Nicole King

December 10, 2012

MATH 3220

Data Mining Final Report

Missouri School District Funding and Quality

Executive Summary

This experiment will analyze the state of Missouri's educational funding in each school district and compare it to the quality of school. The purpose of the study is to identify the relationship among the funding per student and the graduation rates of school districts in Missouri to assess if there is a positive correlation. This experiment requires the use of four datasets which focus on: educational funding among states in the United States, educational funding among school districts in the United States, populations of Missouri school districts, and graduation rates of Missouri school districts. Through computational work, this experiment allows us to draw conclusions and analyze the relationship between funding and quality of schools. This experiment hones in on urban and rural school districts to make further conclusions and compare three correlation coefficients. Overall, this study found that there is not strong enough evidence that supports a positive correlation among funding per student and quality of all Missouri school districts.

Problem Description

The experiment's objective is to compare the funding per student among the school districts within the state of Missouri to the quality of these schools and assess if there is a correlation between these two factors. Overall, this problem involves working with raw data to

an extent that allows one to make logical and reasonable opinions from it with this research as solid evidence.

Analysis Technique

The objective of this problem regards comparing the quality of the 523 school districts in Missouri in 2007 to the amount of money that each school district receives for each student within its population. This experiment includes taking raw data containing the amount of federal, state, and local funding that Missouri and each of its districts receive and then computing this number into one that represents the financial support per pupil in each district. In this case, the quality of the school district is determined by its graduation rates. The overall goal of this experiment is to carry out computational tasks in order to conclude whether the amount of funding a school district receives per student correlates with the quality of the school district based on graduation rates.

The hypothesis made prior to the execution of this experiment is that the amount of funding that a district receives per student in its population will positively correlate with the quality of the school district according to graduation rates. In all, the more financial support that each student obtains in his/her education affects the quality of the school district and its educational success.

An empirical study performed by Dato and Sebold compared the relationship between school funding and student achievement. In this case, student achievement was measured through standardized test scores. The results of this study concluded that funding per student in the school produces significant advances in the test scores of the students (Sebold & Dato, 1981). Another study that supports this hypothesis describes the process of increasing funding in low-

income school districts to minimize the gap among these schools and other schools that receive more financial support. The SAT test scores of the students attending these low-income school districts were recorded and analyzed to determine if an increase in spending within the school would increase the students' achievement. It was found that the SAT test scores improved in these school districts. Thus, the experiment supports equalization among school district funding because it improves student achievement in the low-income school districts (Card & Payne, January).

In the United States, school districts are systems of public schools that offer education for students in kindergarten through twelfth grade. School districts offer special, vocational, and regular services to students without paying any tuition. These public school systems are locally managed and their geographic size and boundaries are determined by each state and region of the country (U.S. Census Bureau, 2010).

According to the dataset, there were 523 school districts that made up the state of Missouri's public school system in 2007. In 2007, there were 13,754 school districts within the United States (District and School Information, 2012). On average, about 46 percent of revenue to public schools derives from local financial sources. This depends on the tax base that each local district enforces. Some states, like California, monitor the range of funding per pupil across all school districts in order to manage the gap and keep this difference in funding controlled. However, Missouri does not enforce any type of monitoring to cease inequalities among its school districts (Burtless, 1996).

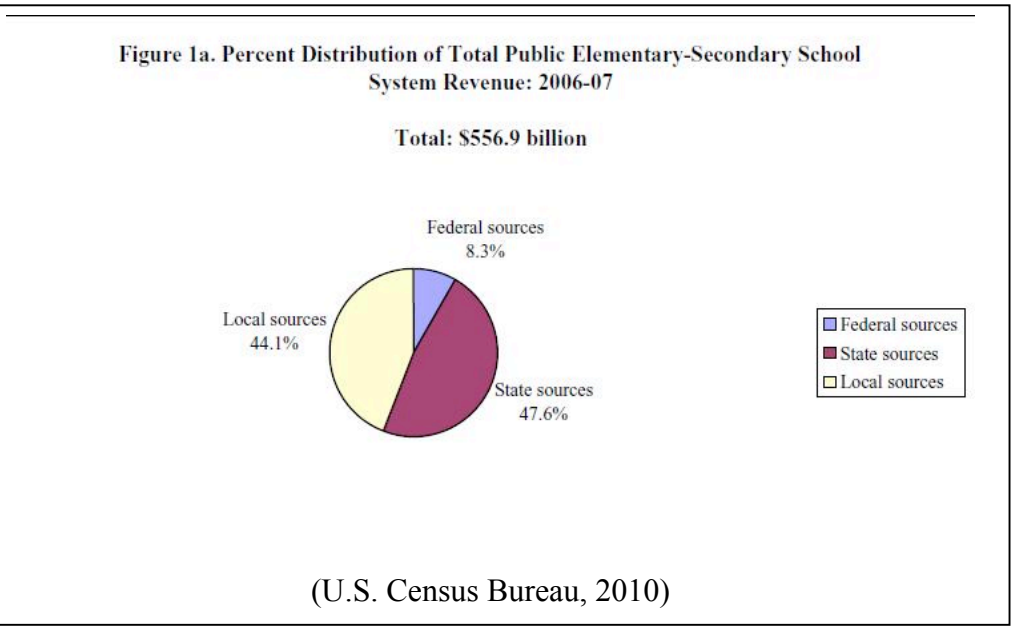
A statistical concept we will be using to analyze our results within this experiment is correlation. Correlation measures the amount of relation between two variables (Lanthier, 2002).

