Encodings

Michael Gleicher
Department of Computer Sciences
University of Wisconsin Madison

The Lunch Problem

Restaurants Have

A Location

A type (French, ...)

Price ($ - $$$$$)

Star rating (★ - ★★★★★)

Different rating (0-10)

Open Mondays

Takes reservations online

(and more...)

Neighborhood

Distance to Apartment

Walking time to Apt

How much on a glyph?

A thought exercise…

We are putting symbols on a map

Only restaurants

(don't even worry about streets)

Glyph Design Problem

A type (French, …)
Price ($ - $$$$$)
Star rating (★-5)
Different rating (0-10)
Open Mondays
Online reservations
(and more…)

Color
Shape
Size
Texture
Angle
Thickness
Elongation
Curvature
Area

Mini-Experiment

Come up with 3 different glyph encodings

You do not need to consider location
You can't use position

Pack as many variables as possible

Glyph Design Problem

A type (French, …)
Price ($ - $$$$$)
Star rating (★-5)
Different rating (0-10)
Open Mondays
Online reservations
(and more…)

Color
Shape
Size
Texture
Angle
Thickness
Elongation
Curvature
Area

Exam Question...

What should you have asked first?

Task!

Where are there many restaurants with …
What properties does this group have …
What's the closest restaurant with …

Is there a restaurant with …
Are there outliers?
(and many more)

Tasks to encodings

If the task doesn't require location
use position for something else

Does the task require all variables at the same time?

Search? Grouping? Averaging?

How much on a glyph?

How much on a glyph?

And, why not more?

What to think about?

What form is the **data**?

What **visual variables** do we have?

What tasks do we need to do?

How will those marks interact?

Encodings

How do we map data to marks?

Data Abstractions

What kind of data

Levels of Measurement

Categorical / Nominal

Ordinal

Interval

Ratio

Continuous vs. Discrete

Finite vs. Infinite

Other properties of types

- set size
- sequential / diverging
- continuous / discrete

Conversions between types

Down-conversions easy

Up-conversions harder

imposed orderings on categorical

Visual Variables

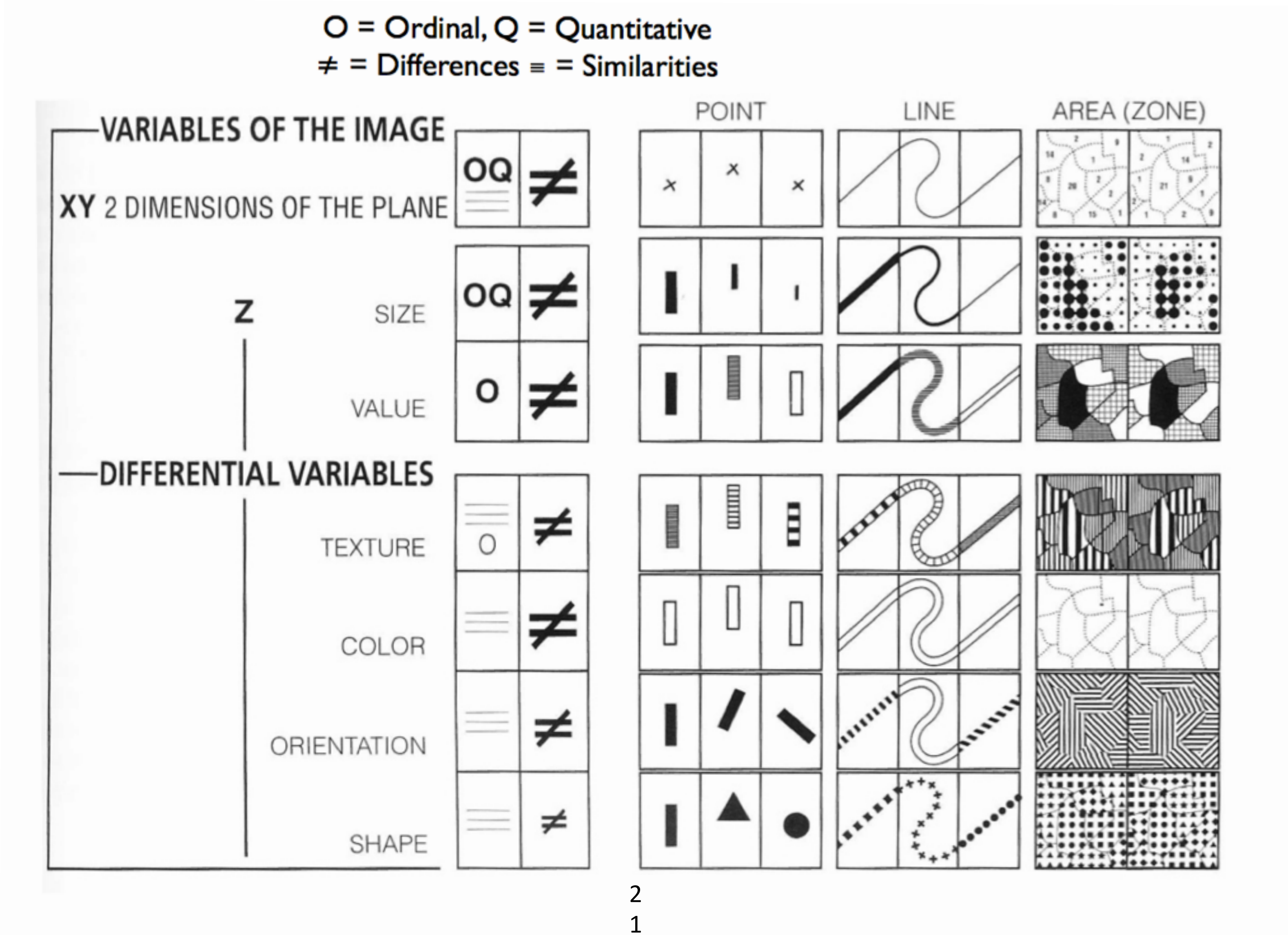
Data Types and Encodings

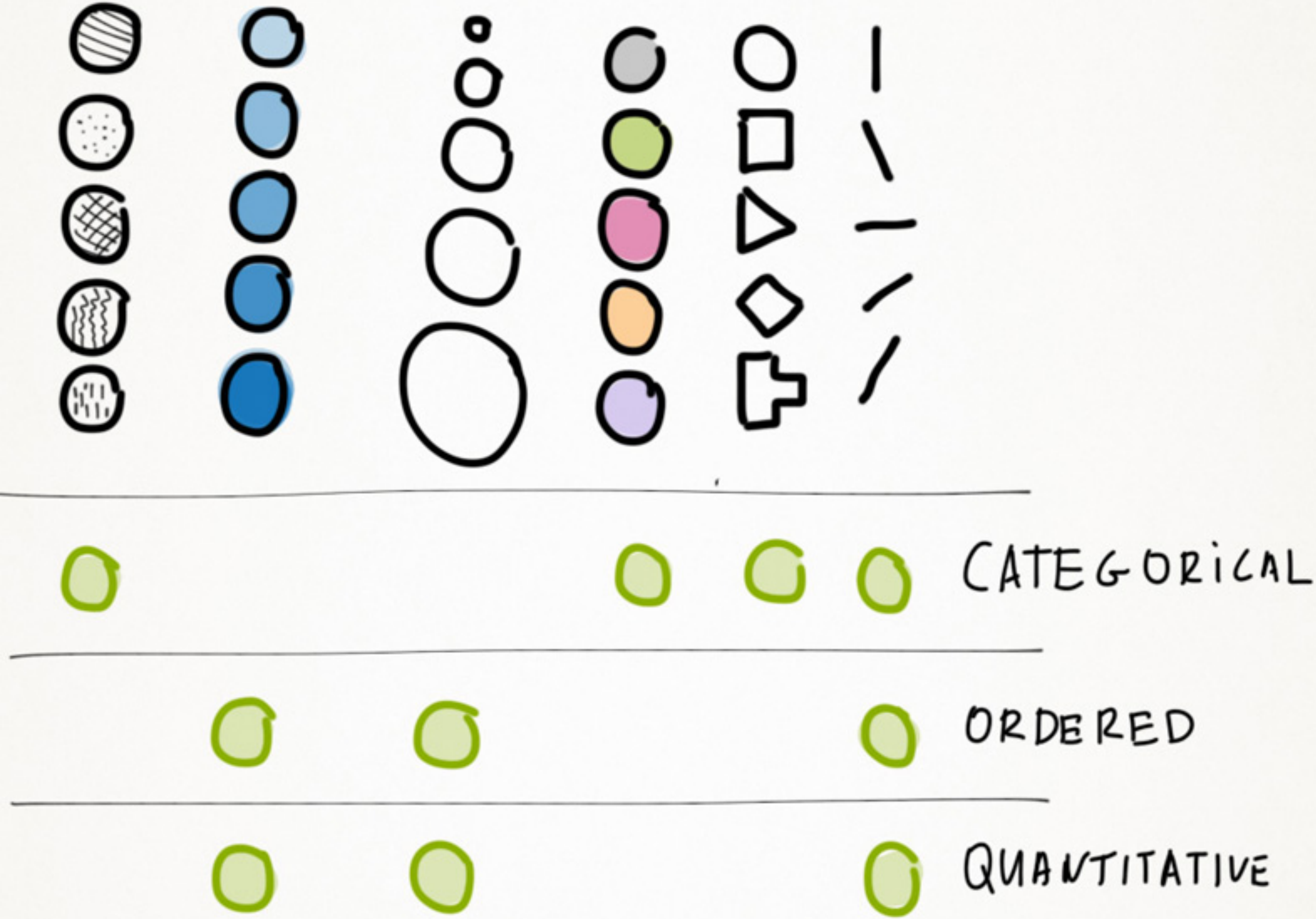
Categorical
Ordinal
Quantitative
- Interval
- Ratio

Small set/Large set
Continuous/Discrete
Diverging / Sequential

Color
Shape
Size
Texture
Angle
Thickness
Elongation
Curvature
Area

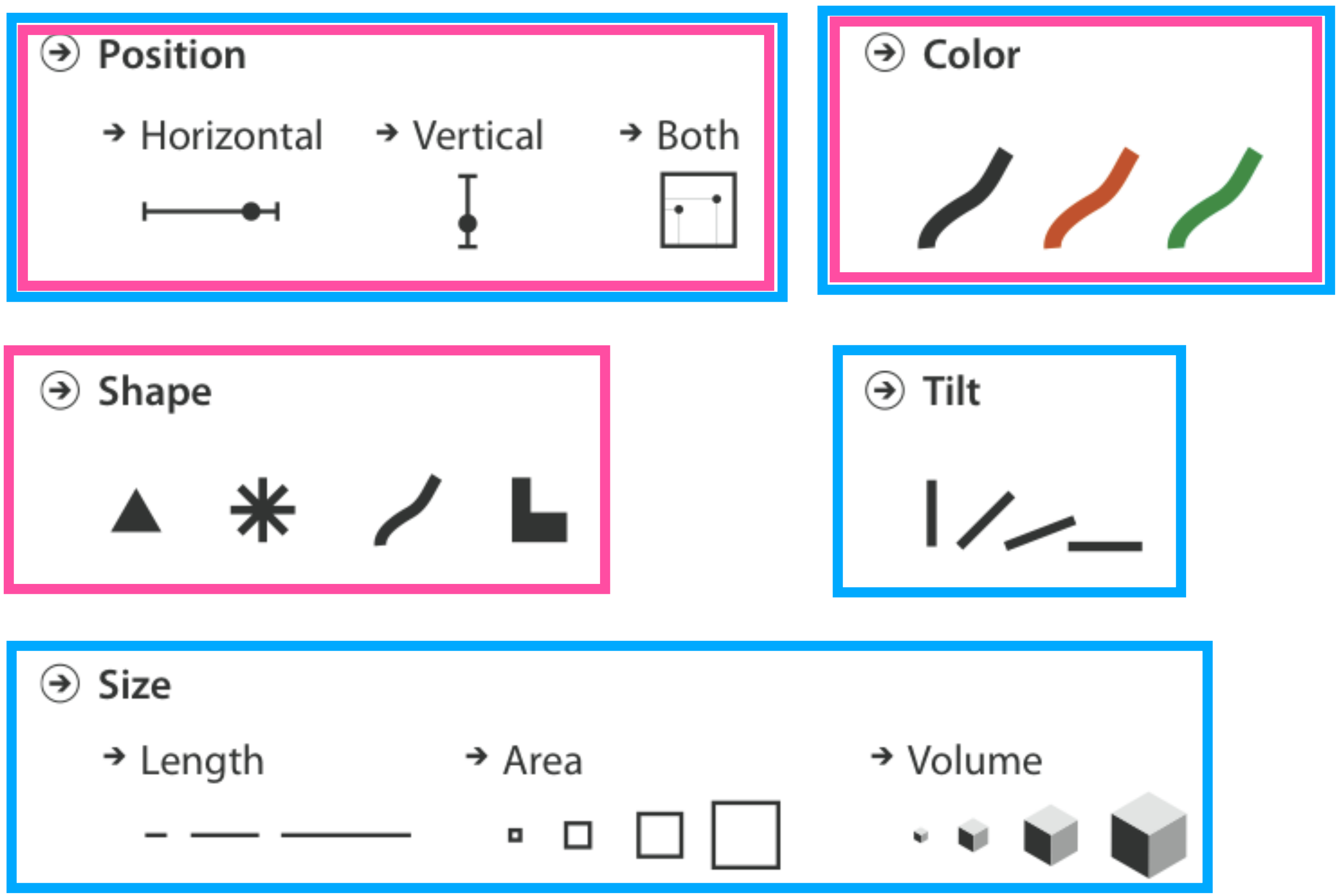
Bertin, 1967


O = Ordinal, Q = Quantitative
≠ = Differences ≡ = Similarities
VARIABLES OF THE IMAGE
XY 2 DIMENSIONS OF THE PLANE
Z
SIZE
VALUE
DIFFERENTIAL VARIABLES
TEXTURE
COLOR
ORIENTATION
SHAPE
OQ
OQ
O
O
POINT
LINE
AREA (ZONE)

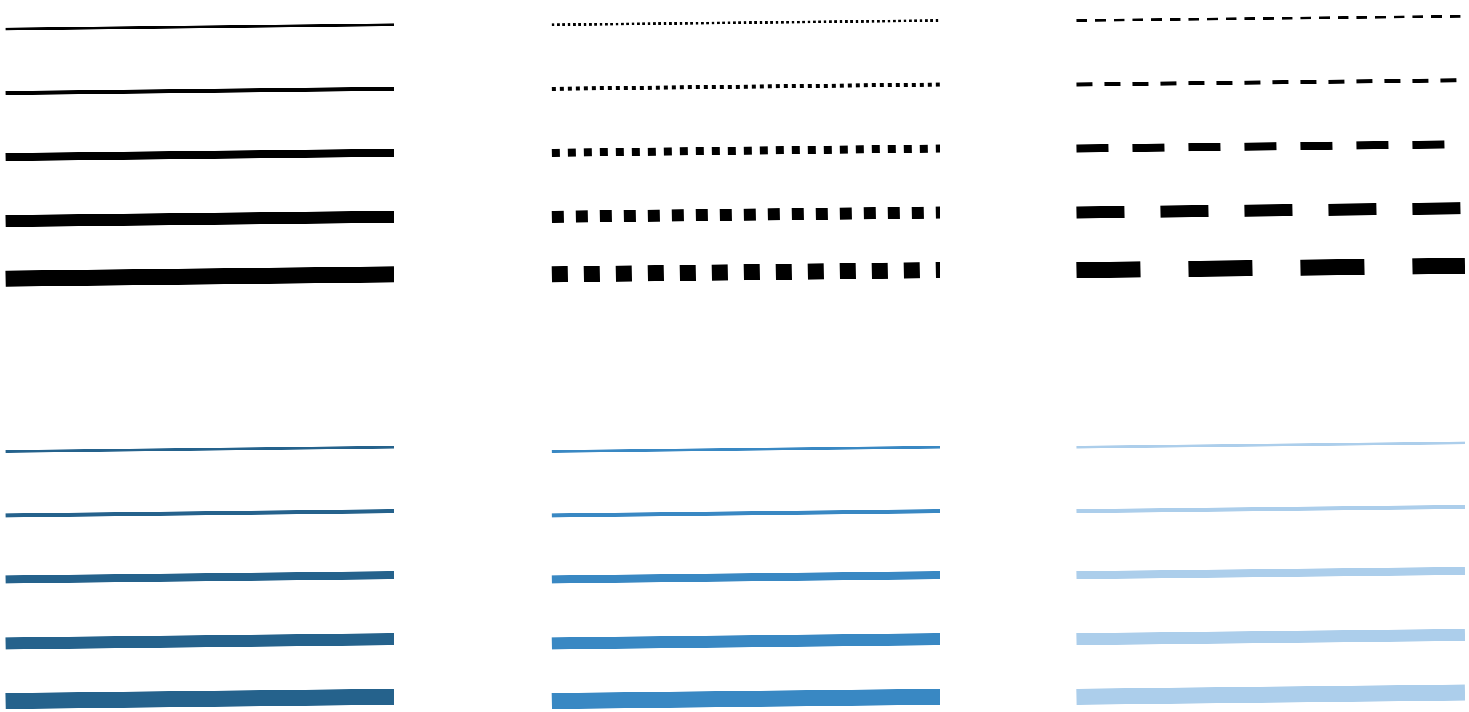

CATEGORICAL
ORDERED
QUANTITATIVE

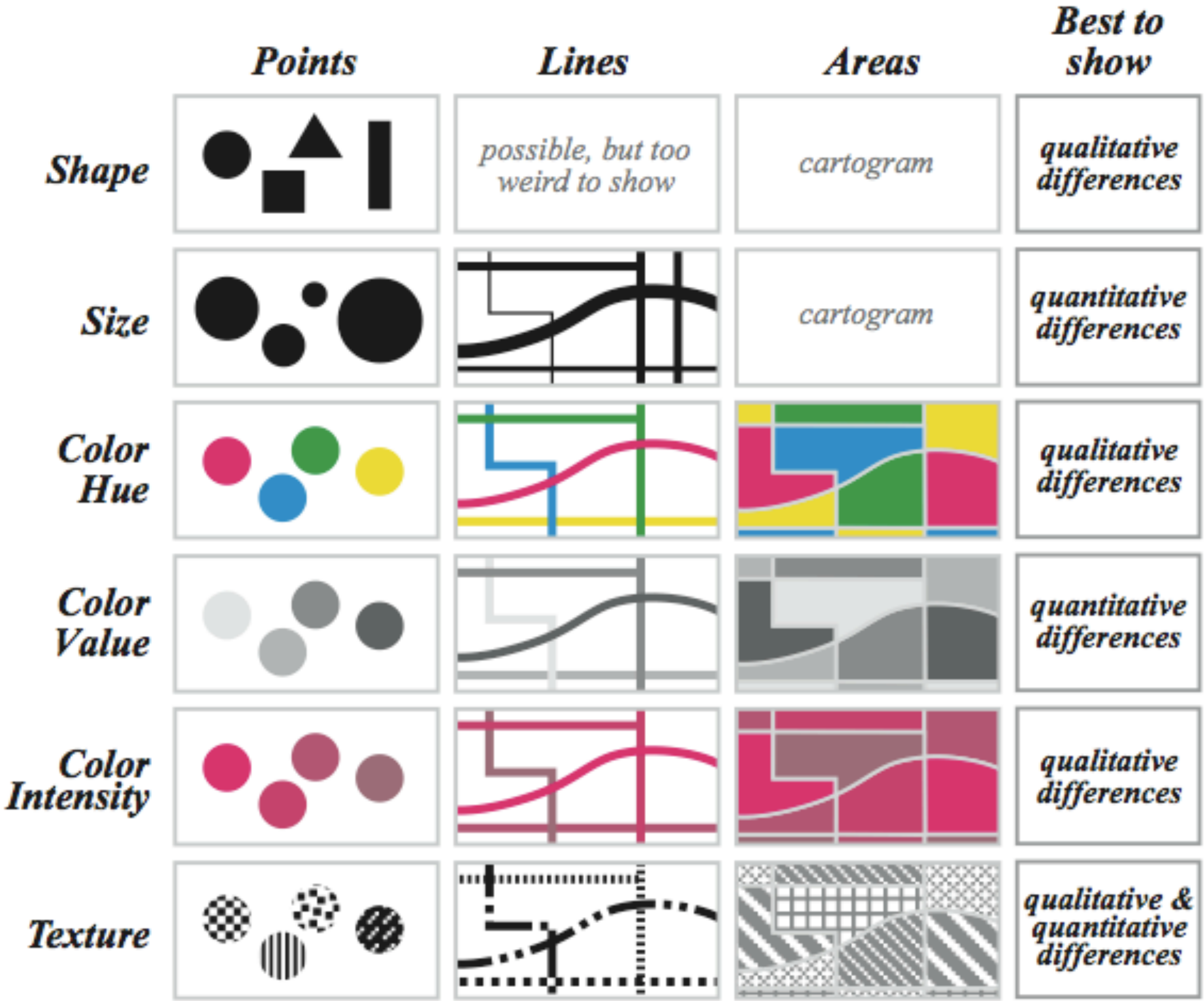
CHANNEL TYPES

identity (what or where) *magnitude (how much)*



A Gallery of Lines




Points
Lines
Areas
Best to show
Shape
possible, but too weird to show
cartogram
qualitative differences
Size
cartogram
quantitative differences
Color Hue
qualitative differences
Color Value
quantitative differences
Color Intensity
qualitative differences
Texture
qualitative & quantitative differences

Types suggest encodings

Continuous1 x Continuous2

Continuous1 -> X Position

Continuous2 -> Y Position

Sample over range

Mappings (values to positions)

Continuous1 x Continuous2

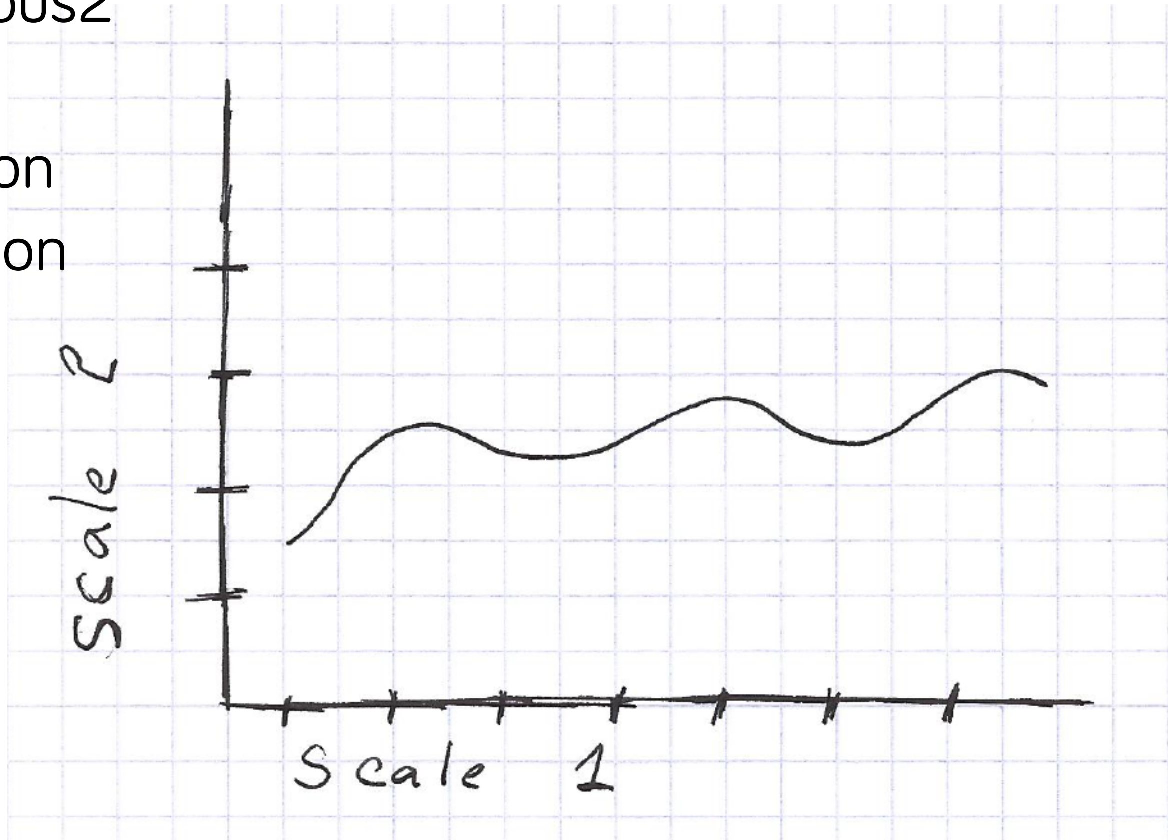
Continuous1 -> X Position

Continuous2 -> Y Position

Sample over range

Mappings
(values to positions)

Interval vs. Ratio?



