Introduction to Visualization

CS594: Big Data Visualization & Analytics

Fabio Miranda

https://fmiranda.me

UIC COMPUTER SCIENCE

Data to knowledge

data → knowledge

Data to **knowledge**

data management

machine learning

statistics

visualization

data integration

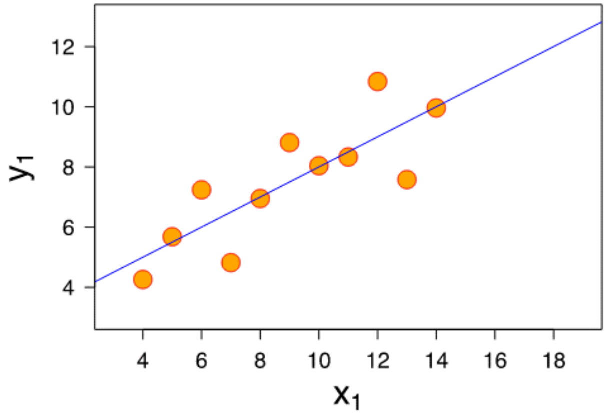
data curation

data structures

data cleaning

data → knowledge

Data to knowledge



data

knowledge

Transform data into visual marks

What is data visualization?

"*Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data*."

Tableau

Data visualization

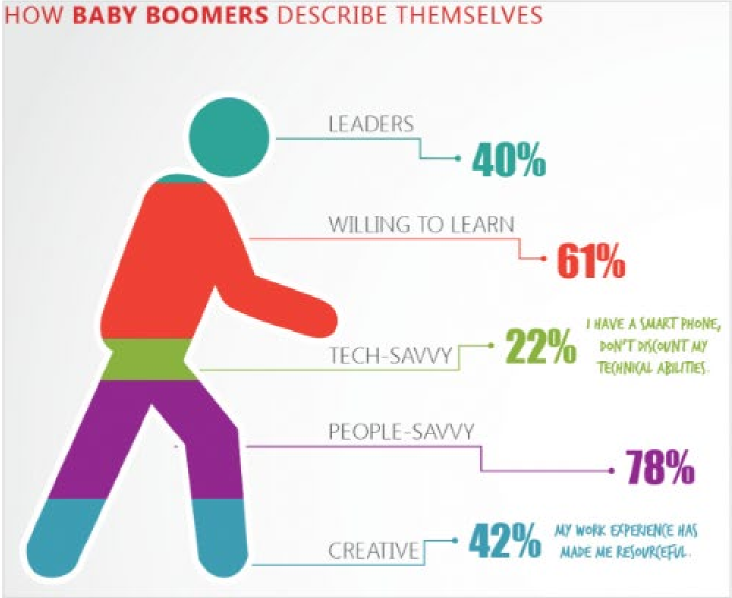
insight $\xrightarrow{\text{Communication}}$ insight

data $\xrightarrow{\text{Exploration / Analysis}}$ insight

Data visualization

insight → insight

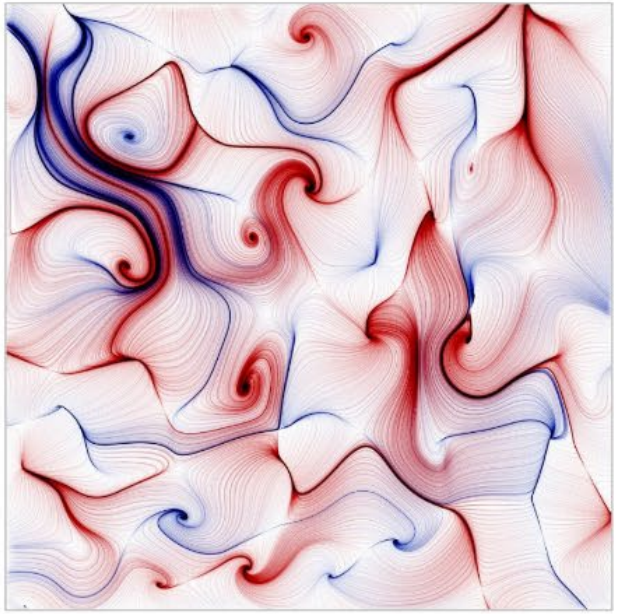
Communication

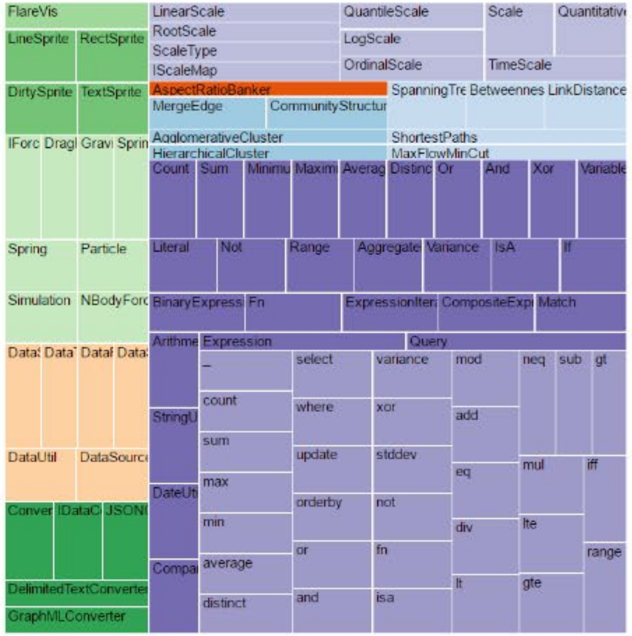


Data visualization

data → insight

Exploration / Analysis

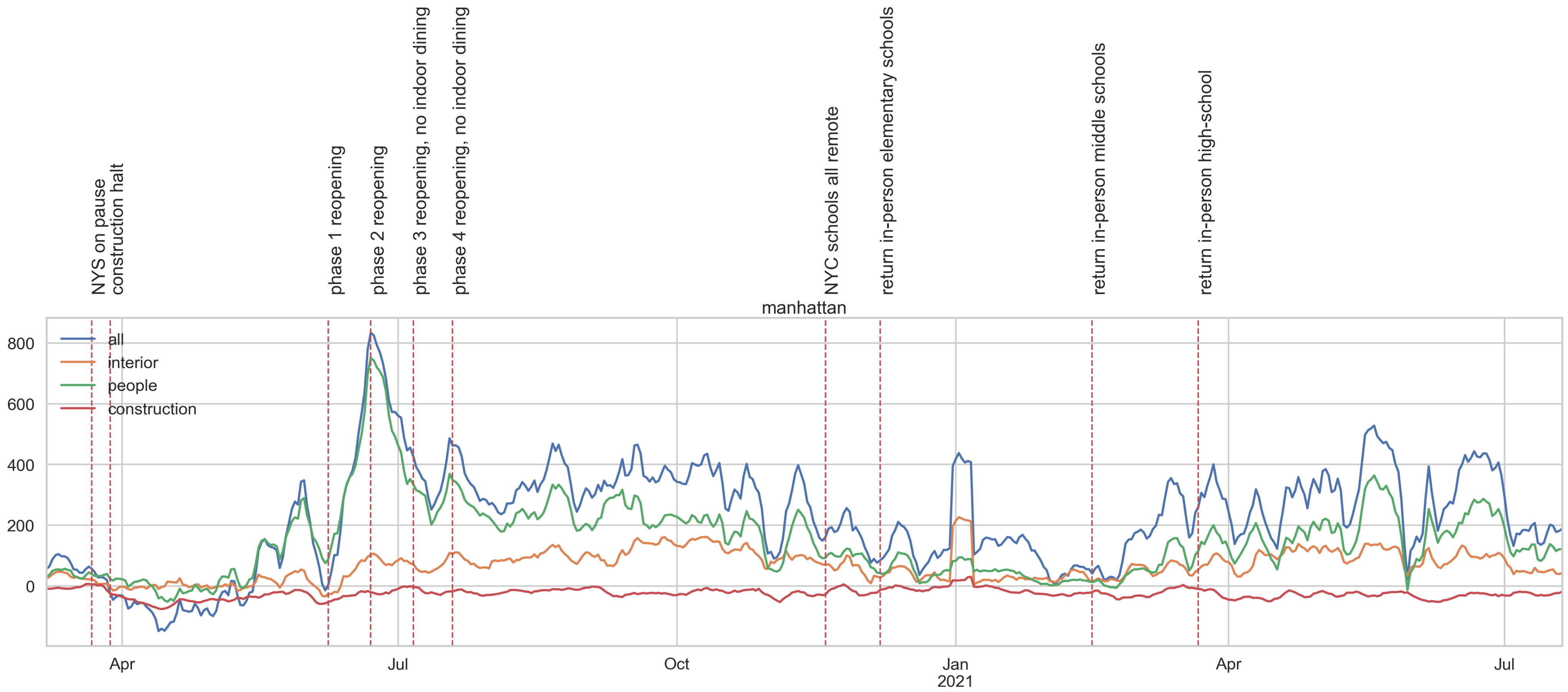




Example:
Noise complaints during pandemic

	date	unique_key	created_date	closed_date	agency	agency_name	complaint_type	descriptor	location_type	incident_zip	...	landmark	date.1	hour_of_day	week	weekday	year	day_of_month	month	aligned_day_index	datetime
0	2017-01-01	35138317	2017-01-01T00:02:54.000	2017-01-01T00:46:54.000	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	11209.0	...	NaN	2017-01-01 00:02:54	0	52	6	2017	1	1	0.0	2017-01-01 00:02:54
1	2017-01-01	35139300	2017-01-01T00:03:41.000	2017-01-01T03:49:13.000	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	10040.0	...	NaN	2017-01-01 00:03:41	0	52	6	2017	1	1	0.0	2017-01-01 00:03:41
2	2017-01-01	35137537	2017-01-01T00:04:01.000	2017-01-01T00:44:40.000	NYPD	New York City Police Department	Noise - Residential	Banging/Pounding	Residential Building/House	11214.0	...	NaN	2017-01-01 00:04:01	0	52	6	2017	1	1	0.0	2017-01-01 00:04:01
3	2017-01-01	35138401	2017-01-01T00:06:04.000	2017-01-01T01:52:03.000	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	11691.0	...	NaN	2017-01-01 00:06:04	0	52	6	2017	1	1	0.0	2017-01-01 00:06:04
4	2017-01-01	35139201	2017-01-01T00:08:24.000	2017-01-01T06:43:42.000	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	10458.0	...	NaN	2017-01-01 00:08:24	0	52	6	2017	1	1	0.0	2017-01-01 00:08:24
5	2017-01-01	35140227	2017-01-01T00:09:08.000	2017-01-01T02:16:21.000	NYPD	New York City Police Department	Noise - Residential	Loud Television	Residential Building/House	11366.0	...	NaN	2017-01-01 00:09:08	0	52	6	2017	1	1	0.0	2017-01-01 00:09:08
6	2017-01-01	35138514	2017-01-01T00:09:22.000	2017-01-01T01:27:35.000	NYPD	New York City Police Department	Noise - Commercial	Loud Music/Party	Club/Bar /Restaurant	11217.0	...	NaN	2017-01-01 00:09:22	0	52	6	2017	1	1	0.0	2017-01-01 00:09:22
7	2017-01-01	35141927	2017-01-01T00:12:02.000	2017-01-01T00:59:53.000	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	11204.0	...	NaN	2017-01-01 00:12:02	0	52	6	2017	1	1	0.0	2017-01-01 00:12:02
8	2017-01-01	35138731	2017-01-01T00:12:36.000	2017-01-01T08:29:48.000	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	10457.0	...	NaN	2017-01-01 00:12:36	0	52	6	2017	1	1	0.0	2017-01-01 00:12:36
9	2017-01-01	35141039	2017-01-01T00:12:44.000	2017-01-01T00:45:47.000	NYPD	New York City Police Department	Noise - Residential	Loud Music/Party	Residential Building/House	10312.0	...	NaN	2017-01-01 00:12:44	0	52	6	2017	1	1	0.0	2017-01-01 00:12:44

Example:
Noise complaints during pandemic



Why visualization?

