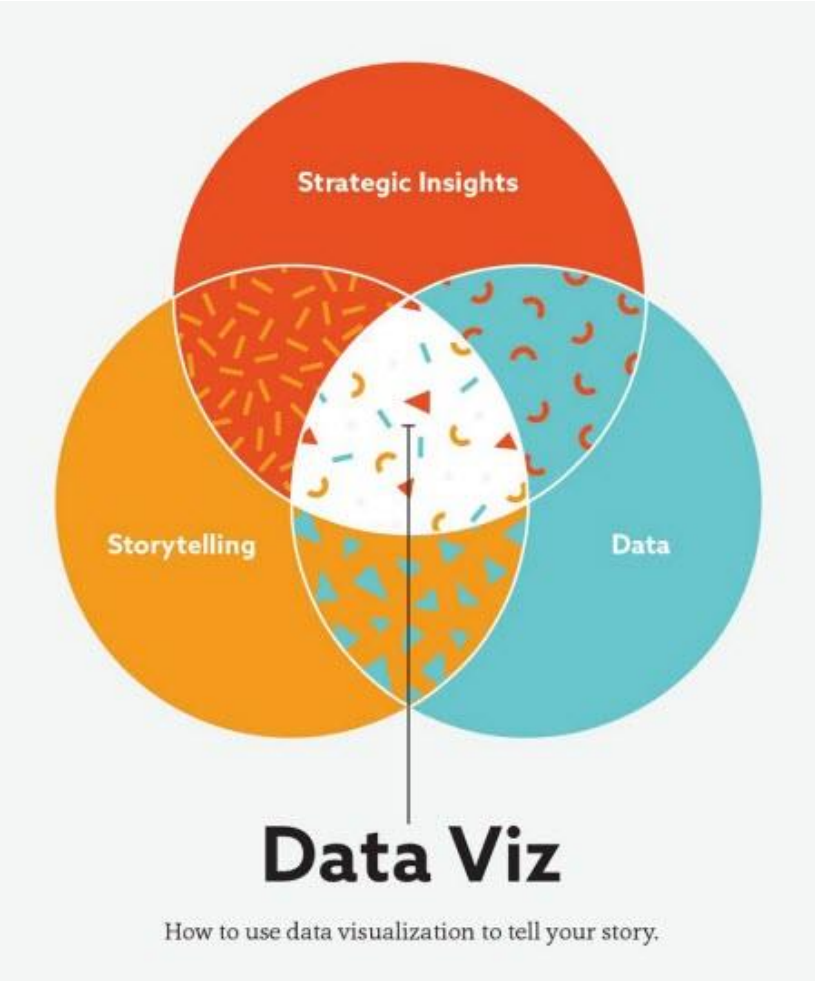
Data visualization: Basic principles

Maria Rivera Araya
mariajose.riveraaraya@jcu.edu.au

Agenda

- Basic steps – How to start?
- Who, why, what
- Colours
- Practices to avoid
- Resources



•Munzner, T. (2015). *Visualization Analysis and Design*. Boca Raton, FL: A K Peters/CRC Press.
•Textbook (Knaflic, 2015) – Introduction (pp. 1–17). Reference: Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Hoboken, NJ: Wiley.

How to start?

Step 1 Who?

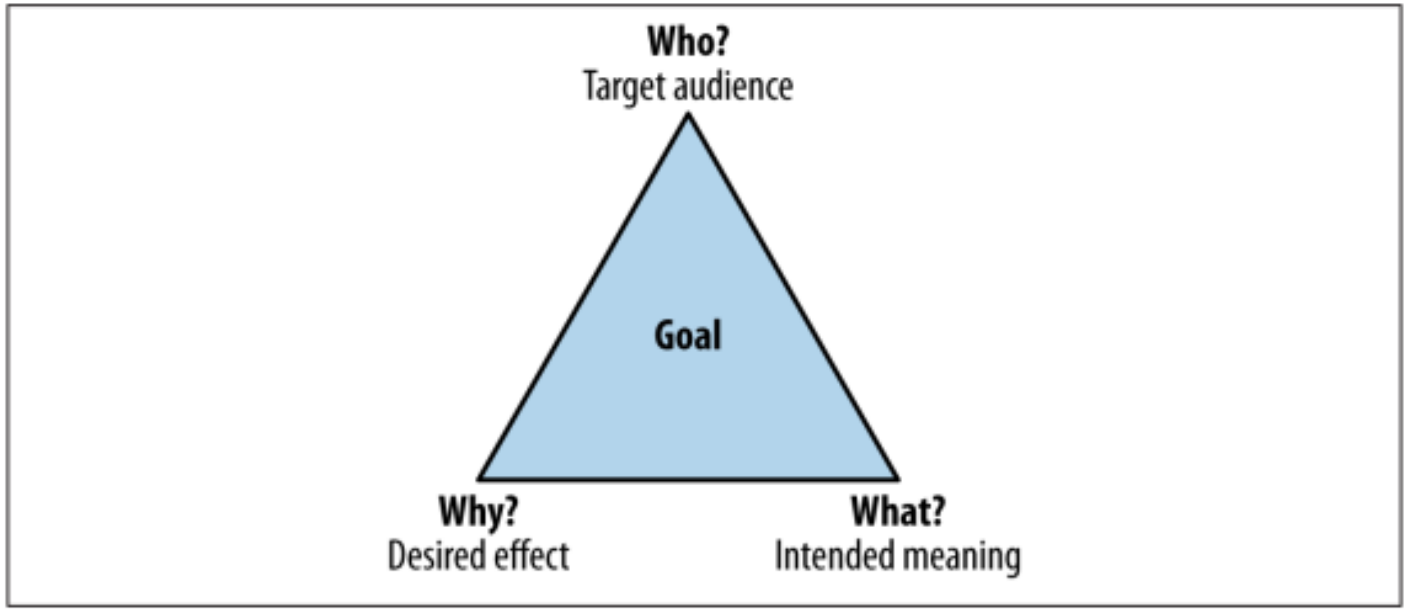


Figure 1-4. Elements of the goal

Identify target audience

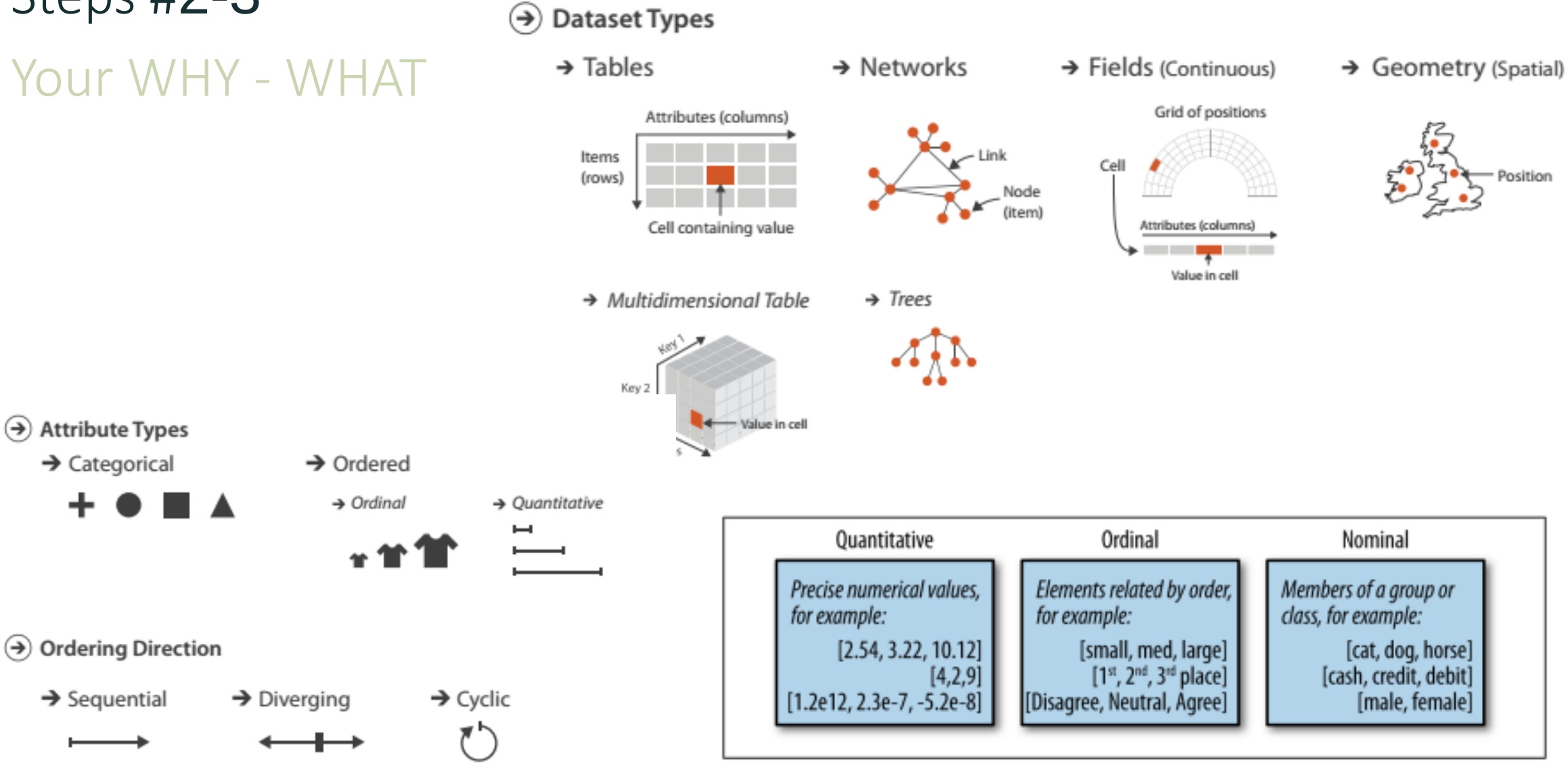
Identify users' questions

Identify users' requirements:

- Previous knowledge
- Vocabulary
- Expectations
- Assumptions

Steps #2-3

Your WHY - WHAT



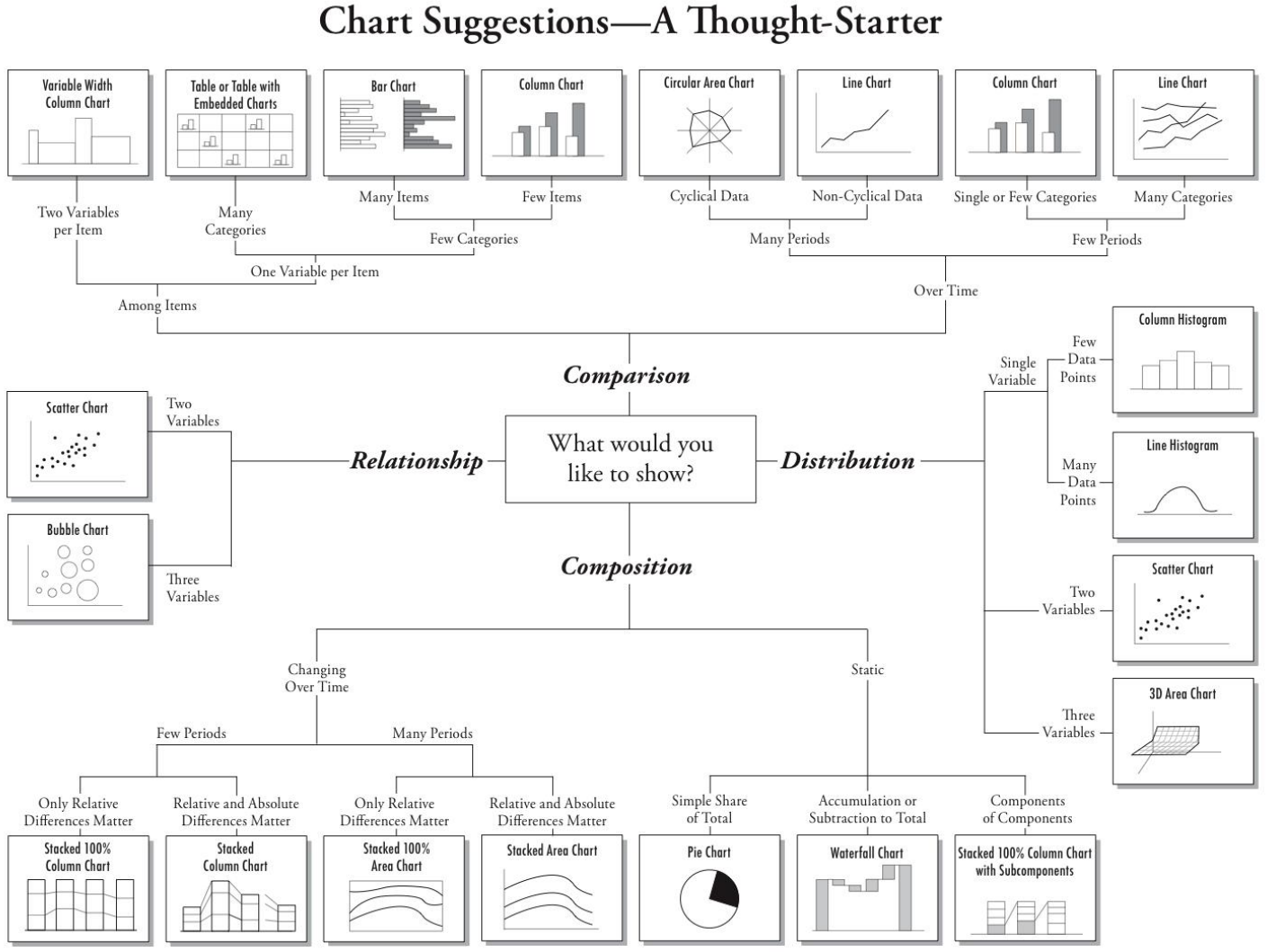
Quantitative	Ordinal	Nominal
Precise numerical values, for example: [2.54, 3.22, 10.12] [4,2,9] [1.2e12, 2.3e-7, -5.2e-8]	*Elements related by order, for example:* [small, med, large] [1st, 2nd, 3rd place] [Disagree, Neutral, Agree]	*Members of a group or class, for example:* [cat, dog, horse] [cash, credit, debit] [male, female]

Figure 1-5. Different types of data

Figure 2.1. *What* can be visualized: data, datasets, and attributes.

Figure 3.1. *Why* people are using vis in terms of actions and targets.

Step 4 HOW?



http://www.infographicsblog.com/chart-suggestions-a-thought-starter-andrew-abela/

How do we build a graph?

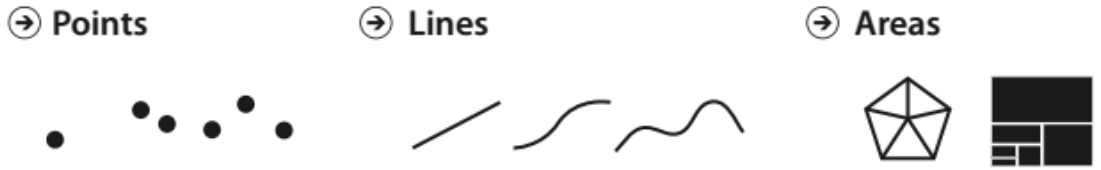
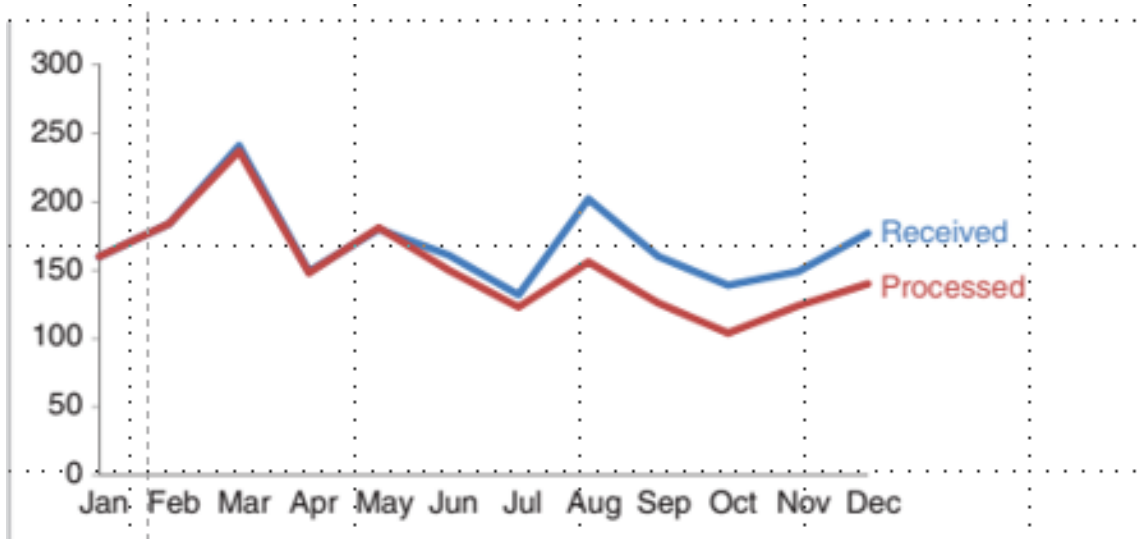


Figure 5.2. Marks are geometric primitives.



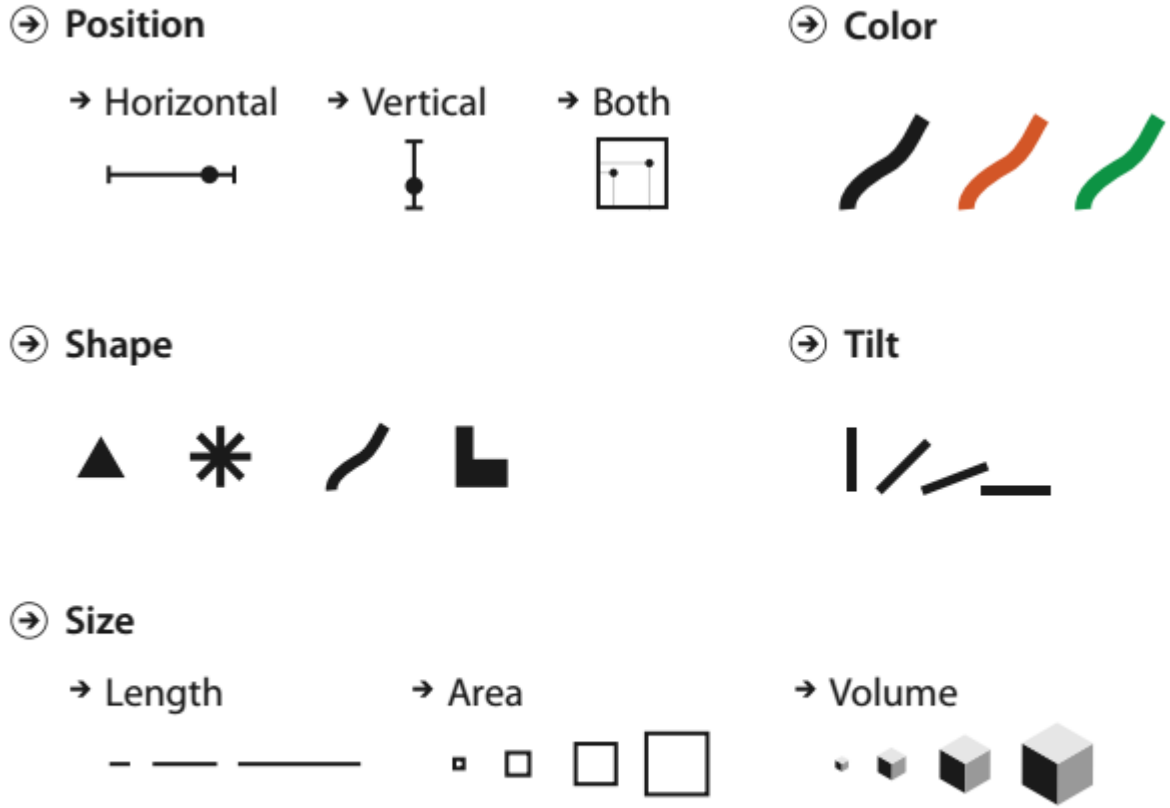


Figure 5.3. Visual channels control the appearance of marks.

Human perception can quickly compare objects and see that one is greater than, equal to or less than the size of another object, but it is very difficult (or takes a long time) for perception to determine the magnitude of difference to any degree of precision

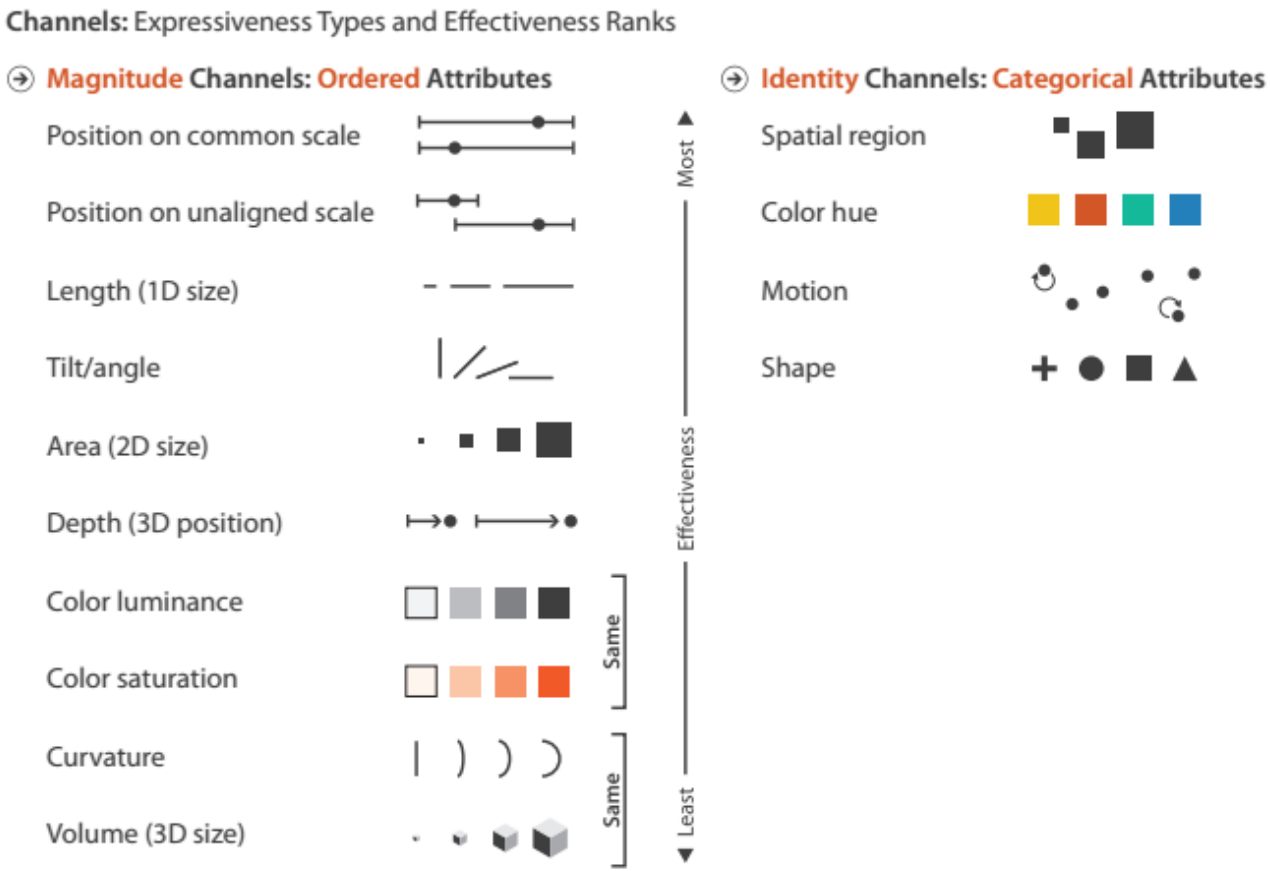
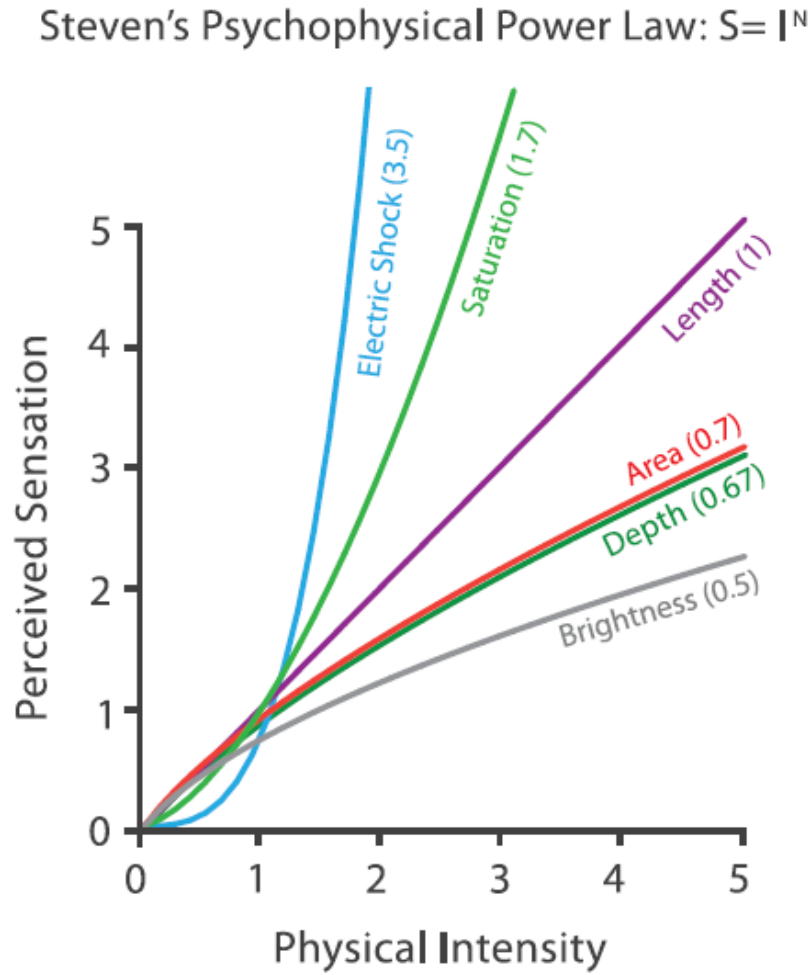


