Katrina Wheelan

December 11, 2020

Thesis Progress Report

**Summary**

This semester, I focused on reading background material, acquiring the appropriate data sources for my analysis, cleaning the data, and doing preliminary investigations into the data. I include here an introduction (from my proposal), some summary statistics from my initial analyses, next steps, and a tentative outline.

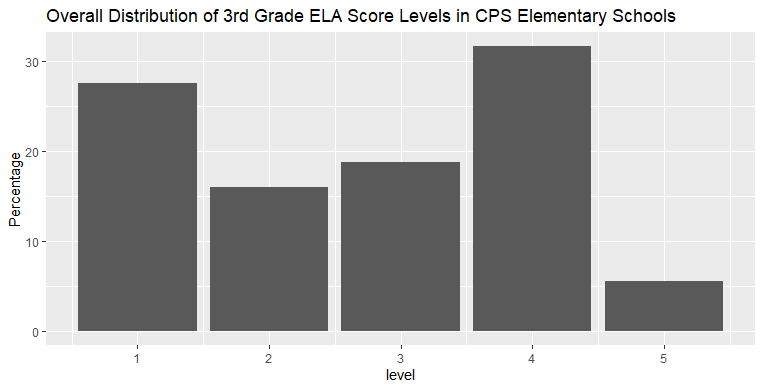
**Introduction**

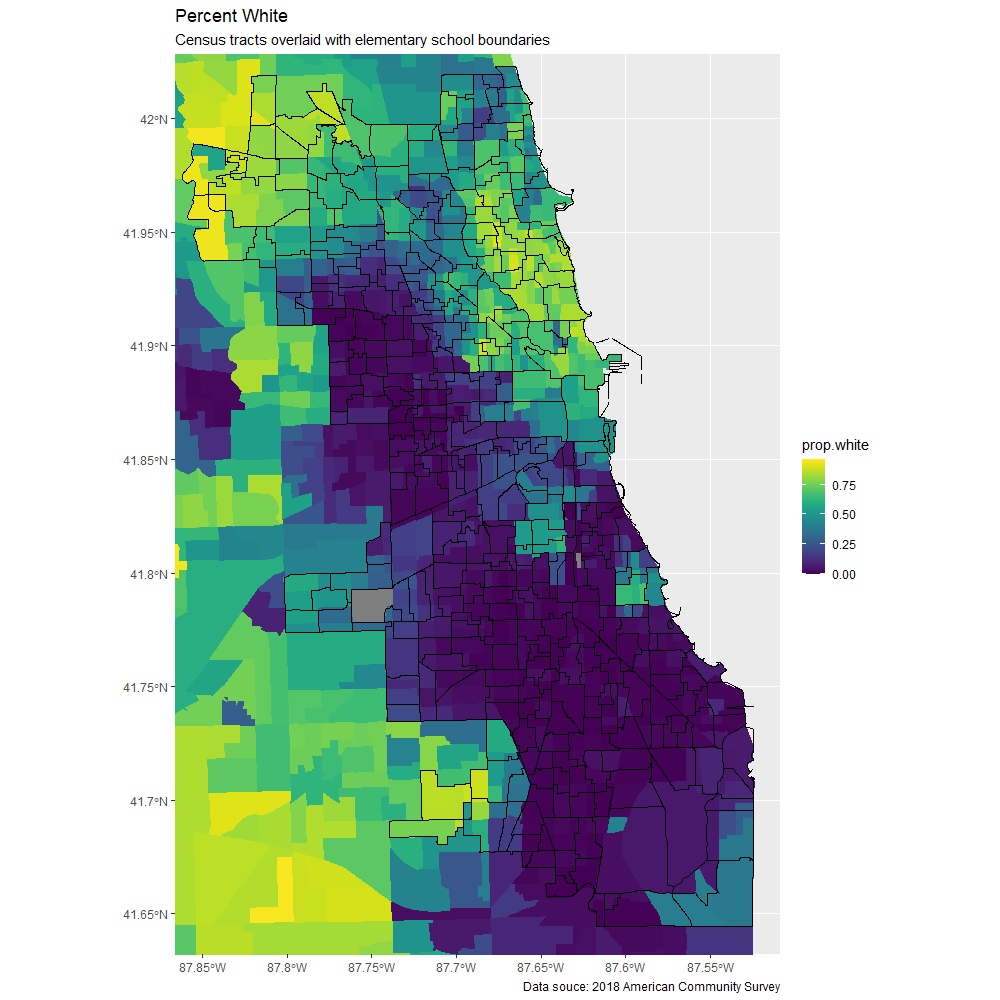
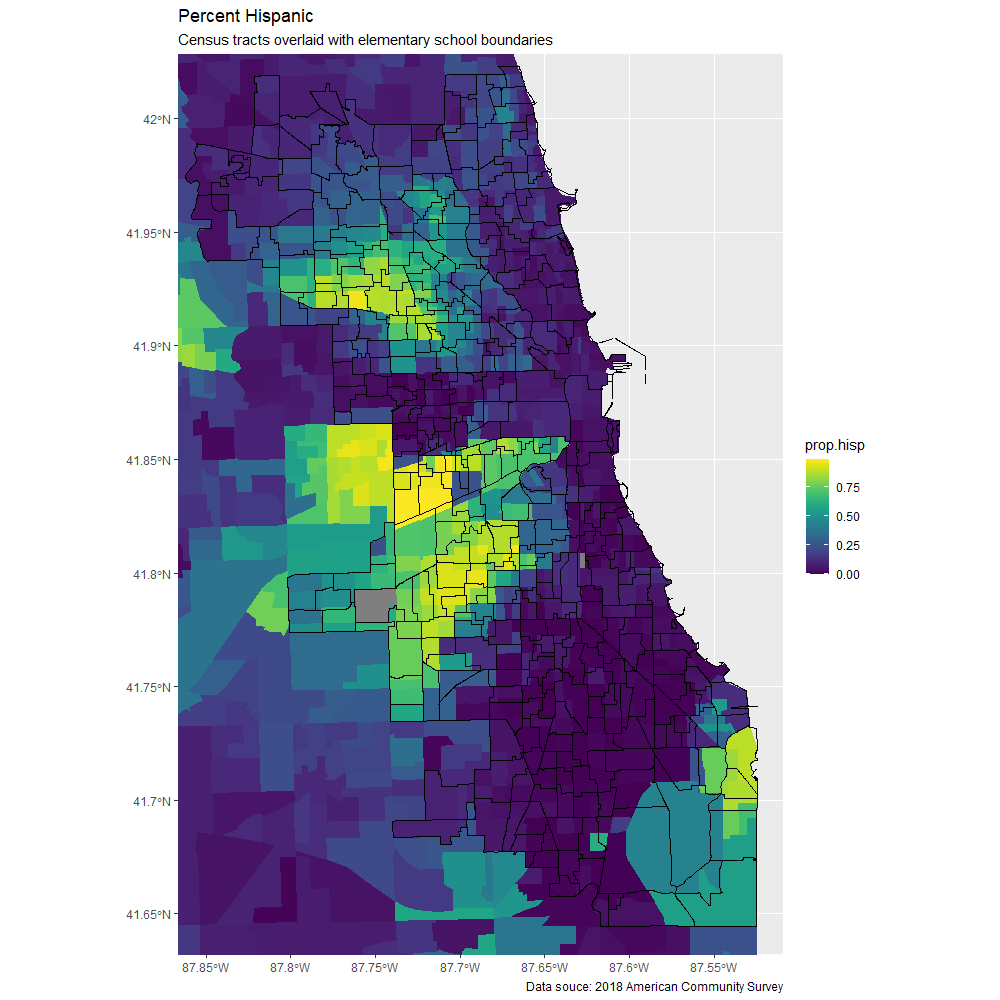
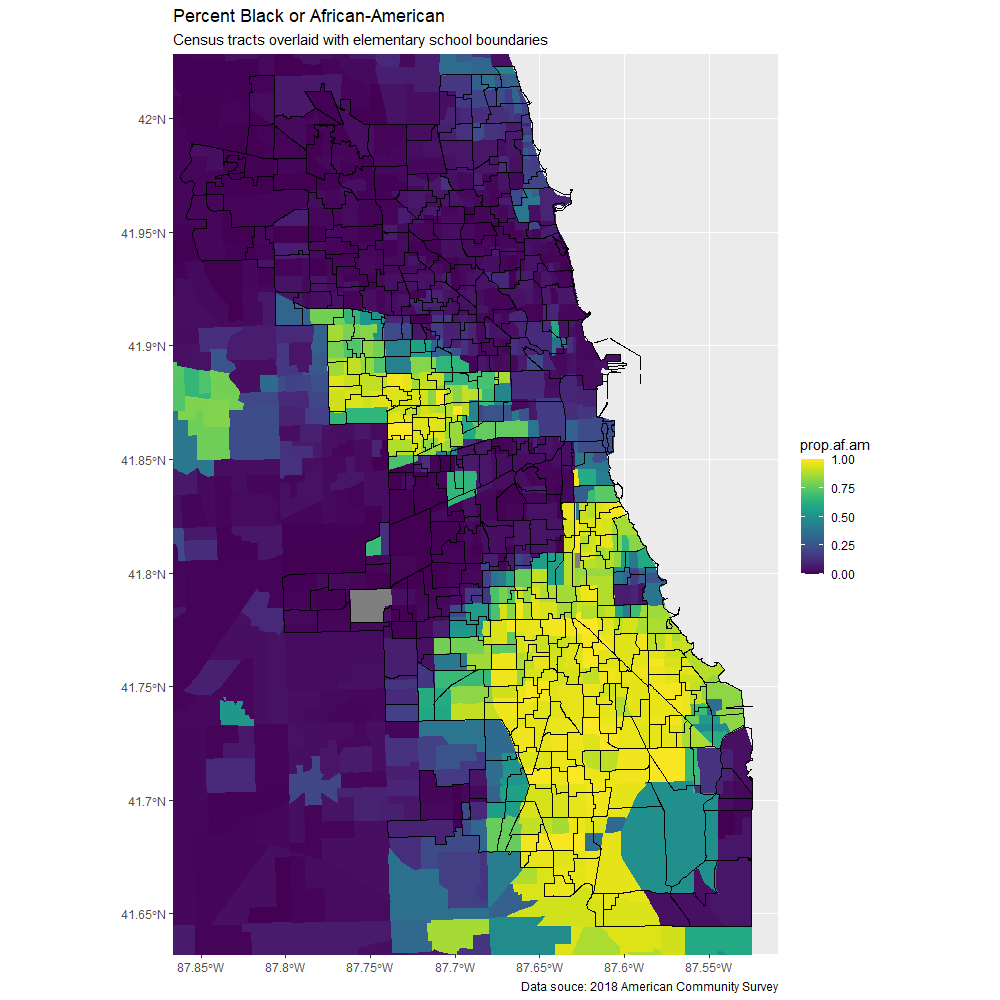
