Multiple linear regression

Jai Jeffryes

11/27/2019

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.

```r
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.

variable	description
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.

ANSWER

This is an observational study, and there are problems with the design.

One

The fundamental problem with this design is the disconnection between class evaluation and beauty score, for the beauty score is assigned by someone different from the person evaluating the class.

Two

Two thirds of the courses have multiple instructors, yet their class evaluations are commingled as well as their beauty scores.

Three

There could be selection bias. The percentage of students who evaluated each class ranges from 10% to 100%.

Four

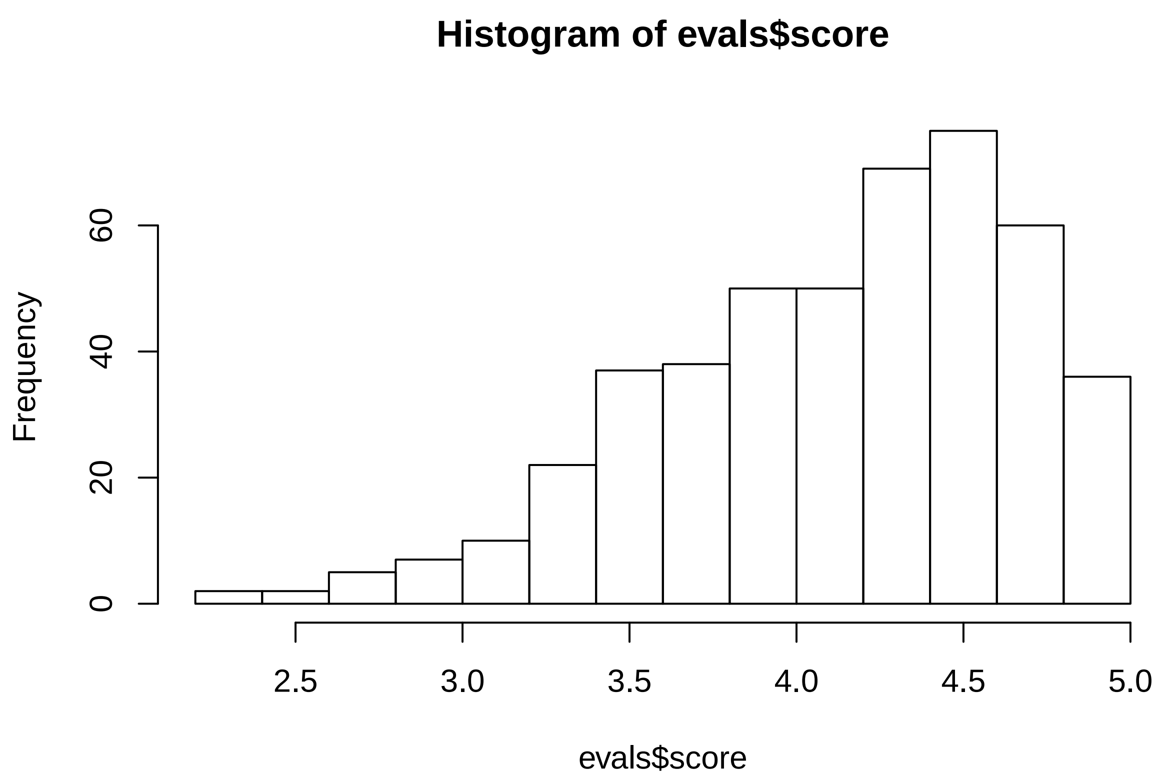
I would like the students evaluating the class also to record a beauty rating. However, if I accepted that, there are still for too few, beauty raters, namely six. By contrast, there are 17,000 class evaluations. Presumably, there are duplicate students among those 17,000 class evaluations, but six raters for beauty is far too few.

Rather than determining a direct connection between beauty and course evaluation, this study may observe a correlation between student course evaluations and third party perception of beauty.

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)
```



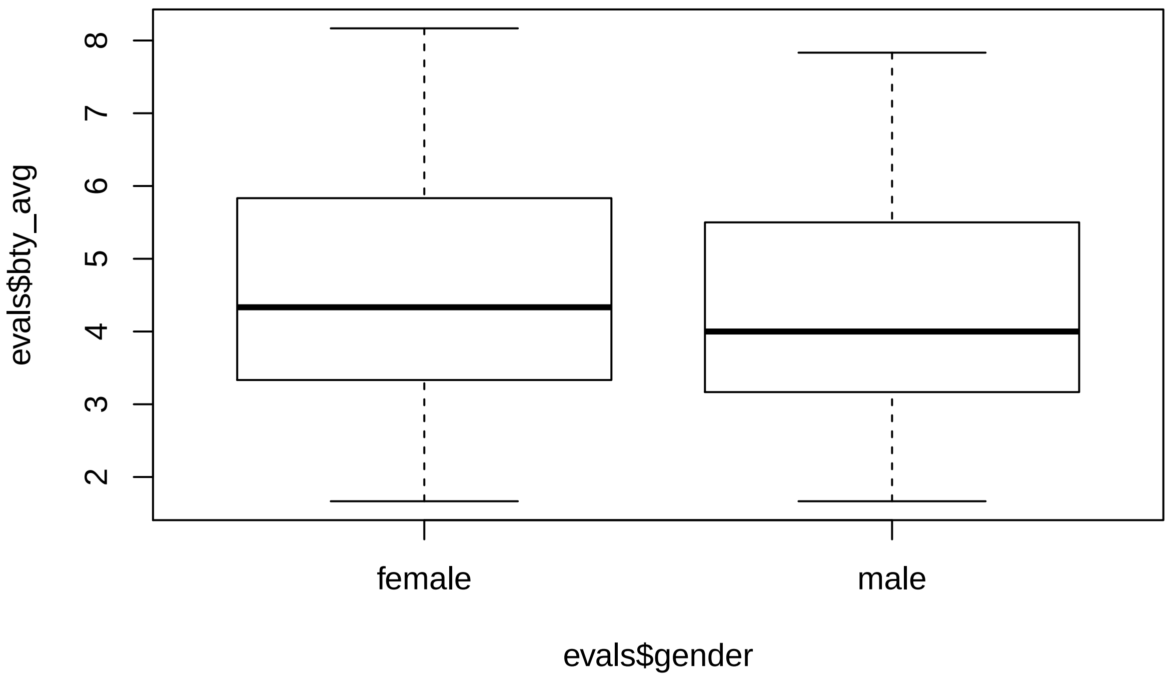
The distribution is symmetric, unimodal, and skewed to the left. It's central tendency is 4.5 out of a positive rating score of 5. I expect this central tendency, since I expect students to rate their instructors well. Otherwise, I expect the administration to take corrective action to achieve better performance.