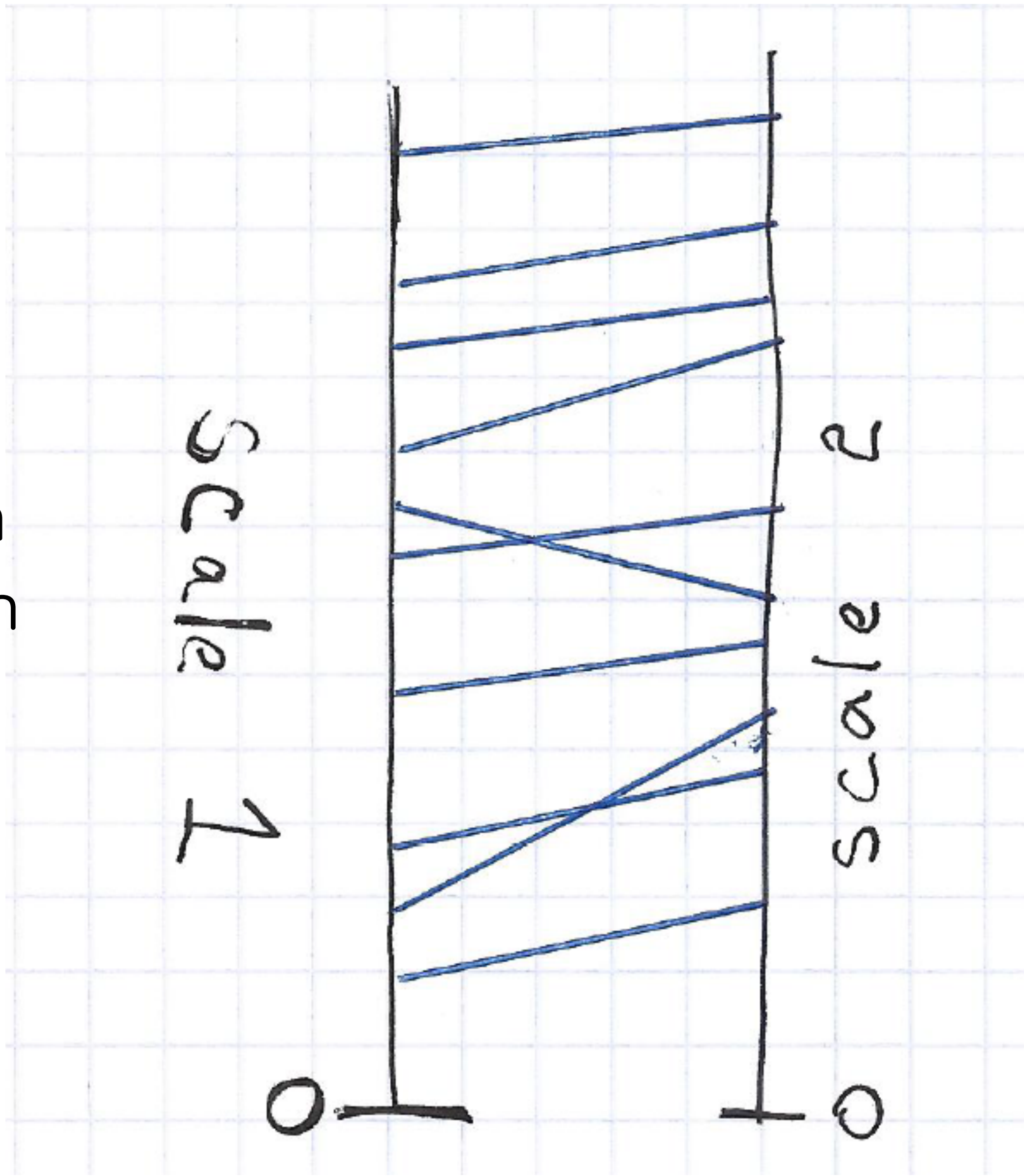
Continuous1 x
Continuous2

Continuous1 -> Y Position

Continuous2 -> Y Position

Specific Samples
(Nx(R,R))

Mappings
(values to positions)



Continuous1 x
Continuous2

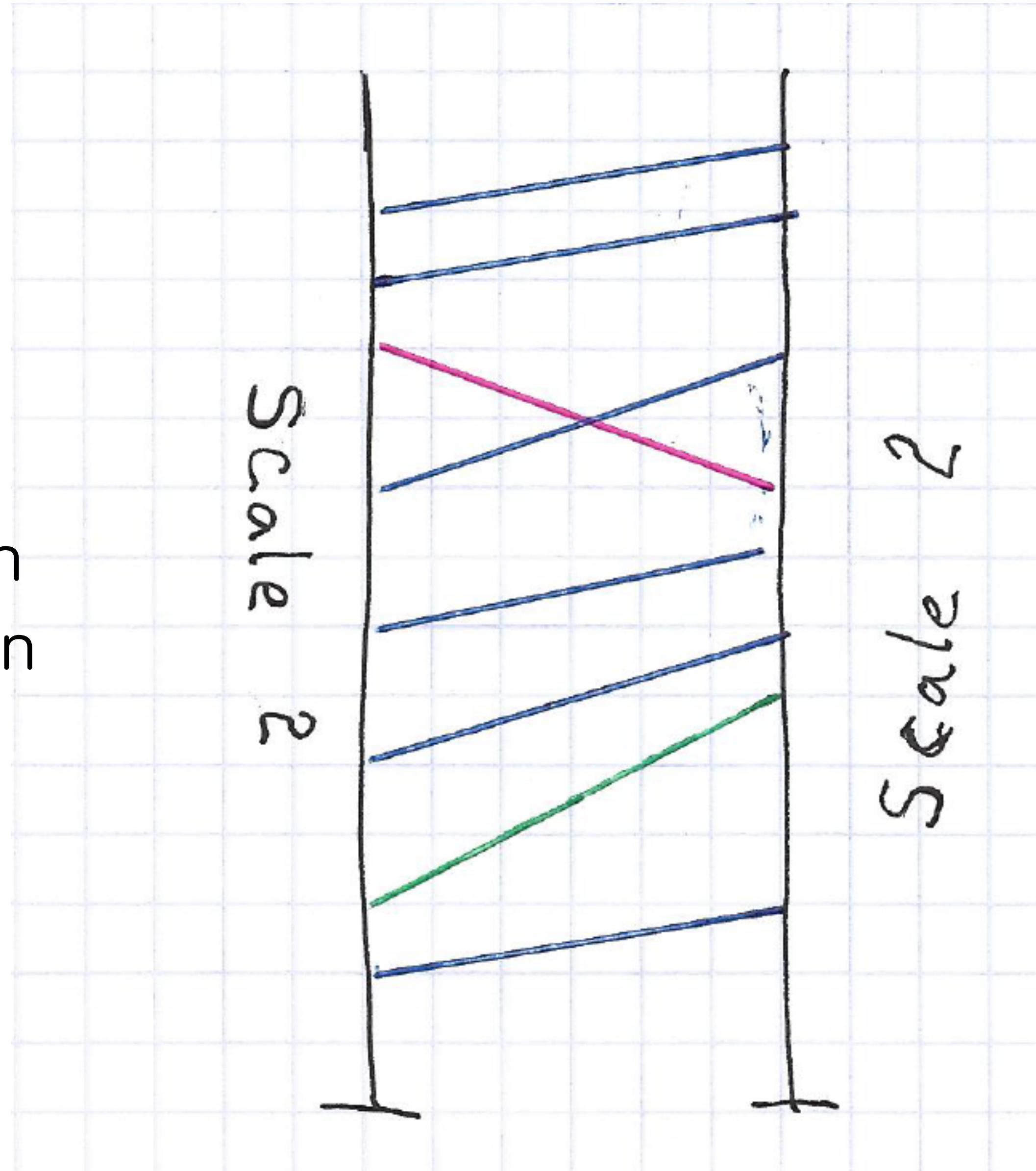
Continuous1 -> Y Position

Continuous2 -> Y Position

Specific Samples (Nx(R,R))

Interval -> Color

Derive!



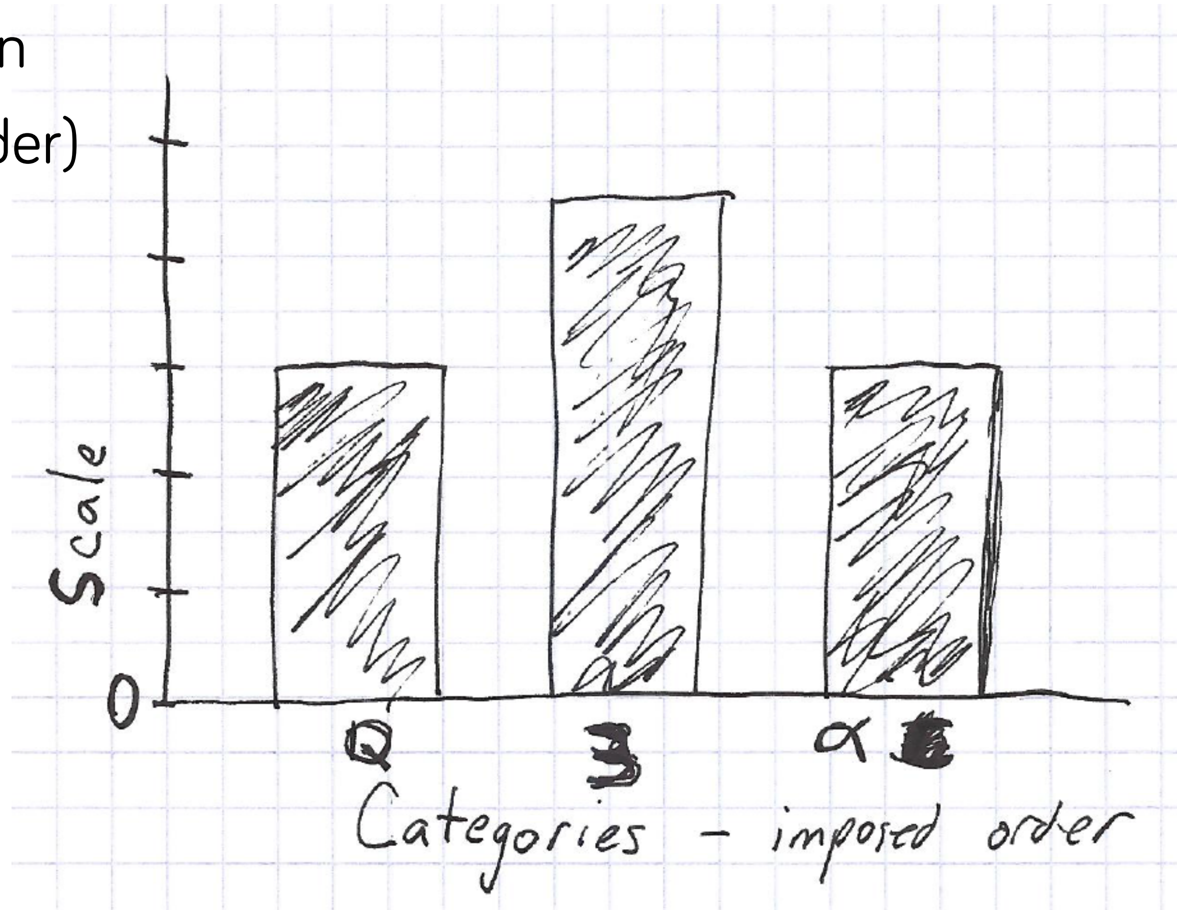
Categorical X Ratio

Categorical -> X position

(mapping? Imposed order)

Ratio -> Y position

(range mapping)



Categorical X Ratio

Categorical –> X position

(mapping? Imposed order)

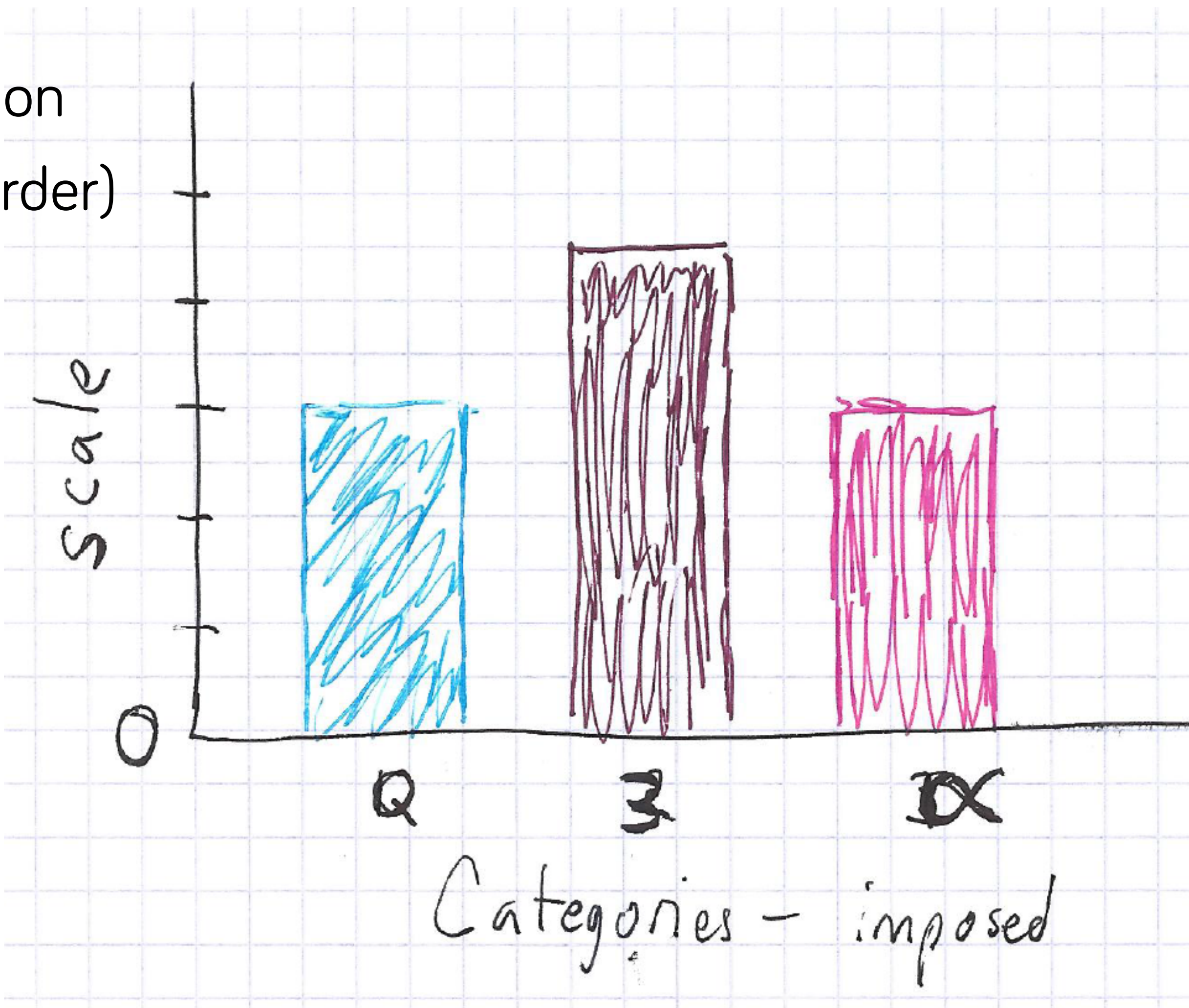
Ratio -> Y position

(range mapping)

Categorical -> Color

(redundant)

(mapping?)

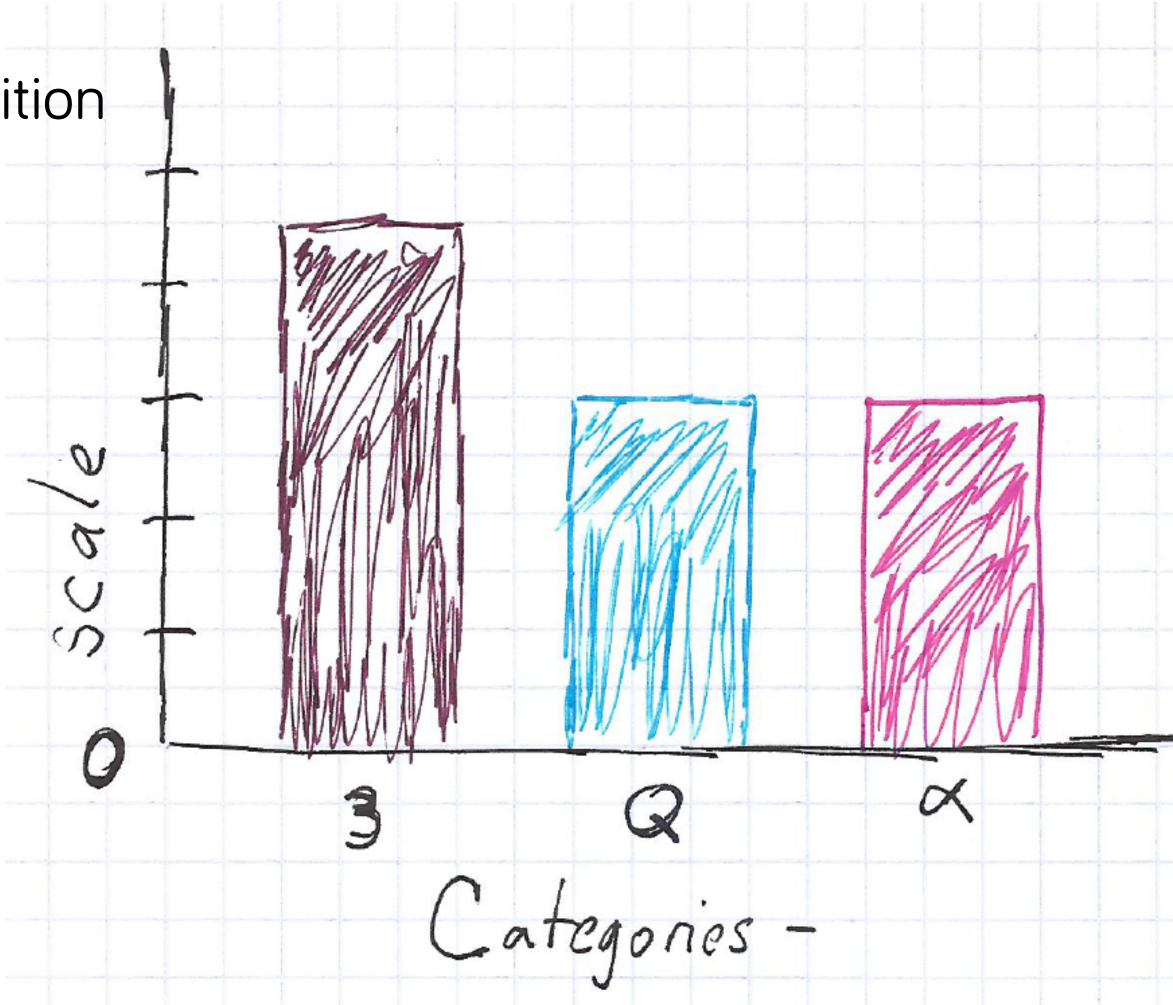


Categorical X Ratio

Ratio/Category-> X position
(sort by rank(Ratio))

Ratio -> Y position
(range mapping)

Categorical -> Color
(redundant)
(mapping?)



Where is this going?

