Open Data for Learning Analytics

Wednesday, March 23 at 3:45 PM

Presented by:

Katherine Rojas <kr2678@tc.columbia.edu>
Maeghan Sill <mms2380@tc.columbia.edu>
Yasemin Gulbahar <yg2918@tc.columbia.edu>

nycsodata24.sched.com #nycSOdata #opendataweek

NYC SCHOOL OF DATA

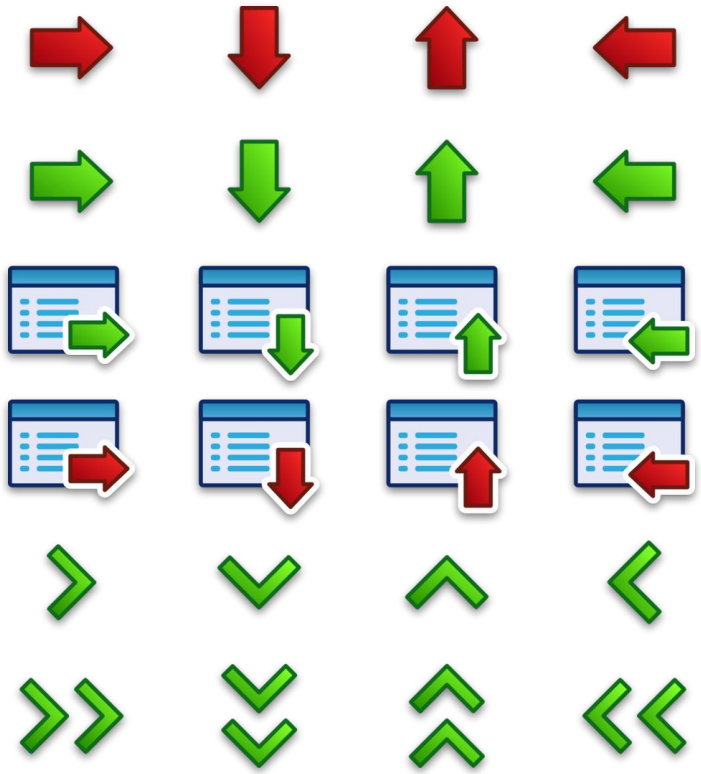
Agenda

- ➔ The Role and Use Cases for Open Data
- ➔ Possible Research Ideas
- ➔ Open Data Analysis
- ➔ Brainstorming Session
- ➔ Wrap-Up

NYC SCHOOL OF DATA

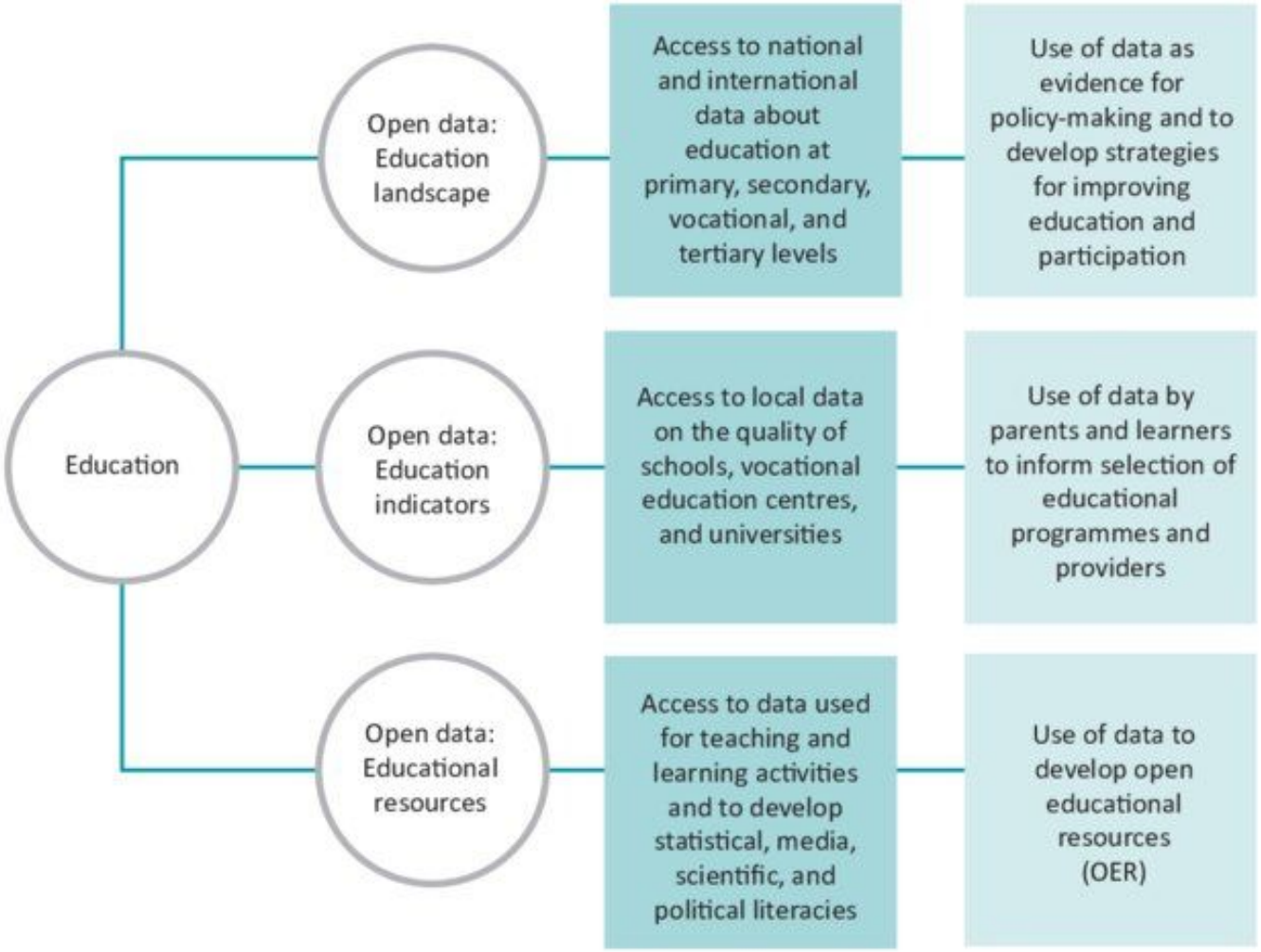
The Role of Open Data for Learning Analytics

1. Accessibility and Transparency
2. Richness and Diversity
3. Fostering Data Literacy Skills
4. Promoting Reproducibility and Innovation
5. Enhancing Institutional Effectiveness



