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" 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
```

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.

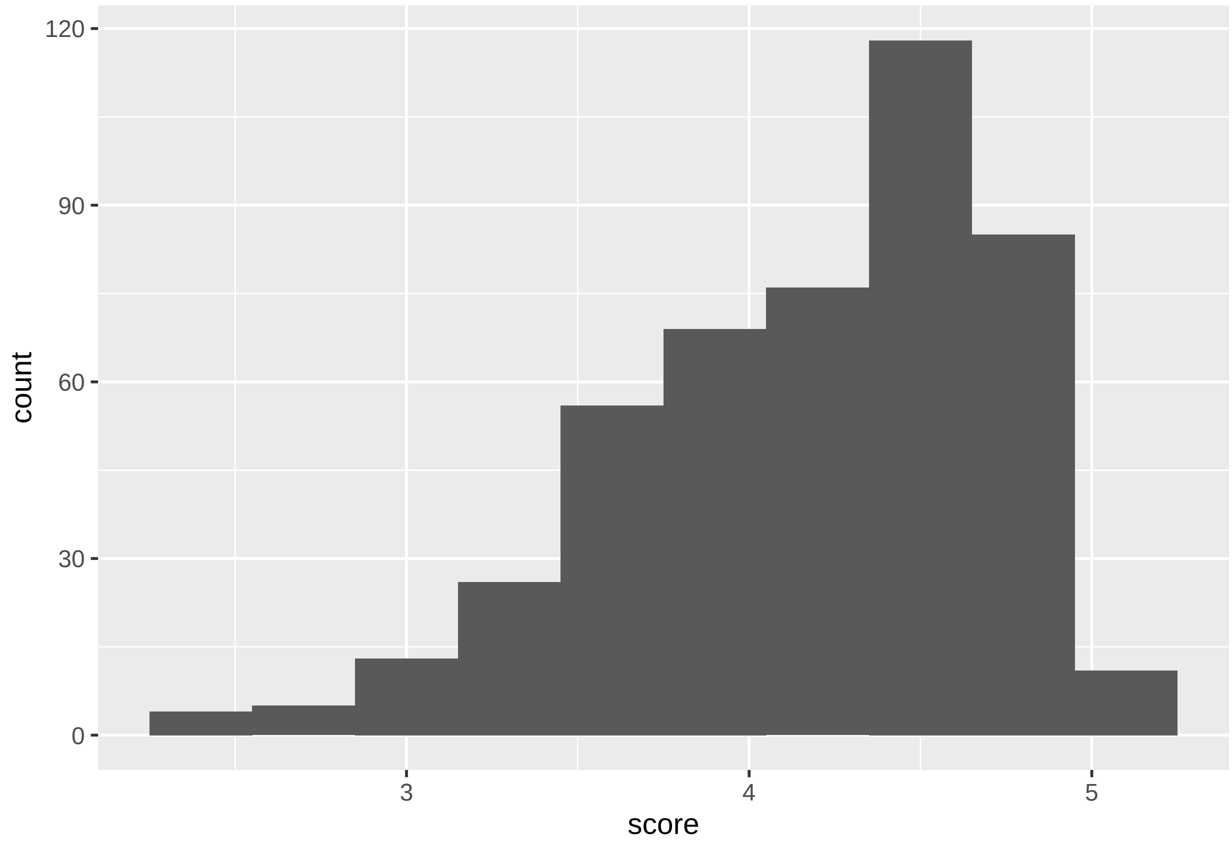
Insert your answer here

This study is observational. The original question cannot be answered by this data. A more accurate question would have been if there is a correlation between the beauty of the professor and the rating his courses receive.

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?

Insert your answer here

```r
ggplot(evals, aes(x = score)) +
  geom_histogram(bins =10)
```



the distribution is left skewed. This is what I expect since, in my experience, the default rating people tend to give is 4 stars. Below expectations gets rated 3 or lower and above expectations is 5 star rating. 3 stars is not the neutral rating it would seem to be on a scale of 1-5. This is particularly true when the person rating has a persona relationship with the thing they are rating such as in this case.

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

Insert your answer here

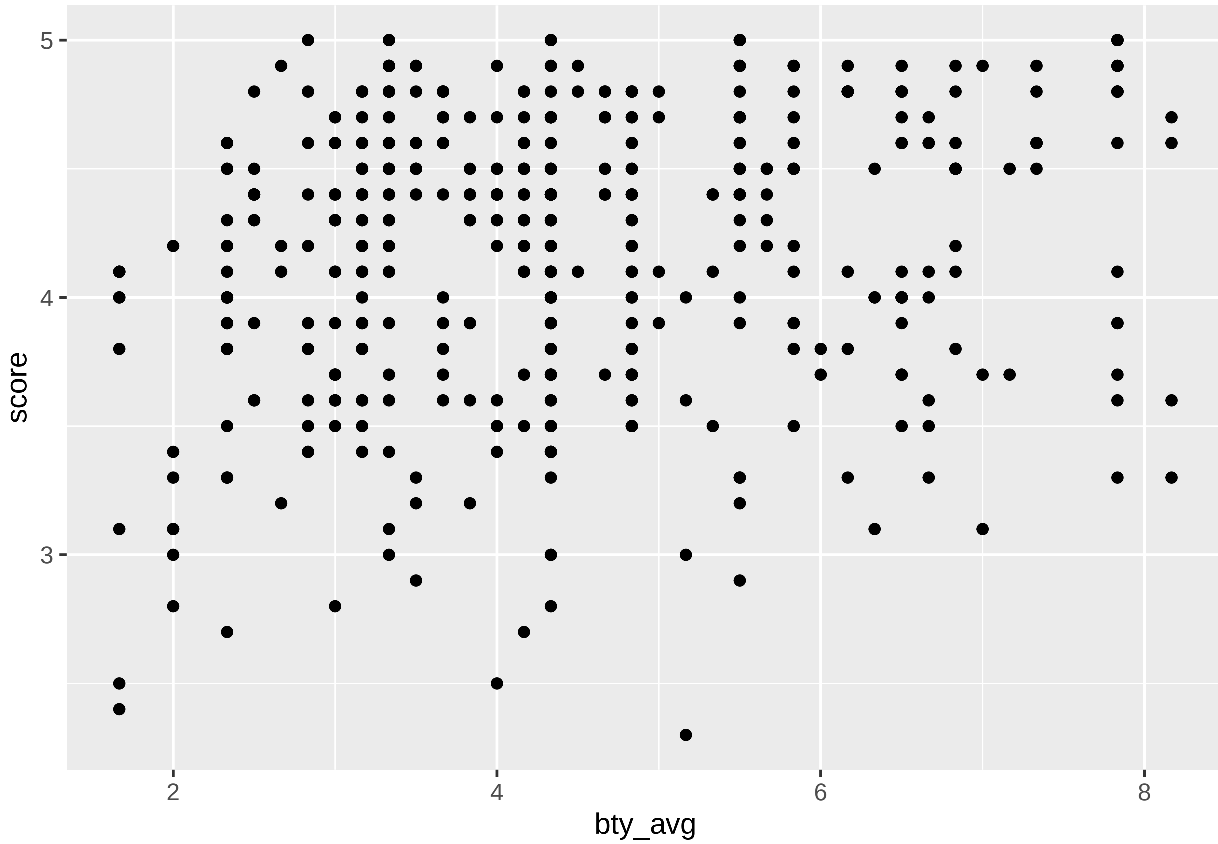
```
ggplot(evals, aes(x = gender, y = bty_avg))+
  geom_violin()
```

The above violin plot illustrates the relationship between the average beauty rating and the gender of the professor. The plot shows a higher median for females and is also much denser at a higher point than the plot for males. Additionally, the female distribution seems to be more evenly distributed. I used a violin plot instead of a boxplot since I think it conveys some more useful information. I use boxplots when I want to easily spot any outliers, something I didn't have to worry about in this scenario.

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:

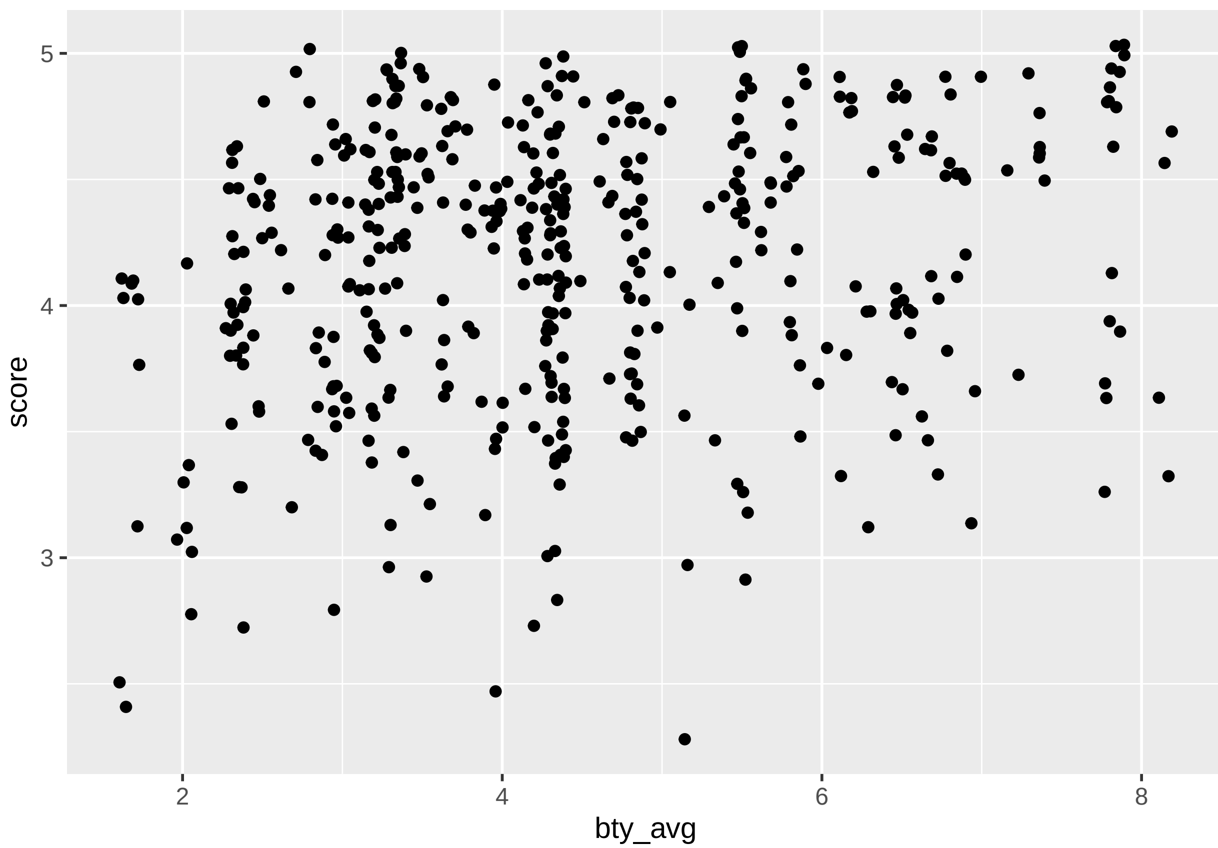
```
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_point()
```



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?

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()
```



