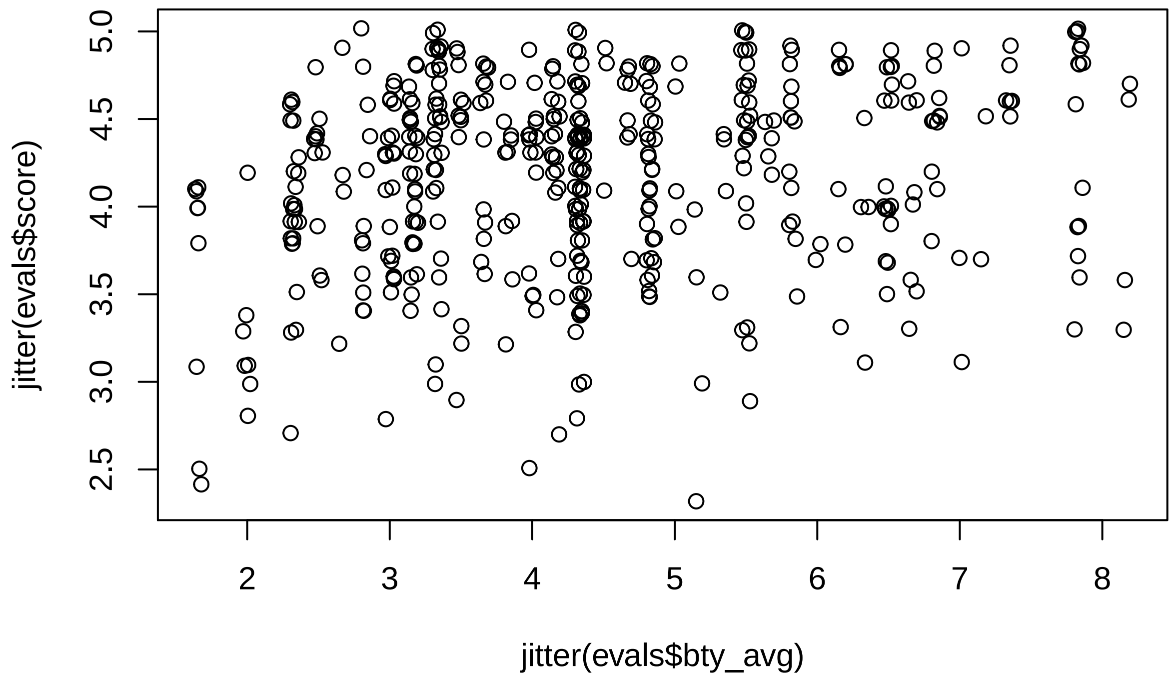


Before we draw conclusions about the trend, compare the number of observations in the data frame with the approximate number of points on the scatterplot. Is anything awry?

The number of observations in the scatter plot seem to be less than the actual number of records (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?

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

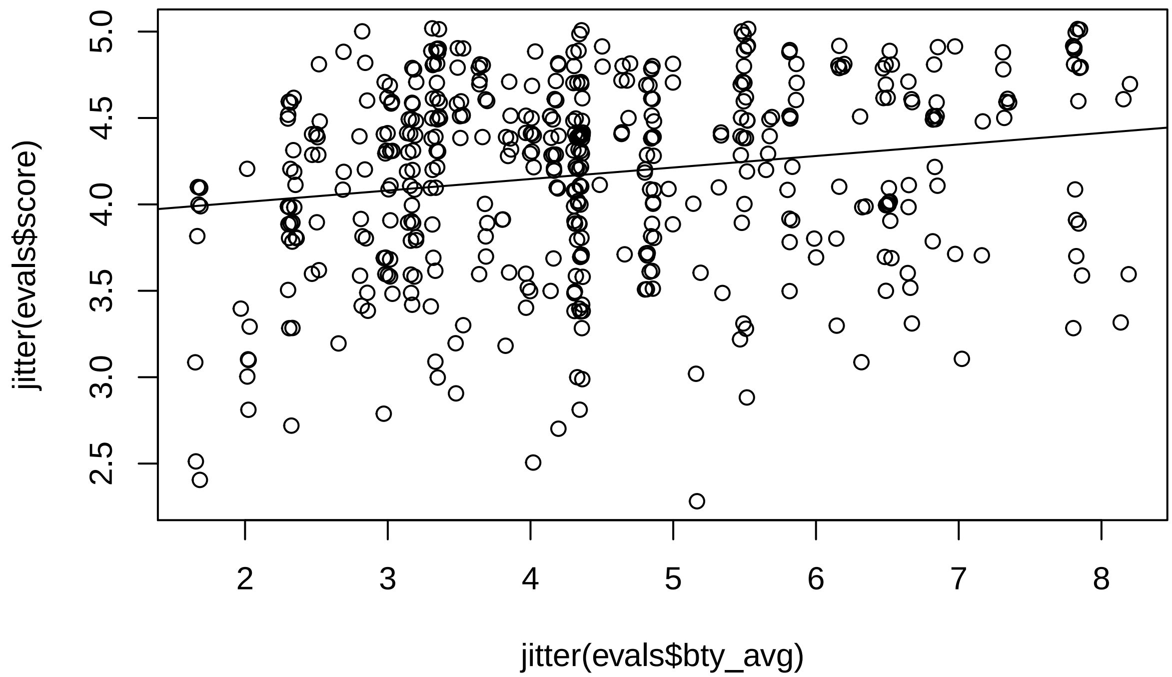


It seems like the duplicate score / beauty average points were not clearly visible in the original plot. Now they seem to be more apparent.

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?

