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 a slightly 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.

variable	description
`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.

variable	description
`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.

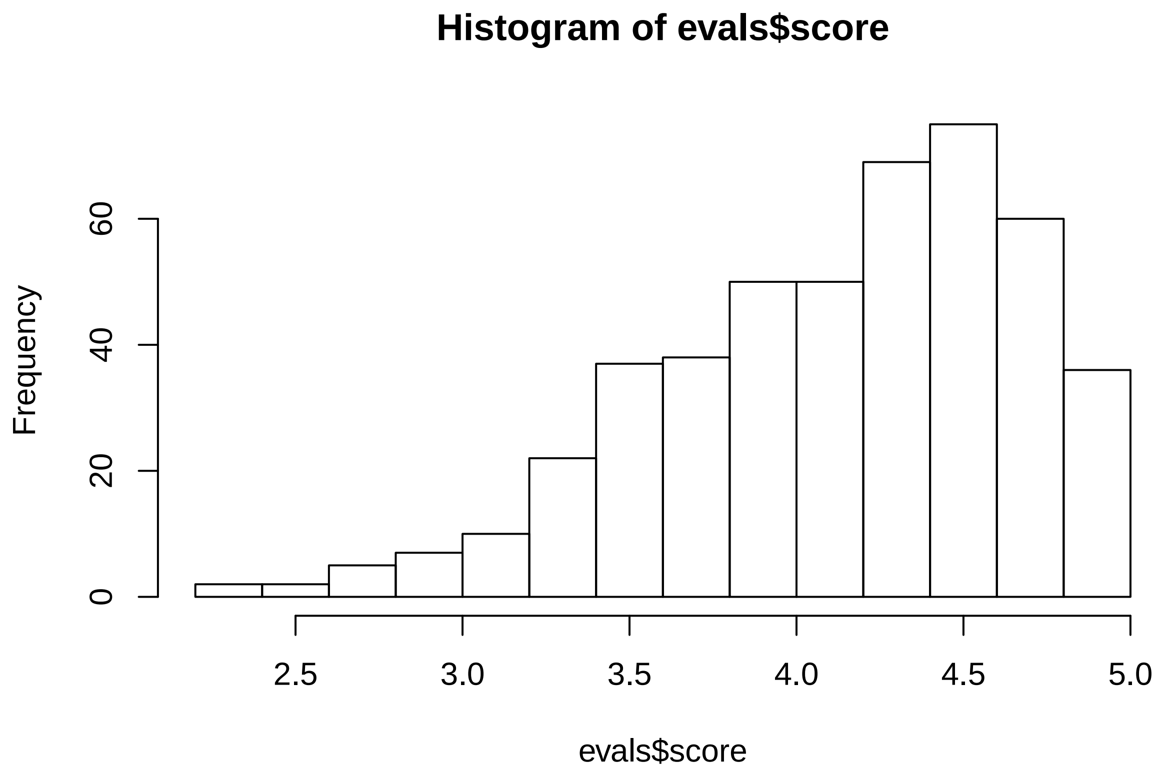
Exploring the data

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.

This appears to be an observational study since we do not have a defined control and experimental group. Since this is an observational study we cannot imply causation, so whether or not beauty leads directly to differences in course evaluations isn't an appropriate question for this design. Instead, we could ask, "Is the professor's beauty correlated with positive or negative student course evaluations?"

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?

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

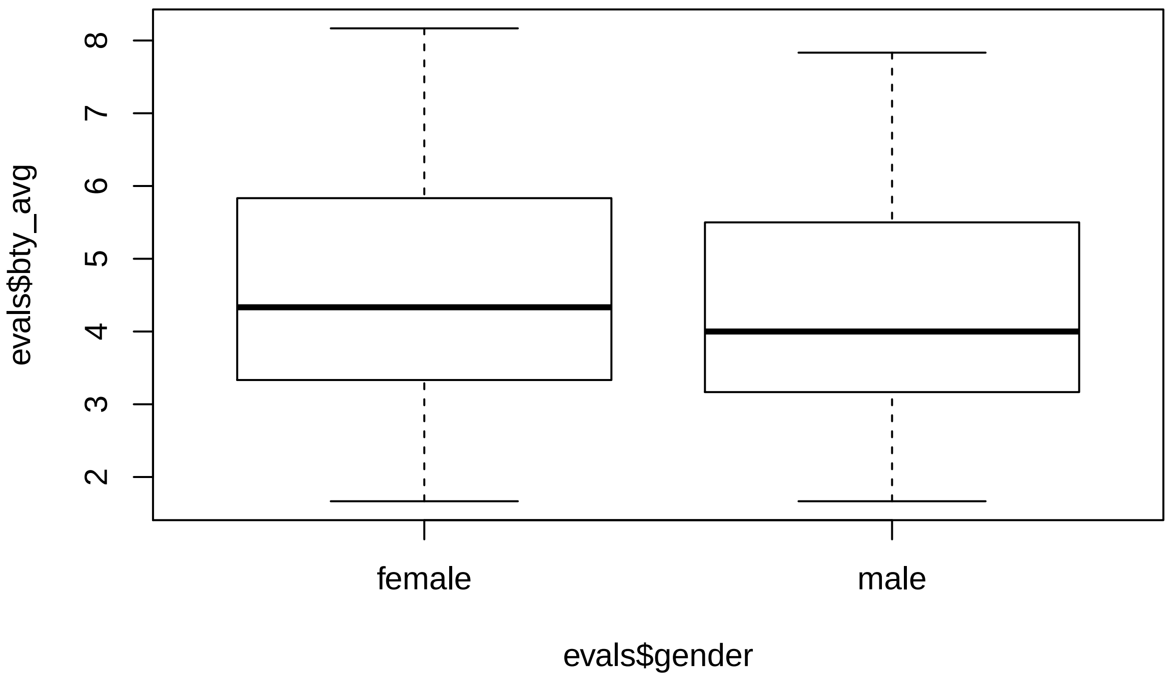


After plotting the distribution of scores, we can see that the student ratings are left skewed. This tells us that generally, students rate professors more positively than they do negatively. This is a bit surprising to me given that I was expecting students to be a bit more critical about their professors, or at least distribute the scores reflecting more of a normal distribution, but it seems that overall, students appear largely satisfied with their instructors.

