

Exploring and Visualizing School Achievement and School Effects

Daniel Anderson¹ & Joseph Stevens¹

¹ University of Oregon

Author Note

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324C110004, awarded to the University of Oregon. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Daniel Anderson, 5262 University of Oregon. E-mail: daniela@uoregon.edu

Abstract

Recent years have seen a push for more open data as a means of facilitating transparency and opening doors for further research. Since the implementation of the No Child Left Behind Act, schools and districts across the country have been required to publicly report the number of students scoring in each performance level classification on the statewide test, disaggregated by student subgroups (e.g., gender, race/ethnicity). These data represent a “coarsening” of the underlying continuous data by binning the distributions into categories. The purpose of this paper is to evaluate the extent to which achievement gap estimates from student-level data align with those estimates from coarsened data, using the approach suggested by Ho and Reardon (2012). Following the evaluation, we use this approach to visualize differences in school-level achievement gaps using publicly-available data, and visualize these estimates geographically using maps. We overlay demographic data as a means to visually evaluate the extent to which demographics of the surrounding area correspond to school-level achievement gaps.

Keywords: Effect Size, Achievement Gaps, Visualization, Maps

Word count: X

Exploring and Visualizing School Achievement and School Effects

Recent years have seen increased attention on open data (i.e., free and publicly available) as one component of open science (e.g., see Dawes, Vidiasova, & Parkhimovich, 2016; Sieber & Johnson, 2015; Zuiderwijk & Janssen, 2014). The National Institute of Health (NIH), for example, recently made 64 data sharing repositories publicly available (Kittrie & Chattopadhyay, 2016). The National Science Foundation (NSF) has placed a similarly high emphasis on making data from funded projects publicly available, as part of their Open Government Plan (National Science Foundation, 2016). The Institute of Education Sciences (IES) has also moved towards greater requirements for data sharing, with Goal 4 Effectiveness grants requiring a data sharing component since 2013, and Goal 3 Efficacy and Replication awards requiring data sharing since 2016 (Institute of Education Sciences, 2016). Despite these movements, however, data that are fully open remain rare, particularly in educational research. For example, the website data.gov represents the “home of the U.S. Government’s open data”, housing nearly a quarter of a million datasets. Of these, only 365 were tagged with the keyword “Education” as of this writing, and the vast majority of these receive an openness score of one star out of a possible five.

Privacy issues abound in educational research, given that the research population generally includes minors. Privacy concerns may therefore explain part of the lack of data sharing (although it should be noted that health care and medical research often face similar privacy concerns). One method of avoiding potential issues of confidentiality is to first aggregate the data in some capacity, before releasing the data publicly. This was the approach taken by the *No Child Left Behind Act* (No Child Left Behind, 2002), with data reported at the overall school and district levels, but not at the individual student level. NCLB required annual testing in reading and mathematics in each of grades 3-8, and once in high school, with the results reported at the aggregate levels by student subgroups (e.g., gender, race/ethnicity). Cut scores were required to be established, delineating students into various performance level classifications (PLCs). The state of California, for example,

included the following four categories in order of increasing competency: “Not Met”, “Nearly Met”, “Met”, and “Exceeds”. The proportion of students scoring in each of these categories, by subgroup, were also required to be reported at each of the school and district levels by NCLB, provided a sufficient number of students were represented within each subgroup, such that individual students could not be identified.

Evaluating Achievement Gaps With Open Data

Aggregating the data to higher levels, combined with reporting test score data by PLC proportions, helps avoid privacy concerns and allows the data to be reported publicly. Despite mandates from NCLB on public reporting, however, the openness of these data vary by state, in terms of accessing the data for research purposes. Some (e.g., Oregon, California) make the process easy, with downloadable comma separated value (CSV) files including the proportion of students scoring in each PLC for all schools and/or districts in the state, by student subgroups. Other states make it relatively easy to access the data for an individual school or district (as mandated by NCLB) but rather difficult to access the data across schools or districts. Further, the utility of these data for research purposes is not particularly straightforward. While raw percentages scoring within specific PLCs can be compared between schools or between student subgroups, these differences are highly dependent upon the placement of the cut scores (Ho, 2008; Holland, 2002).

Ideally, a transformation could be applied to the “coarsened” data (i.e., continuous test score distributions binned into a set of ordinal categories; Reardon & Ho, 2015) to recover the parameters of the full distribution for each subgroup, making the placement of the cut scores inconsequential to evaluating subgroup differences. Ho and Reardon (2012) and Reardon and Ho (2015) propose such an approach, in which the coarsened data are used to construct empirical cumulative distribution functions (ECDFs) for each group. Paired ECDFs for any two groups are then evaluated in the form of a probability-probability (PP) plot, and the area under the PP curve (AUC) provides an estimate of the probability that a randomly selected student from the reference distribution (plotted along the x-axis) would

score higher than a randomly selected student from the focal distribution (plotted along the y-axis; see Ho & Reardon, 2012). This proportion can then be transformed into standard-deviation units, using the following formula, as outlined by Ho (2009)

$$V = \sqrt{2}\Phi^{-1}(AUC) \quad (1)$$

Where Φ^{-1} represents the inverse normal distribution. Under the assumption of respective normality, V is equivalent to Cohen's D , and the dependence upon cut scores is removed, given that it is an estimate of the difference between the paired distributions, rather than the difference between proportions at any particular point on the distribution. Importantly, because V can be estimated from coarsened data, it can be used to estimate the overall achievement gap between student subgroups using data that are mandatorily reported as part of NCLB.

Achievement Gaps in Context

A wealth of previous research has examined achievement differences between student subgroups. Sirin (2005), for example, conducted a meta-analytic review of the relation between socioeconomic status (SES) and academic achievement. The author found that, while there was a medium to strong overall relation, the magnitude was dependent upon (among other factors) the physical location of the school. This is perhaps unsurprising, given that schools draw students from the surrounding neighborhoods, and any disparities between neighborhoods would naturally flow into disparities between schools. When evaluating achievement gaps, it makes sense to consider characteristics of both the school and the surrounding neighborhood/community. For example, it is possible that the racial/ethnic makeup of the surrounding area would relate to achievement gaps at the school, given that identifying as Black in a school with students predominately identifying as White is very different from identifying as Black in a school where students predominately identify as Black (Hanushek, Kain, & Rivkin, 2009).

