Stories from Atlas

IBUS 6101: Big Data for International Business

Assignment – 2 Part - 2

Rachel Aska

Abstract

This analysis examines the intergenerational economic mobility of children based on the neighbourhood's they grew up in and to understand it's effect on their incomes later on in adulthood. Specifically, we examine whether census tracts that produce better-than-expected outcomes for children from families at different income levels. In our case, we intend to find the similarities and differences between low-income (25th percentile) and high-income (75th percentile) families across our census tract (51059461100 Vienna, VA).

We used Linear Regression Model to understand the input and analyze the outcomes using Stata commands. Linear Regression is a supervised machine learning algorithm that learns from the labelled datasets. It helps us to understand the relationship between a dependent and an independent variable. It makes predictions based on the features of our datasets. The dependent variable for our analysis is children who grew up in families at the 75th percentile and the independent variable is 25th percentile of the parental income distribution within each census tract in Fairfax County, VA.

In our dataset (atlas_new_dta) we observe five races. So, we performed analysis on Asians, Whites, Blacks, Hispanic and Native American races. American society is a complex system in which race is deeply intertwined with most aspects of it's functioning. Understanding variation across racial groups is critical because residential segregation, discrimination, and differential access to resources mean that different communities may experience neighbourhoods fundamentally differently. Even families living in the same census tract may have access to different schools, different social networks, and different treatment by institutions.

Race	Coefficient	R^2	P-Value
White	0.4127	0.4151	0.000
Black	0.4155	0.1922	0.000
Asian	0.6784	0.1623	0.000
Hispanic	0.3281	0.1245	0.000
Native American	N/A	N/A	N/A

1. 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?

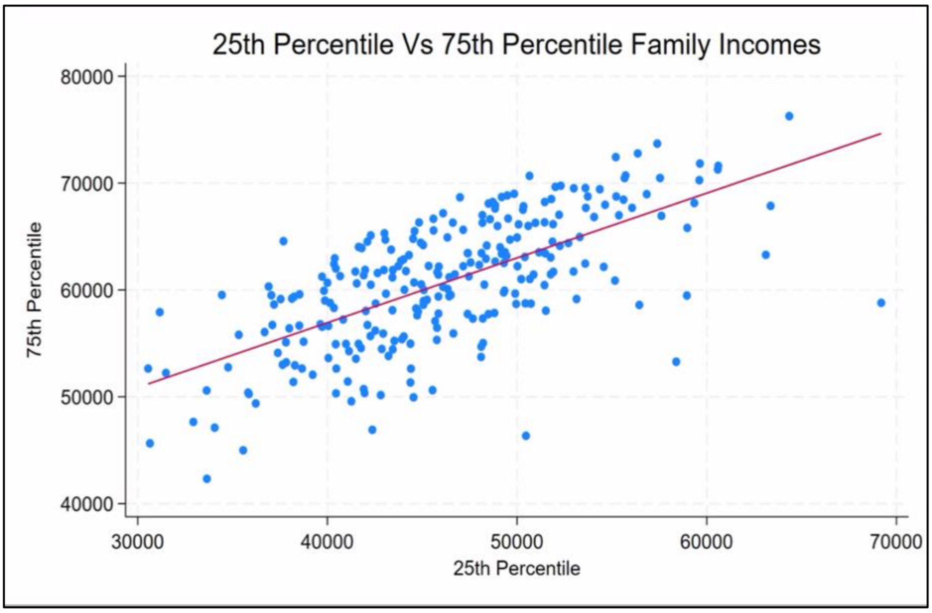
A: In our analysis, we observe the coefficient being less than 1.0 suggesting diminishing returns. The coefficient of 0.61 shows that when low-income children's outcomes improve by 1 percentile, high-

```
. /* Q1 Using a linear regression, estimate the relationship between outcomes of
>  children at the 25th and 75th percentile for the Census tracts in your home c
> ounty.  Generate a scatter plot to visualize this regression*/
.
. * Linear Regression for 25th and 75th Percentiles *
.
. regress kfr_pooled_p75 kfr_pooled_p25 if state=="51" & county=="059"

      Source |       SS           df       MS      Number of obs   =       255
-------------+----------------------------------   F(1, 253)       =    201.91
       Model |  4.1188e+09         1  4.1188e+09   Prob > F        =    0.0000
    Residual |  5.1611e+09       253  20399428.9   R-squared       =    0.4438
-------------+----------------------------------   Adj R-squared   =    0.4416
       Total |  9.2798e+09       254  36534690.7   Root MSE        =    4516.6

------------------------------------------------------------------------------
kfr_pooled~75 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------+----------------------------------------------------------------
kfr_pooled~25 |   .6060875   .0426541    14.21   0.000      .522085    .6900899
        _cons |   32693.54   1991.933    16.41   0.000     28770.66    36616.42
------------------------------------------------------------------------------
```

