

A Primer on Data Visualization



Kokleong Choy
H&B Analytics, ACIA (Singapore)



Why Data Artisans Are The New Data Scientists

Number aren't just numbers. They also tell stories.

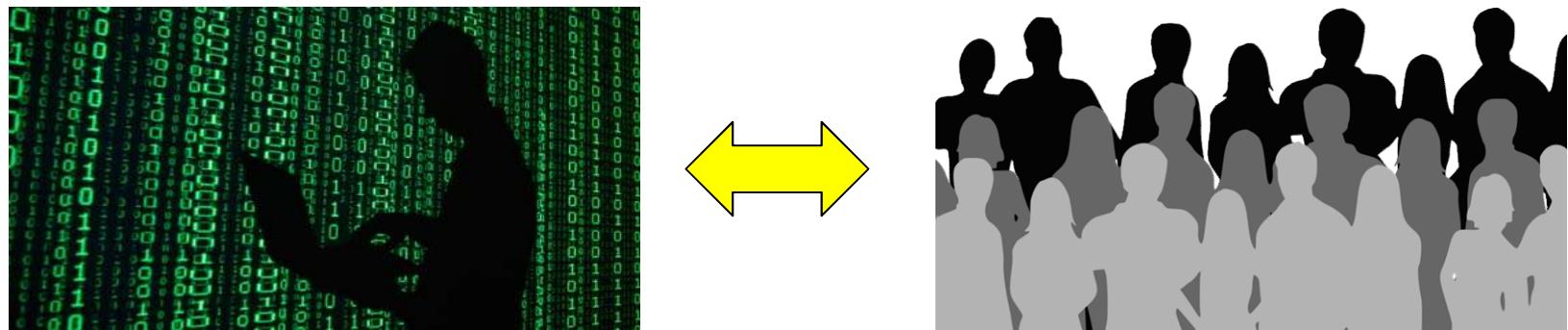
- Data artisans are employees who possess a **blend of technical skills** and **business acumen** that enable them to extract **actionable insight** from the huge volumes of data that exist, demonstrating that businesses don't always need a data scientist to interpret data effectively.

Guidelines on Creating Visualizations

- Come up with a good story
- Identify the message to be conveyed
- Select the visualization type
- Apply Principles of Perception
- Leverage preattentive processing.

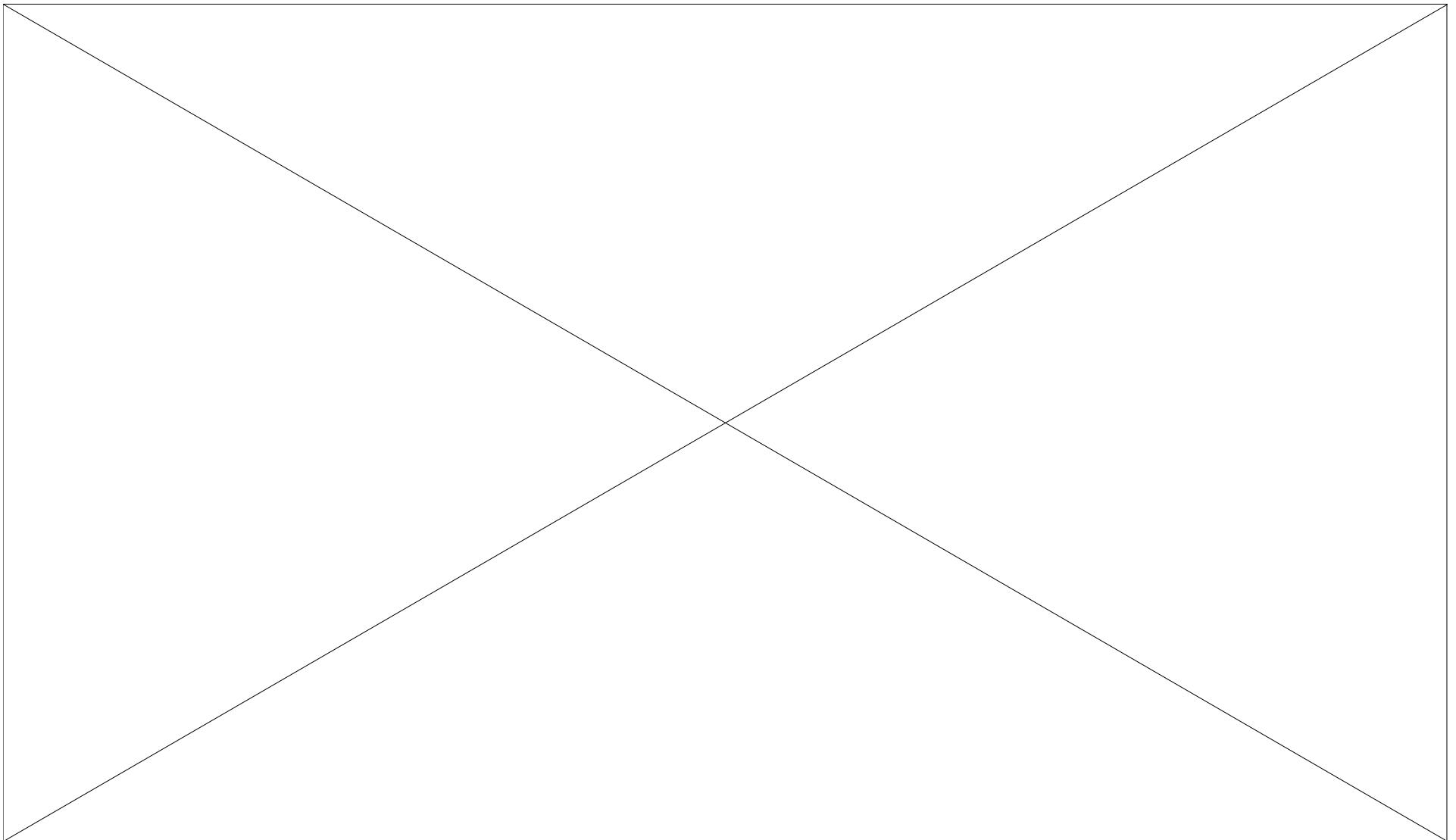
DATA STORY

Data is what you show, the story is what the audience remembers



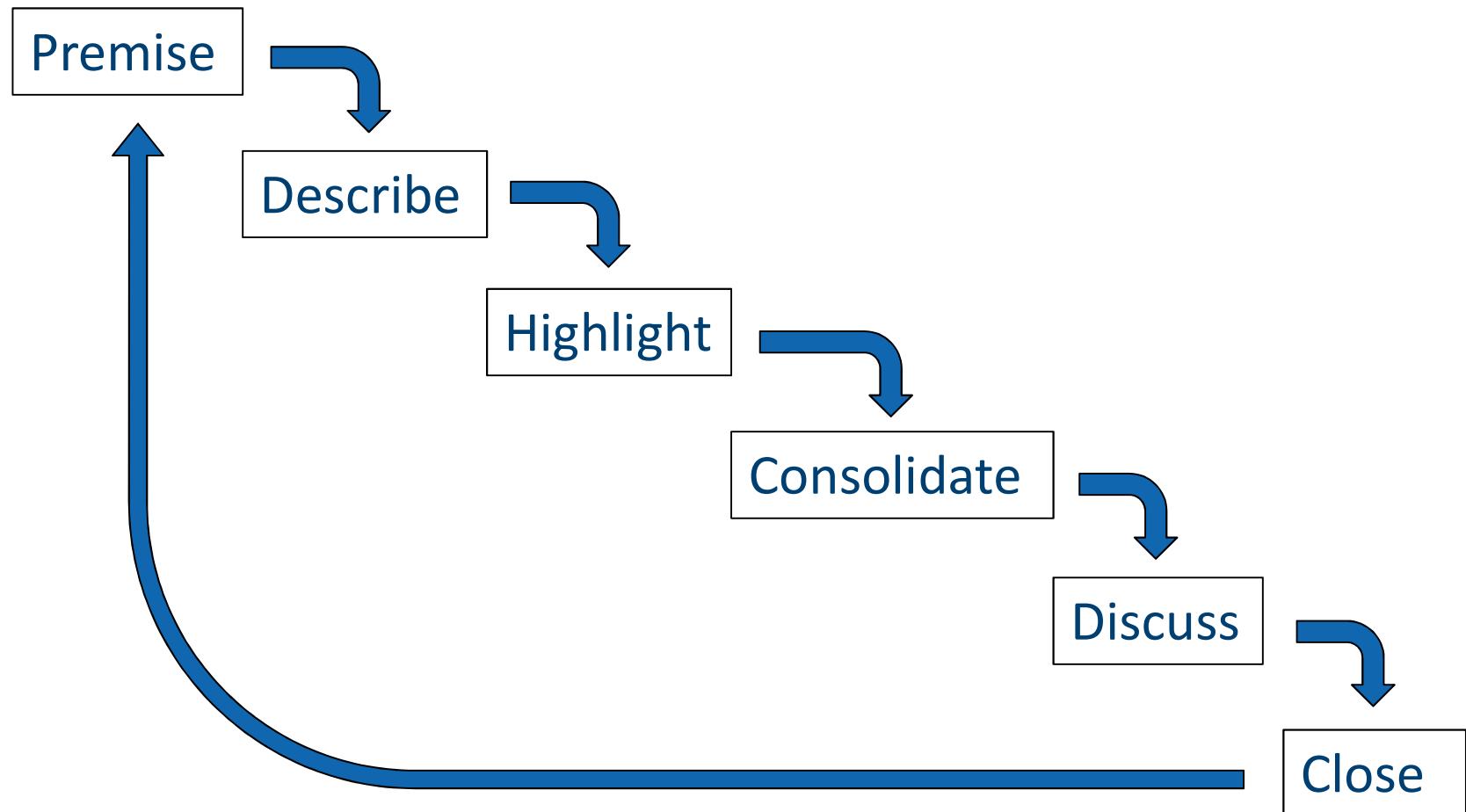
A good story requires understanding both data and audience and connecting one to the other

