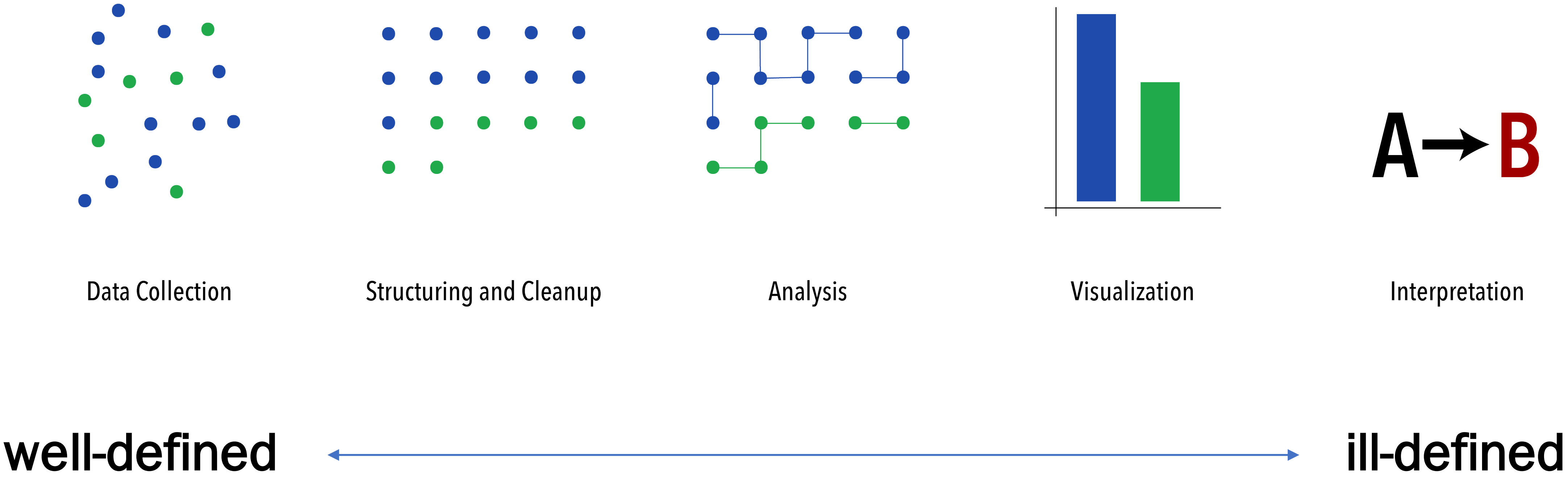
Data Visualization Review

Steven Braun

Data Analytics and Visualization Specialist

May 10, 2017


A→B
Data Collection
Structuring and Cleanup
Analysis
Visualization
Interpretation
well-defined
ill-defined

Why do we visualize data?

To communicate information

To point attention

To tell a story

THREE AXIOMS OF GOOD VISUALIZATION DESIGN

1	Data visualization is intuitive
2	Data visualization engages and invites the viewer to ask more questions
3	Data visualization facilitates discovery

Effective Design

Ethical Design

Universal Design

ELEMENTS OF A VISUALIZATION

Marks

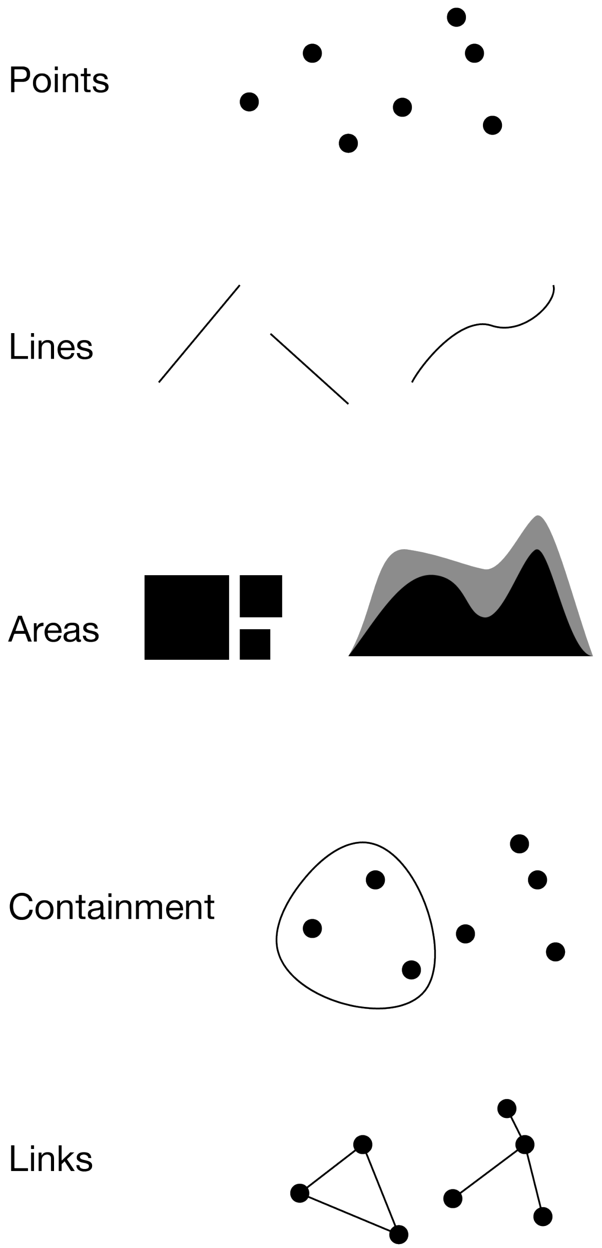
Basic graphical elements

Channels

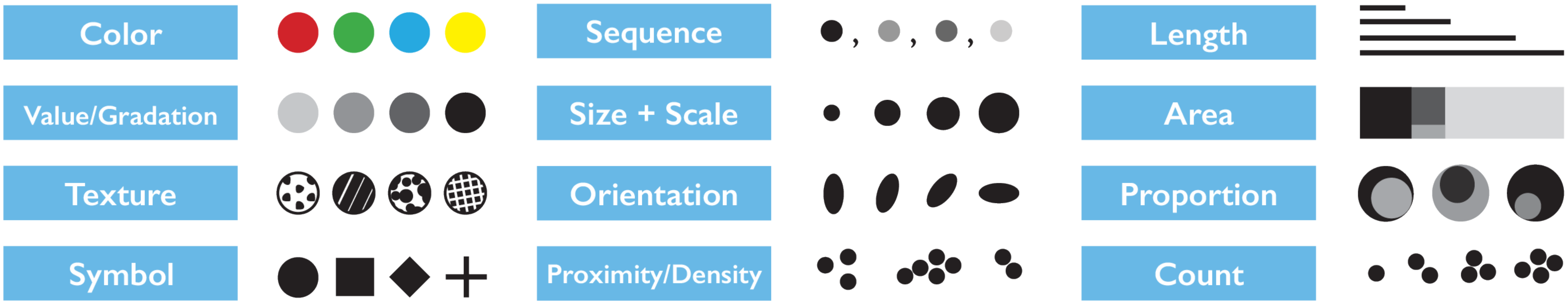
Ways to control mark appearance

Attributes

Dimensions of data


Points
Lines
Areas
Containment
Links

CHANNELS


Color
Sequence
Length
Value/Gradation
Size + Scale
Area
Texture
Orientation
Proportion
Symbol
Proximity/Density
Count

Gestalt laws of perceptual organization dominate in fast visual cognition

There are many ways to visually encode information, and Gestalt laws provide a way to leverage them effectively

LAW OF SIMILARITY

Visually similar objects form a group

LAW OF PROXIMITY

Objects close in space form a group

LAW OF CLOSURE

Objects separated in space are connected in our mind

LAW OF PRÄGNANZ

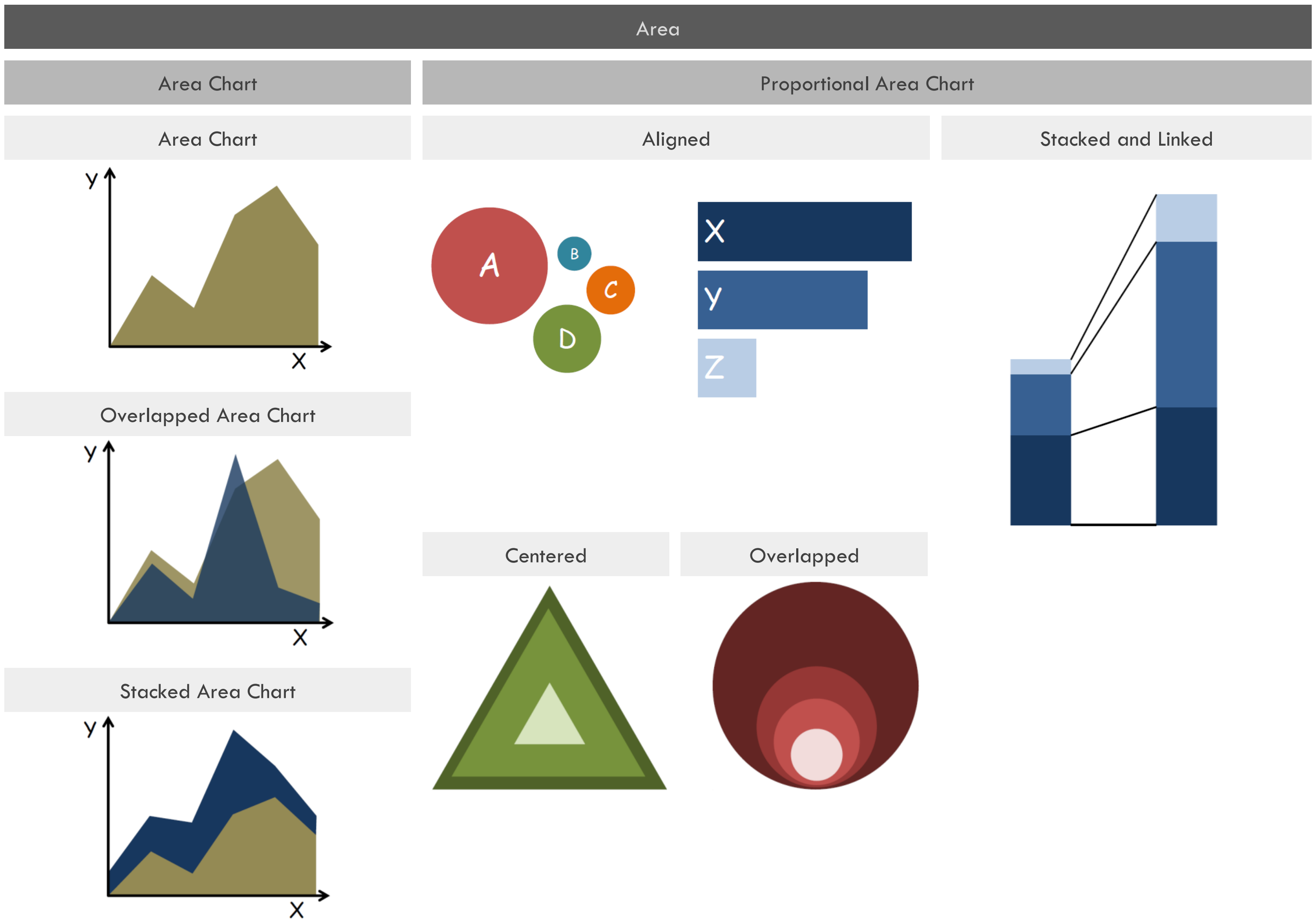
Cognitive load is optimized by reduction to the simplest form

LAW OF ISOMORPHIC CORRESPONDENCE

A viewer’s interpretation of a visualization is informed by past experience

A Taxonomy of Visualizations

Michelle Borkin, Azalea Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, & Hanspeter Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* (Proceedings of InfoVis 2013), 19, 12, 2306-2315.

