Do Smaller Classes Improve Test Scores? Evidence from a Regression Discontinuity Design

Ben Warren

5/28/2022

When investigating the impact of class size on students' academic success, simply comparing smaller and larger classes likely does not capture the true effect. Small classes would be correlated with the amount of funding that a school receives (not to mention the fact that private schools have on average 6 fewer students per class than public schools in the U.S.).

This would mean that a simple comparison might just pick up the effects of household wealth, since wealthier students are more likely to attend private school or reside in districts with more funding. The comparison would likely be biased upwards relative to the true causal effect. If smaller classes indeed correlate to higher test scores, that effect would compound with the effects of household wealth and school funding to create a more significant impact when considering just a simple comparison.

The Tennessee STAR experiment sought to overcome this issue by focusing on what they called "class effects," or the sum effect of a student being assigned to a specific teacher and classroom (this, of course, was the only variable the experiment could truly control). They tested whether outcome variables differed more across the classes than would be predicted by random variance. Through this analysis, they rejected the null hypothesis that there was no class effect on earnings at age 27. The standard deviation of the difference between classes was up to 10% of the income at that age.

One way to look at the potential difference between outcomes between class sizes is to construct a binned scatterplot, an alternative to a traditional scatterplot that breaks data down into "bins" or smaller quantiles of a pre-determined amount and plots them based on a mean characteristic. We can use these observations to begin a regression discontinuity analysis.

The following binned scatterplot of class sizes and school enrollment shows that for fifth-grade classes at public schools in Israel, class sizes display a discontinuity at 40 students in a school. For school enrollments less than 40 and greater than 40, separately, a linear model seems appropriate.
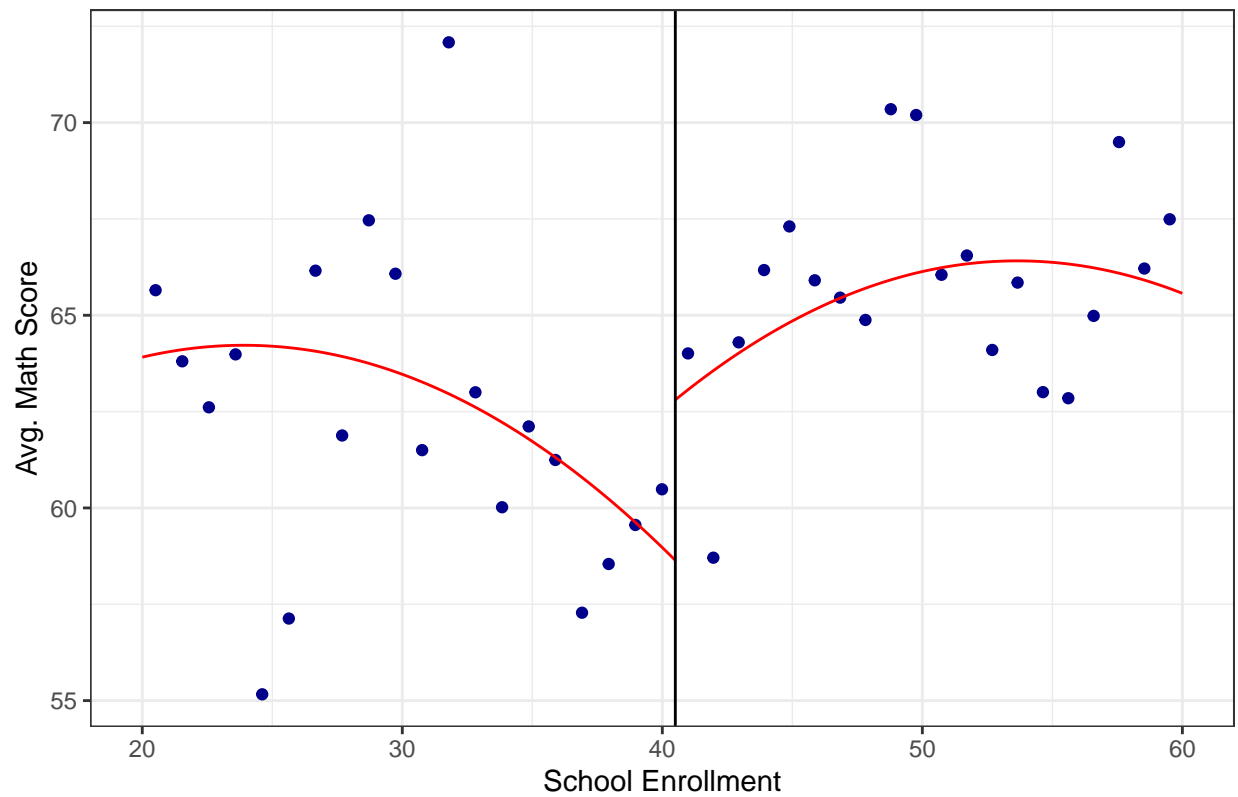
```
## [1] "Mass points detected in the running variable."
```