In addition to school compositional effects, there is considerable evidence that crime

rates in the surrounding area relate to student achievement (e.g., Gonzales, Cauce, Friedman, & Mason, 1996; McCoy, Roy, & Sirkman, 2013). In particular, N. K. Bowen and Bowen (1999) found that measures of neighborhood danger were predictive of a number of outcomes, but particularly measures of behavior and attendance. The authors found that males and students identifying as African American reported higher rates of exposure to both neighborhood and school danger. These results are particularly relevant for achievement gaps when considering Lee and Madyun (2009) found that “Black students were 2.88 times more likely than White peers to reside in the neighborhoods with high crime-high poverty” (p. 165). Further, as noted by Gregory, Skiba, and Noguera (2010), large racial disproportionality exists in terms of school-wide suspension and expulsion rates. This disproportionality may be partially attributable to neighborhood/community differences, which ultimately lead to fewer opportunities for students to learn (i.e., more time spent out of the classroom) and a broadening of the achievement gap.

Summary

The purpose of this paper is to evaluate differences in achievement gaps between schools using open, publicly available data, reported by schools on a mandatory basis as part of NCLB. We begin, however, by first evaluating V using empirical, student-level data collected across the state of Oregon. Specifically, we evaluate the extent to which V corresponds with D when using the full sample. We expect a strong but imperfect relation, given the differing assumption of the two effect sizes. We then manually coarsen the data into the proportion of students scoring in each PLC, re-estimate V , and compare the estimates to those made from the continuous data. Following this investigation, we illustrate how this approach can be useful when working with publicly available data, specifically by using V to calculate achievement gap effect sizes for schools in California and Oregon, combining these data with census data, and producing geo-spatial maps to evaluate both the clustering of school-level estimates of achievement gaps, and the extent to which any clustering relates to the demographics of the surrounding area (e.g., median housing cost).

The following research questions were addressed:

1. To what extent does V estimated with the full data correspond with V estimated from the coarsened data, and to what extent do these estimates correspond with D ?
2. What does the distribution of school-level achievement gaps look like (i.e., shape and variance) in California and Oregon?
3. Do school-level achievement gap estimates tend to be geographically clustered, and if so, does this clustering relate to demographic characteristics of the surrounding area?

Method

Data Sources

Multiple sources of data were used. Student-level data included all students in the state of Oregon who took the reading/language arts or mathematics assessments during the 2012-2013 school year. These data are summarized by content area and race/ethnicity in Table 1, and were collected as part of the National Center on Assessment and Accountability for Special Education (NCAASE; see <http://ncaase.com>), a multi-state collaborative focused on growth modeling and evaluating school effect policies. These data were used to evaluate the correspondence between V and D estimated from the full data, as well as the extent to which V estimated with the discrete (coarsened) data, V_d , corresponded with V estimated from the full, continuous data, V_c .

Publicly-available datasets were used to visualize differences in achievement gaps in both Oregon and California. For both states, data on the proportion of students in each subgroup scoring in each PLC were obtained from the corresponding statewide websites (California Assessment of Student Performance and Progress, 2017; Oregon Department of Education, 2017a). Both states had minimum reporting requirements, although the threshold varied considerably. In California, the minimum reported sample size across student groups was 100, while in Oregon the minimum sample size was only 6. Values below these thresholds in each respective data file were missing.

To map the schools, information on the physical location of the schools, in terms of latitude and longitude, were also necessary. For California, a data file containing this information for all schools in the state was located on the state website (California Department of Education, 2017). For Oregon¹, no specific information on the latitude and longitude of schools could be located. However, a file containing the school name and physical address was located (Oregon Department of Education, 2017b). These addresses were then transformed to latitude and longitude using Google's geocode application programming interface (API; Google, 2018). Geographic information were combined with census data. Specifically, we obtained data on (a) the median housing cost, (b) the number of individuals identifying as Black (for California) or Hispanic/Latino (for Oregon), and (c) the number of individuals with an income to poverty line ratio above 2.0 within each county's census tracts, obtained from the 5-year American Community Survey (ACS) - data collected as part of the nationwide census (United States Census Bureau, 2016).

Analyses

All analyses were conducted within the R statistical computing environment (R Core Team, 2017). The *tidyverse* suite of packages (Wickham, 2017) were used for all data preparation and visualization, along with *patchwork* to combine multiple plots (Pedersen, 2017). Effect sizes were estimated using the *esvis* package (Anderson, 2018), and maps were produced using a combination of the *leaflet* (Cheng, Karambelkar, & Xie, 2017) and *tidycensus* (Walker, 2018) packages. Many of the publicly available datasets were read into R directly from the web using the *rio* package (C.-h. Chan, Chan, Leeper, & Becker, 2018). Finally, in the spirit of open and reproducible research, all the code used to produce this manuscript are publicly available in the form of a GitHub repository². The document itself was produced using the *papaja* package, which is an R Markdown extension for rendering manuscripts formatted according to the standards of the American Psychological Association

¹Although preliminary maps were produced for Oregon, we share only the maps from California in this manuscript. We hope to share maps of both Oregon and California during the conference presentation.

²see https://github.com/DJAnderson07/ncme_18

(see Aust & Barth, 2018).

Effect size comparison. Student-level data from Oregon were used to empirically estimate achievement gap effect sizes by school. Across all effect size analyses, we pooled data across grades to estimate a single effect size, rather than estimating separate effects by grade. For the purposes of this investigation, we evaluated achievement gaps between students identifying as Hispanic/Latino and students identifying as White for all schools that met the minimum reporting size for both groups. Cohen's D was estimated as

$$d = \frac{\bar{X}_{foc} - \bar{X}_{ref}}{\sqrt{\frac{(n_{foc}-1)\sigma_{foc} + (n_{ref}-1)\sigma_{ref}}{n_{foc}+n_{ref}-2}}} \quad (2)$$

where foc represents the focal group, in this case students identifying as Hispanic/Latino, and ref represents the reference group (students identifying as White). The numerator represents the difference in the means of the two distributions, while the denominator represents the pooled standard deviation. The V_c effect size was then estimated with the same data, and for the same set of schools. Comparisons between V_c and D were assessed both in terms of the correlation between the measures, as well as in terms of how discrepant V_c was from D .

Following the evaluation of V_c with D , we collapsed the data into counts by PLC within each school - i.e., we manually coarsened the data. We then estimated V_d (V estimated from discrete data) and compared these estimates with both V_c and D . As stated previously, we expected V_c and D to differ marginally, given the different assumptions of the estimators. However, any differences in V_c and V_d could be interpreted as differences that arose due to the coarsening of the data.

