**Creating CommunityScapes: Using Data Science to Inform Evidence-Based**

**Health Policy in Fairfax County, Virginia**

DSPG team: Eliza Tobin, Cong Cong, Victoria Halewicz, Quinton Neville

SDAD team: Teja Pristavec, Josh Goldstein, Brandon Kramer, Stephanie Shipp, Joy Tobin

[I think this work really hits at the value of using social data for the public good, and I’m glad to see you didn’t overextend your findings to make universal claims about the human condition. And I’m also glad you were though – and specific, one of my favorite attributes – about some of what you found in Fairfax. That said, I do think we need some statements – they can be measured – about how this approach would work on a different or larger stage. And if we did go down that road, I might want a pocket description of Fairfax County (although your wise use of graphics and descriptives does give me a strong sense once I’m done. And I’m particularly comfortable making those suggestions because your already make these links early on (and you’ll see my admiration there!). I just feel we can follow through with that in the finale.]

**Introduction**

Over the past three decades, social scientists and public health scholars documented how [socio-economic factors and poor access to resources](https://www.who.int/social_determinants/corner/en/) affect the health of millions of Americans today. Lower levels of income, education, and wealth, often coupled with limited access to health insurance and quality healthcare facilities, place poorer communities at a higher risk of [health morbidities and mortality.](https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-health/interventions-resources) For example, communities of color, especially those in segregated urban and remote rural areas, often lack preventive health care resulting in higher rates of illness and disease as residents age. These contexts make it harder to acquire quality nutrition and engage in physical exercise, which can contribute to rising obesity rates over time.

The Social and Decision Analytics Division at [University of Virginia](https://biocomplexity.virginia.edu/social-decision-analytics-sdad)' s Biocomplexity Institute and Initiative has partnered with Fairfax County and Inova Translational Medicine Institute to better understand regional health and well-being through a data science framework. [I appreciate that you made explicit the institutional framework under which you operated. This needs to be explicit as each of these are posted and this is much more elegant than an editor’s note.] Prior research made considerable strides exploring how economic vulnerability shapes health outcomes at the national, state, or county level. However, there is a growing need to scale research to inform policies at smaller geographic levels that are actionable for stakeholders. Our research team used an interdisciplinary, data-driven to discover, acquire, and statistically integrate publicly available local data and create [CommunityScapes](https://repository.upenn.edu/admindata_conferences_presentations_2018/44/) – social, economic, environmental, and well-being indicators at sub-county levels.

**Data and Methodology**

Using the literature on the [social determinants of health](https://www.cdc.gov/socialdeterminants/research/index.htm) to guide our analysis, we set out to understand which mechanisms drive health risks in Fairfax County, Virginia. Social determinants of health and health disparities research more generally is typically conducted using national, state, or even country-level datasets. Our project is innovative in that it draws from multiple publicly available data sources to examine three distinct and small sub-county geographical units for patterns of economic vulnerability and obesogenic environment exposure. In doing so, our analyses are honed specifically for policymakers at supervisor districts and high school attendance areas rather than by administrative groupings like census tracts. We are able to provide stakeholders with actionable insights at a more meaningful geographic level. [very smart to make explicit statements of novelty and utility – always super useful for the casual reader (as most policymakers, as opposed to staff, will be).]

Our study drew from three primary data sources: the American Community Survey ([ACS](https://www.census.gov/programs-surveys/acs)), the Fairfax County Housing Stock, and OpenStreetMap ([OSM](https://www.openstreetmap.org/#map=5/38.007/-95.844)) data. Given that federal surveys are typically not available at sub-county levels with sufficient granularity, our first challenge was constructing an ACS-based synthetic population for Fairfax County. The synthetic population enabled us to infer the characteristics and location of individuals in Fairfax County that we could then reaggregate to new geographies. First, we combined the synthetic population data with publicly available Fairfax housing stock data, and reaggregated ACS census-tract level data to high school attendance areas and supervisor districts. Second, we explored the food and physical environment of Fairfax County, including locations of fast food stores, playgrounds, and supermarkets, using OSM. OSM is an open source mapping project that can be thought of as a combination of Wikipedia and Google Maps. We wrote functions to easily retrieve, reshape, and prepare OSM data for use, and then calculated 20-minute bus [travel time isochrones](https://en.wikipedia.org/wiki/Isochrone_map) from each facility or service of interest. The travel time isochrones describe the environment that contributes to obesity by determining whether a property is within a 20-minute bus trip to healthy food sources and physical exercise opportunities. Figure 1 shows an example of such a supermarket access polygon.

