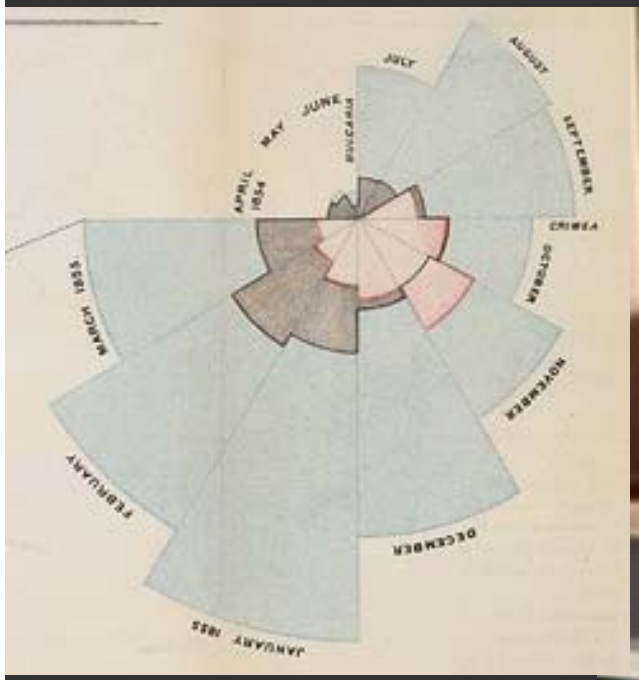
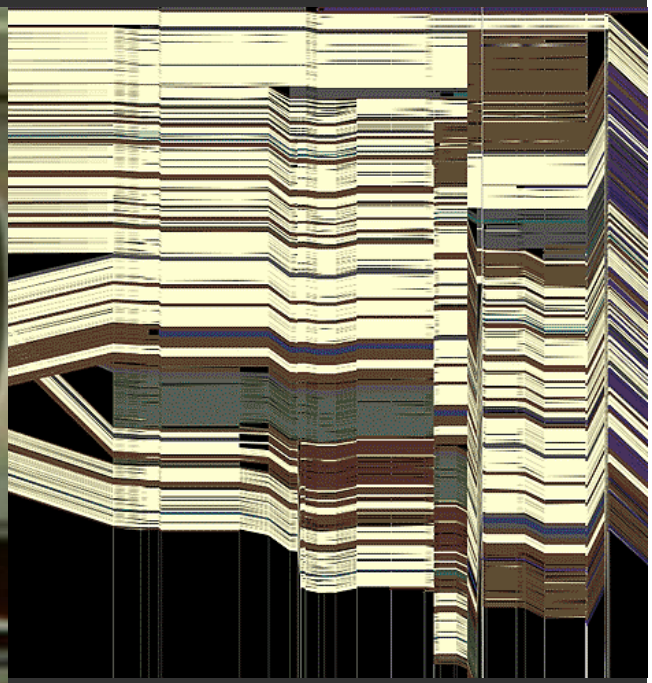
a brief introduction to

Data Visualization





Jeffrey Heer Stanford University

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics

u_X = 9.0 σ_X = 3.317

u_Y = 7.5 σ_Y = 2.03

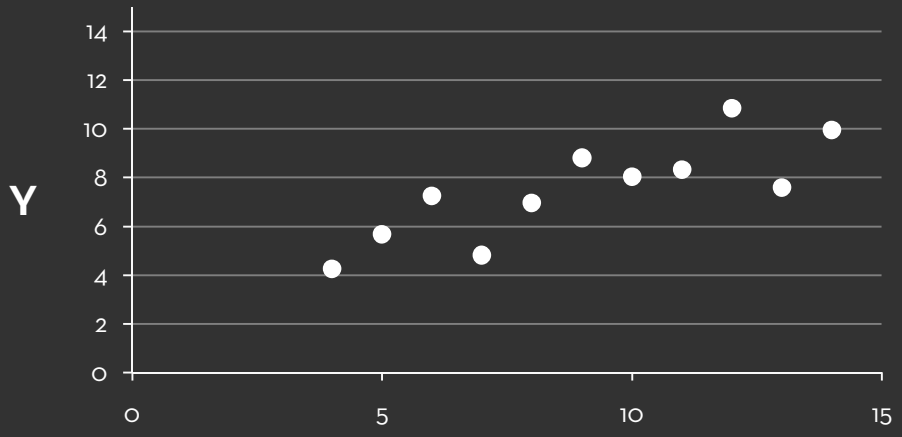
Linear Regression

Y = 3 + 0.5 X

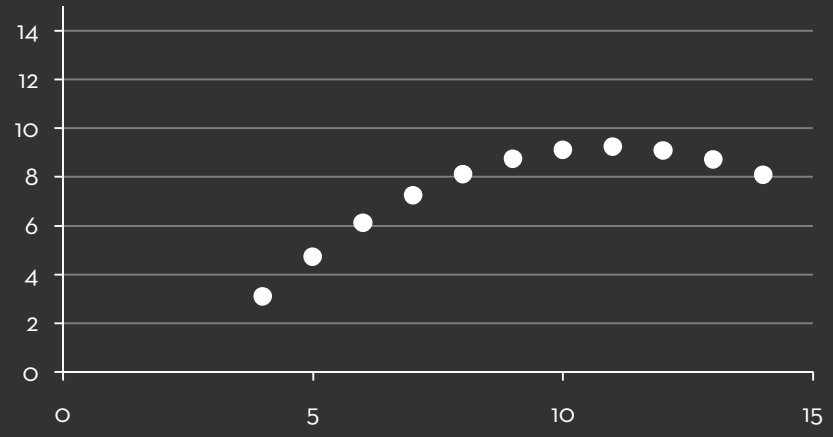
R^2 = 0.67

[Anscombe 73]

