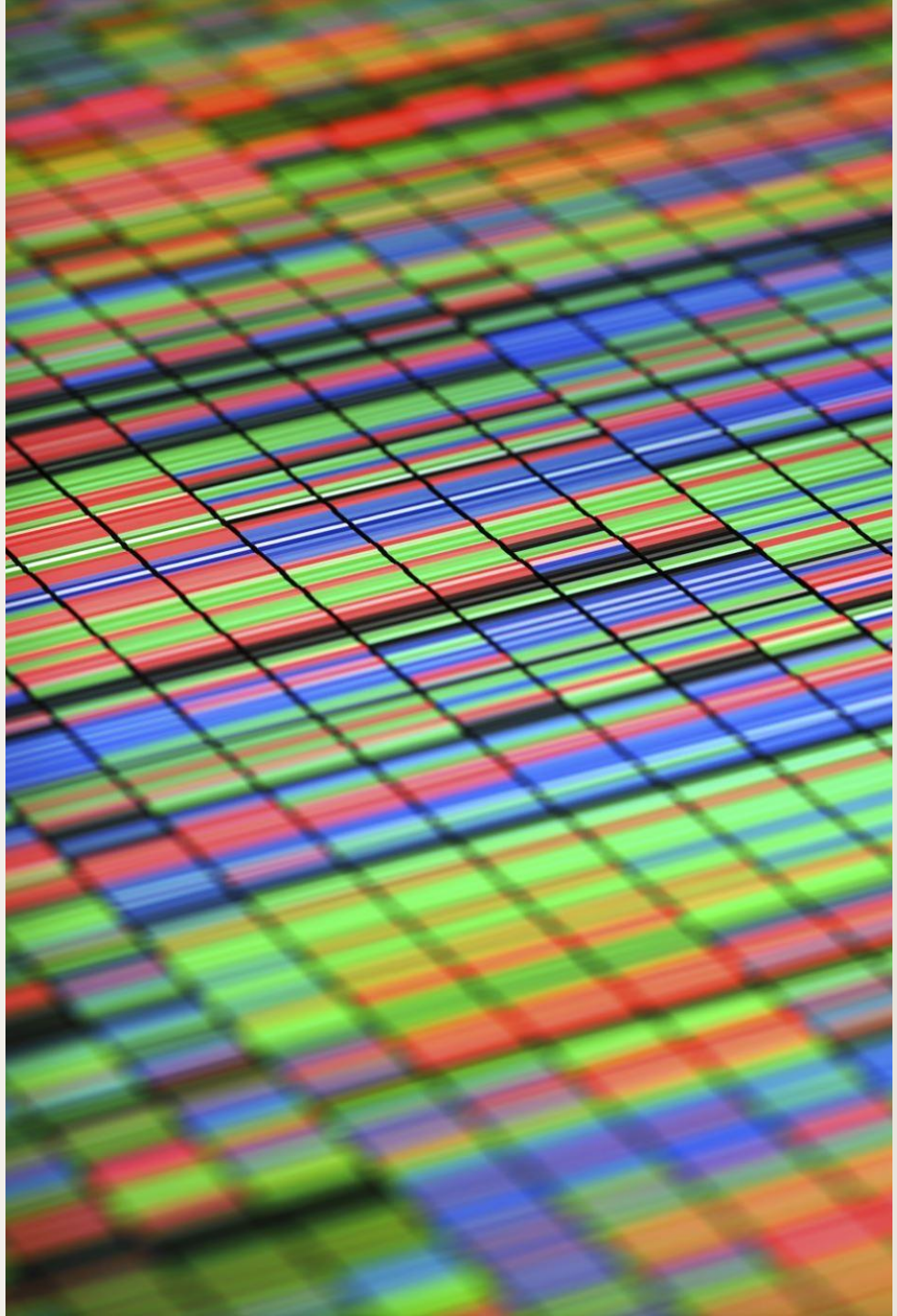


Data Visualization

What is data visualization?

1967

"… finding the artificial memory that best supports our natural means of perception" [Bertin 1967]

1999

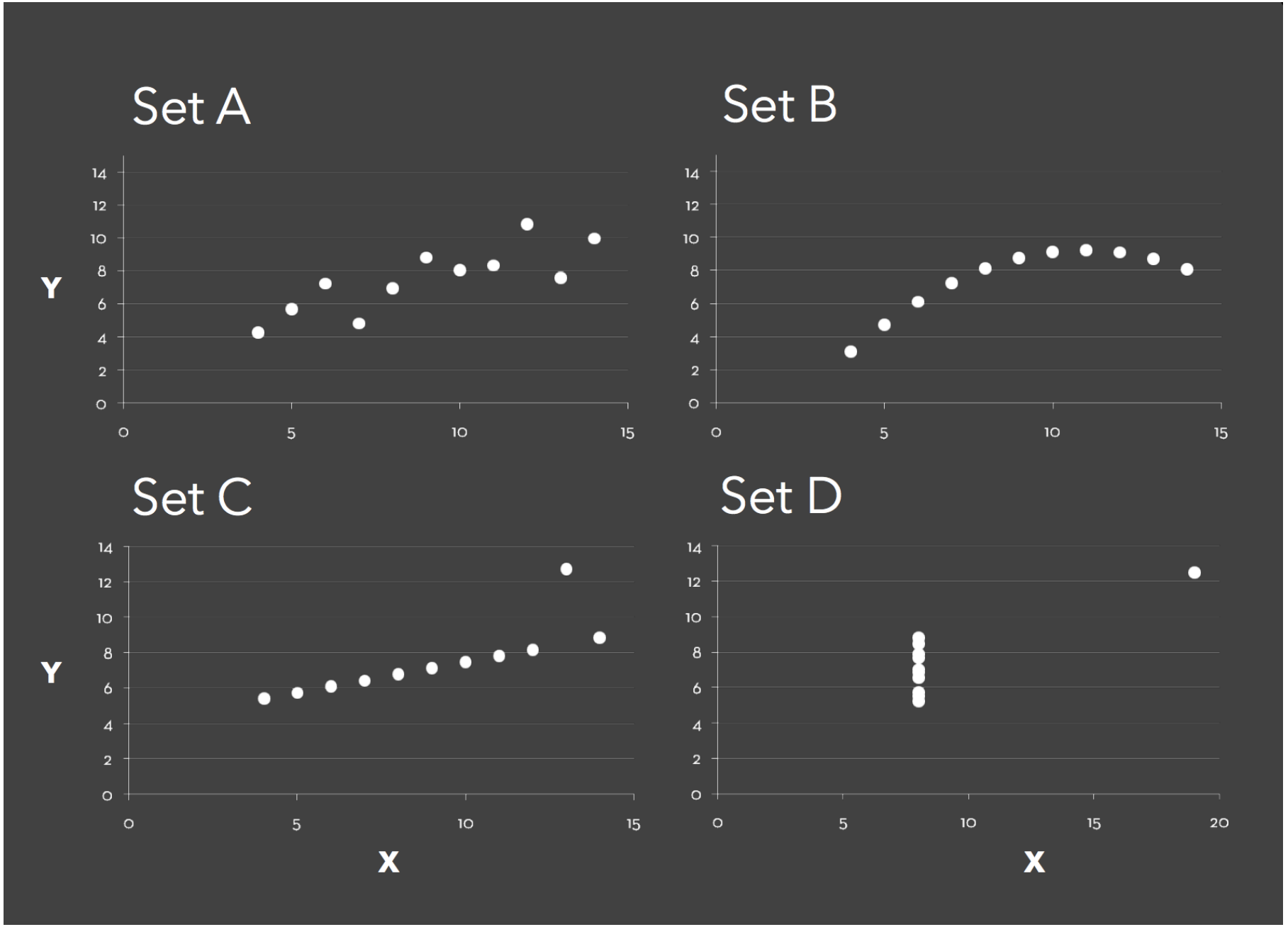
"The use of computer-generated, interactive, visual representations of data to amplify cognition" [Card, Mackinlay, & Schneiderman 1999]

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

Mean X: 9.0, Standard dev X: 3.317
Mean Y: 7.5, Standard dev Y: 2.03

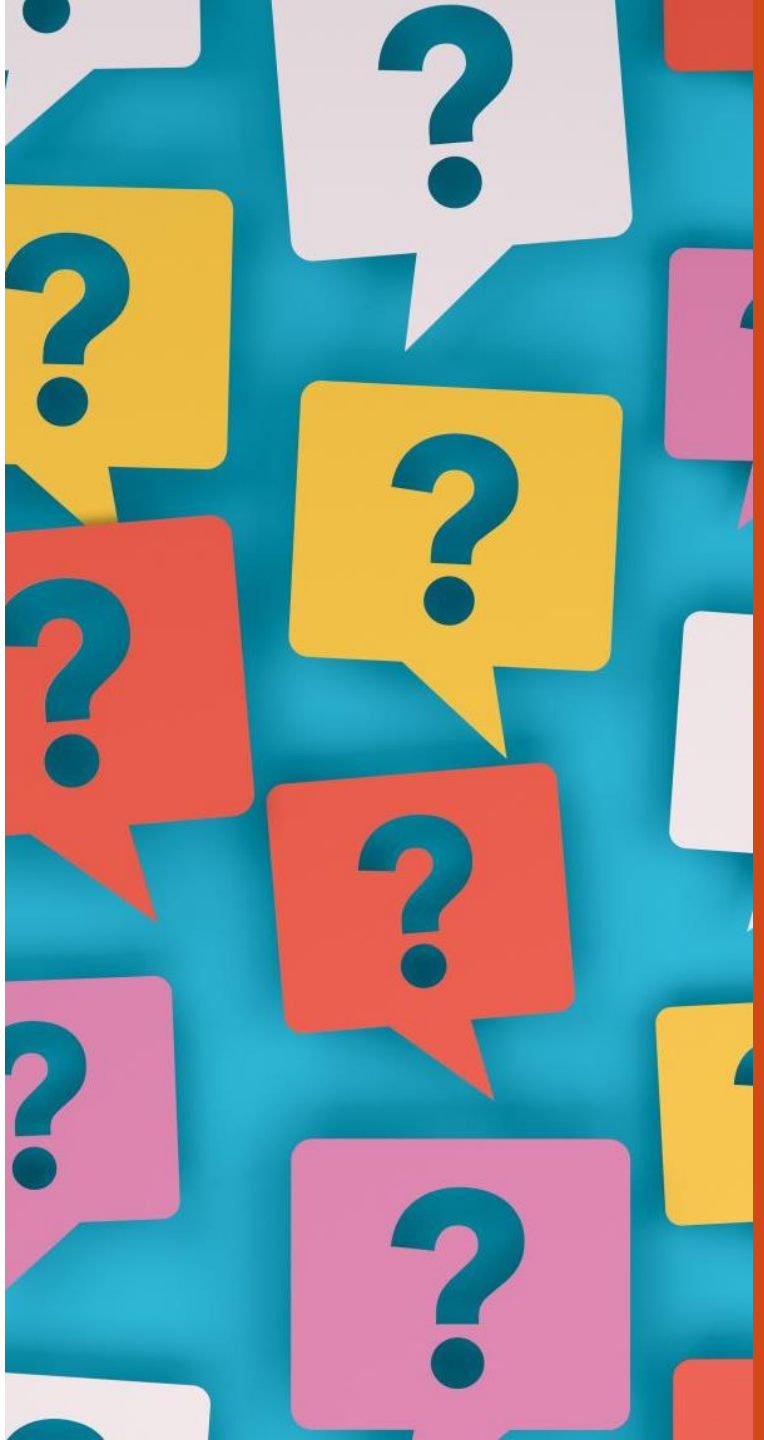
Linear Regression:
$Y = 3 + 0.5X$

[Anscombe 1973]


Set A
Y
14
12
10
8
6
4
2
0
0
5
10
15
Set B
14
12
10
8
6
4
2
0
0
5
10
15
Set C
Y
14
12
10
8
6
4
2
0
0
5
10
15
X
Set D
14
12
10
8
6
4
2
0
0
5
10
15
20
X

Why make data visualizations?

