Visual Model & Visual Encoding

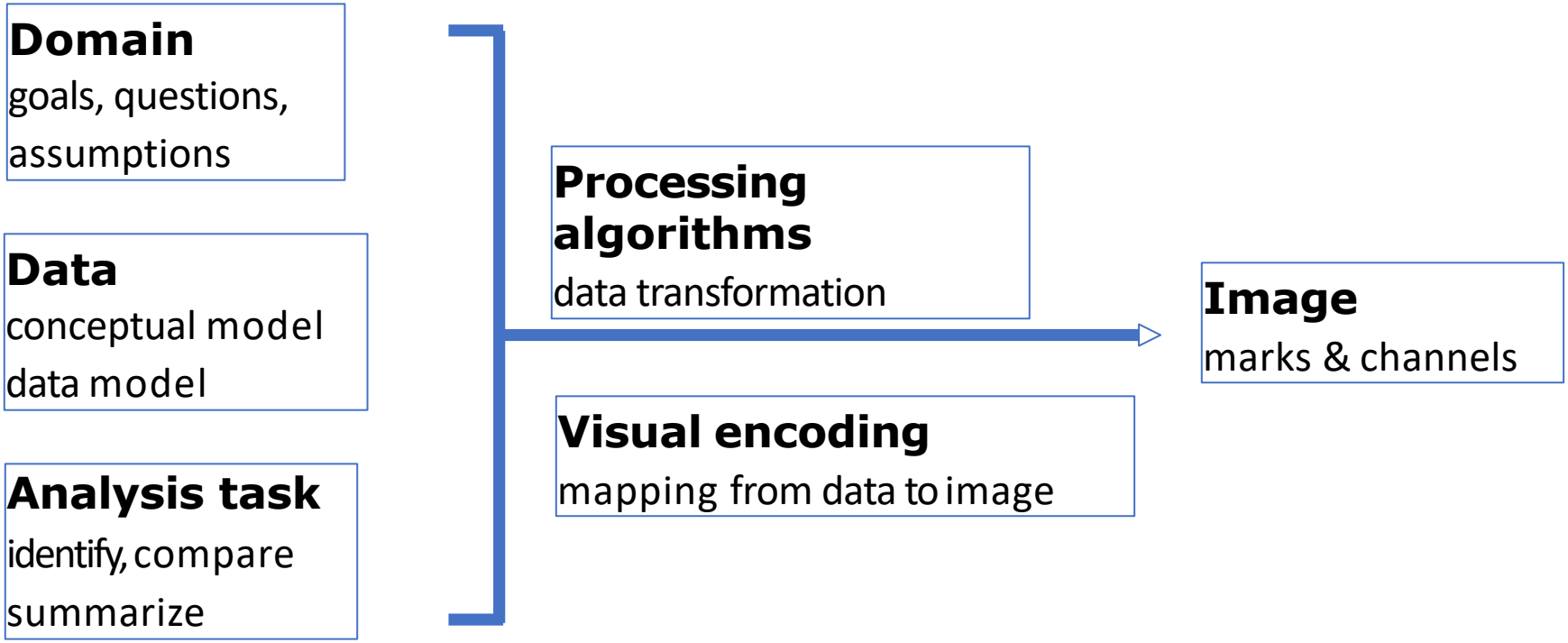
ONE LOVE. ONE FUTURE.

Credit: prof. Nam Wook Kim

Goal

- Learn how data is mapped to images

The big picture



[Slides from J. Heer]

Topics

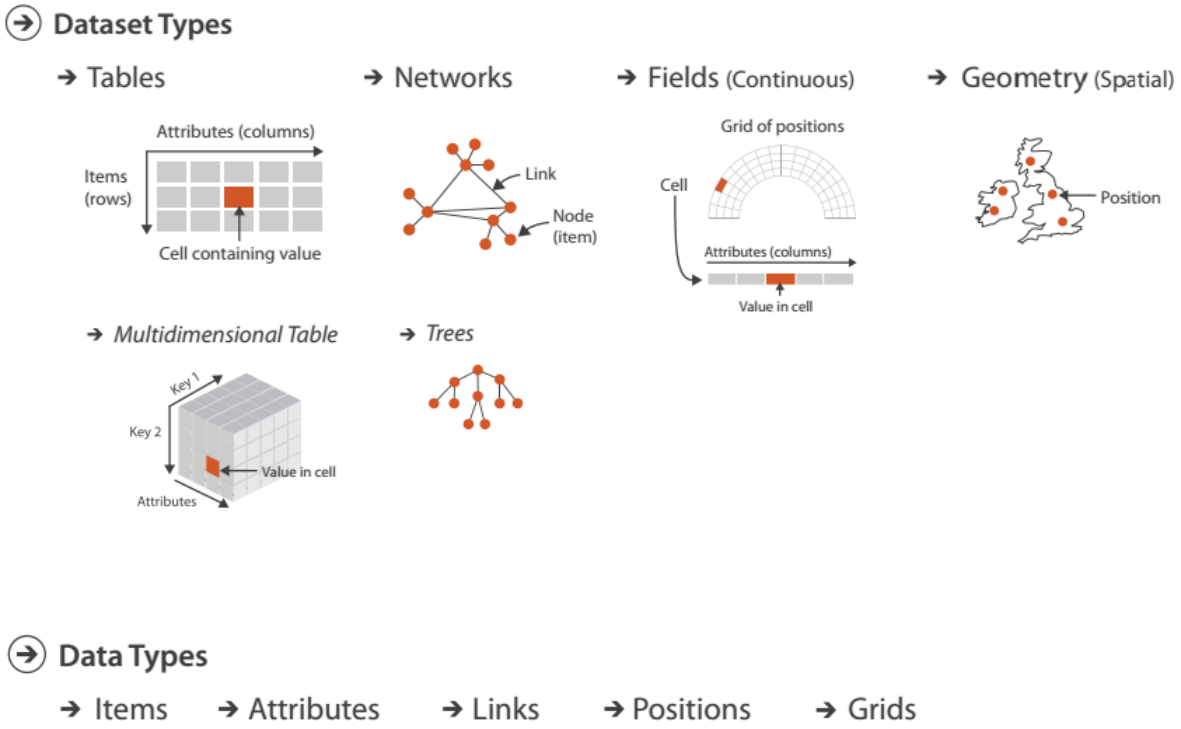
- Properties of Data
- Properties of Images
- Visual encoding: Mapping Data to Images

Properties of data

Taxonomy of datasets

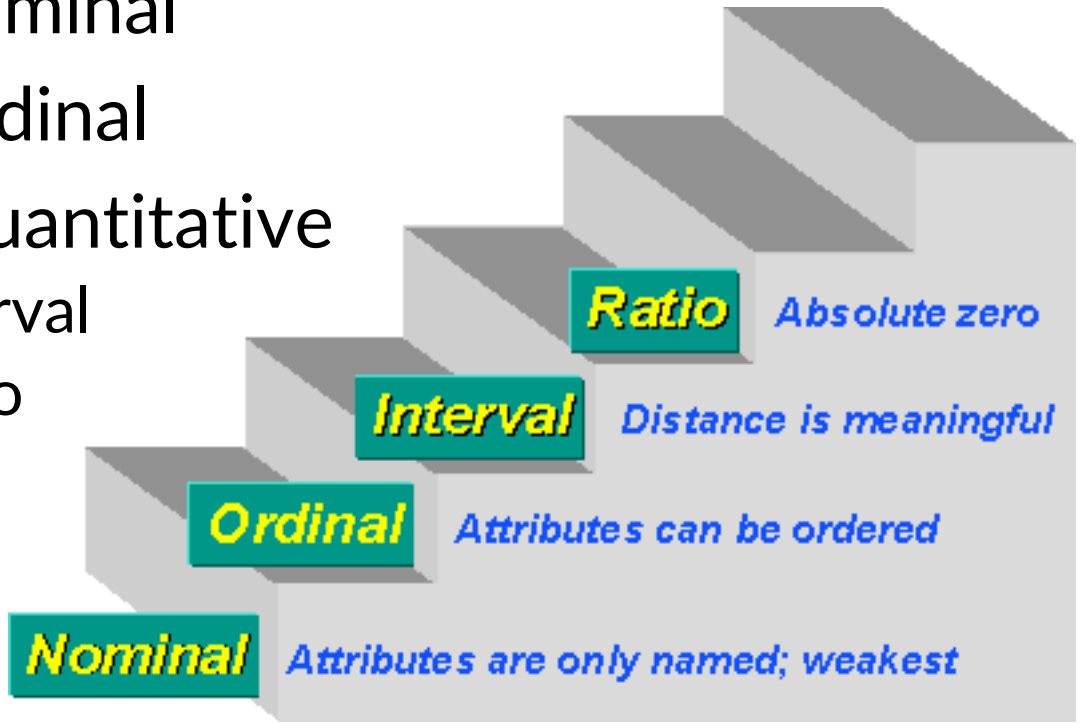
1D (sets and sequences)
Temporal
2D (maps) 3D
(shapes) nD
(relational)
Trees (hierarchies)
Networks (graphs)

and combinations…



Data types: Levels of measurement

- Stevens (1946) classified variables into four levels. These are refererred to as levels of measurement, or levels of data.
- N - Nominal
- O - Ordinal
- Q – Quantitative
  - Interval
  - Ratio



Nominal level measurement

- Merely classifies units into non-ordered categories
  - Distinct groups with no inherent order.
  - Examples: Colors, types of animals, gender.
- No quantitative value
  - Values represent categories, but no mathematical operations are meaningful.
- Examples
  - Male/Female
  - Eye colors
  - Car models

Ordinal level measurement

- Classifies units into ranks or ordered categories
- Examples
  - Ranks: 1st, 2nd, 3rd... place finishers in race
  - Ordered categories: {none = 0, low=1, medium=2, high=3}
- Relative differences
  - Can determine that one value is greater or lesser, but the intervals are not consistent.

Interval level measurement

- Distances (absolute differences) are meaningful
  - any arithmetic operation, such as multiplication, is not
- A fixed difference anywhere on the measurement scale always corresponds to the same difference on the trait being measured
- The zero state of an interval scale is not a true zero value
  - A temperature reading of 0°C does not mean there is no temperature
- Examples
  - Temperature oF
  - A one-degree temperature (°F) difference always means the same thing
  - The absolute difference between 60°F and 61°F is the same as between 100°F and 101°F

Ratio level measurement

- Both differences and ratios are meaningful
- There is a true zero
  - A zero on a ratio scale means there is a total absence of the variable you are measuring.
- Examples
  - Length, area, and population
  - The **relative difference** between a 10 and a 20-year-old is the same as the difference between a 40- and an 80-year-old ('twice as old').
  - Celsius and Fahrenheit are interval scales, Kelvin is a ratio scale
    - Kelvin scale has a true zero (0 K) where nothing can be colder.
    - Zero is just another temperature value in Celsius and Fahrenheit

Significance of Levels of Measurement

- Guiding data analysis and visualization choices
  - Determines statistical analyses
    - Tailor statistical methods based on the level of measurement.
    - Example: Nominal data requires different analyses than ratio data.
  - Informs visualization choices
    - Select visualizations aligned with the nature of your data.
    - Example: Use bar charts for ordinal data, histograms for interval data.
  - Aids in data interpretation
    - Understanding measurement level guides meaningful conclusions.
    - Example: Recognize ratio data for accurate calculations and interpretations.

Properties of images

Image Models

Visual language is a sign system

Images perceived as a set of signs

Sender encodes information in signs

Receiver decodes information from signs

Semiology of Graphics, 1967

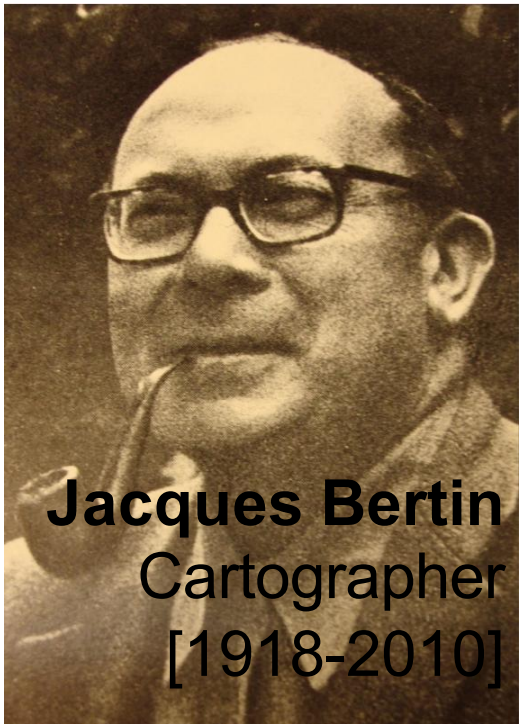


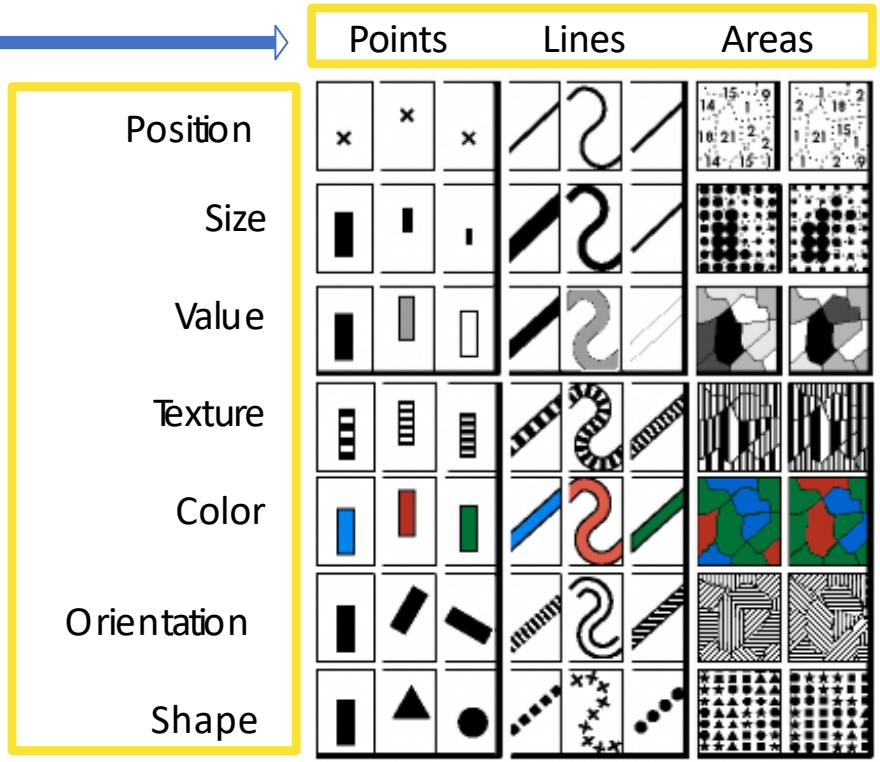
Image models

Visual Marks (Đối tượng biểu diễn)

