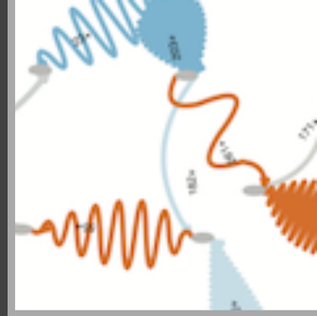
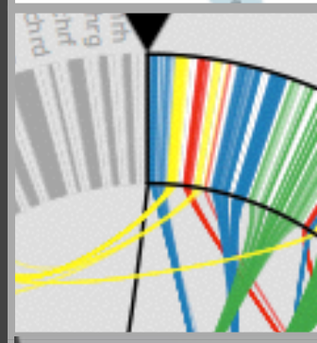
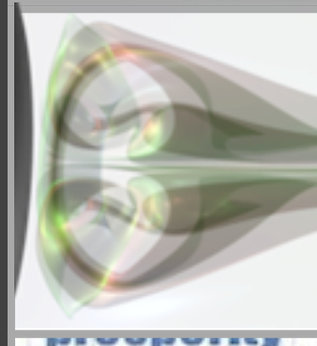
cs6630 | September 4 2014

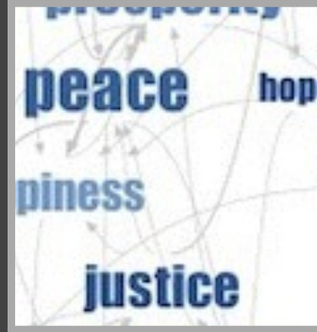
DATA ABSTRACTION & INTRO TO TABLEAU

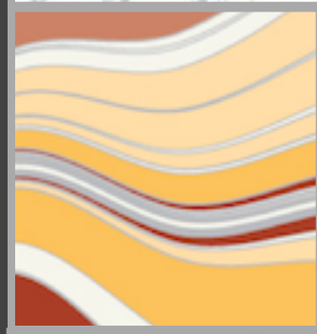
Miriah Meyer
University of Utah











administrivia . . .

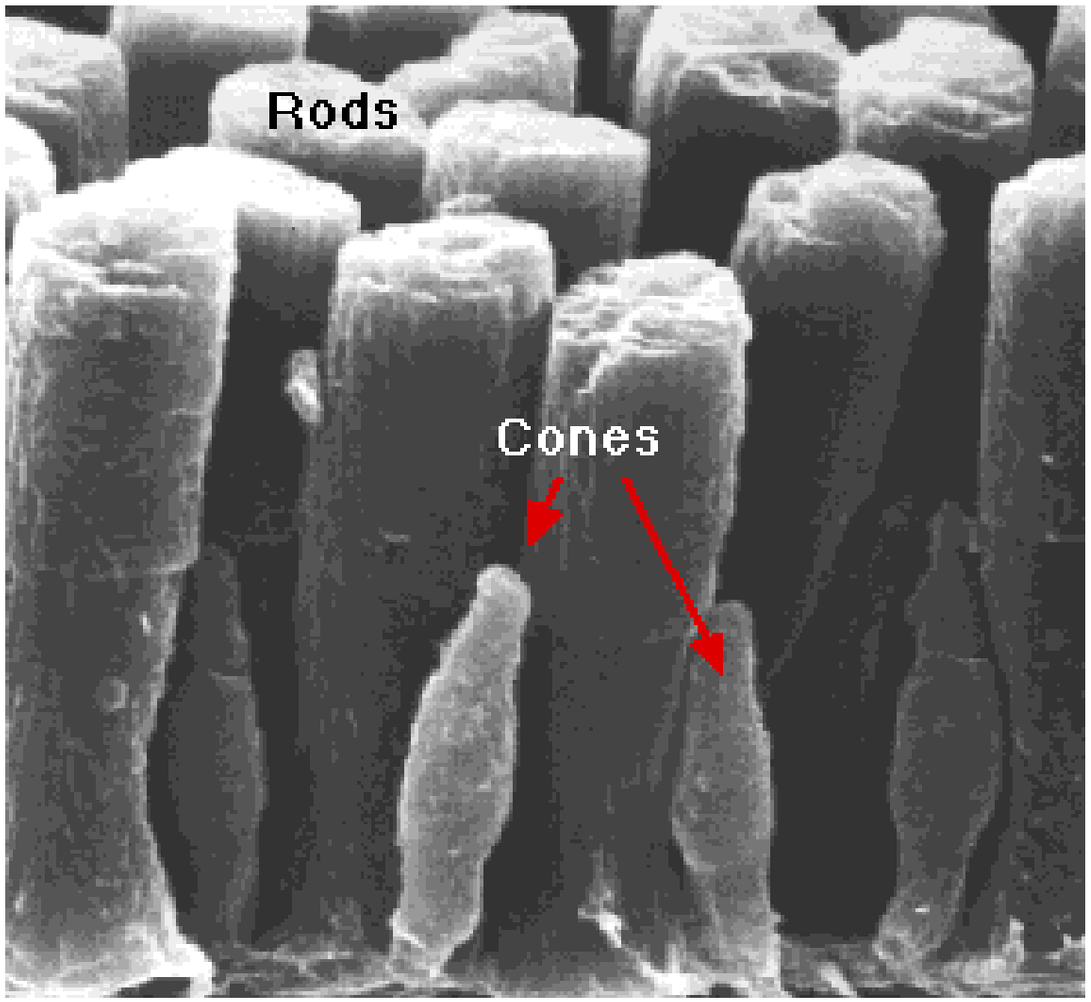
- design critiques due tonight

- first assignment out today

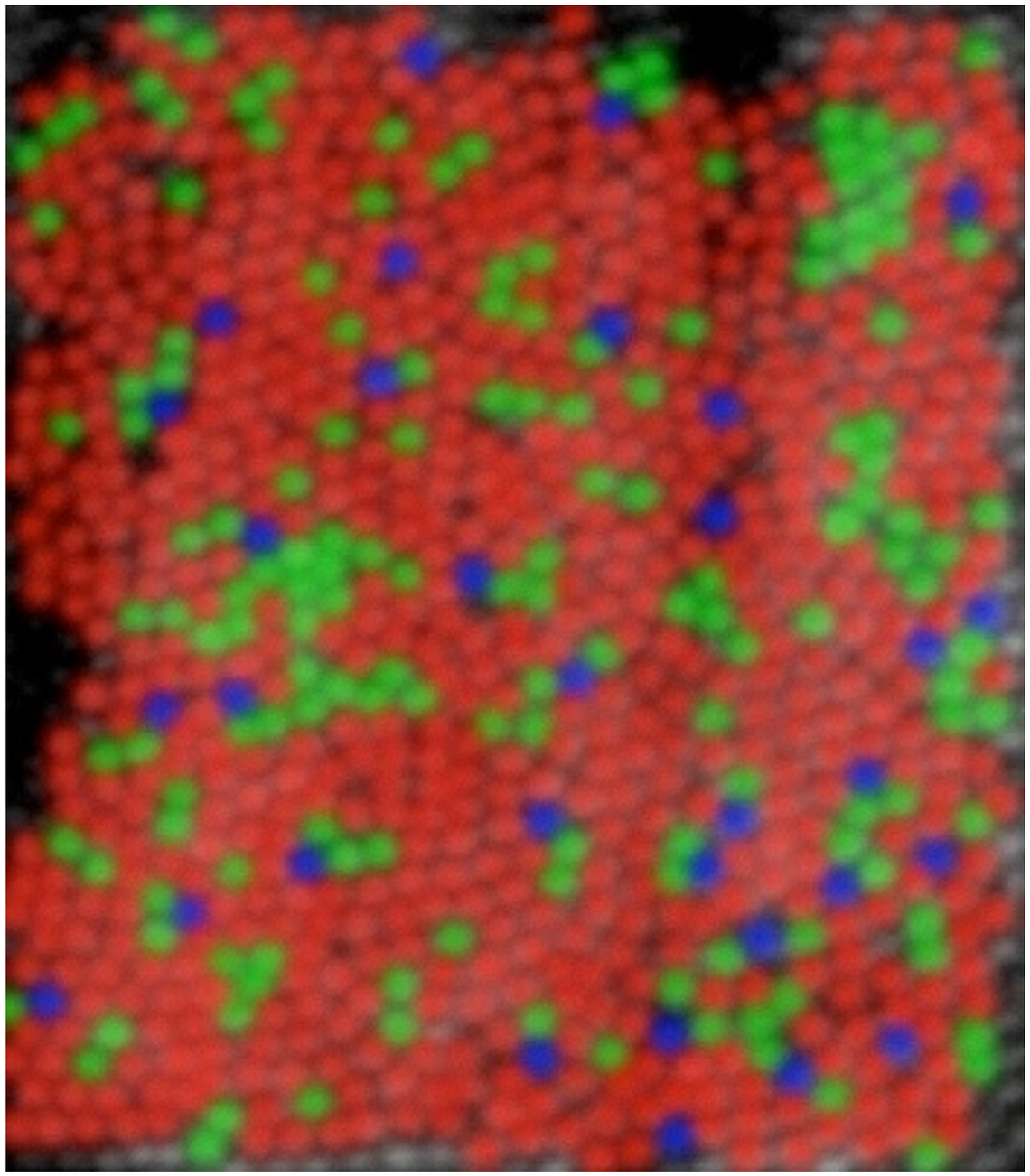
- **there *might* be 3 seats available…**

  - *I will be teaching again next fall!*

last time . . .