- Find patterns
- Answer questions and spur new ones
- Support memory
- Make a decision
- Present an argument
- Engage & inspire your audience



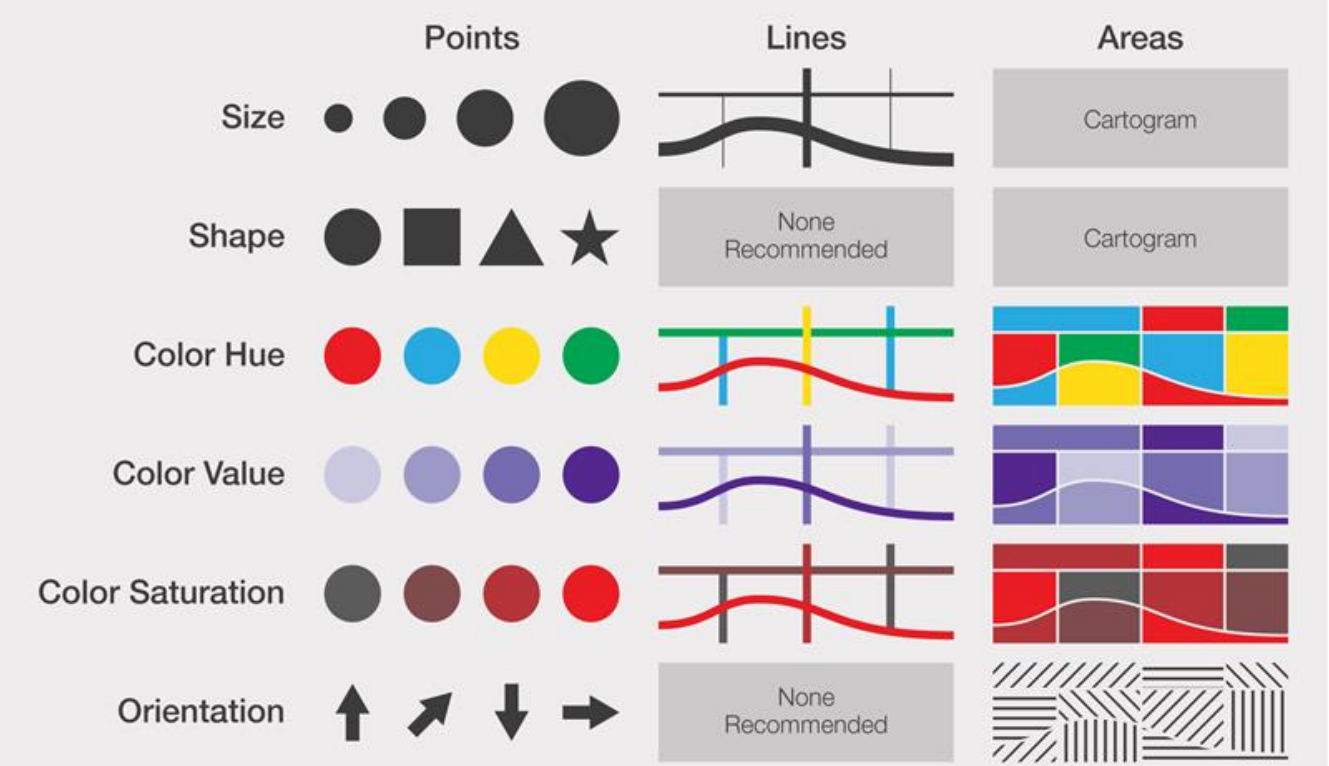
Why study data visualizations?

- The world is creating TONS of data
  - > 90% of data in the world was created in last years(!)
  - 2.5 quintillion (2,500,000,000,000,000,000) bytes of data are created every day
- Relevant for every discipline
- Become informed consumers
- They're super cool & fun to make!

Data Types

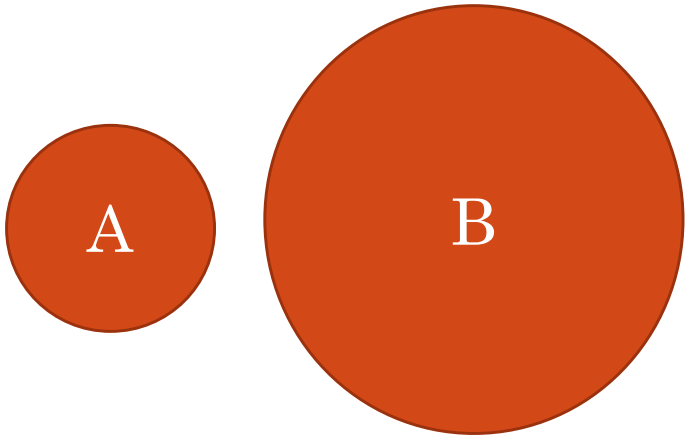
- Nominal
  - Labels or categories
  - Fruits: apples, oranges, kiwi, …
- Ordinal
  - Ordered
  - Apples: US Utility, US No 1, US Fancy, US Extra Fancy
- Quantitative
  - Quantity, amount, or range
  - Length, temperature, location (lat/long), dates

Visual Encodings

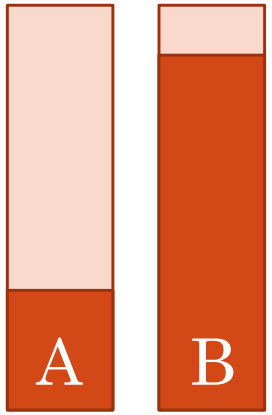


Intuition

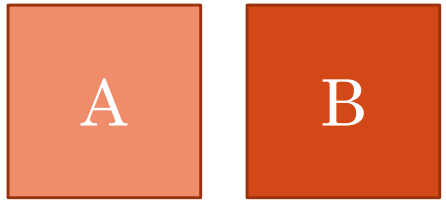
Discuss with your neighbors – How much is A compared to B?



A = 25% B



A = 33% B

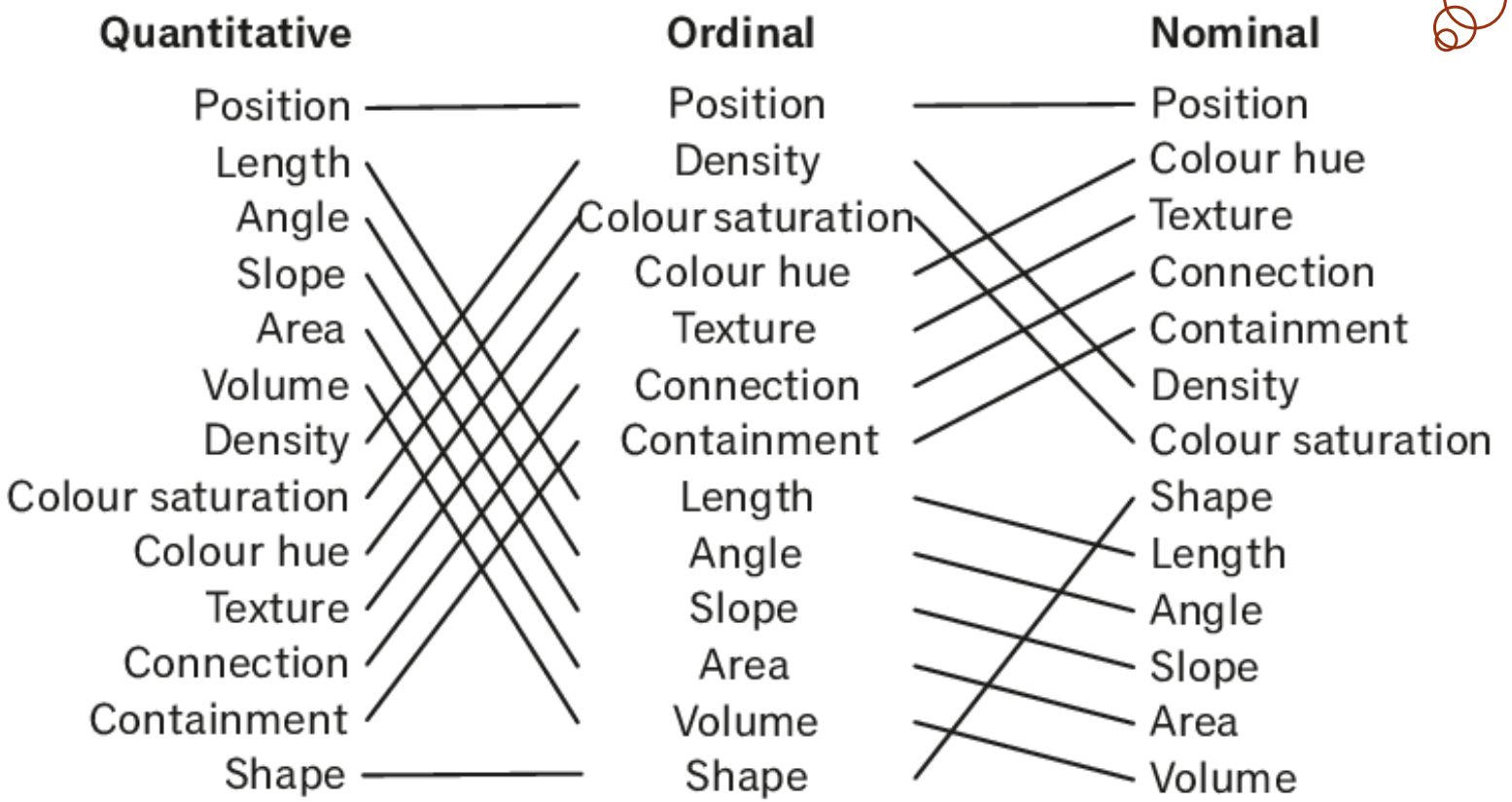


A = 60% B

Mackinlay's ranking

Conjectured effectiveness of encodings by data type.

