

The background image shows a panoramic view of the New York City skyline during sunset. The sky is filled with warm, orange and yellow hues. The Empire State Building stands prominently in the center, its Art Deco spire reaching towards the top. To its right, the One World Trade Center rises from the ground. The Hudson River is visible in the foreground, with several bridges spanning it. The city's dense grid of buildings is visible across the water.

Week 2 Special Data

June 2021

Theory of Viz

W1 - What is Visualization and Why Visualize

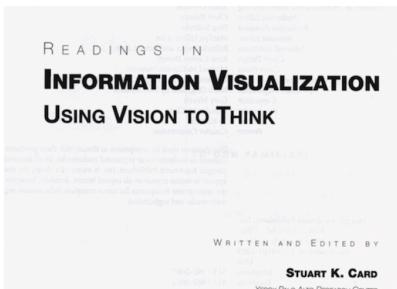
Recommended Reads

- [The Purpose of Visualization \(Stanford CS 448B Slides\)](#) Concise, illustrated slide that covers the definition and rationale of visualization. Recommend to read pdf page 1-25.

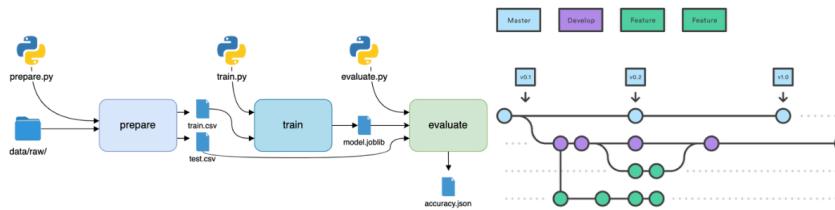


Visual Encoding

- [Information Visualization, in Readings in Information Visualization by C,M,S - Chapter 1](#) Comprehensive, research-oriented book chapter on info viz. Used in many courses, it may give you different angles to think about visualization and how it amplifies our cognition.



Data Cleaning



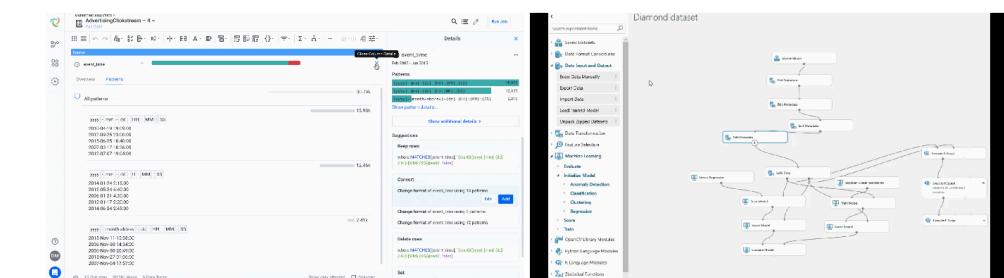
2. Script-based data transformation / cleaning

- Data profiling
- Smart fuzzy matching
- Program synthesis

Master Table								
ID	First Name	Last Name	Address 1	Address 2	Phone	Nat ID	State	ZIP
458	John	O'Neil	16 M. St	Null	516465409	256499458	NY	#0001-1
2398	Jonathan	O'Neil	Null	Norwich	+1578698754	#####	NP	
CL_98	John O	Medal Sweer	NR	00000	256499458	NEW YORK	#000100	OCT.11
Null	Philip	O'neal	Ikeoff	Norwich	256499458	#0001-4381		11/10/1988

3. Automatic/Semi-automatic Data Cleaning

4. Visualization and GUI for the above steps



Data

Cars

Change

Fields

- A Cylinders ▼ +
- A Name ▼ +
- A Origin ▼ +
- Year ▼ +
- # Acceleration ▼ +
- # Displacement ▼ +
- # Horsepower ▼ +
- # Miles per Gallon ▼ +
- # Weight in lbs ▼ +
- # COUNT +

Wildcards

- A Categorical Fields +
- Temporal Fields +
- # Quantitative Fields +

Encoding

Clear

x drop a field here

y drop a field here

column drop a field here

row drop a field here

Marks

auto ▾

size drop a field here

color drop a field here

shape drop a field here

detail drop a field here

text drop a field here

any drop a field here

Filter

Filter invalid numbers

Specified View

No specified visualization yet. Start exploring by dragging a field to encoding pane on the left or examining univariate summaries below.

