TWO SIGMA

www.twosigma.com

The state of open data on school bullying and harassment in NYC

Two Sigma Data Clinic

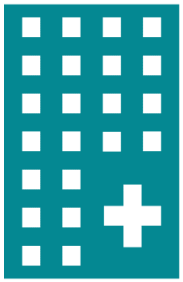
NYC Open Data Week
March 6th, 2018

March 6, 2018

Important legal information

The views expressed herein are not necessarily the views of Two Sigma Investments, LP or any of its affiliates (collectively, “Two Sigma”). The information presented herein is only for informational and educational purposes and is not an offer to sell or the solicitation of an offer to buy any securities or other instruments. Additionally, the information is not intended to provide, and should not be relied upon for investment, accounting, legal or tax advice. Two Sigma makes no representations, express or implied, regarding the accuracy or completeness of this information, and you accept all risks in relying on this information for any purpose whatsoever.

This presentation shall remain the property of Two Sigma and Two Sigma reserves the right to require the return of this presentation at any time. Copyright © 2018 TWO SIGMA INVESTMENTS, LP. All rights reserved.



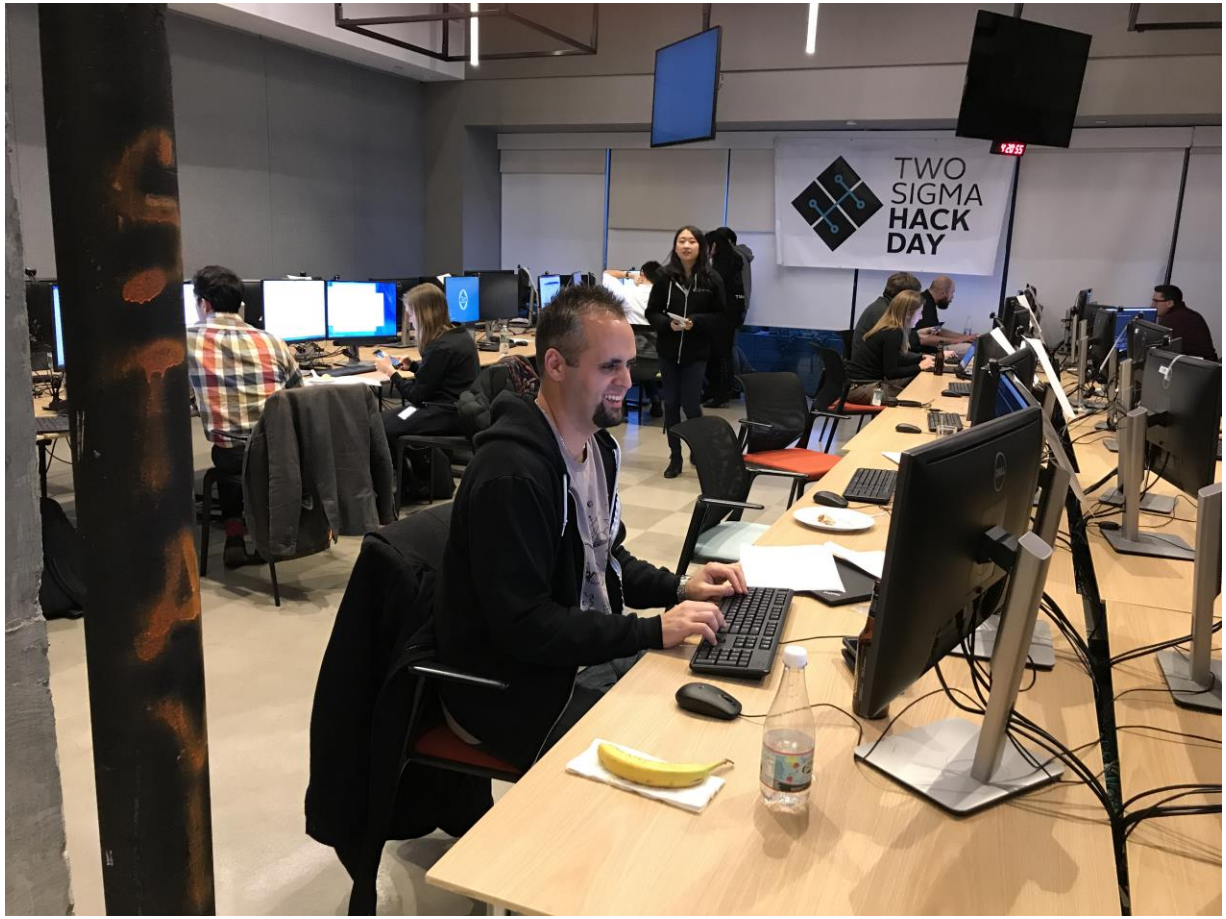
TWO SIGMA
DATA CLINIC

est. 2014

pro-bono data and tech for social good

www.twosigma.com/dataclinic

#nyc #opendata #opendataweek
@twosigma


TWO
SIGMA
HACK
DAY

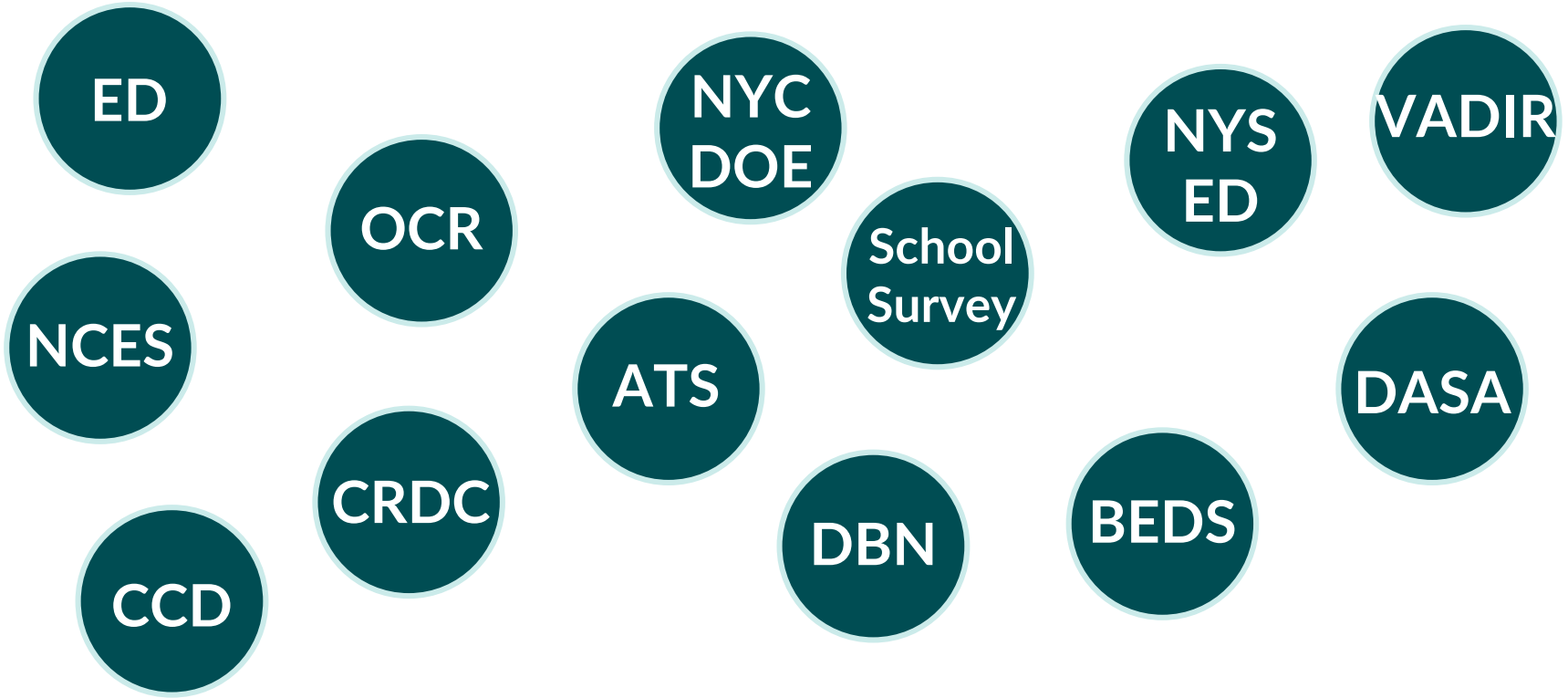
Open Data is an invitation for anyone anywhere anytime to engage with New York City

"Open Data for All" (2015)

http://www1.nyc.gov/assets/home/downloads/pdf/reports/2015/NYC-Open-Data-Plan-2015.pdf

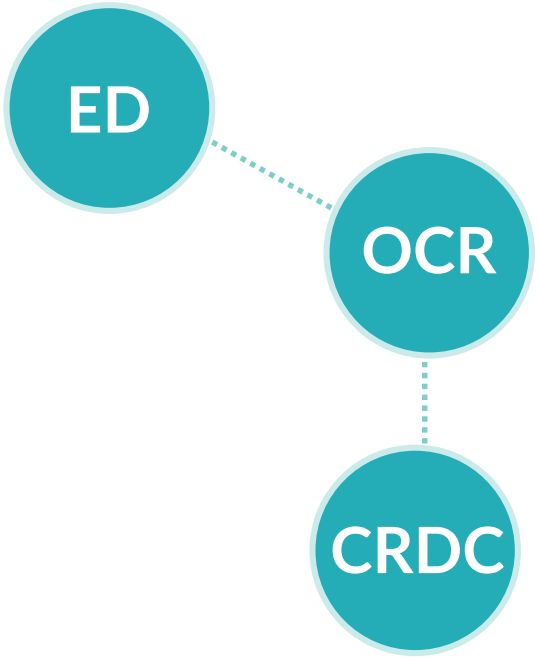
Open data on school bullying and harassment in NYC

Navigating the landscape



Open data on school bullying and harassment in NYC

Federal



- U.S. Department of Education (ED)
  - Office for Civil Rights (OCR)
    - **Civil Rights Data Collection (CRDC)**
      95,507 schools nationwide

Open data on school bullying and harassment in NYC

Federal

Civil Rights Data Collection (CRDC)

2013-14 school year

Allegations of Harassment or Bullying

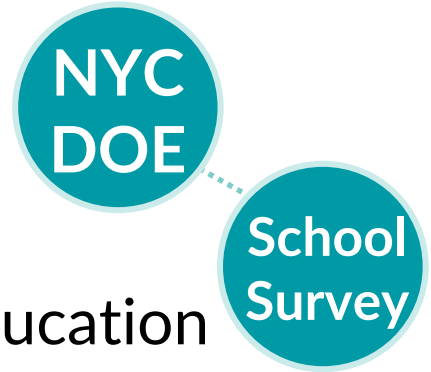
Students Reported as Harassed or Bullied

Students Disciplined for Harassment or Bullying

Race-, Sex-, Disability- based

Open data on school bullying and harassment in NYC

Local



- New York City Department of Education (NYCDOE)
  - **NYC School Survey**
    1,784 schools citywide

Open data on school bullying and harassment in NYC

Local

NYC School Survey

2013-14 school year

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences …

Open data on school bullying and harassment in NYC

Local

NYC School Survey

2013-14 school year

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences …

None of the time ⟷ All of the time

Open data on school bullying and harassment in NYC

Local

NYC School Survey

2013-14 school year

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences …

None of the time ⟷ All of the time

Teachers

At my school, students are often harassed or bullied in school.

At my school, there are conflicts based on differences …

Open data on school bullying and harassment in NYC

Local

NYC School Survey

2013-14 school year

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences ...

None of the time ⟷ All of the time

Teachers

At my school, students are often harassed or bullied in school.

At my school, there are conflicts based on differences ...

Strongly Disagree ⟷ Strongly Agree

Open data on school bullying and harassment in NYC

Local

NYC School Survey

2013-14 school year

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences ...

None of the time ⟷ All of the time

Parents

At my child's school students harass or bully other students.

At my child's school students harass or bully each other based on differences ...

Teachers

At my school, students are often harassed or bullied in school.

At my school, there are conflicts based on differences ...

Strongly Disagree ⟷ Strongly Agree

Open data on school bullying and harassment in NYC

Local

NYC School Survey

2013-14 school year

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences ...

None of the time ⟷ All of the time

Parents

At my child's school students harass or bully other students.

At my child's school students harass or bully each other based on differences ...

Strongly Disagree ⟷ Strongly Agree

Teachers

At my school, students are often harassed or bullied in school.

At my school, there are conflicts based on differences ...

Strongly Disagree ⟷ Strongly Agree

Open data on school bullying and harassment in NYC

Local

NYC School Survey

2013-14 school year

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences ...

None of the time ⟷ All of the time

Parents

At my child's school students harass or bully other students.

At my child's school students harass or bully each other based on differences ...

Strongly Disagree ⟷ Strongly Agree

+ Don't Know

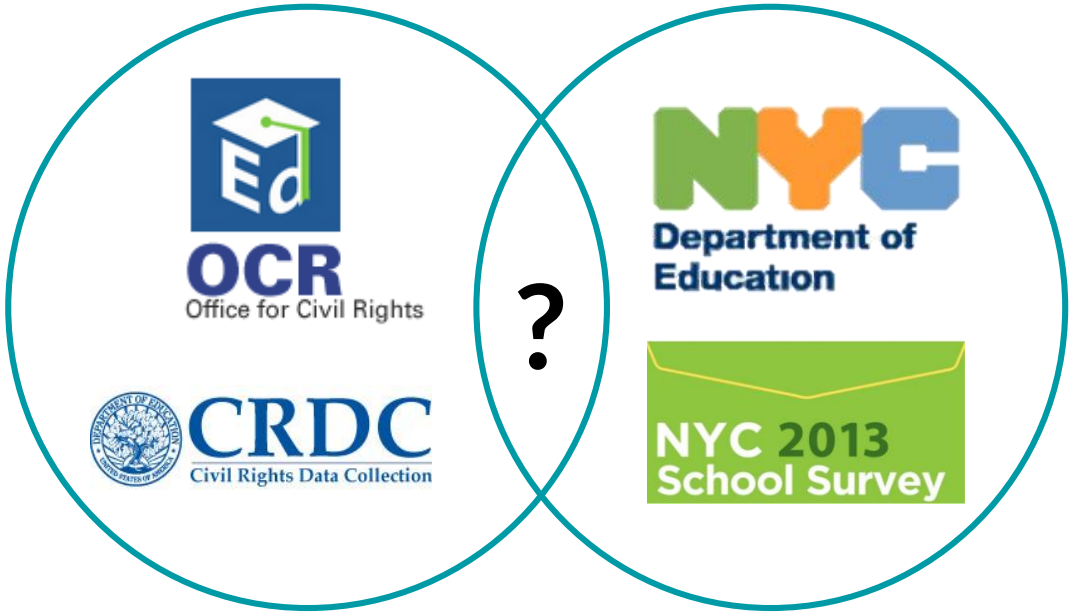
Teachers

At my school, students are often harassed or bullied in school.

At my school, there are conflicts based on differences ...

