Introduction to visualization: marks and channels

CS424: Visualization & Visual Analytics

Fabio Miranda

https://fmiranda.me

UIC COMPUTER SCIENCE

Data to knowledge

data → knowledge

Data to **knowledge**

statistics

data management

machine learning

visualization

data integration

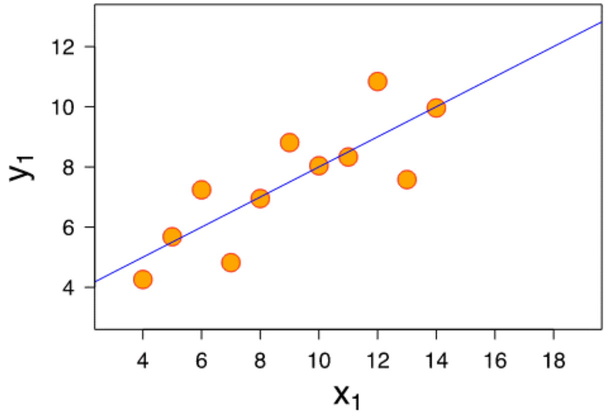
data curation

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

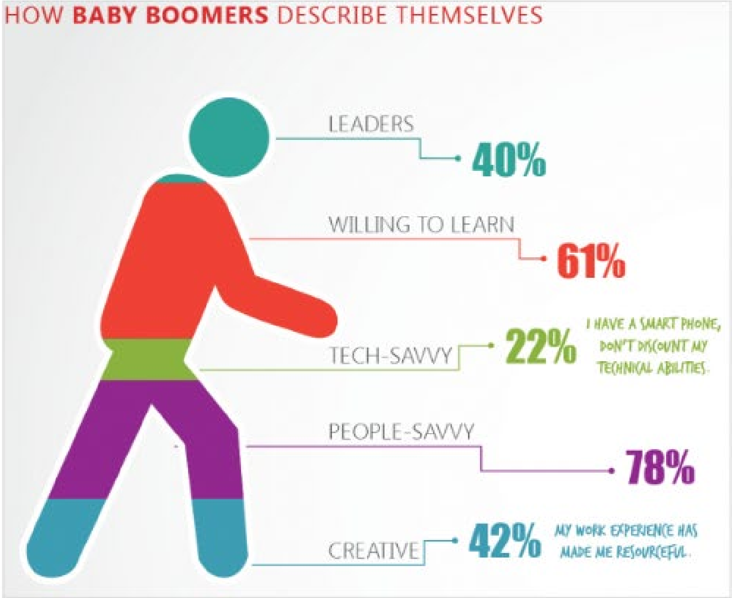
Data visualization

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

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

Data visualization

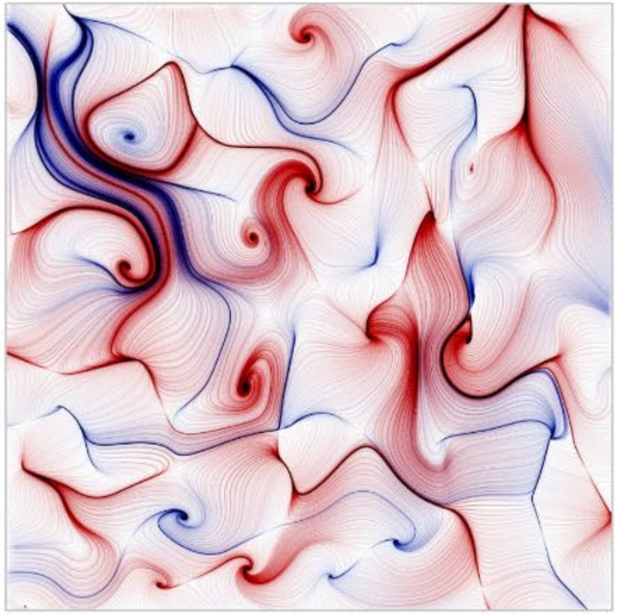
insight —Communication→ insight

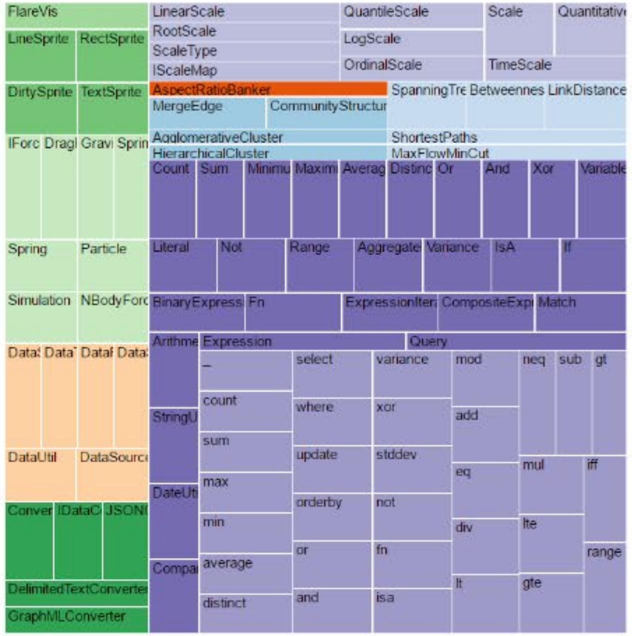


Data visualization

data → insight

Exploration / Analysis



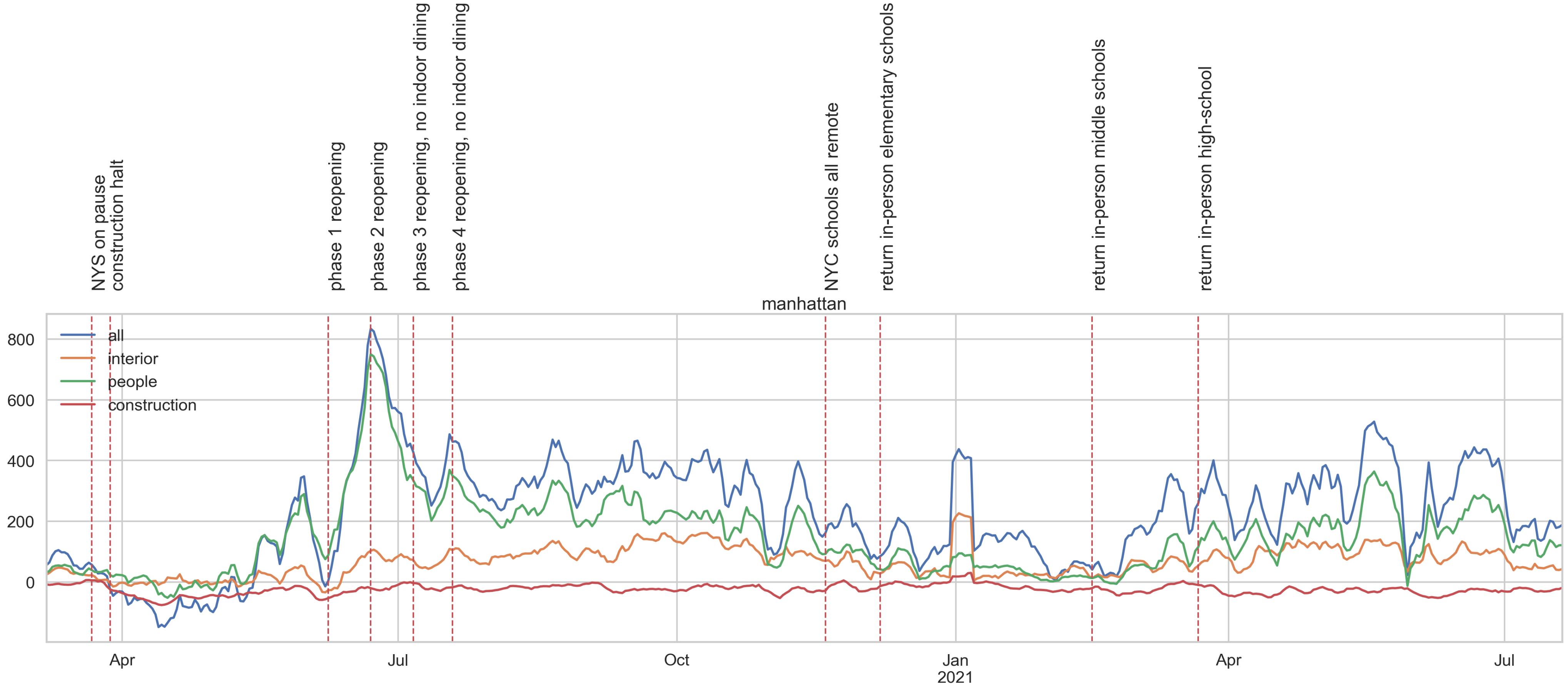


UIC COMPUTER SCIENCE

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.*"

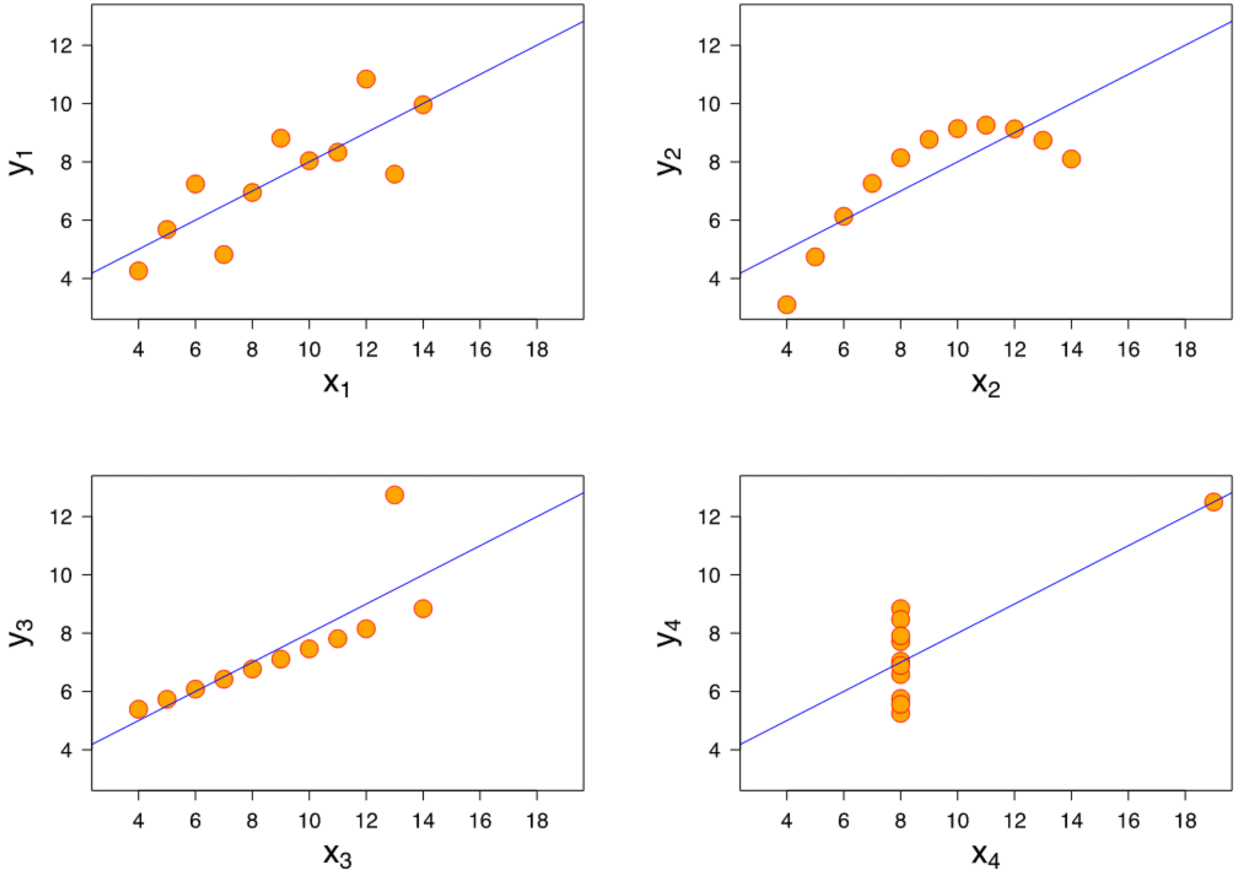
Ben Schneiderman

Importance of visualization

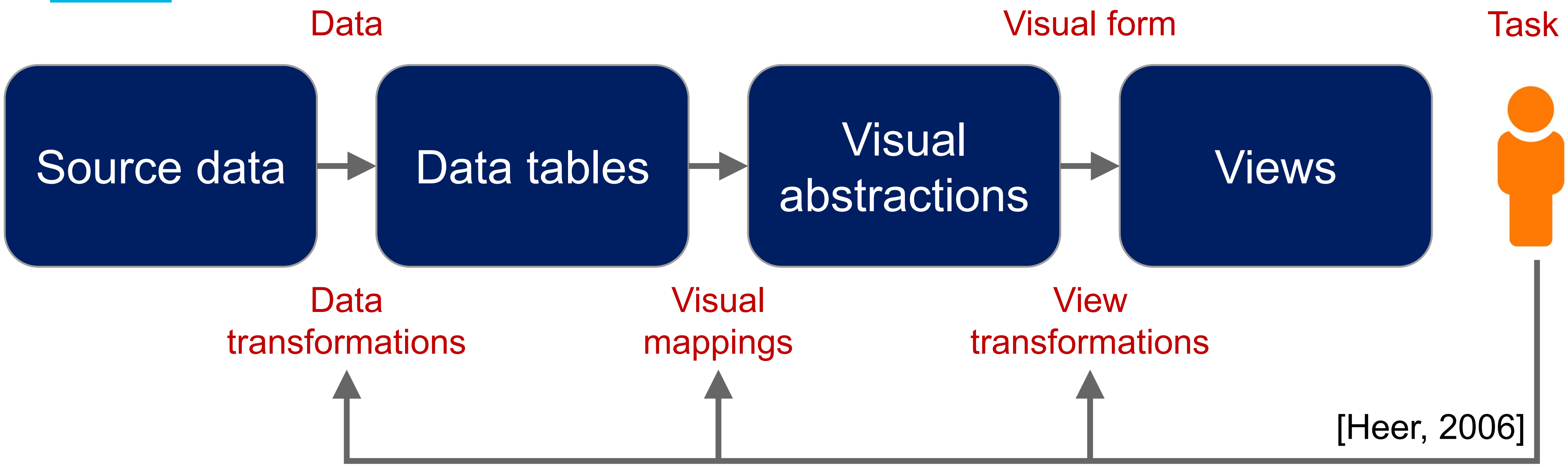
A		B		C		D	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26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

