



ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Visual Model & Visual Encoding

Credit: prof. Nam Wook Kim

Goal

Learn how data is mapped
to images

The Big Picture

Domain

goals, questions,
assumptions

Data

conceptual model
data model

Analysis task

identify, compare
summarize

Processing algorithms

data transformation

Visual encoding

mapping from data to image

Image

marks & channels

[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)

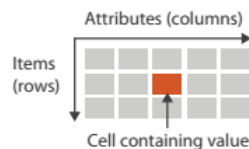
Trees (hierarchies)

Networks (graphs)

and combinations...

Dataset Types

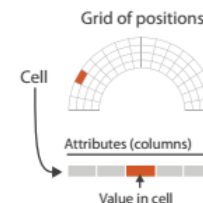
→ Tables



→ Networks



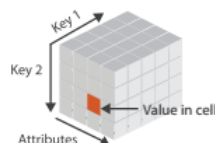
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table



→ Trees



Data Types

→ Items

→ Attributes

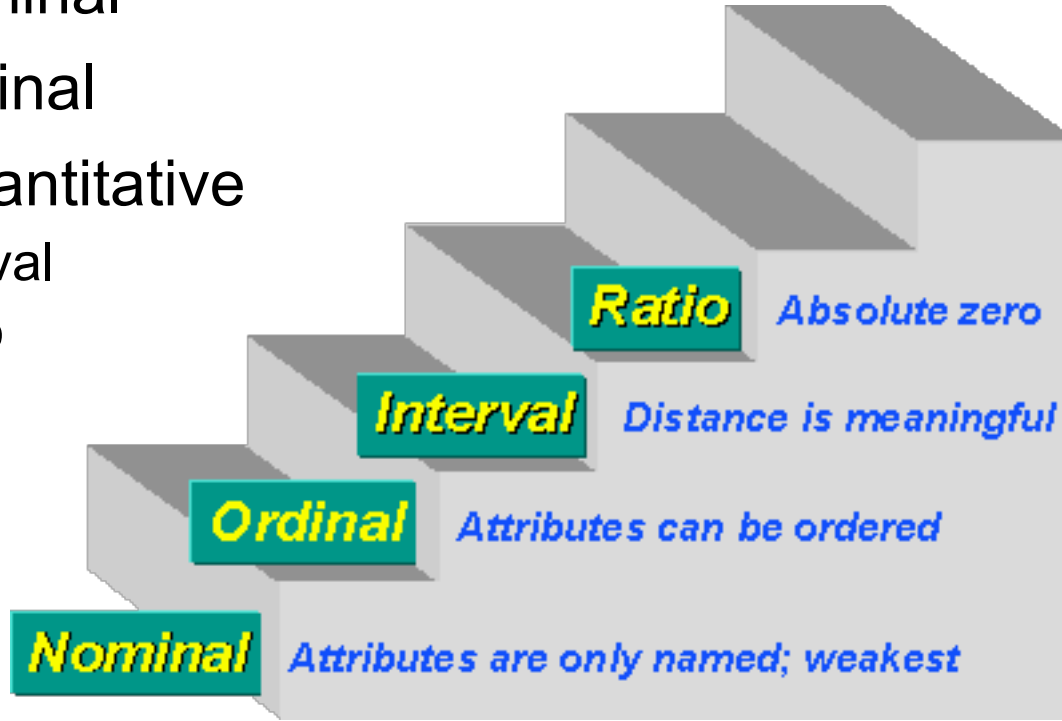
→ Links

→ Positions

→ Grids

Levels of measurement

- Stevens (1946) classified variables into four levels. These are referred 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
- 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}

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 $^{\circ}\text{F}$
 - A one-degree temperature ($^{\circ}\text{F}$) 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

Properties of images

Image Models

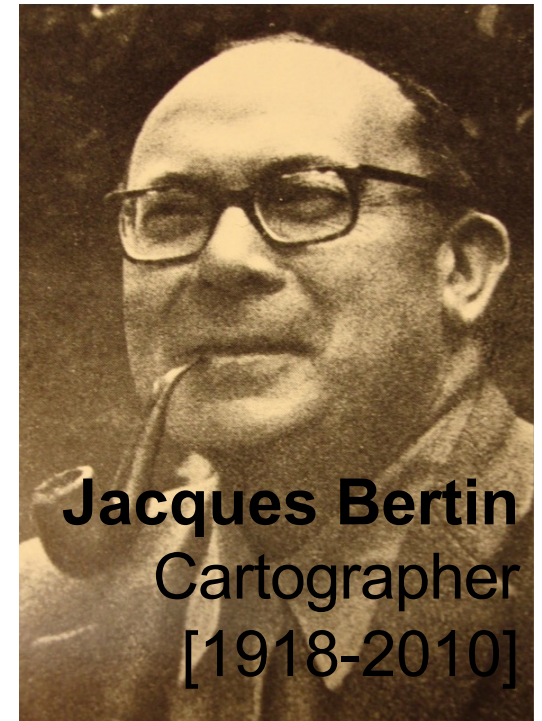
Visual Language is a Sign System

Images perceived as a set of signs

Sender encodes information in signs

Receiver decodes information from signs

Semiology of Graphics, 1967



Jacques Bertin
Cartographer
[1918-2010]

Image Models

Visual Marks (Đối tượng biểu diễn)

Basic graphical elements in an image

Represent information

Perceptual Channels (Biến thị giác)

Control the appearance of marks

Encode information

	Points			Lines			Areas	
Position								
Size								
Value								
Texture								
Color								
Orientation								
Shape								

Visual marks to represent items

→ Points



0D

→ Lines




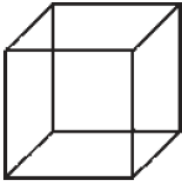


1D

→ Areas



2D

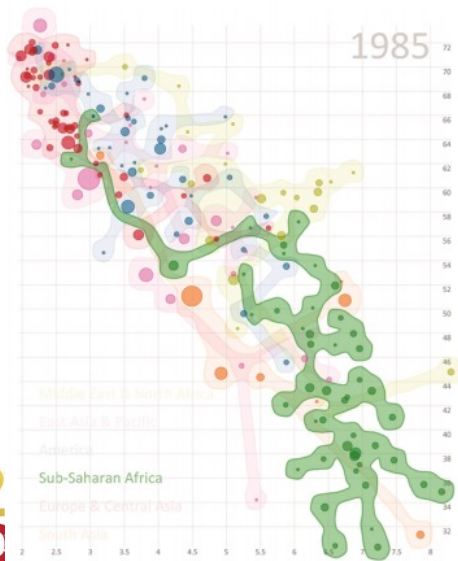
MARK	EXAMPLE	DESCRIPTION
Point		The <i>point</i> mark is commonly used as a marker to represent quantitative values through position on a scale, forming the basis of, for example, the scatter plot.
Line		The <i>line</i> mark is commonly used to represent quantitative values through variation in size (length), forming the basis of, for example, the bar chart.
Shape		The <i>shape</i> mark is commonly used to represent quantitative values through variation in size and position, forming the basis of, for example, the bubble plot.
Form		The <i>form</i> mark is used to represent quantitative values through variation in size (volume), forming the basis of charts that encode 3D representations.

Visual marks to represent links

➔ Containment

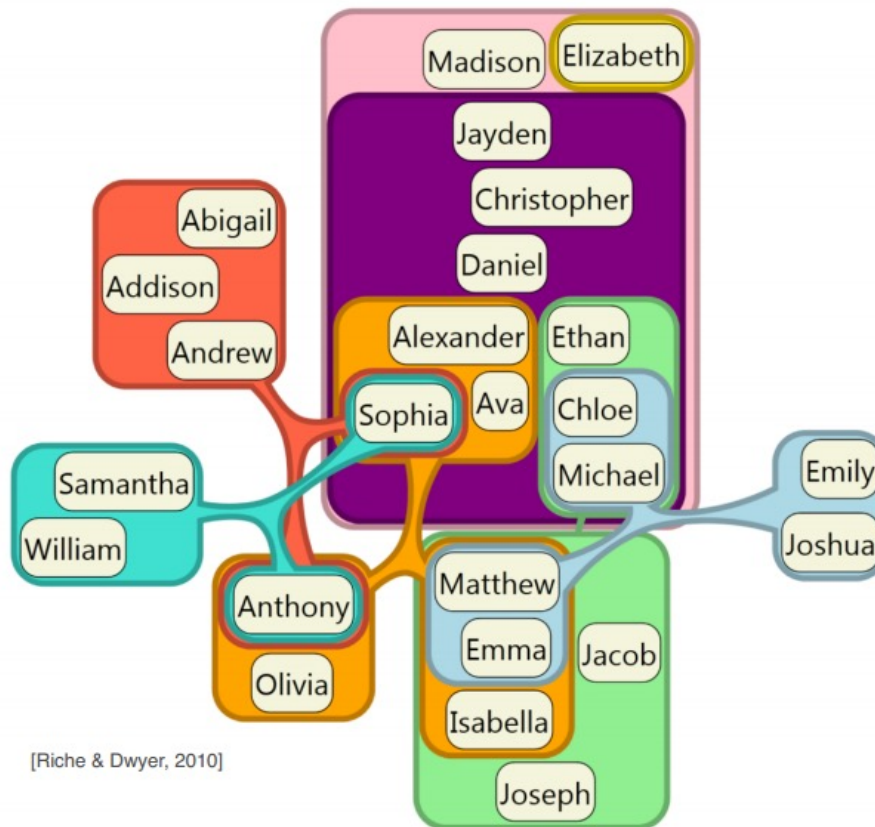


➔ Connection



Visual marks to represent links (2)

Nested containment



[Riche & Dwyer, 2010]

Perceptual Channels

- Control the appearance of marks
- Encode information

→ Position

→ Horizontal



→ Vertical



→ Both



→ Color



→ Shape



→ Tilt



→ Size

→ Length



→ Area



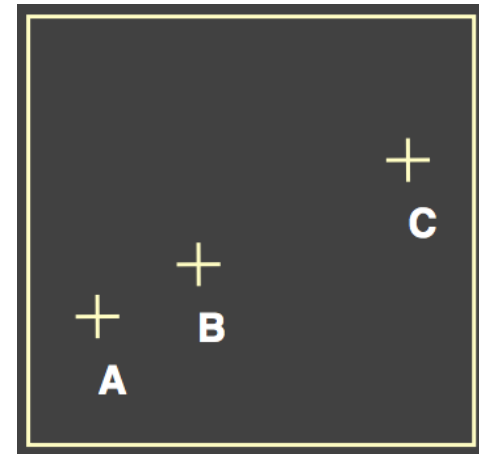
→ Volume



Perceptual channels: Position

- Can encode quantitative variables (Q)