Insert your answer here

The initial scatterplot didn't show us any overlapping points, this can lead not noticing the density is much higher in certain sections of the plot. By adding geom_jitter(), we can now see all the datapoints and draw better conclusions

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?

Insert your answer here

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

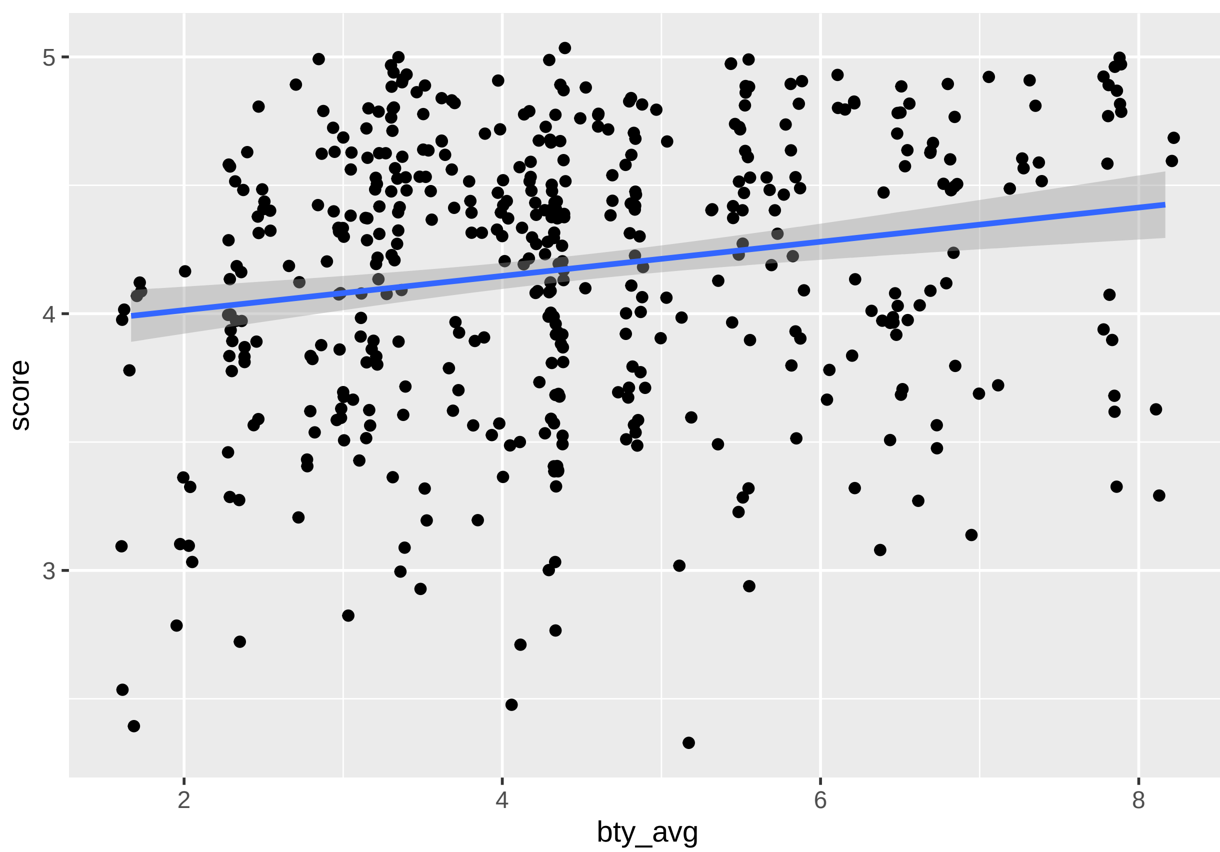
The linear model equation can be written as

$y = 3.88034 + 0.06664(bty_avg)$

The slope is represented as 0.06664. That means that for every increase in 1 for the average beauty ranking we can expect an increse of 0.6664 in the evaluation score. The p value is close to zero which implies a statistically significant predictor, however, the low r squared of 0.035 and the fact that it only predicts an increase of 0.06 implies that it isn't a very practically significant predictor.

