CSE 557A | Sep 26, 2016

INFORMATION VISUALIZATION

Alvitta Ottley

Washington University in St. Louis

Slide Credits:
Mariah Meyer, University of Utah
Remco Chang, Tufts University

Announcements

Assignment 2: Bar and Line Charts
Due: 02-07-2017, 11:59pm (midnight)

In this assignment, you will be using Processing to draw a bar chart and a line chart. This is your first assignment with Processing and you will be learning the basics of Processing such as, handling mouse events, basic intersection detection, and keeping track of the state of the visualization. You will be required to display a visualization based on input data, and there are a few concepts that you need to explore: (1) reading and parsing data; (2) mouse hovering for highlighting visual elements; and (3) animated transition.

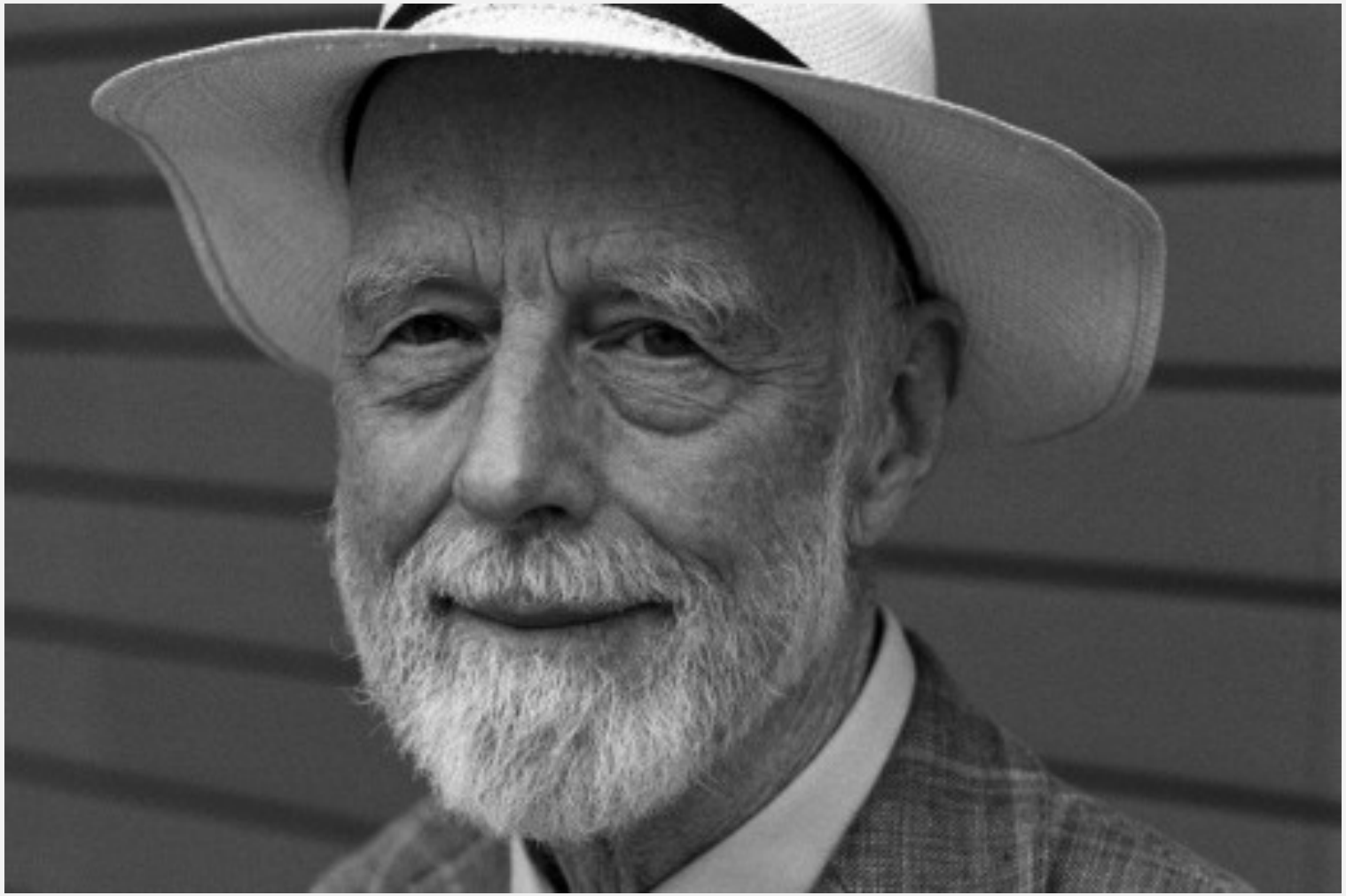
Due Tonight

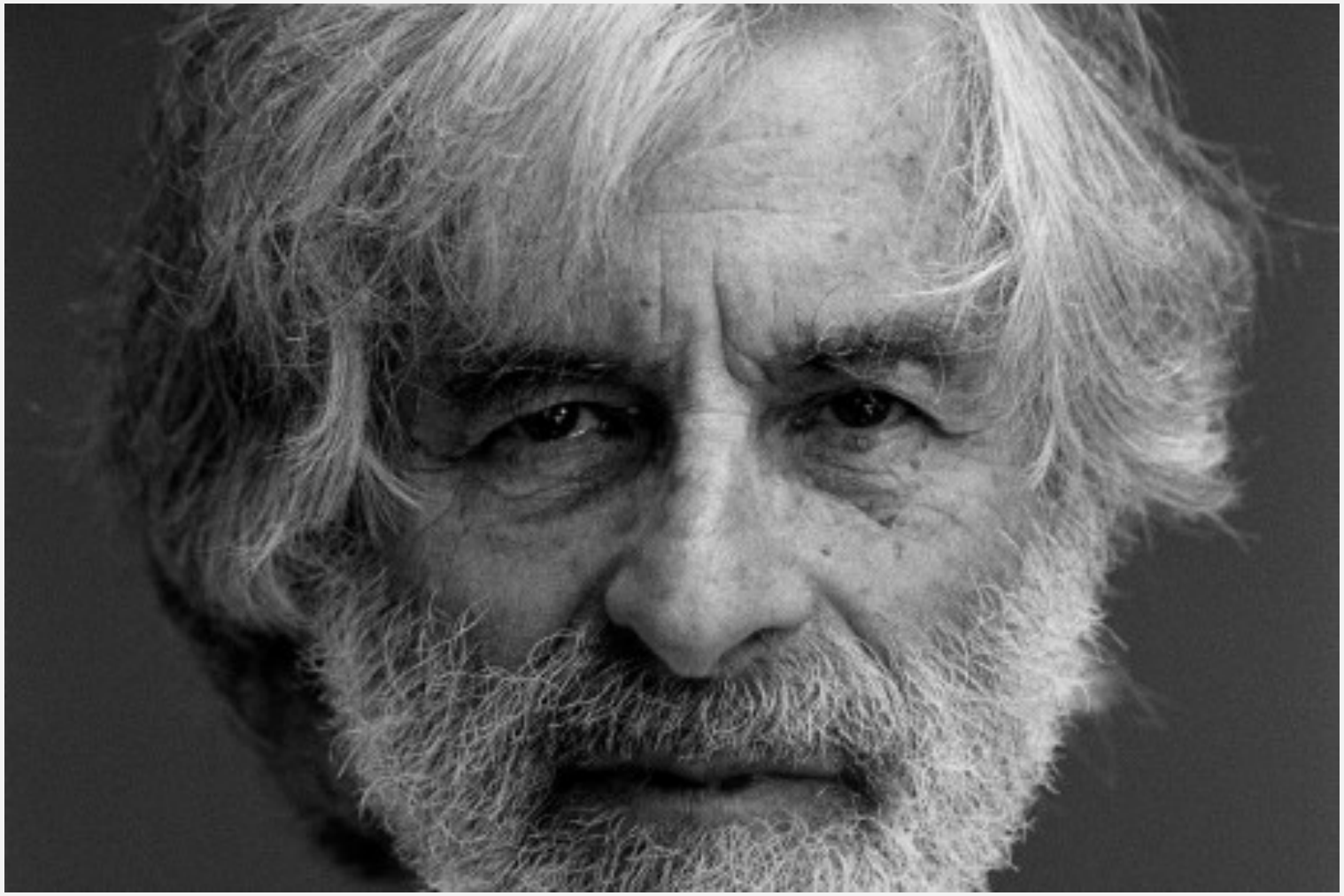
Basic Requirements:

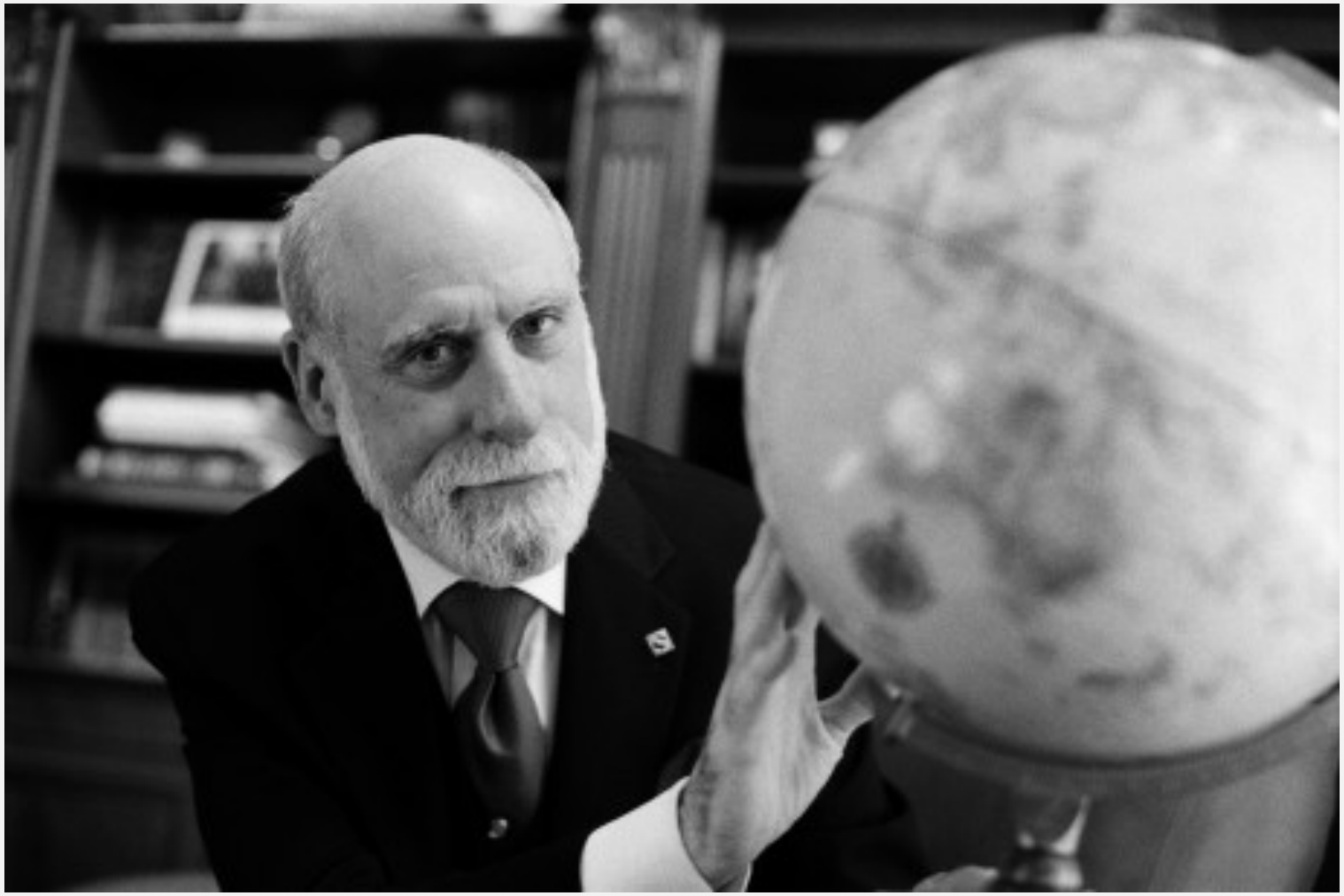
1. **Dataset**
   You will be given a simple comma delimited file (CSV) called "data.csv". This file will have the following properties:
   a. It has two columns, the first column contains categorical (ordinal) data, and the second column contains quantitative data.
   b. The first row has labels for each column.
   c. There are around 10 rows of data.

2. **Bar and Line Charts**
   In Processing, do the following:

HEIDELBERG LAUREATE FORUM



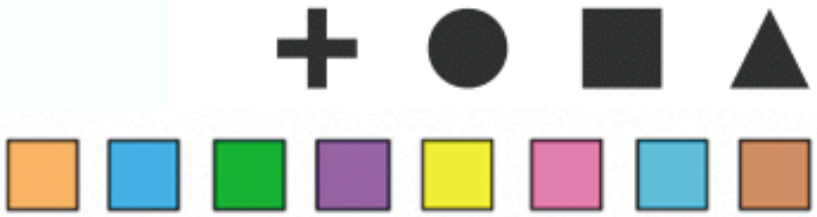




Recap on data mapping

Summary

➔ Categorical
no implicit ordering



➔ Ordered

➔ *Ordinal*

➔ *Quantitative*
meaningful magnitude, can do arithmetic

Slide courtesy of Mariah Meyer

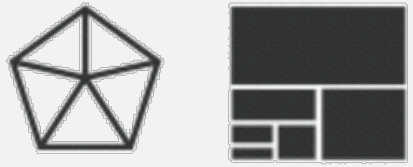
MARKS

-graphical element in an image

-classified according to number of spatial dimensions required

points (0D)

lines (1D)



areas (2D)

CHANNELS

-parameters that control the appearance of marks

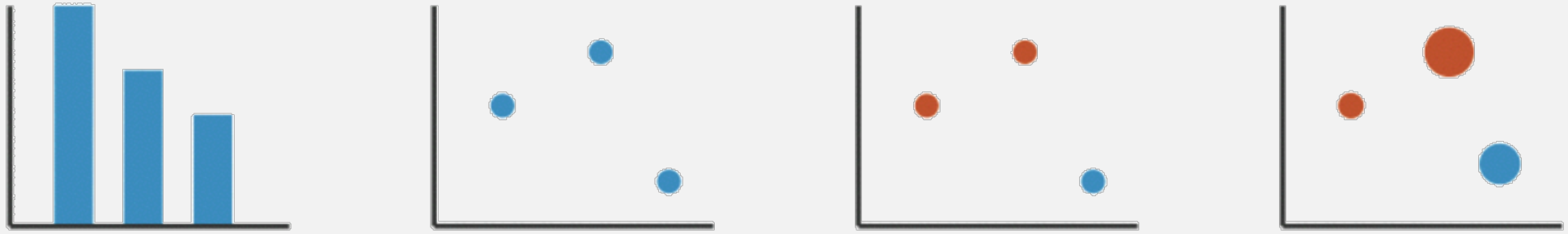
- **Position**
  - Horizontal
  - Vertical
  - Both

