Perception

Maneesh Agrawala

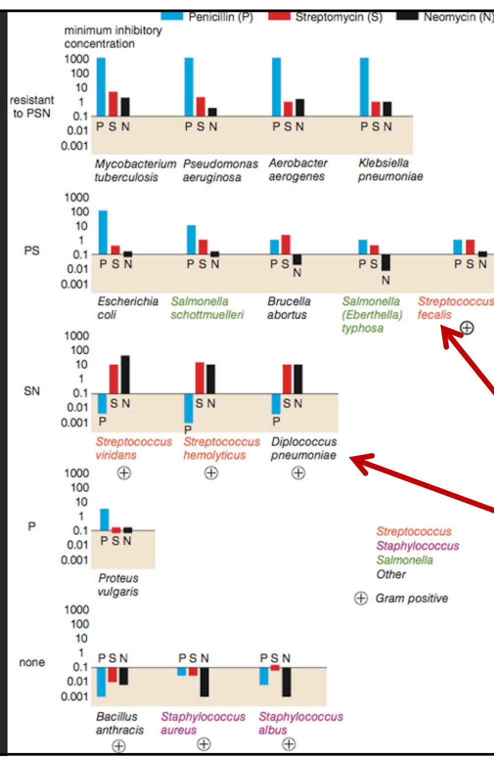
CS 448B: Visualization
Fall 2018

Last Time:
Exploratory Data Analysis

Will Burtin, 1951

Bacteria	Penicillin	Antibiotic Streptomycin	Neomycin	Gram stain
Aerobacter aerogenes	870	1	1.6	–
Brucella abortus	1	2	0.02	–
Bacillus anthracis	0.001	0.01	0.007	+
Diplococcus pneumoniae	0.005	11	10	+
Escherichia coli	100	0.4	0.1	–
Klebsiella pneumoniae	850	1.2	1	–
Mycobacterium tuberculosis	800	5	2	–
Proteus vulgaris	3	0.1	0.1	–
Pseudomonas aeruginosa	850	2	0.4	–
Salmonella (Eberthella) typhosa	1	0.4	0.008	–
Salmonella schottmuelleri	10	0.8	0.09	–
Staphylococcus albus	0.007	0.1	0.001	+
Staphylococcus aureus	0.03	0.03	0.001	+
Streptococcus fecalis	1	1	0.1	+
Streptococcus hemolyticus	0.001	14	10	+
Streptococcus viridans	0.005	10	40	+

How do the drugs compare?



How do the bacteria group with respect to antibiotic resistance?

Not a streptococcus! (realized ~30 yrs later)

Really a streptococcus! (realized ~20 yrs later)

Wainer & Lysen

American Scientist, 2009

Lessons

Exploratory Process

1. Construct graphics to address questions
2. Inspect "answer" and assess new questions
3. Repeat!

Transform the data appropriately (e.g., invert, log)

"Show data variation, not design variation"

-Tufte

Formulating a Hypothesis

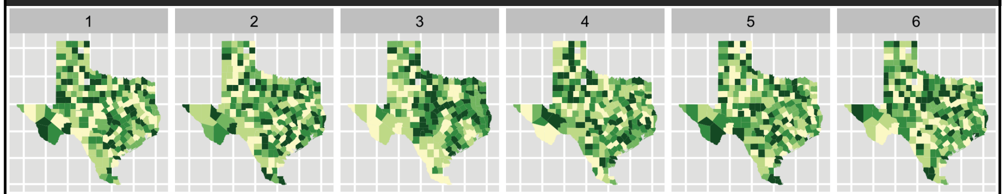
Null Hypothesis (H_0): $\mu_m = \mu_f$
(population)

Alternate Hypothesis (H_a): $\mu_m \neq \mu_f$
(population)

A statistical hypothesis test assesses the likelihood of the null hypothesis.

What is the probability of sampling the observed data assuming population means are equal?

This is called the *p* value

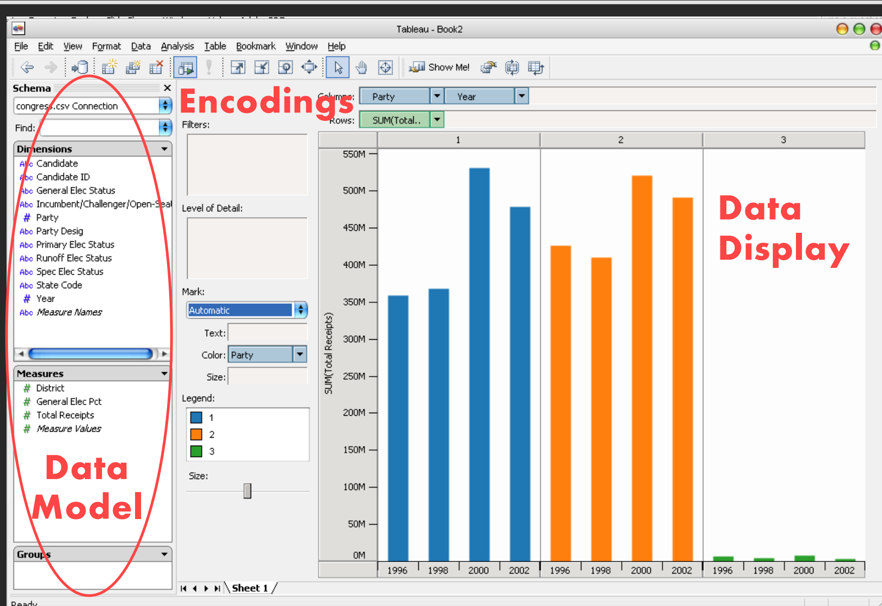


Choropleth maps of cancer deaths in Texas.

