Visualization Basics: Introduction to Tableau

Miriam McGaugh, PhD

Let’s look at some data

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

Let’s look at some data…with statistical 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
9	--	9	--	9	--	9	--
11	--	11	--	11	--	11	--
--	7.5	--	7.5	--	7.5	--	7.5
--	4.122		4.122		4.122		4.122
0.816	0.816	0.816	0.816	0.816	0.816	0.816	0.816
y = 3 + 0.5x	y = 3 + 0.5x	y = 3 + 0.5x	y = 3 + 0.5x	y = 3 + 0.5x	y = 3 + 0.5x	y = 3 + 0.5x	y = 3 + 0.5x

Mean of x

Variance of x

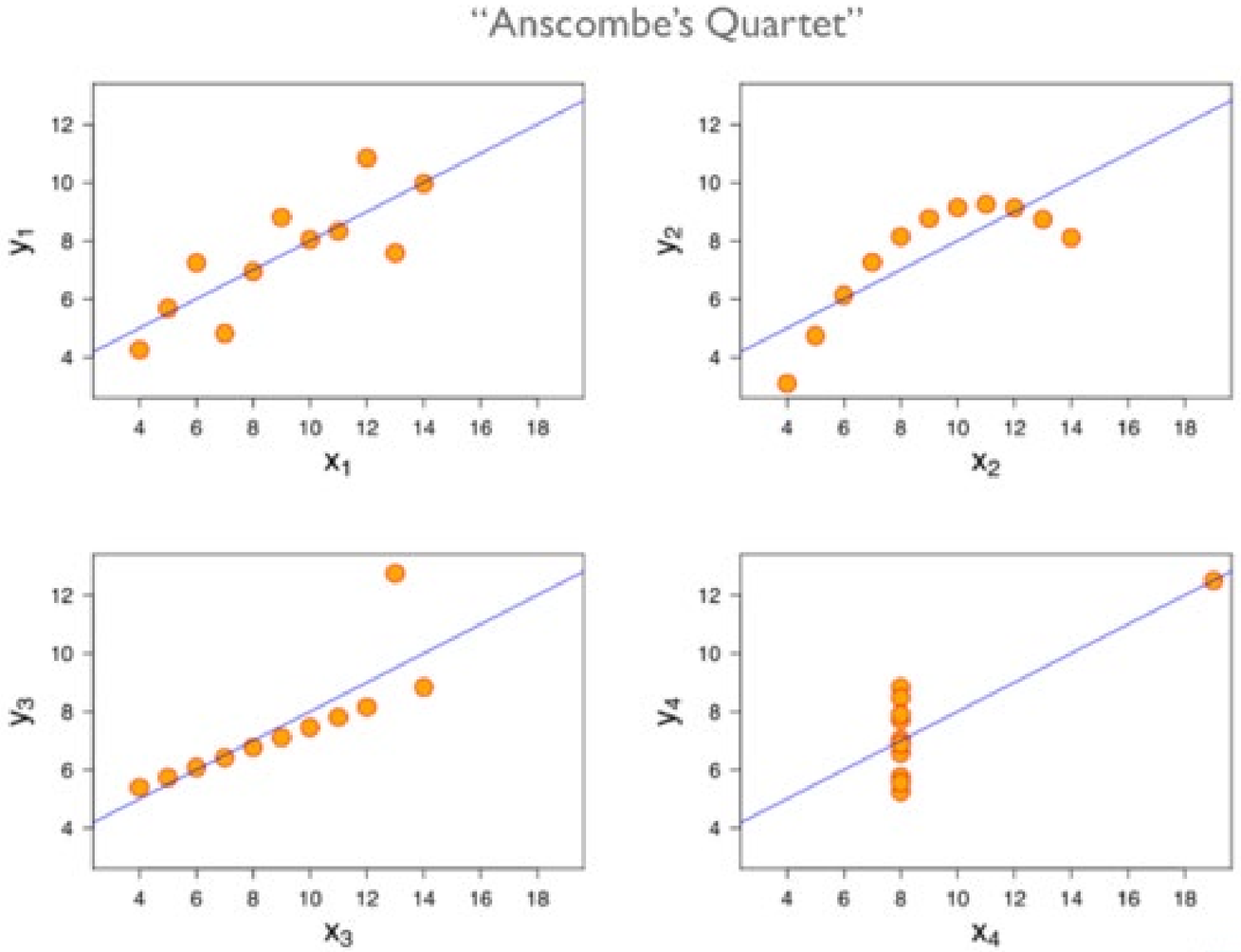
Mean of y

Variance of y

Correlation between x and y

Linear regression line

Let’s look at some data…visually



Source: Wikipedia

Why visualize?

Data visualization is not a personal preference, to be done “just in case some people are more visual.” It is a necessity!

- The previous slide is a classic example of how some insights can *only* be found through visualization.

The first step in understanding your data should always be to examine it visually. Visualization can play a critical role in helping you figure out what the interesting questions are.

Remember the picture superiority effect: pictures are retained at much higher rates than words.

Why visualize?

A good visualization reduces the time to insight. Here's a simple example:

You receive a report of monthly revenues by product and region. Which were your five highest-revenue product/regions last month?

	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Northeast	$ 15,749.00	$ 40,195.00	$ 15,472.00	$ 63,029.00	$ 8,509.00	$ 42,987.00	$ 27,778.00	$ 12,510.00
Southeast	$ 48,044.00	$ 20,741.00	$ 40,643.00	$ 15,687.00	$ 12,342.00	$ 23,297.00	$ 10,401.00	$ 10,522.00
Central	$ 5,240.00	$ 45,296.00	$ 16,114.00	$ 63,359.00	$ 58,198.00	$ 24,191.00	$ 46,826.00	$ 50,278.00
Northwest	$ 12,860.00	$ 11,548.00	$ 43,134.00	$ 19,331.00	$ 60,563.00	$ 51,475.00	$ 28,954.00	$ 14,405.00
Southwest	$ 37,087.00	$ 61,506.00	$ 54,084.00	$ 14,715.00	$ 17,811.00	$ 32,814.00	$ 47,853.00	$ 44,639.00

Why visualize?

Let's add color:

	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Northeast	$ 15,749.00	$ 40,195.00	$ 15,472.00	$ 63,029.00	$ 8,509.00	$ 42,987.00	$ 27,778.00	$ 12,510.00
Southeast	$ 48,044.00	$ 20,741.00	$ 40,643.00	$ 15,687.00	$ 12,342.00	$ 23,297.00	$ 10,401.00	$ 10,522.00
Central	$ 5,240.00	$ 45,296.00	$ 16,114.00	$ 63,359.00	$ 58,198.00	$ 24,191.00	$ 46,826.00	$ 50,278.00
Northwest	$ 12,860.00	$ 11,548.00	$ 43,134.00	$ 19,331.00	$ 60,563.00	$ 51,475.00	$ 28,954.00	$ 14,405.00
Southwest	$ 37,087.00	$ 61,506.00	$ 54,084.00	$ 14,715.00	$ 17,811.00	$ 32,814.00	$ 47,853.00	$ 44,639.00

Color automatically focused your brain – no mental calculations/comparisons needed!

For some applications, eliminating all distractions with a very simple viz can be effective:

	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Northeast								
Southeast								
Central								
Northwest								
Southwest								

How many nines are there?

4	7	7	5	5	2	7	4	7	1
4	9	2	5	7	7	2	6	1	7
1	7	6	9	3	4	7	5	1	2
5	1	6	3	3	8	4	8	6	6
6	5	6	4	9	3	8	9	1	9
3	8	1	5	2	2	3	6	3	9
4	6	4	5	6	3	7	7	9	1
9	1	3	3	6	1	3	3	1	8
8	1	1	8	7	5	8	1	7	4
3	6	9	2	8	9	3	7	5	7
4	4	4	2	8	2	2	9	2	8

How many nines are there?

