lab8

jason grahn

11/24/2018

# The R Part

## 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.

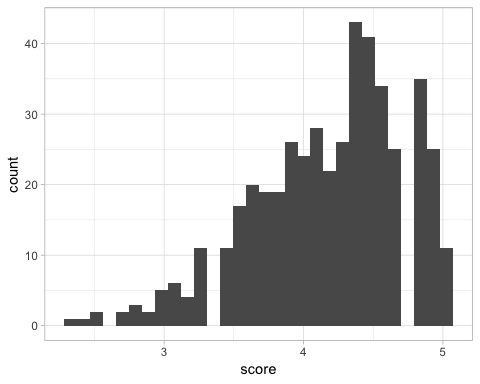
This is an observational study. The primary indicator is that there are no control or experimental groupings. With an observation study we may not be able to provide causation, but we can build a string inference based on correlation. We *can* ask if an instructors beauty has a positive or negative correlation on student course evaluation.

## 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?

evals %>%   
 ggplot() +   
 theme\_light() +   
 geom\_histogram(aes(x = score))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



summary(evals$score)

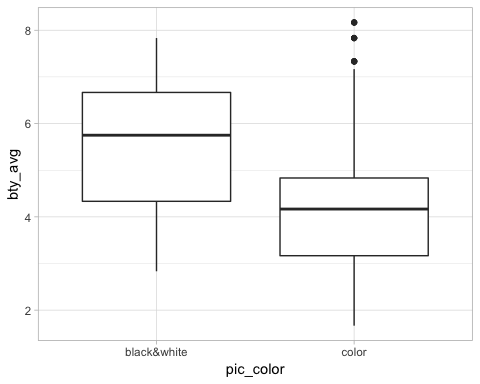
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.300 3.800 4.300 4.175 4.600 5.000

The distribution of score is unimodal with a strong right skew. From this graph, we could infer that students tend to give higher scores more often than lower scores. We would *expect* a normal distribution, but with a mean of 4.175 and a median of 4.3, . Perhaps this is related to some biased questions within the evaluation?

## 3

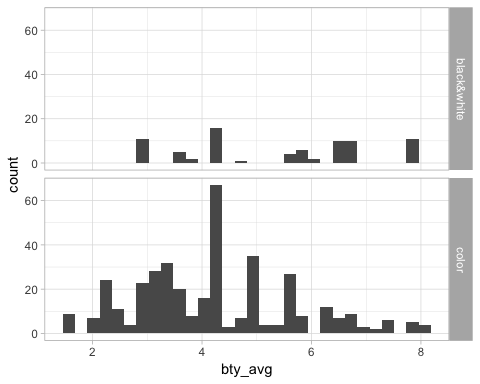
### Excluding score, select two other variables and describe their relationship using an appropriate visualization (scatterplot, side-by-side boxplots, or mosaic plot).

evals\_plot <- evals %>%   
 ggplot() +   
 theme\_light()  
  
evals\_plot +  
 geom\_boxplot(aes(x = pic\_color, y = bty\_avg))



evals\_plot +   
 geom\_histogram(aes(bty\_avg)) +   
 facet\_grid(rows = vars(pic\_color))

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



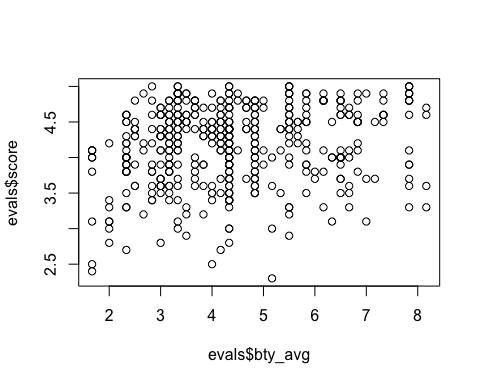
The boxplot showcases that, while the spread of bty\_avg for a color picture is larger with higher uppper scores, a black & white picture has a higher assocated bty\_avg. There seems to be a relationship here.

The side by side histograms reflect the same information that the boxplots related, but *also* show that color pictures simply receive more scores more often.

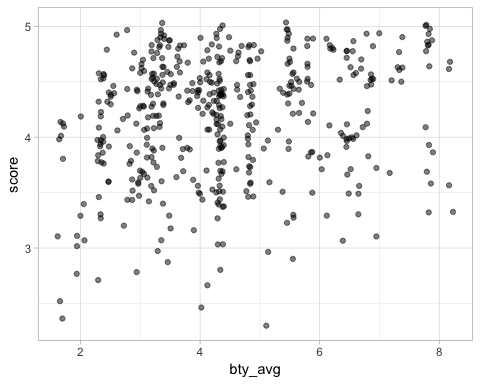
## 3

### Replot the scatterplot, but this time use the function jitter() on the - or the -coordinate. (Use ?jitter to learn more.) What was misleading about the initial scatterplot?

