Empowering Policy Research in the Age of AI, Cloud Computing, and Big Data

Khuloud Odeh, CIO and VP, Technology and Data Science
Graham MacDonald, Chief Data Scientist

URBAN INSTITUTE · ELEVATE · THE · DEBATE

Khuloud Odeh,
CIO and VP, Technology and Data Science

Graham MacDonald,
Chief Data Scientist

OUR MISSION IS TO OPEN MINDS, SHAPE DECISIONS, AND OFFER SOLUTIONS THROUGH ECONOMIC AND SOCIAL POLICY RESEARCH.

UNLOCK

ANALYZE

INFORM

THE URBAN INSTITUTE'S NEXT50

Today's Policy Research Challenge

HOW TO CONTINUE DELIVERING POWER THROUGH KNOWLEDGE SO ALL PEOPLE CAN THRIVE IN A FAST-CHANGING WORLD.

Business as Usual: Technical Challenge

- Traditional ways of working take a lot of time
- Reactive Urban takes a single policy proposal and evaluates the impact
- Size, types and sources of input data
- Complexity of modeling method and calculation
- Dissemination and strategic communication of results

Tech and Data for Policy Impact

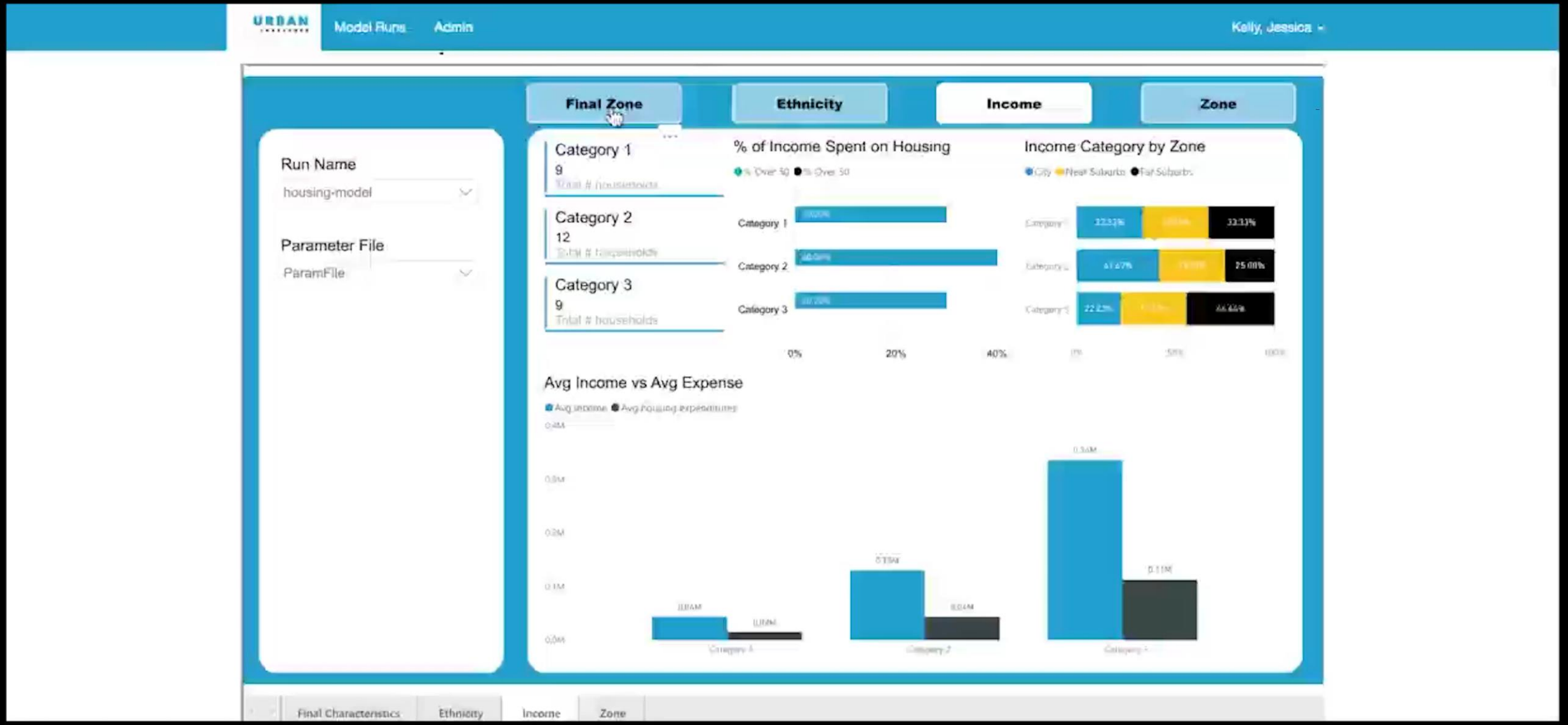
- Rapidly unpack how predictive analytics are shaping society and respond to fast-changing policy decision needs
- Transform the way policies are shaped and designed
- Democratize data and improve Urban's policy research and analysis
- Reduce discrimination in policy decisions and support advocates and community groups in holding their leaders accountable
- Provide companies, governments and others with guidance on what data to share and when, to best serve the public's interest