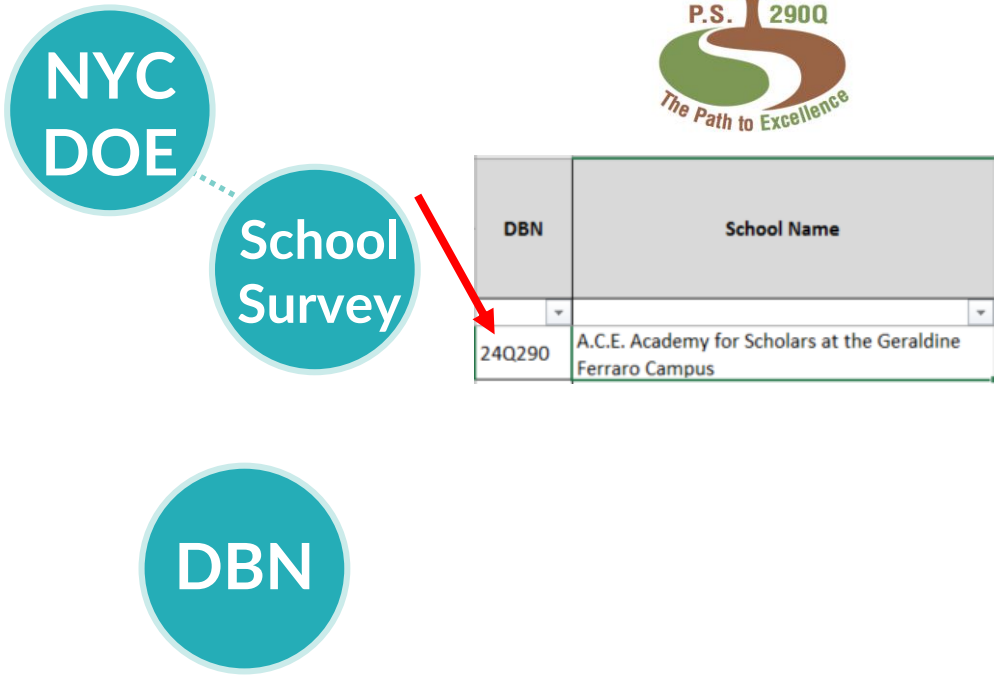
Strongly Disagree ⟷ Strongly Agree

Bringing federal and local together

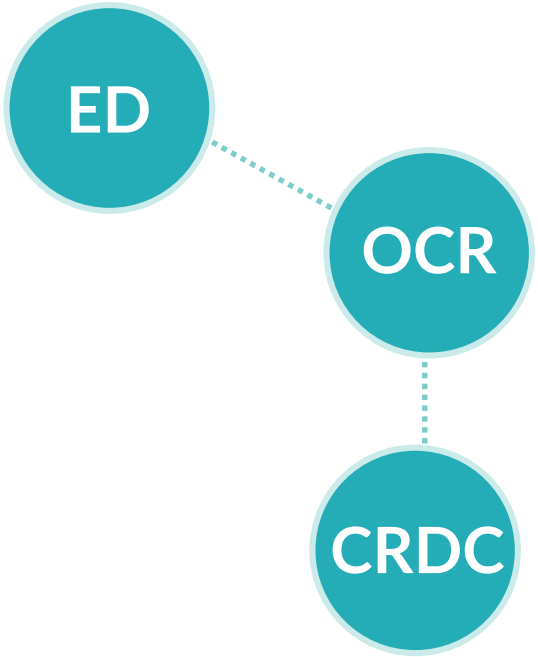


Bringing federal and local together

Schools in the NYC School Survey are identified by a District Borough Number (DBN) …



Bringing federal and local together



SCH_NAME	COMBOKEY
PS 290	362058006223

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

… while schools in the Office for Civil Rights' Civil Rights Data Collection are identified by a 12-digit numeric ID called the "COMBOKEY"

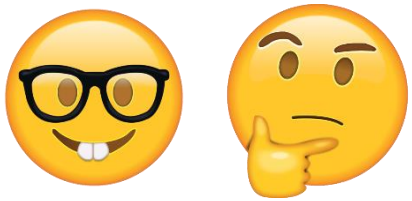
Bringing federal and local together: what we *didn't* do

```
def levenshteinDistance(str1, str2):
    m = len(str1)
    n = len(str2)
    lensum = float(m + n)
    d = []
    for i in range(m+1):
        d.append([i])
    del d[0][0]
    for j in range(n+1):
        d[0].append(j)
    for j in range(1,n+1):
        for i in range(1,m+1):
            if str1[i-1] == str2[j-1]:
                d[i].insert(j,d[i-1][j-1])
            else:
                minimum = min(d[i-1][j]+1, d[i][j-1]+1, d[i-1][j-1]+2)
                d[i].insert(j, minimum)
    ldist = d[-1][-1]
    ratio = (lensum - ldist)/lensum
    return {'distance':ldist, 'ratio':ratio}

print(levenshteinDistance("A.C.E. Academy for Scholars at the Geraldine Ferraro Campus",
                          "PS 290"))
```

```
{'distance': 61, 'ratio': 0.06153846153846154}
```

- P.S. 188 The Island School
- PS 188 ISLAND SCHOOL (THE)



NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

Office for Civil Rights

SCH_NAME	COMBOKEY
PS 290	362058006223

Bringing federal and local together: what we did

NYC DOE

NYC Open Data Portal: School Locations

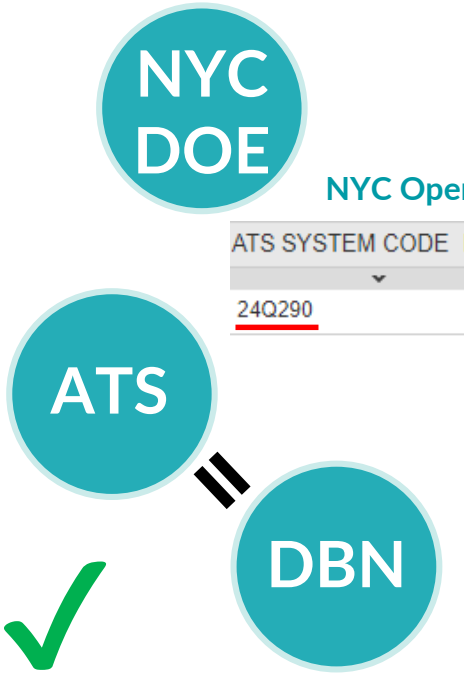
ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

DBN

Bringing federal and local together: what we did

NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus



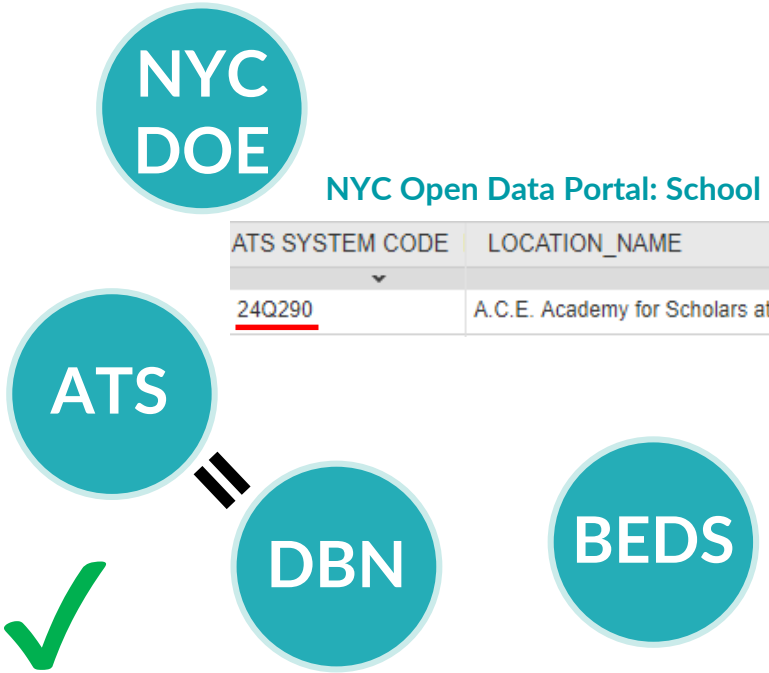
NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

Bringing federal and local together: what we did

NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus



NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

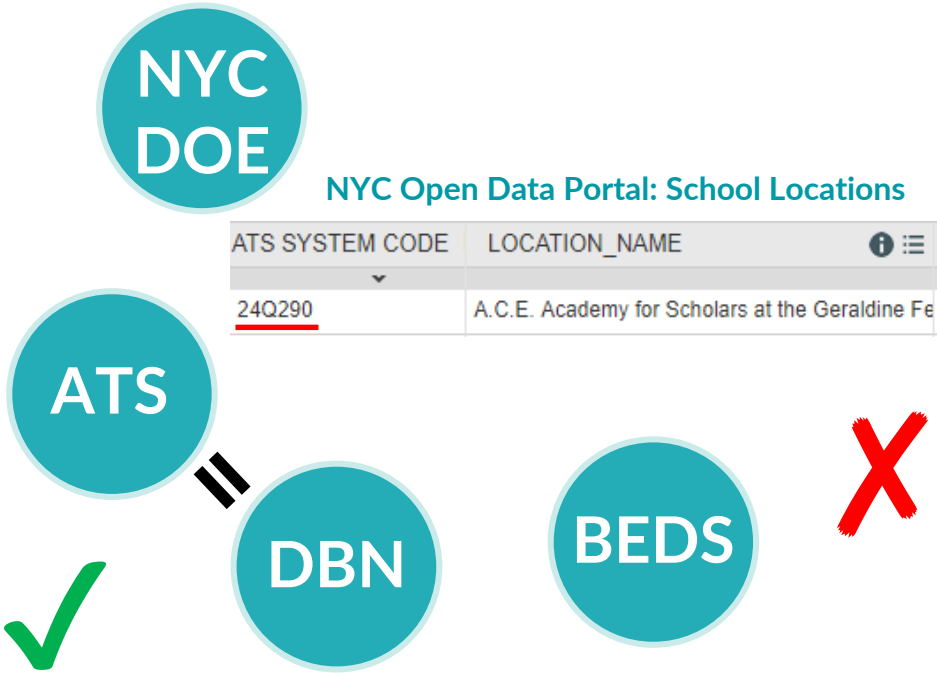
Bringing federal and local together: what we did

NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

Office for Civil Rights

SCH_NAME	COMBOKEY
PS 290	362058006223



NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

Bringing federal and local together: what we did

NYC DOE

NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

DBN

BEDS

Bringing federal and local together: what we did

NYC DOE

NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

DBN

BEDS

Bringing federal and local together: what we did

ED

NYC DOE

DBN

BEDS

Common Core of Data

NCESSCH	SEASCH	SCHNAM
360009806223	342400010290	PS 290

NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

Bringing federal and local together: what we did

ED

NCES

CCD

NYC DOE

DBN

BEDS

Common Core of Data

NCESSCH	SEASCH	SCHNAM
360009806223	342400010290	PS 290

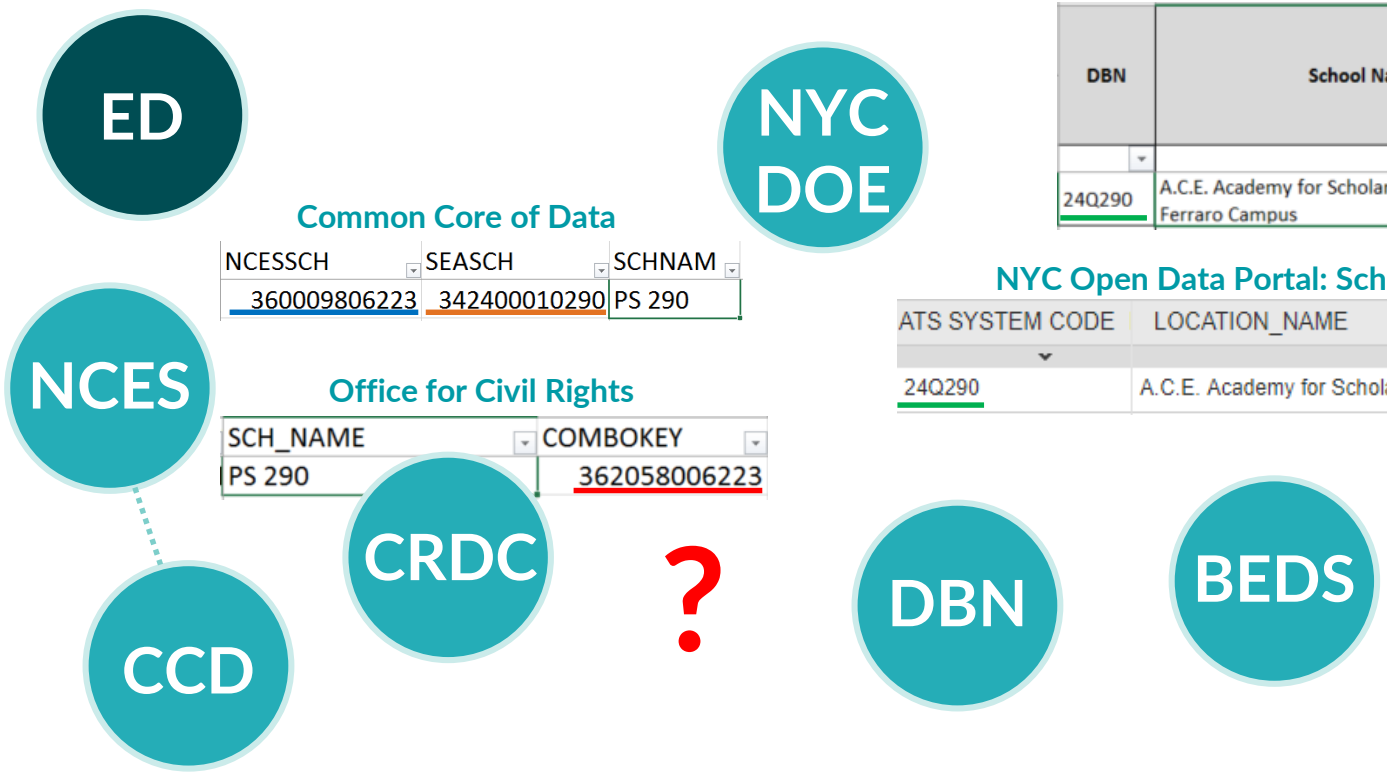
NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

Bringing federal and local together: what we did



Bringing federal and local together: what we did

ED

NCES

CCD

NYC DOE

NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

Common Core of Data

NCESSCH	SEASCH	SCHNAM
360009806223	342400010290	PS 290

Office for Civil Rights

SCH_NAME	COMBOKEY
PS 290	362058006223

CRDC

X

DBN

NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

BEDS

Bringing federal and local together: what we did

ED

NYC DOE

NCES

CCD

CRDC

DBN

BEDS

NYC School Survey

DBN	School Name
24Q290	A.C.E. Academy for Scholars at the Geraldine Ferraro Campus

Common Core of Data

NCESSCH	SEASCH	SCHNAM
360009806223	342400010290	PS 290

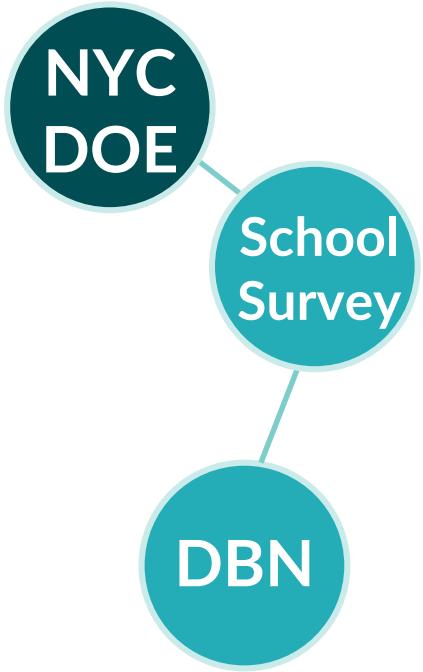
NYC Open Data Portal: School Locations

ATS SYSTEM CODE	LOCATION_NAME	BEDS NUMBER
24Q290	A.C.E. Academy for Scholars at the Geraldine Fe	342400010290

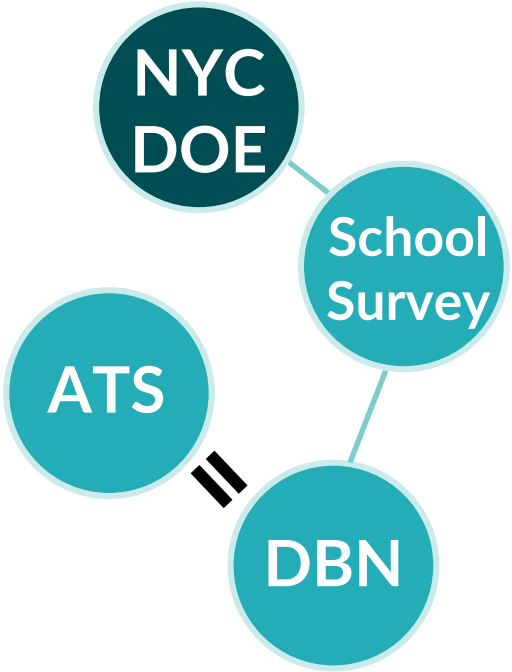
Office for Civil Rights

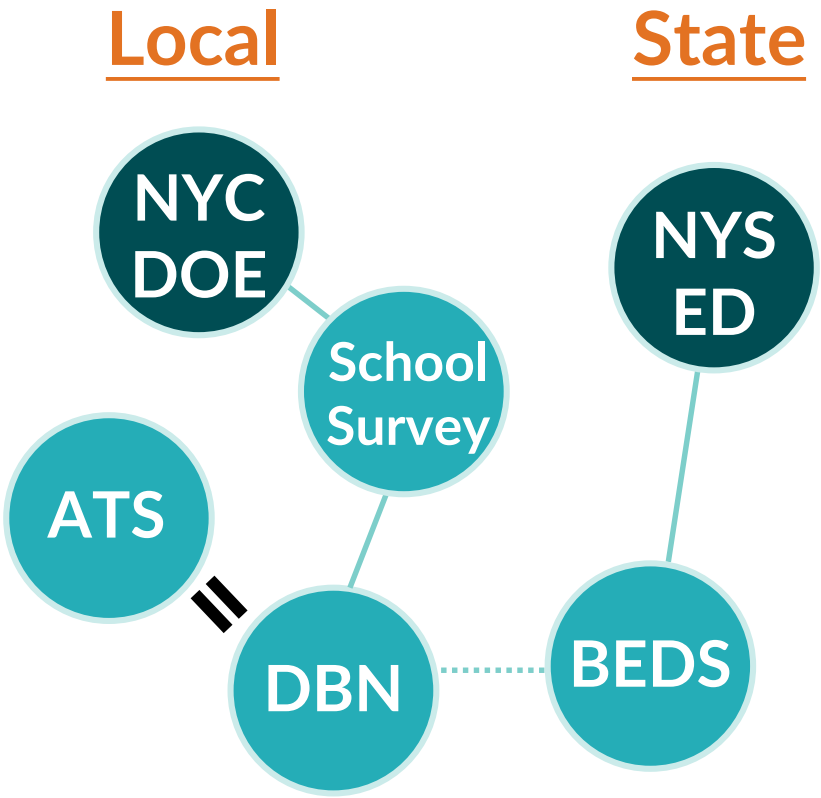
SCH_NAME	COMBOKEY	NCES_SCHOOL_ID
PS 290	362058006223	360009806223

