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.

Creating a reproducible lab report

To create your new lab report, in RStudio, go to New File -> R Markdown... Then, choose From Template and then choose `Lab Report for OpenIntro Statistics Labs` from the list of templates.

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
```

```
evals
```

```
## # A tibble: 463 x 23
##    course_id prof_id score rank    ethnicity gender language   age cls_perc_eval
##        <int>   <int> <dbl> <fct>   <fct>     <fct>  <fct>    <int>         <dbl>
##  1         1       1   4.7 tenure~ minority  female english     36          55.8
##  2         2       1   4.1 tenure~ minority  female english     36          68.8
##  3         3       1   3.9 tenure~ minority  female english     36          60.8
##  4         4       1   4.8 tenure~ minority  female english     36          62.6
##  5         5       2   4.6 tenured not mino~ male   english     59          85
##  6         6       2   4.3 tenured not mino~ male   english     59          87.5
##  7         7       2   2.8 tenured not mino~ male   english     59          88.6
##  8         8       3   4.1 tenured not mino~ male   english     51         100
##  9         9       3   3.4 tenured not mino~ male   english     51          56.9
## 10        10       4   4.5 tenured not mino~ female english     40          87.0
## # ... with 453 more rows, and 14 more variables: cls_did_eval <int>,
## #   cls_students <int>, cls_level <fct>, cls_profs <fct>, cls_credits <fct>,
## #   bty_f1lower <int>, bty_f1upper <int>, bty_f2upper <int>, bty_m1lower <int>,
## #   bty_m1upper <int>, bty_m2upper <int>, bty_avg <dbl>, pic_outfit <fct>,
## #   pic_color <fct>
```

Exploring the data

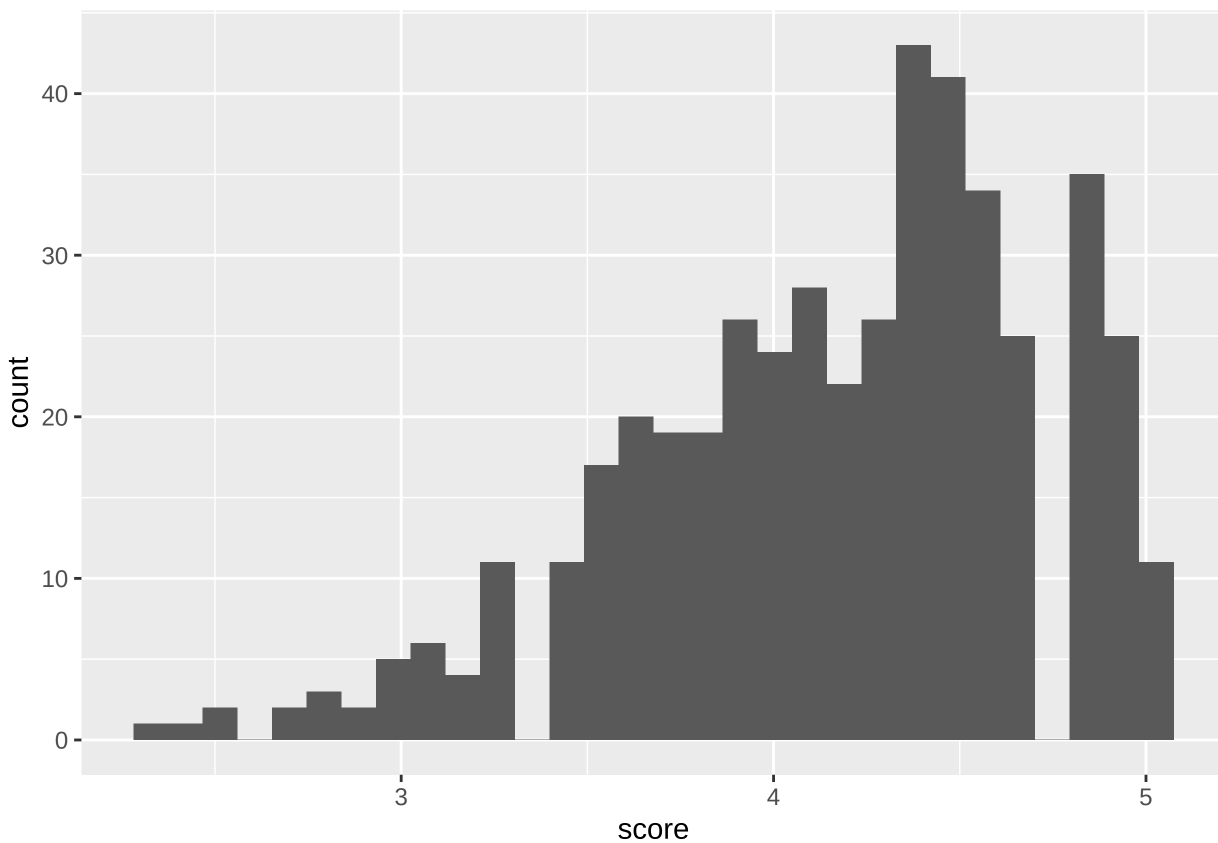
1. Is this an observational study or an experiment? The original research question posed in the paper is whether beauty leads directly to the differences in course evaluations. Given the study design, is it possible to answer this question as it is phrased? If not, rephrase the question.

   Answer: I believe this study to be experimental because the researcher is providing and and controlling the variables. The researcher is not observing the subjects themselves.

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?

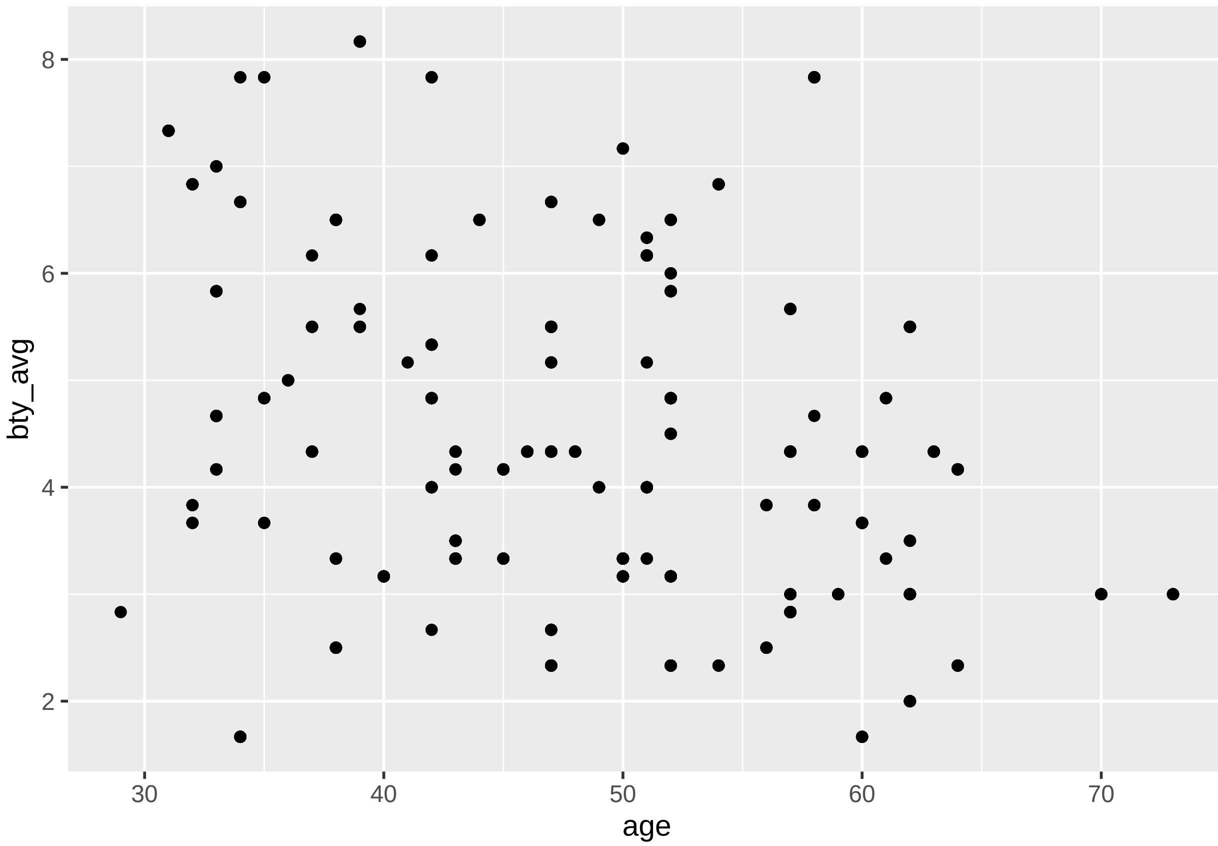
   Answer: From the histogram below the score distribution seems to be close to a normal distribution with a skew to the left.

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

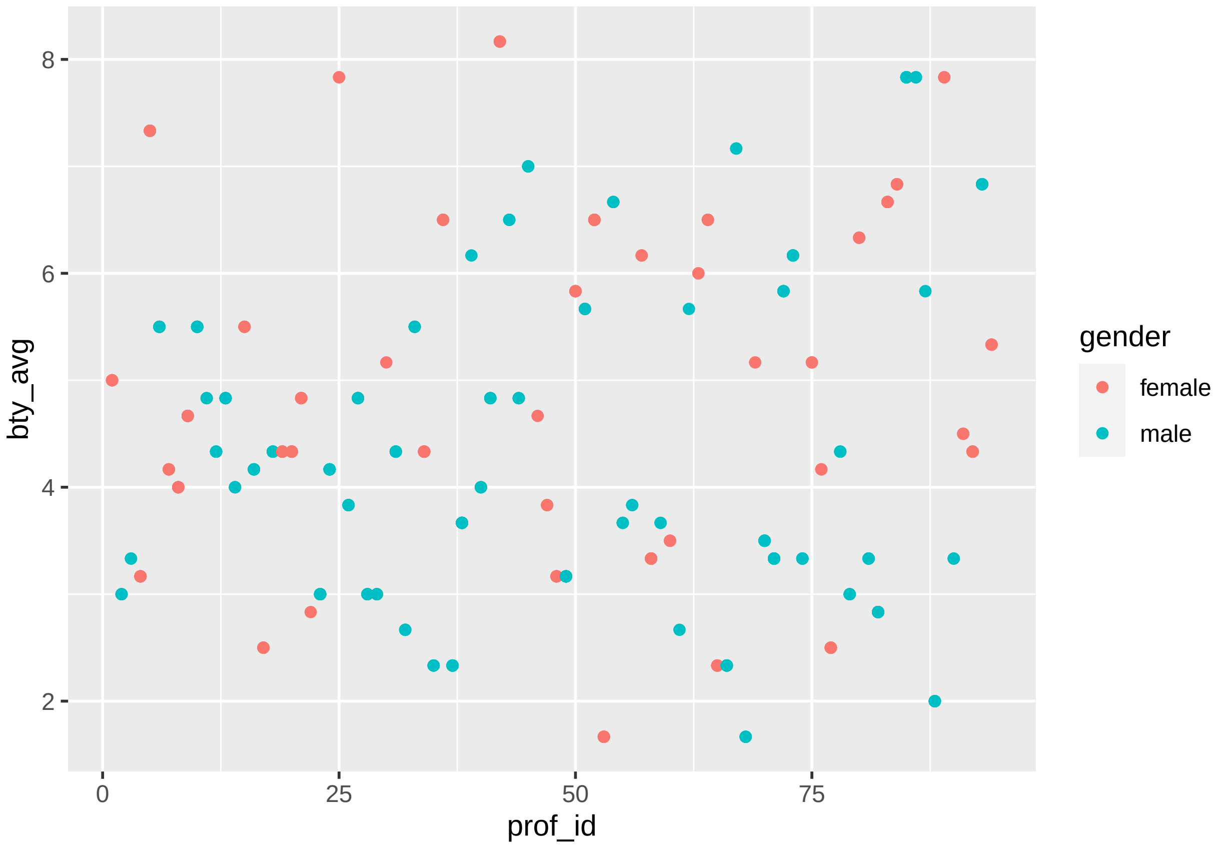


3. Excluding `score`, select two other variables and describe their relationship with each other using an appropriate visualization.

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




