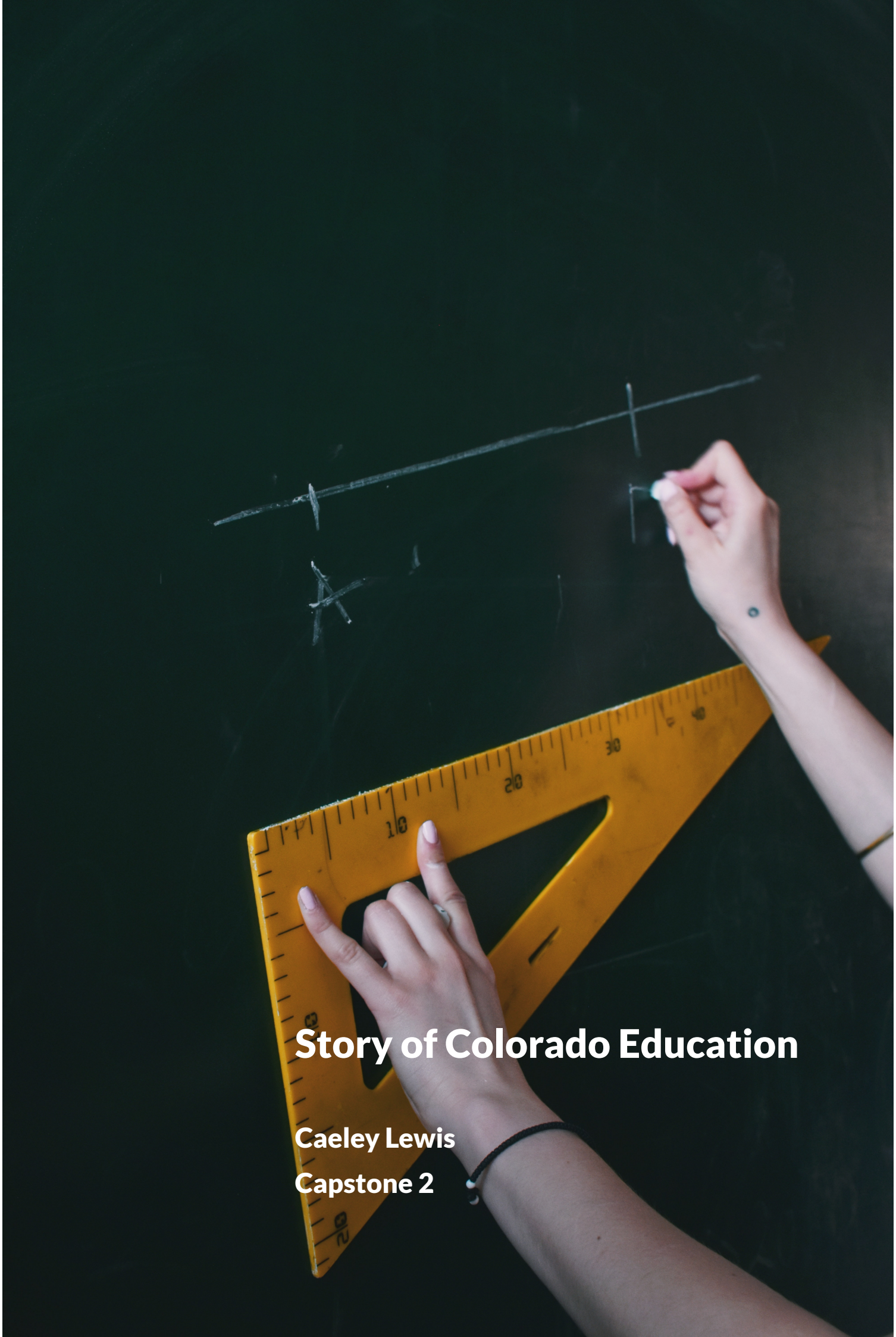

Story of Colorado Education

Caeley Lewis

Capstone 2

Contents

1 INTRODUCTION

According to the American Psychological Association, ”Inadequate education and increased dropout rates affect children’s academic achievement, perpetuating the low-[socioeconomic status] status of the community [1]. Therefore, it is critical that polciy makers, schools, activist groups, and parents understand clearly how our schools are performing and develop ideas of how to improve them .The purpose of our project is to accomplish that by providing a highly visual report of performance of its schools and identify some data-supported ways of providing better education to Colorado’s most vulnerable students.

1.1 GOALS AND OBJECTIVES

We seek to make recommendations

- To improve education for different ethnic communities
- To address child poverty
- To support further analysis

1.2 PROJECT OUTLINE

In order to accomplish our goals and objectives we have created an outline to

1. Perform a thorough visual analysis followed by statistical analysis to test developing hypothese
2. Examine how poverty is changing over time, how we are addressing it, and whether or not our efforts are improving performance
3. Determine if schools’ diversity is representative and how we are meeting the needs of different communitities
4. Analyze spending, determine if schools are spending fairly, and what effect it has on performance
5. Determine how overall performance is changing over time.

2 DATA

We have compiled data data from several sources: performance and socioeconomic data for Colorado schools offered by the non-profit organization, Colorado School Grades [5], census

SAIPE data regarding the number of children in poverty in school districts [2], and district expenditures from the state of Colorado [4]. Our analysis will be limited to the 2010-2012 school years of non-BOCES schools because that is all we have been provided from Colorado School Grades; despite this limitation, we have attempted to make our analysis scalable should more data become available in the future. Sadly, the process of cleaning, joining, and demonstrating the validity of our data took about 90% of the total time devoted to the project.

2.1 CLEANING DATA

The data used for this project was incredibly messy. In particular, the data from Colorado school grades was often inconsistent and redundant, the expenditures data was provided in a report style format that required heavy manipulation to become usable, and the three different data sources were difficult though not impossible to join.

2.1.1 EXPENDITURES FORMAT

In the Appendix Figure 7.6, we see data for the expenditures that is in a report style format. In other words, information for each schools total spending and the spneding per pupil is given in a way that is separated by blank lines in a presentable way. While this data is easy to look at, it is completely unusable for analysis. Furthermore, the tables were given in an outdated .xls format, and some of the spending numbers provided had commas and parentheses in them that had to be removed. For our analysis, we kept the total spending for each district and the spending per pupil. We converted the data into a tall format as shown in the Appendix Table 7.2. This process took an exorbitant amount of time to accomplish correctly.