Figure 5.6. Channels ranked by effectiveness according to data and channel type. Ordered data should be shown with the magnitude channels, and categorical data with the identity channels.



Steps 5-7 CHECK

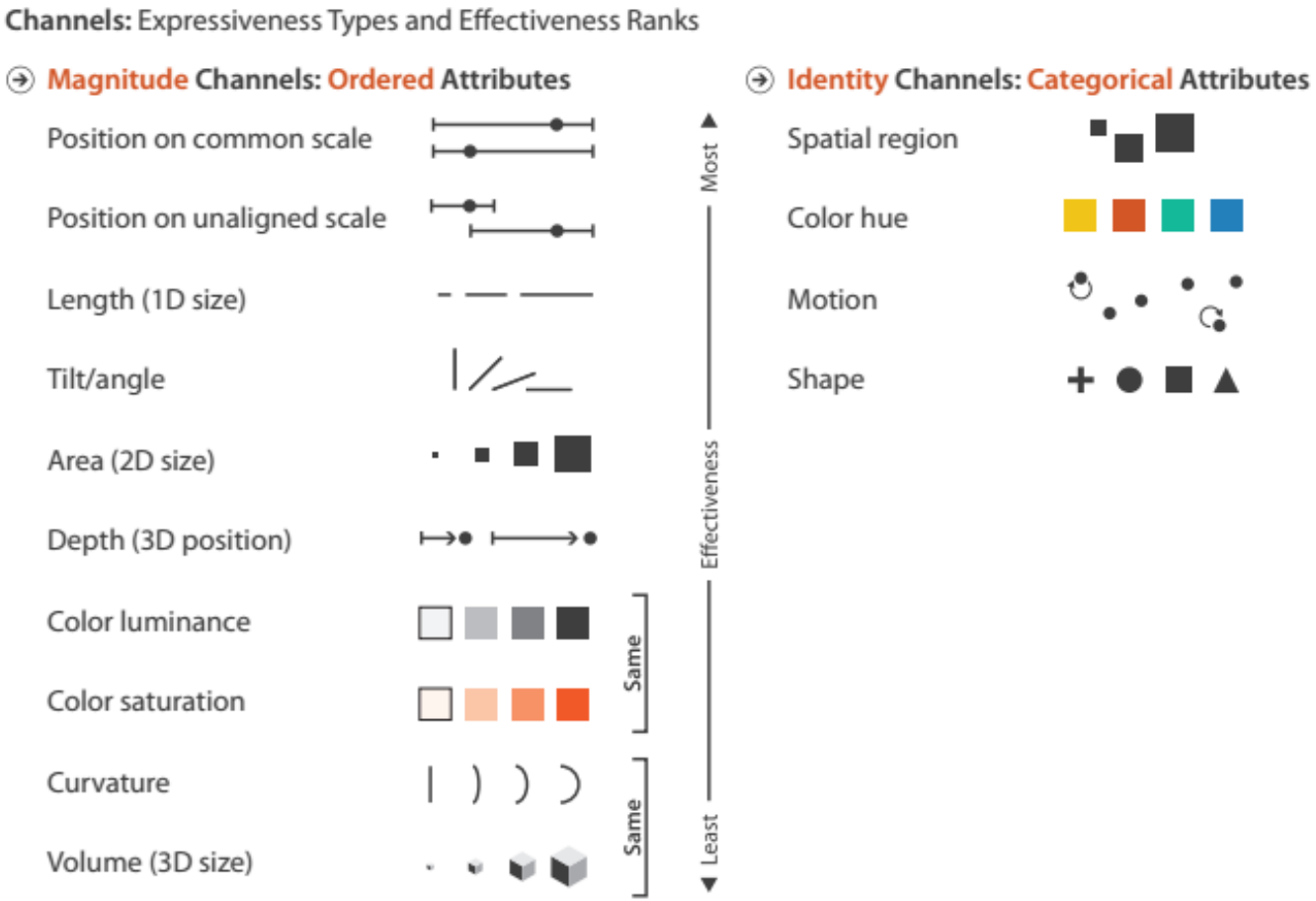


Figure 5.6. Channels ranked by effectiveness according to data and channel type. Ordered data should be shown with the magnitude channels, and categorical data with the identity channels.

- Effectiveness of chosen channels
- Use of 3D graphs
- Clutter
- White Space
- Message – Message – Message
- Can we infer the question by just looking at the graph?

Colour

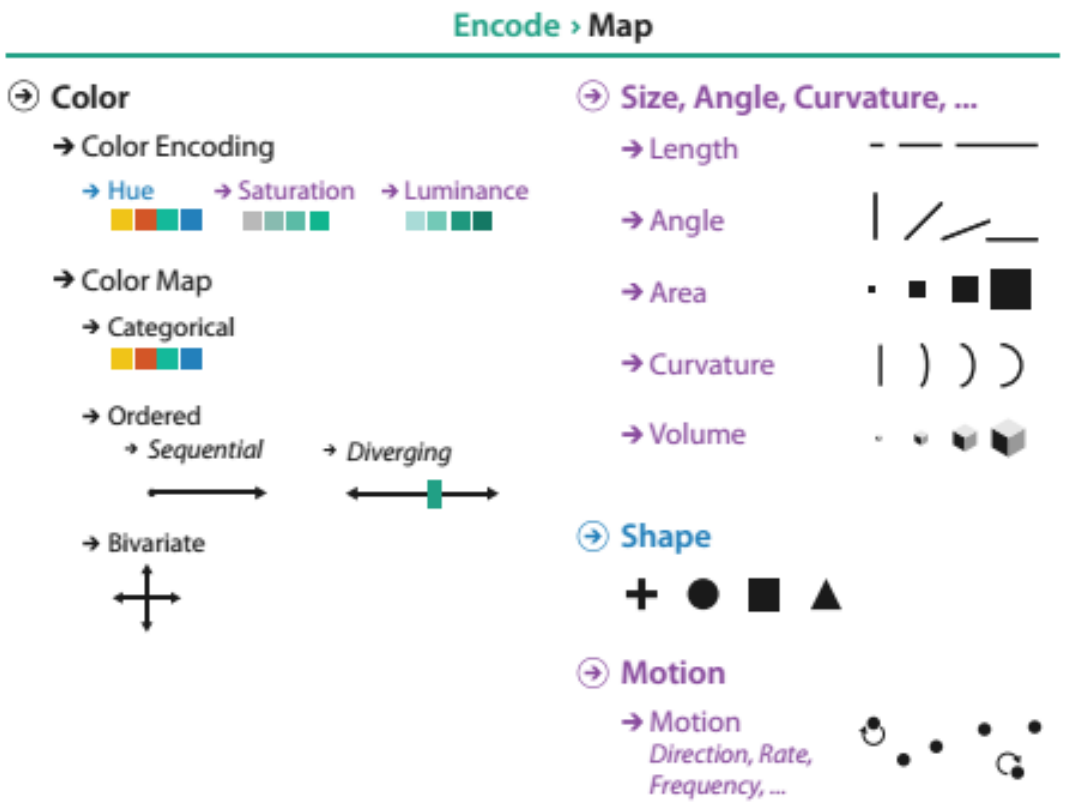


Figure 10.1. Design choices for mapping color and other visual encoding channels.

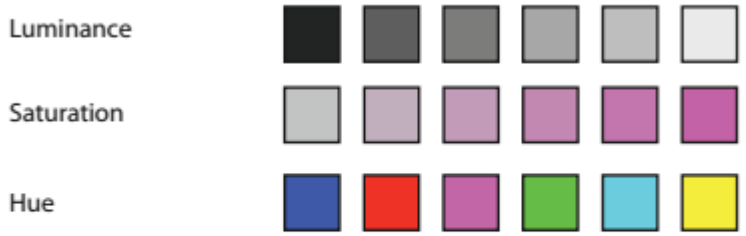


Figure 10.5. The luminance and saturation channels are automatically interpreted as ordered by our perceptual system, but the hue channel is not.

