EAS 503 Final Report Group #10

Our goal was to use the Erie County census data set to create maps which would show people moving to the area the optimal places to live based on your income bracket.

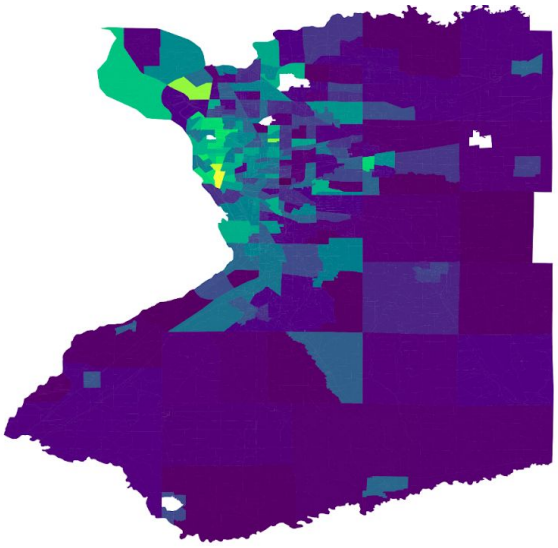
The Erie County census data contained 500 variables and 237 samples. Each sample was a census tract, and the 500 variables included social, economic, housing, and demographic characteristics of each tract. The original data set was in the form of an excel table and we manually deleted several hundred predictors that were clearly not related to our problem. We were able to narrow down the number of predictors to just 40 out of the 500 variables. Our SQL database included these 40 predictors.

For our initial analysis, we started by querying specific columns from our database so that we could create some plots. We plotted the median household income of each area sorted from lowest to highest. We noticed two distinct cut offs for grouping the incomes, which gave us three income brackets of <$40,000, $40,000-70,000, and >$70,000. Then we plotted pairs of variables against one another and made a decision on which variables would have the greatest predictive power when creating a map without being too repetitive. We ended up with 13 predictors. Each predictor value was a number of households in the sample area which answered yes to the census question. We divided all of our predictor values by the total houses in their corresponding areas. This gave us 13 predictor ratios.

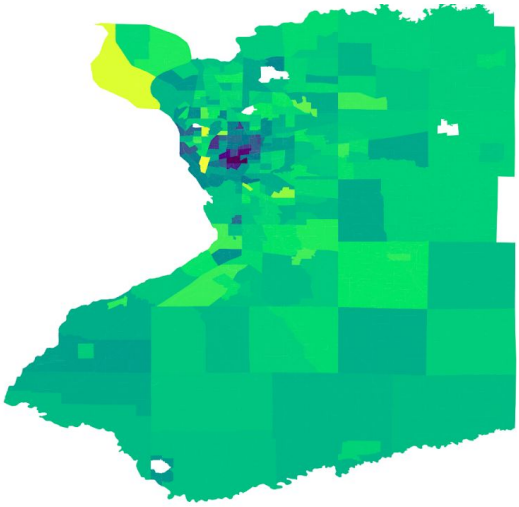
We then decided to calculate a score for each area in order to produce a gradient map. To calculate this score, we had to assign a weight to each predictor ratio. These weights would vary depending on the income bracket. For each predictor ratio, we assigned a weight value from -1 to 2 which represented how important each predictor would be for each of the three income brackets. For example, when we decided on weights for the rent as 30-35% of income, we looked at our plots and noticed that many households which had rent as 30-35% of income were of lower income. We also did some research online and found that if your income is below $40,000, then you should ideally spend 30-35% of your income on rent. Our source also noted that middle income families are renting at much higher rates in recent years. So we gave the weights of 2, 1, 0 to low, medium, and high incomes respectively for the predictor.

We then summed up all of the weighted predictors to get a score for each area at each income bracket. Using python-SQL interface these scores were then saved into a pandas data frame along with the corresponding geographical IDs. We were also able to find a shapefile of Erie County that had correlated geographical IDs to the latitude and longitude locations on a physical map. The Geopandas library allowed us to convert this the shape file into a pandas dataframe. Since our manually created data frame containing the scores and the dataframe extracted from the shape file both had a column containing geographical IDs, we were able to perform an inner join between both to combine them into a single dataframe. Geopandas then allowed us to use the score column of this newly combined dataframe, to create a choropleth map, that was weighted by our assigned score value. We choose the viridis color scheme as the color map. This color map shows the most desirable places to be a bright green color while the non desirable locations to be of a darker shade. This allows us to clearly identify what is considered desirable over not desirable.

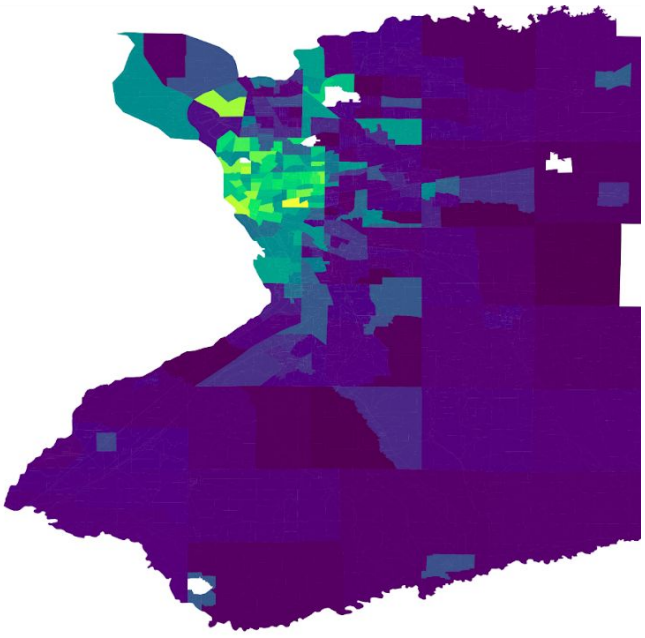
Our results show that the best places for people with lower incomes to move would be closer to the city of buffalo. This makes sense because we had a greater weight on factors like having access to public transportation for people with lower incomes. Our results showed the opposite for the higher income bracket, that areas outside of the city would be more desirable places to live than in the center of the city.



1



2



3

Figures 1-3: Most desirable places to live based on income bracket. Income brackets: 1: <$40k, 2: $40k - $70k, 3: >$70k

Minimum score (least desirable) = black
Maximum score (most desirable) = bright green/ yellow

https://www.jchs.harvard.edu/sites/default/files/05_harvard_jchs_americas_rental_housing_2017.pdf