Visualization Basics

What is data visualization?

“Data visualization is the graphical representation of information and data.” ~*Tableau*

“Data visualization is the representation of data through use of common graphics, such as charts, plots, infographics, and even animations. These visual displays of information communicate complex data relationships and data-driven insights in a way that is easy to understand.” ~*IBM*

“**Data and information visualization** (**data viz** or **info viz**) is an interdisciplinary field that deals with the graphic representation of data and information.” ~*Wikipedia*

Main Subfields

Scientific visualization: model real-world phenomena

Information visualization: map a more abstract concept into 2D or 3D for decision making and analysis purposes

Visual analytics: the frontier of Data mining and Machine Learning

Ref: Key Concepts of Data Visualization

Visualization is Efficient

- Increases memory and leverages available processing resources
- Reduces search for information
- Enhances pattern recognition
- Enables perceptual inference operations
- Uses perceptual attention mechanisms for monitoring
- Encodes information in a manipulable medium

Ref: Key Concepts of Data Visualization

Perception Enablers

Preattentive processing : some visual features can be perceived very rapidly and accurately by our low-level visual system (e.g identifying a red dot in the middle or grey dots)

Gestalt theory: the visual systems understand an image using proximity, similarity, continuity, symmetry, close and relative size features.

Ref: Key Concepts of Data Visualization

Tukey

Consistent with this view, we believe, is a clear demand that pictures based on exploration of data should *force* their messages upon us. Pictures that emphasize what we already know--"security blankets" to reassure us--are frequently not worth the space they take. Pictures that have to be gone over with a reading glass to see the main point are wasteful of time and inadequate of effect. **The greatest value of a picture** is when it *forces* us to notice **what we never expected to see.**

Tufte

Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency. Graphical displays should

- show the data
- induce the viewer to think about the substance rather than abou methodology, graphic design, the technology of graphic pro- duction, or something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.

Graphics *reveal* data. Indeed graphics can be more precise and revealing than conventional statistical computations. Consider Anscombe's quartet: all four of these data sets are described by exactly the same linear model (at least until the residuals are ex- amined).

Data-Ink Ratio

$$\frac{data - ink}{total - ink} = \frac{Elements\ conveying\ data\ information}{All\ elements\ in\ the\ chart}$$

“Above all else show the data.”

~Tufte

When to Visualize Data

- When there is a good underlying structure so that items close to one another can be inferred to be similar
- When users are unfamiliar with a collection's contents
- When users have a limited understanding of how a system is organized and prefer a less cognitively loaded method of exploration
- When users have difficulty verbalizing the underlying information need
- When information is easier to recognize than describe

Ref: Key Concepts of Data Visualization

Mapping Data to a **Representation**

- Define a space, usually using axis
- Define the marks (points, stars, dots…)
- Define the graphical properties of marks (size, position, orientation, color, texture…)

Intent: increase the use of space, encode data, and make the graph efficient.

Ref: Key Concepts of Data Visualization

5 Basic **Representations**

- Data table
- Graphs on rails or planes
- Geospatial
- Network diagram
- Symbols

Ref: Key Concepts of Data Visualization

Data Table

Photos: Add photo/video to iCloud Photo Library																					
Date	Platform	Operating System Version	OS Build Version	Device Family	Device Type	Browser Name	Browser Version	Country Code (IP Address Derived)	City (IP Address Derived)	File Capture Date	File Type	Sub File Type	File Source	Camera Option	Photos Camera Setting	Live Photo Effect	Pixel Width	Pixel Height	Video Duration (seconds)	File Size (KB)
2019-03-22	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-22	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	1372.3
2019-03-22	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-22	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	457.95
2019-03-22	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-22	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	636.59
2019-03-22	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-22	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	818.87
2019-03-23	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr		2019-03-23	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	1200.41
2019-03-25	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	segny	2019-03-25	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	414.73
2019-03-26	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	lyon	2019-03-26	photo	photo	photokit	livephoto_off_hdr_off			750	1334	0.0	156.33
2019-03-26	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-26	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	577.95
2019-03-27	iOS	11.4.1	15g77	iPhone	iPhone 8			fr	paris	2019-03-14	video	video	photokit	livephoto_off_hdr_off			848	480	16.91	9275.85
2019-03-27	iOS	11.4.1	15g77	iPhone	iPhone 8			fr	paris	2019-03-18	photo	photo	photokit	livephoto_off_hdr_off			1600	1200	0.0	133.87
2019-03-27	iOS	11.4.1	15g77	iPhone	iPhone 8			fr	paris	2019-03-18	photo	photo	photokit	livephoto_off_hdr_off			1600	1200	0.0	143.62
2019-03-27	iOS	11.4.1	15g77	iPhone	iPhone 8			fr	paris	2019-03-18	photo	photo	photokit	livephoto_off_hdr_off			1600	1200	0.0	144.13
2019-03-27	iOS	11.4.1	15g77	iPhone	iPhone 8			fr	paris	2019-03-18	photo	photo	photokit	livephoto_off_hdr_off			1600	898	0.0	209.87
2019-03-28	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-28	photo	photo	photokit	livephoto_off_hdr_off			1200	1600	0.0	196.55
2019-03-28	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-28	photo	photo	photokit	livephoto_off_hdr_off			1200	1600	0.0	218.94
2019-03-28	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-28	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	476.22
2019-03-28	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-28	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	485.49
2019-03-28	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-28	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	571.7
2019-03-28	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-28	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	630.73
2019-03-28	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-28	photo	screenshot	photokitInternal	livephoto_off_hdr_off			1242	2208	0.0	760.95
2019-03-29	iOS	12.0	16a366	iPhone	iPhone 6 Plus			fr	paris	2019-03-29	photo	photo	photokit	livephoto_off_hdr_off			750	1334	0.0	160.75

