Energy Use, Income and Population Asthma Hospitalization Rates

Dana Chermesh Reshef
dcr346@nyu.edu
NYU CUSP, Center for Urban Science and Progress, New York
Instructor: Prof. Constantine E. Kontokosta

Abstract – **Demographic and environmental conditions have long been found as strongly related to asthma hospitalization and mortality rates, though most of the relevant studies focus on outdoor air quality as the primary factor affects asthma, overlooking the built environment itself as key factor in shaping public health. Today's Energy disclosure mandate for New York City, Local Low 84 (LL84) gives an opportunity to assess the impact of building energy performance on public health, and on asthma rates in particular. This paper examines two primarily possible relationships, the first is between Energy use and neighborhoods' Asthma rates, while the second relationship to be include is between Income and Energy performance. The study revealed very weak, almost non-existing negative correlations between the variables, which might imply for the insufficiency of the models conducted. At the last part of the study a model was developed for only children asthma rates, as well as multilinear regression that tested the strength of predicting asthma rates by considering total Energy Use and Median Household Income as independent variables. The latest analyses were still vague and / or insufficient and leave questions for a future research.**

I. Introduction

A set of social and physical neighborhood characteristics, such as low median household income, high percentage minority, public and inadequate housing, and multiple environmental pollution burdens were found to have impact on asthma hospitalization rates for urban population in general and for urban children living in New York City neighborhoods in particular (Corburn, Osleeb, and Porter, 2006). Until recently, most of the studies researching possible correlations between environment condition factors and asthma rates usually focus on outdoor air quality as the primary factor affects asthma, overlooking the roll of the built environment itself in determining public health (Colton, Laurent, MacNaughton, Kane, Bennett-Fripp, Spengler and Adamkiewicz, 2015).

New York City's established in 2009 its Local Law 84 data (LL84) energy disclosure mandate, which requires annual energy consumption reporting for buildings covered by the law. The new law was immediately seen as one of the most promising public policy tools to accelerate market transformation around building energy efficiency (Kontokosta, 2013), generating new kind of data and approaches to study urban, social, environmental and health issues.

Using the unique LL84 data, combined with health data and demographic data, modeled primarily with basic linear and multivariate regressions, this research aims to understand the relationships between Energy Use and Asthma rates and between Median Household Income and Energy Use Intensity.

II. Literature Review

A. *Asthma Studies and sociodemographic factors*

Although asthma rates, and in particular asthma hospitalization rate of children is consistently decreasing in the past 20 years, the relationship between low income and asthma rates remains strong. For example, East Harlem in Manhattan, has shown between the years 1997-2000 a decrease of 41% in children asthma rates, but continued to have the highest rate of childhood asthma (New York City Department of Health and Mental Hygiene, 2000). Asthma is the leading cause of emergency room visits, hospitalizations, and missed school days in New York City's poorest neighborhoods (Corburn, Osleeb, and Porter, 2006). In 2000, children 0-4 years of age from low-income areas were more than four times as likely to be hospitalized for asthma than children from high-income areas (New York City Department of Health and Mental Hygiene, 2000).

Methods for assessing the most impactful factors of asthma rates include questionnaires and visual inspections to compare indoor environmental conditions and key health outcomes (Colton, Laurent, MacNaughton, Kane, Bennett-Fripp, Spengler and Adamkiewicz, 2015), a vary of statistical models to be developed and tested such as multiple logistic regression, usually adjusted for control variables of gender, age, smoking, country of birth, income and years in the dwelling, etc (Norbäck, Lampa and Engvall, 2014). Spatial analyses and mapping are also commonly used, visualizing the same factors and their spatially distribution and autocorrelations (Corburn, Osleeb, and Porter, 2006).

The studies that do seek for housing characteristic data as predictors of public health usually obtained for building factors such as data on public housing, the age of the housing stock, the condition of units and buildings, etc. (Norbäck, Lampa and Engvall, 2014). These factors have most probably some correlation least with energy use and performance, though no study founded to be putting the buildings' energy performance factors at the center of its analysis.