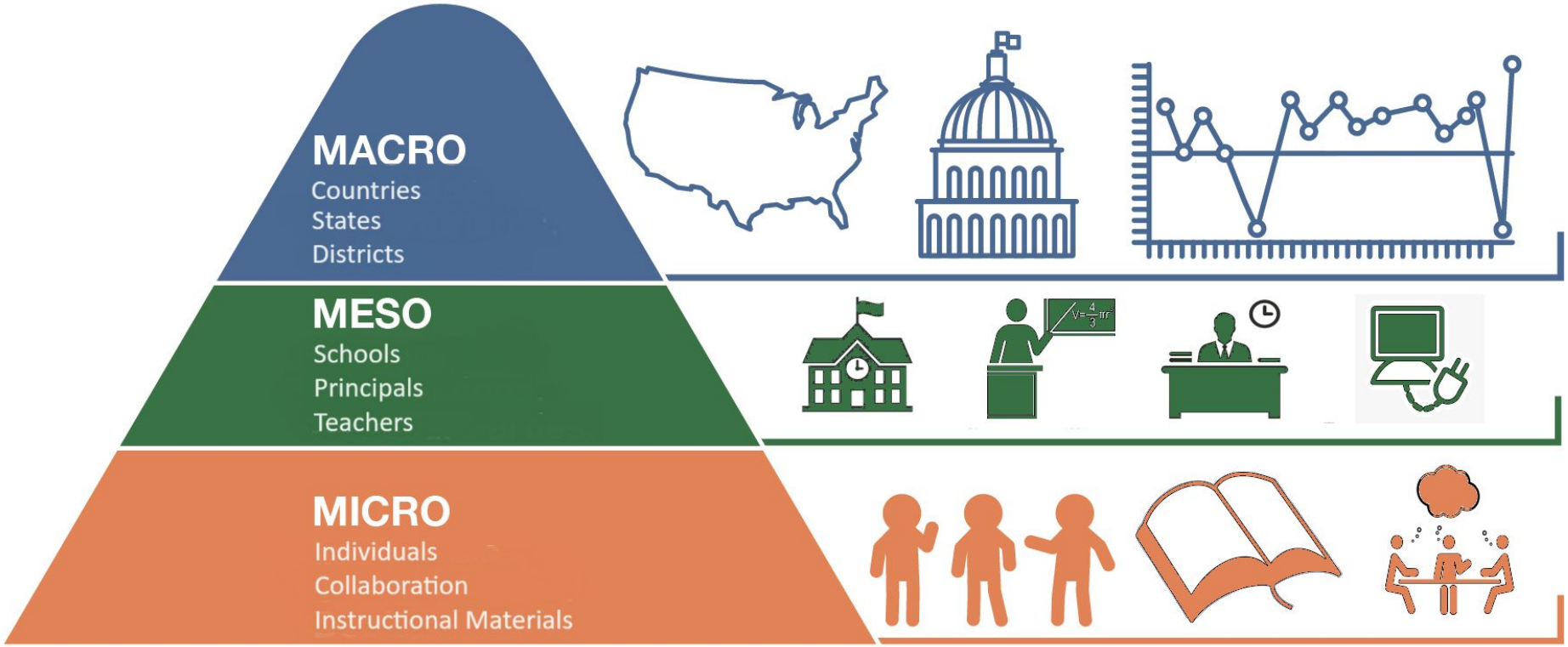
NYC SCHOOL OF DATA

Use Cases of Open Data for Education



NYC SCHOOL OF DATA

Macro, Meso & Micro Levels of Analysis



adapted from https://pressbooks.rampages.us/sociology-research-methods/chapter/2-levels-of-analysis/

Open Data needed for Learning Analytics

1. Administrative Data

Identifying at-risk students based on their past academic performance and attendance patterns

2. Behavioural Data

Analysing student engagement with different learning modules to identify effective and ineffective teaching strategies

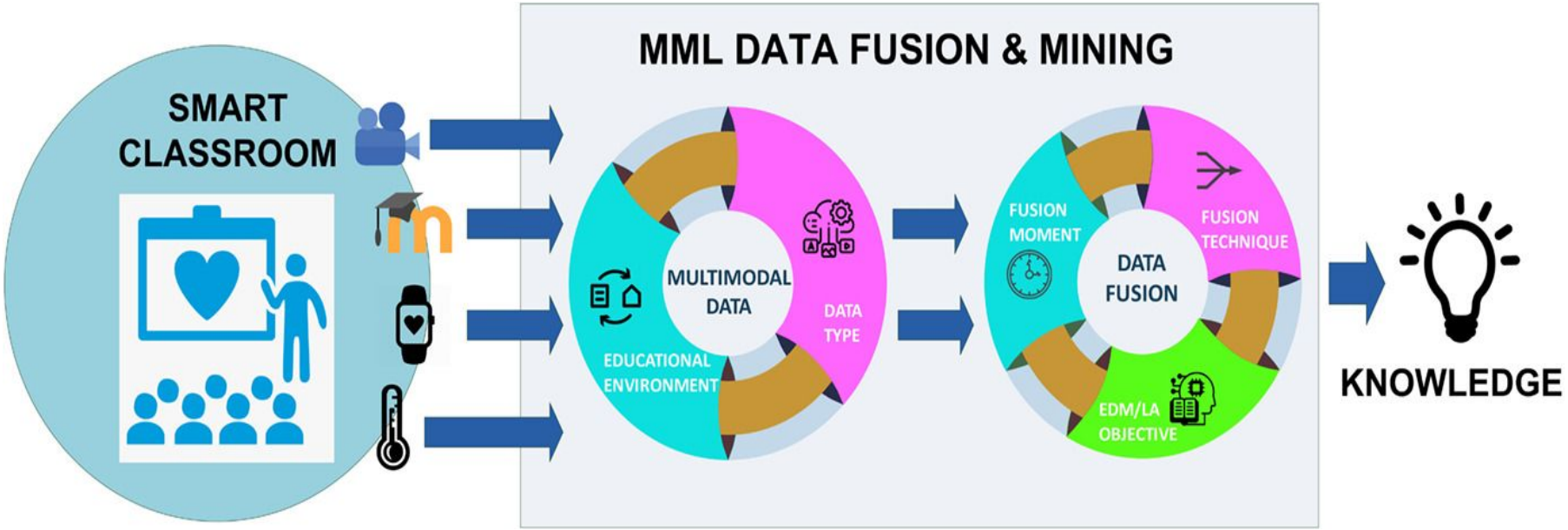
3. Attitudinal Data

Understanding student perceptions of a new instructional approach to inform its refinement or adoption

NYC SCHOOL OF DATA

Open Data needed for Learning Analytics

- The combination of administrative, behavioural, and attitudinal data, in other words fusion of data, provides a comprehensive understanding of the learning process, enabling educators and researchers to identify patterns, develop interventions, & improve educational outcomes.



