DATA 606 Lab 9

Joao De Oliveira

2025-11-29

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

```r
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, ~
```

```r
?evals
```

Exercise 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 1

This is an observational study. Professors weren't randomly assigned to different groups based on their "beauty", which would be a requirement for an experiment. I don't think the question is correctly phrased since we can't check causality without addressing other variables besides beauty. Also, the fact that beauty and course evaluation are correlated doesn't mean one causes the other. I would phrase the question like "After controlling other factors, can we say that there is an association between professors' beauty and their evaluation scores?"

Exercise 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?

```
ggplot(evals, aes(x = score)) +
  geom_histogram(binwidth = 0.2, color='white') +
  labs(title = "Distribution of Course Evaluation Scores",
       x = "Score",
       y = "Count")
```

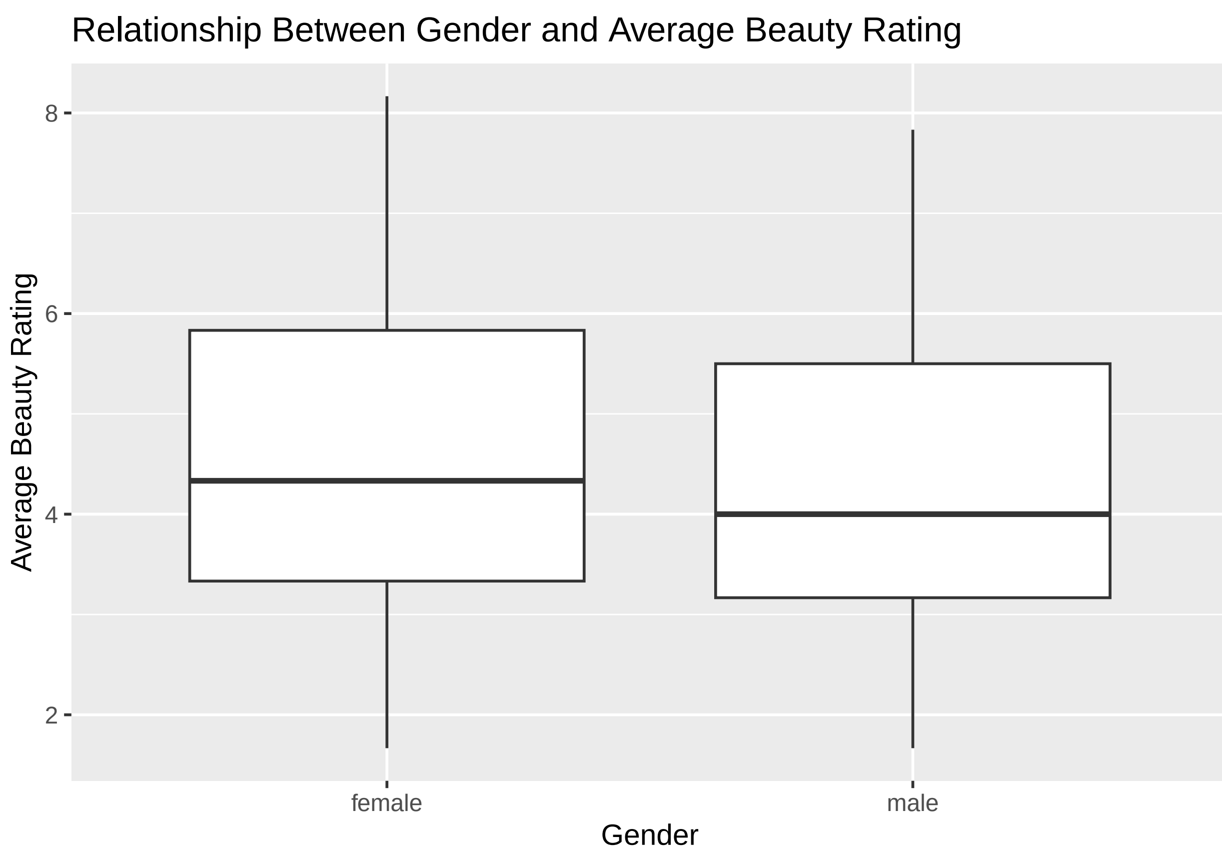
Answer 2

The distribution of scores is left skewed, which means that most of the data is concentrated on the right side (long tail to the left). This means that most students graded professors/courses with high scores, which didn't really surprised me. I think most students tend to give high scores in courses evaluation unless they have something personal against the professor.

Exercise 3

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

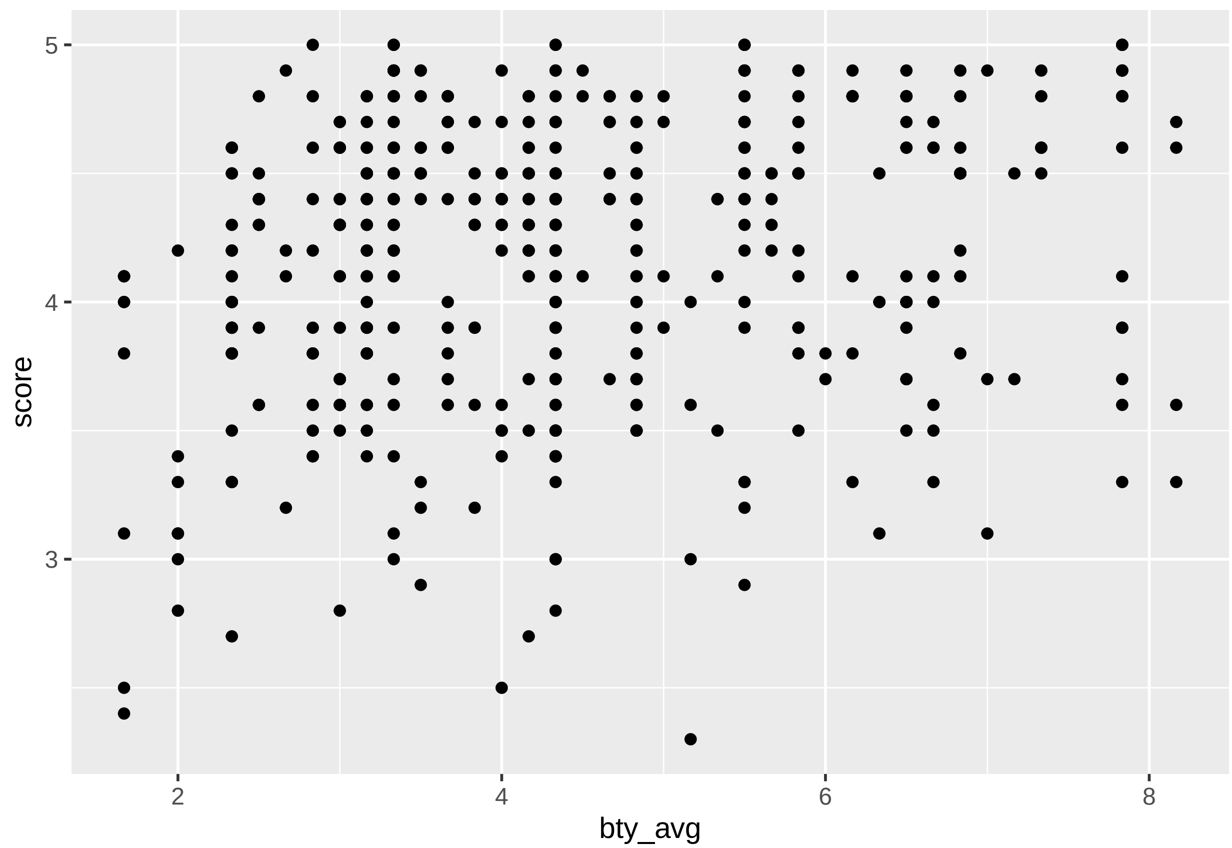
```
ggplot(evals, aes(x = gender, y = bty_avg)) +
  geom_boxplot() +
  labs(title = "Relationship Between Gender and Average Beauty Rating",
       x = "Gender",
       y = "Average Beauty Rating")
```



Answer 3

I decided to check whether there is a relationship between gender and average beauty rating. As we can see in the boxplot, the distribution of female professors' beauty score has higher 1st and 3rd quartiles and median than male professors' distribution. The highest scores are also for female professors. So, there is a relationship between gender and average scores.