120 million rods

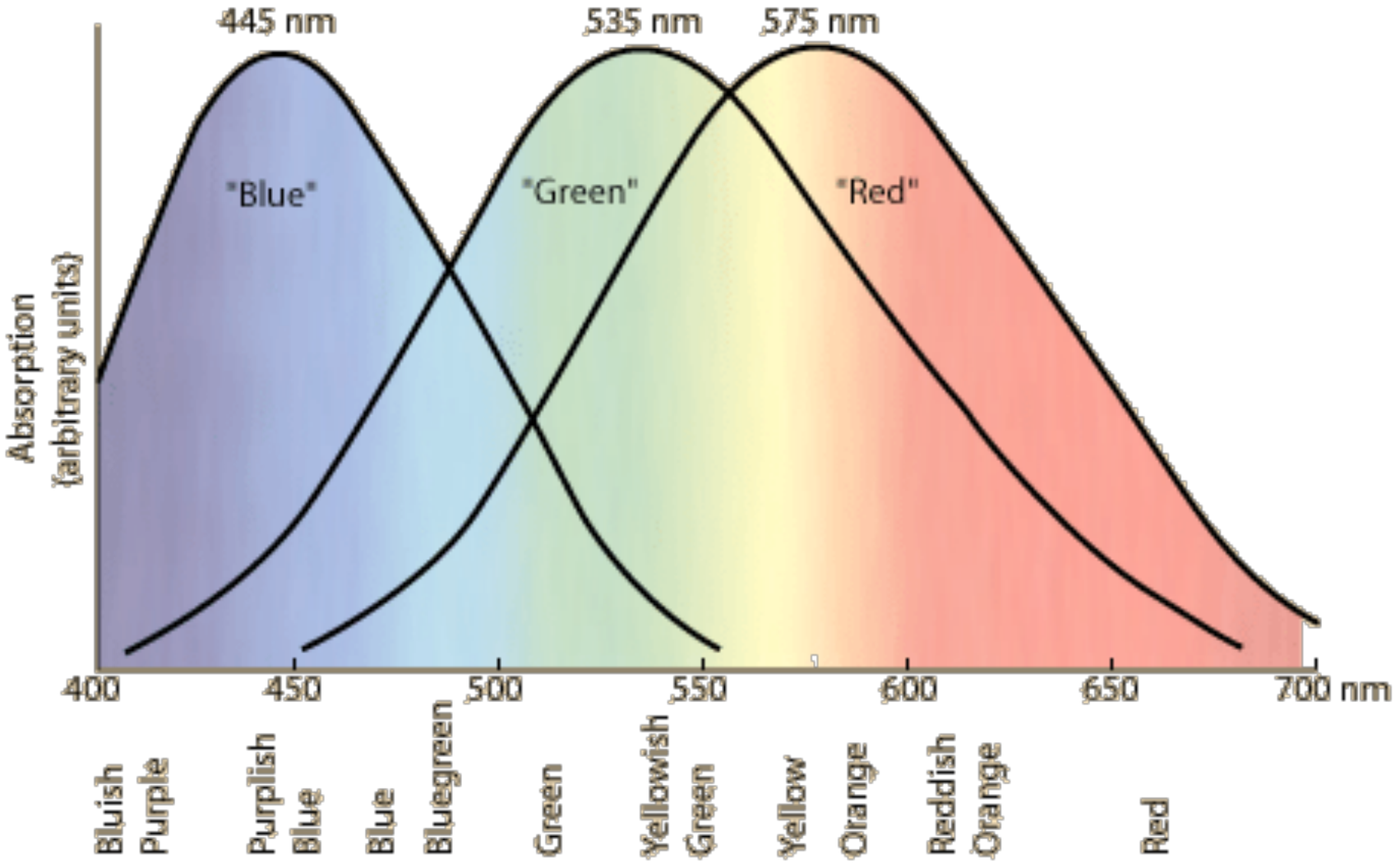


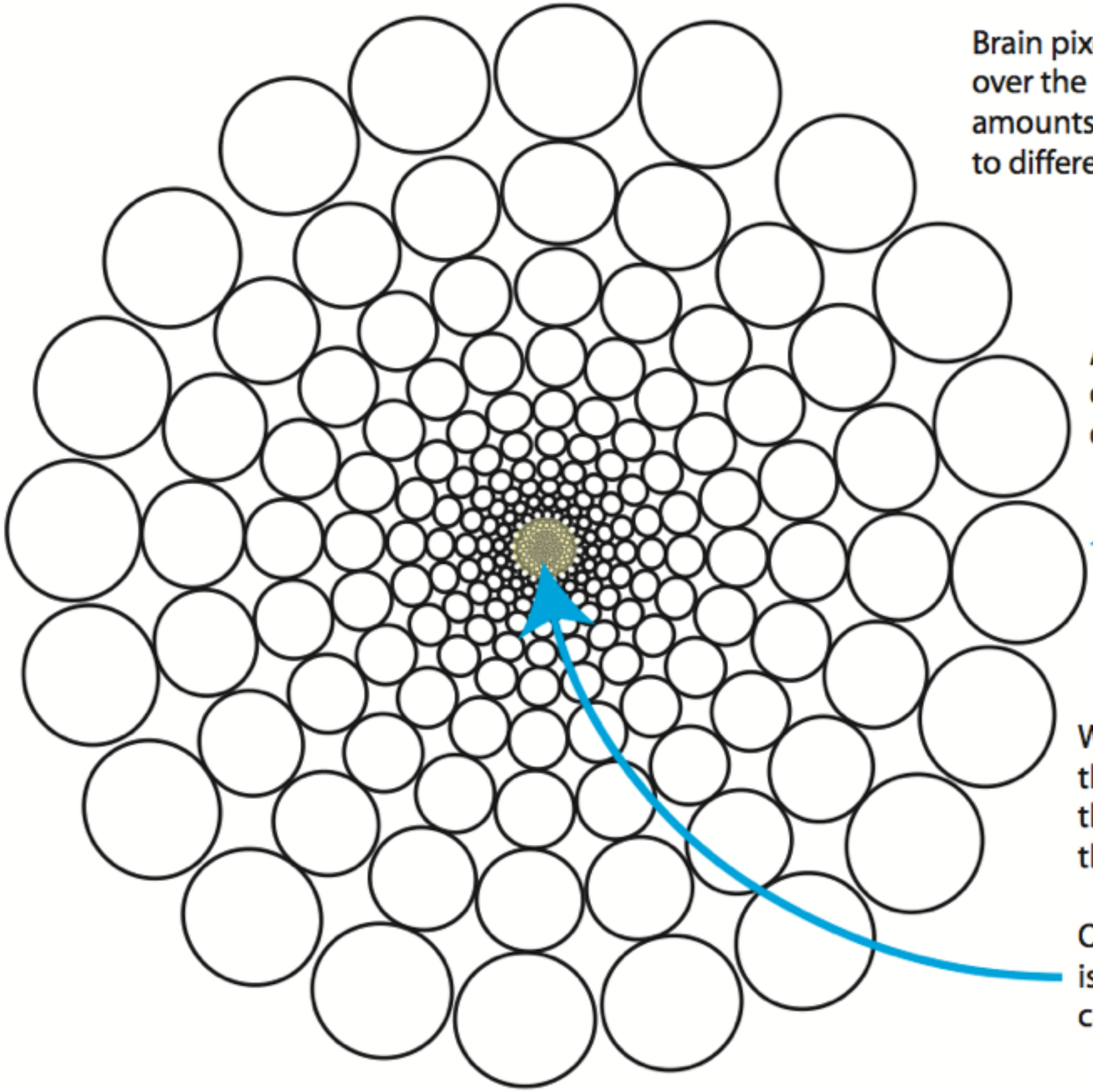
5-6 million cones

Wandell, "Foundations of Vision" (left)

David R. Williams, Univ. of Rochester (right)

Cone Response





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.

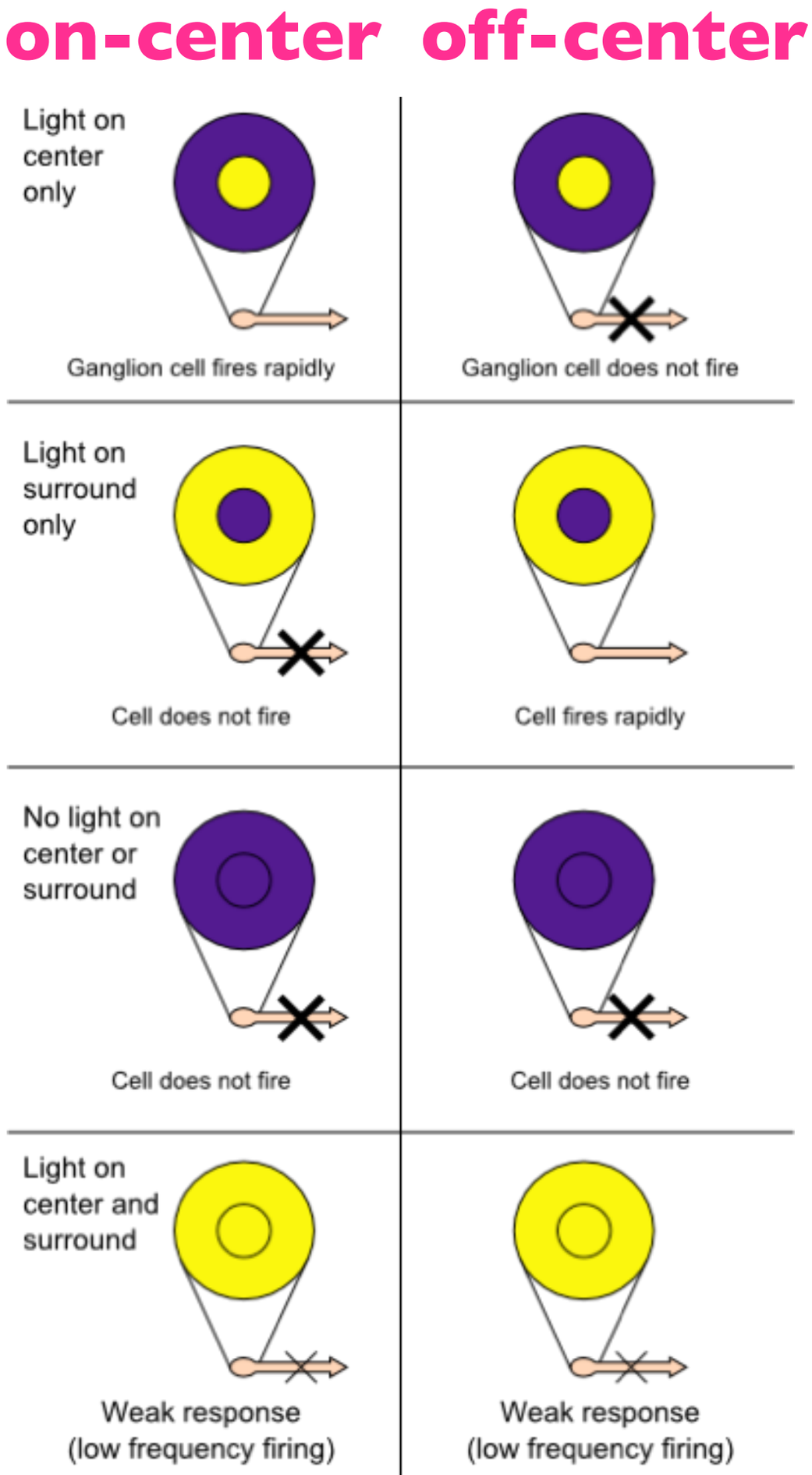
Over half of our visual processing power is concentrated in a slightly larger area called the parafovea.

Takeaway

Our visual system sees differences, not absolute values, and is attracted to edges.