```
ggplot(data = evals, aes(x = prof_id, y = bty_avg, color=gender)) + geom_point()
```

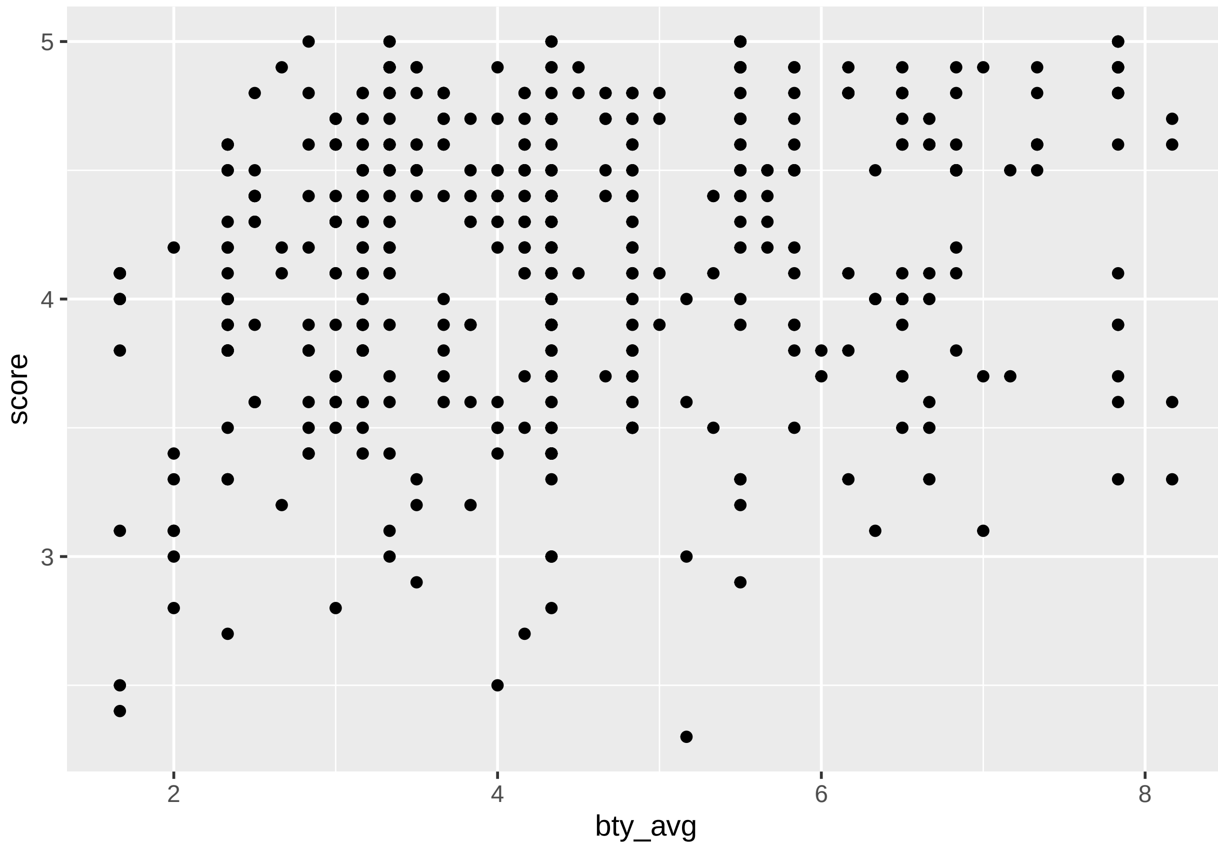



```
Answer:  From the scatterplot above when comparing age to average beauty (bty_avg), the older an
         instructor gets the less attractive they become to their students.  This shows a negative
         linear regression relationship.  And by the second scatterplot, it appears that female
         instructors had a higher beauty average (proportionately) than their male counterparts.
```

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()
```

