

*Mid-Atlantic Statistics Network (MASN)*

*"Data Re-Use in Action"*

*2022-05-06*

# **SOCIAL IMPACT DATA COMMONS**

*an innovative approach to inform equitable growth*

*Dr. Aaron Schroeder*

*Social Decision Analytics Division (SDAD) of the UVA Biocomplexity Institute*



**BIOCOMPLEXITY  
INSTITUTE**



**Center for  
Inclusive Growth**

# Me

## Aaron Schroeder, PhD

Research Associate Professor and Data Scientist, Social Decision Analytics Division (**SDAD**) of the **UVA Biocomplexity Institute**

- Overarching research focus is the enablement of Evidence-Based Policy-Making and Program Evaluation through the **Secure Liberation, Integration and Analysis of All Available Data**
- Background is Public Policy & Public Administration, long-term research focus on **Data**
  - Learned early-on that “good” policy and program-level decision-making relies on good data, but that many times the data was inaccessible, was unusable in its current state, or simply didn’t exist
  - Have spent much of my career trying to address these very issues
    - Creating the privacy-protected probabilistic linking mechanism to federate data between multiple state agencies for the **Virginia Longitudinal Data System (VLDS)**
    - Integrating all available traffic incident, construction, and travel data into a singular access point, **511 Virginia**
    - Creating a data collection system, the **Virginia Early Childhood Professional Registry**, to collect policy-relevant credential data not previously collected
- Lead our **Research Infrastructure, Data & Development Team**
  - Overall Responsibility for Software Development and the Scientific Computing Infrastructure at SDAD
- Co-director of the **Data Science for Public Good (DSPG)** Young Scholars Program
  - immersive summer data science program for undergraduate and graduate students from across the country
- Principal Investigator on many social-good-oriented data integration projects, including today’s focus project, the **Social Impact Data Commons**

# SOCIAL IMPACT DATA COMMONS

## Agenda

- PROJECT VISION
- PROJECT TIMELINE
- PROJECT APPROACH
- AN EXAMPLE FOLLOWING APPROACH FOR BROADBAND EQUITY
  - OTHER EXAMPLES (VERY BRIEFLY)
- ARCHITECTURE

# VISION

## INFORMING EQUITABLE GROWTH



**BIOCOMPLEXITY  
INSTITUTE**



Center for  
Inclusive Growth

The **University of Virginia** and the **Mastercard Center for Inclusive Growth** have a shared vision to use data to inform equitable growth.

Local communities have data on policies, strategies, events and social behaviors but often lack the analytical tools to use their data to drive policy and strategy development.

**Partnering,  
we can make  
a difference.**

# VISION

## WHAT IS A DATA COMMONS?

An open knowledge repository that co-locates data from a variety of sources, builds and curates data insights, and provides tools designed to track issues over time and geography allowing governments and community stakeholders to learn continuously from their own data.



### *Key features:*

- Data sources, collected and created
- Maps reflecting multiple geographies
- Composite metrics
- Navigation and capability to statistically explore the data
- Data download via web or API
- Metadata

**A Data Commons allows multiple audiences to explore issues relevant to their communities.**

# PROJECT TIMELINE

## Project Phases

**Q1-Q2** Ground design and initial deployment on Arlington County focusing on select issues of interest

**Q3-Q4** Expand geographic coverage to City of Alexandria, Washington, D.C., and Fairfax County

**Q5-Q8** Expand to entire Washington, D.C. metro region



# PROJECT APPROACH

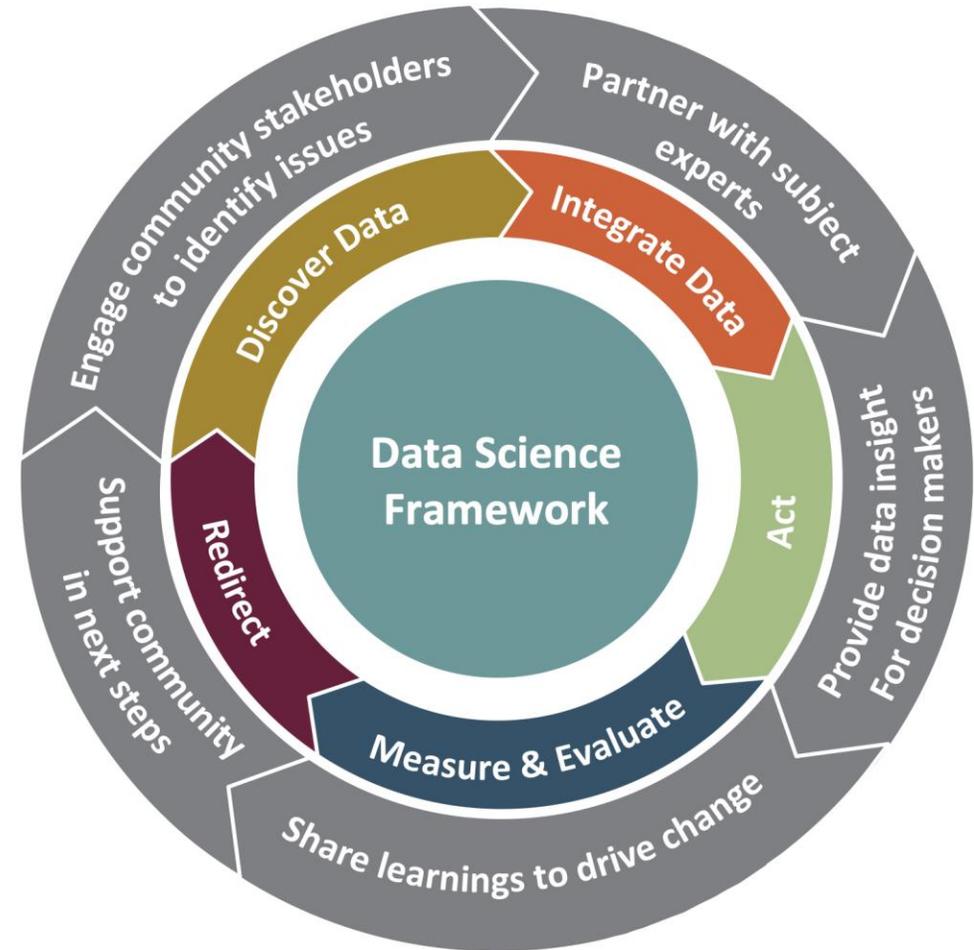
## Initial Grounding in Specific Local Issues

- To kickoff the project, the team worked in partnership with key Arlington County agency personnel, driven by the tireless efforts of CDO Jaime Lees, to identify several local social impact issues
- These issues were selected on the basis that they can **immediately benefit from the provision of new data and metrics** and serve as **powerful exemplars** to showcase the impact and values of the Social Impact Data Commons as the project expands
- Initial Issues:
  - **Equity of Broadband Access**
  - **Equity and Stability of Access to Staple Foods**
  - **“Other-Than Census” Geographies for Equity Analysis**