Maximize the contrast with the background if the outlines of shapes are important.

retinal ganglion cells



source: wikipedia

Cornsweet Illusion

D. Purves and R. B. Lotto

Cornsweet Illusion

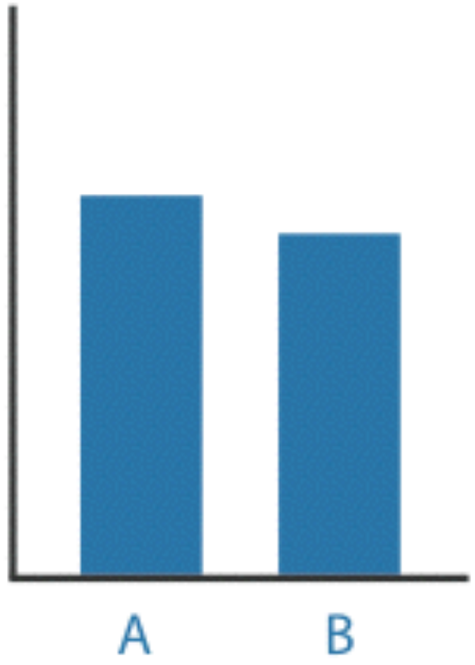
D. Purves and R. B. Lotto

WEBER'S LAW

we judge based on relative, not absolute, differences

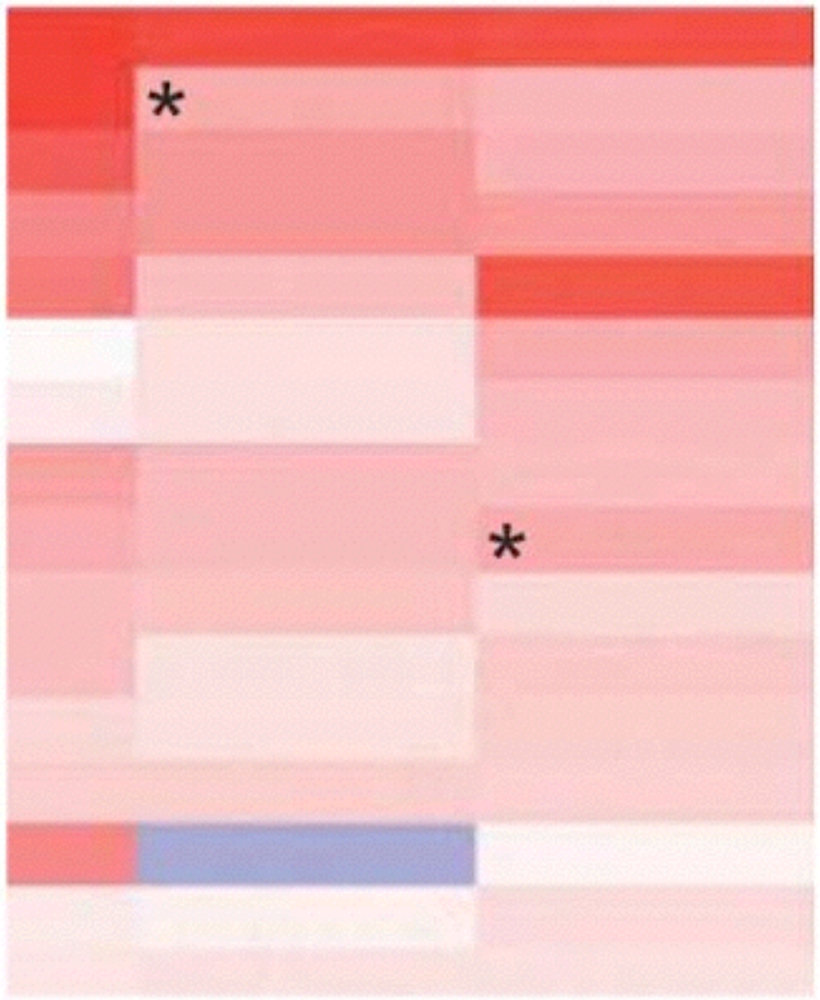
Unframed
Unaligned

Framed
Unaligned



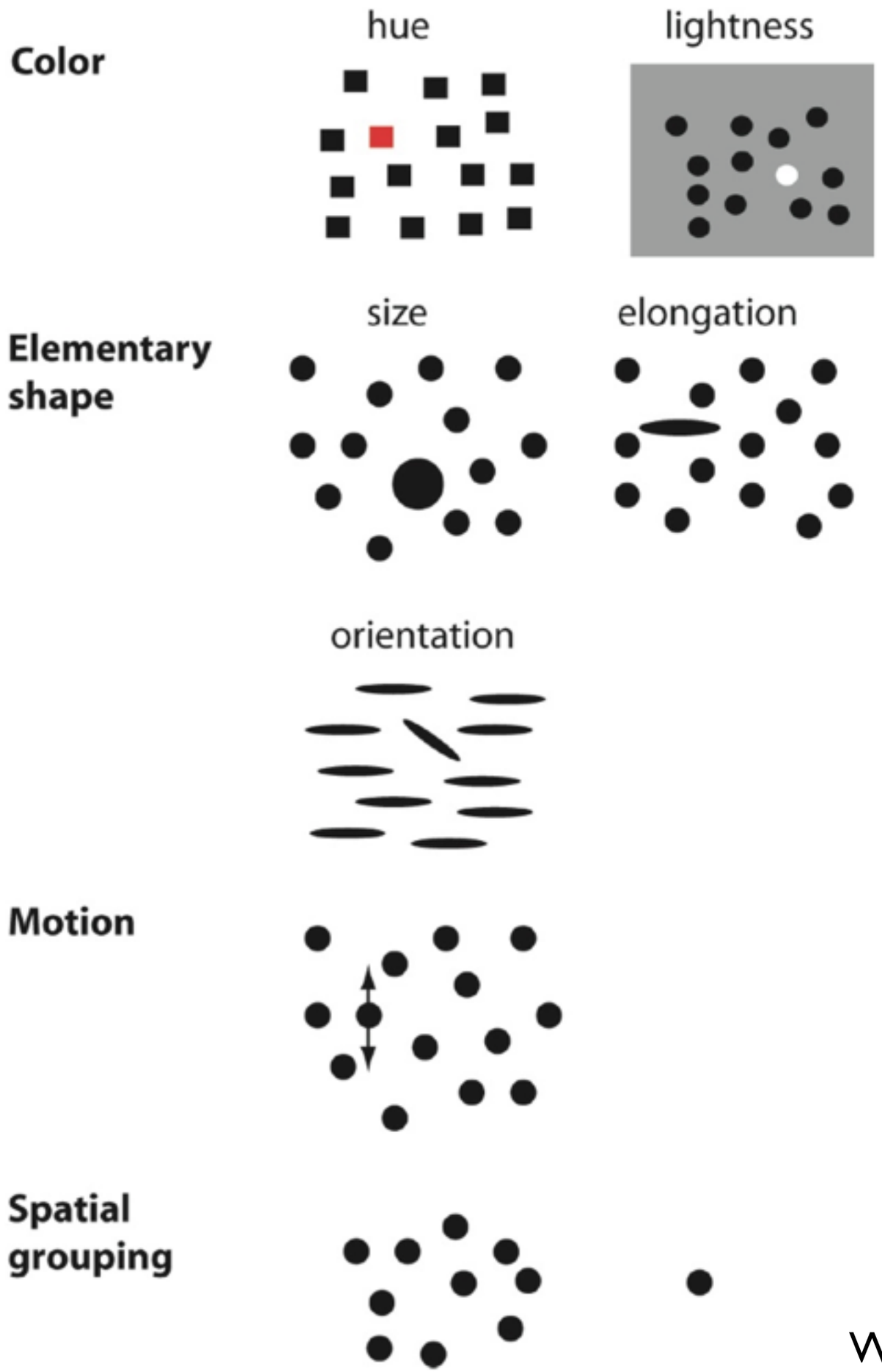
Unframed
Aligned

INTERACTION OF COLOR



Wong 2010

BASIC POPOUT CHANNELS



Ware 2008

Takeaway

We can easily see objects that are different in color and shape, or that are in motion.

Use color and shape sparingly to make the important information pop out.

Gestalt principles

- **similarity:** things that look like each other (size, color, shape) are related

- **proximity:** things that are visually close to each other are related

- **connection:** things that are visually connected are related

- **continuity:** we complete hidden objects into simple, familiar shapes

- **closure:** we see incomplete shapes as complete

- **figure / ground:** elements are perceived as either figures or background

- **common fate:** elements with the same moving direction are perceived as a unit

- data abstraction
- intro to Tableau (by Alex)

data abstraction

the *what* part of an analysis that pertains to the data

translation of domain-specific terms into words that are as generic as possible

type vs *semantics*

data types

- Items
- Attributes
- Links
- Positions
- Grids

dataset types

Tables

Items

Attributes

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box		7/17/07
32	7/16/07	2-High	Medium Box		7/18/07
32	7/16/07	2-High	Medium Box		7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/6/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/6/05	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

