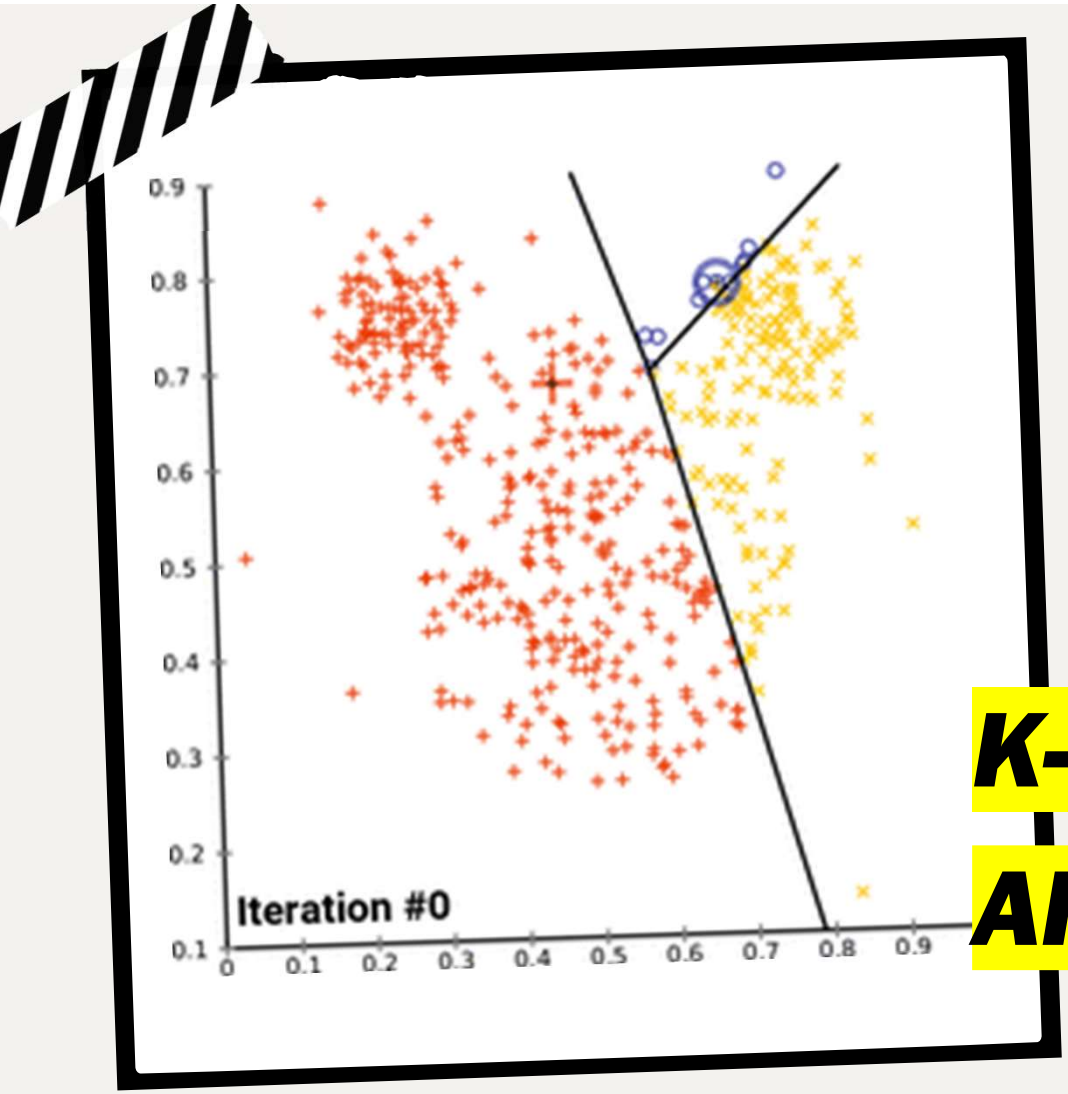
Education Disparity in Chicago Public High Schools

Michael Kralis

Clustering Schools

- I began by clustering schools based on data from ISBE using a K-Means clustering algorithm
- The K-Means algorithm took demographics as inputs like % Low Income, % White, % Black/African American, % Hispanic, etc.