2.1.2 COLORADO SCHOOL GRADES INCONSISTENCY

Colorado School Grades data had even more problems. The most concerning of which was that for every school year and every dataset, different number of schools information were recorded. For example, in the 2010-2011 school year, the dataset that detailed the change and improvement of schools reported 2082 schools while the data that specified the percentage of students receiving free and reduced lunches was 1753.

In addition, performance indicators were sometimes different for each year. The college readiness indicator, for instance, was specified as 0 and 1 for some years and 1 and 2 for others. Each were mapped to 0 and 1 for False and True respectively. Rank was also determined differently for each school year. In one school year, rank for middle and elementary schools was combined and for another separate. Furthermore, some datasets contained information for state and district results while others did not. This information was removed.

2.1.3 COLORADO SCHOOL GRADES REDUNDANCY

Other issues in the Colorado School Grades included redundancies. Most importantly, there are two indicators for growth. There is a growth direction and a growth grade. There are achievement directions and achievement grades, and there are overall directions and overall grades. Because the grades are a broader variable, we consider this more useful.

2.2 JOINING DATA

The biggest issue we encountered in joining data arose from poor similarity of district names. Not only did the data from each source have different formats for the district, but there was vast inconsistency with different datasets from Colorado School Grades. Because of the number of districts and the similarities between distinct districts, we could not employ scalable strategies for text cleaning. As a result, we were only able to apply patchwork by addressing each issue individually. To reduce this amount of work, we found high similarity between the dataset containing demographic information with the census and expenditures datasets. We made a list of changes necessary to join them completely, and then the rest of the Colorado School Grades datasets were joined on school id's.

However, this was also not a simple task. While most schools in the Colorado School Grades are distinctly elementary, middle, or high schools, some are combined schools and this created issues with distinct school id's. For example a K-12 school would have a separate entry for the elementary, middle, and high school components, but each entry was written with the same school id.

2.3 DATA VALIDATION

Finally, we made an effort to demonstrate the validity of our data. The data gathered from the census and the Colorado School Grades data each contained information about the number of students in schools. In order to justify joining this data, we compared their distributions, as shown in Appendix Figure 7.7, and found good consistency.

We also sought to verify that the demographic information provided by Colorado School Grades was reflective of 2010 Census information about race in Colorado. As we can see in Appendix Table 7.4, the relative percentages for our calculated values from the percentages and size of schools are similar to information gathered in the 2010 Census [3].

3 ANALYSIS

In our analysis, we examined, poverty, diversity, spending, and performance in schools. We produced various figures and tables and used statistical tests to make comparisons of distributions. Specifically, we used t-tests for data with skew less than 1, permutation tests for skew greater than 1, and Mann-Whitney U rank testing for imbalanced classes.

3.1 POVERTY

3.1.1 INCREASING POVERTY OVER TIME

From Appendix Figure 7.8, we can see that poverty seems to be increasing over time. A p-value of 0.030 makes us reasonably confident that even if the short three-year time period poverty is increasing.

3.1.2 MEETING NEEDS CHILDREN IN POVERTY

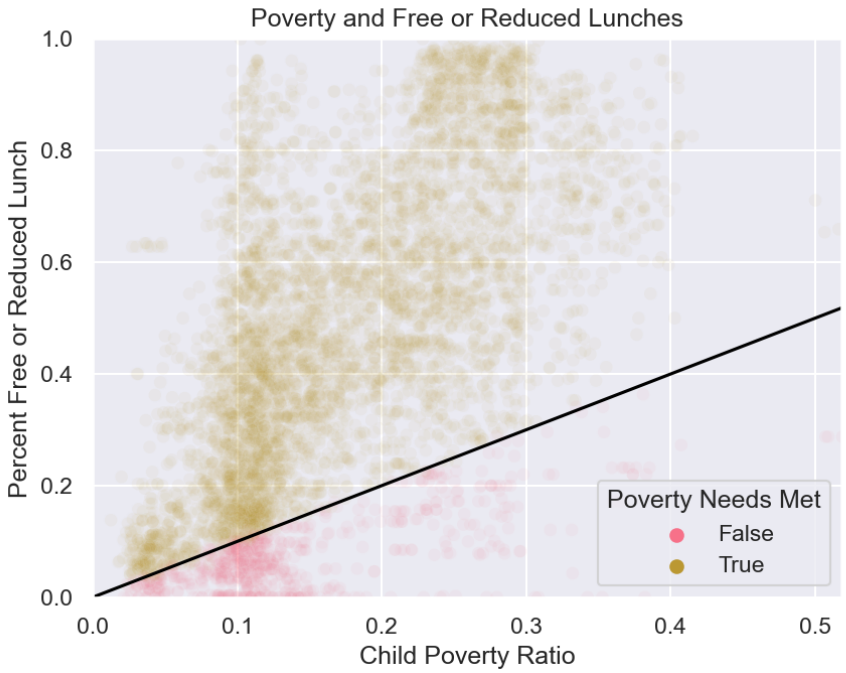

FIGURE 3.1: *0.67% of students in poverty are not receiving free or reduced lunches.*

Because poverty appears to be increasing, we sought to determine if children in poverty are receiving free and reduced lunches in Colorado. Figure 3.1 shows that the percentage of free and reduced lunch increases at a faster rate than child poverty. We can be assured that children in poverty are not going hungry at our schools. Another feature we observe in Figure 3.1 is that there are weak clusters at higher percentages of free and reduced lunches and lower

percentages. This could be worth further investigation, as it suggests that some schools may be more willing to provide free and reduced lunches than others. In fact, There are several schools that offer no percent free and reduced lunches despite child poverty. We find that approximately 0.697% of students in poverty fall beneath the line and are not receiving free or reduced lunches.

We were also interested to see if impoverished communities received more instructional spending on average. Appendix Figure 7.10 shows a funnel effect. Instructional spending in the top 5th percentile decreases as poverty increases while the instructional spending for the lower 95th percentile increases with poverty.