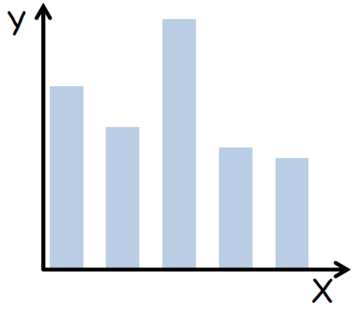


Michelle Borkin, Azalea Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, & Hanspeter Pfister. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* (Proceedings of InfoVis 2013), 19, 12, 2306-2315.

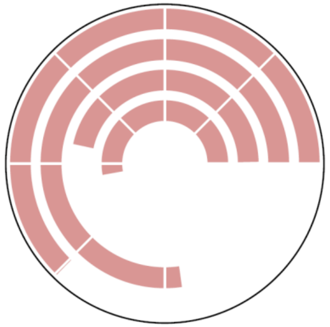
Bar

Bar Chart

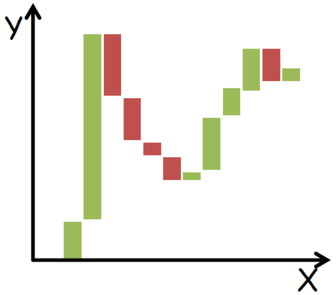
Bar Chart



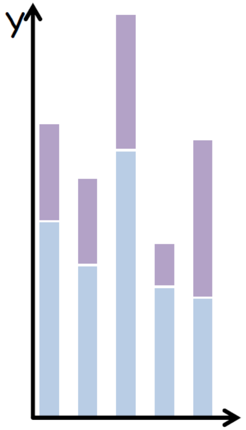
Circular Bar Chart



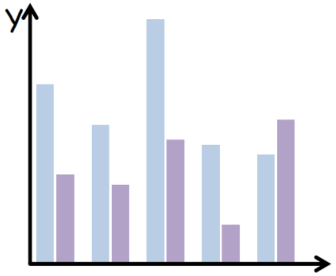
Waterfall Chart



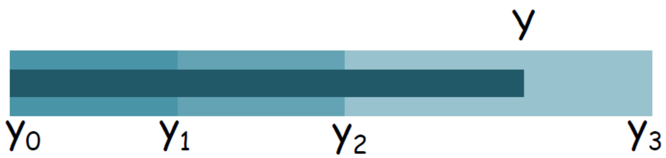
Stacked Bar Chart



Grouped Bar Chart

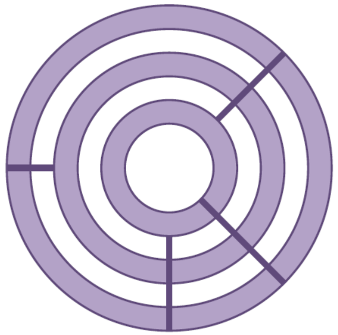


Bullet Graph



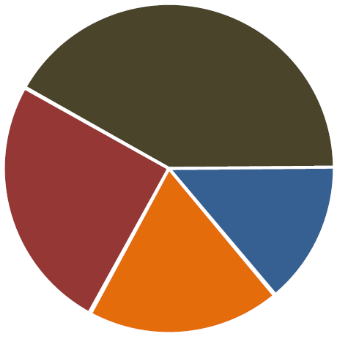
Circle

Belt Chart

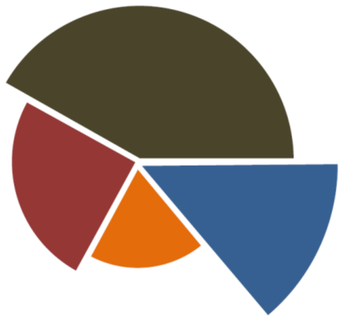


Donut Chart

Pie Chart



Sector Graph



Diagram

Flow Chart

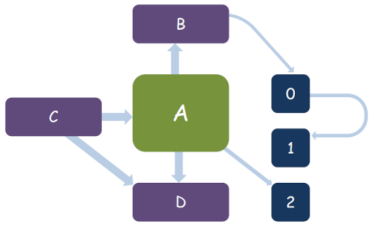
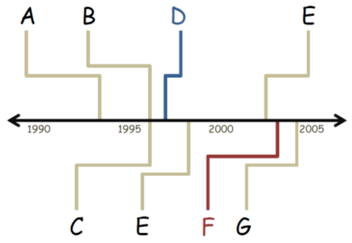


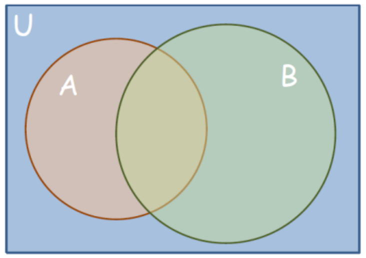
Illustration (or Rendering)

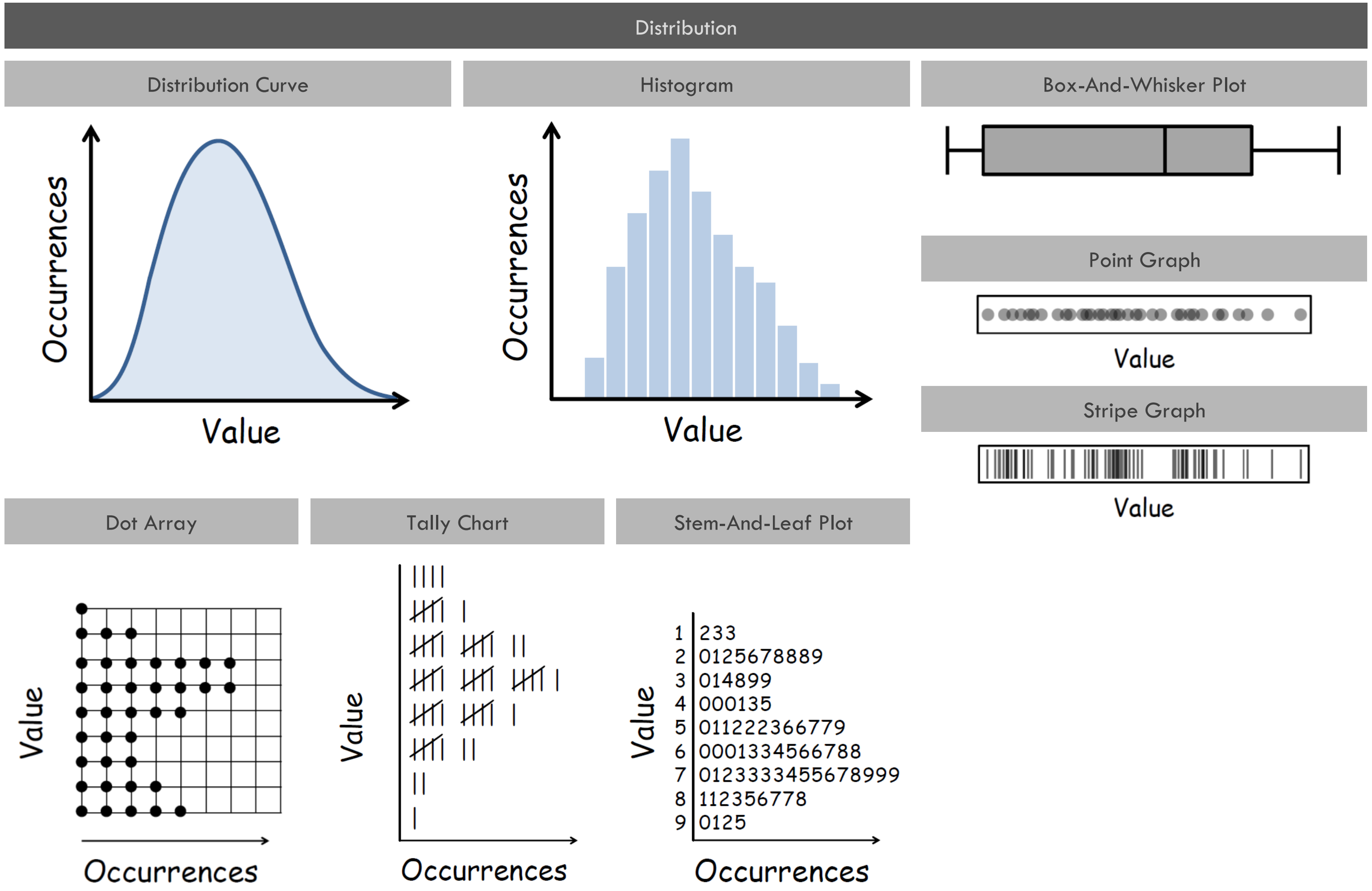
Sankey Diagram

Timeline



Venn Diagram




Distribution
Distribution Curve
Occurrences
Value
Histogram
Occurrences
Value
Box-And-Whisker Plot
Point Graph
Value
Stripe Graph
Value
Dot Array
Value
Occurrences
Tally Chart
Value
Occurrences
Stem-And-Leaf Plot
1 233
2 0125678889
3 014899
4 000135
5 011222366779
6 0001334566788
7 0123333455678999
8 112356778
9 0125
Value
Occurrences

Tree and Network (Graph)

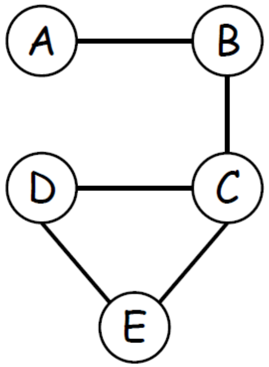
Grid / Matrix

Tree and Network (Graph)