B. *Energy discloser and benchmarking*

Established in 2009, New York City's Local Law 84 data (LL84) energy disclosure mandate requires annual energy consumption reporting for "covered" buildings (buildings subject to the law) as defined:

i. A building that exceeds 50,000 gross square feet
ii. Two or more buildings on the same tax lot that together exceed 100,000 gross square feet, or
iii. Two or more buildings held in the condominium form of ownership that are governed by the same

board of managers and that together exceed 100,000 gross square feet[1].

The reported information includes: (i) the energy utilization index, (ii) the water use per gross square foot, (iii) where available, a rating that compares the energy and water use of the building to that of similar buildings, and (iv) a comparison of data across calendar years for any years such building was benchmarked[2].

LL84 applies in total to more than 15,000 privately-owned properties across New York City, accounting for approximately 45% of the City's total energy consumption (Kontokosta, 2014). Recent years' researches are deepening the understanding of the spatial patterns of energy consumption and the potential negative externalities of inefficient buildings (Kontokosta, 2013). These data is becoming more and more useful to help governments and neighborhoods to better address current urban challenges and to plan a sustainable future development.

III. Data and Methods

A. *Energy Performance Data (Local Law 84)*

The primary dataset was used in this research is the LL84 data reported in year 2016. The first part of this analysis, that seeks for a possible correlation between energy use and asthma rates, includes all property types (residential, offices, commercial, public buildings, etc.). For the second part of the analysis, that concentrates on median household income, only Multifamily housing property type were taken into consideration. The data were cleaned from missing values, non-relevant columns and outliers, which were defined as Energy Use Intensity (EUI) of more than 1,000 kBtu/ft2. Fig1+2 reveal the distribution of Site EUI (kBtu/ft^2) data.

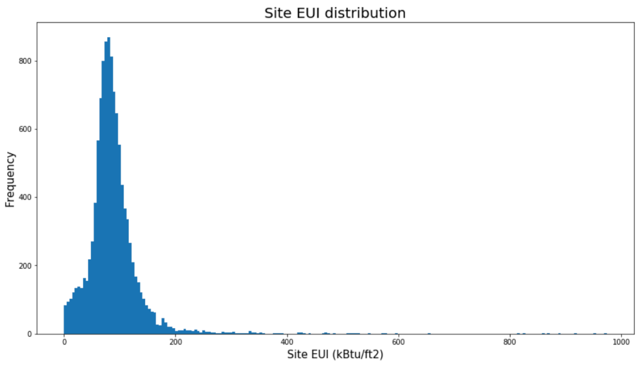


Fig.1 Site EUI distribution (kBtu/ft^2).

The above distribution seems to need more cleaning; Thus, a limit of two standard deviation from the mean, so as observing the log of the distribution, was implied. Fig.2 is the logarithmic distribution of Site EUI (kBtu/ft^2).

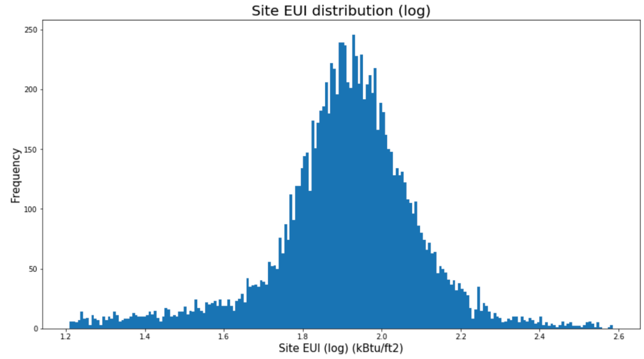


Fig.2 Site EUI log distribution

Prior to the merging of the energy data with other datasets possible relationships within the data were examined.