- Our brains are wired in a visual way.
- Help analysts avoid problems.
- Better communicate findings.
- “*Visualization gives you answer to questions you didn’t know you had.*”

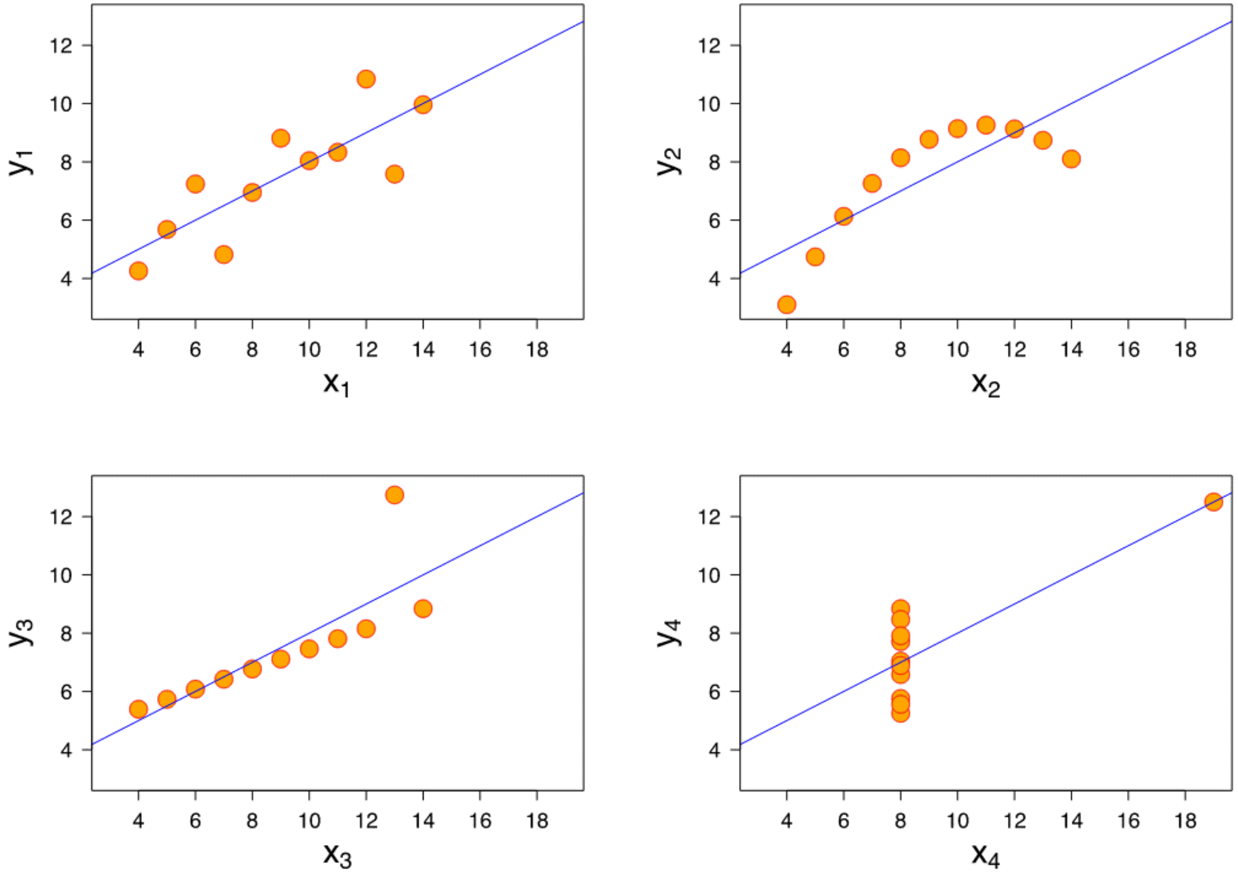
Ben Schneiderman

Importance of visualization

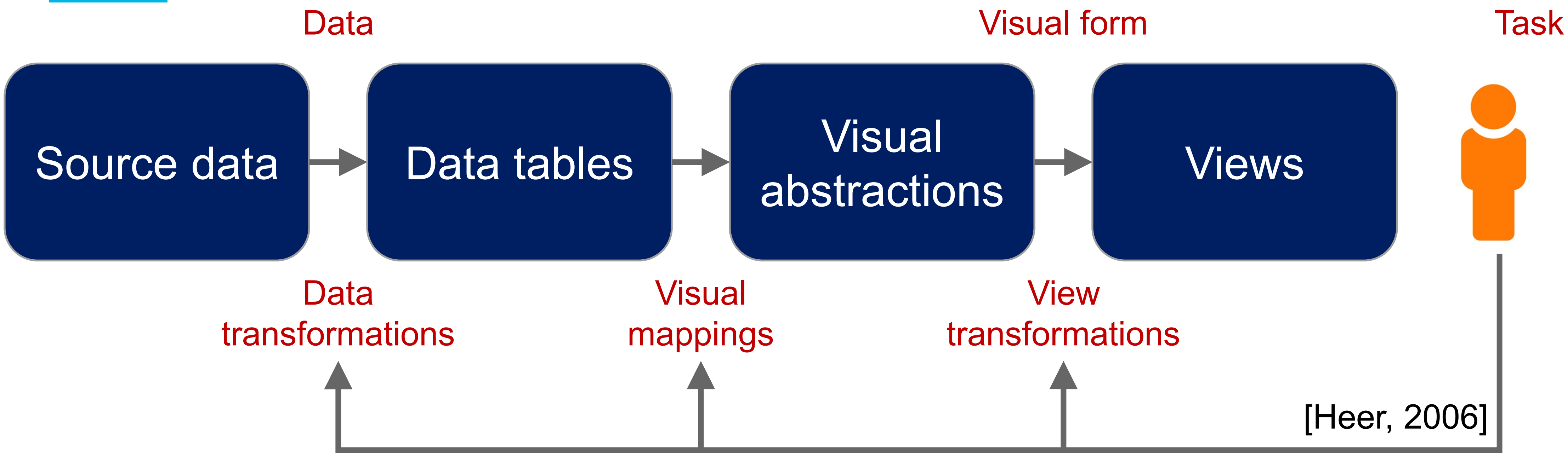
A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	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

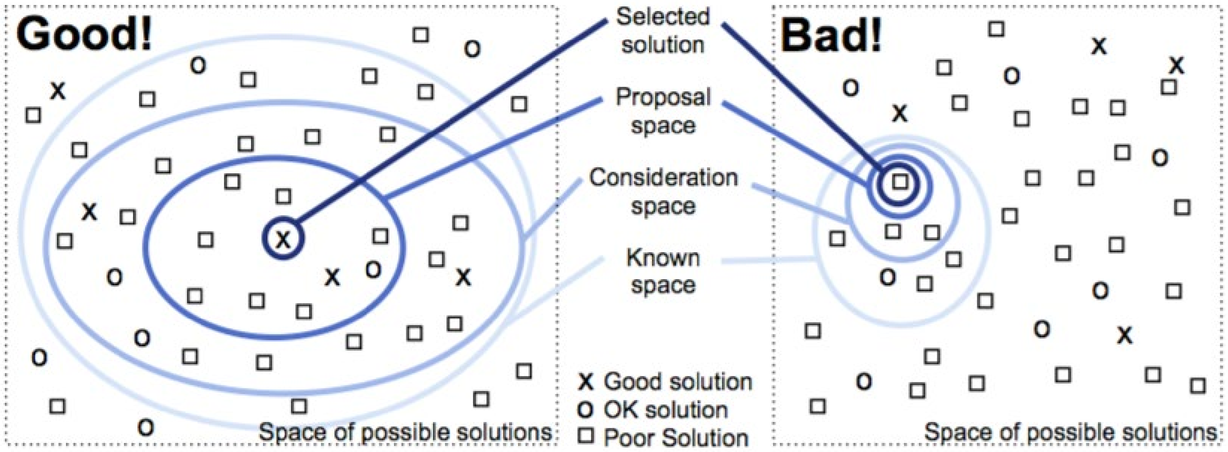


Visualization design



- Creating a data visualization is easy; creating a **good** visualization is hard.
- Visualization design space is huge, it's important to make good choices in each stage.

Visualization design

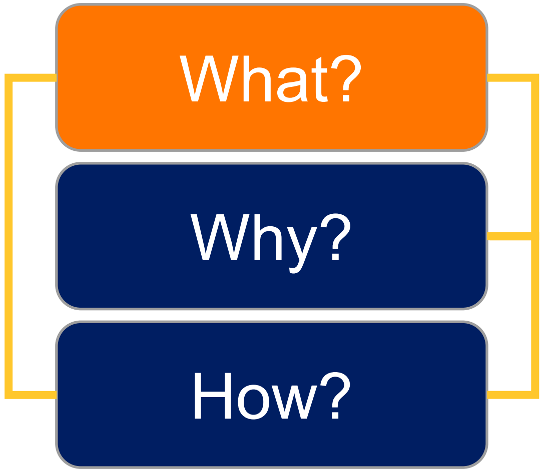


[Munzner, 2015]

Develop principles and techniques to build effective visualizations.

Visualization design

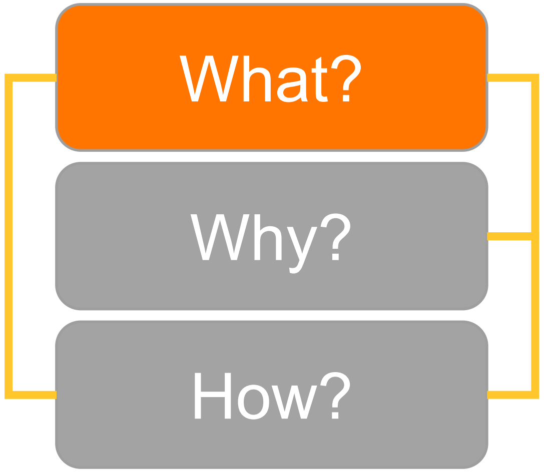
- High-level framework for analyzing vis use:
  - **What** data user sees?
  - **Why** the user intends to use a vis tool?
  - **How** the user intends to use a vis tool?



