

Foundations of Data Visualization

(A Brief Overview)

Cheng Peng

Department of Mathematics



Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding

Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

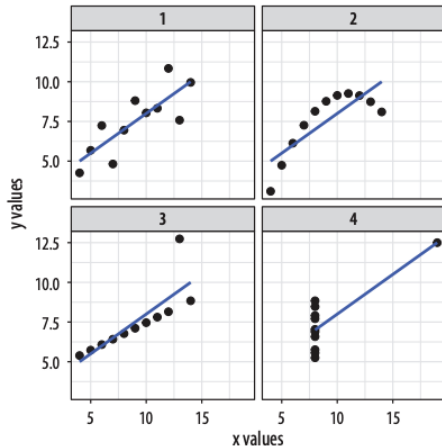
Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

- Plots of Anscombe's quartet.



Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding

Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

- Brief history of data viz.
- Overview of dataviz process
- Principles of dataviz
- Building blocks of dataviz.
- Color coding and usage guidelines
- A general framework for creating better charts
- Purpose and types of dataviz
- Specific visualization Skills

1970s - Foundation of modern dataviz

- John Tukey pioneers the use of visualization with computers (exploratory and confirmatory visualization).

1980s - The science of visualization

- Edward Tufte's work combines statistics with visual design principles
- Cleveland and McGill's work on measuring graphical perception
- Mackinlay's work carries visualization theories to digital age

1990s-2000s: The computer-driven scientific visualization thrives

2010s: The social internet, cheap and easy-to-use software, and massive volumes of data democratize the practice of visualization.

- Rensink and Harrison establish science around graphic perception

Why Visualization

Agenda

Dataviz in past 50 Years

Visualization Process

Gestalt Principles

Gestalt Design Principles

Pre-attentive Attributes

Visual Encoding

Marks

Channels

Expressiveness & Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

- Understand Visualization
 - Dataviz is a process
 - Dataviz is new language built on the science and art.
- Prepare Visualization
 - Manage and clean data
 - Talk and listening
 - Sketch and prototype
- Create Visualization
 - Is the information conceptual or data driven?
 - Am I declaring or exploring something?
 - Types of visualization
- Refine Visualization
 - to make impressive charts
 - to make persuasive charts
- Present and Practice Visualization
 - to persuade audience
 - to make better charts

Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding
Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

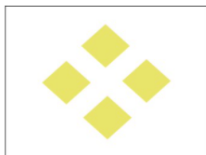
Steps for Good Charts

The principles describe the various ways we tend to visually assemble individual objects into groups and are highly relevant to the design of charts and graphs.

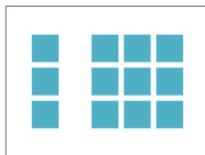
- Objects will be perceived in their simplest form
- Humans naturally follow lines and curves
- The mind will attempt to fill in detail that isn't actually there.



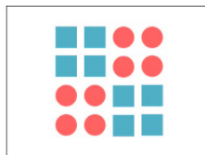
- **Closure:** Elements are typically grouped together if they are a part of an entity
- **Proximity:** Elements are typically grouped together based on their immediacy
- **Similarity:** Elements similar to one another tend to be grouped together



Closure

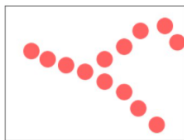


Proximity

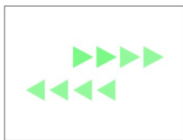


Similarity

- **Continuity:** Elements that are arranged on a line or curve are perceived to be more related than elements not on the line or curve.
- **Common Fate:** When elements coordinate movement together, we tend to relate them to each other.
- **Focal point:** When an element or elements stands out visually, it captures and holds our attention first.



Continuity



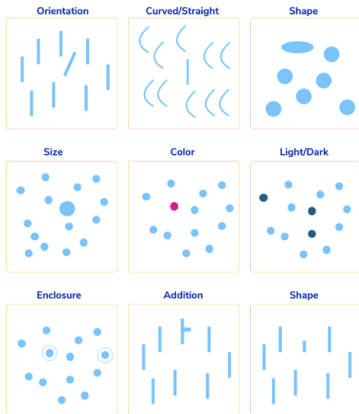
Common Fate



Focal Point

Pre-attentive Attributes

- A preattentive visual property is one which is processed in spatial memory without our conscious action.