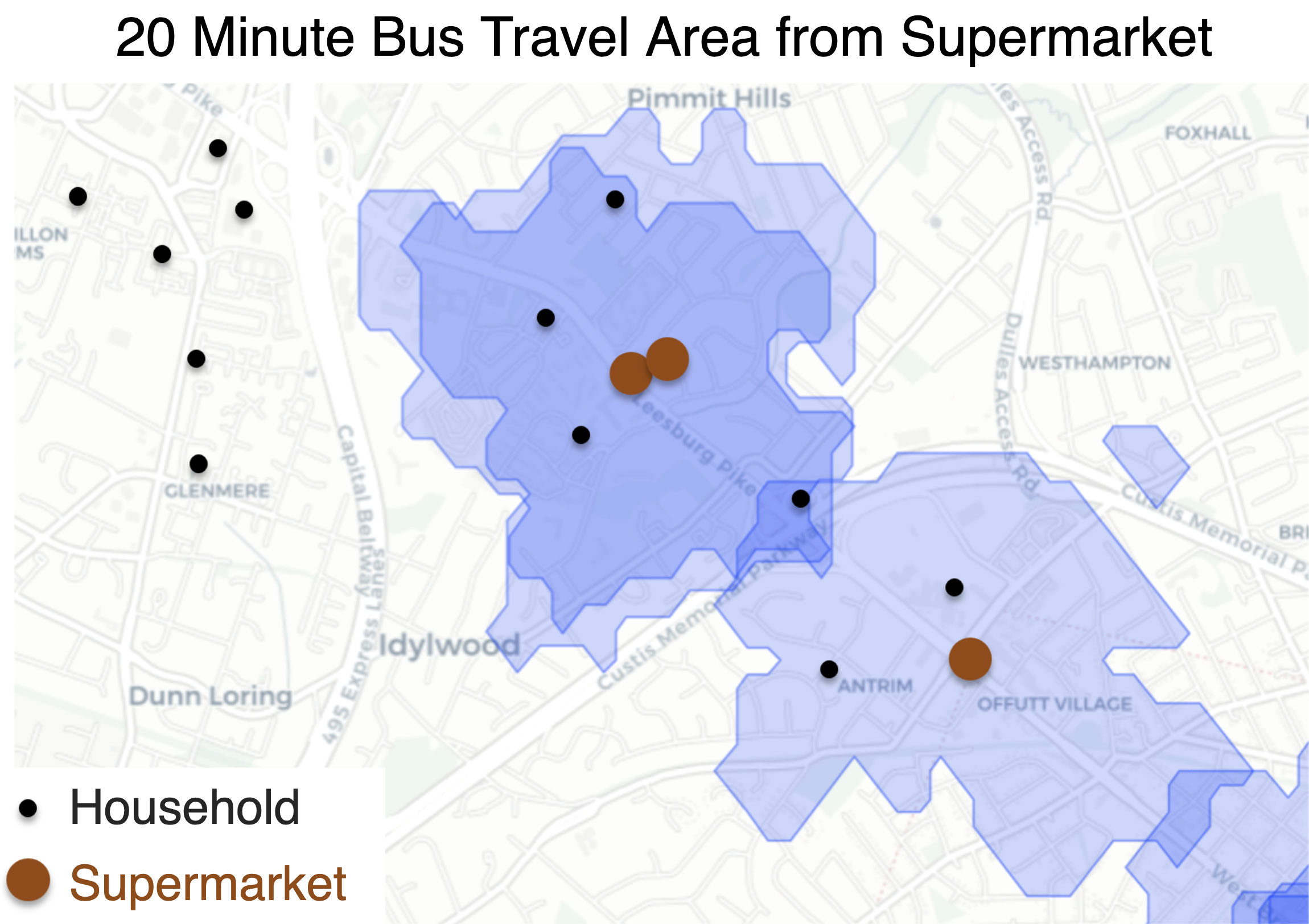


Figure 1: Example of Households Within and Outside a 20-Minute Bus Travel Time Isochrone.