attribute

cell

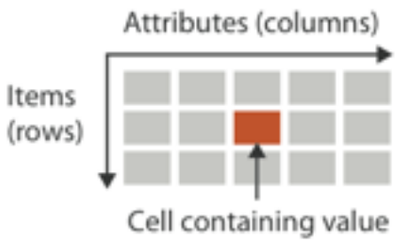
item

dataset types

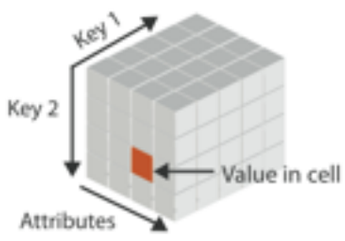
Tables

Items

Attributes



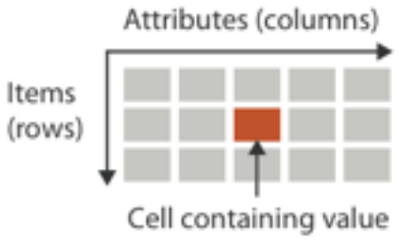
→ *Multidimensional Table*



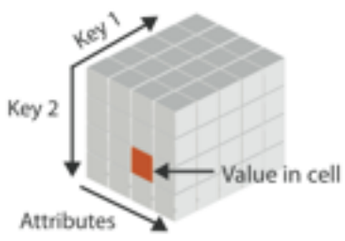
dataset types

Tables

- Items
- Attributes

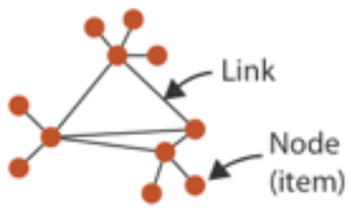


→ *Multidimensional Table*



Networks & Trees

- Items (nodes)
- Links
- Attributes

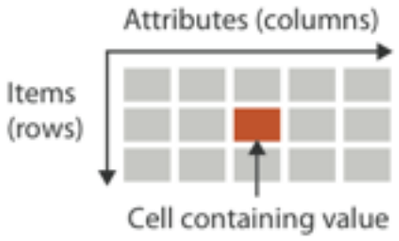


→ *Trees*

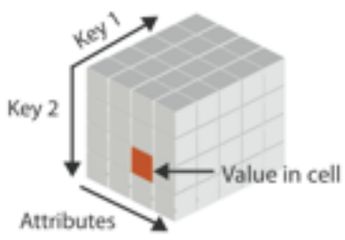
dataset types

Tables

- Items
- Attributes

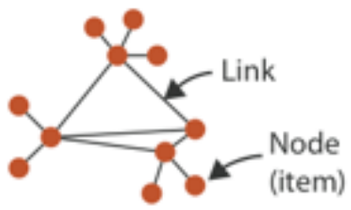


→ *Multidimensional Table*



Networks & Trees

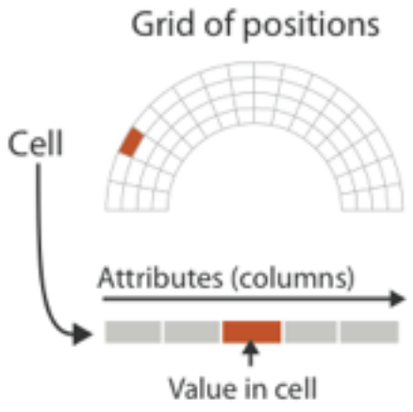
- Items (nodes)
- Links
- Attributes



→ *Trees*

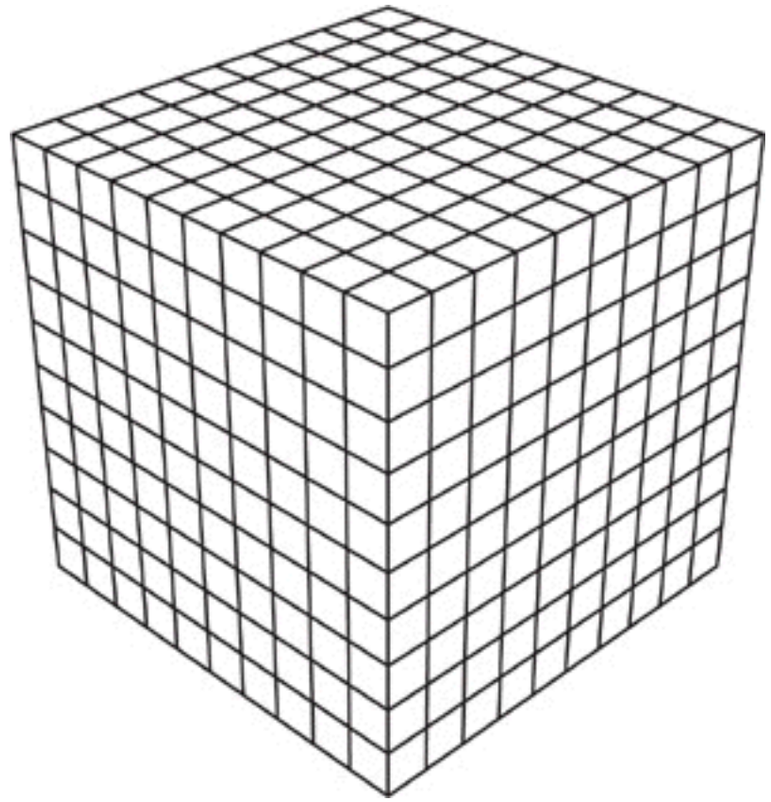
Fields

- Grids
  - Positions
- Attributes

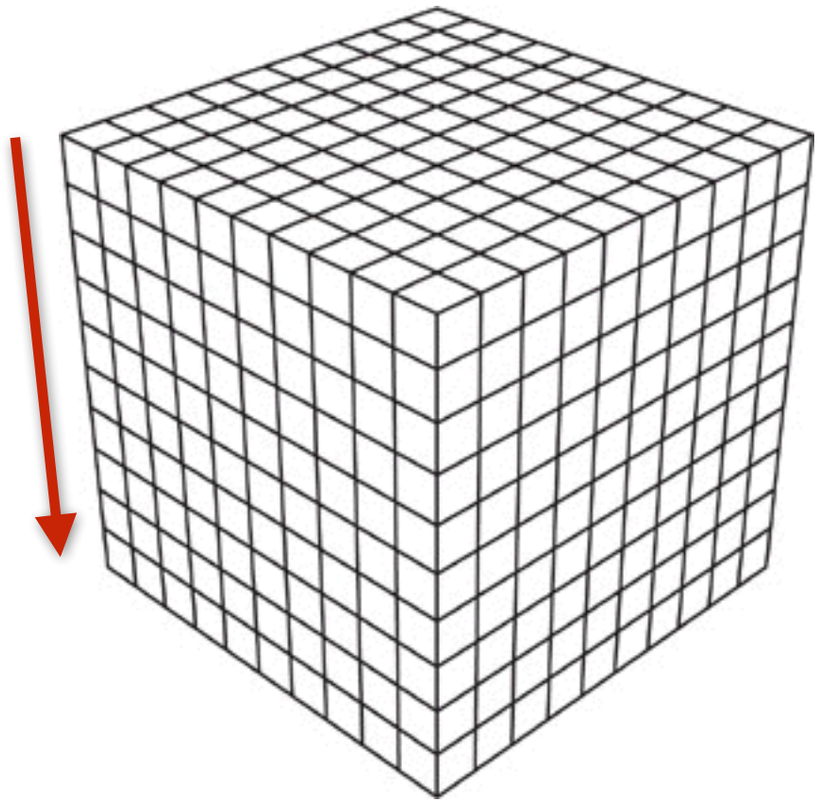


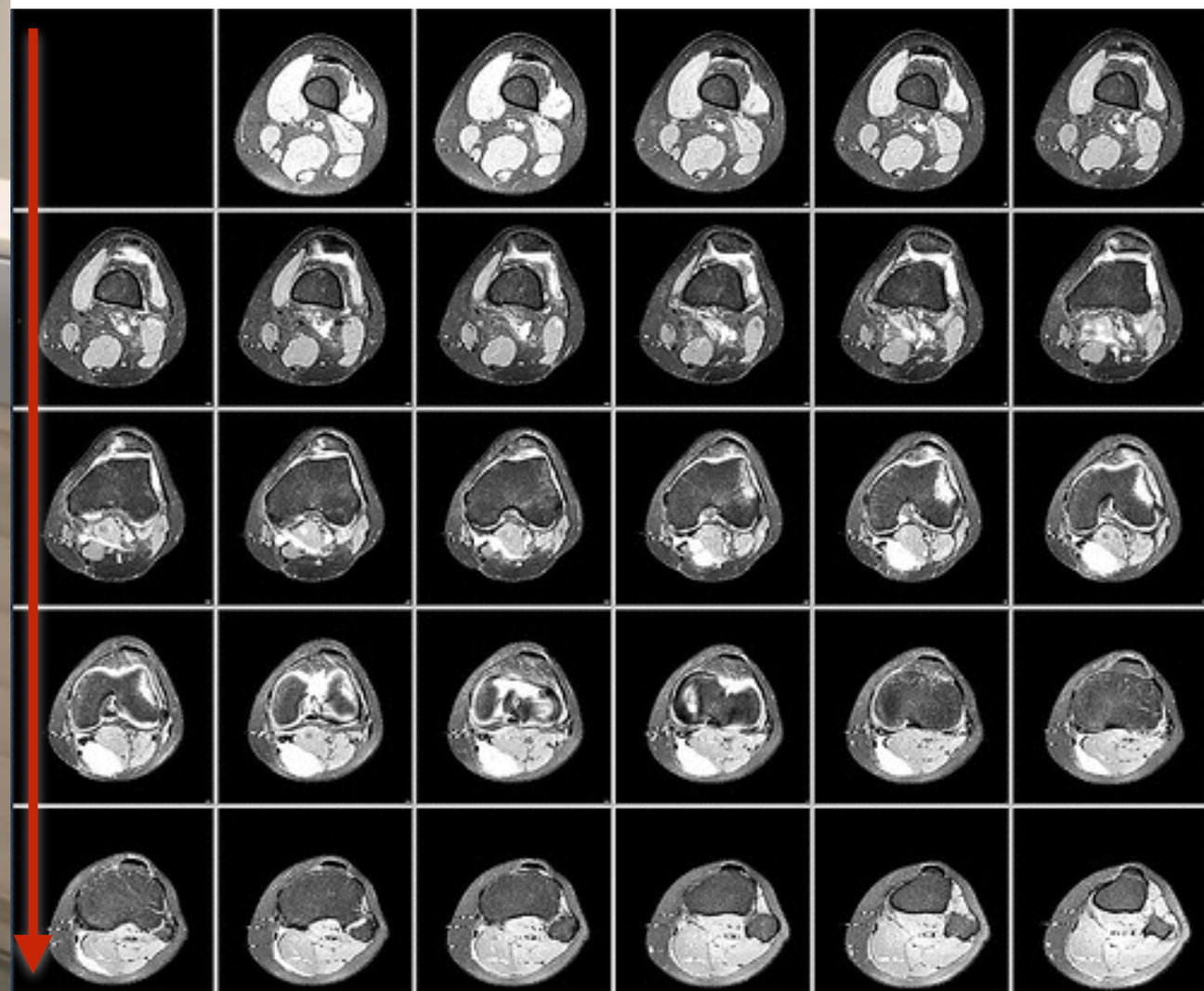

TOSHIBA


TOSHIBA

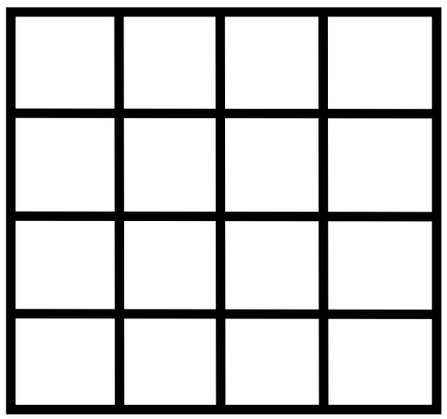



TOSHIBA

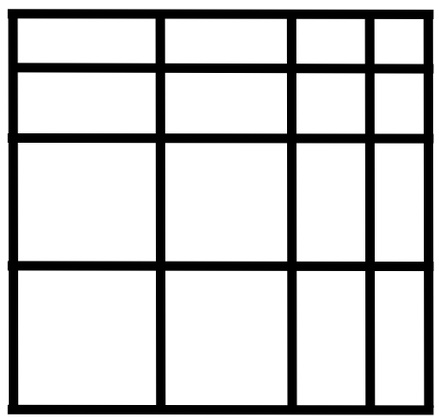




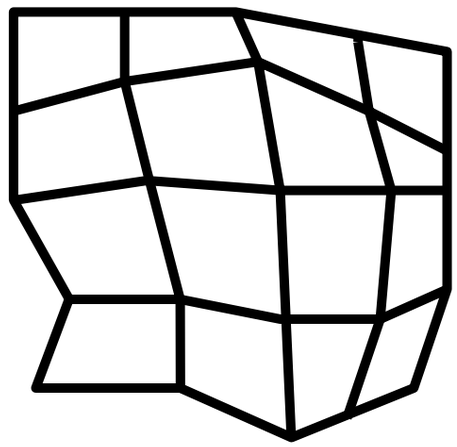
grid types



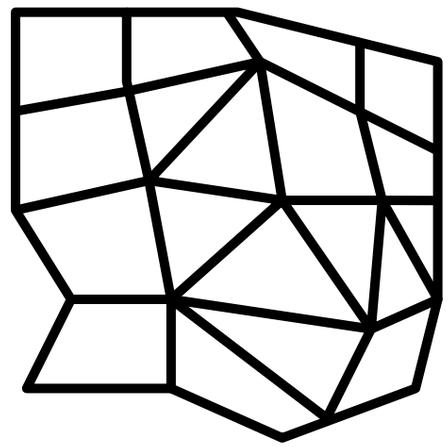
uniform



rectilinear



structured



unstructured

grid choices impact how continuous data is interpreted

two key considerations:

sampling, or the choice of where attributes are measured

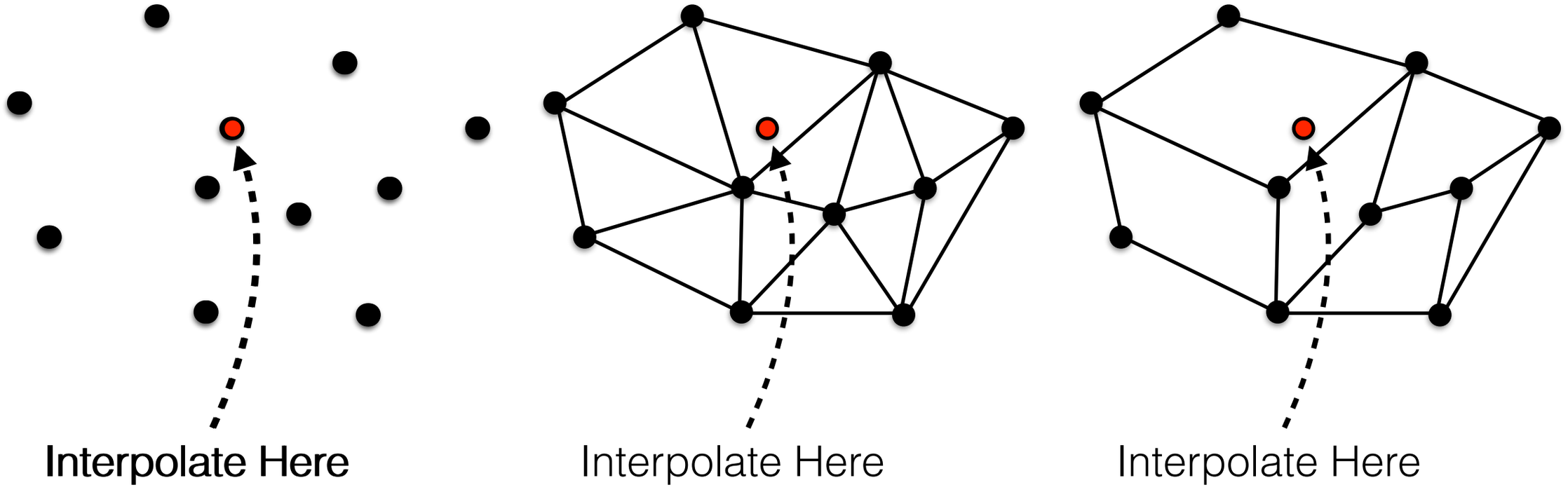
interpolation, or how to model the attributes in the rest of space

grid choices impact how continuous data is interpreted

two key considerations:

sampling, or the choice of where attributes are measured

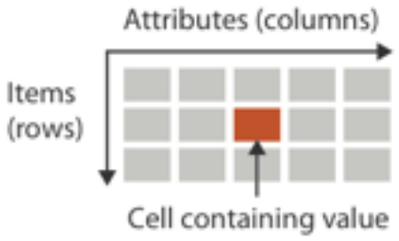
interpolation, or how to model the attributes in the rest of space