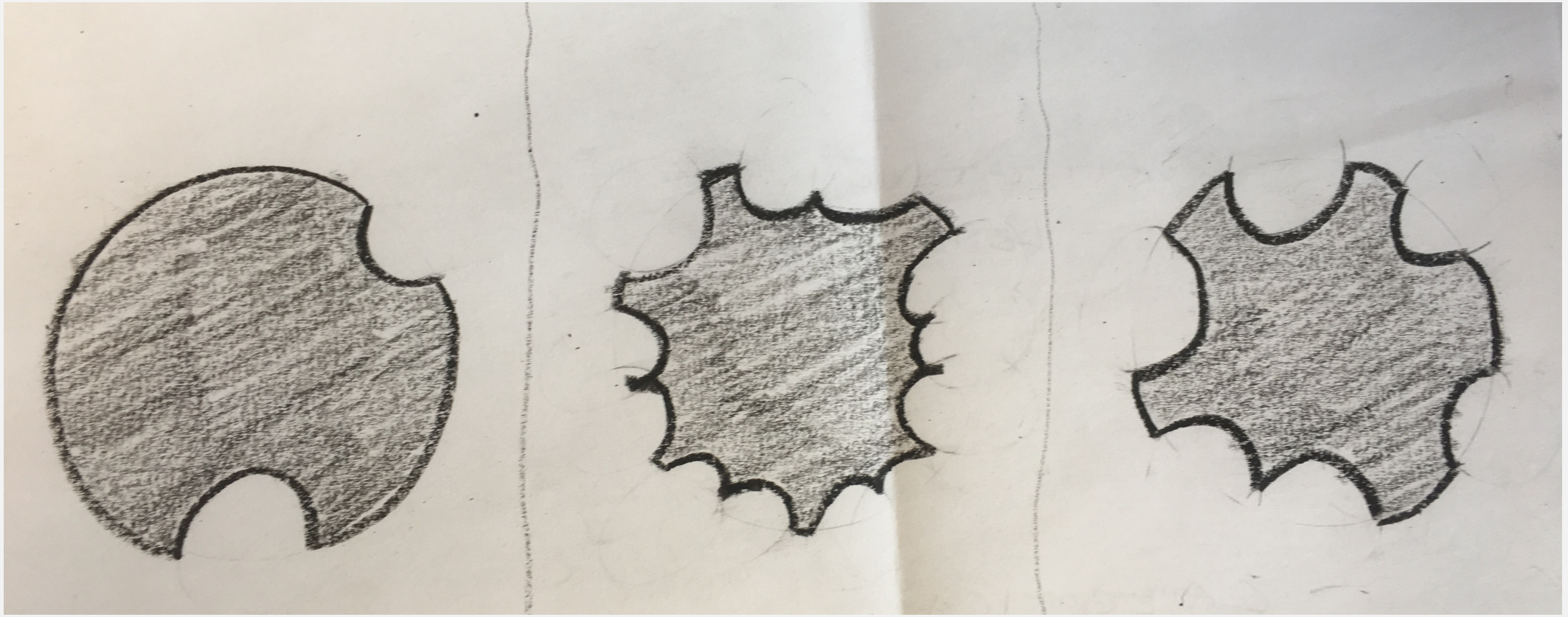
- **Color**

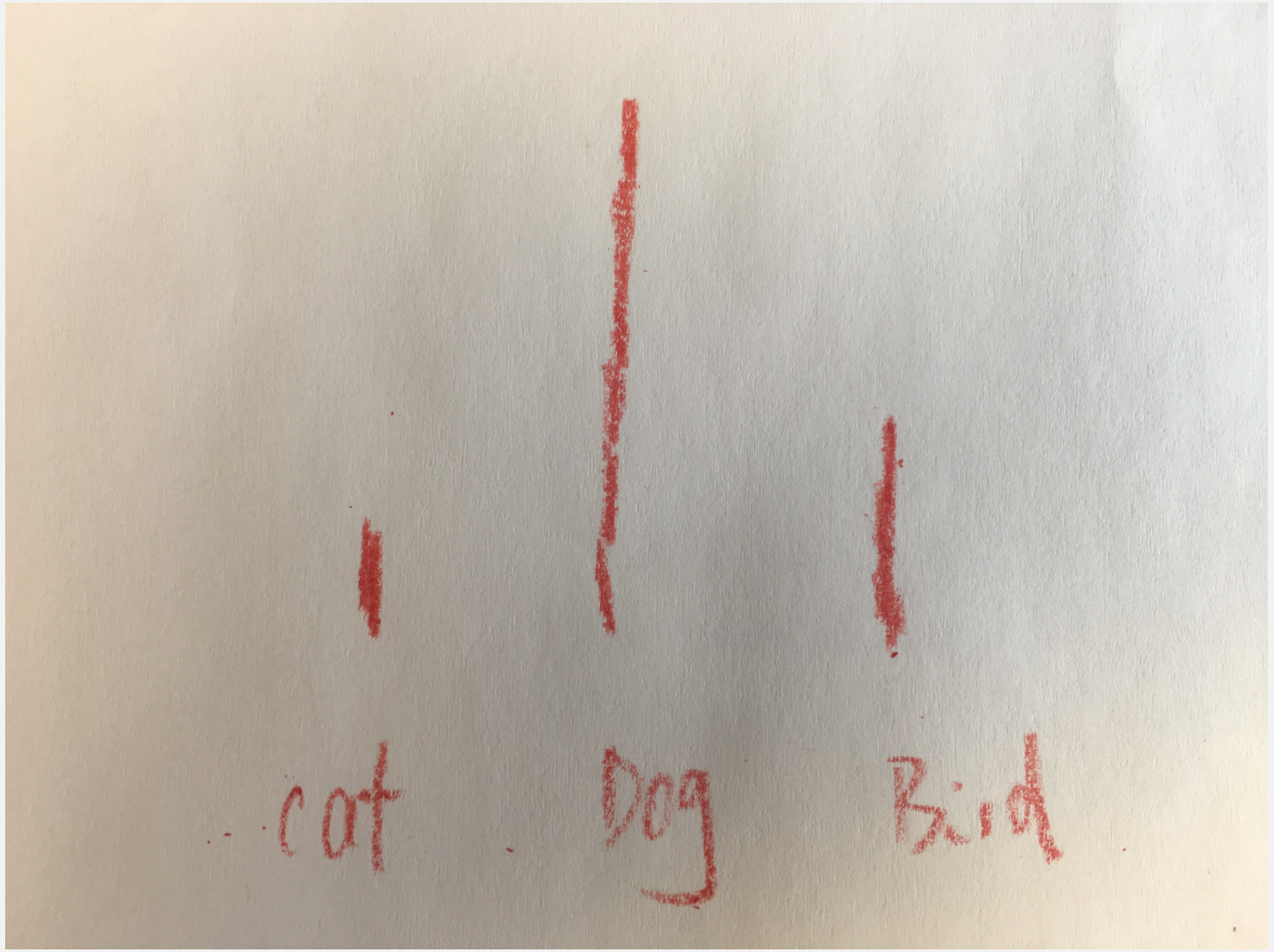
- **Shape**

- **Tilt**

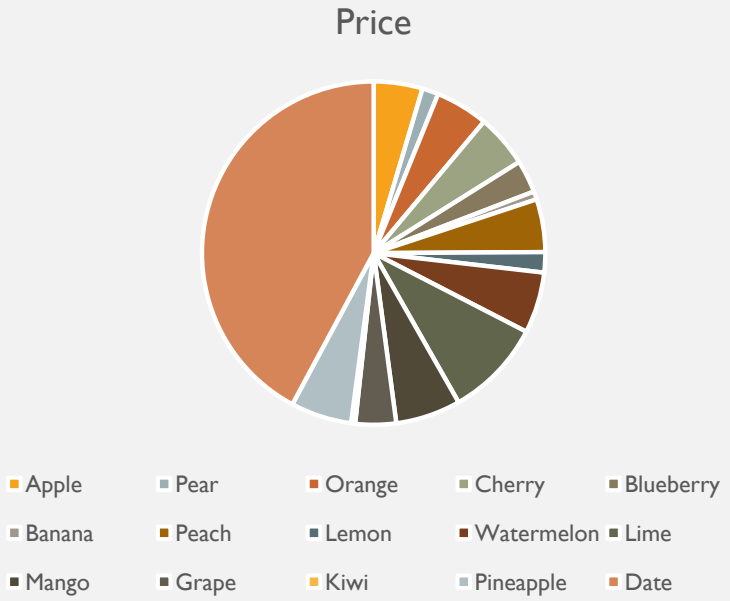
- **Size**
  - Length
  - Area
  - Volume

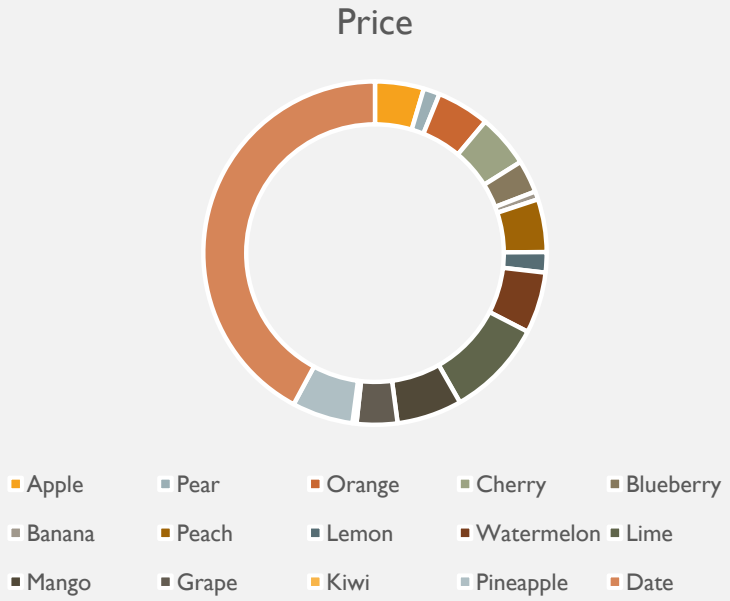





cat
Dog
Bird

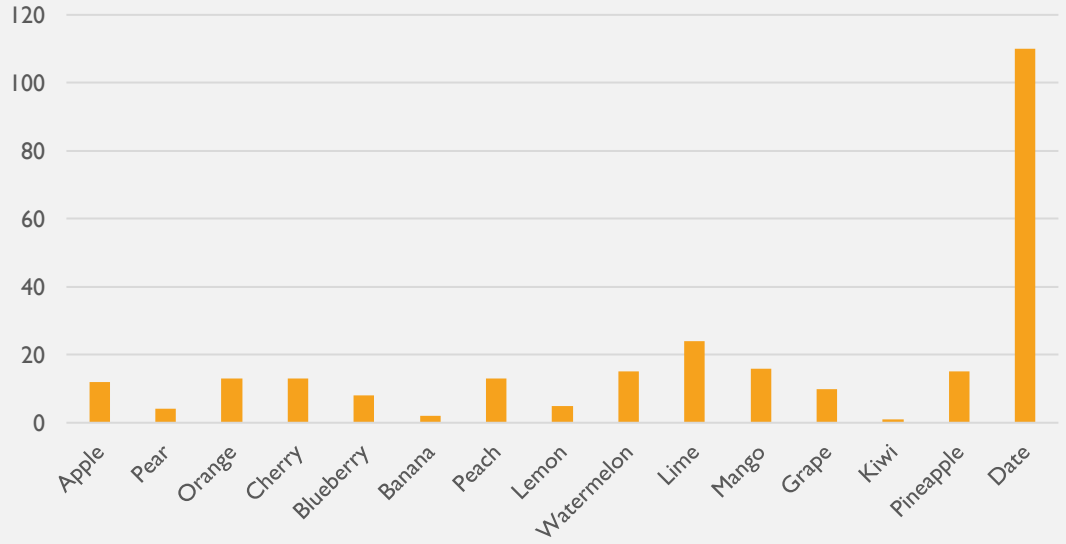
OTHER WAYS TO REPRESENT 2D DATA

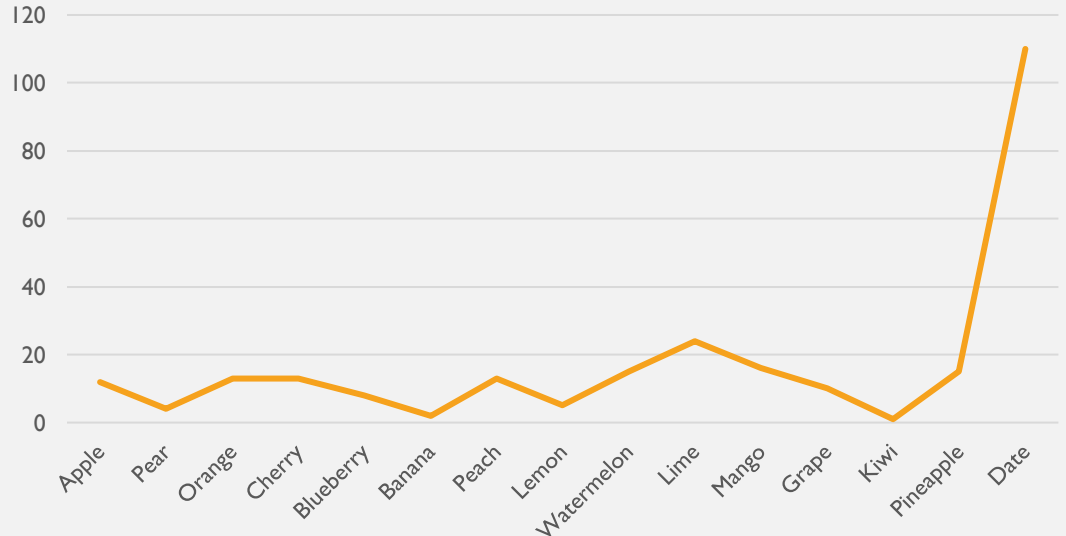




ACROSS CHARTS

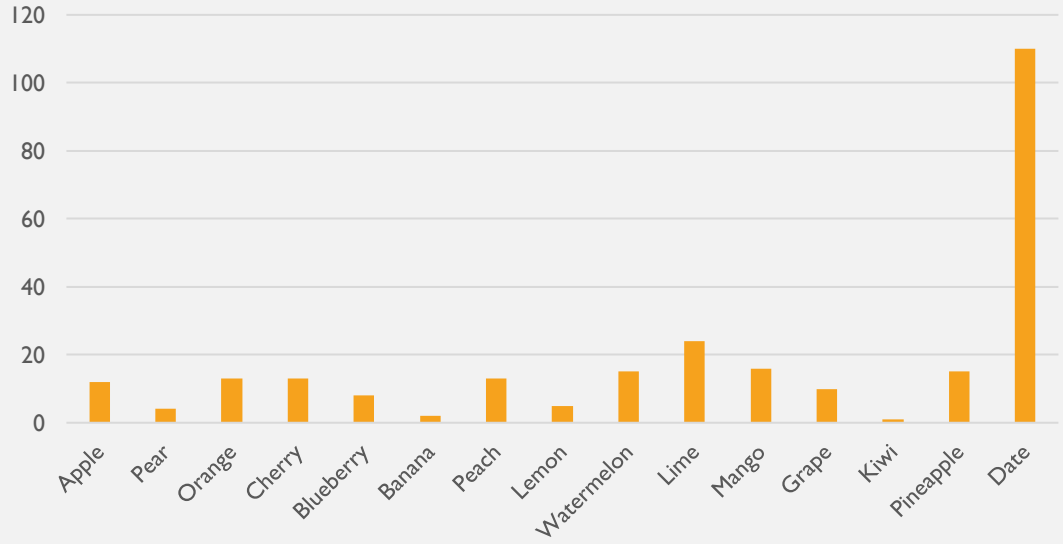
Name	Price
Apple	12
Pear	4
Orange	13
Cherry	13
Blueberry	8
Banana	2
Peach	13
Lemon	5
Watermelon	15
Lime	24
Mango	16
Grape	10
Kiwi	1
Pineapple	15
Date	110

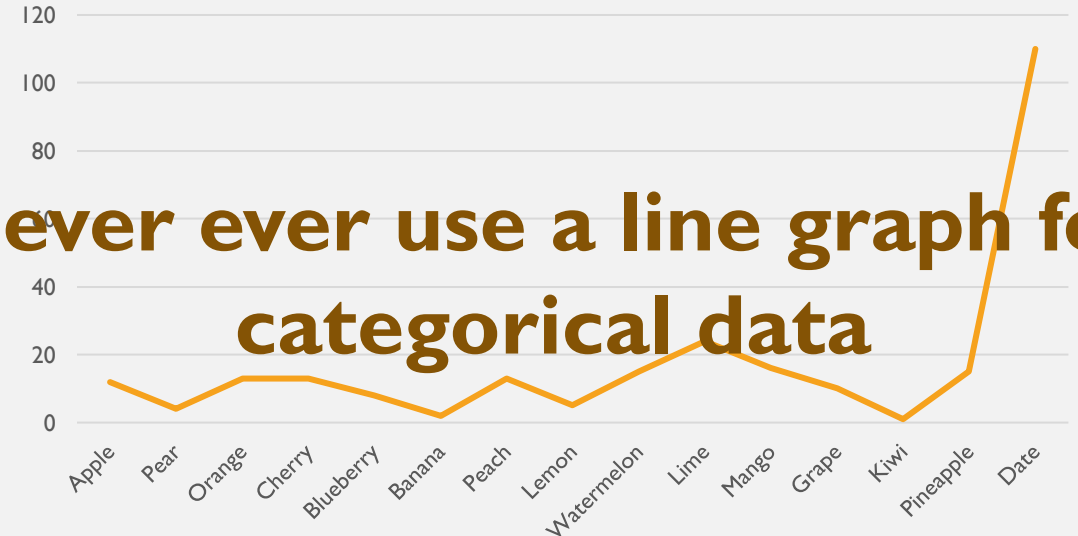




ACROSS CHARTS

Name	Price
Apple	12
Pear	4
Orange	13
Cherry	13
Blueberry	8
Banana	2
Peach	13
Lemon	5
Watermelon	15
Lime	24
Mango	16
Grape	10
Kiwi	1
Pineapple	15
Date	110





Never ever use a line graph for categorical data

TODAY..

Types of marks & channels
Ranking based on effectiveness
Tabular data

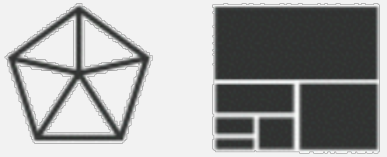
Types of marks & channels

MARK TYPES

marks as nodes (items)

points (0D)

lines (1D)



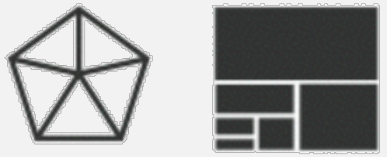
areas (2D)

MARK TYPES

marks as nodes (items)

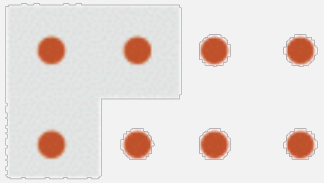
points (0D)

lines (1D)

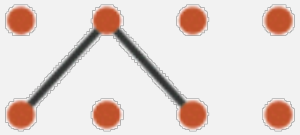


areas (2D)

marks as links



containment



connection

CHANNEL TYPES

identify (what and where)

- Position
  - Horizontal
  - Vertical
  - Both

- Shape

- Size
  - Length
  - Area
  - Volume

magnitude (how much)

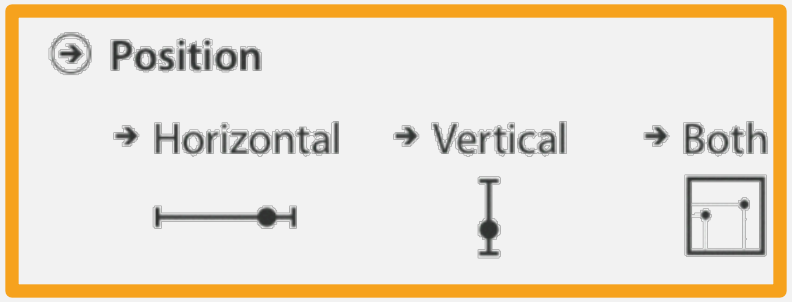
- Color

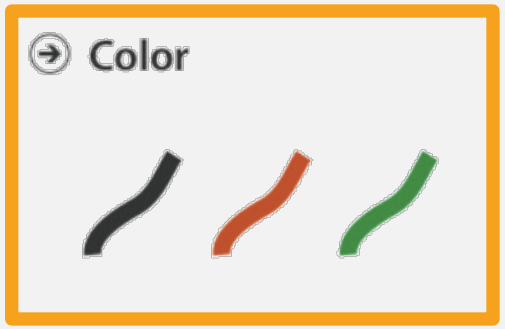
- Tilt

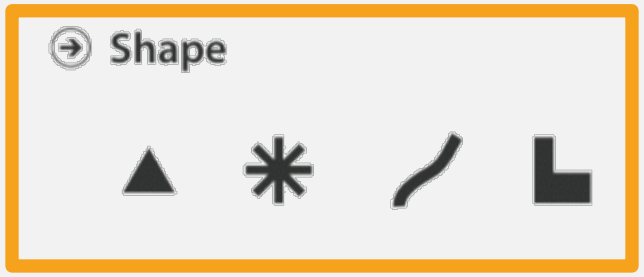
CHANNEL TYPES

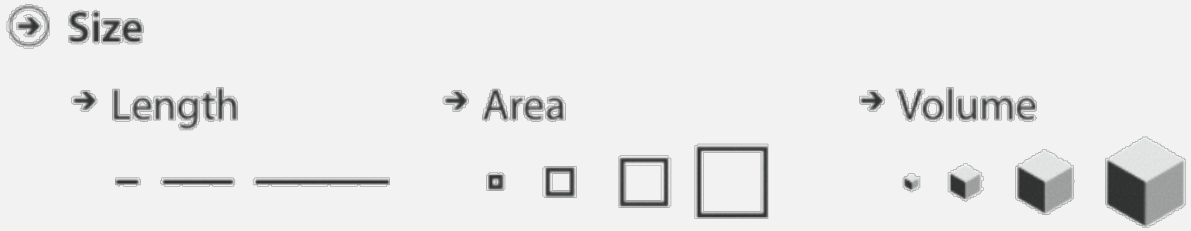
identify (what and where)

magnitude (how much)





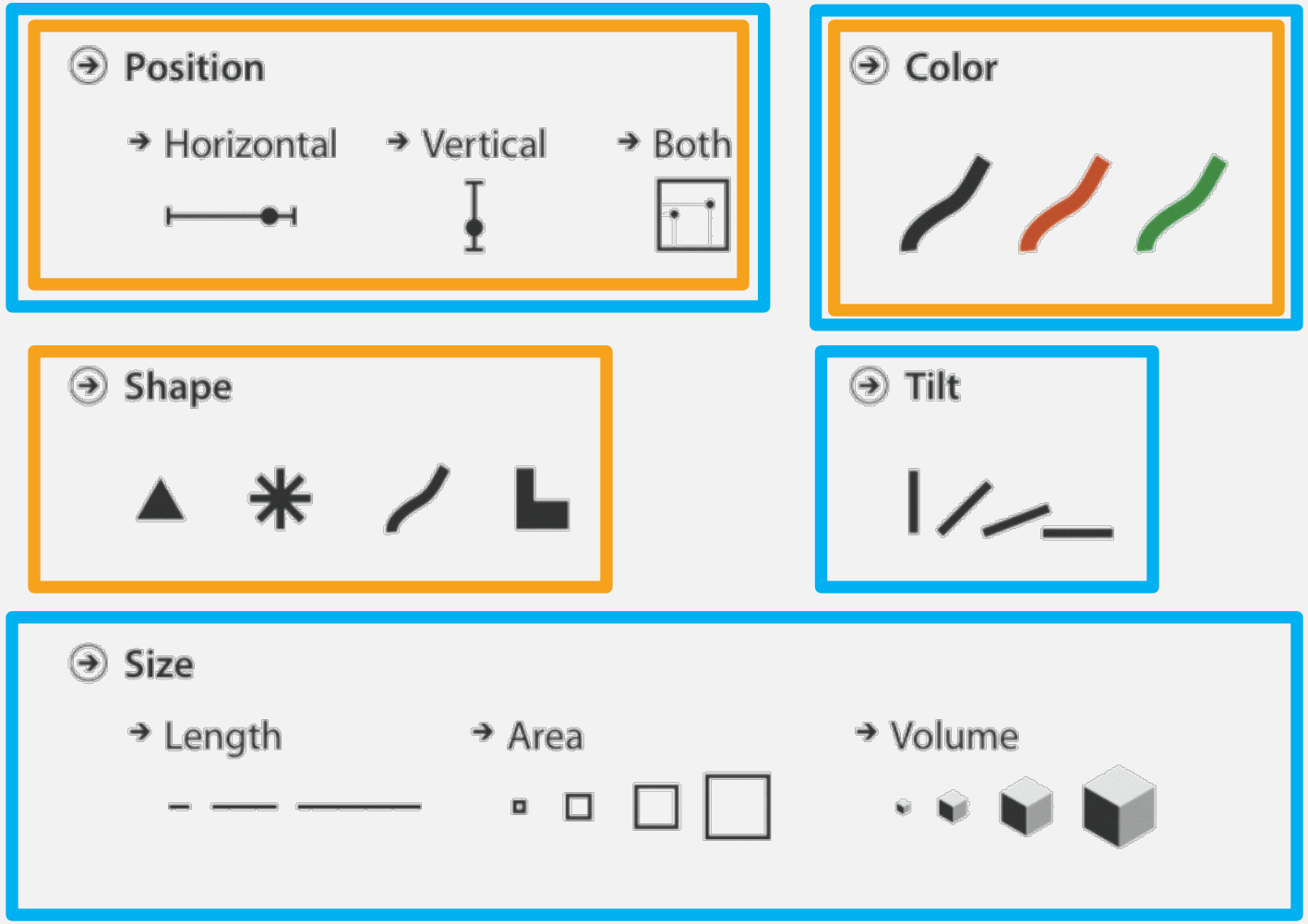




CHANNEL TYPES

identify (what and where)

magnitude (how much)