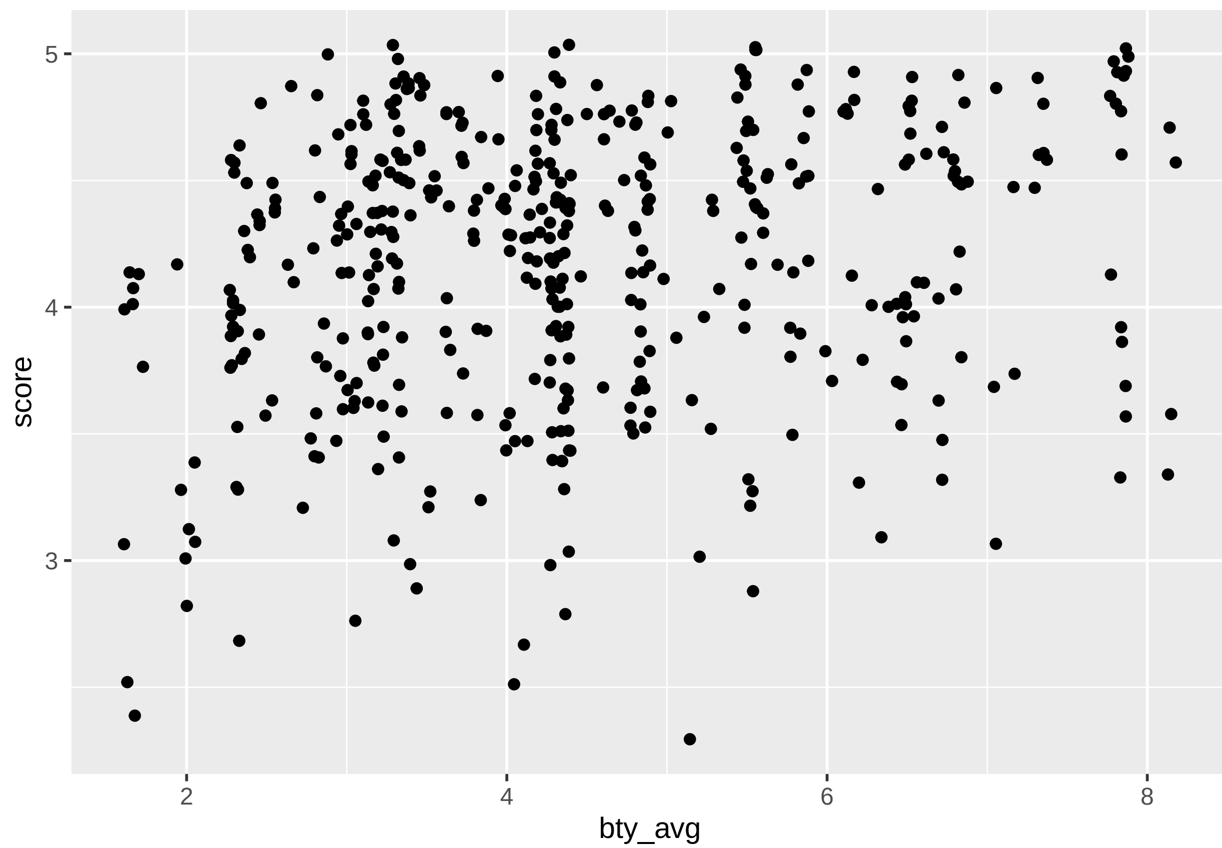
```
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_point()
```



Exercise 4

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

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



Answer 4

The initial scatterplot was misleading because it looked like there were fewer data points than reality. With geom_jitter points are not overlapping, so we can actually see all the observations.