Source: https://wires.onlinelibrary.wiley.com/doi/full/10.1002/widm.1458

Educational Data Mining Techniques

Source: https://www.linkedin.com/pulse/data-mining-its-applications-kunal-sevak/

Improving the Quality of Education

1. Performance Analysis of Educational Programs
2. Level of Equity in Education
3. Effectiveness of Resource Allocation
4. Predictive Analytics for Student Success
5. Learning Environment Optimization
6. Community and Family Engagement
7. Needs Analysis for Curriculum Development
8. Teacher Professional Development
9. Impact of Extracurricular Activities
10. Impact of Digital Divide
11. Effect of Personalized Education
12. Impact of Automated Assessment
13. Use cases for LLMs/Generative AI
14. Assessment of Internet Security
15. Optimizing Decision-Making
16. Ethical Computing
17. Effectiveness of Computing Pedagogy
18. ……

NYC SCHOOL OF DATA

Open Data for Learning Analytics

Analysis Example

A Snapshot of Computer Science Education:
New York City (2015-2022)

Computer Science in NYC

- CS4All: An initiative started in 2015
- Why should students learn CS? Why investigate?

Research Questions

How have the demographics of students taking computer science changed over time?

Do computer science course offerings differ by borough?

What is the relationship between school bandwidth and computer science course offerings?

Does the number of stem teachers in a school in a given year correlate to the number of computer science courses offered?

NYC SCHOOL OF DATA

NYC OpenData

NYC Public Schools InfoHub

Computer Science and Career and Technical Education Reports

Local Law 177 enacted in 2016 requires the Department of Education of the New York City School District to submit to the Council an annual report concerning computer science education for the prior school year.

Local Law 177 of 2016 – Report on Computer Science Instruction

- LL 177 2015-2016
- LL 177 2016-2017
- LL 177 2017-2018
- LL 177 2018-2019
- LL 177 2019-2020
- LL 177 2020-2021
- LL 177 2021-2022

Intergovernmental Affairs

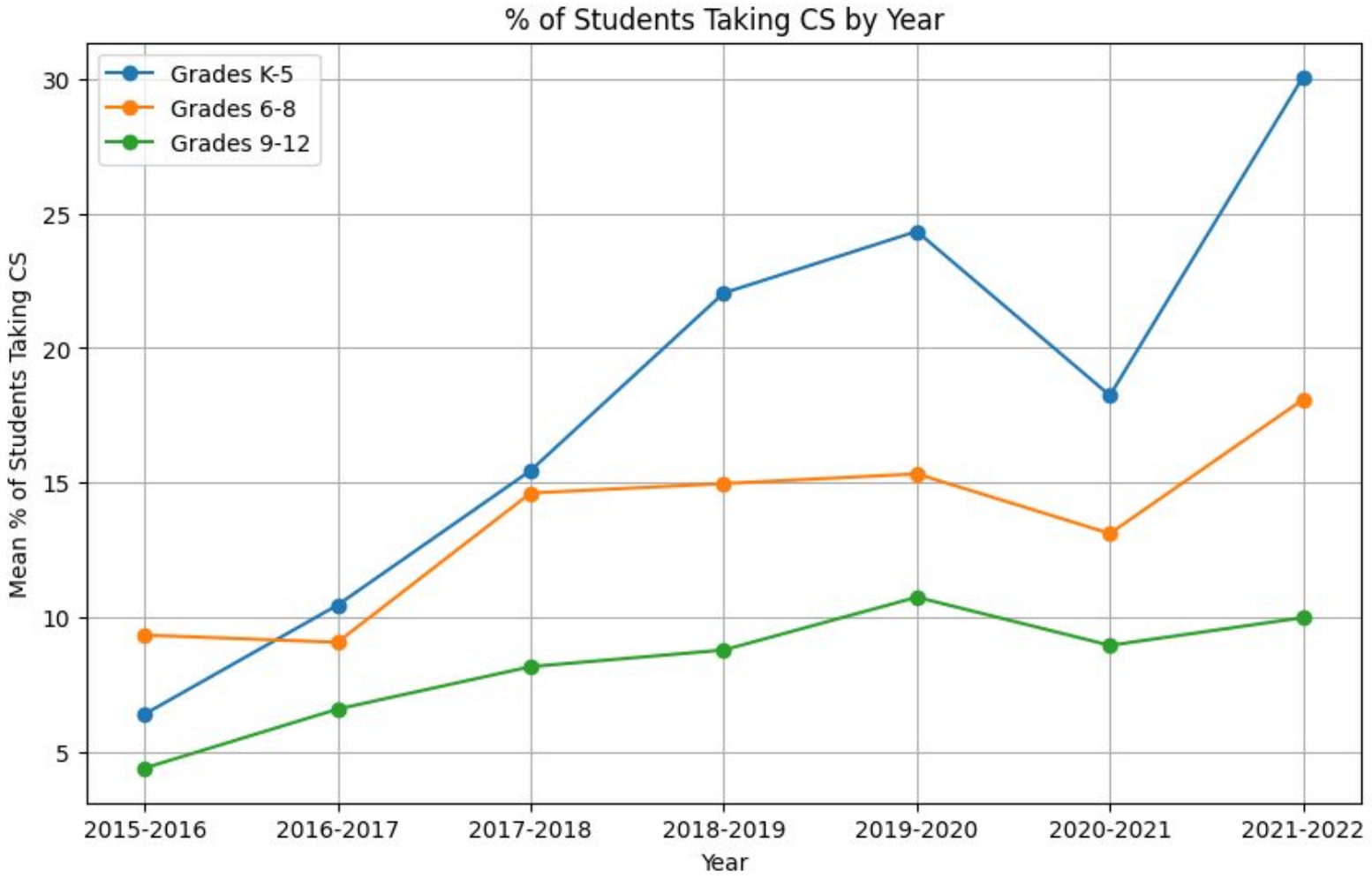
- Intergovernmental Affairs
- Arts Reporting
- Bullying, Harassment and Discrimination
- Class Size Reports

