California Public School College Admission Tests Executive Summary

by Eric Gordon

This is an abbreviated version of this project. To read or access thefull report, click pdf file (https://github.com/ericgordo/P4_EDA_CA_Schools/blob/master/FinalEDA.pdf) or HTML File (https://github.com/ericgordo/P4_EDA_CA_Schools/blob/master/FinalEDA.html)

This project, takes college entrance exam data from California Public Schools to explore, analyze, and find interesting insights into the scope of public education in the nation's largest populated state.

The Data Being Investigated

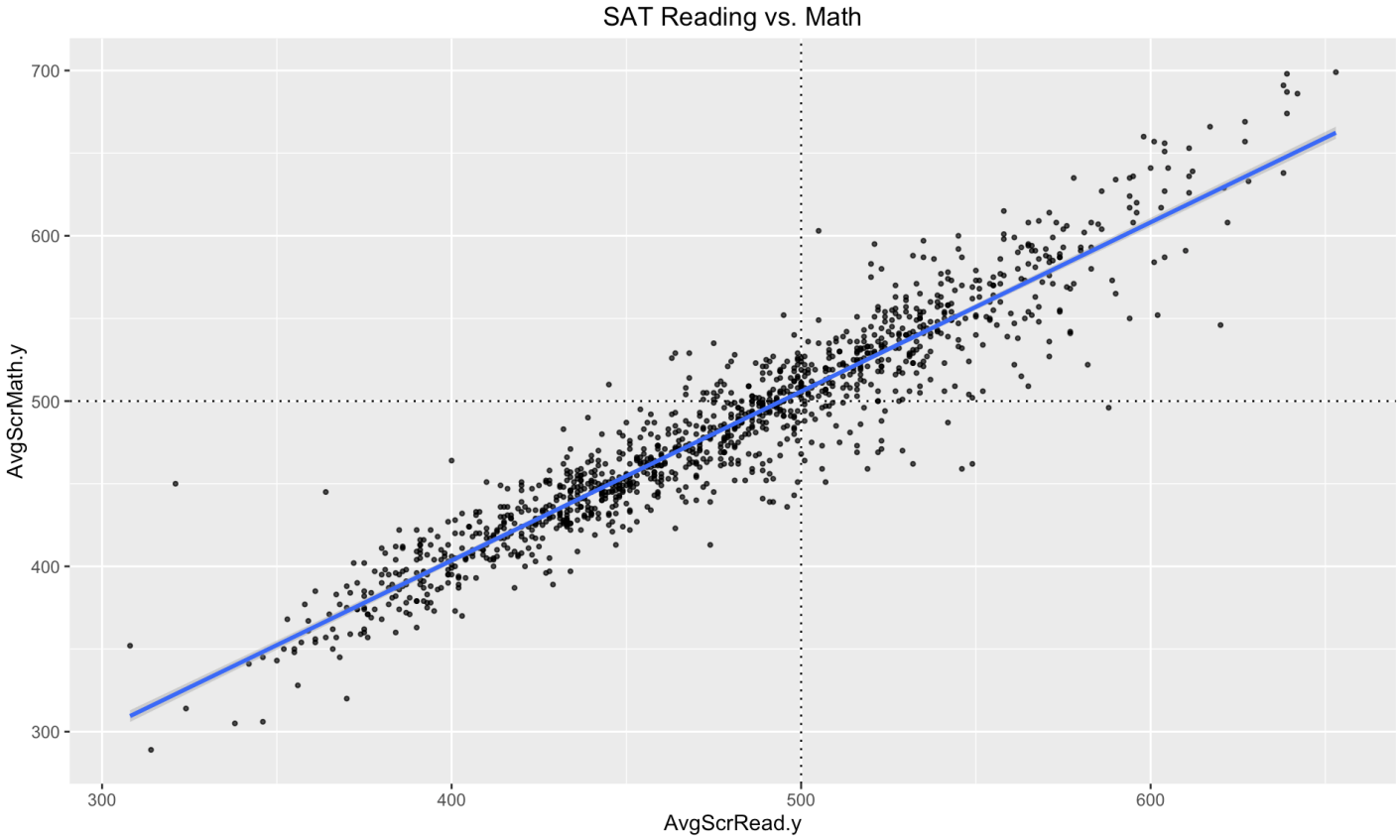
The data in this analysis is the collection of publically accessible files from the California Department of Education (CDE). These CSV files were combined to create a singular data frame containing information on 1252 California public schools, and test score information about these schools. All of this data is recorded from the 2014-2015 school year.