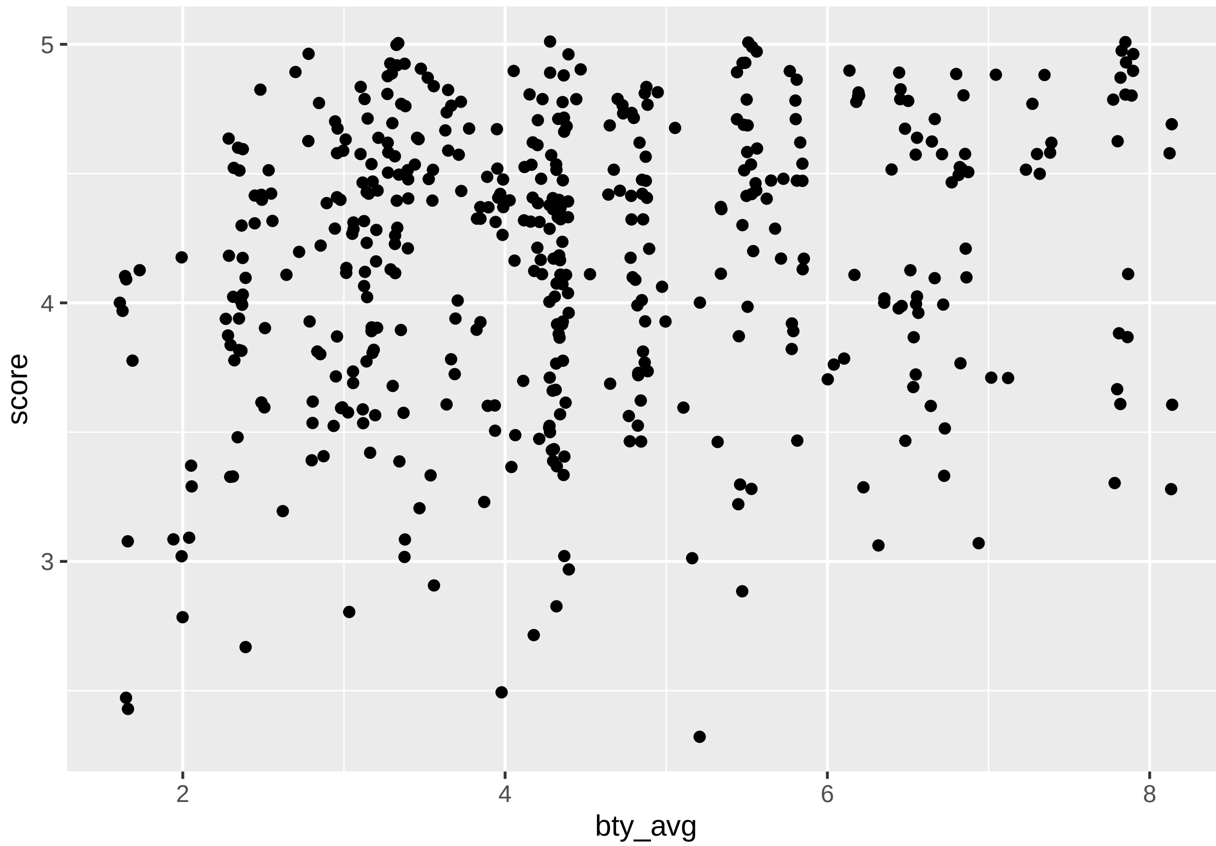


Before you draw conclusions about the trend, compare the number of observations in the data frame with the approximate number of points on the scatterplot. Is anything awry?

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

   Answer: It seems that geom_point is combining scores that are very close in value thus giving it a very clean look. While geom_jitter adds a little variation to each observation thus separating each point and so now each observation is represented in the scatterplot.

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



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?

   Answer: Linear model equation:
   $\hat{y} = \beta_0 + \beta_1$ x bty_avg

```
(m_bty <- lm(score ~ bty_avg, data = evals))
```

```
##
## Call:
## lm(formula = score ~ bty_avg, data = evals)
##
## Coefficients:
## (Intercept)      bty_avg
##     3.88034      0.06664
```

```
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
```

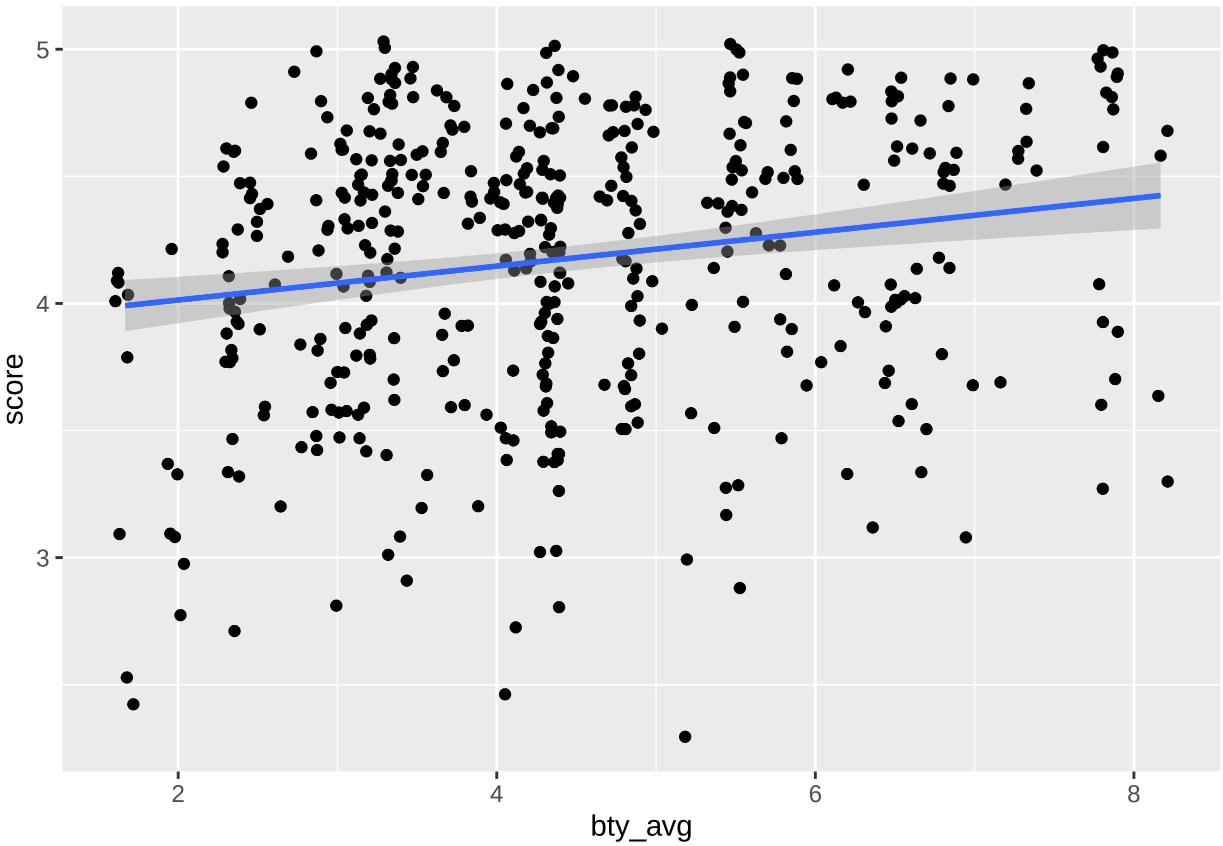
```r
evals %>%
  summarise(cor(score, bty_avg))
```

```
## # A tibble: 1 x 1
##   `cor(score, bty_avg)`
##                   <dbl>
## 1                 0.187
```

```
Answer: Average beauty score does not appear to be a practically significant predictor because
        from initial model calculation the slope (bty_avg) is only 6.7% which is very low.
        From the summary of the model m_bty, one can see the the Adjusted R-sqaured is even
        lower at 3.3% which is a better indicator of relationship strength.
```

