



LEARN.
CONNECT.
ELEVATE.

Capstone Project
Nabil Beitinjaneh
Alejandro Gutierrez Lopez

With Special Thanks to **Tableau** who provided training and hands-on material



McGill

School of
Continuing Studies

[mcgill.ca
/continuingstudies](http://mcgill.ca/continuingstudies)

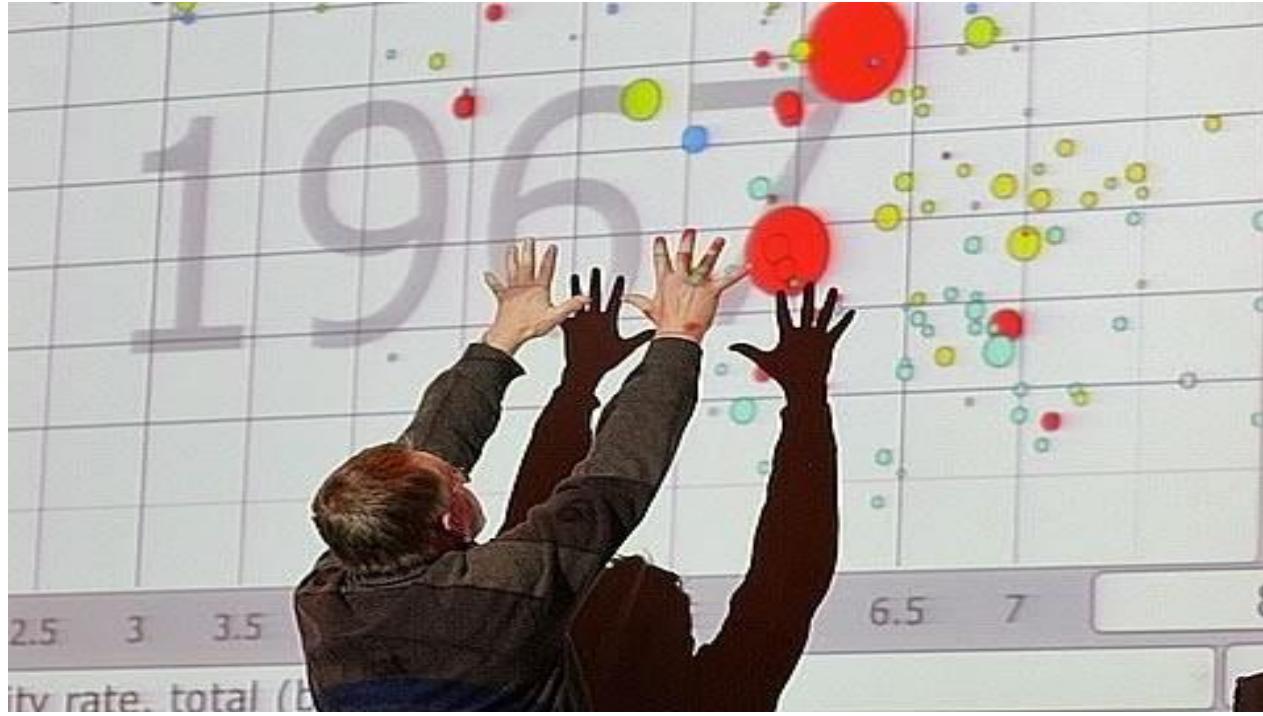
Agenda

- Why Storytelling and Visualization?
- Introduction to data exploration and visualization
 - Exploratory analysis
 - Explanatory analysis
- Which graph?
- Visual Perception
 - On the use of Color
- The 4Cs and examples
- Some hands-on with Tableau

Why visualization and Storytelling?

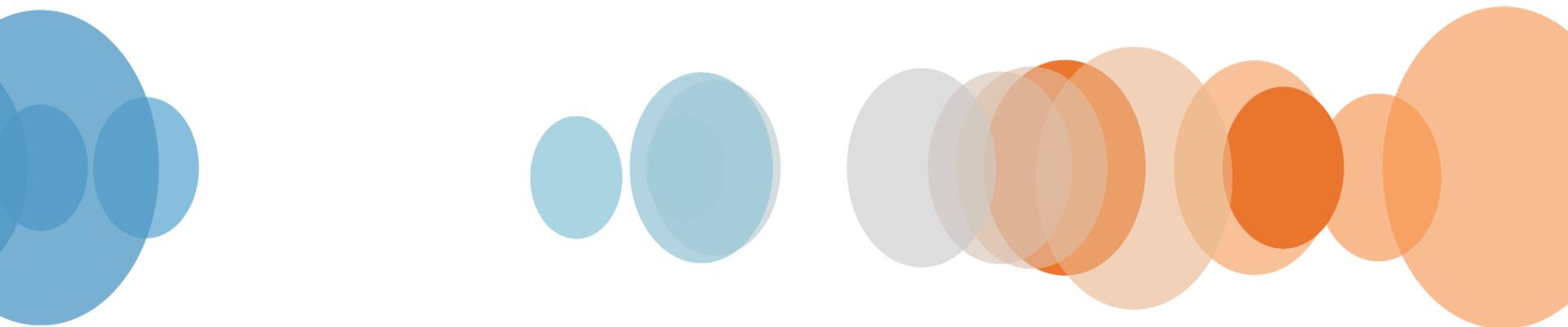
Remember “*The Best Stats you will ever see*”...

~ Hans Rosling – Ted Talk



[Youtube](#)

Pre-attentive Attributes



Visual Perception



Visual Perception

7

3

5

7

8

6

8

5

7

4

1

6

4

7

9

3

1

9

5

4

8

0

8

9

7

8

9

2

6

7

9

2

0

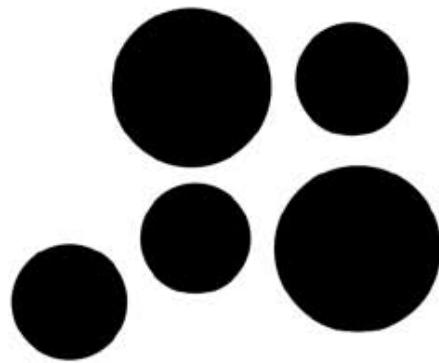
1

2

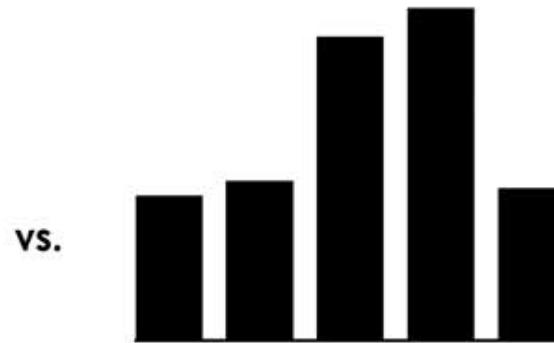
7

Using area vs using length

Using area as visual cue



Using length as visual cue

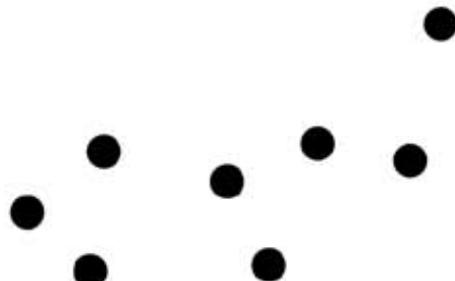


vs.

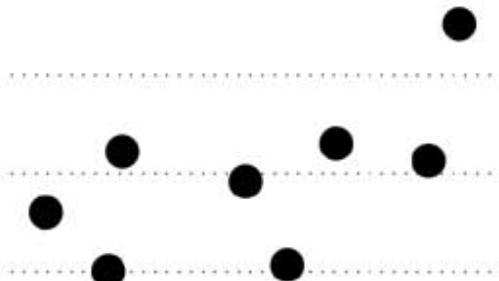
[Yau 2013]

Showing data points only vs additional visual elements

Showing data points only



Additional visual elements

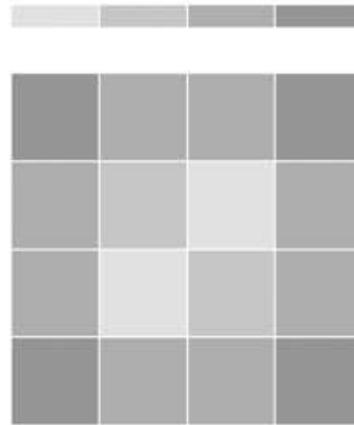


vs.

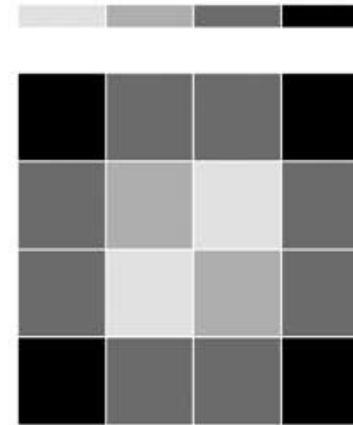
[Yau 2013]

Narrow color scale vs wider color scale

Narrow color scale

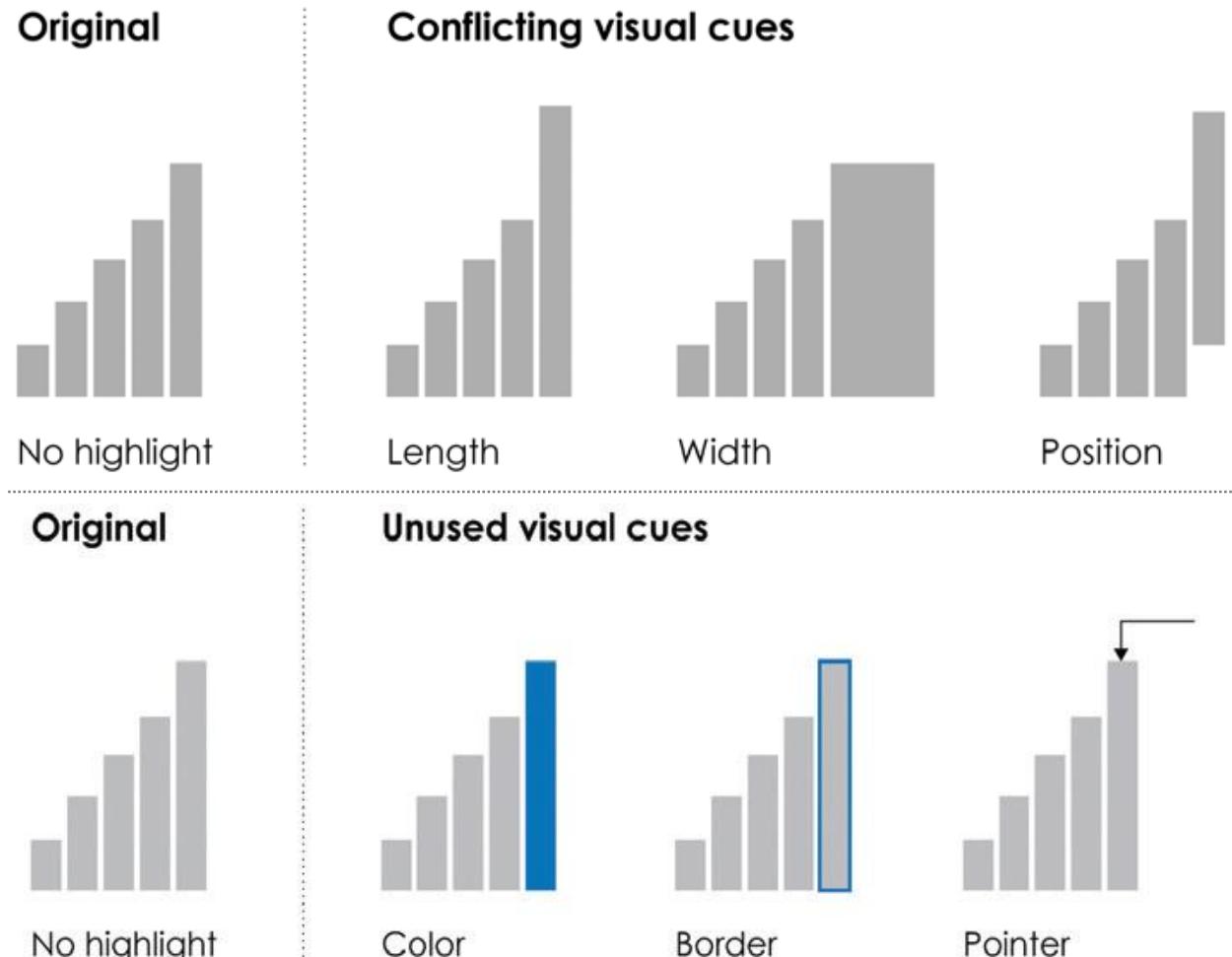


Wide color scale



vs.

