Low Live Birth Weight Prevalence and Socioeconomic Factors in Chicago

https://github.com/aah77/Hingu_ENV872_EDA_FinalProject.git

Aman Hingu

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Low Birth Weight refers to when a baby weighs less than 5 pounds 8 ounces (approximately 2.5 kg) at the time of birth. This definition was adopted as a major benchmark for pregnancy success, as well as a predictor for certain pregnancy complications (Taffel, 1975). For the child, complications associated with low birth weight include Respiratory Distress Syndrome (RDS), Intraventricular Hemmorhages, and a weakened immune system (March of Dimes, 2021). A compromised immune system places the child in a vulnerable state, drastically reducing their ability to fight off infections. Approximately eight percent of all babies born in the United States are born with low birth weight, but this group constitutes over 75% of all neonatal deaths in the United States (Taffel, 1975).

Occurrences of low live birth weight vary widely based on physical characteristics of the mother, pregnancy history, age, and marital status. Many national public health organizations recognize these factors, as well as other factors that pertain to the mother's socioeconomic status. Relevant socioeconomic factors can include household income level, proximity to high traffic volume areas, food insecurity, and educational status. Factors such as income and food insecurity are often intrinsically tied and deter natural maternal weight gain, thereby increasing the likelihood of a low live birth weight (Gourlay, 2021).

The connections between socioeconomic factors and maternal physical health motivates the central research question of the project below:

1. Do socioeconomic factors such as unemployment, educational attainment, and portion below poverty level explain variability in the prevalence of low birth weight?

It is hypothesized that the amalgamation of all three factors will produce the most statistically robust model in comparison to the influence of any of the models alone.

2 Dataset Information

The dataset to be used in this study is the Chicago Health Indicators dataset, originally found on a website called Geo Data and Lab. The Geo Data and Lab website is a repository of many different socioeconomic, health, and environmental datasets. The data was originally provided by the Illinois Department of Public Health and the US Census Bureau.This dataset includes public health data for 77 community areas within the greater Chicago Metropolitan Area, and was aggregated from 2005-2011.

Once attaining the dataset, an initial analysis was completed by running head(), dim(), and summary() functions on the variables within the dataset. After looking through the variables provided in the dataset, a brief and informal literature review was conducted in order to deduce which socioeconomic variables might bear the greatest level of influence on low live birth weight. Parameters that were frequently discussed in the literature included household income, level of education held by the mother, and whether the mother (or other household members) possessed a stable source of income.

Based on these commonly occurring parameters, the dataset was wrangled to omit occurrences of NA, and selected for columns representing "Portion under Poverty Level", "Unemployment Rate", "Community Area", " "Low Birth Weight Prevalence", and "Educational Attainment". A key characteristic in the behavior of socioeconomic variables is their tendency to vary spatially. This phenomena is especially true in urban metropolitan areas where costs of living and income status are major factors in a family's choice of deciding where to live. The area in which a family chooses to reside affects the ease of access they may or may not have to nutrient dense food vendors (such as supermarkets), as opposed to convenience stores. Factors such as these may play larger roles in a mother's likelihood of having a child with a low birth weight given the physiological relationship between a mother's food intake and their child's physical health (Christian, 2020).

Table 1: Variable Descriptions

Variable	Description
Low Live Birth Weight	Percentage of all births in a community area with a birth weight less than 5.5 pounds.
Unemployment Rate	Unemployment as a percentage of persons aged 16 older.
Educational Attainment	Percentage of persons aged 25 or older without a high school diploma.
Portion under Poverty Level	Percentage of households in a community area living under poverty level.

3 Exploratory Analysis

To further confirm whether the socioeconomic variables chosen for in this dataset vary spatially, spatial heat maps of the dependent and independent variables chosen were created and are displayed below using the initial spatial dataframe.

