

Introduction and Overview

CS424: Visualization & Visual Analytics

Fabio Miranda

<https://fmiranda.me>

A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion – fuelled by internet of things and the use of connected devices – are hard to comprehend, particularly when looked at in the context of one day

 500m
tweets are sent every day

Twitter

294bn
billion emails are sent

Radicati Group

3.9bn
people use emails

Radicati Group

320bn

emails to be sent each day by 2021

306bn

emails to be sent each day by 2020

Radicati Group



ACCUMULATED DIGITAL UNIVERSE OF DATA

4.4ZB

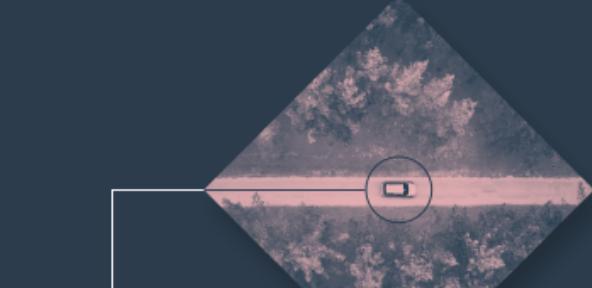
2010

44ZB

2020



Facebook Research



4TB
of data produced by a connected car

Intel

DEMYSTIFYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit	Value	Size
b	0 or 1	1/8 of a byte
B	8 bits	1 byte
KB	1,000 bytes	1,000 bytes
MB	1,000 ³ bytes	1,000,000 bytes
GB	1,000 ⁶ bytes	1,000,000,000 bytes
TB	1,000 ¹² bytes	1,000,000,000,000 bytes
PB	1,000 ¹⁵ bytes	1,000,000,000,000,000 bytes
EB	1,000 ¹⁸ bytes	1,000,000,000,000,000,000 bytes
ZB	1,000 ²¹ bytes	1,000,000,000,000,000,000,000 bytes
YB	1,000 ²⁴ bytes	1,000,000,000,000,000,000,000,000 bytes

A lowercase 'b' is used as an abbreviation for bits, while an uppercase 'B' represents bytes.

65bn

messages sent over WhatsApp and two billion minutes of voice and video calls made

Facebook



5bn
Searches made a day

3.5bn
Searches made a day from Google

Smart Insights



463EB

of data will be created every day by 2025

idc

95m

photos and videos are shared on Instagram

Instagram Business

28PB

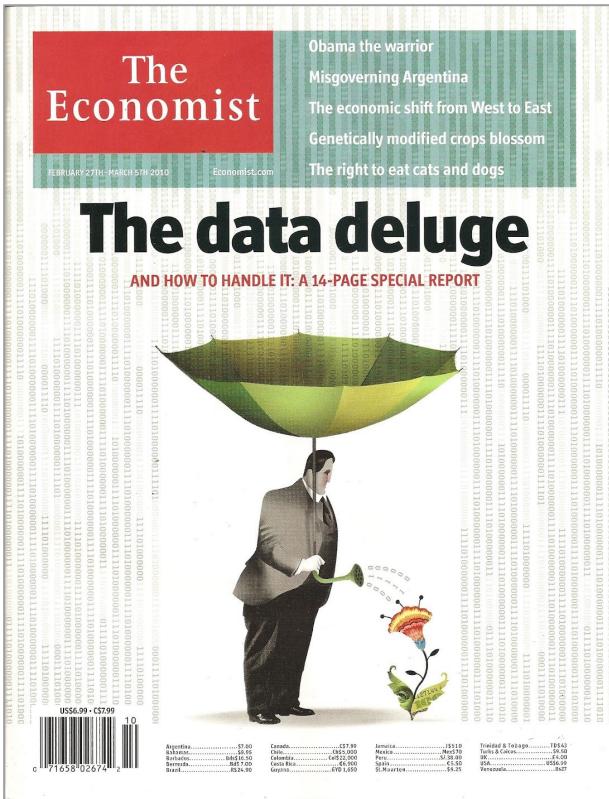
to be generated from wearable devices by 2020

Statista



Source: Raconteur

Data is everywhere



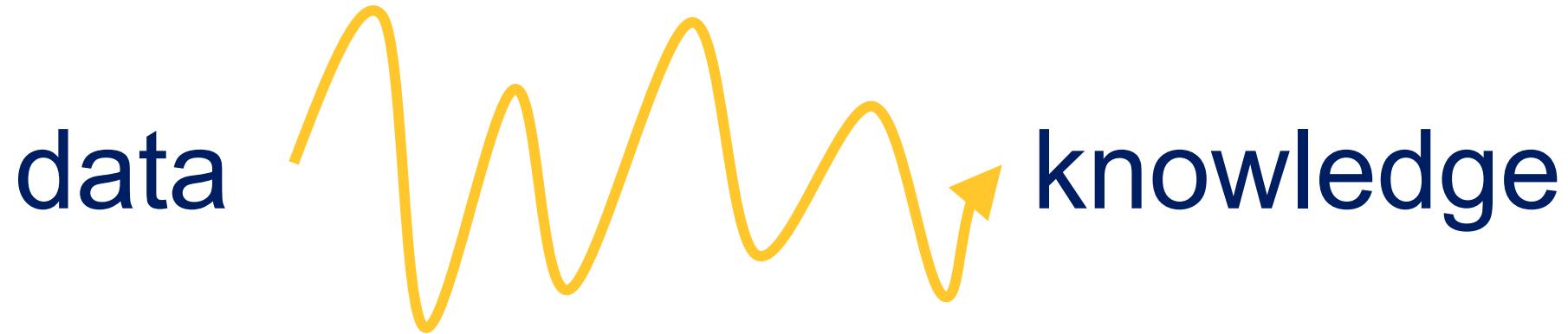
“The ability to take data—to be able to understand it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it—that’s going to be a hugely important skill in the next decades, ... because now we really do have essentially **free** and **ubiquitous data**.”

Hal Varian, Google’s Chief Economist
The McKinsey Quarterly, Jan 2009

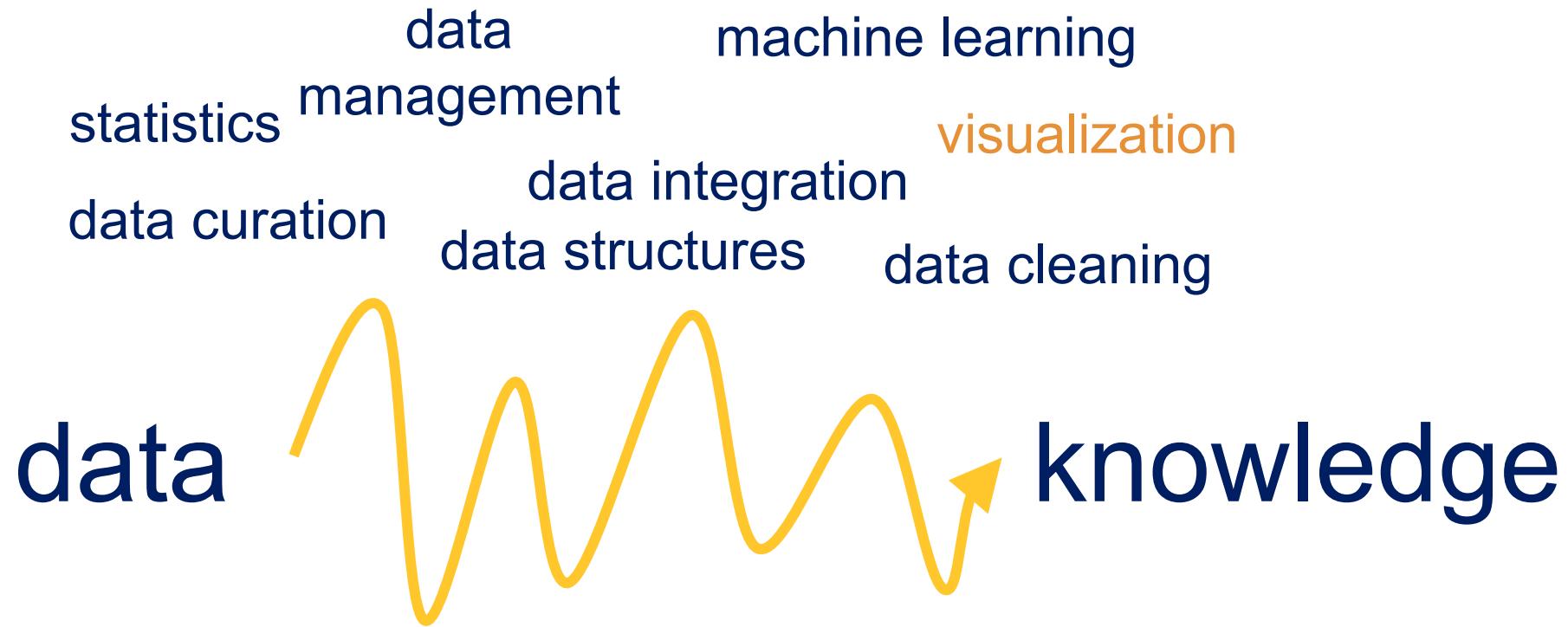
Data to knowledge

data → knowledge

Data to knowledge



Data to knowledge

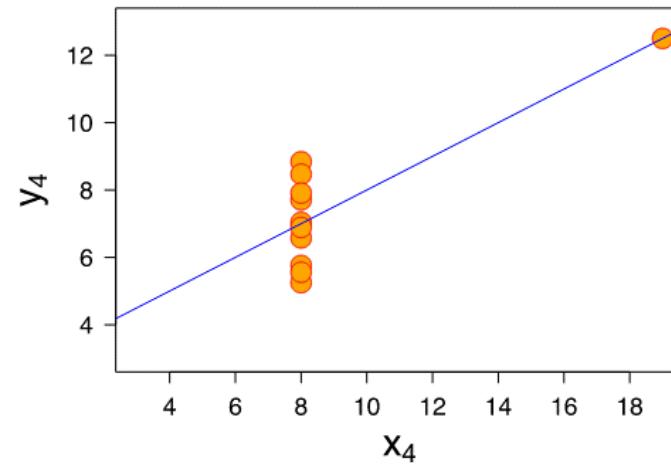
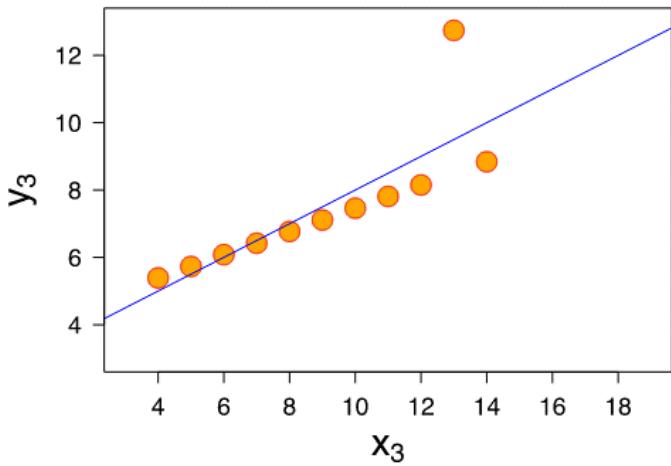
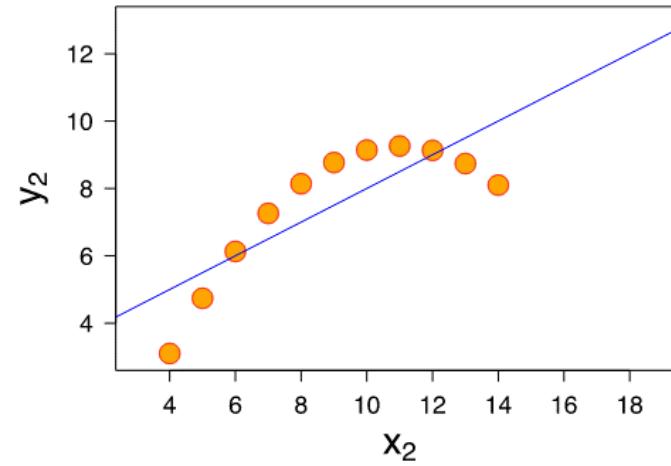
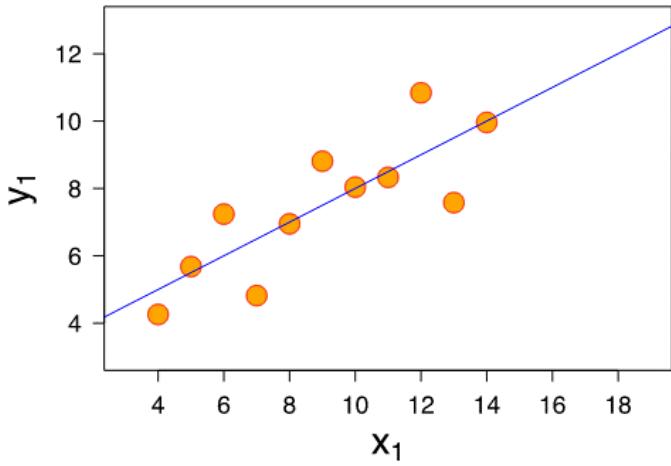