Related Views

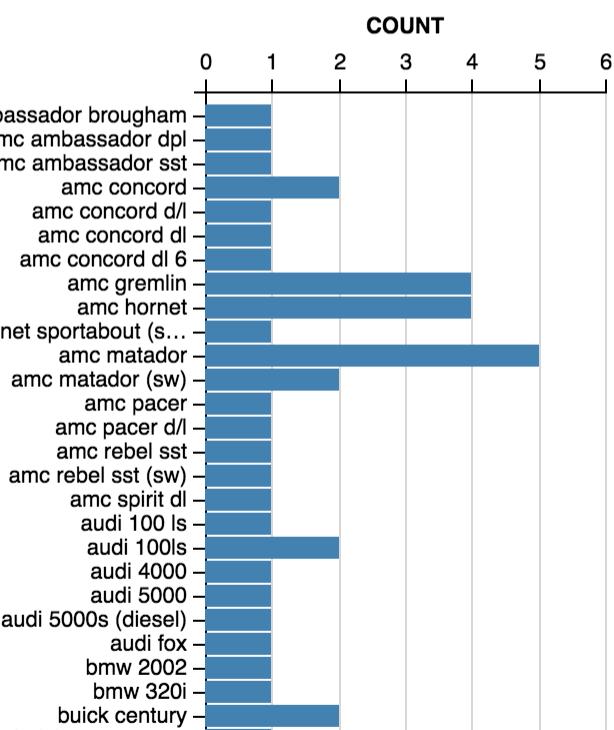
Univariate Summaries Hide

Univariate Summaries

A Cylinders # COUNT

3

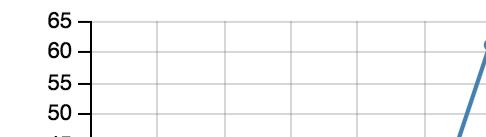
A Name # COUNT

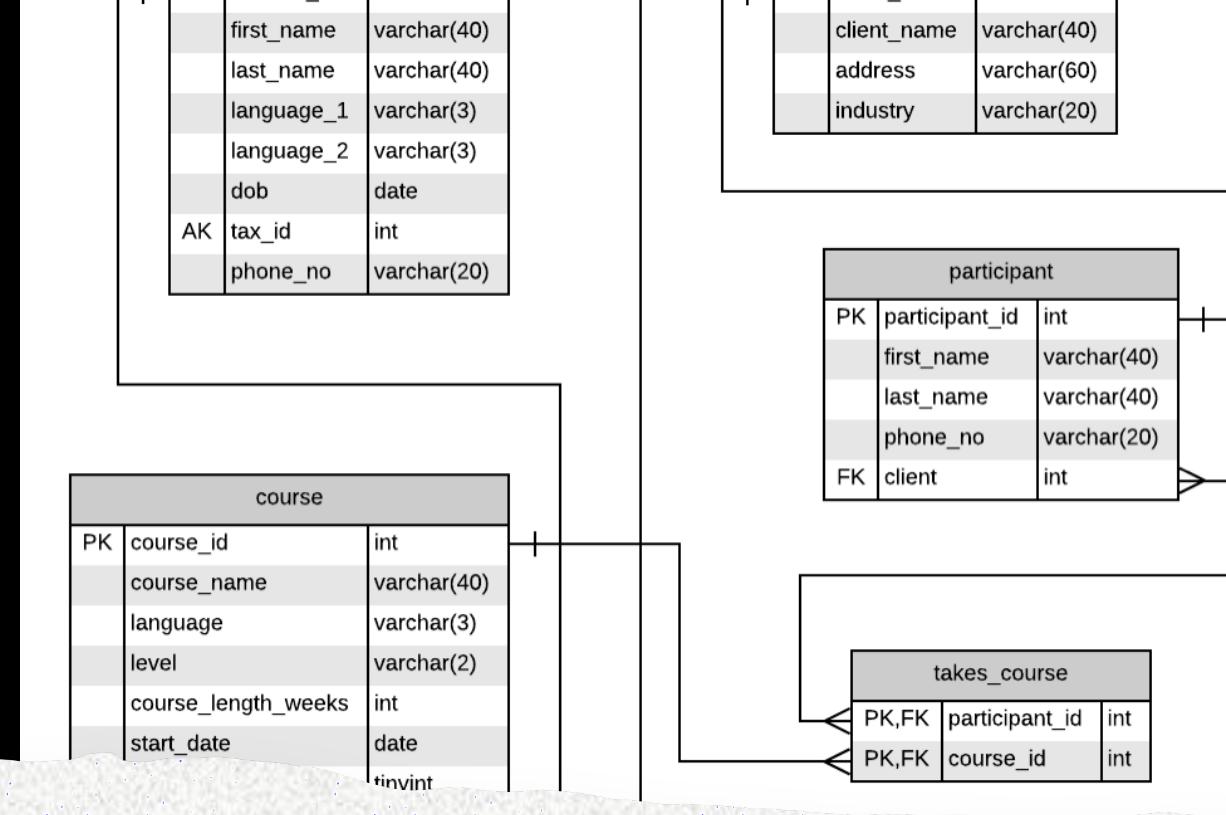


A Origin # COUNT

Origin	COUNT
Europe	15
Japan	75

YEAR (Year) # COUNT

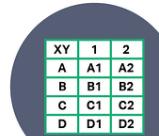




Why these data types?

Structured Data vs Unstructured Data

Can be displayed
in rows, columns and
relational databases



Numbers, dates
and strings

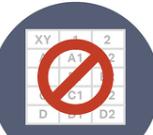
Estimated 20% of
enterprise data (Gartner)



Requires less storage



Cannot be displayed
in rows, columns and
relational databases



Images, audio, video,
word processing files,
e-mails, spreadsheets

Estimated 80% of
enterprise data (Gartner)



Requires more storage



Also, we want to *analyze* the special
data together with our tabular data

Special Data

*What special data
do we have?*



STATISTICS OF OCCUPAT

d to carpenter," "apothecary's ap-
pers should be reported under those

, a manufacturer, has retired from
red lawyer," "retired merchant,"
and life insurance agents.

escribe them as "clerk in store," "store clerk," "bank clerk;" etc. "panic," if it is possible to describe

asons and brick masons.

bonnet manufacturer, a lace maker
maker a chocolate manufacturer.
er" for proprietors of establish-
of manufacture, as cotton manu-
etc.

ers can be reported under a single
line, as "grocer," it should be
s merchant, coal dealer, etc.
all cases where it applies.

guish between farmers and farm
hands this should be one of the points
and especially direct his attention.
“glover,” “hatter,” and “furrier”
make up, in their own establish-
ments and hats or furs which they sell.
Articles should be characterized as
“dealer,” “fur dealer.”
Al or State, whether probate, police,

CENSUS OF 1890.

ATION, AND MONTHS UNEMPLOYED.

pation.—This is a most important question closely, and in reporting occupations terms. A person's occupation

or otherwise) made, therefore, need no
their profession desire,
governor of Massachusetts
in legislature." If a
and put in the occupa-

The organization
in this country as to
the character of wor

Cooks, waiters, etc.
separately from do.

The term "how receive distinct *w* house for their own gainful occupation, w daughters assisting the

You are under no obligation to do just as he expresses it. You can, if you like, find out what he does, and act accordingly.

The inquiry as to o
infants or children too
Neither will the doing of
school be considered as
school" will be the bes
or girl, whatever the ag
contributing to the fami
mechanical or agricultura
stated.

vegetables for market or in nursery products, etc. Two or more of these occupy nurseryman and seed-grower.

Textual Data: Documents and Descriptions

seconds ago

I Ⓜ A ◊ 田 ☰ ▾ ▷ ▶ ▷ ▷

	C
s, etc.	Occupation
pers	Newsboys
	Launderers and laundresses -- not in laundry
	Theater ushers
	Midwives
	Waiters
	Chambermaids
pers	Cooks
	Janitors and sextons
	Coachmen and footmen
	Nurse maids
	Other servants
	Ladies' maids, valets, etc.
	Porters, domestic and professional service
	Teachers -- school
	Butlers
's	Charwomen and cleaners
	Bell boys, chore boys, etc.
	Boarding and lodging house keepers
	Housekeepers and stewards
	Theater ushers
	Laborers -- domestic and professional service
	Farm laborers, working out
	Greenhouse laborers



Bronx
5021
9
4124 De Reimer Avenue



Bronx
2439
53
288 East 169 Street



Bronx
5021
13
4130 De Reimer Avenue



Bronx
5021
14
4132 De Reimer Avenue



Bronx
5021



Bronx
5021



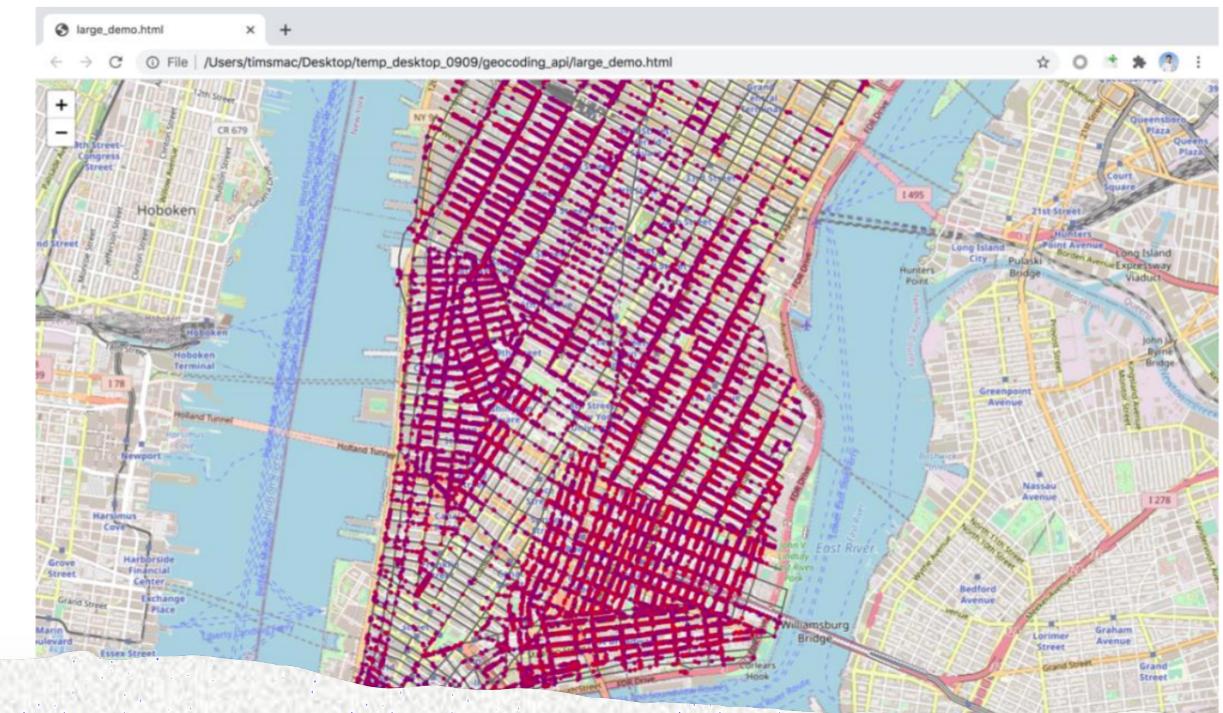
Bronx
5021
33

Visual Data: Street Views and Tax Photos

Geocoding information (E 28 street as example)

	street_name	odd_on	avg_direction	building_num_range	start_end_coordinates	segment_length	segment_direction	building
0	e 28th street	left	120.701027	[1, 20]	[[40.74496, -73.987408], [40.744288, -73.98592]]	146.161254	120.701027	
1	e 28th street	left	118.196871	[21, 50]	[[40.744288, -73.98592], [40.743616, -73.984272]]	157.932656	118.196871	
2	e 28th street	left	116.688941	[51, 129]	[[40.743616, -73.984272], [40.742992, -73.98264]]	154.277624	116.688941	
3	e 28th street	left	118.813089	[130, 166]	[[40.742992, -73.98264], [40.742336, -73.981072]]	151.149160	118.813089	
4	e 28th street	left	119.630543	[167, 249]	[[40.742336, -73.981072], [40.74136, -73.978816]]	219.215487	119.630543	
5	e 28th street	left	118.909944	[250, 316]	[[40.74136, -73.978816], [40.740368, -73.976432]]	128.357902	119.318798	
6	e 28th street	left	118.909944	[317, 433]	[[40.740368, -73.976432], [40.739802, -73.9751...]]	229.525402	118.681300	