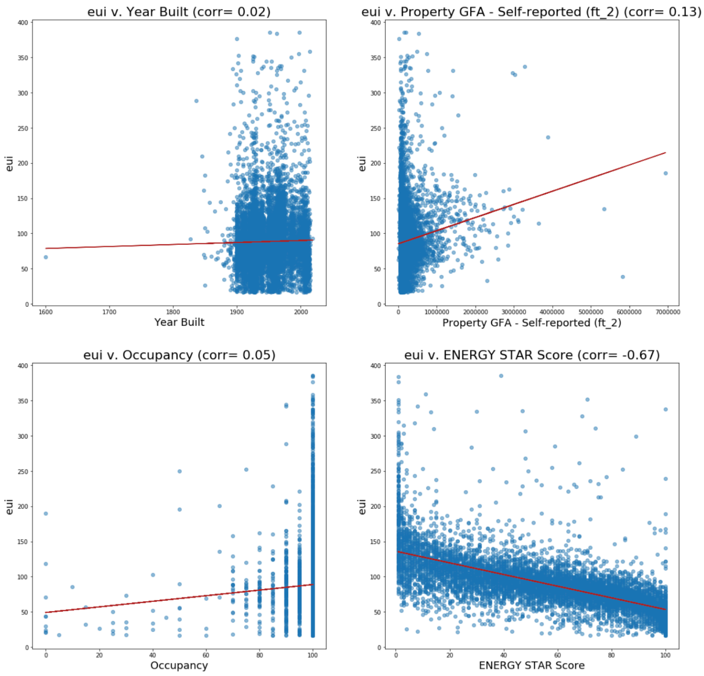


Fig.3 Relationships between factors and Energy use intensity

Possible relationships are shown between Property GFA (ft^2) and Site ETU (kBtu/ft^2), although one might benefit from observing these variables in their logarithmic dimension, and/or from developing a polynomial model that will better explain the data. A strong relationship is shown between ENERGY STAR Score and Site ETU (kBtu/ft^2), as well as a strong predictive linear model. On the other hand, it seems hard to linearly predict Site EUI (kBtu/ft^2) from Year Built / Occupancy data. Indeed, researches on energy efficiency and benchmarking had shown that the assumption of linear correlations between energy performance variables often fails, and that more complex models are usually better in explaining energy use (Kontokosta, 2014).

B. *Asthma hospitalization Data*

The second dataset was collected for this study is Asthma hospitalization data, in particular discharged from hospitals from years 2012-2014. Although the data is not from the same year as the LL84 data it is not likely to be very different and can still provide a reliable analysis. The data were organized and merged

[1] Local Law 84, City of New York, 2009
[2] Local Law 84, City of New York, 2009

with the LL84 dataset to analyze the possible correlation of total Energy Use (kBtu) and the rates of asthma of the population. The LL84 data were grouped by zip code and the median of each zip code was extracted, this to make the merging possible. In accordance, the two datasets were merged based on zip code. Energy use was calculated from EUI (kBtu/ft^2) and Gross Floor Area of the building. The data included from the asthma data set were the rates of discharged patients from hospitals, the number is revealing discharges per 1,000 people.

C. *Median Household Income data*

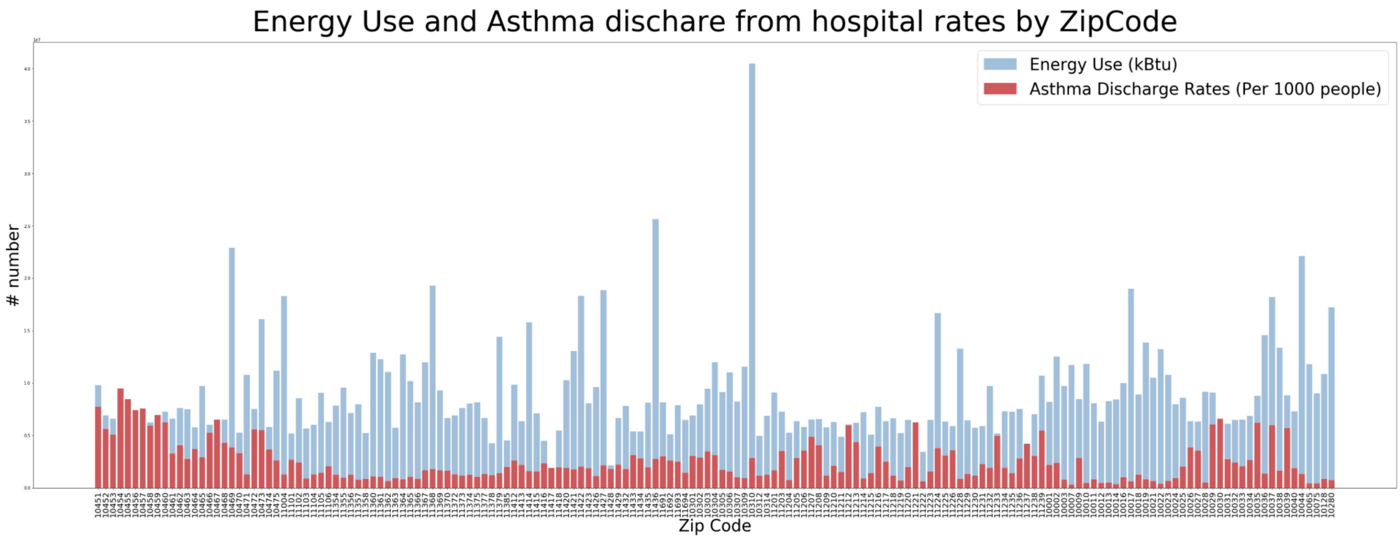
For the second part of the analysis, Median Household Income and Energy Use Intensity relationship, data of Income were also included. The demographic data for all census tracts in New York City were obtained from the American Community Survey (ACS) of year 2015. Data were cleaned from any private and sensitive information prior importing. The data used were Median Household Income, grouped by community district.