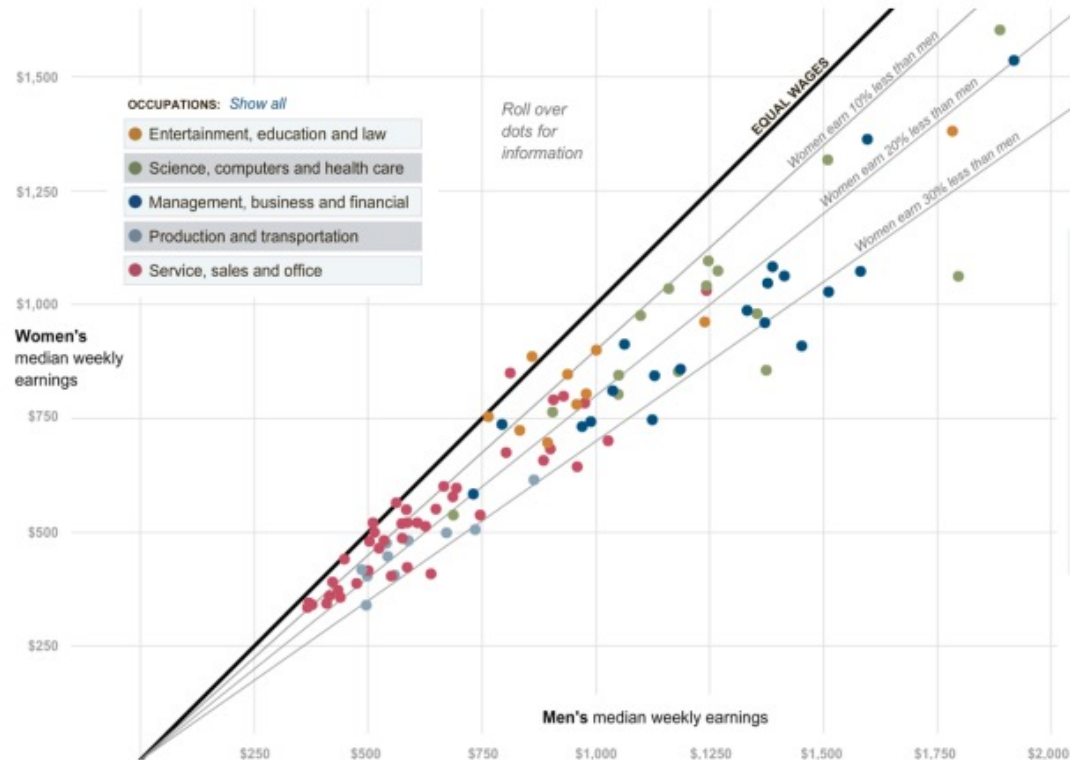
1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.



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

Perceptual channels: Position

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

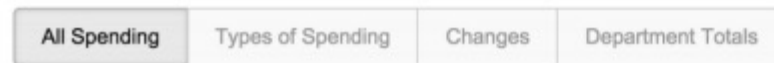


Perceptual channels: Size

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

Four Ways to Slice Obama's 2013 Budget Proposal

Explore every nook and cranny of President Obama's federal budget proposal.



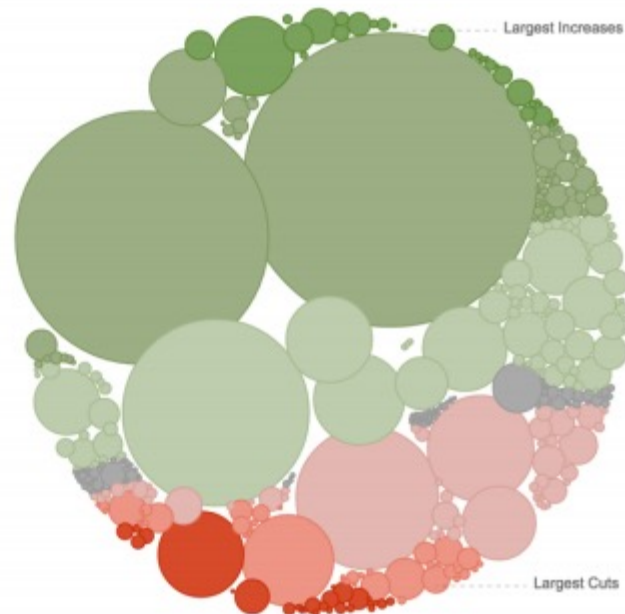
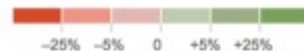
How \$3.7 Trillion Is Spent

Mr. Obama's budget proposal includes \$3.7 trillion in spending in 2013, and forecasts a \$901 billion deficit.

Circles are sized according to the proposed spending.



Color shows amount of cut or increase from 2012.



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

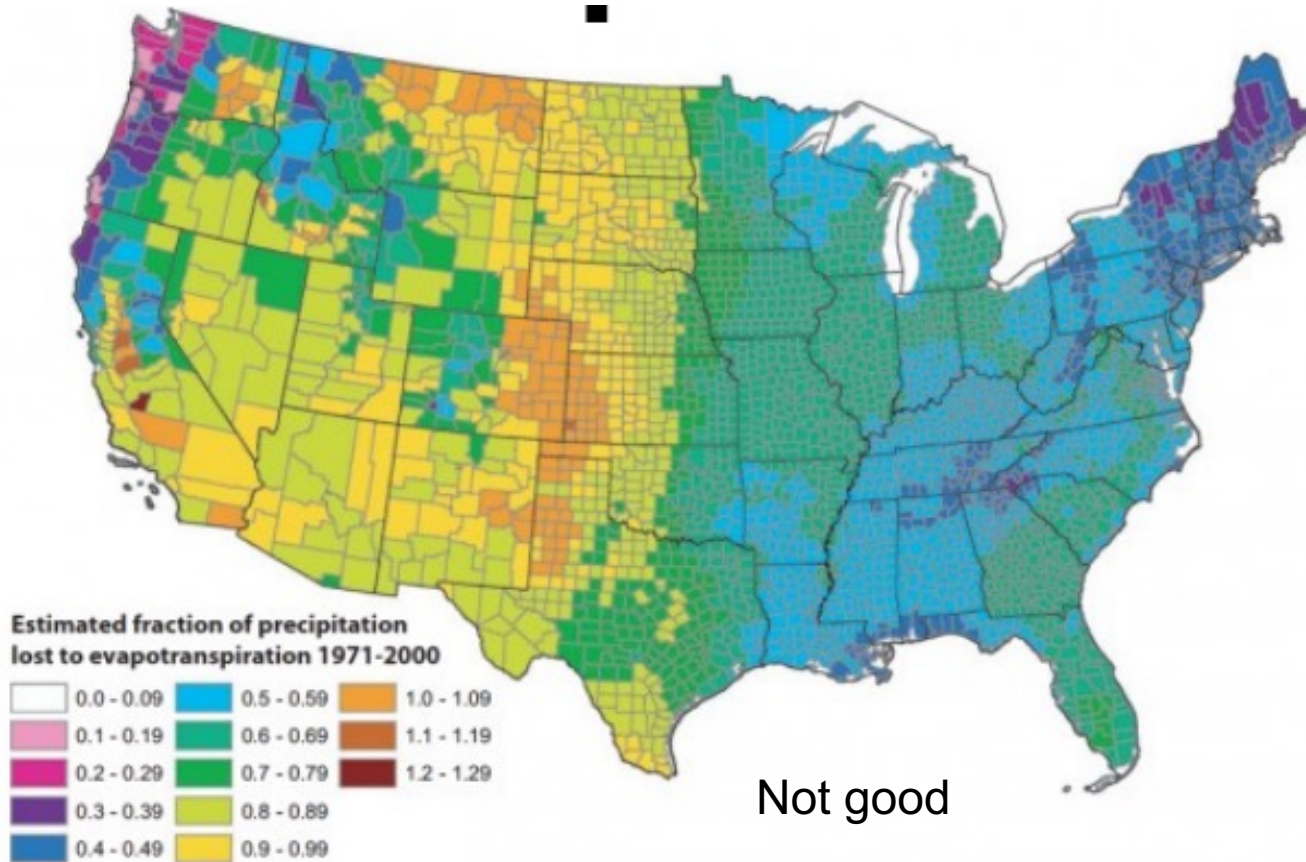


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM 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.

Perceptual channels: Shape

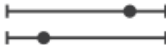

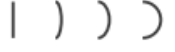

- Encode nominal variables (N)
- No ordered



Encoding information in perceptual channels

Quantitative/Ordered

➔ Magnitude Channels: Ordered Attributes

Position on common scale	
Position on unaligned scale	
Length (1D size)	
Tilt/angle	
Area (2D size)	
Depth (3D position)	
Color luminance	
Color saturation	
Curvature	
Volume (3D size)	

Nominal

➔ Identity Channels: Categorical Attributes

Spatial region	
Color hue	
Motion	
Shape	

Bertin's Levels of Organization

Position	N	O	Q	N ominal
Size	N	O	Q	O rdinal
Value	N	O	q	Q uantitative
Texture	N	o		Note: Q \subset O \subset N
Color	N			
Orientation	N			
Shape	N			

Mackinlay's Ranking

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

[Mackinlay 86]

Quantitative		Ordinal		Nominal
Position	————	Position	————	Position
Length		Density		Hue
Angle		Saturation		Texture
Slope		Hue		Connection
Area		Texture		Containment
Volume		Connection		Density
Density		Containment		Saturation
Saturation		Length		Shape
Hue		Angle		Length
Texture		Slope		Angle
Connection		Area		Slope
Containment		Volume		Area
Shape	————	Shape		Volume