- It takes less than 500 milliseconds for the eye and the brain to process a preattentive property

- The visual encoding is the way in which data is mapped into visual structures made of marks and channels.
- Data visualization is the graphical representation of information and data built based on visual structures.
- Marks and Channels - Building blocks of visualization
 - **Marks** (geometric primitives) represent items or links - basic graphical element in an image.
 - **Channels** (aka channel variables) change appearance of marks based on attributes - independent of the dimensionality of the geometric primitive.
- By using marks and channels to create visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding

Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

- Marks for items

→ Points



0D

→ Lines



1D

→ Areas



2D

- Marks for links

→ Containment



→ Connection



Control appearance proportional to or based on attributes.

→ Position

→ Horizontal



→ Vertical



→ Both



→ Color



→ Shape



→ Tilt



→ Size

→ Length



→ Area



→ Volume



Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding
Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

Expressiveness

- visual encoding should express all of, and only, the information in the dataset attributes
- simple one - lie factor (the ratio of the information “in the chart” and the information “in data”)

Effectiveness

- importance of the attribute should match the salience of the channel
- simple one - data-ink ratio (the ratio of “ink in data” and “ink in the chart”)

Chart Junk

- Unnecessary visual elements in charts that distracts the viewer from the information

Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding

Marks

Channels

**Expressiveness &
Effectiveness**

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity


Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

More on Channels: Expressiveness Types and Effectiveness Ranks

➔ 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  Same

Color saturation  Same

Curvature  Same

Volume (3D size)  Same

Most

Effectiveness

Least

➔ Identity Channels: Categorical Attributes

Spatial region 

Color hue 

Motion 

Shape 

- Color improves a chart's aesthetic quality, expressiveness, hence, its ability to effectively communicate about its data.
- **Categorical Pallet:** Categorical colors help users map non-numeric meaning to objects in a visualization.



- **Sequential Pallet:** Sequential colors have numeric meaning.



- **Divergent Pallet:** Diverging colors also have numeric meaning.



Guidelines of using colors with different types of data.

CATEGORICAL SCALE



ORDINAL SCALE



INTERVAL SCALE



RATIO SCALE



CATEGORICAL SCALE



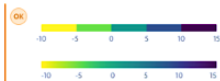
ORDINAL SCALE



INTERVAL SCALE



RATIO SCALE



CATEGORICAL SCALE



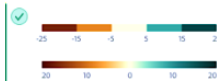
ORDINAL SCALE



INTERVAL SCALE



RATIO SCALE



Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding

Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

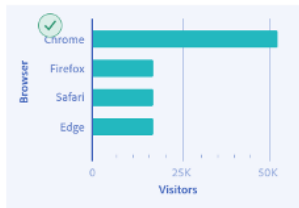
Crafting for Clarity

Choosing Chart Type

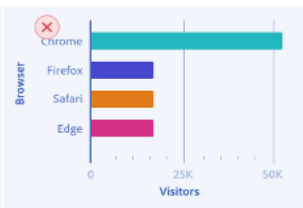
Practicing Persuasion

Steps for Good Charts

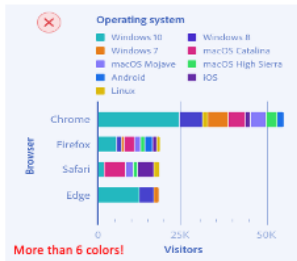
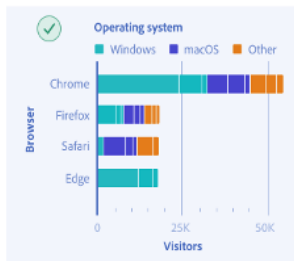
Do's & Don'ts



Use color to separate items into categories.



Don't use color to separate items.



More than 6 colors!

