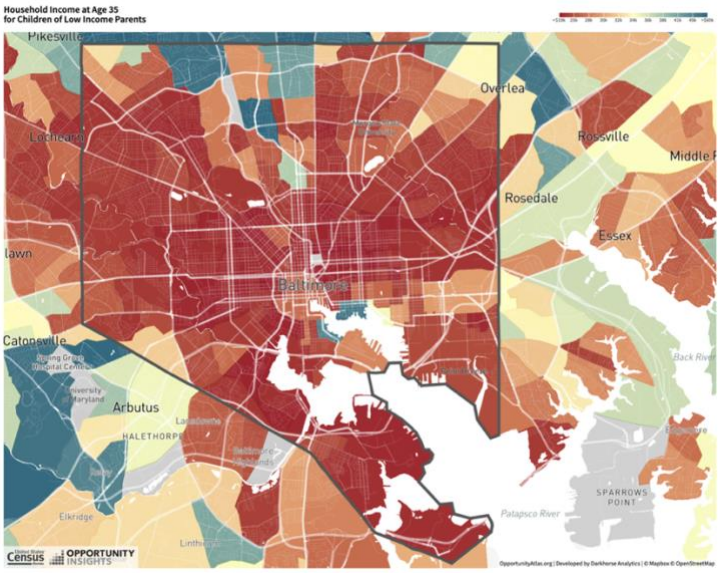
**Empirical Project 1**
**Stories from the Atlas: Describing Data using Maps, Regressions, and Correlations**
Helen (Yingying) Huang

1. **Start by looking up the city where you grew up on the Opportunity Atlas. Zoom in to the Census tracts around your home.**

   **Figure 1 in your narrative should be a map of the Census tracts in your hometown from the Opportunity Atlas. Examples for Milwaukee, WI (where Professor Chetty grew up) and Los Angeles, CA (discussed in Lecture 1) are shown on the next page. The text of your narrative should describe what you see, and what data are being visualized.**

   **Examine the patterns for a number of different groups (e.g., lowest income children, high income children) and outcomes (e.g., earnings in adulthood, incarceration rates). Only choose one or two of these to include in your narrative.**

Figure 1Household Income in Adulthood for Children Raised in Low-Income Households in Baltimore, MD




The Figure 1 presented is the Opportunity Atlas analytical tool that delineates the trajectory of economic mobility in Baltimore, Maryland—a city where I pursued my graduate studies. It visualizes the average annual household incomes of individuals who, now in their mid-thirties, grew up in low-income families, using data from 2014-2015. The

income levels are color-coded: red indicates lower incomes, while green denotes higher incomes, underscoring the economic divides across various census tracts within the city.

Central Baltimore reveals a preponderance of darker red hues, signifying a persistent cycle of poverty, which could stem from systemic issues such as inadequate schooling, limited financial resources, and a lack of job opportunities (The Baltimore Story, n.d.). In contrast, as moves towards the city's edges, a tapestry of green unfolds, signaling a transition to areas where individuals from low-income families appear to achieve greater financial success by age 35, likely benefiting from enhanced access to quality education and supportive community initiatives.

This map's portrayal of income distribution reflects deeper societal cleavages shaped by historical segregation and urban policy decisions (The Baltimore Story, n.d.). It serves as a stark reminder of the city's socio-economic challenges and potential, which stresses the need for policies tailored to the nuanced conditions of each area, aiming to bridge disparities and enhance opportunities for Baltimore's future generations.

2. **(To answer this question, read the Opportunity Atlas manuscript) What period do the data you are analyzing come from? Are you concerned that the neighborhoods you are studying may have changed for kids now growing up there? What evidence do Chetty et al. (2018) provide suggesting that such changes are or are not important? What type of data could you use to test whether your neighborhood has changed in recent years?**

The data under analysis originate from individuals born between 1978 and 1983 in Baltimore, Maryland. The analysis traces their annual household incomes in their mid-thirties, specifically during the years 2014-2015, utilizing Federal income tax records (Chetty et al., 2018). The possibility of neighborhood changes is indeed an important consideration, since neighborhoods can transform over time due to economic, social, and

political factors. These changes could potentially alter the conditions that previously influenced the studied outcomes. However, the authors, Chetty and his colleagues, provide evidence to suggest that while neighborhoods indeed change over time, these changes do not drastically affect the predictive power of their historical measures for current outcomes. They find that poverty rates from 23 years ago are 91% as predictive of current poverty rates as those from five years ago (Chetty et al., 2018). This suggests a relative stability in the factors associated with economic mobility over time, indicating that past measures of upward mobility can remain applicable over extended periods. To test whether a neighborhood has changed in recent years, one could use recent American Community Survey (ACS) data or other current socioeconomic and demographic data to compare with historical data. This would allow for an analysis of changes in key predictive factors such as poverty rates, employment rates, educational attainment, and housing values.

