

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 site of the former Twin Towers. Other recognizable buildings like the Chrysler Building and various skyscrapers are visible across the cityscape. In the distance, a bridge spans a body of water, and the horizon shows the silhouette of the Jersey City and Hoboken waterfronts.

Week 2 Special Data

June 2021

Theory of Viz

W1 - What is Visualization and Why Visualize

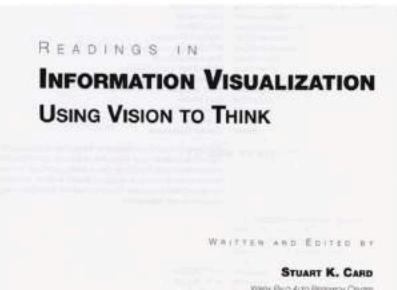
Recommended Reads

- [The Purpose of Visualization \(Stanford CS 448B Slides\)](#) Concise, illustrated slide that covers the definition and rationale of visualization. Recommend to read pdf page 1-26.

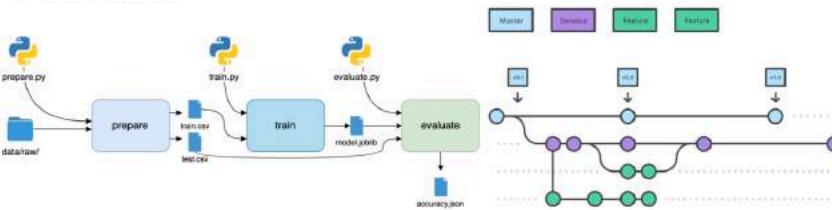


Visual Encoding

- [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

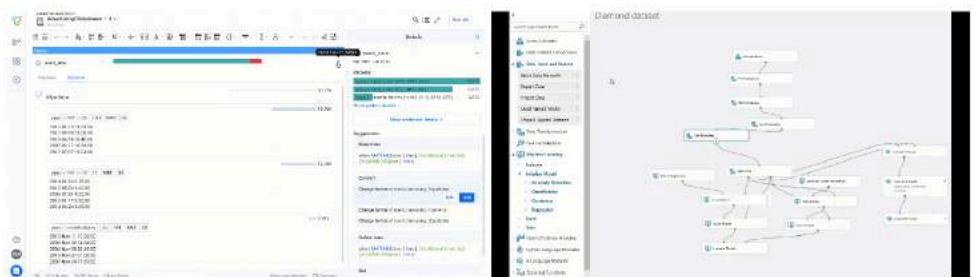
- Data profiling
- Smart fuzzy matching
- Program synthesis

Master Table

ID	First Name	Last Name	Address 1	Address 2	Phone	Nat ID	State	ZIP	Birthday
438	John	Doe	123 Main St		(555) 123-4567	1234567890	NY	100-12345	1985-01-01
2345	Jane	Doe	123 Main St		(555) 123-4567	1234567890	NY	100-12345	1985-01-01
1234	John	Doe	123 Main St		(555) 123-4567	1234567890	NY	100-12345	1985-01-01
9876	John	Doe	123 Main St		(555) 123-4567	1234567890	NY	100-12345	1985-01-01

