

## Last Week's Challenges





Faulty Data - Sampling Bias?



Including data in the repository

Current Research Question How do the metropolitan school districts in the US compare in demographic, social, and economic household conditions, and what factors play a leading role in those discrepancies?

Race Data vs our District Data Over the Entire US

	pct_pov	pct_SP	pct_HHVJ	pct_CC	pct_NCI	pct_CD	pct_CLI	pct_hisp_latinc	pct_white	pct_black	pct_native	pct_asian
pct_pov	1.00	0.62	0.15	0.35	0.62	0.27	0.19	0.22	-0.44	0.41	0.22	-0.22
pct_SP	0.62	1.00	0.26	0.22	0.38	0.26	0.10	0.16	-0.46	0.50	0.23	-0.15
pct_HHVJ	0.15	0.26	1.00	0.04	-0.01	0.11	0.07	0.12	-0.15	0.11	-0.04	0.00
pct_CC	0.35	0.22	0.04	1.00	0.37	-0.04	0.53	0.59	-0.58	0.00	0.33	0.08
pct_NCI	0.62	0.38	-0.01	0.37	1.00	0.13	0.21	0.16	-0.30	0.22	0.30	-0.23
pct_CD	0.27	0.26	0.11	-0.04	0.13	1.00	-0.09	-0.06	0.02	0.09	-0.01	-0.15
pct_CLI	0.19	0.10	0.07	0.53	0.21	-0.09	1.00	0.67	-0.52	0.03	-0.06	0.19
pct_hisp_latino	0.22	0.16	0.12	0.59	0.16	-0.06	0.67	1.00	-0.71	-0.03	-0.07	0.09
pct_white	-0.44	-0.46	-0.15	-0.58	-0.30	0.02	-0.52	-0.71	1.00	-0.49	-0.29	-0.25
pct_black	0.41	0.50	0.11	0.00	0.22	0.09	0.03	-0.03	-0.49	1.00	-0.08	-0.01
pct_native	0.22	0.23	-0.04	0.33	0.30	-0.01	-0.06	-0.07	-0.29	-0.08	1.00	-0.06
pct_asian	-0.22	-0.15	0.00	0.08	-0.23	-0.15	0.19	0.09	-0.25	-0.01	-0.06	1.00

Additional data to possibly look at

Retention rates

Tenure of teachers

Graduation rates (proxy for quality of education?)

Household income

School funding

## Difficulties & Next Steps

Look into / decide on what cities we want to focus on for our research question. (& are we going to do this manually?)

Research what data is out there on each school districts located in the chosen cities.

Finding market race percentages (if districts span multiple counties etc.)

Joining race data with chosen school districts/cities to start analysis.

Look into cloud storage for data or writing scripts to unzip large files.