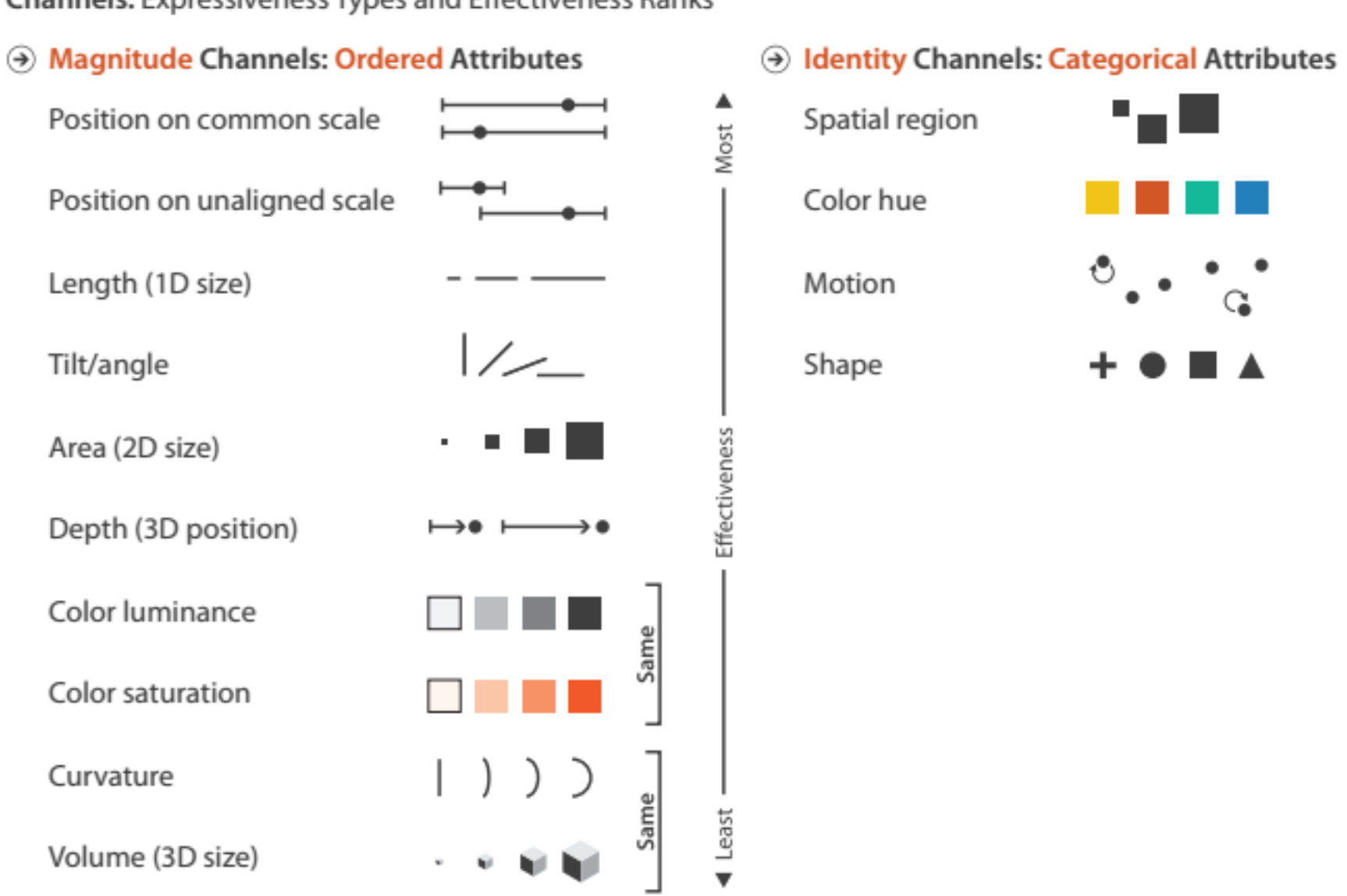
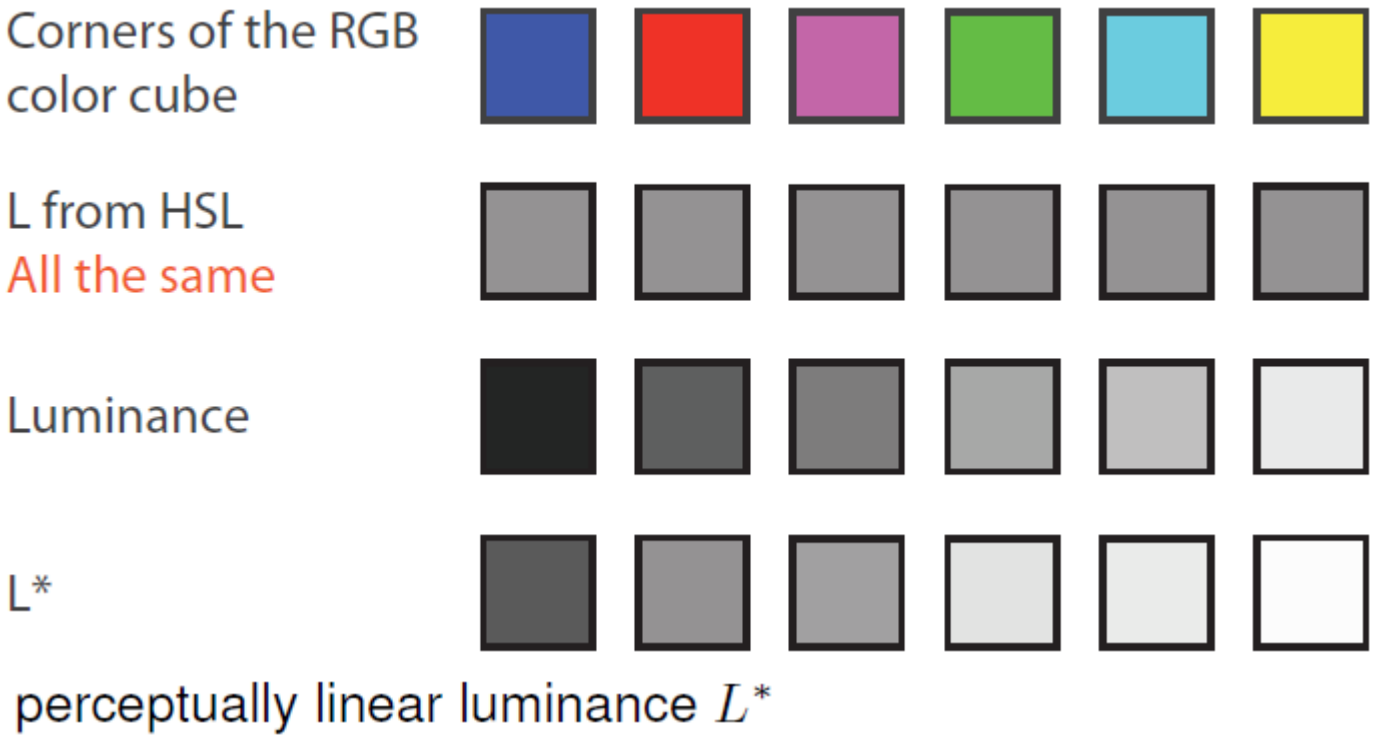


Figure 5.6. Channels ranked by effectiveness according to data and channel type. Ordered data should be shown with the magnitude channels, and categorical data with the identity channels.

Good colour maps

HOW DO WE SEE?

Red to Green
Blue to Yellow
Black to White



http://colorbrewer2.org/#

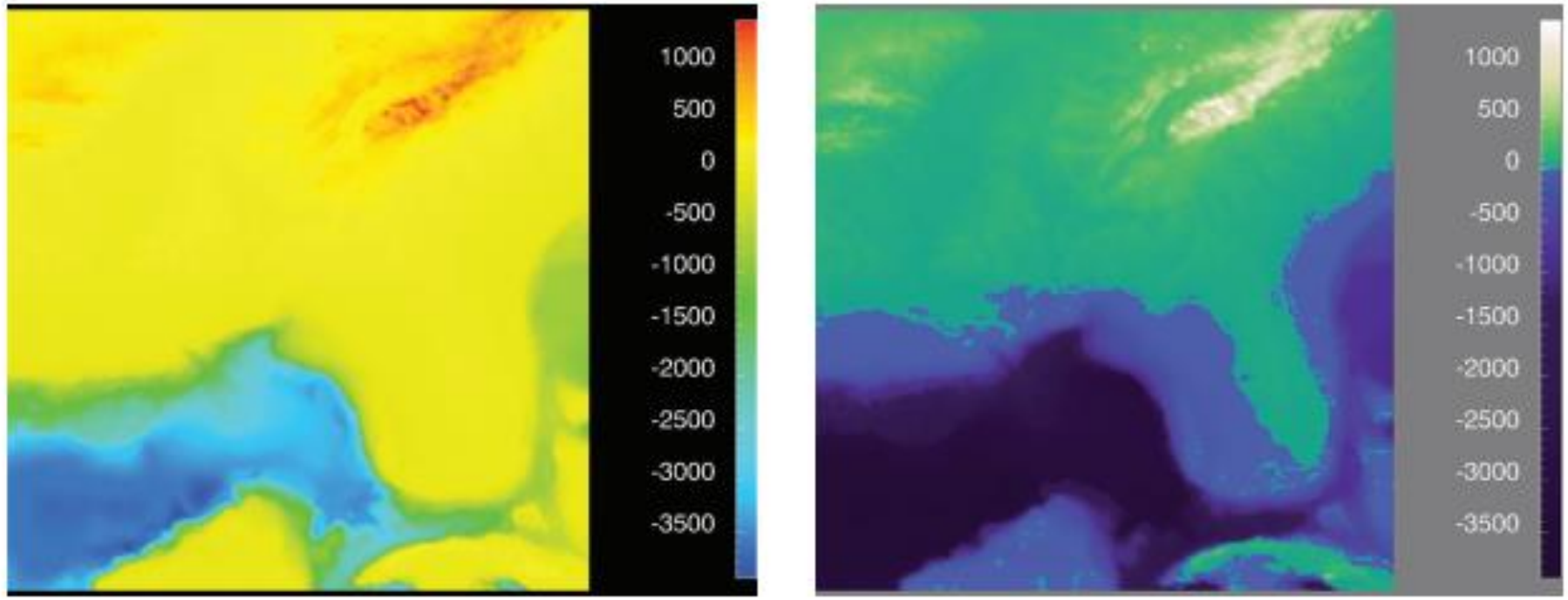


Figure 10.12. Rainbow versus multiple-hue continuous colormap with monotonically increasing luminance. (a) Three major problems with the common continuous rainbow colormap are perceptual nonlinearity, the expressivity mismatch of using hue for ordering, and the accuracy mismatch of using hue for fine-grained detail. (b) A colormap that combines monotonically increasing luminance with multiple hues for semantic categories, with a clear segmentation at the zero point, succeeds in showing high-level, mid-level, and low-level structure. From [Rogowitz and Treinish 98, Figure 1].

Spatial data

Arrange Spatial Data

- **Use Given**
  - Geometry
    - *Geographic*
    - *Other Derived*
  - Spatial Fields
    - *Scalar Fields (one value per cell)*
      - *Isocontours*
      - *Direct Volume Rendering*
    - *Vector and Tensor Fields (many values per cell)*
      - *Flow Glyphs (local)*
      - *Geometric (sparse seeds)*
      - *Textures (dense seeds)*
      - *Features (globally derived)*

Figure 8.1. Design choices for using given spatial data: geometry or spatial fields.

Spatial data

GEOMETRY

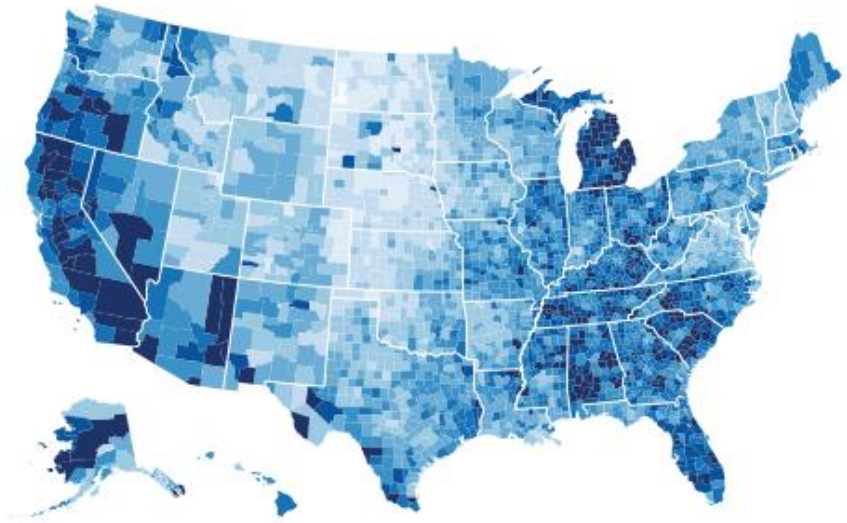


Figure 8.2. Choropleth map showing regions as area marks using given geometry, where a quantitative attribute is encoded with color. From http://bl.ocks.org/mbostock/4060606.

SCALAR/VECTORS/TENSOR

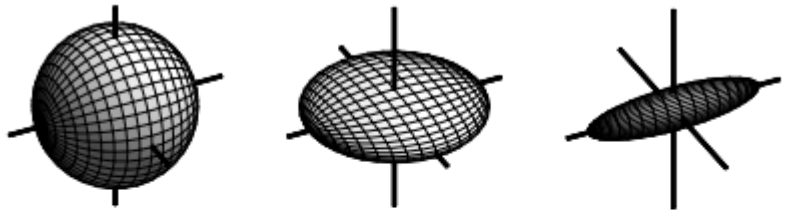


Figure 8.11. Ellipsoid glyphs can show three basic shapes. (a) Isotropic: sphere. (b) Partially anisotropic: planar. (c) Fully anisotropic: linear. From [Kindlmann 04, Figure 1].