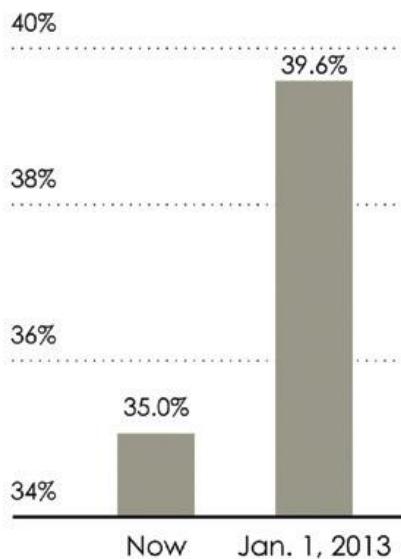
Visual cues to display relevance



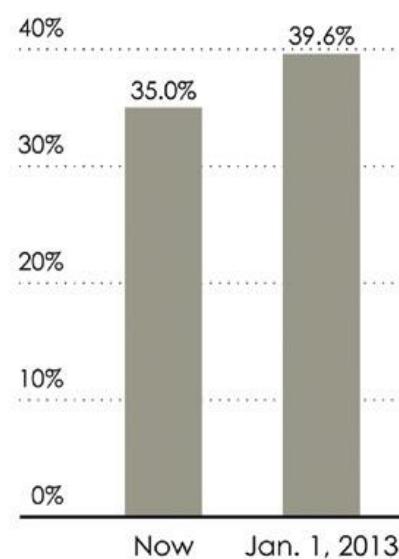
[Yau 2013]

Scale

Axis starting at 34 percent



Axis starting at 0 percent



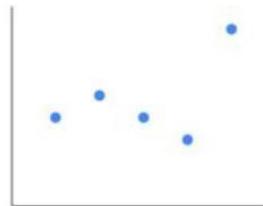
Incorrect scale

[Yau 2013]

Common visual displays

91%

Simple text



Scatterplot



Vertical bar



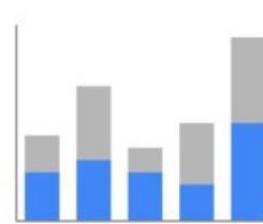
Horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

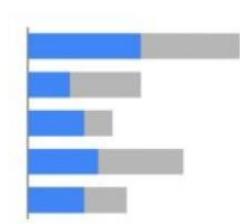
Table



Line



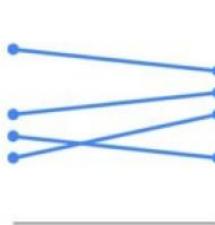
Stacked vertical bar



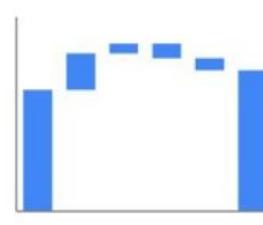
Stacked horizontal bar

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

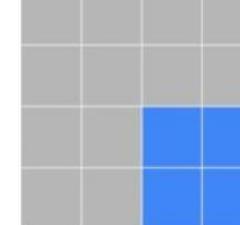
Heatmap



Slopegraph



Waterfall



Square area

[Nussbaumer 2015]

What is data visualization?

- Data visualization is the graphical display of abstract information for sense-making or data analysis, and communication in a way that leads to **understanding for action**.
- Its goal is to visualize data in a way that leads to understanding.

[Stephen Few, 2014]

- Data visualization is a fundamental product from the visual analytics process

Why data visualization?

- People make better decisions when they're based on understanding.
- For information to be understood, it must often be presented in visual form because patterns, trends, and outliers require a picture for the human brain to see and comprehend.
- Data visualization is essential for:
 - Data exploration and understanding
 - Communicating data
 - Making better decisions

Let's do some analysis

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

For each series, provide the mean of x, the mean of y, the variance of x and the variance of y

Now let us graph each series...

How similar are they?

Two different but complementary stages

- Exploration
 - Figuring out what could be relevant
 - Test different combinations
 - Incorrectly, we usually give our exploratory analysis to the user
- Explanation
 - Performed after you find what you want to say
 - Rethink the visualization generated during data exploration based on the user's needs

Explore

- Limits
- Value ranges
- Data types
- Inconsistencies
- Errors
- Missing data
- Trends and patterns
- Sort and organize the data
- Apply formulas (Average, mode, etc.)
- Make comparisons

Explanatory analysis – Audience Analysis

- Who is your audience?
 - Be specific
 - Create different communications for different audiences
 - Who is the decision maker?

[Nussbaumer 2015]

Explanatory analysis: Understand the context

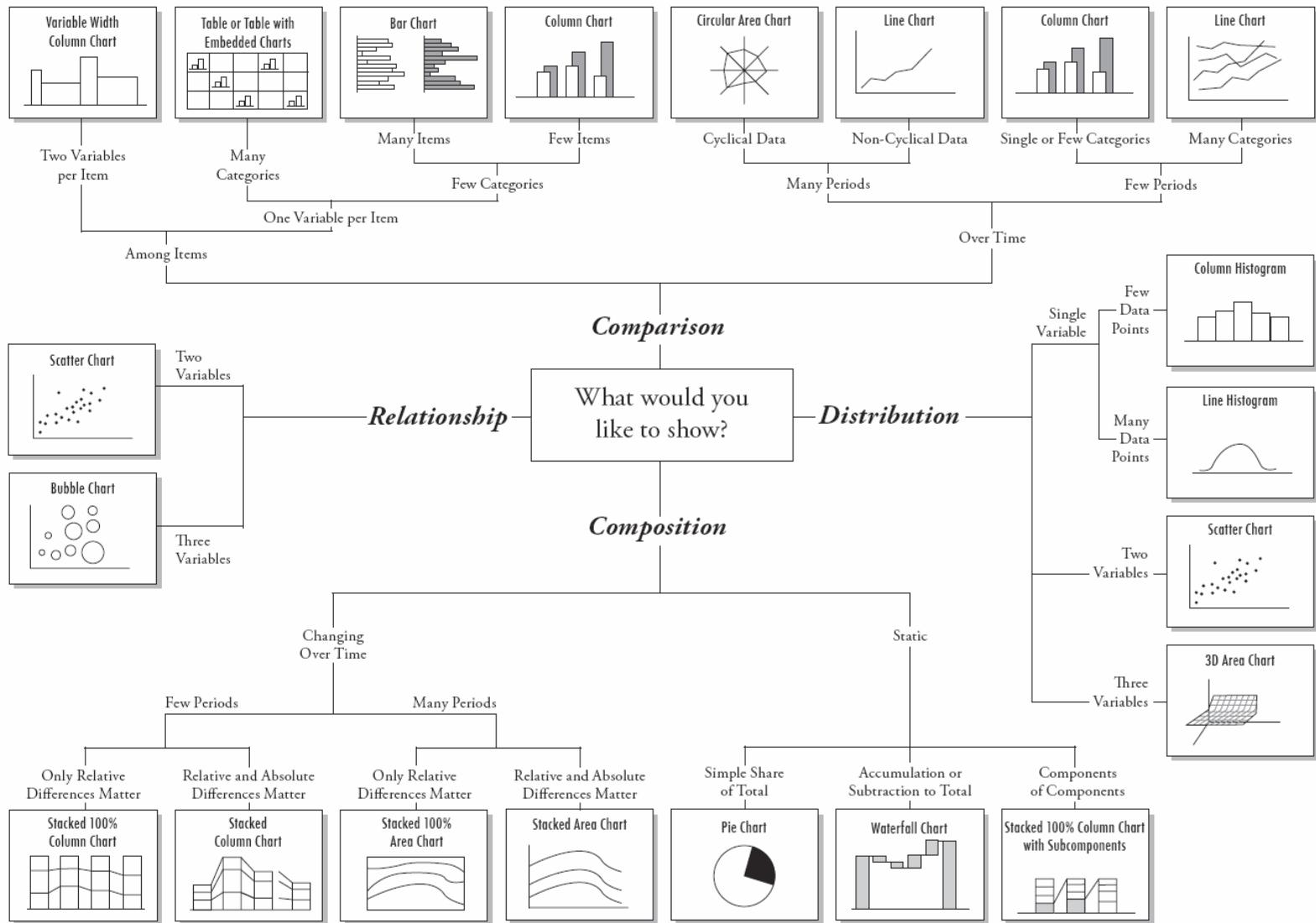
- What do you need them to know or do?
 - Why they should care?
 - Consider that you might know more about the issue than your audience
 - If an action recommendation is not possible consider opening a discussing
- What data is available to make your point?
- You might need to consult with your client to understand the context

[Nussbaumer 2015]

Explanatory analysis – which graph?

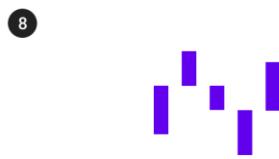
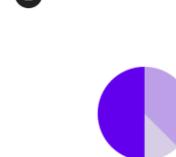
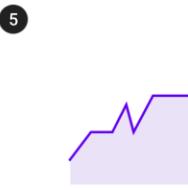
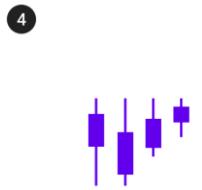
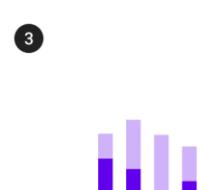
- Choose the appropriate visual display
 - If possible avoid displaying tables in live presentations
 - Color saturation provides visual clues helping the brain (Tip: Excel conditional formatting)
 - Avoid using 3D visualizations displayed on 2D formats

[Nussbaumer 2015]



