

Data questions & transformations

CS424: Visualization & Visual Analytics

Fabio Miranda

<https://fmiranda.me>

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	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

How to visualize this data?

How to visualize data?

1. Decide what to visualize → select & transform data
2. Decide how to visualize it → choose or design representations

How to visualize data?

1. Decide what to visualize → select & transform data
2. Decide how to visualize it → choose or design representations

Possible questions

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_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	5.8	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

What are the most popular regions during the day?

What regions have the highest average trip fare?

Do all regions have the same weekday / weekend pattern?

Vehicle collision in NYC

Did Vision Zero improve the situation of vehicle collisions in NYC since it was implemented?

NYC OpenData

Home Data About Learn Alerts Contact Us Blog | Sign In

Motor Vehicle Collisions - Crashes Public Safety View Data Visualize Export API ...

The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police reported motor vehicle collisions in NYC. The police report (MV104-AN) is required to be filled out for collisions where someone is injured or killed, or where there is at least one fatality.

Updated September 13, 2021
Data Provided by Police Department (NYPD)

About this Dataset

Updated **September 13, 2021**

Data Last Updated September 13, 2021 Metadata Last Updated April 19, 2021

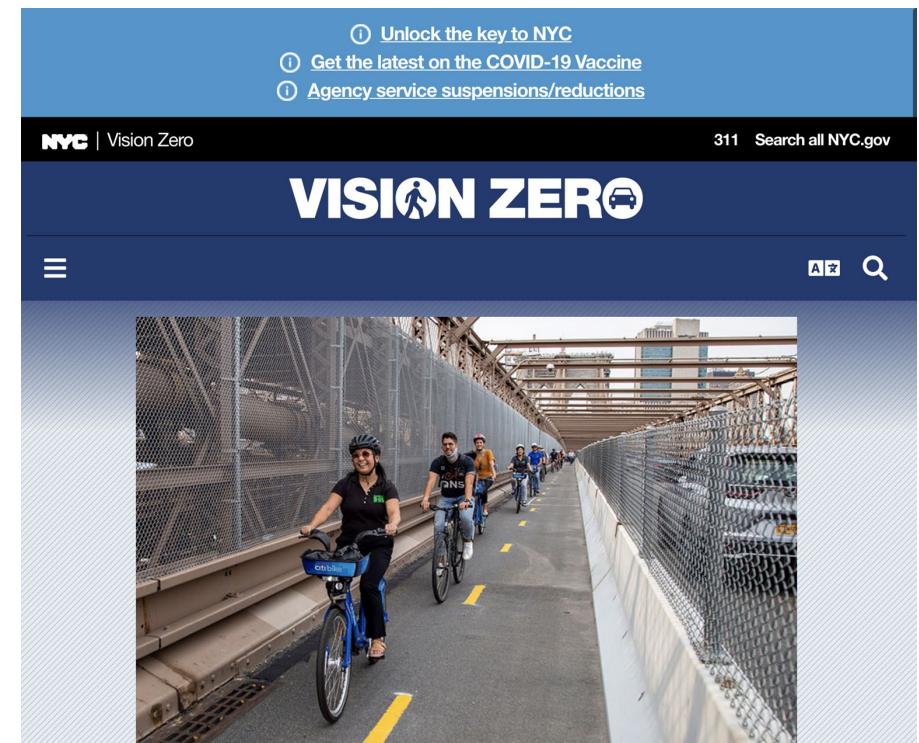
Date Created April 28, 2014

Mute Dataset

Update

Update Frequency Daily
Automation Yes
Date Made Public 5/7/2014

Dataset Information



[Enrico Bertini]

Vehicle collision data

CRASH DATE	Occurrence date of collision
CRASH TIME	Occurrence time of collision
BOROUGH	Borough where collision occurred
ZIP CODE	Postal code of incident occurrence
LATITUDE	Latitude coordinate for Global Coordinate System, WGS 19..
LONGITUDE	Longitude coordinate for Global Coordinate System, WGS 1.
LOCATION	Latitude , Longitude pair
ON STREET NAME	Street on which the collision occurred
CROSS STREET NAME	Nearest cross street to the collision
OFF STREET NAME	Street address if known
NUMBER OF PERSONS INJURED	Number of persons injured
NUMBER OF PERSONS KILLED	Number of persons killed
NUMBER OF PEDESTRIANS INJURED	Number of pedestrians injured
NUMBER OF PEDESTRIANS KILLED	Number of pedestrians killed

NUMBER OF CYCLIST INJURED	Number of cyclists injured
NUMBER OF CYCLIST KILLED	Number of cyclists killed
NUMBER OF MOTORIST INJURED	Number of vehicle occupants injured
NUMBER OF MOTORIST KILLED	Number of vehicle occupants killed
CONTRIBUTING FACTOR VEHICLE 1	Factors contributing to the collision for designated vehicle
CONTRIBUTING FACTOR VEHICLE 2	Factors contributing to the collision for designated vehicle
CONTRIBUTING FACTOR VEHICLE 3	Factors contributing to the collision for designated vehicle
CONTRIBUTING FACTOR VEHICLE 4	Factors contributing to the collision for designated vehicle
CONTRIBUTING FACTOR VEHICLE 5	Factors contributing to the collision for designated vehicle
COLLISION_ID	Unique record code generated by system. Primary Key for ..
VEHICLE TYPE CODE 1	Type of vehicle based on the selected vehicle category (ATV).
VEHICLE TYPE CODE 2	Type of vehicle based on the selected vehicle category (ATV).
VEHICLE TYPE CODE 3	Type of vehicle based on the selected vehicle category (ATV).
VEHICLE TYPE CODE 4	Type of vehicle based on the selected vehicle category (ATV).

Vehicle collision data

CRASH DATE	Occurrence date of collision	NUMBER OF CYCLIST INJURED	Number of cyclists injured
CRASH TIME	Occurrence time of collision	NUMBER OF CYCLIST KILLED	Number of cyclists killed
BOROUGH	Borough where collision occurred	NUMBER OF MOTORIST INJURED	Number of vehicle occupants injured
ZIP CODE	Postal code of incident occurrence	NUMBER OF MOTORIST KILLED	Number of vehicle occupants killed
LATITUDE	Latitude coordinate for Global Coordinate System, WGS 19..	CONTRIBUTING FACTOR VEHICLE 1	Factors contributing to the collision for designated vehicle
LONGITUDE	Longitude coordinate for Global Coordinate System, WGS 1.	CONTRIBUTING FACTOR VEHICLE 2	Factors contributing to the collision for designated vehicle
LOCATION	Latitude, Longitude, Borough, Zip Code, Street Name, Cross Street Name, Off Street Name, Number of Persons Injured, Number of Persons Killed, Number of Pedestrians Injured, Number of Pedestrians Killed	CONTRIBUTING FACTOR VEHICLE 3	Factors contributing to the collision for designated vehicle
ON STREET NAME	Street on which the collision occurred	CONTRIBUTING FACTOR VEHICLE 4	Factors contributing to the collision for designated vehicle
CROSS STREET NAME	Nearest cross street to the collision	CONTRIBUTING FACTOR VEHICLE 5	Factors contributing to the collision for designated vehicle
OFF STREET NAME	Street address if known	COLLISION_ID	Unique record code generated by system. Primary Key for ..
NUMBER OF PERSONS INJURED	Number of persons injured	VEHICLE TYPE CODE 1	Type of vehicle based on the selected vehicle category (ATV).
NUMBER OF PERSONS KILLED	Number of persons killed	VEHICLE TYPE CODE 2	Type of vehicle based on the selected vehicle category (ATV).
NUMBER OF PEDESTRIANS INJURED	Number of pedestrians injured	VEHICLE TYPE CODE 3	Type of vehicle based on the selected vehicle category (ATV).
NUMBER OF PEDESTRIANS KILLED	Number of pedestrians killed	VEHICLE TYPE CODE 4	Type of vehicle based on the selected vehicle category (ATV).

Did Vision Zero improve the situation of vehicle collisions in NYC since it was implemented?

Did Vision Zero **improve the **situation** of vehicle collisions in NYC since it was implemented?**

Can this be answered from
the data?

Did Vision Zero **improve the **situation** of vehicle collisions in NYC since it was implemented?**

Before and after a date or over time?
Everywhere or in some specific areas?

Number of collisions
Number of injuries
Number of deaths
Number of injuries / Number of collisions
Number of speed tickets

Domain question vs. data question

Did Vision Zero **improve** the **situation** of vehicle collisions in NYC since it was implemented?



How did the **number of weekly injuries** change during the last ten years?

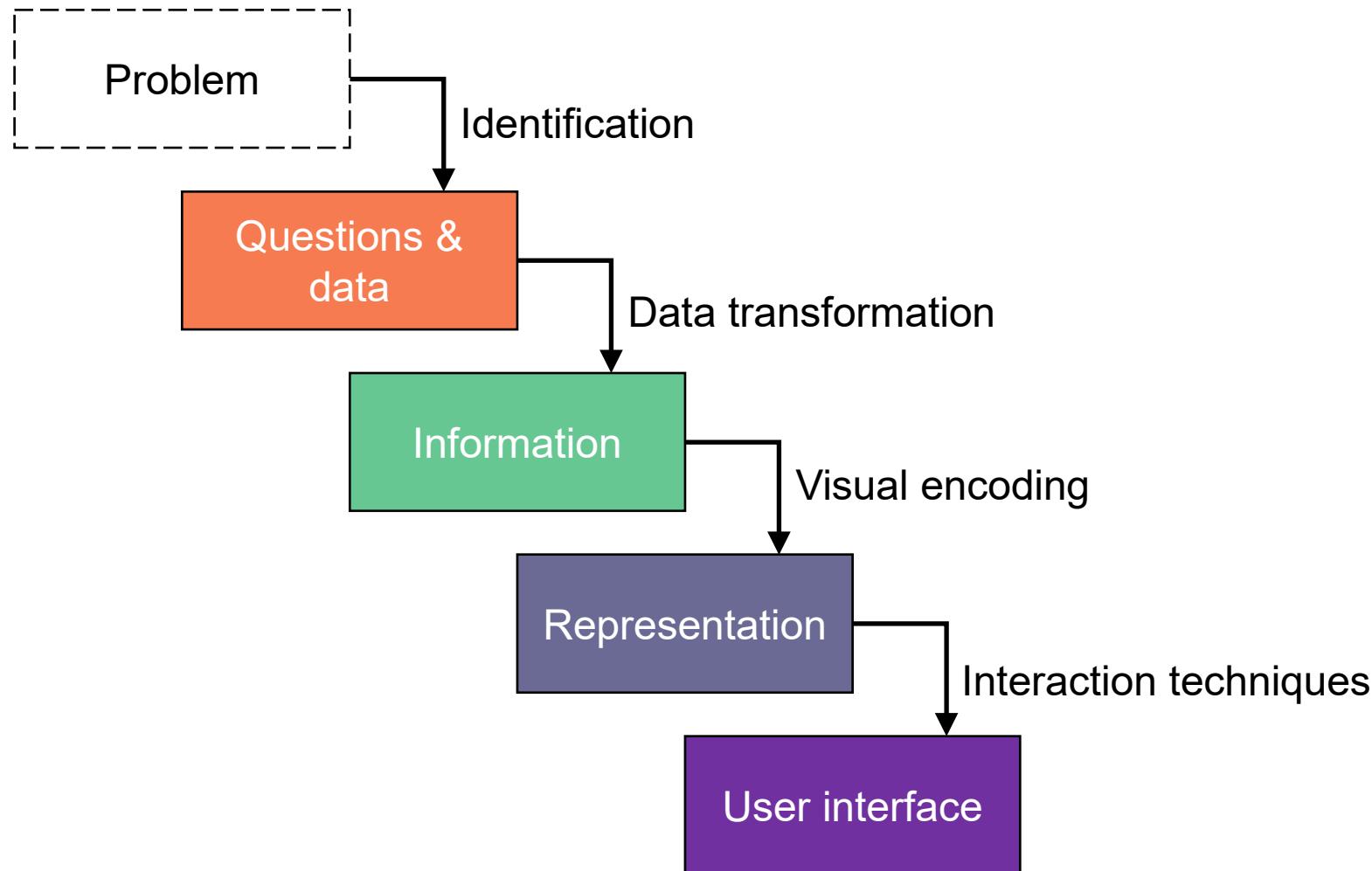
Domain questions vs data questions

Domain Questions	Data Questions
Use domain language	Use data language
Ambiguous	Unambiguous
Imprecise	Specific
Not actionable	Actionable

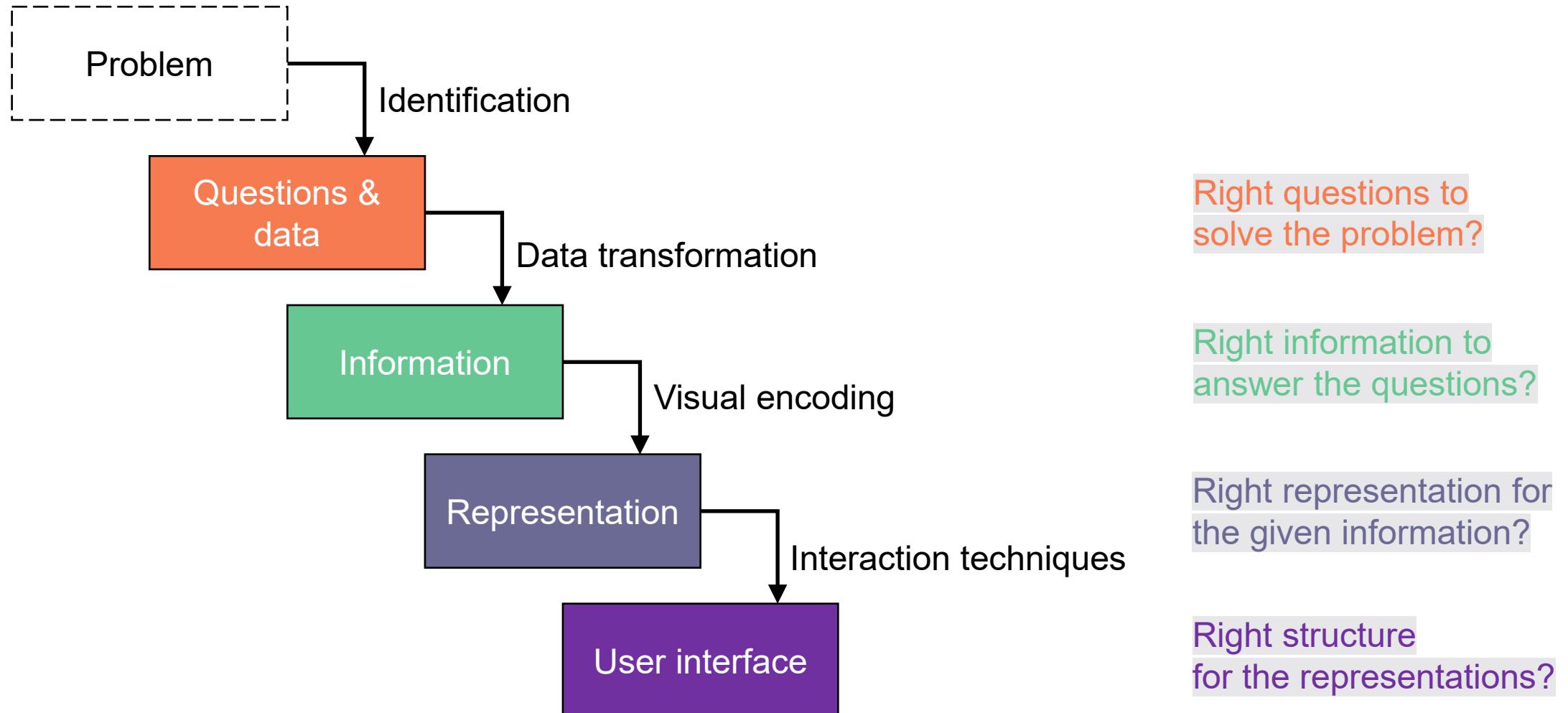
Creating visualizations

- Formulate questions.
- Domain questions → data questions.
- Design transformations and visualizations.
- Answer questions with the visualizations.
 - Test the visualizations.

Creating visualizations



Creating visualizations



Data characterization

- Characterizing the data can help us have a better understanding of **what kind of information** can be extracted from it.
 1. What type of information we can extract.
 2. What type of questions we can ask.
 3. What type of visual representations we can use.

Data characterization

- Characterizing the data can help us have a better understanding of **what kind of information** can be extracted from it.
 1. Fundamental types of information.
 2. Fundamental types of questions.
 3. Fundamental types of visualizations.

