Data Representation

Sungahn Ko

HAiV

0

Disclaimer

HAiV

- **The slides cannot be distributed, posted or used outside of this class**

- **Slides in this course courtesy of**
  - Dr. Abish Malik (Purdue)
  - Dr. Yun Jang (Sejong Univ.)
  - Dr. Ross Maciejewski (ASU)
  - Dr. Niklas Elmqvist (UMD)
  - Dr. David Ebert (Purdue)

1

1

Data Models

- Data models/representations are structured forms suitable for computer-based transformations
- These structures exist in the original data or are derivable from the original data
- Structures retain the information and knowledge content and the related context within the original data
- These structures are transformable into lower-dimensional representations for visualization and analysis

2

2

Data Models vs. Conceptual Models

- **Data models are mathematical abstractions**
  - We can perform numerical operations, addition, subtraction, etc.
- **Conceptual models are our mental constructs**
  - These contain semantic structure and support reasoning
  - Think of giving directions to someone using landmarks

3

3

What Types of Data Models Do We Have?

HAiV

- **Relational Data Models**
  - These are data records that you deal with in excel, or typical database records
  - Data records are of fixed length and well-defined
  - Data is in rows and columns where each column has a domain type

	A	B	C
1	Animal	Brain Weight (kg)	Body Weight (kg)
2	Mountain beav	1.35	465
3	Cow	465	423
4	Grey wolf	36.33	119.5
5	Goat	27.66	115
6	Guinea pig	1.04	5.5
7	Dipliodocus	11700	50
8	Asian elephant	2547	4603

Codd, E.F. (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM 13(6): 37 7-387

4

4

What Types of Data Models Do We Have?

HAiV

- **Relational Data Models**
  - There may be a relationship defined in a schema between tables
  - A database is a collection of these relations

	A	B	C
1	Animal	Brain Weight (kg)	Body Weight (kg)
2	Mountain beav	1.35	465
3	Cow	465	423
4	Grey wolf	36.33	119.5
5	Goat	27.66	115
6	Guinea pig	1.04	5.5
7	Dipliodocus	11700	50
8	Asian elephant	2547	4603

Codd, E.F. (1970). "A Relational Model of Data for Large Shared Data Banks". Communications of the ACM 13(6): 37 7-387

5

5

What Types of Data Models Do We Have?

HAiV

- **Statistical Data Models**
  - This consists of variables or measurements (think of things like the census)
  - Categories are factors relating to the measurements
  - Observations or cases

Patient ID	Date	Chief Compl aints	Location
9398	4/16/09	Ear pain	Karachi
10816	4/16/09	Stuffy nose	Lebanon
1491	4/16/09	Fever	Allepo
16237	4/16/09	Head bleed	Yemen

6

6

Data Types

HAiV

- **Nominal**
  - Data whose categories have no implied ordering
  - Examples include political affiliations of a population
- **Ordinal**
  - Data that has a specified order, but no specified distance metric
  - Examples include beverage sizes at McDonalds (Small, medium, large)

1- SS Stevens, "On the Theory of Scales of Measurement," *Science*, 103(2684):677-680, 1946.

7

7

Data Types

- **Interval**
  - Data that has measurable distances
  - Examples include periods of time (second, minute, etc.) – the zero point is arbitrary
- **Ratio**
  - Same as interval, but include a zero point
  - Example include Celsius scale, height above sea level

1- SS Stevens, "On the Theory of Scales of Measurement," *Science*, 103(2684):677-680, 1946.

8

8

Mapping Data

- **We need to know how to assign quantitative dimensions of our data to *aesthetic attributes*[1] of the data**

Form	Surface	Motion	Sound	Text

Position
Size
Shape
  polygon
  glyph
  image
Rotation
Resolution
	Color
	  hue
	  brightness
	  saturation
	Texture
	  pattern
	  granularity
	  orientation
	Blur
	Transparency
		Direction
		Speed
		Acceleration
			Tone
			Volume
			Rhythm
			Voice
				Label

L Wilkinson (2005) *The Grammar of Graphics*

9

9

Aesthetic Attributes

- An attribute must be capable of representing both continuous and categorical variables
- When representing a continuous variable, an attribute must vary primarily on one psychophysical dimension
- In order to use multidimensional attributes (such as color), we must scale them on a single dimension
- An attribute does not imply a linear perceptual scale
- Much of the skill in graphic design is knowing what combination of attributes should be avoided[1]