Exercise 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
```

Answer 5

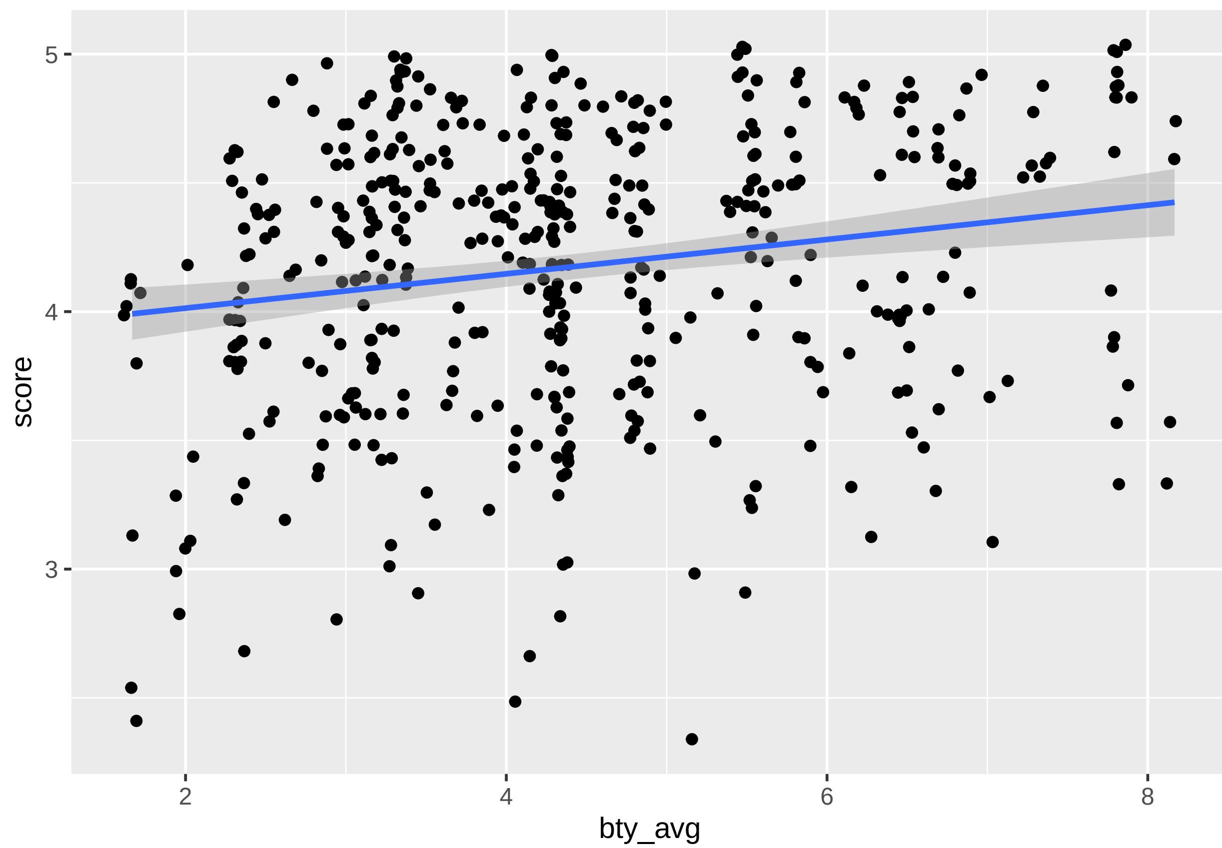
equation: score_hat = 3.88034 + 0.06664 x bty_avg

Since the slope is 0.0664, the model predicts that for each point increase in average beauty rating, the score increases 0.0664. Average beauty rating is a statistically significant predictor because the p-value for bty_avg is 5.083e-05, which is much smaller than 0.05 (common significance level). Since the increase in score per point increase in average beauty rating is pretty small (0.0664), then despite bty_avg being statistically significant, it is not practically significant. The impact is small.

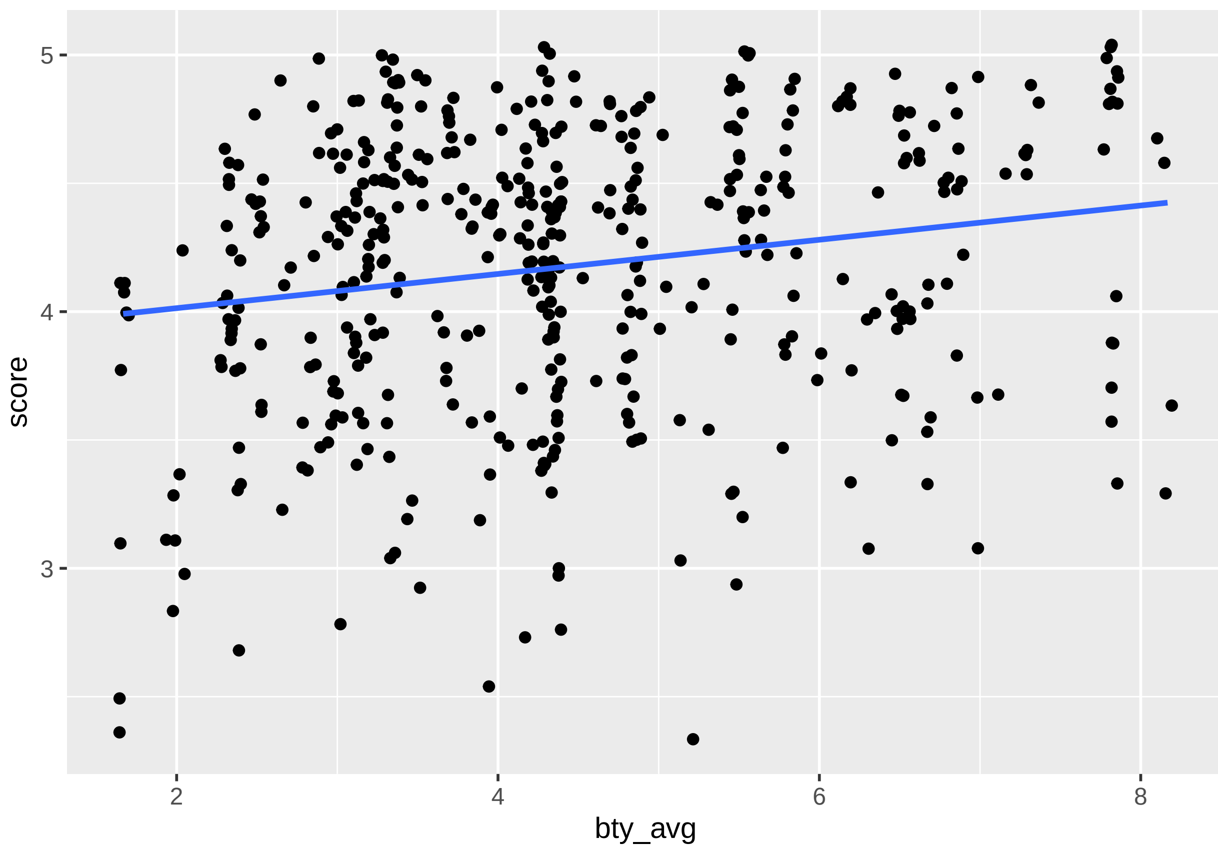
Exercise 6

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).

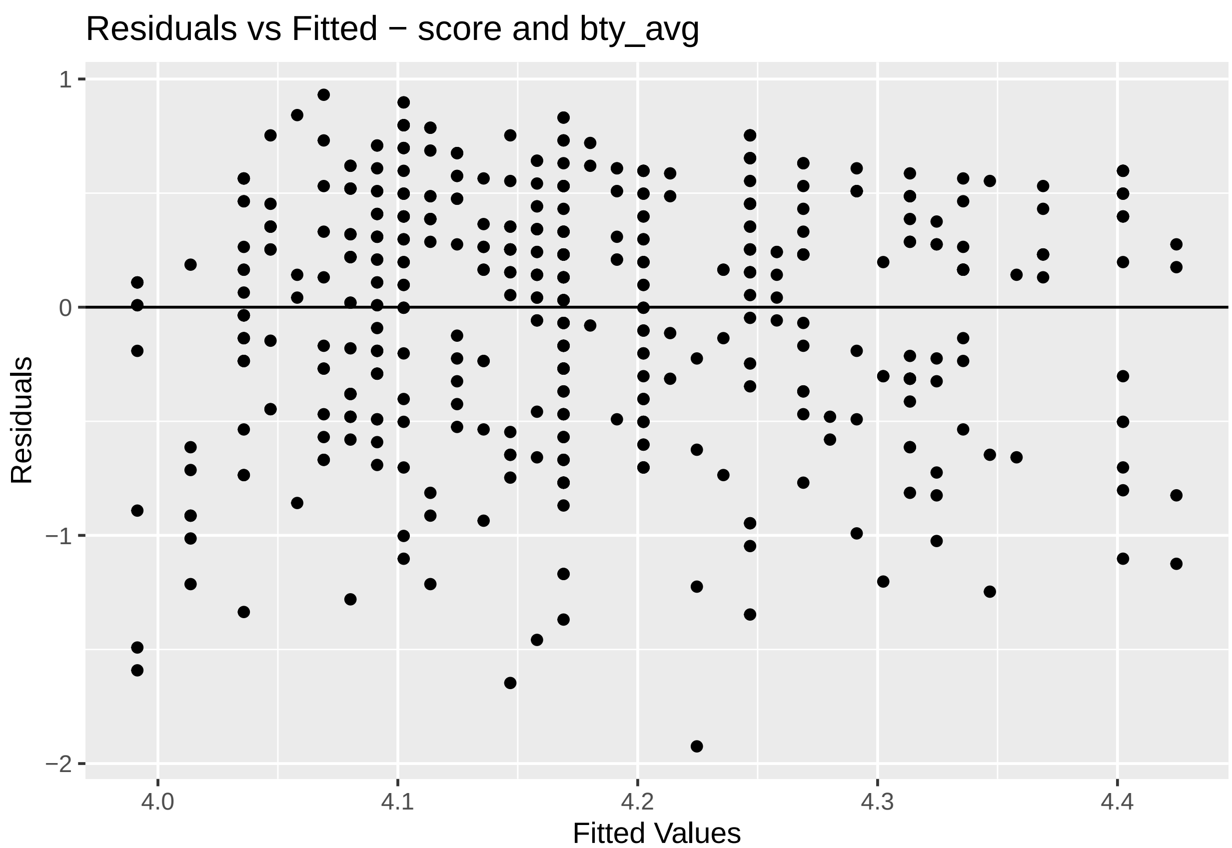
```r
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_jitter() +
  geom_smooth(method = "lm")
```



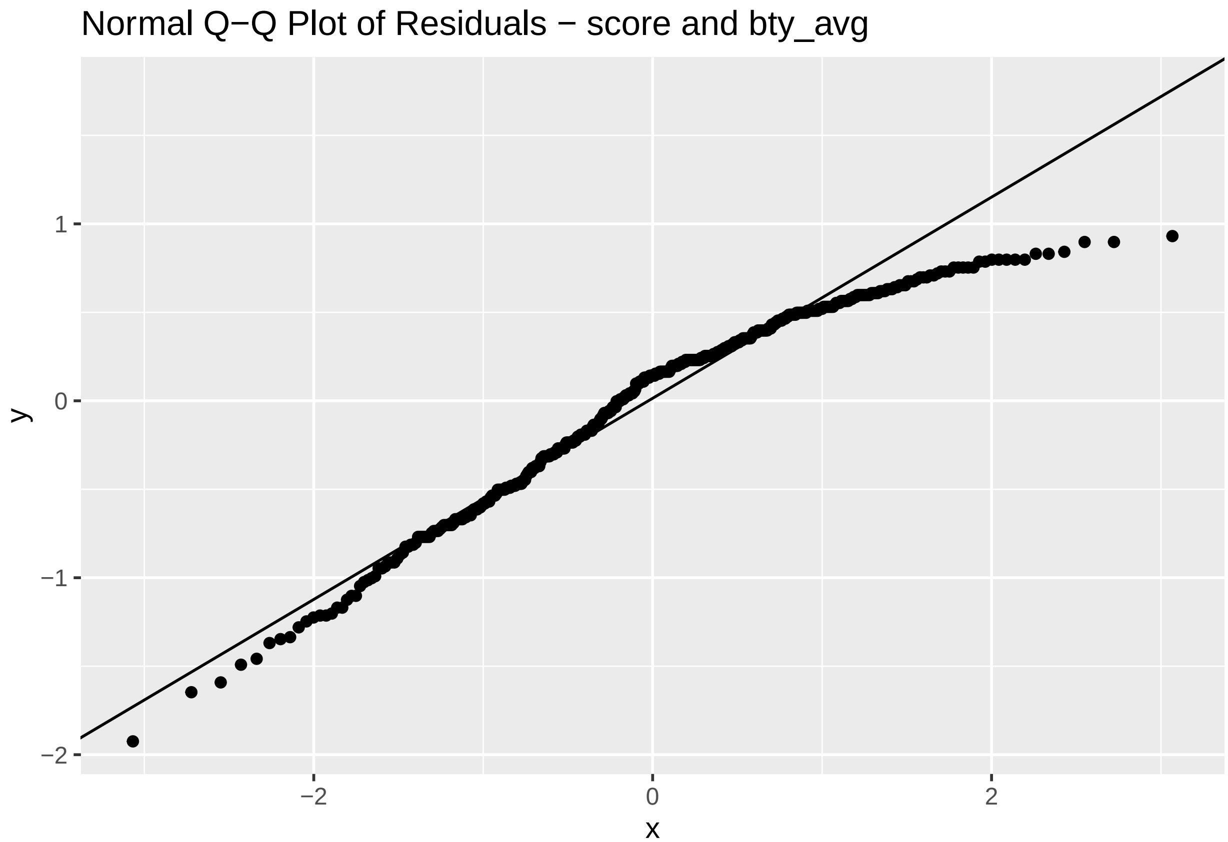

