

Until now

1. **Jupyter** and **Python** basic knowledge
2. Importing / Cleaning data
3. **Pandas** for Exploratory Data Analysis
4. SciPy, Matplotlib, NumPy, Folium and API's
5. **Presenting results and story telling**



Today

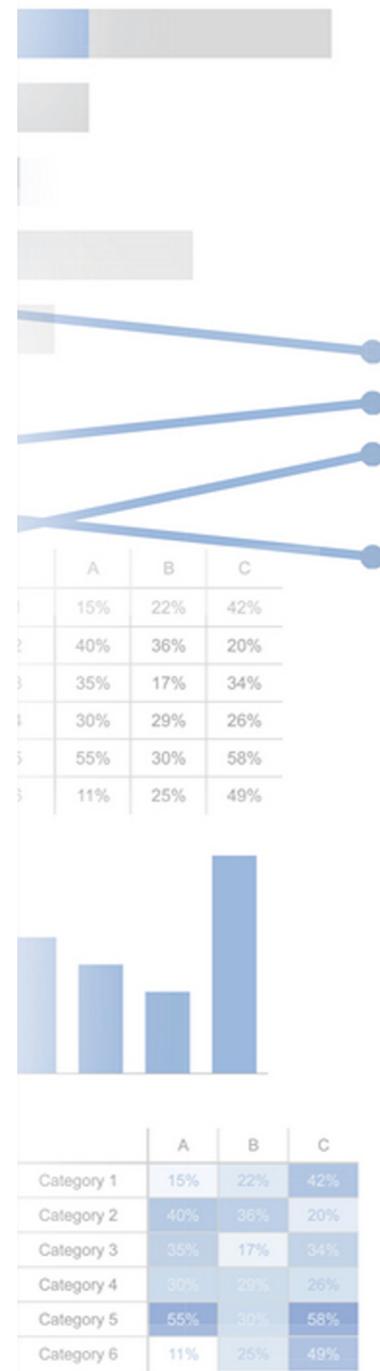
1. **explain data with good visualizations**

