Madison K-12 Students Distribution Report

Team member: Zhuocheng Sun, Yueyu Wang
Instructor: Tyler Carazar-Harter

Introduction:

When the COVID-19 pandemic hit Madison, classes at the Madison Metropolitan School District became entirely online. Students without reliable home access to the internet are less able to participate in their classes and are at risk of falling behind their peers. Therefore, to ensure more K-12 students could have stable access to the Internet, residential internet assistance becomes necessary. In order to deploy the resources most effectively, we need to identify and reach the target population. In this project, we would provide probabilistic estimates of the residential addresses at which internet access could be provided to have the greatest effect.

Internet access is just one use case of knowing where all the K-12 students live in Madison. There are many more scenarios in which knowing where K-12 students are located is very helpful. Therefore, in this report, we mainly focus on different aspects of the places that K-12 students live in, and use the problem of Internet access as an example.

Data source:

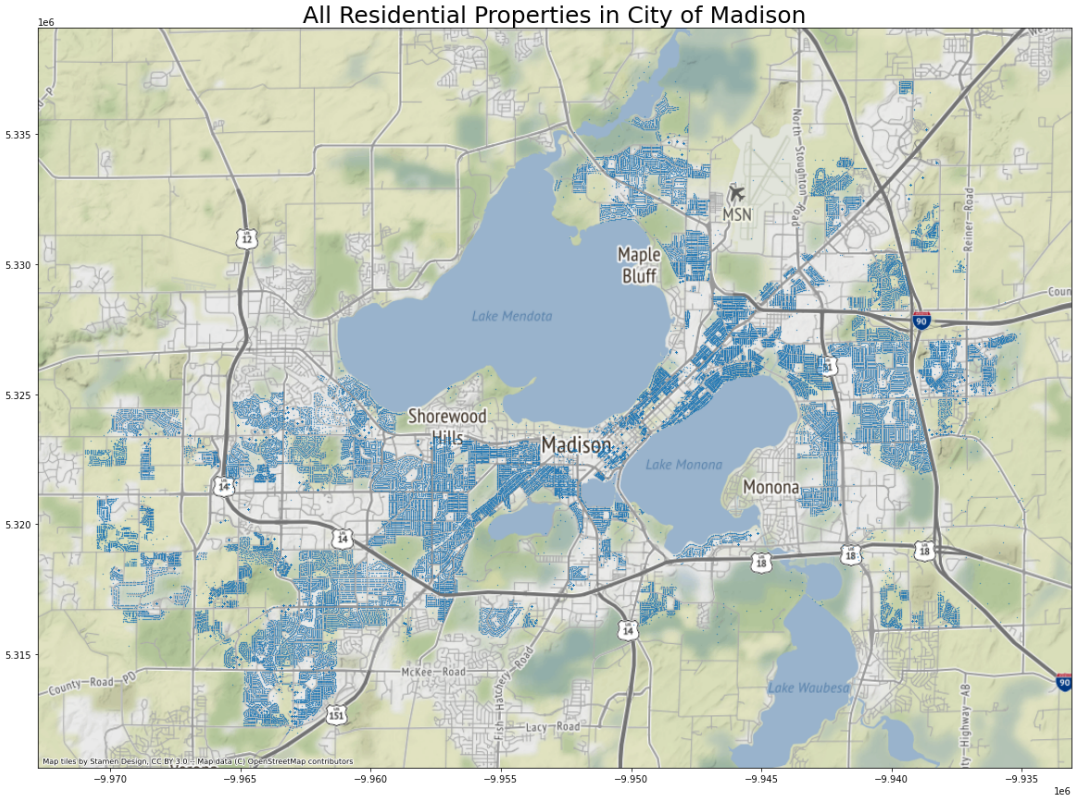
- City of Madison tax parcels: This dataset lists every property in the City of Madison, what Census tract it is in, its address, and, if it is a multifamily dwelling, how many units it contains.
- Census Data(2010 Decennial data, P12 table): There is a broad array of Census data. This can provide helpful information, such as the rate of internet access per track, number of school-age children per tract and household, the incidence of subsidized housing, and household income.