Why Visualization

Agenda

Dataviz in past 50 Years

Visualization Process

Gestalt Principles

Gestalt Design Principles

Pre-attentive Attributes

Visual Encoding Marks

Channels

Expressiveness & Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

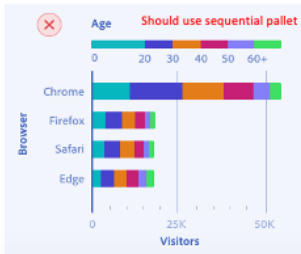
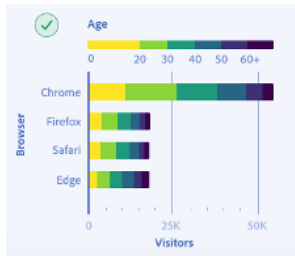
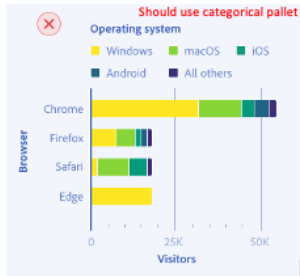
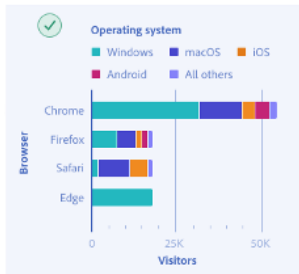
Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts



Why Visualization

Agenda

Dataviz in past 50 Years

Visualization Process

Gestalt Principles

Gestalt Design Principles

Pre-attentive Attributes

Visual Encoding

Marks

Channels

Expressiveness & Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

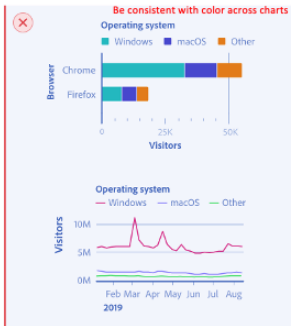
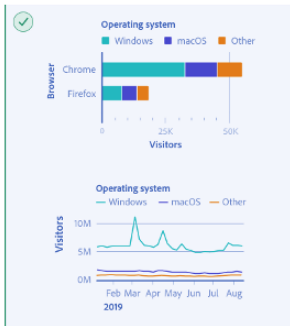
Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts



Why Visualization

Agenda

Dataviz in past 50 Years

Visualization Process

Gestalt Principles

Gestalt Design Principles

Pre-attentive Attributes

Visual Encoding

Marks

Channels

Expressiveness & Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

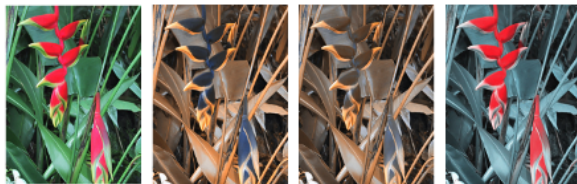
Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

- About 8% - 10 % of men and 1% of women have color vision deficiency.
- Red-green is common (deuteranope and protanope two subcategories).
- Blue-yellow is possible (tritanope is most common in this category)



Normal

Protanope

Deuteranope

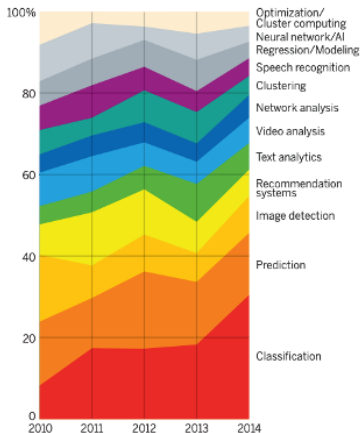
Tritanope

Practical guidelines of using colors in dataviz.

- **Use less color** - keep the number of colors minimum.
- **Use gray** - It doesn't draw the eye the way stronger colors do and is the default color in software.
- **Complement / contrast** - When variables are inherently similar, use similar or complementary colors. When they are in opposition, use contrasting colors.
- **Stick to the variables** - using color for text decoration is distracting.
- **Think how, not which** - It is more important to think about how to use color than which color is used.
- **Consider the color-blind.**

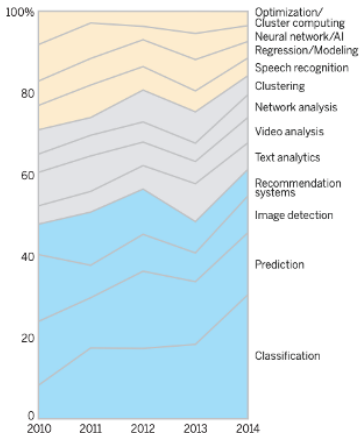
Comparing poor chart and better chart.