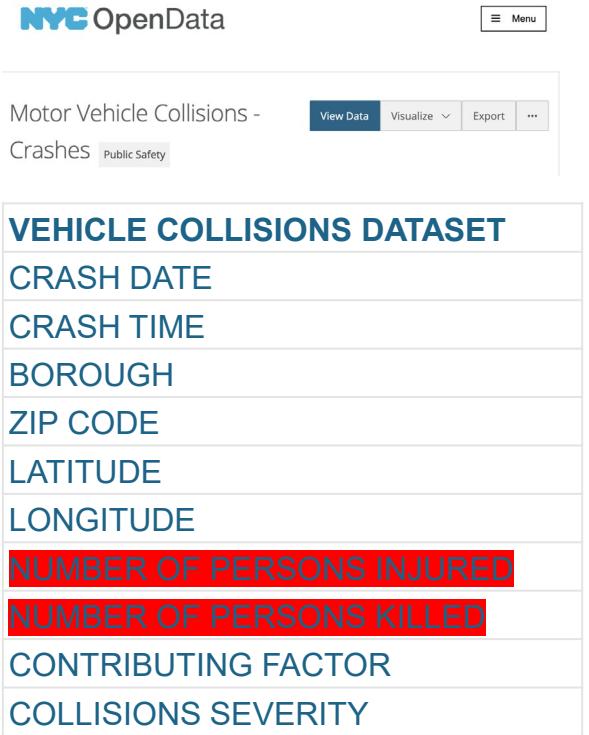
Fundamental types of information

- Categories (ordered, unordered)
- Measures
- Locations and regions
- Times and dates

The screenshot shows the NYC OpenData website interface. At the top, there is a header with the NYC OpenData logo and a menu icon. Below the header, a navigation bar includes 'View Data' (which is highlighted in blue), 'Visualize', 'Export', and an ellipsis button. The main content area displays the 'Motor Vehicle Collisions - Crashes' dataset under the 'Public Safety' category. On the right side of the content area, there is a sidebar titled 'VEHICLE COLLISIONS DATASET' containing a list of fields: CRASH DATE, CRASH TIME, BOROUGH, ZIP CODE, LATITUDE, LONGITUDE, NUMBER OF PERSONS INJURED, NUMBER OF PERSONS KILLED, CONTRIBUTING FACTOR, and COLLISIONS SEVERITY. The 'CONTRIBUTING FACTOR' and 'COLLISIONS SEVERITY' fields are highlighted with red boxes.

Fundamental types of information

- Categories (ordered, unordered)
- Measures
- Locations and regions
- Times and dates

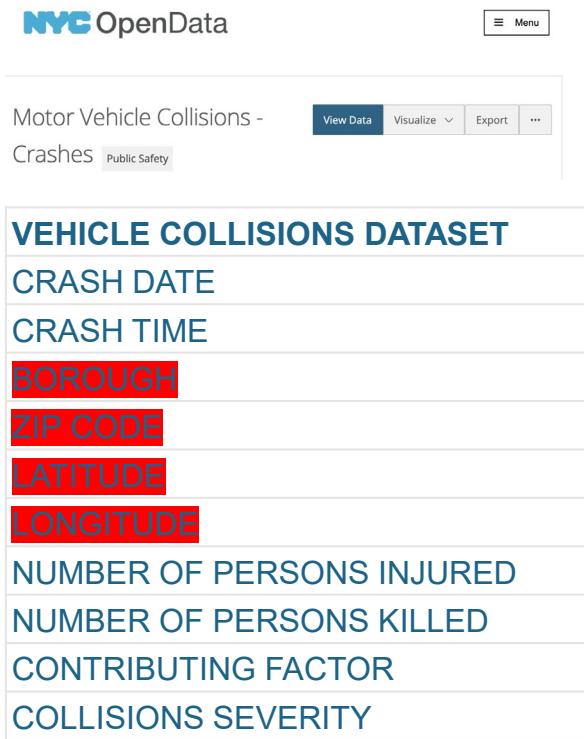


The screenshot shows the NYC OpenData website interface. At the top, there is a logo for "NYC OpenData" and a "Menu" button. Below the header, the dataset title "Motor Vehicle Collisions - Crashes" is displayed, along with buttons for "View Data", "Visualize", "Export", and an ellipsis. The main content area is titled "VEHICLE COLLISIONS DATASET" and lists the following columns: CRASH DATE, CRASH TIME, BOROUGH, ZIP CODE, LATITUDE, LONGITUDE, NUMBER OF PERSONS INJURED, NUMBER OF PERSONS KILLED, CONTRIBUTING FACTOR, and COLLISIONS SEVERITY.

VEHICLE COLLISIONS DATASET
CRASH DATE
CRASH TIME
BOROUGH
ZIP CODE
LATITUDE
LONGITUDE
NUMBER OF PERSONS INJURED
NUMBER OF PERSONS KILLED
CONTRIBUTING FACTOR
COLLISIONS SEVERITY

Fundamental types of information

- Categories (ordered, unordered)
- Measures
- Locations and regions
- Times and dates



The screenshot shows the NYC OpenData website interface. At the top, there is a logo for "NYC OpenData" and a "Menu" button. Below the header, the dataset title "Motor Vehicle Collisions - Crashes" is displayed, along with buttons for "View Data", "Visualize", "Export", and "...". The main content area is titled "VEHICLE COLLISIONS DATASET" and lists various fields: CRASH DATE, CRASH TIME, BOROUGH, ZIP CODE, LATITUDE, LONGITUDE, NUMBER OF PERSONS INJURED, NUMBER OF PERSONS KILLED, CONTRIBUTING FACTOR, and COLLISIONS SEVERITY. The "BOROUGH", "ZIP CODE", "LATITUDE", and "LONGITUDE" fields are highlighted in red.

VEHICLE COLLISIONS DATASET
CRASH DATE
CRASH TIME
BOROUGH
ZIP CODE
LATITUDE
LONGITUDE
NUMBER OF PERSONS INJURED
NUMBER OF PERSONS KILLED
CONTRIBUTING FACTOR
COLLISIONS SEVERITY

Fundamental types of information

- Categories (ordered, unordered)
- Measures
- Locations and regions
- Times and dates

The screenshot shows the NYC OpenData website interface. At the top, there is a logo for 'NYC OpenData' and a 'Menu' button. Below the header, the dataset title 'Motor Vehicle Collisions - Crashes' is displayed, along with a 'Public Safety' tag. A navigation bar below the title includes 'View Data' (which is highlighted in blue), 'Visualize', 'Export', and a '...' button. The main content area is titled 'VEHICLE COLLISIONS DATASET' and lists the following fields:

- CRASH DATE
- CRASH TIME
- BOROUGH
- ZIP CODE
- LATITUDE
- LONGITUDE
- NUMBER OF PERSONS INJURED
- NUMBER OF PERSONS KILLED
- CONTRIBUTING FACTOR
- COLLISIONS SEVERITY

Fundamental types of information: relevant properties

- Categories
 - Ordered / unordered
 - Cardinality
 - Hierarchy
- Measures
 - Unidirection / bi-directional
 - Statistical summaries
 - Part-to-whole
- Locations and regions
 - Individual locations vs. regions
 - Locations as categories vs. as spatial coordinates
 - Granularity / hierarchy
- Times and dates
 - Times and dates are composite objects:
MM:DD:YYY
 - Granularity / hierarchy
 - Cyclical structure

Fundamental types of questions

- Combination of fundamental types of information offers a glimpse of the kind of data questions you can typically ask with data visualization.

Categories

Locations and regions

Times and dates

x

Measures

Directionality
Statistical summary
Part-to-whole

Fundamental types of questions

- Category x measures → ranking and comparison
 - *How do the contributing factors compare in terms of number of collisions?*

Fundamental types of questions

- Category x measures → ranking and comparison
 - *How do the contributing factors compare in terms of number of collisions?*
- Location / region x measures → spatial distribution
 - *How do the number of collisions distribute across the NYC area?*

Fundamental types of questions

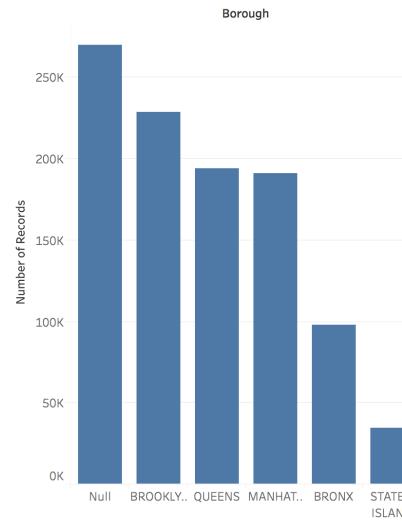
- Time / date x measures → temporal trends
 - *How did the weekly number of collisions change over last year?*
- Measure x measure or category x category → relationships and correlations
 - *Is there a correlation between the number of people injured and the number of people killed?*

Fundamental types of visualization

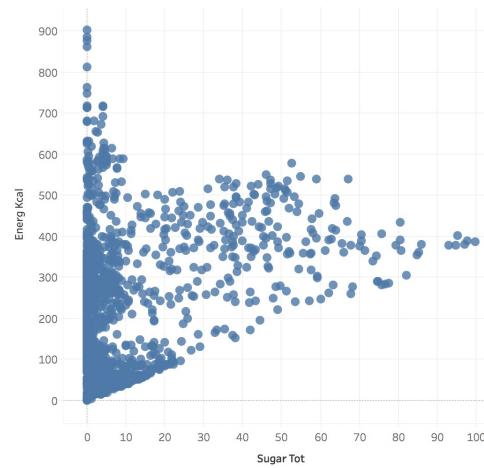
- You can use the data abstraction step to make choices on visual representation.
- Fundamental types of visualization:
 - Widely adopted, effective, useful.
 - Solve very large percentage of visualization problems.
 - Training ground for more sophisticated graphs.
 - Can be described as combinations of two or more attributes.

Fundamental types of visualizations

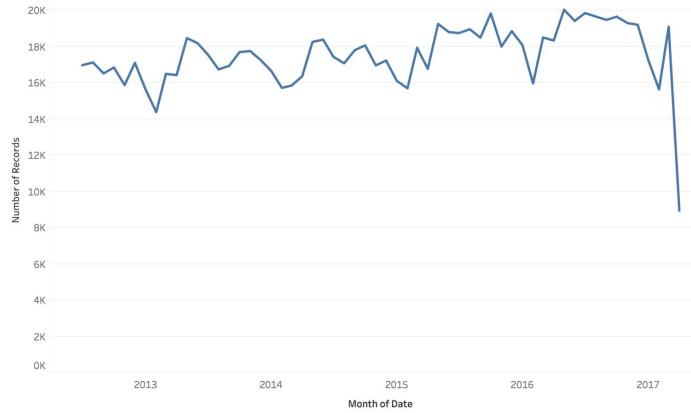
Bar chart



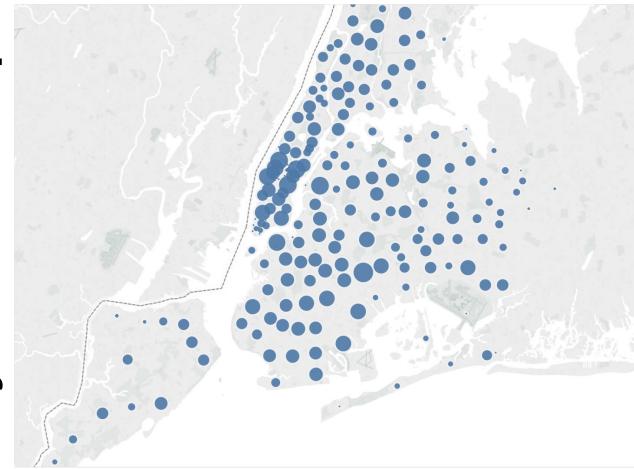
Scatter plot



Line chart



Symbol map

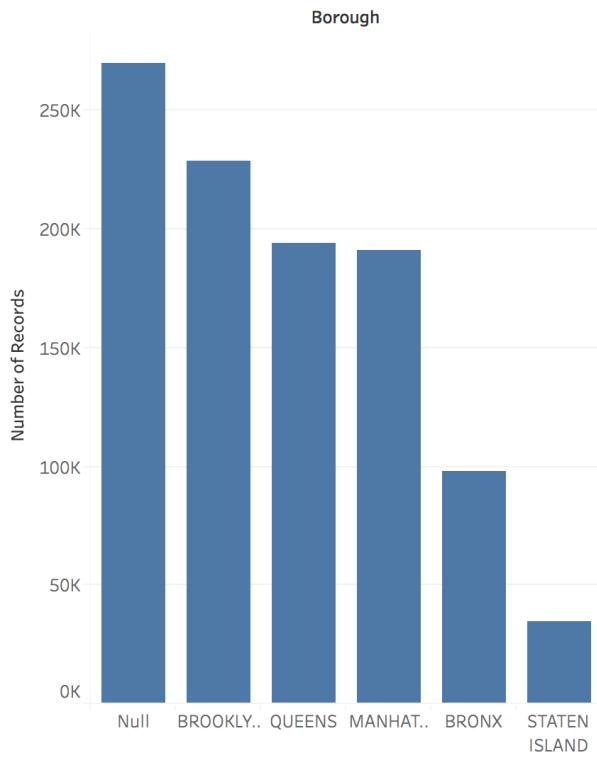


Matrix



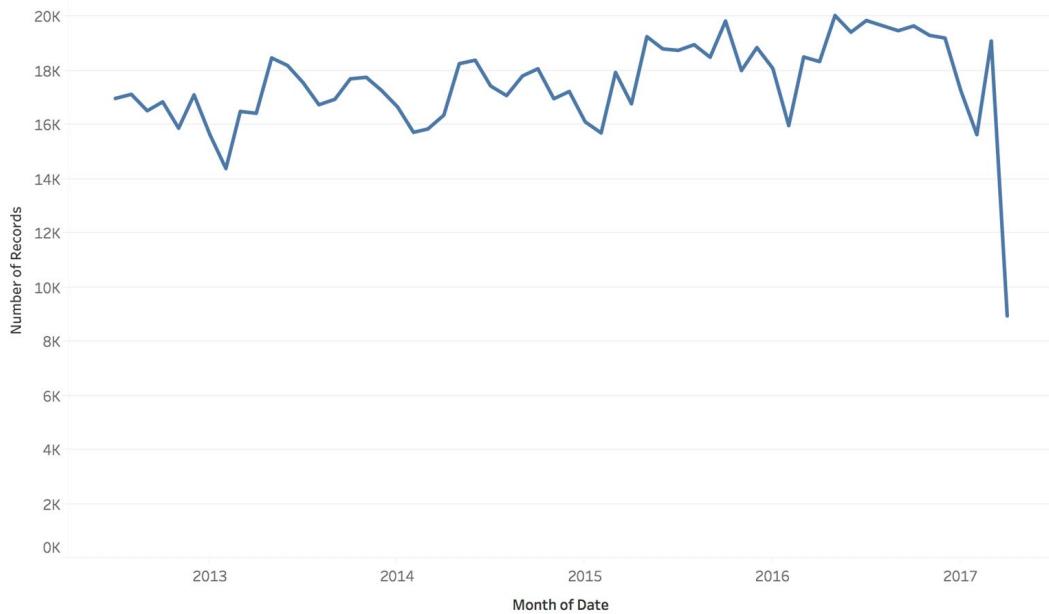
[Enrico Bertini]

Bar chart



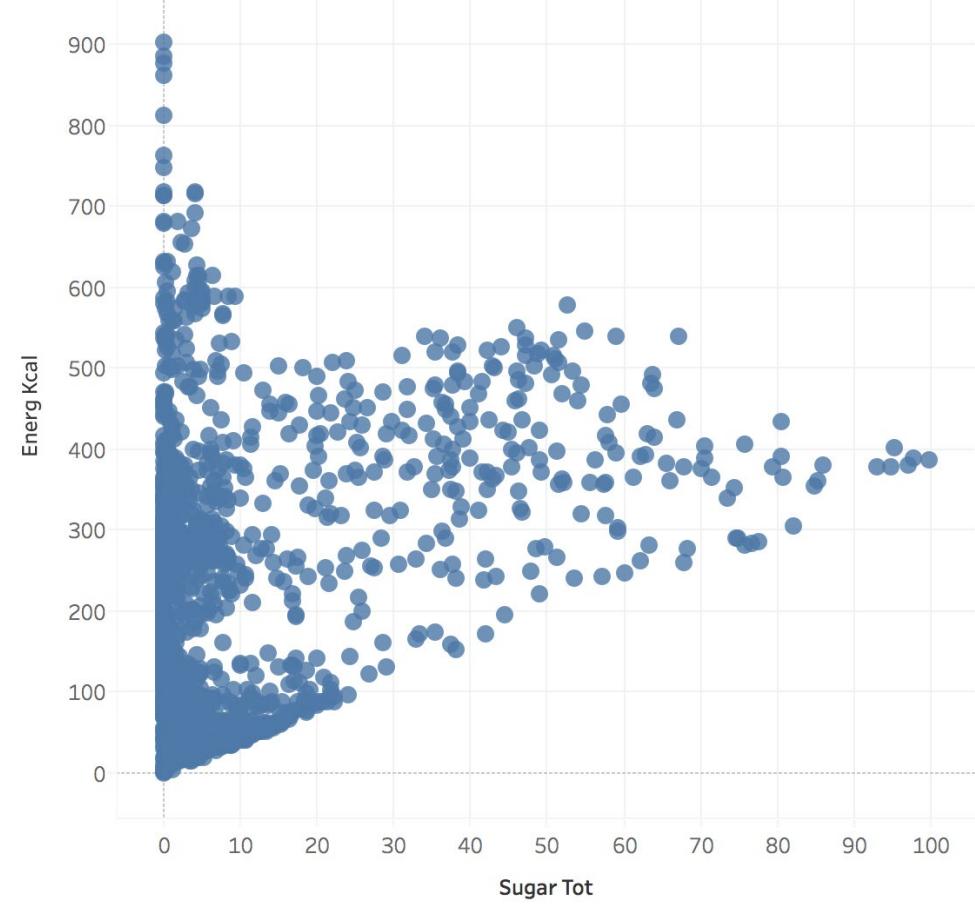
Visualize how a **quantity** distributes across a set of **categories**

Line chart



Visualize how a **quantity** changes in relation to another **quantity** (e.g., time)