A, Abela [2009] <http://www.mymarketresearchmethods.com/types-of-charts-choose/>

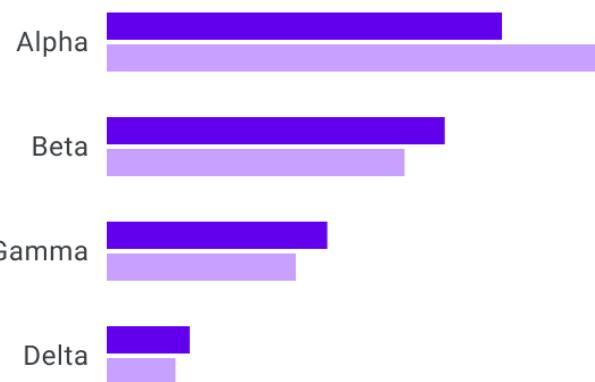
Google just came out with a guide



<https://material.io/design/communication/data-visualization.html#>

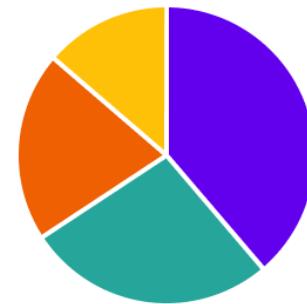
Most interesting are the Do – Don't

Sales by product

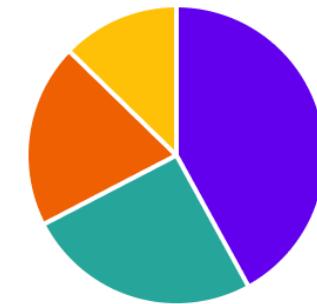


Sales by product

This quarter



Last quarter



● This quarter ● Last quarter

● Alpha ● Beta ● Gamma ● Delta



Do

Don't

<https://material.io/design/communication/data-visualization.html#>

Explanatory analysis – making it better

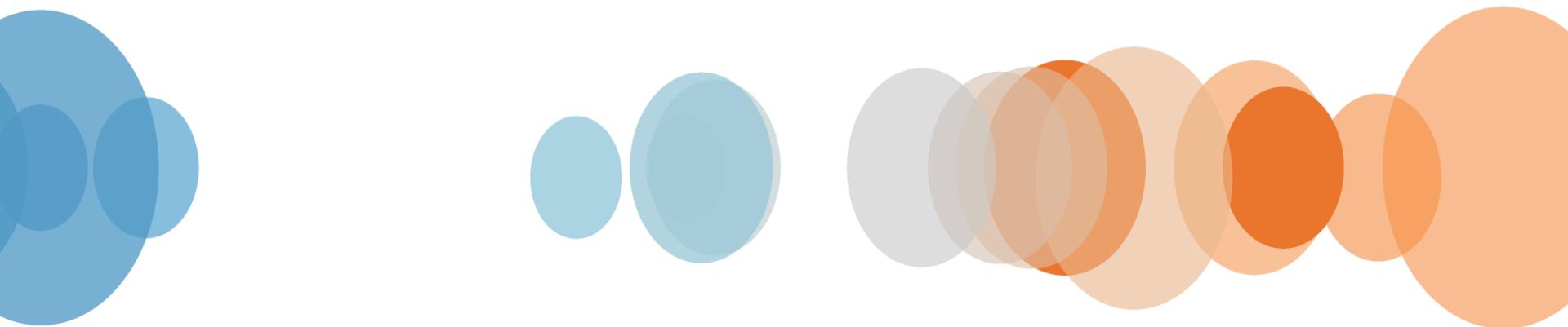
- Eliminate clutter
 - Reduce brain power to process your visualization
- Focus attention where you want it
 - Use color strategically

Explanatory analysis – Design and story tell

- Think like a designer
 - What do you want the audience to be doing with the data?
- Tell a story
 - Stories resonate with us
 - A story has a beginning, middle, and end
 - Narrative flow

[Nussbaumer 2015]

The use of color in visualizations



Earths tones

Gentle browns, blues. Calming, sinks into the page



Cool

Soothing, restful, calm



Unnatural colors

Alarming, unnerving, draws attention.



Warm

Optimistic, active, vivid



Increasing color intensity

Increasing saturation and brightness draws the eye and means the point is more important



Source: *Juice Analytics Whitepaper (part 3)*

THE USE OF COLOR IN DATA VISUALIZATION

SEQUENTIAL

color is ordered from low to high



DIVERGING

two sequential colors with a neutral midpoint



CATEGORICAL

contrasting colors for individual comparison



HIGHLIGHT

color used to highlight something



ALERT

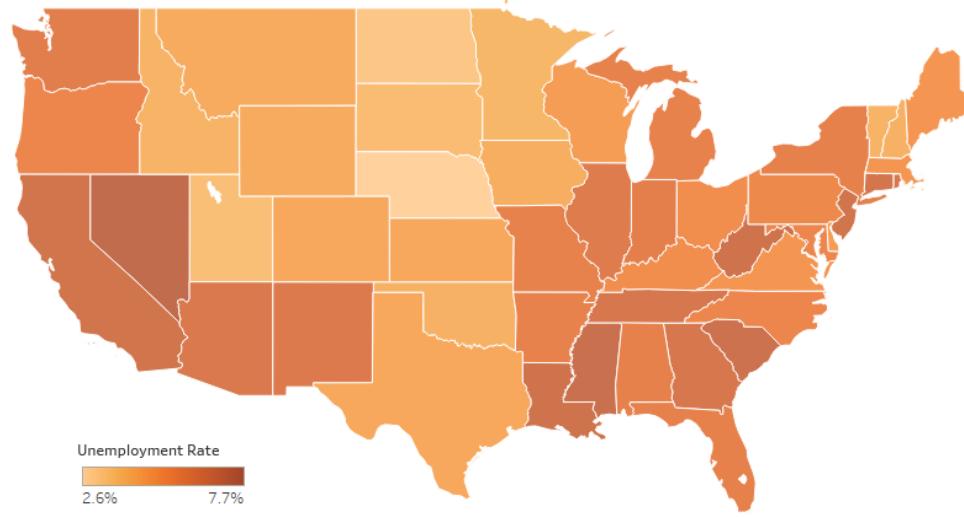
color used to get reader's attention



Source: *The Big Book of Dashboards* (Figure 1.16)

Sequential Color

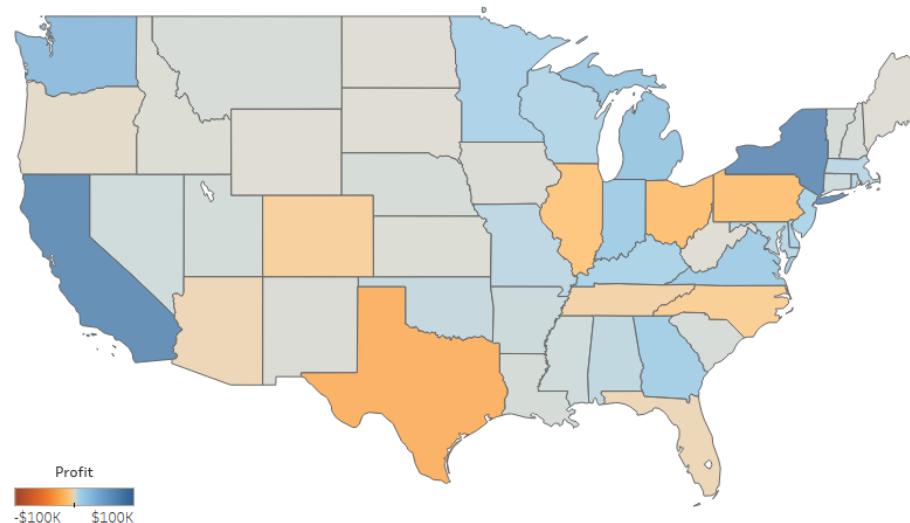
Unemployment Rate by State



Source: *The Big Book of Dashboards* (Figure 1.17)

Diverging Color

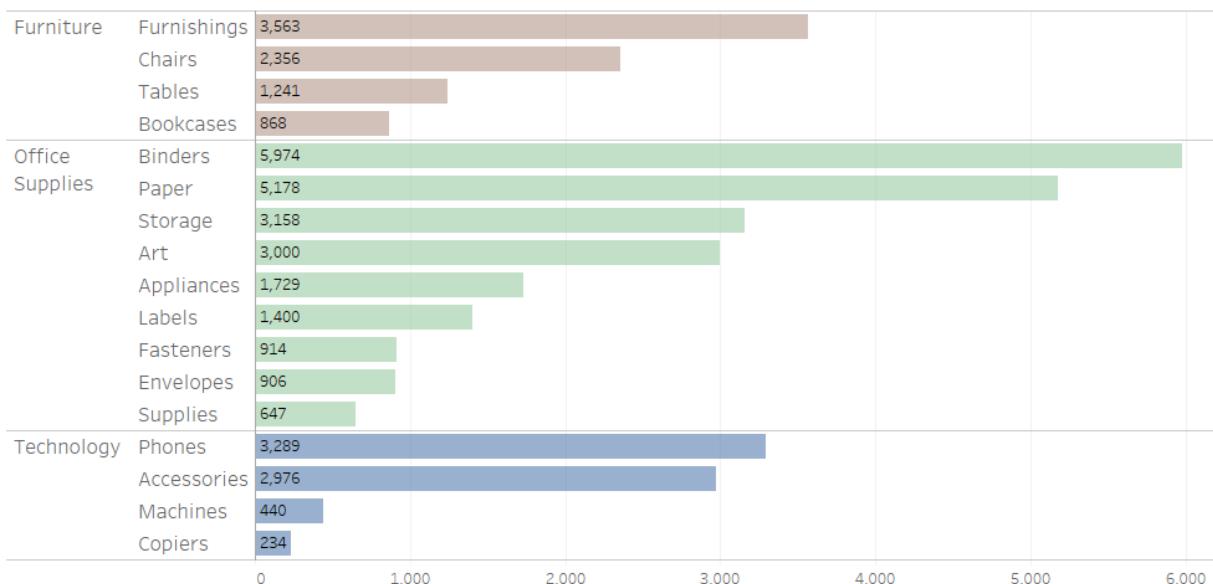
Profit by State



Source: *The Big Book of Dashboards* (Figure 1.19)

Categorical Color

Quantity by Category and Subcategory

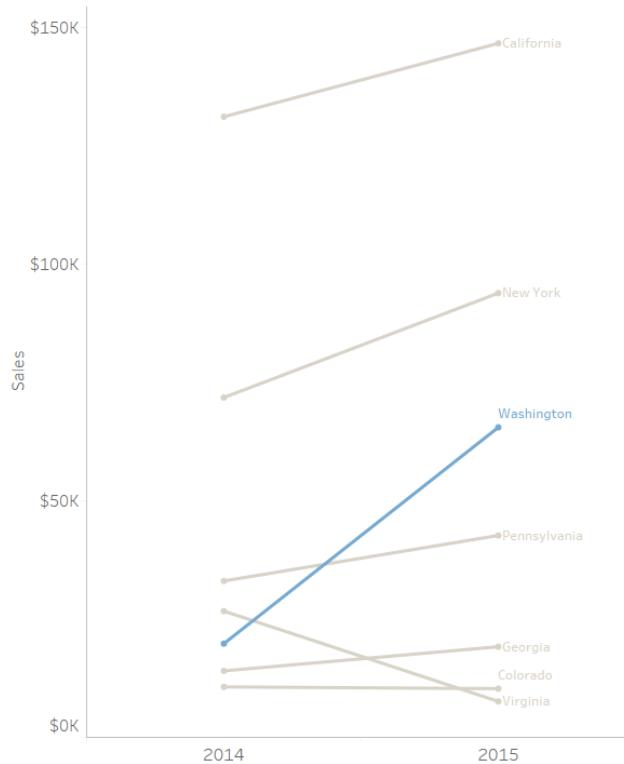


Source: *The Big Book of Dashboards* (Figure 1.20)