	Year	Category Type	Category Values	# of Students All Grades	# Students Taking CS All Grades	% Students Taking CS (CityWide) All Grades	% Within CS All Grades	S
0	2015-2016	District	1	11406	320	2.8	0.6	
1	2015-2016	District	2	61496	3503	5.7	6.5	
2	2015-2016	District	3	21948	324	1.5	0.6	
3	2015-2016	District	4	13093	226	1.7	0.4	
4	2015-2016	District	5	11381	283	2.5	0.5	
...	...	...	...	...	...	...	...	
351	2021-2022	Ethnicity	White	155236	30747	20	17	
352	2021-2022	Gender	Female	465173	85080	18	47	
353	2021-2022	Gender	Male	494604	94864	19	53	
354	2021-2022	Student with Disability Status	Students with Disabilities	201267	35495	18	20	
355	2021-2022	Student with Disability Status	Students without Disabilities	758510	144449	19	80	

	dbn	num_stud	ftstemteach	ptstemteach	year	school_type	totstemteach	ftstemper	ptstemper	totstemper	bandwidth	numcs	numcsap	numcsap	numcsfull	numcspartial	borough
0	01M015	176.0	0.0	0.0	2016		0.0	0.000000	0.0	0.000000	10.0	NaN	NaN	NaN	NaN	NaN	Manhattan
1	01M019	270.0	0.0	0.0	2016		0.0	0.000000	0.0	0.000000	20.0	NaN	NaN	NaN	NaN	NaN	Manhattan
2	01M020	581.0	0.0	0.0	2016		0.0	0.000000	0.0	0.000000	10.0	NaN	NaN	NaN	NaN	NaN	Manhattan
3	01M034	394.0	4.0	0.0	2016		4.0	0.010152	0.0	0.010152	10.0	NaN	NaN	NaN	NaN	NaN	Manhattan
4	01M063	203.0	0.0	0.0	2016		0.0	0.000000	0.0	0.000000	10.0	NaN	NaN	NaN	NaN	NaN	Manhattan

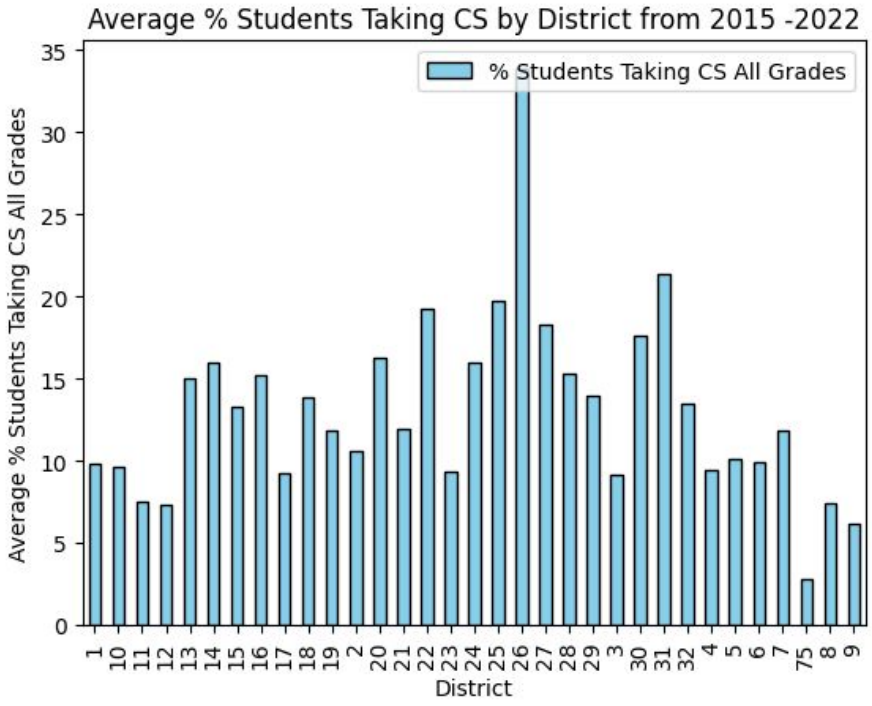
NYC SCHOOL OF DATA

RQ1 - How have the demographics of students taking computer science changed over time?

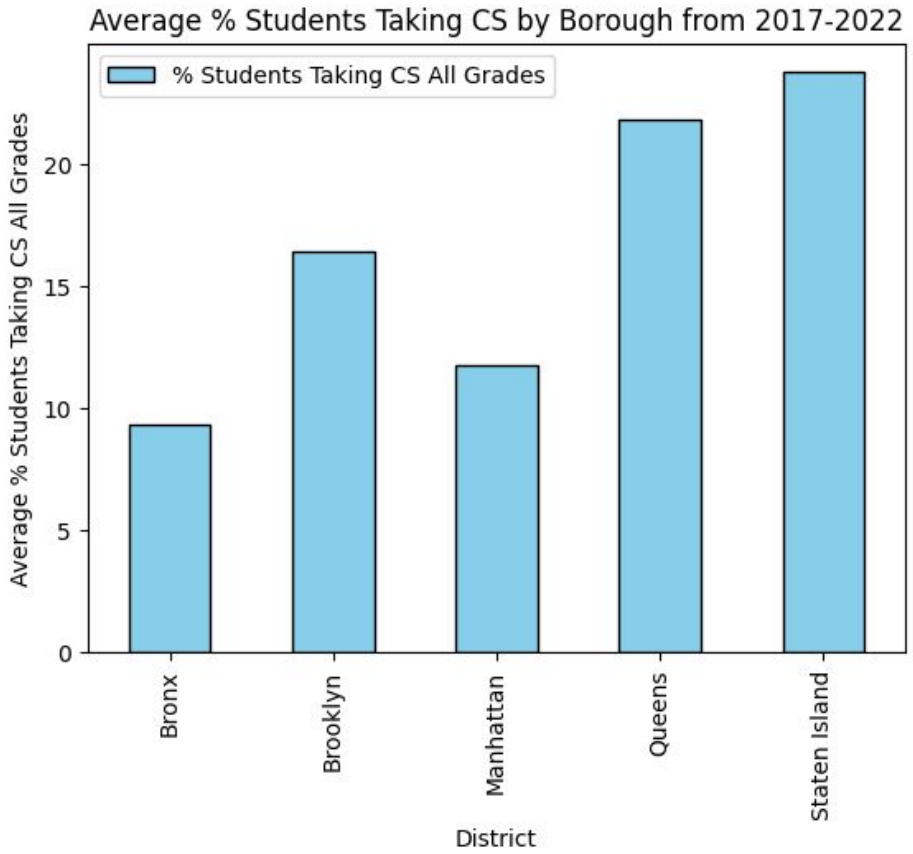


NYC SCHOOL OF DATA

RQ1 - How have the demographics of students taking computer science changed over time?