Importance of visualization

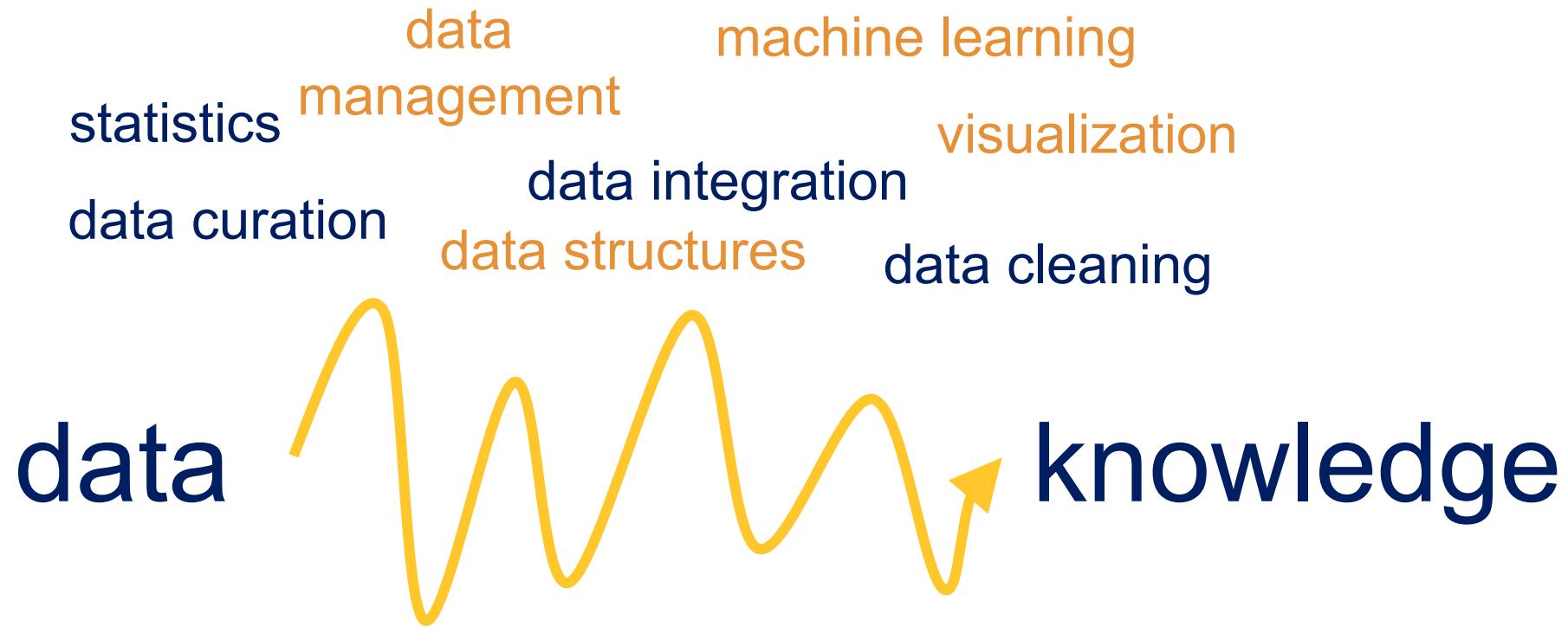
A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.7 4	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.5 0
12.0	10.8 4	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Property	A	B	C	D
Mean of x	9	9	9	9
Mean of y	7.5	7.5	7.5	7.5
Std of x	3.32	3.32	3.32	3.32
Std of y	2.03	2.03	2.03	2.03

Importance of visualization



Data to knowledge



Data analysis: Common practices

1. Domain experts and policy makers formulate hypotheses.
 2. Computer scientists or data scientists select data sets and slices, perform analyses, and derive plots.
 3. Domain experts examine the plots, go to step 1.
- Issues:
 - Dependency on computer scientists or data scientists distance domain experts from the data.
 - Batch-oriented analysis pipeline hampers exploration – mostly confirmatory analyses.
 - Data are complex – often multivariate spatiotemporal.
 - Analysis limited to samples.

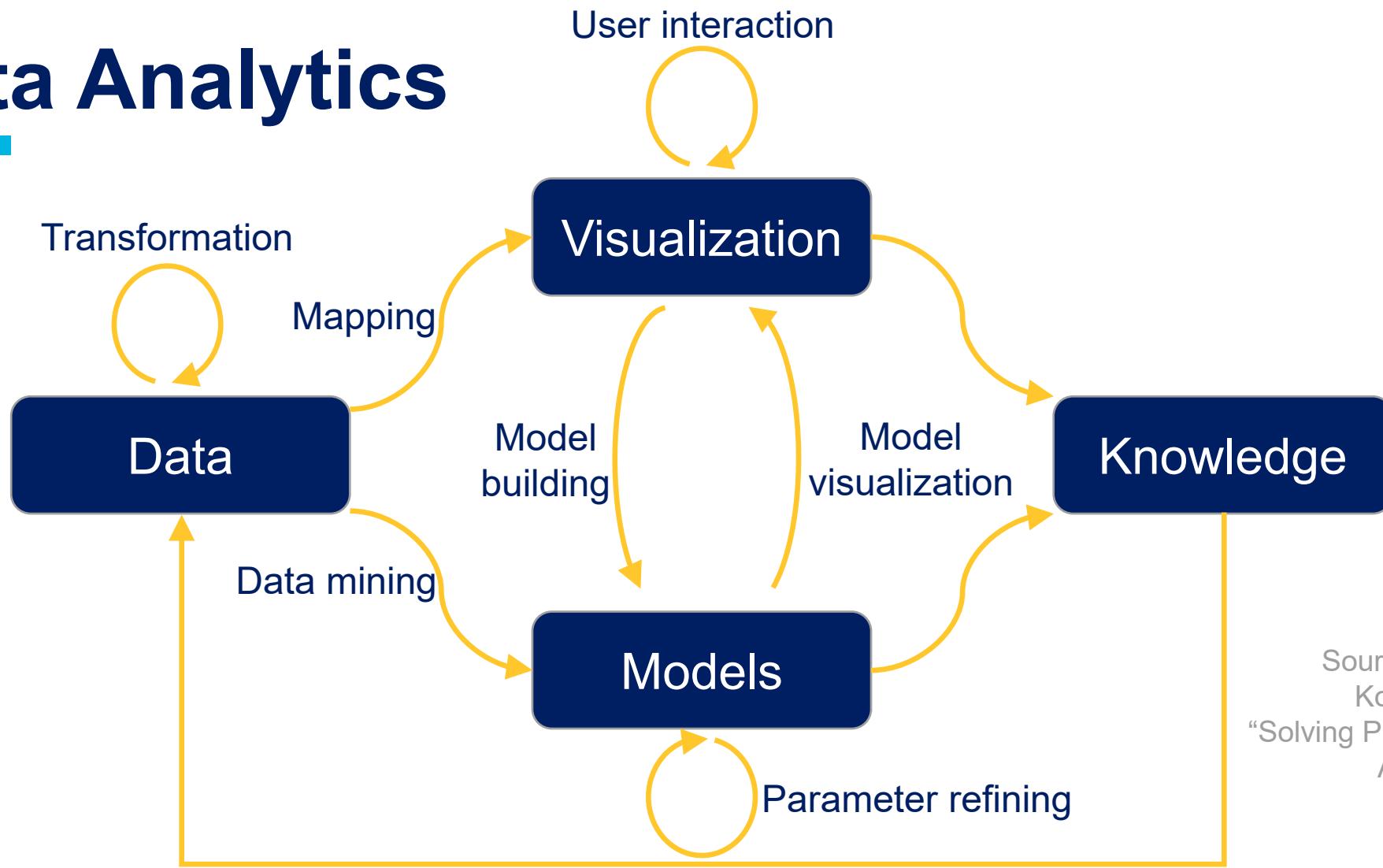
Data analysis: What we need

- Scalable tools and techniques that help domain experts find, clean, integrate, ***interactively*** explore, and explain data.
- Guide users in the exploration process.
- Support interactive queries:
 - “*increased latency reduces the rate at which users make observations, draw generalizations, and generate hypotheses*”.
- Interdisciplinary:
 - “*as data scale and complexity increases, the novel solutions that will ultimately enable interactive, large-scale exploratory data analysis will have to come from truly interdisciplinary work*”.

[Liu and Heer, IEEE TVCG 2014]

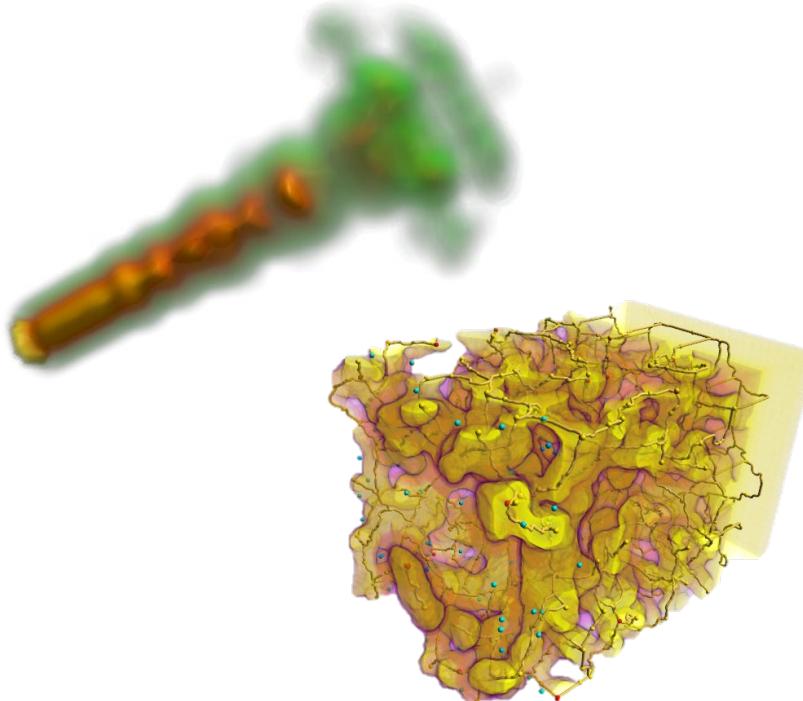
[Chang, Fekete, Freire and Scheidegger, Dagstuhl Reports 2017]

Data Analytics



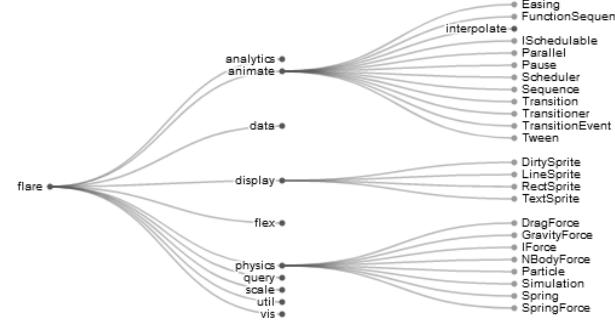
Source: Kleim and Kohlhammer,
“Solving Problems with Visual
Analytics”

Visualization



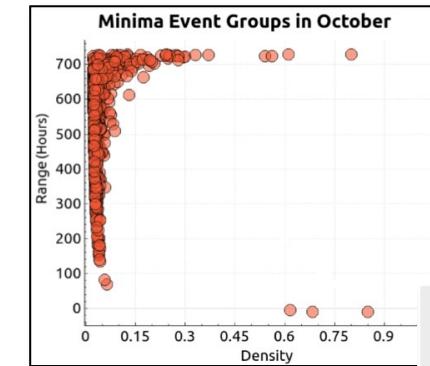
Scientific visualization

Data that has a natural geometric structure (wind flows, MRI).



Information visualization

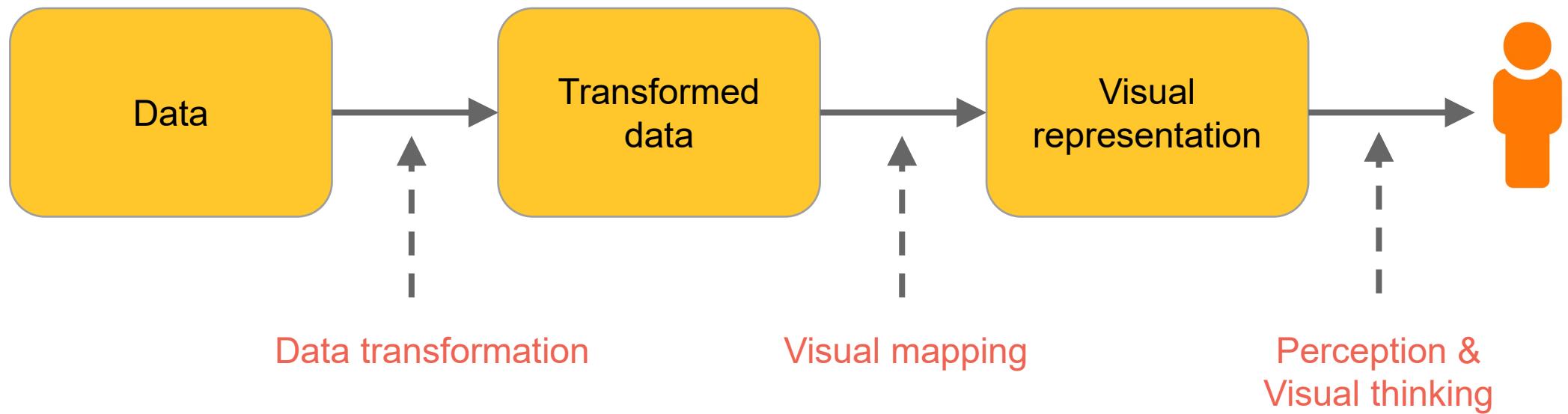
Abstract representations (trees, graphs)



Visual analytics

Interactive visual representations and underlying analytical processes (data mining, stats, etc.)

How to visualize data



Taxi example



Distribution of
NYC Taxi Pickups
and Dropoffs in
Midtown

Taxi example

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fatigue	PULocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount
1	1/1/2018 0:21	1/1/2018 0:24	1	0.5	1 N		41	24	2	4.5	0.5	0.5	0	0	0.3	5.8
1	1/1/2018 0:44	1/1/2018 1:03	1	2.7	1 N		239	140	2	14	0.5	0.5	0	0	0.3	15.3
1	1/1/2018 0:08	1/1/2018 0:14	2	0.8	1 N		262	141	1	6	0.5	0.5	1	0	0.3	8.3
1	1/1/2018 0:20	1/1/2018 0:52	1	10.2	1 N		140	257	2	33.5	0.5	0.5	0	0	0.3	34.8
1	1/1/2018 0:09	1/1/2018 0:27	2	2.5	1 N		246	239	1	12.5	0.5	0.5	2.75	0	0.3	16.55
1	1/1/2018 0:29	1/1/2018 0:32	3	0.5	1 N		143	143	2	4.5	0.5	0.5	0	0	0.3	5.8
1	1/1/2018 0:38	1/1/2018 0:48	2	1.7	1 N		50	239	1	9	0.5	0.5	2.05	0	0.3	12.35
1	1/1/2018 0:49	1/1/2018 0:51	1	0.7	1 N		239	238	1	4	0.5	0.5	1	0	0.3	6.3
1	1/1/2018 0:56	1/1/2018 1:01	1	1	1 N		238	24	1	5.5	0.5	0.5	1.7	0	0.3	8.5
1	1/1/2018 0:17	1/1/2018 0:22	1	0.7	1 N		170	170	2	5.5	0.5	0.5	0	0	0.3	6.8
1	1/1/2018 0:41	1/1/2018 0:46	1	0.6	1 N		162	229	1	5.5	0.5	0.5	1.35	0	0.3	8.15

Data transformation

Filter the data;

- Only rows within Manhattan.
- Only rows within certain blocks of Manhattan.