Remember!
Quantitative = amount
Ordinal = ordered
Nominal = label

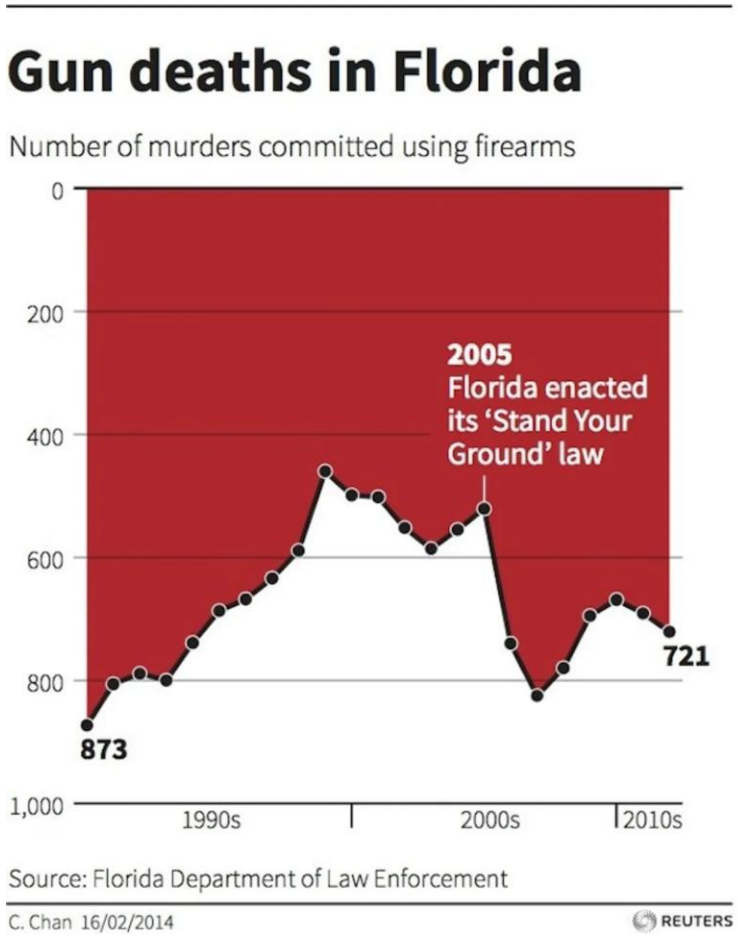


[Mackinlay 1986]

Design criteria

- Tell the whole truth & nothing but the truth
  - Consider the impact of your visualization
- Use encodings that people decode better
  - Better = faster and/or more accurately
  - Think about accessibility!
- Encode the most important information using the most effective encodings

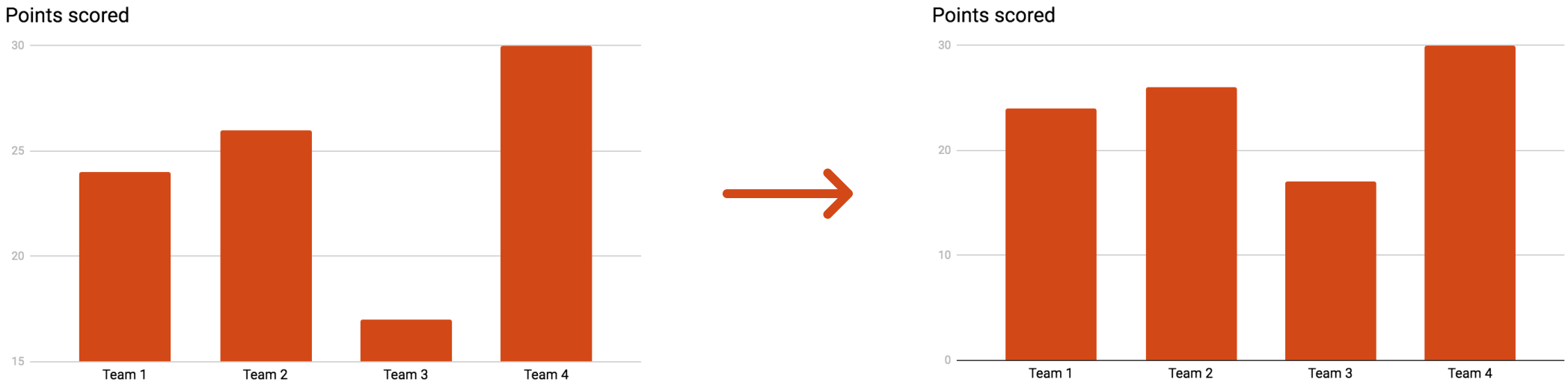
Misleading Visualization Practices



After 2005, did gun deaths in Florida increase or decrease?

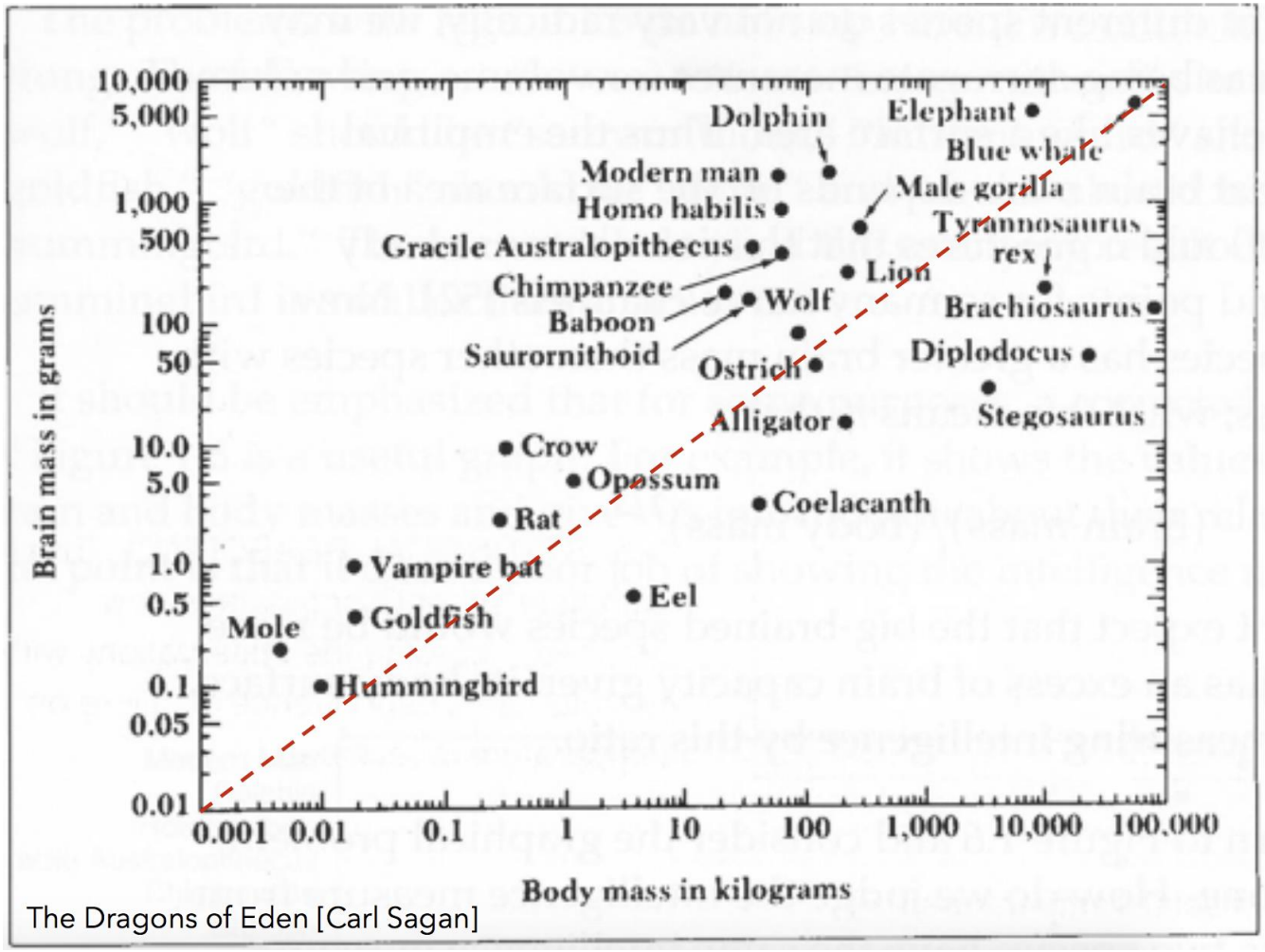
Misleading Visualization Practices

This is the same data!



Team 2 got 26 points – how many points did Team 3 get?

Misleading Visualization Practices

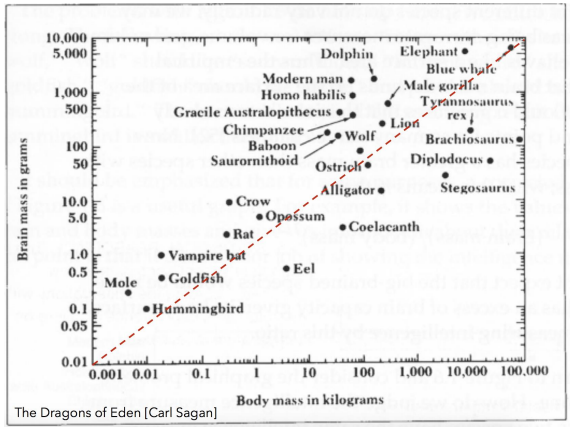


The Dragons of Eden [Carl Sagan]

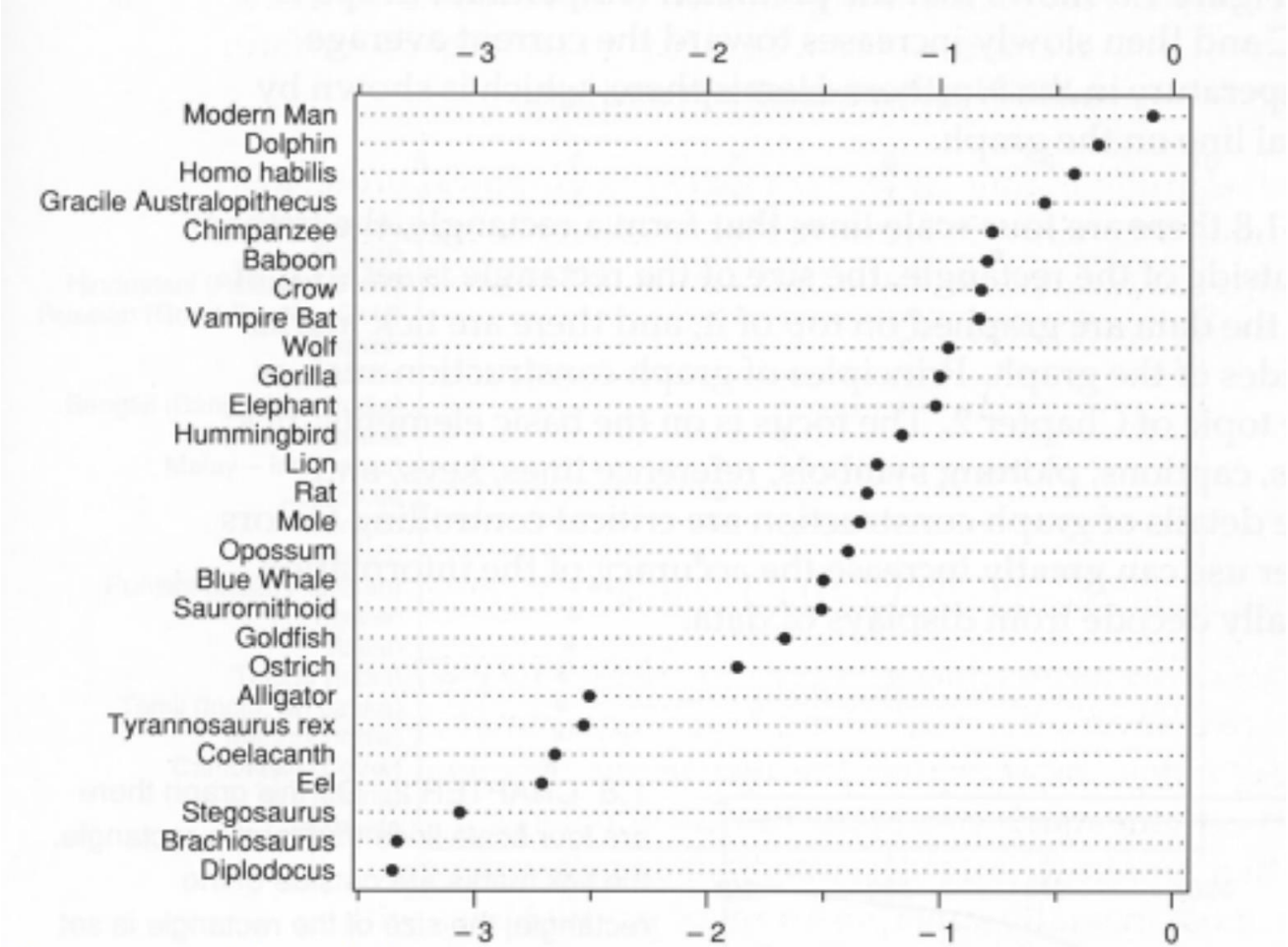
Which two animals have the highest brain mass to body mass ratio?