4	7	7	5	5	2	7	4	7	1
4	9	2	5	7	7	2	6	1	7
1	7	6	9	3	4	7	5	1	2
5	1	6	3	3	8	4	8	6	6
6	5	6	4	9	3	8	9	1	9
3	8	1	5	2	2	3	6	3	9
4	6	4	5	6	3	7	7	9	1
9	1	3	3	6	1	3	3	1	8
8	1	1	8	7	5	8	1	7	4
3	6	9	2	8	9	3	7	5	7
4	4	4	2	8	2	2	9	2	8

Using Color

FACT: approximately 8% of men worldwide are color blind.

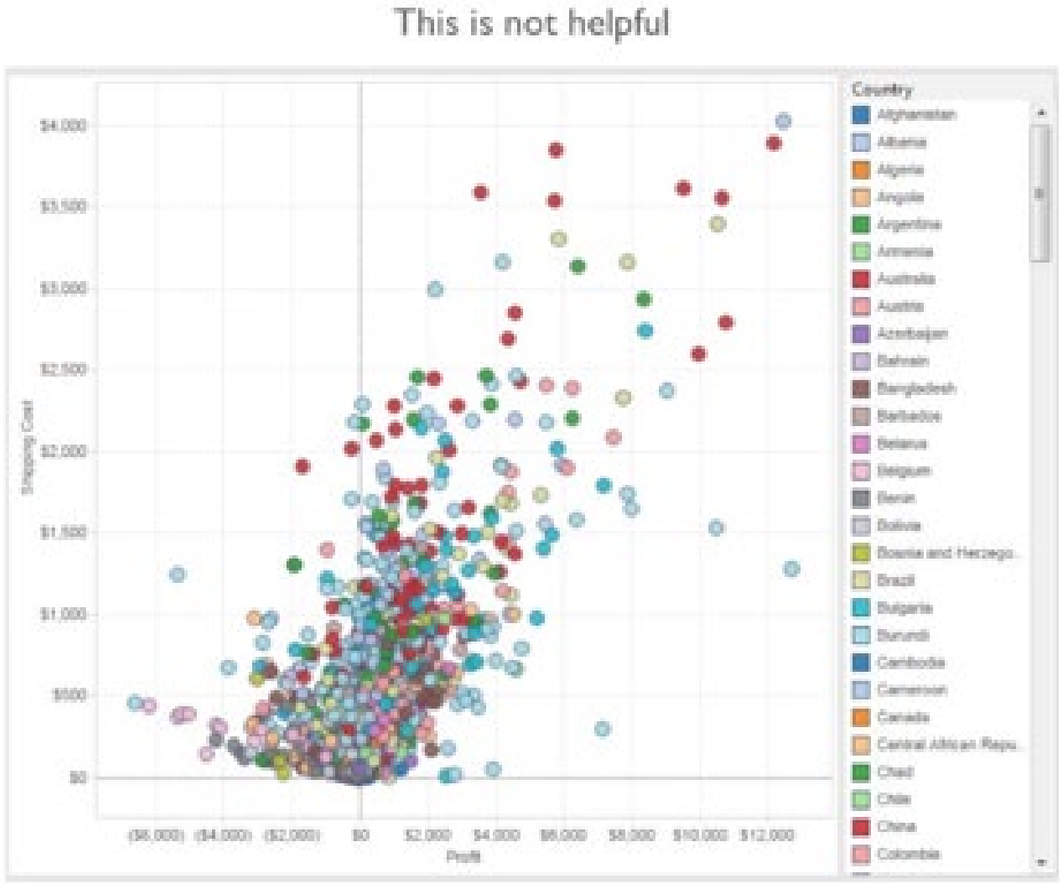
- Avoid red/green palettes! Blue/orange is a good alternative.

- For continuous data, color ramps are effective.
- For discrete data, always try to limit colors (under 5 is ideal).
  - The use of too many colors makes it hard to distinguish, and also requires frequent referencing of the legend.
  - If you limit yourself to just a few colors, then your audience can actually remember what's what. Be kind to your reader!

Using Color

Color me impressed

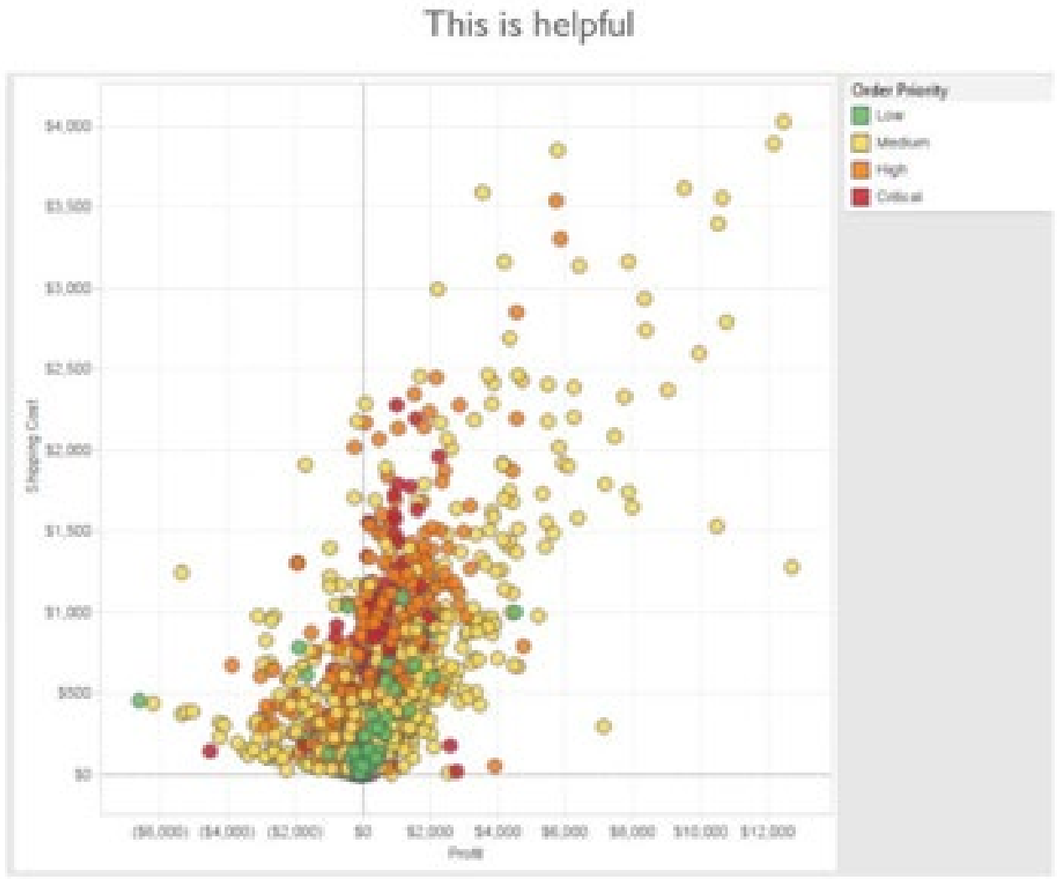
Humans can only distinguish ~8 colors at a time