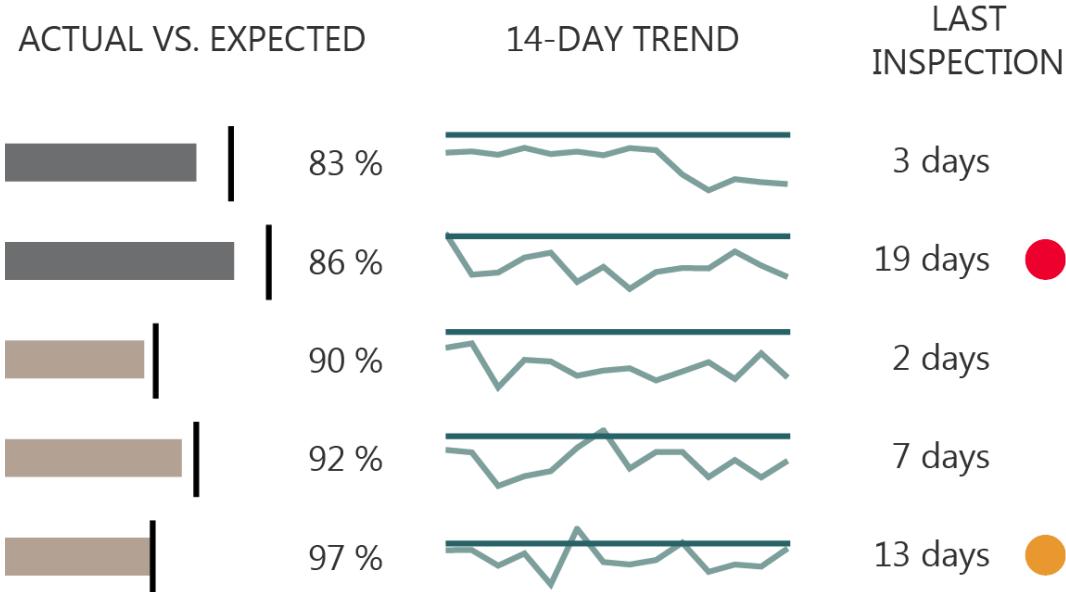
Highlight Color

Source: *The Big Book of Dashboards (Figure 1.21)*

Sales by State, 2014-2015



Alerting Color



Source: *The Big Book of Dashboards* (Figure 1.22)

- Accessories
- Appliances
- Art
- Binders
- Bookcases
- Chairs
- Copiers
- Envelopes
- Fasteners
- Furnishings
- Labels
- Machines
- Paper
- Phones
- Storage
- Supplies
- Tables

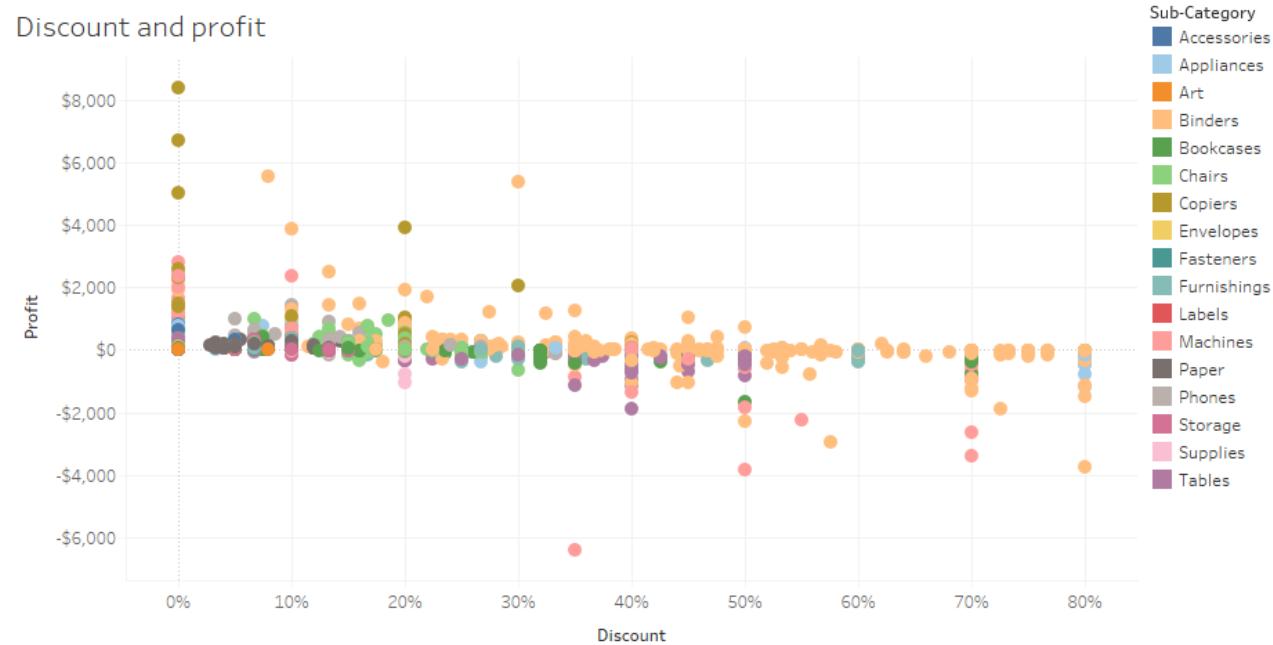
Too Much Color

- Short-term Memory = “small chunks of information”
- Requires reusing the same or similar color
- Requires frequent reference to the legend

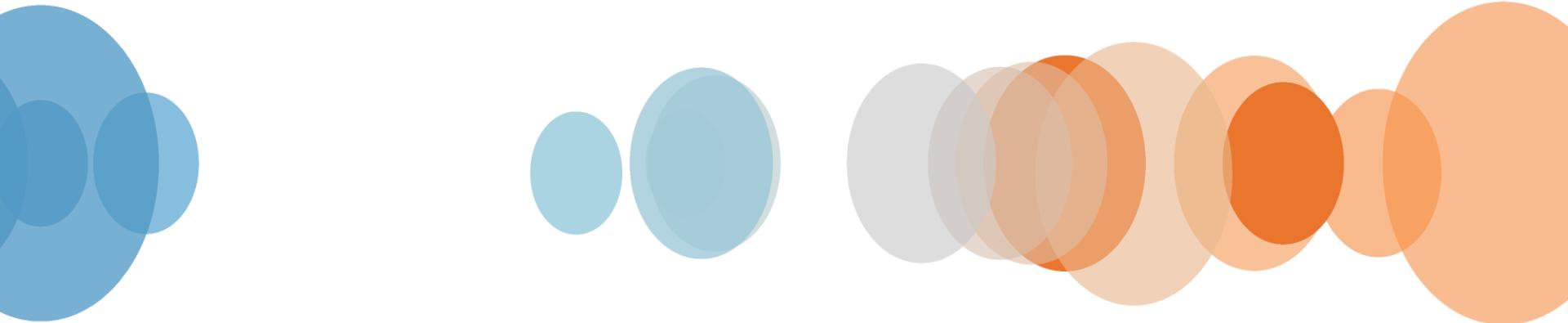


Too Much Colors

Discount and profit



Color Vision Deficiency (aka Colorblind)



Color
Normal



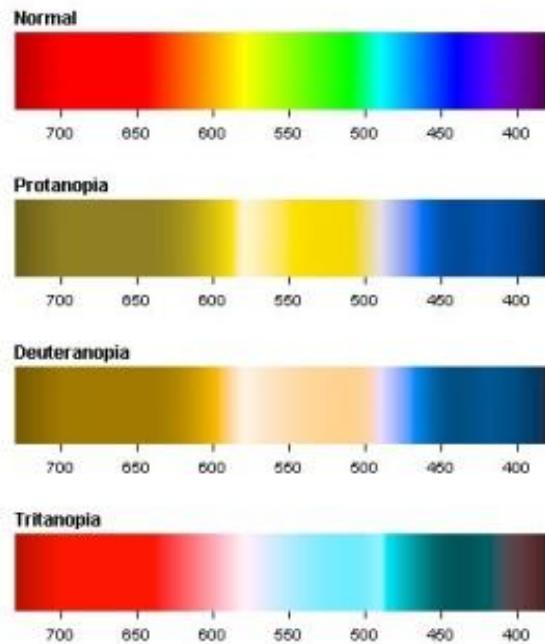
Color
Vision
Deficiency

The Eye with Normal Color Vision

Three types of color sensitive cones

1. Short (S) – respond to short wave lengths
2. Medium (M) - respond to medium wave lengths
 - more sensitive to green colors
3. Long (L) - respond to long wave lengths
 - more sensitive to red colors

How Color is Perceived by Someone with CVD

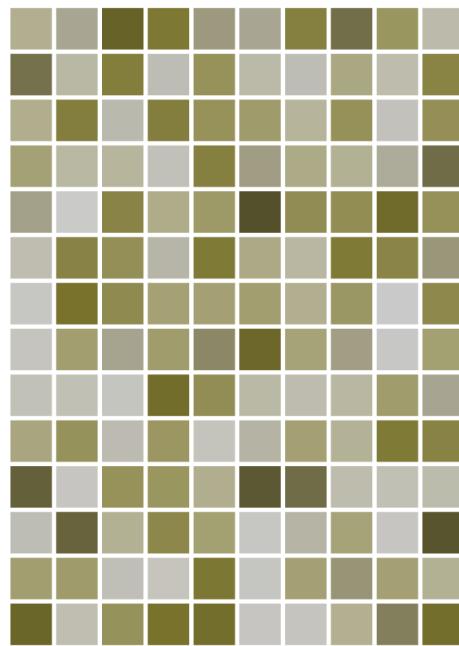


Source: www.colblindor.com

Traffic Light
Colors



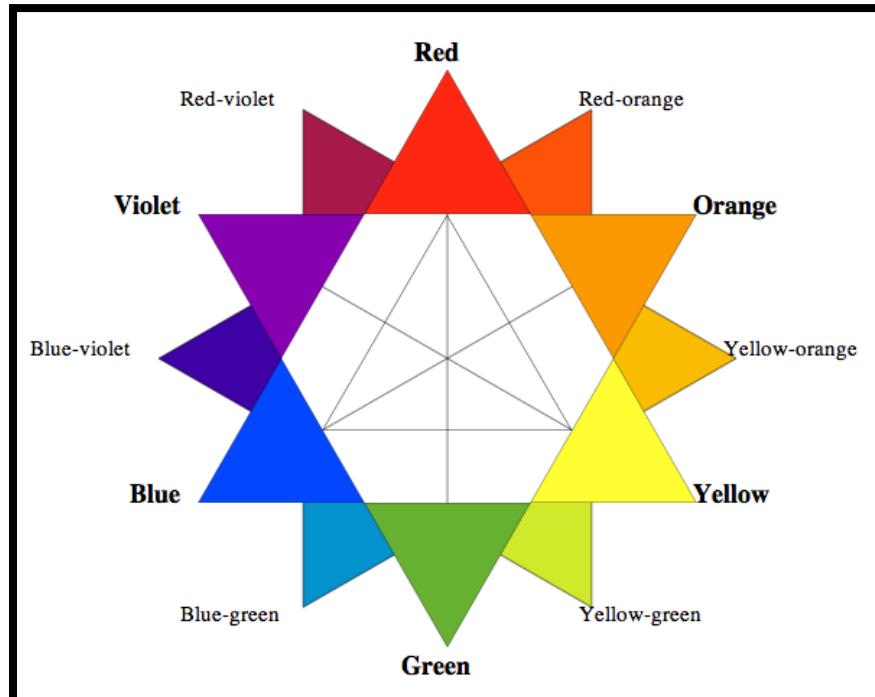
Deutanopia
Simulation



Which are green and
which are red?