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

Excluding 'score' I decided to create a scatterplot of 'age' and 'bty_avg', to see if there is a relationship between the professor's age and the student's average beauty rating.

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



By looking at the boxplot above of student's beauty rating by gender, it appears that females may have a higher median beauty rating than males. However, more statistical testing would be needed to confirm that is true.

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:

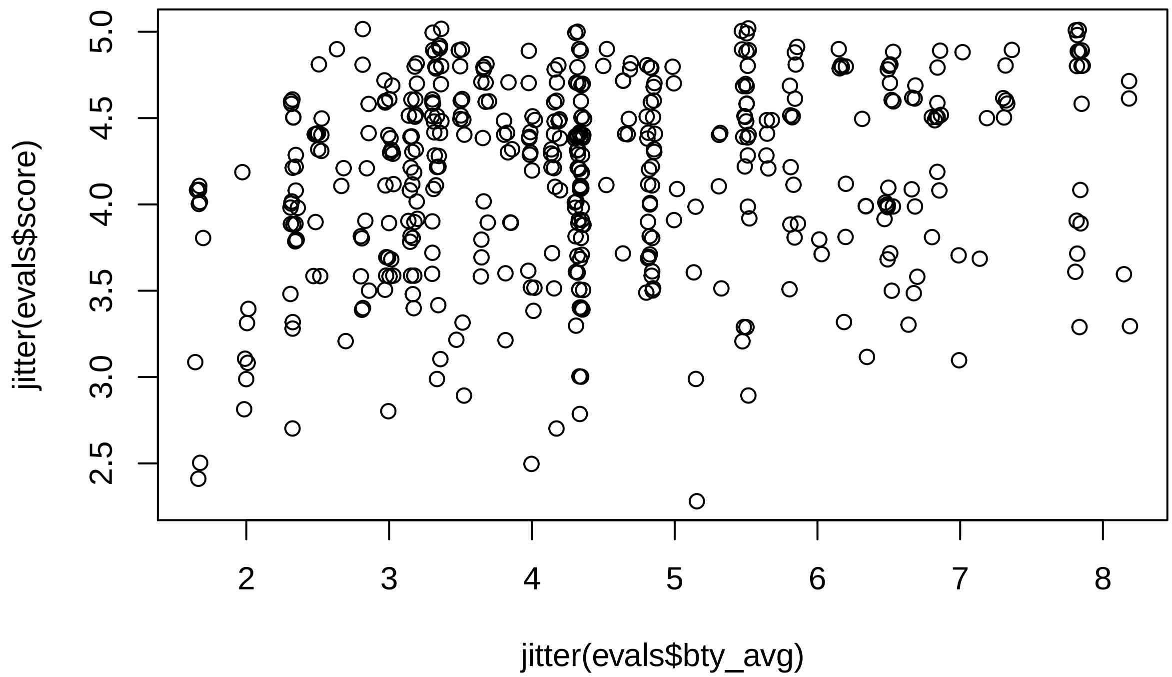
```
plot(evals$score ~ evals$bty_avg)
```

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?

It looks like there are more observations than the number of points on this scatterplot. From the dataset, it appears that there should be 463 points, but the scatterplot above doesn't seem to represent all of these points.

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?

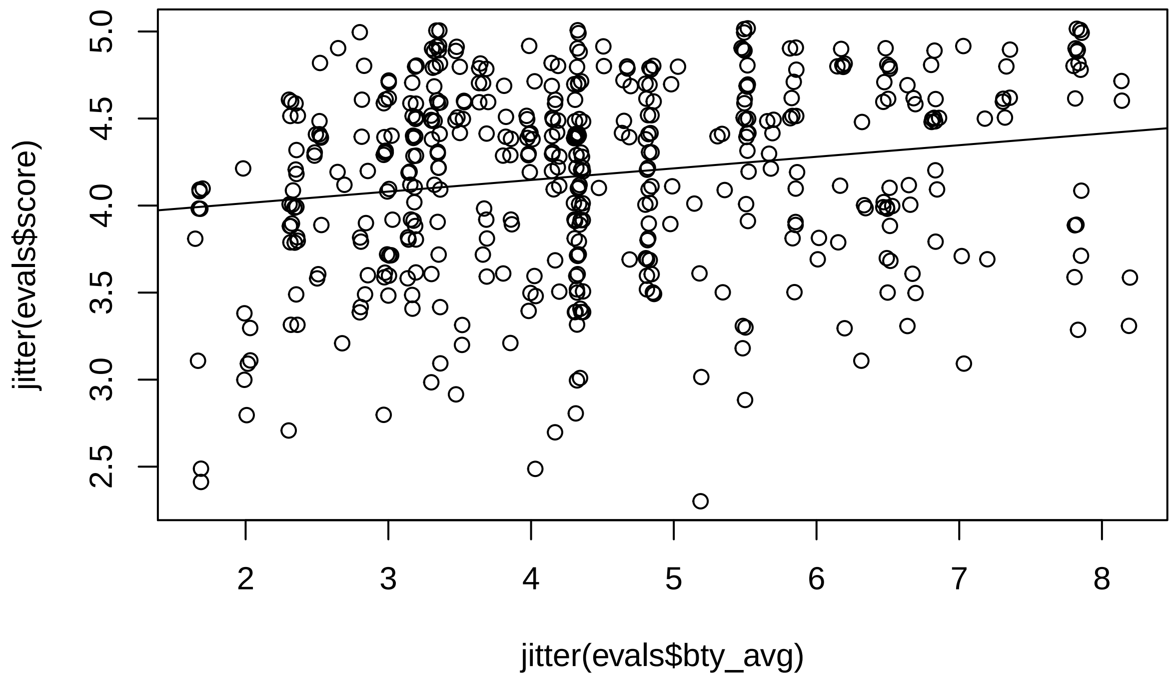
```
plot(jitter(evals$score) ~ jitter(evals$bty_avg))
```



It appears that many of the beauty scores were equal to one another, so when plotting this on a two-dimensional plot, we were only able to see one instance of each beauty score without taking into consideration the frequency of these equal scores. By using jitter, we can now see where there is overlap in scores on the scatterplot.