In this experiment, the variables are funding per student and graduation rate. This experiment will produce a correlation coefficient which will be a number in the range of -1 to 1. A correlation coefficient from 0 to 1 is a positive correlation. This means that as the values of one increases, the values of the other variable increases. A positive correlation also means that if one variable decreases, the other variable will also decrease. A negative correlation, which is a coefficient that lies from -1 to 0, means that as one variable increases, the other decreases. Or it also means that if one variable decreases, the other increases (Lanthier, 2002). Additionally, a 0 correlation means that there isn't a recognizable relationship. A number lying near zero may be a positive or negative correlation, but it is not a very strong correlation if it remains close to 0.

The data sources utilized in this experiment contain information regarding public school districts and their funding in the United States of America (U.S. Census Bureau, 2010). The first data set contains the federal, state, and local funding provided to the country as a whole and to each state among elementary through secondary education in America in 2007. The division of the federal, state, and local educational funding in the United States of America is displayed in Figure 1. In total for the United States, 8.3 percent of educational revenue comes from federal funding, 47.6 percent of funding originates from state sources, and 44.1 percent of funding is sources of local support. Furthermore, the total amount of educational funding that each state receives is listed along with each state's current educational spending and debt. The funding is broken down into property taxes, school lunch charges, tuition and transportation. In fact, the data set includes the amount of funding spent per student in each state. There is a graph representing the current spending per state by student in Figure 2. This data set continues to divide the financial expenditures throughout different areas within education including employee salaries and wages, employee benefits, administration salaries, operations and maintenance, and

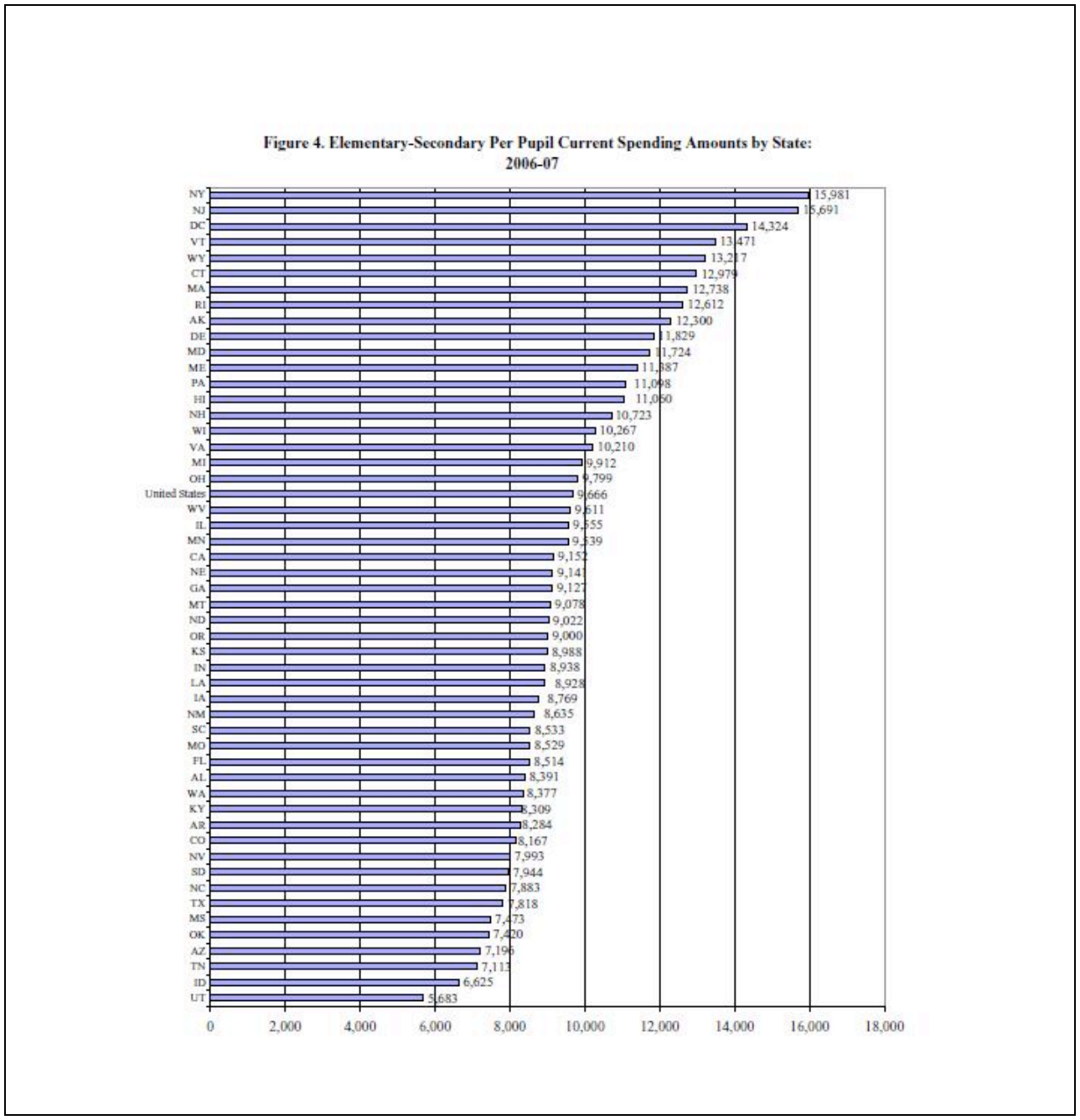
pupil transition wages. The overall concept of this data set is to determine and organize the various funding and expenditures by each state in the United States of America.