Spatial Visualization of the geocoded addresses



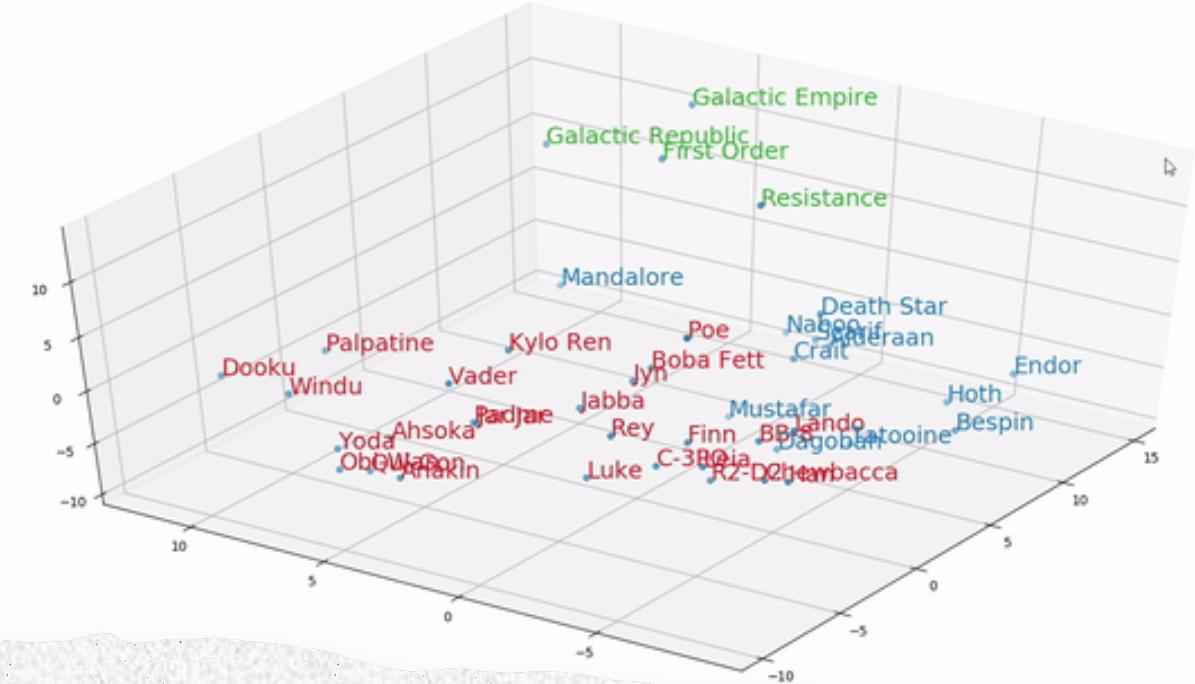
Spatial Data:
Data with Addresses and Maps

Special Data Transformation



Are you going to be on campus
anytime tomorrow? Let me
know if you want to have a
meeting tomorrow afternoon,
instead of Friday afternoon.

1	afternoon
1	are
1	be
0	best
0	go
0	gone
0	monday
1	to
:	
1	tomorrow
0	tuesday
1	you