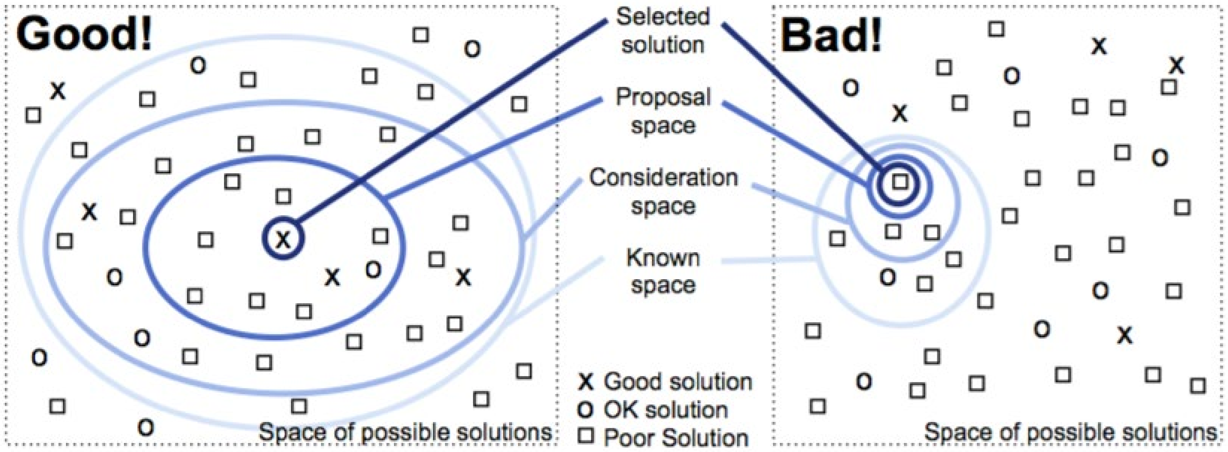


Visualization design



- 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

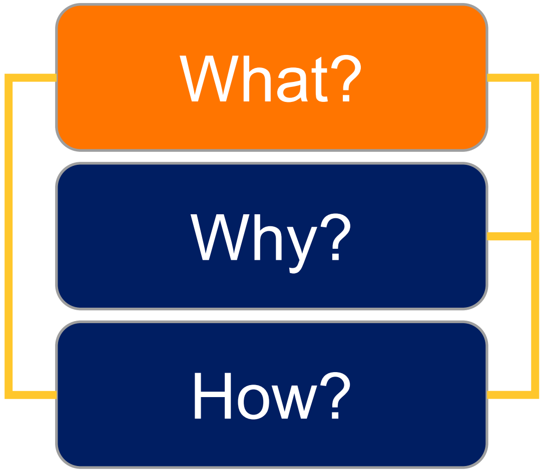


[Munzner, 2015]

Develop principles and techniques to build effective visualizations.

Visualization design

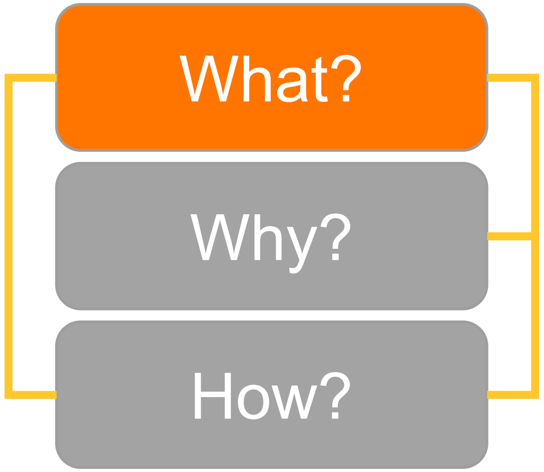
- High-level framework for analyzing vis use:
  - **What** data user sees?
  - **Why** the user intends to use a vis tool?
  - **How** the user intends to use a vis tool?



[Munzner, 2015]

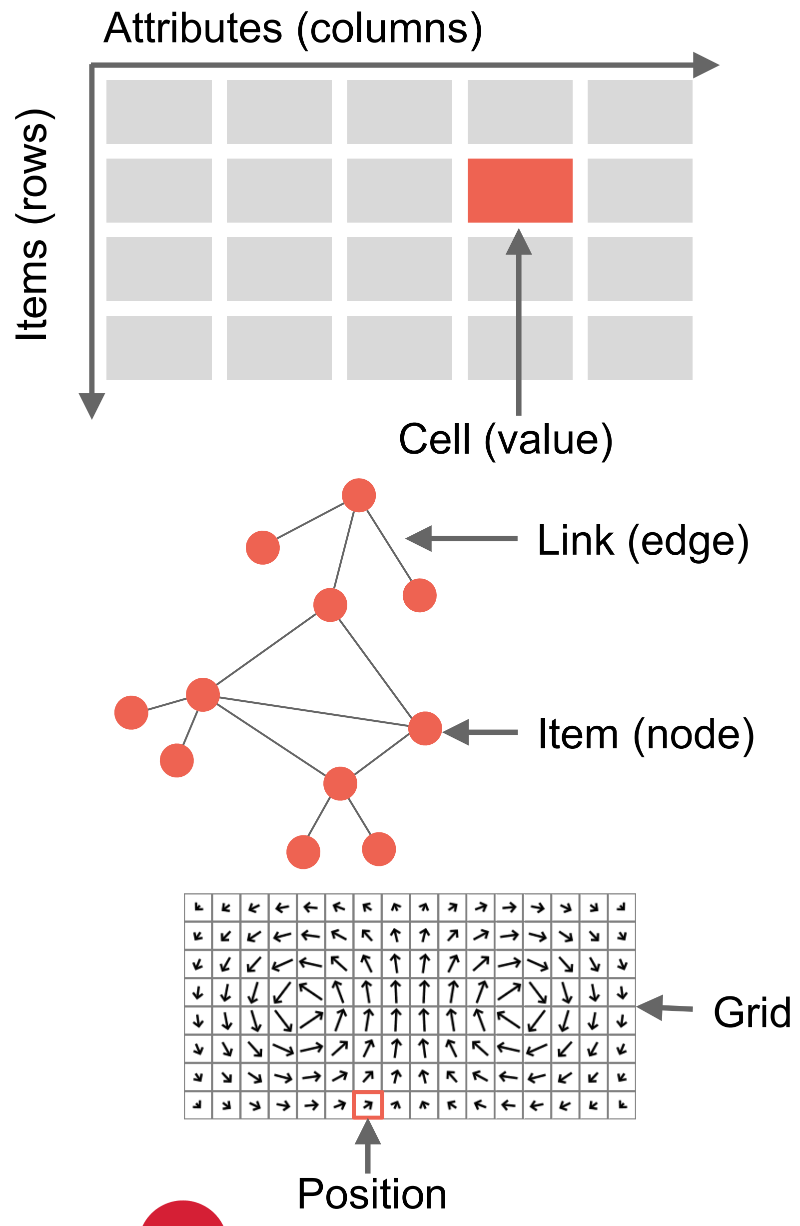
Principles of visualization

- Data
- Visual marks
- Visual channels
- Interaction



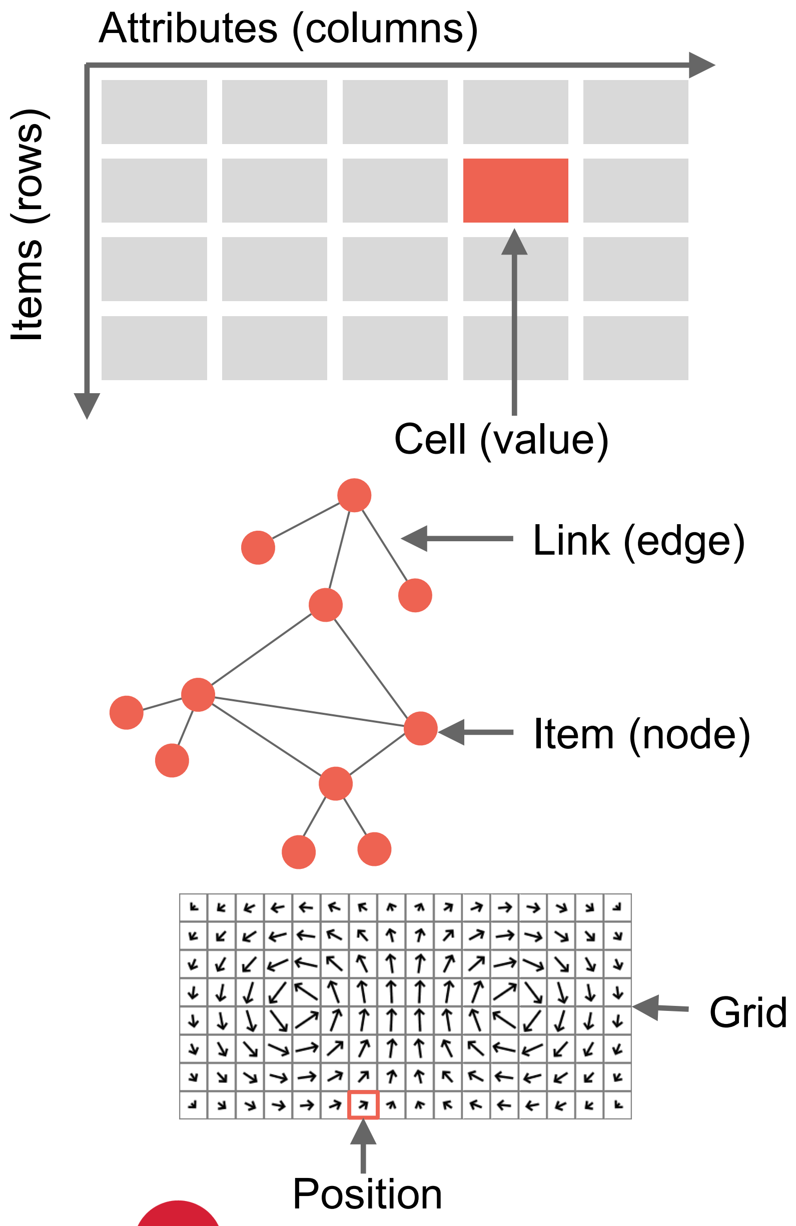
Dataset types

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



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.



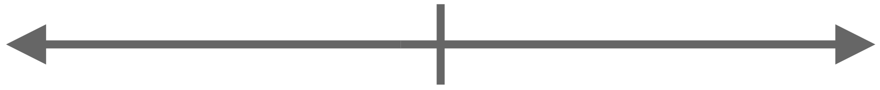
Attribute types

- 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, …

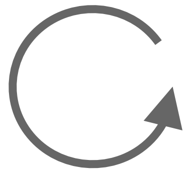
Attribute types

- 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.