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)
```



To find the equation for the linear model, we can do the following:

```
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
```

After identifying the intercept and slope, the equation can be written as $\hat{score} = 3.88034 + 0.06664 \times AverageBeauty$**.**

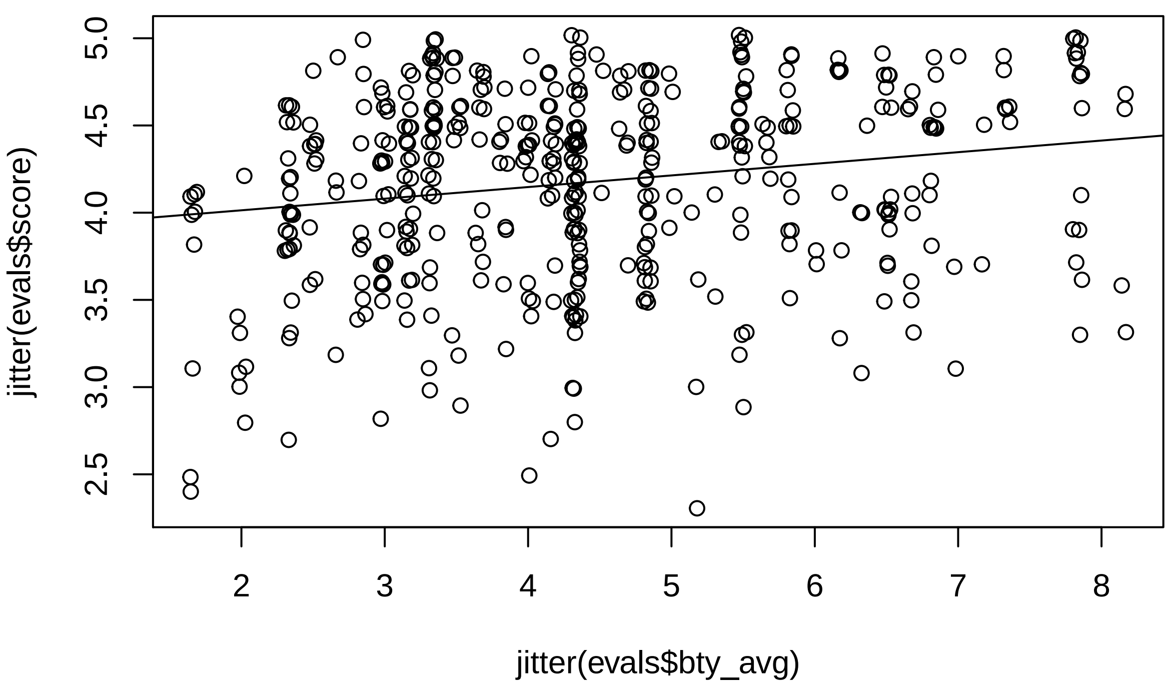
By looking at the p-value from the summary, we can see that it is less than 0.001, and very close to zero, which indicates that the slope is not equal to zero. In other words, the beauty average is a statistically significant predictor of their evaluation score. However, it doesn't really seem to be a very practical predictor given that the adjusted R-squared value is very low (0.03293). This means that very little of the variability in the evaluation score can be accounted for by the professor's beauty score (around 3%).