Using Color

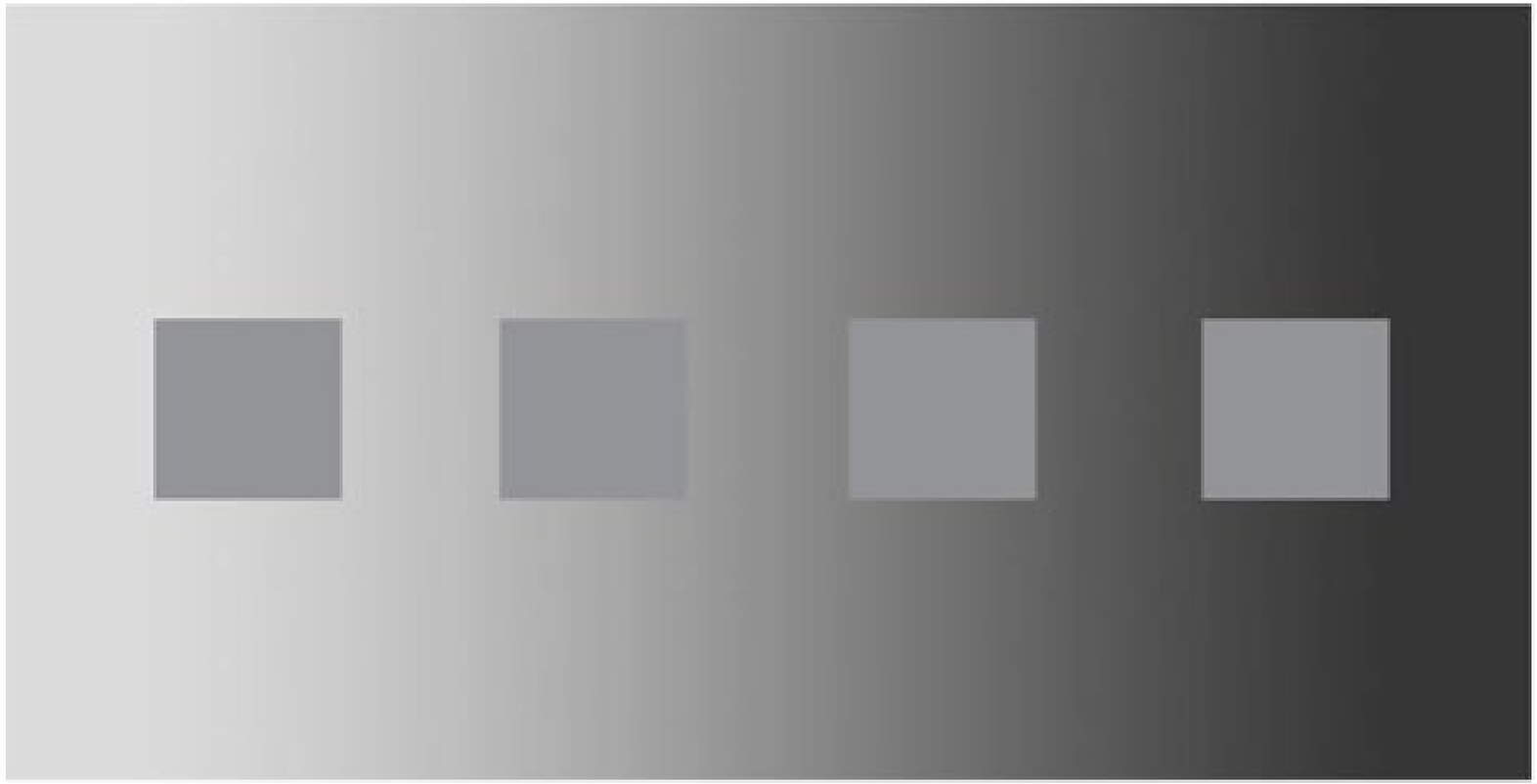
Color me impressed

Humans can only distinguish ~8 colors at a time



Color Me Impressed

Color perception is relative, not absolute

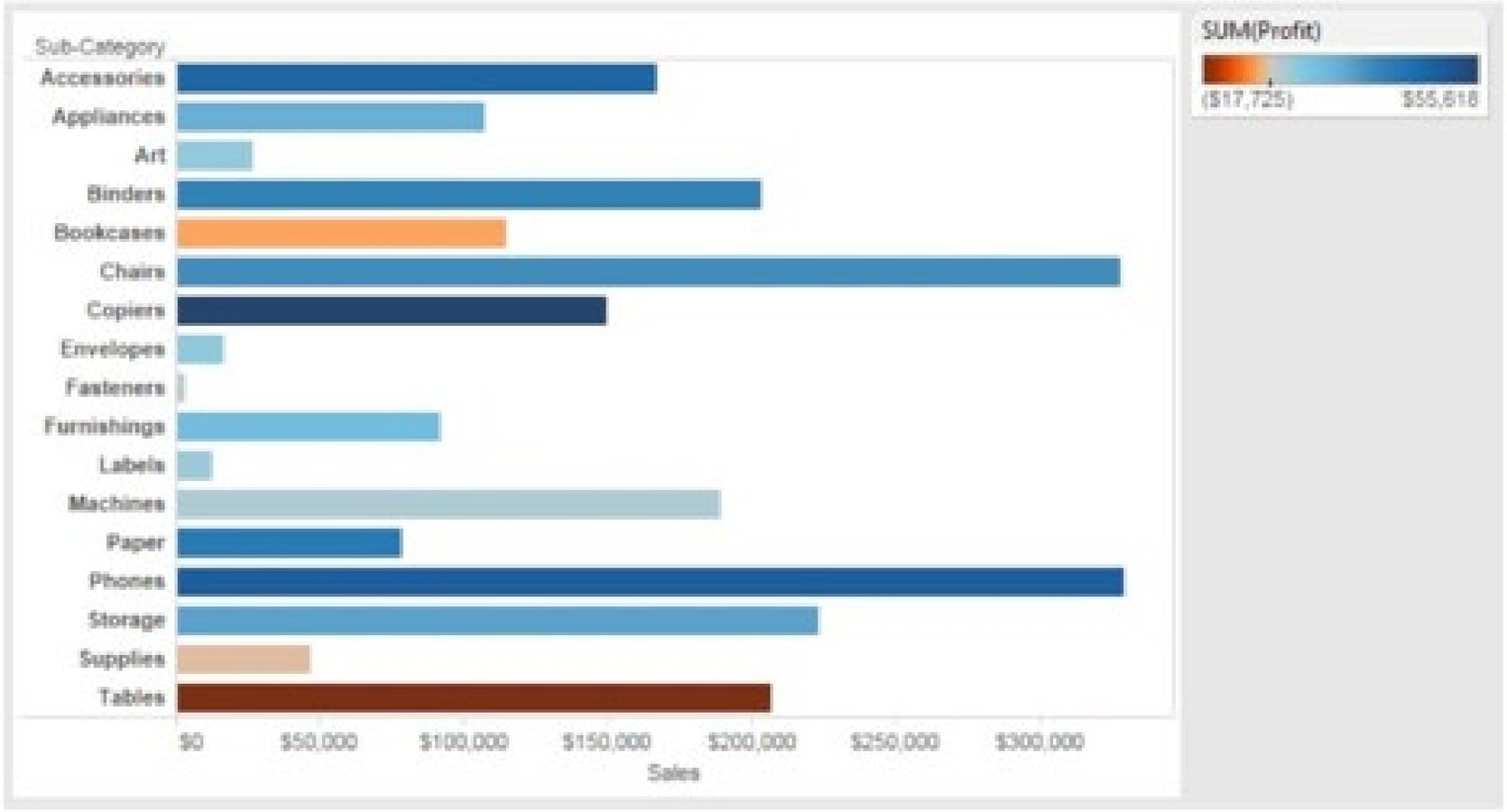


Color Me Impressed

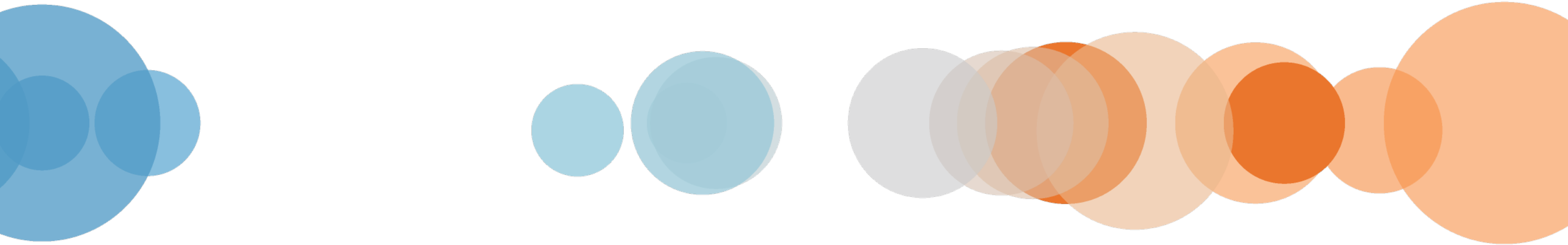
Color perception is relative, not absolute