Merge the data with other data;

- Traffic accidents within 100 meters and 1 hour of pick or dropoff.

Aggregate the data;

- Number of pickups in each hour.
- Number of picks in each day of the week.

Data transformation

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_PUlocationID	DOlocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount	
1	1/1/2018 0:21	1/1/2018 0:24	1	0.5	1 N		41	24	2	4.5	0.5	0.5	0	0	0.3	5.8
1	1/1/2018 0:44	1/1/2018 1:03	1	2.7	1 N	239	140	2	14	0.5	0.5	0	0	0.3	15.3	
1	1/1/2018 0:08	1/1/2018 0:14	2	0.8	1 N	262	141	1	6	0.5	0.5	1	0	0.3	8.3	
1	1/1/2018 0:20	1/1/2018 0:52	1	10.2	1 N	140	257	2	33.5	0.5	0.5	0	0	0.3	34.8	
1	1/1/2018 0:09	1/1/2018 0:27	2	2.5	1 N	246	239	1	12.5	0.5	0.5	2.75	0	0.3	16.55	
1	1/1/2018 0:29	1/1/2018 0:32	3	0.5	1 N	143	143	2	4.5	0.5	0.5	0	0	0.3	5.8	
1	1/1/2018 0:38	1/1/2018 0:48	2	1.7	1 N	50	239	1	9	0.5	0.5	2.05	0	0.3	12.35	
1	1/1/2018 0:49	1/1/2018 0:51	1	0.7	1 N	239	238	1	4	0.5	0.5	1	0	0.3	6.3	
1	1/1/2018 0:56	1/1/2018 1:01	1	1	1 N	238	24	1	5.5	0.5	0.5	1.7	0	0.3	8.5	
1	1/1/2018 0:17	1/1/2018 0:22	1	0.7	1 N	170	170	2	5.5	0.5	0.5	0	0	0.3	6.8	
1	1/1/2018 0:41	1/1/2018 0:46	1	0.6	1 N	162	229	1	5.5	0.5	0.5	1.35	0	0.3	8.15	
1	1/1/2018 0:52	1/1/2018 1:17	1	3.5	1 N	141	113	2	16.5	0.5	0.5	0	0	0.3	17.8	
2	1/1/2018 0:17	1/1/2018 0:22	1	1.04	1 N	137	224	2	5.5	0.5	0.5	0	0	0.3	6.8	
2	1/1/2018 0:24	1/1/2018 0:34	1	1.22	1 N	224	79	2	7.5	0.5	0.5	0	0	0.3	8.8	
2	1/1/2018 0:37	1/1/2018 0:53	1	1.92	1 N	234	100	2	10	0.5	0.5	0	0	0.3	11.3	
1	1/1/2018 0:35	1/1/2018 0:52	1	5.7	1 N	13	189	1	19	0.5	0.5	4.05	0	0.3	24.35	
2	1/1/2018 0:30	1/1/2018 1:13	1	3.74	1 N	48	236	1	25.5	0.5	0.5	6.7	0	0.3	33.5	
1	1/1/2018 0:21	1/1/2018 0:25	2	0.6	1 N	163	162	1	4.5	0.5	0.5	1.7	0	0.3	7.5	
1	1/1/2018 0:31	1/1/2018 1:07	1	10.9	1 N	229	61	2	35	0.5	0.5	0	0	0.3	36.3	
2	1/1/2018 0:15	1/1/2018 0:21	5	1.22	1 N	236	75	2	6	0.5	0.5	0	0	0.3	7.3	
2	1/1/2018 0:25	1/1/2018 0:45	5	3.13	1 N	263	143	2	13	0.5	0.5	0	0	0.3	14.3	
2	1/1/2018 0:51	1/1/2018 1:04	5	2.22	1 N	239	24	2	9.5	0.5	0.5	0	0	0.3	10.8	
2	1/1/2018 0:09	1/1/2018 0:30	1	2.93	1 N	90	233	1	14.5	0.5	0.5	2	0	0.3	17.8	
2	1/1/2018 0:32	1/1/2018 0:58	1	3.52	1 N	233	125	2	18	0.5	0.5	0	0	0.3	19.3	
1	1/1/2018 0:41	1/1/2018 0:54	4	3	1 N	161	146	1	12	0.5	0.5	2.65	0	0.3	15.95	
2	1/1/2018 0:17	1/1/2018 0:21	5	0.25	1 N	234	234	2	4.5	0.5	0.5	0	0	0.3	5.8	
2	1/1/2018 0:24	1/1/2018 0:46	5	3.31	1 N	234	143	1	16	0.5	0.5	3.46	0	0.3	20.76	
2	1/1/2018 0:48	1/1/2018 0:51	5	0.57	1 N	142	239	1	4	0.5	0.5	1.06	0	0.3	6.36	
1	1/1/2018 0:24	1/1/2018 0:31	2	0.7	1 N	170	162	2	6	0.5	0.5	0	0	0.3	7.3	
1	1/1/2018 0:36	1/1/2018 0:43	1	1.8	1 N	233	263	2	7.5	0.5	0.5	0	0	0.3	8.8	
1	1/1/2018 0:49	1/1/2018 0:57	2	1.2	1 N	236	237	2	7.5	0.5	0.5	0	0	0.3	8.8	
1	1/1/2018 0:13	1/1/2018 0:23	1	2.7	1 N	142	166	1	10.5	0.5	0.5	2.35	0	0.3	14.15	
1	1/1/2018 0:33	1/1/2018 1:18	2	4.3	1 N	238	249	2	27.5	0.5	0.5	0	0	0.3	28.8	
2	1/1/2018 0:15	1/1/2018 0:22	1	0.89	1 N	151	238	2	5.5	0.5	0.5	0	0	0.3	6.8	
2	1/1/2018 0:25	1/1/2018 0:29	1	0.49	1 N	238	238	1	4.5	0.5	0.5	1.45	0	0.3	7.25	
2	1/1/2018 0:32	1/1/2018 0:36	2	0.8	1 N	238	151	1	5	0.5	0.5	1.26	0	0.3	7.56	
2	1/1/2018 0:45	1/1/2018 0:58	1	2.09	1 N	238	143	1	11	0.5	0.5	2.46	0	0.3	14.76	
2	1/1/2018 0:31	1/1/2018 0:45	1	2.32	1 N	186	231	1	11	0.5	0.5	3.08	0	0.3	15.38	
2	1/1/2018 0:47	1/1/2018 1:26	1	9.49	1 N	231	116	1	35	0.5	0.5	9.08	0	0.3	45.38	
1	1/1/2018 0:21	1/1/2018 0:28	2	2.5	1 N	141	145	1	9.5	0.5	0.5	2.7	0	0.3	13.5	
1	1/1/2018 0:32	1/1/2018 0:47	1	4.6	1 N	145	263	1	15.5	0.5	0.5	4.2	0	0.3	21	
1	1/1/2018 0:54	1/1/2018 1:03	1	3	1 N	141	146	2	10.5	0.5	0.5	0	0	0.3	11.8	
1	1/1/2018 0:23	1/1/2018 0:52	1	7.3	1 N	90	82	1	26.5	0.5	0.5	1	5.76	0.3	34.56	
1	1/1/2018 0:04	1/1/2018 0:15	1	1.3	1 N	144	234	1	9	0.5	0.5	2.05	0	0.3	12.35	
1	1/1/2018 0:17	1/1/2018 0:41	1	0.8	1 N	234	164	2	14.5	0.5	0.5	0	0	0.3	15.8	
1	1/1/2018 0:42	1/1/2018 0:44	1	0.1	1 N	164	164	2	3	0.5	0.5	0	0	0.3	4.3	
1	1/1/2018 0:48	1/1/2018 0:55	2	0.2	1 N	164	164	1	6	0.5	0.5	1.45	0	0.3	8.75	



Data transformation

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_t	PUlocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount
1	1/1/2018 0:21	1/1/2018 0:24	1	0.5	1 N		41	24	2	4.5	0.5	0.5	0	0	0.3	5.8
1	1/1/2018 0:44	1/1/2018 1:03	1	2.7	1 N		239	140	2	14	0.5	0.5	0	0	0.3	15.3
1	1/1/2018 0:08	1/1/2018 0:14	2	0.8	1 N		262	141	1	6	0.5	0.5	1	0	0.3	8.3
1	1/1/2018 0:20	1/1/2018 0:52	1	10.2	1 N		140	257	2	33.5	0.5	0.5	0	0	0.3	34.8
1	1/1/2018 0:09	1/1/2018 0:27	2	2.5	1 N		246	239	1	12.5	0.5	0.5	2.75	0	0.3	16.55
1	1/1/2018 0:29	1/1/2018 0:32	3	0.5	1 N		143	143	2	4.5	0.5	0.5	0	0	0.3	5.8
1	1/1/2018 0:38	1/1/2018 0:48	2	1.7	1 N		50	239	1	9	0.5	0.5	2.05	0	0.3	12.35
1	1/1/2018 0:49	1/1/2018 0:51	1	0.7	1 N		239	238	1	4	0.5	0.5	1	0	0.3	6.3
1	1/1/2018 0:56	1/1/2018 1:01	1	1	1 N		238	24	1	5.5	0.5	0.5	1.7	0	0.3	8.5
1	1/1/2018 0:17	1/1/2018 0:22	1	0.7	1 N		170	170	2	5.5	0.5	0.5	0	0	0.3	6.8
1	1/1/2018 0:41	1/1/2018 0:46	1	0.6	1 N		162	229	1	5.5	0.5	0.5	1.35	0	0.3	8.15
1	1/1/2018 0:52	1/1/2018 1:17	1	3.5	1 N		141	113	2	16.5	0.5	0.5	0	0	0.3	17.8
2	1/1/2018 0:17	1/1/2018 0:22	1	1.04	1 N		137	224	2	5.5	0.5	0.5	0	0	0.3	6.8
2	1/1/2018 0:24	1/1/2018 0:34	1	1.22	1 N		224	79	2	7.5	0.5	0.5	0	0	0.3	8.8
2	1/1/2018 0:37	1/1/2018 0:53	1	1.92	1 N		234	100	2	10	0.5	0.5	0	0	0.3	11.3
1	1/1/2018 0:35	1/1/2018 0:52	1	5.7	1 N		13	189	1	19	0.5	0.5	4.05	0	0.3	24.35
2	1/1/2018 0:30	1/1/2018 1:13	1	3.74	1 N		48	236	1	25.5	0.5	0.5	6.7	0	0.3	33.5
1	1/1/2018 0:21	1/1/2018 0:25	2	0.6	1 N		163	162	1	4.5	0.5	0.5	1.7	0	0.3	7.5
1	1/1/2018 0:31	1/1/2018 1:07	1	10.9	1 N		229	61	2	35	0.5	0.5	0	0	0.3	36.3
2	1/1/2018 0:15	1/1/2018 0:21	5	1.22	1 N		236	75	2	6	0.5	0.5	0	0	0.3	7.3
2	1/1/2018 0:25	1/1/2018 0:45	5	3.13	1 N		263	143	2	13	0.5	0.5	0	0	0.3	14.3
2	1/1/2018 0:51	1/1/2018 1:04	5	2.22	1 N		239	24	2	9.5	0.5	0.5	0	0	0.3	10.8
2	1/1/2018 0:09	1/1/2018 0:30	1	2.93	1 N		90	233	1	14.5	0.5	0.5	2	0	0.3	17.8
2	1/1/2018 0:32	1/1/2018 0:58	1	3.52	1 N		233	125	2	18	0.5	0.5	0	0	0.3	19.3

