US Education Funding

Joel Robles, Eric Dally, Kenneth Priester, Alejandra Alvarez

Description of the project:

Analyzing the relationship between the revenue and expenditures in education by state, and their average NAEP Math/ Reading exam scores.

We will compare information between high and low performing states to find how school performance and/or students grades are affected by school funding, and teacher-student ratio.

We understand that each state represented within the dataset receives a different level of revenue with respect to their federal, state and local funding streams. Further, we also understand that each state has a different level of spending as it relates to instruction, support services, capital outlay and other items. Understanding then that because there are different levels of funding/expenditures, we are looking to explore if there exists trends that can be derived from the data with respect to the outcomes of the students within the state. By exploring this, we will be able to create a template to inform future practices as it relates to the funding and spending of states on their education as they develop their budgets.

Description of the source data:

Utilizing Kaggle, we will work with the following data set.

To ensure that our exploration of data is specific and measurable, we will be looking exclusively at the Nation Assessment of Education Progress (NAEP) Math and Reading scores for 4th and 8th grades.

As NAEP is congressionally mandated project that is conducted nationwide, it will likely provide the best data comparison across the states.

What is NEAP?

The National Assessment of Educational Progress (NEAP) is the only assessment that measures what U.S. students know and can do in various subjects across the nation, states and in some urban districts.

NEAP is given to a representative sample of students across the country rather than the entire student population.

*From nces.ed.gov

Questions to answer:

1. Are there correlations between funding and/or expenditures and the average state NAEP Math/Reading test scores for 4th and 8th grade?
2. How does the funding and expenditures of higher performing states (as defined by higher average NAEP Math/Reading scores) compare to lower performing states?

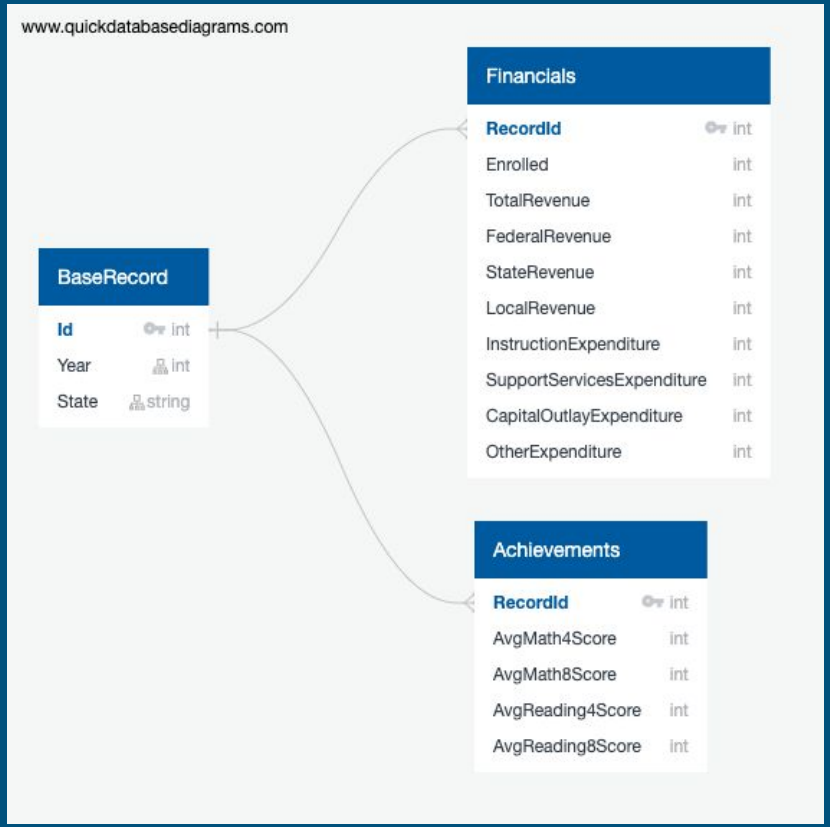
By exploring the questions, we hope to provide advice/guidance to states as they advocate for funding or spend dollars on school budgets.