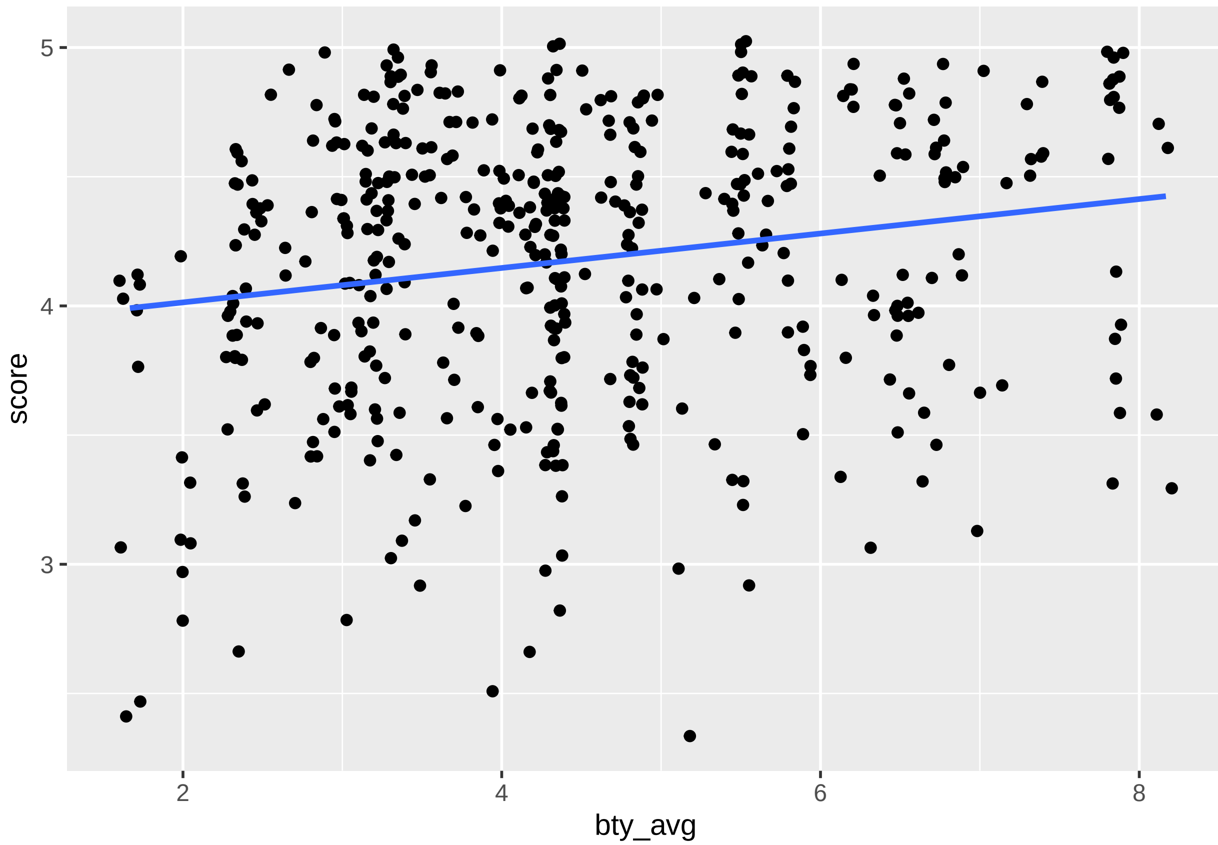
Add the line of the bet fit model to your plot using the following:

```r
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`.

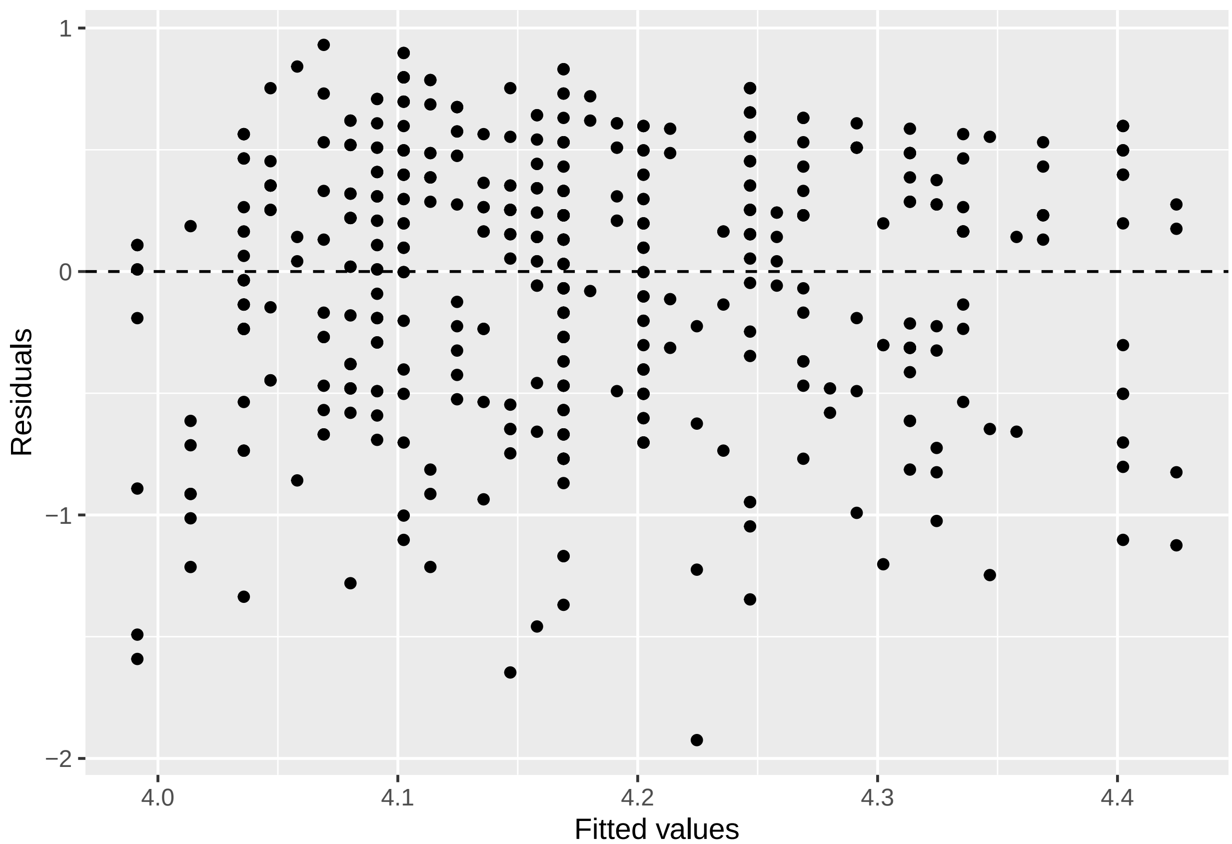
```r
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_jitter() +
  geom_smooth(method = "lm", se = FALSE)
```



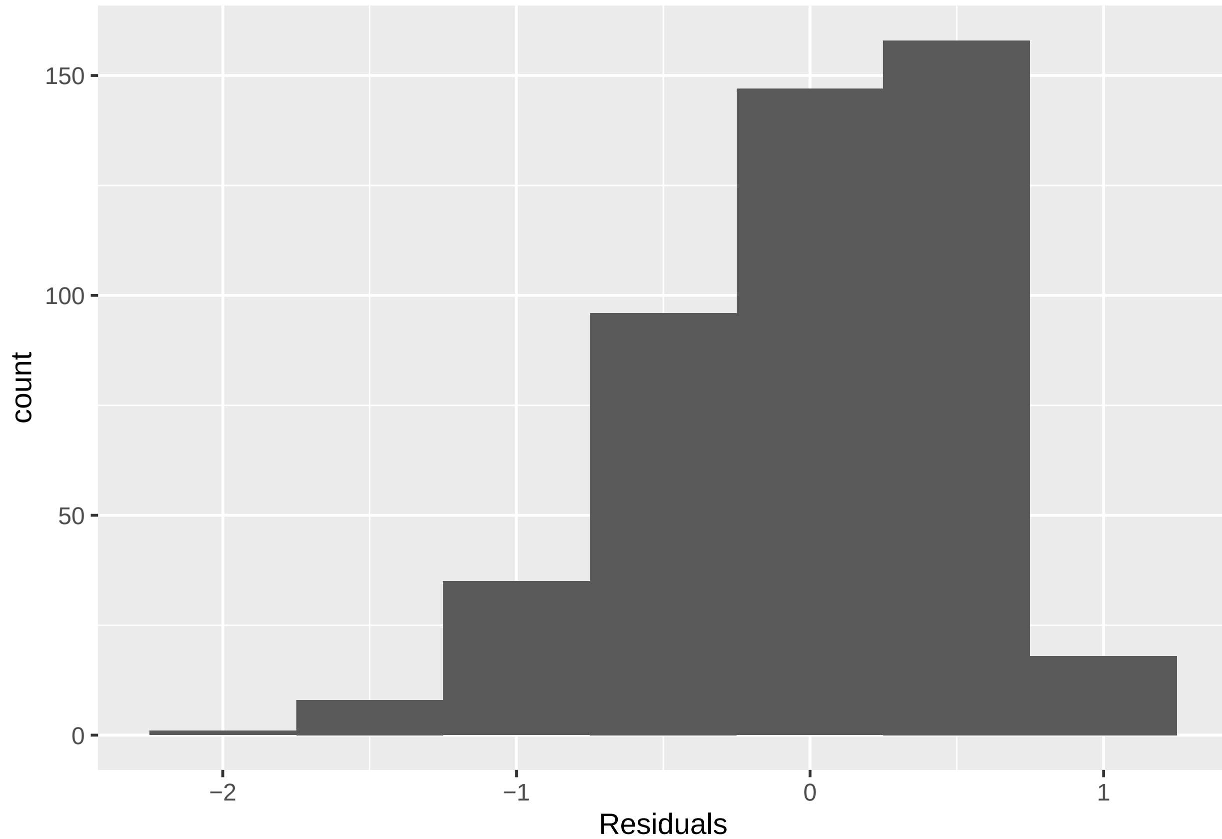
6. Use residual plots to evaluate whether the conditions of least squares regression are reasonable. Provide plots and comments for each one (see the Simple Regression Lab for a reminder of how to make these).

Insert your answer here

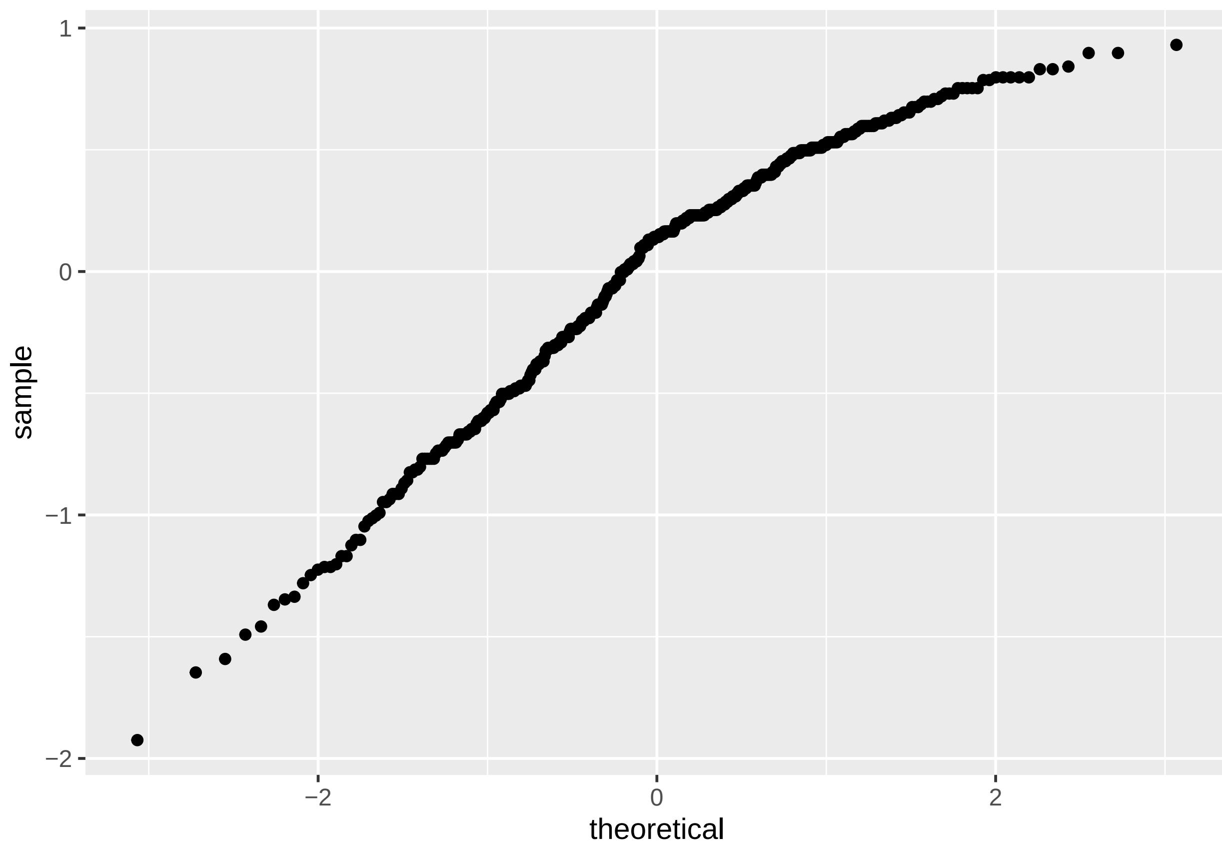
```r
#check for linearity of the residuals
ggplot(data = m_bty, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```




```
#check the distribution of the residuals
ggplot(data = m_bty, aes(x = .resid)) +
  geom_histogram(binwidth = 0.5) +
  xlab("Residuals")
```