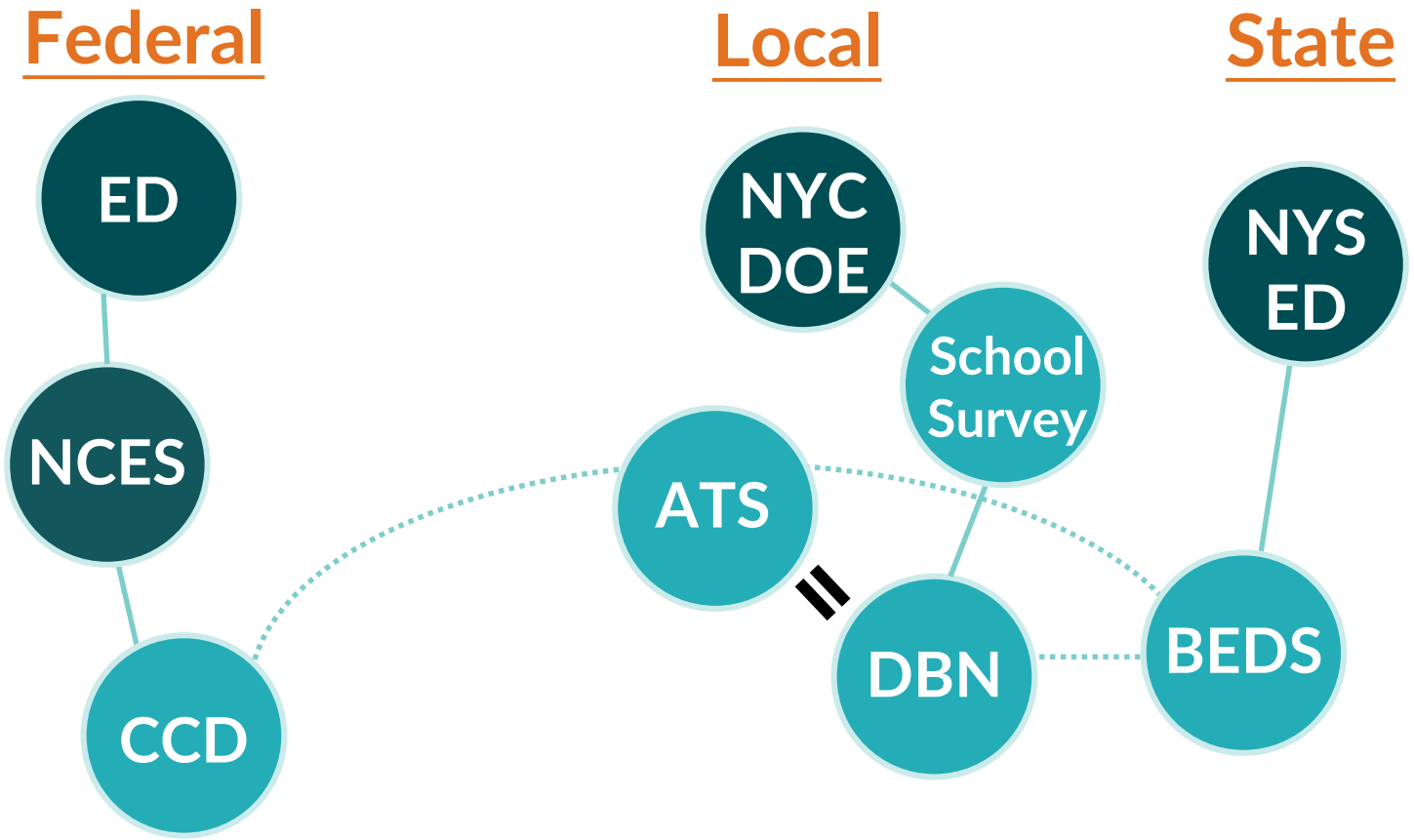
Local

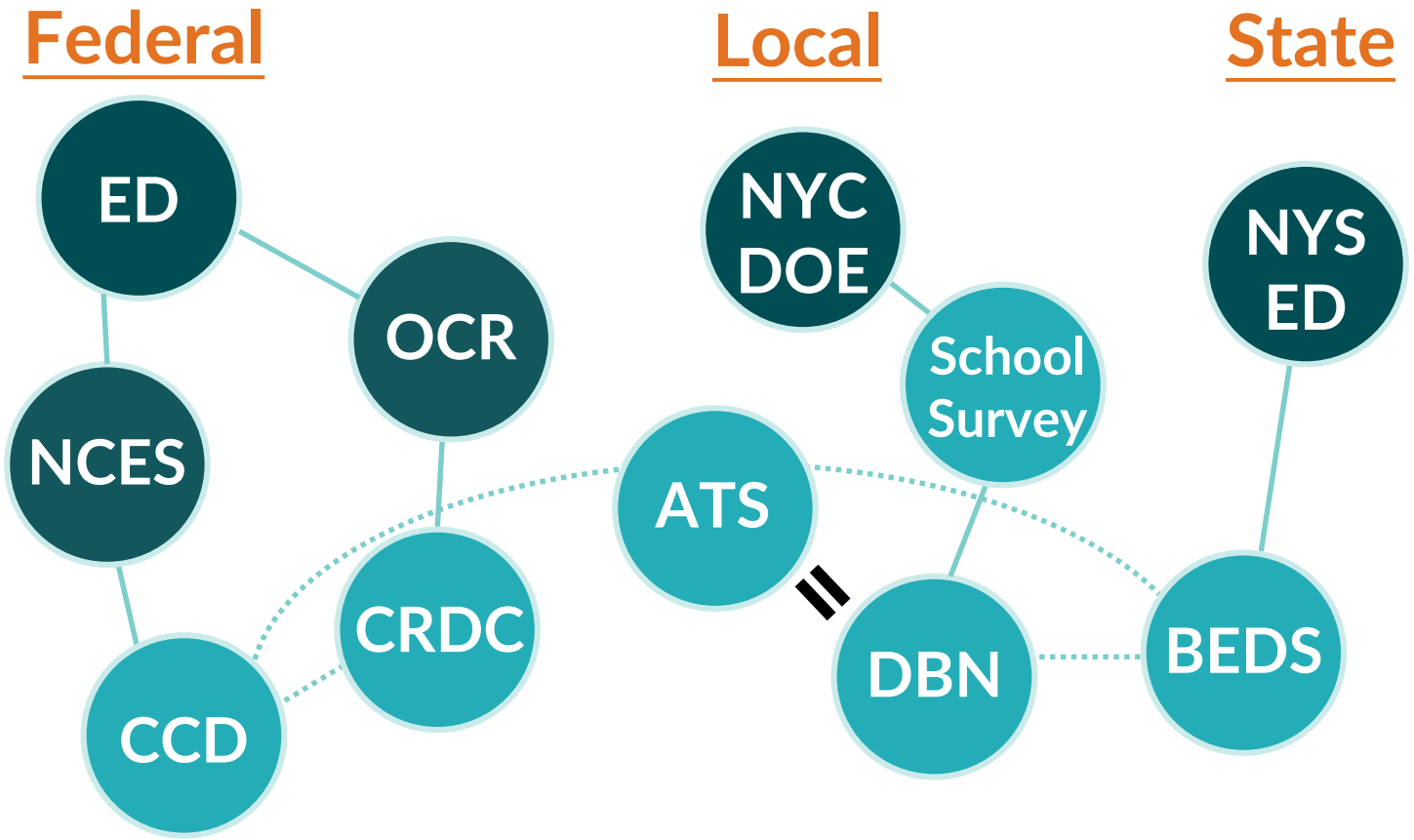


Local




Local
State
NYC DOE
School Survey
NYS ED
ATS
DBN
BEDS


Federal
Local
State
ED
NCES
CCD
NYC DOE
School Survey
ATS
DBN
NYS ED
BEDS


Federal
Local
State
ED
OCR
NCES
CRDC
CCD
NYC DOE
School Survey
ATS
DBN
NYS ED
BEDS

Bringing federal and local together: what we did

Local (DBN) → State (BEDS) → Federal (NCESSCH) → Federal (COMBOKEY)

one school, many "unique" school IDs:

dbn	beds	ncessch	combokey	school_name
24Q290	342400010290	360009806223	362058006223	A.C.E. Academy for Scholars at the Geraldine F...

dnainfo | NEW YORK

How to Navigate New York's Murky System for Tracking Bullying in Schools

By Amy Zimmer and Nigel Chiwaya | September 9, 2015 7:42am

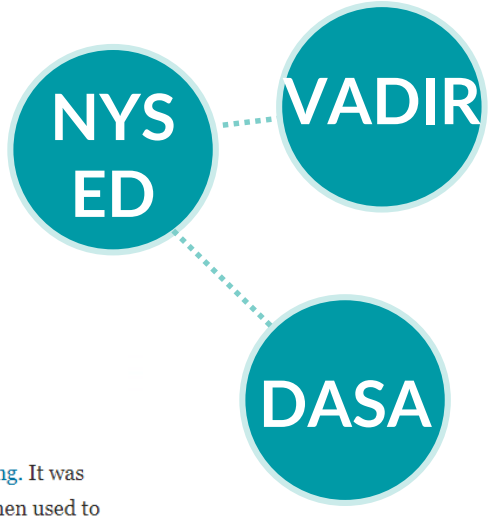
► How city schools are (or aren't) tracking bullying

Two separate state Education Department databases track bullying.

The first is called VADIR — or Violent and Disruptive Incident Reporting. It was created in 2000 and tracks bullying using 20 categories. A formula is then used to crunch the data from the various categories to determine whether a school should be on a "persistently dangerous" list.

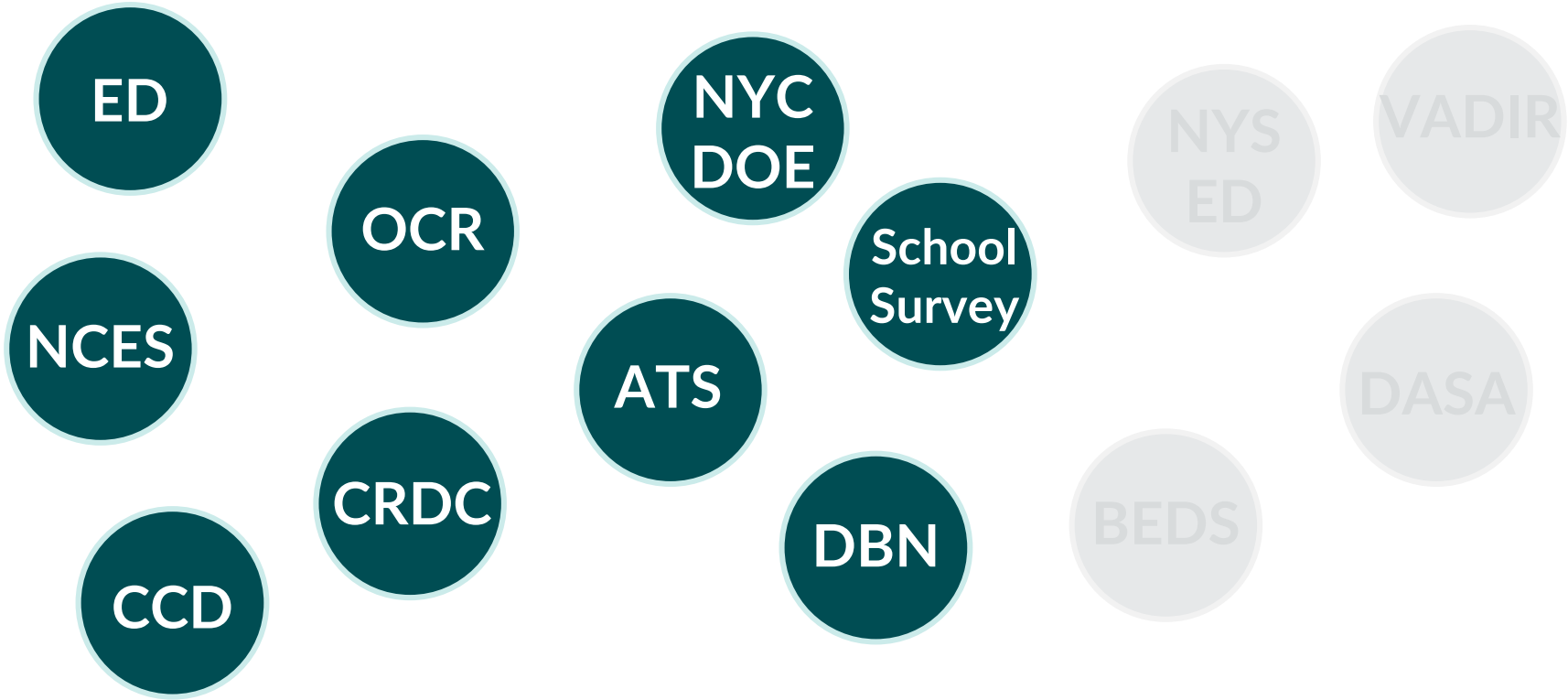
The second, created by the Dignity for all Students Act, was implemented by the state in 2012 to address cyberbullying and bias-related bullying (targeting race, religion, gender or weight, for example) as part of a bigger effort to help make schools safe environments free from discrimination, harassment and bullying.

State



Open data on school bullying and harassment in NYC

Federal + Local



Goals of Analysis

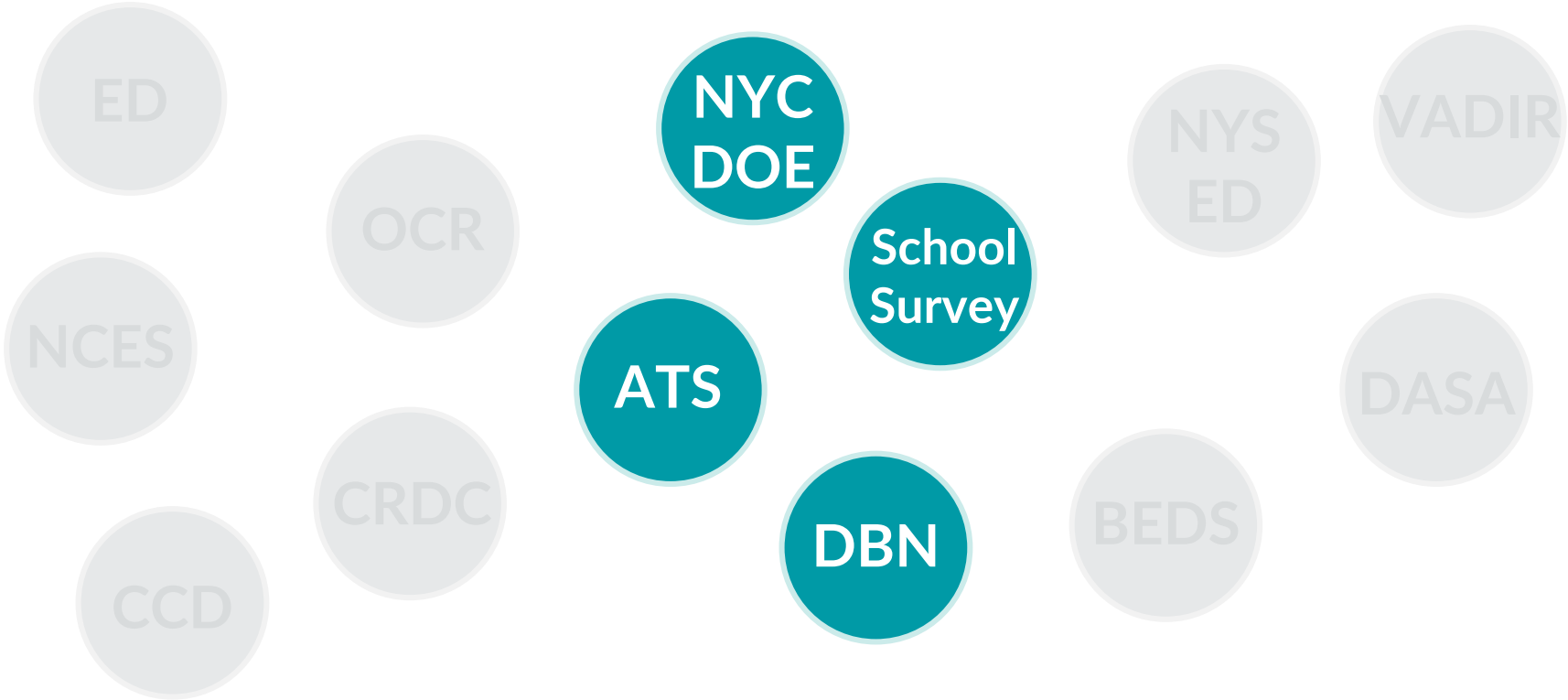
- Local **NYC School Survey** – differences in perceptions of bullying and harassment between students, parents, and teachers.