Principles for choosing

    encodings

    mappings

Match designs to data

Not necessarily designs to task

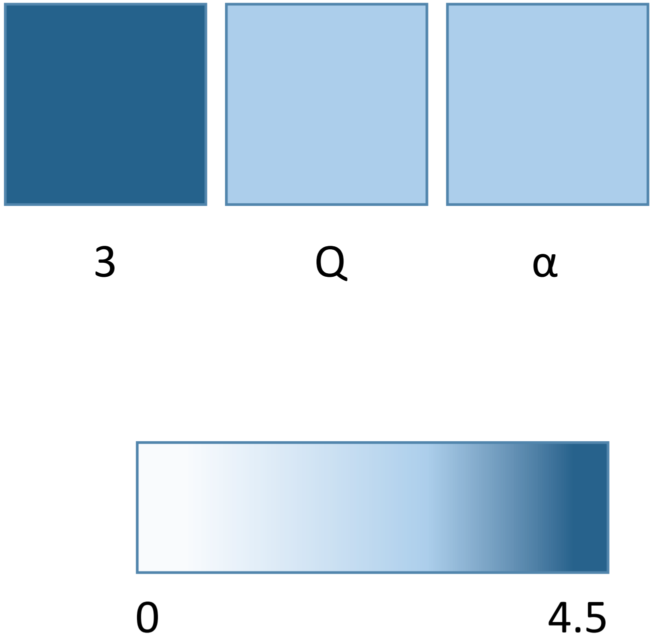
Categorical X Ratio

Categorical -> X position

(mapping? **Order by Ratio**)

Ratio -> Color

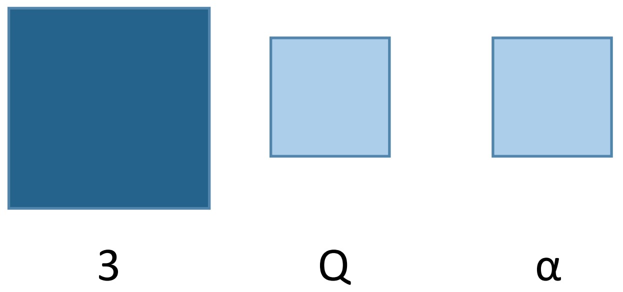
(color encoding)



Categorical X Ratio

Categorical -> X position

(mapping? **Order by Ratio**)



Ratio -> Color

(color encoding)

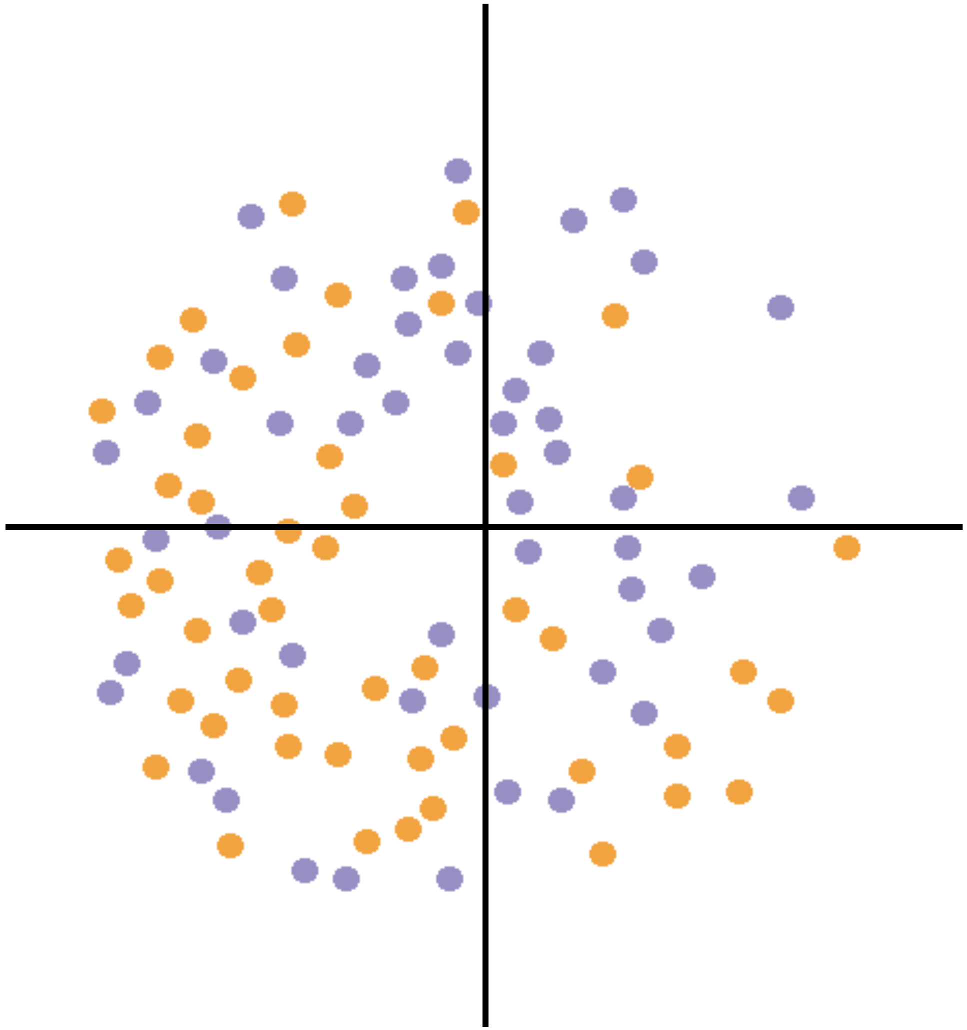
Ratio -> Size / Area

(mapping / ramp)

Categorical X

(Categorical, Ratio, Ratio)

Multi-class scatterplot

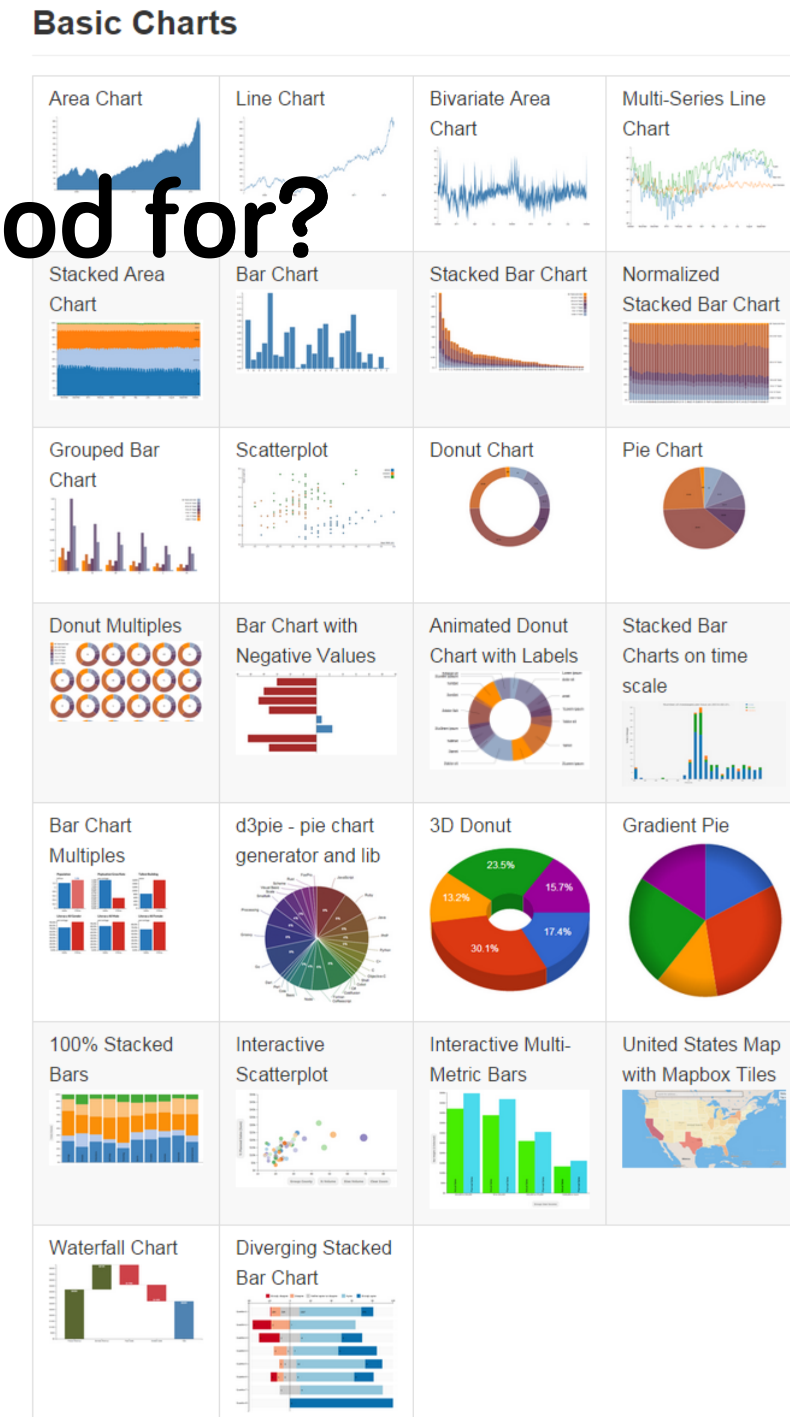


What data is this good for?

Even more data distinctions

Part/Whole

Regions



What do you want to do with these?

Absolute judgment

Relative judgment

Identify / find / match against key

Form groups / regions

Count / quantify

Average / estimate statistics

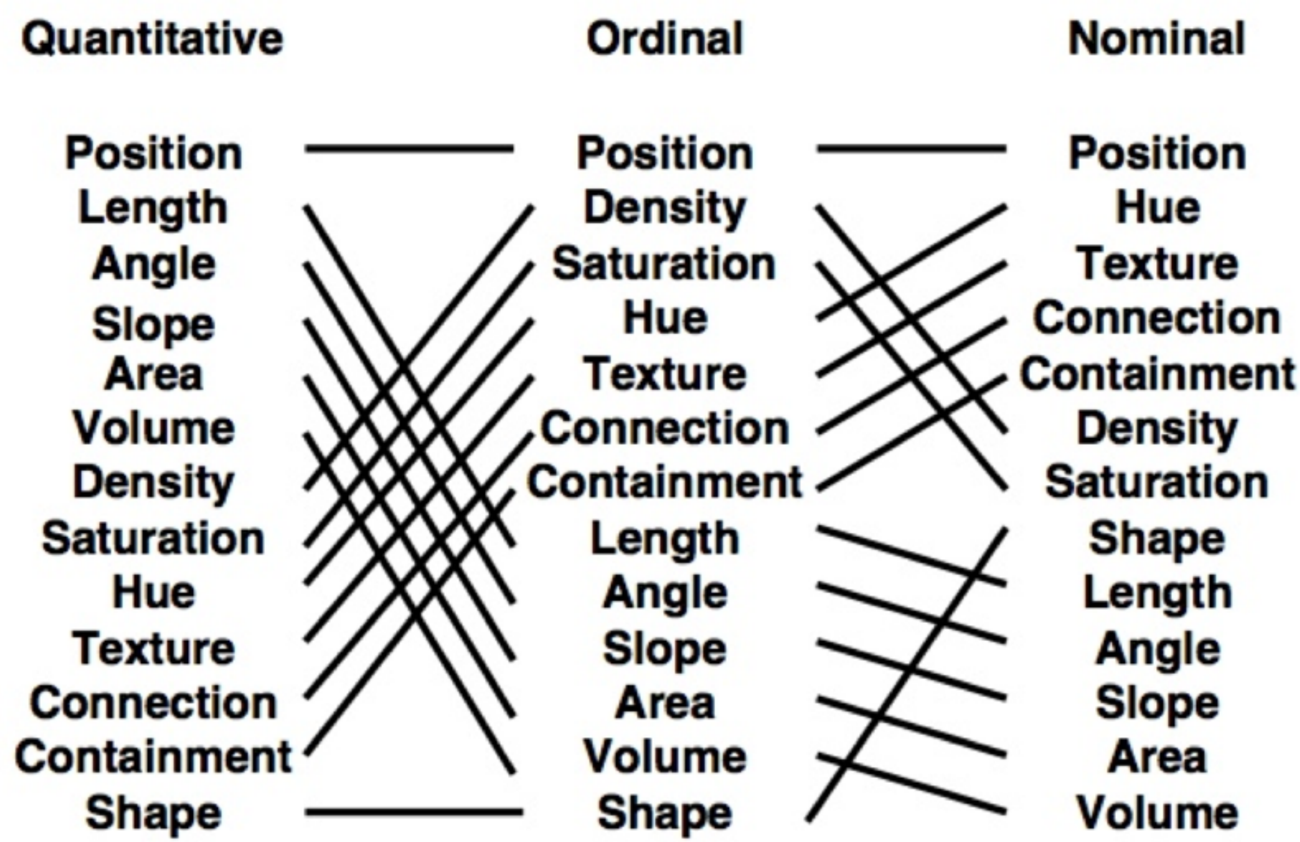
Pick encodings that...

Match your data type

Are good for the low-level task

Work well with other things

Mackinlay, 1986



Cleveland & McGill, 1984

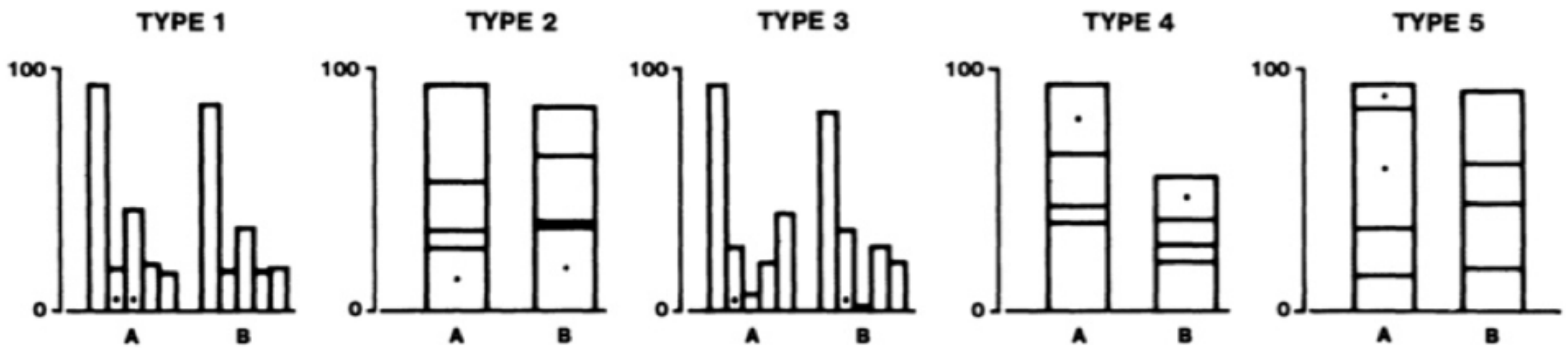


Figure 4. Graphs from position–length experiment.

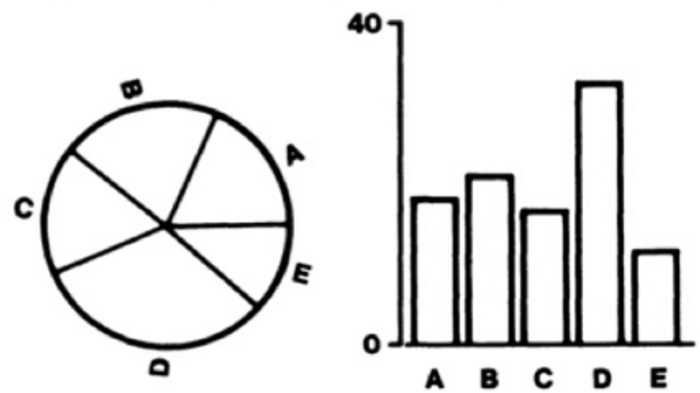


Figure 3. Graphs from position–angle experiment.

Heer & Bostock, 2010

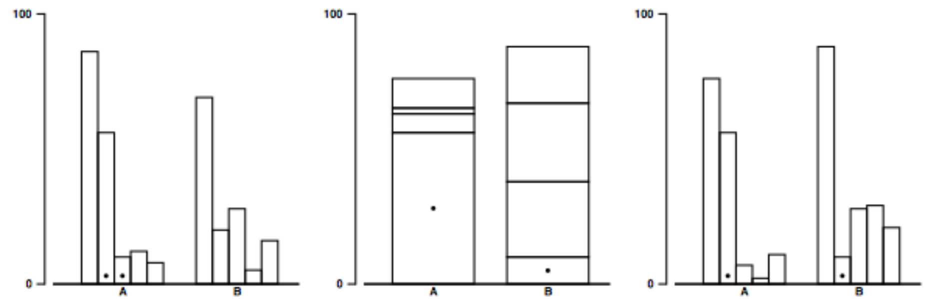


Figure 1: Stimuli for judgment tasks T1, T2 & T3. Subjects estimated percent differences between elements.

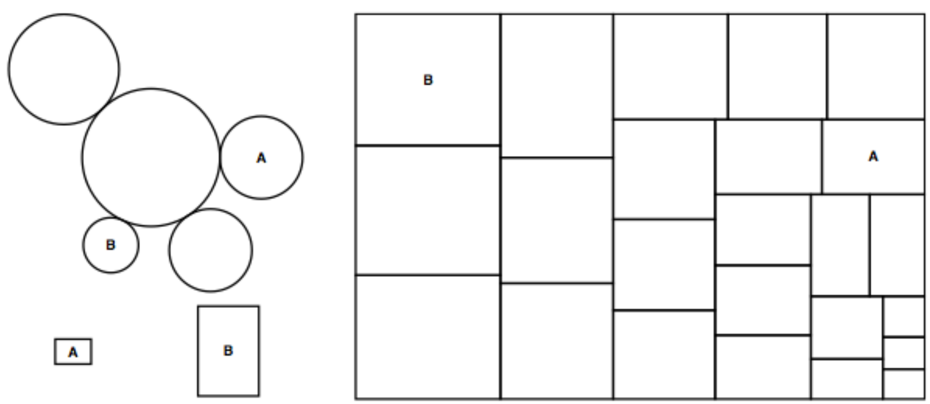


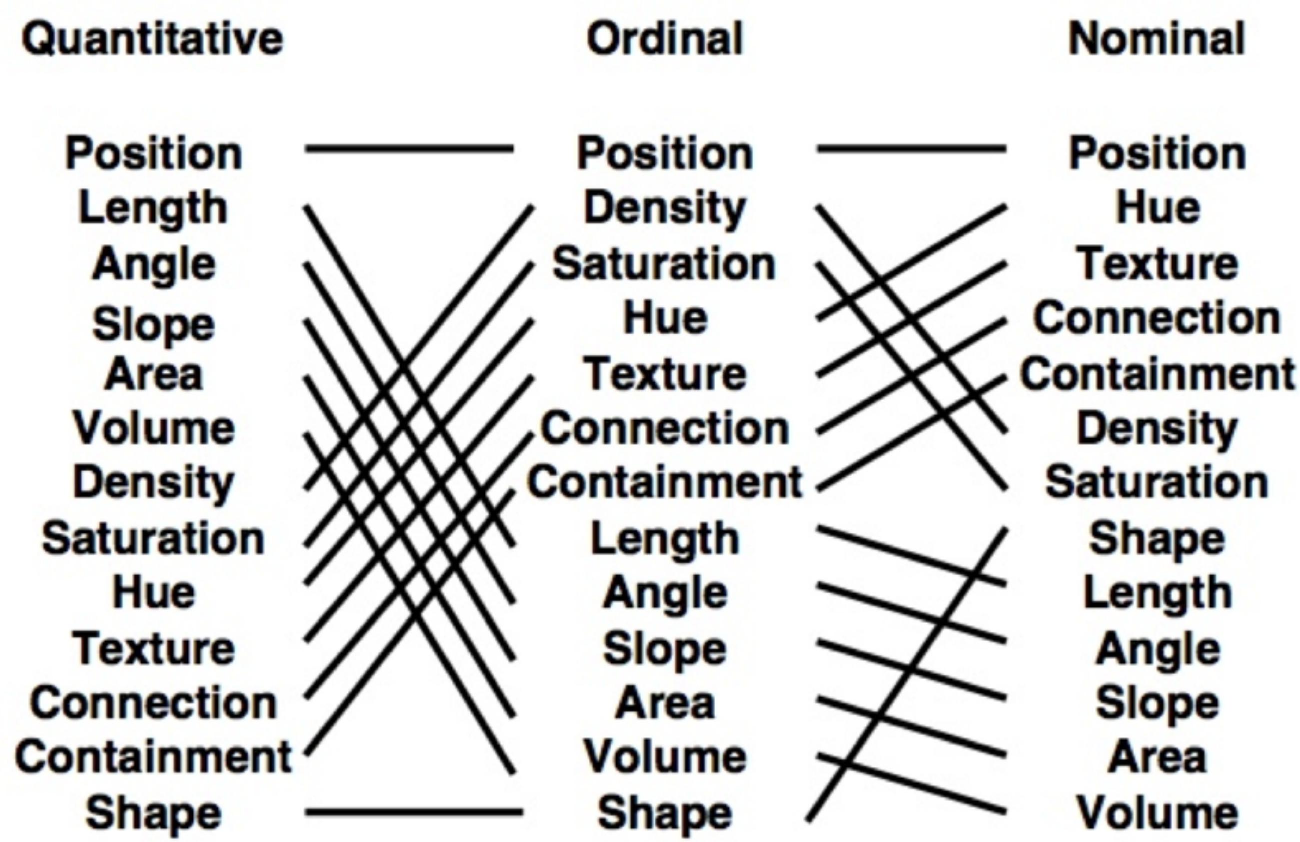
Figure 2: Area judgment stimuli. Top left: Bubble chart (T7), Bottom left: Center-aligned rectangles (T8), Right: Treemap (T9).

What do you want to do with these?

Absolute judgment
Relative judgment
Identify / find / locate
match against key
Form groups / regions

Count / quantify
Average

Mackinlay, 1986

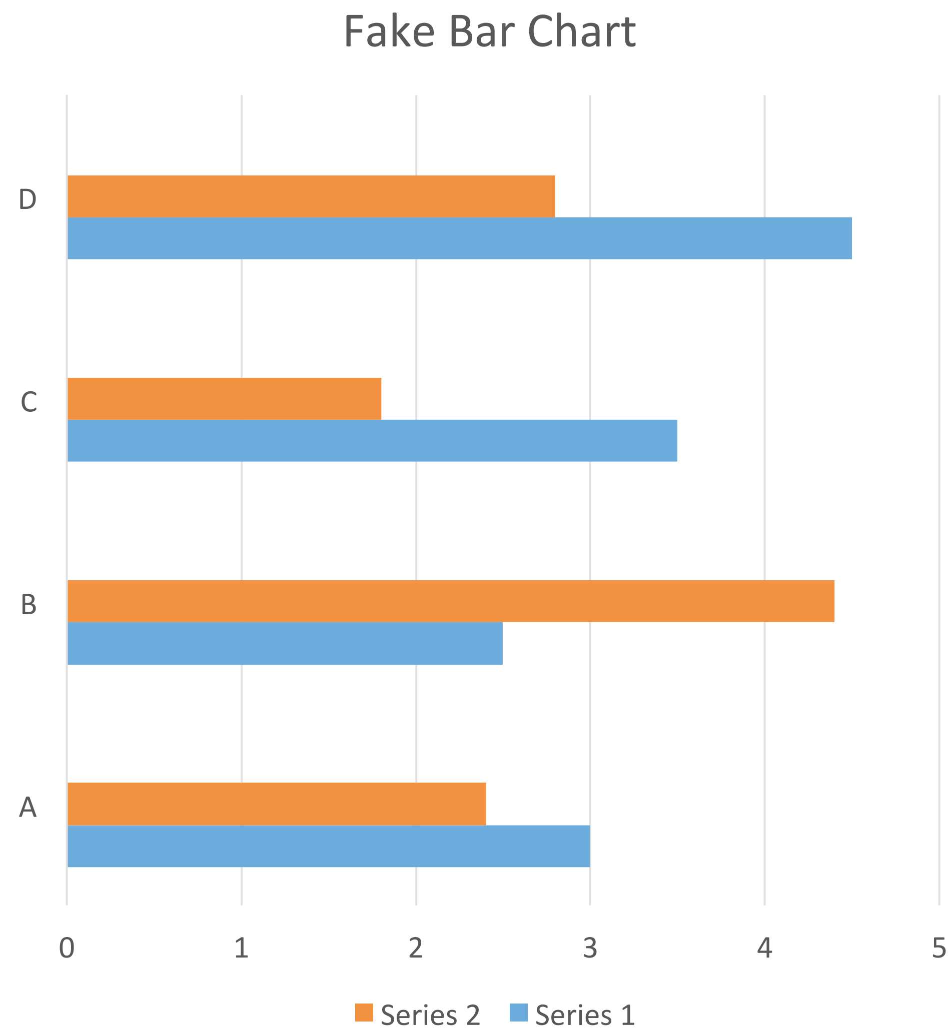


What **task(s)** is each of these good for?