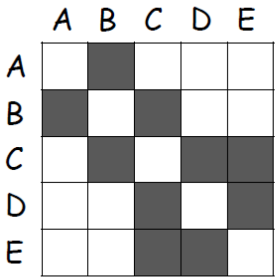
Hive Graph

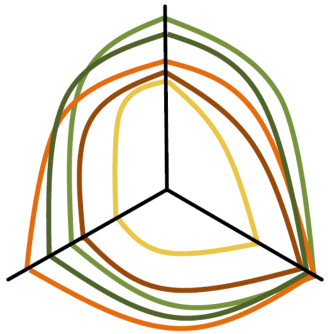
Heat Map

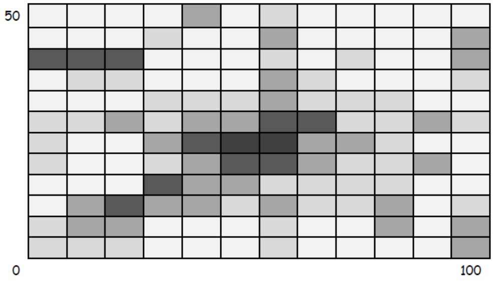
Graph



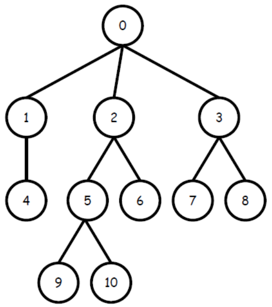
Matrix Representation



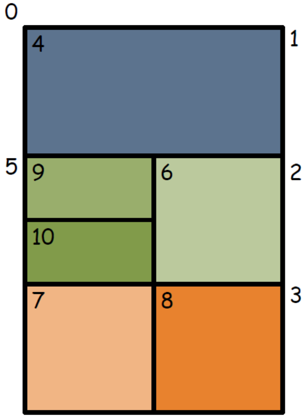




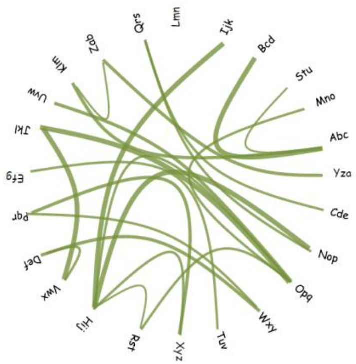
Tree

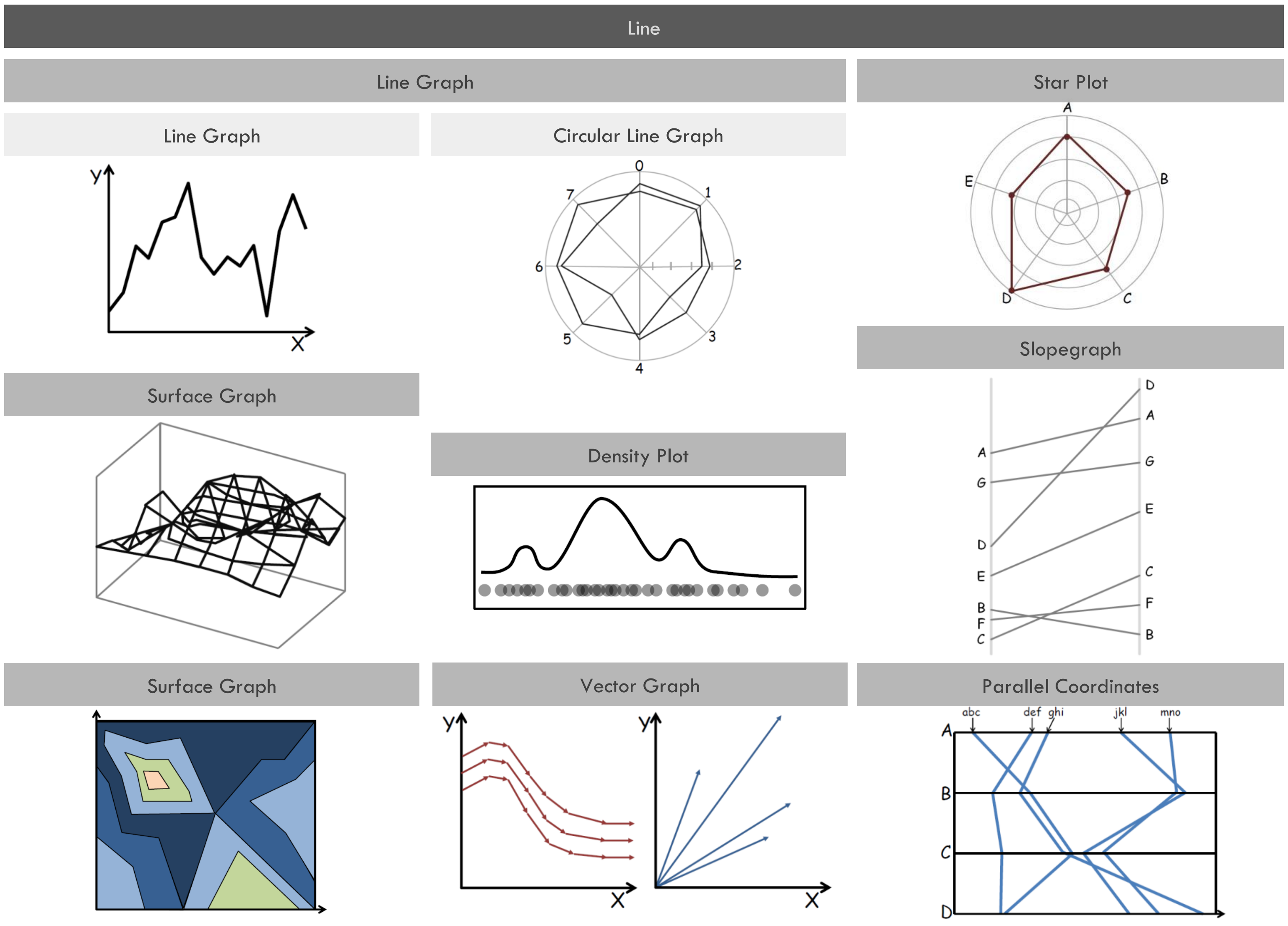


Treemap



Hierarchical Edge Bundling



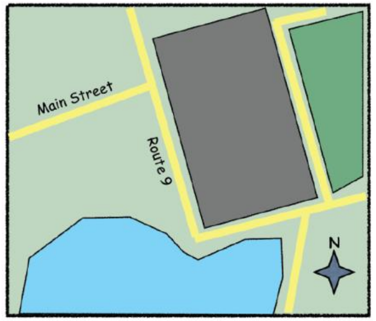

Line
Line Graph
Star Plot
Line Graph
Circular Line Graph
Y
X
0
1
2
3
4
5
6
7
A
B
C
D
E
Slopegraph
Surface Graph
Density Plot
A
G
D
E
B
F
C
D
A
G
E
C
F
B
Surface Graph
Vector Graph
Parallel Coordinates
Y
X
Y
X
abc
def
ghi
jkl
mno
A
B
C
D

Map

Geographic Map

Geographic Map

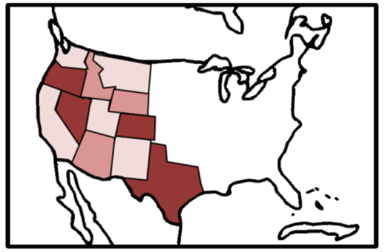
Street Map



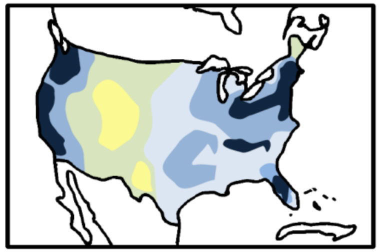
Flow Map

Statistical Map

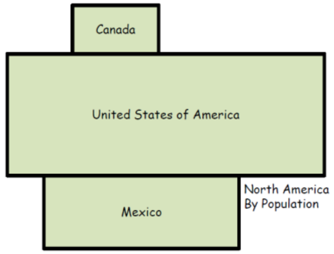
Choropleth Map



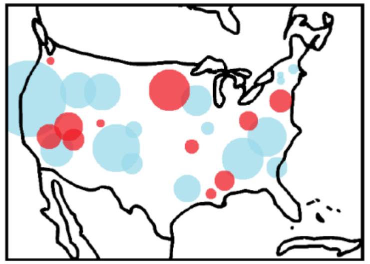
Contour Map (Isopleth)



Distorted Map (Cartogram)



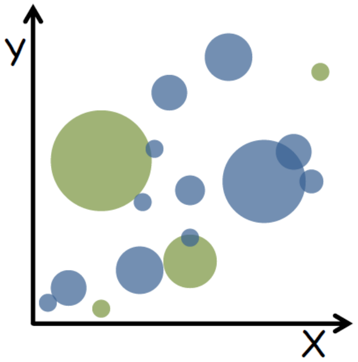
Statistical Plot Map



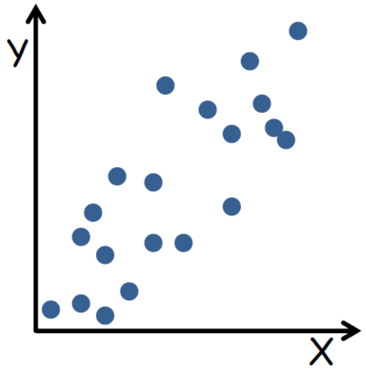
Point

Scatter Plot

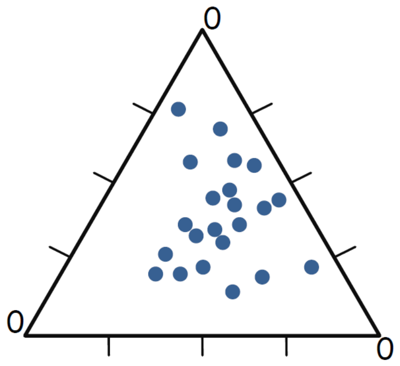
Bubble Chart



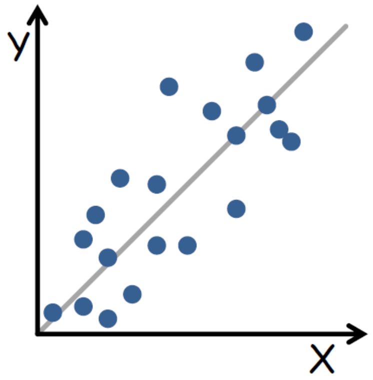
Scatter Plot



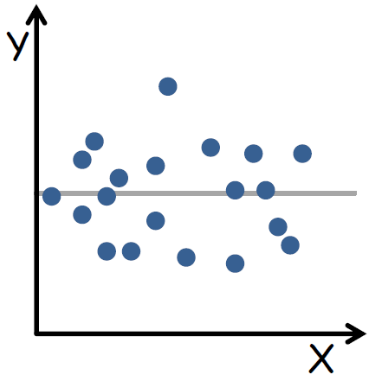
Trilinear Scatter Plot



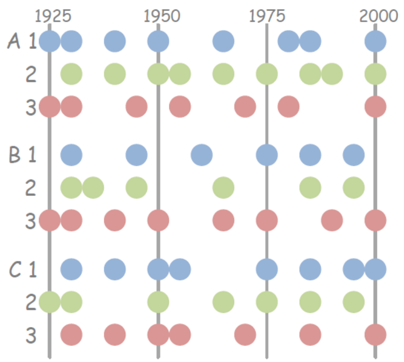
Trend Line



Residual Graph



Dot Plot



Table

Table

ABC	1234	X45
Category	543.2109	7%
Group	45.67	45%
Unit	9876	98%
Class	123.78	12%

Text Chart

Title

- Sed dignissim vehicula
- Nisl quis congue
- Sed vitae rhoncus odio
- Integer at odio

Heading 1

"Nunc aliquam turpis at tellus varius hendrerit. Ut nec magna tortor. Proin adipiscing dolor eget odio semper ut commodo lacus imperdiet."

- Lorem

Heading 2

Aenean tincidunt sem vel massa cursus non tempus quam auctor. In nisi mi, commodo sit.

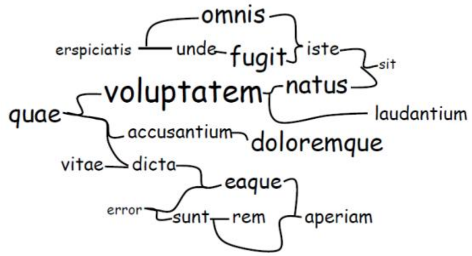
Amet rutrum vitae, fringilla non urna. Quisque sagittis ultrices sapien, quis posuere massa interdum quis.

Heading 3

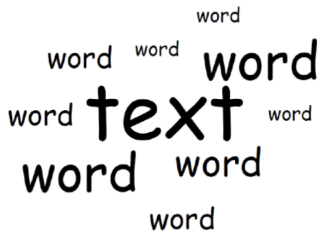
- ✓**Chart 1**
- ✓**Chart 2**
- ✓**Chart 3**
- ✓**Chart 4**