Data processing:

By definition, K-12 students include students from kindergarten for 5 to 6 years old through twelfth grade for 17 to 18 years old, but 18 year-olds could be either twelfth-grade students or college freshmen. From the census data we got, 18 and 19 year-olds are in the same column and we couldn't separate them out. Since we found many addresses with majorities of 18-19 year-olds, and many of those properties are located near UW Madison campus, we think it makes more sense to regard them as college students rather than 12th-grade students. Thus, to exclude the above scenarios, we only count 5 to 17 year-olds as K-12 students in this analysis.

Methods:

Then we use the GeoPandas package to do a spatial join between the census data and property data in tax parcels. Here is what it looks like (we use a small blue dot to represent each building/house):



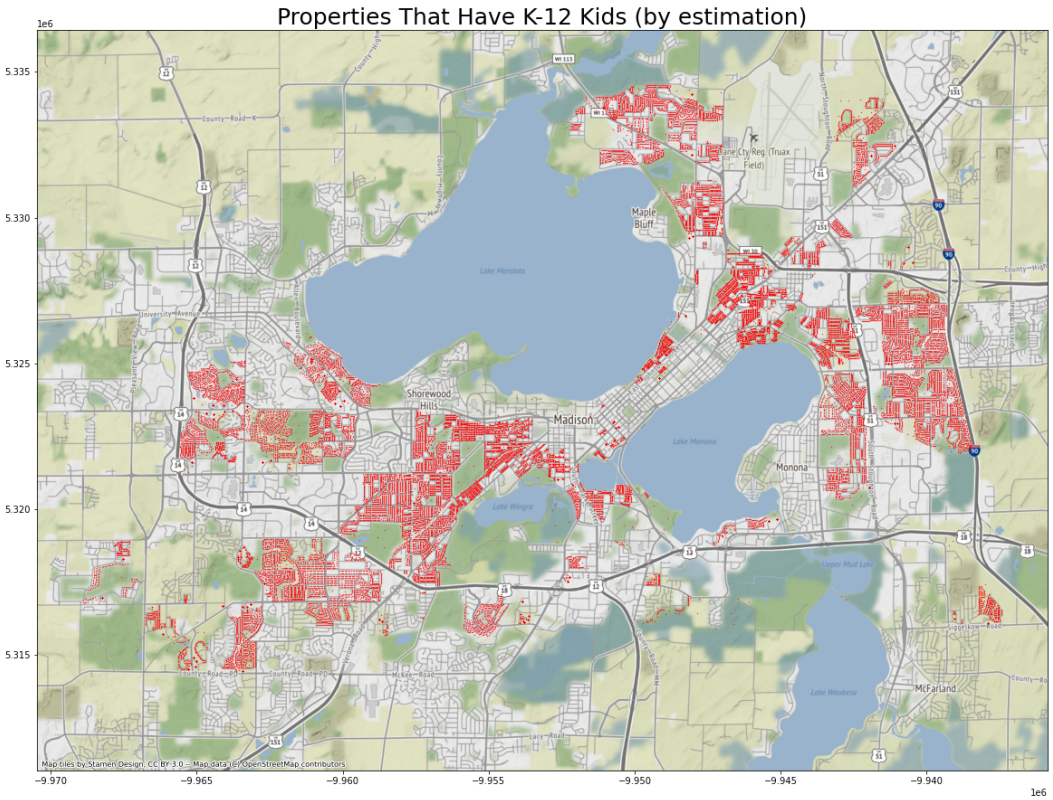
Estimate the number of K-12 children in each property:

We tried several different approaches to estimate the number of K-12 students in each property. Using linear regression did not perform very well because most fields we have don't have a correlation with the number of K-12 students. It turns out that the number of bedrooms in a property is the most indicative field that we have to estimate the number of K-12 children in each property. Since we have the census data per tract, we can estimate the number of K-12 students in each property using:

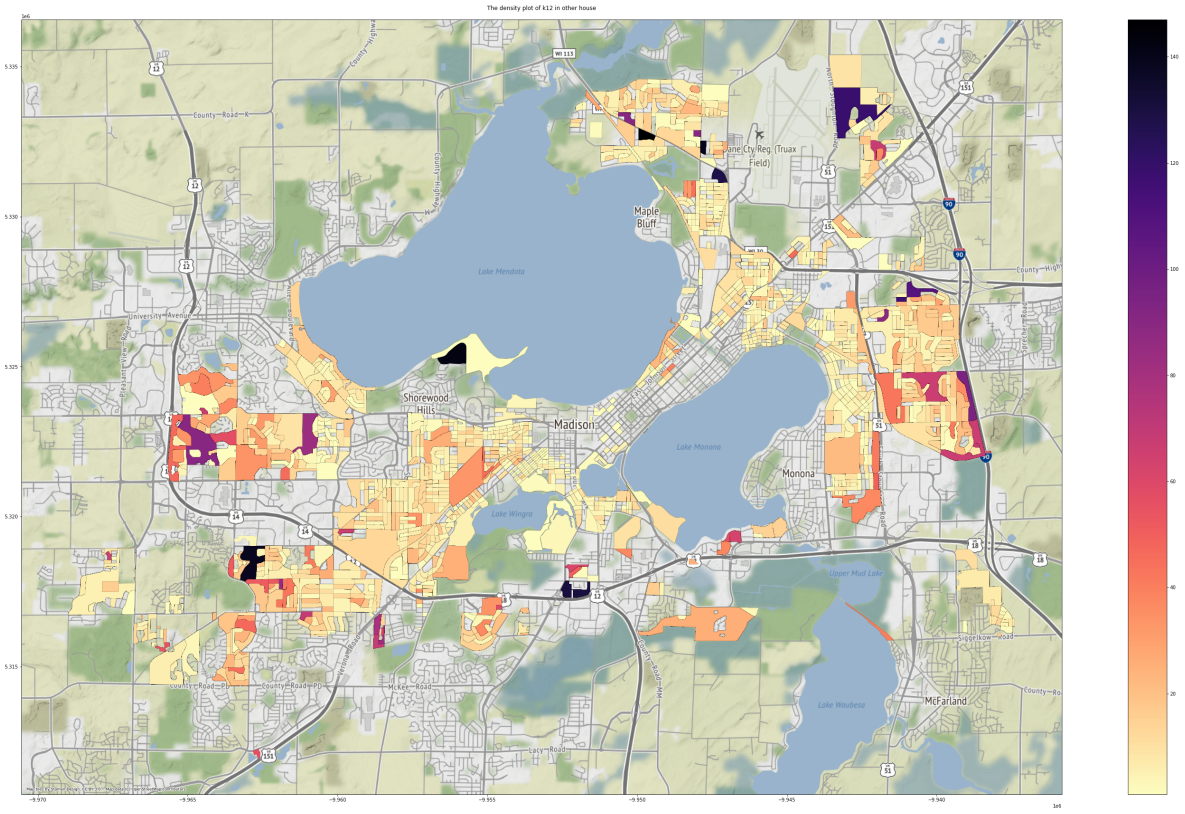
$$K-12_{est} = \frac{(\#\ \text{bedrooms in a property}) * (\#\ \text{K-12 children in the tract})}{\#\ \text{bedrooms in the tract}}$$

Results and plots:

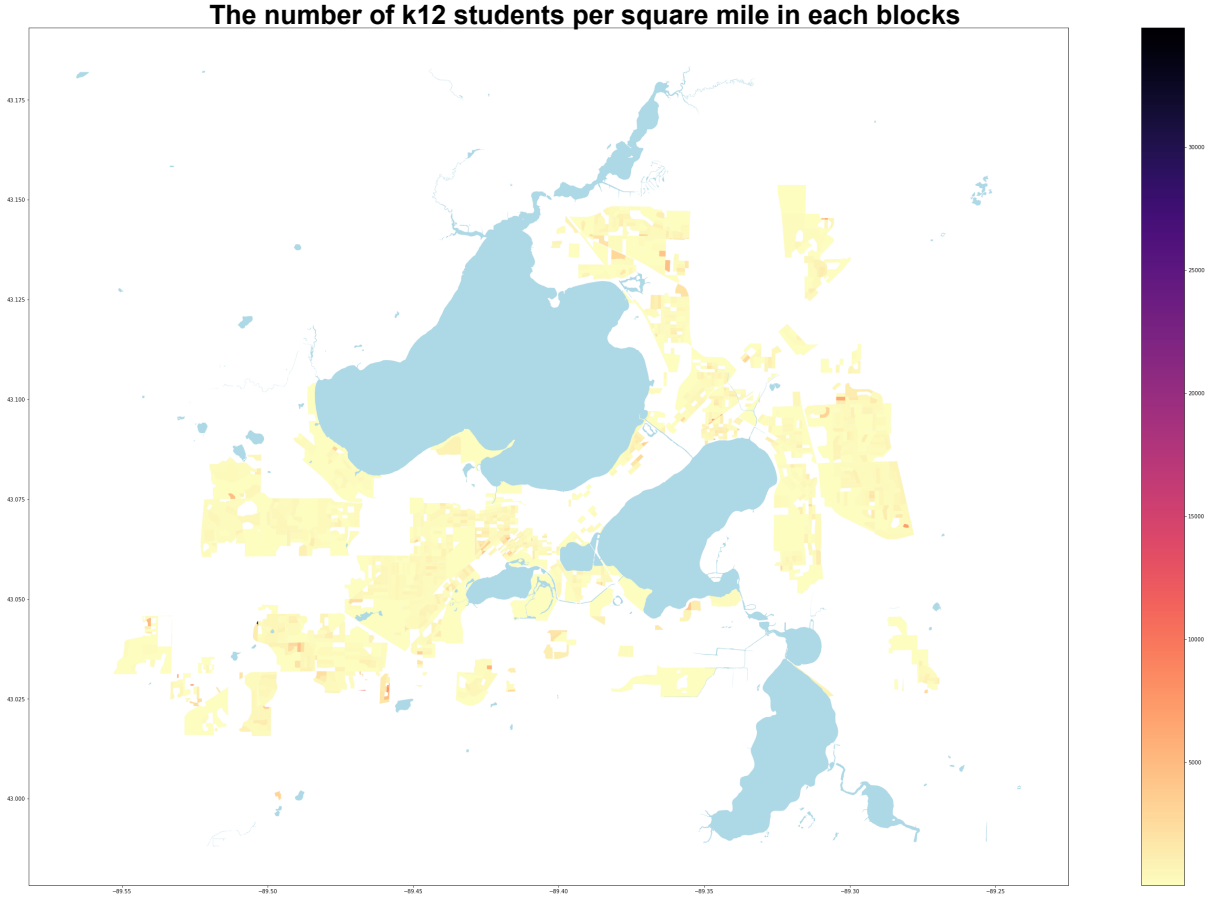
After we estimated the number of K-12 students in each unit, we summarized the total number of K-12 students in different buildings. Then, we use the estimated number of K-12 students to draw the following plot that shows the buildings/houses that might have K-12 students.