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

abline(m_bty)
```




```r
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
```

Equation for the Linear Line:

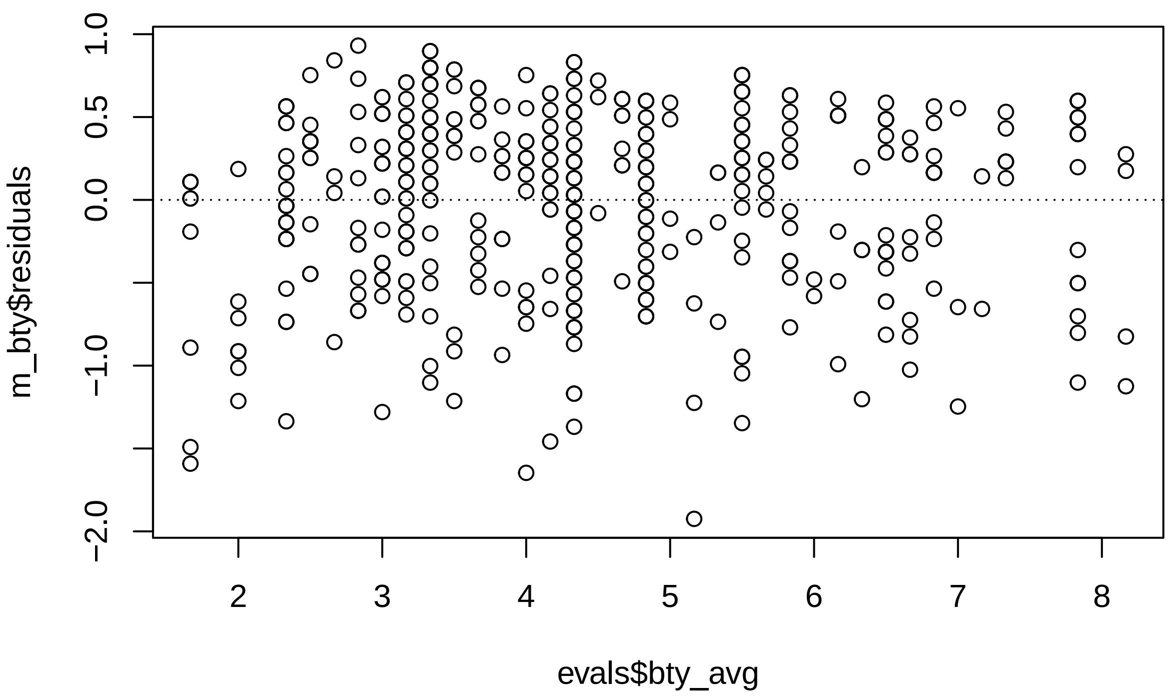
$\hat{score} = 3.880338 + 0.066637 * \text{bty\_avg}$

The slope indicates that for every 1 unit increase in average beauty, there is an increase of 0.066637 in the score.

The p-value is less than 0.05. However, the R2 value practically indicates that it is not a significant predictor.

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)
```



The following are the conditions for least squares regression:

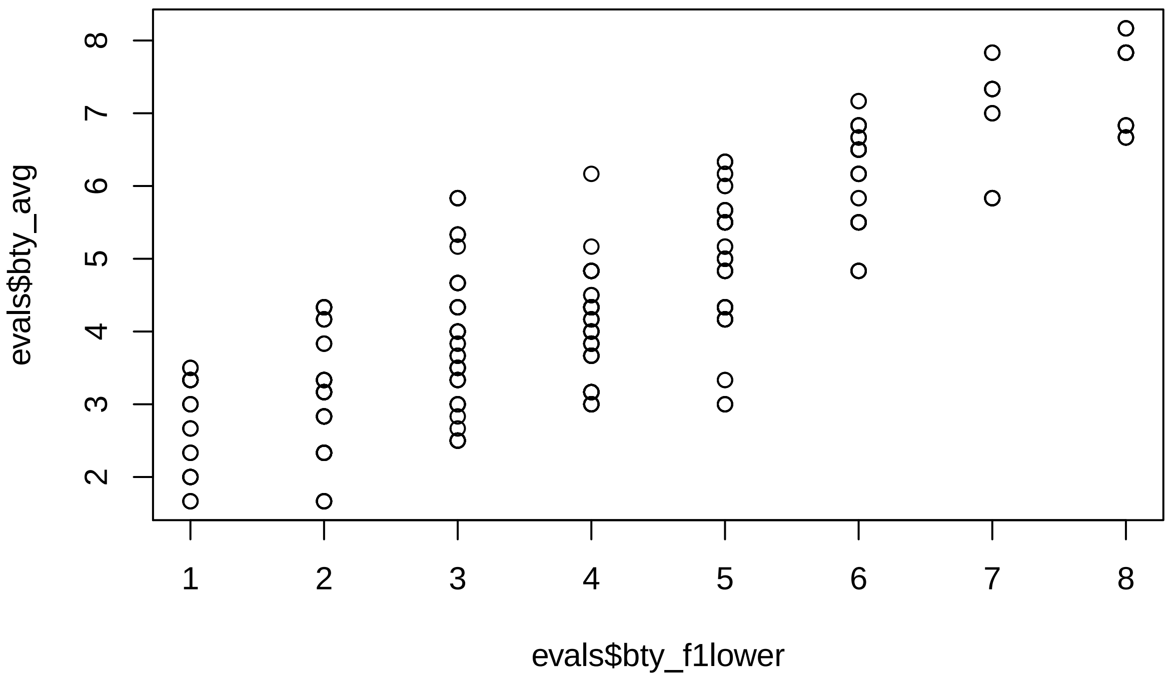
1. Nearly normal residuals - The plot above seems to indicate a nearly normal residual distribution even though there are a few outliers.