One plot shows a real data sets. The others are simulated under the null hypothesis of spatial independence.

Can you spot the real data? If so, you have some evidence of spatial dependence in the data.

Tableau



Polaris/Tableau Approach

Insight: simultaneously specify both database queries and visualization

Choose data, then visualization, not vice versa

Use smart defaults for visual encodings

Recently: automate visualization design (ShowMe – Like APT)

Specifying Table Configurations

Operands are names of database fields

Each operand interpreted as a set {…}

Data is either Ordinal or Quantitative

Three operators:

- concatenation (+)
- cross product (x)
- nest (/)

Table Algebra: Operands

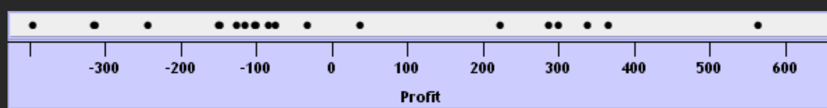
Ordinal fields: interpret domain as a set that partitions table into rows and columns

Quarter = {(Qtr1),(Qtr2),(Qtr3),(Qtr4)} →

Qtr1	Qtr2	Qtr3	Qtr4
95892	101760	105282	98225

Quantitative fields: treat domain as single element set and encode spatially as axes

Profit = {(Profit[-410,650])} →



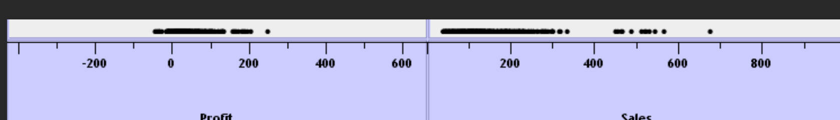
Concatenation (+) Operator

Ordered union of set interpretations

Quarter + Product Type
= {(Qtr1),(Qtr2),(Qtr3),(Qtr4)} + {(Coffee), (Espresso)}
= {(Qtr1),(Qtr2),(Qtr3),(Qtr4),(Coffee),(Espresso)}

Qtr1	Qtr2	Qtr3	Qtr4	Coffee	Espresso
48	59	57	53	151	21

Profit + Sales = {(Profit[-310,620]),(Sales[0,1000])}



Cross (x) Operator

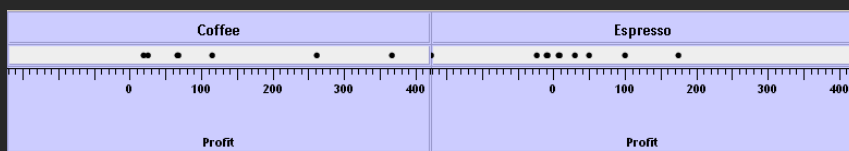
Cross-product of set interpretations

Quarter x Product Type

= {(Qtr1,Coffee), (Qtr1, Tea), (Qtr2, Coffee), (Qtr2, Tea), (Qtr3, Coffee), (Qtr3, Tea), (Qtr4, Coffee), (Qtr4,Tea)}

Qtr1		Qtr2		Qtr3		Qtr4	
Coffee	Espresso	Coffee	Espresso	Coffee	Espresso	Coffee	Espresso
131	19	160	20	178	12	134	33

Product Type x Profit =



Nest (/) Operator

Cross-product filtered by existing records

Quarter x Month

creates twelve entries for each quarter. i.e., (Qtr1, December)

Quarter / Month

creates three entries per quarter based on tuples in database (not semantics)

Polaris/Tableau Table Algebra

The operators (+, x, /) and operands (O, Q) provide an *algebra* for tabular visualization.

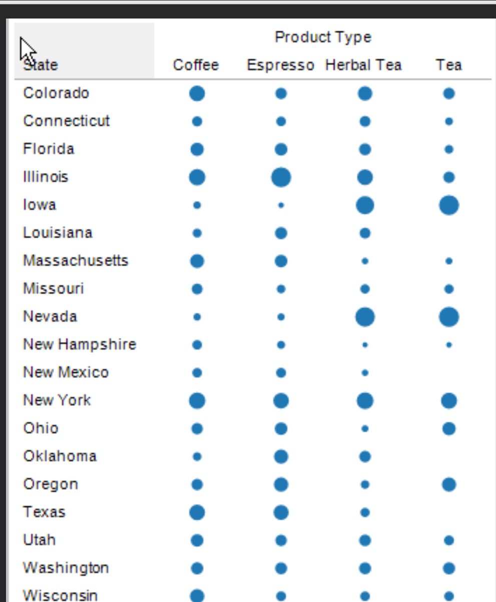
Algebraic statements are then mapped to:

- Queries - selection, projection, group-by aggregation
- Visualizations - trellis plot partitions, visual encodings