In order to combine the Census data with the primary LL84 dataset it was required to convert Zip Code information to Community District terms. To this objective Primary Land Use Tax Lot Output (PLUTO) data set for all New York five boroughs were collected. Using Bin, Block and Lot (BBL) data the two datasets were combined and could be analyzed. The relationship to be examine was between median household income and energy use intensity (EUI) (kBtu/ft^2), with income as the independent variable.

D. *Methodes*

The methods of this study were similar to both parts of the analysis. First, I layered each two variables one over the other to detect possible relationships and / or trends. The first part's layered bar plot, Energy Use and Asthma discharge from hospitals rates, is shown in Fig.5 for all zip codes included in the merged dataset. Any clear relationship nor correlation between the two was emerging from this plot and it didn't turn to provide much essence. Similarly, the second part of this analysis, Median household income and EUI data, were layered one over the other in a bar plot, as shown in Fig.6. Again, it was hard to understand the nature of the relationship between the two variables and whether any correlation exists between them.

To better understand the analyzed variables' relationships simple linear regressions were modeled (Model 1, Model 2), with Energy Use (kBtu) as the independent variable and the Asthma discharge rates as the dependent variable, and Median Household Income as the independent variable for understanding Site EUI (kBtu/ft^2), respectively. Fig.7 shows the data + the fitted line of Model 1; Fig.8 reveals the data + the fitted line of Model 2.