How are schools funded?

According to Education Week, public school funding comes from a variety of sources at the local, state and federal level. Approximately 48% of a school's budget comes from state resources, including income taxes, sales taxes, and fees. Another 44% is contributed locally, primarily through the property taxes of hometowners in the area. The last 8% of the public education budget comes from federal sources, with an emphasis on grants for specific programs and services for students that need them.

*From publicschoolreview.com

After investigating the CSV file (and the respective columns contained within), the team decided to create the following ERD:



Original dataset:

	Id	State	Year	Enrolled	TotalRevenue	FederalRevenue	StateRevenue	LocalRevenue	InstructionExpenditure
0	2003_ALABAMA	ALABAMA	2003	727900.0	5196054.0	567704.0	2966981.0	1661369.0	2817111.0
1	2003_ALASKA	ALASKA	2003	133303.0	1425948.0	259423.0	813371.0	353154.0	763525.0
2	2003_ARIZONA	ARIZONA	2003	875111.0	6529894.0	740579.0	2912629.0	2876686.0	2810907.0
3	2003_ARKANSAS	ARKANSAS	2003	450158.0	3241275.0	379947.0	2394336.0	466992.0	1768713.0
4	2003_CALIFORNIA	CALIFORNIA	2003	6226552.0	59815855.0	5795655.0	33617766.0	20402434.0	29561563.0
...	...	...	...	...	...	...	...	...	...
352	2015_VIRGINIA	VIRGINIA	2015	1279867.0	15857524.0	1012205.0	6240349.0	8604970.0	8755896.0
353	2015_WASHINGTON	WASHINGTON	2015	1072359.0	13709442.0	1036422.0	8293812.0	4379208.0	6508964.0
354	2015_WEST_VIRGINIA	WEST_VIRGINIA	2015	279565.0	3478401.0	362959.0	1979466.0	1135976.0	1819903.0
355	2015_WISCONSIN	WISCONSIN	2015	861813.0	11637376.0	814385.0	5869265.0	4953726.0	5723474.0
356	2015_WYOMING	WYOMING	2015	93867.0	1962874.0	120290.0	1116917.0	725667.0	895910.0

SupportServicesExpenditure	CapitalOutlayExpenditure	OtherExpenditure	AvgMath4Score	AvgMath8Score	AvgReading4Score	AvgReading8Score
1521462.0	432279.0	424662.0	223.0	262.0	207.0	253.0
514052.0	247005.0	55715.0	233.0	279.0	212.0	256.0
1964229.0	864642.0	303924.0	229.0	271.0	209.0	255.0
972598.0	247245.0	170767.0	229.0	266.0	214.0	258.0
17030335.0	7341541.0	2995767.0	227.0	267.0	206.0	251.0
...	...	...	...	...	...	...
5075509.0	1086722.0	627473.0	247.0	288.0	229.0	267.0
4510672.0	1601069.0	546926.0	245.0	287.0	226.0	267.0
1161944.0	232738.0	233836.0	235.0	271.0	216.0	260.0
3691809.0	894823.0	513402.0	243.0	289.0	223.0	270.0
565489.0	409425.0	52520.0	247.0	287.0	228.0	269.0

```
# create database engine
db_url = f"postgresql://{config.DB_USERNAME}:{config.DB_PASSWORD}@{config.DB_HOST}/{config.DB_NAME}"
engine = create_engine(db_url)

# load all the raw data into a datafame
query = text('''SELECT * FROM "RawRecords";''')
df = pd.read_sql_query(query,con=engine, index_col='Id')

# create a df for the stats with the same index as the raw data
stats_df = pd.DataFrame(index=df.index)
# rename the index to match the schema
stats_df.index.rename('RecordId', inplace=True)

# calculate the stats
stats_df['FederalFundingPercent'] = df['FederalRevenue']/df['TotalRevenue']
stats_df['StateFundingPercent'] = df['StateRevenue']/df['TotalRevenue']
stats_df['LocalFundingPercent'] = df['LocalRevenue']/df['TotalRevenue']
stats_df['RevenuePerStudent'] = df['TotalRevenue']/df['Enrolled']
stats_df['InstructionalExpensePercent'] = df['InstructionExpenditure']/df['TotalRevenue']
```