Expressiveness & effectiveness

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

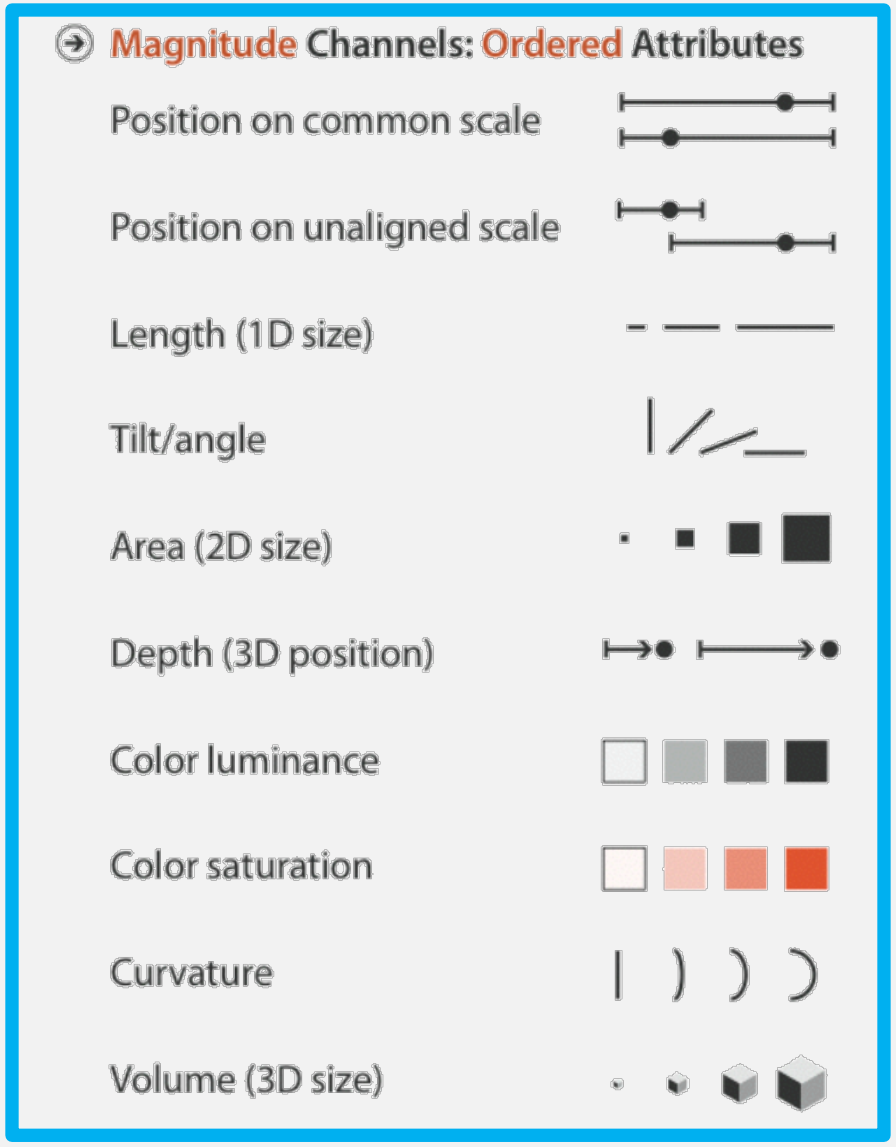
Identity Channels: Categorical Attributes

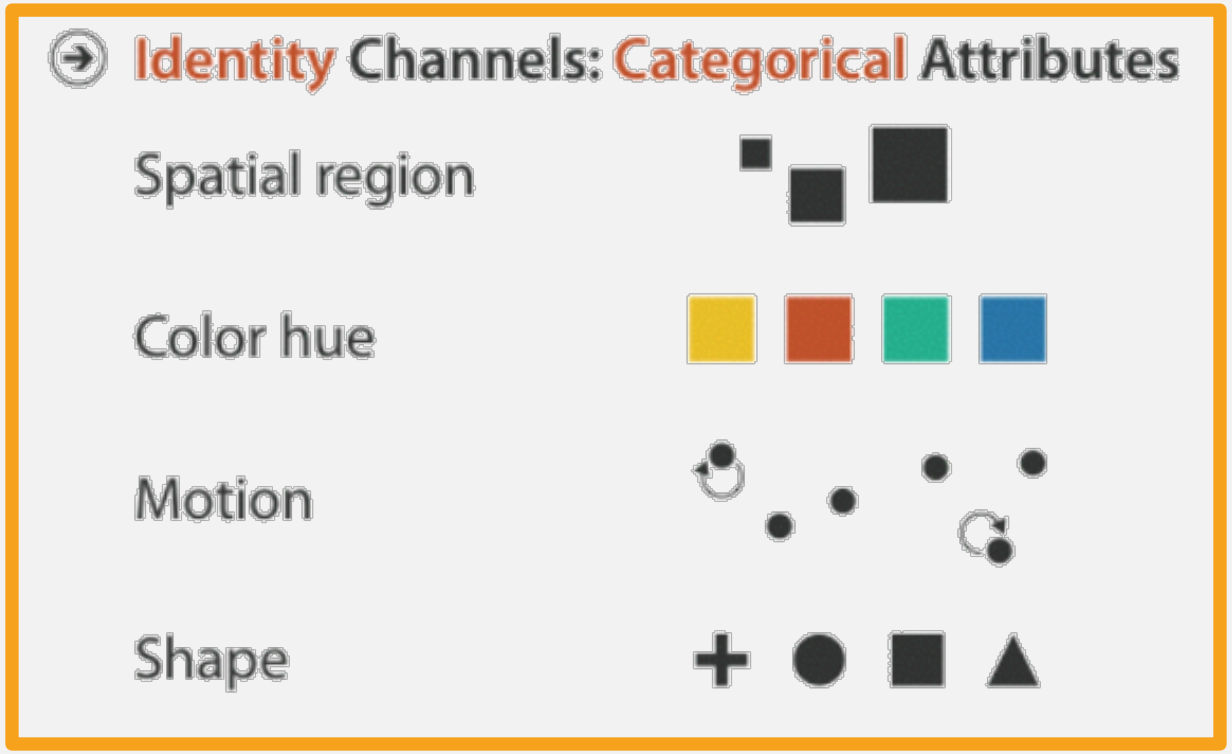
- Spatial region
- Color hue
- Motion
- Shape

expressiveness

magnitude (how much)

identify (what or where)

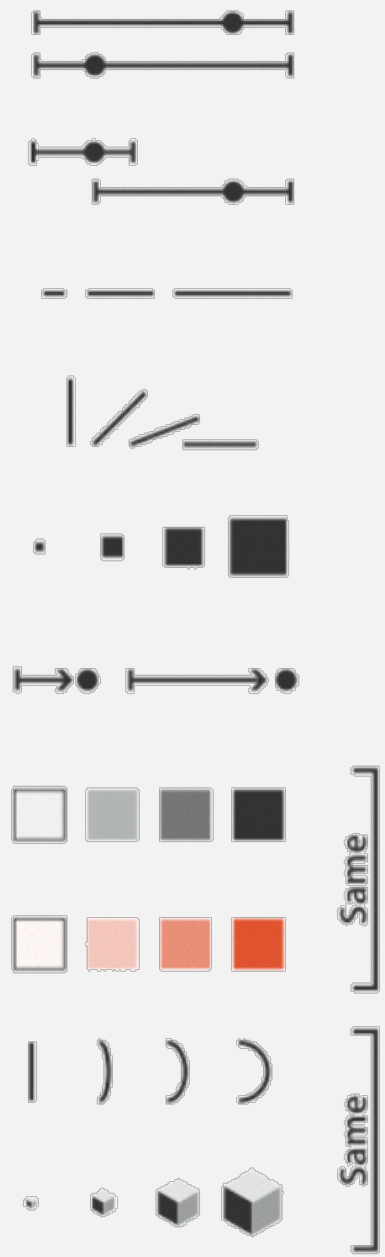




expressiveness

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)



Most

Effectiveness

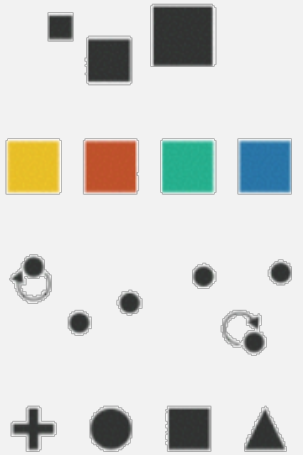
Least

Same

Same

Identity Channels: Categorical Attributes

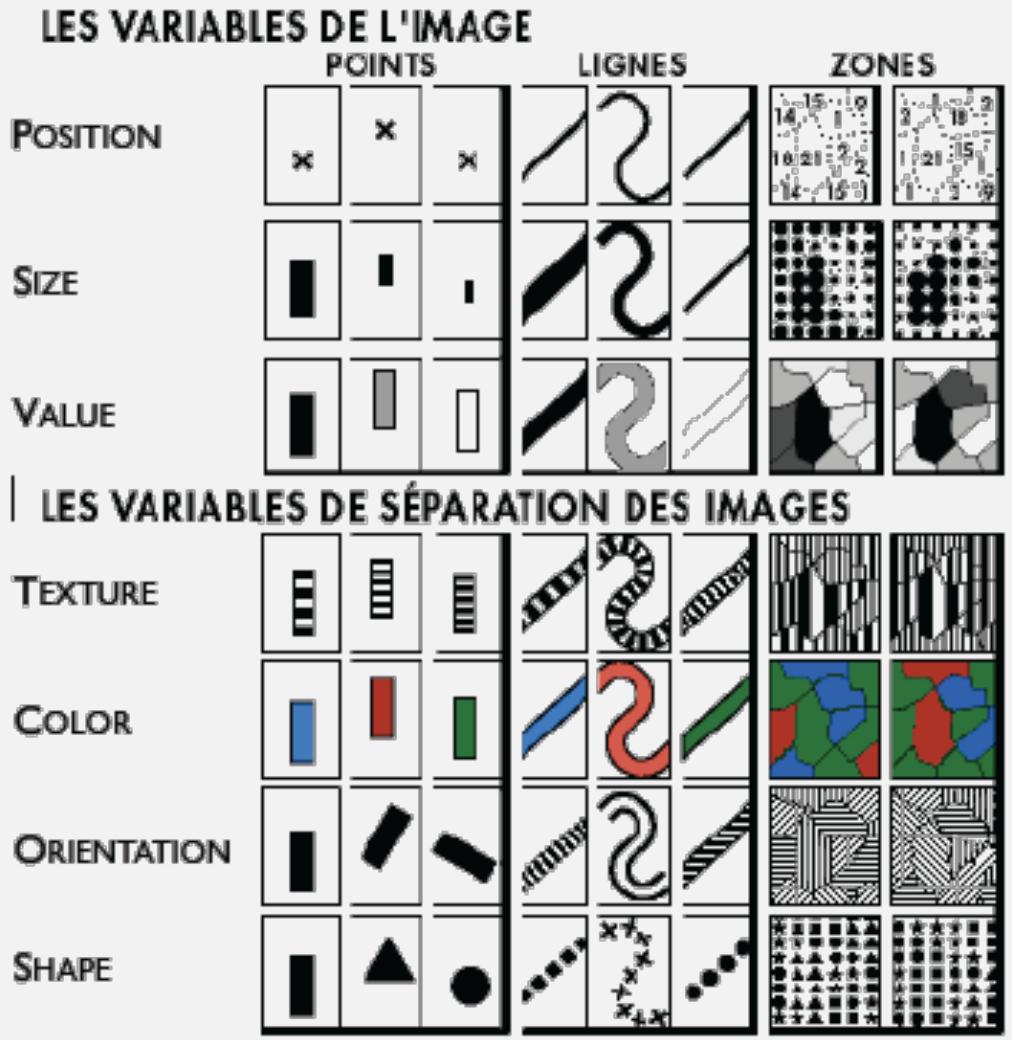
- Spatial region
- Color hue
- Motion
- Shape



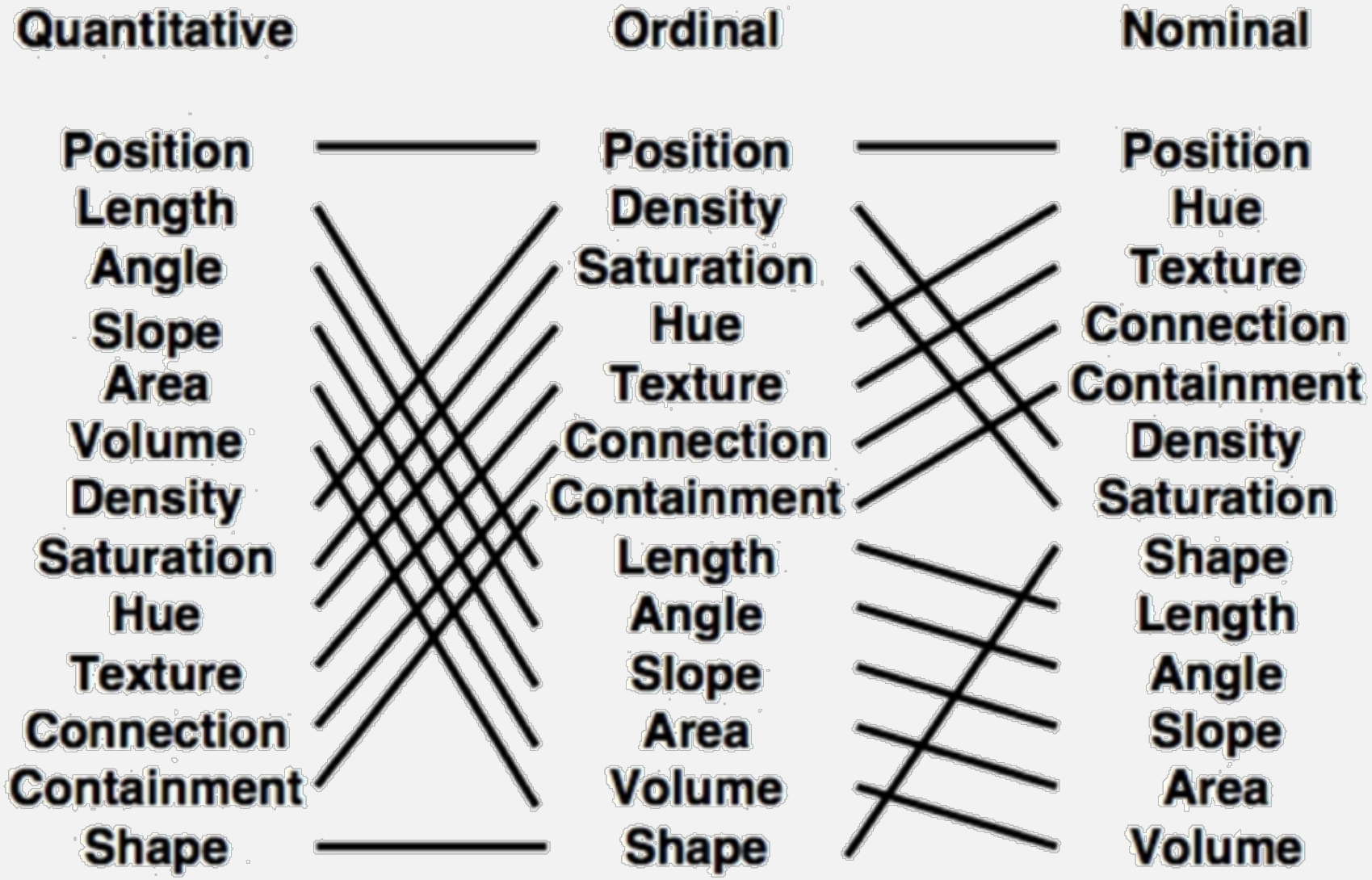
effectiveness

Where do the ranking come from?

Bertin, 1967

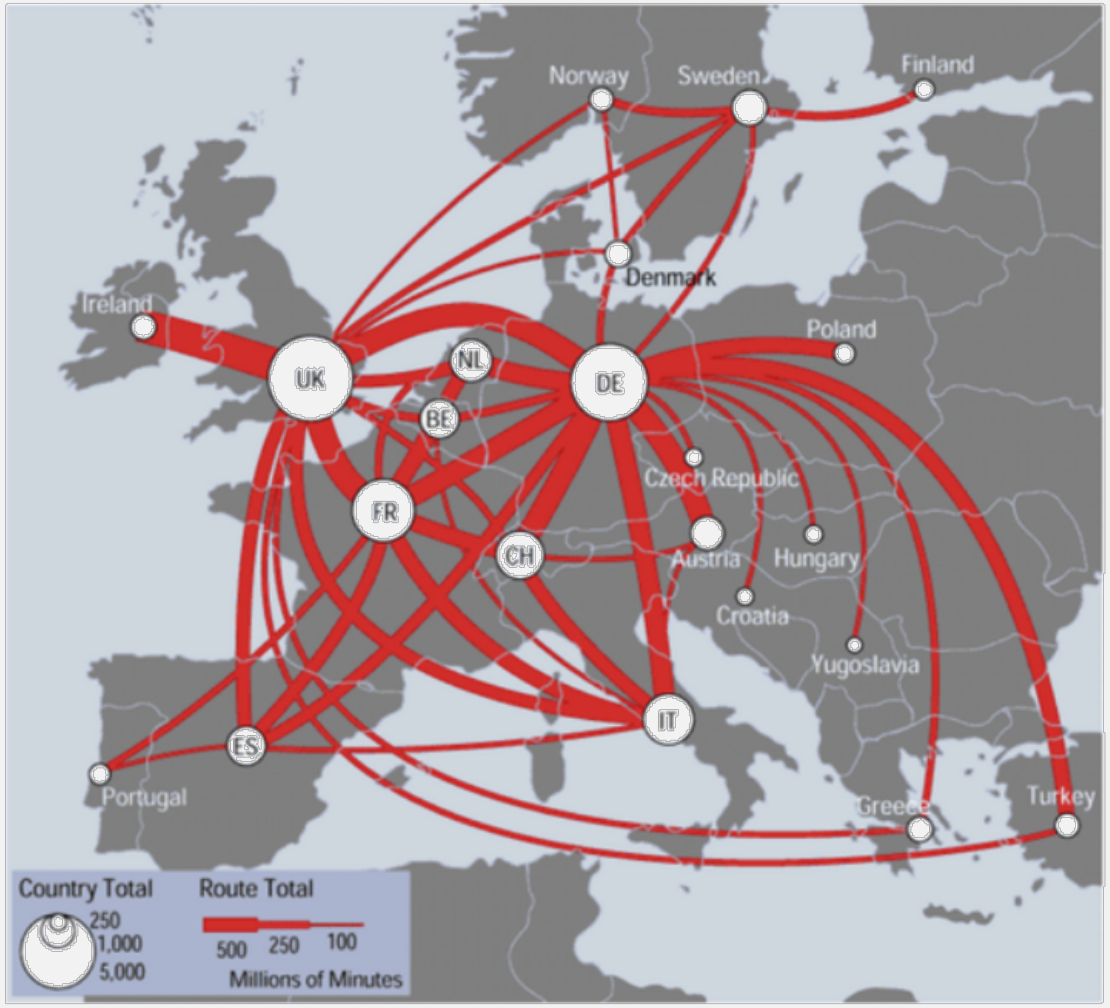

LES VARIABLES DE L'IMAGE
POINTS
LIGNES
ZONES
POSITION
SIZE
VALUE
LES VARIABLES DE SÉPARATION DES IMAGES
TEXTURE
COLOR
ORIENTATION
SHAPE

Mackinlay, 1986



DISCRIMINABILITY

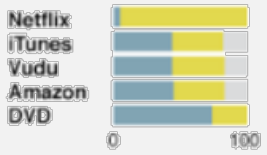
can channel differences be discerned?