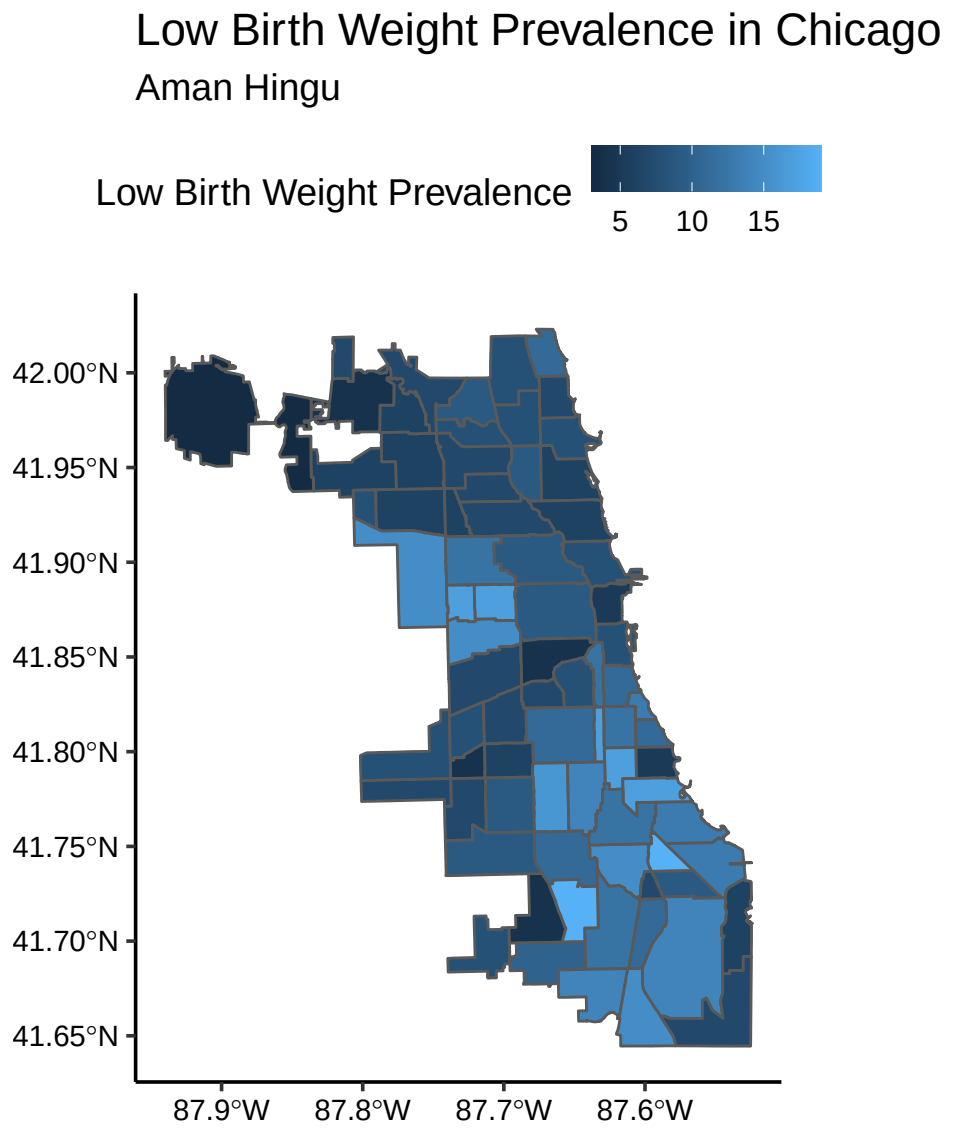

Figure 1: The figure depicts a spatial distribution map of Low Birth Weight Prevalence across the Chicago Metropolitan Area

Observation of the spatial plots above indicates disparities and an obvious variability in the distribution of these socioeconomic variables. For the spatial plot of low live birth weight prevalence, there is more pronounced density in the central eastern and southwest community areas of the Chicago Metropolitan area. The next section proceeds with linear regression analyses to quantitatively explore the influence of the socioeconomic variables of interest on low live birth weight prevalence.