RD Plot
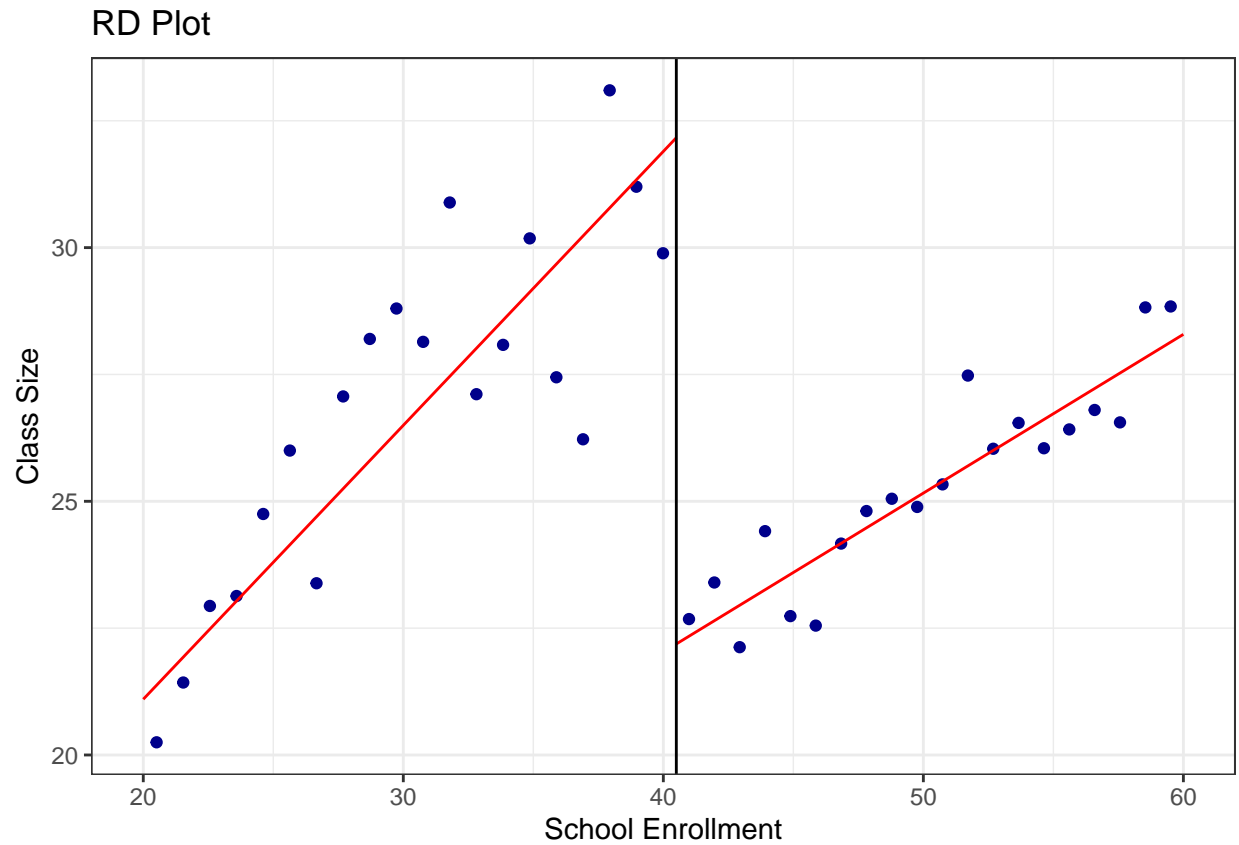
A similar analysis shows that a quadratic fit is appropriate for modeling math scores in RD.

```
## [1] "Mass points detected in the running variable."
```
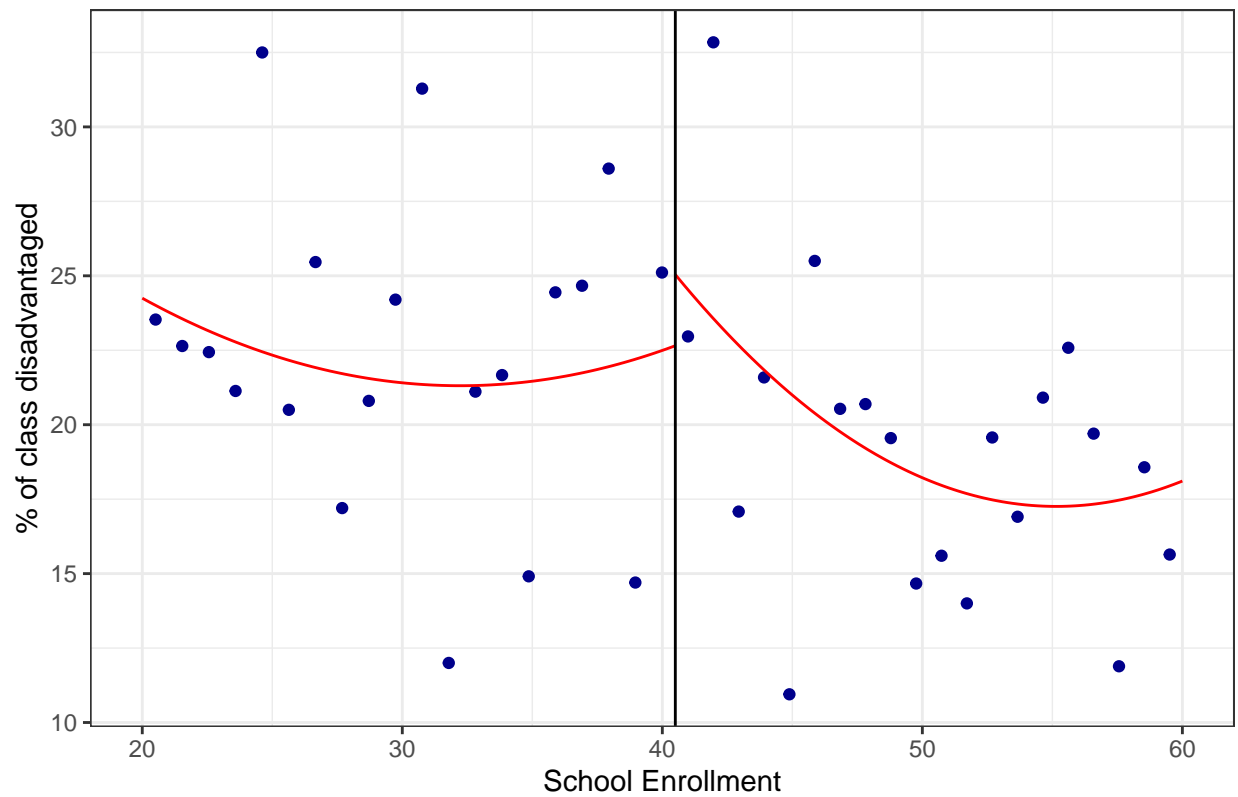
Avg. Math Scores vs. School Enrollment