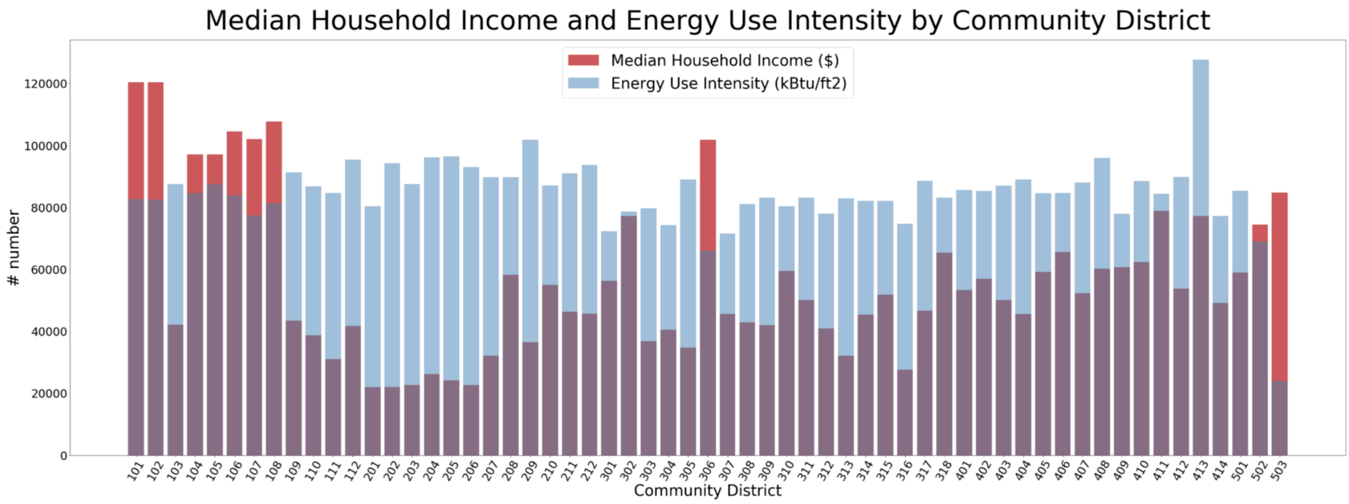


Fig.5 Energy Use (kBtu) and Asthma discharge from hospital rates (per 1,000 people) by Zip Code. **Fig.6** Median Household Income Energy Use Intensity (EUI) (kBtu/ft^2) by Community District

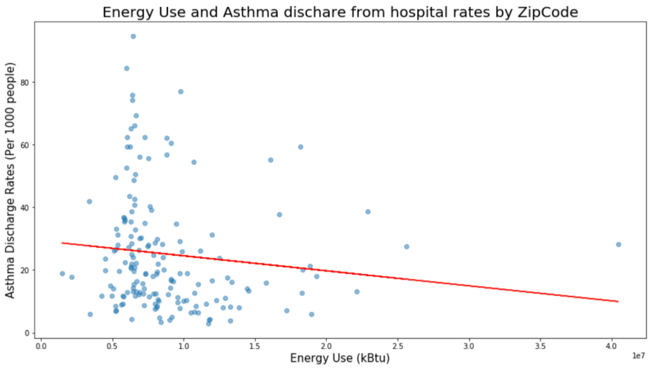


Fig.7 Asthma rates over Energy Use + Model1 fitted line. R-squared = 0.014, slope = -4.791500e-07

The fitted line of Model 1 does not seem to explain at all the scatter plot behind it. The observed data might indicate a possible correlation that is obviously not linear. The R-squared of Model 1 is indeed very small, equals to 0.014, and the slope of the model is -4.791500e-07. These results indicate an extremely weak negative linear correlation between the variables and the lack of ability to predict Asthma rates based on Energy Use. It is possible that the relationship is not linear or that other factors taken into consideration alongside with Energy Use will provide a better model.

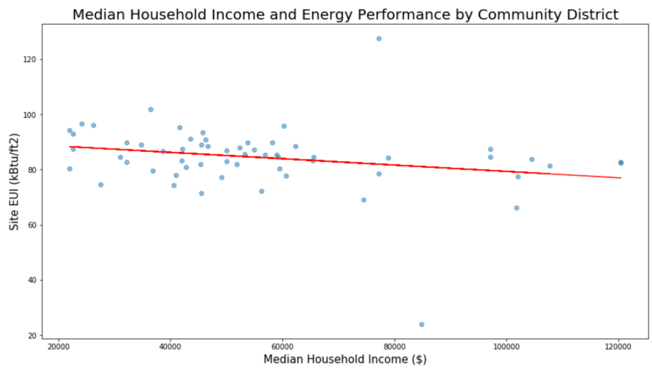


Fig.8 Site EUI (kBtu/ft^2) over Median Household income + Model2 fitted line. R-squared = 0.058, slope = -0.000115

Model 2 reveals a weak negative correlation, with R-squared = 0.058 and a slope of -0.000115. The model's fitted line, though, seems to make a slightly better sense with explaining the data as observed in the scatter plot. Still, due to the low results of the model it is hard to predict Energy Use Intensity of a building according to the median household income. As in the first part of the analysis, it is possible that the relationship is not linear or that other factors taken into consideration together alongside with income will provide a better model.

E. Research limitations and further analysis

Both analyses were insufficient and revealed very weak linear correlations between the variables analyzed. A possible reason to the lack of success of the models is that when looking on zip codes as well as on community districts, we practically assume that people spend most of their time where they live, which is obviously a wrong assumption.

One way to deal with that limitation and to isolate the place a person lives as the main factor when predicting asthma is to limit the population to children only. Children are more likely to spend most of their days and nights in / close to their homes. The available Asthma dataset is divided to ages group, what made the suggested task possible.