Misleading Visualization Practices

This is the same data!



The Dragons of Eden [Carl Sagan]



The Elements of Graphing Data [Cleveland]

Goals of visualization research

1. **Understand** how visualizations convey information

   What do people perceive/comprehend?

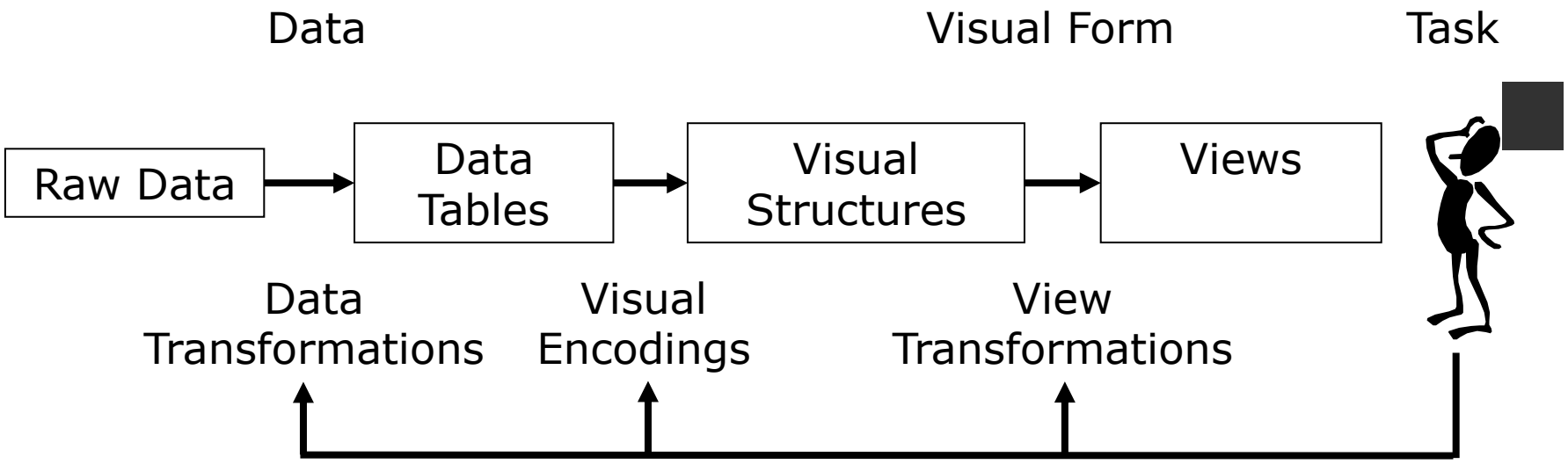
   How do visualizations correspond with mental models?

2. **Develop principles and techniques** for creating effective visualizations and supporting analysis

   Amplify perception and cognition

   Strengthen tie between visualization and mental models

Visualization Reference Model



Which best encodes quantities?

Position

Length

Area

Volume

Value (Brightness)

Color Hue

Orientation (Angle)

Shape

Detecting Brightness

Which is brighter?

Detecting Brightness

(128, 128, 128)

(144, 144, 144)

Which is brighter?

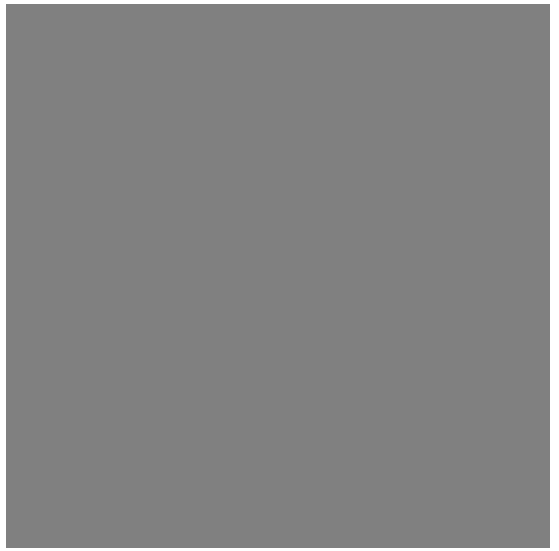
Detecting Brightness

Which is brighter?

Detecting Brightness

(134, 134, 134)

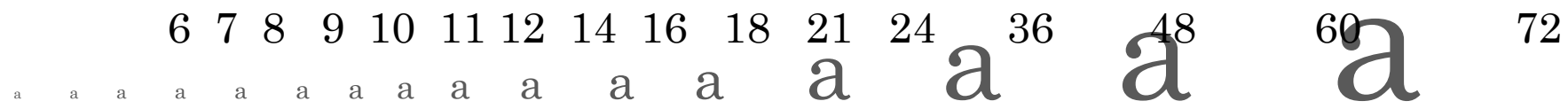
(128, 128, 128)



Which is brighter?

Steps in font size

Sizes standardized in 16th century



Information in color and value

Value is perceived as ordered

∴ Encode ordinal variables (O)

∴ Encode continuous variables (Q) [not as well]

Hue is normally perceived as unordered

∴ Encode nominal variables (N) using color

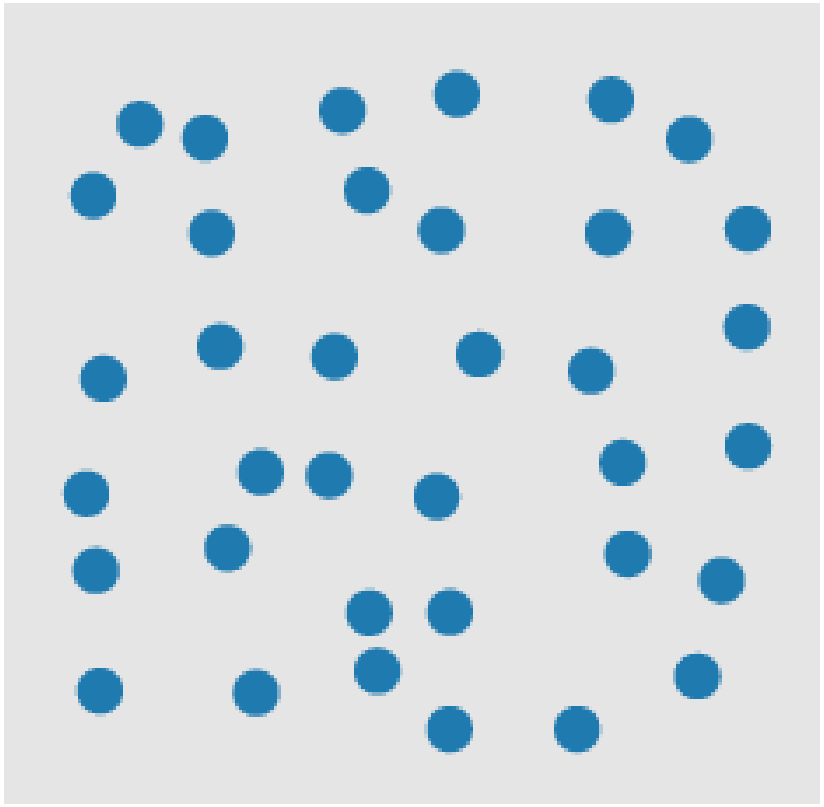
How many 3's

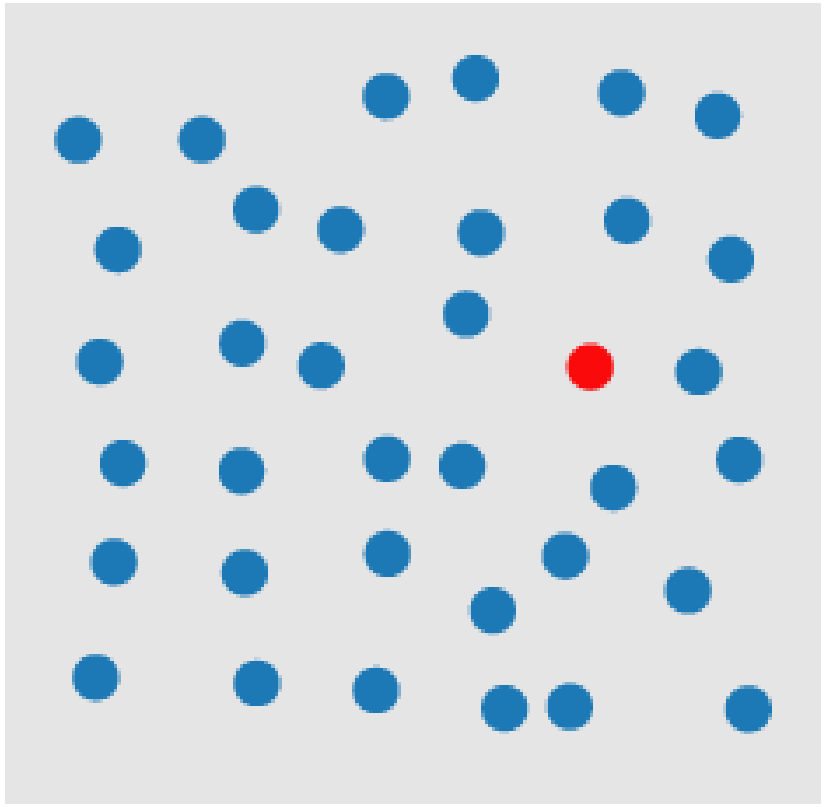
128176875613897654698450698560498282676298098584582245098564589450984509809435859091030209905959595772564675050678904567884578980982167765487636490856091294968 6

How many 3's

128176875613897654698450698560498282676298098584582245098564589450984509809435859091030209905959595772564675050678904567884578980982167765487636490856091294968 6

