

SOCIAL IMPACT DATA COMMONS

Supporting Equity-Informed Decision-Making at the Local Level

DATA - APPLICATIONS - TOOLS - METHODS

Dr. Aaron D. Schroeder
Social & Decision Analytics Division
Biocomplexity Institute, University of Virginia



**BIOCOMPLEXITY
INSTITUTE**



Center for
Inclusive Growth

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.**

Social Impact Data Commons

OBJECTIVE

To make impactful equity-informed decisions at the local and regional level, decision-makers require data and indicators that

- Triangulate on their policy challenges and questions
- Are at a geographic level that informs their decision making
- In a geographic shape that is helpful (e.g., a metropolitan planning corridor, by school boundaries (e.g., elementary schools), and other sub-areas of interest
- Validated and timely

Social Impact Data Commons

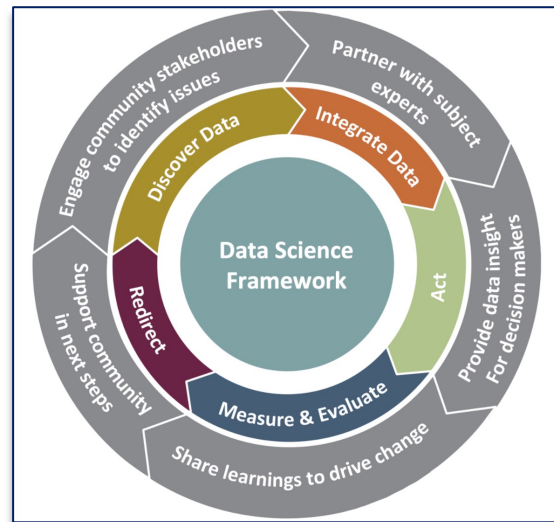
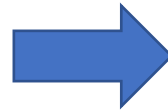
Final Dataset Requirements

Once appropriate datasets are discovered, vetted, created, synthesized and validated, they must be easily:

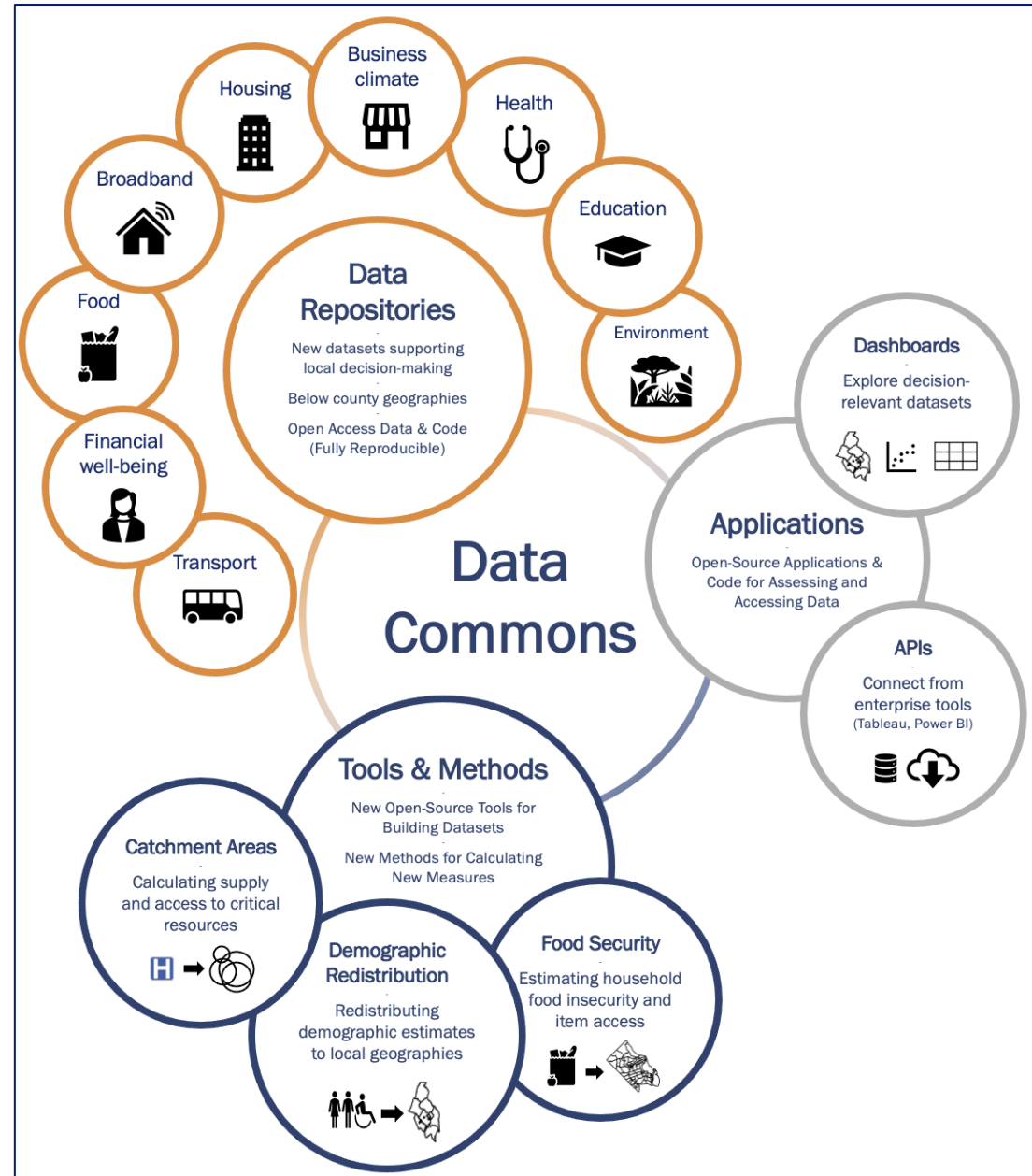
- **Assessable – are these the right data for the policy question?**
 - Can we view data maps, tables, charts, and metadata to assess the data?
- **Accessible - are the data downloadable?**
 - Can we obtain the data through direct website download, programmatic API access?
- **Analyzable – are the data easily integrated into user analytic systems?**
 - Can we standardize the dataset and geographic file formats?

The Solution

- A System of Open-Source Resources for Local Decision-Making (The Data Commons)
- A Process for Cooperative Iteration with Local Stakeholders (CLD3)