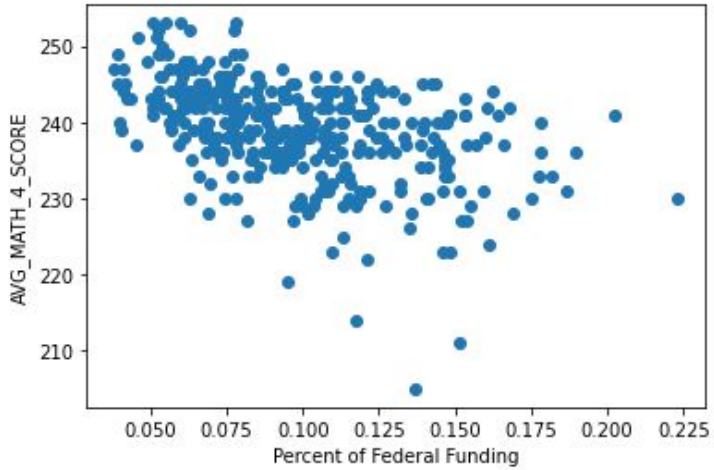
How does different percentages of funding affect 4th Grade Math Scores ?

Schools have 3 sources of funding:

- Federal
- State
- Local

Federal Funding

```
# How does the percentage of federal funding correlate to math grades?
plt.scatter(df2['FEDERAL_REVENUE']/df2['TOTAL_REVENUE'], df2['AVG_MATH_4_SCORE'])
plt.xlabel('Percent of Federal Funding')
plt.ylabel('AVG_MATH_4_SCORE')
plt.show()
```



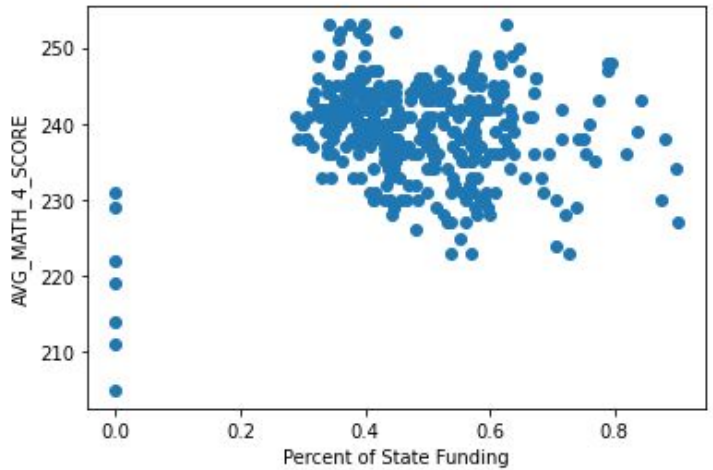
How does different percentages of funding affect 4th Grade Math Scores ?

Schools have 3 sources of funding:

- Federal
- State
- Local

State Funding

```
# How does the percentage of state funding correlate to math grades?
plt.scatter(df2['STATE_REVENUE']/df2['TOTAL_REVENUE'], df2['AVG_MATH_4_SCORE'])
plt.xlabel('Percent of State Funding')
plt.ylabel('AVG_MATH_4_SCORE')
plt.show()
```