```r
ggplot(data = evals, aes(x = bty_avg, y = score)) +
  geom_jitter() +
  geom_smooth(method = "lm", se = FALSE)
```




```r
# residuals and fitted vals
evals <- evals %>%
  mutate(
  bty_fit = fitted(m_bty),
  bty_res = resid(m_bty)
  )

ggplot(evals, aes(x = bty_fit, y = bty_res)) +
  geom_hline(yintercept = 0) +
  geom_point() +
  labs(
    title = "Residuals vs Fitted - score and bty_avg",
    x = "Fitted Values",
    y = "Residuals"
  )
```



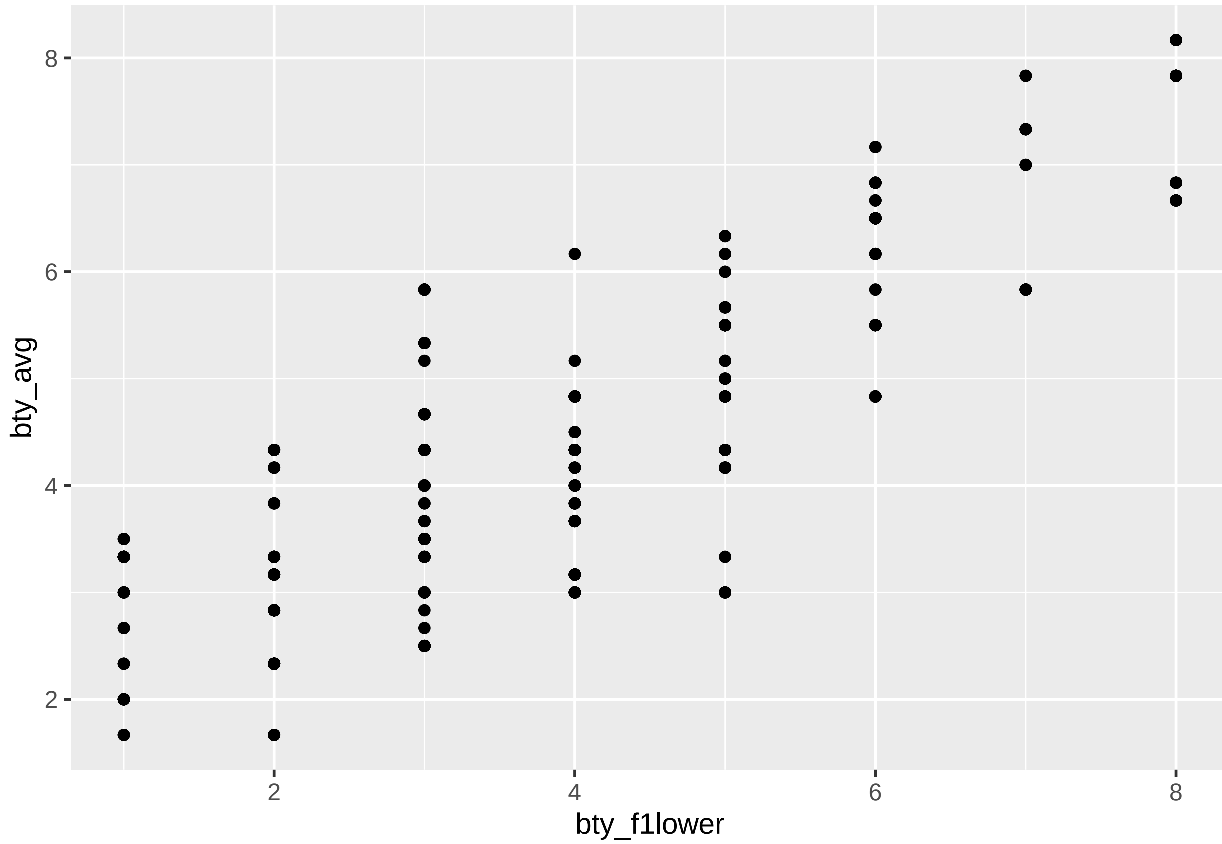