Streaming the Box Office

Top 100 in 2011

AVAILABLE
NOT AVAILABLE
PURCHASE ONLY



Tristan Louis compiled a list of the top 100 movies at the box office, according to Box Office Mojo, that were available streaming. This is a graphical version of that list.

Source: Tristan Louis
By: Nathan Yau

1–50

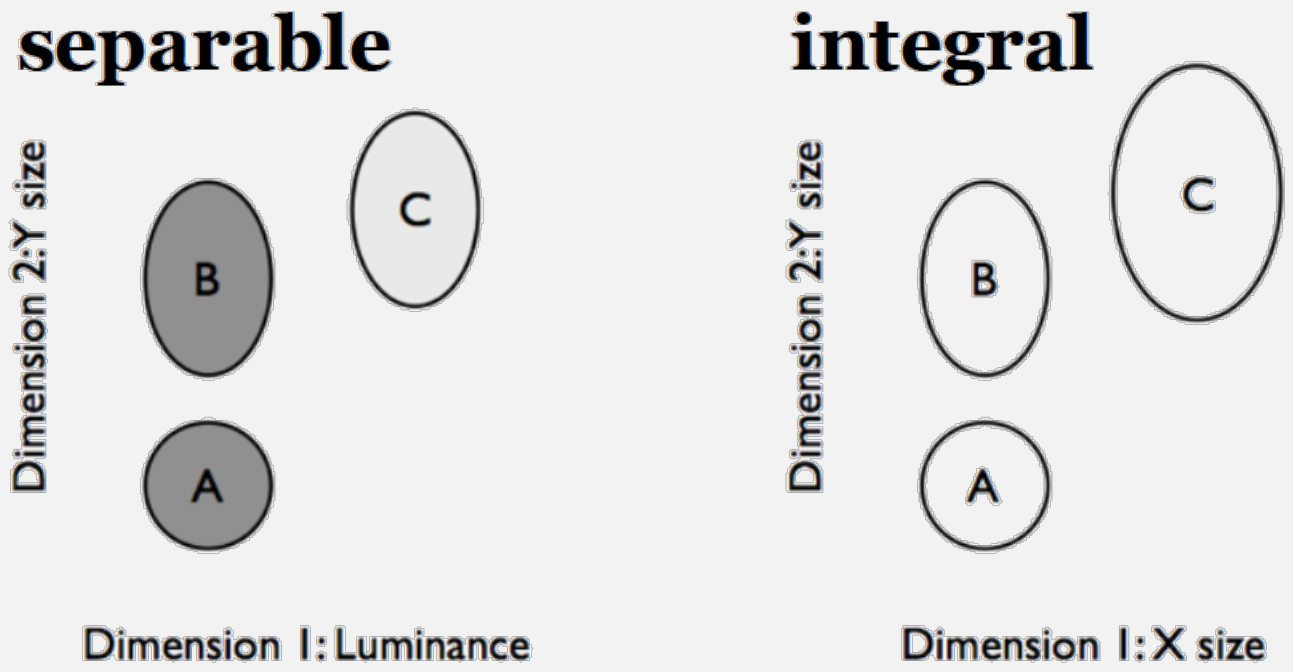
		Netflix	Amazon	iTunes	Vudu	DVD
1	Harry Potter and the Deathly Hallows Part 2	■	■	■	■	■
2	Transformers: Dark of the Moon	■	■	■	■	■
3	The Twilight Saga: Breaking Dawn Part 1	■	■	■	■	■
4	The Hangover Part II	■	■	■	■	■
5	Pirates of the Caribbean: On Stranger Tides	■	■	■	■	■
6	Fast Five	■	■	■	■	■
7	Cars 2	■	■	■	■	■
8	Thor	■	■	■	■	■
9	Rise of the Planet of the Apes	■	■	■	■	■
10	Captain America: The First Avenger	■	■	■	■	■
11	The Help	■	■	■	■	■
12	Bridesmaids	■	■	■	■	■
13	Kung Fu Panda 2	■	■	■	■	■
14	X-Men: First Class	■	■	■	■	■
15	Puss in Boots	■	■	■	■	■
16	Rio	■	■	■	■	■
17	The Smurfs	■	■	■	■	■
18	Mission: Impossible — Ghost Protocol	■	■	■	■	■
19	Sherlock Holmes: A Game of Shadows	■	■	■	■	■
20	Super 8	■	■	■	■	■
21	Rango	■	□	□	□	■
22	Horrible Bosses	■	■	■	■	■
23	Green Lantern	■	■	■	■	■
24	Hop	■	■	■	■	■
25	Paranormal Activity 3	■	■	■	■	■
26	Just Go With It	■	□	□	□	■
27	Bad Teacher	■	■	■	■	■
28	Cowboys & Aliens	■	■	■	■	■
29	Gnomeo and Juliet	■	□	□	□	■
30	The Green Hornet	■	□	□	□	■
31	Alvin and the Chipmunks: Chipwrecked	■	■	■	■	■
32	The Lion King (in 3D)	■	□	□	□	■
33	Real Steel	■	■	□	■	■
34	Crazy, Stupid, Love.	■	■	■	■	■
35	The Muppets	■	■	■	■	■
36	Battle: Los Angeles	■	□	□	□	■
37	Immortals	■	■	■	■	■
38	Zookeeper	■	■	■	■	■
39	Limitless	■	■	■	■	■
40	Tower Heist	■	■	■	■	■
41	Contagion	■	□	■	■	■
42	Moneyball	■	■	□	□	■
43	Justin Bieber: Never Say Never	■	□	□	□	■
44	Dolphin Tale	■	■	■	■	■
45	Jack and Jill	■	■	■	■	■
46	No Strings Attached	■	□	□	□	■
47	Mr. Popper's Penguins	■	■	■	■	■
48	Unknown	■	■	■	■	■
49	The Adjustment Bureau	■	■	■	■	■
50	Happy Feet Two	■	■	■	■	■

51–100

		Netflix	Amazon	iTunes	Vudu	DVD
51	The Girl with the Dragon Tattoo (2011)	■	■	■	■	■
52	Water for Elephants	■	■	■	■	■
53	The Lincoln Lawyer	■	□	□	□	■
54	Midnight in Paris	■	■	■	■	■
55	Friends with Benefits	■	■	■	■	■
56	I Am Number Four	■	□	□	□	■
57	Source Code	■	■	■	■	■
58	Insidious	□	□	□	□	■
59	Tyler Perry's Madea's Big Happy Family	■	■	■	■	■
60	Diary of a Wimpy Kid: Rodrick Rules	■	■	■	■	■
61	Footloose (2011)	■	■	■	■	■
62	The Adventures of Tintin	■	■	■	■	■
63	Hugo	■	■	■	■	■
64	The Dilemma	■	■	■	■	■
65	New Year's Eve	■	■	■	■	■
66	Arthur Christmas	■	■	■	■	■
67	War Horse	■	■	■	■	■
68	Hall Pass	■	■	■	■	■
69	We Bought a Zoo	■	■	■	■	■
70	Soul Surfer	■	□	□	□	■
71	Final Destination 5	■	■	■	■	■
72	The Ides of March	■	■	□	■	■
73	The Descendants	■	■	■	■	■
74	Hanna	■	■	■	■	■
75	Something Borrowed	■	■	■	■	■
76	Spy Kids: All the Time in the World	■	■	■	■	■
77	Scream 4	■	■	■	■	■
78	Big Mommas: Like Father, Like Son	■	■	■	■	■
79	Red Riding Hood	■	■	■	■	■
80	Paul	■	■	■	■	■
81	The Roommate	■	□	□	□	■
82	Jumping the Broom	■	□	□	□	■
83	The Change-Up	■	■	■	■	■
84	30 Minutes or Less	■	■	■	■	■
85	In Time	■	■	■	■	■
86	Colombiana	■	■	■	■	■
87	J. Edgar	■	■	■	■	■
88	Sucker Punch	■	■	■	■	■
89	Larry Crowne	■	■	■	■	■
90	50/50	■	■	■	■	■
91	Drive (2011)	■	■	■	■	■
92	A Very Harold & Kumar 3D Christmas	■	■	■	■	■
93	Courageous	■	□	■	■	■
94	The Rite	■	■	■	■	■
95	Arthur (2011)	■	■	■	■	■
96	The Debt	■	■	■	■	■
97	Priest	■	□	□	□	■
98	The Mechanic	■	■	■	■	■
99	Abduction	■	■	■	■	■
100	Beastly	■	■	■	□	■

SEPARABLE vs INTEGRAL

- **separable:** can judge each channel individually
- **integral:** two channels are viewed holistically



Ware 2004

SEPARABLE vs INTEGRAL

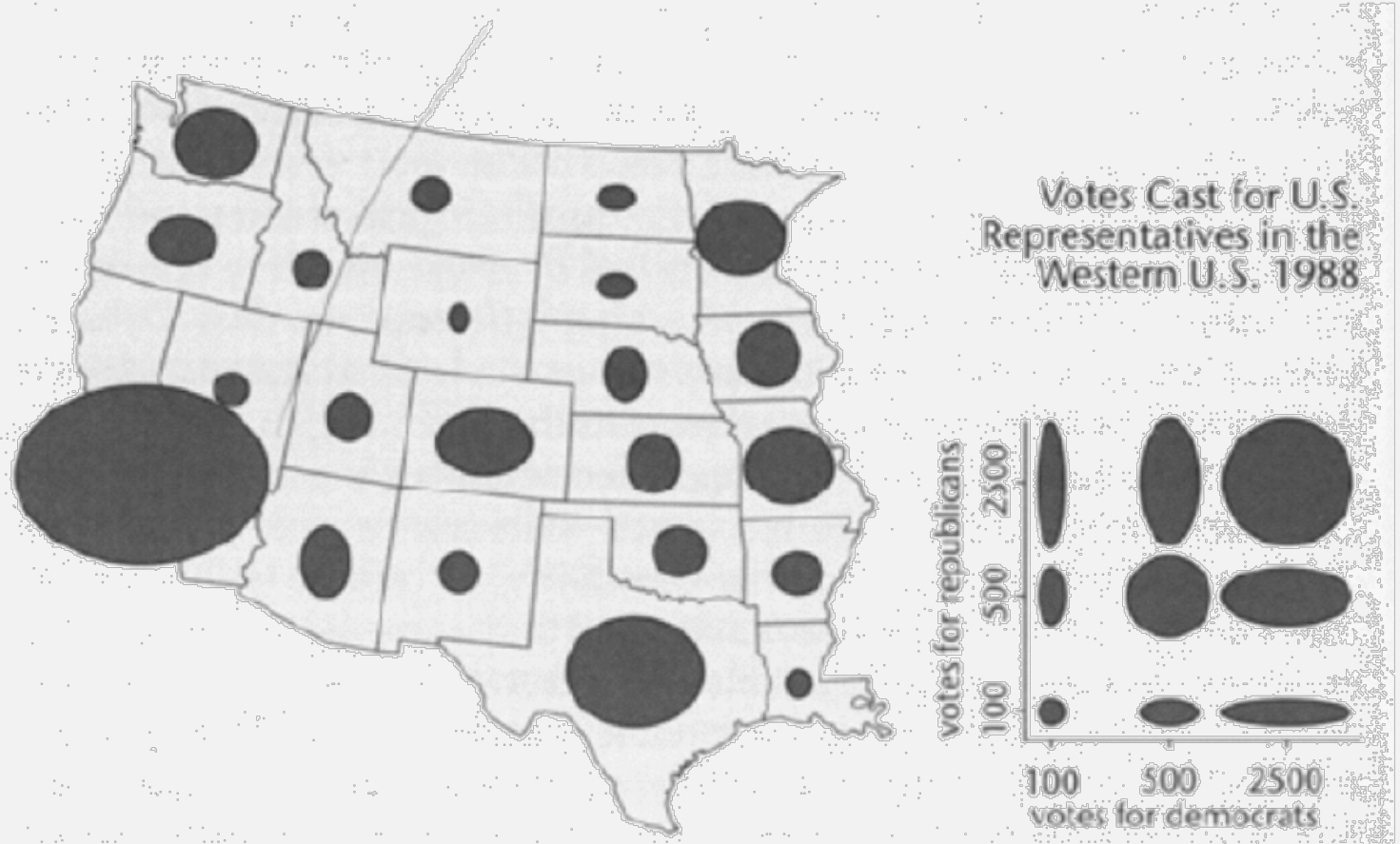
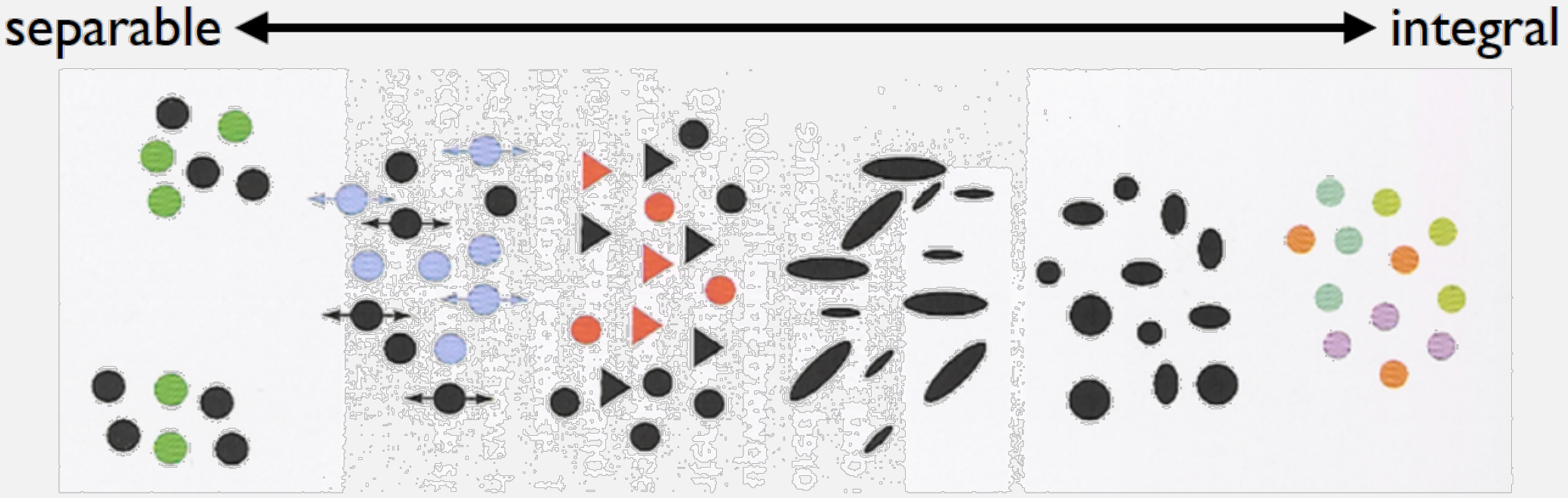


FIGURE 3.38. An example of the use of an ellipse as a map symbol in which the horizontal and vertical axes represent different (but presumably related) variables.

MacEachren 1995

SEPARABLE vs INTEGRAL



Ware 2004

encoding semantics

Graphical Code	Semantics
Small shapes defined by closed contour, texture, color, shaded solid.	Object, idea, entity, node.
Spatially ordered graphical objects.	Related information or a sequence. In a sequence the left-to-right ordering convention borrows from the western convention for written language.
Graphical objects in proximity.	Similar concepts, related information.
Graphical objects having the same shape, color, or texture.	Similar concepts, related information.
Size of graphical object Height of graphical object.	Magnitude, quantity, importance.
Shapes connected by contour.	
Thickness of connecting contour.	
Color and texture of connecting contour.	
Shapes enclosed by a contour, or a common texture, or a common color.	
Nested regions, partitioned regions.	
Attached shapes.	

Ware 2010

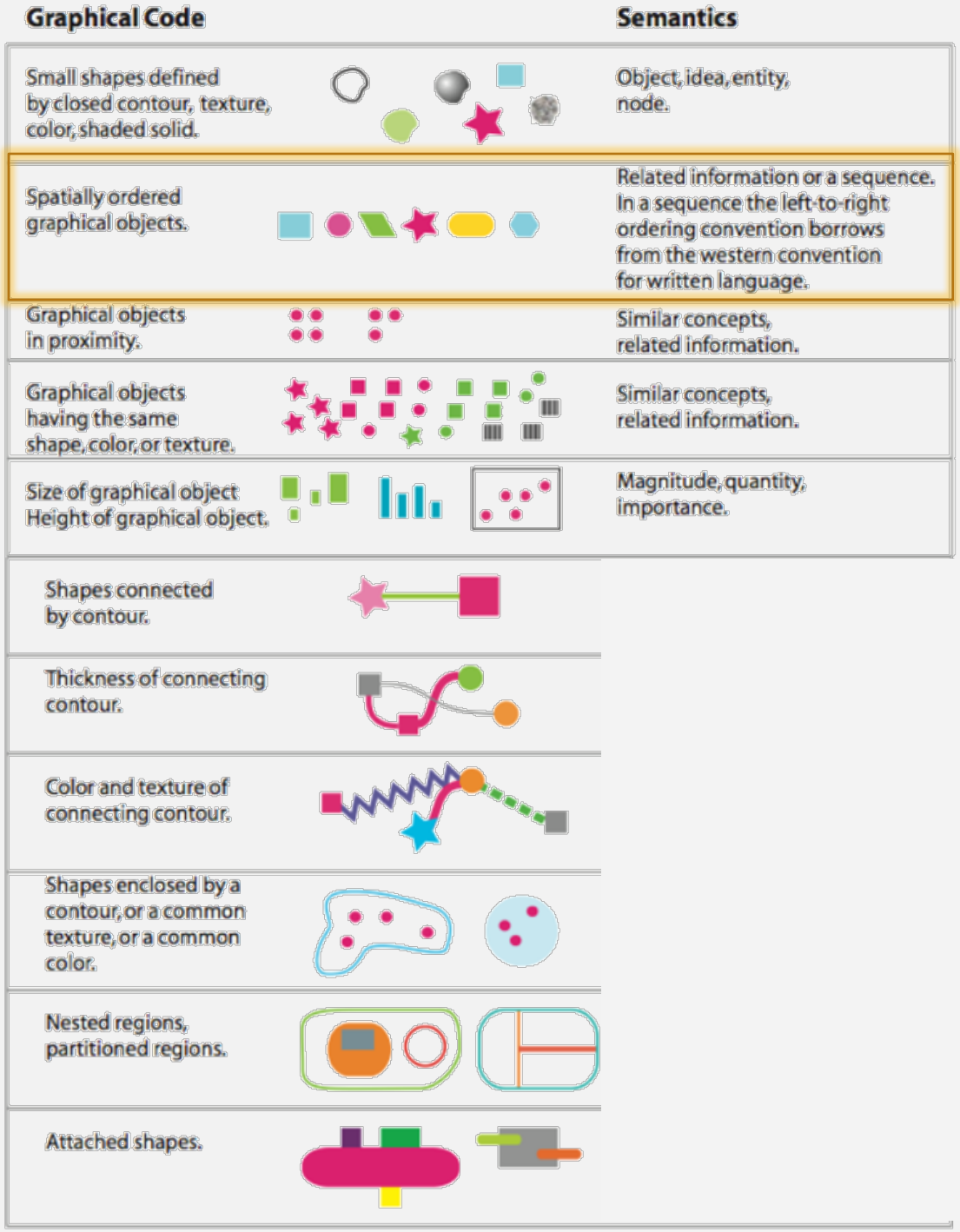
encoding semantics

Graphical Code	Semantics
Small shapes defined by closed contour, texture, color, shaded solid.	Object, idea, entity, node.
Spatially ordered graphical objects.	Related information or a sequence. In a sequence the left-to-right ordering convention borrows from the western convention for written language.
Graphical objects in proximity.	Similar concepts, related information.
Graphical objects having the same shape, color, or texture.	Similar concepts, related information.
Size of graphical object Height of graphical object.	Magnitude, quantity, importance.
Shapes connected by contour.	
Thickness of connecting contour.	
Color and texture of connecting contour.	
Shapes enclosed by a contour, or a common texture, or a common color.	
Nested regions, partitioned regions.	
Attached shapes.	