# PROJECT APPROACH

## Iterative Process with Decision Makers Getting to Required Data & Metrics

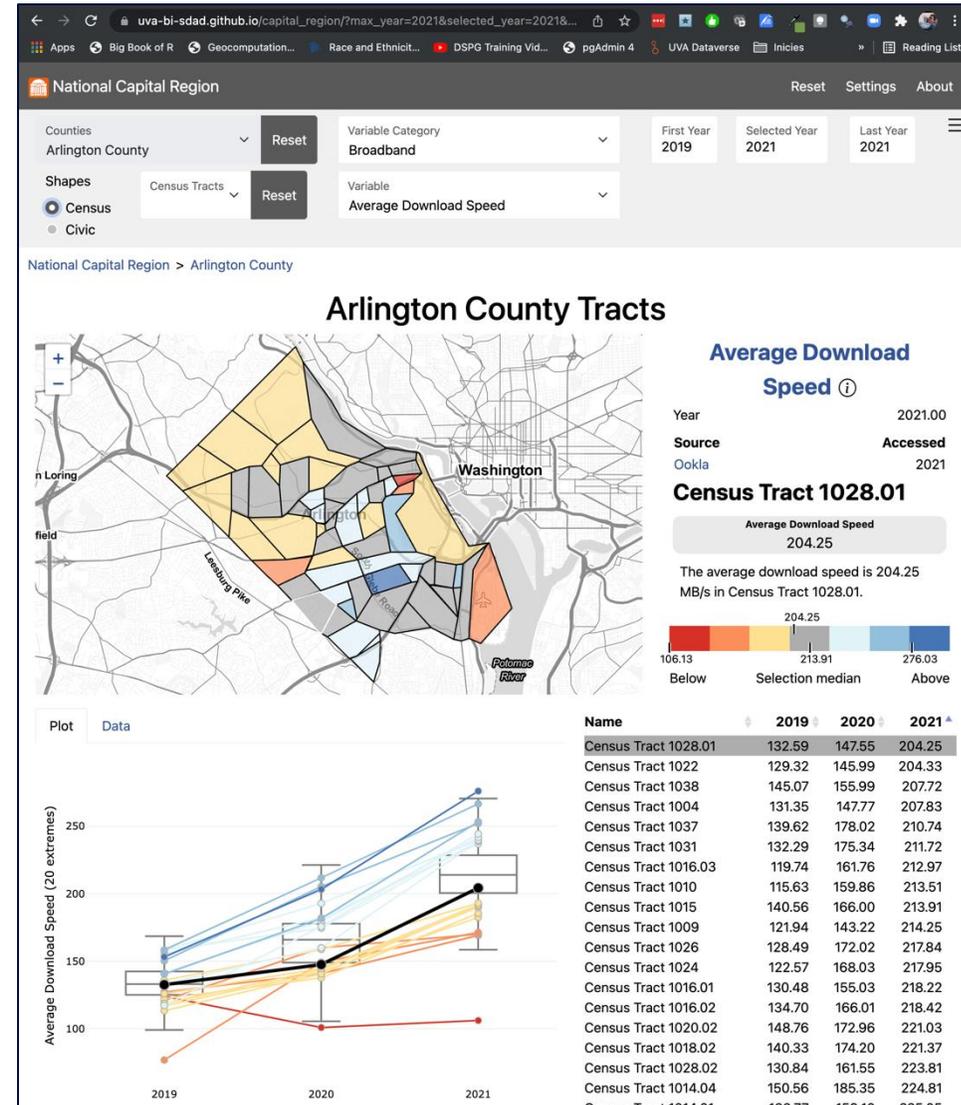
- **Community Learning Through Data Driven Discovery (CLD3)**
- **CLD3** goes beyond traditional organizing aspects of collective action programs and supports communities in building capacity for data-informed decision making.
  - Outer wheel: continuous interaction and communication across stakeholders
  - Middle wheel: data-driven learning process
  - Frontier: between the outer and middle wheels is active collaboration between all partners
  - Inner circle: rigorous research framework to guide the data science



# EQUITY OF ACCESS TO BROADBAND

## Multiple Measures to Tell the Story

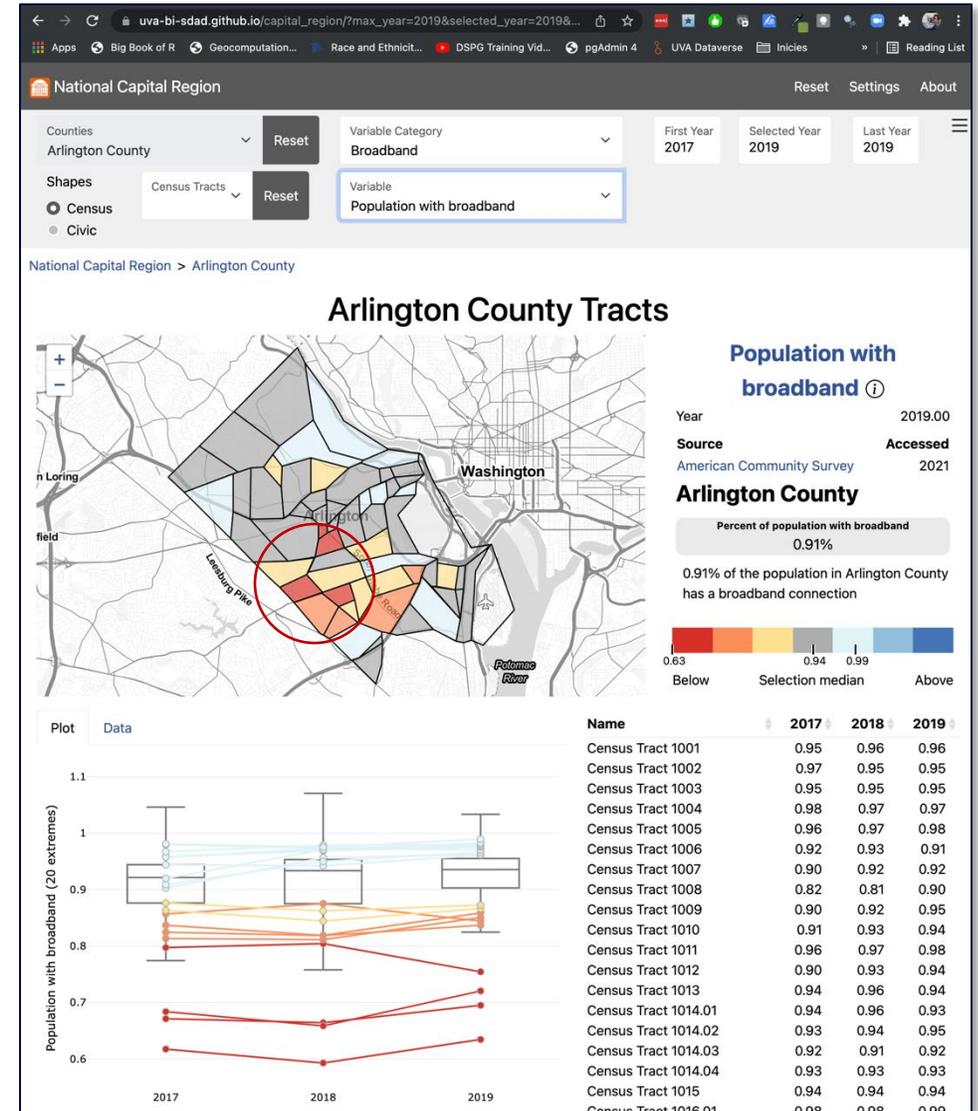
- Average download speeds (from Ookla) are relatively high across Arlington with the slowest average still above 100Mb (the newer standard for “broadband”)
- Ookla data only recently made available in 600-meter squares that we translated to block groups