Scatter plot



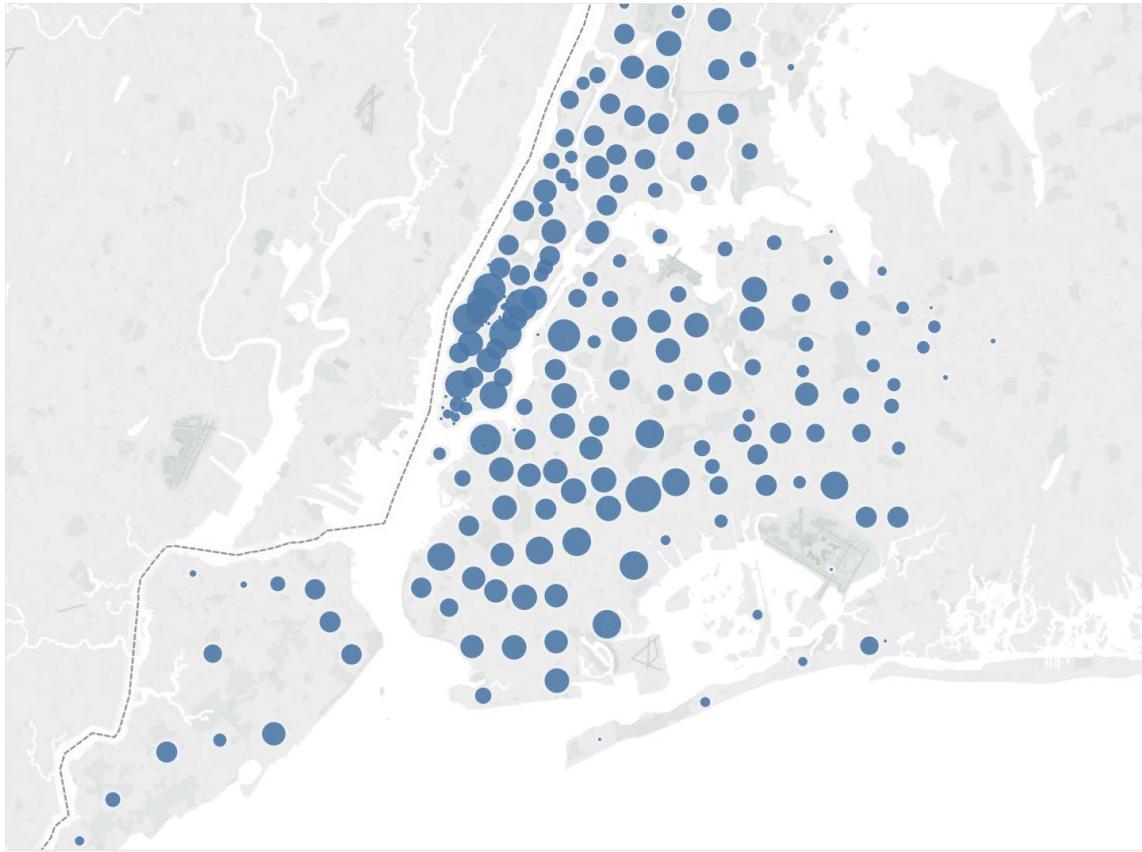
Visualize how a **quantity** relates to another **quantity**

Matrix

Vehicle Type Code 2	Vehicle Type Code 1									
	AMBULA..	BICYCLE	BUS	FIRE TRUCK	LARGE COM VE..	LIVERY VEHICLE	MOTORC..	OTHER	PASSENG ER VEH..	PEDICAB
AMBULANCE
BICYCLE
BUS
FIRE TRUCK
LARGE COM..
LIVERY VEH..
MOTORCYC..
OTHER
PASSENGE..
PEDICAB
PICK-UP TR..
SCOOTER
SMALL COM..
SPORT UTIL..
TAXI
UNKNOWN
VAN

Visualize how a **quantity** distributes across two **categories**

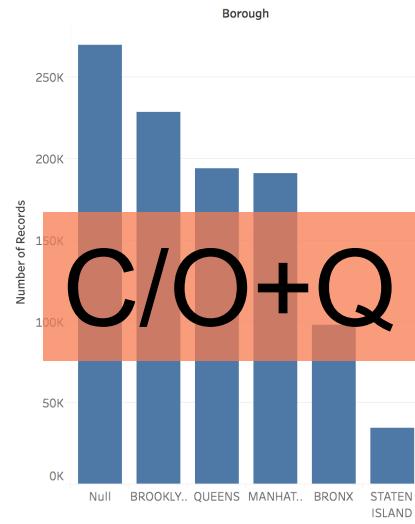
Matrix



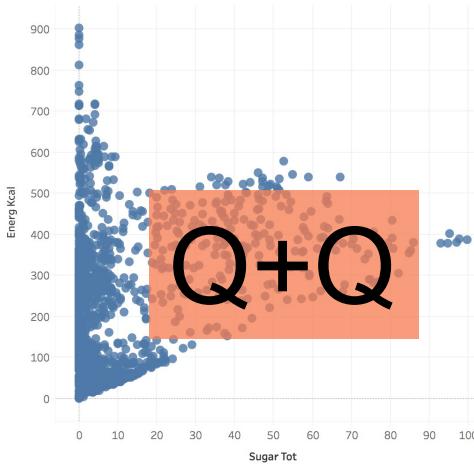
Visualize how a **quantity** distributes
across two **spatial coordinates**

Fundamental types of visualizations

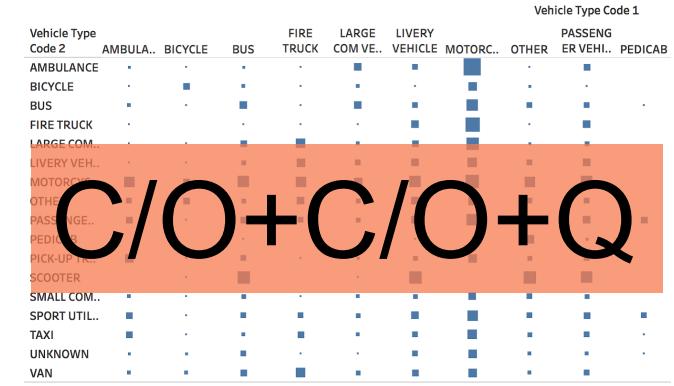
Bar chart



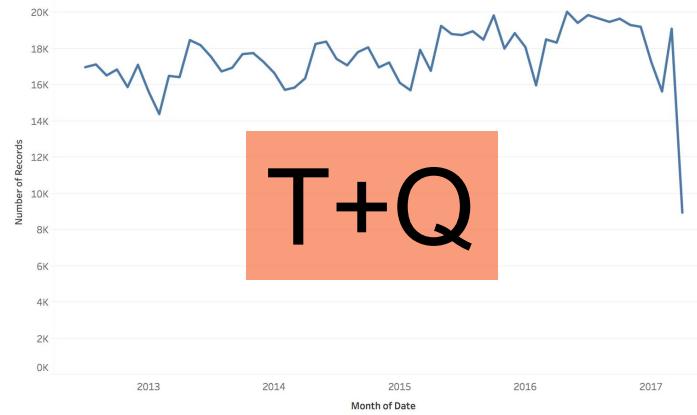
Scatter plot



Matrix



Line chart



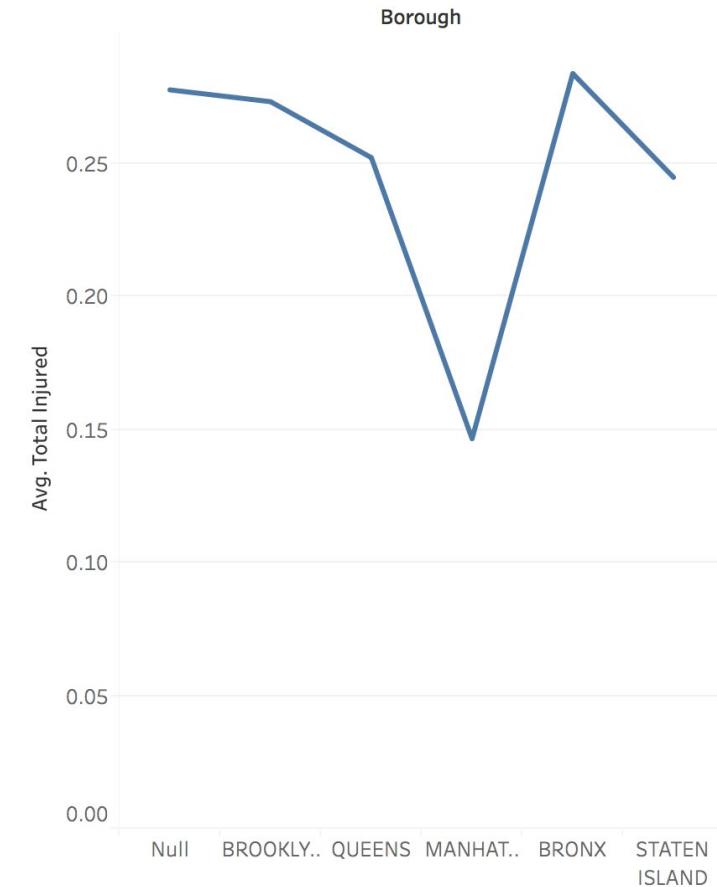
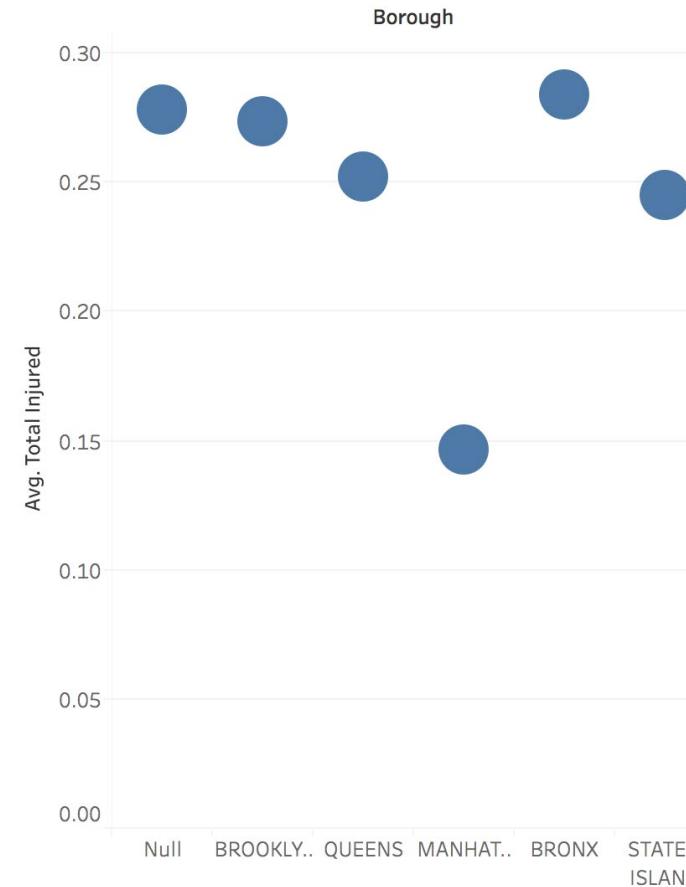
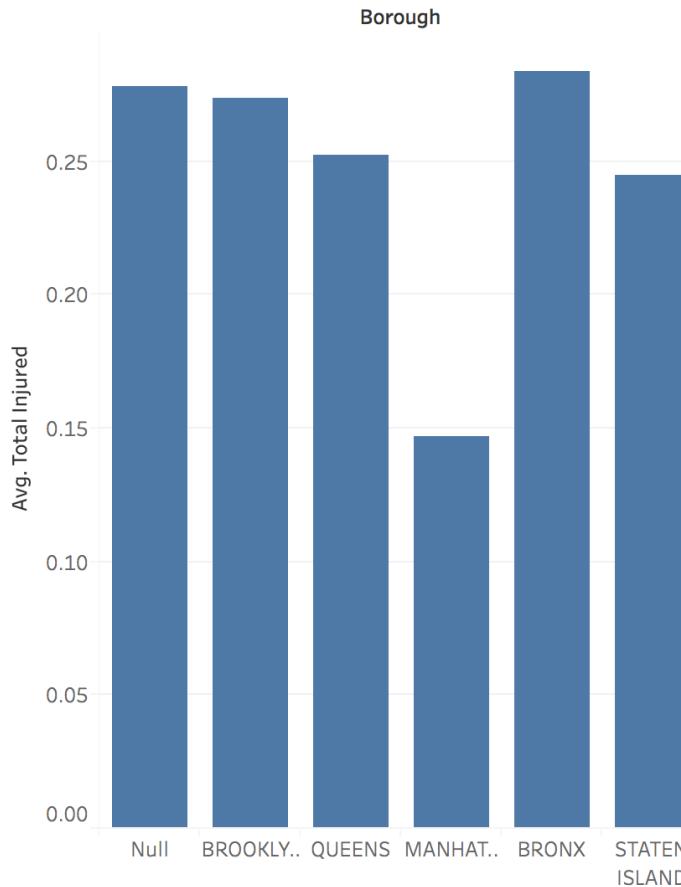
Symbol map



Alternate representations

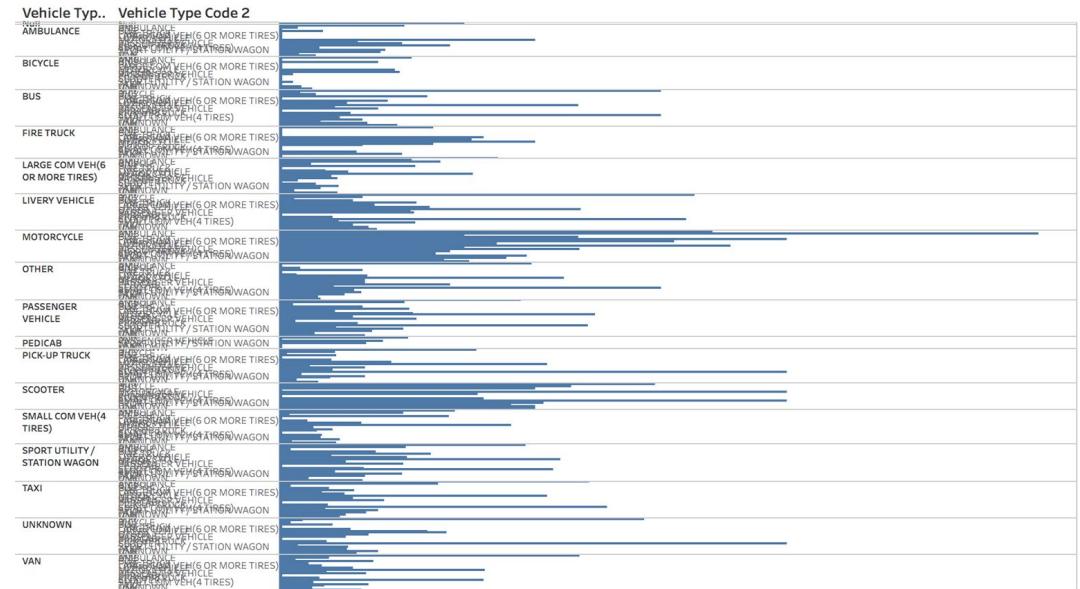
- Creating different representations of the same data can be useful to reason about graphical encodings and advantage / disadvantages of alternate designs.