What's a good story? 200 Countries, 200 Years, 4 Minutes



Link: <https://www.youtube.com/watch?v=jbkSRLYSojo>

How to tell a good story



Exploratory vs Explanatory Data Visualization

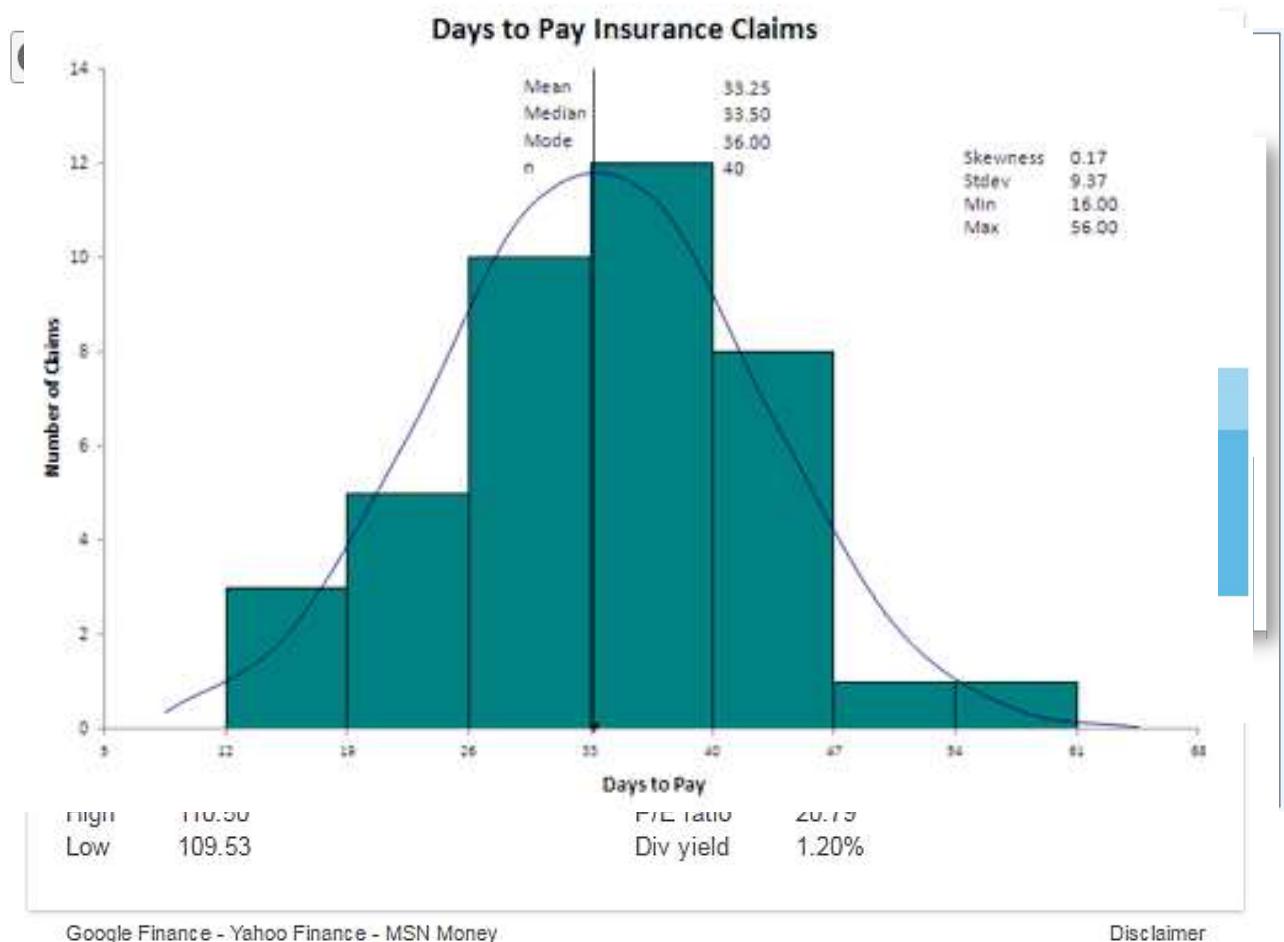
 Exploratory	 Explanatory
New data	Familiar data (to designer)
Multiple chart views	One best view
Little effort to pivot chart views	High interactivity in one display
Focus is on switching perspectives & making sense of the data	Focus is on communication
Story-in-progress	Clear and decided message
Internal facing	External facing

Guidelines on Creating Visualizations

- Come up with a good story
- Identify the message to be conveyed
- Select the visualization type
- Apply Principles of Perception
- Leverage preattentive processing.

8 Types of Quantitative Messages

- Time-Series
- Ranking
- Part-to-whole
- Deviation
- Frequency Distribution
- Correlation
- Nominal Comparison
- Geographic / Geospatial

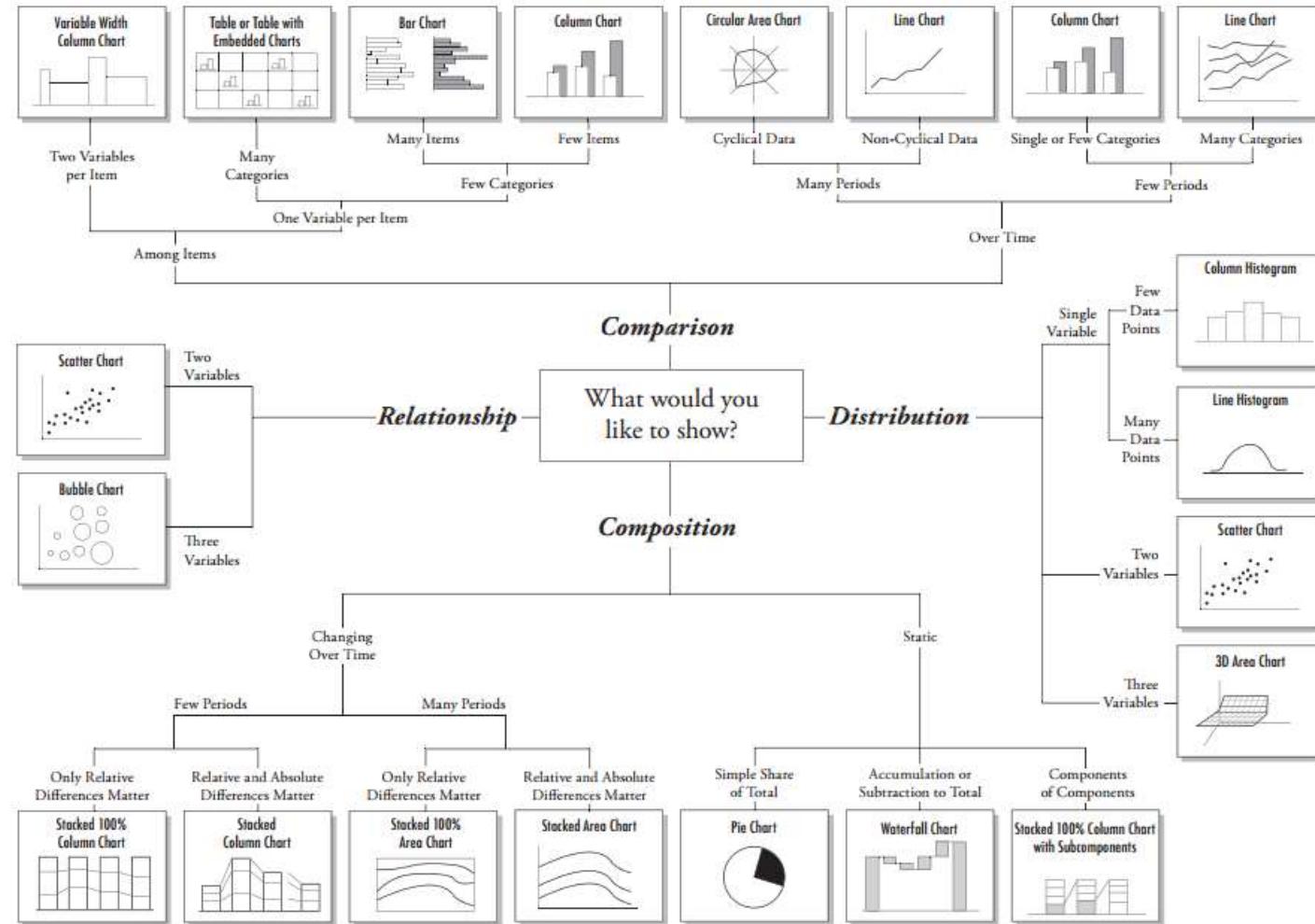


Guidelines on Creating Visualizations

- Come up with a good story
- Identify the message to be conveyed
- Select the visualization type
- Apply Principles of Perception
- Leverage preattentive processing.