income children's outcomes improve by 0.6 percentiles. It indicates that children from low-income families acutely endure the effects of the neighbourhood. The R-squared value of 44.3% indicates that neighbourhood has a significant effect on the quality of life, though this is not the only factor. This effect is observed in more than just children from higher-income families. The remaining 56% reflects family-specific resources such as parental education and wealth, the individuals ability and network pools. As we see in the above interpretation, there are other underlying issues to factor in including neighbourhoods. This includes uniform access to education, safety and proximity to employment. The public and private schools of Fairfax County in Virginia are highly ranked according to Niche and there are reports of low crime rates in our tract according to Neighbourhood Scout. The access to the District's market is an additional positive add on. This inference aligns with our graph, we observe a positive slope with an upward trend. We see tight cluster's around the line with a couple of outliers. It is important to note that this



analysis reveals correlation rather than causality. Families pick where to reside based on their finances and preferences. Certain communities may be home to more motivated or well-off families, and rather than neighbourhood characteristics, family advantages may be the reason why their children excel.

2. Next, examine whether the patterns you have looked at above are similar by race. Draw a scatter plot for each race.

A:

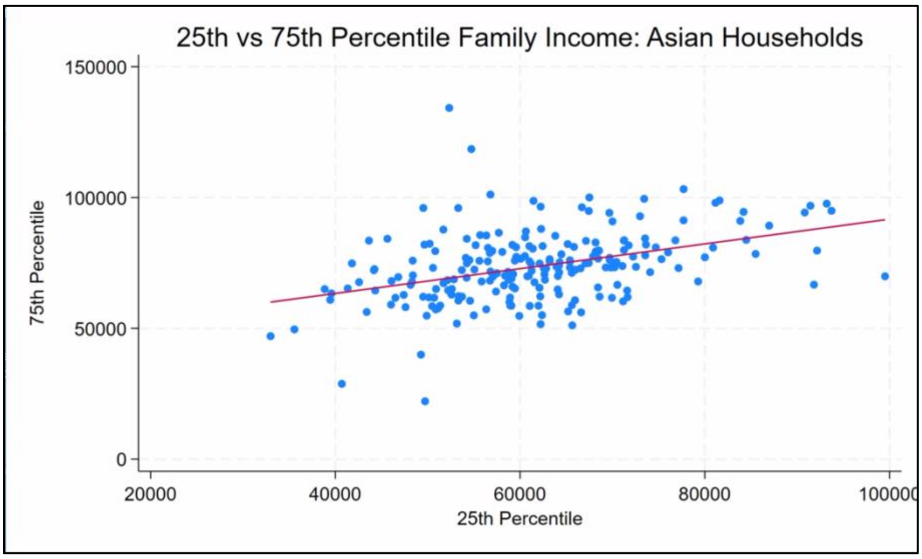
a. Asian: Asian households show something interesting; we see the highest coefficient (0.68) but a

```
. regress kfr_asian_p75 kfr_asian_p25 if state=="51" & county=="059"

      Source |       SS           df       MS      Number of obs   =       217
-------------+----------------------------------   F(1, 215)       =     42.84
       Model |  6.5989e+09         1  6.5989e+09   Prob > F        =    0.0000
    Residual |  3.3117e+10       215   154031031   R-squared       =    0.1662
-------------+----------------------------------   Adj R-squared   =    0.1623
       Total |  3.9716e+10       216   183868600   Root MSE        =     12411

------------------------------------------------------------------------------
kfr_asian_p75 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------+----------------------------------------------------------------
kfr_asian_p25 |    .473437   .0723318     6.55   0.000     .3308669    .6160072
        _cons |   44403.33   4535.723     9.79   0.000     35463.15    53343.51
------------------------------------------------------------------------------
```