#the original plot  
plot(evals$score ~ evals$bty\_avg)



#ggplot is easier to modify than plot() when it comes to additional commands.  
evals\_plot +   
 geom\_jitter(aes(bty\_avg, score), alpha = 0.5)

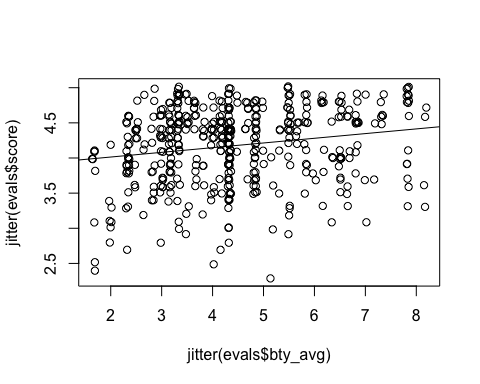


The original plot was misleading because it didn’t showcase all available points. Where the original simply overlays points on top of each other, adding jitter to a scatterplot we can differentuate these overlapping points. Another way of doing this would be heavy use of alpha which shows positions with more dense point population by making them darker.

## 4

### 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 and add the line to your plot using abline(m\_bty). 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?

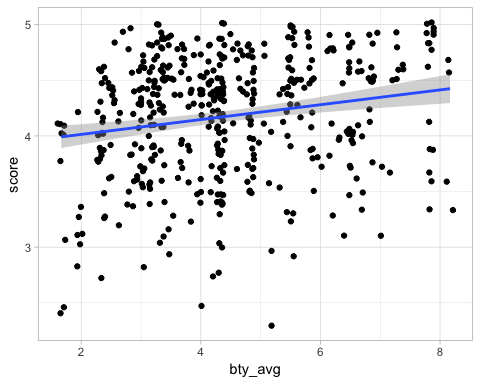
#according to the book  
m\_bty <- lm(score ~ bty\_avg, data = evals)  
  
plot(jitter(evals$score) ~ jitter(evals$bty\_avg))  
abline(m\_bty)



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

#and the more usable version using geom\_smooth for the LM line of that   
#relationship:   
evals %>%   
 ggplot(aes(x = bty\_avg, y = score)) +  
 theme\_light() +  
 geom\_jitter() +  
 geom\_smooth(method = lm)



cor(evals$score, evals$bty\_avg)

## [1] 0.1871424

get\_regression\_table(m\_bty)

## # A tibble: 2 x 7  
## term estimate std\_error statistic p\_value lower\_ci upper\_ci  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 intercept 3.88 0.076 51.0 0 3.73 4.03   
## 2 bty\_avg 0.067 0.016 4.09 0 0.035 0.099

The linear model for this relationship is

score = 0.06664\*bty\_avg + 3.88034

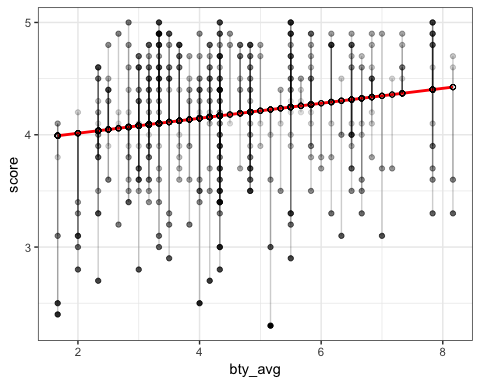
The interpretation for the slope *0.06664* is that for every positive increment change in bty\_avg, we reflect a *0.06664* positive change in evaluation score.

Running anova(m\_bty) reflects that using bty\_avg is a statisically significant predictor with a p-value of approximately zero. However, the slope is so low that this does not appear to be a *practically* significant predictor.

## 5

### 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).

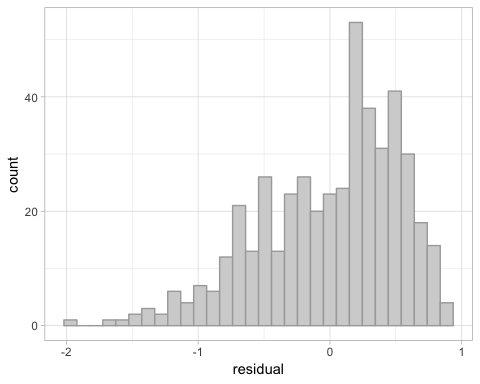
#borrowed this code from https://drsimonj.svbtle.com/visualising-residuals to build a linear regression plot  
# Steps 1 and 2 build the model and make it a data frame  
d <- lm(score ~ bty\_avg, data = evals) %>%   
 broom::augment()  
  
# Steps 3 and 4 build the plot  
ggplot(d, aes(x = bty\_avg, y = score)) +  
 geom\_smooth(method = "lm", se = FALSE, color = "red") +  
 geom\_segment(aes(xend = bty\_avg, yend = .fitted), alpha = .2) +   
 # Note `.fitted`  
 geom\_point(aes(alpha = abs(.resid))) + # Note `.resid`  
 guides(alpha = FALSE) +  
 geom\_point(aes(y = .fitted), shape = 1) + # Note `.fitted`  
 theme\_bw()



This is merely a visualization for the gut-check. My initial response is that thre are a huge collection of potential outliers and we’ll end up rejecting this model.