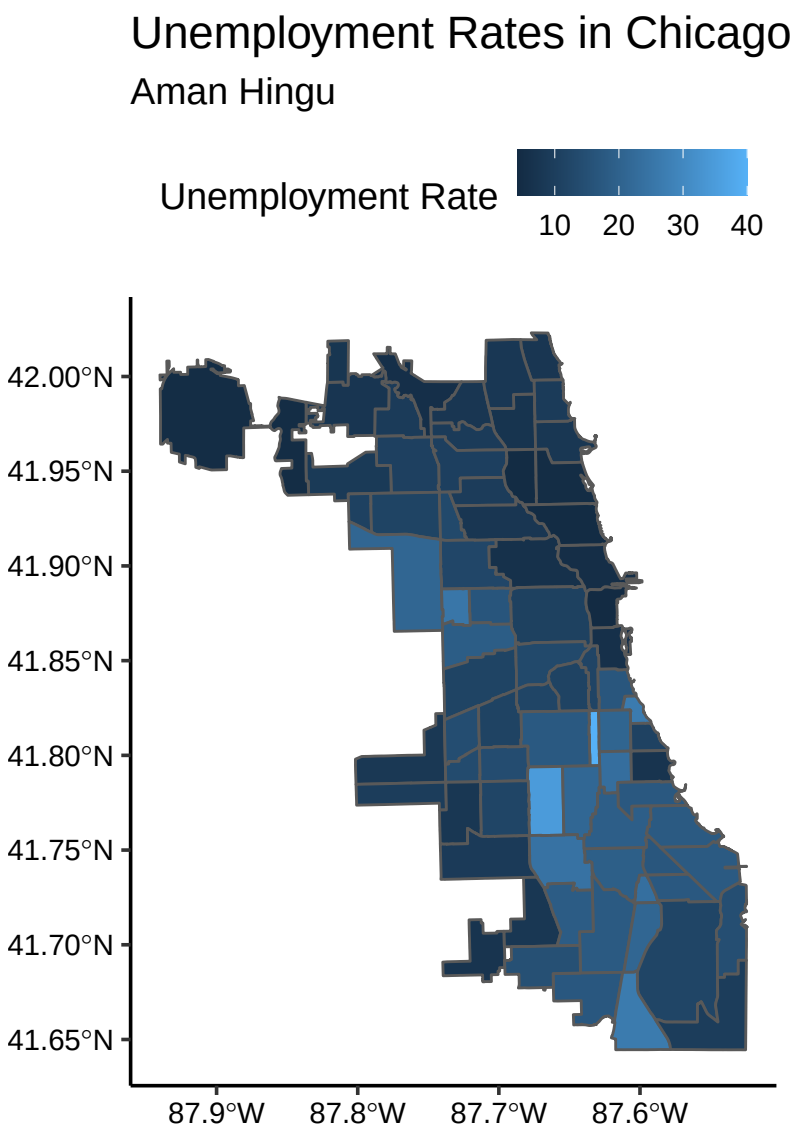

Figure 2: The figure depicts a spatial distribution of unemployment rates across the Chicago Metropolitan Area