3. **Now turn to the atlas.dta data set. How does average upward mobility, pooling races and genders, for children with parents at the 25th percentile (`kfr pooled_p25`) in your home Census tract compare to mean (population-weighted, using `count_pooled`) upward mobility in your state and in the U.S. overall? Do kids where you grew up have better or worse chances of climbing the income ladder than the average child in America?**

   ***Hint*: The Opportunity Atlas website will give you the tract, county, and state FIPS codes for your home address. For example, searching for "Lynwood Road, Verona, New Jersey" will display Tract 34013021000, Verona, NJ. The first two digits refer to the state code, the next three digits refer to the county code, and the last 6 digits refer to the tract code. In Stata, listing this observation can be done as follows:**

```
list kfr_pooled_p25 if state == 34 & county == 013 & tract == 021000
```

The average upward mobility for children whose parents is at the 25th percentile (kfr_pooled_p25) in Tract 24510130600, Hampden, Baltimore, MD, is approximately $25,110.69. In comparison, the state of Maryland's average, calculated on a population-weighted basis using the count_pooled as a weighting factor, is about $34,359.27. For the

U.S. as a whole, it is around $34,293.48. Therefore, children in Tract 24510130600 have lower chances of climbing the income ladder compared to their state and national peers.

4. **What is the standard deviation of upward mobility (population-weighted) in your home county? Is it larger or smaller than the standard deviation across tracts in your state? Across tracts in the country? What do you learn from these comparisons?**

The standard deviation within Baltimore County stands at approximately $4,881.61. In contrast, the state of Maryland has a significantly higher standard deviation of approximately $8,200.24, while the national figure is marginally lower at around $7,897.46. These figures suggest a more homogeneous economic landscape within the county, which could be attributed to factors such as localized economic policies, demographic characteristics, or an equitable distribution of resources that contribute to leveling economic opportunities and fostering uniform economic outcomes.

5. **Now let's turn to downward mobility: repeat questions (3) and (4) looking at children who start with parents at the 75th and 100th percentiles. How do the patterns differ?**

In Baltimore's Hampden tract 24510130600, children from households in the 75th percentile typically face an income decline of $53,955.42. This decrease is slightly larger than the Maryland average of $51,459.32 and the national average of $51,257.34. For children from the wealthiest families, in the 100th percentile, the average income reduction in the tract is $89,514.55. This figure is considerably higher than the state and national averages, which are $69,302.75 and $69,168.07, respectively.

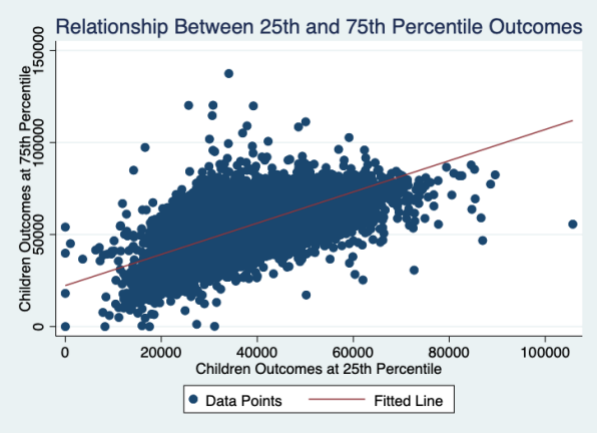
Regarding the standard deviation, which indicates the variability of income, it stands at $9,014.04 at the 75th percentile and $15,834.47 at the 100th percentile in Baltimore County. This suggests there is greater income fluctuation at the highest income level. By contrast, in Maryland, the variability is $9,984.78 at the 75th percentile and a bit

lower at the 100th percentile, at $15,224.32. Nationally, the standard deviations are $9,323.64 at the 75th percentile and the most considerable, $16,356.85 at the 100th percentile, indicating the widest income variability for the wealthiest across the U.S.