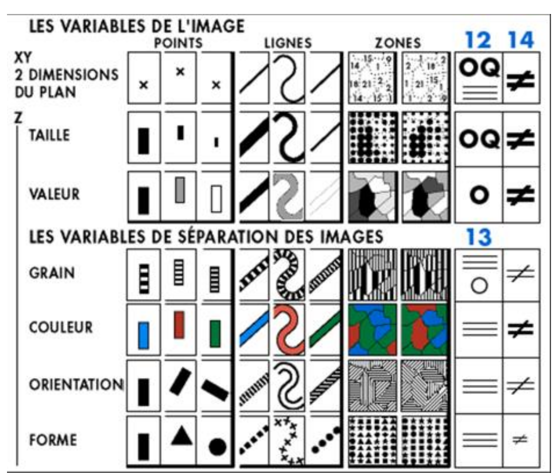
1-SM Kosslyn (1994), *The Elements of Graph Design*

10

10

8 Visual Variables

- **Visualization is concerned primarily with a mapping to visual form**
- **[x,y]**
  - Position
- **[z]**
  - Size (Taille)
  - Value (Valeur)
  - Color (Couleur)
  - Texture (Grain)
  - Orientation
  - Shape (Forme)



J Bertin (1967), *The Semiology of Graphics*

11

11

Position

HAiV

- Position refers to a location in a multi-dimensional space
- **Continuous variables** map to densely distributed locations
- **Categorical variables** map to a lattice
- Positions are ordered, but the ordering may or may not have meaning in terms of what is being measured
- Sometimes, position is just a way to keep things from overlapping

L Wilkinson (2005) *The Grammar of Graphics*

12

12

Position[1]

HAiV

- Cleveland[2] rates position on a common scale as the **best way to represent a quantitative dimension visually**
- This reflects research findings that points or line lengths placed adjacent to a common axis enable judgments with the least bias or error
- However, this recommendation has a caveat, as it depends on how far the graphic primitive (point, line, etc.) is from a reference axis[3]
- If a graphic is far from an axis, the multiple steps needed to store and decode the variation can impair judgment

1 - L Wilkinson (2005) *The Grammar of Graphics*
2 – WS Cleveland, *The Elements of Graphing Data*, 1985
3 – D Simkin and R Hastie (1987). An information processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454-465

13

13

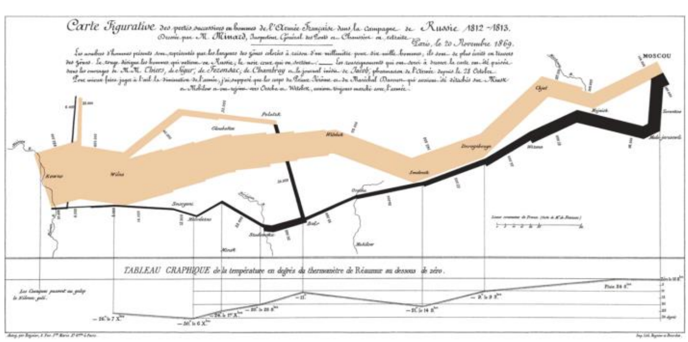
Size[1]

- Bertin[2] defines size variation in terms of length or area
- For three dimensions we have volume
- Cleveland[3] ranks area and volume representations among the **worst attributes** to use for graphing data
- Some designers assign size to only one dimension of an object
  - Think of the bar chart where the width of the bar is typically constant, but the height is varied

1 - L Wilkinson (2005) *The Grammar of Graphics*
2 – J Bertin (1967), The Semiology of Graphics
3 – WS Cleveland, *The Elements of Graphing Data*, 1985

Size

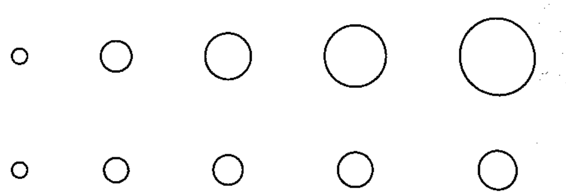
- Size for lines is usually equivalent to thickness
- This is less likely to induce perceptual distortion
- Size can be used to great effect with path



1 – Charles Joseph Minard: Mapping Napoleon's March, 1861 by John Corbett, Center for Spatially Integrated Social Science

Size

- For objects with rotational symmetry, we can map size to the diameter rather than area
- Representing data through area or volume should probably be confined to positively skewed data that can benefit from the perceptual equivalent of the square root transformation