Bar chart alternatives



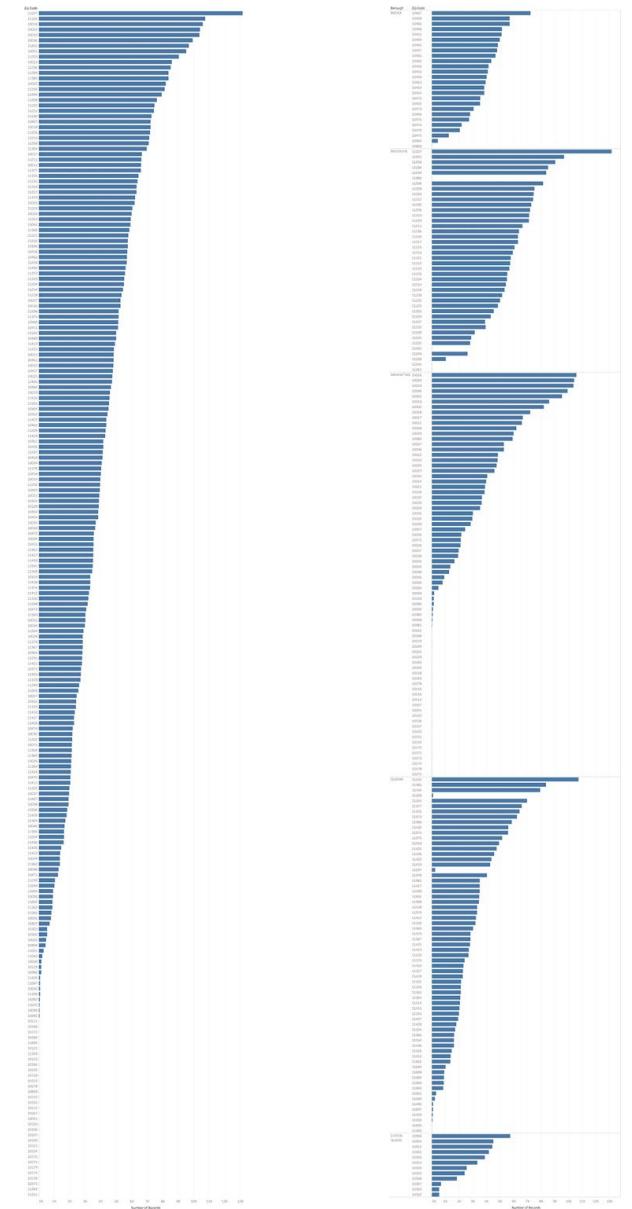
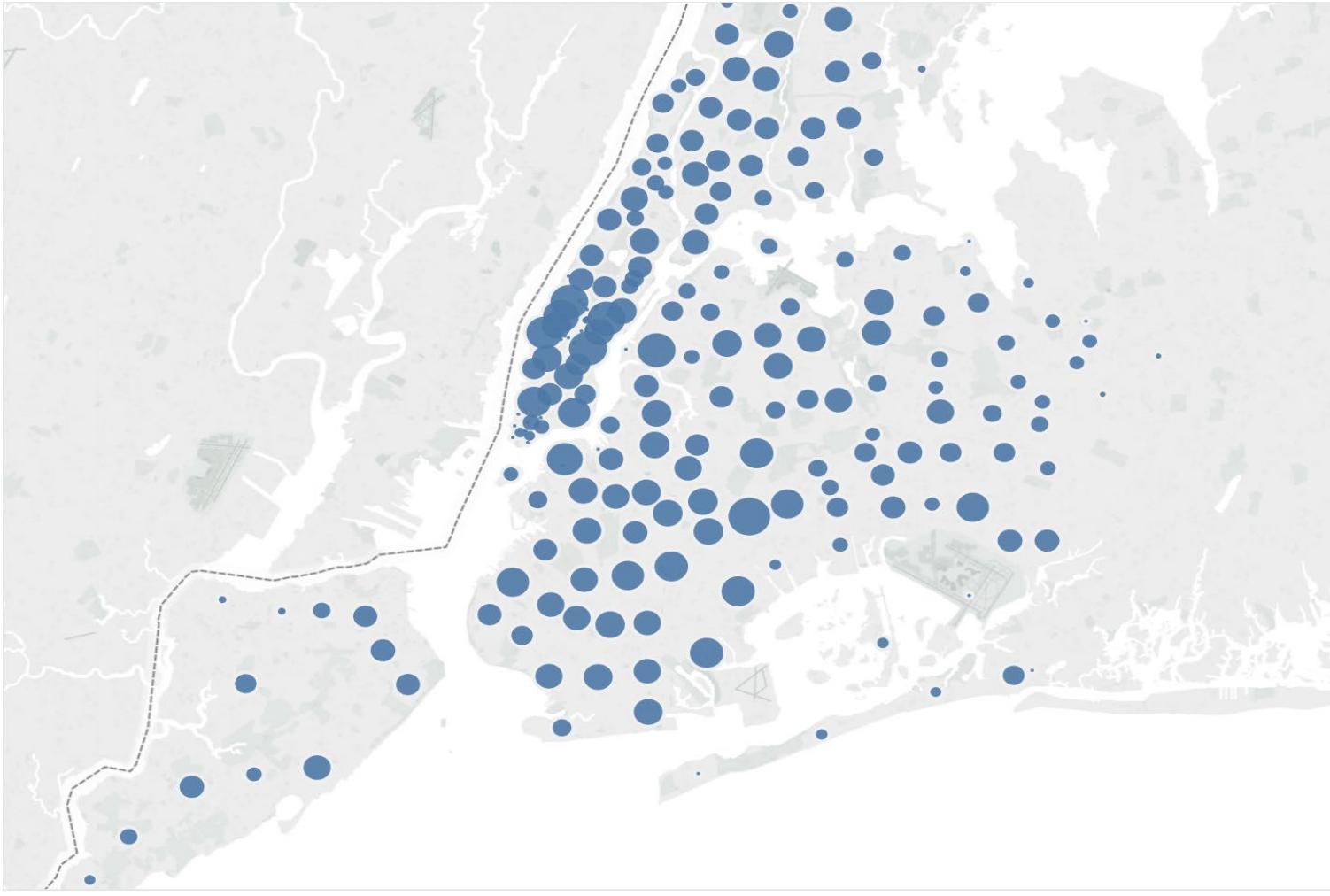
[Enrico Bertini]

Matrix alternatives



[Enrico Bertini]

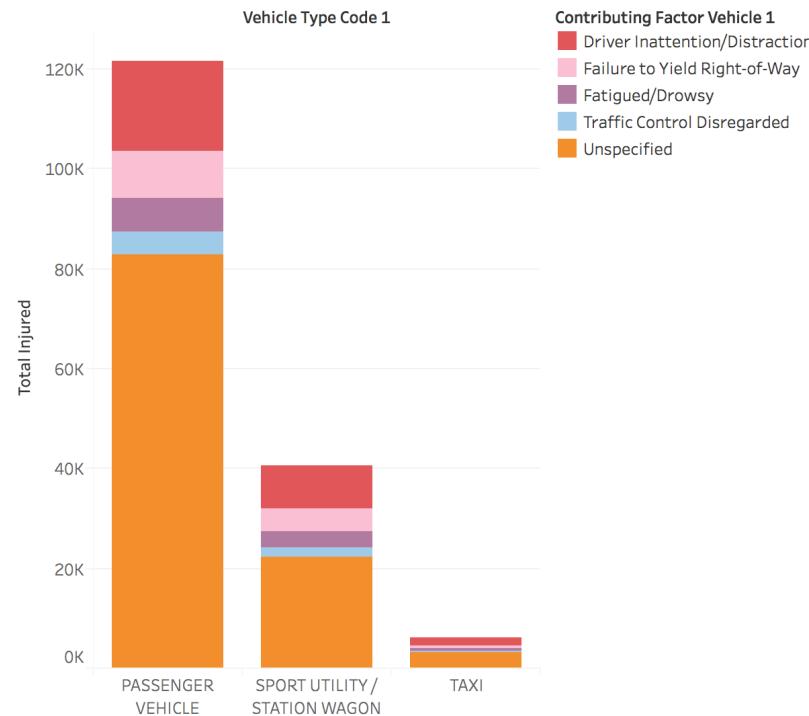
Symbol map alternatives



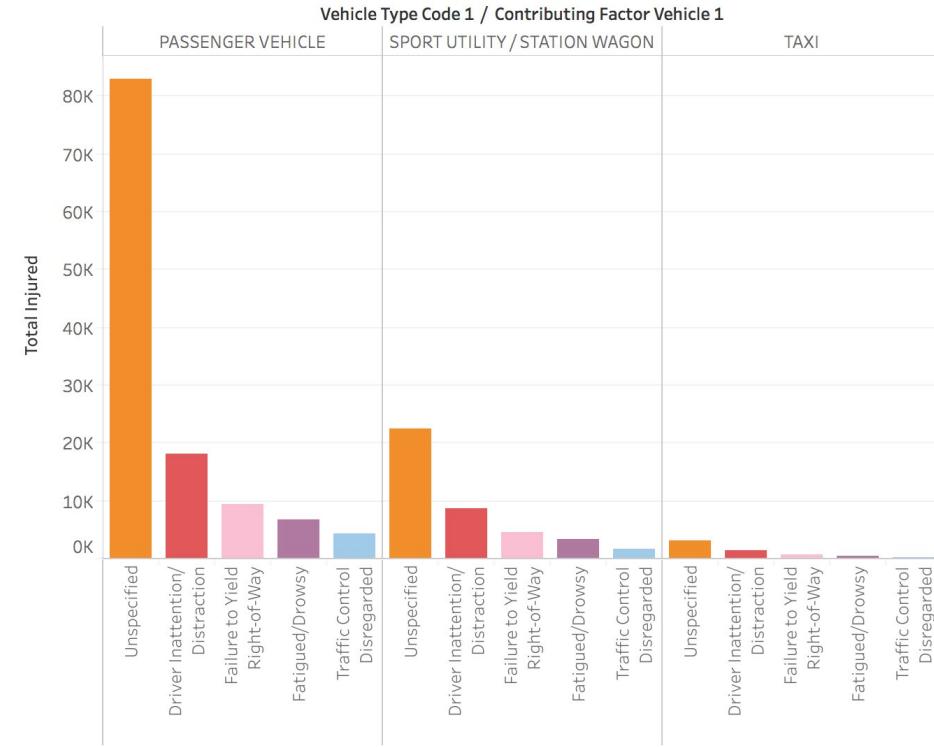
Beyond two attributes

- What if we need to visualize more than two attributes at the same time?

Bar chart



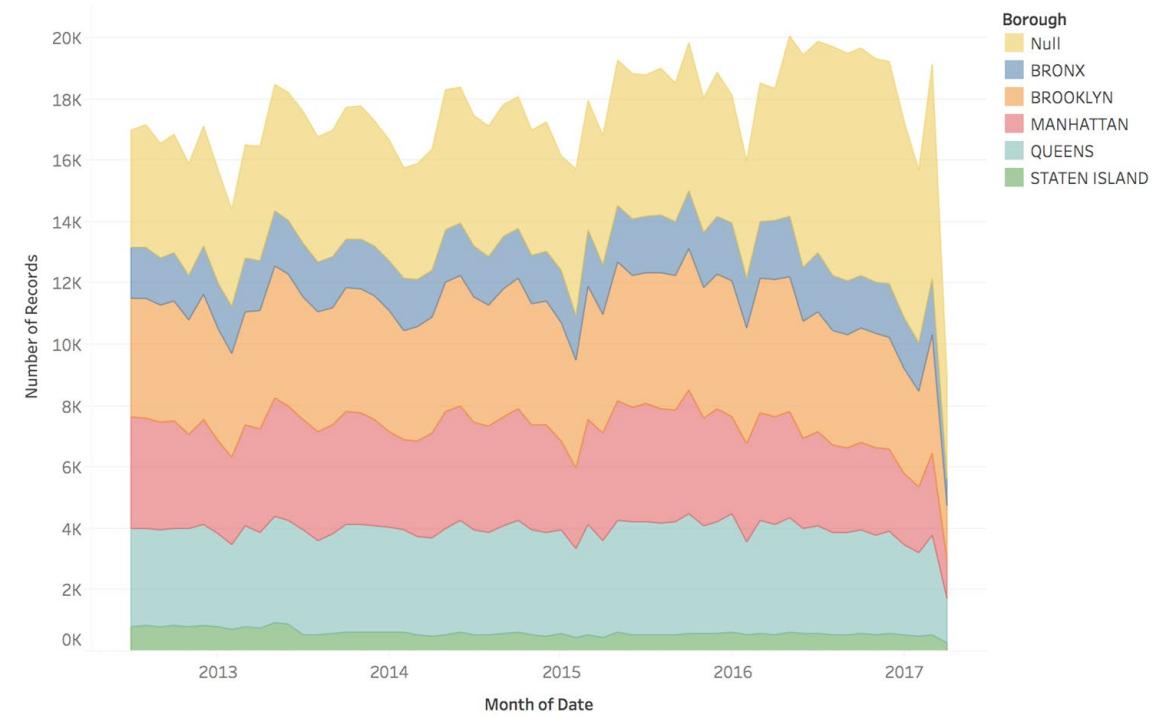
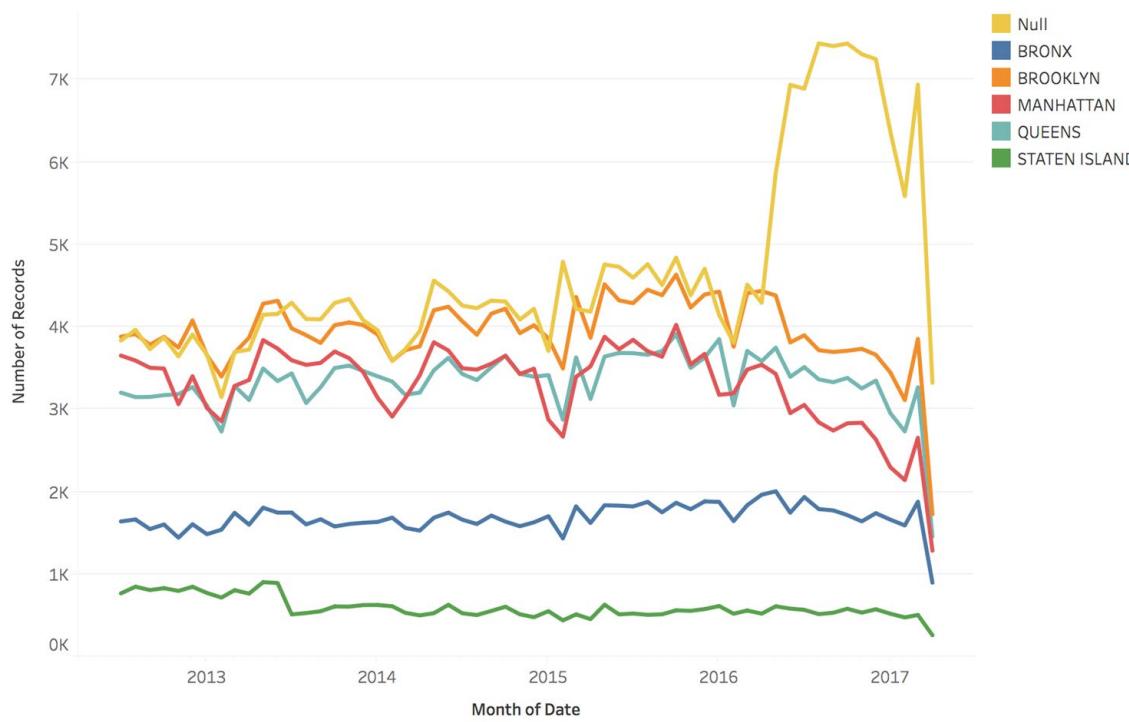
Stacked



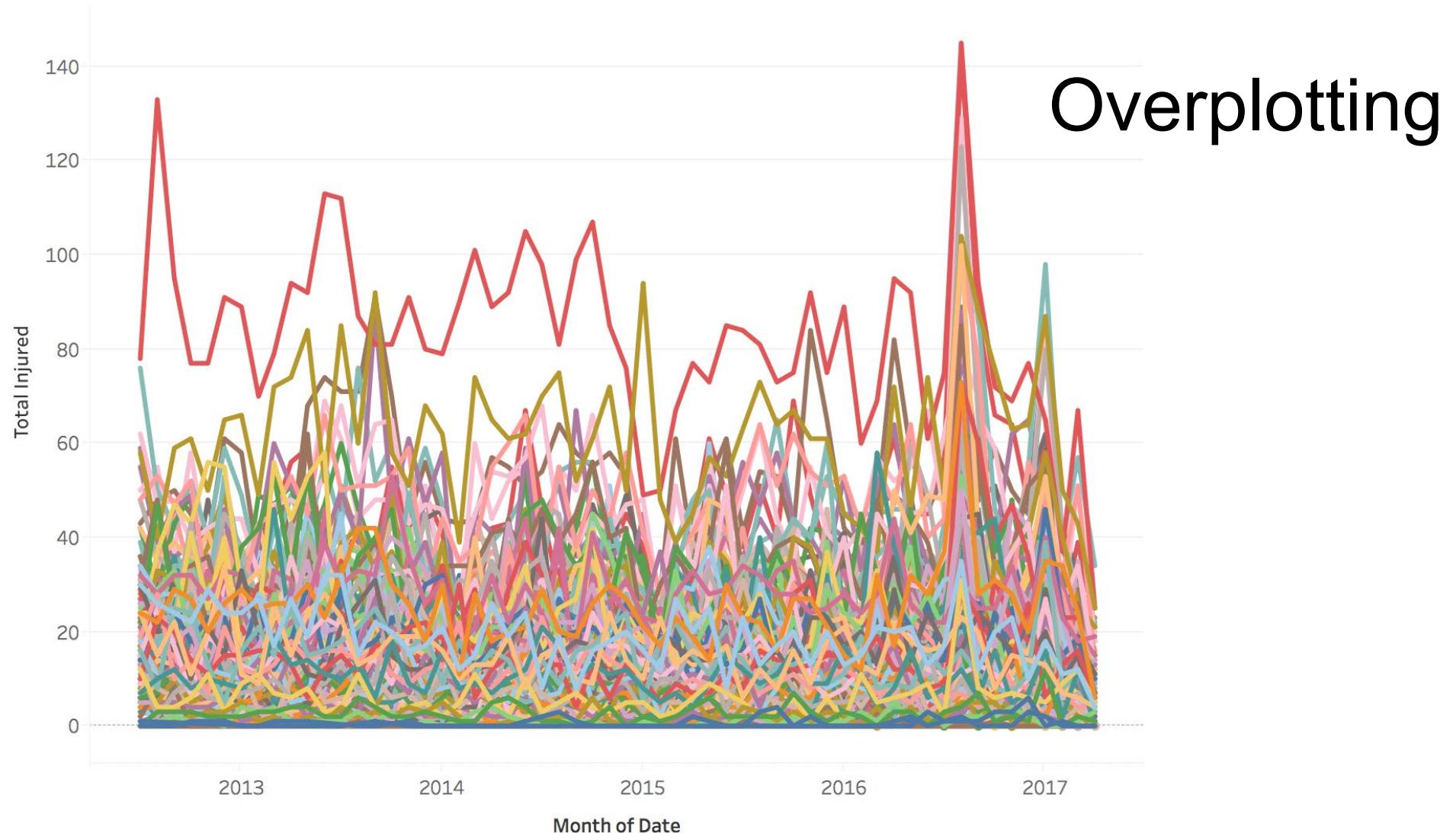
Grouped

[Enrico Bertini]

Line chart



Line chart



Faceting

- How to facet data across multiple views?
 - Juxtapose them side by side.
 - Superimpose the views as layers on top of each other.

