Lab 10 - Multiple Linear Regression

Sanaz Saadatifar

Date of lab session

Lab report

Load data

```
evals <- read_csv("https://dyurovsky.github.io/85309/data/lab10/evals.csv")
```

```
## Rows: 463 Columns: 21
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (9): rank, ethnicity, gender, language, cls_level, cls_profs, cls_credi...
## dbl (12): score, age, cls_perc_eval, cls_did_eval, cls_students, bty_f1lower...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

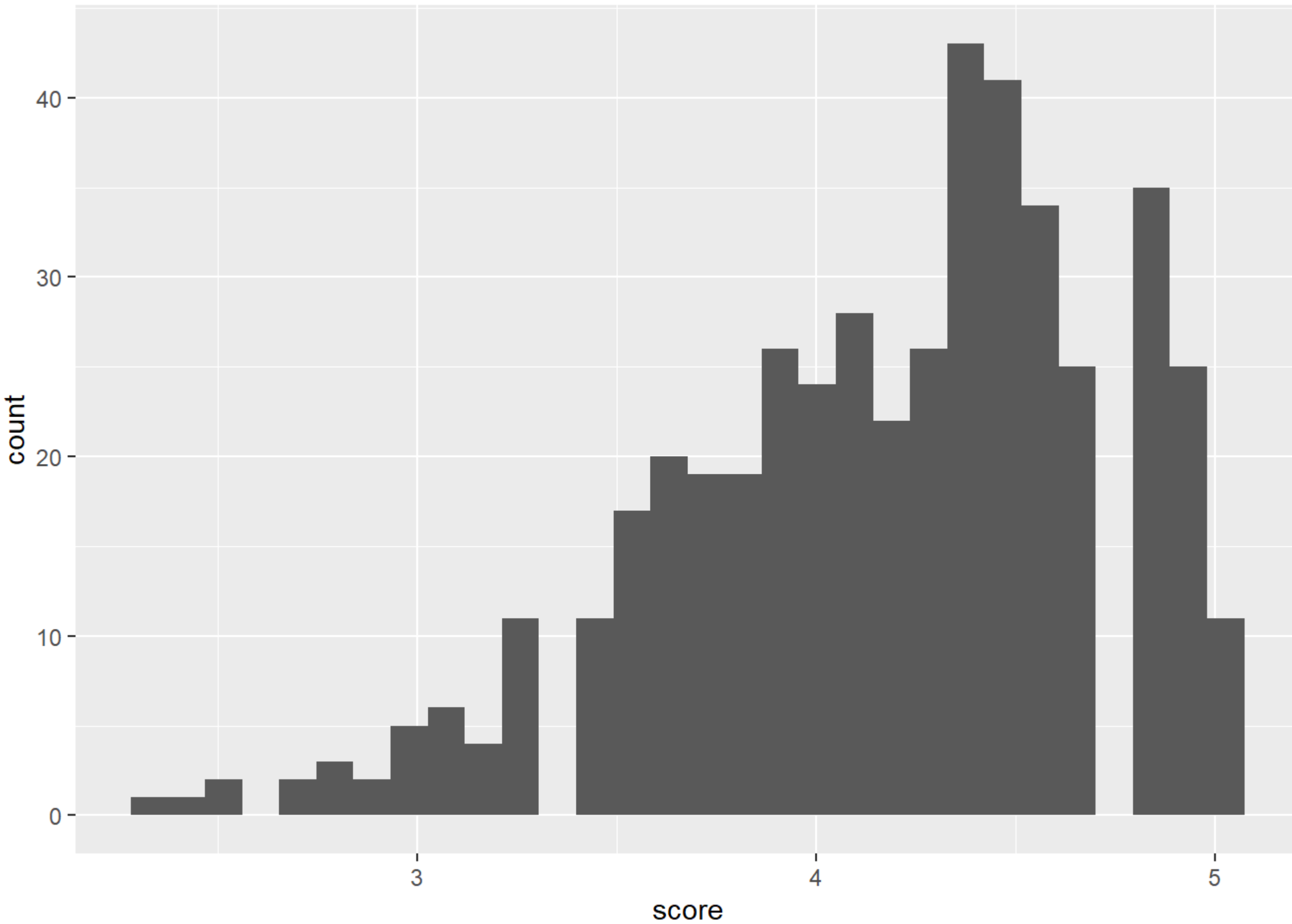
Exercise 1:

this is an observational study because beauty was not randomly assigned to professors. As a result, we can't make a causal claim from this data. A better way of phrasing this research question would be to say something like "is there an association between beauty and course ratings?"