These patterns suggest that the tract in Hampden may have certain advantages or conditions that support children from wealthier families in maintaining or improving their economic status compared to the state and national averages. Additionally, the data reveals significant variation in outcomes within Baltimore County, particularly at the highest income level. Moreover, for children from the wealthiest families (those at the 100th percentile), there is a significant disparity in outcomes are evident both within Baltimore County and across the United States, with the national level exhibiting the greatest variability. Such disparities may reflect a broader spectrum of opportunities or economic risks present across different regions, potentially affecting the economic prospects of children from affluent families.

6. **Using a linear regression, estimate the relationship between outcomes of children at the 25th and 75th percentile for the Census tracts in your home county. Generate a scatter plot to visualize this regression. Do areas where children from low-income families do well generally have better outcomes for those from high-income families, too?**

Figure 2 Relationship Between 25th and 75th Percentile Outcomes




The linear regression analysis (Figure 3) reveals a positive and statistically significant relationship between the outcomes of children at the 25th percentile and those at the 75th percentile within the Census tracts in Baltimore County. The model has an R-squared value of 0.533, suggesting that about 53.3% of the variation in outcomes for children at the 75th percentile can be attributed to the variation in outcomes at the 25th percentile. However, this also means that a significant portion of the variance is not explained by the model, highlighting the presence of other influencing factors. The positive coefficient of 0.8482 for the 25th percentile outcomes implies that areas where children from lower-income families fare better are also associated with better outcomes for children from higher-income families, indicating a shared influence of certain factors across the socioeconomic spectrum. The intercept, representing the expected outcome for children at the 75th percentile when the 25th percentile outcome is zero, provides a baseline of $22,285.25 for such outcomes. While this analysis reveals a strong association, it does not imply causation, and it should be noted that the underlying factors driving this relationship require further investigation.

7. **Next, examine whether the patterns you have looked at above are similar by race. If there is not enough racial heterogeneity in the area of interest (i.e., data is missing for most racial groups), then choose a different area to examine.**

Figure 4 White Children: 25th vs. 75th Percentile Outcomes

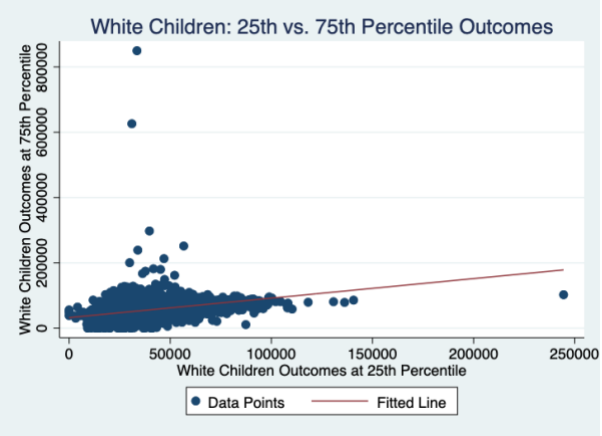
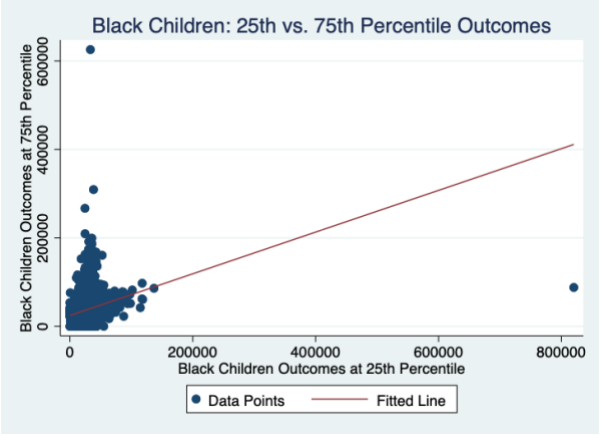



Figure 5 Black Children: 25th vs. 75th Percentile Outcomes




The regression results derived from Stata (Figure 3 & Figure 4) reveal modest associations between the 25th and 75th percentile economic outcomes within the populations of white and black children in Baltimore County. The determination coefficients (R-squared), 0.269 for white children and 0.0857 for black children, signify that the models explain 26.9% and 8.57% of the variance in the 75th percentile outcomes based on the 25th percentile outcomes, respectively. These values indicate weak predictive

power for both models, with the model for white children exhibiting a marginally stronger association compared to that for black children. This implies that other unmeasured factors are likely having a significant impact on the economic outcomes at the 75th percentile for both racial groups.