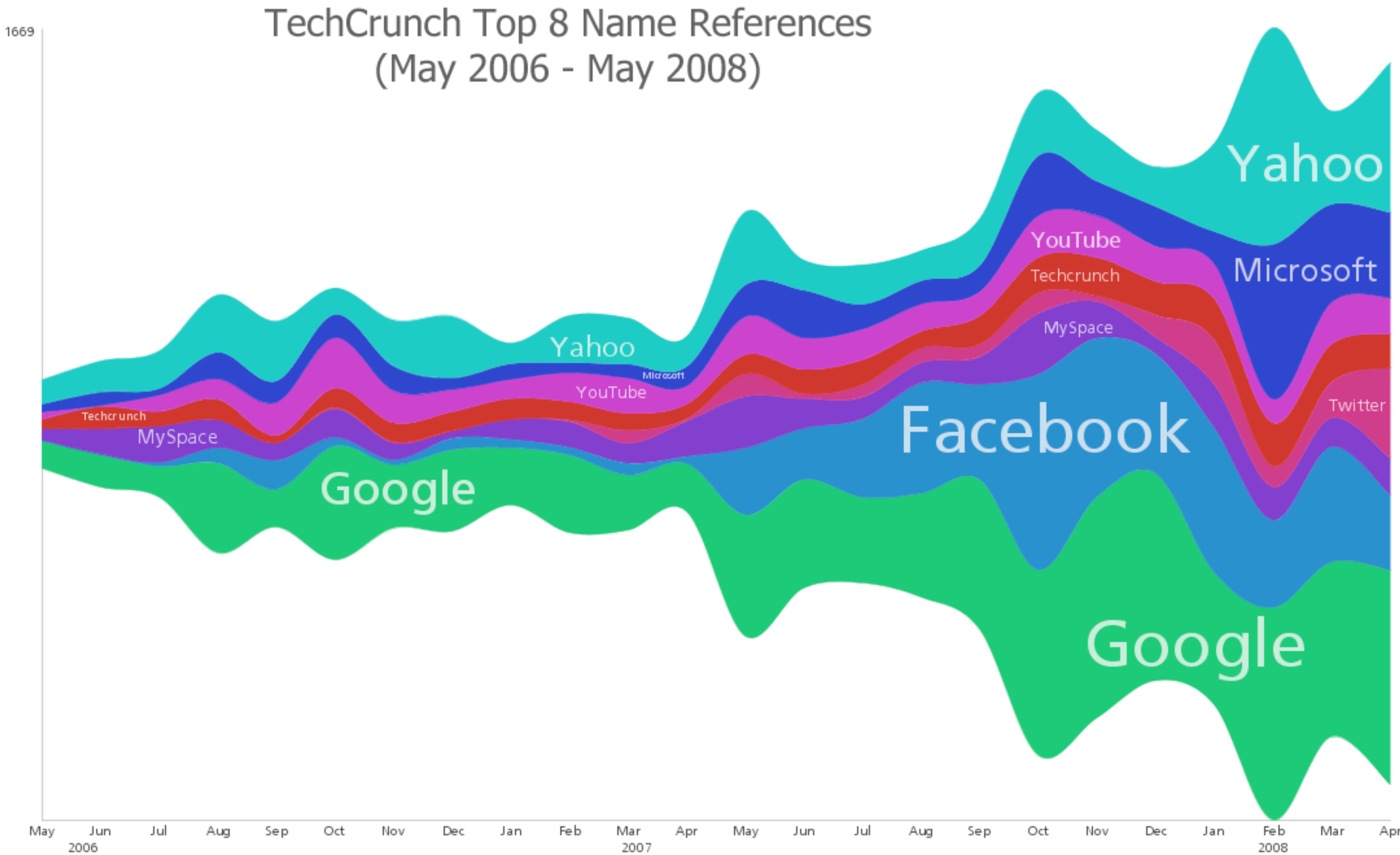
Position is good for most things

Absolute judgment
Relative judgment
Identify / find / locate
match against key
Form groups / regions

Count / quantify
Average



Position: Relative vs. Absolute



Neoformix

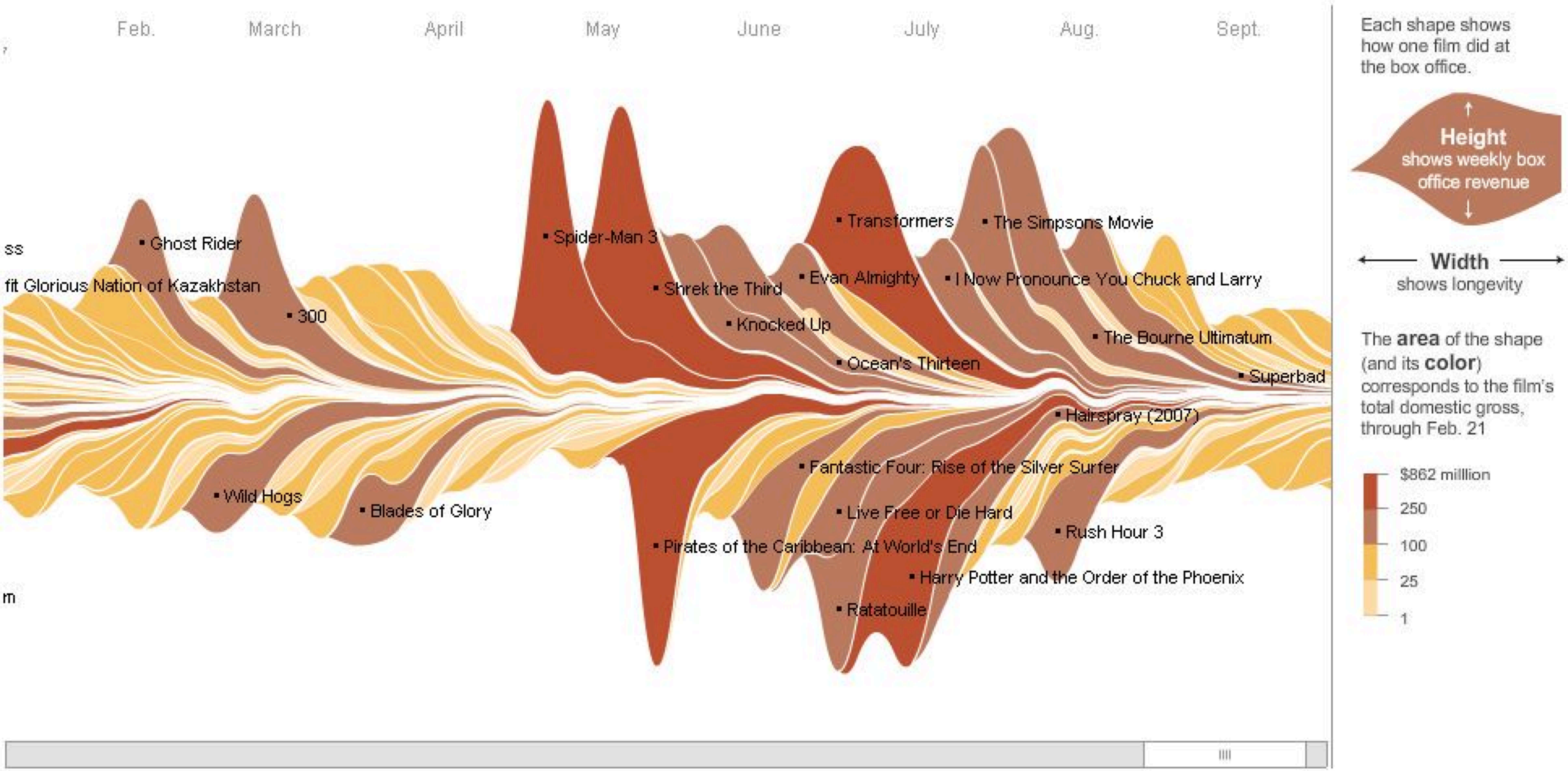
February 23, 2008

SIGN IN TO E-MAIL OR SAVE THIS | FEEDBACK

The Ebb and Flow of Movies: Box Office Receipts 1986 — 2008

Summer blockbusters and holiday hits make up the bulk of box office revenue each year, while contenders for the Oscars tend to attract smaller audiences that build over time. Here's a look at how movies have fared at the box office, after adjusting for inflation.

Find Movie Go

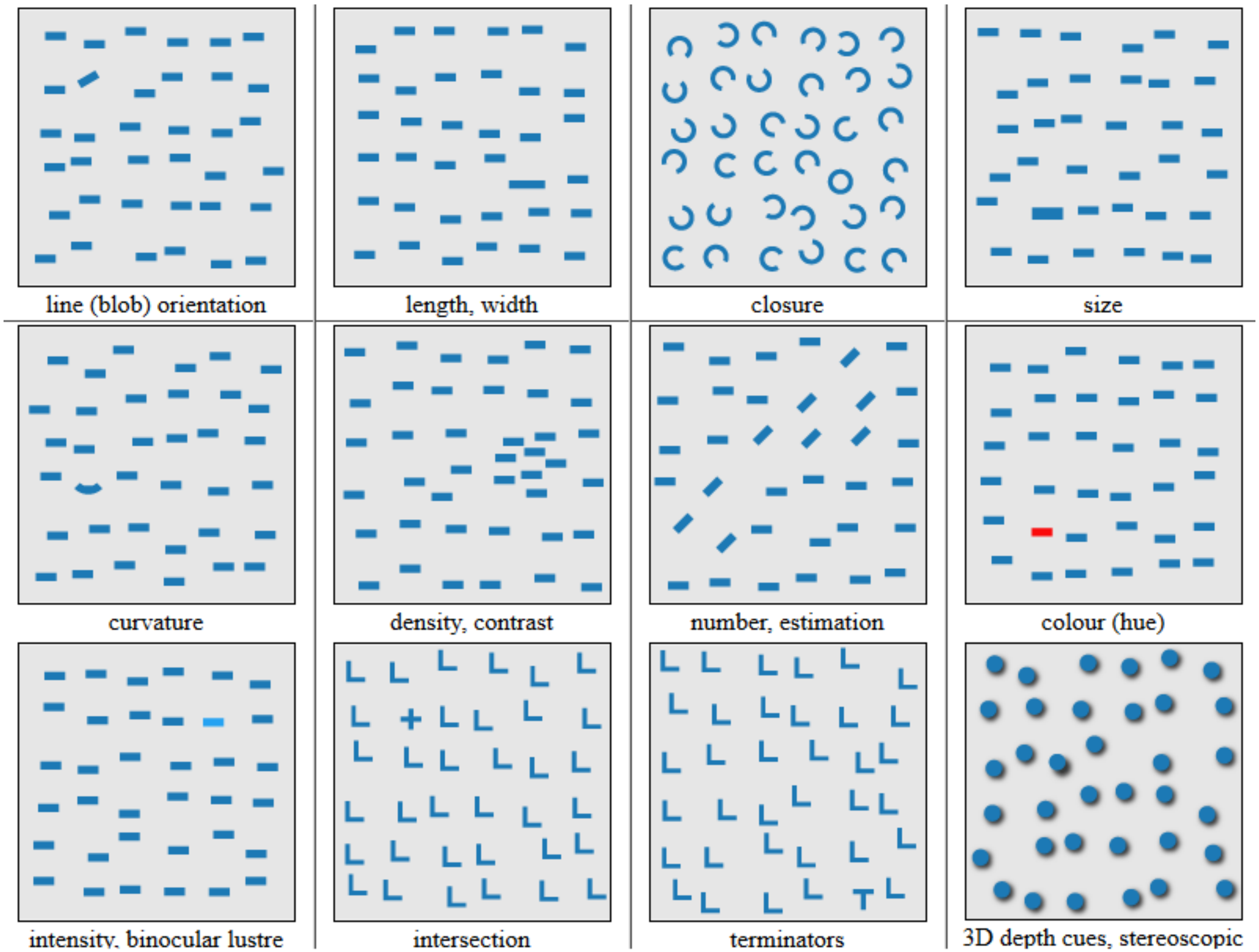


What do you want to do with these?

Absolute judgment
Relative judgment
Identify / find / locate
match against key
Form groups / regions

Count / quantify
Average

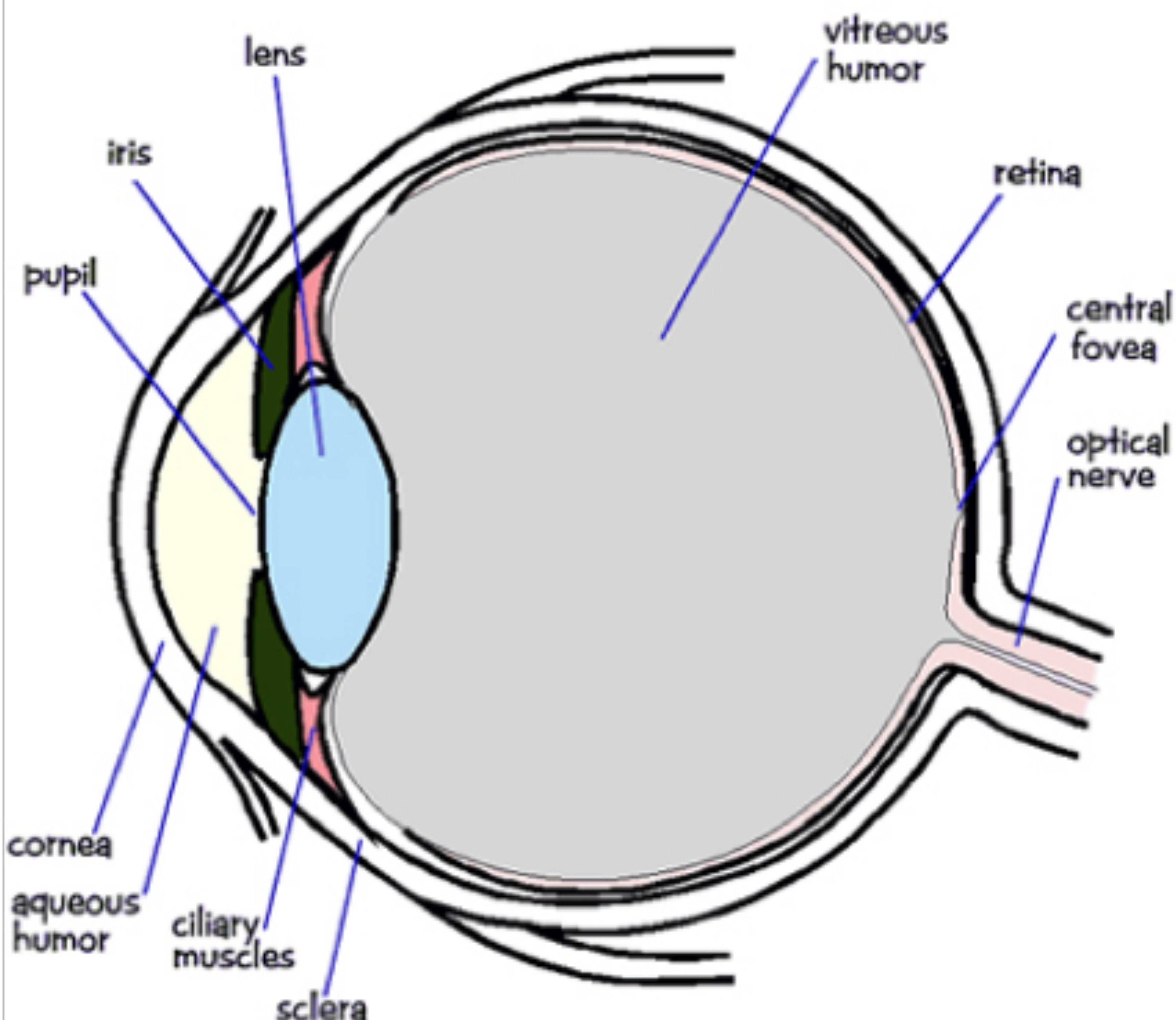
Most things OK for identification
(given sufficient contrast)

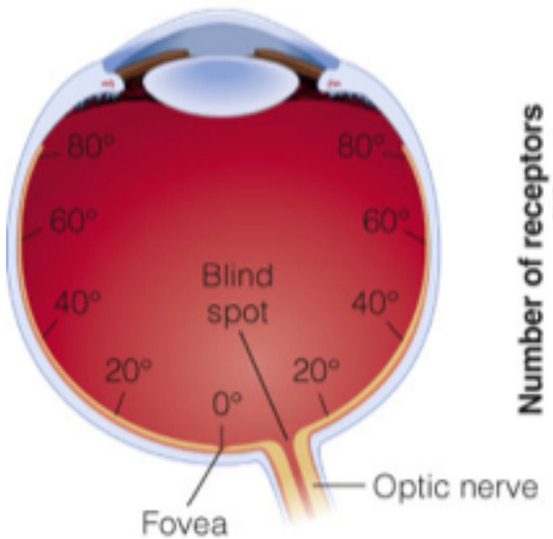


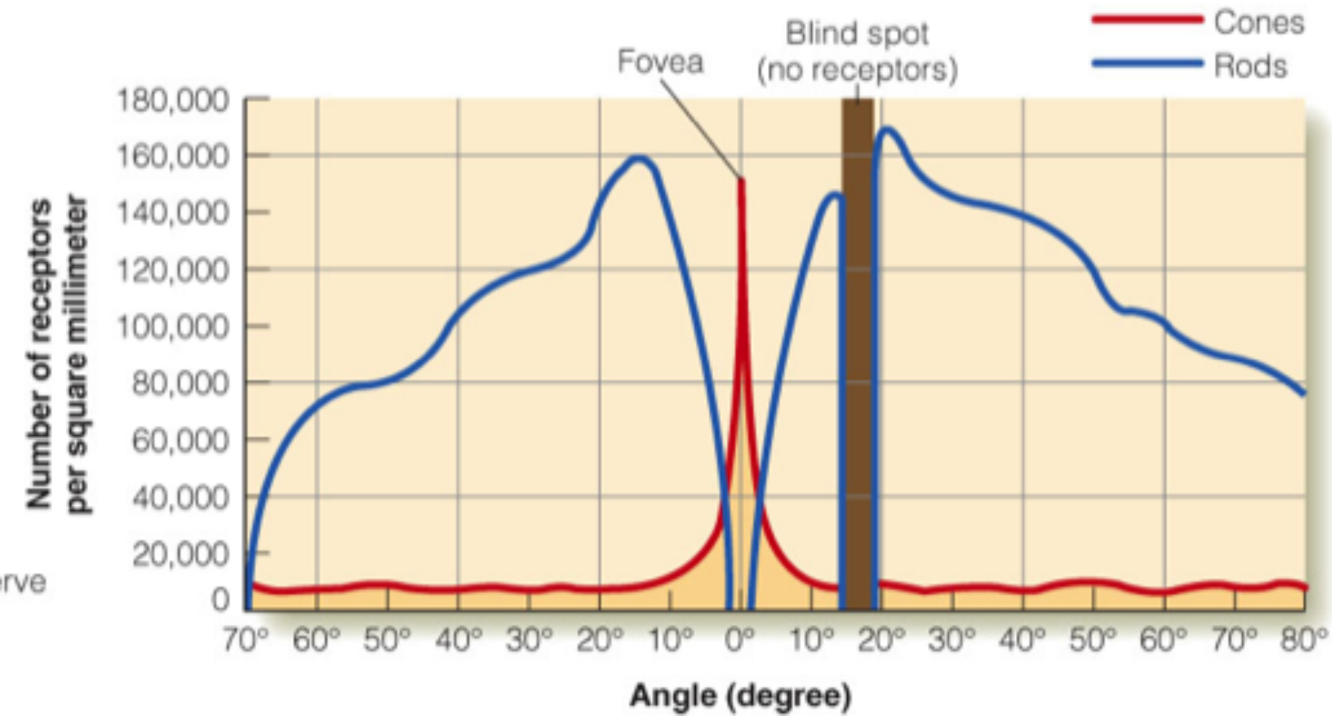
Healy web page
Popout examples

Human Perception

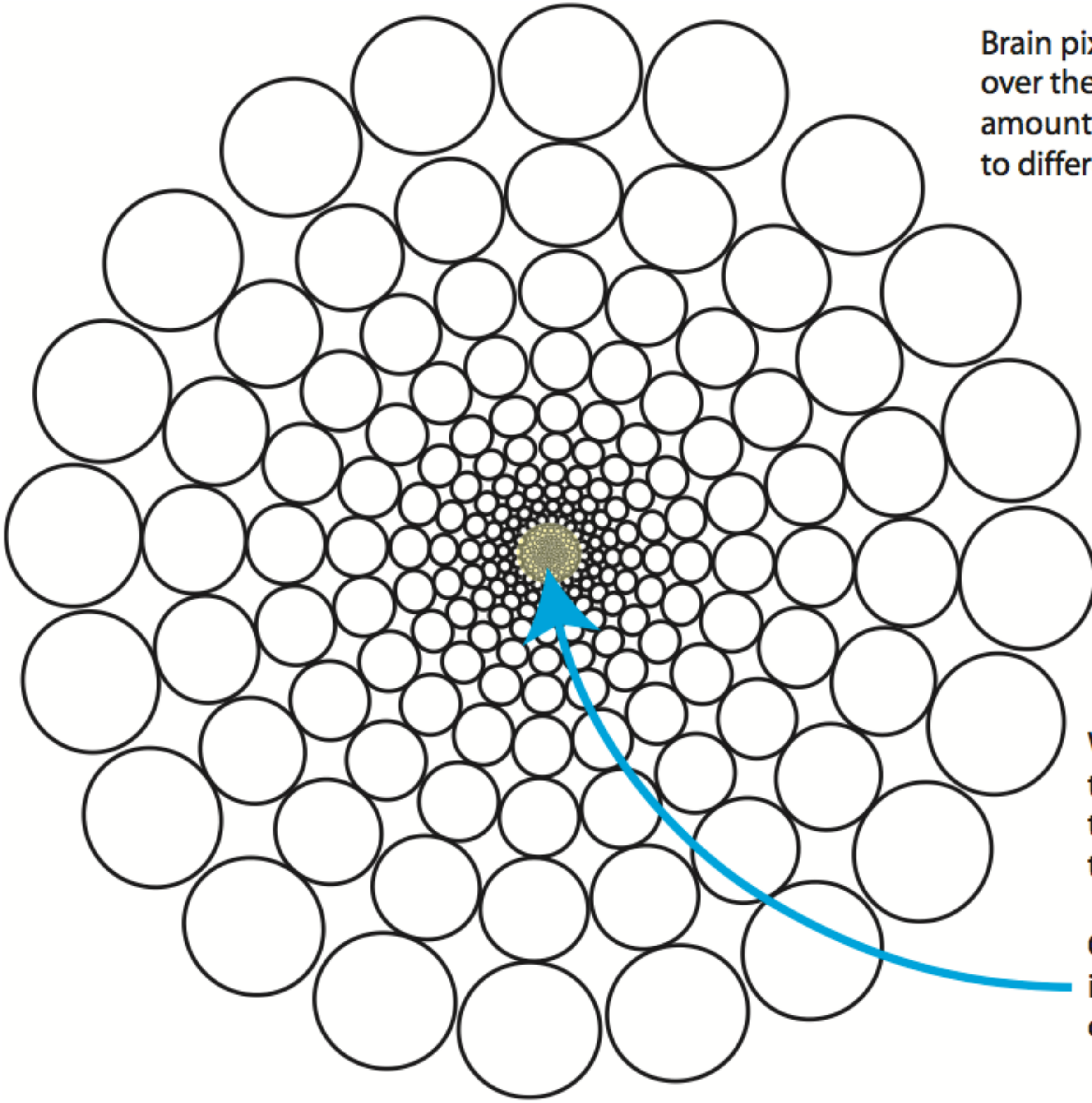
A little bit on how we see


lens
vitreous humor
iris
retina
pupil
central fovea
optical nerve
cornea
aqueous humor
ciliary muscles
sclera





E. B. Goldstein "Sensation and Perception"
(Adapted from Lindsay & Norman, 1977)



Brain pixels vary enormously in size over the visual field. This reflects differing amounts of neural processing power devoted to different regions of visual space.