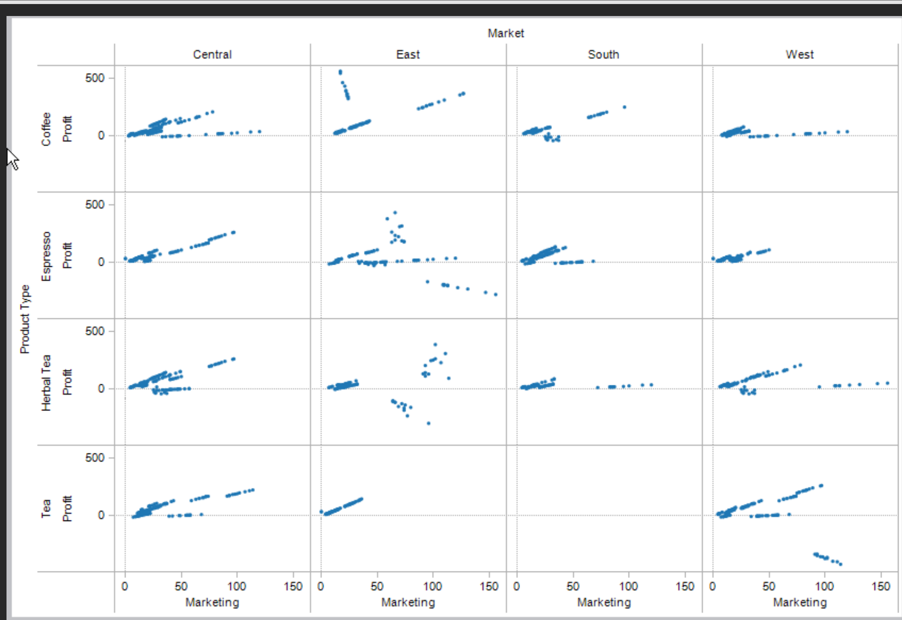
In Tableau, users make statements via drag-and-drop

- Note that this specifies operands NOT operators!
- Operators are inferred by data type (O, Q)

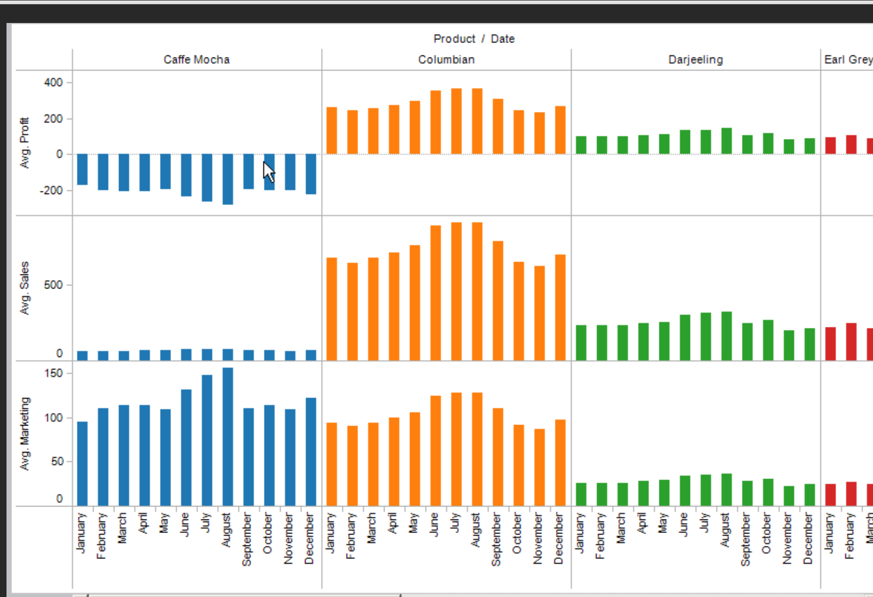
Ordinal - Ordinal



Quantitative - Quantitative



Ordinal - Quantitative



Summary

Exploratory analysis may combine graphical methods, and statistics

Use questions to uncover more questions

Formal methods may be used to confirm

Interaction is essential for exploring large multidimensional datasets

Announcements

A2: Exploratory Data Analysis

Use Tableau to formulate & answer questions

First steps

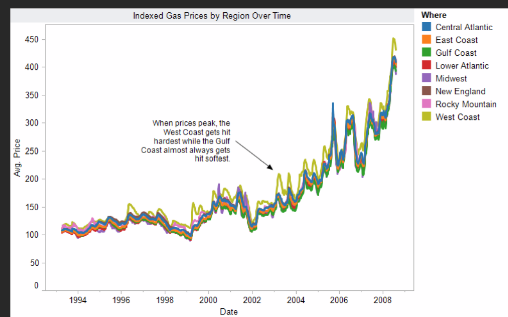
- Step 1: Pick a domain
- Step 2: Pose questions
- Step 3: Find data
- Iterate

Create visualizations

- Interact with data
- Question will evolve
- Tableau

Make wiki notebook

- Keep record of all steps you took to answer the questions



Due before class on Oct 15, 2018

Perception

Mackinlay's ranking of encodings

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Val)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Val)
Density (Val)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Topics

Signal Detection
Magnitude Estimation
Pre-Attentive Visual Processing
Using Multiple Visual Encodings
Gestalt Grouping
Change Blindness

Detection

Detecting brightness

Which is brighter?

Detecting brightness

(128, 128, 128) (130, 130, 130)

Which is brighter?