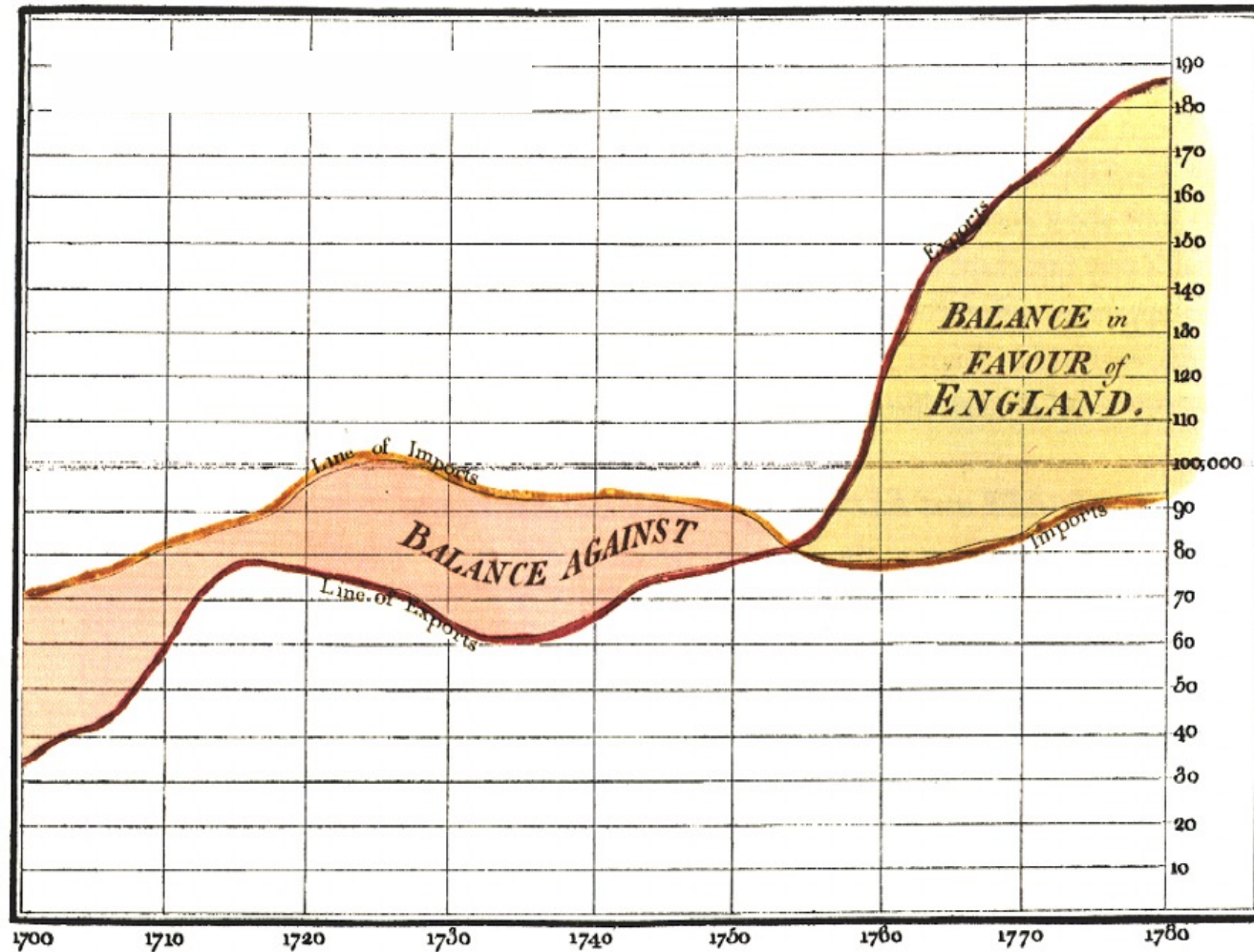


Jock D. Mackinlay
Vice President
Tableau Software

Example: Deconstructions

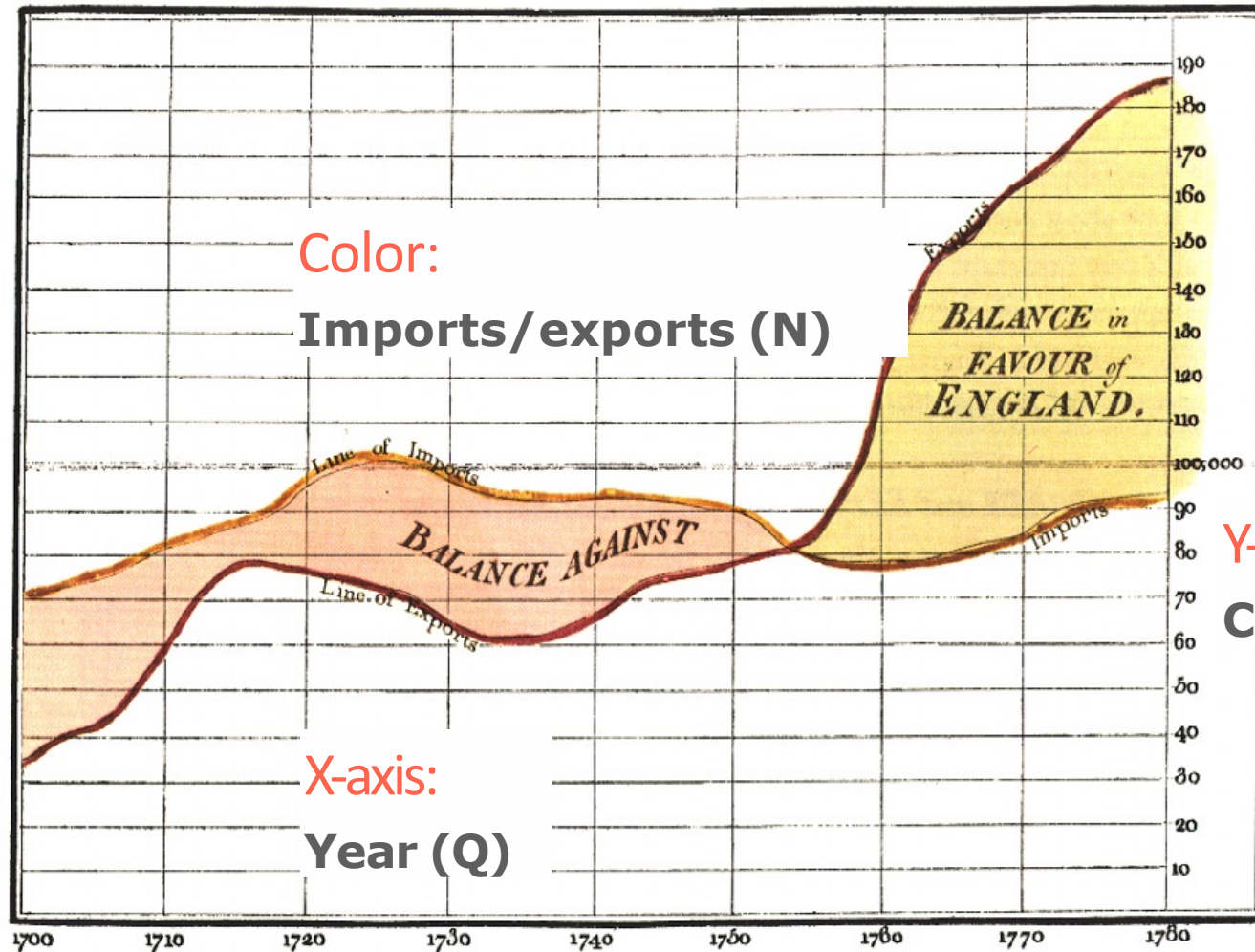
William Playfair, 1786

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



William Playfair, 1786

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



Color:

Imports/exports (N)

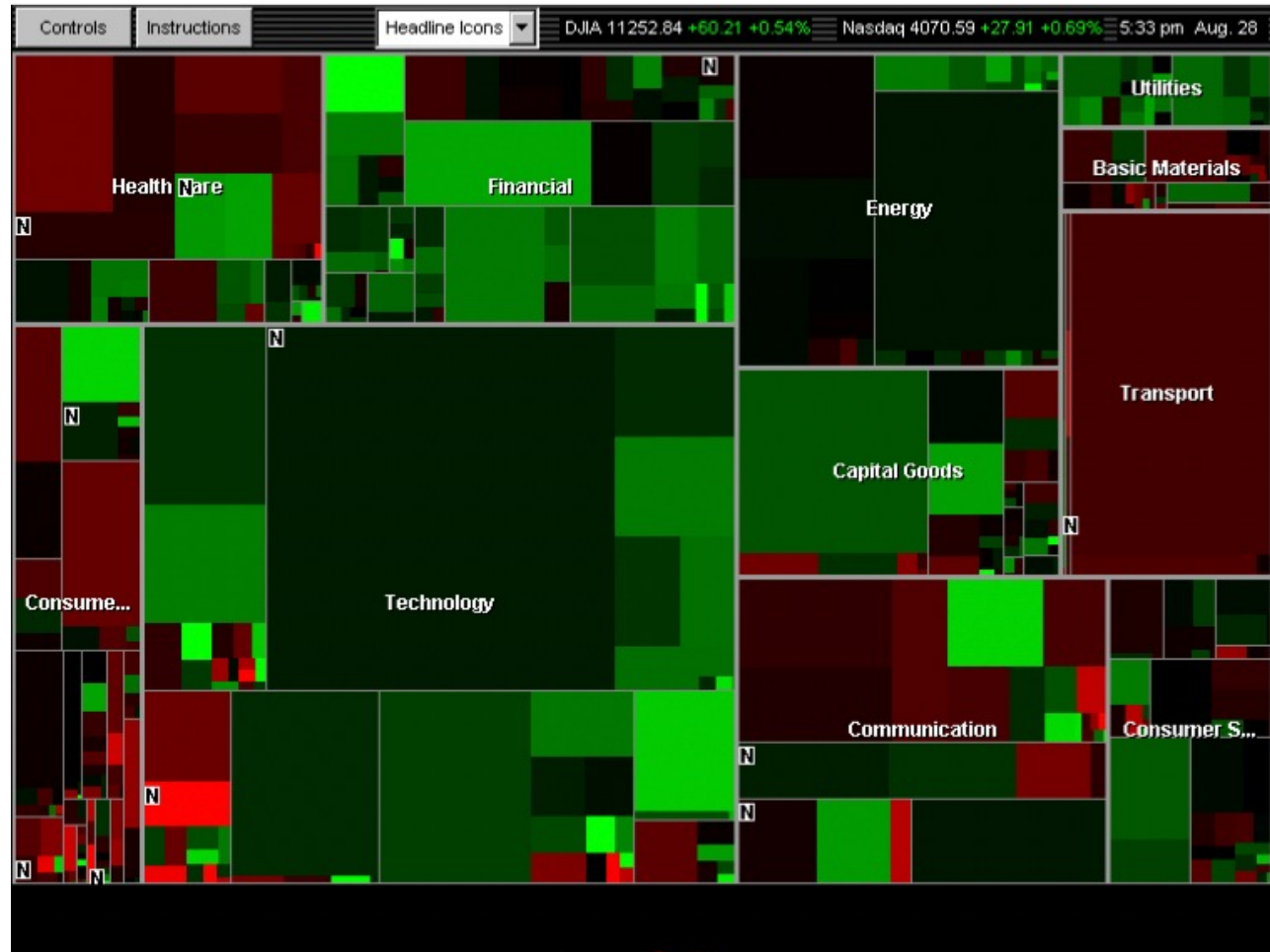
X-axis:

Year (Q)

Y-axis:

Currency (Q)

Wattenberg's Map of the Market



Wattenberg's Map of the Market

Rectangle

Area: market cap

(Q)