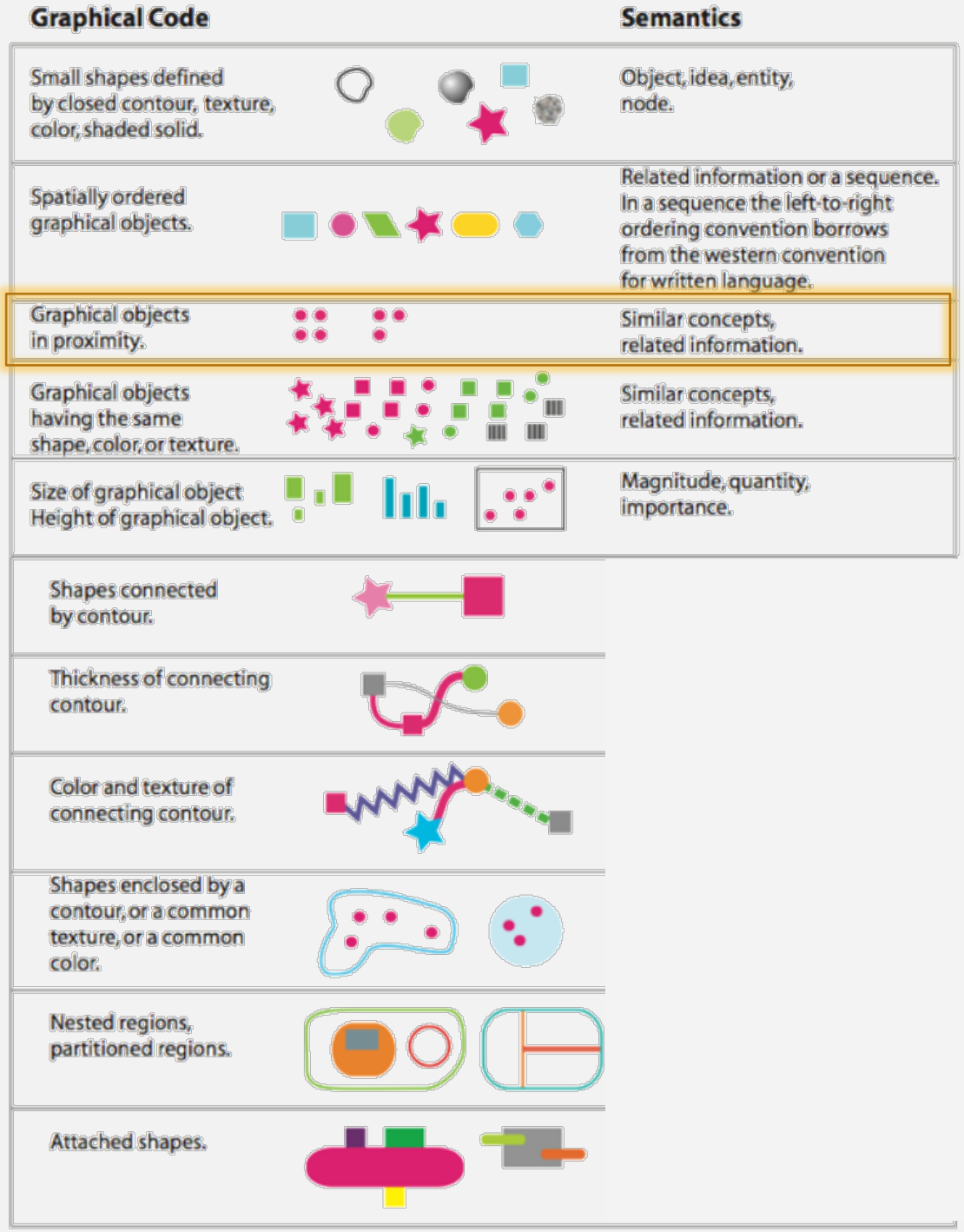
Ware 2010

encoding semantics

Ware 2010

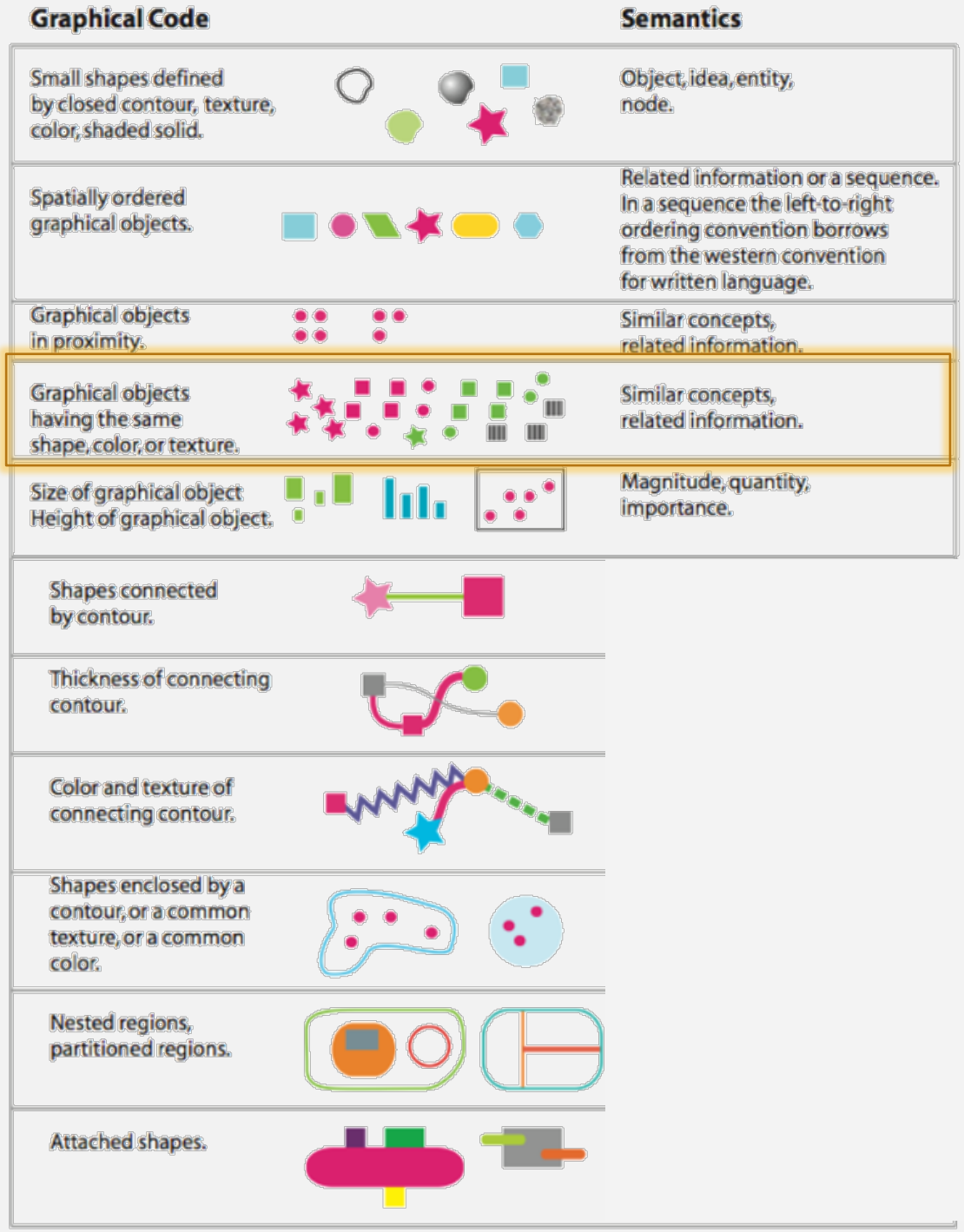


encoding semantics



Ware 2010

encoding semantics



Ware 2010

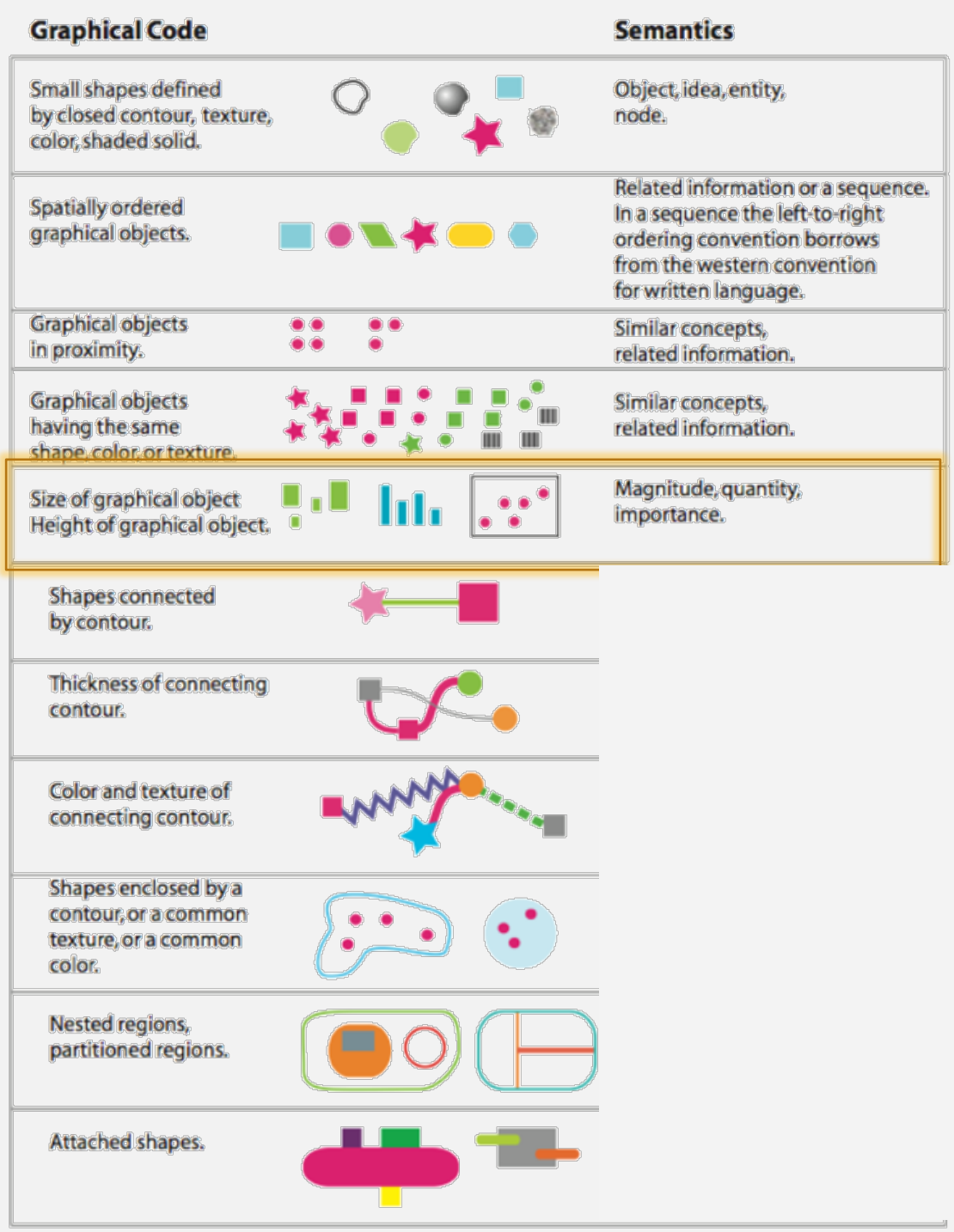
encoding semantics

Graphical Code	Semantics
Small shapes defined by closed contour, texture, color, shaded solid.	Object, idea, entity, node.
Spatially ordered graphical objects.	Related information or a sequence. In a sequence the left-to-right ordering convention borrows from the western convention for written language.
Graphical objects in proximity.	Similar concepts, related information.
Graphical objects having the same shape, color, or texture.	Similar concepts, related information.
Size of graphical object Height of graphical object.	Magnitude, quantity, importance.
Shapes connected by contour.	
Thickness of connecting contour.	
Color and texture of connecting contour.	
Shapes enclosed by a contour, or a common texture, or a common color.	
Nested regions, partitioned regions.	
Attached shapes.	

Ware 2010

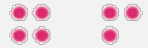
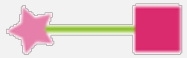
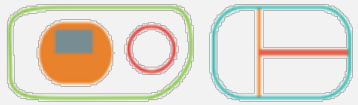
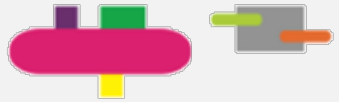
What do the remaining encodings mean to you?

Ware 2010



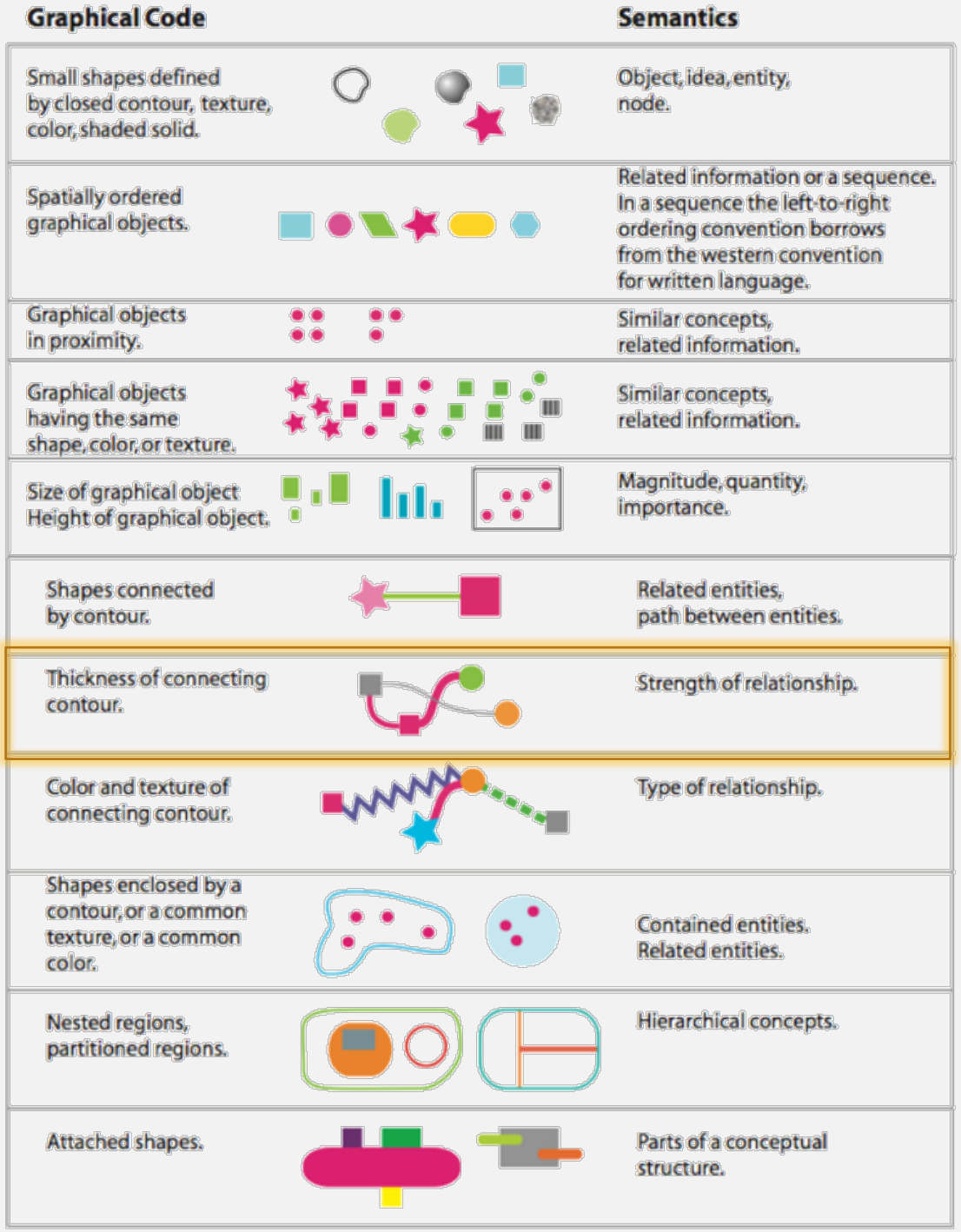
encoding semantics

Ware 2010

Graphical Code		Semantics
Small shapes defined by closed contour, texture, color, shaded solid.		Object, idea, entity, node.
Spatially ordered graphical objects.		Related information or a sequence. In a sequence the left-to-right ordering convention borrows from the western convention for written language.
Graphical objects in proximity.		Similar concepts, related information.
Graphical objects having the same shape, color, or texture.		Similar concepts, related information.
Size of graphical object Height of graphical object.		Magnitude, quantity, importance.
Shapes connected by contour.		Related entities, path between entities.
Thickness of connecting contour.		Strength of relationship.
Color and texture of connecting contour.		Type of relationship.
Shapes enclosed by a contour, or a common texture, or a common color.		Contained entities. Related entities.
Nested regions, partitioned regions.		Hierarchical concepts.
Attached shapes.		Parts of a conceptual structure.

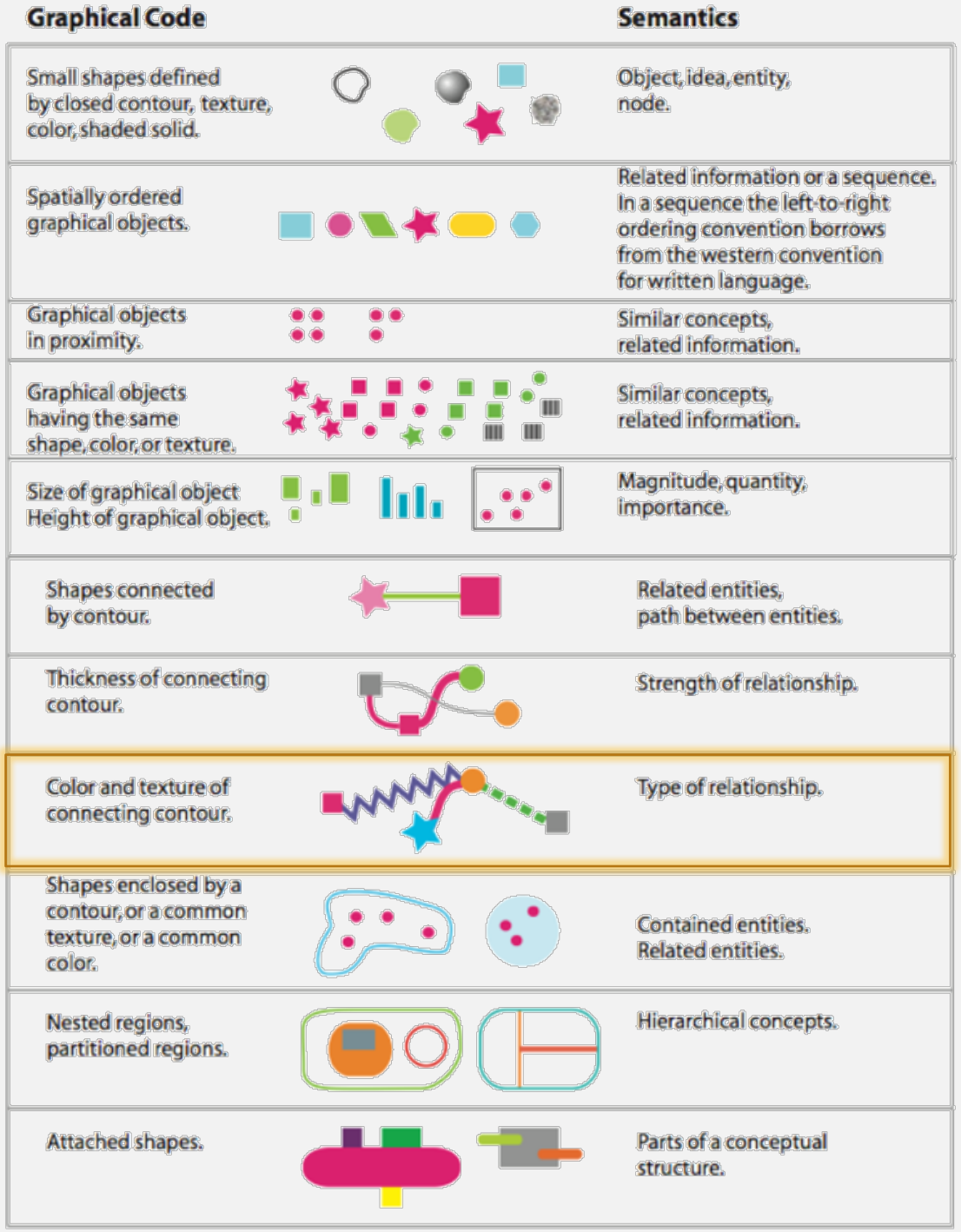
encoding semantics

Ware 2010



encoding semantics

Ware 2010



encoding semantics

Ware 2010

Graphical Code	Semantics
Small shapes defined by closed contour, texture, color, shaded solid.	Object, idea, entity, node.
Spatially ordered graphical objects.	Related information or a sequence. In a sequence the left-to-right ordering convention borrows from the western convention for written language.
Graphical objects in proximity.	Similar concepts, related information.
Graphical objects having the same shape, color, or texture.	Similar concepts, related information.
Size of graphical object Height of graphical object.	Magnitude, quantity, importance.
Shapes connected by contour.	Related entities, path between entities.
Thickness of connecting contour.	Strength of relationship.
Color and texture of connecting contour.	Type of relationship.
Shapes enclosed by a contour, or a common texture, or a common color.	Contained entities. Related entities.
Nested regions, partitioned regions.	Hierarchical concepts.
Attached shapes.	Parts of a conceptual structure.

