Analyzing Disparities in Special Education Programs in California Schools

Xin Huang, Eric Sun,
Jasper Dong, Allison Peng
STA 160

Background

- Special/Specialized Education (SPED) programs are important in supporting students with disabilities succeed in academic communities
- The “effectiveness” of a program can be measured by how well a disabled student is integrated into the general education (Brock & Shaefer, 2015)
- Disparities in general education inclusion can occur due to various factors, including geographic, demographic, socioeconomic, etc. (Harper, 2017)
- Students of color are often overrepresented in sped classrooms, causing a significant difference in racial proportions. (Harper, 2017)

Literature Review

Cooc (2021)

Disparities in general education inclusion for disabled students exist between racial groups in varying degrees, but the study is limited to districts that may have possible similar characteristics

Brock & Schaefer (2015)

MANOVA and Tukey's HSD are used to show how urbanicity can have a negative impact on general education inclusion in Ohio schools

Original Question:

“Analyze the accessibility of special education (SPED) programs across different counties in California, with a breakdown by charter and non-charter school status and race.”

New Question:

“What factors significantly contribute to disparities in special education (SPED) effectiveness across California school districts?”

Data Source: California Department of Education

The California Department of Education has a section called Accessing Education Data, whose goal is to obtain data collected by the California Department of Education (CDE), including enrollment data, demographic and student group data, staff data, school and district accountability, statewide assessment results, and directory information.

We chose a dataset that allowed us to analyze special education enrollment broken down into different categories, called SPED Data by Program Setting.

Data characteristics:

- 17 variables (both quantitative and qualitative)
- 115,424 observations
- Subsets of data based on the values of different variables, several overlapping observations

Data Description

Variables of interest:

<u>District Name</u>: Lists of school districts (235 districts)

<u>SPED Enrollment (PROP)</u>: Proportion of total enrolled students placed in a special education program

<u>Program Type</u>: Proportion of the day where SPED students are integrated into the regular classroom

- **PS_RCGT80 = Regular class $\geq$ 80% of the time**
- **PS_RC4079 = Regular class between 40-70% of the time**
- **PS_RCLT40 = Regular class $\leq$ 40% of the time**
- **PS_SSOS = Separate school & other settings**
- **PS_PSS = Preschool setting**
- **PS_MUK = Missing or unknown**

Data Description

Variables of interest:

Ethnicity Enrollment (PROP): Proportion of SPED enrollment that are represented by each ethnic category

- **RB = African American**
- **RI = American Indian or Alaska Native**
- **RA = Asian**
- **RF = Filipino**
- **RH = Hispanic or Latino**
- **RD = Not Reported**
- **RP = Pacific Islander**
- **RT = Two or More Races**
- **RW = White**

Data Preprocessing

- We removed NA values from the data
- To adjust for higher populations in certain districts, we created race proportions by dividing the total race count per district by the total SPED enrollment per district

Exploratory Data Analysis

Summary Statistics for Special Ed Enrollment, Race, and Programs Settings

	SPED Enrollment	Total Enrollment	African American	American Indian or Alaska Native	Asian	Filipino	Hispanic or Latino	Not Reported	Pacific Islander	Two or More Races	White	>80%	40%-79%	<39%	Separate School	Preschool	Unknown
Count	235	235	235	235	235	235	235	235	235	235	235	235	235	235	235	235	235
Mean	2398.668	16364.425	0.064	0.004	0.060	0.015	0.535	0.011	0.001	0.048	0.234	58.783	15.764	17.209	2.596	5.640	0
Std	5736.214	36416.496	0.070	0.014	0.081	0.023	0.193	0.016	0.003	0.030	0.162	13.736	9.101	10.887	3.849	4.335	0
Min	163	327	0	0.003	0.072	0.022	0.687	0.017	0	0.064	0.337	65.950	21.150	21.550	2.550	8.050	0
Max	83528	529902	0.365	0.158	0.445	0.022	0.958	0.094	0.021	0.149	0.685	94.300	43.600	78.900	33.900	32.500	0