Color Me Impressed

For quantitative data, color intensity and diverging palettes work well



Visualization Best Practices

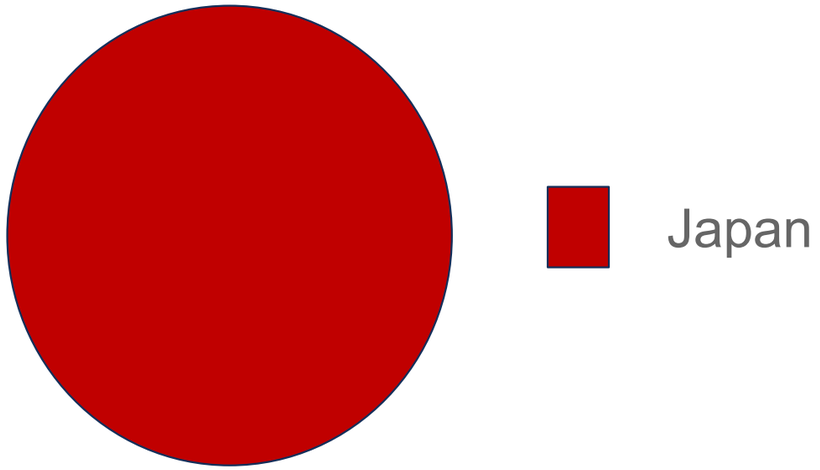


Your brain on autopilot (Kahneman's "System 1")

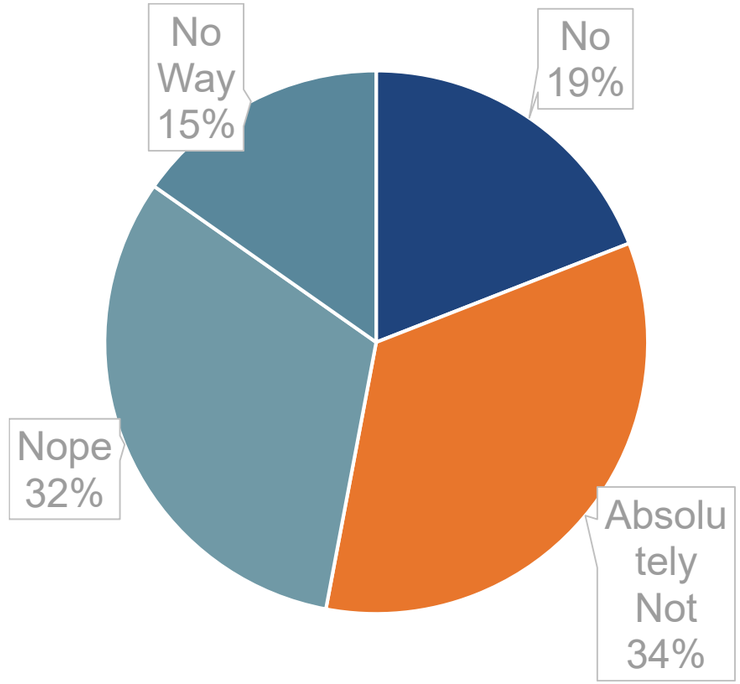
Evryneoe can raed txet, eevn if the letetrs are scrmeblad. The olny tnihg taht mtetras is taht the fsrit and lsat leertts are in the rgiht pcales. Tihs is one of the mnay enctixig tgnihs you wlil laern in tihs cslas!

Pie Charts – NO!

The only effective pie charts that I have seen:

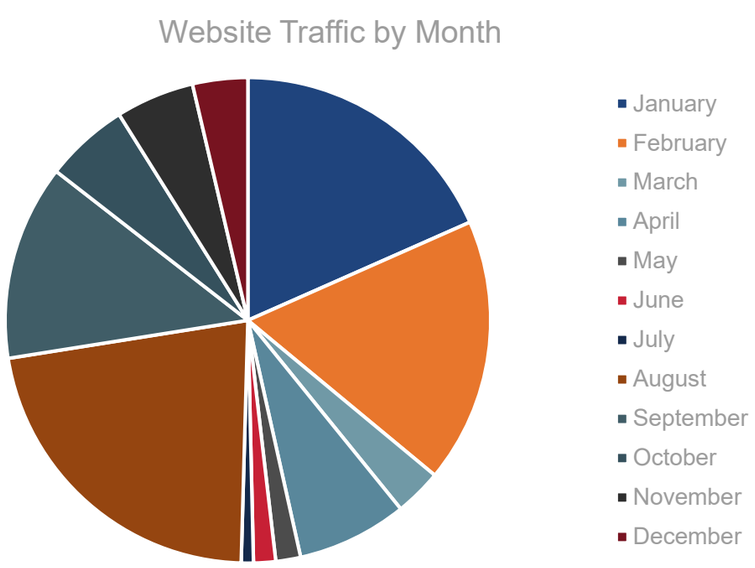


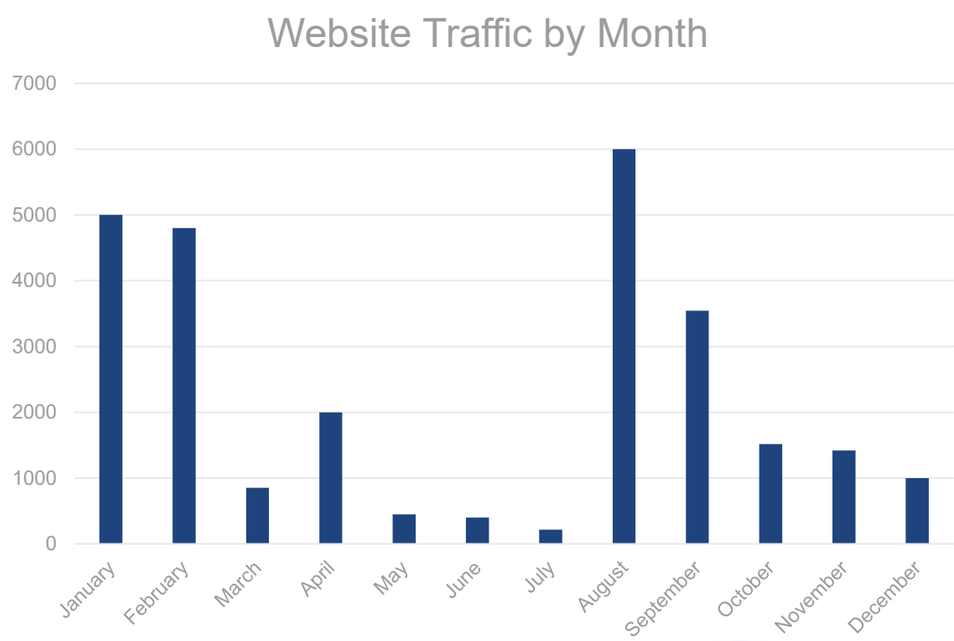
Is a pie chart a good idea?



Pie Charts – NO!