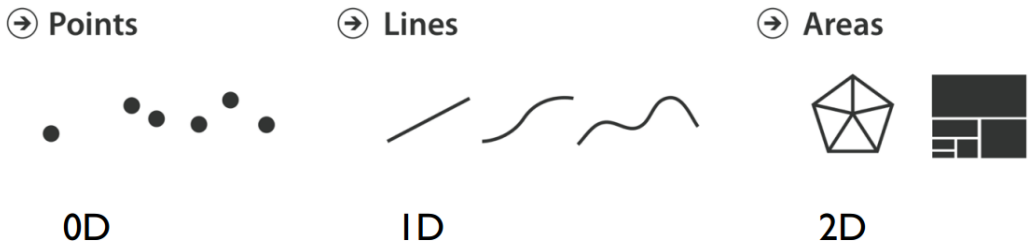
- Basic graphical elements in an image
- **Represent information (items or links)**

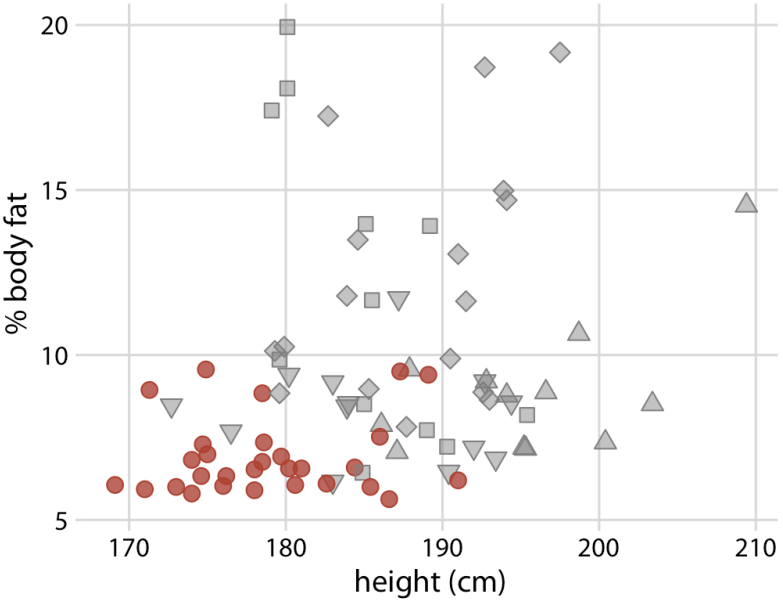
Perceptual Channels (Biến thị giác)

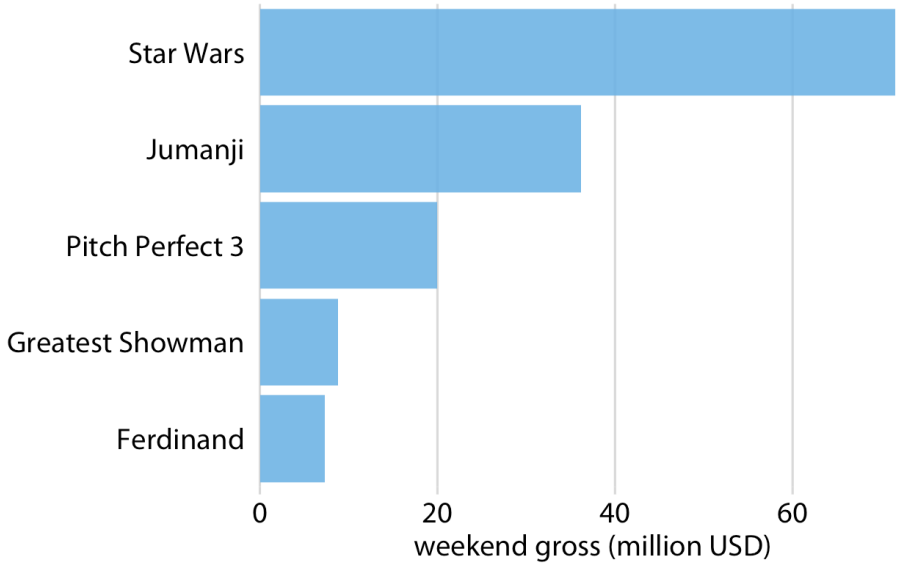
- Control the appearance of marks
  - Change appearance based on attribute value
- Encode information
- **Channel = Visual Variables**



Visual marks to represent items

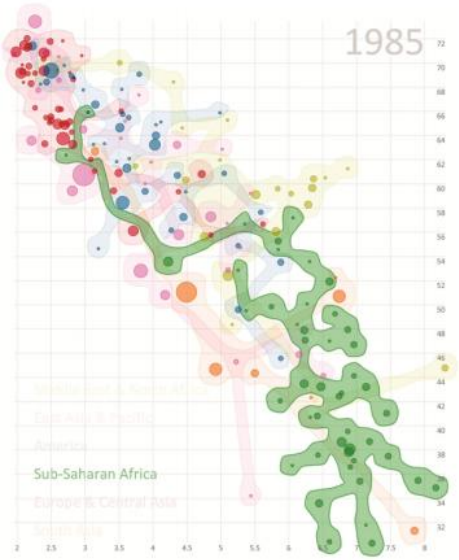




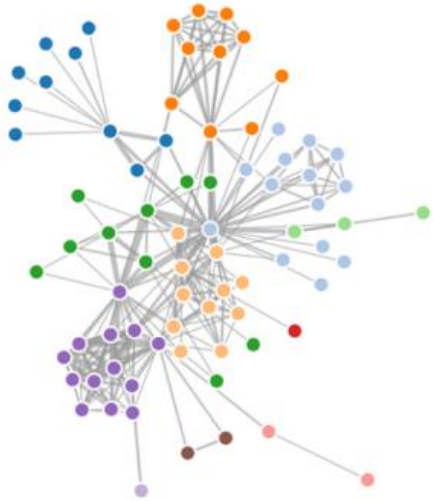


Visual marks to represent links

Containment

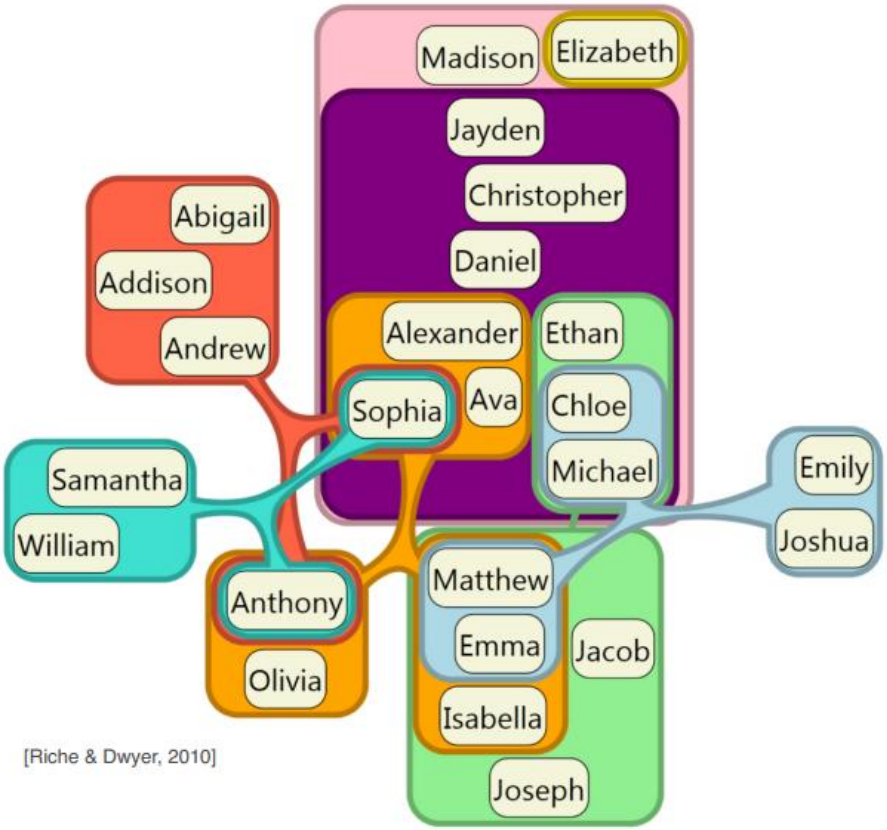


Connection



Visual marks to represent links (2)

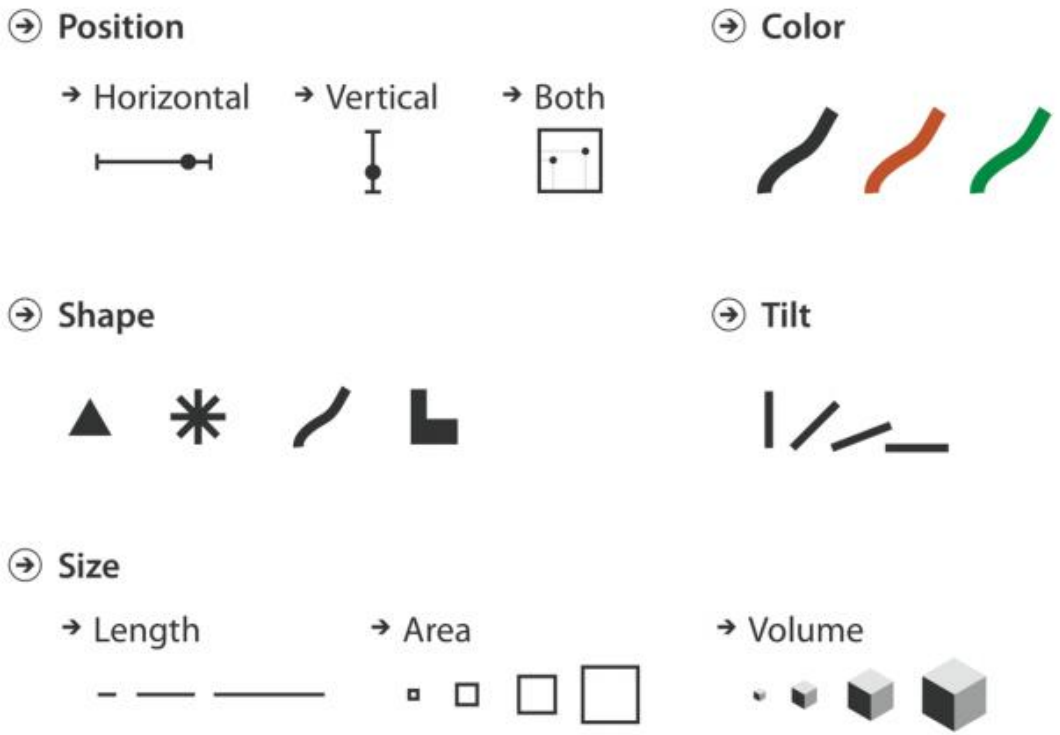
Nested containment



[Riche & Dwyer, 2010]

Perceptual channels

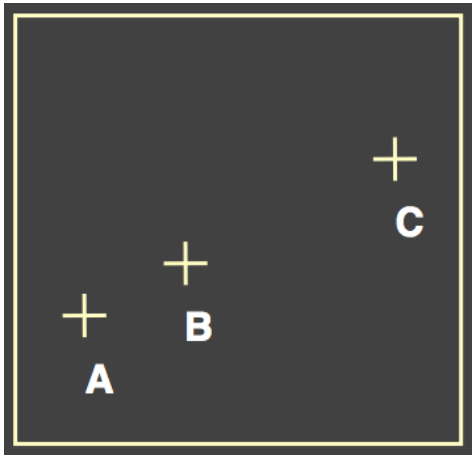
- Control the appearance of marks
- Encode information



Perceptual channels: Position

- Can encode quantitative variables (Q)