Evaluating school-level achievement gap estimates. Publicly available data reporting on the percentage of students scoring in each PLC were obtain for California and Oregon, as described previously. Both files reported percentages by grade and across grades, by student subgroups. Similar to the student-level analyses, we used the pooled percentages across grades, rather than conducting separate analyses by grade. The Black-White

achievement gap was estimated for California, while the Hispanic-White achievement gap was estimated for Oregon. In both cases, schools were only included if they reported percentages for both groups (i.e., met the minimal reporting threshold).

State level evaluations were followed up with county-level investigations using geographic and census data. Census tracts represent geographic areas with between approximately 1,200 to 8,000 people, with an optimum size of 4,000 people. Geographic boundaries of the census tracts were also obtained from the ACS survey. Maps were then produced with the census tracts overlayed and colored according to the respective variable. Schools were displayed as “pins” on the map, and were colored according to the magnitude of the estimated achievement gap. The relation between achievement gaps at the school level and demographic characteristics of the surrounding area were then visually examined. All maps were interactive, helping to facilitate exploration (e.g., zoom, pan). In our results, we share screen captures from these maps, although it is recognized that some of the features are difficult to portray through static images.

Results

Effect Size Comparisons

Figure 1 displays the relation between V_c and D both in terms of the bivariate relation and the distribution of differences with respect to V_c for both mathematics and reading/language arts. For the bivariate relation, the x-axis was restricted to span -2 to 1.5, which resulted in the removal of one outlier point (which had little effect on the overall trend, but nonetheless made the bulk of the data more difficult to see). Topographical lines are displayed on top of the points to help display the density of the data, given that many points overlapped. The majority of the estimates were negative, as measured by either V_c or D , indicating that students identifying as Hispanic/Latino performed, on average, lower than students identifying as White in most schools across Oregon. Note that there were a few schools for both content areas, however, where the achievement gap was reversed (i.e., students identifying as White scored, on average, lower than students identifying as

Hispanic/Latino). Overall, V_c and D correlated at 0.87 for mathematics and 0.86 for reading/language arts. For the histograms, the vertical line represents the mean of the differences for the corresponding content area. Note that the mean for each content area was slightly negative ($M_{mth} = -0.12$, $M_{rdg} = -0.16$), indicating that, on average for this sample, V_c represented a slightly lower estimate of the effect size (less positive/more negative) relative to D . The distributions had a standard deviation of 0.15 and 0.16 for mathematics and reading/language arts respectively.

Figure 2 displays the bivariate relation and differences between V_d estimated with discrete (coarsened) data, and Cohen's D and V_c estimated with the full data. As would be expected, the relation between V_d and D was less strong than the relation between V_c and D . However, the overall relation was still quite strong, correlating at 0.73 and 0.72 for mathematics and reading/language arts, respectively. The standard deviations of the differences were also marginally more broad, at 0.21 and 0.23, respectively. The coarsening of the data did also lead to some differences in the estimated effect between V_c and V_d . The correlations were similar in magnitude to V_c and D , correlating at 0.83 and 0.89 for mathematics and reading/language arts, respectively. The distribution of differences centered on 0.02 and 0.03 for mathematics and reading/language arts, respectively, implying that, on average, there was very little difference between V calculated with the continuous data (V_c) and V calculated with the discrete data (V_d). The standard deviations of these distributions were 0.17 and 0.23.

School-Level Achievement Gaps

Overall Estimates by State. Using publicly available data, we calculated the Black-White achievement gap for all schools meeting minimum reporting criteria for both groups of students in California, and the Hispanic/Latino achievement gap for all schools meeting minimum reporting criteria for both groups in Oregon. For both states, we limited the analysis to only mathematics. Figure 3 displays the distribution of effects for each state, with the mean displayed by the blue vertical line. Schools with an estimated effect above

zero represent schools in which the average achievement of White students was below the corresponding focal group. This portion of each corresponding distribution is displayed by a dark gray color and, for both states, represented very few schools relative to the total distribution. The overall mean Hispanic/Latino - White achievement gap in Oregon was -0.41 with a standard deviation of 0.29. In California, the average Black-White achievement gap was estimated at -0.54 with a standard deviation of 0.34. When interpreting these numbers, it is important to keep in mind that these represent the mean and variance of school-level achievement gaps. In other words, the overall achievement gap for each state may be marginally different, given that these estimates represent the mean of group means (which are unbalanced).

Mapping the Estimates. Following the estimation of all achievement gaps within each state, we restricted the analyses to Alameda County in California³. Maps were then produced to examine any geographical clustering of school-level achievement gap estimates, as well as features of the surrounding area that may correspond with this clustering. Figures 4, 5, and 6 displays screen captures of an interactive map, with census tracts colored according to the median housing cost, number of individuals within the tract identifying as Black, and the number of individuals in the tract with an income to poverty line ratio above 2.0, respectively. Note that in addition to providing information about the surrounding area, census tracts also help describe population density, with smaller polygons representing high population density (i.e., more people per square mile). Schools (pins on the map) colored darker red represent larger estimated achievement gaps. In the screen capture, it is somewhat difficult to discern some of the census tracts, particularly smaller tracts (high population density), which tend to correspond with many overlapping schools. However, the clustering of schools corresponding to the estimated achievement gap is still readily apparent. The Northwest section of the map includes a clear cluster of schools with large achievement

³As mentioned previously, preliminary maps were also produced for Multnomah County in Oregon and we hope to share these maps during the conference. Readers interested in the Oregon results can find similar screen captures in the GitHub repository from the link noted previously.

gaps, with students identifying as Black in these schools performing, on average, more than 1.5 standard deviations below their White peers. The more central portion of the map includes mostly schools with relatively moderate achievement gaps, colored pink ($V_d = -0.3$ to -0.9), or essentially no achievement gap, colored white ($V_d = -0.3$ to 0.3). However, even within this cluster of schools with more moderate achievement gaps, there are a few schools with very large achievement gaps.

In terms of using the surrounding demographics of the area to predict school-level achievement gaps, there are no clearly discernible patterns among any of the maps displayed. Figure 6, displaying the number of people within each tract with an income to poverty ratio above 2.0, has the most variability and there does appear to be a modest relation. In particular, the cluster of schools with high estimated achievement gaps in the Northwest are generally in areas with fewer individuals with an income to poverty ratio over 2.0.