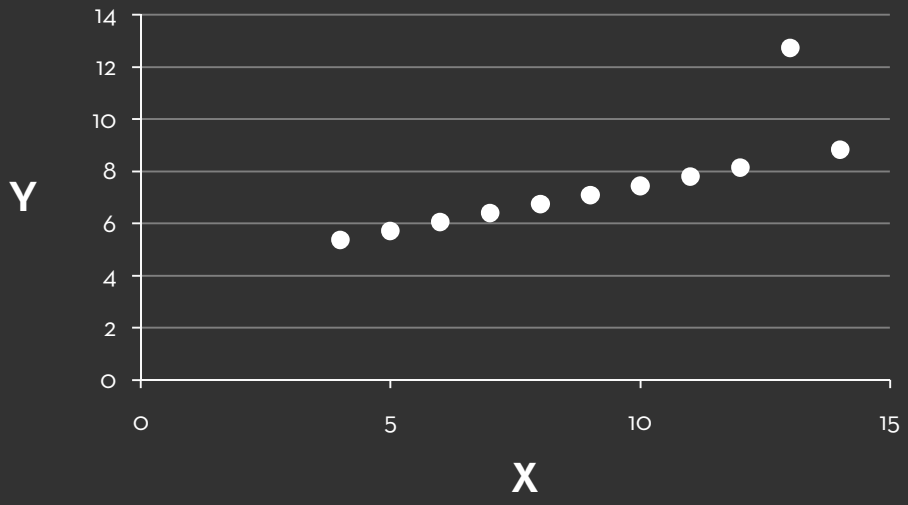
Set A



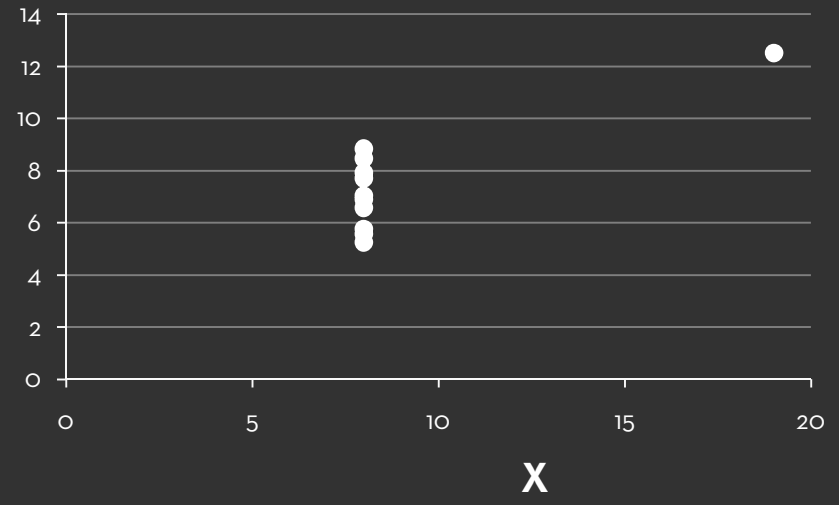
Set B

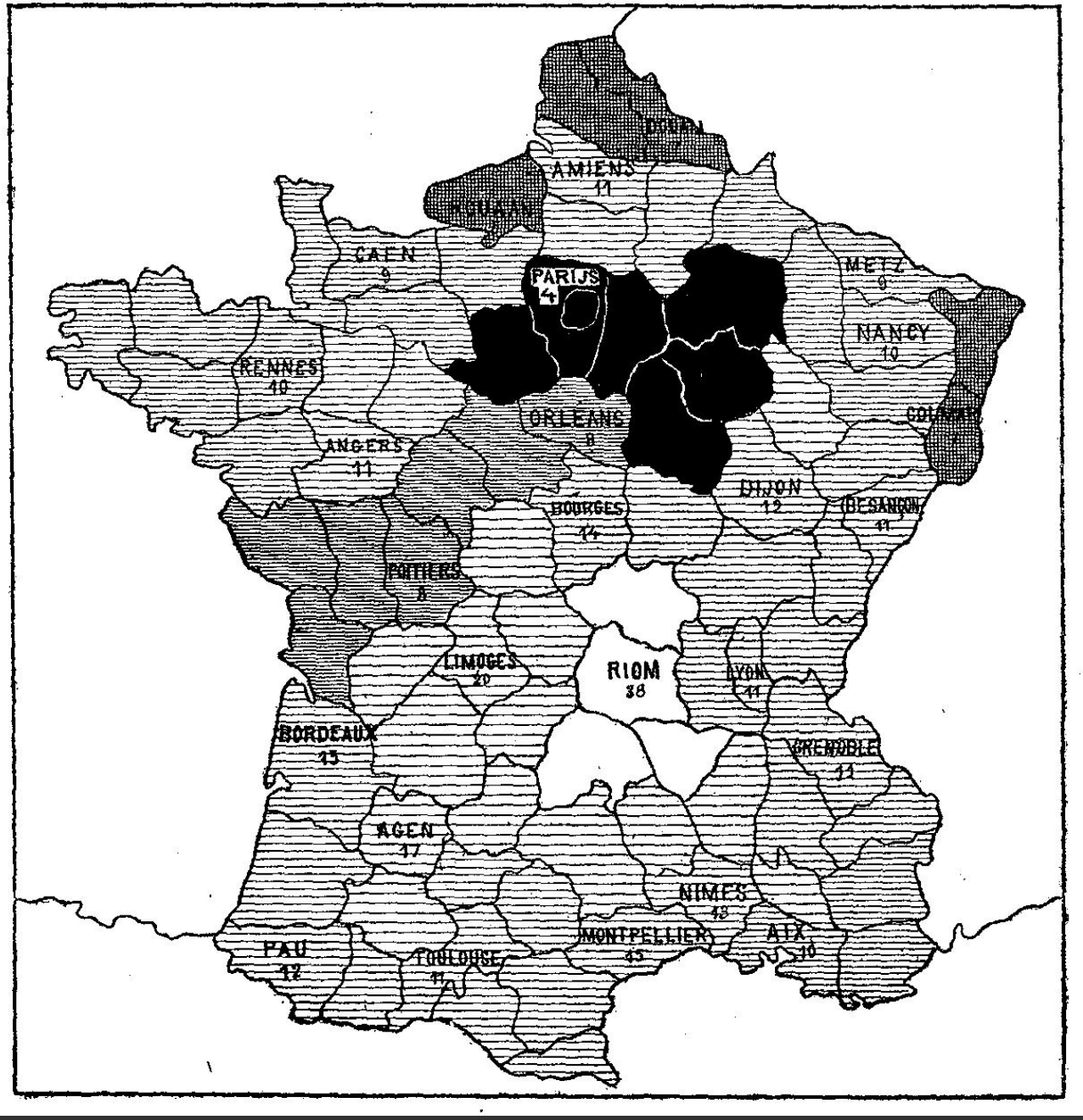


Set C



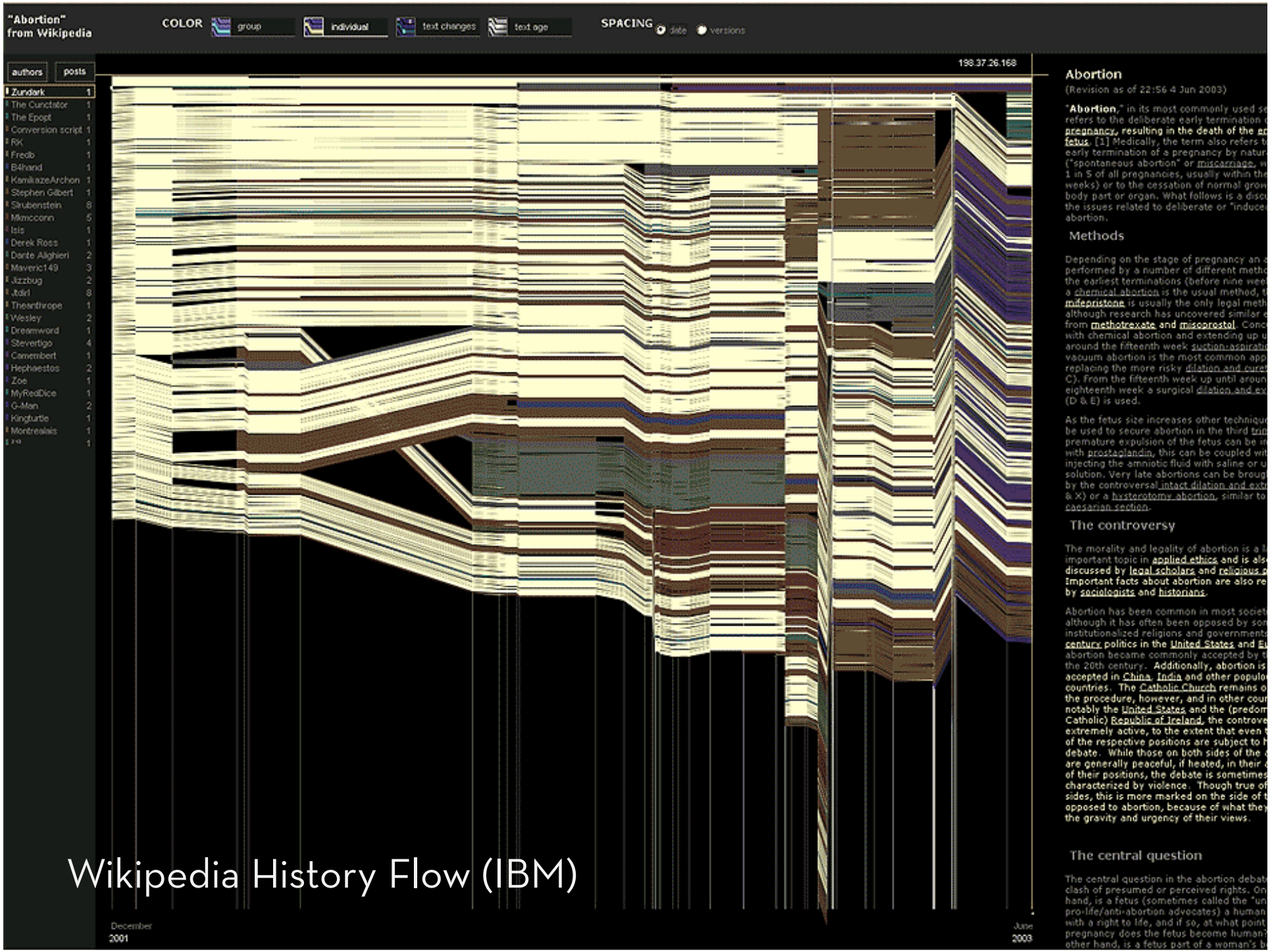
Set D



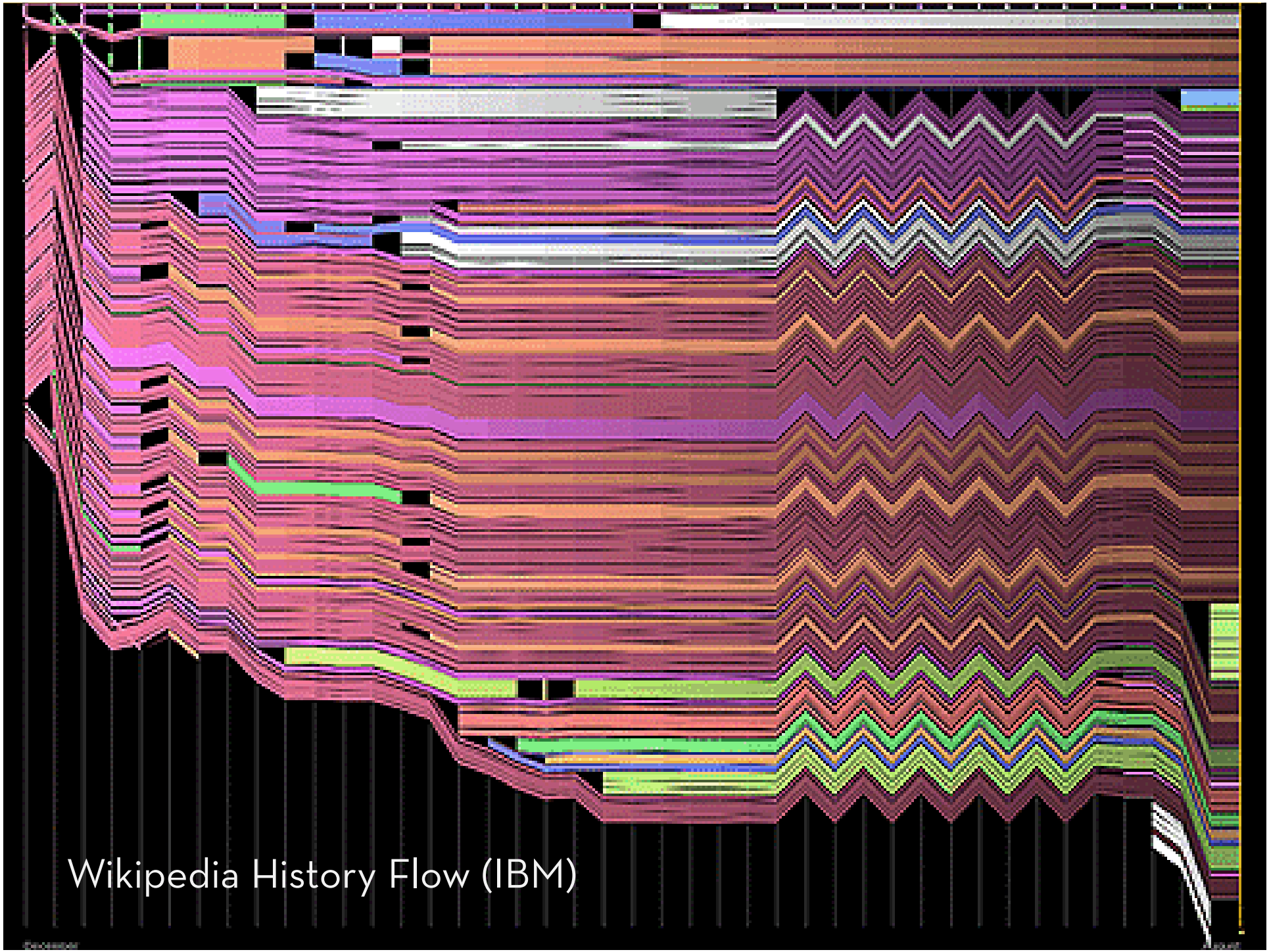


1826(?) Illiteracy in France, Pierre Charles Dupin

cabspotting.org



Wikipedia History Flow (IBM)