Just noticeable difference

JND (Weber's Law)

$$\Delta S = k \frac{\Delta I}{I}$$

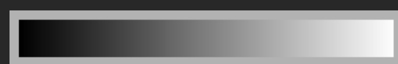
- Ratios more important than magnitude
- Most continuous variations in stimuli are perceived in discrete steps

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

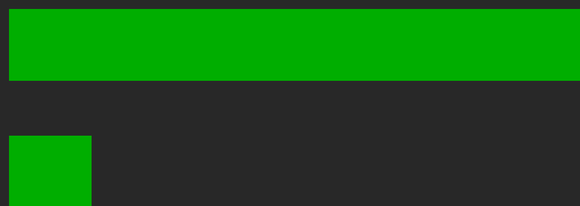
∴ Encode nominal variables (N) using color

Steps in font size

Sizes standardized in 16th century

Estimating Magnitude

Compare areas of circles

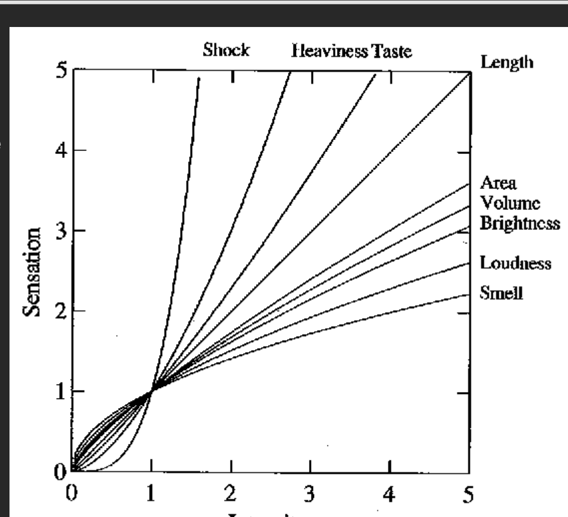


Compare lengths of bars

Steven's power law

$$S = I^p$$

p < 1 : underestimate
p > 1 : overestimate



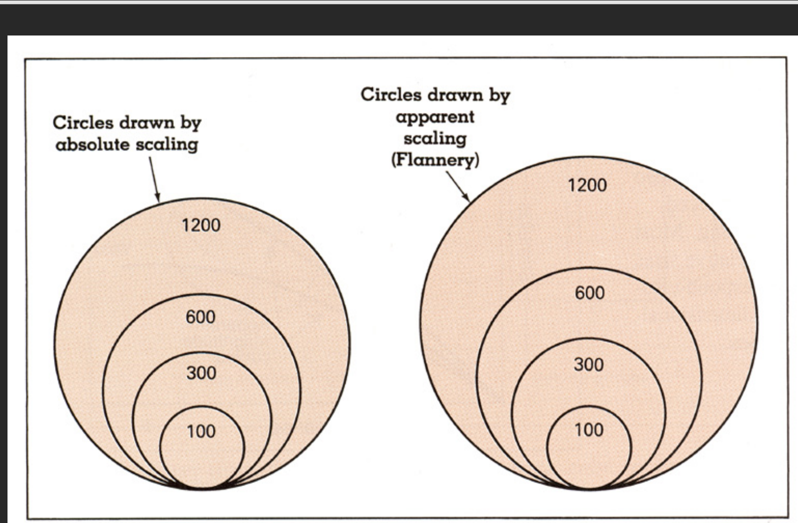
[graph from Wilkinson 99, based on Stevens 61]

Exponents of power law

Sensation	Exponent
Loudness	0.6
Brightness	0.33
Smell	0.55 (Coffee) - 0.6 (Heptane)
Taste	0.6 (Saccharine) -1.3 (Salt)
Temperature	1.0 (Cold) – 1.6 (Warm)
Vibration	0.6 (250 Hz) – 0.95 (60 Hz)
Duration	1.1
Pressure	1.1
Heaviness	1.45
Electic Shock	3.5

[Psychophysics of Sensory Function, Stevens 61]

Apparent magnitude scaling

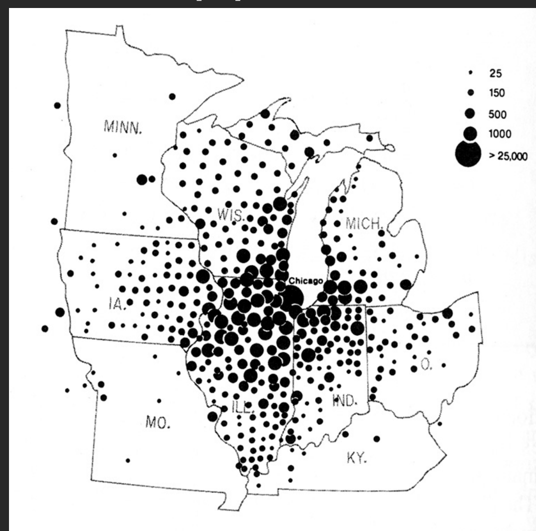


[Cartography: Thematic Map Design, Figure 8.6, p. 170, Dent, 96]

$S = 0.98A^{0.87}$ [from Flannery 71]

Proportional symbol map

Newspaper Circulation



[Cartography: Thematic Map Design, Figure 8.8, p. 172, Dent, 96]

Graduated sphere map

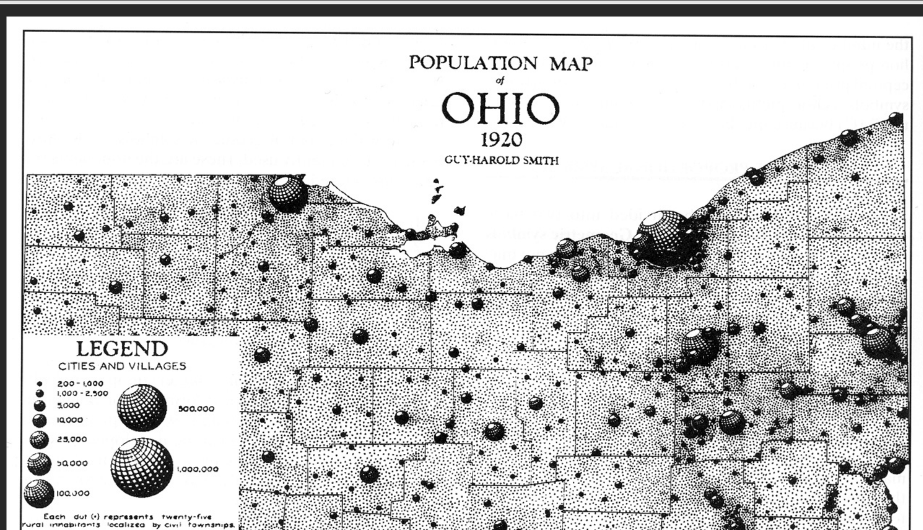


FIGURE 7.4. An eye-catching map created using three-dimensional geometric symbols. (After Smith, 1928. First published in *The Geographical Review,* 18(3), plate 4. Reprinted with permission of the American Geographical Society.)