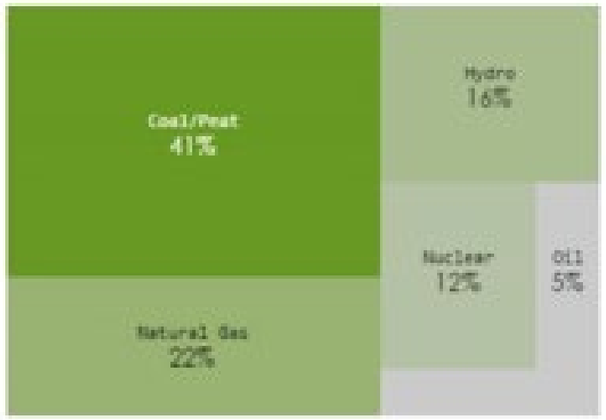
From visualization guru Edward Tufte: "*the only worse design than a pie chart is several of them*."
Pie charts are tempting because everyone understands what they are meant to convey (the various parts of a whole), but it is much harder to compare slices of a pie than it is to compare length or height (especially when slices become very thin):



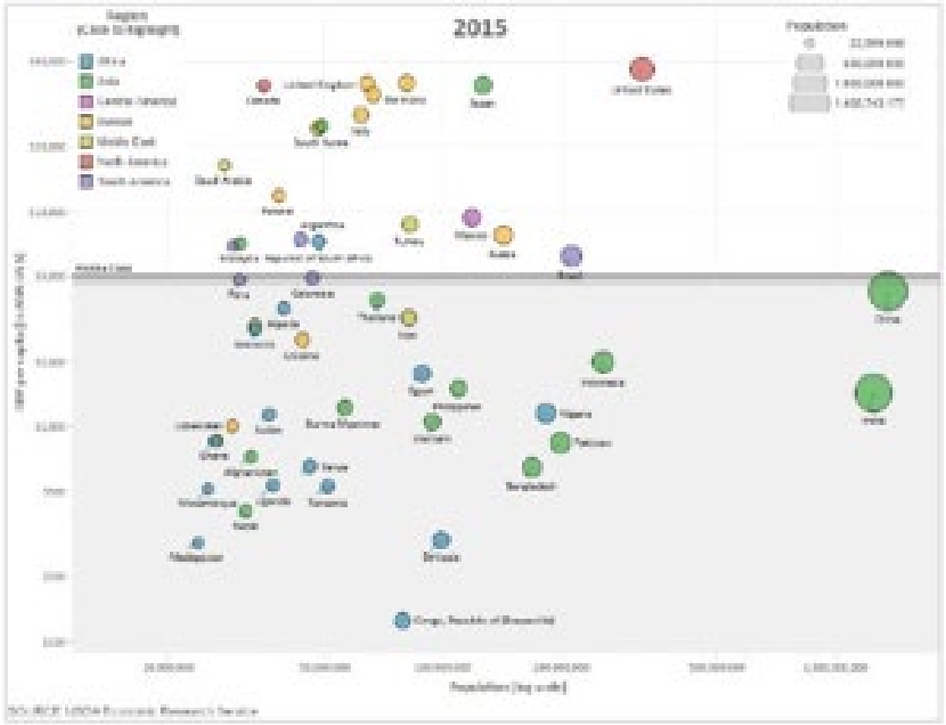


Best Practices

Relative proportions:
treemap



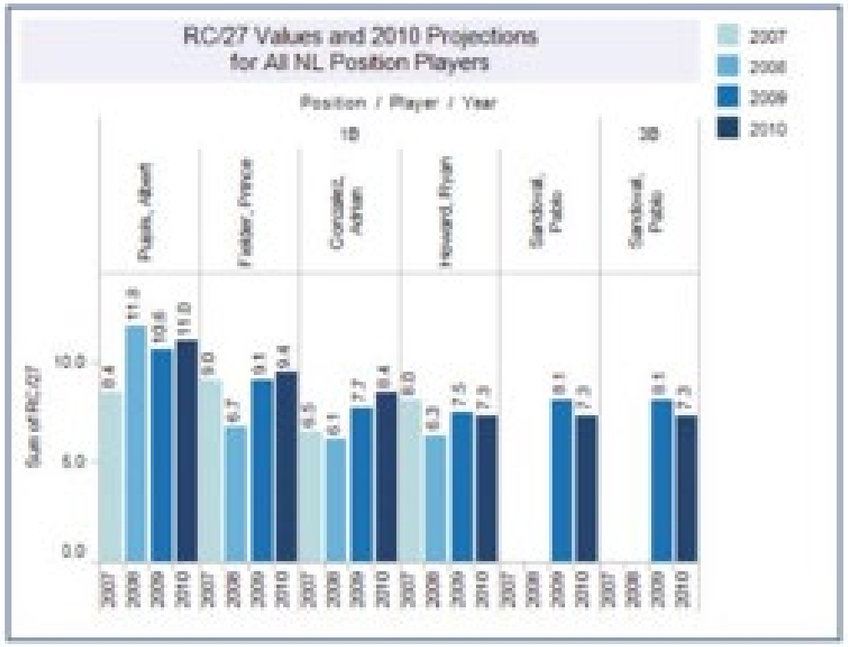
Exploring relationships:
scatter plot



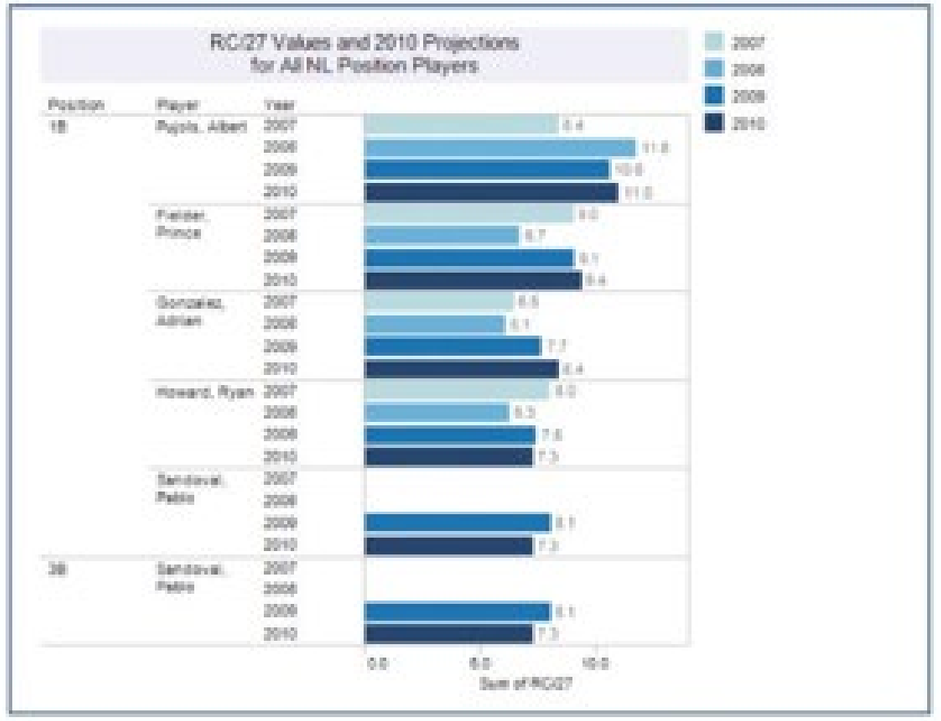
Best Practices

Orient data so people can read it

