Fundamental Graphs and Data Transformation

Thursday, October 11, 2018 17:53

1 Data abstraction -> Graph design
2 Fundamental graphs and variations
3 Data transformation

How do I visualize this?

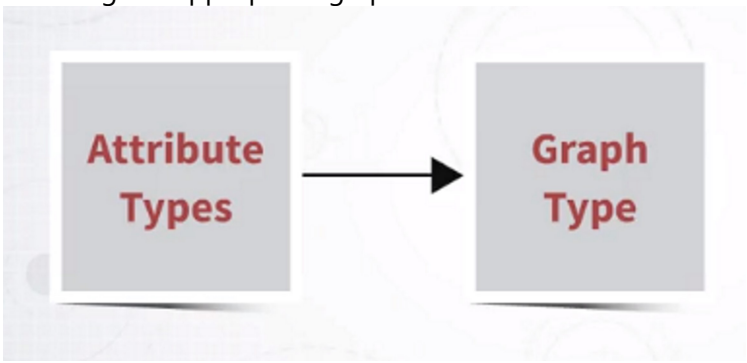
We can break the problem:

1. Decide WHAT to visualize
    a. Select: visualizations usually use a small subset of the data
    b. Transform: the data must be in an specific form

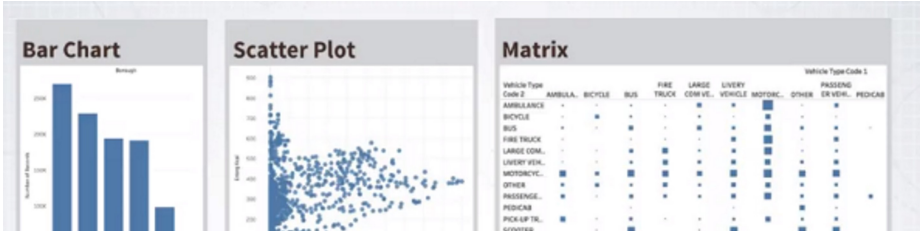
2. Decide HOW to visualize
    a. Choose: selecting an existing format that is appropriate for a goal
    b. Design: trying to create a novel representation for a specific problem