Borough with the most significant change over years: Bronx

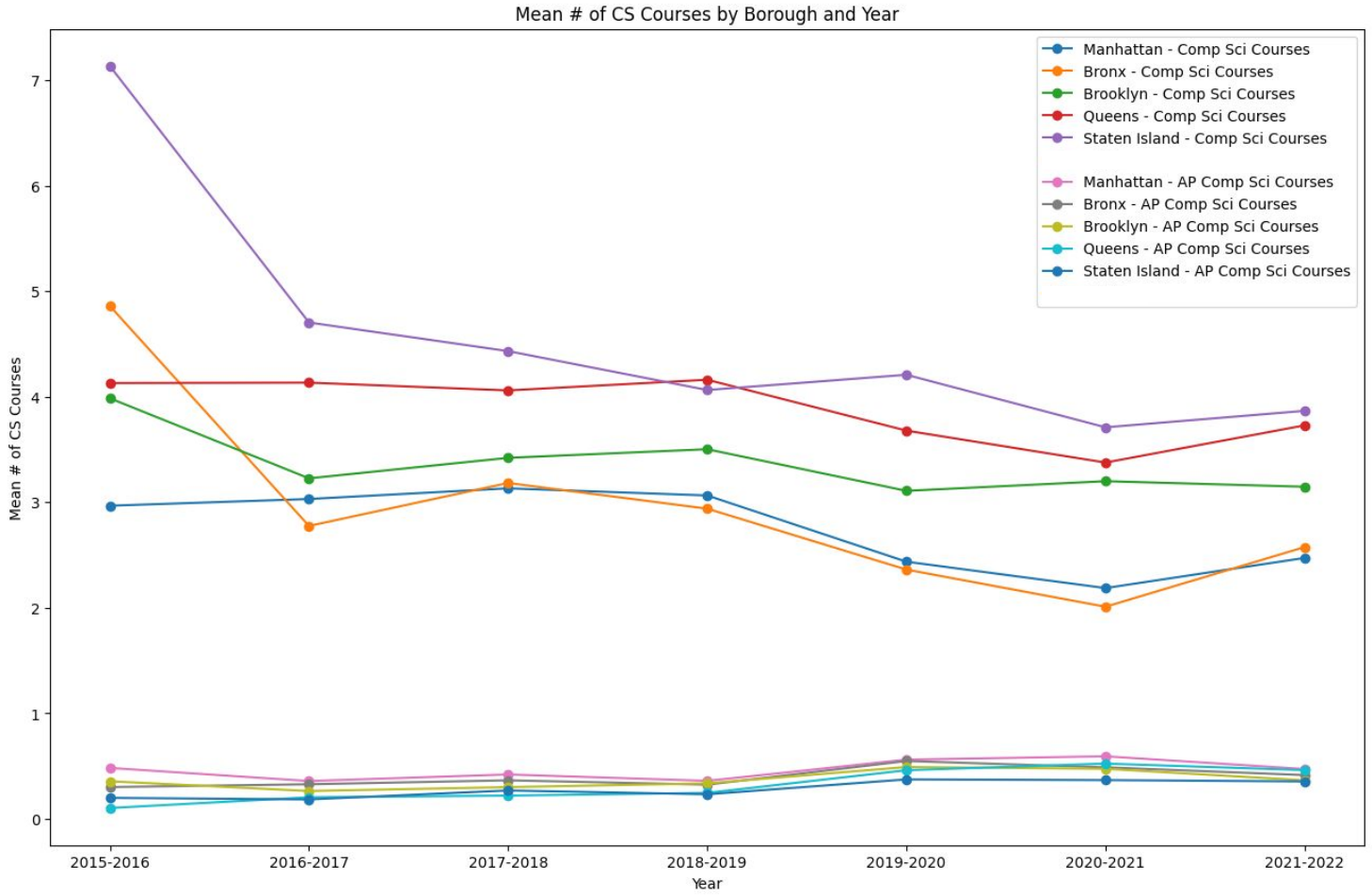


RQ1 - How have the demographics of students taking computer science changed over time?

Demographics	2015 (% of students)	2022 (% of students)	Significant Results
English Language Learners	4.5%	22%	Equalized 20-21 Shifted in 21-22
Non English Language Learners	5.7%	18%	
Economically Disadvantaged	8.9%	19%	Overall Decrease 20-21 Shifted 21-22
Non- Economically Disadvantaged	11.3%	18%	
Male	5.9%	19%	M > F in CS by an avg of 1%
Female	5.1%	18%	
Race	Asian - 8.9% White 7.3% Multiracial - 5.0% Native Am - 4.7% Hispanic - 4.7% Black - 3.6%	Asian - 25% White 20% Multiracial - 14% Native Am - 21% Hispanic - 18% Black - 14%	Inc across all groups Significant dip in Multiracial during pandemic
Students with Disabilities	4.3%	18%	Gap Inc from 2015 - 2019. In 2020 CS slowly becoming more accessible
Students w/o Disabilities	5.9%	19%	

NYC SCHOOL OF DATA

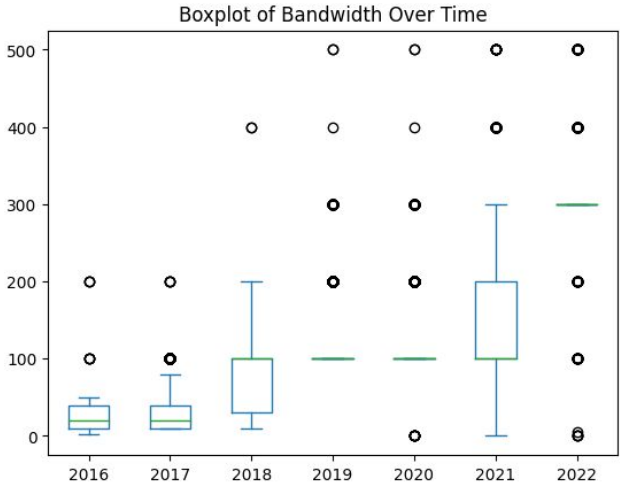
RQ2 - Do computer science course offerings differ by borough?