8. **Using the Census tracts in your home county, can you identify any covariates which help explain some of the patterns you have identified above? Some examples of covariates you might examine include housing prices, income inequality, fraction of children with single parents, job density, etc. For 2 or 3 of these, report estimated correlation coefficients along with their 95% confidence intervals.**

In examining the socioeconomic factors within the Census tracts of Baltimore County, it appears that certain covariates affect children's economic outcomes. A higher average rent for a two-bedroom apartment in 2015 was correlated with better outcomes for children at both the 25th percentile (coefficient: 0.3387; 95% CI: 0.3295 to 0.3479) and the 75th percentile (coefficient: 0.1284; 95% CI: 0.1202 to 0.1366). This suggests that areas with higher rents may provide environments conducive to economic mobility. Conversely, a greater proportion of single-headed households with children in 2010 was negatively associated with children's outcomes at the 25th percentile (coefficient: -0.5650; 95% CI: -0.5712 to -0.5588) and at the 75th percentile (coefficient: -0.5457; 95% CI: -0.5528 to -0.5386), indicating that single parenthood may contribute to lower economic outcomes. Job density (measured in square miles) in 2013 had no significant impact on outcomes at the 25th percentile but was slightly negatively related to the 75th percentile outcomes (coefficient: -0.0243; 95% CI: -0.0425 to -0.0060), which could reflect the complex relationship between job availability and economic success.

9. **Open question: formulate a hypothesis for why you see the variation in upward mobility for children who grew up in the Census tracts near your home and provide correlational evidence testing that hypothesis.**

**For this question, many covariates have been provided to you in the atlas.dta file, which are described under the "Characteristics of Census tracts" header in Table 1.**

**You are welcome to use outside data that are not included in atlas.dta, but this is *not* required. Diane Sredl has created a research guide for our class that contains links to other data sources. You may wish to read this tutorial on how to add variables to a data set in Stata.**

Hypothesis: "Educational attainment in 2010, measured by the proportion of residents with at least a college degree (frac_coll_plus2010), significantly predicts upward mobility for children who grew up near Census Tract 24510130600 in Hampden, Baltimore, MD."

Support for this hypothesis is provided by the correlational analysis: The results from Stata indicate a positive relationship between the educational attainment of adults and the economic advancement of children, with a coefficient of 0.498 at the 25th income percentile (kfr_pooled_p25) and 0.392 at the 75th percentile (kfr_pooled_p75), each situated within a narrow 95% confidence interval. Notably, the more pronounced correlation for lower-income children suggests that educational attainment is particularly influential for children from this income bracket. However, it is crucial to acknowledge that educational attainment, while significant, is just one of the many factors influencing upward mobility. This indicates the necessity for a more detailed investigation into the various elements that shape a child's economic prospects.

10. **Putting together all the analyses you did above, what have you learned about the determinants of economic opportunity where you grew up? Identify one or two key lessons or takeaways that you might discuss with a policymaker or journalist if asked about your hometown. Mention any important caveats to your conclusions; for example, can we conclude that the variable you identified as a key predictor in the question above has a causal effect (i.e., changing it would change upward mobility) based on that analysis? Why or why not?**

In analyzing the determinants of economic opportunity in Baltimore, it becomes evident that family structure and the educational environment play pivotal roles. There is a correlation between higher household incomes and lower rates of single-parent households with greater upward mobility, suggesting that policies promoting economic stability and supportive family environments may bolster economic prospects. Similarly, a higher percentage of college-educated adults within a community correlates with better outcomes for children, highlighting the significance of an educational culture. However, it is important to note that these findings are correlational and do not necessarily imply causation. They indicate potential areas for policy intervention, but warrant caution, as improving these indicators may not directly lead to better economic outcomes. Effective policies should be multifaceted, addressing the complex interplay of social and economic factors that influence opportunity. Policymakers must understand that while education and family support are significant, they are just parts of a broader network of variables that affect economic mobility.

## References


The Baltimore Story. (n.d.). 1910: Residential Segregation. https://www.thebaltimorestory.org/history-1/1910-residential-segregation

Chetty, R., Friedman, J., Hendren, N., Jones, M., & Porter, S. (2018). *The Opportunity Atlas: Mapping the childhood roots of social mobility*. National Bureau of Economic Research. https://doi.org/10.3386/w25147

Researchers at the Census Bureau, Harvard University, & Brown University. (n.d.). *The Opportunity Atlas.* https://www.opportunityatlas.org/