[Munzner, 2015]

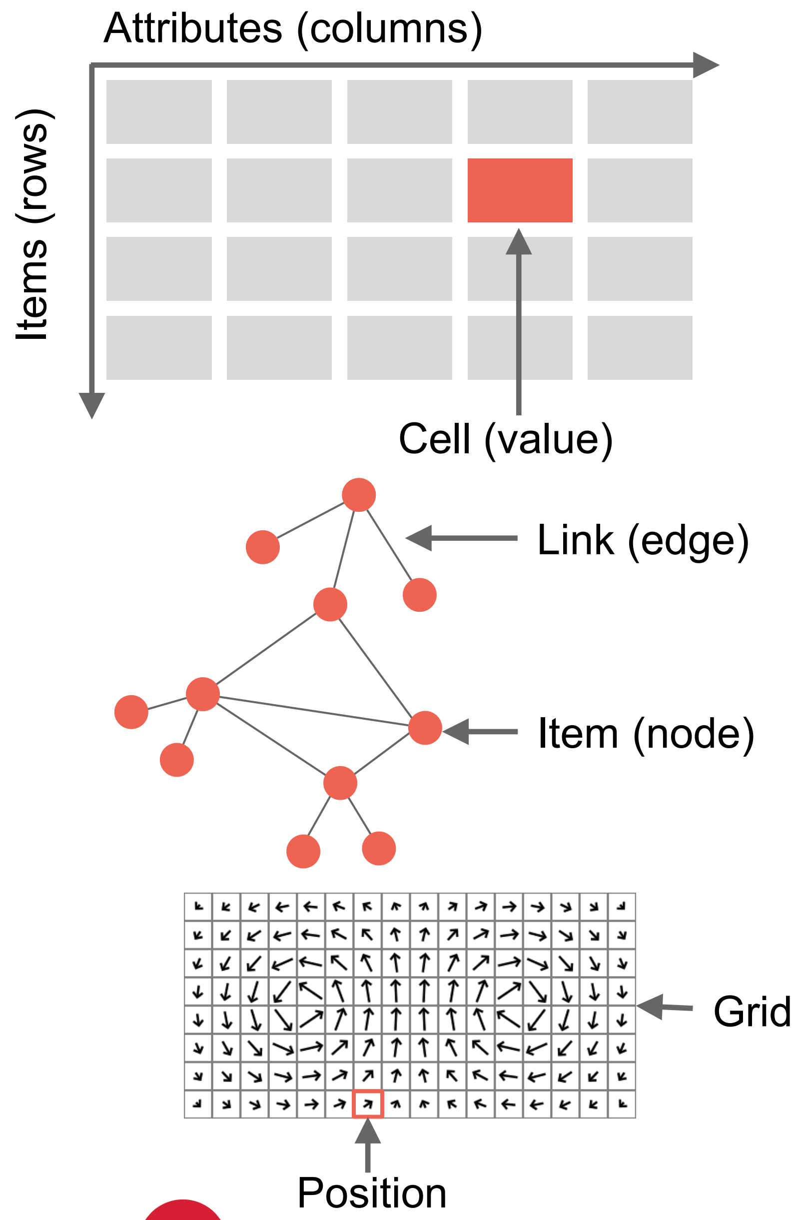
Principles of visualization

- Data
- Visual marks
- Visual channels
- Interaction



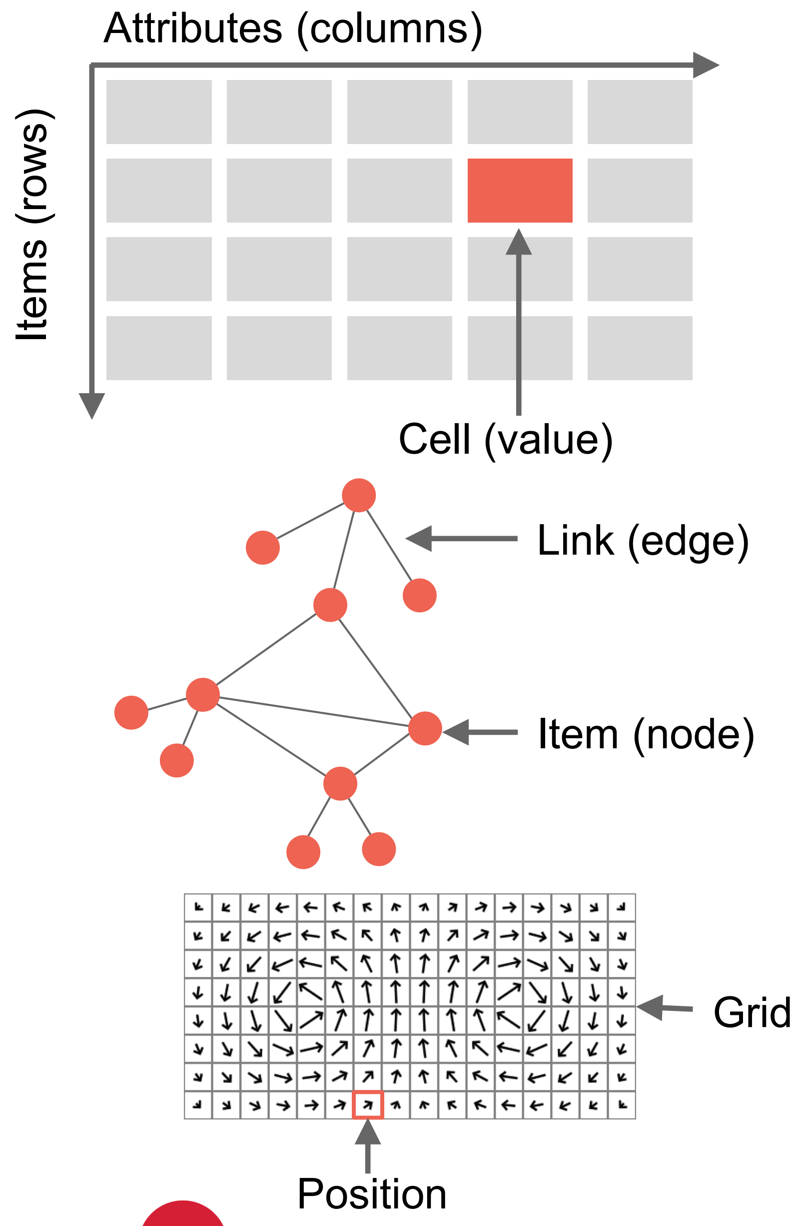
Data types

- Items: individual, discrete entity – record, data point, etc.
- Attributes: item property that can be measured, observed, logged.
- Links: relationship between entities.
- Position: spatial location.
- Grids: strategy for sampling continuous data.



Dataset types

- Table: items and attributes
- Networks & trees: items (nodes), links, attributes
- Fields: grids, positions, attributes.
- Clusters, sets, lists: items.



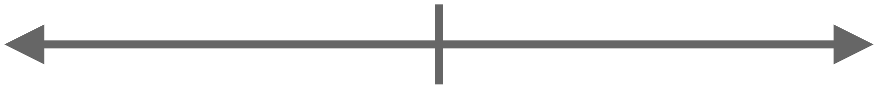
Attribute types

- Categorical: attributes draw from a discrete set, but there may exist hierarchical structure.
  - Fruits, vegetables, furniture type, car type, …
- Ordered: attributes with a natural *ordering*.
  - Ordinal: well-defined ordering, but we cannot do mathematical operations.
    - T-Shirt size (large, medium, small), ranks.
  - Quantitative: measurement of magnitude that supports comparison / mathematical operations.
    - Height, temperature, density, …

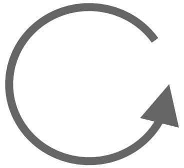
Attribute types

- Ordered: different ordering directions.
  - Sequential: homogeneous range from minimum to maximum value.

  - Diverging: can be deconstructed into two sequences pointing in opposite directions that meet at a common zero point.



  - Cyclic: values wrap around back to starting point.



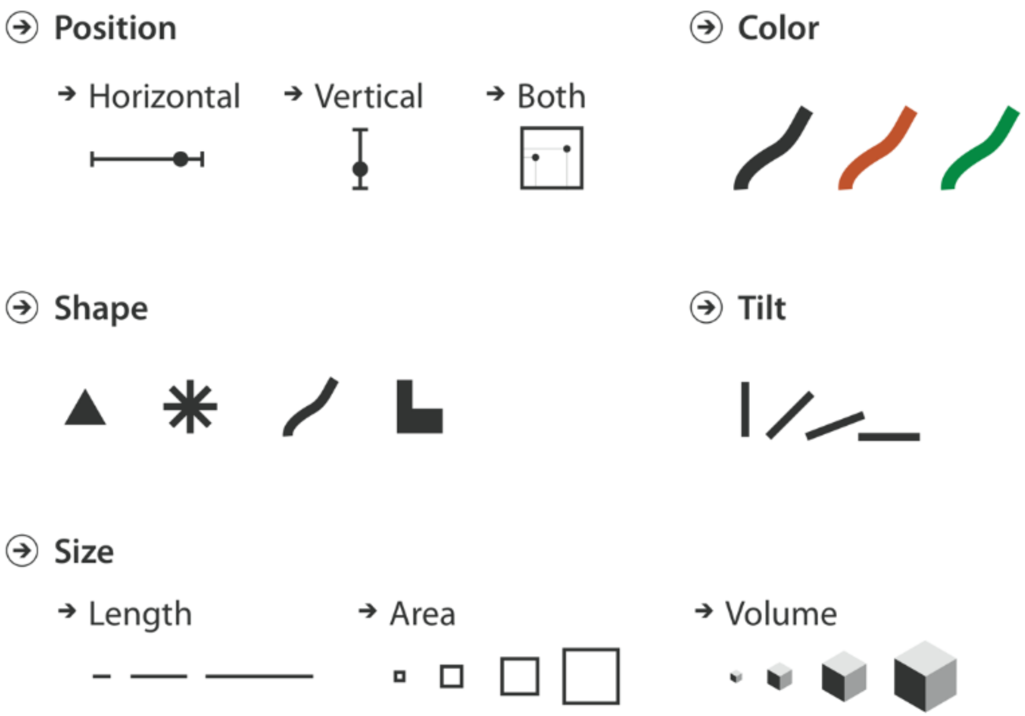
Visual marks

- Represent items and links.
- Geometric primitives, can be classified according to their spatial dimensions: 0D (points), 1D (lines), 2D (areas), etc.

[Munzner, 2014]

Visual channels

- Encode properties of a mark.
- Control appearance based on data attributes.



