Implementing Housing Analytics locally

Georgia Smart Webinar. May 9, 2019

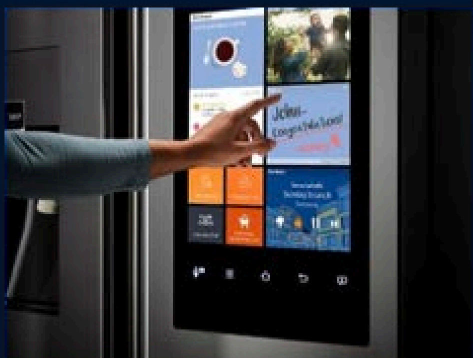
Dr. Omar Isaac Asensio
Assistant Professor, School of Public Policy
Georgia Institute of Technology

Overview

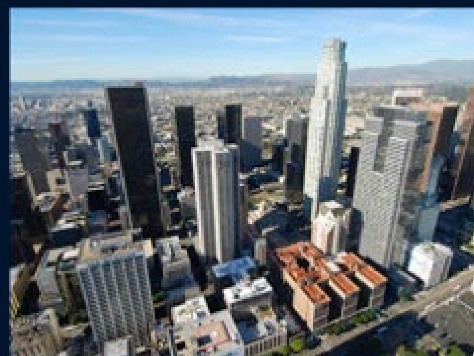
- Research overview: smart data housing analytics (open data versus analytics)
- Albany, GA
- Some considerations on implementing effective data transitions and cultural change around big data

About the speaker

Randomized Policy Experiments



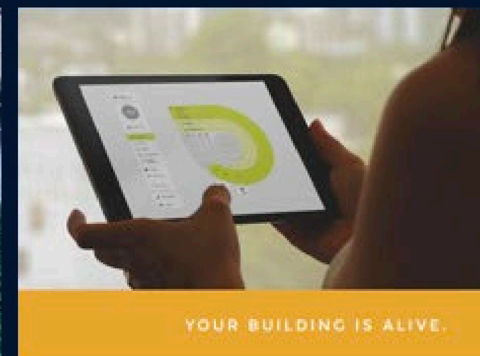
Energy Conservation Randomized Controlled Trials (RCTs)



Smart Grid and Energy Efficiency Program Evaluation

Big Data Analytics and Smart Mobility

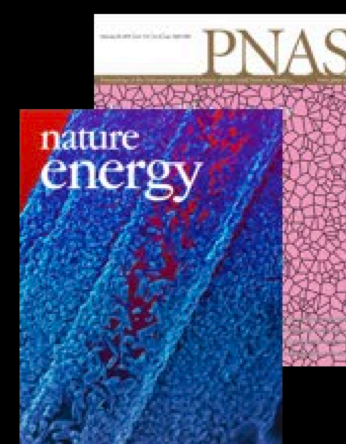
Machine Learning and Real-time Intelligence in Sustainable Transportation Infrastructure



Civic Data Science and Urban Sustainability & Computational and Statistical Models

Social and Behavioral Dimensions of Energy Use

SCIENTIFIC AMERICAN
NBC NEWS
The Washington Post
THE ECONOMIC TIMES
YAHOO! News
Utility DIVE
CBS RADIO
Los Angeles Times
npr



The Objective

1. Through the Georgia Smart program, Georgia tech has partnered with the City of Albany

2. To build a housing data inventory to better manage housing investments and address issues of neighborhood blight

The Objective

1. Through the Georgia Smart program, Georgia tech has partnered with the City of Albany

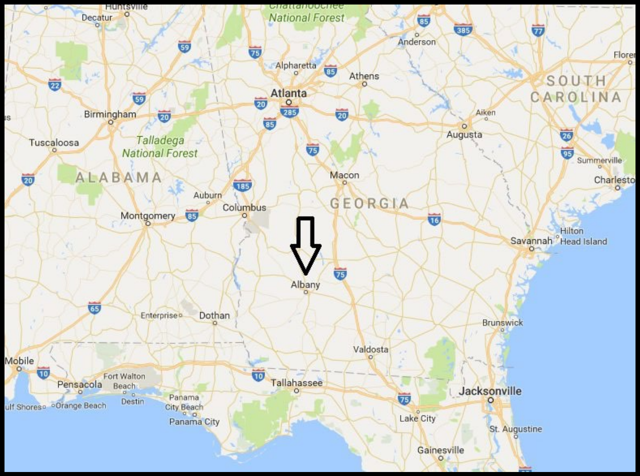
2. To build a housing data inventory to better manage housing investments and address issues of neighborhood blight