Discussion

Education has somewhat of a unique challenge when it comes to making data open and accessible, given individuals of a protected class (minors) are included in the vast majority of cases. Aggregating data helps make these data publicly available and more easy to share, but comes with challenges in reducing the overall utility of the data for research purposes. This paper illustrated one method of using publicly available data, which have been reported by all states on a mandatory basis since the implementation of NCLB, to address a substantively meaningful question: How do achievement gaps vary between schools? Overall, the method used to estimate achievement gaps appeared to work well, in terms of the extent to which achievement gap estimates from a full dataset with individual student-level records were recovered when using coarsened data. Further, we were able to combine these effect size estimates with geo-spatial and census data to evaluate the extent to which school-level achievement gaps related to the demographics of the surrounding area.

In this study, the largest gains to evaluating achievement gaps via mapping was the clear cluster of schools found with large achievement gaps. Without looking at the data

spatially, or being intimately familiar with the set of schools, we likely would have observed the schools as all having large achievement gaps, but missed that they were nearly all clustered in the same geographical area. The census data did not, in this case, appear to explain much of the underlying mechanisms behind this clustering. However, it is important to note that clear patterns have been found in prior investigations of school variability using census data, using different school-level outcomes (see Anderson, 2017). Deeper investigations into the census data may find trends that were not apparent here.

Limitation and Directions for Future Research

In this study, we evaluated effect sizes with empirical data. Although this is helpful in determining how a statistic behaves in applied use, it is also a limitation in that “truth” was not known. That is, while Cohen’s D and V_c were moderately different, with V_c being marginally lower on average, we cannot state which is more “correct”. It is also important to note, however, that any measure of effect size is a summary statistic, summarizing the difference between two distributions. Important differences at specific points in the distribution may therefore be missed by these summary measures. For example, it is possible for large achievement gaps to be apparent at one point in the distribution, and not others, including the directionality of the effect even reversing (Anderson, 2018; Anderson, Stevens, & Nese, 2017; Stevens, Anderson, Nese, & Tindal, 2017). Despite using empirical data, however, we were able to know “truth” relative to the extent that V_d recovered the estimate from V_c .

This research also only preliminarily examined the ACS data, but the ACS, and the census bureau generally, has a wealth of data that have generally been underutilized in educational research. This manuscript investigated three of potentially thousands of variables that could account for variability among schools, not only in achievement gaps, but a host of other outcomes (e.g., graduation rates or behavioral data). A logical extension of this work would be to include data on criminal activity in the area, given the results of N. K. Bowen and Bowen (1999) and others (e.g., McCoy et al., 2013) indicate that neighborhood

danger relates to school-wide outcomes. These data are not available as part of the ACS, which is the most up-to-date census dataset, but are available through other US Census Bureau datasets.

Conclusions

The approach to estimating school-level achievement gaps illustrated here can lead to potentially rich datasets that can be shared publicly without concerns of jeopardizing student privacy, given that summary-level data are used to produce the estimates. In this study, we estimated achievement gaps for every school in both California and Oregon that met minimum reporting requirements. Future research could use these estimates for further research. For example, it would be interesting to actually model the spatial features, rather than just explore them visually, which could be done by estimating a multilevel model with census tract entered as a random effect. The variability in school-level achievement gaps lying between census tracts could then be estimated, and census variables could be used to account for this variability. An incredible array of questions could be addressed within this framework, while using only publicly available data.

References

- Anderson, D. (2017, November). *Mapping statewide school ratings with us census tracts*. Retrieved March 24, 2018, from <http://www.dandersondata.com/post/mapping-statewide-school-ratings-with-us-census-tracts/>
- Anderson, D. (2018). *Esvis: Visualization and estimation of effect sizes*. Retrieved from <https://github.com/DJAnderson07/esvis>
- Anderson, D., Stevens, J. J., & Nese, F. T., J. (2017). *Visualizing achievement gaps across the full distribution*. Retrieved from https://djanderson07.github.io/ncme_2017/
- Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Bowen, N. K., & Bowen, G. L. (1999). Effects of crime and violence in neighborhoods and schools on the school behavior and performance of adolescents. *Journal of Adolescent Research*, 14(3), 319–342.
- California Assessment of Student Performance and Progress. (2017). *Test results for english language arts/literacy and mathematics*. Retrieved March 24, 2018, from <https://caaspp.cde.ca.gov/sb2017/ResearchFileList>
- California Department of Education. (2017). *Public schools and districts files*. Retrieved March 24, 2018, from <https://www.cde.ca.gov/ds/si/ds/pubschls.asp>
- Chan, C.-h., Chan, G. C., Leeper, T. J., & Becker, J. (2018). *Rio: A swiss-army knife for data file i/o*.
- Cheng, J., Karambelkar, B., & Xie, Y. (2017). *Leaflet: Create interactive web maps with the javascript 'leaflet' library*. Retrieved from <https://CRAN.R-project.org/package=leaflet>
- Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1), 15–27.
- Gonzales, N. A., Cauce, A. M., Friedman, R. J., & Mason, C. A. (1996). Family, peer, and neighborhood influences on academic achievement among african-american adolescents:

- One-year prospective effects. *American Journal of Community Psychology*, 24(3), 365.
- Google. (2018, March 15). *Google maps api*. Retrieved March 24, 2018, from
<https://developers.google.com/maps/documentation/geocoding/start>
- Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher*, 39(1), 59–68.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2009). New evidence about brown v. Board of education: The complex effects of school racial composition on achievement. *Journal of Labor Economics*, 27(3), 349–383.
- Ho, A. D. (2008). The problem with “proficiency”: Limitations of statistics and policy under no child left behind. *Educational Researcher*, 37(6), 351–360.
- Ho, A. D. (2009). A nonparametric framework for comparing trends and gaps across tests. *Journal of Educational and Behavioral Statistics*, 34(2), 201–228.
- Ho, A. D., & Reardon, S. F. (2012). Estimating achievement gaps from test scores reported in ordinal “proficiency” categories. *Journal of Educational and Behavioral Statistics*, 37(4), 489–517.
- Holland, P. W. (2002). Two measures of change in the gaps between the cdfs of test-score distributions. *Journal of Educational and Behavioral Statistics*, 27(1), 3–17.
- Institute of Education Sciences. (2016, October). *IES policy regarding public access to research*. Retrieved March 24, 2018, from <https://ies.ed.gov/funding/researchaccess.asp>
- Kittrie, E., & Chattopadhyay, S. (2016). *NIH makes data sharing repositories publicly viewable on healldata.gov*. Retrieved March 22, 2018, from
https://datascience.nih.gov/Blog_HealthData.gov
- Lee, M., & Madyun, N. (2009). The impact of neighborhood disadvantage on the black–white achievement gap. *Journal of Education for Students Placed at Risk*, 14(2), 148–169.
- McCoy, D. C., Roy, A. L., & Sirkman, G. M. (2013). Neighborhood crime and school climate as predictors of elementary school academic quality: A cross-lagged panel analysis. *American Journal of Community Psychology*, 52(1-2), 128–140.