[Munzner, 2014]

Visual marks & channels

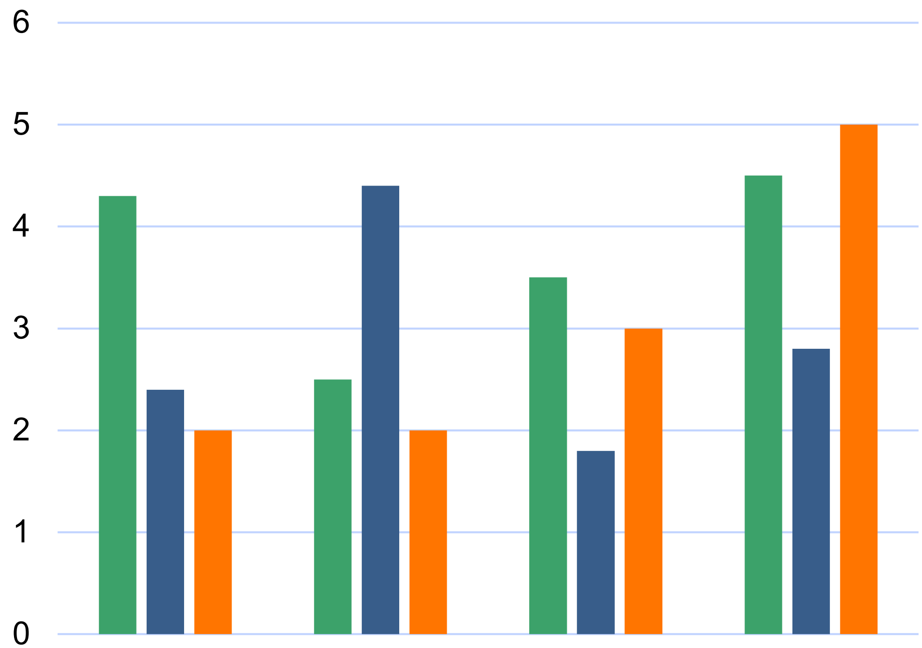
- We can associate tabular data with visual marks and channels as follows:

Attributes → Channels

Items → Marks

Car	Horsepower	Year	Color
Car 1	60	2013	Silver
Car 2	86	2015	Green
Car 3	55	1999	Red
Car 4	50	1990	Blue

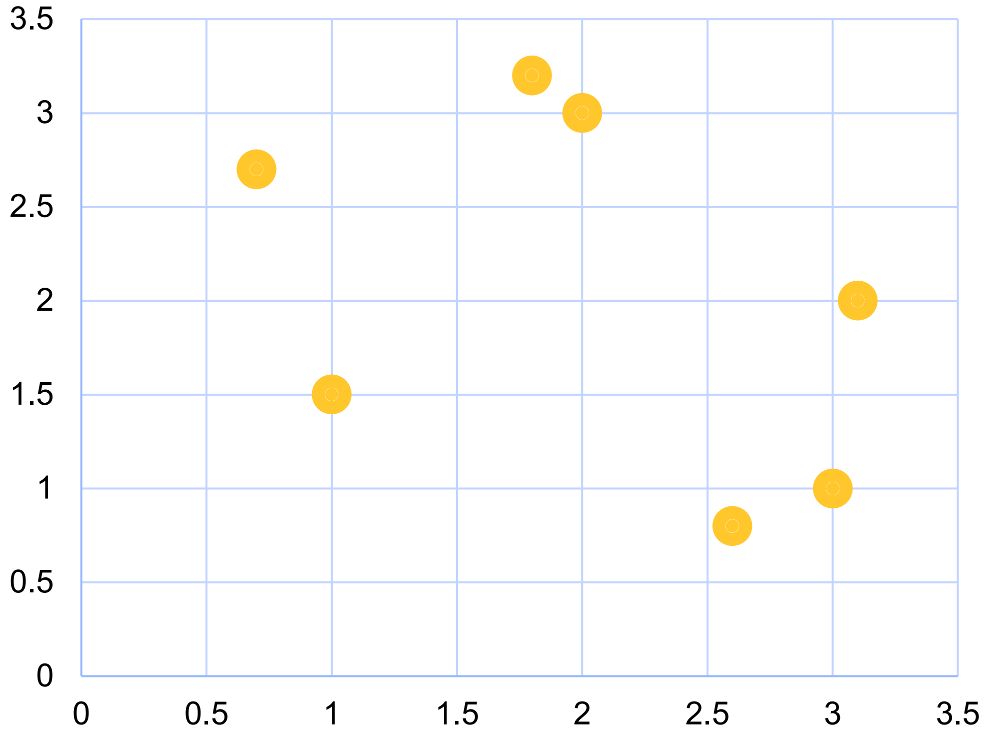
Visual marks & channels: example 1



Bar charts:

- Marks: lines
- Channels: vertical lengths and horizontal positions.
- Each bar is an item, with the quantitative attribute mapped to y spatial channel and categorical attribute mapped to x spatial channel.

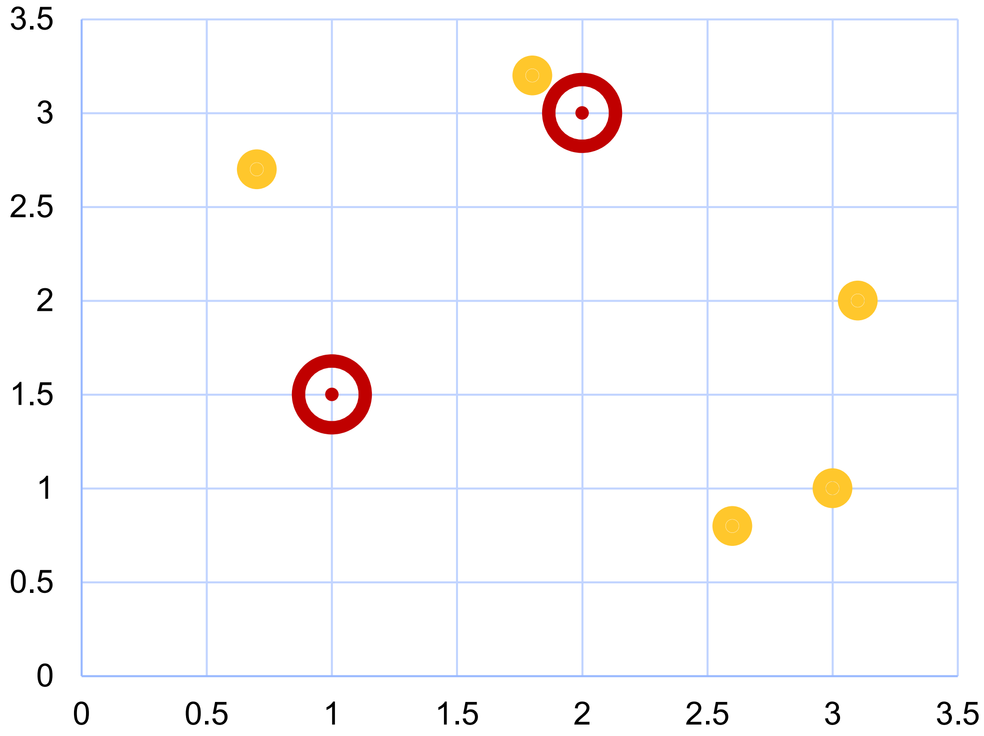
Visual marks & channels: example 2



Scatterplots:

- Marks: points
- Channels: vertical and horizontal positions.
- Each point is an item, with the quantitative attributes mapped to x and y spatial channels.

Visual marks & channels: example 3



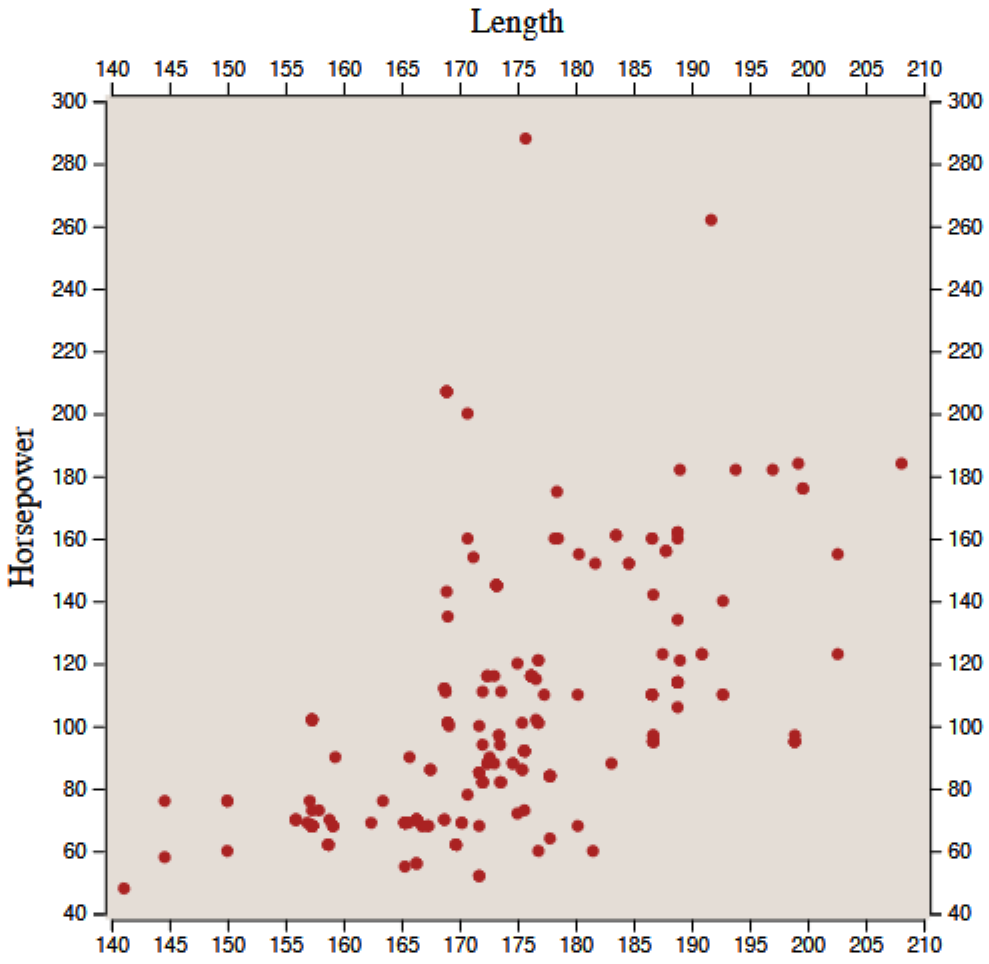
Scatterplots:

- Marks: points
- Channels: vertical and horizontal positions, color, size.
- Each point is an item, with the quantitative attributes mapped to x and y spatial channels, and color and size.