3. Automatic/Semi-automatic Data Cleaning

4. Visualization and GUI for the above steps



Data Type Constraints

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

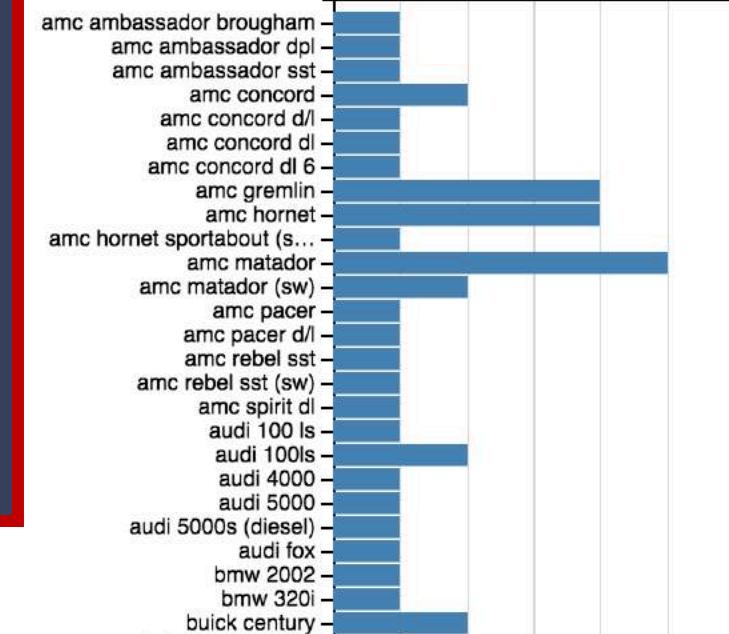
A Cylinders # COUNT

3

A Name # COUNT

COUNT

0 1 2 3 4 5 6

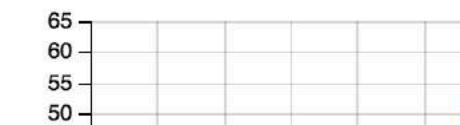


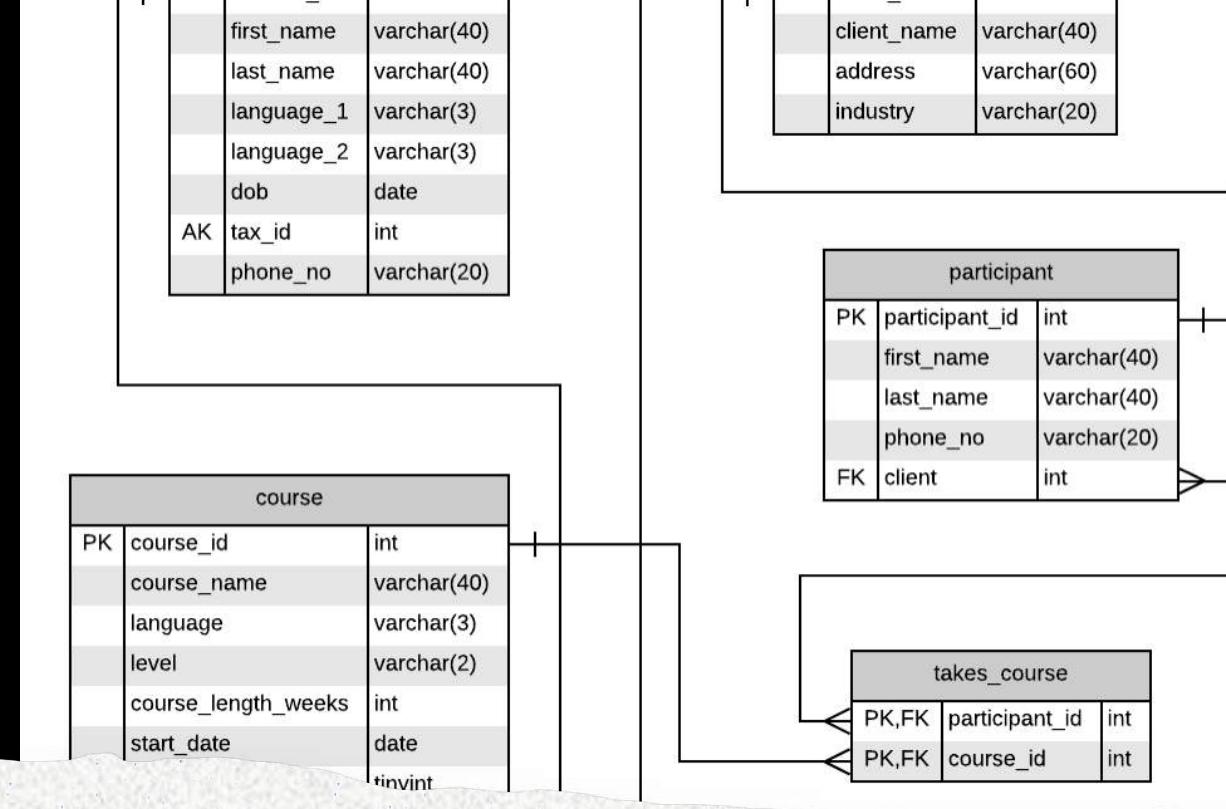
A Origin # COUNT

Origin

Europe Japan

YEAR (Year) # COUNT

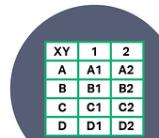




Why these data types?

Structured Data vs Unstructured Data

Can be displayed
in rows, columns and
relational databases



Numbers, dates
and strings

Estimated 20% of
enterprise data (Gartner)



20%

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)



80%

Requires more storage



Special Data

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

A collage of New York City street signs at night. In the foreground, a green street sign for "PARK AV" is visible, along with a "ONE WAY" sign pointing right. Behind it, another "ONE WAY" sign points left, and a green sign for "E 33 ST" is partially visible. The background shows the Empire State Building and other city buildings with illuminated windows.

What *special data*
do we have?

STATISTICS OF OCCUPATION

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

masons 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
as this should be one of the points
d especially direct his attention.
“glover,” “hatter,” and “furrier”
make up, in their own establish-
es and hats or furs which they sell.
ticles should be characterized as
dealer,” “fur dealer.”

CENSUS OF 1890.

ATION, AND MONTHS UNEMPLOYED.

pation.—This is a most important
ions closely, and in reporting occu-
ning terms. A person's occupation

or otherwise) may therefore, need no their profession designating governor of Massachusetts in legislature." If a man put in the occupa-

The organization
in this country as t
the character of wor

Cooks, waiters, etc.,
separately from do.

The term "home" 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. If or more of these occupy nurseryman and seed-grower.

Textual Data: Documents and Descriptions

seconds ago

I Ⓜ A ◊ 田 [?] ▾ ▾ ▾ ▾ ▾

	C
	Occupation
s, etc.	Newsboys
pers	Launderers and laundresses -- not in laundry
	Theater ushers
	Midwives
	Waiters
	Chambermaids
	Cooks
pers	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