- Federal **Civil Rights Data** – prevalence of bullying and harassment allegations.

- Local-Federal **survey agreement** – investigating school characteristics associated with (dis)agreement.

Open data on school bullying and harassment in NYC

Local



NYC School Survey
2013-14 school year

bullying/harassment

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences ...

Parents

At my child's school students harass or bully other students.

At my child's school students harass or bully each other based on differences ...

Teachers

At my school, students are often harassed or bullied in school.

At my school, there are conflicts based on differences ...

NYC School Survey
2013-14 school year

bullying/harassment
based on differences

Students

At my school students harass or bully other students.

At my school students harass or bully each other based on differences ...

Parents

At my child's school students harass or bully other students.

At my child's school students harass or bully each other based on differences ...

Teachers

At my school, students are often harassed or bullied in school.

At my school, there are conflicts based on differences ...

NYC School Survey
2013-14 school year

Students

bullying/harassment

bullying/harassment
based on differences

Parents

bullying/harassment

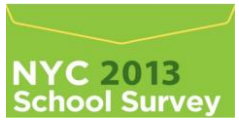
bullying/harassment
based on differences

Teachers

bullying/harassment

bullying/harassment
based on differences

NYC School Survey
2013-14 school year



Students

bullying/harassment

bullying/harassment
based on differences

None of the time ⟷ All of the time

Parents

bullying/harassment

bullying/harassment
based on differences

Strongly Disagree ⟷ Strongly Agree

+ Don't Know

Teachers

bullying/harassment

bullying/harassment
based on differences

Strongly Disagree ⟷ Strongly Agree

NYC School Survey
2013-14 school year

Students

bullying/harassment

bullying/harassment
based on differences

None of the time → All of the time

1 → 4

Parents

bullying/harassment

bullying/harassment
based on differences

Strongly Disagree → Strongly Agree + Don't Know

1 → 4 + X

Teachers

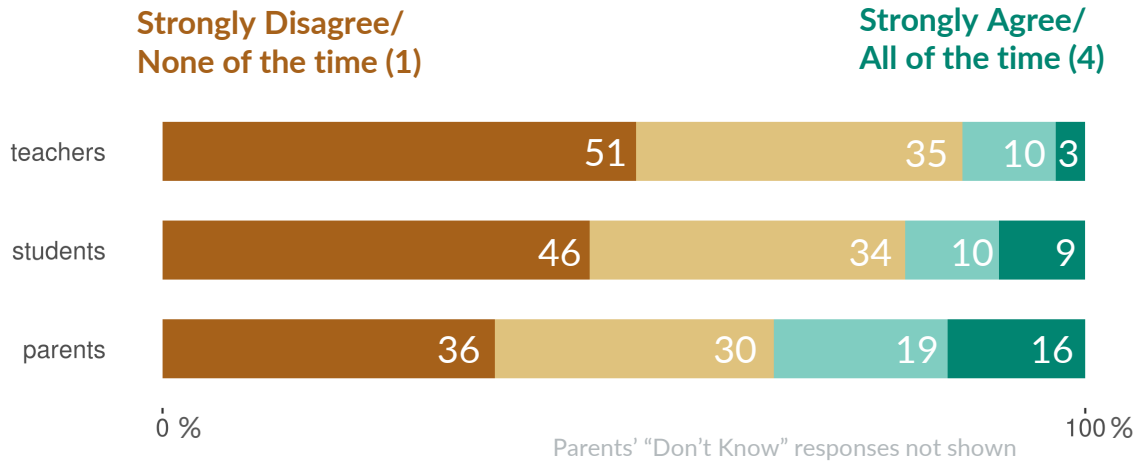
bullying/harassment

bullying/harassment
based on differences

Strongly Disagree → Strongly Agree

1 → 4

Students bully or harass each other / there are conflicts based on differences



- The vast majority of teachers say they disagree (35%) **or** strongly disagree (51%) there are *conflicts* based on differences.
- Parents are the most likely (35%) to feel that students harass/bully each other based on differences.
- Students' feelings about bullying/harassment are "in the middle," between parents' and teachers'.

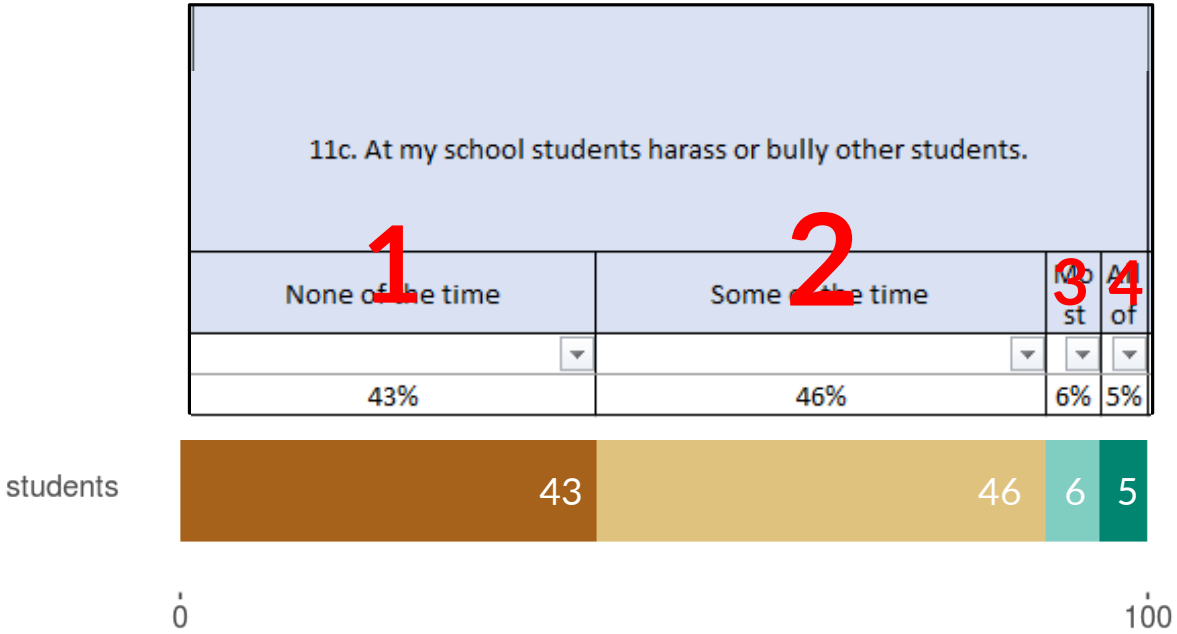
NYC School Survey

Creating a common scale for comparing schools *to each other*

11c. At my school students harass or bully other students.			
1	2	3	4
None of the time	Some of the time	Most of the time	All of the time
43%	46%	6%	5%

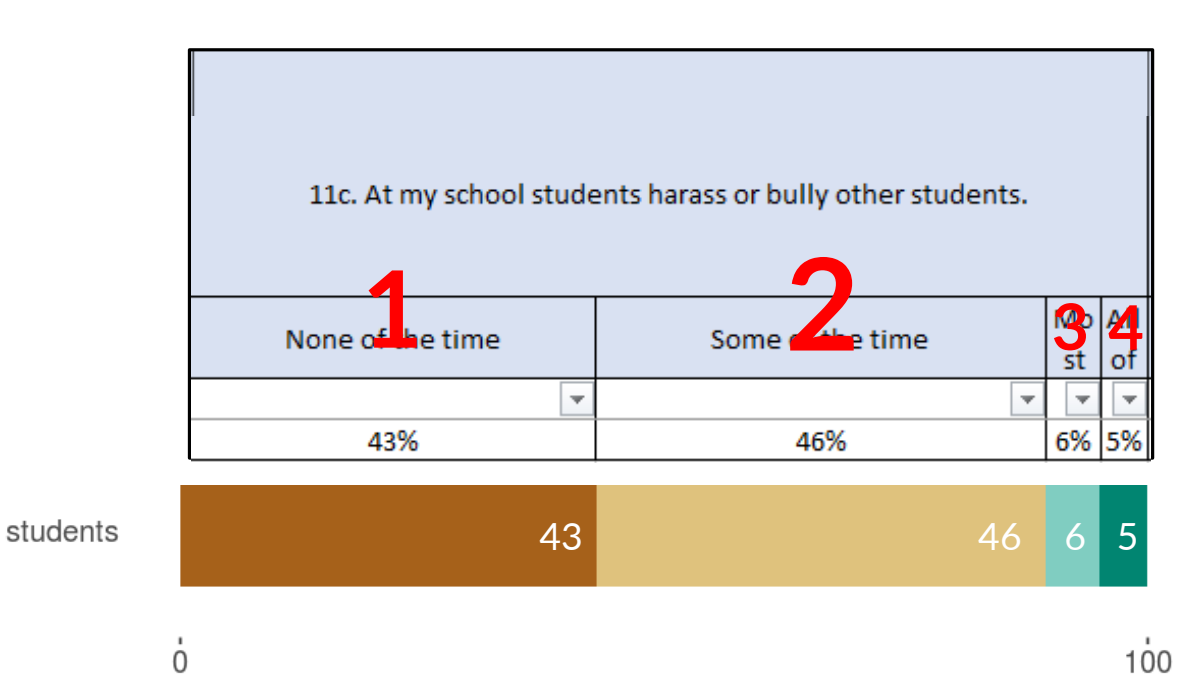
NYC School Survey

Creating a common scale for comparing schools *to each other*



NYC School Survey

Creating a common scale for comparing schools *to each other*



Average school "score" =
1.73

NYC School Survey

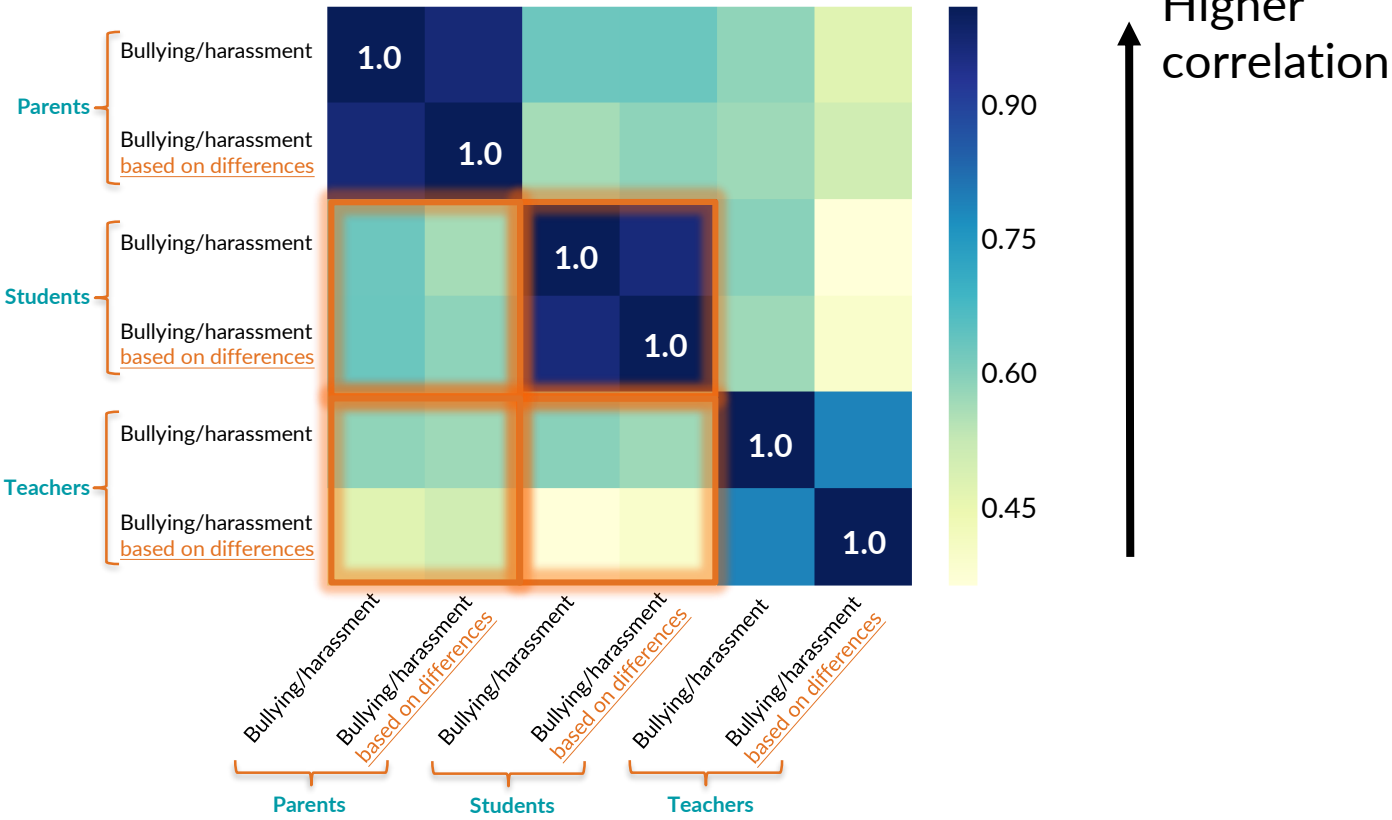
Creating a common scale for comparing schools *to the citywide average*

11c. At my school students harass or bully other students.			
None of the time (1)	Some of the time (2)	Most of the time (3)	All of the time (4)
43%	46%	6%	5%

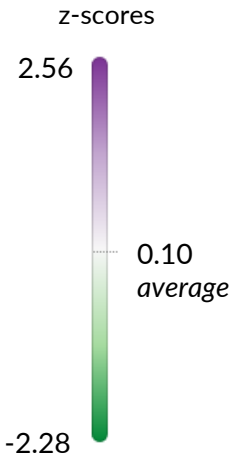
average school "score"	average NYC "score"	school "z-score"
1.73	1.96	-0.72

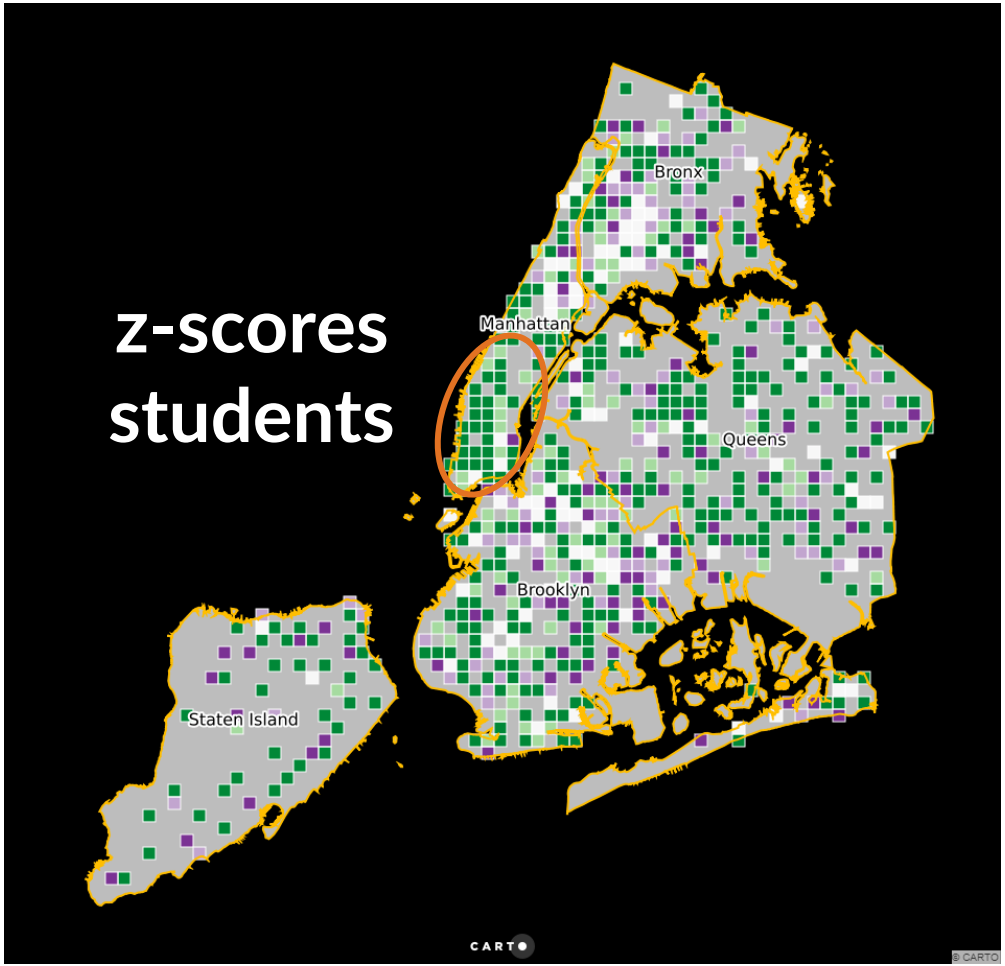
- Given the average school "score" for each school…
- Calculate the citywide average "score"
- Compare each school's "score" to the citywide average
- We do this for both bullying/harassment questions
  - Parents
  - Students
  - Teachers