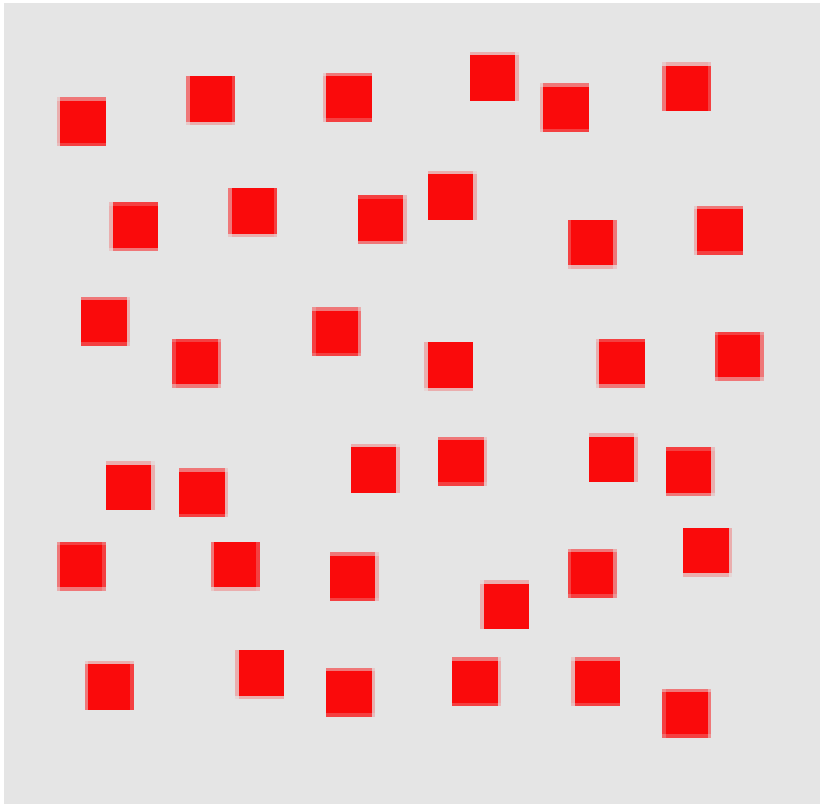
Visual pop-out: Color

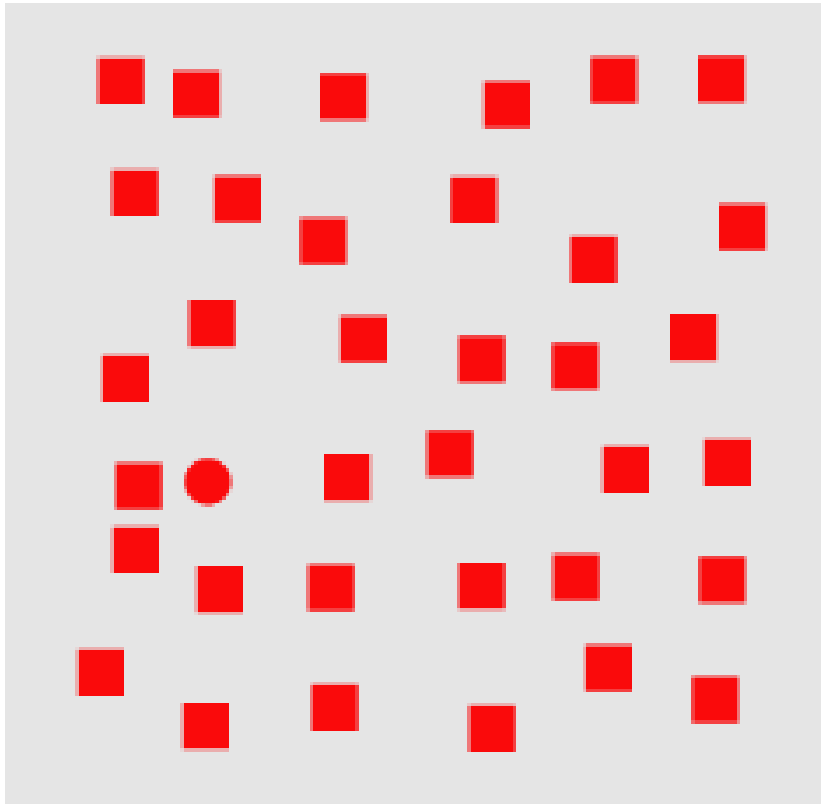




http://www.csc.ncsu.edu/faculty/healey/PP/index.html

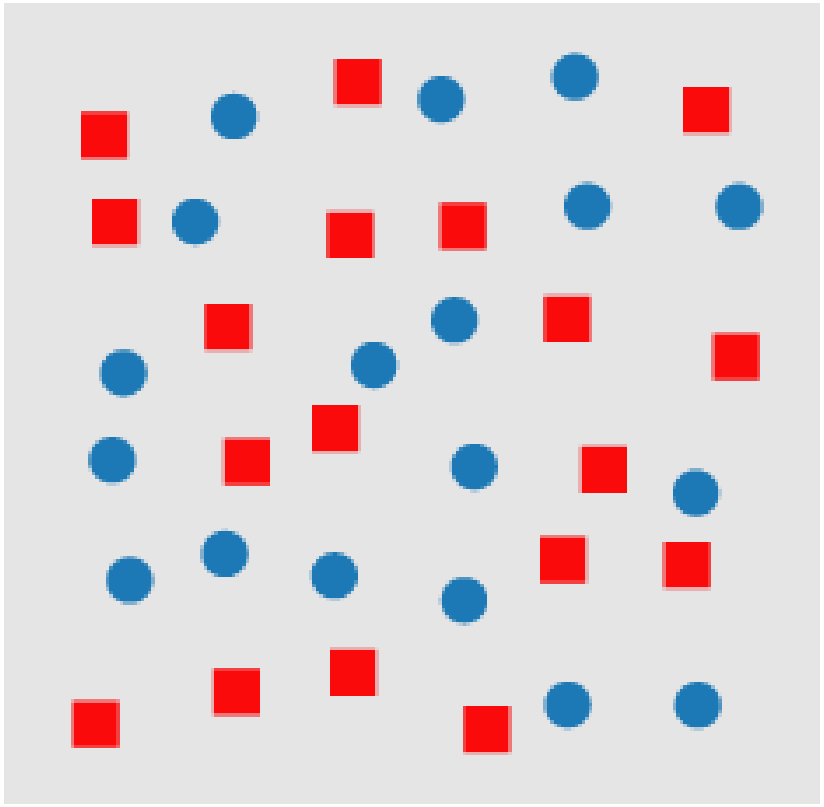
Visual pop-out: Shape

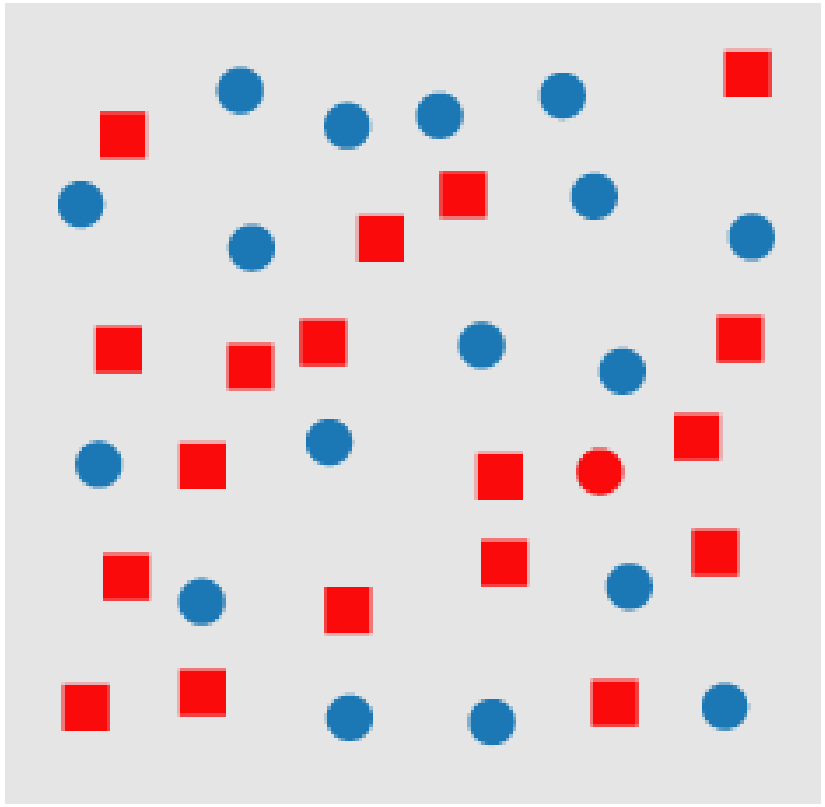




http://www.csc.ncsu.edu/faculty/healey/PP/index.html

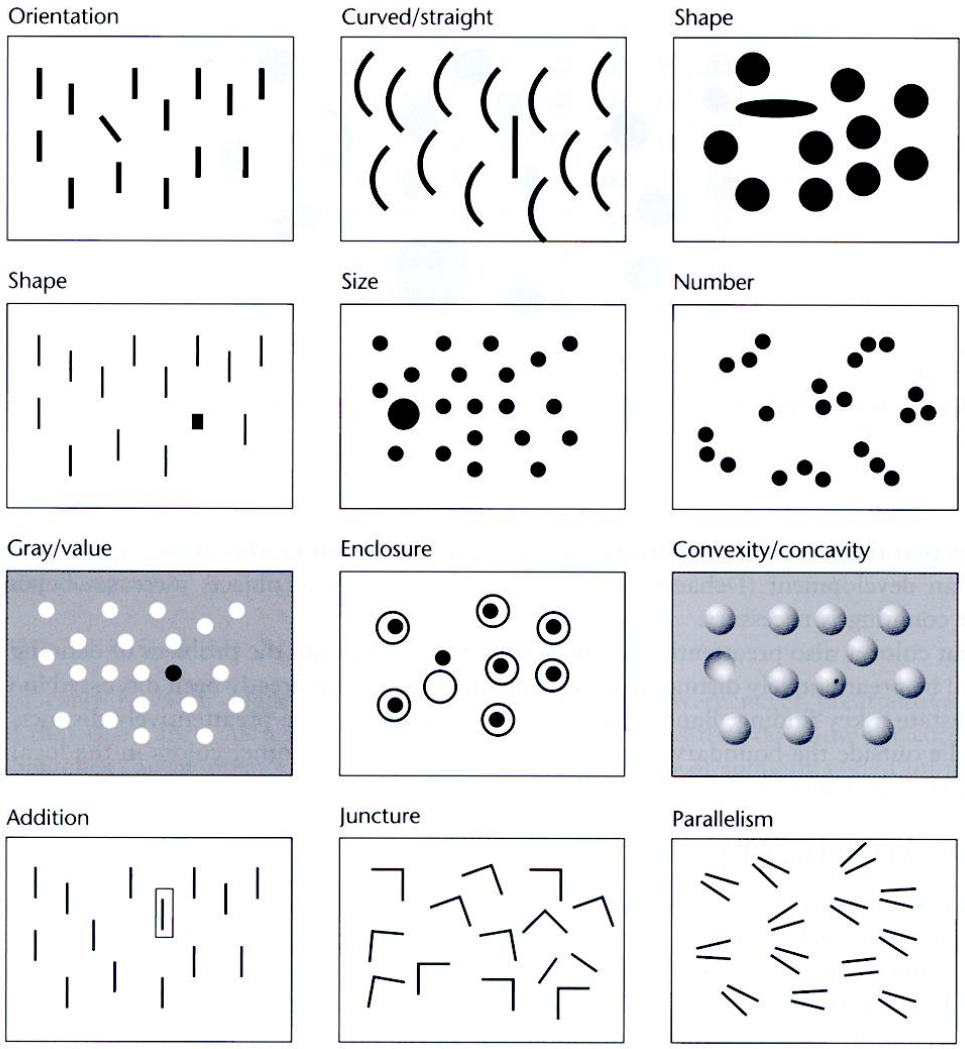
Feature Conjunctions





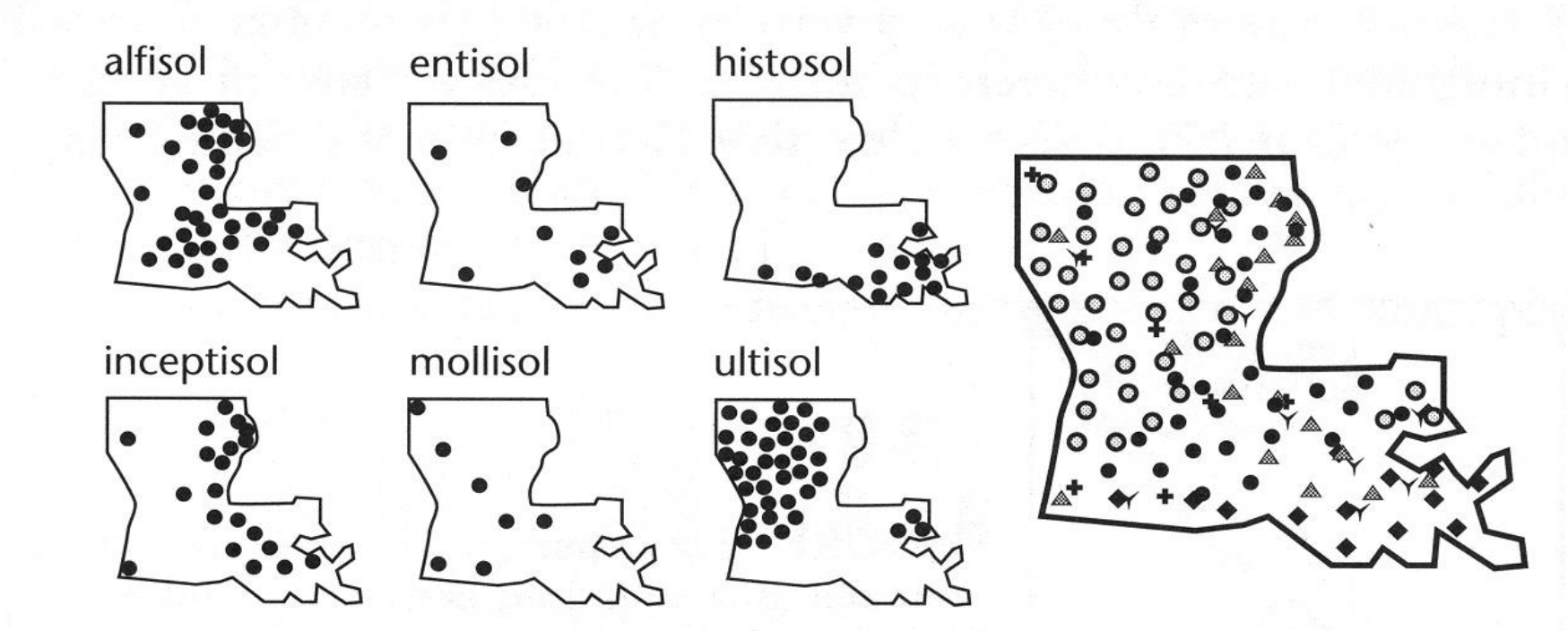
http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Pre-Attentive features



[Information Visualization. Figure 5. 5 Ware 04]

Small Multiples



[Figure 2.11, p. 38, MacEachren 95]

Which best encodes quantities?