Source: *The Big Book of Dashboards* (Figure 33.3)

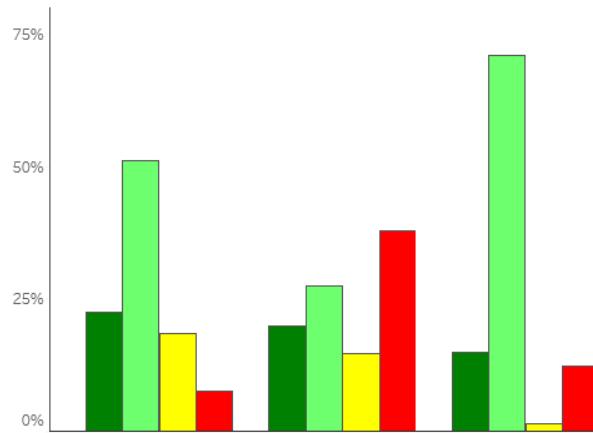
Color Wheel



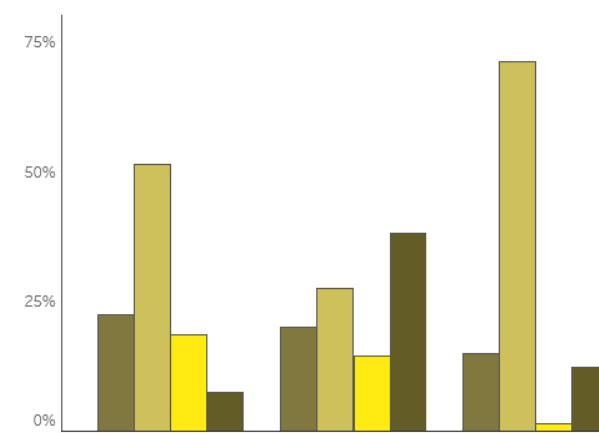
Blue is the most friendly color for people with CVD. Orange or Red could be used for “bad”

<https://www.color-blindness.com/coblis-color-blindness-simulator/>

Traffic Light Colors



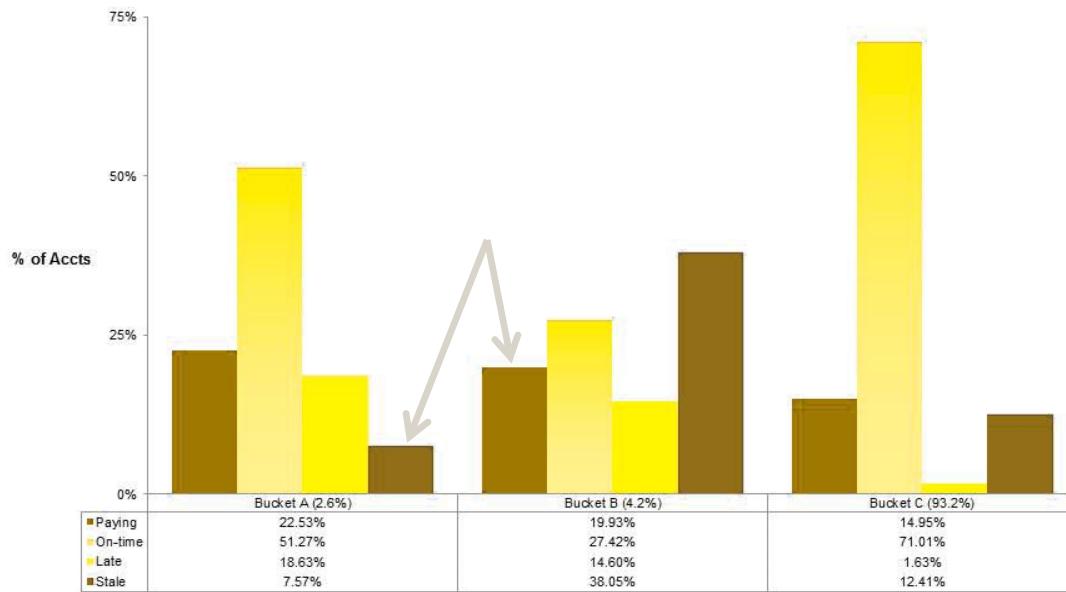
Protanopia Simulation



Source: *The Big Book of Dashboards* (Figure 1.24)

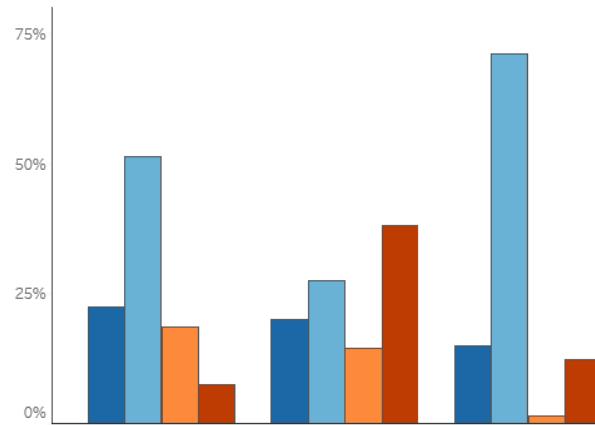
Using the traffic light colors and a simulator

Protanope Simulation

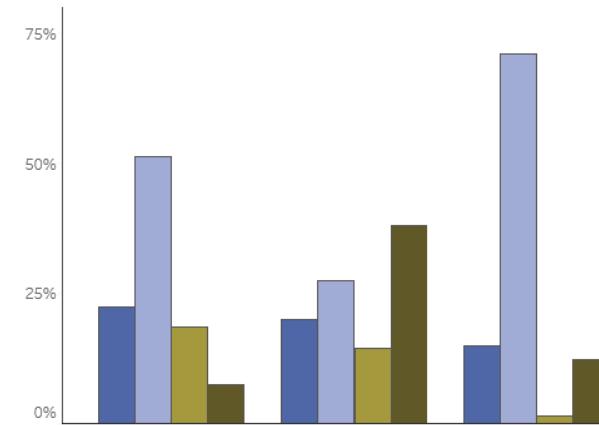


<https://www.color-blindness.com/coblis-color-blindness-simulator/>

Colorblind-Friendly Blue and Orange



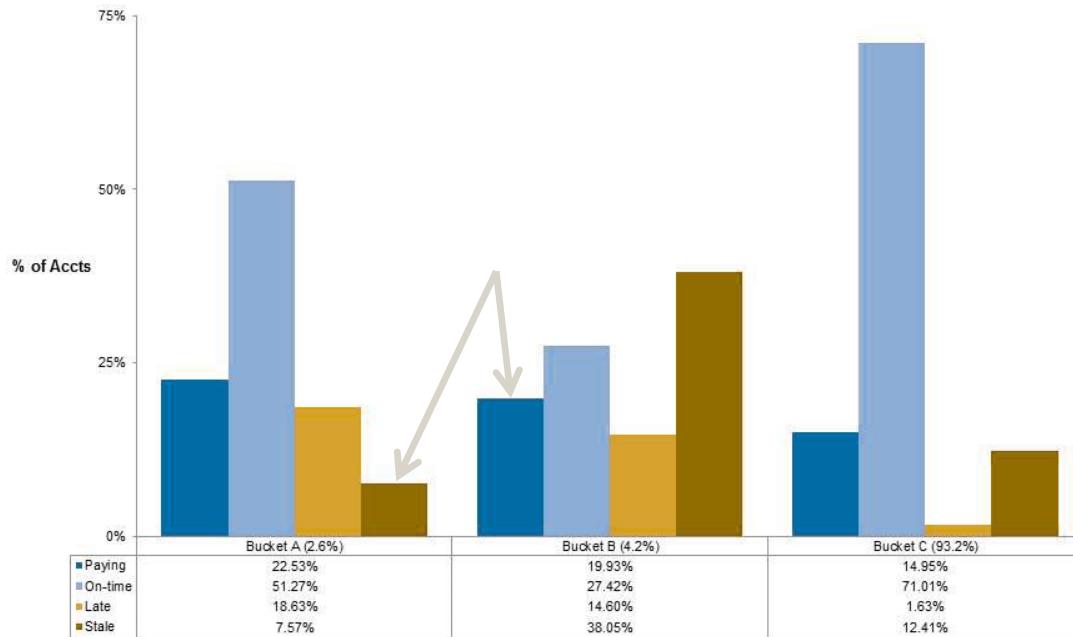
Protanopia Simulation



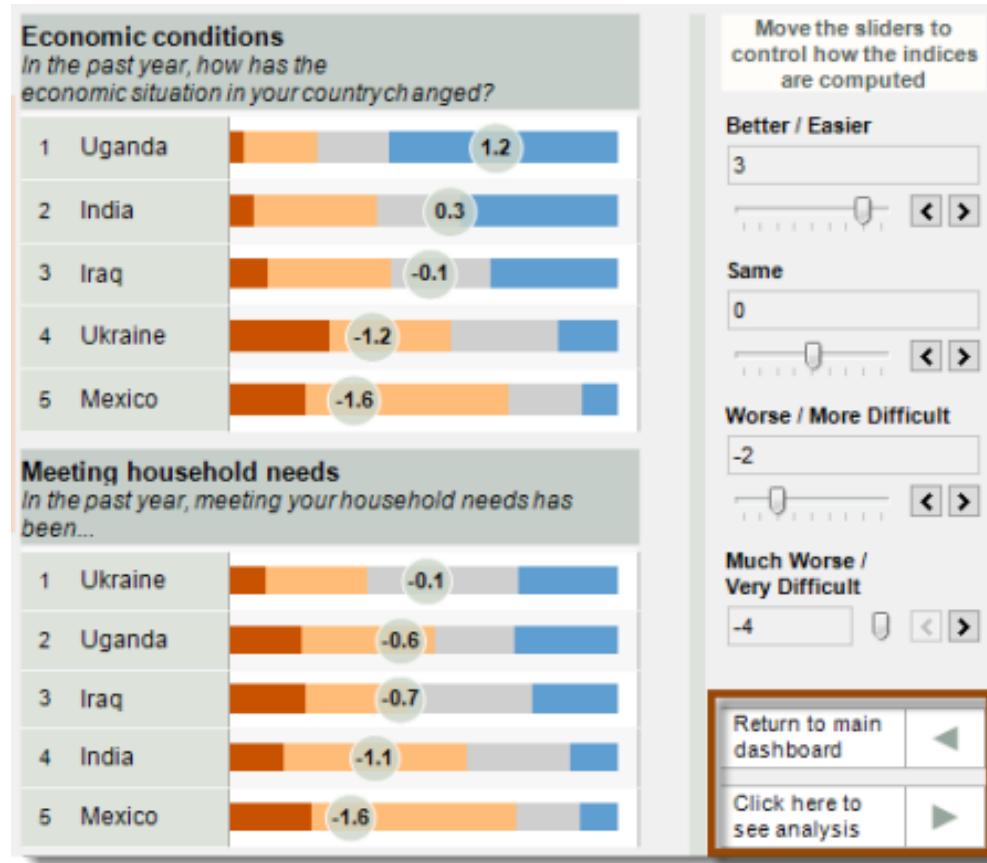
Source: *The Big Book of Dashboards* (Figure 1.25)

Using the colorblind friendly palette built into Tableau

Protanope Simulation

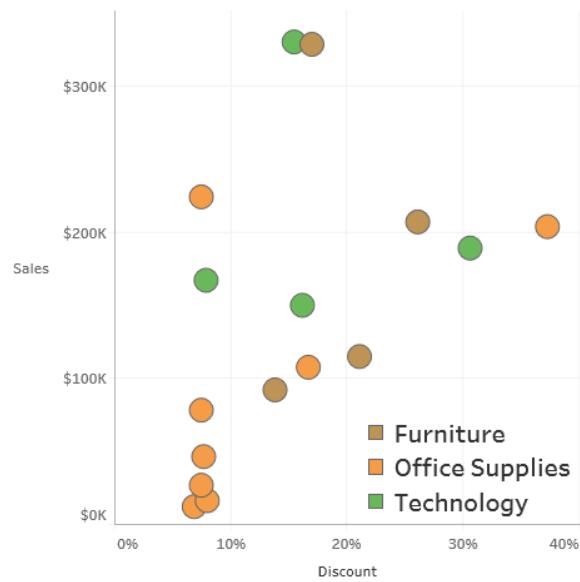


Blue is easy to see even if Orange turns to a shade of brown.