```r
# normal Q-Q plot
ggplot(evals, aes(sample = bty_res)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = "Normal Q-Q Plot of Residuals - score and bty_avg")
```



Answer 6

Linear The residuals appear somewhat randomly scattered around the y=0 horizontal line, which suggests that is reasonable to assume a linear relationship. **Q-Q plot** we can see that most points fall on the straight line, however, there are deviations on the tails (mainly upper right) indicating that the residuals are not perfectly normal, but not enough to dismiss the model. **Constant variance** The spread of residuals seems roughly constant across the range of fitted values, which supports the constant-variance assumption. **Independence** Since some professors appear multiple times, that can be an issue regarding independence. At the same time, the scores are from various courses.

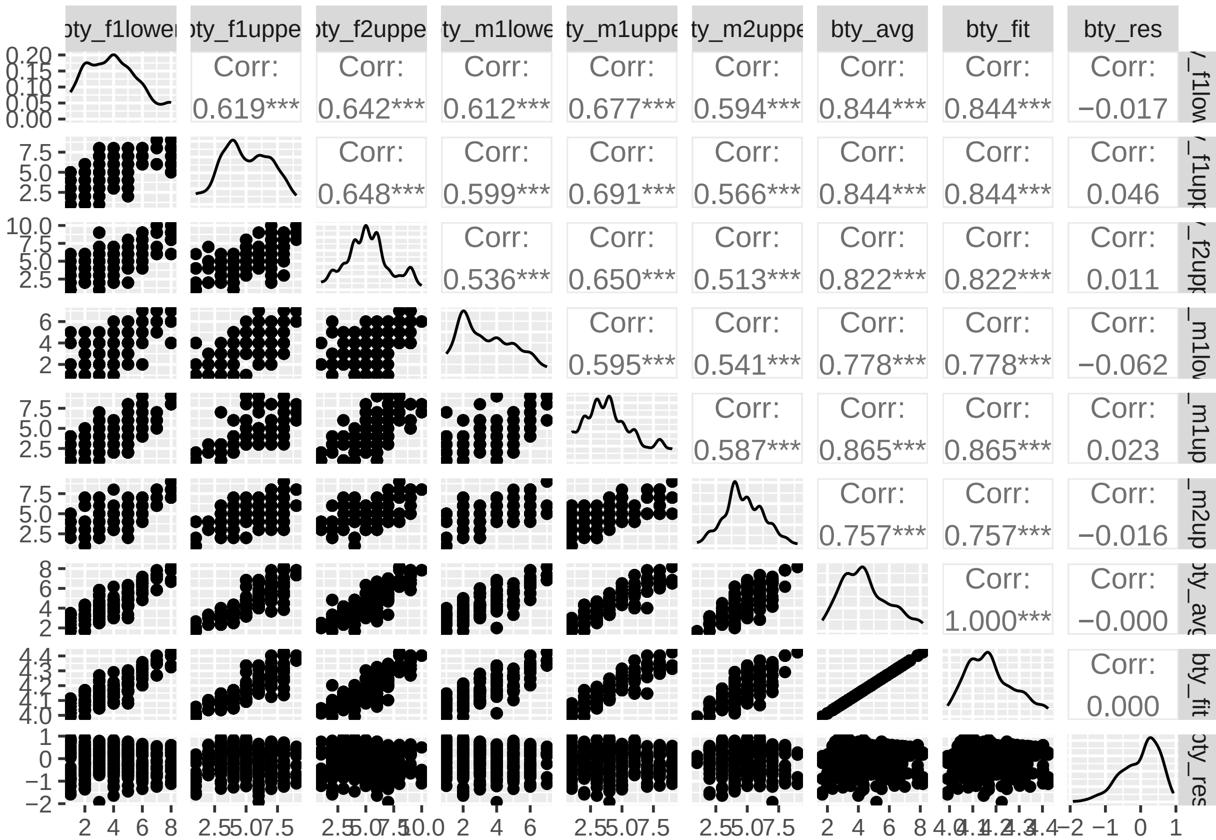
```
ggplot(data = evals, aes(x = bty_f1lower, y = bty_avg)) +
  geom_point()
```




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

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

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



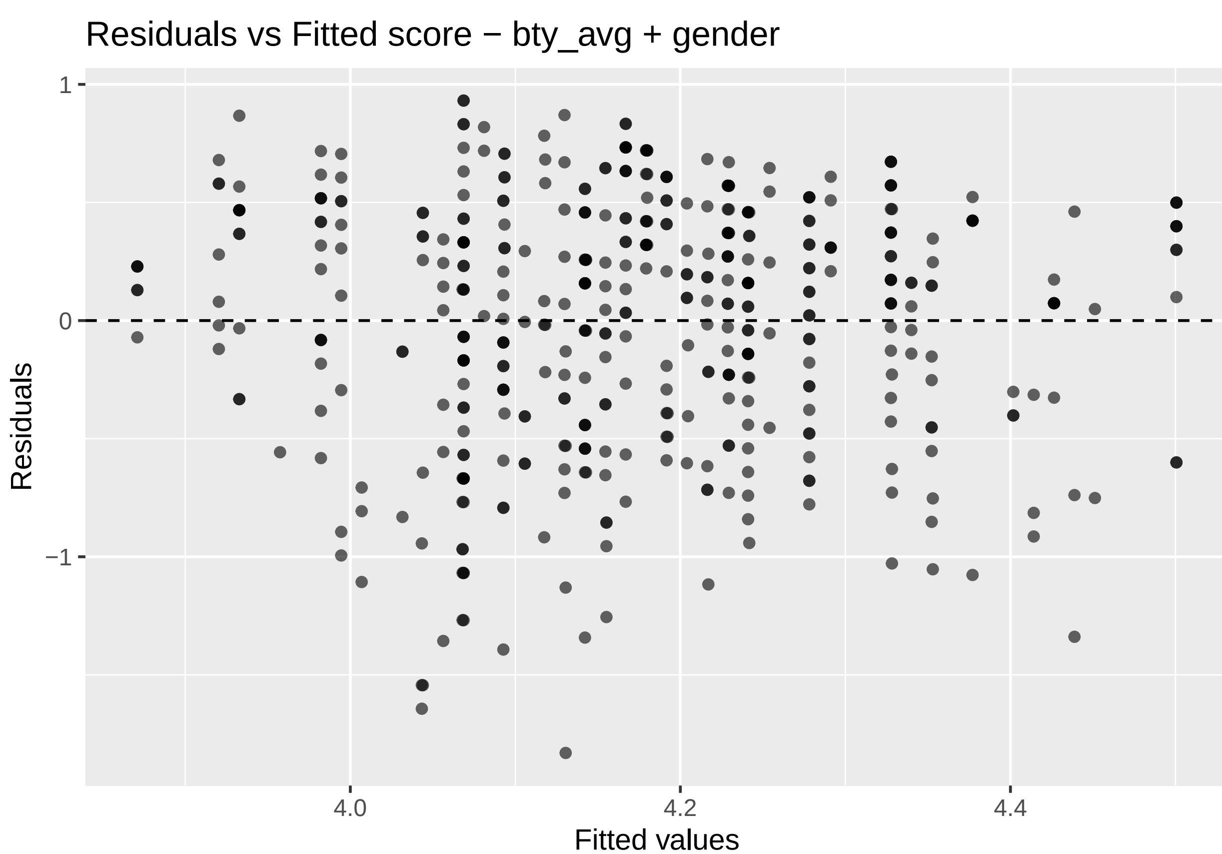