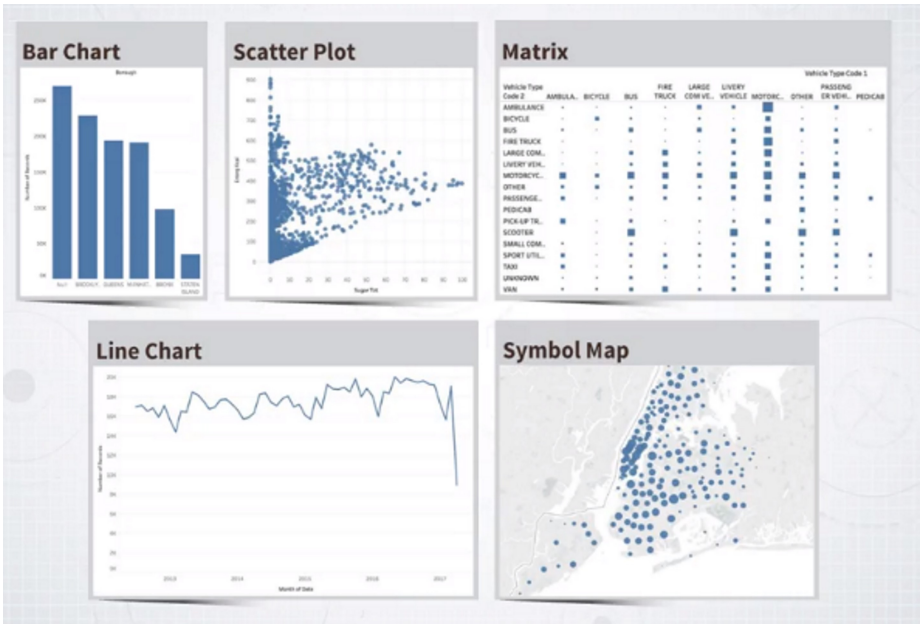
Fundamental graphs

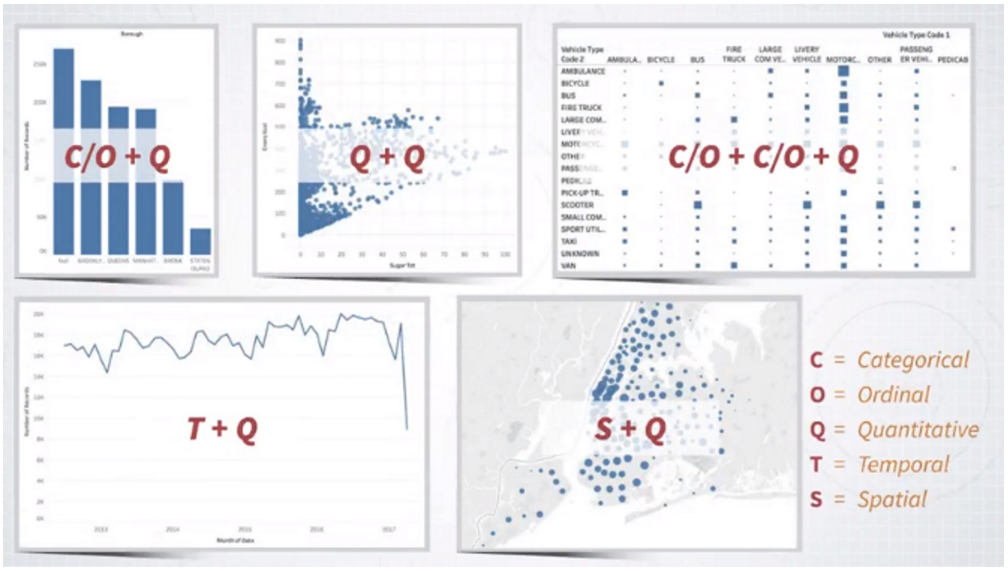
Selecting the appropriate graphic:



- Widely adopted, effective and useful.
- Solve very large percentage of visualization problems.
- Training ground for more sophisticated graphs.







Graphs can be described as combinations of two or more attributes.

Bar chart

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

Bars -> categories
Bar length -> quantity

Line chart

Visualize how a **quantity** changes in relation to another **quantity** (usually time).

Scatter plot

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

Matrix

Visualize how a **quantity** distribute (relates) across two **categories**.

Symbol Map

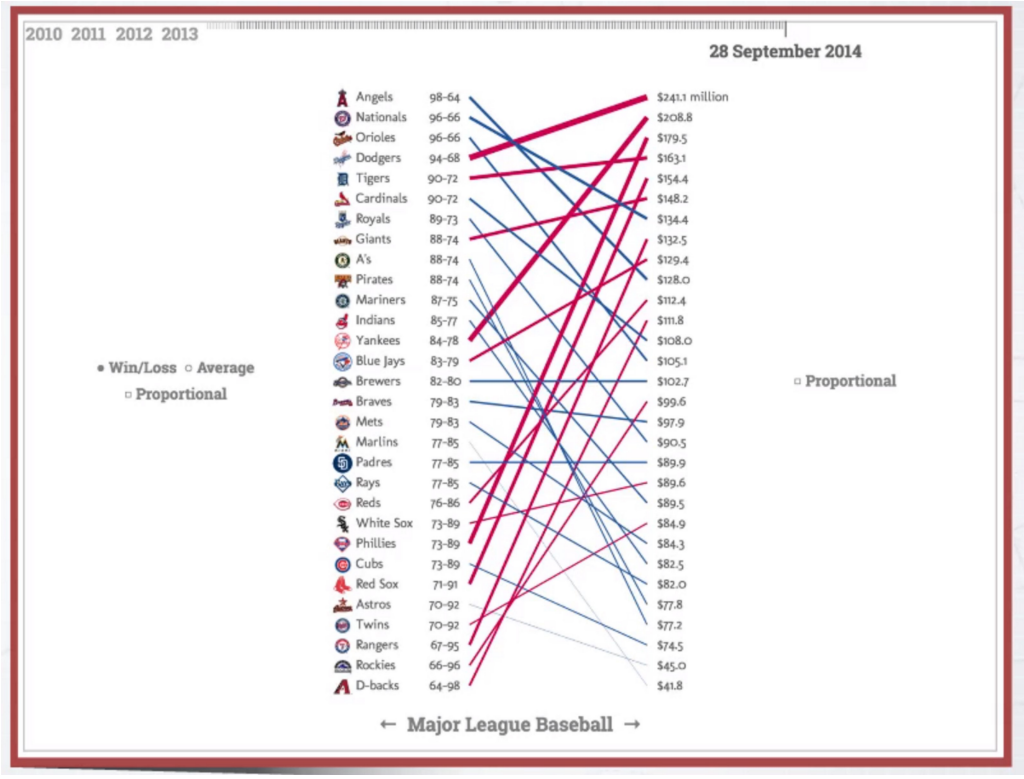
Visualize how a **quantity** distributes across two **spatial coordinates**.