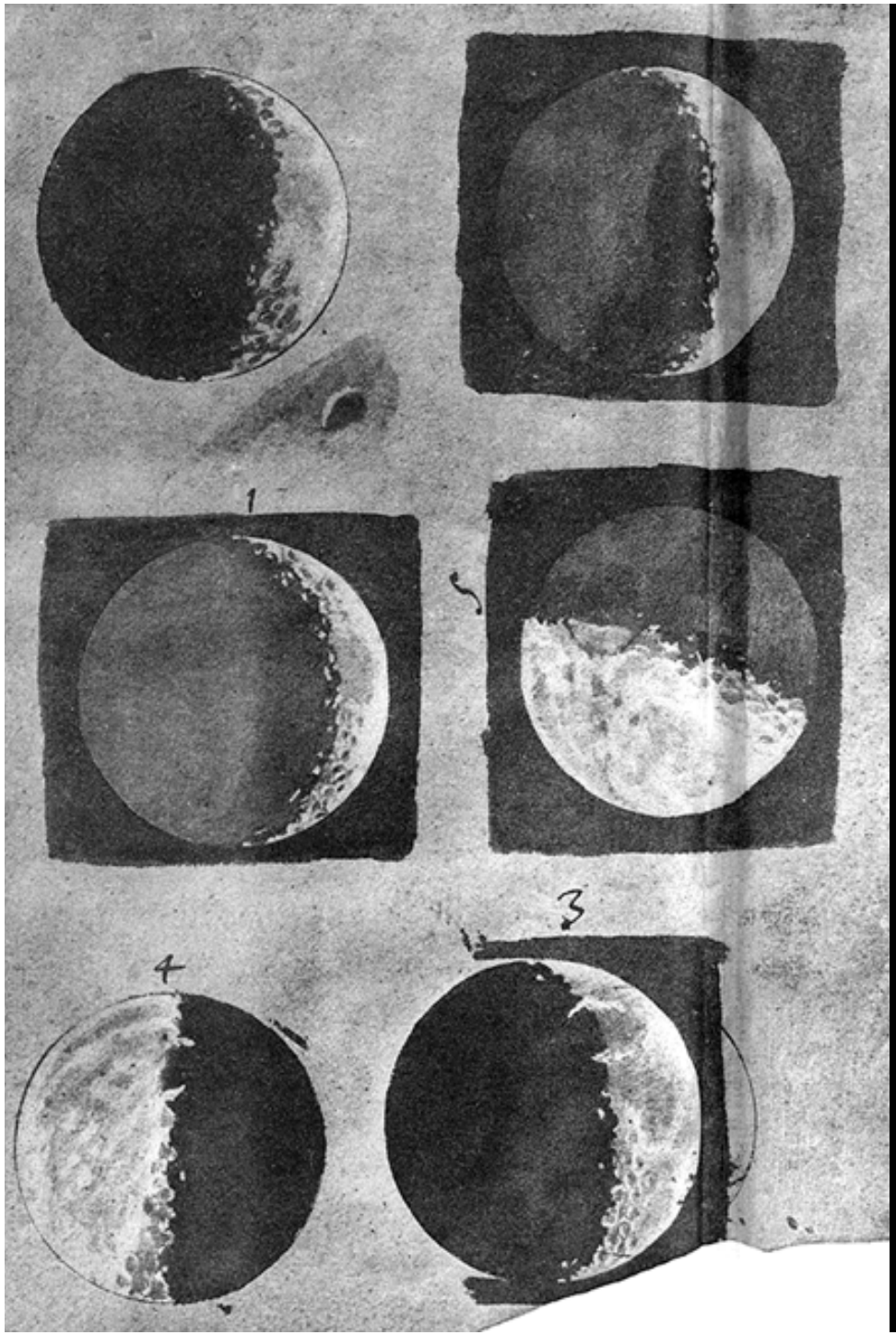

Wikipedia History Flow (IBM)

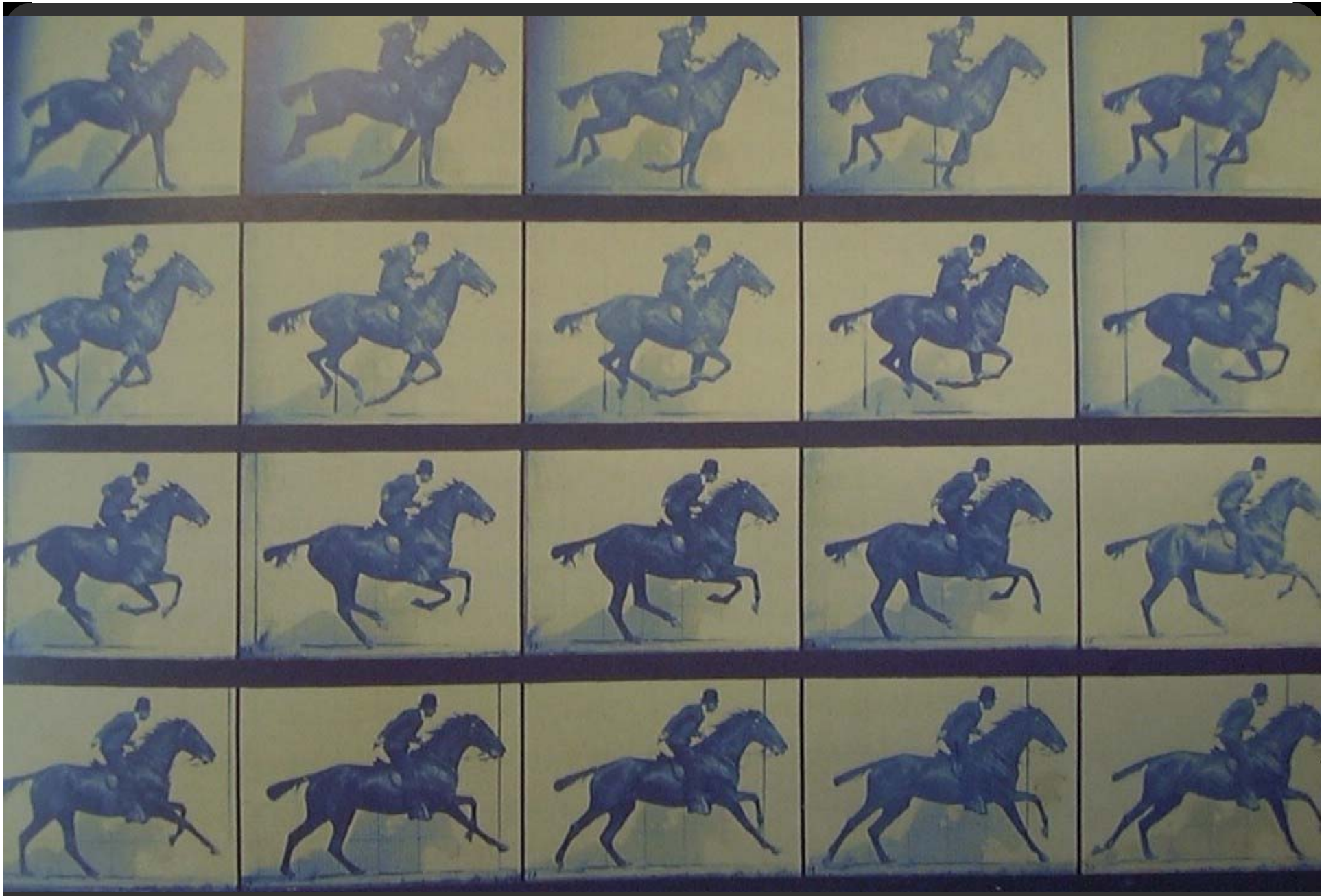
The Value of Visualization

Three functions of visualizations

Record: store information

- Photographs, blueprints, ...


1
3
4



Gallop, Bay Horse “Daisy” [Muybridge 1884-86]

Three functions of visualizations

Record: store information

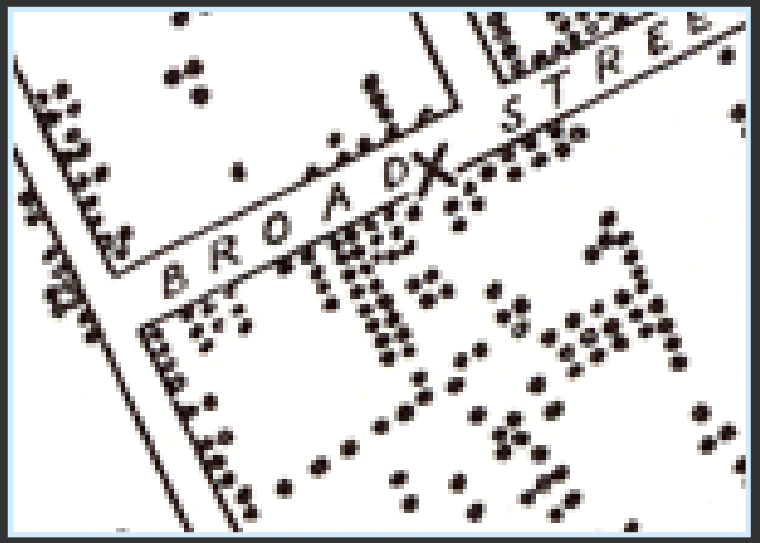
- Photographs, blueprints, ...