Exercise 2:

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

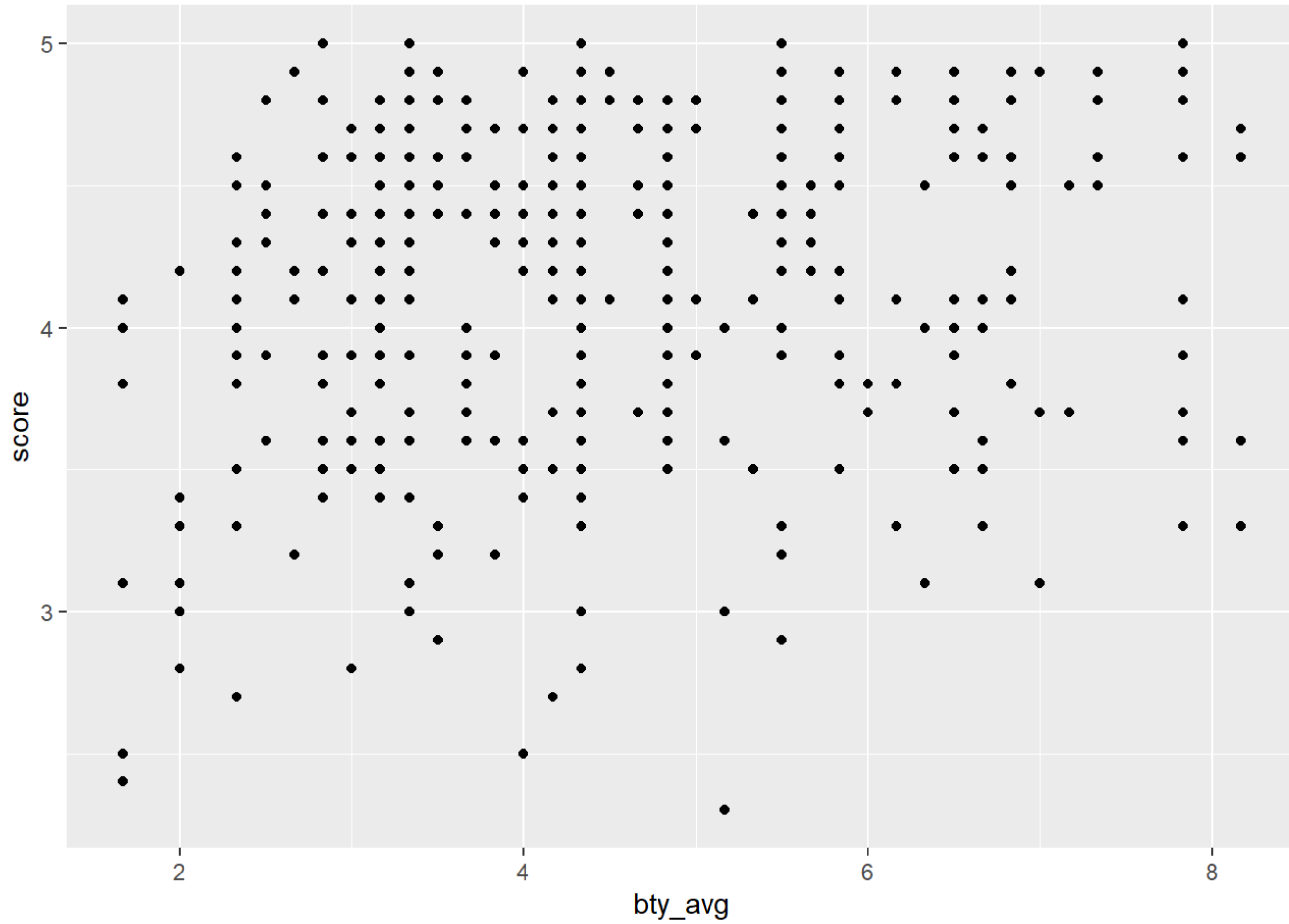
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



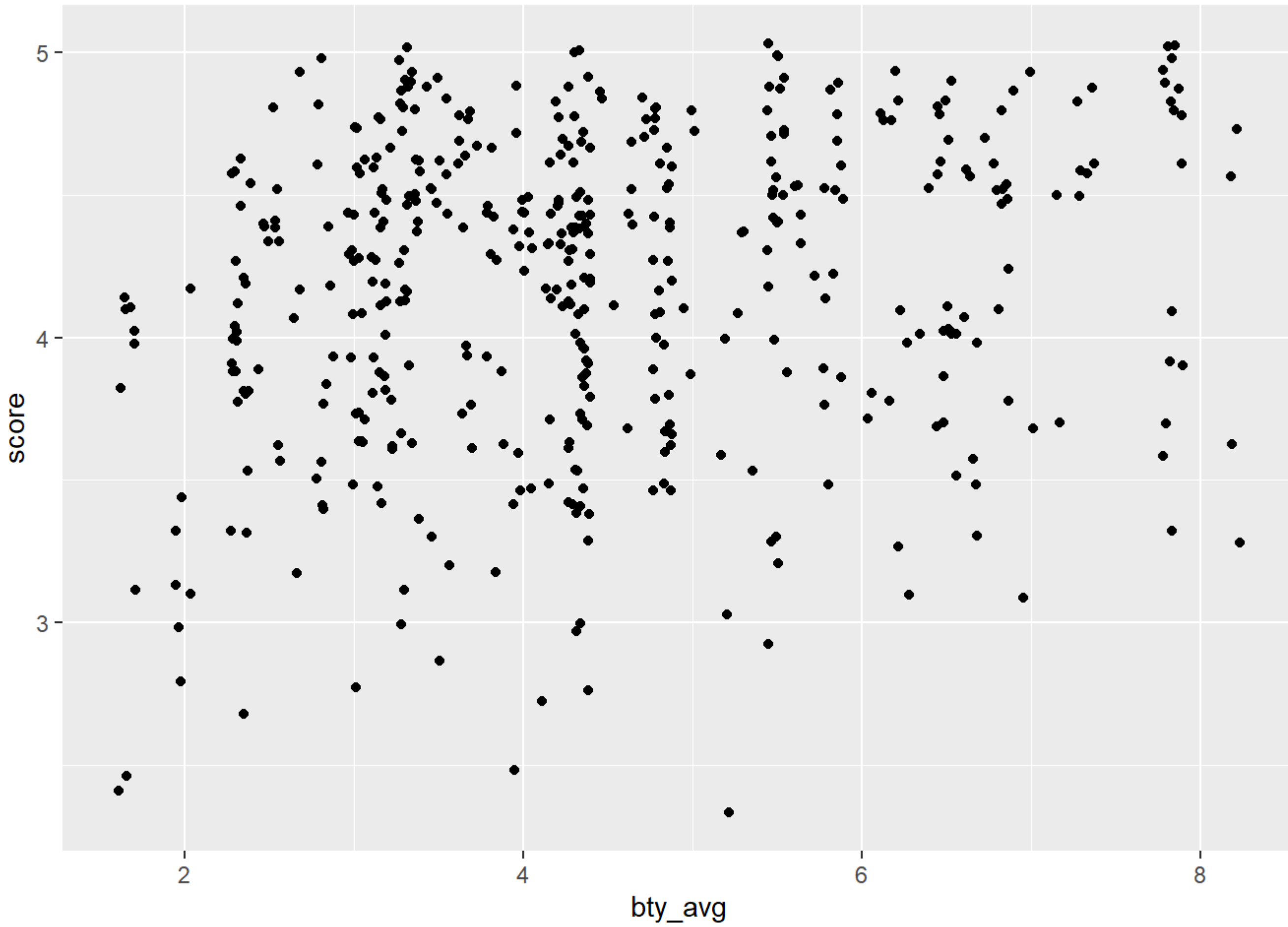
The distribution of scores look unimodal and left-skewed. most scores are high rather than in the middle. Maybe this is surprising if we expect most courses to be average. But maybe people either really like their courses on the whole or maybe feel bad about giving bad scores

Exercise 3:

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




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



What was misleading about the original plot is that the points were stacked on top of each other (called overplotting). Geom_jiter fixes this by moving the points by a small random amount.

Exercise 4:

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

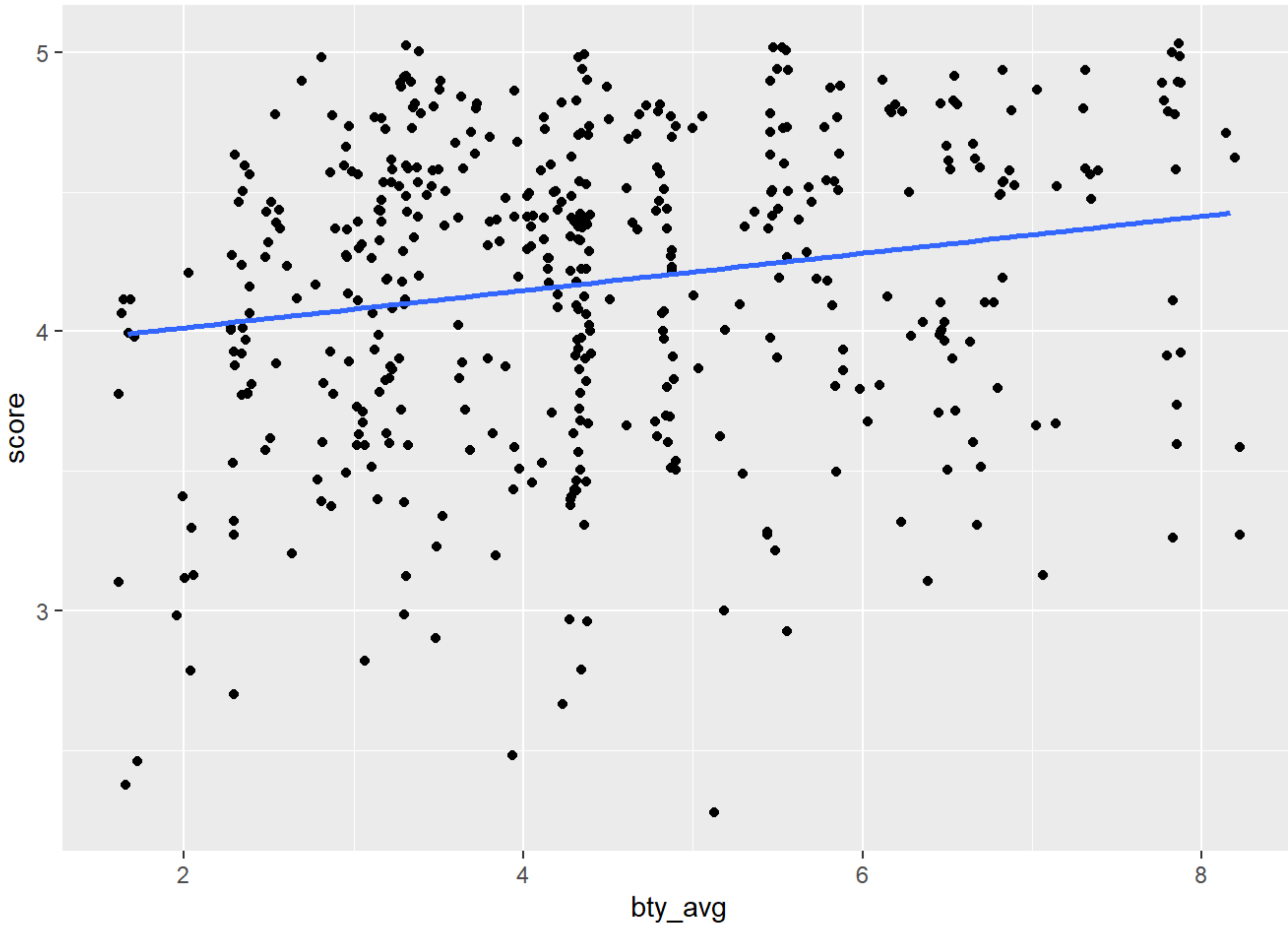
```
##
## Call:
## lm(formula = score ~ bty_avg, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9246 -0.3690  0.1420  0.3977  0.9309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.88034    0.07614   50.96  < 2e-16 ***
## bty_avg      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
```

Score = 0.066*bty_avg + 3.88 In this model, bty_avg is there reliable predictor because its P value is a small. exchange of 0.066 in score for each point of bty_avg might or might not be practically significant. it seems small relative to the variability in score, but it might matter in some kind of evaluation.

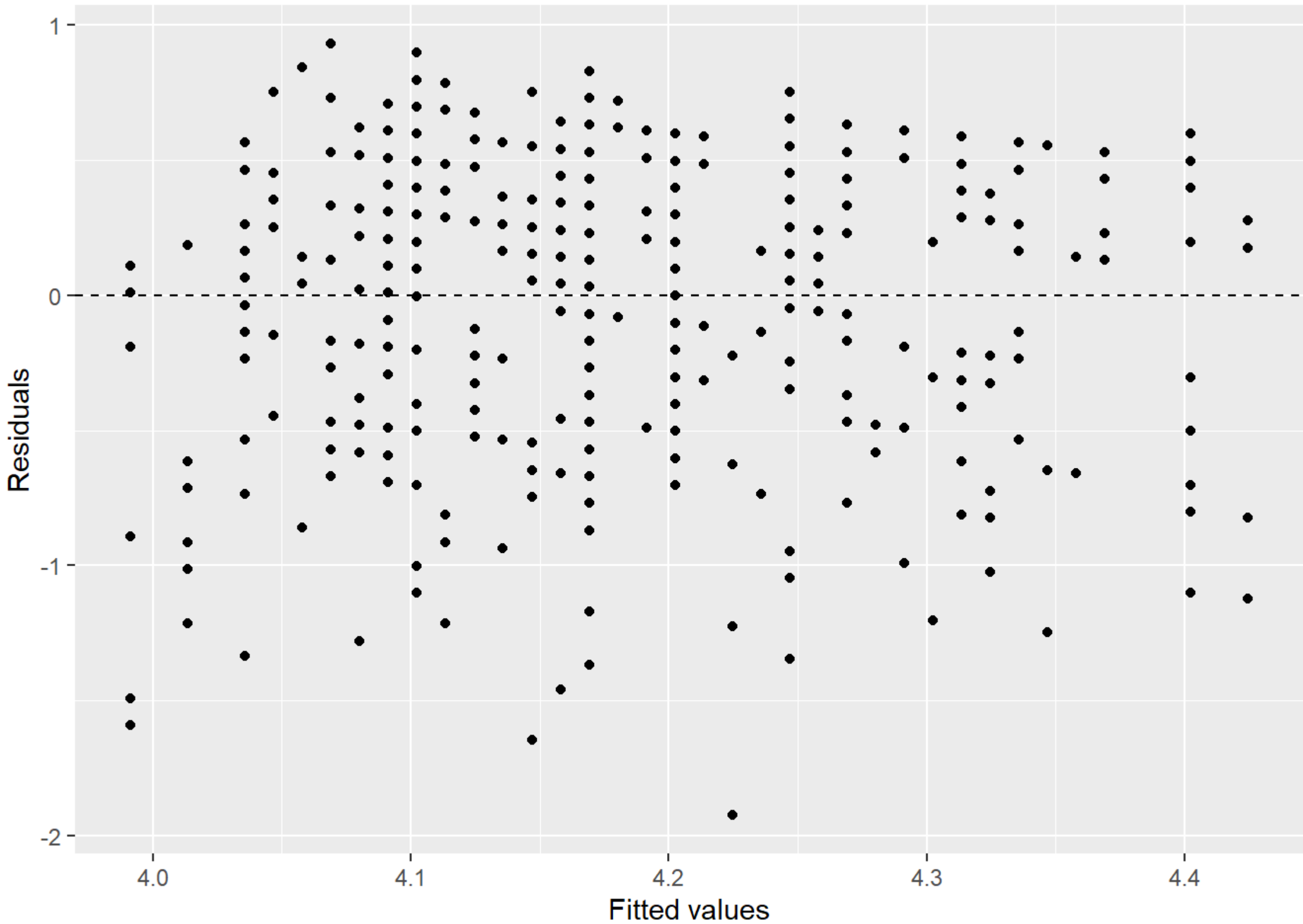
Exercise 5:

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