Charting Decision Tree – (These are suggestions only!)

Chart Suggestions—A Thought-Starter

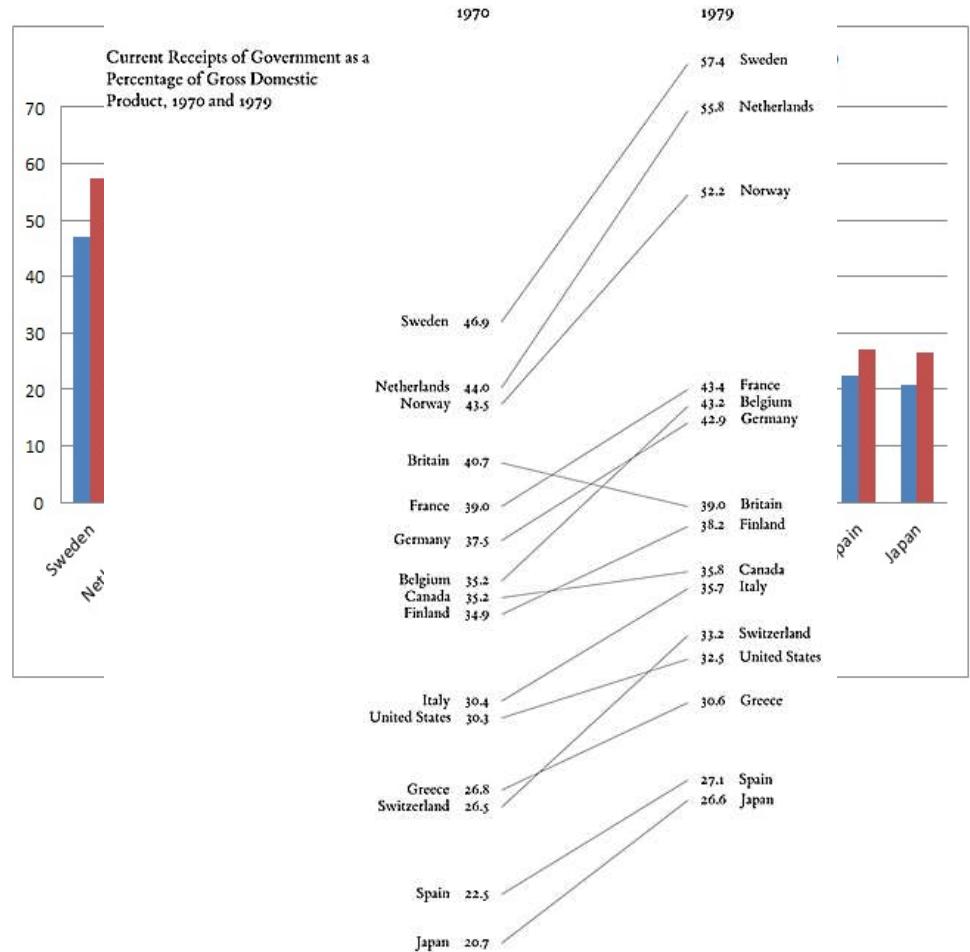


Link: <http://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf>

Choose carefully. Every visualization's emphasis is different.

3 representations of the same data

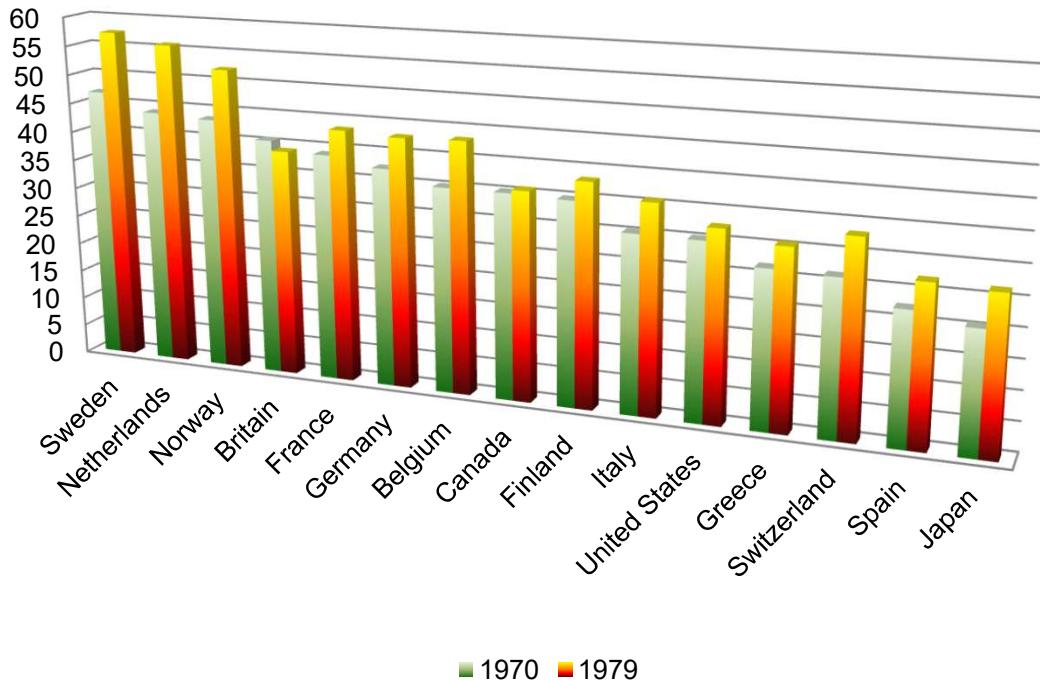
- A table provides a high level of detail with particular emphasis on the outlier.
- A standard column chart permits visual between-year and between-country comparisons.
- A slope chart does both, and eliminates non-data visual elements (e.g. axes).



Tufte, Edward. *The Visual Display of Quantitative Information*.
Cheshire, Connecticut: Graphics Press; 1983; p. 158

Maximize the “data-ink” ratio

Current Receipts of Government as a Percentage of GDP



Unnecessary use of:

- 3D bars
- Gradient fill
- Too dark grid lines
- Too many divisions on y-axis

No extra information is conveyed using these non-data elements.

Guidelines on Creating Visualizations

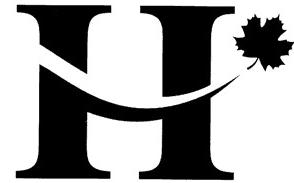
- Come up with a good story
- Identify the message to be conveyed
- Select the visualization type
- Apply Principles of Perception
- Leverage preattentive processing.