Next, to build the CommunityScape, we created two [composite indices](https://composite-indicators.jrc.ec.europa.eu/?q=10-step-guide) using our variables to measure *economic vulnerability* and *obesogenic environments*. Rather than simply aggregating measurable variables to generate a relative index score, composite indices provide a robust strategy for identifying unobserved factors and adjusting features by their underlying dimensional importance. We used principal component and factor analyses to quantify both feature and factor importance, using factor loadings and percentage of variance explained respectively. We obtained final results using a weighted aggregation scheme, normalized to a scaled score from zero to one, where one describes the most vulnerable area and zero the least.

Our economic vulnerability index highlights areas with high proportions of individuals who are, for example, unemployed, live in poverty, or do not own a vehicle. Our [obesogenic environments](https://www.ncbi.nlm.nih.gov/pubmed/28449616) exposure index identifies areas where residents do not have convenient access to supermarkets and team sports facilities, or live in the vicinity of fast food restaurants. We generated maps to visually decompose these results at the census tract, high school attendance area, and supervisor district levels, showing the most and least vulnerable areas in Fairfax for each respective index. Figures 2 and 3 show examples of economic vulnerability mapped to Fairfax County [high school attendance areas](https://data-fairfaxcountygis.opendata.arcgis.com/datasets/high-school-attendance-areas) and of obesogenic environment exposure mapped to [supervisor districts](https://data-fairfaxcountygis.opendata.arcgis.com/datasets/1e026c9a9605454ab6b1515e6896216a_17). The CommunityScape points to Arlington-adjacent areas and those south of Arlington as areas with at risk populations. Residents of Annandale, Justice, and Mount Vernon high school attendance areas are most economically vulnerable; those living in Mason, Providence, and Lee supervisor districts have the highest obesogenic environment exposure.