3.1.3 POVERTY AND PERFORMANCE

In order to determine if child poverty is creating an issue for education, we looked at all performance metrics. While strong correlations did not exist in achievement and growth grades. Poverty was well reflected in college readiness and ACT grades as shown in Figure 3.2.

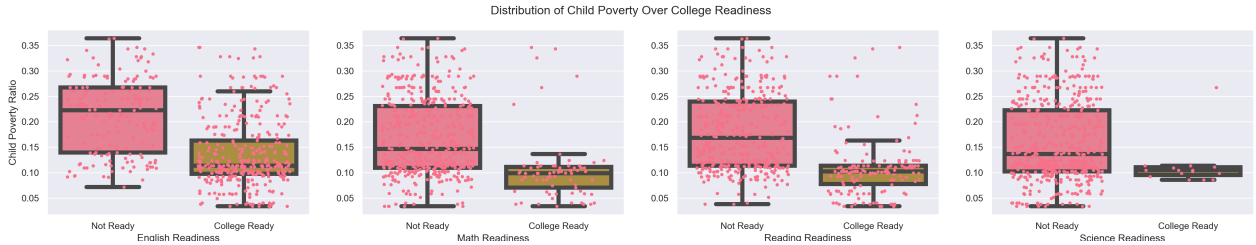

FIGURE 3.2: *In every category, increased poverty is linked with less college readiness.*

3.2 DIVERSITY

3.2.1 LACK OF REPRESENTATIVE SCHOOLS

We were interested in diversity in schools, and we sought to understand if the diversity in schools was representative of state level demographics. Looking at Figure 7.12, we can see that schools are not representative. We would hope to see a normal distribution around the state demographics.

3.2.2 FAILURE TO MEET NEEDS OF HISPANIC COMMUNITIES

Ideally, diversity would not be strongly correlated with performance metrics. However, in the data found that ethnicity was highly correlated. In Appendix Figure 7.13, we can see that among every ethnicity, college readiness was distinctly separated over time. In general, white communities had positive correlations while non-white communities struggle. Hispanic communities struggle the most. In fact, across every single performance metric, the percentage of students identifying as Hispanic is negatively correlated as shown in Figure 3.3.

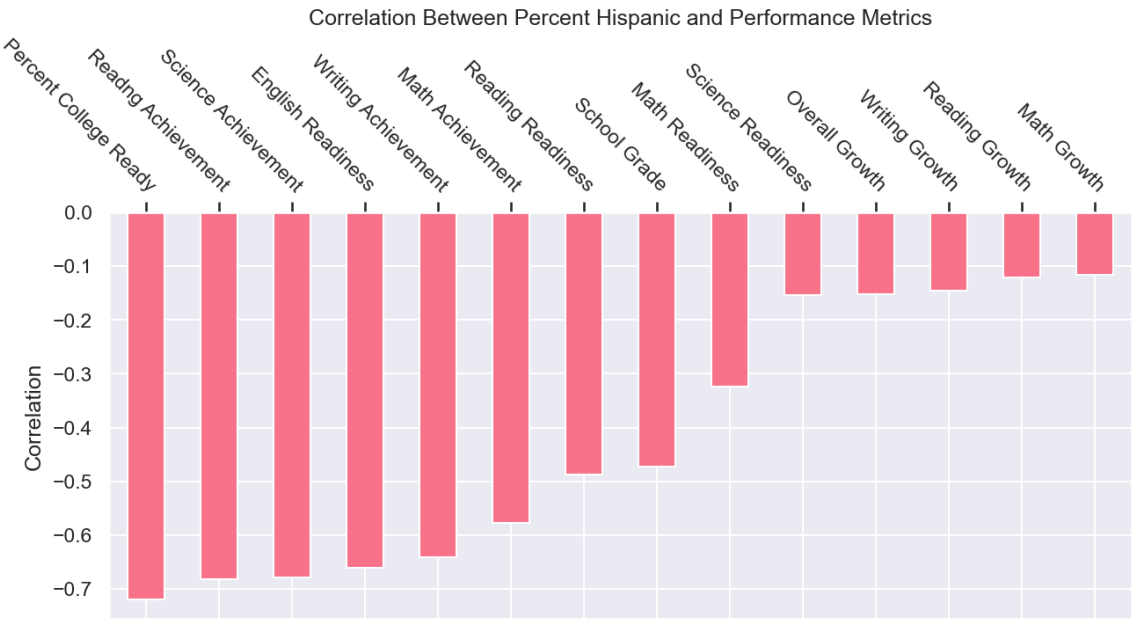

FIGURE 3.3: *Across every performance metric, schools with a higher percentage of Hispanic student are negatively correlated with their performance metrics.*

3.3 SPENDING

3.3.1 UNFAIR SPENDING

One of our goals was to see how much schools were spending per pupil and determine if that spending was indeed fair. The first thing that we noticed in our analysis was that the distribution of spending per pupil is highly skewed. In Table 3.1, we can see that there is about $30,000 difference between the minimum and 97.5 percentile and $70,000 difference between the 97.5 percentile and the maximum.

TABLE 3.1

Quantile	Total Spending Per Pupil
0.0%	$6090.00
2.5%	$8346.90
25.0%	$10116.25
50.0%	$12401.50
75.0%	$15827.00
97.5%	$36131.82
100.0%	$108789.00

The question that follows is which communities are receiving more funds per pupil. In a comparison between majority and minority white communities, as shown in Figure 3.4, we find that majority white communities receive more spending on average than minority white communities with confidence in the p-value of 2.08e-9.