2. Linearity - The points are scattered across and are not confined to a small band.

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)
```

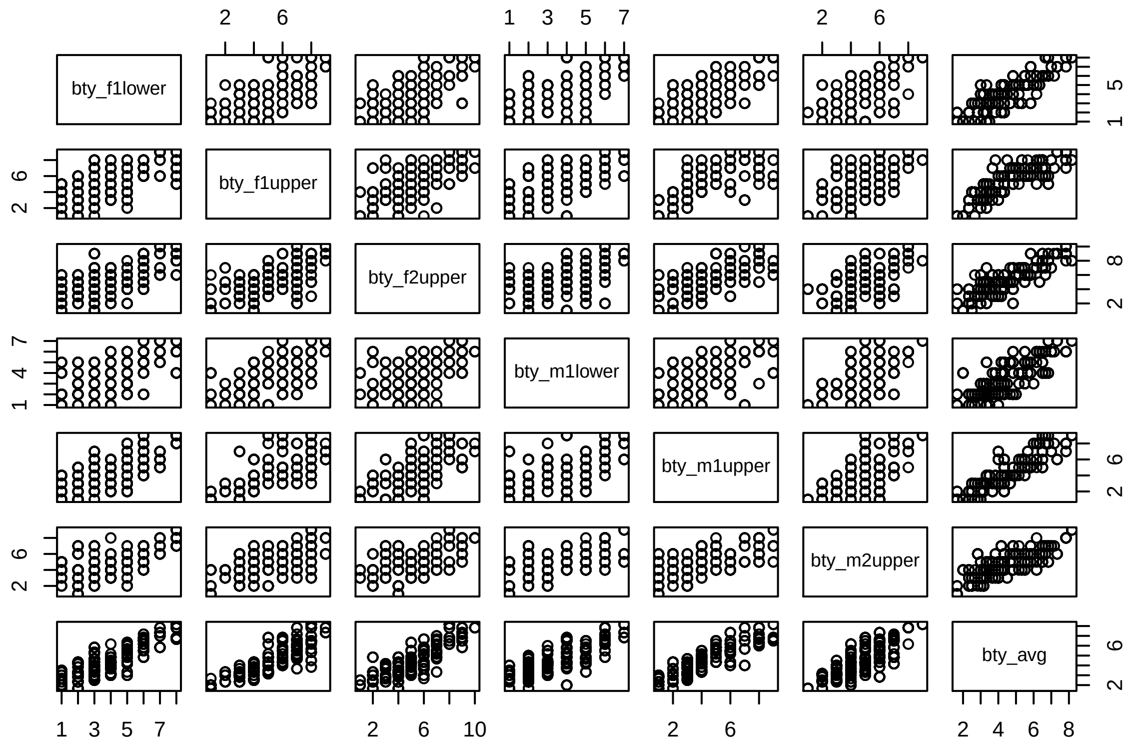



```
cor(evals$bty_avg, evals$bty_f1lower)
```

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

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.

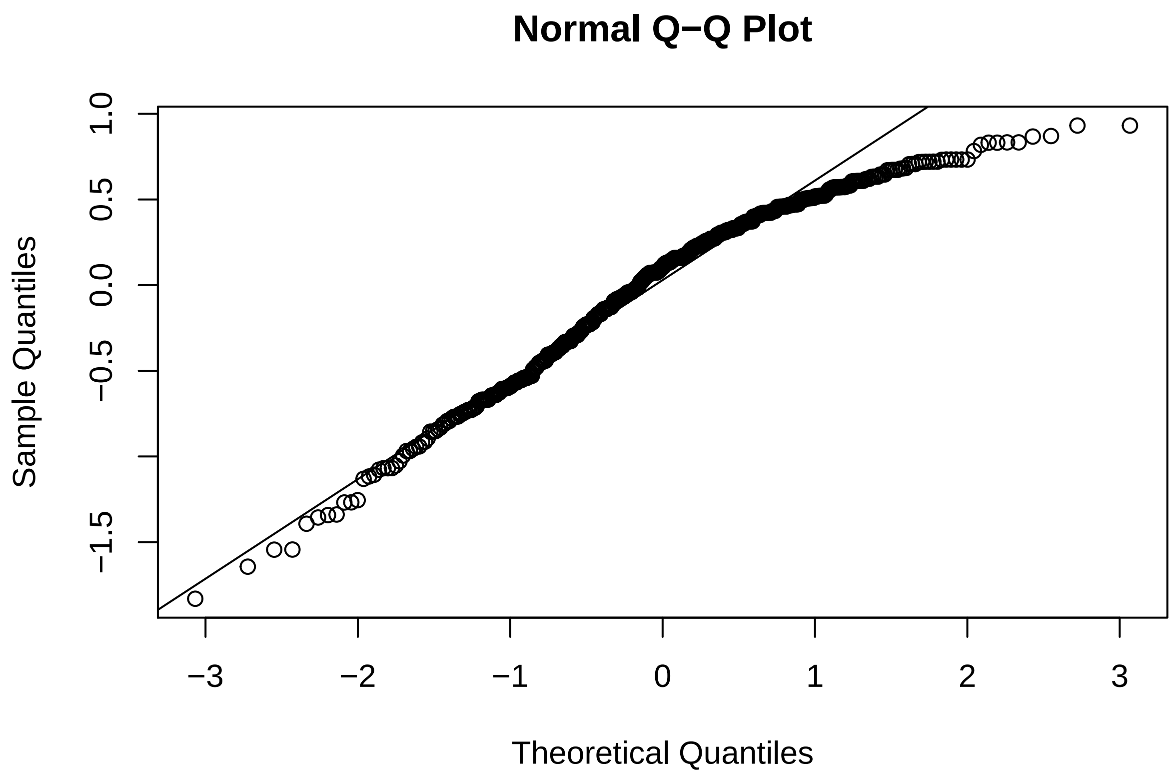
```
m_bty_gen <- lm(score ~ bty_avg + gender, data = evals)
summary(m_bty_gen)
```