Figure 1: Division of Education Funding in the United States



(U.S. Census Bureau, 2010)

Figure 2: Spending Per Student among States in the U.S.A.



(U.S. Census Bureau, 2010)

The second data source displays the school districts in the United States by each ID census. The amount of total revenue given to each state is listed for the year of 2007 (U.S. Census Bureau, 2010). The total federal revenue, state revenue, and local revenue specified to

each school district in the country is also illustrated on a scale of 1,000 in this data set. Within each type of financial support (federal, state, and local), the total revenue in each is separated into particular types of funding according to each district. This data set also determines each school district's expenditures. In all, this data set distinguishes the funding and expenses for all school districts in the country from 2007.

The third data set utilized in this experiment withholds the populations of residents who lived in each school district in 2007 (U.S. Census Bureau, 2010). The school districts are organized by state locations and District ID numbers. Moreover, the initial populations listed are divided into the residents in these districts from five years of age to seventeen years of age. This population determines the amount of students who attend each school district. Lastly, this data set illustrates the number of residents from five years of age to seven years of age who are living in poverty within each school district's population.

The final data set taken into account for this experiment includes the public school districts in the state of Missouri and their dropout rates and graduation rates (District and School Information, 2012). Dropout rates are listed for each district in the state from the years of 2007, 2008, 2009, 2010, and 2011. Graduation rates are also listed among each district in Missouri from the years of 2007, 2008, 2009, 2010, and 2011. Near the top of this data set, the state average dropout rates and graduation rates are displayed from 2007 through 2011. This data set will be the parameter within the experiment that will determine the quality of each public school district within the state of Missouri. There are various metrics of school districts that can be viewed to assess the quality of that district. However, graduation rates data is the metric that this experiment will be reviewing. Since completing graduation is highly significant to all school districts and students, and since it is measureable, dissimilar to a quality metric like teacher

performance, this will be the quality measure to compare the per student funding among the districts in the experiment.

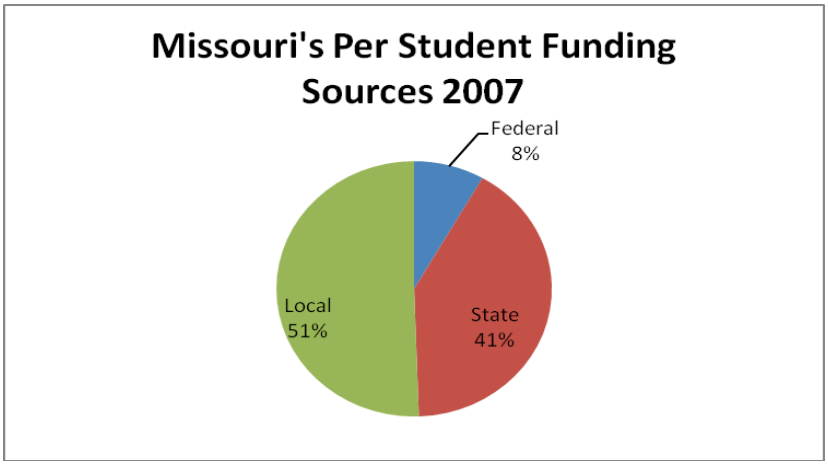
The tools used throughout this experiment include the Excel program and the formulas that this program contains. The experiment calls for the use of online educational resources related to the United States of America and the state of Missouri which are listed on the Bibliography page. This experiment requires the use of a calculator regarding the computational work involved in checking the Excel program formulas used. Additionally, teachers at Grand Center Arts Academy and Professor Aleshunas will all be of help and support throughout this experiment.