<u>Proposal Lead</u>
City of Albany, GA

<u>In Collaboration With</u>
The Department of Community & Economic Development (DCED)
Dougherty County, GA
Albany Georgia Initiative for Community Housing (GICH)
Fight Albany Blight (FAB)
Albany Works! (New Community Partner)
Dr. Omar I. Asensio (Georgia Tech)

Albany, GA

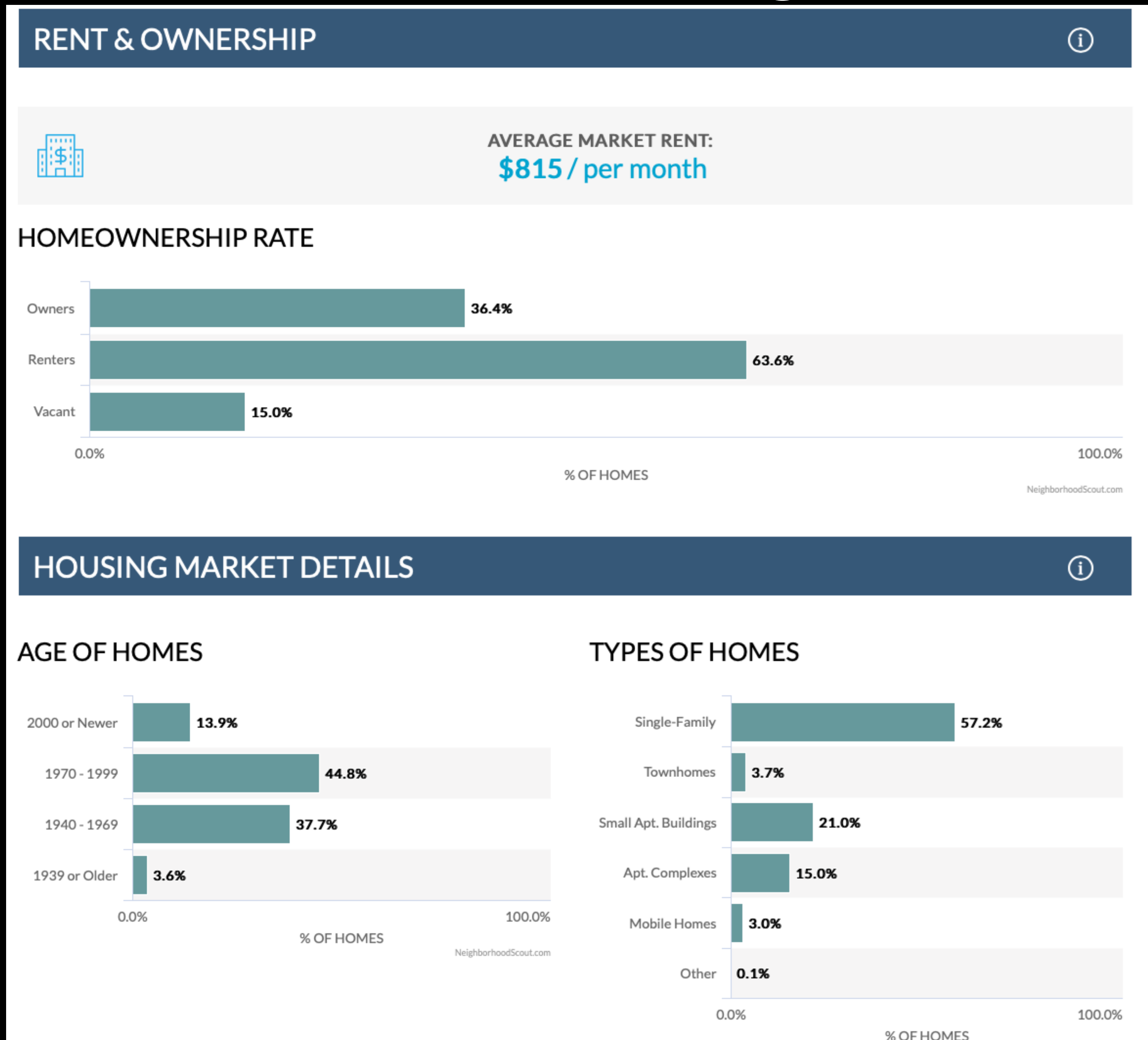
CITY OF Albany GEORGIA



A community in South Georgia,
“Big city amenities, small town feel.”