Alternate representations

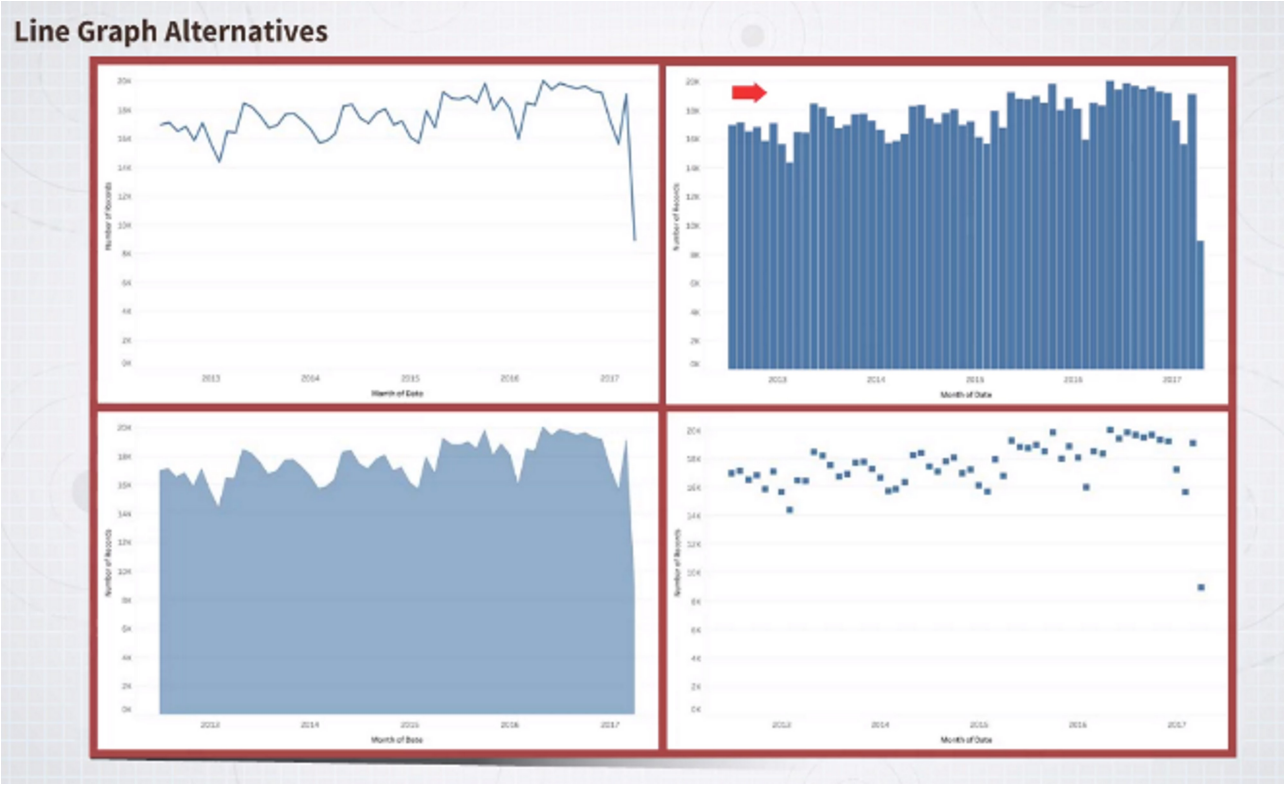
Is it possible to create **different representations** of the **same data**?

How to reason about the quality and adv/disadvantages of alternate visual encodings.

Bar chart ---> Slope chart



Line charts --> bar chart / area chart / dots



Matrix alternatives ---> / Nested categories

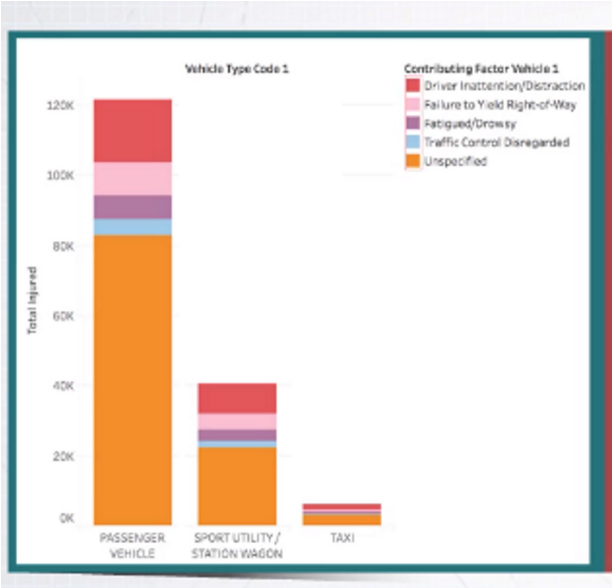
Symbol map ---> Get rid of spatial metaphor and have a bar chart

Going beyond two attributes

Stacked bar chart

- As many bars as many categories in cat 1.
- AS many segments (colors) as many categories in cat 2.