Network data

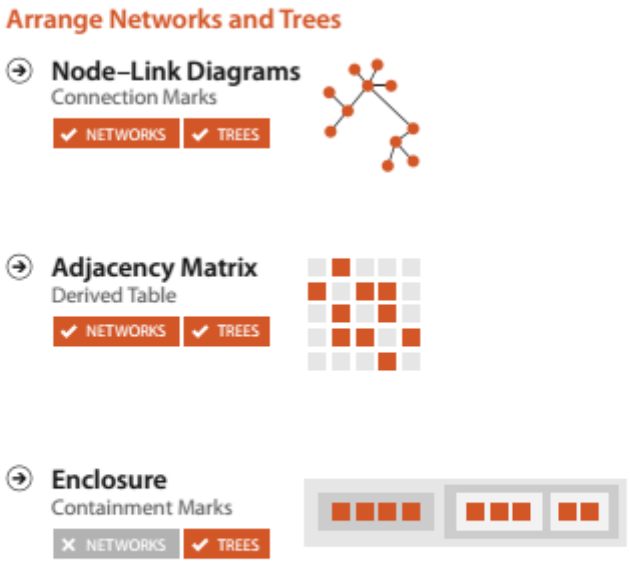


Figure 9.1. Design choices for arranging networks.

(a) (b)

Figure 9.2. Node–link layouts of small trees. (a) Triangular vertical for tiny tree. From [Buchheim et al. 02, Figure 2d]. (b) Spline radial layout for small tree. From http://mbostock.github.com/d3/ex/tree.html.

Facet (split)/multiple views

JUXTAPOSE

Figure 12.3. Overview–detail example with geographic maps, where the views have the same encoding and dataset; they differ in viewpoint and size. Made with Google Maps, http://maps.google.com.

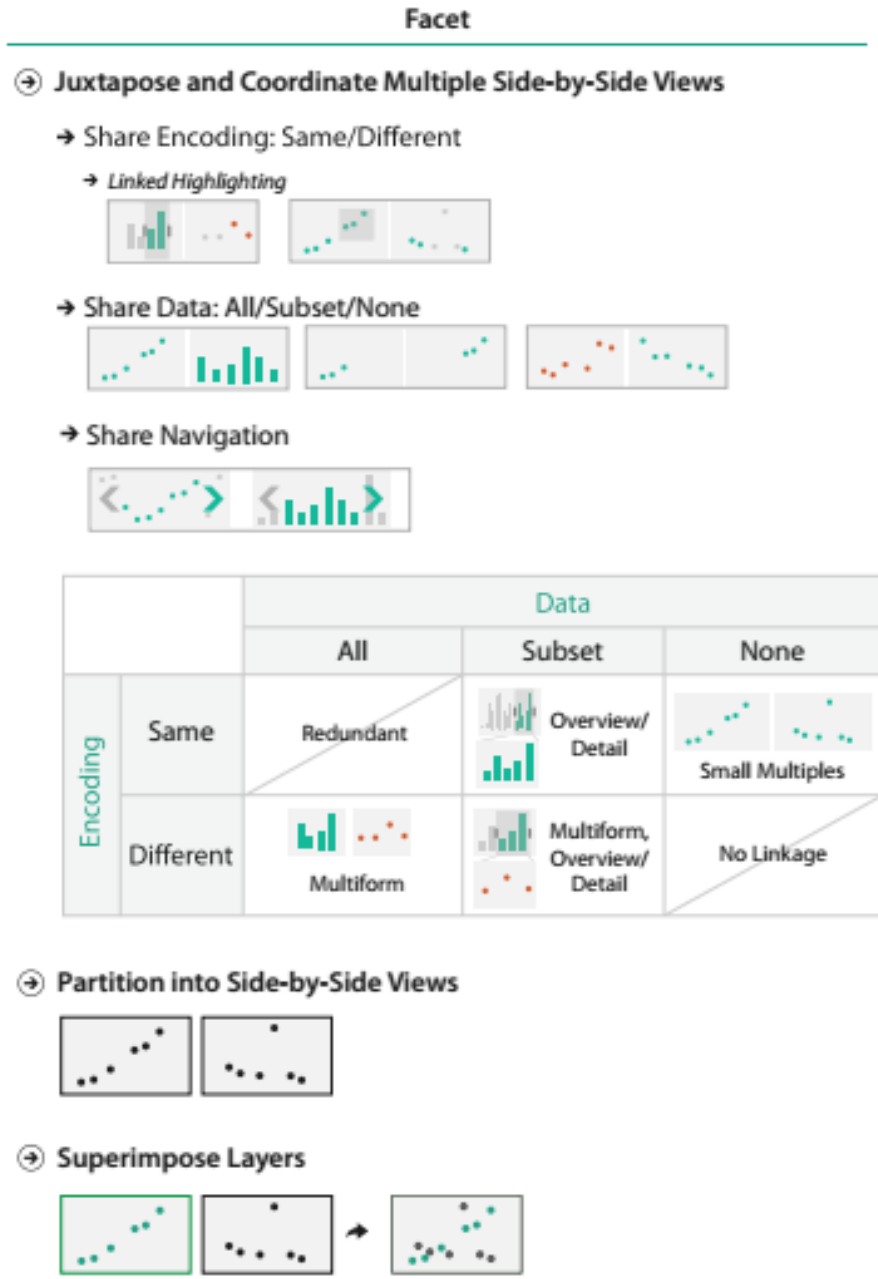


Figure 12.1. Design choices of how to facet information between multiple views.

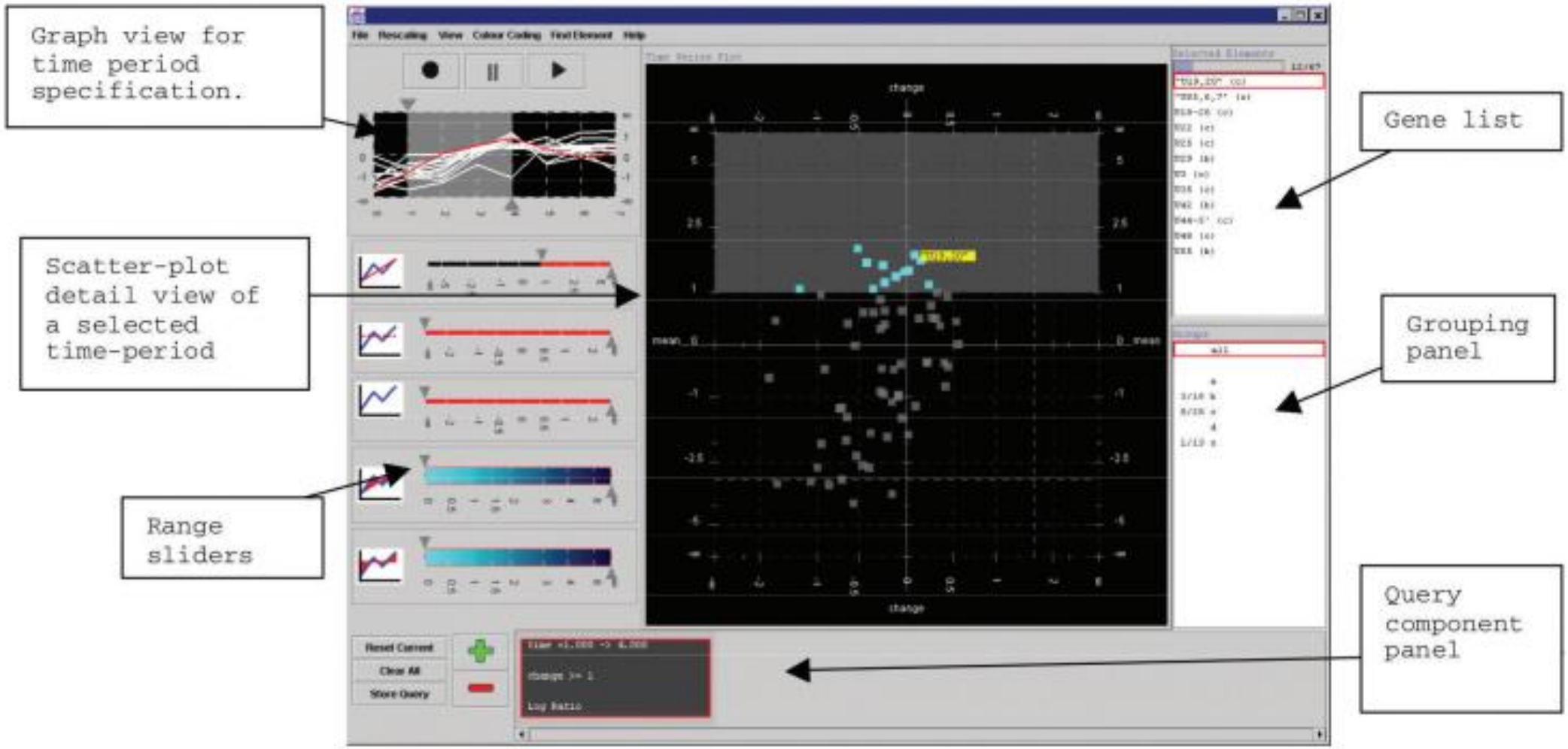


Figure 12.4. Multiform overview–detail vis tool for microarray exploration features a central scatterplot linked with the graph view in the upper left. From [Craig and Kennedy 03, Figure 3].