Visual marks & channels: Cars

Car	HP	Price	Length	Style	Maker
Car 1	60	10000	130	Convertible	BMW
Car 2	86	12000	100	Hatchback	Audi
Car 3	55	11000	120	Wagon	Audi
Car 4	50	20000	80	Hatchback	Dodge

Marks: points
Channels: vertical and horizontal positions

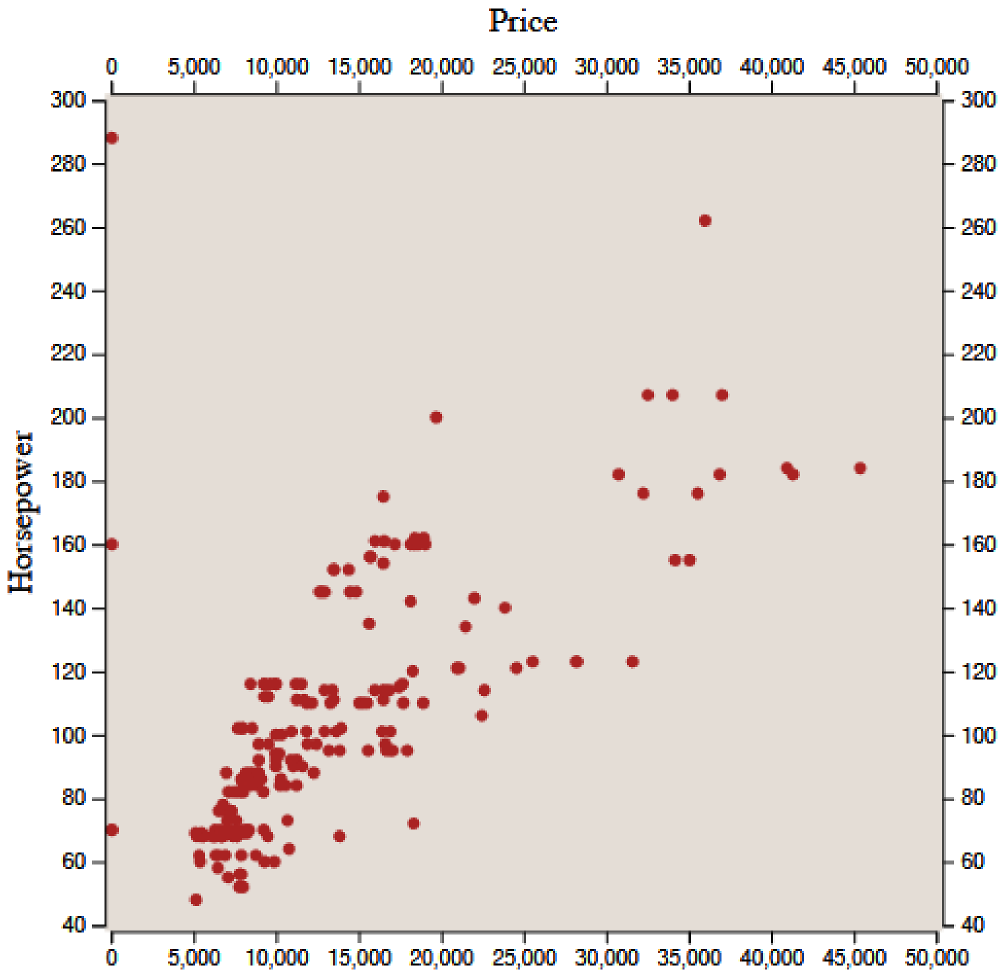


[Freire, 2019]

Visual marks & channels: Cars

Car	HP	Price	Length	Style	Maker
Car 1	60	10000	130	Convertible	BMW
Car 2	86	12000	100	Hatchback	Audi
Car 3	55	11000	120	Wagon	Audi
Car 4	50	20000	80	Hatchback	Dodge

Marks: points
Channels: vertical and horizontal positions

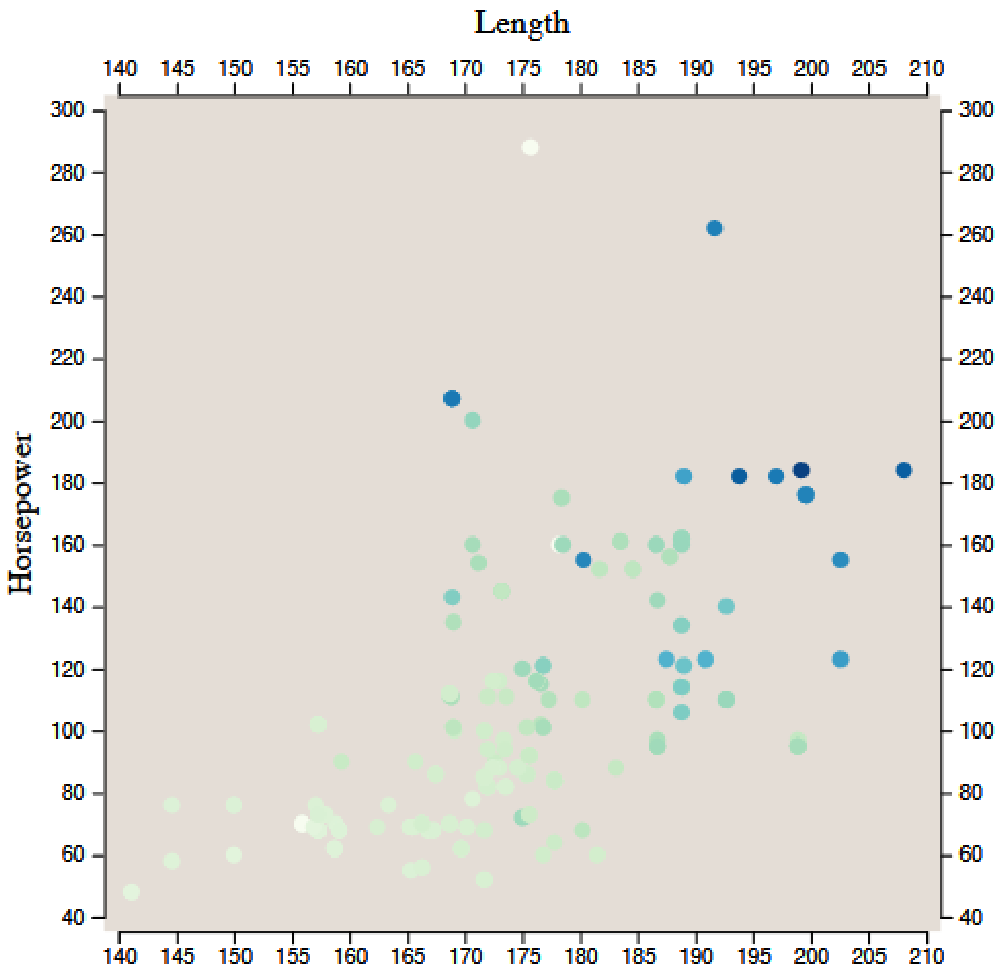


[Freire, 2019]

Visual marks & channels: Cars

Car	HP	Price	Length	Style	Maker
Car 1	60	10000	130	Convertible	BMW
Car 2	86	12000	100	Hatchback	Audi
Car 3	55	11000	120	Wagon	Audi
Car 4	50	20000	80	Hatchback	Dodge

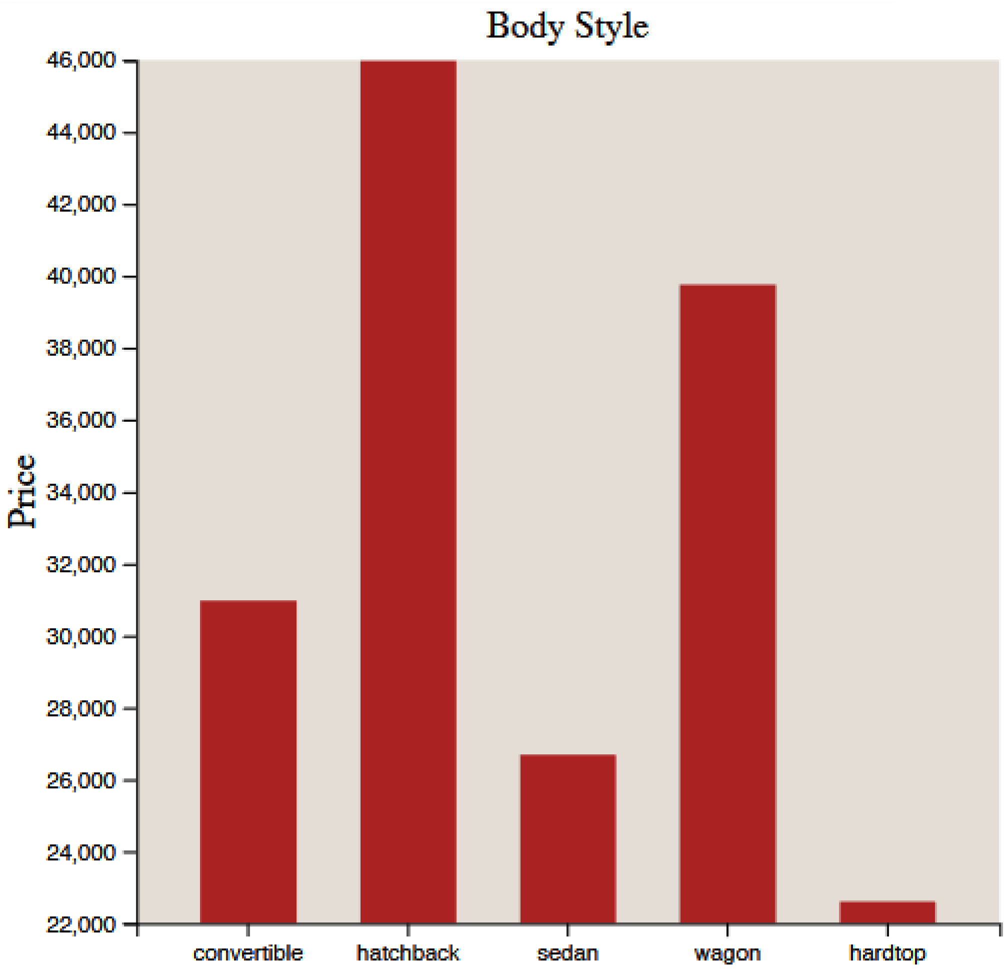
Marks: points
Channels: vertical and horizontal positions, color



[Freire, 2019]

Visual marks & channels: Cars

Car	HP	Price	Length	Style	Maker
Car 1	60	10000	130	Convertible	BMW
Car 2	86	12000	100	Hatchback	Audi
Car 3	55	11000	120	Wagon	Audi
Car 4	50	20000	80	Hatchback	Dodge



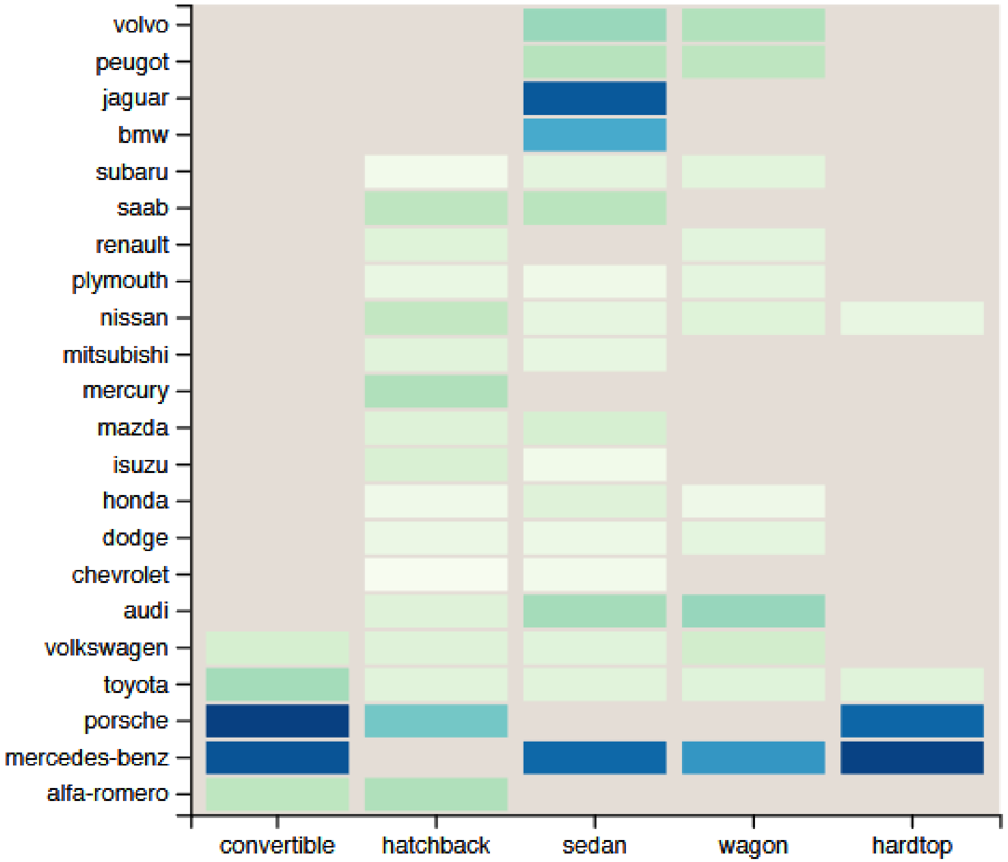
Marks: lines
Channels: vertical lengths and horizontal positions