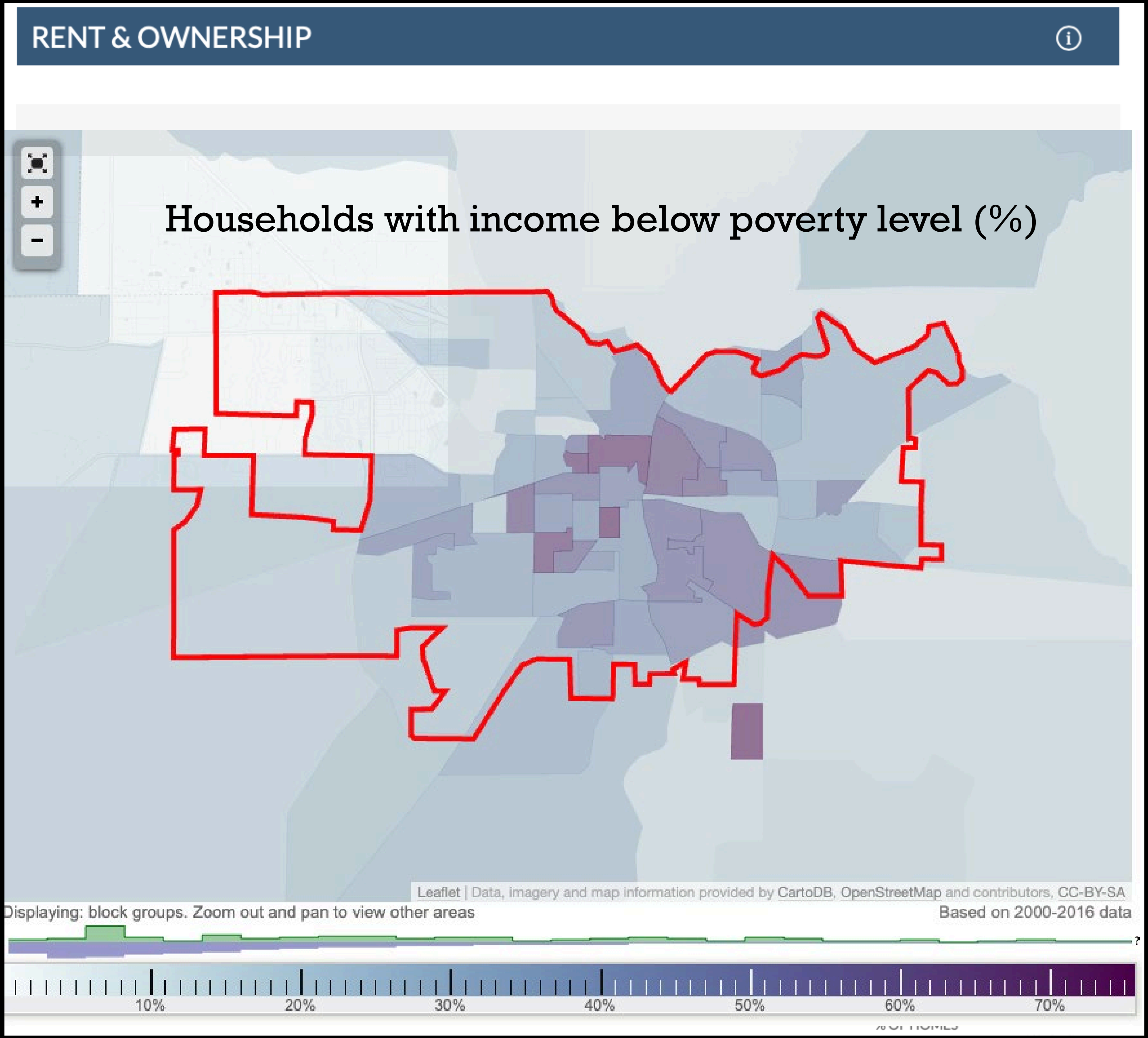
Population: ~73,000

Why housing?