# EQUITY OF ACCESS TO BROADBAND

Multiple Measures to Tell the Story

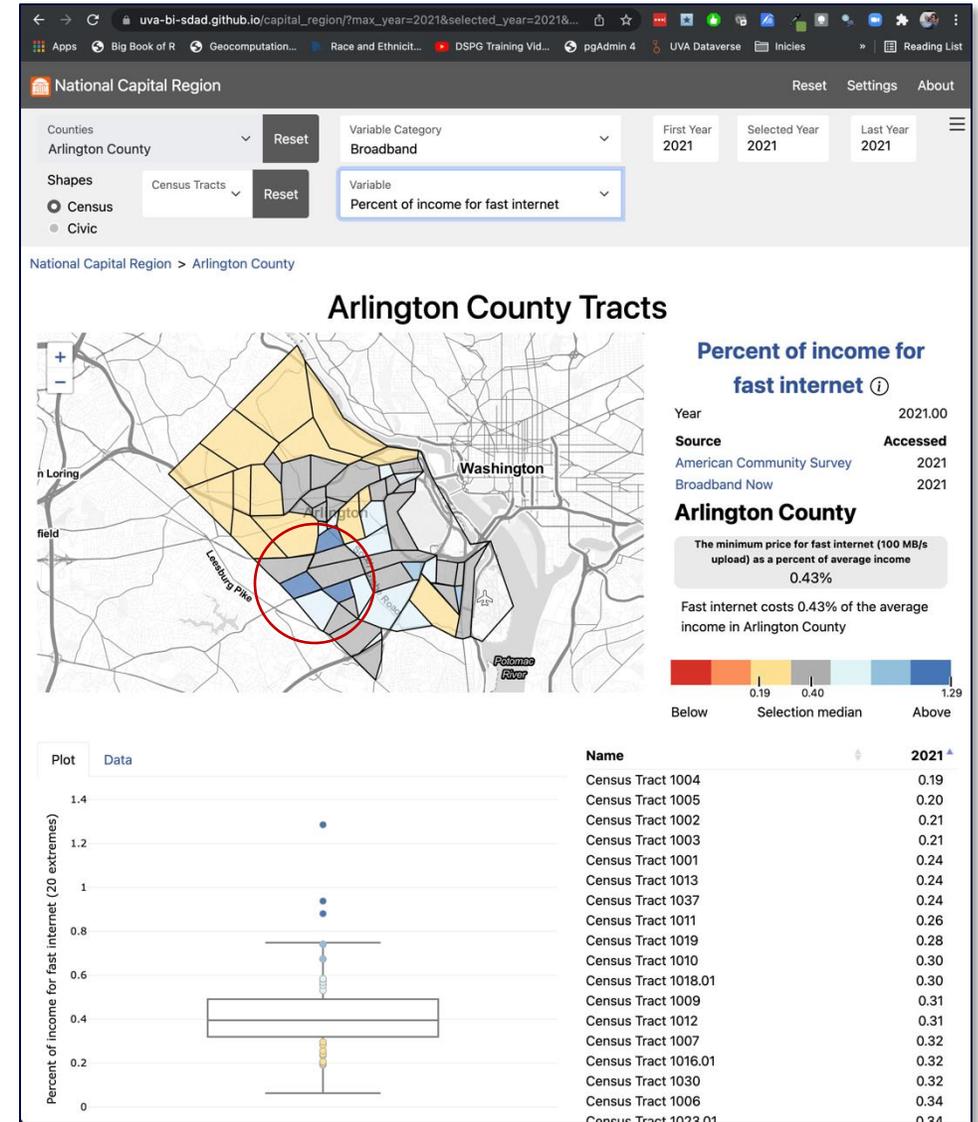
- However, specific areas can be identified that have a significantly lower level of broadband adoption than the rest of Arlington



# EQUITY OF ACCESS TO BROADBAND

## Multiple Measures to Tell the Story

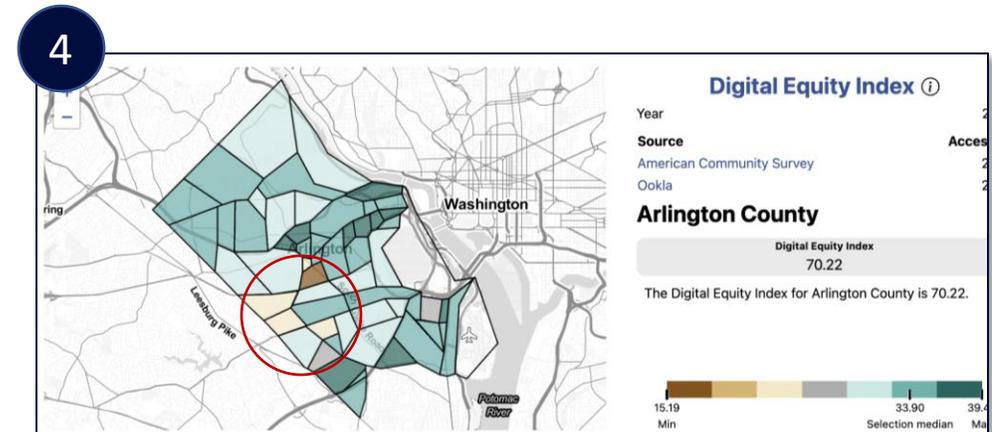
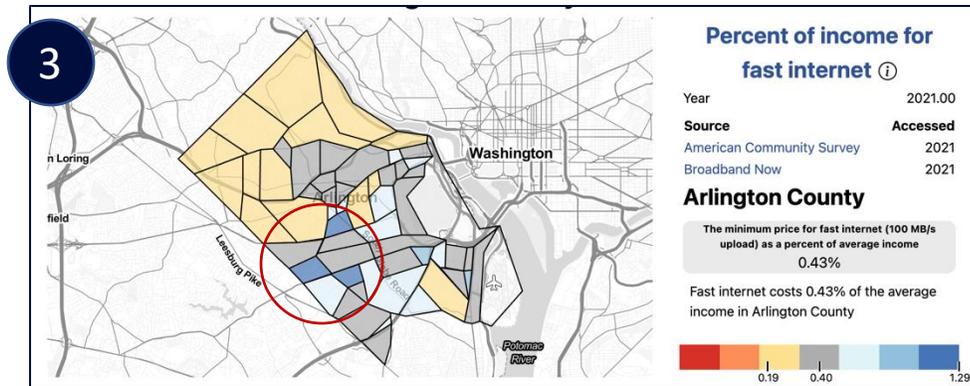
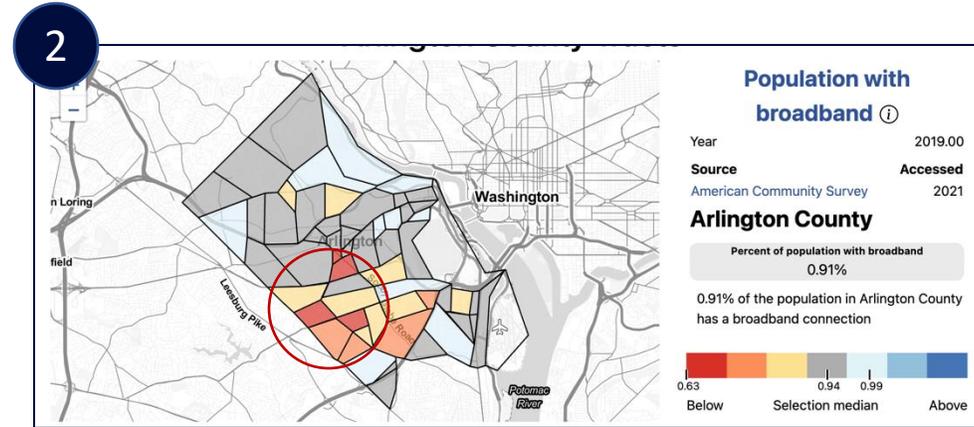
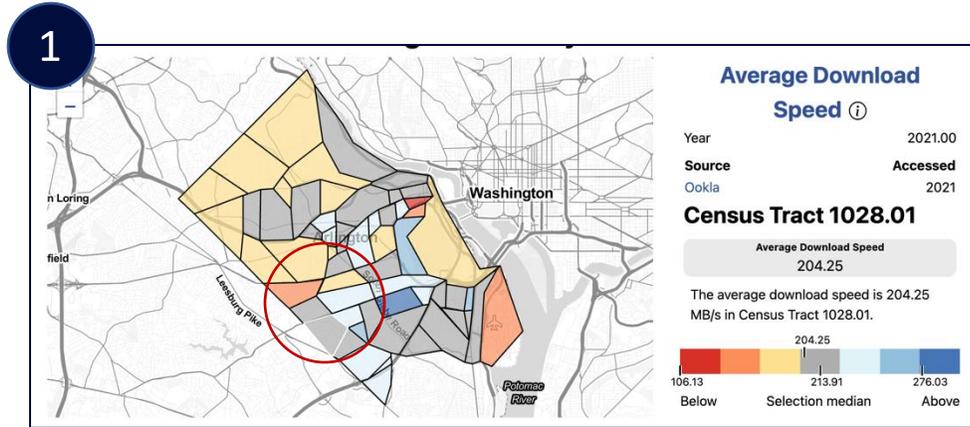
- Calculated % of household income vs cost of 100Mb/s in every block group
- These areas of lowest broadband adoption appear to directly correlate with the areas having the highest ratio of household income to the cost of broadband, indicating an economic issue, as opposed to an issue of availability.