Position

Length

Area

Volume

Value (Brightness)

Color Hue

Orientation (Angle)

Shape

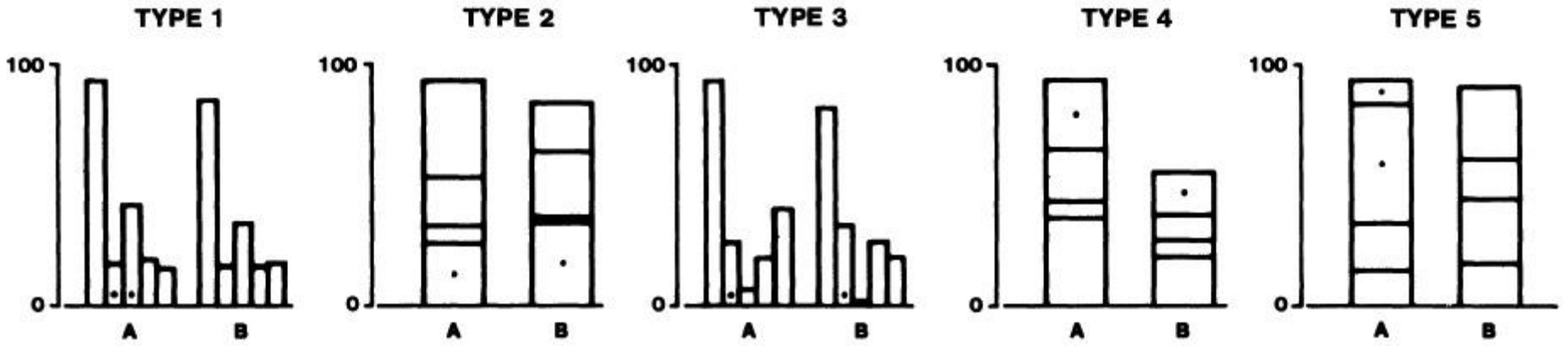


Figure 4. Graphs from position–length experiment.

Cleveland & McGill, Graphical Perception 1984

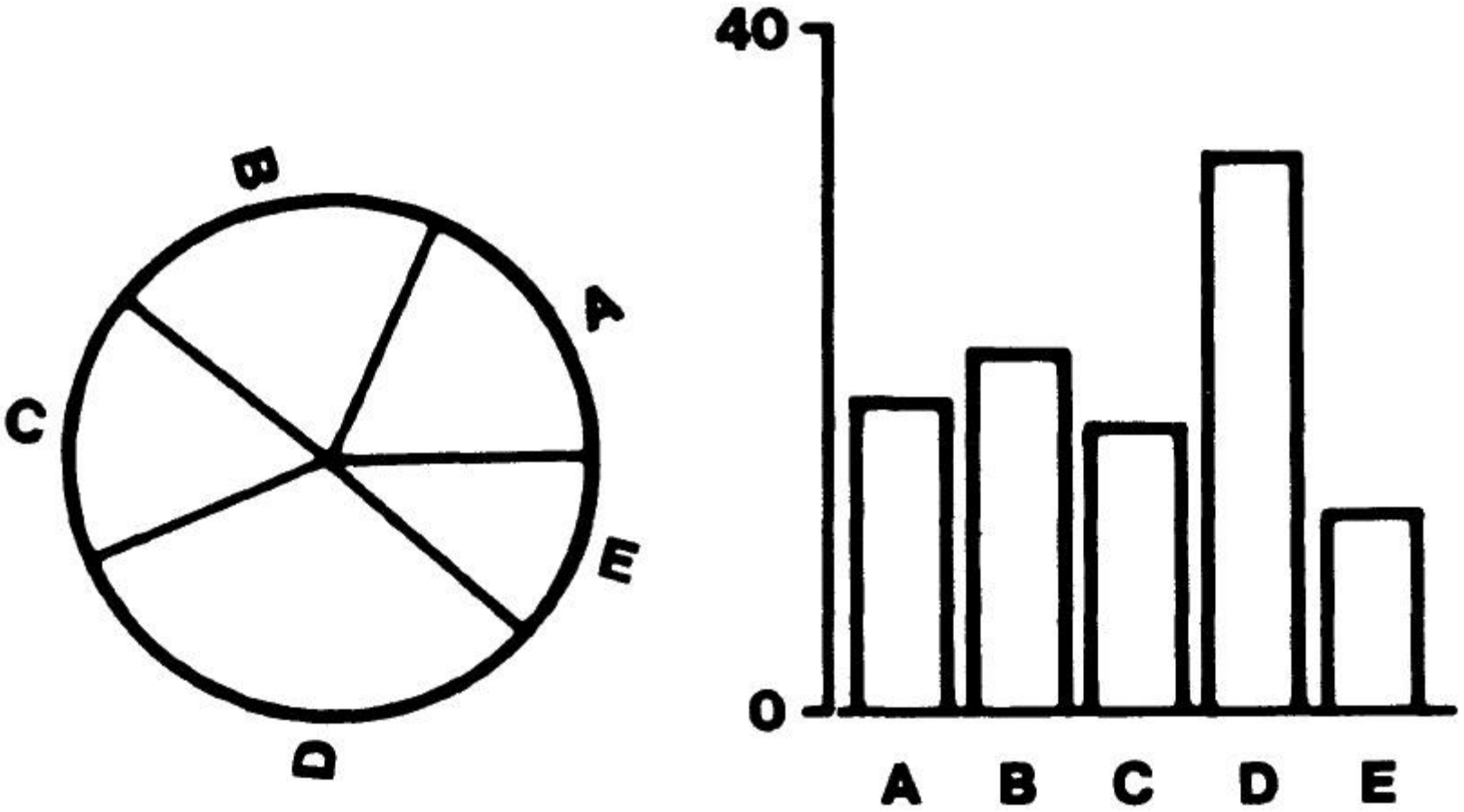


Figure 3. Graphs from position–angle experiment.

[Cleveland and McGill 84]

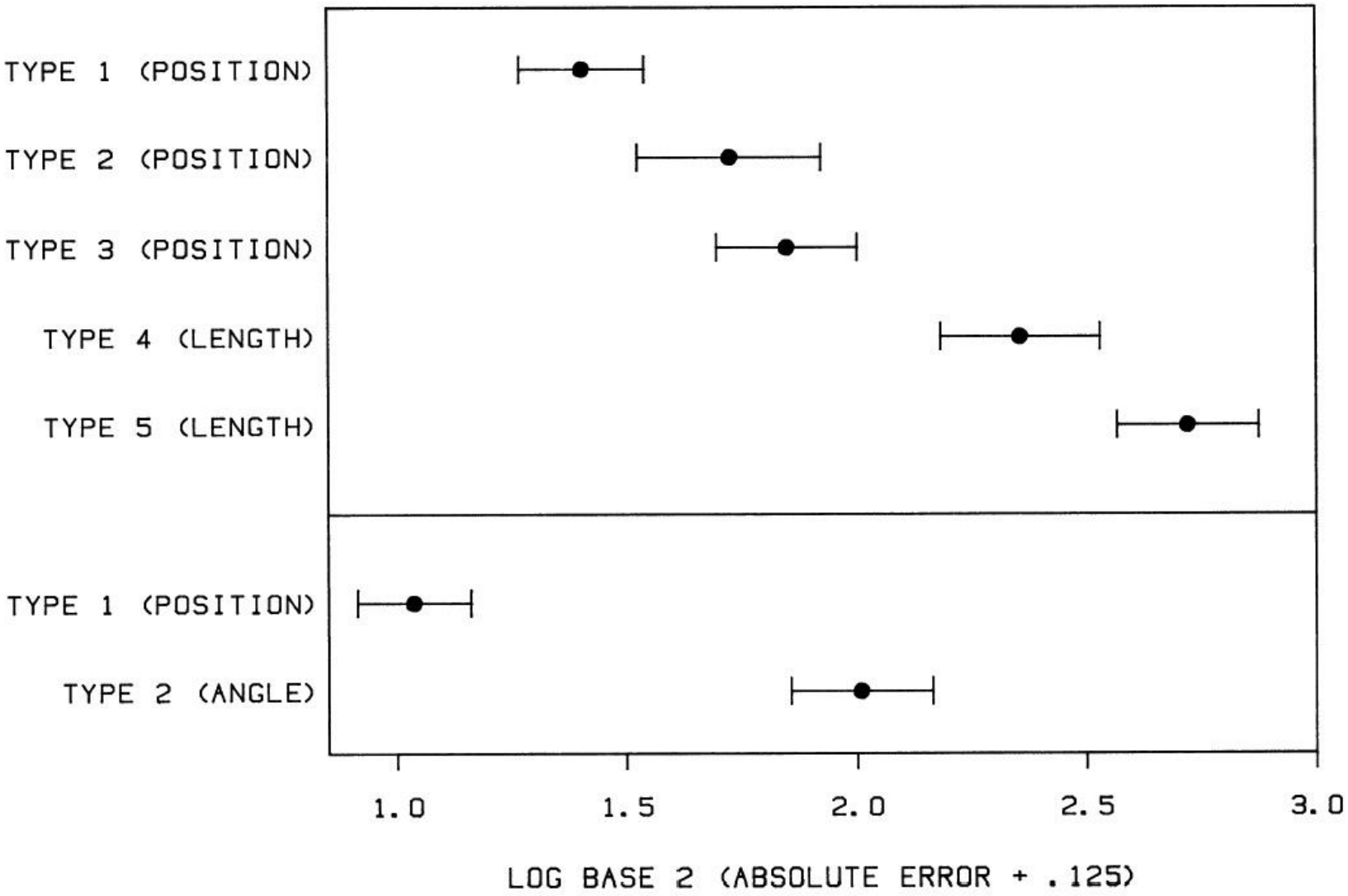
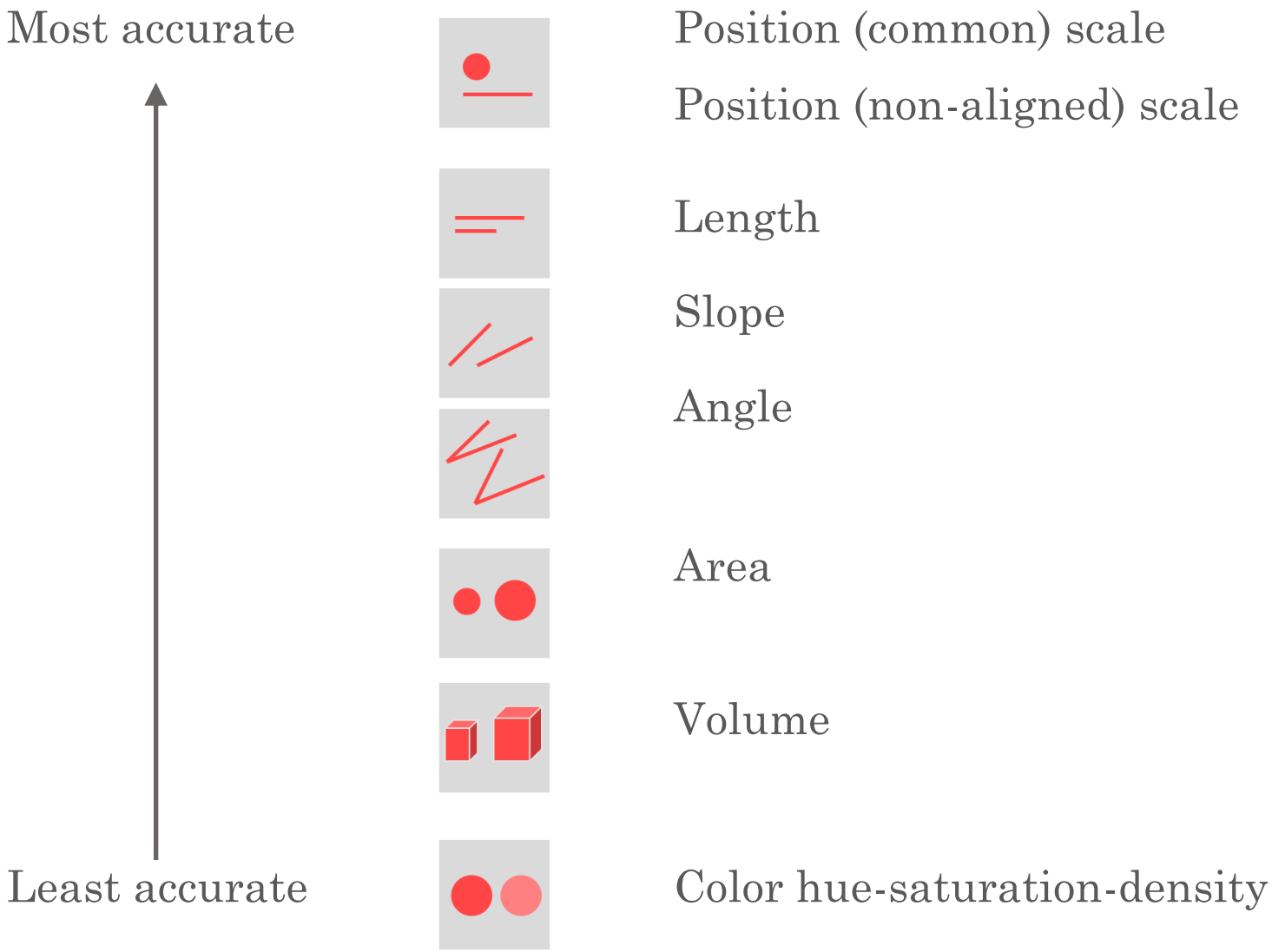