Text Based

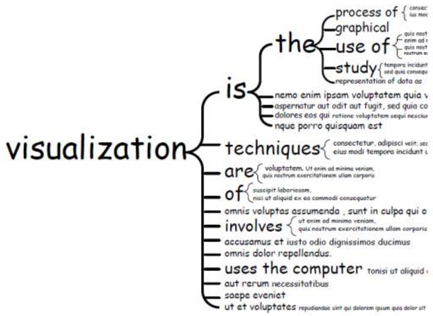
Phrase Net

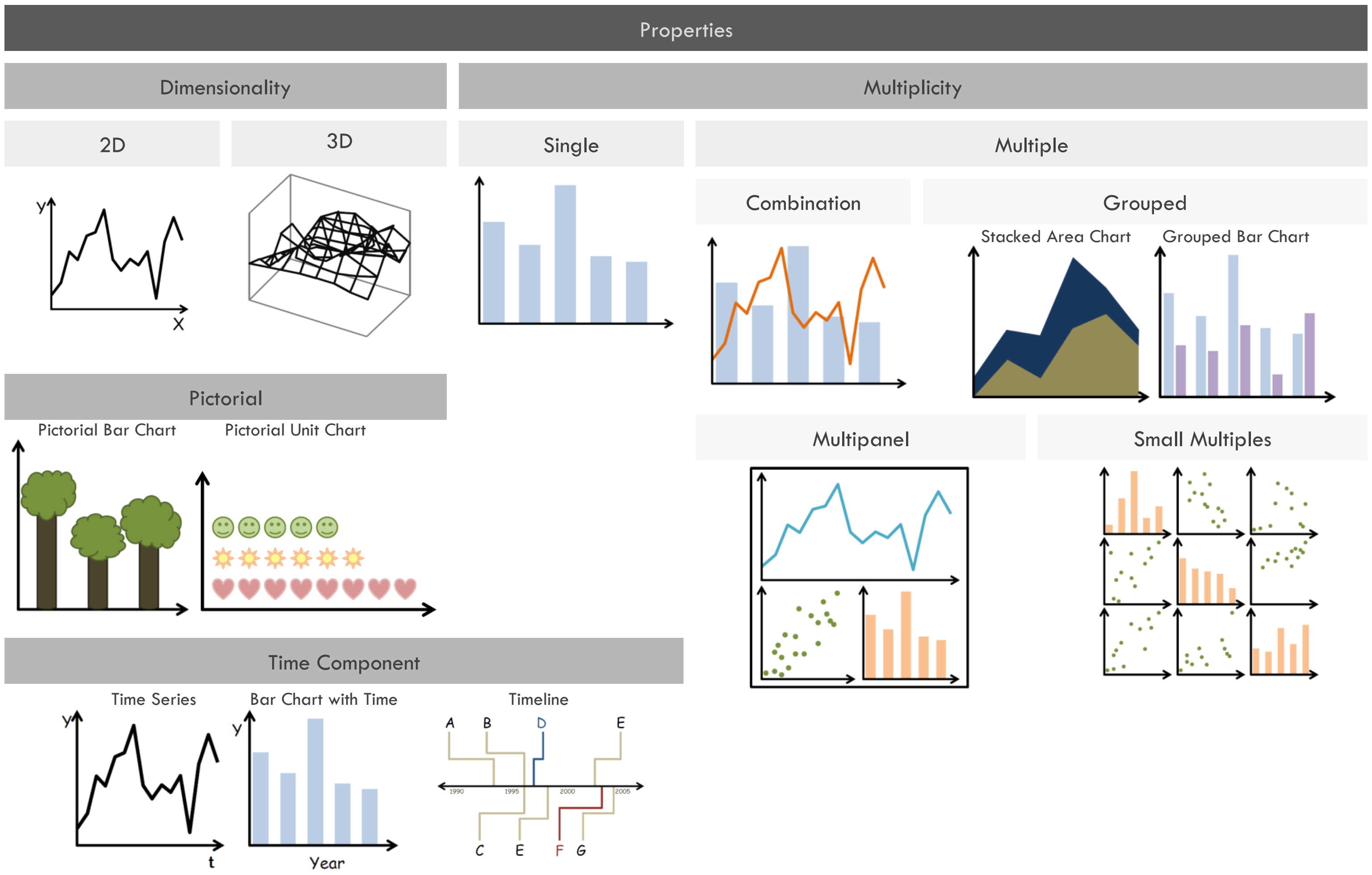


Word Cloud



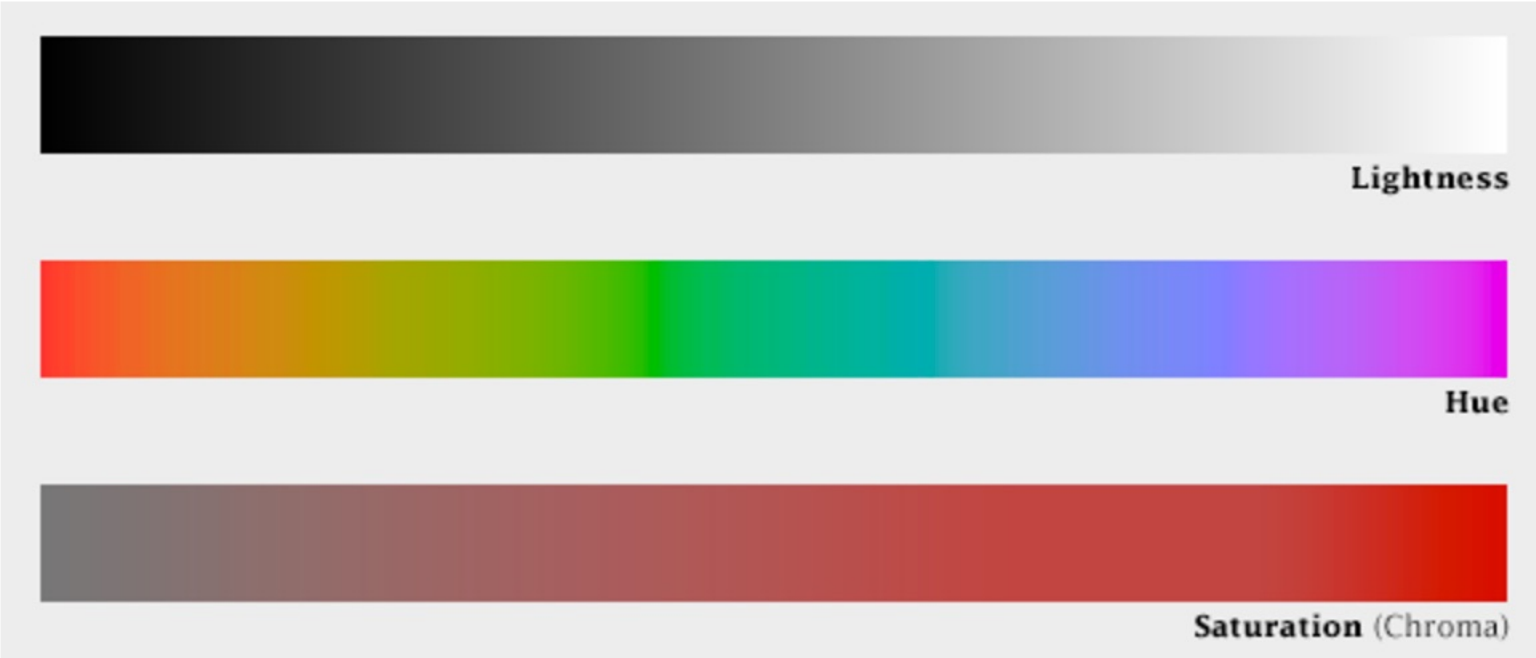
Word Tree





Properties
Dimensionality
Multiplicity
2D
3D
Single
Multiple
Combination
Grouped
Stacked Area Chart
Grouped Bar Chart
Pictorial
Pictorial Bar Chart
Pictorial Unit Chart
Multipanel
Small Multiples
Time Component
Time Series
Bar Chart with Time
Timeline
Year

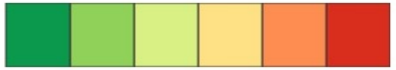
COMPONENTS OF COLOR



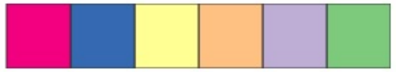
TYPES OF COLOR SCHEMES

Sequential

Diverging



Qualitative



Using a total of 4 – 6 different colors is ideal

Beyond 12 colors, a palette becomes difficult to discern

If necessary, transform your data to reduce the number of data classes

Use semantically relevant colors that are consistent with the tone and content of the data your visualization represents

Color palette choices should demonstrate parity in comparisons between both relative and absolute magnitudes

Simplify charts and graphs
to aid communication

GRAPHICAL INTEGRITY

"Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data."

Edward Tufte, "Visual Display of Quantitative Information"

Often, we need to work within and against **software defaults** that may impede chart simplification

SIMPLIFYING CHARTS AND GRAPHS

Reduce number of elements

Fewer elements means fewer bits to process

Use Gestalt principles

Alignment, spacing, grouping, and hierarchy

Use strong visual encodings

Position, length, proportion, hue, and more

When is it appropriate to create a visualization in 3D?

Only when the 3rd dimension is inherently spatial

(rarely ever)

Two-dimensional representations of multidimensional data use

visual depth cues

to represent higher dimensions

EXAMPLES OF MULTIDIMENSIONAL PLANE VISUALIZATIONS

Heatmap
Scatter plot matrix
Parallel coordinates
Network
Trellis
Small multiples

Why do we use data visualization
to explore our data?

Visually displaying quantitative data is
a bootstrapping mechanism that
makes it easier to find patterns

One way we can use visualization to explore data is through making charts and graphs interactive

DATA EXPLORATION TECHNIQUES

Zooming and panning
Highlighting
Brushing
Filtering (faceting)
Eliding
Details on demand
Animation

Linking

STRATEGIES FOR STATIC VISUALIZATIONS

Small multiples

Callouts (selection/highlighting)