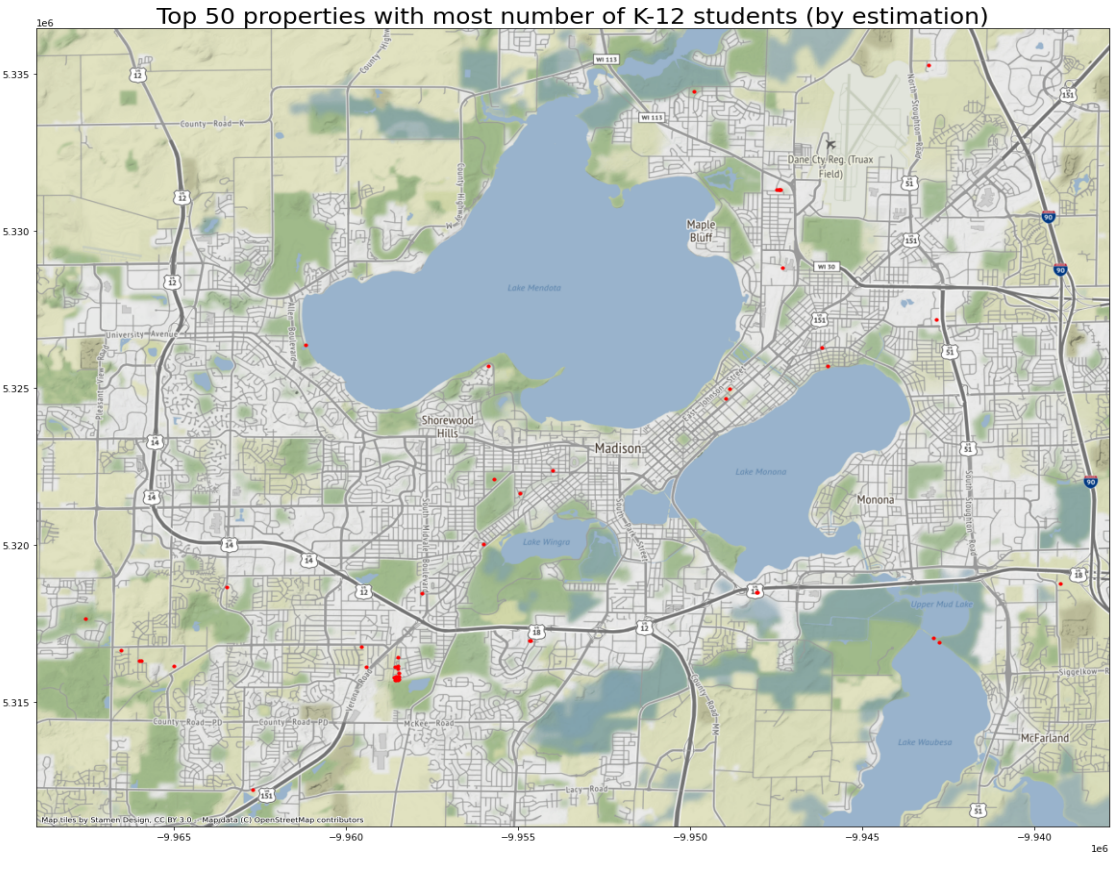
Then, we use different colors to differentiate the number of K-12 Kids in each census block. The darker the block is, the more K-12 students in that block. In this way, we could get a general idea of which block has more K-12 students.



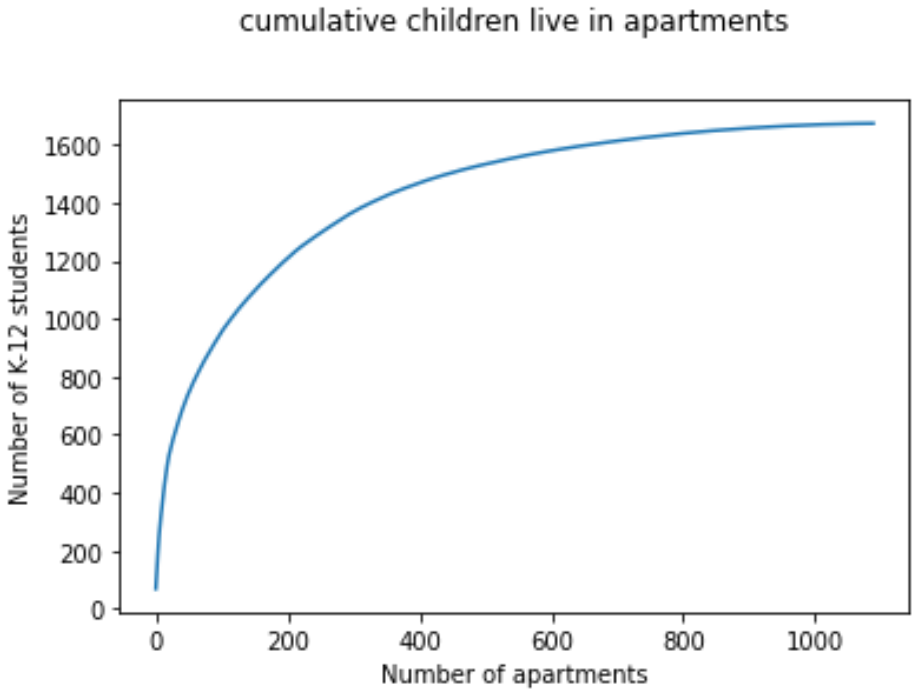
The number of k12 students per square mile in each blocks