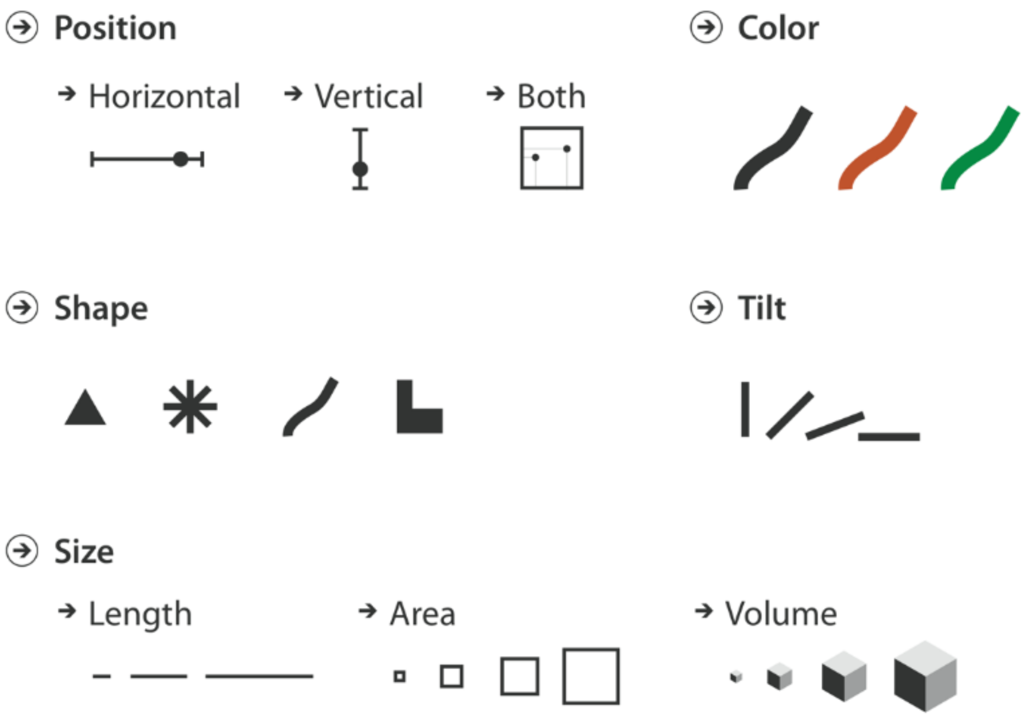
Visual marks

- 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

Channel and mark types

- Channel types:
  - Identify channels: what something is and where it is (circle, triangle, cross, etc.)
    - *What? Where?*
  - Magnitude channels: how much something there is (length, luminance, etc.)
    - *How much?*
- Mark types:
  - Item marks
  - Link marks: show relationship between items

Recap

- Visual marks: geometric elements that depict items and links.

What something is and where it is

- Visual channels: control marks' appearance.
  - Magnitude for ordered data.
  - Identify for categorial data.

How much something there is

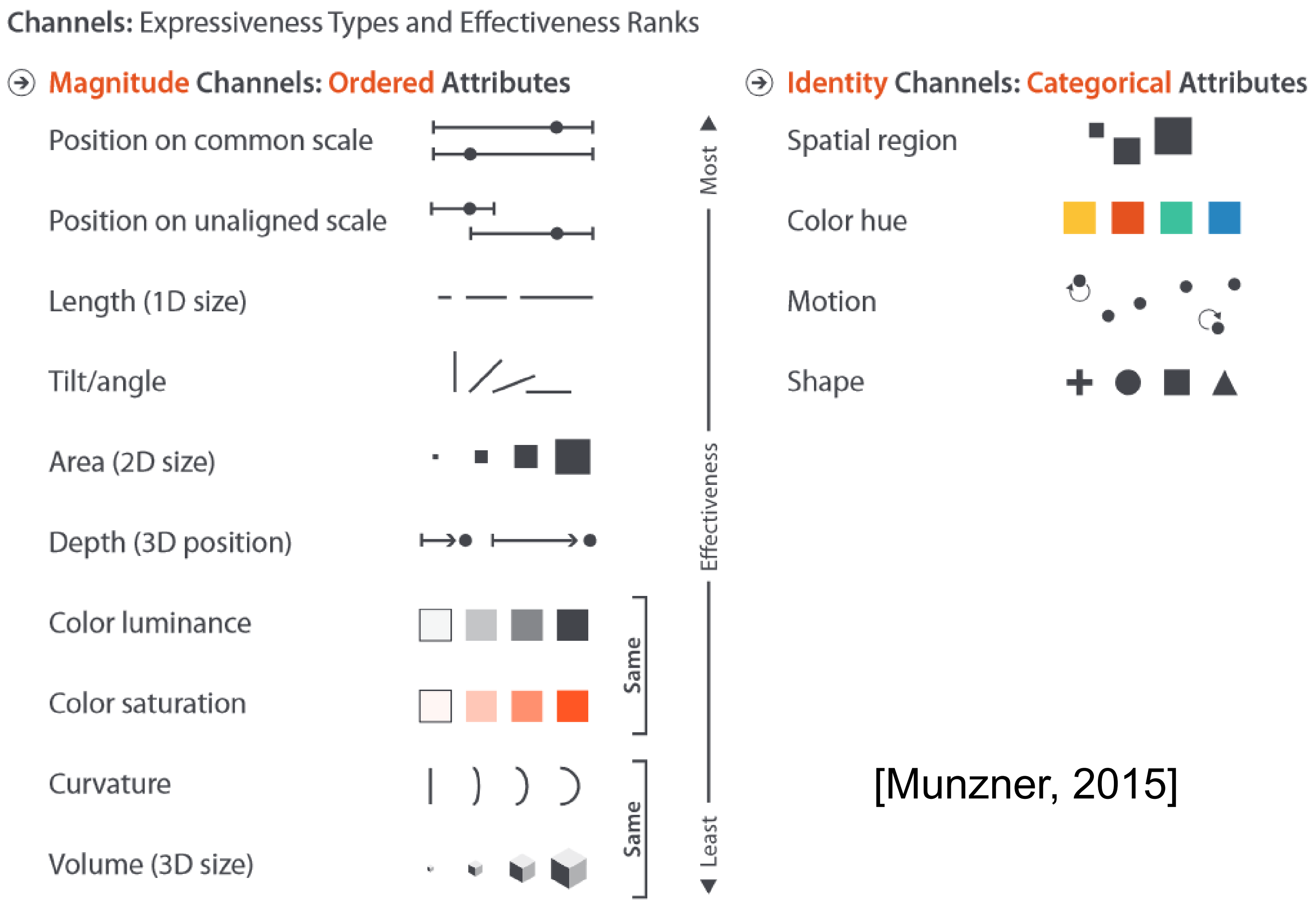
- Building blocks for visual encoding.

Choice of marks and channels

- Expressiveness: visual encoding should express all of the information in the dataset.

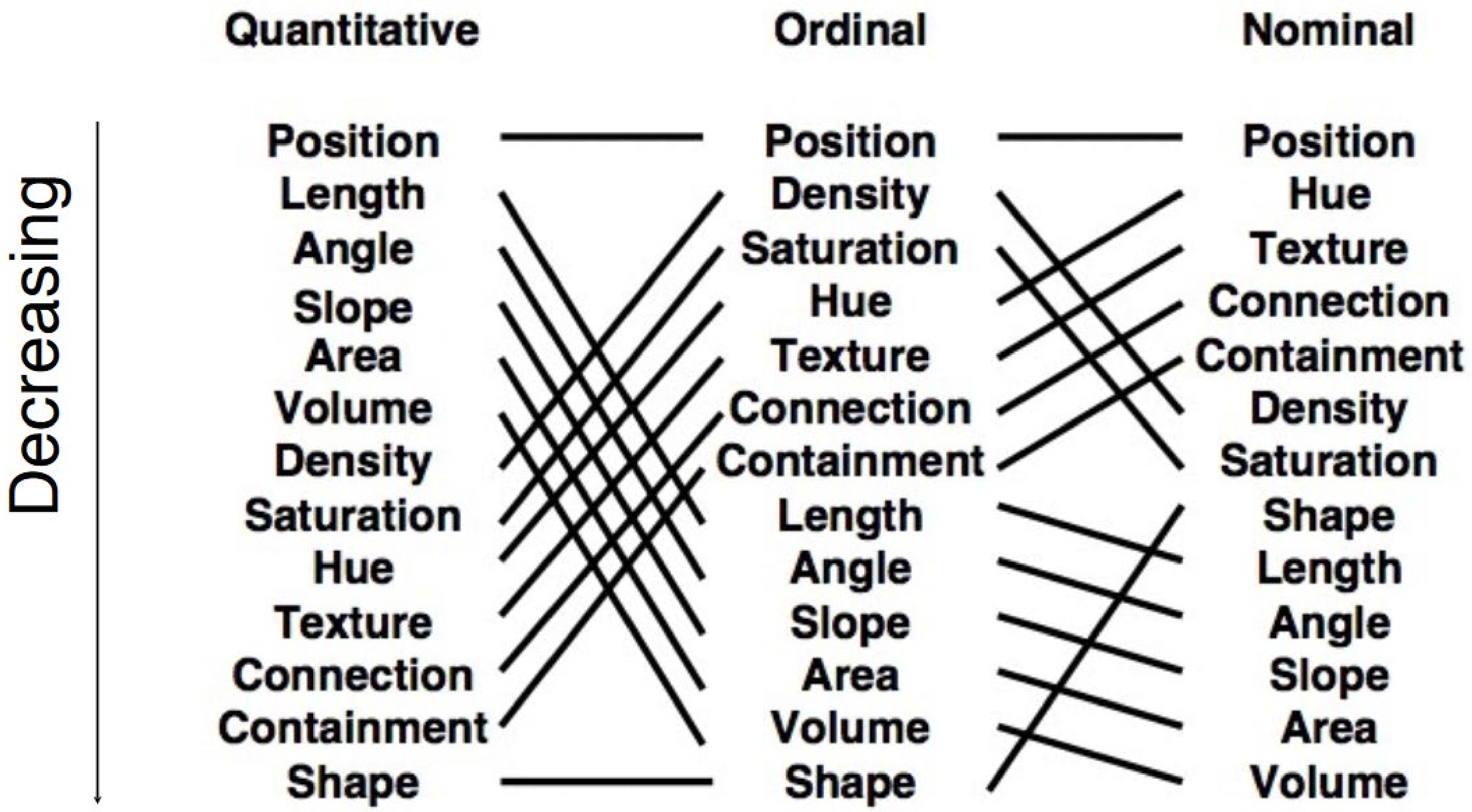
- Effectiveness: importance of the attribute should match the salience of the channel. Important items are the most salient.

Expressiveness types and effectiveness ranks



[Munzner, 2015]

Jock Mackinlay, 1986

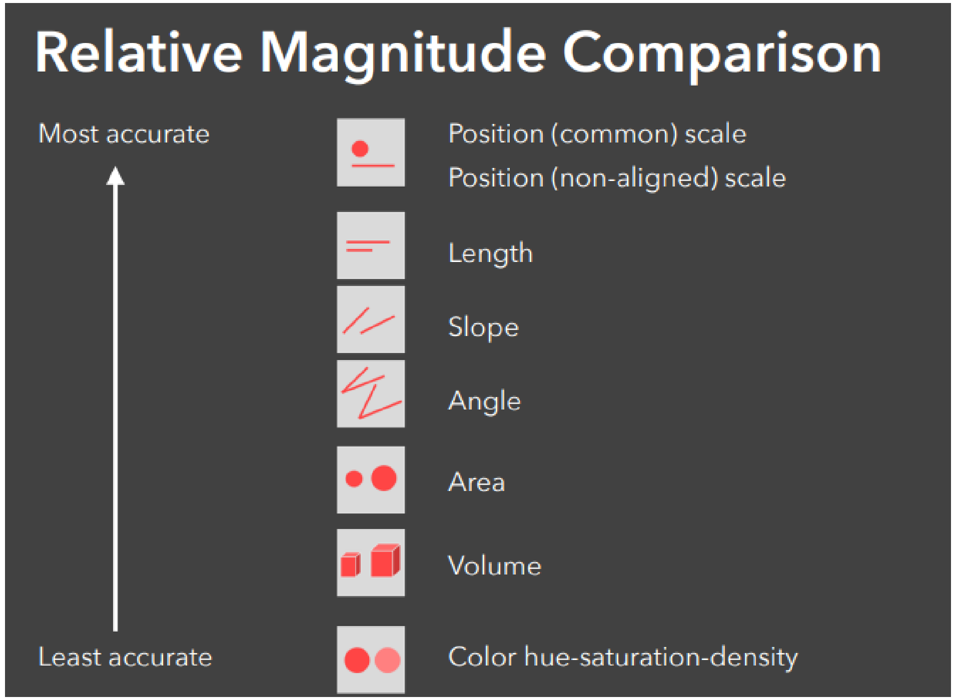


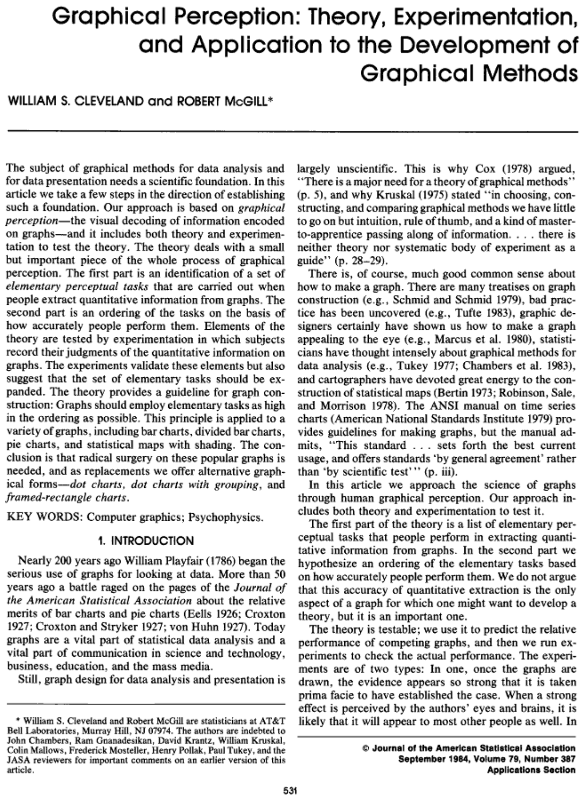
Channel effectiveness

- Ranking determined by:
    1. Accuracy: how well can a viewer decode the information in the channel?
    2. Discriminability: how easily can differences between attribute levels be perceived?
    3. Separability: can channels be used independently?
    4. Popup: can a channel provide popout in the visualization?
    5. Grouping: can a channel show perceptual grouping?