RENT & OWNERSHIP
AVERAGE MARKET RENT:
$815 / per month
HOMEOWNERSHIP RATE
Owners
36.4%
Renters
63.6%
Vacant
15.0%
0.0%
100.0%
% OF HOMES
NeighborhoodScout.com
HOUSING MARKET DETAILS
AGE OF HOMES
2000 or Newer
13.9%
1970 - 1999
44.8%
1940 - 1969
37.7%
1939 or Older
3.6%
0.0%
100.0%
% OF HOMES
NeighborhoodScout.com
TYPES OF HOMES
Single-Family
57.2%
Townhomes
3.7%
Small Apt. Buildings
21.0%
Apt. Complexes
15.0%
Mobile Homes
3.0%
Other
0.1%
0.0%
100.0%
% OF HOMES

Why housing?



RENT & OWNERSHIP
Households with income below poverty level (%)
Leaflet | Data, imagery and map information provided by CartoDB, OpenStreetMap and contributors, CC-BY-SA
Displaying: block groups. Zoom out and pan to view other areas
Based on 2000-2016 data
10%
20%
30%
40%
50%
60%
70%

The Challenge

1. Important city data is typically siloed across various information systems and city departments

2. To increase transparency, many governments are moving towards open data, but users have diverse interests in the usage of the data and requirements for data presentation

3. Open data alone is often insufficient to paint a picture of how well our policies and programs are working

1. Tell us about your data in as few words as possible. What are the basics?

Our data is eclectic.

1. Tell us about your data in as few words as possible. What are the basics?

Our data is eclectic.

Our data typically has one or more of the three main elements: an x-y coordinate, an address, or a parcel ID. We anticipate aggregating several hundreds of thousands of records.

1. Tell us about your data in as few words as possible. What are the basics?

Our data is eclectic.

Our data typically has one or more of the three main elements: an x-y coordinate, an address, or a parcel ID. We anticipate aggregating several hundreds of thousands of records.

Actual number: now estimated to be 150+ million records across 21 city departments; with real-time automated data updates such as utilities, crime, or disaster recovery data.