At the edge of the visual field we can only barely see something the size of a fist at arm's length.

We can resolve about 100 points on the head of a pin held at arm's length in the very center of the visual field called the fovea.

Over half of our visual processing power is concentrated in a slightly larger area called the parafovea.

Ware 2010

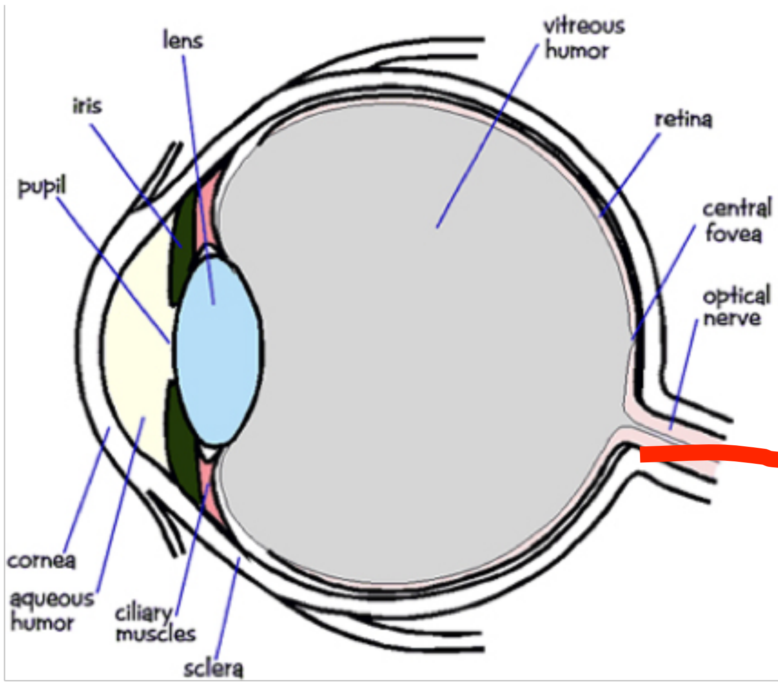
Some implications...

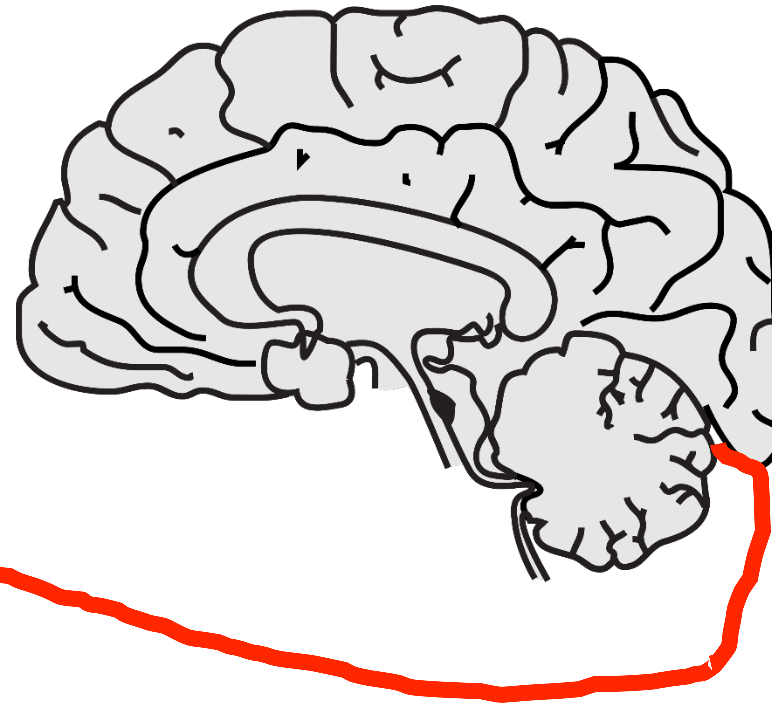
We see very little at a time

We build up an image over time
using memory

Fast parallel sensing
Slow serial search

Some computer systems thinking...





Lots of sensors
(retinal cells)

Lots of computing
(neurons...)

And a thin cable connected them!

Theory:

Clever coding gets lots of information in a little space

Practice:

Some things can be very efficient

Others have work-arounds

Other bottlenecks

Throughout the stages…

- Competition for Identification
- Competition for selection

Very limited short term memory

Bad news: resource limited

Good news: some parallel mechanisms

Your eyes do some things really well

Your brain needs help with some simple things

Attention

We control where to direct those resources

Pre-Attention

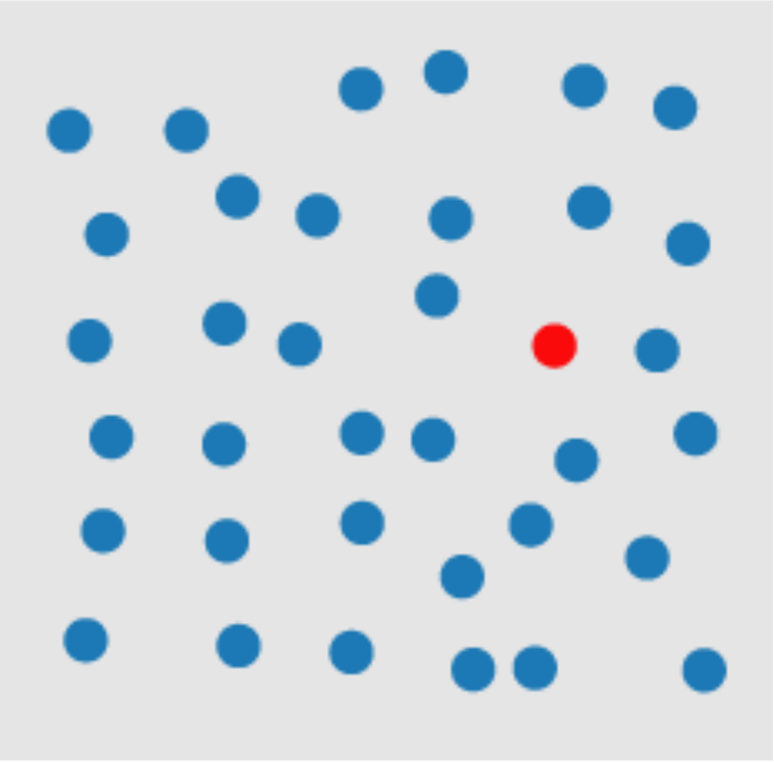
A misnomer!

Involuntary?

Pre-cognitive?

Efficient?

Parallel?



Pre-Attention

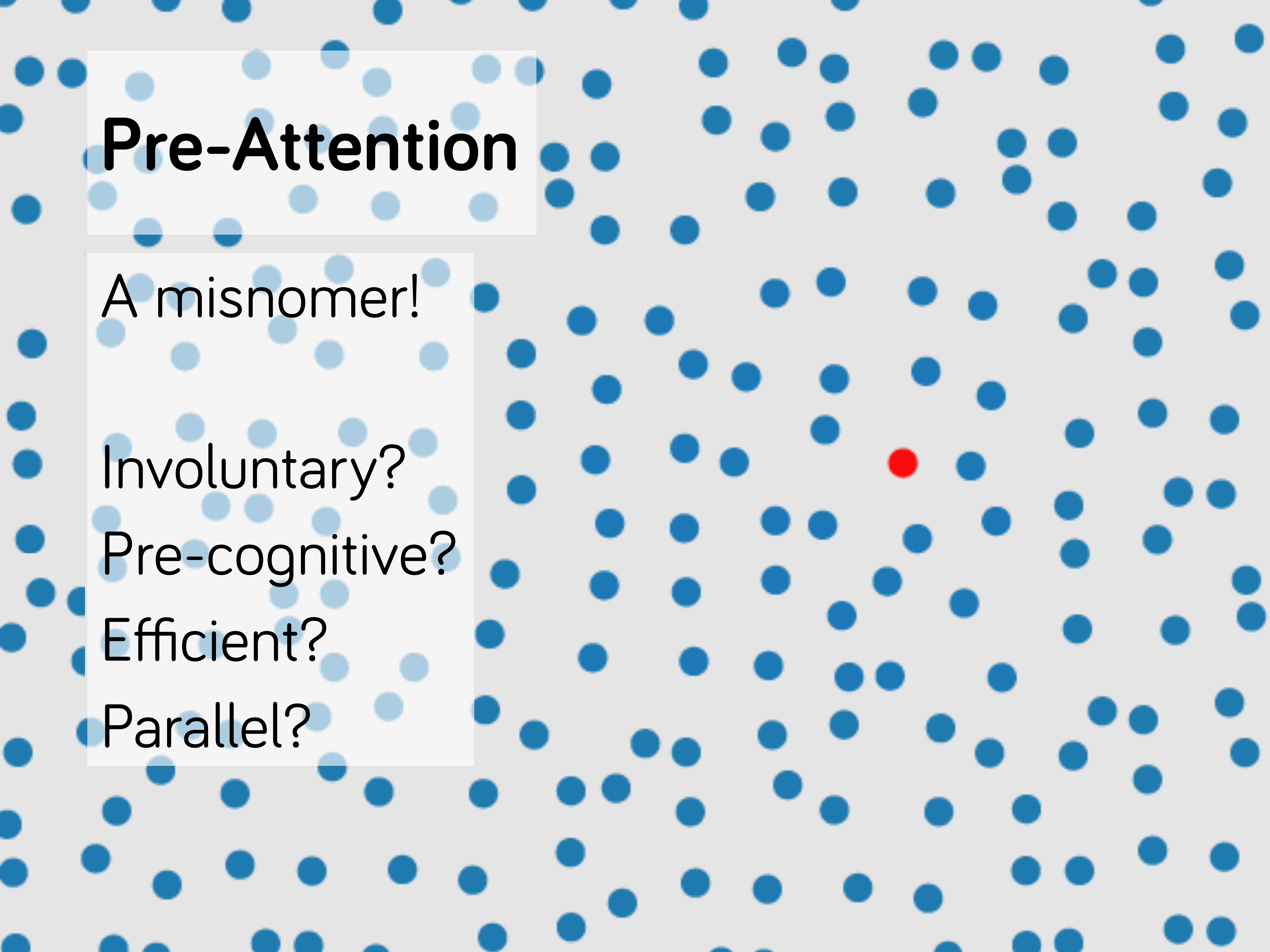
A misnomer!

Involuntary?

Pre-cognitive?

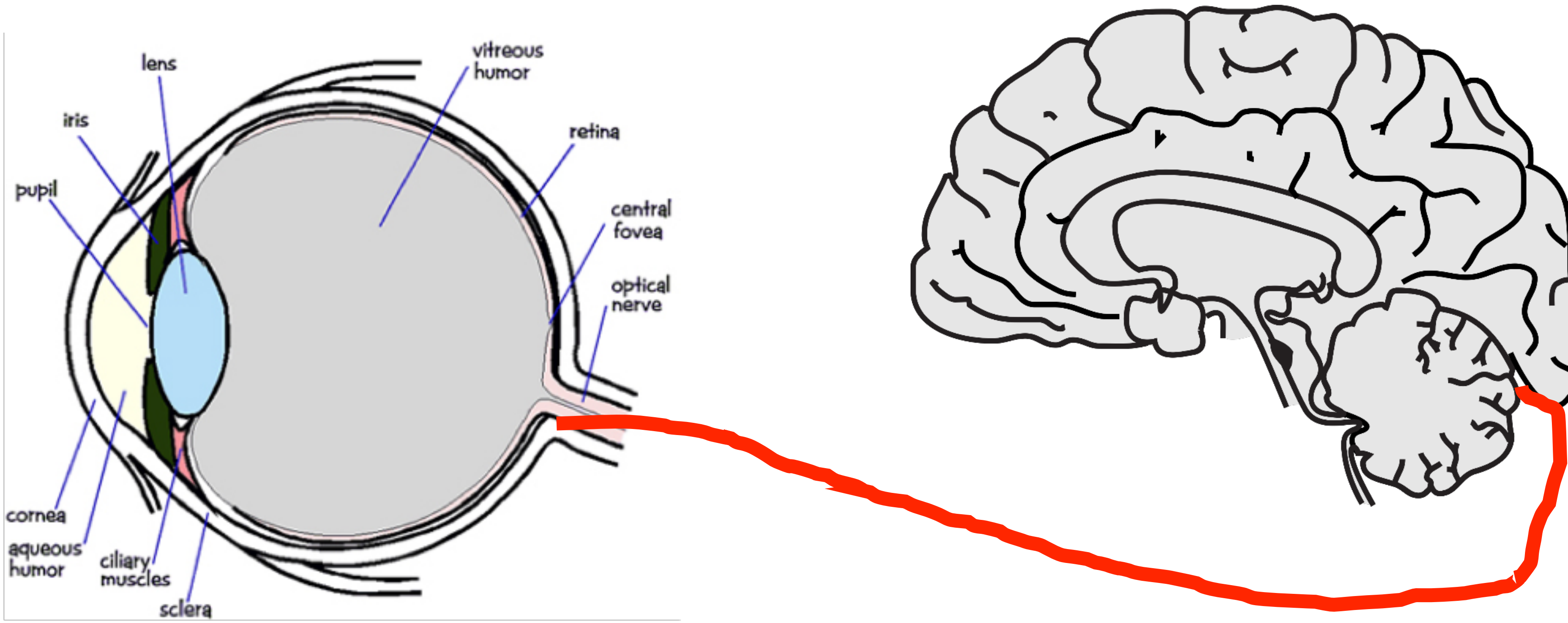
Efficient?

Parallel?



Other Perception Tricks

Some computer systems thinking...



Lots of sensors
(retinal cells)

Lots of computing
(neurons...)

What can you do efficiently?

Locate something that pops out

Sense ensemble properties

estimate numerousity

estimate statistics

Encodings for estimation?

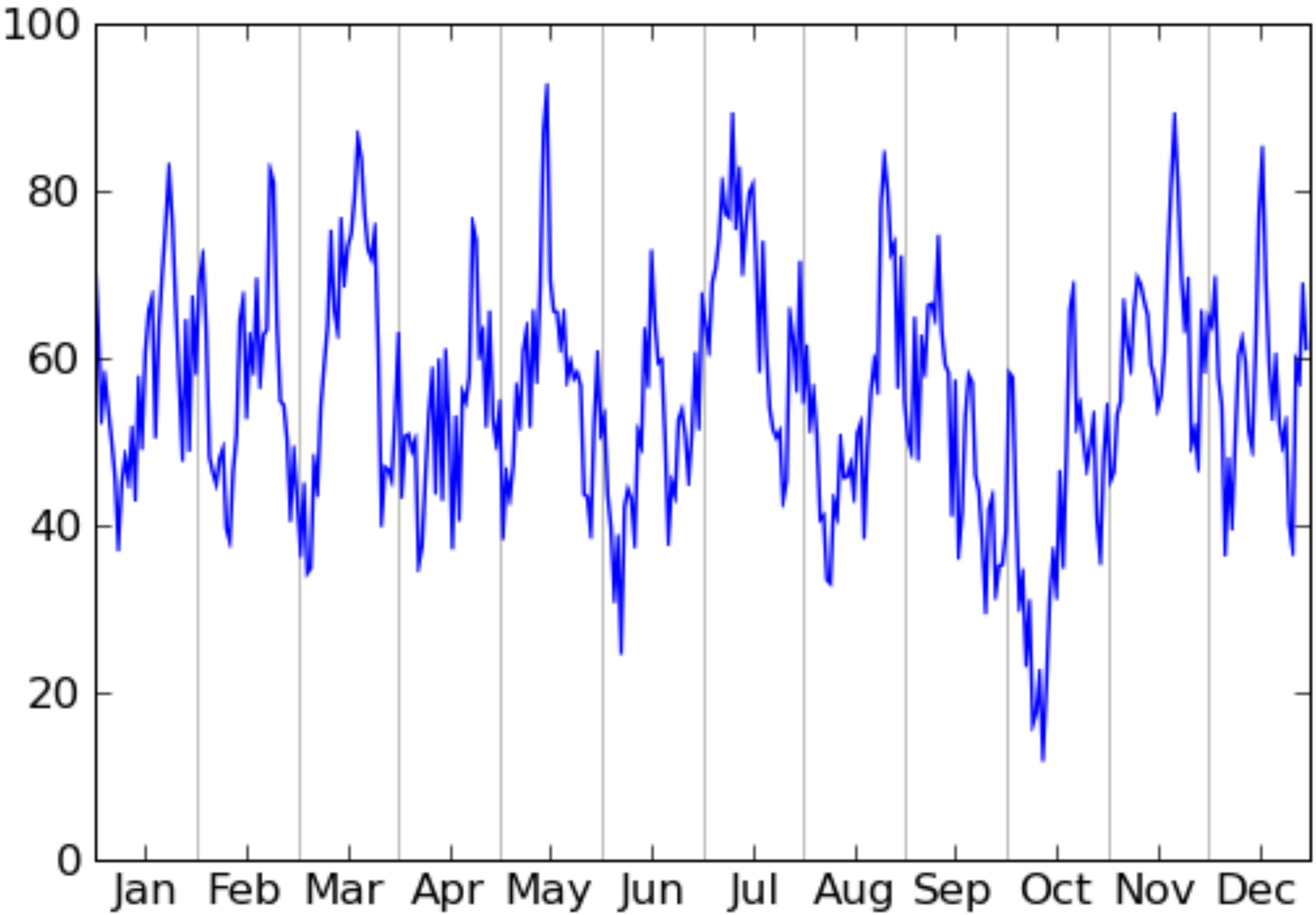
Estimating values

Estimating relative differences

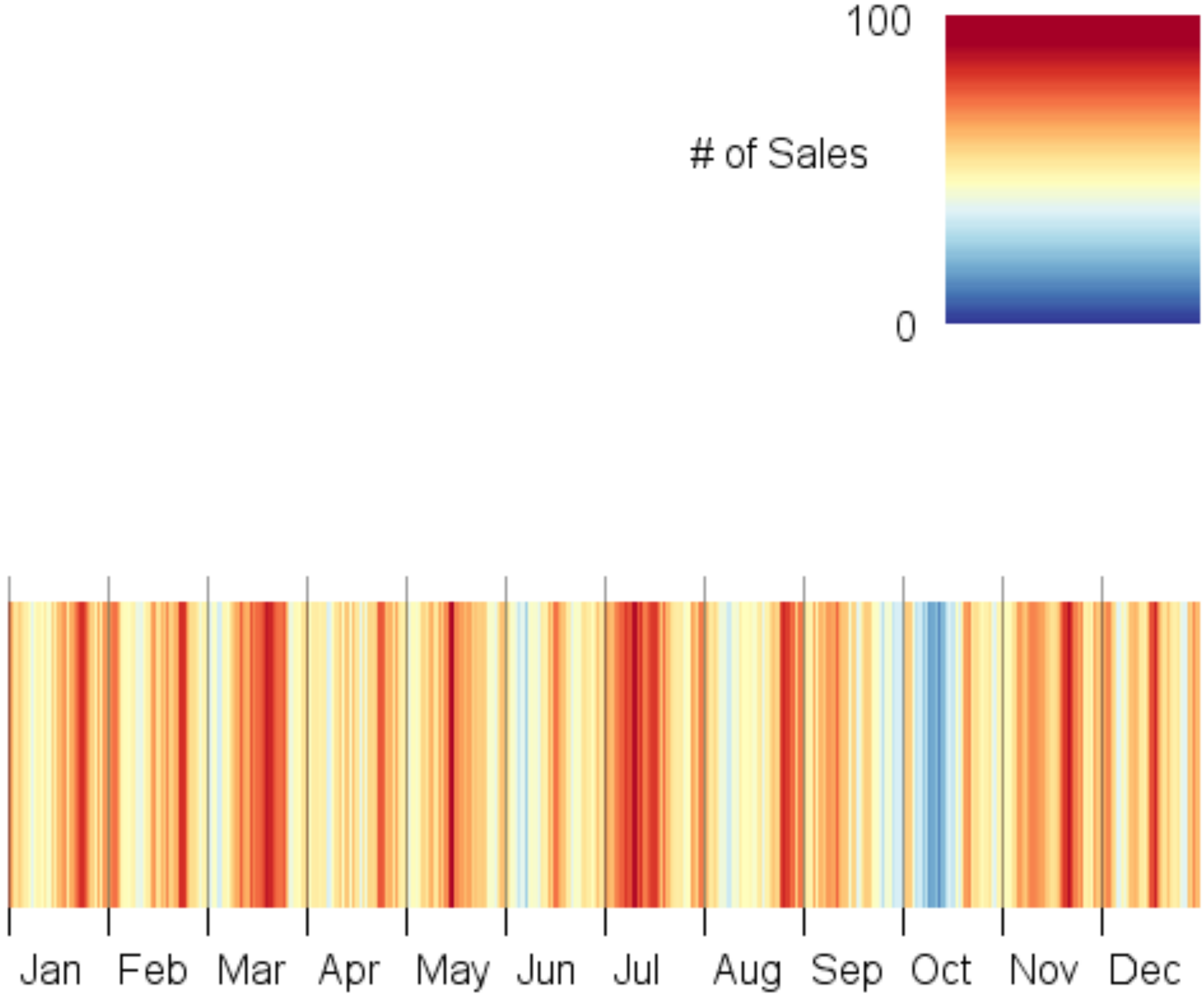
Estimating ensemble properties (!)

