

Process Book – New York City Schools

Overview and Motivation

Consider this inquiry: where is the best place to send a child to school in New York City? How do you define best? How do you define place? How do you define school?

New York City has the largest system of schools in the United States. New York City's 5 boroughs are divided into 32 geographical school districts, which have 1,707 schools that serve over 1 million students. Students in grades K-8 are assigned to a zoned school, and students in grades 9-12 have the option of applying to over 400 high schools. Determining the best place to send a child to school in New York City can be a daunting task.

I selected this topic because I have a passion for using technology to enhance education, and I thought a data visualization on this topic could enhance the experience of comparing boroughs, school districts, and schools on a variety of metrics that the user chooses. The goal is to enable the user to input his/her definition of best, geospatially visualize all of the boroughs, school districts, or schools that meet this definition, and select specific points of interest for detailed comparison so the user can decide which are the best fit. This visualization also supports scenarios of determining which boroughs, school districts, or schools are the worst fit and determining which metrics influence other metrics, for instance, whether student to teacher ratio influences graduation rate.

Overall, this data visualization should provide a user the opportunity to get a holistic picture of the New York City school system. All of the data was sourced from New York City Open Data for the school year 2010-2011. The granular data available through New York City Open Data enabled creating a complete picture of the New York City school system.

In addition to the social motivations for pursuing this topic, I would also like to highlight three personal motivations for completing this project: (a) I aim to use this project to build my data visualization portfolio, with the hope to work in this field, (b) I aim to create this project well enough to be featured on the class hall of fame, and (c) I aim to create this project well enough to get this project highlighted on the New York City Open Data page as a featured visualization.

Related Work

When considering potential topics for the project, I knew that I wanted to represent data geospatially so that I could gain experience in how to create a map in D3. I followed tutorials from Mark Dimarco (<http://datamaps.markmarkoh.com/using-custom-map-data-w-datamaps/>) and Mike Bostock (<http://bost.ocks.org/mike/map/#converting-data>) to download all of the necessary software, run all of the necessary conversions, and create a map for a custom shape file in D3.

Inspiration and coloring for the choropleth map came from the Color Brewer examples in the Design for Information book (<http://isabelmeirelles.com/book-design-for-information/>) and the Color Brewer scale picker (<http://colorbrewer2.org/>).

Inspiration for the parallel coordinates chart came from Mike Bostock (<http://bl.ocks.org/mbostock/1341021>) and the chart was derived from a Jason Davies example (<http://bl.ocks.org/jasondavies/1341281>).

Questions

From the Proposal

- Which school districts and schools within New York City have the best/worst:
 - Overall Rating
 - Average Class Size
 - Safety Rating
 - Quality Rating
 - Environment Rating
 - Graduation Rate
 - Drop Out Rate
 - Attendance Rate
 - English Test Scores
 - Math Test Scores
 - Teacher Scores
- How do some of the factors listed in the previous bullet point influence some of the other factors in the previous bullet point? For instance, do schools with the highest attendance have the highest graduation outcomes and does the overall rating influence the graduation outcome.
- What is the teacher, student, and parent perception of the school district?

Since the Proposal

Imagine the following use cases:

- As a parent, I am moving to New York City and want to know the best place to live to send my child to the best schools in New York City that fit my criteria.
- As a parent, I want to know how my child's school compares to other schools based on my criteria.
- As a New York City citizen, I want to know how boroughs compare to each other and school districts compare to each other based on my criteria.

On a Humorous Note

After hours of data scrubbing, the following idea for a metric comes to mind: ridiculousness of school name factor. There are many school names that describe the type of students they aim to produce to attract students. My hypothesis is that the more the school name sounds like a marketing tactic, the worse the school performance would be.

Data

Creating a holistic picture of the New York City school system required a lot of data. I found it very important to keep all of the datasets to the same school year (2010-2011) so as not to confuse or unintentionally mislead the user. The only dataset that does not meet this criteria is the School Zones shape file, which is for the school year 2011-2012, as it was the oldest shape file available.

Source Data

Data was sourced from the following Department of City Planning, City of New York datasets:

- Geospatial Data
 - Borough Boundaries (Clipped to Shoreline) Shape File
 - http://www.nyc.gov/html/dcp/download/bytes/nybb_15a.zip
 - School Districts (Clipped to Shoreline) Shape File
 - http://www.nyc.gov/html/dcp/download/bytes/nysd_15a.zip

Data was sourced from the following New York City Open Data datasets:

- Geospatial Data
 - School Point Locations Shape File
 - <https://data.cityofnewyork.us/Education/School-Point-Locations/jfju-ynrr>
 - School Zones 2011-2012 Shape File
 - <https://data.cityofnewyork.us/Education/School-Zones-2011-2012/dqkt-8x6u>
- Borough Level Data
 - 2010-2011 Class Size – Borough Summary
 - <https://data.cityofnewyork.us/Education/2010-2011-Class-Size-Borough-Summary/7yds-6i8e>
 - English Language Arts (ELA) Test Results by Grade 2006-2011 – Boro – All Students
 - <https://data.cityofnewyork.us/Education/English-Language-Arts-ELA-Test-Results-by-Grade-20/4kse-vfnd>
 - NYS Math Test Results By Grade 2006-2011 – Boro – All Students
 - <https://data.cityofnewyork.us/Education/NYS-Math-Test-Results-By-Grade-2006-2011-Boro-All-/a2nf-cvfm>
 - Graduation Outcomes – Borough – Classes of 2005-2011 – Total Cohort
 - <https://data.cityofnewyork.us/Education/Graduation-Outcomes-Borough-Classes-of-2005-2011-E/qs5h-jhhg>

- District Level Data
 - School Attendance and Enrollment by District 2010-11
 - <https://data.cityofnewyork.us/Education/School-Attendance-And-Enrollment-By-District-2010-/rfpq-hs49>
 - 2010-2011 Class Size – District-level Summary
 - <https://data.cityofnewyork.us/Education/2010-2011-Class-Size-District-level-Summary/82rt-zc4y>
 - English Language Arts (ELA) Test Results by Grade 2006-2011 – District – All Students
 - <https://data.cityofnewyork.us/Education/English-Language-Arts-ELA-Test-Results-by-Grade-20/85ty-ti6v>
 - NYS Math Test Results by Grade 2006-2011 – District – All Students
 - <https://data.cityofnewyork.us/Education/NYS-Math-Test-Results-By-Grade-2006-2011-District-/gyaz-82xj>
- School Level Data
 - Quality Review 2005-2012
 - <https://data.cityofnewyork.us/Education/Quality-Review-2005-2012/piri-jns7>
 - School Progress Reports – All Schools - 2010-11
 - <https://data.cityofnewyork.us/Education/School-Progress-Reports-All-Schools-2010-11/yig9-9zum>
 - NYC School Survey – 2011
 - <https://data.cityofnewyork.us/Education/NYC-School-Survey-2011/mnz3-dyi8>
 - English Language Arts (ELA) Test Results by Grade 2006-2011 – School level – All Students
 - <https://data.cityofnewyork.us/Education/English-Language-Arts-ELA-Test-Results-by-Grade-20/yu9n-iqyk>
 - NYS Math Test Results by Grade 2006-2011 – School Level – All Students
 - <https://data.cityofnewyork.us/Education/NYS-Math-Test-Results-By-Grade-2006-2011-School-Le/jufi-gzgp>
 - Graduation Outcomes – School Level – Classes of 2005-2011 – Total Cohort
 - <https://data.cityofnewyork.us/Education/Graduation-Outcomes-School-Level-Classes-of-2005-2/cma4-z18m>
 - School Safety Report
 - <https://data.cityofnewyork.us/Education/School-Safety-Report/qybk-bjic>
 - 2010-2011 Class Size – School-level detail
 - <https://data.cityofnewyork.us/Education/2010-2011-Class-Size-School-level-detail/urz7-pzb3>
 - NYC Public High School Program Data
 - <https://data.cityofnewyork.us/Education/NYC-Public-High-School-Program-Data/mreg-rk5p>

The School District Breakdowns dataset (<https://data.cityofnewyork.us/Education/School-District-Breakdowns/g3vh-kbnw>) was also attempted to be sourced from New York City Open Data. This dataset was intended to have demographic statistics broken down by school districts. However, there were only 2,103 total respondents across over 1 million students, so the sample was much too small to be usable as representative of the demographics of New York City students. Therefore, this dataset was excluded from the project.

Cleansed Data

Geospatial

All of the geospatial datasets were downloaded as .shp files and opened with QGIS. There was no cleansing required for the Borough and School District shape files. There were two cleansing activities that were required for the School Point shape file: (a) remove 2 school data points that were not New York City schools, (b) fix the borough identifier for all of the Brooklyn schools, as it was incorrectly listed as Borough 2 – Bronx instead of Borough 3 – Brooklyn. The issue with the borough identifiers was only found once filtering was implemented in the data visualization and no Brooklyn schools appeared when filtering on Brooklyn.