Borrowing from Gestalt Psychology

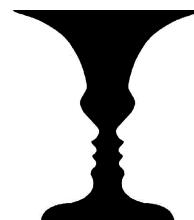
- Similarity / Anomaly



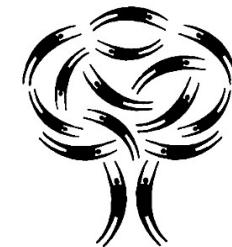
- Continuation



- Figure/Ground



- Proximity

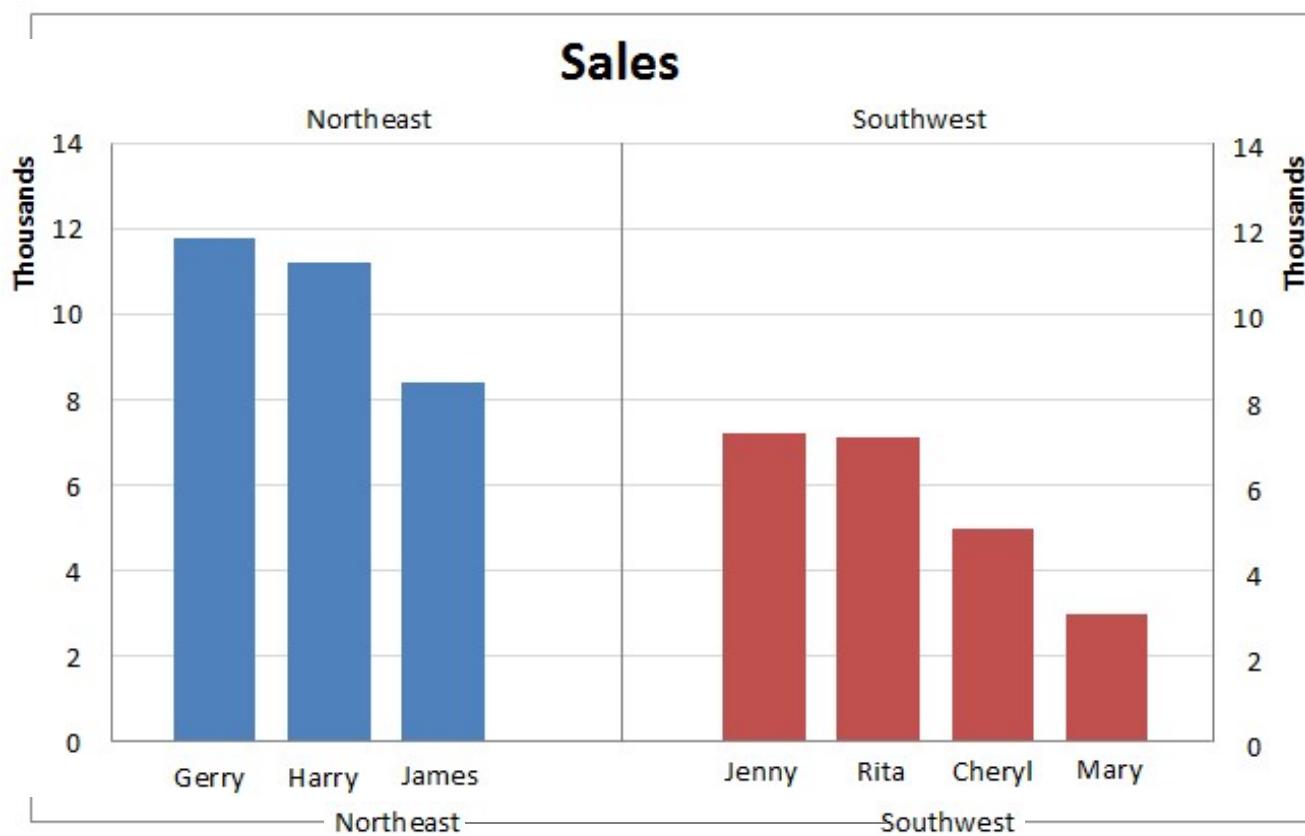


- Closure / Enclosure



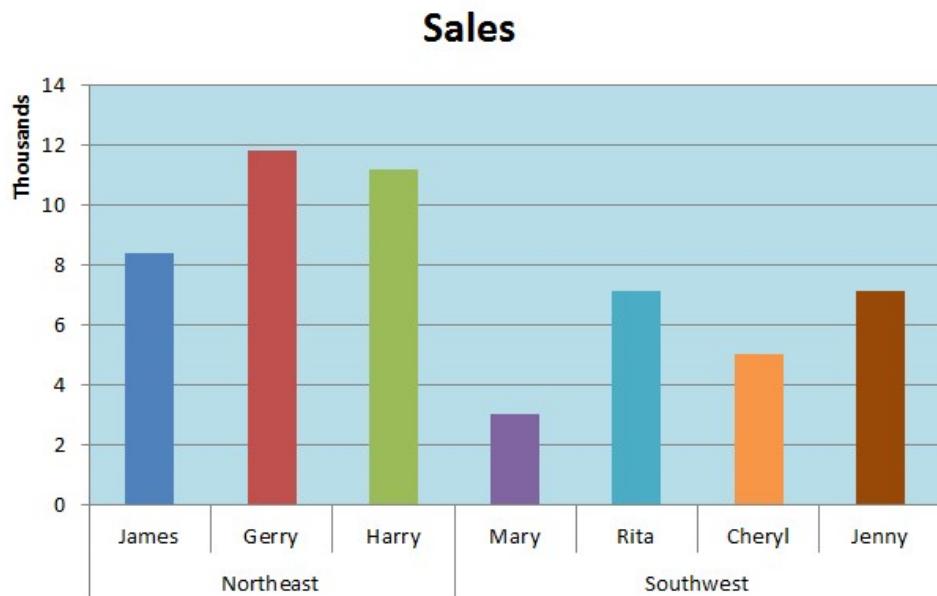
- Common Fate
- Parallelism
- Connectedness

Gestalt Principles in Practice



- Similarity
- Continuity
- Figure / Ground
- Eliminate / Fade Grid Lines
- Proximity
- Enclosure

Before and after applying Gestalt principles



Gestalt principles provide an organized framework for thinking about how to refine a visualization.

Guidelines on Creating Visualizations

- Come up with a good story
- Identify the message to be conveyed
- Select the visualization type
- Apply Principles of Perception
- Leverage preattentive processing.

Let's talk about Preattentive Processing

Now try again. How many 5's are there in this block of numbers?

9873497902756479028972862409240603707057027907280
3208029007302501270237008374082078720272007083247
8026027037937757097073779706674620970947027809279
7970972309723097959275092727979873497260802747903

Preattentive Processing: Definition

Preattentive processing of visual information is performed automatically on the entire visual field detecting basic features of objects in the display.

It is high bandwidth, massively parallel processing as opposed to low bandwidth, serial processing.

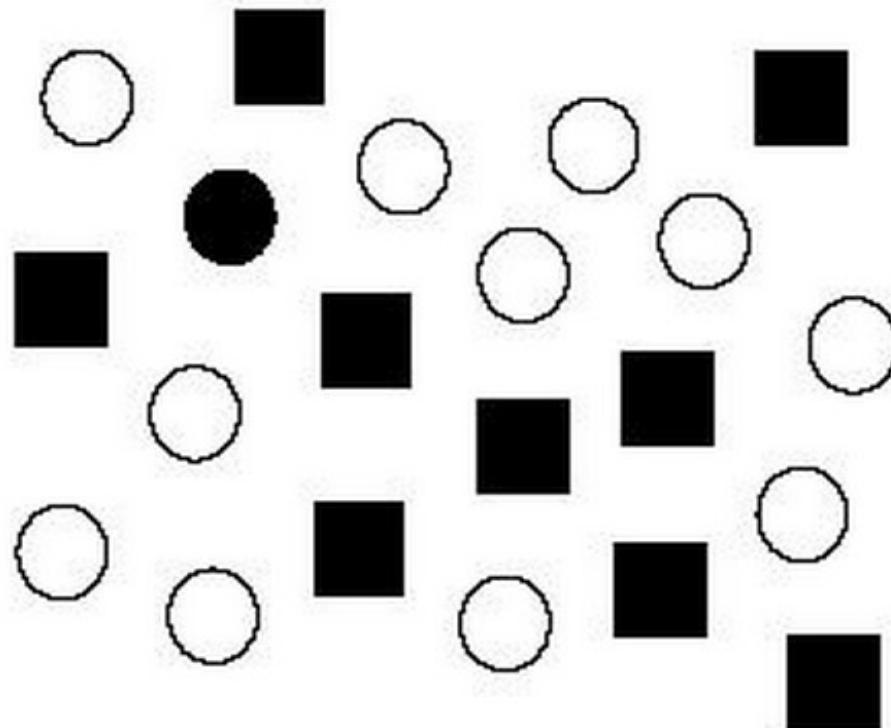
In English, that means you could

- Tell quickly at a glance how many 5's there were...
- Without looking at the other numbers individually.

Without the highlighting, you had to look at each number individually to figure out if it was a 5 or not.

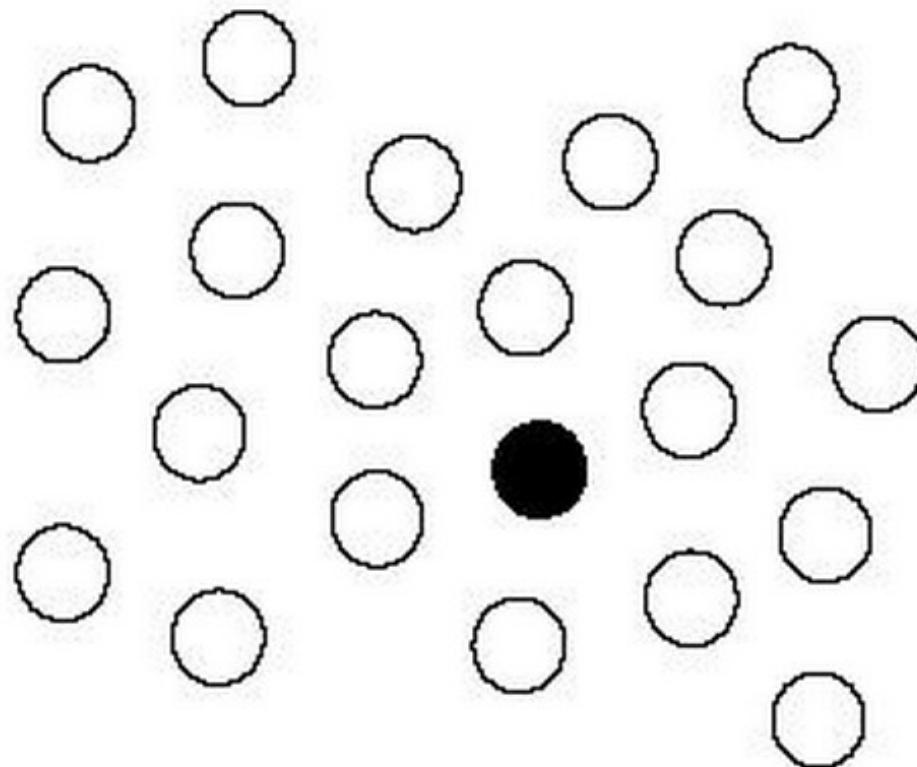
A few more examples of preattentive processing