```
## `geom_smooth()` using formula 'y ~ x'
```

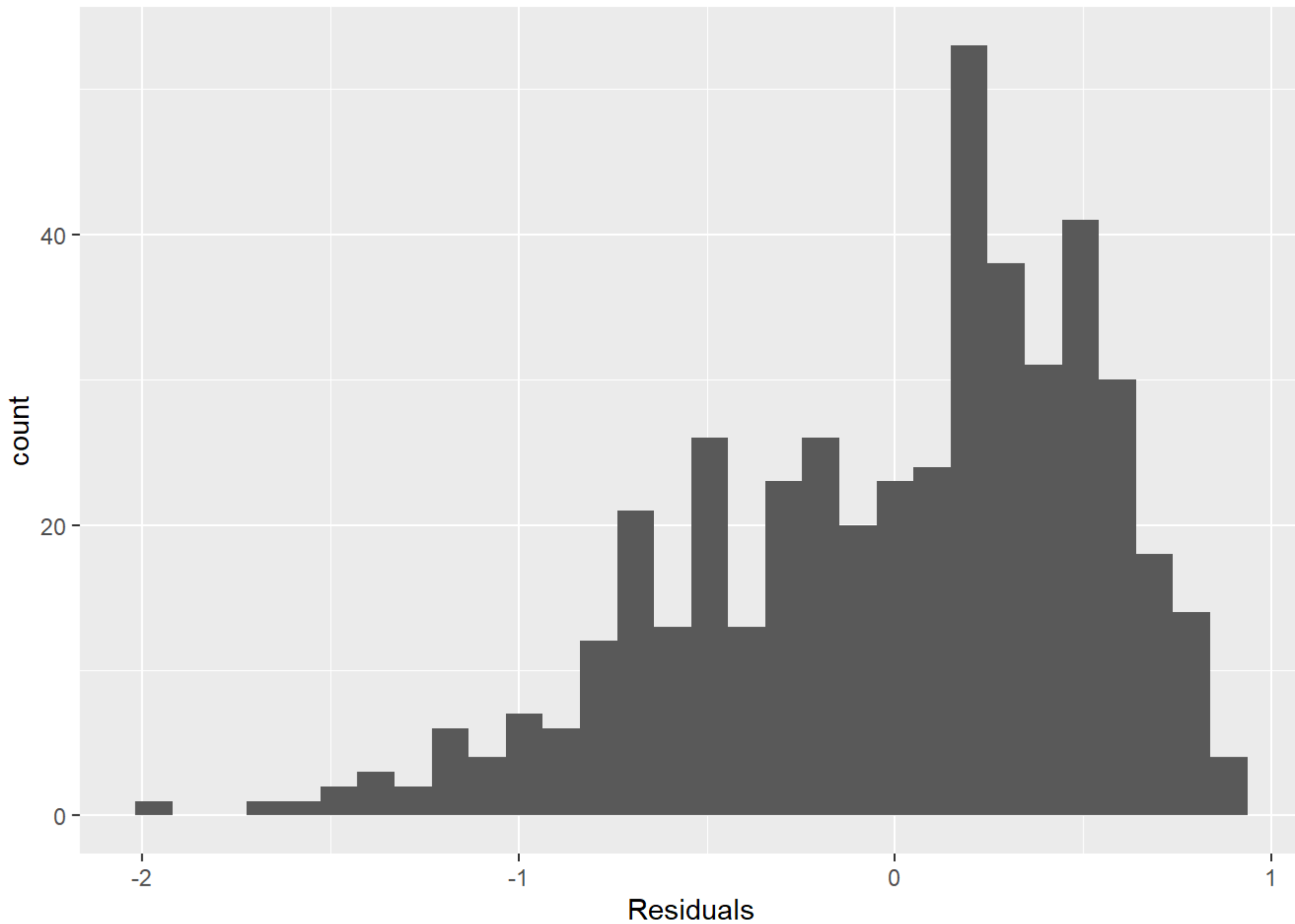