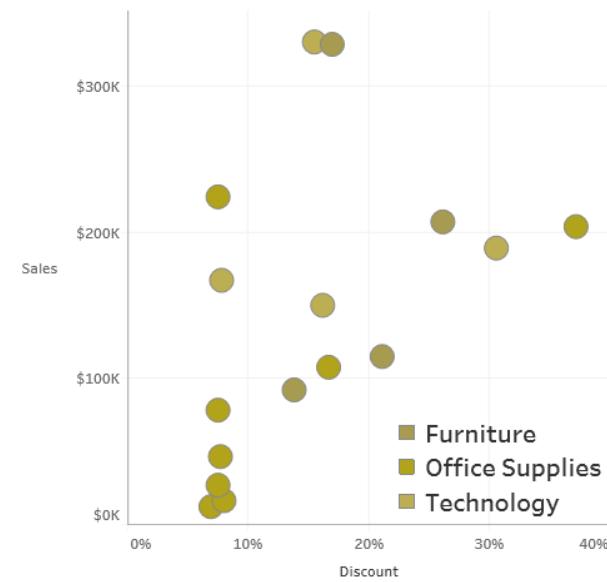


Good color contrast even with someone with CVD

Normal Color



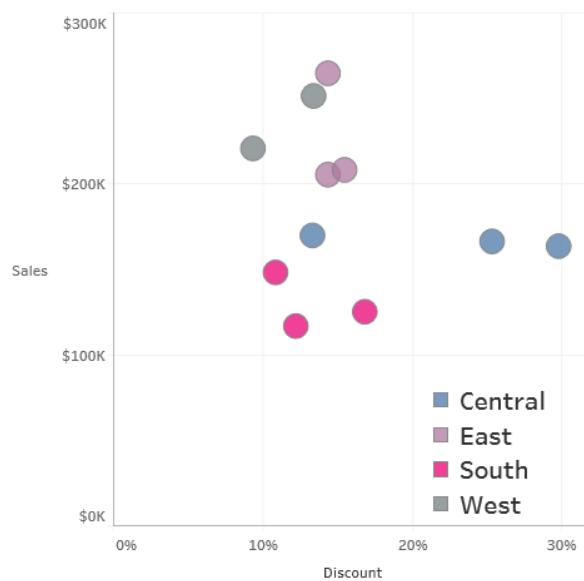
Protanopia CVD Simulation



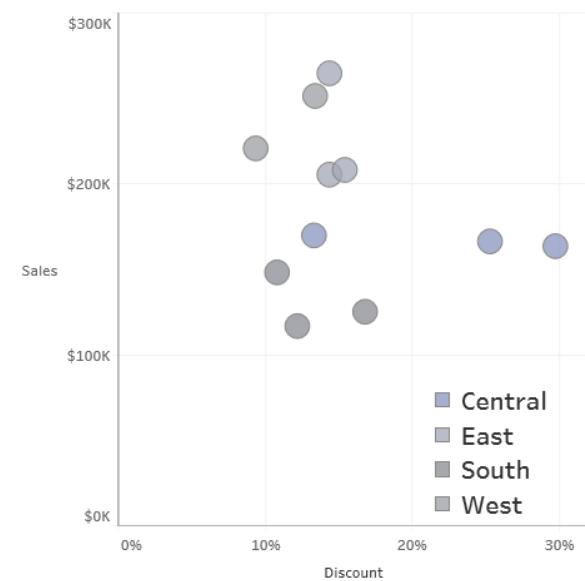
Source: *The Big Book of Dashboards* (Figure 1.26)

Not just red and green! Do not use red, green, orange and brown together.

Normal Color



Deuteranopia CVD Simulation

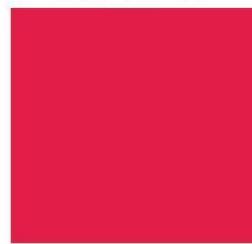


Source: *The Big Book of Dashboards* (Figure 1.27)

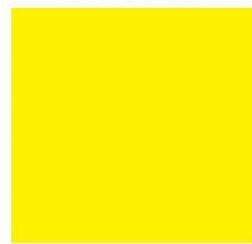
Also, purple and fuchsia with grey are not good together

Traffic Light Color Palette

#E22049



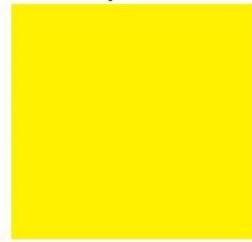
#FFF200



#0D9E49



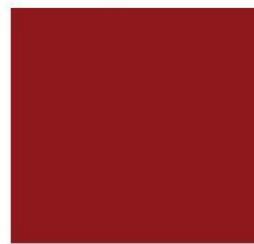
Deuteranopia Simulation



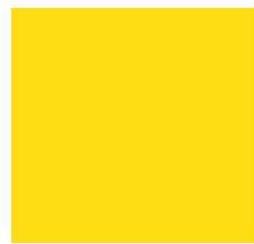
Source: *The Big Book of Dashboards* (Figure 33.8)

Alternate Traffic Light Color Palette

#8E191C



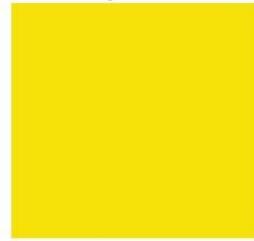
#FFDE17



#83C775

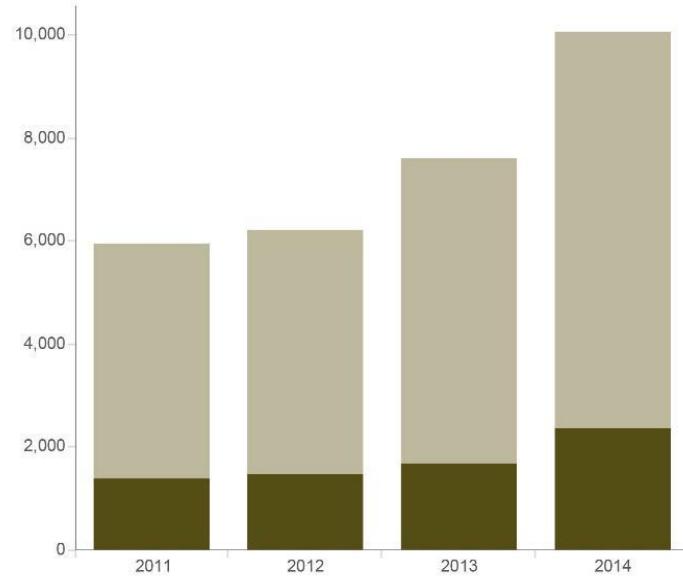
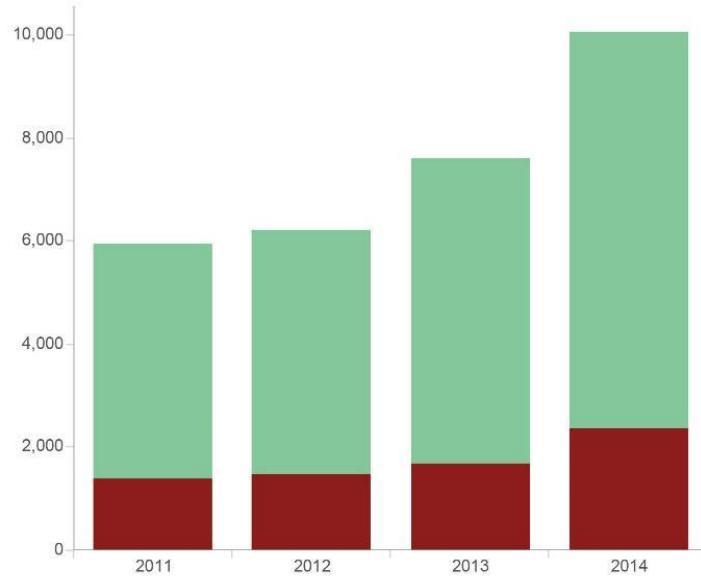


Deutanopia Simulation



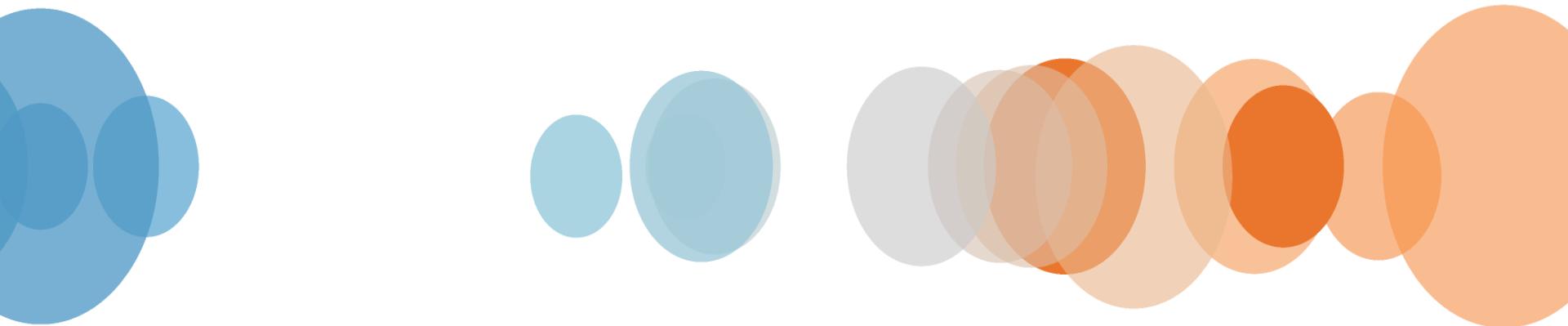
Source: *The Big Book of Dashboards* (Figure 33.9)

Example in Practice



Source: *The Big Book of Dashboards* (Figure 33.11)

To summarize with the 4Cs and a Tool



The Shaffer 4 C's of Data Visualization

Clear - easily seen; sharply defined

- who's the audience? what's the message?
- clarity more important than aesthetics

Clean - thorough; complete; unadulterated

- labels, axis, gridlines, formatting, right chart type, color choice, etc.

Concise - brief but comprehensive

- not minimalist but not verbose

Captivating - to attract and hold by beauty or excellence

- does it capture attention? is it interesting? does it tell the story?

Data Viz Checklist

Data Visualization Checklist

by Stephanie Evergreen & Ann K. Emery
February 2018

This checklist is meant to be used as a guide for the development of high impact data visualizations. Rate each aspect of the data visualization by circling the most appropriate number, where 2 points means the guideline was fully met, 1 means it was partially met, and 0 means it was not met at all. n/a should not be used frequently, but reserved for when the guideline truly does not apply. For example, a pie chart has no axes lines or tick marks to rate. If the guidelines has been broken intentionally to make a point, rate it n/a and deduct those points from the total possible. Refer to the Data Visualization Anatomy Chart on the last page for guidance on vocabulary and the Resources at the end for more details.

	Guideline	Rating
Text	6-12 word descriptive title is left-justified in upper left corner Short titles enable readers to comprehend takeaway messages even while quickly skimming the graph. Rather than a generic phrase, use a descriptive sentence that encapsulates the graph's finding or "so what?" Western cultures start reading in the upper left, so locate the title there.	2 1 0 n/a
	Subtitle and/or annotations provide additional information Subtitles and annotations (call-out text within the graph) can add explanatory and interpretive power to a graph. Use them to answer questions a viewer might have or to highlight specific data points.	2 1 0 n/a
	Text size is hierarchical and readable Titles are in a larger size than subtitles or annotations, which are larger than labels, which are larger than axis labels, which are larger than source information. The smallest text - axis labels - are at least 9 point font size on paper, at least 20 on screen.	2 1 0 n/a
	Text is horizontal Titles, subtitles, annotations, and data labels are horizontal (not vertical or diagonal). Line labels and axis labels can deviate from this rule and still receive full points. Consider switching graph orientation (e.g.,	2 1 0 n/a

Data Viz Checklist - Site

 Data Visualization Checklist

Rate your visualization

Data Visualization Checklist

The Data Visualization Checklist is a compilation of 24 guidelines on how graphs should be formatted to best show the story in your data. The 24 guidelines are broken down into 5 sections: Text, Arrangement, Color, Lines, and Overall.

In the Rate your visualization page, you can upload your visualization (jpg, gif, png) and the page will walk you through each checkpoint so that you can score your visualization and identify its strengths and places where you could make improvements.

Is This Site Confidential?