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).

In order to determine whether the conditions of least squares regression are reasonable, we need to confirm the following:

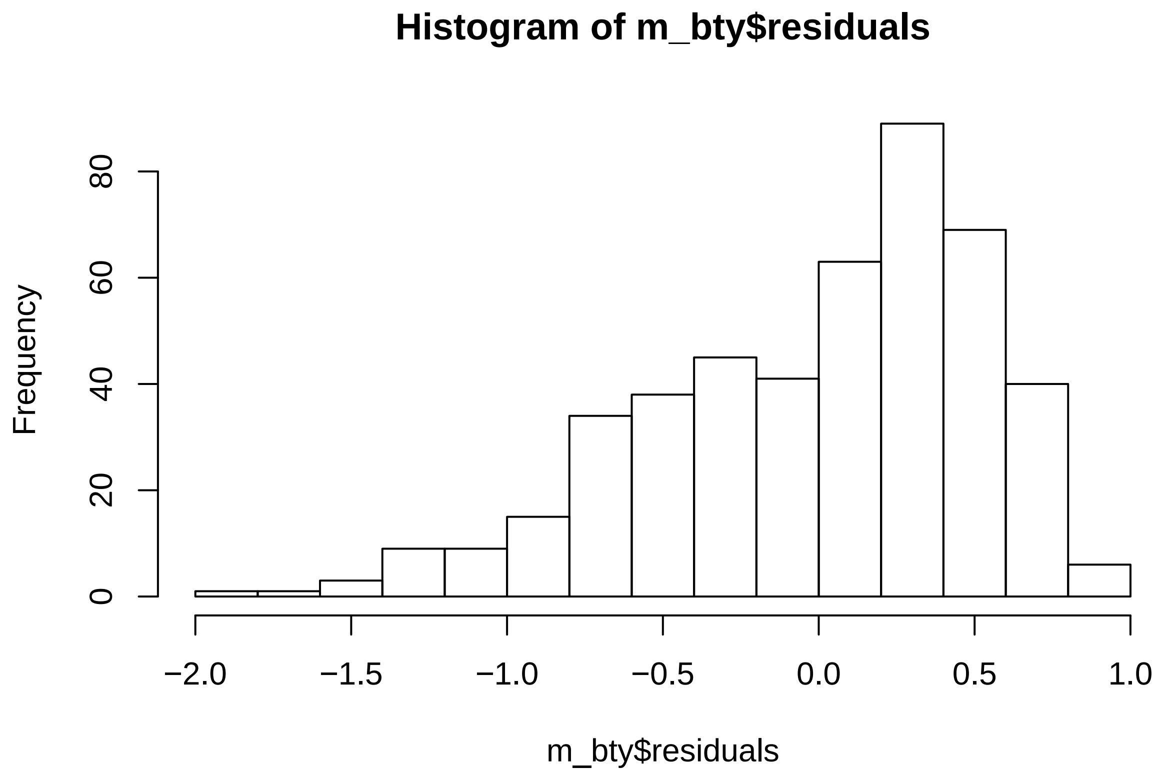
- **Linearity – from the plot below, this relationship does appear to show a linear trend.**

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



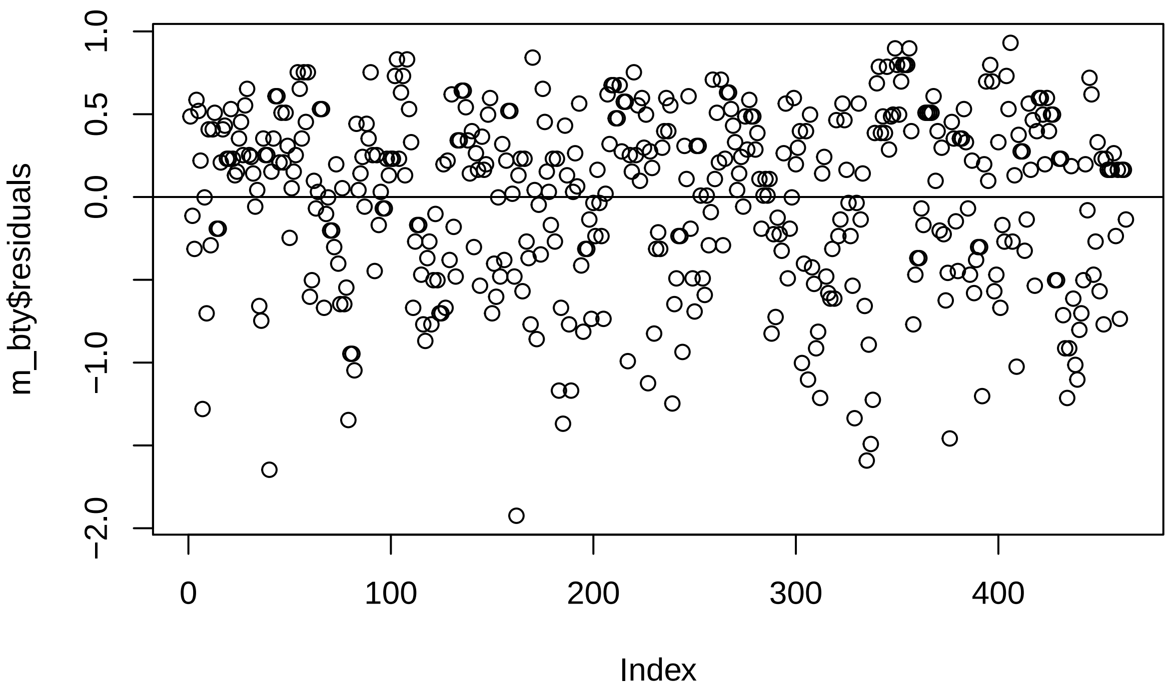
- **Nearly normal residuals – by looking at the histogram of residuals, it appears that the distribution is a bit left skewed, but is somewhat approaching a normal distribution. If some outliers are removed, it would appear to be close to a normal distribution of residuals.**