There is a vague quadratic relationship between the % of a school disadvantaged and the school enrollment, but it is weak.
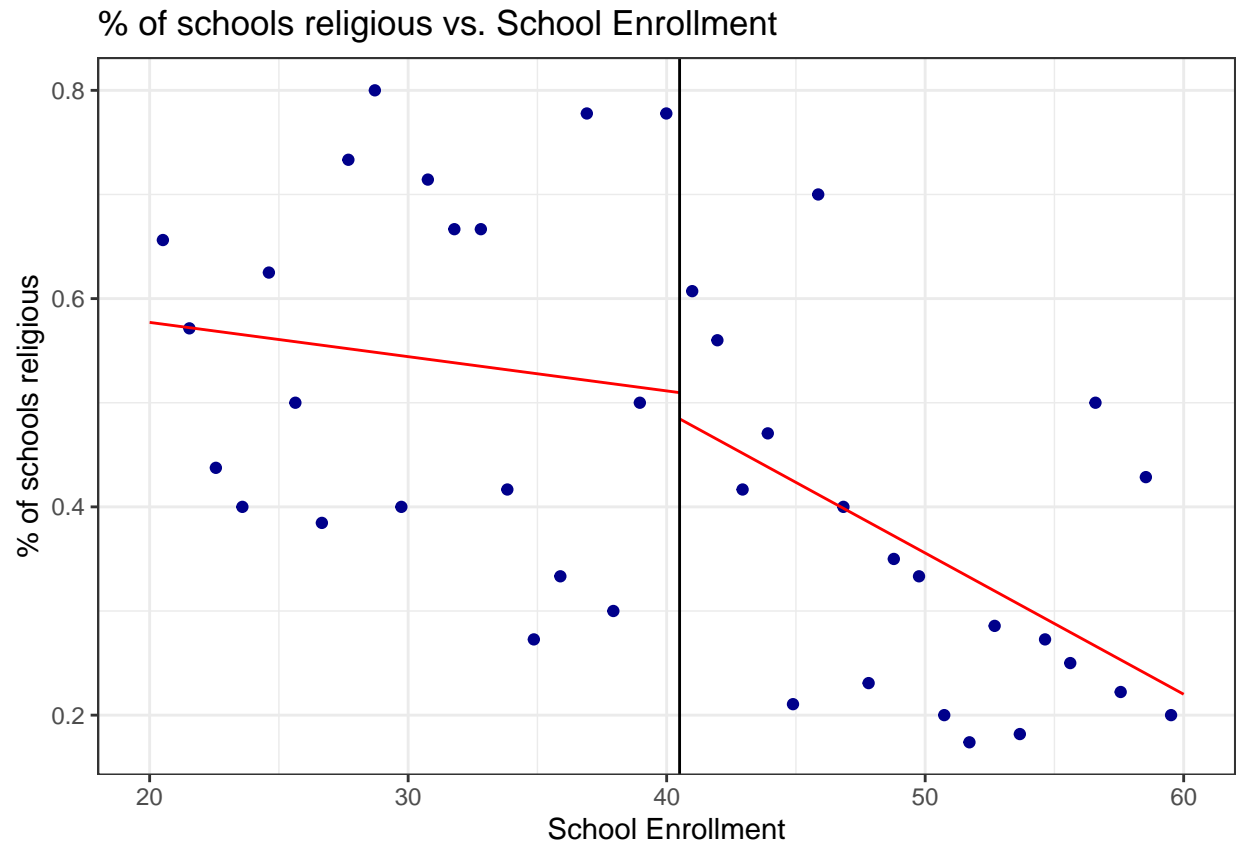
```
## [1] "Mass points detected in the running variable."
```

Disadvantaged vs. School Enrollment

There appears to be a linear relationship between school enrollment and the % of the school that is religious.

```
## [1] "Mass points detected in the running variable."
```

## % of schools religious vs. School Enrollment

There does not appear to be a relationship between the proportion of the students that are girls and school enrollment.

```
## [1] "Mass points detected in the running variable."
```