STORYTELLING

Story 1: Housing Market Forecaster Model


1970's


Final Zone
Ethnicity
Income
Zone
Run Name
housing-model
Parameter File
All
Category 1
9
Total # households
Category 2
12
Total # households
Category 3
% of Income Spent on Housing
% Over 30
% Over 50
Category 1
30.00%
Category 2
40.00%
Category 3
30.00%
0%
20%
Income Category by Zone
City
Near Suburbs
Far Suburbs
Category 1
33.33%
33.33%
33.33%
Category 2
41.67%
33.33%
25.00%
Category 3
22.22%
33.33%
44.44%
0%
50%
100%
2019
0.4M
0.3M
0.2M
0.1M
0.0M
0.04M
0.02M
0.13M
0.04M
0.34M
0.11M
Category 1
Category 2
Category 3


URBAN
Model Runs
Admin
Kelly, Jessica
Final Zone
Ethnicity
Income
Zone
Run Name
housing-model
Parameter File
ParamFile
Category 1
9
Total # households
Category 2
12
Total # households
Category 3
9
Total # households
% of Income Spent on Housing
% Over 30
% Over 50
Category 1
Category 2
Category 3
0%
20%
40%
Income Category by Zone
City
Near Suburbs
Far Suburbs
33.33%
75.00%
66.67%
22.22%
Avg Income vs Avg Expense
Avg Income
Avg Housing Expenditure
0.4M
0.3M
0.2M
0.1M
0.0M
0.11M
Final Characteristics
Ethnicity
Income
Zone

Story 2: Data Bias Assessment Tool

The challenge with biased data

Governments want to make equitable policy decisions using reliable, representative data

Biased data leads to biased algorithms and biased results

Availability of data with spatial information

Examples of policy decisions impacted by biased data

Measuring equity in allocation of local infrastructure, investments, services, and hazards	Park locations, affordable housing, building permits, trash collection centers
Measuring under/over-reporting in resident generated data	311 data, Snowplow request data
Any geographic data	