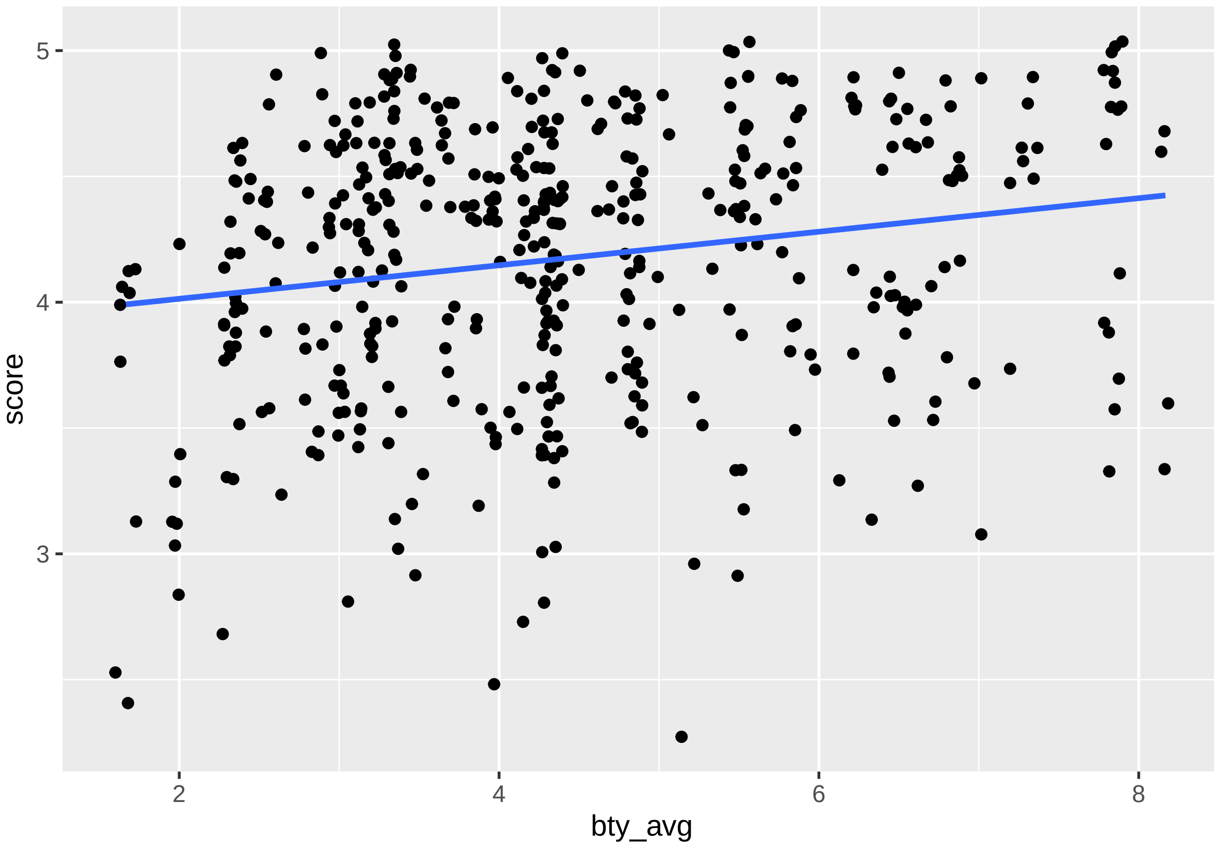
Add the line of the bet fit model to your plot using the following:

```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`.

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



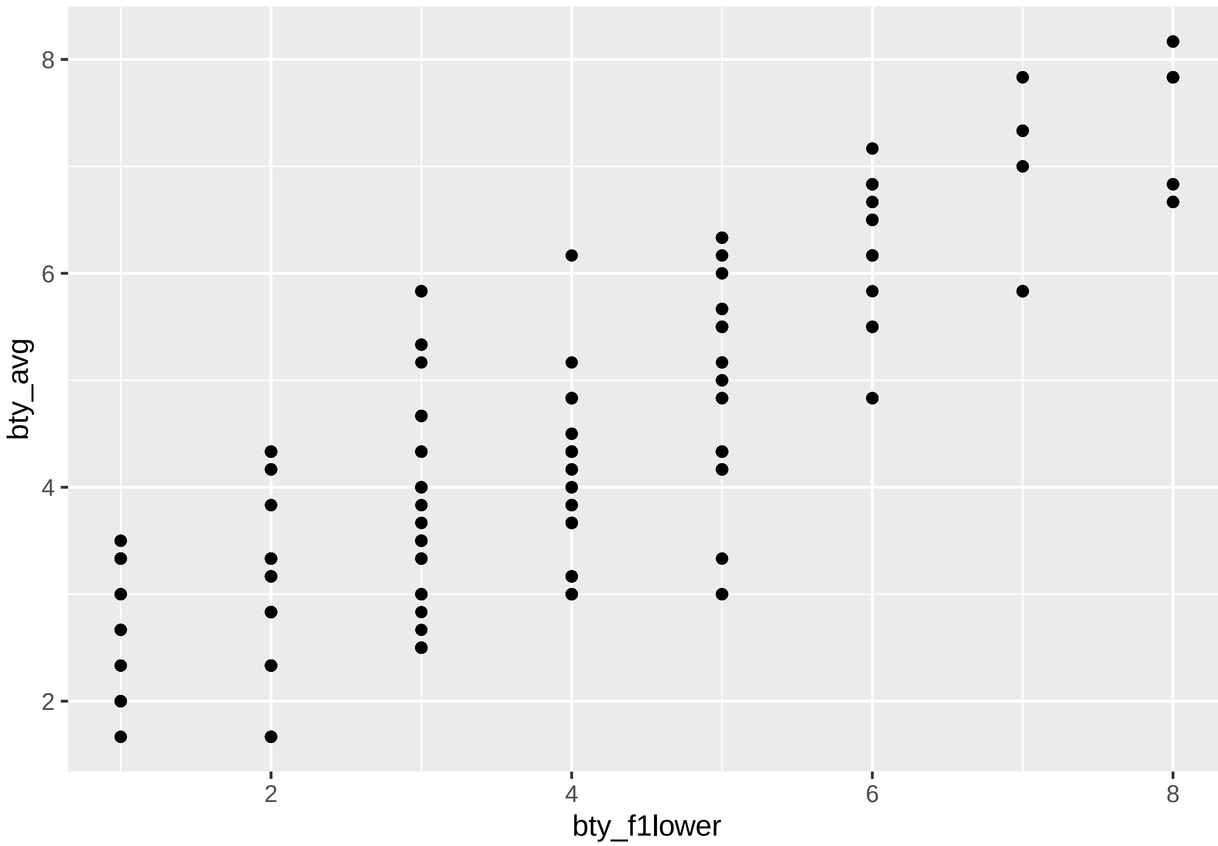
6. Use residual plots to evaluate whether the conditions of least squares regression are reasonable. Provide plots and comments for each one (see the Simple Regression Lab for a reminder of how to make these).

   Answer: The linear regression plot with least sum of the squares line supports the assumption that bty_avg is not a significant predictor of score because the positive and negative residuals are not tightly packed along the blue line.

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.

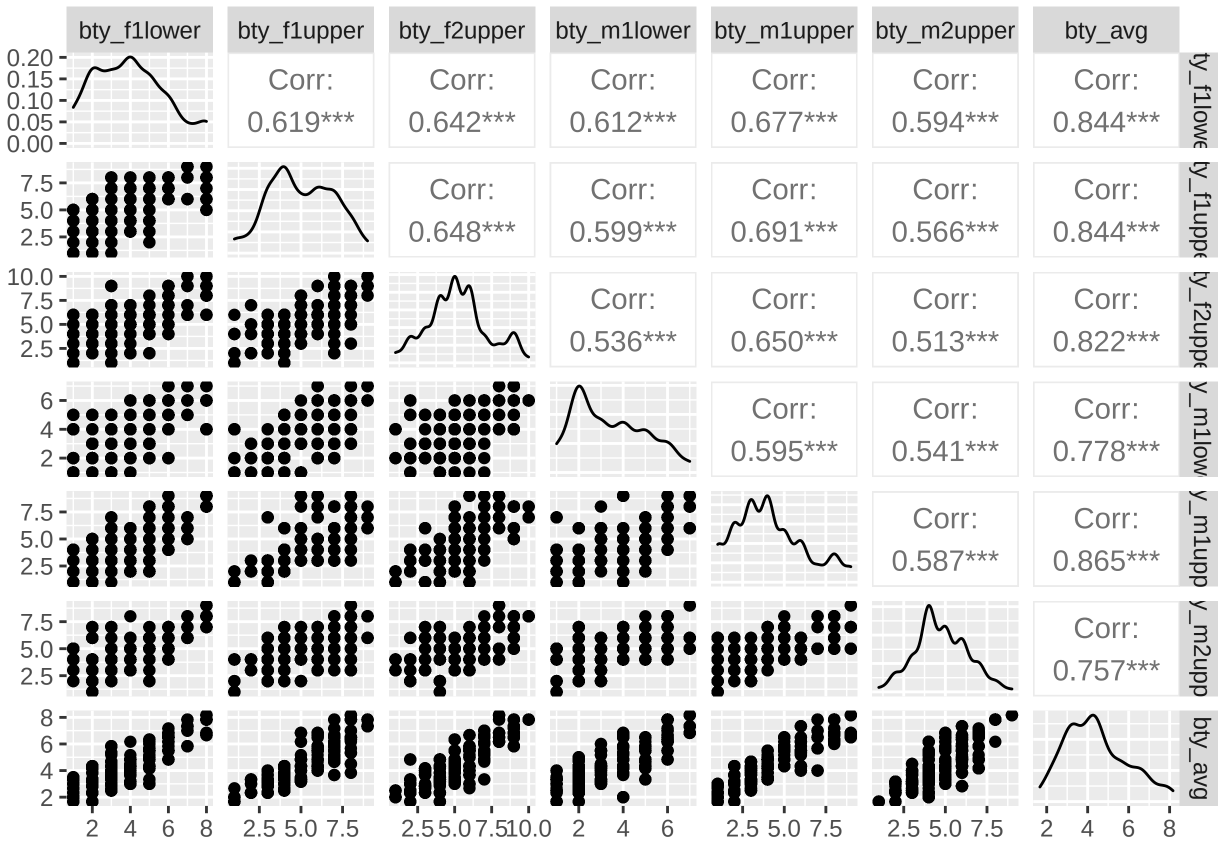
```
ggplot(data = evals, aes(x = bty_f1lower, y = bty_avg)) +
  geom_point()
```