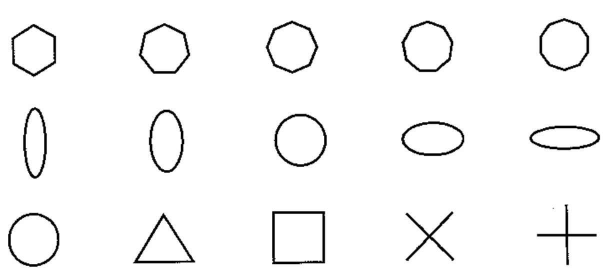


Top row changes diameter from 1-5
Bottom row changes area from 1-5

L Wilkinson (2005) *The Grammar of Graphics*

Shape

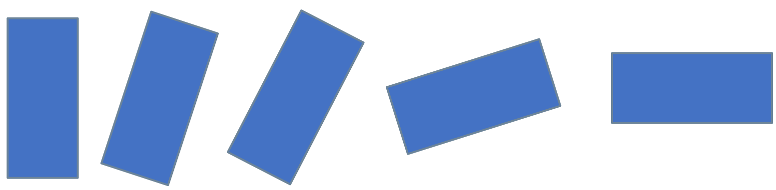
- Shape refers to the shape or boundary of an object
- Examples would include map symbols
- Shape must vary without affecting size, rotation and other attributes



L Wilkinson (2005) *The Grammar of Graphics*

Rotation

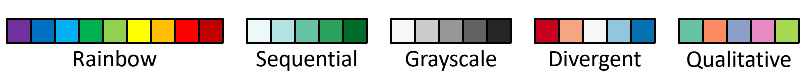
- This is the rotational angle of the graphic primitive
- Lines, areas and surfaces can only rotate if they are positionally unconstrained



L Wilkinson (2005) *The Grammar of Graphics*

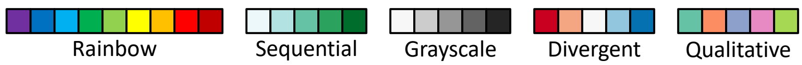
Color

- Color is really a psychological phenomenon
- The physical stimulus for color is light
- We see color because of the photoreceptors in our retina



Color

- Use of Rainbow?



Texture

- Texture includes pattern, granularity and orientation
- Granularity is the repetition of a pattern per unit of area
- Orientation is the angle of the pattern
- Textures can be described in a variety of ways
  - Fourier transform – decomposes a grid of brightness values into sums of trigonometric components
  - This decomposition is orientation dependent
  - Auto-correlogram – characterize the spatial moments of a texture

L Wilkinson (2005) *The Grammar of Graphics*

Texture

- Texture alone can be a basis for perception
- Two gray areas that have the same overall level of brightness can be discriminated if their texture is different

L Wilkinson (2005) *The Grammar of Graphics*

Texture

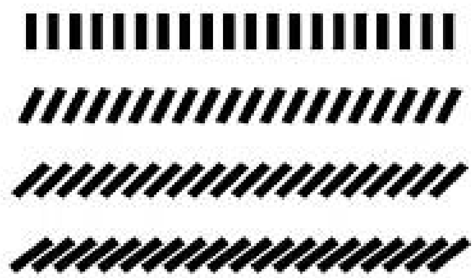
- Granularity – Changing the size or resolution of patterns in a texture changes the granularity
  - Less grainy patterns (those with low-frequency spatial components) are more difficult to resolve
- Pattern – Patterns make use of increasing degrees of randomness to encode data

L Wilkinson (2005) *The Grammar of Graphics*

Texture

HAiV

- Orientation – Variation in texture orientation can introduce visual illusions (making lines not seem parallel) and is typically avoided as an encoding in textures

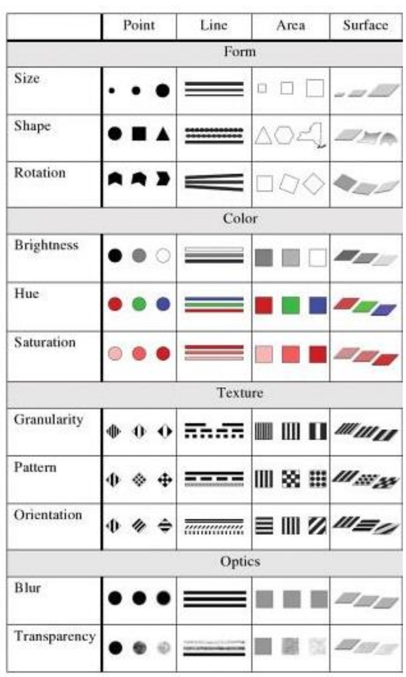


L Wilkinson (2005) *The Grammar of Graphics*

24

24

HAiV



L Wilkinson (2005) *The Grammar of Graphics*, p. 317

25

25

Groups of Representation

- Bertin describes four groups of representations
- *Diagrams* – correspondences are between a single component (think of things like a bar chart or a line graph)
- *Networks* – correspondences between different components are linked together
- *Maps* – correspondences between components are arranged according to geography
- *Symbols* – correspondences between components is linked to a symbol or a shape