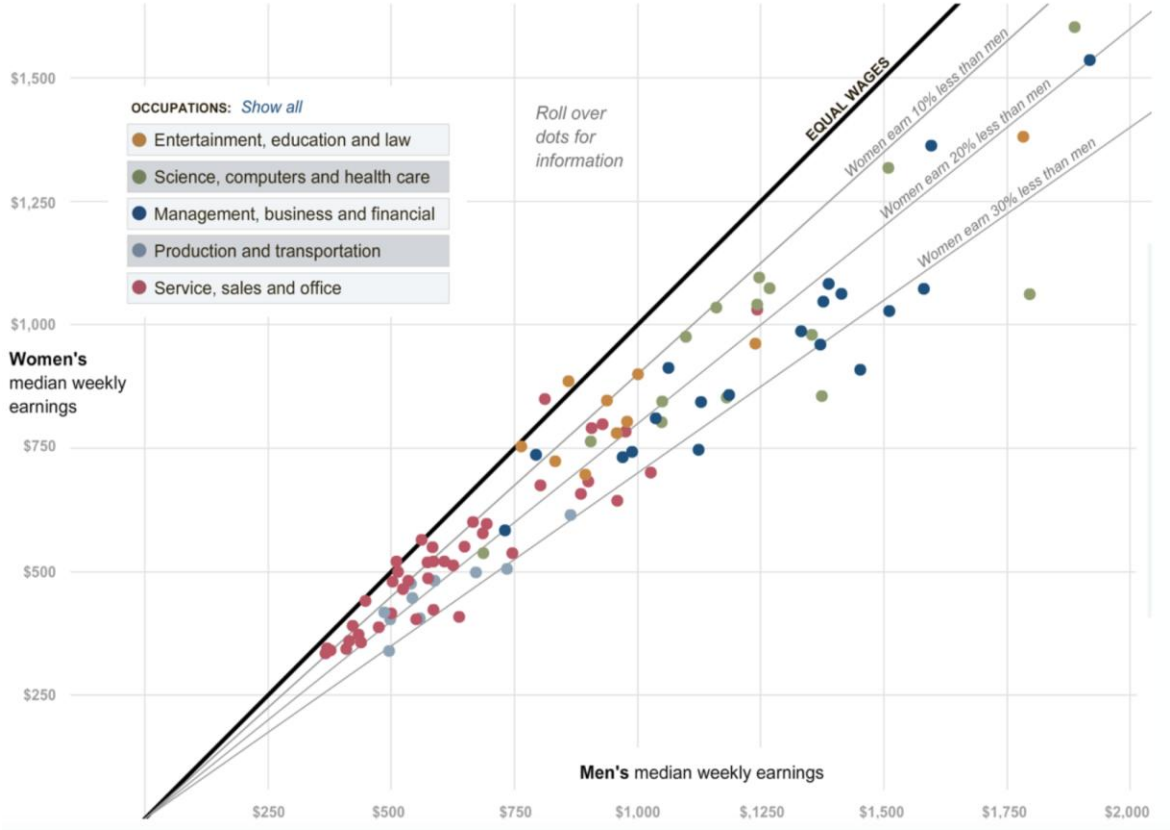
1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.



"Resemblance, order and proportional are the three signfields in graphics." — Bertin

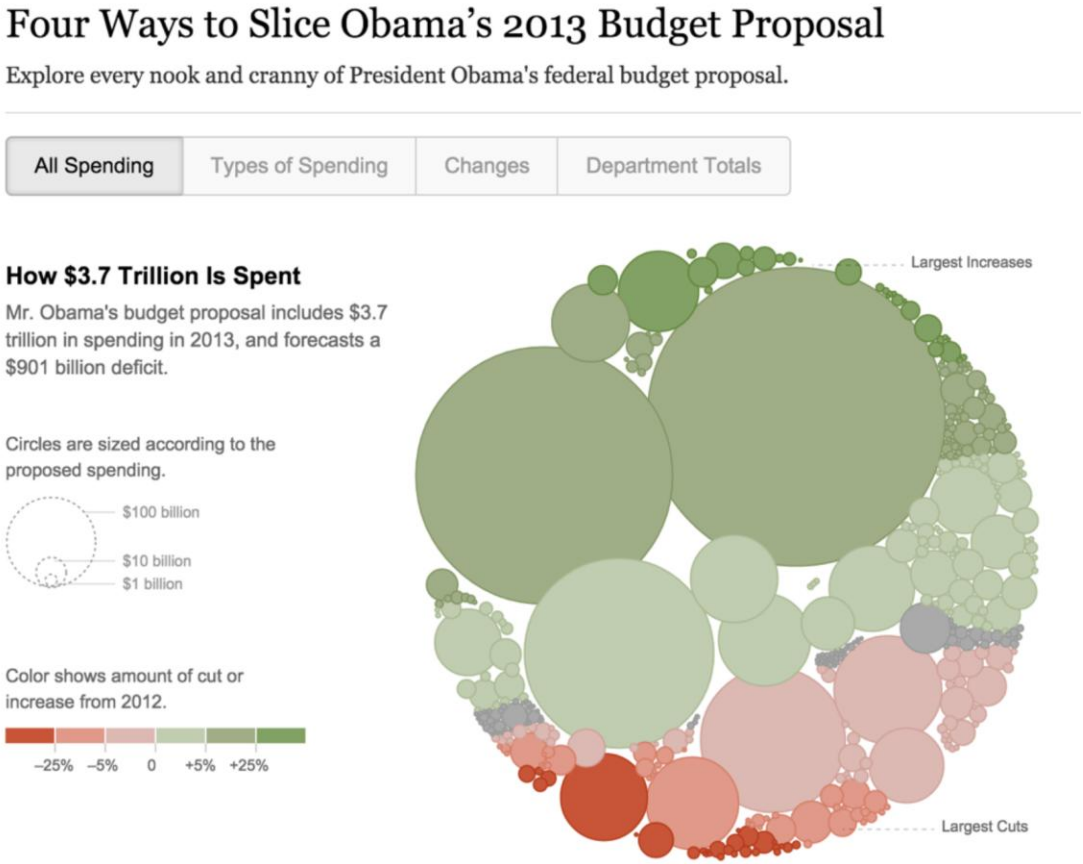
Perceptual channels: Position

- The most used perceptual channel
- Suitable for most data types



Perceptual channels: Size

- Length, area, volume
  - Good for 1D, 2D
  - Easy to compare



Encoding Information in Color and Value

- Value (lightness) is perceived as ordered
  - Encode ordinal variables (O) [better]
  - Encode continuous variables (Q)

- Hue is normally perceived as unordered
  - Encode nominal variables (N)

Perceptual channels: Color

- Should limit the number of colors

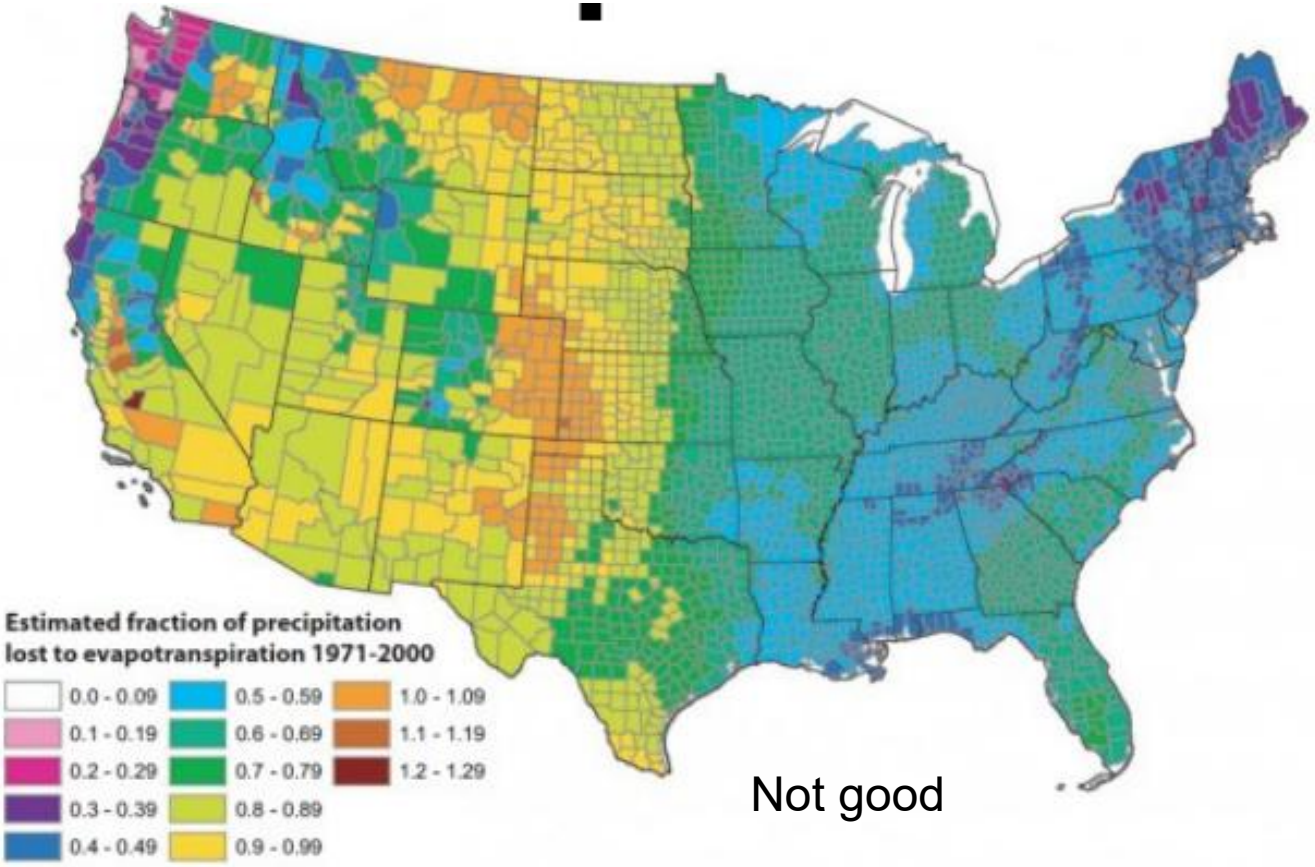


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

Cliff Mass

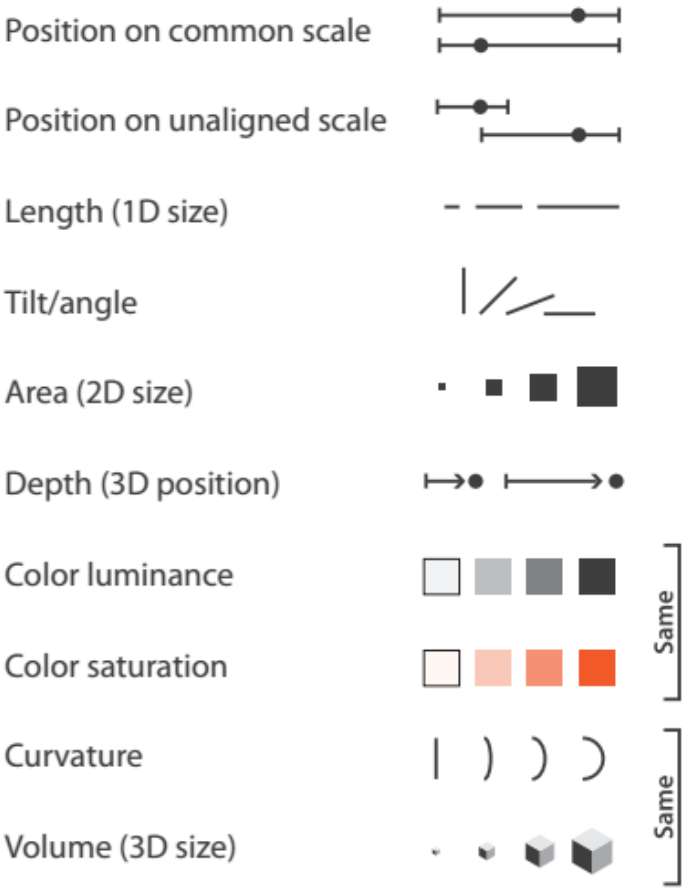
Perceptual channels: Shape

- Encode nominal variables (N)
- No ordered

Encoding information in perceptual channels

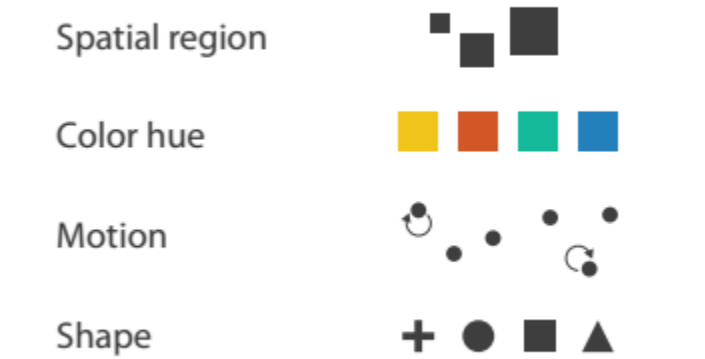
Quantitative/Ordered

Magnitude Channels: Ordered Attributes

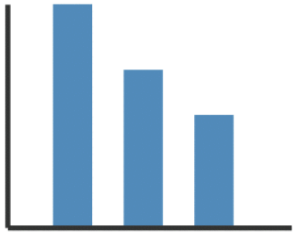


Nominal

Identity Channels: Categorical Attributes



Using marks and channels



Mark: Line

Channel: Length/Position

1 quantitative attribute

1 categorical attribute

Mark: Point

Channel: Position

2 quantitative attr.