PERCENTAGE OF TOTAL RESEARCH PAPERS



SOURCE: ACCENTURE INSTITUTE FOR HIGH PERFORMANCE, ANALYSIS OF A STANFORD MACHINE LEARNING RESEARCH PAPER DATABASE

PERCENTAGE OF TOTAL RESEARCH PAPERS



SOURCE: ACCENTURE INSTITUTE FOR HIGH PERFORMANCE, ANALYSIS OF A STANFORD MACHINE LEARNING RESEARCH PAPER DATABASE

Some guidelines to achieve a clear design.

- **Take stuff away** - think about every mark on your chart and ask, Is this necessary to make your point?
- **Remove redundancy** - A headline that reads “Sales vs. Revenue” just repeats the axis labels.
- **Limit color and eye travel** - Color is powerful—and distracting.
- **Know how people think** - The brain works on heuristics. It is important to respect convention — and take advantage of it.
- **Describe ideas, not structure** - Use text, headlines, captions, and other visual markers to highlight ideas or insights rather than to describe the visualization’s architecture.
- **Align everything** - This simple guideline is supremely effective at creating visual order.

Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding

Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

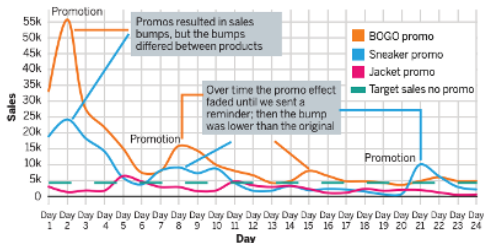
Choosing Chart Type

Practicing Persuasion

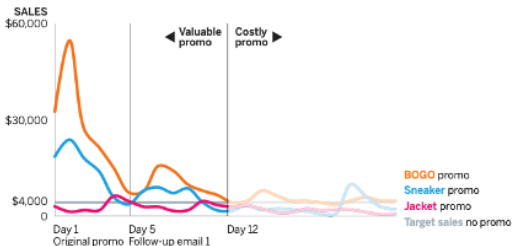
Steps for Good Charts

Comparing poor chart and better chart.

STORE PROMOTIONS AND SALES OVER TWO-PUS WEEKS



STORE PROMOTIONS ARE WORTH IT—FOR 12 DAYS



Guidelines for selecting appropriate chart type.

- **Know the basic categories** - The simplest way to begin is to understand your intent.
- **Listen to how you describe things** - Find someone to chat with about your data and the idea you want to convey.
- **Rely on your workhorses** - Understand that more specialized and unusual chart types will require more effort on the part of your viewers.
- **Don't forget tables** - Sometimes all the individual data points in a set matter more than a trend or what comprises them.
- **Good writers are great readers** - good chart makers are great chart consumers. It is important to find inspiration in others' visualizations to improve you visualizations.

Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding
Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

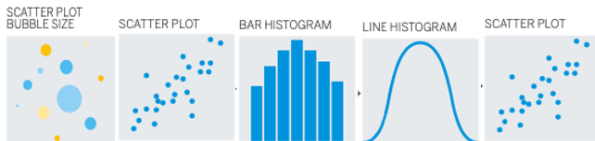
Steps for Good Charts

According to the applications, there are four basic types of charts.

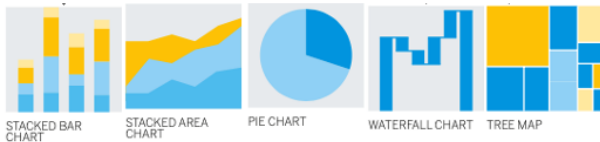
- **Comparisons** - Some keywords: before/after, categories compare, contrast over time, peaks, rank, trend, types, etc.



- **Comparisons and Distributions** - Some keywords: alluvial, cluster, distributed, from/to, plotted, points, spread, spread over, relative to, transfer, etc.



- **Compositions** - Some Keywords: components, divided up, group, makes up, of the whole, parts, percentage, pieces, portion, proportion, slices, subsections, total, etc.



- **Maps, Networks, and Logics** - Some keywords: cluster, complex connections, group, hierarchy, if/then, network, organize, paths, places, relationships, routes, structure, space, yes/no, etc.