bty\_regression\_points <- get\_regression\_points(  
 lm(score ~ bty\_avg, data = evals)  
 )  
  
bty\_regression\_points %>%   
 ggplot() +   
 theme\_light() +  
 geom\_histogram(aes(residual), fill = "lightgrey", color = "darkgrey")

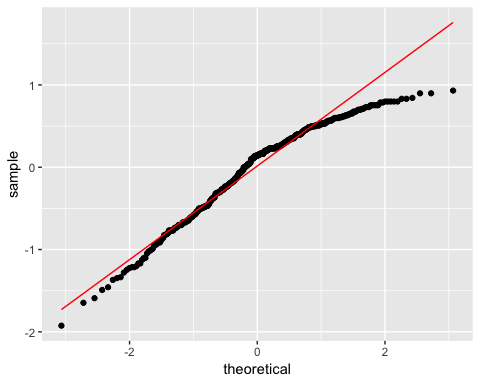
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#describe(bty\_regression\_points)

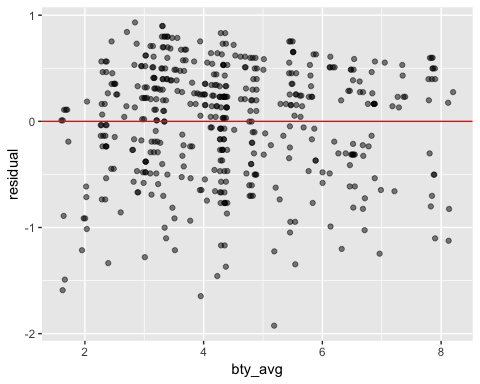
The histogram of residuals shows left skewness, which is reaffirmed with a skew factor of -0.71. This isn’t terrible, but it also isn’t great.

ggplot(bty\_regression\_points, aes(sample = residual)) +  
 stat\_qq() +  
 stat\_qq\_line(color = "red")



Though points generally hug the line, this QQ plot drops off on the upper end by quite a bit.

bty\_regression\_points %>%   
 ggplot() +  
 geom\_jitter(aes(bty\_avg, residual), alpha = 0.5) +  
 geom\_hline(yintercept = 0, color = "red")



The residual plot for bty shows weight under the zero residual line, though they DO appear generally constant, without any specific sort of pattern.

Conditions for the least squares line are Linearity, Nearly Normal residuals, and Constant Variability.

\* Linearity: The data show generally positive linearity.  
\* Nearly Normal residuals: The Histogram of residuals show a left skewed distribution. The residuals normal probability (qq) plot show that the points fall off the line at the bottom by a bit, and most definitely fall from the line at the upper end.  
\* Constant Variability: From the residual plot, we can observe that there seems to have constant variability.

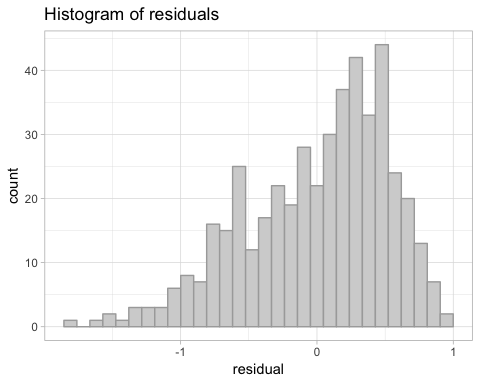
We have to assume independence given we are not provided inforamation regarding the sampling methods.

## 7

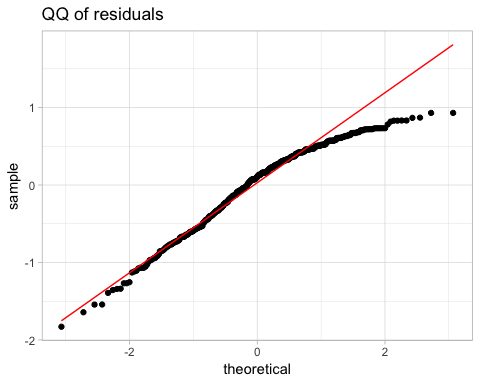
### P-values and parameter estimates should only be trusted if the conditions for the regression are reasonable. Verify that the conditions for this model are reasonable using diagnostic plots.

m\_bty\_gen <- lm(score ~ bty\_avg + gender, data = evals)  
  
m\_bty\_regression\_points <- get\_regression\_points(m\_bty\_gen)  
  
m\_bty\_regression\_points %>%   
 ggplot() +   
 theme\_light() +  
 geom\_histogram(aes(residual), fill = "lightgrey", color = "darkgrey") +  
 labs(title = "Histogram of residuals")

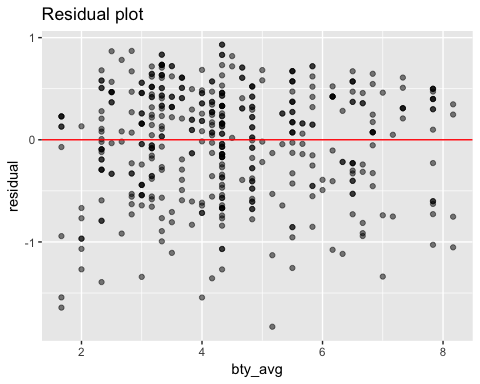
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