```r
m_bty_gen <- lm(score ~ bty_avg + gender, data = evals)
summary(m_bty_gen)
```

```
##
## 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
```

Exercise 7

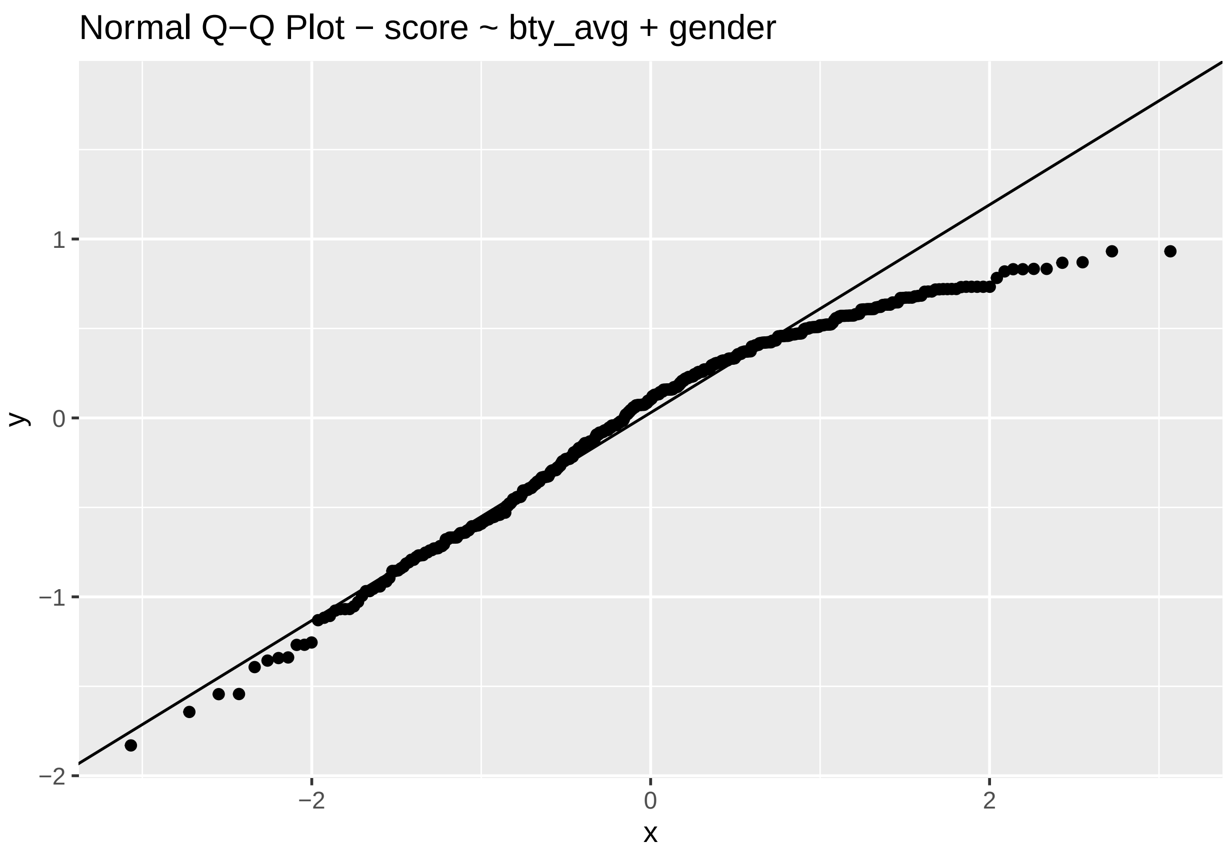
P-values and parameter estimates should only be trusted if the conditions for the regression are reasonable. Verify that the conditions for this model are reasonable using diagnostic plots.

```
evals <- evals %>%
mutate(
bty_gen_fit = fitted(m_bty_gen),
bty_gen_res = resid(m_bty_gen)
)

# residuals and fitted vals

ggplot(evals, aes(x = bty_gen_fit, y = bty_gen_res)) +
geom_hline(yintercept = 0, linetype = "dashed") +
geom_point(alpha = 0.6) +
labs(
title = "Residuals vs Fitted score - bty_avg + gender",
x = "Fitted values",
y = "Residuals"
)
```




```
# normal Q-Q plot

ggplot(evals, aes(sample = bty_gen_res)) +
```

```r
stat_qq() +
stat_qq_line() +
labs(
title = "Normal Q-Q Plot - score ~ bty_avg + gender"
)
```



Answer 7

Linear The residuals appear somewhat randomly scattered around the y=0 horizontal line, which suggests that is reasonable to assume a linear relationship. **Q-Q plot** SQ–Q plot shows approximate normality with mild tail deviations. **Constant variance** The spread of residuals seems roughly constant across the range of fitted values, which supports the constant-variance assumption. **Independence** Since some professors appear multiple times, that can be an issue regarding independence. At the same time, the scores are from various courses.

So, the conditions for this multiple linear regression appear reasonably satisfied.

Exercise 8

Is `bty_avg` still a significant predictor of `score`? Has the additionof `gender` to the model changed the parameter estimate for `bty_avg`?

Answer 8

Yes. The p-value for bty_avg is very small (on the order of 10^-7), so bty_avg is still a statistically significant predictor of score even after controlling for gender. The estimate for bty_avg changes only slightly compared to the simple model (0.07416 ~ 0.06664), meaning the relationship between beauty and score is not fully explained by gender. The coefficient gendermale is positive and statistically significant.

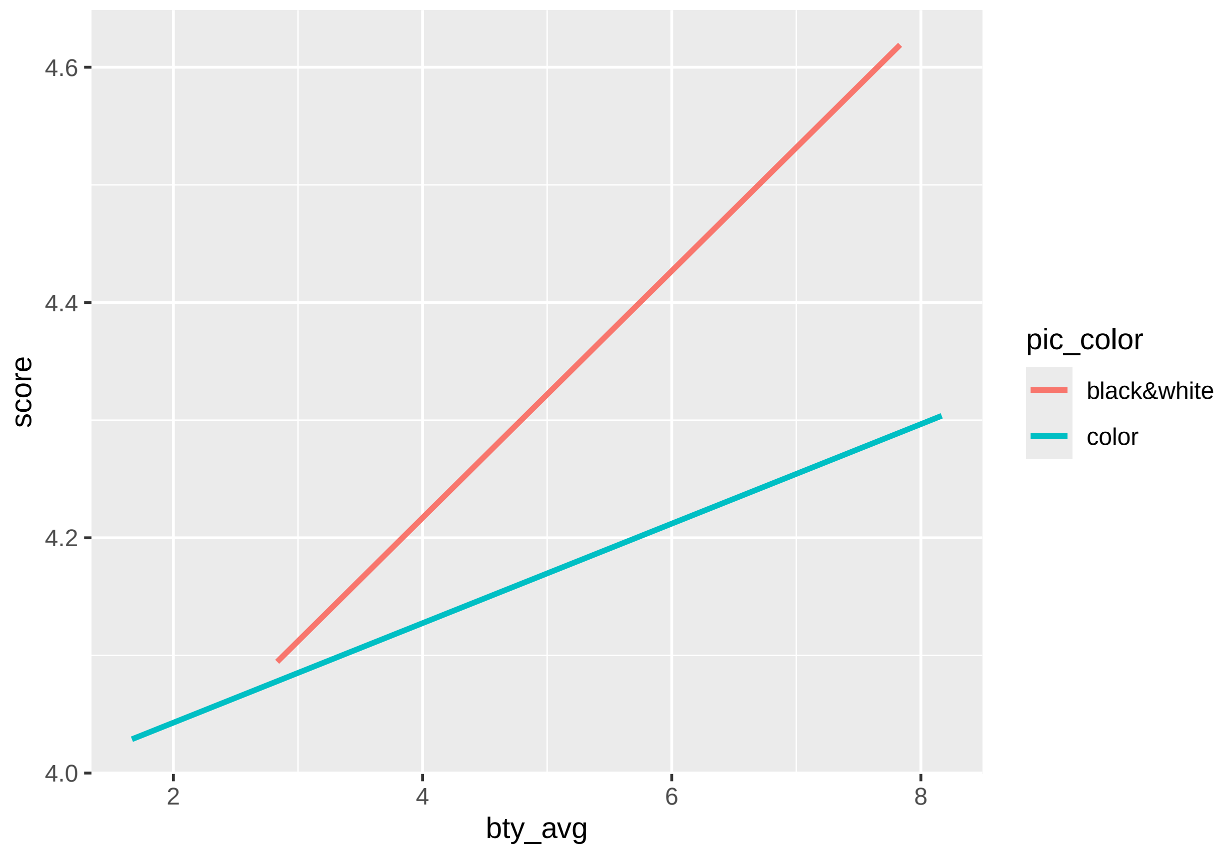
Exercise 9

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?