Cleveland and McGill

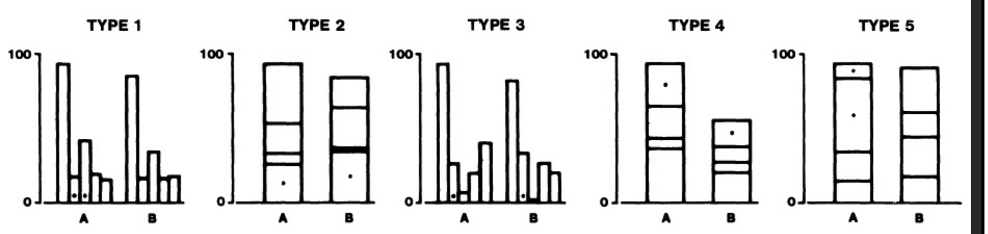


Figure 4. Graphs from position–length experiment.

[Cleveland and McGill 84]

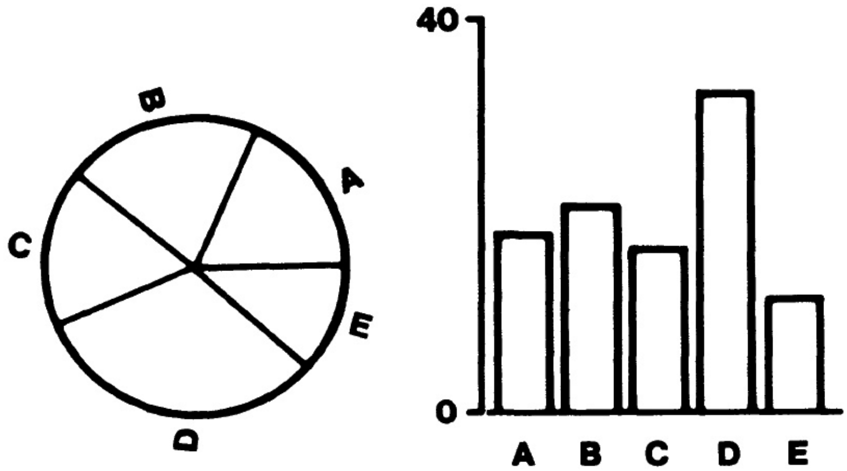


Figure 3. Graphs from position–angle experiment.

[Cleveland and McGill 84]

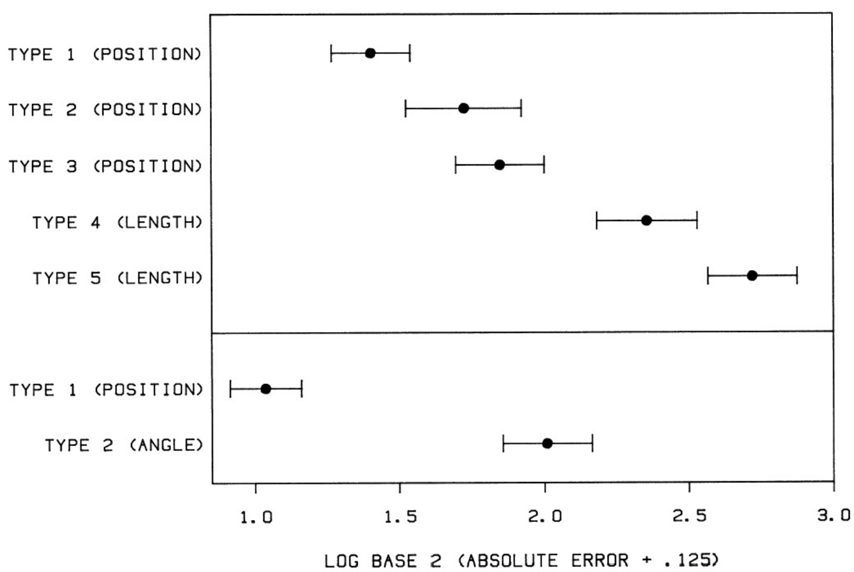


Figure 16. Log absolute error means and 95% confidence intervals for judgment types in position–length experiment (top) and position–angle experiment (bottom).