```r
hist(m_bty$residuals, breaks = 20)
```



- **Constant variability – there does appear to be pretty constant variability of the residuals when looking at the residuals plot. There doesn't seem to be a lot of clustering, and the residuals seem to be scattered fairly equally, indicating constant variability.**

```r
plot(m_bty$residuals)
abline(h = 0)
```



- **Independent observations – this is not time series data, and each observation is independent from one another.**

From the plots and tests, it appears that the only test that I'd be a little worried about when evaluating whether the conditions of least squares regression are reasonable would be the distribution of residuals, and it being nearly normal. There is a pretty strong indication that these residuals are left skewed. However, moving some outliers may help.

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)
```

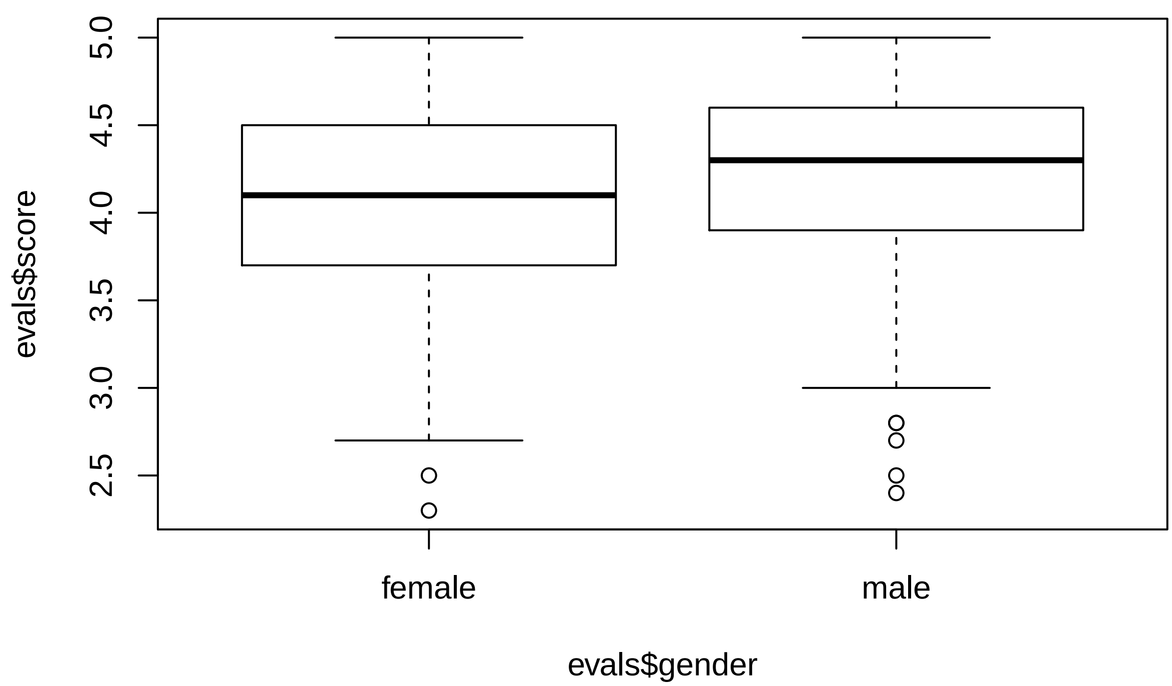
```
##
## Call:
## lm(formula = score ~ bty_avg + gender, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8305 -0.3625  0.1055  0.4213  0.9314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.74734    0.08466  44.266  < 2e-16 ***
## 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.

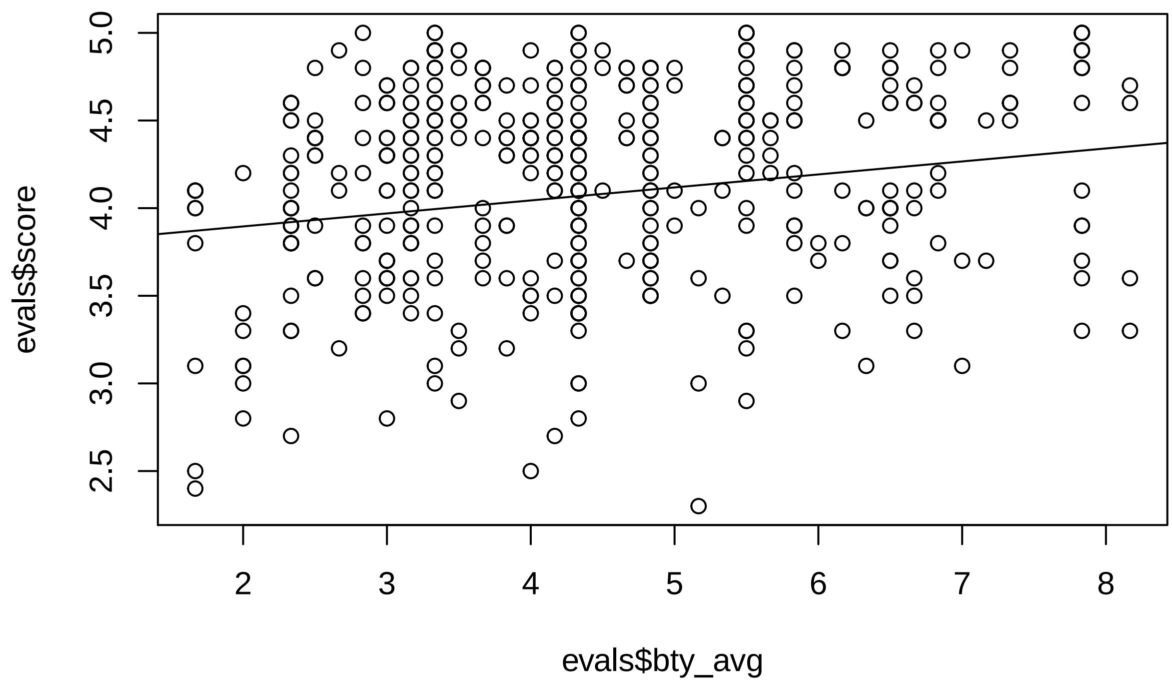
Similar to question 6, I'll check the same conditions:

- **Linearity – from the plots below, this relationship does appear to show a linear trend for both variables and their relationship with 'scores'. It looks like the median score is higher for males than females (p value is less than 0.05, indicating the slope is not equal to 0 and has a positive value of 0.14151), and takes on a linear increase, while this is also shown with the average beauty score as well.**

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




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