Raise your hand once you find the black filled circle

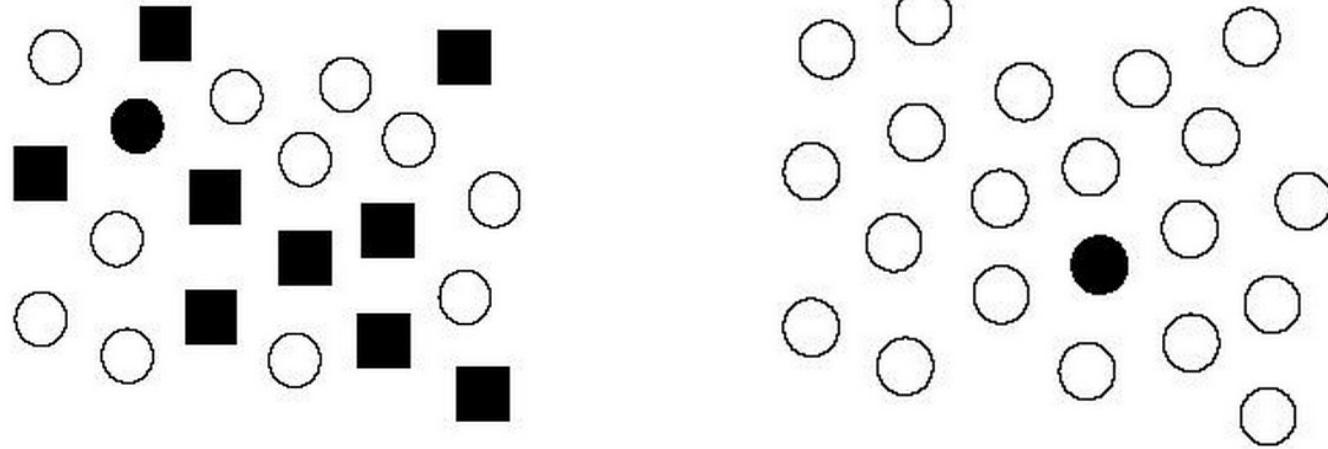


A few more examples of preattentive processing

Raise your hand once you find the black filled circle



Lower cognitive load = faster perception

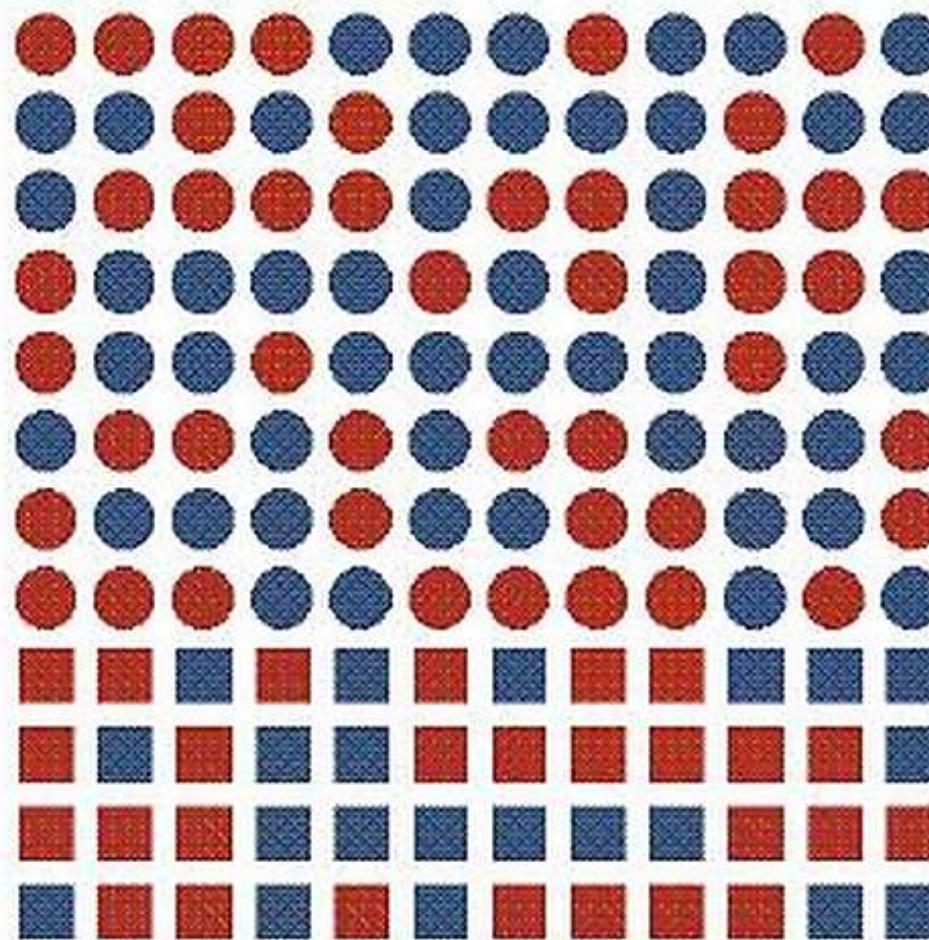


Perceiving the black circle was faster on the right because only color was needed in discriminating between target and non target elements.

Both color and shape needed to be considered on the left.

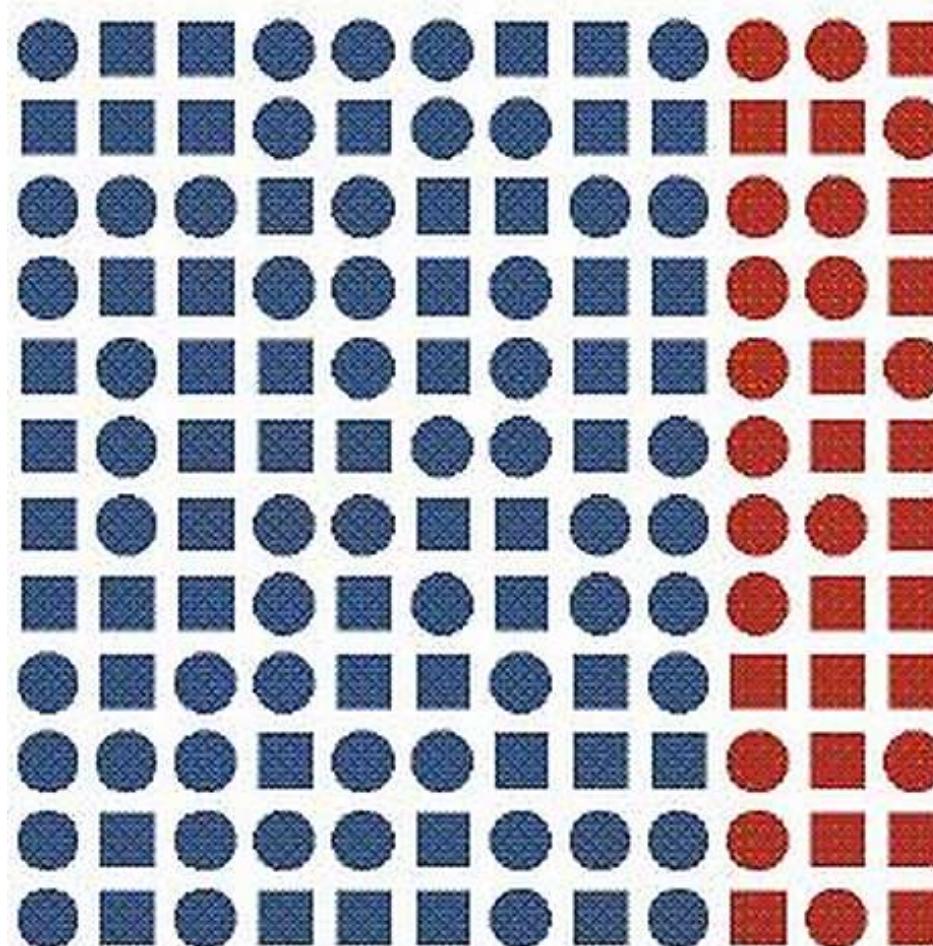
A few more examples of preattentive processing

Raise your hand once you see the “boundary”.