National Science Foundation. (2016). *The national science foundation open government plan* (No. MSU-CSE-06-2).

No Child Left Behind. (2002). Act of 2001, 20 usca 6301 et seq.

Oregon Department of Education. (2017a). *Student assessment*. Retrieved March 24, 2018, from
<http://www.oregon.gov/ode/educator-resources/assessment/Pages/Assessment-Group-Reports.aspx>

Oregon Department of Education. (2017b). Retrieved March 24, 2018, from
<http://www.ode.state.or.us/pubs/labels/SchoolMail.xls>

Pedersen, T. L. (2017). *Patchwork: The composer of ggplots*. Retrieved from
<https://github.com/thomasp85/patchwork>

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from
<https://www.R-project.org/>

Reardon, S. F., & Ho, A. D. (2015). Practical issues in estimating achievement gaps from coarsened data. *Journal of Educational and Behavioral Statistics, 40*(2), 158–189.

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly, 32*(3), 308–315.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.

Stevens, J. J., Anderson, D., Nese, F. T., J, & Tindal, G. (2017). *Using effect size measures to estimate and report achievement gaps*. Retrieved from
http://www.ncaase.com/docs/Stevens_et_al_NCME_2017-1.pdf

United States Census Bureau. (2016). *American community survey*. Retrieved March 24, 2018, from <https://www.census.gov/programs-surveys/acs/>

Walker, K. (2018). *Tidycensus: Load us census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames*. Retrieved from <https://github.com/walkerke/tidycensus>

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from

<https://CRAN.R-project.org/package=tidyverse>

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17–29.

Table 1

Content Area	ethniccd	n	mean	sd
Mathematics	Am. Indian	6341	217.81	21.67
Mathematics	Asian	13986	229.35	21.79
Mathematics	Black	9689	215.26	21.87
Mathematics	Hispanic	91967	217.60	19.61
Mathematics	Multiethnic	19646	222.63	20.39
Mathematics	Pac Islander	2760	218.01	19.99
Mathematics	White	234136	223.74	19.64
Rdg/Language Arts	Am. Indian	6006	217.47	21.47
Rdg/Language Arts	Asian	13941	225.01	19.27
Rdg/Language Arts	Black	9237	216.00	21.07
Rdg/Language Arts	Hispanic	89071	216.40	19.44
Rdg/Language Arts	Multiethnic	18649	222.51	19.68
Rdg/Language Arts	Pac Islander	2706	217.23	19.35
Rdg/Language Arts	White	220974	223.48	19.26

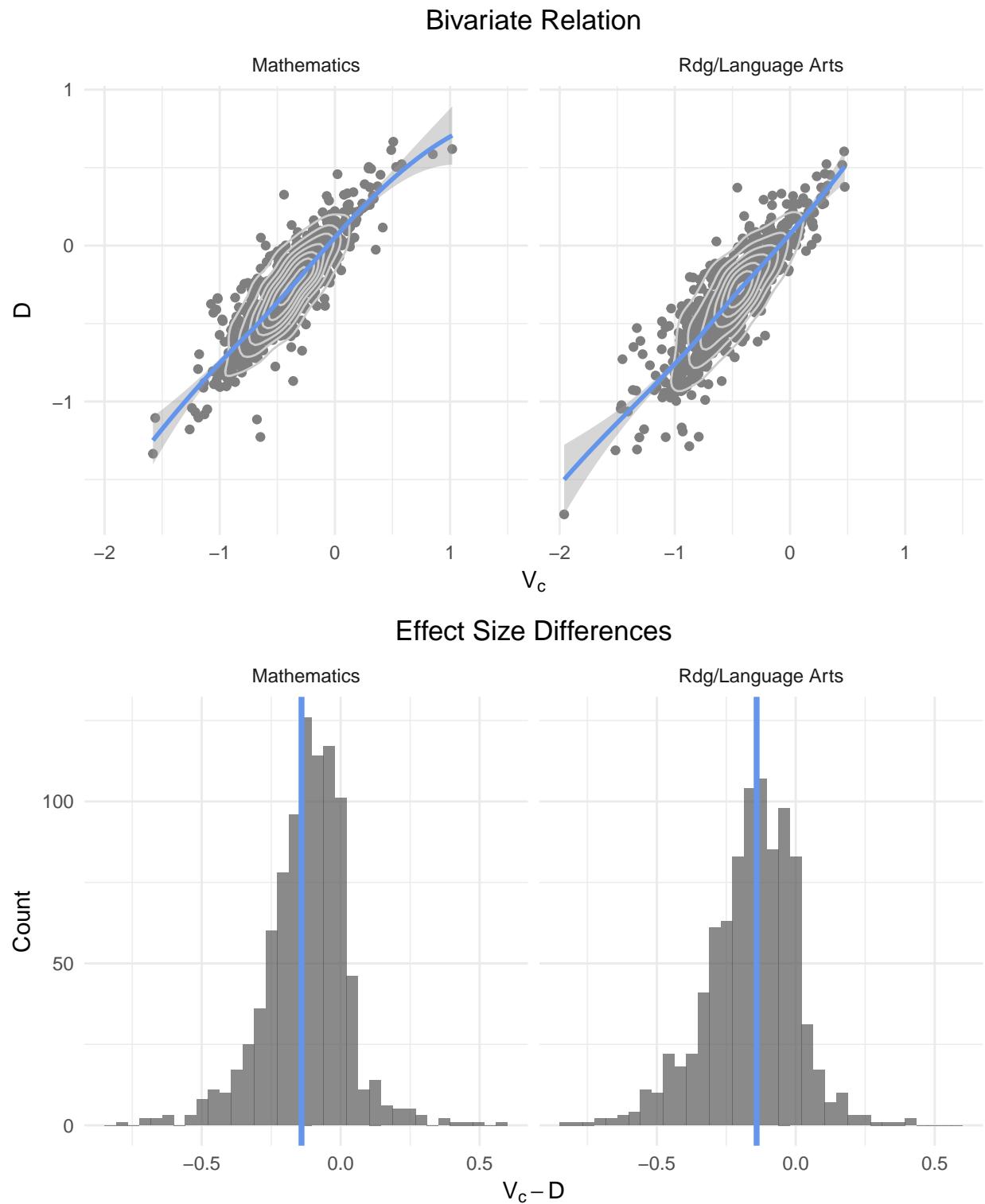
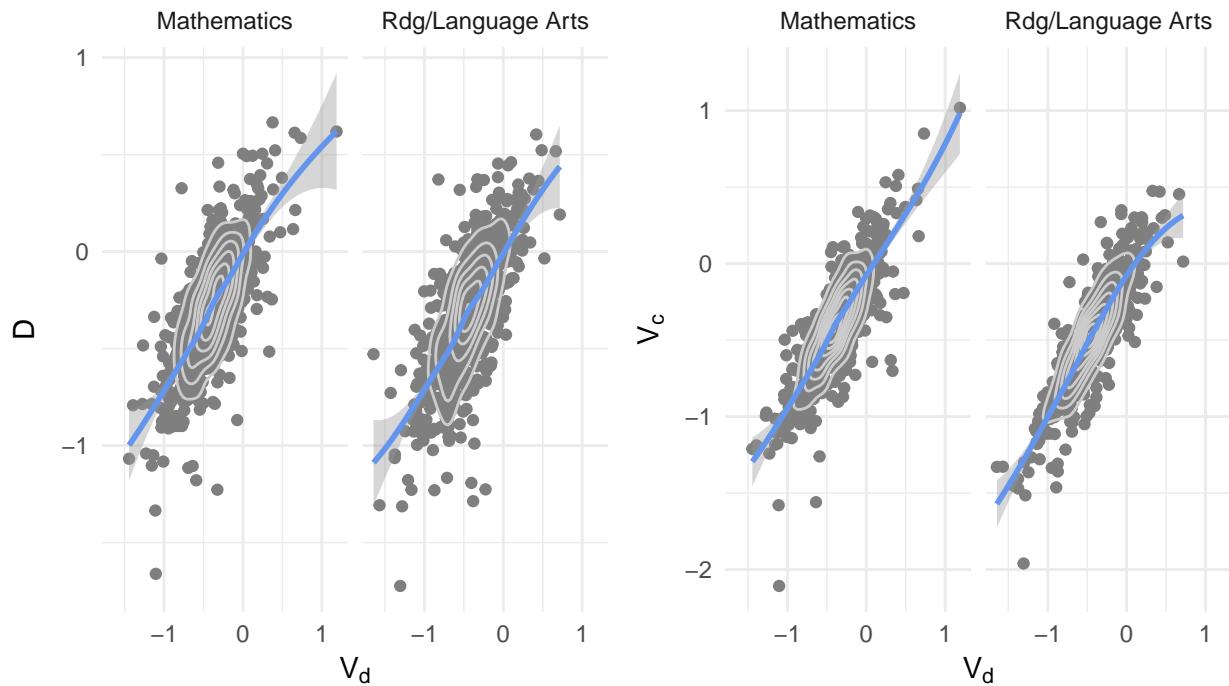