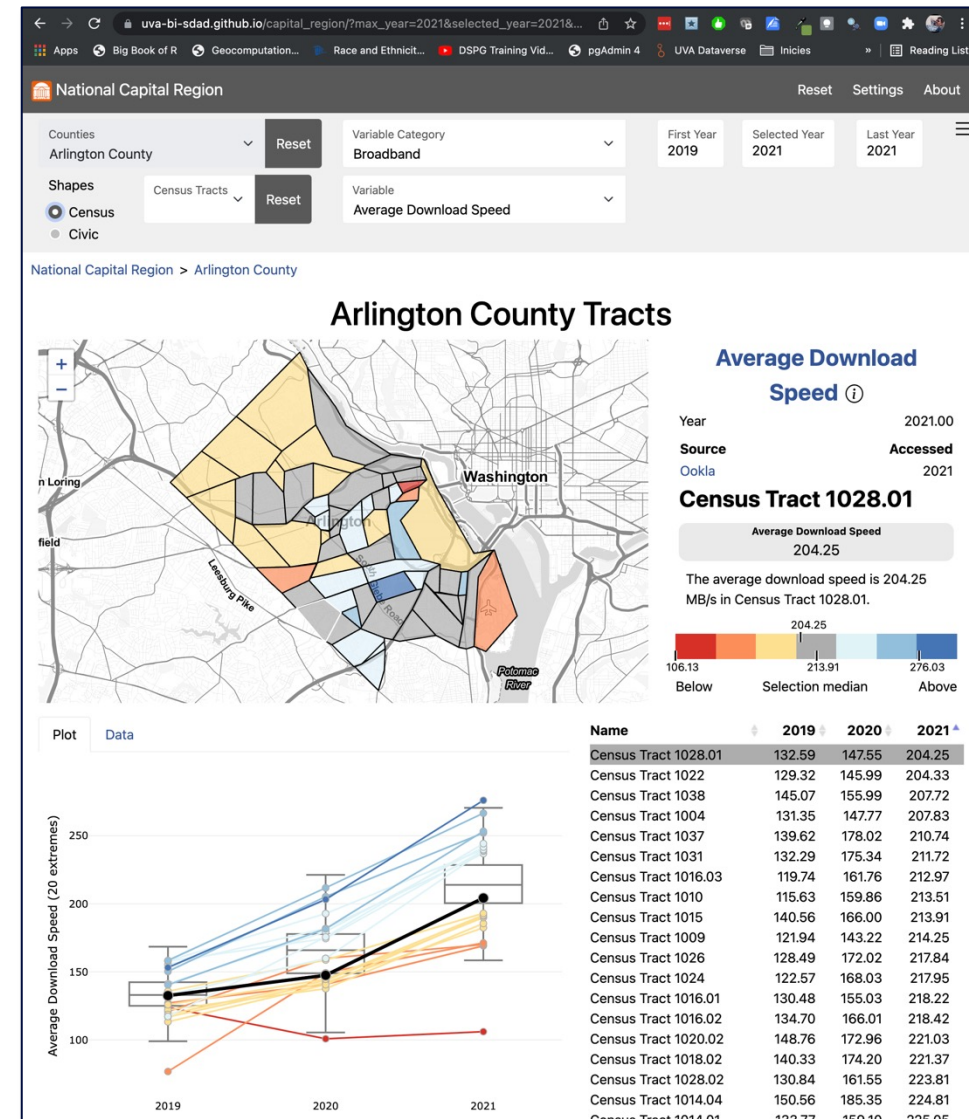
The Community Learning Through Data Driven Discovery (CLD3) Process



EQUITY OF ACCESS TO BROADBAND

Multiple Measures to Tell the Story

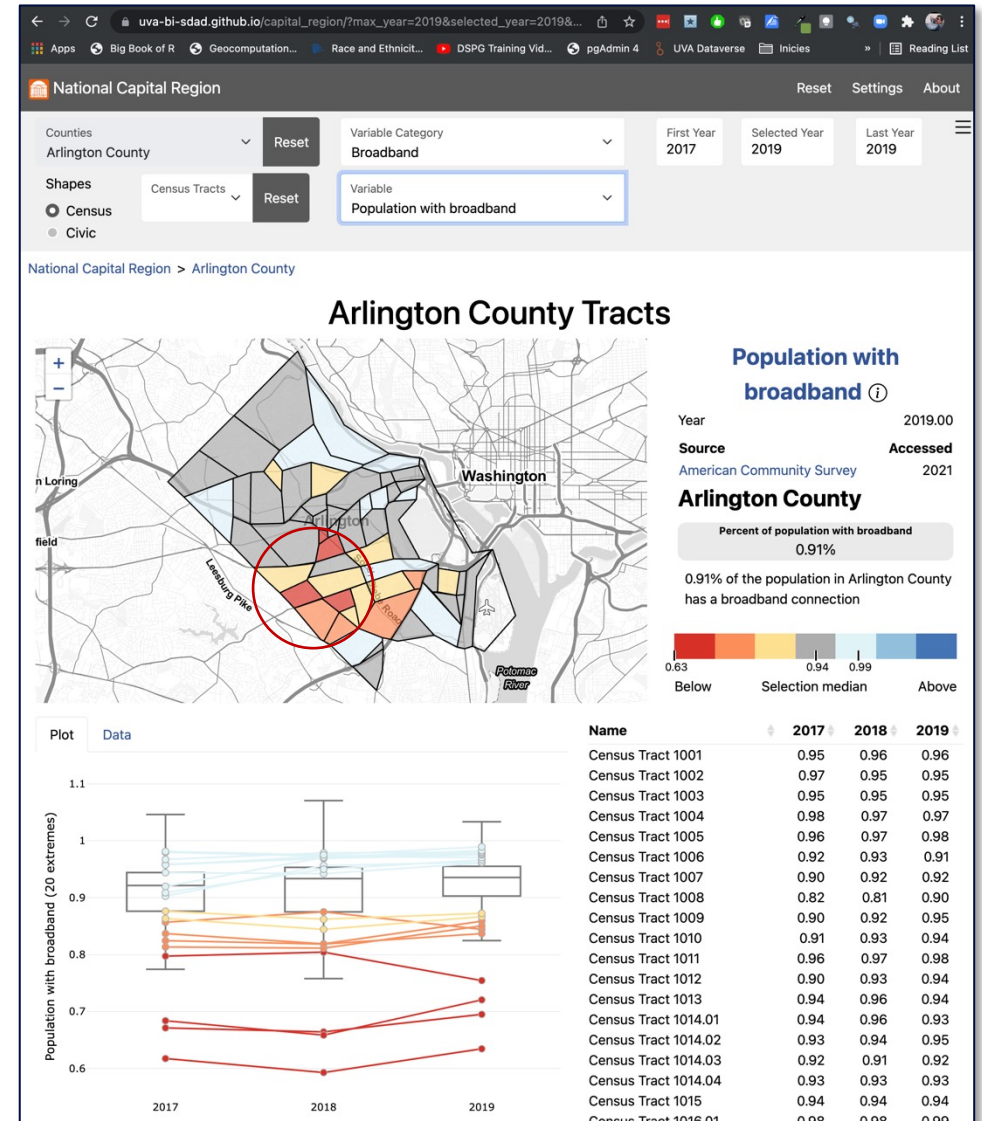
- Average download speeds (from Ookla) are relatively high across Arlington with the slowest average still above 100Mb (the newer standard for "broadband")
- Ookla data only recently made available in 600-meter squares that we translated to block groups