Turning Texts into Numerical/Categorical Data

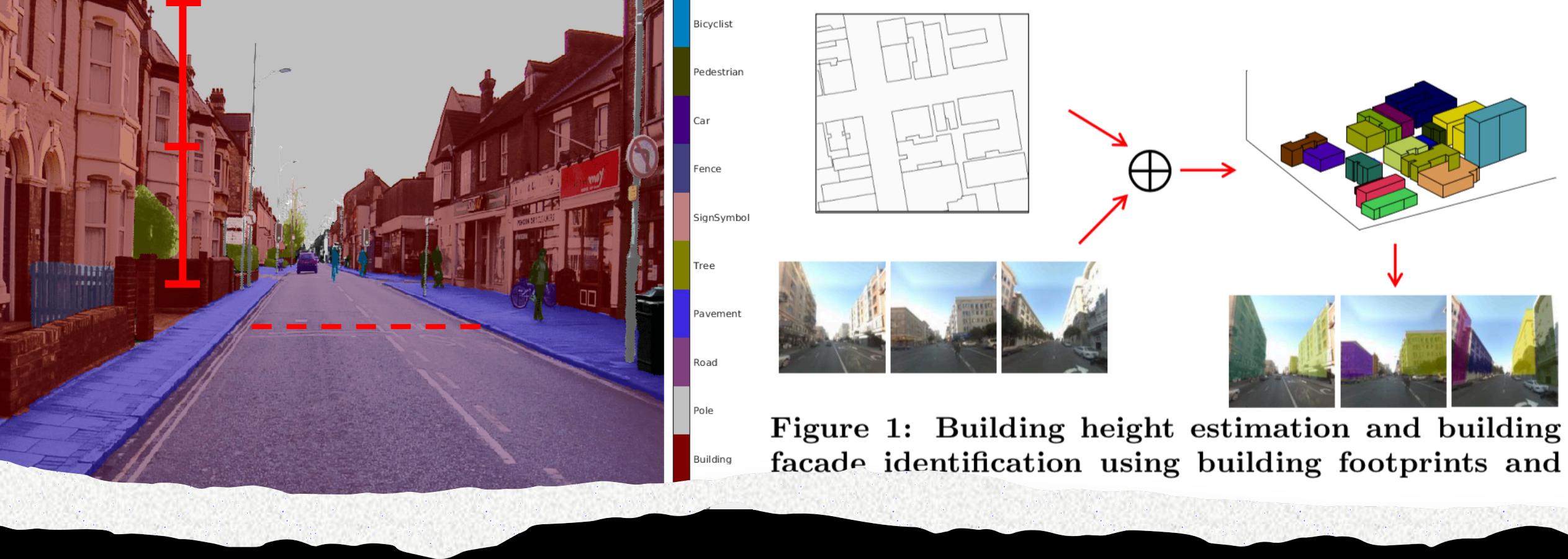
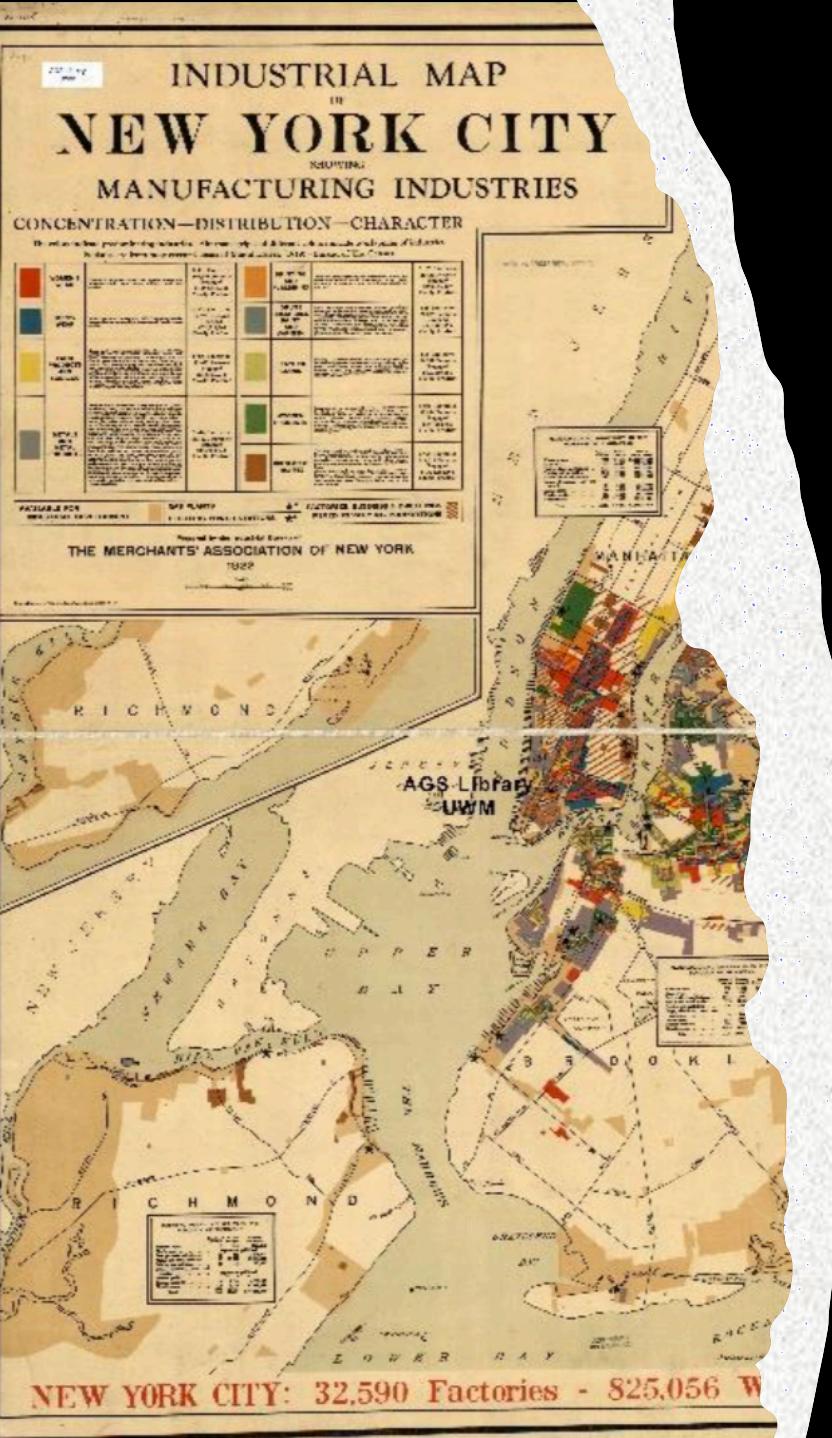
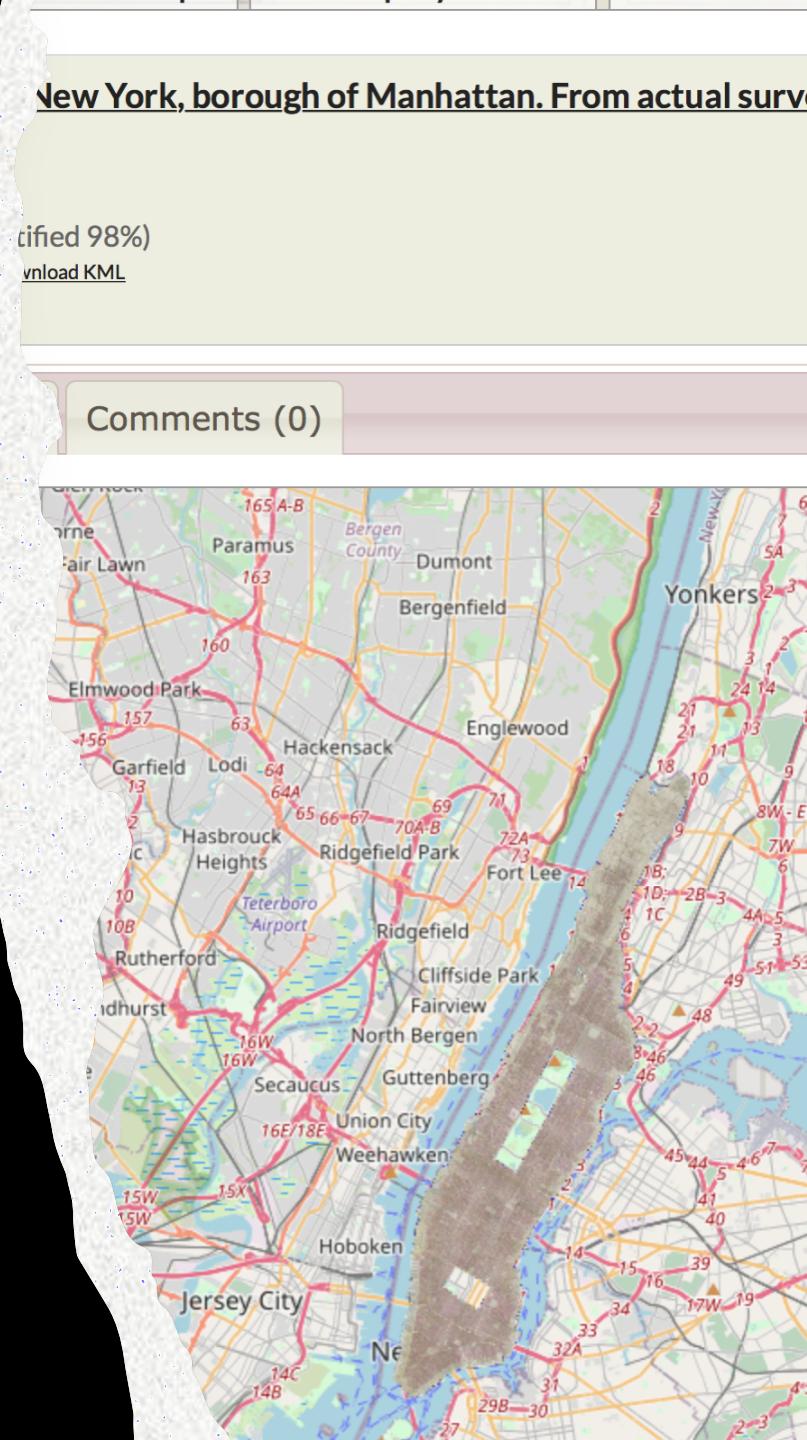