Channel effectiveness: accuracy

- Cleveland & McGill hierarchy:





Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

WILLIAM S. CLEVELAND and ROBERT McGILL*

The subject of graphical methods for data analysis and for data presentation needs a scientific foundation. In this article we take a few steps in the direction of establishing such a foundation. Our approach is based on *graphical perception*—the visual decoding of information encoded on graphs—and it includes both theory and experimentation to test the theory. The theory deals with a small but important piece of the whole process of graphical perception. The first part is an identification of a set of *elementary perceptual tasks* that are carried out when people extract quantitative information from graphs. The second part is an ordering of the tasks on the basis of how accurately people perform them. Elements of the theory are tested by experimentation in which subjects record their judgments of the quantitative information on graphs. The experiments validate these elements but also suggest that the set of elementary tasks should be expanded. The theory provides a guideline for graph construction: Graphs should employ elementary tasks as high in the ordering as possible. This principle is applied to a variety of graphs, including bar charts, divided bar charts, pie charts, and statistical maps with shading. The conclusion is that radical surgery on these popular graphs is needed, and as replacements we offer alternative graphical forms—*dot charts*, *dot charts with grouping*, and *framed-rectangle charts*.

KEY WORDS: Computer graphics; Psychophysics.

1. INTRODUCTION

Nearly 200 years ago William Playfair (1786) began the serious use of graphs for looking at data. More than 50 years ago a battle raged on the pages of the *Journal of the American Statistical Association* about the relative merits of bar charts and pie charts (Eells 1926; Croxton 1927; Croxton and Stryker 1927; von Huhn 1927). Today graphs are a vital part of statistical data analysis and a vital part of communication in science and technology, business, education, and the mass media.

Still, graph design for data analysis and presentation is largely unscientific. This is why Cox (1978) argued, "There is a major need for a theory of graphical methods" (p. 5), and why Kruskal (1975) stated "in choosing, constructing, and comparing graphical methods we have little to go on but intuition, rule of thumb, and a kind of master-to-apprentice passing along of information. . . . there is neither theory nor systematic body of experiment as a guide" (p. 28–29).

There is, of course, much good common sense about how to make a graph. There are many treatises on graph construction (e.g., Schmid and Schmid 1979), bad practice has been uncovered (e.g., Tufte 1983), graphic designers certainly have shown us how to make a graph appealing to the eye (e.g., Marcus et al. 1980), statisticians have thought intensely about graphical methods for data analysis (e.g., Tukey 1977; Chambers et al. 1983), and cartographers have devoted great energy to the construction of statistical maps (Bertin 1973; Robinson, Sale, and Morrison 1978). The ANSI manual on time series charts (American National Standards Institute 1979) provides guidelines for making graphs, but the manual admits, "This standard . . . sets forth the best current usage, and offers standards 'by general agreement' rather than 'by scientific test'" (p. iii).

In this article we approach the science of graphs through human graphical perception. Our approach includes both theory and experimentation to test it.

The first part of the theory is a list of elementary perceptual tasks that people perform in extracting quantitative information from graphs. In the second part we hypothesize an ordering of the elementary tasks based on how accurately people perform them. We do not argue that this accuracy of quantitative extraction is the only aspect of a graph for which one might want to develop a theory, but it is an important one.

The theory is testable; we use it to predict the relative performance of competing graphs, and then we run experiments to check the actual performance. The experiments are of two types: In one, once the graphs are drawn, the evidence appears so strong that it is taken prima facie to have established the case. When a strong effect is perceived by the authors' eyes and brains, it is likely that it will appear to most other people as well. In

* William S. Cleveland and Robert McGill are statisticians at AT&T Bell Laboratories, Murray Hill, NJ 07974.

The authors are indebted to John Chambers, Ram Gnanadesikan, David Krantz, William Kruskal, Colin Mallows, Frederick Mosteller, Henry Pollak, Paul Tukey, and the JASA reviewers for important comments on an earlier version of this article.

© Journal of the American Statistical Association
September 1984, Volume 79, Number 387
Applications Section

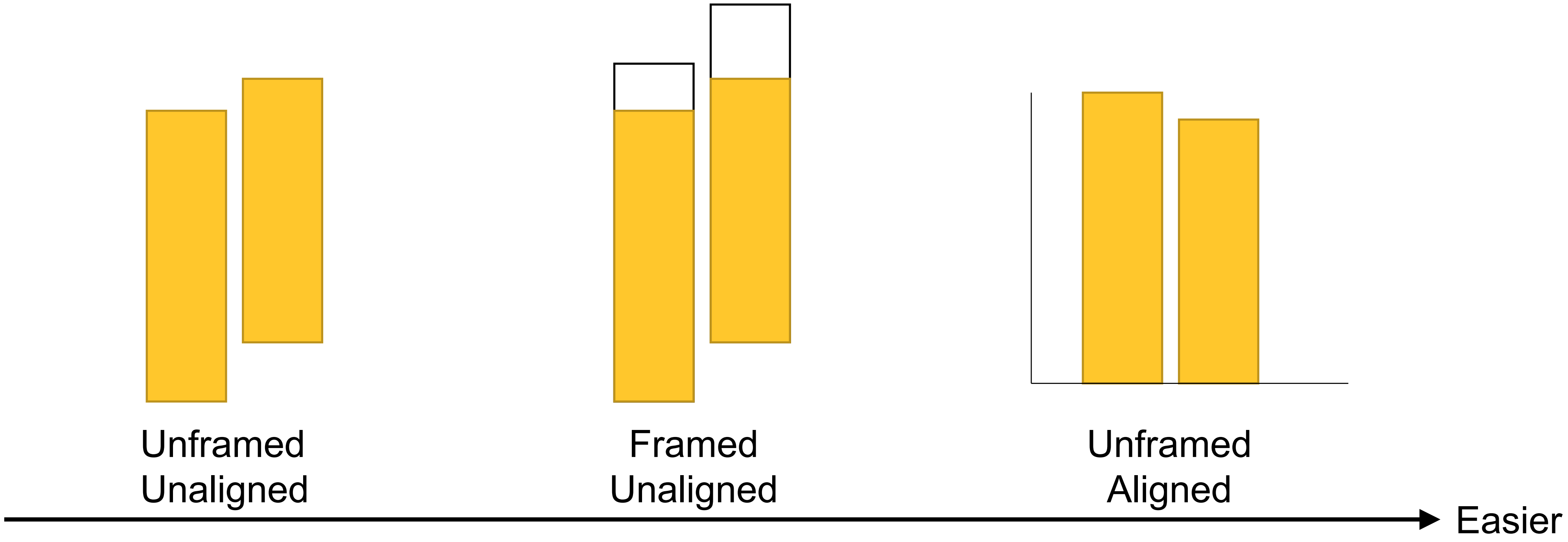
Channel effectiveness: accuracy

Channel effectiveness: accuracy

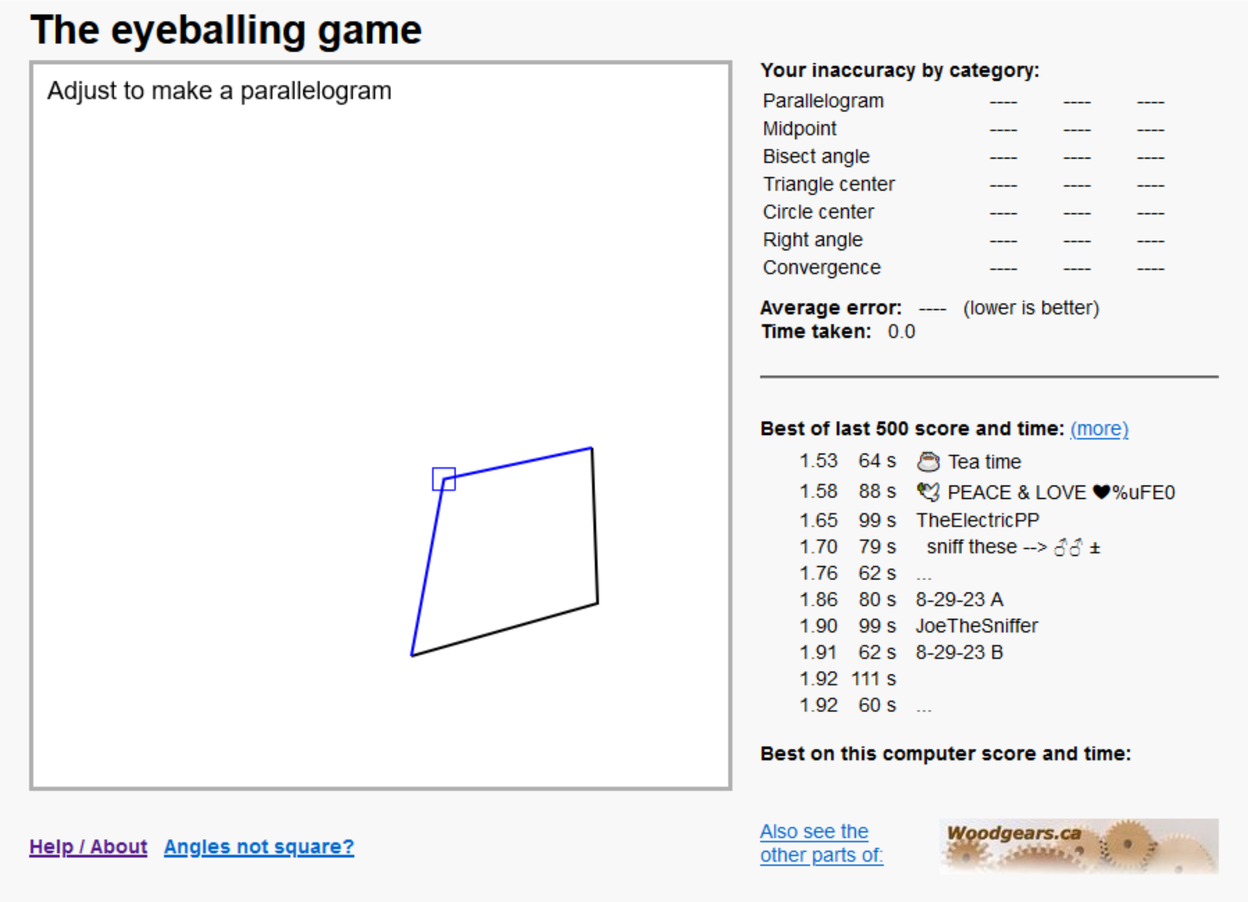
Channel effectiveness: accuracy

Relative vs absolute judgment

- Our perception is based on relative judgment, not absolute.