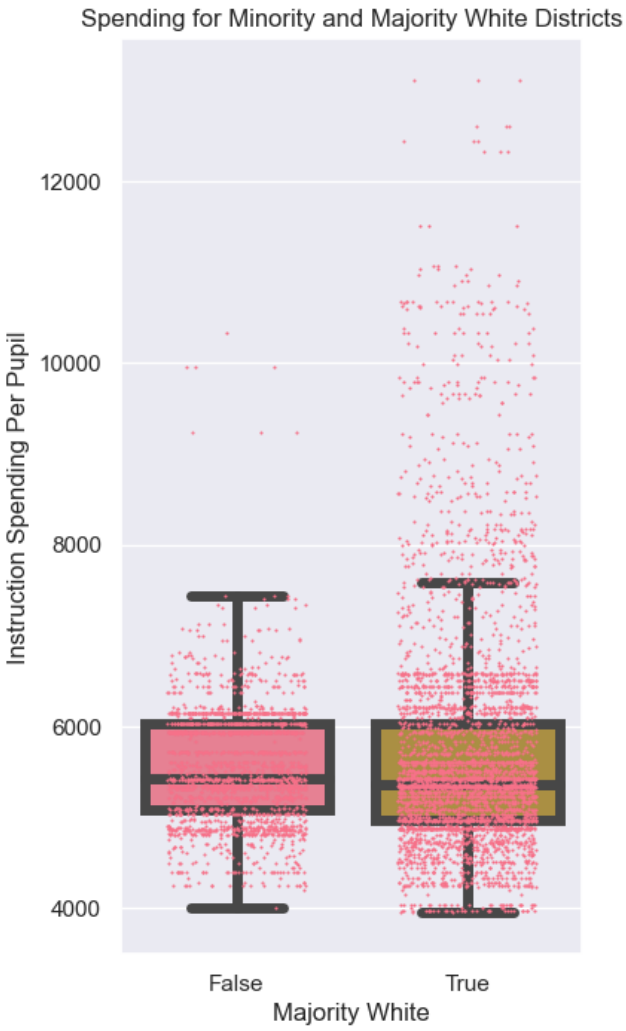

FIGURE 3.4: *Outliers for spending are primarily in the majority white schools. A t-test to determine if majority white schools receive more funds per pupil than minority white districts yielded a p-value 2.08e-9.*

3.3.2 SPENDING IMPROVES COLLEGE READINESS

Now that we have observed that spending is unfair, we would like to understand the effect spending has on performance. In Figure 3.5, we see that spending is indeed less, with the exception of English in 2010, for passing students in every category. We obtain p-values that are statistically significant in math, reading, and science that confirm what we can see visually.

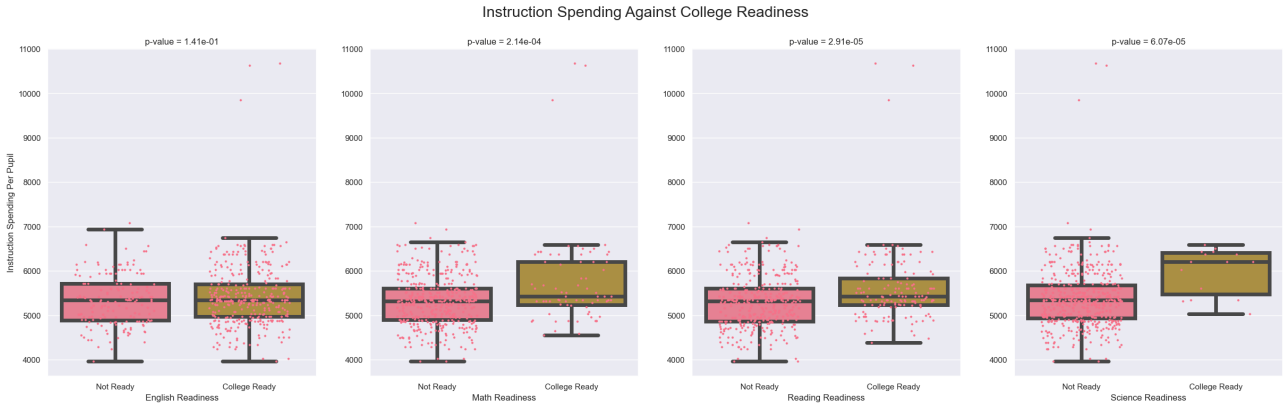

FIGURE 3.5: *In every category, increased spending is linked with more college readiness in every category and nearly every year. Mann-Whitney U rank testing yields p-values of 0.141 for english readiness, 0.00021 for math readiness, 2.91e-5 for reading readiness, 6.07e-5 for science readiness.*

3.4 PERFORMANCE OVER TIME

Finally, we would like to understand how performance is changing over time. Because achievment and growth grades are given on a curve, we cannot observe its change over time. However, we can analyze college readiness. Appendix Figure 7.17 shows that college readiness is not improving over time. It also makes the difference between english and science readiness clearly apparent. While about 70% of students are english ready, less than 5% of students are science ready.

4 RESULTS

With our analysis we can make three important recommendations: reach out to Hispanic communities, address increasing poverty, and improve data quality.

4.1 IMPROVE EDUCATION FOR HISPANIC COMMUNITIES

Our analysis showed that academic performance is highly correlated with ethnic demographic information. In general, majority white communities are statistically shown to receive more spending and have higher metrics than minority white communities. We also showed, that increased spending in and of itself is statistically linked with improved college readiness. High majority Hispanic schools are the most affected by these disparities. In fact, increasing percentage of Hispanic students is negatively correlated with every performance metric and they are the highest correlations among all variables in the data.

Without knowing the inner workings of Colorado Schools, we suspect two barriers Hispanic

students are facing, language and poverty. After all, ACT's cannot be taken in other languages and professional ACT prep cannot happen if students cannot afford tutors. We also suspect that these issues are exacerbated by a lack of diversity in schools. Instead of observing a normal distribution of ethnicity around the state average, we observe that schools are more segregated. Despite making up about 25% of the population there are over sixty schools that are more than 90% Hispanic. There is evidence that diversity improves performance and we would seek to understand the underlying causes for such unintentional segregation such as affordable housing. The Board of Education's ability to address these problems, but non-profits, parents, and other government officials can.

4.2 ADDRESS POVERTY

We also showed that child poverty is increasing in Colorado even within the short three-year time period examined and that poverty is linked with with performance despite the fact that impoverished communities tend to receive more funds to spend on instruction per pupil. Without information about teaching quality and school resources available, we can only recommend from our data that the distribution of spending be revisited. It is imperative that intentional measures are taken to improve the quality of education for communities of low socioeconomic status.

On the other hand, we also concluded that there is good policy in place to provide free and reduced lunches with only about 0.6% students in poverty failing to receive free and reduced lunches.