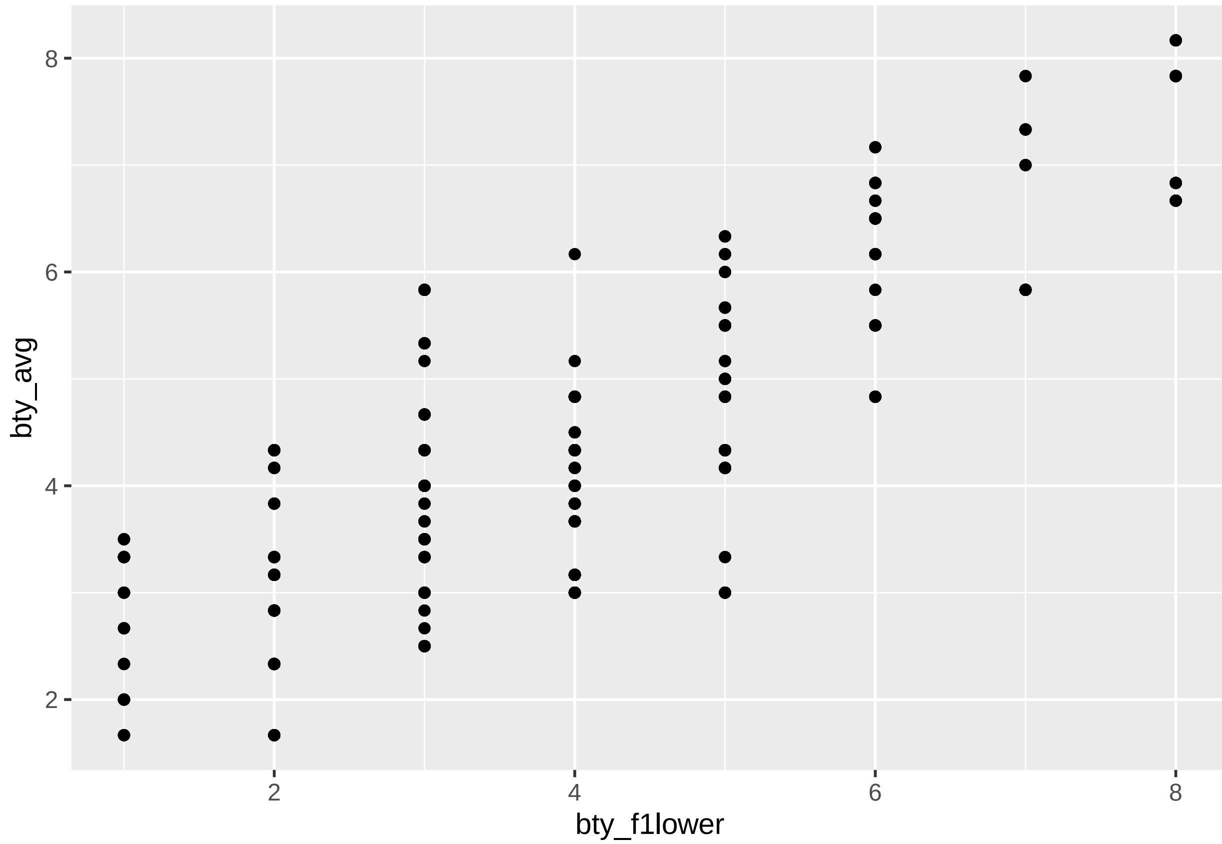

```
#create a normal probability plot
ggplot(data = m_bty, aes(sample = .resid)) +
  stat_qq()
```