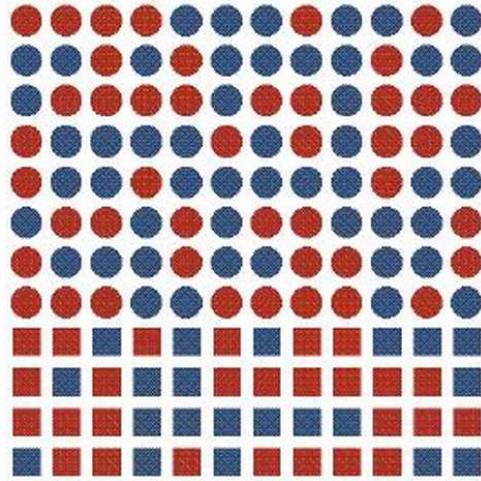


A few more examples of preattentive processing

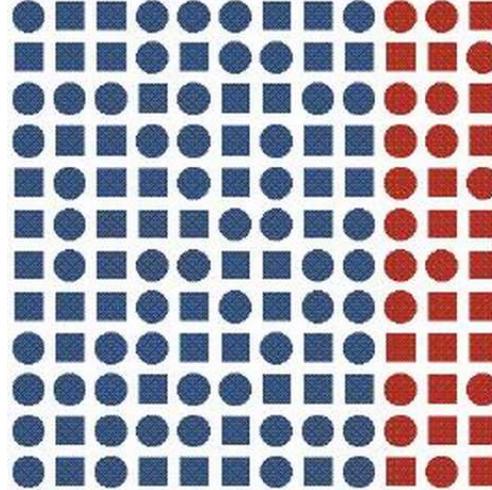
Raise your hand once you see the “boundary”.



Preattentive attributes are not created equally. Each is unique.



Demarcated by shape
Colors are mixed

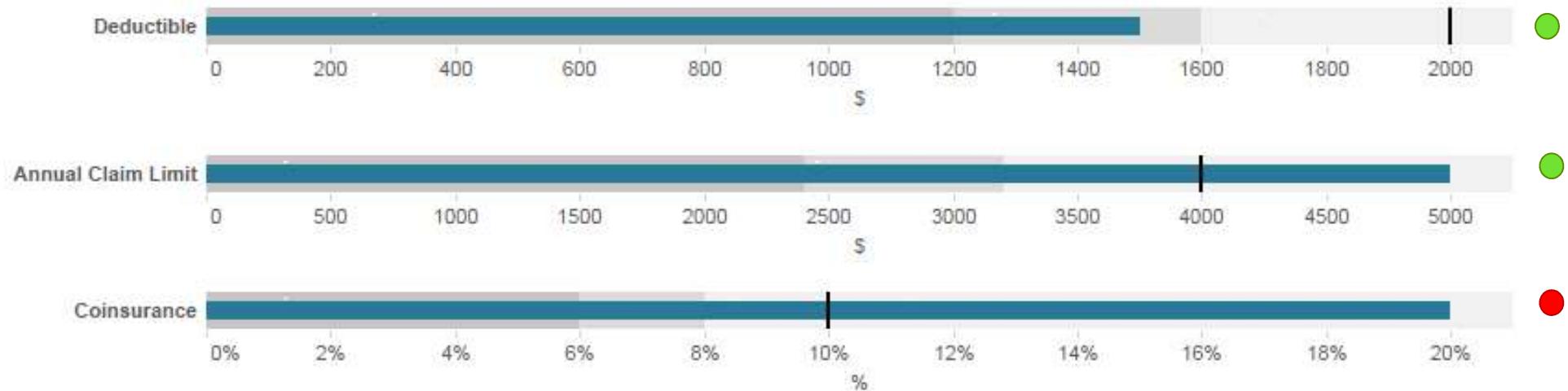


Demarcated by color
Shapes are mixed

Both color and shape are preattentive...but humans respond more strongly to color as an attribute for **qualitative** discrimination.

Color should be preferred, and **reserved**, for qualitative discrimination.

An example of using color judiciously



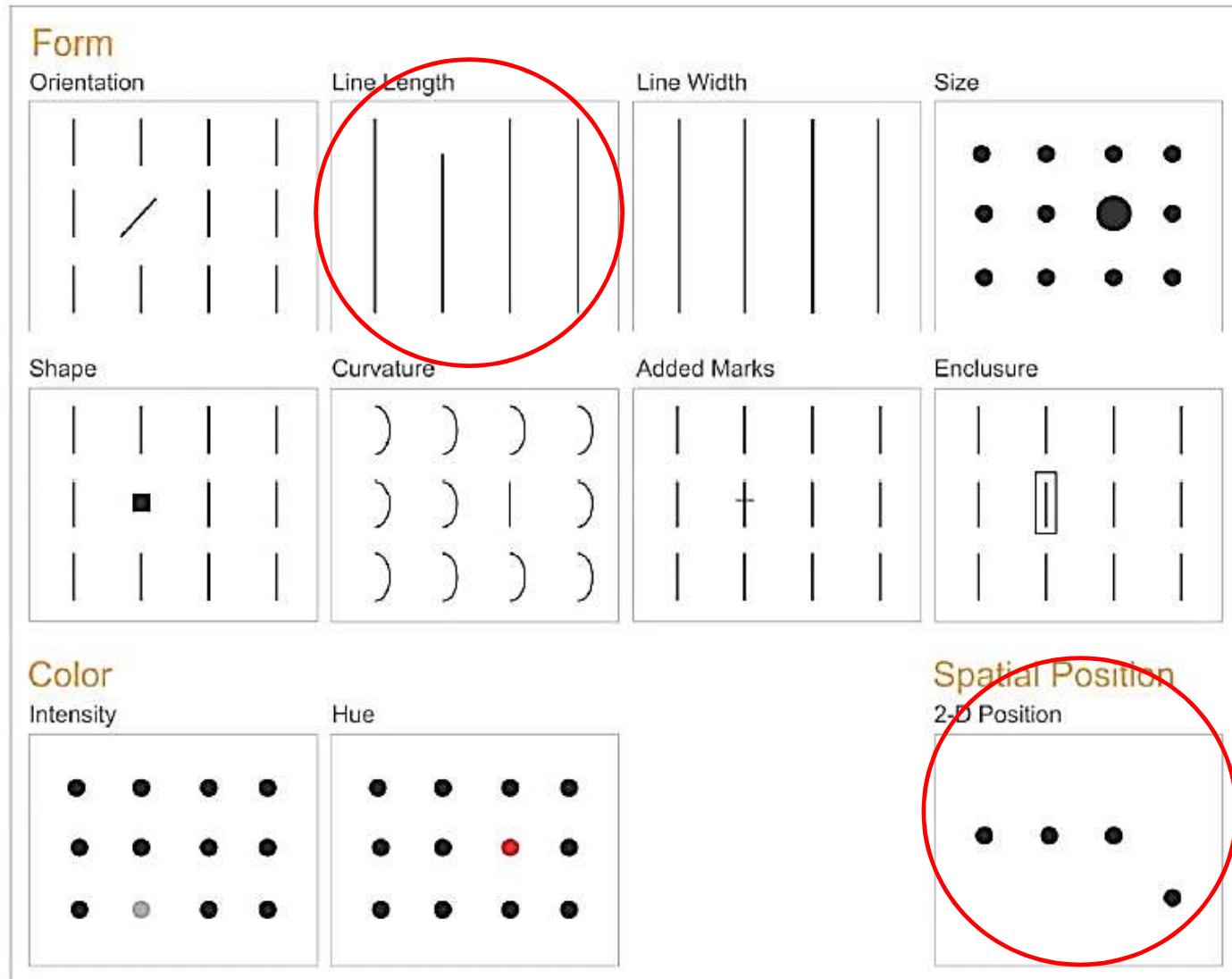
A health plan's design attributes are compared against the market norm.

Green dots indicate attributes that are **more** competitive in the market.

Red dots indicate attributes that are **less** competitive in the market.

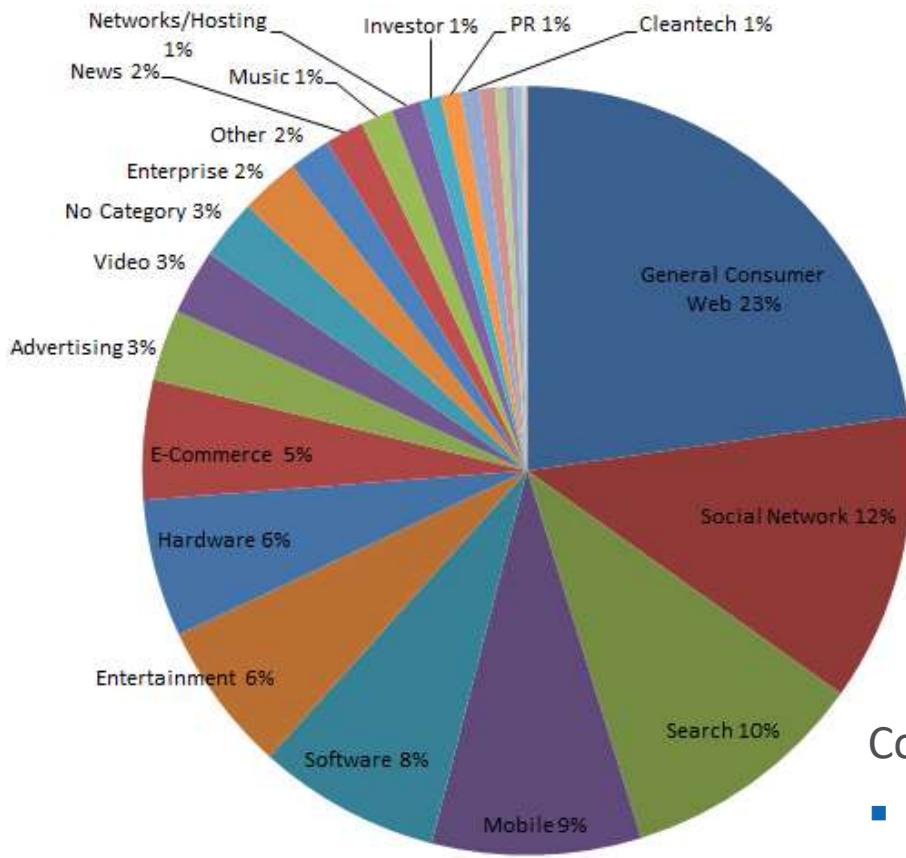
Using only black and grey in the chart bars improves contrast, making it easier for the viewer to immediately see which plan attributes are competitive vs the general market.