4.3 IMPROVE DATA QUALITY AND RANGE OF AVAILABLE DATA

The data we gathered from the Census and the State of Colorado improved our analysis of the data from Colorado School Grades; however, it significantly increased the difficulty of cleaning and joining data. This could have been made easier simply if the data from Colorado School Grades had been more self-consistent. This could likely be improved by making the setup this data more relational.

Furthermore, with the data provided we have quite a few performance metrics available, but we have very few features evaluate them with. Important features that would be worth analyzing is the educational strategies employed by the schools, teacher churn, teacher qualifications, changes in leadership, and descriptions of whether or not English is a first language for non-white ethnic groups.

5 CONCLUSION

Altogether, our work has shown that Colorado schools are prioritizing providing free and reduced lunches, majority Hispanic schools and schools with more poverty are not having their needs met for college readiness; however, our work is not done. As mentioned earlier, we need to better understand the mechanisms for why high percentage Hispanic schools and more impoverished schools are performing more poorly. Right now, the only mechanism we can claim is a spending one. In addition, we may want to consider ways to make information easily accessible

6 BIBLIOGRAPHY

References

[1] American Psychological Association. "Education and Socioeconomic Status". In: *Public Interest Directorate* (2017). URL: https://www.apa.org/pi/ses/resources/publications/education.

[2] US Census. *SAIPE Datasets*. Tech. rep. US Census, 2010-2012. URL: https://www.census.gov/programs-surveys/saipe/data/datasets.All.List_1743592724.html#list-tab-List_1743592724.

[3] *Colorado*. Tech. rep. US Census, 2010. URL: https://www.census.gov/geographies/reference-files/2010/geo/state-local-geo-guides-2010/colorado.html.

[4] *District Revenues And Expenditures [Comparison of All Expenditures, Table IVC*. Tech. rep. Colorado Department of Education, 2010-2012. URL: http://www.cde.state.co.us/cdefinance/revexp.

[5] *Visualize the State of Education in Colorado [Kaggle Competition]*. Tech. rep. Colorado School Grades, 2012. URL: https://www.kaggle.com/competitions/visualize-the-state-of-education-in-colorado/overview.

7 APPENDIX

7.1 TABLES

	district_name	county	instruction_total	support_total	...	community_per_pupil	other_per_pupil	sum_per_pupil	year
0	MAPLETON 1	ADAMS	39962942	23760636	...	33	761	9653	2010
1	ADAMS 12 FIVE STAR	ADAMS	220263102	130356958	...	23	2087	10837	2010
2	ADAMS COUNTY 14	ADAMS	34792431	34598159	...	100	1120	11104	2010
3	BRIGHTON 27J	ADAMS	62557927	50921503	...	15	1296	9287	2010
4	BENNETT 29J	ADAMS	4853879	3540284	...	0	939	8784	2010

TABLE 7.2: *Expenditures data after being manipulated into a tall data format.*

Quantile	Total Spending Per Pupil
0.0%	$6090.00
2.5%	$8346.90
25.0%	$10116.25
50.0%	$12401.50
75.0%	$15827.00
97.5%	$36131.82
100.0%	$108789.00

TABLE 7.3: *Quantiles for spending per pupil show a heavily skewed dataset. Between the max and the 97.5% there is an order of magnitude difference*

(a) *Calculated*

Race	Statewide Percentage
American Indian	0.008598
Asian	0.030238
Black	0.049605
Hispanic	0.312047
White	0.568823
Pacific Islander	0.002154
Two Or More	0.028535

(b) *2010 Census Information*

Race	Statewide Percentage
American Indian	0.017
Asian	0.038
Black	0.047
Hispanic	0.225
White	0.665
Pacific Islander	0.002
Two Or More	0.034

TABLE 7.4: *Statewide demographics determined from taking the sum of the multiplication of percentages in each school by the total number of students and dividing by the sum of all students*

7.2 FIGURES

TABLE IVC COMPARISON OF ALL PROGRAM EXPENDITURES 2012-2013

	COUNTY	DISTRICT/ BOCES	Instruction Services	Total Support Services	Community Services	Other Expenditures	Total Expenditures
	ADAMS	MAPLETON 1					
$	Amount		38,592,373	25,709,054	779,878	4,153,285	69,234,590
$	Per Pupil		5,081	3,385	103	547	9,115
%	All Funds		55.7	37.1	1.1	6.0	100.0
	ADAMS	ADAMS 12 FIVE STAR					
$	Amount		221,051,769	120,585,070	1,335,918	51,372,445	394,345,203
$	Per Pupil		5,347	2,917	32	1,243	9,538
%	All Funds		56.1	30.6	0.3	13.0	100.0
	ADAMS	ADAMS COUNTY 14					
$	Amount		38,196,626	36,456,267	413,091	77,970,868	153,036,852
$	Per Pupil		5,516	5,265	60	11,260	22,100
%	All Funds		25.0	23.8	0.3	50.9	100.0
	ADAMS	BRIGHTON 27J					
$	Amount		67,423,272	52,627,926	247,509	24,647,115	144,945,822
$	Per Pupil		4,392	3,428	16	1,605	9,441
%	All Funds		46.5	36.3	0.2	17.0	100.0
	ADAMS	BENNETT 29J					
$	Amount		4,894,710	3,252,791	0	988,729	9,136,230
$	Per Pupil		4,766	3,167	0	963	8,895
%	All Funds		53.6	35.6	0.0	10.8	100.0