The geospatial datasets then needed to be converted to GeoJSON using ogr2ogr. The source shape files used a gridded coordinate system, so they need to be converted back to conventional latitude and longitude with the parameter EPSG:4326. The ogr2ogr conversion statements were as follows:

- ogr2ogr -f GeoJSON -t_srs EPSG:4326 nybb.json nybb.shp
- ogr2ogr -f GeoJSON -t_srs EPSG:4326 nysd.json nysd.shp
- ogr2ogr -f GeoJSON -t_srs EPSG:4326 nysp.json
Public_Schools_Points_2011-2012A.shp

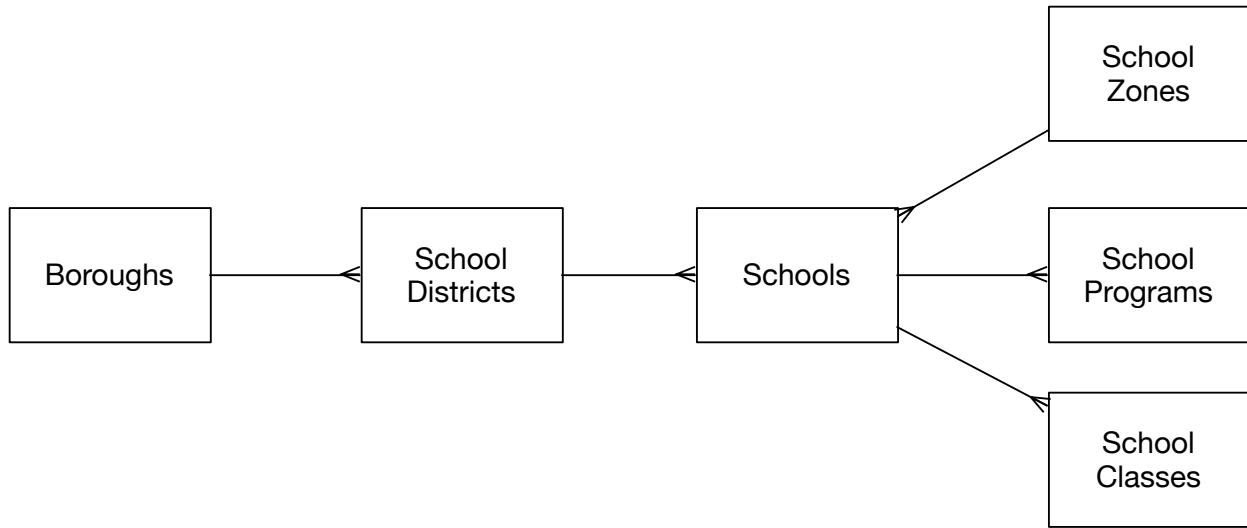
The GeoJSON then needed to be converted to TopoJSON and merged to improve performance. The topojson conversion statement was as follows:

- topojson -o nycs.json -p -- nybb.json nysd.json nysp.json

The final geospatial data file is nycs.json.

Tabular

All of the tabular datasets were downloaded as .csv files and loaded as staging tables into an Access database. Tables were created in the Access Database for the following concepts: Boroughs, School Districts, Schools, School Programs, and School Classes. A conceptual data model is displayed below.



There was significant cleansing required to get the data into a usable common format, which I performed by writing SQL queries against the Access database. This included:

- Standardizing district and school identifiers to a set of common keys, for instance “DISTRICT 01”, “CSD 01 Manhattan”, “M015”, “01M015”, etc.
- Removing text from numerical fields. In many cases, the text values “s” or “-“ were used in place of nulls where no data was available. All of these needed to be removed.
- Removing years of data outside of the school year 2010-2011.
- Separating summary level data from detail level data. For instance, the test results have detail rows by grade and summary rows across all grades all in the same tabular dataset. This same behavior was repeated across several other datasets. The summary and detail needed to be split apart so they could be landed on the appropriate tables.
- Cleanup of source data errors. For instance, update of the mysterious borough “OUEENS” back to “QUEENS”.
- Trimming of extra spaces at the beginning or end of text fields

Two key challenges of interpreting the data was the New York City Open Data page has virtually no metadata documentation and column names listed were often physical names that were not consistent across datasets (lacked business context that logical naming would have). The New York City Department of Education website (<http://schools.nyc.gov/default.htm>) needed to be referenced to interpret the meaning of many of the data points.

After all of the data was cleansed, it was loaded to the appropriate table in the target model. All borough metrics were loaded to the Boroughs table, all school district metrics were loaded to the School Districts table, and all Schools metrics were loaded to the Schools table. School detail level metrics were loaded to the School Programs and School Classes tables.

After all of the target tables were loaded, I created a copy of the Access database, deleted all of the staging tables, and ran a JSONUtils command line utility (<http://www.divconq.com/wp->

content/uploads/2013/05/JSONUtils_dotnet_v1001.zip) to export the Access database to JSON. This is file nycsa.json.

One opportunity I need to explore is denormalizing and nesting the JSON for School Programs and School Classes. Also, the school zone files were a late addition and need to be merged into the nycs.json file.

JSON Data Structure

After merging the nycs.json and nycsa.json files in JavaScript, the resulting JSON has the following structure (only key areas of interest are expanded with detail shown). Items with red text have been determined to be unnecessary and can be deleted from the JSON. Items with green text will be added into the JSON – elementary and middle school zones.

```
{
  arcs
  objects
  {
    nybb
    {
      crs
      geometries
      {
        BoroCode
        BoroName
        Shape_Area
        Shape_Leng
        GraduationOutcomeTotalCohortNumber
        GraduationOutcomeTotalGradNumber
        GraduationOutcomeTotalGradPercentage
        GraduationOutcomeTotalRegentsNumber
        GraduationOutcomeTotalRegentsPercentageCohort
        GraduationOutcomeTotalRegentsPercentageGrad
        GraduationOutcomeAdvancedRegentsNumber
        GraduationOutcomeAdvancedRegentsPercentageCohort
        GraduationOutcomeAdvancedRegentsPercentageGrad
        GraduationOutcomeRegentsWithoutAdvancedNumber
        GraduationOutcomeRegentsWithoutAdvancedPercentageCohort
        GraduationOutcomeRegentsWithoutAdvancedPercentageGrad
        GraduationOutcomeLocalNumber
        GraduationOutcomeLocalPercentageCohort
        GraduationOutcomeLocalPercentageGrad
        GraduationOutcomeStillEnrolledNumber
        GraduationOutcomeStillEnrolledPercentageCohort
        GraduationOutcomeDroppedOutNumber
        GraduationOutcomeDroppedOutPercentageCohort
        MathNumberTested
        MathMean
        MathLevelOneNum
        MathLevelOnePct
      }
    }
  }
}
```

```
    MathLevelTwoNum
    MathLevelTwoPct
    MathLevelThreeNum
    MathLevelThreePct
    MathLevelFourNum
    MathLevelFourPct
    EnglishNumberTested
    EnglishMean
    EnglishLevelOneNum
    EnglishLevelOnePct
    EnglishLevelTwoNum
    EnglishLevelTwoPct
    EnglishLevelThreeNum
    EnglishLevelThreePct
    EnglishLevelFourNum
    EnglishLevelFourPct
}
type
}
nysd
{
    crs
    geometries
    {
        SchoolDist
        Shape_Area
        Shape_Leng
        DistrictAttendance
        DistrictEnrollment
        MathNumberTested
        MathMean
        MathLevelOneNum
        MathLevelOnePct
        MathLevelTwoNum
        MathLevelTwoPct
        MathLevelThreeNum
        MathLevelThreePct
        MathLevelFourNum
        MathLevelFourPct
        EnglishNumberTested
        EnglishMean
        EnglishLevelOneNum
        EnglishLevelOnePct
        EnglishLevelTwoNum
        EnglishLevelTwoPct
        EnglishLevelThreeNum
        EnglishLevelThreePct
        EnglishLevelFourNum
        EnglishLevelFourPct
    }
    type
}
```

```
nysze
{
    Shape_Area
    Shape_Leng
}
nyszm
{
    Shape_Area
    Shape_Leng
}
nysp
{
    crs
    geometries
    {
        coordinates
        properties
        {
            ADDRESS
            ADMIN_DIST
            ATS_CODE
            BORO
            BORONUM
            FAX
            GEO_DISTRI
            GRADES
            LOC_CODE
            MANAGED_BY
            PRINCIPAL
            PRIN_PH
            SCHOOLNAME
            SCH_TYPE
            STATE_CODE
            ZIP
            City

            HasGraduationOutcomeData
            GraduationOutcomeAdvancedRegentsNumber
            GraduationOutcomeAdvancedRegentsPercentageGrad
            GraduationOutcomeDroppedOutNumber
            GraduationOutcomeDroppedOutPercentageCohort
            GraduationOutcomeLocalNumber
            GraduationOutcomeLocalPercentageCohort
            GraduationOutcomeRegentsWithoutAdvancedPercentageGrad
            GraduationOutcomeStillEnrolledNumber
            GraduationOutcomeStillEnrolledPercentageCohort
            GraduationOutcomeTotalCohortNumber
            GraduationOutcomeTotalGradNumber
            GraduationOutcomeTotalGradPercentage
            GraduationOutcomeTotalRegentsNumber
            GraduationOutcomeTotalRegentsPercentageCohort
            GraduationOutcomeTotalRegentsPercentageGrad
        }
    }
}
```

HasClassSizeData
StudentToTeacherRatio

HasQualityScoreData
QualityScore

HasProgressReportData
ProgressReportAdditionalCredit
ProgressReportEnvironmentCategoryGrade
ProgressReportEnvironmentCategoryScore
ProgressReportOverallGrade
ProgressReportOverallScore
ProgressReportPeerIndex
ProgressReportPerformanceCategoryGrade
ProgressReportPerformanceCategoryScore
ProgressReportProgressCategoryGrade
ProgressReportProgressCategoryScore