```
mbty_residuals <- tibble(x = nrow(evals),
                       fitted = fitted(m_bty),
                       resid = residuals(m_bty))

#linearity #constant variability
ggplot(mbty_residuals, aes(x = fitted, y = resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```




```
#Normal residuals
ggplot(mbty_residuals, aes(x = resid)) +
  geom_histogram() +
  xlab("Residuals")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

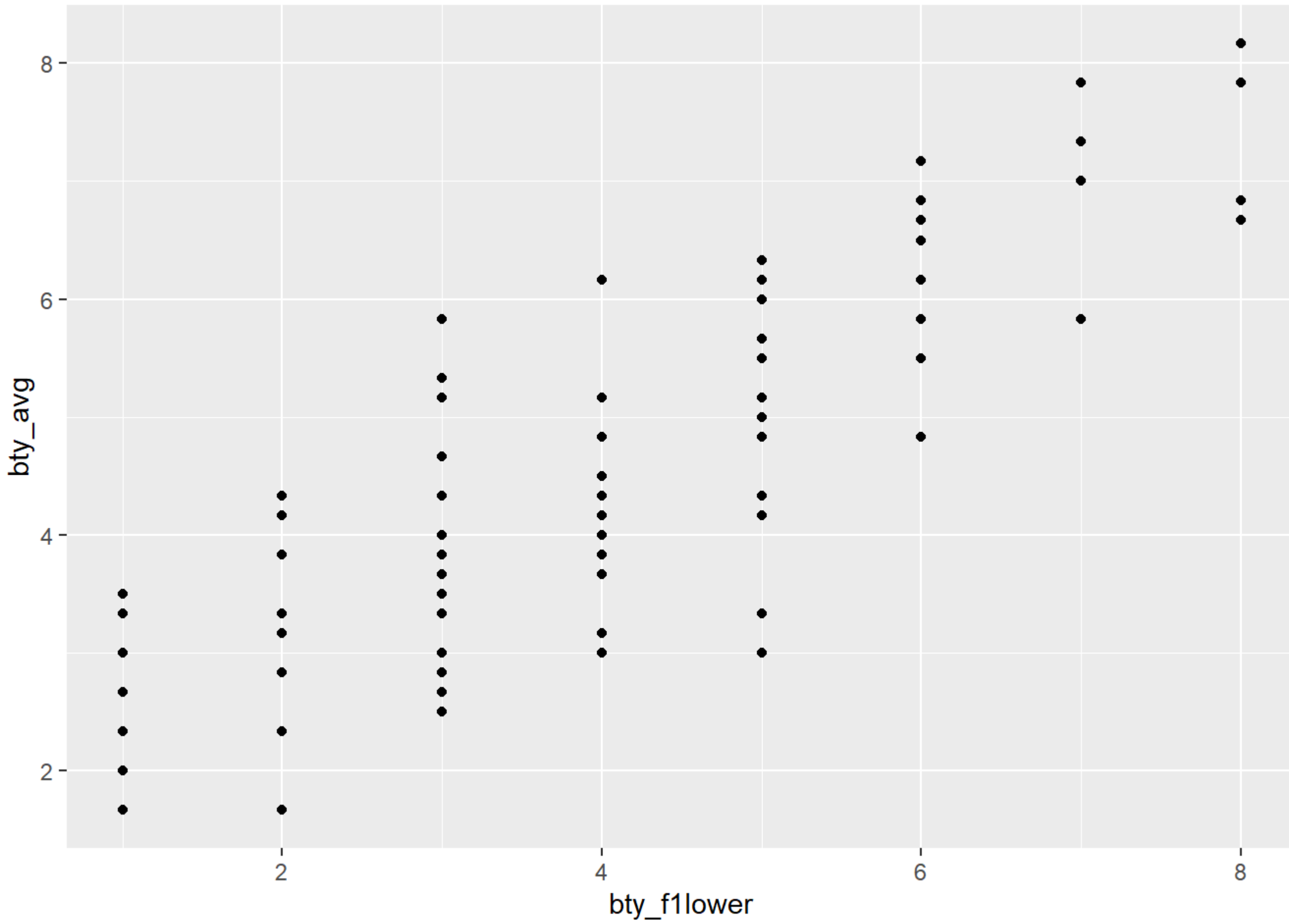


The

normal residuals condition appears to be violated the residuals are left skewed. the other two conditions look pretty OK the residuals don't have an obvious pattern and appear roughly constant variance (although maybe a little bit less at the high end of the range).

Exercise 6:

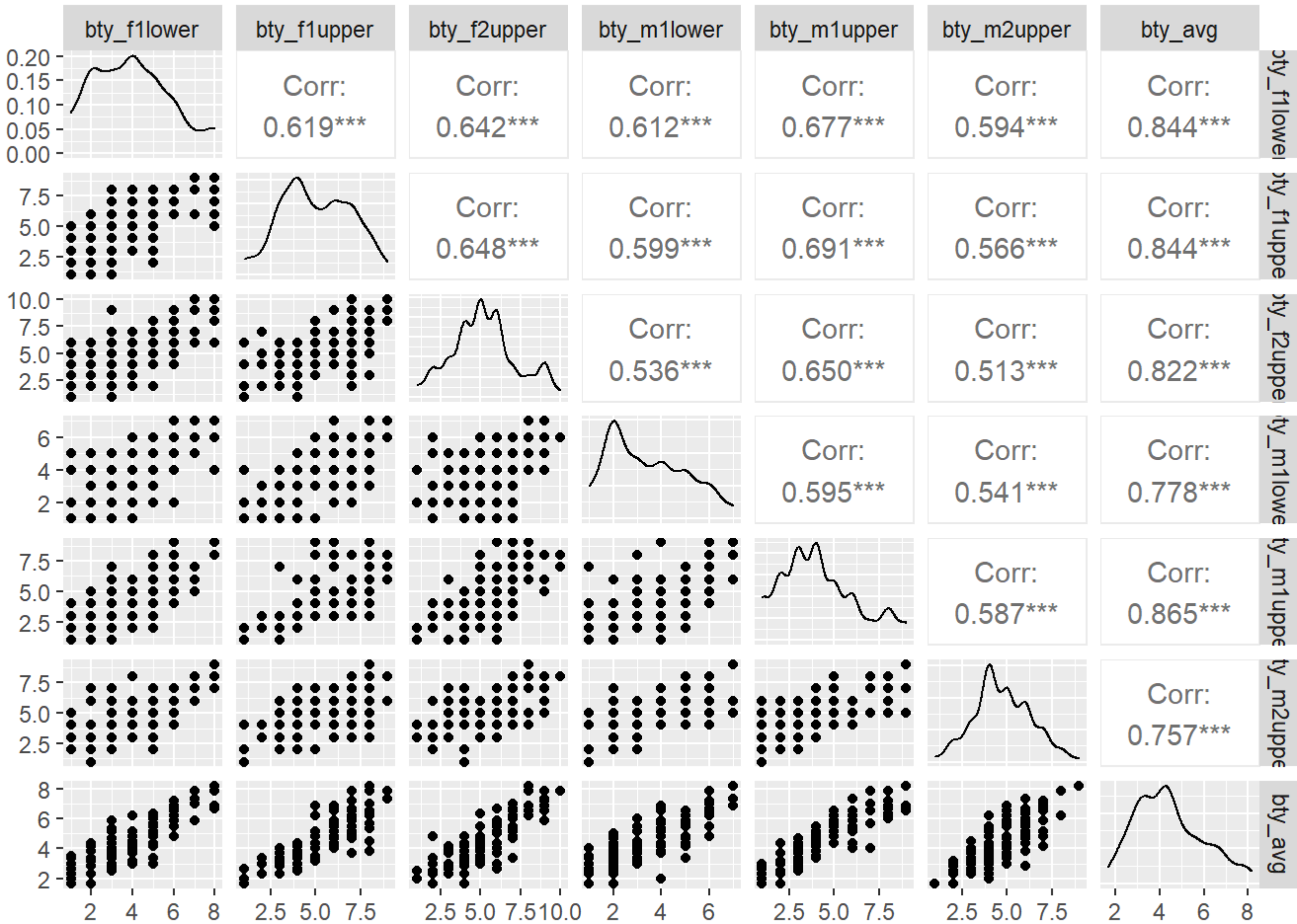
```
ggplot(evals, aes(x = bty_f1lower, y = bty_avg)) +
  geom_point()
```



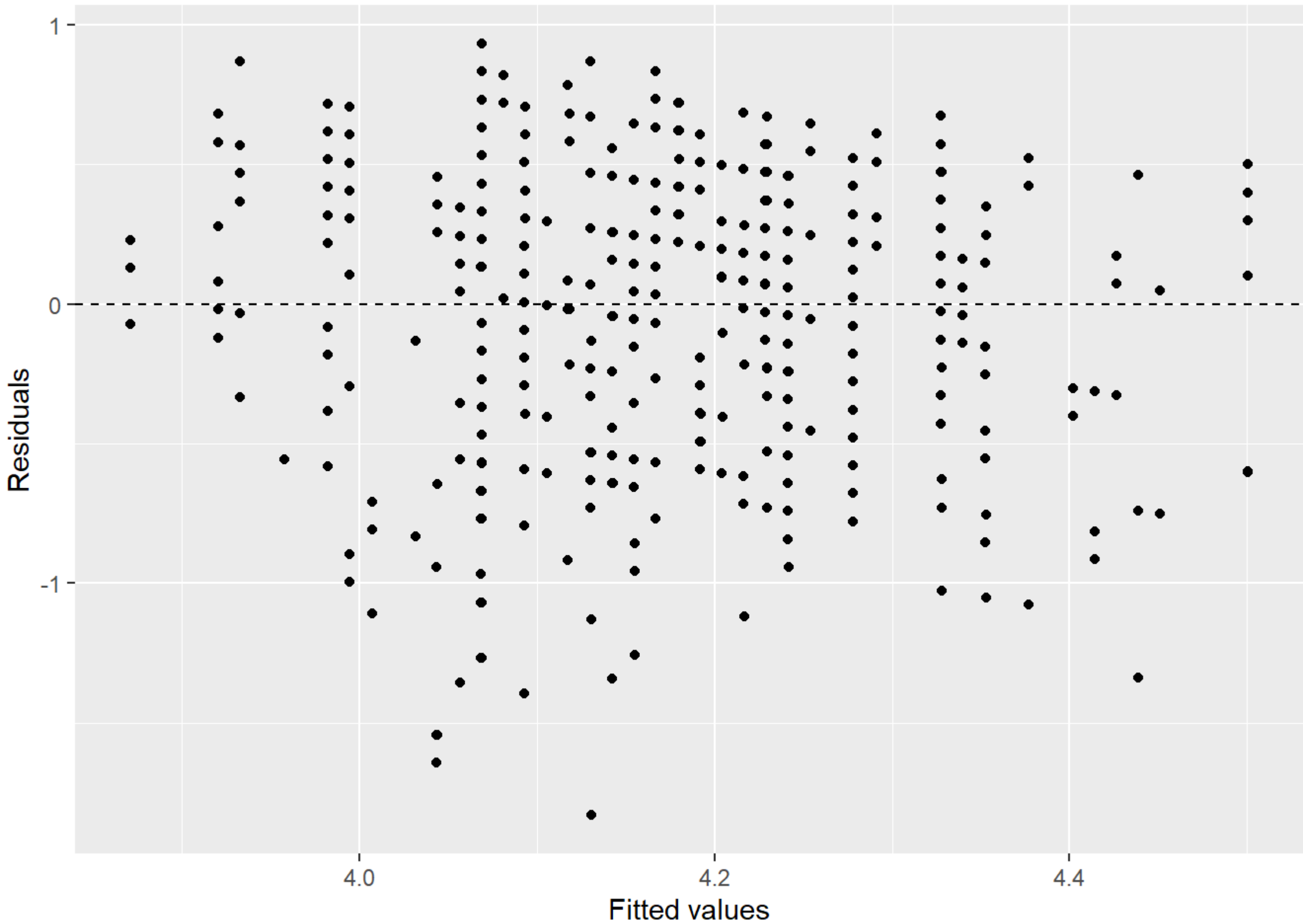