```
##
## 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
```

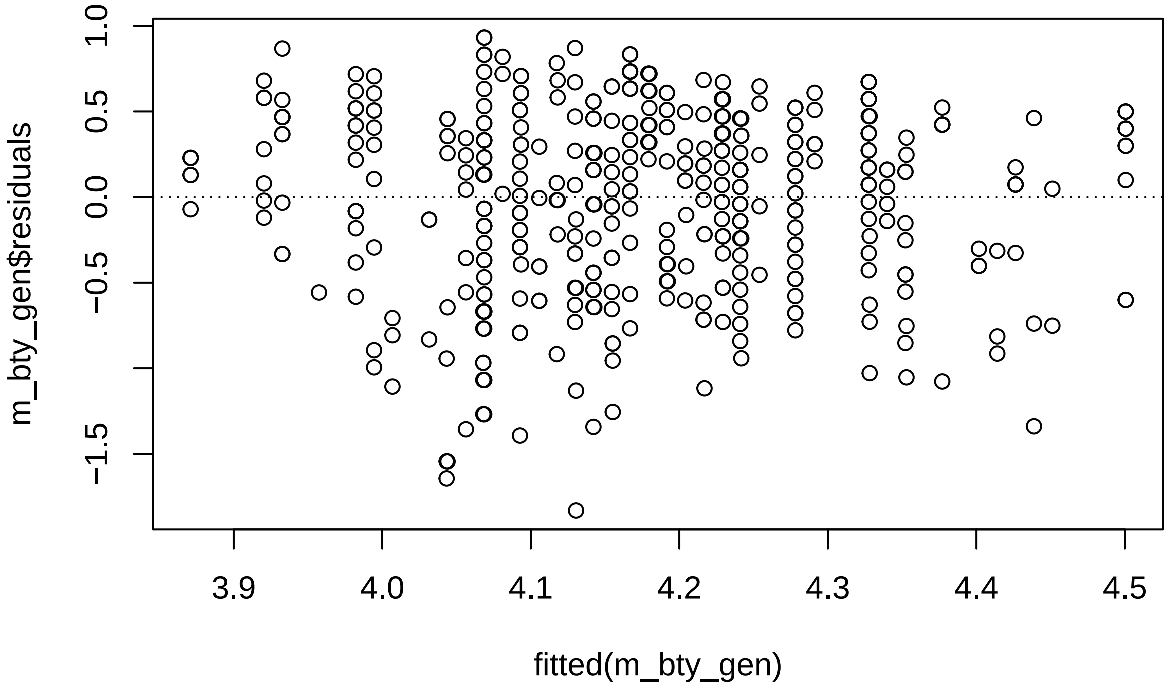
7. P-values and parameter estimates should only be trusted if the conditions for the regression are reasonable. Verify that the conditions for this model are reasonable using diagnostic plots.