```
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:



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)
```

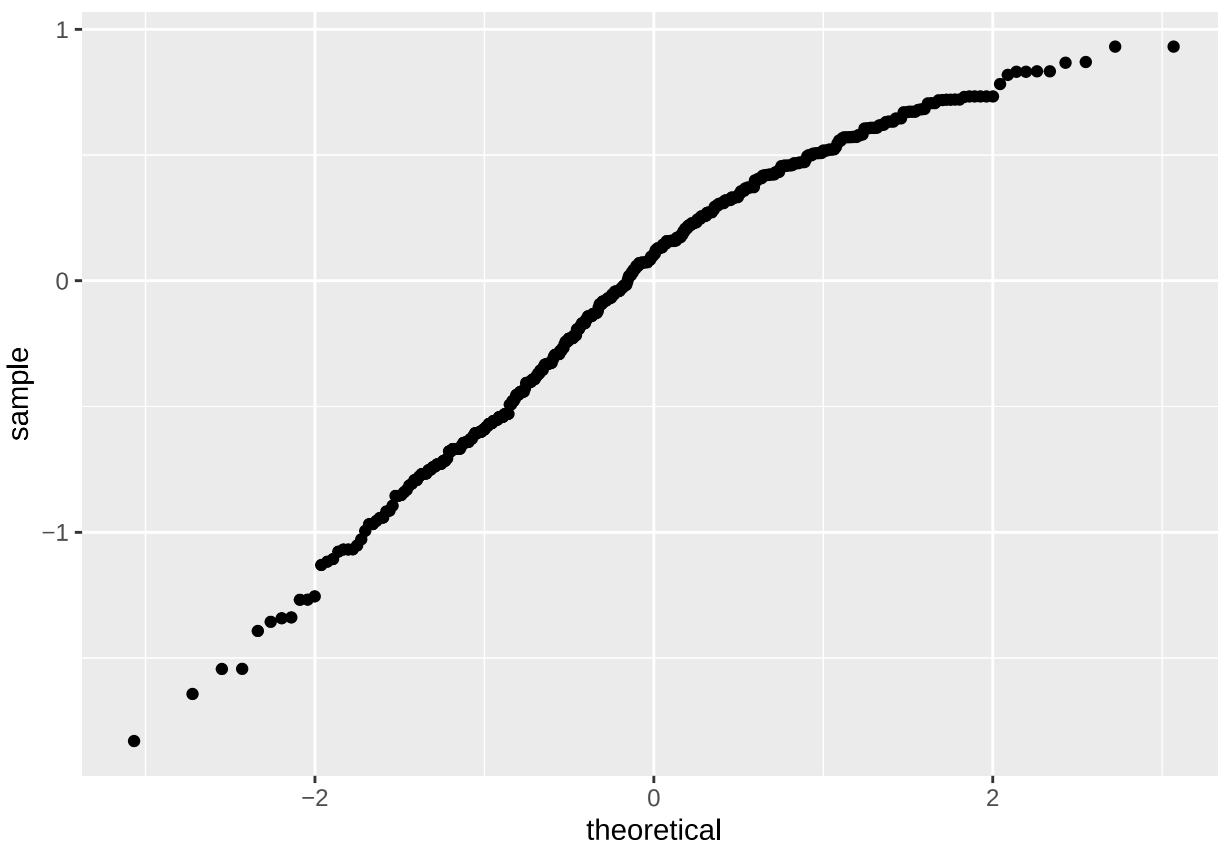
```
##
## Call:
## lm(formula = score ~ bty_avg + gender, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8305 -0.3625  0.1055  0.4213  0.9314
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.74734    0.08466  44.266  < 2e-16 ***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

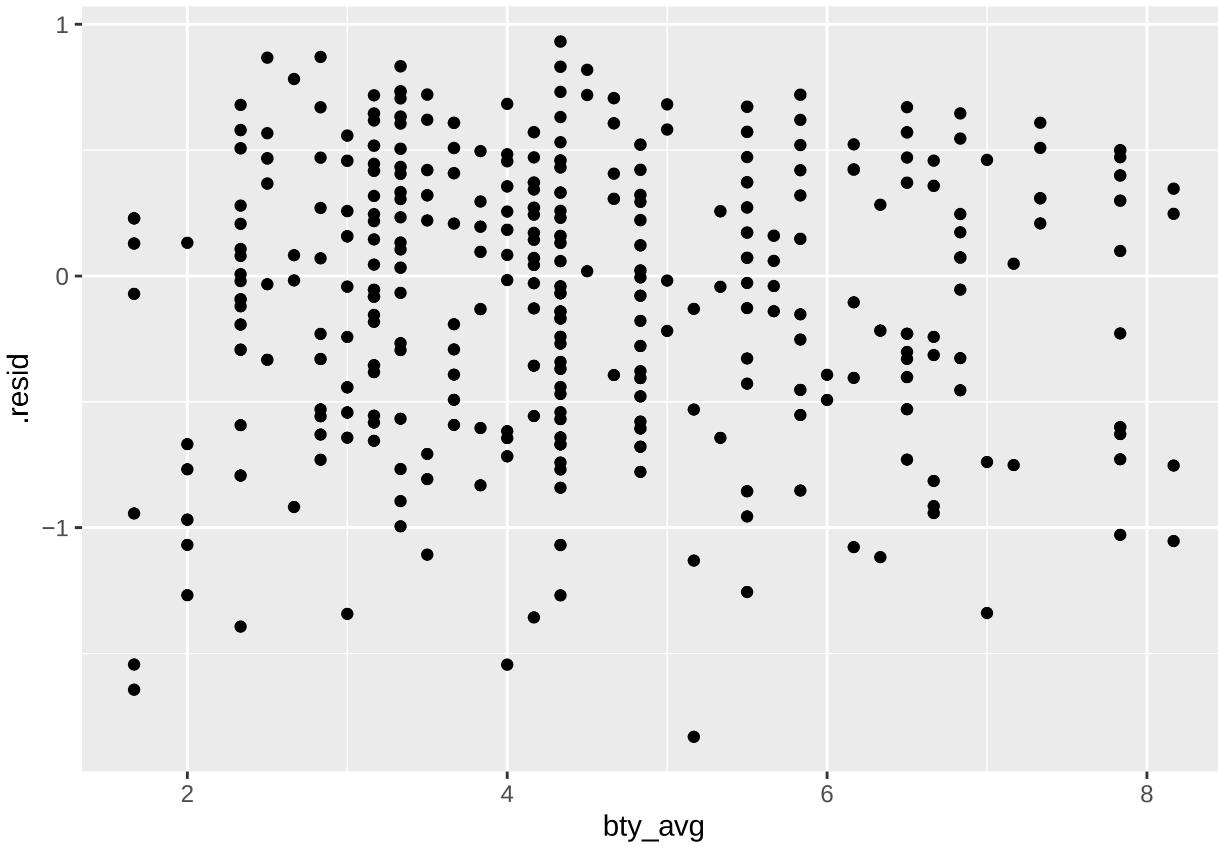
   Assumption 1: Are residuals nearly normal. It appears that the residuals are not nearly as they seem to bow out from what would be the least sum of the squares line.