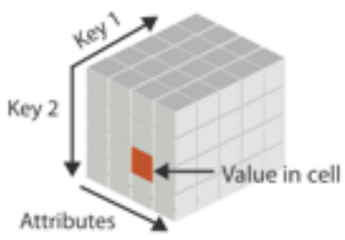
dataset types

Tables

- Items
- Attributes

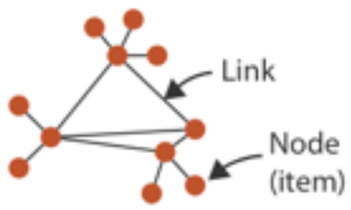


→ *Multidimensional Table*



Networks & Trees

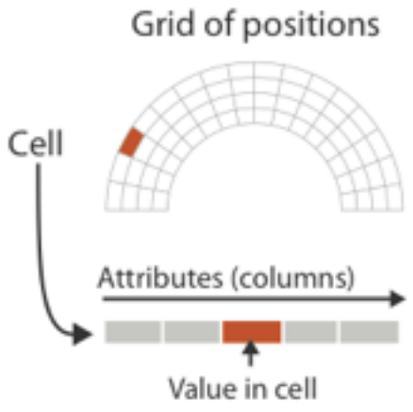
- Items (nodes)
- Links
- Attributes



→ *Trees*

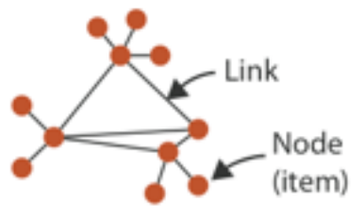
Fields

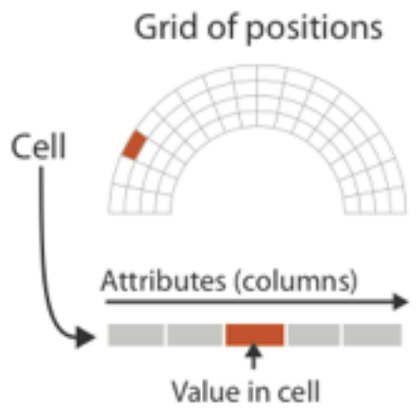
- Grids
  - Positions
- Attributes



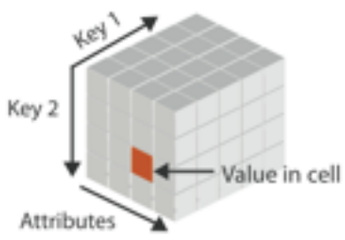
dataset types

Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes



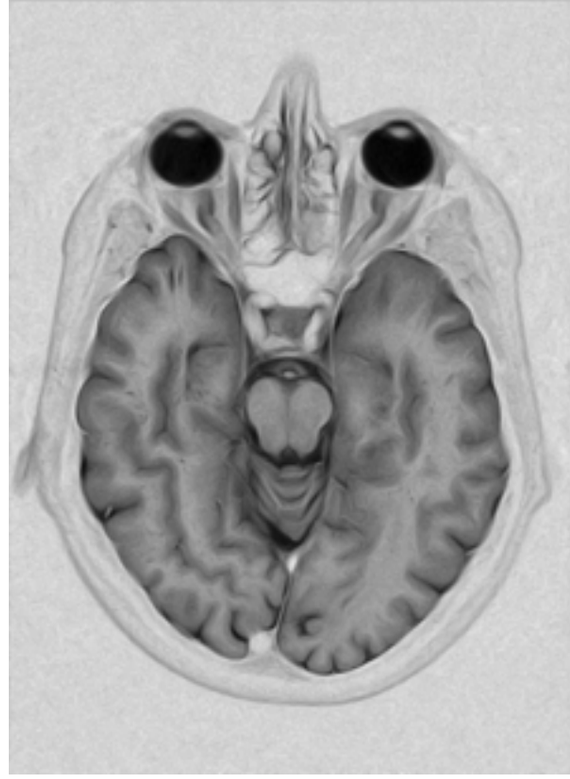


→ *Multidimensional Table*



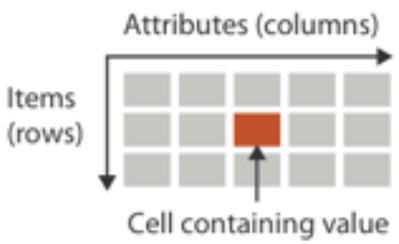
→ *Trees*

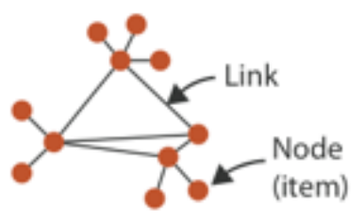
scalar

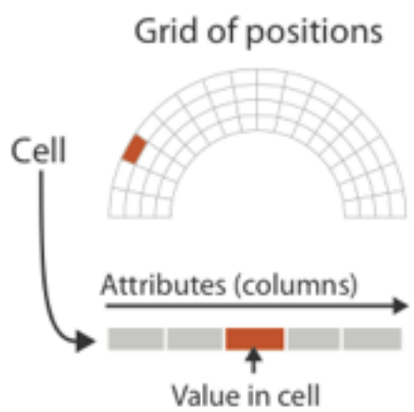


dataset types

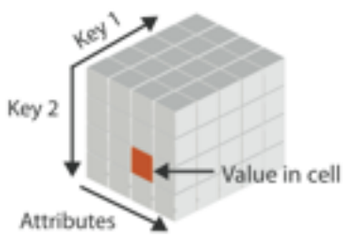
Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes





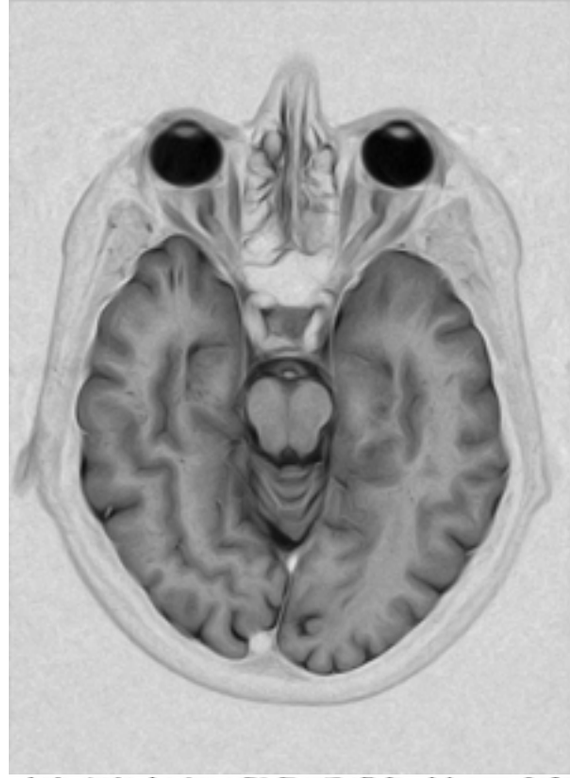


→ *Multidimensional Table*



→ *Trees*

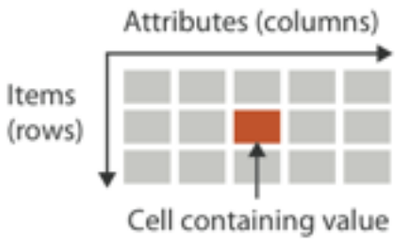
scalar

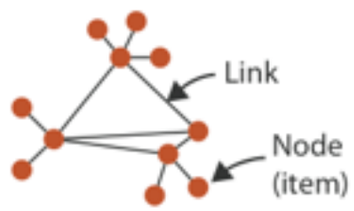


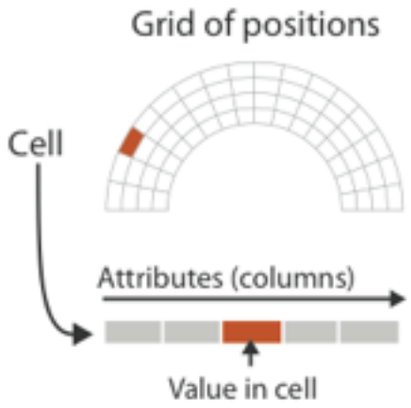
vector

dataset types

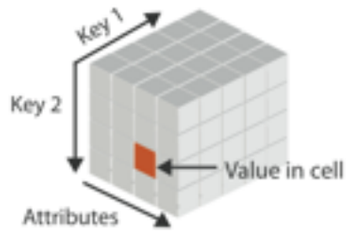
Tables	Networks & Trees	Fields
Items	Items (nodes)	Grids
Attributes	Links	Positions
	Attributes	Attributes