Figure 16. Log absolute error means and 95% confidence intervals for judgment types in position–length experiment (top) and position–angle experiment (bottom).

[Cleveland and McGill 84]


Most accurate
Least accurate
Position (common) scale
Position (non-aligned) scale
Length
Slope
Angle
Area
Volume
Color hue-saturation-density

Combinatorics of Encodings

Challenge:

Pick the best encoding from the exponential number of possibilities $(n+1)^8$

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.

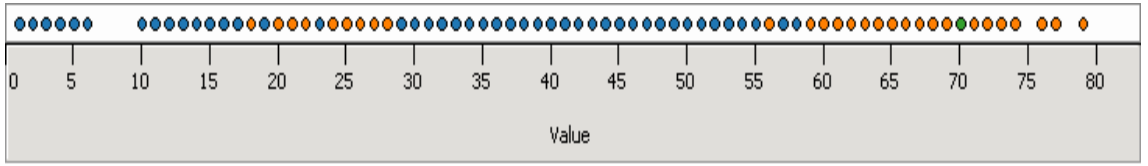
Design Criteria (Mackinlay)

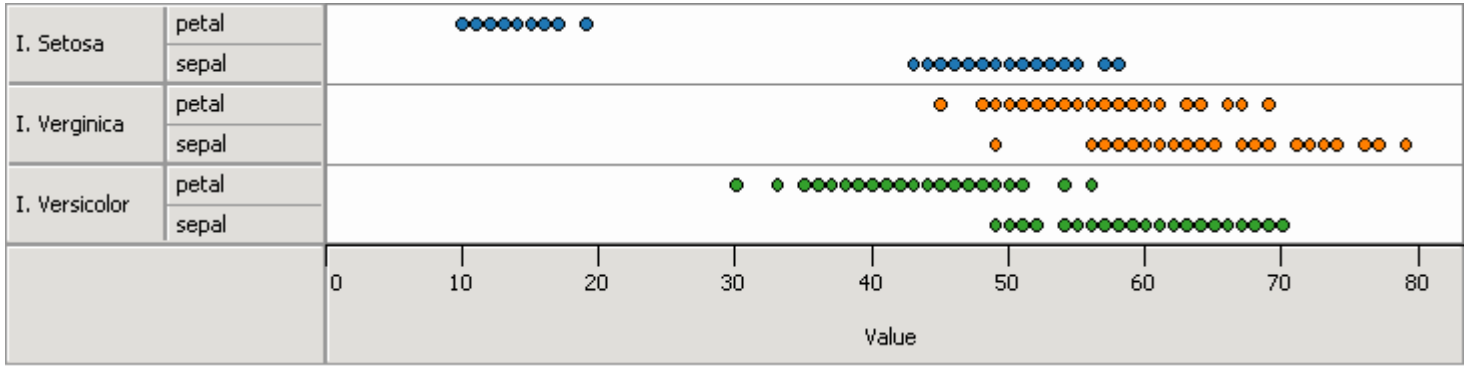
Expressiveness

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express *all* the facts in the set of data, and *only* the facts in the data.

Cannot express the facts

A one-to-many (1 → N) relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position





Expresses facts not in the data

A length is interpreted as a quantitative value;

∴ Length of bar says something untrue about N data

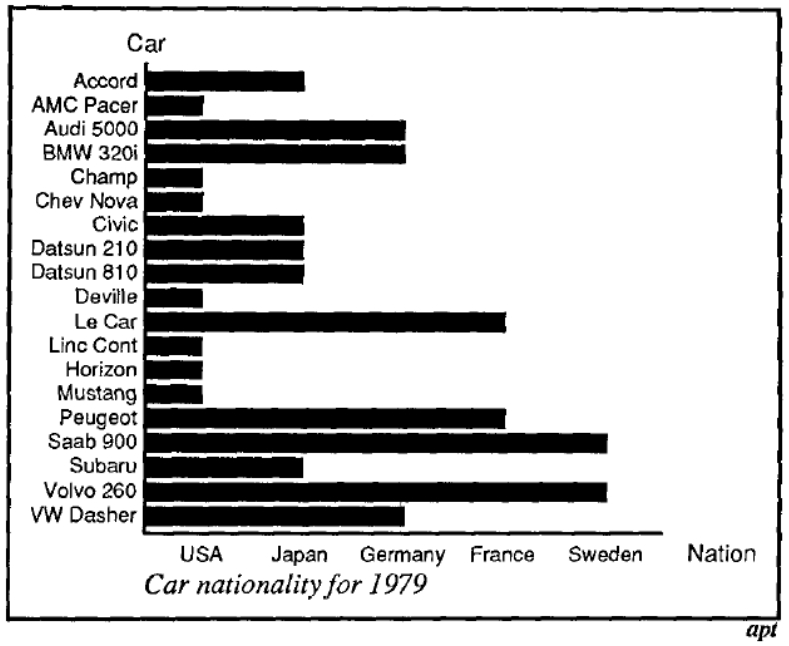


Fig. 11. Incorrect use of a bar chart for the *Nation* relation. The lengths of the bars suggest an ordering on the vertical axis, as if the USA cars were longer or better than the other cars, which is not true for the *Nation* relation.

[Mackinlay, APT, 1986]

Design Criteria (Mackinlay)

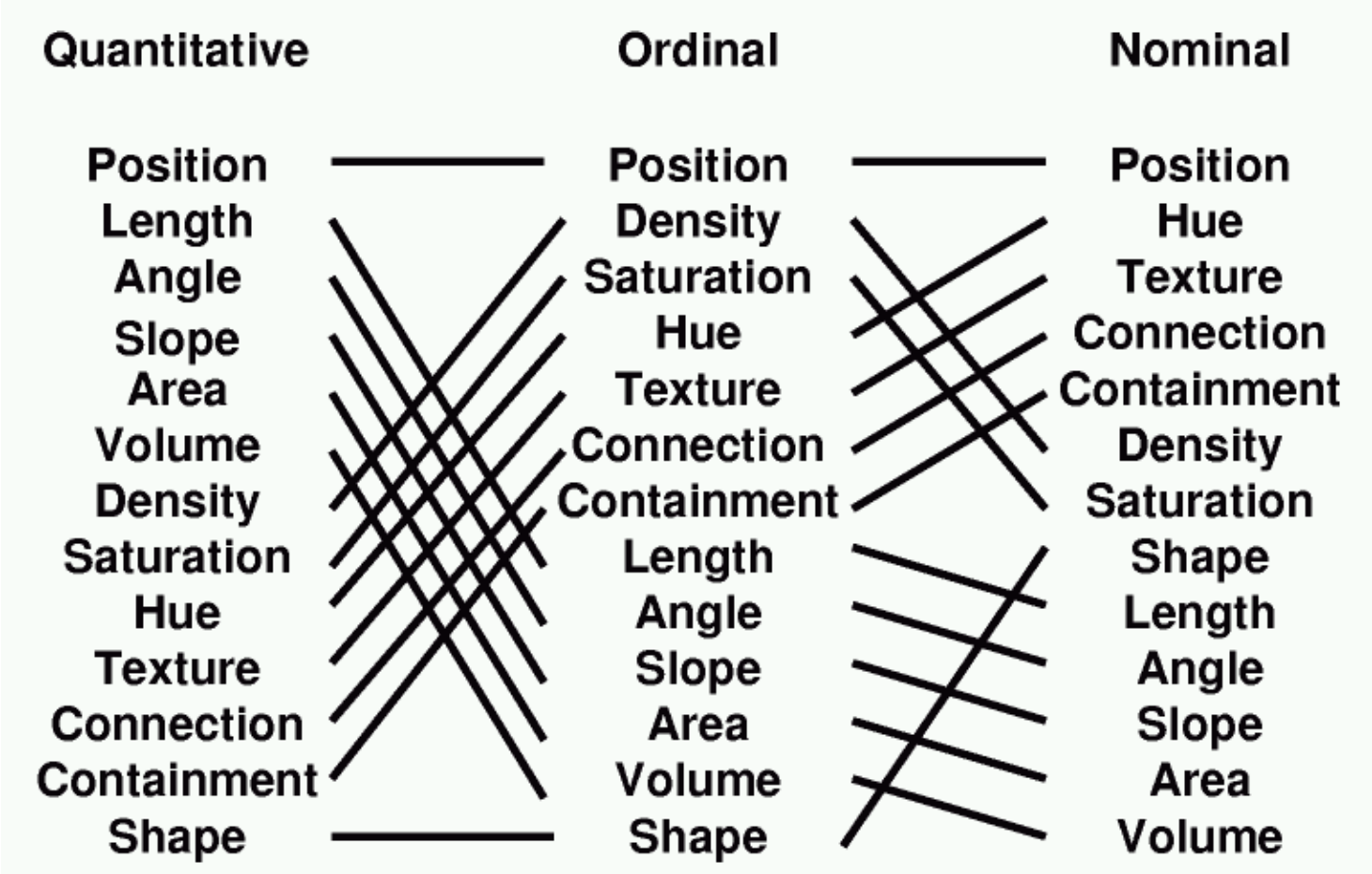
Expressiveness

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express *all* the facts in the set of data, and *only* the facts in the data.

