Multiple linear regression

Andrew Bowen

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" by Hamermesh and Parker found that instructors who are viewed to be better looking receive higher instructional ratings.

Here, you will analyze the data from this study in order to learn what goes into a positive professor evaluation.

Getting Started

Load packages

In this lab, you will explore and visualize the data using the **tidyverse** suite of packages. The data can be found in the companion package for OpenIntro resources, **openintro**.

Let's load the packages.

```
library(tidyverse)
library(openintro)
library(GGally)
```

This is the first time we're using the `GGally` package. You will be using the `ggpairs` function from this package later in the lab.

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. The result is a data frame where each row contains a different course and columns represent variables about the courses and professors. It's called `evals`.

```
glimpse(evals)
```

```
## Rows: 463
## Columns: 23
## $ course_id     <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1~
## $ prof_id       <int> 1, 1, 1, 1, 2, 2, 2, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5,~
## $ score         <dbl> 4.7, 4.1, 3.9, 4.8, 4.6, 4.3, 2.8, 4.1, 3.4, 4.5, 3.8, 4~
## $ rank          <fct> tenure track, tenure track, tenure track, tenure track, ~
## $ ethnicity     <fct> minority, minority, minority, minority, not minority, no~
## $ gender        <fct> female, female, female, female, male, male, male, male, ~
## $ language      <fct> english, english, english, english, english, english, en~
## $ age           <int> 36, 36, 36, 36, 59, 59, 59, 51, 51, 40, 40, 40, 40, 40, ~
## $ cls_perc_eval <dbl> 55.81395, 68.80000, 60.80000, 62.60163, 85.00000, 87.500~
## $ cls_did_eval  <int> 24, 86, 76, 77, 17, 35, 39, 55, 111, 40, 24, 24, 17, 14,~
```

```
## $ cls_students  <int> 43, 125, 125, 123, 20, 40, 44, 55, 195, 46, 27, 25, 20, ~
## $ cls_level     <fct> upper, upper, upper, upper, upper, upper, upper, upper, ~
## $ cls_profs     <fct> single, single, single, single, multiple, multiple, mult~
## $ cls_credits   <fct> multi credit, multi credit, multi credit, multi credit, ~
## $ bty_f1lower   <int> 5, 5, 5, 5, 4, 4, 4, 5, 5, 2, 2, 2, 2, 2, 2, 2, 2, 7, 7,~
## $ bty_f1upper   <int> 7, 7, 7, 7, 4, 4, 4, 2, 2, 5, 5, 5, 5, 5, 5, 5, 5, 9, 9,~
## $ bty_f2upper   <int> 6, 6, 6, 6, 2, 2, 2, 5, 5, 4, 4, 4, 4, 4, 4, 4, 4, 9, 9,~
## $ bty_m1lower   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 7, 7,~
## $ bty_m1upper   <int> 4, 4, 4, 4, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 6, 6,~
## $ bty_m2upper   <int> 6, 6, 6, 6, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 6, 6,~
## $ bty_avg       <dbl> 5.000, 5.000, 5.000, 5.000, 3.000, 3.000, 3.000, 3.333, ~
## $ pic_outfit    <fct> not formal, not formal, not formal, not formal, not form~
## $ pic_color     <fct> color, color, color, color, color, color, color, color, ~
```

We have observations on 21 different variables, some categorical and some numerical. The meaning of each variable can be found by bringing up the help file:

```
?evals
```

Loading the df into my local env

```
evals <- evals
```

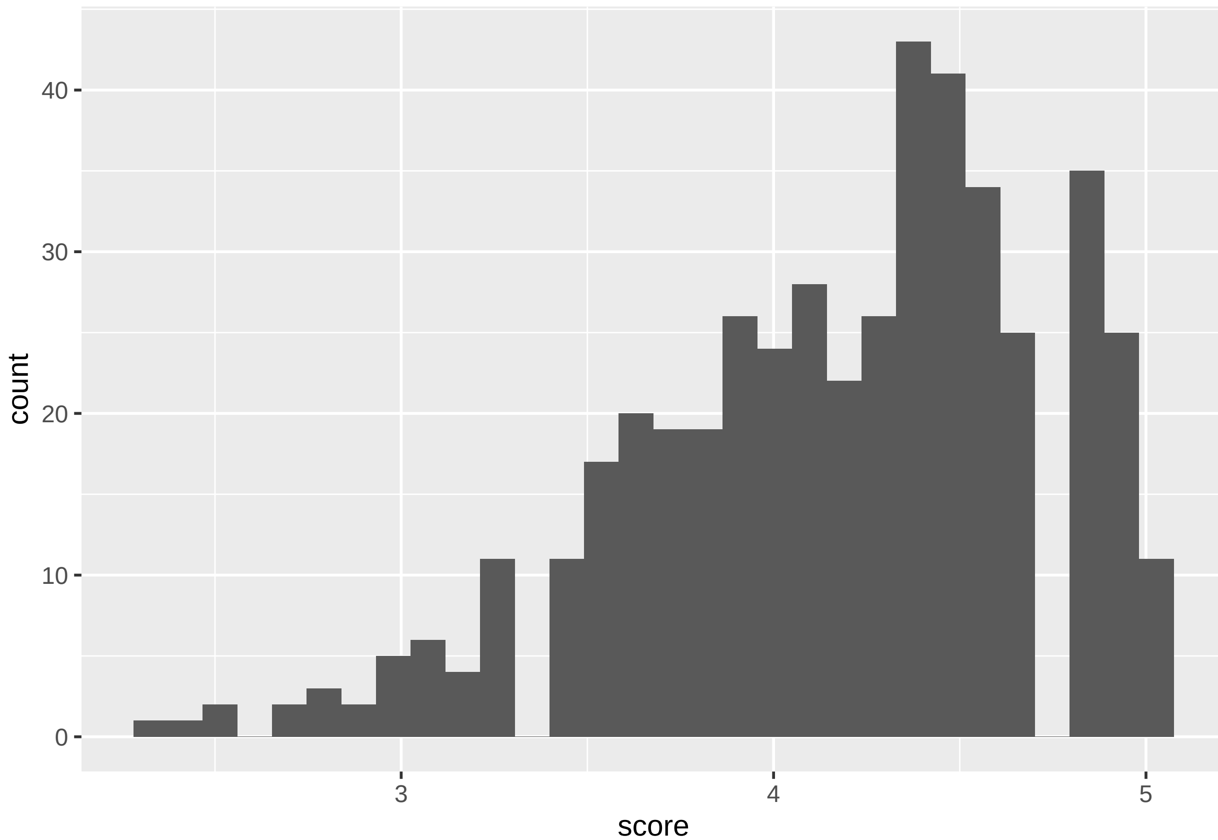
Exploring the data

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.

Assuming the experimenters did not change any conditions that would impact a student's evaluation, this would be an *observational study*. This is because the experimenters are only comparing changes in behavior, and not applying treatment to subjects in order to influence/change their behavior.

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? Plotting the `score` distribution via a `ggplot` histogram.

```
ggplot(evals, aes(x=score)) + geom_histogram()
```




```
# Adding in some summary stats from  our distribution
evals %>%
    summarise(mean_score = mean(score),
              median_score = median(score),
              stddev_score = sd(score),
              iqr_score = IQR(score))
```