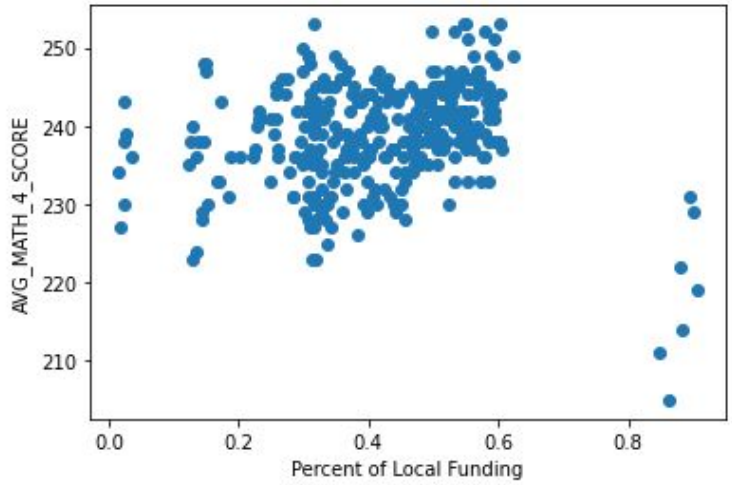
How does different percentages of funding affect 4th Grade Math Scores ?

Schools have 3 sources of funding:

- Federal
- State
- Local

Local Funding

```
# How does the percentage of local funding correlate to math grades?
plt.scatter(df2['LOCAL_REVENUE']/df2['TOTAL_REVENUE'], df2['AVG_MATH_4_SCORE'])
plt.xlabel('Percent of Local Funding')
plt.ylabel('AVG_MATH_4_SCORE')
plt.show()
```



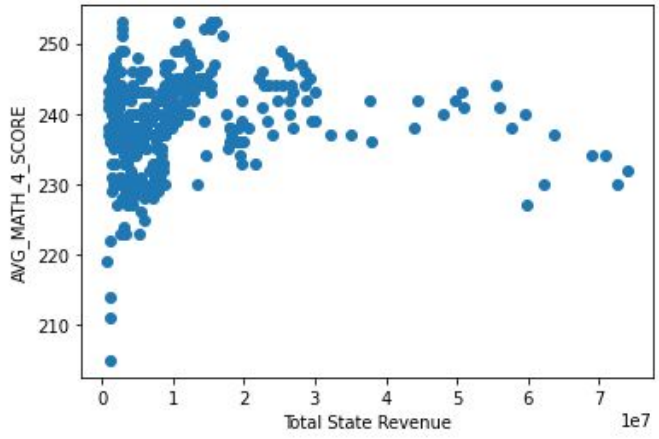
How does different percentages of funding affect 4th Grade Math Scores ?

Schools have 3 sources of funding:

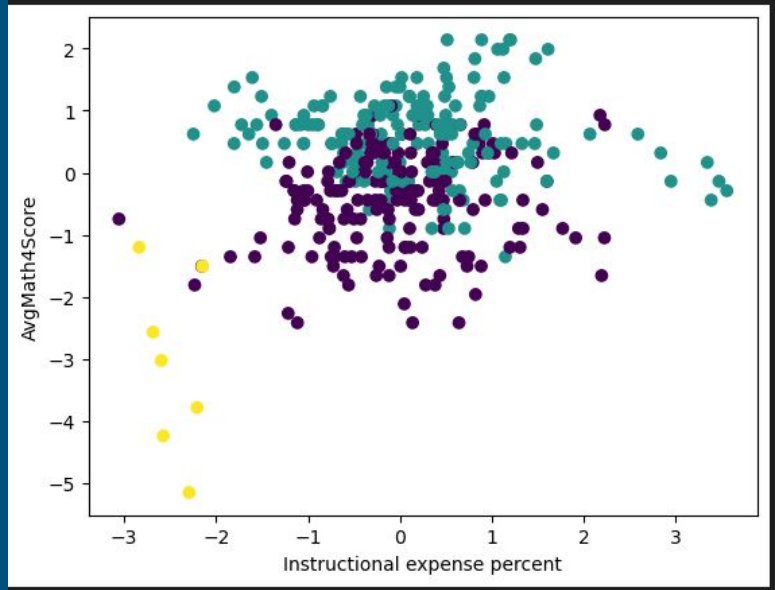
- Federal
- State
- Local

Total Funding

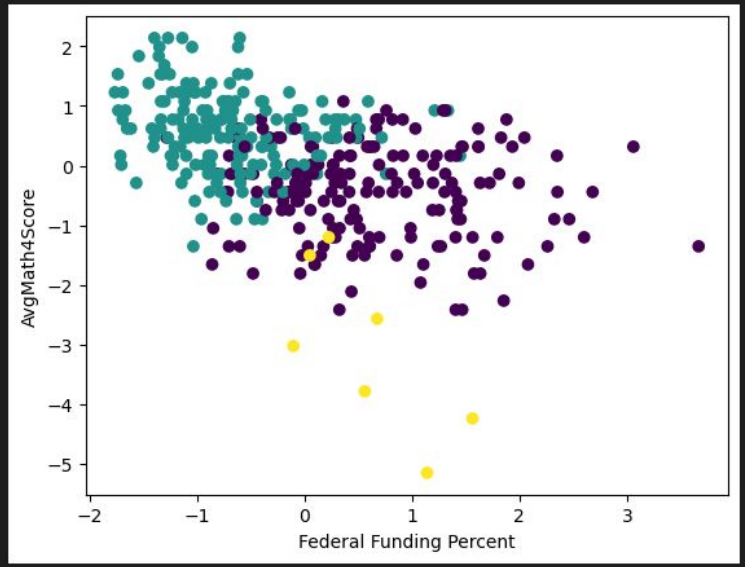
```
#Looking for relationship between total state revenue and average 4th Grade Math Scores
plt.scatter(df2['TOTAL_REVENUE'], df2['AVG_MATH_4_SCORE'])
plt.xlabel('Total State Revenue')
plt.ylabel('AVG_MATH_4_SCORE')
plt.show()
```



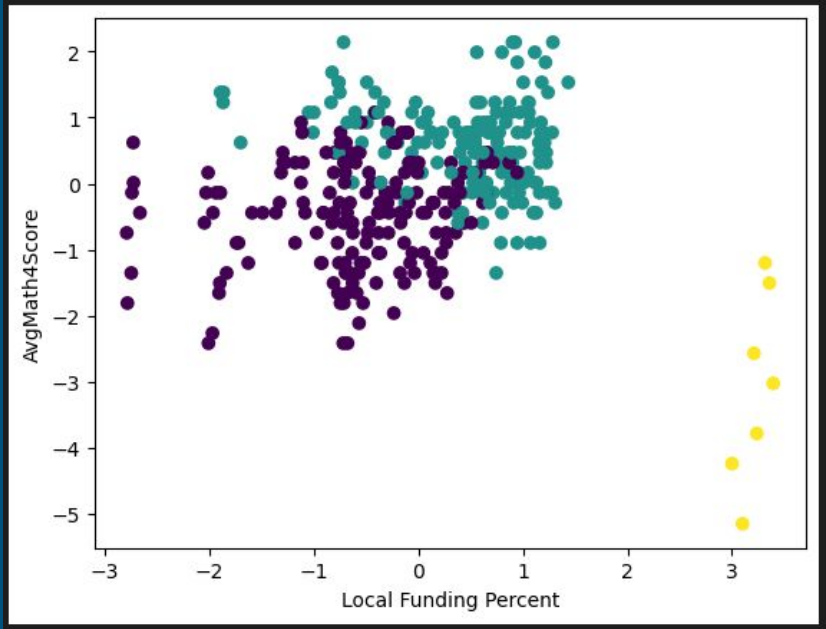
```python
#Instructional Funding
plt.scatter(funding_df["Instructional expense percent"], funding_df["AvgMath4Score"], c=funding_df['cluster'])
plt.xlabel('Instructional expense percent')
plt.ylabel('AvgMath4Score')
plt.show()
```