Good



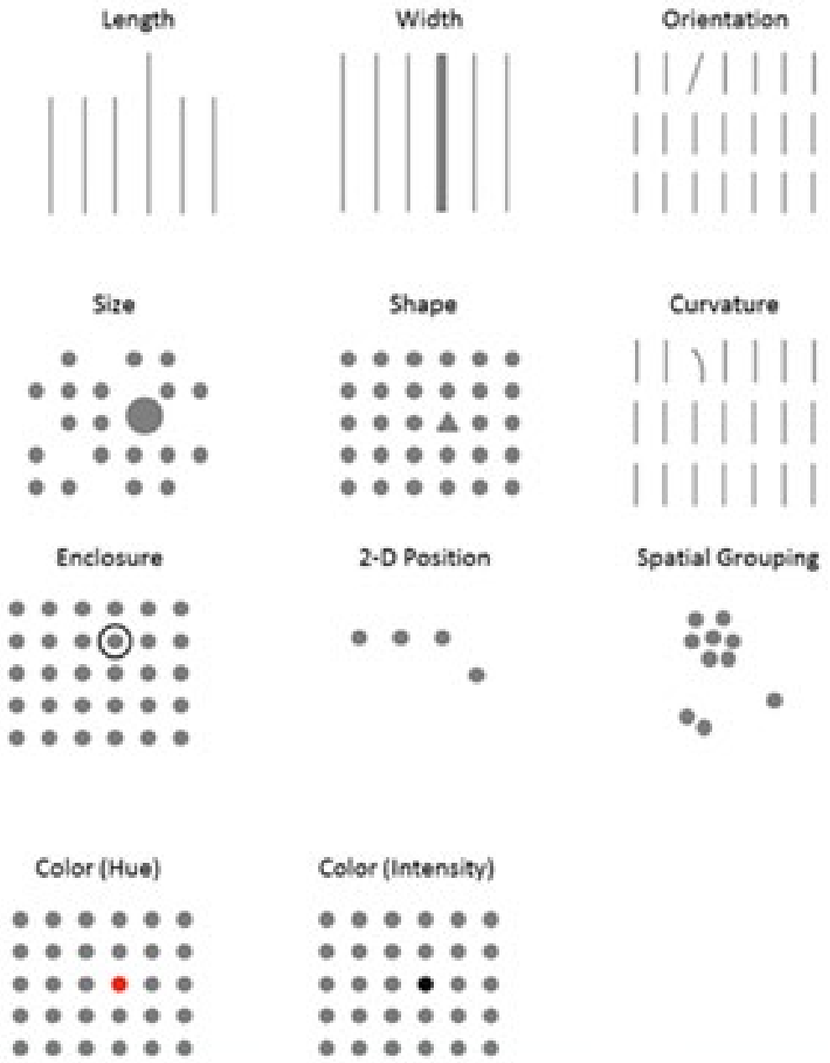
Better



Type of data

Qualitative (categorical)	Qualitative (ordinal)	Quantitative
Arizona, New York, Texas	Gold, silver, bronze	Weight (10 lbs, 20 lbs, 5000 lbs)
Sarah, John, Maria	Excellent health, good health, poor health	Cost ($50, $100, $0.05)
Coors, Bud Light, Stella Artois	Love it, like it, hate it	Discount (5%, 10%, 12.8%)

Pre-attentive Visual Attributes

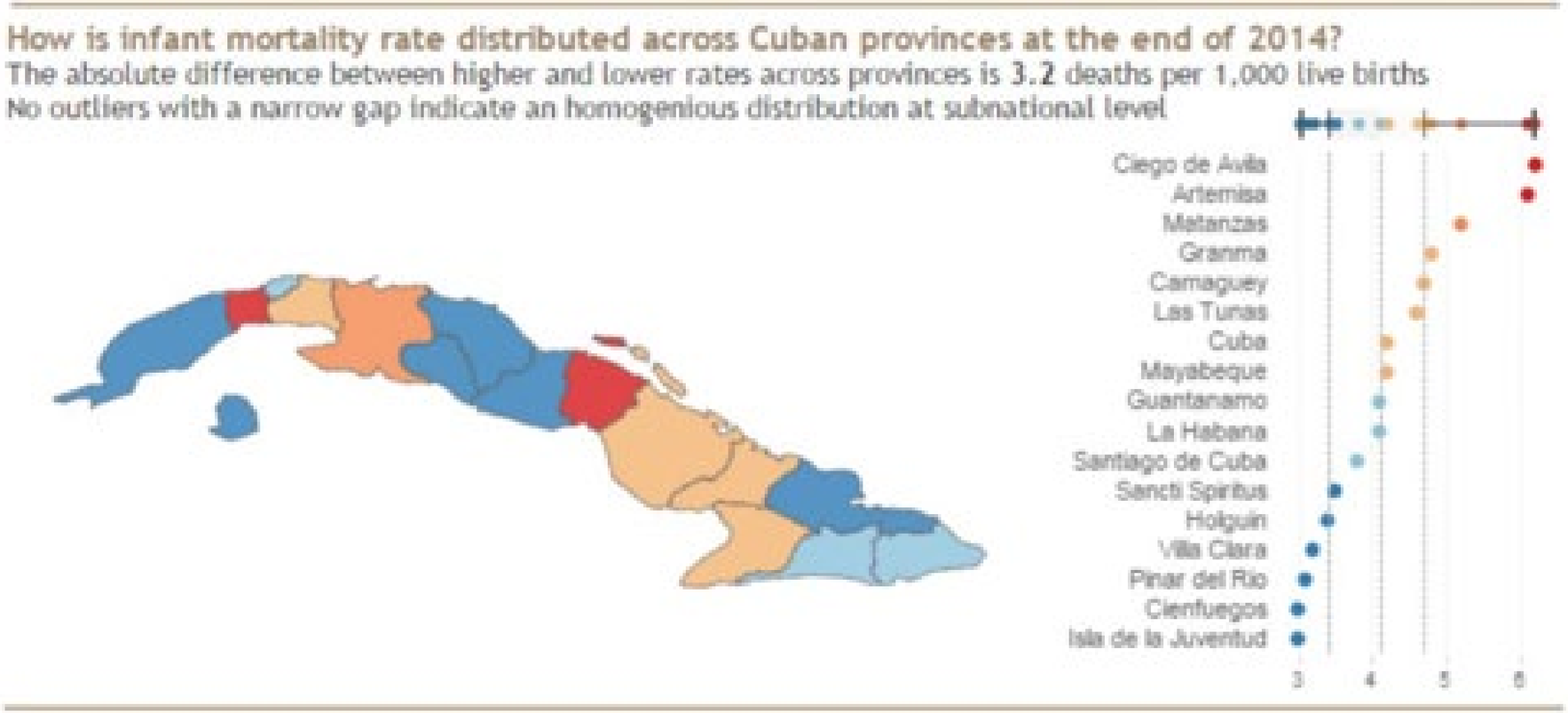


How do Humans Like Their Data?

Categorical	Ordinal	Quantitative
Position	Position	Position
Shape	Size	Length
Different Colors	Color Intensity	Size
	Different Colors	Color Intensity
	Shape	