```r
lm_gend <- lm(evals$score ~ evals$gender)
summary(lm_gend)
```

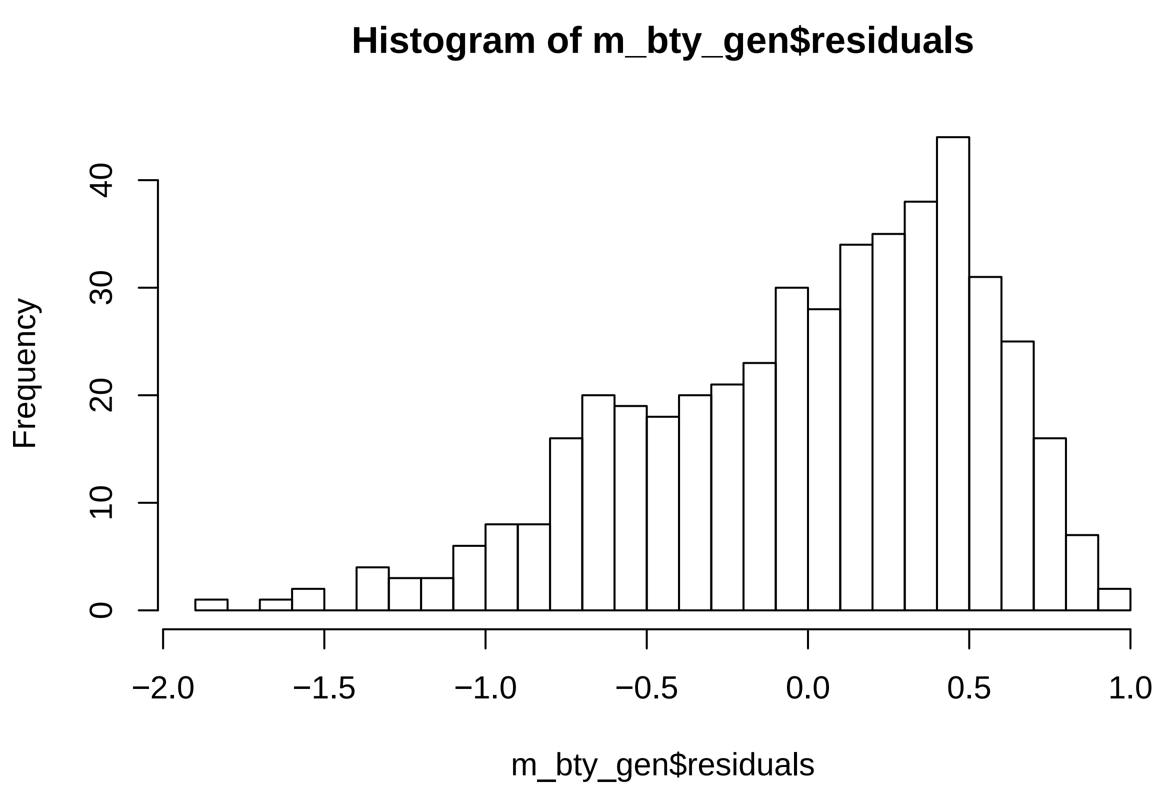
```
##
## Call:
## lm(formula = evals$score ~ evals$gender)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.83433 -0.36357  0.06567  0.40718  0.90718
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.09282    0.03867 105.852  < 2e-16 ***
## evals$gendermale  0.14151    0.05082   2.784  0.00558 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5399 on 461 degrees of freedom
## Multiple R-squared:  0.01654,    Adjusted R-squared:  0.01441
## F-statistic: 7.753 on 1 and 461 DF,  p-value: 0.005583
```

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

```
##
## 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
```

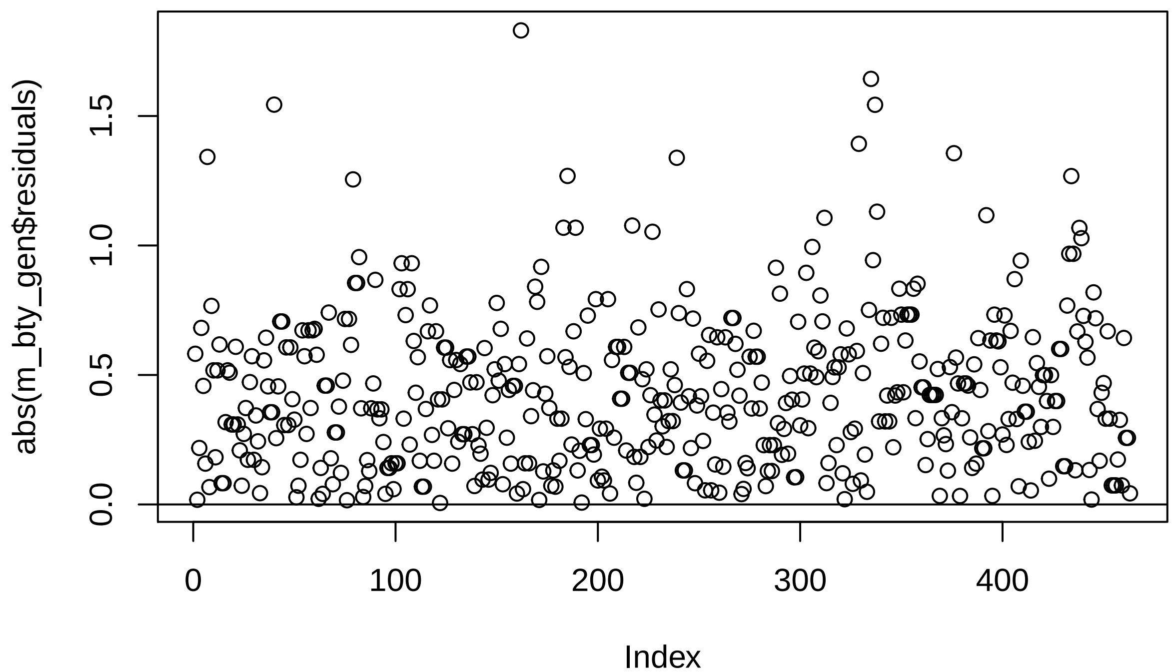
- **Nearly normal residuals (checking outliers) – by looking at the histogram of residuals, it appears that the distribution is a bit left skewed, similar to question 6, but is somewhat approaching a normal distribution. If some outliers are removed, it would appear to be close to a normal distribution of residuals.**

```r
hist(m_bty_gen$residuals, breaks = 20)
```



- **Constant variability – there does appear to be pretty constant variability of the residuals when looking at the residuals plot. There doesn't seem to be a lot of clustering, and the residuals seem to be scattered fairly equally, indicating constant variability.**

```
plot(abs(m_bty_gen$residuals))
abline(h = 0)
```



- **Independent observations – this is not time series data, and each observation is independent from one another.**

From the plots and tests, it appears that the conditions for regression are met, with the exception of a normal distribution of residuals. However, one could argue that because the distribution is unimodal and could be approaching normal, then it meets the condition.

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, bty_avg is still a signficant predictor of 'score'. To check to see if the addition of 'gender' to the model has changed the parameter estimate for bty_avg, we can use the following:

```
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
```

We can see that the addition of 'gender' has changed the parameter estimate for 'bty_avg' from 0.06664 to 0.17239, and the p-value has decreased from 5.083×10^{-5} to 8.177×10^{-7}. It also appears that adding gender made the model account for slightly more of the score's variability, from roughly 3% to 5%.

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 equation of the line corresponding to males is $\hat{score}_{male} = 3.74734 + 0.07416 \times bty_{average} + 0.17239$

For two professors who received the same beauty rating, the gender that tends to have the higher course evaluation score is males.

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`.