Rectangle

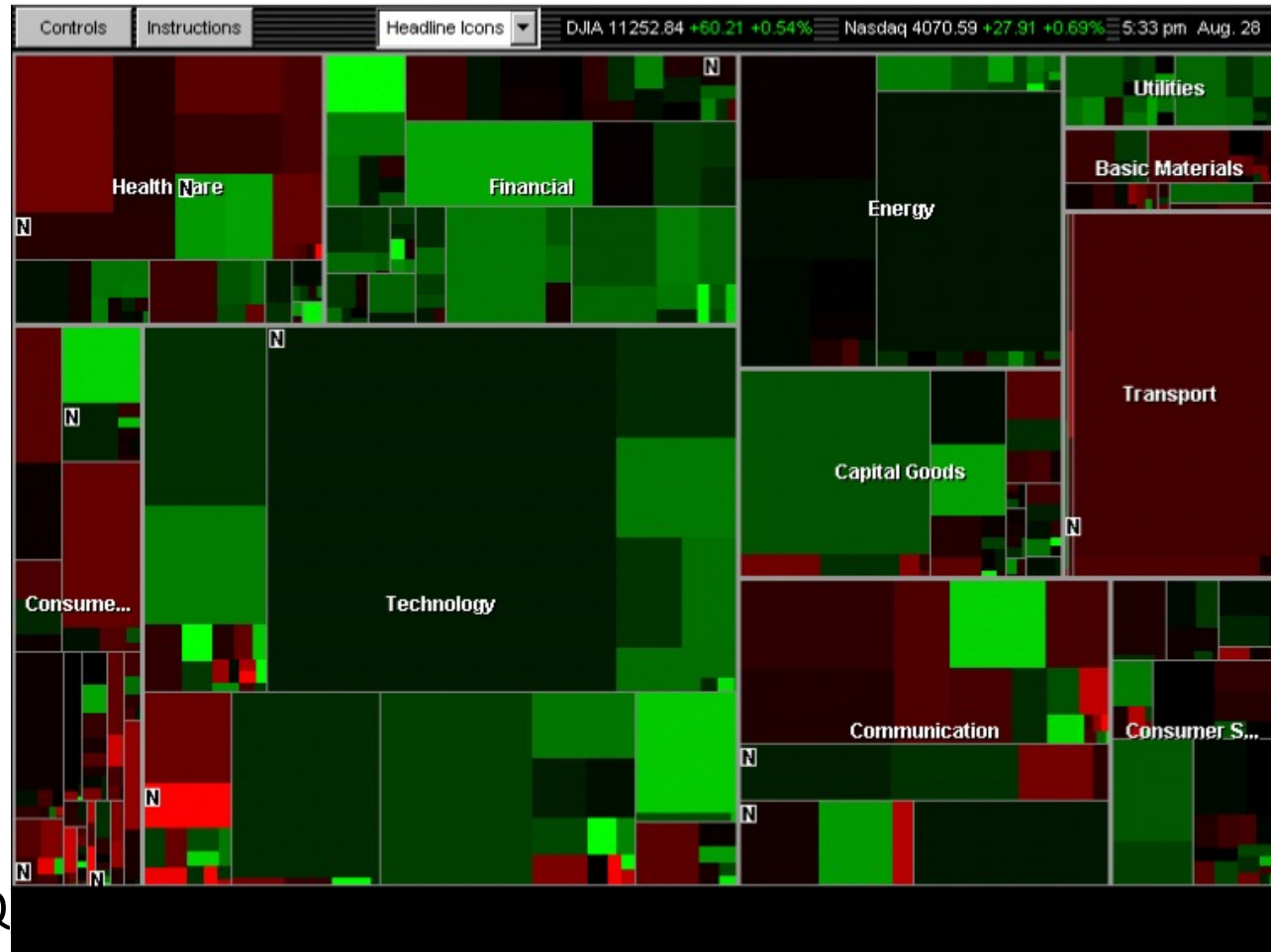
Position: market sector (N), market cap (Q)

Color Hue:

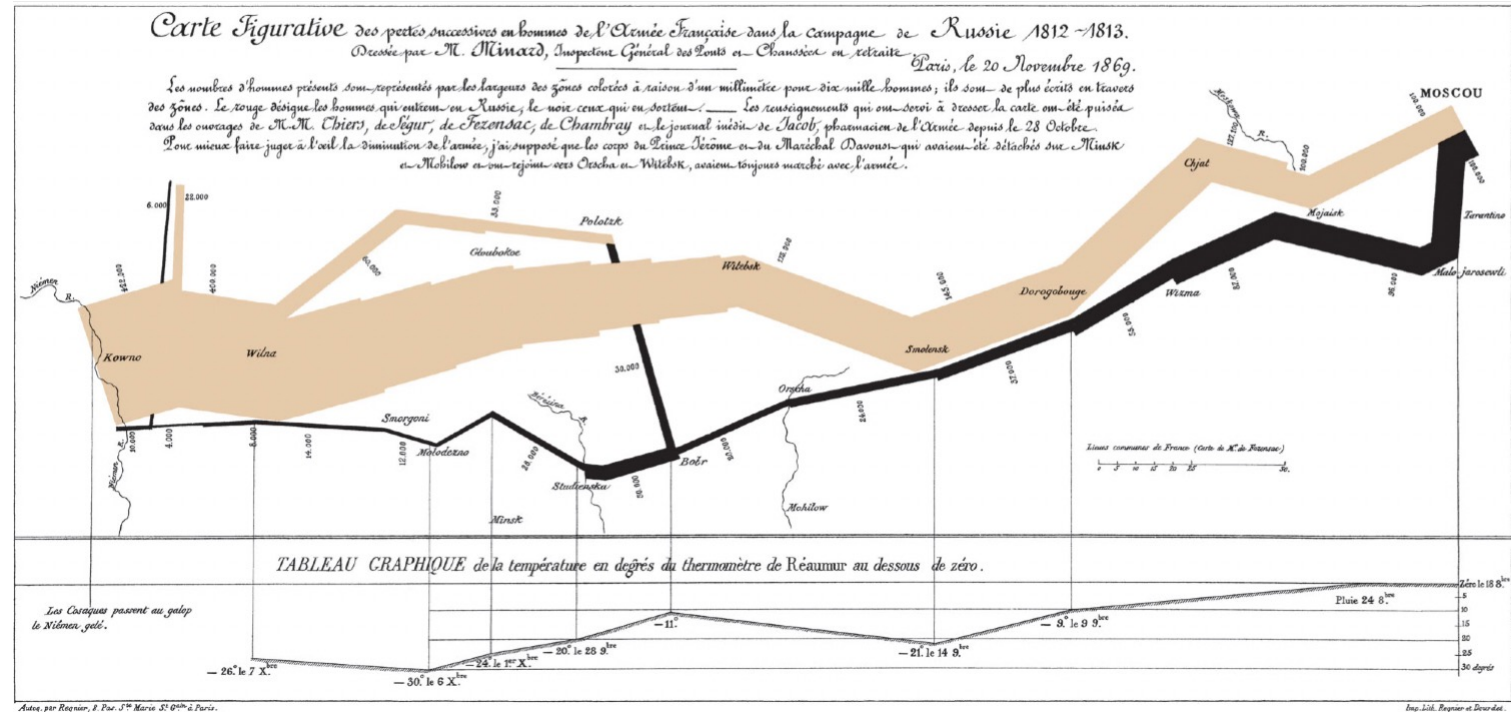
loss vs. gain (N)

Color Value:

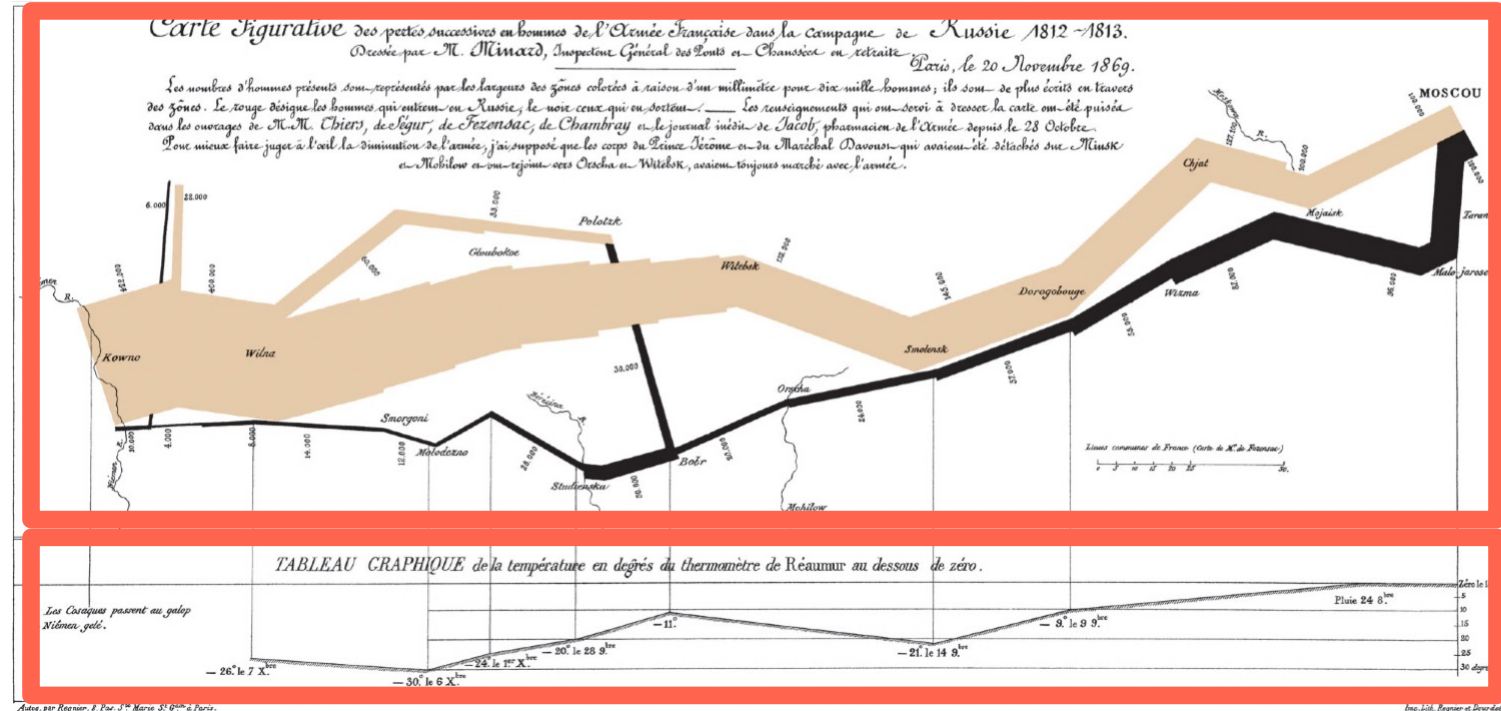
magnitude of loss or gain (Q)