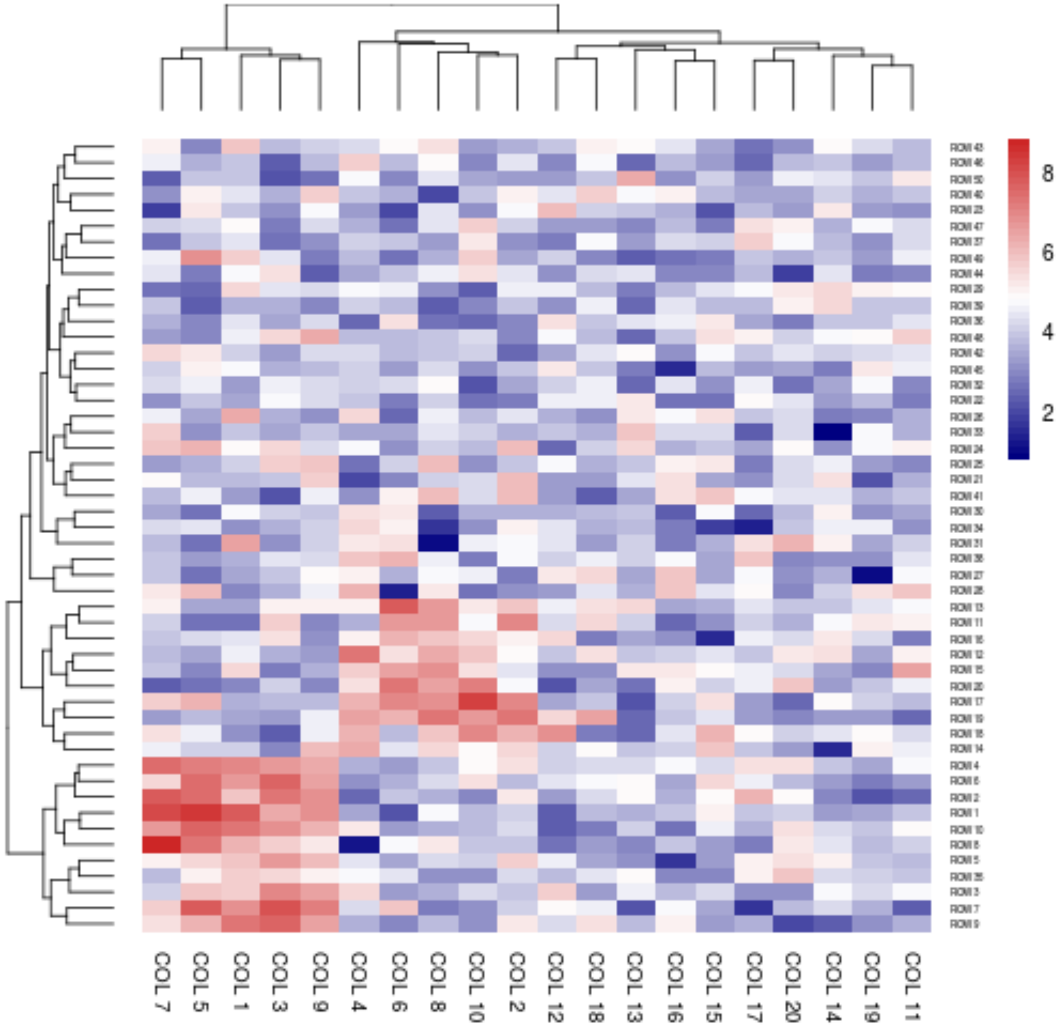
Filtering and faceting

Multiform encoding

VISUALIZATION BINGO

When creating the visualization shown here, it is best to use

this kind of color scheme



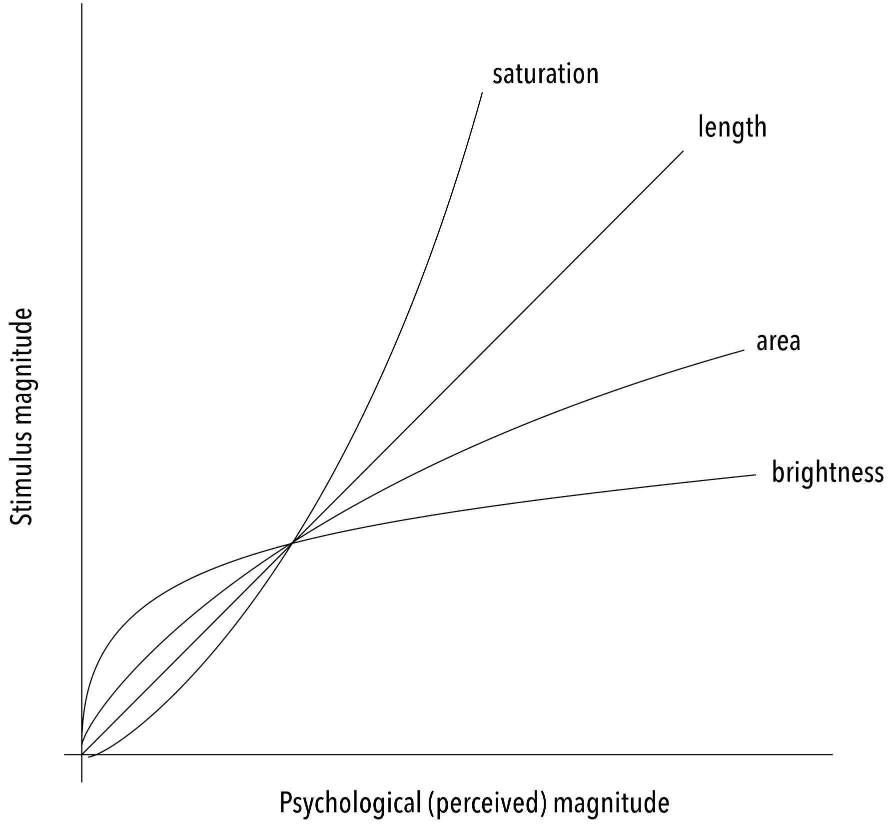
This concept, also known as the "just noticeable difference," says that the perceived change in a stimulus is proportional to the actual magnitude of the stimulus

STEVEN'S POWER LAW

The perceived magnitude or intensity of a stimulus

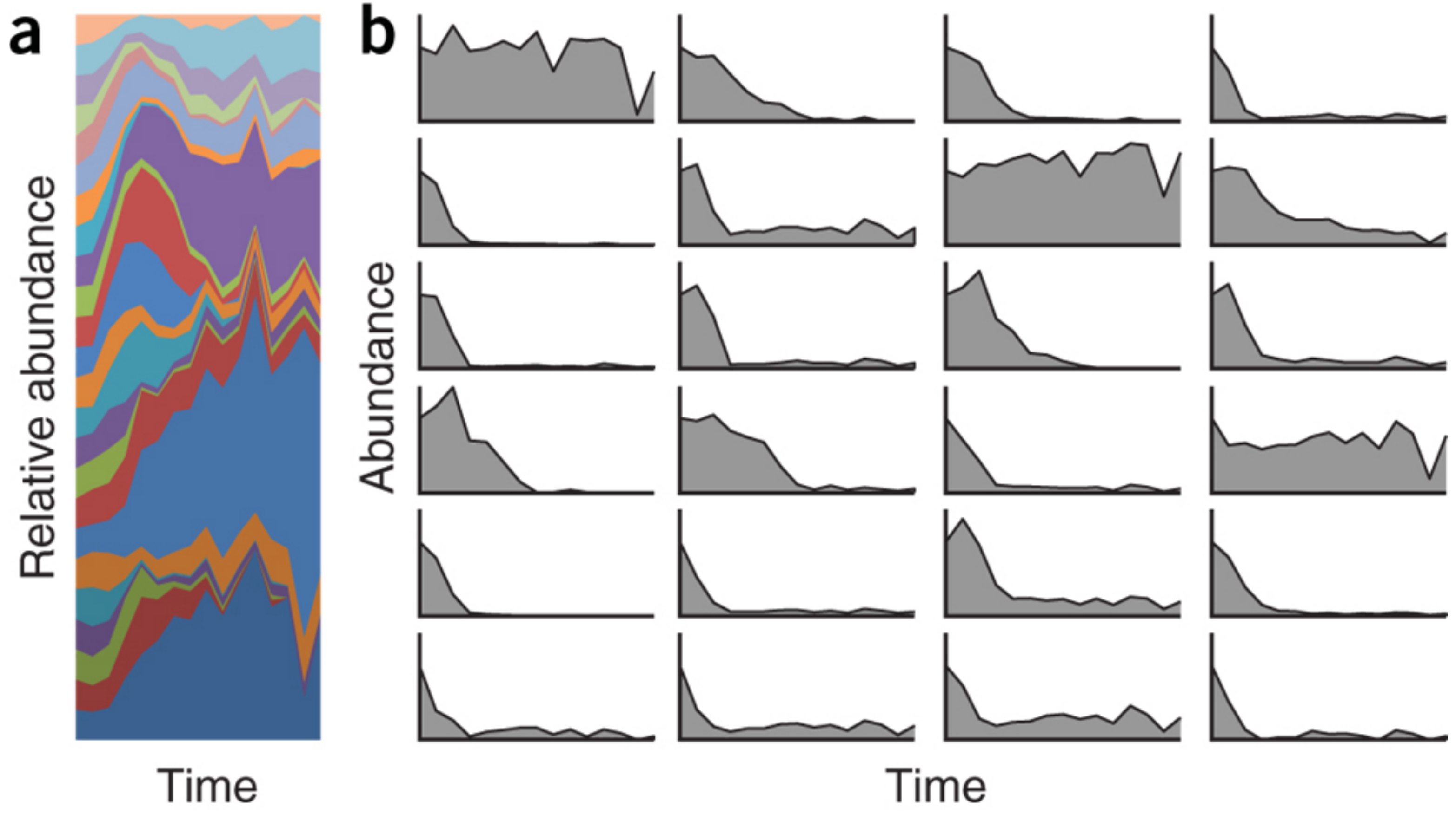
is related to the actual stimulus intensity

by a power law



These describe the psychology behind how we perceive whole forms that arise from combinations of smaller parts

In this form of visualization, the same chart is repeated in sequence across different slices of a shared dimension