The methodology included in this experiment begins with creating a hypothesis that is plausible to be tested and its strength must be able to be determined with the results. Secondly, it is essential to locate research that supports this hypothesis and provides a foundation to make assumptions from. The next step is to research and learn about public education and Missouri public school districts. This background knowledge will facilitate throughout the experiment by providing a deeper understanding of where educational funding comes from, why it is essential within school districts, and where the funding goes after it is received. The following step includes locating raw data sets that may be manipulated and converted into information that one may use to determine opinions and to make decisions. These raw data sets should be received from reliable sources and each must apply to the year of 2007 to avoid any inconsistencies.

The next step within this methodology is to save and keep the original copies of every data set utilized within this experiment. At this time, we will compare the total educational state funding for Missouri to the total amount of funding listed for other states in the country. More

specifically, we will review how Missouri ranks per student funding among the country. The total financial support and federal, state, and local funding divisions per student in Missouri will be compared and matched within each division to the previously listed pie chart in Figure 1. The results of Missouri's per student funding is illustrated in Figure 3. The dataset containing information related to every state and the amount of educational funding each receives shows that Missouri obtained $730,811,000 in 2007. The state of Missouri received about $9,691.00 per student in 2007 which ranks it 38th in the nation according to this dataset.

Figure 3: Division of Education Funding in Missouri



Next, the public school districts in the state of Missouri will be listed in a new Excel file along with their total funding and federal, state, and local funding. These attributes must be sorted among all school districts in the country and pasted correctly to the new Excel file. At this time, the attributes must be sorted alphabetically according to the name of the school district. Now, the total population of students will be listed in a column according to the district. It is vital that the districts and populations correspond and this must be doubled checked. If there is missing data (if there are no population data for certain school districts) this must be highlighted red to alert us that it contains missing values. These steps within the process are significant and

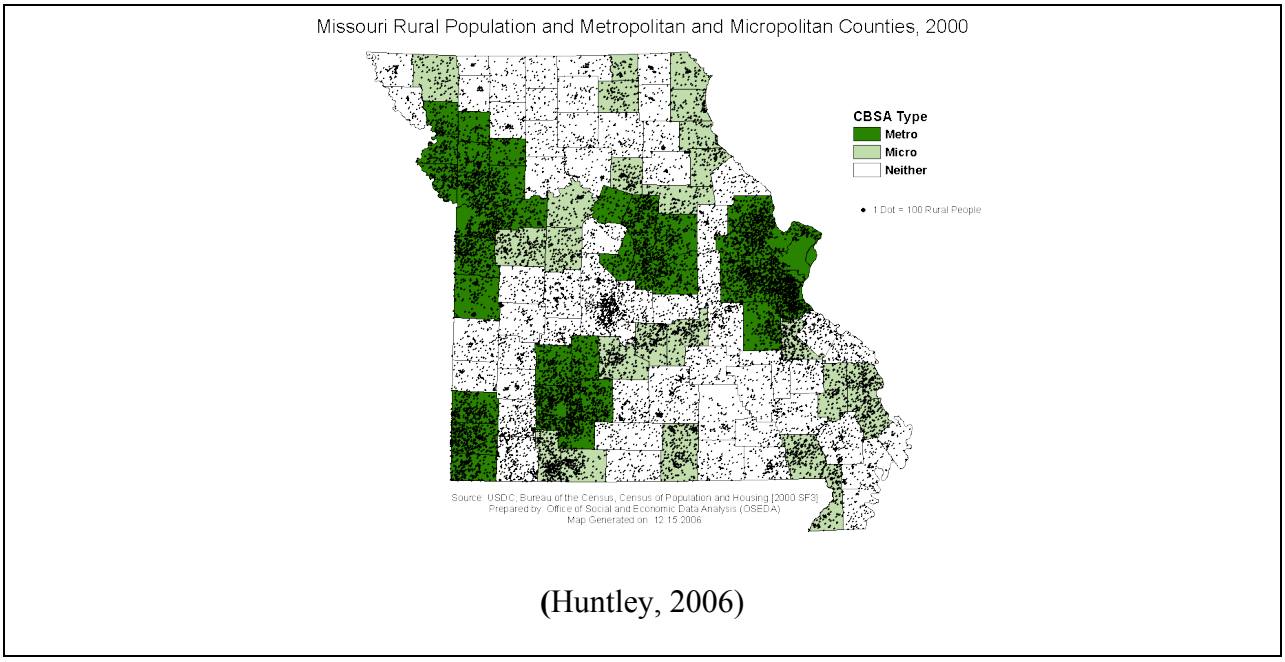
should be diligently completed to assure that the funding data values align correctly with the corresponding school district's population. The datasets must align perfectly and each of the 523 attributes must be evaluated. At this moment, we will dedicate the following column to the division of the total funding for each public school district by the population of students who attend. This number will determine the financial support dedicated to each student in the district.