Figure 1: Summary statistics for our variables of interest

Exploratory Data Analysis

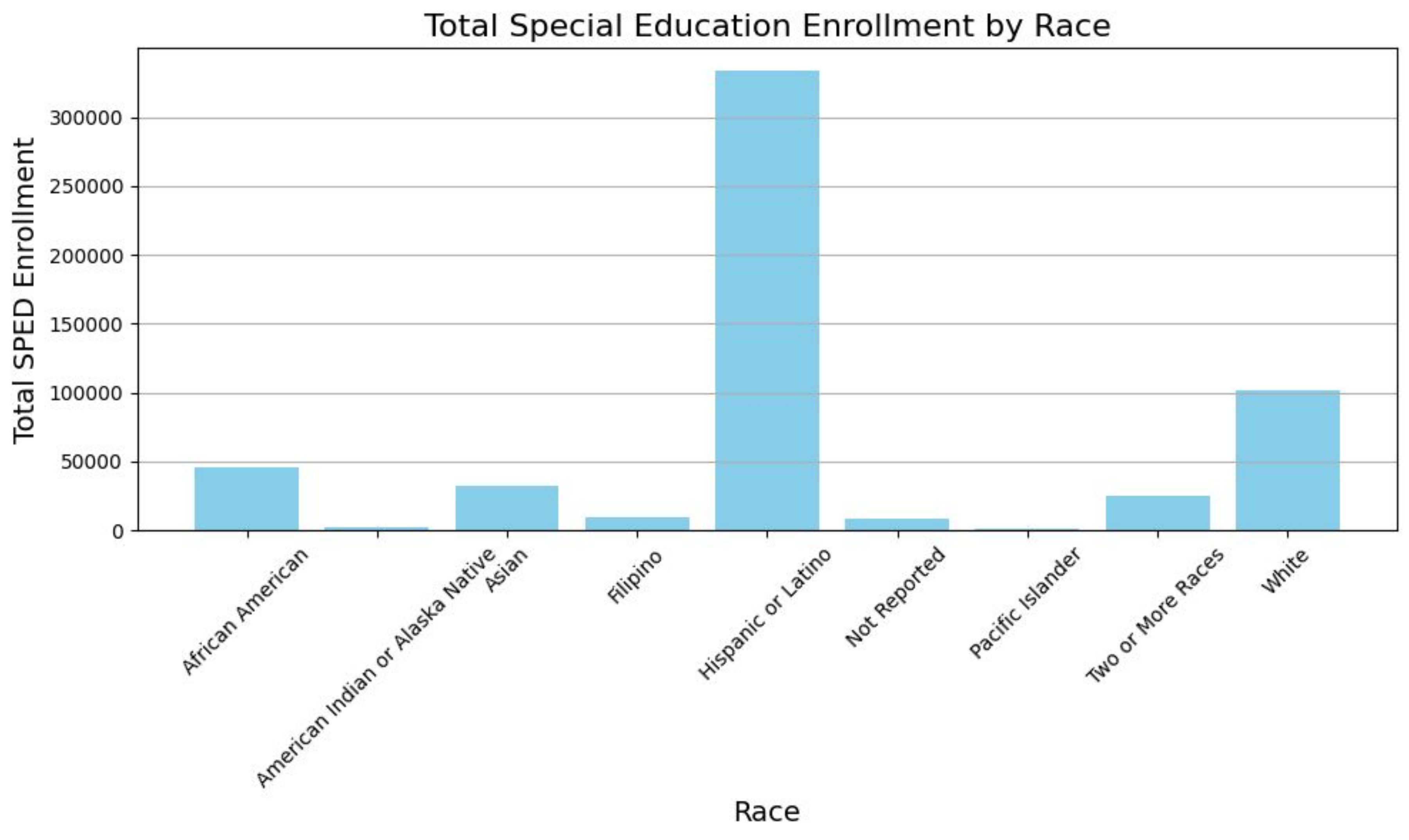

Figure 2: Total SPED enrollment by race

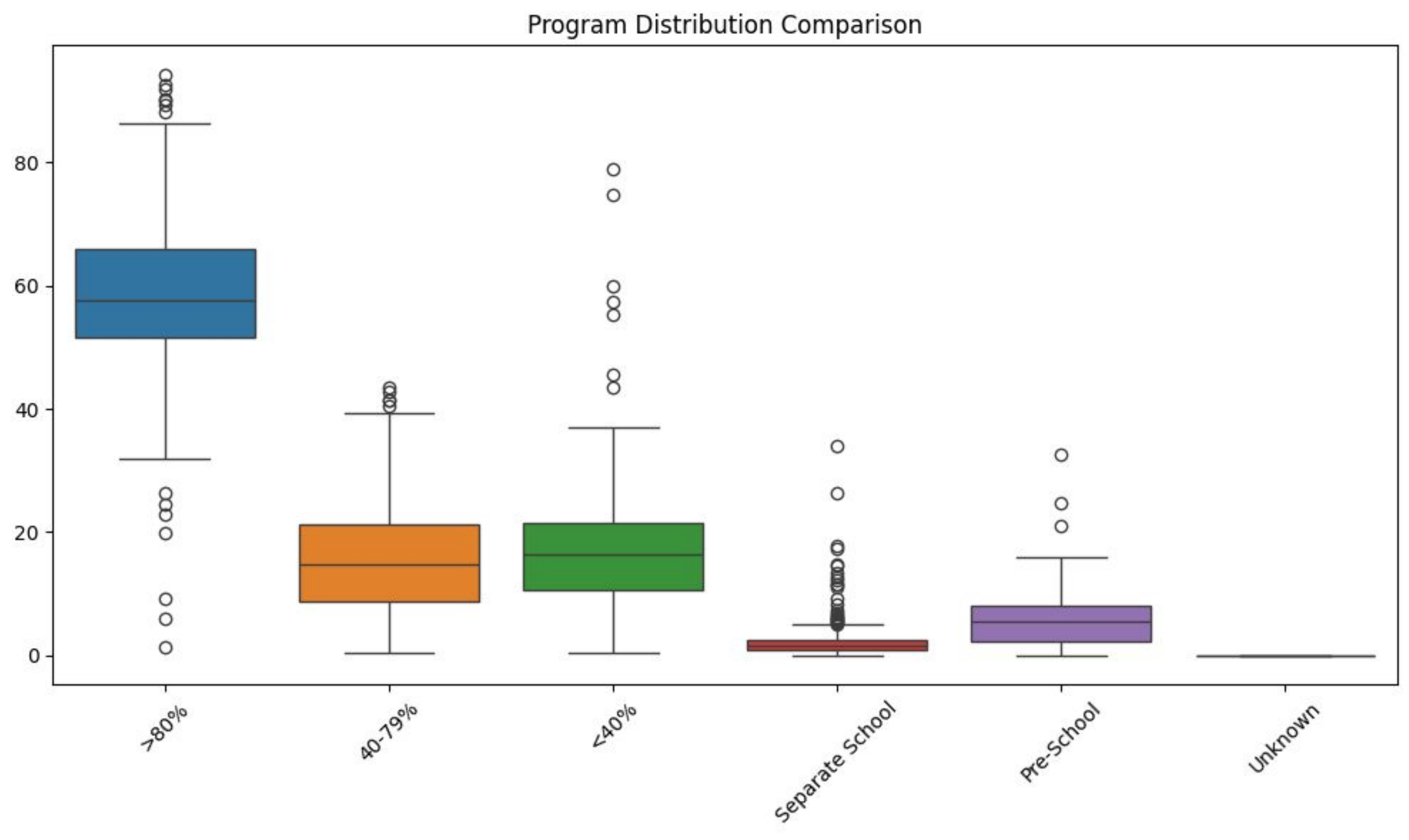

Figure 3: Box plot of distributions of different program types

Exploratory Data Analysis

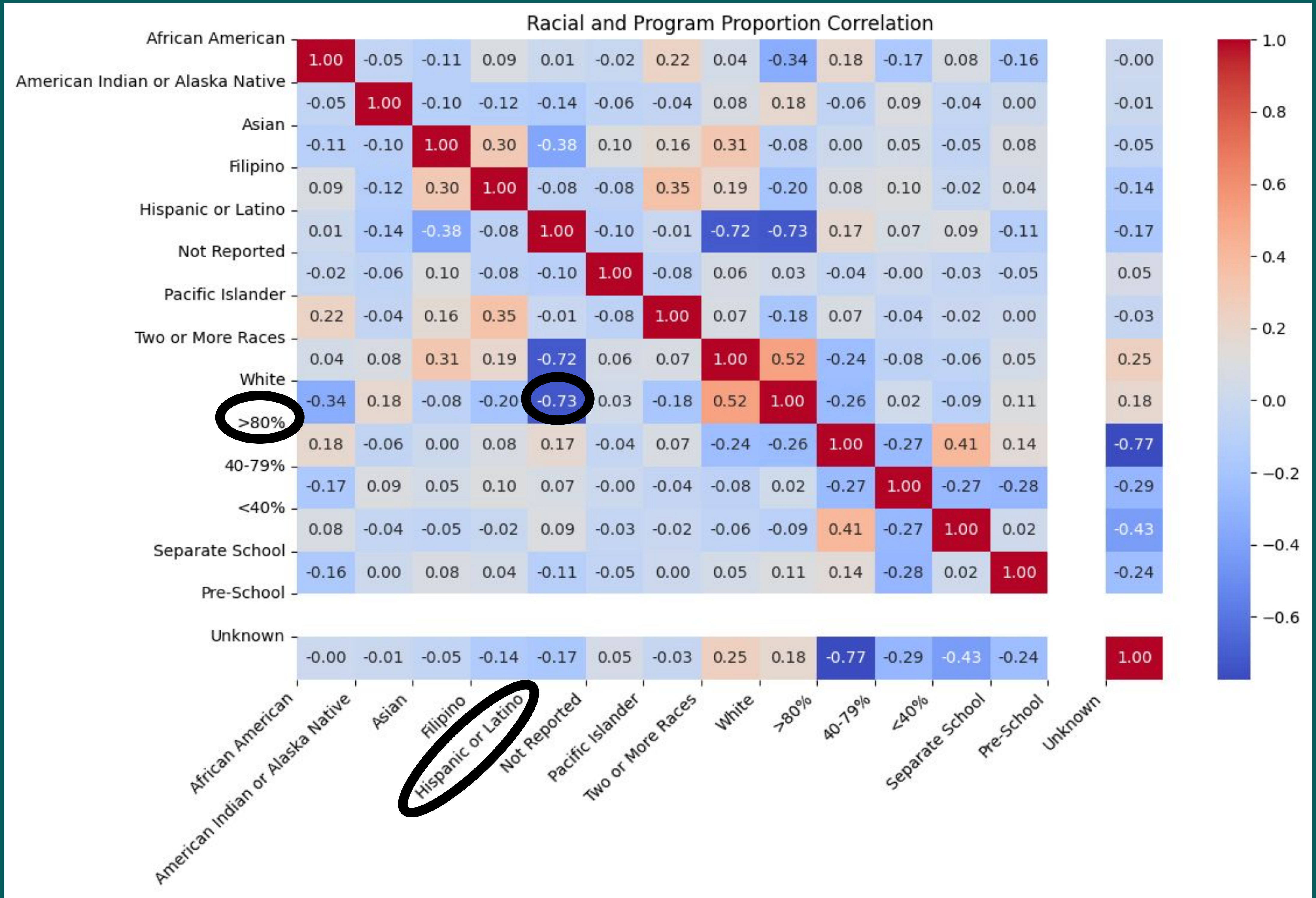

Figure 4: Correlation matrix of race and program type

Plan - Methods

Unsupervised Machine Learning: Cluster the aggregate data with K-means using program type

→

ANOVA: Test for significant difference for program type between clusters, then test for significant difference for races between program type

K-means Clustering

- **K-means clustering** is used to group together similar districts based on program type distributions
- It is a unsupervised machine learning method and given a number of clusters, uses euclidean distance to cluster similar points together
- <u>By creating groups of similar clusters, we can compare distributions of various factors across clusters, which gives us a better interpretation on which factors significantly separate each cluster</u>