HasSurveyData
SurveyParentAcademicExpectationsScore
SurveyParentCommunicationScore
SurveyParentEngagementScore
SurveyParentResponseRate
SurveyParentSafetyRespectScore
SurveyStudentAcademicExpectationsScore
SurveyStudentCommunicationScore
SurveyStudentEngagementScore
SurveyStudentResponseRate
SurveyStudentSafetyRespectScore
SurveyTeacherAcademicExpectationsScore
SurveyTeacherCommunicationScore
SurveyTeacherEngagementScore
SurveyTeacherResponseRate
SurveyTeacherSafetyRespectScore
SurveyTotalAcademicExpectationsScore
SurveyTotalCommunicationScore
SurveyTotalEngagementScore
SurveyTotalResponseRate
SurveyTotalSafetyRespectScore

HasSafetyData
MajorRate
NoCrimRate
OtherRate
PropCount
VioCount
BuildingCode
ENGGroupA
RangeA
Register

MathNumberTested
MathMean

```
    MathLevelOneNum
    MathLevelOnePct
    MathLevelTwoNum
    MathLevelTwoPct
    MathLevelThreeNum
    MathLevelThreePct
    MathLevelFourNum
    MathLevelFourPct
    EnglishNumberTested
    EnglishMean
    EnglishLevelOneNum
    EnglishLevelOnePct
    EnglishLevelTwoNum
    EnglishLevelTwoPct
    EnglishLevelThreeNum
    EnglishLevelThreePct
    EnglishLevelFourNum
    EnglishLevelFourPct
}
}
type
}
}
transform
type
}
```

Data Dictionary

Each of the elements of interest are defined in the data dictionary. Items labeled key metric are expected to be available in the parallel coordinate line chart or the map. Metrics that are not key may be available for detailed comparison views in the small multiple bar chart comparison visualizations.

Category	Element	Type	Values	Source	Definition	Scale	Key Metric?
Borough	BoroCode	Integer	1, 2, 3, 4, 5	Borough Shape File	Unique identifier for the borough		
Borough	BoroName	Text	Manhattan, Bronx, Brooklyn, Queens, Staten Island	Borough Shape File	Text name for the borough		
Borough	GraduationOutcome DroppedOutPercent ageCohort	Double		Graduation Outcomes	Percentage of the Cohort that did not graduate because of drop out	0-100	
Borough	GraduationOutcome StillEnrolledPercent ageCohort	Double		Graduation Outcomes	Percentage of the Cohort that did not graduate, yet are still enrolled	0-100	
Borough	GraduationOutcome TotalGradPercentag e	Double		Graduation Outcomes	Percentage of the Cohort that graduated	0-100	Y
Borough	MathMean	Integer		Math Test Results	Average scale score for math test	min-max	Y
Borough	MathLevelOnePct	Double		Math Test Results	Students performing at this level are well below proficient in standards for their grade and have knowledge, skills, and practices that are insufficient for the expectations at their grade	0-100	
Borough	MathLevelTwoPct	Double		Math Test Results	Students performing at this level are partially proficient in standards for their grade and have knowledge, skills, and practices that are partial yet insufficient for the expectations at their grade	0-100	
Borough	MathLevelThreePct	Double		Math Test Results	Students performing at this level are proficient in standards for their grade and have knowledge, skills, and practices that are sufficient for the expectations at their grade	0-100	
Borough	MathLevelFourPct	Double		Math Test Results	Students performing at this level excel in standards for	0-100	

					their grade and have knowledge, skills, and practices that are more than sufficient for the expectations at their grade		
Borough	EnglishMean	Integer		ELA Test Results	Average scale score for English test	min-max	Y
Borough	EnglishLevelOnePct	Double		ELA Test Results	Students performing at this level are well below proficient in standards for their grade and have knowledge, skills, and practices that are insufficient for the expectations at their grade	0-100	
Borough	EnglishLevelTwoPct	Double		ELA Test Results	Students performing at this level are partially proficient in standards for their grade and have knowledge, skills, and practices that are partial yet insufficient for the expectations at their grade	0-100	
Borough	EnglishLevelThreePct	Double		ELA Test Results	Students performing at this level are proficient in standards for their grade and have knowledge, skills, and practices that are sufficient for the expectations at their grade	0-100	
Borough	EnglishLevelFourPct	Double		ELA Test Results	Students performing at this level excel in standards for their grade and have knowledge, skills, and practices that are more than sufficient for the expectations at their grade	0-100	
School District	SchoolDist	Integer	1-32	School District Shape File	Unique identifier for the school district		
School District	DistrictAttendance	Double		School Attendance and Enrollment by District	Average district attendance for the school district	0-100	Y
School District	DistrictEnrollment	Integer		School Attendance and Enrollment by District	Total enrollment for the school district	min-max	Y
School	MathMean	Integer		Math Test	Average scale score for math	min-	Y

District				Results	test	max	
School District	MathLevelOnePct	Double		Math Test Results	Students performing at this level are well below proficient in standards for their grade and have knowledge, skills, and practices that are insufficient for the expectations at their grade	0-100	
School District	MathLevelTwoPct	Double		Math Test Results	Students performing at this level are partially proficient in standards for their grade and have knowledge, skills, and practices that are partial yet insufficient for the expectations at their grade	0-100	
School District	MathLevelThreePct	Double		Math Test Results	Students performing at this level are proficient in standards for their grade and have knowledge, skills, and practices that are sufficient for the expectations at their grade	0-100	
School District	MathLevelFourPct	Double		Math Test Results	Students performing at this level excel in standards for their grade and have knowledge, skills, and practices that are more than sufficient for the expectations at their grade	0-100	
School District	EnglishMean	Integer		ELA Test Results	Average scale score for English test	0-100	Y
School District	EnglishLevelOnePct	Double		ELA Test Results	Students performing at this level are well below proficient in standards for their grade and have knowledge, skills, and practices that are insufficient for the expectations at their grade	0-100	
School District	EnglishLevelTwoPct	Double		ELA Test Results	Students performing at this level are partially proficient in standards for their grade and have knowledge, skills, and practices that are partial yet insufficient for the expectations at their grade	0-100	
School District	EnglishLevelThreePct	Double		ELA Test Results	Students performing at this level are proficient in standards for their grade and have knowledge, skills, and practices that are sufficient	0-100	

					for the expectations at their grade		
School District	EnglishLevelFourPct	Double		ELA Test Results	Students performing at this level excel in standards for their grade and have knowledge, skills, and practices that are more than sufficient for the expectations at their grade	0-100	
School	ADDRESS	Text		School Point Shape File	Address of the school		
School	ATS_CODE	Text		School Point Shape File	Unique identifier for a school		
School	BORONUM	Integer	1, 2, 3, 4, 5	School Point Shape File	Unique identifier for the borough; Used to join back to the borough data		
School	GEO_DISTRI	Integer	1-32	School Point Shape File	Unique identifier for the school district; Used to join back to the school district data		
School	GRADES	Text		School Point Shape File	Concatenated list of all of the grades the school serves		
School	SCHOOLNAME	Text		School Point Shape File	Text name for the school		
School	SCH_TYPE	Text		School Point Shape File	The level of the school (Elementary, High, etc.)		
School	GraduationOutcome DroppedOutPercent ageCohort	Double		Graduation Outcomes	Percentage of the Cohort that did not graduate because of drop out	0-100	
School	GraduationOutcome StillEnrolledPercent ageCohort	Double		Graduation Outcomes	Percentage of the Cohort that did not graduate, yet are still enrolled	0-100	
School	GraduationOutcome TotalGradPercentag e	Double		Graduation Outcomes	Percentage of the Cohort that graduated	0-100	Y
School	StudentToTeacherRa tio	Integer		Class Size	Overall ratio of student to teacher at the school	min-max	Y
School	QualityScore	Text	U, D, P, WD	Quality Review	Underdeveloped, Developing, Proficient, Well Developed	U, D, P, WD	Y
School	ProgressReportAddit ionalCredit	Double		School Progress Reports	Progress report additional credit for helping special education students, English Language Learners, and other high need students make exemplary progress.	min-max	
School	ProgressReportEnvir	Double		School	Progress report score based	0-15	

	onmentCategoryScore			Progress Reports	on attendance and results of parent, student, and teacher surveys. 15% of total score.		
School	ProgressReportOverallScore	Double		School Progress Reports	Overall combined progress report score.	0-100	Y
School	ProgressReportPerformanceCategoryScore	Double		School Progress Reports	Progress report score based on (for elementary/middle) English and Math Tests (for high school) Graduation Rate. 25% of total score.	0-25	
School	ProgressReportProgressCategoryScore	Double		School Progress Reports	Progress report score based on (for elementary/middle) student improvement on English and Math Tests (for high school) credit accumulation and regents completion and pass rates. 60% of total score.	0-60	
School	SurveyParentAcademicExpectationsScore	Double		NYC School Survey		0-10	
School	SurveyParentCommunicationScore	Double		NYC School Survey		0-10	
School	SurveyParentEngagementScore	Double		NYC School Survey		0-10	
School	SurveyParentResponseRate	Double		NYC School Survey		0-10	
School	SurveyParentSafetyRespectScore	Double		NYC School Survey		0-10	
School	SurveyStudentAcademicExpectationsScore	Double		NYC School Survey		0-10	
School	SurveyStudentCommunicationScore	Double		NYC School Survey		0-10	
School	SurveyStudentEngagementScore	Double		NYC School Survey		0-10	
School	SurveyStudentResponseRate	Double		NYC School Survey		0-10	
School	SurveyStudentSafetyRespectScore	Double		NYC School Survey		0-10	
School	SurveyTeacherAcademicExpectationsScore	Double		NYC School Survey		0-10	