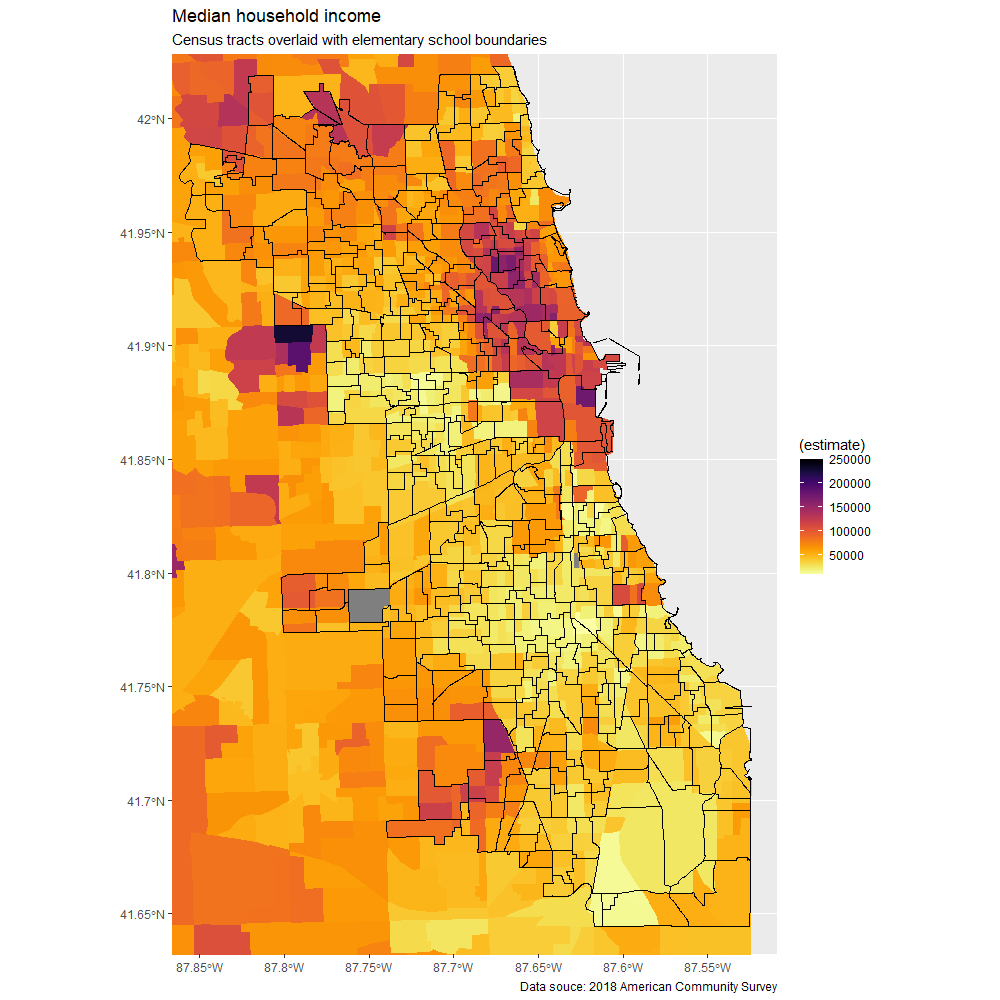
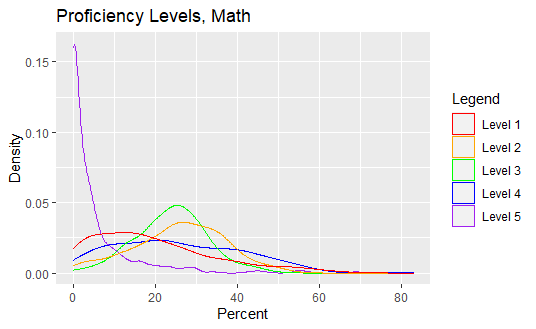
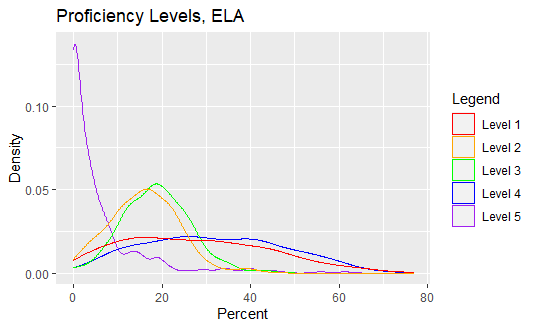
Social scientists have done extensive research on the effects of wealth and racial disparities between districts on lifetime outcomes, but there has been comparatively little research on how intra-school disparities affect educational outcomes. I plan to focus my thesis on precisely these distributional effects. I am interested in the extent to which the income and racial make-ups of schools and the districts they serve affects the distribution of outcomes over time. In essence, I want to know whether traditional factors of positive educational outcomes -- like an influx of wealth to a neighborhood -- masks adverse *distributional* effects. I chose to focus on Chicago Public Schools because it comprises a single school district and because there are readily available and extensive data. Chicago schools are also notorious for their racial homogeneity; in more than half of Chicago public elementary schools, at least 90% of the students identify as the same race.[[1]](#footnote-0)