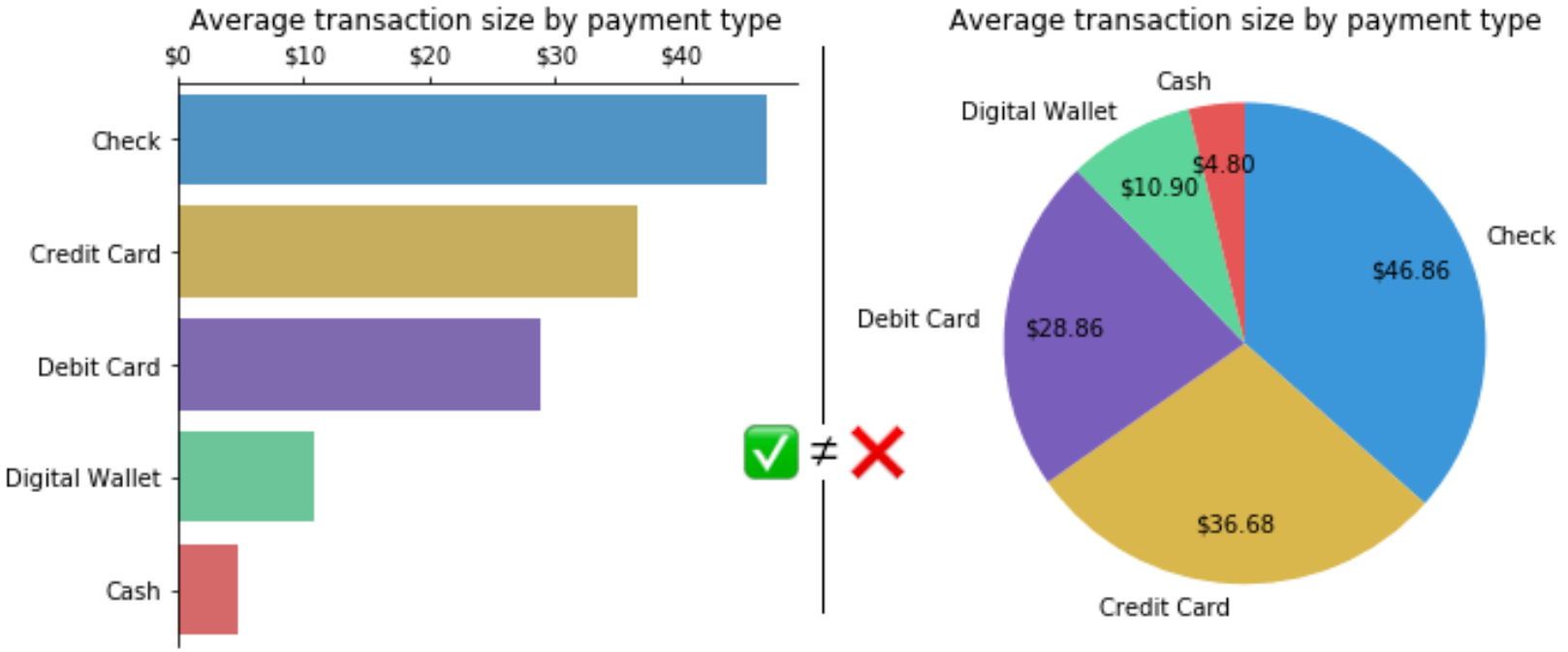
Stem and Leaf Plot

STEMS	LEAVES
0	5 8
1	2 3 5 7
2	0 0 0 5 8 8 9
3	0 0 1 3 3 3 6 6 7 7 7 7 7 8 8 8 8 9 9
4	1 3 5 5 5 6 7 7 8 8 8 9 9
5	0 0 0 1 1 1 1 2 6 8
6	0 0 1 1 2 4 4 4 4 8 8 9
7	0 5 5 5 5 7
8	3 4 4 5 6 6 6 7 8 9
9	0 1 2 2 2 2 5 5 5 6 8 9 9
10	2 2 2 5 7

Raw Data in Row:
102, 102, 102, 105, 107

Ref: Data Table
Stem and Leaf Plot

Rails or Planes



Bar Chart

Pie Chart

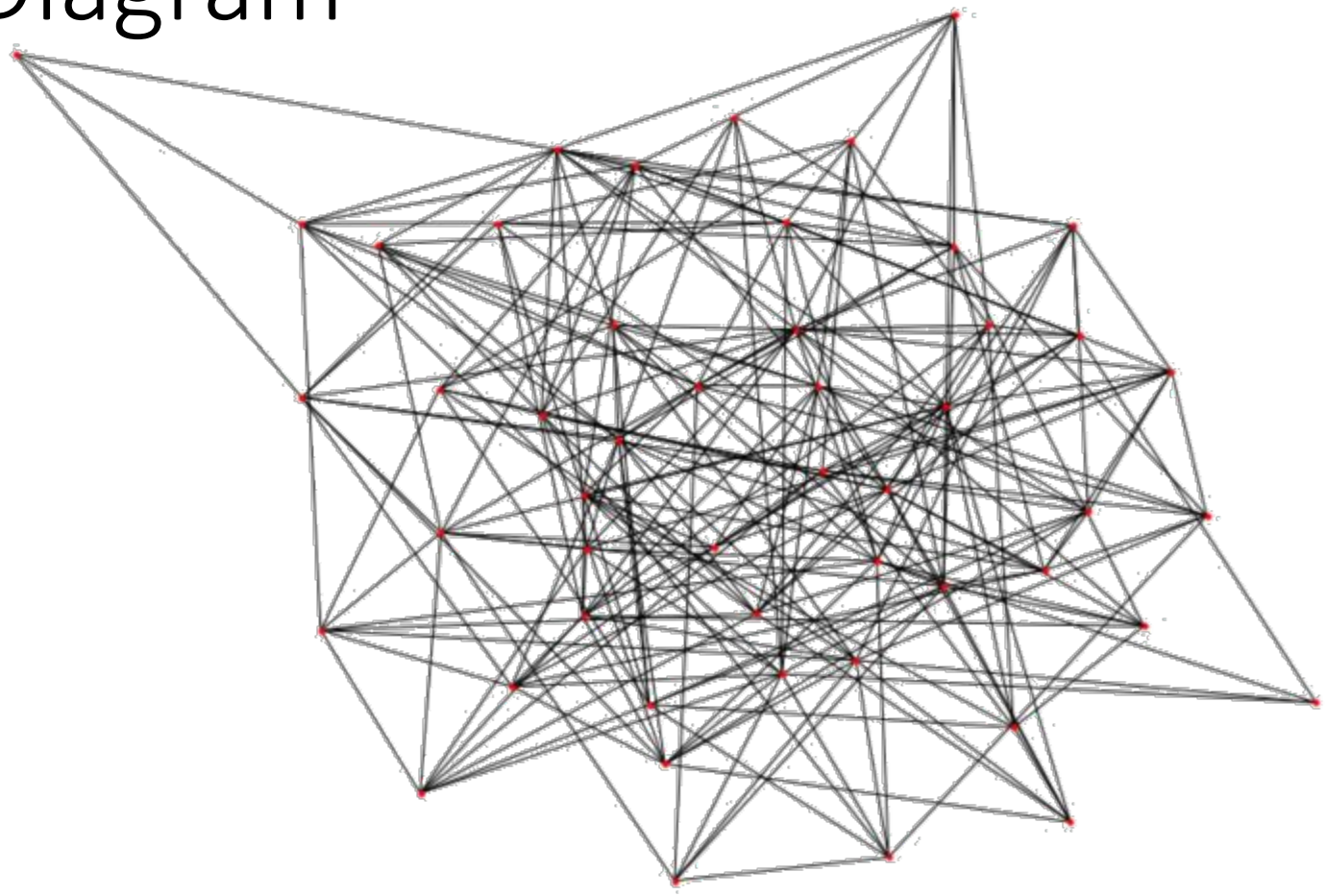
Ref: [Bar Chart, Pie Chart]