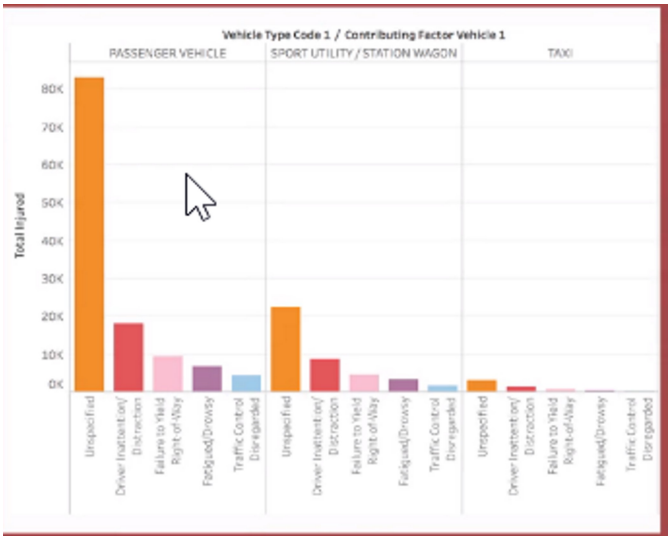
Very good when the question is regarding to the proportion -> *communicate proportions. (part-to-whole information)*



Grouped bar chart

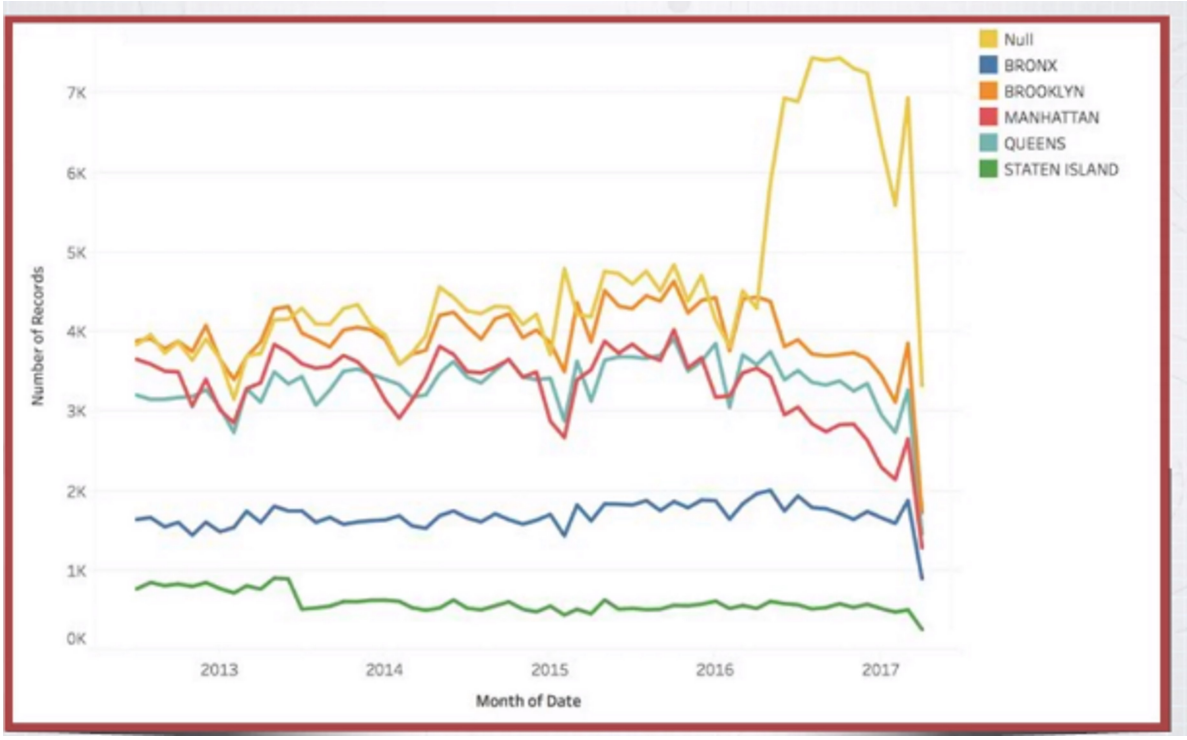
Same bar chart is repeated multiple times (as many times as the number of categories in cat 2)