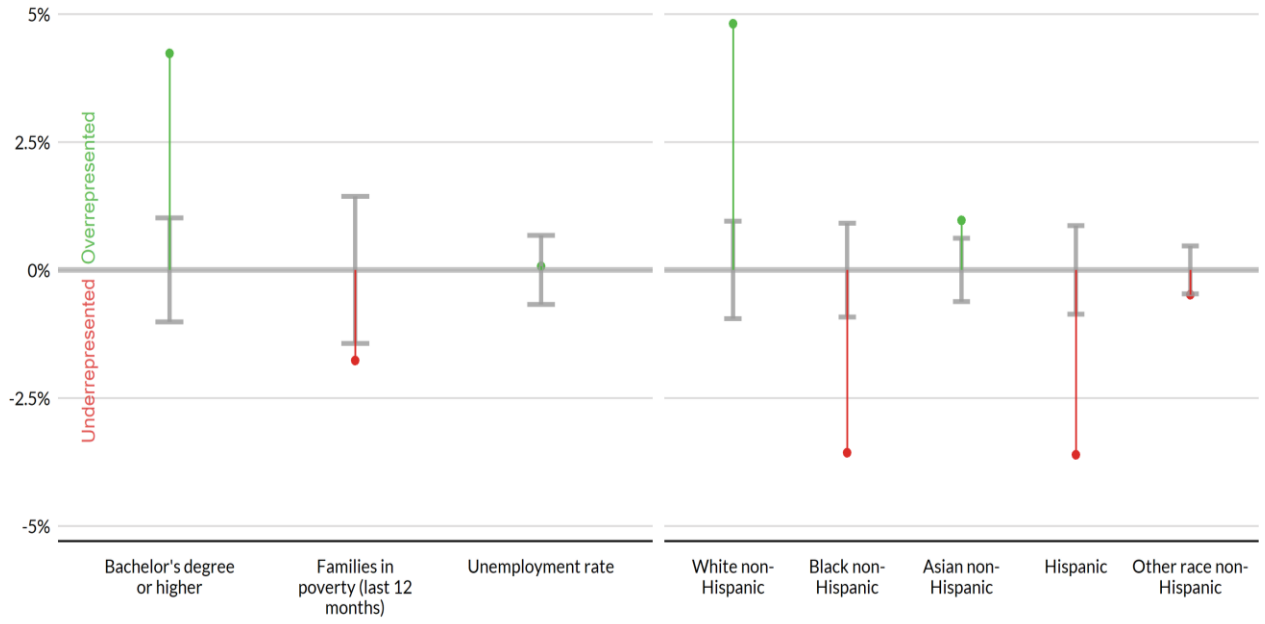
New Data Bias Assessment Tool

INITIAL ALGORITHM WAS BUILT OVER 8 WEEKS SUMMER INTERNSHIP

REFINE ALGORITHM OVERTIME IN ITERATIVE APPROACH

HAVE WORKING PROTOTYPE WITH VALUE TO POLICY DECISION MAKERS AND INTEREST IN BUILDING A PUBLIC INTERFACE

Bias Assessment Tool

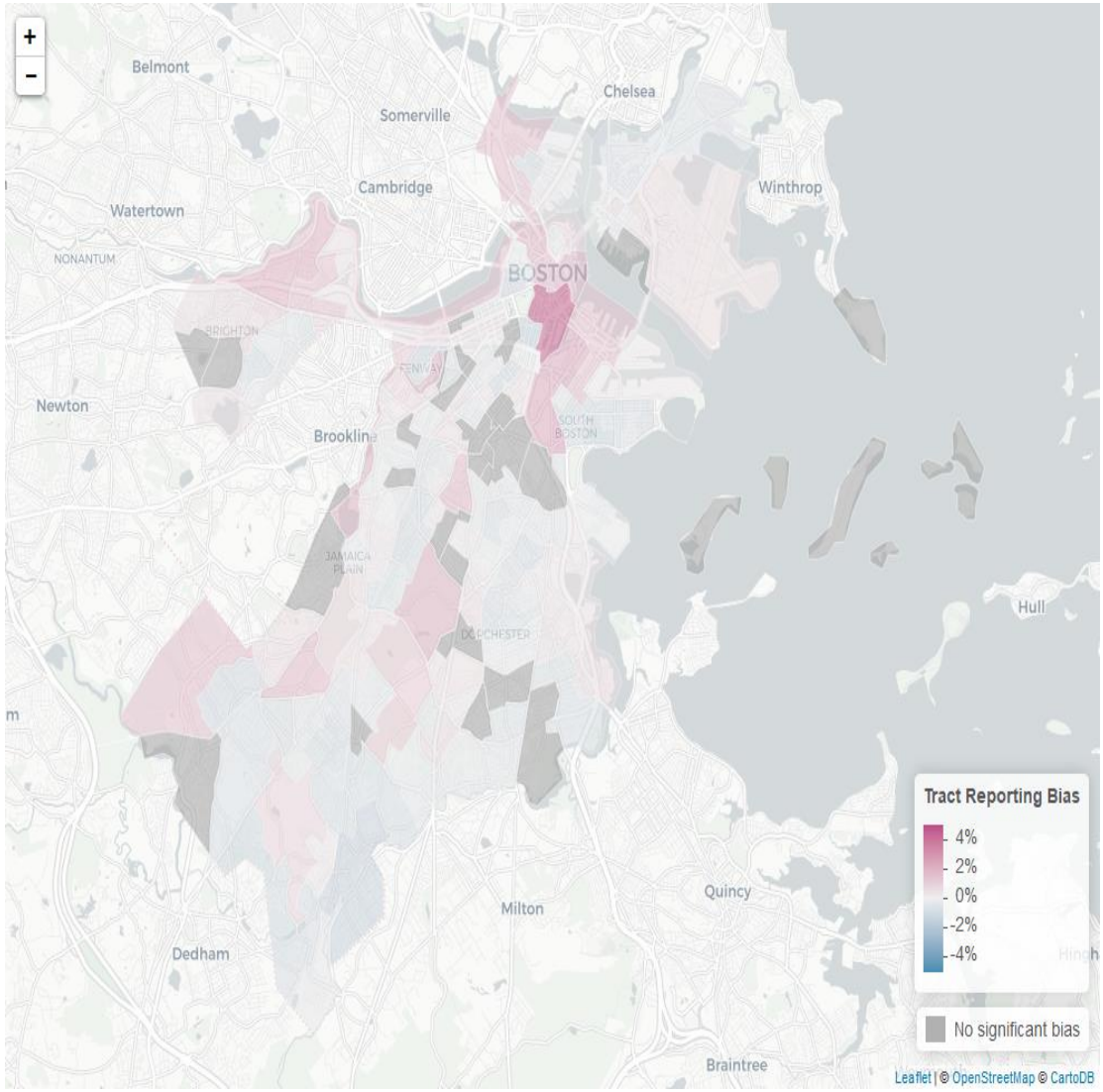


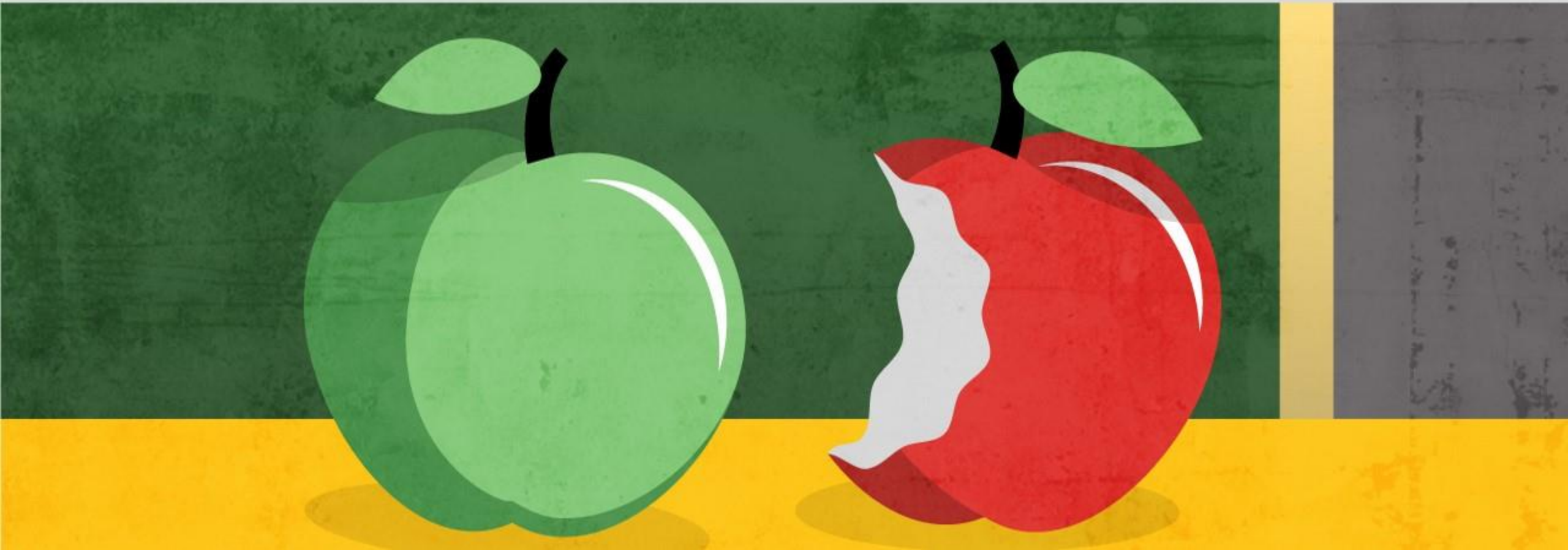
Source: Hubway Stations - Boston Open Data Portal

URBAN INSTITUTE

Upload Data

Export Data





Story 3: Education Data Portal

2013-14 (v.2a)	ZIP (14.0 MB) Flat file (99 MB)	SAS Format ZIP (2.07 KB)	PDF (1.41 MB) ZIP (483 KB)[3]	Flat file (41 KB)	See the file specifications for this year on this site: https://www2.ed.gov/about/inits/ed/edfacts/archived-file-specifications.html
	ZIP (21.3 MB) SAS file (257 MB)	SPSS Code ZIP (5.30 KB)			
2012-13 (v.2a)	ZIP (14.0 MB) Flat file (97 MB)	SAS Format ZIP (1.91 KB)	PDF (1.31 MB) ZIP (488 KB)[3]	Flat file (41 KB)	See the file specifications for this year on this site: https://www2.ed.gov/about/inits/ed/edfacts/archived-file-specifications.html