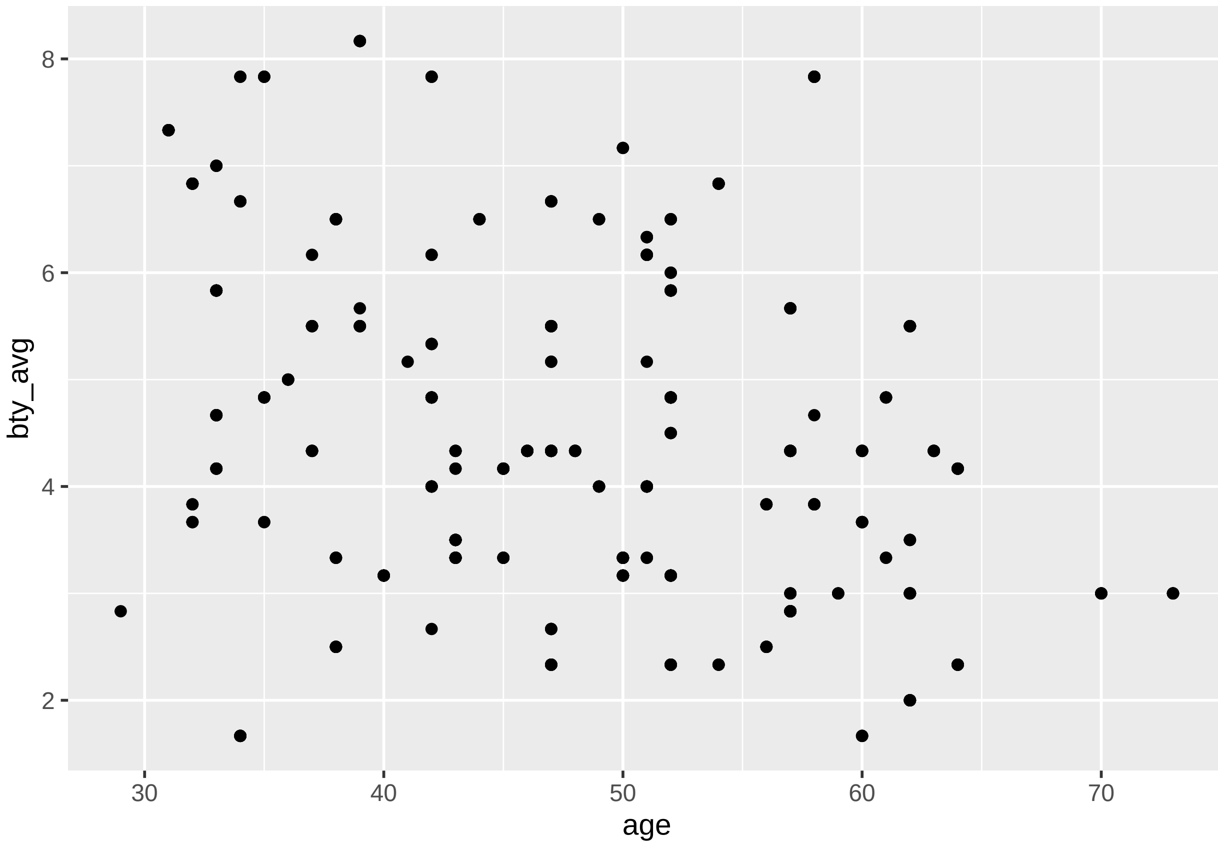
```
## # A tibble: 1 x 4
##   mean_score median_score stddev_score iqr_score
##        <dbl>        <dbl>        <dbl>     <dbl>
## 1       4.17          4.3        0.544       0.8
```

This distribution appears to be skewed-left. The median of the distribution is at 4.3, with an IQR of 0.8. This shows that students are generally "easy graders" when it comes to courses and are more likely to rate a course highly. I'd expect to see this based on personal experience. When I rate courses, I tend to give the instructor the benefit of the doubt more often than not.

3. Excluding `score`, select two other variables and describe their relationship with each other using an appropriate visualization.

Would like to take a look at a professor's age (`age`) and its relationship to the average beauty rating of the professor `bty_avg`. Luckily, this data is from the University of Texas at Austin, and not CUNY SPS as all the professors at CUNY are devilishly handsome, so the CUNY `bty_avg` distribution would be severely skewed to the point of being unusable. Going to plot their relationship via scatter plot:

```r
ggplot(evals, aes(x=age, y=bty_avg)) + geom_point()
```



Eye-test wise, there's a slight negative correlation in this data, but it doesn't appear very strong. Let's use the `cor` function in R to get the Pearson's correlation coefficient.

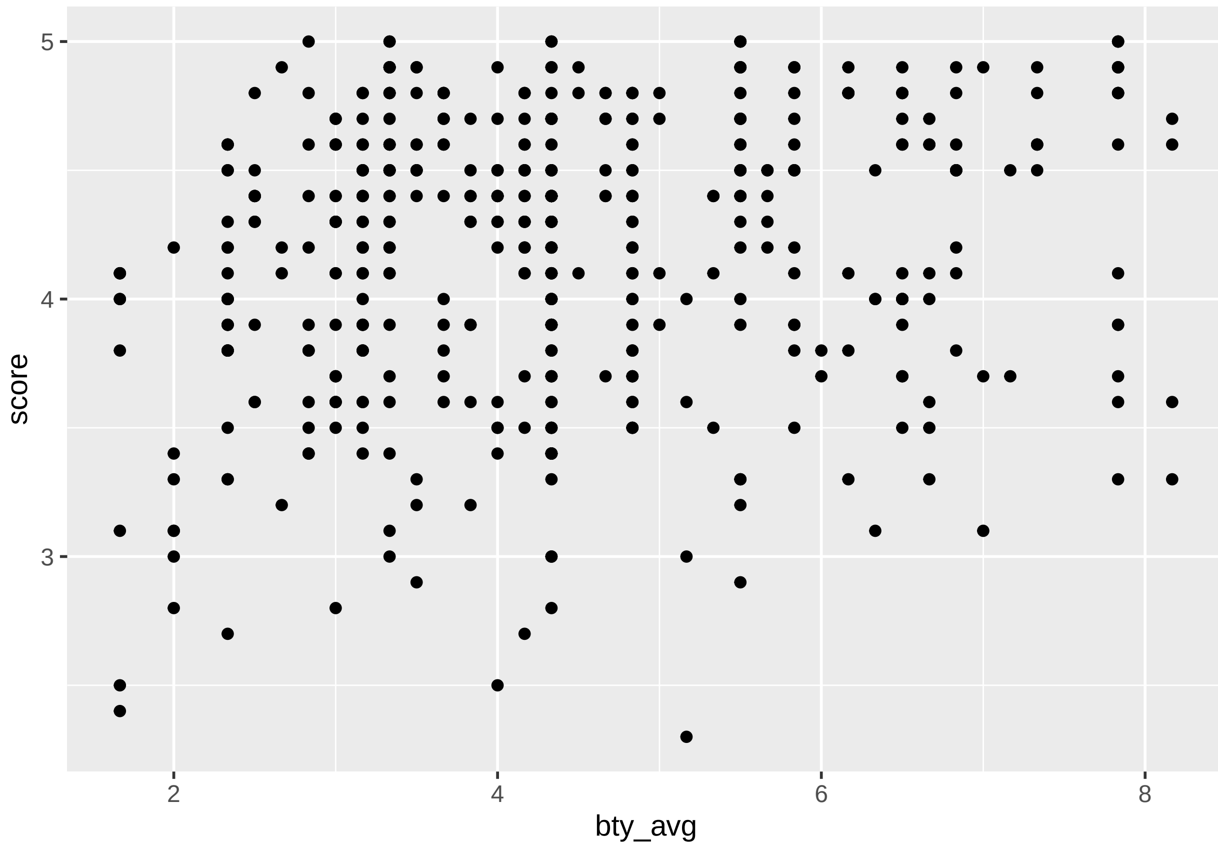
```r
cor(evals$age, evals$bty_avg)
```

```
## [1] -0.3046034
```