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

ANSWER

`bty_avg` and `gender`. Women appear to inspire a higher perception of beauty than men.

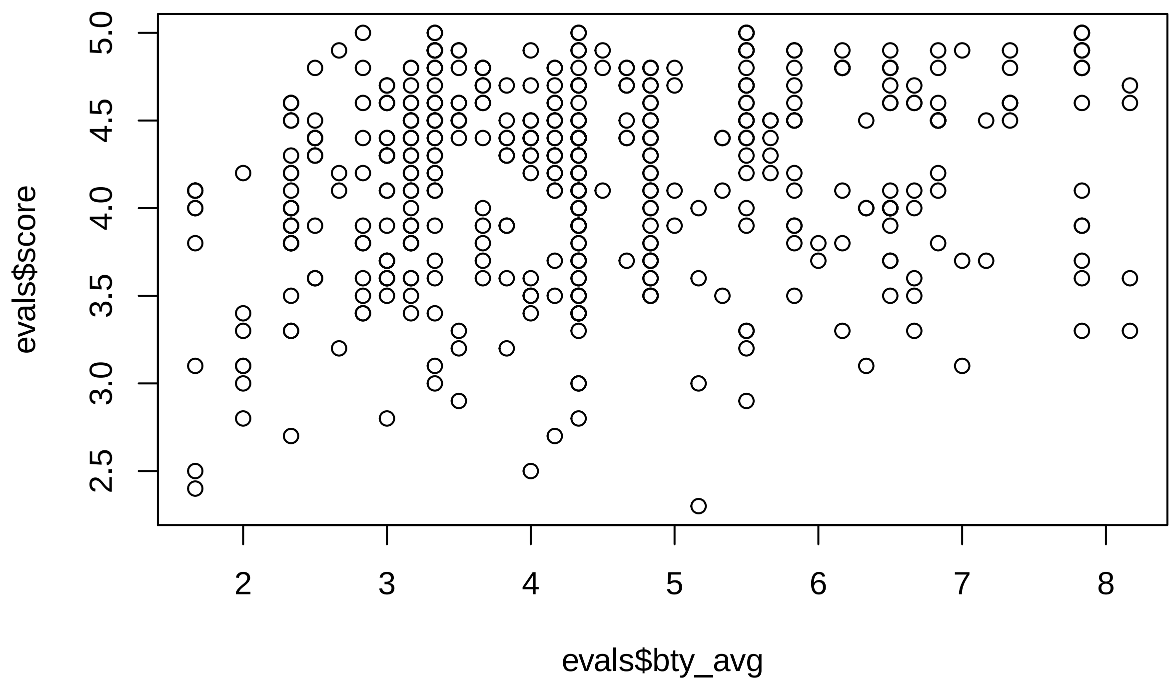
```
boxplot(evals$bty_avg ~ evals$gender)
```



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?

ANSWER

It seems like a leading question suggesting I'm supposed to notice something awry. However, I don't see anything. Yes, there appear to be fewer than 463 points on the plot. I looked for NA values in the data frame and there weren't any. I counted points on the plot by eye and found about 100 in the right half. All that says to me, given absence of NA values, is many points on this plot must overlap.

```
sum(is.na(evals$score))
```

```
## [1] 0
```

```
sum(is.na(evals$bty_avg))
```

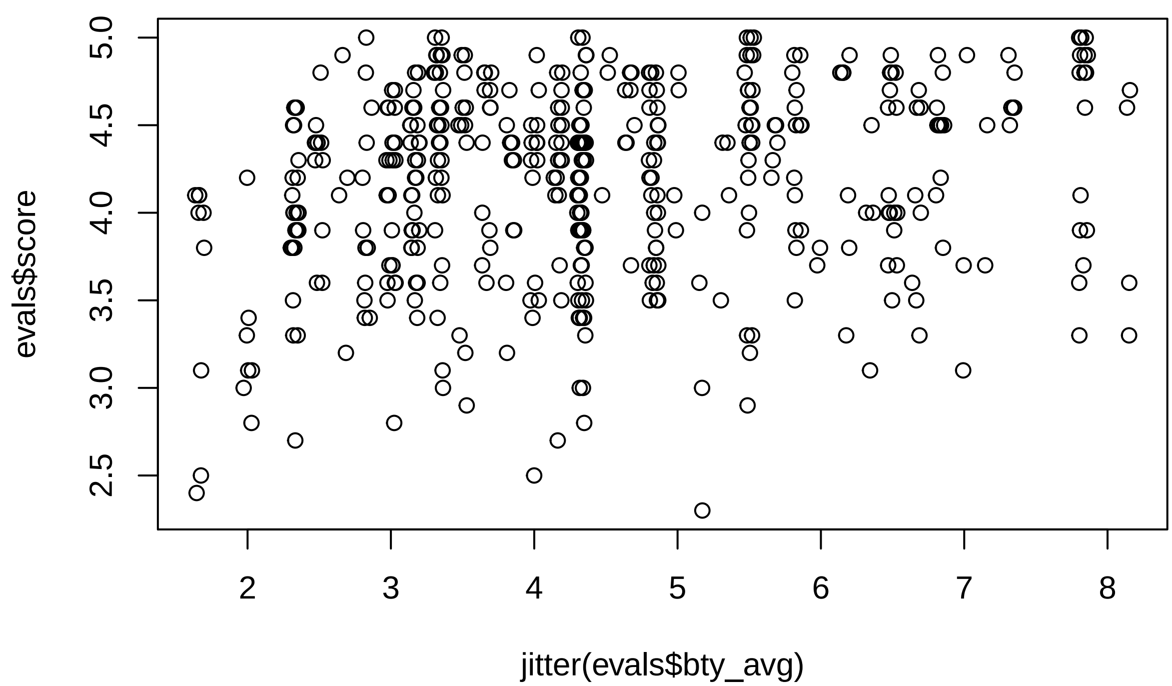
```
## [1] 0
```

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

My guess was right. Many points overlapped.