Geospatial

Heat Map

Ref: [Heat Map]

Network Diagram

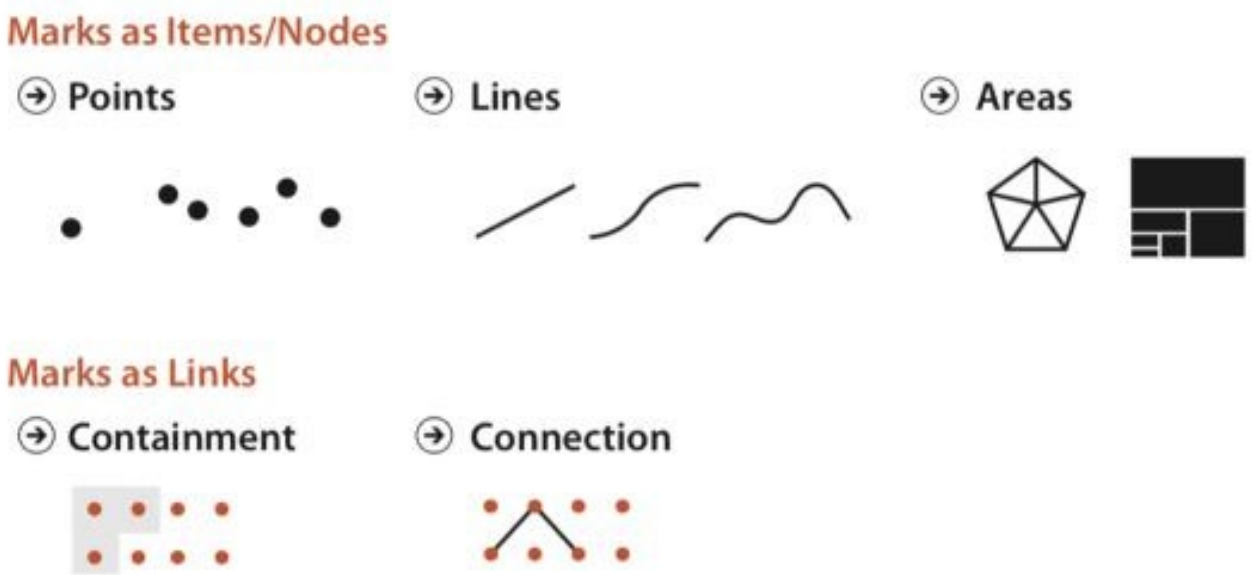


Ref: Network Diagram

Symbols

Marks

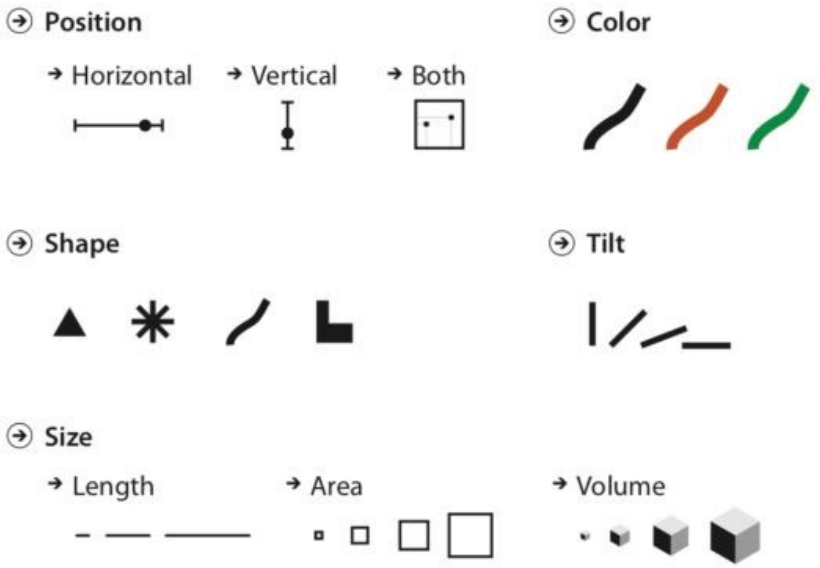
Marks depict items or links.