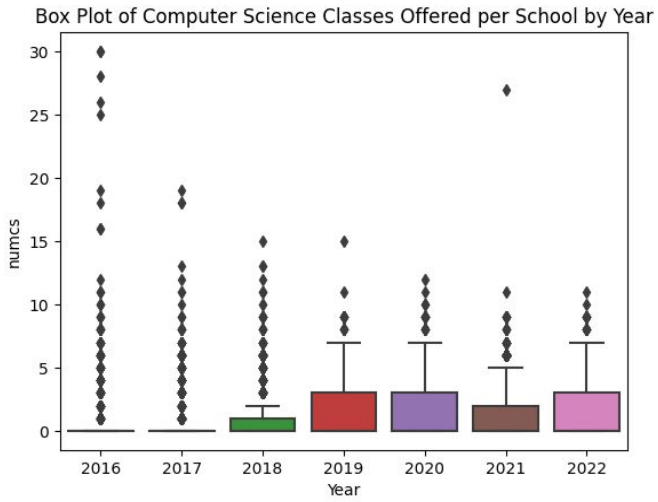


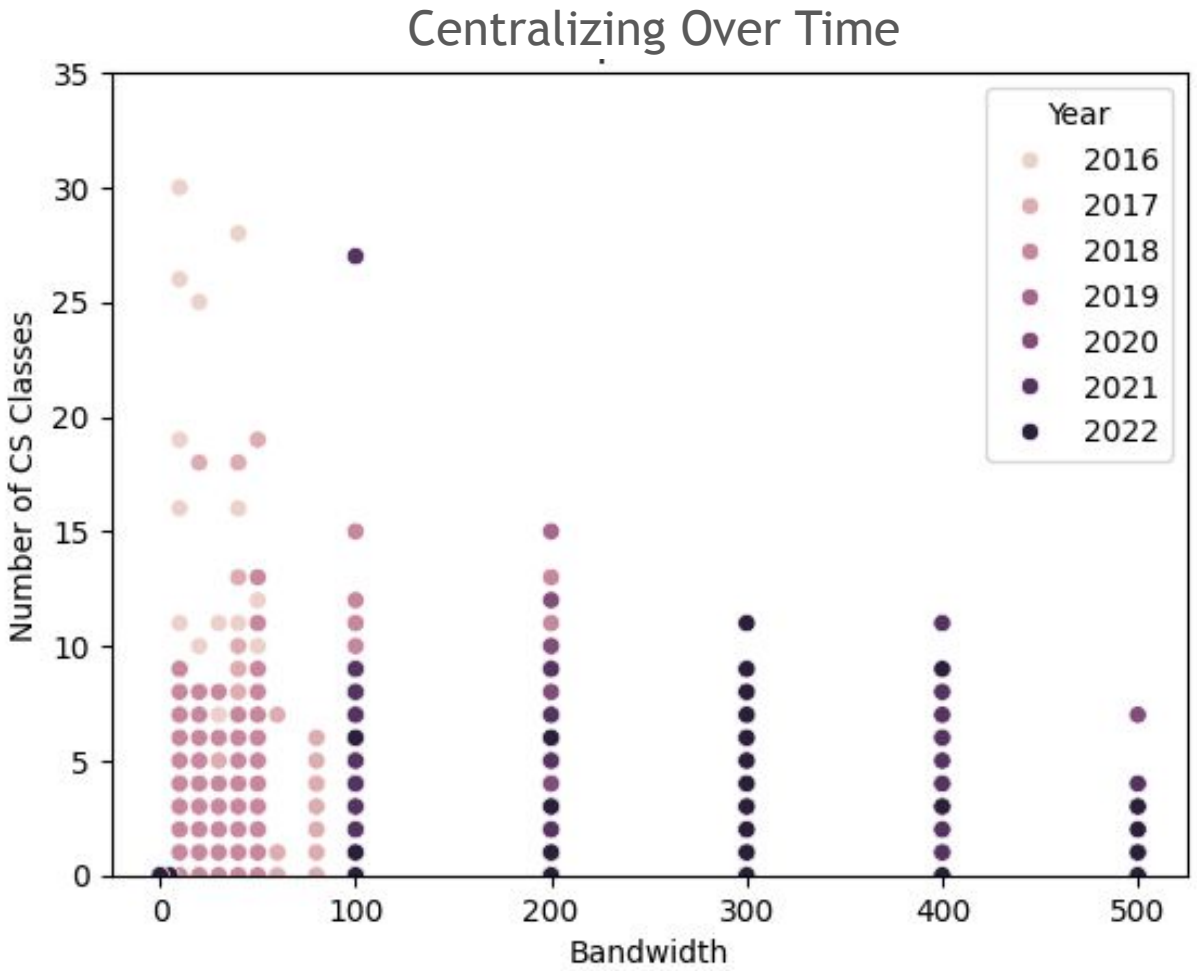
These are the average number of computer science courses per school by borough that were offered each year.
These averages are for schools that offered computer science courses.
The number of schools offering computer science courses has increased by about 500 from 2015-2022.

NYC SCHOOL OF DATA

RQ3 - What is the relationship between school bandwidth and computer science course offerings?