```
evals %>%
  summarise(cor = cor(bty_avg, bty_f1lower)) %>%
  pull()
```

```
## [1] 0.8439112
```

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

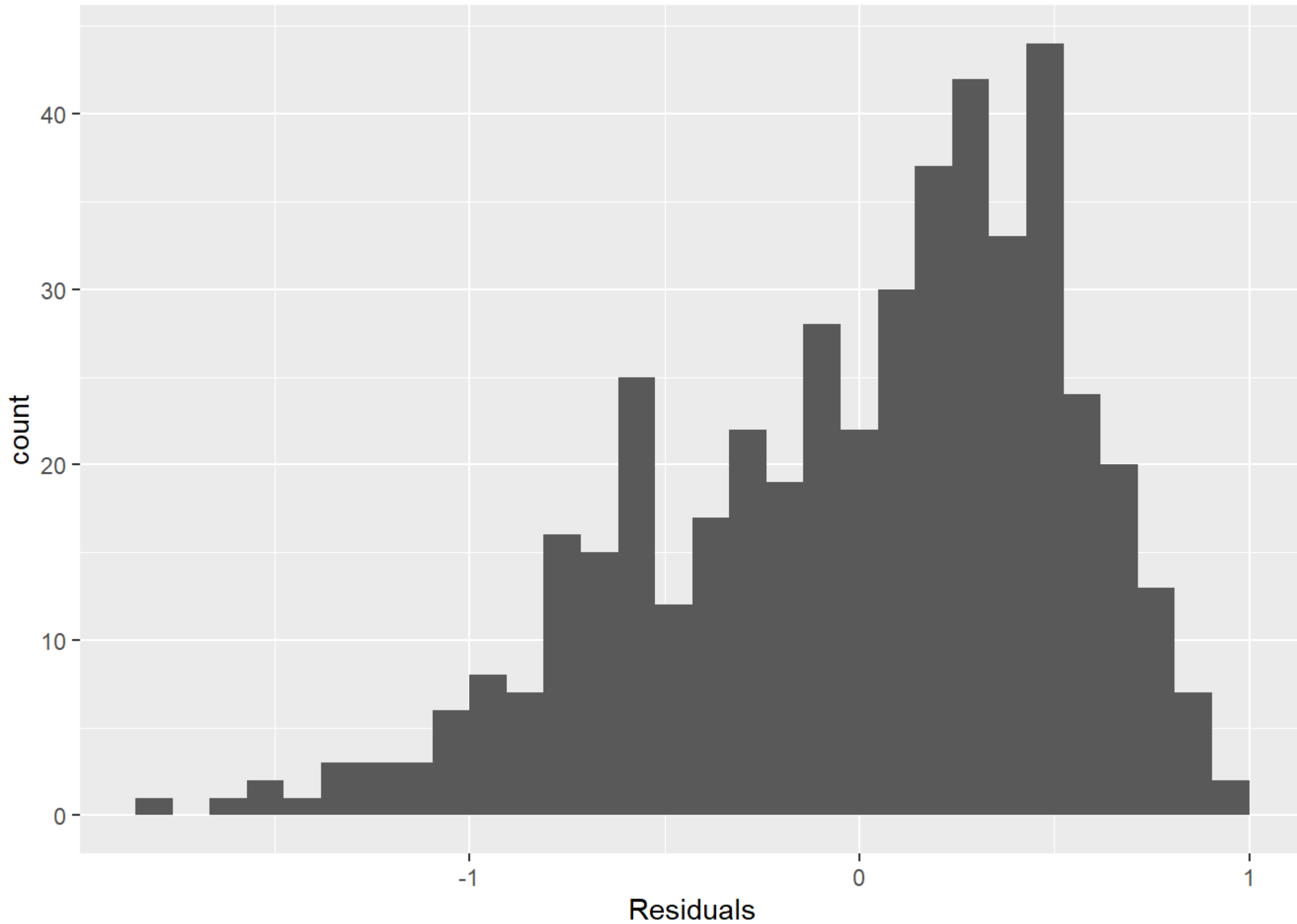



```r
m_bty_gen <- lm(score ~ bty_avg + gender, data = evals)


m_bty_gen_residuals <- tibble(x = nrow(evals),
                       fitted = fitted(m_bty_gen),
                       resid = residuals(m_bty_gen))

#linearity #constant variability
ggplot(m_bty_gen_residuals, aes(x = fitted, y = resid)) +
  geom_jitter() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  xlab("Fitted values") +
  ylab("Residuals")
```




```
#Normal residuals
ggplot(m_bty_gen_residuals, aes(x = resid)) +
  geom_histogram() +
  xlab("Residuals")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#summary(m_bty_gen)
```

Normality is again violated the distribution of residuals is left skewed. this time there is also a stronger violation of constant variability. it looks like there is more variability in the middle part of the range than on the edges.

Exercise 7:

```
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
```

Yes, beauty average is this still a reliable predictor as the P values are small for both gender and beauty average. including gender in the model did change its estimate though.

Exercise 8:

Score = 3.91 + .074*bty_avg for males Score = 3.74 + .074*bty_avg for females

This model says that men get better ratings on average, holding beauty constant.

Exercise 9:

```
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
```

We have turned these tree levels of one this variable into two separate binary variables. The intercept is there non tenure track is going. Because of that we can compare each of these two (tenure track and tenured) two non tenure track because that is what the reference category is, but we cannot compare them with each other. The numbers show that tenure track and tenured professors are doing worse compared to the not tenure track ones. But because of the way that model is set up we cannot compare the tenure track with tenured.

Exercise 10:

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

summary(m_full)
```

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_students + cls_level + cls_profs + cls_credits +
##     bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77397 -0.32432  0.09067  0.35183  0.95036
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.0952141  0.2905277  14.096  < 2e-16 ***
## ranktenure track      -0.1475932  0.0820671  -1.798  0.07278 .
## ranktenured           -0.0973378  0.0663296  -1.467  0.14295
## ethnicitynot minority  0.1234929  0.0786273   1.571  0.11698
## gendermale             0.2109481  0.0518230   4.071 5.54e-05 ***
## languagenon-english   -0.2298112  0.1113754  -2.063  0.03965 *
## age                   -0.0090072  0.0031359  -2.872  0.00427 **
## cls_perc_eval          0.0053272  0.0015393   3.461  0.00059 ***
## cls_students           0.0004546  0.0003774   1.205  0.22896
## cls_levelupper         0.0605140  0.0575617   1.051  0.29369
## cls_profssingle       -0.0146619  0.0519885  -0.282  0.77806
## cls_creditsone credit  0.5020432  0.1159388   4.330 1.84e-05 ***
## bty_avg                0.0400333  0.0175064   2.287  0.02267 *
## pic_outfitnot formal  -0.1126817  0.0738800  -1.525  0.12792
## pic_colorcolor        -0.2172630  0.0715021  -3.039  0.00252 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.498 on 448 degrees of freedom
## Multiple R-squared:  0.1871, Adjusted R-squared:  0.1617
## F-statistic: 7.366 on 14 and 448 DF,  p-value: 6.552e-14
```

Ethnicity has two categories, nonminority and minority. The reference category is minority. All else being equal non minority professors R rated on average 0.12 points higher than minority professors. If we want to look to the P value it is not that small so overall ethnicity white not be a very good variable to include in the model for prediction of perfessor rates.

Exercise 11:

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

summary(m_full)
```