[Cleveland and McGill 84]

Relative magnitude estimation

Most accurate

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

Least accurate

Mackinlay's ranking of encodings

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density (Val)	Color Hue
Angle	Color Sat	Texture
Slope	Color Hue	Connection
Area (Size)	Texture	Containment
Volume	Connection	Density (Val)
Density (Val)	Containment	Color Sat
Color Sat	Length	Shape
Color Hue	Angle	Length
Texture	Slope	Angle
Connection	Area (Size)	Slope
Containment	Volume	Area
Shape	Shape	Volume

Preattentive vs. Attentive

How many 3's

1281768756138976546984506985604982826762
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

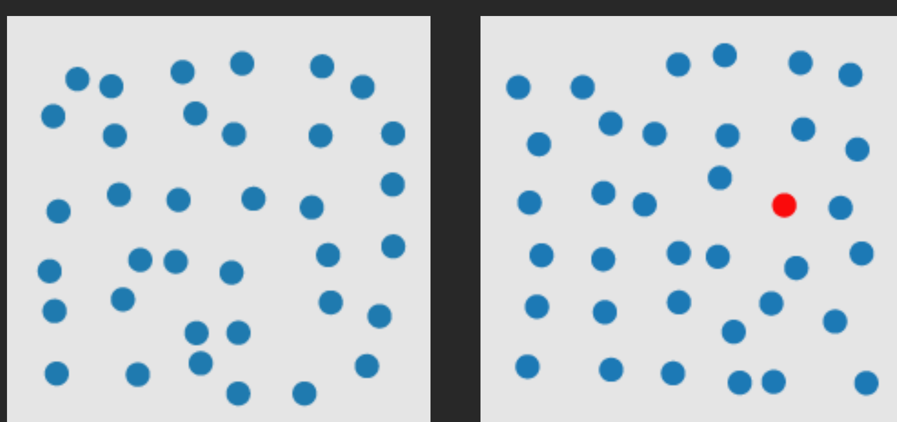
[based on slide from Stasko]

How many 3's

1281768756138976546984506985604982826762
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

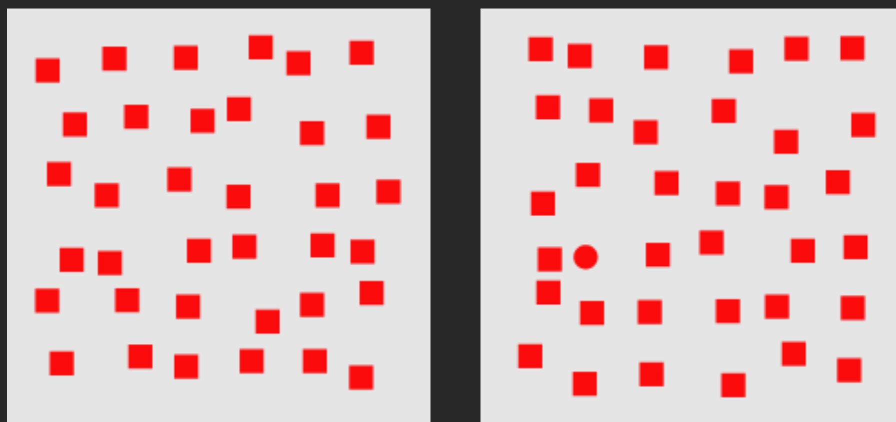
[based on slide from Stasko]

Visual pop-out: Color



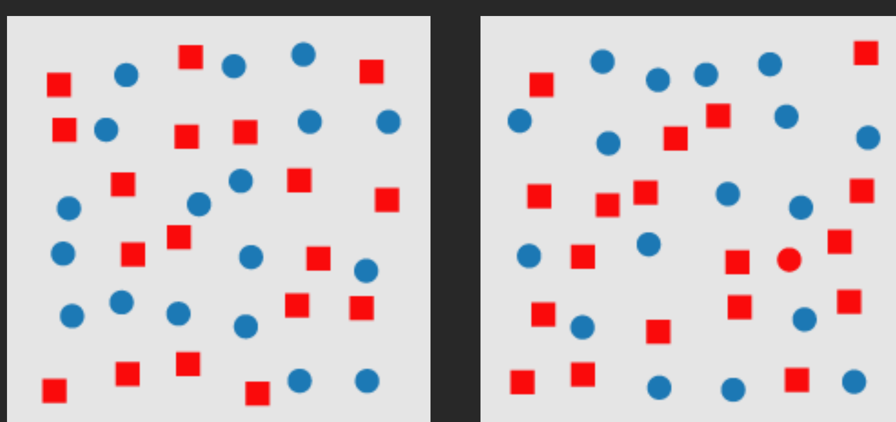
http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Visual pop-out: Shape



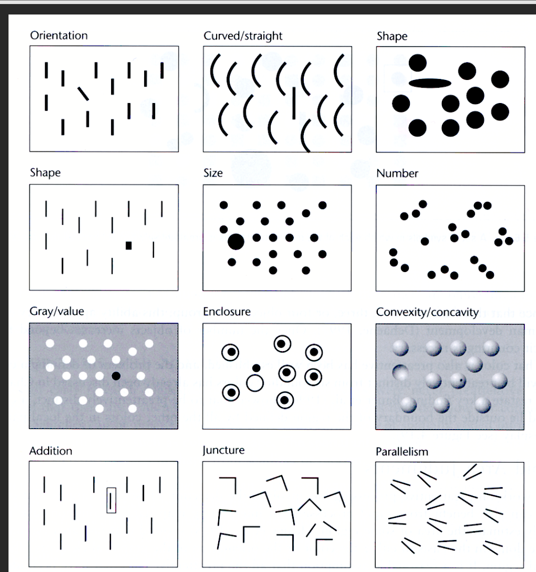
http://www.csc.ncsu.edu/faculty/healey/PP/index.html