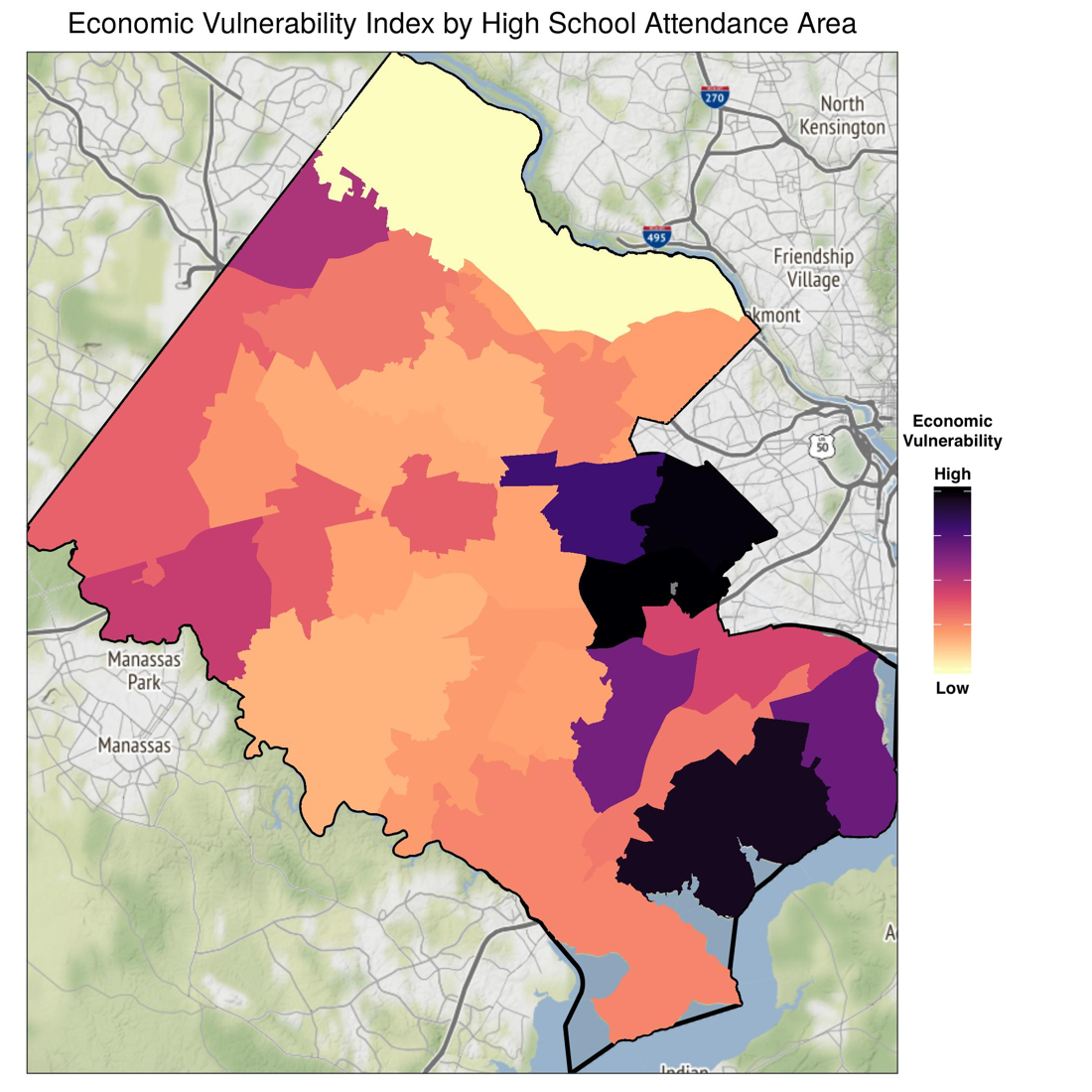


Figure 2: Economic Vulnerability in Fairfax County by High School Attendance Area.