```
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
```

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



Answer 9

score= $^{0+}$ 1×bty_avg+ ˆ2×pic_colorcolor

For black and white, the last term becomes zero, so: score= $^{0+}$ 1×bty_avg For color pictures (1), since pic_colorcolor =1, then the last term is just ˆ2, so the equation for color pictures is score= $^{0+}$ 1×bty_avg+ ˆ2

Since ˆ2 is negative, for two professors with the same beauty rating, the professor with a black-and-white picture tends to have a slightly higher predicted score than the one with a color picture.

Exercise 10

Create a new model called m_bty_rank with gender removed and rank added in. How does R appear to handle categorical variables that have more than two levels? Note that the rank variable has three levels: teaching, tenure track, tenured.

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

```
##
## Call:
## lm(formula = score ~ bty_avg + rank, data = evals)
##
## Residuals:
```

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

Answer 10

R handles this by creating dummy variables: ranktenure track (1 if the professor is tenure track, 0 otherwise) and ranktenured (1 if the professor is tenured, 0 otherwise). The the reference level (rankteaching) is implicitly coded as 0 for both dummy variables. By default, R uses the first category in alphabetical order as the baseline. The coefficients for ranktenure track and ranktenured show how the expected score differs from the reference group, after controlling for beauty.

Exercise 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 11

I would expect variables like number of professors in the course (cls_profs) or instructor outfit in the picture (pic_outfit) to have little or no association with evaluation scores, once other variables are controlled for. For example, whether the course has a single professor or multiple professors doesn't obviously seem tied to how students rate teaching quality (different than quantity).

Exercise 12

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

```r
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
```

Answer 12

Looking at the model output, the variable with the largest p-value is cls_profssingle, with a p-value of 0.77806, confirming that this predictor is not contributing much to explaining variation in scores.

Exercise 13

Interpret the coefficient associated with the ethnicity variable.

Answer 13

The ethnicity variable has two levels: minority (reference level) and not minority. The coefficient in the summary is ethnicitynot minority=0.1234929. This mean that holding all other variables in the model constant, professors who are not minority are expected to have a course evaluation score that is 0.1234929 points higher on average compared to professors who are a minority.

Exercise 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?

```r
m_drop1 <- 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_drop1)
```

```
##
## 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
```

Answer 14

When we compare summary(m_full) and summary(m_drop1) we can see that the coefficients and p-values for the remaining variables barely change. The same predictors remain significant, which suggests that cls_profs was not strongly related to either score or the other explanatory variables. As a result, removing it does not materially affect the estimates of the remaining coefficients.

Exercise 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.

```r
m_final <- lm(score ~ gender + language + age + cls_perc_eval + cls_credits
+ bty_avg + pic_color, data = evals)