A new dataset of children aged 0-17 hospital discharge rates was collected and merged with the LL84 original data, in the same workflow as describe earlier in the method section of this research. A linear model (Model 3) has been formulated again to these new data, with R-squared of 0.032. Though more than twice as high as the total population rates (Model 1, R squared = 0.014), it is still a very weak correlation that cannot explain the observed data in a sufficient way. Fig.9 shows the scatter plot of the data of Children Asthma rates over Total Energy Use and the fitted line of Model 3.

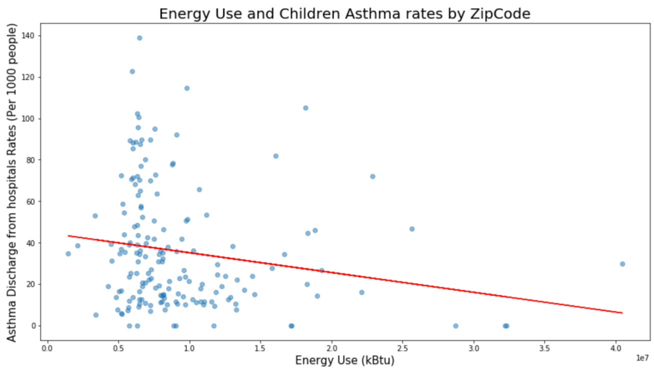


Fig.9 Site EUI (kBtu/ft^2) over Median Household income + Model2 fit ted line. R-squared = 0.032, slope = -9.537701e-07

The insufficient results of the three models were strengthening the assumption that the relationships could be better described by more complexed models. The final step of this research was to try to build a multilinear model to predict Asthma discharge rates with two independent variables: (i) Total Energy Use and (ii) Median Household Income. For this purpose, the two datasets of the first and second part of the analysis were merged, to have Income, Energy Use and Asthma rates organized and grouped by zip codes. The model (Model 4) gave a much better R-squared: 0.408, but still very small numbers of coefficients to its independent variables: 3.274e-07 For Energy Use and -0.0008 For income. These results may indicate that there is a strong multicollinearity or other numerical issues, which make this model unreliable. Although we saw in Model 2 that a correlation between Income and Site EUI (kBtu/ft^2) barely exists, it was wrong to infer the relationship between Income and total Energy Use (kBtu) is also negligible.

IV. Results

All models were executed in this analysis didn't sufficiently predict or explain the observed data, whether they had very small R-squared values or had multicollinearity issues. Fig 10-14 are maps of New York City, revealing the spatial distribution of the factors were included in the analysis.