School	SurveyTeacherCommunicationScore	Double		NYC School Survey		0-10	
School	SurveyTeacherEngagementScore	Double		NYC School Survey		0-10	
School	SurveyTeacherResponseRate	Double		NYC School Survey		0-10	
School	SurveyTeacherSafetyRespectScore	Double		NYC School Survey		0-10	
School	SurveyTotalAcademicExpectationsScore	Double		NYC School Survey		0-10	Y
School	SurveyTotalCommunicationScore	Double		NYC School Survey		0-10	Y
School	SurveyTotalEngagementScore	Double		NYC School Survey		0-10	Y
School	SurveyTotalResponseRate	Double		NYC School Survey		0-10	
School	SurveyTotalSafetyRespectScore	Double		NYC School Survey		0-10	Y
School	MajorRate	Double		School Safety	Ratio of number of major crimes to registered students. It includes the most serious personal and property crimes. The property crimes are burglary, grand larceny and grand larceny auto. The crimes against persons are murder, rape, robbery and felony assault.	min-max	Y
School	NoCrimRate	Double		School Safety	Ratio of number of non-criminal incidents to registered students. This category includes actions which are not classified as crimes but are nevertheless disruptive to the school environment. It includes disorderly conduct, harassment, loitering, possession of marijuana, dangerous instruments and trespass.	min-max	Y
School	OtherRate	Double		School	Ratio of number of other	min-	Y

				Safety	crimes to registered students. This category is composed of many crimes and incidents that range in severity. It includes reports of incidents such as arson/explosion, misdemeanor assault, criminal possession or sale of a controlled substance, sale of marijuana, criminal mischief, petit larceny, reckless endangerment, sex offenses (not including rape, which is included in the Major Crimes), and weapons possession.	max	
School	MathMean	Integer		Math Test Results	Average scale score for math test	min-max	Y
School	MathLevelOnePct	Double		Math Test Results	Students performing at this level are well below proficient in standards for their grade and have knowledge, skills, and practices that are insufficient for the expectations at their grade	0-100	
School	MathLevelTwoPct	Double		Math Test Results	Students performing at this level are partially proficient in standards for their grade and have knowledge, skills, and practices that are partial yet insufficient for the expectations at their grade	0-100	
School	MathLevelThreePct	Double		Math Test Results	Students performing at this level are proficient in standards for their grade and have knowledge, skills, and practices that are sufficient for the expectations at their grade	0-100	
School	MathLevelFourPct	Double		Math Test Results	Students performing at this level excel in standards for their grade and have knowledge, skills, and practices that are more than sufficient for the expectations at their grade	0-100	
School	EnglishMean	Integer		ELA Test Results	Average scale score for English test	min-max	Y
School	EnglishLevelOnePct	Double		ELA Test Results	Students performing at this level are well below	0-100	

					proficient in standards for their grade and have knowledge, skills, and practices that are insufficient for the expectations at their grade		
School	EnglishLevelTwoPct	Double		ELA Test Results	Students performing at this level are partially proficient in standards for their grade and have knowledge, skills, and practices that are partial yet insufficient for the expectations at their grade	0-100	
School	EnglishLevelThreePct	Double		ELA Test Results	Students performing at this level are proficient in standards for their grade and have knowledge, skills, and practices that are sufficient for the expectations at their grade	0-100	
School	EnglishLevelFourPct	Double		ELA Test Results	Students performing at this level excel in standards for their grade and have knowledge, skills, and practices that are more than sufficient for the expectations at their grade	0-100	

Exploratory Data Analysis

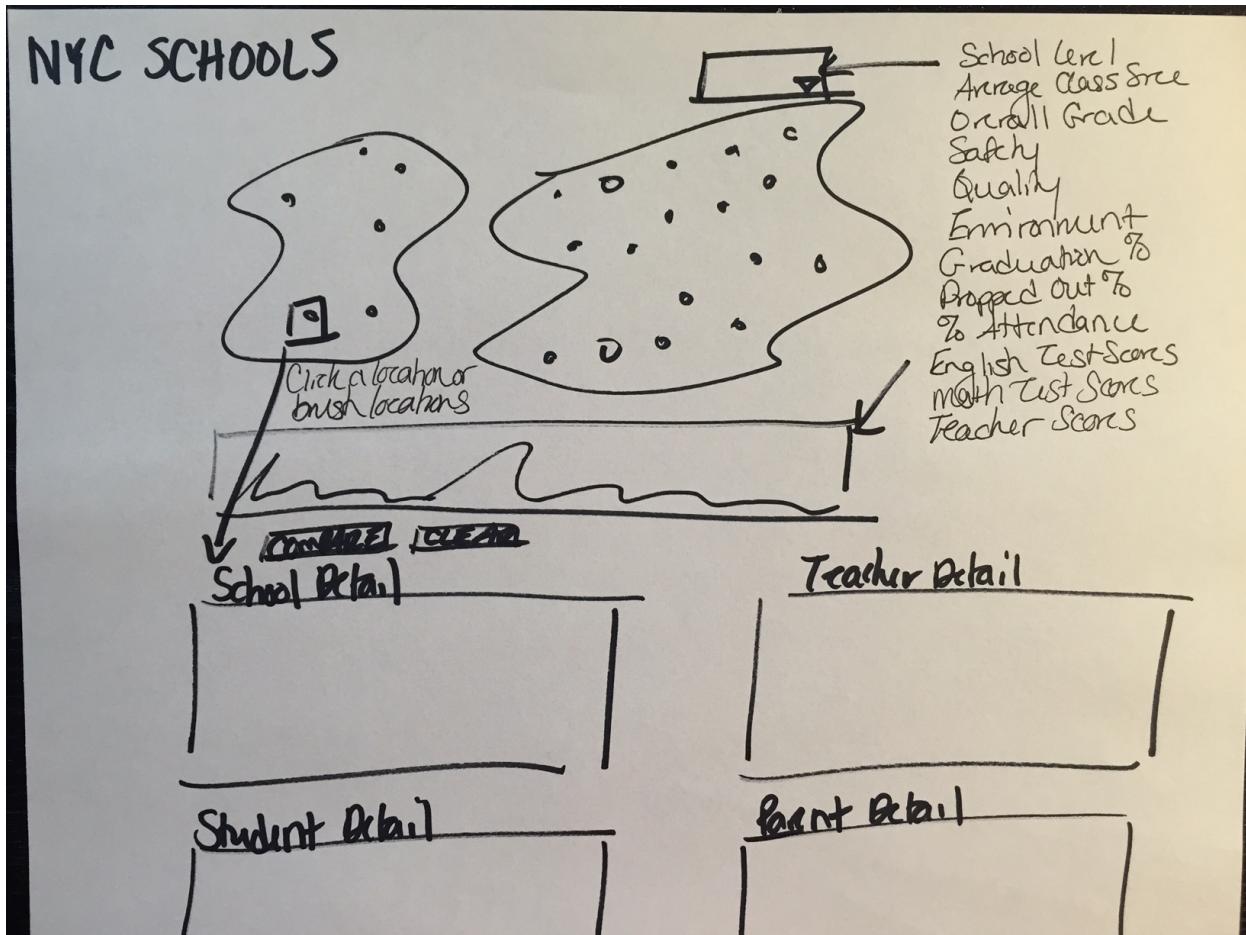
I explored my data in Access tables, with Tableau, and with a prototype for how to create a map. One of the biggest things I learned is that the schools do not always have every single metric available. I used Access to setup the data so that I could handle scenarios of missing data gracefully. I used Tableau to determine how I would create my percentile thresholds in D3.

Design Evolution

1. The first drawing of my design that I included in my proposal is below. It focused on enabling the user to select a metric type to use to color the boroughs, school districts, or schools on the map. After selecting a metric type, the user could brush the metric values of interest, which would then filter the items on the map. Clicking on the map item would update the detail views.

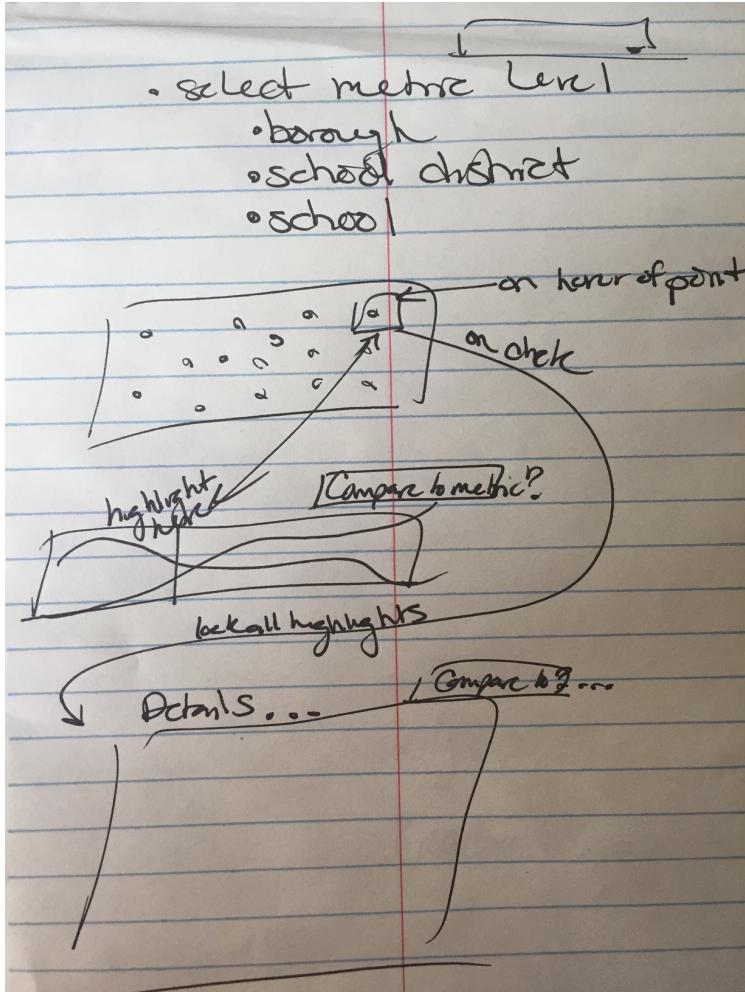
A key limitation of this design was that a user could only brush one metric. An additional key limitation was that the user could not see any of the relationships between the metrics, for instance, whether student to teacher ratio influences

graduation rate. The visualization was also missing a way to intuitively swap between borough, school district, and school metrics



