

Task abstraction

CS424: Visualization & Visual Analytics

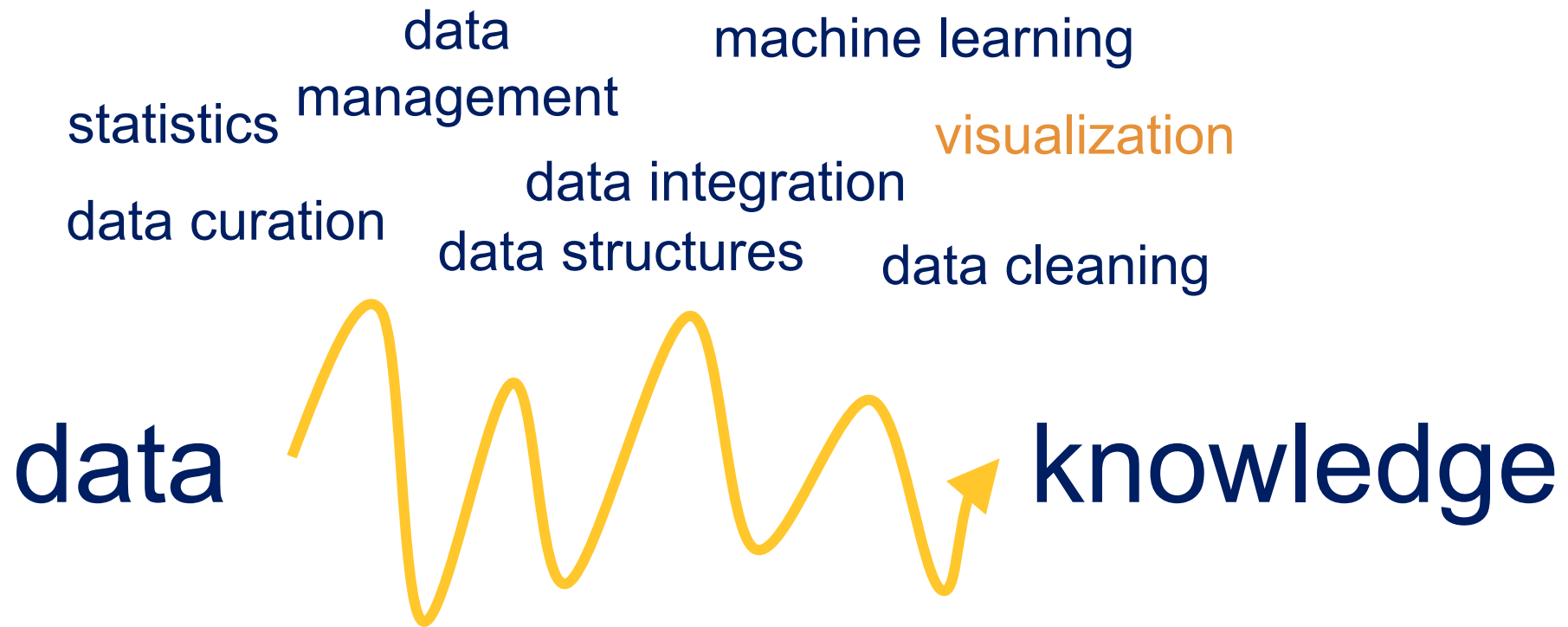
Fabio Miranda

<https://fmiranda.me>

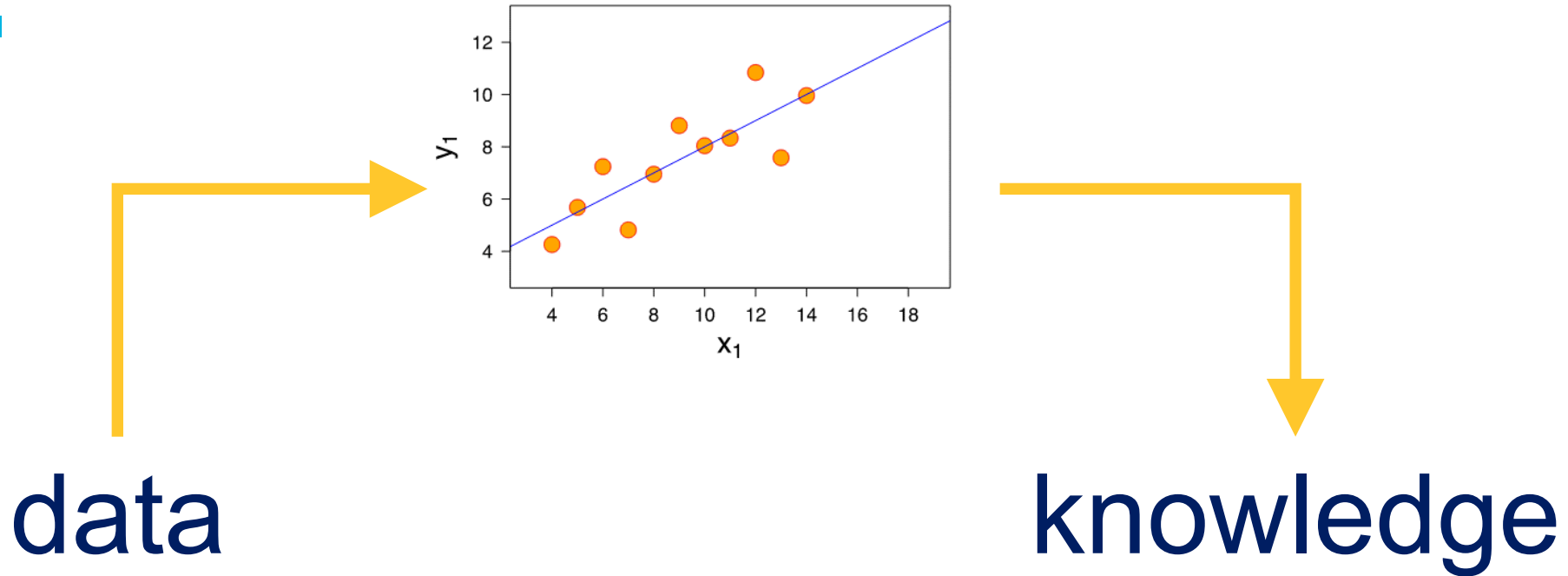
Data to knowledge

data → knowledge

Data to knowledge



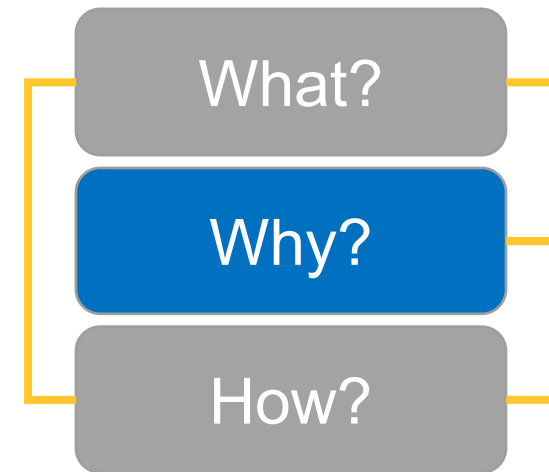
Data to knowledge



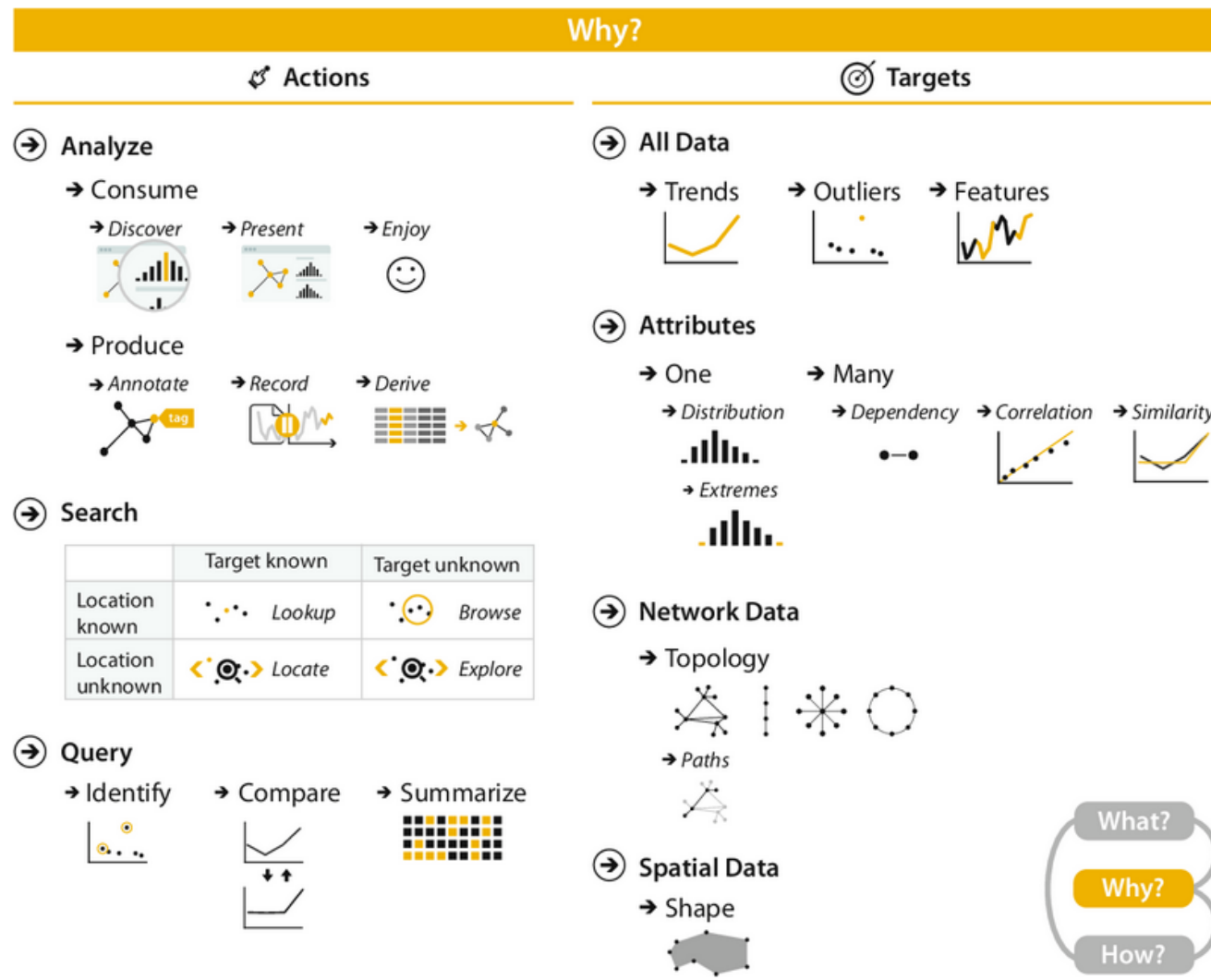
Transform data into visual marks

Task abstraction

- Analyzing tasks abstractly – rather than thinking of domain-specific tasks, think of abstract tasks.
- Domain-specific task: “contrast the prognosis of patients who were intubated in the ICU more than one month to patients hospitalized within the first week.”
- Abstract tasks: “compare values between two groups.”