Analyze: support reasoning about information

- Process and calculate
- Reason about data
- Develop models and hypotheses

In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]

Cholera outbreak



Used map to hypothesize that pump on Broad St. was the cause. [from Tufte 83]

Three functions of visualizations

Record: store information

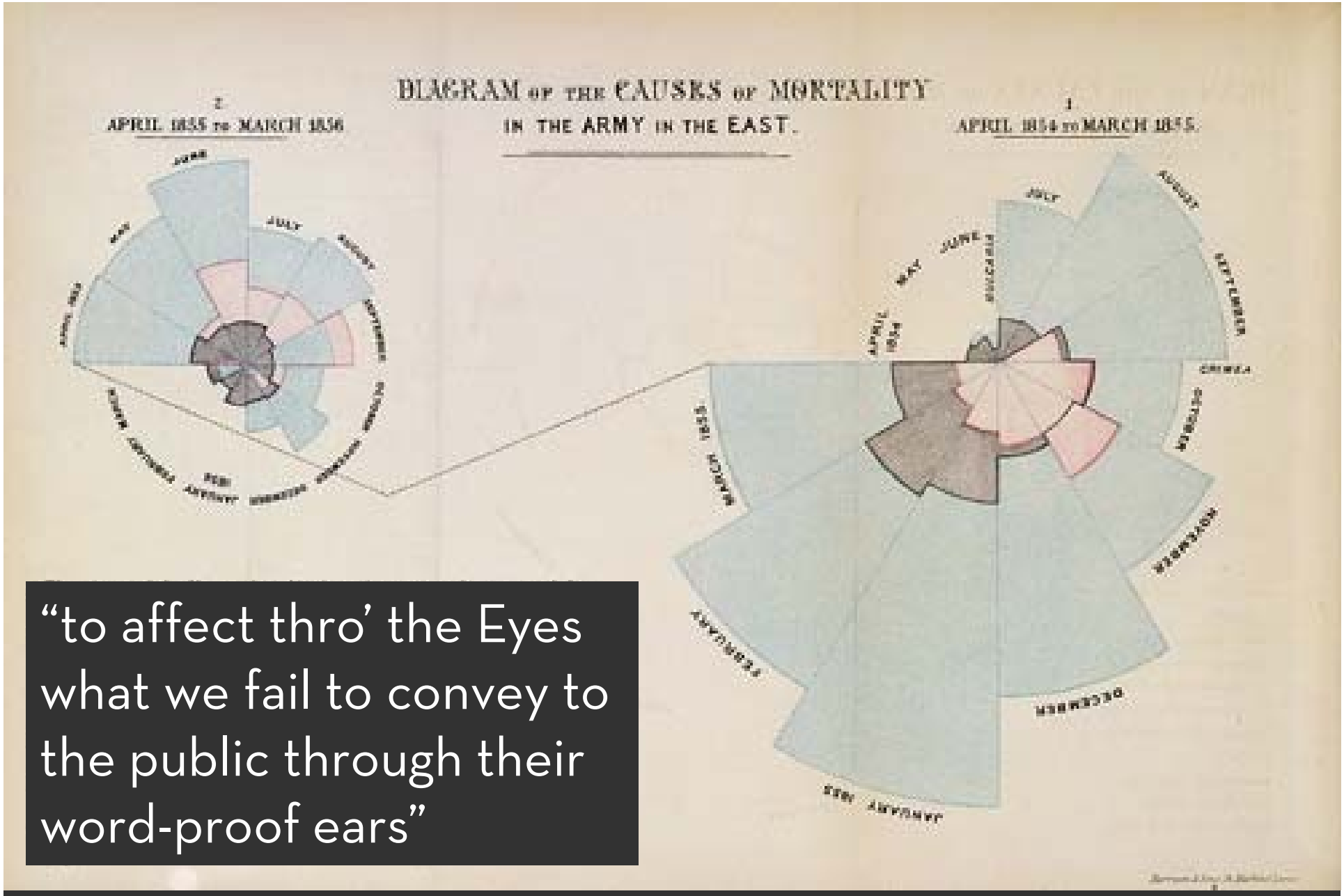
- Photographs, blueprints, ...

Analyze: support reasoning about information

- Process and calculate
- Reason about data
- Feedback and interaction

Communicate: convey information to others

- Share and persuade
- Collaborate and revise
- Emphasize important aspects of data

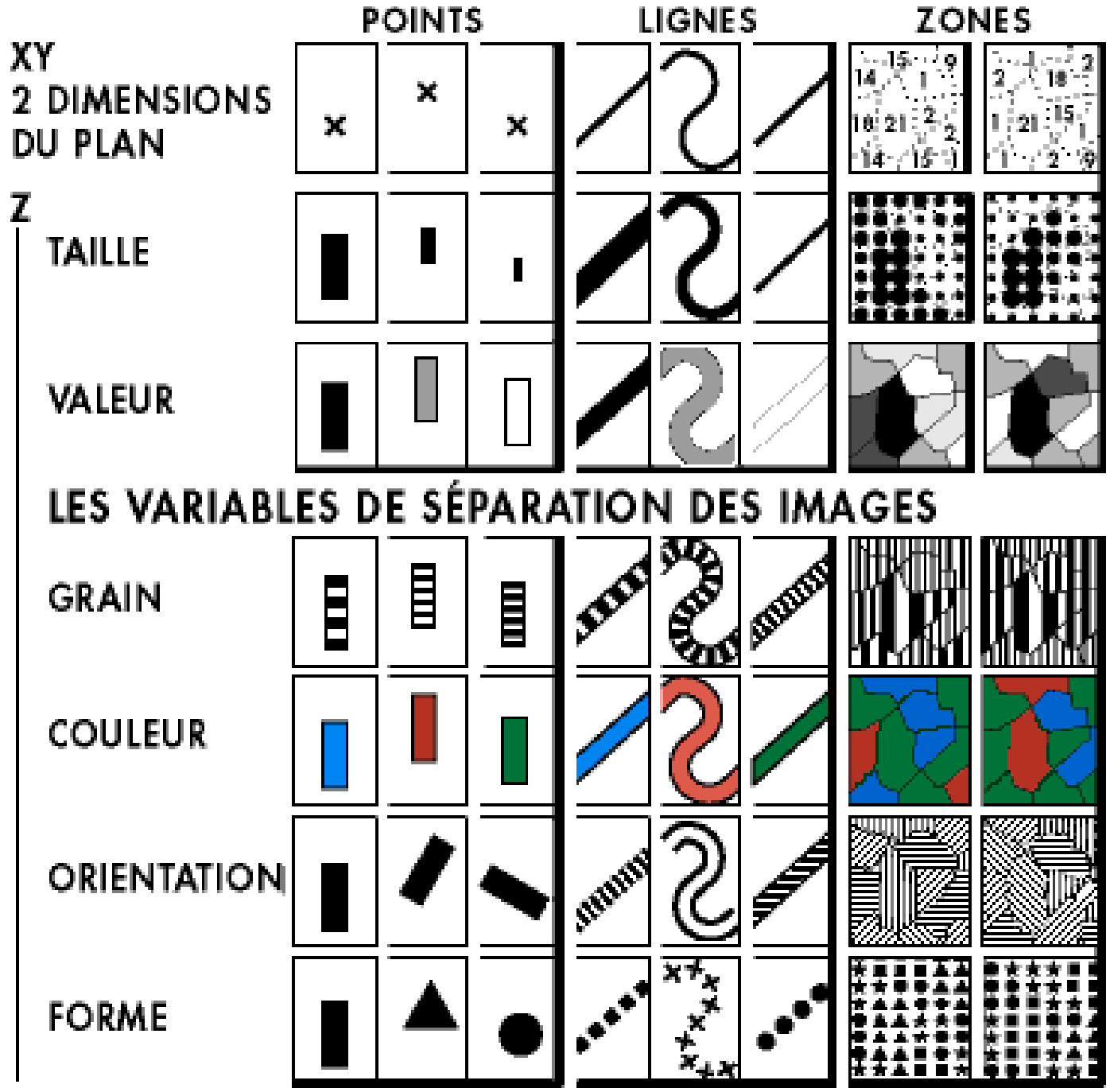


"to affect thro' the Eyes what we fail to convey to the public through their word-proof ears"

1856 "Coxcomb" of Crimean War Deaths, Florence Nightingale

The Language of Visualization

LES VARIABLES DE L'IMAGE

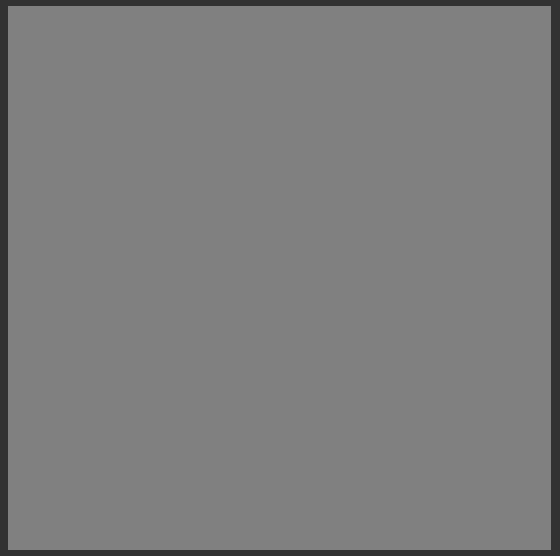


Detecting Brightness

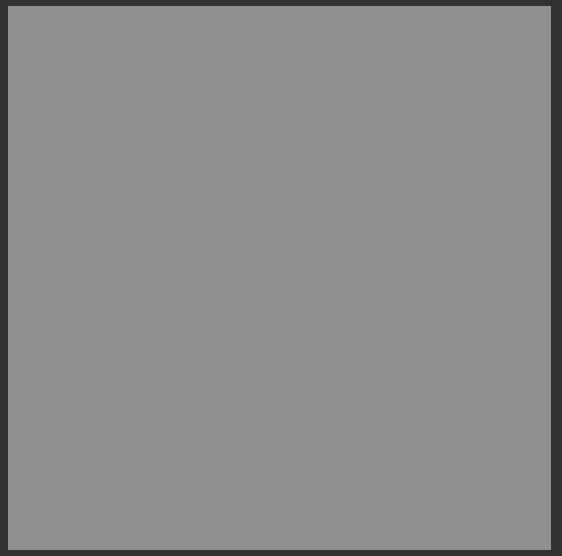
Which is brighter?

Detecting Brightness

(128, 128, 128)



(144, 144, 144)



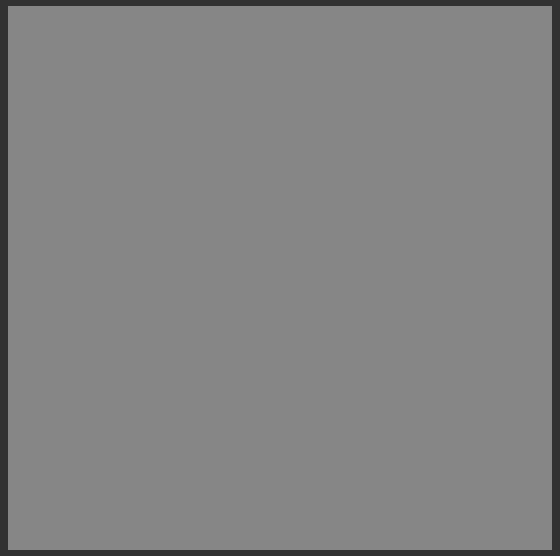
Which is brighter?