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()
```

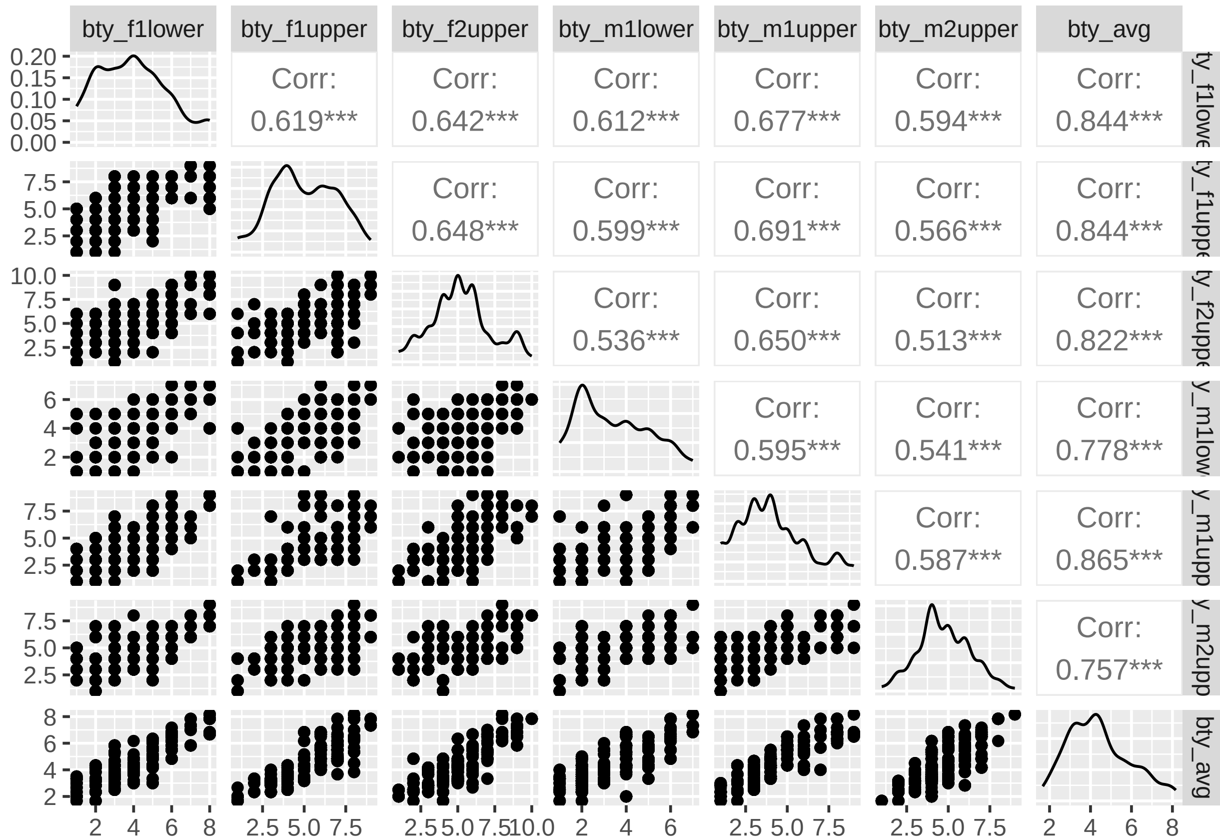



```
evals %>%
  summarise(cor(bty_avg, bty_f1lower))
```

```
## # 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:

```
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
##
```

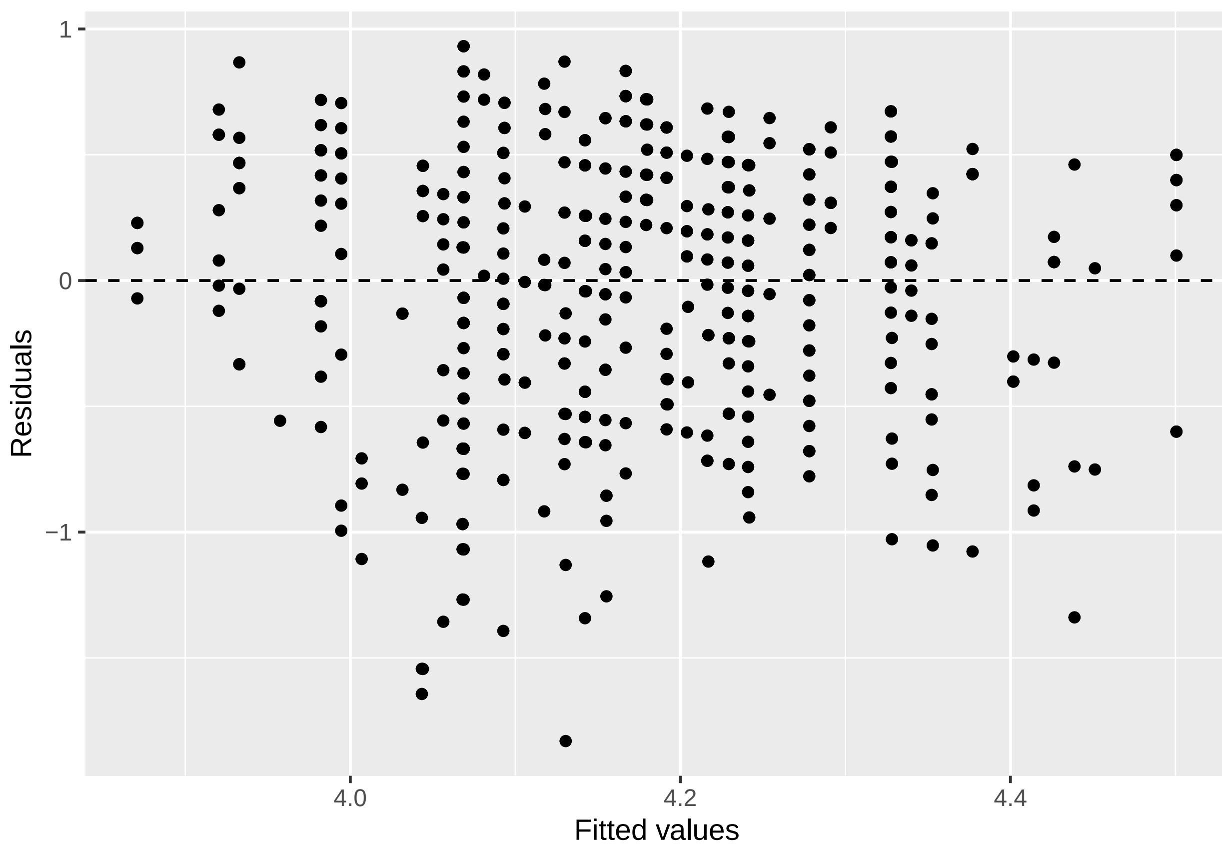
```
## Residual standard error: 0.5287 on 460 degrees of freedom
## Multiple R-squared:  0.05912,    Adjusted R-squared:  0.05503
## F-statistic: 14.45 on 2 and 460 DF,  p-value: 8.177e-07
```

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.

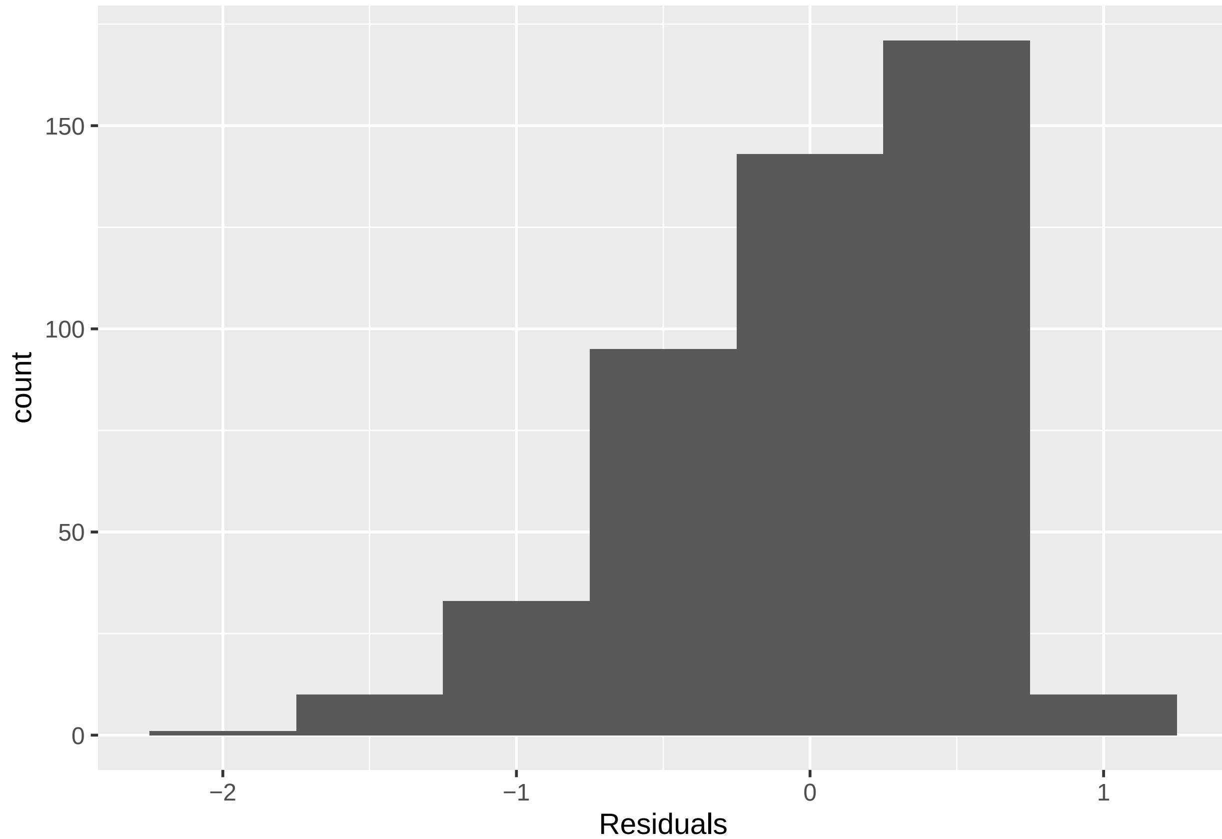
Insert your answer here

The below plots do seem to show that the conditions for the model are reasonable. The residuals are normally distributed and the points don't seem to have a pattern in the scatterplot.