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
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_point()
```

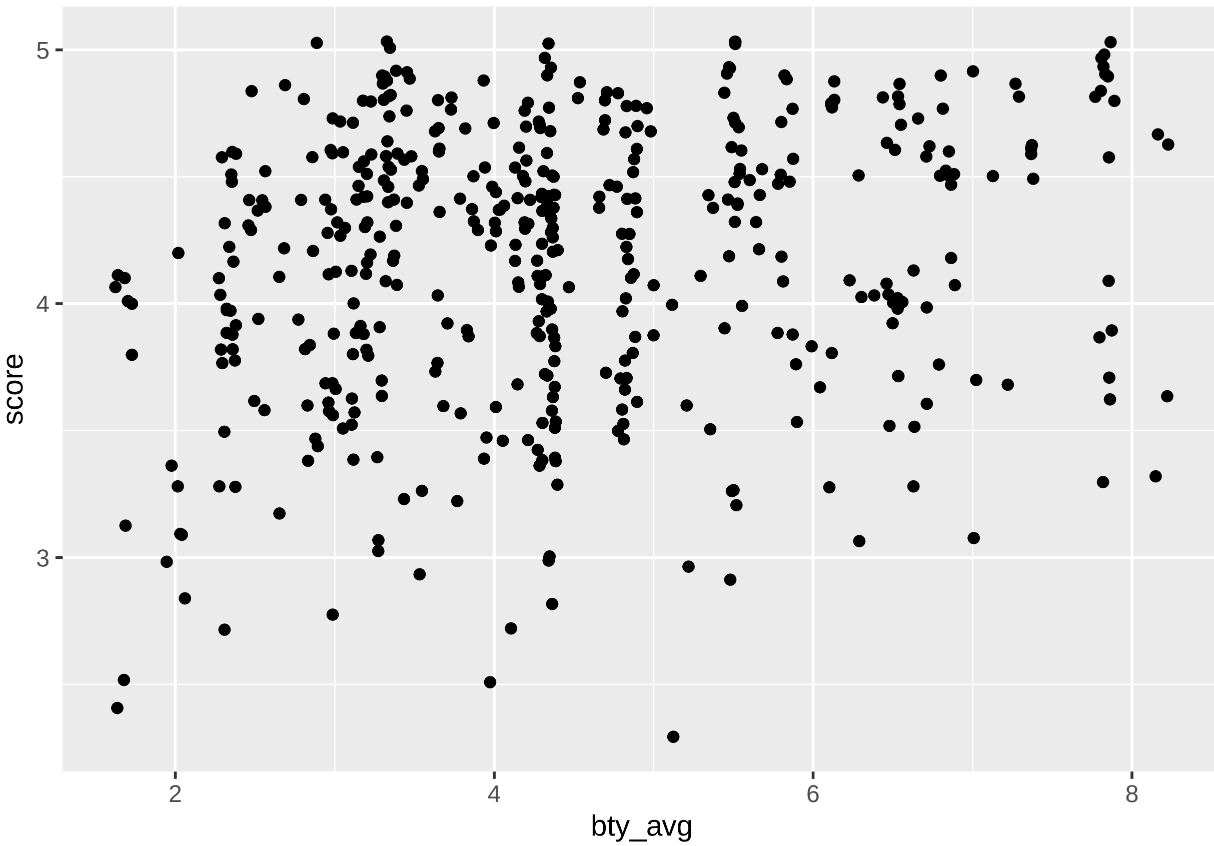


Before you draw conclusions about the trend, compare the number of observations in the data frame with the approximate number of points on the scatterplot. Is anything awry?

The scatter plot is “stacked” due to the granularity of our data. For instance, two professrs could have the same beauty rating and review score, which would not be distinguished on this plot

4. Replot the scatterplot, but this time use `geom_jitter` as your layer. What was misleading about the initial scatterplot?

```
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_jitter()
```



The initial scatter plot had its points aligned to the same granularity of the data (in this case one decimal point). Identical points were overlaid and hard to distringuish

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. 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(score ~ bty_avg, data=evals)

summary(m_bty)
```

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

```
## Multiple R-squared:  0.03502,    Adjusted R-squared:  0.03293
## F-statistic: 16.73 on 1 and 461 DF,  p-value: 5.083e-05
```

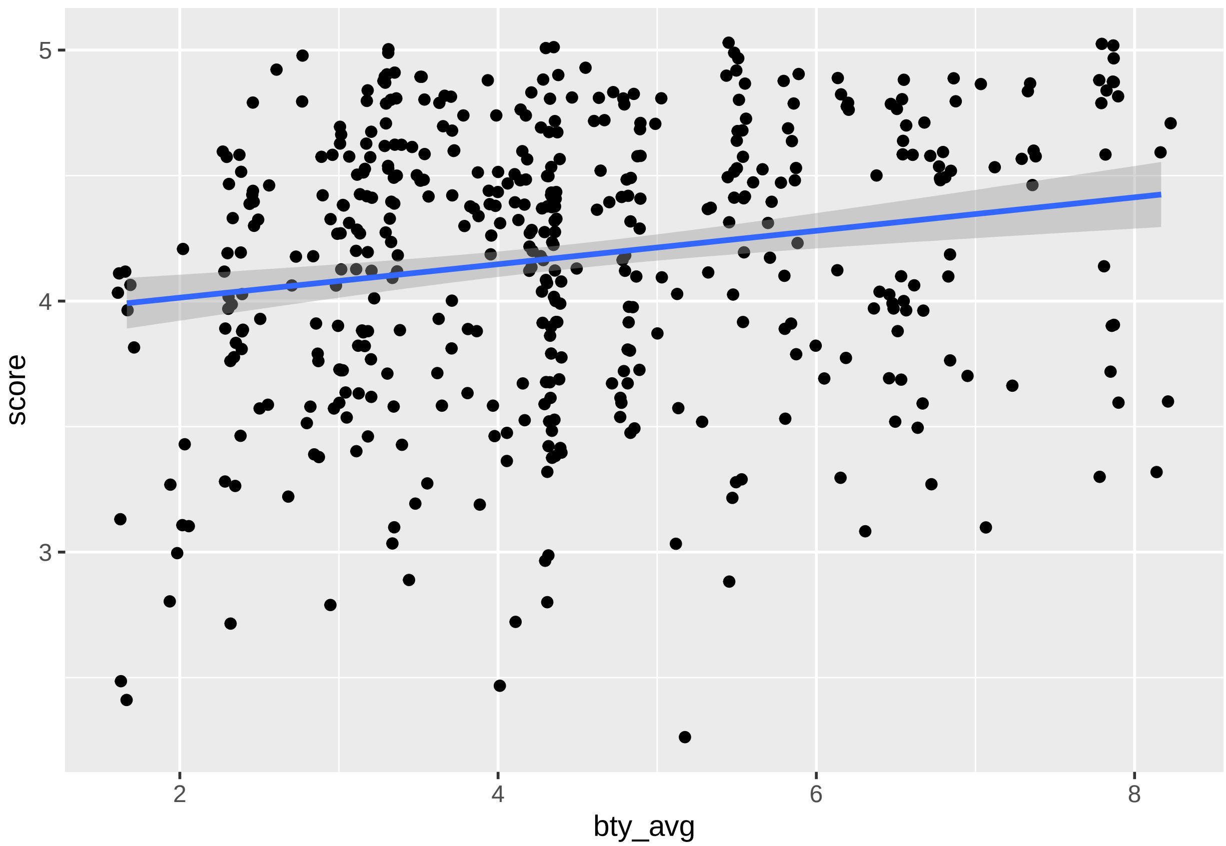
The formula for this linear model would be

$score = 0.0664 * (bty\_avg) + 3.88034$

With a p-value less than a 5% significance level, beauty average is a statistically significant predictor of a professor's review score.