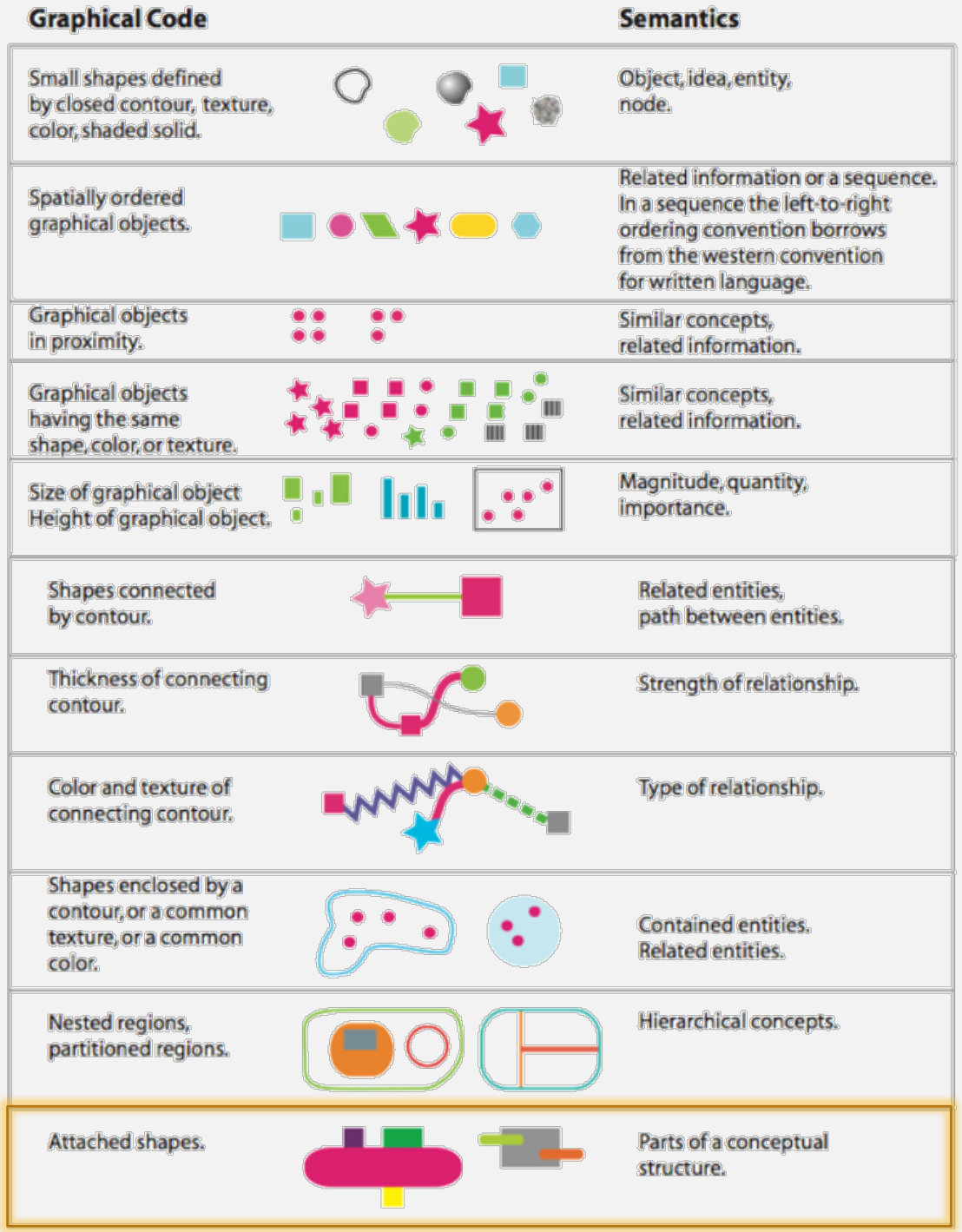
encoding semantics

Ware 2010

Graphical Code	Semantics
Small shapes defined by closed contour, texture, color, shaded solid.	Object, idea, entity, node.
Spatially ordered graphical objects.	Related information or a sequence. In a sequence the left-to-right ordering convention borrows from the western convention for written language.
Graphical objects in proximity.	Similar concepts, related information.
Graphical objects having the same shape, color, or texture.	Similar concepts, related information.
Size of graphical object Height of graphical object.	Magnitude, quantity, importance.
Shapes connected by contour.	Related entities, path between entities.
Thickness of connecting contour.	Strength of relationship.
Color and texture of connecting contour.	Type of relationship.
Shapes enclosed by a contour, or a common texture, or a common color.	Contained entities. Related entities.
Nested regions, partitioned regions.	Hierarchical concepts.
Attached shapes.	Parts of a conceptual structure.

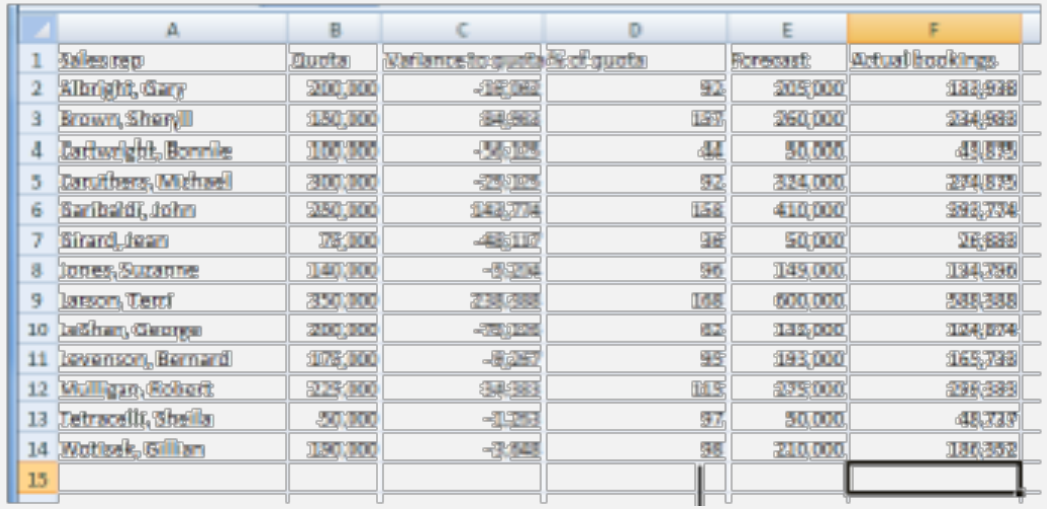
encoding semantics

Ware 2010



Tabular data

WHAT ABOUT A TABLE WITH MORE DIMENSIONS?



	A	B	C	D	E	F
1	Sales rep	Quota	Variance to quota	% of quota	Forecast	Actual bookings
2	Albright, Gary	200,000	-16,062	92	205,000	183,938
3	Brown, Sharyll	150,000	84,483	157	260,000	234,483
4	Cartwright, Bonnie	100,000	-56,125	44	50,000	43,875
5	Castiberg, Michael	300,000	-25,125	92	324,000	274,875
6	Garibaldi, John	250,000	143,774	158	410,000	393,774
7	Girard, Jean	75,000	-48,117	36	50,000	26,883
8	Jones, Suzanne	140,000	-5,264	96	145,000	134,736
9	Larson, Terri	350,000	238,388	168	600,000	588,388
10	Lothan, George	200,000	-75,926	62	132,000	124,074
11	Lovenson, Bernard	175,000	-9,257	95	193,000	165,743
12	Mulligan, Robert	325,000	58,583	118	375,000	383,583
13	Petrocelli, Sheila	50,000	-1,251	97	50,000	48,749
14	Wottask, Gillian	190,000	-3,648	98	210,000	186,352
15						

TABLE LENS

- Rao, Card (1994)

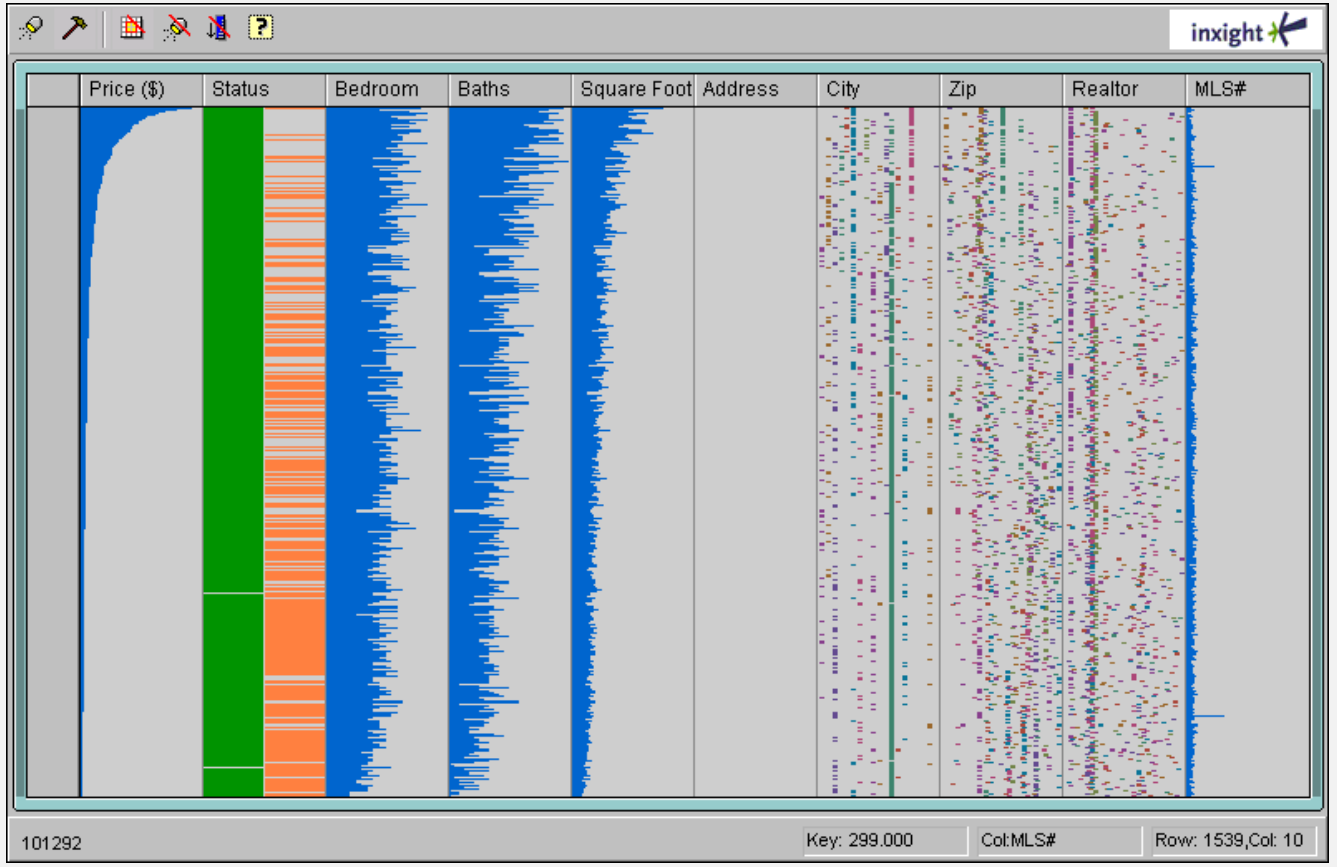
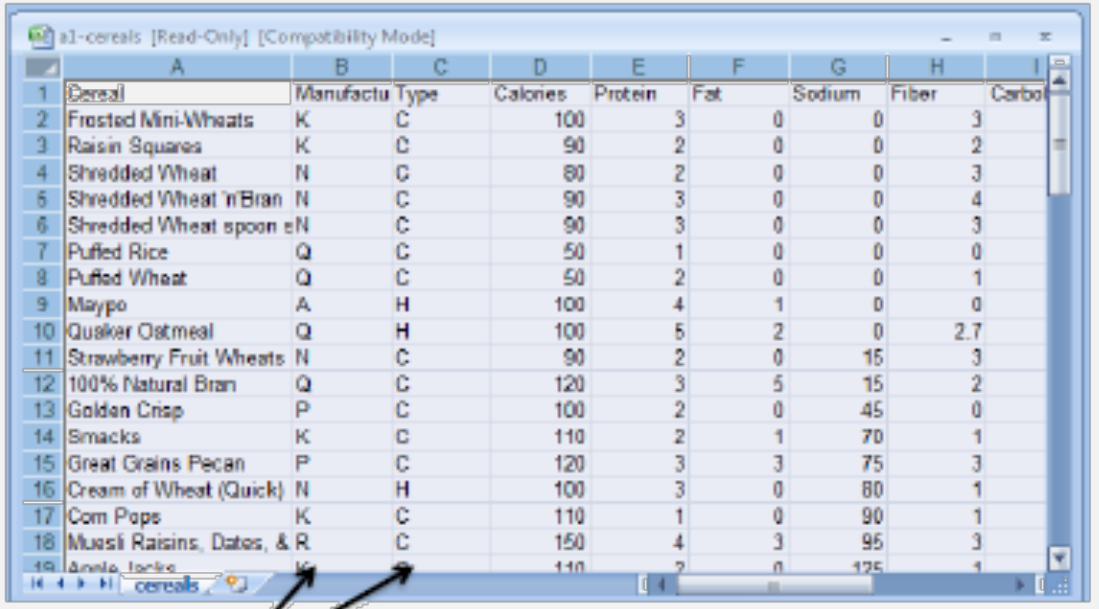


TABLE LENS

	A	B	C	D	E	F
1	Sales rep	Quota	Variance to quota	% of quota	Forecast	Actual bookings
2	Albright, Gary	200,000	-16,062	92	205,000	183,938
3	Brown, Sheryll	150,000	84,983	157	260,000	234,983
4	Cartwright, Bonnie	100,000	-56,125	44	50,000	43,875
5	Carruthers, Michael	300,000	-25,125	92	324,000	274,875
6	Garibaldi, John	150,000	143,774	158	410,000	293,774
7	Girard, Jean	75,000	-48,117	36	50,000	26,883
8	Jones, Suzanne	140,000	-5,204	96	149,000	134,796
9	Larson, Terri	350,000	238,388	168	600,000	588,388
10	Lothan, George	200,000	-75,126	62	132,000	124,874
11	Severson, Bernard	175,000	-9,267	95	193,000	165,733
12	Mulligan, Robert	225,000	84,383	115	375,000	309,383
13	Petrocelli, Sheila	50,000	-1,262	97	50,000	48,737
14	Plotsek, Gillian	190,000	-3,648	98	210,000	186,352
15						

Change quantitative values to bars

TABLE LENS



	A	B	C	D	E	F	G	H	I
1	Cereal	Manufactu	Type	Calories	Protein	Fat	Sodium	Fiber	Carbol
2	Frosted Mini-Wheats	K	C	100	3	0	0	3	
3	Raisin Squares	K	C	90	2	0	0	2	
4	Shredded Wheat	N	C	80	2	0	0	3	
5	Shredded Wheat 'n'Bran	N	C	90	3	0	0	4	
6	Shredded Wheat spoon s	N	C	90	3	0	0	3	
7	Puffed Rice	Q	C	50	1	0	0	0	
8	Puffed Wheat	Q	C	50	2	0	0	1	
9	Maypo	A	H	100	4	1	0	0	
10	Quaker Oatmeal	Q	H	100	5	2	0	2.7	
11	Strawberry Fruit Wheats	N	C	90	2	0	15	3	
12	100% Natural Bran	Q	C	120	3	5	15	2	
13	Golden Crisp	P	C	100	2	0	45	0	
14	Smacks	K	C	110	2	1	70	1	
15	Great Grains Pecan	P	C	120	3	3	75	3	
16	Cream of Wheat (Quick)	N	H	100	3	0	80	1	
17	Corn Pops	K	C	110	1	0	90	1	
18	Muesli Raisins, Dates, &	R	C	150	4	3	95	3	
19	Apple Jacks	K	C	110	2	0	125	1	

What do you do for nominal data?

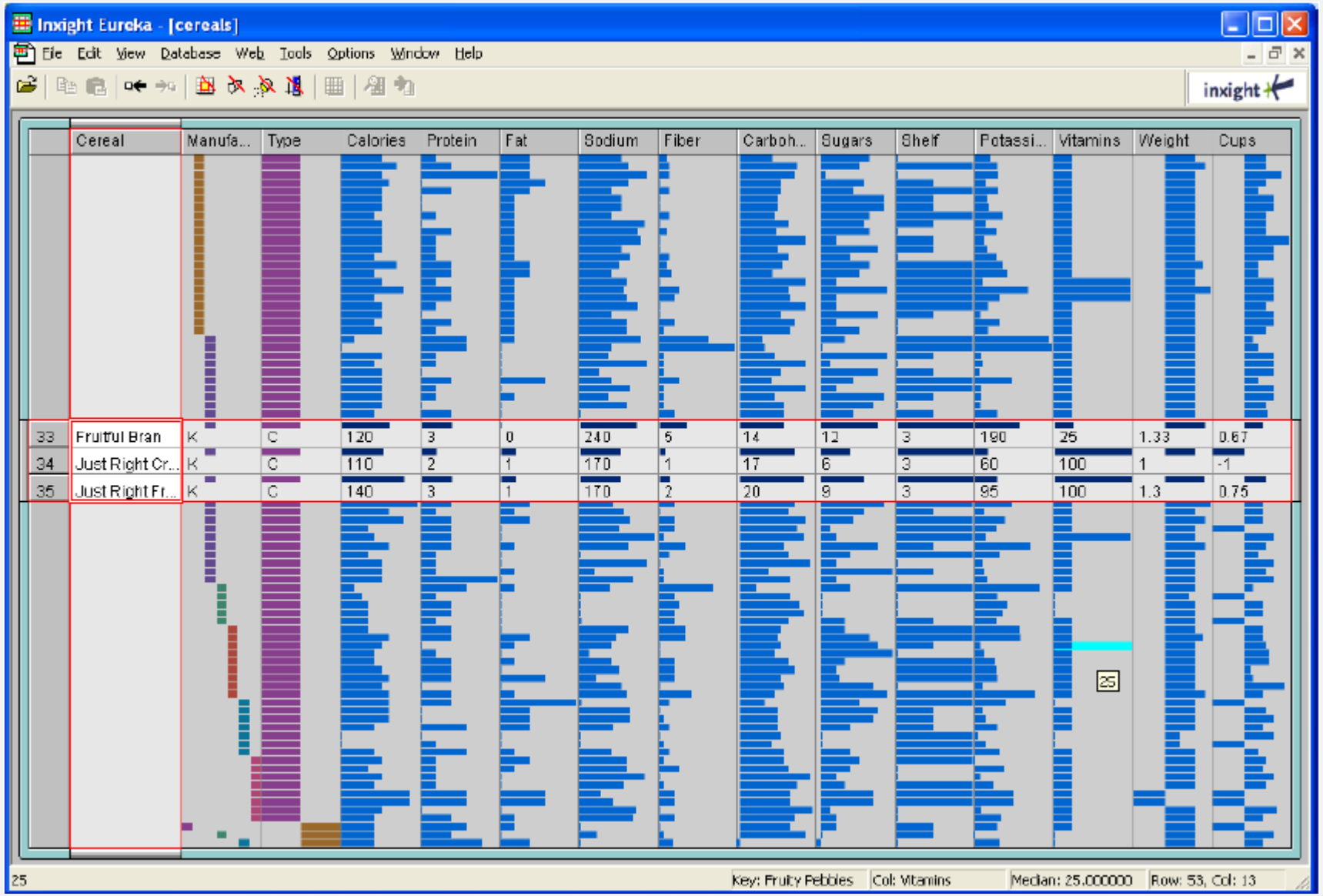
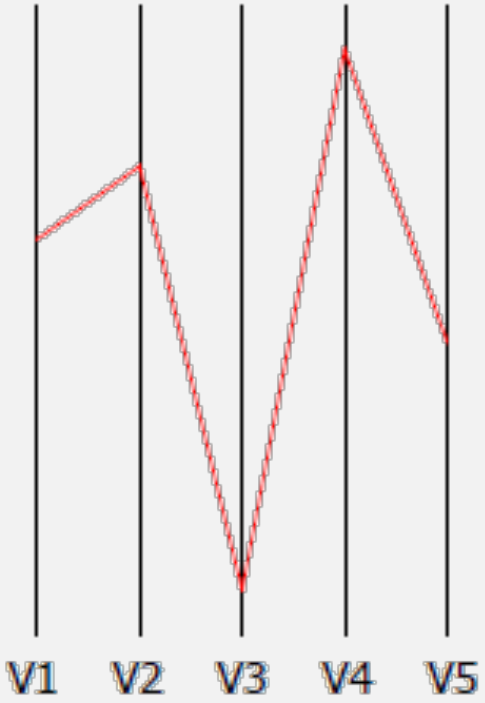