Adding Hue

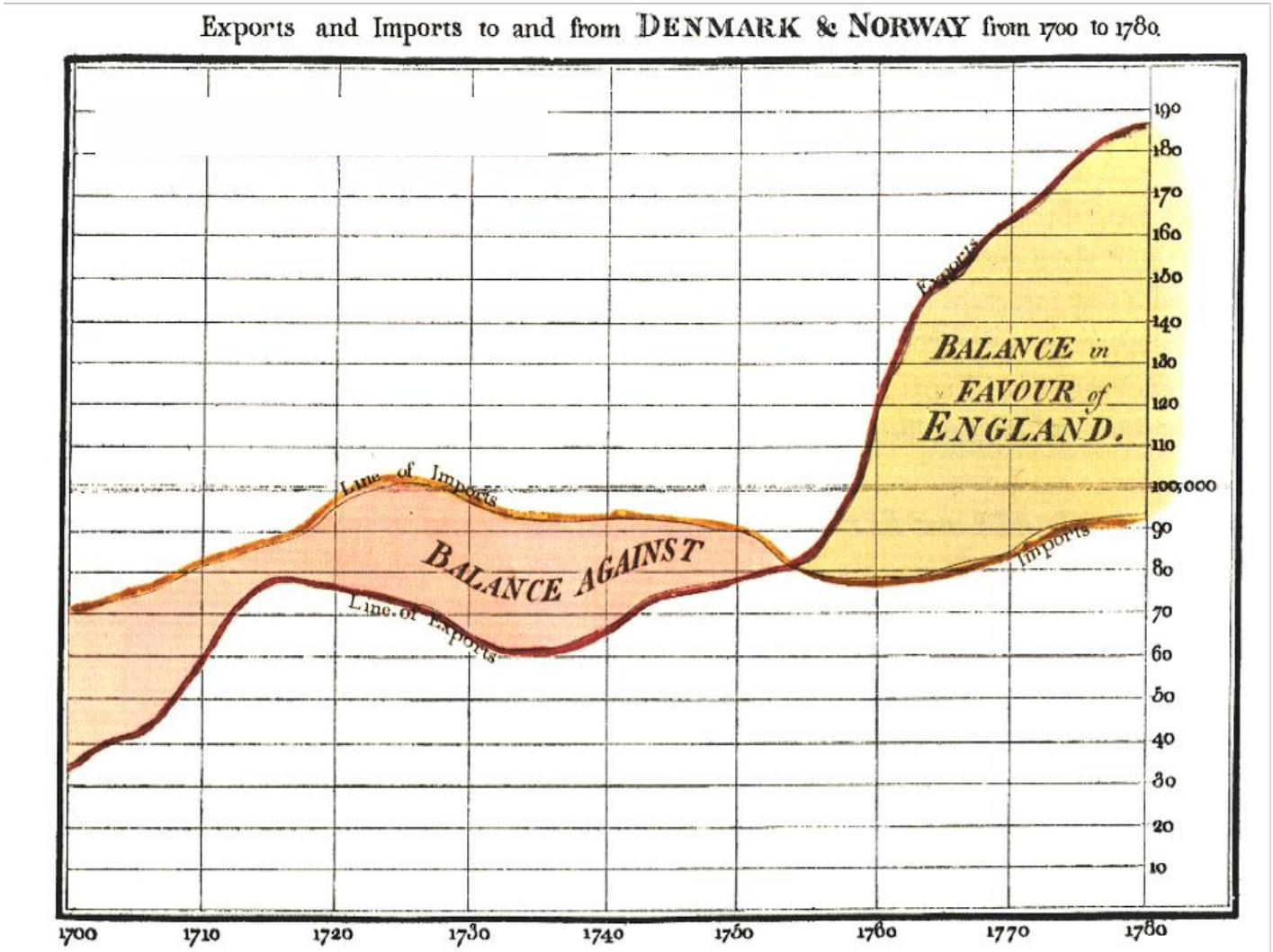
+1 categorical attr.

Adding Size

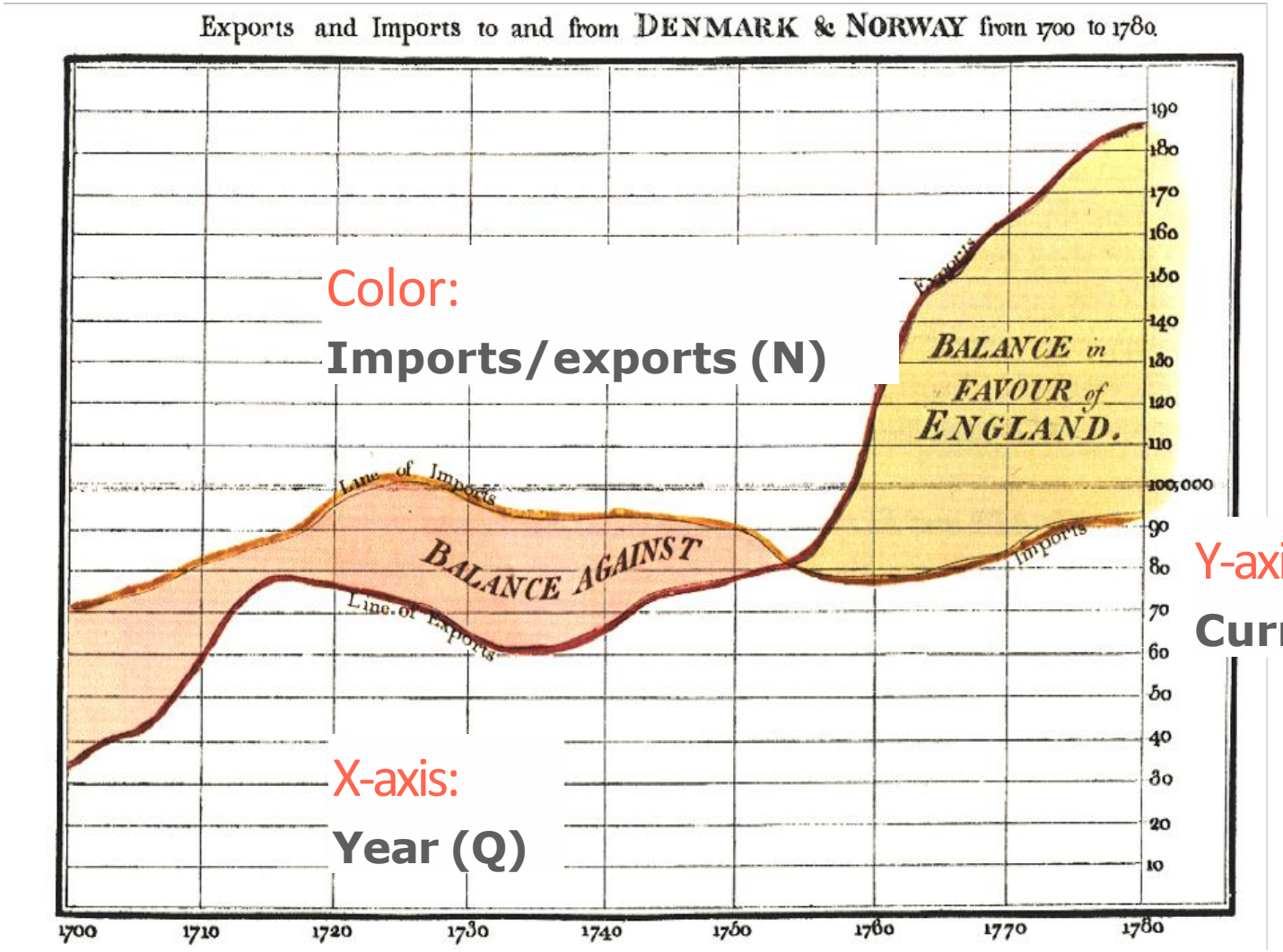
+1 quantitative attr.

Example: Deconstructions

William Playfair, 1786



William Playfair, 1786



Color:

Imports/exports (N)

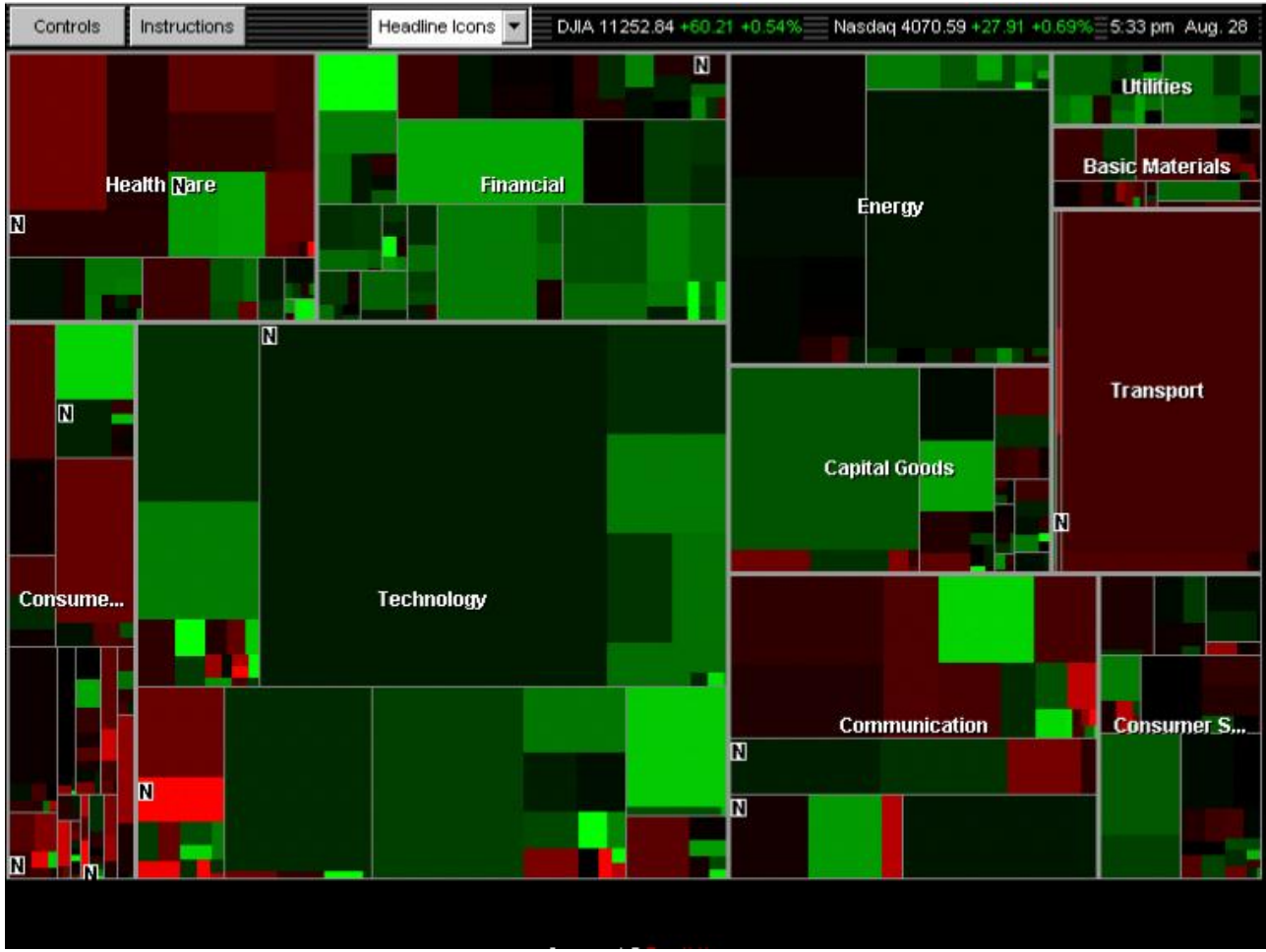
X-axis:

Year (Q)

Y-axis:

Currency (Q)

Wattenberg's Map of the Market



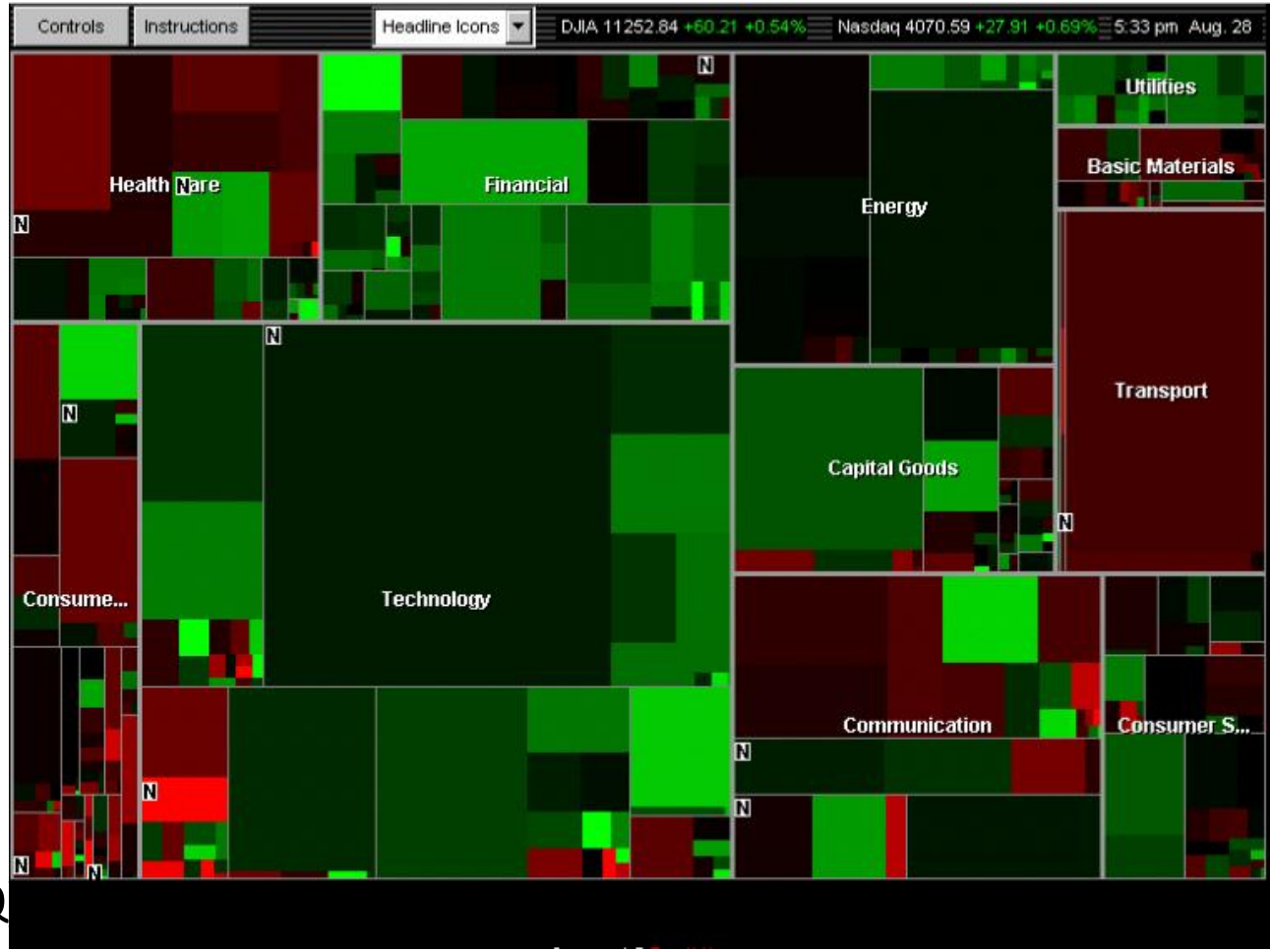
Wattenberg's Map of the Market

Rectangle Area: market cap (Q)