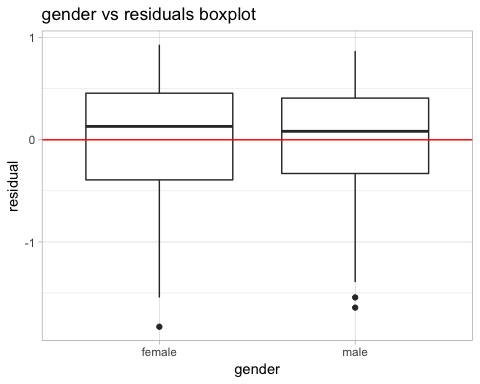
m\_bty\_regression\_points %>%   
 ggplot(aes(sample = residual)) +  
 theme\_light() +  
 stat\_qq() +  
 stat\_qq\_line(color = "red") +  
 labs(title = "QQ of residuals")



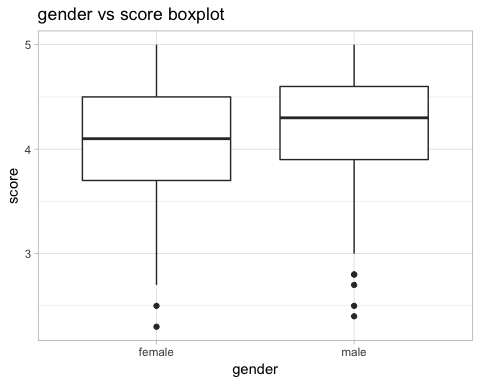
m\_bty\_regression\_points %>%   
 ggplot() +  
 geom\_point(aes(bty\_avg, residual), alpha = 0.5) +  
 geom\_hline(yintercept = 0, color = "red") +  
 labs(title = "Residual plot")



m\_bty\_regression\_points %>%   
 ggplot() +  
 theme\_light() +   
 geom\_boxplot(aes(gender, residual)) +  
 geom\_hline(yintercept = 0, color = "red") +  
 labs(title = "gender vs residuals boxplot")



m\_bty\_regression\_points %>%   
 ggplot() +  
 theme\_light() +   
 geom\_boxplot(aes(gender, score)) +  
 #geom\_hline(yintercept = 0, color = "red") +  
 labs(title = "gender vs score boxplot")



There’s at least 16 different plots we could make of these data, so I hope the above suffice. What we learn from looking at them can be simplified into a few succinct statements:

\* Residuals appear to be nearly normal.  
\* Variability within bty\_avg and residuals is still nearly constant.  
\* Each variable seems to be related to the outcome with some degree of linearity.

That said, some potential faults:

1. The histogram of residuals still suggest a left skew.  
2. The residuals still drop off the line on the upper end of the normal probability(qq) plot.

## 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?

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

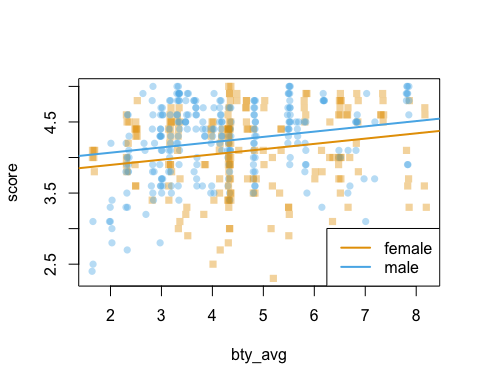
Yes, bty\_avg is still a *statistically* valid predictor for score. Adding the gendermale parameter to the model has increasedthe parameter estimate for bty\_avg from 0.06664 to 0.07416. It still isn’t very practial.

The Multiple R-Squared is rather poor at 0.059. While we might have proper statistical significance, I wouldn’t use this model.

## 9

### What is the equation of the line corresponding to males? (Hint: For males, the parameter estimate is multiplied by 1.) For two professors who received the same beauty rating, which gender tends to have the higher course evaluation score?

multiLines(m\_bty\_gen)



get\_regression\_table(m\_bty\_gen)

## # A tibble: 3 x 7  
## term estimate std\_error statistic p\_value lower\_ci upper\_ci  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 intercept 3.75 0.085 44.3 0 3.58 3.91   
## 2 bty\_avg 0.074 0.016 4.56 0 0.042 0.106  
## 3 gendermale 0.172 0.05 3.43 0.001 0.074 0.271

The equation for the line corresponding to males is

Score = 3.747 + 0.172 + 0.074(avg\_bty).

The equation for females is

Score = 3.747 + 0.074(avg\_bty)

The effect here is that for any given point on the line, a male professor will have a score approximately 0.172 points higher than a given female.