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



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

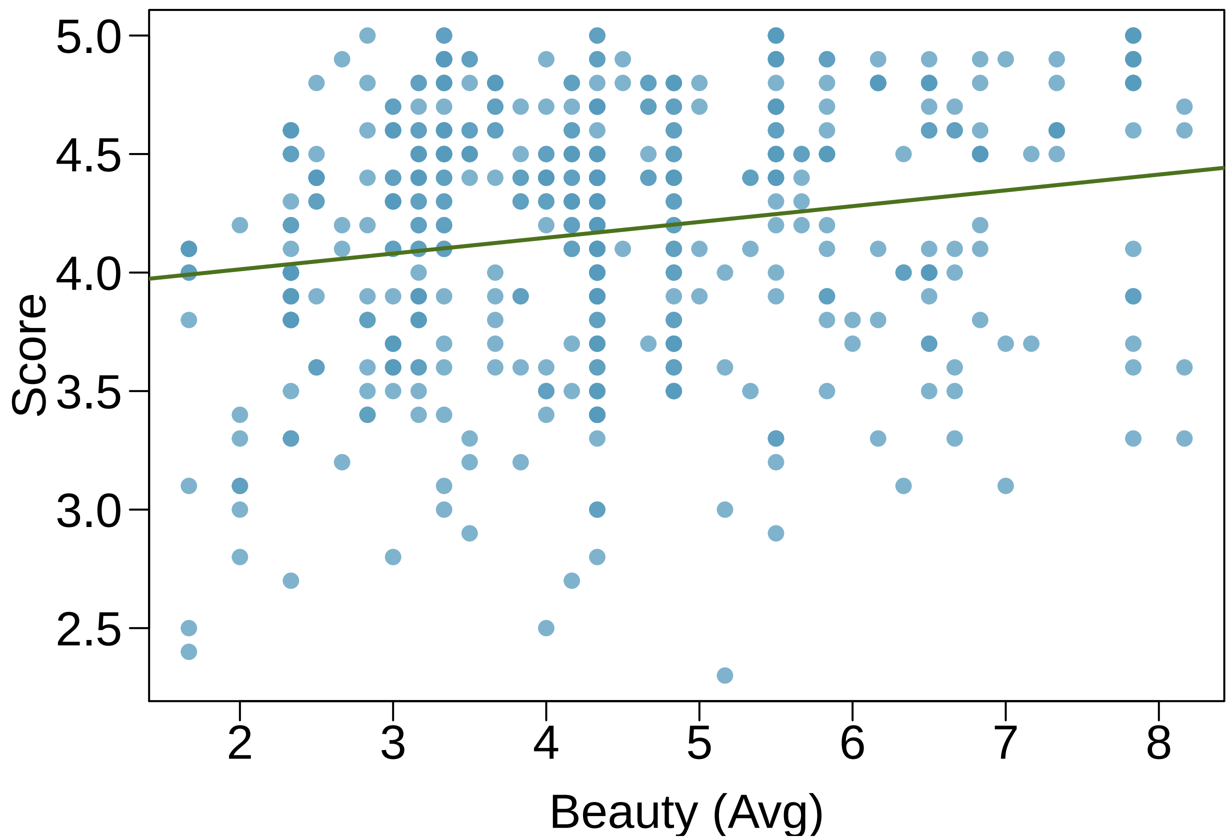
```r
library(openintro)
```

```
## Please visit openintro.org for free statistics materials

##
## Attaching package: 'openintro'

## The following objects are masked from 'package:datasets':
##
##     cars, trees
```

```r
m_bty <- lm(score ~ bty_avg, data = evals)
# plot bty_avg vs. score -----------------------------------------
par(mar = c(3.5, 4, 1, 0.5), las = 1, mgp = c(2.5, 0.7, 0),
   cex.lab = 1.5, cex.axis = 1.5)
plot(score ~ bty_avg, data = evals,
     pch = 19, col = COL[1,2],
     xlab = "Beauty (Avg)", ylab = "Score", axes = FALSE)
axis(1)
axis(2)
# axis(2, at = seq(20, 80, 20))
box()
abline(m_bty, col = COL[2], lwd = 2)
```




```r
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
```

$\widehat{score} = 3.88 + 0.07 * bty\_avg$

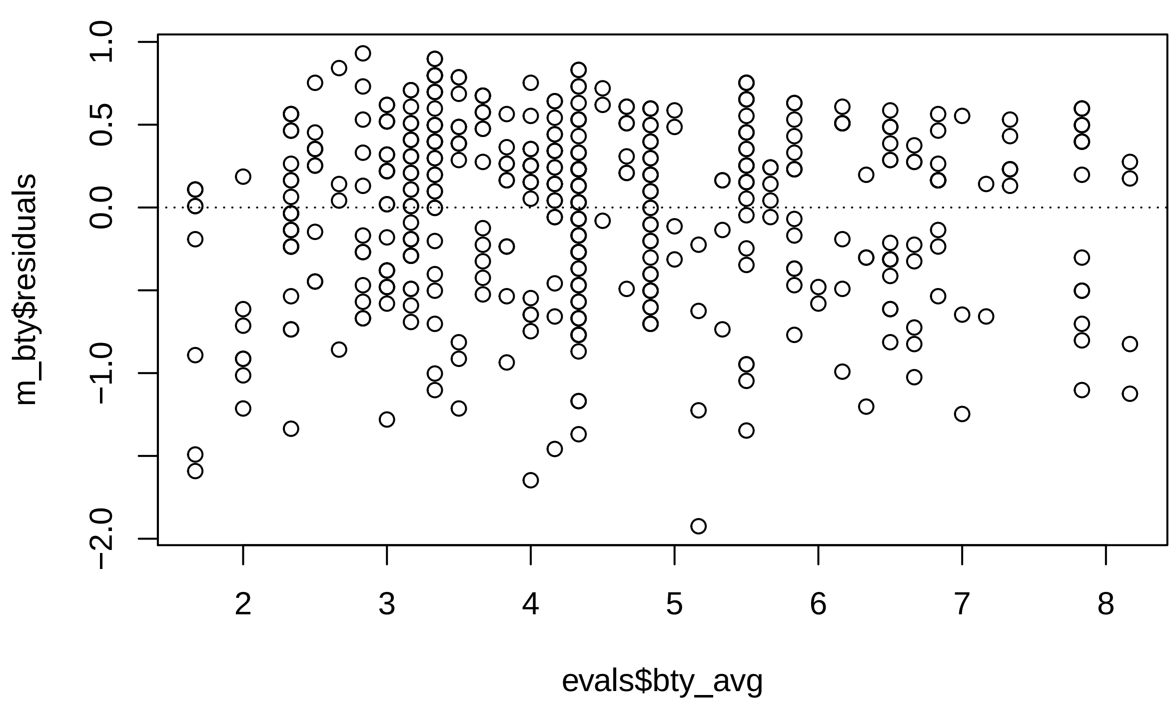
`bty_avg` appears to be a statistically significant predictor, although I noted problems with the experimental design.