% of Students Female vs. School Enrollment
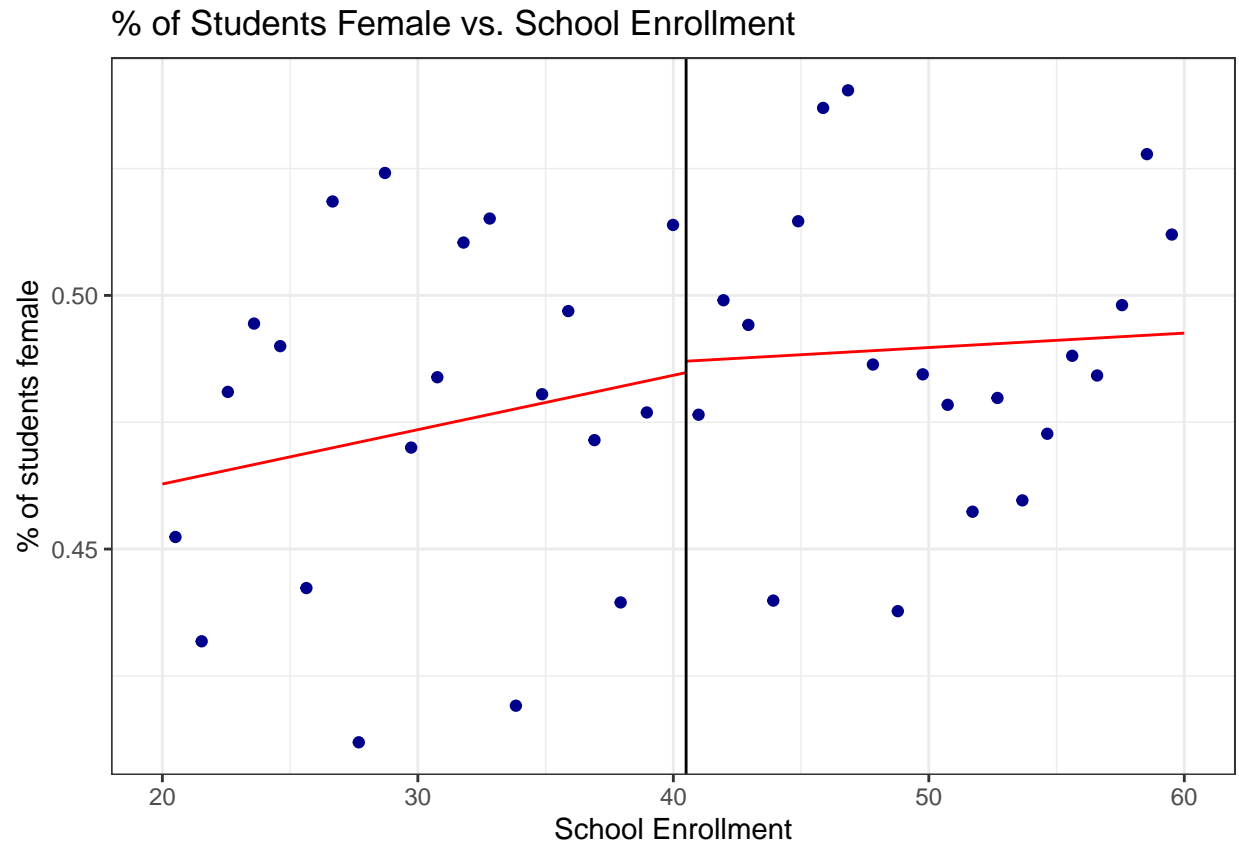
A histogram of schools by enrollment numbers shows that the distribution is fairly uniform, but with modes at the low end of the 40-students-and-under group and the high end of the over-40 students group.
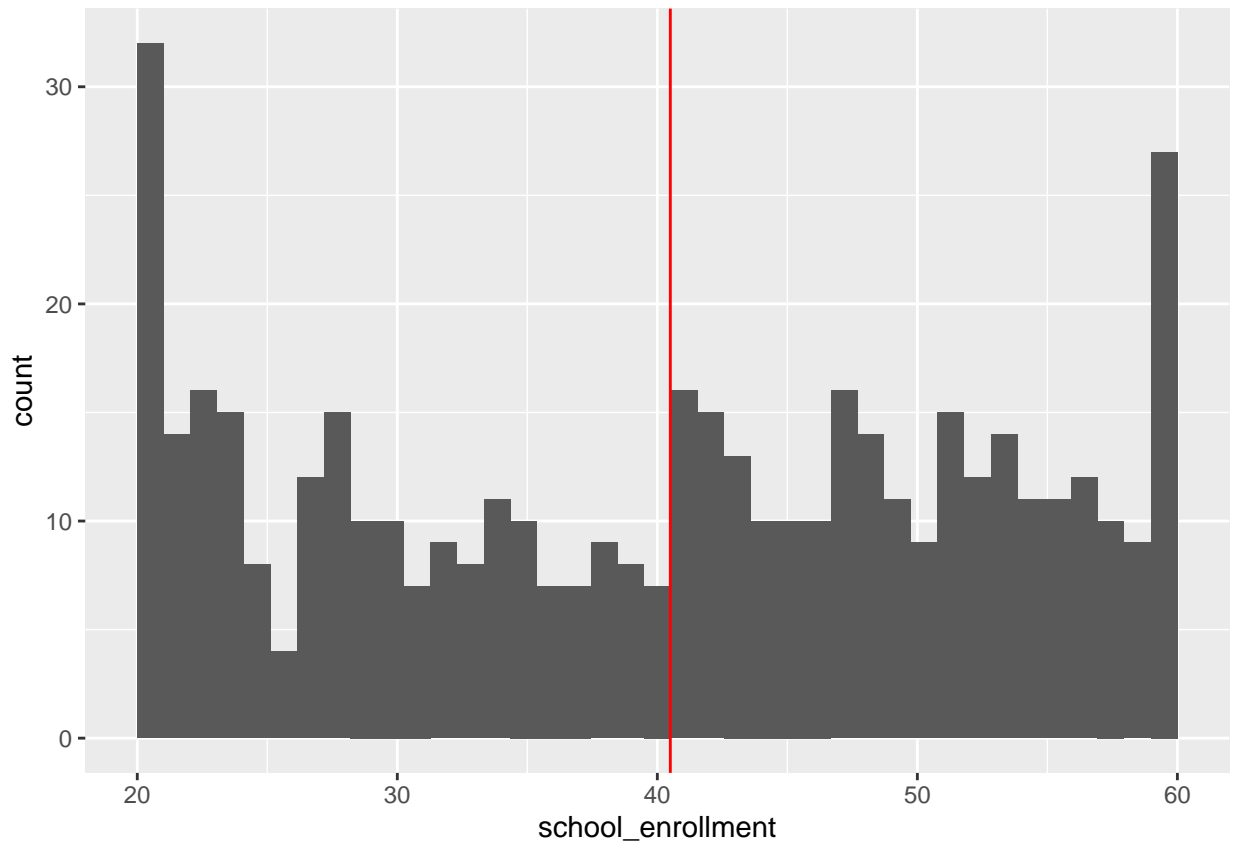
Informed by observations from the graphs, we can now run regressions to quantify the relationship between school enrollment and the other variables.