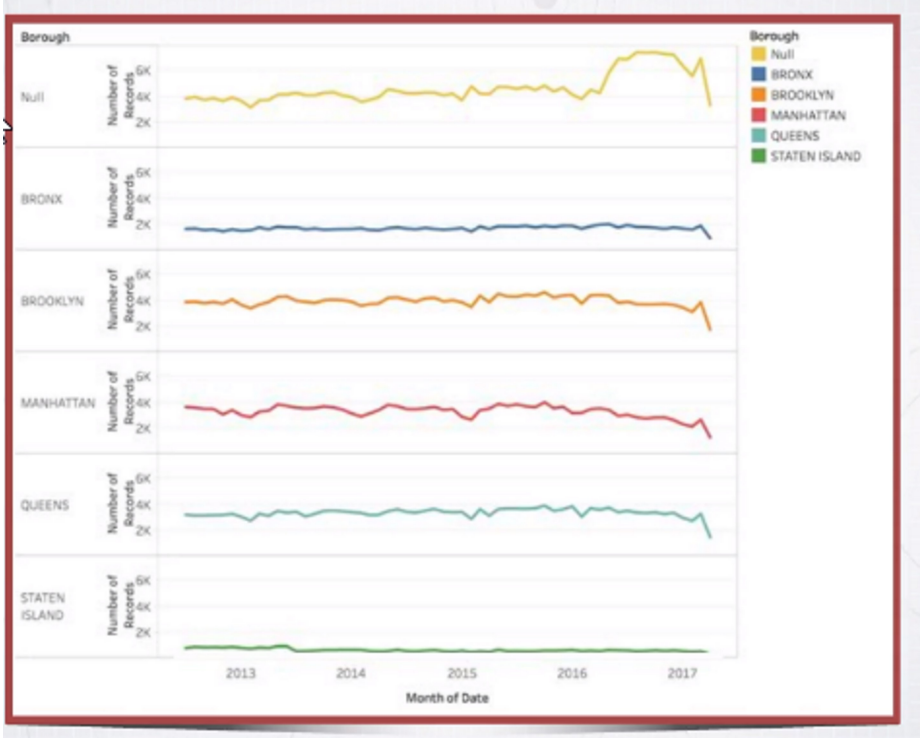
Better when the goal is to compare every single value one to another.



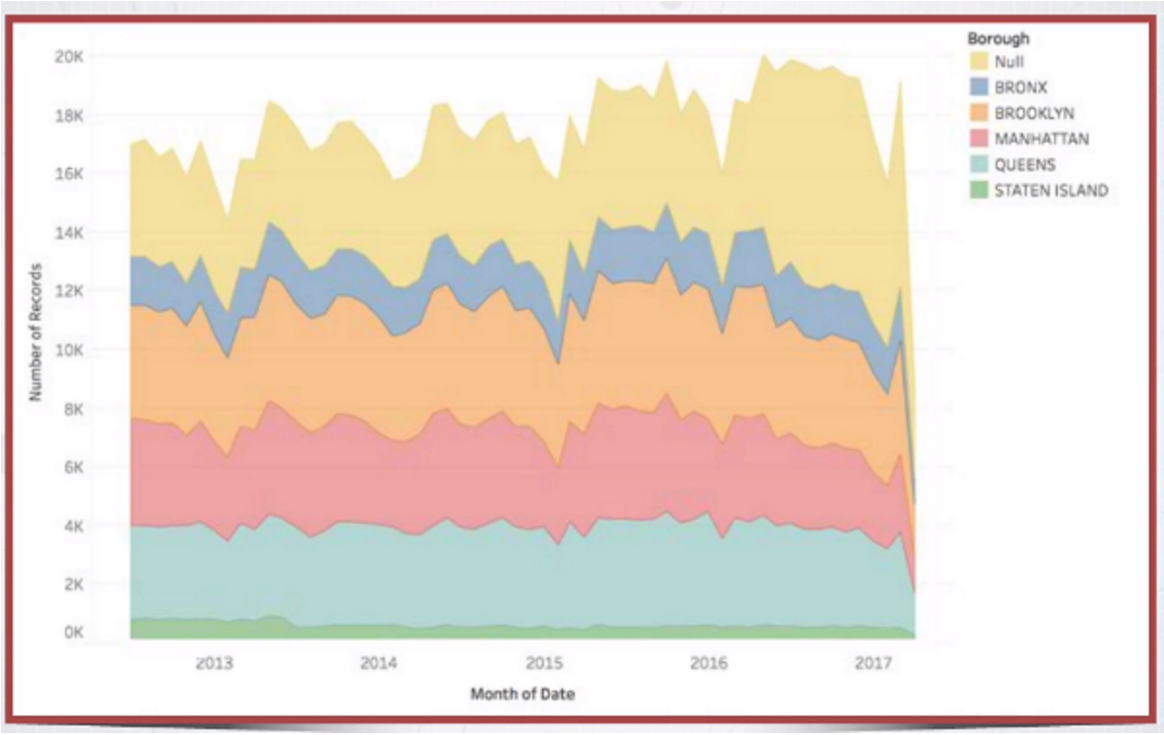
Line charts

One possibility: esasy to compare





Stacked area chart: good for reading the individual values over time. *Patterns are affected by the shape of the line below the one being observed.*