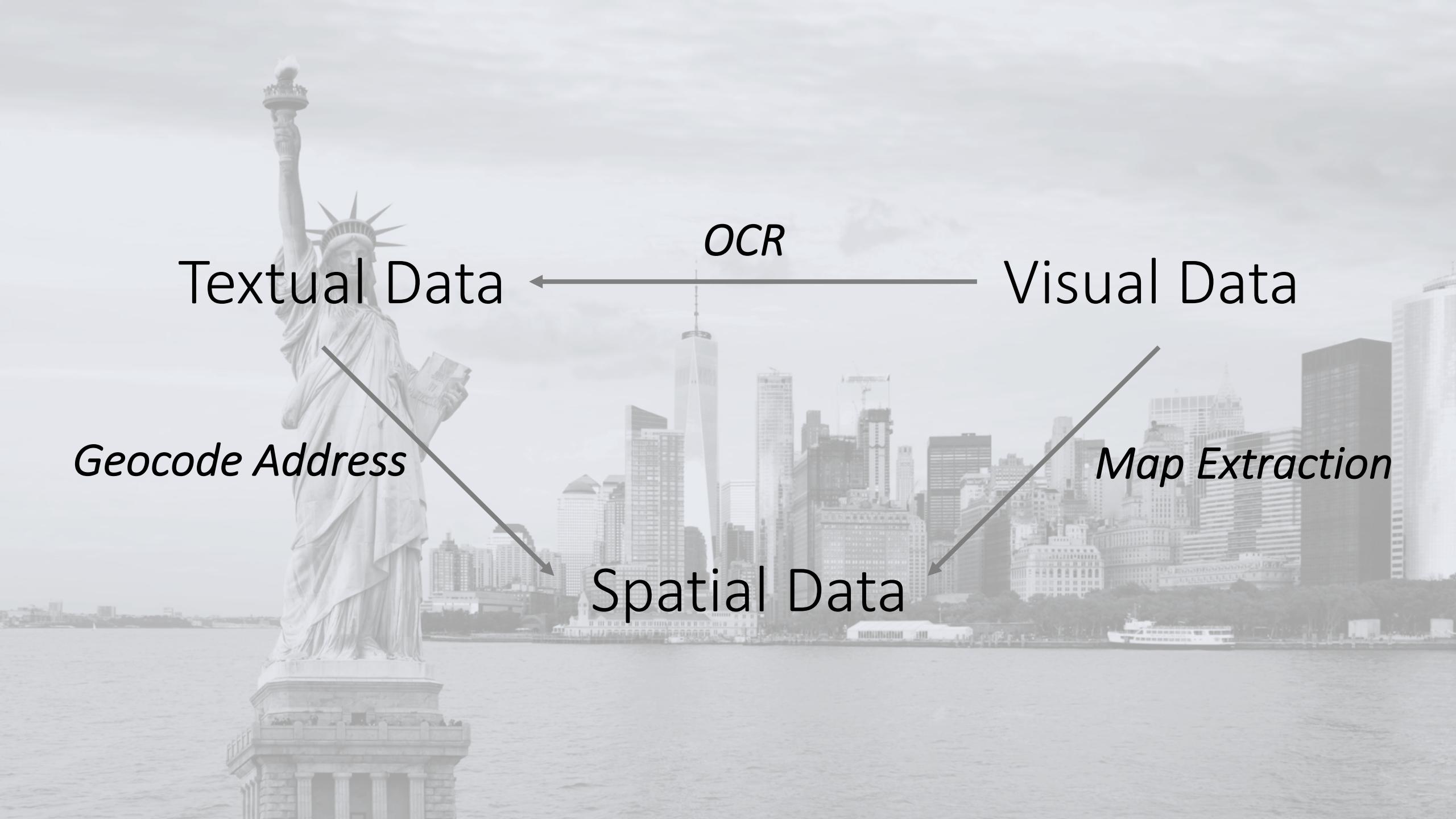
Figure 1: Building height estimation and building facade identification using building footprints and