Partitioning

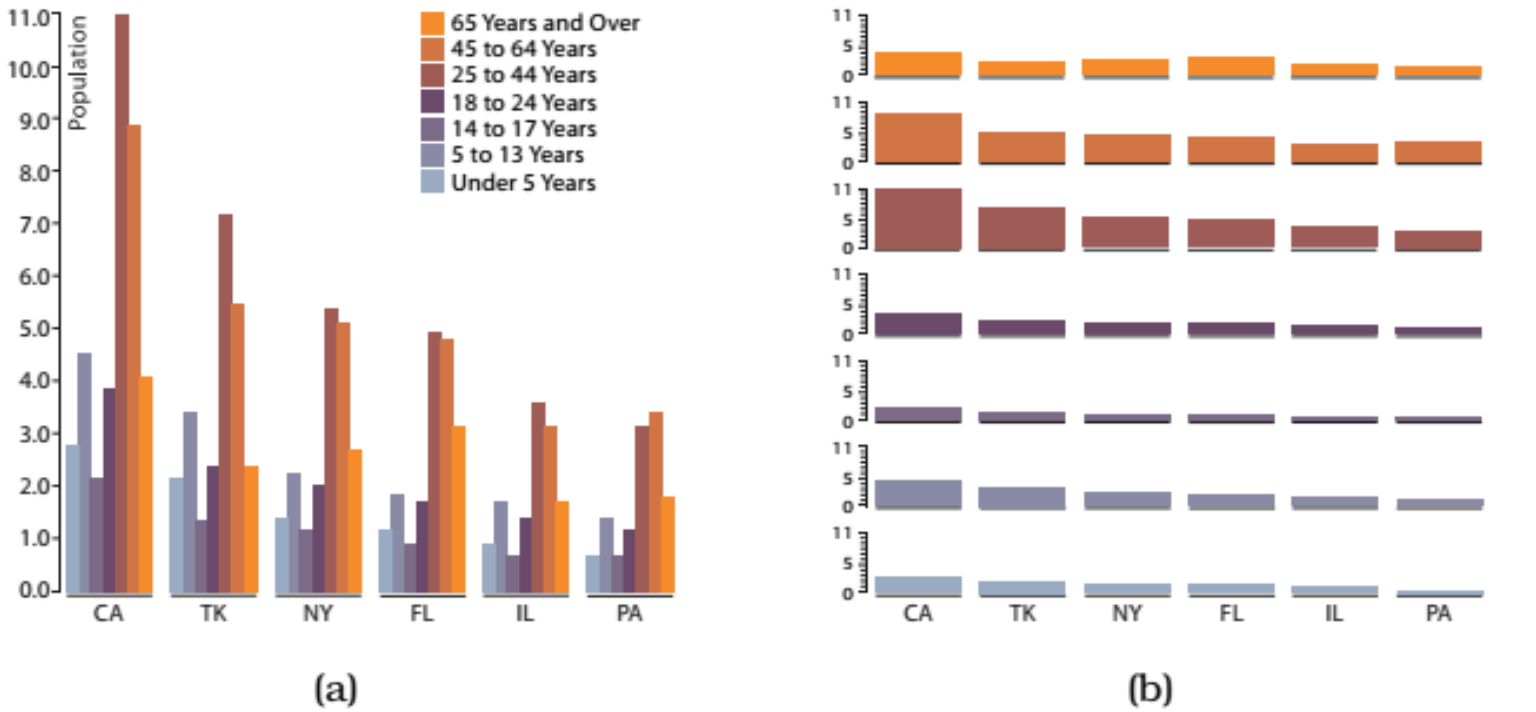
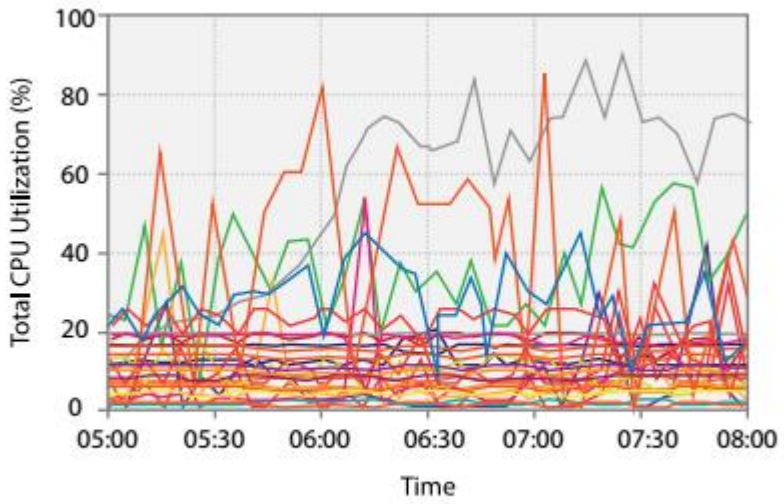


Figure 12.8. Partitioning and bar charts. (a) Single bar chart with grouped bars: separated by *state* key into regions, using seven-mark glyphs within each region. (b) Four aligned small-multiple bar chart views: separated by *group* key into vertically aligned list of regions, with a full bar chart in each region. From http://bl.ocks.org/mbostock/3887051, after http://bl.ocks.org/mbostock/4679202.

Superimpose



(c)

Reduction/simplification

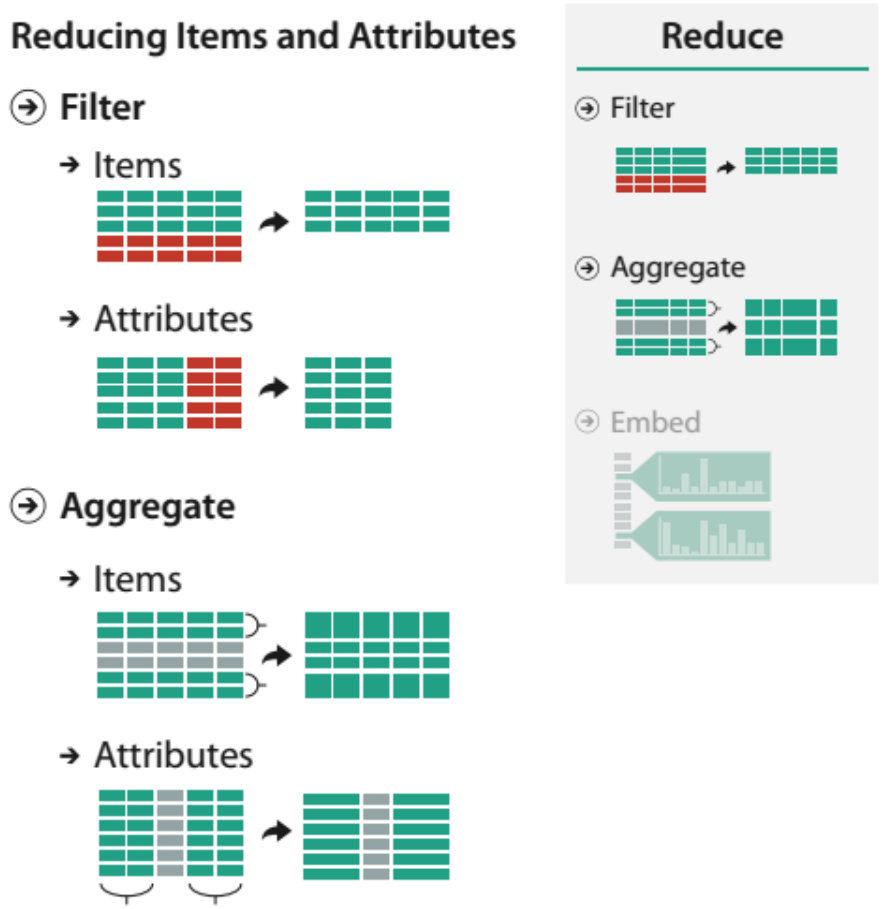


Figure 13.1. Design choices for reducing (or increasing) the amount of data items and attributes to show.

Focus/context

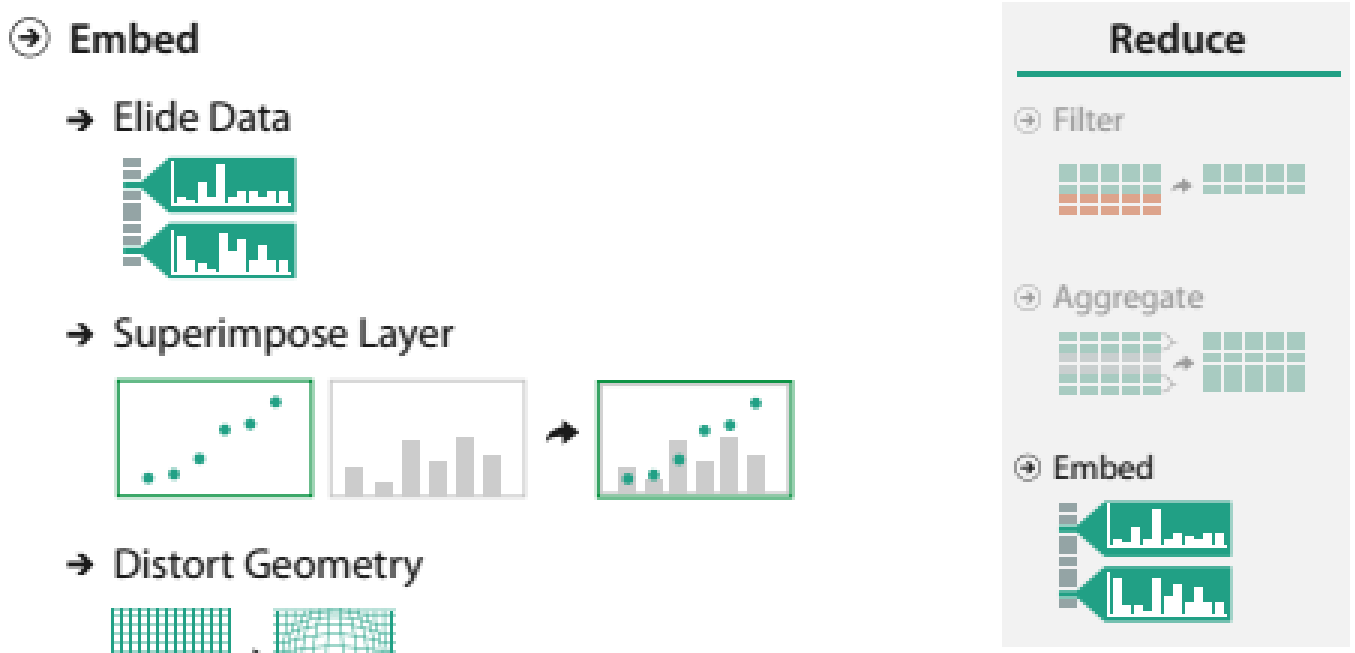


Figure 14.1. Design choices for embedding focus information within context.

Search Engine Marketing

Campaign: PPC | Country: Canada | Year of Date: 2015

interworks

Metric: vs. Goal

Cost per Click	Cost per Conversion	Sales per Conversion	A-Conversion Rate	B-Conversion Rate
+0.01	-0.02	+1.50	1.00%	2.64%

Impressions
262,568,188

Sessions
6,560,690

Conversions
87,177

Click Thru Rate
2.50%

Conversion Rate
1.33%

Impressions to Conversions
0.03%

Inbound Search Queries

Search Query	
soccer socks	189,450
baseball cleats	159,893
goalie mask	137,961
golf club	131,4
hockey stick	102,6
baseball glove	100,2
goalkeeper gloves	99,8
roller blades	92,6
soccer jersey	91,4
bowling ball	85,0
football helmet	81,0
basketball ball	77,6
cleats	58,0
cheerleading	48,5
wrestling shoes	32,0

Visitors over Time POC

Select Measure: Impressions

Conversion % by Time & Weekday

Conversion Rate: 1.17% – 1.

Hour (group)

Weekday of..	12AM	3AM	6AM	9AM	12PM	3PM	6PM	9PM
Sunday	1.41%	1.40%	1.43%	1.39%	1.29%	1.32%	1.23%	1.4
Monday	1.23%	1.35%	1.30%	1.32%	1.29%	1.36%	1.38%	1.2
Tuesday	1.37%	1.35%	1.39%	1.48%	1.35%	1.30%	1.25%	1.2
Wednesday	1.34%	1.42%	1.41%	1.30%	1.34%	1.24%	1.30%	1.2
Thursday	1.27%	1.39%	1.39%	1.29%	1.27%	1.31%	1.33%	1.3
Friday	1.25%	1.39%	1.38%	1.40%	1.42%	1.37%	1.21%	1.4
Saturday	1.20%	1.34%	1.27%					1.1

Hour (group): 6AM
Weekday of Date: Saturday
Conversion Rate: 1.27%

+ableau

Search Engine Marketing

PERFORMANCE DASHBOARD

interworks

Campaign: PPC | Country: Canada | Year: 2015

Core Metrics

vs. Goal

A/B Conversion Testing

+0.01	-0.02	+1.50	1.00%	2.64%
Cost per Click	Cost per Conversion	Sales per Conversion	Champion (80%)	Challenger (20%)

Conversion Analysis

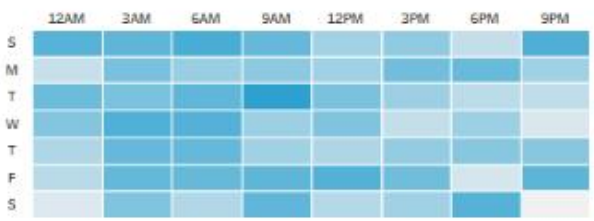
Inbound Search Queries

1	soccer socks	189,450
2	baseball cleats	159,893
3	goalie mask	137,961
4	golf club	131,496
5	hockey stick	102,688
6	baseball glove	100,227
7	goalkeeper gloves	99,897
8	roller blades	92,600
9	soccer jersey	91,447
10	bowling ball	85,037
11	football helmet	81,084
12	basketball ball	77,614
13	cleats	58,014
14	cheerleading	48,562
15	wrestling shoes	32,052

Weekly Performance

Impressions

Conversion % by Time & Weekday



+ableau

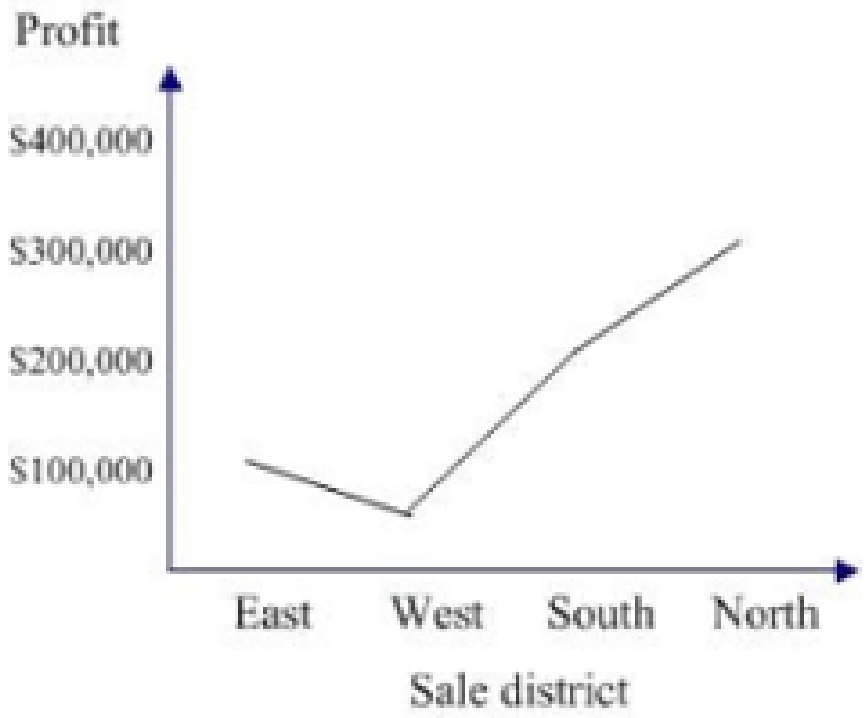
Removing the clutter!