There are many interesting insights to be noted about this data, and for a full report follow one of the links mentioned above. The following results are in my opinion the most enlightening insights realized through the investigation.

SAT Comparing Subjects

The average score of a student's math scores at a school seems to be directly relational to the average reading score of the same school. Even though these are very distinct subjects, it seems highly relational the average scores of these two subjects seems to go hand and hand at different schools. This is to say, there are few to no schools that excel in one of these subjects and not the other.

This graph below seems to make it clear that a school either does well in preparing students for both Math and Reading subjects of the SAT, or that these two aspects of the SAT are highly related. That is the skills it takes to do well on one section of the test actually are the same as the other, even though they are different "subjects."




```
## [1] 0.9452091
```

This graph shows that how a school does in preping their students for one section of the SAT is highly relational to how that same school does in preparing their students for the other section. The correlation between these two variables is 0.945. This is shocking, especially when you consider how distinct people often view these subjects.

While just looking at the graph shows the clear relationship, it is interesting to investigate the linear model (shown above with the blue line) to predict a school's average SAT Math Score based on its average writing and reading score, to see how strong this correlation is. Let's create this model and evaluate it.

```
##
## Call:
## lm(formula = AvgScrMath.y ~ AvgScrRead.y + AvgScrWrite, data = school.df)
##
## Coefficients:
##  (Intercept)  AvgScrRead.y   AvgScrWrite
##     -23.2402        0.1390        0.9339
```

This linear model created takes a school's average SAT reading and writing scores, to predict what the school's average math section score is. We will evaluate the model using actual data. To test this model, I found the average scores of the reading and writing sections for a school named "Agoura", and we will input these values into our linear model, then compare our result with the actual school's Average SAT Math Score.

```
Agoura <- data.frame (AvgScrRead.y=571, AvgScrWrite=573)

modelEstimate = predict(
  score_model, newdata = Agoura,
  interval="prediction", level = .95)
modelEstimate
```

```
##       fit      lwr      upr
## 1 591.245 555.5662 626.9238
```