Detecting Brightness

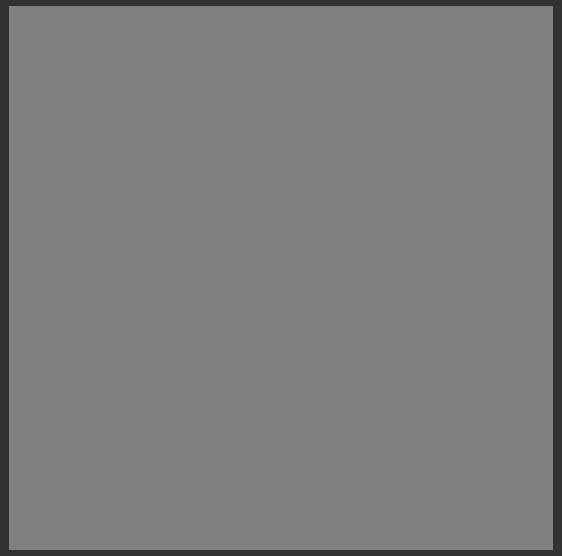
Which is brighter?

Detecting Brightness

(134, 134, 134)

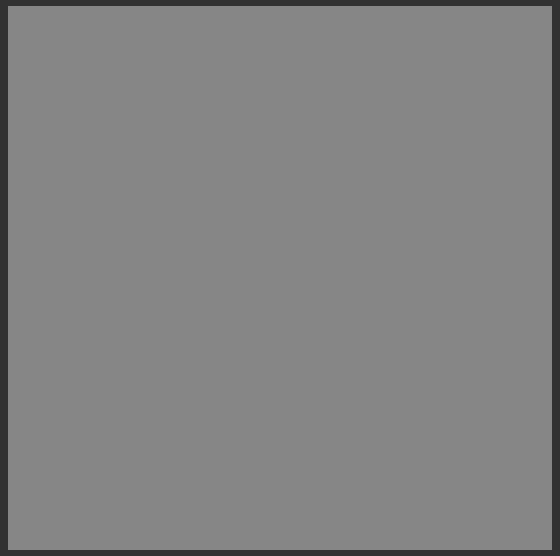


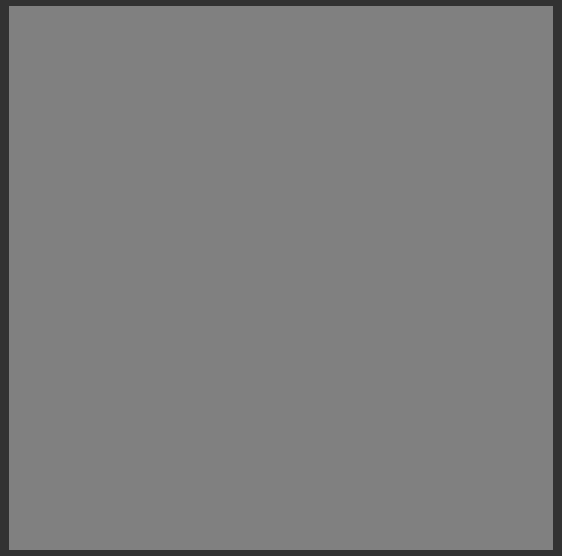
(128, 128, 128)



Which is brighter?

Just Noticeable Difference

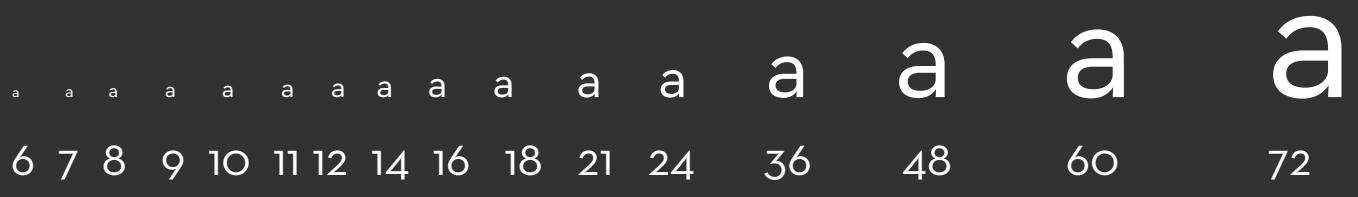




Many continuous variations are perceived in discrete steps

Steps in font size

Sizes standardized in 16th century

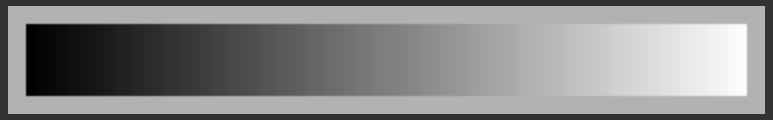


Information in color and value

Value is perceived as ordered

→ Encode ordered values

→ Encode continuous variables [but not as well]



Hue is normally perceived as unordered

→ Encode categories using color

Which best encodes quantities?

Position

Length

Area

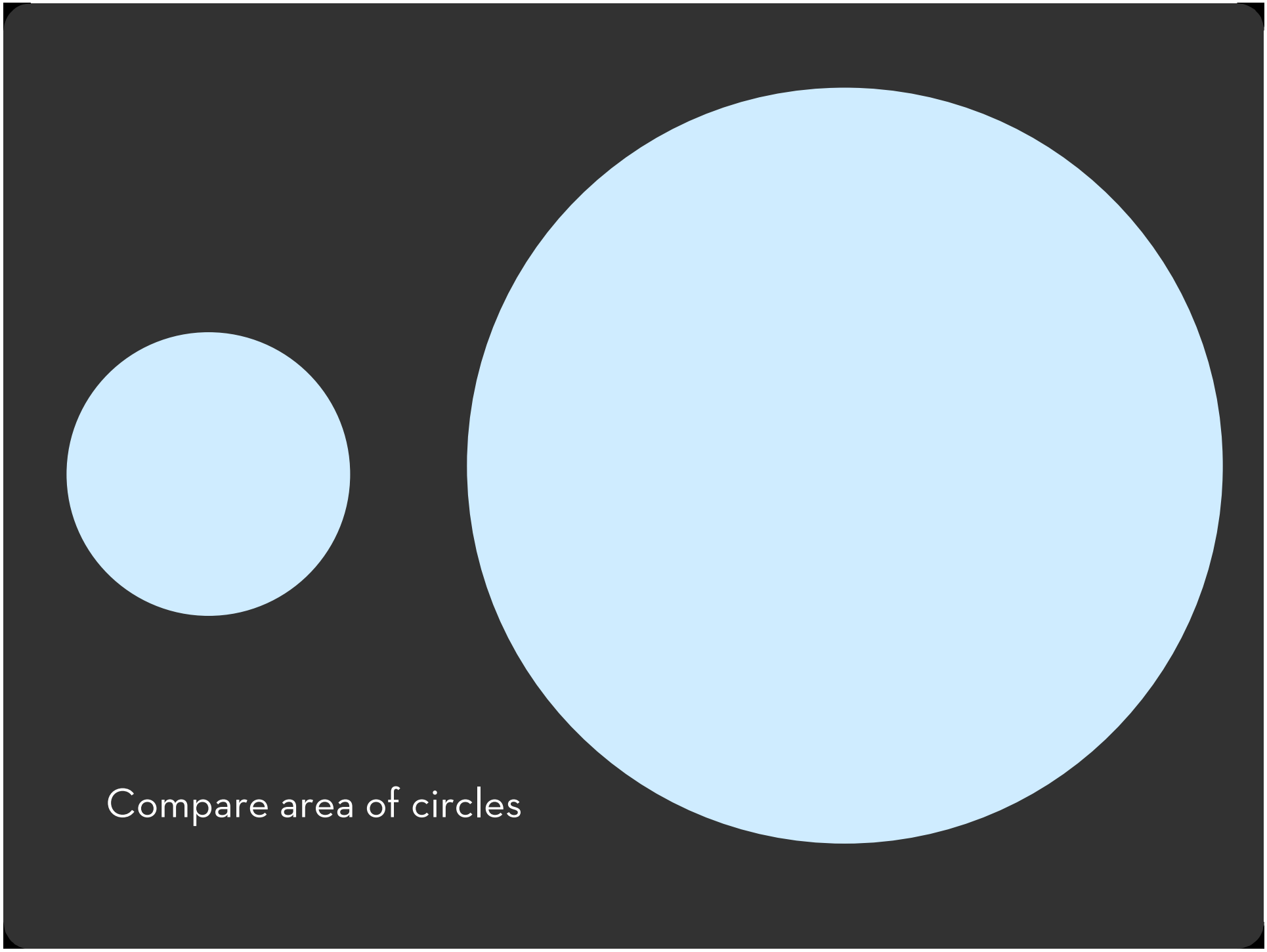
Volume

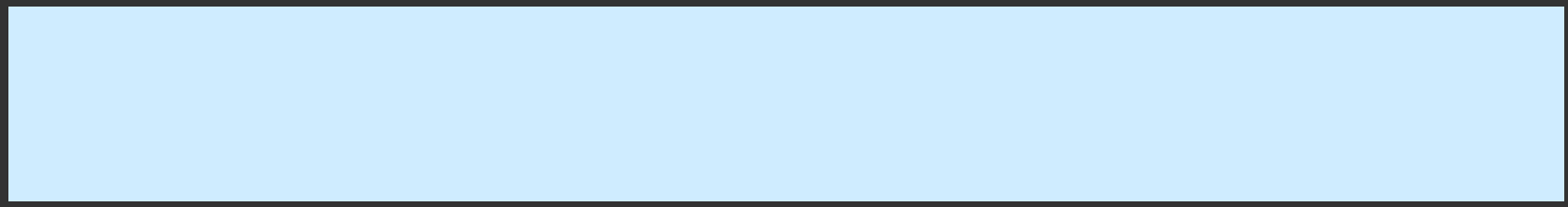
Value (Brightness)

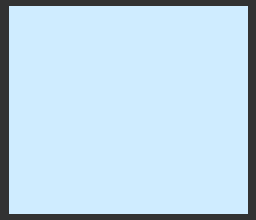
Color Hue

Orientation (Angle)