FIGURE 7.6: *Expenditures data was initially given in a report style format.*

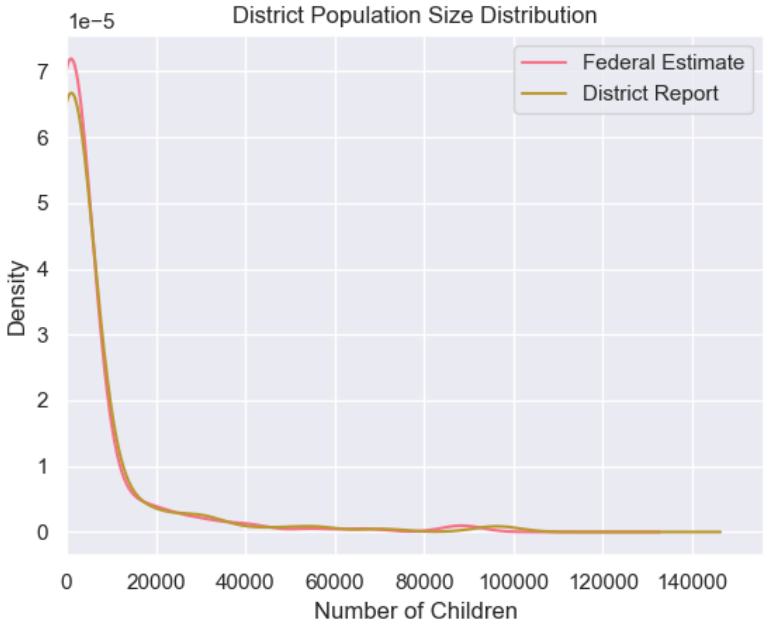

FIGURE 7.7: *A comparison between the distribution of district sizes given by the Census and that provided by Colorado School Grades.*

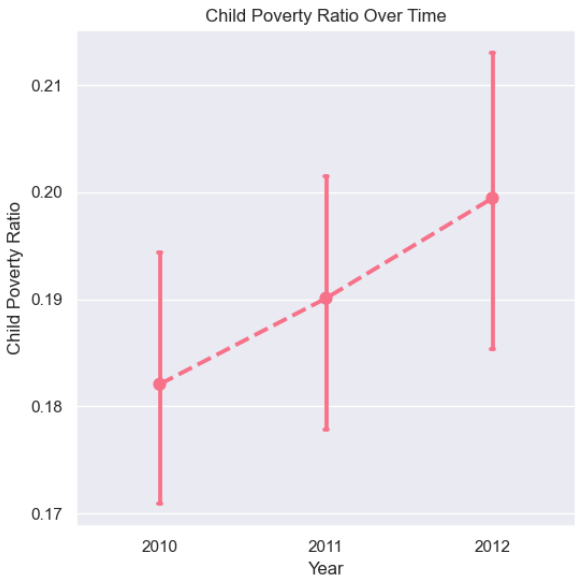

FIGURE 7.8: *Poverty is increasing at a rate that is statistically significant, with a p-value of 0.030, in the three year time frame.*

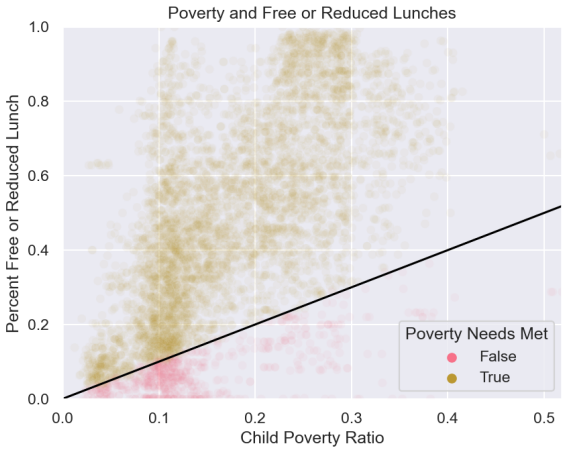

FIGURE 7.9: *0.67% of students in poverty are not receiving free or reduced lunches.*

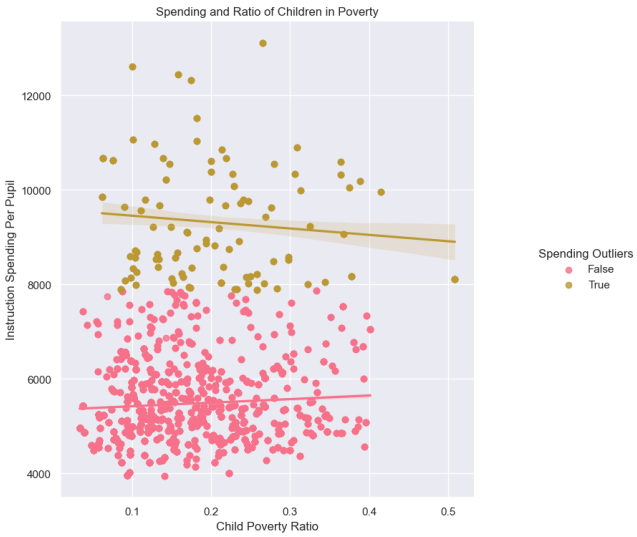

FIGURE 7.10: *There is a funnel effect that spending for the bottom 95 percentile increases as child poverty increases while the top 5th percentile decreases.*

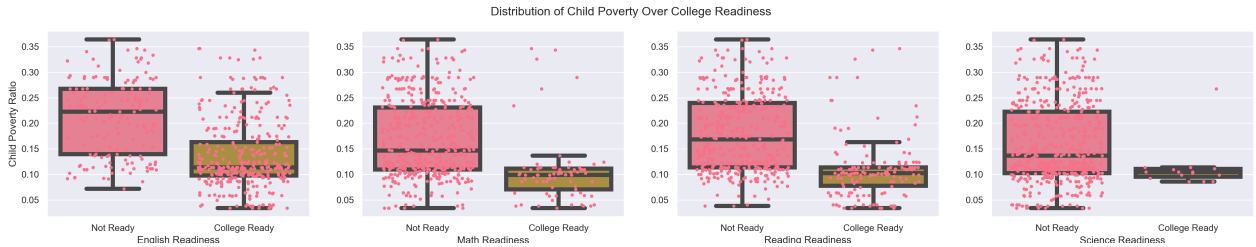

FIGURE 7.11: *In every category, increased poverty is linked with less college readiness.*

