Visualizations of Colorado State Education System

Team members:

Spears Paul - spearspa@indiana.edu

Shruthi Ramakrishnan - shrurama@iu.edu

Jaideep Patel - jaipatel@iu.edu

CSCI-B565, Data Mining Final Project

Due 4/29/2016

Objectives and Significance:

Education is at the center of every society. It helps provide children a better future, giving them access to culture knowledge and future jobs. Building stronger communities and more advances societies. A good education system should be able to provide high performing schools to every child regardless of their:

- o Location
- o Income
- o Race

The underlying question is: Are these goals being achieved by the Colorado State education system in United States? The Colorado state education system and Colorado school grades have been collection data for the last three years in order to evaluate measure and improve the public education system. Based on the data our team has developed some key visualizations making it easy for everyone to view and analyze the performance. Our primary goal is to uncover the trends hidden in the data. This will help detect opportunities, recommendations, weaknesses and strengths. Every child deserves a high performing school to explore, learn and engage.

By creating these visualizations, it will provide parents with a simpler method to look at the data that has been gathered by Colorado. Having this ability will allow the parents to more effectively select which school they will send their child, and drive up the overall quality of education in the State. This promotes a system where schools are actively attempting to increase their overall quality in an attempt to attract more students, while also allowing parents to be more easily and actively involved in their children's education.

Background

The State of Colorado has collected data on their public school system for multiple years. Now with such a repository of data, the challenge is in understanding what that data means and then in applying that knowledge to solve potential problems in the school system. While a pioneer in school grading systems, there is still much space left for Colorado in the understanding of the data which they collect. This has led the

Colorado Department of Education to submit their large data source to Kaggle for individuals to assist in creating an interpretation. The Colorado Department of Education is a non-profit community organization that believes all children deserve access to a high-performing school. This would be more easily accomplished if the large data source could be transformed such that it was more easily understood by individuals, such as in a visual format. Part of this project is to create such a visual representation of the data based on the rating systems created by the Colorado Department of Education and their other collected data.

Many rating systems are based solely on academic achievement, or one snapshot in time of student performance. However, Colorado School Grades includes student academic growth in its rating of overall school performance. Colorado School Grades worked with the Center for Education Policy Analysis at the University of Colorado at Denver and R-Squared Research, LLC to calculate the grades using statewide test score data as well as the exact same variables and weights as the Colorado Department of Education's School Performance Framework. Colorado School Grades replaces fuzzy categories such as "performance" and "priority improvement" with universally understood A-F letter grades (Table 1).

Table 1 Colorado Schools Grading Model

Colorado School Grades Ratings Model (highest to lowest)	
Category	*Distribution*
A (plus)	98.0-100
A	92.0-97.9
A-	90.0-91.9
B (plus)	85.0-89.9
B	70.0-84.9
B-	65.0-69.9
C(plus)	55.0-64.9
C	25.0-54.9
C-	15.0-24.9
D(plus)	13.0-14.9
D	7.0-12.9
D-	5.0-6.9
F	4.9 and below

Other similar projects have been done on this same data set. Our goal is to take a fresh look at the data and be able to confirm or refute some of the previously made claims about relationships between various attributes of this data set.

Methods

The data was provided by the Colorado State Department of Education to Kaggle as part of a competition to create these visualizations. By using this data, we produced a variety of visualizations from which we were able to draw some conclusions that were not obvious by just a cursory inspection of the data. This process mostly involved using tools, such as Tableau, to easily manipulate into these visualizations. A critical step in accomplishing this manipulation was some level of pre-processing of the data. In particular, the data was very distributed, and an aggregation of the data was critical in being able to compare various features that were in different initial data files.

Results

As a general introduction to the school grades, figure 1 shows the relative amount of schools at each grade level, 'A' though 'F'. This shows that there is a reasonable distribution of schools at each grade, but that the distribution is slightly skewed toward the higher ranked schools. Overall, this is a reasonable result, as school are likely to grade themselves up, to make themselves more appealing to potential students. This means that lower ranked schools are expected to provide a markedly lower quality of education than the higher ranked schools.
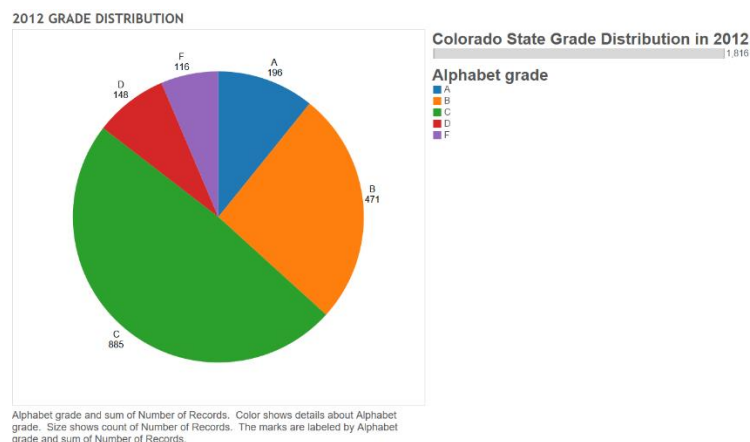

Figure 1: Overall Grade Distribution 2012

Geographic Outline of School Ratings in Colorado:

This section will help to provide an overview of which geographical locations provide both the highest and lowest rated schools. By visualizing this data, parents will be able to easily understand which locations they should aim to both travel to, and stay away from. Figure 2 shows a general map with all of the schools color coordinated by grade at their location on a map of the State of Colorado. This also shows a particular focus on the Denver area where there is a high concentration of schools. Figure 3 shows a heat map of the highest concentrations of high and low graded schools respectively. As can be seen from these figures, it is possible to attend a high rated school no matter where in the State a person in located. With this in mind, the highest concentration of 'A' graded schools is found in the Denver area, where the concentration of 'F' graded schools is more spread out over the state.
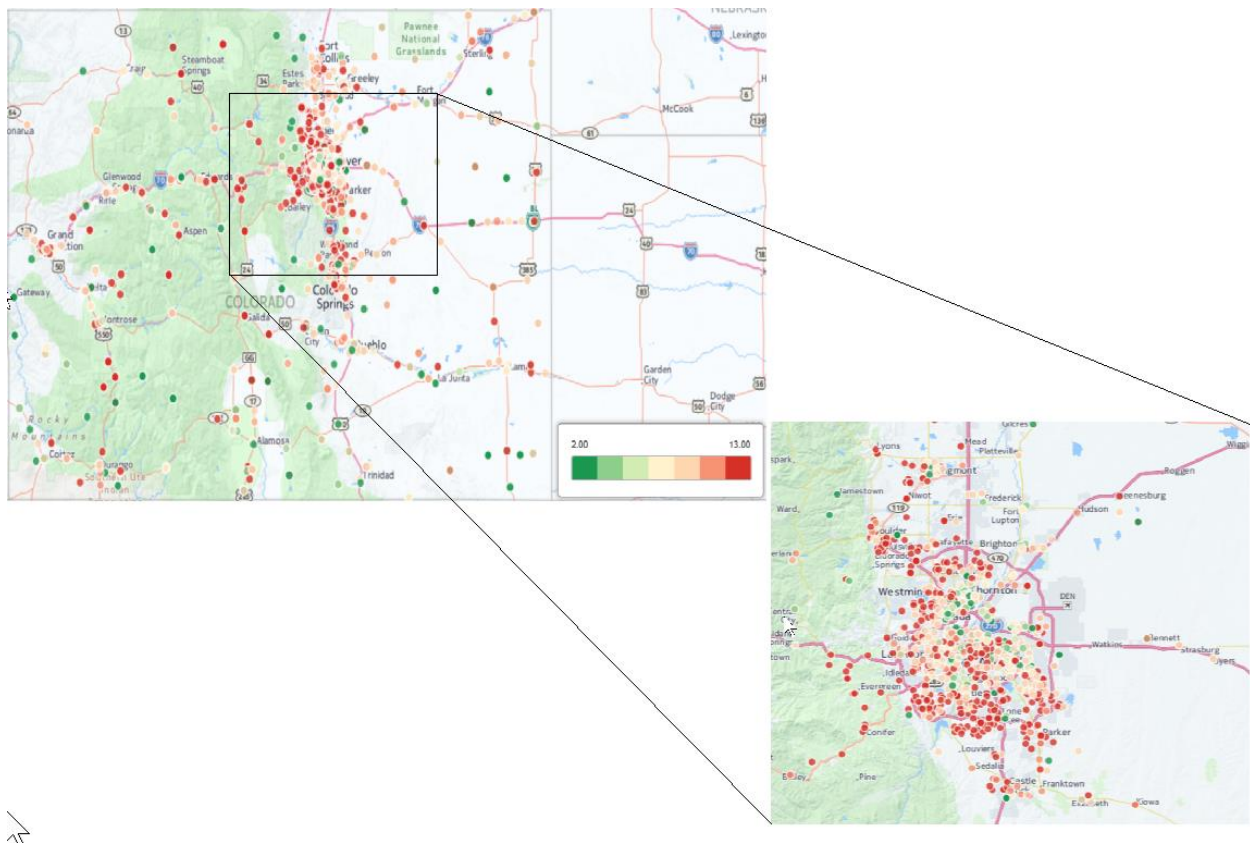

Figure 2: Colormap grading of all schools in Colorado, focus on Denver

It is interesting to note that there are hotspots for both 'A' and 'F' graded schools in the Denver area. This is due to the very high concentration of schools in this area overall. In particular, this can be seen in the number of 'A' and 'F' schools in the most populated districts, figure 4. Denver is the top of both these lists in having more schools at both 'A' and 'F' rating. There is a much higher population density in Denver than in other location in the state, so there are many more schools overall to keep the number of students per school reasonably low. This ratio of students per school, along with which districts are the most highly concentrated can be seen in figure 5. From these geographic visualizations, it is great to know that the State of Colorado provides opportunities to individuals regardless of their physical location in the State. While some schools may have higher or lower grades overall, it is possible for the active parent to send their child to a highly ranked school regardless of their physical location.
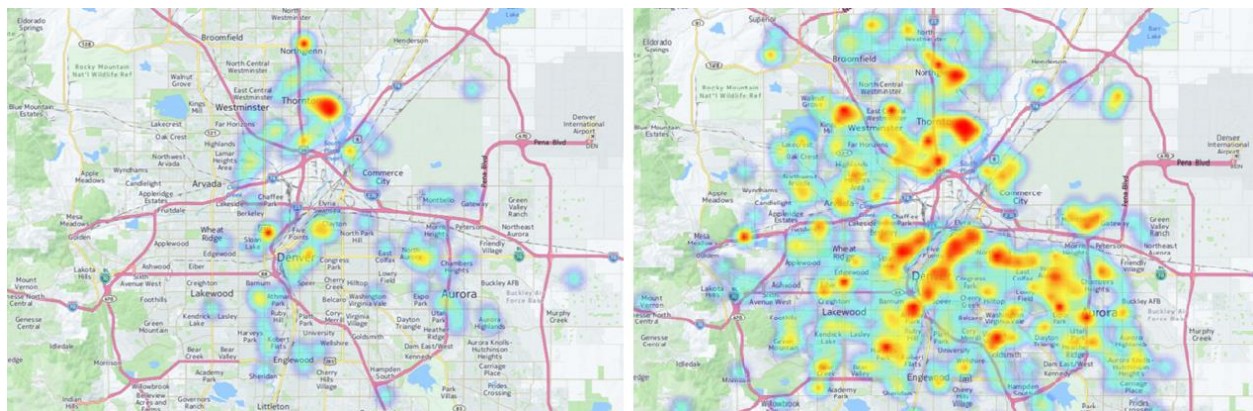

Figure 3: Heatmap displaying location of 'A' Schools (left) and 'F' Schools (right)

Income and Demographic effect on School ratings

Across the entire State, there is a wide variety of demographics in the various schools, as can be seen in figure 6, along with a large disparity in the income of the families of the students attending those schools. Many individuals are from low income families, and this can most easily be measured by looking at the percentage of students on Colorado's free lunch program. Many different claims on the effects of demographics and income on the quality of a school's provided education have been made in the past, we took a fresh look at the data in an attempt to discover which features are correlated. As is shown in the race distributions, a large portion of the State of Colorado falls under

the category of "White". With this knowledge, the demographics can be approximated by measuring the percentage of students in this category.

Top Districts with A Grade Schools
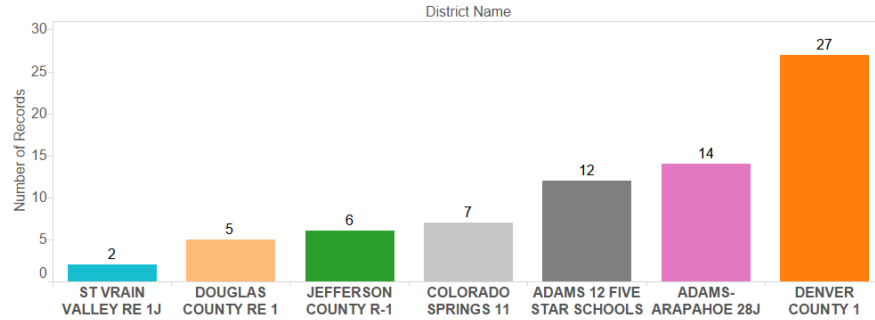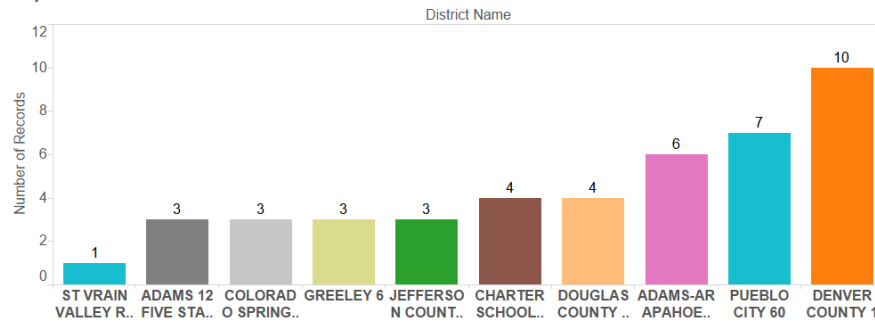
Top Districts with F Grade Schools