[Freire, 2019]

Visual marks & channels: Cars

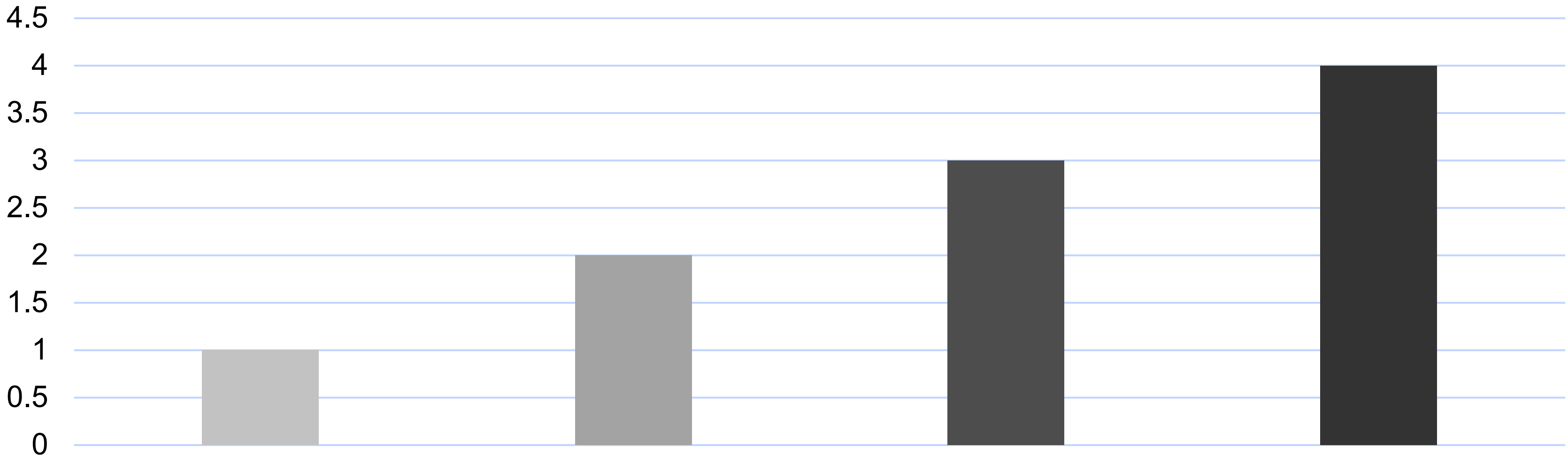
Car	HP	Price	Length	Style	Maker
Car 1	60	10000	130	Convertible	BMW
Car 2	86	12000	100	Hatchback	Audi
Car 3	55	11000	120	Wagon	Audi
Car 4	50	20000	80	Hatchback	Dodge

Marks: area (simple box)
Channels: vertical and horizontal positions, color



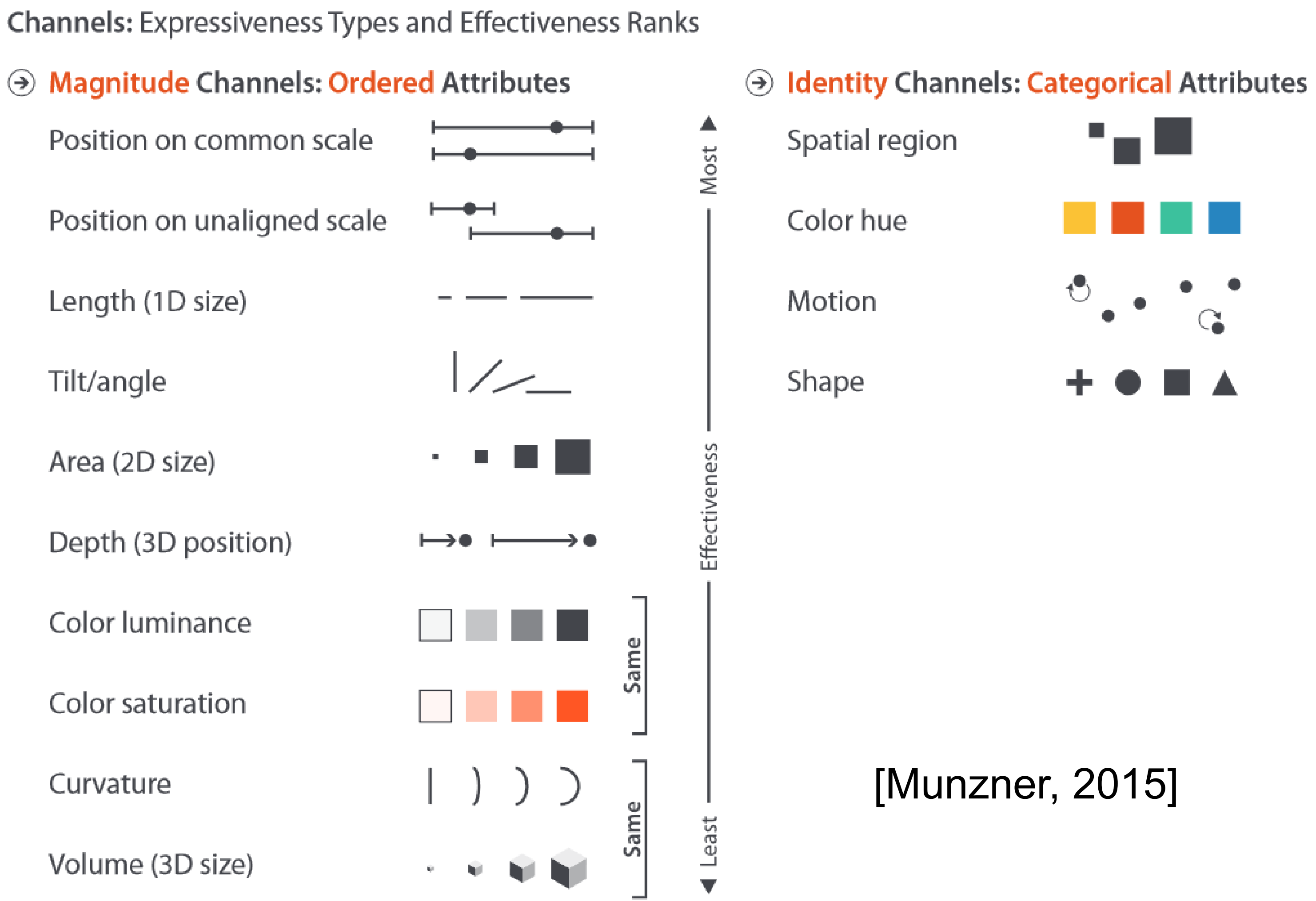
[Freire, 2019]

Redundant encoding



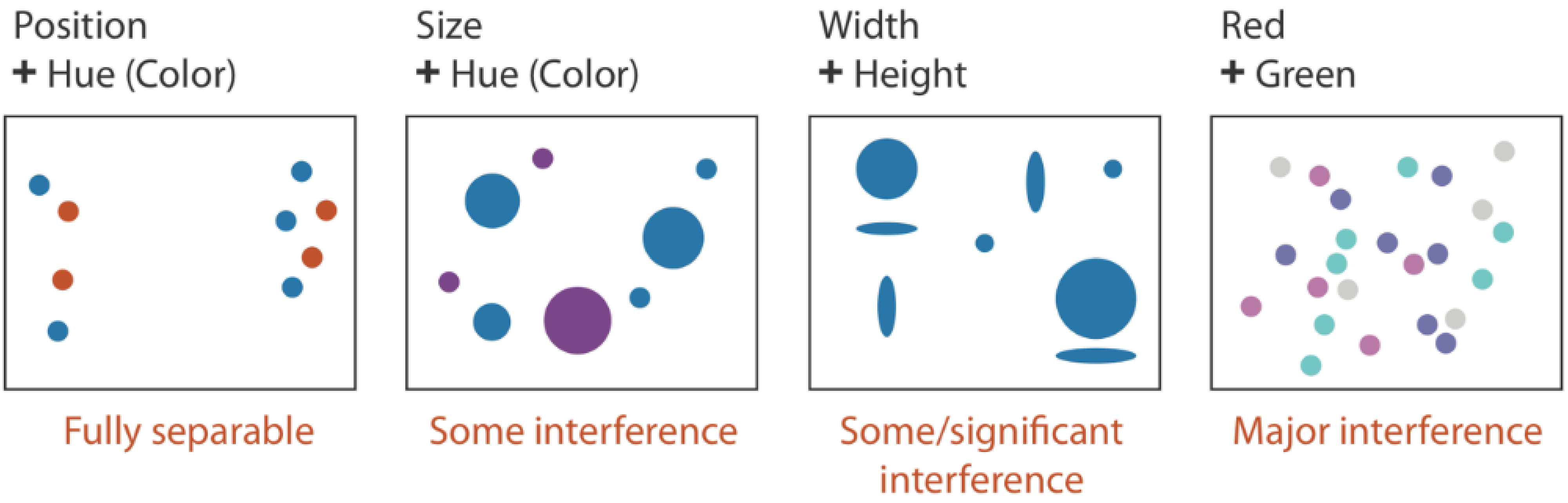
Length, position, color

Expressiveness types and effectiveness ranks



[Munzner, 2015]

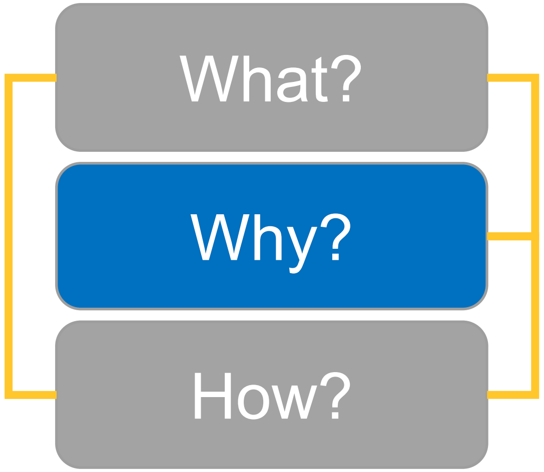
Separability of attributes



[Munzner, 2015]

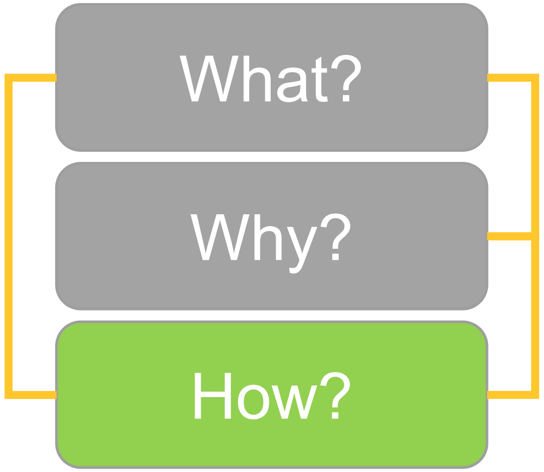
Task abstraction

- 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.”