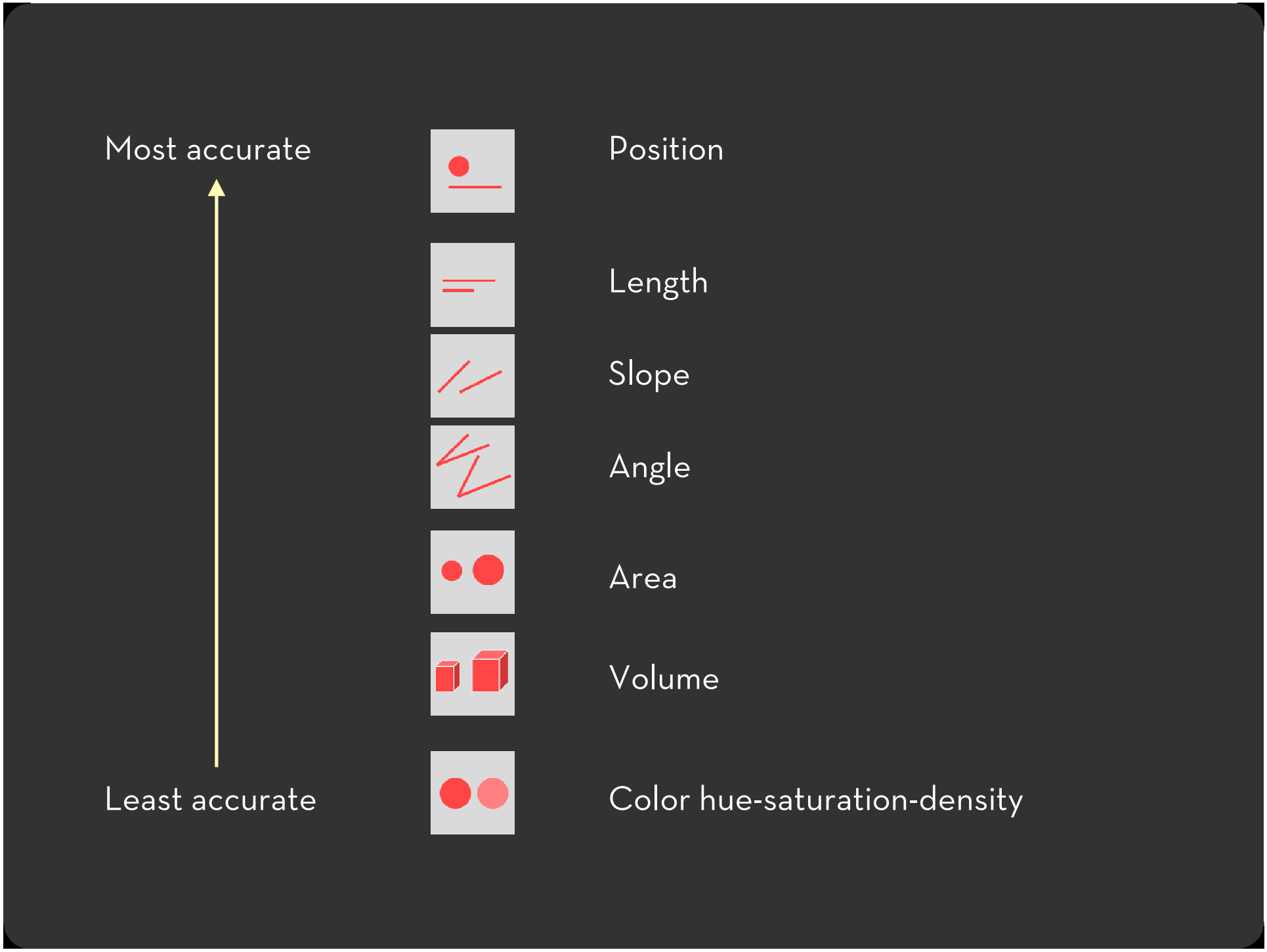
Shape


Compare area of circles



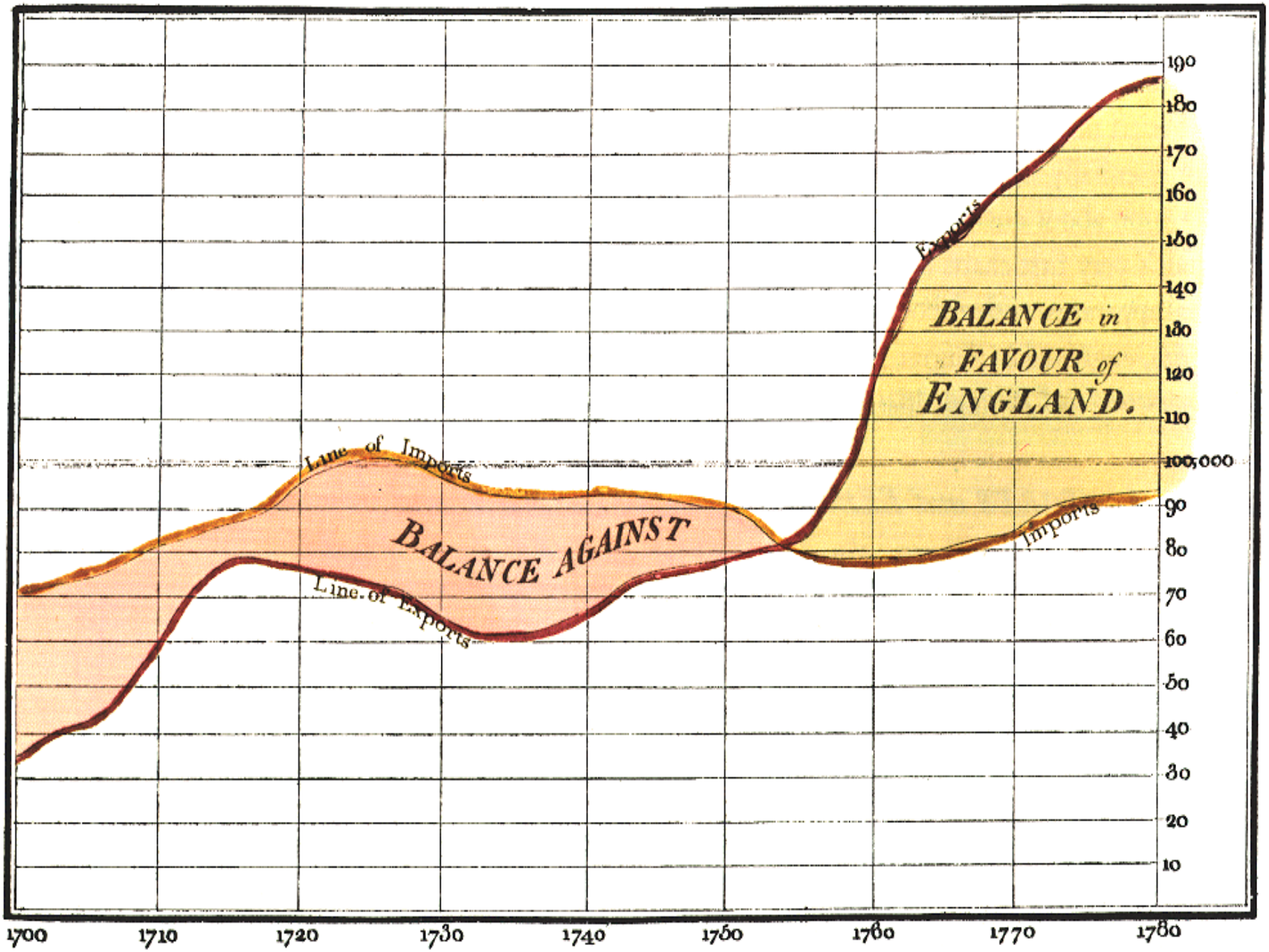


Compare length of bars

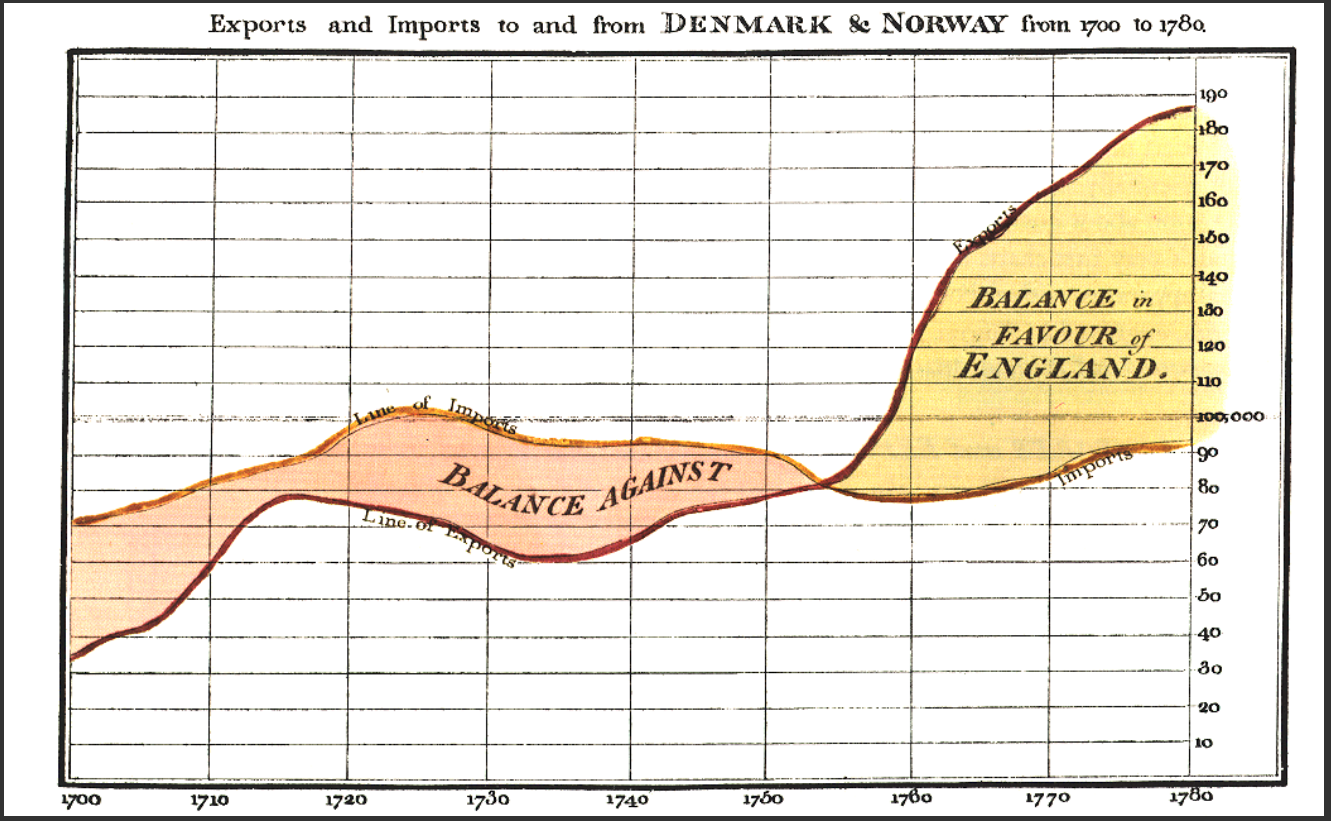

Most accurate
Position
Length
Slope
Angle
Area
Volume
Color hue-saturation-density
Least accurate