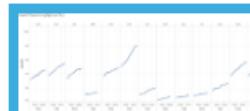
After you rate your graph you'll see a page of your summary scores. You will have options there to:

- Make your visual public so that anyone can see it
- Save your scores so you get a URL you can return to
- Or you can simply not click Submit this report and your chart and data will disappear when you close your browser

To get oriented to the checklist and its guidance, click through the training below.

Rate your visualization

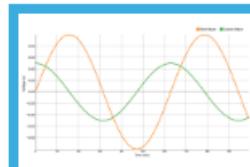
Recent submissions



Score: 95.8%



Score: 81.3%

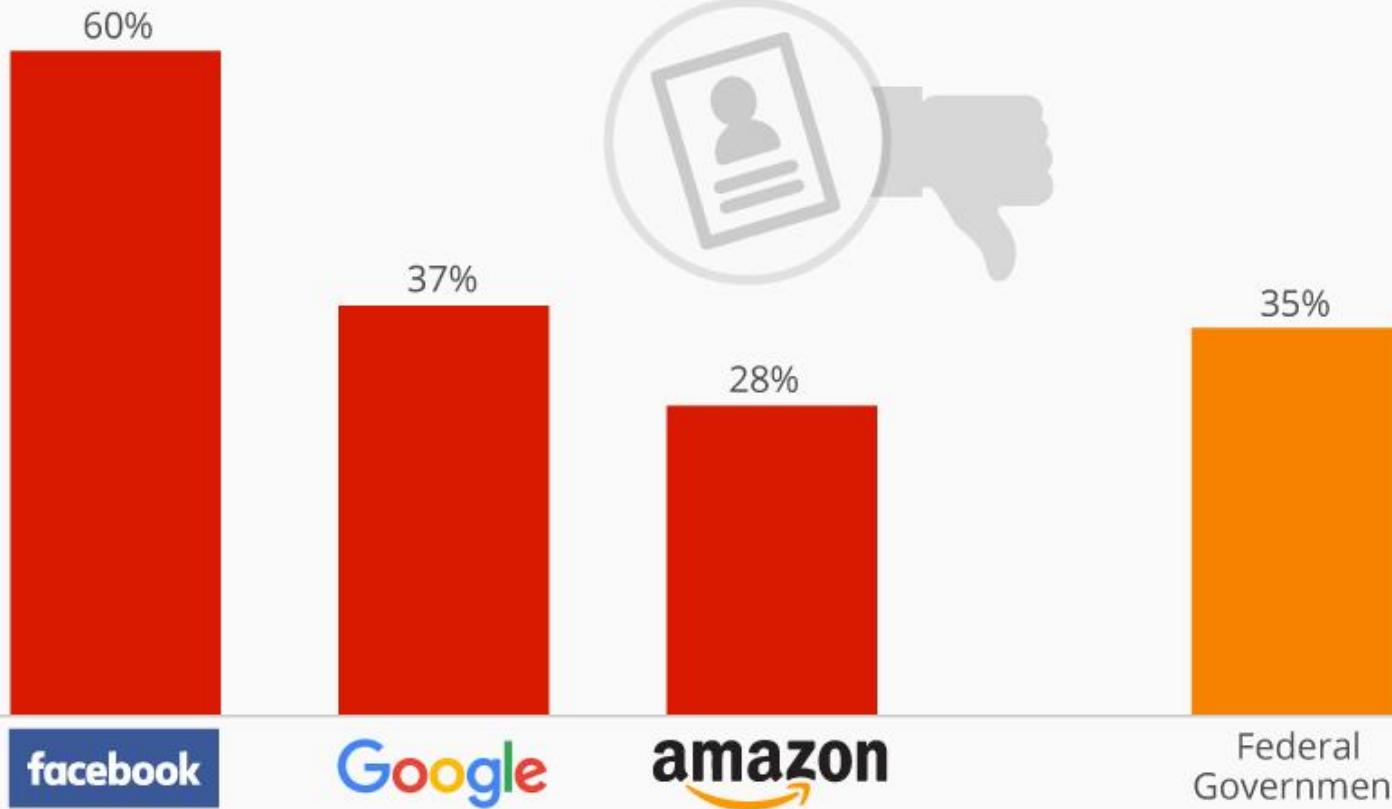


Score: 58.3%



Americans Do Not Trust Facebook with Personal Info

Share of respondents who do not trust selected organizations with their personal information

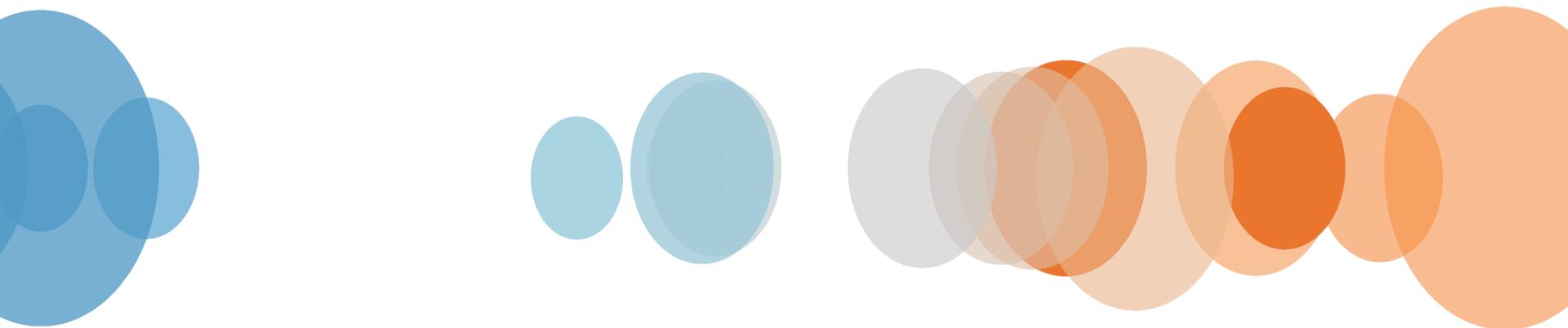


Survey conducted between March 23-27, 2019
with 1,000 U.S. adults over the phone

Source: NBC Wall Street Journal March Poll

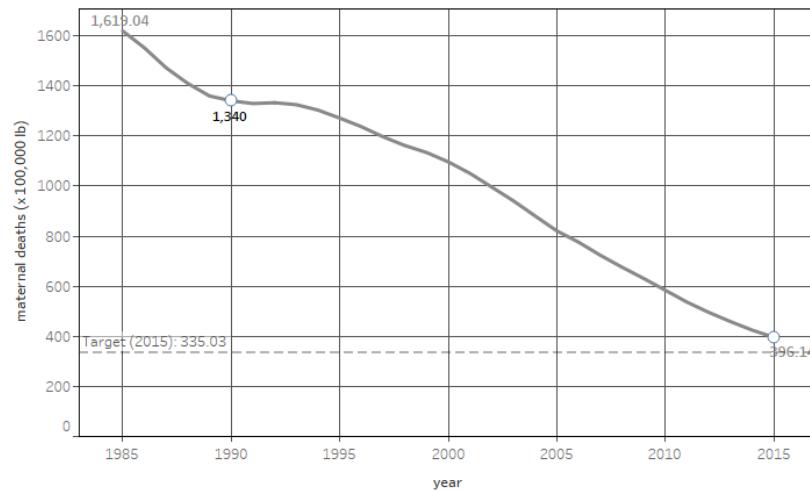
statista

Examples

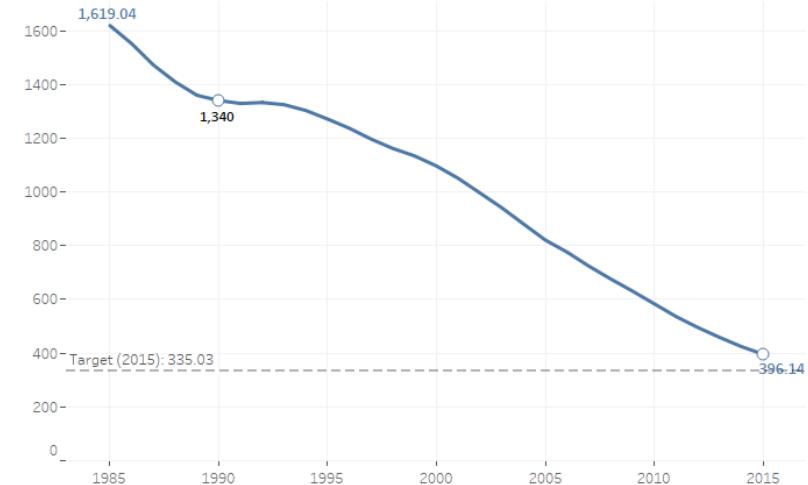


Which one is easier to understand? Why?

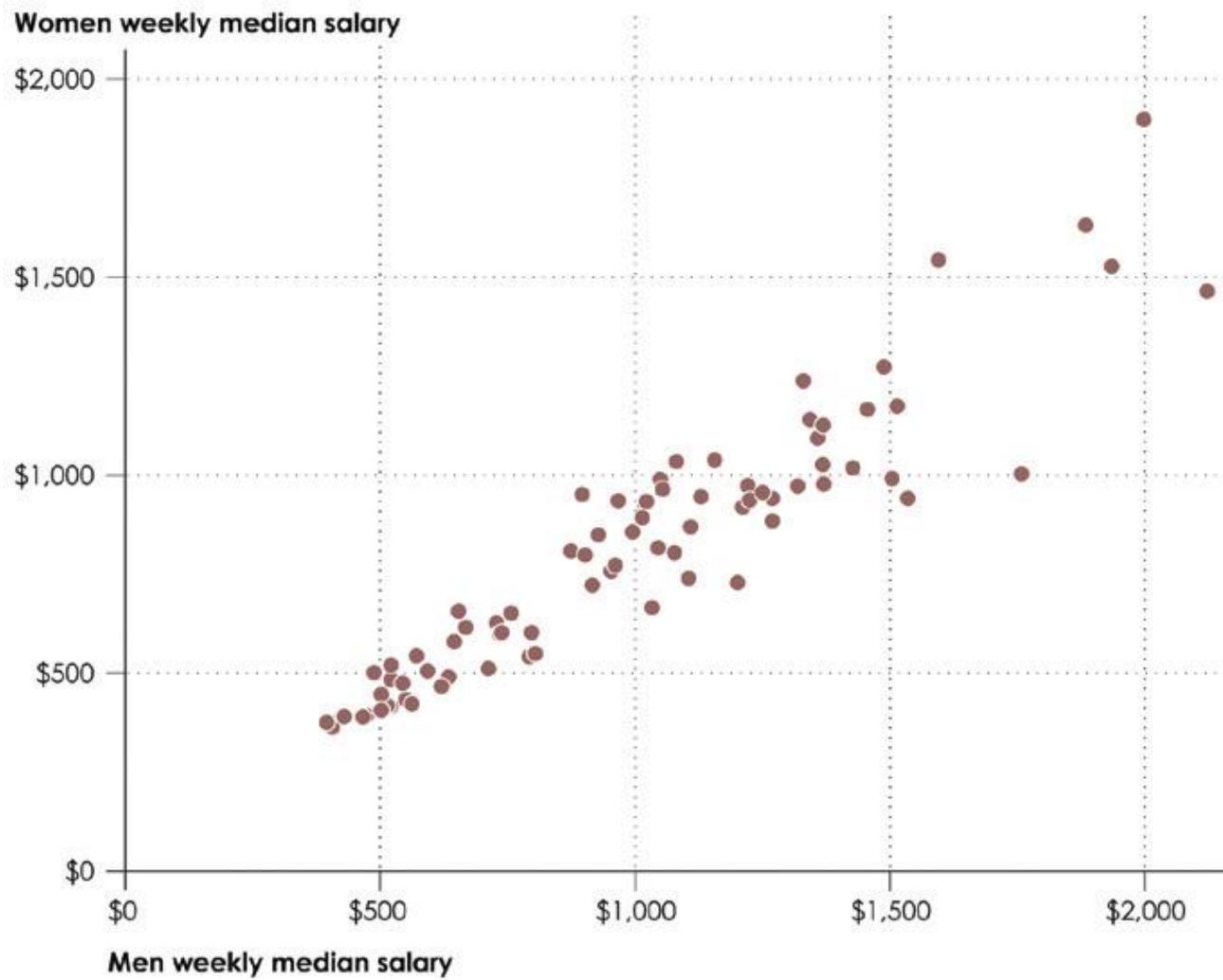
Maternal mortality ratio time series, baseline 1990 and target 2015



