Madison K-12 Students Distribution Report

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

Introduction:

When the COVID-19 pandemic hit Madison, WI, classes at the Madison Metropolitan School District became entirely online, and students continued to have virtual classes in early 2021.[1] Unfortunately, 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, residential Internet assistance becomes necessary to ensure more K-12 students could have stable access to the Internet. In order to deploy the resources most effectively, we need to identify and reach the target population. This project provides estimates of the residential addresses at which Internet access could be provided to have the greatest effect.

The city of Madison has a population of 233 thousand (2010 census), and there are about 27 thousand children in K-12 age. Among the 27 thousand K-12 students, about 4200 of them live in apartments, while others live in non-apartment properties like houses. The number of bedrooms in a property is a great indicator of the number of K-12 students living in that property. There are about 130 thousand apartment bedrooms (70% of them are based on our estimation) and 169 thousand non-apartment bedrooms (1% of them are based on our estimation) in the City of Madison.

Internet access is just one use case of estimating where all the K-12 students live in Madison, WI. 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 K-12 students live in and use the problem of Internet access as an example to illustrate.

Data Sources:

City of Madison tax parcels: This dataset lists properties 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): A broad array of Census data. This dataset 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.

Dane county Shapefiles: These Shapefiles store the geometric location and attribute information of geographic features of Dane county at the block level.

---

[1] WKOW, *Madison schools announce all-virtual start to third quarter*, Jan 2021, https://www.wkow.com/coronavirus/madison-schools-announce-all-virtual-start-to-third-quarter/article_799ec314-260f-58a2-a1a7-c3a993d6d6fe.html

Exclude 18-year-olds to Avoid Misrepresentation:

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

Data Imputation for 'Bedrooms' in Commercial Property Class:

1164 properties in the Commercial property class are apartments or have at least some apartments inside that property, but they all miss the 'Bedroom' data.

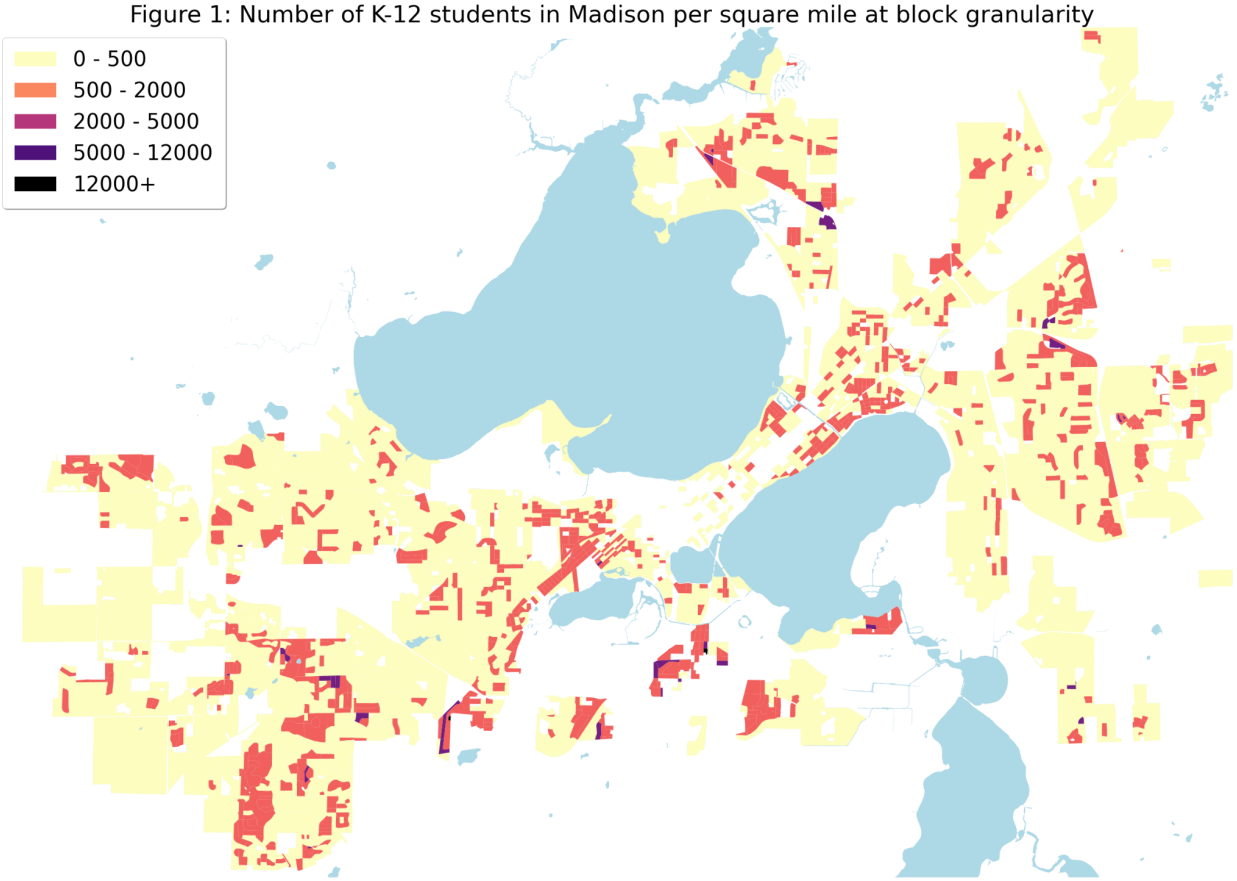
1. For properties that have a number of units information: We calculated the average number of bedrooms per unit (1.9503 bedrooms per unit) using the 'Bedrooms' data from apartments in the Residential property class. Then we multiplied the number of units with the average number of bedrooms per unit to fill in the missing 'Bedrooms' field.
2. For properties without units: We divided 'NetTaxes' by the number of bedrooms for all the apartments in the residential property class, and used the average number (1493.3027 net tax per bedroom) to calculate 'Bedrooms' data for those properties using their 'NetTaxes.'