Data transformation

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_taxis_in	PULocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount	Area
1	1/1/2018 0:21	1/1/2018 0:24	1	0.5	1 N		41	24	2	4.5	0.5	0.5	0	0	0.3	5.3	Midtown
1	1/1/2018 0:44	1/1/2018 1:03	1	2.7	1 N		239	140	2	14	0.5	0.5	0	0	0.3	15.3	Chelsea
1	1/1/2018 0:08	1/1/2018 0:14	2	0.8	1 N		262	141	1	6	0.5	0.5	1	0	0.3	8.3	Downtown
1	1/1/2018 0:20	1/1/2018 0:52	1	10.2	1 N		140	257	2	33.5	0.5	0.5	0	0	0.3	34.3	Downtown
1	1/1/2018 0:09	1/1/2018 0:27	2	2.5	1 N		246	239	1	12.5	0.5	0.5	2.75	0	0.3	16.5	Downtown
1	1/1/2018 0:29	1/1/2018 0:32	3	0.5	1 N		143	143	2	4.5	0.5	0.5	0	0	0.3	5.3	Midtown
1	1/1/2018 0:38	1/1/2018 0:48	2	1.7	1 N		50	239	1	9	0.5	0.5	2.05	0	0.3	12.3	Downtown
1	1/1/2018 0:49	1/1/2018 0:51	1	0.7	1 N		239	238	1	4	0.5	0.5	1	0	0.3	6.3	Downtown
1	1/1/2018 0:56	1/1/2018 1:01	1	1	1 N		238	24	1	5.5	0.5	0.5	1.7	0	0.3	8.3	Downtown
1	1/1/2018 0:17	1/1/2018 0:22	1	0.7	1 N		170	170	2	5.5	0.5	0.5	0	0	0.3	6.3	Downtown
1	1/1/2018 0:41	1/1/2018 0:46	1	0.6	1 N		162	229	1	5.5	0.5	0.5	1.35	0	0.3	8.1	Midtown
1	1/1/2018 0:52	1/1/2018 1:17	1	3.5	1 N		141	113	2	16.5	0.5	0.5	0	0	0.3	17.3	Downtown
2	1/1/2018 0:17	1/1/2018 0:22	1	1.04	1 N		137	224	2	5.5	0.5	0.5	0	0	0.3	6.3	Downtown
2	1/1/2018 0:24	1/1/2018 0:34	1	1.22	1 N		224	79	2	7.5	0.5	0.5	0	0	0.3	8.3	Downtown
2	1/1/2018 0:37	1/1/2018 0:53	1	1.92	1 N		234	100	2	10	0.5	0.5	0	0	0.3	11.3	Downtown
1	1/1/2018 0:35	1/1/2018 0:52	1	5.7	1 N		13	189	1	19	0.5	0.5	4.05	0	0.3	24.3	Downtown
2	1/1/2018 0:30	1/1/2018 1:13	1	3.74	1 N		48	236	1	25.5	0.5	0.5	6.7	0	0.3	33.3	Downtown
1	1/1/2018 0:21	1/1/2018 0:25	2	0.6	1 N		163	162	1	4.5	0.5	0.5	1.7	0	0.3	7.3	Midtown
1	1/1/2018 0:31	1/1/2018 1:07	1	10.9	1 N		229	61	2	35	0.5	0.5	0	0	0.3	36.3	Midtown
2	1/1/2018 0:15	1/1/2018 0:21	5	1.22	1 N		236	75	2	6	0.5	0.5	0	0	0.3	7.3	Midtown
2	1/1/2018 0:25	1/1/2018 0:45	5	3.13	1 N		263	143	2	13	0.5	0.5	0	0	0.3	14.3	Midtown
2	1/1/2018 0:51	1/1/2018 1:04	5	2.22	1 N		239	24	2	9.5	0.5	0.5	0	0	0.3	10.3	Midtown
2	1/1/2018 0:09	1/1/2018 0:30	1	2.93	1 N		90	233	1	14.5	0.5	0.5	2	0	0.3	17.3	Midtown
2	1/1/2018 0:32	1/1/2018 0:58	1	3.52	1 N		233	125	2	18	0.5	0.5	0	0	0.3	19.3	Midtown

Visual mapping

Visual marks:

- Represent items.

④ Points



④ Lines



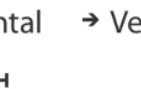
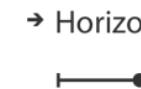
④ Areas



Visual channels:

- Control appearance based on data attributes.

④ Position



④ Color



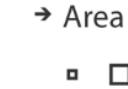
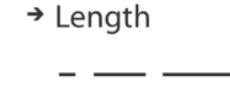
④ Shape



④ Tilt



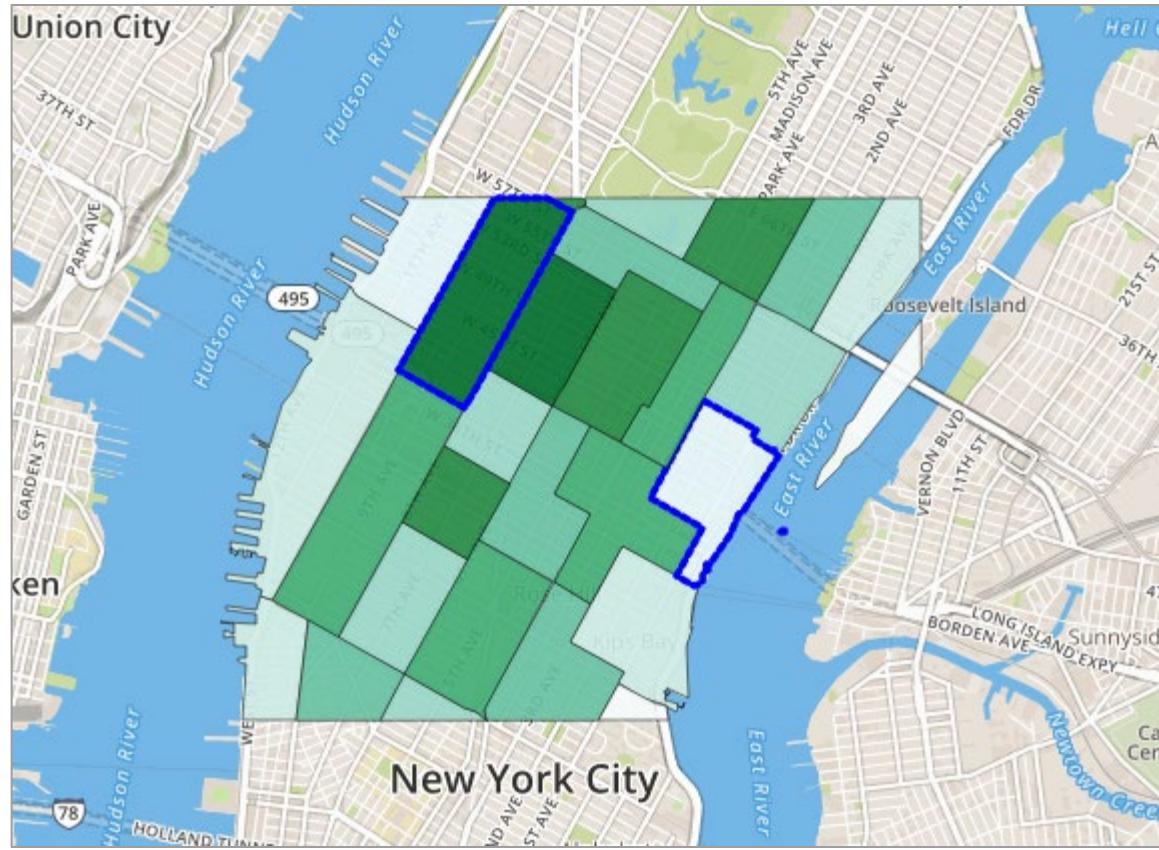
④ Size



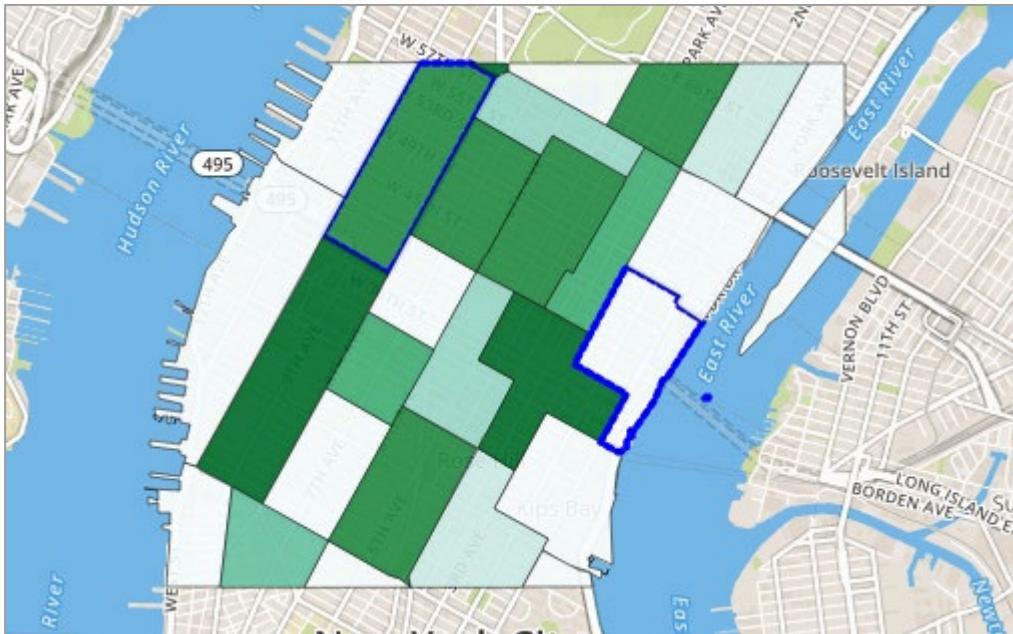
From: CS 171 – Harvard University

Visual mapping

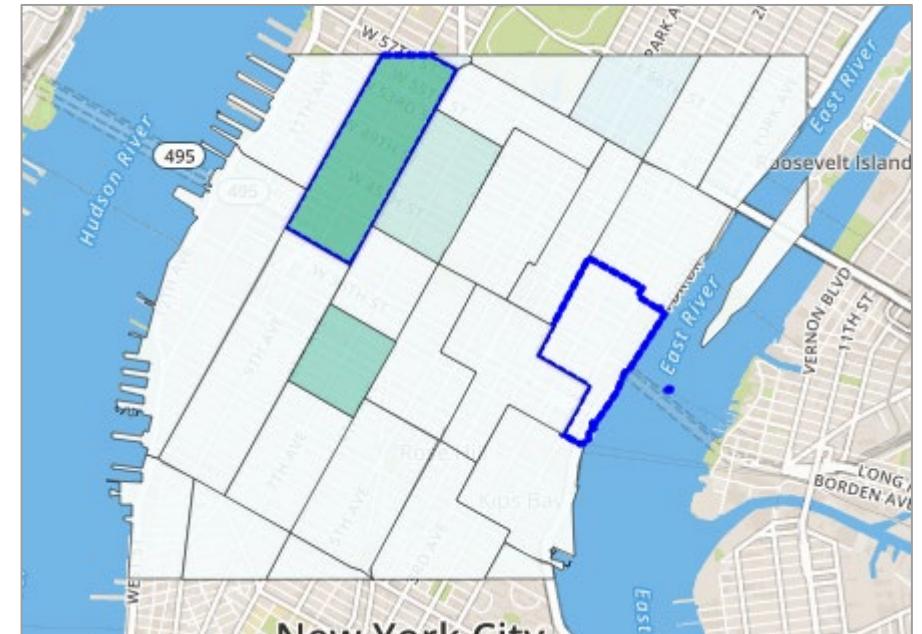
Quantitative data
Mark: polygon areas
Channel: color