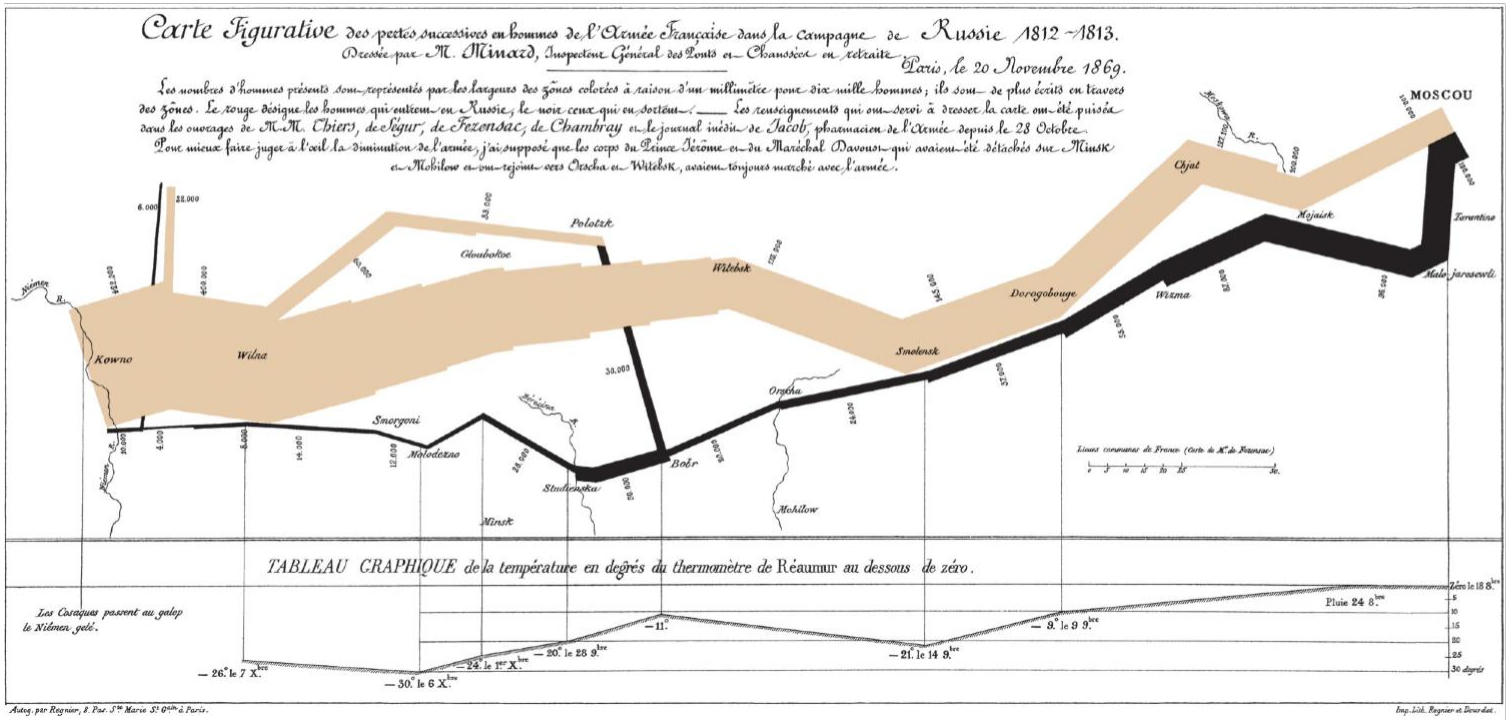
Rectangle Position: market sector (N), market cap (Q)

Color Hue: loss vs. gain (N)

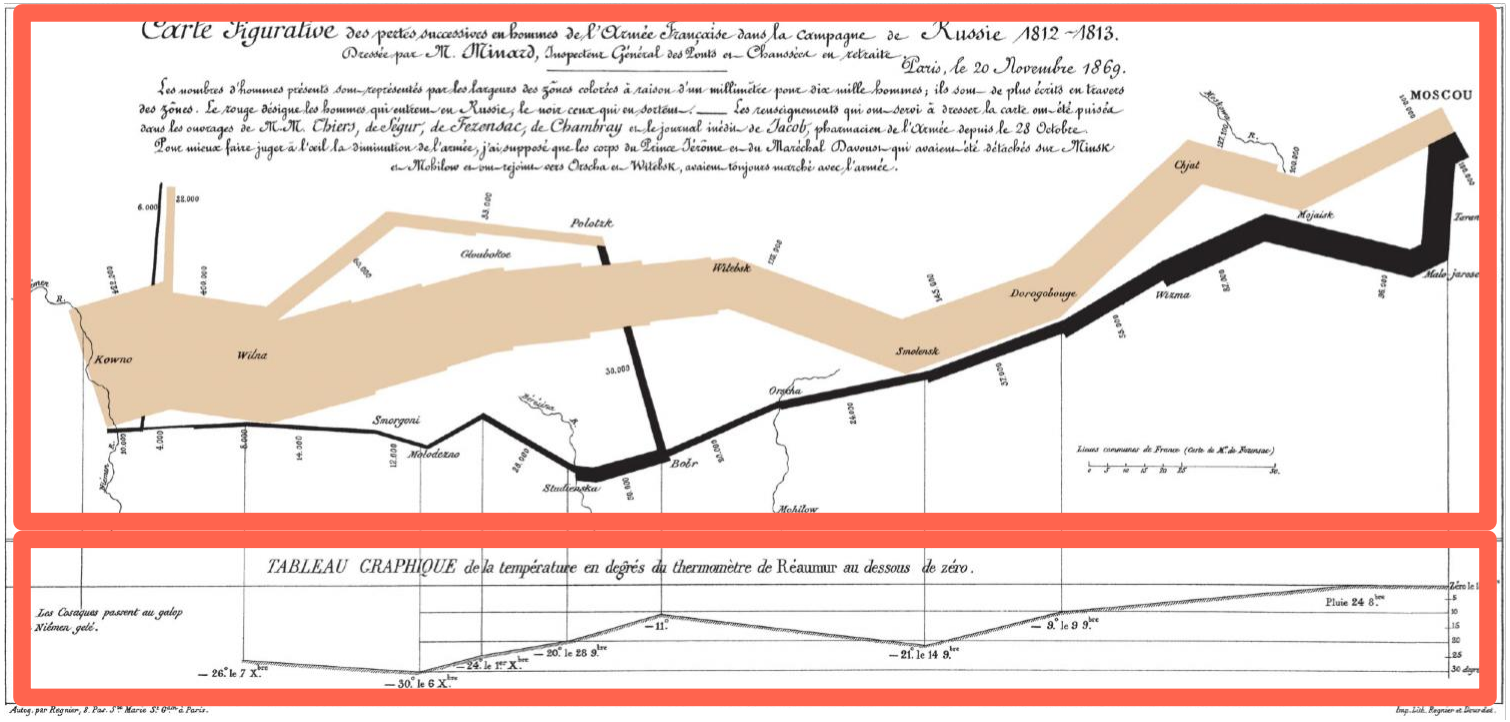
Color Value: magnitude of loss or gain (Q)



Minard 1869: Napoleon's March



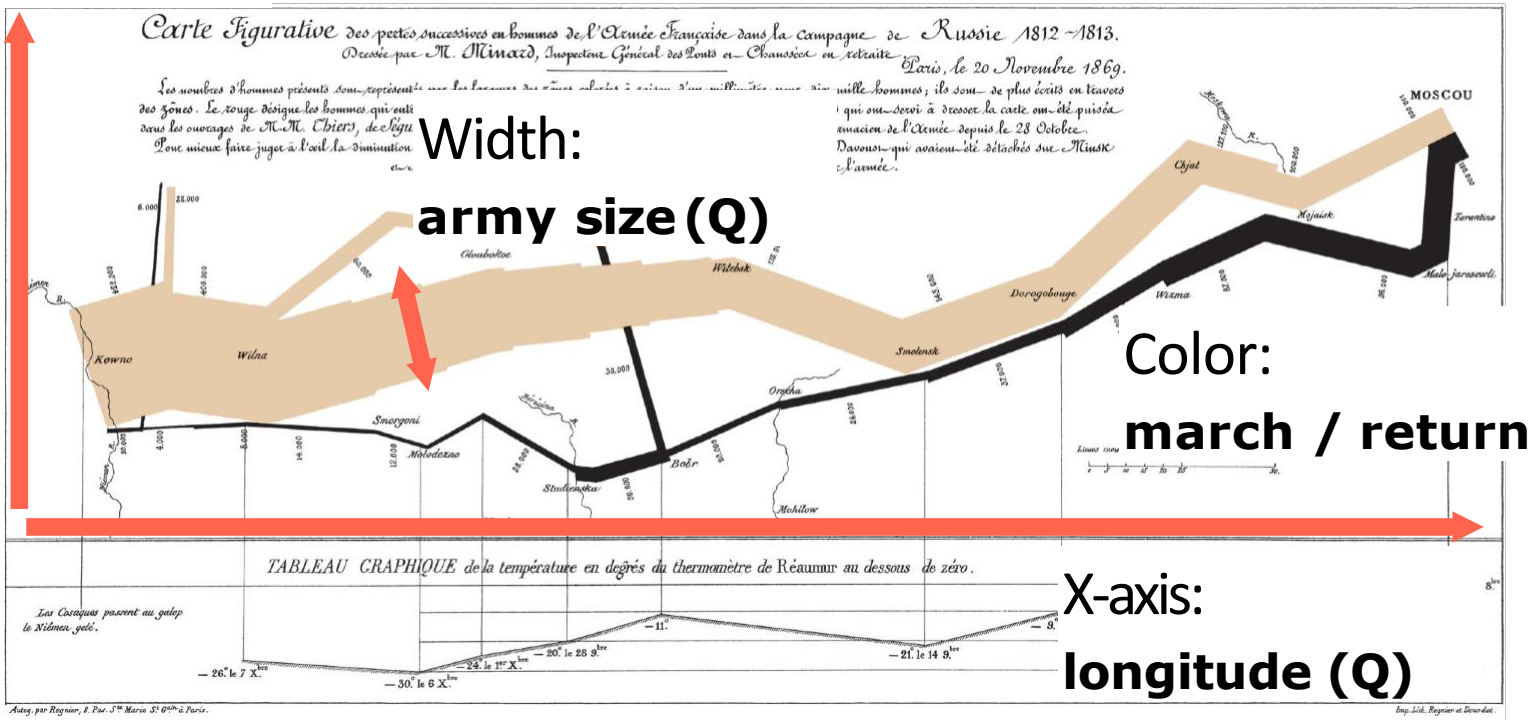
Minard 1869: Napoleon's March



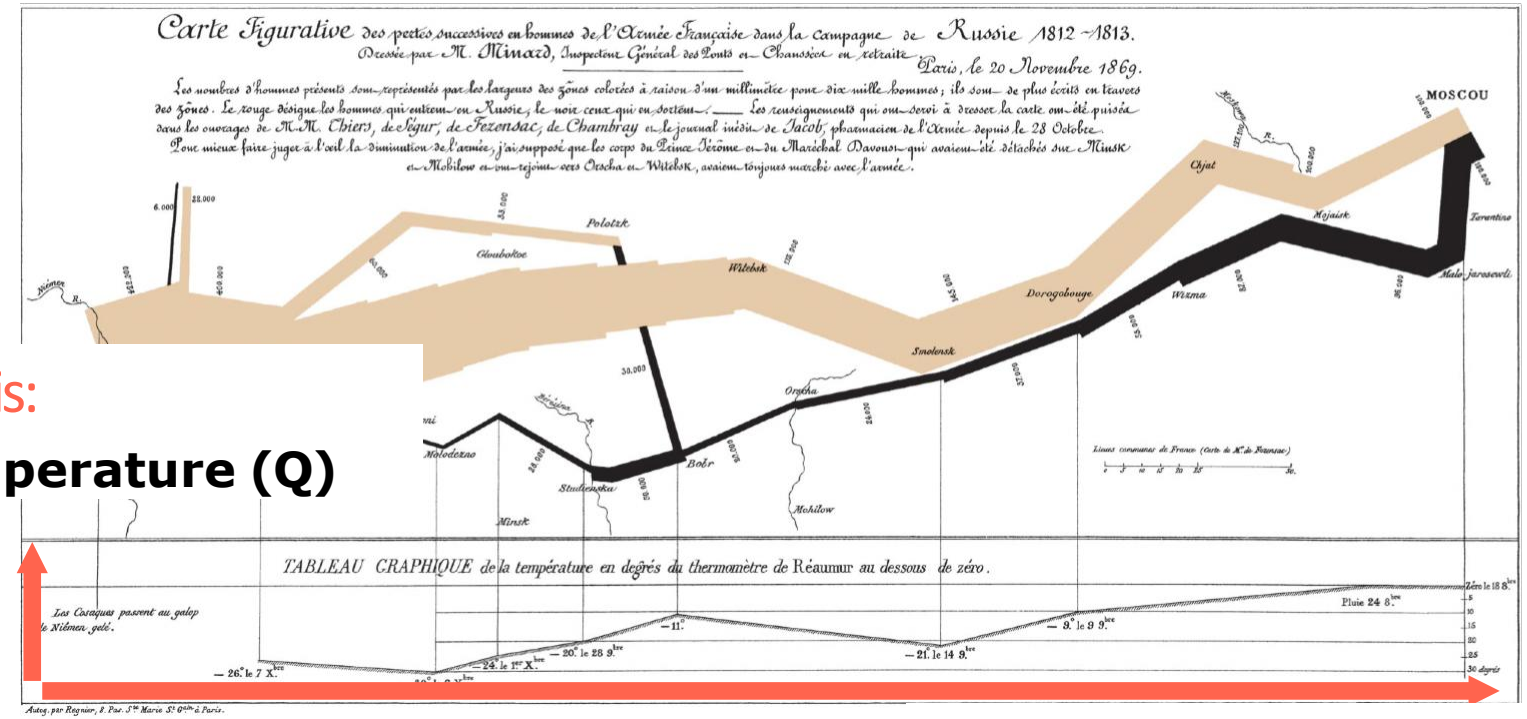
Minard 1869: Napoleon's March

Y-axis:

latitude (Q)