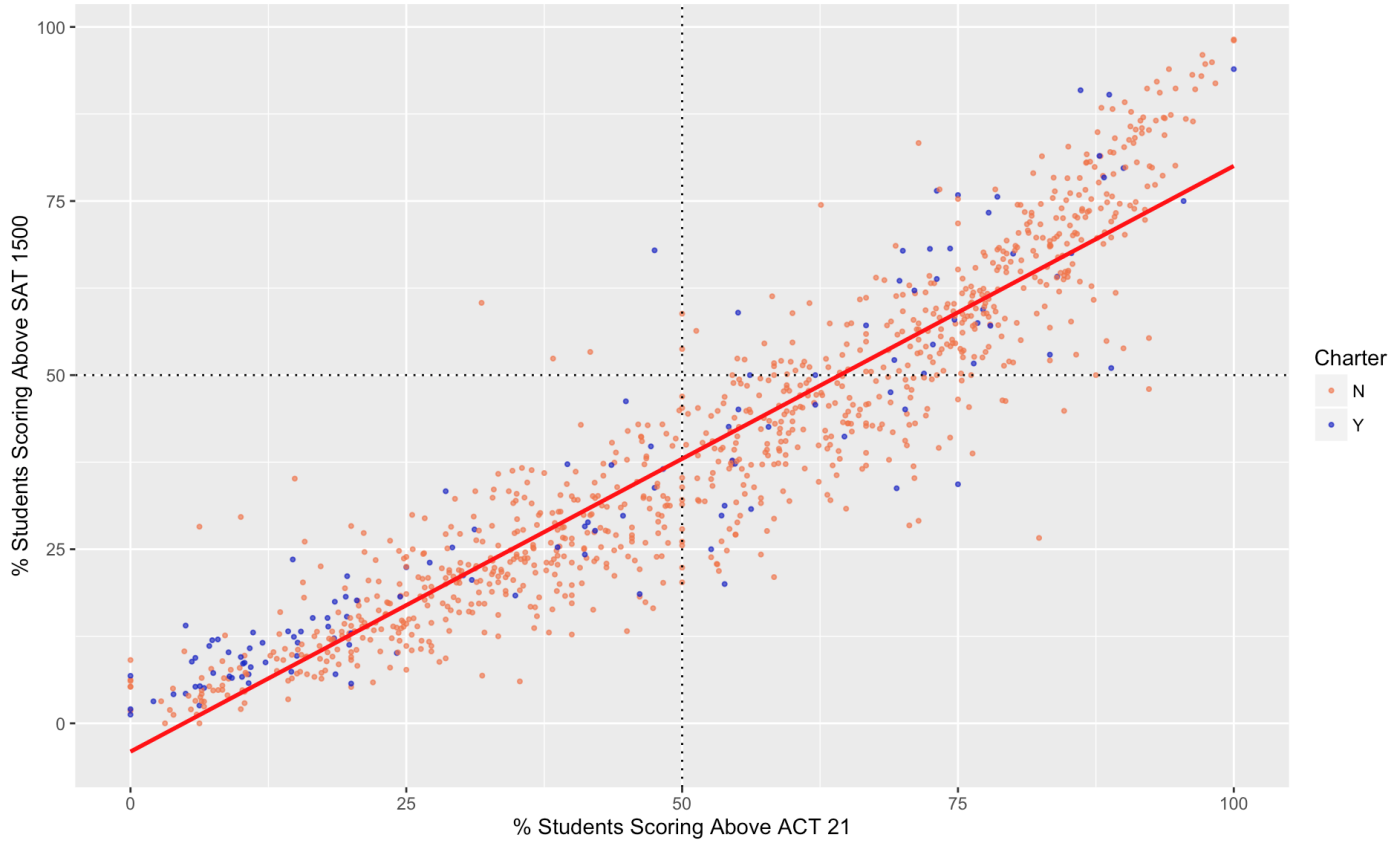
```
print (subset (school.df,
               School=="Agoura High")$AvgScrMath.y)
```

```
## [1] 576
```

Our model put the predicted Math score to be about 591, while the school actually had an average of 576. Our model came close, and the 95% confidence level actually included this value. So while our model is not perfect, it actually can give us a very good sense of what a school's average math score could be simply based on the school's reading and writing scores.

What is worth noting is how highly correlated the seaperate sections of the SAT are, basically shwoing that these sections may not be as seaperate and unique as one may think.

Students Scoring Above Average on Both Tests, and Charter Schools



```
## [1] 0.9327578
```

```
## [1] "ACT Summary All Schools"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   29.87   53.33   51.43   74.29  100.00     167
```

```
## [1] "SAT Summary All Schools"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   20.16   37.13   38.93   56.36   98.20      15
```

```
## [1] "Are Schools That Fall Below Avg For Both Tests Charters?"
```

```
##
##   N   Y
## 414  74
```

```
## [1] "How Many are Charters on Entire Plot?"
```

```
##
##   N   Y
## 945 125
```

```
## [1] "ACT Summary Charters"
```

```
## $N
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   32.81   54.79   53.08   75.00  100.00      88
##
## $Y
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   14.29   34.88   39.61   68.89  100.00      79
```

```
## [1] "SAT Summary Charters"
```

```
## $N
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   21.76   37.98   40.04   57.25   98.20       7
##
## $Y
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   12.02   29.34   33.33   50.43   93.94       8
```