	ZIP (19.9 MB) SAS file (247 MB)	SPSS Code ZIP (5.10 KB)			
20 (v.2a)*	ZIP (15.1 MB) file (101	SAS Format ZIP (2.16 KB)	PDF (1.63 MB)	Flat file (41 KB)	See the file specifications for this year on this site: https://www2.ed.gov/about/inits/ed/edfacts/archived-file-specifications.html
	ZIP MB) SAS fi 69 MB	SPS 6.42 K			
2010-11 (v.2a)*	ZIP (14.8 MB) Flat file (97.7 MB)	Format ZIP (1.94 KB)	PDF (3.6 MB) ZIP (612 KB)[3]	Flat file (40.5 KB)	See the file specifications for this year on this site: https://www2.ed.gov/about/inits/ed/edfacts/archived-file-specifications.html
	ZIP (20.8 MB) SAS file (270 MB)	SPSS Code ZIP (5.01 KB)			
2009-10 (v.2a)*	ZIP (15 MB) Flat file (95.7 MB)	SAS Format ZIP (1.5 KB)	PDF (2.79 MB)[3]	Flat file (39 KB)	See the file specifications for this year on this site: https://www2.ed.gov/about/inits/ed/edfacts/archived-file-specifications.html
	ZIP (22.3 MB) SAS file (270 MB)	SPSS Code ZIP (4.67 KB)			
2008-09 (v.1b)*	ZIP (14 MB) Flat file (96 MB)	SAS Format ZIP (1.5 KB)	PDF (1.9 MB)	Flat file	See the file specifications for this year on this site: https://www2.ed.gov/about/inits/ed/edfacts/archived-file-