Task abstraction



[Munzner, 2014]

Task = action + target

- Action verbs and target nouns

- Action verbs:

- Analyze
 - Consume
 - Produce
 - Search
 - Lookup, brose, locate, explore
 - Query
 - Identify, compare, summarize

- Target nouns:

- All data
 - Trends, outliers, features
 - Attributes
 - One
 - Distribution: extremes
 - Many
 - Dependency, correlation, similarity
 - Network data
 - Spatial data

Three levels of actions

- Analyze: high-level goals of the user
- Search: need to locate interesting items within the displayed data
- Query: different aspects of the extraction of information

Actions

➔ Analyze

➔ Consume

➔ Discover



➔ Present



➔ Enjoy



➔ Produce

➔ Annotate



➔ Record



➔ Derive



➔ Search

	Target known	Target unknown
Location known	Lookup	Browse
Location unknown	Locate	Explore

➔ Query

➔ Identify



➔ Compare



➔ Summarize



Three levels of actions: Analyze

- Analyze: consume information already generated or produce new material for some purpose.
- Consume:
 - Discover: visualization to find new knowledge that was not previously known.
 - Present: help user communicate something known to someone.
 - Enjoy: casual interaction with a visualization.
- Produce:
 - Annotate: add graphical or textual annotations to pre-existing visualization.
 - Record: persist visualization elements.
 - Derive: produce new data elements based on existing data elements (i.e., transformation).

Three levels of actions: Search

- Each analyze goal requires the user to search for items of interest within the visualization.

	Location known	Location unknown
Target known	Lookup	Locate
Target unknown	Browse	Explore

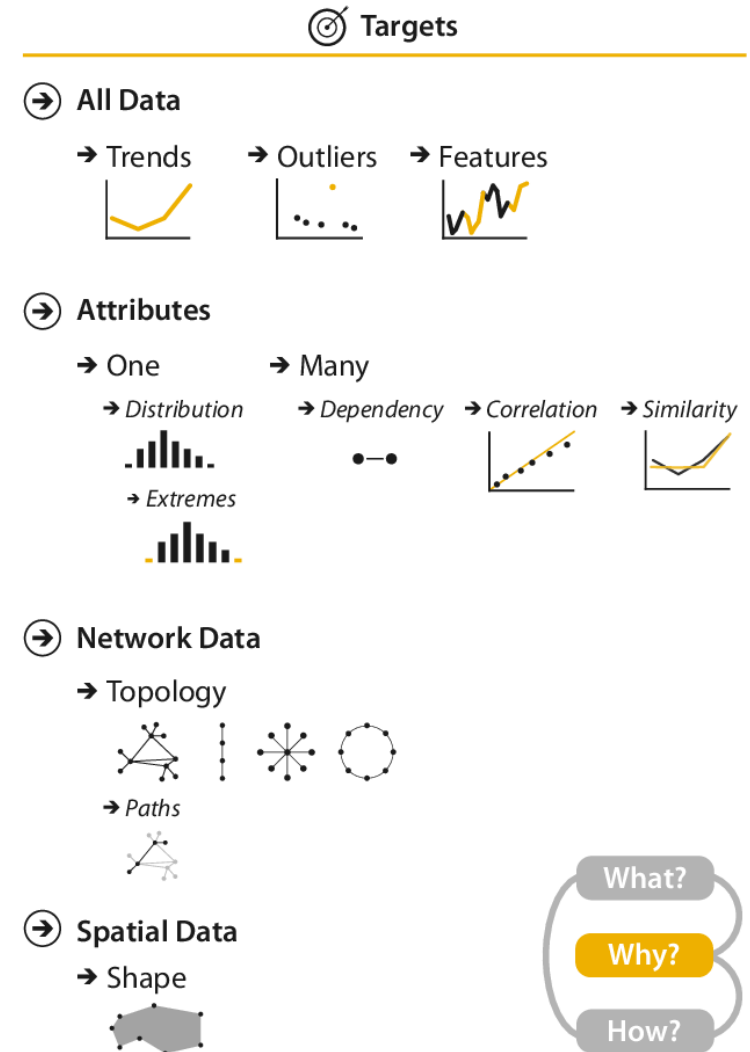
Target: some aspect of the data that is of interest to the user.

Three levels of actions: Query

- Once items of interest have been found, user likely wants to extract some information from the visualization.
- Identify: extract characteristics or references from a single target.
- Compare: compare multiple targets to each other.
- Summarize: provide a comprehensive view of all the data (i.e., overview).

Targets

- Targets are the aspects of the data of interest to the user.
- Targets that apply to all data:
 - Trends (or patterns)
 - Outliers
 - Features
- If data carries attributes:
 - Single attribute (e.g., distribution, extremes)
 - Many (dependencies, correlation, similarity)



Targets

- Targets are the aspects of the data of interest to the user.
- Targets that apply to all data:
 - Trends (or patterns)
 - Outliers
 - Features
- If data carries attributes:
 - Single attribute (e.g., distribution, extremes)
 - Many (dependencies, correlation, similarity)

Targets

➔ All Data

➔ Trends



➔ Outliers



➔ Features



➔ Attributes

➔ One

➔ Distribution



➔ Extremes



➔ Many

➔ Dependency



➔ Correlation

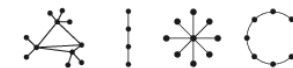


➔ Similarity



➔ Network Data

➔ Topology



➔ Paths



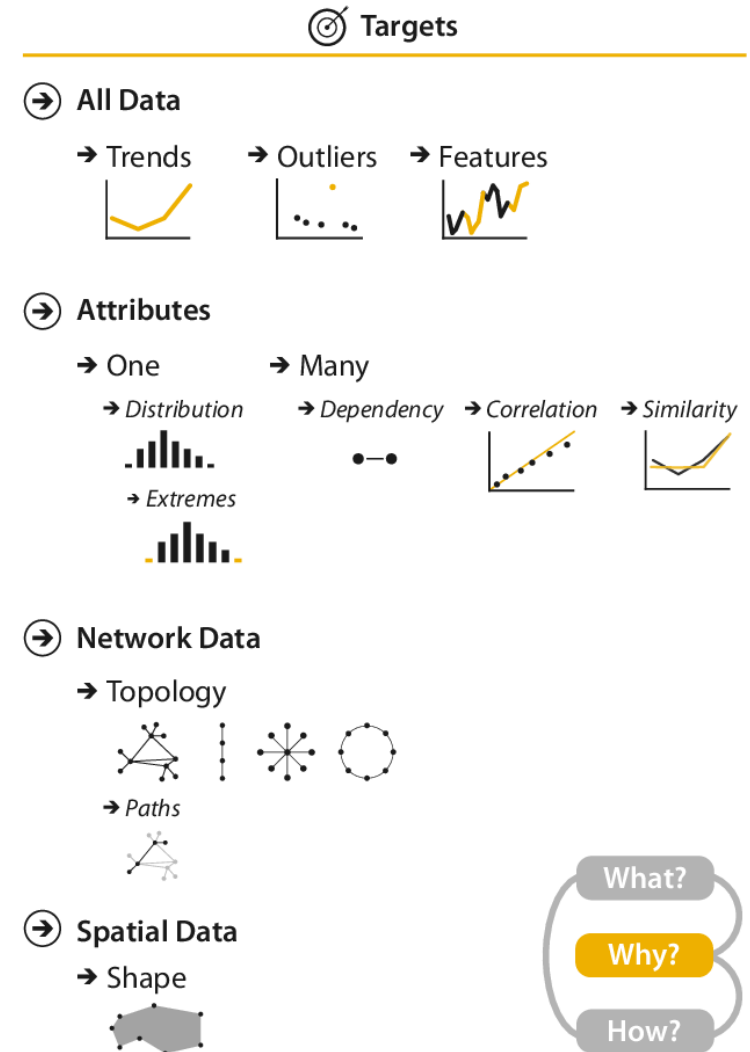
➔ Spatial Data

➔ Shape



Targets (cont.)

- Domain target types:
- Network data: user may be interested in studying aspects of the topology (connection patterns) or analyzing paths through the topology.
- Spatial data: user interaction with spatial data may be interesting in studying shapes.

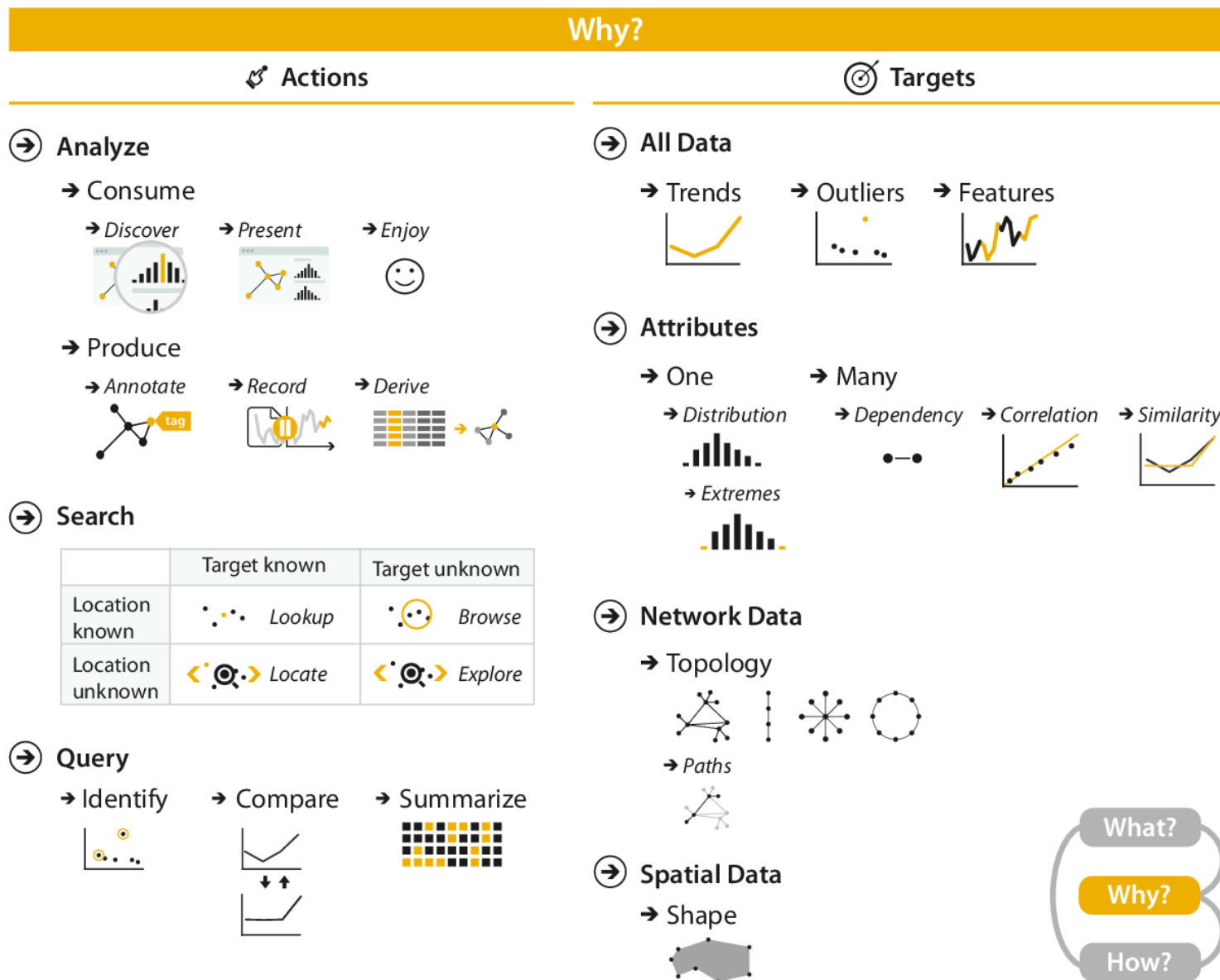


Task abstraction exercise

Visualization:

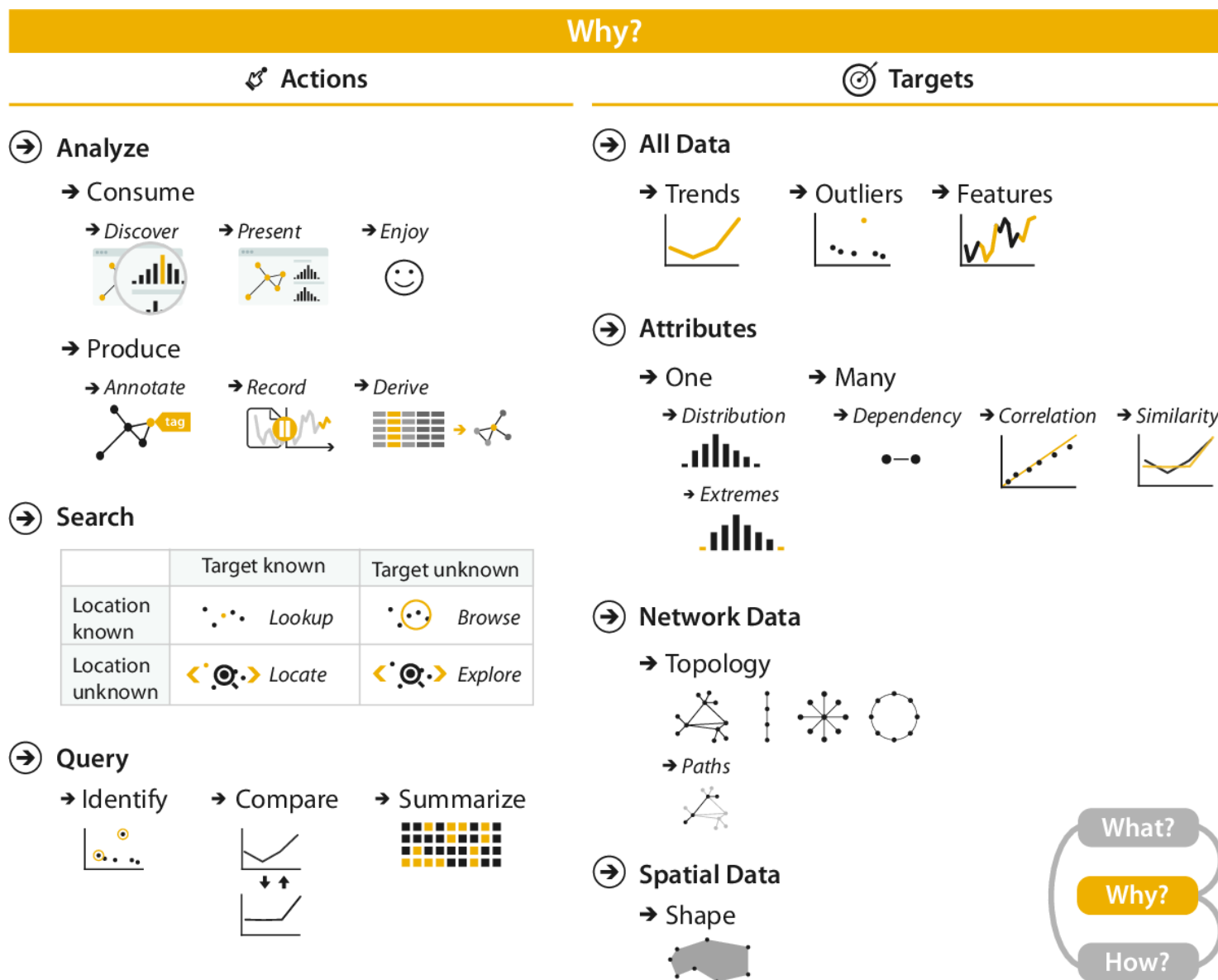
<https://new.mta.info/document/59281>

- Discuss 4 tasks that the map helps a user perform.
- Task: action + target



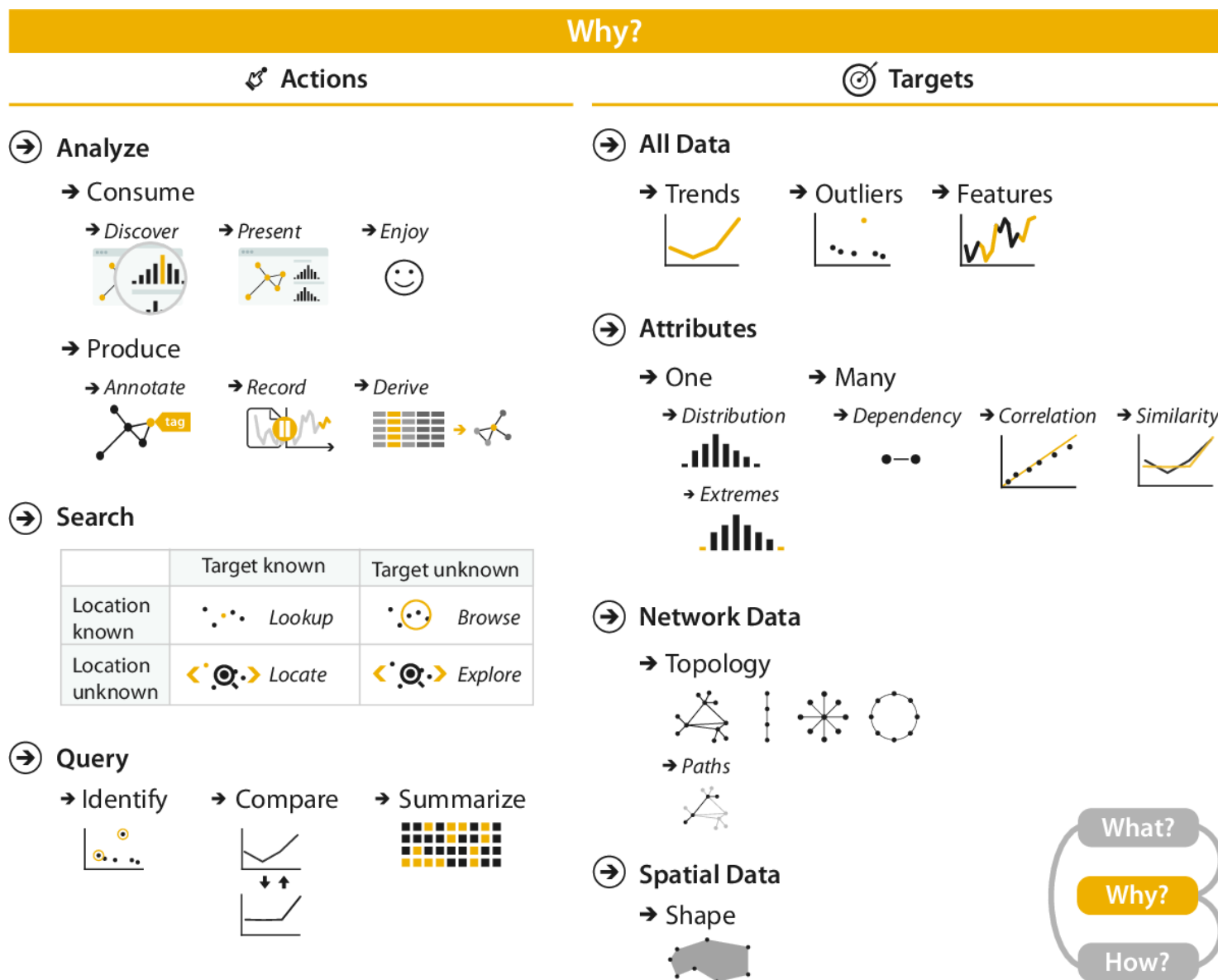
Task abstraction exercise

- “Figuring out how to get from A to B”
- “Explaining to someone how to get from A to B”
- “Finding the nearest station”