A regression run on class size returns a strong model utility test and a fairly high R-squared value of over 0.5, which shows that school enrollment appears to have a strong relationship with class size.

```
##
## Call:
## lm(formula = classize ~ above40 + x + x_above, data = narrow)
##
## Residuals:
##       Min      1Q   Median      3Q      Max
## -20.6288  -2.0062   0.4046   2.6970  21.2934
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.04583    0.51468  64.207   <2e-16 ***
## above40     -11.00075    0.59509 -18.486   <2e-16 ***
## x            0.67252    0.03172  21.204   <2e-16 ***
## x_above     -0.34178    0.03424  -9.981   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.47 on 1176 degrees of freedom
## Multiple R-squared:  0.5411, Adjusted R-squared:  0.5399
## F-statistic: 462.2 on 3 and 1176 DF,  p-value: < 2.2e-16
```

7

```
## $vcov
##              (Intercept)     above40           x      x_above
## (Intercept)    1.3041391 -1.30413910  0.06475370 -0.064753704
## above40       -1.3041391  1.41848159 -0.06475370  0.058886477
## x              0.0647537 -0.06475370  0.00329797 -0.003297970
## x_above       -0.0647537  0.05888648 -0.00329797  0.003682584
##
## $dof
## (Intercept)     above40           x      x_above
##    98.19551   166.99029   117.51752   157.47507
##
## $adj.se
## (Intercept)     above40           x      x_above
##  1.15623747  1.19969459  0.05802547  0.06115426
##
## $se
## (Intercept)     above40           x      x_above
##  1.14198910  1.19100025  0.05742796  0.06068429
##
## $se.Stata
## (Intercept)     above40           x      x_above
##  1.13377222  1.18299036  0.05699266  0.06025655
```

The relationship between average math scores and higher school enrollment is tenuous at most. The model utility test returns a p-value less than 0.5, but the R-squared value is just 0.03, indicating that though higher school enrollment may help explain some variation in math scores, that connection is small.

```
##
## Call:
## lm(formula = avgmath ~ above40 + x + x_above, data = narrow)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.546  -6.311   0.720   6.795  29.937
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 60.82845    1.14174  53.277  < 2e-16 ***
## above40      3.42890    1.32012   2.597  0.00951 **
## x           -0.19263    0.07036  -2.738  0.00628 **
## x_above      0.29810    0.07596   3.924 9.2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.916 on 1176 degrees of freedom
## Multiple R-squared:  0.03229,    Adjusted R-squared:  0.02982
## F-statistic: 13.08 on 3 and 1176 DF,  p-value: 2.098e-08


## $vcov
##              (Intercept)     above40           x      x_above
## (Intercept)   1.88173532 -1.88173532  0.099010642 -0.099010642
## above40      -1.88173532  2.61686061 -0.099010642  0.070952753
## x             0.09901064 -0.09901064  0.006750153 -0.006750153
```

```
## x_above      -0.09901064   0.07095275  -0.006750153   0.008167906
##
## $dof
## (Intercept)      above40            x       x_above
##     98.19551    166.99029    117.51752    157.47507
##
## $adj.se
## (Intercept)      above40            x       x_above
##   1.38887881   1.62948042   0.08301415   0.09107639
##
## $se
## (Intercept)      above40            x       x_above
##   1.37176358   1.61767136   0.08215932   0.09037647
##
## $se.Stata
## (Intercept)      above40            x       x_above
##   1.36579952   1.61214280   0.08177869   0.09000218
```

Similar to math scores, verbal test scores show a weak relationship with high school enrollment, with an R-squared value of just over 0.01.

```
##
## Call:
## lm(formula = avgverb ~ above40 + x + x_above, data = narrow)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.622  -5.404   0.915   6.226  18.840
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 69.82073    0.96571  72.300  < 2e-16 ***
## above40      2.62813    1.11658   2.354 0.018750 *
## x           -0.16438    0.05951  -2.762 0.005831 **
## x_above      0.22585    0.06425   3.515 0.000456 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.387 on 1176 degrees of freedom
## Multiple R-squared:  0.01811,    Adjusted R-squared:  0.01561
## F-statistic: 7.231 on 3 and 1176 DF,  p-value: 8.267e-05


## $vcov
##              (Intercept)      above40            x       x_above
## (Intercept)   1.70274368  -1.70274368   0.09065305  -0.090653051
## above40      -1.70274368   2.25982553  -0.09065305   0.070503311
## x             0.09065305  -0.09065305   0.00603460  -0.006034600
## x_above      -0.09065305   0.07050331  -0.00603460   0.006963421
##
## $dof
## (Intercept)      above40            x       x_above
##     98.19551    166.99029    117.51752    157.47507
##
```

```
## $adj.se
## (Intercept)      above40            x      x_above
##  1.32117309   1.51424555   0.07849094   0.08409337
##
## $se
## (Intercept)      above40            x      x_above
##  1.30489221   1.50327161   0.07768269   0.08344712
##
## $se.Stata
## (Intercept)      above40            x      x_above
##  1.29814223   1.49704755   0.07726855   0.08304460
```

The identification assumption for regression discontinuity design is that a) there is no other discontinuity except the variable in question and b) the agents involved in the experiment cannot influence the discontinuity.