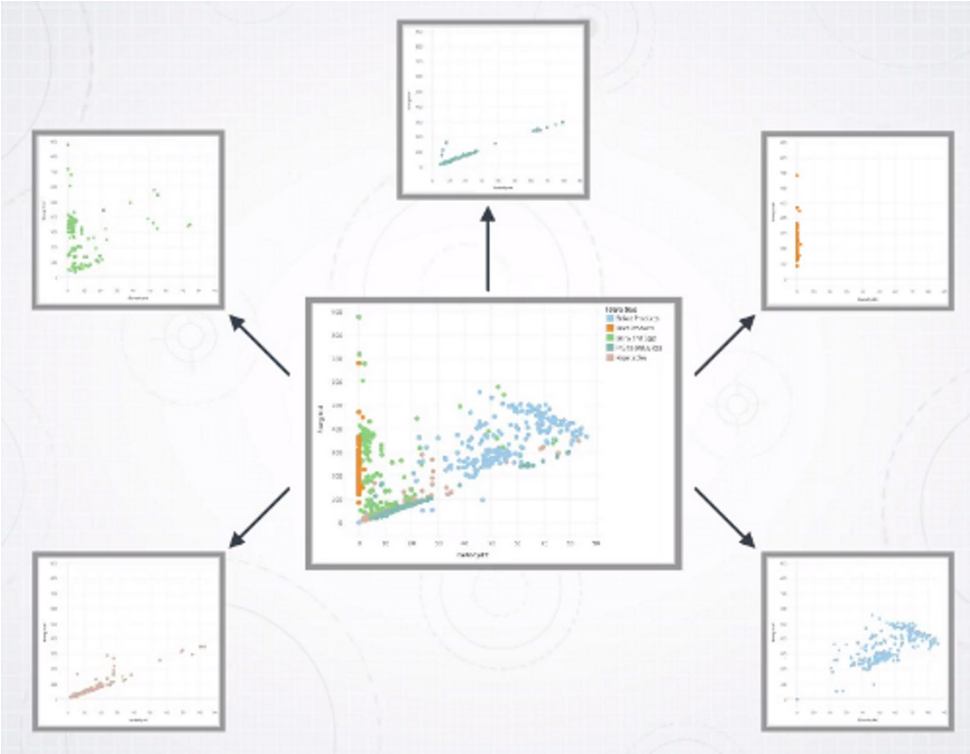
Scatter plots + Faceting

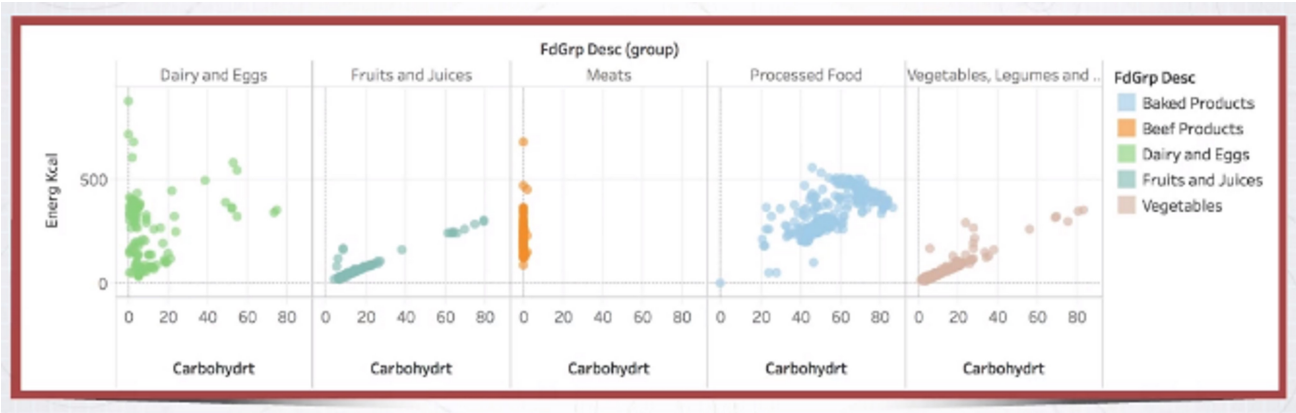
Color: encode information about an extra attribute (categorical)

- Scatter plot can encode up to 4 attributes.
- The more channels or attributes you add to the chart, the harder it is to decode the information.