Bronx

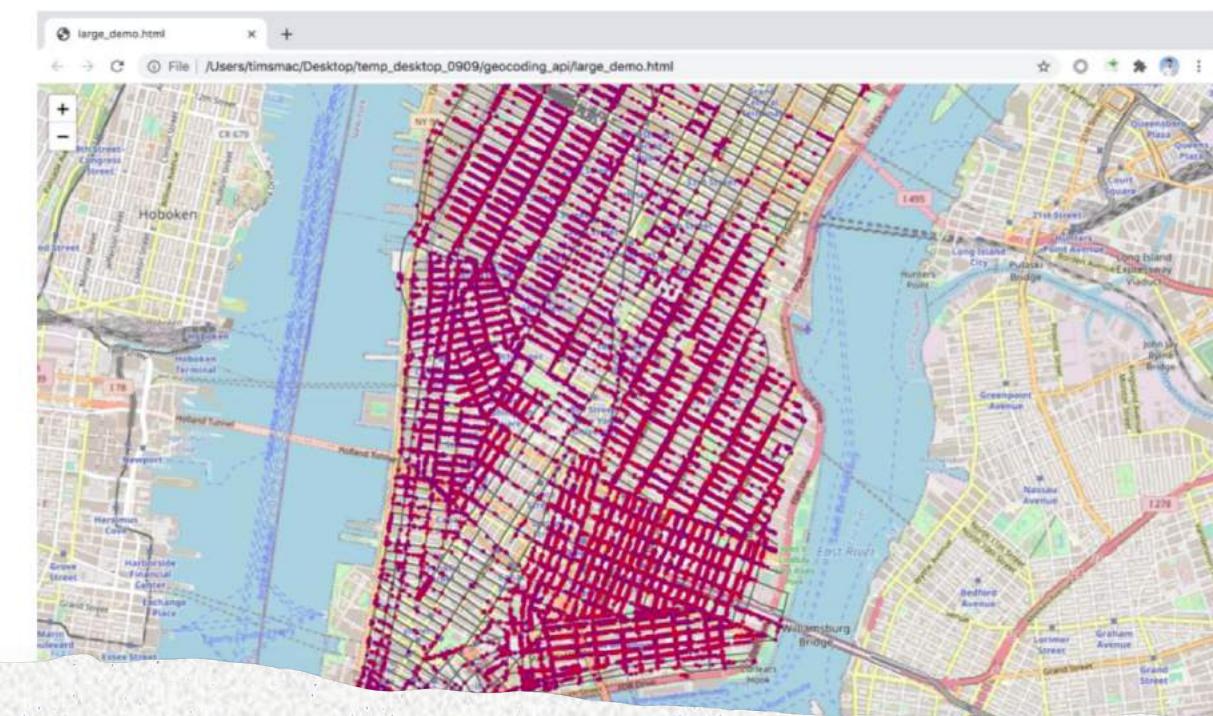
Bronx

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



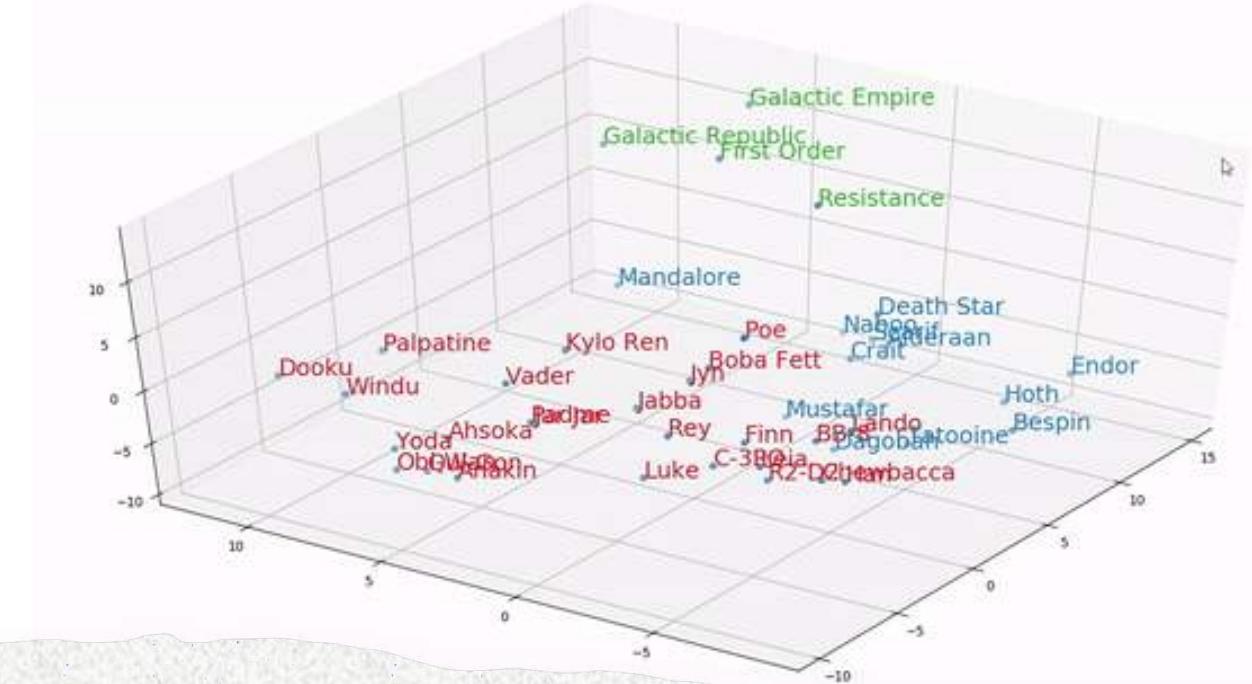
Spatial Data:
Data with Addresses and Maps