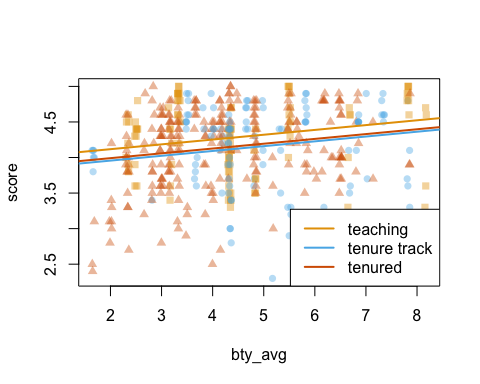
## 10

### 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 <- evals %>%   
 lm(score ~ bty\_avg + rank, data = .)  
summary(m\_bty\_rank)

##   
## Call:  
## lm(formula = score ~ bty\_avg + rank, data = .)  
##   
## 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

multiLines(m\_bty\_rank)



For a categorical variable with greater than two levels, R creates additional variables for every level greater than 1. Here, we’ve created ranktenure track and ranktenured. As a result, we must interpret the coefficients a little differently. With “teaching” rank as the base 0, knowing what rank-level a given professor has (apparently) negative consequences to their output score.

## 11

### Which variable would you expect to have the highest p-value in this model? Why? Hint: Think about which variable would you expect to not have any association with the professor score.

I think pic\_color has the least association with score because the color of their picture would have nothing to do with how well (or not) they teach. I choose that.

## 12

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

kitchen.sink <- lm(score ~ ., data = evals)  
  
get\_regression\_table(kitchen.sink) %>%   
 arrange(desc(p\_value)) %>%   
 select(term, p\_value) %>%   
 mutate(is\_max = if\_else(p\_value == (max(p\_value)),  
 TRUE,  
 FALSE)) %>%   
 filter(is\_max == TRUE)

## # A tibble: 1 x 3  
## term p\_value is\_max  
## <chr> <dbl> <lgl>   
## 1 cls\_profssingle 0.86 TRUE

summary(kitchen.sink)

##   
## Call:  
## lm(formula = score ~ ., data = evals)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.65706 -0.29305 0.08147 0.35007 0.90405   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.148734 0.320175 12.958 < 2e-16 \*\*\*  
## ranktenure track -0.173006 0.084516 -2.047 0.04125 \*   
## ranktenured -0.085964 0.068284 -1.259 0.20873   
## ethnicitynot minority 0.162897 0.084109 1.937 0.05341 .   
## gendermale 0.257620 0.053573 4.809 2.09e-06 \*\*\*  
## languagenon-english -0.250299 0.111415 -2.247 0.02516 \*   
## age -0.008470 0.003313 -2.557 0.01090 \*   
## cls\_perc\_eval 0.003259 0.002133 1.528 0.12723   
## cls\_did\_eval 0.002831 0.003148 0.899 0.36897   
## cls\_students -0.001472 0.001998 -0.737 0.46175   
## cls\_levelupper 0.033277 0.059815 0.556 0.57826   
## cls\_profssingle 0.009305 0.052743 0.176 0.86004   
## cls\_creditsone credit 0.513176 0.119785 4.284 2.25e-05 \*\*\*  
## bty\_f1lower 13.161137 15.972662 0.824 0.41040   
## bty\_f1upper 13.204997 15.969921 0.827 0.40876   
## bty\_f2upper 13.179555 15.971314 0.825 0.40970   
## bty\_m1lower 13.125270 15.968941 0.822 0.41156   
## bty\_m1upper 13.142450 15.972988 0.823 0.41107   
## bty\_m2upper 13.111995 15.968308 0.821 0.41202   
## bty\_avg -78.897853 95.826573 -0.823 0.41076   
## pic\_outfitnot formal -0.086384 0.074635 -1.157 0.24772   
## pic\_colorcolor -0.232676 0.077173 -3.015 0.00272 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4941 on 441 degrees of freedom  
## Multiple R-squared: 0.2122, Adjusted R-squared: 0.1747   
## F-statistic: 5.658 on 21 and 441 DF, p-value: 1.374e-13

I was wrong; pic\_color is not the highest (worst) p-value for the model. The highest p-value comes from cls\_profssingle with a p-value of *0.778*. This originally comes from the cls\_prof variable from the data.

## 13

### Interpret the coefficient associated with the ethnicity variable.

get\_regression\_table(kitchen.sink) %>%   
 filter(term == "ethnicitynot minority")

## # A tibble: 1 x 7  
## term estimate std\_error statistic p\_value lower\_ci upper\_ci  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 ethnicitynot min… 0.163 0.084 1.94 0.053 -0.002 0.328

The ethnicity variable was changed to ethnicitynot minority because it is a categorical variable and selected *minority* as the base. If other components of the model are equal, non-minority professors will score approximately 0.123 higher than a minority professor. Note that this also has a fairly high p\_value and may be removed from the model with further analysis.