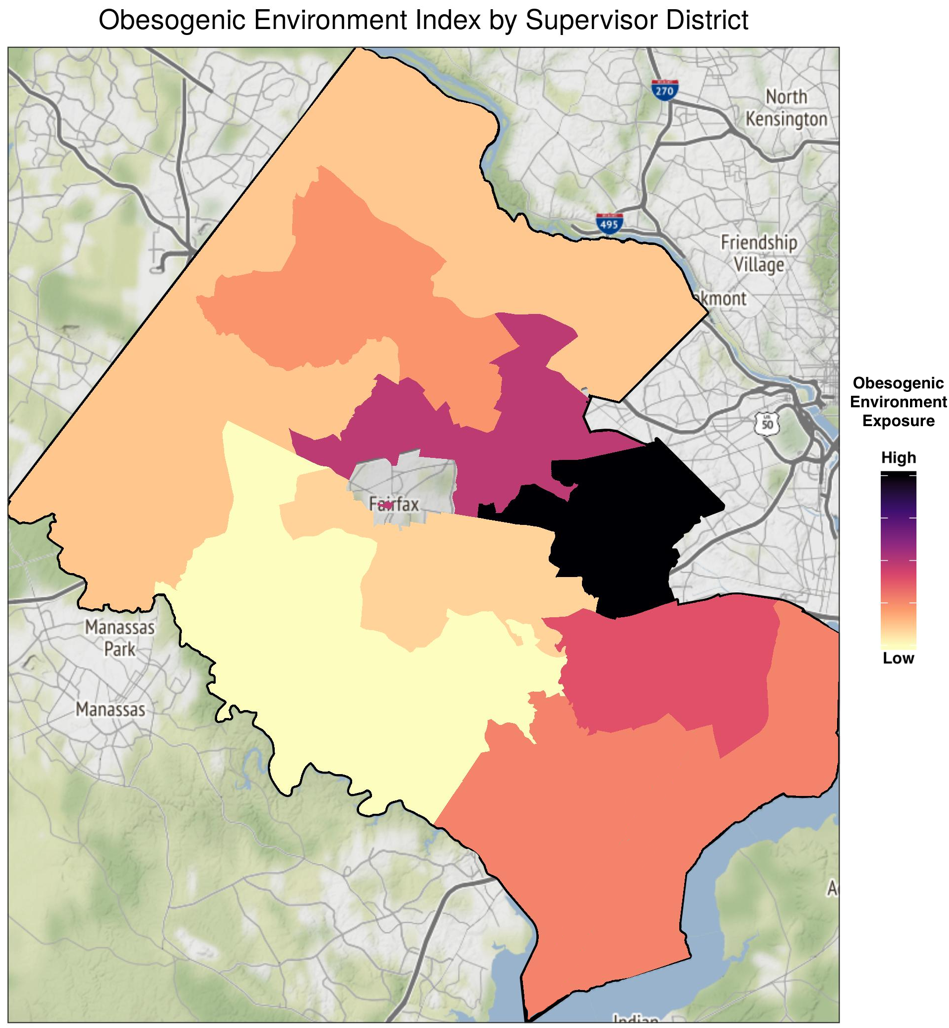


Figure 3: Obesogenic Environment Exposure in Fairfax County by Supervisor District.

**Findings and Future Directions**

Our CommunityScapes offer legislators a novel tool for developing policies that span multiple geographic levels. While past research typically explores disparities at the national or state level, our work highlights economic and social risk factors at actionable sub-county geographies. Similarly, policy is often created at a national and state level, but local policymakers like boards of supervisors have the ability to influence resource distribution across sub-county populations. For example, both the economic vulnerability and obesogenic indices suggest that the Mason and Lee supervisor districts in Fairfax County are areas with the highest risk of poor outcomes, suggesting that legislators should tailor their policies to address mechanisms that reproduce health inequalities in those specific areas.

Moreover, our factor analyses suggest that multiple intersecting social and economic factors, such as living in poverty, not having health insurance, and not having a high school degree increase the health burdens of individuals living in those districts. Our model also points to challenges for racial and ethnic minorities, especially those without good ability to speak English proficiently, as these groups have an increased risk of being both economically vulnerable and living in obesogenic environments.

These factors provide lawmakers important insights into policies that directly address the needs of at-risk populations. For example, Fairfax County policymakers may consider implementing English as a Second Language programs, low or zero cost childcare initiatives, as well as GED programs for parents that are available in both English and Spanish. Our sub-county analyses illustrate the need for policymakers to look at populations from various viewpoints to gain a full understanding of the resources distribution across the county. Analyzing health inequities using quantitative methods provides decision-makers with new, vital tools for evidence-based policymaking.

Moving forward, we plan to enhance the CommunityScape to better inform policymaking for Fairfax County residents. To verify the robustness of our composite indices, the team will conduct sensitivity analyses of our model findings. We will continue to collaborate with our sponsors, [Fairfax County](https://www.fairfaxcounty.gov/health-humanservices/) and the [Inova Translational Medicine Institute](https://www.inova.org/itmi/home) to contextualize the priorities of residents and policymakers and further refine our interpretations of which areas are most in need of support.

Additionally, through discussions with sponsors and deeper policy research, we can gain a better understanding of current policy measures that are working and past initiatives that succeeded or failed to address health and prosperity. Overlaying Inova healthcare facilities’ locations on the CommunityScape will create a more informed picture of individuals’ ability to access care. The current CommunityScape provides an overview of the distribution of opportunity across Fairfax County, but further measures will provide even more insight into the resources and demographic factors that impact one’s ability to [live, work, and play in a healthy environment in Fairfax County](http://www.livehealthyfairfax.org/).