Governmental assistance funds usually are limited. It might be too expensive to provide stable internet access to all the dark regions in the above plot. To make governmental assistance more effective, we select 50 buildings with the most number of K-12 students by our estimation.

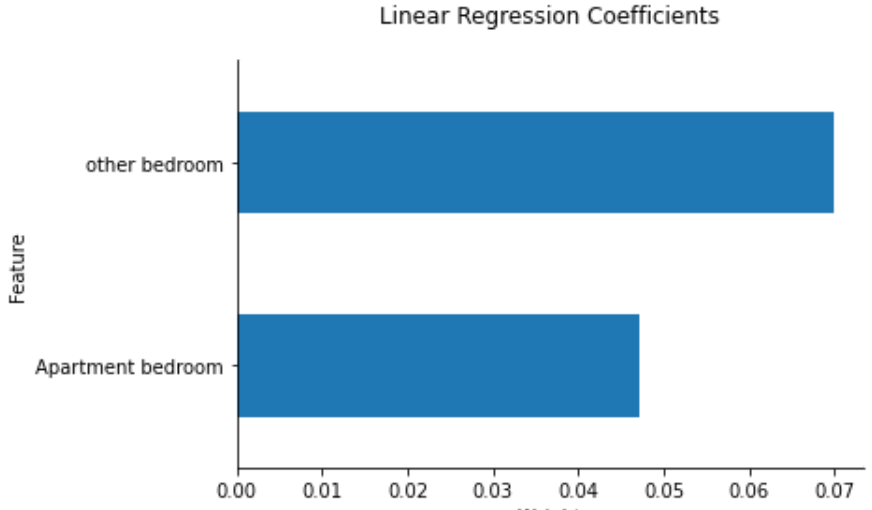


Another interesting question we want to find out is whether K-12 childrens are gathered in some specific apartments, or are they evenly distributed among all apartments. To answer this question, we generate a cumulative plot, where the X-axis represents the number of apartments, and the Y-axis represents the cumulative K-12 students living in those apartments (by estimation).



Based on the plot above, we can see that the cumulative line has many curves, and there are some apartments that have remarkably more K-12 students than other apartments. Therefore, K-12 students are not evenly distributed among these apartments and it is important for us to know which apartments are. So we extracted the top 20 apartments with the most number of K-12 students into a csv file: Top20_Apartments.csv.