The following step involved in this methodology is to list the graduation rates in the Excel file according to the school district. It is essential that each school district's information is aligned across the row in order for all the information to appropriately apply. After the school districts are aligned in the Excel file, we must sort through the 523 attributes to see if there is missing data. Since there is missing data (school districts listed without funding, population, or graduation rates listed), these school districts need to be deleted. This is not the most complete method to perform this experiment but it is adequate due to time constraints. Once the rows are deleted containing missing data, we must also delete the red text from the previous step. Now, there will be a school district listed and three sets of important information: funding per student, district population, and graduation rates. We will have 431 school districts to work with. At this time, we will find the correlation of the funding per student row and the graduation rates row.

The following step includes locating twenty "urban" school districts and twenty "rural" school districts and placing these different types of districts in separate Excel sheets. Since there are various measures to define a school district as urban or rural, we will base our results on the information available by the Missouri Data Census Center. The map of Missouri located in Figure 4 shows the counties in the state and rural areas are depicted by dots. As you can see, most of the rural dots are found in dark green areas that represent a metropolitan area. We will locate urban school districts by researching districts within dark green counties with less black

dots. However, we will be conducting the rural school districts from the white areas that are neither metropolitan nor micropolitan counties on this map. With the use of the map in Figure 4 and a map labeling Missouri's counties, we will determine appropriate rural and urban school districts (Huntley, 2006). Once we create two sheets full of twenty rural school districts' information and twenty urban school districts' information, we will formulate the correlation between the funding per student and graduation rates. At this point, we will have a correlation of the entire state's school districts, an urban school district correlation, and a rural school district correlation. We will compare these results and review how they relate to each other.

Figure 4: Missouri Rural, Metropolitan, and Micropolitan Counties



(Huntley, 2006)

After calculating the correlation of the funding per student and graduation rates among 431 school districts in the state of Missouri, there is a 0.243973 correlation. A correlation coefficient of 0.24 is a weak positive correlation among the two metrics that the experiment measured. There are multiple different potential causes for this result. First of all, this correlation

may show that funding in school districts is not as important as our society thinks. Moreover, this could mean that there is another factor of greater influence than funding on school districts. In fact, there are school districts among the data that are provided with lower per student funds but continue to have high graduation rates. This could also mean that the quality metric we used in this experiment, graduation rates, isn't the best category to use to determine how a school is performing. This weak positive correlation does not lead us into any extreme conclusions at this point. Thus, it makes sense that the correlation isn't strong because there are 431 school districts which are provided various amounts of financial support and remain at different quality levels. Further research on this topic would fulfill this curiosity and likely result in more strong conclusions.

The correlation coefficient found among the twenty distinguished rural districts is 0.540777. Since this correlation is positive and much stronger than the result found among the entire state, this supports that the more investment in a rural school there is, the more positive influence this funding has on the quality of education and the success of the students who attend. Now, this correlation is relatively strong, and it is a solid result that allow for conclusions to be made. The results of the rural school districts are located in Figure 5.

After formulating the correlation among the twenty school districts identified as urban, we found 0.295153 as a result. This is an interesting conclusion because even though it is not a particularly strong positive correlation, the number is stronger than the original correlation coefficient calculated among all 431 school districts in Missouri. This could be due to the coincidence of the urban schools utilized within this portion of the experiment or the fact that there are only twenty schools in consideration at this time. However, this result proves that there is consistency because the urban district's correlation is only 0.05118 greater than the correlation

coefficient among the state-wide portion of the study. Figure 6 illustrates the results of the urban school districts that were utilized.

Figure 5: Results of Twenty Rural School Districts

School District	Total Fund	Federal	State	Local	Dist Pop	Per Student Funding	Grad Rate
FREDERICKTOWN SCH DIST R1	13249	1466	7193	4590	1777	7.455824423	87.7
IRON CO SCH DIST C 4	3688	438	999	2251	535	6.893457944	90.5
STE GENEVIEVE CO SD R2	16999	1495	4993	10511	2561	6.637641546	84.6
WORTH SCHOOL DIST R 3	3494	336	1786	1372	326	10.71779141	97.1
THAYER SCH DIST R2	5315	684	2981	1650	584	9.101027397	95.2
PIKE SCH DIST R3	6402	656	2282	3464	504	12.70238095	91.1
MORGAN CO SCH DIST R-2	11581	1388	3565	6628	2010	5.761691542	83.1
MORGAN COUNTY SCHOOL DIST R 1	6191	900	2740	2551	810	7.643209877	84.1
MACKS CREEK SCH DIST 5	2924	209	1477	1238	433	6.752886836	84.6
CLIMAX SPRINGS SCHOOL DIST R4	2940	206	854	1880	273	10.76923077	100.0
STEELVILLE SCH DIST R 3	8285	817	4256	3212	1001	8.276723277	75.3
CARUTHERSVILLE SCHOOL DIST 18	11132	1578	6877	2677	1376	8.090116279	69.7
COOTER SCH DIST R 4	2065	155	1297	613	144	14.34027778	100.0
DELTA SCHOOL DISTRICT C-7	2066	239	1233	594	234	8.829059829	100.0
LAMAR SCH DIST R-1	10725	566	4811	5348	1393	7.699210337	85.6
LIBERAL SCH DIST R-2	3833	253	2161	1419	580	6.60862069	88.4
BUCKLIN SCH DIST R 2	1904	173	1041	690	170	11.2	100.0
BROOKFIELD SCH DIST R3	10037	826	5641	3570	1066	9.415572233	88.5
IBERIA SCH DIST R5	5630	531	3377	1722	714	7.885154062	91.0
ST ELIZABETH SCH DIST R-4	2055	115	1258	682	253	8.122529644	100.0