## 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?

#dropping cls\_prof from the model  
no.cls\_profssingle <- lm(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, data = evals)  
  
summary(no.cls\_profssingle)

##   
## Call:  
## lm(formula = 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,   
## data = evals)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.65231 -0.29316 0.08609 0.35534 0.90369   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.153239 0.318805 13.028 < 2e-16 \*\*\*  
## ranktenure track -0.172918 0.084422 -2.048 0.04112 \*   
## ranktenured -0.086056 0.068207 -1.262 0.20772   
## ethnicitynot minority 0.159842 0.082216 1.944 0.05251 .   
## gendermale 0.257977 0.053476 4.824 1.94e-06 \*\*\*  
## languagenon-english -0.250910 0.111239 -2.256 0.02458 \*   
## age -0.008486 0.003308 -2.565 0.01064 \*   
## cls\_perc\_eval 0.003306 0.002114 1.564 0.11854   
## cls\_did\_eval 0.002791 0.003136 0.890 0.37398   
## cls\_students -0.001456 0.001994 -0.730 0.46558   
## cls\_levelupper 0.033179 0.059747 0.555 0.57895   
## cls\_creditsone credit 0.510496 0.118687 4.301 2.09e-05 \*\*\*  
## bty\_f1lower 12.754596 15.788232 0.808 0.41961   
## bty\_f1upper 12.797991 15.785080 0.811 0.41794   
## bty\_f2upper 12.772603 15.786532 0.809 0.41890   
## bty\_m1lower 12.718383 15.784190 0.806 0.42081   
## bty\_m1upper 12.735597 15.788303 0.807 0.42030   
## bty\_m2upper 12.705531 15.783901 0.805 0.42127   
## bty\_avg -76.456952 94.718540 -0.807 0.41998   
## pic\_outfitnot formal -0.089291 0.072714 -1.228 0.22011   
## pic\_colorcolor -0.231293 0.076689 -3.016 0.00271 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4935 on 442 degrees of freedom  
## Multiple R-squared: 0.2122, Adjusted R-squared: 0.1765   
## F-statistic: 5.952 on 20 and 442 DF, p-value: 5.578e-14

Yes, dropping cls\_prof from the model changed the p-value and estimates of the variables, some greater than others. I also notice that our Adjusted R-sq has increased.

## 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.

The final model, using backward steps to p-hack our way to significance under approx 0.05 is:

#get\_regression\_table(lm(score ~ ethnicity + gender + language + cls\_perc\_eval + cls\_credits + bty\_f1upper + pic\_color,data = evals))  
#arrange(desc(p\_value))  
p.hack.backstep <- lm(score ~ ethnicity + gender + language + cls\_perc\_eval + cls\_credits + bty\_f1upper + pic\_color,data = evals)  
summary(p.hack.backstep)

##   
## Call:  
## lm(formula = score ~ ethnicity + gender + language + cls\_perc\_eval +   
## cls\_credits + bty\_f1upper + pic\_color, data = evals)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.83433 -0.32930 0.09688 0.36927 0.95780   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.385372 0.169659 19.954 < 2e-16 \*\*\*  
## ethnicitynot minority 0.179364 0.075101 2.388 0.0173 \*   
## gendermale 0.207292 0.048848 4.244 2.67e-05 \*\*\*  
## languagenon-english -0.219810 0.103074 -2.133 0.0335 \*   
## cls\_perc\_eval 0.004591 0.001428 3.214 0.0014 \*\*   
## cls\_creditsone credit 0.497623 0.103101 4.827 1.90e-06 \*\*\*  
## bty\_f1upper 0.061983 0.012676 4.890 1.40e-06 \*\*\*  
## pic\_colorcolor -0.184878 0.064919 -2.848 0.0046 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4969 on 455 degrees of freedom  
## Multiple R-squared: 0.1779, Adjusted R-squared: 0.1653   
## F-statistic: 14.07 on 7 and 455 DF, p-value: < 2.2e-16

Alternatively, we can use fastbw() from the rms package to automate this:

library(rms)

## Loading required package: Hmisc

## Loading required package: lattice

##   
## Attaching package: 'lattice'

## The following object is masked from 'package:corrgram':  
##   
## panel.fill

## Loading required package: survival

## Loading required package: Formula

##   
## Attaching package: 'Hmisc'

## The following object is masked from 'package:psych':  
##   
## describe

## The following objects are masked from 'package:dplyr':  
##   
## src, summarize

## The following objects are masked from 'package:base':  
##   
## format.pval, units

## Loading required package: SparseM

##   
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':  
##   
## backsolve

kitchen.sink.2 <- ols(score ~ ., data=evals)  
  
rms.phack.ols <- fastbw(fit=kitchen.sink.2,   
 rule="p",   
 sls=0.5)  
rms.phack.ols