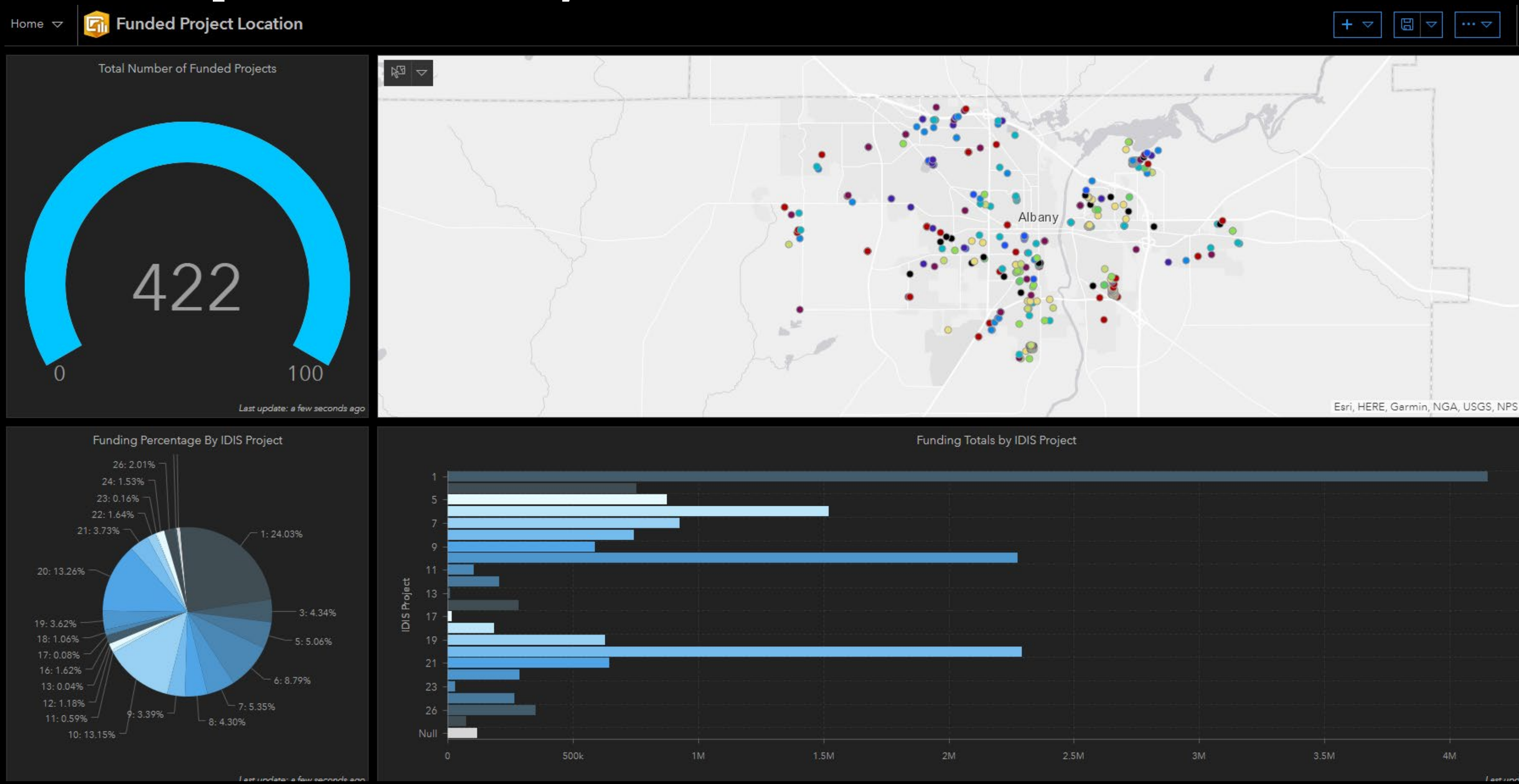
1. Tell us about your data in as few words as possible. What are the basics?

Share a known pattern in your data set using numbers or a simple visualization.

One data pattern in our initial dataset is that the funds allocated for most housing projects are in impoverished portions of the city.

Share a known pattern in your data set using numbers or a simple visualization.

One data pattern in our initial dataset is that the funds allocated for most housing projects are in impoverished portions of the city.



Share a known pattern in your data set using numbers or a simple visualization.

What is the biggest opportunity you see in your data?
Who could make use of them and to what end?

The biggest opportunity in our data is simply data access. This will be a game changer in the ability of the city leaders as well as citizens to have access to data that could assist them to make better informed and data driven decisions.

What is the biggest opportunity you see in your data?
Who could make use of them and to what end?

I actually have two major concerns. Data accuracy and the possibility of disclosing something that is too personal for public consumption.

I wouldn't want any data that we provide to be used to intentionally or even unintentionally harm someone else.

What is the greatest risk you see in your data?
What wouldn't you want them to be used for?

Data Story: An example in Housing and energy consumption in LA county

Residential Per Square Foot Energy Consumption

	Per Sq. Ft. Ranking	Cities	Population (2010)	Income	Total Usage (Residential)	Median Per Sq Ft Consumption
	1	Compton	96,986	$44,851	1.6 trillion BTU	55,058
	2	Unincorporated Metro	302,152	$38,151	4.1 trillion BTU	54,958
	3	Hidden Hills	1,985	$250,001	163 billion BTU	53,015
	4	San Fernando	23,645	$47,125	334 billion BTU	51,956
	5	Hawaiian Gardens	14,860	$50,909	142 billion BTU	50,421
	6	Rolling Hills	1,860	$219,688	141 billion BTU	49,920
	7	Burbank	103,340	$67,156	2.5 trillion BTU	49,920
	8	Inglewood	111,800	$44,146	1.9 trillion BTU	49,693
	9	Lynwood	70,699	$44,896	800 billion BTU	49,242
	10	La Canada Flintridge	21,405	$156,583	901 billion BTU	48,955
...	...	...	...	...	...	...
	84	Lawndale	32,769	$52,447	475 billion BTU	36,334
	85	Diamond Bar	46,826	$88,760	1.2 trillion BTU	33,440
	86	Cudahy	23,805	$41,713	247 billion BTU	32,821
	87	Walnut	29,156	$97,885	564 billion BTU	31,809
	88	Bell Gardens	42,072	$38,917	438 billion BTU	31,316
	89	Uninc. Santa Monica Mountains	30,804	$126,594	536 billion BTU	30,844
	90	Uninc. San Fernando Valley	3,918	$97,656	49 billion BTU	30,132
	91	Westlake Village	6,022	$94,420	331 billion BTU	27,244
	92	Avalon	3,569	$51,160	19 billion BTU	12,358
	93	Signal Hill	11,077	$66,250	166 billion BTU	11,361