Minard 1869: Napoleon's March



Y-axis:

temperature (Q)

X-axis:

longitude (Q) / time (O)

Bertin's Levels of Organization

Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	o	
Color	N		
Orientation	N		
Shape	N		

Nominal

Ordinal

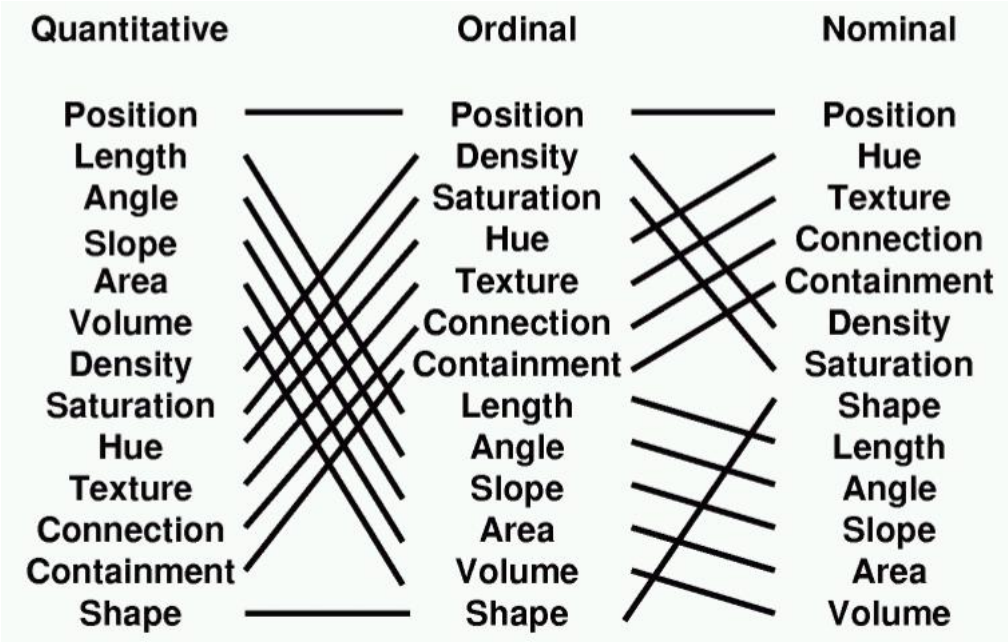
Quantitative

Note: **Q** ⊂**O** ⊂**N**

Mackinlay's Ranking

- Expanded Bertin's variables and conjectured effectiveness of encodings by data type.

[Mackinlay 86]

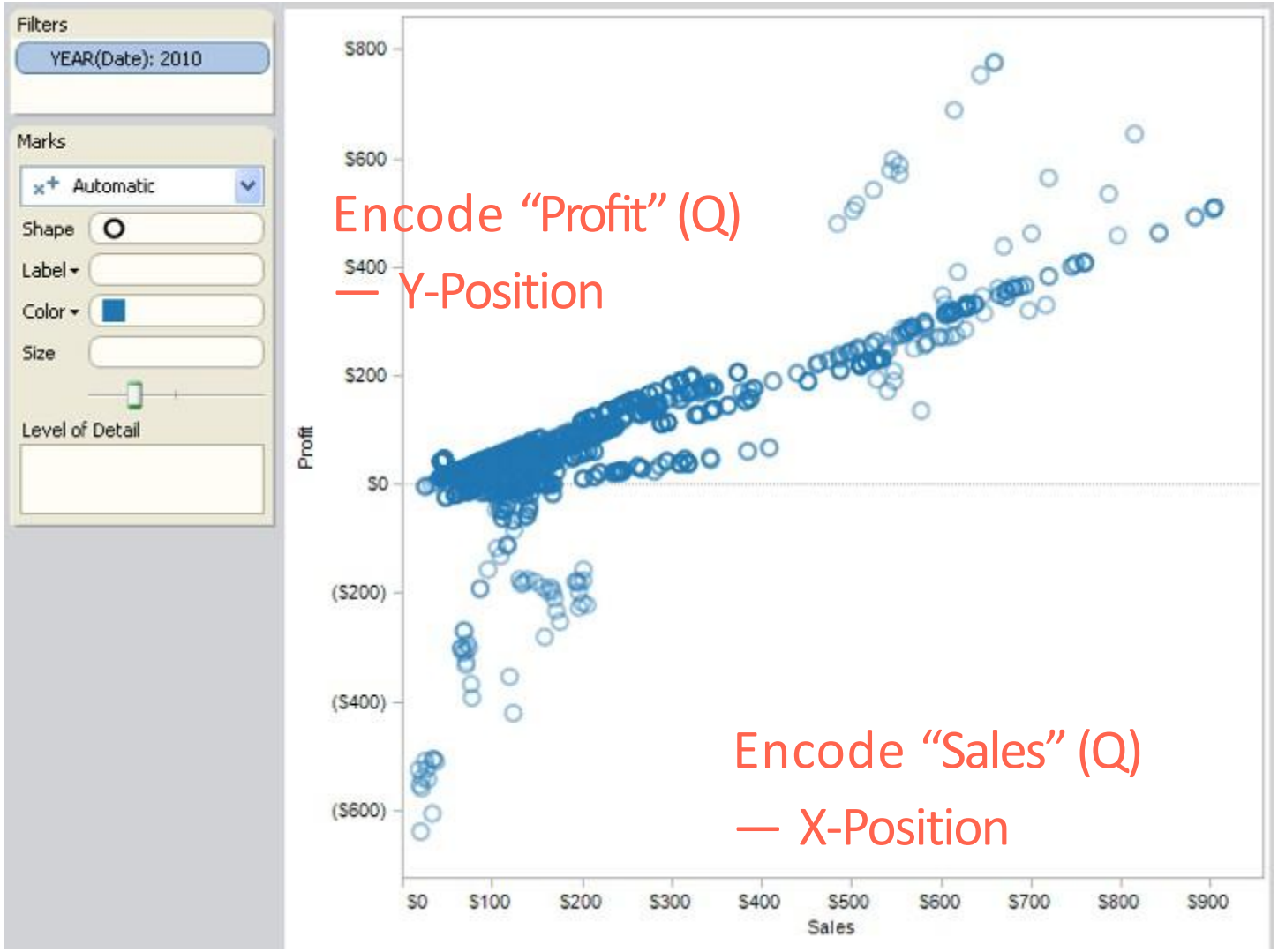


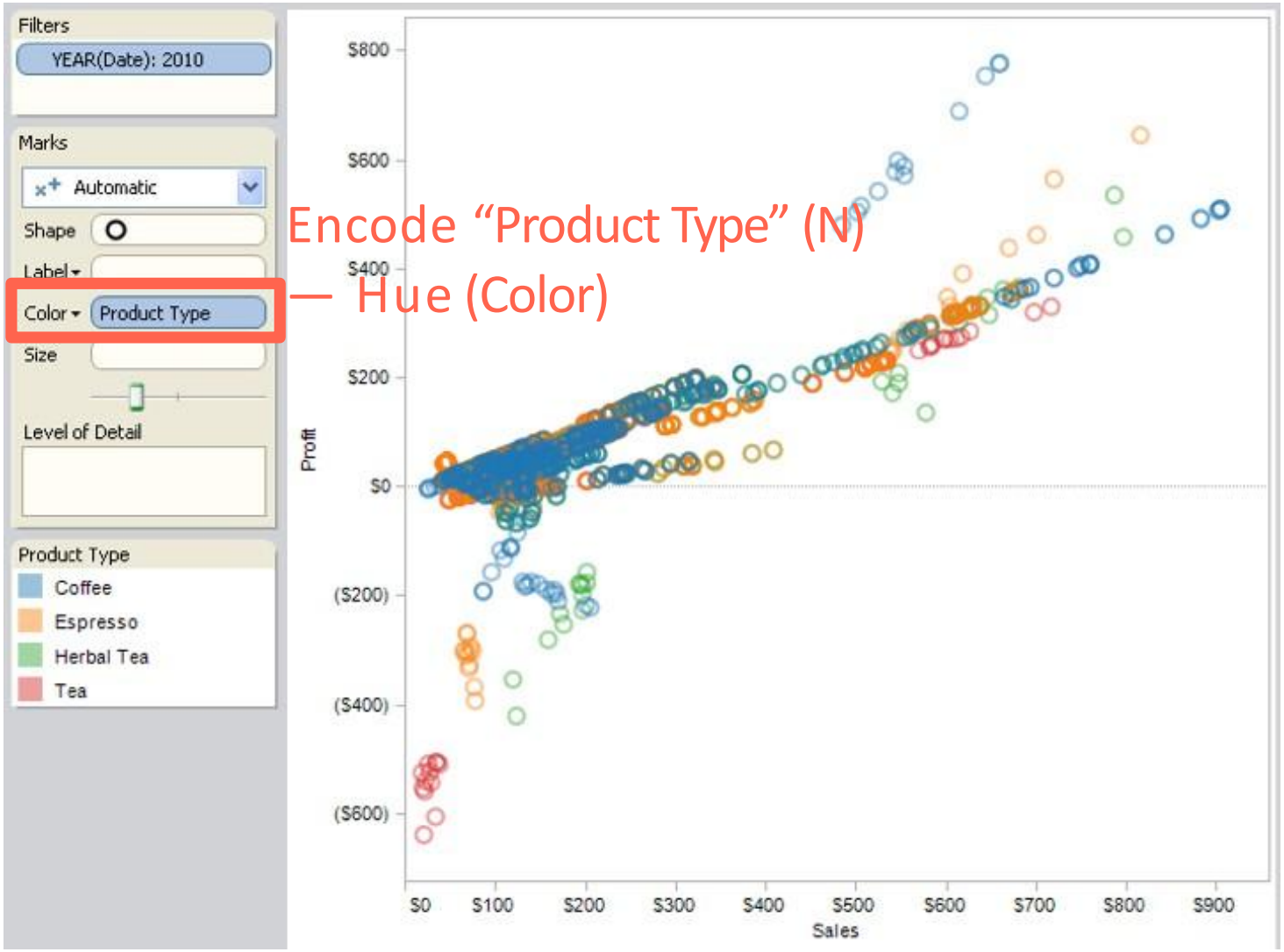
Jock D. Mackinlay
Vice President
Tableau Software