EQUITY OF ACCESS TO BROADBAND

Multiple Measures to Tell the Story

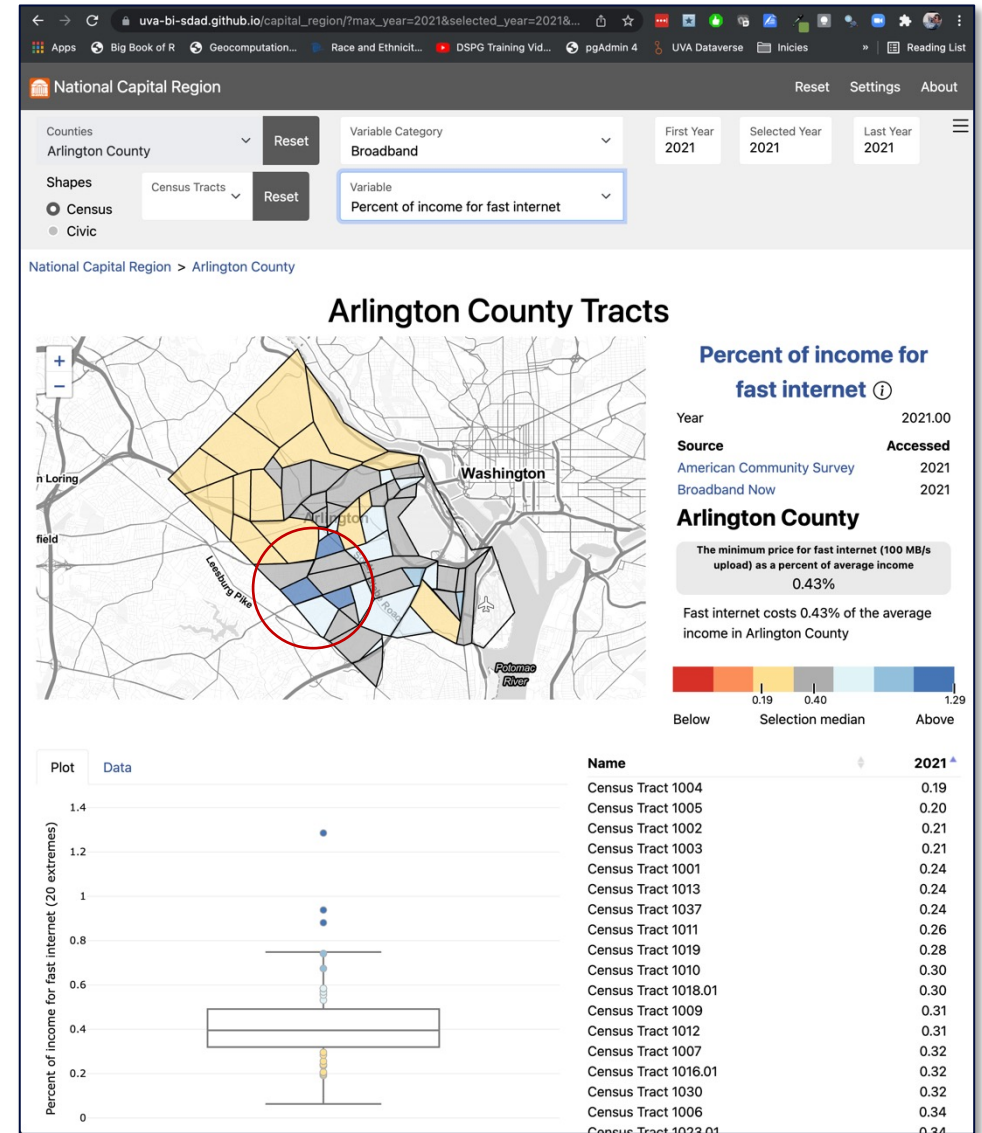
- However, specific areas can be identified that have a significantly lower level of broadband adoption than the rest of Arlington



EQUITY OF ACCESS TO BROADBAND

Multiple Measures to Tell the Story

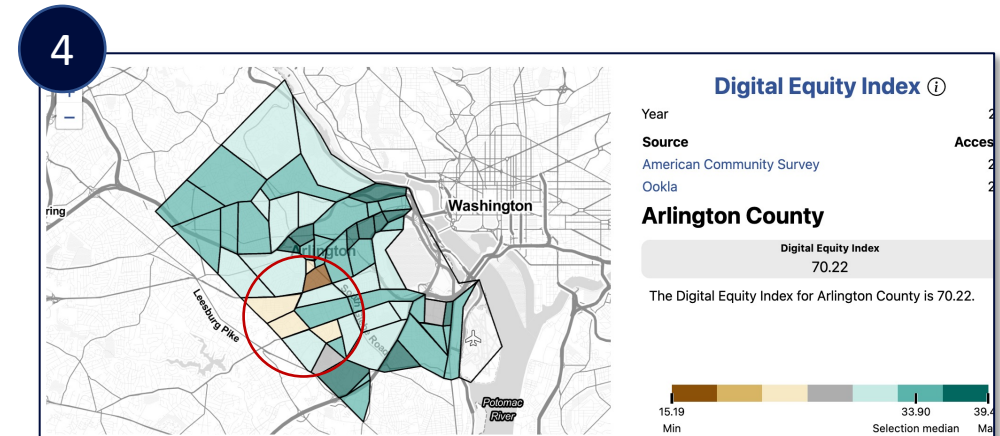
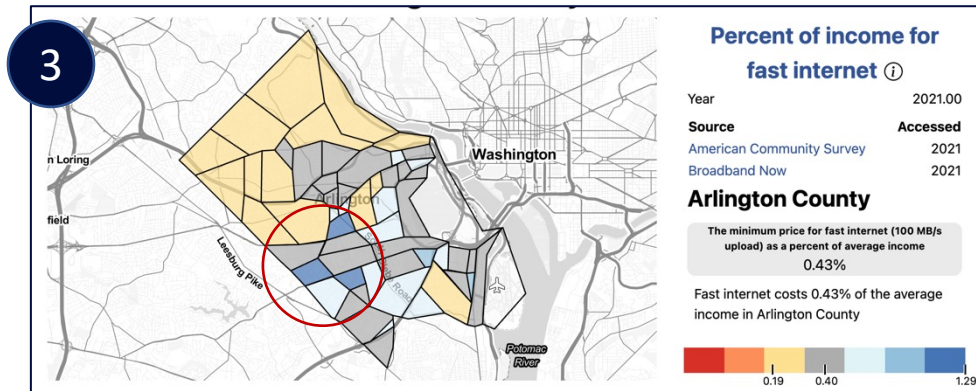
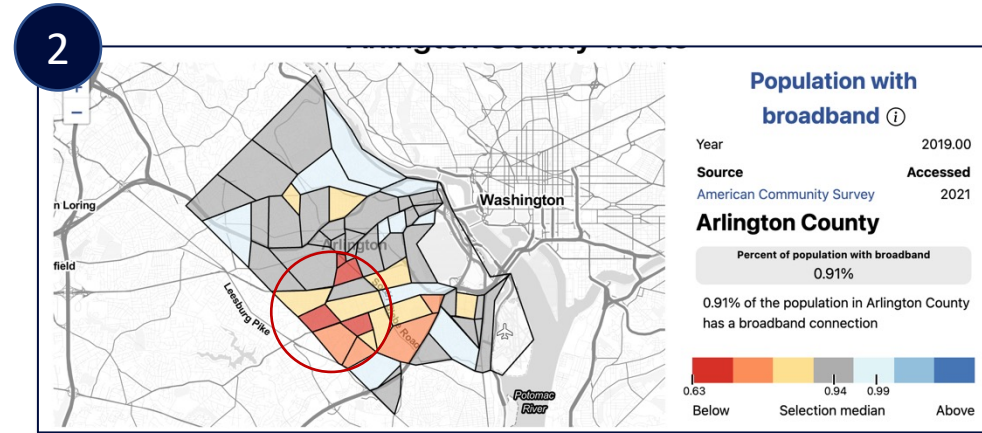
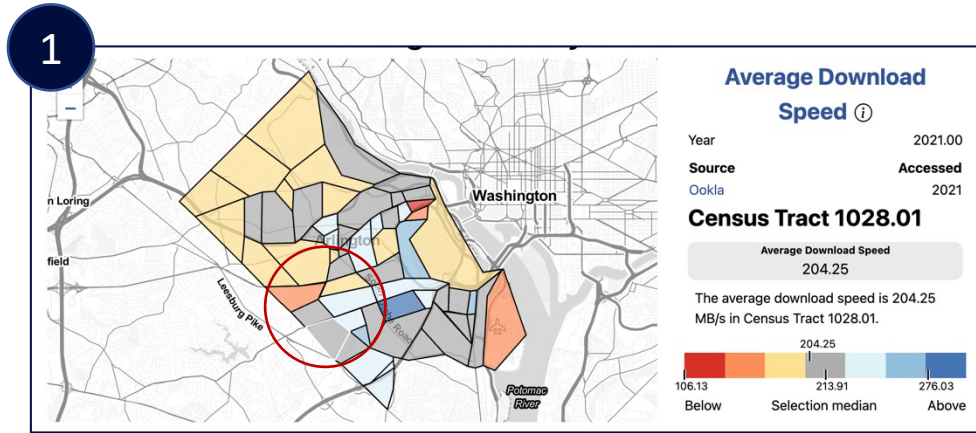
- Calculated % of household income vs cost of 100Mb/s in every block group
- Scraped cost of every level of data service for every census block
- These areas of lowest broadband adoption appear to directly correlate with the areas having the highest ratio of household income to the cost of broadband,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