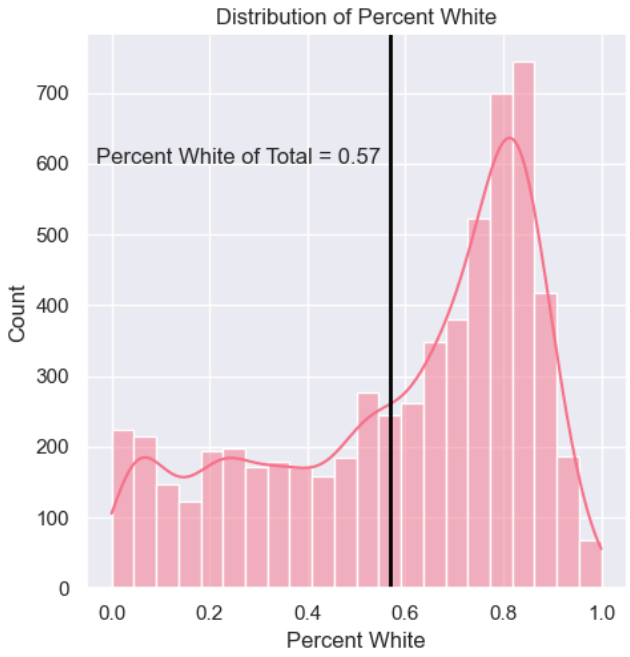

FIGURE 7.12: *The distribution of percent white schools is not centered around the expected mean of 57%.*

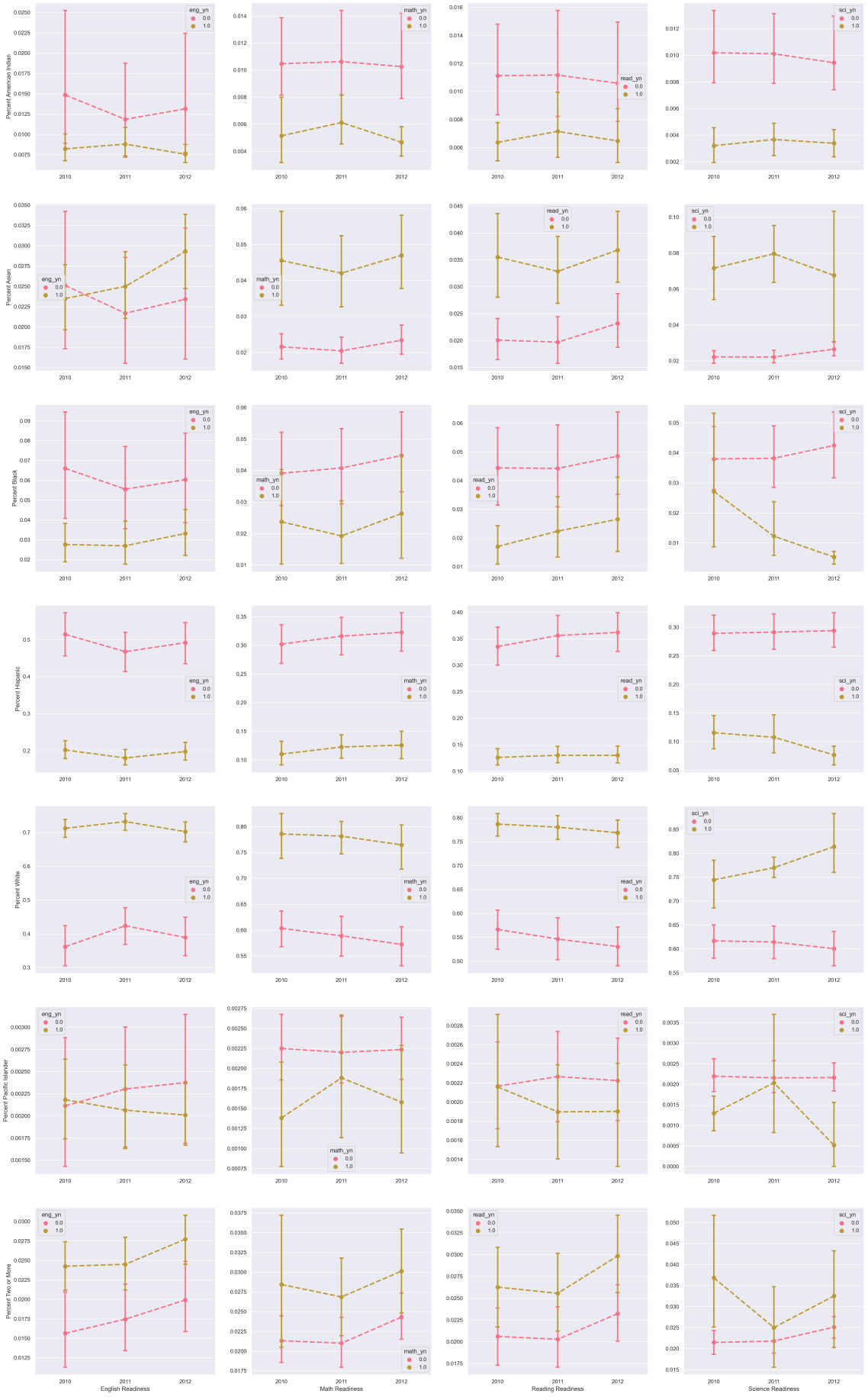

FIGURE 7.13: *College readiness and diversity are highly linked in every category and year.*

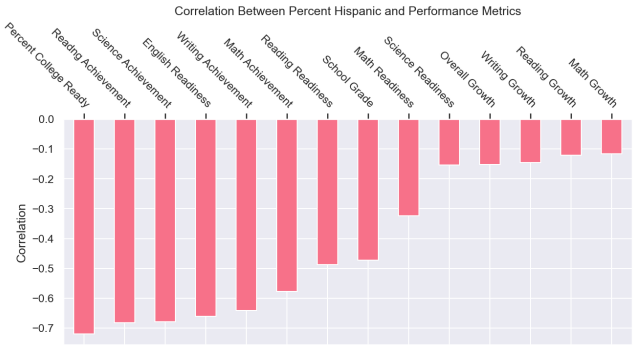

FIGURE 7.14: *Across every performance metric, schools with a higher percentage of Hispanic student are negatively correlated with their performance metrics.*