Turning Images into Numerical Data



Turning Maps into Categorical Data





Textual Data

OCR

Visual Data

Geocode Address

Spatial Data

Map Extraction

 Facilities and Program Sites 

 FILTERS  ABOUT

 17,817 of 30,930 records 

 PROPERTY TYPES

Facility Category 

All 

Education, Child Welfare, and Youth 

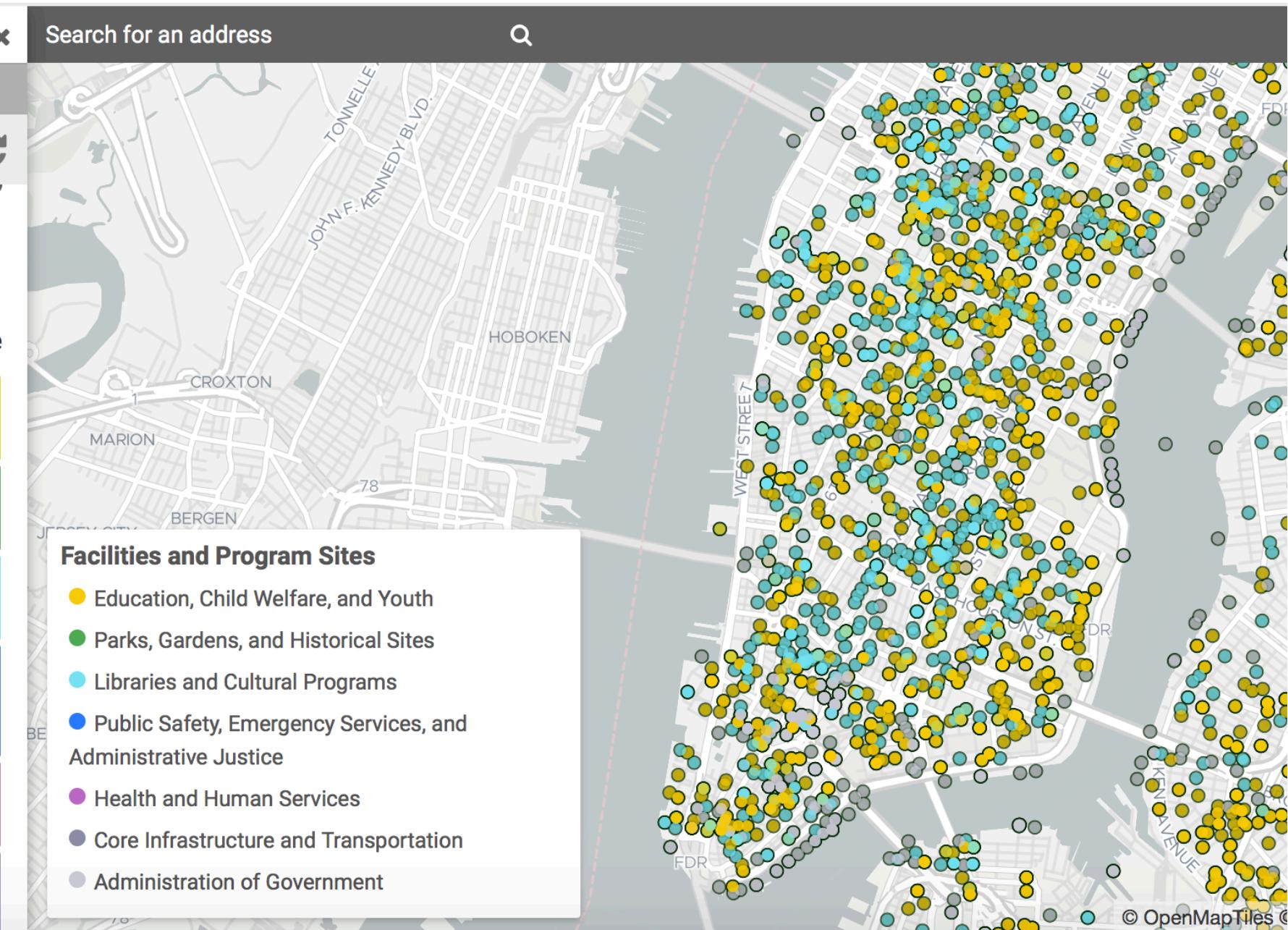
Parks, Gardens, and Historical Sites 

Libraries and Cultural Programs 

Public Safety, Emergency Services, and Administration of Justice 

Health and Human Services 

Core Infrastructure and Transportation 



OCR -> Preprocessing + Recognition + Postprocessing

Borough of Manhattan.			
Food and Drugs			
166557	Liberty Pension Co Inc.	Render Edib.	
166558	Darmas & Psillas	Manufacture	
166559	Joseph Vogt		
166560	Cullen & Coliani	Carbonate	
166761	United Cigar Store Co	"	
166562	Louis K. Higg Co Inc.	"	
166563	Loft Inc.		
166564	Gem Fountain Corp.	"	
166565	Henry Ruge	"	
166566	Anna M. Dagenais	"	
166567	William Doscher	"	
166568	Dominick Freda	Sell oysters	are obtained
		Preserve & Prep.	approved by the Board.
166569	Sapeta Bros.	Mfg.	
166570	Josef Gorsky	& Maintain a Smoke Hou	
166571	A.Cappella Provision Co	Prep. & Pres.	
166572	William Plath	Sausage & Mai	
166573	Liberty Provision Co	"	
166574	Novara & Longo	Mfg. Sausag	
190840	C.C. & S. Restaurant Inc.	Con. a Rest	
190841	Boulevard Restaurant	"	
190842	Charles Mateon	"	
190843	Emil Ascherman	"	
190844	Frank Busto	"	
190845	Arnold Schleifer	"	
190846	Samuel Shapiro	"	
190847	Nicholas Goerg	"	
190848	Broadway, Rest. Co. Inc.	"	
190849	Klein and Stein	"	
190850	Benjamin Lopatkin	"	
190851	Shefrin Bros.	"	
190852	Joseph Picorino	"	
190853	Decourty & Mennier	"	
190854	Manos & Philindras	"	
190855	Comins & Heiligman	"	
190856	Louis Cassas	"	
190857	Henry W. Thoelen	"	
190858	Cohen & Feinian	"	
190859	Louis Lindenbaum	"	
190960	Kohl Brosl	"	
190861	Edmond Sens	"	
190862	Benjamin Weise	"	
190863	Ciprios Bros.	"	
190864	Gus Scarmalis	"	
190865	Samuel Wilfand	"	
190866	Frank Menke	"	
190867	Miller & Knihinitzky	"	
190888	Michael E. Sexton	"	
190869	M. & S. Lunch	"	

BOROUGH OF BROOKLYN

Food and Drugs			
169599	Delaware Cheese Co.	Sell milk & milk Prods.	383 So. 1st St.
4160	Stella Stanich	Keep 1 cow	1637 Ocean Ave.
4161	Frank Cammerer	"	1752 80th St.
169565	William Schebler	Break out eggs for tanning purposes	1100 Grand St.
169566	Fred. Dickhuth	Manf. ice cream	27 Alabama Ave.
169567	Michal Beltran	"	51 Arlington Ave.
169568	Joseph M. Gottlieb	"	433 Bedford Ave.
169569	John Schmidt	"	109 Bridge St.
169570	Alfred During	"	57 Broadway
169571	Argyros & Co.	"	1356 Broadway
169572	Harms & Kroll	"	129 Ditmars Ave.
169573	Charles O. Selka	"	8415 18th Ave.
169574	Louis Strauss	"	5404 5th Ave.
169575	Kope & Windeler	"	5821 5th Ave.
169576	August Reiner	"	7308 "
169577	John D. Jark	"	7504 "
169578	Meyer & Hashagen	"	7402 15th Ave.
169579	Henry Horwege	"	877 Flatbush Ave.
169580	Charles Heins	"	938 "
169581	Lorenzen Bros.	"	6906 Ft. Hamilton P'kway
169582	Matthias Meyn	"	2807 "
169583	Frederick Hollwedel	"	3356 Fulton St.
169584	Herman N. Rickett	"	228 Kings Highway
169585	Charles Cordts	"	751 Manhattan Ave.
169586	Chas. H. Intermann	"	135 Meserole St.
169587	Pietro Cordaro	"	1080 Nostrand Ave.
169588	John H. Seebeck	"	230 A St. Nicholas Ave.
169589	August H. Siedentopf	"	239 7th Ave.
169590	Iatridis & Lazar	"	412 Schenck Ave.
169591	Harry Levy	"	275 Wyckoff Ave.
169592	Ferdinand Richter	"	85 Montrose Ave.
169593	Acme Bottling Co. Inc.	Mfg Carbonated water	Prepare and preserve
169594	Peter Olsen	"	meats & manf. sausages
192723	Longeren & Swanson	Cond. a restaurant	4206 8th Ave.
192724	Accles & Cahill	"	477 Atlantic Ave.
192725	Barnet Abramowitz	"	1202 Bedford Ave.
192726	John Schmitt	"	119 Belmont Ave.
192727	Arthur Johnson	"	1398 Bergen St.
192728	Carl Horstmann	"	1408 "
192729	Ethel Moreira	"	134 Buffalo Ave.
192730	Julius Wolff	"	110 Broadway
192731	Mick Koss & Frank Karlam	"	26 Church Ave.
192732	Hugh Bond	"	624 Classon Ave.
192733	William Strauss	"	Ft. of Conover St.
192734	Walter McCorry	"	241 Court St.
192735	Emil Probst	"	S.W. Cor. E. H.Y. Ave.
192736	Abraham Pifko	"	* Van Sinderen Ave.
192737	Oscar Ostenen	"	8440 18th Ave.
192738	Joe Bayer & Prushansky	"	4912 5th Ave.
192739	Haese & Kimm	"	5121 "
192740	Sidney Goldman	"	5311 "
192741	Lung & Gum	"	7130 "
192742	Christian Poulsen	"	8522 "
192743	Ex-Joseph O'Brien &	"	873 Flatbush Ave.
192744	Andrew O'Brien	"	514 Flushing Ave.
192745	D. & B. Incorporated	"	1120 "
192746	Henry W. Higham	"	577 Fulton St.
192747	Kurt Weingartner	"	2356 "
192748	Meyer Kalman	"	2572 "
			18 Graham Ave.