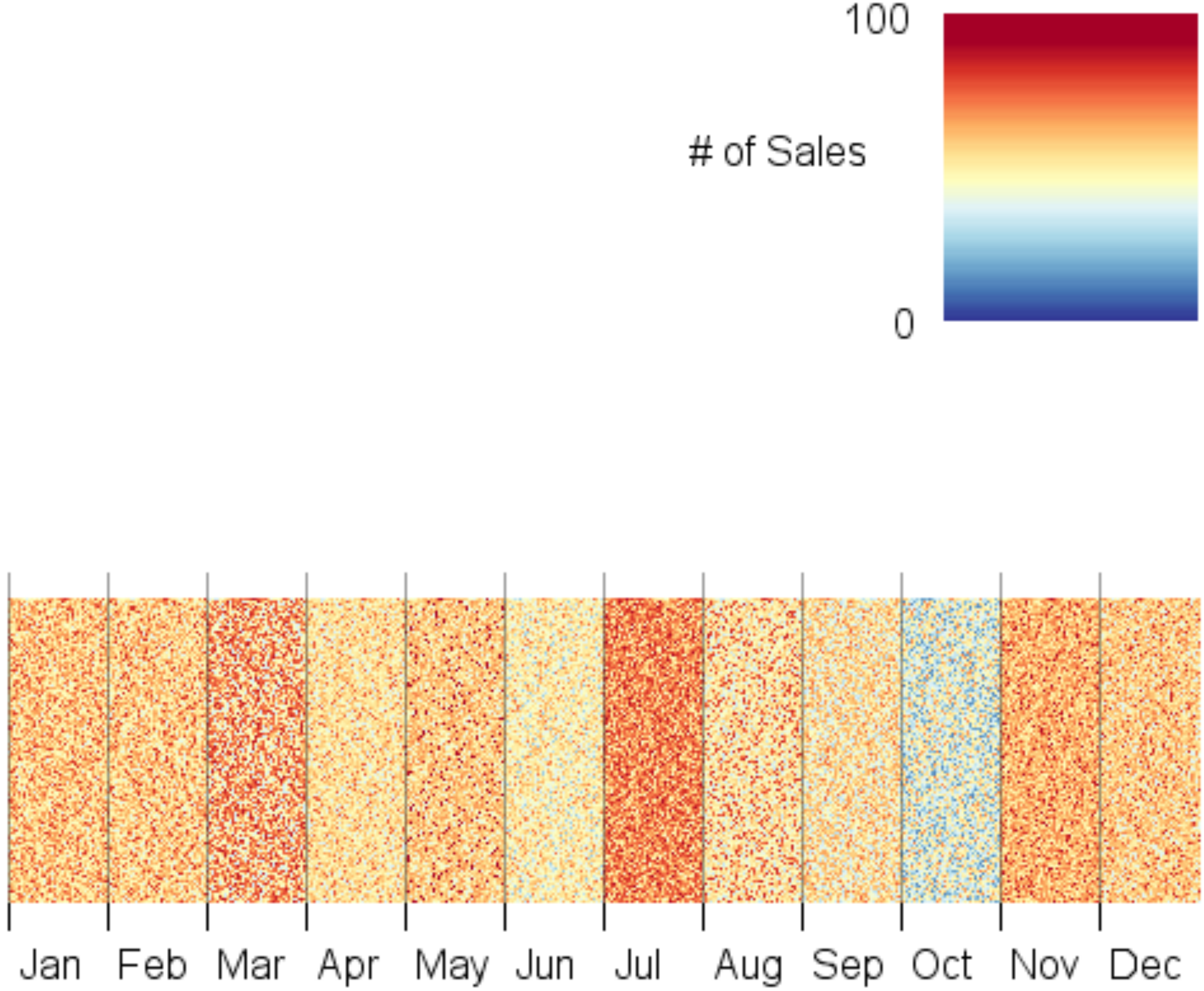
a diversion – to my work

Ensemble Encoding

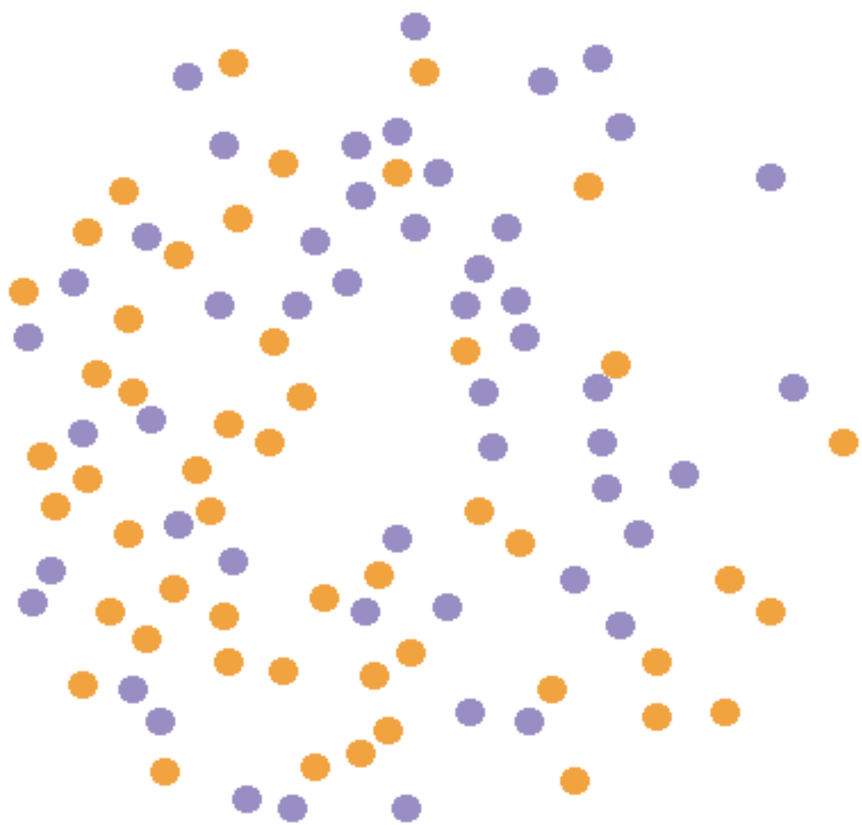


100
80
60
40
20
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec


100
# of Sales
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

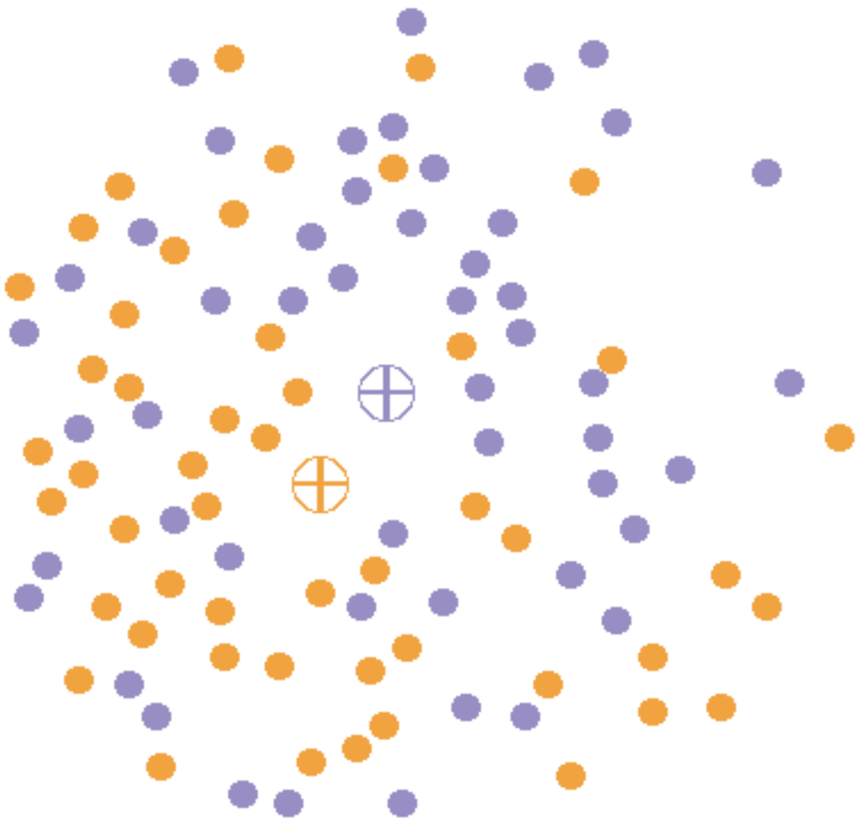

100
# of Sales
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

Which Color Point is Higher on Average?



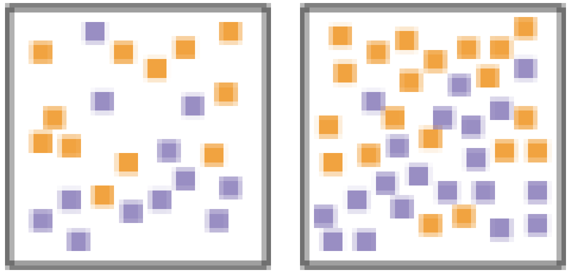
Gleicher, M., Correll, M., Nothelfer, C. and Franconeri, C. “Perception of Average Value in Multiclass Scatterplots.” InfoVis 2013

How did you do that?

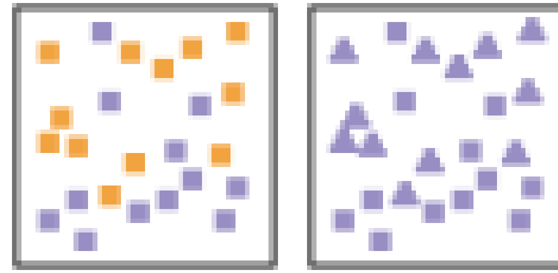


Key Results

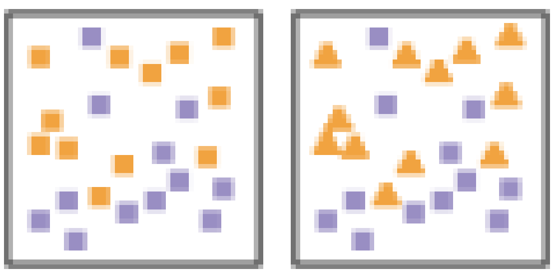
Larger differences gives better performance



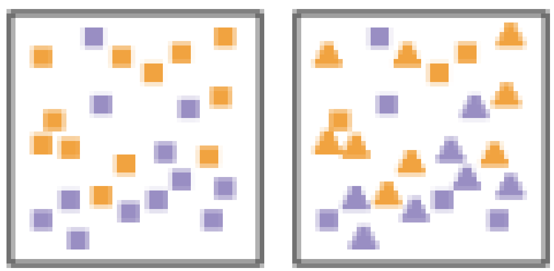
More points do not hurt performance



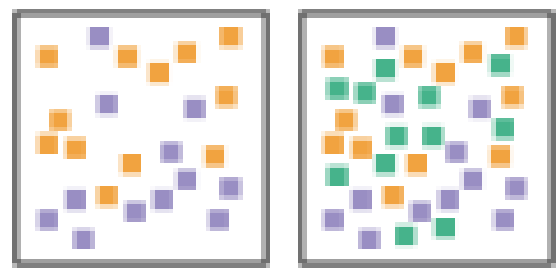
Stronger cues (color) outperform weaker ones



Redundant cues do not help performance



Conflicting cues do not hurt performance



Distractor class does not hurt performance

Gleicher, M., Correll, M., Nothelfer, C. and Franconeri, C. "Perception of Average Value in Multiclass Scatterplots." InfoVis 2013

Factors to consider

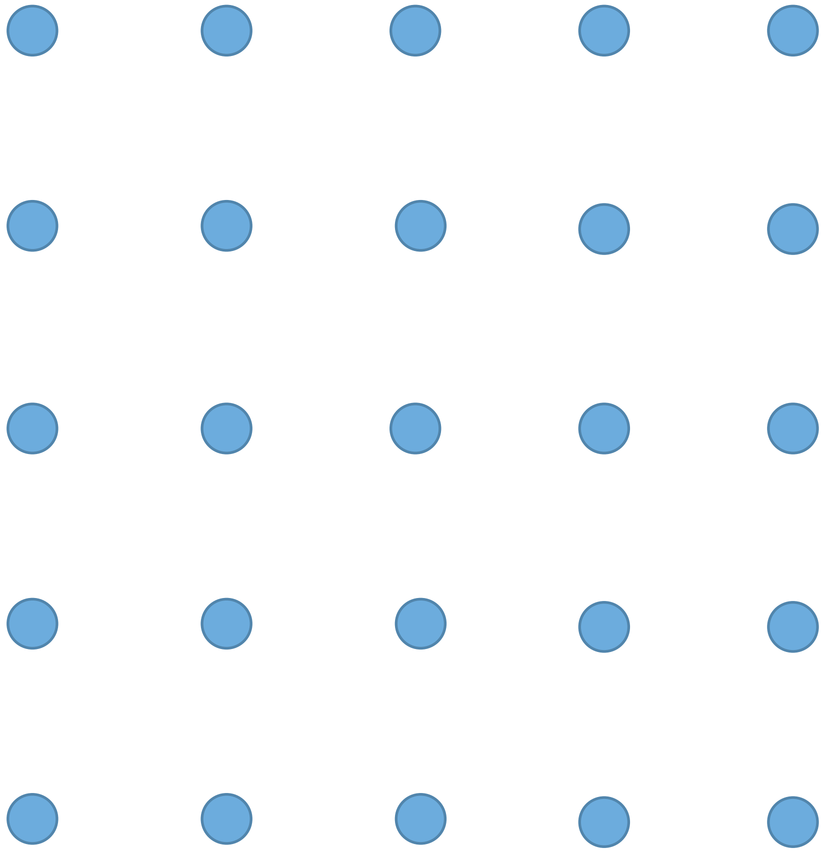
Accuracy (estimate value)

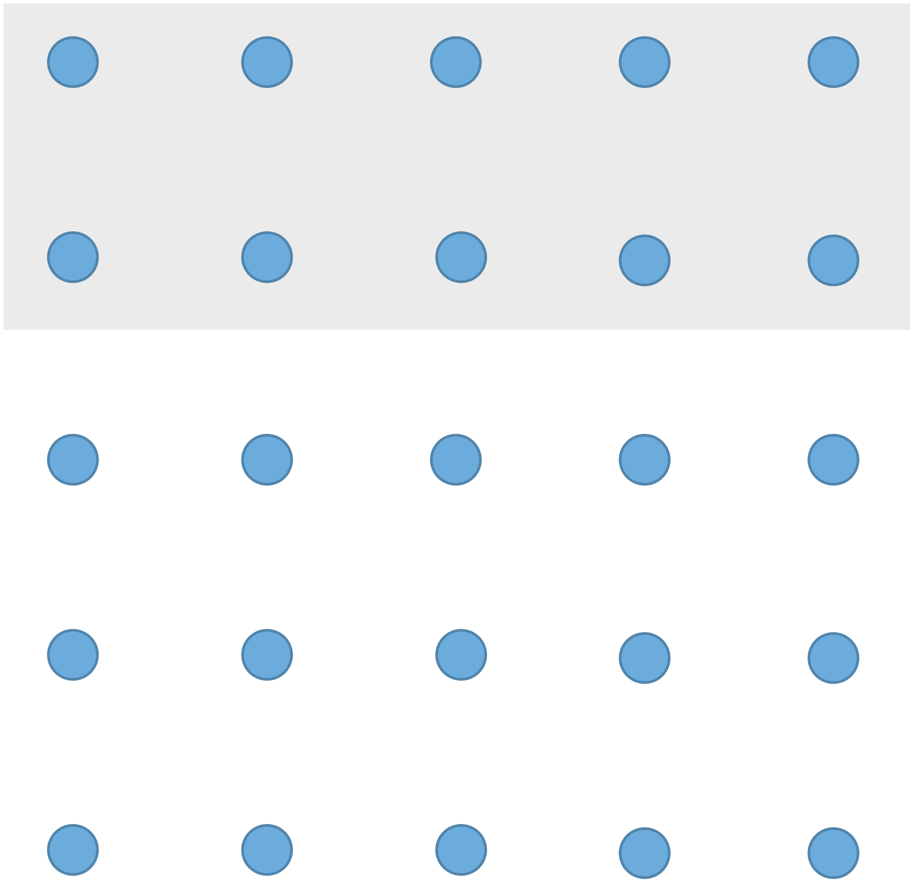
Discriminability (tell apart)

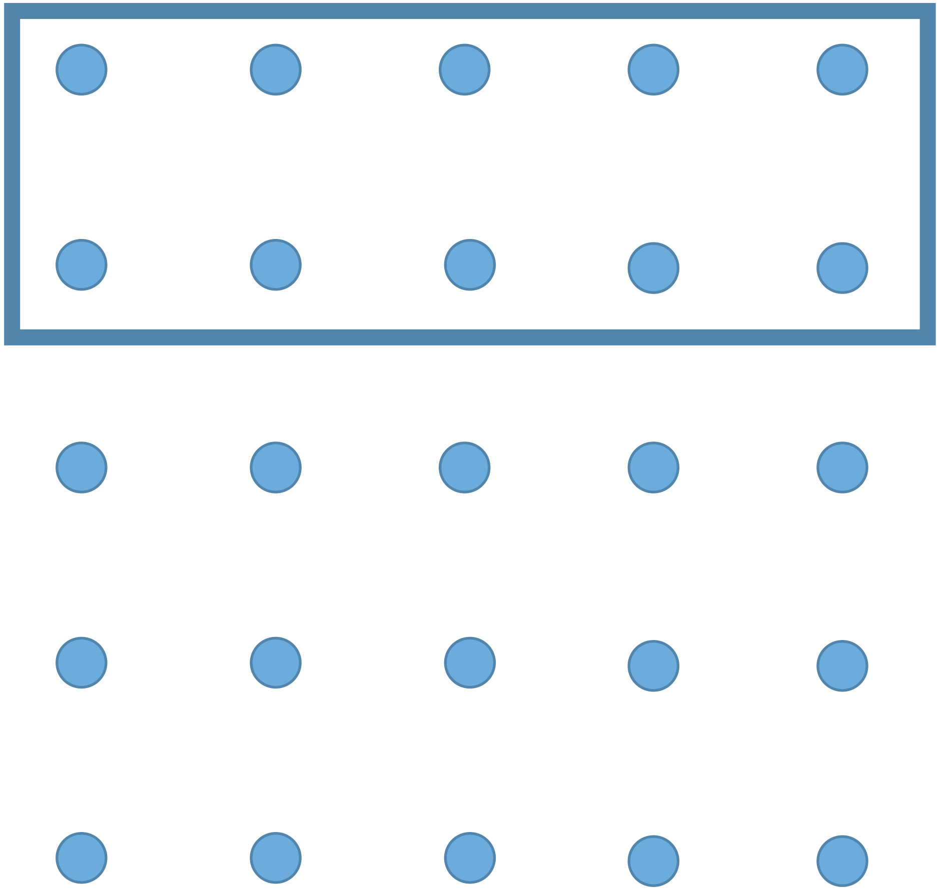
Separability (see different things)