very low R-squared (16%) value. This seems contradictory but actually reveals something important. The high coefficient means that when a neighbourhood is better (measured by 25th percentile outcomes), both poor and rich Asian children benefit roughly equally. But the low R-squared means that neighbourhood quality doesn't tell us much about which Asian families will succeed. This happens because Asian neighbourhoods contain families with very different outcomes living side by side. . Asian neighbourhoods often group people by ethnicity rather than income. We might see some recent immigrants have advanced degrees but work in low-paying jobs while they establish themselves. Others that have been here for generations and are very successful, so you might have a restaurant owner living next to a software engineer. Familial education and cultural attitudes toward schooling seem to matter more than the neighbourhood itself. This indicates that there are other corresponding factors



which are affecting the model which are not being used as a part of our analysis. The scatter plot for

Asian households shows points spread widely around the trend line with many outliers. This pattern shows what the low R-square tells us; outcomes vary dramatically even within the same neighbourhoods. This indicates In Fairfax County, Asian households show income convergence, areas where low-income (25%) Asian families do better also tend to be areas where higher-income (75%) Asian families do better. Asian families have diverse economic conditions. The tight-knit nature of Asian communities means neighbourhoods often group by ethnicity and culture rather than by income, creating this diversity. The importance to education and rapport within their communities is a driving force for Asian families rather than their neighbourhood of residence.

b. White:

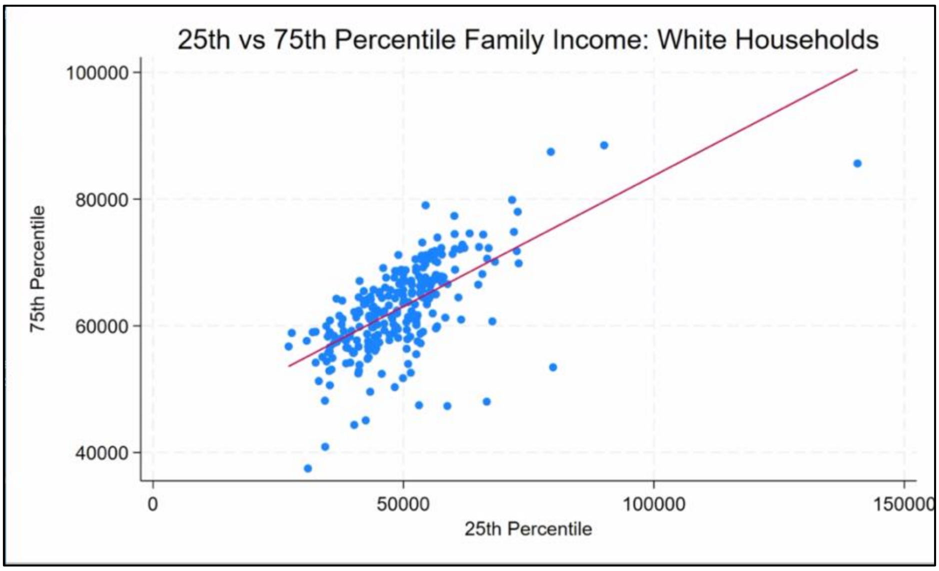
White households show the strongest R-square value of 41%, this means that if you know a neighbourhood's characteristics, you can predict individual White families' outcomes well. This is substantially higher than what we observed for Asian households (16.23%), suggesting that neighbourhood level 25th percentile income is a much stronger predictor of 75th percentile outcomes for White families. They face less discrimination in housing and can afford to move to neighbourhoods that match their income level. This creates neighbourhoods where most White families have similar incomes and resources. The coefficient of 0.41 shows that when outcomes for low-income White children improve by 10

```
. regress kfr_white_p75 kfr_white_p25 if state=="51" & county=="059"
```

Source	SS	df	MS
Model	5.4485e+09	1	5.4485e+09
Residual	7.6759e+09	253	30339502.1
Total	1.3124e+10	254	51670913.7

Number of obs	=	255
F(1, 253)	=	179.58
Prob > F	=	0.0000
R-squared	=	0.4151
Adj R-squared	=	0.4128
Root MSE	=	5508.1

kfr_white_p75	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
kfr_white_p25	.4127293	.0307986	13.40	0.000	.3520751	.4733835
_cons	42430.64	1558.889	27.22	0.000	39360.59	45500.7



percentiles, high-income White children's outcomes improve by about 4 percentiles. We see a positive upward trend with notably tighter clustering around the regression line compared to the Asian household scatter plot, with fewer outliers and less dispersion. This indicates that in Fairfax County, White households show stronger income convergence patterns areas where low-income (25%) White families do better are much more reliably areas where higher-income (75%) White families also do better. Unlike the Asian households, the 25th percentile income provides considerably more predictive power about what the 75th percentile will be in any given neighbourhood.