```r
qqnorm(m_bty_gen$residuals)
qqline(m_bty_gen$residuals)
```



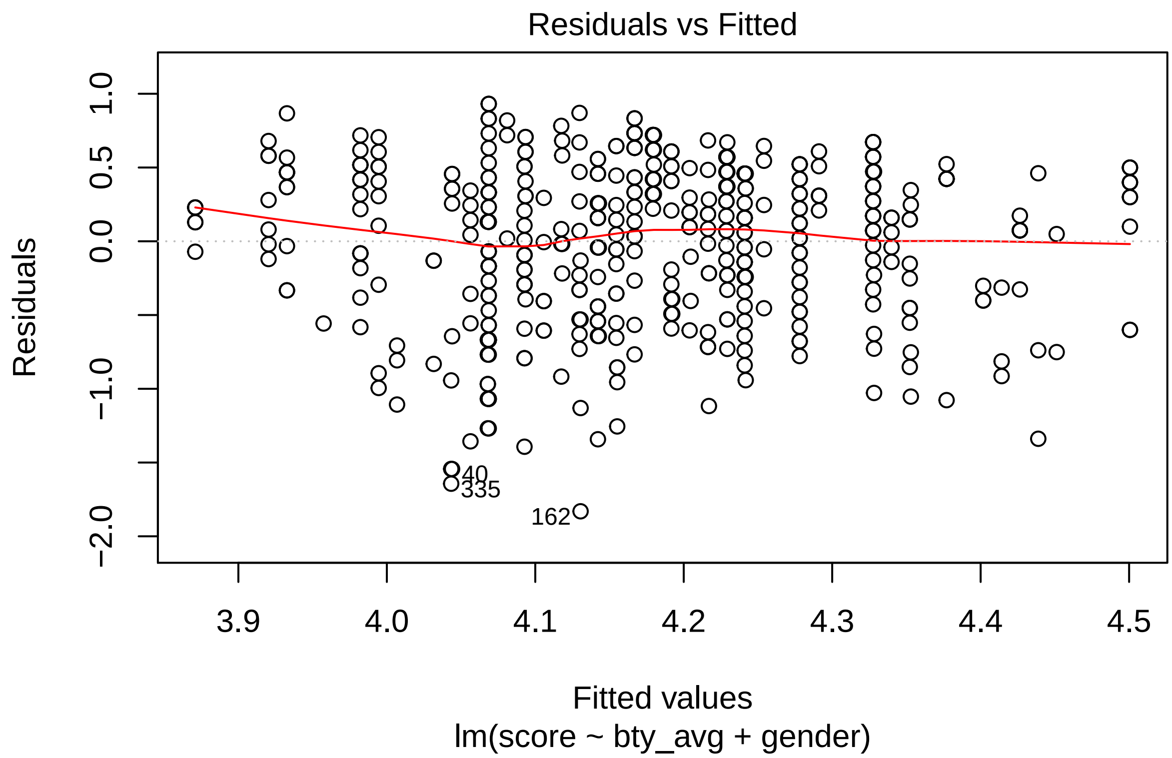
Based on the plot above, we can say that the distribution is nearly normal.

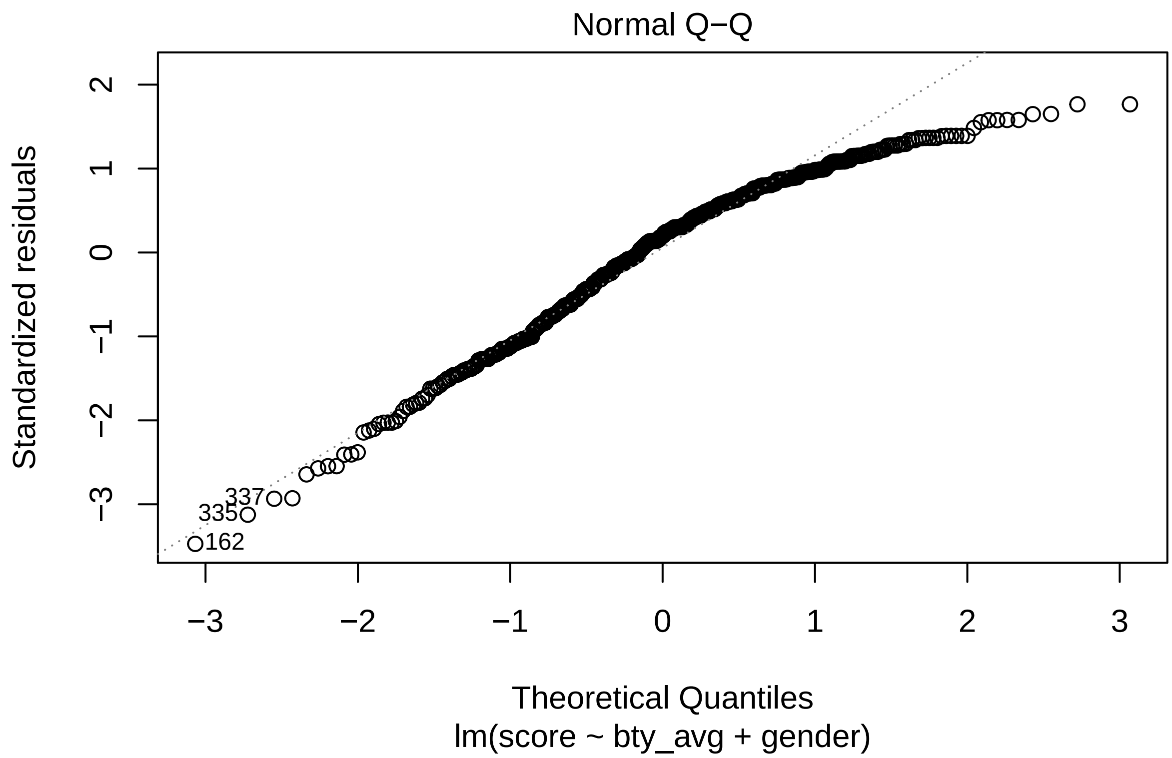
```r
plot(m_bty_gen$residuals ~ fitted(m_bty_gen))
abline(h = 0, lty = 3)
```

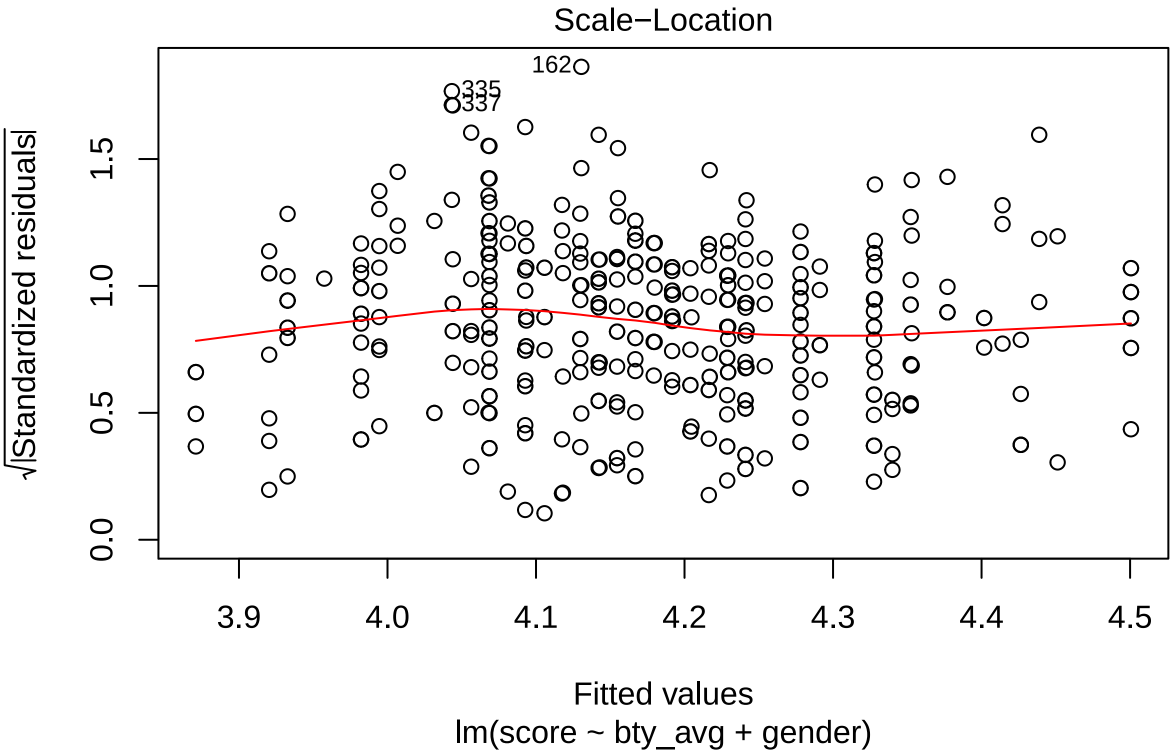


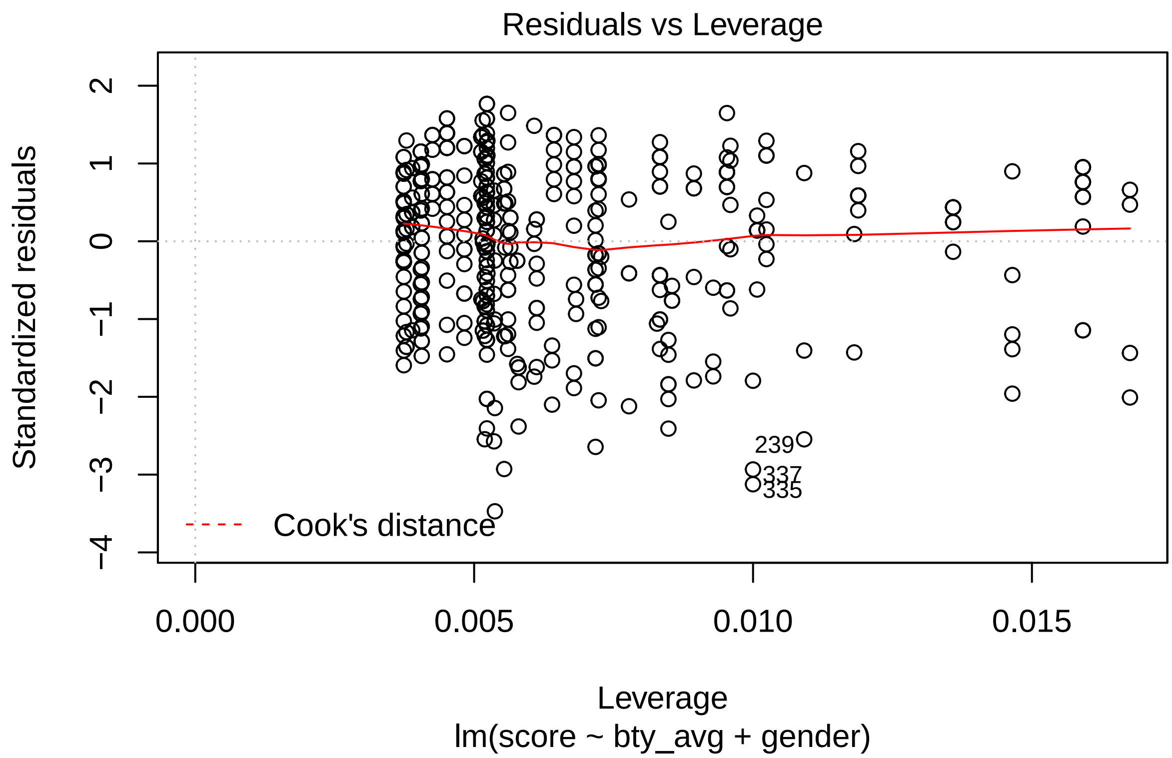
Based on the above plot we can assume the condition for constant variability also holds good. Below are the diagnostic plots.