EQUITY OF ACCESS TO BROADBAND

The Digital Equity Index (DEI)

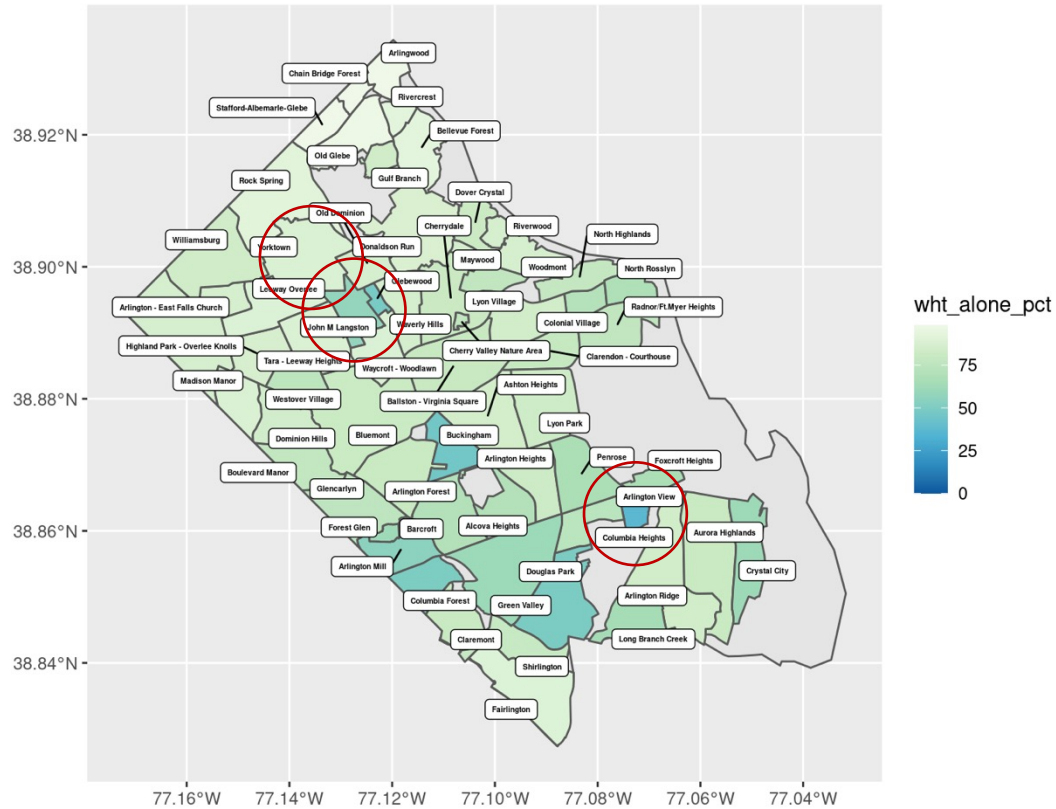
Combine these with other socio-economic variables to create completely new metric



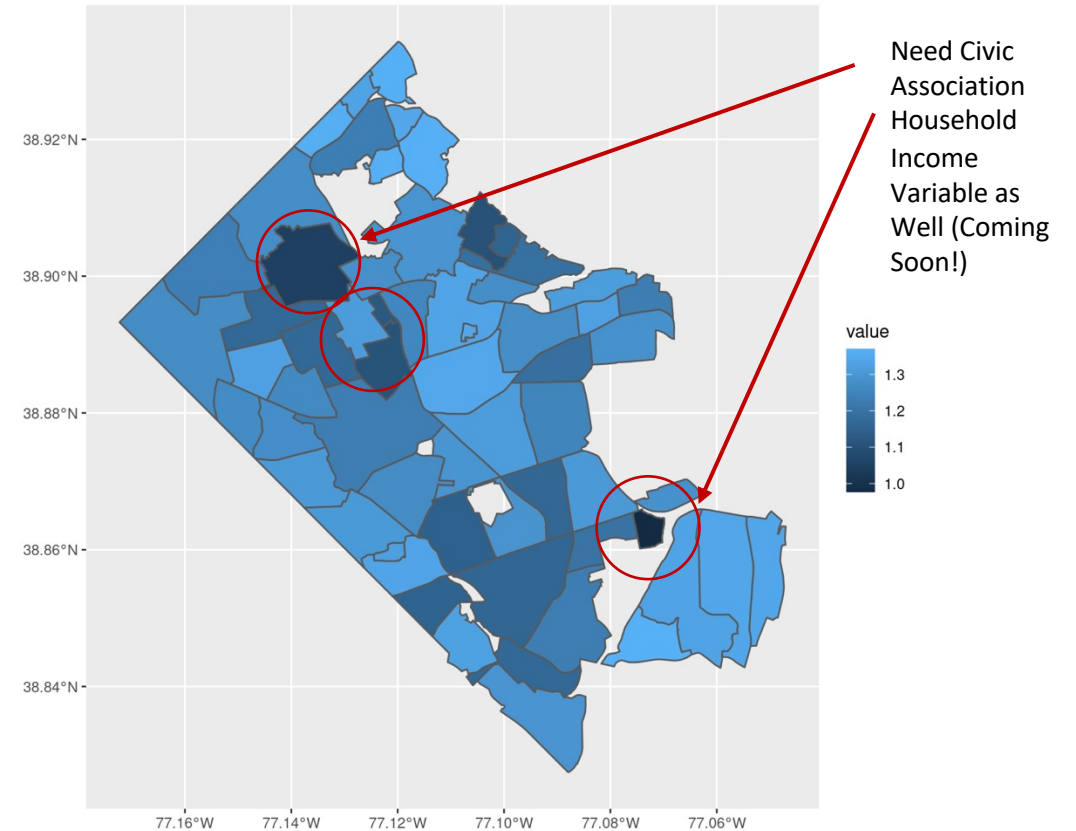
GEOGRAPHIES FOR EQUITY ANALYSIS

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]

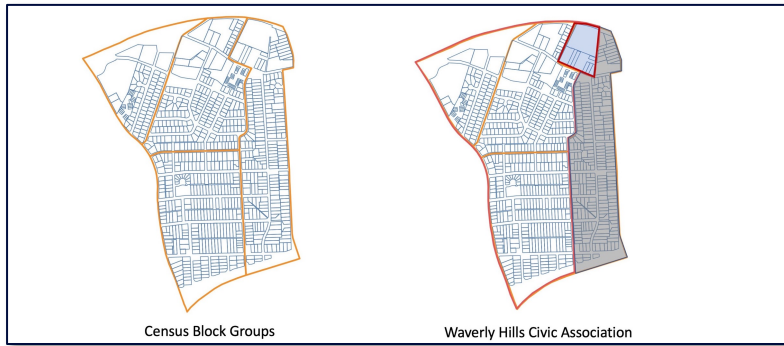


Census Demographics Redistributed + Relative Primary Care Physician Access (Catchment Area)
for Arlington Civic Associations