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

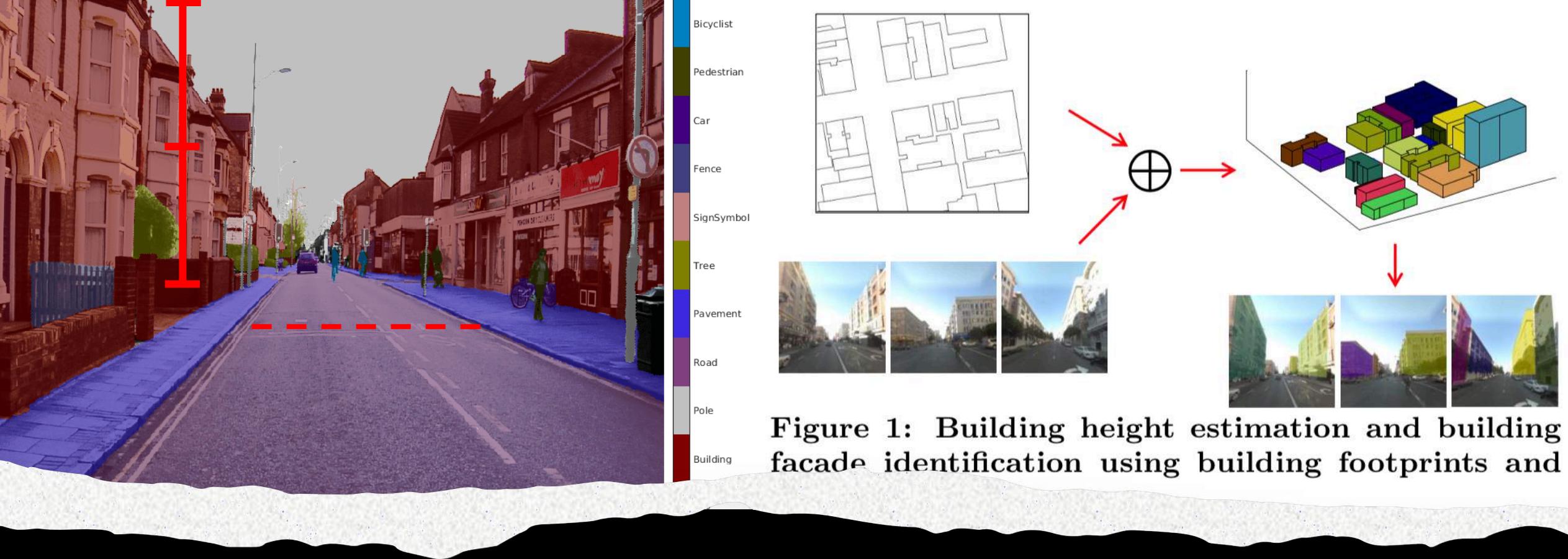
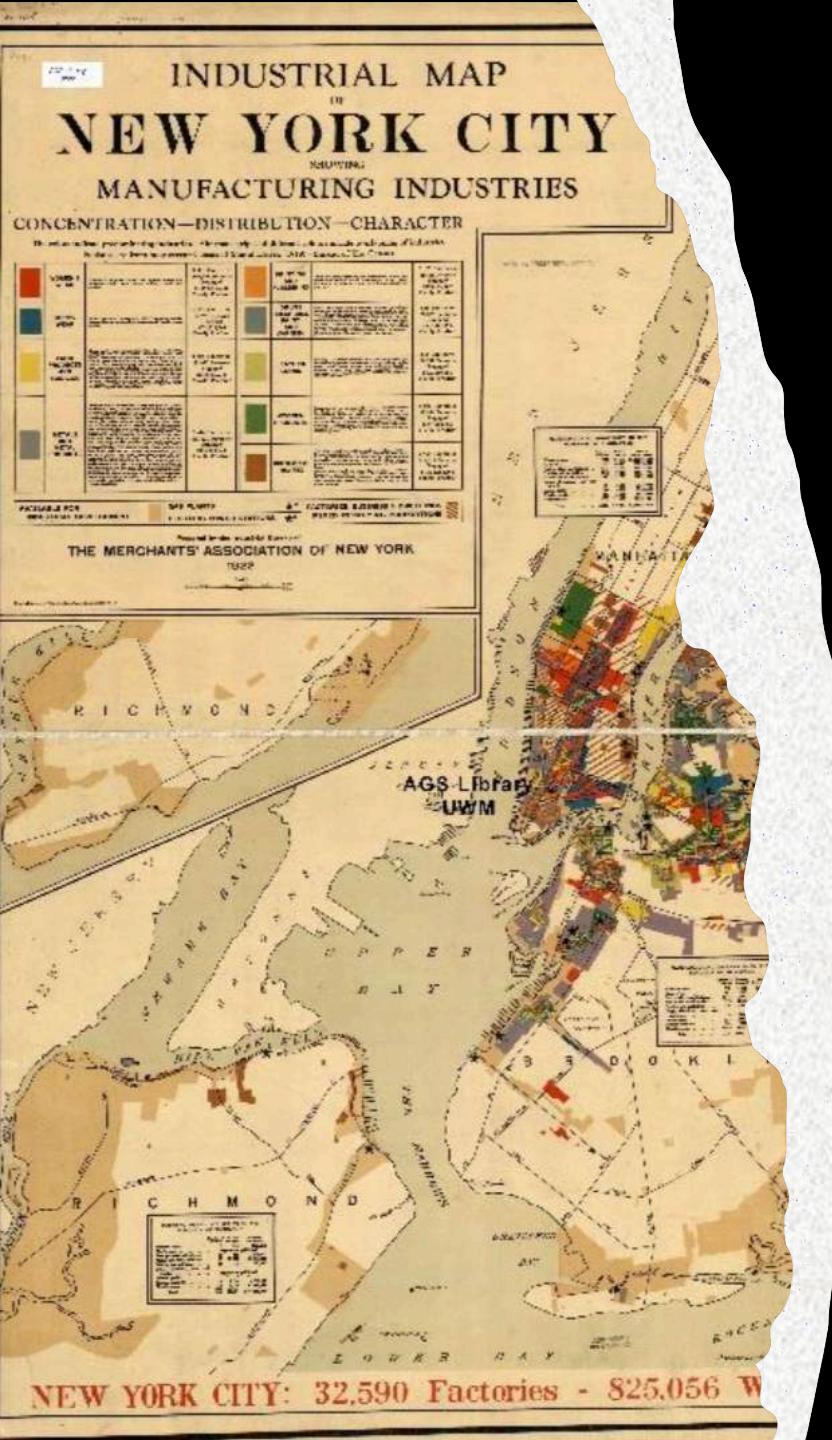
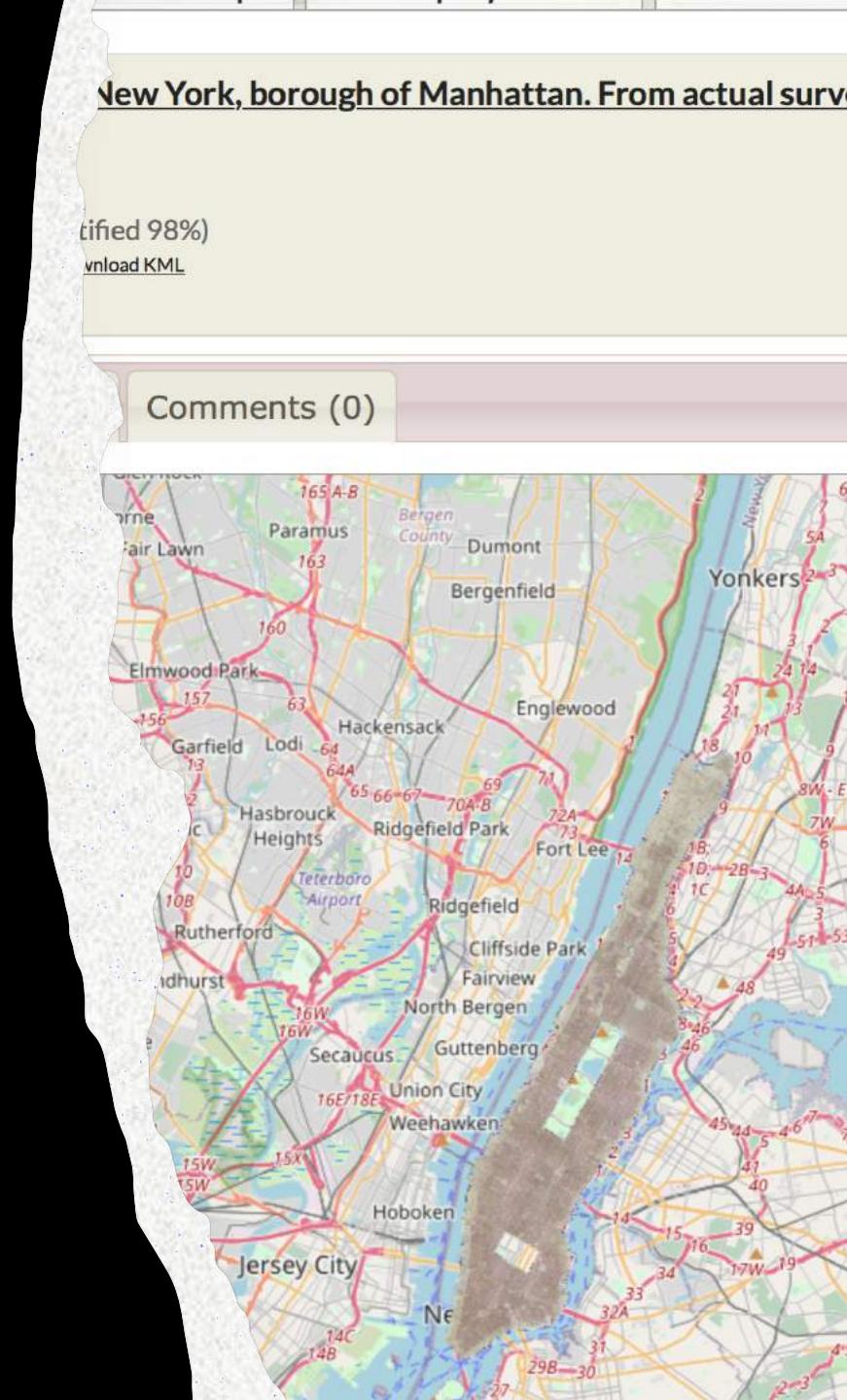