5 Years ago…

URBAN INSTITUTE

EDUCATION DATA EXPLORER

SCHOOLS | DISTRICTS | COLLEGES | ABOUT | FULL API

2019

K–12 Schools

Publicly available data on K–12 schools include basics, such as location and grade offerings, along with information about enrollment demographics, staffing levels, student discipline, and more. The data in this section come from the National Center for Education Statistics' Common Core of Data and the Civil Rights Data Collection.

Data Explorer

Build your own dataset using our guided editor. Select the schools, districts, or states you're interested in, and then choose the information you want about those schools or the schools in those districts or states.

PICK YOUR SCHOOLS

First, select the schools you're interested in.

PICK YOUR DATA

Next, tell us what information you want about those schools.

EXPORT YOUR DATA

The data you select will be displayed in a downloadable CSV file. Complex requests might generate multiple files.

USE YOUR DATA

Every request comes with a data dictionary to help you make sense of the data.

Build your own dataset using our guided editor. Select the schools, districts, or states you're interested in, and then choose the information you want about those schools or the schools in those districts or states.

PICK YOUR SCHOOLS

First, select the schools you're interested in.

PICK YOUR DATA

Next, tell us what information you want about those schools.

EXPORT YOUR DATA

The data you select will be displayed in a downloadable CSV file. Complex requests might generate multiple files.

USE YOUR DATA

Every request comes with a data dictionary to help you make sense of the data.