J Bertin (1967), *The Semiology of Graphics*

26

26

Taxonomy by Data Type[1]

- 1D (sets and sequences)
- Temporal
- 2D (maps)
- 3D (shapes)
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)
- Text and Documents[2]

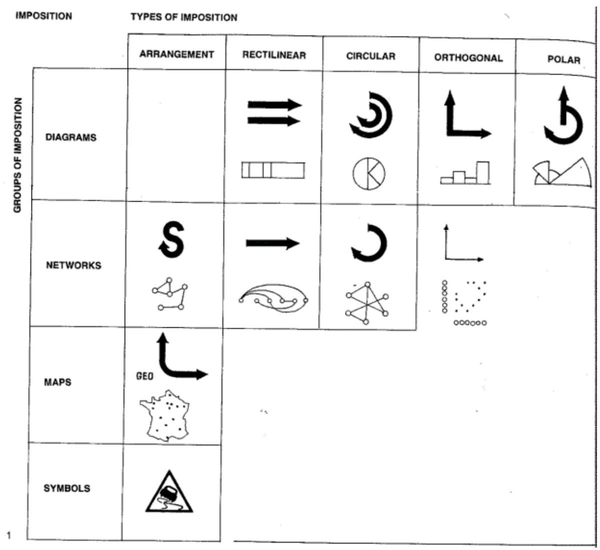
1 – B. Shneiderman. *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. In the Proceedings of the IEEE Symposium on Visual Languages, pp. 336-343, 1996.
2 –Borrowed from Pat Hanrahan's "From Data to Image" lecture: http://graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding/

27

27

Types of Graphic Representation

HAiV



J Bertin (1967), *The Semiology of Graphics*, p. 52

28

28

Cleveland's Hierarchy[1]

HAiV

- Cleveland evaluated elements when isolated
- Tasks were restricted to magnitude and ratio comparisons
- Research indicates this hierarchy may be best in pre-attentive stages or when focusing only on portions of a graphic[2,3,4]

Best

1 - Position along a common scale

2 – Position along nonaligned scales

3 – Length

4 – Angle/Slope

5 – Area

6 – Volume

7 - Color

Worst

1 – WS Cleveland, *The Elements of Graphing Data*, 1985
2 – Spence, I., and Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology* 5, 61-77
3 – Simkin, D., and Hastie, R. (1987). An information processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454-465
4 – Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors*, 4, 535-554

29

29

What if I Combine Encodings?

HAiV

- So, what happens when we combine several scales in a single display?
- Can we represent one quantitative dimension with color and another with orientation and expect a perceiver to respond to both dimensions?
- Do these things make psychological sense?

L Wilkinson (2005) *The Grammar of Graphics*

30

30

Integral Versus Separable Dimensions

HAiV

- "A configuration has properties that have to be expressed as some form of interaction or interrelation between the components, be they features or dimensions."[1]
- *Integral dimensions* are not as easily decomposable by perceivers as *separable dimensions*
- Separating hue from brightness in a color (integral dimensions) are harder to decompose than say size and texture (separable dimensions)

1 – Garner, W. R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy and J. R. Pomerantz (Eds.), *Perceptual Organization* (pp. 119-139). Hillsdale, NJ: Lawrence Erlbaum.

31

31

Integral Versus Separable Dimensions

- Discriminations between classes defined by multiple integral stimuli are more difficult than those between classes defined by separable stimuli
- Selective attention to only one dimension is more difficult with integral stimuli
- Redundant cues in classification improve performance for integral stimuli and degrade performance for separable

32

32

Combinatorics of Encodings

- **Principle of Consistency:** The properties of the image should match the properties of the data
- **Principle of Importance Ordering:** Encode the most important information in the most effective way

1 – This slide is borrowed from Pat Hanrahan's "From Data to Image" lecture: http://graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding/