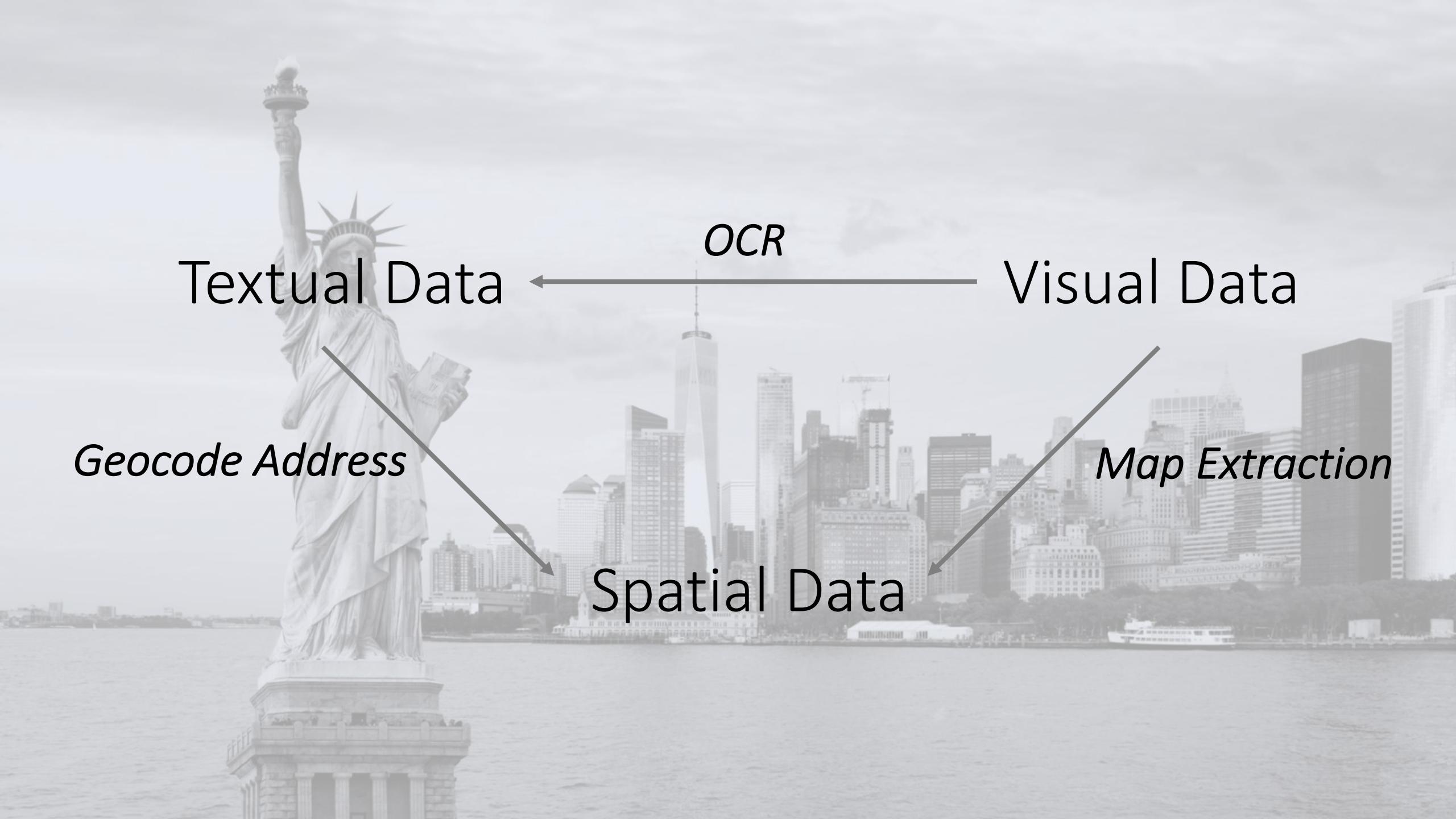
Figure 1: Building height estimation and building facade identification using building footprints and

Turning Images into Numerical Data



Turning Maps into Categorical Data





Textual Data

OCR

Visual Data

Geocode Address

Spatial Data

Map Extraction

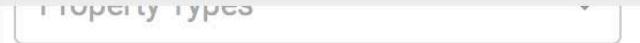
**Facilities and Program Sites**

 FILTERS

 ABOUT

 17,817 of 30,930 records







All [Expand](#) | [Collapse](#)

