Madison Internet Access Analysis Project 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.

Data source:

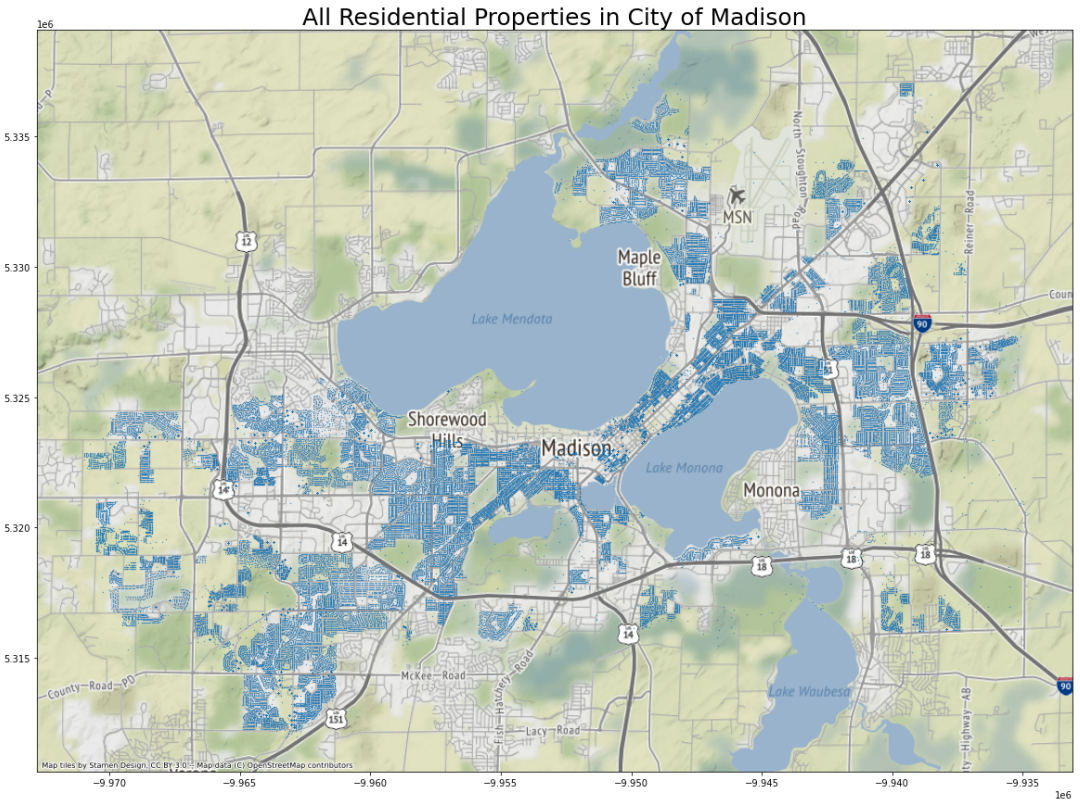
- City of Madison tax parcels: This dataset lists every property in the City of Madison, what Census tract it is in, its address, and, if it is a multifamily dwelling, how many units it contains.
- Census Data: There is a broad array of Census data. This can provide helpful information, such as the rate 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):