`bty_avg` is not a practical predictor. The correlation is pretty flat. At a slope of 0.07, movement of a score by just a quarter point would require a difference in average beauty of almost 4. All other things being equal, an instructor with a beauty rating of 8 would be predicted to have only a quarter point higher course evaluation score than an instructor with a beauty rating of 4.

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).

Linearity

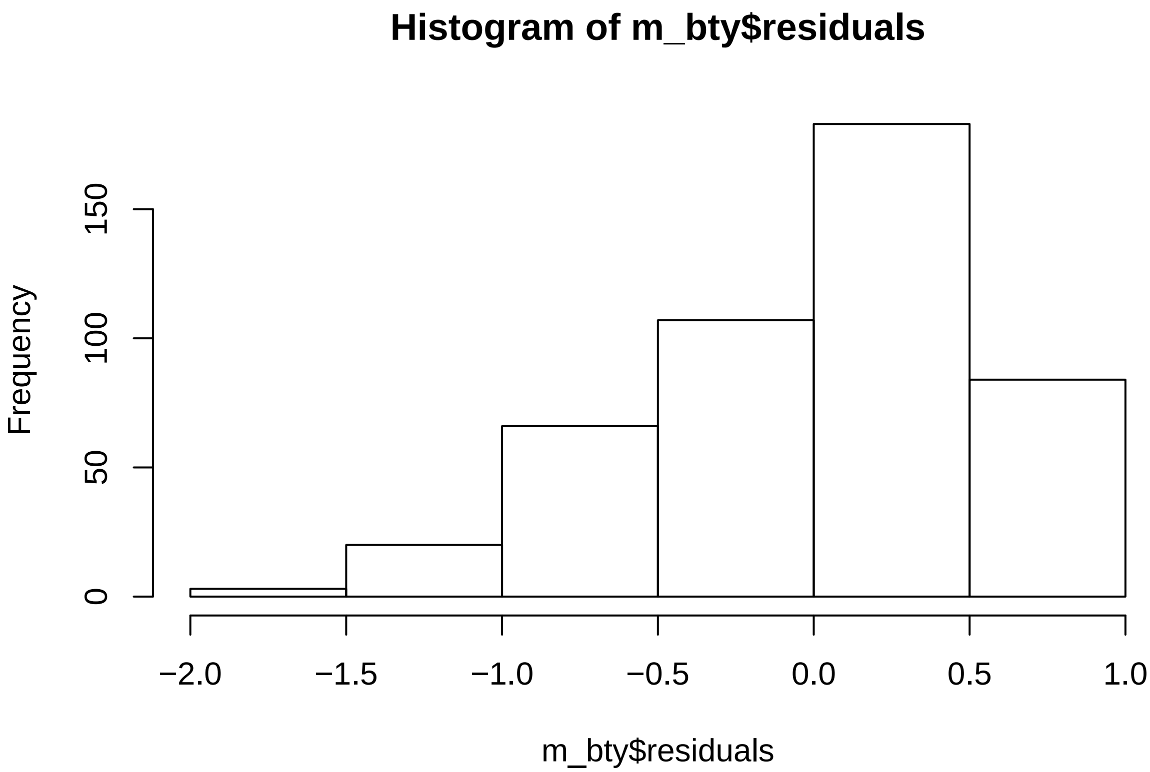
```r
plot(m_bty$residuals ~ evals$bty_avg)
abline(h = 0, lty = 3)
```



I don't see patterns in the residual plot, so it appears to satisfy the requirement of linearity.