Feature conjunctions



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

Preattentive features



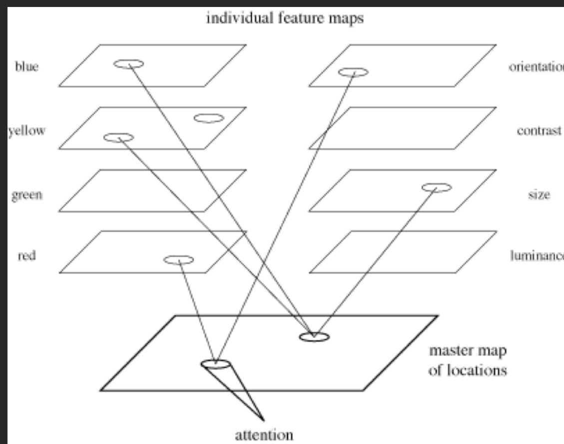
[Information Visualization. Figure 5. 5 Ware 04]

More preattentive features

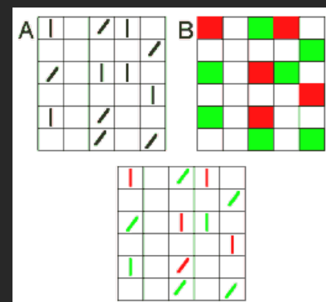
Line (blob) orientation	Julesz & Bergen [1983]; Wolfe et al. [1992]
Length	Triesman & Gormican [1988]
Width	Julesz [1985]
Size	Triesman & Gelade [1980]
Curvature	Triesman & Gormican [1988]
Number	Julesz [1985]; Trick & Pylyshyn [1994]
Terminators	Julesz & Bergen [1983]
Intersection	Julesz & Bergen [1983]
Closure	Enns [1986]; Triesman & Souther [1985]
Colour (hue)	Nagy & Sanchez [1990, 1992]; D'Zmura [1991]; Kawai et al. [1995]; Bauer et al. [1996]
Intensity	Beck et al. [1983]; Triesman & Gormican [1988]
Flicker	Julesz [1971]
Direction of motion	Nakayama & Silverman [1986]; Driver & McLeod [1992]
Binocular lustre	Wolfe & Franzel [1988]
Stereoscopic depth	Nakayama & Silverman [1986]
3-D depth cues	Enns [1990]
Lighting direction	Enns [1990]

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

Feature-integration theory



Treisman's feature integration model [Healey04]

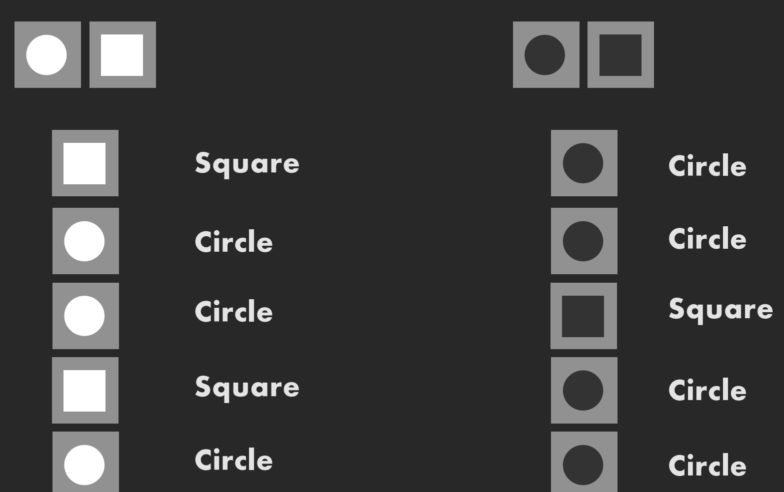


Feature maps for orientation & color [Green]

Multiple Attributes

One-dimensional: Lightness

One-dimensional: Shape



Correlated dims: Shape or lightness

- Circle
- Square
- Square
- Circle
- Square

- Circle
- Square
- Square
- Square
- Circle

Orthogonal dims: Shape & lightness

- Circle
- Square
- Square
- Circle
- Square

Speeded classification

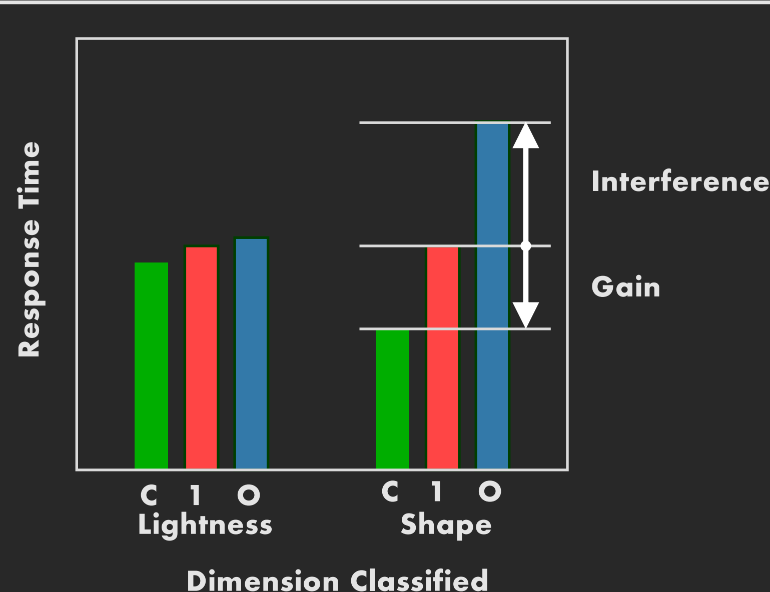
Redundancy gain

Facilitation in reading one dimension when the other provides redundant information