Ref: Key Concepts of Data Visualization

Channels

Channels control appearance independent of the geometric primitive.



Ref: [Key Concepts of Data Visualization]

Choosing Channels

Channels that show *magnitude* information are good for *ordered data*.

- Quantity
- Duration

Channels that show *identity* information are good for *categorical data*.

- Identity
- Location
- Type

Ref: Key Concepts of Data Visualization

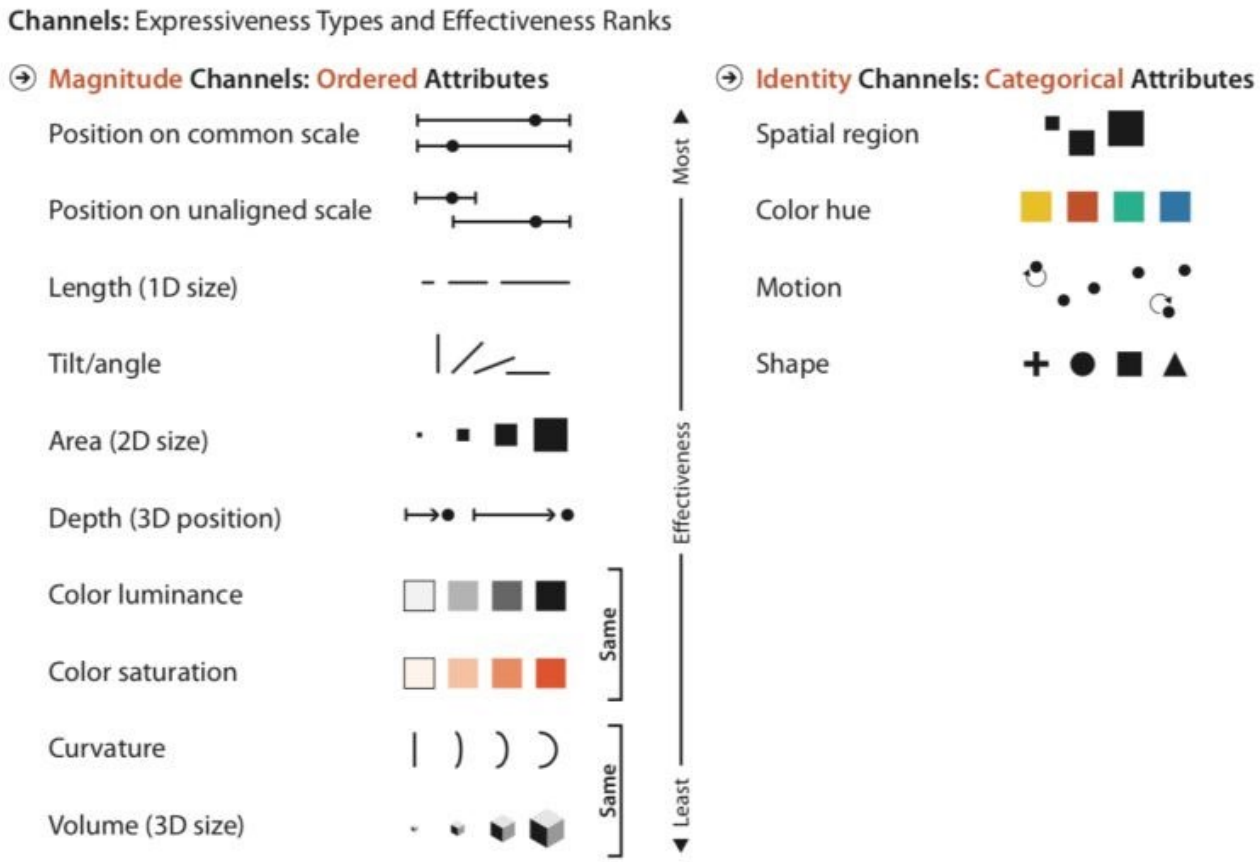
Choosing Channels

Expressiveness: visual encoding should express only the information in the dataset attributes. Ordered data should be shown as ordered, and vice versa

Effectiveness: the most important attributes should be encoded with the most effective channels to be most noticeable

Ref: Key Concepts of Data Visualization

Choosing Channels



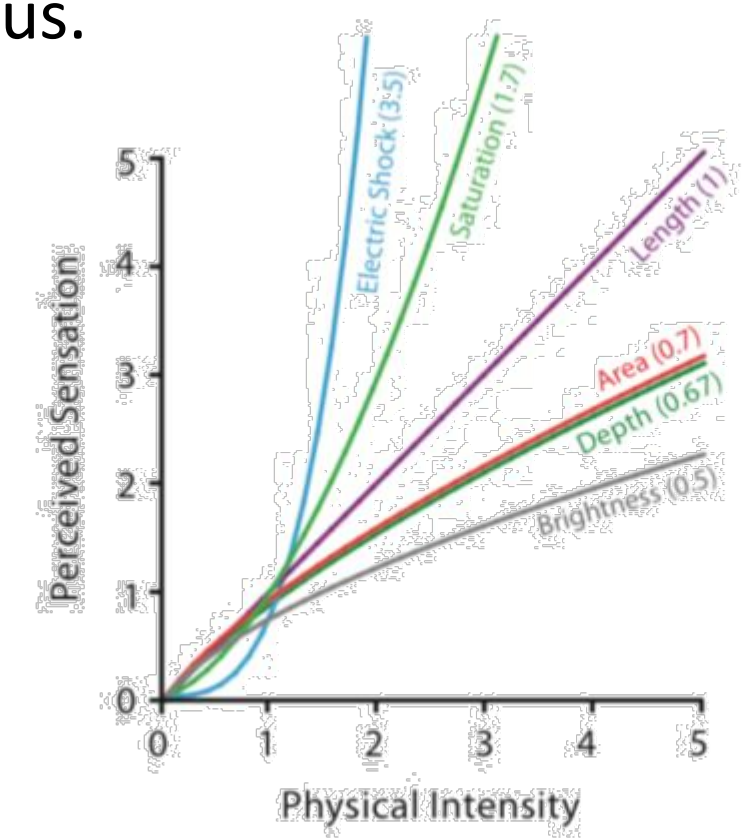
Ref: [Key Concepts of Data Visualization]

Measuring Effectiveness

Accuracy: Relationship of perceptual judgement to objective measurement of stimulus.