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




```
Assumption 2: The variablility of the residuals is nearly constant.  Although you can some
    variability among the residuals there is still a lot of scattering so I would have to
    assess that variability is not very constant.
```

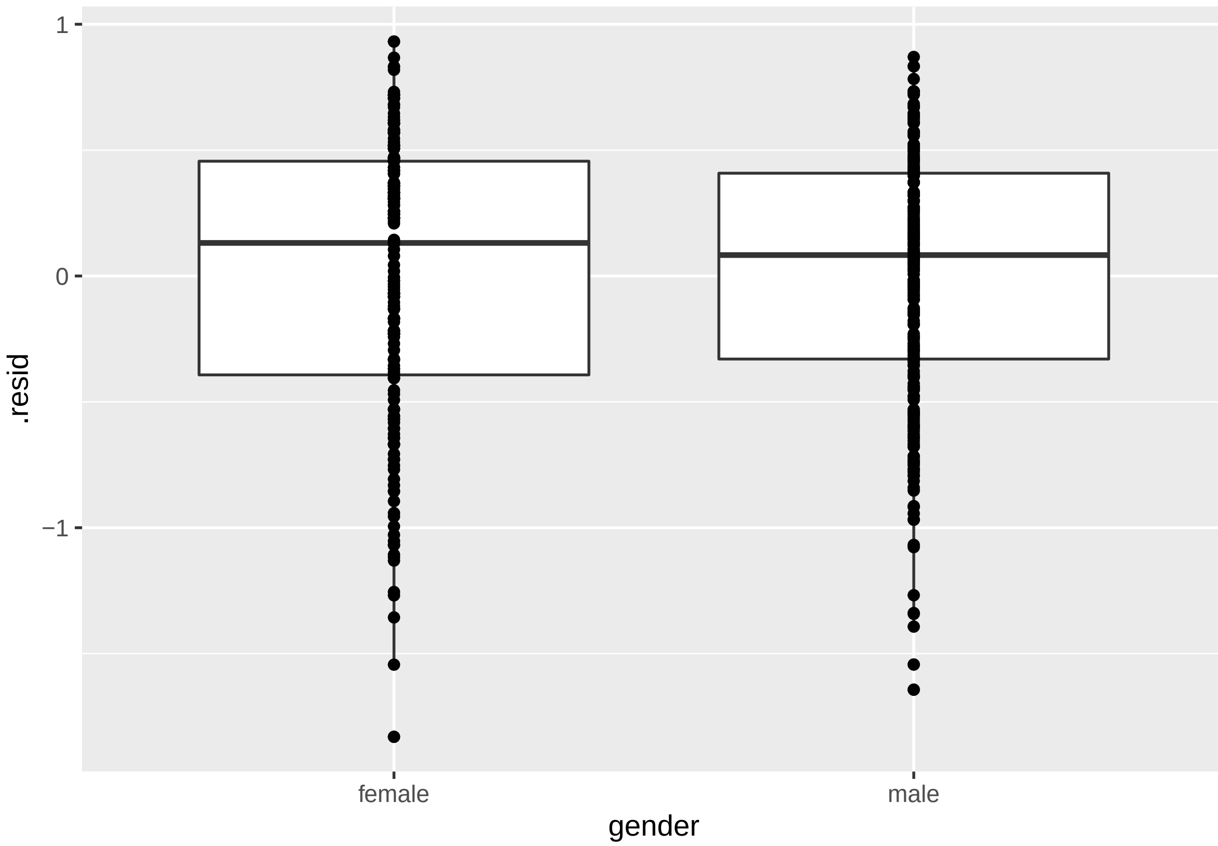
```
ggplot(data = m_bty_gen, aes(x = bty_avg , y = .resid)) +
  geom_point()
```




```
Assumption 3:  I cannot definitively say that the residuals are independent because I can not
    figure out how to plot the variables in order of their data collection
```

```
Assumption 4:  From the obox plots and scatter plots below it appears that each varibale is
    linearly related to the outcome.
```

```
ggplot(data = m_bty_gen, aes(x = gender, y = .resid)) + geom_boxplot() + geom_point()
```



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`?

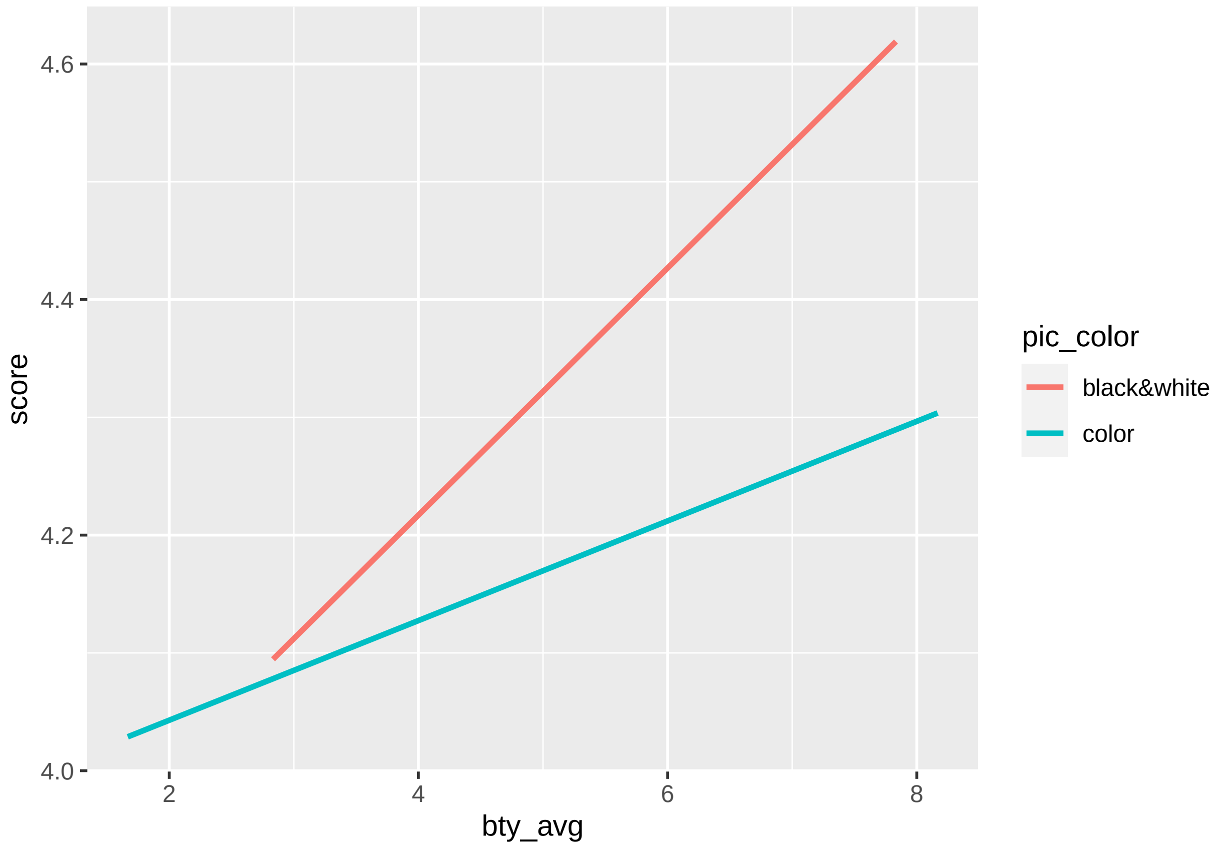
   Answer: I still do not believe that bty_avg is a significant predictor if an instructor's score because all four diagnotistic assumptions are not true as well as the adjusted R squared value did not increase sig- nificantly by adding gender into the equation.

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?