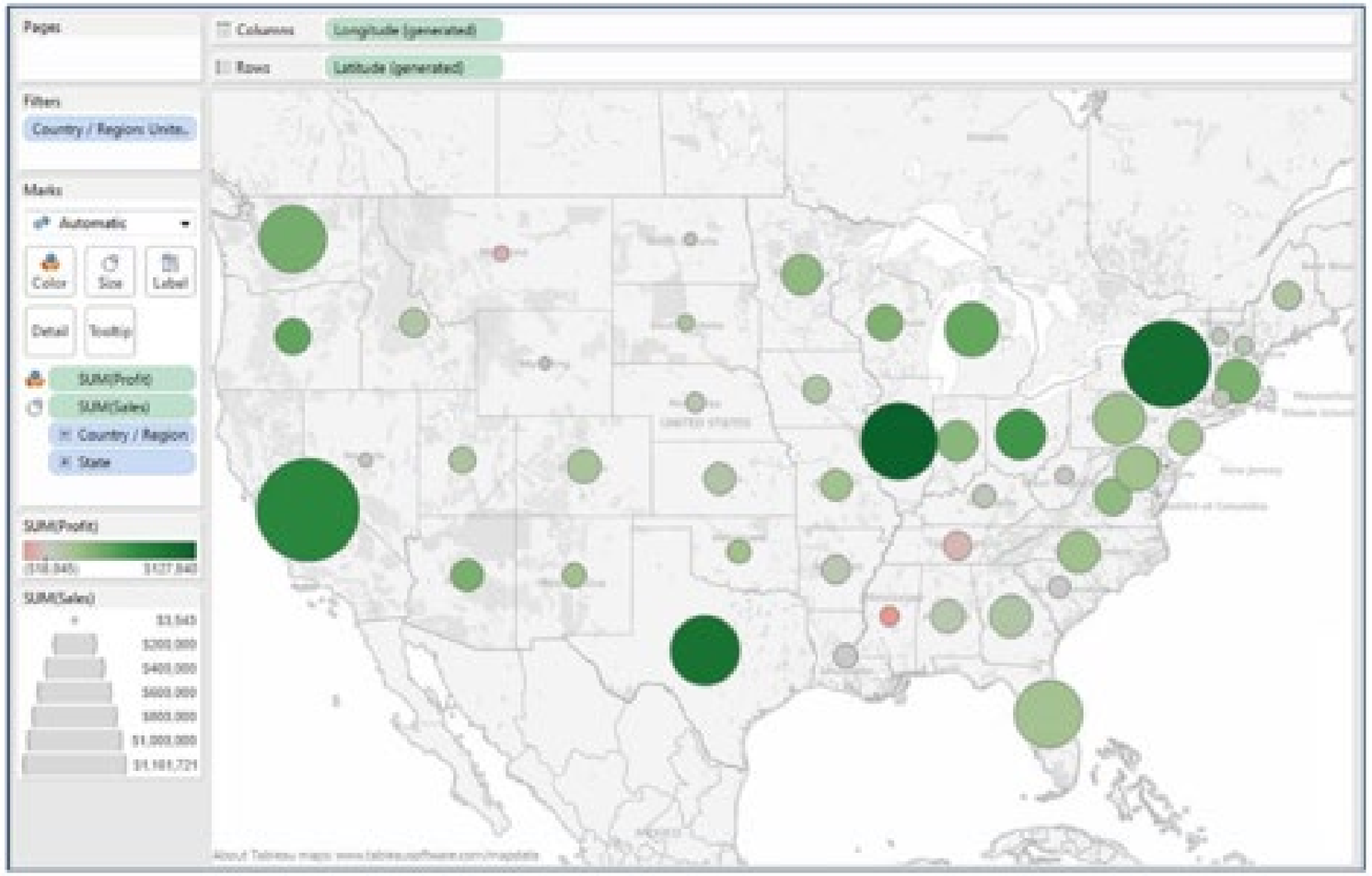
Mapping to Insight

Use filled maps for defined areas and only ONE measure



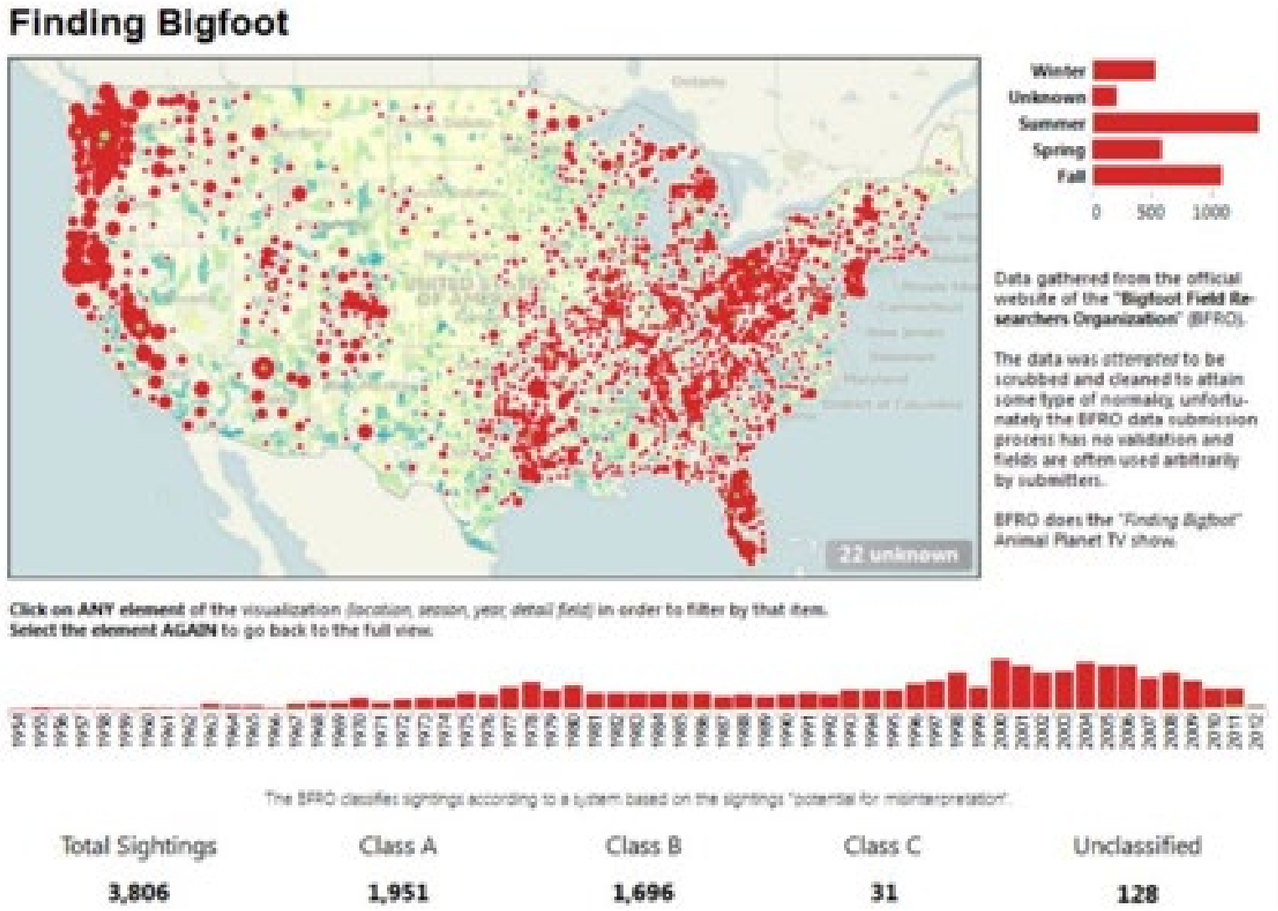
Mapping to Insight

Filled maps won’t work for multiple measures



Dashboards

Dashboards should pass the five second test



Dashboarding for the Five Second Test

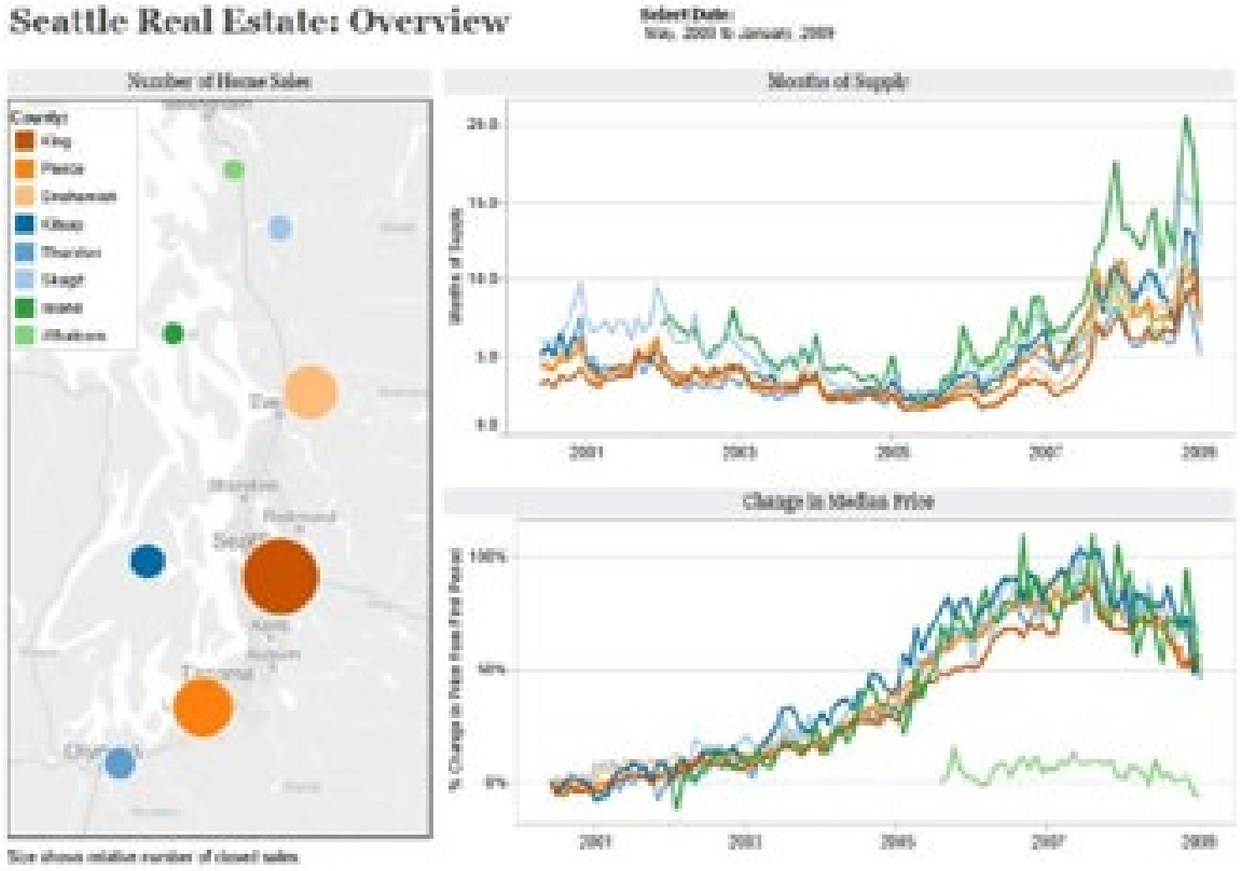
Most important view goes on top or top-left

Legends go near their views

Avoid using multiple color schemes on a single dashboard

Use 5 views or fewer in dashboards

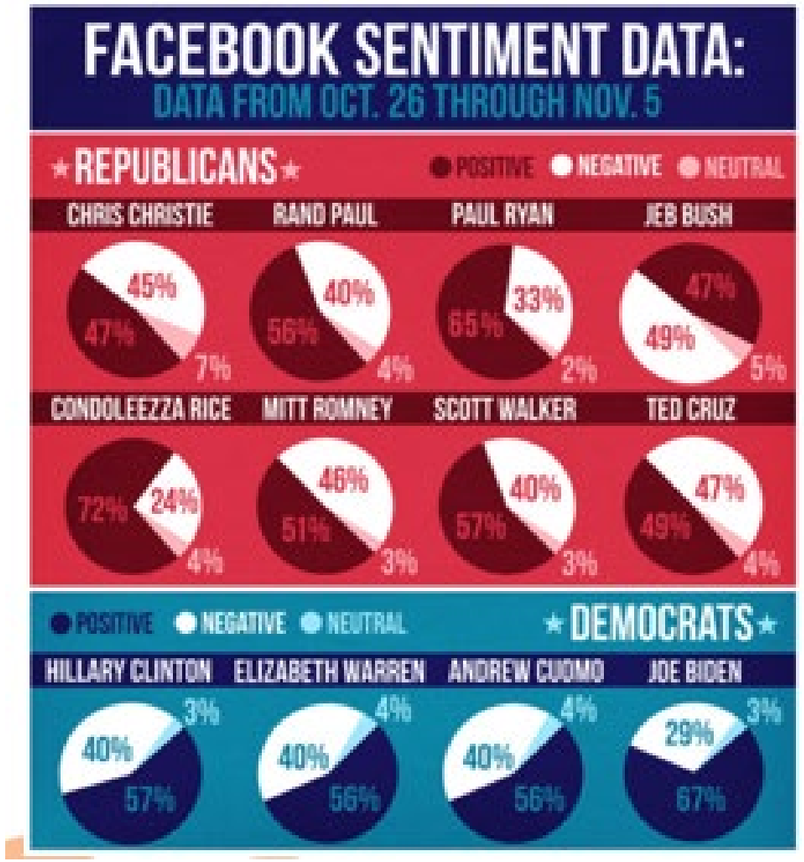
Provide interactivity



Five Questions to Consider in a Dashboard Review

1. What problem/question does this solve/answer?