Stevens psychophysical power law:

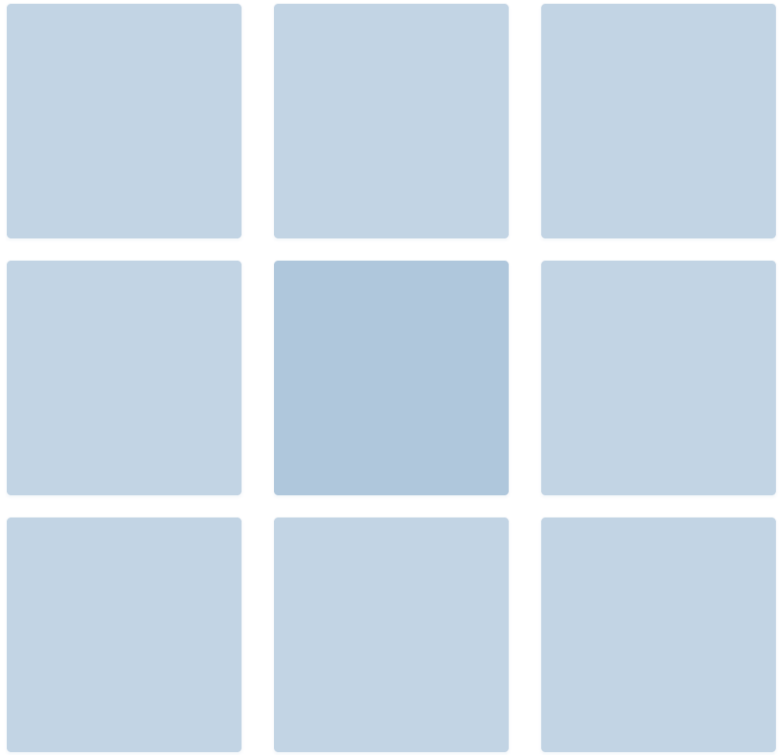
$$S = I^n$$



Ref: Key Concepts of Data Visualization

Measuring Effectiveness

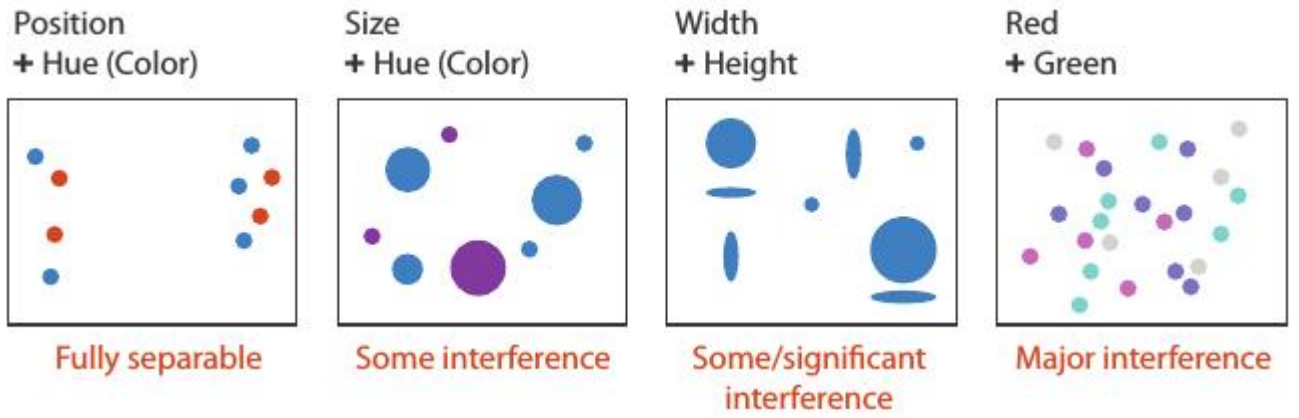
Discriminability: Can the items be differentiated by humans as intended?



Ref: Key Concepts of Data Visualization

Measuring Effectiveness

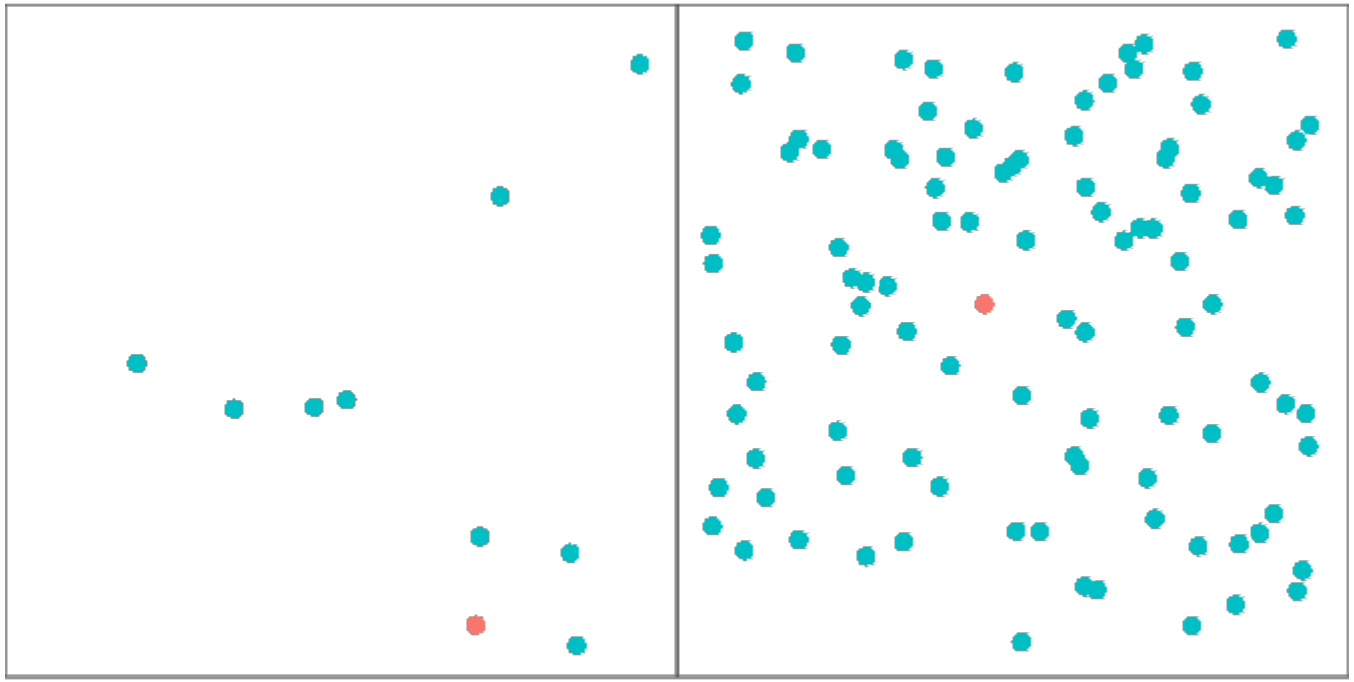
Separability: Can encoding channels be used independently?