Minard 1869: Napoleon's March



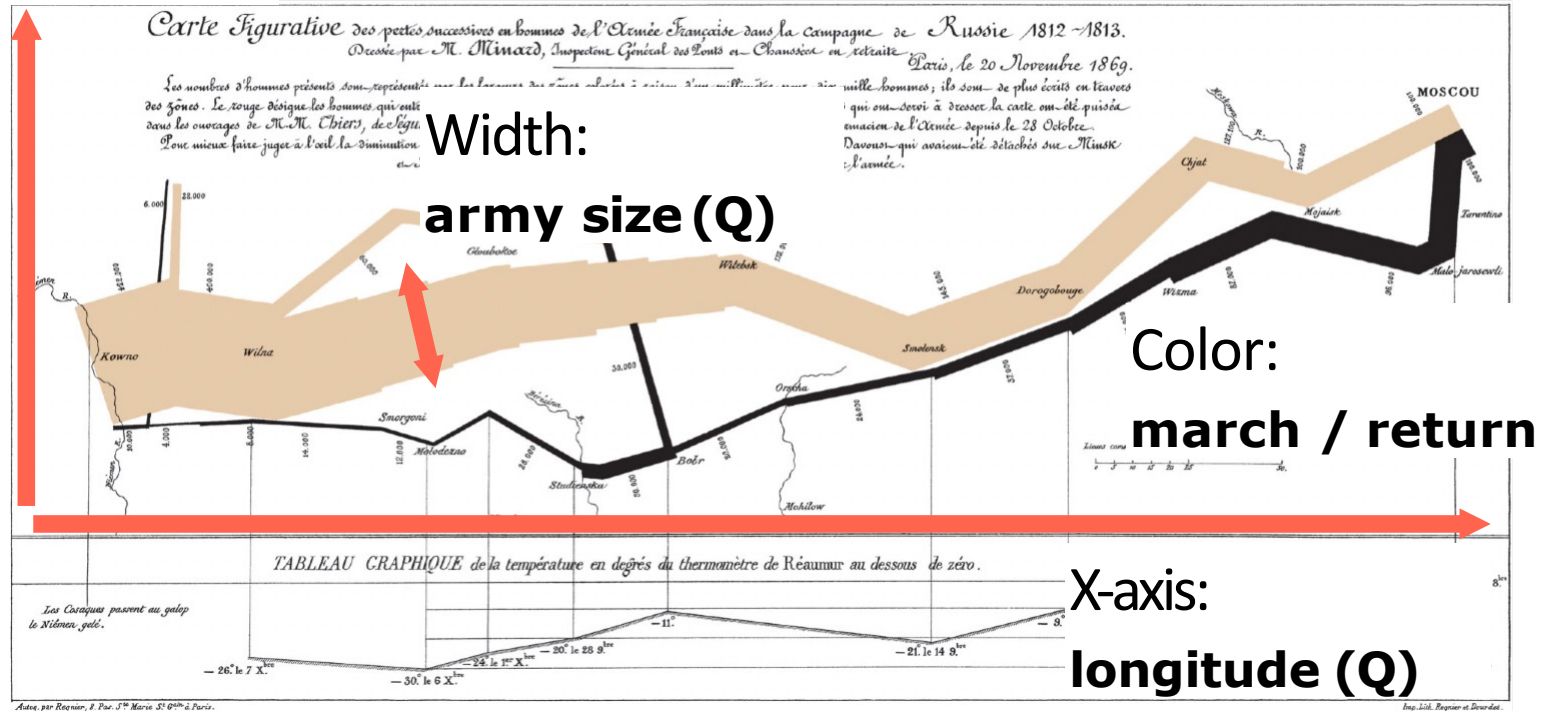
Minard 1869: Napoleon's March



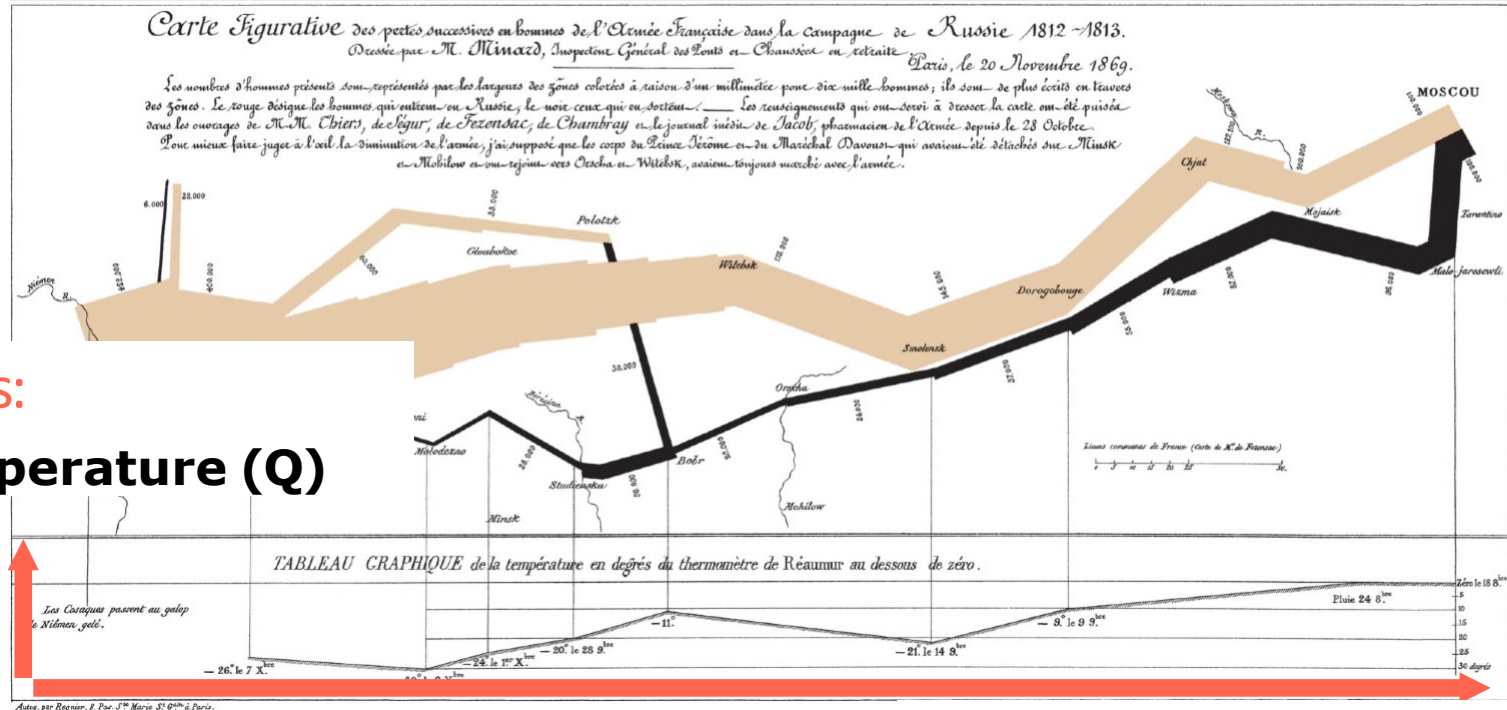
Minard 1869: Napoleon's March

Y-axis:

latitude (Q)