# EQUITY OF ACCESS TO BROADBAND

## The Digital Equity Index (DEI)

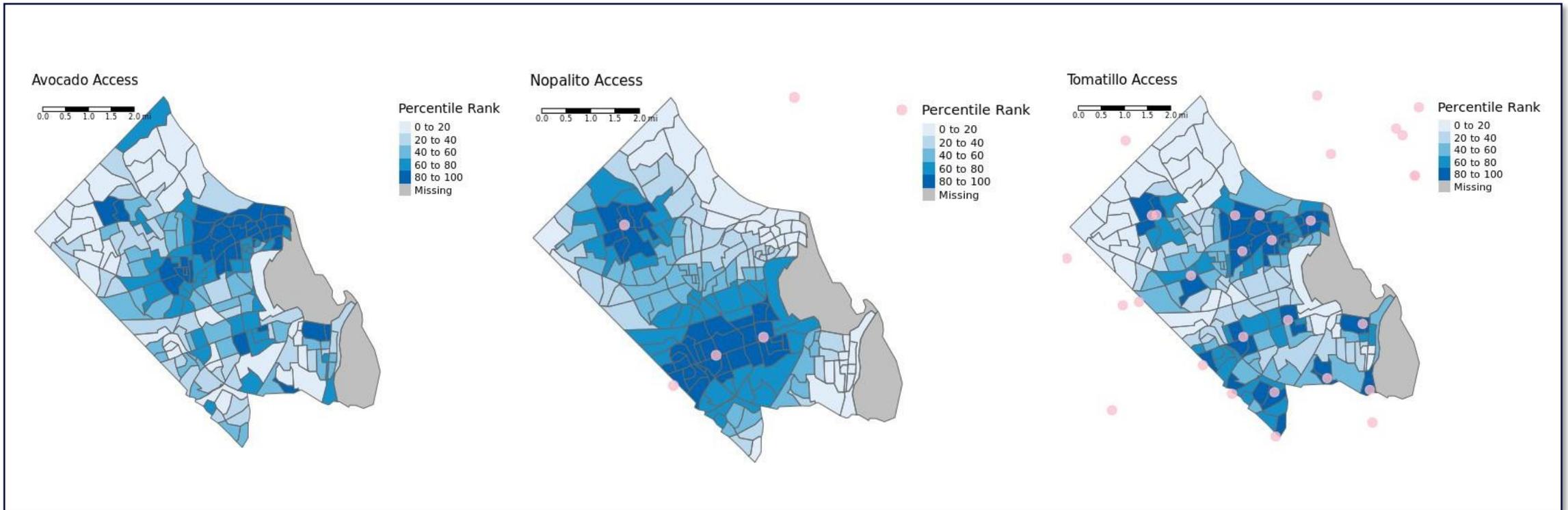
Combine these with other socio-economic variables to create completely new metric



# EQUITY & STABILITY OF STAPLE FOODS ACCESS

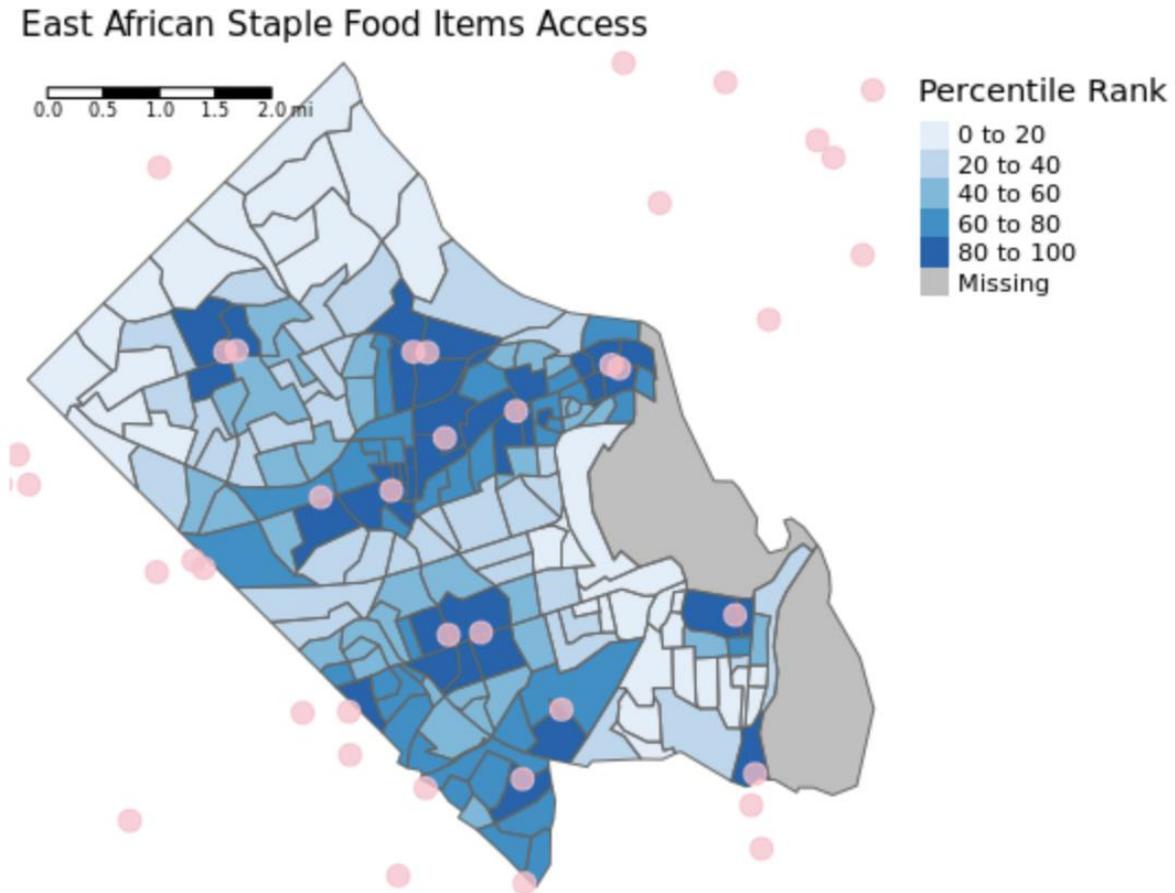
Relative access to individual food items using a gravity index