Faceting + scatter plots



[Enrico Bertini]

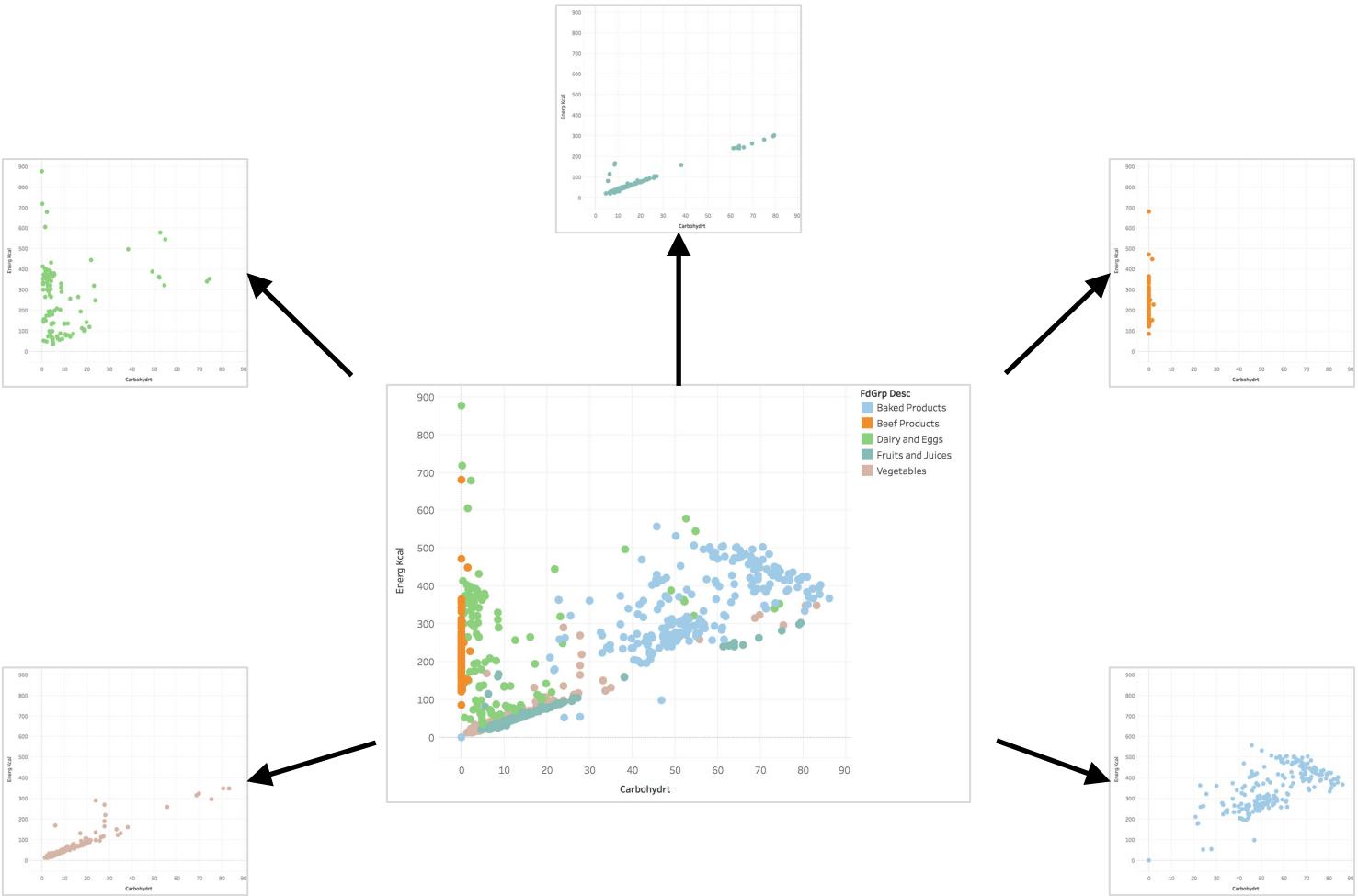
Faceting + scatter plots



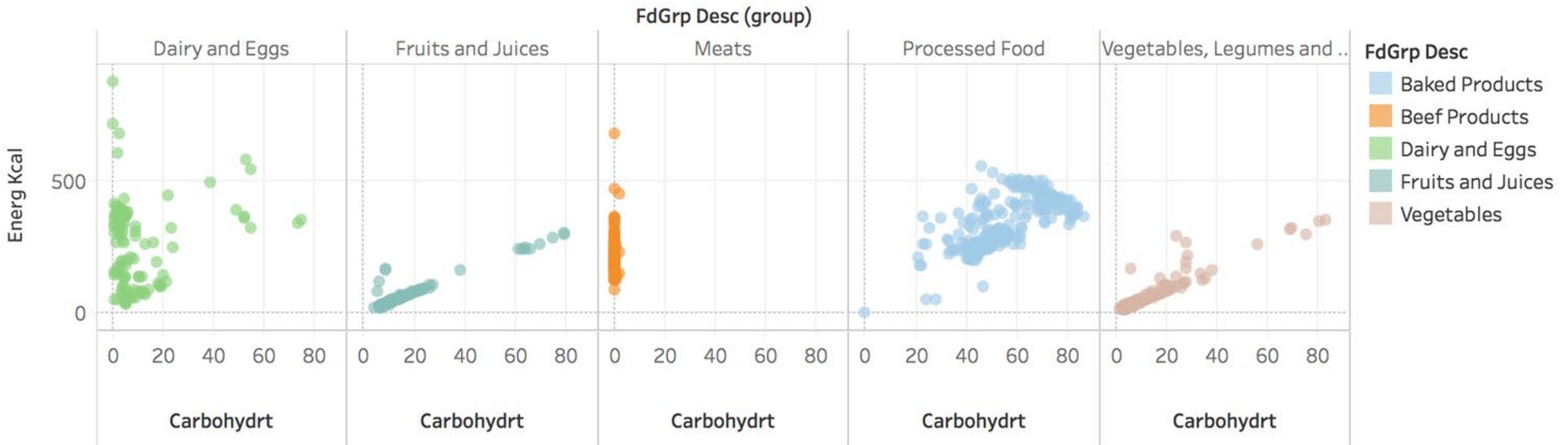
Faceting

1. Select one categorial / ordinal attribute.
2. Create as many sets as the number of values.
3. Create one plot for each value.

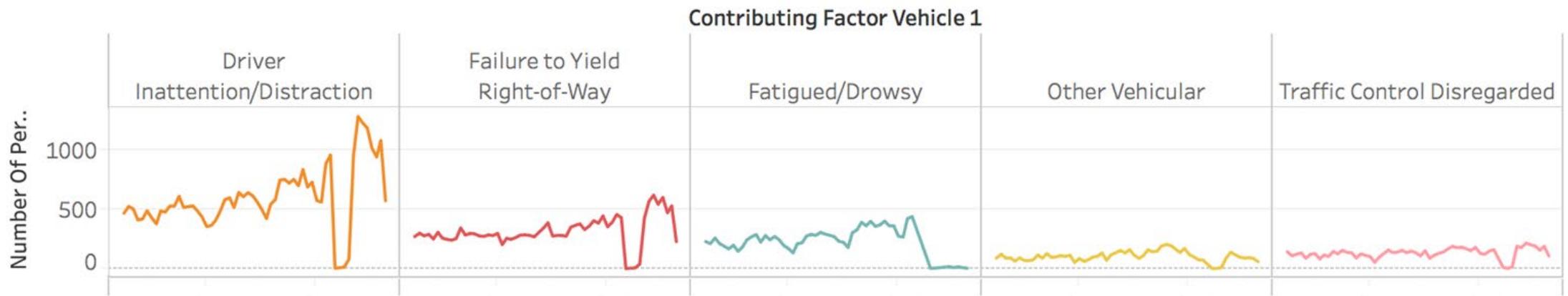
Faceting + scatter plots



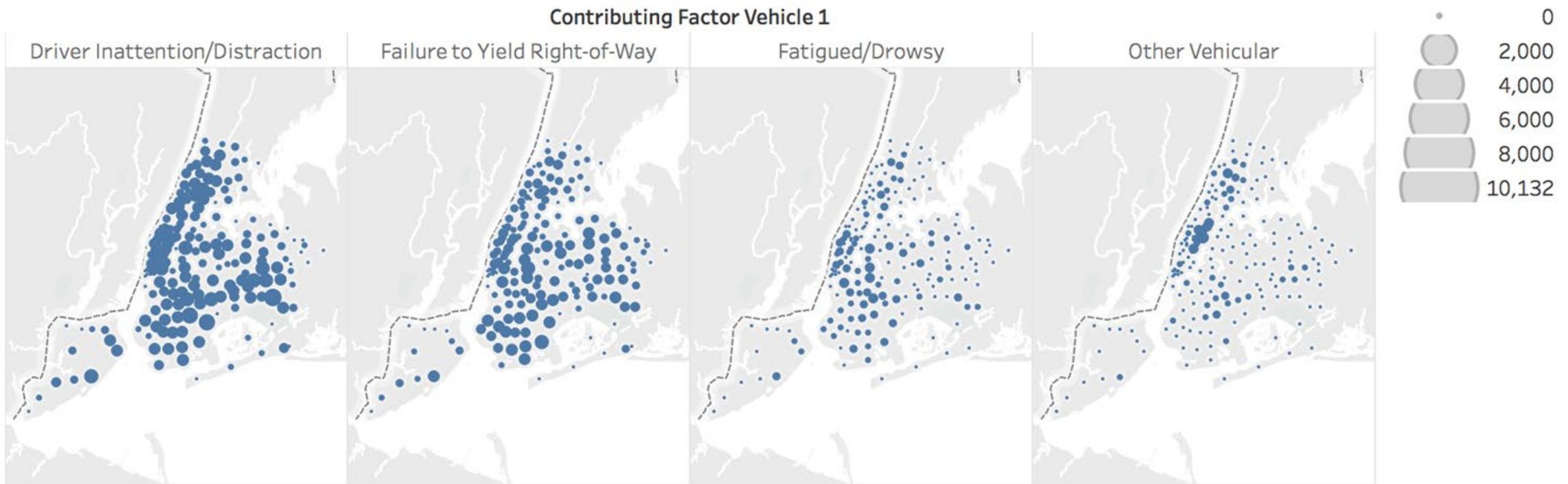
Faceting + scatter plots



Faceting + line chart

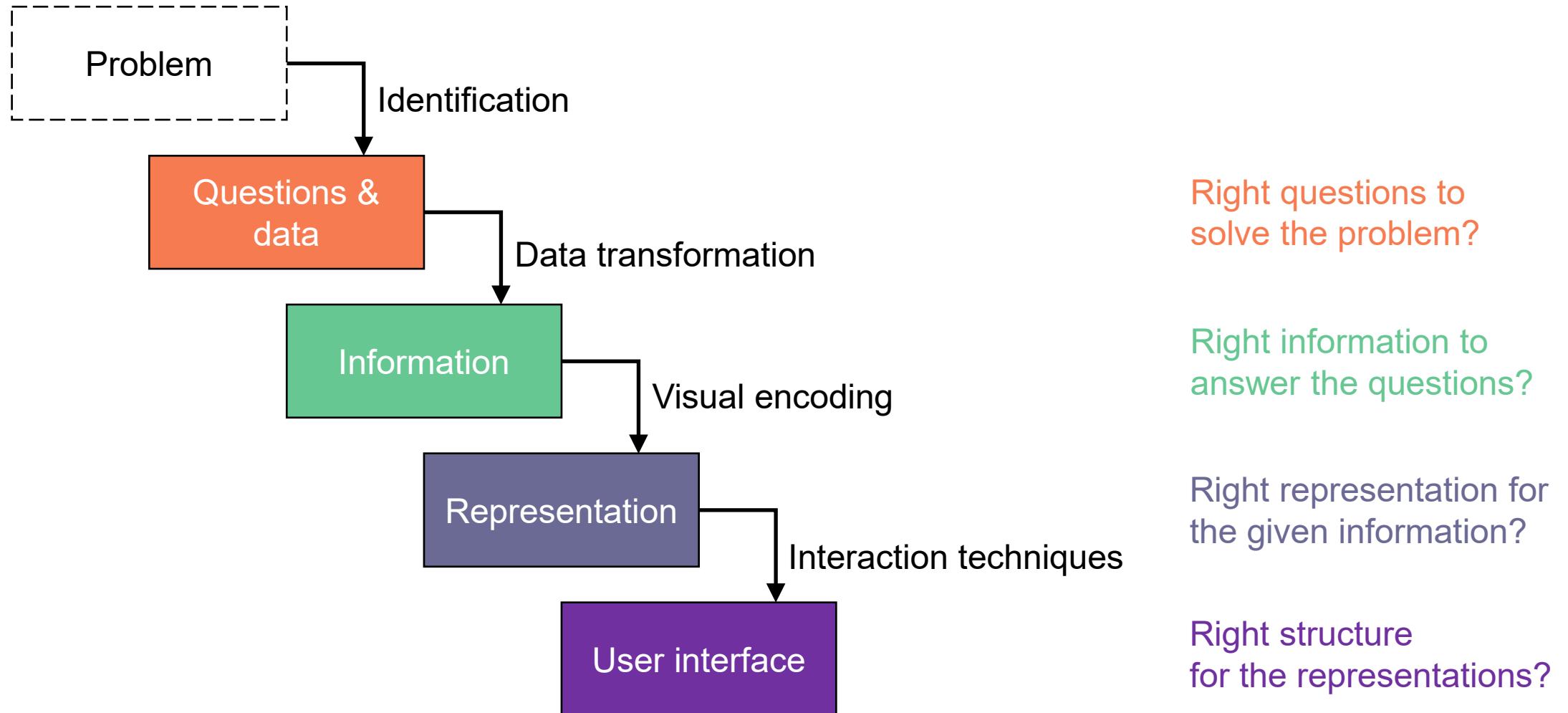


Faceting + maps



[Enrico Bertini]

Creating visualizations



How to visualize data?

1. Decide what to visualize → select & transform data
2. Decide how to visualize it → choose or design representations

Data transformation

- Selection
- Aggregation:
 - Temporal
 - Spatial
 - Binning

Data transformation: selection

Selection: select from the table the attributes needed for the visualization.

1) Food dataset: visualize the relationship between **carbs** and **calories** and see how it is affected by food category.

B	C	D	E	F	G	H	I	J	K	L	M
NDB_No	FdGrp_CD	FdGrp_Desc	Shrt_Desc	Water	Energ_Kcal	Protein	Lipid_Tot	Ash	Carbohydrt	Fiber_TD	Sugar_Tot
1001	100	Dairy and Egg	BUTTER,WIT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1002	100	Dairy and Egg	BUTTER,WH	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1003	100	Dairy and Egg	BUTTER OIL,	0.24	876	0.28	99.48	0	0	0	0
1004	100	Dairy and Egg	CHEESE,BLU	42.41	353	21.4	28.74	5.11	2.34	0	0.5
1005	100	Dairy and Egg	CHEESE,BRIC	41.11	371	23.24	29.68	3.18	2.79	0	0.51
1006	100	Dairy and Egg	CHEESE,BRIG	48.42	334	20.75	27.68	2.7	0.45	0	0.45
1007	100	Dairy and Egg	CHEESE,CAN	51.8	300	19.8	24.26	3.68	0.46	0	0.46
1009	100	Dairy and Egg	CHEESE,CHE	36.75	403	24.9	33.14	3.93	1.28	0	0.52
1011	100	Dairy and Egg	CHEESE,COL	38.2	394	23.76	32.11	3.36	2.57	0	0.52
1012	100	Dairy and Egg	CHEESE,COT	79.79	98	11.12	4.3	1.41	3.38	0	2.67
1013	100	Dairy and Egg	CHEESE,COT	79.64	97	10.69	3.85	1.2	4.61	0.2	2.38
1014	100	Dairy and Egg	CHEESE,COT	81.01	72	10.34	0.29	1.71	6.66	0	1.85
1015	100	Dairy and Egg	CHEESE,COT	80.69	86	11.83	2.45	1.36	3.66	0	3.67
1016	100	Dairy and Egg	CHEESE,COT	82.48	72	12.39	1.02	1.39	2.72	0	2.72
1017	100	Dairy and Egg	CHEESE,CRE	54.44	342	5.93	34.24	1.32	4.07	0	3.21
1018	100	Dairy and Egg	CHEESE,EDA	41.56	357	24.99	27.8	4.22	1.43	0	1.43
1019	100	Dairy and Egg	CHEESE,FET	55.22	264	14.21	21.28	5.2	4.09	0	4.09
1020	100	Dairy and Egg	CHEESE,FON	37.92	389	25.6	31.14	3.79	1.55	0	1.55
1022	100	Dairy and Egg	CHEESE,GOL	41.46	356	24.94	27.44	3.94	2.22	0	2.22
1023	100	Dairy and Egg	CHEESE,GRL	33.19	413	29.81	32.34	4.3	0.36	0	0.36
1024	100	Dairy and Egg	CHEESE,LIM	48.42	327	20.05	27.25	3.79	0.49	0	0.49
1025	100	Dairy and Egg	CHEESE,MO	41.01	373	24.48	30.28	3.55	0.68	0	0.5
1026	100	Dairy and Egg	CHEESE,MO	50.01	300	22.17	22.35	3.28	2.19	0	1.03
1027	100	Dairy and Egg	CHEESE,MO	48.38	318	21.6	24.64	2.91	2.47	0	1.01
1029	100	Dairy and Egg	CHEESE,MO	46.46	302	25.96	20.03	3.72	3.83	0	0.6
1030	100	Dairy and Egg	CHEESE,MU	41.77	368	23.41	30.04	3.66	1.12	0	1.12
1032	100	Dairy and Egg	CHEESE,PAR	20.84	431	38.46	28.61	8.03	4.06	0	0.9
1033	100	Dairy and Egg	CHEESE,PAR	29.16	392	35.75	25.83	6.04	3.22	0	0.8
1034	100	Dairy and Egg	CHEESE,POR	45.45	352	23.78	28.2	2	0.57	0	0.57
1035	100	Dairy and Egg	CHEESE,PRC	40.95	351	25.58	26.62	4.71	2.14	0	0.56
1036	100	Dairy and Egg	CHEESE,RICO	71.7	174	11.26	12.98	1.02	3.04	0	0.27
1037	100	Dairy and Egg	CHEESE,RICO	74.41	138	11.39	7.91	1.15	5.14	0	0.31
1038	100	Dairy and Egg	CHEESE,RON	30.91	387	31.8	26.94	6.72	3.63	0	0.73
1040	100	Dairy and Egg	CHEESE,SWI	37.12	380	26.93	27.8	2.77	5.38	0	1.32
1042	100	Dairy and Egg	CHEESE,PAS	39.61	366	18.13	30.71	6.78	4.78	0	2.26
1043	100	Dairy and Egg	CHEESE,PAS	39.08	375	22.13	31.2	5.84	1.73	0.1	0.62
1044	100	Dairy and Egg	CHEESE,PAS	42.31	334	24.73	25.01	5.85	2.1	0	1.23
1046	100	Dairy and Egg	CHEESE FD,FL	44	330	16.86	25.63	4.96	8.56	0	5.59
1048	100	Dairy and Egg	CHEESE SPR	47.65	290	16.41	21.23	5.98	8.73	0	7.32
1049	100	Dairy and Egg	CREAM,FLUI	80.57	130	2.96	11.5	0.67	4.3	0	0.16
1050	100	Dairy and Egg	CREAM,FLUI	73.75	195	2.7	19.31	0.58	3.66	0	0.14
1052	100	Dairy and Egg	CREAM,FLUI	63.5	292	2.17	30.91	0.46	2.96	0	0.11
1053	100	Dairy and Egg	CREAM,FLUI	57.71	345	2.05	37	0.45	2.79	0	0.11

Data transformation: selection

Selection: select from the table the attributes needed for the visualization.