##   
## Deleted Chi-Sq d.f. P Residual d.f. P AIC R2   
## cls\_profs 0.03 1 0.8600 0.03 1 0.8600 -1.97 0.212  
## cls\_level 0.31 1 0.5791 0.34 2 0.8442 -3.66 0.212  
## bty\_m2upper 0.55 1 0.4566 0.89 3 0.8271 -5.11 0.211  
## bty\_m1lower 0.13 1 0.7177 1.02 4 0.9062 -6.98 0.210  
## bty\_m1upper 0.55 1 0.4575 1.58 5 0.9042 -8.42 0.209  
## cls\_students 0.86 1 0.3524 2.44 6 0.8751 -9.56 0.208  
## cls\_did\_eval 0.58 1 0.4464 3.02 7 0.8831 -10.98 0.207  
## pic\_outfit 2.08 1 0.1497 5.10 8 0.7474 -10.90 0.203  
## bty\_f1lower 2.17 1 0.1403 7.27 9 0.6090 -10.73 0.199  
##   
## Approximate Estimates after Deleting Factors  
##   
## Coef S.E. Wald Z P  
## Intercept 3.879399 0.262823 14.761 0.000e+00  
## rank=tenure track -0.176805 0.082231 -2.150 3.155e-02  
## rank=tenured -0.065734 0.063878 -1.029 3.035e-01  
## ethnicity=not minority 0.177498 0.075600 2.348 1.888e-02  
## gender=male 0.254748 0.052426 4.859 1.179e-06  
## language=non-english -0.192999 0.105650 -1.827 6.773e-02  
## age -0.007384 0.003076 -2.401 1.635e-02  
## cls\_perc\_eval 0.004007 0.001432 2.798 5.149e-03  
## cls\_credits=one credit 0.473139 0.110045 4.300 1.712e-05  
## bty\_f1upper 0.076414 0.024786 3.083 2.050e-03  
## bty\_f2upper 0.053085 0.021435 2.477 1.326e-02  
## bty\_avg -0.084013 0.040923 -2.053 4.008e-02  
## pic\_color=color -0.209377 0.067299 -3.111 1.863e-03  
##   
## Factors in Final Model  
##   
## [1] rank ethnicity gender language age   
## [6] cls\_perc\_eval cls\_credits bty\_f1upper bty\_f2upper bty\_avg   
## [11] pic\_color

This results in the same final model with a lot less coding, resulting in less errors.

rms.phack.model <- lm(score ~ ethnicity + language + gender + cls\_perc\_eval + cls\_credits + bty\_f1upper + pic\_color, data = evals)  
summary(rms.phack.model)

##   
## Call:  
## lm(formula = score ~ ethnicity + language + gender + cls\_perc\_eval +   
## cls\_credits + bty\_f1upper + pic\_color, data = evals)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.83433 -0.32930 0.09688 0.36927 0.95780   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.385372 0.169659 19.954 < 2e-16 \*\*\*  
## ethnicitynot minority 0.179364 0.075101 2.388 0.0173 \*   
## languagenon-english -0.219810 0.103074 -2.133 0.0335 \*   
## gendermale 0.207292 0.048848 4.244 2.67e-05 \*\*\*  
## cls\_perc\_eval 0.004591 0.001428 3.214 0.0014 \*\*   
## cls\_creditsone credit 0.497623 0.103101 4.827 1.90e-06 \*\*\*  
## bty\_f1upper 0.061983 0.012676 4.890 1.40e-06 \*\*\*  
## pic\_colorcolor -0.184878 0.064919 -2.848 0.0046 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4969 on 455 degrees of freedom  
## Multiple R-squared: 0.1779, Adjusted R-squared: 0.1653   
## F-statistic: 14.07 on 7 and 455 DF, p-value: < 2.2e-16

But we all know that P-hacking our way to significance is bad practice and can result in over-fitting. Alternatively, we use step() from the stats package to run: *step(kitchen.sink, direction = “backward”)* which gives us a model with the lowest AIC of -642.28 and a higher Adjusted R-squared of 0.1821.

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  
  
Call:  
lm(formula = score ~ rank + ethnicity + gender + language + age +   
 cls\_perc\_eval + cls\_credits + bty\_f1lower + bty\_f1upper +   
 bty\_f2upper + bty\_avg + pic\_outfit + pic\_color, data = evals)

summary(  
 lm(formula = score ~ rank + ethnicity + gender + language + age +   
 cls\_perc\_eval + cls\_credits + bty\_f1lower + bty\_f1upper +   
 bty\_f2upper + bty\_avg + pic\_outfit + pic\_color, data = evals)  
)

##   
## Call:  
## lm(formula = score ~ rank + ethnicity + gender + language + age +   
## cls\_perc\_eval + cls\_credits + bty\_f1lower + bty\_f1upper +   
## bty\_f2upper + bty\_avg + pic\_outfit + pic\_color, 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 \*\*   
## cls\_creditsone credit 0.502732 0.111631 4.504 8.54e-06 \*\*\*  
## bty\_f1lower 0.038838 0.026460 1.468 0.14287   
## bty\_f1upper 0.079899 0.025975 3.076 0.00223 \*\*   
## bty\_f2upper 0.058486 0.021868 2.674 0.00776 \*\*   
## bty\_avg -0.140359 0.057725 -2.432 0.01543 \*   
## pic\_outfitnot formal -0.100300 0.069317 -1.447 0.14860   
## pic\_colorcolor -0.205474 0.067209 -3.057 0.00237 \*\*   
## ---  
## 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

But fine, since we’re using p-hacking, the rest of the code will be based on: lm(score ~ ethnicity + language + cls\_perc\_eval + cls\_credits + bty\_f1upper + pic\_color).