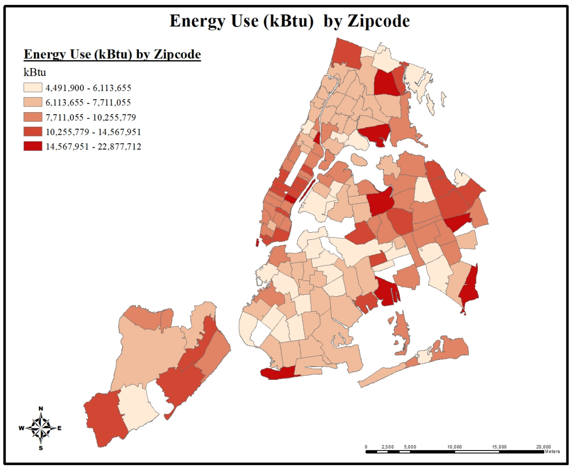


Fig.10 Energy Use (kBtu) distribution by Zip Code

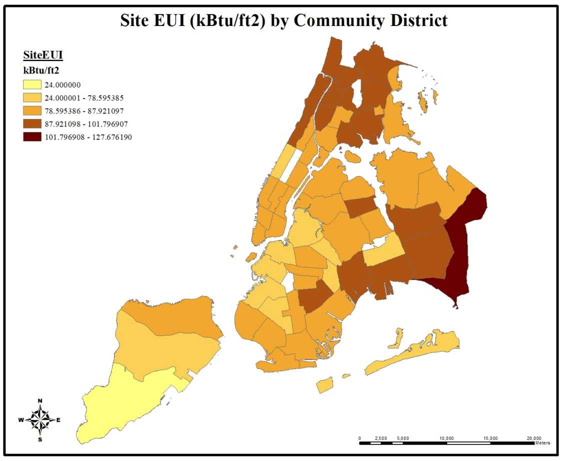


Fig.13 Site EUI (kBtu/ft^2) by Community district

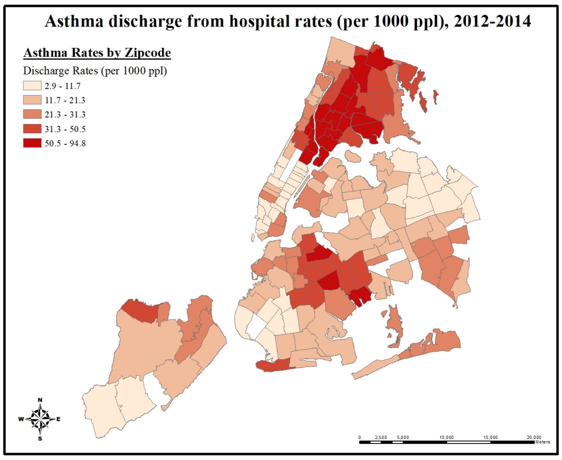


Fig.11 Total Asthma rates (per 1,000 ppl) distribution by Zip Code

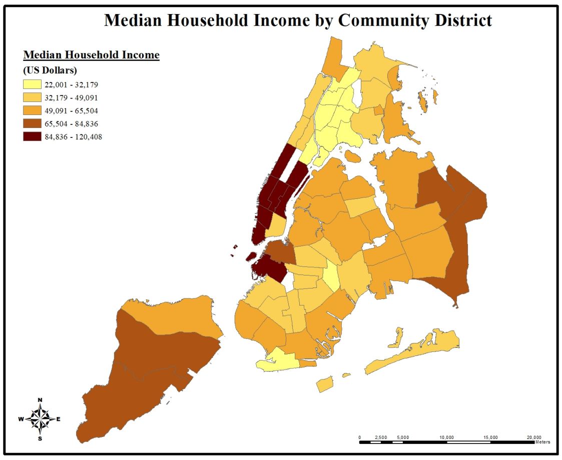


Fig14 Median Household Income (US Dollars) by Community district

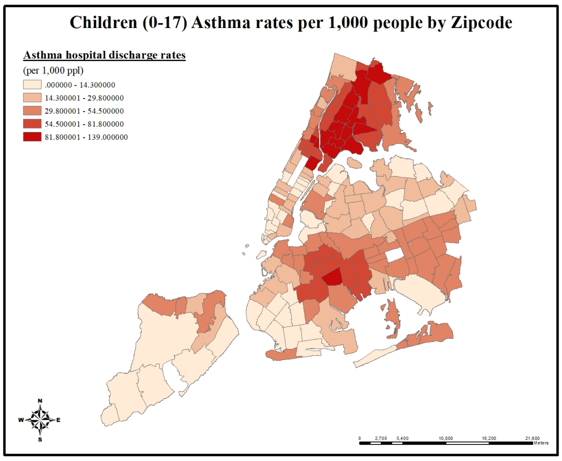


Fig.12 Children (aged 0-17) Asthma rates distribution (per 1,000 ppl) by Zip Code

The spatial views also don't give much more understanding of the correlations, which is reasonable due to the extremely we ek correlations had emerged from the models. Comparing Fig.1 0 and Fig.14 might imply the correlation between high income a nd high total energy use although it is not what we were trying t o assess. Fig.13 and Fig.14 do not reveal a clear conclusion reg arding higher income as a predictor of higher / lower EUI (kBtu /ft^2), which fits the ambiguous results of Model 2 that indicated a weak, non-compelling, negative correlation between these var iables. Moreover, one could see the obvious similarity between the distribution of poverty (low-income households as revealed in Fig.14 as the lighter colors of Community Districts) and the s patial concentration of Asthma as revealed in Fig.11 and Fig.12 , with the Children Asthma rates as the more strongly spatially c lustered than the total population clustering. This similarity as w ell as the correlation between low income and high asthma rates is well known and broadly discussed and researched, and was n ot the purpose of this study. The maps on Fig.10-12 might impl y some negative correlations between total Energy Use and Ast hma rates, though it seems, as confirmed by Models 1 and 3, thi s correlation is scientifically weak.