Estimate the Number of K12 Students in Each Property:

We first plotted the census data of K-12 students using geometry information. Different blocks have various sizes. Therefore, a big block with many K-12 students does not necessarily mean it has a denser population because of its size. Therefore, we plot each block based on the number of K-12 students per square mile.

In Figure 1, we use different colors for various ranges of numbers of K-12 Kids in each census block per square mile. The darker the block is, the higher the K-12 student population density in that block. In this way, we could get a general idea of which block has a denser K-12 students population. We can see in Figure 1 that most census blocks are colored yellow or orange, so they have similar K-12 students population densities.

Figure 1: Number of K-12 students in Madison per square mile at block granularity



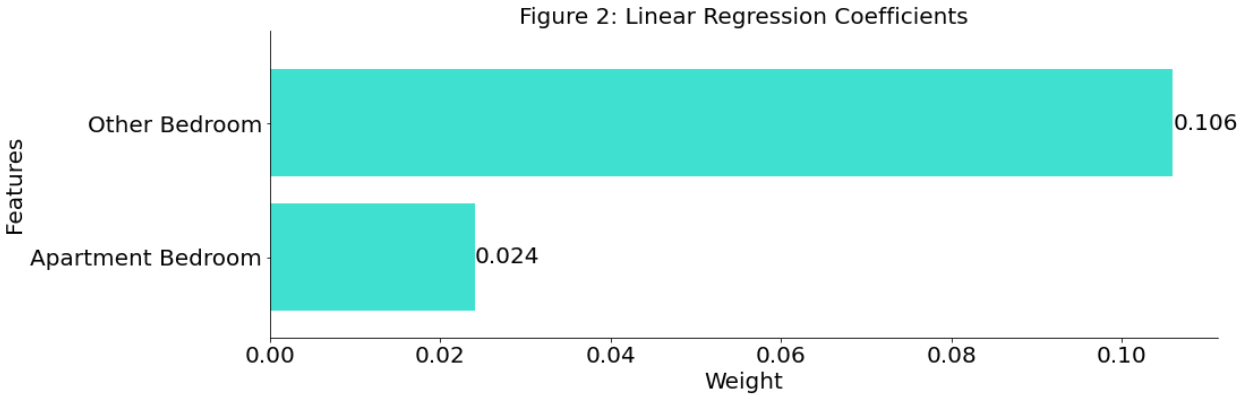
Estimate the number of K-12 children in each property:

We have a variety of fields for each property in our datasets. However, most fields in the dataset cannot help us estimate the number of K-12 students in that property. We found the number of bedrooms and the living area in a block has a positive correlation with the number of K-12 children in the block.