Maternal mortality ratio time series, baseline 1990 and target 2015
maternal deaths per 100,000 live births



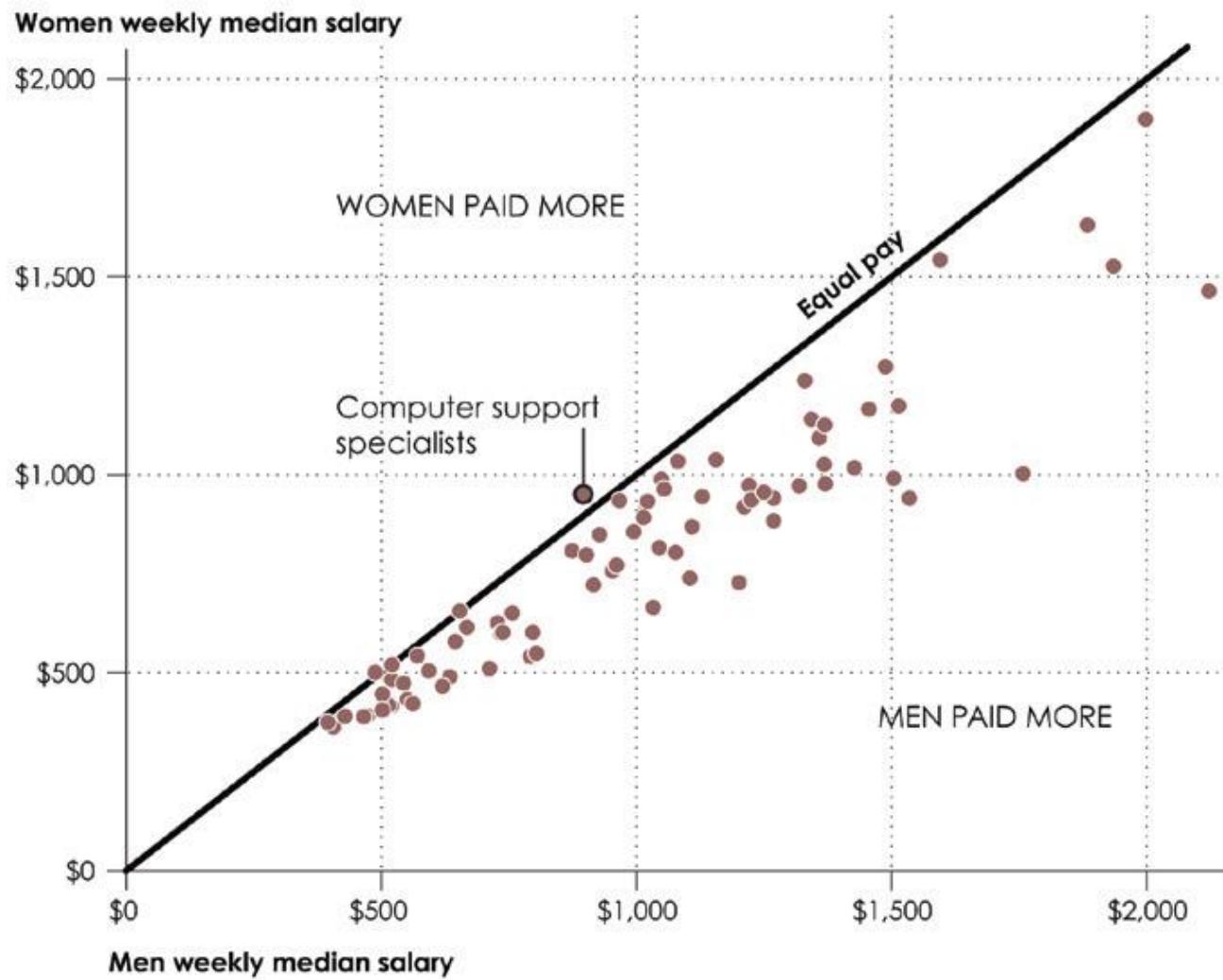
Gender pay gap in 2011



Source: Bureau of Labor Statistics

[Yau 2013]

Gender pay gap in 2011



Source: Bureau of Labor Statistics

[Yau 2013]

	Scenario #1	Scenario #2	Scenario #3
Sales	\$782,193	\$782,193	\$782,193
Royalty	173,965	173,965	173,965
Lease	1,089,496	1,089,496	1,089,496
Total Revenue	2,045,654	2,045,654	2,045,654
Management Fees	(590,550)	(590,550)	(514,285)
Direct Expenses:	(73,695)	(73,695)	(73,695)
SGA expenses	(3,750)	(3,750)	(3,750)
Interest Expenses	(105,650)	(105,650)	(105,650)
Total Expenses	(773,645)	(773,645)	(697,380)
Cash Available	1,272,009	1,272,009	1,348,274
Principal Payment	(2,273,950)	(1,125,718)	(1,125,717)
Net Cash	(\$1,001,941)	\$146,291	\$222,557

	Scenario #1	Scenario #2	Scenario #3
Sales	\$782,193	\$782,193	\$782,193
Royalty	173,965	173,965	173,965
Lease	1,089,496	1,089,496	1,089,496
Total Revenue	2,045,654	2,045,654	2,045,654
Management Fees	(590,550)	(590,550)	(514,285)
Direct Expenses:	(73,695)	(73,695)	(73,695)
SGA expenses	(3,750)	(3,750)	(3,750)
Interest Expenses	(105,650)	(105,650)	(105,650)
Total Expenses	(773,645)	(773,645)	(697,380)
Cash Available	1,272,009	1,272,009	1,348,274
Principal Payment	(2,273,950)	(1,125,718)	(1,125,717)
Net Cash	(\$1,001,941)	\$146,291	\$222,557

Flag of Mali

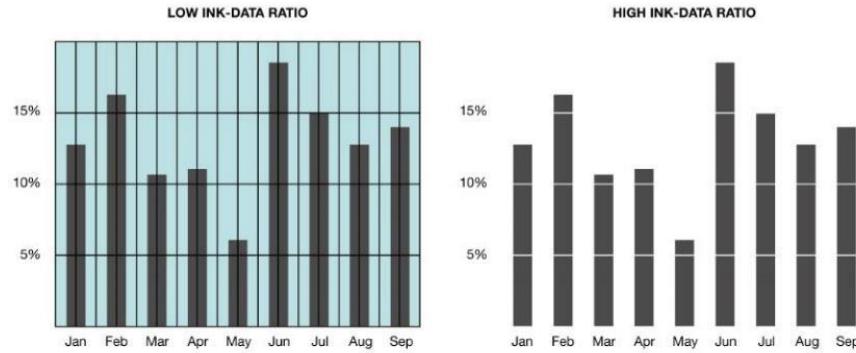


	Scenario #1	Scenario #2	Scenario #3
Sales	\$782,193	\$782,193	\$782,193
Royalty	173,965	173,965	173,965
Lease	1,089,496	1,089,496	1,089,496
Total Revenue	2,045,654	2,045,654	2,045,654
Management Fees	(590,550)	(590,550)	(514,285)
Direct Expenses:	(73,695)	(73,695)	(73,695)
SGA expenses	(3,750)	(3,750)	(3,750)
Interest Expenses	(105,650)	(105,650)	(105,650)
Total Expenses	(773,645)	(773,645)	(697,380)
Cash Available	1,272,009	1,272,009	1,348,274
Principal Payment	(2,273,950)	(1,125,718)	(1,125,717)
Net Cash	(\$1,001,941)	\$146,291	\$222,557

	Scenario #1	Scenario #2	Scenario #3
Sales	\$782,193	\$782,193	\$782,193
Royalty	173,965	173,965	173,965
Lease	1,089,496	1,089,496	1,089,496
Total Revenue	2,045,654	2,045,654	2,045,654
Management Fees	(590,550)	(590,550)	(514,285)
Direct Expenses:	(73,695)	(73,695)	(73,695)
SGA expenses	(3,750)	(3,750)	(3,750)
Interest Expenses	(105,650)	(105,650)	(105,650)
Total Expenses	(773,645)	(773,645)	(697,380)
Cash Available	1,272,009	1,272,009	1,348,274
Principal Payment	(2,273,950)	(1,125,718)	(1,125,717)
Net Cash	(\$1,001,941)	\$146,291	\$222,557

Tufte's Five Laws of Data-Ink:

1. Above all else show the data.
2. Maximize the data-ink ratio
3. Erase non-data-ink.
4. Erase redundant data-ink.
5. Revise and edit.



Remove to improve the **data tables** edition

Created by Darkhorse Analytics

www.darkhorseanalytics.com

<https://www.darkhorseanalytics.com/portfolio/2016/1/7/data-looks-better-naked-clear-off-the-table>

Maximizing the Data Ink Ratio

Show the Data - Maximize the Data Ink Ratio

Watch later Share

PlayfairData presents

Show the Data

Maximize the Data Ink Ratio

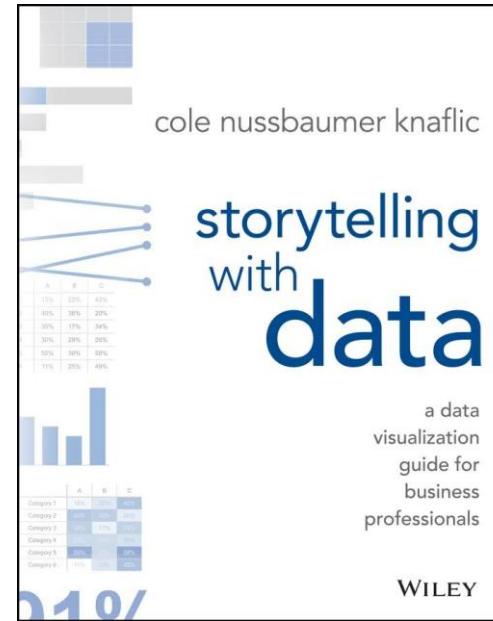
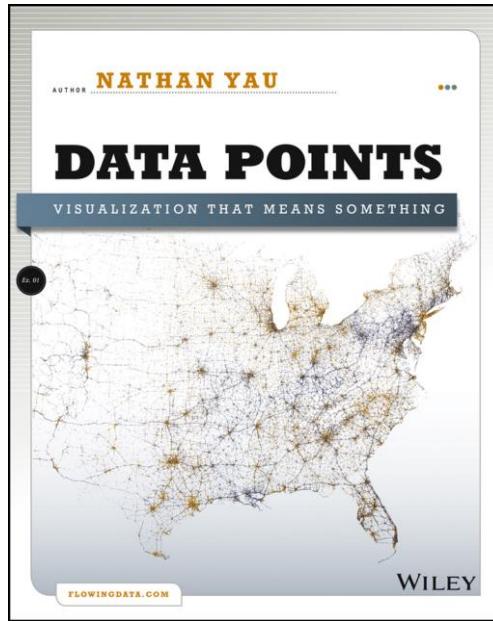
MORE VIDEOS

An animation of the "Data-Ink Ratio" found in
The Visual Display of Quantitative Information, by Edward R. Tufte

0:02 / 1:00

YouTube

Two great references



[Storytelling with Data](#)
[Data Points](#)