Ref: Perception and Visualization

Measuring Effectiveness

Popout: Do outliers stand out?

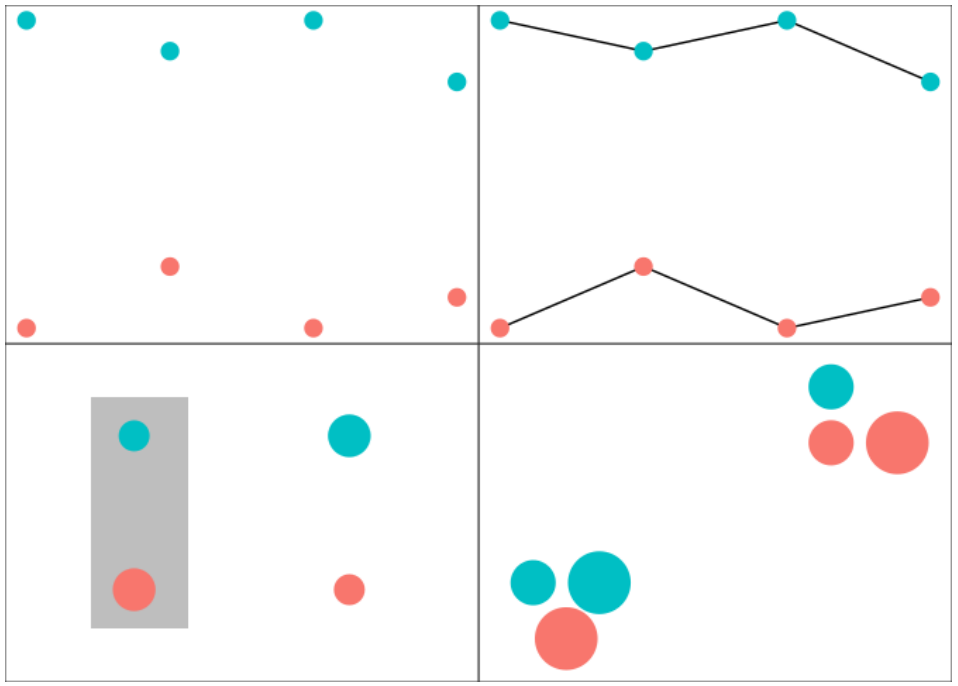


Ref: Perception and Visualization

Measuring Effectiveness

Grouping: The extent to which related items may be perceived as such.

Identity Channel



Links

Enclosure

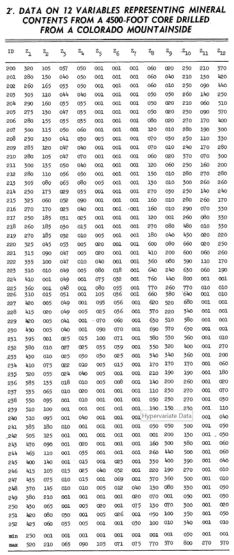
Proximity

Ref: Perception and Visualization

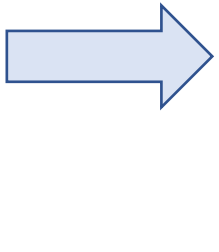
Hypervariate Data

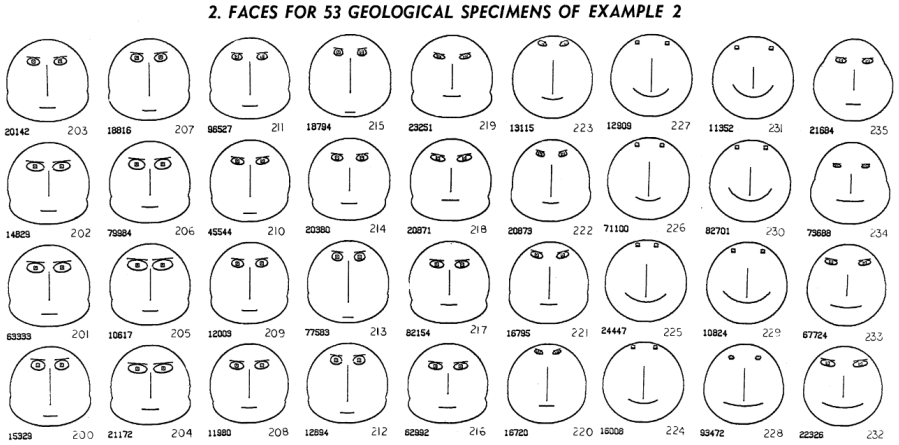
- Highly dimensional, **hypervariate,** data is challenging to represent in a way that can be interpreted easily.
  - Data table - inefficient
  - Dimensionality reduction – data loss
  - Scatterplot matrix – limited dimensions, overwhelming
  - Feature encoding – limited number of features