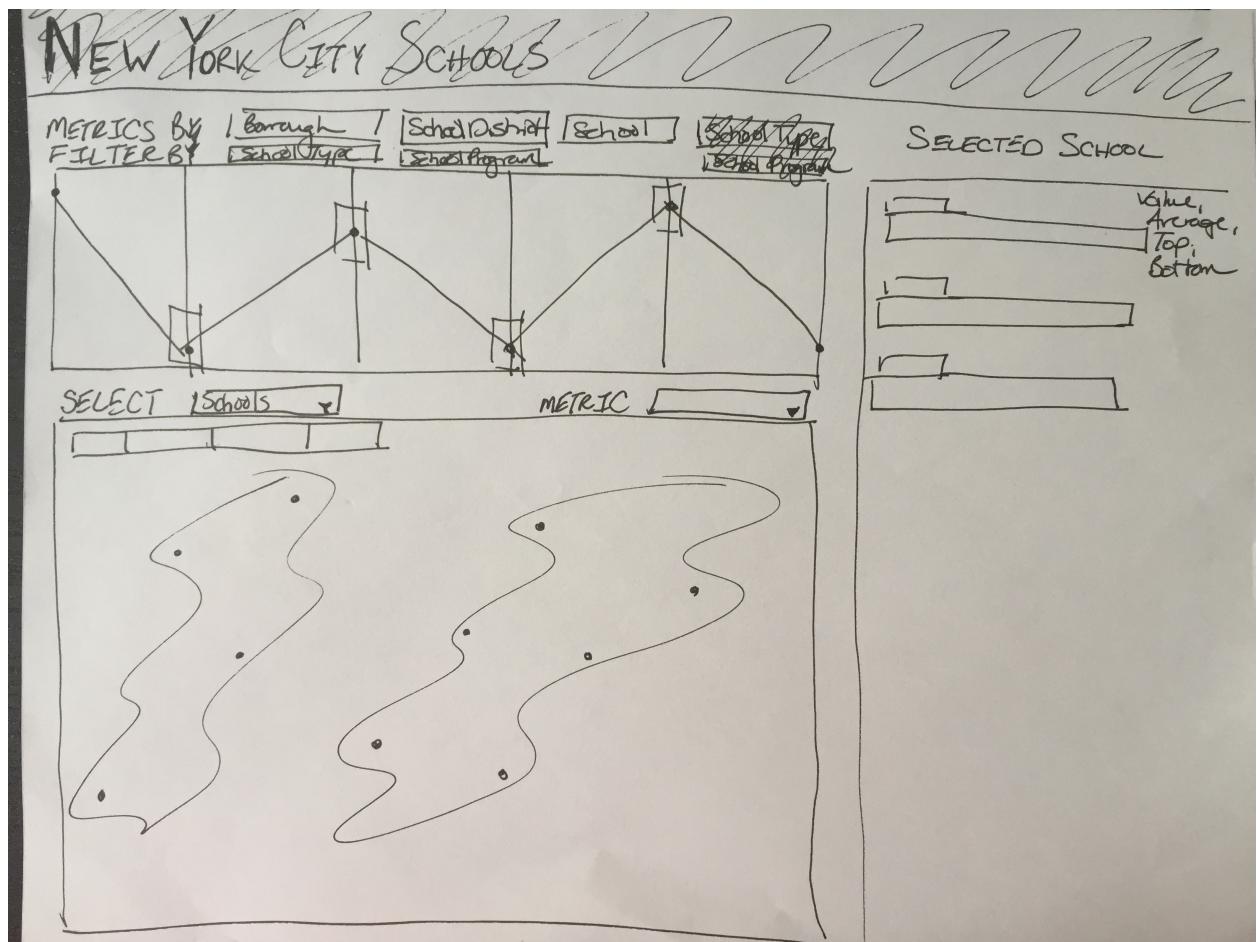
2. I consulted the Visualizing Time book (<http://www.springer.com/us/book/9780387779065>) and the D3 Gallery (<https://github.com/mbostock/d3/wiki/Gallery>) and reviewed a plethora of options to brush multiple metrics. At first, I was leaning toward a bar chart. However, this still prevented seeing relationships between the metrics. After much consideration, I selected the parallel coordinate line chart because it enabled seeing relationships between the metrics and brushing multiple metrics.

When I first setup the parallel coordinate line chart, I had the map on top, and the line chart on bottom. The workflow at that point was select the metric of interest, select the boroughs, school districts, or schools of interest, then narrow down the selected items using the parallel coordinate line chart.

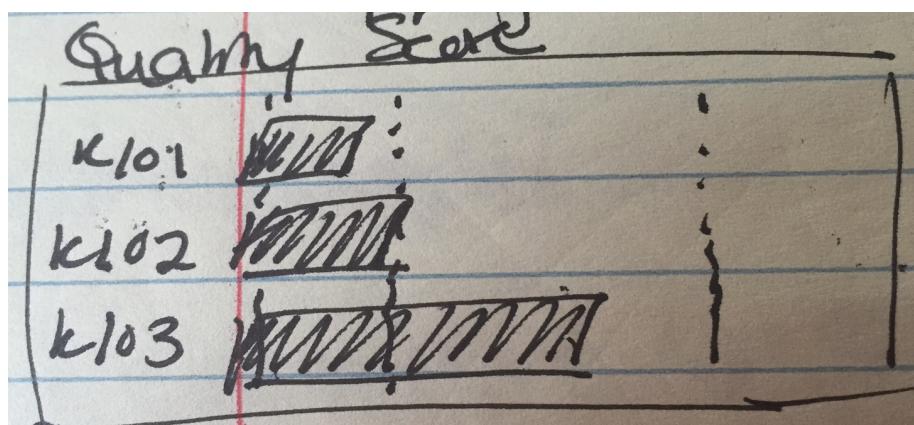


After implementing this, I realized that the order was backwards; the parallel coordinate line chart needed to be on top and the map needed to be on the bottom because the parallel coordinate line chart was the master and the map was the detail. It made much more sense from the user perspective to establish some filter parameters first, then pick some schools of interest

3. The next iteration of the design had the parallel coordinate line chart on top and the map on the bottom.



- The last major piece to setup was how to view the detail on the selected boroughs, school districts, or schools. Adhering to the principle of showing the data with charts instead of telling the data with text, I knew I would need to setup some small multiples. I chose to put them on the right hand side of the page because it made sense from the perspective of working from left to right. The lines represent the percentile markers – 75-100%, 50-75%, 25-50%, 0-25%.



5. The map is going to get a legend that updates with a percentile or ordinal scale based upon what metric is selected. Below is an example of the ordinal version. Points above 50% will be colored with positive color, blue, points below 50% will be colored with negative color, and points with no data will be invisible or white. The more positive or negative the value (i.e. closer to 100% or closer to 0%), the darker the color. Colors from the Color Brewer scale for diverging data were chosen, keeping in mind that it needed to look good for folks who are or are not colorblind.

If there is a need to show sequential data, a single color, multiple shade scale will be used.