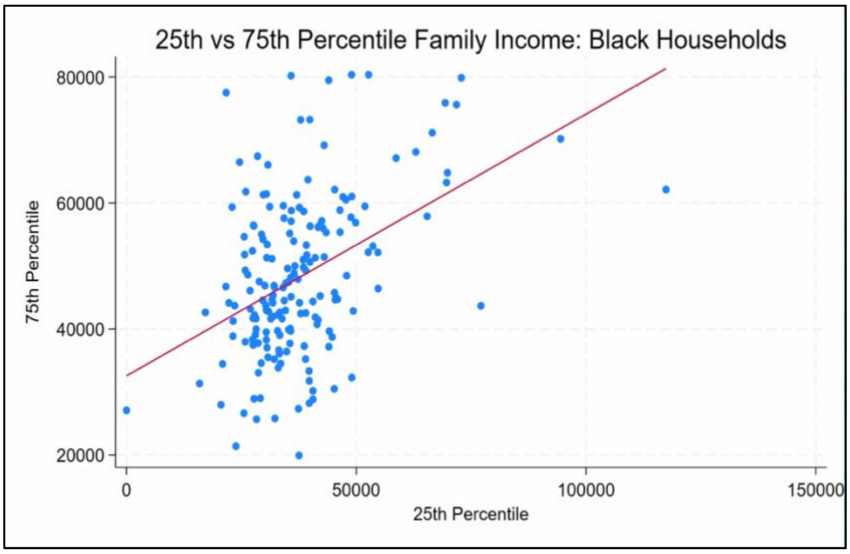
c. Black:

Black households show moderate R-squared (19%), like Asian families but for different reasons. Black families across all income levels often live in similar neighbourhoods because of housing discrimination. Even wealthy Black families face barriers moving to predominantly wealthy neighbourhoods, while poor Black families may live in the same areas due to limited options. This means Black neighbourhoods contain families with a wider range of incomes than White neighbourhoods. Some families are struggling while others are quite successful, all living in the same area. The coefficient of 0.42 shows neighbourhoods do matter - better neighbourhoods help both poor and rich Black children. But the low R-squared shows that family characteristics and individual factors matter more than neighbourhood in explaining why some Black children succeed while others don't. Research shows Black

```
. regress kfr_black_p75 kfr_black_p25 if state=="51" & county=="059"

      Source |       SS           df       MS      Number of obs   =       180
-------------+----------------------------------   F(1, 178)       =     42.34
       Model |  5.5849e+09         1  5.5849e+09   Prob > F        =    0.0000
    Residual |  2.3480e+10       178   131910208   R-squared       =    0.1922
-------------+----------------------------------   Adj R-squared   =    0.1876
       Total |  2.9065e+10       179   162373927   Root MSE        =     11485

-------------------------------------------------------------------------------
kfr_black_p75 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------+----------------------------------------------------------------
kfr_black_p25 |   .4154556   .0638492     6.51   0.000     .2894568    .5414544
        _cons |    32558.5   2550.385    12.77   0.000     27525.62    37591.38
-------------------------------------------------------------------------------
```



families often need much higher incomes than White families to access the same neighbourhood quality, and even in good neighbourhoods, they may face continued discrimination. The scatter plot for Black households shows considerable spread around the regression line, indicating high diversity within neighbourhoods. Black families at different income levels often end up in similar neighbourhoods, not by choice but because of discrimination in housing markets.

d. Hispanic:

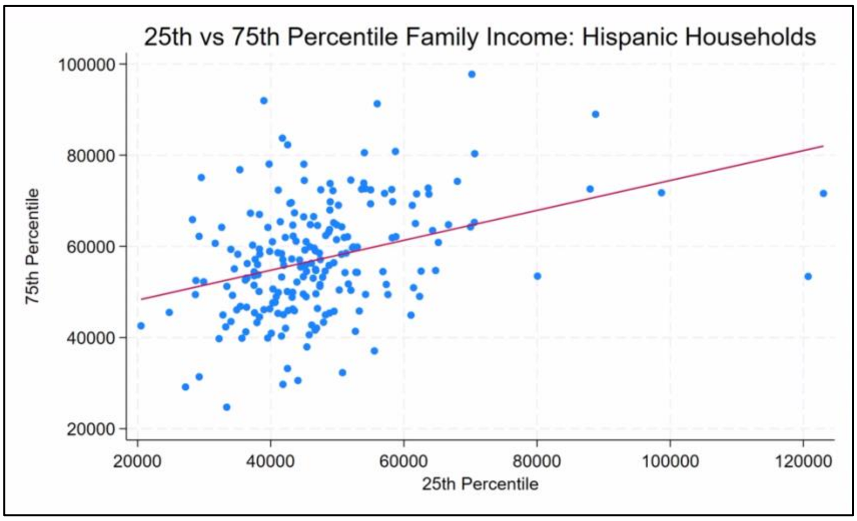
Hispanic households show the weakest neighbourhood effects with only 12.4% (R-square) of outcomes explained by neighbourhood characteristics. This means neighbourhoods tell us very little

```
. regress kfr_hisp_p75 kfr_hisp_p25 if state=="51" & county=="059"

      Source |       SS           df       MS      Number of obs   =       205
-------------+----------------------------------   F(1, 203)       =     28.86
       Model |  3.9266e+09         1  3.9266e+09   Prob > F        =    0.0000
    Residual |  2.7624e+10       203   136080590   R-squared       =    0.1245
-------------+----------------------------------   Adj R-squared   =    0.1201
       Total |  3.1551e+10       204   154661693   Root MSE        =     11665

------------------------------------------------------------------------------
kfr_hisp_p75 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
kfr_hisp_p25 |   .3280866   .0610769     5.37   0.000     .2076602     .448513
       _cons |   41644.13   3000.639    13.88   0.000     35727.71    47560.54
------------------------------------------------------------------------------
```

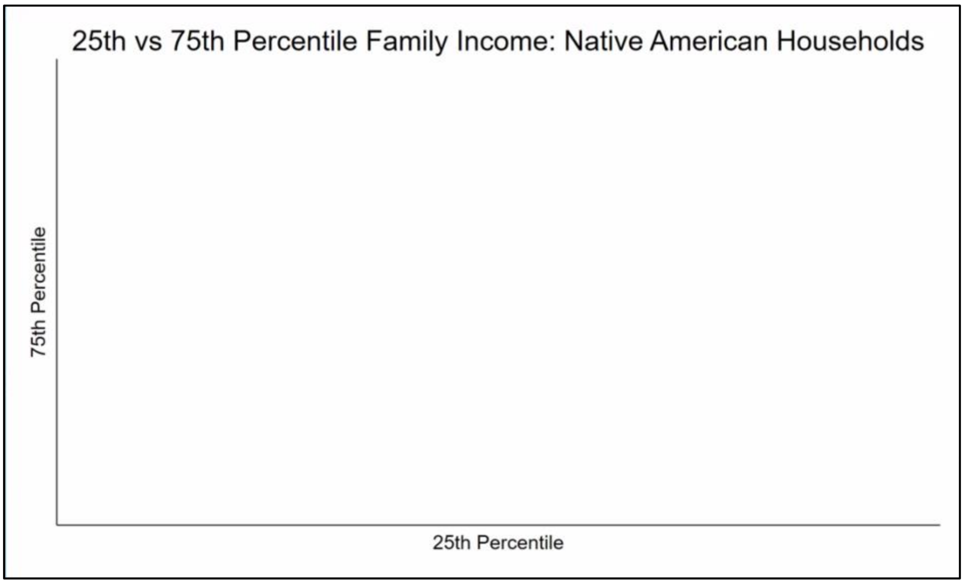
about which Hispanic children will succeed. Their children's success depends more on parents' education than the neighbourhood. Second or third-generation families might have very different experiences. Hispanic neighbourhoods often have strong business and social networks that help people find jobs and support each other, regardless of what the neighbourhood looks like on paper. The coefficient of 0.33 indicates that when low-income Hispanic children's outcomes improve by 10 percentiles, high-income Hispanic children's outcomes improve by about 3 percentiles. This is the lowest R-squared value among all racial groups examined (Asian: 16.23%, Black: 19.22%, White: 41.51%), suggesting that neighbourhood-level 25th percentile income is the weakest predictor of 75th percentile outcomes for Hispanic families. We see a positive upward trend but with substantial scatter throughout the distribution, with numerous outliers dispersed across all income levels, creating a notably diffuse cloud of data