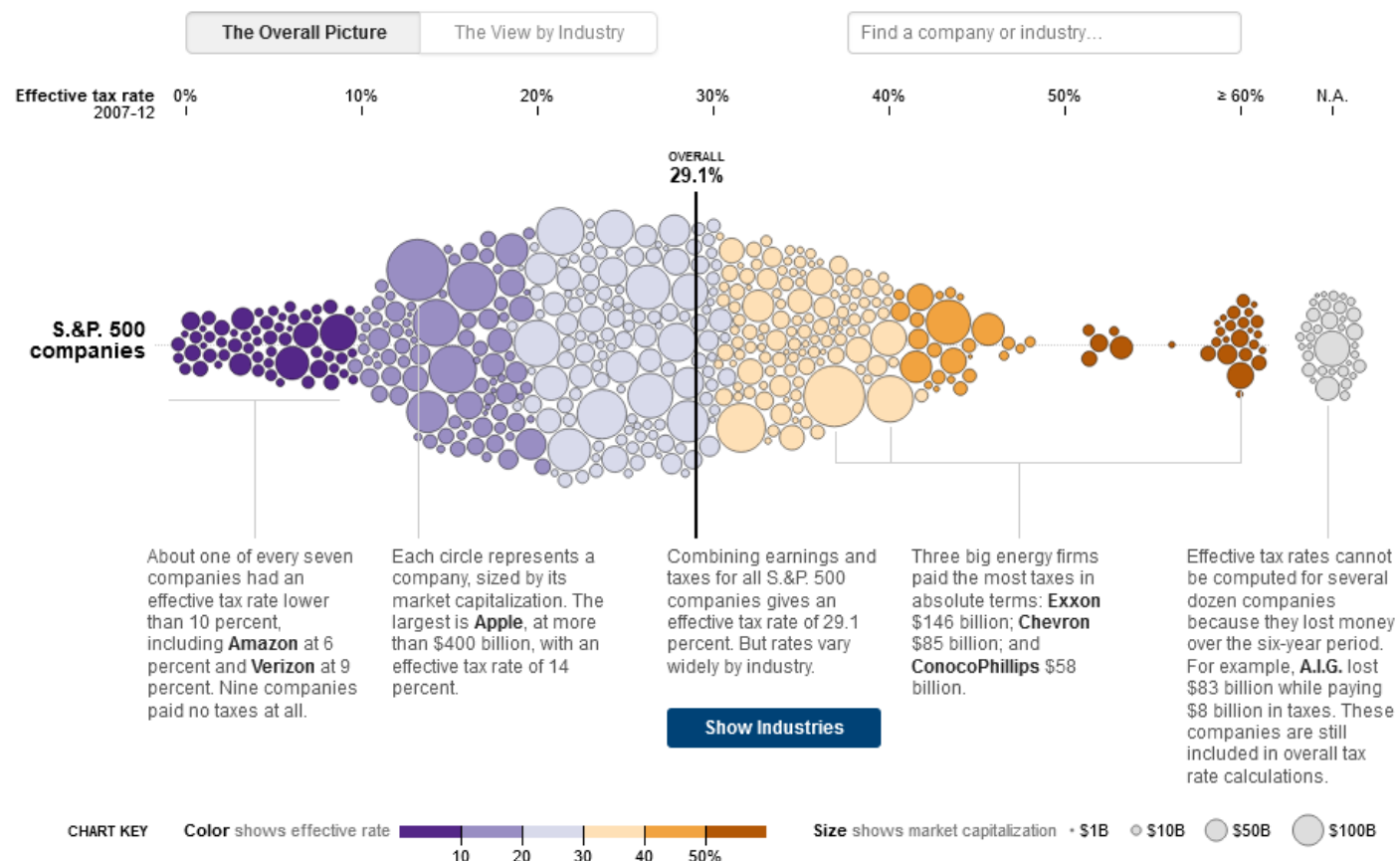


Task abstraction exercise

- “Figuring out how to get from A to B”
 - Action: Discover, locate path, identify
 - Target: Path (network data)
- “Explaining to someone how to get from A to B”
 - Action: Present
 - Target: Path (network data)
- “Finding the nearest station”
 - Action: Locate (target known, location unknown)
 - Target: Single attribute

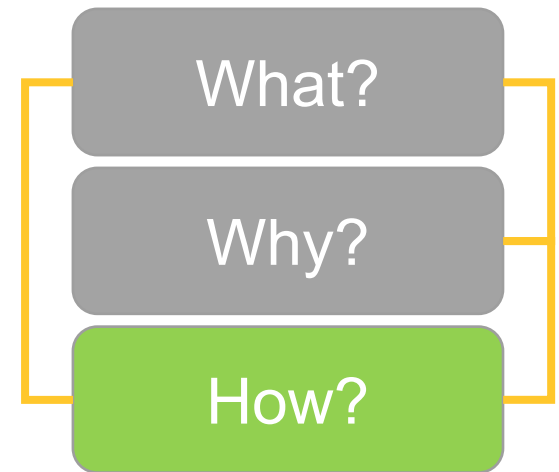


Task abstraction exercise



How to design vis idioms

- How a vis idiom can be constructed out of a set of design choices?
 - Encode
 - Manipulate: change, select, navigate
 - Facet: coordinate multiple views
 - Reduce: filter, aggregate



Big data example



Distribution of NYC Taxi
Pickups and Dropoffs in
Midtown Manhattan

Big data example

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_f	PULocationID	DOLocationID	payment	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement	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 inside certain blocks of Manhattan.
- Merge data with other data:
 - Traffic accidents within 100 meters and 1 hour of pickup and dropoff.
- Aggregate the data:
 - Number of pickups in each hour.
 - Number of pickups in each day of the week..

Data transformation

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fPULocationID	DOLocationID	payment_fare_amo	extra	mta_tax	tip_amo	tolls_amo	improvement	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	0.3	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	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	0.3	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	0.3	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	0.3	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	0.3	8.75

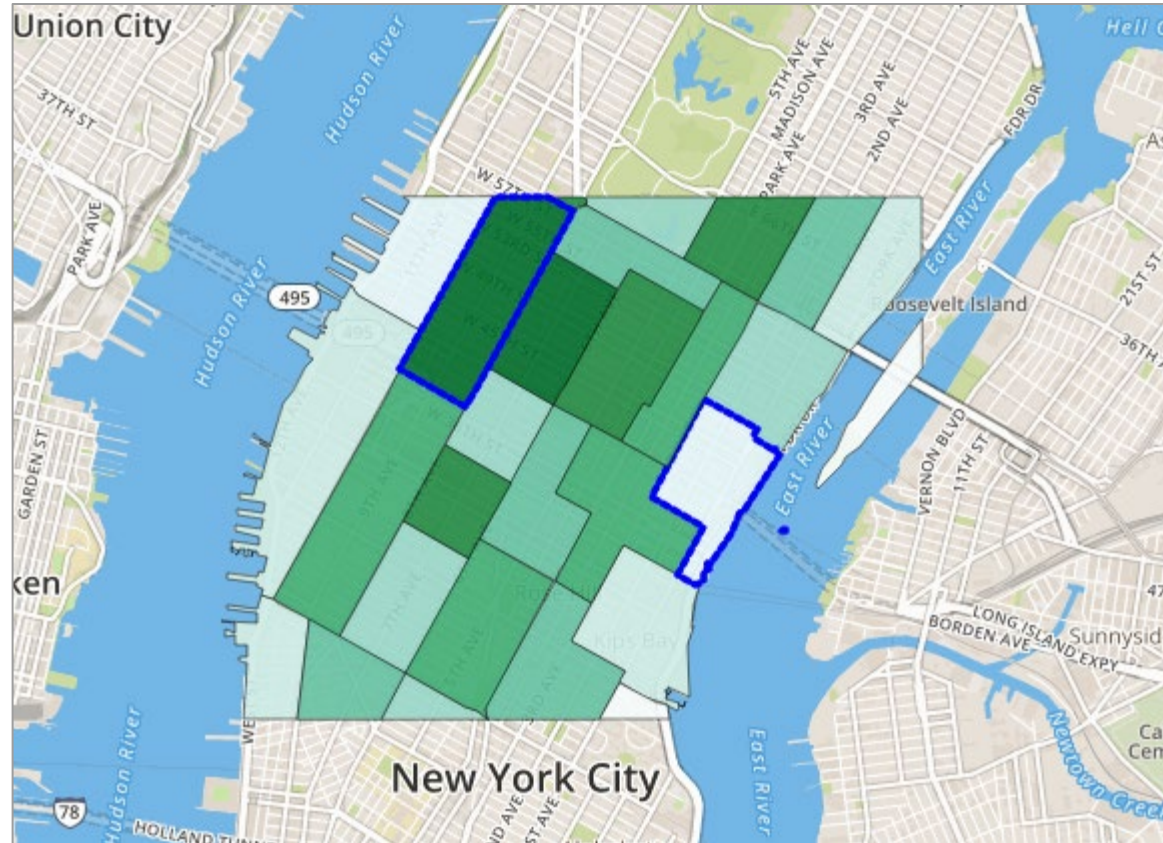


Data transformation

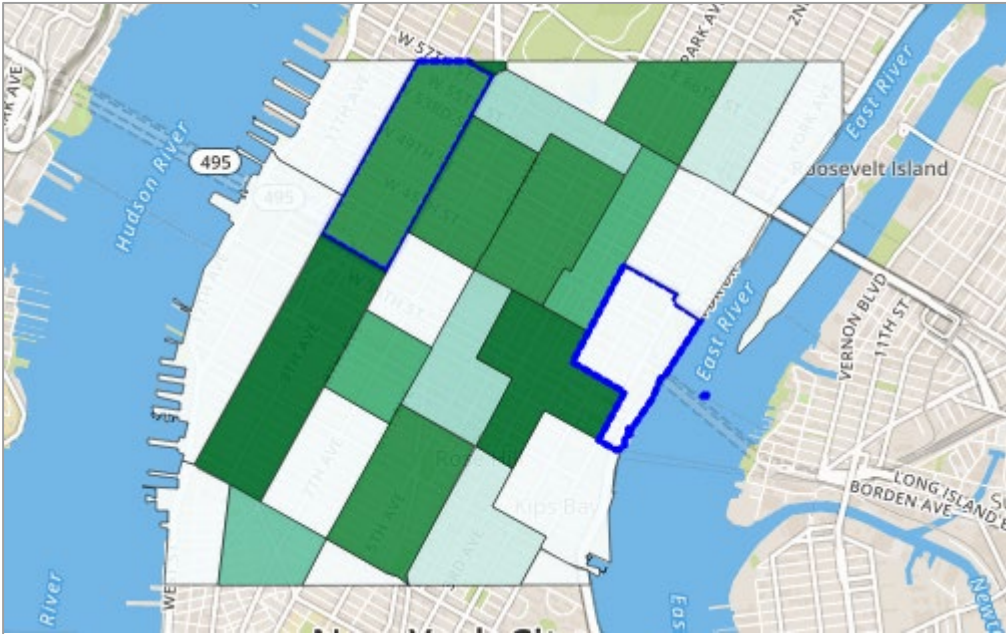
VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_f	PULocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement	total_amount	Area
1	1/1/2018 0:21	1/1/2018 0:24	1	0.5	1	N	41	24	2	4.5	0.5	0.5	0	0	0.3	5.3	Midtown
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	33.5	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

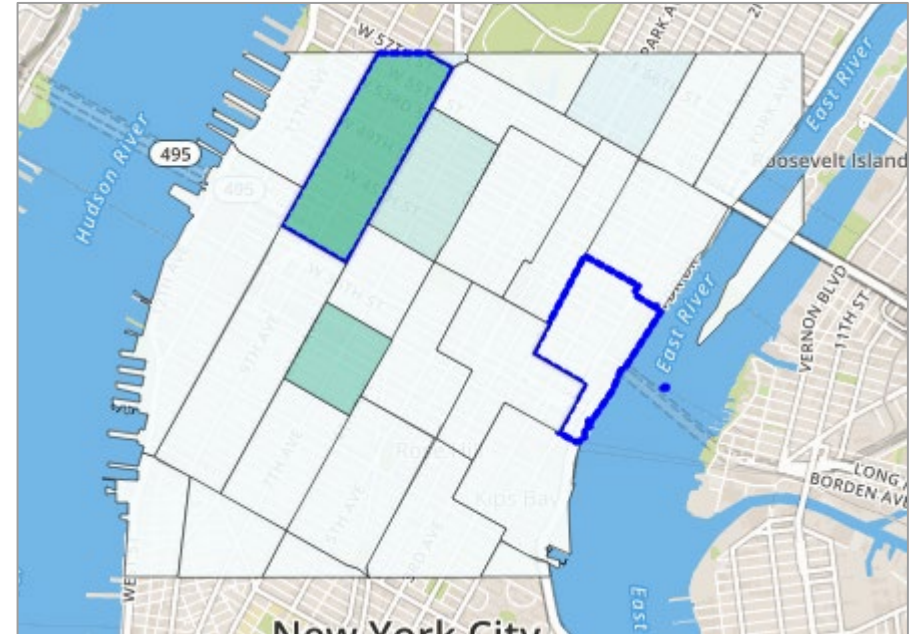
Quantitative data
Mark: polygon areas
Channel: color