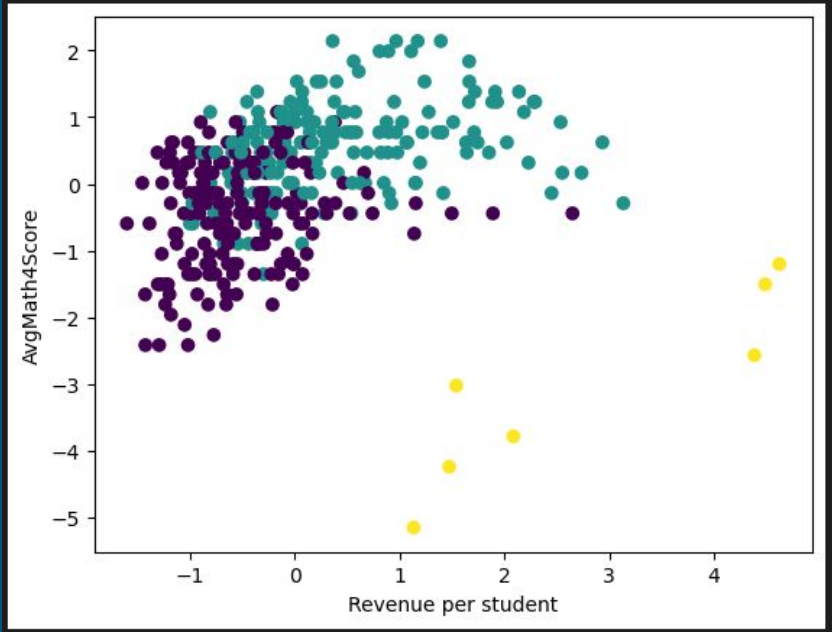
```
#Federal Funding
plt.scatter(funding_df["Federal Funding Percent"], funding_df["AvgMath4Score"], c=funding_df['cluster'])
plt.xlabel('Federal Funding Percent')
plt.ylabel('AvgMath4Score')
plt.show()
```



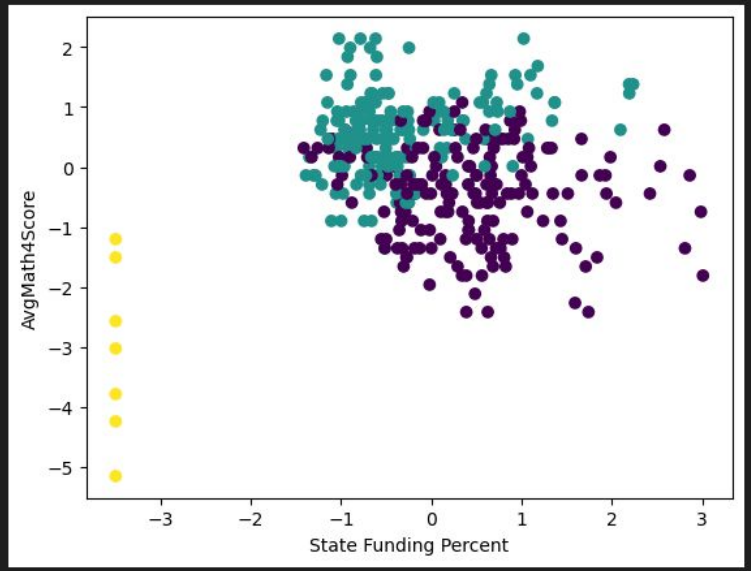
```
#Local Funding
plt.scatter(funding_df["Local Funding Percent"], funding_df["AvgMath4Score"], c=funding_df['cluster'])
plt.xlabel('Local Funding Percent')
plt.ylabel('AvgMath4Score')
plt.show()
```