Items can be combined to create culture-specific food baskets



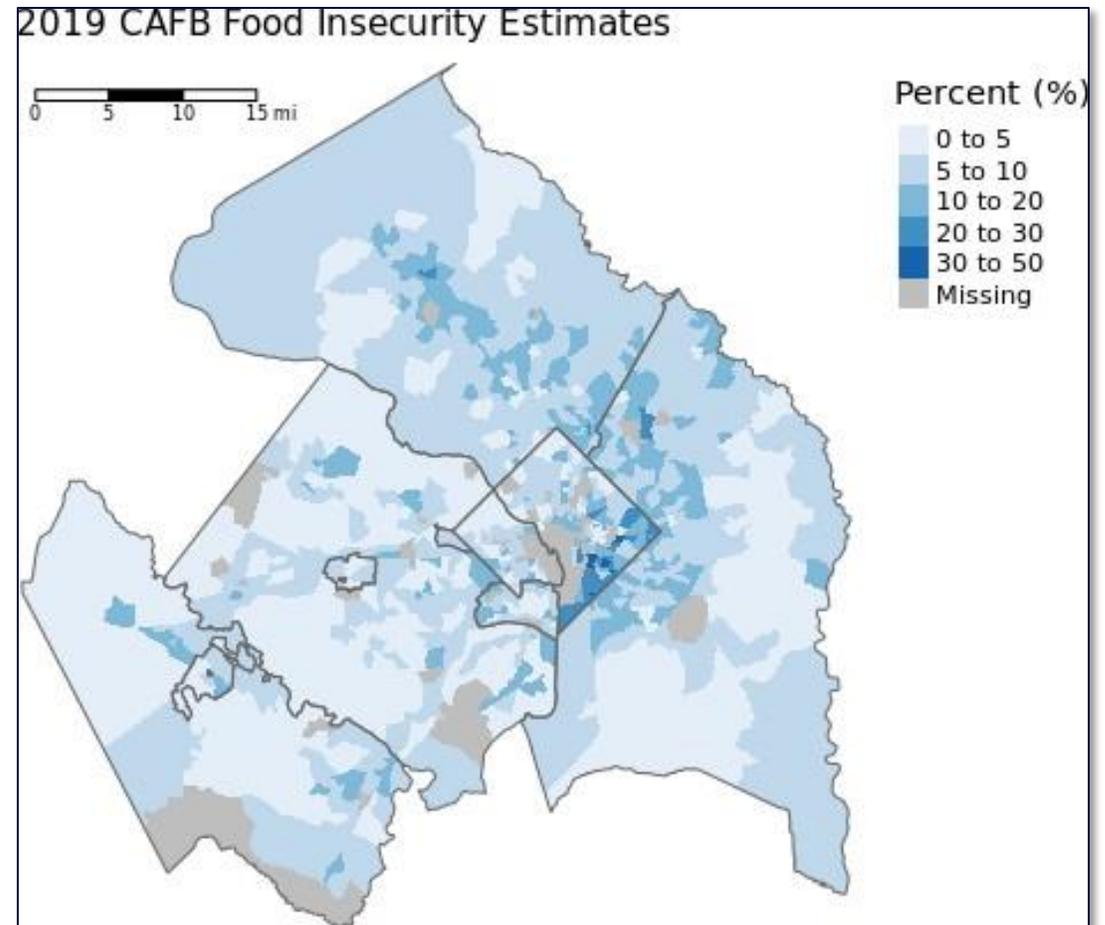
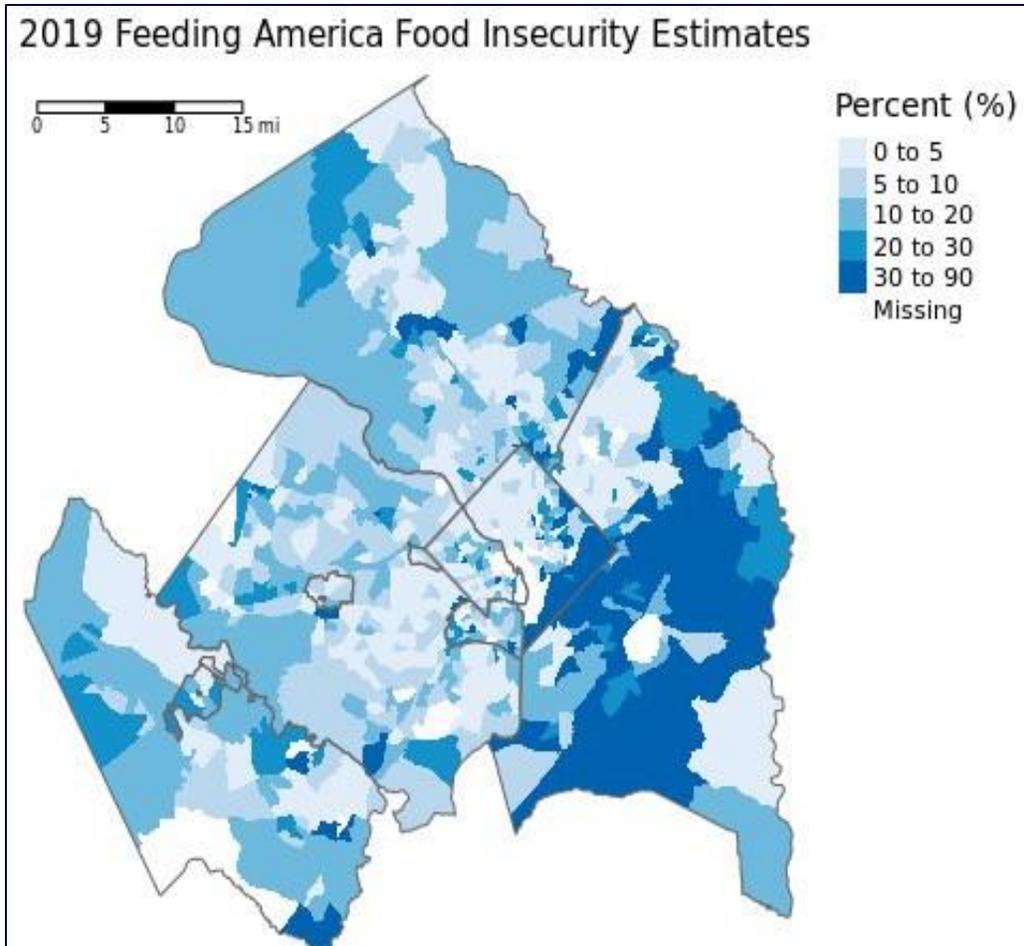
# EQUITY & STABILITY OF STAPLE FOODS ACCESS

Relative access using travel times and a gravity index



# EQUITY & STABILITY OF STAPLE FOODS ACCESS

Two Measures of Food Insecurity: Feeding America (National – adjusted to Census Tracts) and Capital Area Food Bank (Feeding America Local)



# GEOGRAPHIES FOR EQUITY ANALYSIS

Translation of Census Demographics to New Geographies



Census Block Groups

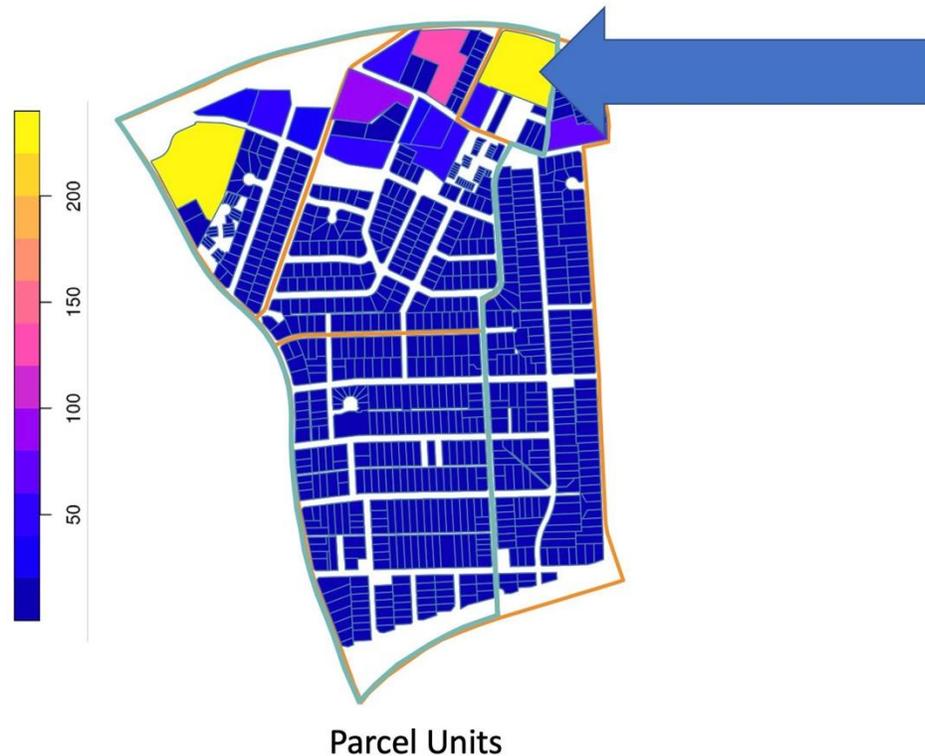


Waverly Hills Civic Association

# GEOGRAPHIES FOR EQUITY ANALYSIS

Translation of Census Demographics to New Geographies

Reviewing multiple methods – creating new approaches



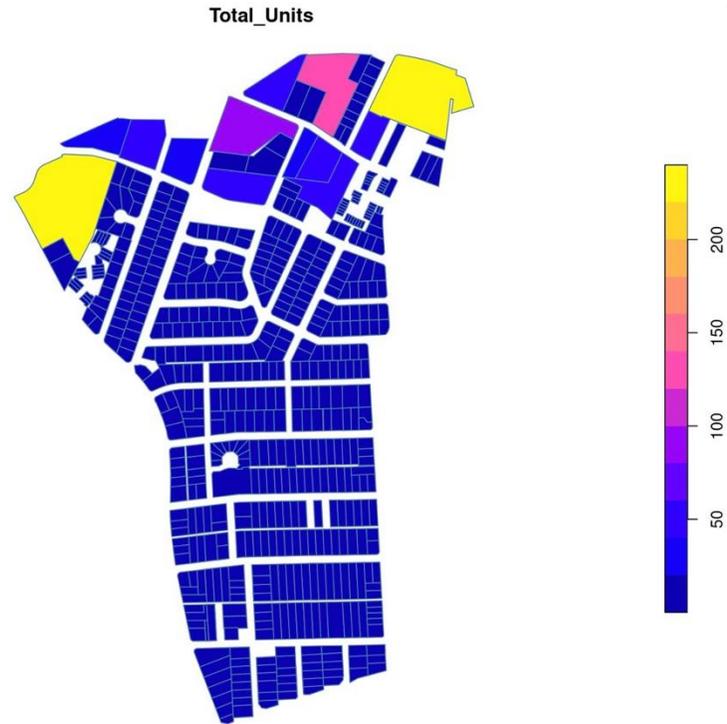
## Partial Overlap

- Leave it Off? (Policy Map 50% Approach)
- Use Percentage Overlap? (City-Data, ESRI, others)
- Our approach, step 1 – Allocate demographics to parcels according to number of housing units