Deconstructing Visualizations

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



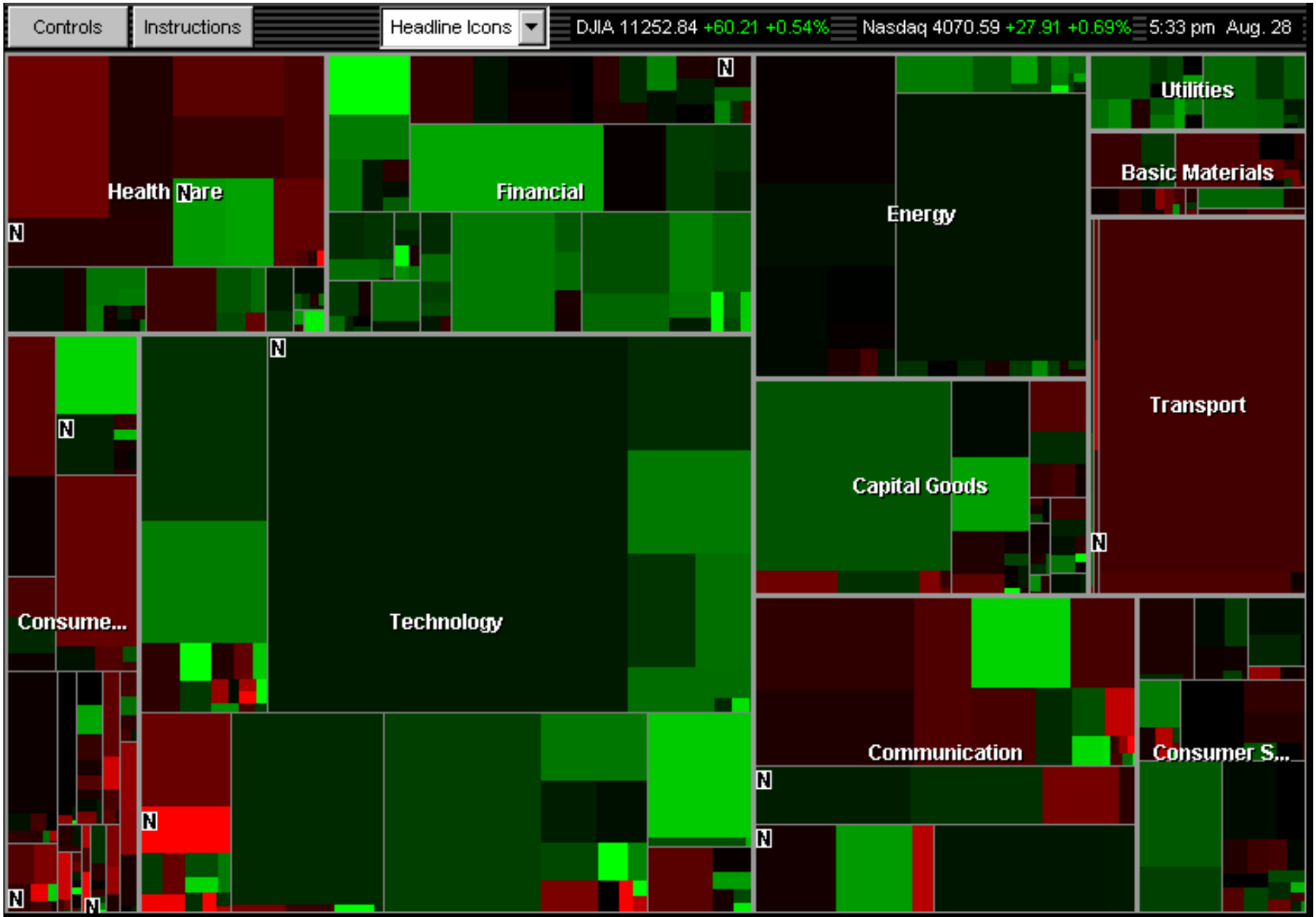
Time-Series Chart



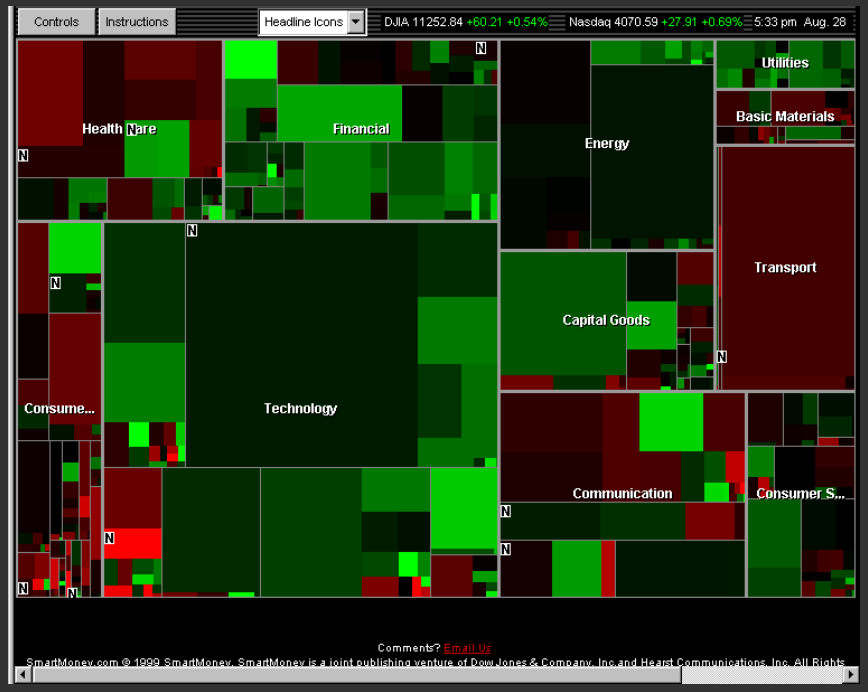
x-axis position: year

y-axis position: currency

color: imports/exports ratio


Controls
Instructions
Headline Icons
DJIA 11252.84 +60.21 +0.54%
Nasdaq 4070.59 +27.91 +0.69%
5:33 pm Aug. 28
Health Care
Financial
Energy
Utilities
Basic Materials
Transport
Capital Goods
Consume...
Technology
Communication
Consumer S...
http://www.smartmoney.com/marketmap/