BOROUGH OF MANHATTAN

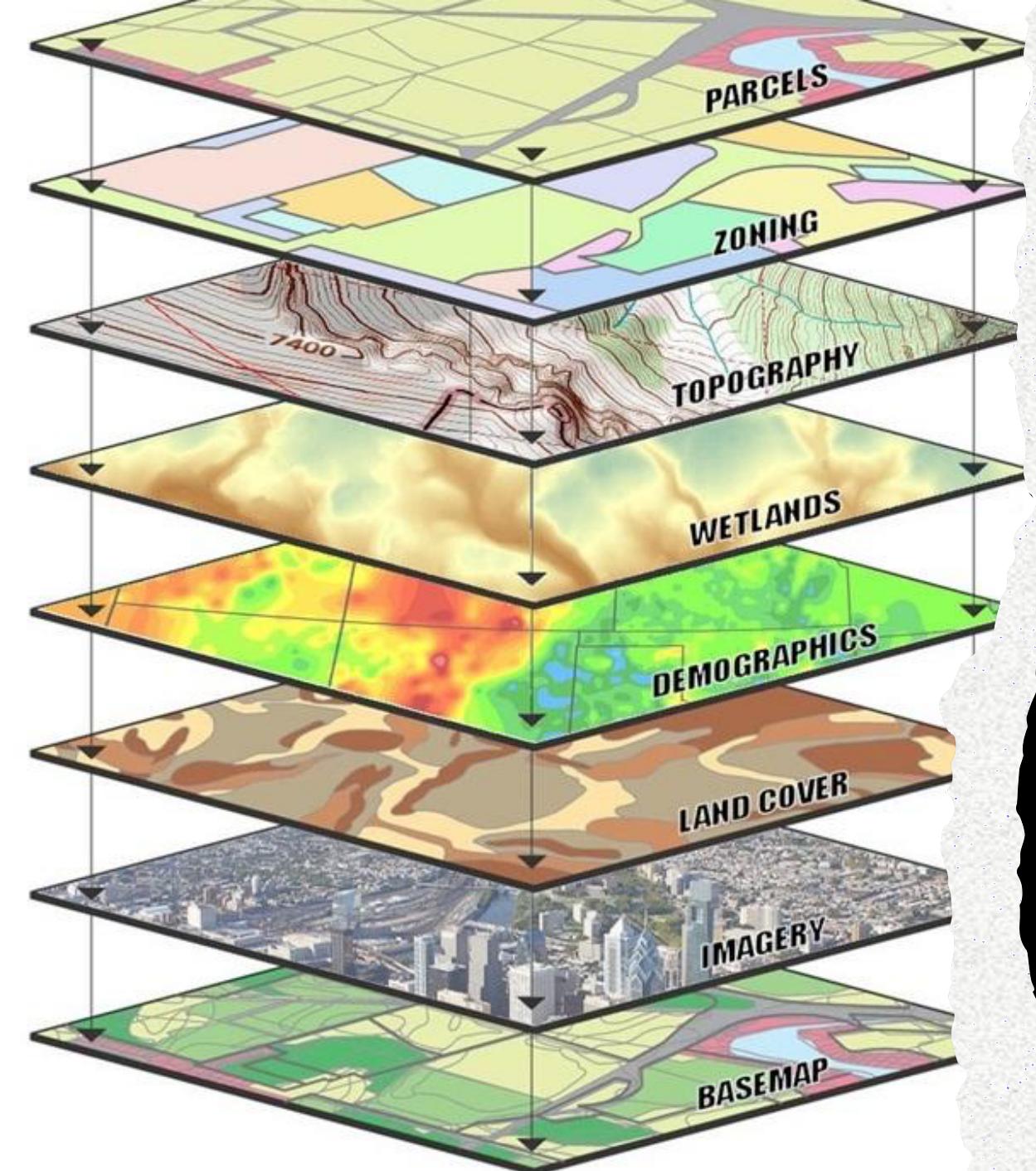
BOROUGH OF MANHATTAN			
2907	number	Bernlohr & Katz	name Bell Milk nature
1931	number	Hechman & Drach	name Bell Milk nature
576	number	Theobald Schreder	name
5197	number	Louis Spitzer	name
4988	number	Edw. Tannenbaum	name
2917	number	Max Moskowitz	name
1231	number	Bridget Nefforn	name
9056	number	Bush & Somerville	name
3885	number	Herman Alperstein	name
3886	number	Joseph Feinberg	name
4687	number	Daniel Reeves, Inc.	name
1827	number	Goldberg & Reich	name
414	number	Lee, Shifron	name
524	number	Fred Nusselman	name
2002	number	Joe, Waksigian	name
2972	number	Atlas Economy Stores, Inc.	name
1820	number	Henry Wisch	name
1400	number	Paul Struck	name
4679	number	Harry Fleishman	name
1959	number	Chas Cronquist	name
2880	number	Fettus Penn	name
3905	number	Joseph Marin	name
1978	number	Daniel Reeves, Inc.	name
1987	number	John Ferre	name
422	number	Chas. Gerstner	name
474	number	Frank Mutterer	name
1431	number	Papke & Brandt	name
1443	number	Nikolas Pensch	name
491	number	Jacob Steinberg	name
504	number	Henry Weintraub	name
512	number	Hyman Mecklowitz	name
487	number	Wagner & Gross	name
1448	number	David Fichtbaum	name
2915	number	Mill Mans	name
5793	number	Groka & Groka	name
2468	number	Max, L. Kerschen	name
3939	number	Hollman Bros.	name
3943	number	Wittmann & Igors	name
446	number	Krich Hoffman	name
2943	number	Atlas Economy Stores	name
552	number	Christian Botscheller	name
516	number	Isak Kawut	name
4117	number	Barnett Schneller	name
4810	number	Maple Lodge Farms Co., Inc.	name
4823	number	Pasquale Femese	name
4824	number	Joe, Schuman	name
3558	number	Sacks & Blumenthal	name
3570	number	Andrew Davey Inc.	name
4153	number	Max Bernstein	name
5832	number	Commodore Bakery Inc.	name
1257	number	Theo. Hasselbach	name
1838	number	Max Gold	name
2045	number	Chas. Gottlein	name
276	number	John Stechmann	name
1847	number	Podhorzer & Spitzer	name

370 - 3rd Avenue address
 61 - 2nd Ave. address
 624 - 2nd Ave. address
 2195 - 3rd Ave. address
 1857 - 3rd Ave. address
 1101 address
 1044 - 1st Ave. address
 200 - 1st Ave. address
 1708 - 3rd Ave. address
 1665 - 3rd Ave. address
 1934 address
 527 W. 133rd St. address
 1824 - 2nd Ave. address
 176 - 3rd Avenue address
 397 - 3rd Ave. address
 432 - 3rd Ave. address
 551 - 3rd Ave. address
 712 - 3rd Ave. address
 749 address
 215 address
 905 address
 1052 address
 1143 address
 1240 address
 1332 address
 1408 address
 1591 address
 1660 address
 1909 address
 31 Washington St. address
 1 West 125th St. address
 501 West 125th St. address
 227 West 145th St. address
 226 West 146th St. address
 550 West 207th St. address
 565 West 207th St. address
 565 West 207th St. address
 565 West 207th St. address
 37 Grand St. address
 287 Henry St. address
 197 Hester St. address
 343 Hudson St. address
 506 Hudson St. address