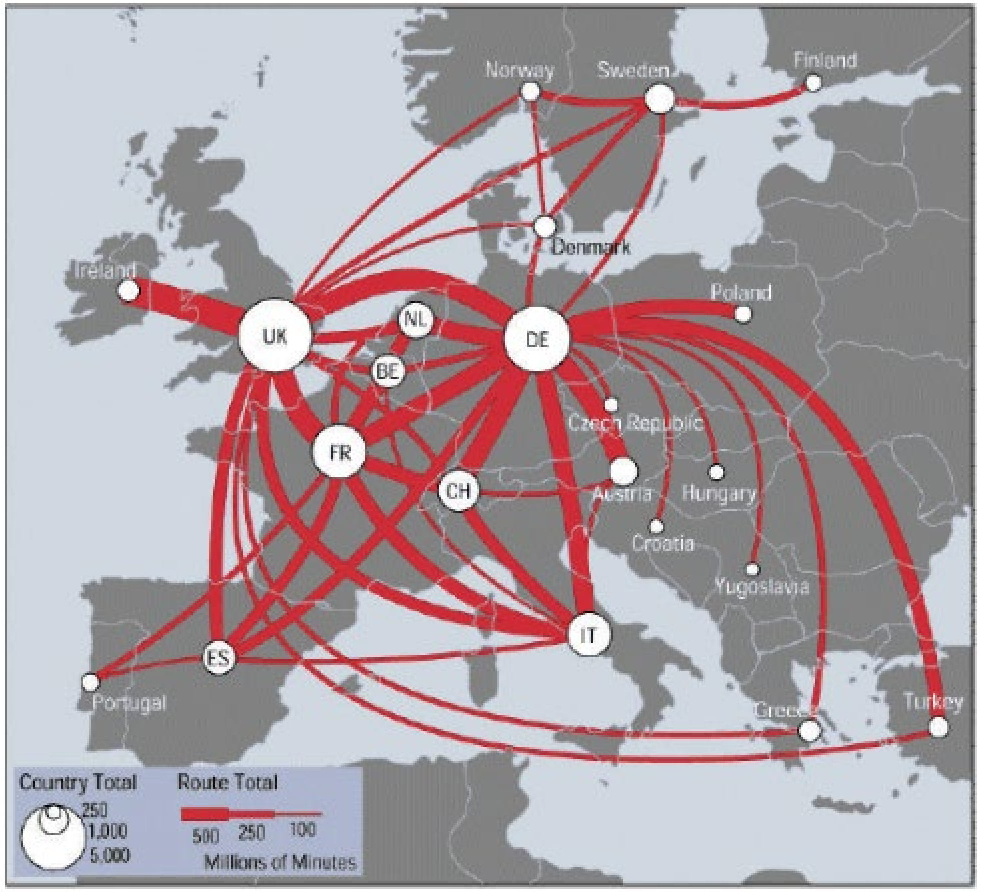


Position, length & angle



Channel effectiveness: discriminability

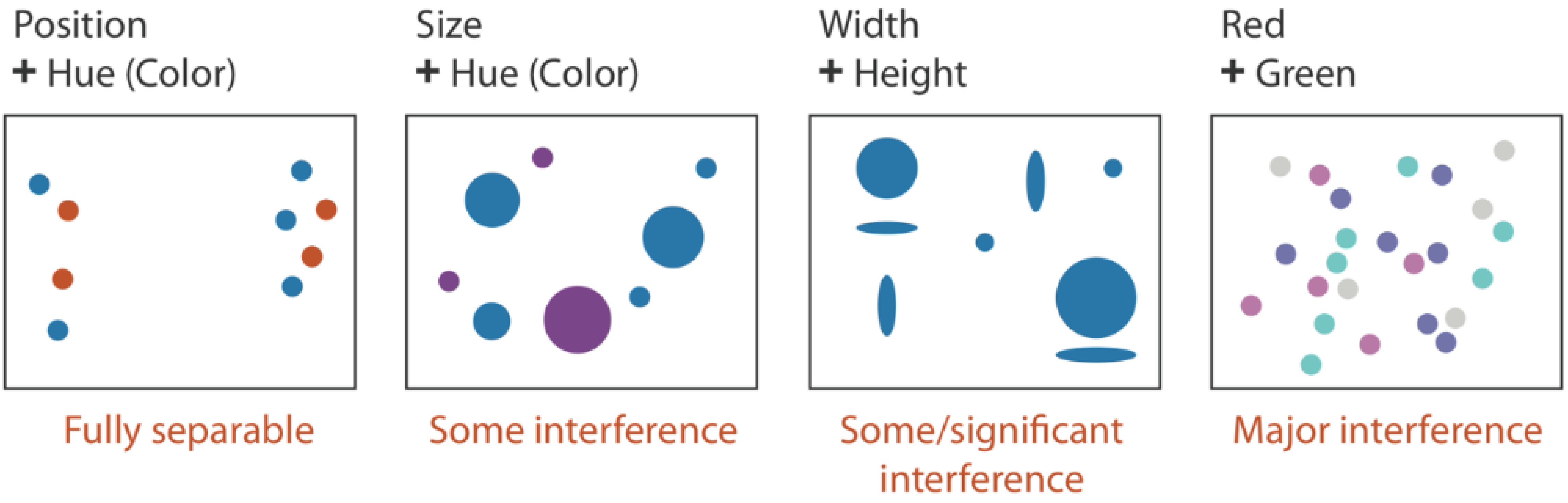
- Many channels can only support a limited number of discriminable (distinguishable) levels / bins.
  - Line width: up to 3 or 4
  - Color hues: up to 5 or 6



Channel effectiveness: separability

- Some encodings can be used independently of each other:
  - Vertical and horizontal position can be used independently.
  - Color (hue) and position can be used independently.
- Some encodings interfere with each other:
  - Width and height do not function well independently.
  - Two different values in the red and green channels does not work well.

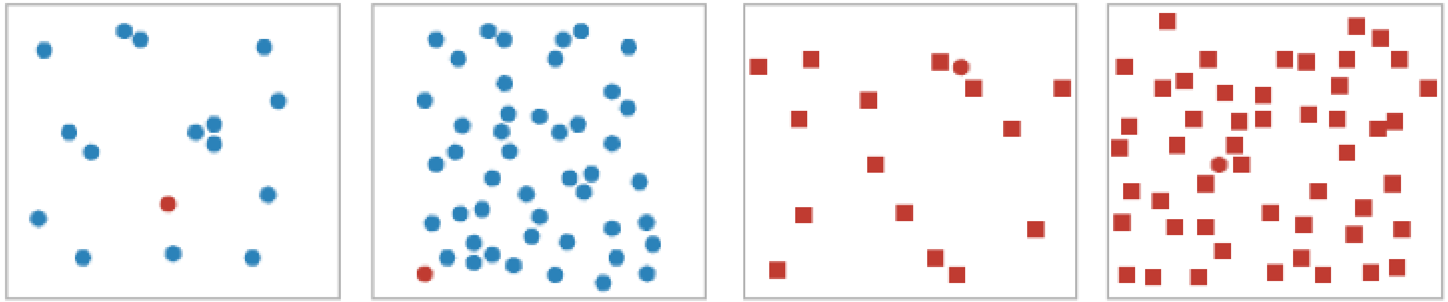
Channel effectiveness: separability



[Munzner, 2015]

Channel effectiveness: popout

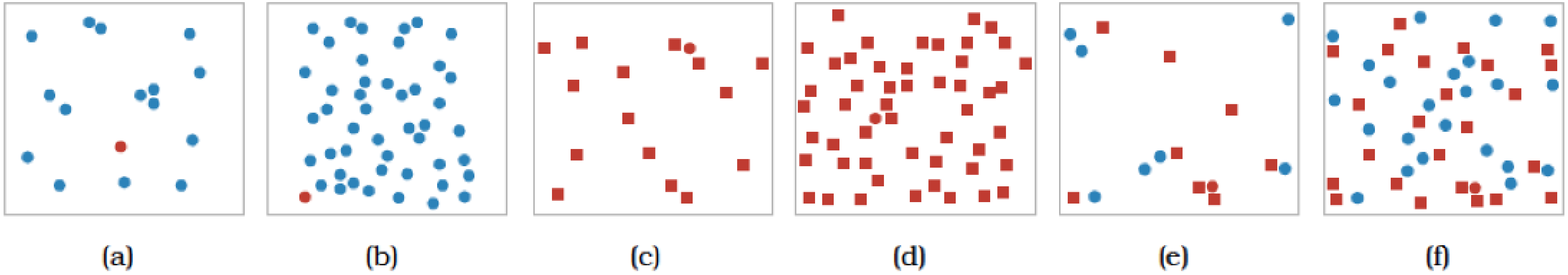
- Many channels support visual popout: one or few items stand out from others.



(a) (b) (c) (d)

Channel effectiveness: popout

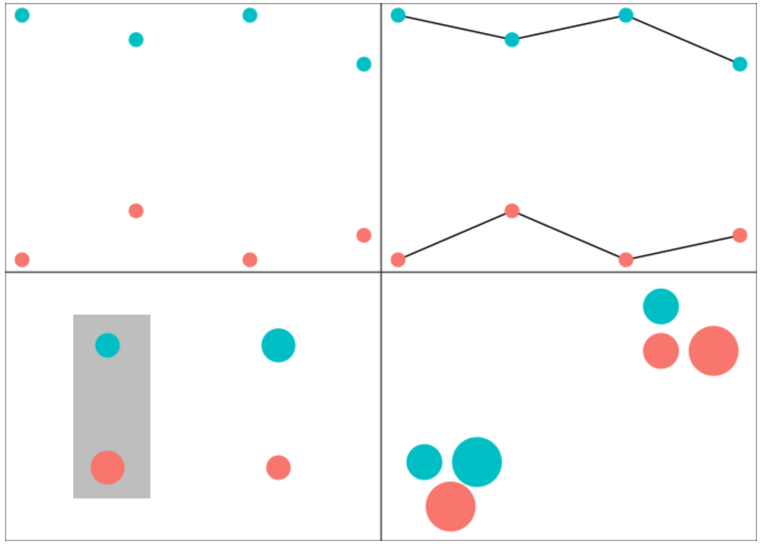
- Many channels support visual popout: one or few items stand out from others.



More difficult with multiple channels

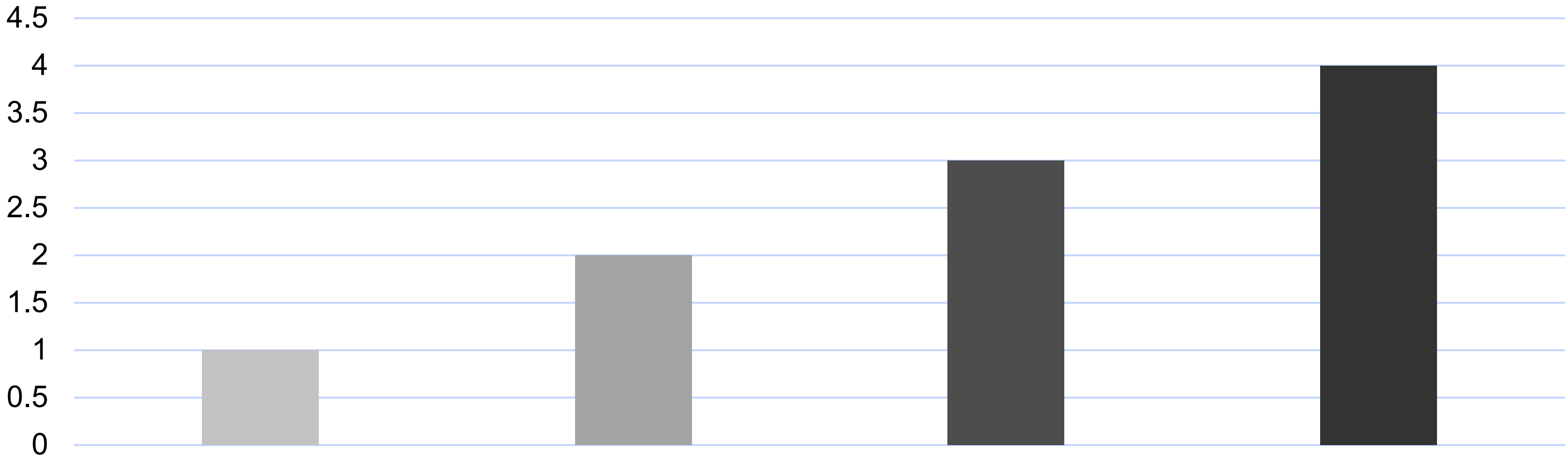
Channel effectiveness: grouping

- Perceptual grouping can be achieved by:
  - Identity channel to represent items as groups.
  - Using link marks.
  - Enclosure.
  - Spatial proximity.