CHOOSE SCHOOLS | CHOOSE YOUR DATA

SELECT DATA

Time Frame

Start year: Select

End year: 2016

Geography

State *

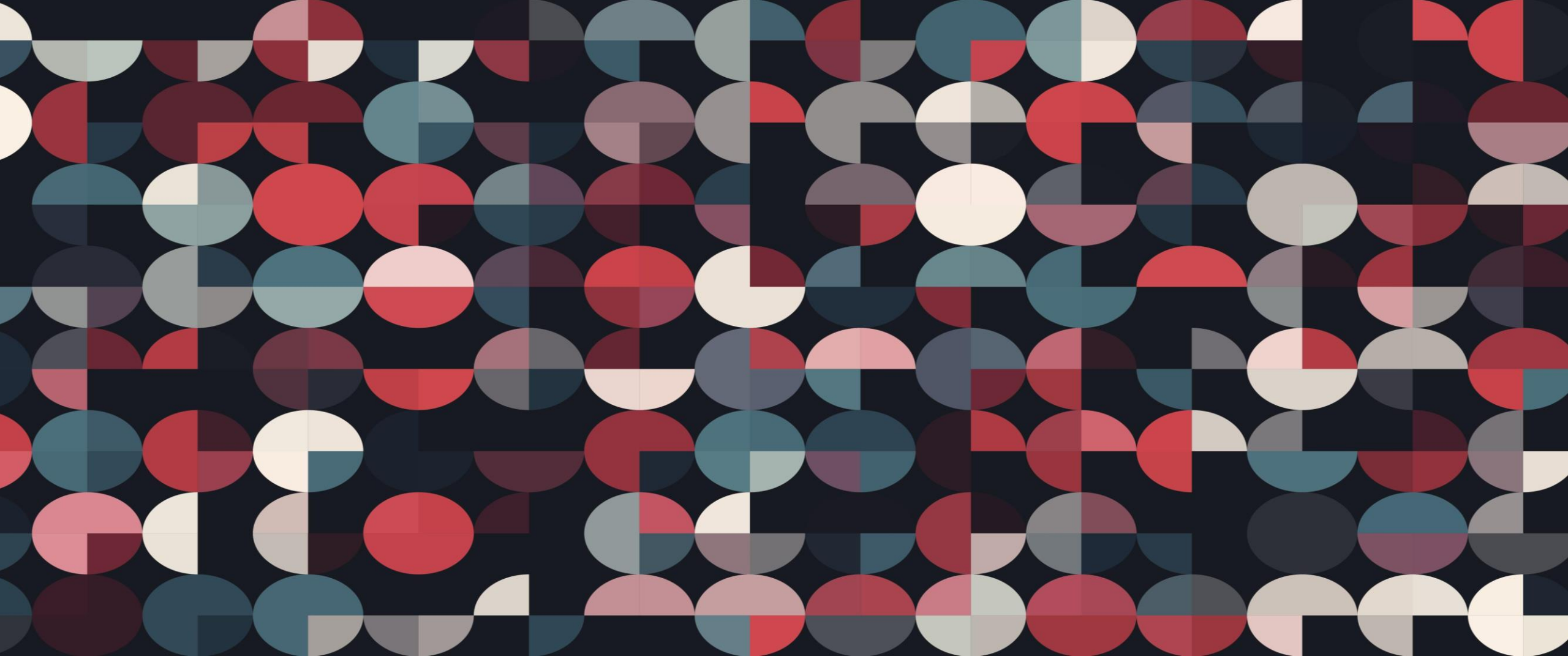
CLEAR

Search for one or more states

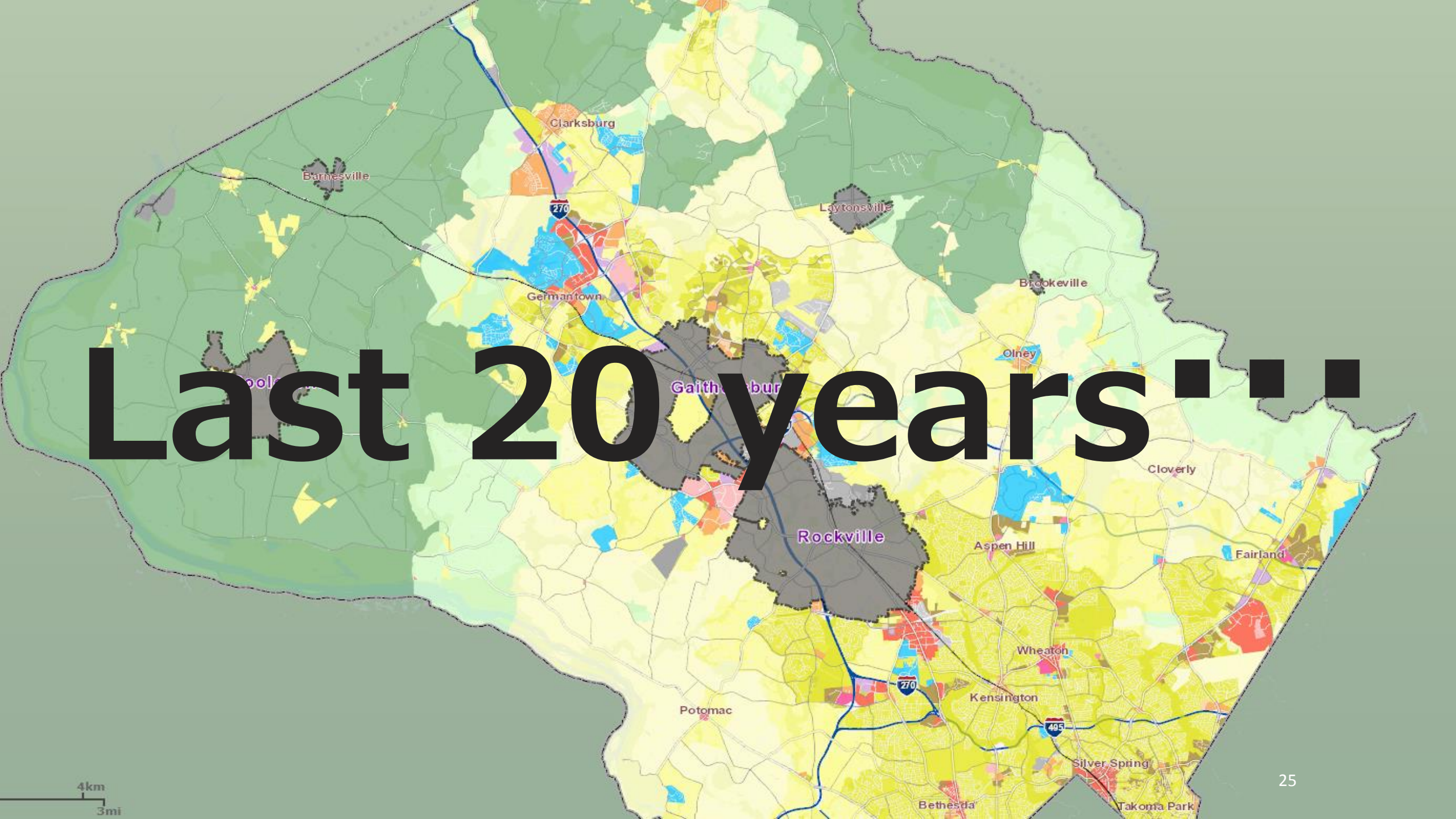
Approximate results: 0 records from 0 schools