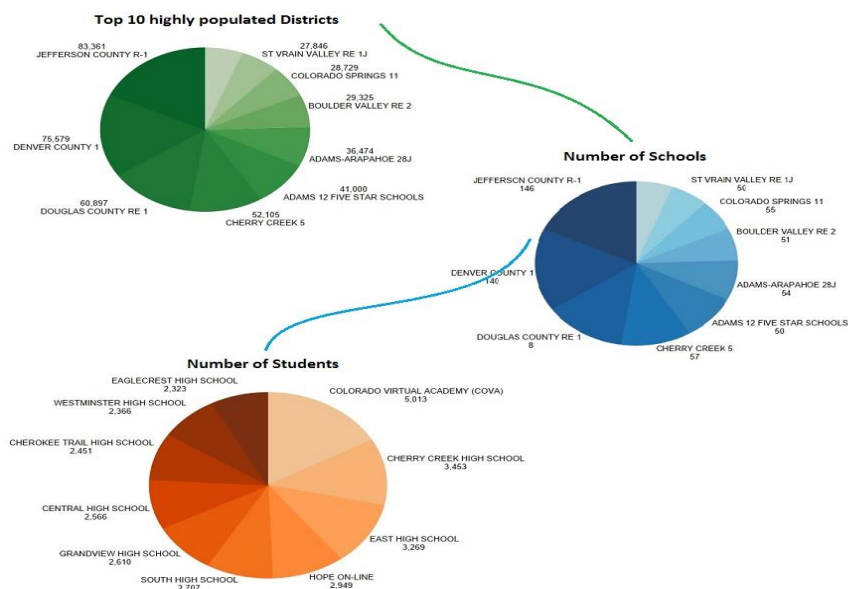*Figure 4: Number of Schools with 'A' Grade (top) and 'F' Grade (bottom)*

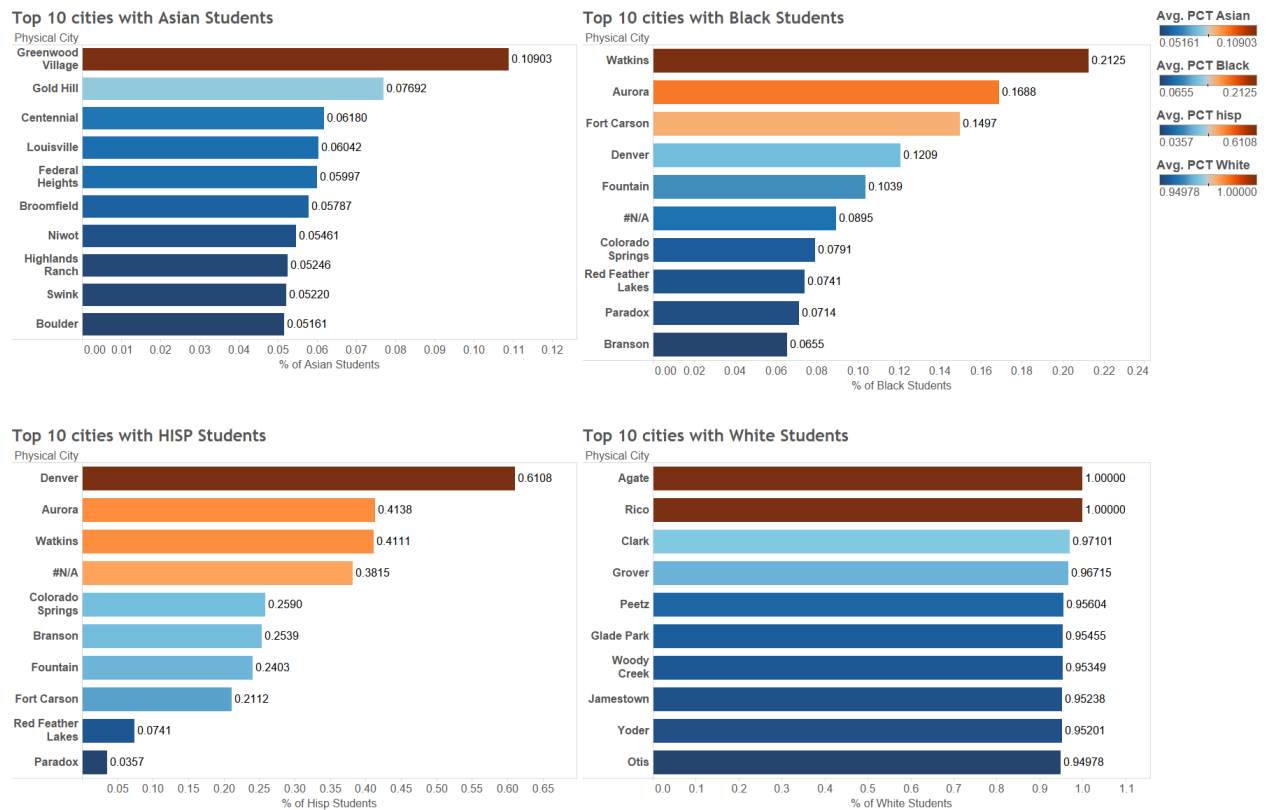*Figure 5: Population distribution per school in most populated districts*

Figure 6: Top cities with varied demographic percentages

A first exploration of the data compares the percentage of students in each school that are white, an approximation for demographic data, and the percentage of students that are on the free or reduced lunch program, an approximation for low income students. These results can be seen in figure 7, where there is clearly a negative correlation between these two factors. This means that there is a correlation where as a school has a larger percentage of white students, then there are fewer low income students. This trend is particularly true when the school has a low percentage of 'White' students.