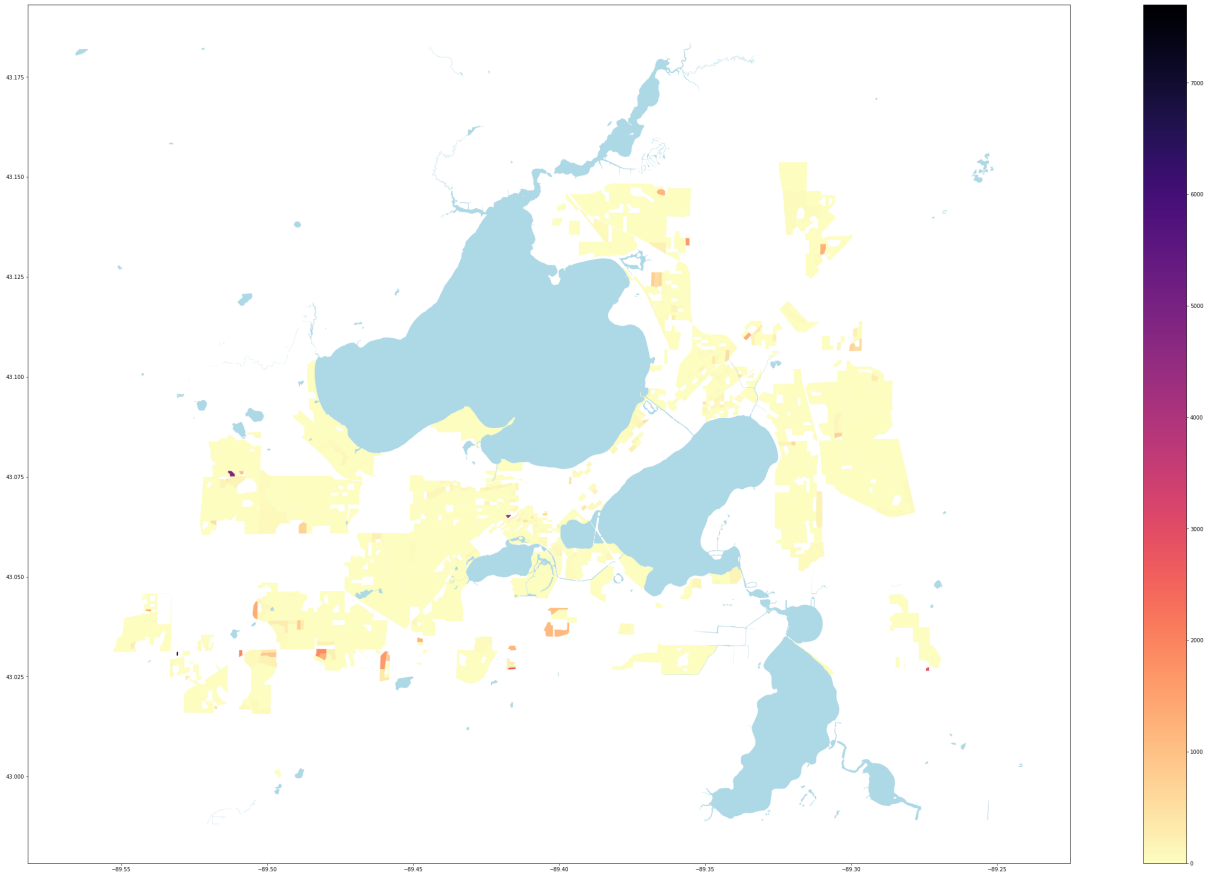
To find out whether these K-12 students are more likely to live in apartments or houses, we apply a linear regression model (y is the number of K-12 students, x1 is the number of apartment bedrooms, x2 is the number of other bedrooms). The coefficients of x1 and x2 we got are as follows:



According to the coefficient plot, we find out that the coefficient of “other bedrooms” is larger than the coefficient of Apartment bedrooms, which means that K-12 students are more likely to live in a bedroom in houses than a bedroom in apartments. Therefore, it would make more sense to estimate the number of K-12 students in each bedroom based on the coefficients we got.