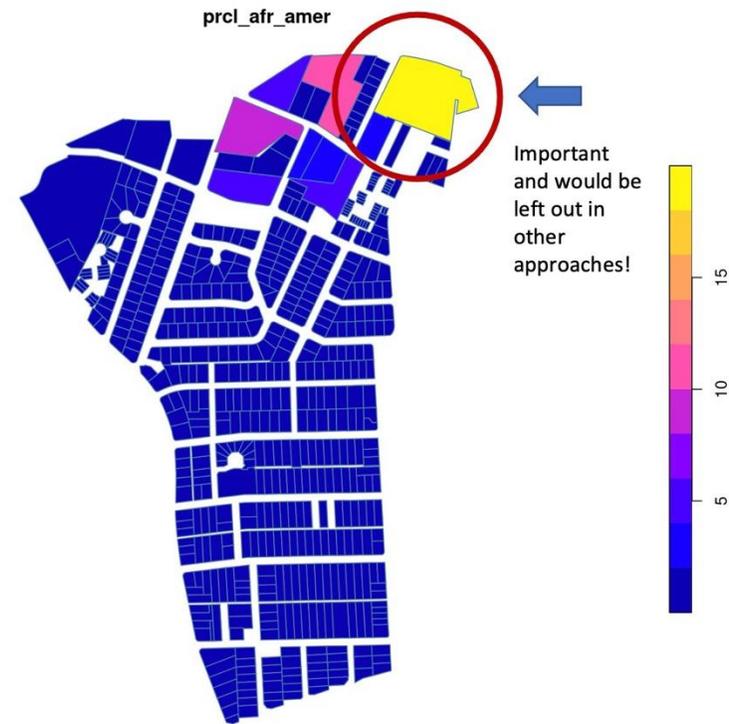
# GEOGRAPHIES FOR EQUITY ANALYSIS

Translation of Census Demographics to New Geographies

Reviewing multiple methods – creating new approaches



Waverly Hills Parcel Units

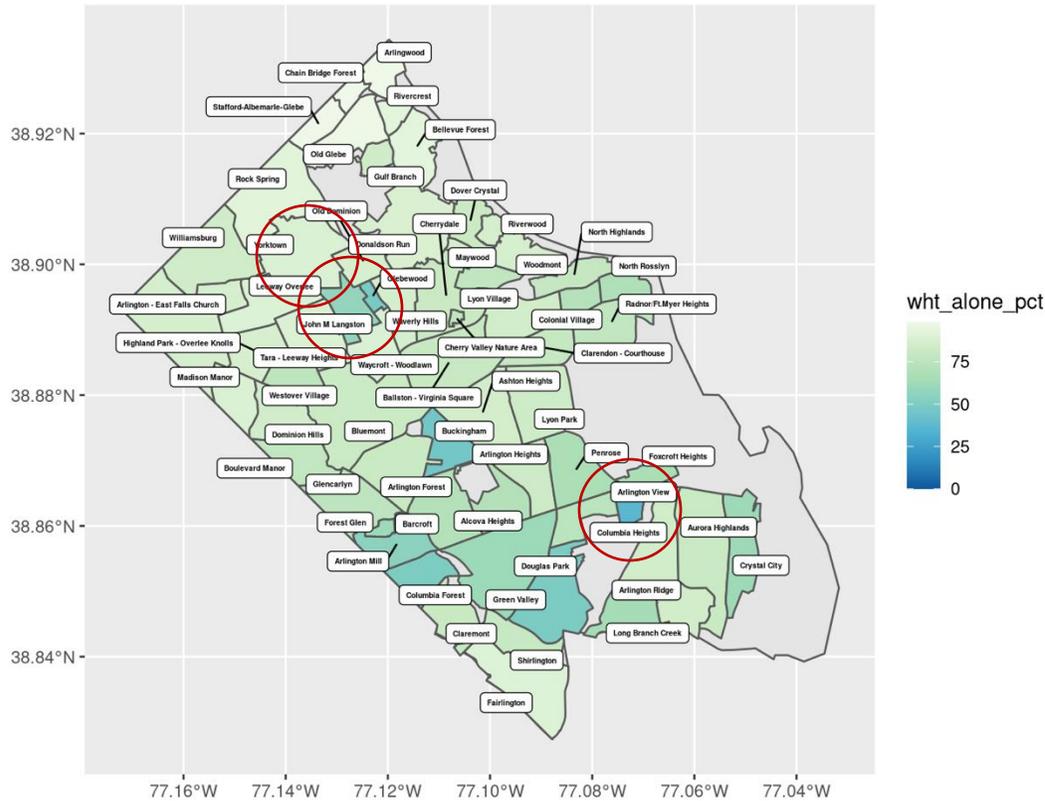


Waverly Hills Estimated Demographic per Parcel

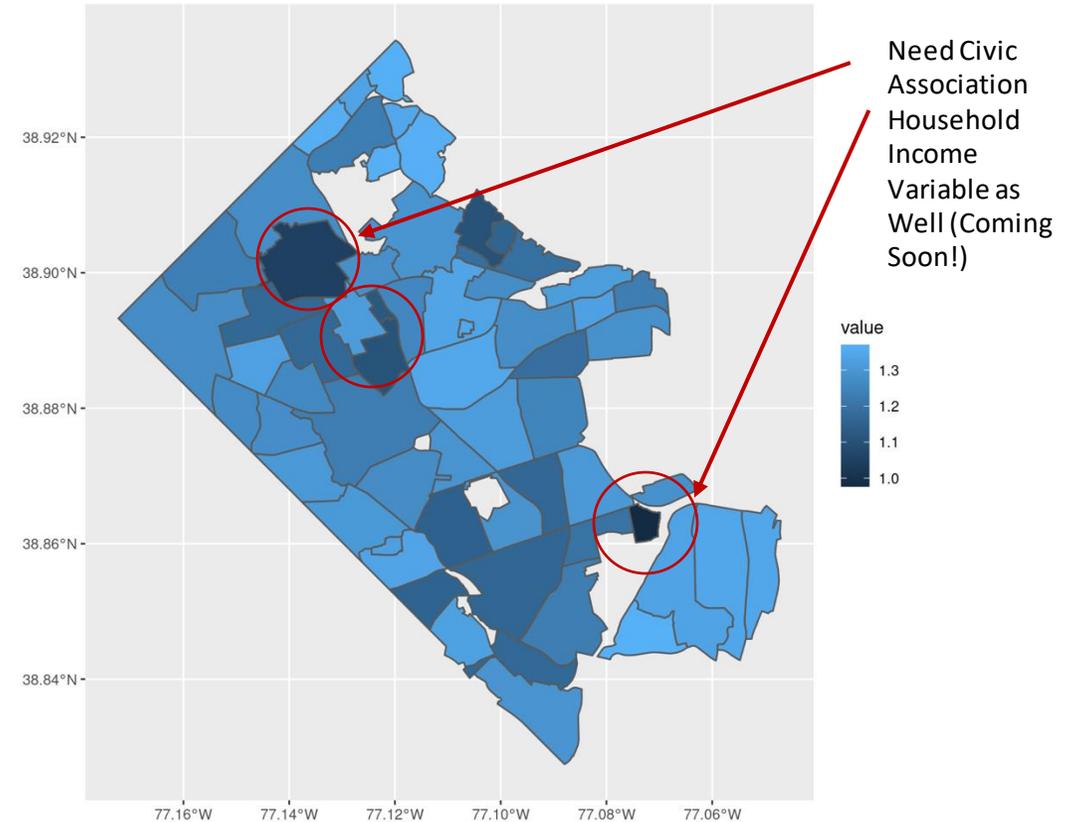
# GEOGRAPHIES FOR EQUITY ANALYSIS

## ACS Block Group Demographics Translated to Arlington Civic Associations

Arlington Civic Association Demographics  
Percent White 2019 [ACS Redistribution]