Responses are highly correlated

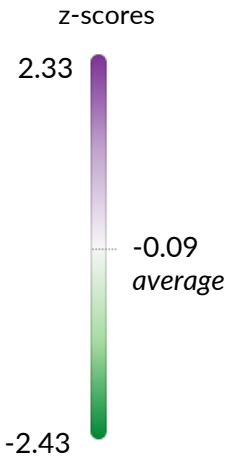


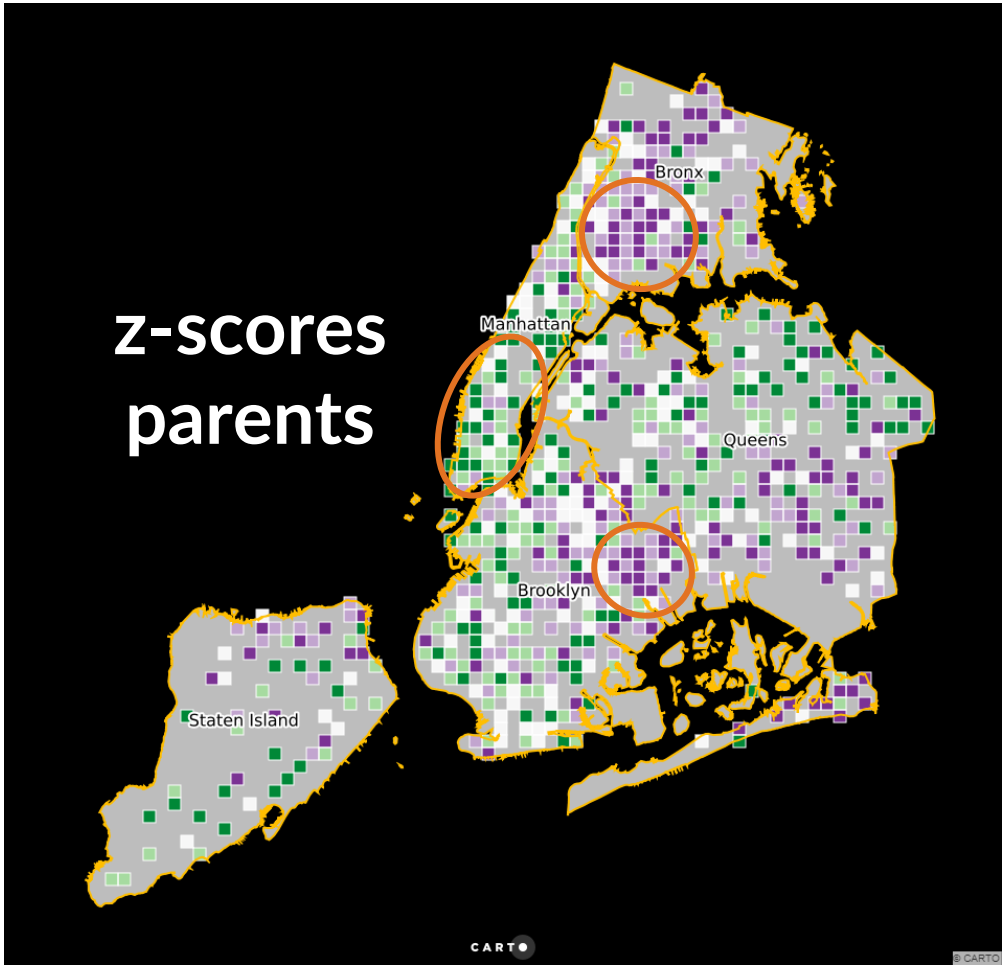
Bullying/harassment based on differences: student responses



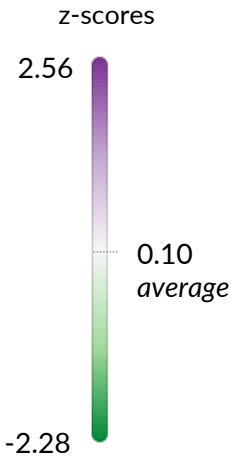


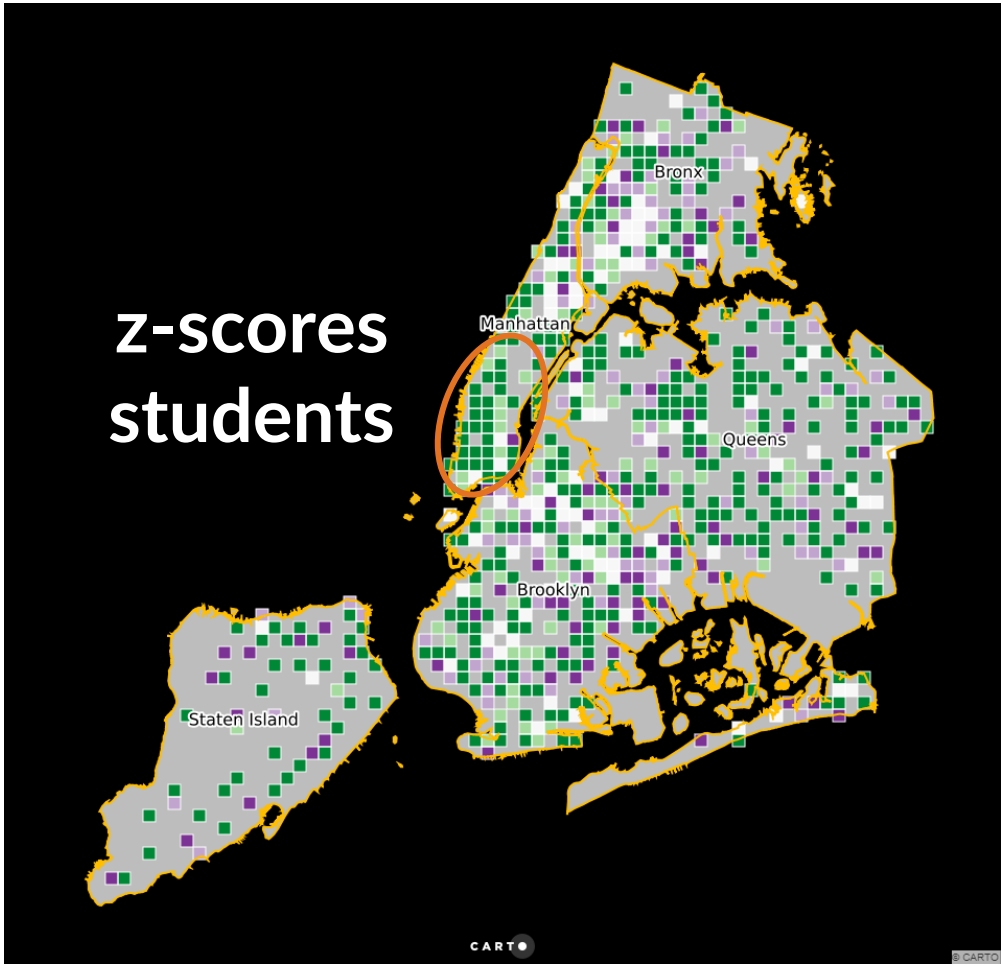
Bullying/harassment based on differences: parent responses



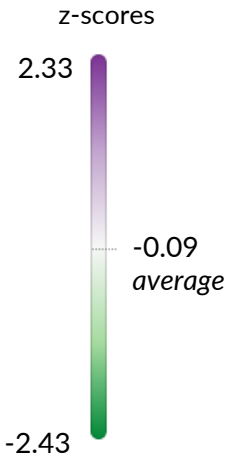


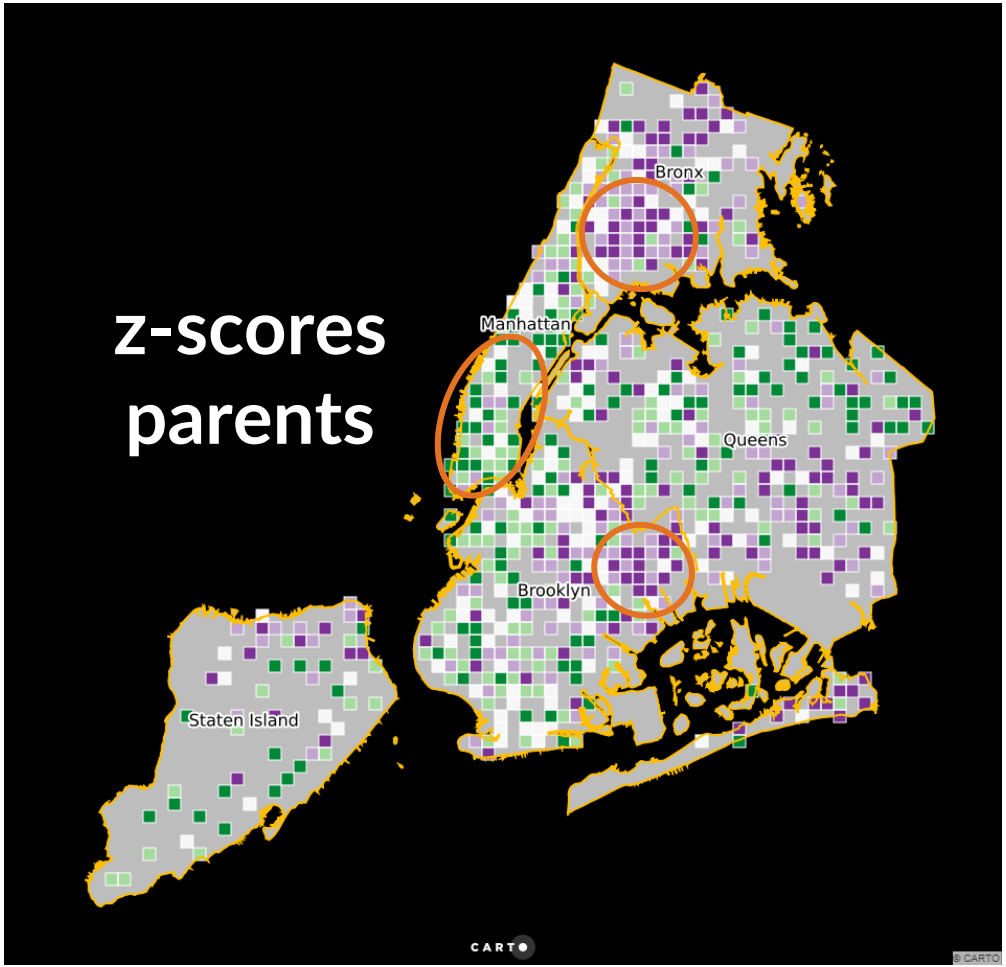
Bullying/harassment based on differences: student responses



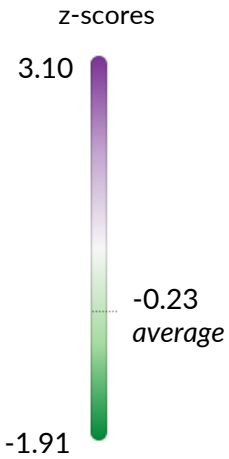


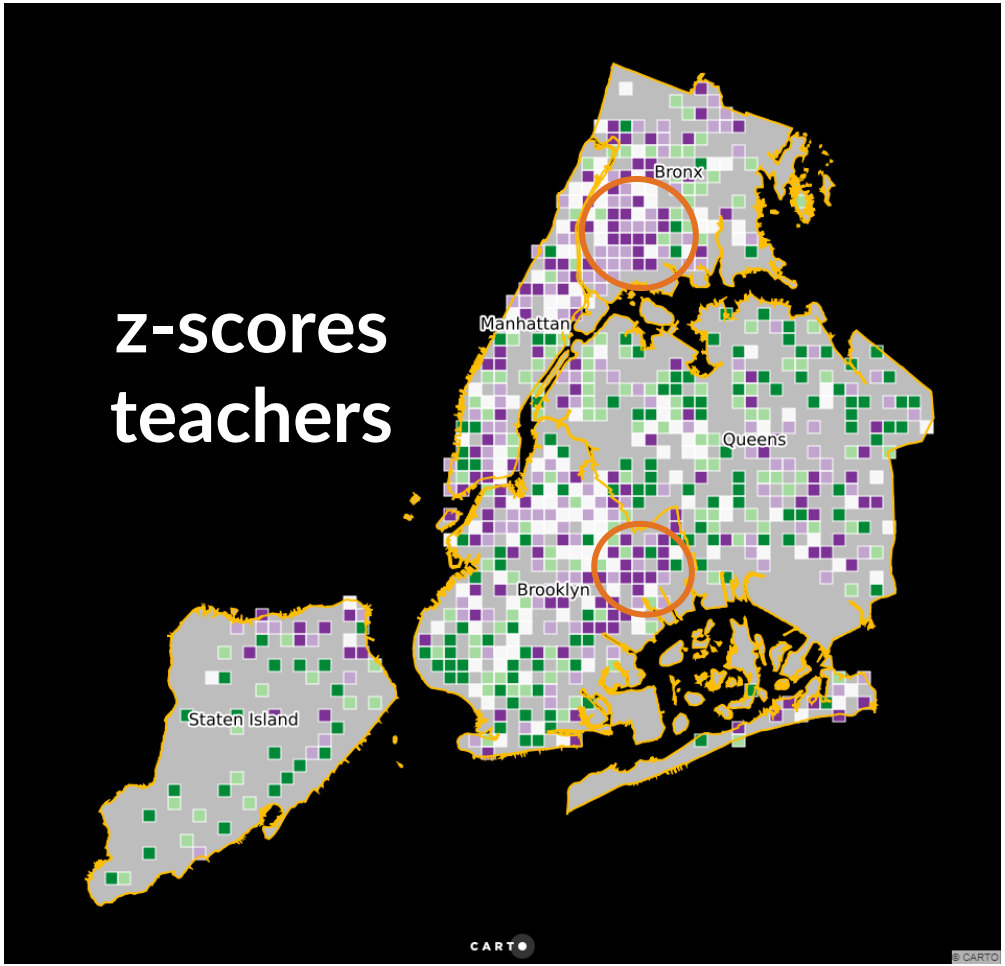
Bullying/harassment based on differences: parent responses