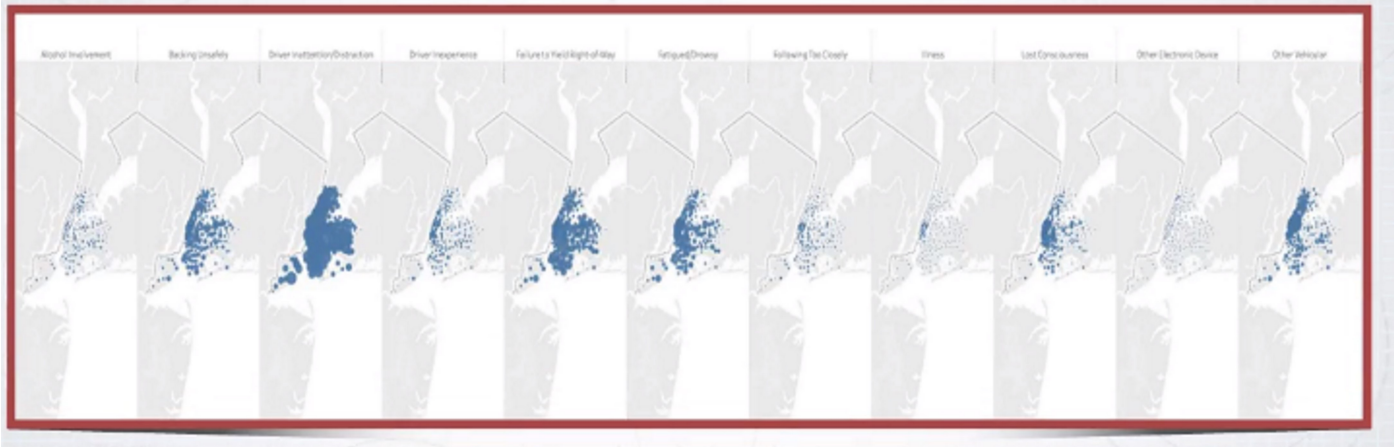
Faceting (small multiples)
Can be applied to any graph.

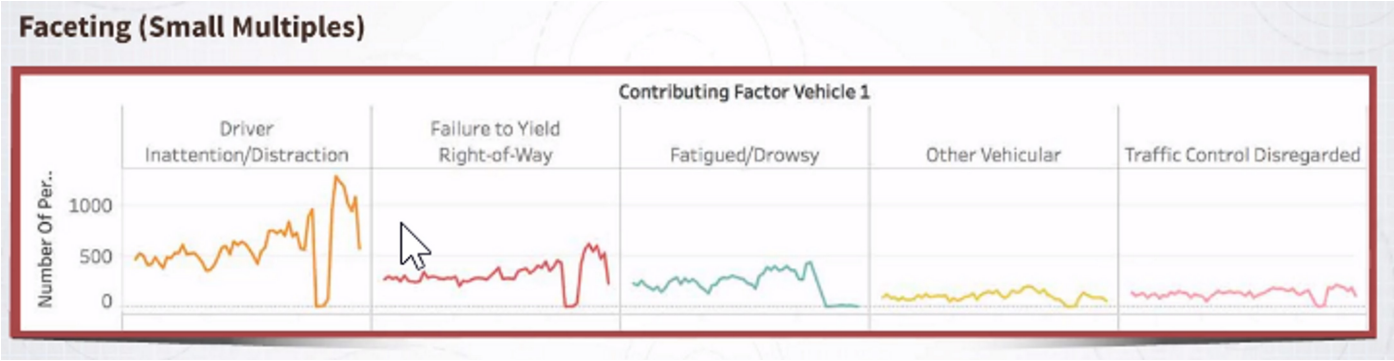
- Select one categorical/ordinal attribute
- Create as many sets as the number of values
- Create one plot for each of these values