Primary Care Physician Access by Civic Association  
2019 [3-Step Catchment Area Calculation]

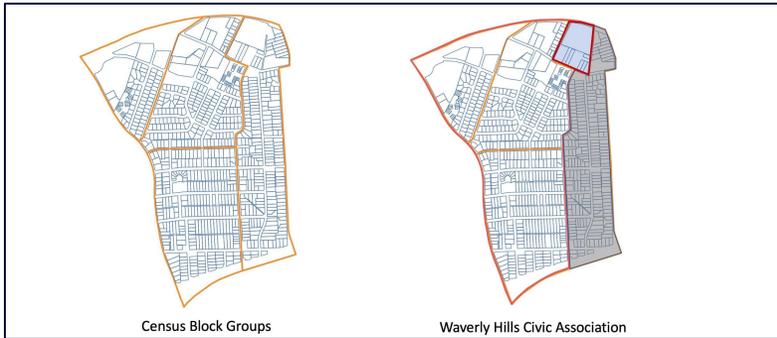


Census Demographics Redistributed + Relative Primary Care Physician Access (Catchment Area)  
for Arlington Civic Associations

# GEOGRAPHIES FOR EQUITY ANALYSIS

Creating data and metrics in geographies that matter locally

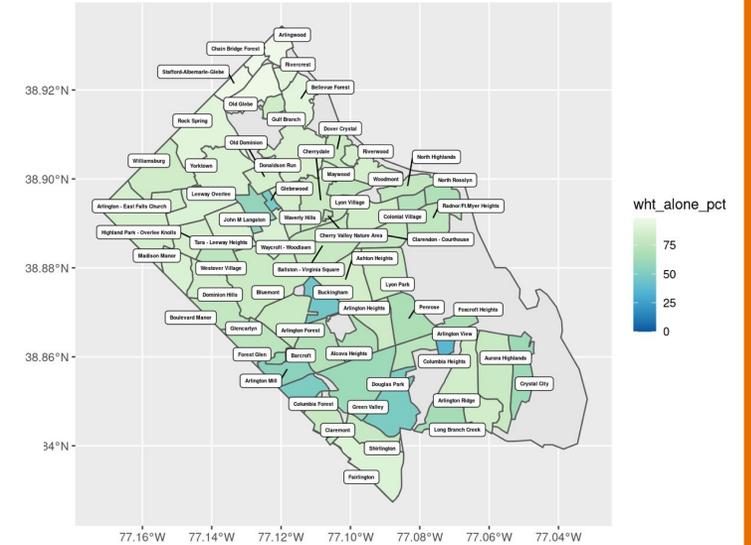
Translation of Census Demographics to New Geographies



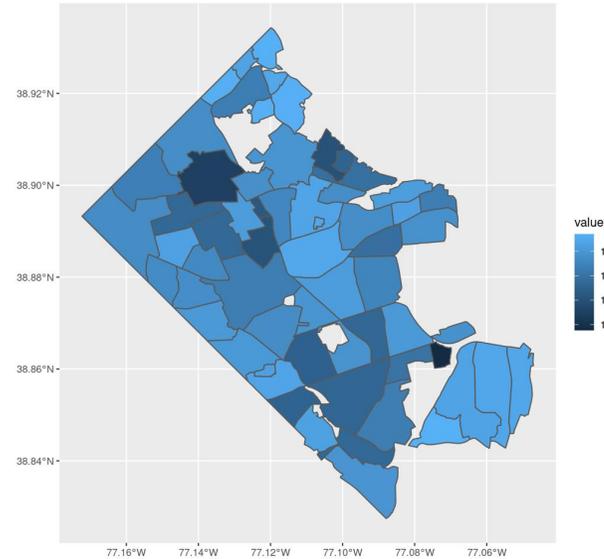
Enabling Analysis by Neighborhood, rather than Census Geographies



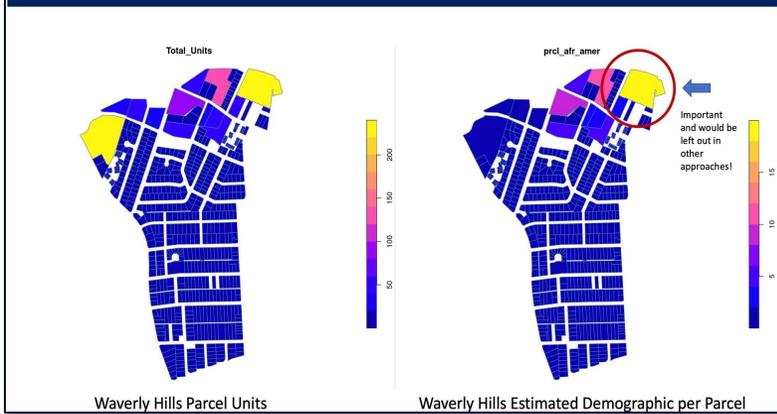
Arlington Civic Association Demographics  
Percent White 2019 [ACS Redistribution]



Primary Care Physician Access by Civic Association  
2019 [3-Step Catchment Area Calculation]