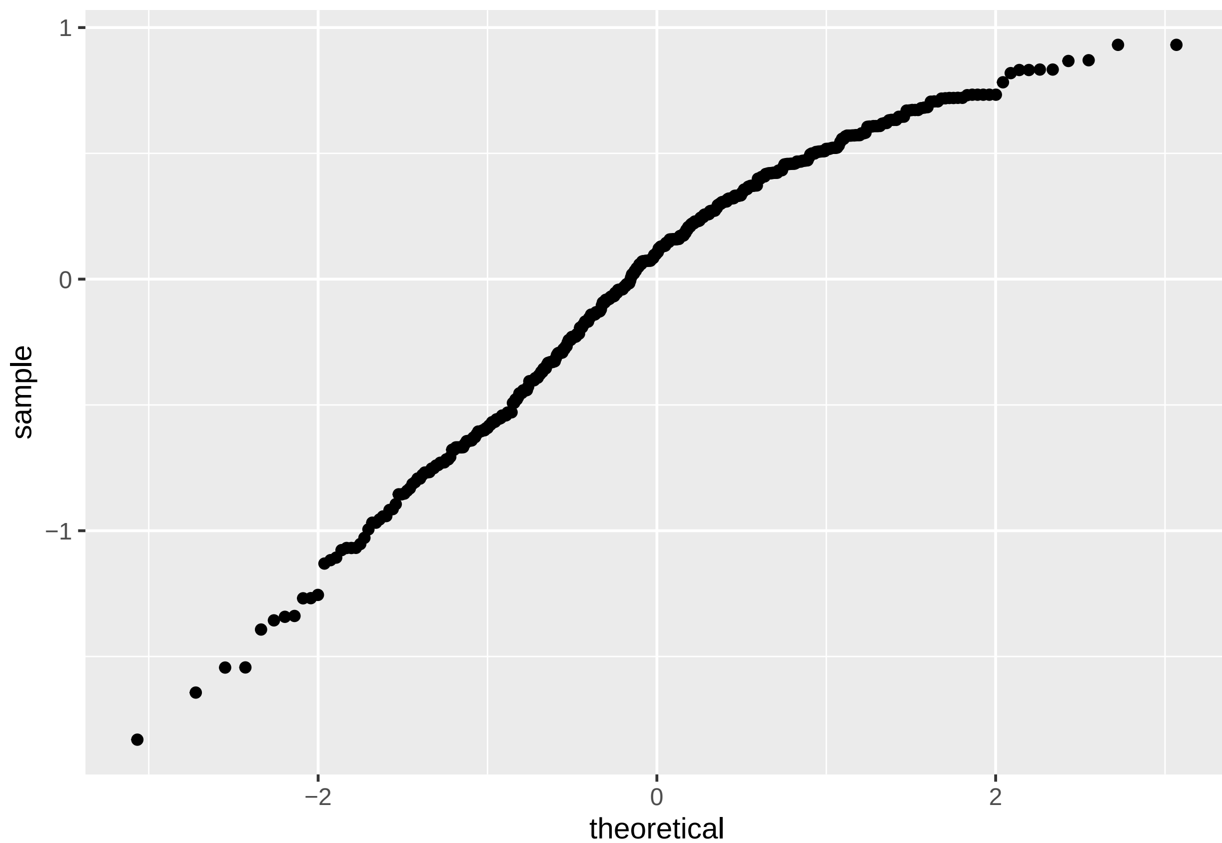
```
ggplot(data = m_bty_gen, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```




```
ggplot(data = m_bty_gen, aes(x = .resid)) +
  geom_histogram(binwidth = 0.5) +
  xlab("Residuals")
```




```
ggplot(data = m_bty_gen, aes(sample = .resid)) +
  stat_qq()
```



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`?

Insert your answer here

The below summary shows us that the slope of average beauty rises to 0.07416 in the new model. bty_avg is still a significant predictor and actually slightly increased the slope.

```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
```

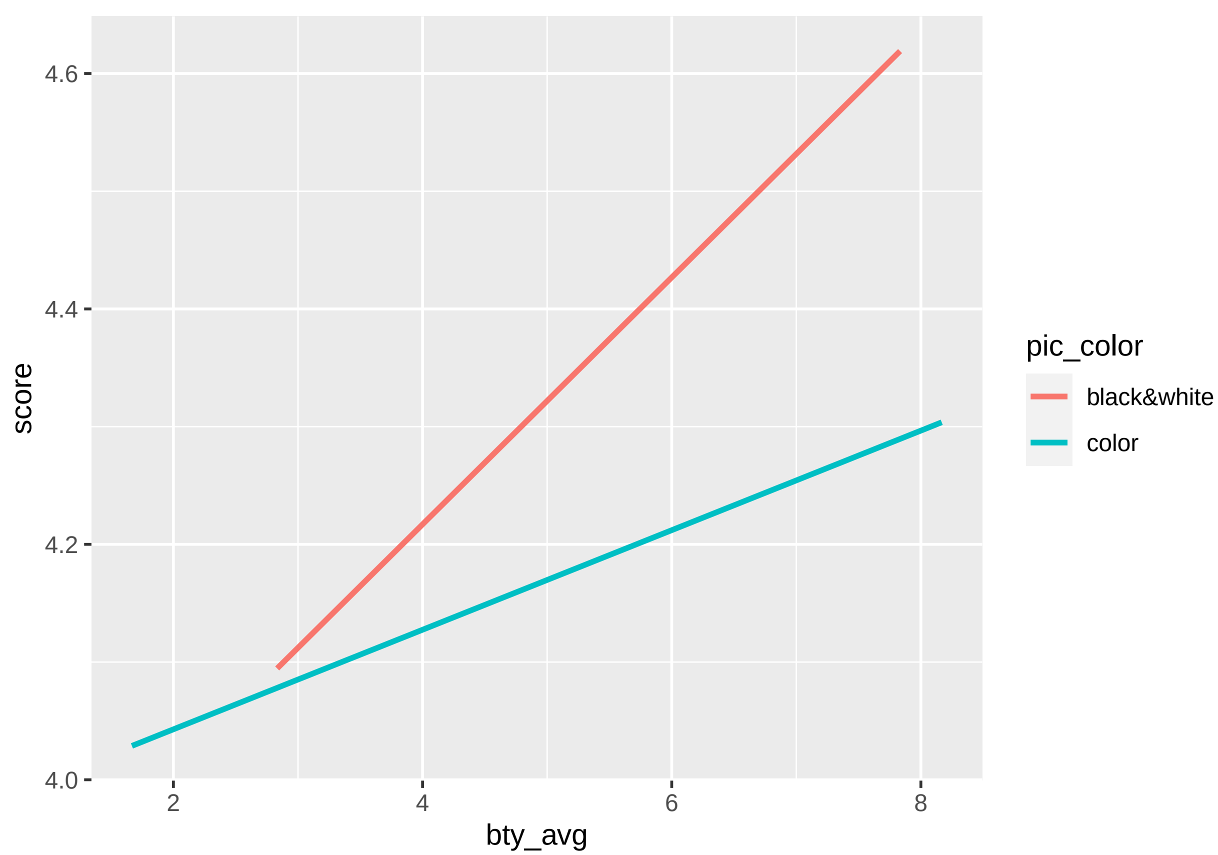
```
##
## Residual standard error: 0.5287 on 460 degrees of freedom
## Multiple R-squared:  0.05912,    Adjusted R-squared:  0.05503
## F-statistic: 14.45 on 2 and 460 DF,  p-value: 8.177e-07
```

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.

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

```r
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?

Insert your answer here

Using the summary from the earlier code, we can write the equation for those with color pictures as: Score=3.74734+0.07416 x btyavg+0.17239(1). Interestingly, professors with black and white photos get higher scores on average.

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`.

Insert your answer here

As can seen below, R just adds a line for every level of the variable. It always results in one less line than the amount of levels since one level gets categorized as 0.

```
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
```

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.

Insert your answer here

I would assume that language of the university where they got their degree would be the least significant and therefore have the highest p-value.

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.