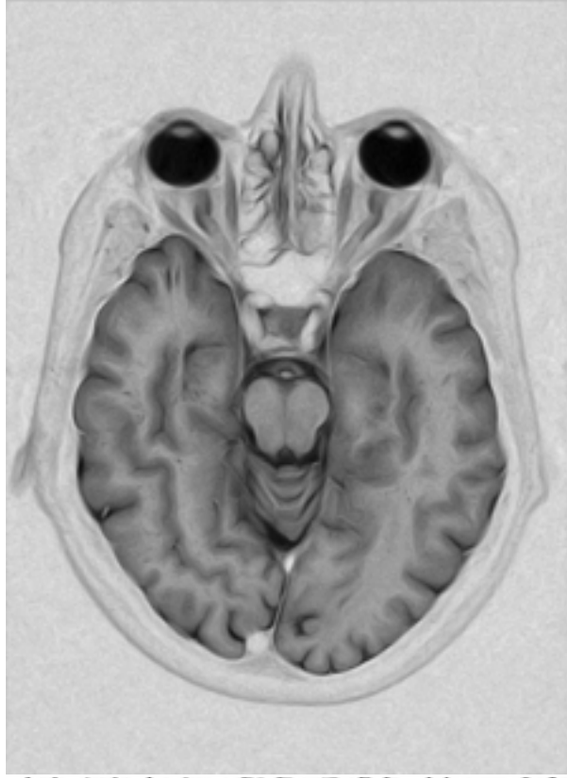


→ *Multidimensional Table*



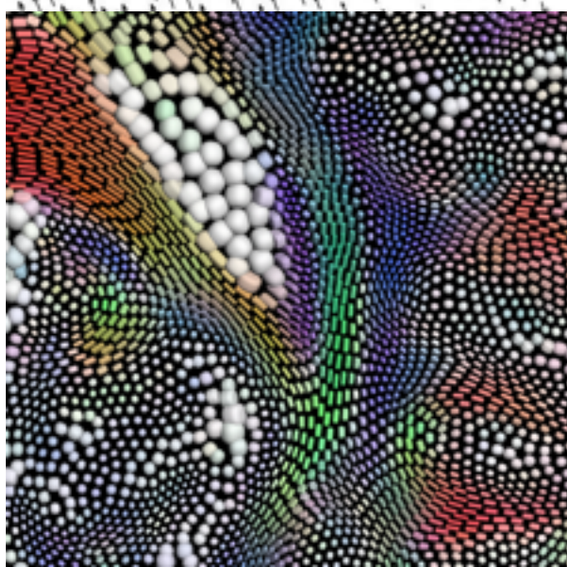
→ *Trees*

scalar



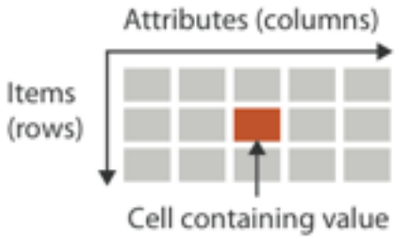
vector

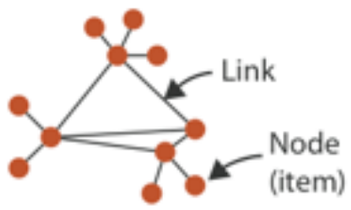
tensor

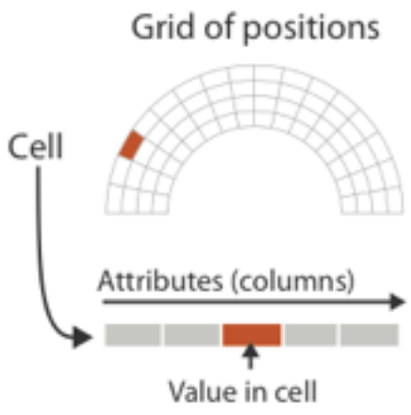


dataset types

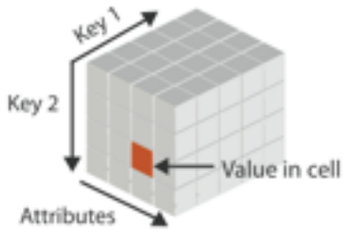
Tables	Networks & Trees	Fields	Geometry
Items	Items (nodes)	Grids	Items
Attributes	Links	Positions	Positions
	Attributes	Attributes	



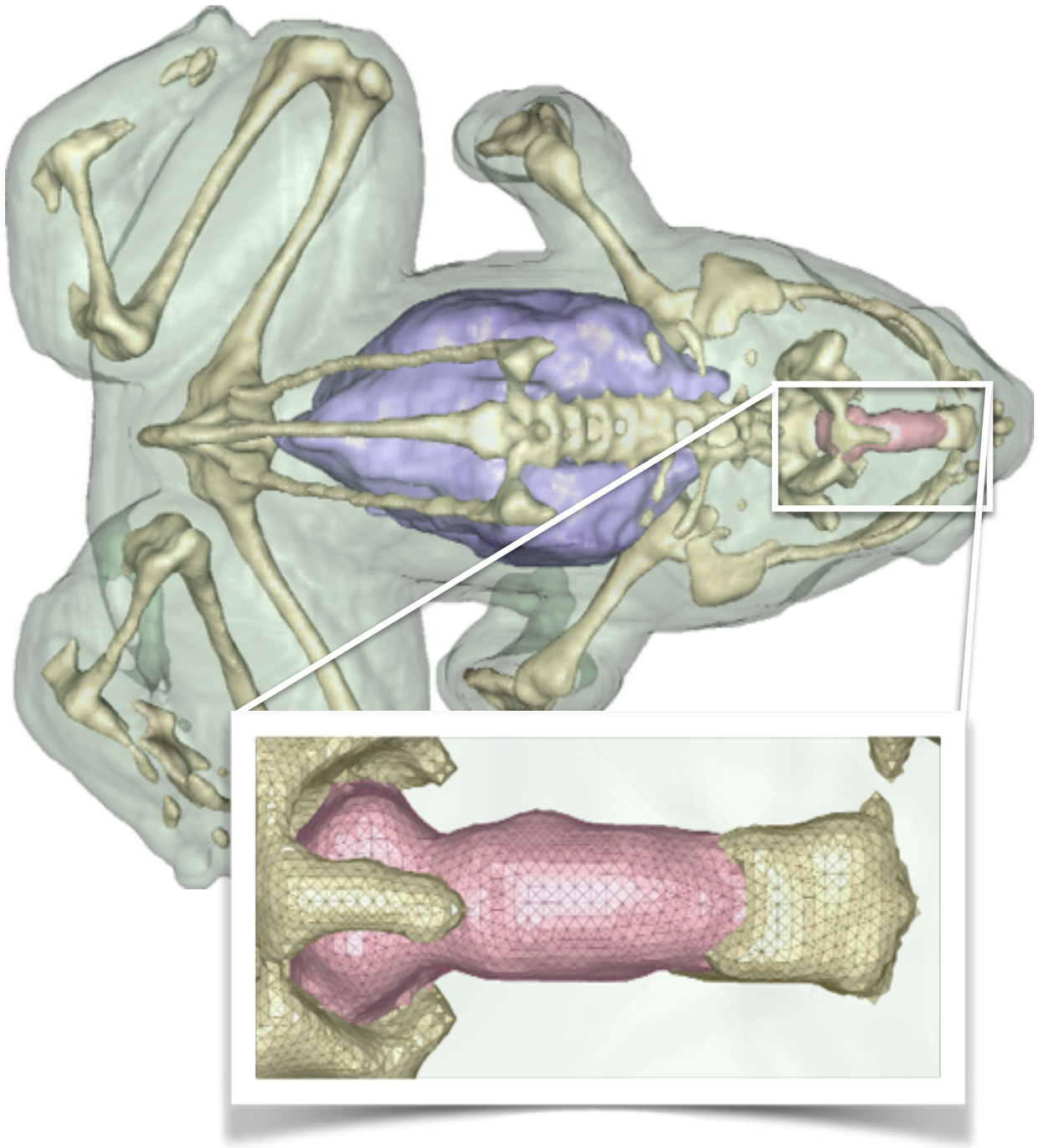




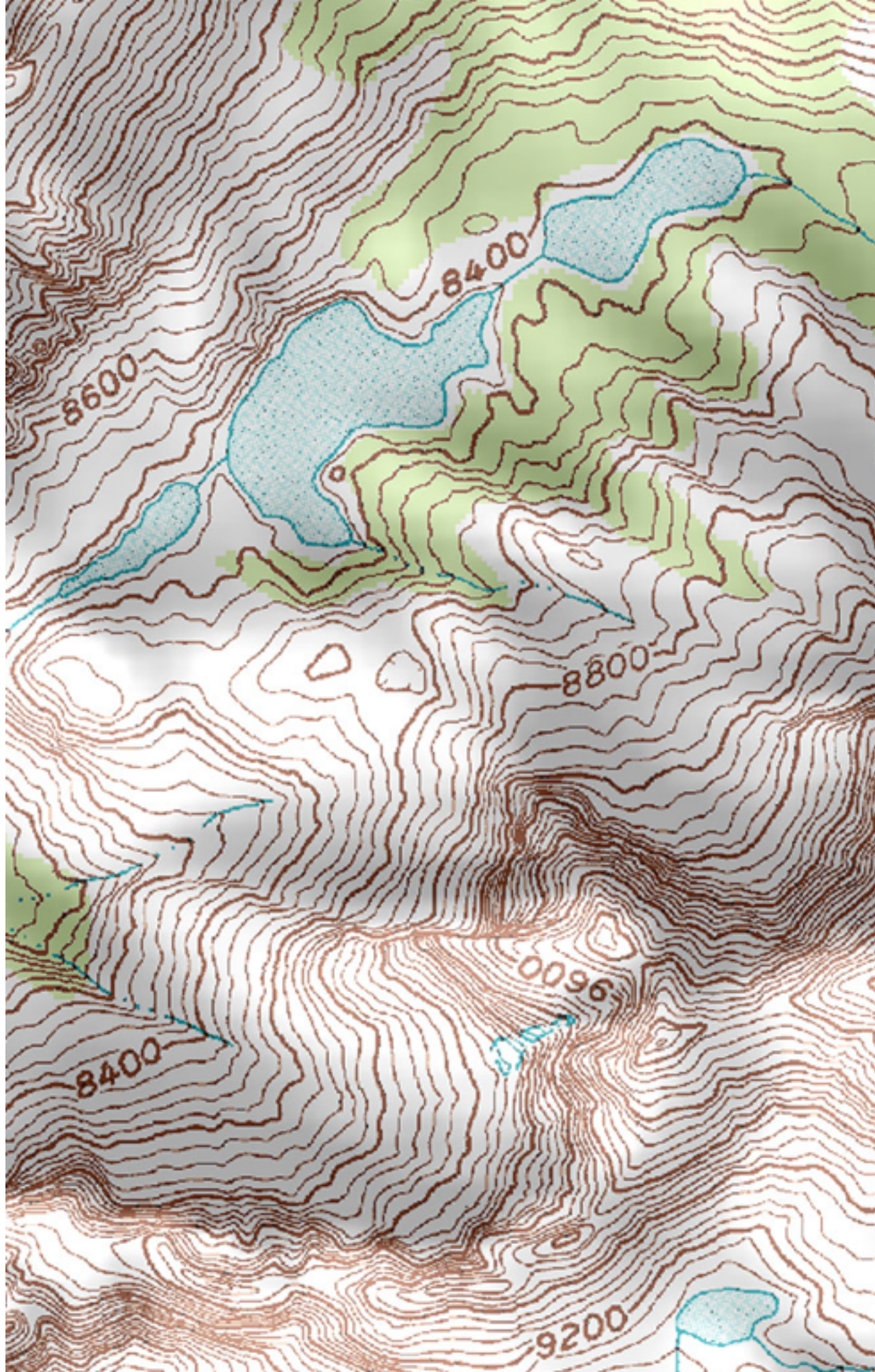
→ *Multidimensional Table*



→ *Trees*

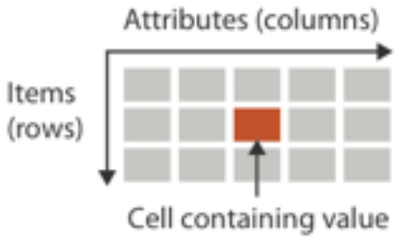


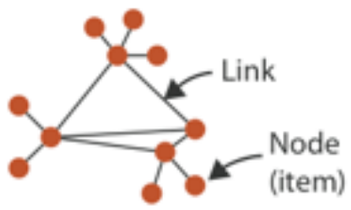
[Bronson 2014]