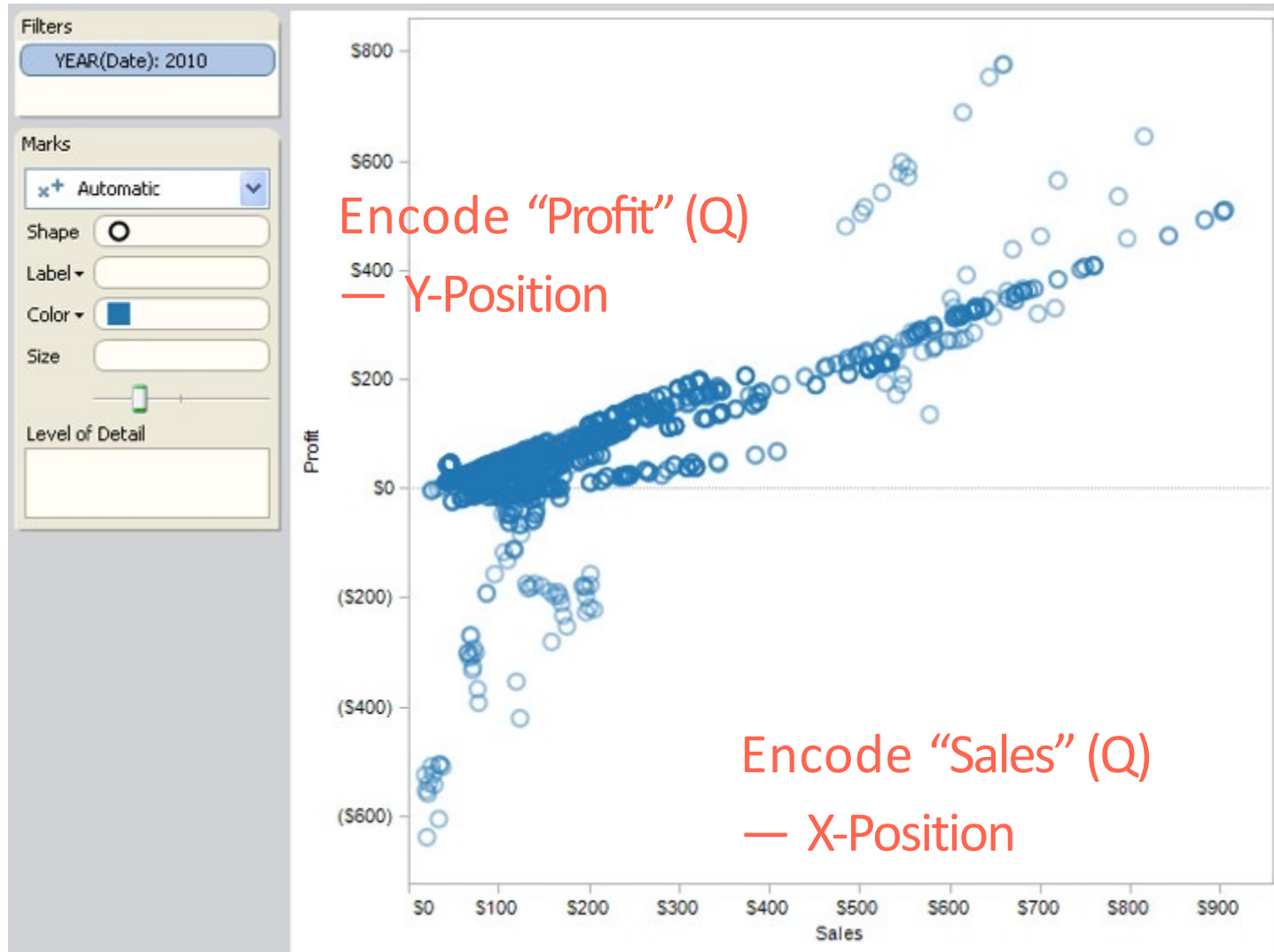
Minard 1869: Napoleon's March

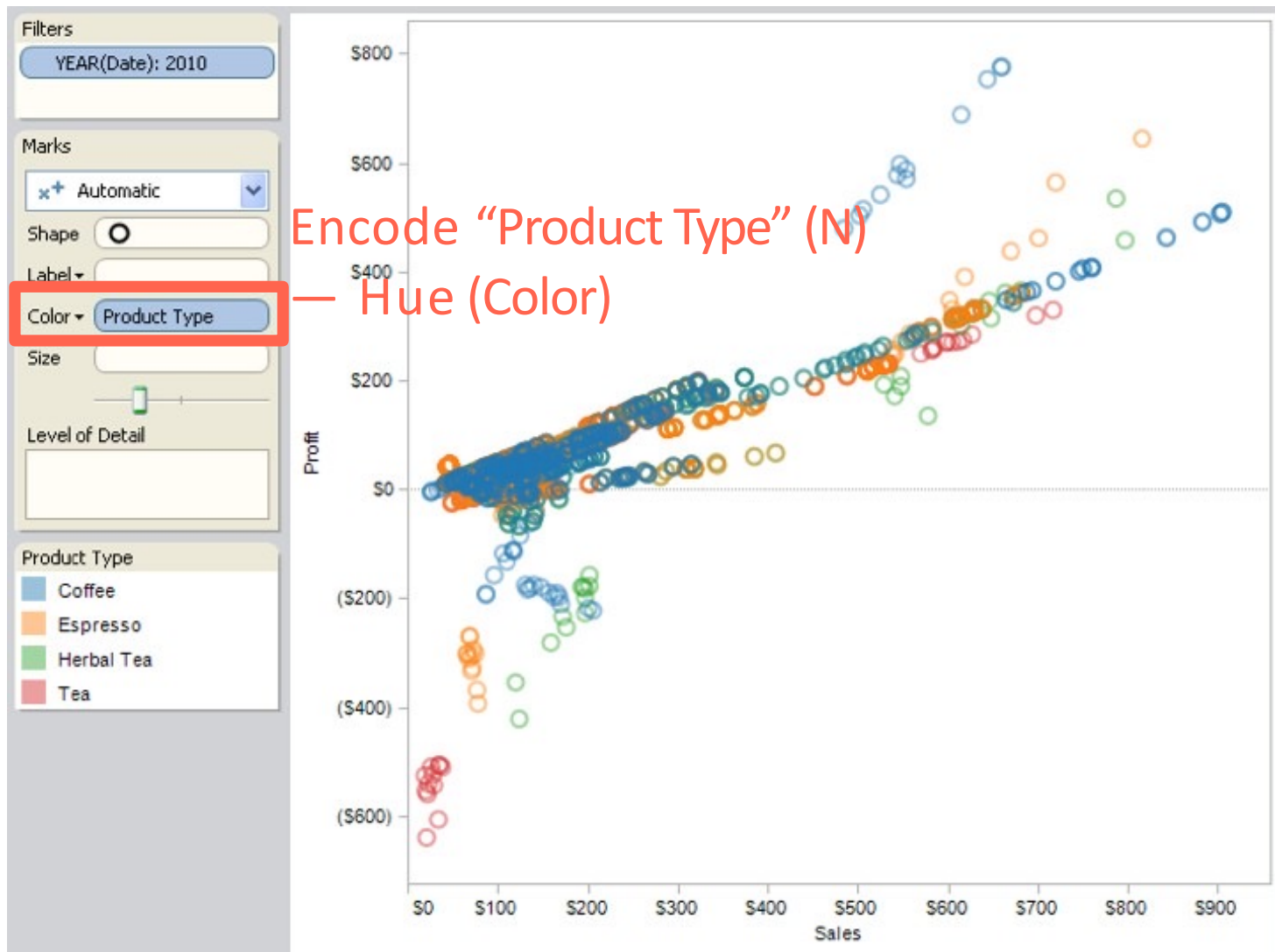


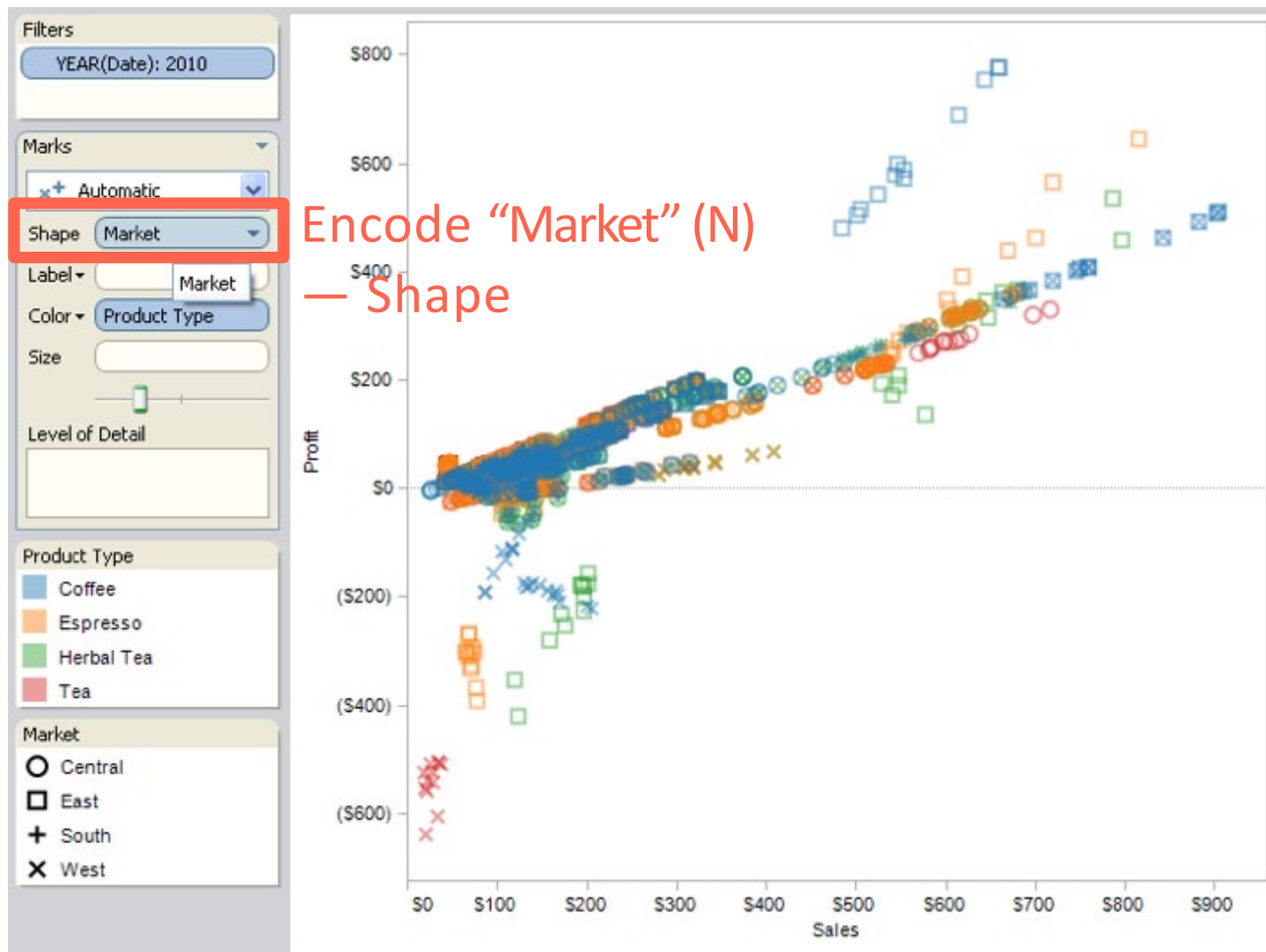
Example: Encoding Data

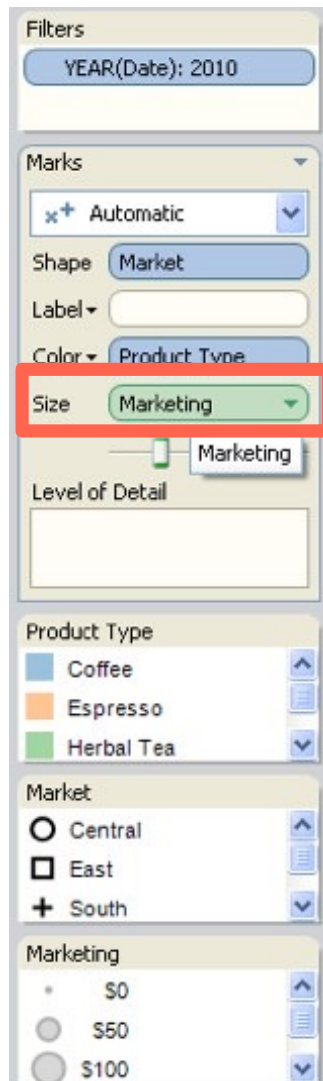
Example: Coffee Sales

- 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}

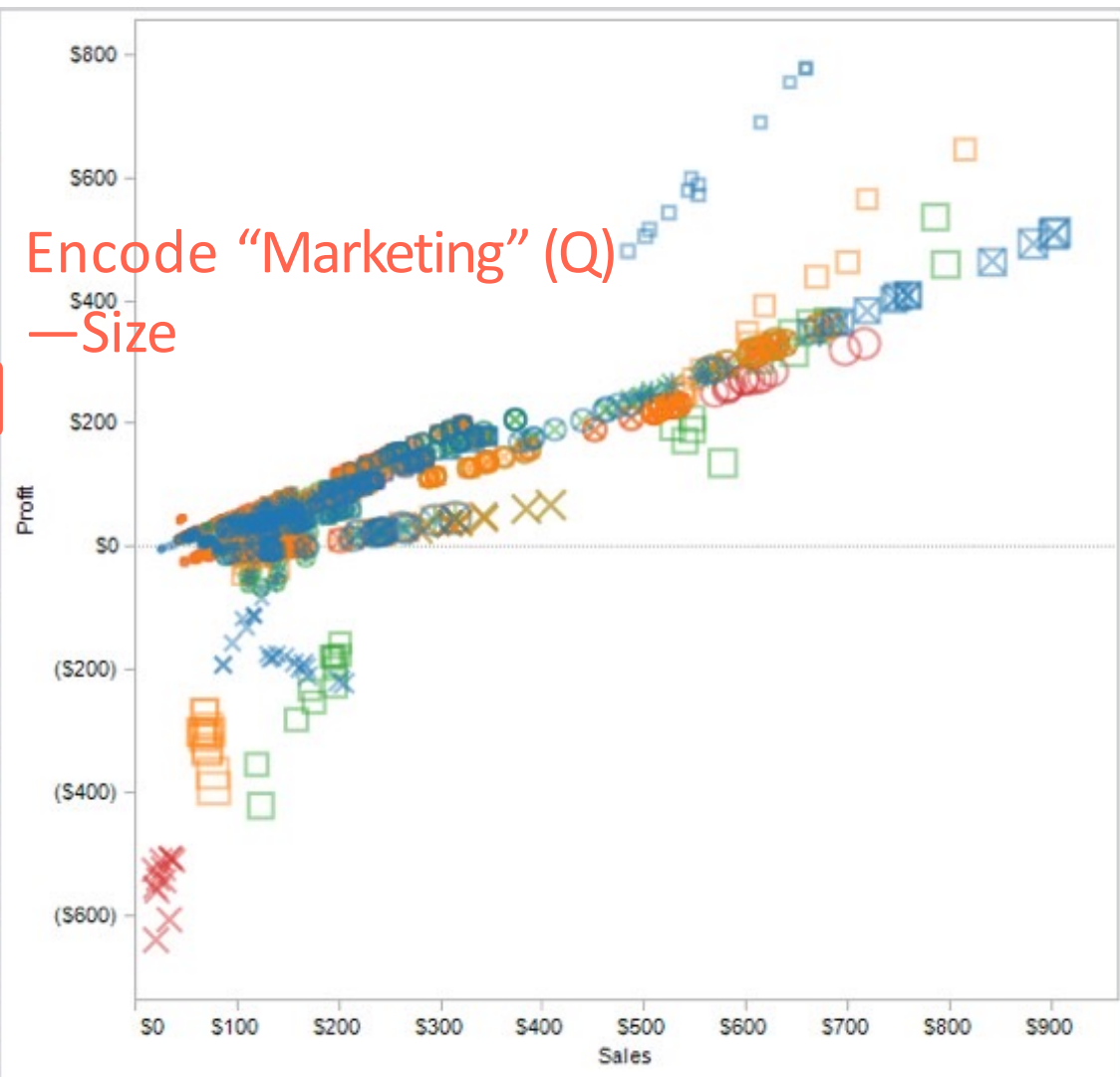






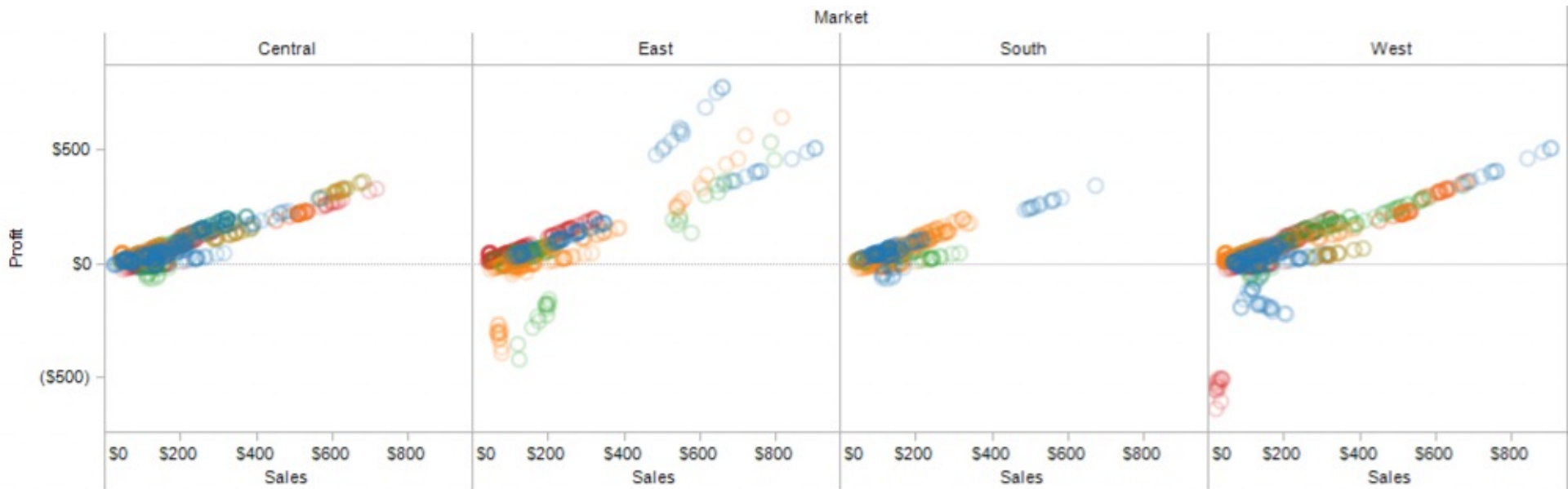


Encode "Marketing" (Q)
—Size



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 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 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

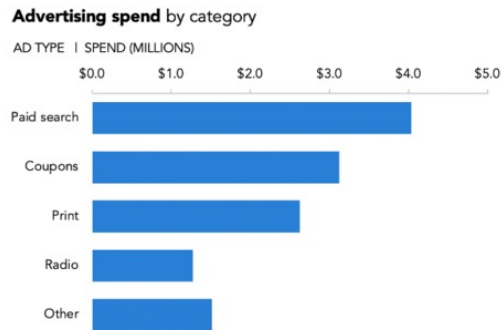
Design Criteria [Mackinlay 86]

- Expressiveness

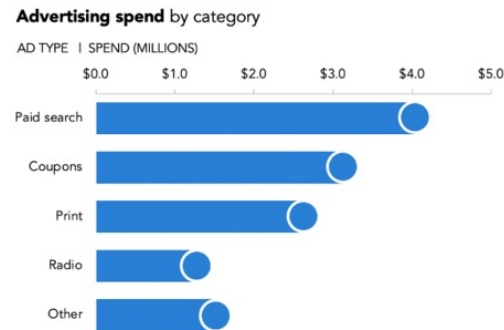
- A set of facts is expressible in a visualization if it expresses all the facts and only the facts in the data.
- Tell the truth and nothing but the truth
 - (don't lie, and don't lie by omission)