Insert your answer here

The below summary shows us that the variable with the highest p-value is the number of professors for that class, or cls_profssingle. It has a p-value of 0.77, implying very low correlation. My hypothesis from earlier is very incorrect, language of the professors university has a relatively low p-value.

```r
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
```

13. Interpret the coefficient associated with the ethnicity variable.

Insert your answer here

The estimated effect of being in the "not minority" group in the ethnicity variable is an increase of 0.1234929 in score, but since it has a p-value of 0.11698 which is above the usual metric of 0.05, it might not be considered statistically significant.

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?

Insert your answer here

After dropping the variable with the highest p-value, the coefficients of the other variables changed very slightly at the later decimal points and the adjusted r squared rose a bit, suggesting a better model fit. This implies heavy colinearity between the dropped variable and the model.

```r
m_mostly_full <- 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_mostly_full)
```

```
##
## 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
```

```r
#summary(m_full)
```

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.

Insert your answer here

The equation of the line with all the variables having a p-value of 0.05 or lower is $Score = 3.907030 + (genderx0.202597) + (ethnicityx0.163818) + (languagex - 0.246683) + (agex - 0.006925) + (clas_{p}erc_{e}valx0.004942) + (cls_{c}reditsx0.517205) + (bty_{a}vgx0.046732) + (pic_{o}utfitx - 0.113939) + (pic_{c}olorx - 0.18087)$

```r
final_model <- step(m_full, direction = "backward", criterion = "p-value")
```

```
## Start:  AIC=-630.9
## score ~ rank + gender + ethnicity + language + age + cls_perc_eval +
##     cls_students + cls_level + cls_profs + cls_credits + bty_avg +
##     pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - cls_profs      1    0.0197 111.11 -632.82
## - cls_level      1    0.2740 111.36 -631.76
## - cls_students   1    0.3599 111.44 -631.40
## - rank           2    0.8930 111.98 -631.19
## <none>                      111.08 -630.90
## - pic_outfit     1    0.5768 111.66 -630.50
## - ethnicity      1    0.6117 111.70 -630.36
## - language       1    1.0557 112.14 -628.52
## - bty_avg        1    1.2967 112.38 -627.53
## - age            1    2.0456 113.13 -624.45
## - pic_color      1    2.2893 113.37 -623.46
## - cls_perc_eval  1    2.9698 114.06 -620.69
## - gender         1    4.1085 115.19 -616.09
## - cls_credits    1    4.6495 115.73 -613.92
##
## Step:  AIC=-632.82
## score ~ rank + gender + ethnicity + language + age + cls_perc_eval +
##     cls_students + cls_level + cls_credits + bty_avg + pic_outfit +
##     pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - cls_level      1    0.2752 111.38 -633.67
## - cls_students   1    0.3893 111.49 -633.20
## - rank           2    0.8939 112.00 -633.11
## <none>                      111.11 -632.82
## - pic_outfit     1    0.5574 111.66 -632.50
## - ethnicity      1    0.6728 111.78 -632.02
## - language       1    1.0442 112.15 -630.49
## - bty_avg        1    1.2872 112.39 -629.49
## - age            1    2.0422 113.15 -626.39
## - pic_color      1    2.3457 113.45 -625.15
## - cls_perc_eval  1    2.9502 114.06 -622.69
## - gender         1    4.0895 115.19 -618.08
## - cls_credits    1    4.7999 115.90 -615.24
##
## Step:  AIC=-633.67
## score ~ rank + gender + ethnicity + language + age + cls_perc_eval +
##     cls_students + cls_credits + bty_avg + pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - cls_students   1    0.2459 111.63 -634.65
## - rank           2    0.8140 112.19 -634.30
## <none>                      111.38 -633.67
## - pic_outfit     1    0.6618 112.04 -632.93
## - ethnicity      1    0.8698 112.25 -632.07
```

```
## - language       1    0.9015 112.28 -631.94
## - bty_avg        1    1.3694 112.75 -630.02
## - age            1    1.9342 113.31 -627.70
## - pic_color      1    2.0777 113.46 -627.12
## - cls_perc_eval  1    3.0290 114.41 -623.25
## - gender         1    3.8989 115.28 -619.74
## - cls_credits    1    4.5296 115.91 -617.22
##
## Step:  AIC=-634.65
## score ~ rank + gender + ethnicity + language + age + cls_perc_eval +
##     cls_credits + bty_avg + pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - rank           2    0.7892 112.42 -635.39
## <none>                       111.63 -634.65
## - ethnicity      1    0.8832 112.51 -633.00
## - pic_outfit     1    0.9700 112.60 -632.65
## - language       1    1.0338 112.66 -632.38
## - bty_avg        1    1.5783 113.20 -630.15
## - pic_color      1    1.9477 113.57 -628.64
## - age            1    2.1163 113.74 -627.96
## - cls_perc_eval  1    2.7922 114.42 -625.21
## - gender         1    4.0945 115.72 -619.97
## - cls_credits    1    4.5163 116.14 -618.29
##
## Step:  AIC=-635.39
## score ~ gender + ethnicity + language + age + cls_perc_eval +
##     cls_credits + bty_avg + pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## <none>                       112.42 -635.39
## - pic_outfit     1    0.7141 113.13 -634.46
## - ethnicity      1    1.1790 113.59 -632.56
## - language       1    1.3403 113.75 -631.90
## - age            1    1.6847 114.10 -630.50
## - pic_color      1    1.7841 114.20 -630.10
## - bty_avg        1    1.8553 114.27 -629.81
## - cls_perc_eval  1    2.9147 115.33 -625.54
## - gender         1    4.0577 116.47 -620.97
## - cls_credits    1    6.1208 118.54 -612.84
```

```r
summary(final_model)
```

```
##
## Call:
## lm(formula = score ~ gender + ethnicity + 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|)
```

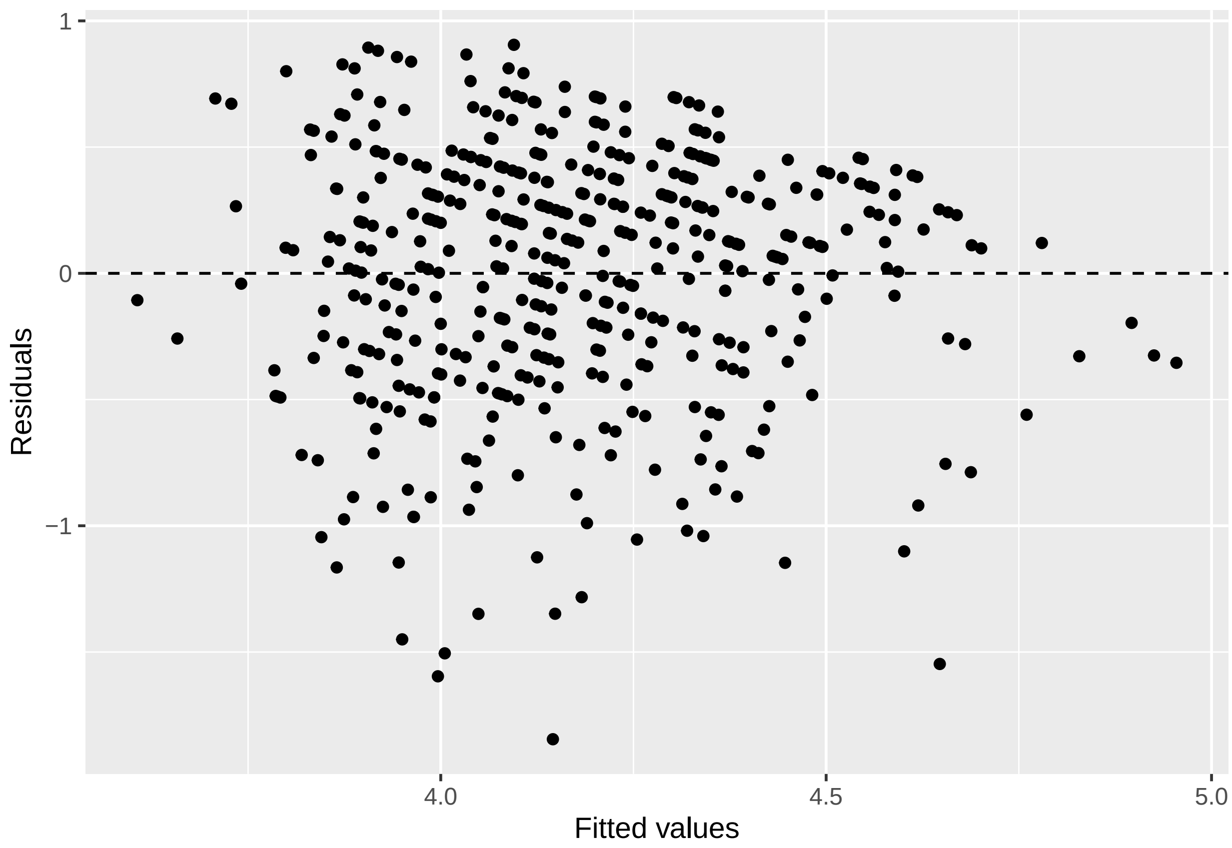
```
## (Intercept)            3.907030   0.244889  15.954  < 2e-16 ***
## gendermale             0.202597   0.050102   4.044 6.18e-05 ***
## ethnicitynot minority  0.163818   0.075158   2.180 0.029798 *
## languagenon-english   -0.246683   0.106146  -2.324 0.020567 *
## age                   -0.006925   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
```