Visual interaction



12pm – 2pm pickups



6am pickups

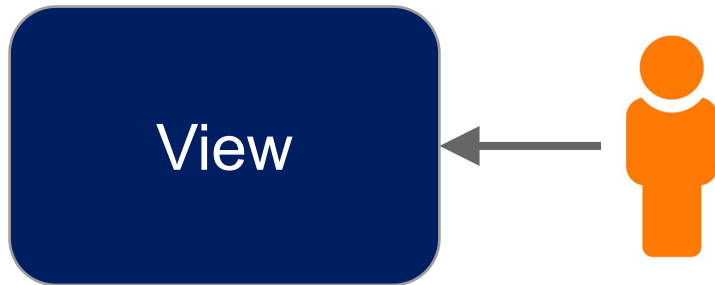
Interaction



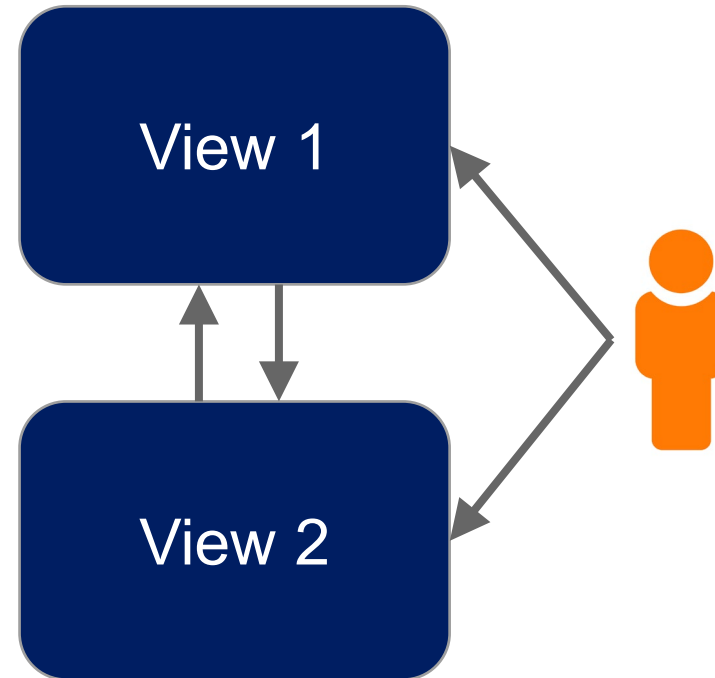
- Interaction can be used to manipulate:
 - Data
 - Visual mapping
 - View
- Why manipulate visualizations?
 - Often not possible to visualize all the information needed to answer all questions in one single static view.
 - Interaction permits to adapt / change the visualization so that it's possible (or easier) to answer multiple questions.
 - Especially useful when visualization is used as a general-purpose application for data analysis and exploration.

Interaction

Single view



Multiple views



[Bertini, 2020]

Single view interactions

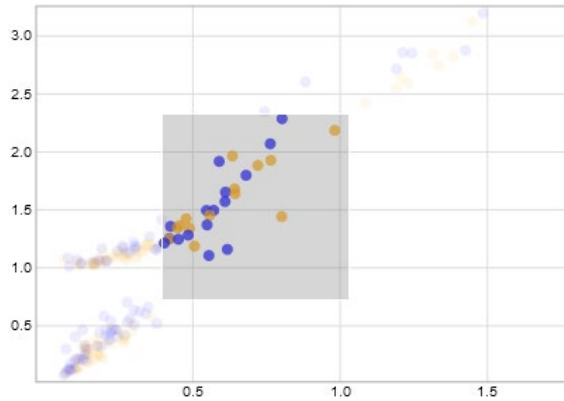
Manipulate	Methods
View	Selection Navigation Spatial arrangement
Mapping	Change mapping
Data	Aggregation Filtering

View interaction methods

- Selection: any action aimed at selecting one or more elements of the visualization.
 - Click → highlight (change color and/or borders, grey out the rest, etc.)
 - Hover → show more info (labels, info in linked view, etc.)
 - Click + drag → apply operation
- Navigation: changing the level of details and moving the viewport.
- Spatial arrangement: changing the way elements of the visualization are arranged / ordered.
 - Reordering → make visual patterns apparent.

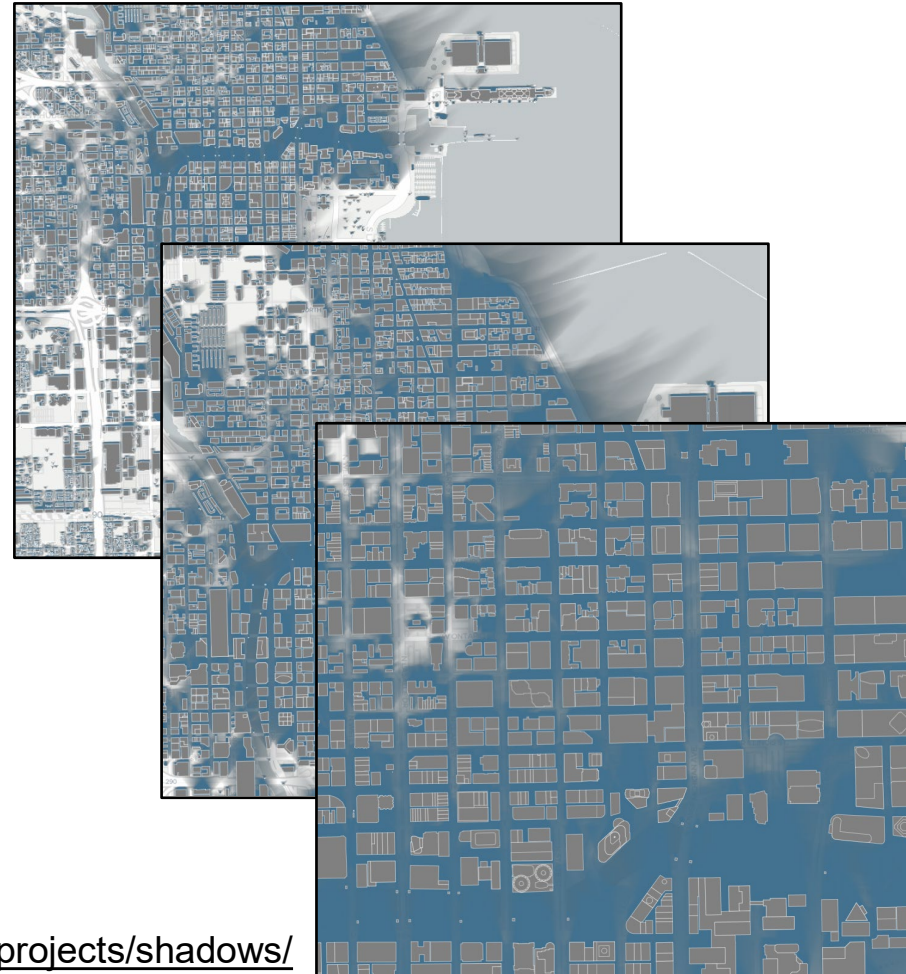
View interaction methods

Selection



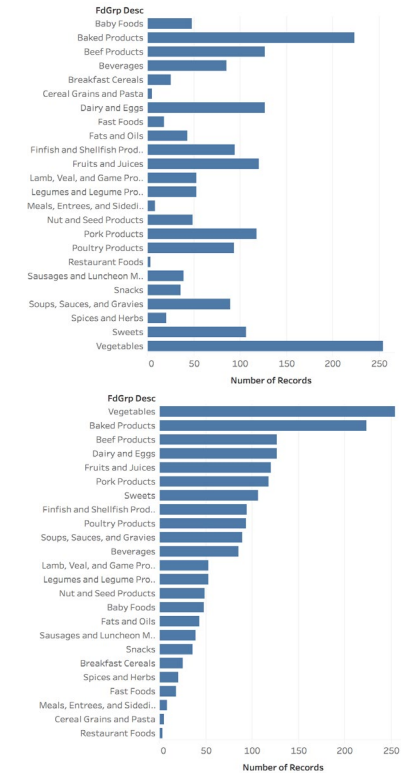
<https://vgc.poly.edu/projects/urban-pulse/>

Navigation



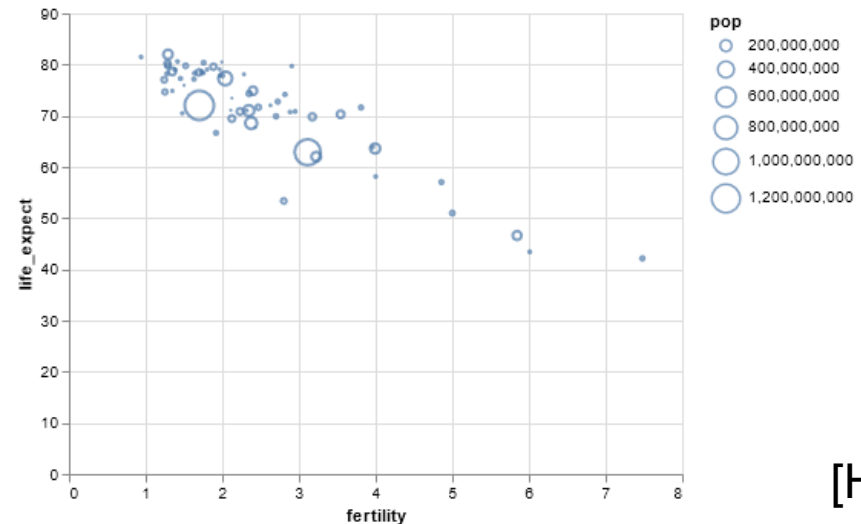
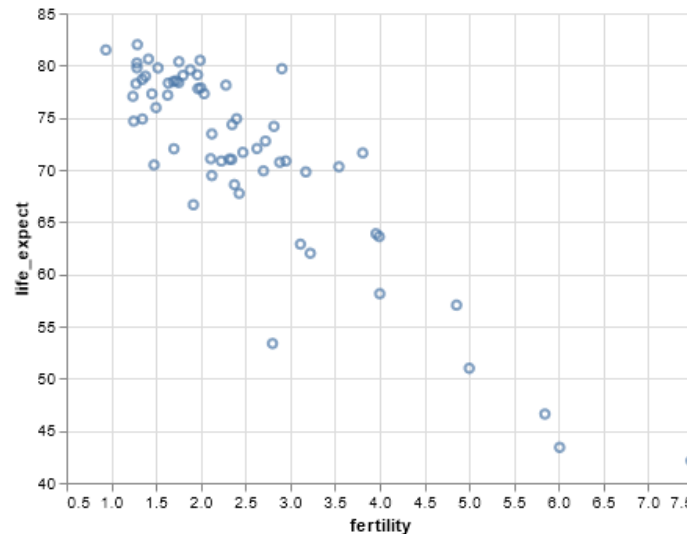
<https://vgc.poly.edu/projects/shadows/>

Spatial arrangement



Mapping interaction method

- Change mapping: changing the way attributes are encoded with visual channels.
 - Completely different plot or changes in properties of a given plot.