* You must select at least one state before seeing results

FEEDBACK +



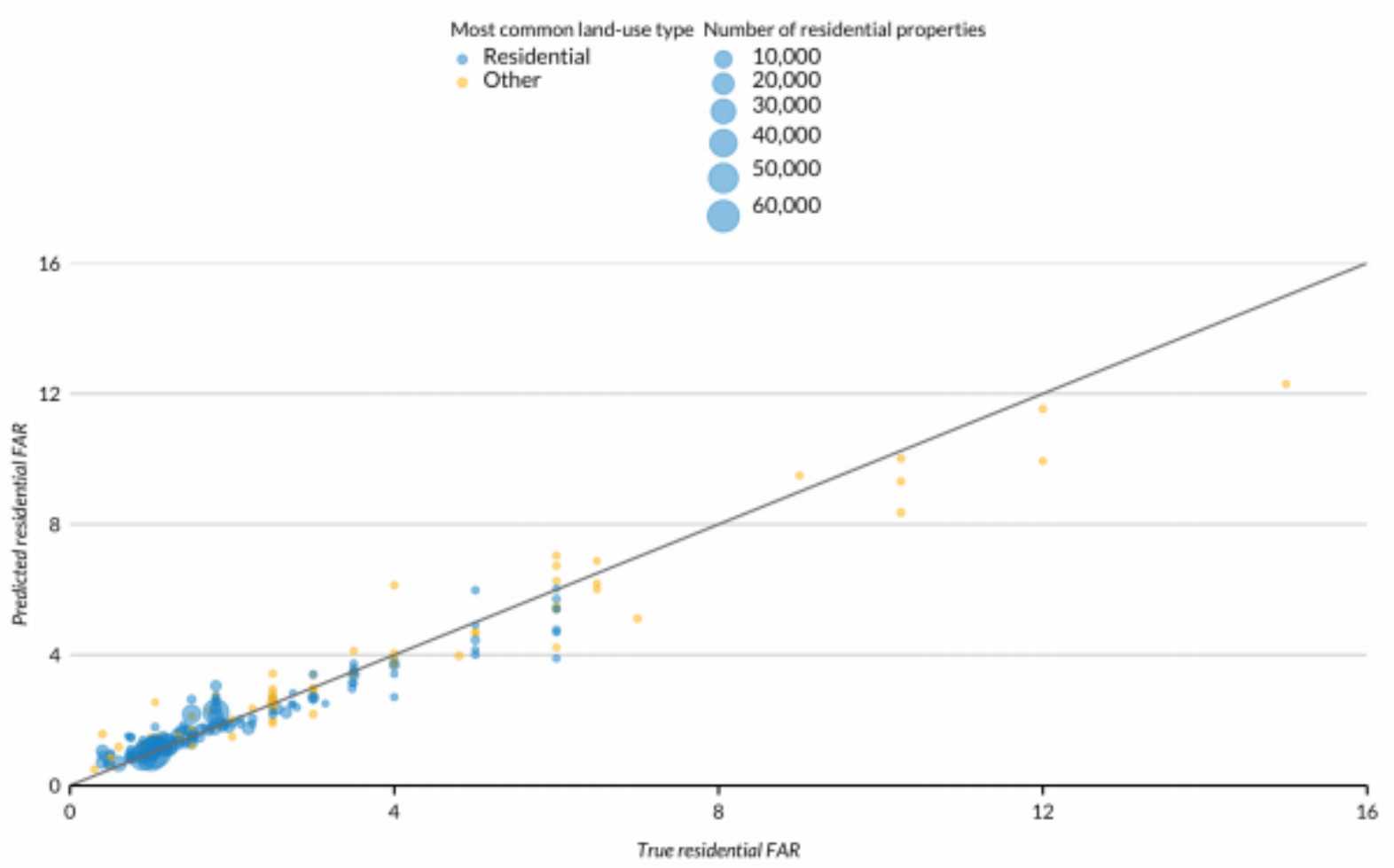
Story 4: AI in policy research


Last 20 years…
Clarksburg
Barnesville
Laytonsville
Brookeville
Germantown
Olney
Rockville
Cloverly
Aspen Hill
Fairland
Wheaton
Kensington
Potomac
Silver Spring
Bethesda
Takoma Park
270
495
4km
3mi

2019

FIGURE 2

Predicted and True FAR: In-Sample DC, Montgomery County, and Arlington County Model