GEOGRAPHIES FOR EQUITY ANALYSIS

Creating data and metrics in geographies that matter locally

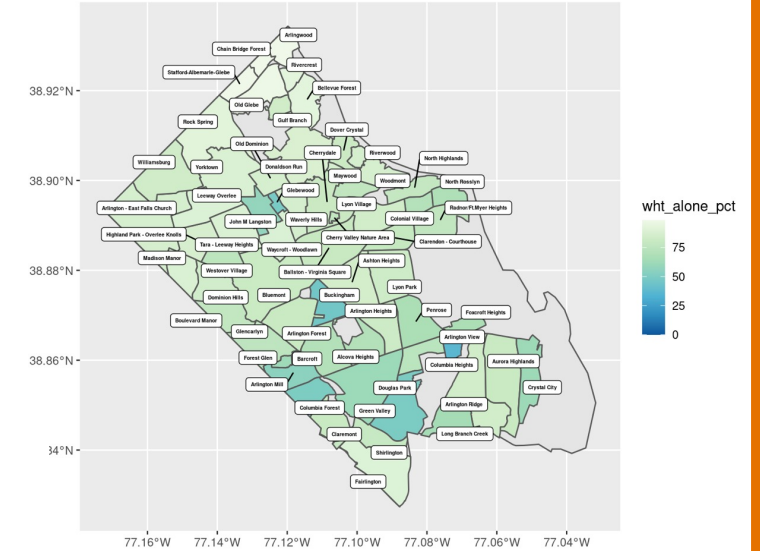
Translation of Census Demographics to New Geographies



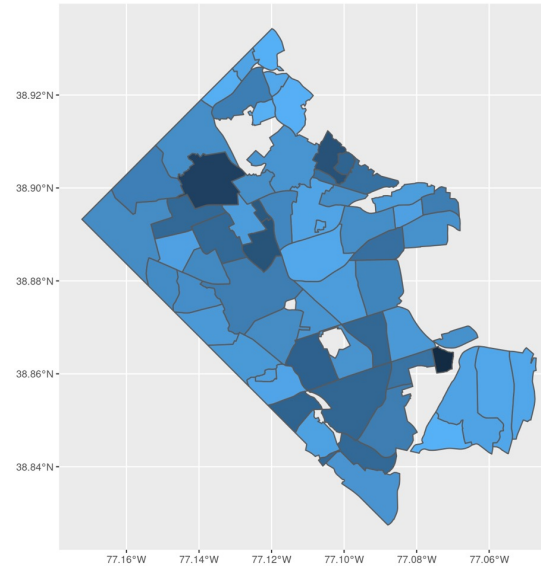
Enabling Analysis by Neighborhood, rather than Census Geographies



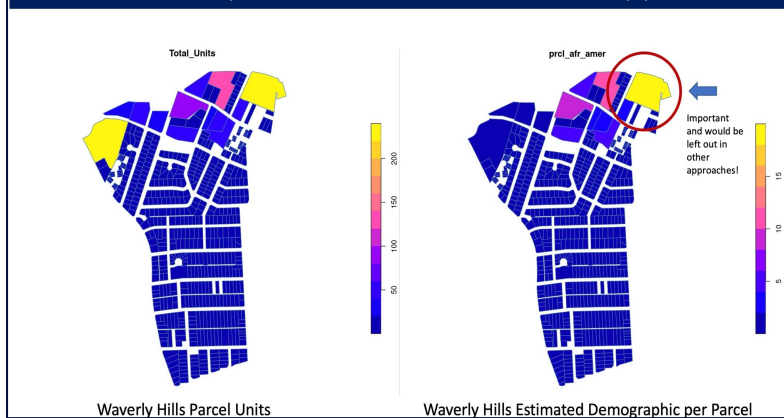
Arlington Civic Association Demographics
Percent White 2019 [ACS Redistribution]



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

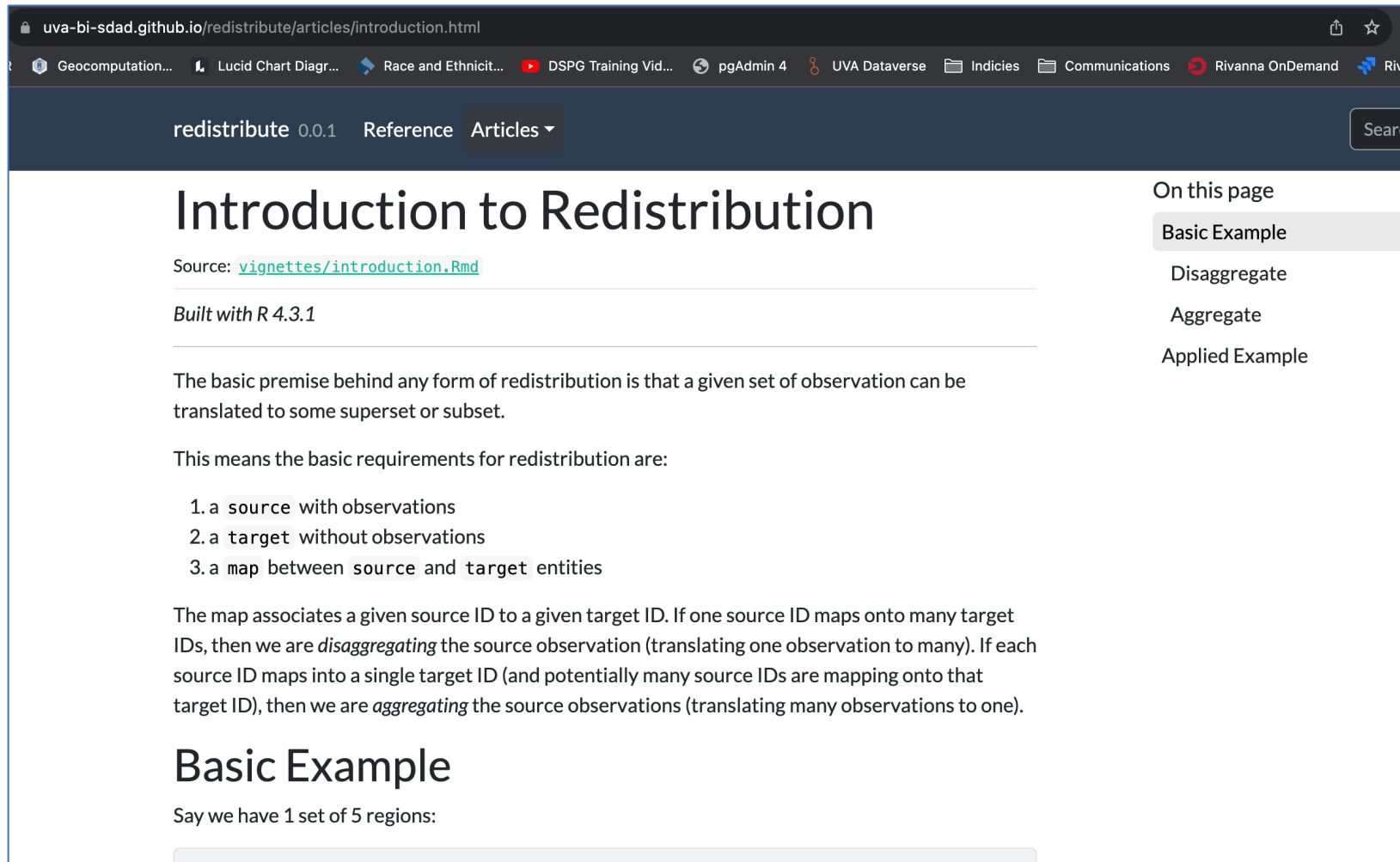


Reviewing multiple methods – creating new approaches



GEOGRAPHIES FOR EQUITY ANALYSIS

Creating open-source tools for dataset creation



The screenshot shows a web browser window with the URL `uva-bi-sdad.github.io/redistribute/articles/introduction.html`. The browser's address bar and tabs are visible at the top. The page content includes a navigation bar with 'redistribute 0.0.1', 'Reference', and 'Articles' (with a dropdown arrow). A search bar is located on the right side of the navigation bar. The main heading is 'Introduction to Redistribution'. Below the heading, the source is listed as `vignettes/introduction.Rmd` and it is noted that the page is 'Built with R 4.3.1'. The text explains the basic premise of redistribution: a given set of observations can be translated to a superset or subset. It then lists three basic requirements for redistribution: 1. a source with observations, 2. a target without observations, and 3. a map between source and target entities. The text further explains that the map associates a source ID to a target ID, and discusses the concepts of disaggregating (one observation to many) and aggregating (many observations to one) source observations. A section titled 'Basic Example' begins with the text 'Say we have 1 set of 5 regions:'. On the right side of the page, there is a 'On this page' section with a list of links: 'Basic Example' (highlighted), 'Disaggregate', 'Aggregate', and 'Applied Example'.

uva-bi-sdad.github.io/redistribute/articles/introduction.html

Geocomputation... Lucid Chart Diagr... Race and Ethnicit... DSPG Training Vid... pgAdmin 4 UVA Dataverse Indices Communications Rivanna OnDemand Riva

redistribute 0.0.1 Reference Articles ▾ Search

Introduction to Redistribution

Source: [vignettes/introduction.Rmd](#)

Built with R 4.3.1

The basic premise behind any form of redistribution is that a given set of observation can be translated to some superset or subset.