PCT low income vs PCT White
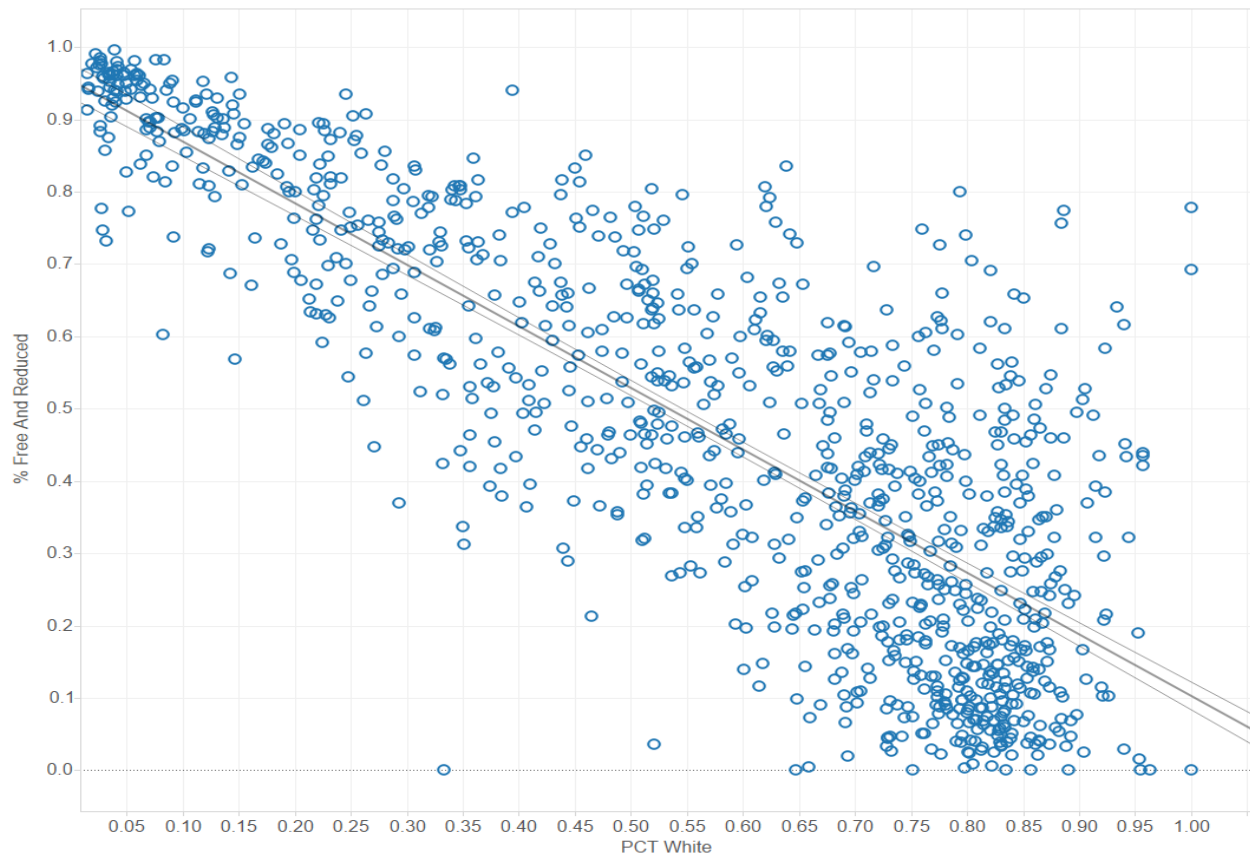
PCT White vs. % Free And Reduced.

Figure 7: Correlation between Demographics and Low Income

The next step is to compare the grades of the schools to each of these factors, the results can be seen in figure 8. While there is no strong correlation between the school grade and the demographic or income level of the students, there is a slight trend that as more students are 'White', the grade of the school will be likely higher, and as more students are low income, the grade will be likely lower. There are many schools which do not follow this general trend, but there is a weak correlation between these factors.
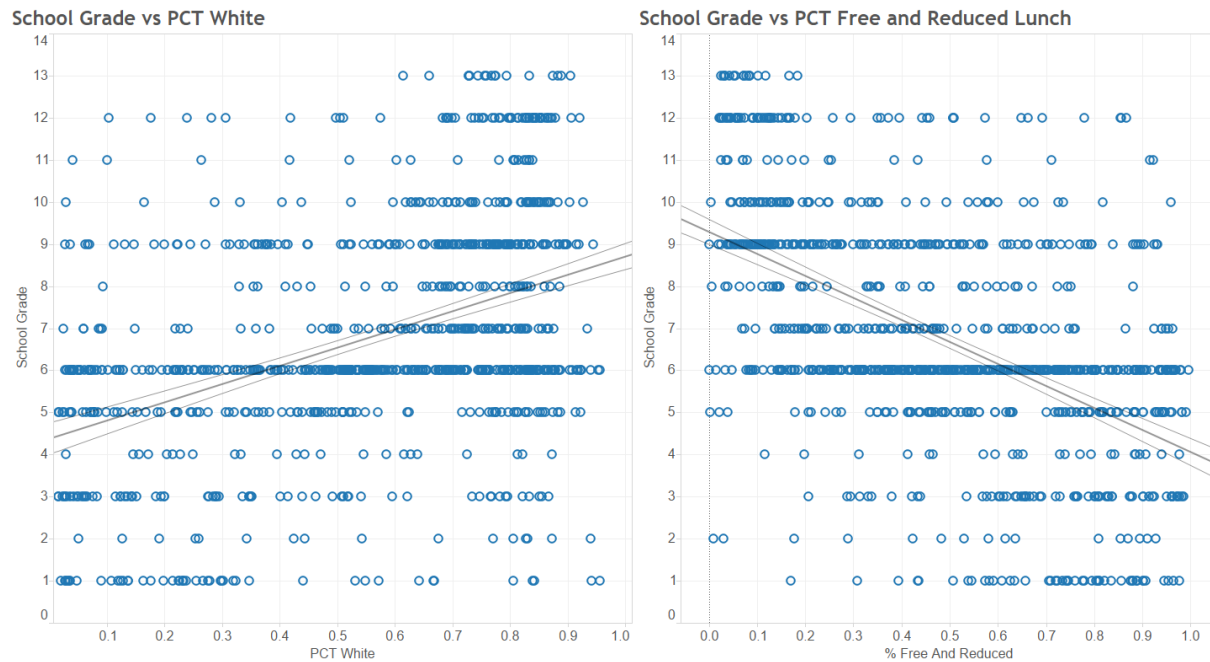
Figure 8: School grades correlation to Demographis (left) and Income level (right)