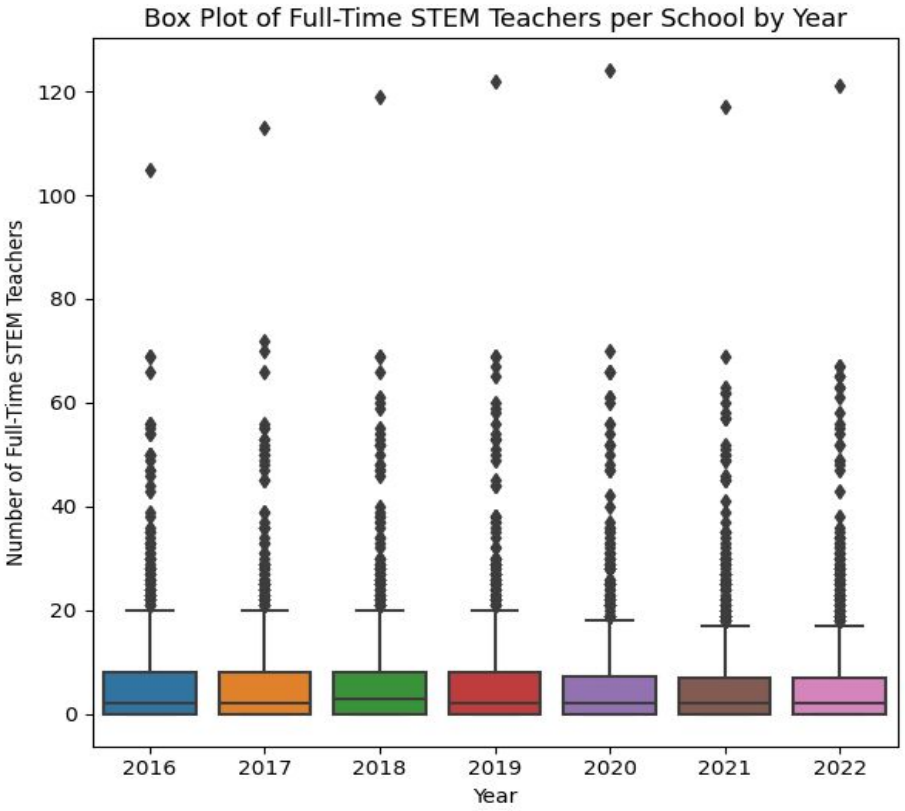


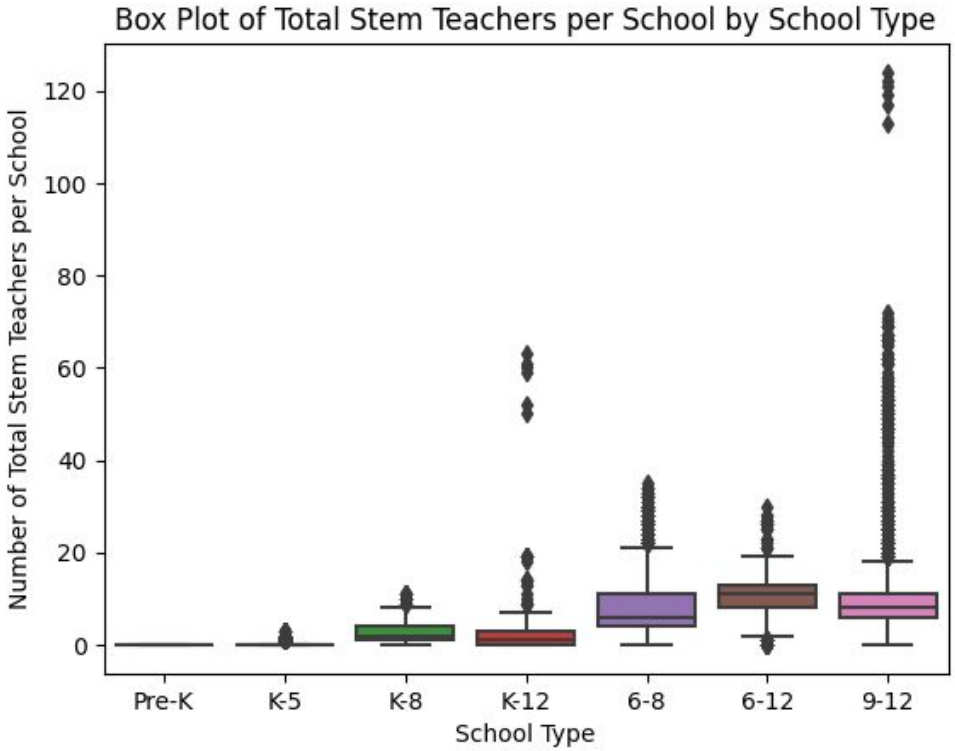




NYC SCHOOL OF DATA

RQ4 - Does the number of stem teachers in a school in a given year correlate to the number of computer science classes offered?