- The goal of the algorithm is to have the “K means” converge to a “stable” point.
- The clusters are the partitions created by determining which mean each point is closest to.
- In this example, the black dots are the means and K=3

K-Means Clustering Algorithm

Summary Statistics (Cluster 0)

% Student Enrollment - White	1.11
% Student Enrollment - Black or African American	89.20
% Student Enrollment - Hispanic or Latino	8.38
% Student Enrollment - Asian	0.45
% Student Enrollment - Low Income	84.46
Student Attendance Rate	91.07
High School Dropout Rate - Total	5.44
High School 4-Year Graduation Rate - Total	80.92
% Graduates enrolled in a Postsecondary Institution within 12 months	57.32

Summary Statistics (Cluster 1)

% Student Enrollment - White	4.15
% Student Enrollment - Black or African American	11.15
% Student Enrollment - Hispanic or Latino	81.30
% Student Enrollment - Asian	2.17
% Student Enrollment - Low Income	89.21
Student Attendance Rate	91.74
High School Dropout Rate - Total	4.41
High School 4-Year Graduation Rate - Total	81.06
% Graduates enrolled in a Postsecondary Institution within 12 months	63.23

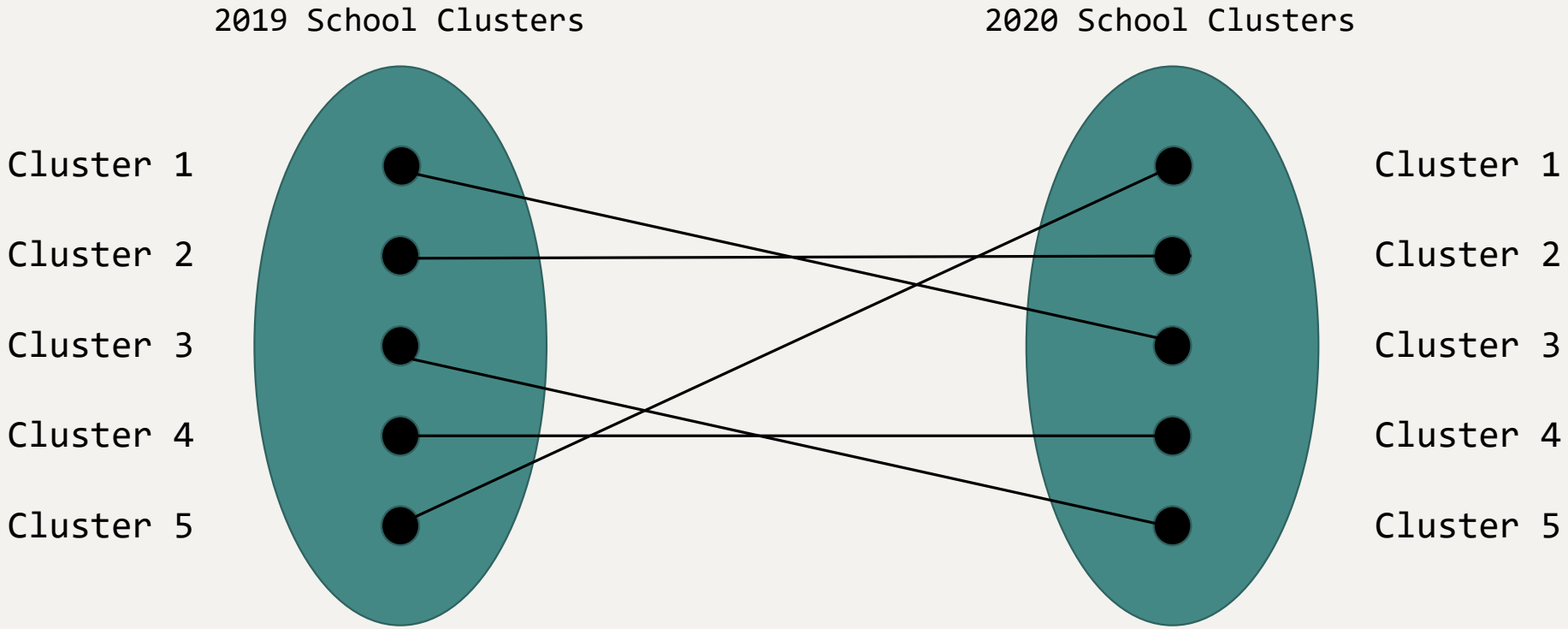
Summary Statistics (Cluster 2)