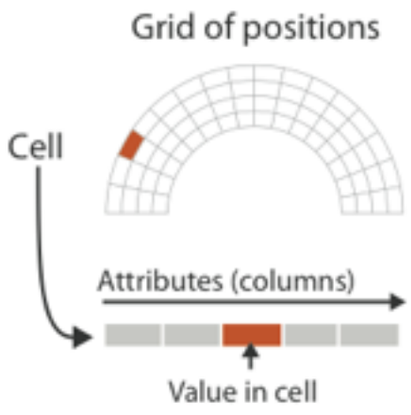


dataset types

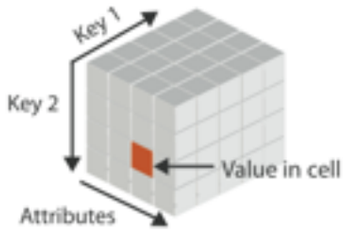
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		







→ *Multidimensional Table*



→ *Trees*

attribute types

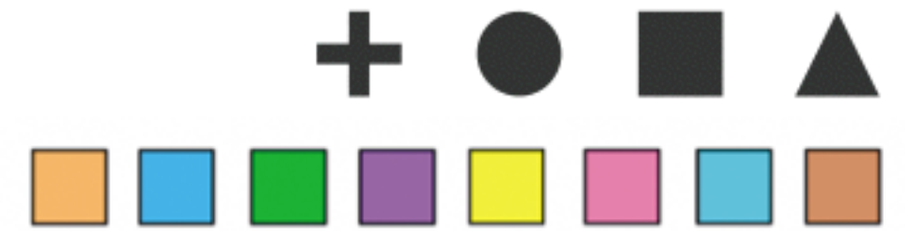
➔ Categorical

attribute types

- **Categorical**
  no implicit ordering

attribute types

➔ Categorical
no implicit ordering



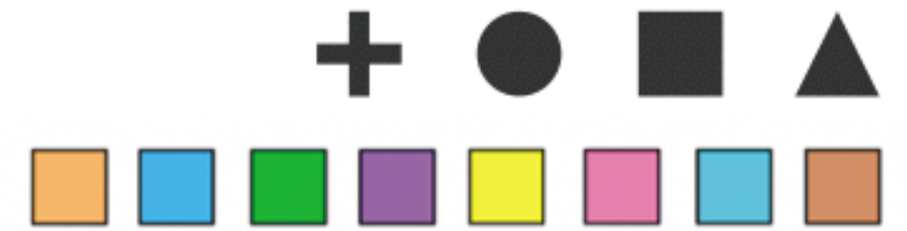
attribute types

- Categorical
  no implicit ordering

- Ordered
  - *Ordinal*
  - *Quantitative*

attribute types

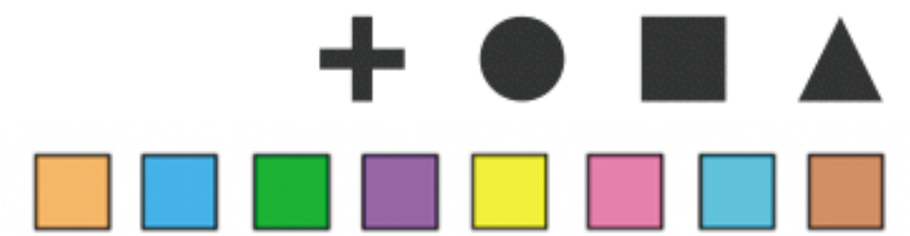
- Categorical
  no implicit ordering



- Ordered
  - *Ordinal*
  - *Quantitative*
    meaningful magnitude, can do arithmetic

attribute types

- Categorical
  no implicit ordering



- Ordered
  - *Ordinal*

  - *Quantitative*
    meaningful magnitude, can do arithmetic

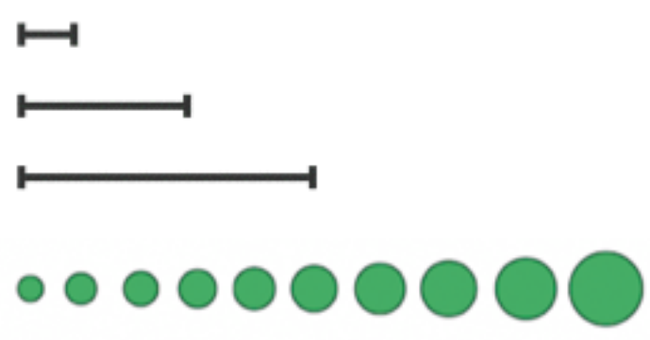
attribute types

➔ Categorical
no implicit ordering

➔ Ordered

➔ *Ordinal*

➔ *Quantitative*
meaningful magnitude,
can do arithmetic



A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Spec		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

quantitative
ordinal
categorical

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Spec		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

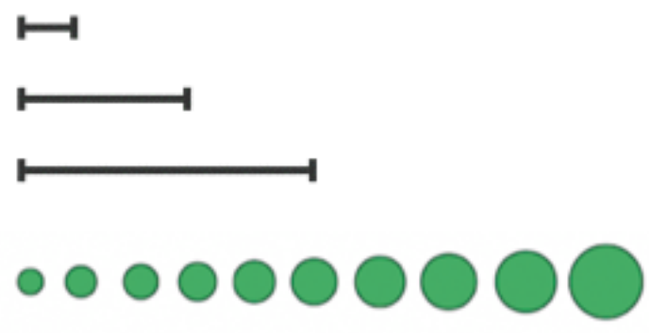
quantitative
ordinal
categorical

attribute types

- Categorical
  no implicit ordering

- Ordered
  - *Ordinal*

  - *Quantitative*
    meaningful magnitude, can do arithmetic

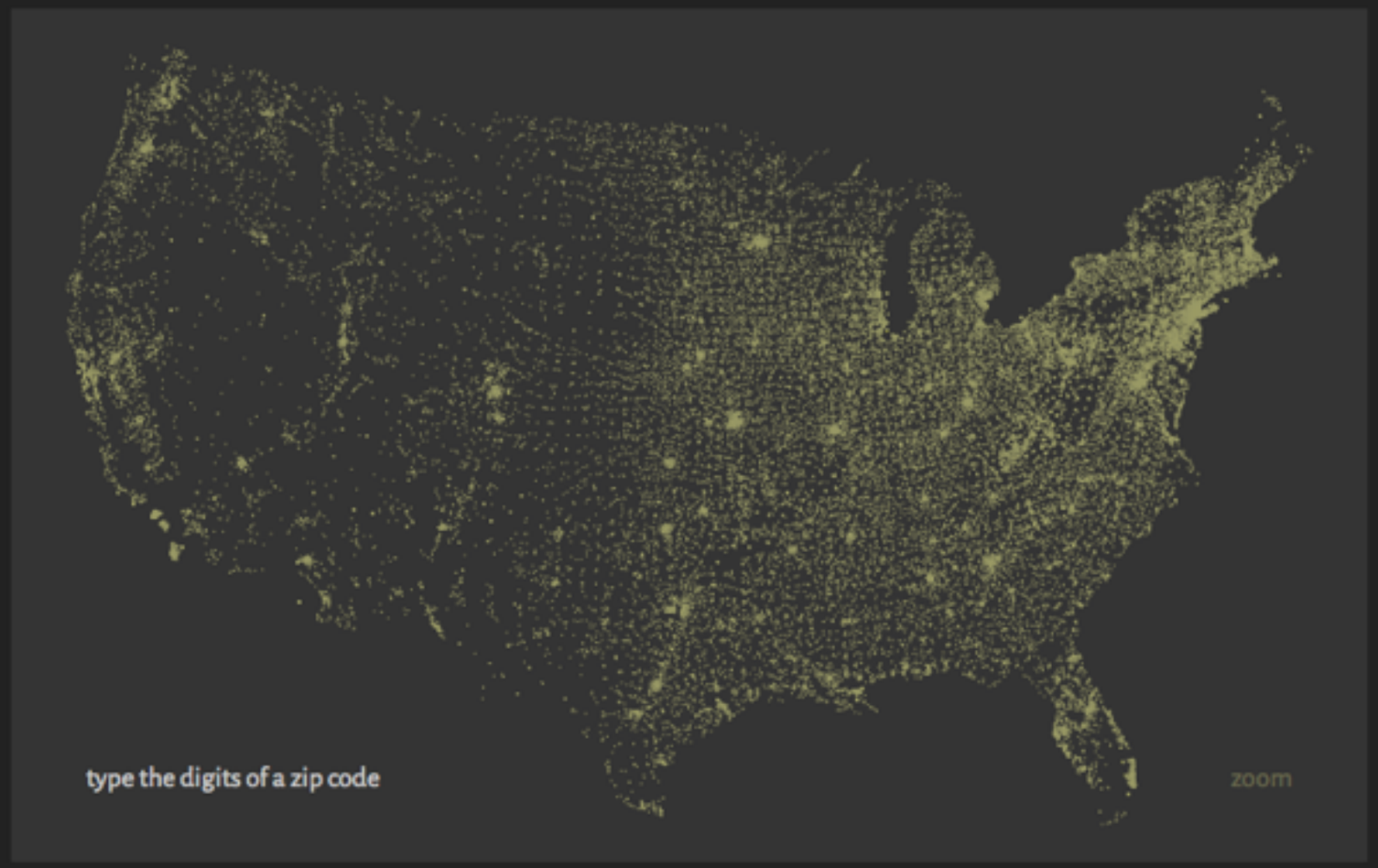


- Sequential

- Diverging

- Cyclic

- Hierarchical



Hit the letter **z**, or click the word **zoom** to enable or disable zooming.

zipdecode

This project began a very short sketch (a few hours) that I created in 1999 because I was curious about how the numbering works for postal codes in the states.

A detailed description of this project (and source code built with Processing) can be found in my book *Visualizing Data*.