Also, to check the Colorado Department of Education's grading system, we compared the individual grades of the High Schools with their college readiness scores, a combination of individual scores in different subjects, Math, Science, Reading, and Writing. This discovered a very strong correlation, figure 9, between the grade of the school and the college readiness. This correlation is less useful to all parents, as it only is applicable to high schools which prepare their students for college, as opposed to elementary schools which prepare their students for middle school. This is a promising result, as it confirms our expectation that higher rated schools would more effectively prepare students for their next academic steps. A more specific version of these results can be seen in figure 10 where the readiness in the different categories is displayed over multiple years. These results are split and only shown for the A ranked and F ranked schools. Overall, the A ranked schools have much higher scores on the college preparedness than the F ranked schools.
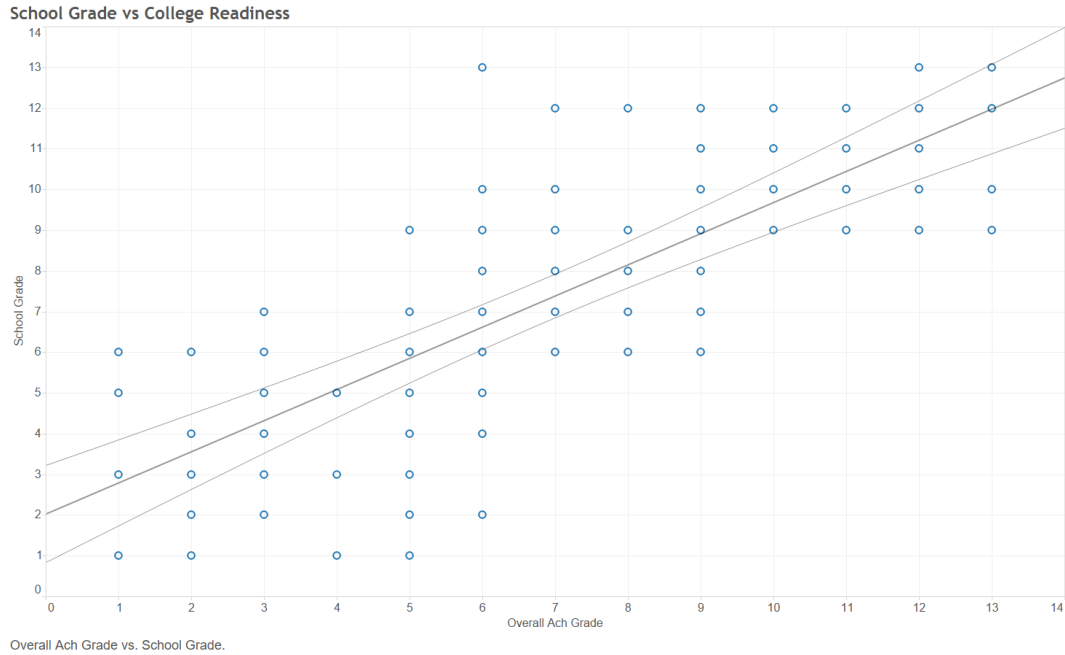
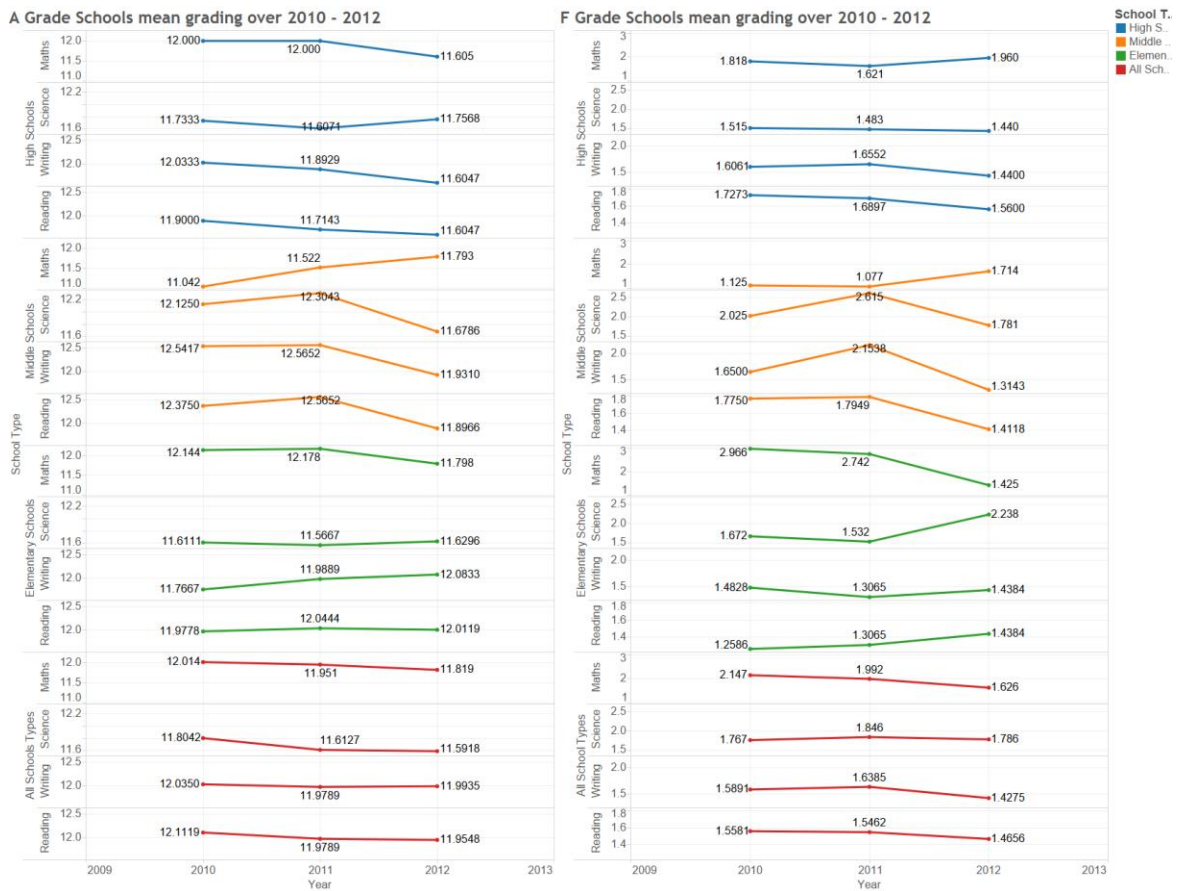*Figure 9: School Grade vs College Readiness*

Figure 10: College Readiness for A (left) and F (right) schools

Overall, this leads us to conclude that the income and demographics do have a slight impact on the grade of a school. This will allow prospective parents to try to move away from these trends, while simultaneously allowing for educators to focus their efforts on improving the quality of these lower income schools. It cannot be determined if either of these attributes are a cause of the lower grades, but there is a weak correlation that the data displays.