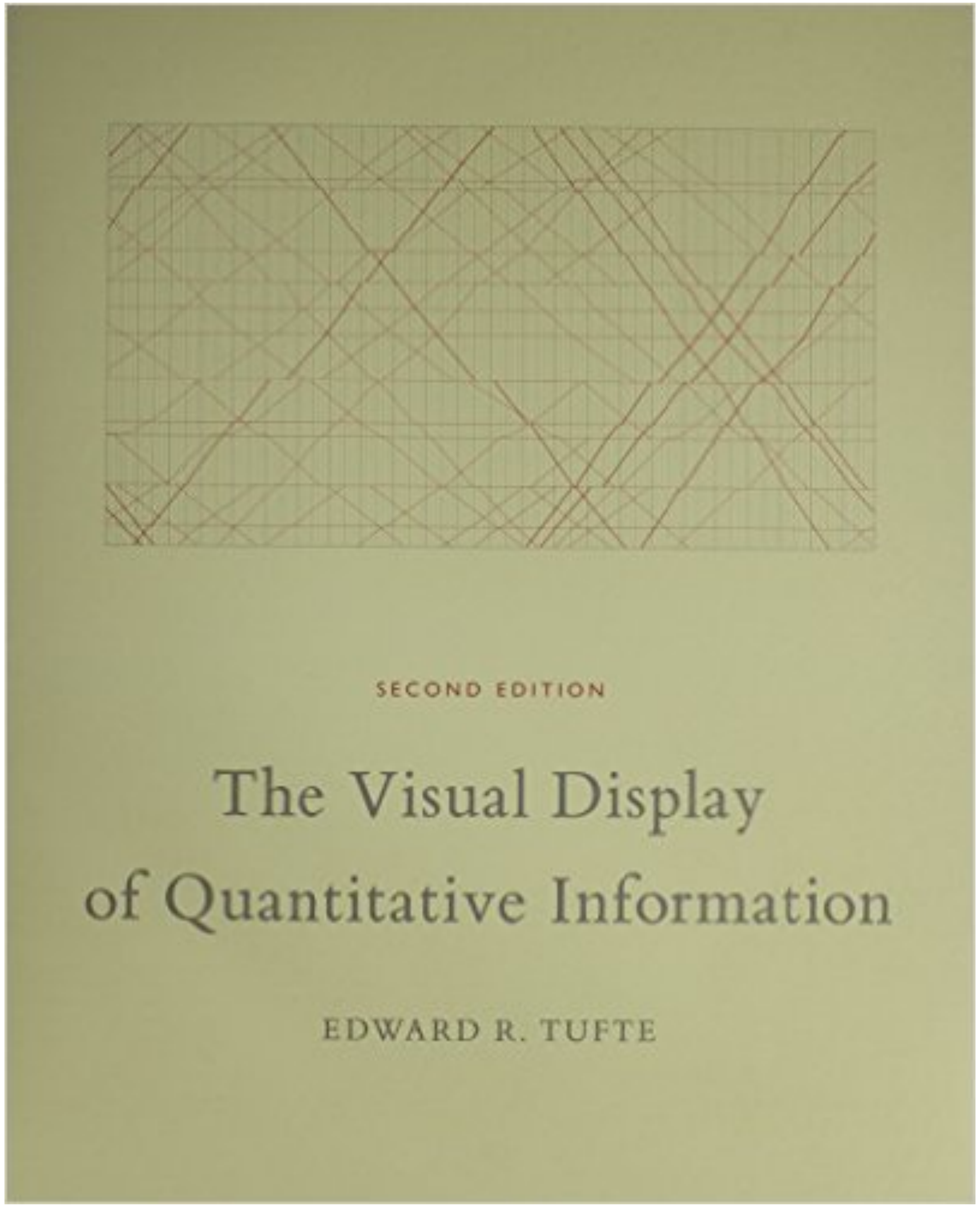


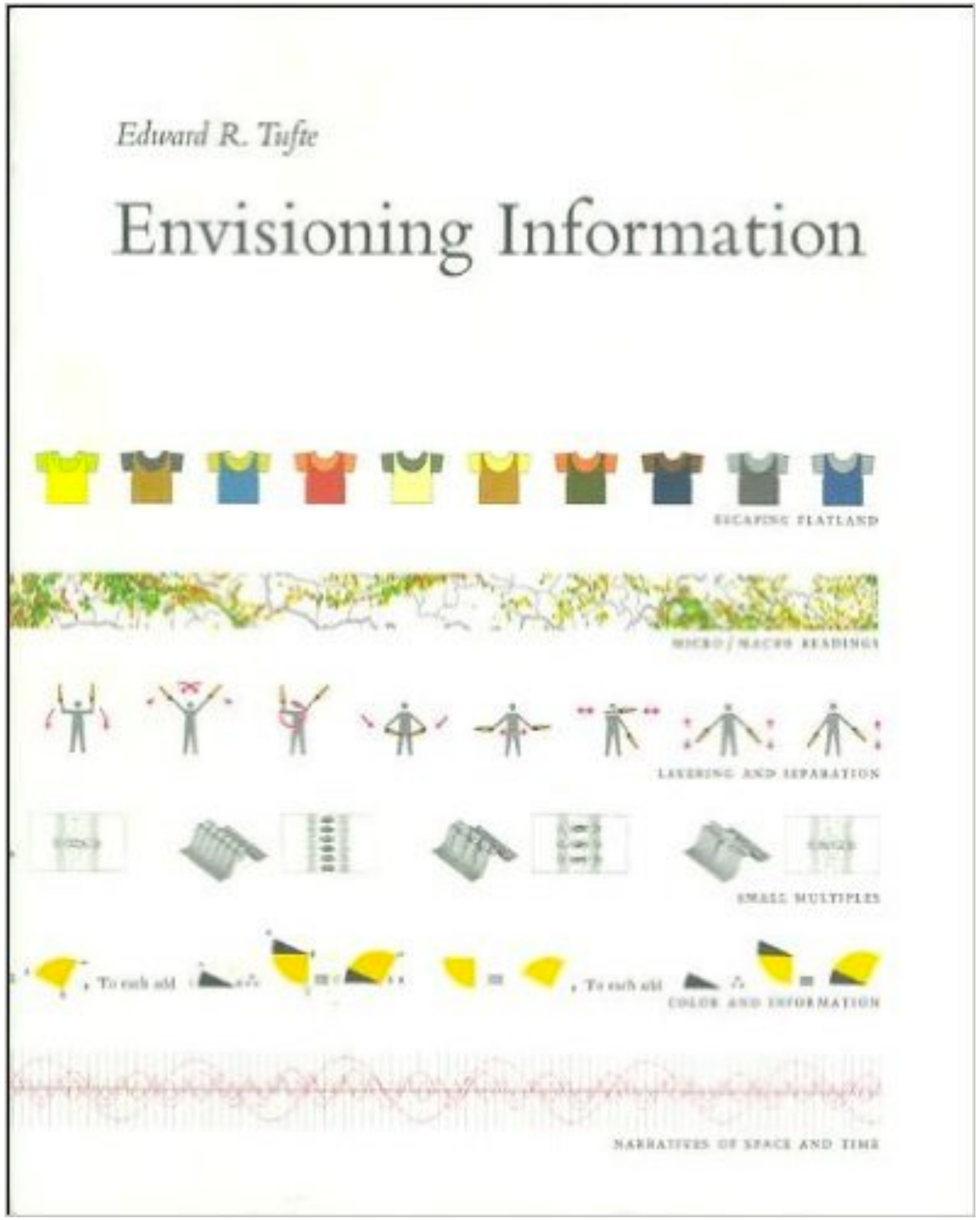
Of the three components of color, human perception is best at discerning variations in

this

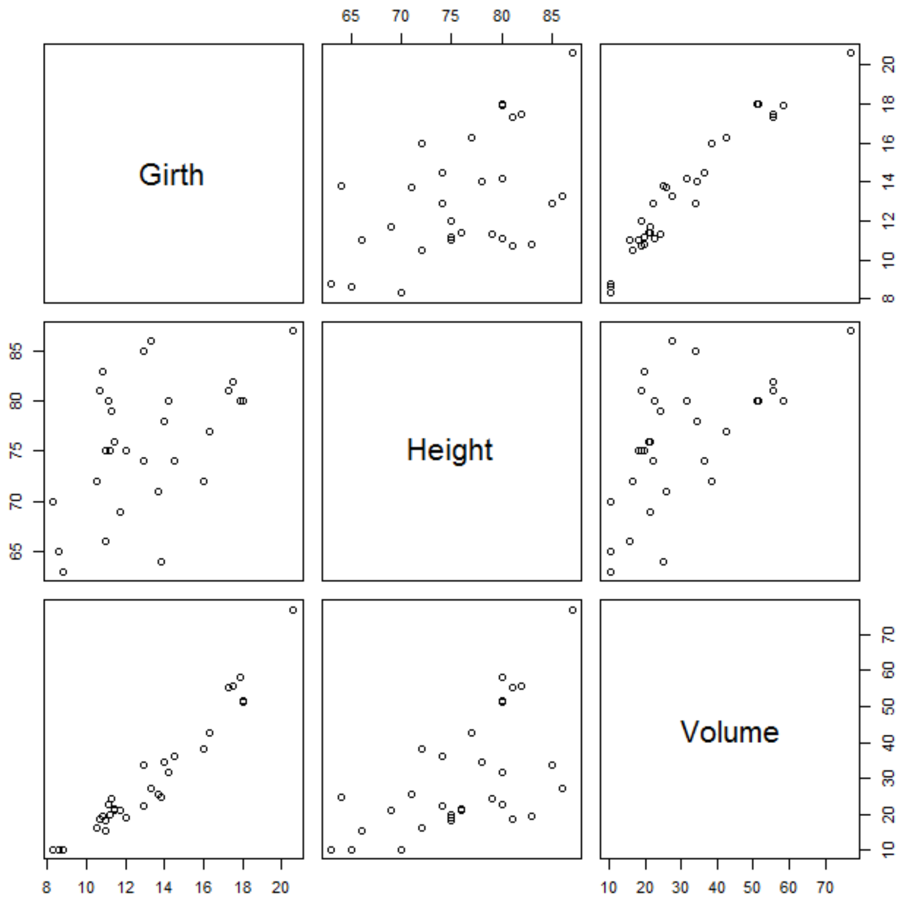
In this kind of qualitative data scale, there are distinct classes that have no natural hierarchy

This "father of visualization" is often quoted for saying "Above all else, show the data"

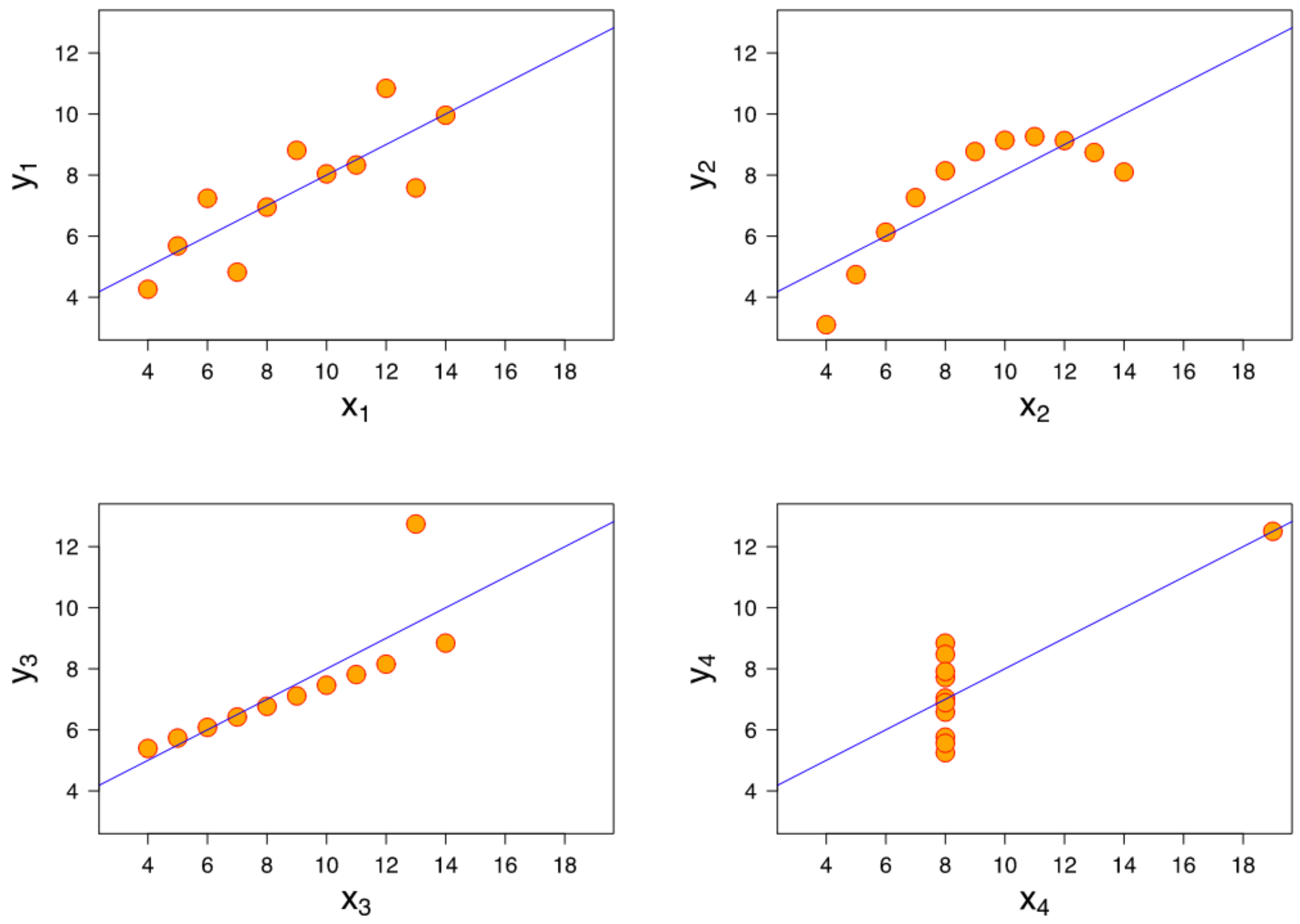

SECOND EDITION
The Visual Display of Quantitative Information
EDWARD R. TUFTE


Edward R. Tufte
Envisioning Information
ESCAPING FLATLAND
MICRO / MACRO READINGS
LAYERING AND SEPARATION
SMALL MULTIPLES
COLOR AND INFORMATION
NARRATIVES OF SPACE AND TIME

In the kind of visualization depicted here, multiple scatter plots are arranged to share common axes across different attributes



In this, multiple sets of data have the same summary statistics but differ drastically when graphed



y_1
x_1
y_2
x_2
y_3
x_3
y_4
x_4
4 6 8 10 12 14 16 18
4 6 8 10 12

In ggplot2, we use this function to add a small amount of random variation to the location of points in a chart

In this kind of visualization, changes in one chart correspond to simultaneous changes in another chart

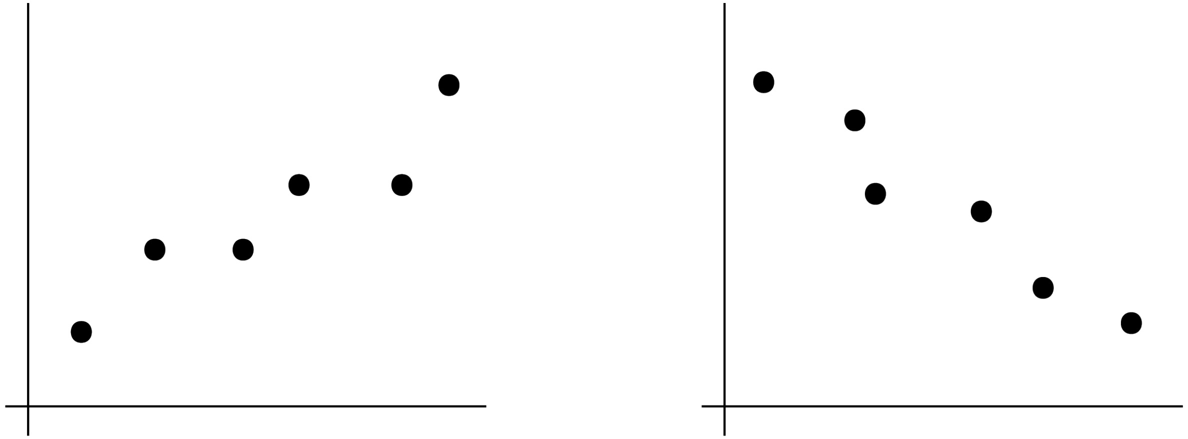
This attribute of a color describes the vividness and perceptibility of a color at different resolutions/sizes

According to this “law,” we perceptually group like elements based on size, color, shape, and other properties

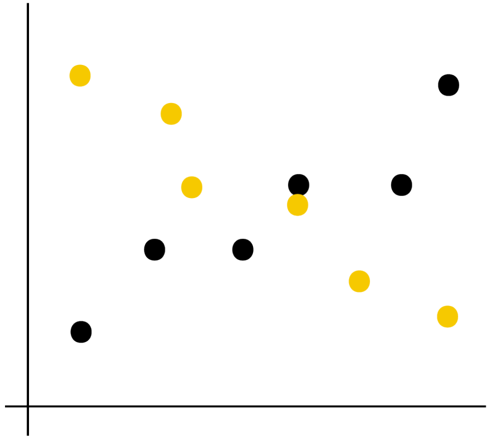
In this exploratory technique, data classes may be juxtaposed, partitioned, or superimposed