```r
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
```

After outputting the summary of the new model, we can see that it creates dummy variables similar to two levels, based on alphabetical order. Therefore, it appears that 'teaching', since it's first in alphabetical order is given the dummy variable of 0, while 'tenure track' and 'tenured' are given 1 when being compared to 'teaching' and then given 0 when being compared amongst each other.

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 expect that the number of credits variable would have the highest p-value, since it doesn't seem to have any association with the professor's 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)
```

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

It appears that my suspicions were very wrong, it looks like professors teaching a one credit class has a very low p-value $p = 1.84 \times 10^{-5}$.

13. Interpret the coefficient associated with the ethnicity variable.

With a coefficent value of 0.1234929, we can see that if all other variables remain constant, if a professor is categorized as a non-minority, then their score, on average will increase by 0.12.

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?

It appears that the number of professors had the highest p-value. After dropping this value, I'll re-fit the model below:

```
m_full_drop1 <- 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_drop1)
```

```
##
## 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
```

After removing the number of professors variable from the model, it does appear that the other coefficients and significance values of all other variables did change. Since they did change, it shows that there wasn't a lot of colinearity of the number of professors variable with others in the model.

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.

Using the step function, I was able to use backward-selection and p-value as the selection criterion to determine the best model.

```
step(m_full, direction = "backward", trace = TRUE)
```

The best model that I could create from the existing variables is below:

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

```
##
## 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
```

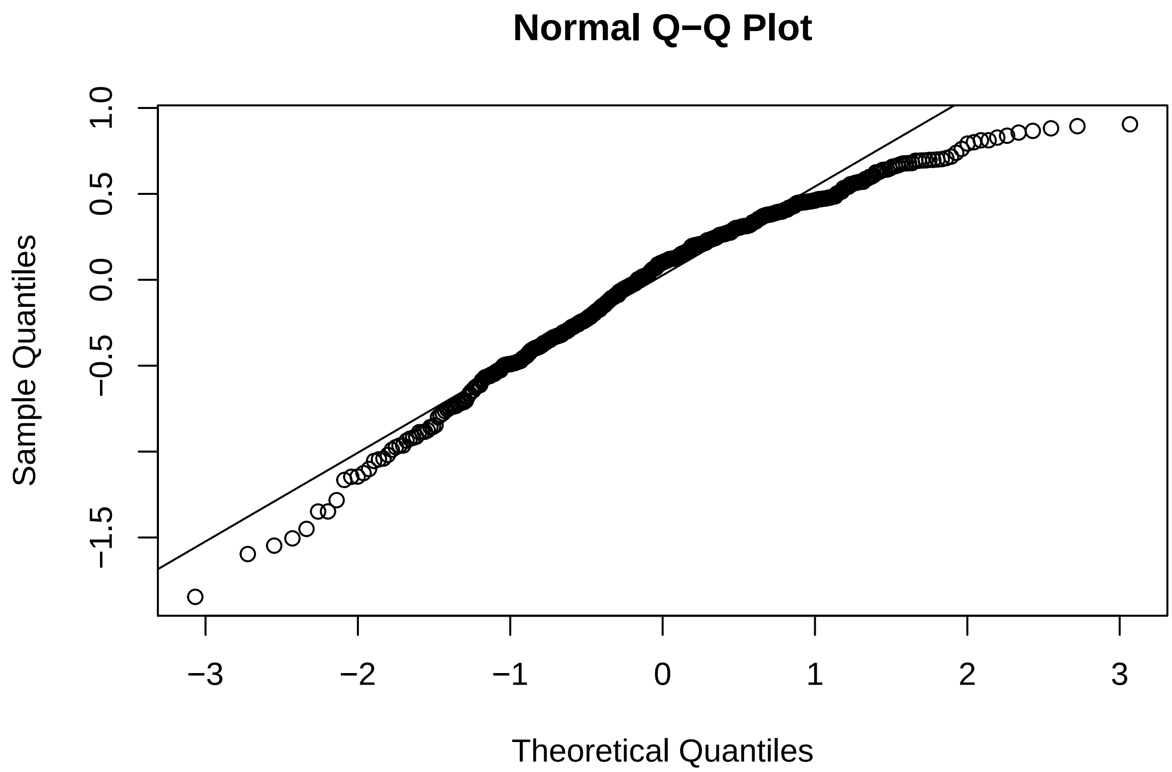
```
## -1.8455 -0.3221  0.1013  0.3745  0.9051
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.907030   0.244889  15.954  < 2e-16 ***
## ethnicitynot minority  0.163818   0.075158   2.180 0.029798 *
## gendermale             0.202597   0.050102   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
```

The equation for this model can be written as $\hat{score} = 3.90 + (0.16 * ethnicity) + (0.20 * gender) - (0.25 * language) - (0.007 * age) + (0.005 * cls_{perceval}) + (0.52 * cls_{credits}) + (0.05 * bty_{avg}) - (0.11 * pic_{outfit}) - (0.18 * pic_{color})$