Education, Child Welfare, and Youth  

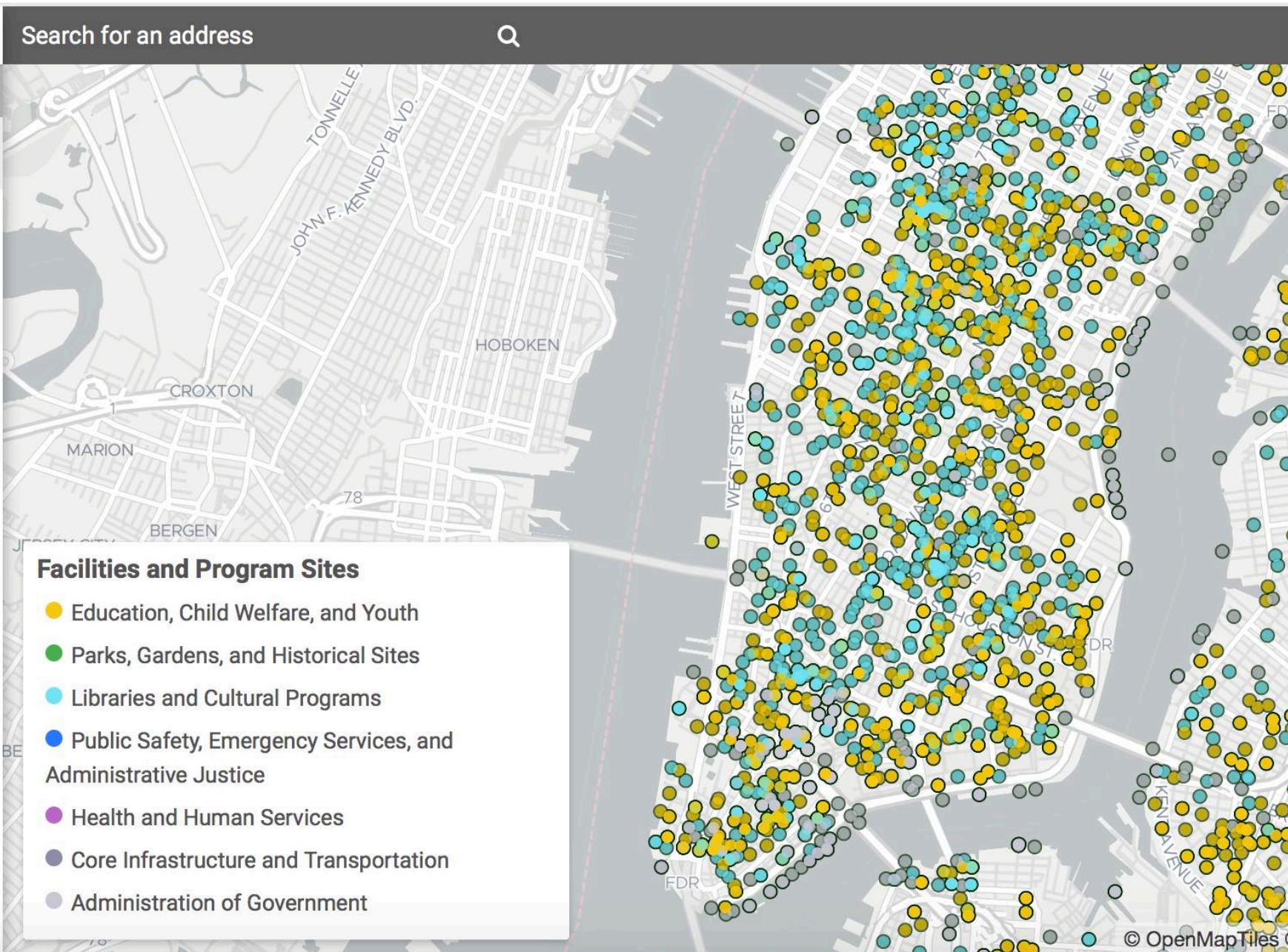
Parks, Gardens, and Historical Sites  

Libraries and Cultural Programs  

Public Safety, Emergency Services, and Administration of Justice  

Health and Human Services  

Core Infrastructure and Transportation  



OCR -> Preprocessing + Recognition + Postprocessing

Food and Drugs			
Borough of Manhattan			
Food and Drugs			
166557	Liberty Pension Co Inc.	Render Edib.	
166558	Darmas & Peillas	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
		approved ad-	
166569	Sapeta Bros.	Preserve&Prep.	Meat, Mf-
		& Maintain a Smoke Hou-	
166570	Josef Gorsky	Preserve & Prep.	Meat
166571	A.Cappella Provision Co	Prep.&Pres.	
166572	William Plath	Sausage&Meat	
166573	Liberty Provision Co	"	
166574	Movara & Longo	Mfg. Sausage	
190840	C.O.&S. Restaurant Inc.	Con. a Rest	
190841	Boulevard Restaurant	"	
190842	Charles Mateon	"	
190843	Emil Ascherman	"	
190844	Frank Rusto	"	
190845	Arnold Schleifer	"	
190846	Samuel Shapiro	"	
190847	Nicholas Goerg	"	
190848	Broadway, Rest. Co. Inc.	"	
190849	Klein and Stein	"	
190850	Benjamin Lopatkina	"	
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	"	
190980	Kohl Brosl	"	
190881	Edmond Sens	"	
190882	Benjamin Weiss	"	
190883	Ciprius Bros.	"	
190884	Gus Scarmalis	"	
190885	Samuel Wilfand	"	
190886	Frank Menke	"	
190887	Miller & Knihinitzky	"	
190888	Michael E. Sexton	"	
190889	H. & S. Lunch	"	

BOROUGH OF BROOKLYN			
Food and Drugs			
169599	Delaware Cheese Co.	Sell milk & milk Prods.	383 So. 1st St.
4150	Stella Stanich	Keep 1 cow	1537 Ocean Ave.
4161	Frank Cammerer		1752 80th St.
169565	William Schebler	Break out eggs for	
		tanning purposes	
169566	Fred. Dickhuth	Manif. ice cream	1100 Grand St.
169567	Michal Beltran	"	27 Alabama Ave.
169568	Joseph M. Gottlieb	"	51 Arlington Ave.
169569	John Schmidt	"	433 Bedford Ave.
169570	Alfred During	"	109 Bridge St.
169571	Argyros & Co.	"	57 Broadway
169572	Harms & Kroll	"	577 Broadway
169573	Charles O. Selka	"	1356 Broadway
169574	Louis Strauss	"	129 Ditmas Ave.
169575	Kope & Windeler	"	8415 18th Ave.
169576	August Reiner	"	5404 5th Ave.
169577	John D. Jack	"	5221 5th Ave.
169578	Meyer & Hashagen	"	7308 "
169579	Henry Horwege	"	7504 "
169580	Charles Heins	"	7402 15th Ave.
169581	Lorenzen Bros.	"	877 Flatbush Ave.
169582	Matthias Meyn	"	938 "
169583	Frederick Hollwedel	"	6906 Ft. Hamilton P'kway
169584	Herman N. Rickett	"	2807 "
169585	Charles Cordts	"	3356 Fulton St.
169586	Chas. H. Intermann	"	228 Kings Highway
169587	Pietro Cordero	"	751 Manhattan Ave.
169588	John H. Seeseck	"	135 Meserole St.
169589	August H. Siedentopf	"	1080 Nostrand Ave.
169590	Iatridis & Lazar	"	220 A St. Nicholas Ave.
169591	Harry Levy	"	297 7th Ave.
169592	Ferdinand Richter	"	412 Schenck Ave.
169593	Aome Bottling Co. Inc.	Mfg Carbonated water	275 Wyckoff Ave.
169594	Peter Olsen	Prepare and preserve	85 Montrose Ave.
		meats & manf. sausages	1080 Nostrand Ave.
192723	Longeren & Swanson	Cond. a restaurant	477 Atlantic Ave.
192724	Accles & Cahill	"	1302 Bedford Ave.
192725	Barnet Abramowitz	"	119 Belmont Ave.
192726	John Schmitt	"	1398 Bergen St.
192727	Arthur Johnson	"	1408 "
192728	Carl Horstmann	"	134 Buffalo Ave.
192729	Ethel Moreira	"	110 Broadway
192730	Julius Wolff	"	26 Church Ave.
192731	Mick Koss & Frank Karlak	"	624 Classon Ave.
192732	Hugh Bond	"	Ft. of Conover St.
192733	William Strauss	"	241 Court St.
192734	Walter McCorry	"	S.W. Cor. E. N.Y. Ave.
			* Van Sinderen Ave.
192735	Emil Probst	"	8440 18th Ave.
192736	Abraham Pifko	"	4912 5th Ave.
192737	Oscar Osten	"	5121 "
192738	Joe Bayer & Prushansky	"	5311 "
192739	Haase & Kimm	"	7130 "
192740	Sidney Goldman	"	8522 "
192741	Lung & Gum	"	873 Flatbush Ave.
192742	Christian Poulsen	"	514 Flushing Ave.
192743	Ex-Joseph O'Brien &	"	1120 "
	Andrew O'Brien	"	577 Fulton St.
192744	D. & B. Incorporated	"	2356 "
192745	Henry W. Higham	"	2572 "
192746	Kurt Weingartner	"	18 Graham Ave.
192747	Meyer Kalman	"	