points around the regression line. This indicates that Hispanic households show the weakest income convergence patterns among all groups studied, while areas where low-income (25%) Hispanic families do better tend somewhat to be areas where higher-income (75%) Hispanic families also do better, this relationship is considerably weaker. While there's an upward trend, it's weak and unreliable for predicting individual outcomes. This shows that Hispanic children's success depends much more on family and individual factors than on neighbourhood characteristics.

e. Native American:

```
. regress kfr_natam_p75 kfr_natam_p25 if state=="51" & county=="059"
no observations
r(2000);
```



We were unable to perform analysis for Native American households in our tract due to insufficient data. The dataset contains few observations for this demographic group in this county, preventing any analysis of the relationship between 25th and 75th percentile income outcomes. This absence itself tells us something important about inequality and exclusion. Native American families might be missing from Fairfax County because of historical displacement from their ancestral lands, colonization, forced removal or because data collection methods miss this population.

Washington D.C. region was originally Native American land before forced removal. The fact that we can't even study Native American families' economic mobility today reflects centuries of marginalization and prejudices.

Introspection

Our analysis reveals that neighbourhoods affect racial groups differently. White families show strong neighbourhood effects (41%), Hispanic families show the weakest neighbourhood effects at only 12%, followed by Asian (16%) and Black (19%) families. This variation reflects differences in residential sorting patterns, housing discrimination, and the relative importance of family versus place-based factors.

We observe a consistent pattern of coefficients below 1.0 across all racial groups provides strong support for progressive geographical investment. As mentioned previously, low-income children are more sensitive to neighbourhood quality than high-income children regardless of race. This means that investing in disadvantaged neighbourhoods produces the greatest benefits for children who need help most.

Neighbourhoods matter for children's economic mobility, but they matter differently for different communities. Effective policy cannot be a one-size-fits-all approach, we need to understand the strong and weak points of each race. The goal should not be forcing all communities into a single pattern but rather ensuring that all children, regardless of race or family income, have access to the resources and opportunities needed for upward mobility.

References:

- American Library Association. (n.d.). *Indigenous tribes of Washington, DC*. https://www.ala.org/aboutala/indigenous-tribes-washington-dc
- Chetty, R., Friedman, J. N., Hendren, N., Jones, M. R., & Porter, S. R. (2018). *The Opportunity Atlas: Mapping the childhood roots of social mobility* (NBER Working Paper No. 25147). National Bureau of Economic Research. https://doi.org/10.3386/w25147
- City of Fairfax Police Department. (n.d.). *Crime map*. https://www.fairfaxva.gov/Government/Departments/Police/Crime-Map
- Garsd, J. (2018, October 4). *The American dream: One block can make all the difference*. NPR. https://www.npr.org/2018/10/04/654085265/the-american-dream-one-block-can-make-all-the-difference
- Guerrieri, V. (2020). *Urban economics lectures* [Lecture slides]. Urban Economics Association. https://urbaneconomics.org/workshops/lectures2020/slides/UEA_Lectures_Guerrieri.pdf
- U.S. Census Bureau. (2010). *2010 Census summary file 1*. U.S. Department of Commerce. https://www.census.gov/data.html
- NeighborhoodScout. (n.d.). *Fairfax, VA crime rates*. https://www.neighborhoodscout.com/va/fairfax/crime
- Niche. (n.d.). *Fairfax County Public Schools rankings*. https://www.niche.com/k12/d/fairfax-county-public-schools-va/rankings/
- U.S. Census Bureau & Opportunity Insights. (2018). *The Opportunity Atlas: Mapping the childhood roots of social mobility in America*. https://opportunityatlas.org
- StataCorp. (2023). *Stata Statistical Software: Release 18*. College Station, TX: StataCorp LLC.