FILTERING AND FACETING

Side-by-side



Layered

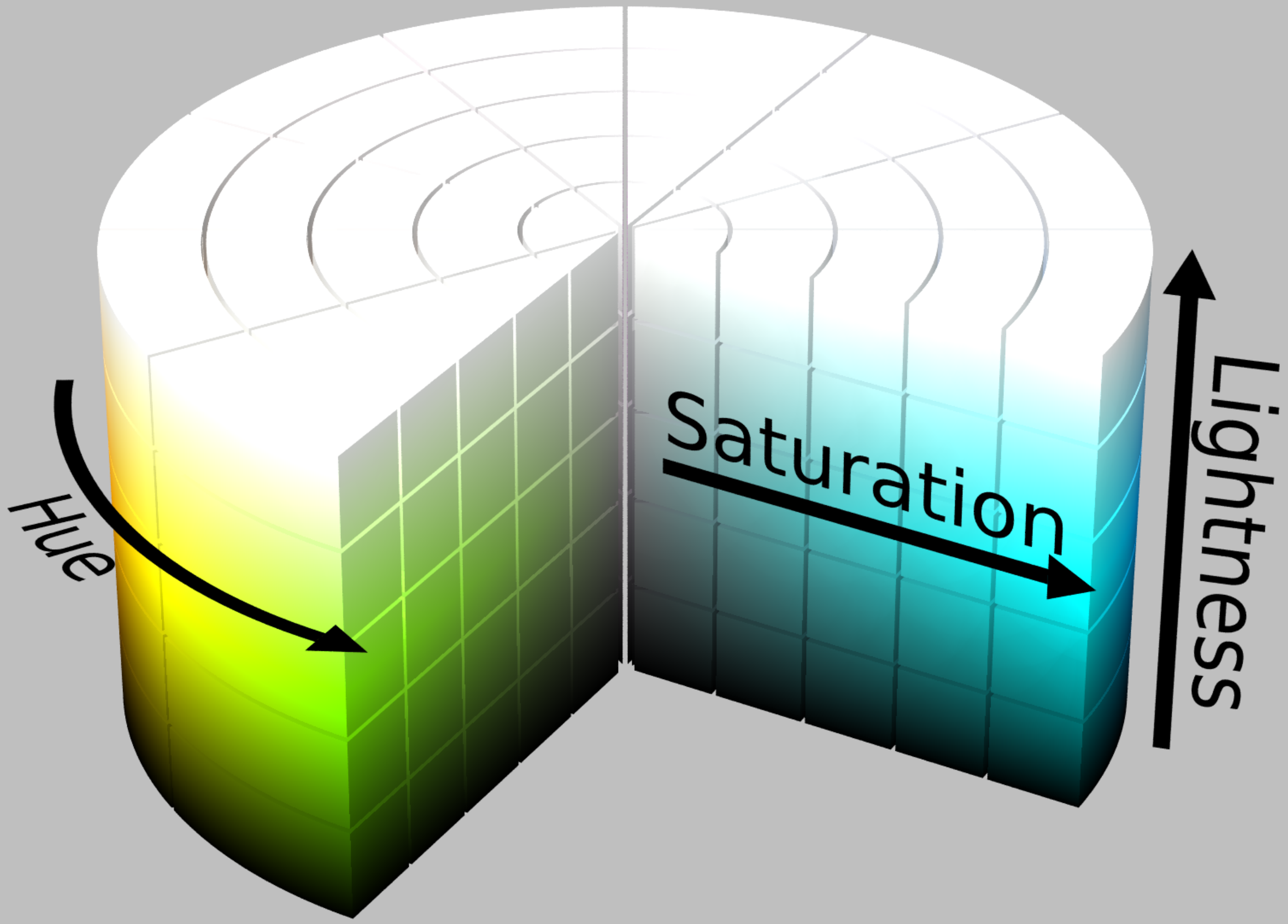


In the color illusion shown here, identical colors may appear to be different because of perceptual effects caused by adjacent colors

This concept says that in any visualization, the quantity of ink used to depict the data should be as equal to the total ink used to print the graphic as possible

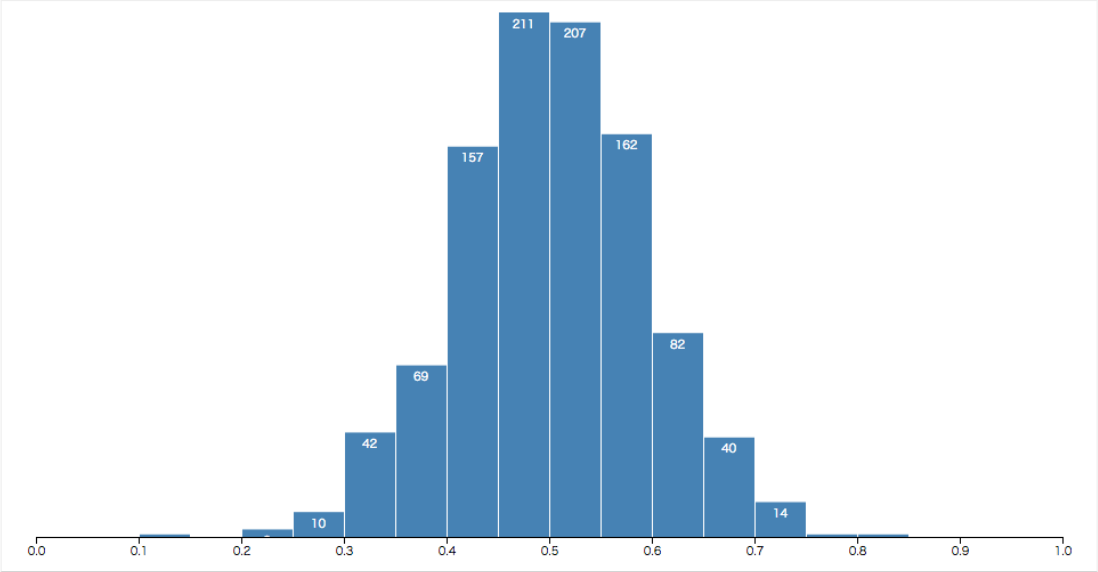
$$\text{“data-ink ratio”} = \frac{\text{data ink}}{\text{total ink used to print the graphic}}$$

In one color model commonly employed in visualization, the three components of color include lightness, hue, and this

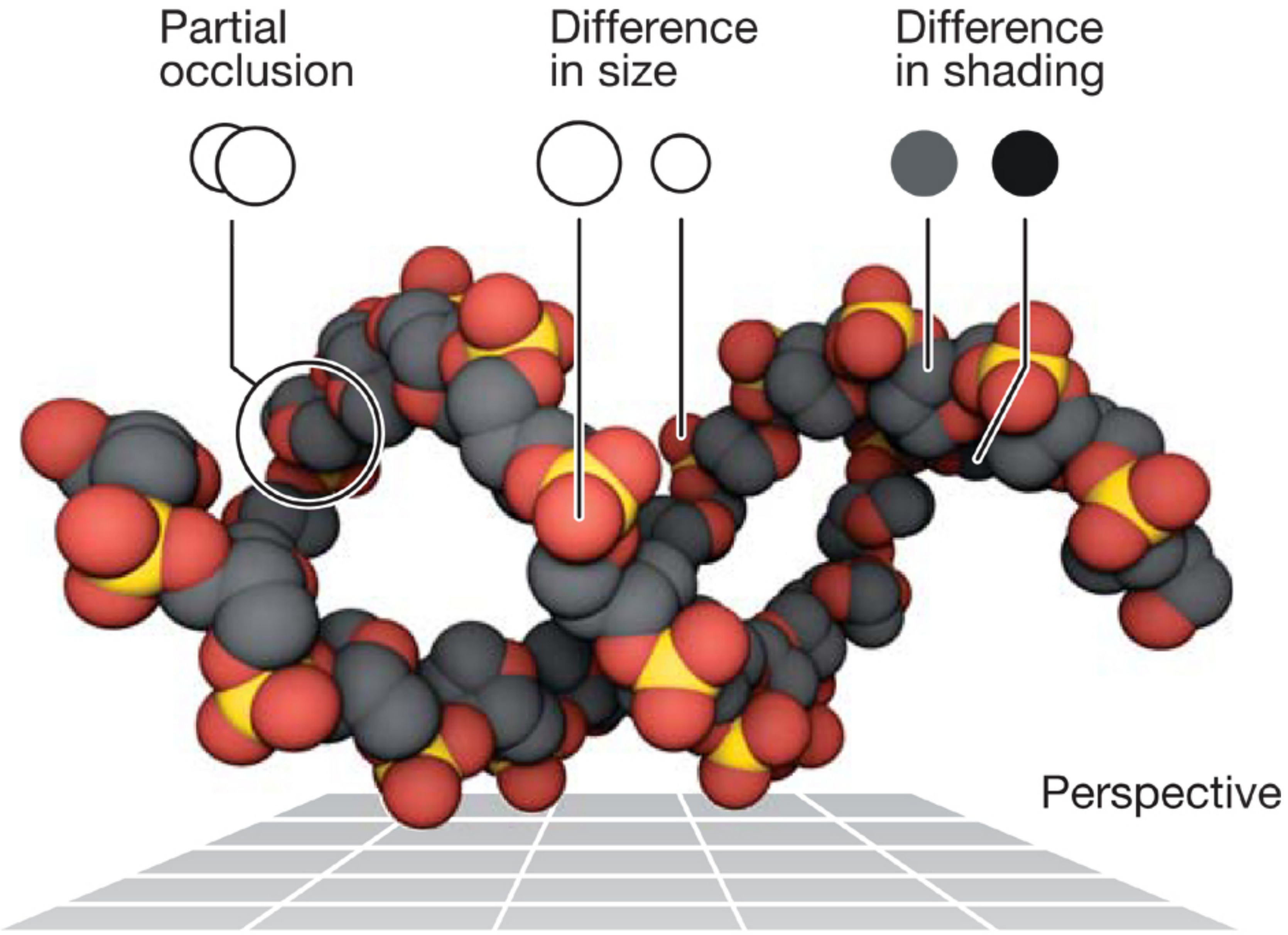


https://en.wikipedia.org/wiki/HSL_and_HSV

Although similar in appearance to a bar chart, this kind of chart, shown here, is used to represent continuous distributions of data



Visual depth cues, such as color, size, shading, and this, enable us to depict higher-dimensional data in 2 dimensions



Gehlenborg, Nils and Wong, Bang. Points of View: Into the third dimension. *Nature Methods* 9, 851 (2012)

In this kind of visual representation, the same data are presented multiple times using different visual encodings

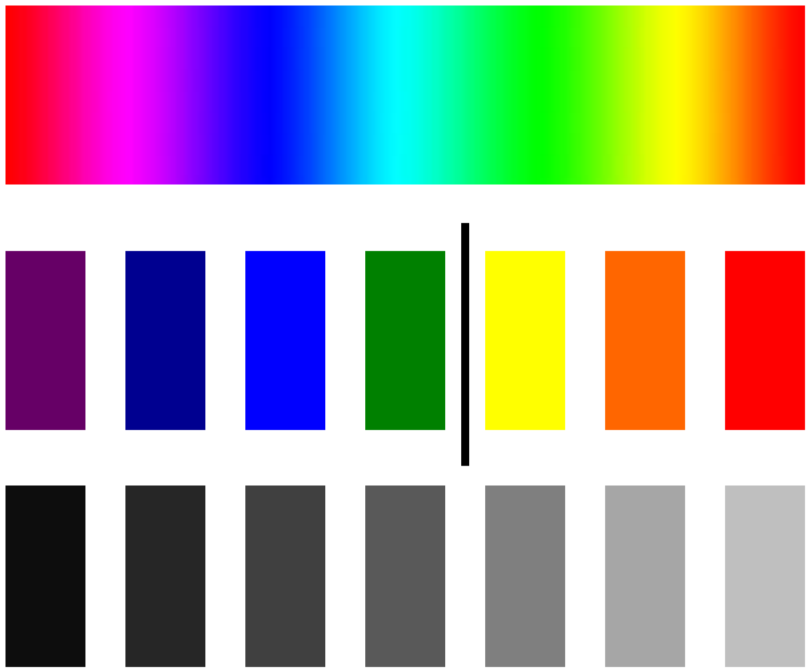
In ggplot2, we use this function to change the size, color, and rotation of plots