% Student Enrollment - White	28.25
% Student Enrollment - Black or African American	17.06
% Student Enrollment - Hispanic or Latino	38.19
% Student Enrollment - Asian	11.95
% Student Enrollment - Low Income	54.30
Student Attendance Rate	94.63
High School Dropout Rate - Total	2.24
High School 4-Year Graduation Rate - Total	91.25
% Graduates enrolled in a Postsecondary Institution within 12 months	81.36

Trends over Time

- In addition to identifying summary statistics of the clusters, we can see how these clusters change over time
- More specifically, we are looking to analyze the demographic changes as well as academic changes

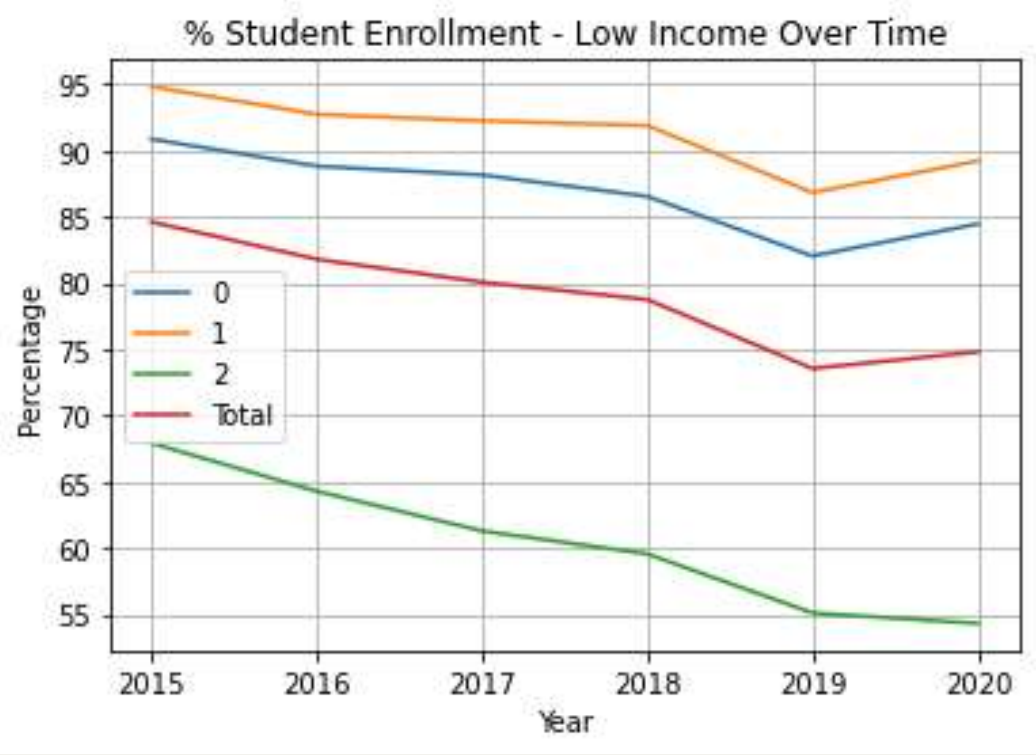
Graphical Model



Create stable matching between years

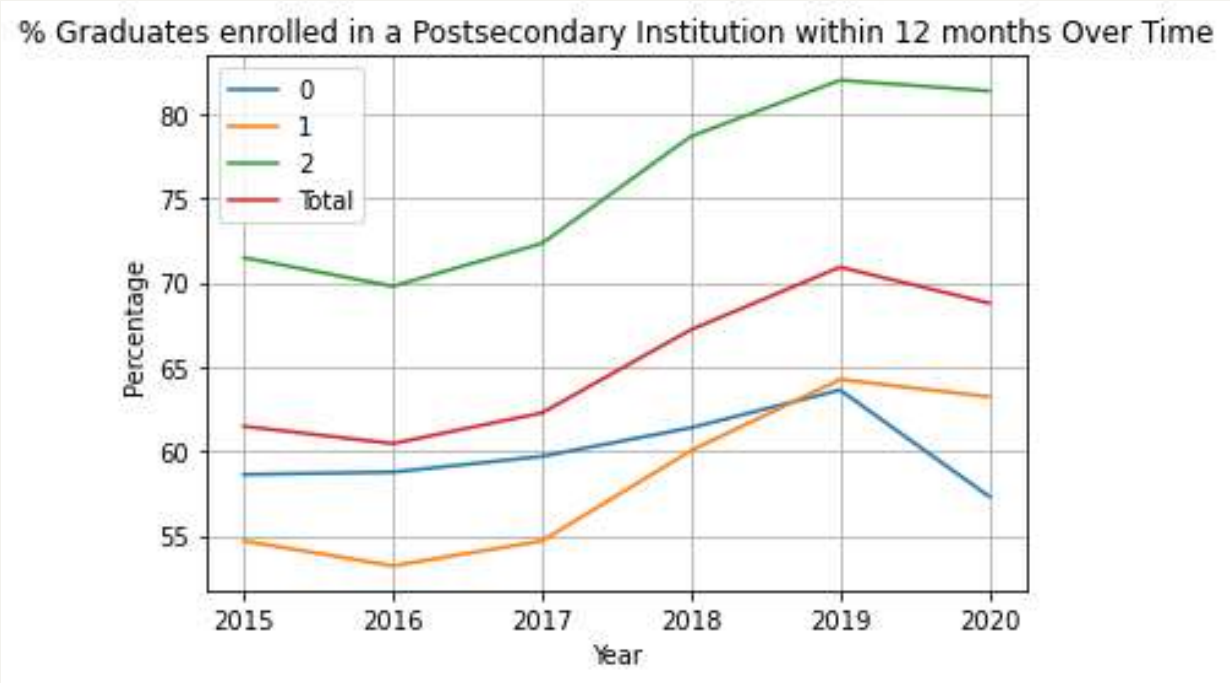
Low Income Proportions Over Time

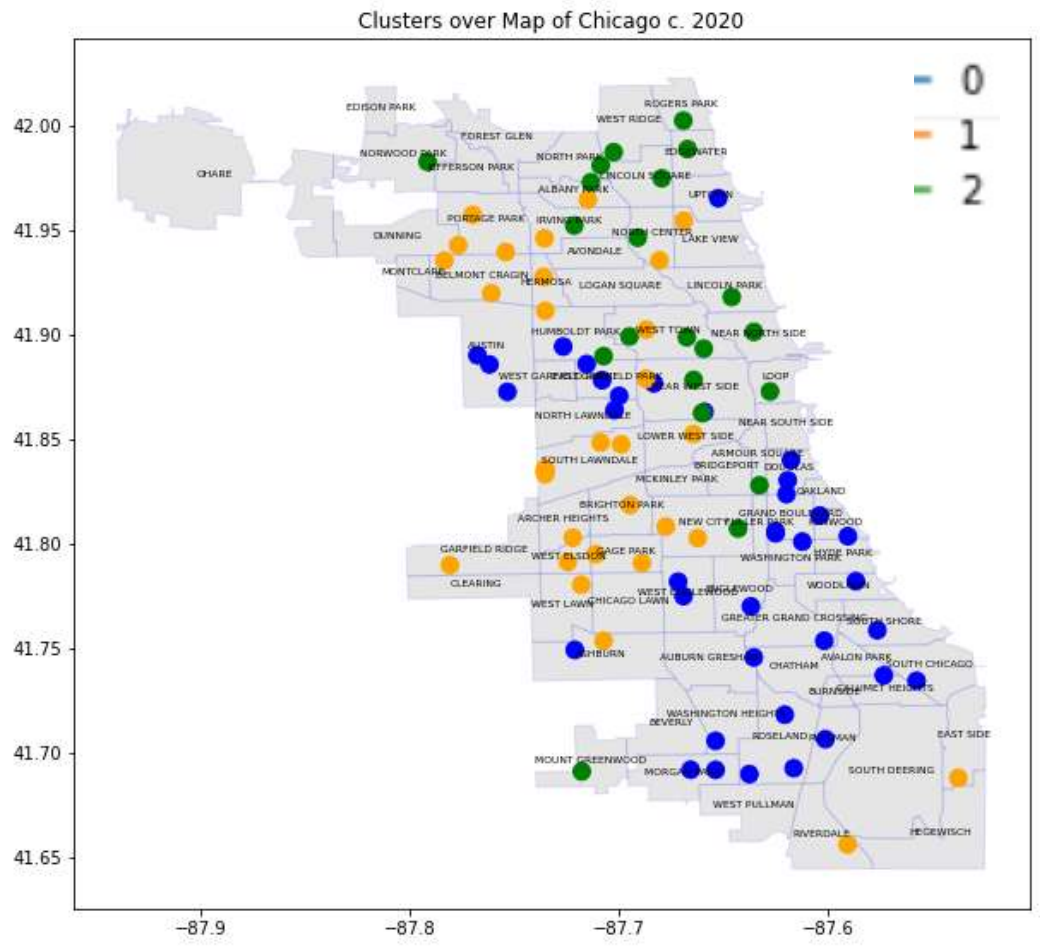
- As we can see, the gap between low-income student enrollment has gotten larger between the clusters.
- More specifically, this gap is mainly attributed to the decline in cluster 2.