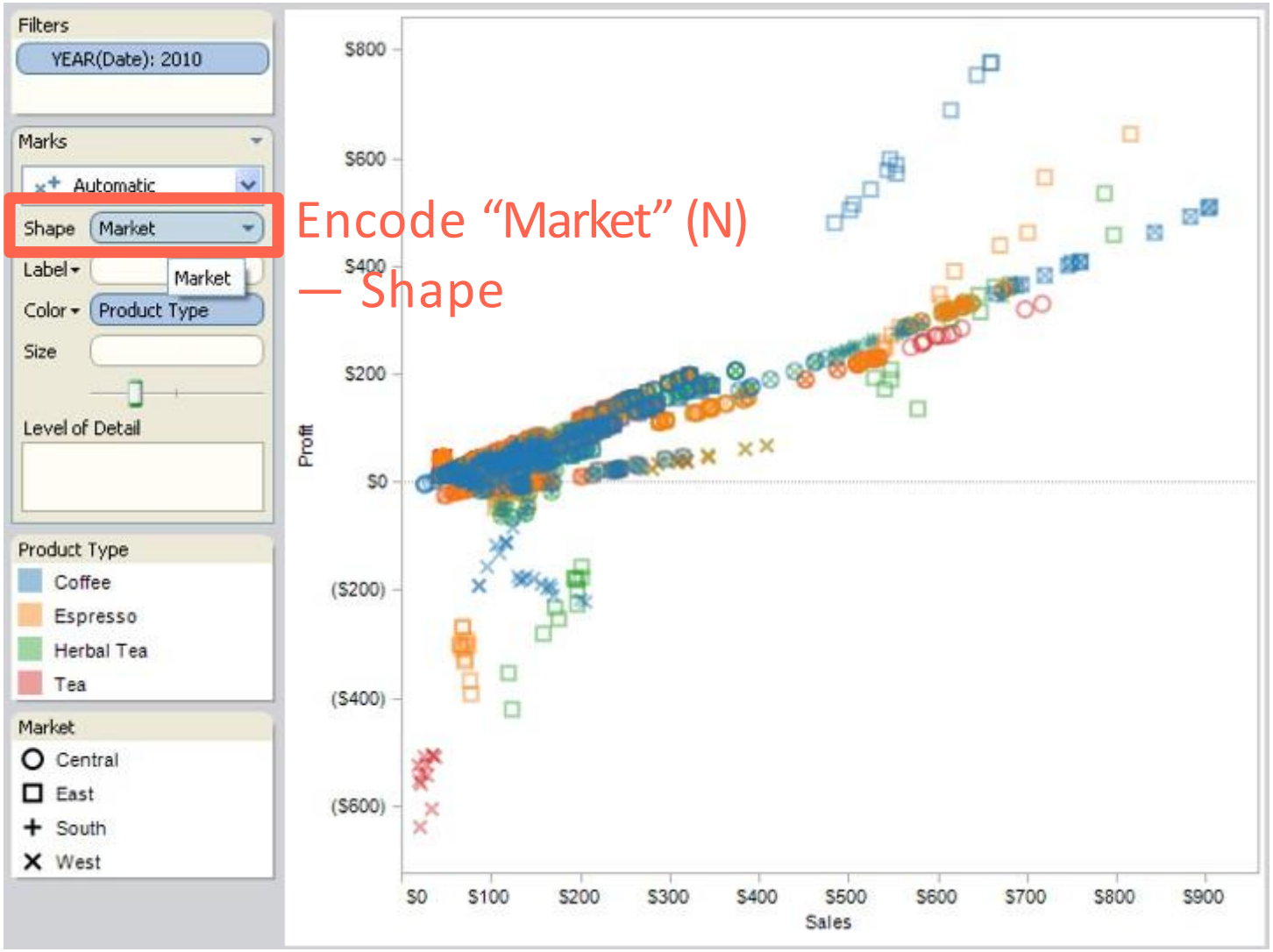
Example: Encoding Data

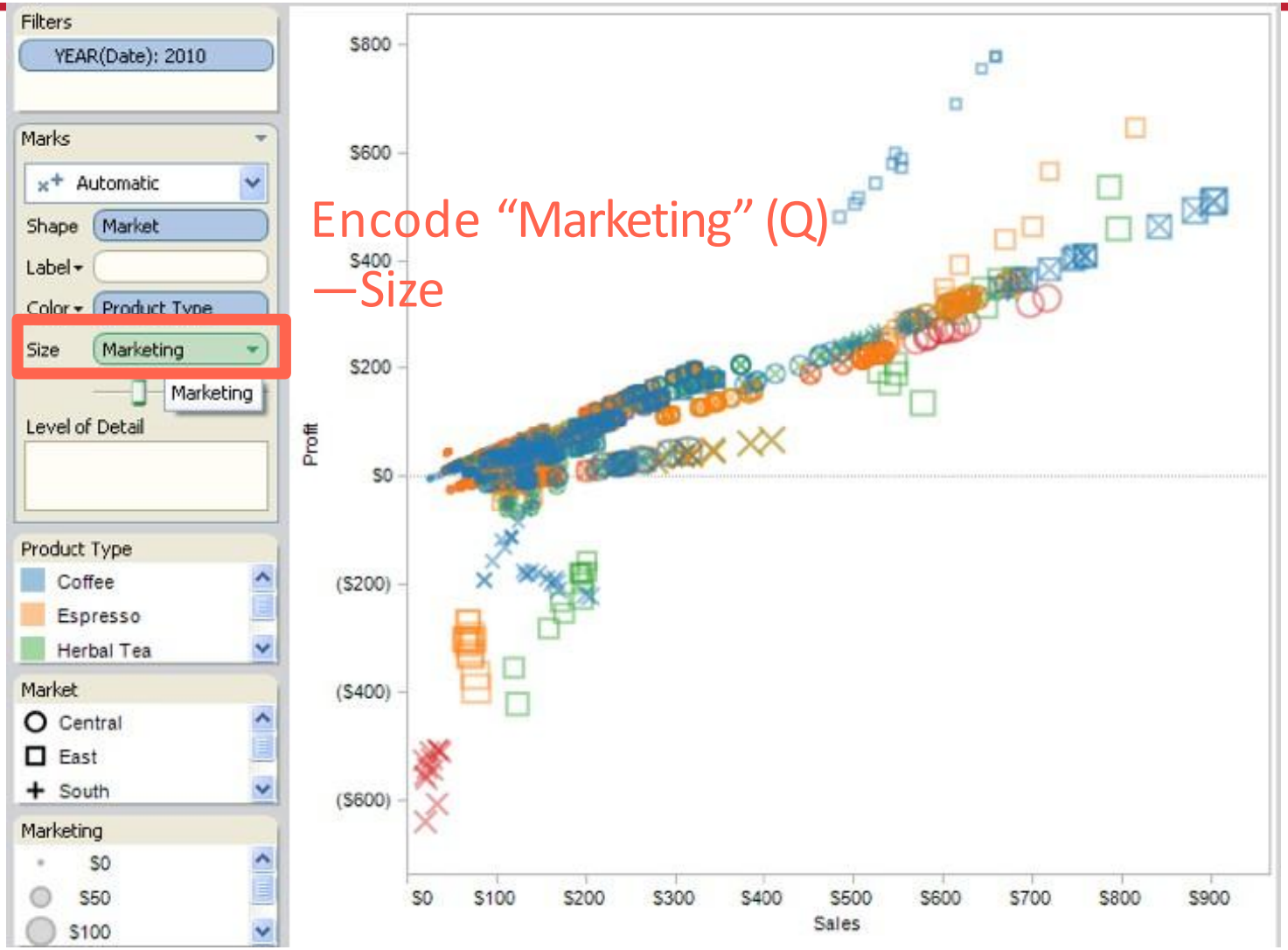
Example: Coffee Sales

- Sales figures for a fictional coffee chain
  - Sales Q-Ratio
  - Profit Q-Ratio
  - Marketing Q-Ratio
  - Product Type N {Coffee, Espresso, Herbal Tea, Tea}
  - Market N {Central, East, South, West}


Filters
YEAR(Date): 2010
Marks
Automatic
Shape
Label
Color
Size
Level of Detail
Encode "Profit" (Q)
— Y-Position
Encode "Sales" (Q)
— X-Position
Profit
$800
$600
$400
$200
$0
($200)
($400)
($600)
$0
$100
$200
$300
$400
$500
$600
$700
$800
$900
Sales

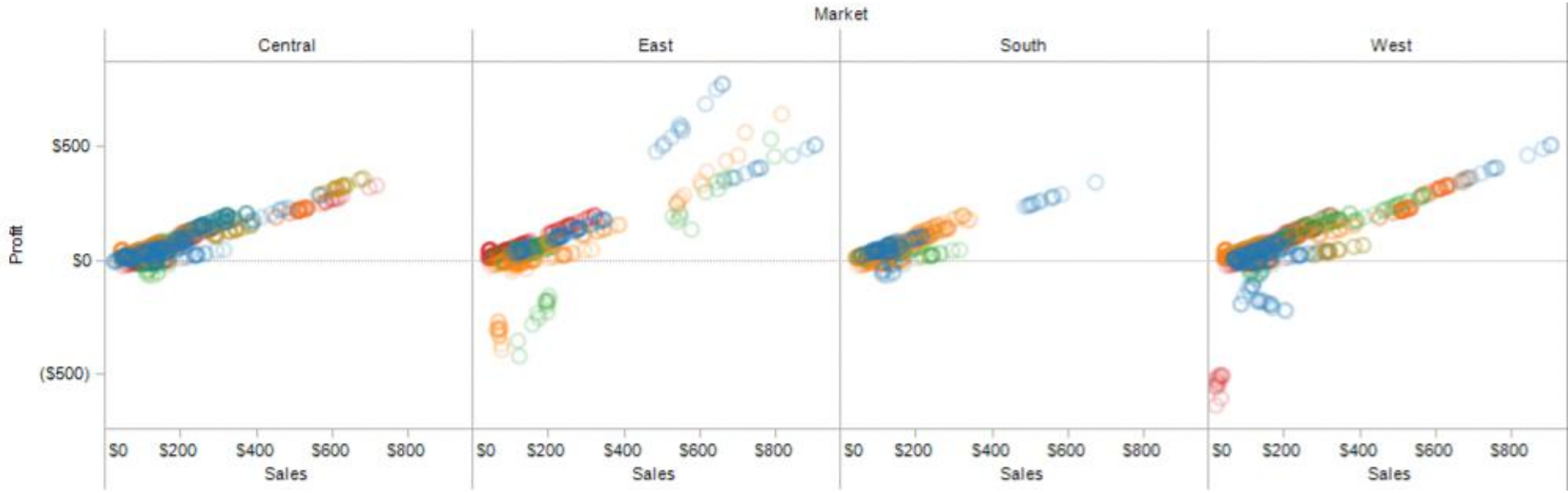

Filters
YEAR(Date): 2010
Marks
Automatic
Shape
Label
Color
Product Type
Size
Level of Detail
Product Type
Coffee
Espresso
Herbal Tea
Tea
Encode “Product Type” (N)
— Hue (Color)
$800
$600
$400
$200
$0
($200)
($400)
($600)
Profit
$0
$100
$200
$300
$400
$500
$600
$700
$800
$900
Sales


Filters
YEAR(Date): 2010
Marks
Automatic
Shape
Market
Label
Market
Color
Product Type
Size
Level of Detail
Product Type
Coffee
Espresso
Herbal Tea
Tea
Market
Central
East
South
West
Encode “Market” (N)
— Shape
$800
$600
$400
$200
$0
($200)
($400)
($600)
Profit
$0
$100
$200
$300
$400
$500
$600
$700
$800
$900
Sales


Filters
YEAR(Date): 2010
Marks
Automatic
Shape
Market
Label
Color
Product Type
Size
Marketing
Marketing
Level of Detail
Product Type
Coffee
Espresso
Herbal Tea
Market
Central
East
South
Marketing
$0
$50
$100
Encode “Marketing” (Q)
—Size
$800
$600
$400
$200
$0
($200)
($400)
($600)
Profit
$0
$100
$200
$300
$400
$500
$600
$700
$800
$900
Sales

Avoid over-encoding

- Use trellis plots (small multiples/facets) that subdivide space to enable comparison across multiple plots.



Visual encoding

Choosing visual encodings

- Assume ***k*** visual channels and ***n*** data attributes. We would like to pick the “best” encoding among a combinatorial set of possibilities of size $\mathbf{n^k}$

Choosing visual encodings

- Assume ***k*** visual encodings and ***n*** data attributes. We would like to pick the "best" encoding among a combinatorial set of possibilities of size $\mathbf{n^k}$

- **Principle of Consistency**
  - The properties of the image (visual variables) should match the properties of the data.

- **Principle of Importance Ordering**
  - Encode the most important information in the most effective way.

Design Criteria [Mackinlay 86]

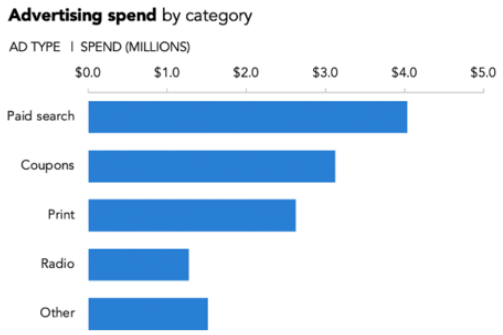
- Expressiveness
- Effectiveness

Design Criteria [Mackinlay 86]

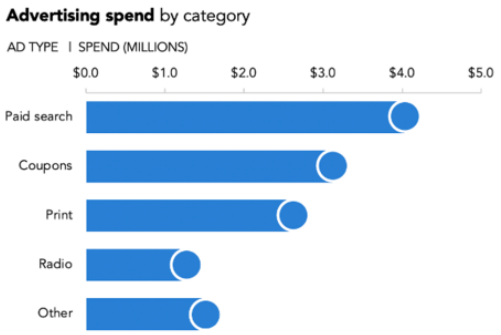
- Expressiveness
  - A set of facts is expressible in a visualization if it expresses all the facts and only the facts in the data.
  - Tell the truth and nothing but the truth
    - (don’t lie, and don’t lie by omission)

Dot plot

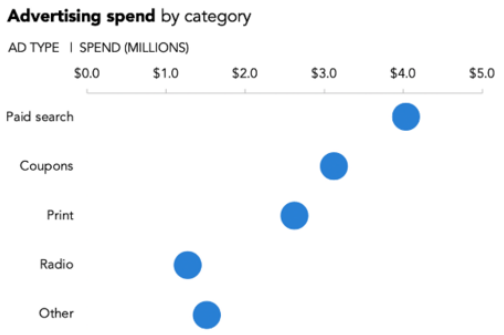
A typical horizontal bar chart…



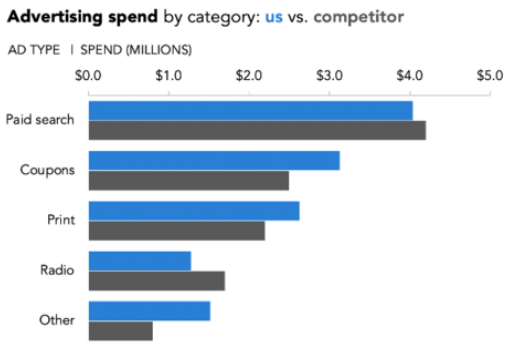
Encode end points with dots.



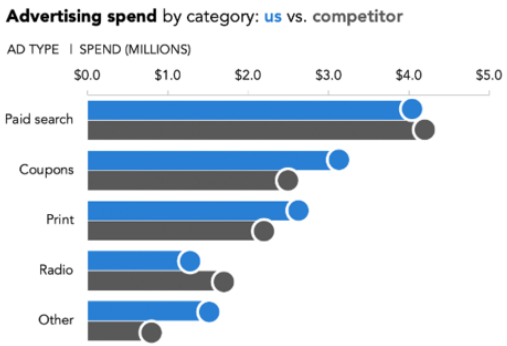
Remove the bars. Voila! A dot plot.