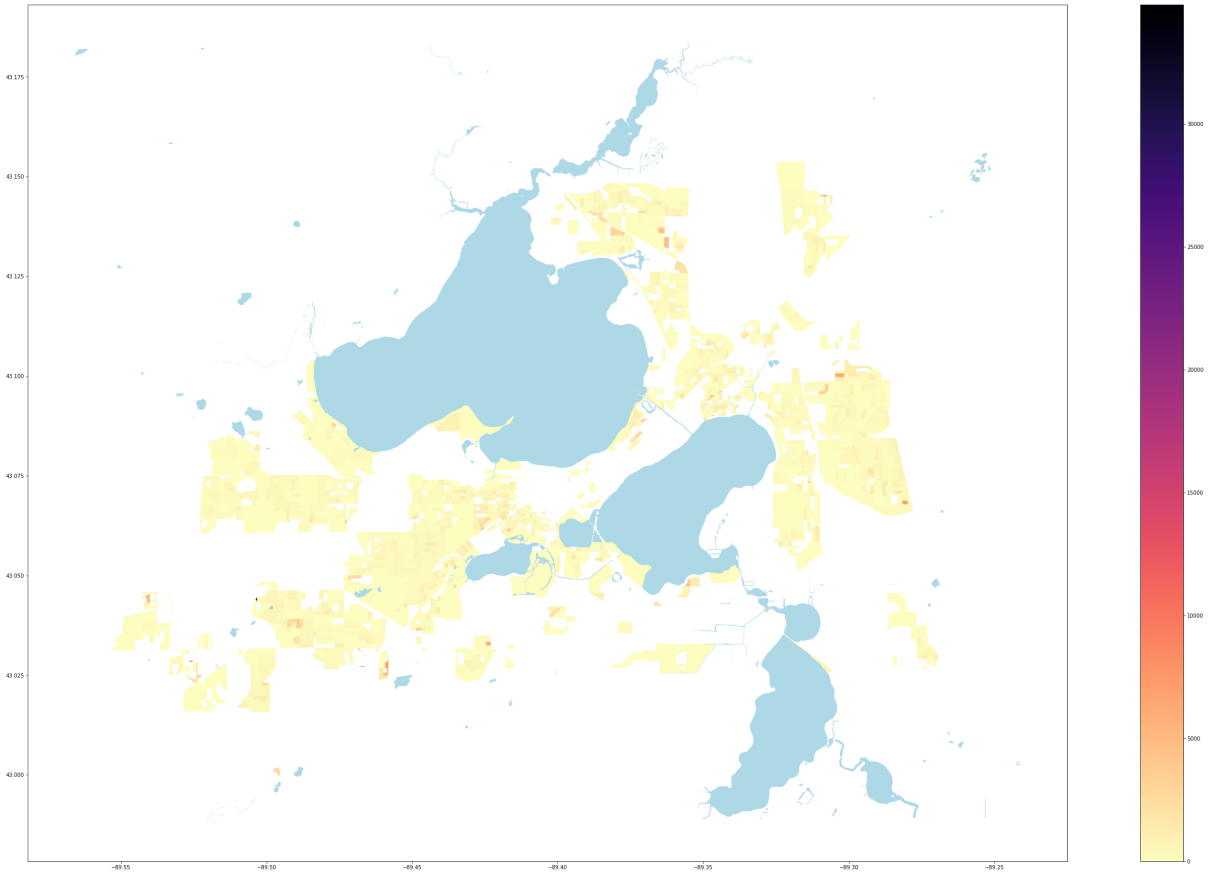
After we updated our estimation, we made several density plots to show the density of K-12 students who live in apartments to discover how many students per square mile. According to the plot, we can find that apartments in some blocks have relatively high K-12 students density, which matches with the cumulative plot above.

The density plot of k12 in apartment



We also draw the plots to show the number of K-12 students who live in non-apartments per square mile.

The density plot of k12 in non-apartment properties



According to the density plot of K-12 students in non-apartments in each block and the density plot of K-12 students in apartments in each block, we can find that K-12 students who live in houses per square mile are very similar. Some blocks have higher density of K-12 students living in non-apartments, but the difference is not as significant as that in the density plot of K-12 students living in apartments.

Problems:

Theoretically, we could apply our model to larger areas and generate similar reports. But there are some problems in the analysis that are worth mentioning. First of all, the related data we have is very limited, and we are just using the number of bedrooms to estimate. If we could have more data related to this topic, we could try different models and possibly get a better estimation. Some data entries have a value of 0 for the number of bedrooms. We add a small value of 0.01 to all #bedroom fields to make it bigger than 0. However, predictions for these properties would be off since the number of bedrooms is not shown correctly. Another problem is that K-12 students are constantly changing. The plots we draw might be totally different by using the 2020 census data (we used the 2010 census data).