1) Food dataset: visualize the relationship between **carbs** and **calories** and see how it is affected by food category.

Q+Q+C

B	C	D	E	F	G	H	I	J	K	L	M
NDB_No	FdGrp_CD	FdGrp_Desc	Shrt_Desc	Water	Energ_Kcal	Protein	Lipid_Tot	Ash	Carbohydrt	Fiber_TD	Sugar_Tot
1001	100	Dairy and Egg	BUTTER,WIT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1002	100	Dairy and Egg	BUTTER,WH	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1003	100	Dairy and Egg	BUTTER OIL,	0.24	876	0.28	99.48	0	0	0	0
1004	100	Dairy and Egg	CHEESE,BLU	42.41	353	21.4	28.74	5.11	2.34	0	0.5
1005	100	Dairy and Egg	CHEESE,BRIC	41.11	371	23.24	29.68	3.18	2.79	0	0.51
1006	100	Dairy and Egg	CHEESE,BRIG	48.42	334	20.75	27.68	2.7	0.45	0	0.45
1007	100	Dairy and Egg	CHEESE,CAN	51.8	300	19.8	24.26	3.68	0.46	0	0.46
1009	100	Dairy and Egg	CHEESE,CHE	36.75	403	24.9	33.14	3.93	1.28	0	0.52
1011	100	Dairy and Egg	CHEESE,COL	38.2	394	23.76	32.11	3.36	2.57	0	0.52
1012	100	Dairy and Egg	CHEESE,COT	79.79	98	11.12	4.3	1.41	3.38	0	2.67
1013	100	Dairy and Egg	CHEESE,COT	79.64	97	10.69	3.85	1.2	4.61	0.2	2.38
1014	100	Dairy and Egg	CHEESE,COT	81.01	72	10.34	0.29	1.71	6.66	0	1.85
1015	100	Dairy and Egg	CHEESE,COT	80.69	86	11.83	2.45	1.36	3.66	0	3.67
1016	100	Dairy and Egg	CHEESE,COT	82.48	72	12.39	1.02	1.39	2.72	0	2.72
1017	100	Dairy and Egg	CHEESE,CRE	54.44	342	5.93	34.24	1.32	4.07	0	3.21
1018	100	Dairy and Egg	CHEESE,EDA	41.56	357	24.99	27.8	4.22	1.43	0	1.43
1019	100	Dairy and Egg	CHEESE,FET	55.22	264	14.21	21.28	5.2	4.09	0	4.09
1020	100	Dairy and Egg	CHEESE,FON	37.92	389	25.6	31.14	3.79	1.55	0	1.55
1022	100	Dairy and Egg	CHEESE,GOU	41.46	356	24.94	27.44	3.94	2.22	0	2.22
1023	100	Dairy and Egg	CHEESE,GRU	33.19	413	29.81	32.34	4.3	0.36	0	0.36
1024	100	Dairy and Egg	CHEESE,LIM	48.42	327	20.05	27.25	3.79	0.49	0	0.49
1025	100	Dairy and Egg	CHEESE,MO	41.01	373	24.48	30.28	3.55	0.68	0	0.5
1026	100	Dairy and Egg	CHEESE,MO	50.01	300	22.17	22.35	3.28	2.19	0	1.03
1027	100	Dairy and Egg	CHEESE,MO	48.38	318	21.6	24.64	2.91	2.47	0	1.01
1029	100	Dairy and Egg	CHEESE,MO	46.46	302	25.96	20.03	3.72	3.83	0	0.6
1030	100	Dairy and Egg	CHEESE,MU	41.77	368	23.41	30.04	3.66	1.12	0	1.12
1032	100	Dairy and Egg	CHEESE,PAR	20.84	431	38.46	28.61	8.03	4.06	0	0.9
1033	100	Dairy and Egg	CHEESE,PAR	29.16	392	35.75	25.83	6.04	3.22	0	0.8
1034	100	Dairy and Egg	CHEESE,POR	45.45	352	23.78	28.2	2	0.57	0	0.57
1035	100	Dairy and Egg	CHEESE,PRC	40.95	351	25.58	26.62	4.71	2.14	0	0.56
1036	100	Dairy and Egg	CHEESE,RICO	71.7	174	11.26	12.98	1.02	3.04	0	0.27
1037	100	Dairy and Egg	CHEESE,RICO	74.41	138	11.39	7.91	1.15	5.14	0	0.31
1038	100	Dairy and Egg	CHEESE,ROM	30.91	387	31.8	26.94	6.72	3.63	0	0.73
1040	100	Dairy and Egg	CHEESE,SWI	37.12	380	26.93	27.8	2.77	5.38	0	1.32
1042	100	Dairy and Egg	CHEESE,PAS	39.61	366	18.13	30.71	6.78	4.78	0	2.26
1043	100	Dairy and Egg	CHEESE,PAS	39.08	375	22.13	31.2	5.84	1.73	0.1	0.62
1044	100	Dairy and Egg	CHEESE,PAS	42.31	334	24.73	25.01	5.85	2.1	0	1.23
1046	100	Dairy and Egg	CHEESE FD,FL	44	330	16.86	25.63	4.96	8.56	0	5.59
1048	100	Dairy and Egg	CHEESE SPR	47.65	290	16.41	21.23	5.98	8.73	0	7.32
1049	100	Dairy and Egg	CREAM,FLUI	80.57	130	2.96	11.5	0.67	4.3	0	0.16
1050	100	Dairy and Egg	CREAM,FLUI	73.75	195	2.7	19.31	0.58	3.66	0	0.14
1052	100	Dairy and Egg	CREAM,FLUI	63.5	292	2.17	30.91	0.46	2.96	0	0.11
1053	100	Dairy and Egg	CREAM,FLUI	57.71	345	2.05	37	0.45	2.79	0	0.11

Data transformation: selection

Selection: select from the table the attributes needed for the visualization.

1) Food dataset: visualize the relationship between **carbs** and **calories** and see how it is affected by food category.

Q+Q+C



Data transformation: selection

Selection: select from the table the attributes needed for the visualization.

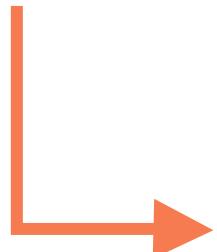
2) Food dataset: visualize how the average amount of calories distributes across food categories.

B	C	D	E	F	G	H	I	J	K	L	M
NDB_No	FdGrp_CD	FdGrp_Desc	Shrt_Desc	Water	Energ_Kcal	Protein	Lipid_Tot	Ash	Carbohydrt	Fiber_TD	Sugar_Tot
1001	100	Dairy and Egg	BUTTER,WIT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1002	100	Dairy and Egg	BUTTER,WH	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1003	100	Dairy and Egg	BUTTER OIL,	0.24	876	0.28	99.48	0	0	0	0
1004	100	Dairy and Egg	CHEESE,BLU	42.41	353	21.4	28.74	5.11	2.34	0	0.5
1005	100	Dairy and Egg	CHEESE,BRIC	41.11	371	23.24	29.68	3.18	2.79	0	0.51
1006	100	Dairy and Egg	CHEESE,BRIG	48.42	334	20.75	27.68	2.7	0.45	0	0.45
1007	100	Dairy and Egg	CHEESE,CAN	51.8	300	19.8	24.26	3.68	0.46	0	0.46
1009	100	Dairy and Egg	CHEESE,CHE	36.75	403	24.9	33.14	3.93	1.28	0	0.52
1011	100	Dairy and Egg	CHEESE,COL	38.2	394	23.76	32.11	3.36	2.57	0	0.52
1012	100	Dairy and Egg	CHEESE,COT	79.79	98	11.12	4.3	1.41	3.38	0	2.67
1013	100	Dairy and Egg	CHEESE,COT	79.64	97	10.69	3.85	1.2	4.61	0.2	2.38
1014	100	Dairy and Egg	CHEESE,COT	81.01	72	10.34	0.29	1.71	6.66	0	1.85
1015	100	Dairy and Egg	CHEESE,COT	80.69	86	11.83	2.45	1.36	3.66	0	3.67
1016	100	Dairy and Egg	CHEESE,COT	82.48	72	12.39	1.02	1.39	2.72	0	2.72
1017	100	Dairy and Egg	CHEESE,CRE	54.44	342	5.93	34.24	1.32	4.07	0	3.21
1018	100	Dairy and Egg	CHEESE,EDA	41.56	357	24.99	27.8	4.22	1.43	0	1.43
1019	100	Dairy and Egg	CHEESE,FET,	55.22	264	14.21	21.28	5.2	4.09	0	4.09
1020	100	Dairy and Egg	CHEESE,FON	37.92	389	25.6	31.14	3.79	1.55	0	1.55
1022	100	Dairy and Egg	CHEESE,GOU	41.46	356	24.94	27.44	3.94	2.22	0	2.22
1023	100	Dairy and Egg	CHEESE,GRL	33.19	413	29.81	32.34	4.3	0.36	0	0.36
1024	100	Dairy and Egg	CHEESE,LIM	48.42	327	20.05	27.25	3.79	0.49	0	0.49
1025	100	Dairy and Egg	CHEESE,MO	41.01	373	24.48	30.28	3.55	0.68	0	0.5
1026	100	Dairy and Egg	CHEESE,MO	50.01	300	22.17	22.35	3.28	2.19	0	1.03
1027	100	Dairy and Egg	CHEESE,MO	48.38	318	21.6	24.64	2.91	2.47	0	1.01
1029	100	Dairy and Egg	CHEESE,MO	46.46	302	25.96	20.03	3.72	3.83	0	0.6
1030	100	Dairy and Egg	CHEESE,MU	41.77	368	23.41	30.04	3.66	1.12	0	1.12
1032	100	Dairy and Egg	CHEESE,PAR	20.84	431	38.46	28.61	8.03	4.06	0	0.9
1033	100	Dairy and Egg	CHEESE,PAR	29.16	392	35.75	25.83	6.04	3.22	0	0.8
1034	100	Dairy and Egg	CHEESE,POR	45.45	352	23.78	28.2	2	0.57	0	0.57
1035	100	Dairy and Egg	CHEESE,PRC	40.95	351	25.58	26.62	4.71	2.14	0	0.56
1036	100	Dairy and Egg	CHEESE,RICO	71.7	174	11.26	12.98	1.02	3.04	0	0.27
1037	100	Dairy and Egg	CHEESE,RICO	74.41	138	11.39	7.91	1.15	5.14	0	0.31
1038	100	Dairy and Egg	CHEESE,RON	30.91	387	31.8	26.94	6.72	3.63	0	0.73
1040	100	Dairy and Egg	CHEESE,SWI	37.12	380	26.93	27.8	2.77	5.38	0	1.32
1042	100	Dairy and Egg	CHEESE,PAS	39.61	366	18.13	30.71	6.78	4.78	0	2.26
1043	100	Dairy and Egg	CHEESE,PAS	39.08	375	22.13	31.2	5.84	1.73	0.1	0.62
1044	100	Dairy and Egg	CHEESE,PAS	42.31	334	24.73	25.01	5.85	2.1	0	1.23
1046	100	Dairy and Egg	CHEESE FD,FL	44	330	16.86	25.63	4.96	8.56	0	5.59
1048	100	Dairy and Egg	CHEESE SPR	47.65	290	16.41	21.23	5.98	8.73	0	7.32
1049	100	Dairy and Egg	CREAM,FLUI	80.57	130	2.96	11.5	0.67	4.3	0	0.16
1050	100	Dairy and Egg	CREAM,FLUI	73.75	195	2.7	19.31	0.58	3.66	0	0.14
1052	100	Dairy and Egg	CREAM,FLUI	63.5	292	2.17	30.91	0.46	2.96	0	0.11
1053	100	Dairy and Egg	CREAM,FLUI	57.71	345	2.05	37	0.45	2.79	0	0.11

Data transformation: selection

Selection: select from the table the attributes needed for the visualization.

2) Food dataset: visualize how the average amount of calories distributes across food categories.



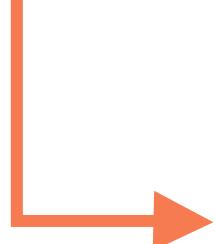
Intermediate aggregation step is necessary!

B	C	D	E	F	G	H	I	J	K	L	M
NDB_No	FdGrp_CD	FdGrp_Desc	Shrt_Desc	Water	Energ_Kcal	Protein	Lipid_Tot	Ash	Carbohydrt	Fiber_TD	Sugar_Tot
1001	100	Dairy and Egg	BUTTER,WIT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1002	100	Dairy and Egg	BUTTER,WH	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1003	100	Dairy and Egg	BUTTER OIL,	0.24	876	0.28	99.48	0	0	0	0
1004	100	Dairy and Egg	CHEESE,BLU	42.41	353	21.4	28.74	5.11	2.34	0	0.5
1005	100	Dairy and Egg	CHEESE,BRIC	41.11	371	23.24	29.68	3.18	2.79	0	0.51
1006	100	Dairy and Egg	CHEESE,BRIG	48.42	334	20.75	27.68	2.7	0.45	0	0.45
1007	100	Dairy and Egg	CHEESE,CAN	51.8	300	19.8	24.26	3.68	0.46	0	0.46
1009	100	Dairy and Egg	CHEESE,CHE	36.75	403	24.9	33.14	3.93	1.28	0	0.52
1011	100	Dairy and Egg	CHEESE,COL	38.2	394	23.76	32.11	3.36	2.57	0	0.52
1012	100	Dairy and Egg	CHEESE,COT	79.79	98	11.12	4.3	1.41	3.38	0	2.67
1013	100	Dairy and Egg	CHEESE,COT	79.64	97	10.69	3.85	1.2	4.61	0.2	2.38
1014	100	Dairy and Egg	CHEESE,COT	81.01	72	10.34	0.29	1.71	6.66	0	1.85
1015	100	Dairy and Egg	CHEESE,COT	80.69	86	11.83	2.45	1.36	3.66	0	3.67
1016	100	Dairy and Egg	CHEESE,COT	82.48	72	12.39	1.02	1.39	2.72	0	2.72
1017	100	Dairy and Egg	CHEESE,CRE	54.44	342	5.93	34.24	1.32	4.07	0	3.21
1018	100	Dairy and Egg	CHEESE,EDA	41.56	357	24.99	27.8	4.22	1.43	0	1.43
1019	100	Dairy and Egg	CHEESE,FET,	55.22	264	14.21	21.28	5.2	4.09	0	4.09
1020	100	Dairy and Egg	CHEESE,FON	37.92	389	25.6	31.14	3.79	1.55	0	1.55
1022	100	Dairy and Egg	CHEESE,GOU	41.46	356	24.94	27.44	3.94	2.22	0	2.22
1023	100	Dairy and Egg	CHEESE,GRL	33.19	413	29.81	32.34	4.3	0.36	0	0.36
1024	100	Dairy and Egg	CHEESE,LIM	48.42	327	20.05	27.25	3.79	0.49	0	0.49
1025	100	Dairy and Egg	CHEESE,MO	41.01	373	24.48	30.28	3.55	0.68	0	0.5
1026	100	Dairy and Egg	CHEESE,MO	50.01	300	22.17	22.35	3.28	2.19	0	1.03
1027	100	Dairy and Egg	CHEESE,MO	48.38	318	21.6	24.64	2.91	2.47	0	1.01
1029	100	Dairy and Egg	CHEESE,MO	46.46	302	25.96	20.03	3.72	3.83	0	0.6
1030	100	Dairy and Egg	CHEESE,MU	41.77	368	23.41	30.04	3.66	1.12	0	1.12
1032	100	Dairy and Egg	CHEESE,PAR	20.84	431	38.46	28.61	8.03	4.06	0	0.9
1033	100	Dairy and Egg	CHEESE,PAR	29.16	392	35.75	25.83	6.04	3.22	0	0.8
1034	100	Dairy and Egg	CHEESE,POR	45.45	352	23.78	28.2	2	0.57	0	0.57
1035	100	Dairy and Egg	CHEESE,PRC	40.95	351	25.58	26.62	4.71	2.14	0	0.56
1036	100	Dairy and Egg	CHEESE,RICO	71.7	174	11.26	12.98	1.02	3.04	0	0.27
1037	100	Dairy and Egg	CHEESE,RICO	74.41	138	11.39	7.91	1.15	5.14	0	0.31
1038	100	Dairy and Egg	CHEESE,RON	30.91	387	31.8	26.94	6.72	3.63	0	0.73
1040	100	Dairy and Egg	CHEESE,SWI	37.12	380	26.93	27.8	2.77	5.38	0	1.32
1042	100	Dairy and Egg	CHEESE,PAS	39.61	366	18.13	30.71	6.78	4.78	0	2.26
1043	100	Dairy and Egg	CHEESE,PAS	39.08	375	22.13	31.2	5.84	1.73	0.1	0.62
1044	100	Dairy and Egg	CHEESE,PAS	42.31	334	24.73	25.01	5.85	2.1	0	1.23
1046	100	Dairy and Egg	CHEESE FD,FL	44	330	16.86	25.63	4.96	8.56	0	5.59
1048	100	Dairy and Egg	CHEESE SPR	47.65	290	16.41	21.23	5.98	8.73	0	7.32
1049	100	Dairy and Egg	CREAM,FLUI	80.57	130	2.96	11.5	0.67	4.3	0	0.16
1050	100	Dairy and Egg	CREAM,FLUI	73.75	195	2.7	19.31	0.58	3.66	0	0.14
1052	100	Dairy and Egg	CREAM,FLUI	63.5	292	2.17	30.91	0.46	2.96	0	0.11
1053	100	Dairy and Egg	CREAM,FLUI	57.71	345	2.05	37	0.45	2.79	0	0.11

Data transformation: aggregation

Aggregation: grouping multiple rows to form a summary value.
Avg, sum, min, max, std. dev. ...