[Chernoff Faces](#)



2'. DATA ON 12 VARIABLES REPRESENTING MINERAL CONTENTS FROM A 4500-FOOT CORE DRILLED FROM A COLORADO MOUNTAINSIDE



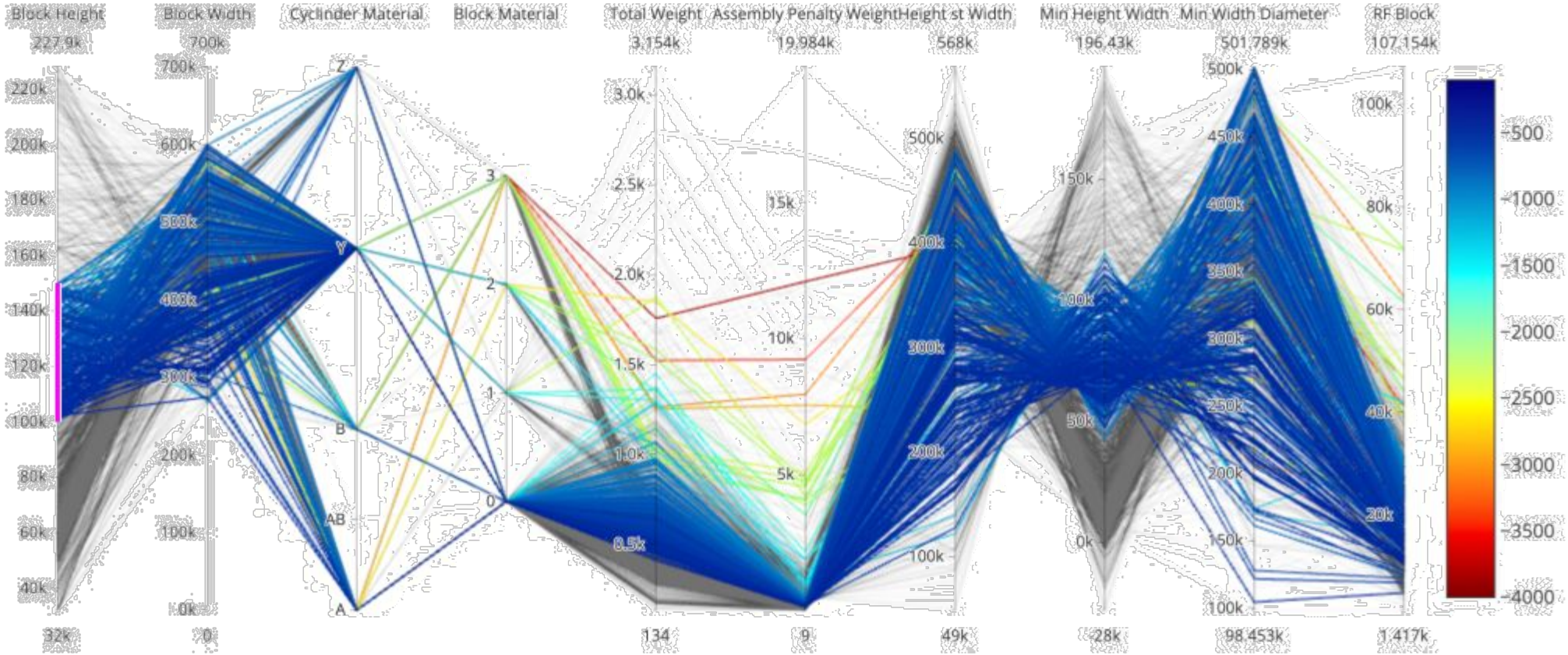


2. FACES FOR 53 GEOLOGICAL SPECIMENS OF EXAMPLE 2

Ref: [Key Concepts of Data Visualization](#)