Figure 1. Differences in Cohen's D and V estimated with continuous data

Bivariate Relation



Effect Size Differences

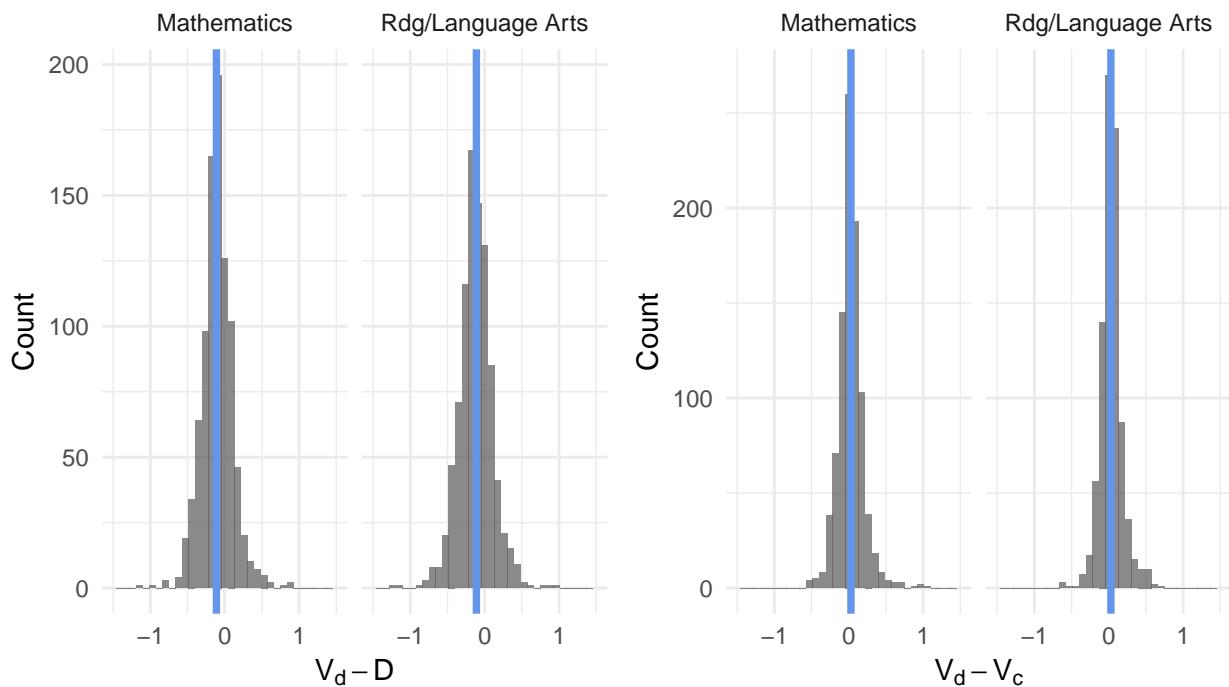


Figure 2. Differences in V estimated from coarsened data, and V and Cohen's D estimated from continuous data