Dot plot

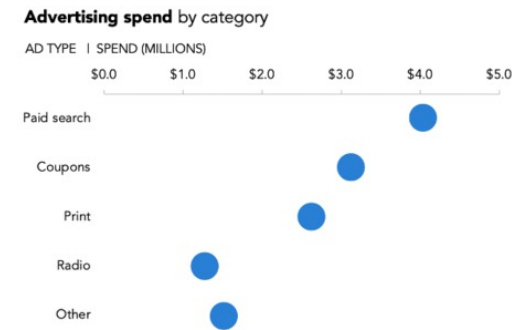
A typical horizontal bar chart...



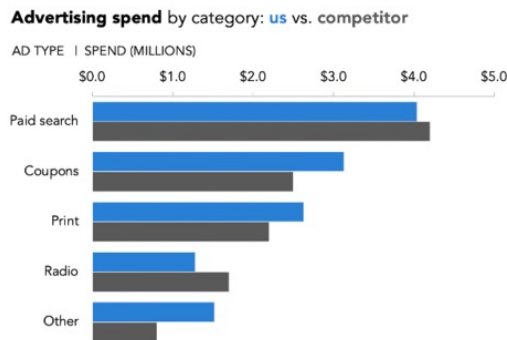
Encode end points with dots.



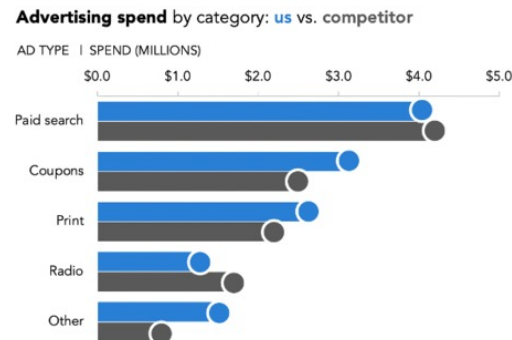
Remove the bars. Voila! A dot plot.



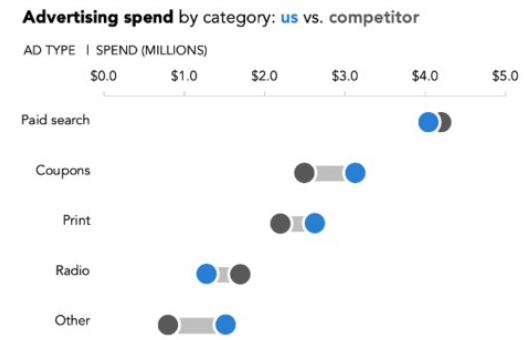
A dual-series bar chart...



Encode end points with dots.

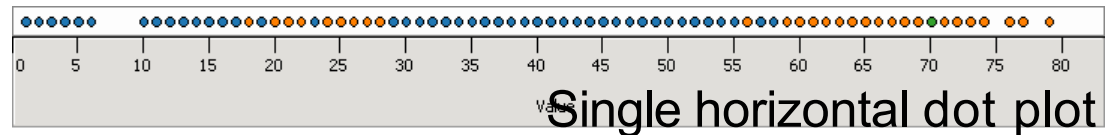


Collapse, remove bars, connect.



Can not express the facts

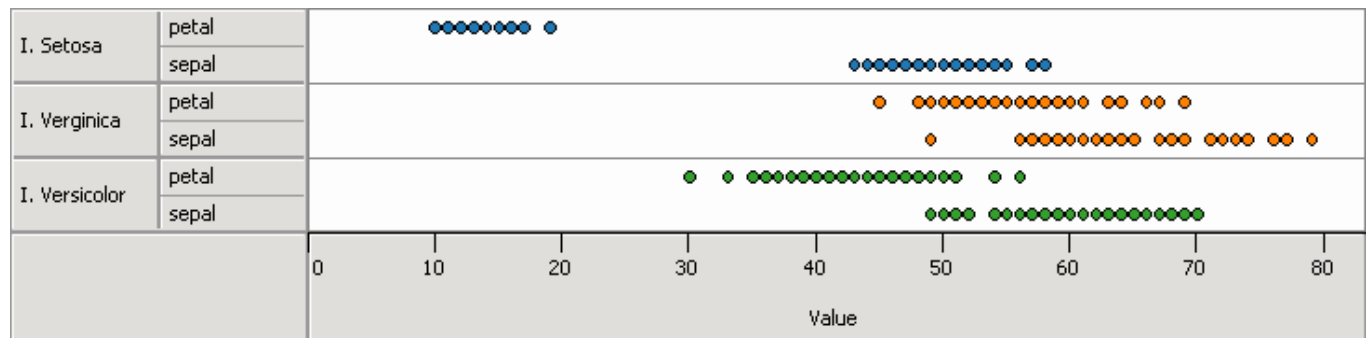
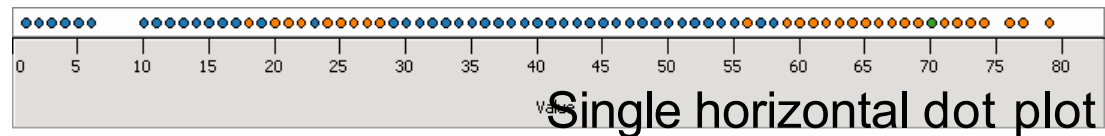
The relationship among multiple data attributes may not be expressed in a single horizontal dot plot.



Single horizontal dot plot

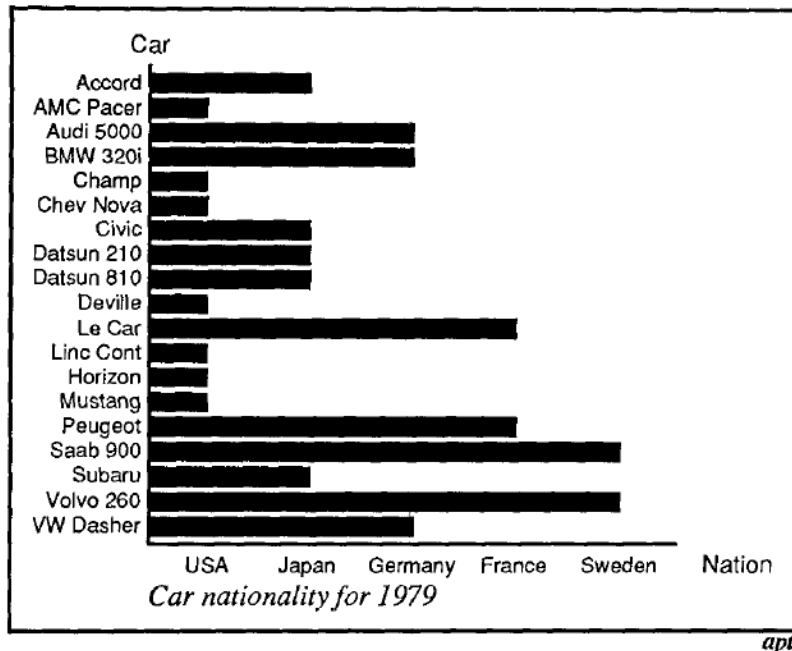
Can not express the facts

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



Categories in different positions

Expresses facts not in the data



A length is interpreted as a quantitative value.