To answer these questions, I am using several high dimensional data sources. I have access to the Illinois Department of Education data for each public elementary school in Chicago. These data include detailed demographic data, standardized testing scores, and other indicators at the school-level. I also have access to the American Community Survey data, which provides detailed metrics about the characteristics of each census tract. I will employ several spatio-temporal methods to analyze the effects of racial and income disparities on the distribution of testing outcomes across space (school zones) and time (school years).

**Descriptive Statistics in Graphs**

(Note: ELA means English Language, for the reading portion of the Illinois standardized test.)

****

****

**Next Steps**

I’ve hit two major limitations. On the explanatory side, the Chicago elementary school districts do not align perfectly with U.S. Census tracts. I am working on developing an appropriate method to weight demographic data from the American Community Survey (which are at the Census tract-level) to get a good estimate of the underlying demographics of individual elementary school districts. It likely doesn’t make sense to weight by geographic area because we don’t expect homogeneity across census tracts, especially near school district boundaries. I am looking into finding finer, block-level demographic data. I also plan to use elementary school demographics as an approximate check on district demographics (while keeping in mind that private school enrollment may drive a wedge between school and district demographics in some areas).

The second limitation is on the response side. For privacy reasons, the Illinois Department of Education does not publish individual test score data. The most detailed information is the proficiency level distributions within various student subsets. There are five proficiency levels, determined by test score thresholds, and the bins are not evenly spaced. Since this data is ordinal, it will be difficult to assess spread. Currently, I am considering using a single metric like the ratio of students in the lowest proficiency level to those in the highest two levels. I may also try random sampling to simulate continuous testing data. In January, I will speak to Professor Klingenberg for general advice on handling ordinal data and for specific advice on measuring spread in ordinal datasets.

After addressing these issues, I will also have to decide how to express other indicators of interest. Currently, I plan to calculate and use the GINI index to measure economic inequality. I am looking into a similar metric for categorical distributions to quantify the racial heterogeneity of schools and districts. Then, I will be able to run initial models.

**Outline:**

Introduction

Brief Review of Literature

Note on Data Sources

Summary Statistics

Discussion of the Model

Analysis

Results

Conclusions

1. My analysis of Illinois Department of Education data. Data fetched from <https://www.isbe.net/Pages/Illinois-State-Report-Card-Data.aspx> [↑](#footnote-ref-0)