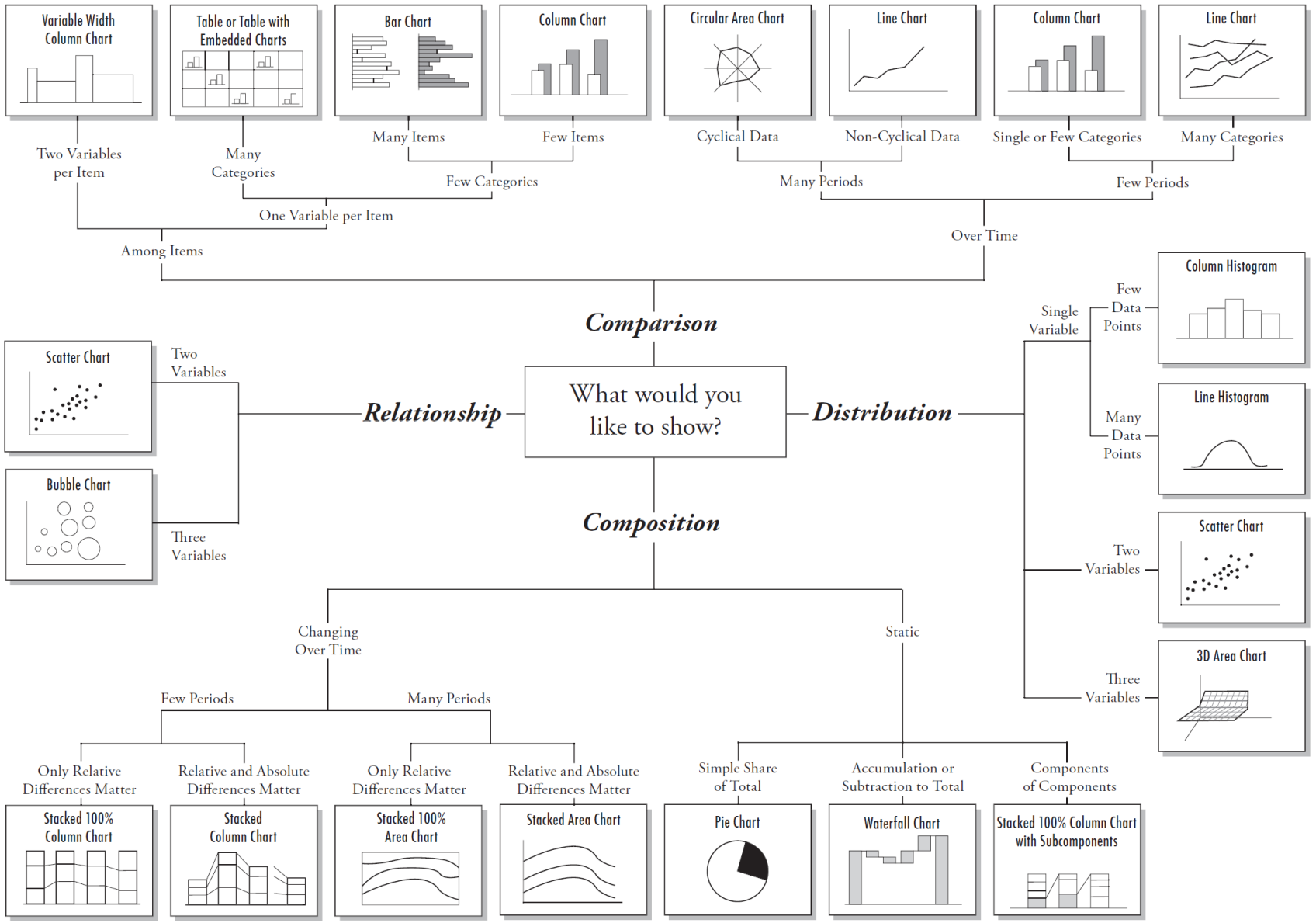
Hypervariate Data

- Parallel Coordinate Plot



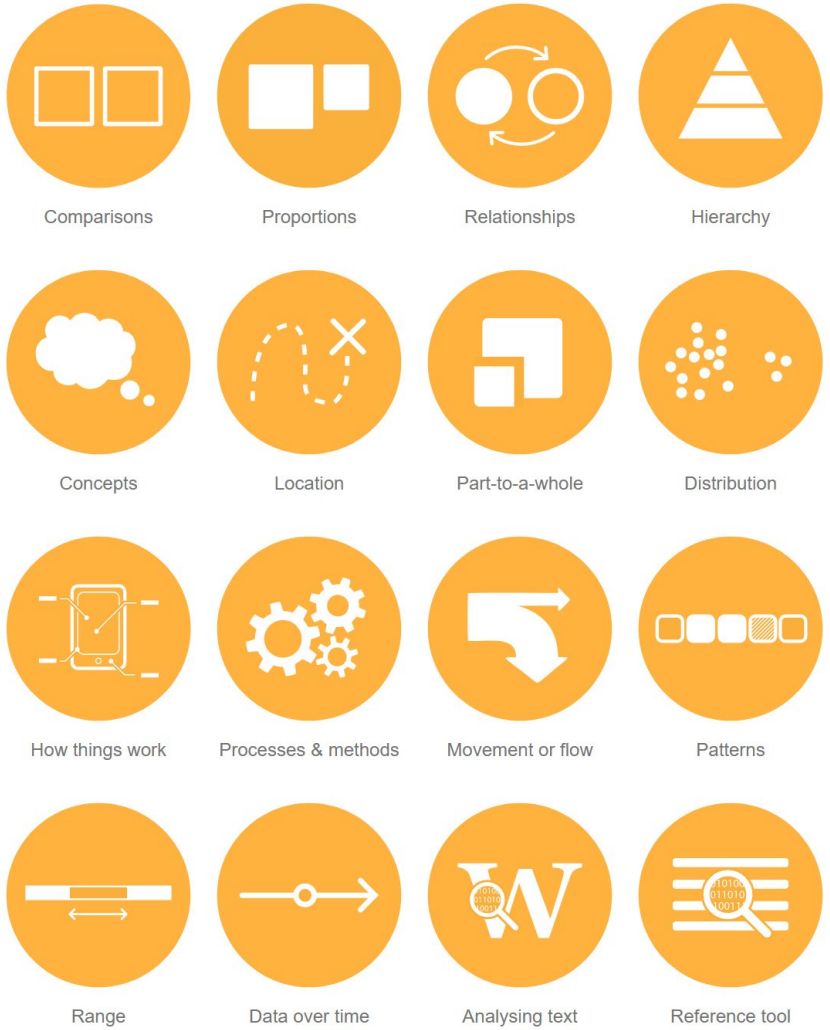
Ref: [Key Concepts of Data Visualization]

Chart Suggestions—A Thought-Starter



Flow
Sankey
Flow Map
Chord Diagram

My *New Favorite* Data Visualization Reference



Further Reading

Five Ways to Lie with Charts

Homework 4 - Plan your visualizations!

When exploring and analyzing your data you should be interpreting what the data tell you in order to draw conclusions.

One of the best ways to do this is through visualization. Especially if you plan to share your findings.

Based on your initial effort, describe the types of visualization you intend to include in your final presentation. Plan at least 3. Include:

- Visualization type
- Data to be used
- How dimensions will be expressed
- Purpose

Due 9/16.

Homework 5 – Script your narrative.

The final project is 10 minutes of presentation and 5 minutes of Q&A.

Without planning, you either won't fill 10 minutes or you will run over.

Develop a script or detailed outline for your presentation.

Include:

- Your interest in the dataset. (Feel free to get creative.)
- How you obtained it.
- How you cleaned it.
- How you organized it.
- How you analyzed it.
- Summarize your compelling findings.
- Draw relevant conclusions.
- A takeaway or call to action.

Remember: You are the domain expert.

Due 9/16