(**All funding is scaled in thousands.)

Figure 6: Results of Twenty Urban School Districts

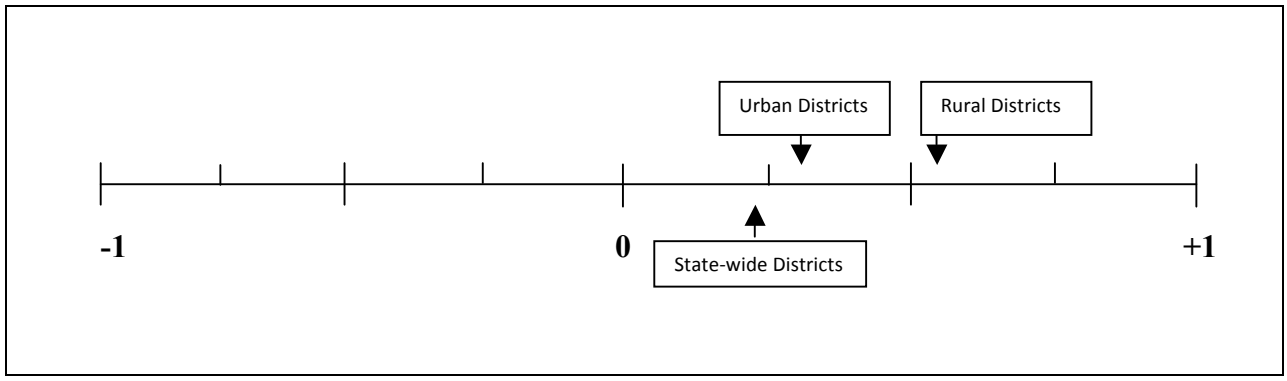
School District	Total Fund	Federal	State	Local	Dist Pop	Per Student Funding	Grad Rate
ST LOUIS CITY BD OF EDUCATION	443778	59992	179131	204655	62138	7.141813383	54.8
KANSAS CITY SCH DIST 33	346064	46744	137182	162138	41488	8.341303509	75.7
NORMANDY SCH DIST	56317	5433	32719	18165	7248	7.770005519	60.1
RIVERVIEW GARDENS SCHOOL DIST	71144	6737	44422	19985	9343	7.614684791	82.7
RITENOUR SCH DIST	57913	3703	24030	30180	7457	7.76625989	85.9
UNIVERSITY CITY SCH DIST	42741	3277	14027	25437	5416	7.89161743	76.7
JEFFERSON CITY SCH DIST	75909	7103	22424	46382	11415	6.649934297	83.5
FRANCIS HOWELL SCHOOL DIST R 3	181070	6780	64097	110193	26390	6.861311103	90.1
ST CHARLES CO. SCHOOL DIST. R-6	67445	3621	21807	42017	9735	6.928094504	84.3
LEES SUMMIT SCH DIST R-7	176228	5201	66240	104787	15009	11.74148844	91.7
FT ZUMWALT SCH DIST R II	170204	6256	60836	103112	21396	7.95494485	89.6
BAYLESS SCH DIST	12068	691	4190	7187	1879	6.422565194	92.4
AFFTON SCH DIST 101	26802	537	5938	20327	3345	8.012556054	90.3
HAZELWOOD SCH DIST R-1	191771	7523	72975	111273	22498	8.523913237	84.3
FAYETTE SCH DIST R 3	6054	475	2940	2639	711	8.514767932	86.2
NEW FRANKLIN SCH DIST R-1	3850	687	1729	1434	449	8.574610245	76.5
POLO SCH DIST R 7	3765	223	2008	1534	349	10.78796562	97.4
BRAYMER SCH DIST C 4	2946	276	1465	1205	356	8.275280899	91.2
LAWSON SCH DIST R-14	11012	548	5799	4665	1198	9.191986644	92.9
INDEPENDENCE SCH DIST 30	126987	14694	59938	52355	11918	10.65505957	82.7

(**All funding is scaled in thousands.)