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	DOLocationII	payment_	fare_amou	extra	mta_tax	tip_amou	tolls_amo	improvem	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

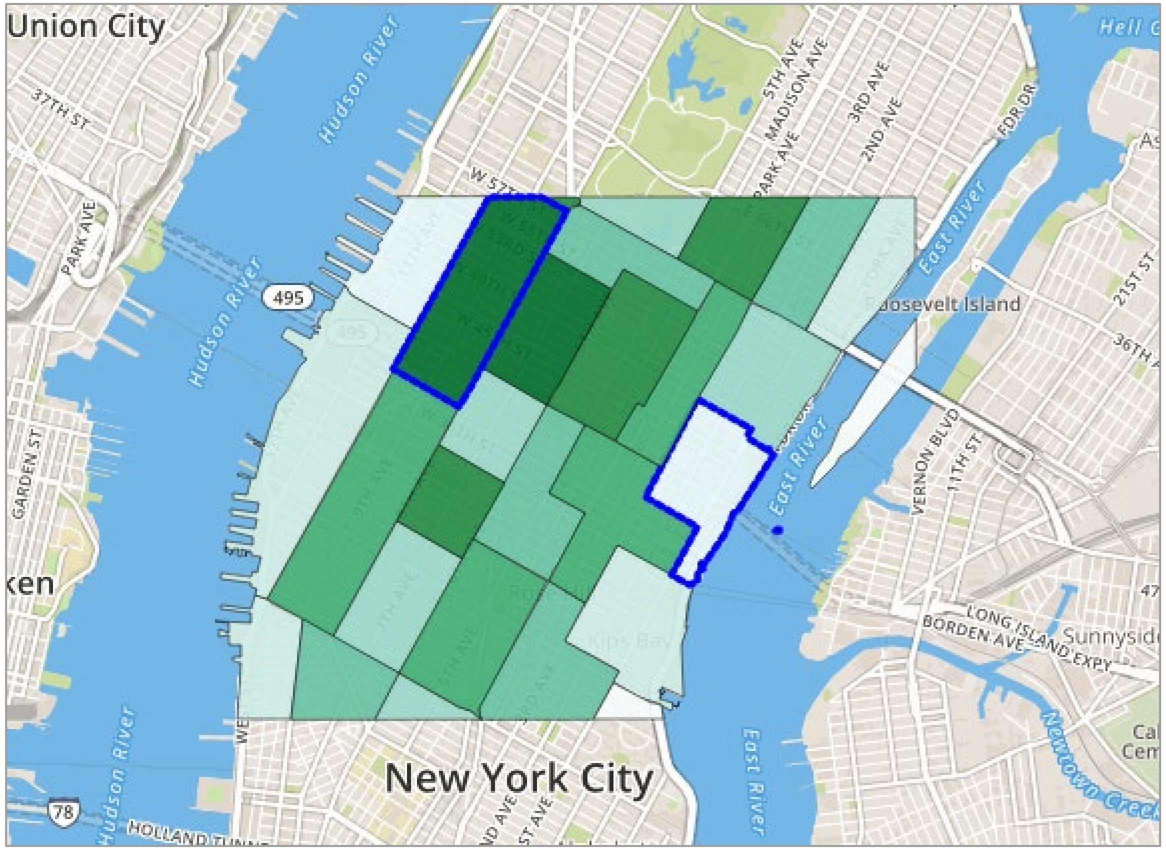
VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_f	PULocationID	DOLocationII	payment_	fare_amo	extra	mta_tax	tip_amou	tolls_amo	improvem	total_amount
1	1/1/2018 0:21	1/1/2018 0:24	1	0.5	1	N	41	24	2	4.5	0.5	0.5	0	0	0.3	5.8
1	1/1/2018 0:44	1/1/2018 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

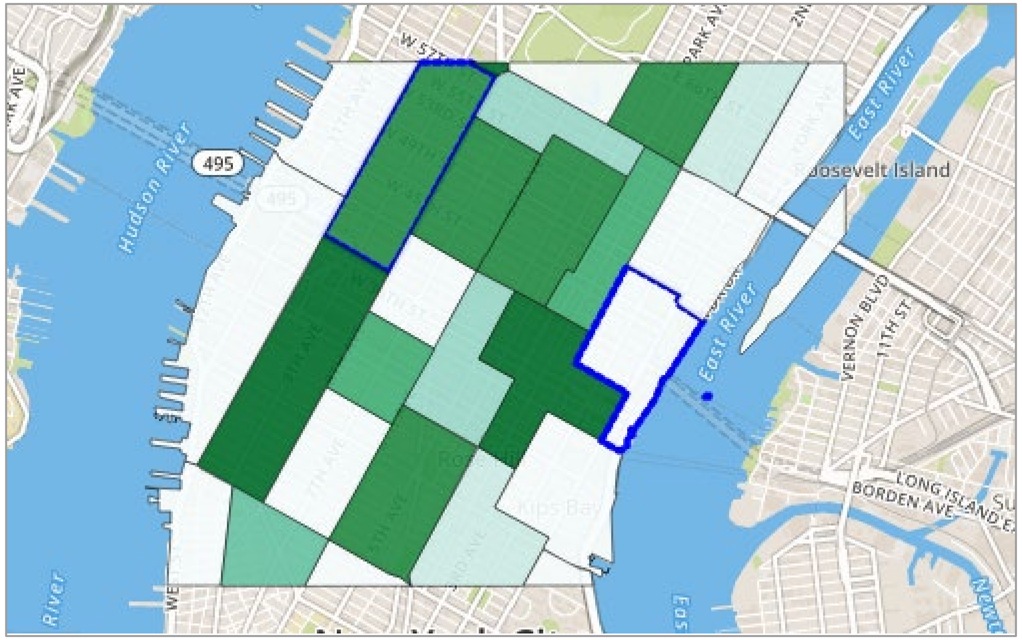
VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_f	PULocationID	DOLocationII	payment_	fare_amou	extra	mta_tax	tip_amou	tolls_amo	improvem	total_amount	Area
1	1/1/2018 0:21	1/1/2018 0:24	1	0.5	1	N	41	24	2	4.5	0.5	0.5	0	0	0.3	5.8	Midtown
1	1/1/2018 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	34.8	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	16.55	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	5.8	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	12.35	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	8.5	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	6.8	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	8.15	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	17.8	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	6.8	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	8.8	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	24.35	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	7.5	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	10.8	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	17.8	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

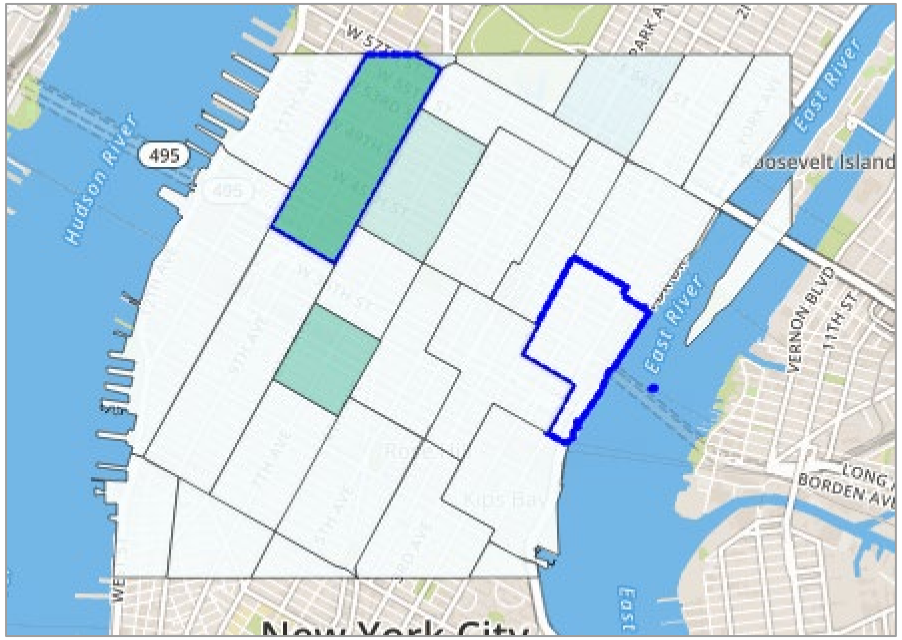
Quantitative data
Mark: polygon areas
Channel: color