College Enrollment Over Time

- Again, the gap between college enrollment statistics has become even larger.
- More importantly, we can see how the pandemic halted growth rate of CPS students enrolling in college.
- We can also see how the pandemic affected students of color.





Mapping these Clusters

- After seeing these clusters projected onto a map of Chicago, we can see just how segregated the city is.
- Not only do the clusters give good indication of academic factors, but also partitions Chicago well.

Traditional Approach: Correlation Matrix

Index	ıt Enrollmen	ent - Black c	ollment - His	ıt Enrollmer	Enrollment - L	ıt Attendanc	ol Dropout I	Year Graduat	stsecondary
% Student Enrollment - White	1	-0.459694	0.0413344	0.670453	-0.881595	0.485443	-0.31436	0.313289	0.507657
% Student Enrollment - Black or Afric...	-0.459694	1	-0.897257	-0.420106	0.341374	-0.430198	0.269335	-0.232523	-0.432968
% Student Enrollment - Hispanic or La...	0.0413344	-0.897257	1	0.0520113	0.0392081	0.236293	-0.145118	0.101725	0.22719
% Student Enrollment - Asian	0.670453	-0.420106	0.0520113	1	-0.569967	0.413457	-0.250182	0.25526	0.434239
% Student Enrollment - Low Income	-0.881595	0.341374	0.0392081	-0.569967	1	-0.642394	0.447039	-0.483383	-0.644065
Student Attendance Rate	.485443	-0.430198	0.236293	0.413457	-0.642	1	-0.722903	0.784383	0.726142
High School Dropout Rate - Tota	-0.31436	0.269335	-0.145118	-0.250182	0.447039	-0.722903	1	-0.741412	-0.547078
High School 4-Year Graduation Ra -...	0.313289	-0.232523	0.101725	0.25526	-0.483383	0.784383	-0.741412	1	0.771031
% Graduates enrolled in a Postsecond...	.507657	-0.432968	0.22719	0.434239	-0.644	0.726142	-0.547078	0.771031	1

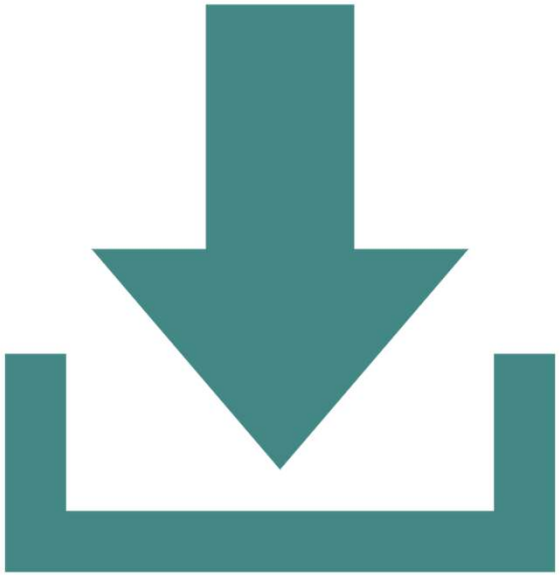
Index	Student Enrollment - Whi	nrollment - Black or Africa	nt Enrollment - Hispanic c	Student Enrollment - As	dent Enrollment - Low In
Student Attendance Rate	0.485443	-0.430198	0.236293	0.413457	-0.642394
High School Dropout Rate - Total	-0.31436	0.269335	-0.145118	-0.250182	0.447039
High School 4-Year Graduation Rate -...	0.313289	-0.232523	0.101725	0.25526	-0.483383
% Graduates enrolled in a Postsecond...	0.507657	-0.432968	0.22719	0.434239	-0.644065

Inference

- From this we can see that in every academic factor, proportion of low-income and African American or Black students is the deciding factor.

Possible future work

- Using other methods to cluster?
- Analyzing different features?
- Look into funding of schools?
- Using methods in algebraic statistics
  - Compute Markov basis functions from elimination Gröbner basis
  - Construct Markov chain to “walk” along fibers
  - Test for independence



Github Repo!

Progress can be found in my github repository here: https://github.com/mkralis123/SoReMo