In this quasi-experiment, these assumptions appear to hold.

If schools followed the class size rule described in Angrist and Lavy (1999) - that once schools reach 40 students, they should split students into multiple classes - class size should roughly half on average at 40 students. Since it is likely the case that not all schools follow this rule directly, the true effect is less than this.

This phenomenon is evident when considering a plot of average class size versus school enrollment. Though the data appear to be linear in various ranges throughout the chart, a linear model considering the whole data set does not fit.

Avg. Class Size vs. School Enrollment
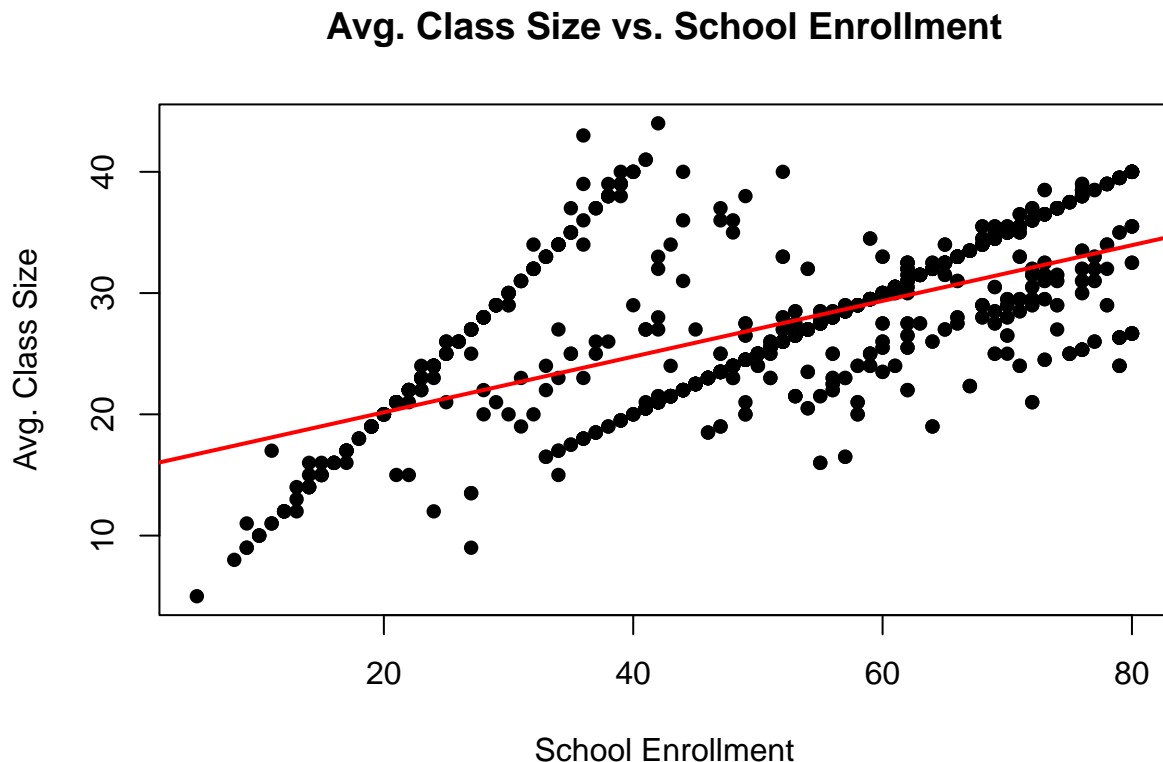
When we see split the data at school enrollment = 40, the pattern is more clear. Still, it appears that some

10

schools make the switch to more, smaller classes before they reach 40 students. This further obfuscates the relationship.