In conclusion, when we compare all three of these results to one another, it helps us put the entire experiment into perspective. Figure 7 shows a scale which features the correlation coefficients throughout the three components of this study. As you can see, all of the correlations are positive- even though some are weaker than others. This means that the more funding a school district is given, the better quality of school it will be. The correlation coefficient among the 431 school districts among the state is the weakest of the three. This number and the coefficient of the urban school districts are quite close to zero, concluding that the funding and quality may have no relationship in these studies. Yet, the coefficient of the rural school districts is over 0.5 which is a stronger foundation to draw conclusions off of that were previously

mentioned. However, keep in mind that correlation is not always causation and that some other factor may be involved that produces these conclusions.

Figure 7: Scale of Correlation Coefficients of State-Wide, Rural, and Urban Districts



Assumptions

The initial assumption in this experiment is that the datasets are reliable and accurate data from 2007. Another assumption is that using graduation rates as a quality school district metric is appropriate and will provide adequate results. The final assumption is that the method for determining urban and rural school districts using the map in Figure 4 is an appropriate method. Even if there are better means of identifying urban and rural school districts, the method used in this experiment is consistent and is based on information from the United States Census Bureau.

Results

This experiment has provided us with conclusions after working with raw data. As a result, we found that there is a weak positive correlation of .243973 among the 431 school districts in the state was studied. This shows that the relationship among the funding per student and quality of school is not significant. It could also prove that there may be another factor that holds more importance regarding the quality of schools. This set of our results may also show

that the quality of school metric we used (graduation rates) may not be the most beneficial measure to assess the quality of school districts. There is a positive correlation of .540777 among the rural school districts. This stronger, positive correlation coefficient supports the conclusions that when a rural school district receives more financial support, it will perform higher. Lastly, this experiment found consistency after we tested the urban school districts. The correlation coefficient regarding urban school districts is 0.295153. This number is near the coefficient found among all school districts in the state which illustrates consistency in the experiment. Overall, these two correlation coefficients are not particularly strong. However, this leads us to draw conclusions that point us in a direction to improve this experiment and perform further research in order to use metrics that will provide strong results in future work.

In all, the results from the state-wide and urban school district studies do provide strong enough correlation between funding per student and graduation rates to draw conclusions. At this point, it appears as if there is a minimal relationship between these two factors. However, there is a positive correlation among the rural school districts. Thus, when a rural school district receives more funding, it is likely that the quality of the district will improve.

Issues

I encountered various issues while carrying out this experiment. The first issue I underwent was that there was missing data in my school population attributes. There were only about three schools that were listed so I just went on and thought I would research these numbers soon. However, there were almost 100 missing school districts' graduation rates in the third dataset. My experiment began by testing 523 school districts and concluded by considering 431 school districts. In the future, I would need to research to enter in the missing data values instead

of simply deleting the rows with missing data like I had to due to a time constraint. Another issue I had was that I copied the incorrect population numbers and found the per student funding with the wrong population. This altered my results until I realized that I copied over the population of the district instead of the population of 5-17 year olds in the district. Additionally, I had an issue with Excel where I wasn't able to pull down a formula. Therefore, I had to manually input the formula to find the per student funding for approximately 500 school districts twice. Then, I realized I had missing graduation rates data and had to delete some of the values I previously typed in. In the future, I would wait to formulate the per student funding until the end so I knew which school districts I was deleting before manually typing in the formula. However, this process taught me a great deal. Another issue I had concerned identifying urban and rural school districts. I did not realize how confusing or subjective this distinction is until I began researching them. I expected to find a clear list of urban and rural schools. Yet, this list does not exist. Hopefully, the method I used to determine urban and rural school districts is appropriate and reliable. The final issue I encountered pertained to the scale of numbers in the datasets. Some of the numbers were presented in dollars and others were presented in thousands. It took me a short period of time to realize my numbers should be multiplied by 1,000 and then the results appeared plausible.