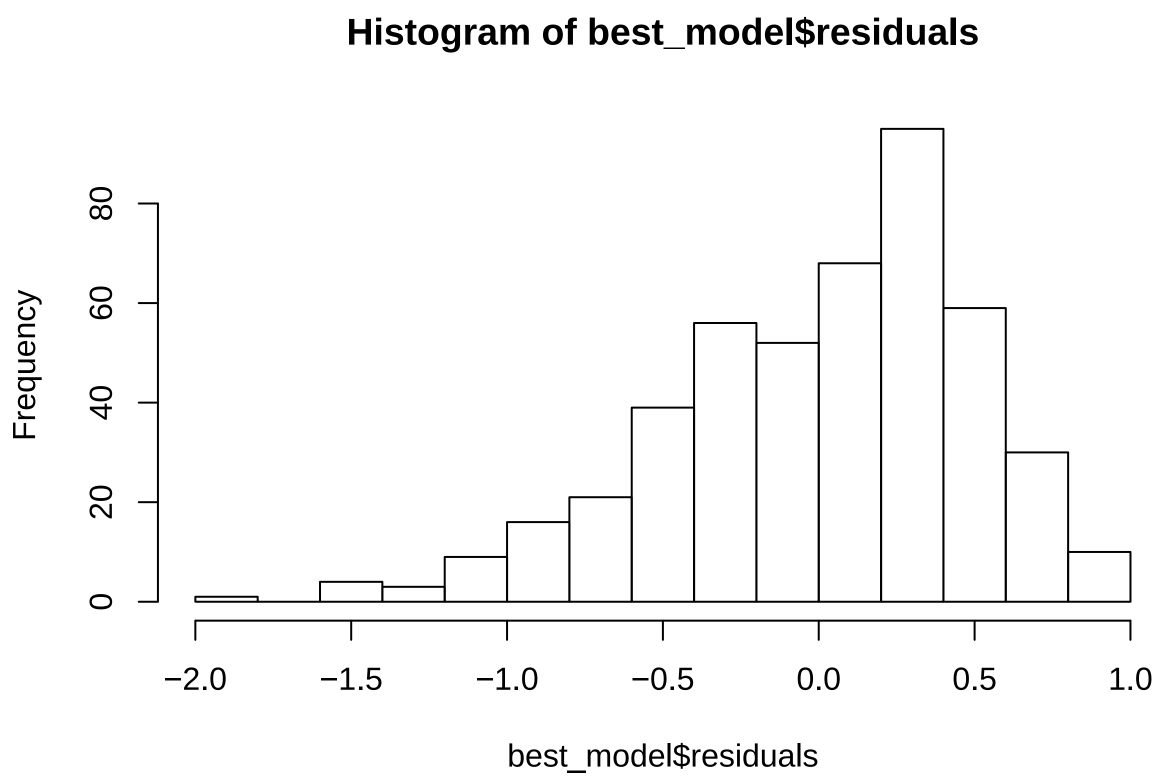
16. Verify that the conditions for this model are reasonable using diagnostic plots.

To verify that the conditions for this model are reasonable, we'll need to check the following:

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

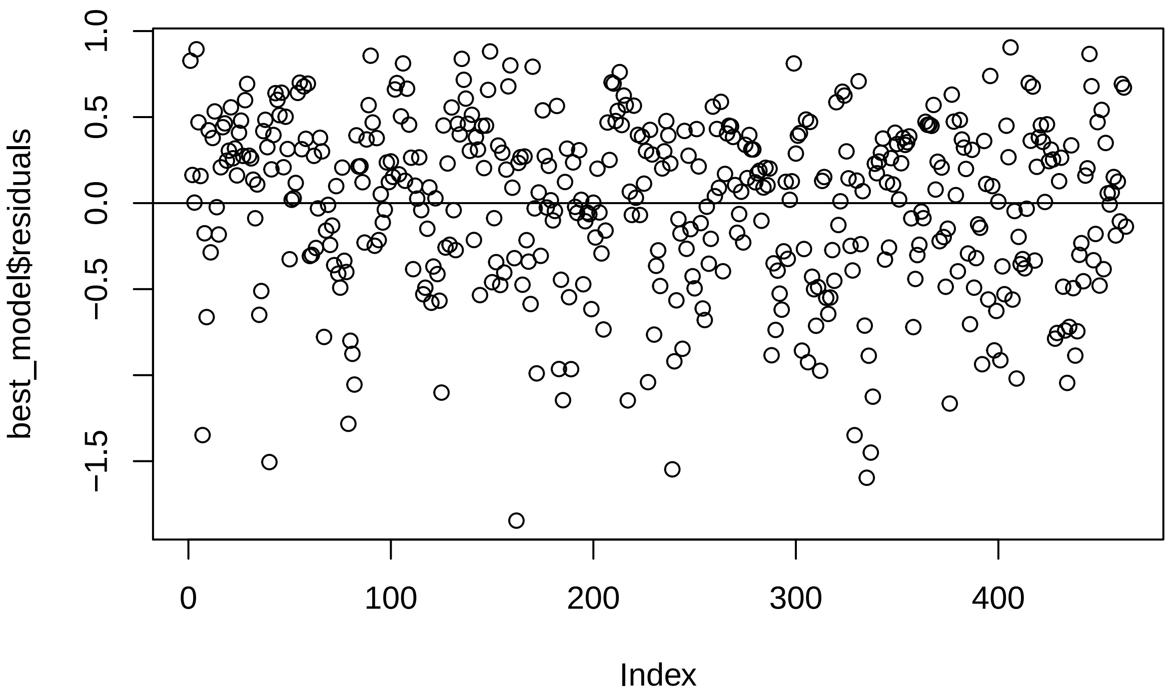



```r
hist(best_model$residuals)
```



Given that the histogram and q-q plot seem to indicate that the distribution is a bit left skewed, I worry that the condition for this model is reasonable. There does appear to be a few outliers here. This is similar to my concern in the above exercises and models.

```
plot(best_model$residuals)
abline(h = 0)
```



The residuals do seem to have pretty constant variability, and the observations seem to be independent from one another. We have also been able to fit a linear model to this data, so the rest of these diagnostic tests seem reasonable – with the exception of a nearly normal residual distribution. Because of this last case, I'd argue that this is not reasonable.

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?

Considering that each row represents a course, this new information will have an impact on the independence of observations criteria.

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 my final model, the characteristics of a professor and course at the University of Texas at Austin that would be associated with a high evaluation score would be a professor that is:

- **Not a minority in identified ethnicity**
- **A male**
- **Graduated from an English-speaking school**
- **Is younger in age**
- **Has a high percentage of students fill out evaluations**

- **Teaches a one-credit course**
- **Has a high average beauty rating**
- **Wears formal clothing in their picture**
- **Has a black and white picture**

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

I would not feel comfortable generalizing my conclusions to apply to professors generally (at any university). There may be significant regional variations in perceptions of course/professor ratings – especially when thinking about things like different educational/cultural norms by region, demographics of students, types of courses offered at universities, etc.