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

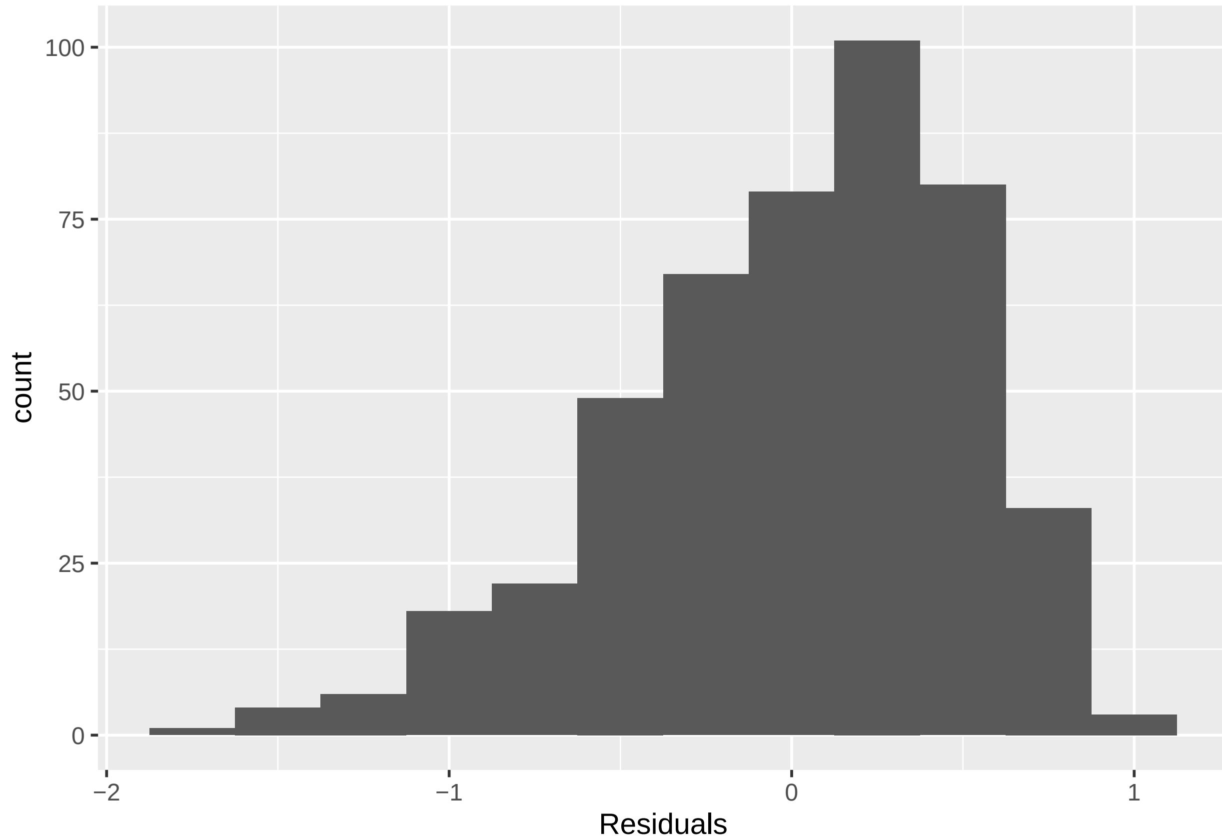
Insert your answer here

The scatterplot below seems to show that the conditions are not met. There is higher density above the x axis and towards the right of the y axis. The distribution is fairly normal.

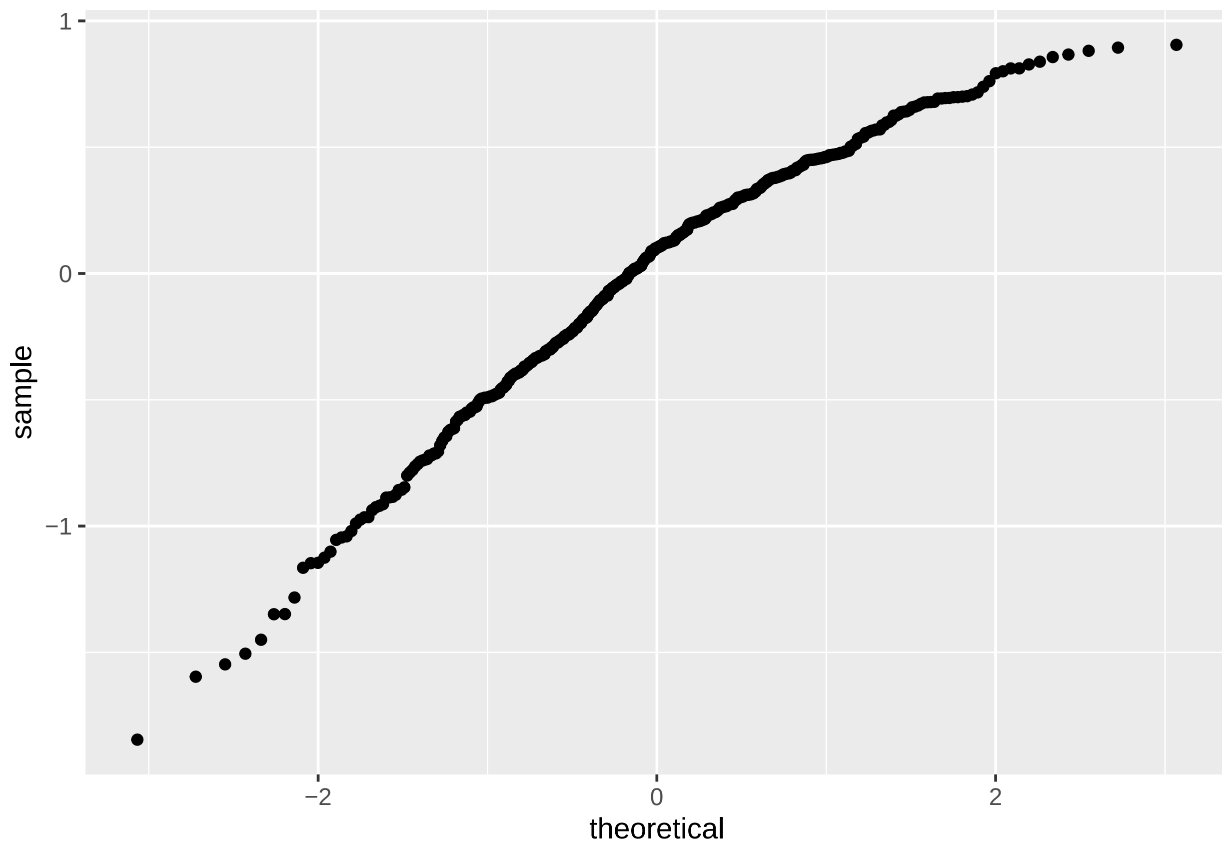
```
ggplot(data = final_model, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```




```
ggplot(data = final_model, aes(x = .resid)) +
  geom_histogram(binwidth = 0.25) +
  xlab("Residuals")
```




```
ggplot(data = final_model, aes(sample = .resid)) +
  stat_qq()
```



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?

Insert your answer here

Yes, if a professor has taught multiple courses. This would make the variables dependent on each other (since the same professor would probably receive similar ratings across classes). It may help to average every professors' scores across classes and rerun the analysis.

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.

Insert your answer here

Based on the best fitting model, one would expect a professor who is male, non-minority, speaks English, is young, higher percentage of completed evaluations, teaches one credit classes, has a higher beauty ranking, has formal clothes in his picture and uses black and white pictures to get higher evaluation scores.

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

Insert your answer here

I would not apply this model to other universaties since many of the variables can be perceived as positive or negative depending on the culture of the area around it. Obviously, Engish-speaking would be much less of a factor in non-English speaking countries, age can be perceived differently in different cultures and other such considerations.