```r
plot(m_bty_gen)
```


Residuals vs Fitted
Residuals
1.0
0.5
0.0
−1.0
−2.0
40
335
162
3.9
4.0
4.1
4.2
4.3
4.4
4.5
Fitted values
lm(score ~ bty_avg + gender)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(score ~ bty_avg + gender)
337
335
162
−3
−2
−1
0
1
2
3


Scale−Location
162
335
337
√|Standardized residuals|
0.0
0.5
1.0
1.5
3.9
4.0
4.1
4.2
4.3
4.4
4.5
Fitted values
lm(score ~ bty_avg + gender)



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`?

```r
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
```

```r
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
```

In the previous model, the parameter estimate for bty_avg was 0.06664. Now with the addition of gender, the estimate for bty_avg is 0.07416. So the parameter estimate has increased by 0.00752.

The t-value for bty_avg is almost identical and stronger than gender.

In the previous model, the R^2 was 0.03502. Now with the addition of gender, the R^2 is 0.05912. So there is an increase of 0.0241

Adding gender to the model did help, but bty_avg by itself is not a significant predictor of score.

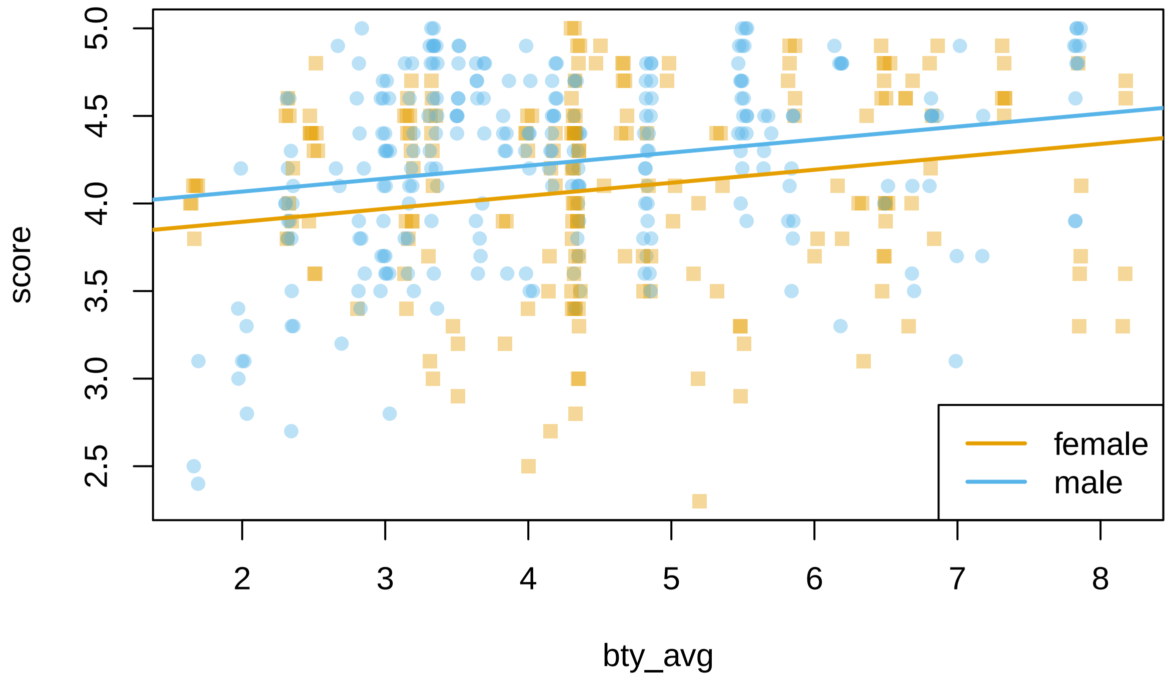
Note that the estimate for `gender` is now called `gendermale`. You'll see this name change whenever you introduce a categorical variable. The reason is that R recodes `gender` from having the values of `female` and `male` to being an indicator variable called `gendermale` that takes a value of 0 for 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.

```r
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?

The following is the equation:

$\hat{score} = 3.7473382 + 0.0741554 \times \text{bty\_avg} + 0.0741554 \times 1$

Adding 0.0741554 for men increases the score for men by that many units. So for the same beauty ratings, men will have a higher course evaluation score.

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`.

```
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
```

R handles categorical variables with more than 2 levels by creating (levels - 1) number of variables. For the rank, it creates "ranktenure track"" abd "ranktenured"" variables. It does not create a variable for "rankteaching". Rather teaching is denoted by having B__1 = 0 and B__2 = 0. Ranktenured track is denoted by B__1 = 1 and B__2 = 0 and RankTenured is denoted by B__1=0 and B__2=1.