   Answer: $\hat{y} = \beta_0 + \beta_1$ x pic_color + β_2 x 0

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

```
##
## Call:
## lm(formula = score ~ bty_avg + pic_color, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8892 -0.3690  0.1293  0.4023  0.9125
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.06318    0.10908  37.249  < 2e-16 ***
## bty_avg          0.05548    0.01691   3.282  0.00111 **
## pic_colorcolor -0.16059    0.06892  -2.330  0.02022 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5323 on 460 degrees of freedom
```

```
## Multiple R-squared:  0.04628,    Adjusted R-squared:  0.04213
## F-statistic: 11.16 on 2 and 460 DF,  p-value: 1.848e-05
```

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`.

    Answwer: R appears to consolidate categorical values that have a very similar meaning or cannotation. In this case, teaching and tenure track are not associated with tenure.

```r
m_bty_rank <- lm(score ~ bty_avg + rank, data = evals)
summary(m_bty_rank)
```

```
##
## Call:
## lm(formula = score ~ bty_avg + rank, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8713 -0.3642  0.1489  0.4103  0.9525
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3.98155    0.09078  43.860  < 2e-16 ***
## bty_avg           0.06783    0.01655   4.098 4.92e-05 ***
## ranktenure track -0.16070    0.07395  -2.173   0.0303 *
## ranktenured      -0.12623    0.06266  -2.014   0.0445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5328 on 459 degrees of freedom
## Multiple R-squared:  0.04652,    Adjusted R-squared:  0.04029
## F-statistic: 7.465 on 3 and 459 DF,  p-value: 6.88e-05
```

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.

Answer: My choice would be cls_level. Because I would think that upper level students would be a little harder on there instructors. I think I believe would consider class level in my grading an instructor.

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.

    Answer: Removing class level did not have much of an affect on the over all instructors score.

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

```
##
## Call:
## lm(formula = score ~ rank + gender + ethnicity + language + age +
##     cls_perc_eval + cls_students + cls_profs + cls_credits +
##     bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.76617 -0.31818  0.09325  0.35169  0.93485
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.0938164  0.2905587  14.089  < 2e-16 ***
## ranktenure track      -0.1419977  0.0819039  -1.734 0.083656 .
## ranktenured           -0.0895639  0.0659238  -1.359 0.174957
## gendermale             0.2046338  0.0514798   3.975  8.2e-05 ***
## ethnicitynot minority  0.1383380  0.0773580   1.788 0.074405 .
## languagenon-english   -0.2109231  0.1099296  -1.919 0.055655 .
## age                   -0.0087375  0.0031258  -2.795 0.005407 **
## cls_perc_eval          0.0053940  0.0015382   3.507 0.000499 ***
## cls_students           0.0003430  0.0003622   0.947 0.344081
## cls_profssingle       -0.0150776  0.0519931  -0.290 0.771956
## cls_creditsone credit  0.4692494  0.1116766   4.202  3.2e-05 ***
## bty_avg                0.0412068  0.0174728   2.358 0.018785 *
## pic_outfitnot formal  -0.1216791  0.0733912  -1.658 0.098026 .
## pic_colorcolor        -0.1955247  0.0684550  -2.856 0.004486 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.498 on 449 degrees of freedom
## Multiple R-squared:  0.1851, Adjusted R-squared:  0.1615
## F-statistic: 7.845 on 13 and 449 DF,  p-value: 3.699e-14
```

13. Interpret the coefficient associated with the ethnicity variable.

    Answer: Although it is one of the lower p-values which is a good sign, I am surprised that it is not lower especially with number of minority students that attend UT, but, then again, I'm in Texas.

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?

    Answwer: Dropping the highest p-value had little affect on the adjusted R squared value. This tells me that the cls_profs is collinear to many of the other explanatory variables.

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

```
##
## Call:
```

```
## lm(formula = score ~ rank + gender + ethnicity + language + age +
##     cls_perc_eval + cls_students + cls_credits + bty_avg + pic_outfit +
##     pic_color, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7761 -0.3187  0.0875  0.3547  0.9367
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           4.0856255  0.2888881  14.143  < 2e-16 ***
## ranktenure track     -0.1420696  0.0818201  -1.736 0.083184 .
## ranktenured          -0.0895940  0.0658566  -1.360 0.174372
## gendermale            0.2037722  0.0513416   3.969 8.40e-05 ***
## ethnicitynot minority  0.1424342  0.0759800   1.875 0.061491 .
## languagenon-english  -0.2093185  0.1096785  -1.908 0.056966 .
## age                  -0.0087287  0.0031224  -2.795 0.005404 **
## cls_perc_eval         0.0053545  0.0015306   3.498 0.000515 ***
## cls_students          0.0003573  0.0003585   0.997 0.319451
## cls_creditsone credit  0.4733728  0.1106549   4.278 2.31e-05 ***
## bty_avg               0.0410340  0.0174449   2.352 0.019092 *
## pic_outfitnot formal  -0.1172152  0.0716857  -1.635 0.102722
## pic_colorcolor       -0.1973196  0.0681052  -2.897 0.003948 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4975 on 450 degrees of freedom
## Multiple R-squared:  0.185,  Adjusted R-squared:  0.1632
## F-statistic:  8.51 on 12 and 450 DF,  p-value: 1.275e-14
```

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.

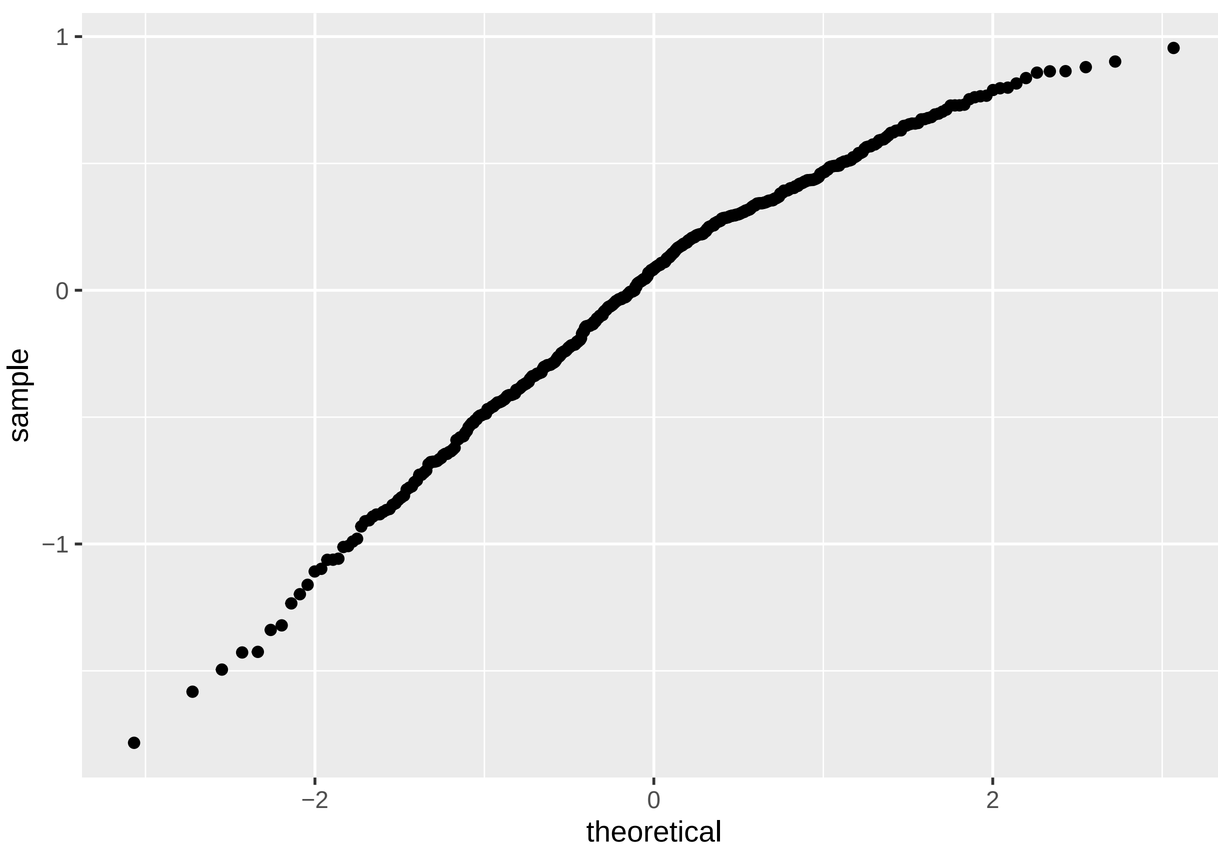
```
m_best_model <- lm(score ~ rank + gender + ethnicity + language + age
             + cls_perc_eval + cls_students + cls_level + cls_credits + bty_avg
             + pic_outfit + pic_color, data = evals)
summary(m_best_model)
```

```
##
## 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 .
```