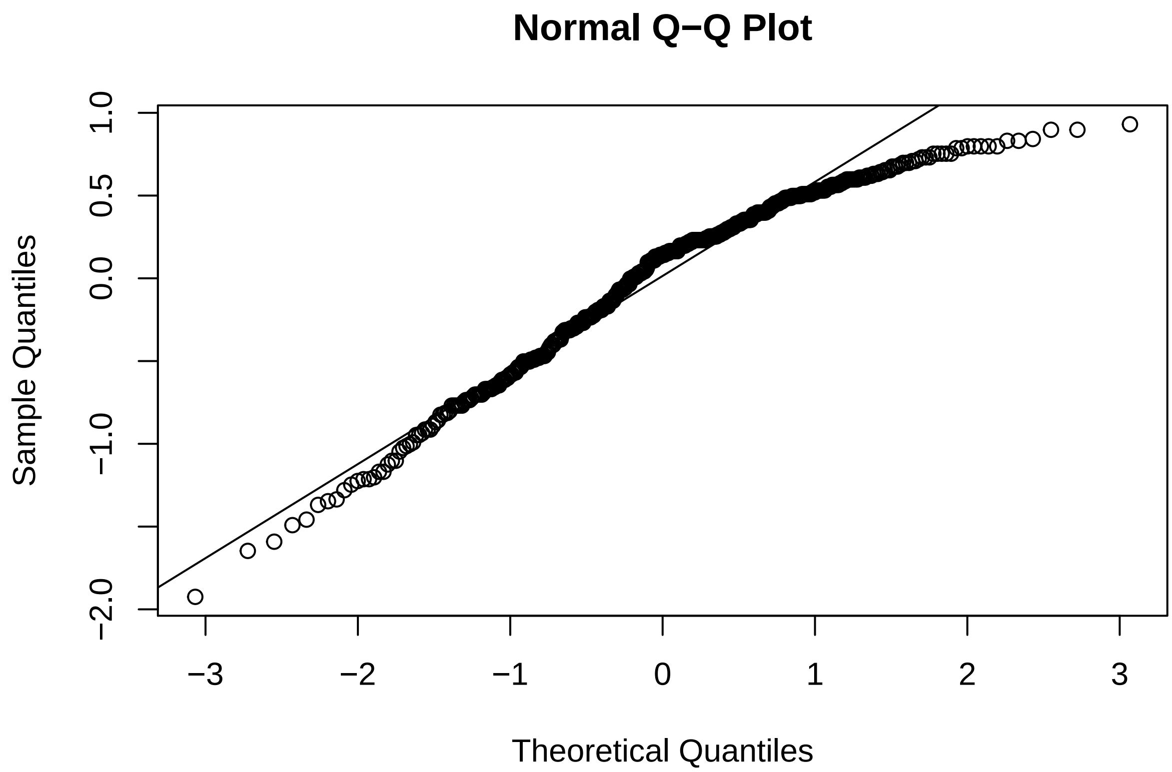
Nearly normal

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



Residuals are skewed to the left. I don't yet understand fully when I should reject based on skewedness.

```
qqnorm(m_bty$residuals)
qqline(m_bty$residuals)
```



I'm going to go with rejection. I don't think the residuals satisfy the requirement of being nearly normal.

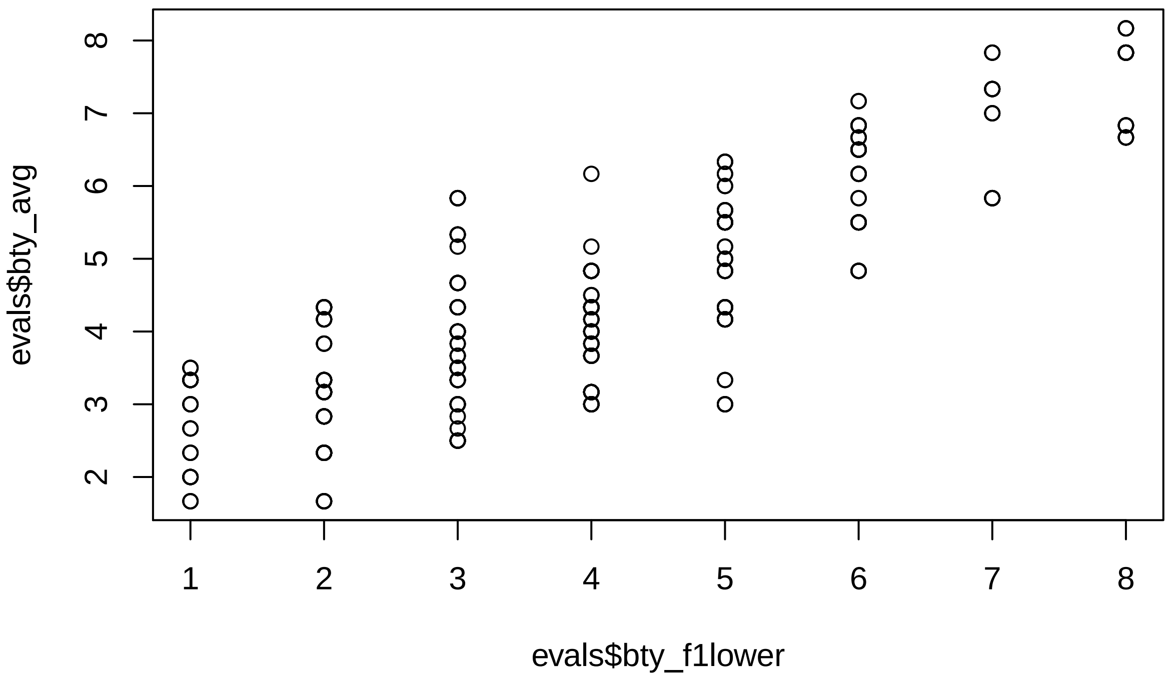
Constant variability

The first plot does not present constant variability (does it?). I'm saying no, because there are is greater variance in residuals below the correlation than above. Is that the right distinction?

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)
```

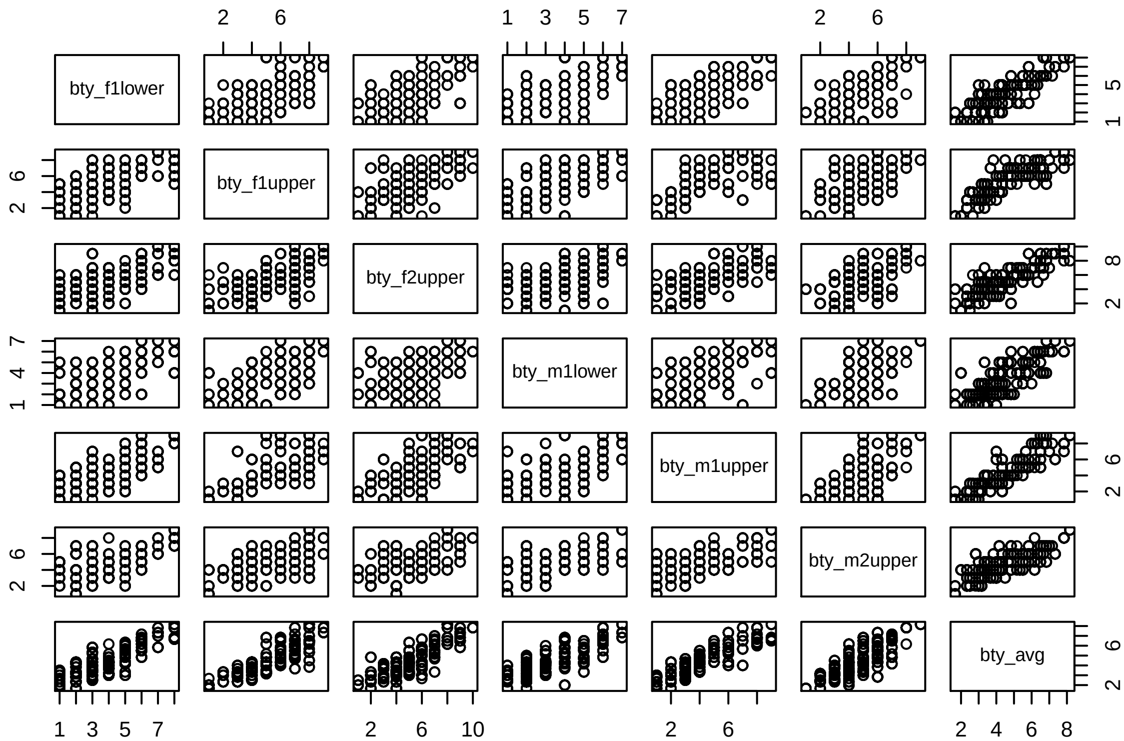