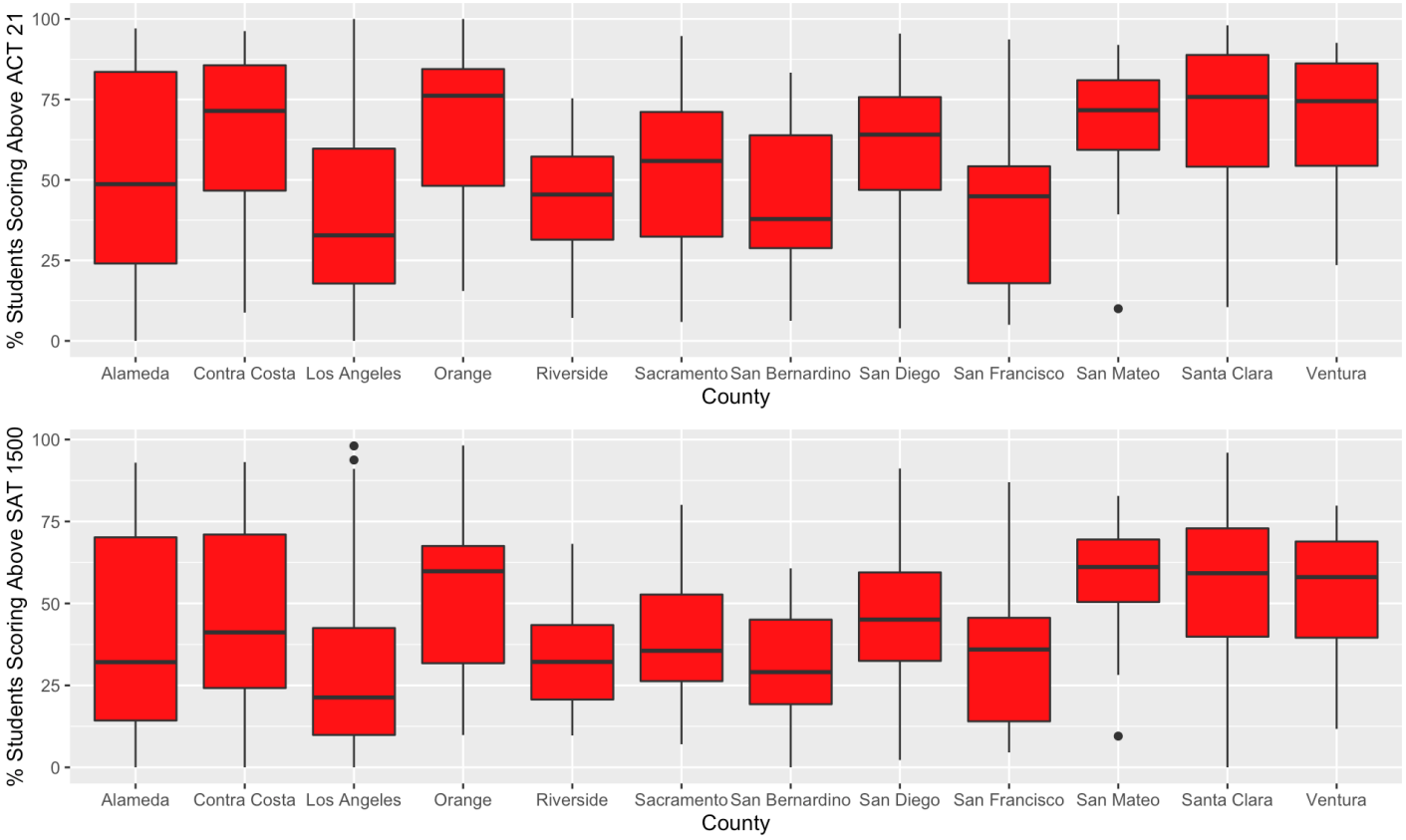
There is a clear strong correlation between the percentage of students who will score above a 21 on the ACT, and the percent of students who will score above a 1500 on the SAT, as seen on the red line on the graph and the 0.93 calculated correlation. This graph shows that at a particular school, roughly the same percentage of students score above average on both the SAT and ACT, not just one or the other.

Also, as we can see from the red trend line, the summary stats, and the points themselves on the graph, students seem to achieve overall at higher rates in California on the ACT versus the SAT. The median percent of students scoring above average at a single school on the ACT was 53%, versus only 37% for the SAT.

Another aspect of this graph that is vital is the coloring of Charter schools and Non-Charter schools. The recent expansion of charter schools through out the country has seemed to bring lively debate in the field of education, but here it seemed to be clear that there were a lot of charter schools that had the majority of their students score below average in both the SAT and the ACT. Looking at the specific numbers now, of the 125 charter schools that had both scores recorded for the ACT and SAT, 74 of them had less than 50% of their students performing above average on either tests. Additionally, the mean of non-charter schools on the ACT was around 53% of students scoring above average with a median of 54%, while charter schools only having a mean of 39% and median of 34%. The numbers for SAT are a little bit lower, but again show normal public schools with a higher percentage of students scoring above average in both the mean and median.