In R, we can use this package for visualizing sets and their intersections

This kind of color scheme is problematic for many reasons, including its imperceptibility by people with color blindness and its lack of a natural ordering of values

CONSIDERATIONS IN UNIVERSAL DESIGN

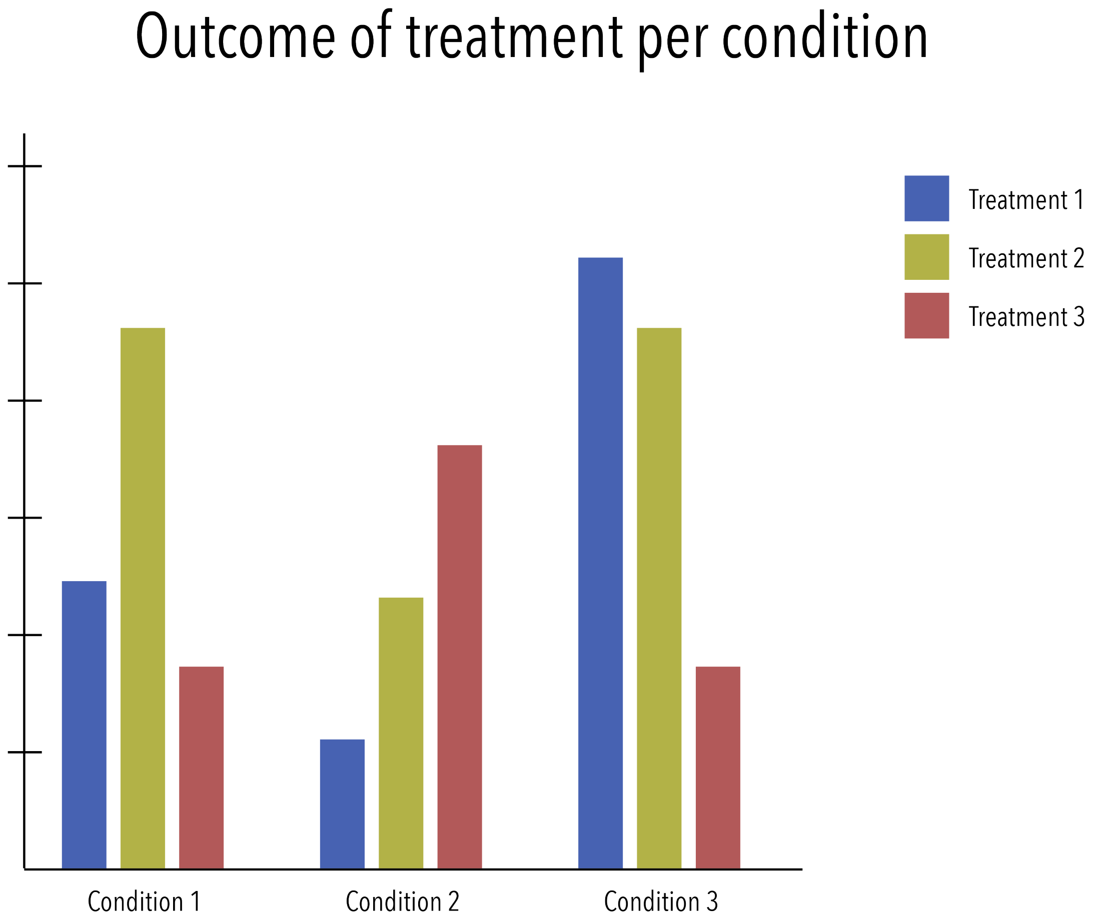
Color perception insensitivities



These refer to dimensions of data that you communicate using visual encodings

In this mode of interaction, details are depicted for a selected data item while remaining data are displayed in summary

When creating the visualization shown here, it is best to use **this kind of color scheme**



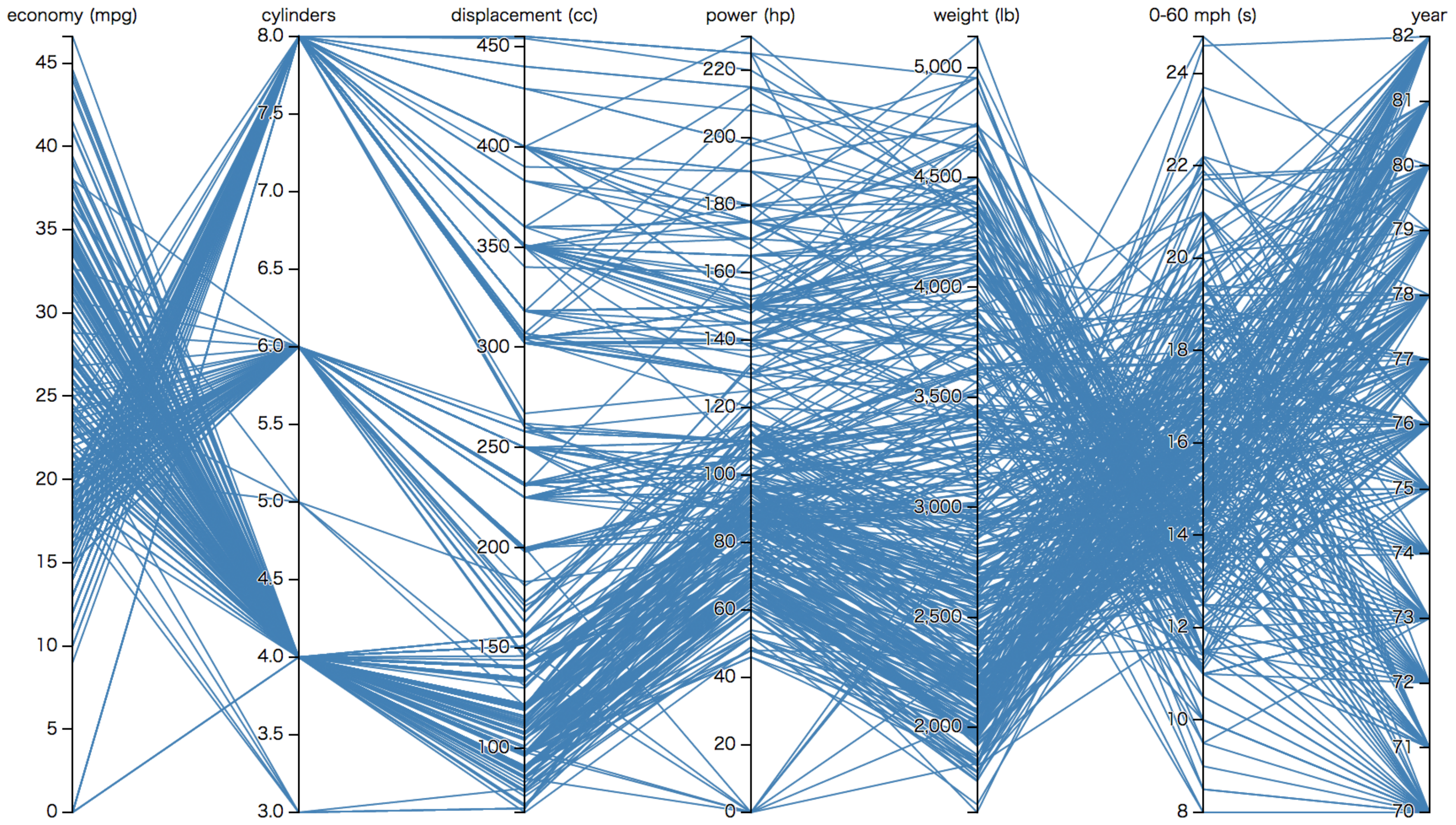
This mathematically describes the relationship between the height and width of a visualization

Color palette choices should demonstrate this in comparisons between both relative and absolute magnitudes

We can improve the readability of a visualization by enhancing **this**, which describes the extent to which a visual encoding is designed to pop out in our visual search processes

In ggplot2, we can use **this** plot function for visualizing data with one discrete variable and one continuous variable

In the kind of visualization depicted here, multiple dimensions are plotted across multiple parallel axes



In this kind of qualitative data scale, there are distinct classes that possess a natural hierarchy

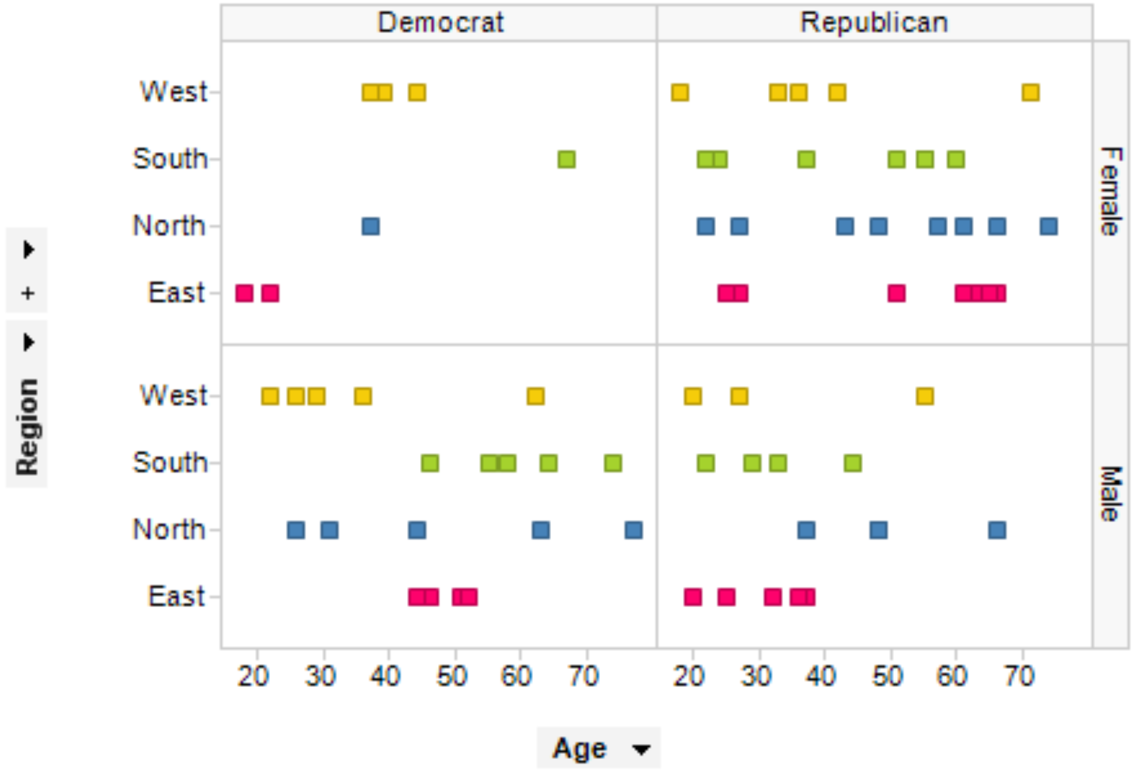
According to this “law,” our interpretation of visual forms is informed by past experience

When creating visualizations with lines as marks, an aspect ratio should be selected to “bank” line segments to this

In the heatmap() function, this cluster method is used

These refer to the basic graphical elements of a chart or figure

In this kind of display shown here, data are faceted out across different dimensions



In this form of dimensional reduction, we identify dimensions of greatest variability

Of all possible visual encodings (channels),
this is the most accurately perceived

Position, brightness/saturation, and this represent the most common ways of encoding temporal data in visualizations

According to this “law,” cognitive load is optimized by reducing visual forms to their simplest forms

These refer to ways (encodings) in which you can control the appearance of marks

In the heatmap() function, we use this argument for centering and scaling for color