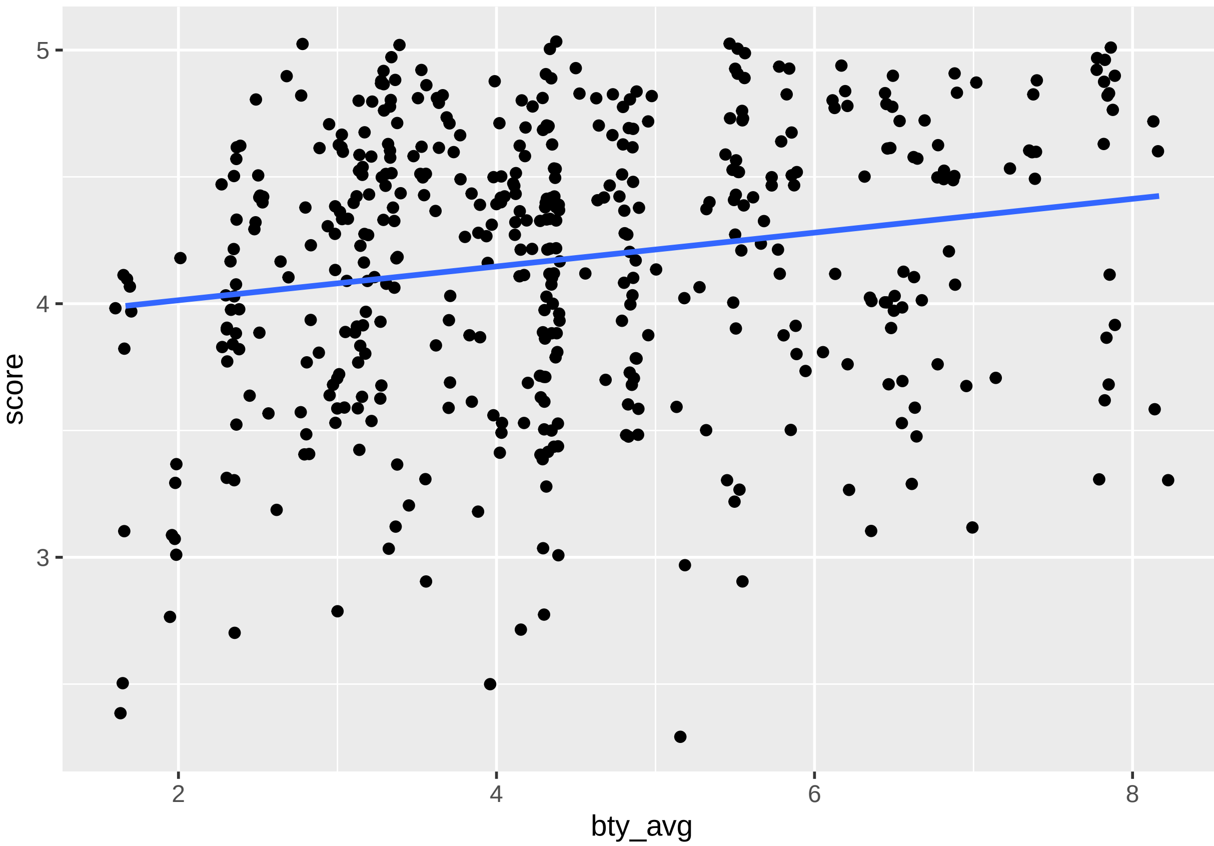
Add the line of the bet fit model to your plot using the following:

```
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_jitter() +
  geom_smooth(method = "lm")
```



The blue line is the model. The shaded gray area around the line tells you about the variability you might expect in your predictions. To turn that off, use `se = FALSE`.

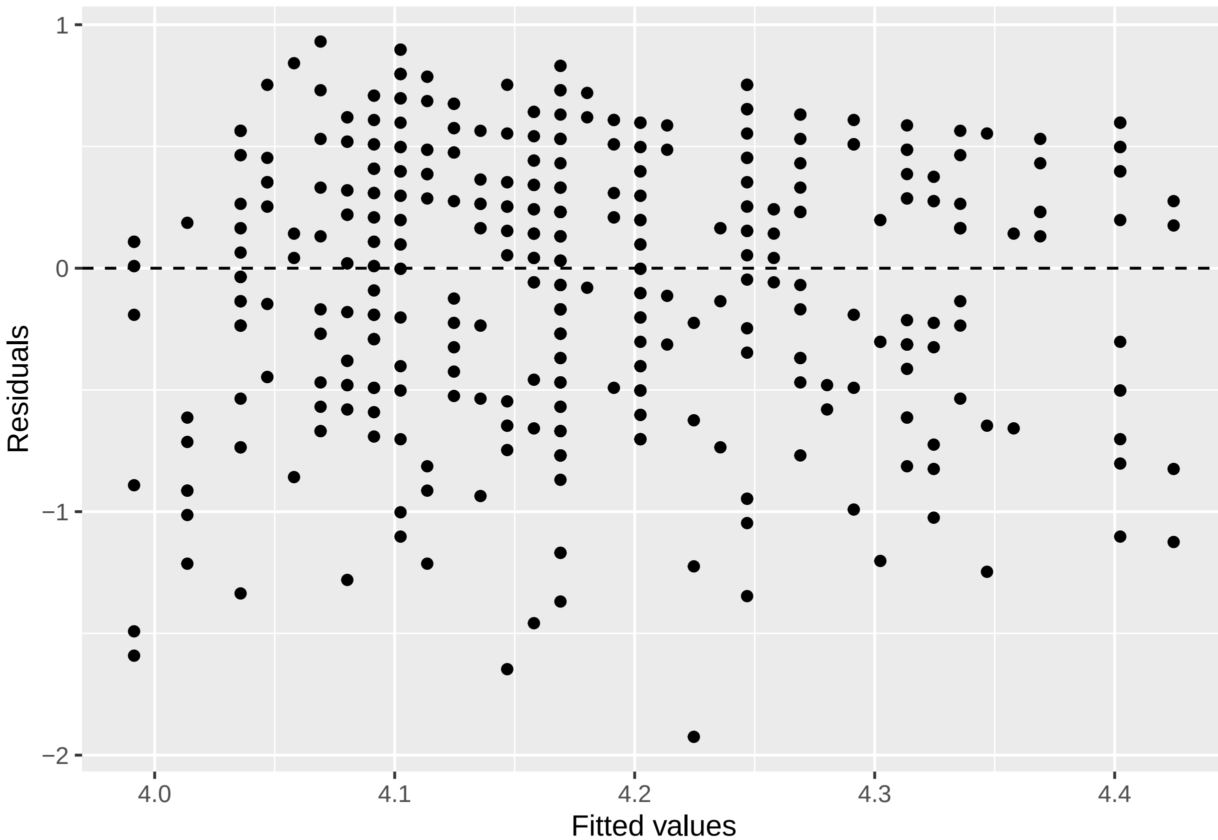
```
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_jitter() +
  geom_smooth(method = "lm", se = FALSE)
```



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).

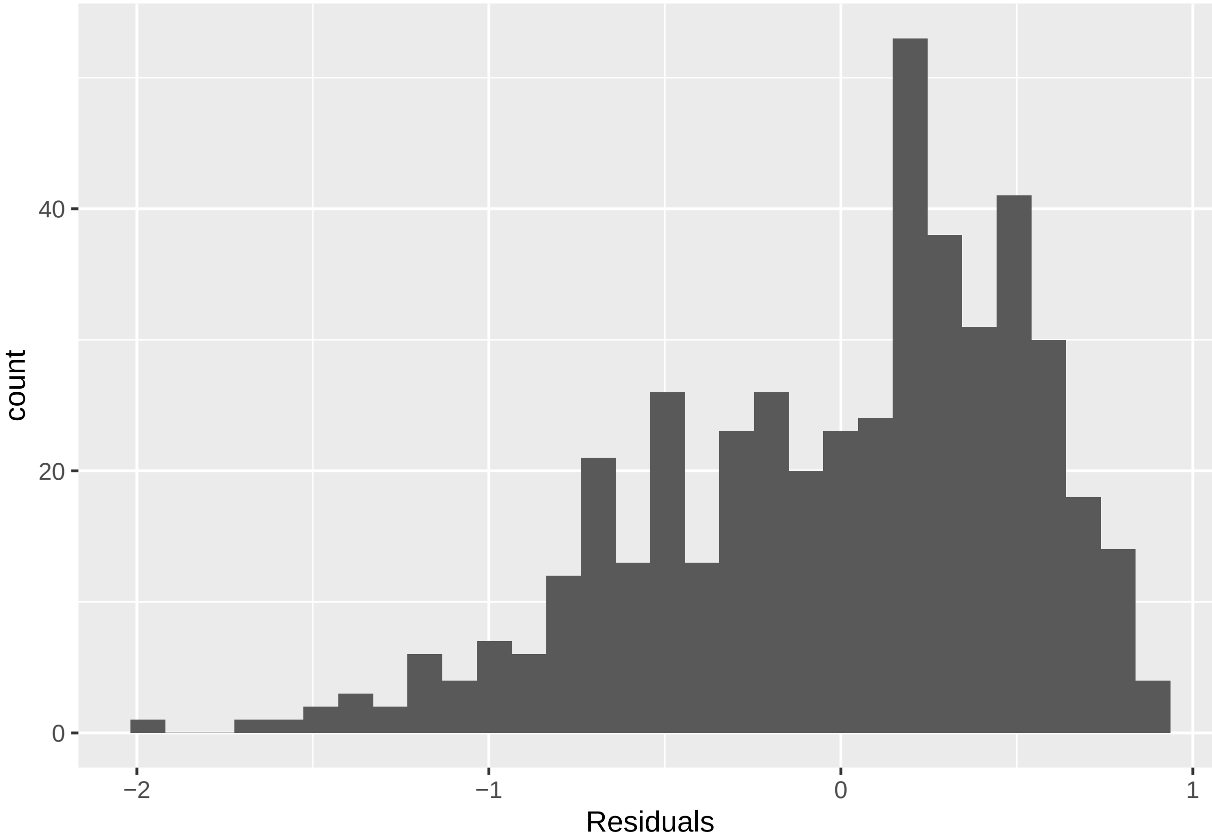
Linearity Will plot the residuals vs fitted values to see if there is a relationship between the residuals and our fitted values. From the below plot this condition looks to be met.

```
#  Using the syntax from our simple regression lab
ggplot(data = m_bty, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```



Normality Checking that our residuals are nearly normally distributed. These look to be slightly skewed left