Avg. Class Size vs. School Enrollment

Avg. Class Size vs. School Enrollment
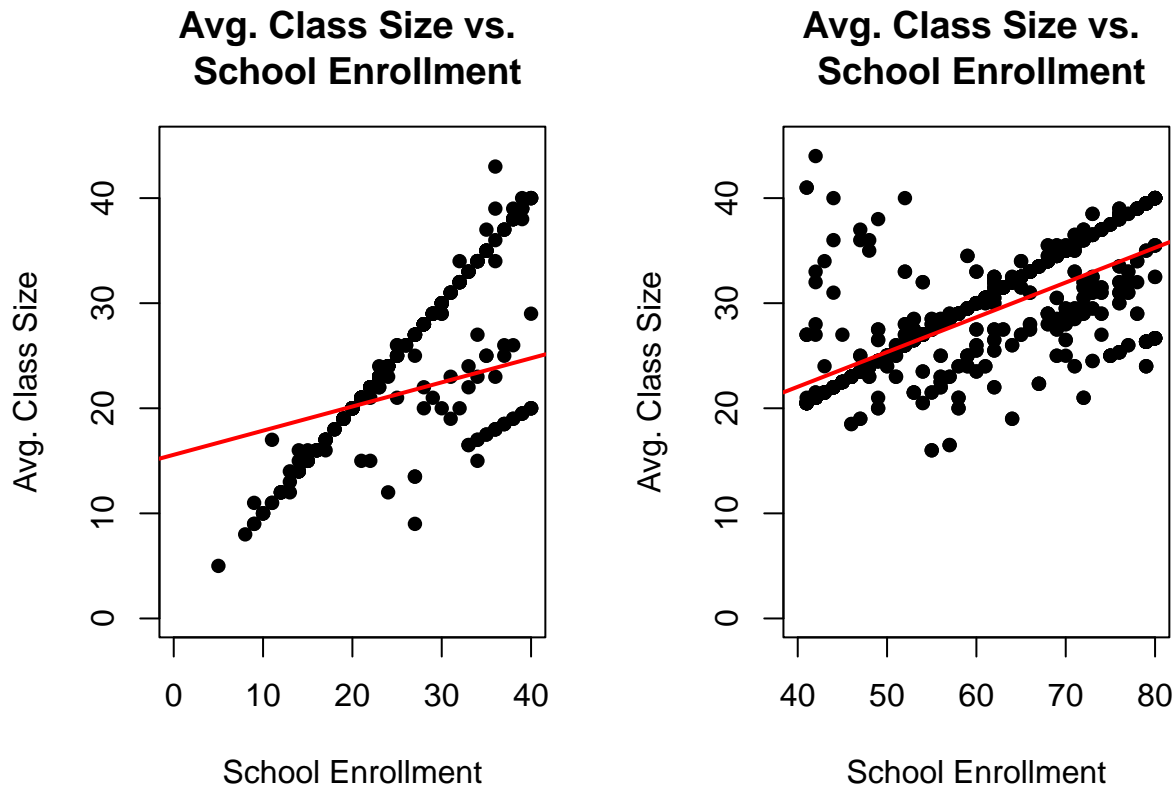
If all schools followed the class size rule as it is stated in Angrist and Lavy (1999), then we would see an immediate drop to an average of 20.5 students/class upon the school enrolling 41 students. In our data, it seems like the discontinuity begins to appear more around the 30 student mark, but it varies significantly by school. In any case, the average doesn't appear to drop as drastically as the rule would suggest.

Question 8: Suppose your school superintendent is considering a reform to reduce class sizes in your school from 40 to 35. Use your estimates above to predict the change in math and verbal test scores that would result from this reform.

If we want to understand how student outcomes would change if they were in smaller classes, we can divide the RD estimate of the change in test scores by the change in number of students per class at the threshold.

First, we have to calculate the average class size for schools with enrollments of 40 students and those with enrollments of 41.

The estimated discontinuity from the RD analysis was 3.4289 for average math score and 2.62813 for average verbal score. When school enrollment increased from 40 to 41 (where we measured the discontinuity), average class size decreased from 29.89 students/class to 22.68 students/class, or by 7.21 students/class. The proposed students/class decrease is 5 students, from 40 to 35. Thus, we can estimate that the change in average math score per decrease in student/class is 3.4289/7.21 = 0.48 and the change in average verbal score is 2.62813/7.21 = 0.36.

With this in mind, we can predict that the change in students/class from 40 to 35 (a 5-student decrease) will result in a change in average math score of 0.48*5 = 2.4 and a change in average verbal score of 0.36*5 = 1.8.

However, this phenomenon may not hold for other 5-student decreases. Intuitively, smaller classes should still increase test scores, but we have to test that on the data.

```
## [1] Verbal Mean, class size between 12.5 and 17.5:
```

```
## [1] 72.58139
```

```
## [1] Verbal Mean, class size between 17.5 and 22.5:
```

```
## [1] 71.23868
```

```
## [1] Math Mean, class size between 12.5 and 17.5:
```

```
## [1] 64.72139
```

```
## [1] Math Mean, class size between 17.5 and 22.5:

## [1] 62.71374

## [1] Math Median, class size between 12.5 and 17.5:

## [1] 64.6

## [1] Math Median, class size between 17.5 and 22.5:

## [1] 63.83

## [1] Verbal Median, class size between 12.5 and 17.5:

## [1] 73.27

## [1] Verbal Median, class size between 17.5 and 22.5:

## [1] 72.39
```
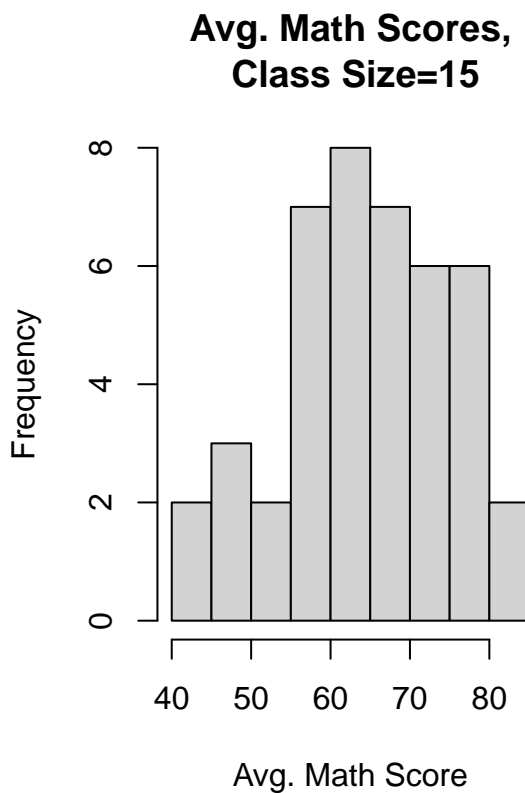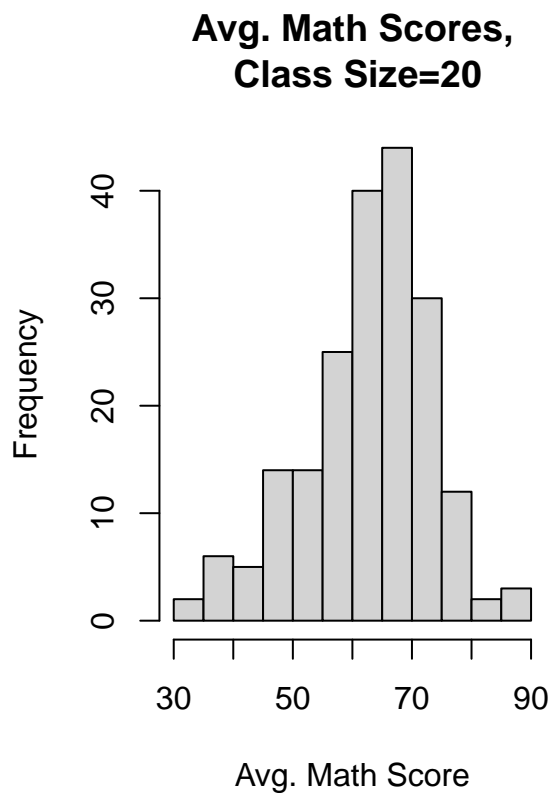
Avg. Math Scores, Class Size=20