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.

I would assume that language would not have any association with the professor score.

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)
```

```
##
## 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
```

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

It seems I am wrong in my assumption. Language have a p-value of "0.03965" whereas, cls_prof seems to have a higher p-value at "0.77806".

13. Interpret the coefficient associated with the ethnicity variable.

Keeping all other variables as constant, an ethnicity value of "not minority" tends to increase the score by 0.1234929 units

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?

```
m_without_profs <- 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_without_profs)
```

```
## 
## 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
```

The p-values and coefficients changed by a small margin. This indicates that cls_prof was collinear with the other variables

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.

Moving backward from the previous step, we find the below model to be the best:

```r
# removing cls_levels
m_inter <- lm(score ~ rank + ethnicity + gender + language + age + cls_perc_eval
             + cls_students + cls_credits + bty_avg
             + pic_outfit + pic_color, data = evals)
summary(m_inter)
```

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_students + cls_credits + bty_avg + pic_outfit +
##     pic_color, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -1.7761 -0.3187  0.0875  0.3547  0.9367
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.0856255  0.2888881  14.143  < 2e-16 ***
## ranktenure track      -0.1420696  0.0818201  -1.736 0.083184 .
## ranktenured           -0.0895940  0.0658566  -1.360 0.174372
## ethnicitynot minority  0.1424342  0.0759800   1.875 0.061491 .
## gendermale             0.2037722  0.0513416   3.969 8.40e-05 ***
## languagenon-english   -0.2093185  0.1096785  -1.908 0.056966 .
## age                   -0.0087287  0.0031224  -2.795 0.005404 **
## cls_perc_eval          0.0053545  0.0015306   3.498 0.000515 ***
## cls_students           0.0003573  0.0003585   0.997 0.319451
## cls_creditsone credit  0.4733728  0.1106549   4.278 2.31e-05 ***
## bty_avg                0.0410340  0.0174449   2.352 0.019092 *
## pic_outfitnot formal  -0.1172152  0.0716857  -1.635 0.102722
## pic_colorcolor        -0.1973196  0.0681052  -2.897 0.003948 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4975 on 450 degrees of freedom
## Multiple R-squared:  0.185,  Adjusted R-squared:  0.1632
## F-statistic:  8.51 on 12 and 450 DF,  p-value: 1.275e-14
```

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

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_credits + bty_avg + pic_outfit + pic_color,
##     data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78424 -0.31397  0.09261  0.35904  0.92154
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.152893   0.280892  14.785  < 2e-16 ***
## ranktenure track      -0.142239   0.081819  -1.738 0.082814 .
## ranktenured           -0.083092   0.065532  -1.268 0.205469
## ethnicitynot minority  0.143509   0.075972   1.889 0.059535 .
## gendermale             0.208080   0.051159   4.067 5.61e-05 ***
## languagenon-english   -0.222515   0.108876  -2.044 0.041558 *
## age                   -0.009074   0.003103  -2.924 0.003629 **
## cls_perc_eval          0.004841   0.001441   3.359 0.000849 ***
## cls_creditsone credit  0.472669   0.110652   4.272 2.37e-05 ***
## bty_avg                0.043578   0.017257   2.525 0.011903 *
## pic_outfitnot formal  -0.136594   0.068998  -1.980 0.048347 *
## pic_colorcolor        -0.189905   0.067697  -2.805 0.005246 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4975 on 451 degrees of freedom
## Multiple R-squared:  0.1832, Adjusted R-squared:  0.1632
## F-statistic: 9.193 on 11 and 451 DF,  p-value: 6.364e-15
```

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

```
##
## Call:
## lm(formula = score ~ ethnicity + gender + language + age + cls_perc_eval +
##     cls_credits + bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8455 -0.3221  0.1013  0.3745  0.9051
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.907030   0.244889  15.954  < 2e-16 ***
## ethnicitynot minority  0.163818   0.075158   2.180 0.029798 *
## gendermale             0.202597   0.050102   4.044 6.18e-05 ***
## languagenon-english   -0.246683   0.106146  -2.324 0.020567 *
## age                   -0.006925   0.002658  -2.606 0.009475 **
## cls_perc_eval          0.004942   0.001442   3.427 0.000666 ***
## cls_creditsone credit  0.517205   0.104141   4.966 9.68e-07 ***
## bty_avg                0.046732   0.017091   2.734 0.006497 **
## pic_outfitnot formal  -0.113939   0.067168  -1.696 0.090510 .
## pic_colorcolor        -0.180870   0.067456  -2.681 0.007601 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4982 on 453 degrees of freedom
## Multiple R-squared:  0.1774, Adjusted R-squared:  0.161
## F-statistic: 10.85 on 9 and 453 DF,  p-value: 2.441e-15
```

```
# removing pic_outfit
m_inter <- lm(score ~ ethnicity + gender + language + age + cls_perc_eval
             + cls_credits + bty_avg
             + pic_color, data = evals)
summary(m_inter)
```