Effectiveness

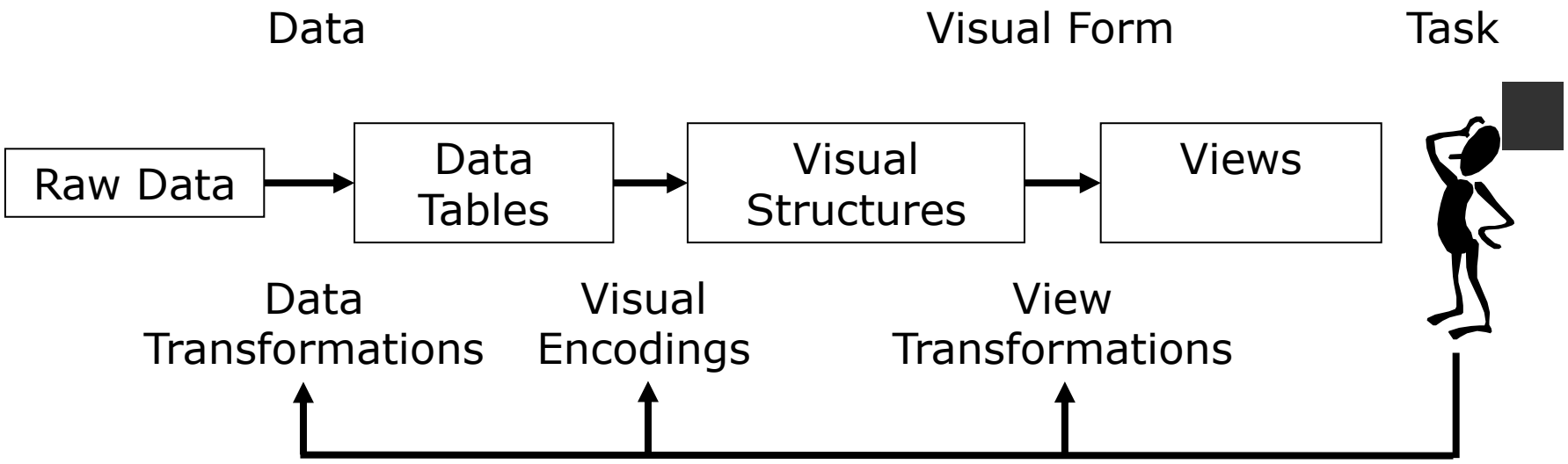
A visualization is more effective than another visualization if the information conveyed by one visualization is more readily *perceived* than the information in the other visualization.

Mackinlay's Ranking

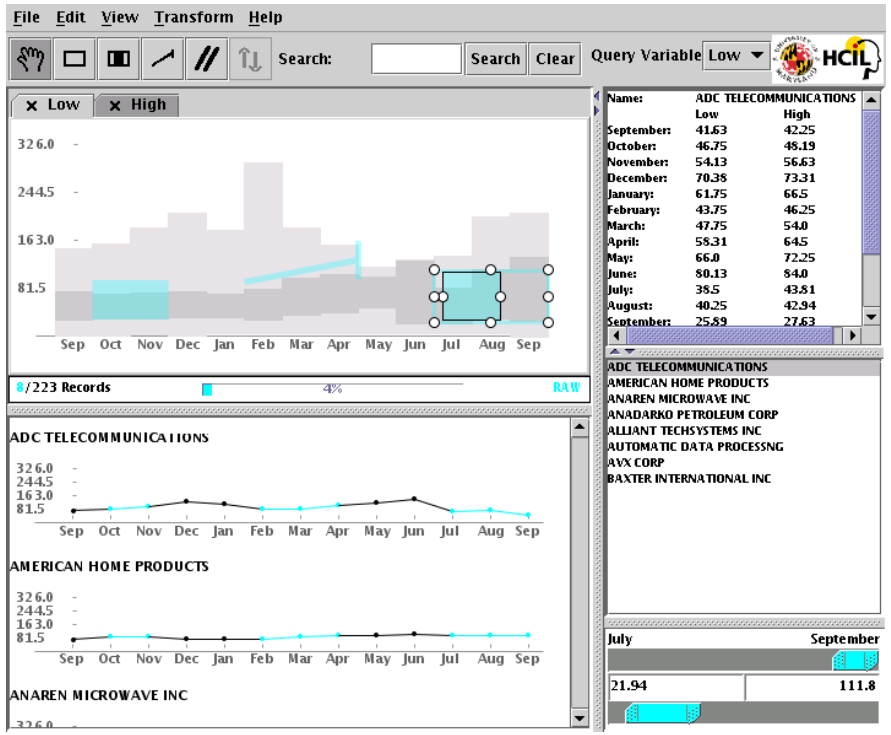


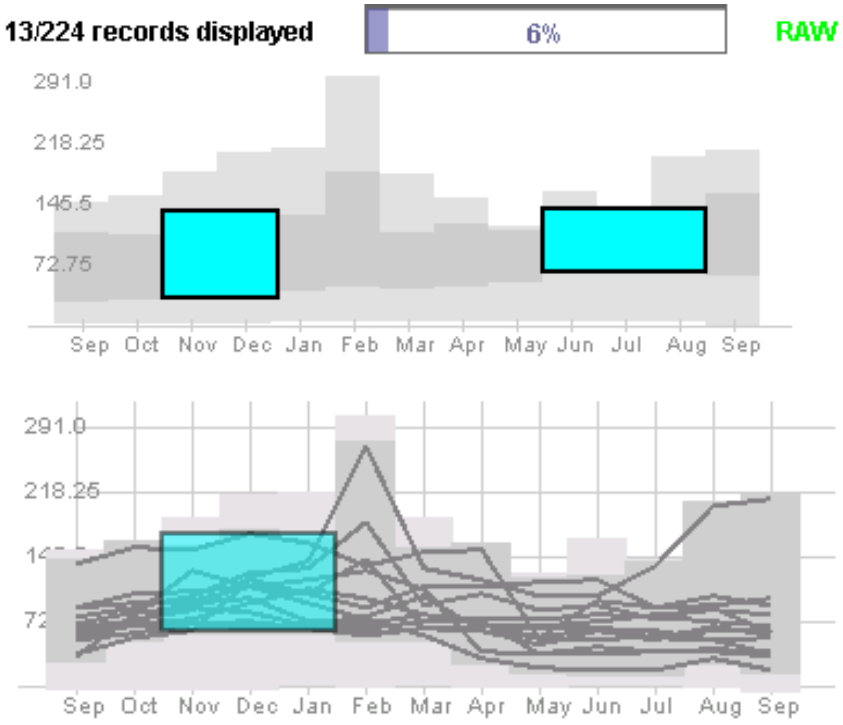
Conjectured *effectiveness* of the encoding

Visualization Reference Model



TimeSearcher [Hochheiser & Shneiderman 02]





Based on Wattenberg's [2001] idea for sketch-based queries of time-series data.

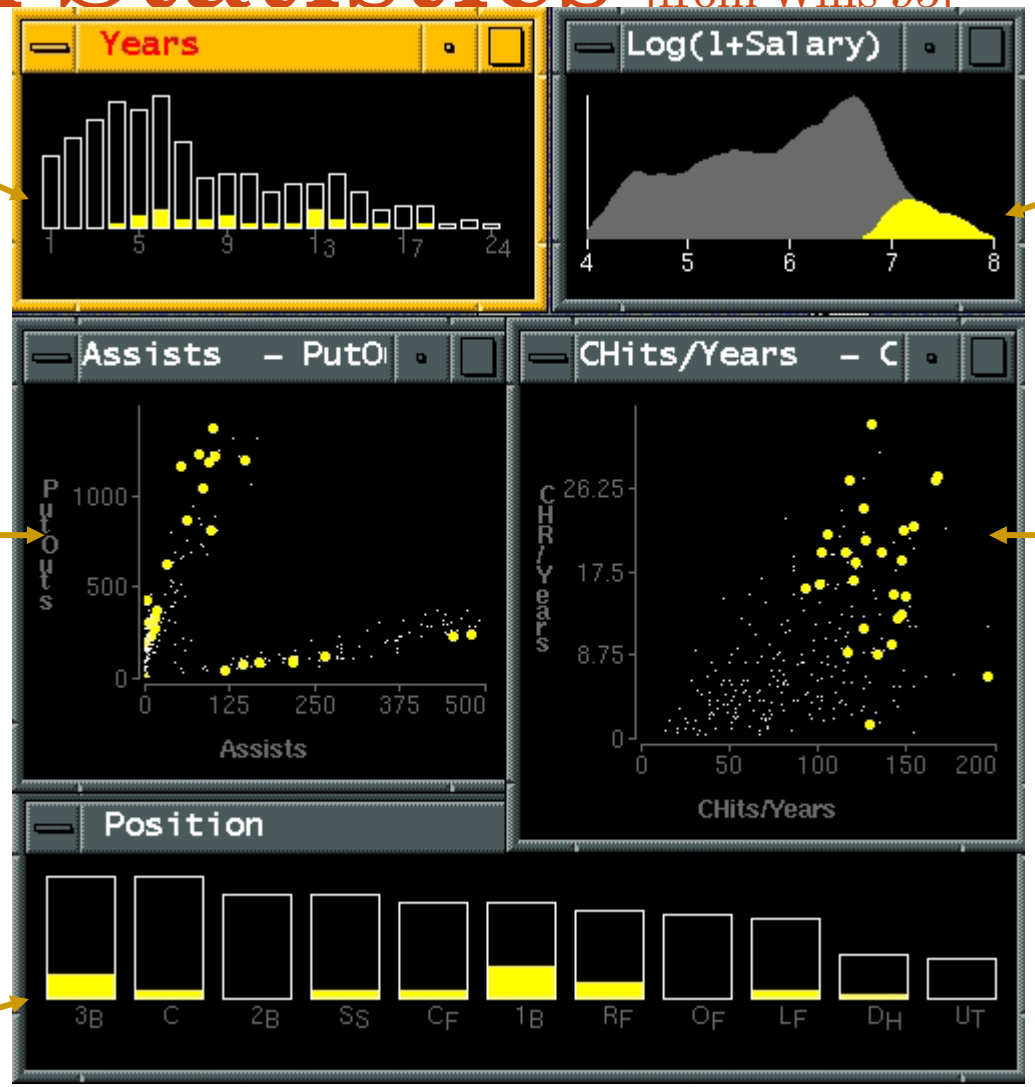
Interaction Techniques

Dynamic Queries

Filter a visualization through direct, reversable actions that avoid complex syntax.

Baseball Statistics [from Wills 95]

how long in majors



select high salaries

avg assists vs avg putouts (fielding ability)

avg career HRs vs avg career hits (batting ability)

distribution of positions played

Interaction Techniques

Dynamic Queries

Filter a visualization through direct, reversable actions that avoid complex syntax.

Brushing and Linking

Highlight relationships between related items across multiple visualization views.