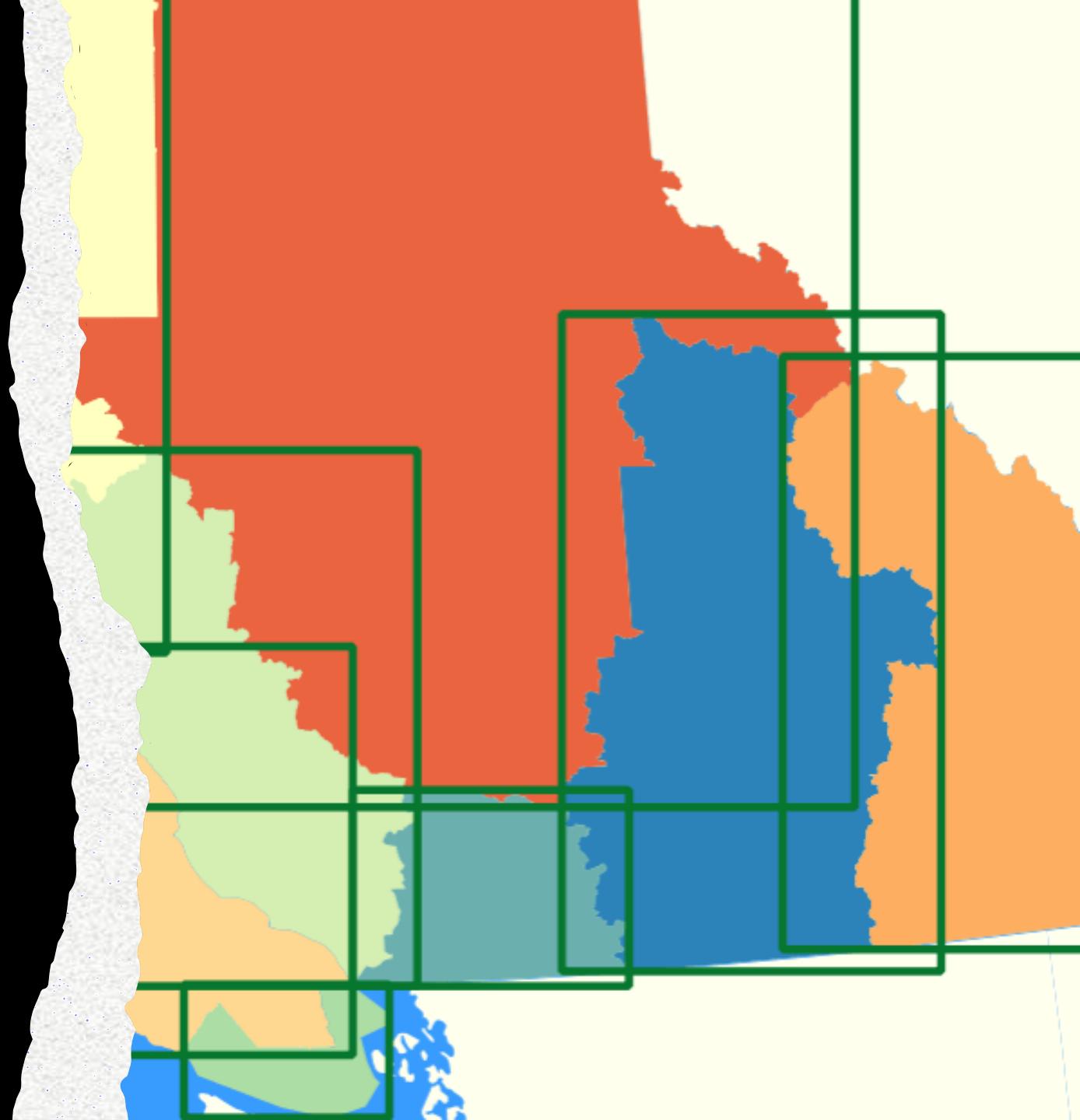


After data transformation,
special data still require *special handling*

Changes how
we *store*:
GIS data layers

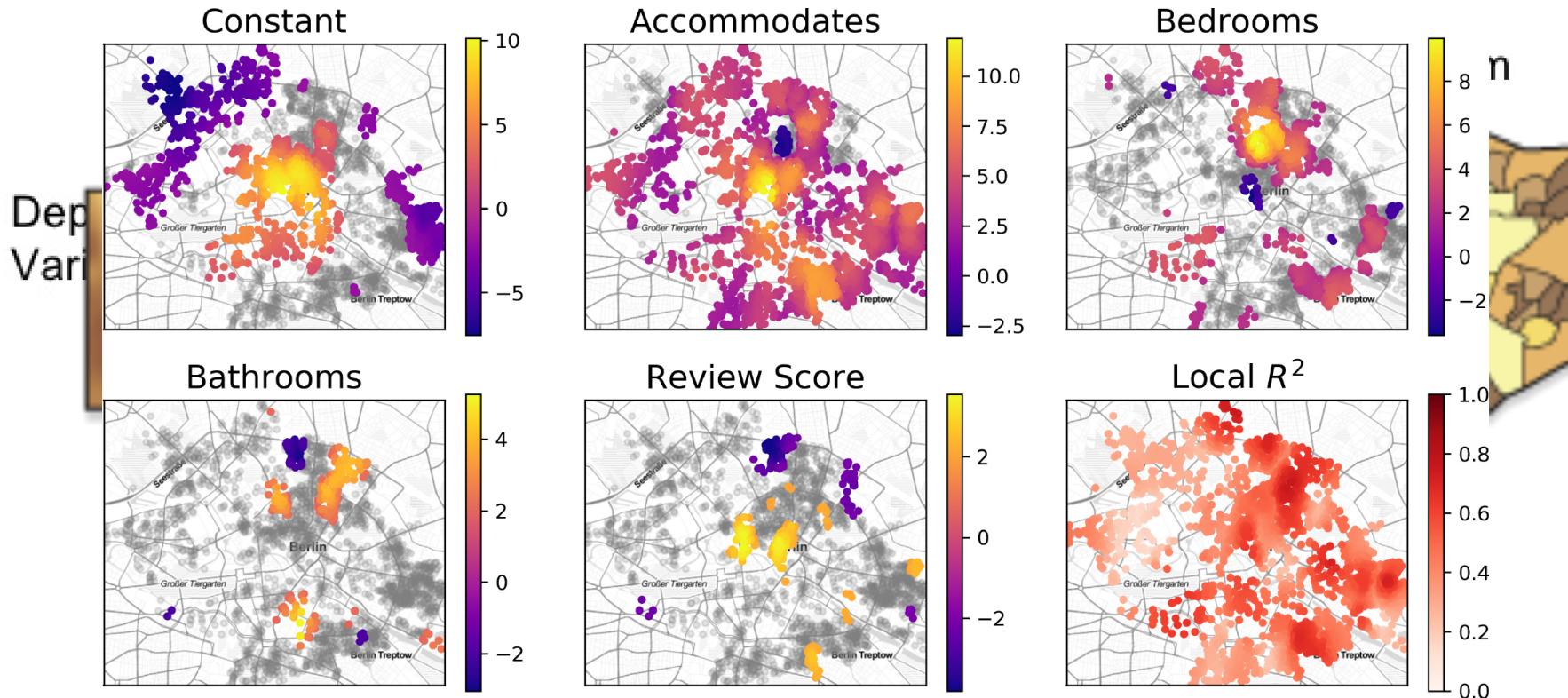


Changes how
we *search*:
Spatial index
bounding boxes

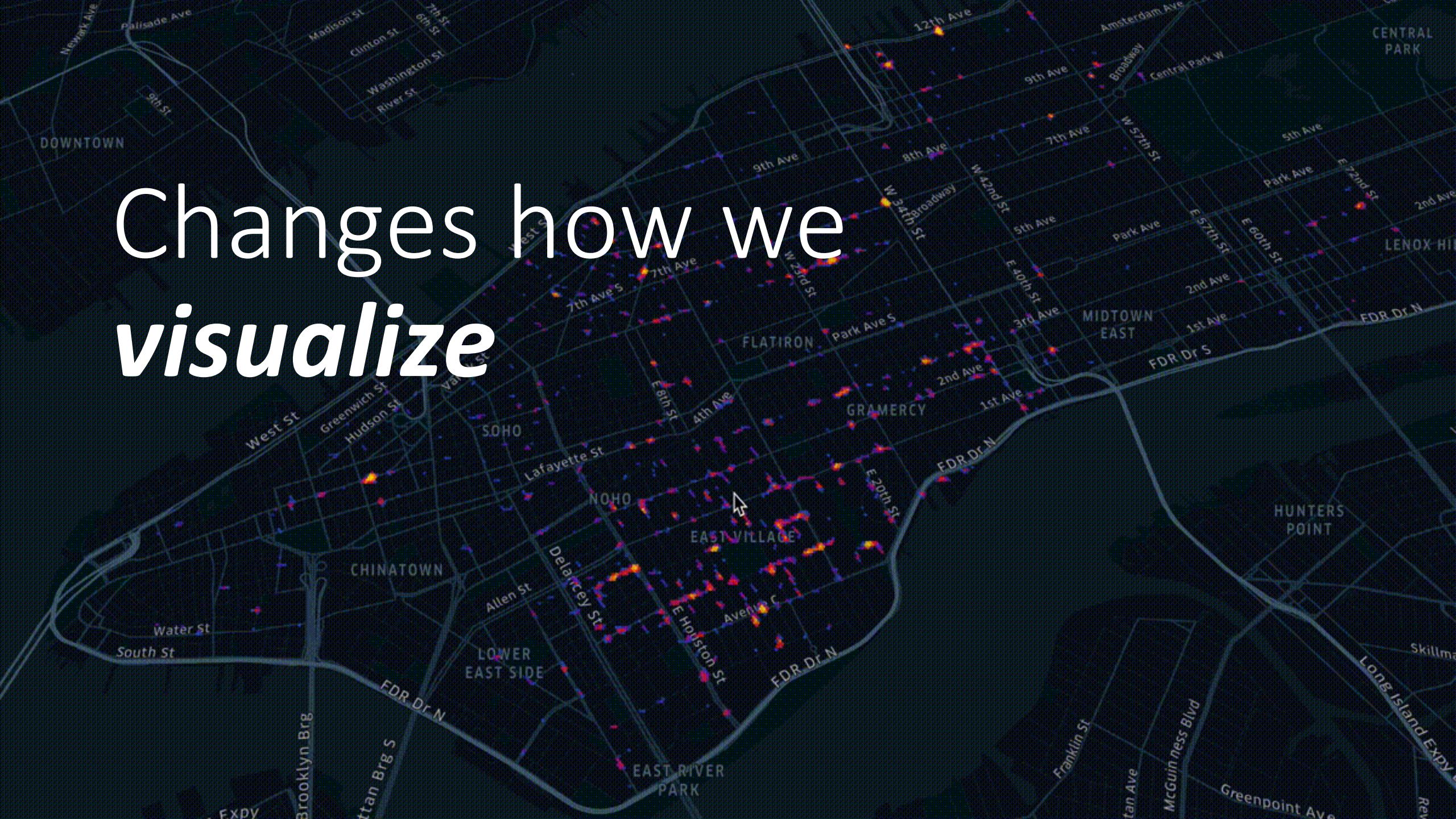


Changes how we *analyze*:

Geographically weighted regression



Changes how we *visualize*



Summary

- Special data can be transformed to integrate better with traditional data.
- Creative conversions among special data open up new opportunities.
- Special data requires special ways to store, search, analyze and visualize.