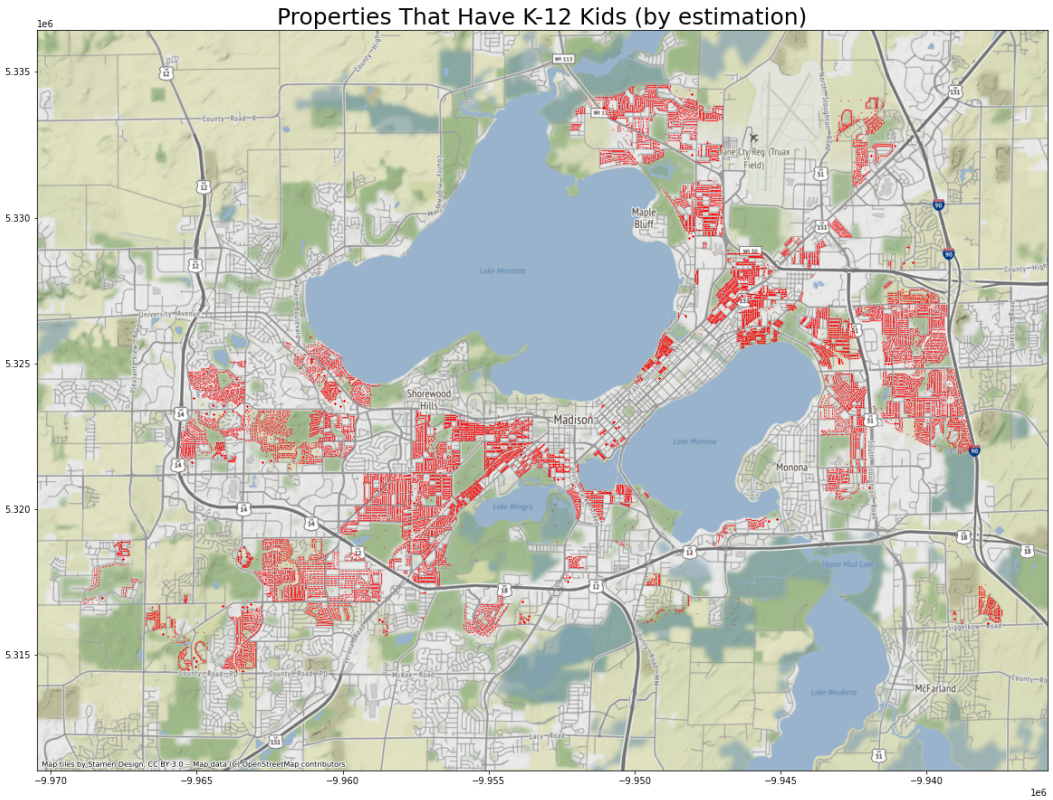
Estimate the number of K-12 children in each property:

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:

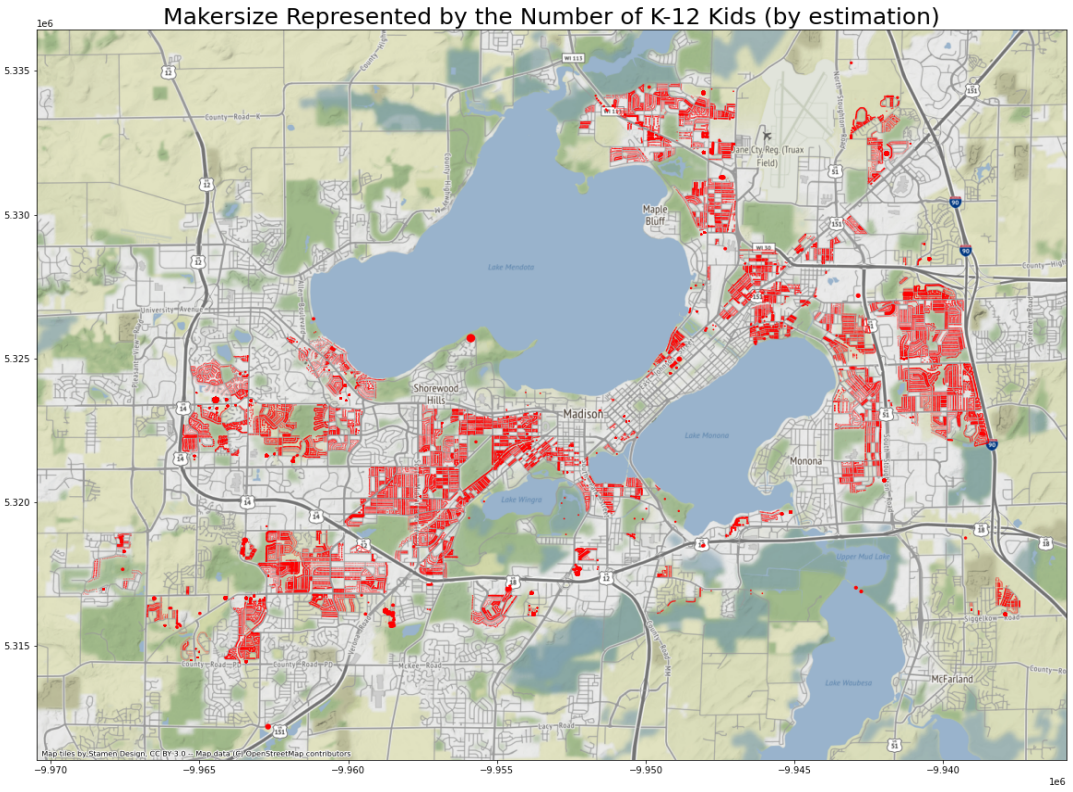
$$K-12_{est} = \frac{(\#\ \text{bedrooms in a property}) * (\#\ \text{K-12 children in the tract})}{\#\ \text{bedrooms in the tract}}$$

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.