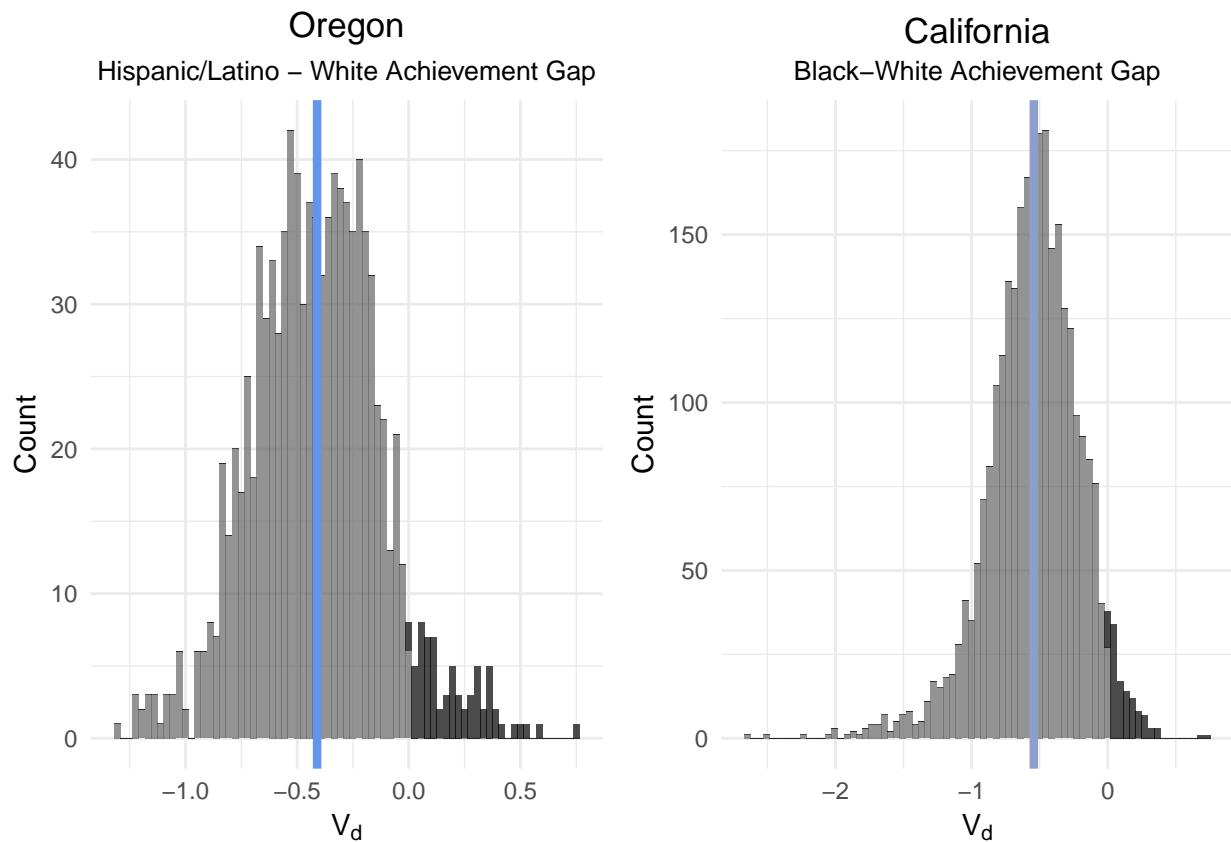


Figure 3. Distribution of Achievement Gap Effect Sizes in Oregon and California

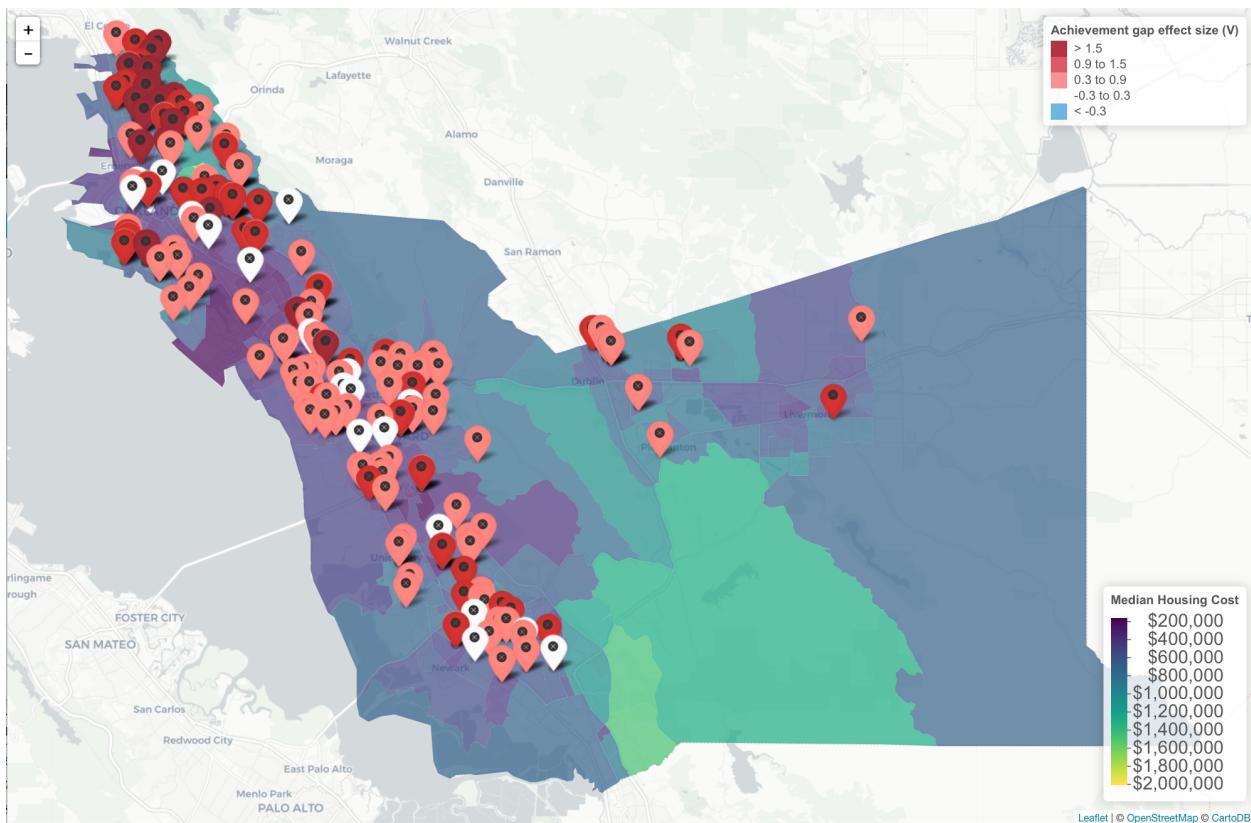


Figure 4. Alameda County School-Level Achievement Gaps and Median Housing Cost