[Heer, 2020]

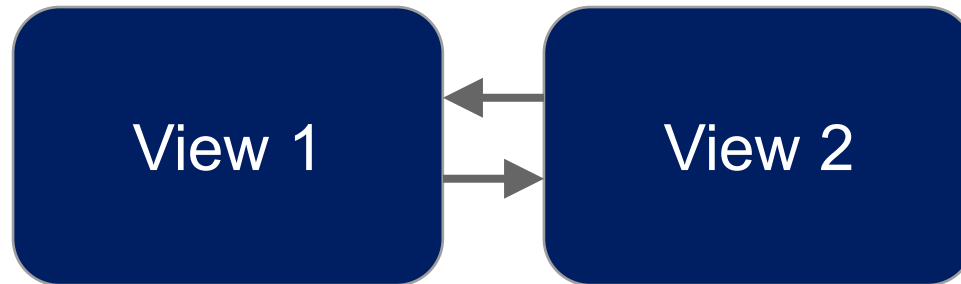
Data interaction methods



- Aggregation: changing the level of granularity of a given data set.
 - Space and time are hierarchical and often require observing patterns at different resolutions.
- Filtering: filtering data interactively according to some criteria or constraints.

Multiple linked views

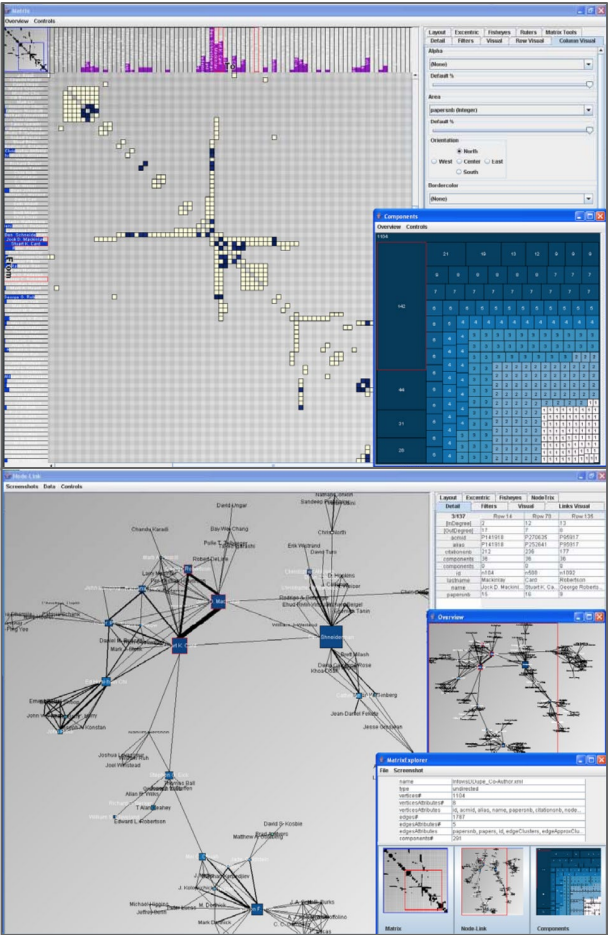
- Why multiple linked views?
 - Show different properties of the same data simultaneously.
 - Use one view to navigate, select, filter information in the other view.



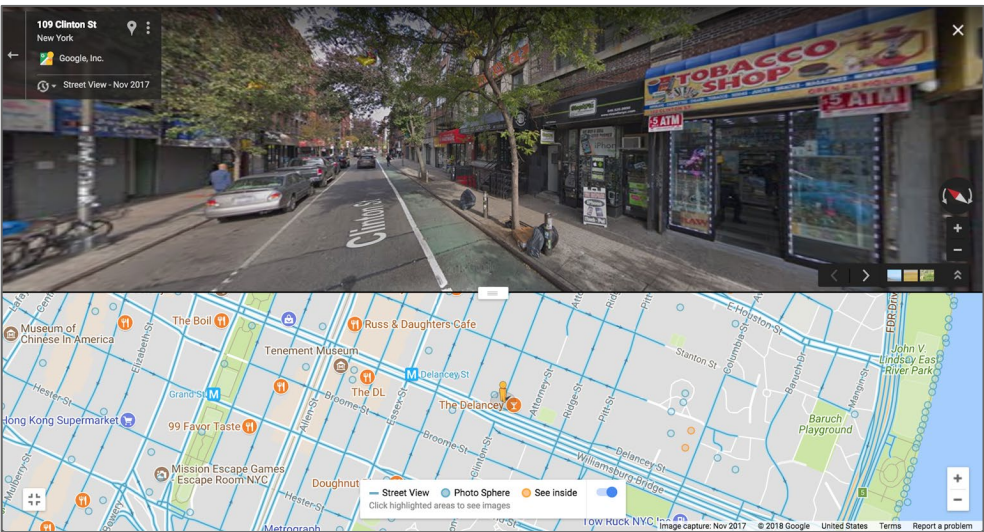
How to show different properties?

- Different information
 - Subset of data
 - Different attributes
 - Different granularity
 - Transformation
- Different representation

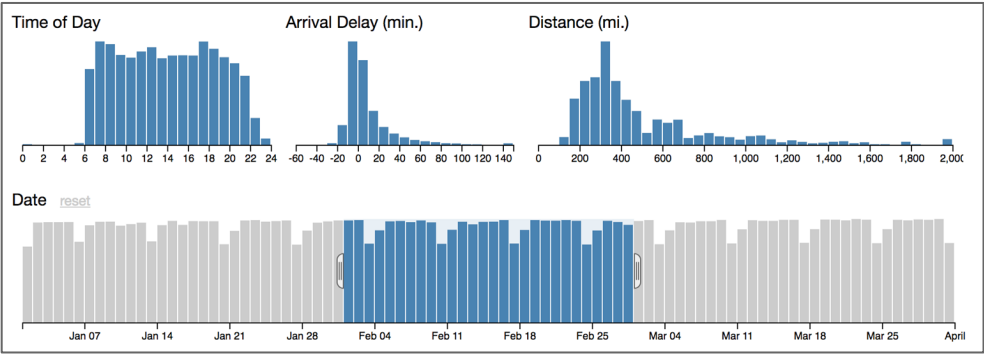
Multiple linked views



Same information,
different representation
[Riche, 2006]