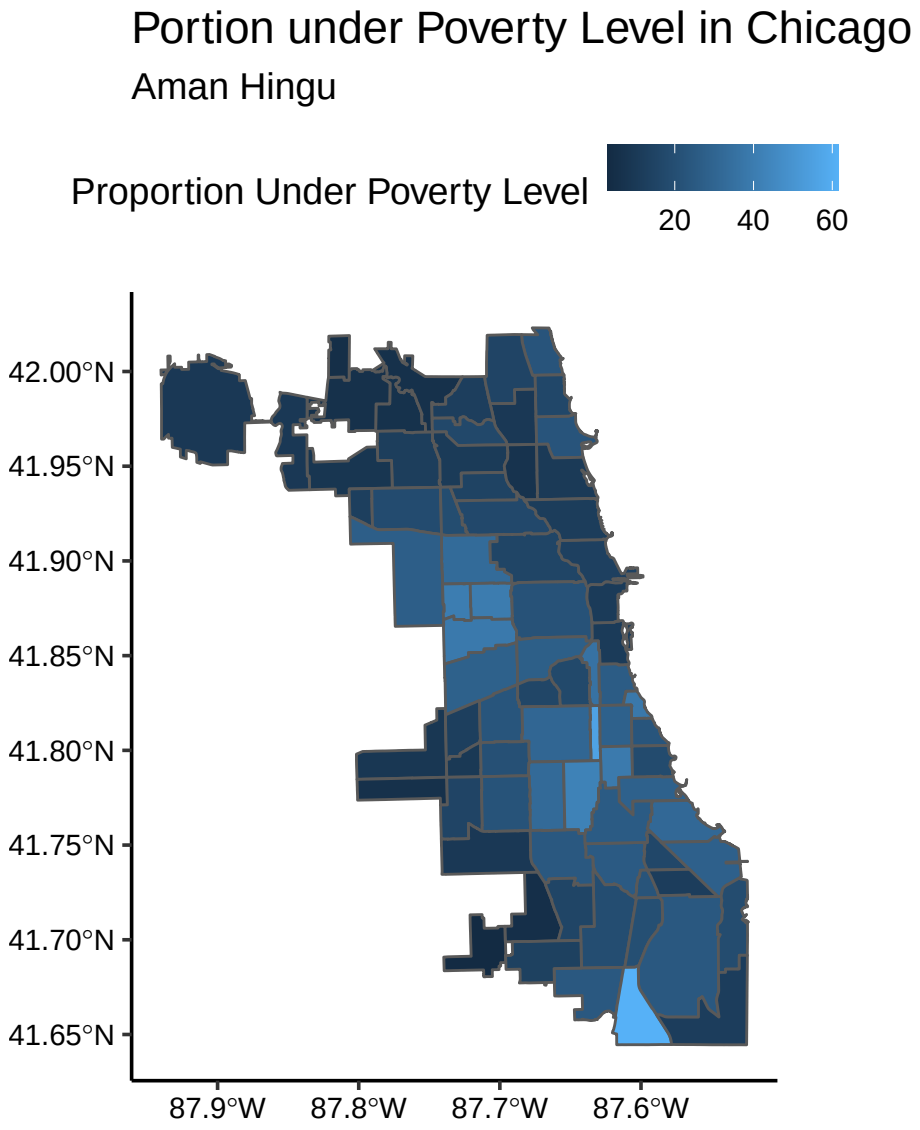

Figure 3: The figure depicts a spatial distribution of portion of population under poverty level across the Chicago Metropolitan Area

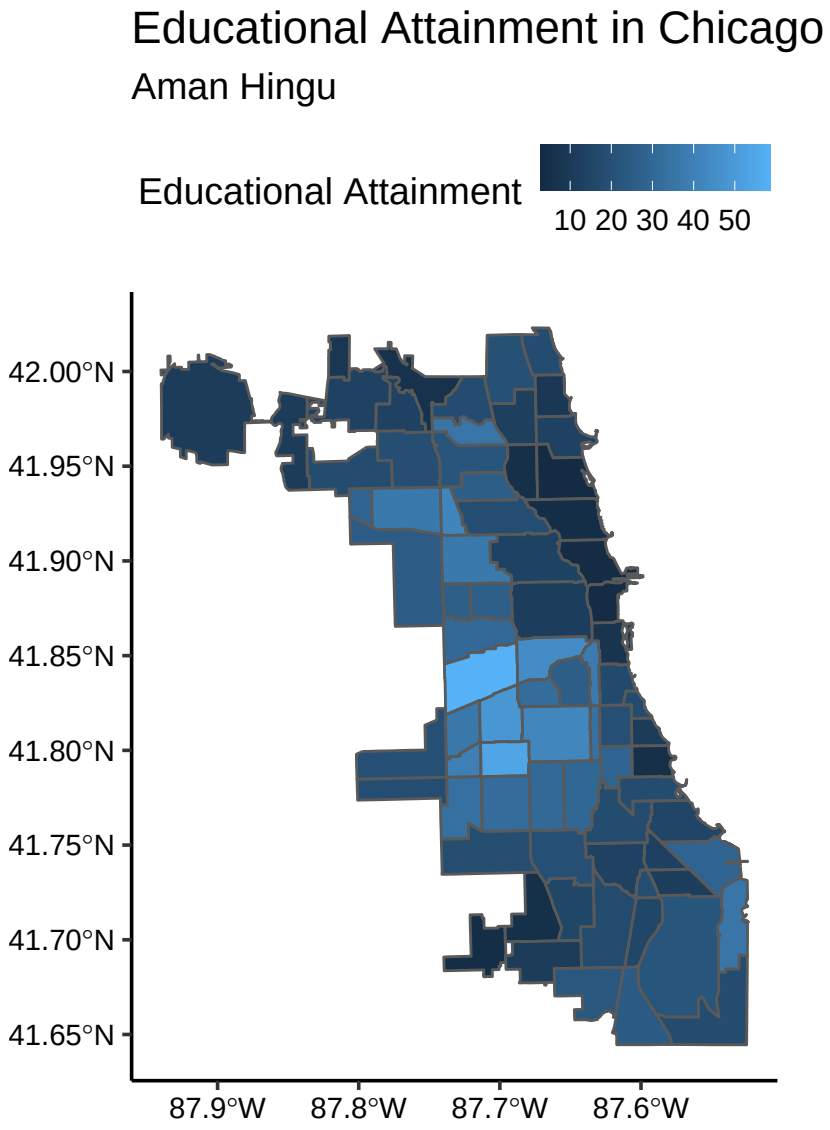

Figure 4: The figure depicts a spatial distribution of educational attainment across the Chicago Metropolitan Area

4 Analysis

4.1 Question: Do socioeconomic factors such as unemployment, educational attainment, and portion below poverty level explain variability in the prevalence of low birth weight?