Results - K-means Clustering

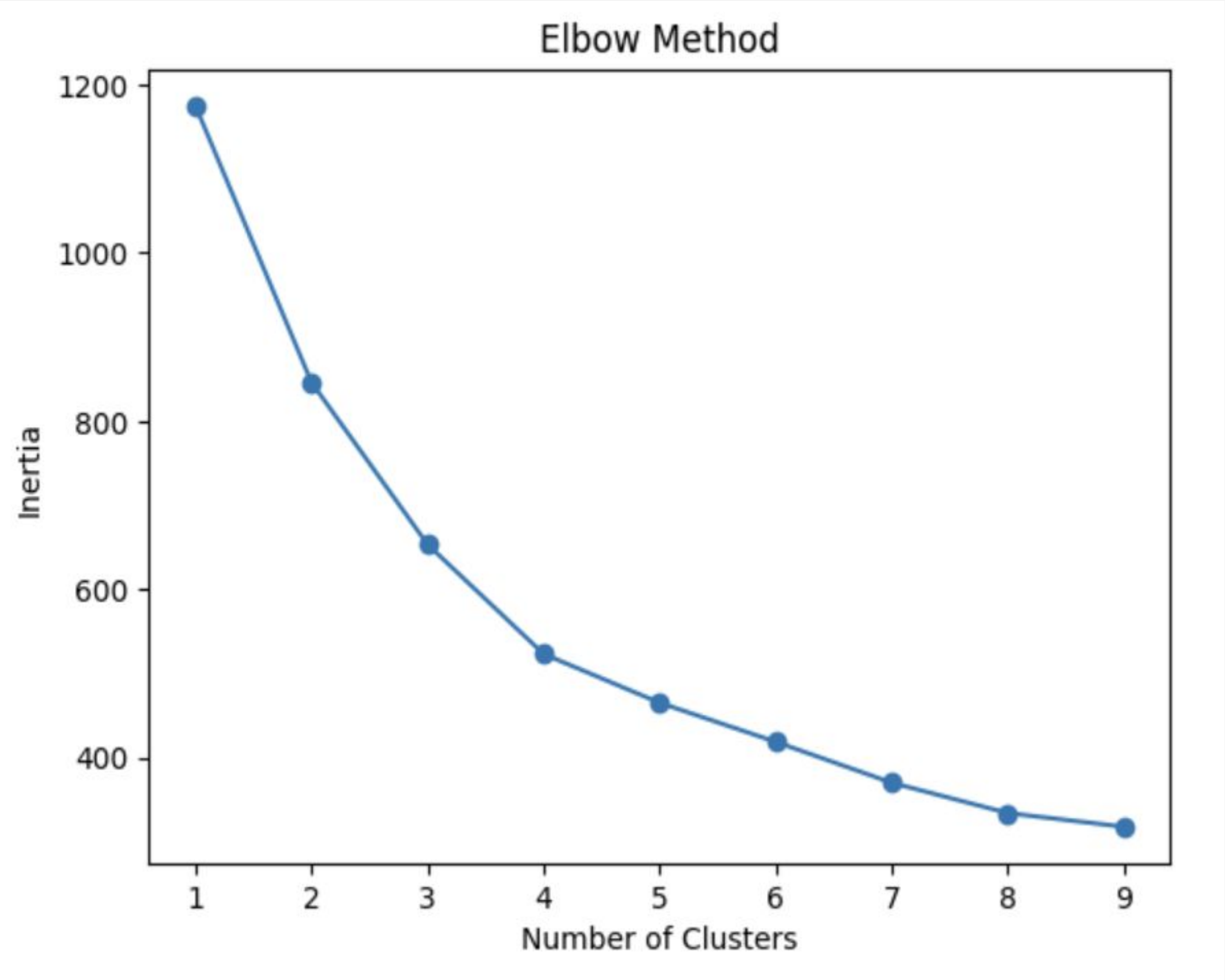

Figure 5: Line graph showing elbow method for selecting clusters

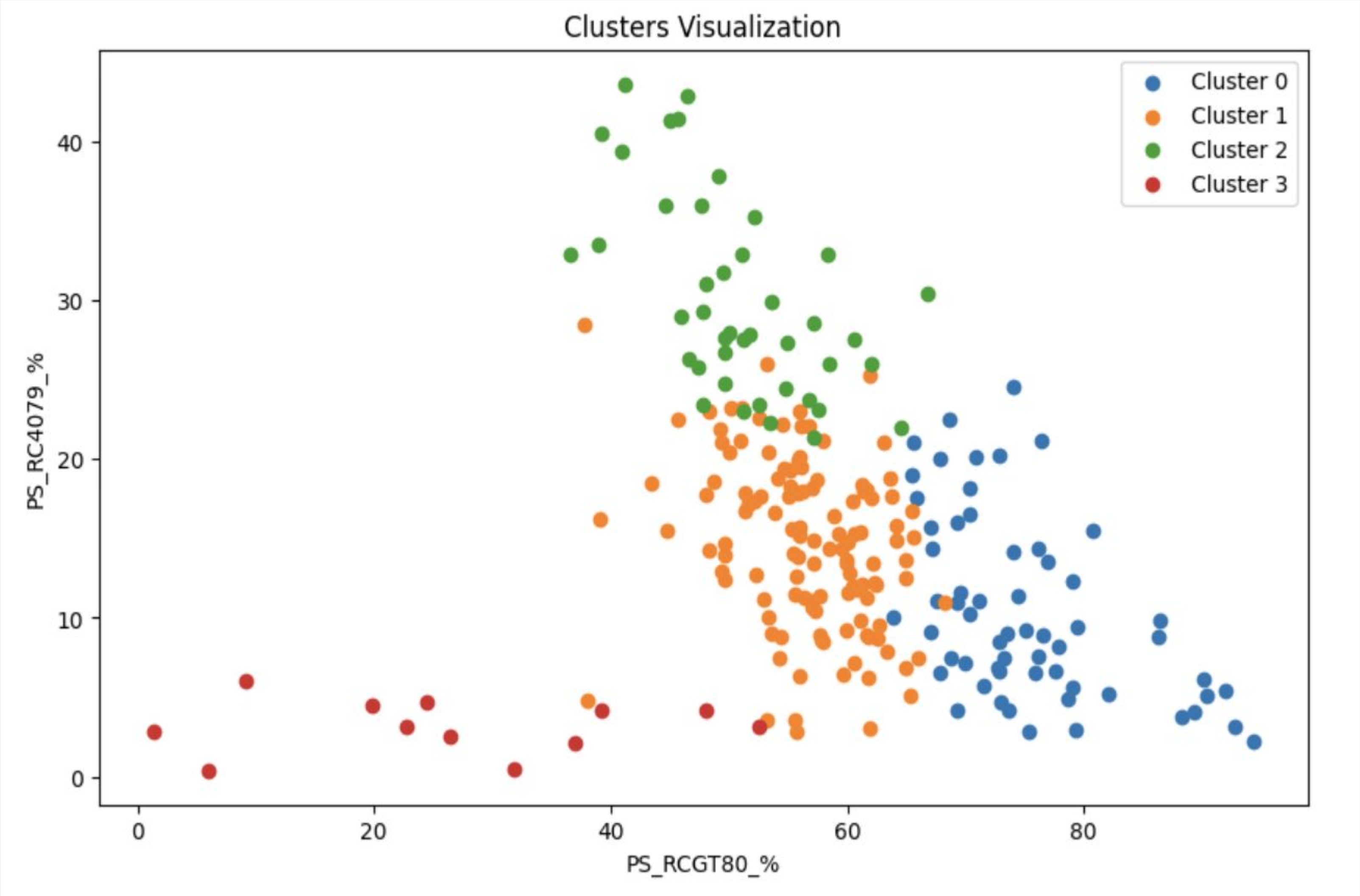

Figure 6: Scatter plot visualizing two dimensions of clusters

- Elbow method is used to assess an appropriate # of clusters to use based on total explained variance; the choice of K = 4 clusters is used

Results - Comparing Across Clusters

Results comparing across clusters using ANOVA. α = 0.05

	F Stat	P-value
>80%	202.023200	2.716e-64
40%-79%	135.566714	1.128e-50
<40%	123.851422	7.743e-48
Separate School	97.208419	1.014e-40
Pre-school	24.154733	1.241e-13
Unknown	NAN	NAN