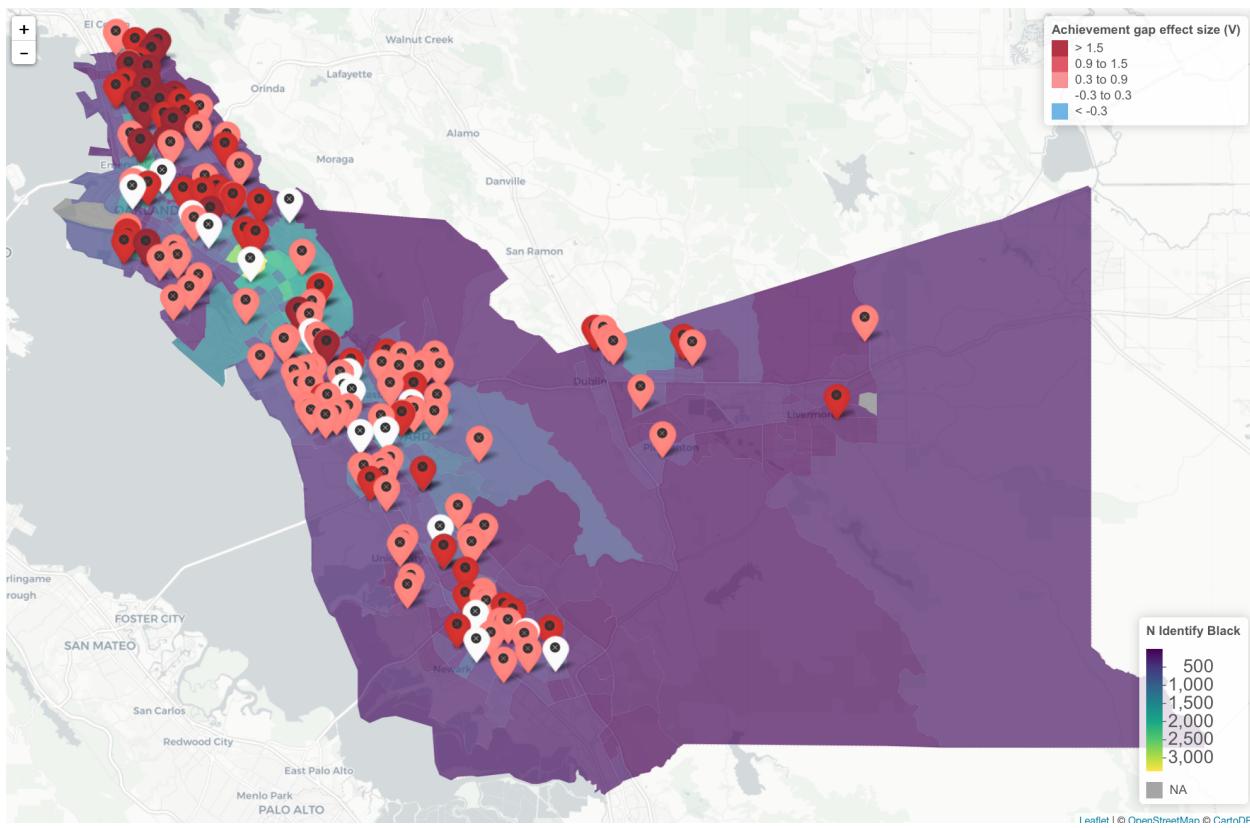


Figure 5. Alameda County School-Level Achievement Gaps and Number of Individuals Identifying as Black

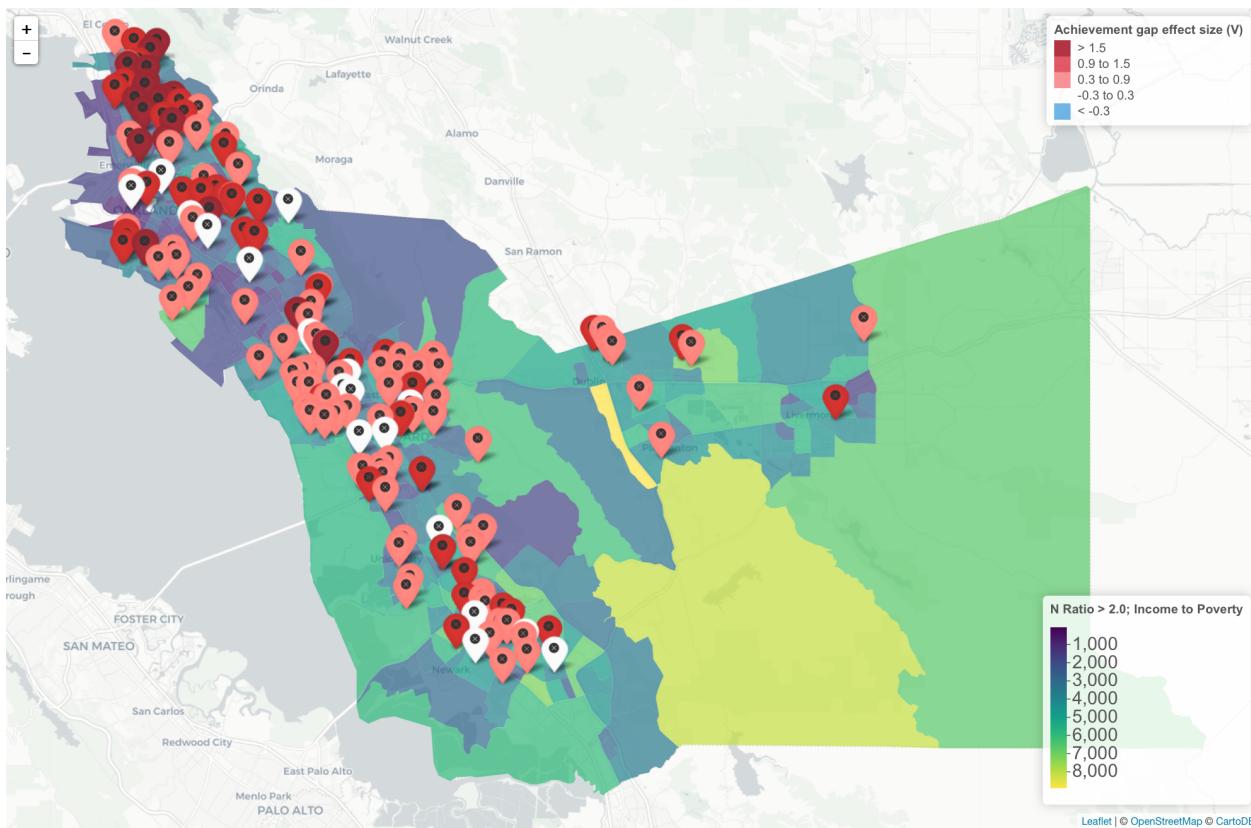


Figure 6. Alameda County School-Level Achievement Gaps and Median Housing Cost