Visual interaction

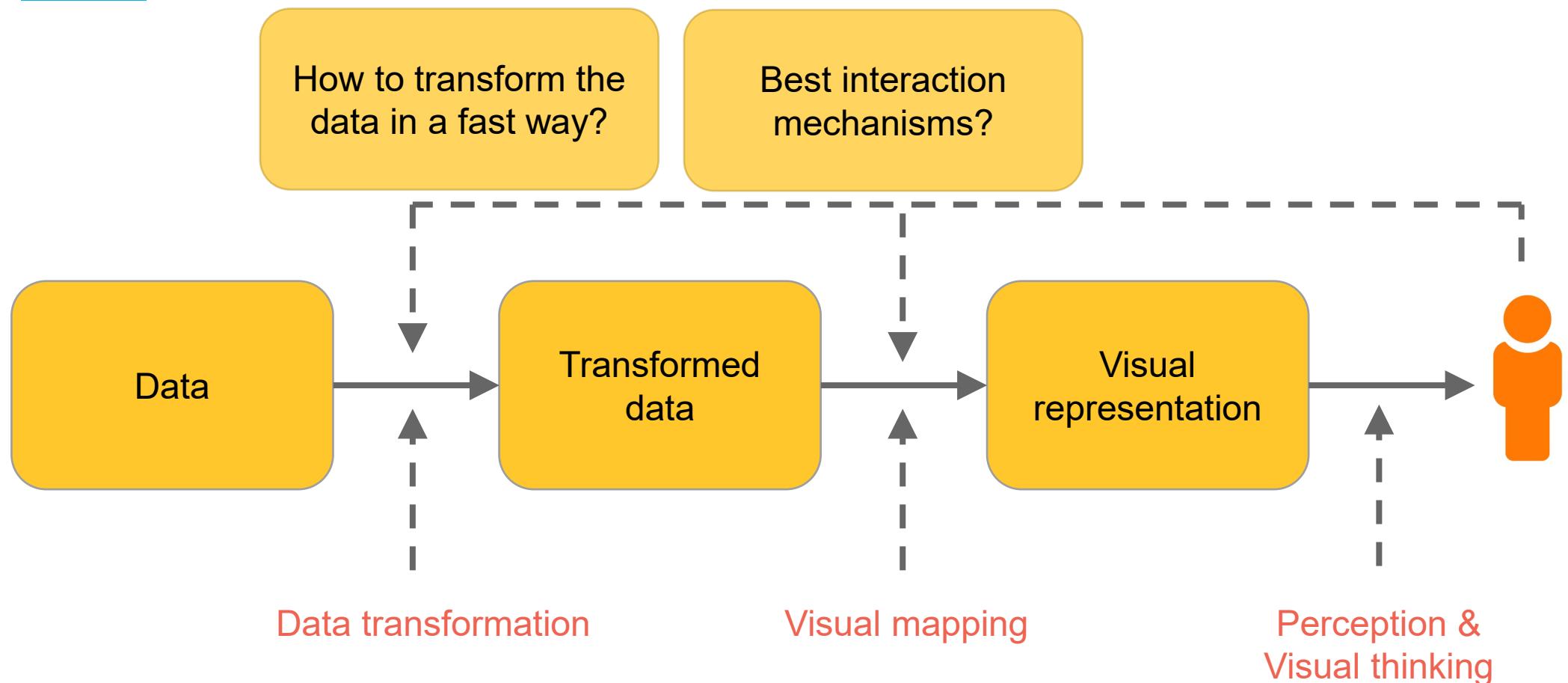


12pm – 2pm pickups



6am pickups

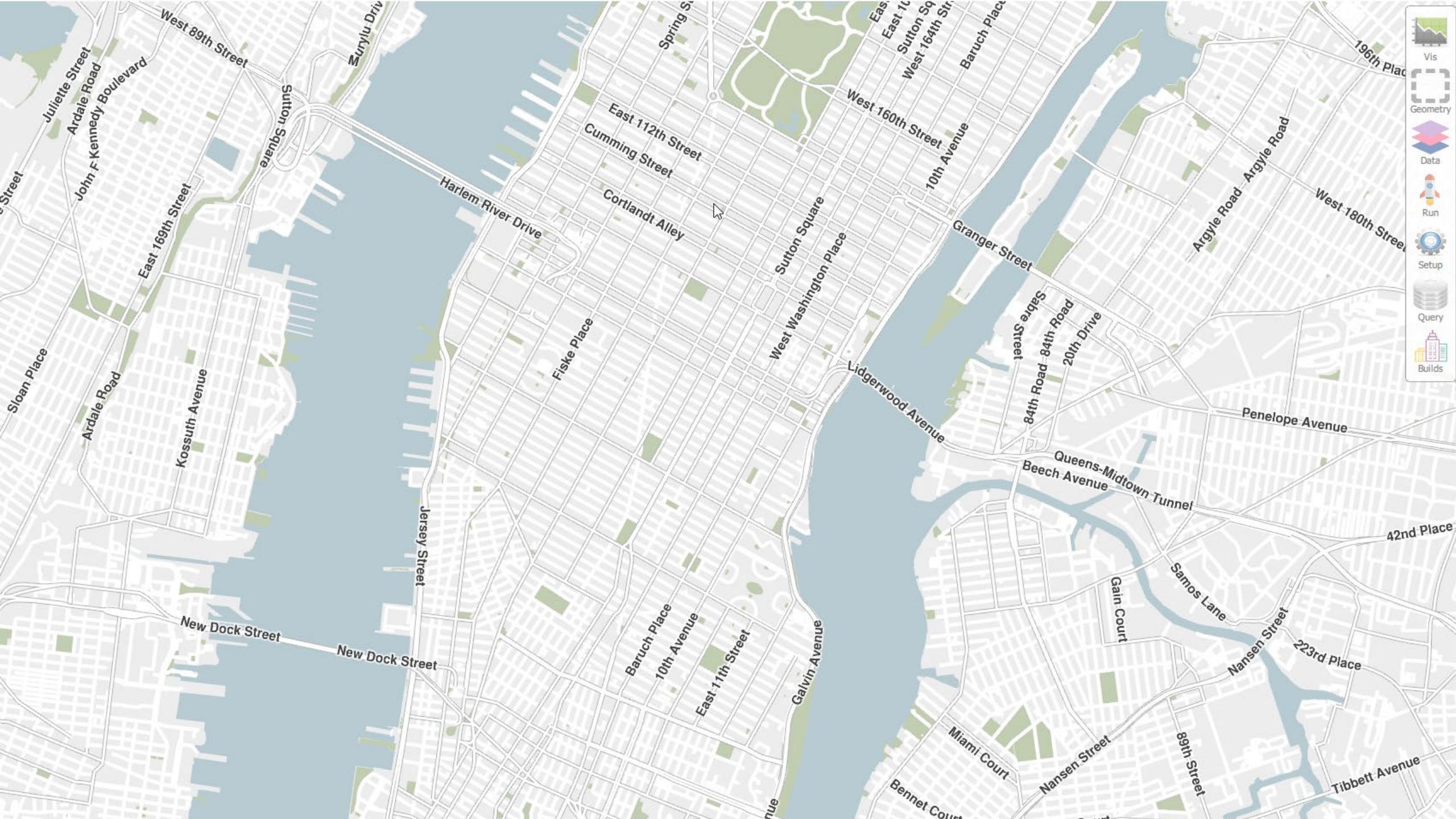
How to visualize data



Data-driven science

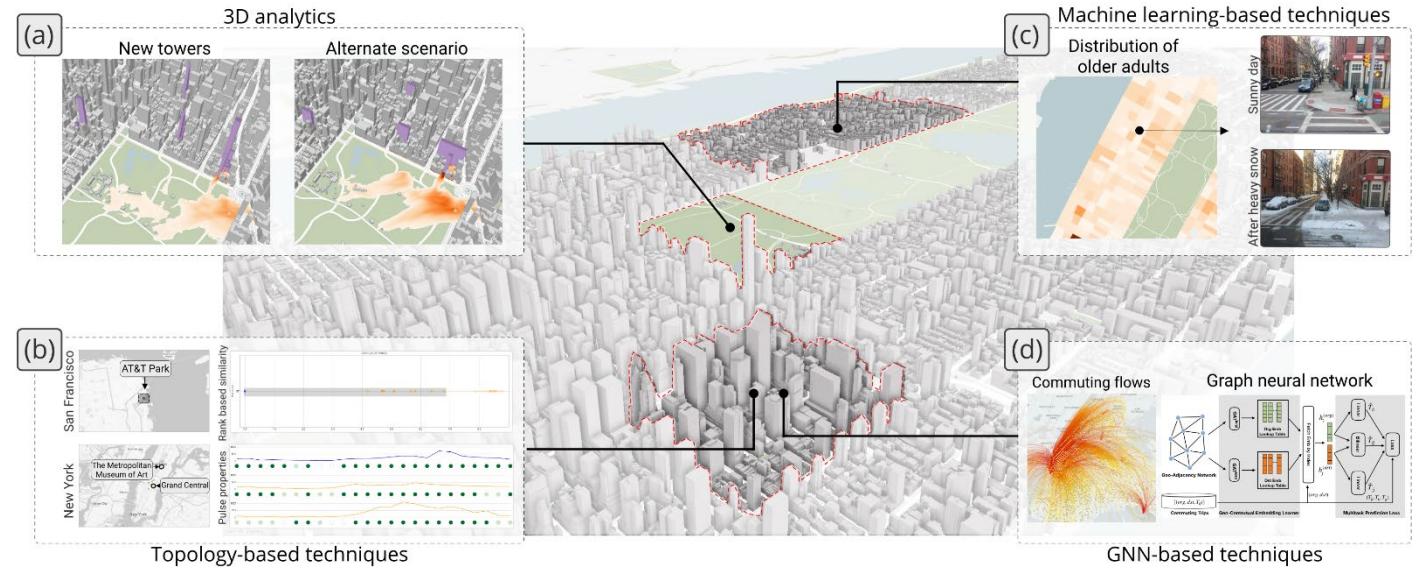
**“Visualization gives you answer to
questions you didn’t know you had.”**

Ben Schneiderman



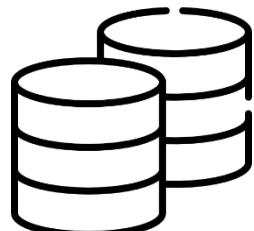
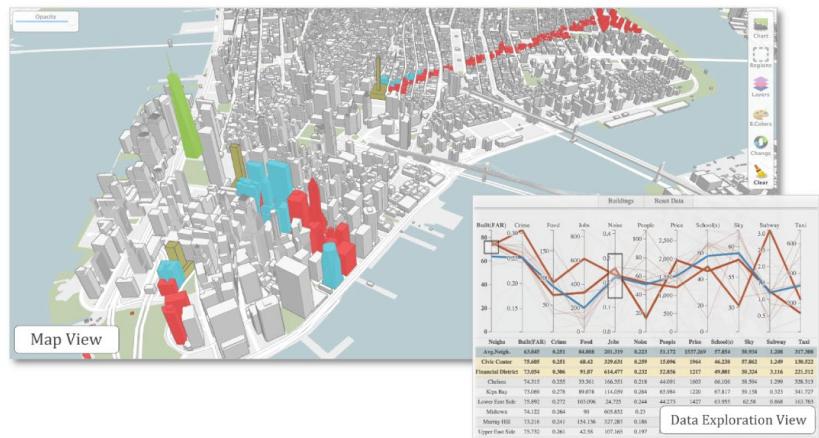
Research

Assistant Professor, CS (UIC)
PhD, CS 2018 (NYU)
MSc, CS 2012 (PUC-Rio)
BSc, CS 2009 (UFMG)

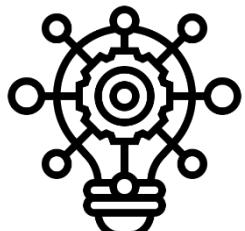


- Methods and techniques that follow a **human-centered approach to data science**, fostering the involvement of domain experts in the analysis process of big data.
- **Interactive tools and frameworks** that combine visualization, data management, human-computer interaction, and machine learning to support data-driven decision making by domain experts.

Visualization & Visual analytics



Data storage

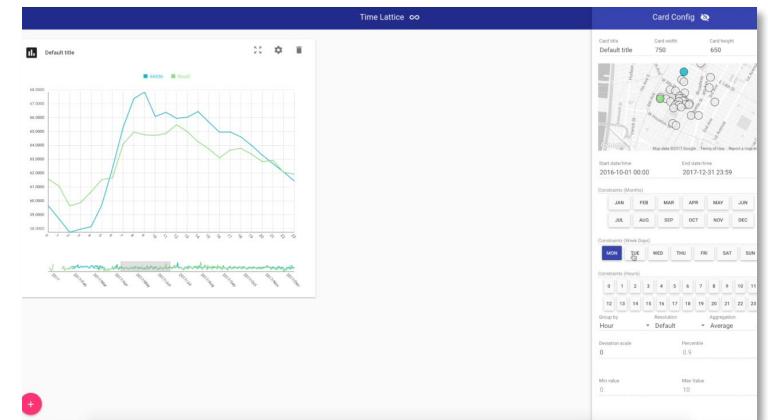
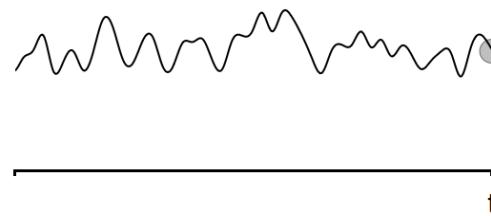


Data analytics



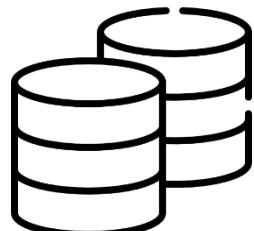
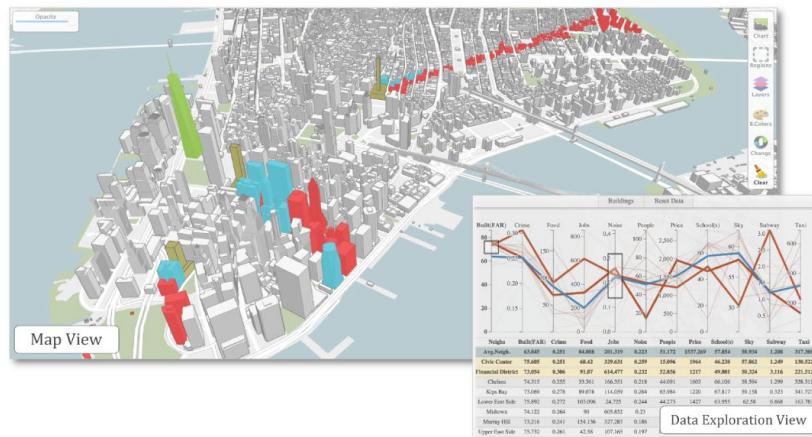
Data visualization

[Best SIGMOD'18 paper]

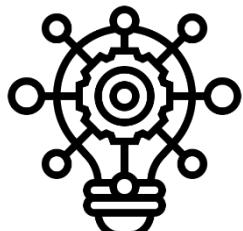


Time Lattice [CGF 2017]

Visualization & Visual analytics



Data storage

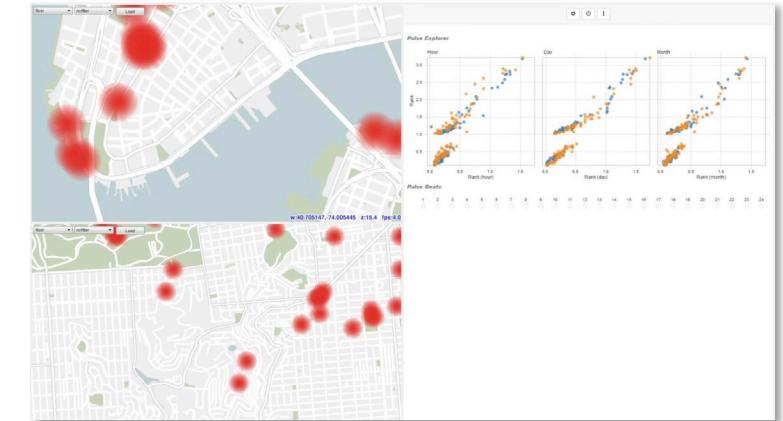
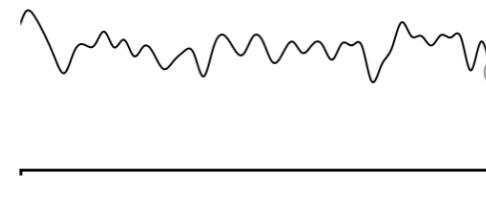


Data analytics



Data visualization

[Best SIGMOD'18 paper]

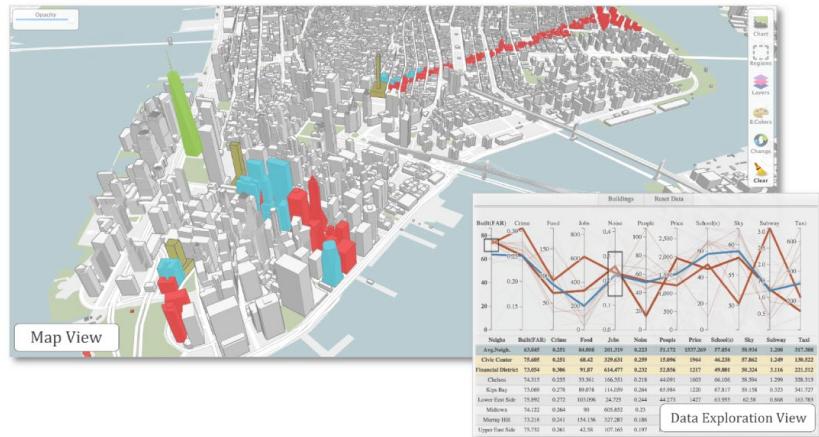


Urban Pulse [TVCG 2016]