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.

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)

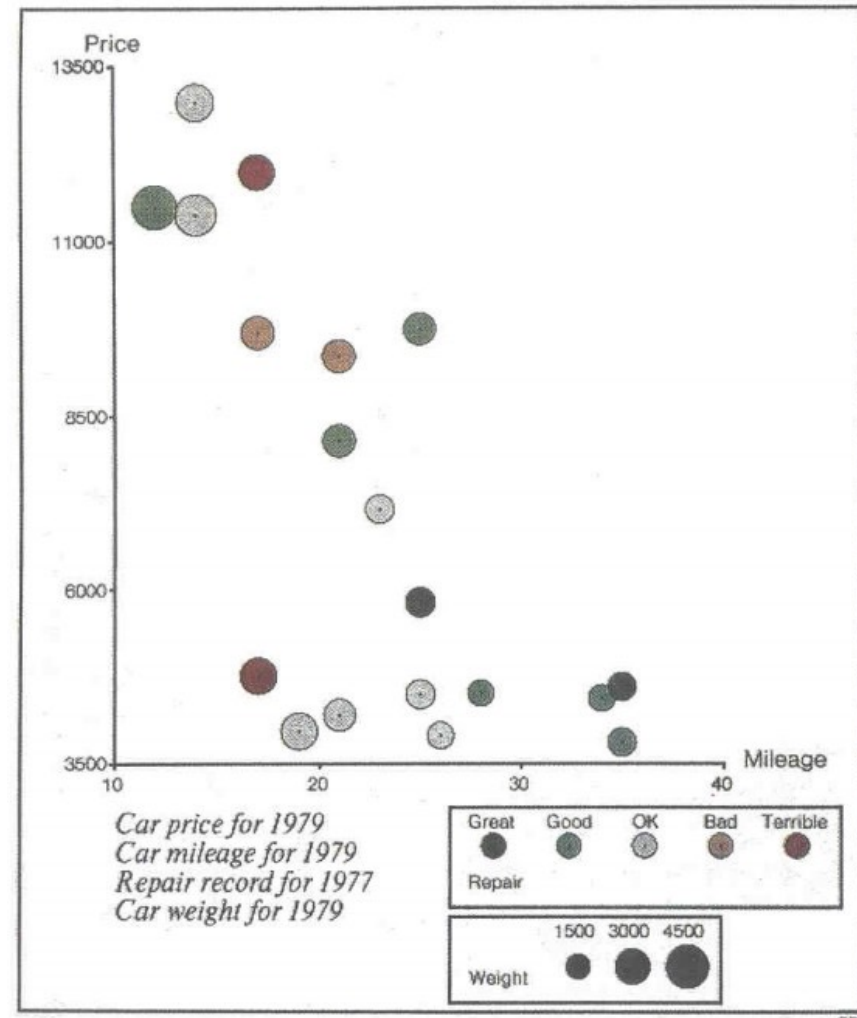
Mackinlay's Design Algorithm

- 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**

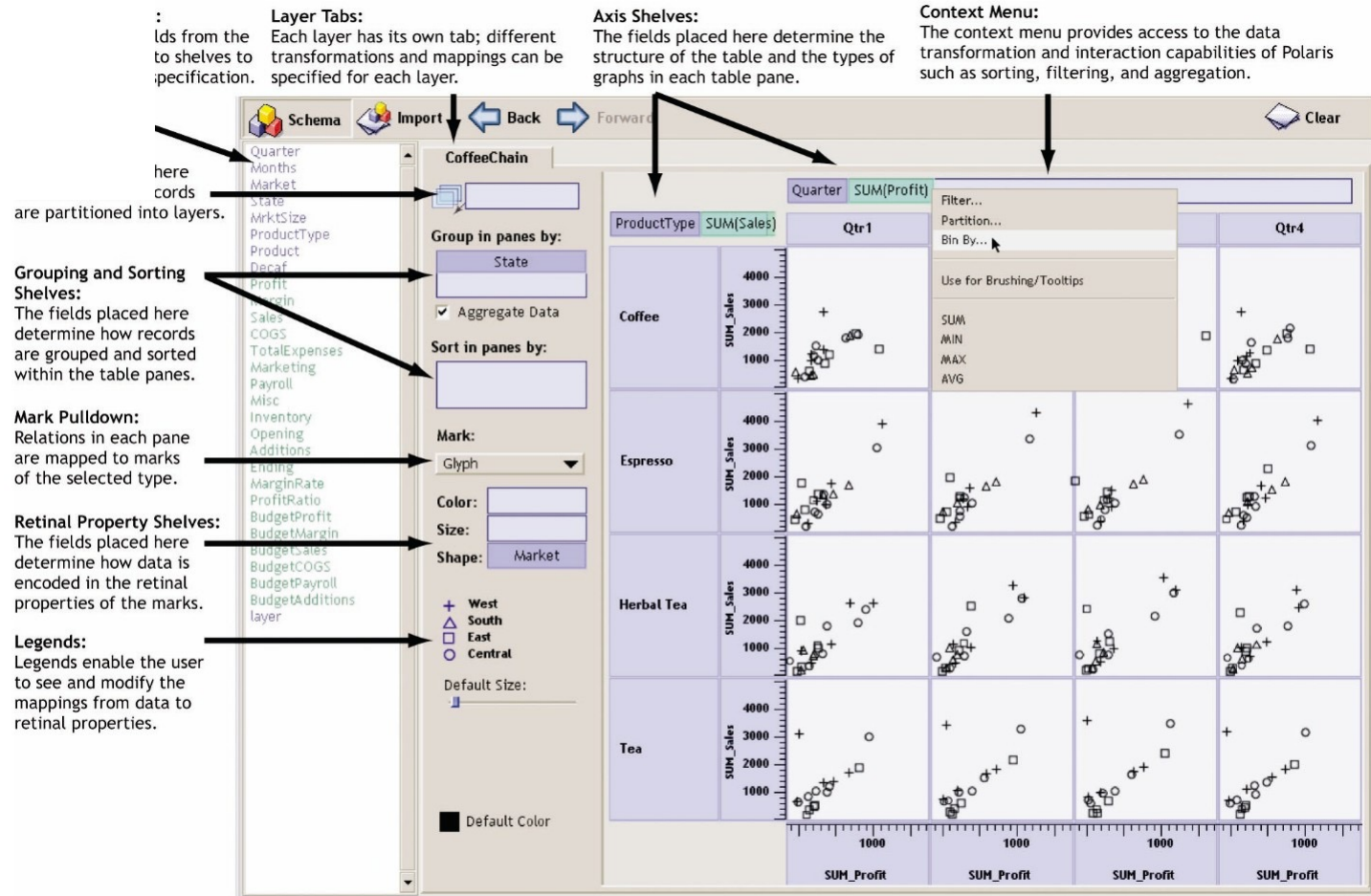
APT

- Automatically generate a chart for input variables:

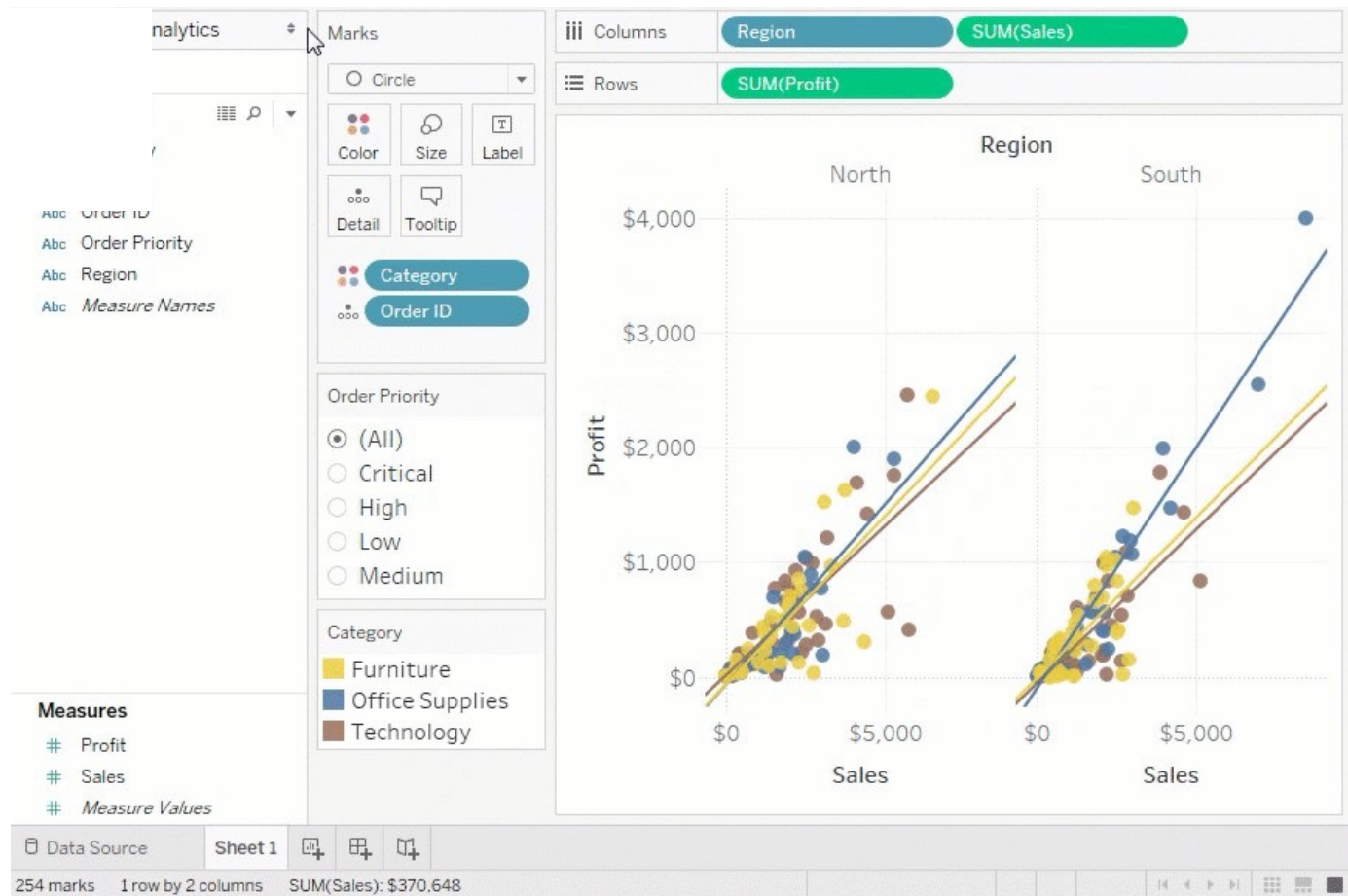
1. Price
2. Mileage
3. Repair
4. Weight



Polaris [Stolte et al 2002]



Tableau



Take away: Visual Encoding Design

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