Avg. Math Scores, Class Size=15

The mean math scores for schools with class sizes of exactly 20 were higher than those with exactly 15 students/class by 2.6 points. But if we expand this to the *range* of schools between class sizes of 12.5 - 17.5 vs. 17.5-22.5, the mean math scores in the lower range are on average 2 points higher than the upper range.

The histogram above shows that there is a larger range of average scores for schools in the upper range, but the median average score for schools in the upper range is 63.83, while the median for the lower range is 64.6. This would suggest that though a specific change from 20 students to 15 students didn't appear to have a positive impact on test scores, moving from the upper range (17.5 to 22.5 students/class) to the lower range (12.5 to 17.5 students/class) had a slight positive impact on test scores.

This phenomenon holds for verbal scores as well.
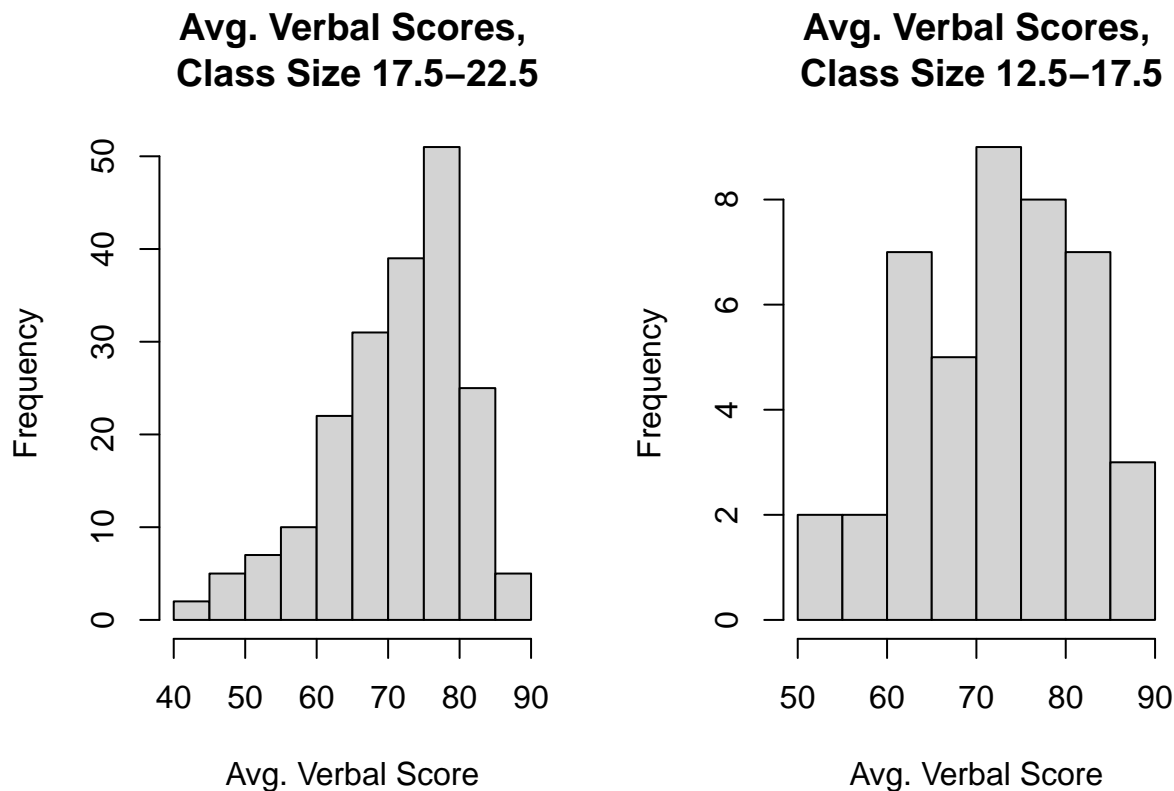
Based on this study, you can expect a small increase in average and median math and verbal scores by moving from the 17.5-22.5 students/class range to the 12.5-17.5 students/class range. I would be less confident repeating the same prediction I made to the school moving from 40-35 students/class, but I would still be fairly confident in saying that a decrease in students/class would still have a positive impact, albeit a smaller one.

We've seen that higher scores in kindergarten have been shown to improve outcomes in the long run, so this is still a valuable effect to seek out. This may mean that a district would be best-served to minimize the class sizes in its kindergarten classes, but not worry so much about class size in the later grades.