This means the basic requirements for redistribution are:

1. a source with observations
2. a target without observations
3. a map between source and target entities

The map associates a given source ID to a given target ID. If one source ID maps onto many target IDs, then we are *disaggregating* the source observation (translating one observation to many). If each source ID maps into a single target ID (and potentially many source IDs are mapping onto that target ID), then we are *aggregating* the source observations (translating many observations to one).

Basic Example

Say we have 1 set of 5 regions:

On this page

- Basic Example
- Disaggregate
- Aggregate
- Applied Example

Data in action: data stories

- Applies the Social Impact Data Commons to real local issues
- How can measures be triangulated to tell a story?
 - Access to broadband: Multiple measures tell the story in Arlington
 - Health care: Is access to urgent care equitable in the National Capital Region?
 - Business diversity: How do measures of equity change over space and time in Fairfax?
- Available on the [Social Impact Data Commons website](#)

Health Equity in the National Capital Region

Is access to urgent care equitable in Arlington and Fairfax County?

Issue overview

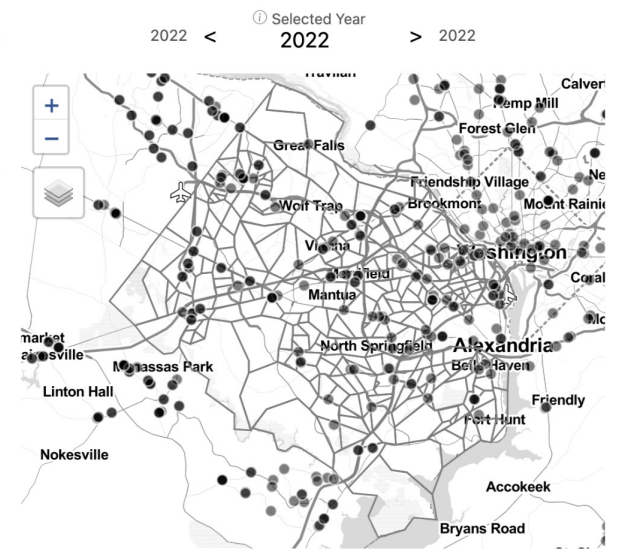
Our stakeholders in Arlington and Fairfax Counties were interested in understanding the equity of access to health services by neighborhood, by race, by household income, and by housing type. We began analyzing access to urgent care facilities.

We inventoried a variety of urgent care location data sources for accuracy and quality. Given that urgent care is a rapidly growing health care service, we found that administrative datasets were incomplete by a wide margin. We found that [Google Maps](#) provided the most complete picture of urgent care facility locations in the Capital Region. To get a better understanding of the idea of access, we compared several measures.

Where are urgent cares in Arlington and Fairfax?

First, we began by examining the locations of urgent care. We found that there are 113 urgent care facilities in Fairfax and 18 facilities in Arlington. By number of facilities, Fairfax has the greatest access to urgent care in the National Capital region. We calculated access to urgent care by count, or presence of an urgent care in a given geography. For most census tracts, there is no urgent care present. Fairfax and Arlington residents who live in a census tract without an urgent care may be able to easily drive to one nearby, though.

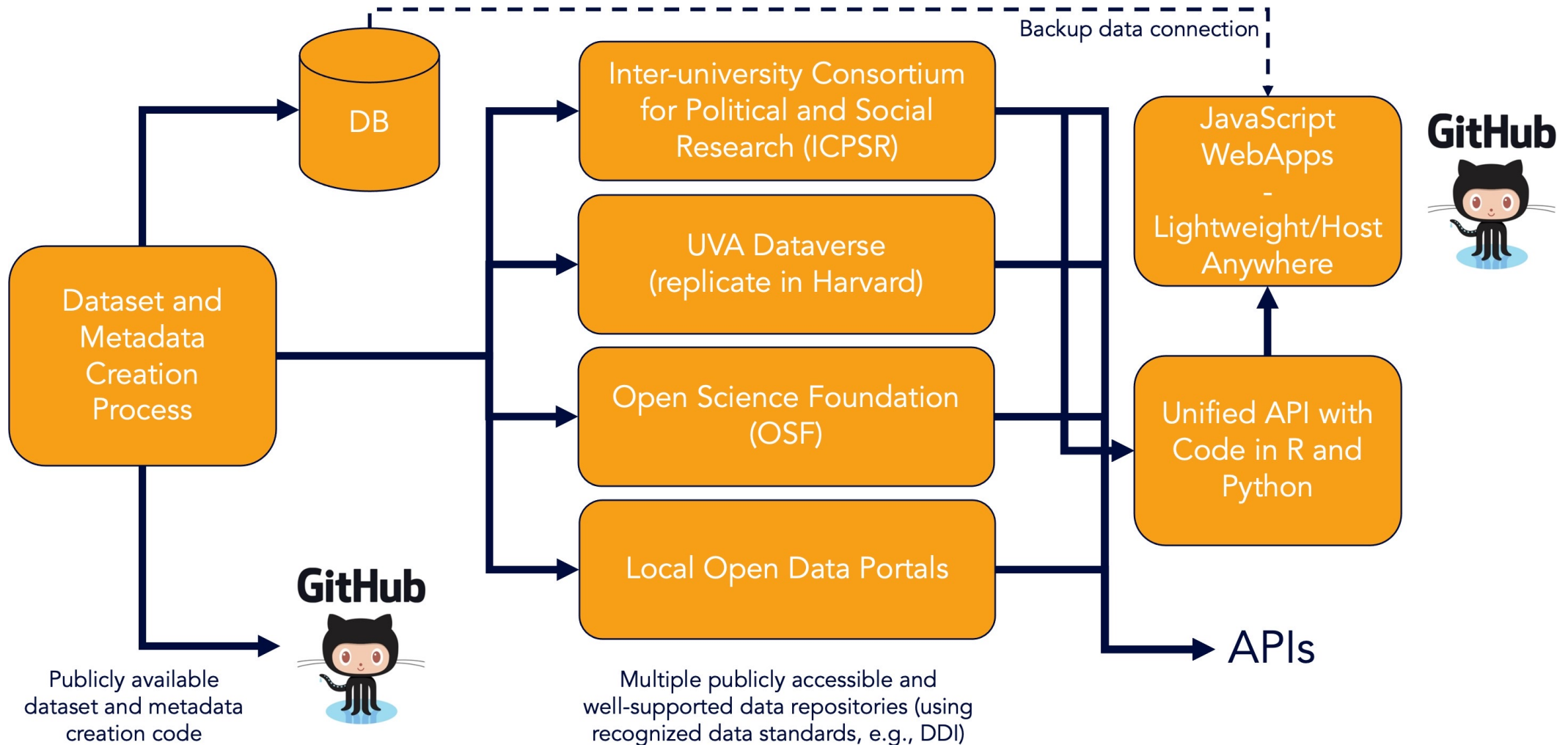
National Capital Region Census Tracts



Urgent care availability (count) ①

DATA COMMONS GENERALIZED ARCHITECTURE

Modular, Sustainable, Expandable



Data Repository Example: Health

uva-bi-sdad / sdc.health Public

Watch 1 Fork 0

Code Issues Pull requests Actions Projects Security Insights

main sdc.health / Health Care Services /

Go to file Add file ...