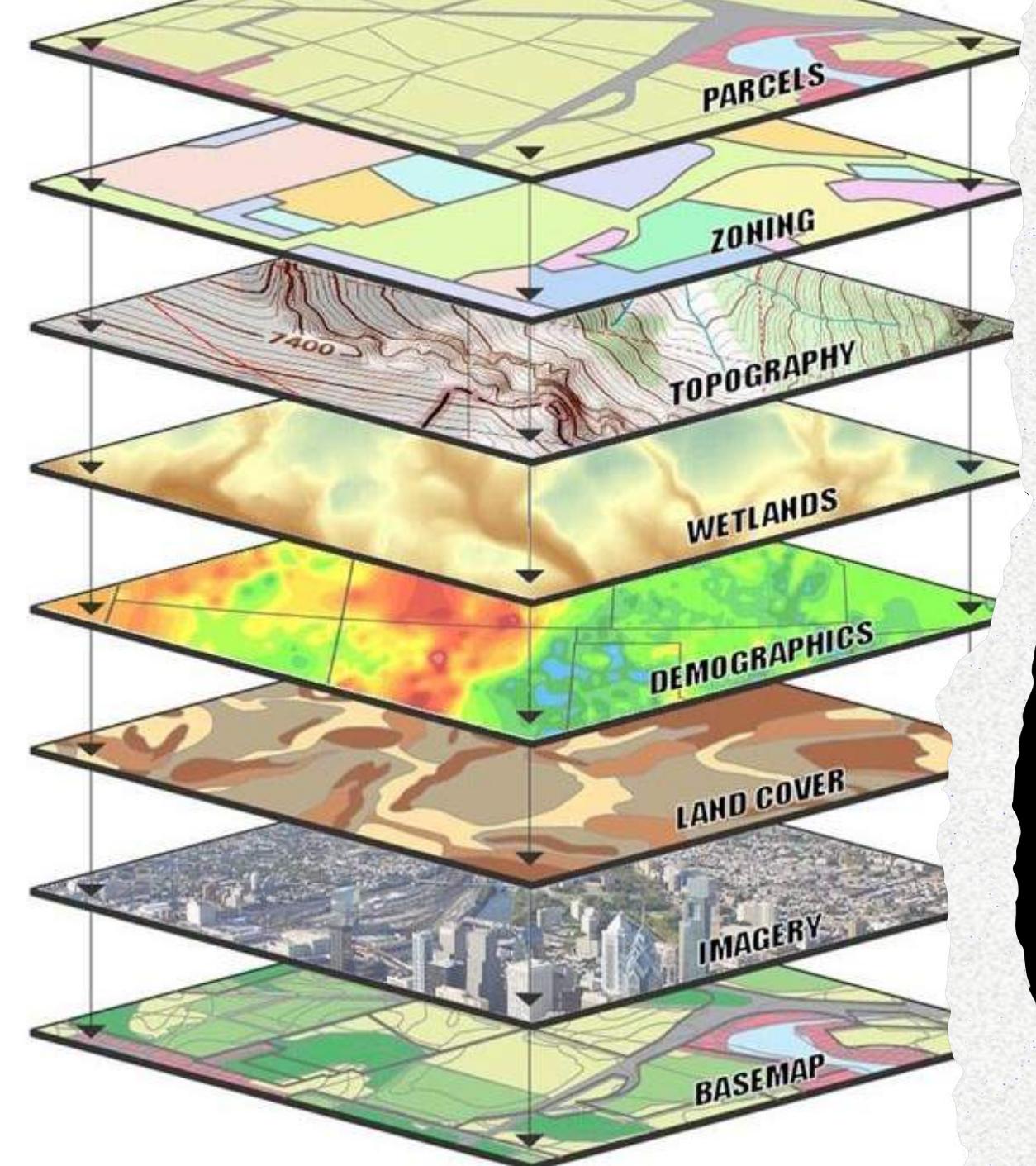
BOROUGH OF MANHATTAN			
Food and Drugs			
2907	number	Bernlohr & Katz	name Bell Milk
1931	number	Hechman & Drack	name Bell Milk
576	number	Julius Schreder	nature
2197	number	Theobald Bros.	name
4938	number	Louis Spitzer	name
2917	number	Edw. Tannenbaum	name
1231	number	Max Moskowitz	name
		Bridget Neffors	name
		Bush & Somerville	name
		Herman Alperstein	name
		Joseph Feinberg	name
		Daniel Reeves, Inc.	name
		Goldberg & Reich	name
		Lee, Shifron	name
		Fred Neisselman	name
		Joe, Wazigan	name
		Atlas Economy Stores, Inc.	name
		Henry Wisch	name
		Paul Struck	name
		Harry Fleishman	name
		Chas Cronquist	name
		Fettus Fenn	name
		Joseph Marin	name
		Daniel Reeves, Inc.	name
		John Ferre	name
		Chas. Gerstner	name
		Frank Mutterer	name
		Pape & Brandt	name
		Nikolas Fensch	name
		Jacob Sheinberg	name
		Henry Weintraub	name
		Hyman Neckowitz	name
		Wagner & Grobe	name
		David Fichtenbaum	name
		Bull Mani	name
		Broka & Srokah	name
		En, L. Korsch	name
		Hollmen Bros.	name
		Wittmann & Igors	name
		Eric Hoffman	name
		Atlas Economy Stores	name
		Christian Boscneider	name
		Isak Kawut	name
		Barnett Schneider	name
		Maple Lodge Farms Co., Inc.	name
		Pasquale Femese	name
		Joe, Schuman	name
		Sacks & Blumenthal	name
		Andrew Davy Inc.	name
		Max Bernstein	name
		Commodore Bakery Inc.	name
		Thos. Hasselbach	name
		Max Gold	name
		Chas. Gottlieb	name
		John Stechmann	name
		Podhorser & Spitzer	name

210 - 3rd Avenue address
 11 - 2nd Ave. address
 624 - 2nd Ave. address
 2195 - 3rd Ave. address
 1257 - 3rd Ave. address
 1101 address
 1044 - 1st Ave. address
 418 W. 39th St. address
 562 W. 52nd St. address
 418 W. 39th St. address
 1402 - 1st Ave. address
 1502 - 1st Ave. address
 1402 address
 108 Henry St. address
 200 - 7th Ave. address
 1708 - 3rd Ave. address
 1655 - 3rd Ave. address
 1934 address
 527 W. 133rd St. address
 1824 - 2nd Ave. address
 176 - 3rd Avenue address
 2451 - 2nd Ave. address
 2525 - 7th Ave. address
 148 Sherman Ave. address
 151 - 6th Ave. address
 571 - 1st Ave. address
 22 Stanton St. address
 140 Stanton St. address
 175 - 3rd Avenue address
 397 - 3rd Ave. address
 432 - 3rd Ave. address
 551 - 3rd Ave. address
 712 - 3rd Ave. address
 194 address
 215 address
 905 address
 1052 address
 1141 address
 1240 address
 1332 address
 1408 address
 1599 address
 1660 address
 1905 address
 317 Washington St. address
 1 West 11th St. address
 501 West 125th St. address
 227 West 145th St. address
 226 West 146th St. address
 550 West 207th St. address
 505 West 207th St. address
 505 West 208th St. address
 37 Grand St. address
 287 Henry St. address
 197 Hester St. address
 343 Hudson St. address
 506 Hudson St. address