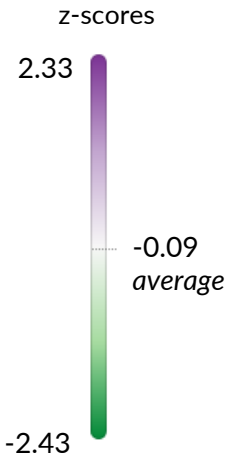


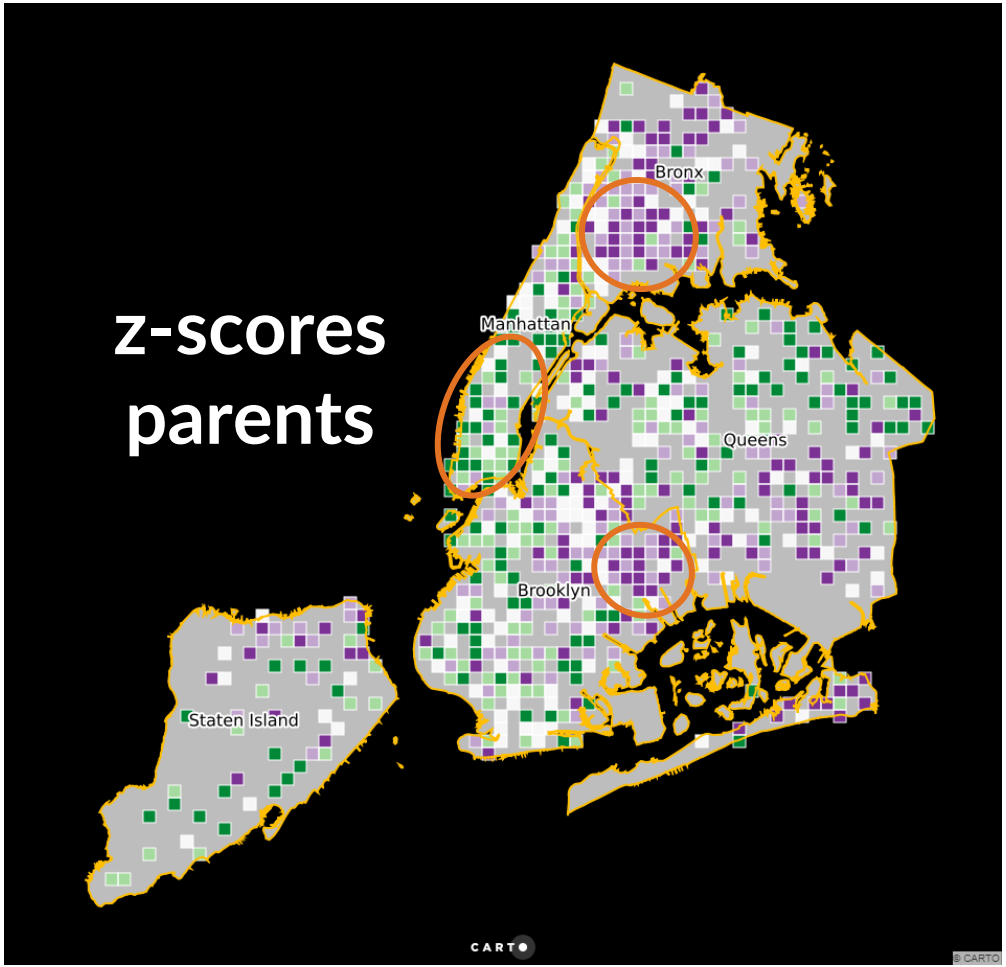
Bullying/harassment based on differences: teacher responses



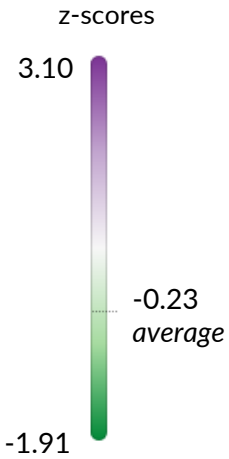


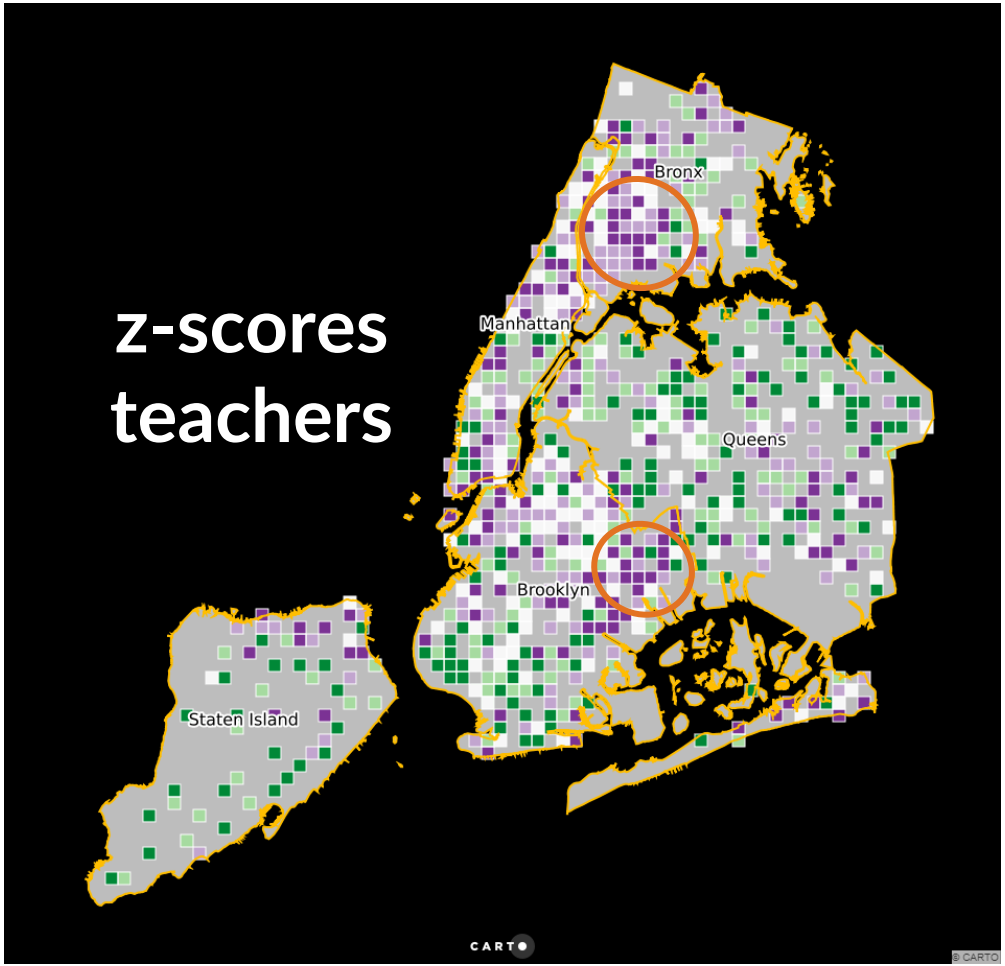
Bullying/harassment based on differences: parent responses





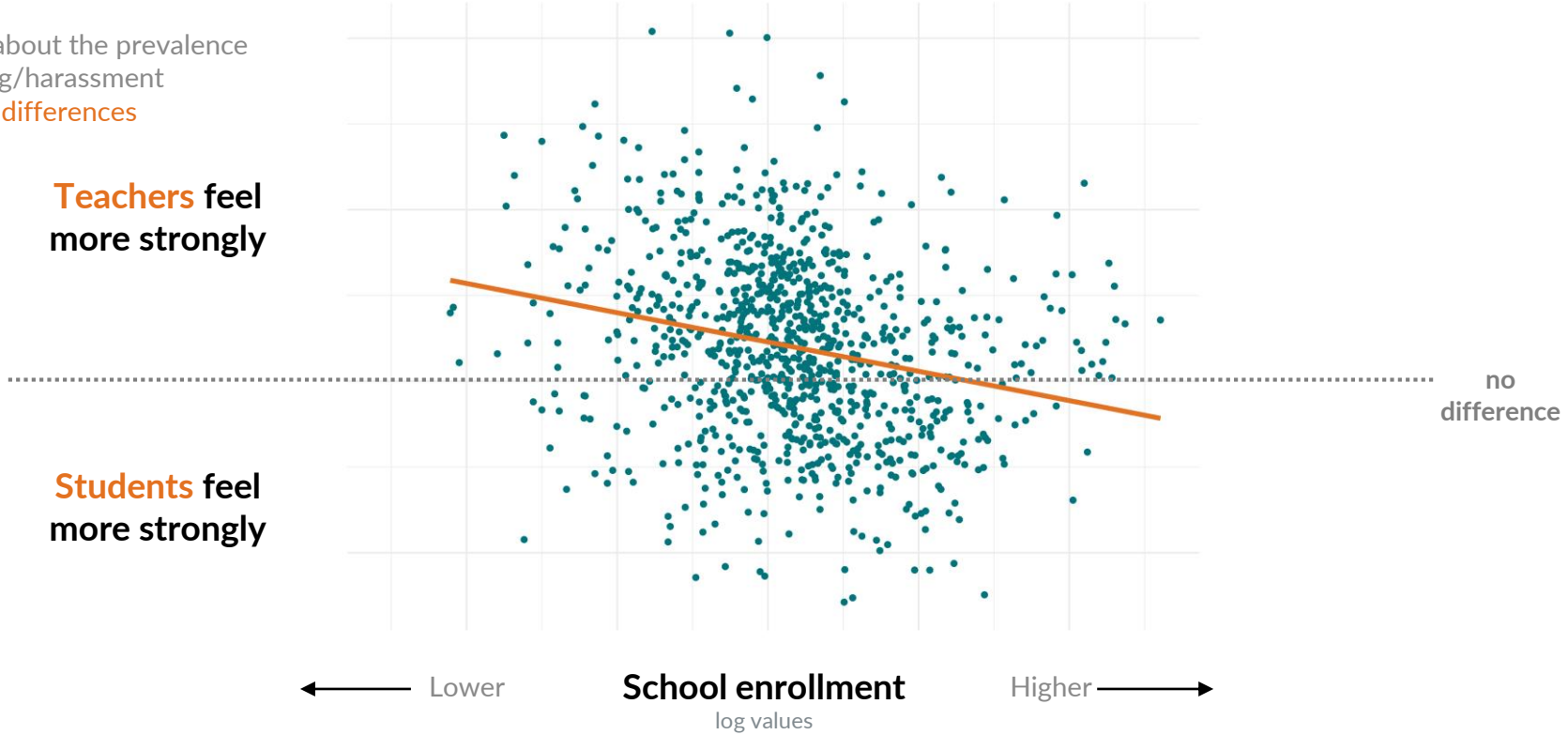
Bullying/harassment based on differences: teacher responses





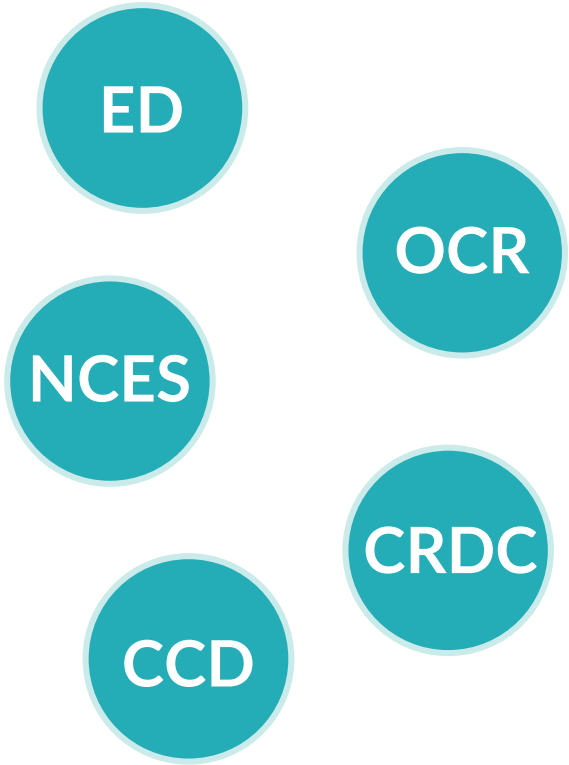
Teachers and students appear to disagree
in the largest and smallest schools (though this correlation is small)

Feelings about the prevalence of bullying/harassment based on differences



Open data on school bullying and harassment in NYC

Federal



Open data on school bullying and harassment in NYC

Federal

Civil Rights Data Collection (CRDC)

2013-14 school year

Allegations of Harassment or Bullying

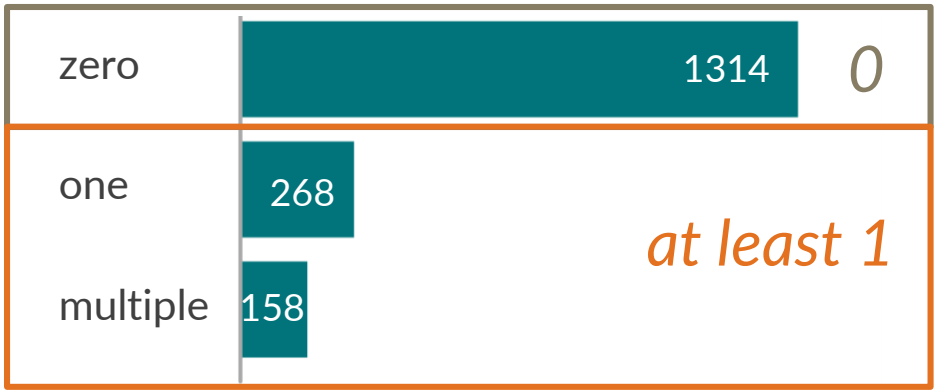
Students Reported as Harassed or Bullied

Students Disciplined for Harassment or Bullying

Race-, Sex-, Disability- based

Civil Rights Data Collection
Office for Civil Rights
2013-14 school year

# of schools reporting _____
allegations of bullying/harassment



- Sex-based bullying/harassment is the most commonly alleged category, disability-based is the least
- 75% of NYC schools report zero allegations (of any type) of bullying/harassment to the federal OCR. Nationally, this number is 80%.

Characteristics of schools with …

	0 allegations	1+ allegations
Total Enrollment	576	713
Students with disabilities (%)	18	19
English Language Learners (%)	14	12
Parent responses z-scores*	-0.12	0.37
Student responses z-scores*	-0.15	0.27
Teacher responses z-scores*	-0.15	0.45

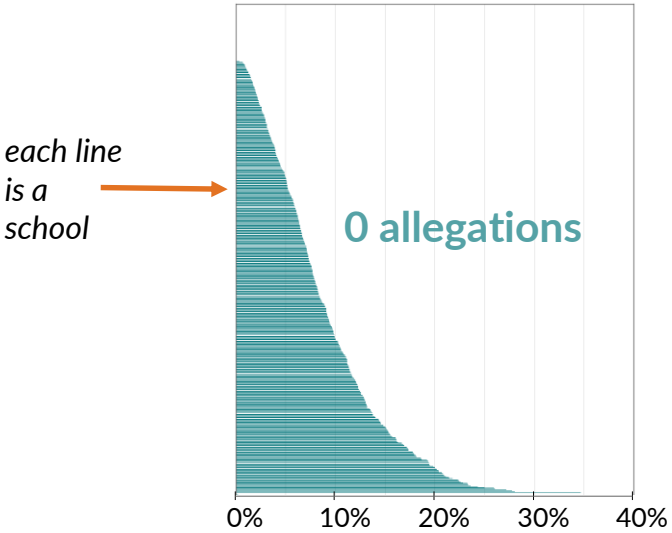
1314 schools (75%) reporting 0 allegations

426 schools reporting 1+ allegations

Additional characteristics not shown as there were no significant differences. *z-scores are shown for the bullying/harassment due to differences question (results for responses the general question were similar)

How much do the two surveys (dis)agree?

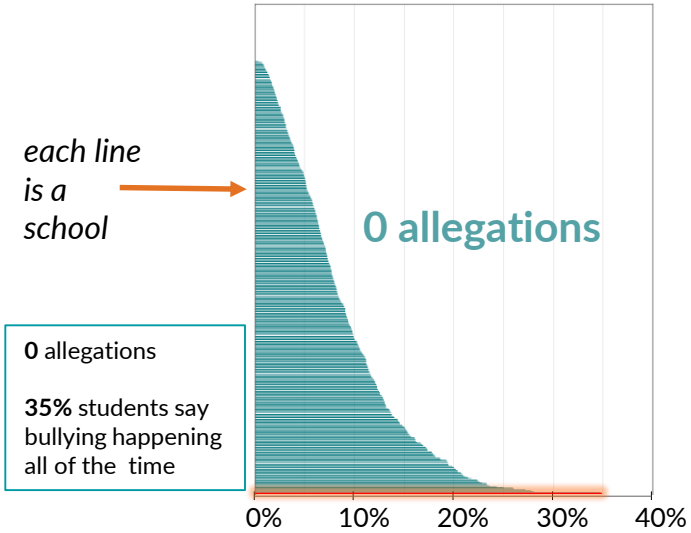
How much do the two surveys (dis)agree?



% of students who say bullying/ harassment based on differences happens all of the time in …

- Schools that reported **zero** allegations

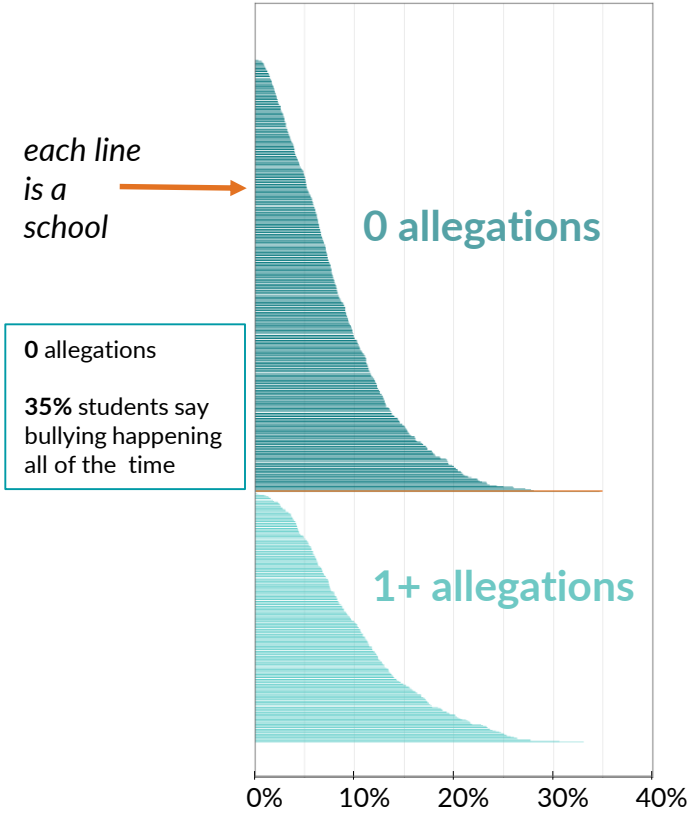
How much do the two surveys (dis)agree?