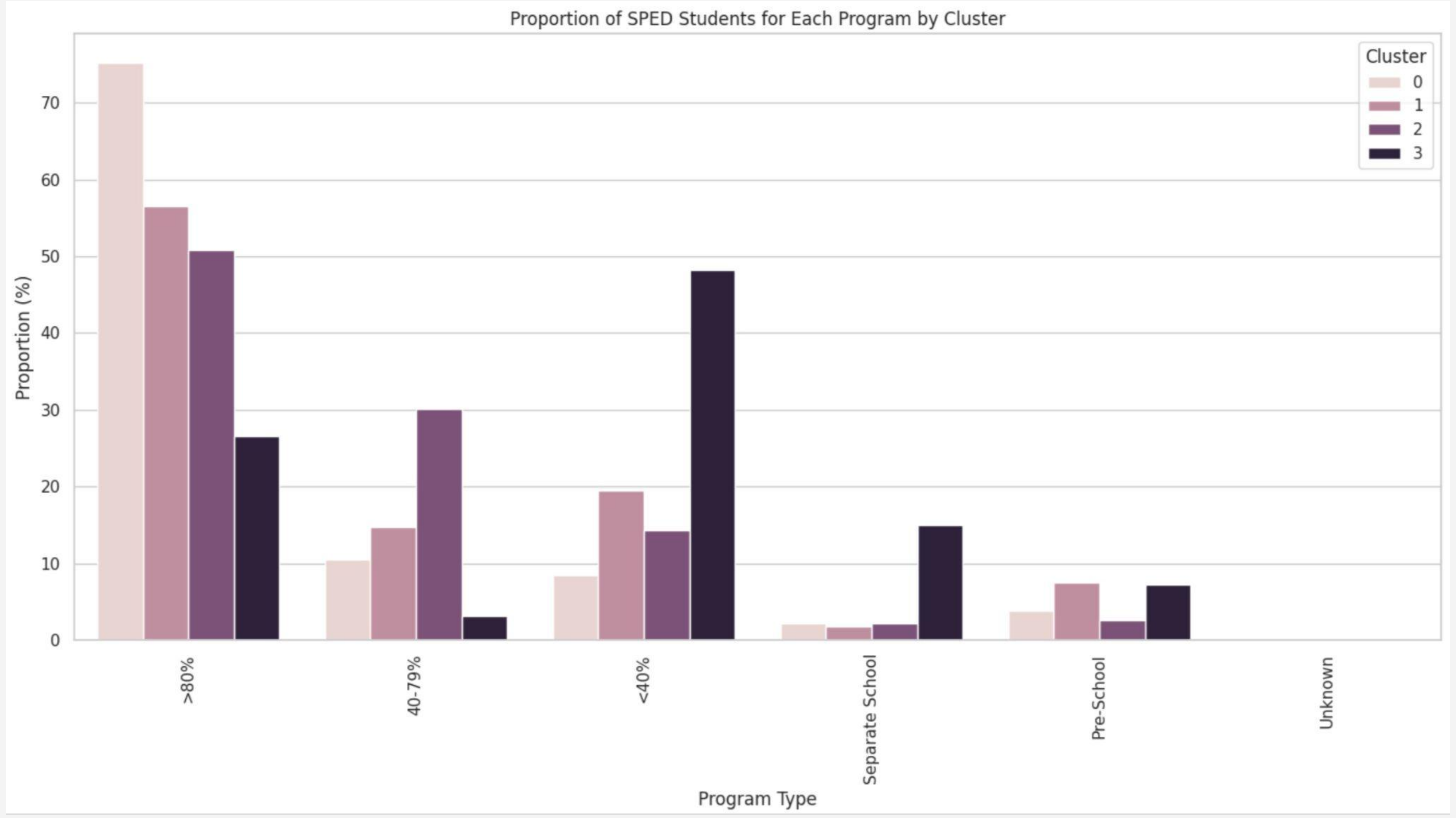

Figure 7: Bar chart showing different program proportions for each cluster

Results - Comparing Across Races

Result for ANOVA testing across clusters. α = 0.05

	F Stat	P-value
African American	1.484361	0.219518
American Indian or Alaska Native	0.391764	0.759044
Asian	2.067739	0.105243
Filipino	4.673547	0.003437
Pacific Islander	4.431138	0.005340
Two or More Races	3.870056	0.009966
White	5.256679	0.001586
Hispanic or Latino	3.197518	0.024209
Not Reported	0.605153	0.612269