Updated September 2004 to add several features over the original, including zoom, some new colors (thanks to Eugene Kuo), and a better zip code database (because of all the people who emailed and were sad that they couldn't find themselves).

special

attribute semantics

key *vs* value

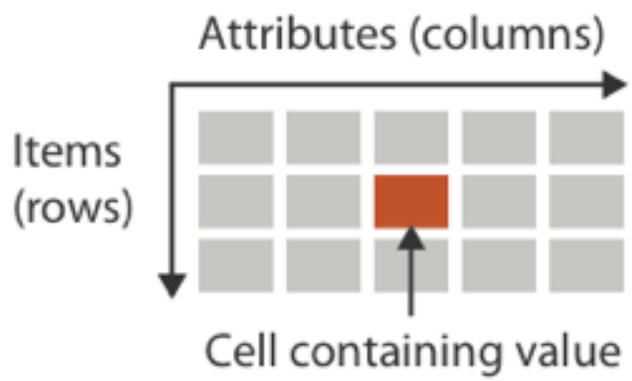
special

attribute semantics

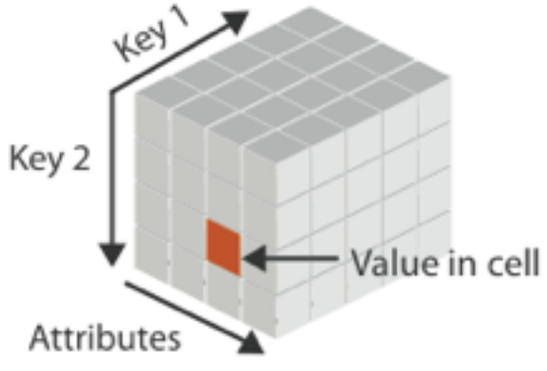
key *vs* value

tables

flat



multidimensional



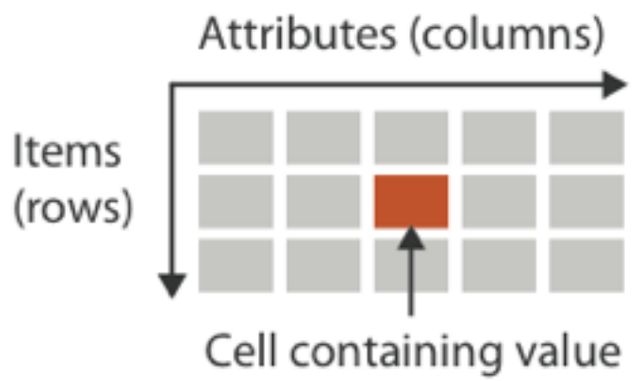
special

attribute semantics

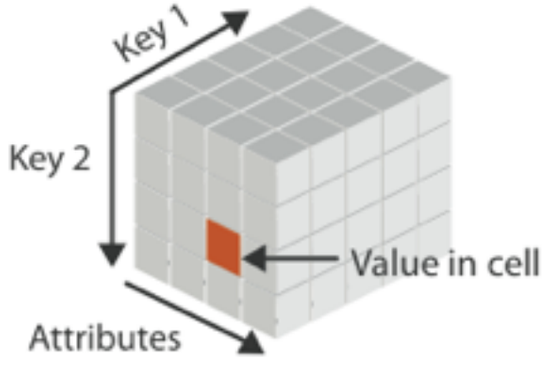
key *vs* value

tables

flat

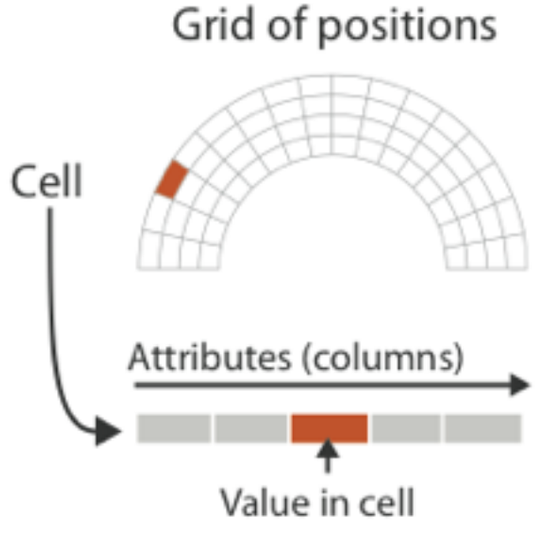


multidimensional



fields

Grid of positions



special

attribute semantics

temporal

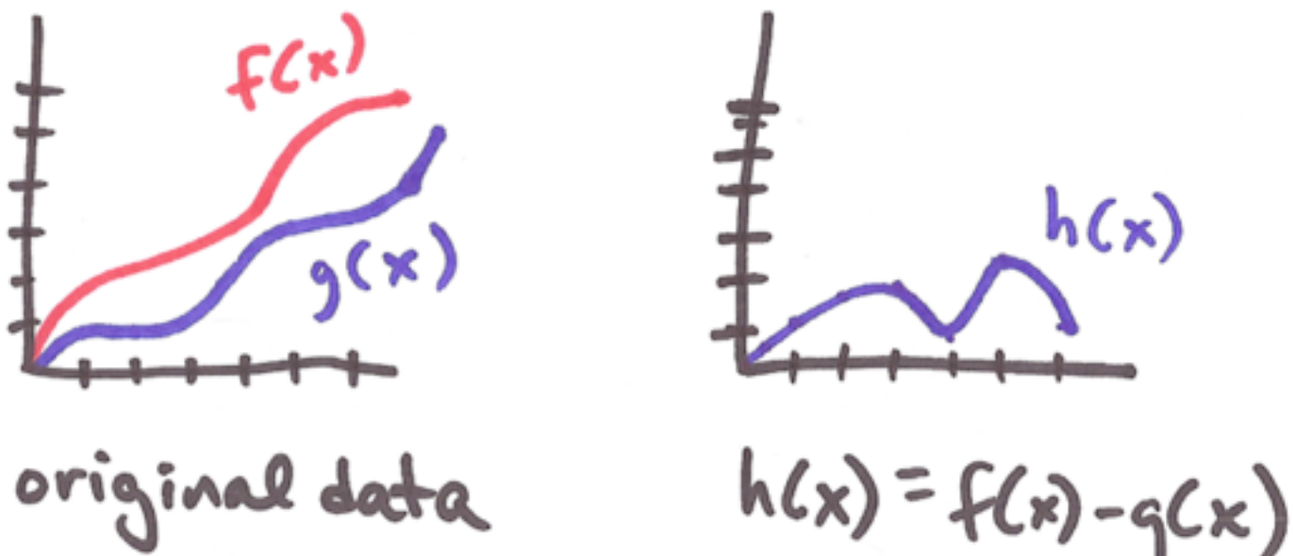
what makes time special?

abstraction exercise . . .


ESPRESSO
& HOT TEA
8oz SMALL
12oz MEDIUM
16oz LARGE
GUATEMALA CASI CIELO
Floral, lemon & cocoa
1.50 1.70 1.90
GUATEMALA CASI CIELO de-caf
Floral, lemon & cocoa
1.50 1.70 1.90
SUMATRA
spicy, herbal & earthy
1.50 1.70 1.90
ESPRESSO 1.45 1.70 1.85
AMERICANO 1.60 1.80 2.00
CAFÉ LATTE 2.15 2.75 3.20
CAPPUCCINO 2.15 2.75 3.20
CAFÉ MOCHA 2.65 3.05 3.55
ORGANIC BREAKFAST 1.70 1.90 2.10
ORGANIC LONG LIFE GREEN TEA 1.70 1.90 2.10
MONSOON CHAI 1.70 1.90 2.10
CHAI TEA LATTE 2.40 2.95 3.35
BLACK TEA LATTE 2.20 2.55 3.20
HOT CHOCOLATE 2.50 2.75 3.00
HOMEMADE SYRUP FLAVORS .50 each

DERIVED ATTRIBUTES

- **derived attribute: compute from originals**
  - simple change of type
  - acquire additional data
  - complex transformation
  - transformation is abstraction choice



DATA MODEL vs CONCEPTUAL MODEL

- **data model:** mathematical abstraction (data abstraction)
  - set with operations, eg. floats with * / - +

- **conceptual model:** mental construction (semantics)
  - includes semantics, supports reasoning

- **conceptual model motivates derived data (data abstraction choices)**

EXAMPLE

- **from data model . . .**
  - 32.52, 54.06, -17.35, . . . (floats)
- **using conceptual model . . .**
  - temperature
- **to new data abstraction.**
  - continuous to 2 significant figures (Q)
  - hot, warm , cold (O)
  - above freezing, below freezing (C)

another abstraction exercise . . .


Arena
301 W. South Temple
Temple Square
132 W. South Temple
City Center
100 S. Main St
Planetarium
125 S. 400 West
Gallivan Plaza
300 S. Main St
Old Greektown
525 W. 200 South
Central Station
250 S. 600 West
Courthouse
450 S. Main St
Library
225 E. 400 South
Trolley
625 E. 400 South
900 East
875 E. 400 South
Stadium
1349 E. 500 South
South Campus
1790 E. South Campus Dr
Fort Douglas
200 S. Mario Capecchi Dr
University Medical Center
10 N. Mario Capecchi Dr
900 South
860 S. 200 West
Ball Park
180 W. 1300 South
Central Pointe
221 W. 2100 South
River Trail
2340 S. 1070 West
Redwood Junction
1740 W. Research Way
Decker Lake
3070 S. Decker Lake Blvd
West Valley Central
2750 W. Lehman Ave
Millcreek
210 W. 3300 South
Meadowbrook
188 W. 3900 South
Murray North
71 W. Fireclay Ave
Murray Central
140 W. Vine St
Fashion Place West
222 W. Winchester St
Midvale Fort Union
180 W. 7250 South
Midvale Center
95 W. Center St
Historic Sandy
165 E. 9000 South
Sandy Expo
115 E. 9400 South
Sandy Civic Center
115 E. Sego Lily Dr
Bingham Junction
7387 S. Bingham Jct Blvd
Historic Gardner
1127 W. 7800 South
West Jordan City Center
8021 S. Redwood Rd
2700 West
8351 S. 2700 West
Jordan Valley
3400 W. 8600 South
4800 West
4773 W. Old Bingham Hwy
5600 West
5651 W. Old Bingham Hwy
South Jordan Parkway
10605 S. Grandville Ave
Daybreak Parkway
11405 S. Grandville Ave
UTAH TRANSIT AUTHORITY
TRAX
LINES
Blue (701)
(Downtown – Sandy)
Red (703)
(University of Utah – Daybreak)
Green (704)
(Downtown – West Valley City)
Station
Station with park-and-ride
AUGUST 2011

L5. Visual Encodings

REQUIRED READING

Chapter 5

Marks and Channels

5.1 The Big Picture

Marks are basic geometric elements that depict items or links, and channels control their appearance. The effectiveness of a channel for encoding data depends on its type: the channels that perceptually convey magnitude information are a good match for ordered data, and those that convey identity information with categorical data. Figure 5.1 summarizes the channel rankings.

5.2 Why Marks and Channels?

Learning to reason about marks and channels gives you the building blocks for analyzing visual encodings. The core of the design space of visual encodings can be described as an orthogonal combination of two aspects: graphical elements called marks, and visual channels to control their appearance. Even complex visual encodings can be broken down into components that can be analyzed in terms of their marks and channel structure.

5.3 Defining Marks and Channels

Chapter 6

Rules of Thumb

6.1 The Big Picture

This chapter contains **rules of thumb**: advice and guidelines. Each of them has a catchy title in hopes that you'll remember it as a slogan. The rules of thumb covered in this chapter are

- No Unjustified 3D
  - The Power of the Plane
  - The Disparity of Depth
  - Occlusion Hides Information
  - Perspective Distortion Dangers
  - Tilted Text Isn't Legible
- No Unjustified 2D
- Eyes Beat Memory
- Resolution over Immersion
- Overview First, Zoom and Filter, Detail on Demand
- Responsiveness Is Required
- Get It Right in Black and White
- Function First, Form Next

Chapter 10

Map Color and Other Channels

10.1 The Big Picture

This chapter covers the mapping of color and other nonspatial channels in visual encoding design choices, summarized in Figure 10.1. The colloquial term *color* is best understood in terms of three separate channels: luminance, hue, and saturation. The major design choice for colormap construction is whether the intent is to distinguish between categorical attributes or to encode ordered attributes. Sequential ordered colormaps show a progression of an attribute from a minimum to a maximum value, while diverging ordered colormaps have a visual indication of a zero point in the center where the attribute values diverge to negative on one side and positive on the other. Bivariate colormaps are designed to show two attributes simultaneously using carefully designed combinations of luminance, hue, and saturation.

The characteristics of several more channels are also covered: the magnitude channels of size, angle, and curvature and the identity channels of shape and motion.

10.2 Color Theory

Intro to Tableau