Remove the Clutter

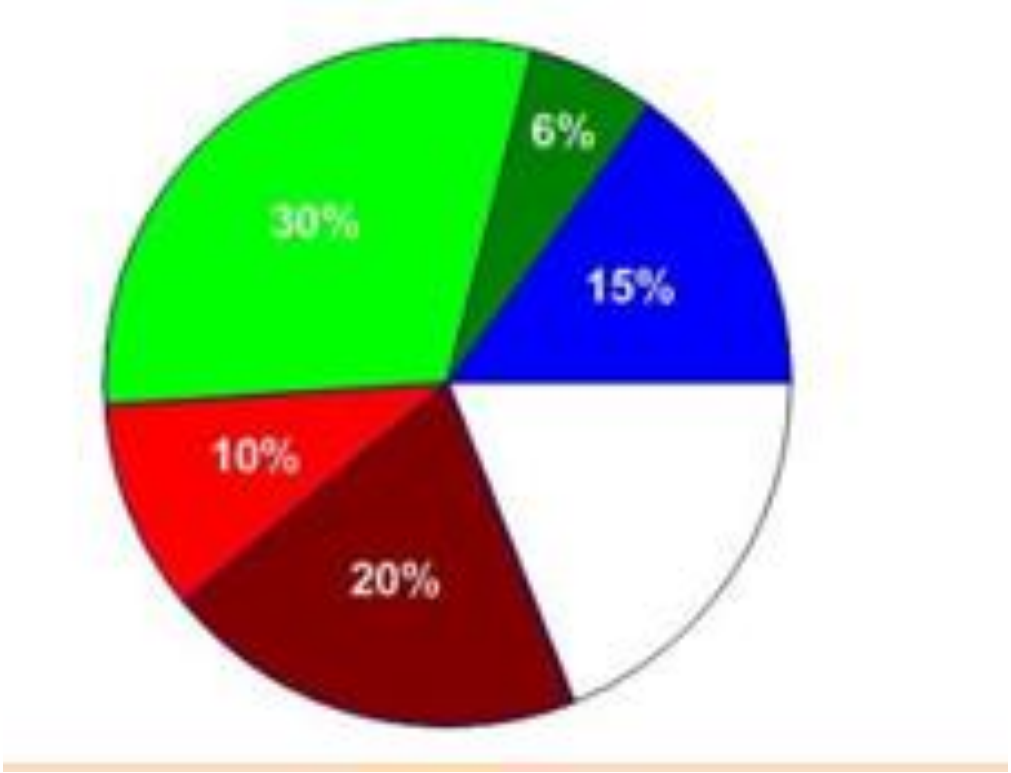
There's a lot of clutter here that I can remove. I'll show these changes as a bulleted list:

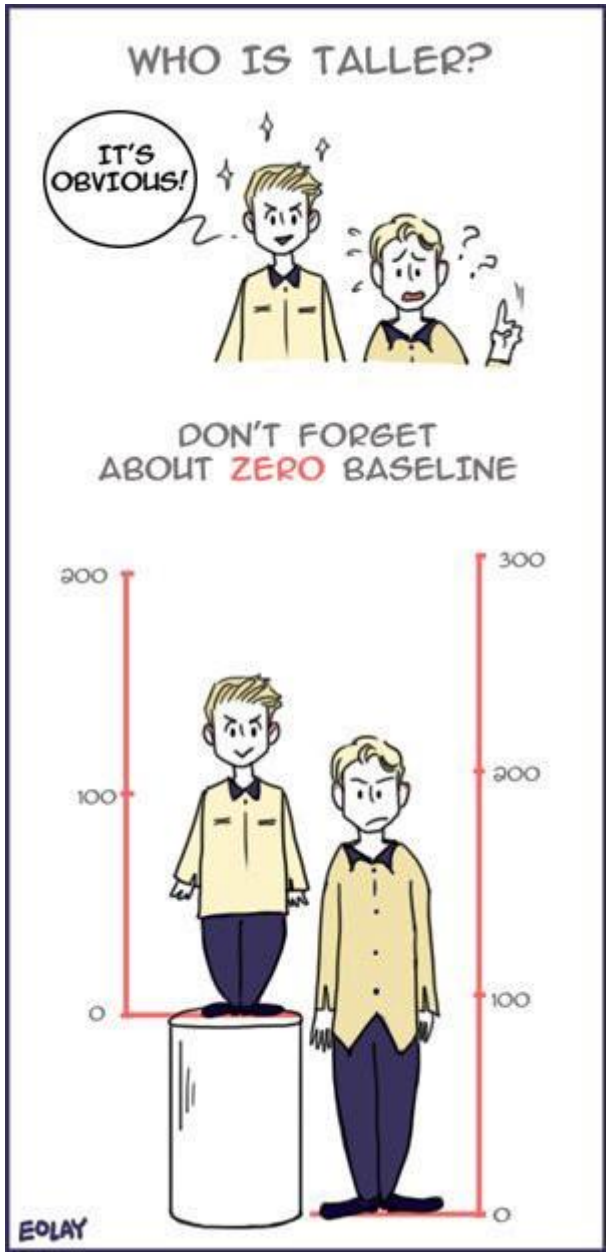
- **Reduce** the font size of the titles for my measure totals represented by a single worksheet
- **Move** labels as needed for emphasis on the numbers
- **Remove** chart borders
- **Re-label** my weekdays to an abbreviated format (i.e. Friday is F)
- **Remove** field labels
- **Lighten** my text across the entire chart
- **Remove** all % labels from my Conversion % by Time & Weekday and **convert** it from a highlight table to a heat map
- **Remove** the color legend
- **Change** "Year of Date" on the filter to simply "Year"