The model is: 3.385 + (ethnicitynot minority) x0.179 + (languagenon-english) x -0.22 + (cls\_perc\_eval) x 0.005 + (cls\_creditsone credit) x 0.498 + (bty\_f1upper) x 0.062 + (pic\_colorcolor) x -0.185

score= +β^0 +β^1×ethnicity\_not\_minority +β^3×language\_non−english +β^4+×class\_perc\_eval +β^5×class\_credits\_one +β6×bty\_f1upper +β7×picture\_color\_colored

## 16

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

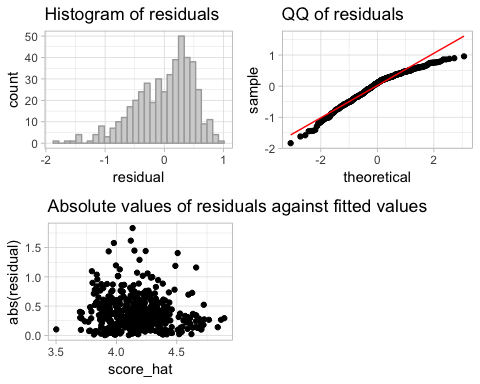
library(cowplot) #gunna need to combine some plots here.

##   
## Attaching package: 'cowplot'

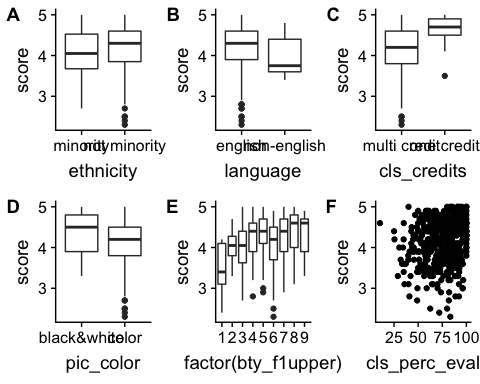
## The following object is masked from 'package:ggplot2':  
##   
## ggsave

p\_hack\_regression\_points <- get\_regression\_points(rms.phack.model)  
  
histo <- p\_hack\_regression\_points %>%   
 ggplot() +   
 theme\_light() +  
 geom\_histogram(aes(residual), fill = "lightgrey", color = "darkgrey") +  
 labs(title = "Histogram of residuals")  
  
qqplot <- p\_hack\_regression\_points %>%   
 ggplot(aes(sample = residual)) +  
 theme\_light() +  
 stat\_qq() +  
 stat\_qq\_line(color = "red") +  
 labs(title = "QQ of residuals")  
  
absolute <- p\_hack\_regression\_points %>%   
 ggplot() +   
 theme\_light() +  
 geom\_point(aes(score\_hat, abs(residual))) +   
 labs(title = "Absolute values of residuals against fitted values")  
  
plot\_grid(histo, qqplot, absolute)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



ethn <- evals %>%   
 ggplot() +  
 geom\_boxplot(aes(ethnicity, score))  
   
lang <- evals %>%   
 ggplot() +  
 geom\_boxplot(aes(language, score))   
  
cls\_cre <- evals %>%   
 ggplot() +  
 geom\_boxplot(aes(cls\_credits, score))   
  
cls\_perc <- evals %>%   
 ggplot() +  
 geom\_point(aes(cls\_perc\_eval, score))  
  
pic\_col <- evals %>%   
 ggplot() +  
 geom\_boxplot(aes(pic\_color, score))   
  
bty\_f1 <- evals %>%   
 ggplot() +  
 geom\_boxplot(aes(factor(bty\_f1upper), score))   
  
plot\_grid(ethn, lang, cls\_cre, pic\_col, bty\_f1, cls\_perc, labels = c("A", "B", "C", "D", "E", "F"))



\* The residuals of the model is nearly normal as visible in the histogram and QQ plot; with a left skew.  
\* Fitted vs absolute values are close.   
\* We assume independence based on study method.  
\* Boxplots show linearity related to categorical variables (bty\_f1upper is a bit off, but it's nearly linear, which is ok).   
\* Cls\_perc\_eval is a fuzzy scatter plot, and not very linear, but

## 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?

Not likely. If the same course is taught by the same professor, the courses are independent of one another and thus evaluations are independent. That said, a worthwhile variable may be something that relates a student having had the instructor previously; or a longevity study for subset of teachers over time (something that might be scrubbed from ratemyprofessors.com).

## 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.

A non minority male, who speaks English, teaches a one credit class, has a picture in black & white, has high ratings from upper level *females* and has a high percent of students who evaluated. This makes me wonder if females drive the beauty scoring, or if female students drive evaluation scores in general.

## 19

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

No, I most definitely would not as I can think of at least a few additional variables that might have an impact the scores of instructors outside predominantly white / male Austin, Texas. For example, per-capita of minorities in a university city might have an impact on the scores of minority professors. The per-capita *gender of students* may play an impact of scores for the gender of professors. Just these two alone give me reason to pause. The sampling of professors from Austin is not representative of the country as a whole; and most *certainly* the students at a university in Austin are not representative of the country as a whole!