Inxight Eureka - [cereals]
File Edit View Database Web Tools Options Window Help
inxight
Cereal Manufa... Type Calories Protein Fat Sodium Fiber Carboh... Sugars Shelf Potassi... Vitamins Weight Cups
33 Fruitful Bran K C 120 3 0 240 6 14 12 3 190 25 1.33 0.67
34 Just Right Cr... K C 110 2 1 170 1 17 6 3 60 100 1 -1
35 Just Right Fr... K C 140 3 1 170 2 20 9 3 95 100 1.3 0.75
25
25
Key: Fruity Pebbles
Col: Vitamins
Median: 25.000000
Row: 53, Col: 13

TABLE LENS

- Video: http://www.open-video.org/details.php?videoid=8304

PARALLEL COORDINATES

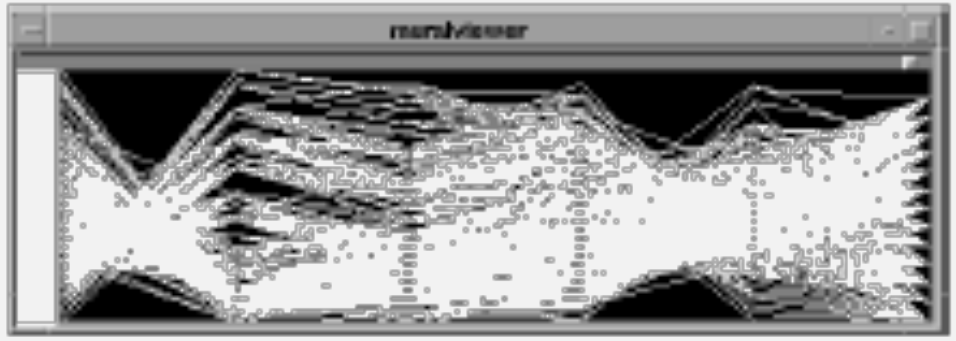


Encode variables along a horizontal row

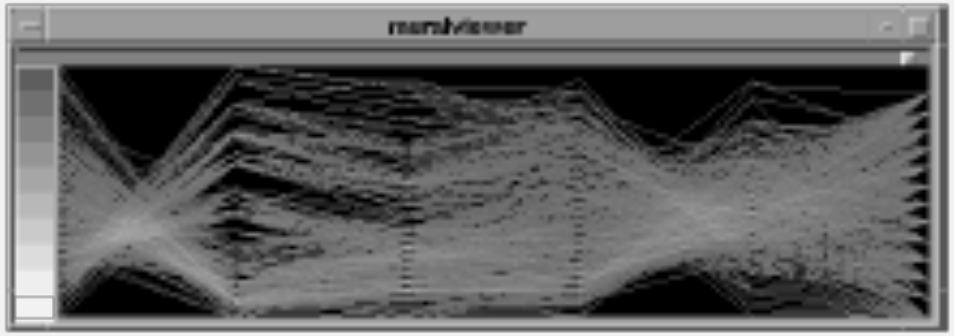
Vertical line specifies different values that variable can take

Data point represented as a polyline

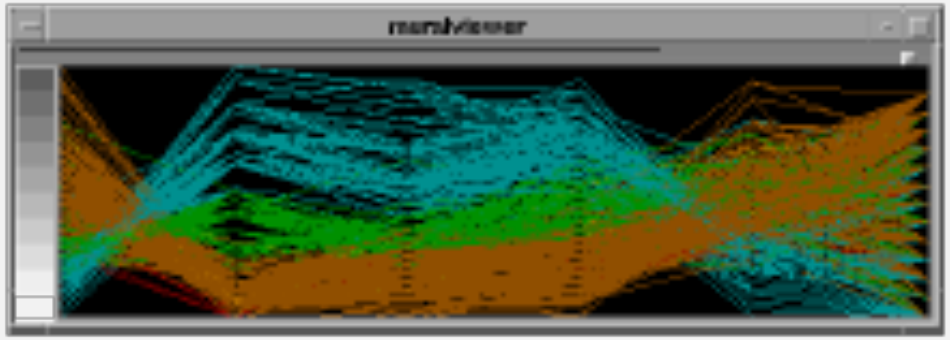
PARALLEL COORDINATES



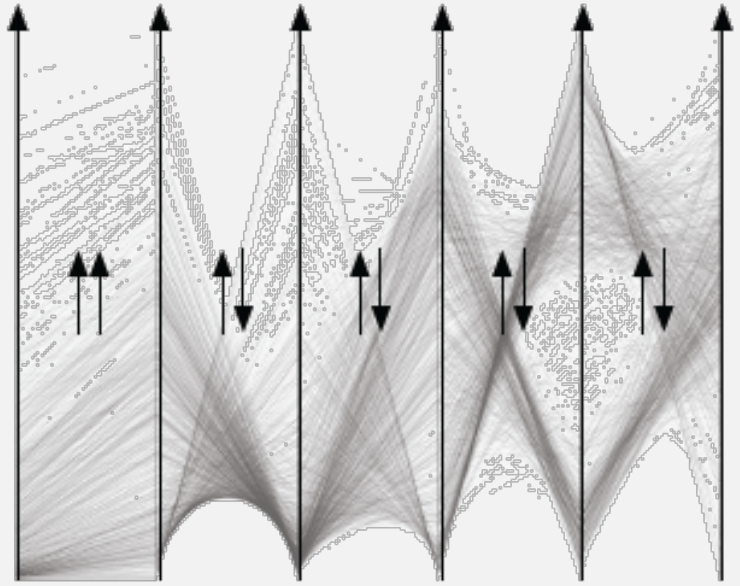
Basic

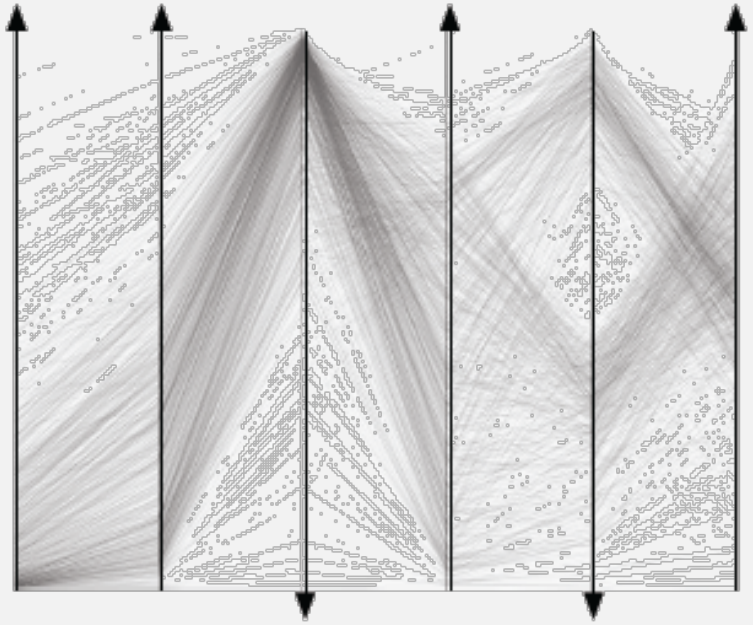


Grayscale



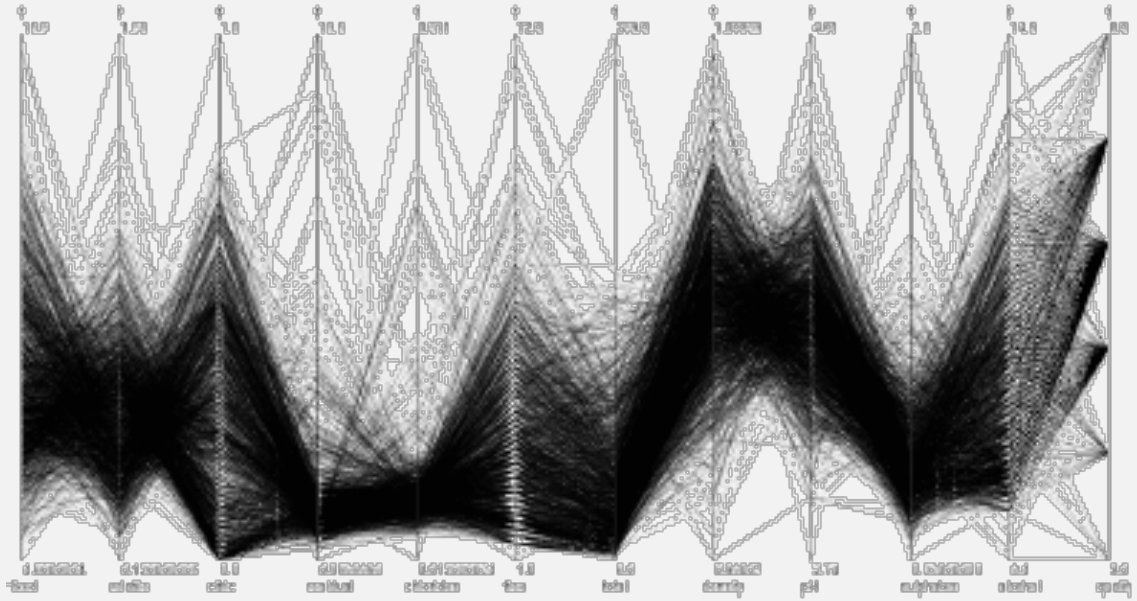
DIMENSION ORIENTATION

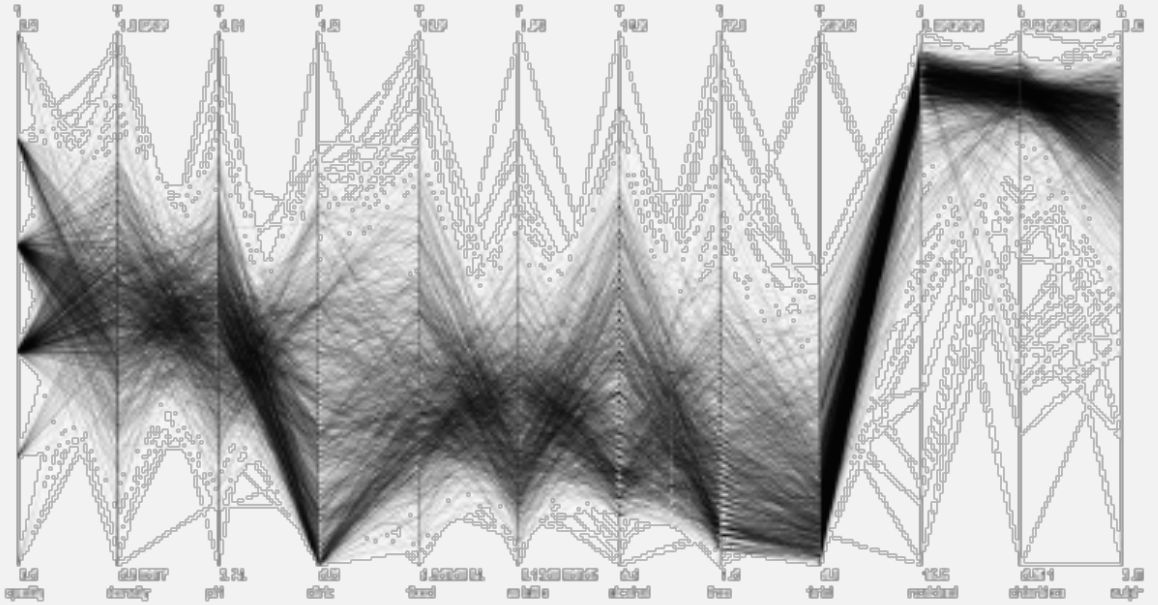




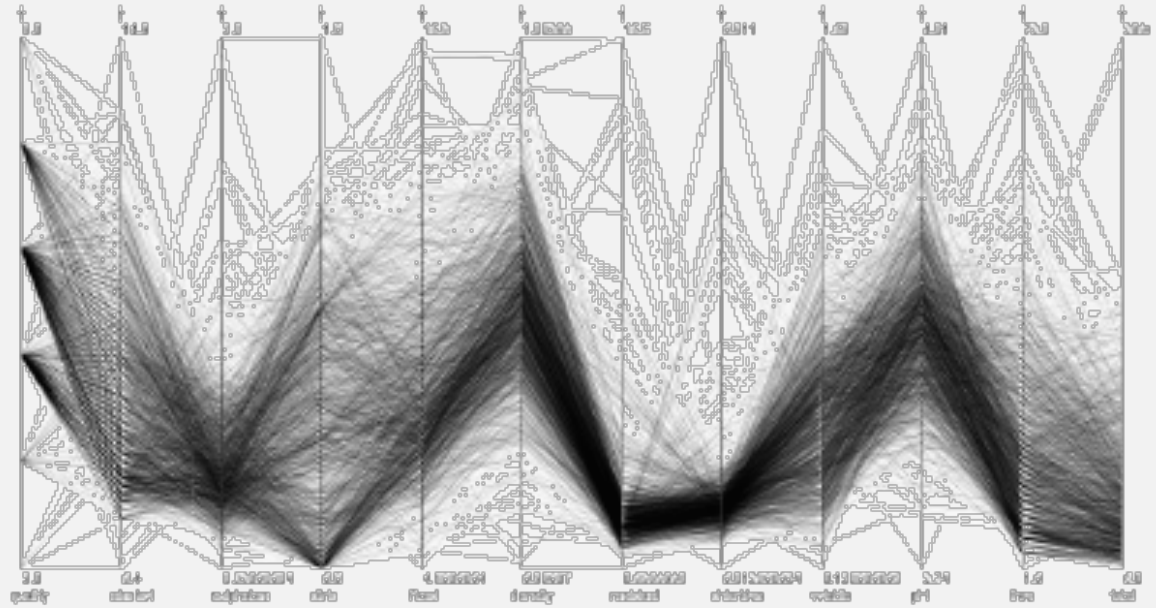
Dasgupta InfoVis 2010

DIMENSION ORDERING



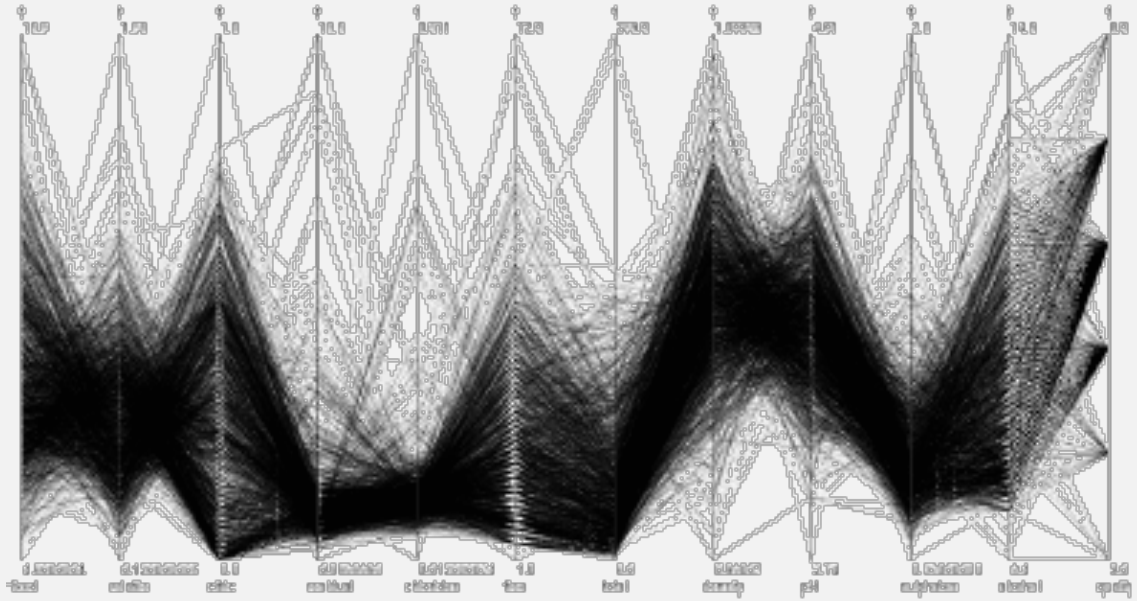


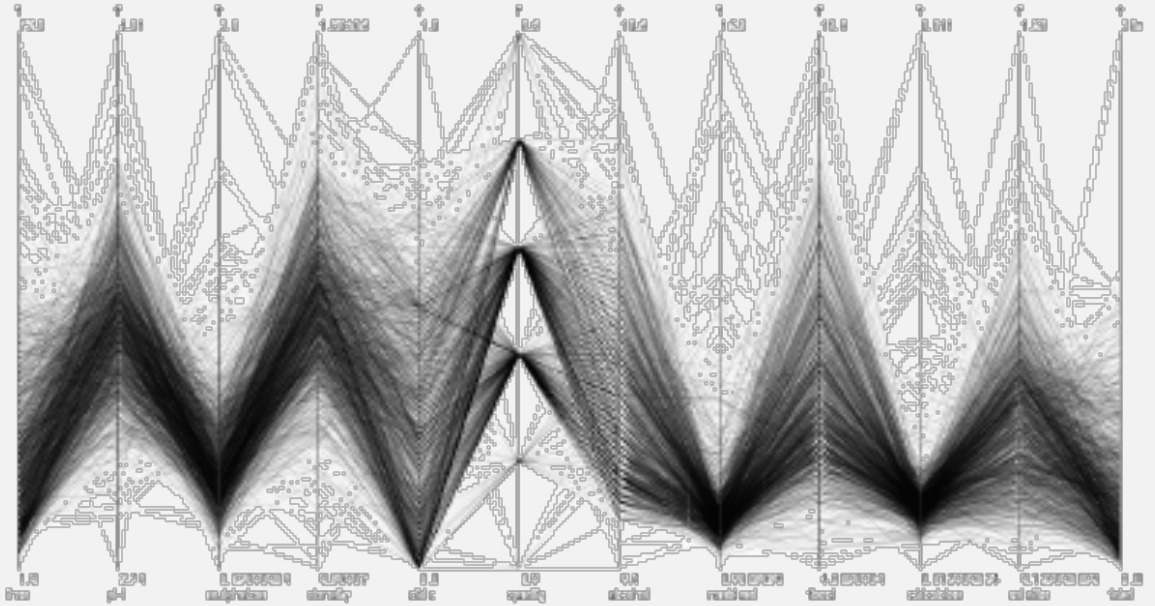
(a) View with maximized angles of crossing, including axis inversions (last three axes).