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + cls_students + cls_level + cls_credits +
##     bty_avg + pic_outfit + pic_color, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7836 -0.3257  0.0859  0.3513  0.9551
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.0872523  0.2888562  14.150  < 2e-16 ***
## ranktenure track      -0.1476746  0.0819824  -1.801 0.072327 .
## ranktenured           -0.0973829  0.0662614  -1.470 0.142349
## ethnicitynot minority  0.1274458  0.0772887   1.649 0.099856 .
## gendermale             0.2101231  0.0516873   4.065 5.66e-05 ***
## languagenon-english   -0.2282894  0.1111305  -2.054 0.040530 *
## age                   -0.0089992  0.0031326  -2.873 0.004262 **
## cls_perc_eval          0.0052888  0.0015317   3.453 0.000607 ***
## cls_students           0.0004687  0.0003737   1.254 0.210384
## cls_levelupper         0.0606374  0.0575010   1.055 0.292200
## cls_creditsone credit  0.5061196  0.1149163   4.404 1.33e-05 ***
## bty_avg                0.0398629  0.0174780   2.281 0.023032 *
## pic_outfitnot formal  -0.1083227  0.0721711  -1.501 0.134080
## pic_colorcolor        -0.2190527  0.0711469  -3.079 0.002205 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4974 on 449 degrees of freedom
## Multiple R-squared:  0.187,  Adjusted R-squared:  0.1634
## F-statistic: 7.943 on 13 and 449 DF,  p-value: 2.336e-14
```

the coefficients and significance of the other explanatory variables slightly changed, the changes are very small so we can ignore it and we can consider that as no change. so because it is considered as no change then the dropped variable was not collinear with the other explanatory variables because the correlation is not significant.

Exercise 12:

```
lm_all <- lm(score ~ . , data = evals)

#summary(lm_all)
lm_step <- step(lm_all)
```

```
## Start:  AIC=-631.44
## score ~ rank + ethnicity + gender + language + age + cls_perc_eval +
##     cls_did_eval + cls_students + cls_level + cls_profs + cls_credits +
##     bty_f1lower + bty_f1upper + bty_f2upper + bty_m1lower + bty_m1upper +
##     bty_m2upper + bty_avg + pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - cls_profs      1    0.0076 107.66 -633.41
## - cls_level      1    0.0756 107.73 -633.12
## - cls_students   1    0.1325 107.78 -632.87
## - bty_m2upper    1    0.1646 107.81 -632.73
## - bty_m1lower    1    0.1649 107.82 -632.73
## - bty_m1upper    1    0.1653 107.82 -632.73
## - bty_avg        1    0.1655 107.82 -632.73
## - bty_f1lower    1    0.1657 107.82 -632.73
## - bty_f2upper    1    0.1662 107.82 -632.73
## - bty_f1upper    1    0.1669 107.82 -632.72
## - cls_did_eval   1    0.1974 107.85 -632.59
## - pic_outfit     1    0.3270 107.98 -632.04
## <none>                       107.65 -631.44
## - rank           2    1.0319 108.68 -631.02
## - cls_perc_eval  1    0.5699 108.22 -631.00
## - ethnicity      1    0.9156 108.57 -629.52
## - language       1    1.2320 108.88 -628.17
## - age            1    1.5956 109.25 -626.63
## - pic_color      1    2.2190 109.87 -623.99
## - cls_credits    1    4.4803 112.13 -614.56
## - gender         1    5.6448 113.30 -609.78
##
## Step:  AIC=-633.41
## score ~ rank + ethnicity + gender + language + age + cls_perc_eval +
##     cls_did_eval + cls_students + cls_level + cls_credits + bty_f1lower +
##     bty_f1upper + bty_f2upper + bty_m1lower + bty_m1upper + bty_m2upper +
##     bty_avg + pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - cls_level      1    0.0751 107.73 -635.09
## - cls_students   1    0.1299 107.79 -634.85
## - bty_m2upper    1    0.1578 107.82 -634.73
## - bty_m1lower    1    0.1581 107.82 -634.73
## - bty_m1upper    1    0.1585 107.82 -634.73
## - bty_avg        1    0.1587 107.82 -634.73
## - bty_f1lower    1    0.1590 107.82 -634.72
## - bty_f2upper    1    0.1594 107.82 -634.72
## - bty_f1upper    1    0.1601 107.82 -634.72
## - cls_did_eval   1    0.1929 107.85 -634.58
## - pic_outfit     1    0.3673 108.03 -633.83
## <none>                       107.66 -633.41
## - rank           2    1.0311 108.69 -632.99
## - cls_perc_eval  1    0.5958 108.25 -632.85
## - ethnicity      1    0.9207 108.58 -631.47
## - language       1    1.2392 108.90 -630.11
```

```
## - age            1    1.6029 109.26 -628.57
## - pic_color      1    2.2156 109.87 -625.98
## - cls_credits    1    4.5061 112.16 -616.42
## - gender         1    5.6685 113.33 -611.65
## 
## Step:  AIC=-635.09
## score ~ rank + ethnicity + gender + language + age + cls_perc_eval + 
##     cls_did_eval + cls_students + cls_credits + bty_f1lower + 
##     bty_f1upper + bty_f2upper + bty_m1lower + bty_m1upper + bty_m2upper + 
##     bty_avg + pic_outfit + pic_color
## 
##                 Df Sum of Sq    RSS     AIC
## - bty_m2upper    1    0.1353 107.87 -636.50
## - bty_m1lower    1    0.1356 107.87 -636.50
## - bty_m1upper    1    0.1359 107.87 -636.50
## - bty_avg        1    0.1361 107.87 -636.50
## - bty_f1lower    1    0.1364 107.87 -636.50
## - bty_f2upper    1    0.1368 107.87 -636.50
## - bty_f1upper    1    0.1374 107.87 -636.49
## - cls_students   1    0.1910 107.92 -636.27
## - cls_did_eval   1    0.2465 107.98 -636.03
## - pic_outfit     1    0.4078 108.14 -635.34
## <none>                      107.73 -635.09
## - rank           2    1.0007 108.73 -634.80
## - cls_perc_eval  1    0.5664 108.30 -634.66
## - ethnicity      1    0.9869 108.72 -632.86
## - language       1    1.1731 108.91 -632.07
## - age            1    1.5633 109.30 -630.41
## - pic_color      1    2.1435 109.88 -627.96
## - cls_credits    1    4.4849 112.22 -618.20
## - gender         1    5.6057 113.34 -613.60
## 
## Step:  AIC=-636.5
## score ~ rank + ethnicity + gender + language + age + cls_perc_eval + 
##     cls_did_eval + cls_students + cls_credits + bty_f1lower + 
##     bty_f1upper + bty_f2upper + bty_m1lower + bty_m1upper + bty_avg + 
##     pic_outfit + pic_color
## 
##                 Df Sum of Sq    RSS     AIC
## - bty_m1lower    1    0.0319 107.90 -638.37
## - bty_m1upper    1    0.1637 108.03 -637.80
## - cls_students   1    0.2207 108.09 -637.56
## - cls_did_eval   1    0.2710 108.14 -637.34
## - pic_outfit     1    0.4084 108.28 -636.75
## <none>                      107.87 -636.50
## - bty_f1lower    1    0.4873 108.36 -636.42
## - cls_perc_eval  1    0.5417 108.41 -636.18
## - bty_avg        1    0.6009 108.47 -635.93
## - rank           2    1.1260 109.00 -635.70
## - ethnicity      1    0.8591 108.73 -634.83
## - language       1    1.1373 109.01 -633.65
## - bty_f2upper    1    1.1690 109.04 -633.51
```