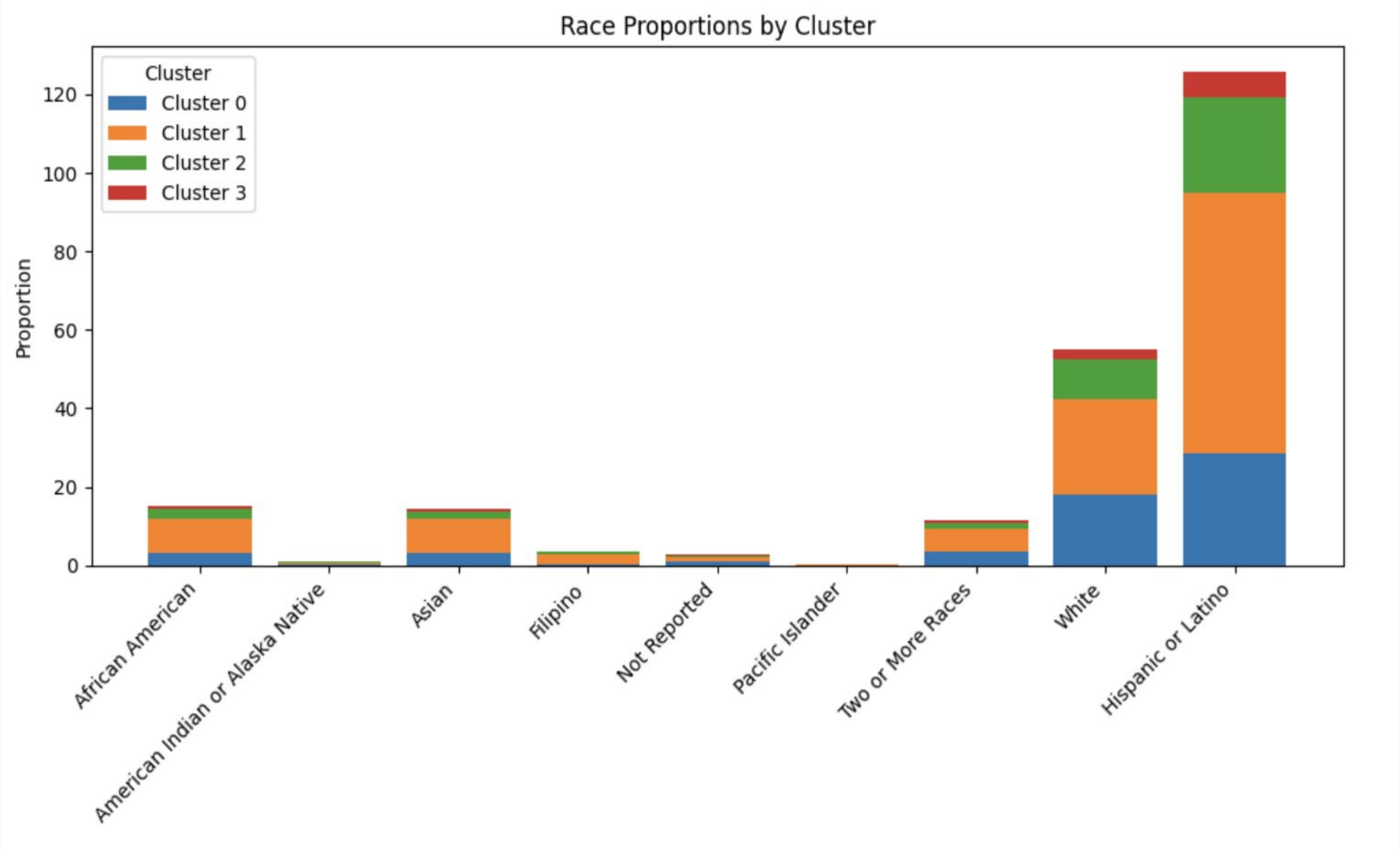

Figure 8: Bar chart showing different race proportions for each cluster

Results - Tukey's HSD

Results using Tukey's HSD to get pairwise differences between race proportions

Pairwise Comparison	RB = African American	RI = American Indian or Alaska Native	RA = Asian	RF = Filipino	RH = Hispanic or Latino	RP = Pacific Islander	RT = Two or more Races	RW = White	RD = Not Reported
0 - 1				0.0132			-0.0131	-0.0988	
0 - 2					0.1088		-0.0178		
0 - 3									
1 - 2						-0.0019			
1 - 3									
2 - 3									

Discussion/Setbacks

- The integration %s of Special Education Students indicate the extent of inclusion but further research is needed to test integration effectiveness, i.e. social development, learning speed, overall progress, etc.
- Although we've shown that there are significant differences in race proportions across clusters, it is not exactly clear how they exactly contribute to cluster separation
- Current results are limited to ethnicities. There are a lot of other factors that we could consider adding, such as urbanicity, disability category, gender, charter school status, etc.

Conclusion/Next steps

- Significant differences between clusters for of school districts for race proportions and program type exist, indicating a possible racial bias when integrating students in the regular classroom

- Latent variable models such as principal component analysis or factor analysis can help explain how much each variable contributes to each individual cluster

- Assumptions for ANOVA testing may be considered violated, so using a nonparametric version may be more appropriate

THANK YOU!