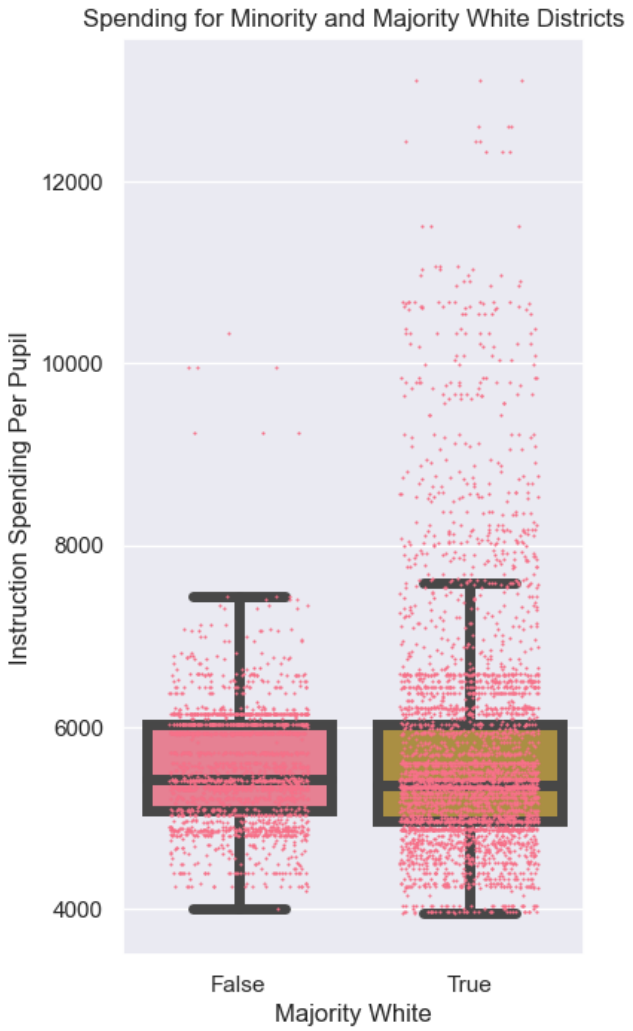

FIGURE 7.15: *The majority of outliers for spending are in the majority white schools. A t-test to determine if majority white schools receive more funds per pupil than minority white districts yielded a p-value 2.08e-9.*

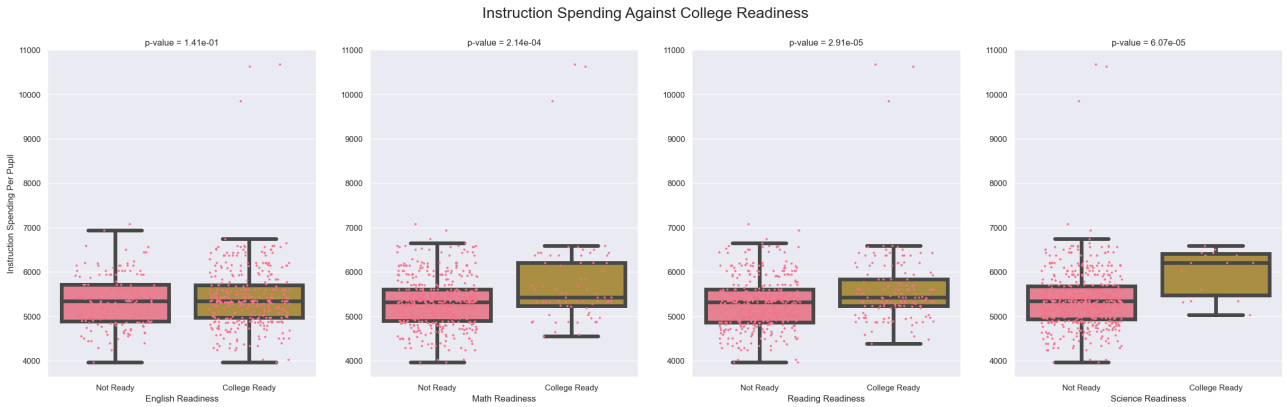

FIGURE 7.16: *In every category, increased spending is linked with more college readiness in every category and nearly every year. Mann-Whitney U rank testing yields p-values of 0.141 for english readiness, 0.00021 for math readiness, 2.91e-5 for reading readiness, 6.07e-5 for science readiness.*

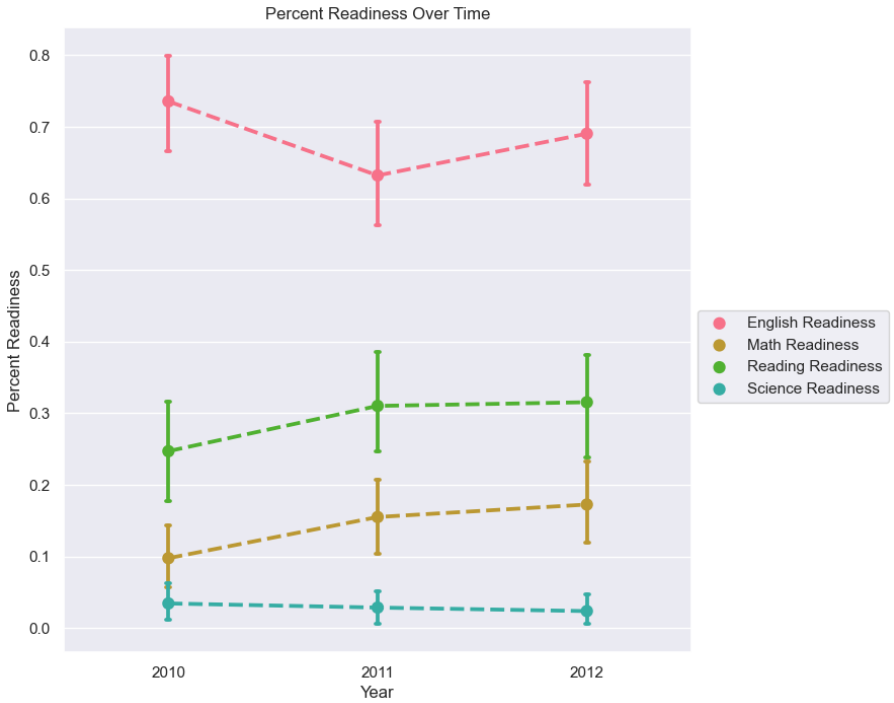

FIGURE 7.17: *Mean college readiness is not changing significantly over time. The percentage of students receiving high english ACT scores is signifcantly higher than any other category particularly science.*