Since a large living area usually means more bedrooms, using both of them to estimate is redundant. So we decided to use the number of bedrooms in a property, the most indicative field we have, to estimate the number of K-12 children in each property.

However, a bedroom in an apartment and a bedroom in a non-apartment property are not the same. Therefore, we need to have different estimates for the number of K-12 children depending on the property type. There are about 299 thousand bedrooms in our dataset (130 thousand apartment bedrooms, 169 thousand house bedrooms), so differentiating apartment bedrooms and non-apartment bedrooms in estimation are vital for our estimation.

To determine whether K-12 students are more likely to live in apartments or non-apartments, we apply a linear regression model (Response variable is the number of K-12 students per square mile. Two explanatory variables are the number of apartment bedrooms at block granularity and the number of non-apartment bedrooms at block granularity).



According to Figure 2, we find out that the coefficient of “other bedroom” (non-apartment bedrooms) is larger than the coefficient of “Apartment Bedroom,” which means that K-12 students are less likely to live in a bedroom in apartments than a bedroom in other properties like houses. The coefficient of “Apartment bedroom” is 0.024, and the coefficient of “other bedroom” is 0.106. The R-squared for the linear regression model is 0.28. These coefficients tell that, on average, there is one more K-12 child for every 40 more non-apartment bedrooms and one more K-12 child for every increment of 10 apartment bedrooms. Therefore, it would make more sense to estimate the number of K-12 students in each bedroom based on linear regression coefficients. Since we know the exact number of K-12 children in each block from census data, we apply the ratio between two coefficients to improve our estimates (if we directly multiply the coefficients with the number of bedrooms for each type, the total number of K-12 children in each block would not be identical to that from census data).

Therefore, the formula to estimate the number of K-12 students in each property is:

R : ratio, $0.106/0.024 = 4.38$

K_b : number of K12 students in block b

B_a : number of apartment bedrooms in block b

B_n : number of non-apartment bedrooms in block b

B : number of bedrooms in the target property

Estimate for an apartment property:

$$Apartment_{k12} = K_b * \frac{B}{B_a + R * B_n} \quad (1)$$

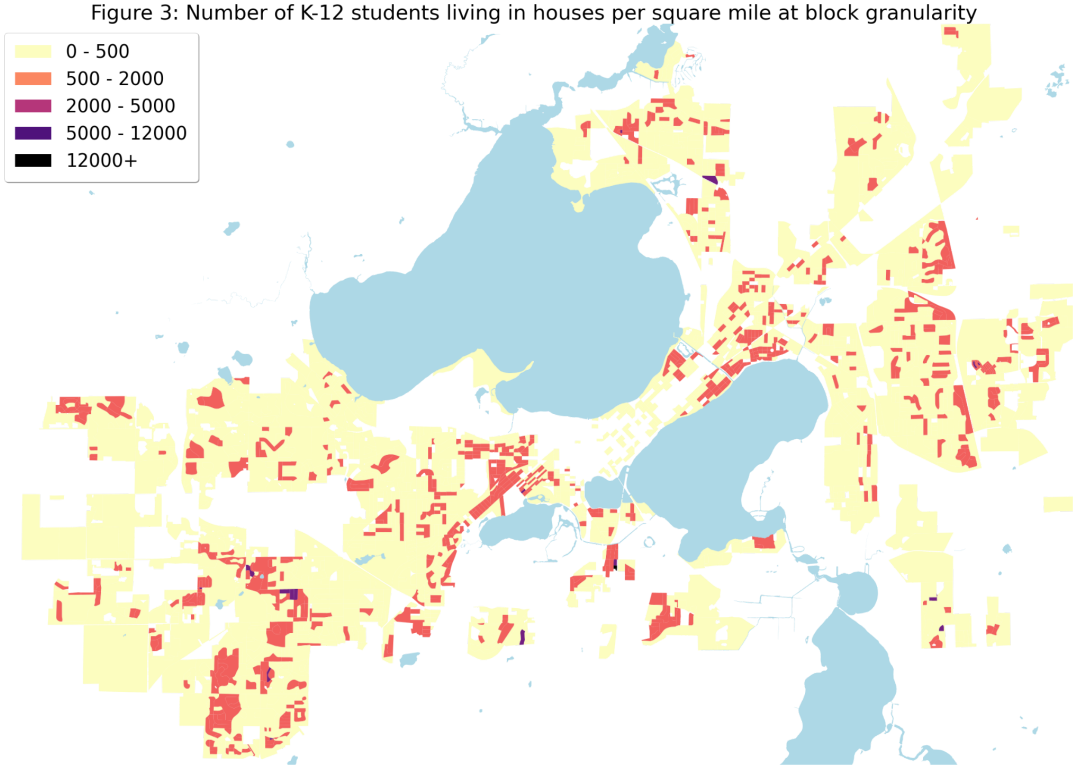
Estimate for a non-apartment property:

$$Non - apartment_{k12} = K_b * \frac{R * B}{B_a + R * B_n} \quad (2)$$

Results and plots:

Based on our updated estimation, we first made a density plot to show our estimate of how many K-12 students live in non-apartments per square mile.

Figure 3: Number of K-12 students living in houses per square mile at block granularity



According to Figure 3, we can find that the number of K-12 students who live in non-apartment per square mile is very similar to what we have in Figure 1. Of course, some blocks have a higher density than others, but most of them have less than 2000 K-12 children living in non-apartments per square mile.

Correspondingly, we made a density plot to show an estimate of how many K-12 students live in apartments per square mile.

Figure 4: Number of K-12 students living in apartments per square mile at block granularity

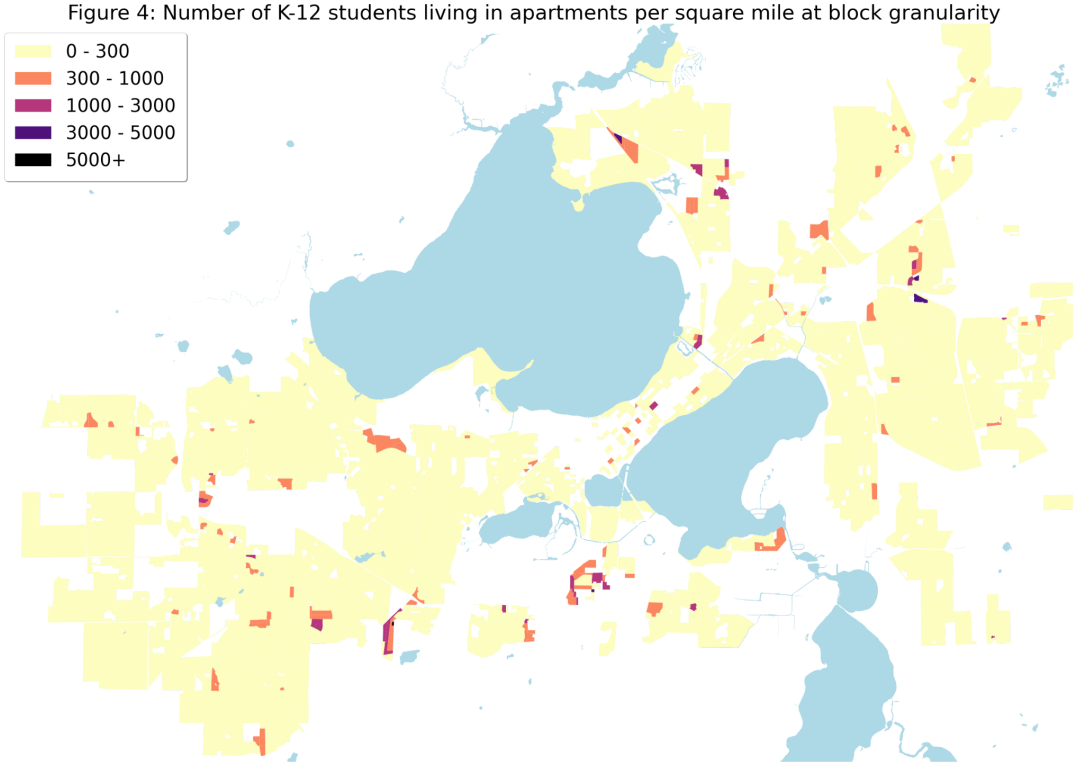
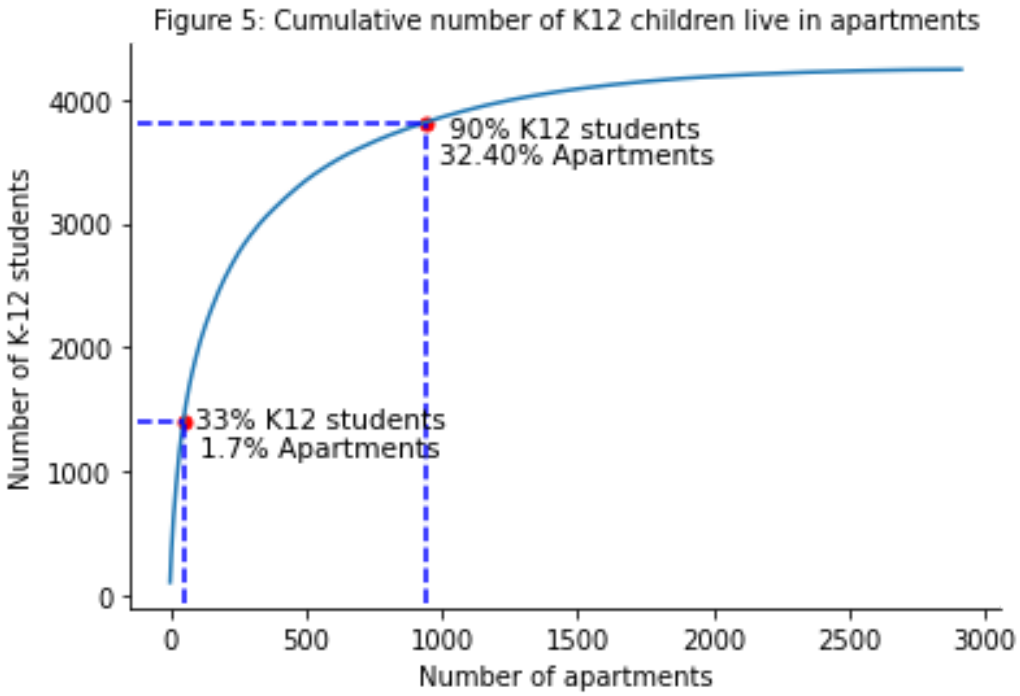


Figure 4 indicates that some blocks have a relatively high K-12 students density of K-12 students living in apartments. This leads to the question of whether K-12 children are gathered in some specific apartments or are evenly distributed among all apartments. To answer this question, we generate a cumulative plot, where the X-axis represents the number of apartments, and the Y-axis represents the cumulative number of K-12 students living in those apartments (by estimation).



We sort all the apartments by the number of K-12 students. Based on Figure 5, more than one-third of K-12 students live in the top 50 apartments, and about 90% of K-12 students live in the top 32% of apartments. Therefore, K-12 students are not evenly distributed among all apartments, and it is crucial for us to know which apartments have a vast number of K-12 children. So we extracted the top 50 apartments with the most K-12 students into a CSV file: Top50_Apartments.csv.