Growth over time

Data has been collected over multiple years, and it is possible to inspect for trends over that time. This section looks into these trends and inspects the data to see what could be expected over the next few years if these trends continue. Figure 11 shows the overall percentage of low income students at the various school grading levels from both 2010 and 2012. In particular, there is an increase in the percentage of students in higher graded schools who are low income students. This indicates that low income schools are becoming more successful in the grading scale, which is a promising result. This is even more promising when seen in conjunction with figure 12 which shows that the overall distribution of schools with high percentage of low income students is relatively constant over the three years of data. Unfortunately, as can be seen in figure 13, there is very little change overall in the number of schools that are at a higher rating.
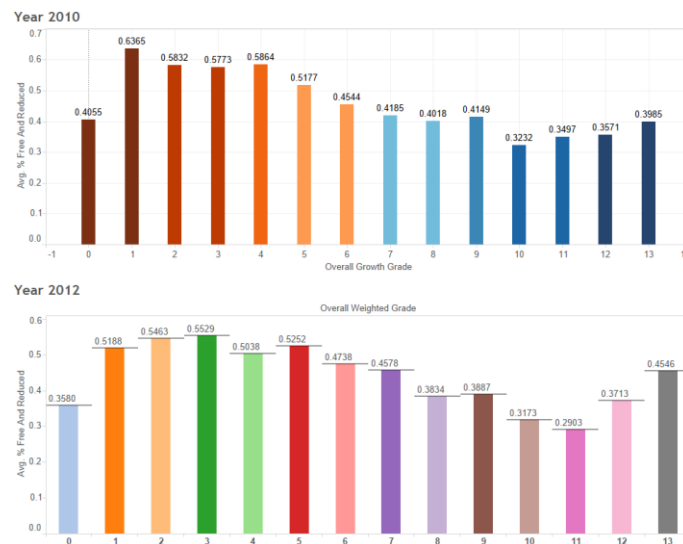
Figure 11: Percentage of low income students at each grade from 2010 and 2012

District Name	
ADAMS COUNTY 14	0.85133
AGUILAR REORGANIZ..	0.82500
CENTENNIAL R-1	0.83500
MANZANOLA 3J	0.83800
SHERIDAN 2	0.83325

District Name	
ASPEN 1	0.0553
DOUGLAS COUNTY RE..	0.0909
EXPEDITIONARY BOCE..	0.0000
LEWIS-PALMER 38	0.1139
MOUNTAIN BOCES	0.0000
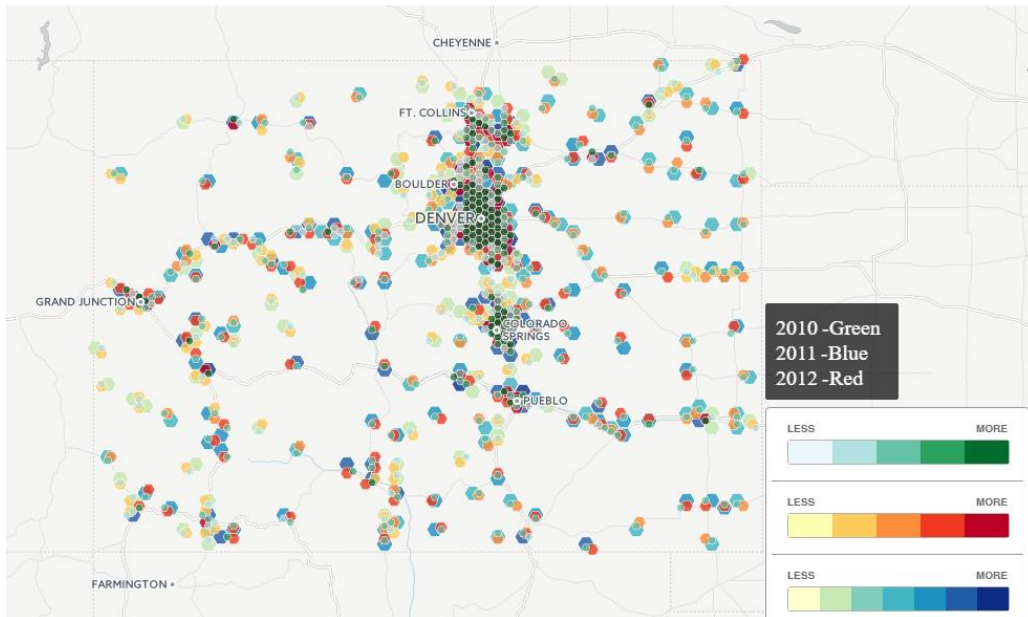
Figure 12: Districts with low income students over three years
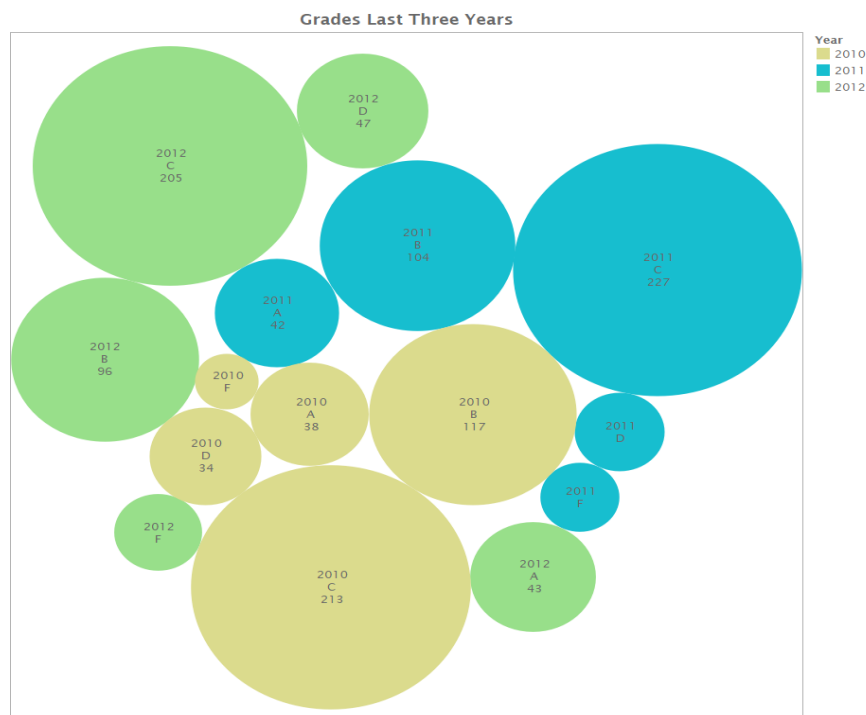
Figure 13: number of schools at each grade level each year

Another feature that was tracked over the three years was the graduation rate of students, this data is visualized in figure 14. These rates are fairly similar over the three recorded years, with a slight increase in 2012. Hopefully, this is a sign of an increased effectiveness of the school system in producing graduates and that the trend from 2011 to 2012 will continue in later years.