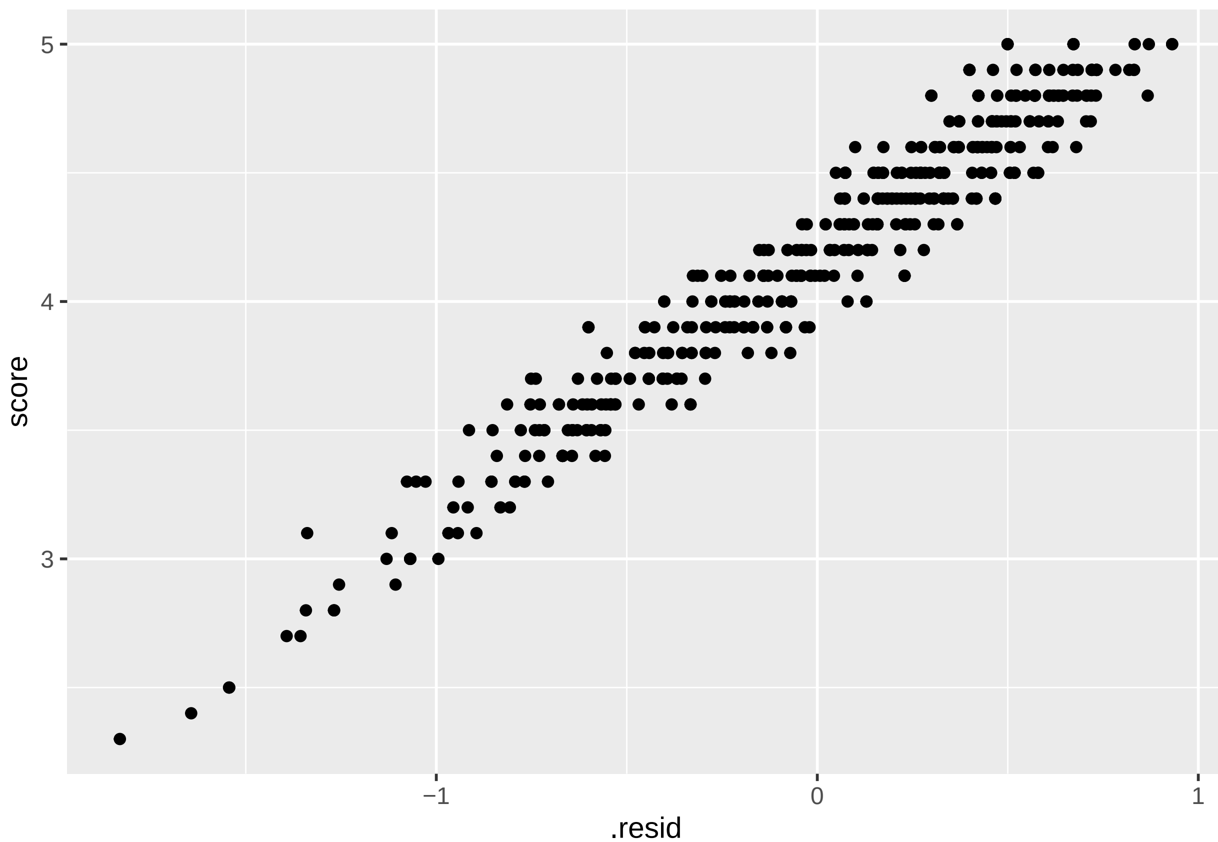
```
## ranktenured           -0.0973829  0.0662614  -1.470 0.142349
## gendermale              0.2101231  0.0516873   4.065 5.66e-05 ***
## ethnicitynot minority  0.1274458  0.0772887   1.649 0.099856 .
## languagenon-english   -0.2282894  0.1111305  -2.054 0.040530 *
## age                   -0.0089992  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
```

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

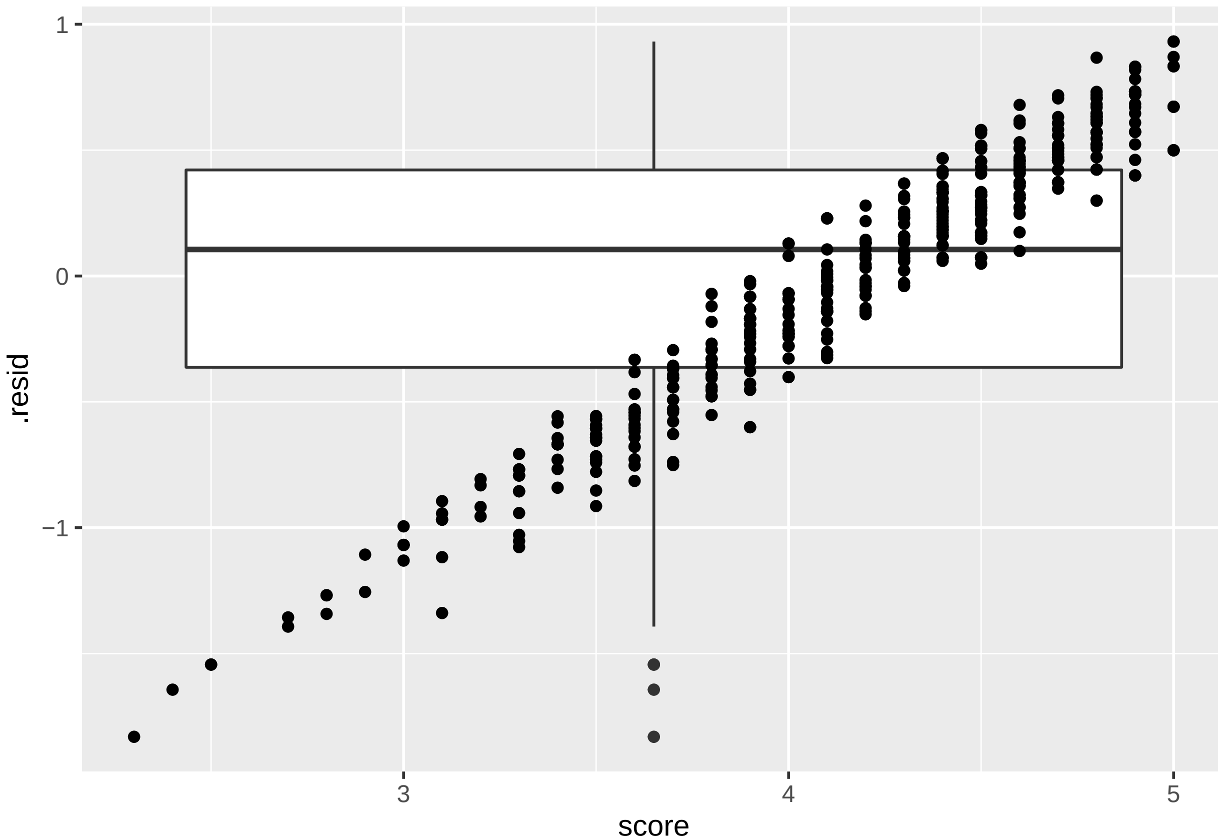
```
ggplot(data = m_best_model, aes(sample = .resid)) +
  stat_qq()
```



```r
ggplot(data = m_bty_gen, aes(x = .resid , y = score)) +
  geom_point()
```




```r
ggplot(data = m_bty_gen, aes(x = score, y = .resid)) + geom_boxplot() + geom_point()
```



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 I am interpreting this correctly, it seems that some of the variables on the evaluation do not have a significant impact on the score given to a professor. That being said I would probably take out many of the variables having to do with appearance and focus more on course material, presentation, average grade in the class, etc.

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: based on my final model the characteristics most important to getting a positive score are rank (tenure or not), gender, language, age, number of students taking the evaluation, class size, class level, number of class credit, physical appearance, attire and the type of picutre posted (color or black and white)

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

References