Implementation

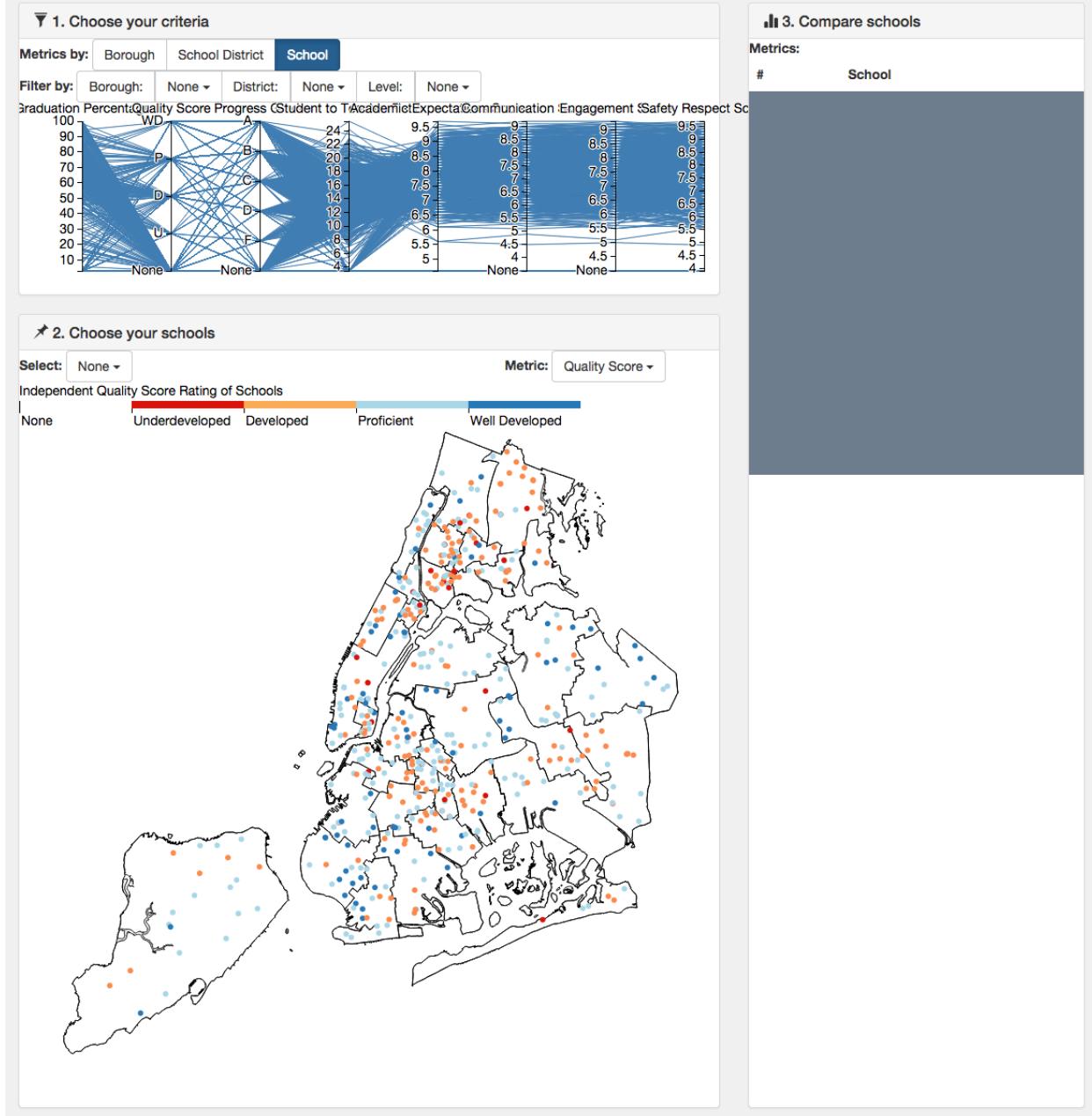
My implementation plan was as follows:

- Fail fast and try to implement the map, which I perceived to be the most difficult item to tackle. I got the map working on the first day, so I proceeded down the path of using the map for the visualization.
 - The map requires: topojson.v1.min.js
- Next I implemented some basic multiselect checkbox dropdowns for metric selection.
 - The multiselect checkbox dropdowns require: bootstrap-multiselect.js
- Then I implemented the parallel coordinate line chart.
- Then I added all of the additional buttons and dropdowns depicted. I also started to clean up all of the styling
- Then I started the small multiple bar chart implementation. This item remains a work in progress.

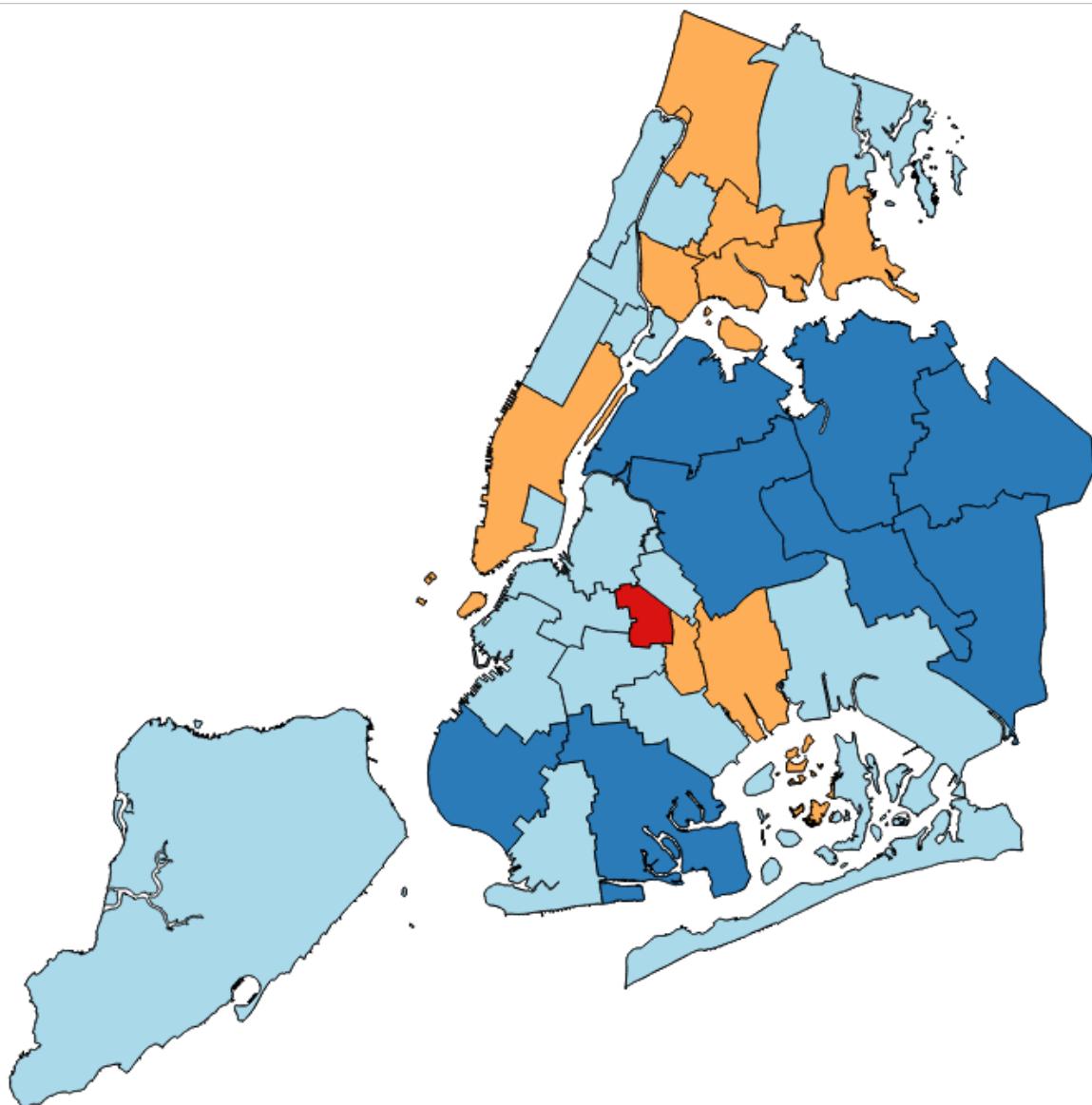
The remaining implementation tasks are detailed as “not yet implemented” in the rest of this section. Description of issues that need to be resolved and remaining additional work that needs to be completed follow in the evaluation section of this document.

Below is a full screenshot of the current state of the implementation.