Filtering interference

Difficulty in ignoring one dimension while attending to the other

Speeded classification



Types of dimensions

Integral

Filtering interference and redundancy gain

Separable

No interference or gain

Configural

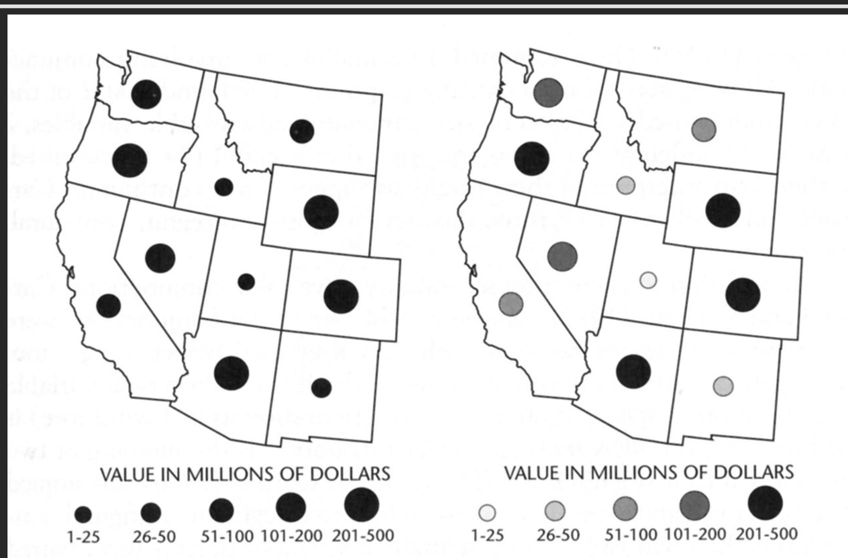
Only interference, but no redundancy gain

Asymmetrical

One dimension separable from other, not vice versa

Stroop effect – Color naming influenced by word identity, but word naming not influenced by color

Correlated dims: Size and value



W. S. Dobson, Visual information processing and cartographic communication: The role of redundant stimulus dimensions, 1983 (reprinted in MacEachren, 1995)

Othogonal dims: Aspect ratio

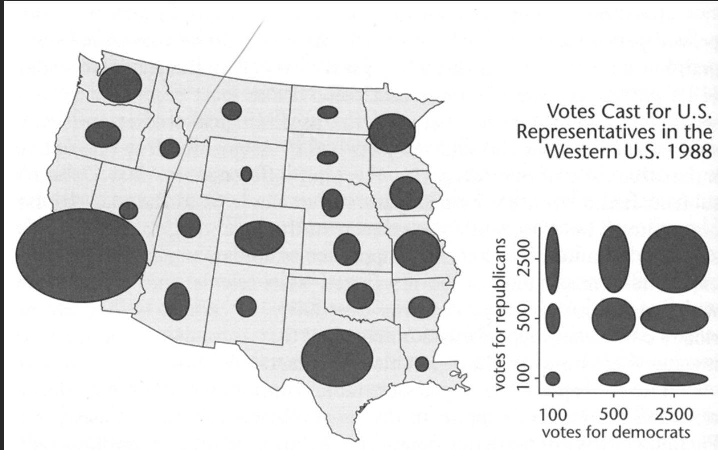


FIGURE 3.38. An example of the use of an ellipse as a map symbol in which the horizontal and vertical axes represent different (but presumably related) variables.

[MacEachren 95]

Orientation and Size (Single Mark)

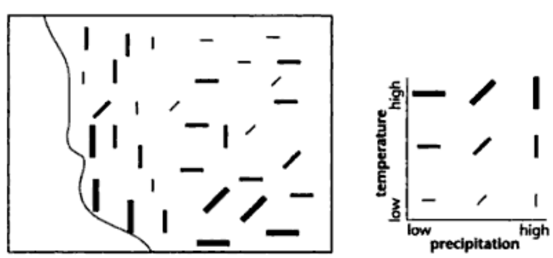


FIGURE 3.36. A map of temperature and precipitation using symbol size and orientation to represent data values on the two variables.

How well can you see temperature or precipitation?
Is there a correlation between the two?

[MacEachren 95]

Shape and Size

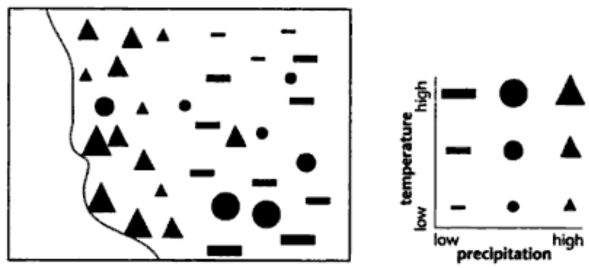
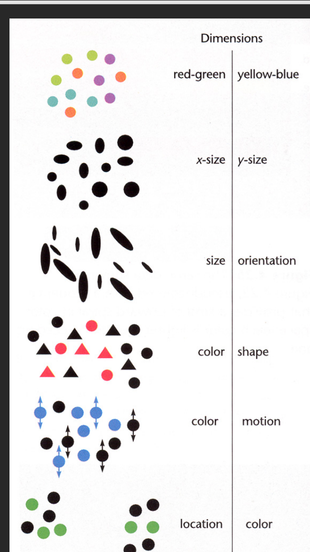


FIGURE 3.40. The bivariate temperature–precipitation map of Figure 3.36, this time using point symbols that vary in shape and size to represent the two quantities.

Easier to see one shape across multiple sizes than one size of across multiple shapes?

[MacEachren 95]

Summary of Integral-Separable



Integral

Separable

[Figure 5.25, Color Plate 10, Ware 00]