Treemap of the Stock Market



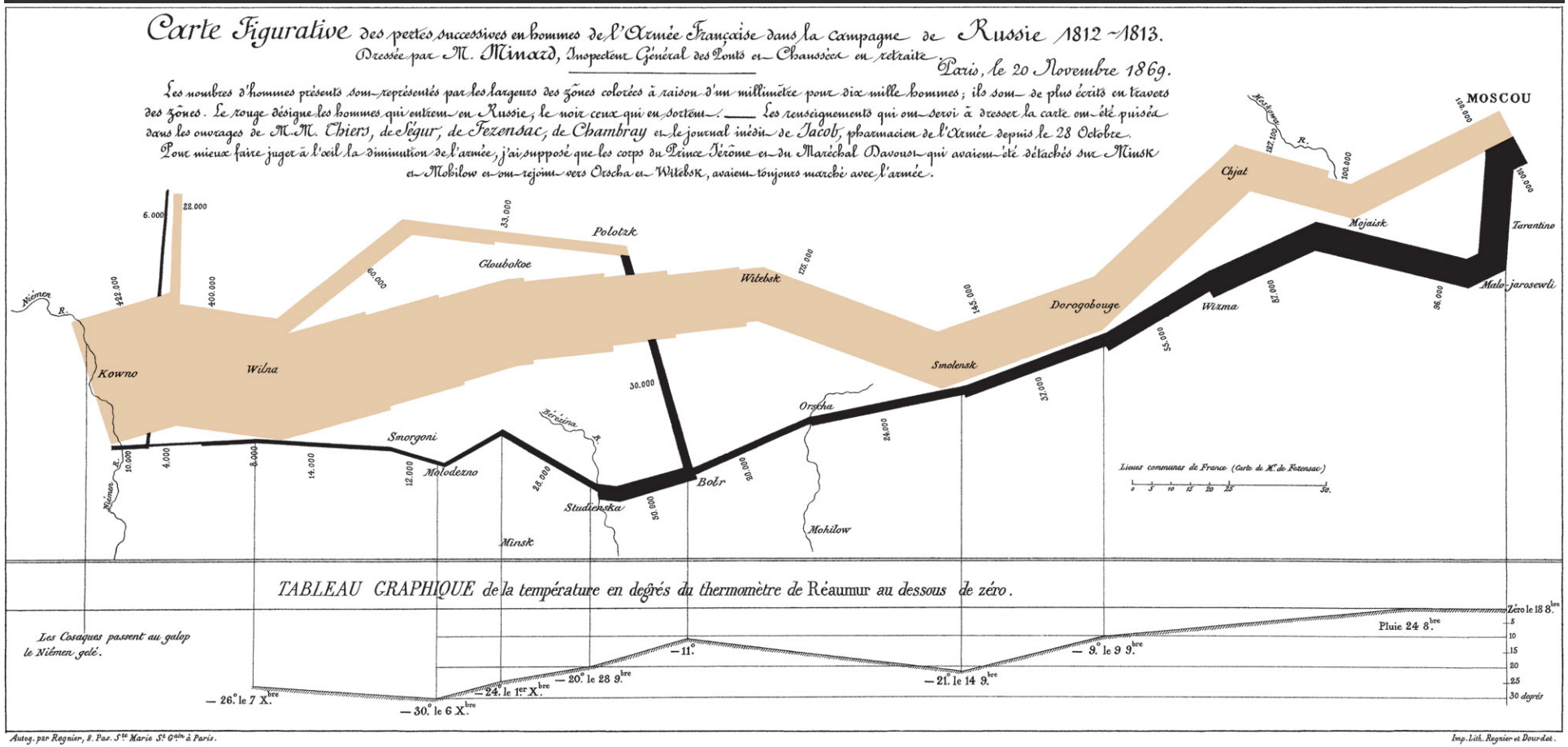
rectangle area: market capitalization

rectangle position: market sector, market cap

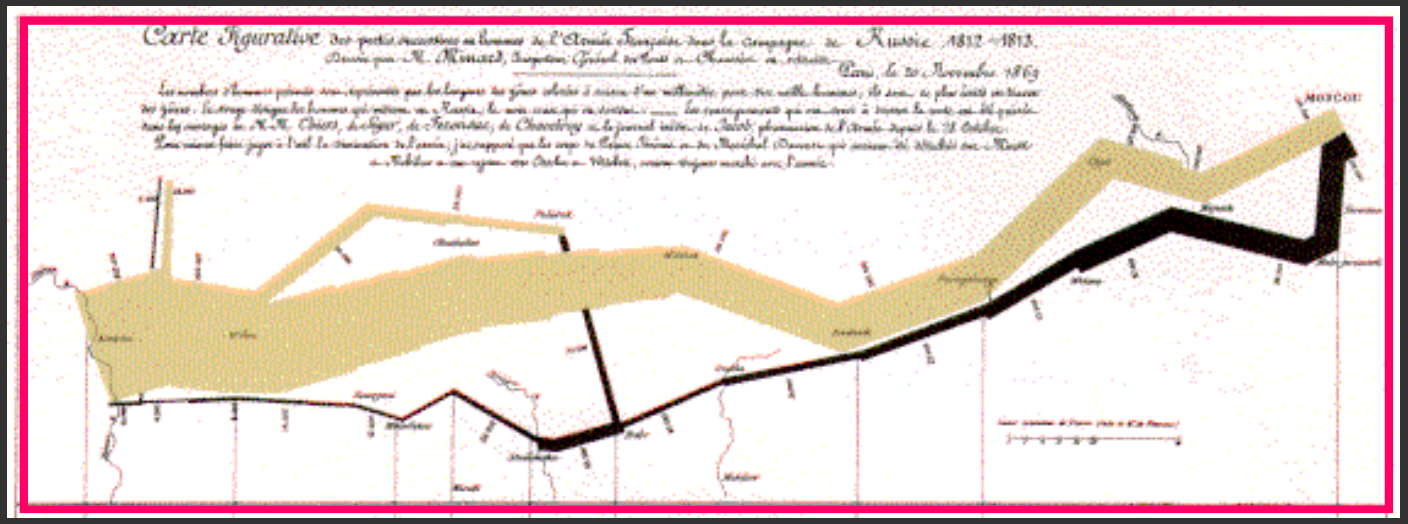
color hue: loss vs. gain

color value: magnitude of loss or gain

Minard 1869: Napoleon's march

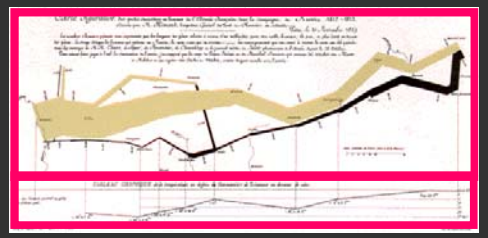


Single axis composition



+

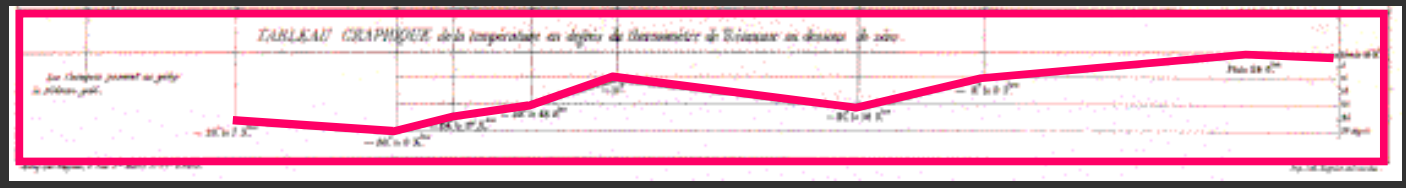
=



Mark composition

y-axis: temperature

\+ x-axis: longitude / time

\= 

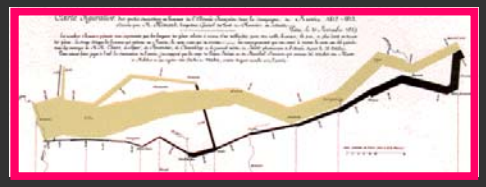
temperature over space/time

Mark composition

y-axis: longitude

\+ x-axis: latitude

\+ width: army size

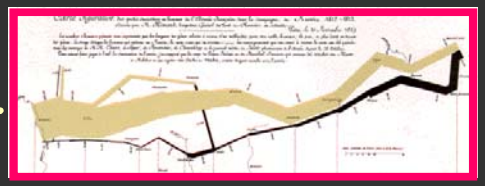
\= 

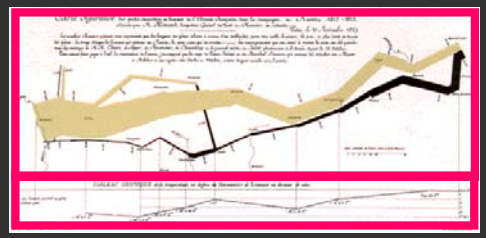
army position (lat/lon) and army size (width)

longitude

latitude

army size

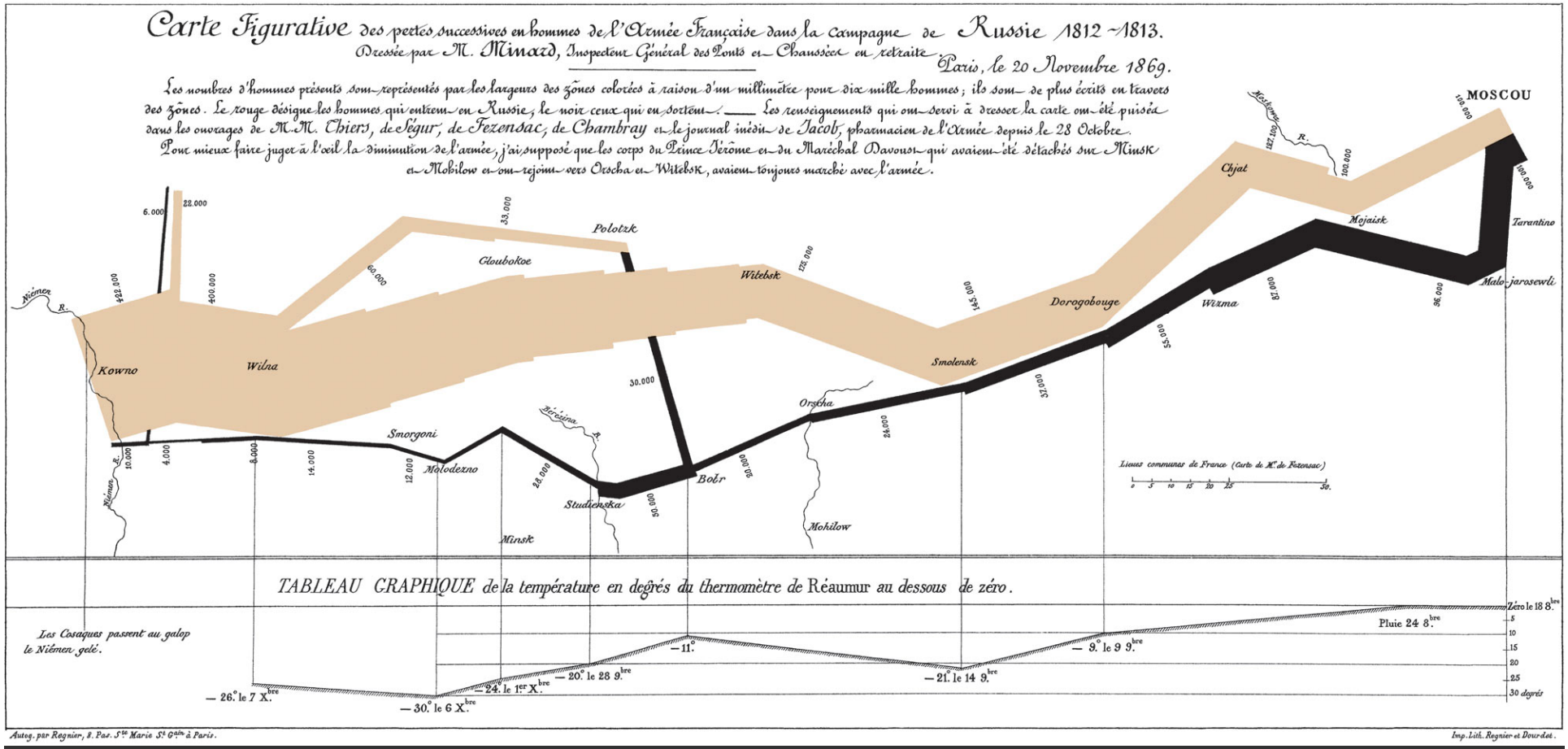




temperature

latitude / time

Minard 1869: Napoleon's march



Depicts at least 5 quantitative variables. Any others?

Summary

Effective visualizations can help us **record**, **analyze**, and **communicate** data

Creating visualizations requires **modeling data** and applying **appropriate visual encodings** and **interaction techniques**.

As you encounter visualizations in the world, try to **deconstruct what you see**. What are the underlying data and visual encodings?