Preattentive Attributes for Data Visualization



Link: http://www.perceptualedge.com/articles/ie/visual_perception.pdf

Encoding quantitative information preattentively



- Consider the pie chart: easy to tell large slices from small slices
- But not easy to tell how **much** larger
- Quantitative differences are best encoded using line length or spatial position, e.g. bar or scatter charts.

Color is also used poorly here:

- Too many categories for color to work well
- Shades of the same color lead to poor contrast

Exceptions to using line length and spatial position to encode quantities

- Bubble charts, as in the video shown earlier.
 - Both axes already used in scatter plot for life expectancy versus year
 - Using size of bubbles to encode population was better than showing a third axis on a 3D chart, which is generally not recommended.

- Color intensity can be used to encode quantities, as in well-designed heat maps.

Leverage color's preattentive power appropriately

- Hue:

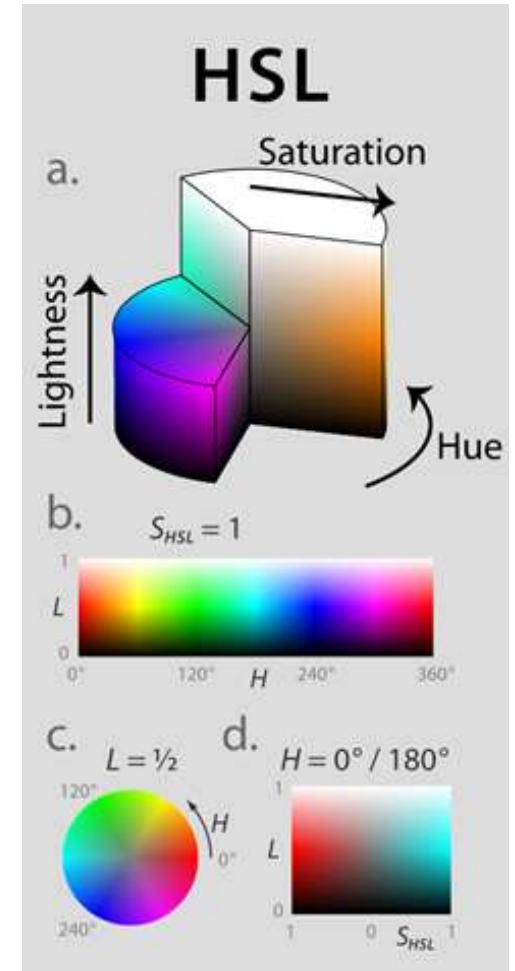
Red versus **Blue**

- Saturation:

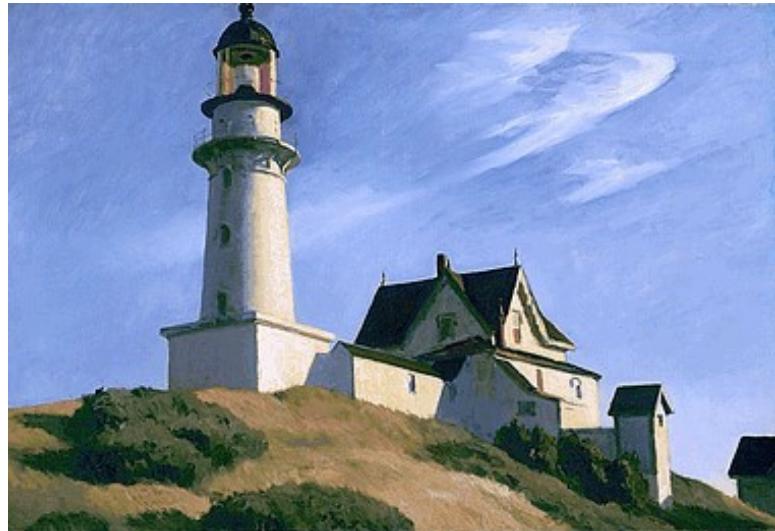
Orange versus **ORANGE**

- Lightness / Luminosity:

Light Green versus **Dark Green**



Lightness is the component that carries the most information.



"Lighthouse at Two Lights"
Edward Hopper, 1929
New York Metropolitan Museum of Art



Good heat map designers understand the importance of lightness...

Even without hue, lightness is useful for conveying quantitative info

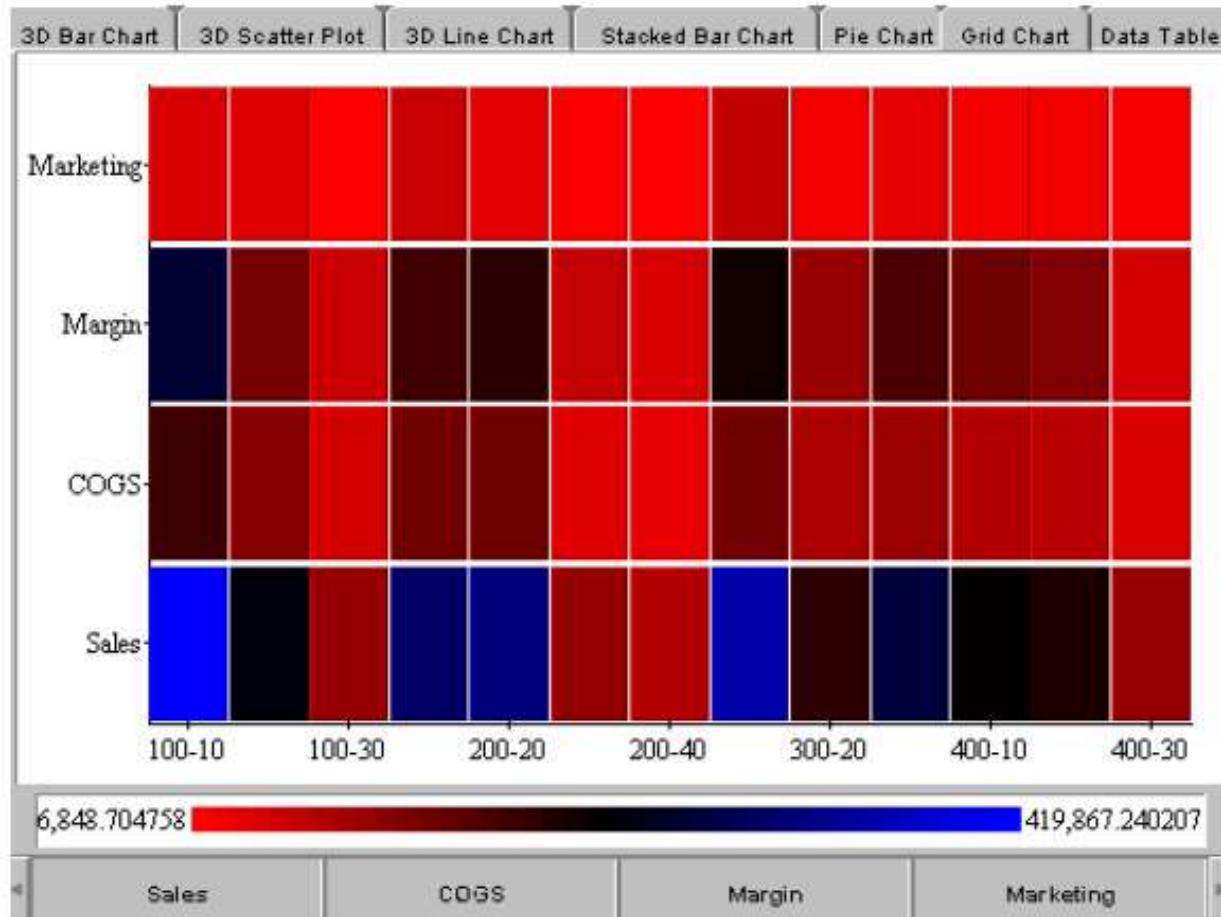
- Grey Scale works well even without hue.



- This palette deviates from the ideal (for reasons we will discuss later).
- It remains readable because **lightness**, the component of color perceived most strongly, changes continuously.



While bad heat maps use hue or saturation to encode quantitative info.



It is very difficult to tell apart dark red, black and dark blue. The scale becomes useless toward the middle.