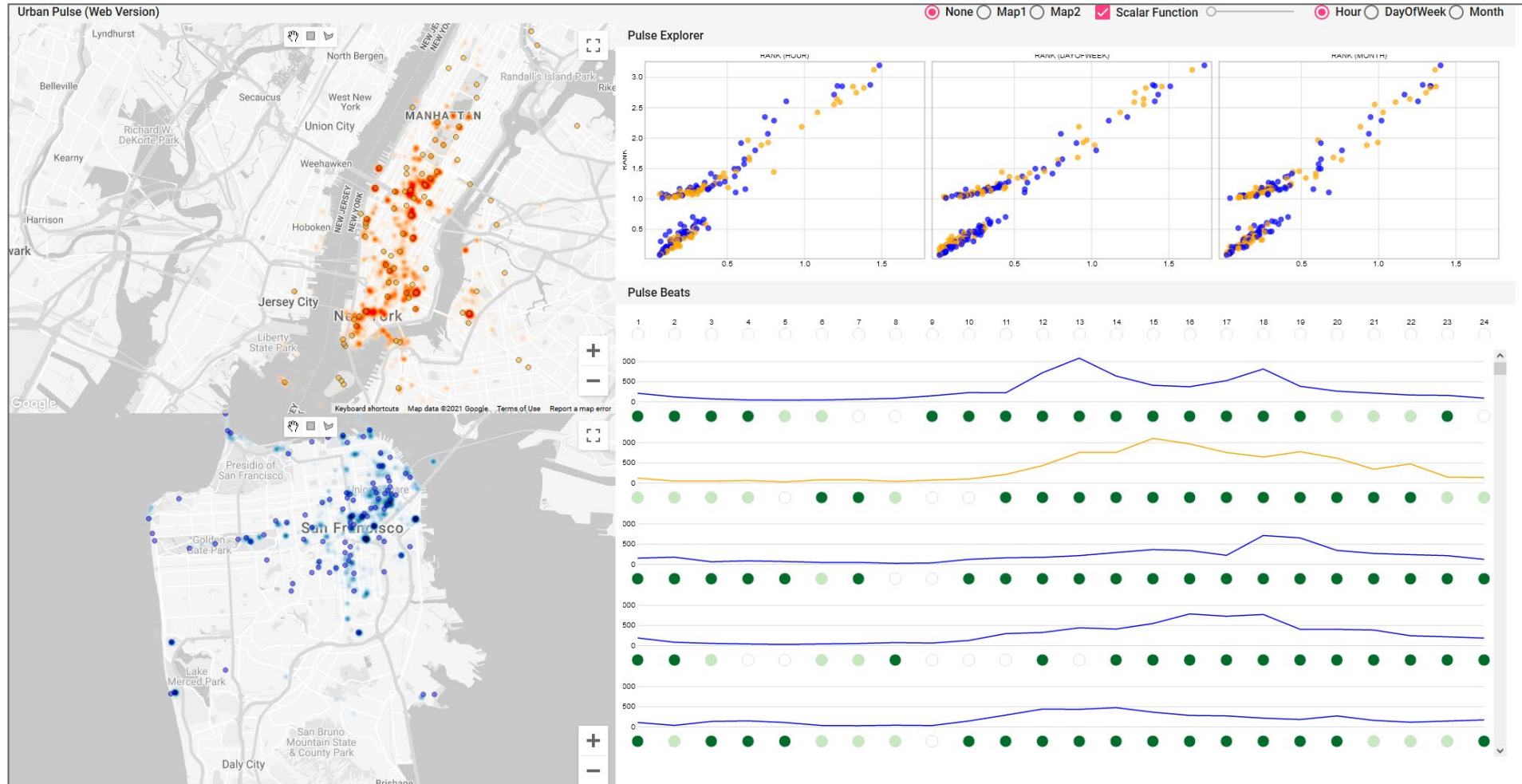


Different information & representation



Different information,
same representation

Multiple linked views

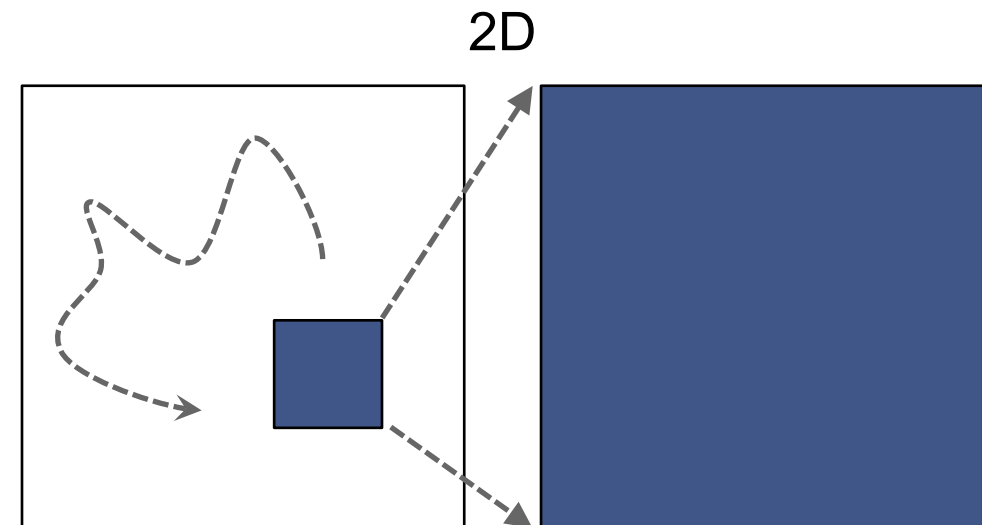
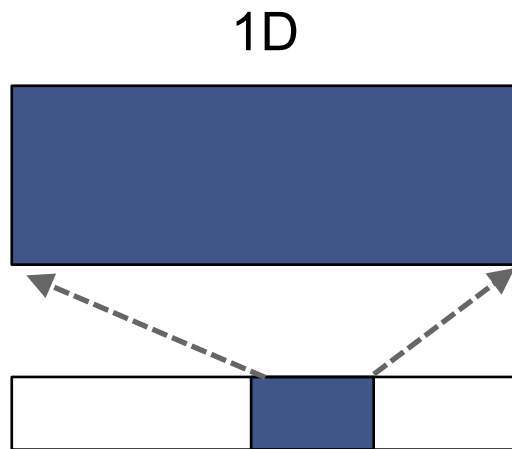


<http://vgc.poly.edu/projects/urban-pulse/>

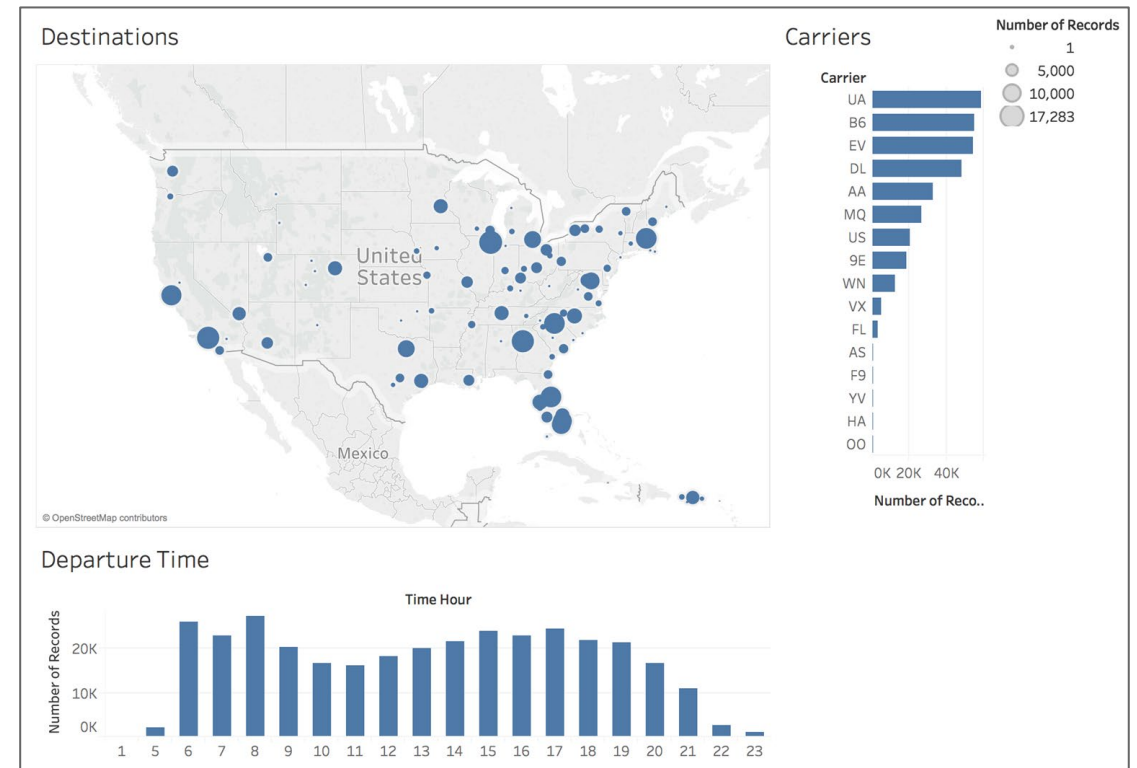
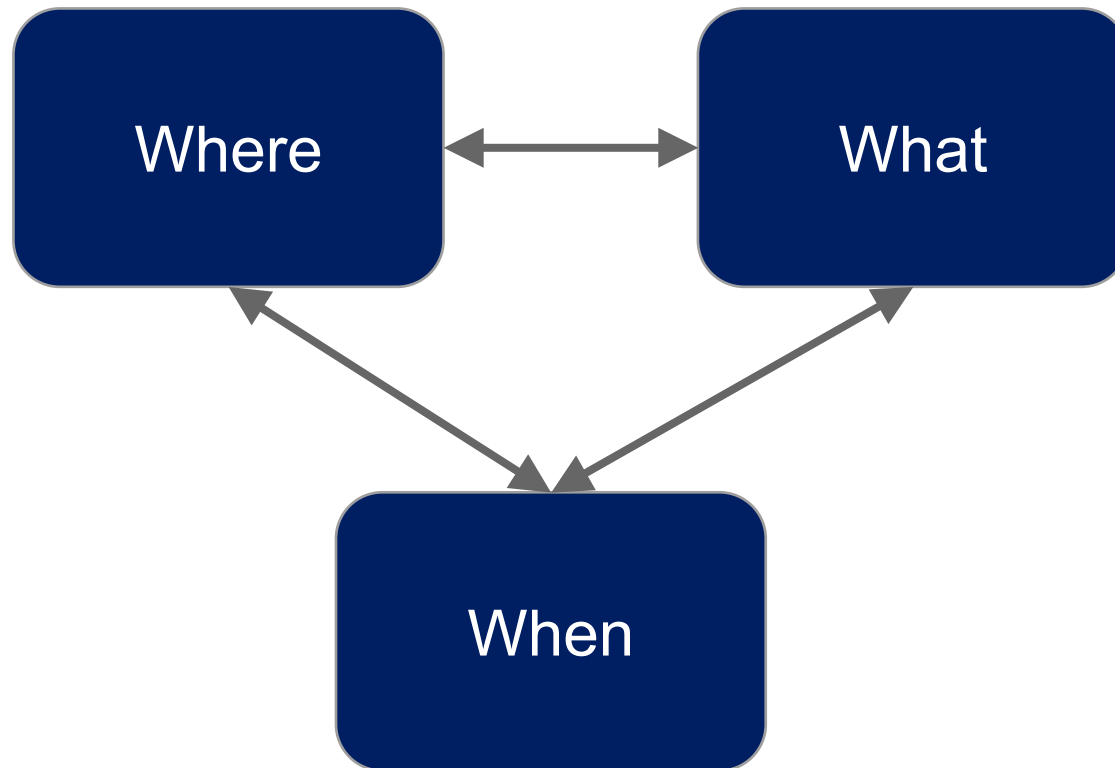
Overview + detail

- Visualization mantra:
“Overview first, zoom and filter, then details on demand”

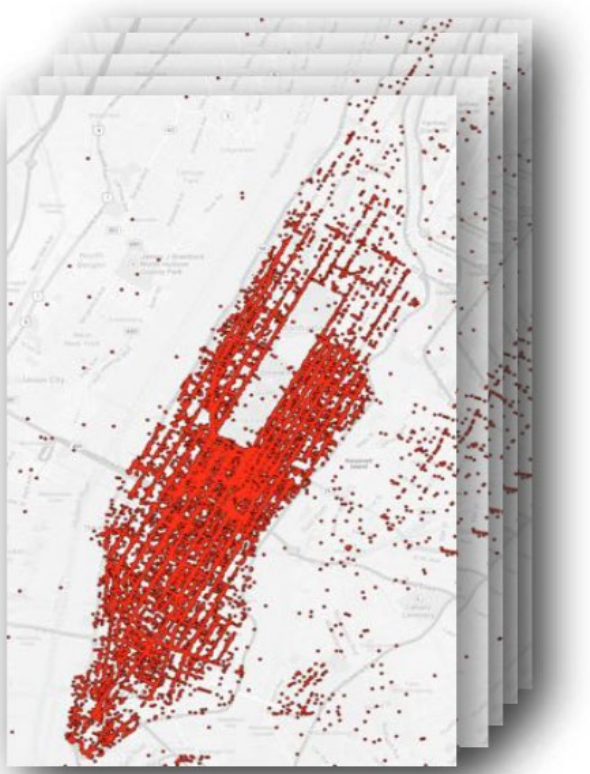
[Shneiderman, 1996]



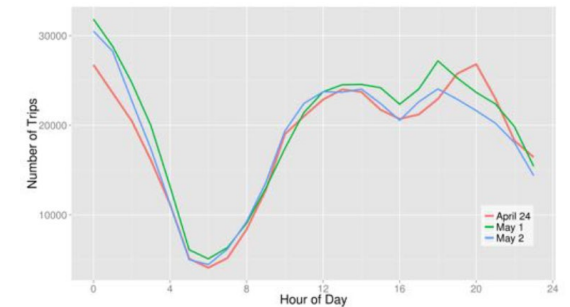
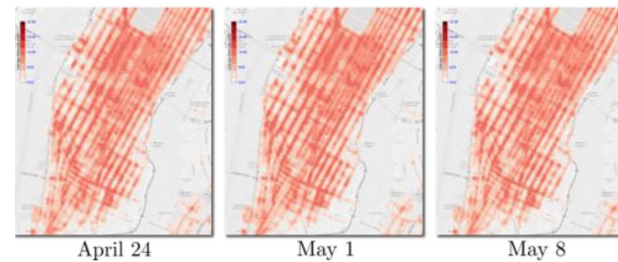
Where, what, when



Big data challenges

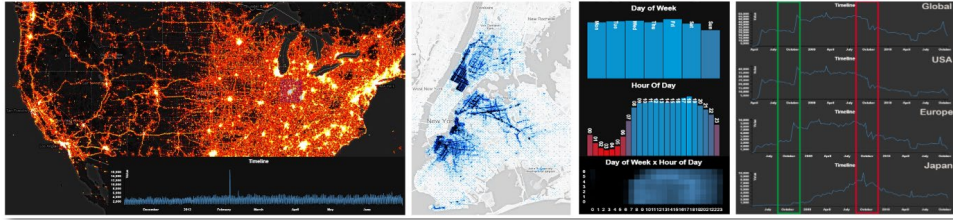


- 365×24 1-hour slices in one year.
- Which slides are interesting?