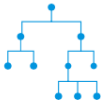
FLOW CHART



GEOGRAPHY



HIERARCHIES



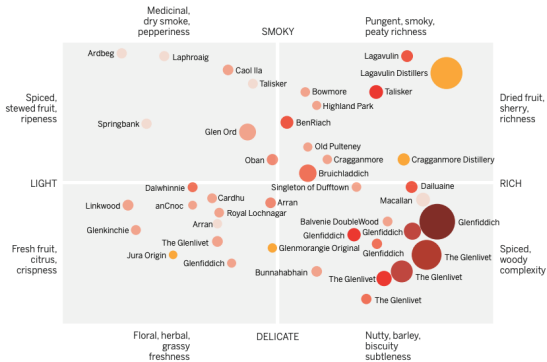
2 X 2



NETWORKS



Whisky data contains information of 42 records of whiskys brands. Six variables including age, cost, character, flavor, and region. Choose an appropriate chart to visualize the data.



SOURCE: THE MALT WHISKY FLAVOUR MAP. INSPIRED BY DIAGIO, BASED ON SVG CREATED BY UISCE BEATHA; PRICES FROM THEWHISKYEXCHANGE.COM

Data Source: <https://raw.githubusercontent.com/pengdsci/sta553/main/DatavizPrinciple/whisky.csv>

The guidelines for building persuasion into your charts.

- **Shift the context question** - before making a chart, asking yourself what you try to say, to whom, and where.
- **Emphasize and isolate** - shine a bright light on the most salient information. Limit the number of places an audience can focus. Move their eyes to where you want them to go.
- **Consider your reference points** - The ultimate form of isolation is to remove any information that doesn't directly support your point. Try to avoid multiple interpretations.
- **Point things out** - It doesn't take much to move someone's eyes. Pointers, demarcations, and simple labels signal to an audience what matters.
- **Lure** - Upending expectations can be powerfully persuasive. Evidence to the contrary is challenging and will foster discussion: Here's what you think our data looks like; here's how it actually looks.

Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding
Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

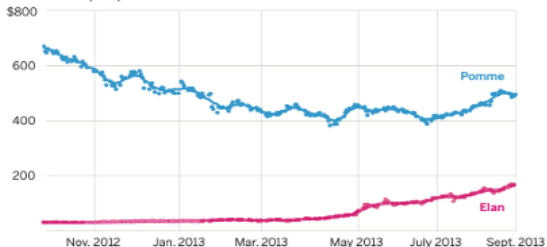
Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

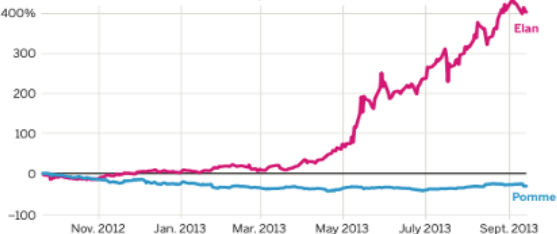
ELAN INC. STRONGER THAN POMME CO.

SHARE PRICE (USD)



ELAN INC. RISING, POMME CO. EBBING

% CHANGE IN STOCK PRICE SINCE SEPT. 17, 2012



Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding
Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color
Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts

Talk and Listen - Put aside your data and find someone in the domain to have a conversation to set your context. Address questions like

- Who is this for?
- What do you want them to do after seeing this?
- How will it be displayed?
- If you could show them only one thing, what would it be?
- Will it be surprising or affirming?

Sketch - As you're talking, start to sketch possible approaches. Go fast. The key is to keep moving. You want to be generative, creating ideas rapidly. Continue talking through the process, and as new ideas and visual words come up, jot them down.

Prototype - Whereas sketching is fast and open, prototyping is a bit slower and more deliberate. Use color purposefully. Sketching is generative; prototyping is iterative. Hone your chart until it approaches good.

Why Visualization

Agenda

Dataviz in past 50
Years

Visualization Process

Gestalt Principles

Gestalt Design
Principles

Pre-attentive
Attributes

Visual Encoding
Marks

Channels

Expressiveness &
Effectiveness

More on Channels

Color for Viz

Color Guidelines

Do's & Don'ts

Color Deficiency

Controlling Color

Crafting for Clarity

Choosing Chart Type

Practicing Persuasion

Steps for Good Charts