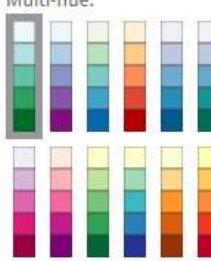
Choosing Appropriate Color Schemes, Advice from Cartography

Number of data classes: 3 ▾ [i](#)

Nature of your data: [i](#)

sequential diverging qualitative

Pick a color scheme:

Multi-hue: 

Single hue: 

Only show: [i](#)

colorblind safe print friendly photocopy safe

Context: [i](#)

roads cities borders

Background: [i](#)

solid color terrain

color transparency 

how to use | updates | downloads | credits

COLORBREWER 2.0
color advice for cartography

3-class BuGn

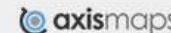
   

EXPORT

HEX ▾

#e5f5f9
#99d8c9
#2ca25f

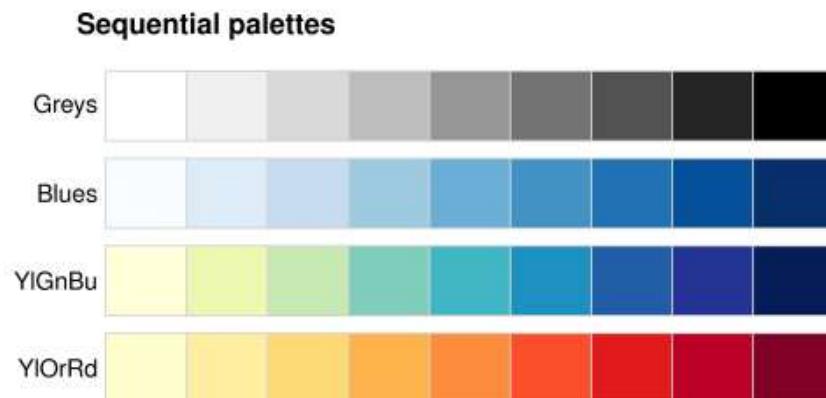
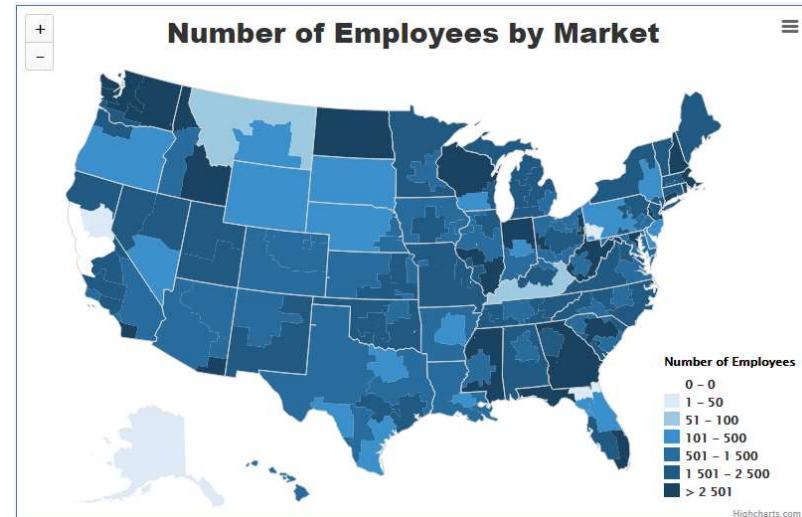
© Cynthia Brewer, Mark Harrower and The Pennsylvania State University
[Source code and feedback](#)
[Back to Flash version](#)
[Back to ColorBrewer 1.0](#)



Different Data Types, Different Color Palettes

■ Sequential Data

- Ordinal; low to high.
- Use a single hue and vary lightness and saturation continuously.
- Prototypical situation for a heat map.



Color Palettes for Sequential, Divergent and Qualitative Data

■ Divergent Data

- Ordinal, but equal emphasis on extremes at both ends of the data range.
- Use 2 contrasting hues
 - Darkest at both extremes;
 - Decrease in saturation as they approach the middle;
 - Lightness increases to white in the middle.



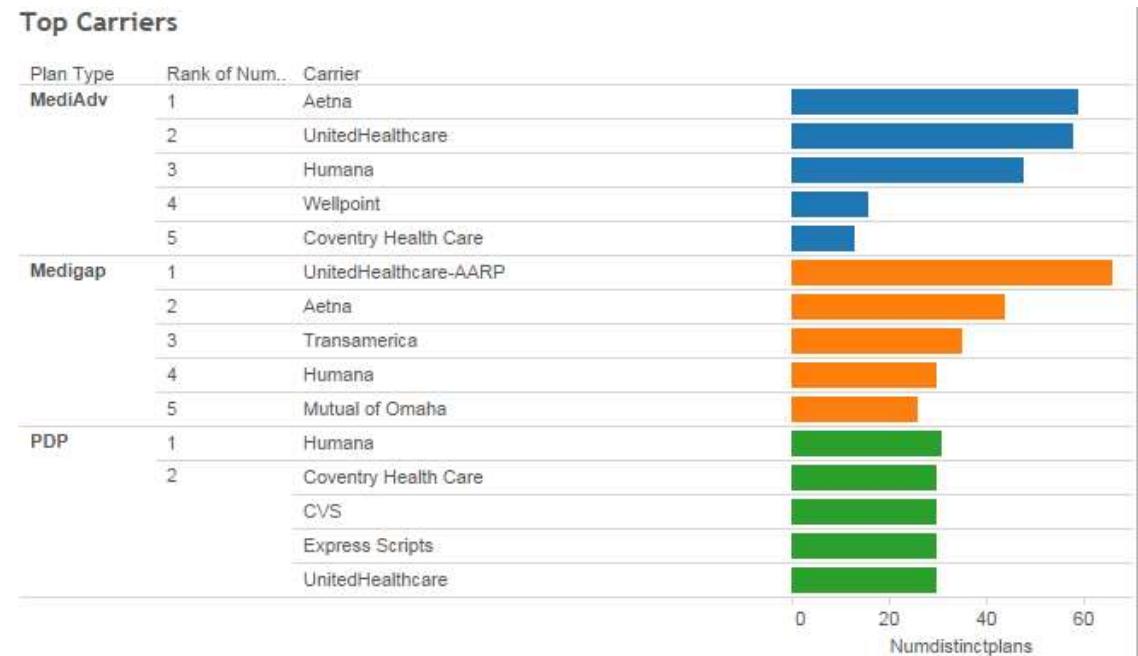
Diverging palettes



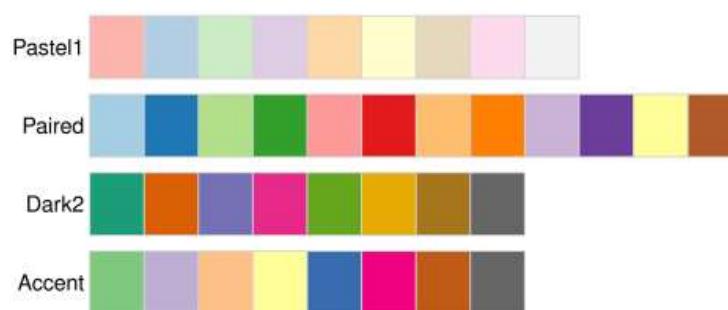
Color Schemes for Sequential, Divergent and Qualitative Data

▪ Qualitative

- Categorical or nominal data.
- Hues chosen should be contrasting.
- No magnitude differences should be implied, so normalize for lightness and saturation.



Categorical palettes



Recap and Take Home Message

- Discover and tell your story!
- Use your understanding of the data to pick the appropriate visuals.
- Organize and curate each element of your charts.
- Be **mindful** of the brain's processing power, and leverage accordingly.

Well-designed software like Tableau often makes good default choices, but software is not infallible. Knowing the core principles prevents software from misleading you.

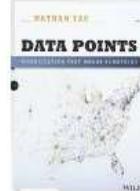


Resources

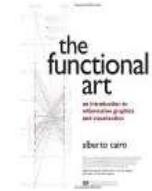
- <http://www.storytellingwithdata.com/>
- <http://www.perceptualedge.com/>
- <http://www.colorbrewer2.org/>
- <http://extremepresentation.typepad.com/>



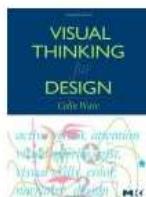
[Show Me the Numbers: Designing Tables and G...](#)
by Stephen Few
\$27.76



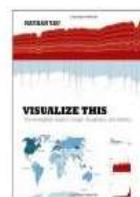
[Data Points: Visualization That Means Somet...](#)
by Nathan Yau
\$28.00



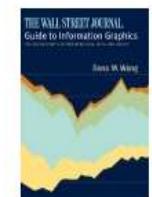
[The Functional Art: introduction to info...](#)
by Alberto Cairo
\$32.47



[Visual Thinking: for Design \(Morgan Kaufman...](#)
by Colin Ware
\$42.40



[Visualize This: The FlowingData Guide to De...](#)
by Nathan Yau
\$24.51



[The Wall Street Jou Guide to Informatio...](#)
by Dona M. Wong
\$20.85



[Beautiful Visualization: Looking at Data th...](#)
\$50.48



[The Visual Display of Quantitative Informat...](#)
by Edward R. Tufte
\$29.62



[Resonate: Present \ Stories that Trans...](#)
by Nancy Duarte
\$17.42