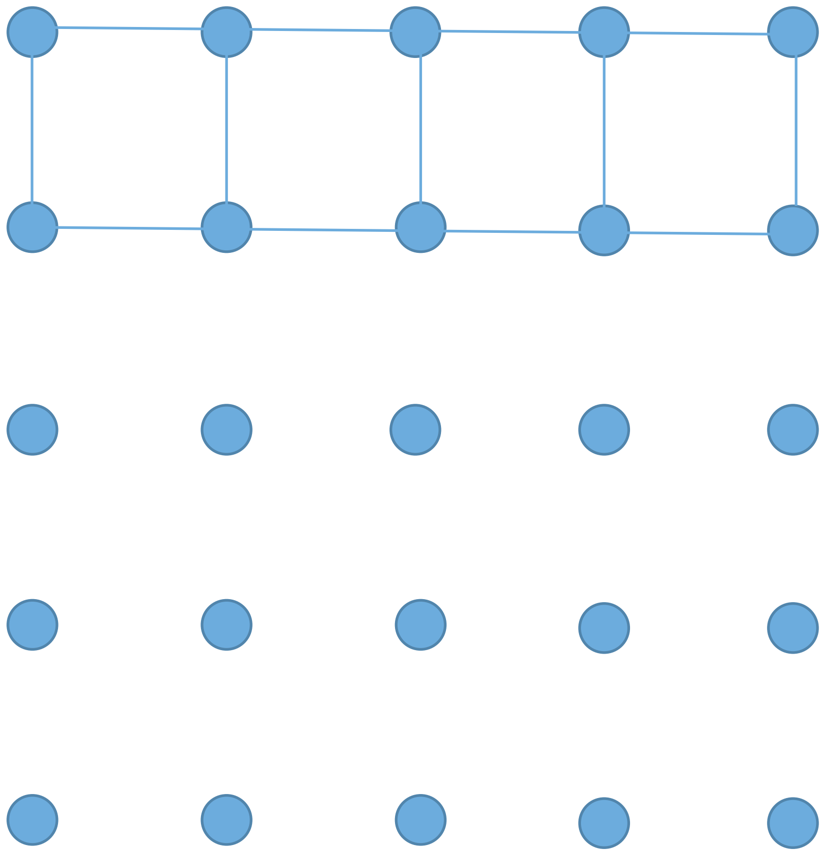
And others…

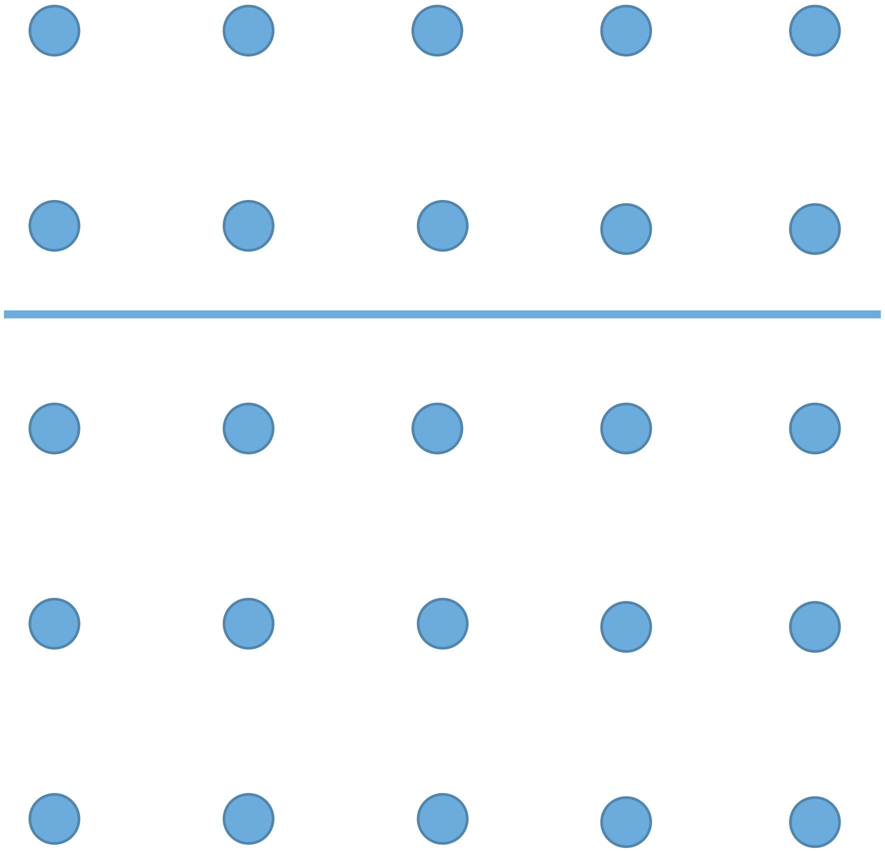
Grouping

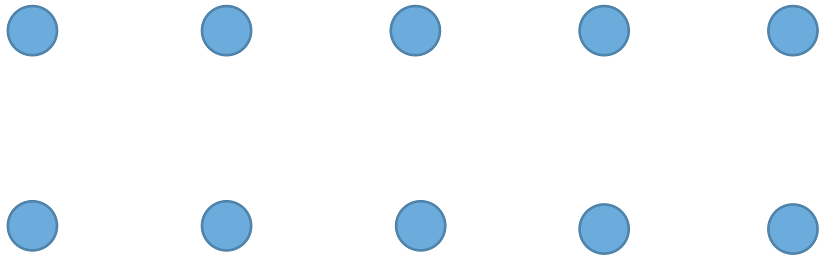


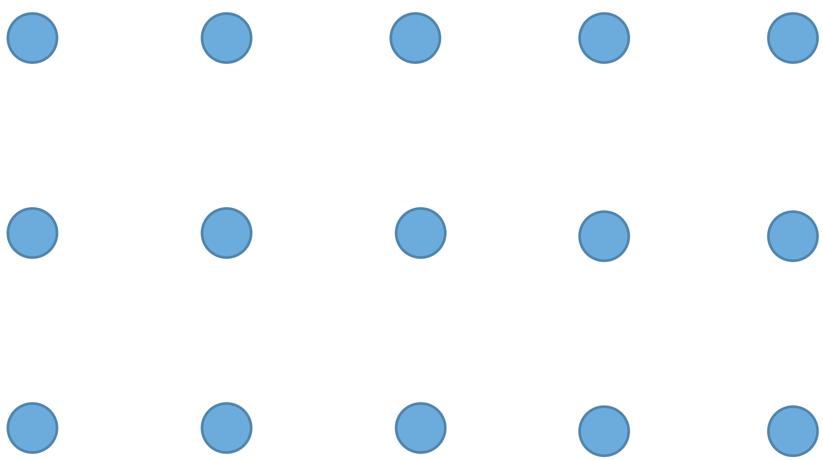


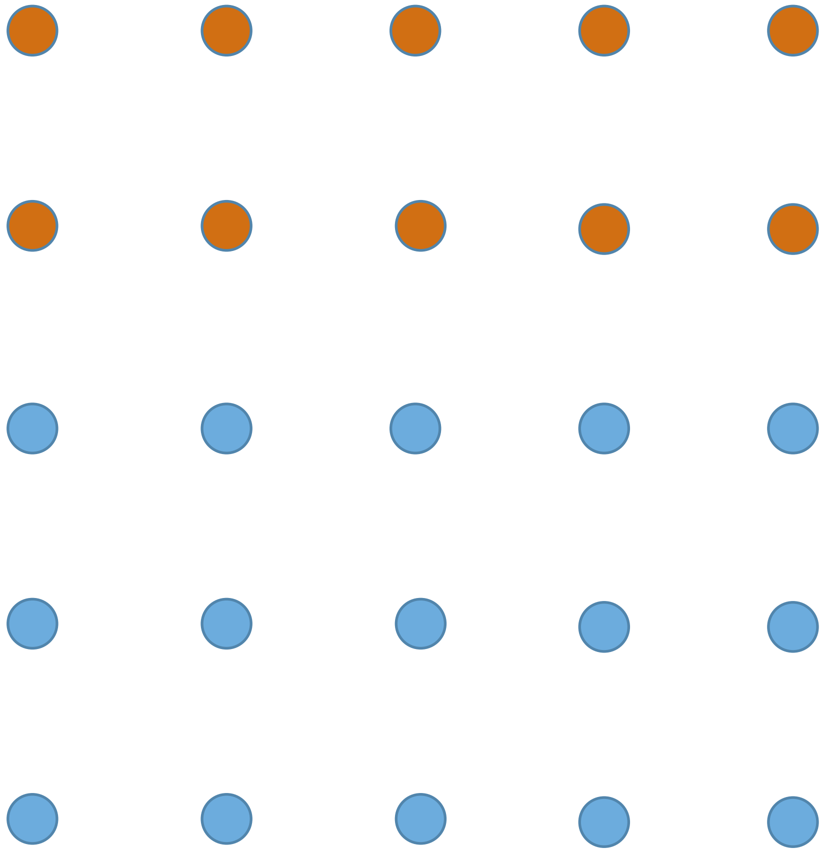


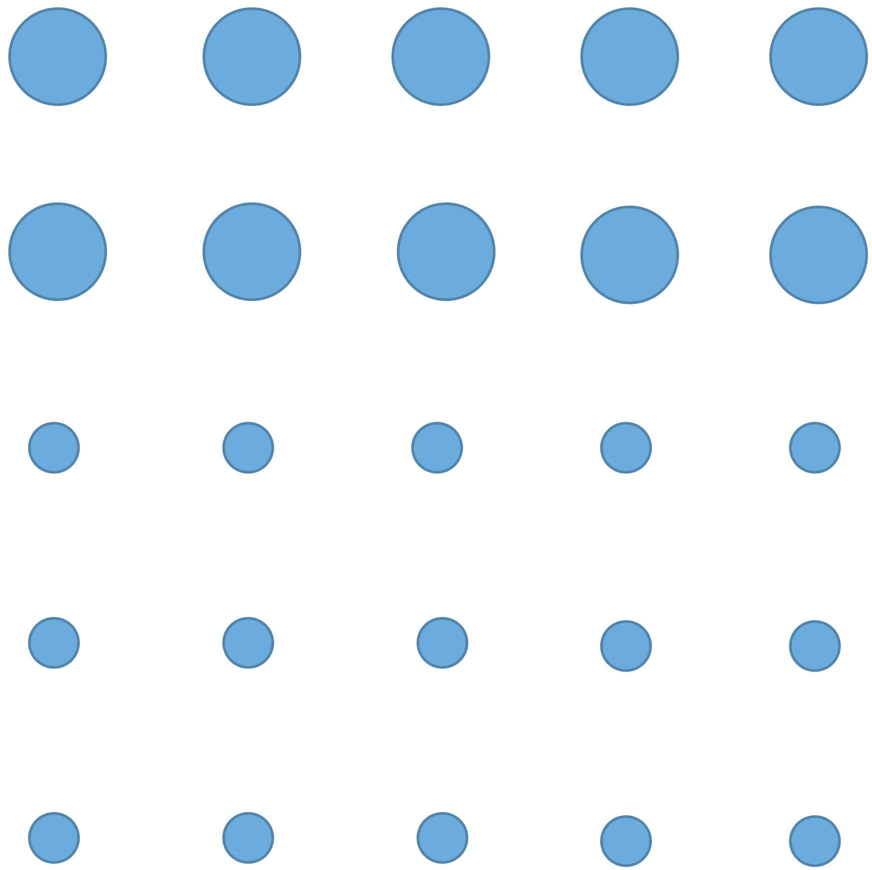












Multi-Variate Glyph
(from the Paris Apartment Problem)

Position = Location

Price (ordinal) -> Size

Type (categorical) -> Hue

Rating (ordinal) -> Lightness/Saturation

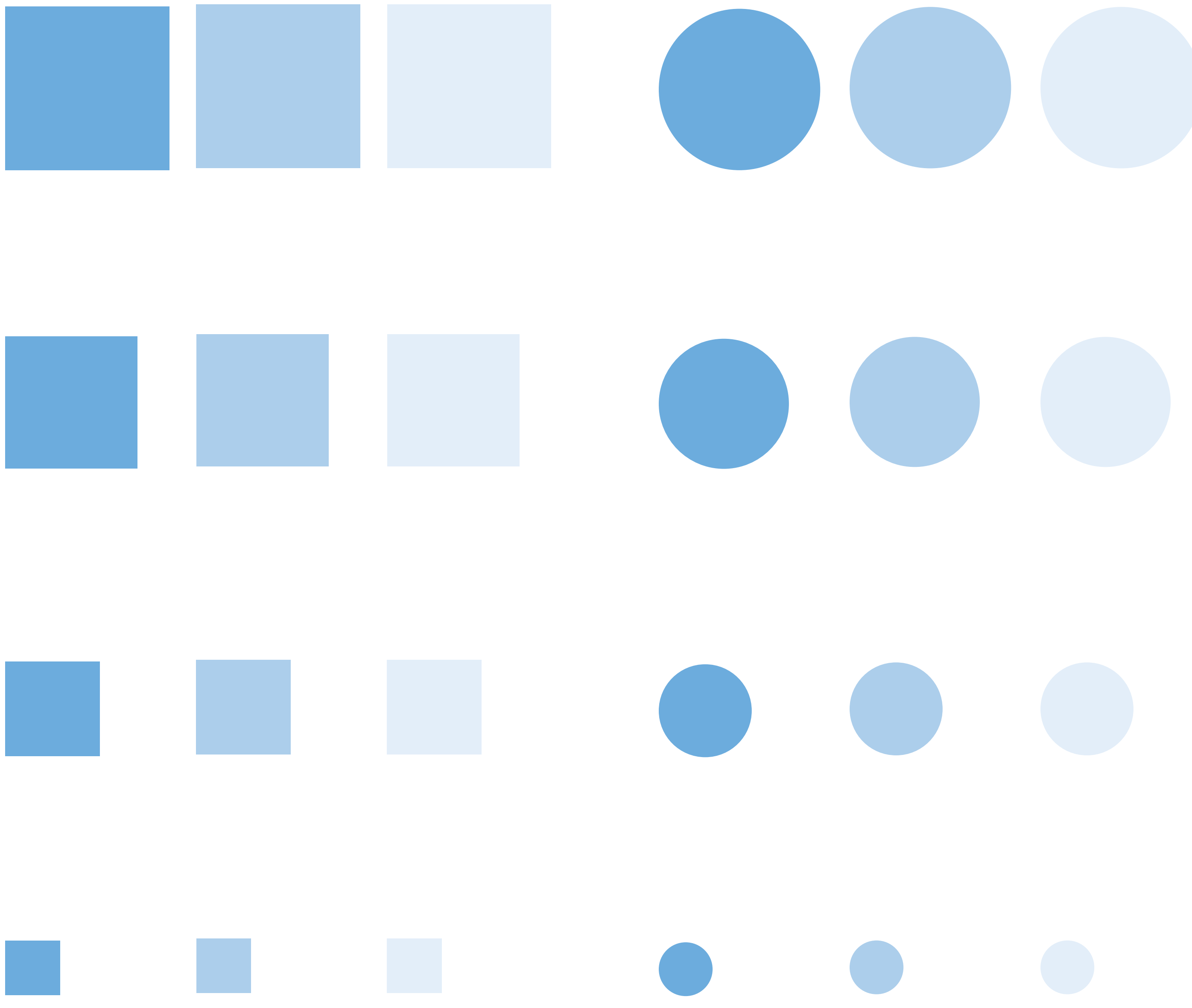
Open Monday (categorical/binary) -> Shape

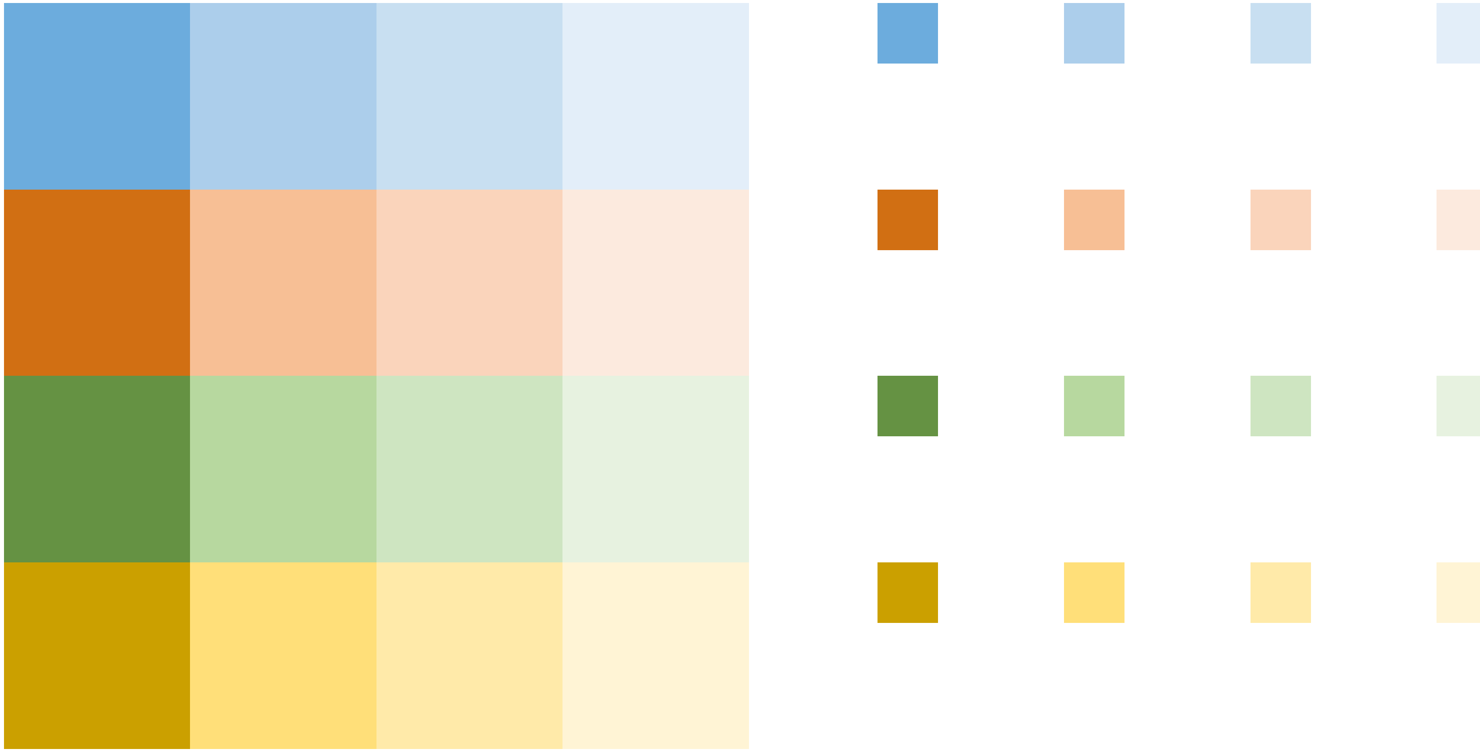
How much can you encode?

Restaurant

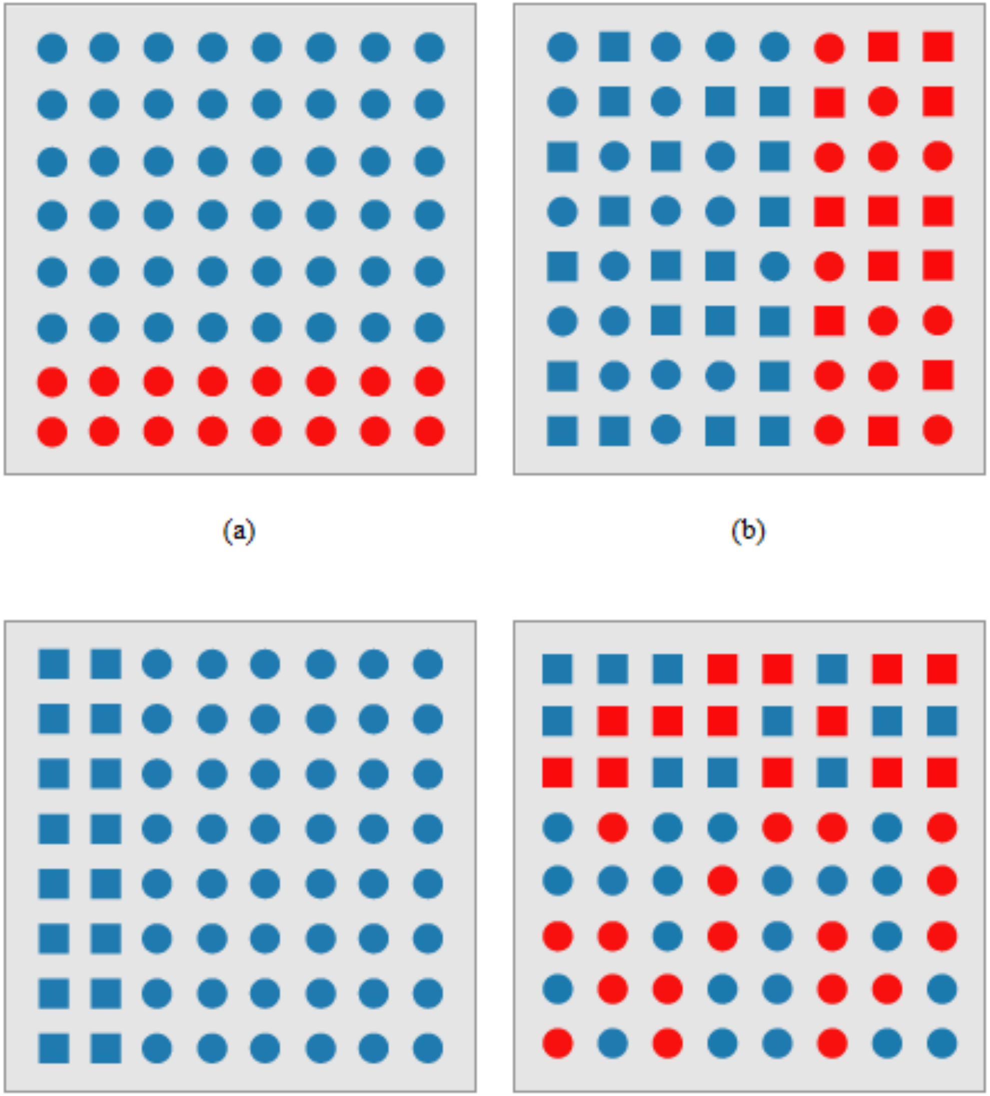
- location
- price ($, $$, $$$, $$$$)
- type (French, Italian, ...)
- rating (★, ★★, ★★★, ★★★★)
- open Mondays

Cue Interactions





Cue Interactions

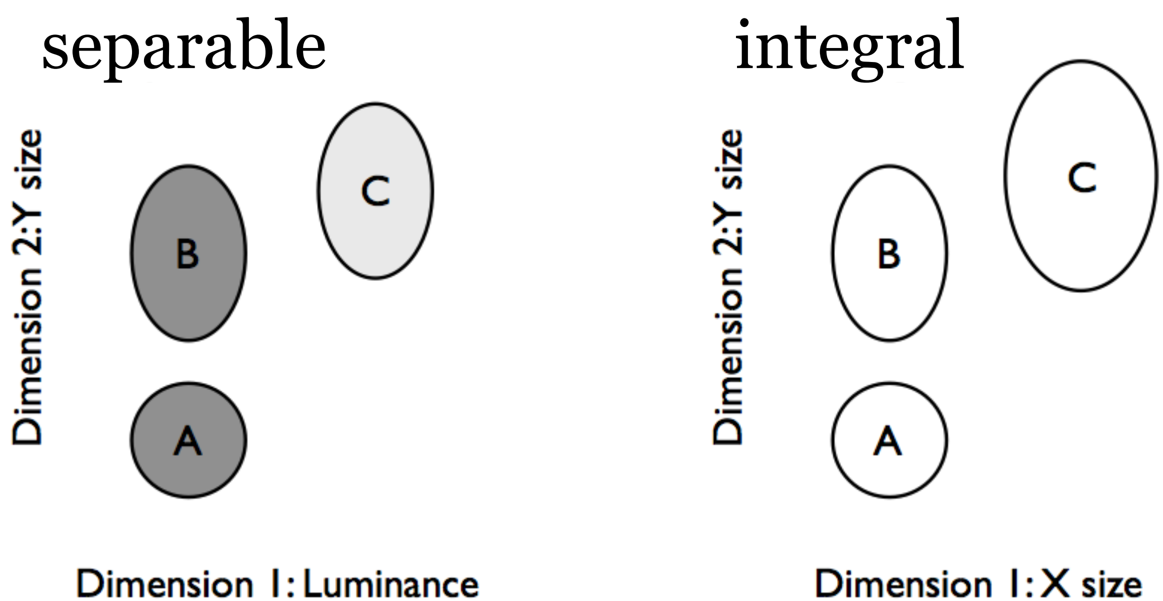


(a) (b) (c) (d)