V. Discussion and Conclusions

A. *Discussion*

As discussed earlier in this paper, the poor results of the models do not necessarily infer the lack of any relationship between Energy Use and Asthma rates as well as between Income and Site EUI. They do mean that the models are too simple and / or that the geographical references we picked, zip codes and community districts, are too broad or misleading, due to the argue by which people do not necessarily primarily expose to the air in the building they live in or its immediate surrounding. To this objective, including only children aged 0-17 in the analysis strengthen the model but only marginally, and turned to provide not a much better explanation of the asthma rates data. It is possible that calculating only young children, aged 0-4 for example, would have given a deeper understanding, a more reliable model and would revealed a stronger correlation. Unfortunately, these data was available only in the County level, which would be too broad for inferring any reasonable conclusion.

Moreover, as has been argued in latest studies, reliance on simple metrics such as EUI often fails to account for significant effect on elements of a building that influence building energy consumption (Kontokosta, 2014). This suggests that there might be a better variable from the energy performance data for predicting public health as well. A proper factor to be considered in this case could be Total GHG Emissions (Metric Tons CO_2e), that indicates the more polluting amount of the energy use, that aren't generated by heat or electricity.

Taking all the above into consideration, there is of course the possibility that there really is no strong correlation between Energy Use and Performance and between Asthma hospitalization rates and / or Median Household Income.

B. *Conclusions and next steps*

The recent proliferation of energy disclosure policies in U.S. and global cities is generating significant new streams of data on patterns of energy consumption in buildings, which can help address environmental and social challenges. This research tried to define the correlation between Energy Use and Asthma rates and between Median Household Income and Energy Use Intensity, though the analysis did not succeed with developing well predictive models for the dependent variables. One reason to the lack of success as argues before could be the limitation of treating zip codes and community districts as proxies for a person's air and environmental consumption. Another reason Finally, could be the simplicity of the model

In future analyses, it could be beneficial to tackle the limitations acknowledged above and by either adding more variables to the model including seeking for possible multicollinearities, reducing the data to a more specific group of population to be examined, and / or applying polynomial model to the analyses.

VI. References

[1] J. Cayla, N. Maizi and C. Marchand, "The role of income in energy consumption behaviour: Evidence from French households data", Energy Policy, vol. 39, no. 12, pp. 7874-7883, 2011.

[2] M. Colton, J. Laurent, P. MacNaughton, J. Kane, M. Bennett-Fripp, J. Spengler and G. Adamkiewicz, "Health Benefits of Green Public Housing: Associations With Asthma Morbidity and Building-Related Symptoms", American Journal of Public Health, vol. 105, no. 12, pp. 2482-2489, 2015.

[3] J. Corburn, J. Osleeb, and M. Porter, "Urban asthma and the neighbourhood environment in New York City" Heakth and Place, vol. 12 no.2, pp. 167-179, 2006.

[4] C. Kontokosta, "Energy disclosure, market behavior, and the building data ecosystem", Annals of the New York Academy of Sciences, vol. 1295, no. 1, pp. 34-43, 2013.

[5] C. Kontokosta, "A Market-Specific Methodology for a Commercial Building Energy Performance Index", The Journal of Real Estate Finance and Economics, vol. 51, no. 2, pp. 288-316, 2014.

[6] New York City Department of Health and Mental Hygiene, "Asthma Facts, second edition", 2000.

[7] D. Norbäck, E. Lampa and K. Engvall, "Asthma, Allergy and Eczema among Adults in Multifamily Houses in Stockholm (3-HE Study) - Associations with Building Characteristics, Home Environment and Energy Use for Heating", PLoS ONE, vol. 9, no. 12, p. e112960, 2014.

[8] R. Sari and U. Soytas, "The growth of income and energy consumption in six developing countries", Energy Policy, vol. 35, no. 2, pp. 889-898, 2007.

[9] A. Sharpe, C.R. Thornton, V. Nikolaou, N.J. Osborne. "Higher energy efficient homes are associated with increased risk of doctor diagnosed asthma in a UK subpopulation", Environment International, vol. 75, pp. 234-244, 2015.

[10] Ipython Notebook of this analysis and all data that were included in it can be found on GitHub/danachermesh/CivicAnalytics2017_dcr346/Problem_Set2