Los Angeles Energy Atlas

Data Story: An example in Housing and energy consumption in LA county

Residential per Capita Energy Consumption Top and Bottom 10 Residential Per Capita Cities

	Per Capita Ranking	Cities	Population (2010)	Income	Total Usage (Residential)	Per capita consumption
	1	Mailibu	8,097	$115,168	771 billion BTU	95 million BTU
	2	Hidden Hills	1,985	$250,001	163 billion BTU	82 million BTU
	3	Rolling Hills	190,790	$219,688	141 billion BTU	76 million BTU
	4	Westlake Village	302,152	$94,419	331 billion BTU	55 million BTU
	5	Beverly Hills	211,278	$88,589	1.8 trillion BTU	53 million BTU
	6	Calabasas	151,745	$132,583	960 billion BTU	50 million BTU
	7	Rolling Hills Estates	149,746	$146,296	341 billion BTU	49 million BTU
	8	Santa Clarita	149,343	$84,375	9.1 trillion BTU	49 million BTU
	9	Palos Verdes Estates	145,438	$160,432	621 billion BTU	46 million BTU
	10	Bradbury	103,340	$124,167	59 billion BTU	46 million BTU
...	...	...	...	...	...	...
	89	South Gate	5,591	$44,278	1.1 trillion BTU	12 million BTU
	90	Lynwood	11,077	$44,896	800 billion BTU	11 million BTU
	91	Paramount	1,985	$42,440	606 billion BTU	11 million BTU
	92	Maywood	13,158	$36,455	293 billion BTU	11 million BTU
	93	Bell Gardens	14,860	$38,917	438 billion BTU	10 million BTU
	94	Cudahy	1,860	$41,713	247 billion BTU	10 million BTU
	95	Bell	1,292	$38,563	364 billion BTU	10 million BTU
	96	Huntington Park	3,918	$39,040	610 billion BTU	10 million BTU
	97	Hawaiian Gardens	1,422	$50,909	142 billion BTU	10 million BTU
	98	Avalon	3,569	$51,160	19 billion BTU	5 million BTU

Los Angeles Energy Atlas

Using Analytics Dashboards for Policy and Program Evaluation

1. A Housing data inventory needs record linkage to other possible outcomes (Housing investment > Utility consumption)

2. Open Data Dashboards can be misleading and promote misinformation about performance

   We often need statistical methods to uncover hidden patterns in the data

   Housing starts may co-vary with redevelopment projects, so the true benefits are unclear without a reference case or baseline (counterfactual)

3. We cannot study only those housing investment projects that choose to participate e.g. self selection bias (Asensio and Delmas 2016)

Types of Data Sources for Housing Analytics

1. Curated data

   Data sources under direct operational control by the City (traditional administrative data)

2. Automated data

   Data sources generated as inherent to the automatic function of a system (e.g. transit data, social data, smart meter utility data)

3. Volunteered data

   Data sources collected manually or automatically (on a platform) as a result of public participation (e.g. Arc GIS Hub)

Next steps 2019-2020

- Integrate housing data with all other available city data using ArcGIS Hub
- Public engagement workshops to understand the needs of a public facing "Data Hub"
- Deploy statistical and computational methods to evaluate the performance of 20 years of housing investments in Albany
- Summer 2019: Civic Data Science students (Chris LeDantec, Ellen Zegura)

The Research Team

Asensio Lab May 2019

Q&A