COMPUTER SCIENCE

Big data vis



Data storage

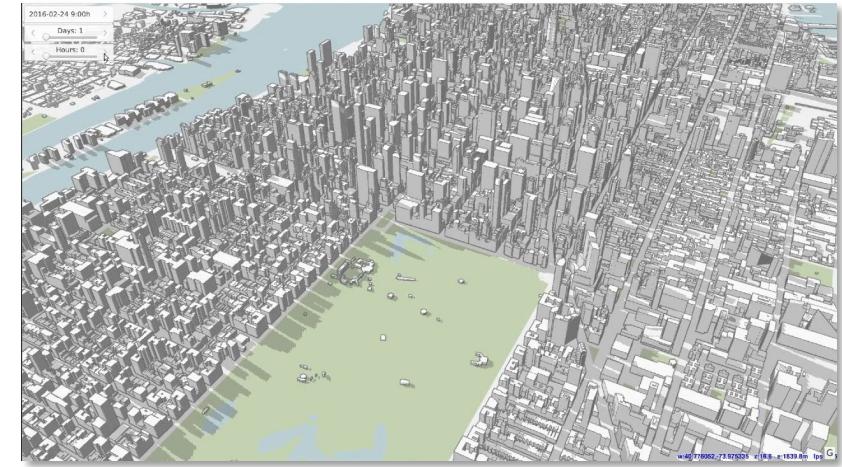
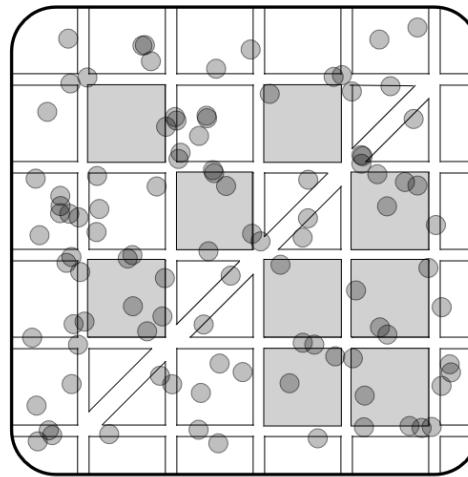


Data analytics



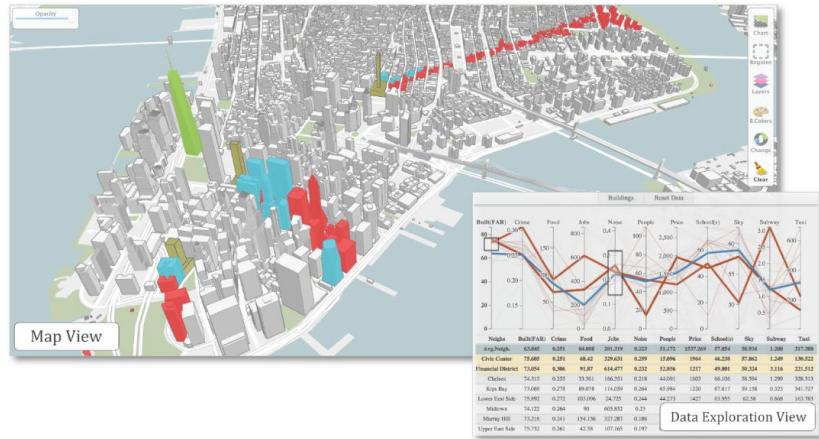
Data visualization

[Best SIGMOD'18 paper]



Shadow Accrual Maps [TVCG 2019, New York Times 2018]

Big data vis



Data storage

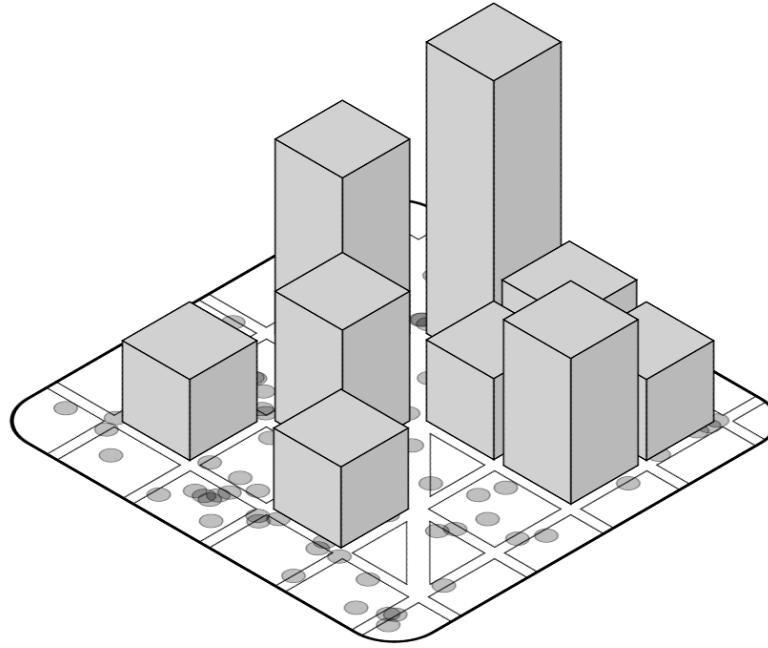


Data analytics



Data visualization

[Best SIGMOD'18 paper]

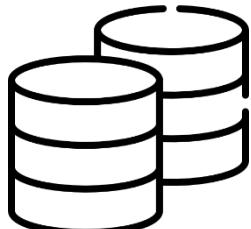


Urban Mosaic [CHI 2020]

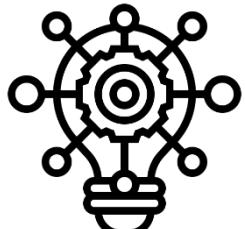


COMPUTER SCIENCE

Big data vis



Data storage

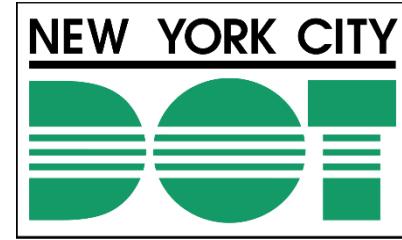


Data analytics



Data visualization

[Best SIGMOD'18 paper]



KPF

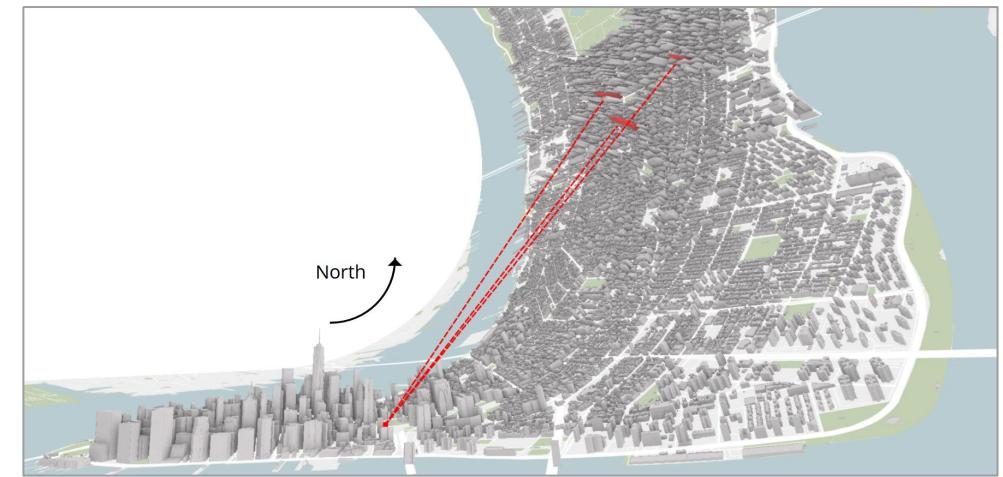
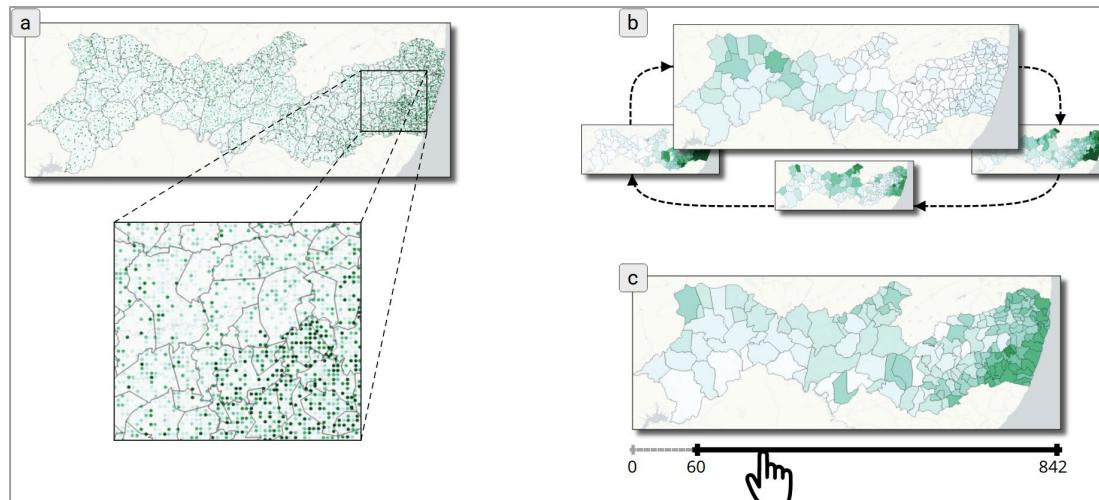
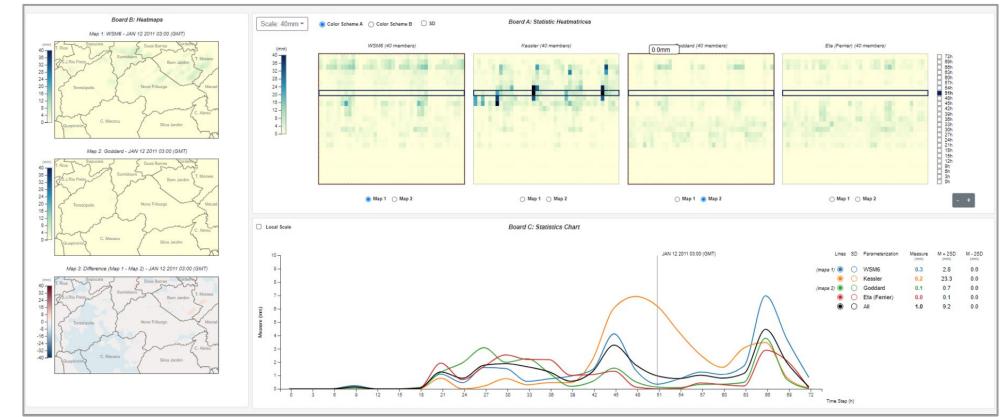
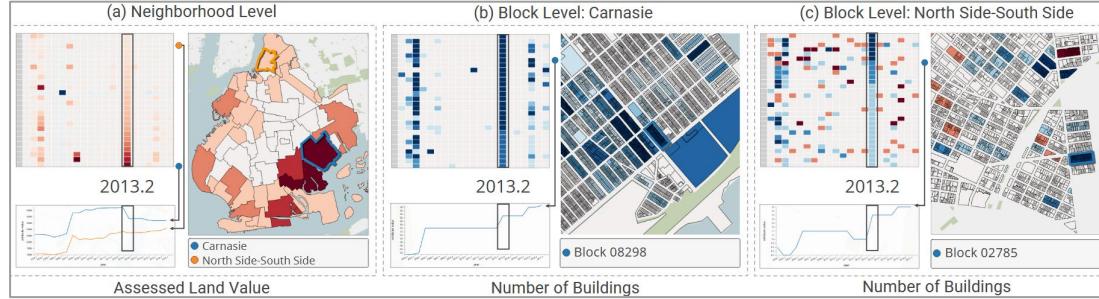
The New York Times



DRAW
BROOKLYN



Big data vis



CS424: Visualization & Visual Analytics

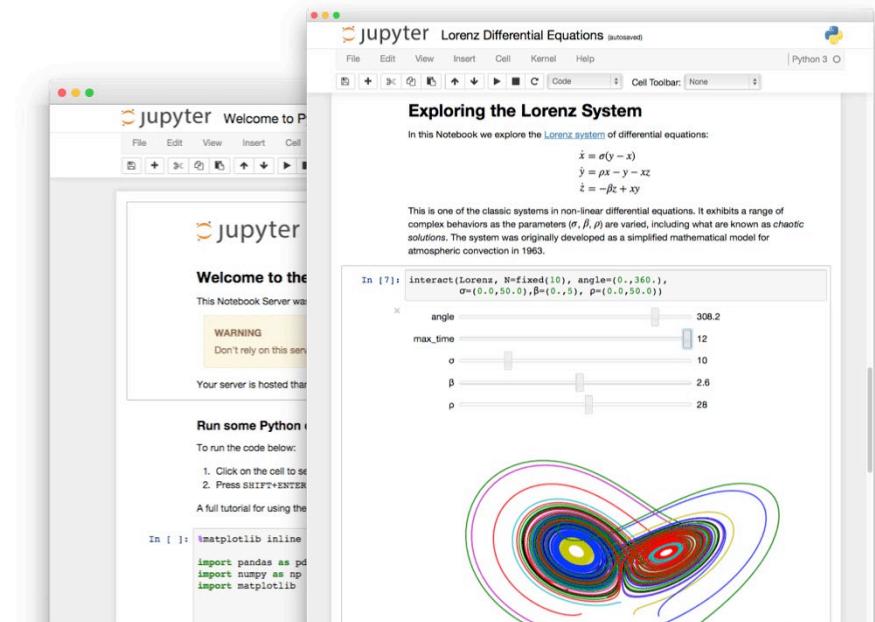
- Introduction to the field of visualization:
 - Scientific visualization
 - Information visualization
 - Visual analytics
- Design and implementation aspects of visual analytics systems.
 - Principles of data visualization
 - Exploratory data analysis
 - Data management for visualization
 - Visual analytics systems
 - ML for visualization
 - Uncertainty
 - Big data vis
 - ...
- At the end of the course, you will be able to design and implement web-based visual analytics systems by combining visualization, data management, data mining and machine learning techniques that work in tandem to enable interactive exploration of multidimensional and heterogeneous datasets.