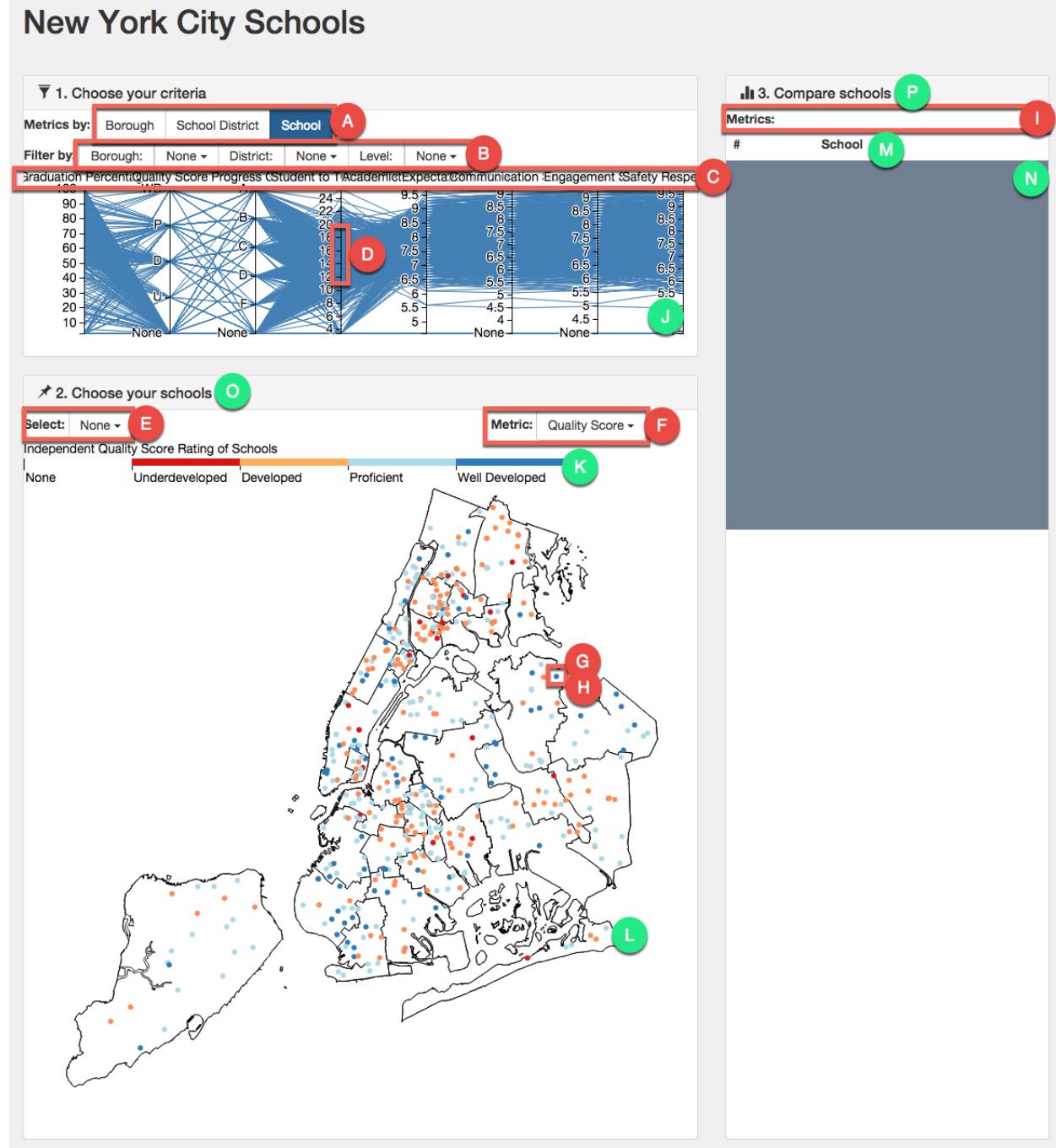
New York City Schools



Below is a screenshot example of the coloring when a school district metric is selected.



Below is a view of all of the user actions that can be performed and all of the components that can be updated. The color coding is: (a) red circles indicate a user action can be made, (b) green circles indicate areas that get updated, yet do not have any user actions that can be made. The discussion that follows this image describes all of the possible user actions and how they impact the visualization.

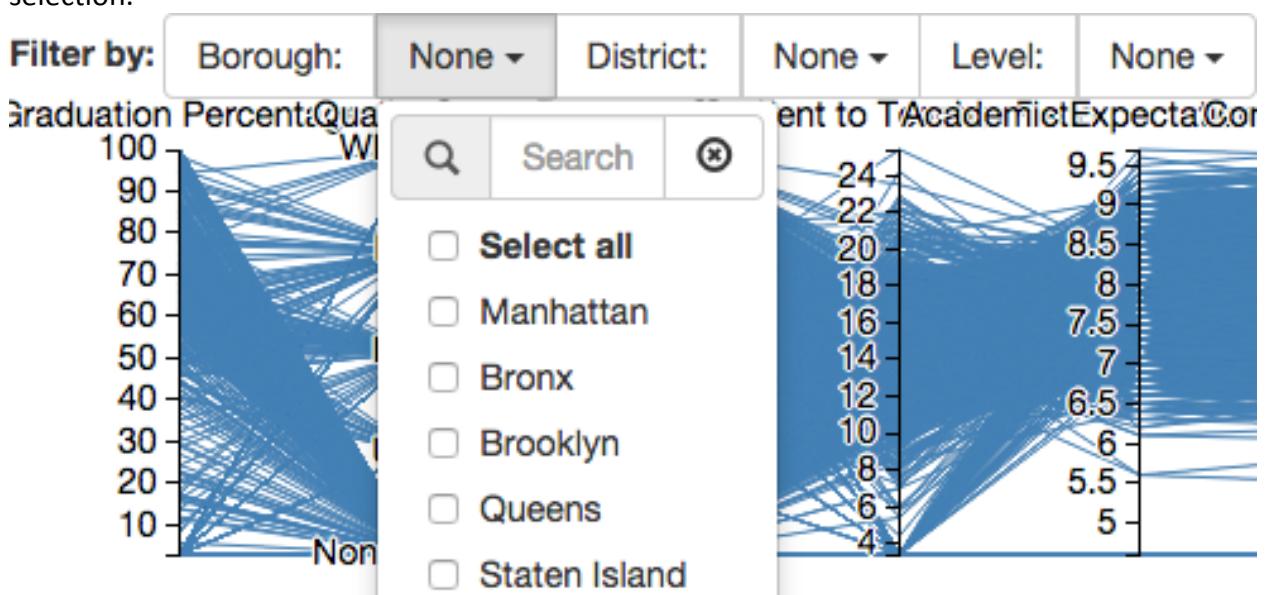


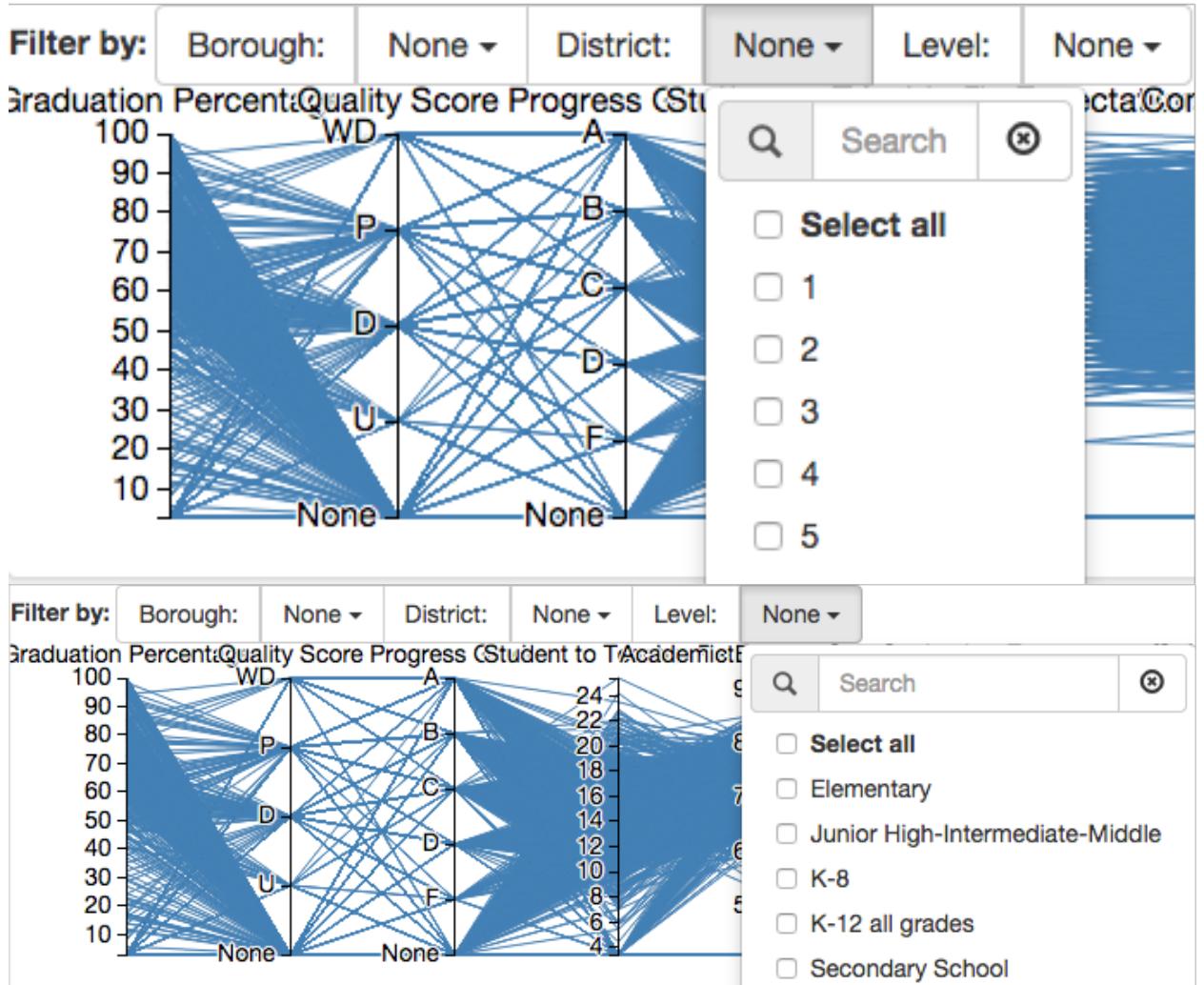
- A. The user can choose whether he/she wants to see metrics for boroughs, school districts, or schools. This is a button group.