33

33

The Expressiveness & Effectiveness Criteria

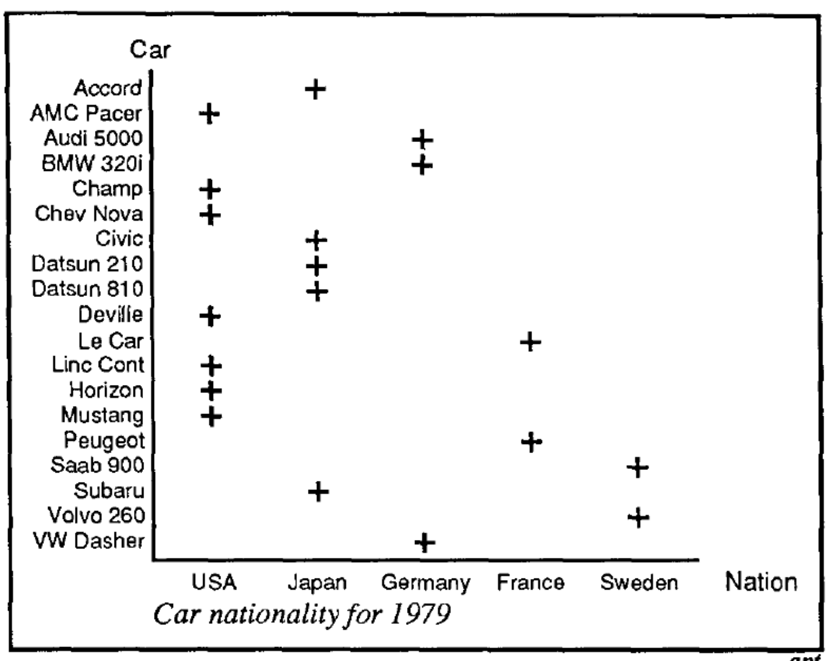
HAiV

- **Expressiveness:** A set of facts is expressible in a visual language if the visualizations express all the facts in the set of data, and only the facts in the data

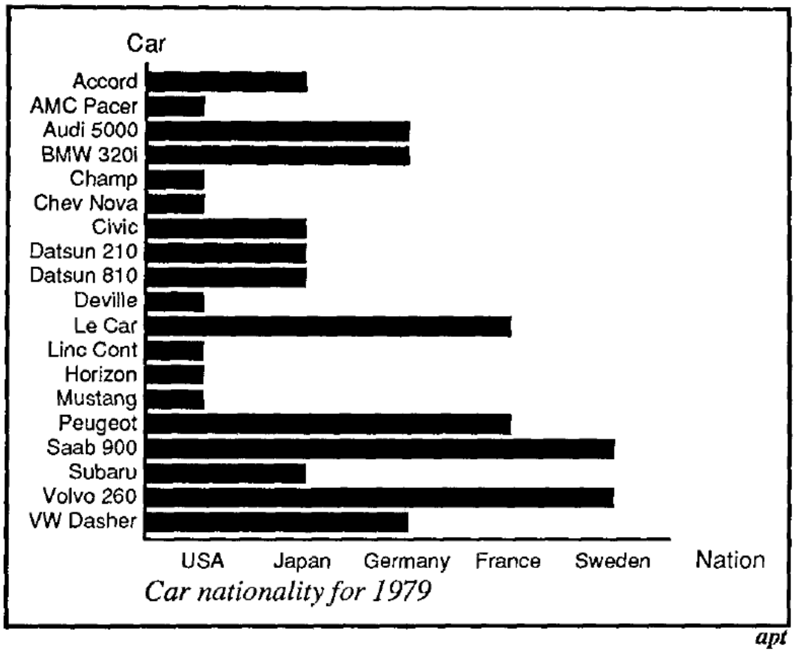
- **Effectiveness:** A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization

1 – J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics, 5(2): 110-141, 1986.

HAiV



1 – J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics, 5(2): 110-141, 1986.



1 – J. Mackinlay, Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics, 5(2): 110-141, 1986.

36

36

Potentially Expressive/Effective

- Extensions to the expressive/effectiveness criteria are necessary under interaction
- Potential Expressiveness: A visualization is potentially expressive if it has the potential, under user control, to display all its assigned information over time
- Potential Effectiveness: A visualization is potentially effective if over time it can present the information sufficiently clearly

1 – C. Beshers and S. Feiner, Auto Visual: Rule-Based Design of Interactive Multivariate Visualizations, IEEE Computer Graphics & Applications, 13(4): 41-49, 1993.
2 - H Senay and E Ignatius, A Knowledge-Based System for Visualization Design, IEEE Computer Graphics & Applications, 14(6): 36-47, 1994

37

37