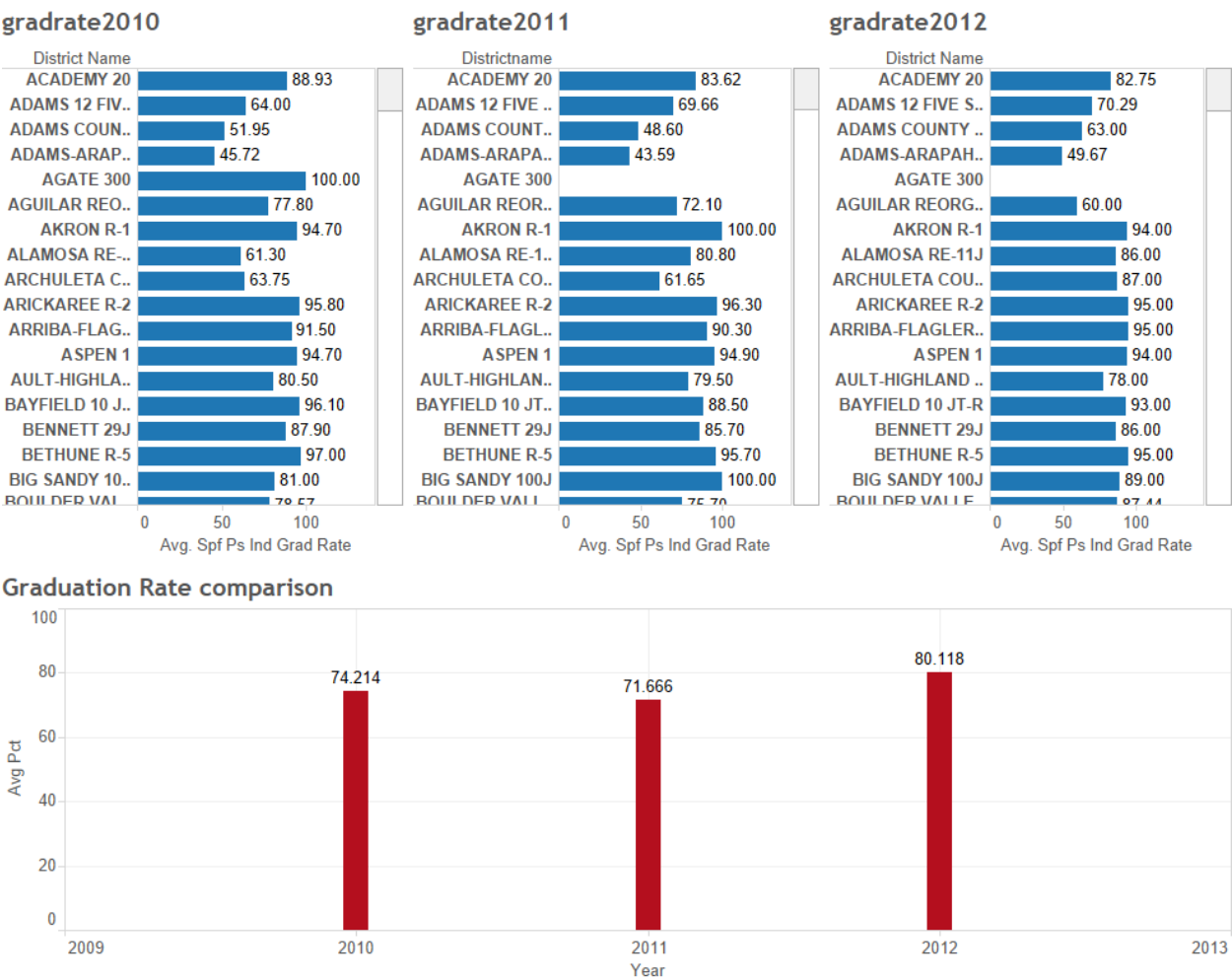| gradrate2010		gradrate2011		gradrate2012	
District Name		Districtname		District Name	
ACADEMY 20	88.93	ACADEMY 20	83.62	ACADEMY 20	82.75
ADAMS 12 FIV..	64.00	ADAMS 12 FIVE ..	69.66	ADAMS 12 FIVE S..	70.29
ADAMS COUN..	51.95	ADAMS COUNT..	48.60	ADAMS COUNTY ..	63.00
ADAMS-ARAP..	45.72	ADAMS-ARAPA..	43.59	ADAMS-ARAPAH..	49.67
AGATE 300	100.00	AGATE 300		AGATE 300	
AGUILAR REO..	77.80	AGUILAR REOR..	72.10	AGUILAR REORG..	60.00
AKRON R-1	94.70	AKRON R-1	100.00	AKRON R-1	94.00
ALAMOSA RE-..	61.30	ALAMOSA RE-1..	80.80	ALAMOSA RE-11J	86.00
ARCHULETA C..	63.75	ARCHULETA CO..	61.65	ARCHULETA COU..	87.00
ARICKAREE R-2	95.80	ARICKAREE R-2	96.30	ARICKAREE R-2	95.00
ARRIBA-FLAG..	91.50	ARRIBA-FLAGL..	90.30	ARRIBA-FLAGLER..	95.00
ASPEN 1	94.70	ASPEN 1	94.90	ASPEN 1	94.00
AULT-HIGHLA..	80.50	AULT-HIGHLAN..	79.50	AULT-HIGHLAND ..	78.00
BAYFIELD 10 J..	96.10	BAYFIELD 10 JT..	88.50	BAYFIELD 10 JT-R	93.00
BENNETT 29J	87.90	BENNETT 29J	85.70	BENNETT 29J	86.00
BETHUNE R-5	97.00	BETHUNE R-5	95.70	BETHUNE R-5	95.00
BIG SANDY 10..	81.00	BIG SANDY 100J	100.00	BIG SANDY 100J	89.00
BOULDER VAL..	78.57	BOULDER VALL..	75.70	BOULDER VALLE..	87.44