uva-bi-sdad Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/6cc6... 6140fb2 last week History

..			
code			
data			
docs			
Dentists/Service Catchment Scores	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/d791...		last year
Drug and Rehab/Service Catchment Scores	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/c2f8...		last week
EMS/Service Catchment Scores	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/d791...		last year
Health Professionals/VA Graduates	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/1aeb...		last year
Hospitals and Emergency Rooms/Service Access Scor...	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/d6fb...		2 weeks ago
Mental Health/Service Access Scores	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/6cc6...		last week
Nursing Homes	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/258f...		last year
PCNA Measures/Check-up and Dental Visits	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/d5f6...		last week
Physicians	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/d791...		last year
Urgent Care Centers/Service Access Scores	Update from https://github.com/uva-bi-sdad/sdc.health_dev/commit/f77e...		2 weeks ago

Standardization & Metadata

Full Metadata Record

Name
..
measure_info.json
va_hdcttr_2015_2021_employment_access_index.csv.xz

```
{
  "employment_access_index": {
    "aggregation_method": "weighted sum",
    "categories": "",
    "category": "Employment/Workforce Development",
    "citations": "",
    "data_type": "numeric",
    "equity_category": "Accessibility",
    "layer": "",
    "long_description": "Employment access measures the accessibility of jobs in a particular area. Poor j
employment access index is obtained from the Housing + Transportation (H+T) Affordability Index data provided
is calculated by summing the total number of jobs divided by the square of the distance to those jobs. The ind
H+T index website for years 2015, 2019, and 2020. To fill in the missing data for the years 2016-2018, we perf
2021, we estimated employment access values by multiplying the rate of change observed from 2019 to 2020 by th
district level for the state of Virginia.",
    "long_name": "Employment Access Index",
    "measure_type": "",
    "short_description": "Employment access is the job accessibility at a location",
    "short_name": "Employment Access Index",
    "sources": [
      {
        "name": "Housing + Transportation Affordability Index, Center for Neighborhood Technology",
        "url": "https://htaindex.cnt.org/",
        "location": "2015, 2019, and 2020 Datasets",
        "location_url": "https://htaindex.cnt.org/download/",
        "date_accessed": "2023"
      }
    ],
    "statement": "There are {value} jobs per square mile in {region_name}.",
    "type": "",
    "unit": "job",
    "variants": ""
  }
}
```

Standardized Data File Names

- Coverage: Virginia
- Resolutions: Health Districts, Counties, Census Tracts
- Years: 2015 – 2021
- Topic: Employment Access Index

Everything must be both
Machine AND Human Readable

Rich and accessible metadata

Social Data Commons Data Library

Search all measures

Show 10 entries

Search: health

	category	long_name	short_description	short_name	statement	type
36	Health	Dental care geographic availability (2 step-enhanced floating catchment areas)	Index of dental care availability based on supply and demand of providers	Dental care geographic availability	The dental care availability for {features.name} is {value}.	index
37	Health	Dental care availability by count	Count of dentists based on provider locations	Dental care availability (count)	There are {value} dentists in {features.name}.	count
38	Health	Dental care geographic availability (2 step-enhanced floating catchment areas)	Index of dental care availability based on supply and demand of providers	Dental care geographic availability	The dental care availability for {features.name} is {value}.	index
39	Health	Emergency medical services geographic availability (3-step floating catchment areas)	Index of emergency medical services availability based on supply and demand of facilities	Emergency medical services geographic availability	The emergency medical services availability for {features.name} is {value}.	index

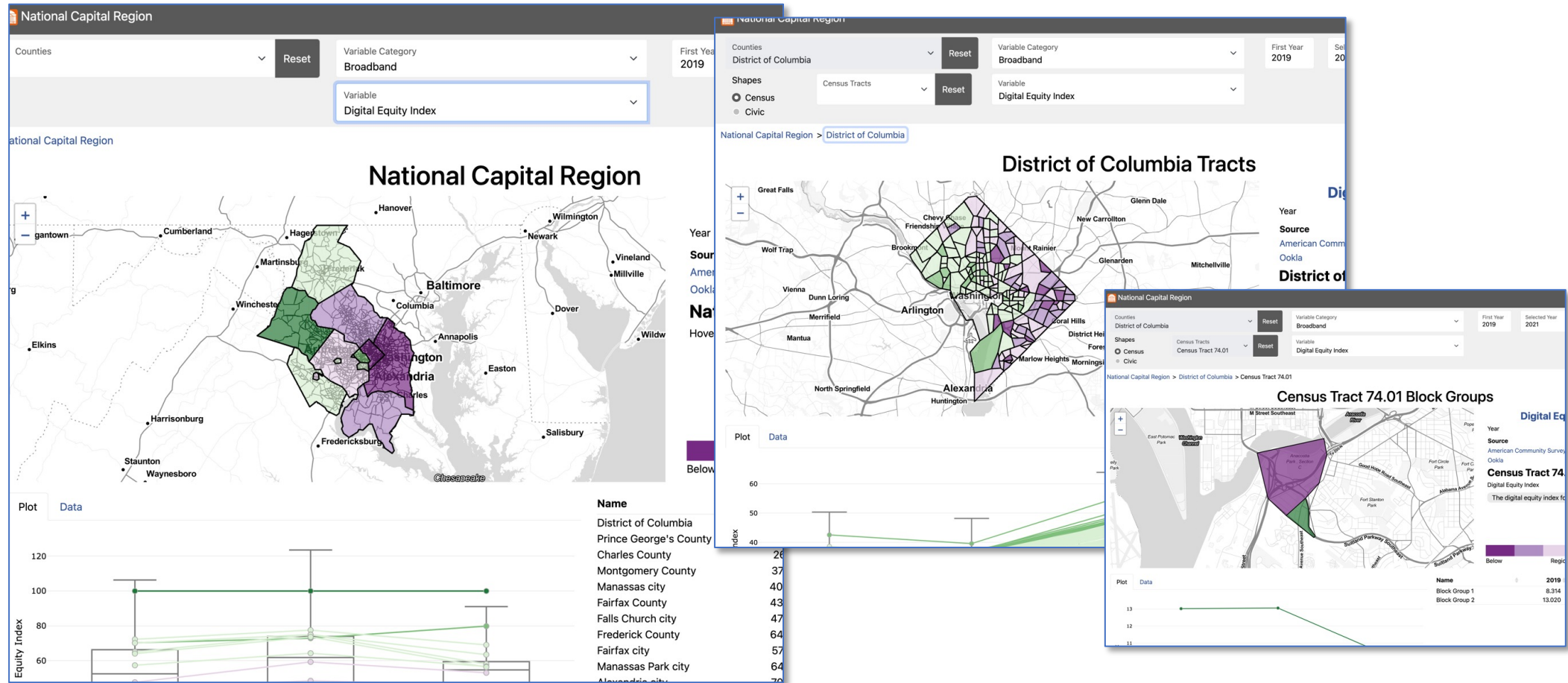
- Describes and contextualizes measures on the dashboard
- Adheres to standards (e.g. FAIR)
- Openly available online as an [interactive data library](#)