```
ggplot(data = m_bty, aes(x = .resid)) +
  geom_histogram() +
  xlab("Residuals")
```

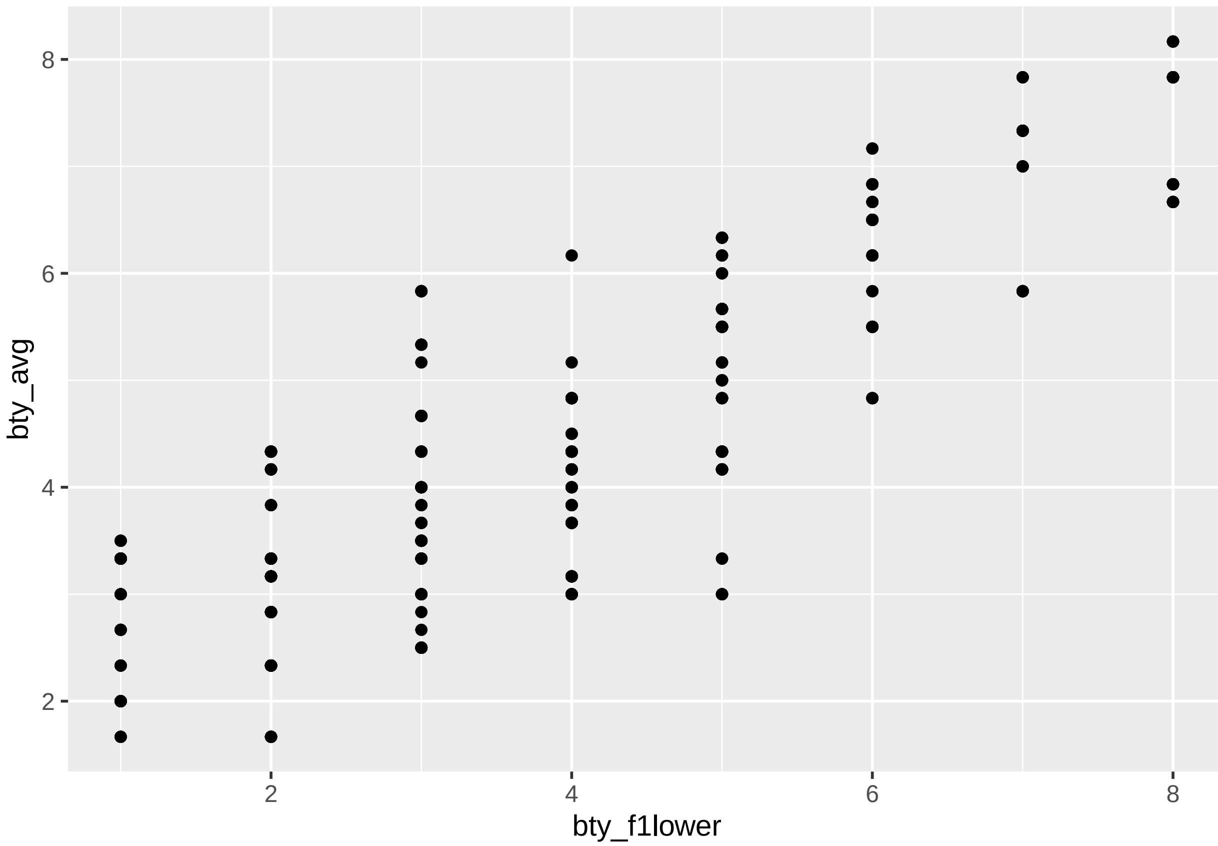


Additionally, the observations should be independent as, in theory, one student's review shouldn't be impacted by another student's review. In practice one could challenge this assumption as student's likely discuss professors with each other and opinions could transfer between students in a class.

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.

```
ggplot(data = evals, aes(x = bty_f1lower, y = bty_avg)) +
  geom_point()
```

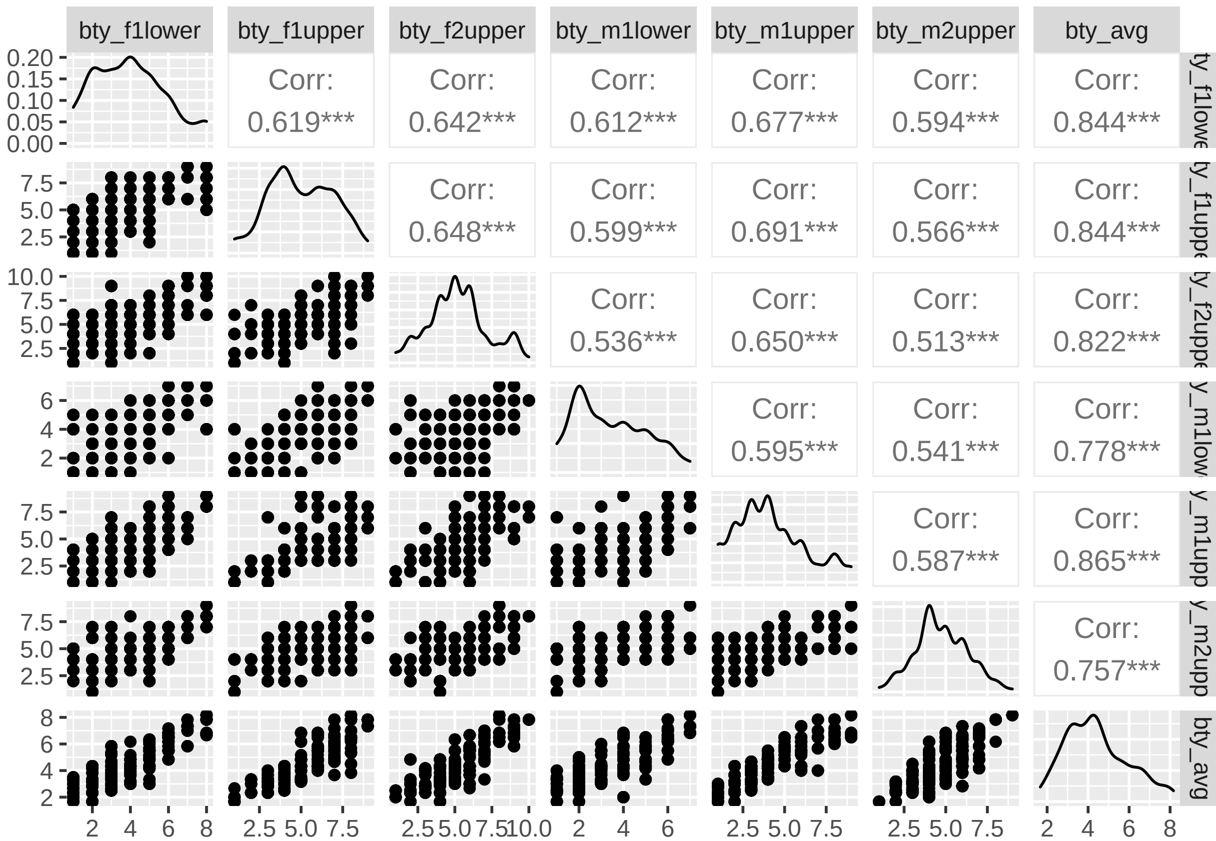



```r
evals %>%
  summarise(cor(bty_avg, bty_f1lower))
```

```
## # A tibble: 1 x 1
##   `cor(bty_avg, bty_f1lower)`
##                         <dbl>
## 1                       0.844
```

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