Visual interaction

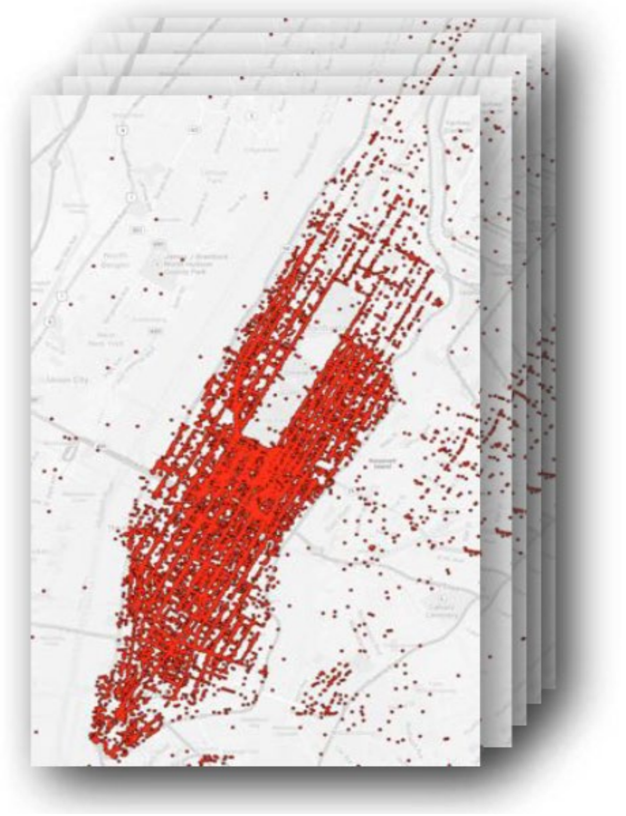


12pm – 2pm pickups

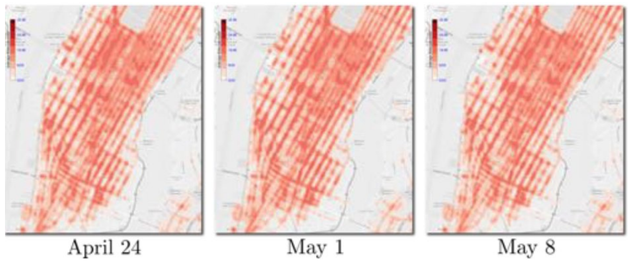


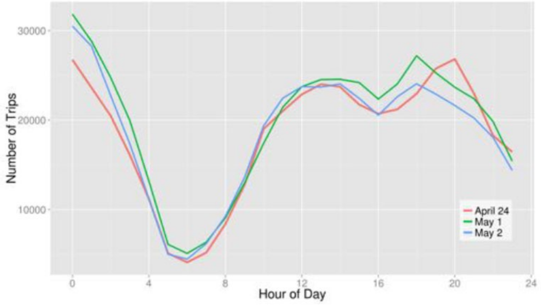
6am pickups

Big data challenges

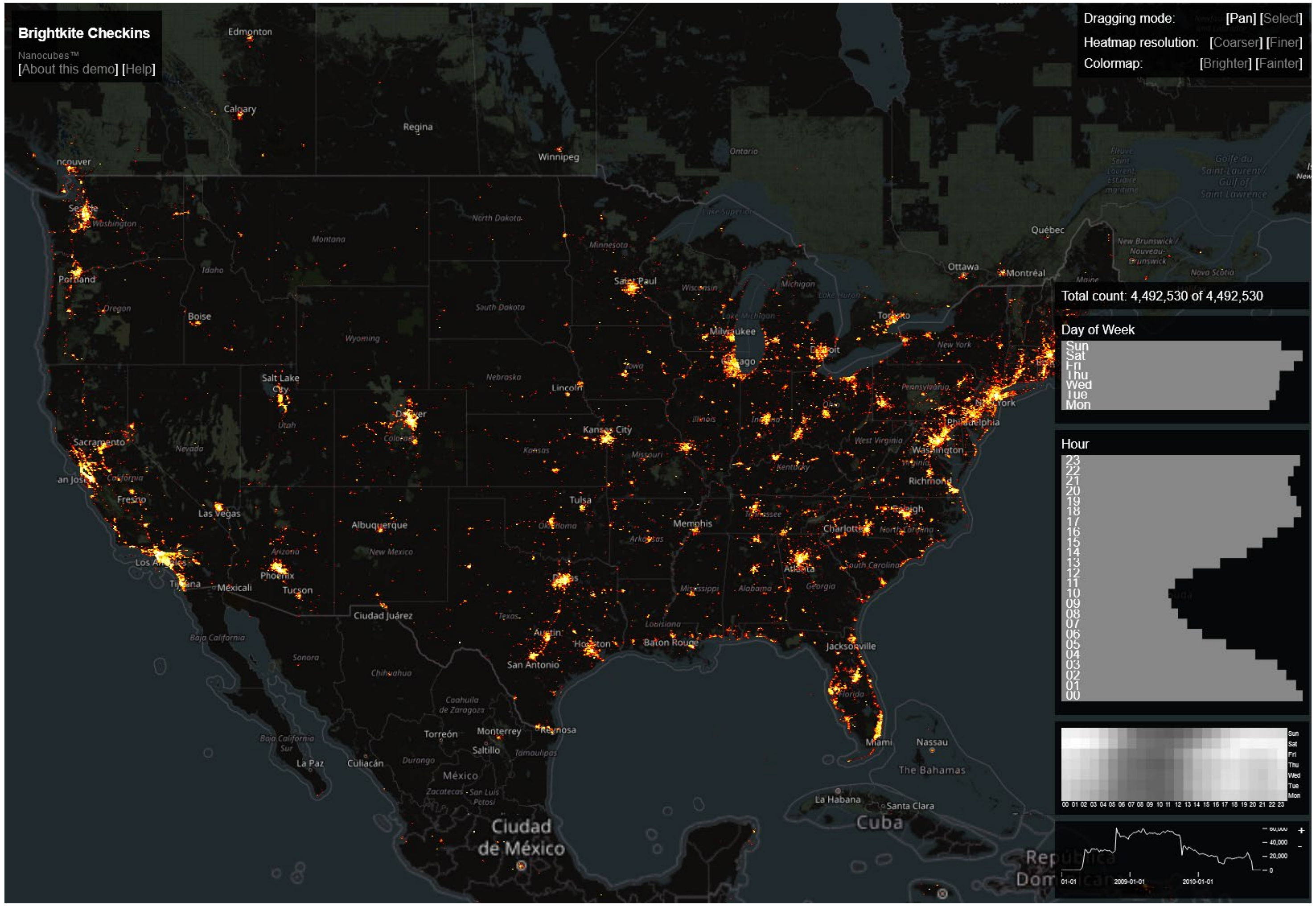


- 365*24 1-hour slices in one year.
- Which slides are interesting?





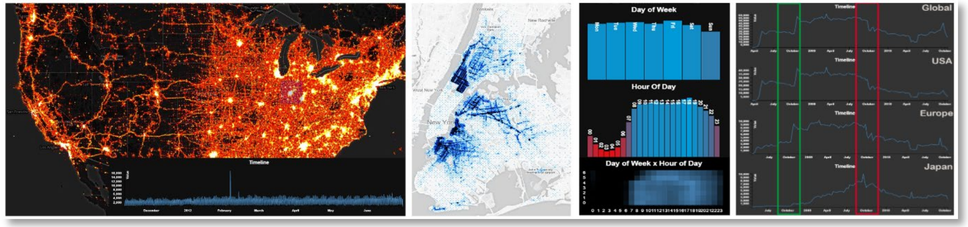
Facet and reduce



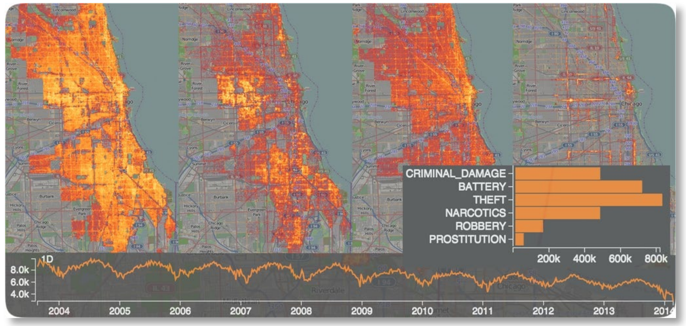
[Lins, 2012]

Techniques for Interactive Visual Analysis

OLAP queries

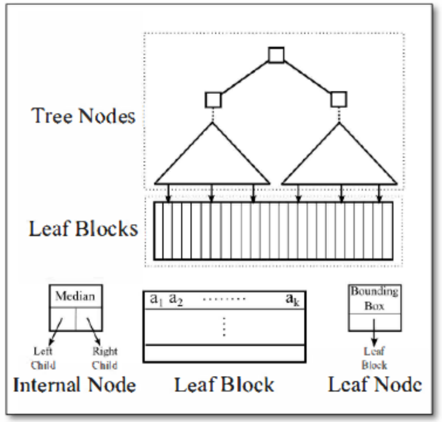


Hashedcubes [Pahins et al., 2017]



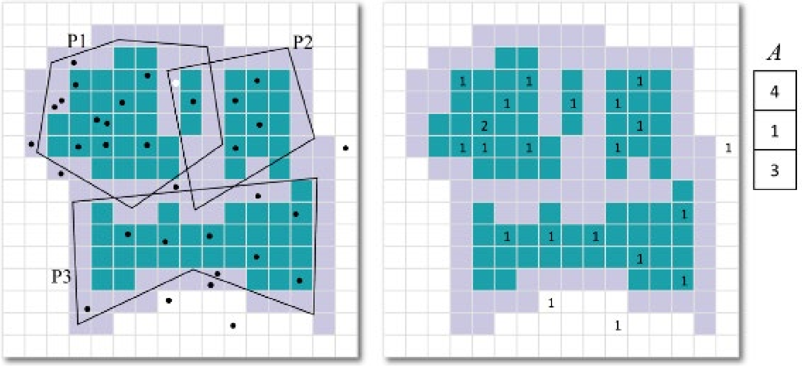
Nanocube [Lins et al., 2013]
TopKube [Miranda et al., 2018]

Selection



STIG [Doraiswamy et al., 2015]

Spatiotemporal joins



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

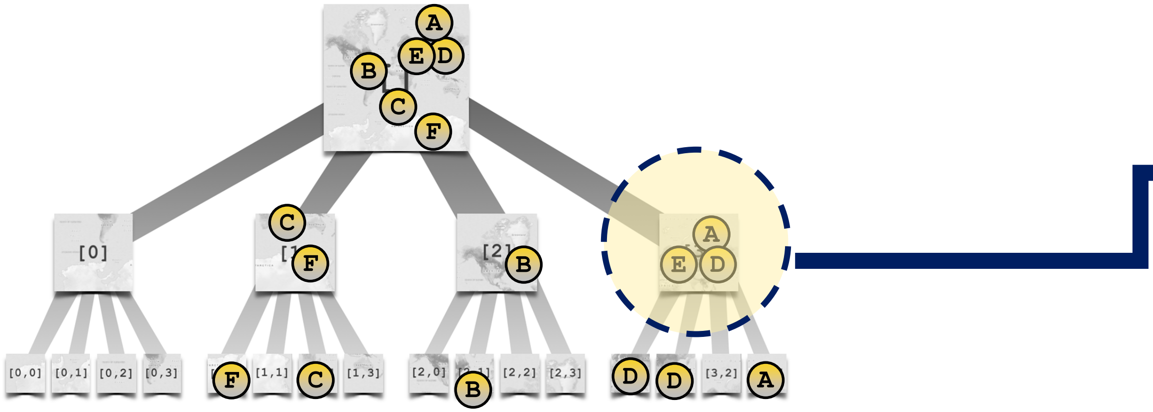
Datacube model

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).

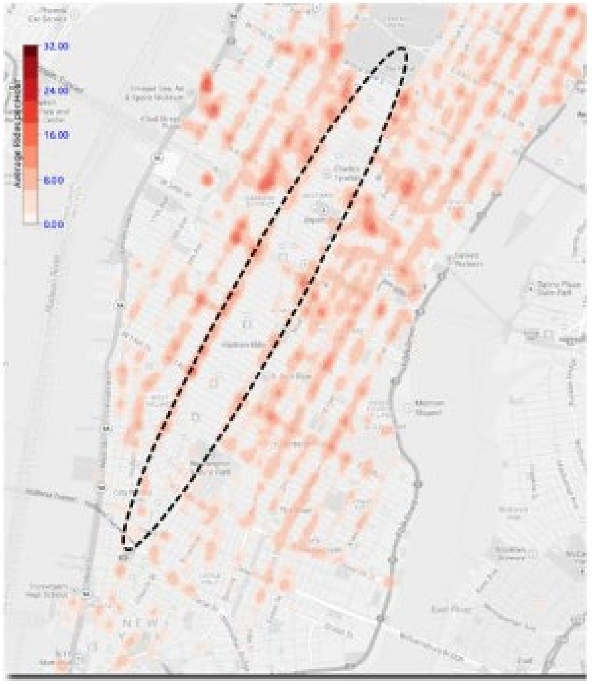
latitude	longitude	keyword
42.102908	-73.242852	#phoenix
29.617161	-81.636398	#phoenix
23.014051	75.120052	#la
26.014051	75.120052	#nyc
28.014051	74.120052	#la
23.014051	75.120052	#phoenix

	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.