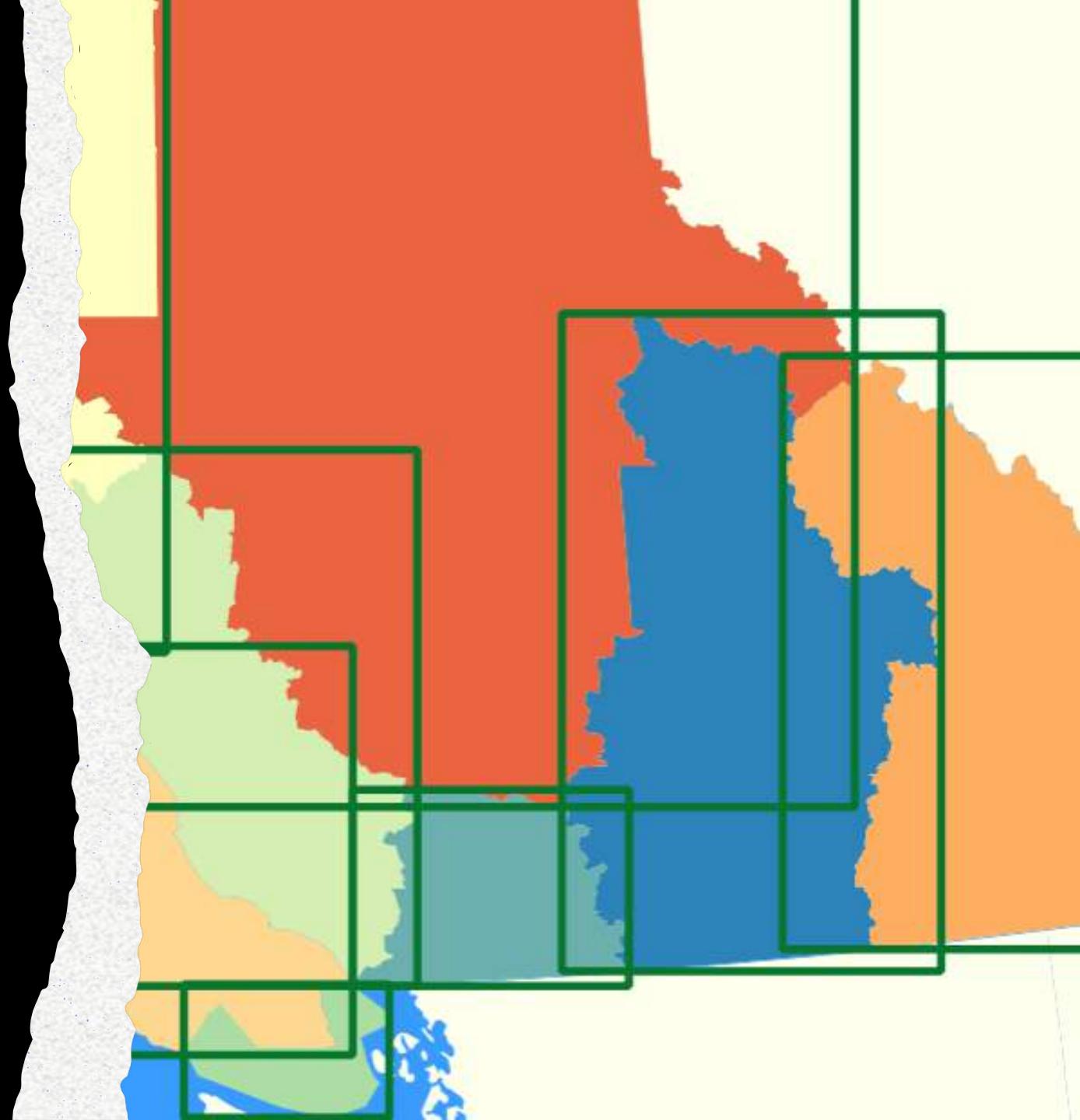


After data transformation,
special data still require *special handling*

Changes how
we *store*:
GIS data layers

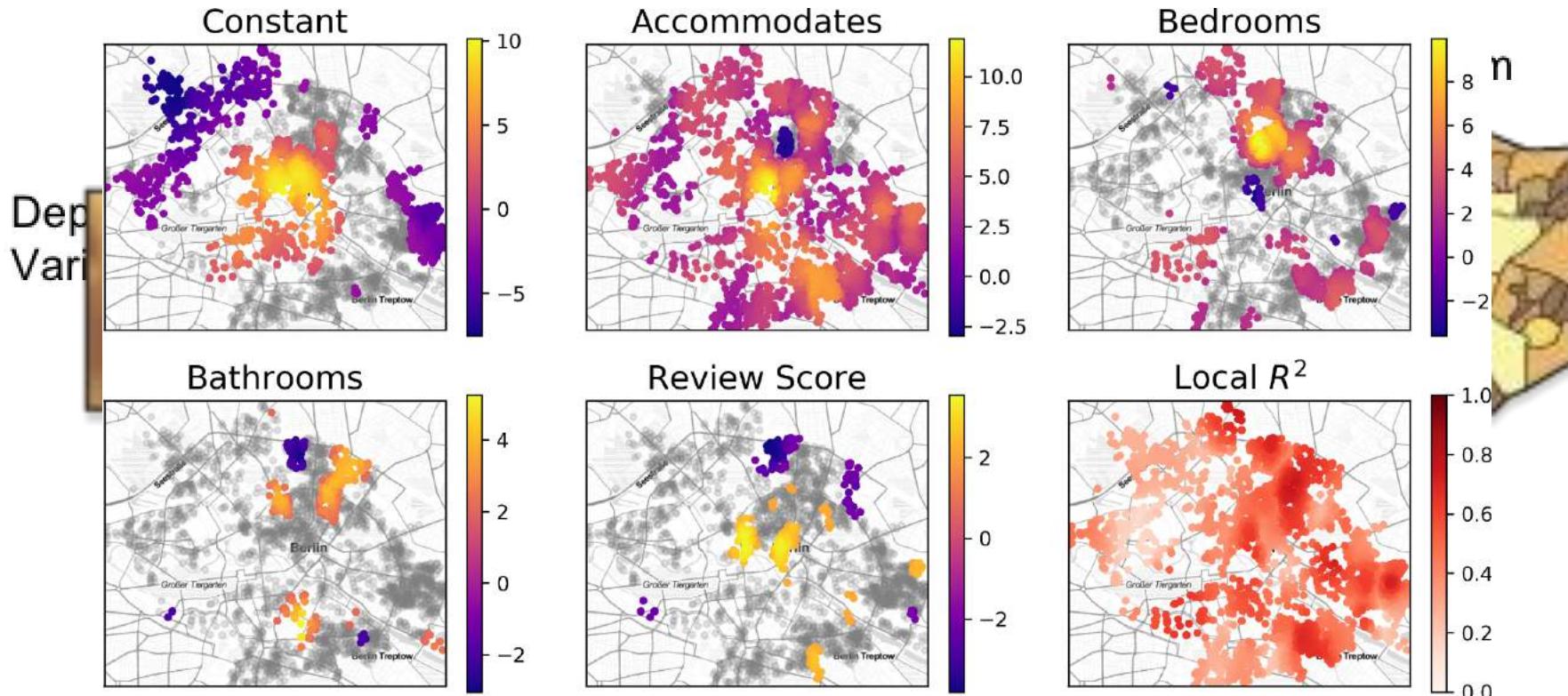


Changes how
we *search*:
Spatial index
bounding boxes



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.