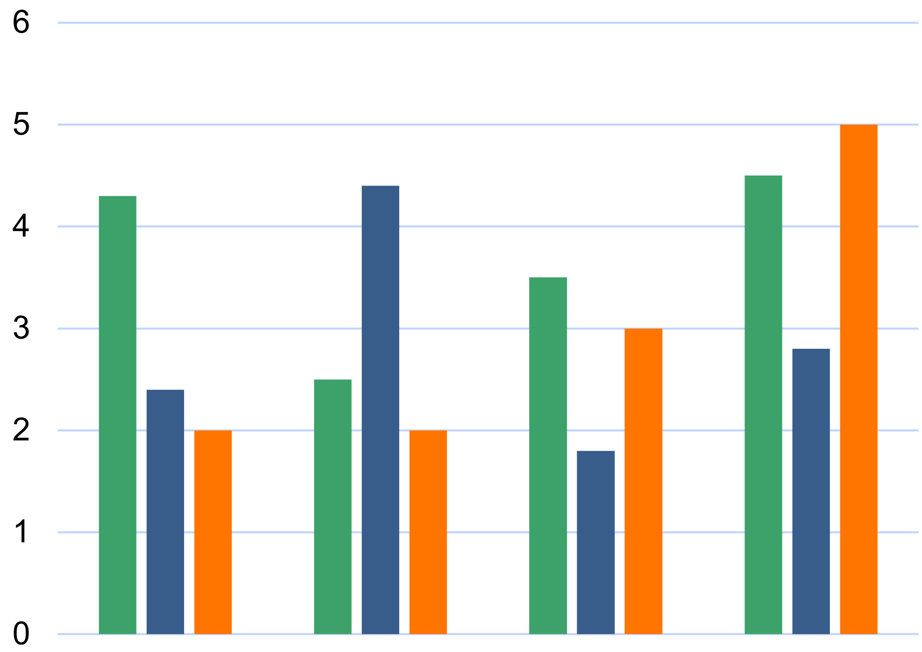
[Tierney, 2019]

Redundant encoding



Length, position, color

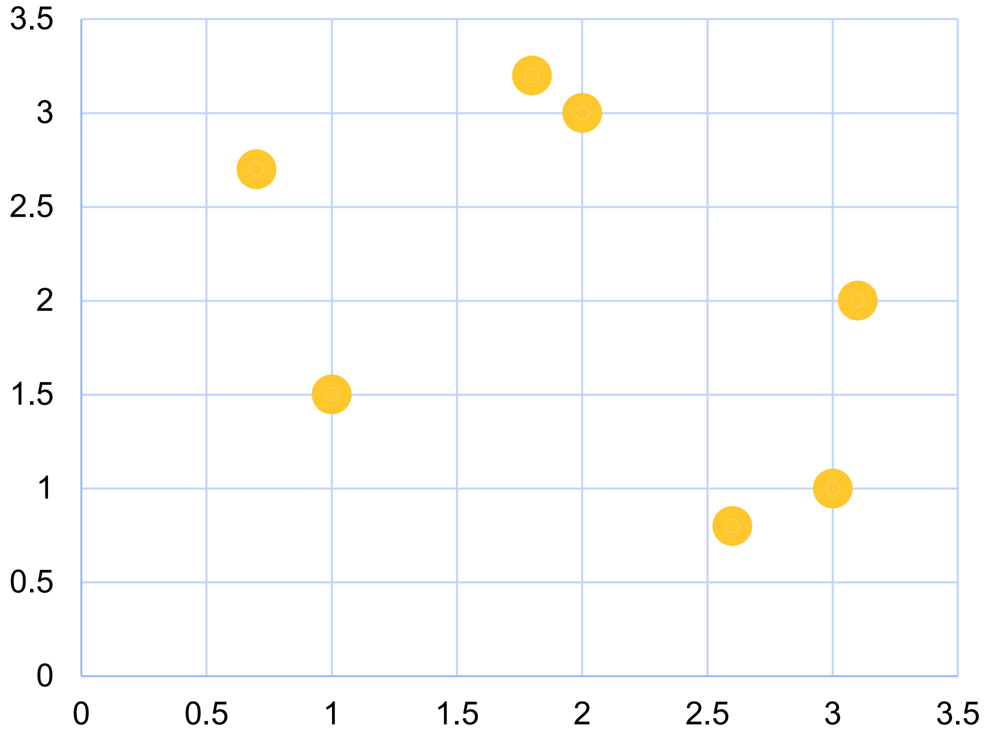
Visual marks & channels: example 1



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.

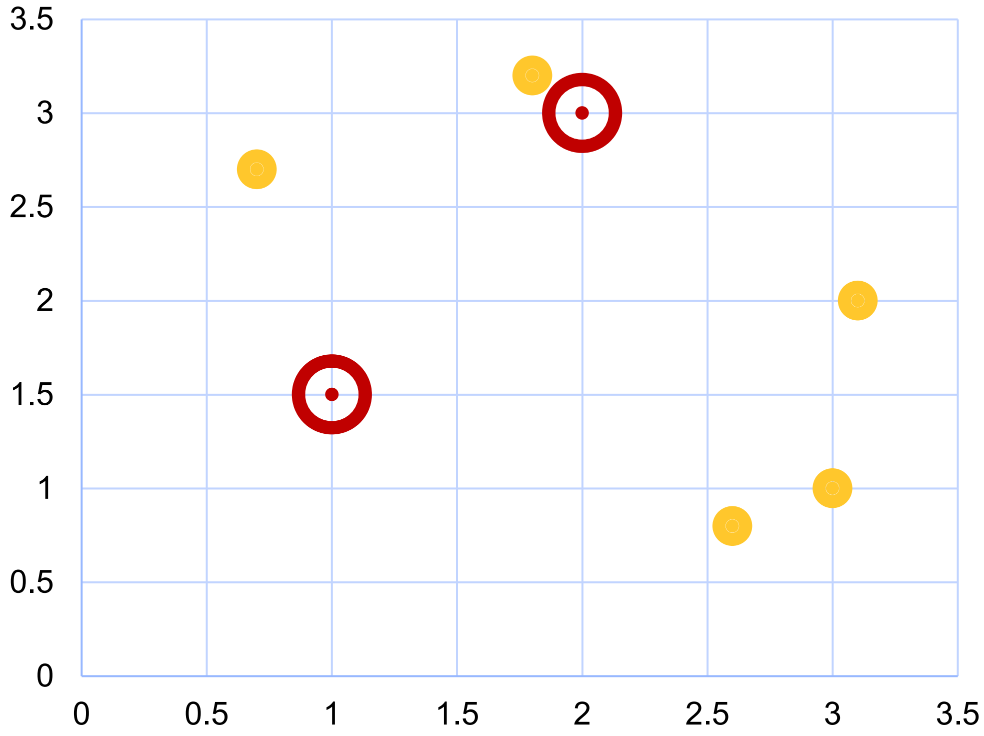
Visual marks & channels: example 2



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

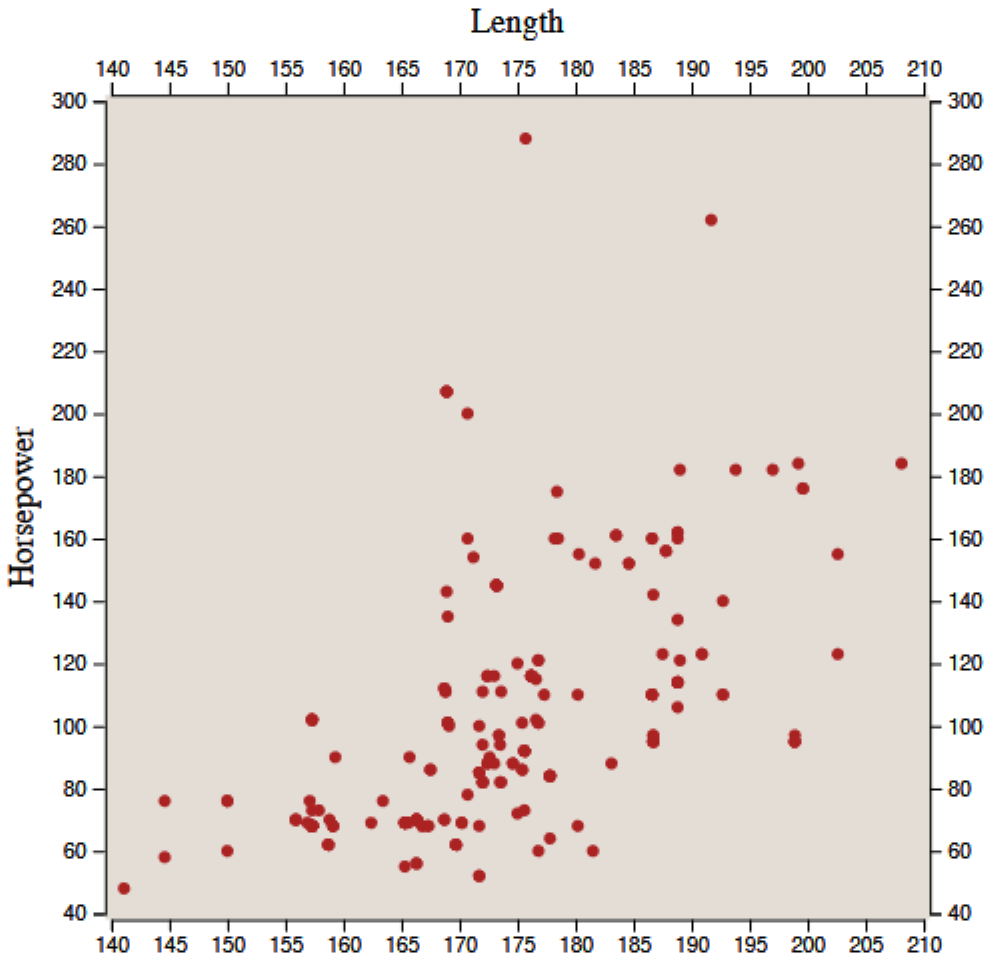


Scatterplots:

- 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

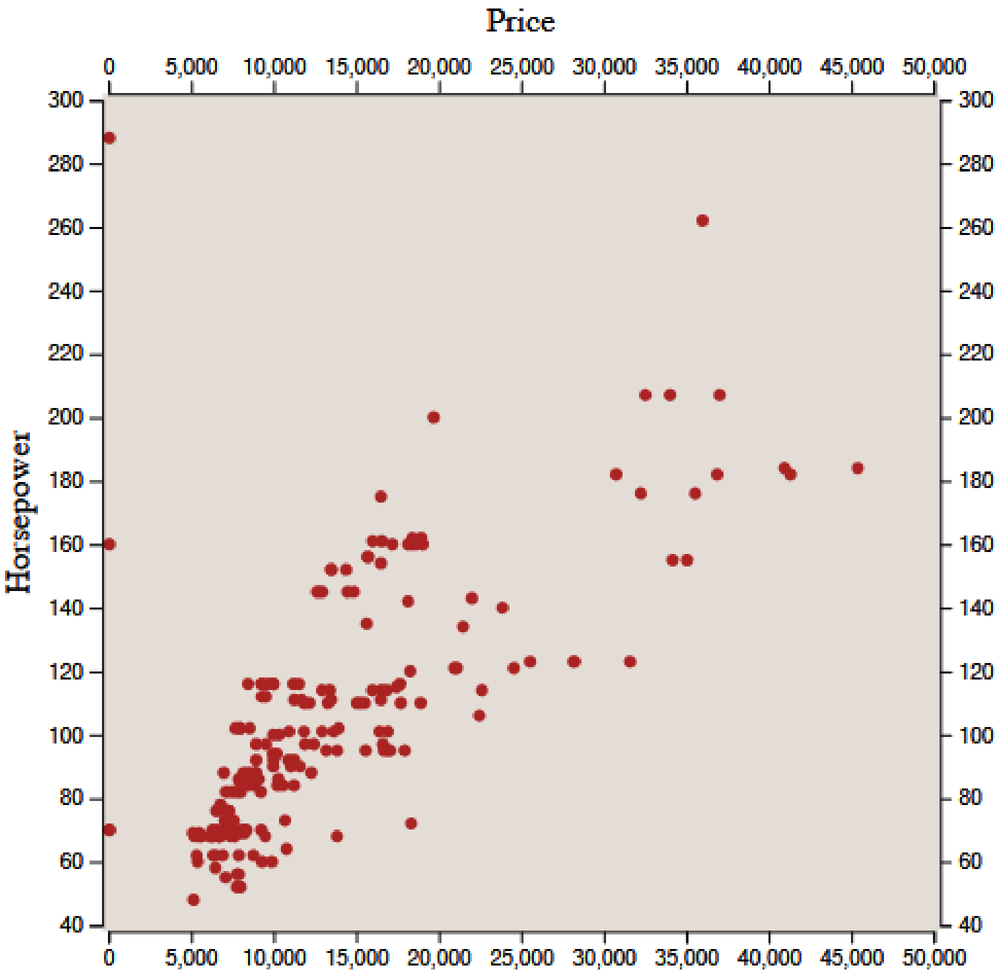


[Freire, 2019]

Marks: points
Channels: vertical and horizontal positions

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



[Freire, 2019]