summary(m_final)
```

```
##
## Call:
## lm(formula = score ~ gender + language + age + cls_perc_eval +
##     cls_credits + bty_avg + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.81919 -0.32035  0.09272  0.38526  0.88213
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.967255   0.215824  18.382  < 2e-16 ***
## gendermale             0.221457   0.049937   4.435 1.16e-05 ***
## languagenon-english   -0.281933   0.098341  -2.867  0.00434 **
## age                   -0.005877   0.002622  -2.241  0.02551 *
## cls_perc_eval          0.004295   0.001432   2.999  0.00286 **
## cls_creditsone credit  0.444392   0.100910   4.404 1.33e-05 ***
## bty_avg                0.048679   0.016974   2.868  0.00432 **
## pic_colorcolor        -0.216556   0.066625  -3.250  0.00124 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5014 on 455 degrees of freedom
## Multiple R-squared:  0.1631, Adjusted R-squared:  0.1502
## F-statistic: 12.67 on 7 and 455 DF,  p-value: 6.996e-15
```

Answer 15

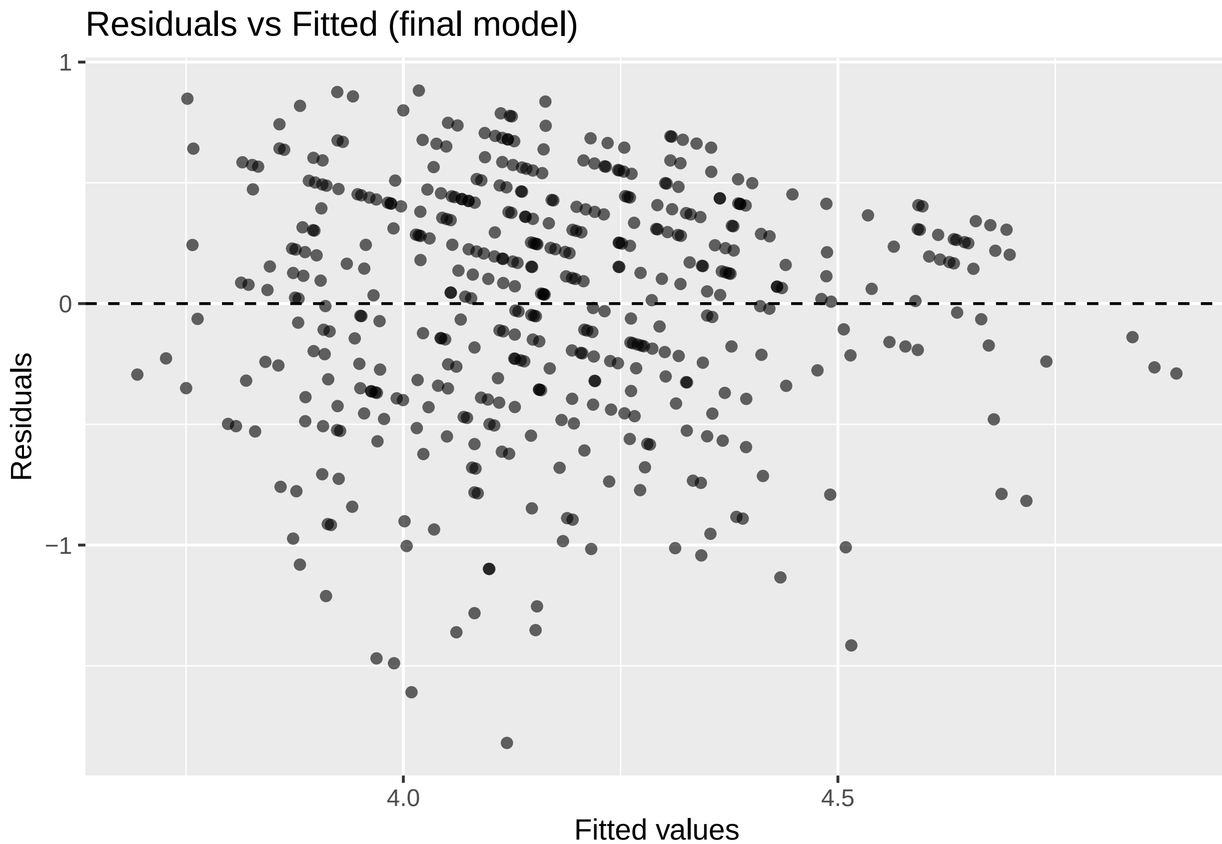
score = 3.967255 + 0.221457 x gendermale - 0.281933 x languagenon-english -0.005877 x age + 0.004295 x cls_perc_eval + 0.444392 x cls_creditsone credit + 0.048679 x bty_avg -0.216556 x pic_colorcolor

Exercise 16

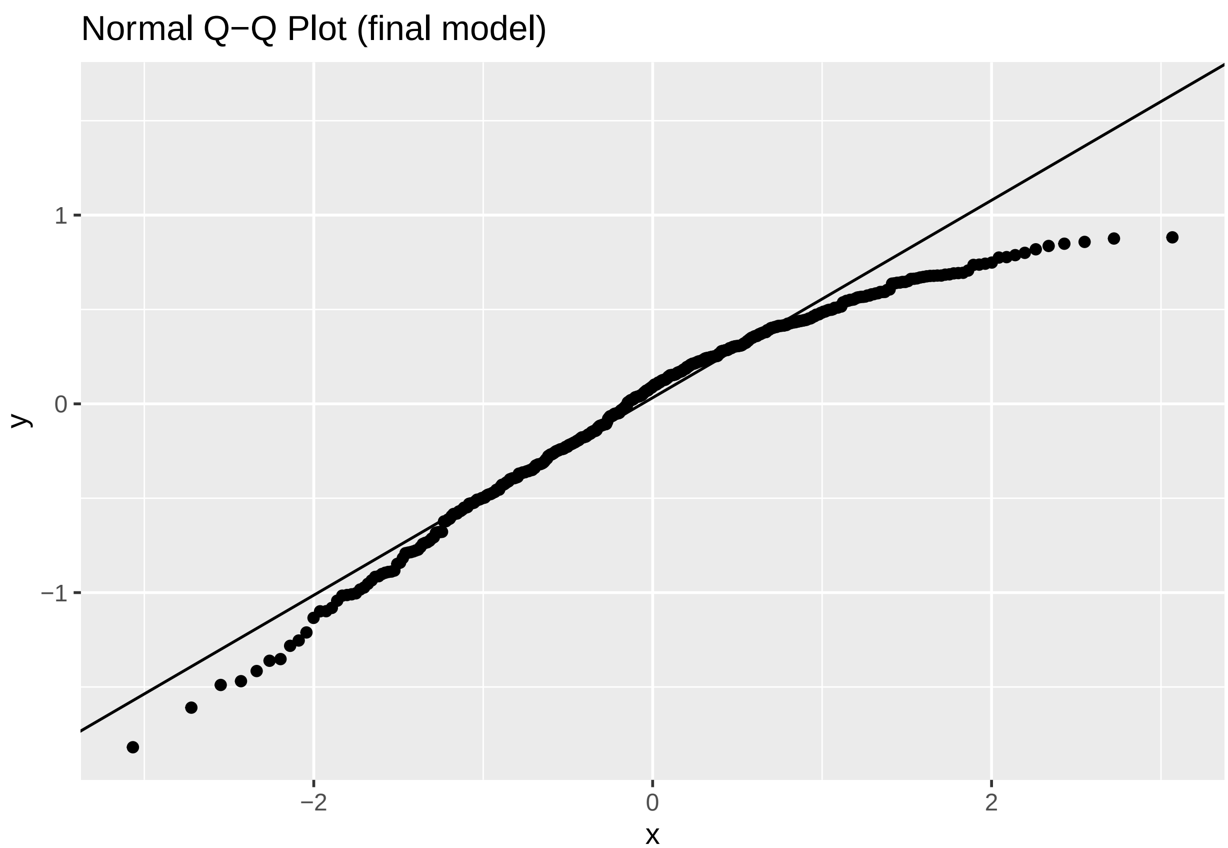
Verify that the conditions for this model are reasonable using diagnostic plots.

```r
evals <- evals %>%
mutate(
final_fit = fitted(m_final),
final_res = resid(m_final)
)

# residuals and fitted
ggplot(evals, aes(x = final_fit, y = final_res)) +
geom_hline(yintercept = 0, linetype = "dashed") +
geom_point(alpha = 0.6) +
labs(
```

```r
title = "Residuals vs Fitted (final model)",
x = "Fitted values",
y = "Residuals"
)
```




```r
# normal Q-Q plot
ggplot(evals, aes(sample = final_res)) +
stat_qq() +
stat_qq_line() +
labs(
title = "Normal Q-Q Plot (final model)"
)
```



Answer 16

Linear The residuals are generally scattered around the y=0 horizontal line, which suggests that is reasonable to assume a linear relationship. **Q-Q plot** SQ–Q plot shows approximate normality with some tail deviations. **Constant variance** The spread of residuals seems roughly constant across the range of fitted values, which supports the constant-variance assumption. **Independence** Same potential issue as before: multiple courses per professor may introduce clustering

Overall, the linear model assumptions look reasonably satisfied, with independence being the main possible limitation.

Exercise 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 17

Since each row represents a course and some professors teach more than one course, then there will be multiple rows for the same professor (one for each course taught). This directly impacts the independence of observations condition since evaluations from different courses taught by the same professor are likely not independent. For example, a professor's personal teaching style, beauty rating, and overall rank are constant across all their courses in the dataset. If a professor is generally rated highly, all their courses will be rated highly.Our linear regression treats each row as independent, so the standard errors for our coefficients may be underestimated, leading to p-values that are somewhat too small.

Exercise 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 18

Based on my final model, a professor and course with high evaluation score has a high beauty rating, is a young male, is not from a minority, has high percentage of students completing evaluations, teach multi-credit courses, and has a black-and-white picture.

Exercise 19

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

Answer 19

I wouldn't feel comfortable generalizing the conclusions to all professors at all universities. The data come from a single institution (UT Austin), which has its own culture, student body, and evaluation practices, so the relationships between beauty, gender, ethnicity, and scores may not hold elsewhere. The sample is also not a random sample of all professors, and important factors (like subject area or departmental culture) are not included. Finally, because multiple courses come from the same professor, the independence assumption is likely violated, which can make the standard errors and p-values too optimistic. Overall, the results are very informative for this context but they should be interpreted cautiously rather than as universal truths.