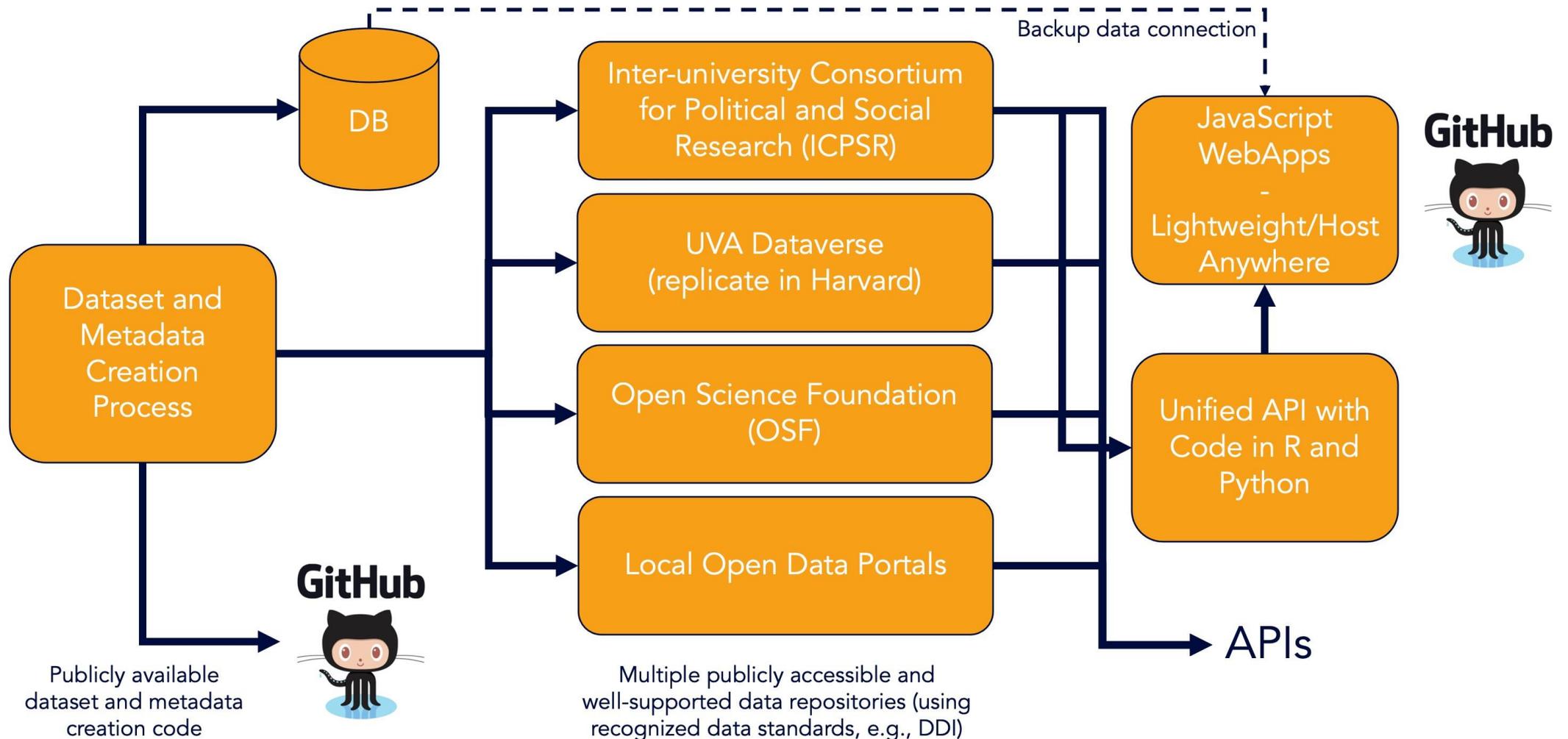


Reviewing multiple methods – creating new approaches



# DATA COMMONS GENERALIZED ARCHITECTURE

Modular, Sustainable, Expandable



# DATA COMMONS GENERALIZED ARCHITECTURE

Dataverse Storage: Leveraging free, well-supported, data and metadata storage and download

The screenshot shows the UVA Dataverse interface for a draft dataset. The page title is "Social Data Commons: Education & Training: Virginia SOL Scores". The status is "Draft" and "Unpublished". The dataset description is "Virginia Standards of Learning Test Scores collected by the Virginia Department of Education (VDOE)". The subject is "Social Sciences (Ex: Education, Politics, Sociology, Economics, Psychology)". The page includes a search bar, filters for file type and access, and a list of two files: "va\_ct\_vdoe\_2019\_2021\_3rd\_grade\_mean\_median\_read\_score.csv.xz" (3.6 KB) and "va\_hd\_vdoe\_2019\_2021\_3rd\_grade\_mean\_median\_read\_score.csv.xz" (1.4 KB). The page is powered by the Dataverse Project v. 5.6 built 477-5fe0150.

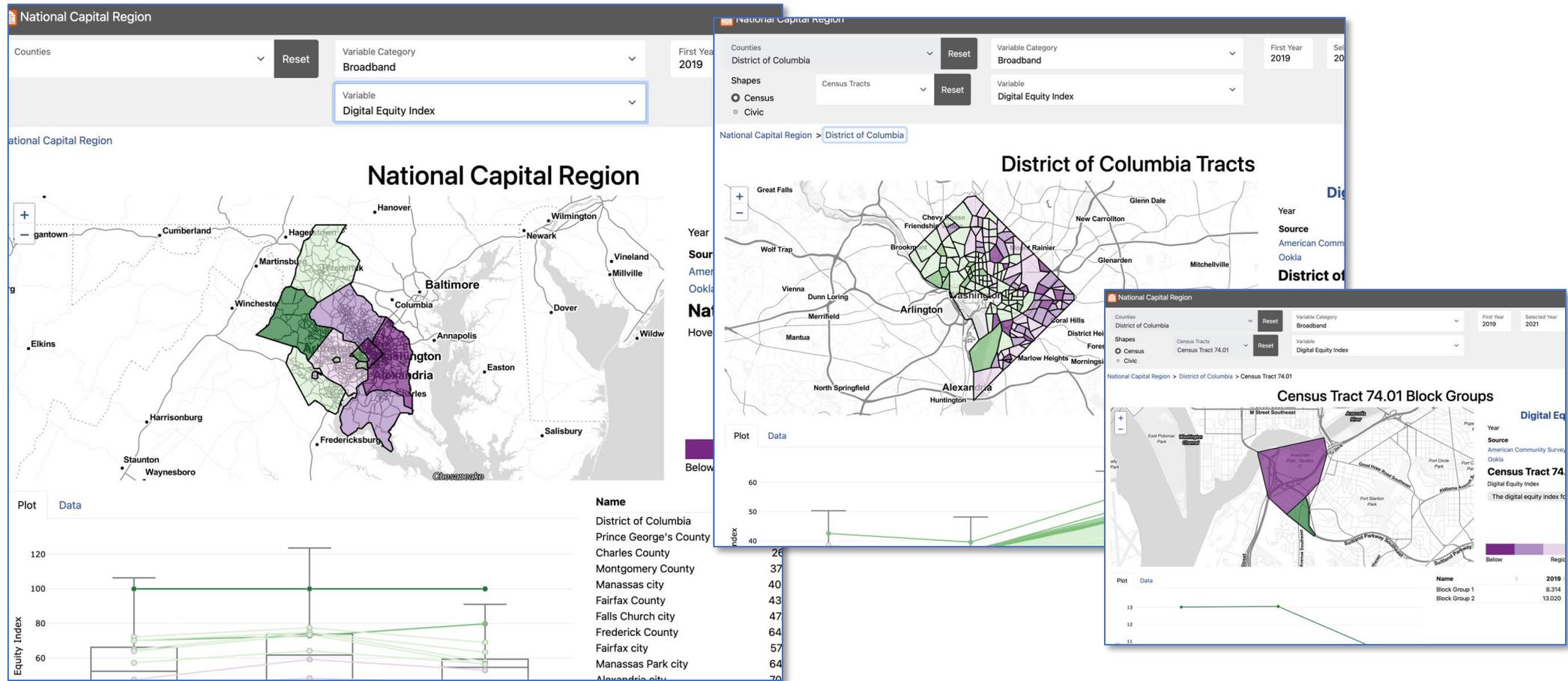
The screenshot shows the UVA Dataverse interface for the same dataset, displaying the citation metadata. The dataset persistent ID is "doi:10.18130/V3/LYK9RZ". The title is "Social Data Commons: Education & Training: Virginia SOL Scores". The authors are "Schroeder, Aaron (University of Virginia), Charankevich, Hanna (University of Virginia), Niagra, Bob (University of Waterloo)". The contact information is "Schroeder, Aaron, Charankevich, Hanna". The dataset description is "Virginia Standards of Learning Test Scores collected by the Virginia Department of Education (VDOE)". The subject is "Social Sciences (Ex: Education, Politics, Sociology, Economics, Psychology)". The data creation date is "2021-11-17". The geographic coverage is "United States, Virginia" and the geographic unit is "State Health District; County".

# DATA COMMONS GENERALIZED ARCHITECTURE

Lightweight JavaScript WebApp, Universally Deployable:

NCR Version (this project: beta): [https://uva-bi-sdad.github.io/capital\\_region/](https://uva-bi-sdad.github.io/capital_region/)

VDH Version (sister project: deployed, VA only): [https://uva-bi-sdad.github.io/vdh\\_rural\\_health\\_site/](https://uva-bi-sdad.github.io/vdh_rural_health_site/)



# NEXT STEP - FAIRFAX POLICY AREA GROUP MEETINGS

## Some Ideas from Initial Meetings – Just Getting Started!

- Housing Group
  - Market Affordability Index
  - Housing Age and Condition Index
  - Housing Heterogeneity Measure
  - Neighborhood amenities (transport, food, parks, etc.)
  - HOA geographies
- Food Security Group
  - Benchmark a “validatable” measure of Food Insecurity
  - Measuring impact of SNAP on Food Insecurity
  - Percent of Food Insecure who are eligible for SNAP
  - Number eligible but not participating in SNAP in multiple geographies
- Economic Health & Growth Group
  - Up-to-date business size and demographics
  - Measuring the ongoing health of business districts
  - Understanding migration flows
  - Planning Unit, COG Traffic Analysis Zone geographies
- Technology Group
  - Broadband access at multiple geographies
  - Broadband access by demographic groups