Address	#Bedrooms	#K12Est	Est K12 per bedroom
225 N Thompson Dr	632	109	0.172
2002 Elka Ln	144	63	0.438
501 Northport Dr	256	58	0.227
57 Northridge Ter	96	51	0.531
210-236 & 301-349 & 401-448 & 501-544 East Bluff	441	44	0.100
2501 City View Dr	130	40	0.308
201 N Walbridge Ave	220	40	0.182
3502 Straubel St	96	39	0.406
302 Parkwood Ln	544	38	0.070
2801 Dryden Dr	80	35	0.438
4733 Crescent Rd	72	34	0.472
1124-1148 (every four) Morraine View Dr	649	32	0.049
6421 Bridge Rd	232	30	0.129
7601 Raddliffe Dr	128	30	0.234
415 Moorland Rd	280	29	0.104
1661 Lake Point Dr	216	29	0.134
4629 Atticus Way	208	27	0.130
1108 Moorland Rd	540	27	0.050
1502 Troy Dr	256	26	0.102
302 Kent Ln	244	25	0.102
2201 Cypress Way	144	25	0.174
2301 Cypress Way	144	25	0.174
702 Bear Claw Way	392	24	0.061
420 N Segoe Rd Unit EXP	288	24	0.083
2809 Curry Pkwy	476	24	0.050
8002 Starr Grass Dr	372	24	0.065
6733 Schroeder Rd	224	24	0.107
5806 Raymond Rd	180	23	0.128
700 W Badger Rd	48	22	0.458
2330-2354 (every four) Allied Dr	66	22	0.333
621-629 (odd) E Mifflin St	186	22	0.118
1622 Fordem Ave	216	22	0.102
1602 Fordem Ave	384	21	0.055
5001 Sheboygan Ave	468	21	0.045
1511 Wright St	48	19	0.396
1501 Wright St	48	19	0.396
6401-6409 (odd) & 6439-6443 (odd) Bridge Rd	151	19	0.126
1002 N Pleasant View Rd	292	19	0.065
801 N Thompson Dr	124	19	0.153
4310-4336 (even) Melody Ln & 2402-2454 (even) Independence Ln	638	19	0.030
9001 & 9002 & 9023 & 9045 & 9067 Hawks Reserve Ln & 1802-1806 (even) Maplecrest Dr & 9012-9020 (every four) Dominus Ln & 9002 & 9010-9030 Royal Oaks Cir & 1801 & 1825 Opus Ln	273	18	0.066
2520 S Stoughton Rd	186	18	0.097
1116 Catalpa Cir	12	17	1.417
2720 McDivitt Rd	68	17	0.250
6801 Schroeder Rd	32	17	0.531
1-147 (odd) Craig Ave	148	16	0.108
4711 Jenewein Rd	64	16	0.250
7502 Westward Way	64	16	0.250
2309-2345 (every four) Carling Dr Unit 4	81	15	0.185
3702 Packers Ave	128	15	0.117

Note:
Some addresses are grouped together because they belong to the same geometry point.
*(even): only include even address number. *(odd): only include odd address number.
*(every four): only include one address number for every four (e.g. 1004, 1008, 1012, ...)

Known issues:

Theoretically, we could apply our model to larger areas and generate similar reports. However, there are some issues in this analysis that are worth mentioning. First of all, data that can help us to estimate K-12 students for each property is minimal, and we are just using the number of bedrooms to estimate. If we could have more data related to this topic, we could try more models and estimate better. Another issue is that the K-12 student population is constantly changing. Therefore, the plots we draw might differ using the 2020 census data (we used the 2010 census data).

Conclusion:

In conclusion, K-12 children in Madison are more likely to be in non-apartment properties (houses, etc.) than apartments. Some census blocks have a very high density of K-12 children living in non-apartment properties, but since no single non-apartment property has a large number of bedrooms, they do not have a very significant number of K-12 children in estimation. Therefore, it is better to focus on areas with a higher density of K-12 children living in non-apartments rather than finding a non-apartment property with many K-12 children. For K-12 children living in apartment properties, they are much more gathered. More than one-third of K-12 children living in apartments live in the top 2 percent of apartments with the most K-12 students. Therefore, it would be much more effective and efficient to distribute resources towards areas with higher K-12 children density or apartments with a large amount of K-12 children.