The overall analysis goal was to create a linear regression model with a significant p-value that optimizes the adjusted R-squared value. Thus, the process here was to first create a correlation plot to get an idea of which of the variables of interest had the strongest correlation value to low live birth weight and then proceed with creating a model from that. Next step in the process included creating multiple linear regression models, visualizing their model fit through residual plots & QQ plots, and looking at their coefficients, p-values, and R-squared values. The last step of the process included creating another model to see if a higher R-squared value could be achieved while maintaining significance.

Unemployment and Portion under Poverty Level had the two highest correlation values, and thus my first multiple linear regression model consisted of looking at those 2 variables as predictors of low live birth weight prevalence. When visualizing the model fit, the QQ plot yielded most of the points on the 1/1 line, but there was assymetry in the residuals plot indicating poor model fit. The p-value was significant for the model, and the adjusted R-squared value was 0.532, meaning that just about half of the variability in low live birth weight can be explained by unemployment and poverty level.

Given the lower R-squared value, another multiple linear regression model was created that included the third variable, Education Level, that was originally left out. This model had a very similar poor model fit as evidenced by the skewed residuals vs. fitted plot. The model was also significant, but had a slightly higher adjusted R-squared value. The results of both models are organized into Table 2.

Table 2: Model Results

Model	Variables	P	R
1	Unemployment, Portion Under Poverty Level	***2.35E-13	0.5320
2	Unemployment, Portion Under Poverty Level, Educational Attainment	***6.36E-15	0.5941

5 Summary and Conclusions

The model that was both significant and resulted in the highest adjusted R-squared value was Model 2. Model 2, which included all three independent variables of interest, resulted in an R-squared value of 0.5941. This means that approximately 59% of the variability in Low Birth Weight Prevalence can be explained by Unemployment Rate, Educational Attainment, and Portion Under Poverty Level. While this was a greater fit in comparison to Model 1, Model 2 is still relatively weak and could be improved.

For further improvements, researchers should consider exploring other variables which are known to have some level of correlation to low birth weight from the literature. One example of such a variable is air quality (Sarizadeh, 2020). This study indicated that, due to physiological changes in a mother's body, mothers were particularly more sensitive to higher concentrations of common air pollutants. Common air pollutants included in this study were Ozone, PM 10, and PM 2.5. Air quality can also vary regionally depending on a particular region's proximity to industrial activity, such as oil and gas plants, paper mills, etc.

Researchers can also consider variables commonly associated with malnutrition in mothers, which is a leading cause of low live birth weight (Sarizadeh, 2020). These variables can include proximity towards supermarkets, or even the concentration of supermarkets within a particular community area. Food insecurity associated with residing in a food desert is known to be associated with malnutrition (Christian, 2020).

Including these variables in a multivariable linear regression may lead to a model in which more of the variability in low live birth weight prevalence is accounted for. Local public health agencies can utilize models such as these to drive policy based changes to support mothers at a higher risk of having a child with a low birth weight by virtue of their socioeconomic status.

6 References

Christian, Vikram J., et al. “Food Insecurity, Malnutrition, and the Microbiome - Current Nutrition Reports.” SpringerLink, Springer US, 10 Nov. 2020, https://link.springer.com/article/10.1007/s13668-020-00342-0.

Sarizadeh, Reihaneh et al. “The Association Between Air Pollution and Low Birth Weight and Preterm Labor in Ahvaz, Iran.” International journal of women’s health vol. 12 313-325. 4 May. 2020, doi:10.2147/IJWH.S227049

March of Dimes. Low Birthweight, https://www.marchofdimes.org/complications/low-birthweight.aspx#.

Coleman-Jensen A, Gregory C, Singh A. Household food security in the United States in 2013. 2014. https://www.ers.usda.gov/webdocs/publications/45265/48787_err173.pdf. Accessed 5 Nov 2020.

Hanson KL, Connor LM. Food insecurity and dietary quality in US adults and children: a systematic review. Am J Clin Nutr. 2014;100(2):684–92. https://doi.org/10.3945/ajcn.114.084525. Comprehensive systematic review on dietary quality in food secure households.

Ritz B, Wilhelm M, Hoggatt KJ, Ghosh JK. Ambient air pollution and preterm birth in the environment and pregnancy outcomes study at the University of California, Los Angeles. Am J Epidemiol. 2007;166(9):1045–1052. doi:10.1093/aje/kwm181

Giovannini N, Schwartz L, Cipriani S, et al. Particulate matter (PM10) exposure, birth and fetal-placental weight and umbilical arterial pH: results from a prospective study. J Matern Fetal Neonatal Med. 2018;31(5):651–655. doi:10.1080/14767058.2017.1293032