Clicking on a button drives the following behavior:

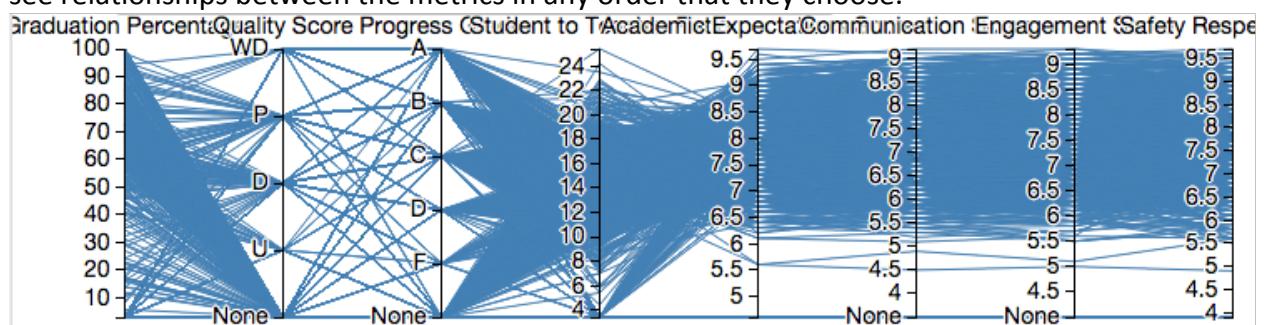
- a. If school is not the clicked button, clear and hide the filter by [B] [not yet implemented], else if school is the clicked button, show the filter by [B] [not yet implemented]
 - b. Update the parallel coordinate line chart [J] [not yet implemented]
 - c. Reset the selected items list [E] [not yet implemented]
 - d. Reset the selected metric list [F] [not yet implemented]
 - e. Update the map legend [K] [not yet implemented]
 - f. Update the map [L] [implemented]
 - g. Reset the metrics compare list [I] [not yet implemented]
 - h. Reset the selected items text list [M] [not yet implemented]
 - i. Reset the small multiples bar chars [N] [not yet implemented]
 - j. Update the “choose your” text [O] [not yet implemented]
 - k. Update the “compare” text [P] [not yet implemented]
- B. If the user has selected to see metrics by school, the user can choose whether he/she wants to filter by Borough [implemented], District [implemented], Level [implemented], or Program [not yet implemented]. All of these filters allow checkbox multiple selection.





Selecting or deselecting a filter drives the following behavior:

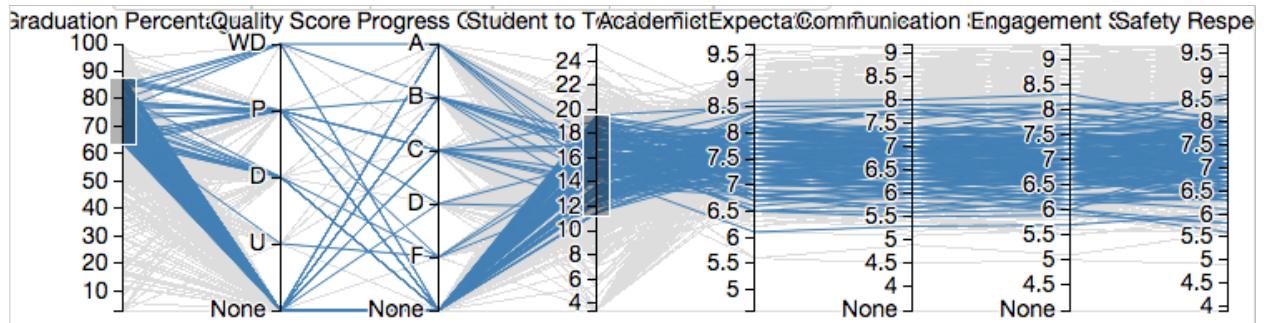
- Filter the parallel coordinate line chart [J] [implemented]
 - Filter the map [L] [implemented]
- C. The user can choose to reorder the parallel coordinate line chart key metric columns to see relationships between the metrics in any order that they choose.



Clicking, dragging, and dropping the column headers drives the following behavior:

- Reorder the columns on the parallel coordinate line chart [J] [implemented]

- D. The user can choose to brush one to many metrics on the parallel coordinate line chart to filter the items to their criteria.



Brushing drives the following behavior:

- Gray out the lines that are outside of the brushed selection on the parallel coordinate line chart [J] [implemented]
 - Filter the map [L] [implemented]
- E. The user can select one to many boroughs, school districts, or schools on the map by choosing the item from the select dropdown. The select dropdown allows checkbox multiple selection. This supports the use case of the user knowing exactly what borough, district, or school he/she wants to view and compare.

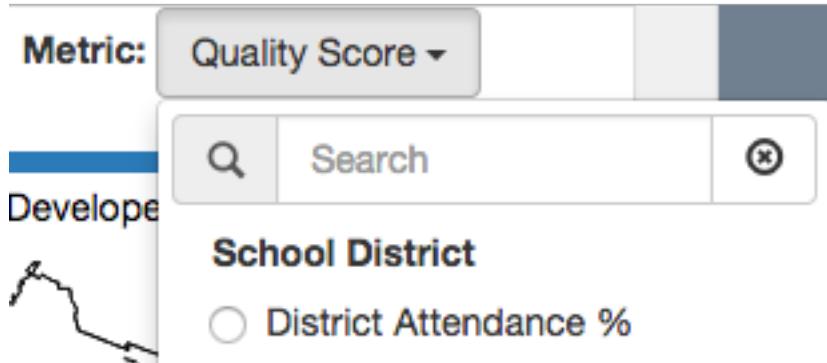
Select: None ▾
Metric:

Independence	<input type="checkbox"/> Select all	<input type="checkbox"/> P.S. 001 THE BERGEN	<input type="checkbox"/> New York City Montessori Charter School	<input type="checkbox"/> P.S. 095 THE GRAVESEND	<input type="checkbox"/> I.S. 096 SETH LOW	<input type="checkbox"/> P.S. 97 THE HIGHLAWN	<input type="checkbox"/> I.S. 98 BAY ACADEMY	<input type="checkbox"/> P.S. 099 ISAAC ASIMOV	<input type="checkbox"/> P.S. 100 THE CONEY ISLAND SCHOOL	<input type="checkbox"/> P.S. 101 THE VERRAZANO
None	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Selecting an item toggles the following behavior:

- Highlight or remove highlight from the selected item on the map [L] [not yet implemented]
- Add or remove the selected item to the selected item list [M] [not yet implemented]
- Draw the small multiple bar charts for the default or selected metrics compare list [N] [not yet implemented]

- F. The user can select which metric to use to color the map. The metric dropdown allows selection of one metric at a time to color the map.



Selecting a metric drives the following behavior:

- Update the map legend [K] [not yet implemented]
- Update the map [L] [implemented]

- G. The user can hover over an item on the map and see a summary of the item (name and value for current selected metric).



Hovering over an item drives the following behavior:

- Hover styling [G] [implemented]
- Show tooltip [G] [not yet implemented]

- H. The user can click an item on the map to select it. This causes the same actions as selecting the item from the select dropdown. This supports the use case of the user not knowing the name of the borough, district, or school he/she wants to view and compare, yet identifying it of interest after exploring the items in the visualization.

Clicking an item on the map toggles the following behavior:

- Highlight or remove highlight from the selected item on the map [L] [not yet implemented]
- Add or remove the selected item to the selected item list [M] [implemented]
- Draw the small multiple bar charts for the default or selected metrics compare list [N] [not yet implemented]

- I. The user can select one to many metrics in the metrics compare list to view small multiple bar charts for the selected items and the selected comparison metrics. The metrics comparison dropdown allows checkbox multiple selection. Selecting a comparison metric drives the following behavior:

- Draw the small multiple bar charts for the selected metrics [N] [not yet implemented]

Evaluation

Issues

There currently exist the following issues that need to be resolved:

- Major performance issues on load/update of visualizations
- Parallel coordinate line chart axes need to brought back to the front after changing filters
- Brushes need to get reset on changing filters

Remaining Work

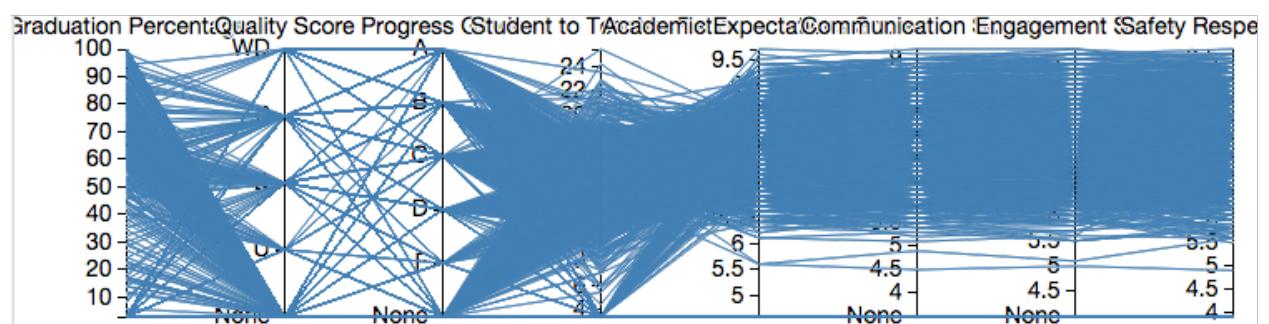
The following work remains needing to be completed:

- Implement all items tagged [not yet implemented] in the implementation section
- Refactor code to promote best practices and optimization, as some code is left over from prototyping and is not ideal
- Execute performance profiling to identify root cause of major performance issues and refactor code to resolve performance issues
- Decide on the final set of metrics to make available and remove data from JSON for metrics that will not be used
- Complete stylesheets for the whole page to make it pretty
- Create landing page for site
- Create pitch video

Learnings

Mapping data on the map helped to visualize errors in the source data and resolve the errors. I found that the source data had Brooklyn coded as the Bronx, and I found that the source data had Queens labeled Oueens.

There are many cases where a school does not have a metric because it was not represented in the source data. Those schools are represented on the visualizations as “None”. Missing metrics are the root cause of the lines shooting down to the min point on the y axis across metrics as depicted below.



Do the schools with low graduation percentages have low academic expectations scores?
Surprisingly, no. The below image has the no data values brushed out, and I can see that even schools with graduation rates as low as 10% have academic expectations scores above 6 out of 10.