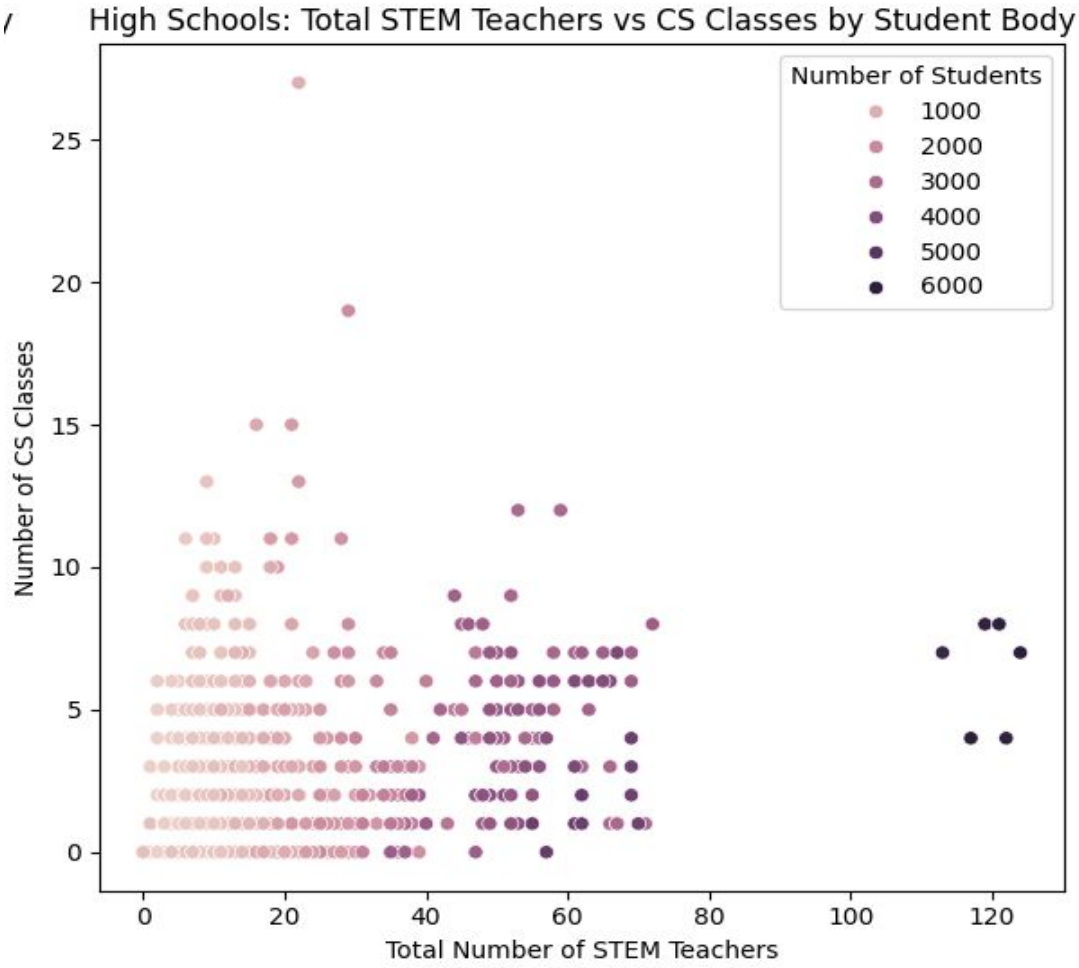




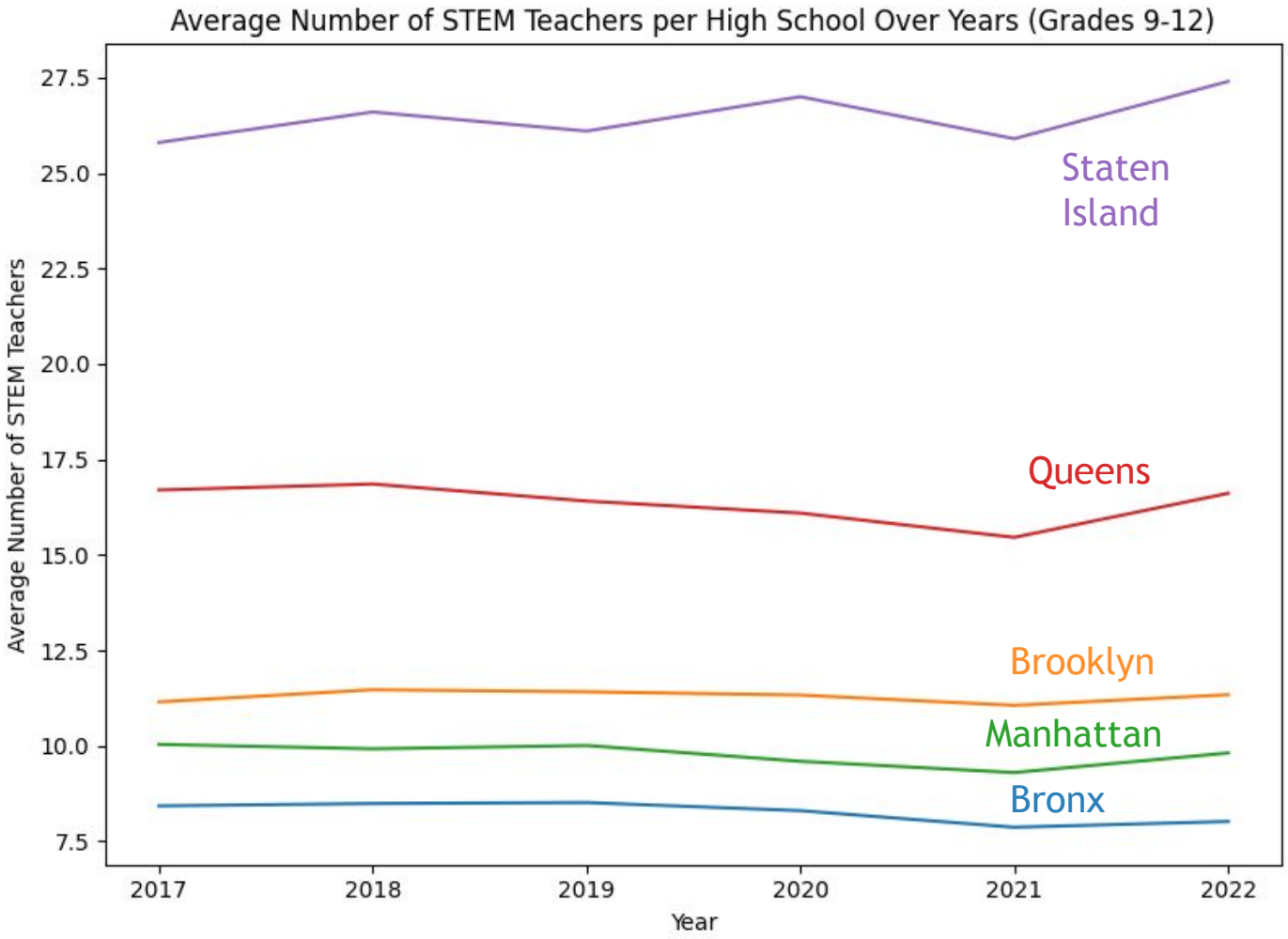
NYC SCHOOL OF DATA

RQ4 - Does the number of stem teachers in a school in a given year correlate to the number of computer science classes offered?

	Number of Students	Number of FT STEM Teachers
Number of Computer Science Courses	.21	.30
Number of AP Computer Science Courses	.33	.48



RQ4 - Does the number of stem teachers in a school in a given year correlate to the number of computer science classes offered?



Average Number of Students Per High School:	
Staten Island	1745.92
Queens	1003.32
Brooklyn	644.72
Manhattan	505.91
Bronx	436.15

NYC SCHOOL OF DATA

Conclusions and Future Work

More complete and interoperable data

Attitudinal data

Additional Research

Behavioral data

NYC SCHOOL OF DATA

Open Data for Learning Analytics

Brainstorming Session

Scenario-based discussion time!

Main Goal

Based on the report prepared by NYC Open Data (2020), we have to create case studies and presentations to share how Open Data can improve service delivery, promote equity, and increase efficiency (Strategic Initiative 11).

Please, work on the following scenarios and discuss possible pros and cons of the provided cases!

NYC SCHOOL OF DATA

Scenario #1 - Increasing Efficiency: Individual Differences

“Sheldon, the principal of a primary school, is curious about the general performance of students in the school. He is not only interested in demographic information about students like socioeconomic status, parents' education, siblings, having a computer at home, early childhood education, etc. but also wants to understand the abilities of students as well as the subject field that the students are good at. He figures out that this data can be collected by administering several surveys and tests (like personality, cognitive ability, motivation, problem solving, computational thinking, social intelligence, etc.). After collecting all of the data, he can figure out what kind of individual differences exist between the students. Moreover, he can use these data to classify the students according to their interests and abilities to provide relevant after school activities.

- What could be some relevant demographic information to collect in order to identify individual differences among students?
- How will the data be treated if there are marginal groups (students having language barriers, or learning difficulties, gifted students etc.)?
- How will the data be treated if there are outliers in the data? To what extent will the results of analysis be reliable with or without outliers?”

Scenario #2 - Improving Service Delivery: Increasing Accessibility

Ansley, a middle school language teacher, is worrying about some of her students that they will possibly fail the course. She thinks that either they are not motivated to learn or they lack access to instructional materials, technology, Internet or similar necessary learning resources. The causes may also due to some challenges about transportation, the quality of food provided at school, the physical conditions of learning environments, etc.

- What are the possible causes of low grades and performance in any course?
- How can we reveal those potential accessibility obstacles students faced?
- What kind of data and sources can help Ansley to figure out the problems?

We can collect some data about accessible technologies, libraries near the school, instructional materials provided by the school and the library, even transportation options, location of the school, etc.

- What kind of data is needed to answer about the possible student failure/success?
- Can we predict possible student failure/success based on these data?
- What could be some serious decisions taken based on the analysis?

Scenario #3 - Promoting Equity: Cultural Differences & Habits

Casey is working at a school where most of the students are from different countries, having English as their second language as well as varied prior knowledge and study habits. Casey, with an understanding of “one size does not fit all”, decides to explore these cultural differences to establish a more inclusive teaching environment for those students. She thinks “Although I cannot name it, the students may emerge in similar behaviours. So, if I come up with clusters, it will be easy for me to provide interventions based on their characteristics”.

- What could be some relevant information to collect in order to identify cultural & habitual differences among students?
- How will the data be treated in terms of finding similarities and differences between different student groups?
- Which clusters did you come up with? How did you name them?
- What could be some possible insight gained based on these efforts?

Wrap-Up the Session

Questions & Comments

Thank you!

NYC SCHOOL OF DATA

Stay in touch at beta.nyc/links

βetaNYC

nycsodata24.sched.com #nycSOdata #opendataweek