% of students who say bullying/ harassment based on differences happens all of the time in …

- Schools that reported zero allegations

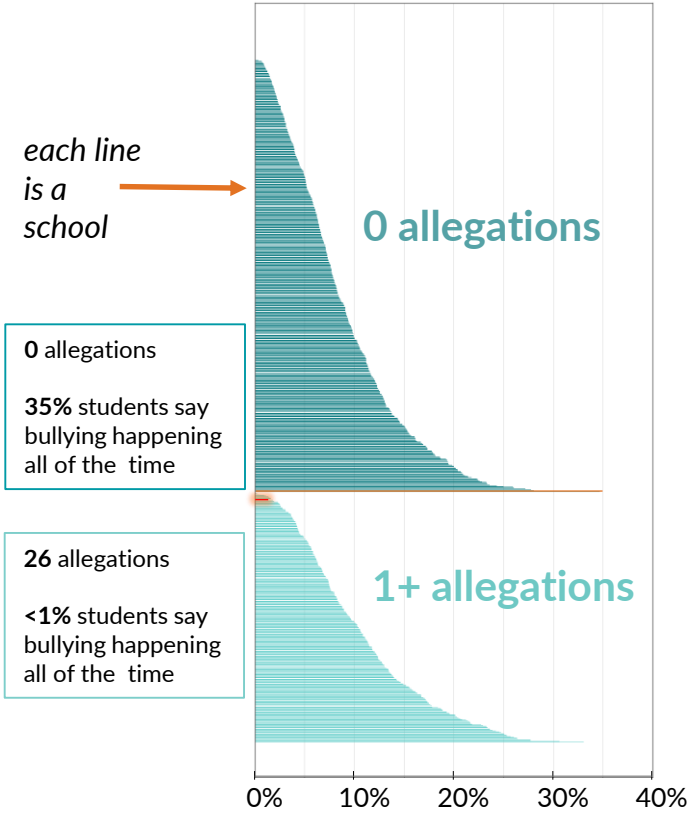
How much do the two surveys (dis)agree?



% of students who say bullying/ harassment based on differences happens all of the time in …

- Schools that reported **zero** allegations
- Schools that reported at least 1 allegation

How much do the two surveys (dis)agree?



0 allegations

35% students say bullying happening all of the time

26 allegations

<1% students say bullying happening all of the time

% of students who say bullying/ harassment based on differences happens all of the time in …

- Schools that reported **zero** allegations
- Schools that reported at least 1 allegation

Zones of (dis)agreement

zero allegations ⟷ **at least one allegation**

Zones of (dis)agreement

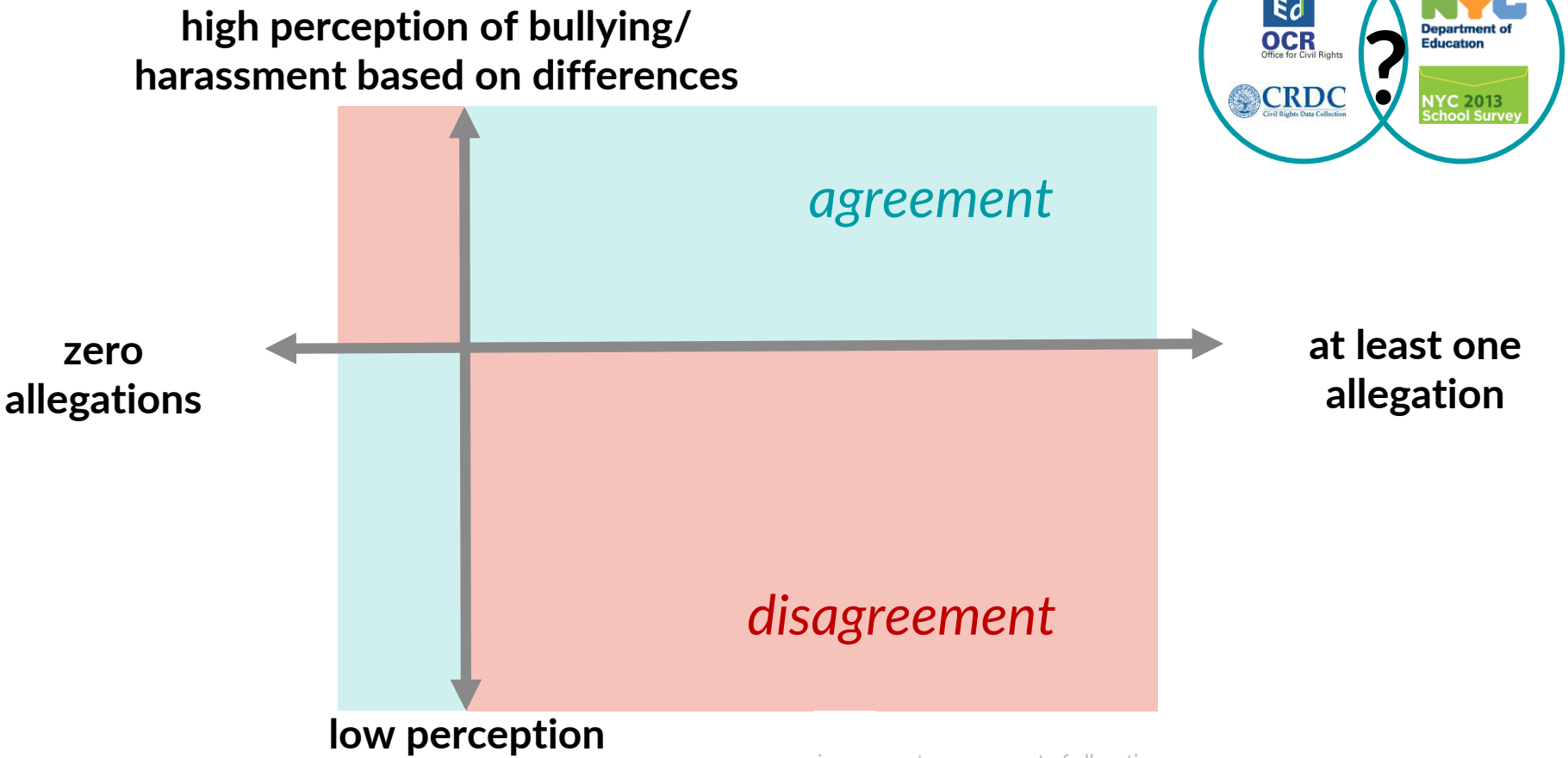
high perception of bullying/ harassment based on differences

zero allegations

at least one allegation

low perception

Zones of (dis)agreement

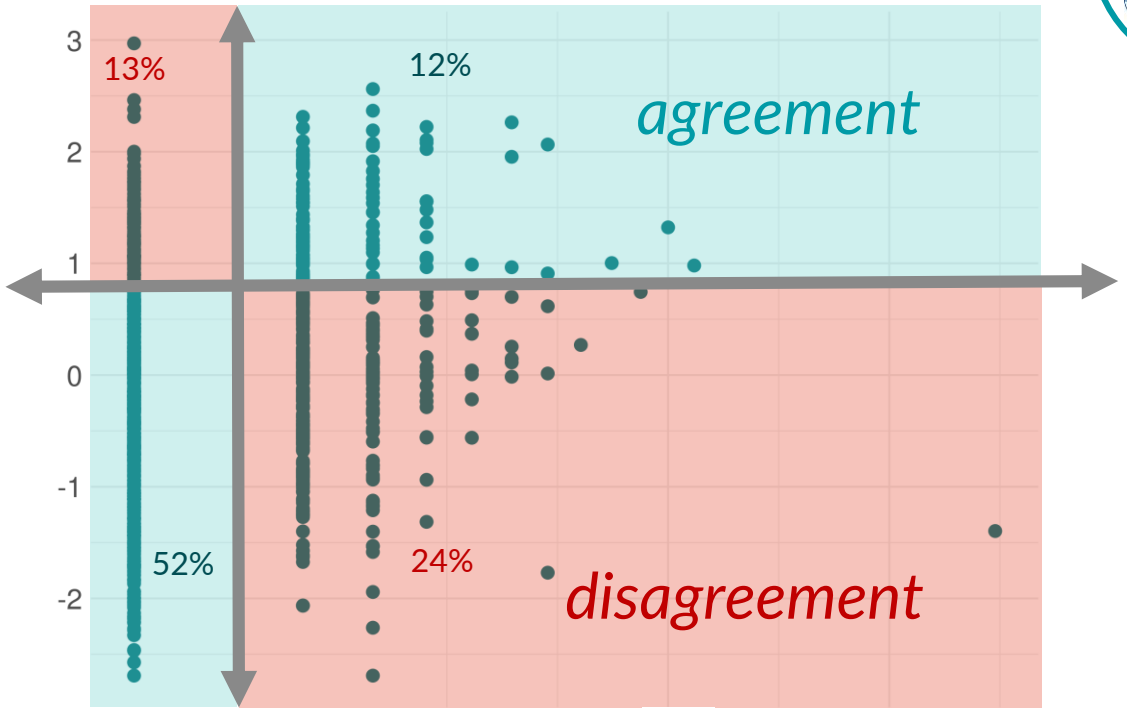


Zones of (dis)agreement

high perception of bullying/ harassment based on differences

zero allegations

at least one allegation



low perception

x-axis represents square-root of allegations

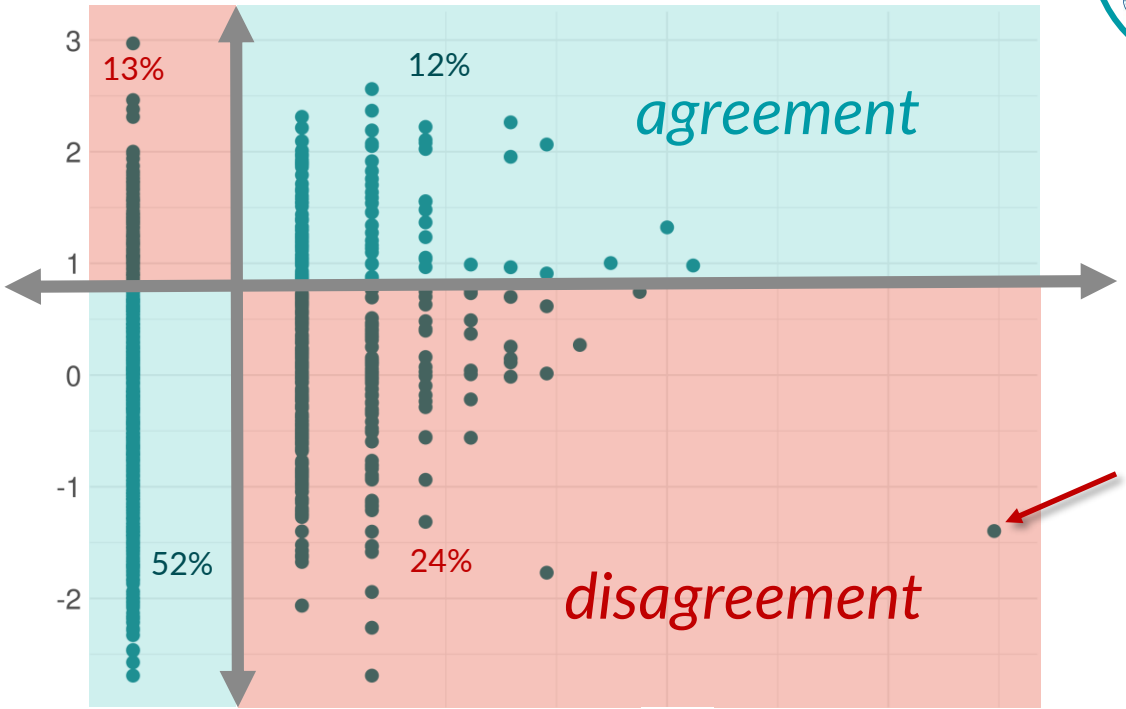
Zones of (dis)agreement

high perception of bullying/ harassment based on differences

zero allegations

at least one allegation

low perception



26 allegations

<1% students say bullying happening all of the time

x-axis represents square-root of allegations

Which factors are associated with survey *dis*agreement?

Parents

- High Schools *vs. Elementary Schools*
- High Schools *vs. Middle Schools*
- High Schools *vs. Mixed-Grade Schools*
- General Education Schools *vs. Career/Technical/etc. Schools*
- Regular Schools *vs. Charter Schools*
- Total Enrollment
- Female (%)
- Black (%)
- Hispanic (%)
- Students with Disabilities (%)
- English Language Learners (%)
- Free Lunch (%)

Students

- High Schools *vs. Elementary Schools*
- High Schools *vs. Middle Schools*
- High Schools *vs. Mixed-Grade Schools*
- General Education Schools *vs. Career/Technical/etc. Schools*
- Regular Schools *vs. Charter Schools*
- Total Enrollment
- Female (%)
- Black (%)
- Hispanic (%)
- Students with Disabilities (%)
- English Language Learners (%)
- Free Lunch (%)

Teachers

- High Schools *vs. Elementary Schools*
- High Schools *vs. Middle Schools*
- High Schools *vs. Mixed-Grade Schools*
- General Education Schools *vs. Career/Technical/etc. Schools*
- Regular Schools *vs. Charter Schools*
- Total Enrollment
- Female (%)
- Black (%)
- Hispanic (%)
- Students with Disabilities (%)
- English Language Learners (%)
- Free Lunch (%)

Which factors are associated with survey *dis*agreement?

Parents	Students	Teachers
High Schools *vs. Elementary Schools*	High Schools *vs. Elementary Schools*	**High Schools *vs. Elementary Schools***
High Schools *vs. Middle Schools*	High Schools *vs. Middle Schools*	**High Schools *vs. Middle Schools***
High Schools *vs. Mixed-Grade Schools*	High Schools *vs. Mixed-Grade Schools*	**High Schools *vs. Mixed-Grade Schools***
General Education Schools *vs. Career/Technical/etc. Schools*	General Education Schools *vs. Career/Technical/etc. Schools*	**General Education Schools *vs. Career/Technical/etc. Schools***
Regular Schools *vs. Charter Schools*	Regular Schools *vs. Charter Schools*	Regular Schools *vs. Charter Schools*
Total Enrollment	**Total Enrollment**	**Total Enrollment**
Female (%)	Female (%)	Female (%)
Black (%)	Black (%)	Black (%)
Hispanic (%)	Hispanic (%)	Hispanic (%)
Students with Disabilities (%)	Students with Disabilities (%)	**Students with Disabilities (%)**
English Language Learners (%)	English Language Learners (%)	English Language Learners (%)
Free Lunch (%)	Free Lunch (%)	Free Lunch (%)

positive association with OCR-NYC School Survey disagreement

not significantly associated with OCR-NYC School Survey (dis)agreement

Which factors are associated with survey *dis*agreement?