```r
evals %>%
  select(contains("bty")) %>%
  ggpairs()
```



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 you've accounted for the professor's gender, you can add the gender term into the model.

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

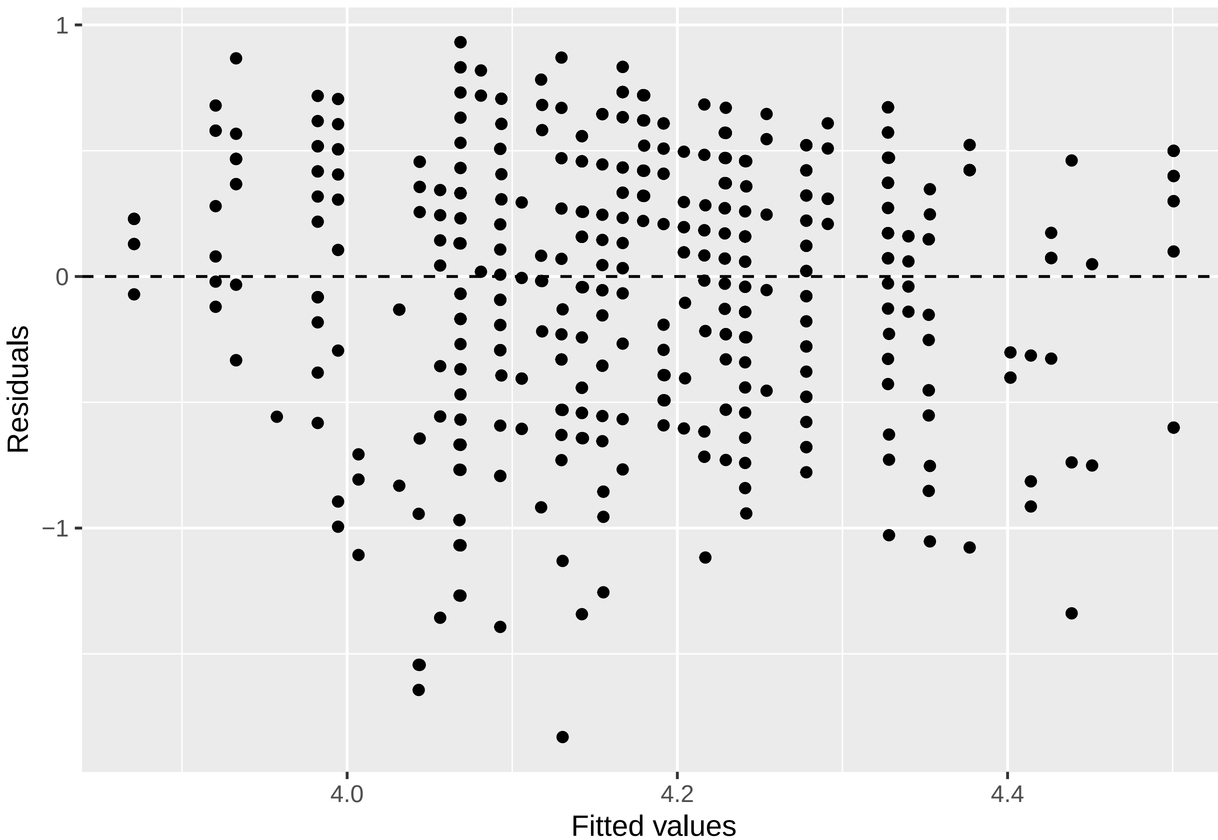
```
##
## Call:
## lm(formula = score ~ bty_avg + gender, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8305 -0.3625  0.1055  0.4213  0.9314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.74734    0.08466  44.266  < 2e-16 ***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

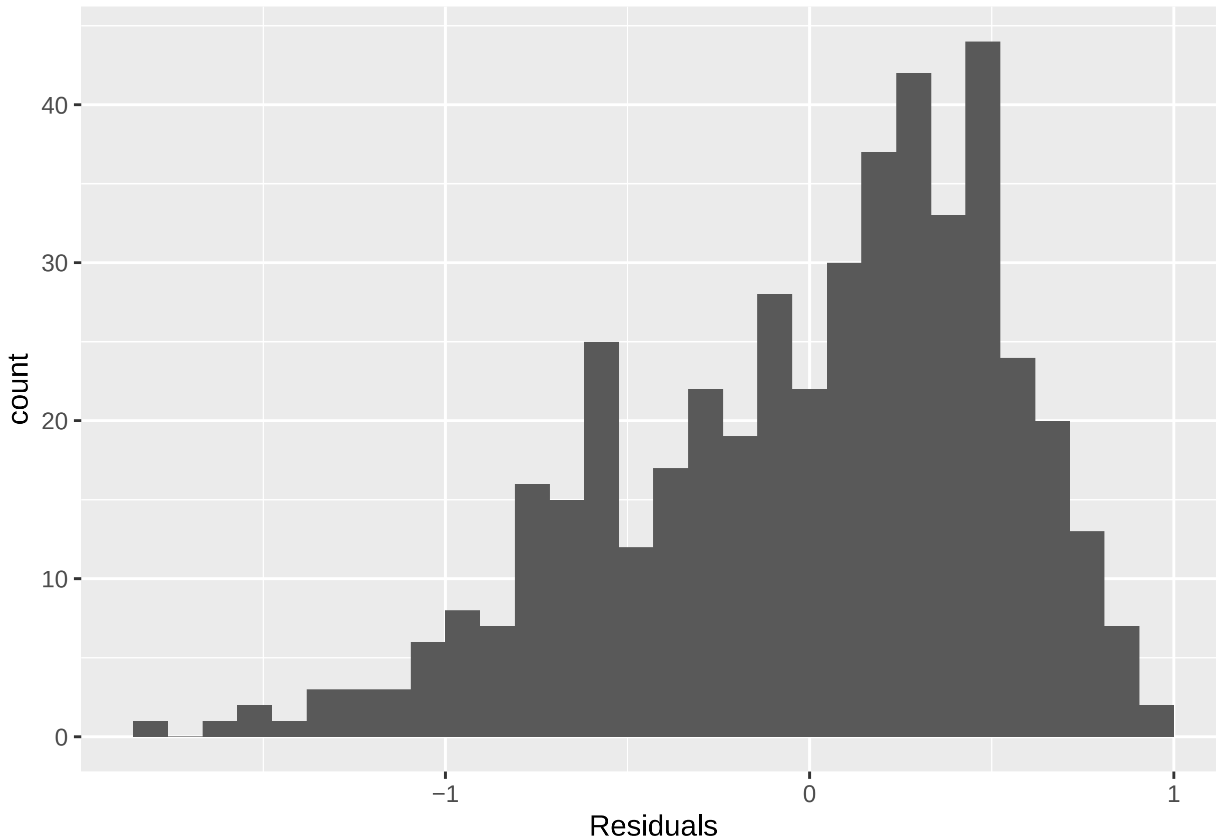
7. P-values and parameter estimates should only be trusted if the conditions for the regression are reasonable. Verify that the conditions for this model are reasonable using diagnostic plots.

Independence is met as above, because student's ratings (beauty and score) should be independent of one another. Assessing Linearity and Normality below

```
#  Using the syntax from our simple regression lab
ggplot(data = m_bty_gen, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```