```
##
## Call:
## lm(formula = score ~ ethnicity + gender + language + age + cls_perc_eval +
##     cls_credits + bty_avg + pic_color, data = evals)
##
## Residuals:
```

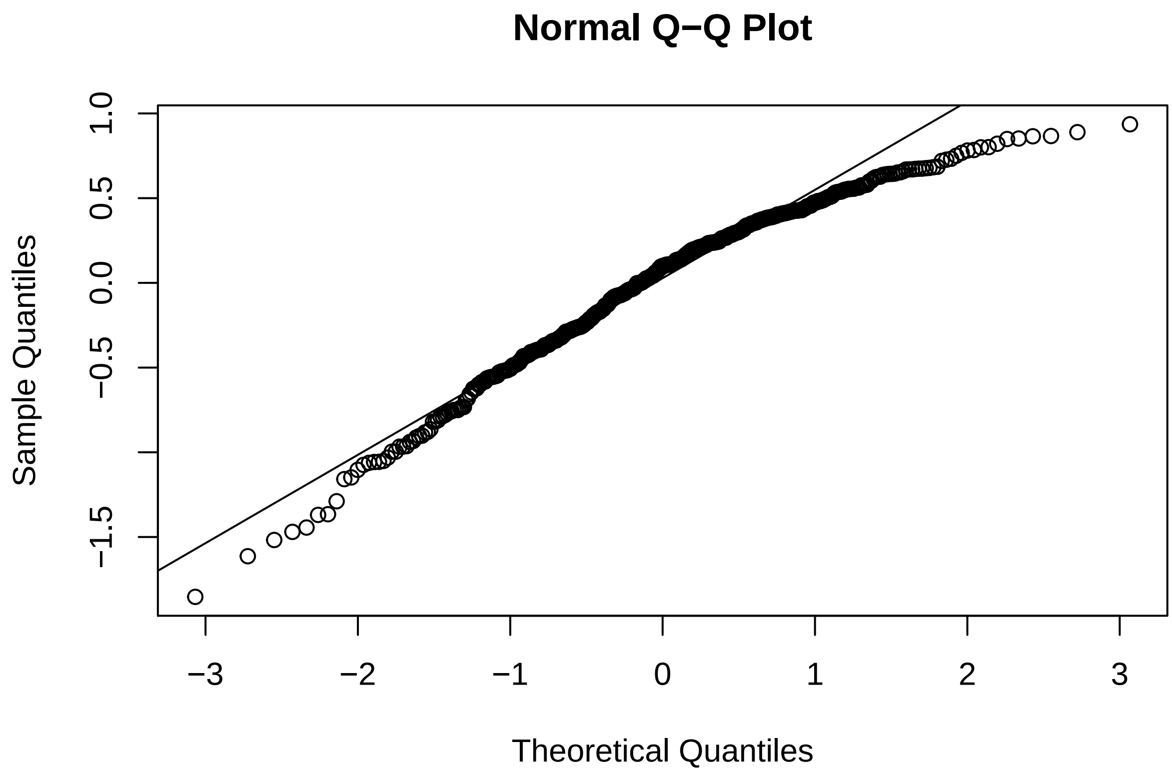
```
##      Min       1Q   Median       3Q      Max 
## -1.85320 -0.32394  0.09984  0.37930  0.93610 
## 
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            3.771922   0.232053  16.255  < 2e-16 ***
## ethnicitynot minority  0.167872   0.075275   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
```

The following is the linear model for predicting the score:

$\hat{score}$ = 3.7719215 + 0.1678723 * ethnicitynot minority + 0.2071121 * gendermale + -0.2061781 * languagenon-english + -0.0060459 * age + 0.0046559 * cls_perc_eval + 0.5053062 * cls_creditsone credit + 0.0510693 * bty_avg + -0.1905788 * pic_colorcolor

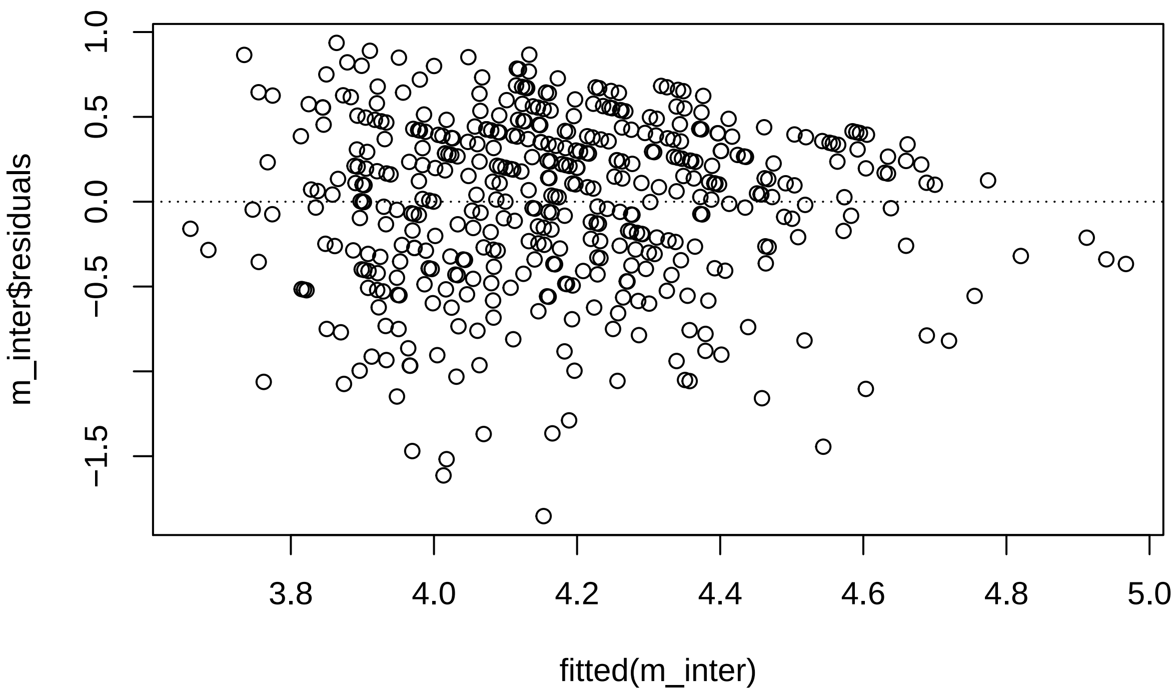
16. Verify that the conditions for this model are reasonable using diagnostic plots.

```r
qqnorm(m_inter$residuals)
qqline(m_inter$residuals)
```



Based on the plot above, we can say that the distribution is nearly normal.

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



Based on the above plot we can assume the condition for constant variability also holds good.

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?

By not sampling for all courses taught in the university (rather sampling only all courses by some sample professors), I can assume that the results may be biased. There may be some courses, maybe fine arts, where beauty may have an impact on the scores. Similarly there may be some courses where students may be impervious to beauty.

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.

Based on the final model, professors who have the following characteristics will get a higher score:

A professor who is: Not Minority, is Male, is English Educated, has Lower Age, has More percent of students in class who completed evaluation, teaching One credit courses, having higher beauty average and having a color picture

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

We cannot generalize this to all universities. The samples are specific to University of Texas, Austin. If samples were collected from different universities across the country and the study conducted then we could have generalized it to all universities