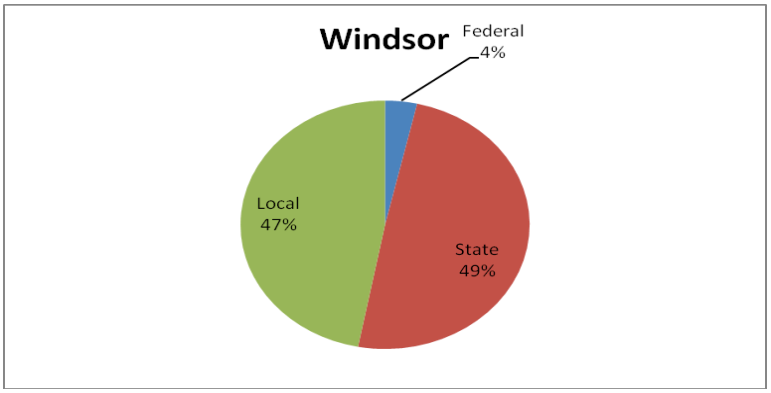
Appendices

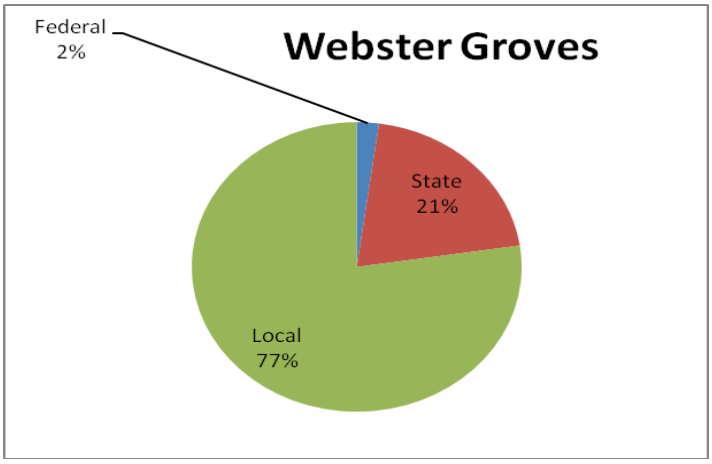
I would like to feature some of the well-known schools and their results.

School District	Total Funds	Federal Funds	State Funds	Local funds	Dist Pop	Per Student Funding	Grad Rate
WINDSOR SCH DIST C 1	25312	903	12512	11897	3449	7.338938823	95.1
WEBSTER GROVES SCH DIST	50157	1083	10268	38806	5312	9.442206325	91.6
ROCKWOOD SCH DIST R-6	217784	4297	49892	163595	22261	9.783208302	94.5
RIVERVIEW GARDENS SCHOOL DIST	71144	6737	44422	19985	9343	7.614684791	82.7
ST LOUIS CITY BD OF EDUCATION	443778	59992	179131	204655	62138	7.141813383	54.8
PARKWAY SCH DIST C-2	200038	5112	38448	156478	23230	8.611192424	93.3
NORMANDY SCH DIST	56317	5433	32719	18165	7248	7.770005519	60.1
LINDBERGH SCH DIST R-8	56867	1024	10816	45027	6635	8.570761115	93.9
KIRKWOOD SCH DIST R 7	63466	959	12864	49643	6493	9.774526413	92.6
FOX CONS SCH DIST C 6	104697	5225	53621	45851	12407	8.438542758	90.6

**All funding is scaled in thousands.

I also thought it would be interesting to look into the funding distribution of sources among school districts. I compared my school district, Windsor C-1, to Webster Groves school district. I noticed that Webster Groves receives more local funding while Windsor is supported by more state funding. I think this is interesting because Webster Groves is more populated, has a higher expense of living, and has higher taxes than Imperial, MO where Windsor is located. The pie charts comparing the sources are located below.




Webster Groves
Federal
2%
State
21%
Local
77%

Bibliography

Burtless, G. (1996). *Does Money Matter? The Effect of School Resources on Student Achievement.* Retrieved from Google Books: http://books.google.com/books?hl=en&lr=&id=N3UIwF9P1WUC&oi=fnd&pg=PA1&dq=education+funding+per+pupil&ots=SUEtQzgfqb&sig=_i6H6Cl1ap_Ru27usNdKlcZMjtk#v=onepage&q=education%20funding%20per%20pupil&f=false

Card, D., & Payne, A. (January). *School finance reform, the distribution of school spending, and the distribution of student test scores.* Retrieved from Science Direct: http://www.sciencedirect.com/science/article/pii/S0047272700001778

Chaffee R-II Comprehensive School Improvement Plan. (2007, May). Retrieved from http://chaffee.k12.mo.us/view/19.pdf

District and School Information. (2012, July 6). Retrieved from Missouri Department of Elementary and Secondary Education: dese.mo.gov/

Huntley, L. (2006, December). *Ten Things to Know About Urban Vs. Rural*. Retrieved from Missouri Census Data Center: http://mcdc.missouri.edu/TenThings/urbanrural.shtml

Lanthier, E. (2002, March). *Correlation*. Retrieved from NVCC: http://www.nvcc.edu/home/elanthier/methods/correlation.htm

MISSOURI- 2010 Census Results. (2010). Retrieved from United Census Bureau: http://2010.census.gov/news/img/cb11cn49_mo_totalpop2010map.jpg

Sebold, F., & Dato, W. (1981, January). *School Funding and Student Achievement: An Empircal Analysis.* Retrieved from Public Finance Review: http://pfr.sagepub.com/content/9/1/91.short

U.S. Census Bureau. (2010, December 8). *Public School Finance Data*. Retrieved from United States Census Bureau: http://www.census.gov

U.S. Census Bureau. (2010, December 8). *School District Estimates for 2007*. Retrieved from United States Census Bureau: http://www.census.gov

U.S. Census Bureau. (2010, December 8). *School Districts*. Retrieved from United States Census Bureau: http://www.census.gov/did/www/schooldistricts/