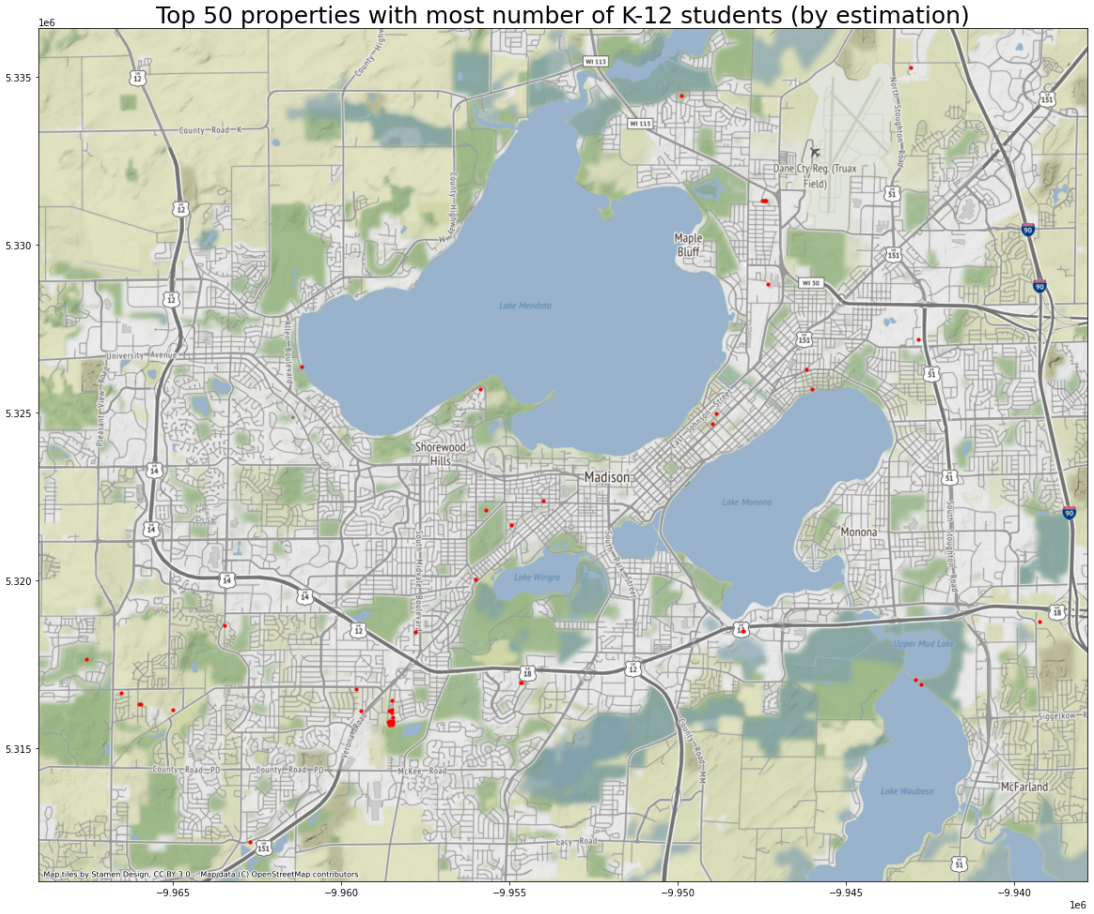


Then, we use different marker sizes to differentiate the number of K-12 Kids in each building. The larger the red marker is, the more K-12 students in that property. In this way, we could get a general idea of which region has a denser population of K-12 students.



Governmental assistance funds usually are limited. It might be too expensive to provide stable internet access to all the “dense” red 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.



According to the above plot, we can see that most red dots are not adjacent to places that have free wifi open to the public, such as Madison Public Library, College Library, etc. Therefore, it would be challenging for these students to continue their studies during the COVID-19 pandemic. Moreover, the Madison area could have days with extreme weather in the winter, and teachers might choose to teach remotely since they have had this experience before during the pandemic. Hence, setting up stable internet connections for these places is crucial even if we don't have a pandemic.

Problems

Theoretically, we could apply our model to larger areas and generate similar reports. But there are some problems in the analysis that is 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 census data years later (we used the 2017 census data).