Practices to avoid

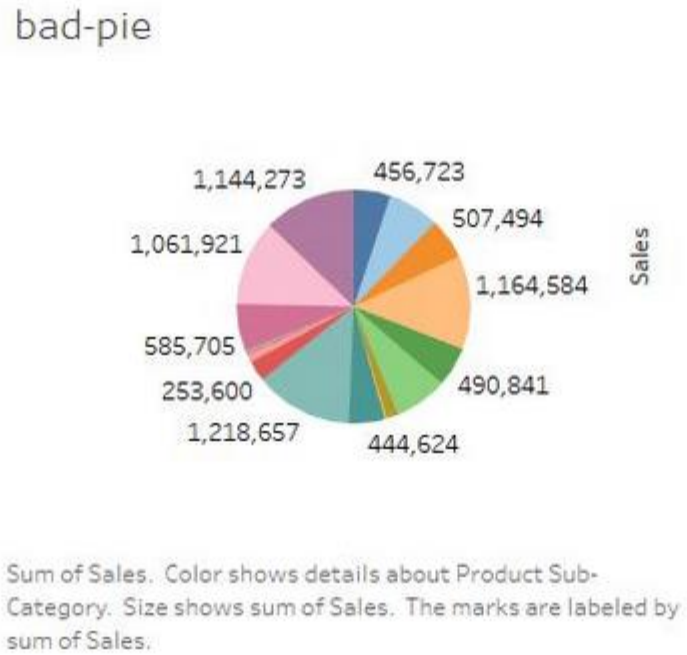


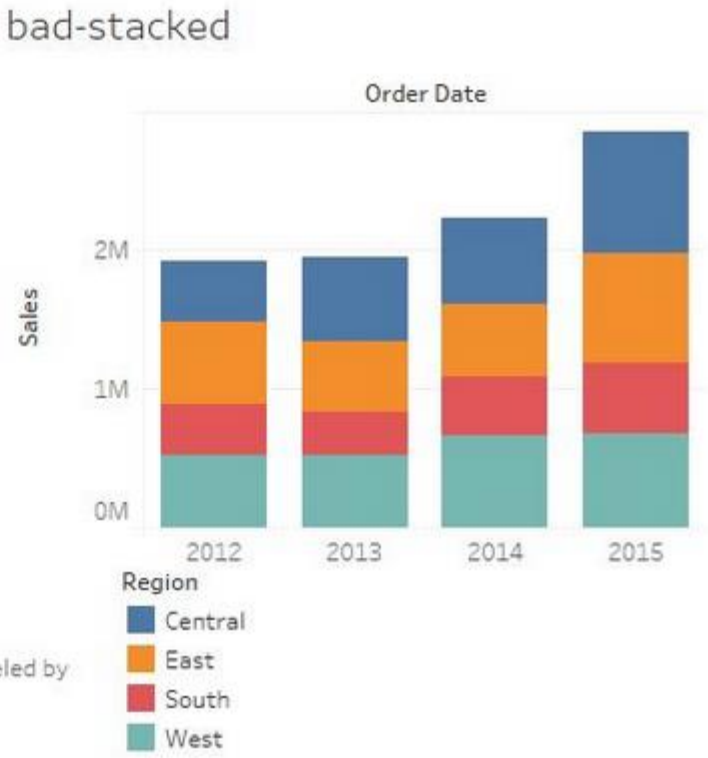
Practices to avoid





Practices to avoid





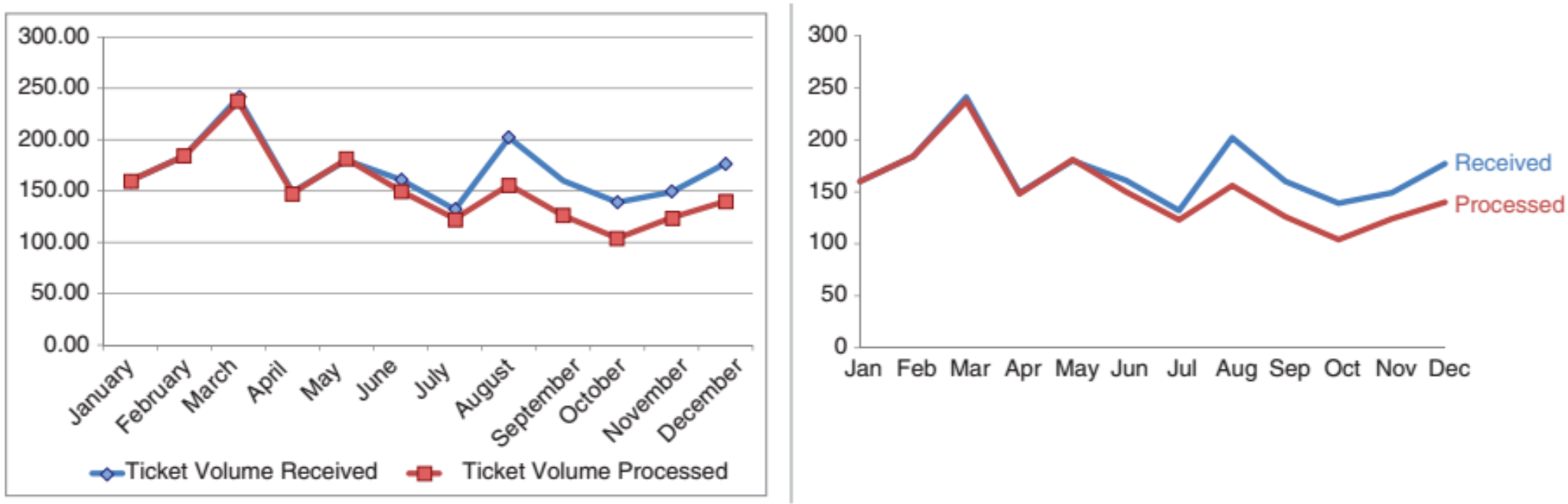
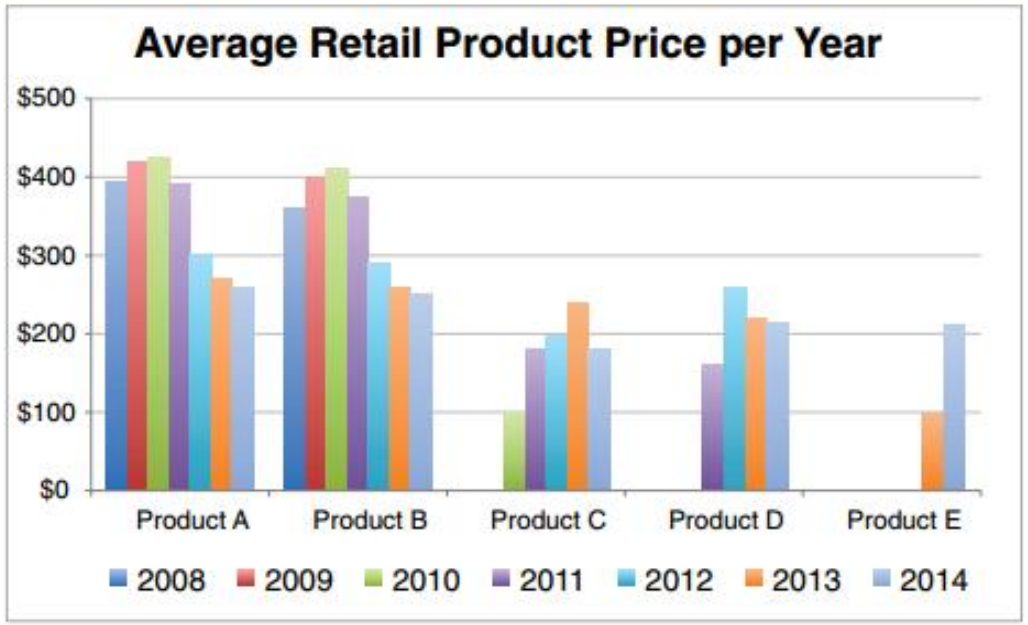


FIGURE 3.24 Before-and-after

Price has declined for all products on the market since the launch of Product C in 2010



To be competitive, we recommend introducing our product *below the $223 average* price point in the **$150–$200 range**

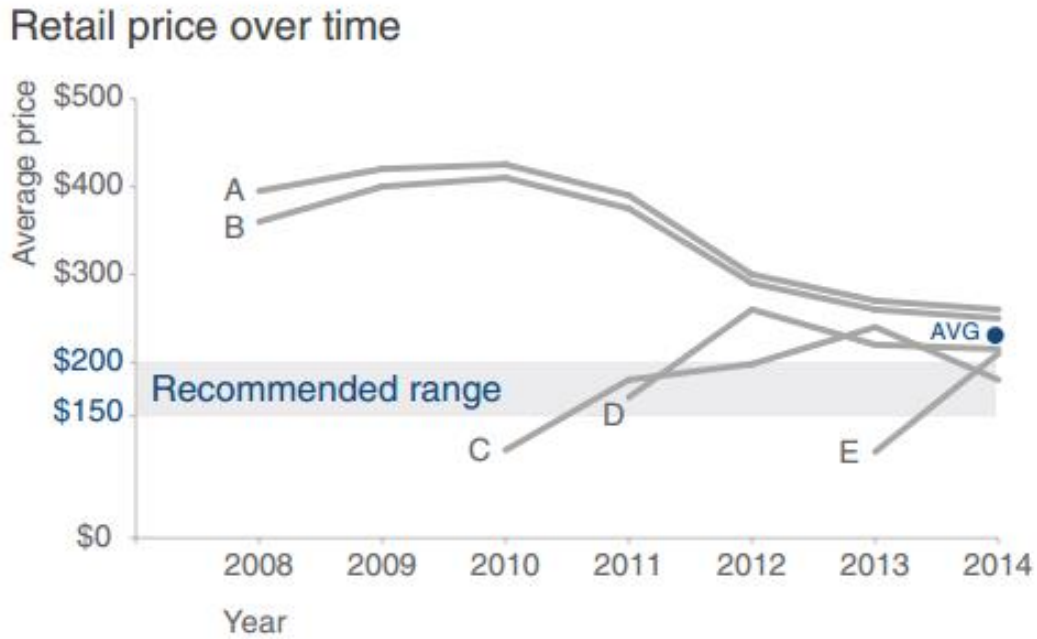


FIGURE 8.20 Before-and-after

Other resources

Linkedin Learning

https://www.jcu.edu.au/library/learn/linkedinlearning

https://www.jcu.edu.au/learn/news/studiosity

https://www.visualisingdata.com/2017/07/best-visualisation-web-june-2017/

http://www.bigbookofdashboards.com/dashboards.html

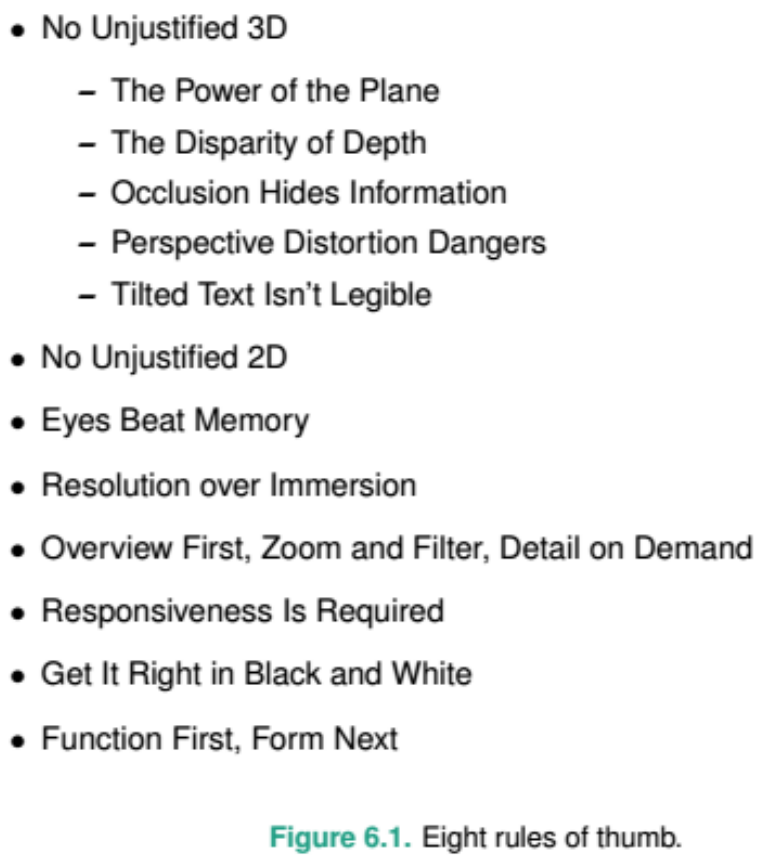
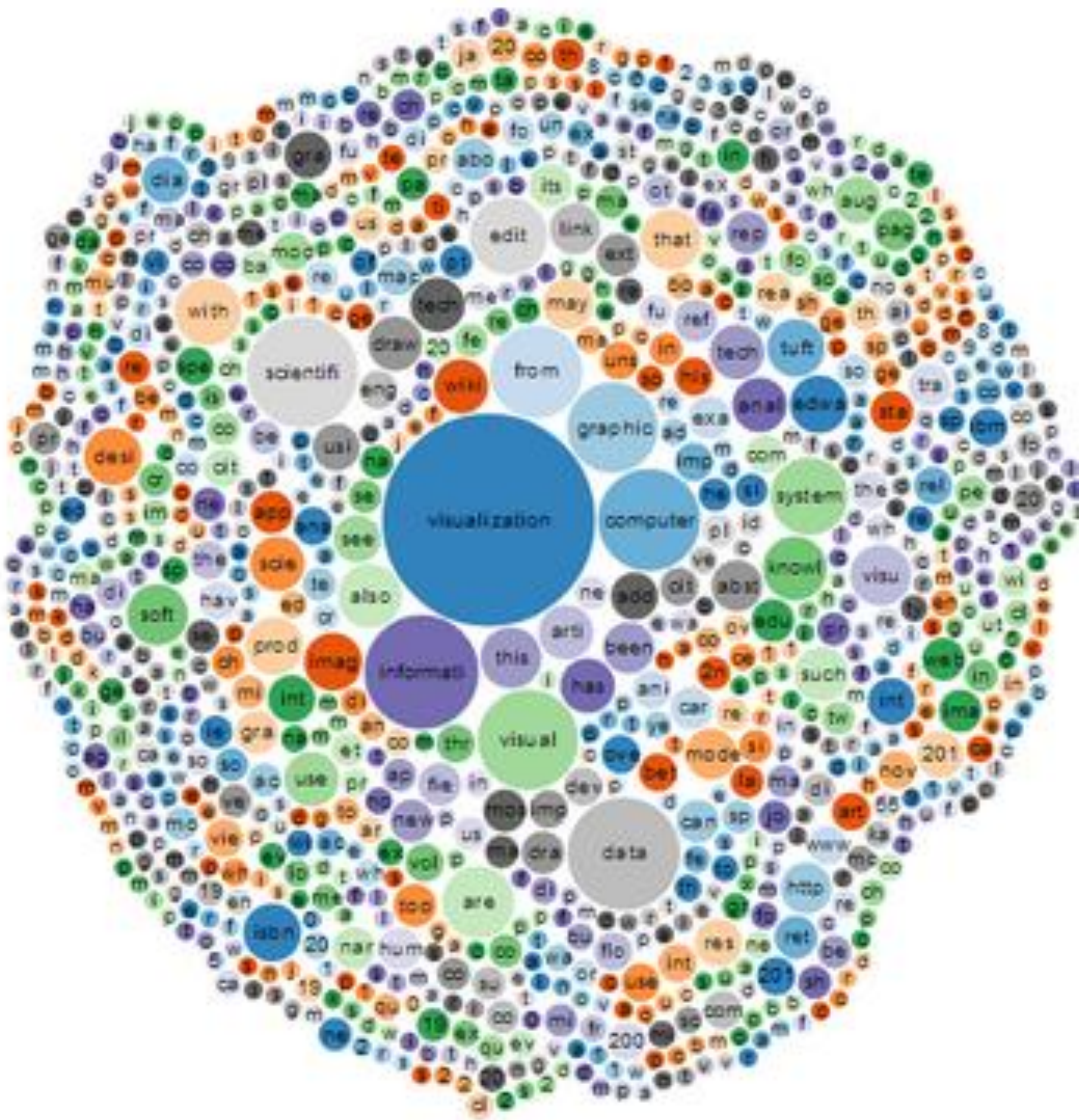


Figure 6.1. Eight rules of thumb.

Thanks!!!


visualization
informati
computer
graphic
scientifi
from
data
visual
system
edit
with
this
also