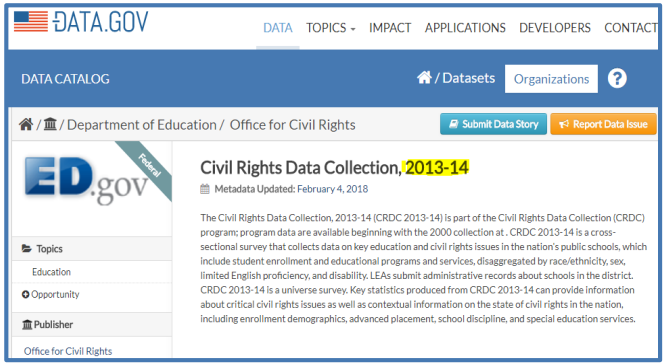
Parents	Students	Teachers
High Schools *vs. Elementary Schools*	High Schools *vs. Elementary Schools*	**High Schools *vs. Elementary Schools***
High Schools *vs. Middle Schools*	High Schools *vs. Middle Schools*	**High Schools *vs. Middle Schools***
High Schools *vs. Mixed-Grade Schools*	High Schools *vs. Mixed-Grade Schools*	**High Schools *vs. Mixed-Grade Schools***
General Education Schools *vs. Career/Technical/etc. Schools*	General Education Schools *vs. Career/Technical/etc. Schools*	**General Education Schools *vs. Career/Technical/etc. Schools***
Regular Schools *vs. Charter Schools*	Regular Schools *vs. Charter Schools*	Regular Schools *vs. Charter Schools*
Total Enrollment	**Total Enrollment**	**Total Enrollment**
Female (%)	Female (%)	Female (%)
Black (%)	Black (%)	Black (%)
Hispanic (%)	Hispanic (%)	Hispanic (%)
Students with Disabilities (%)	Students with Disabilities (%)	**Students with Disabilities (%)**
English Language Learners (%)	English Language Learners (%)	English Language Learners (%)
Free Lunch (%)	Free Lunch (%)	Free Lunch (%)

positive association with OCR-NYC School Survey disagreement

not significantly associated with OCR-NYC School Survey (dis)agreement

Caveats and considerations

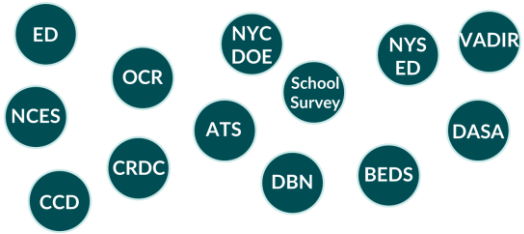
- Our data is not perfect
  - 2013-14 school year, the latest year available on the Civil Rights Data Collection (Office for Civil Rights)
  - Student sample is smaller than teachers and parents
  - Between-survey (local vs. federal) & within-survey (parent vs. teacher vs. student) comparisons required assumptions
- We only looked at federal and citywide open data, not state
- We are not education experts



What we learned about bullying and harassment in NYC schools

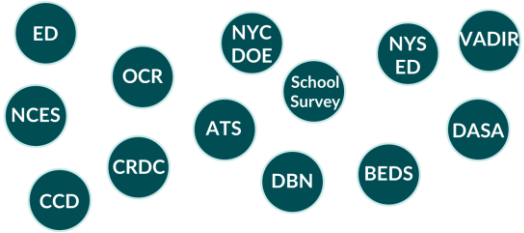
- Teachers' responses are less correlated with students' and parents'. This varies by school enrollment.
- In aggregate, perceptions of school bullying/harassment are related to federal reporting of bullying/harassment.
- But perceptions can vary dramatically, even in schools that report zero allegations—which is the vast majority of schools.
- Bigger schools tend to have more disagreement between local and federal school surveys.

What we learned about NYC open data

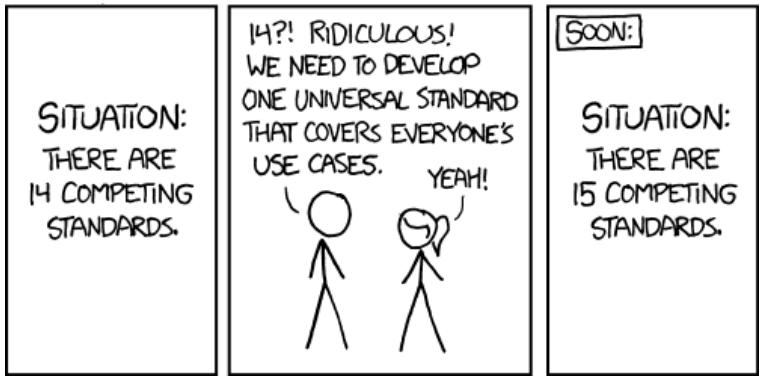


- **Having a "unique" school ID might not be enough …**
  - DBN = "District Borough Number" [Local]
  - ATS = "Automate the Schools" [Local]
  - BEDS = "Basic Education Data System" [State]
  - NCESSCH = "National Center of Education Statistics School ID" [Federal]
  - COMBOKEY = Office for Civil Rights [Federal]

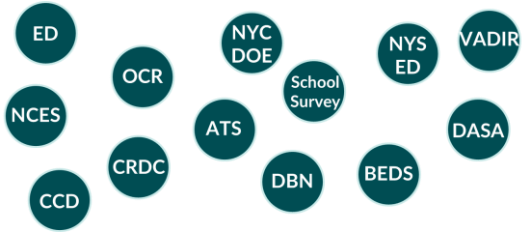
What we learned about NYC open data



- **Having a "unique" school ID might not be enough …**
  - DBN = "District Borough Number" [Local]
  - ATS = "Automate the Schools" [Local]
  - BEDS = "Basic Education Data System" [State]
  - NCESSCH = "National Center of Education Statistics School ID" [Federal]
  - COMBOKEY = Office for Civil Rights [Federal]

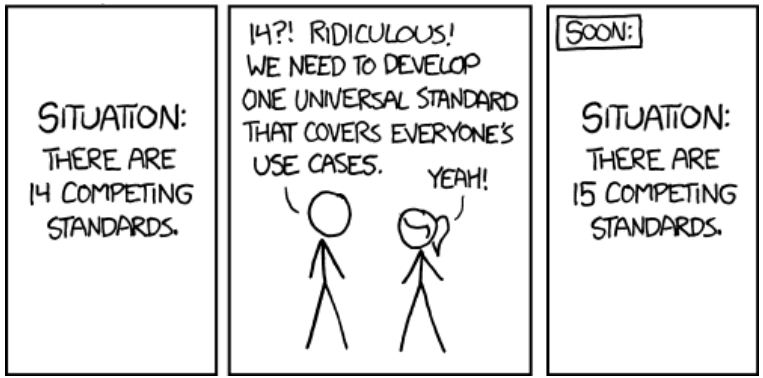


What we learned about NYC open data

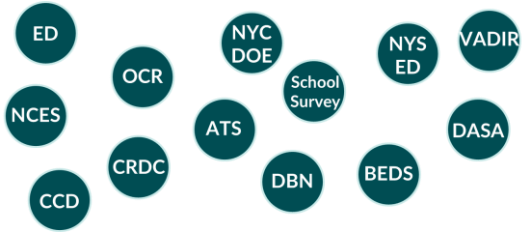


- **Having a "unique" school ID might not be enough …**
  - DBN = "District Borough Number" [Local]
  - ATS = "Automate the Schools" [Local]
  - BEDS = "Basic Education Data System" [State]
  - NCESSCH = "National Center of Education Statistics School ID" [Federal]
  - COMBOKEY = Office for Civil Rights [Federal]

Crosswalks (files that link all the IDs) can help!

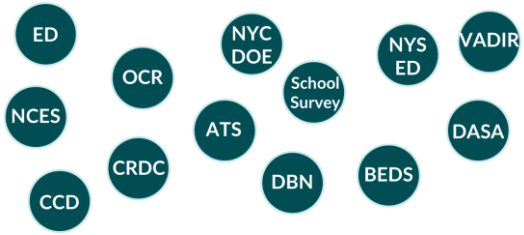


What we learned about NYC open data



- Having a "unique" school ID might not be enough …
- Being publicly available != being "open"
  - Necessary, but not sufficient

What we learned about NYC open data



- Having a "unique" school ID might not be enough …
- Being publicly available != being "open"
  - Necessary, but not sufficient

Secure | https://sunlightfoundation.com/policy/documents/ten-open-data-principles/

SUNLIGHT FOUNDATION

7. Commonly owned or open Standards

Commonly owned or open standards refer to who owns the format in which data is stored. For example, if only one company manufactures the program that can read a file where data is stored, access to that information is dependent upon use of the company's processing program. Sometimes that program is unavailable to the public at any cost, or is available, but for a fee. For example, Microsoft Excel is a fairly commonly-used spreadsheet program which costs money to use. Freely available alternative formats often exist by which stored data can be accessed without the need for a software license. Removing this cost makes the data available to a wider pool of potential users.

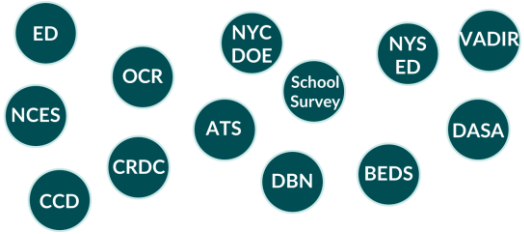
NYC School Survey

2017 School Survey

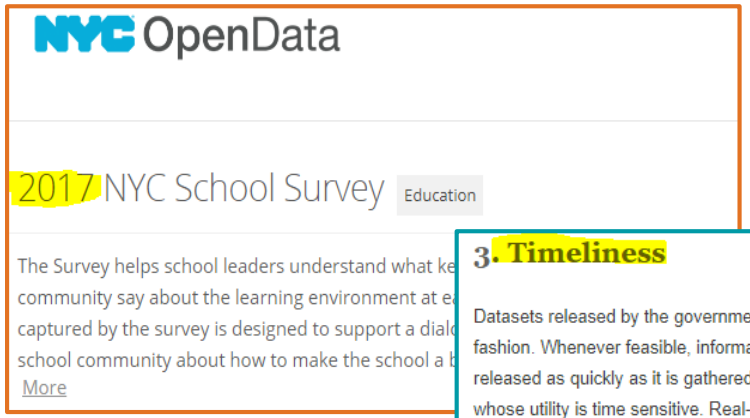
For the second year in a row, survey partici
students, and teachers completing the 201
citywide level and also summarized and rep

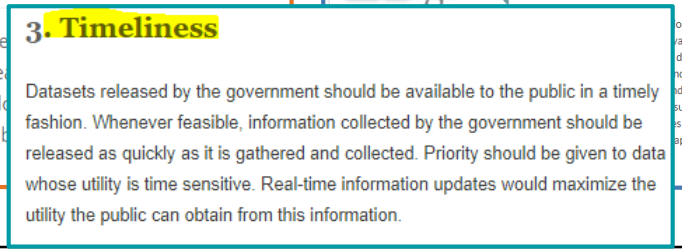
- View the 2017 citywide analysis of survey results. (PDF)
- Download the 2017 parent, teacher, and student survey data for all New York City public schools. (Excel)
- View the 2017 NYC School Survey Data Protocol. (PDF)

What we learned about NYC open data

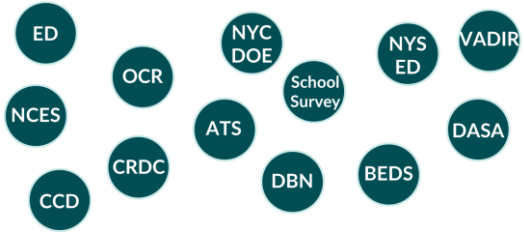


- Having a “unique” school ID might not be enough …
- Being publicly available != being “open”
- Limitations in the availability of open data may hinder timely analyses



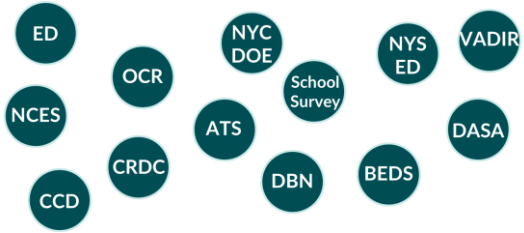


What we learned about NYC open data

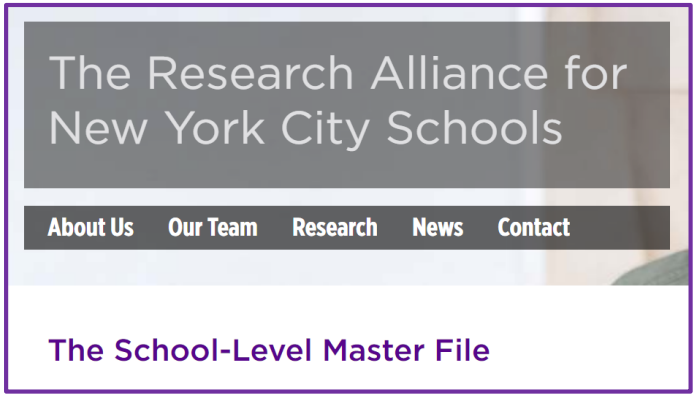


- Having a “unique” school ID might not be enough …
- Being publicly available != being “open”
- Limitations in the availability of open data may hinder timely analyses
- Understanding the original purpose of different data sets can guide research

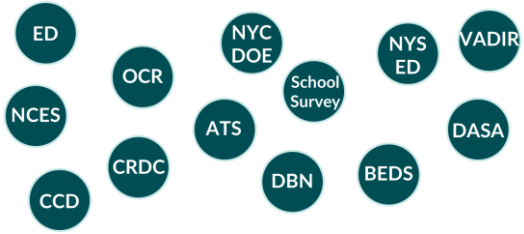
What we learned about NYC open data



- Having a “unique” school ID might not be enough …
- Being publicly available != being “open”
- Limitations in the availability of open data may hinder timely analyses
- Understanding the original purpose of different data sets can guide research
- Someone else probably has (or had) the same data problem as you!



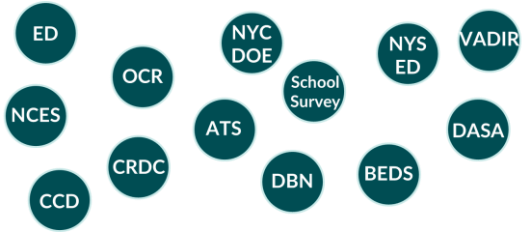
What we learned about NYC open data



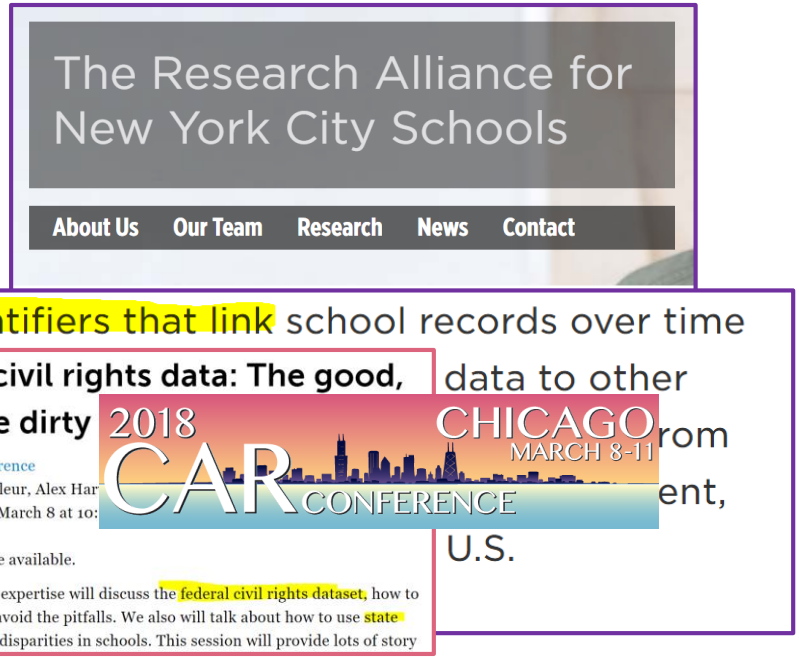
- Having a "unique" school ID might not be enough …
- Being publicly available != being "open"
- Limitations in the availability of open data may hinder timely analyses
- Understanding the original purpose of different data sets can guide research
- Someone else probably has (or had) the same data problem as you!

- Identifiers that link school records over time and that link New York City data to other public sources such as those available from the New York State Education Department, the Census Bureau, and the U.S. Department of Education

What we learned about NYC open data



- Having a "unique" school ID might not be enough …
- Being publicly available != being "open"
- Limitations in the availability of open data may hinder timely analyses
- Understanding the original purpose of different data sets can guide research
- Someone else probably has (or had) the same data problem as you!



John Paul Farmer
@johnpaulfarmer
Following
"#OpenData may be the Central Park of data" — an amazing public resource that we shouldn't take for granted. - @MiguelGamino at #nycSoData on #OpenDataDay
11:16 AM - 3 Mar 2018 from Manhattan, NY

www.twosigma.com/dataclinic

Rachael Weiss Riley & Christine Zhang
Dave Santin
Tiffany Chang
Chris Mulligan

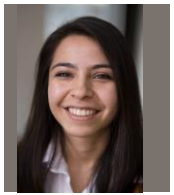
Special thanks to:

Ben Wellington (Two Sigma), Erin Stein (Overdeck Family Foundation)
Matt Chingos & Erica Blom (Urban Institute)
Adrienne Schmoeker (NYC MODA)
Nick Chmura & Two Sigma Q4 2017 Hack Day participants

Julia Bloom-Weltman (julia.bloom-weltman@aemcorp.com)
Director of Research
Applied Engineering Management

Meghan McCormick (meghan.mccormick@mdrc.org)
Research Associate
MDRC

Johanna Miller (@johannawithanh)
Advocacy Director
New York Civil Liberties Union



Jasmine Soltani (jasmine.soltani@nyu.edu)
Data Manager
Research Alliance for NYC Schools

Amy Zimmer (@the_zim | amyz@localize.city)
Journalist, Content Strategist
Localize.city