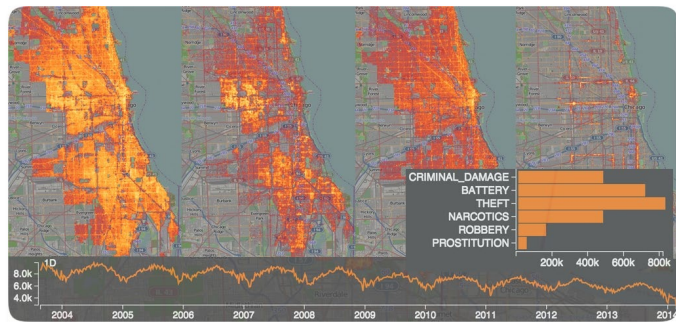


Accelerating data interaction methods

OLAP queries



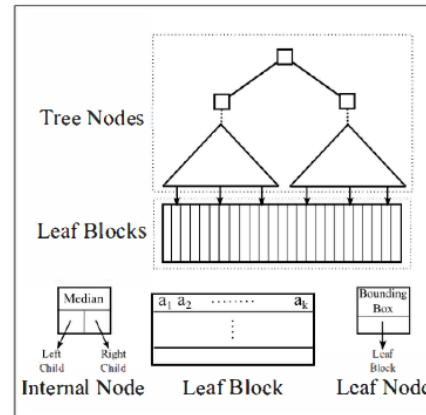
Hashedcubes [Pahins et al., 2017]



Nanocube [Lins et al., 2013]

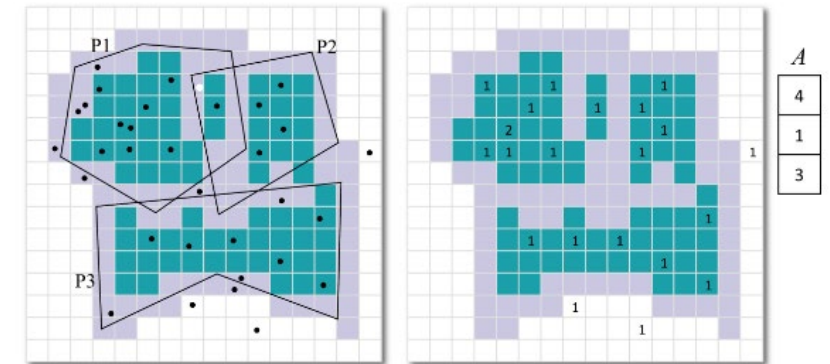
TopKube [Miranda et al., 2018]

Selection



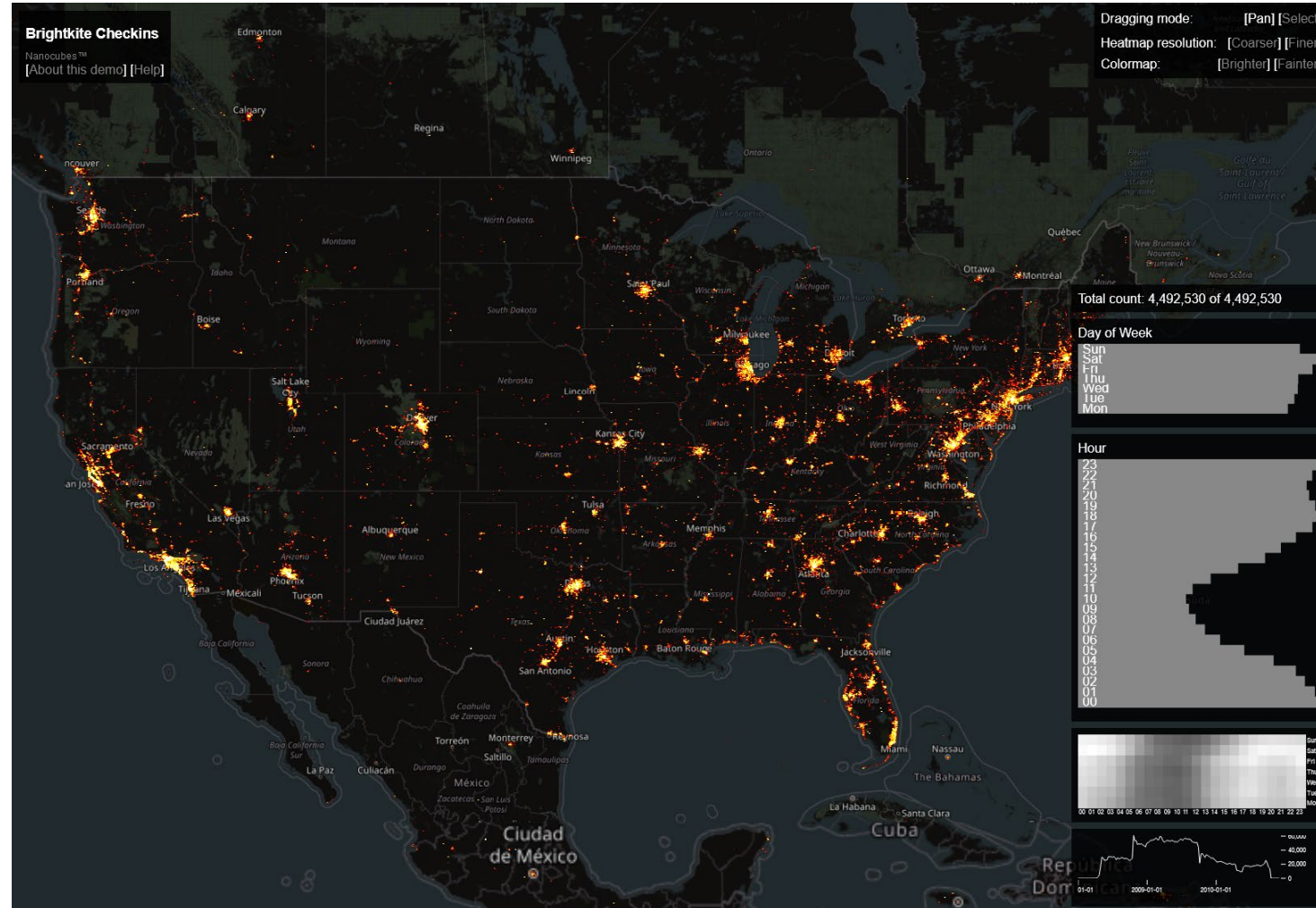
STIG [Doraiswamy et al., 2015]

Spatiotemporal joins



Raster join [Tzirita Zacharitou, Doraiswamy et al., 2018]

Interactive aggregations



[Lins, 2012]

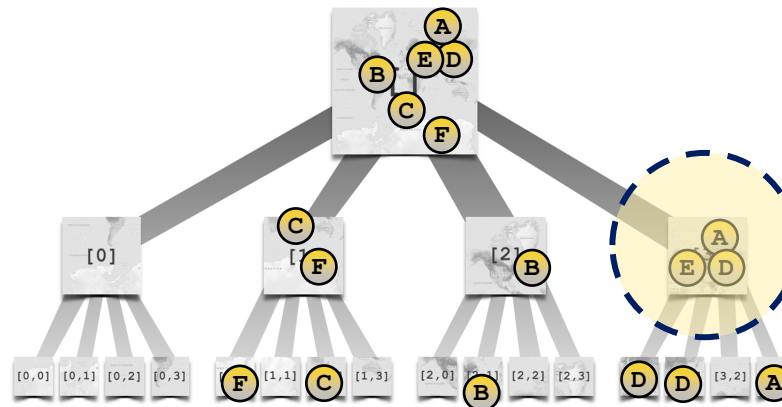
Datacube model

Following datacube model, aggregate every record along a hierarchy of bins.

The data structure is a mapping of bins to a pre-computed summary (e.g., count, timeseries).

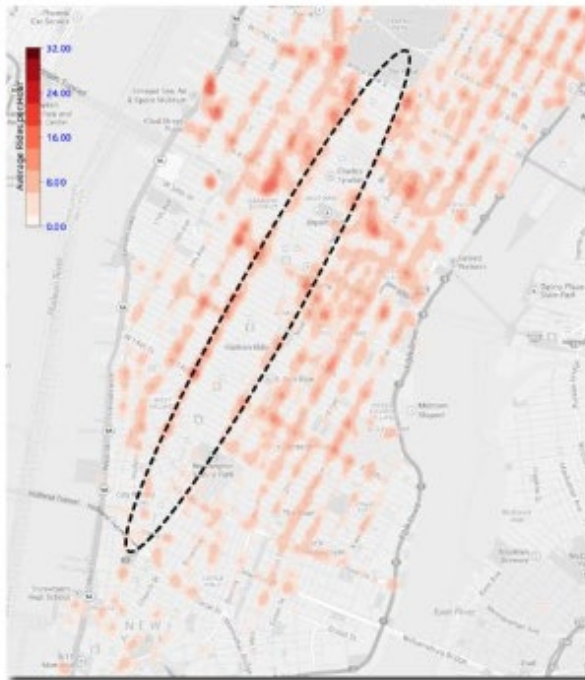
latitude	longitude	keyword
42.102908	-73.242852	#phoenix
29.617161	-81.636398	#phoenix
23.014051	75.120052	#la
26.014051	75.120052	#nyc
28.014051	74.120052	#la
23.014051	75.120052	#phoenix

	latitude	longitude
A	42.102908	-73.242852
B	29.617161	-81.636398
C	23.014051	75.120052
D	26.014051	75.120052
E	28.014051	74.120052
F	29.61161	-81.63638



K	c	p
0	10	1
1	22	2
2	15	0

Missing interesting slices



May 1 (8-9am)

- Data management: ensures operations are performed interactively.
- Analytics: points to interesting patterns or features of the data.