Faceting with Maps



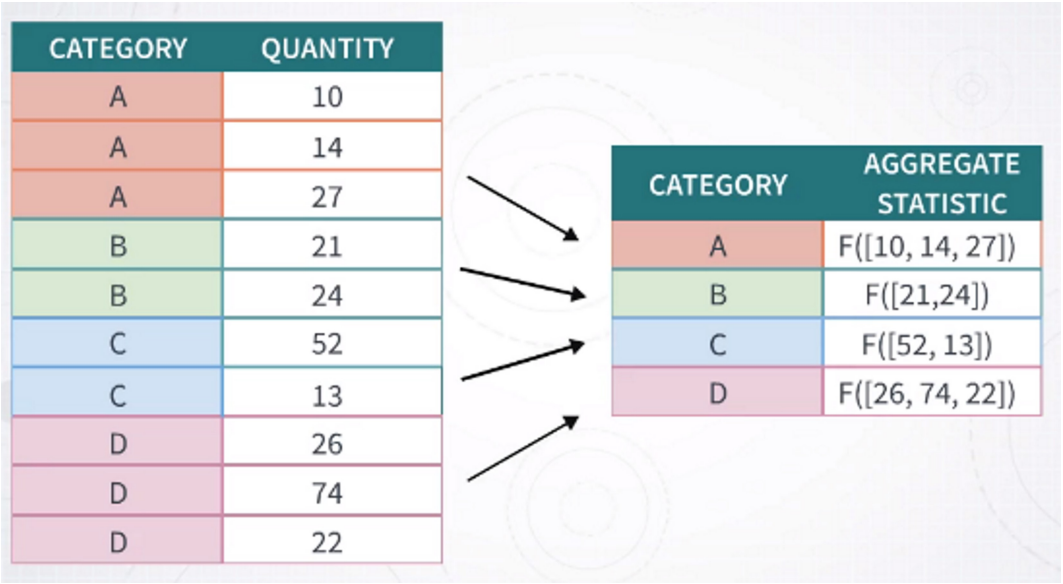


Data transformation

1- Select / transform
2- Choose / design

Selection
Selecting the attributes you need from the table you have.

Virtually all graphs require **selection**
Many require an intermediate **aggregation step**, necessary to answer your question
-> Common aggregation functions:
Sum / Max / Min / Average / Median / Stddev



Common/useful data transformations

- Data transformation is part of the design process.

- Creating the right, effective visual representation for a given problem is not only about finding the right graphical format, but also finding the right information.

- It's almost never the case that you can take the original data and represent it as it is.
- Part of the process is to figure out what is the best transformation to achieve and create an effective visual representation -> think creatively how transformation may lead to better communication and understanding.
- Visualizing data is NOT only about how to visualize data but also WHAT information to visualize -> It's up to you to produce the right information for your problem.

Common data transformations:

Time and date:

- Aggregation by Time resolution (hierarchical structure)
  - Seconds / minutes / hours
  - Day / week / month / year
  - According the time resolution you can stop different trends

Geographical data

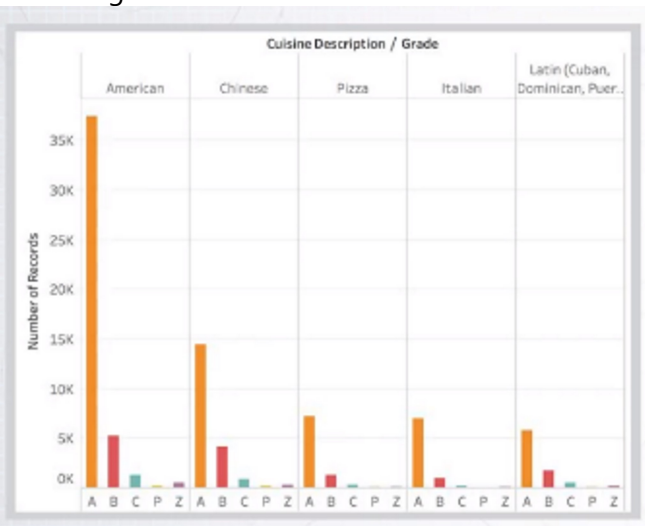
- Hierarchical aggregation
  - Zip code, county, state, country…
- Geocoding and decoding
  - Names to coordinates
  - Coordinates to street names

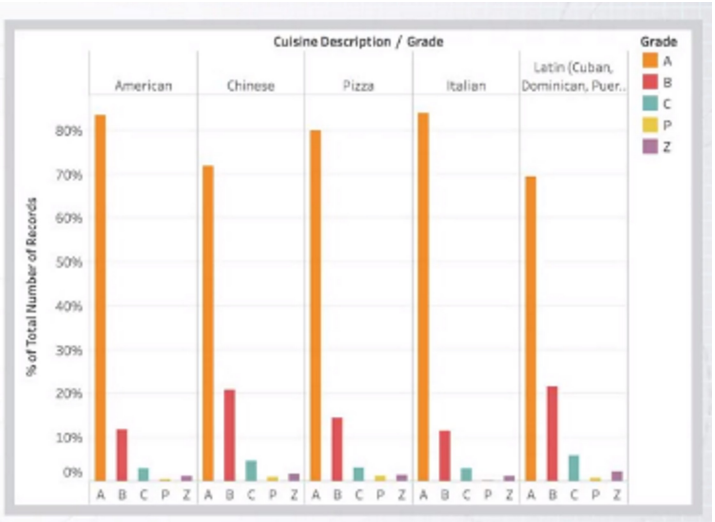
Binning

Quantitative attribute -> ordinal (discrete)

Rescaling / re-expression

- Normalization: rescaling, for instance from [min, max] to [0, 1]
- Percentages





- Distance from reference (e.g. average)