```
ggplot(data = m_bty_gen, aes(x = .resid)) +
  geom_histogram() +
  xlab("Residuals")
```



Normality
This residuals plot is slightly skewed left but is close enoguh to normal that we could meet this 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`?

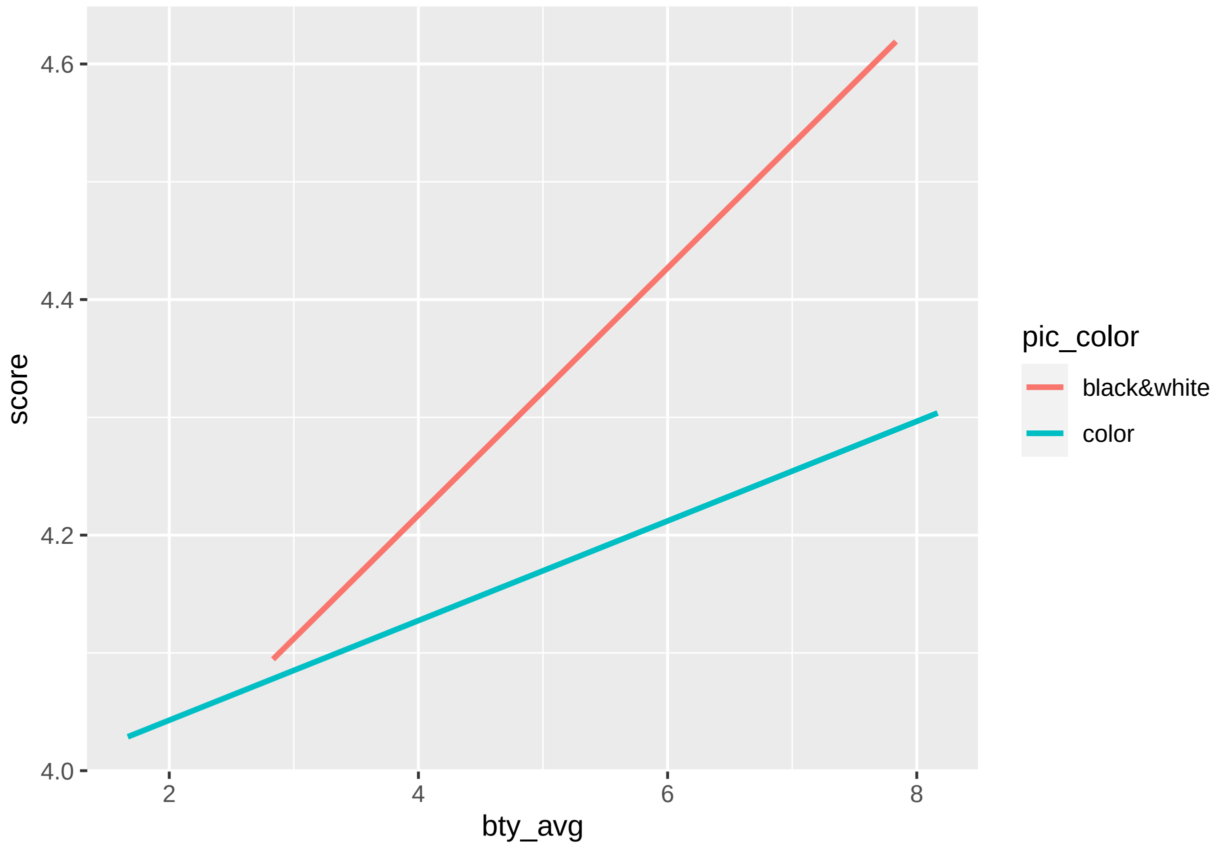
Beauty rating is still a significant predictor of `score`, even with the addition of the `gender` variable. The addition of the `gender` variable has lowered the p-value of our linear model, meaning that our model's statistical significance has improved

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 `male` and `female` to being an indicator variable called `gendermale` that takes a value of 0 for female professors and a value of 1 for male professors. (Such variables are often referred to as "dummy" variables.)

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

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

```
ggplot(data = evals, aes(x = bty_avg, y = score, color = pic_color)) +
 geom_smooth(method = "lm", formula = y ~ x, se = FALSE)
```



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

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

```
##
## Call:
## lm(formula = score ~ bty_avg + gender + pic_color, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7807 -0.3587  0.1142  0.4113  0.8839
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.95031    0.11173  35.354  < 2e-16 ***
## bty_avg          0.06171    0.01676   3.683 0.000258 ***
## gendermale       0.18750    0.05016   3.738 0.000209 ***
## pic_colorcolor -0.18851    0.06837  -2.757 0.006064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5249 on 459 degrees of freedom
## Multiple R-squared:  0.07445,    Adjusted R-squared:  0.0684
## F-statistic: 12.31 on 3 and 459 DF,  p-value: 9.321e-08
```

$$\hat{score} = 3.95031 + 0.06171 * (bty\_avg) + 0.18750 * (gender) - 0.18851 * (pic\_color)$$

Looking at the plot of our models including picture color, professors with black & white photos tend to have a higher score than those with color pictures and 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`.

```
evals$rank <- relevel(evals$rank, ref = "teaching")
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
```

Using `relevel`, R will assign the other categorical variables in that field to ascending values (1,2,3,... until $n - 1$ categories).

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, gender, ethnicity, 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 guess the `cls_credit` variable would have a weak relationship with a professor's score. In general, students likely would not be thinking about how many credits a class counts for when reviewing professors.

Let's run the model…

```
m_full <- lm(score ~ rank + gender + ethnicity + 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 + gender + ethnicity + 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
## gendermale             0.2109481  0.0518230   4.071 5.54e-05 ***
## ethnicitynot minority  0.1234929  0.0786273   1.571  0.11698
## 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.

```
summary(m_full)
```

```
##
## Call:
## lm(formula = score ~ rank + gender + ethnicity + 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
## gendermale            0.2109481  0.0518230   4.071 5.54e-05 ***
## ethnicitynot minority  0.1234929  0.0786273   1.571  0.11698
## 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
```

My prediction was off! Turns out the highest p-value (lowest significance) comes from the `cls_profsingle` variable at $p \approx 0.77$. My prediction of `cls_credit` was off as that looks to be a significant predictor or score.

13. Interpret the coefficient associated with the ethnicity variable.

For the ethnicity variable, the ethnicity of not minority of 0.1234929 shows that a professors score is positively associated with the ethnicity variable being not a minority.

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

```
##
## Call:
## lm(formula = score ~ rank + gender + ethnicity + 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
## gendermale             0.2101231  0.0516873   4.065 5.66e-05 ***
## ethnicitynot minority  0.1274458  0.0772887   1.649 0.099856 .
## 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
```

Dropping the `cls_profs` variable looks to have improved the p-values for most of the variables. This would imply that the `cls_profs` variable is colinear with our other predictors and should be dropped from 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 `regsubsets` function from the `leaps` package to select the best linear model. This does a regressive search for the optimal model given the predictor (independent) variables.

```
library(leaps)
```

```
subset <- regsubsets(score~., y=score,
                     data=evals, method="backward")

subset_summary <- summary(subset)

predictors <- as.data.frame(subset_summary$outmat)
```

```
m_opt <- lm(score ~ prof_id + ethnicity + gender + language + cls_perc_eval + cls_credits + bty_avg + pi

summary(m_opt)
```