DATA COMMONS GENERALIZED ARCHITECTURE

Lightweight JavaScript WebApp, Universally Deployable:

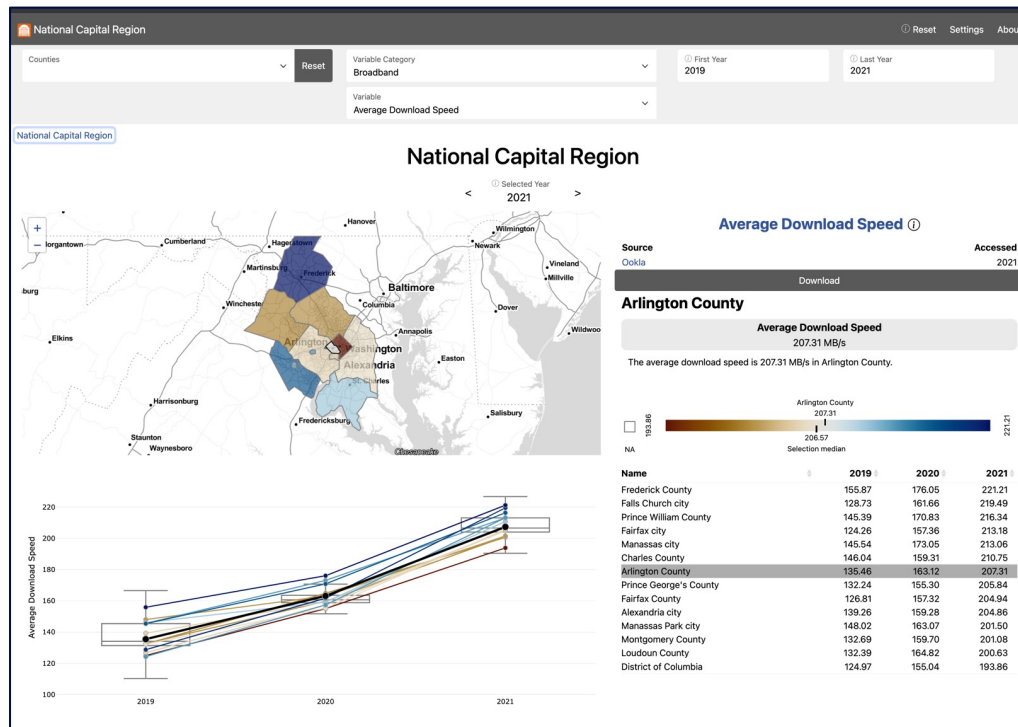
NCR Version (this project: beta): https://uva-bi-sdad.github.io/capital_region/

VDH Version (sister project: deployed, VA only): https://uva-bi-sdad.github.io/vdh_rural_health_site/

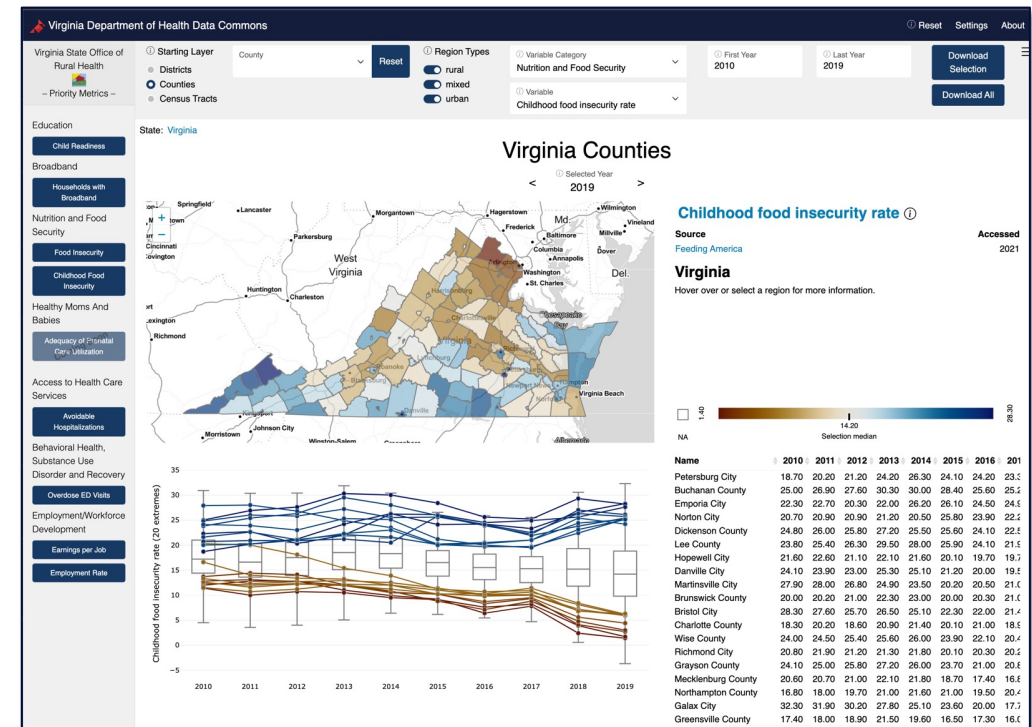


Two Data Commons Projects

1. Social Impact Data Commons to Inform Equitable Growth (Mastercard Center for Inclusive Growth) – National Capital Region
 2. Data Commons to Support Department of Health Strategic Plans (Virginia Department of Health) – State of Virginia
- Both applying the CLD Process to discover data needs



https://uva-bi-sdad.github.io/capital_region



https://uva-bi-sdad.github.io/vdh_rural_health_site

Specific artifacts of the SIDC project

- An Open-Source Data Dashboard (License: CCA 4.0). A lightweight JavaScript-based data dashboard that can run on low-cost/free hosting services like GitHub.
 - NCR Site (https://uva-bi-sdad.github.io/capital_region)
 - VDH Site (https://uva-bi-sdad.github.io/vdh_rural_health_site)
- 11 Open-Source Data Repositories (License: CCA 4.0). Over 150 datasets are currently hosted in 11 repositories freely accessible via GitHub. (<https://github.com/uva-bi-sdad/sdc.all/tree/main/data>)
- Open-Source Dataset Tools (License: CCA 4.0). Multiple R Packages used to create localized datasets.
 - **Catchment** - An R package to calculate spatial access and availability metrics.
 - <https://uva-bi-sdad.github.io/catchment/articles/introduction.html>
 - **Redistribute** - An R package to redistribute population data to alternate geographies.
 - <https://uva-bi-sdad.github.io/redistribute/articles/introduction.html>
 - **Food Security Calculator (Sub-County)** – Created, in process of packaging
 - **Census-tract-level family budget calculator** – Created, in process of packaging
- Data Stories and Walkthroughs
 - https://uva-bi-sdad.github.io/sdc.intro/health_equity.html
 - <https://uva-bi-sdad.github.io/sdc.intro/broadband.html>
 - https://uva-bi-sdad.github.io/sdc.intro/economic_diversity.html

Future Challenges/Opportunities

- Continue Expansion of Data Commons into New Policy Areas and New Geographies (Nationwide)
- Create New Policy Relevant Indicators
- Maintain Data Commons on an Ongoing Basis

Project Information & Contacts

- Project Information:
 - <https://uva-bi-sdad.github.io/sdc.intro>
- NCR Dashboard:
 - https://uva-bi-sdad.github.io/capital_region
- CLD3 Process:
 - <https://datascienceforthepublicgood.org/economic-mobility/research-framework>
- SIDC Measures and Metadata
 - <https://uva-bi-sdad.github.io/sdc.metadata/>
- Contacts:
 - Dr. Aaron Schroeder, Principal Investigator, ads7fg@virginia.edu
 - Dr. Stephanie Shipp, Co-Principal Investigator, sss5sc@virginia.edu