```
cor(evals$bty_avg, evals$bty_f1lower)
```

```
## [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)
```

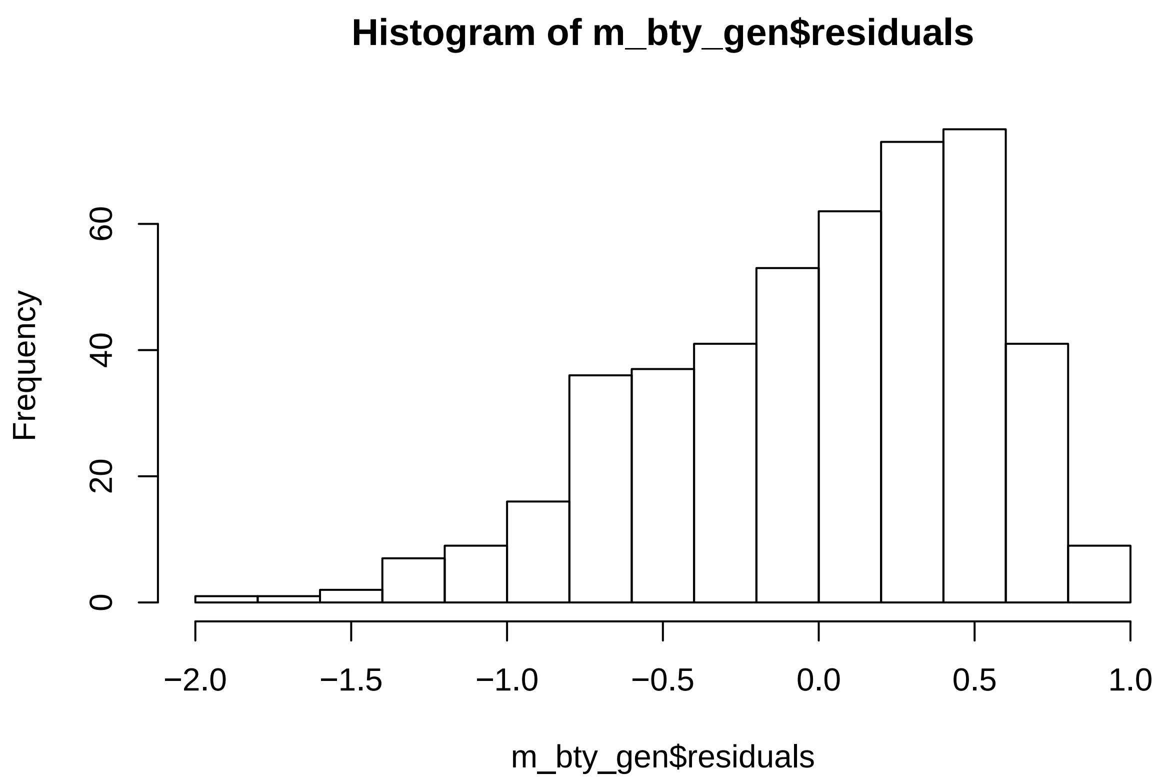
```
##
## Call:
## lm(formula = score ~ bty_avg + gender, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8305 -0.3625  0.1055  0.4213  0.9314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.74734    0.08466  44.266  < 2e-16 ***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

Check for outliers

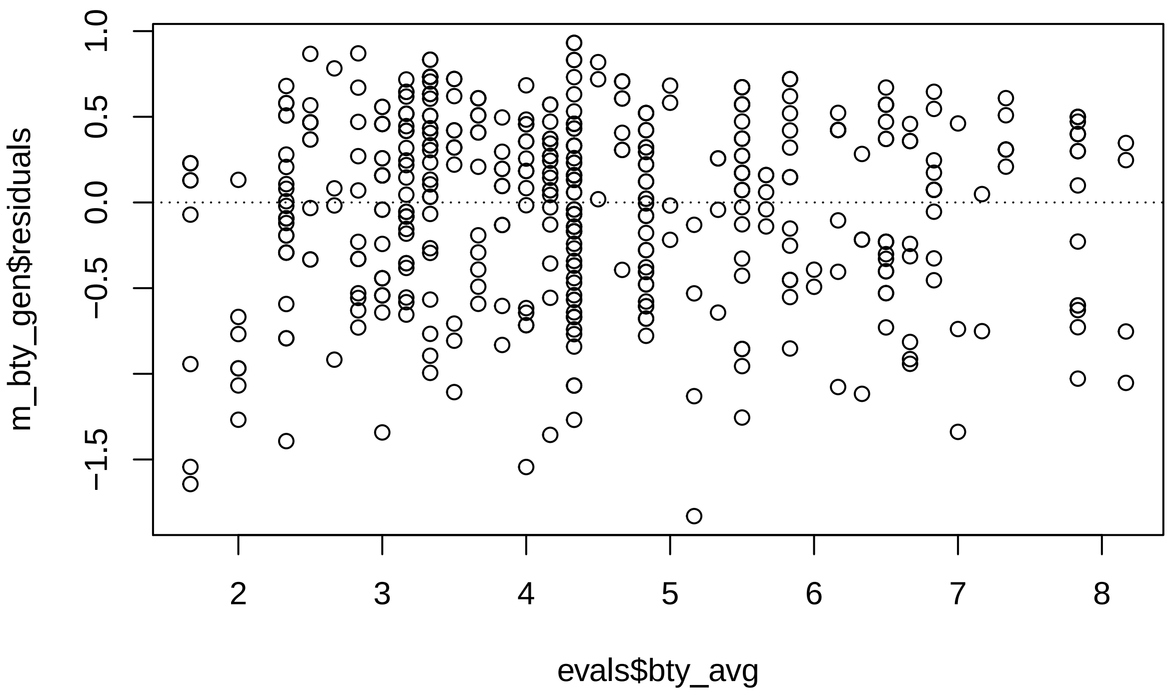
```
hist(m_bty_gen$residuals)
```



Residuals are normal enough and no outliers.