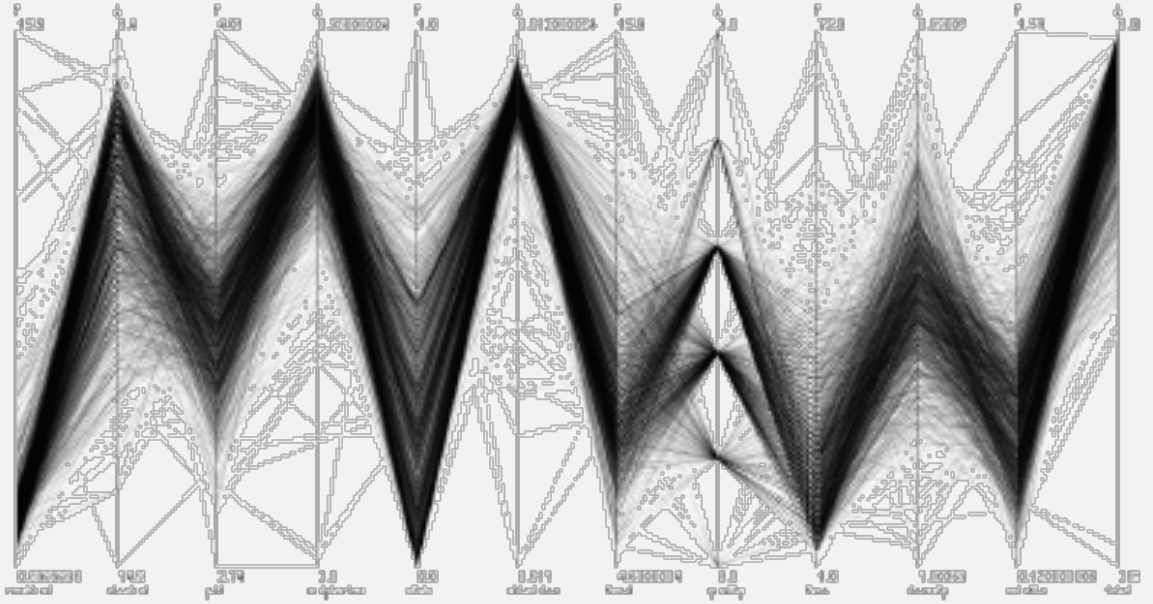
(b) Minimizing the number of crossings.

DIMENSION ORDERING



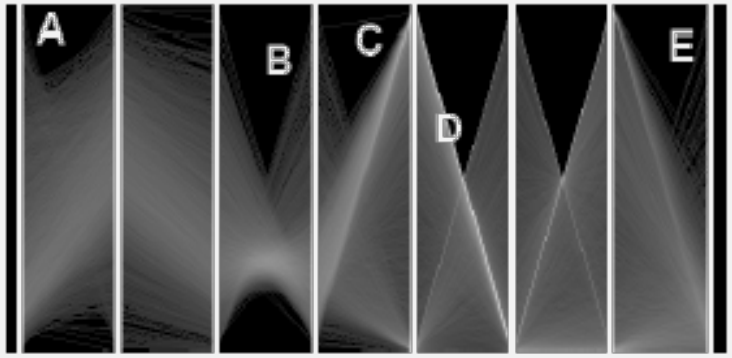


(c) Minimized angles of crossing and maximum parallelism.

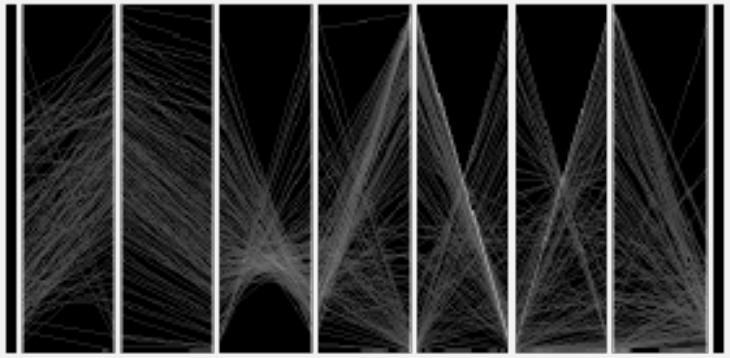


(d) Maximized number of crossings and minimized angles of crossing, and including inversions.

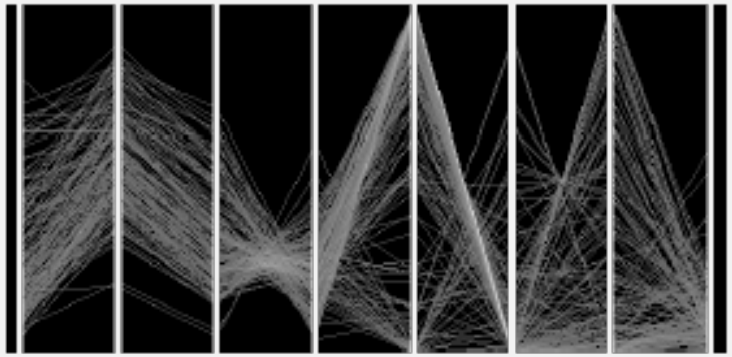
DENSITY REDUCTION



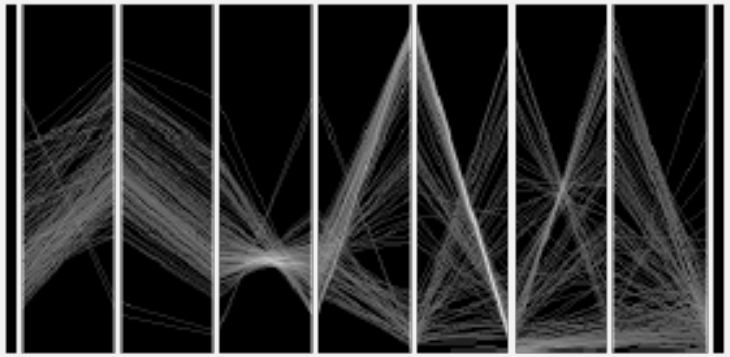
(a) Original data set comprising 7342 items.



(b) Targeting a visual quality of 0.90 retains 155 items.



(c) Using 155 randomly picked items results in a visual quality of 0.74.

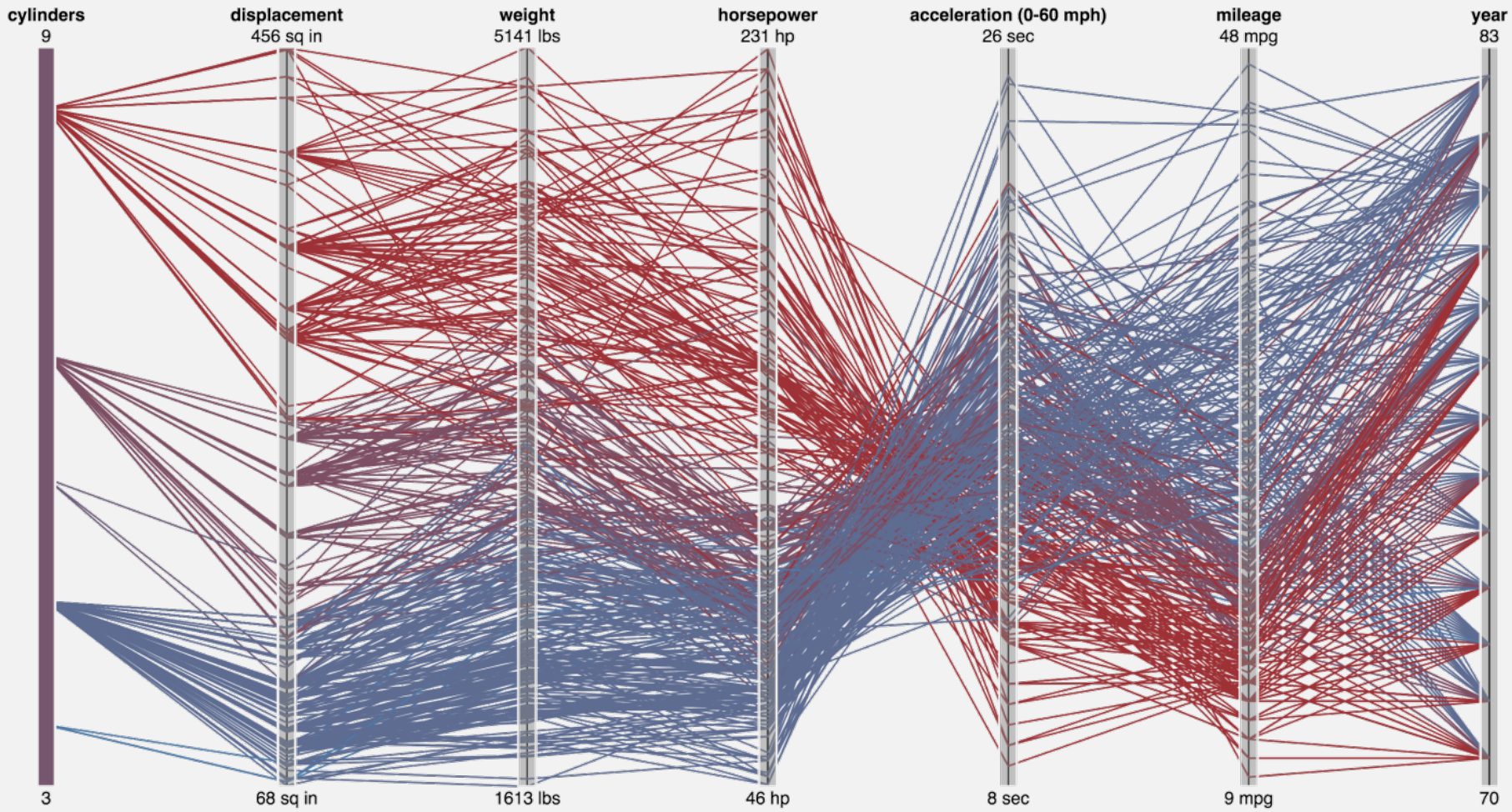


(d) Using the K-means algorithm to construct 155 clusters results in a visual quality of 0.76.

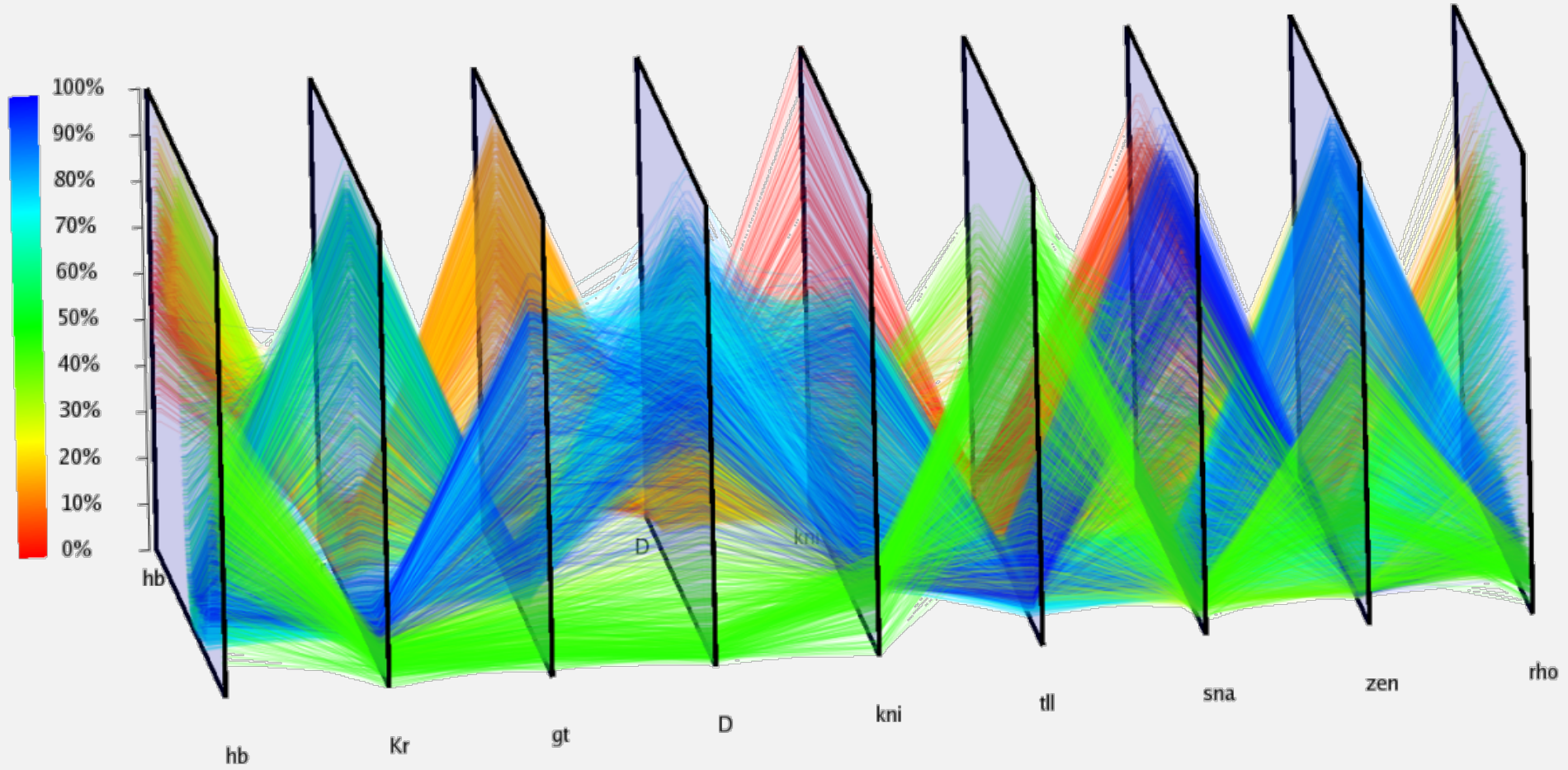
Figure 8: *Producing an abstraction of the household economics data set (a) by (b) targeting a visual quality of 0.90 using sampling retains 155 items. Structures are preserved in* **B**–**E** *and outliers are revealed in* **A**. *Randomly picking 155 items with no quality control (c) only preserves structures in* **D**. *Using 155 cluster centroids (d) also preserves structures in* **A** *and* **E**.

Johansson EuroVis 2008

PROTOVIS DEMO

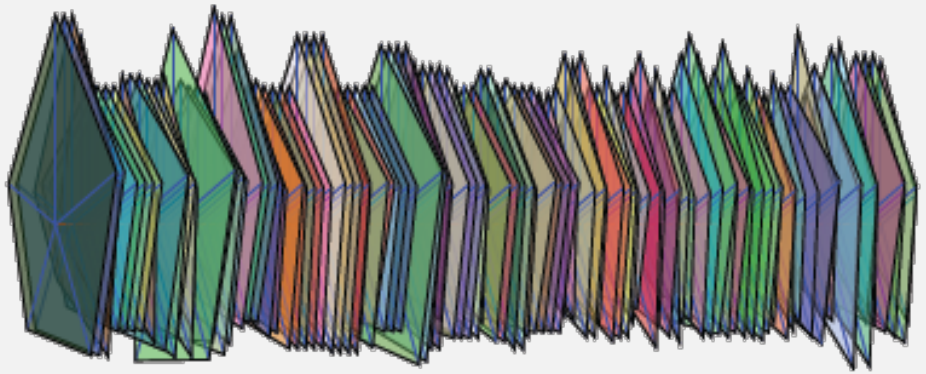


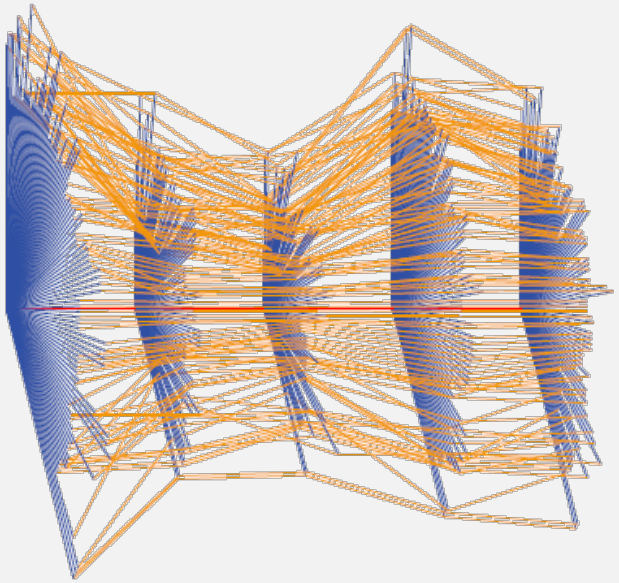
LOTS OF EXTENSIONS…

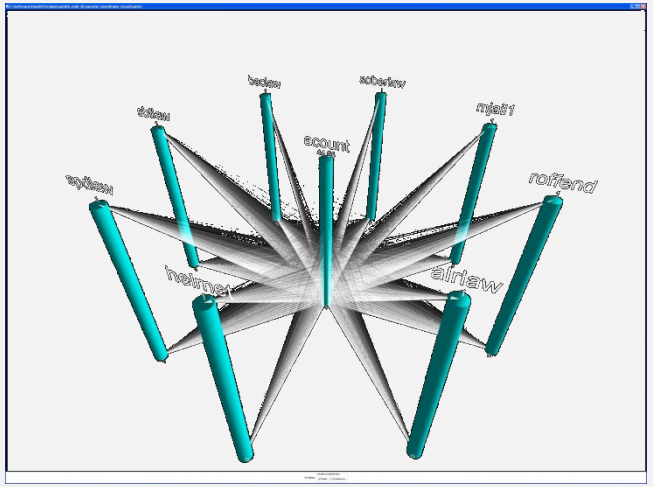


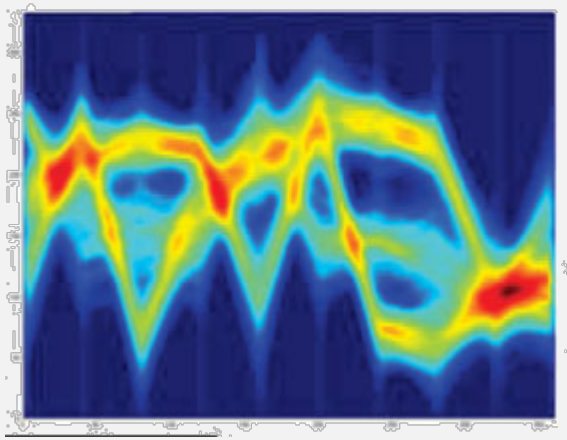
Drosophila Gene Expression Data Exploration and Visualization
Lawrence Berkeley National Lab

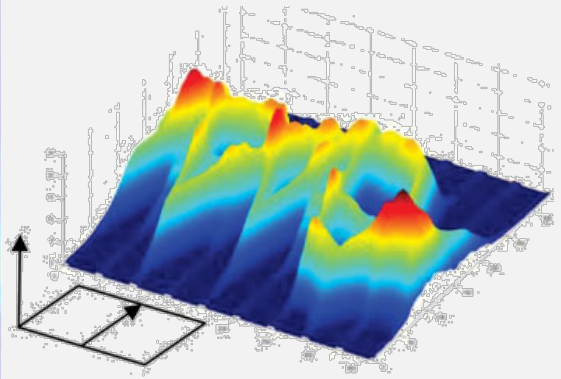
LOTS OF EXTENSIONS…











Fanea et al. InfoVis 05

Villaveces, UNC / Renci

Streit, Cytometry Part A, 2006

Assignment 3: Parallel Coordinates

NEXT TIME…

REQUIRED READING

Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases

Chris Stolte, Diane Tang, and Pat Hanrahan

Abstract—In the last several years, large multidimensional databases have become common in a variety of applications such as data warehousing and scientific computing. Analysis and exploration tasks place significant demands on the interfaces to these databases. Because of the size of the data sets, dense graphical representations are more effective for exploration than spreadsheets and charts. Furthermore, because of the exploratory nature of the analysis, it must be possible for the analysts to change visualizations rapidly as they pursue a cycle involving first hypothesis and then experimentation. In this paper, we present Polaris, an interface for exploring large multidimensional databases that extends the well-known Pivot Table interface. The novel features of Polaris include an interface for constructing visual specifications of table-based graphical displays and the ability to generate a precise set of relational queries from the visual specifications. The visual specifications can be rapidly and incrementally developed, giving the analyst visual feedback as they construct complex queries and visualizations.

Index Terms—Database visualization, database analysis, visualization formalism, multidimensional databases.