SEPARABLE vs INTEGRAL

separable: can judge each channel individually

integral: two channels are viewed holistically



Ware 2004

SEPARABLE vs INTEGRAL

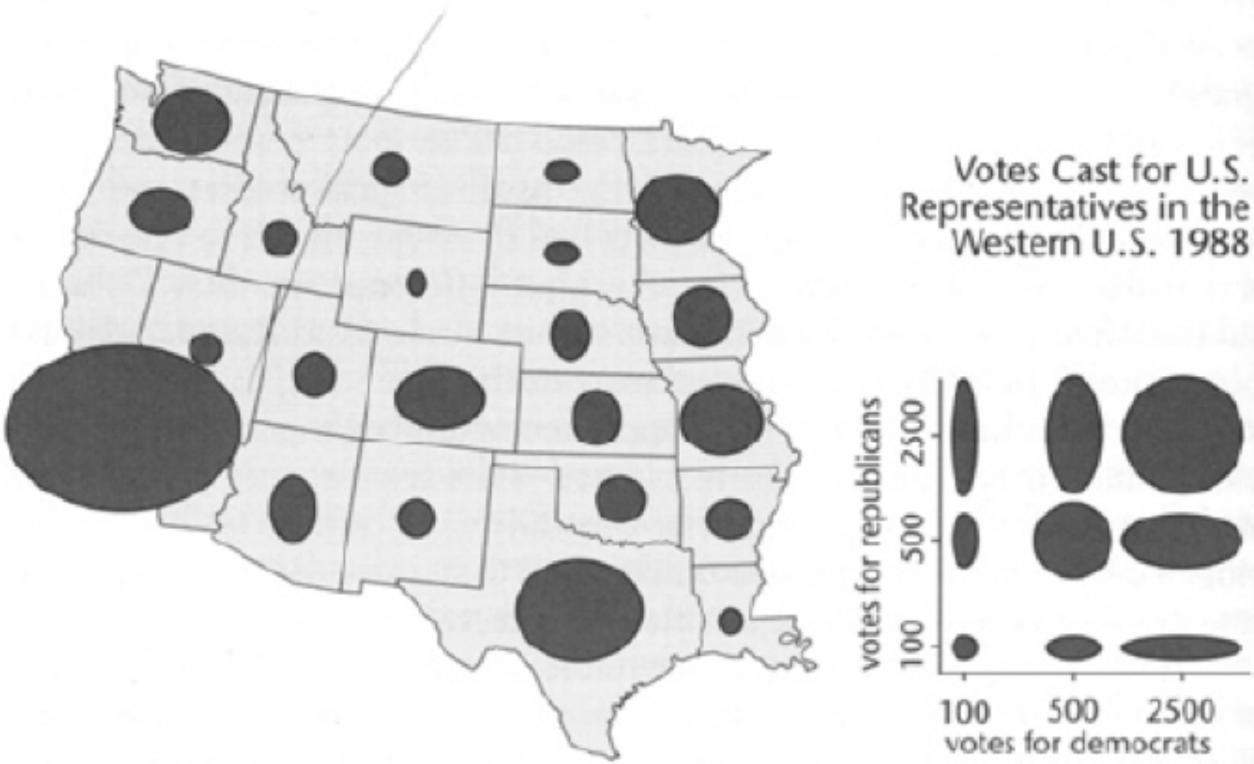
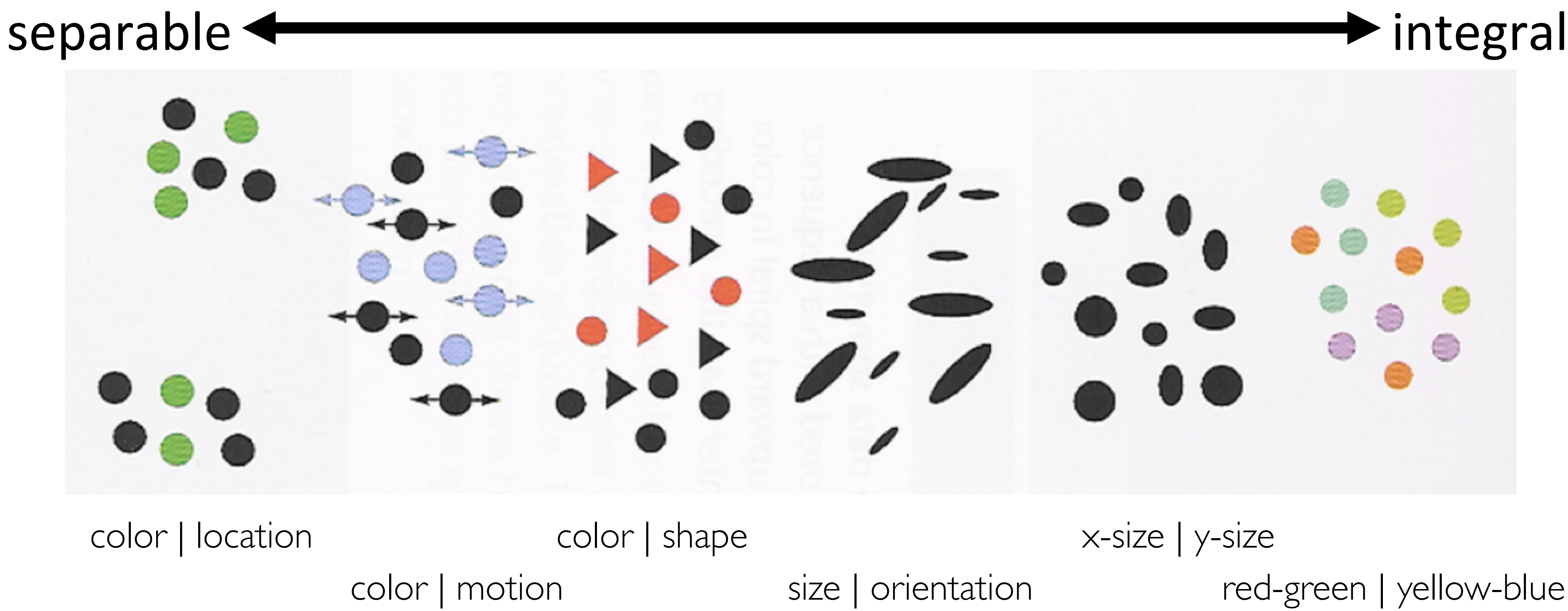


FIGURE 3.38. An example of the use of an ellipse as a map symbol in which the horizontal and vertical axes represent different (but presumably related) variables.

MacEachren 1995

SEPARABLE vs INTEGRAL



Ware 2004

Factors to consider

Accuracy
Discriminability
Separability

Popout, Search, Average, …

Salience, Semantics, …

Taxonomy-Based Glyph Design—with a Case Study on Visualizing Workflows of Biological Experiments

Eamonn Maguire, Philippe Rocca-Serra, Susanna-Assunta Sansone, Jim Davies, and Min Chen

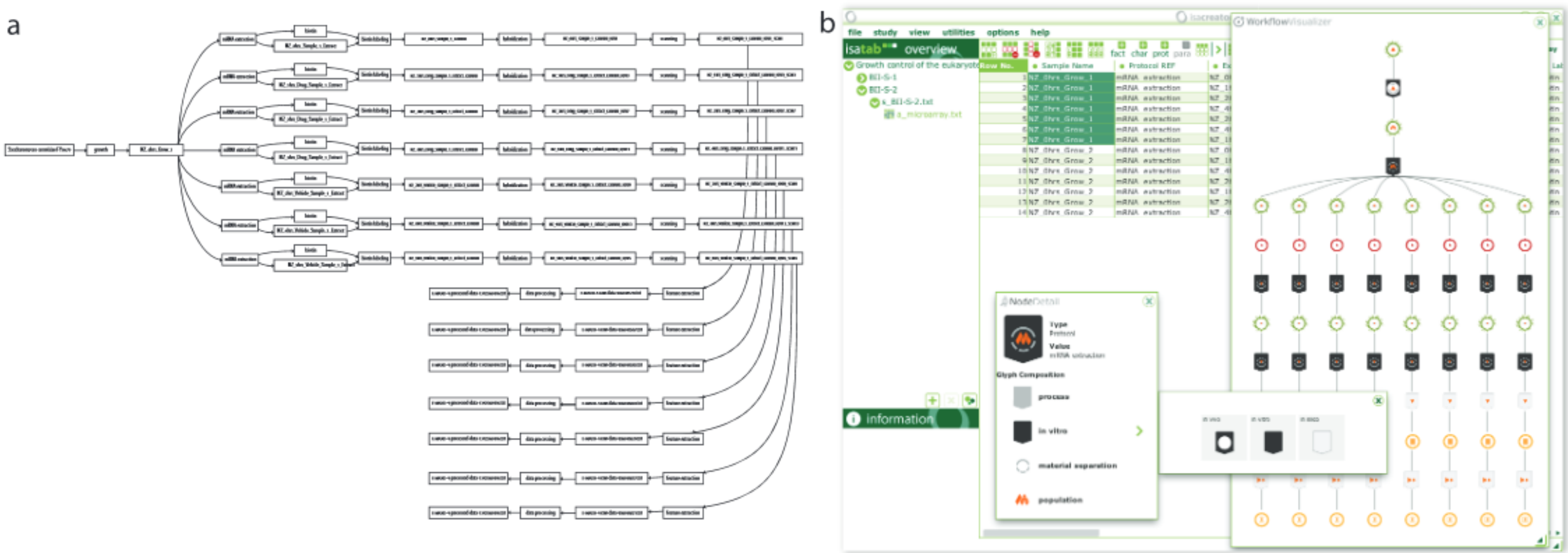
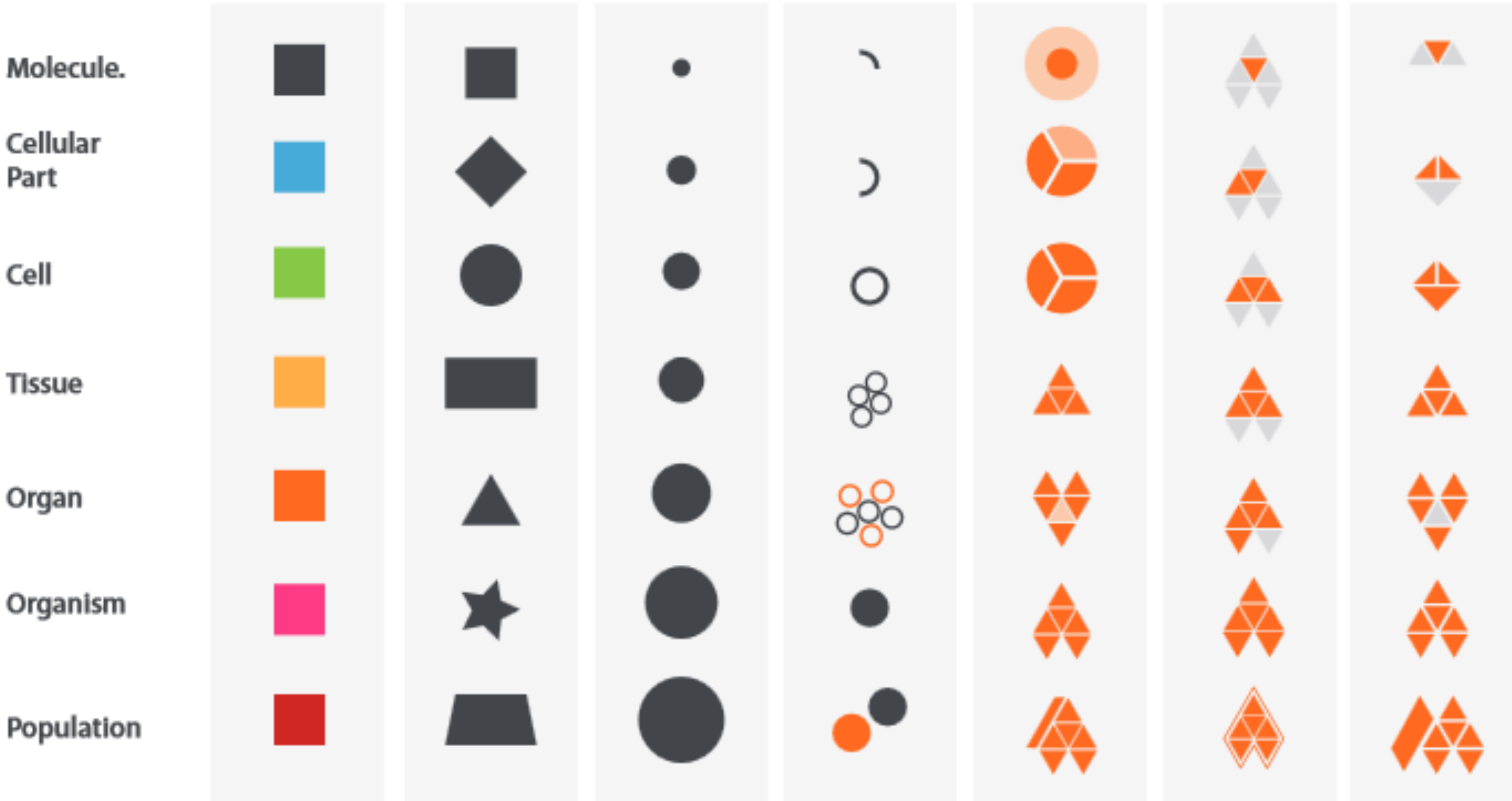


Fig. 1. a) Workflow as rendered currently using toolkits such as GraphViz. b) We propose to replace the textual labels with glyphs, while allowing interactive access to detailed descriptions. This makes it easy to gain an overview, search components and compare workflows. The screenshot shows a prototype developed within ISAcreator, a system for capturing biological experiment metadata.

Abstract—Glyph-based visualization can offer elegant and concise presentation of multivariate information while enhancing speed and ease in visual search experienced by users. As with icon designs, glyphs are usually created based on the designers' experience and intuition, often in a spontaneous manner. Such a process does not scale well with the requirements of applications where a large number of concepts are to be encoded using glyphs. To alleviate such limitations, we propose a new systematic process for glyph design by exploring the parallel between the hierarchy of concept categorization and the ordering of discriminative capacity of visual channels. We examine the feasibility of this approach in an application where there is a pressing need for an efficient and effective means to visualize workflows of biological experiments. By processing thousands of workflow records in a public archive of biological

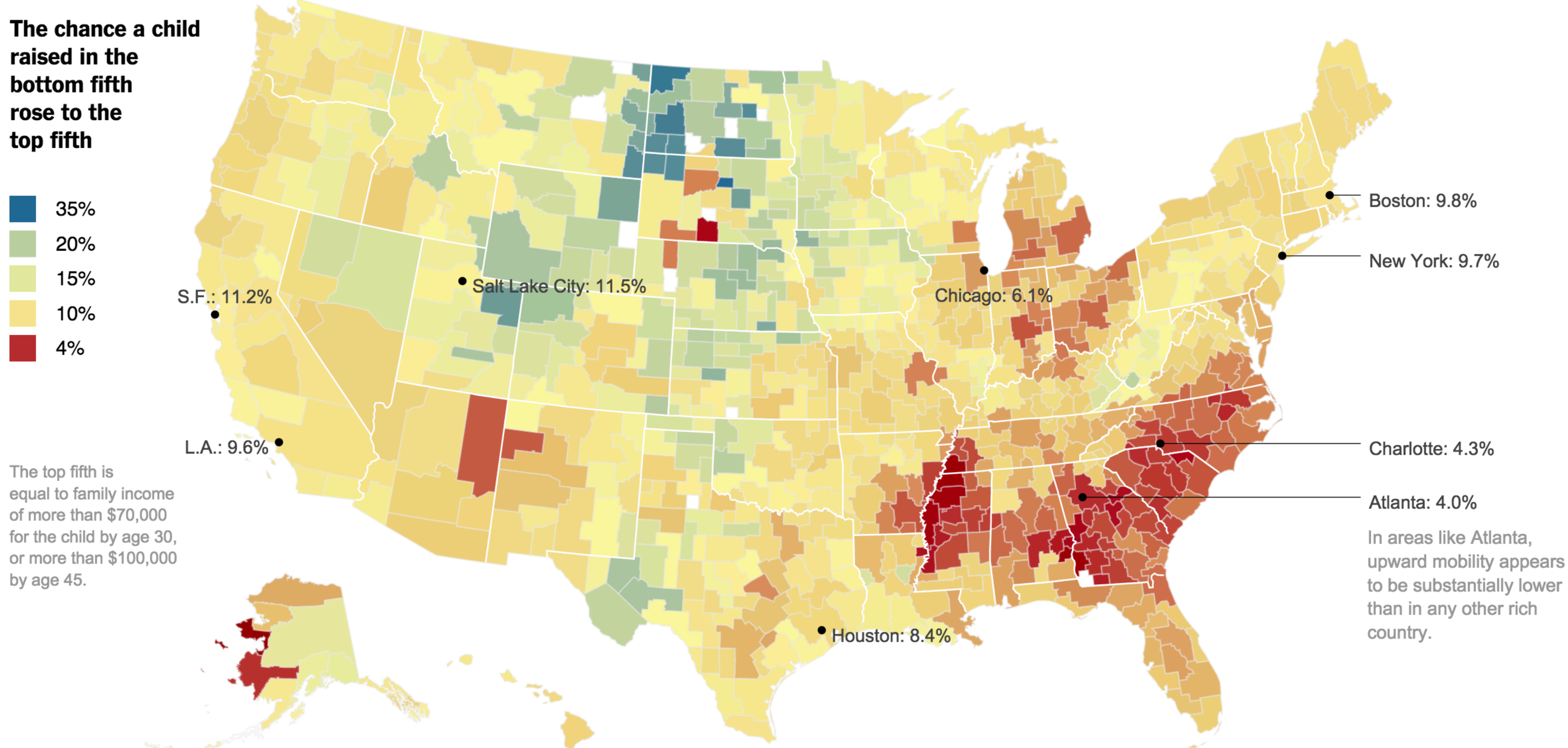

design option 1
design option 2
design option 3
design option 4
design option 5
design option 6
design option 7
S0
Inputs and Outputs
Process
S7
Biological
Device
Chemical
Data
S6
In Vitro
In Vivo
In Silico
S3
Data Collection
Data Processing
Data Analysis
S2
Material perturbation
Material separation
Material amplification
Material combination
Material collection
S5
Molecule.
Cellular Part
Cell
Tissue
Organ
Organism
Population
S4
Material induced pertubation.
Behaviourally induced pertubation.
Physically induced pertubation.


Molecule.
Cellular Part
Cell
Tissue
Organ
Organism
Population

Some Examples

Name that encoding!

Critique that encoding!

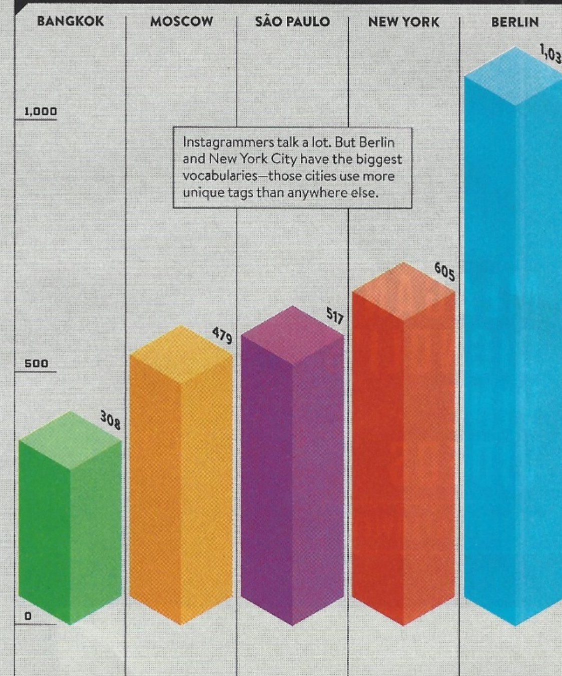


The chance a child raised in the bottom fifth rose to the top fifth
35%
20%
15%
10%
4%
S.F.: 11.2%
Salt Lake City: 11.5%
Chicago: 6.1%
Boston: 9.8%
New York: 9.7%
L.A.: 9.6%
Charlotte: 4.3%
Atlanta: 4.0%
Houston: 8.4%
The top fifth is equal to family income of more than $70,000 for the child by age 30, or more than $100,000 by age 45.
In areas like Atlanta, upward mobility appears to be substantially lower than in any other rich country.

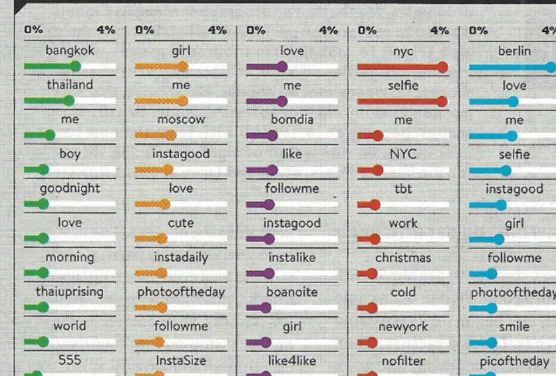
The Quantified Selfie

You've probably often wondered: Who tilts their heads more when taking selfies, the women of Bangkok or São Paulo? To find out, researcher Lev Manovich and his colleagues used Mechanical Turk to crowdsource the analysis of 3,200 selfies posted on Instagram during one week in December 2013. (Result: São Paulans are the head-tiltingest selfie takers.) Now the team has gone back to the data, exploring the tags we attach to our smartphone photos. Here's what's #trending. —RACHEL SOMERSTEIN

WHO USES THE WIDEST VARIETY OF SELFIE TAGS (BY NUMBER OF UNIQUE TAGS)



TOP 10 TAGS IN EACH CITY (BY PERCENTAGE OF SELFIES THAT USE THEM)



Big surprise, "love" makes the top 10 in every city but NYC, the only one where "work" makes the top 10. And self-snappers seem to feel the need to say where they are; less so for Brazilians, though, who really like their "likes."