```
## - age            1    1.5120 109.38 -632.06
## - bty_f1upper    1    1.8540 109.72 -630.61
## - pic_color      1    2.3311 110.20 -628.61
## - cls_credits    1    4.3939 112.26 -620.02
## - gender         1    5.8105 113.68 -614.21
##
## Step:  AIC=-638.37
## score ~ rank + ethnicity + gender + language + age + cls_perc_eval +
##     cls_did_eval + cls_students + cls_credits + bty_f1lower +
##     bty_f1upper + bty_f2upper + bty_m1upper + bty_avg + pic_outfit +
##     pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - bty_m1upper    1    0.1347 108.03 -639.79
## - cls_students   1    0.2258 108.13 -639.40
## - cls_did_eval   1    0.2842 108.19 -639.15
## - pic_outfit     1    0.4271 108.33 -638.54
## <none>                       107.90 -638.37
## - bty_f1lower    1    0.5241 108.42 -638.12
## - cls_perc_eval  1    0.5243 108.42 -638.12
## - rank           2    1.0984 109.00 -637.68
## - ethnicity      1    0.9341 108.83 -636.38
## - language       1    1.1373 109.04 -635.51
## - bty_avg        1    1.1563 109.06 -635.43
## - bty_f2upper    1    1.4875 109.39 -634.03
## - age            1    1.6158 109.52 -633.49
## - bty_f1upper    1    2.3451 110.25 -630.41
## - pic_color      1    2.4870 110.39 -629.82
## - cls_credits    1    4.3675 112.27 -622.00
## - gender         1    5.8001 113.70 -616.13
##
## Step:  AIC=-639.79
## score ~ rank + ethnicity + gender + language + age + cls_perc_eval +
##     cls_did_eval + cls_students + cls_credits + bty_f1lower +
##     bty_f1upper + bty_f2upper + bty_avg + pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - cls_students   1    0.2111 108.25 -640.89
## - cls_did_eval   1    0.2718 108.31 -640.63
## - pic_outfit     1    0.4059 108.44 -640.05
## - bty_f1lower    1    0.4326 108.47 -639.94
## <none>                       108.03 -639.79
## - cls_perc_eval  1    0.6118 108.65 -639.18
## - rank           2    1.1107 109.15 -639.05
## - ethnicity      1    0.8677 108.90 -638.09
## - language       1    1.0897 109.12 -637.14
## - bty_avg        1    1.2696 109.31 -636.38
## - bty_f2upper    1    1.3535 109.39 -636.03
## - age            1    1.8207 109.86 -634.05
## - bty_f1upper    1    2.2334 110.27 -632.32
## - pic_color      1    2.5154 110.55 -631.13
## - cls_credits    1    4.8529 112.89 -621.45
```

```
## - gender         1    5.7018 113.74 -617.98
##
## Step:  AIC=-640.89
## score ~ rank + ethnicity + gender + language + age + cls_perc_eval +
##     cls_did_eval + cls_credits + bty_f1lower + bty_f1upper +
##     bty_f2upper + bty_avg + pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## - cls_did_eval   1    0.1415 108.39 -642.28
## - pic_outfit     1    0.3573 108.60 -641.36
## - bty_f1lower    1    0.4667 108.71 -640.89
## <none>                      108.25 -640.89
## - rank           2    1.1821 109.43 -639.86
## - ethnicity      1    0.9919 109.24 -638.66
## - language       1    0.9963 109.24 -638.64
## - bty_avg        1    1.3404 109.59 -637.19
## - bty_f2upper    1    1.5770 109.82 -636.19
## - age            1    1.8395 110.09 -635.08
## - bty_f1upper    1    2.1827 110.43 -633.64
## - pic_color      1    2.3708 110.62 -632.85
## - cls_perc_eval  1    2.4497 110.70 -632.52
## - cls_credits    1    4.8839 113.13 -622.45
## - gender         1    5.5358 113.78 -619.79
##
## Step:  AIC=-642.28
## score ~ rank + ethnicity + gender + language + age + cls_perc_eval +
##     cls_credits + bty_f1lower + bty_f1upper + bty_f2upper + bty_avg +
##     pic_outfit + pic_color
##
##                 Df Sum of Sq    RSS     AIC
## <none>                      108.39 -642.28
## - pic_outfit     1    0.5065 108.89 -642.12
## - bty_f1lower    1    0.5212 108.91 -642.06
## - rank           2    1.2018 109.59 -641.18
## - ethnicity      1    1.0217 109.41 -639.94
## - language       1    1.0886 109.48 -639.65
## - bty_avg        1    1.4304 109.82 -638.21
## - bty_f2upper    1    1.7306 110.12 -636.95
## - age            1    1.9987 110.39 -635.82
## - pic_color      1    2.2613 110.65 -634.72
## - bty_f1upper    1    2.2891 110.68 -634.60
## - cls_perc_eval  1    2.3203 110.71 -634.47
## - cls_credits    1    4.9069 113.30 -623.78
## - gender         1    5.7484 114.14 -620.35
```

the final model is “score ~ rank + ethnicity + gender + language + age + cls_perc_eval + cls_credits + bty_f1lower + bty_f1upper + bty_f2upper + bty_avg + pic_outfit + pic_color”

```r
Linear_Final_Model <- lm(score ~ rank + ethnicity + gender + language + age + cls_perc_eval
             + bty_f1upper + cls_credits + bty_avg + bty_f2upper
             + pic_outfit + pic_color + bty_f1lower , data = evals)

summary(Linear_Final_Model)
```

```
##
## Call:
## lm(formula = score ~ rank + ethnicity + gender + language + age +
##     cls_perc_eval + bty_f1upper + cls_credits + bty_avg + bty_f2upper +
##     pic_outfit + pic_color + bty_f1lower, data = evals)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7317 -0.3070  0.1020  0.3535  0.9283
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.087168   0.284915  14.345  < 2e-16 ***
## ranktenure track      -0.184119   0.082619  -2.229  0.02634 *
## ranktenured           -0.079796   0.065691  -1.215  0.22511
## ethnicitynot minority  0.156156   0.075989   2.055  0.04046 *
## gendermale             0.254929   0.052300   4.874 1.52e-06 ***
## languagenon-english   -0.228479   0.107710  -2.121  0.03445 *
## age                   -0.009130   0.003177  -2.874  0.00424 **
## cls_perc_eval          0.004470   0.001444   3.097  0.00208 **
## bty_f1upper            0.079899   0.025975   3.076  0.00223 **
## cls_creditsone credit  0.502732   0.111631   4.504 8.54e-06 ***
## bty_avg               -0.140359   0.057725  -2.432  0.01543 *
## bty_f2upper            0.058486   0.021868   2.674  0.00776 **
## pic_outfitnot formal  -0.100300   0.069317  -1.447  0.14860
## pic_colorcolor        -0.205474   0.067209  -3.057  0.00237 **
## bty_f1lower            0.038838   0.026460   1.468  0.14287
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4919 on 448 degrees of freedom
## Multiple R-squared:  0.2068, Adjusted R-squared:  0.1821
## F-statistic: 8.345 on 14 and 448 DF,  p-value: 4.79e-16
```

linear model for predicting score based on the final model: score = 4.087168 - 0.184119* ranktenure track - 0.079796* ranktenured + 0.156156* ethnicitynot minority + 0.254929* gendermale - 0.228479* languagenon-english - 0.009130* age + 0.004470* cls_perc_eval + 0.079899* bty_f1upper + 0.502732* cls_creditsone credit - 0.140359* bty_avg + 0.058486* bty_f2upper - 0.100300* pic_outfitnot formal - 0.205474* pic_colorcolor + 0.038838* bty_f1lower

Exercise 13:

All else being equal, in terms of classes, one credit classes (lab, PE, etc.) and classes with high percent of students in class who completed evaluation are scored higher. In terms of professor, the professors being non-tenure track (teaching), not minority, male, non-English, younger, having black & white picture, and wearing formal outfits are scored higher.