2. Is this really the best way to display the information?
3. Does everything add value?
4. Is there functional interactivity?
5. Are there clear labels?

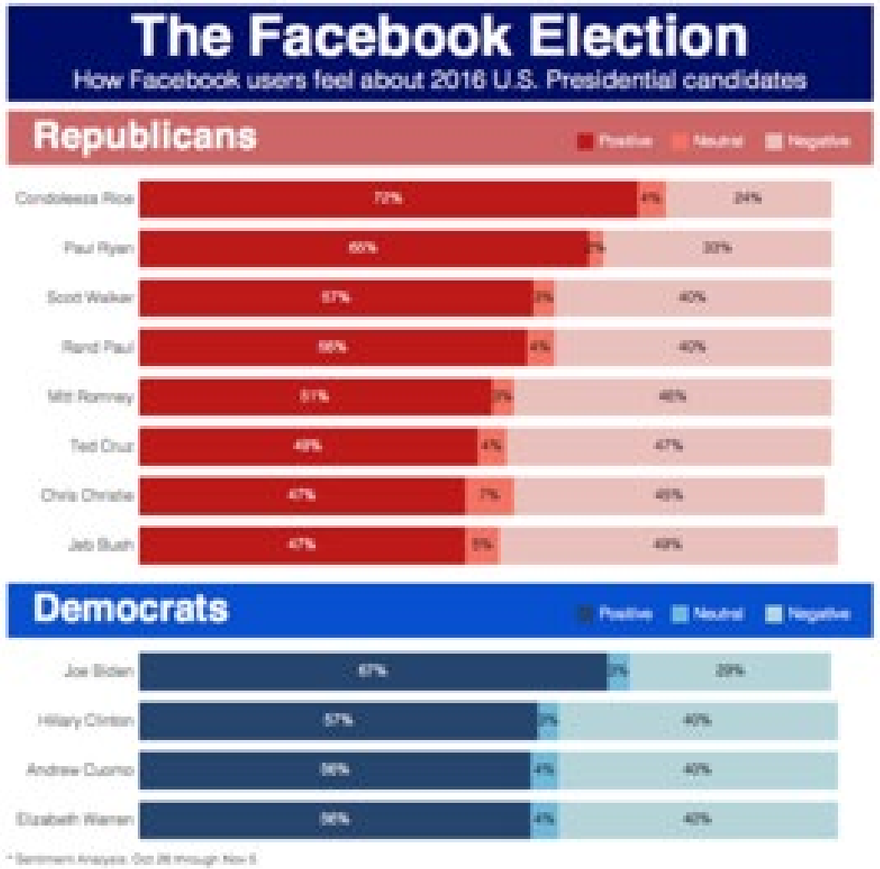
Dashboard Review Example: Before



(2014 poll data)
How can this improve?

1. What problem/question does this solve/answer?
2. Is this really the best way to display the information?
3. Does everything add value?
4. Is there functional interactivity?
5. Are there clear labels?

Dashboard Review Example: After



What's changed?

- Title clearly addresses question at hand
- Bar charts for cleaner comparison
- Labels – showing more with less
- Sorting of bars for stronger visualization of sentiment


+ableau®