Linearity and variability

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



It looks a lot like it did without the gender coefficient. Maybe this is OK.

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`?

ANSWER

Yes, `bty_avg` is still a significant predictor. Adding `gender` to the model changed the parameter estimate for `bty_avg` from 0.06664 to 0.07416.

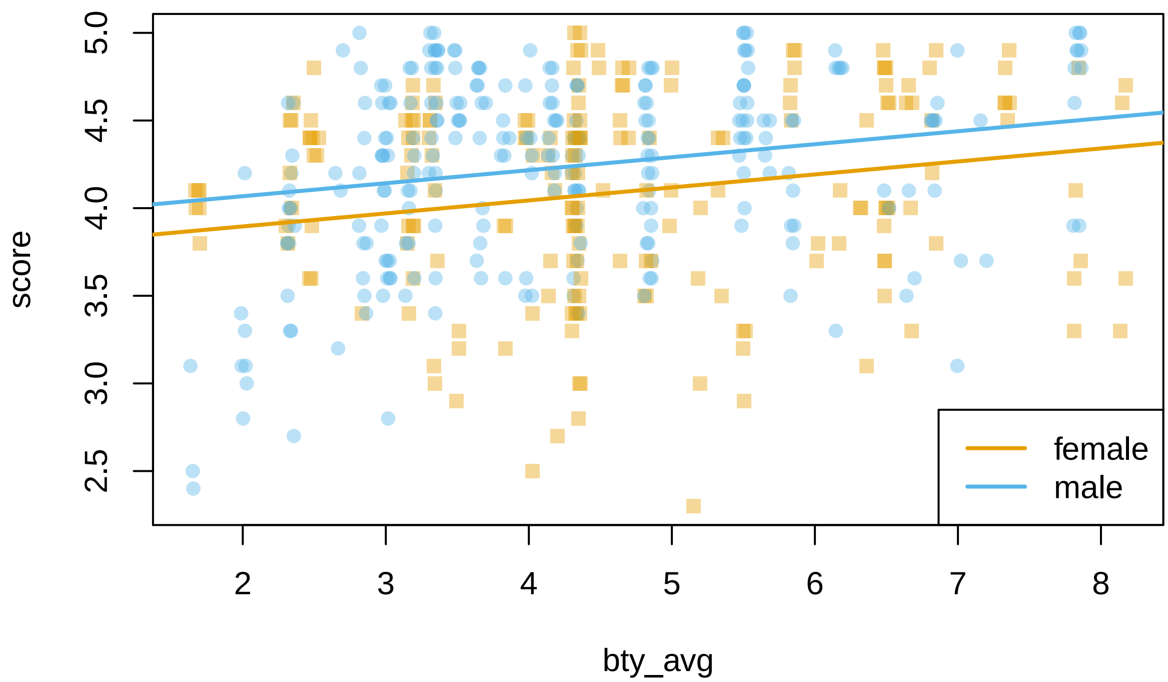
Note that the estimate for `gender` is now called `gendermale`. You'll see this name change whenever you introduce a categorical variable. The reason is that R recodes `gender` from having the values of `female` and `male` to being an indicator variable called `gendermale` that takes a value of 0 for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

$$\widehat{score} = 3.74734 + 0.07416 \times bty\_avg + 0.17239$$

Men tend to have the higher course evaluation for the same beauty rating.

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 creates dummy variables numbering (category levels - 1). The unnamed level is a reference variable.

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

I expect the `cls` variables not to have an association with the professor score. Therefore: `cls_perc_eval`, `cls_students`, `cls_level`, `cls_profs`, or `cls_credits`. To pick one, I guess `cls_credits`. But I always thought there was a problem with the design including some classes with multiple instructors, so `cls_profs` would be a close second for my guess.

Let's run the model…

```r
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.

ANSWER

Output: cls_profssingle has a P value of 0.77806.

So `cls_profs` is it.

13. Interpret the coefficient associated with the ethnicity variable.

ANSWER

All other predictors being equal, an additional point in beauty rating increases the score of someone who is not a minority 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?

```
m_full_less_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_less_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
```

There were some very slight shifts, but no major upset, revealing collinearity with the other predictor 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.

ANSWER

Full model adjusted R^2: 0.1617.

`cls_profsingle` has highest P at 0.77806.

Exclude `cls_prof`, adjusted R^2: 0.1634.

`cls_levelupper` has highest P at 0.292200.

Exclude `cls_level`, adjusted R^2: 0.1632.

That reduction reduced the predictive value of the model, so we keep `cls_level`. This is the same model from a couple questions back.