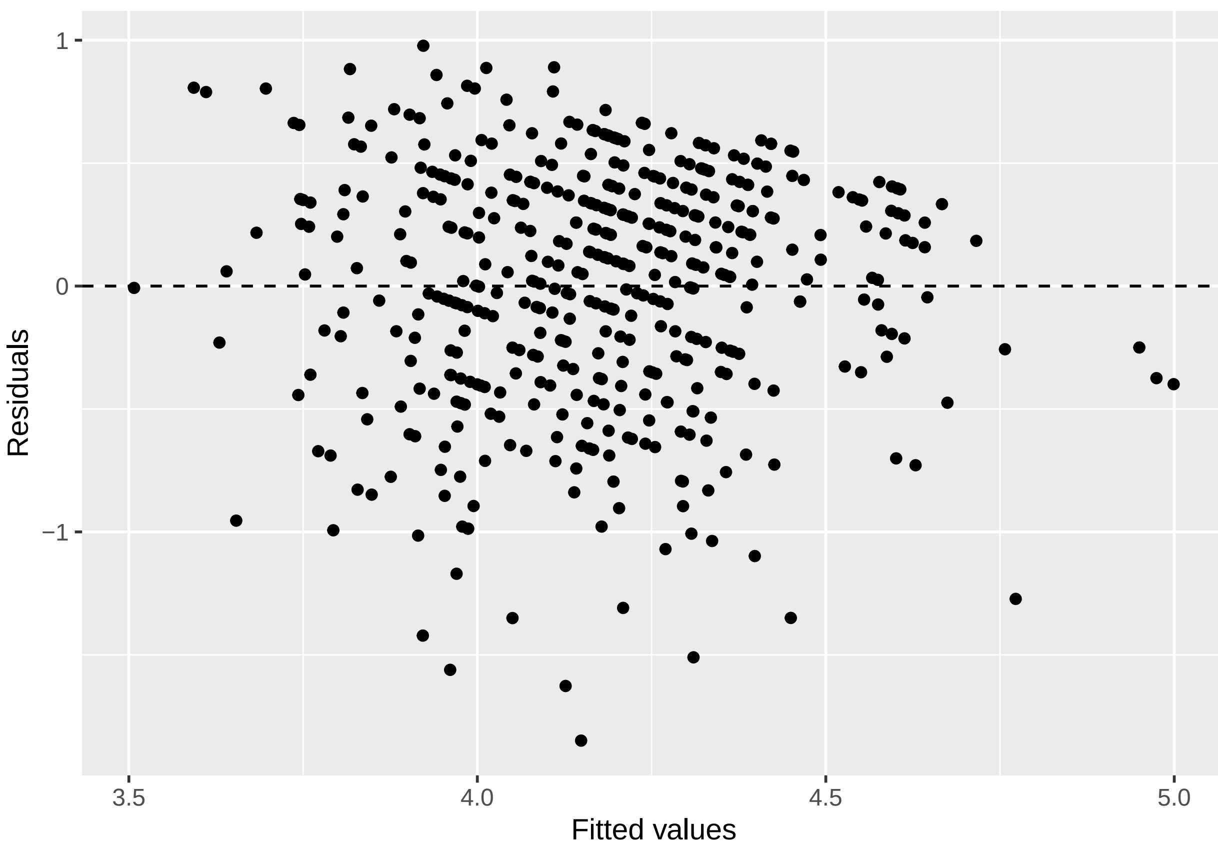
```
##
## Call:
## lm(formula = score ~ prof_id + ethnicity + gender + language +
##     cls_perc_eval + cls_credits + bty_avg + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.84891 -0.34216  0.08869  0.37039  0.97749
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.6626180  0.1900213  19.275  < 2e-16 ***
```

```
## prof_id                -0.0029569  0.0009304  -3.178 0.001584 **
## ethnicitynot minority   0.1520122  0.0749352   2.029 0.043084 *
## gendermale              0.1924636  0.0483446   3.981 7.99e-05 ***
## languagenon-english    -0.2009558  0.1031185  -1.949 0.051937 .
## cls_perc_eval           0.0041858  0.0014400   2.907 0.003831 **
## cls_creditsone credit   0.5922699  0.1062131   5.576 4.23e-08 ***
## bty_avg                 0.0628542  0.0160660   3.912 0.000105 ***
## pic_colorcolor         -0.2506012  0.0711737  -3.521 0.000473 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4966 on 454 degrees of freedom
## Multiple R-squared:  0.1806, Adjusted R-squared:  0.1662
## F-statistic: 12.51 on 8 and 454 DF,  p-value: 2.883e-16
```

16. Verify that the conditions for this model are reasonable using diagnostic plots. The independence condition is met, but let's check on linarity and normality

Linearity

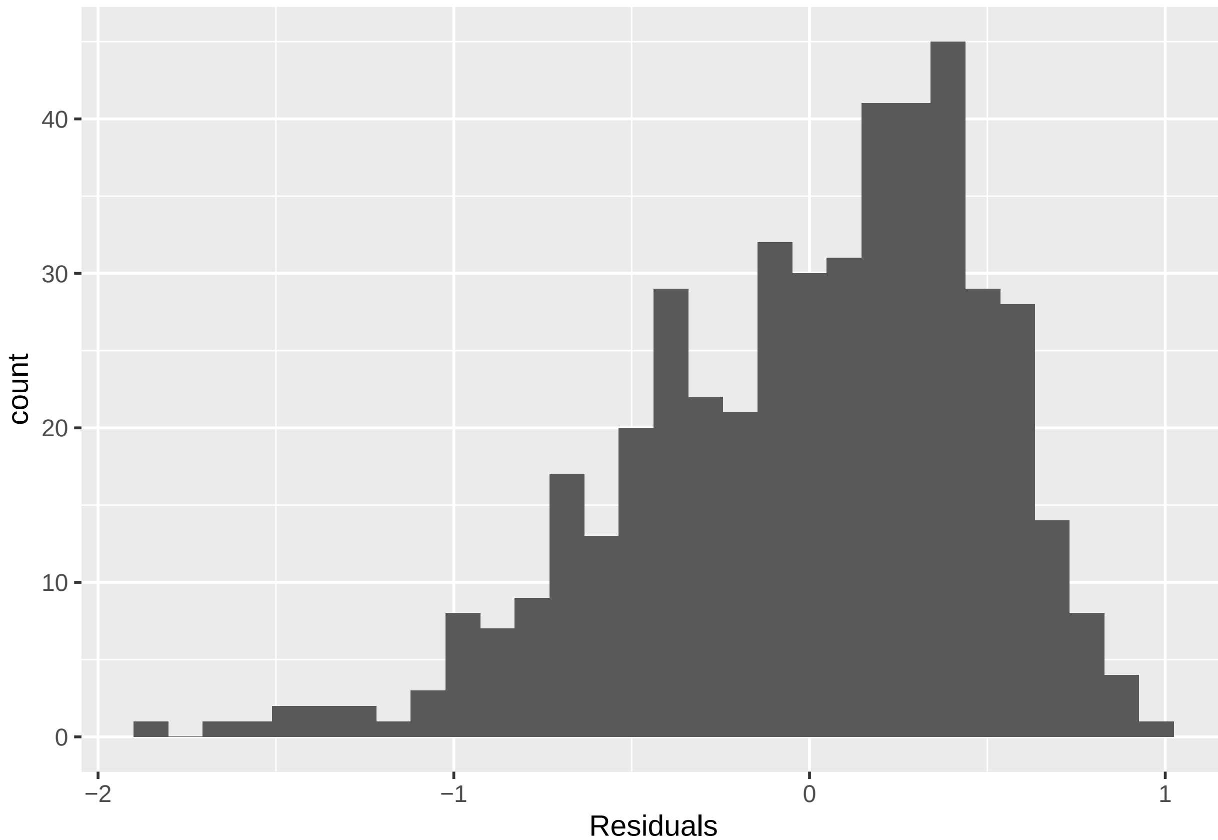
```
#  Using the syntax from our simple regression lab
ggplot(data = m_opt, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```



There doesn't seem to be a strong relationship between the residuals and fitted values.

Normality

```r
ggplot(data = m_opt, aes(x = .resid)) +
  geom_histogram() +
  xlab("Residuals")
```



This distribution of residuals looks to be slightly skewed but the normality condition appears to be met. This is in line with residual distributions plotted above.

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?

Given that this data is at the course level, rather than the individual professor level, that could cause some issues in the groupings. For instance, if a new professor teaches a course in a given year, that could change how that course is scored. This information could also impact the credits level `cls_credits` which is a factor in regression.

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.

```r
summary(m_opt)
```

```
##
## Call:
## lm(formula = score ~ prof_id + ethnicity + gender + language +
##     cls_perc_eval + cls_credits + bty_avg + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.84891 -0.34216  0.08869  0.37039  0.97749
```

```
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           3.6626180  0.1900213  19.275  < 2e-16 ***
## prof_id              -0.0029569  0.0009304  -3.178 0.001584 ** 
## ethnicitynot minority  0.1520122  0.0749352   2.029 0.043084 *  
## gendermale            0.1924636  0.0483446   3.981 7.99e-05 ***
## languagenon-english   -0.2009558  0.1031185  -1.949 0.051937 .  
## cls_perc_eval         0.0041858  0.0014400   2.907 0.003831 ** 
## cls_creditsone credit  0.5922699  0.1062131   5.576 4.23e-08 ***
## bty_avg               0.0628542  0.0160660   3.912 0.000105 ***
## pic_colorcolor       -0.2506012  0.0711737  -3.521 0.000473 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4966 on 454 degrees of freedom
## Multiple R-squared:  0.1806, Adjusted R-squared:  0.1662
## F-statistic: 12.51 on 8 and 454 DF,  p-value: 2.883e-16
```

Given our optimized regression model `m_opt`, you would expect a professor with a high score at UT Austin to be a non-minority, non-english speaking, "beautiful" man with a color picture teaching one credit with a low evaluation percentage.

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

I would be hesitant to generalize these findings to all professors. Firstly, students at other universities might have other inputs that change how they rate a professor. In addition, there could be other independent variables not captured in our dataset that could influence how students are likely to rate a professor. For instance, quality of instruction, humor, and relationship-building with students are variables that, while very difficult to quantify, would likely factor in highly into a student's review of an instructor

That being said, it is documented that people who are rated as highly physically attractive are more likely to be rated well in job interviews. While not the exact same as a course, this is a form of subjective rating that can be skewed by human biases.