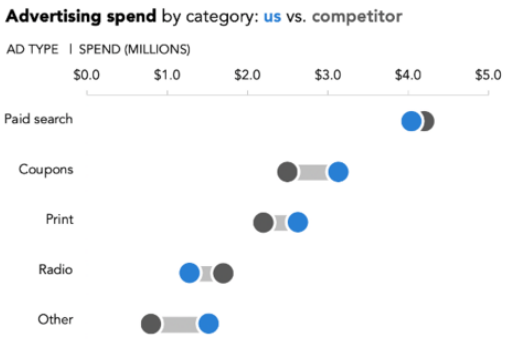
A dual-series bar chart…



Encode end points with dots.

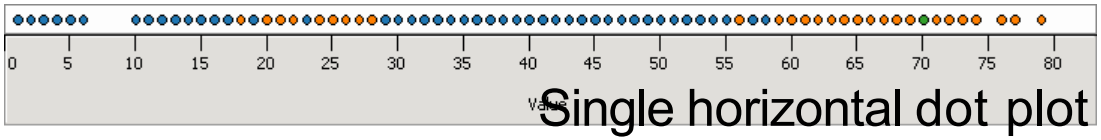


Collapse, remove bars, connect.



Can not express the facts

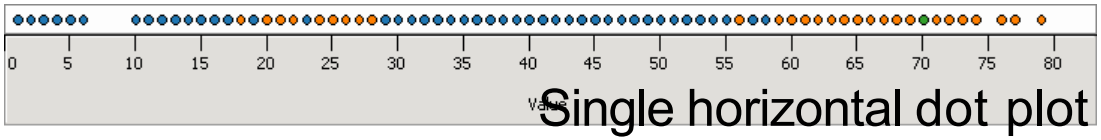
The relationship among multiple data attributes may not be expressed in a single horizontal dot plot.



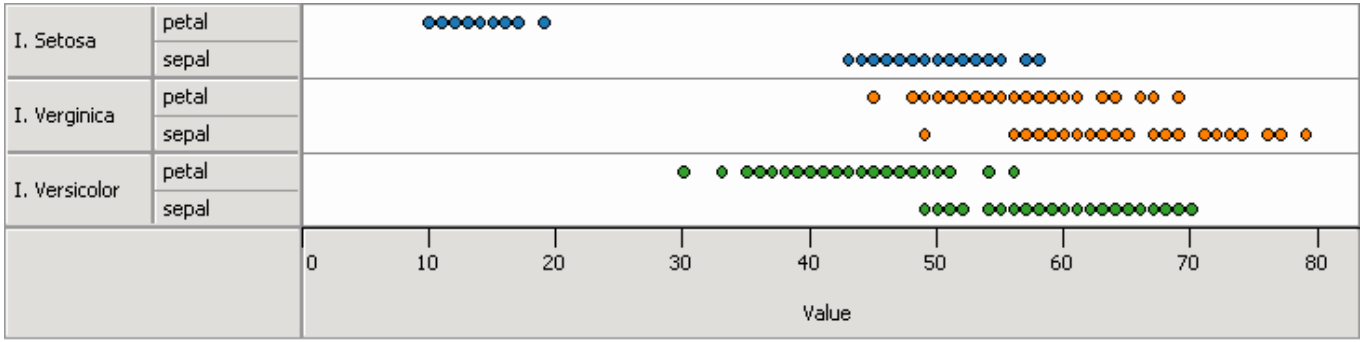
Single horizontal dot plot

Can not express the facts

The relationship among multiple data attributes may not be expressed in a single horizontal dot plot.



Single horizontal dot plot



Categories in different positions

Expresses facts not in the data

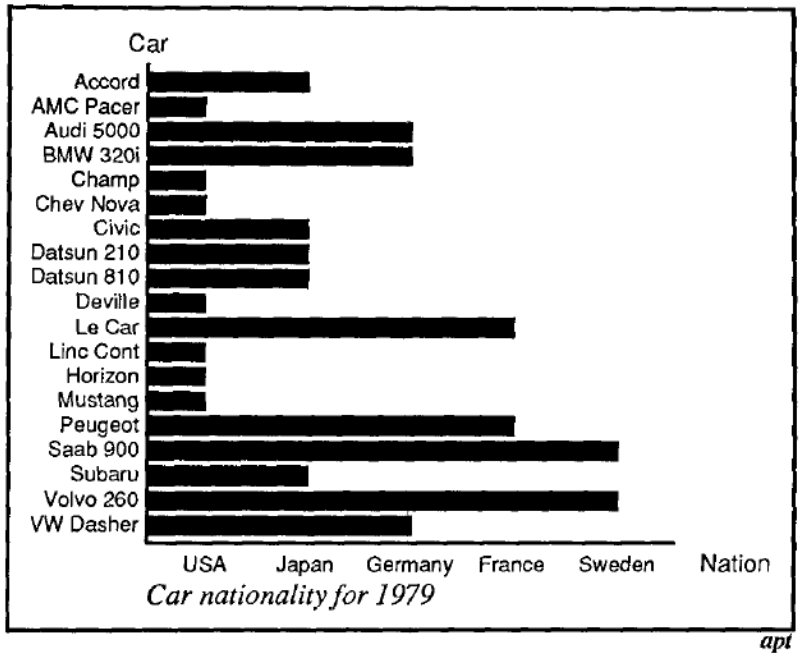


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

A length is interpreted as a quantitative value.

Design Criteria [Mackinlay 86]

- Expressiveness
  - A set of facts is expressible in a visualization if it expresses all the facts and only the facts in the data.
  - **Tell the truth and nothing but the truth**
    - **(don't lie, and don't lie by omission)**
- Effectiveness
  - A visualization is more effective than another one if the information conveyed is more readily perceived.
  - **Use encodings that people decode better**
    - **(where better = faster and/or more accurate)**

Mackinlay's Design Algorithm

- APT - "A Presentation Tool", 1986

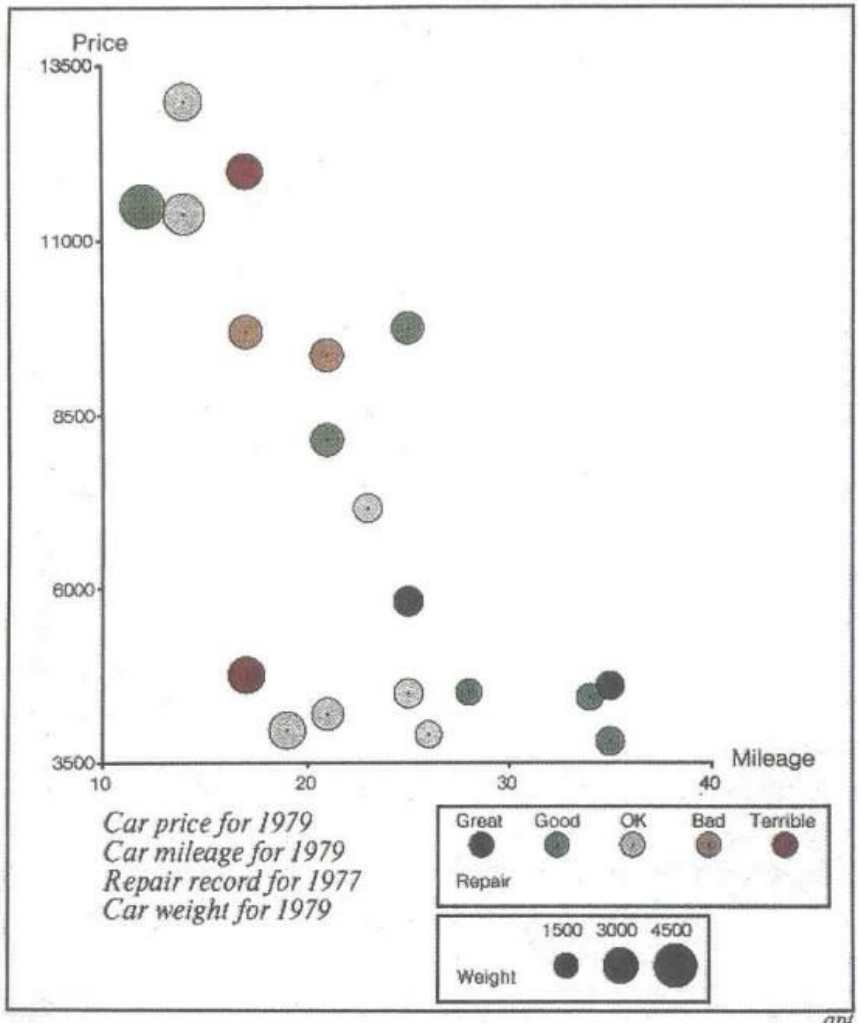
- User formally specifies data model and type
- Input: ordered list of data variables to show

- APT searches over design space
  - Test expressiveness of each visual encoding Generate encodings that pass test
  - Rank by perceptual effectiveness criteria

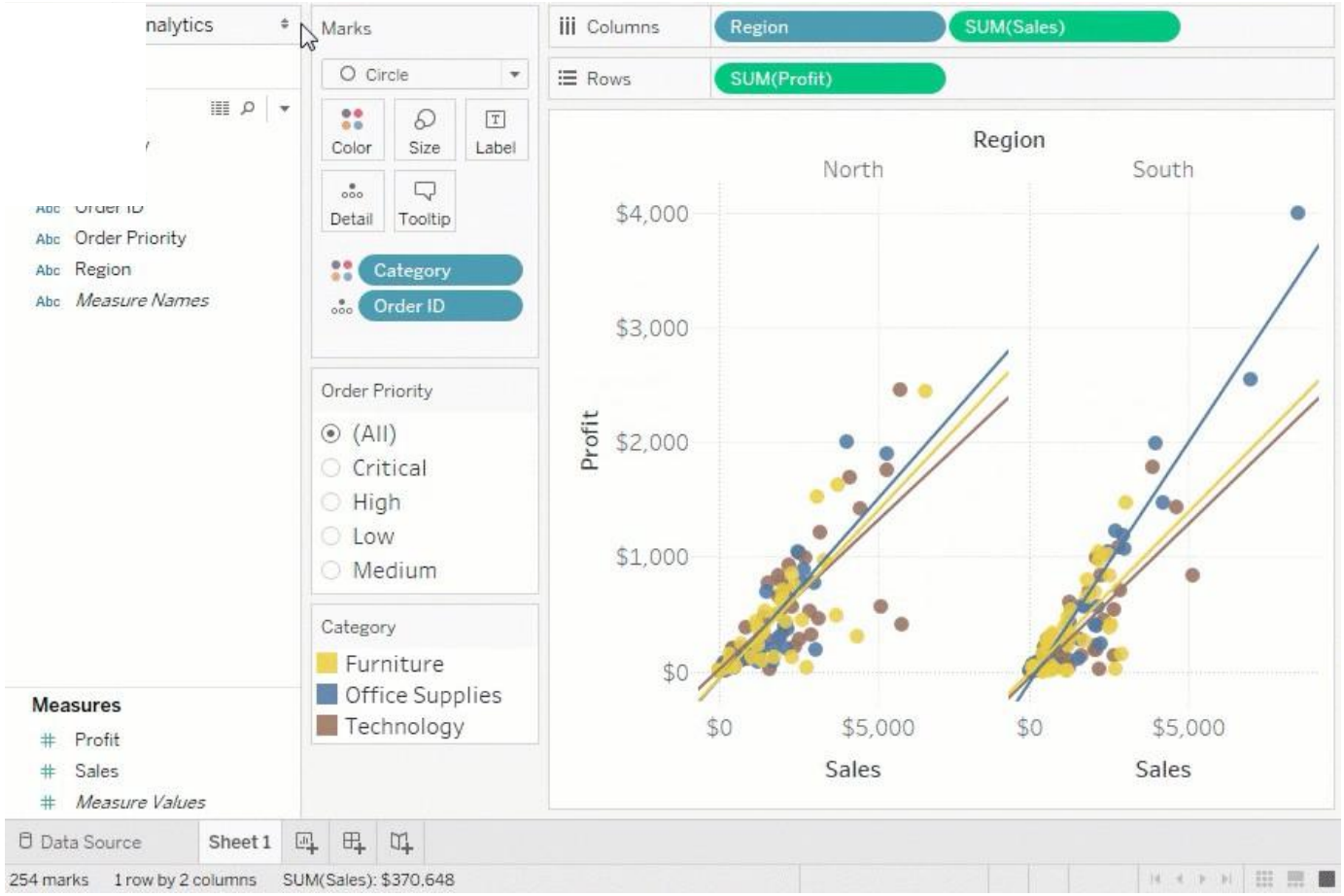
- Output the "most effective" visualization

APT

- Automatically generate a
  variables:
  - Price
  - Mileage
  - Repair
  - Weight



Tableau


nalytics
Abc Order Priority
Abc Region
Abc Measure Names
Measures
Profit
Sales
Measure Values
Marks
Circle
Color
Size
Label
Detail
Tooltip
Category
Order ID
Order Priority
(All)
Critical
High
Low
Medium
Category
Furniture
Office Supplies
Technology
Columns
Region
SUM(Sales)
Rows
SUM(Profit)
Region
North
South
$4,000
$3,000
$2,000
$1,000
$0
Profit
$0
$5,000
$0
$5,000
Sales
Sales
Data Source
Sheet 1
254 marks
1 row by 2 columns
SUM(Sales): $370,648

Take away: Visual Encoding Design

- Use expressive and effective encodings
- Avoid over-encoding
  - Reduce the problem space
  - Use interaction to generate relevant views
- Rarely does a single visualization answer all questions. Instead, the ability to generate appropriate visualizations quickly is critical!