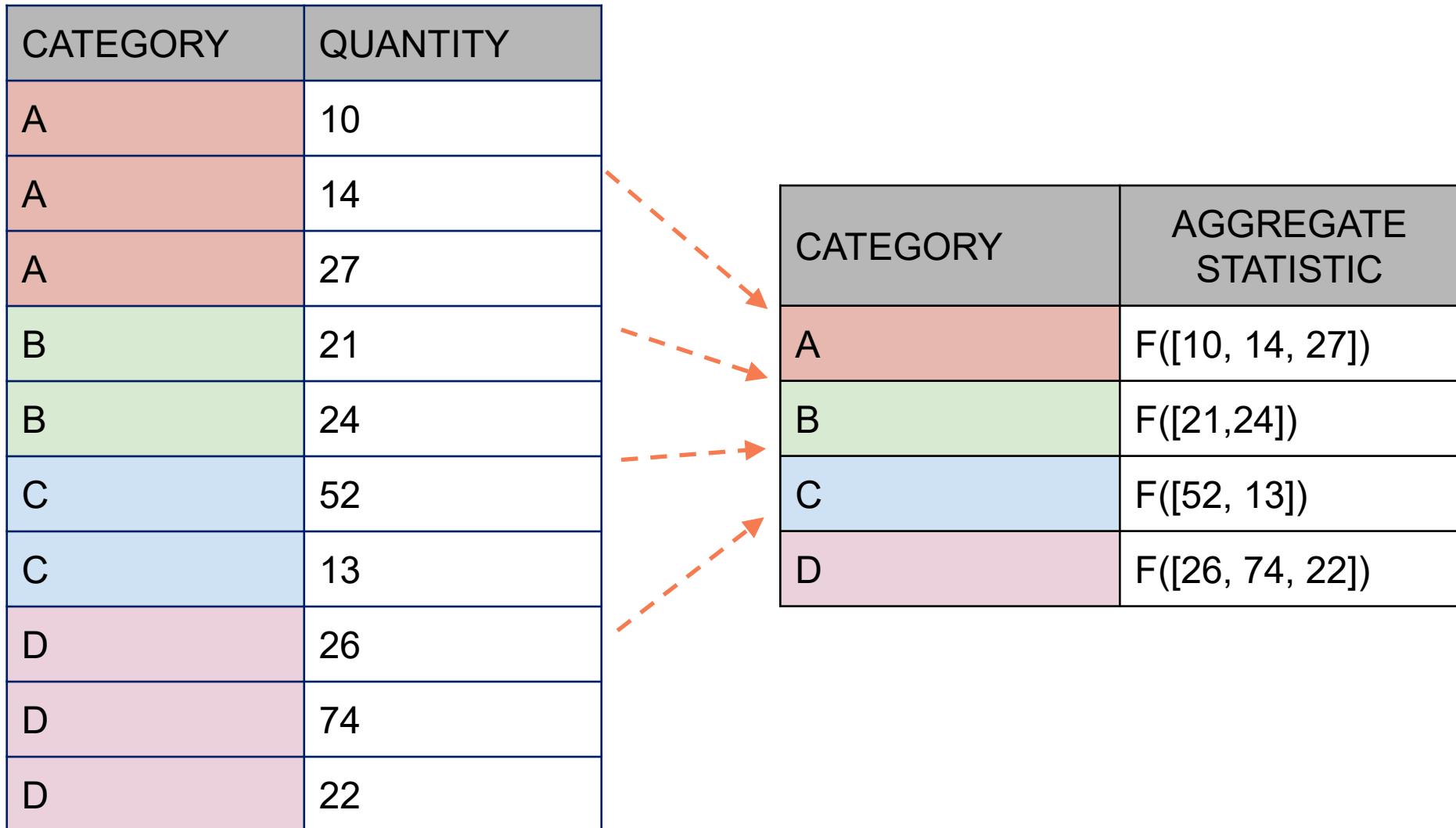
2) Food dataset: visualize how the average amount of calories distributes across food categories.



Intermediate
aggregation step
is necessary!

B	C	D	E	F	G	H	I	J	K	L	M
NDB_No	FdGrp_CD	FdGrp_Desc	Shrt_Desc	Water	Energ_Kcal	Protein	Lipid_Tot	Ash	Carbohydrt	Fiber_TD	Sugar_Tot
1001	100	Dairy and Egg	BUTTER,WIT	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1002	100	Dairy and Egg	BUTTER,WH	15.87	717	0.85	81.11	2.11	0.06	0	0.06
1003	100	Dairy and Egg	BUTTER OIL,	0.24	876	0.28	99.48	0	0	0	0
1004	100	Dairy and Egg	CHEESE,BLU	42.41	353	21.4	28.74	5.11	2.34	0	0.5
1005	100	Dairy and Egg	CHEESE,BRIC	41.11	371	23.24	29.68	3.18	2.79	0	0.51
1006	100	Dairy and Egg	CHEESE,BRIG	48.42	334	20.75	27.68	2.7	0.45	0	0.45
1007	100	Dairy and Egg	CHEESE,CAN	51.8	300	19.8	24.26	3.68	0.46	0	0.46
1009	100	Dairy and Egg	CHEESE,CHE	36.75	403	24.9	33.14	3.93	1.28	0	0.52
1011	100	Dairy and Egg	CHEESE,COL	38.2	394	23.76	32.11	3.36	2.57	0	0.52
1012	100	Dairy and Egg	CHEESE,COT	79.79	98	11.12	4.3	1.41	3.38	0	2.67
1013	100	Dairy and Egg	CHEESE,COT	79.64	97	10.69	3.85	1.2	4.61	0.2	2.38
1014	100	Dairy and Egg	CHEESE,COT	81.01	72	10.34	0.29	1.71	6.66	0	1.85
1015	100	Dairy and Egg	CHEESE,COT	80.69	86	11.83	2.45	1.36	3.66	0	3.67
1016	100	Dairy and Egg	CHEESE,COT	82.48	72	12.39	1.02	1.39	2.72	0	2.72
1017	100	Dairy and Egg	CHEESE,CRE	54.44	342	5.93	34.24	1.32	4.07	0	3.21
1018	100	Dairy and Egg	CHEESE,EDA	41.56	357	24.99	27.8	4.22	1.43	0	1.43
1019	100	Dairy and Egg	CHEESE,FET,	55.22	264	14.21	21.28	5.2	4.09	0	4.09
1020	100	Dairy and Egg	CHEESE,FON	37.92	389	25.6	31.14	3.79	1.55	0	1.55
1022	100	Dairy and Egg	CHEESE,GOU	41.46	356	24.94	27.44	3.94	2.22	0	2.22
1023	100	Dairy and Egg	CHEESE,GRL	33.19	413	29.81	32.34	4.3	0.36	0	0.36
1024	100	Dairy and Egg	CHEESE,LIM	48.42	327	20.05	27.25	3.79	0.49	0	0.49
1025	100	Dairy and Egg	CHEESE,MO	41.01	373	24.48	30.28	3.55	0.68	0	0.5
1026	100	Dairy and Egg	CHEESE,MO	50.01	300	22.17	22.35	3.28	2.19	0	1.03
1027	100	Dairy and Egg	CHEESE,MO	48.38	318	21.6	24.64	2.91	2.47	0	1.01
1029	100	Dairy and Egg	CHEESE,MO	46.46	302	25.96	20.03	3.72	3.83	0	0.6
1030	100	Dairy and Egg	CHEESE,MU	41.77	368	23.41	30.04	3.66	1.12	0	1.12
1032	100	Dairy and Egg	CHEESE,PAR	20.84	431	38.46	28.61	8.03	4.06	0	0.9
1033	100	Dairy and Egg	CHEESE,PAR	29.16	392	35.75	25.83	6.04	3.22	0	0.8
1034	100	Dairy and Egg	CHEESE,POR	45.45	352	23.78	28.2	2	0.57	0	0.57
1035	100	Dairy and Egg	CHEESE,PRC	40.95	351	25.58	26.62	4.71	2.14	0	0.56
1036	100	Dairy and Egg	CHEESE,RICO	71.7	174	11.26	12.98	1.02	3.04	0	0.27
1037	100	Dairy and Egg	CHEESE,RICO	74.41	138	11.39	7.91	1.15	5.14	0	0.31
1038	100	Dairy and Egg	CHEESE,RON	30.91	387	31.8	26.94	6.72	3.63	0	0.73
1040	100	Dairy and Egg	CHEESE,SWI	37.12	380	26.93	27.8	2.77	5.38	0	1.32
1042	100	Dairy and Egg	CHEESE,PAS	39.61	366	18.13	30.71	6.78	4.78	0	2.26
1043	100	Dairy and Egg	CHEESE,PAS	39.08	375	22.13	31.2	5.84	1.73	0.1	0.62
1044	100	Dairy and Egg	CHEESE,PAS	42.31	334	24.73	25.01	5.85	2.1	0	1.23
1046	100	Dairy and Egg	CHEESE FD,FL	44	330	16.86	25.63	4.96	8.56	0	5.59
1048	100	Dairy and Egg	CHEESE SPR	47.65	290	16.41	21.23	5.98	8.73	0	7.32
1049	100	Dairy and Egg	CREAM,FLUI	80.57	130	2.96	11.5	0.67	4.3	0	0.16
1050	100	Dairy and Egg	CREAM,FLUI	73.75	195	2.7	19.31	0.58	3.66	0	0.14
1052	100	Dairy and Egg	CREAM,FLUI	63.5	292	2.17	30.91	0.46	2.96	0	0.11
1053	100	Dairy and Egg	CREAM,FLUI	57.71	345	2.05	37	0.45	2.79	0	0.11

Data transformation: aggregation



Data transformation: aggregation

Aggregation: grouping multiple rows to form a summary value.
Avg, sum, min, max, std. dev. ...

2) *Food dataset: visualize how the average amount of calories distributes across food categories.*

C+Q

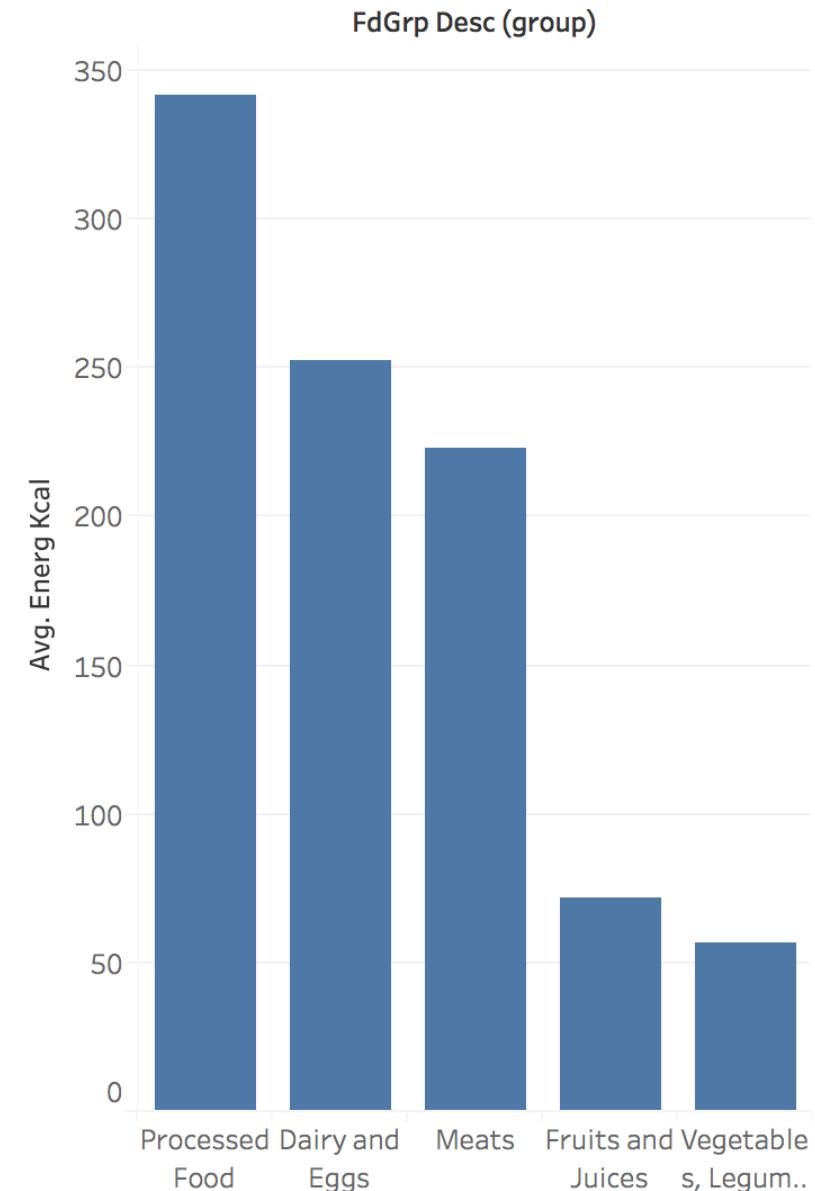
CATEGORY	AGGREGATE STATISTIC
A	F([10, 14, 27])
B	F([21,24])
C	F([52, 13])
D	F([26, 74, 22])

Data transformation: aggregation

Aggregation: grouping multiple rows to form a summary value.
Avg, sum, min, max, std. dev. ...

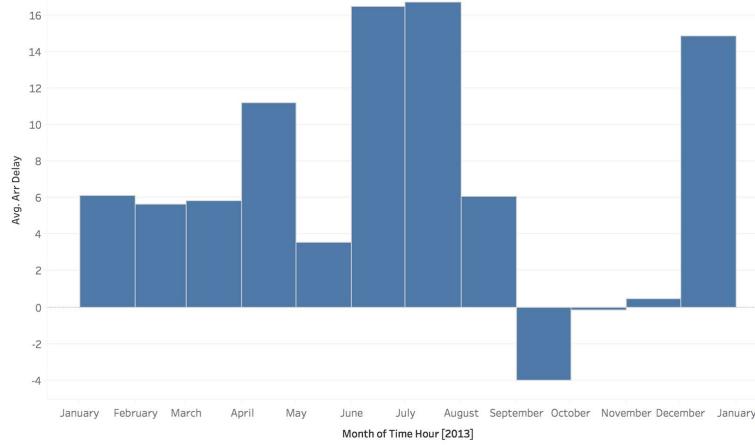
2) Food dataset: visualize how the average amount of calories distributes across food categories.