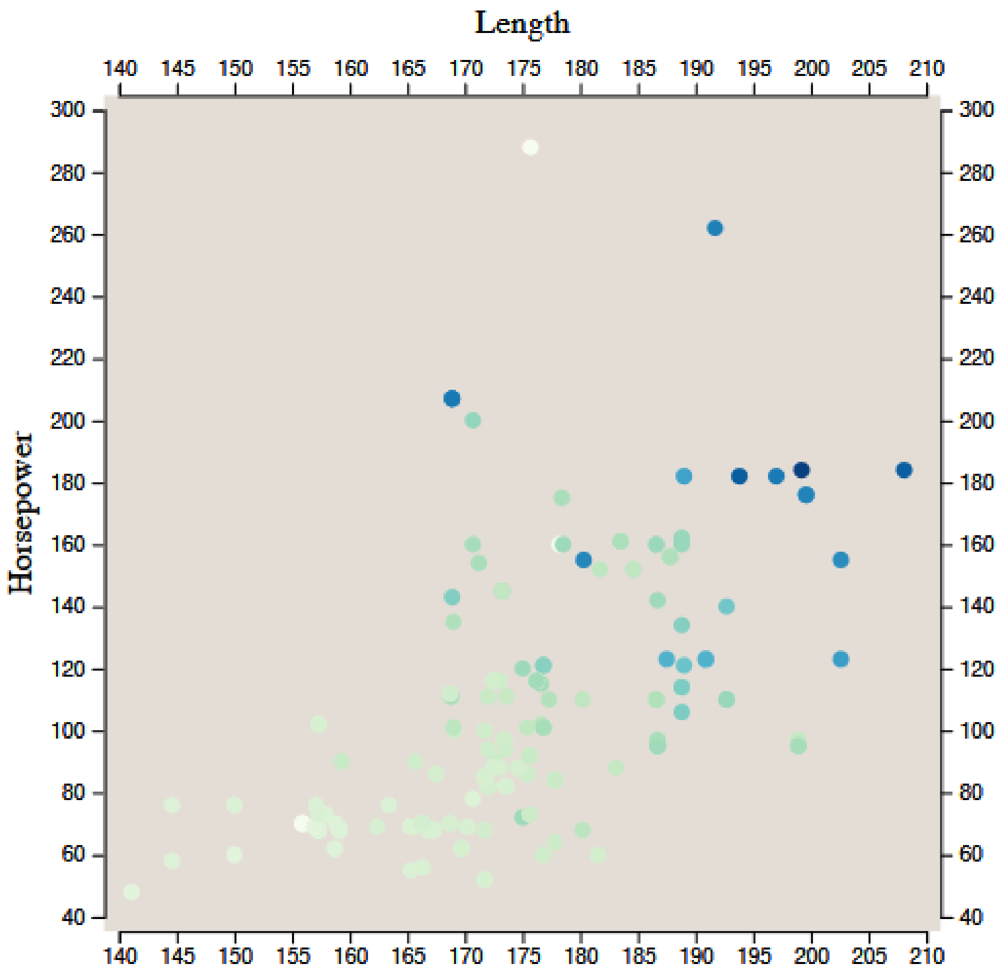
Marks: points
Channels: vertical and horizontal positions

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

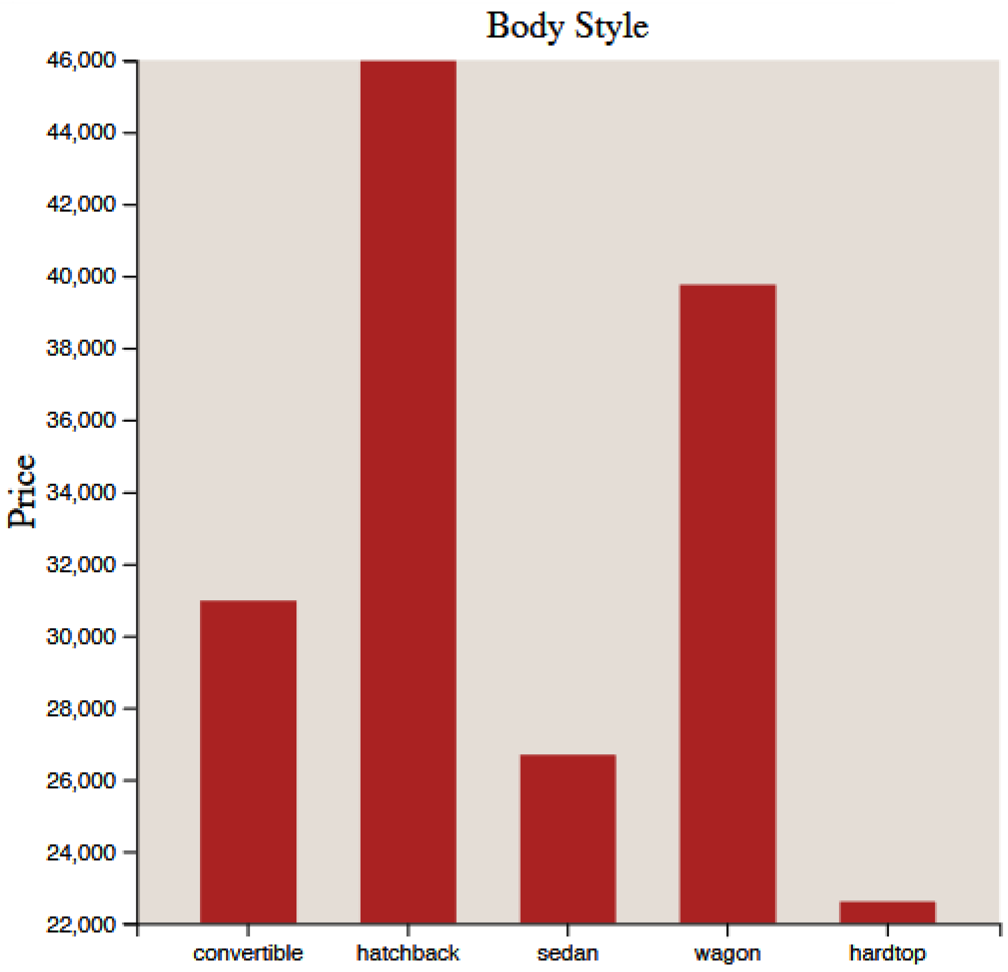
Channels: vertical and horizontal positions, color



[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



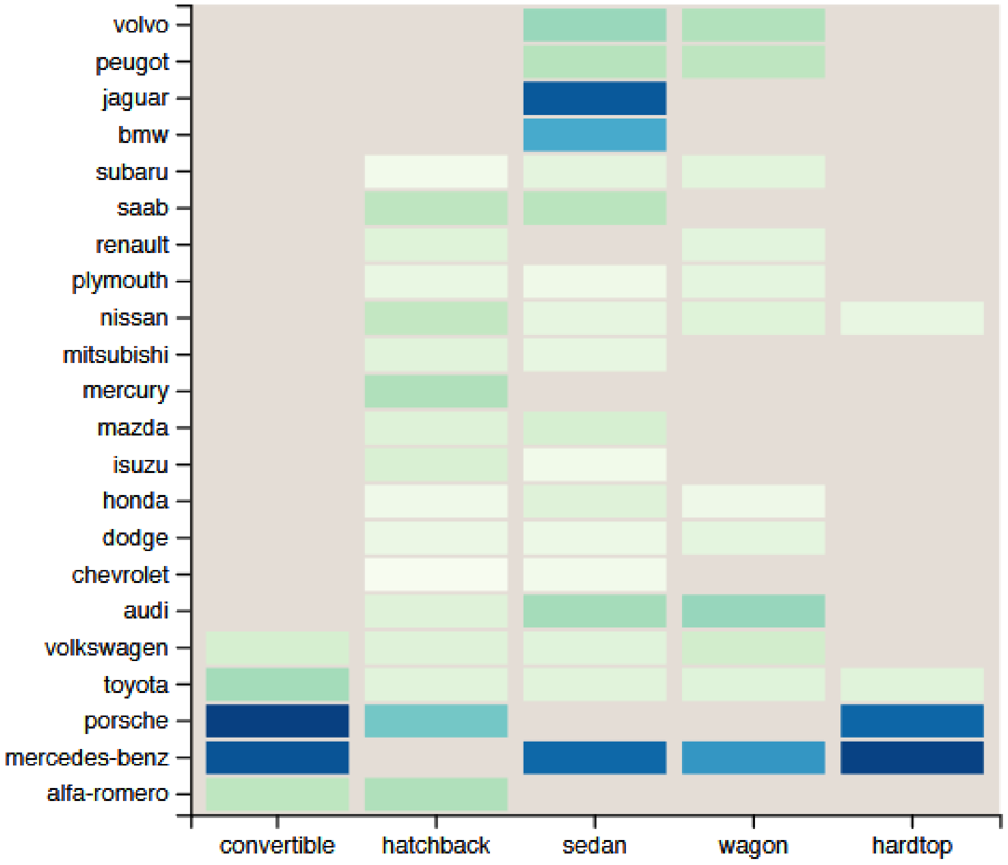
[Freire, 2019]

Marks: lines
Channels: vertical lengths and horizontal positions

Visual marks & channels: Cars

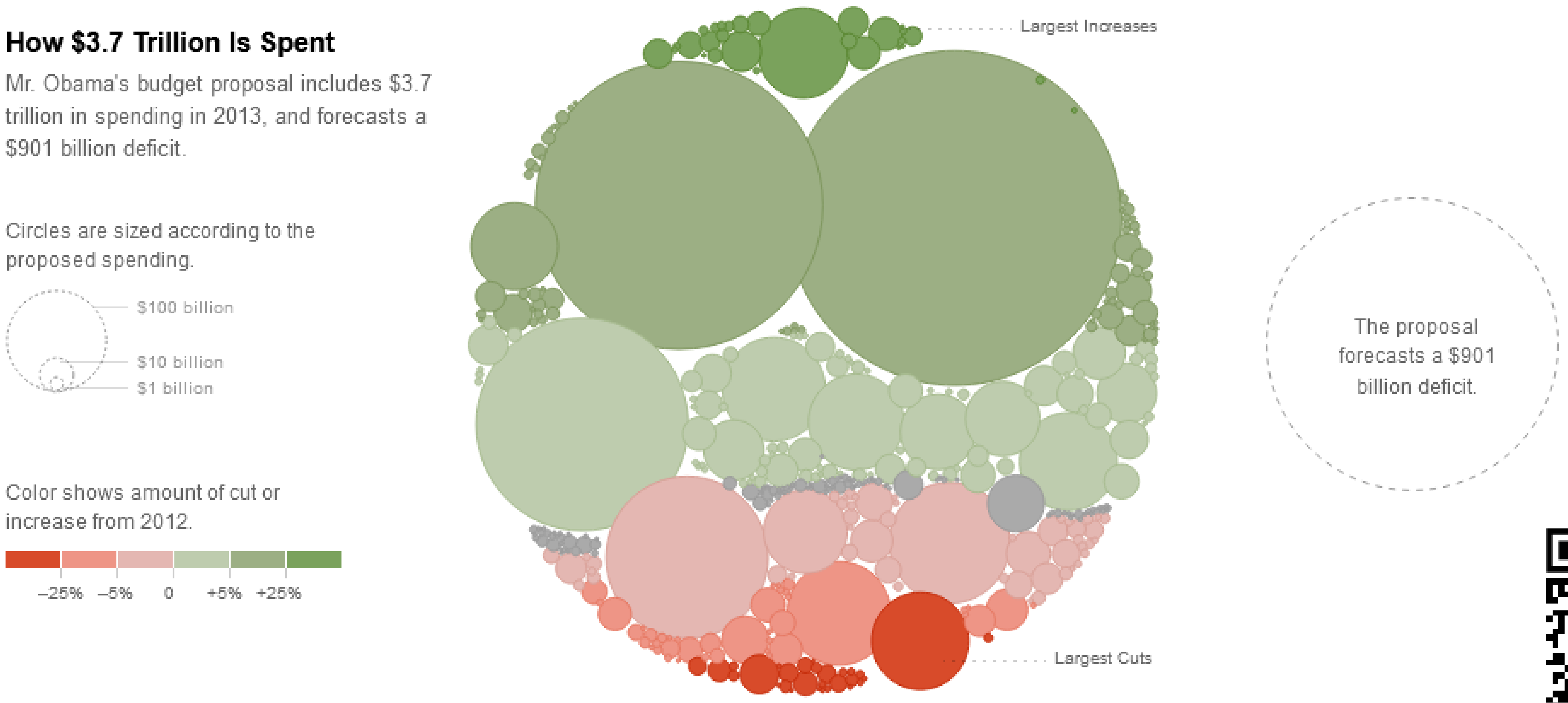
Car	HP	Price	Length	Style	Maker
Car 1	60	10000	130	Convertible	BMW
Car 2	86	12000	100	Hatchback	Audi
Car 3	55	11000	120	Wagon	Audi
Car 4	50	20000	80	Hatchback	Dodge