Many of these California charter schools therefore do not seem to be showing results that they get their students to score above average on college entrance exams. This is shocking, because charter schools are advertised and often considered to be alternatives to offer a better educational outcome than a traditional public school. However, according to this success metric, charter schools do not seem to be outperforming other public scores in this comparison, but instead seem to be underperorming the state averages. This scatterplot really just shows that a title like “Charter School” may not having a uniform effect, but truly schools can vary greatly, and need to be examined an evaluated on a case by case basis.

Urban Test Score Distributions



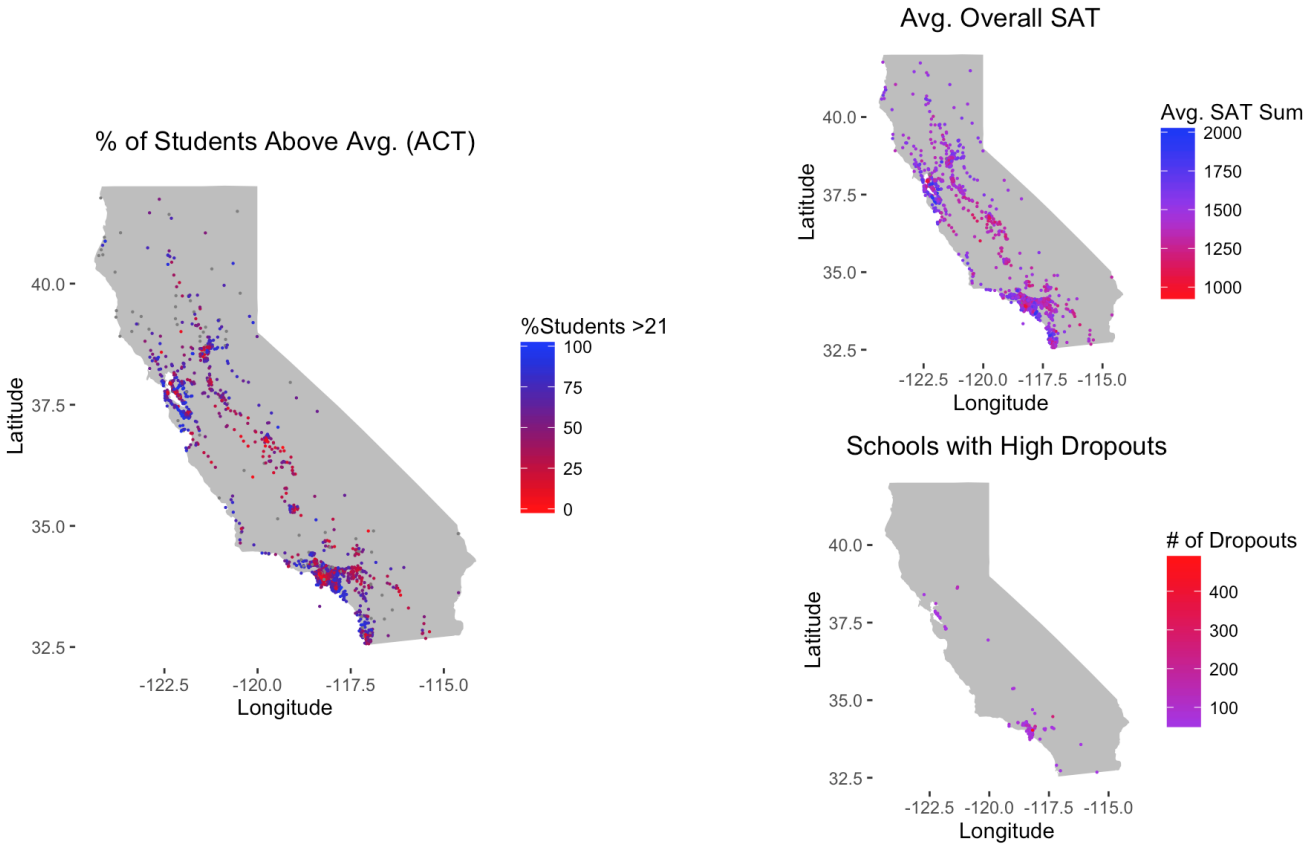
```
## $Alameda
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   24.04   48.68   50.75   83.56   97.06      10
##
## $`Contra Costa`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    8.79   46.67   71.43   63.26   85.61   96.24       4
##
## $`Los Angeles`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   17.82   32.77   38.97   59.72  100.00      43
##
## $Orange
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   15.49   48.15   76.15   66.71   84.44  100.00       1
##
## $Riverside
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    7.14   31.43   45.45   44.62   57.24   75.37       4
##
## $Sacramento
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    5.88   32.37   55.88   51.72   71.11   94.70      10
##
## $`San Bernardino`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    6.25   28.81   37.84   44.59   63.87   83.33      10
##
## $`San Diego`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    3.90   46.88   64.07   61.38   75.72   95.45      15
##
## $`San Francisco`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    5.00   17.91   44.90   38.01   54.23   93.64       3
##
## $`San Mateo`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   59.32   71.66   67.61   80.98   91.95
##
## $`Santa Clara`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   10.48   54.10   75.76   69.93   88.82   98.02       3
##
## $Ventura
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   23.53   54.37   74.47   66.13   86.18   92.59       2
```

```
## $Alameda
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   14.29   32.10   39.48   70.16   92.94       4
##
## $`Contra Costa`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   24.20   41.18   45.23   71.02   93.12
##
## $`Los Angeles`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00    9.88   21.32   27.90   42.50   98.05       4
##
## $Orange
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.84   31.77   59.82   52.96   67.52   98.20
##
## $Riverside
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.72   20.69   32.16   32.63   43.42   68.18
##
## $Sacramento
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    7.05   26.30   35.56   38.90   52.71   80.08       2
##
## $`San Bernardino`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   19.26   29.06   31.74   45.06   60.71
##
## $`San Diego`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.22   32.48   45.08   45.60   59.46   91.16
##
## $`San Francisco`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    4.55   14.04   35.96   34.71   45.62   86.98       1
##
## $`San Mateo`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.52   50.44   61.10   57.64   69.50   82.81
##
## $`Santa Clara`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   39.86   59.22   57.90   72.90   95.98       1
##
## $Ventura
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.72   39.56   58.03   54.07   68.88   79.84
```