Graduation Rate comparison

Figure 14: Graduation rate over multiple years

Overall, some districts had a particularly marked growth rate (figure 15). This is a useful result for parents and administrators. The parents will be able to expect continued growth in these school over the next few years. Administrators can examine what actions these schools took to achieve this growth, and recommend similar actions to other schools in similar situations.
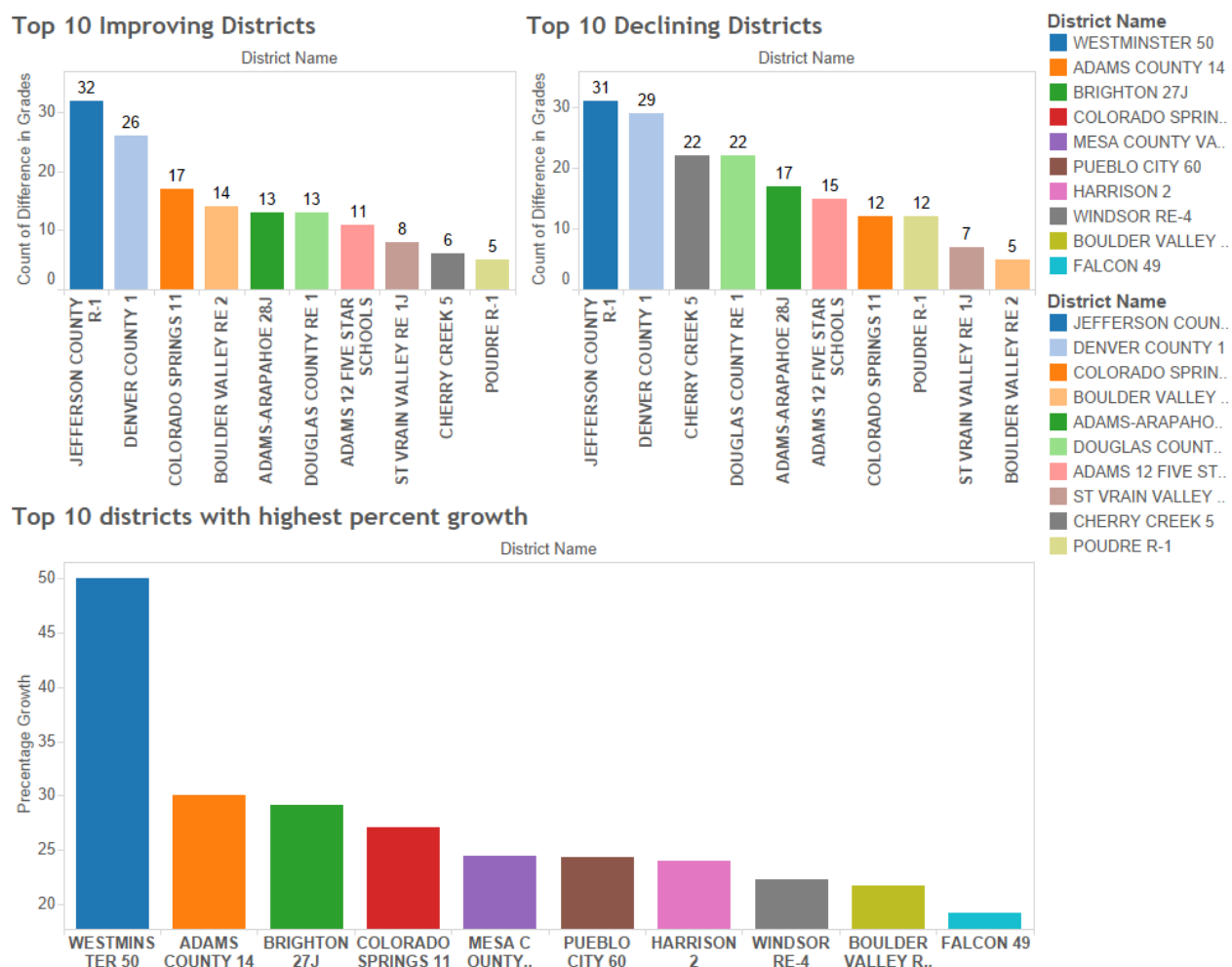
Figure 15: Top 10 growing and declining districts

Conclusions:

Some potential conclusions have been stated throughout the course of the exploration of the data in the results. There are two primary conclusions that this visualization of the data has caused. The first is for the educators and administrators of the Colorado State education system, and the second for parents seeking to enroll their children in this education system.

Educators and administrators should focus their efforts on lower income schools whose populations are primarily from minority demographics. This is because on average, those schools are more likely to be struggling to keep up with higher standards for a quality education. Some of these potential ideas for improving the quality of these

schools can be gained by looking at currently improving schools and examining which actions those schools took to improve their overall grade. We also conclude that while there is a high concentration of low graded schools in Denver, that other locations should not be ignored to provide a higher quality of education overall in the state.

Parents should be assured that regardless of their physical location, it is possible to enroll their child in a high quality school. While there are some areas that are focus points for either lower or higher grades, good schools can be found throughout the State of Colorado. With the visualizations shown in this document, those parents can search for the districts which most synergize with their child. We also have included table 2 to directly display the top 15 schools for each of elementary, middle, and high schools.

Table 2: Top 15 schools in each category

Top 15 Elementary Schools	Top 15 Middle schools	Top 15 High Schools
AURORA QUEST K-8	STARGATE CHARTER SCHOOL	THE VANGUARD SCHOOL (HIGH)
POLARIS AT EBERT ELEMENTARY SCHOOL	THE CLASSICAL ACADEMY MIDDLE SCHOOL	PEETZ JUNIOR-SENIOR HIGH SCHOOL
STECK ELEMENTARY SCHOOL	AURORA FRONTIER K-8	LIBERTY COMMON CHARTER SCHOOL
COYOTE RIDGE ELEMENTARY SCHOOL	HULSTROM OPTIONS K-8 SCHOOL	DSST: STAPLETON
HULSTROM OPTIONS K-8 SCHOOL	AURORA ACADEMY CHARTER SCHOOL	RIDGWAY HIGH SCHOOL
BEAR CREEK ELEMENTARY SCHOOL	AURORA QUEST K-8	THE CLASSICAL ACADEMY HIGH SCHOOL
COTTONWOOD CREEK ELEMENTARY SCHOOL	WESTLAKE MIDDLE SCHOOL	RALSTON VALLEY SENIOR HIGH SCHOOL
BRADFORD INTERMEDIATE SCHOOL	CLYDE MILLER K-8	KIOWA HIGH SCHOOL
ELK CREEK ELEMENTARY SCHOOL	DISCOVERY CANYON CAMPUS SCHOOL	LYONS MIDDLE/SENIOR HIGH SCHOOL
TRAUT CORE ELEMENTARY SCHOOL	STEM SCHOOL	THOMAS MACLAREN STATE CHARTER SCHOOL
ZACH ELEMENTARY SCHOOL	VANGUARD CLASSICAL SCHOOL	CRESTED BUTTE COMMUNITY SCHOOL
CHINOOK TRAIL ELEMENTARY SCHOOL	AKRON ELEMENTARY SCHOOL	D'EVELYN JUNIOR/SENIOR HIGH SCHOOL
BOULDER COMMUNITY SCHOOL/INTEGRATED STUDIES	EAGLEVIEW MIDDLE SCHOOL	RIDGEVIEW CLASSICAL CHARTER SCHOOLS
HORIZONS K-8 SCHOOL	CHALLENGER MIDDLE SCHOOL	PEAK TO PEAK CHARTER SCHOOL
SUPERIOR ELEMENTARY SCHOOL	MOUNTAIN RIDGE MIDDLE SCHOOL	BUENA VISTA HIGH SCHOOL

Individual Tasks:

Data Cleaning and Aggregation was done by each person individually. Visualizations were divided into three sections.

Jaideep focused on analyzing the data of Location data of School Ratings in Colorado. He also built all the visualizations of the same.

I worked on analyzing and building visualizations of the growth of the schools, districts over time and finding out the trends over three years from 2010 to 2012.

Paul investigated how demographics and income affected the ratings of schools and also built visualizations on the same.

The three of us contributed equally on the project report. We developed visualizations on the individual sections and merged all of them together into a document.

References:

Kaggle Competition website:

https://www.kaggle.com/c/visualize-the-state-of-education-in-colorado

Colorado Department of Education website:

http://www.cde.state.co.us/

Colorado School Grades data repository:

http://www.coloradoschoolgrades.com/