Requirements

- CS 251 (Data Structures) and solid grasp of programming in a language like C, C++, Java, Python or JavaScript and basic data structures to be able to implement the visualization projects in the course.

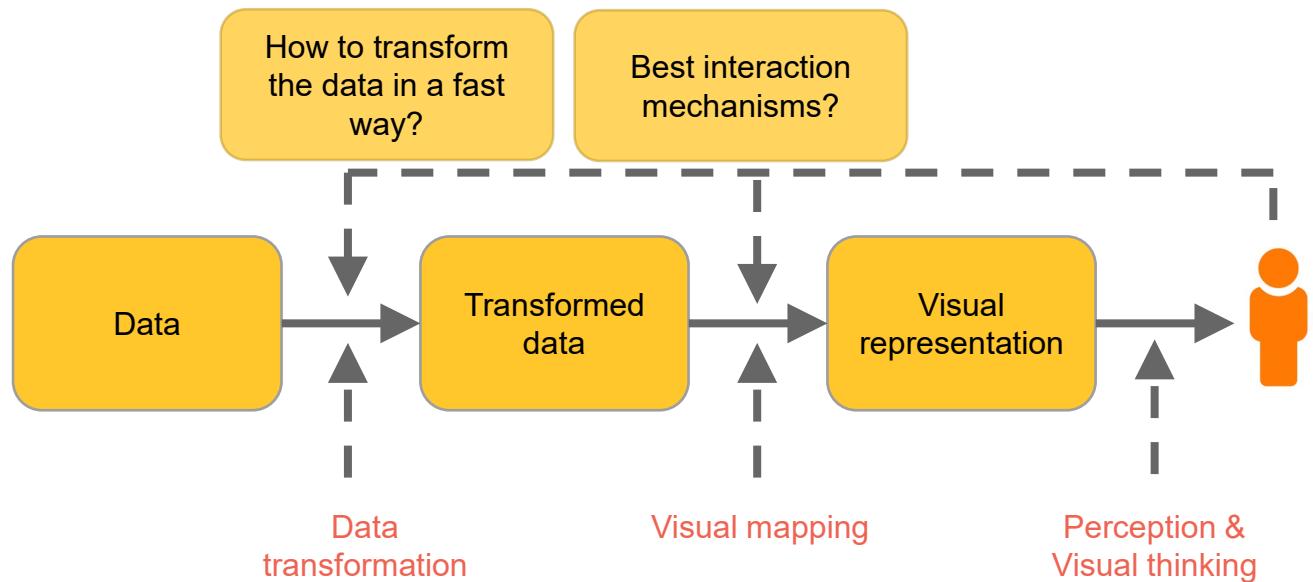
Week 2: Building blocks

- Necessary tools and frameworks for exploratory data analysis:
 - Git
 - Pandas
 - GeoPandas
 - Jupyter



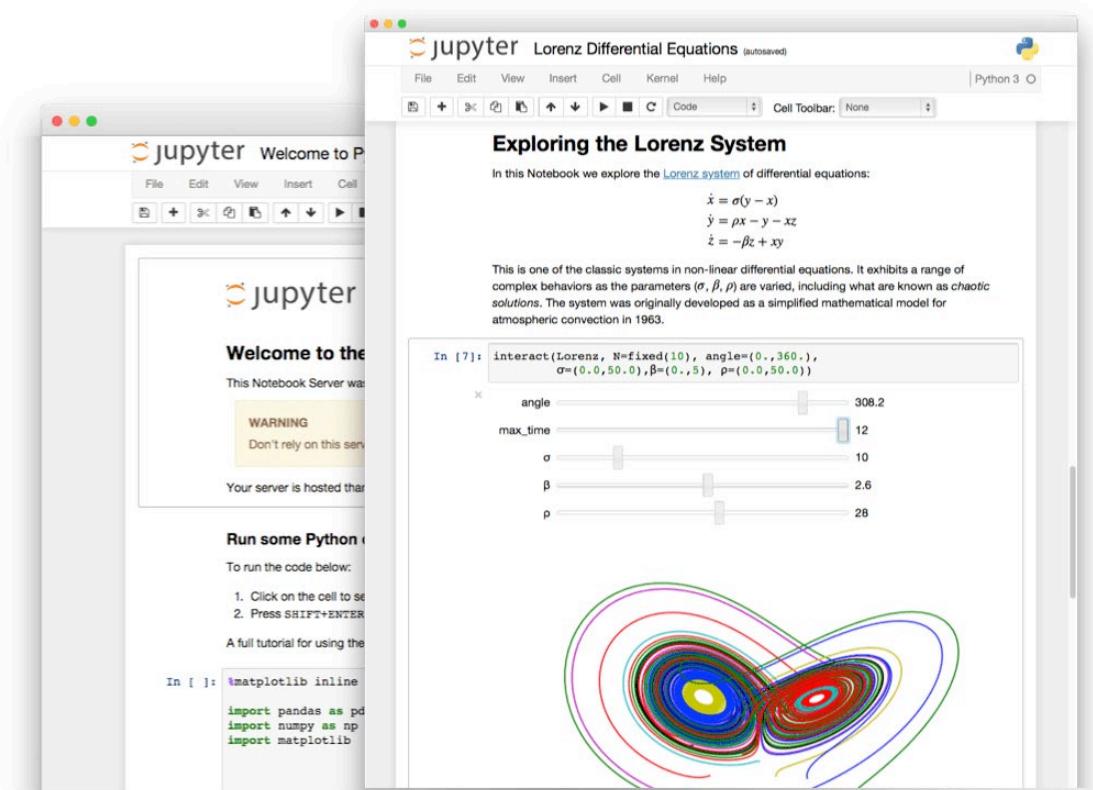
Week 3: Intro to vis

- Fundamentals of data visualization:
 - Data types
 - Attribute types
 - Channels
 - Tasks
 - Analysis loop



Week 4: Exploratory data analysis

- Integrating visualization, data management and analytics techniques to interactive computing environments.
 - Jupyter Notebooks



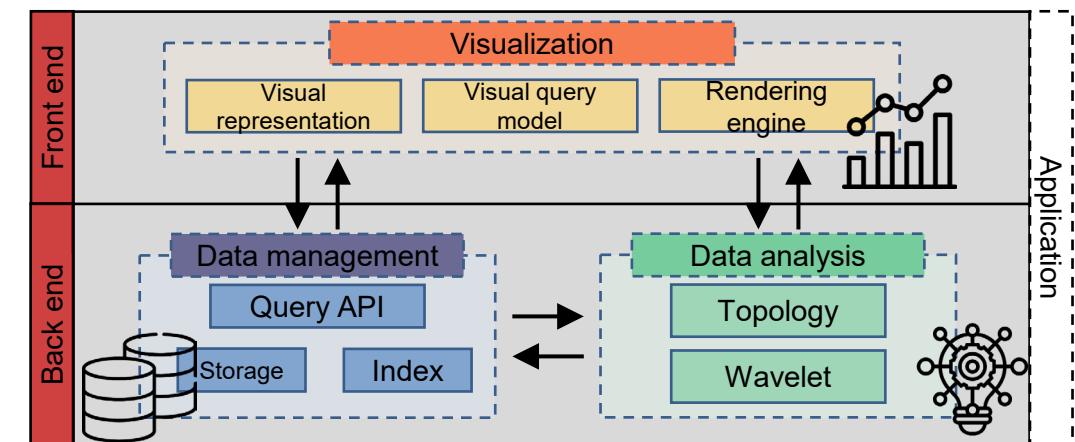
Week 5: Web technologies

- Necessary tools and frameworks for the development of big data analytic systems:
 - React (Javascript)
 - Angular (Javascript)
 - Flask (Python)



Week 7: Data management for vis

- Approximate queries, hashing, learned indices.
- Integration with MySQL, PostGIS.
- Back-end and front-end integration.
- Spatial queries.



Week 8: Coordinated views

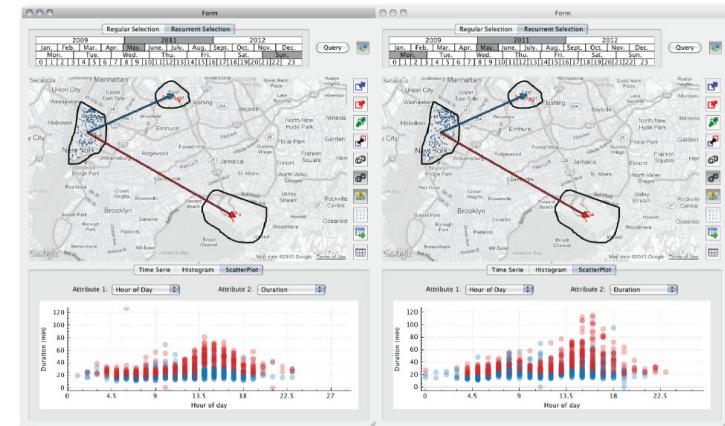
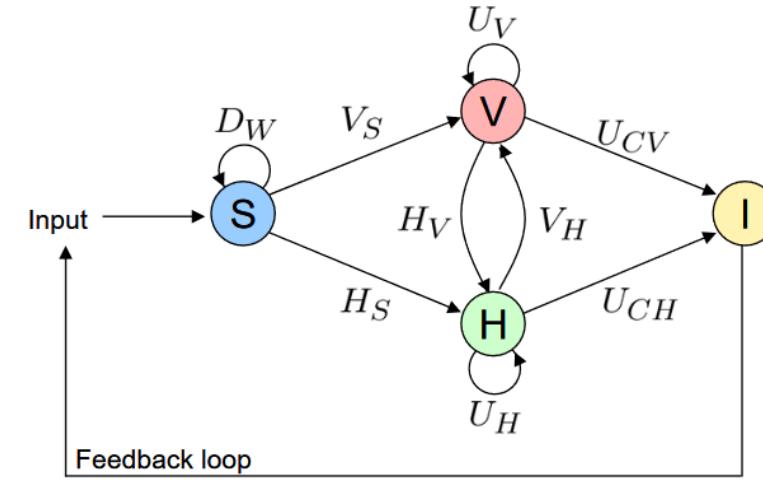
- Multiple views
 - Overview & detail
 - Focus & context
 - Small multiples
- Juxtaposition



“Multiple Coordinated Views at Large Displays for Multiple Users: Empirical Findings on User Behavior, Movements, and Distances”, Langner et al.

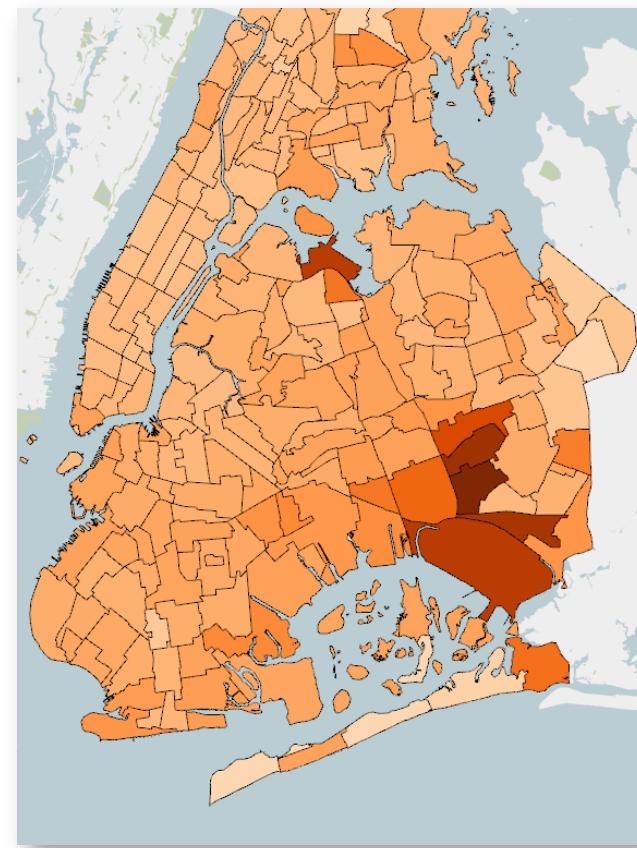
Week 10: Visual analytics systems

- Visual analytics process
- Analytical processes
 - Data mining
 - Stats
 - ML
- Visual analytics mantra



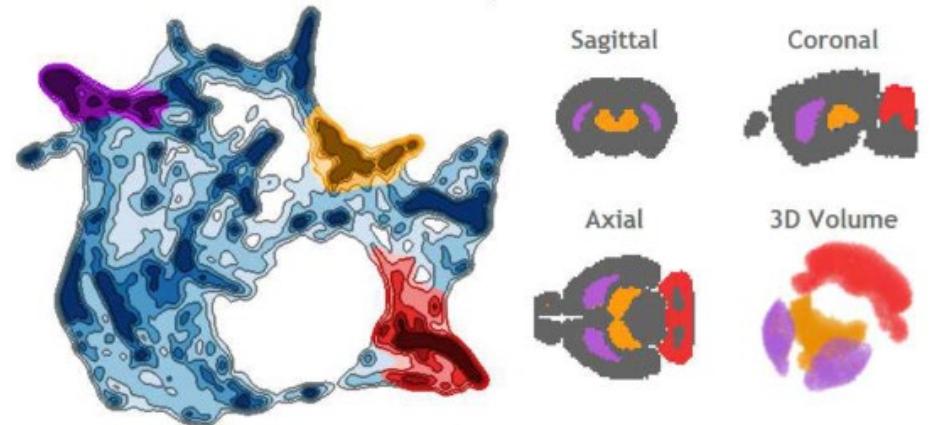
Week 12: Spatial and uncertainty vis

- Spatial data
 - Attributes
 - Data primitives
- Spatial queries
 - Selection
 - Join
 - Aggregation
- Spatial indices