URBAN INSTITUTE

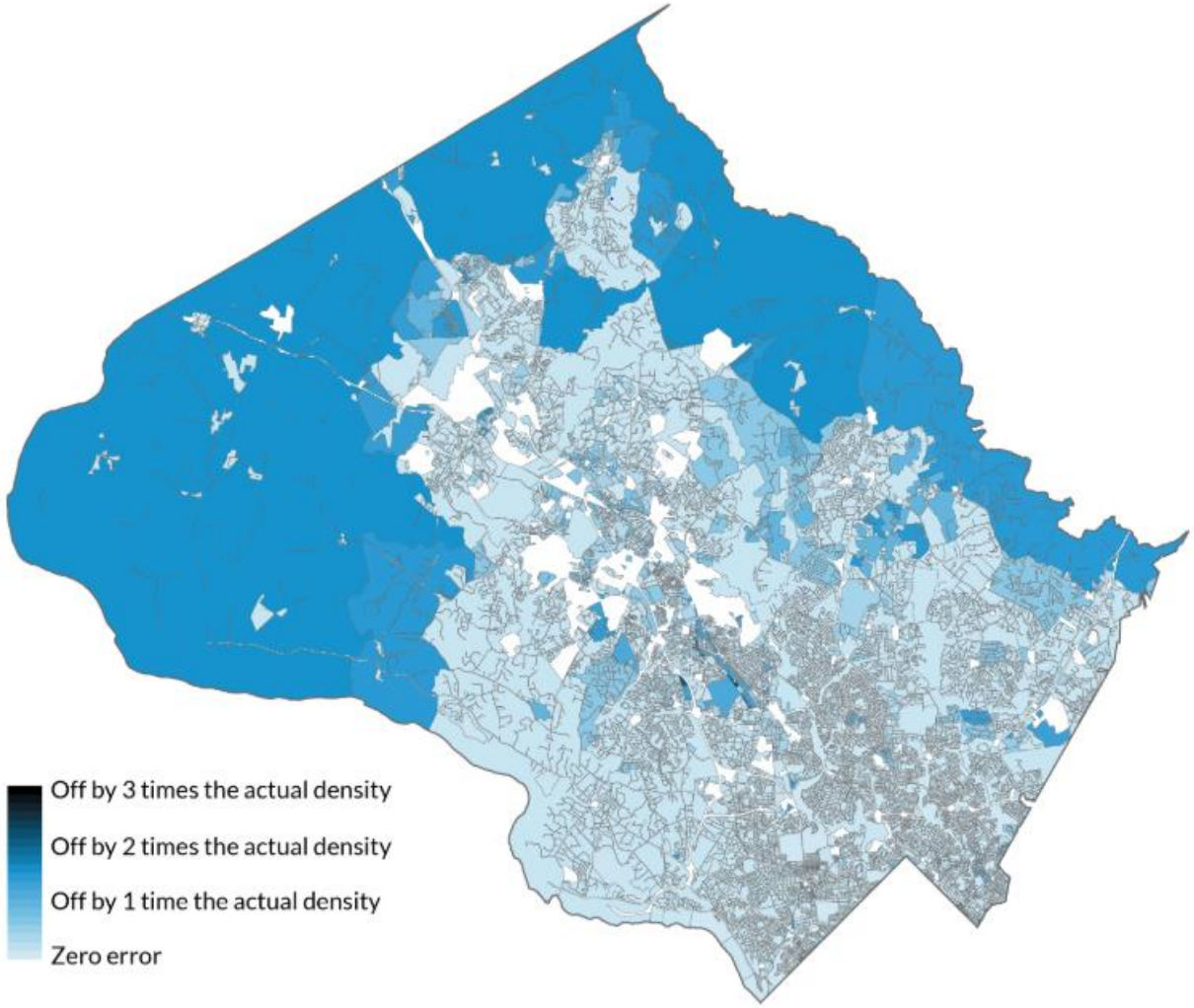
Sources: ZTRAX and local zoning ordinance data. Data were provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the authors and do not reflect the position of Zillow Group.

Note: FAR = floor area ratio.

Machine Generated

We Can Predict Density Limits across Much of Nonrural Montgomery County

Relative error between actual and predicted maximum-allowed FAR



URBAN INSTITUTE

Source: ZTRAX and local zoning ordinance data. Data were provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at http://www.zillow.com/ztrax. The results and opinions are those of the authors and do not reflect the position of Zillow Group.
Notes: FAR = floor area ratio. Relative error reflects the absolute value of the difference between the actual and predicted maximum FAR permitted by right for the average residential property in a given zone, as a ratio of the actual permitted FAR.

OUR PROCESS

“Think Big, Start Small, Fail Fast, Scale Rapidly”

How did we get here?

- Strategic Investment
- Leadership
- Resilient infrastructure
- Cloud Computing & Modern Architecture
- Agility & Rapid Prototyping
- Accelerating Innovation

Benefits of Our Process

Answering complex questions, but doing it quickly and effectively

Collaboration at each level/layer between technologists and researchers

Comfort in change and revision

Speed and independence

Different products for different audiences possible

Components are portable

A series of small investments

More Information

- Website: Urban.org
- Emails and Twitter
  - kodeh@urban.org
    @kodeh
  - gmacdonald@urban.org
    @GrahamIMac
- Data@Urban Blog
  - medium.com/@urban_institute/