C+Q

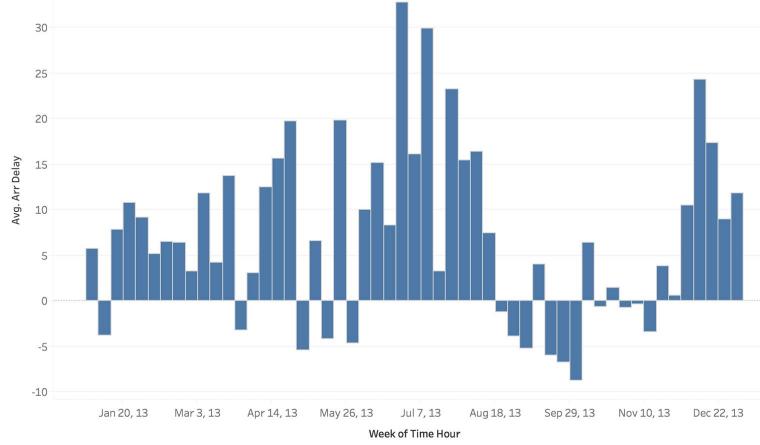


Data transformation: temporal aggregation

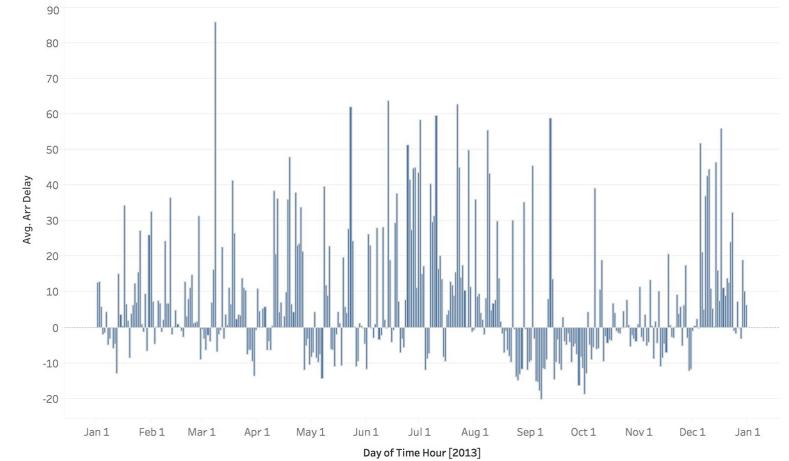
- Time and date:
 - Aggregation by: seconds, minutes, hours, day, week, ...



Months



Weeks

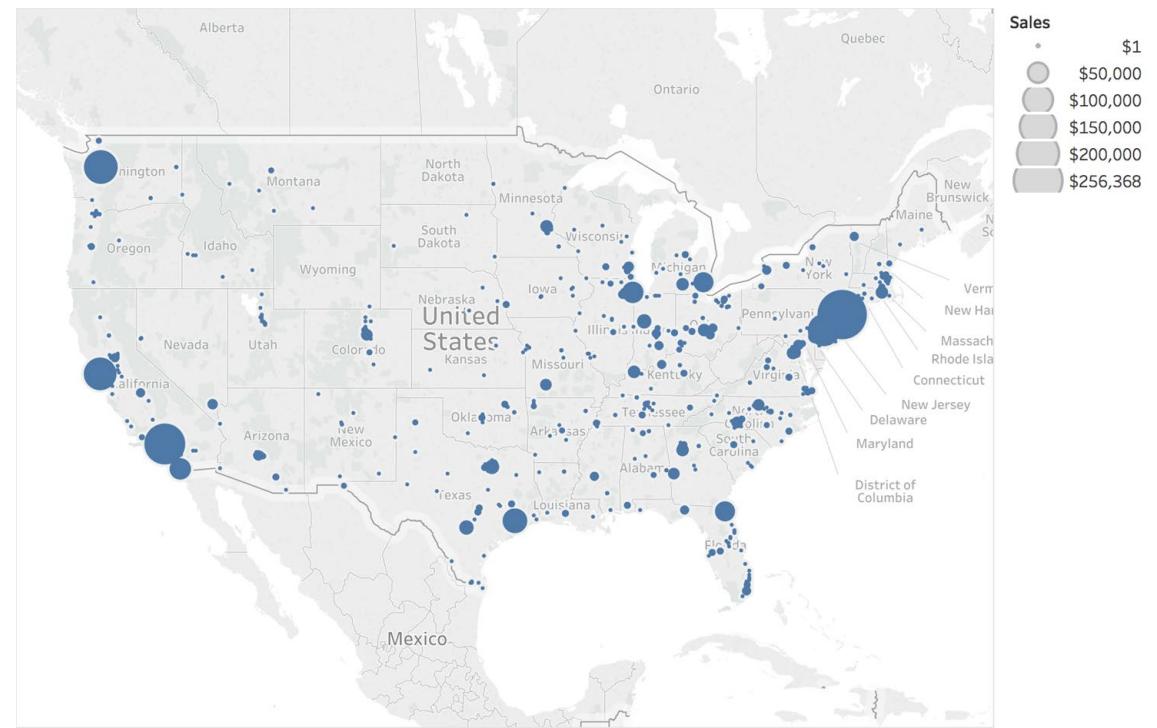
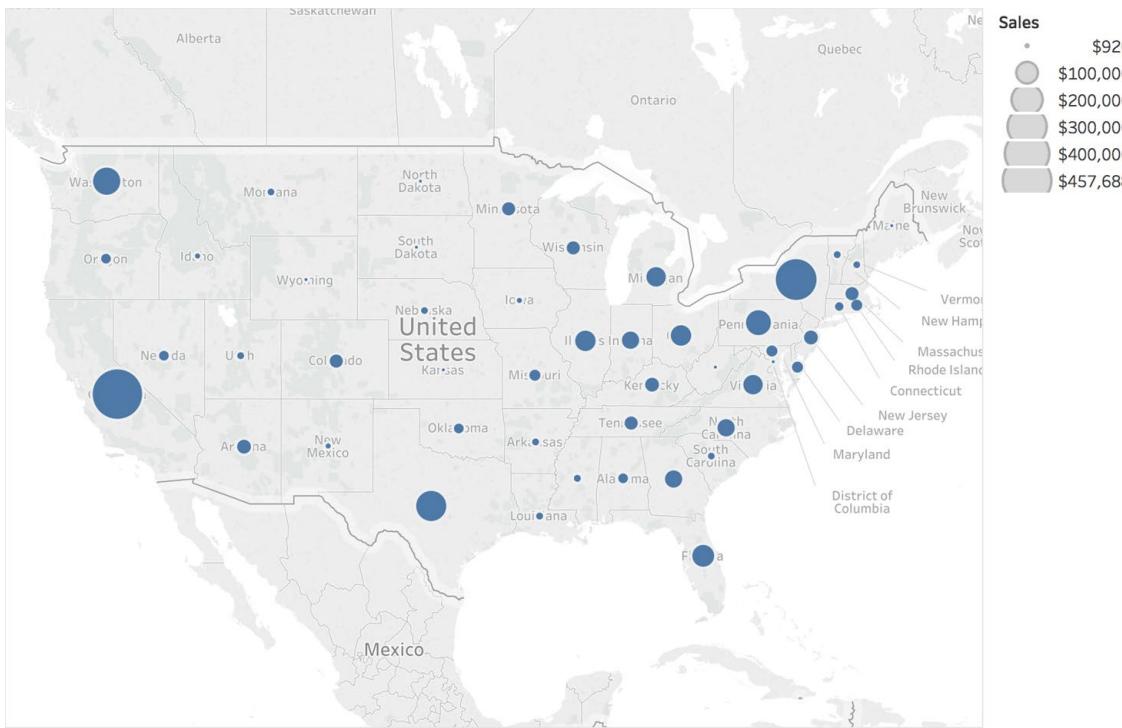


Days

[Enrico Bertini]

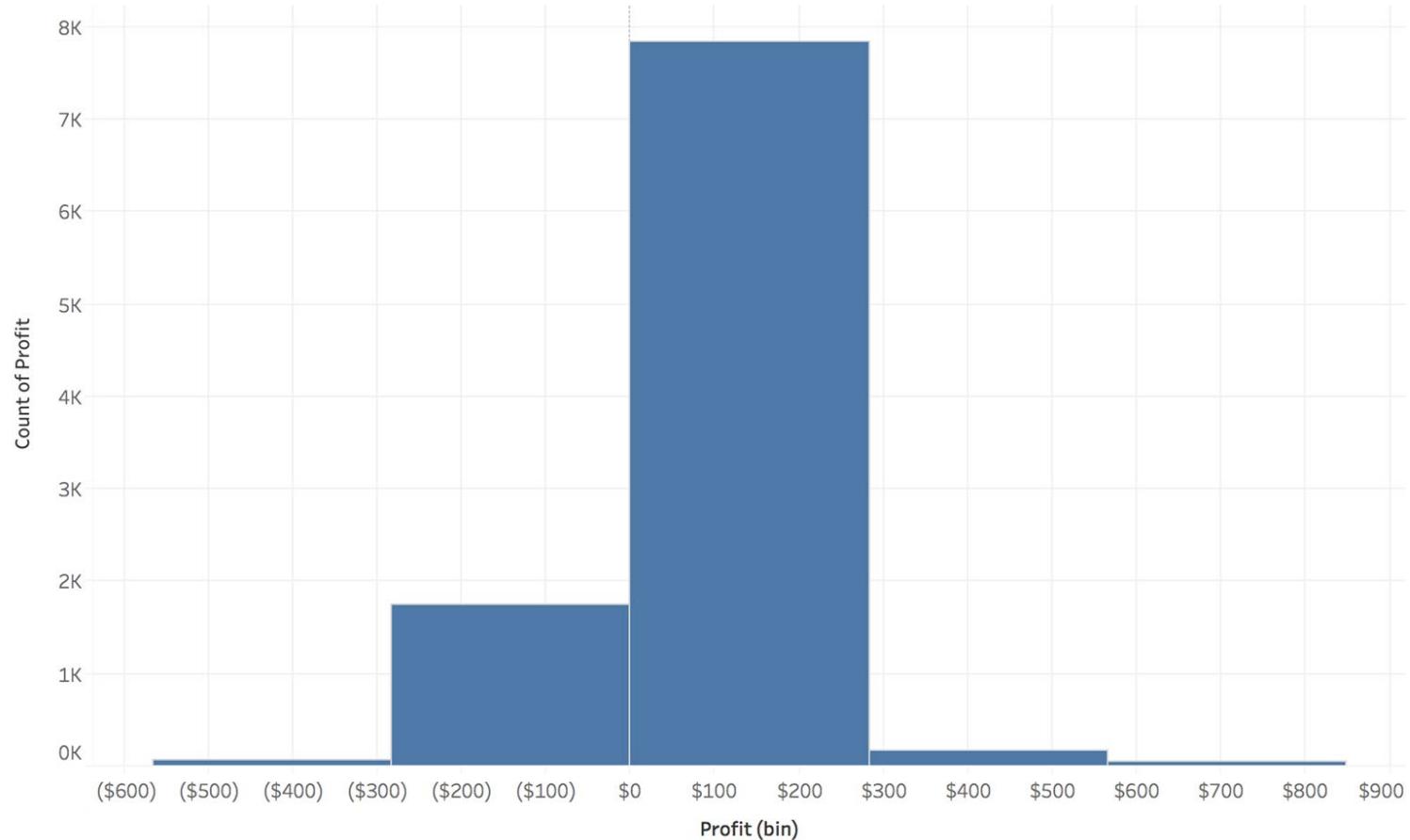
Data transformation: spatial aggregation

- Location and region:
 - Aggregation by: zip code, county, state, ...



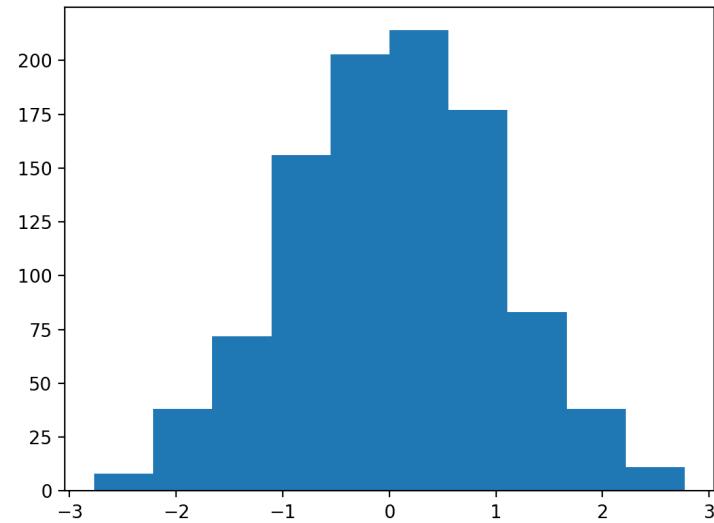
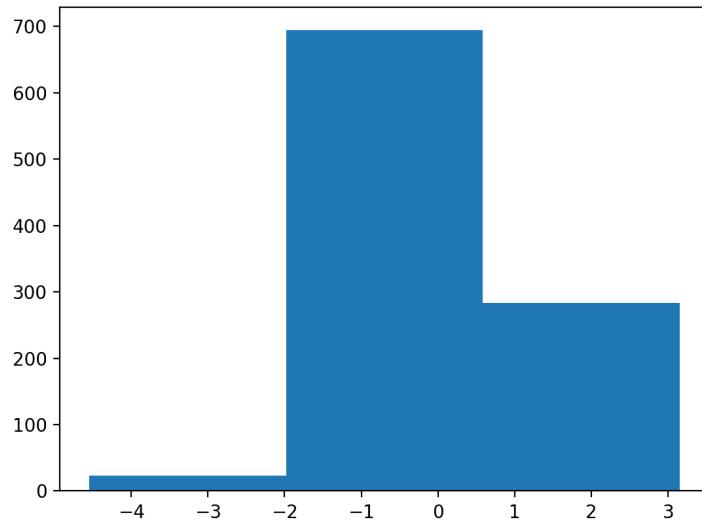
Data transformation: binning

- Transforms quantitative measures to (categorical) ordinal through binning.



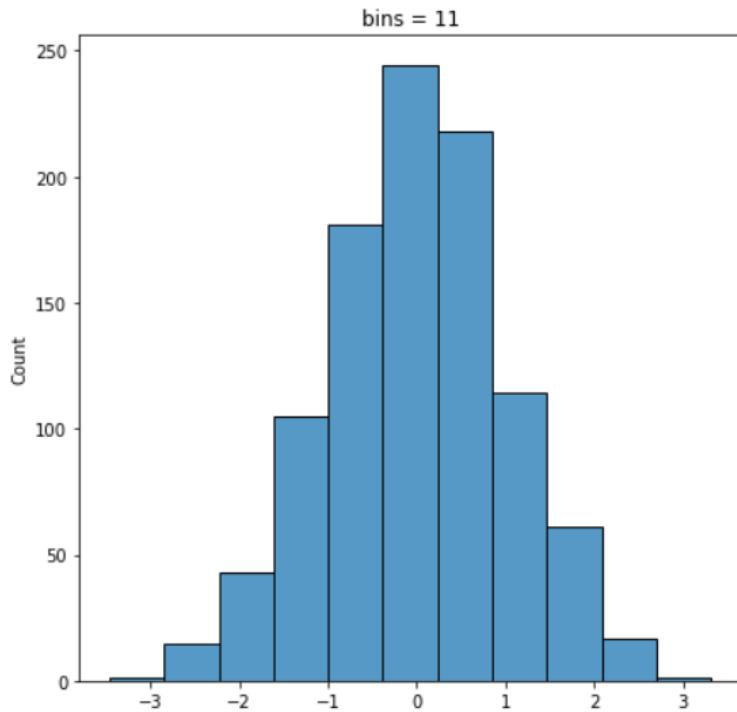
Data transformation: binning

- Fast and reliable way to visualize probability density.
- Impacted by the number of bins, i.e., depends on the width of bins.

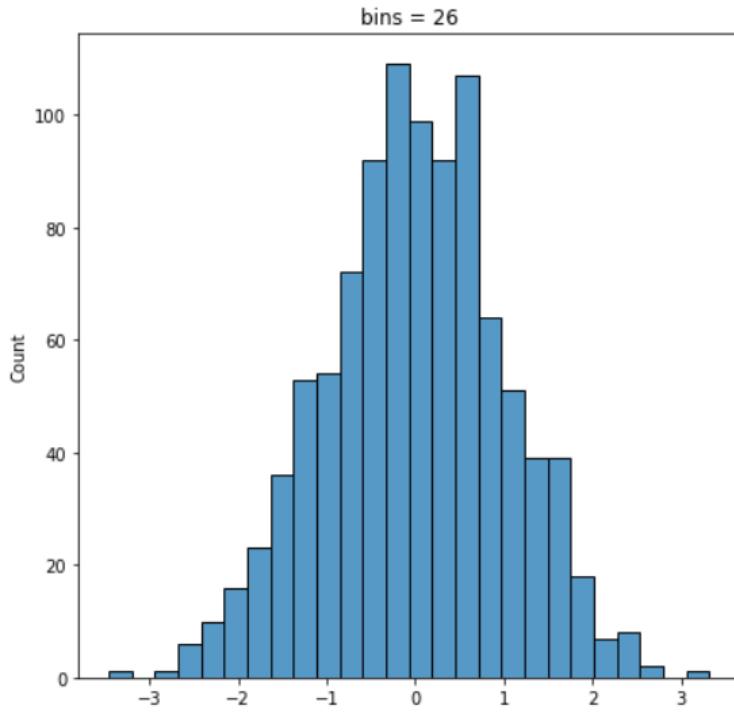


Data transformation: binning

- Deciding the number of bins:



Sturge's rule:
 $bins = 1 + [\log_2 n]$



Freedman-Diaconis rule:
$$h = \frac{2IQR(x)}{\sqrt[3]{n}}$$

$$bins = \left\lceil \frac{\max(x) - \min(x)}{h} \right\rceil$$

Data aggregation generates uncertainty

- Same stats, different graphs [Matejka and Fitzmaurice, 2017]