The urban centers throughout California also seem to greatly differ on their student tests averages. Counties like San Mateo, Ventura, Orange and Santa Clara all have their median school reporting that well over half of their students scoring above the national average on both the SAT and the ACT. Then when you look at the Los Angeles county, the median school reports that only about 32% of its students score above average on the ACT, and only around 21% of the students earn above average marks on the SAT. While the Los Angeles county has a

large range of schools, as seen by the long boxplots, the average school there seems to be below the rest of the California Urban districts with maybe the exception being San Bernardino. It is just interesting that even in a singular state, the urban counties can achieve at such different levels, and it was worth including on this analysis.

Maps of California



There is a large variety of school performance as seen on the first two maps, with some schools having close to or all of their students performing above average on the SAT, and averaging very high combined SAT scores. As seen on the third map, the only large areas that are marked with high dropout rates seem to be the Los Angeles area, followed by the East Bay of San Francisco, then other schools scattered in the state. Thus from all of these maps combined we can see that geography seems to play a vital role in the success of California public schools.

It is hard to generalize results for the entire state, but instead it is vital to compare geographical regions. For example, the geography of the San Francisco Bay Area seems to led to noticeable locations where schools seemed to perform differently, simply depending on what part of the Bay you examine. Thus in general, Calfiornia school performance seems to be highly relational to where in California the school is located.

Conclusion

This Report was just a small sample of the analysis that was done to find these results. For the full report, with deeper data examination as well as a more robust conclusion, please follow one of the links at the top of the page.

Work Cited

All data used in this analysis found at: http://www.cde.ca.gov/ds/ (http://www.cde.ca.gov/ds/)

Definition of “Urban” Counties: http://www.urbancounties.com/about/board/ (http://www.urbancounties.com/about/board/)

General RStudio TroubleShooting

https://www.udacity.com/course/data-analysis-with-r--ud651 (https://www.udacity.com/course/data-analysis-with-r--ud651)

http://stackoverflow.com/ (http://stackoverflow.com/)

http://www.cookbook-r.com/ (http://www.cookbook-r.com/)

https://www.r-bloggers.com/ (https://www.r-bloggers.com/)