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

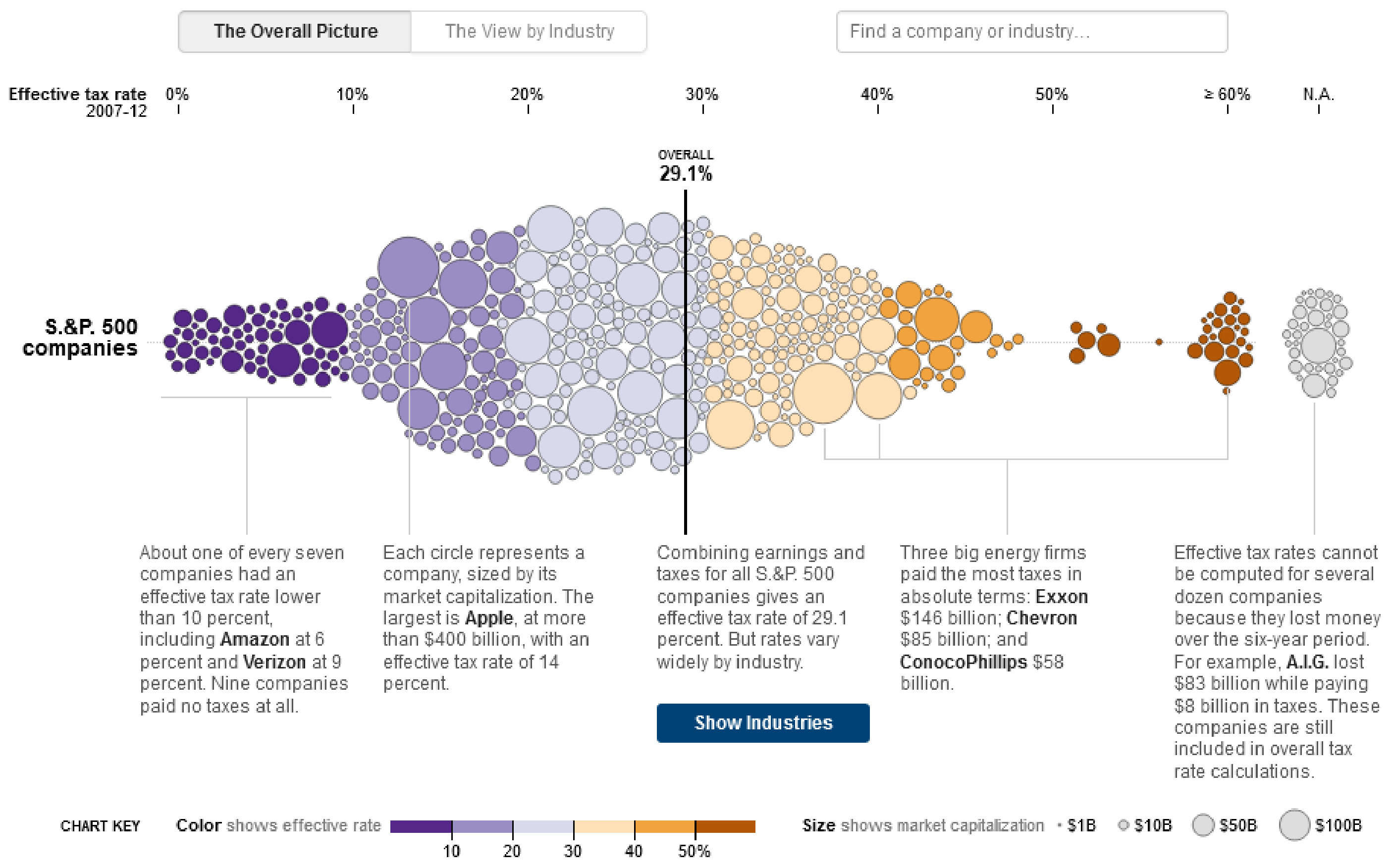


[Freire, 2019]

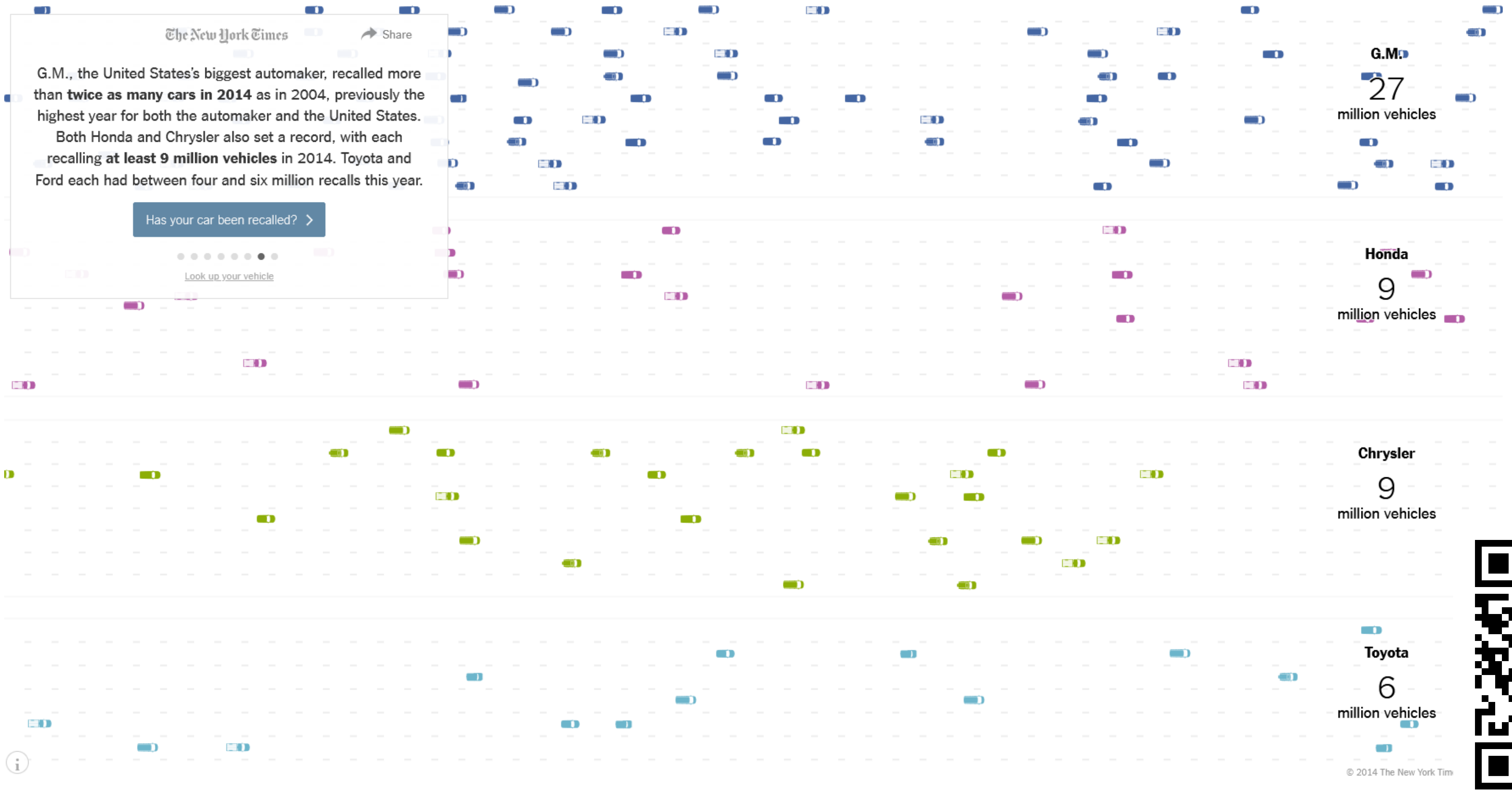
What visual variables are used?



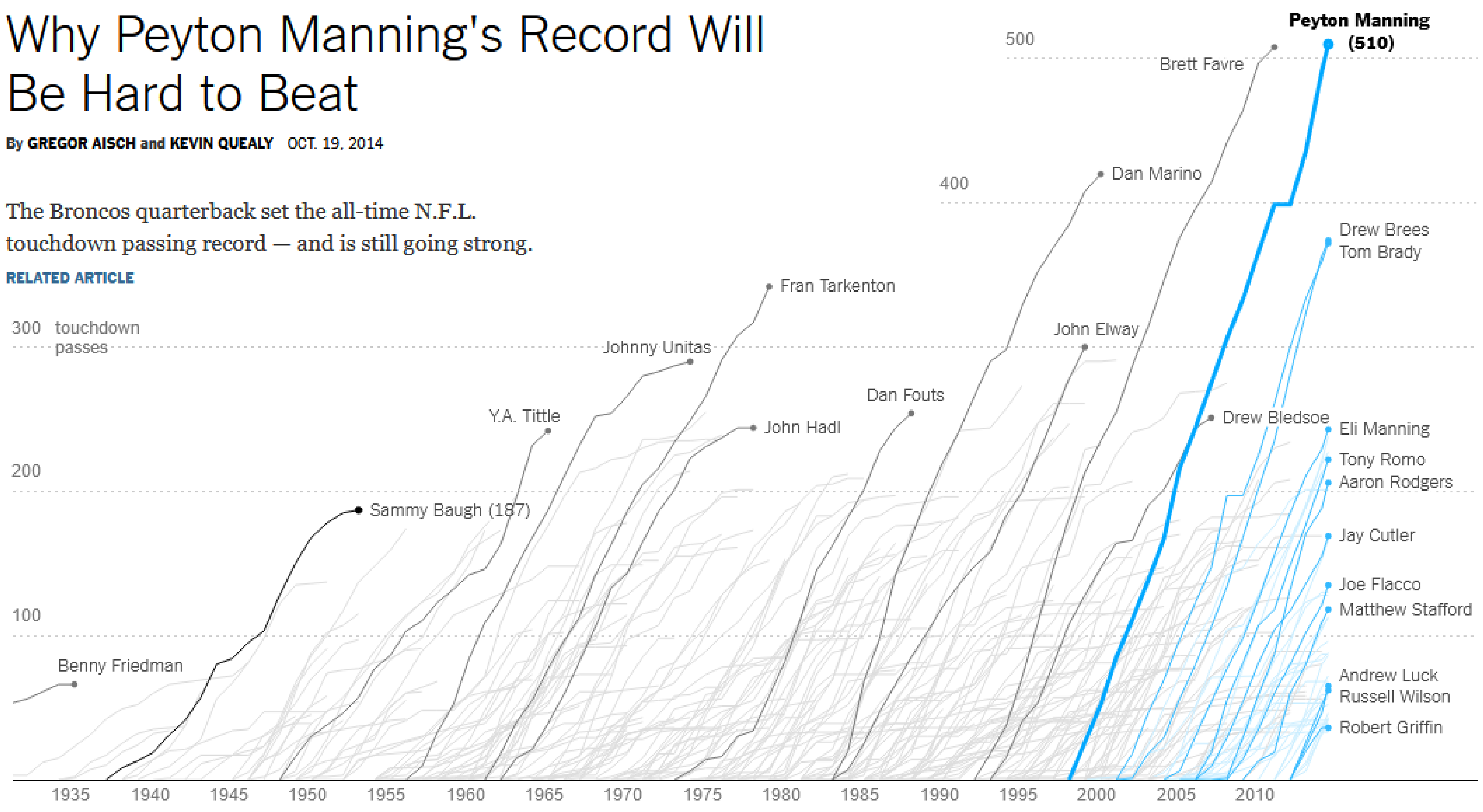
What visual variables are used?



What visual variables are used?

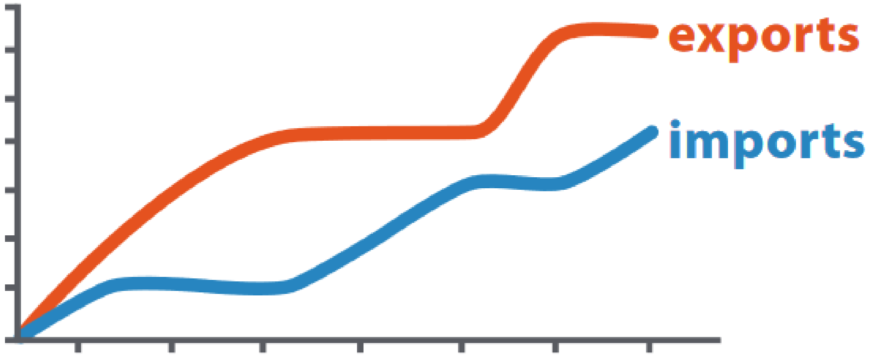


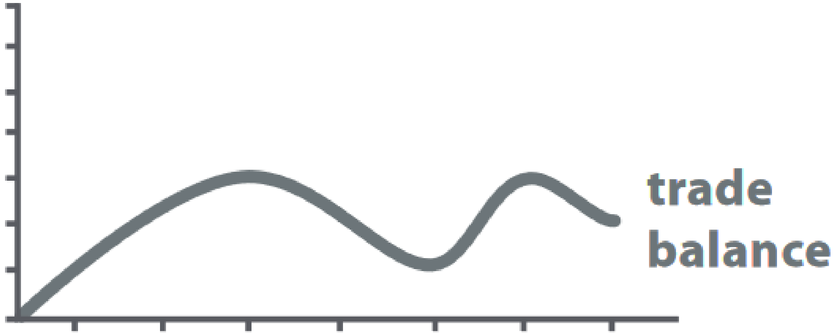
What visual variables are used?



Derive

- Decide what is the right thing to visualize.
- Transformations to extract it from the original dataset.





trade balance = exports − imports