```r
m_full_less_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_less_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
```

$\widehat{score}$ = 4.0872523 - 0.1476746 * $rank_{tenuretrack}$ - 0.0973829 * $rank_{tenured}$ + 0.1274458 * $ethnicity_{notminority}$ + 0.2101231 * $gender_{male}$ - 0.2282894 * $language_{non-english}$ - 0.0089992 * age + 0.0052888 * cls_perc_eval + 0.0004687 * $cls\_students$ + 0.0606374 * cls_level_{upper} + 0.5061196 * $cls\_credits_{onecredit}$ + 0.0398629 * bty_avg- 0.1083227 * $pic\_outfit_{notformal}$ - 0.2190527 * pic_color_{color}

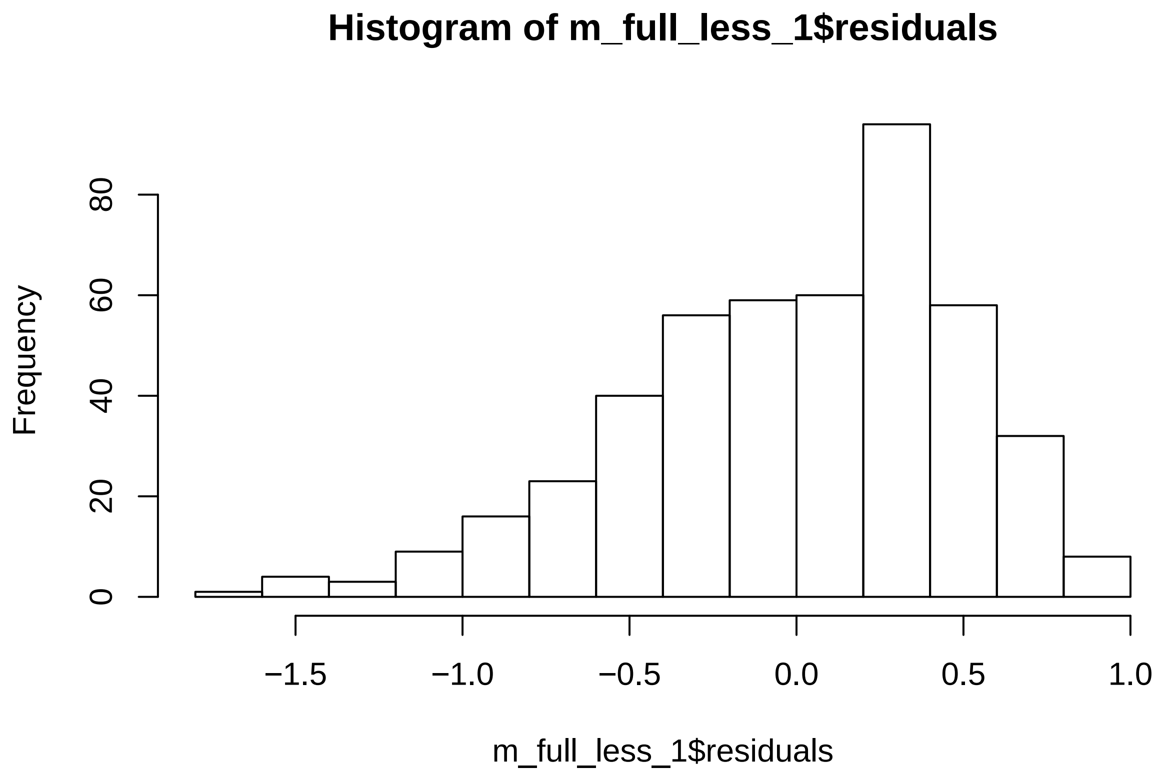
16. Verify that the conditions for this model are reasonable using diagnostic plots.

ANSWER

Seems reasonable. No outliers. Linear. Consistent variability.

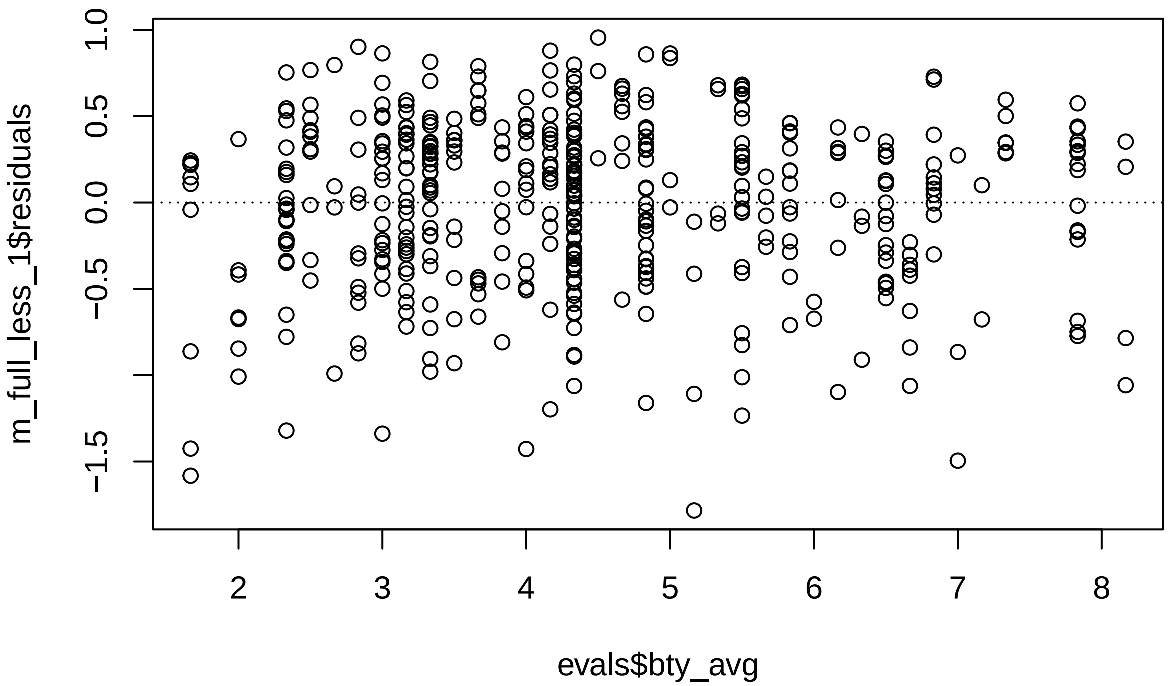
Check for outliers

```r
hist(m_full_less_1$residuals)
```



Linearity and variability

```
plot(m_full_less_1$residuals ~ evals$bty_avg)
abline(h = 0, lty = 3)
```



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

If it is a sample of professors instead of courses, then I think the scores should have been weighted by the number of courses each professor taught. Varibility in the number of courses taught would bias the results.

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

Instructor features.

Rank: Teaching (not tenured or tenure track).

Ethnicity: Not a minority.

Gender: Male.

Language: English.

Age: Younger the better.

Beauty: The more attractive the better.

Photo: Color and formal attire.

Class features.

More evaluations the better.

Higher class size the better.

Upper division.

One-credit class.

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

ANSWER

No. mostly due to differences in studentbody demographics. A school somewhere other than Texas may have a higher proportion of minority students and the coefficient for non-minorities may be higher.