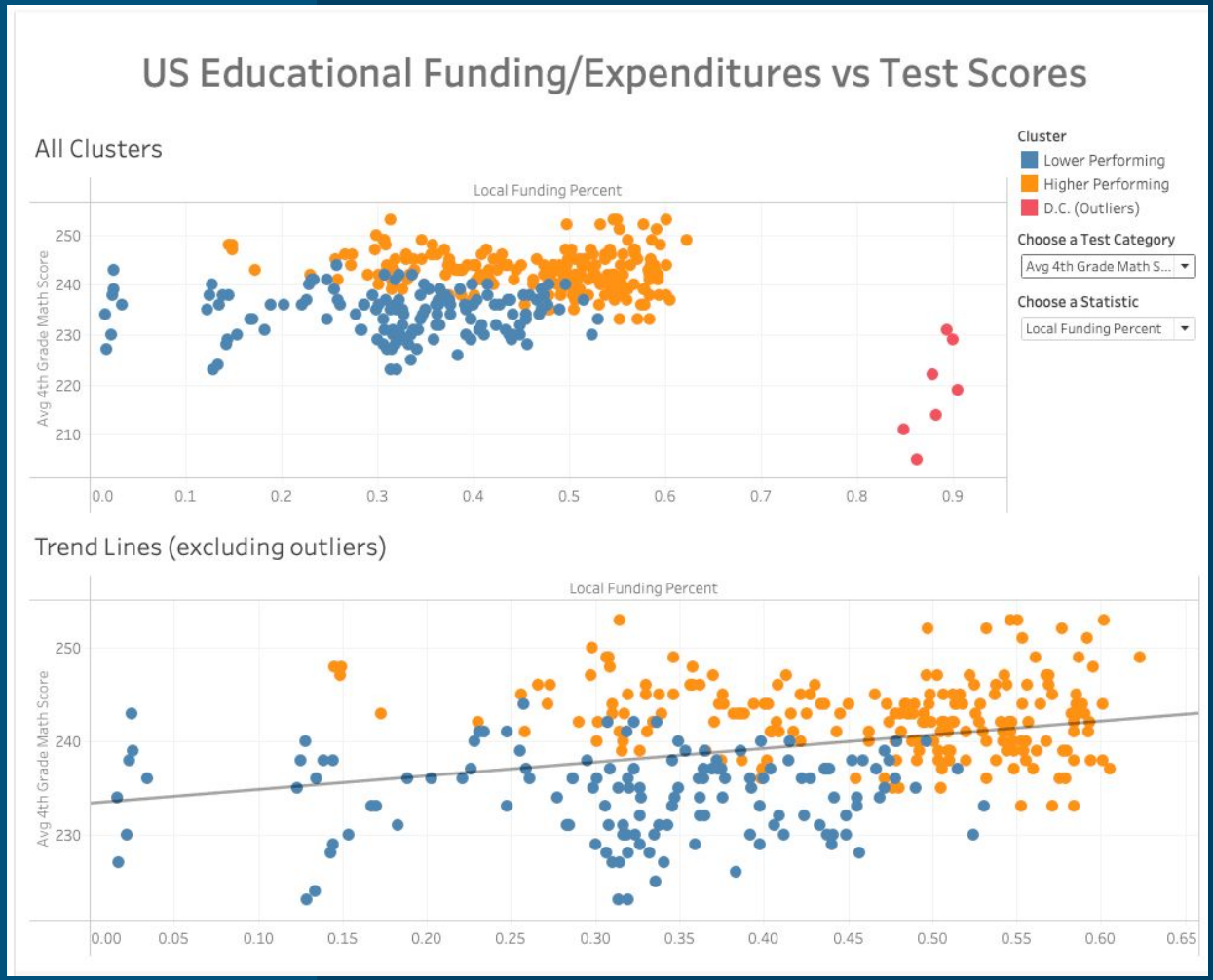
```
# How does the percentage of federal funding correlate to math grades?
plt.scatter(funding_df["Revenue per student"], funding_df["AvgMath4Score"], c=funding_df['cluster'])
plt.xlabel('Revenue per student')
plt.ylabel('AvgMath4Score')
plt.show()
```



```
#State Funding
plt.scatter(funding_df["State Funding Percent"], funding_df["AvgMath4Score"], c=funding_df['cluster'])
plt.xlabel('State Funding Percent')
plt.ylabel('AvgMath4Score')
plt.show()
```



US Educational Funding/Expenditures vs Test Scores



Acknowledgements

- Kaggle.com
- publicschoolreview.com
- nces.ed.gov

*Data is sourced from the U.S. Census Bureau and the National Center for Education Statistics (NCES)