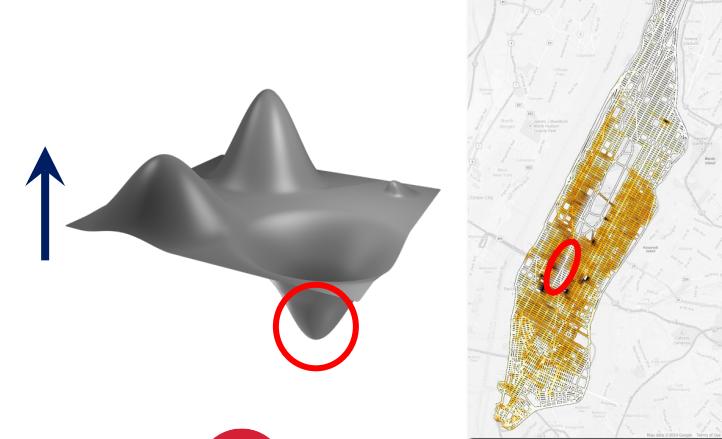


Week 13: ML & data mining for vis

- Machine learning:
 - Dimensionality reduction
 - Regression & classification for vis
- Data mining:
 - Topology-based techniques

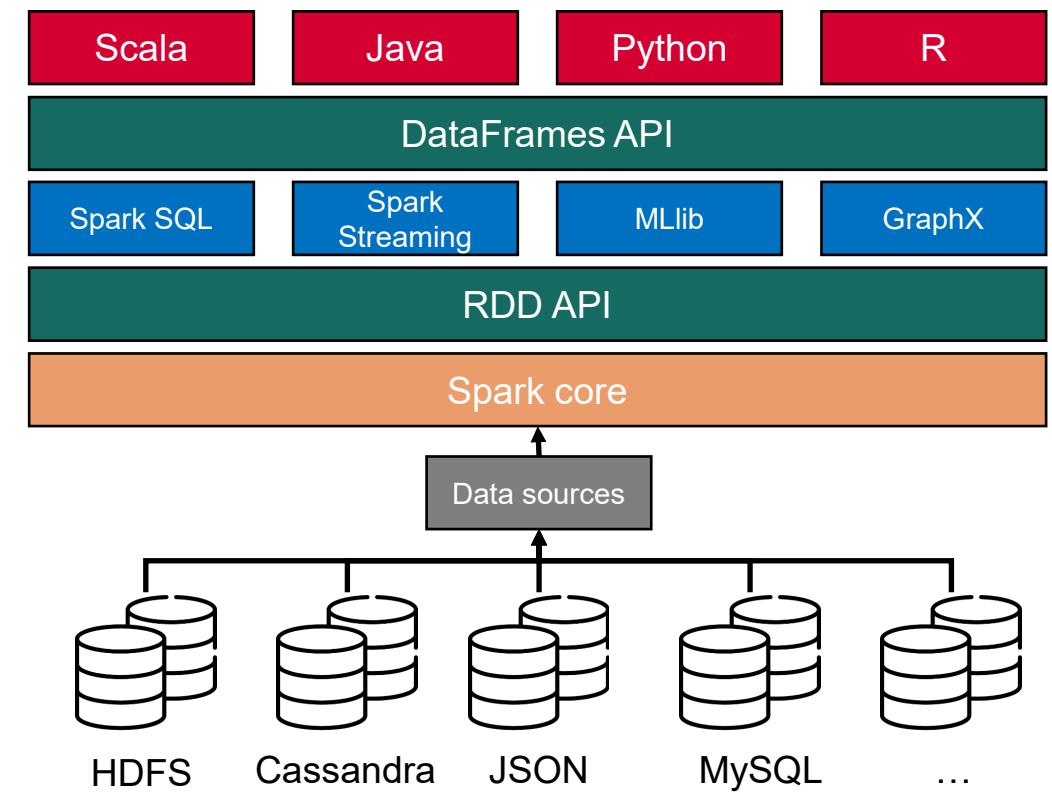


[Pezzotti et al., 2016]



Week 14: Vis for big data

- Large-scale data processing
- GPU approaches



Logistics

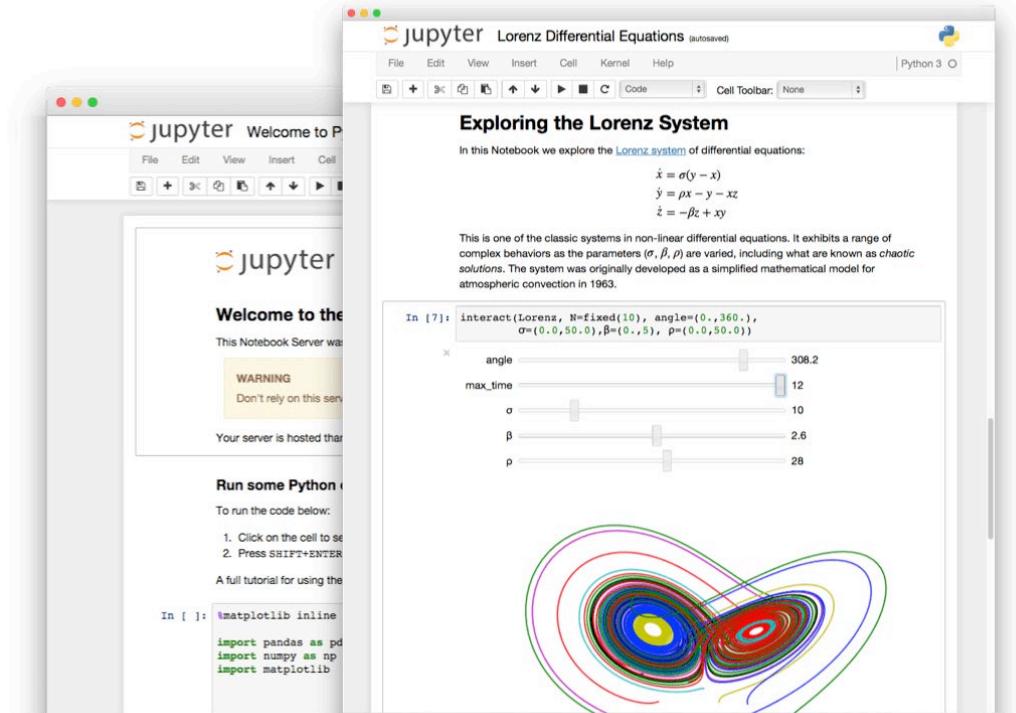
- Syllabus: <https://fmiranda.me/courses/cs424-fall-2022/>
- We will meet twice a week:
 - Tuesday 9:30am – 10:45pm (Central)
 - Thursday 9:30am – 10:45pm (Central)
- In-person: EVL's Continuum.
- Office hours:
 - Friday 11:00am – 1:00pm (Central).
 - Zoom (link on Discord).
- TA: Kazi Omar (komar3)

Grading policy

- Project 1: 20%
- Project 2: 25%
- Final project: 25%
- Vis critique presentation: 15%
- Participation: 15%

Project 1

- Exploratory data analysis
- Jupyter
- Week 3 (Sep 6), due week 6 (Sep 27) – three weeks!

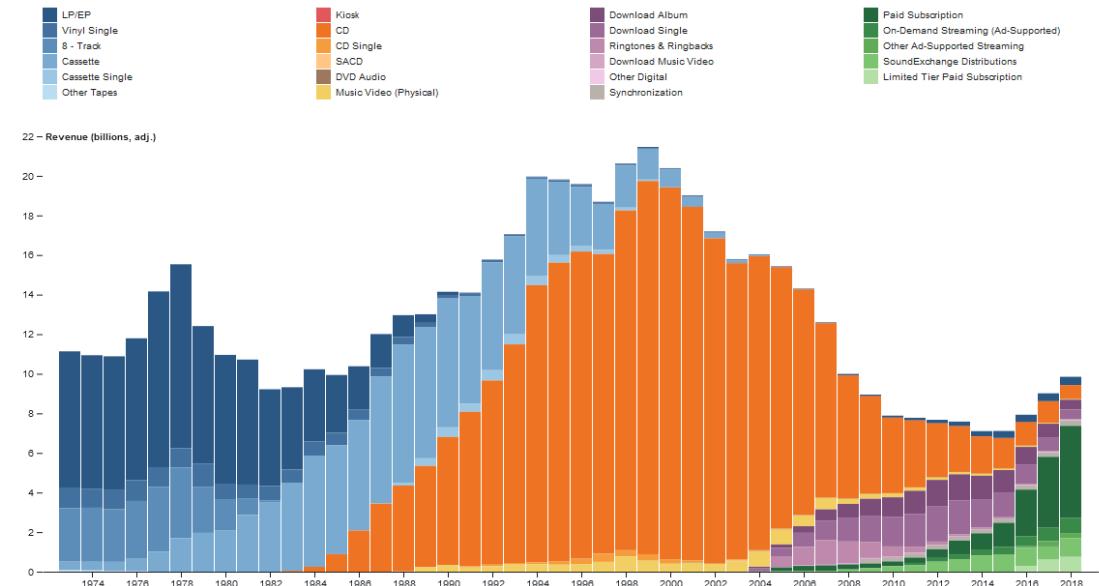


Project 2

- Coordinated views
- Observable
- Week 8 (Oct 11), due week 11 (Nov 1)

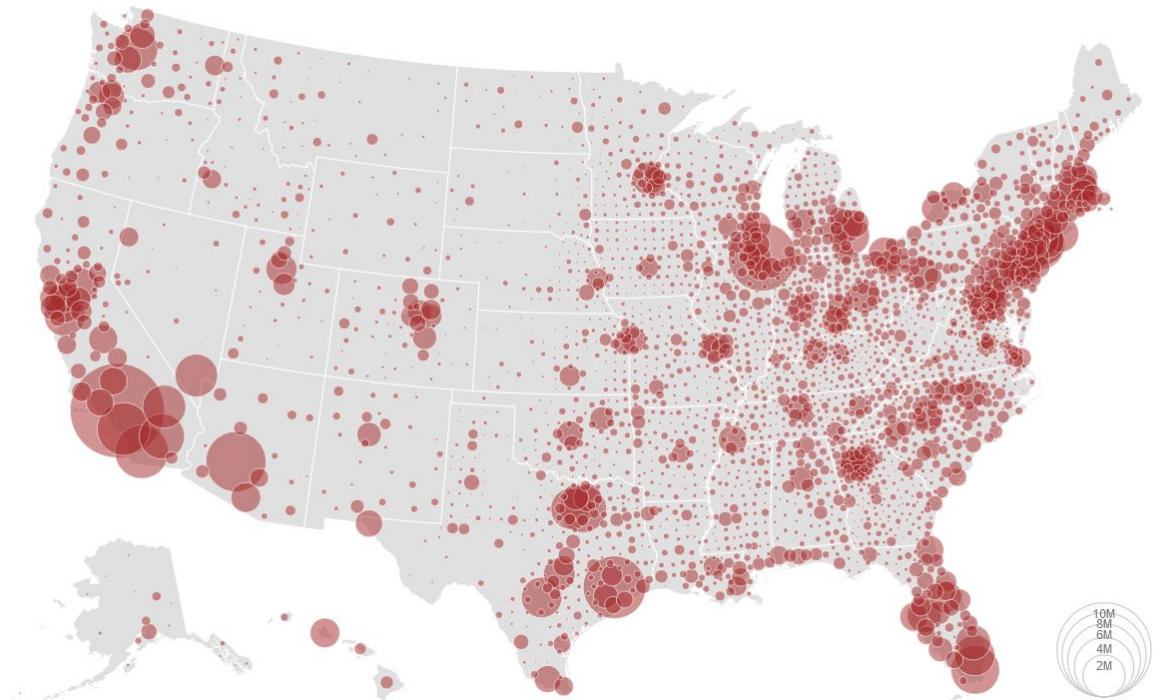
Revenue by Music Format, 1973–2018

Data: RIAA



Final project

- Spatial data visualization
- Javascript + React
- Week 12 (Nov 8), due week 15 (Nov 29)



Projects

- Detailed instructions regarding proposals and presentations will be posted following the course schedule.
- Teams of 2 students.
- Select your own dataset or select from a list of suggestions.
- Deliveries for each project:
 - ~8-minute presentation.
 - Github project, detailing implementation and findings.

Recommended workflow

Day 0: Read the instructions

Day 1-2: Familiarize yourself with the main topics of the assignment + think of potential findings

Day 3-14: Code the different components of the assignment

Day 15-19: Start writing README.md file

Day 19-20: Fix problems, double check README.md

Day 21: Submission and presentation

Vis critique

- Evaluate one visualization, considering how it maps data to visualizations.
- One presentation (week 9)

Final project

Week		Topic
1	Aug 23	Course introduction & Why vis?
2	Aug 30	Building blocks with Git, Pandas, GeoPandas, Jupyter
3	Sep 6	Introduction to visualization: Data & attribute types, channels, tasks, analysis loop
4	Sep 13	Exploratory data analysis + Lab
5	Sep 20	Building blocks: Web technologies for visualization
6	Sep 27	Project 1 presentation: exploratory data analysis
7	Oct 4	Data management for visualization
8	Oct 11	Embedded, juxtaposed and coordinated views + Lab
9	Oct 18	Vis critique (presentation) (Zoom)
10	Oct 25	Visual analytics systems: integrating visualization and analytics (Invited talks, Zoom)
11	Nov 1	Project 2 presentation: coordinated views
12	Nov 8	Spatial and uncertainty visualization
13	Nov 15	Machine learning & data mining for visualization + Lab
14	Nov 22	Visualization for complex and big data: OLAP cubes, MapReduce, Hadoop, Spark, Dask
15	Nov 29	Final project presentation: spatial data visualization

Lab sessions

- Weeks 4, 8 and 13.
- Opportunities to discuss shortcomings, findings, insights, and ask questions about the projects.
- Groups of ~6 students:
 - Introduce your project to the other members of the group.
 - Discuss potential findings.
 - Get their impressions about potential preliminary results.

Other considerations

- Discord will be the main platform for communication, please check Blackboard for the Discord invitation link.
- Github Classroom will be used for assignments and final project.
- Participation grade will take into consideration your contribution to a productive environment, either in the classroom, discord or office hours.