2. Scopus

- For what can I use it?
- Scopus API



explain data with good visualizations



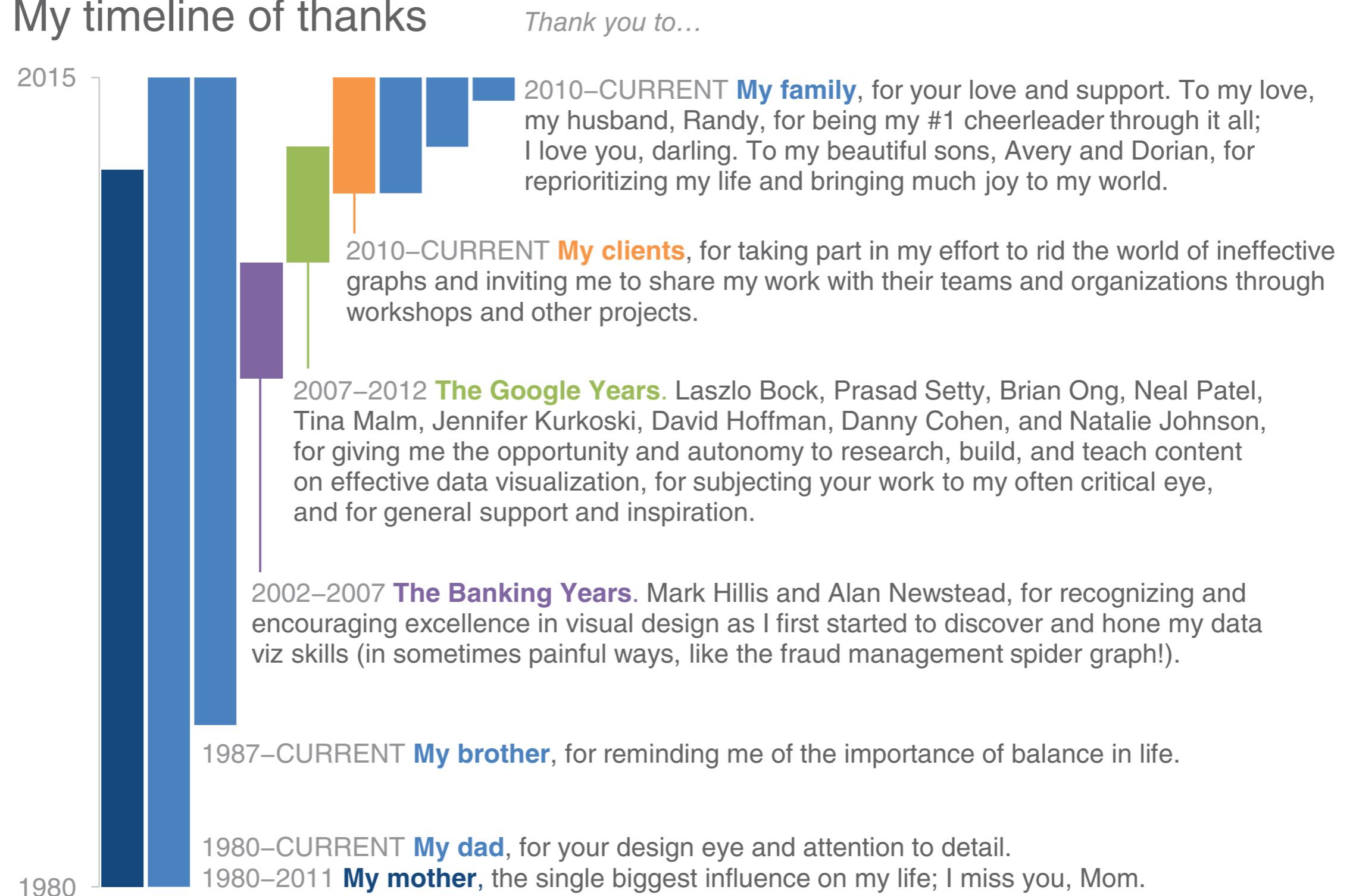
cole nussbaumer knaflic

storytelling with data

a data
visualization
guide for
business
professionals

acknowledgments

My timeline of thanks



explain data with good visualizations

Useful for...

- anyone that needs to communicate using data
- **you** sharing results of your work with your colleges or employees
- **students visualizing** thesis data or preparing presentations for conferences

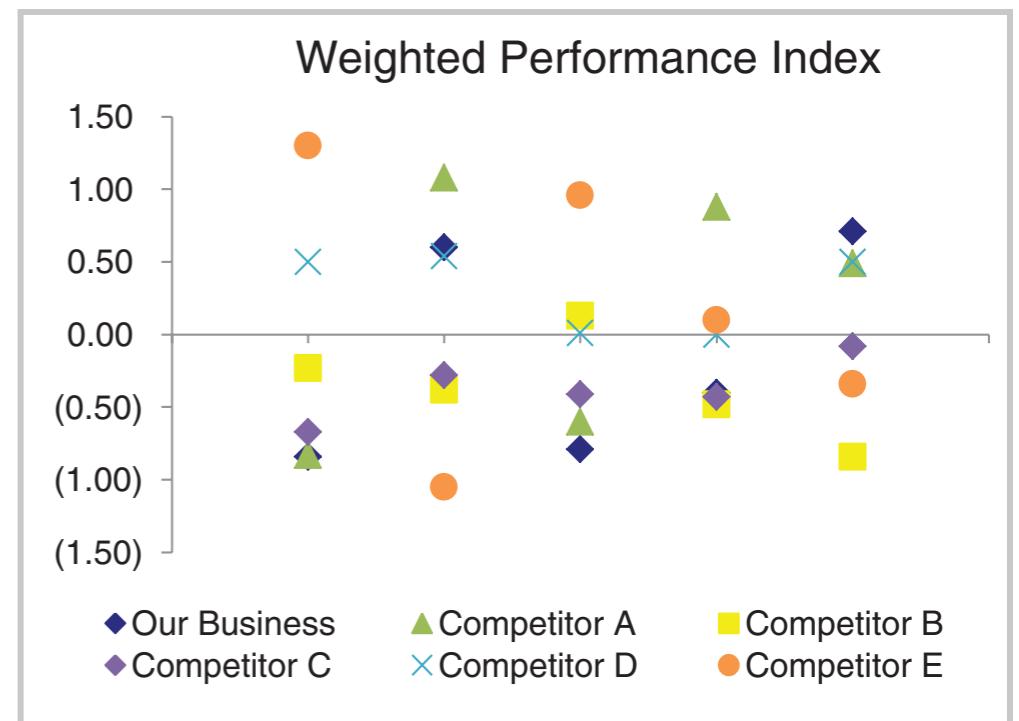
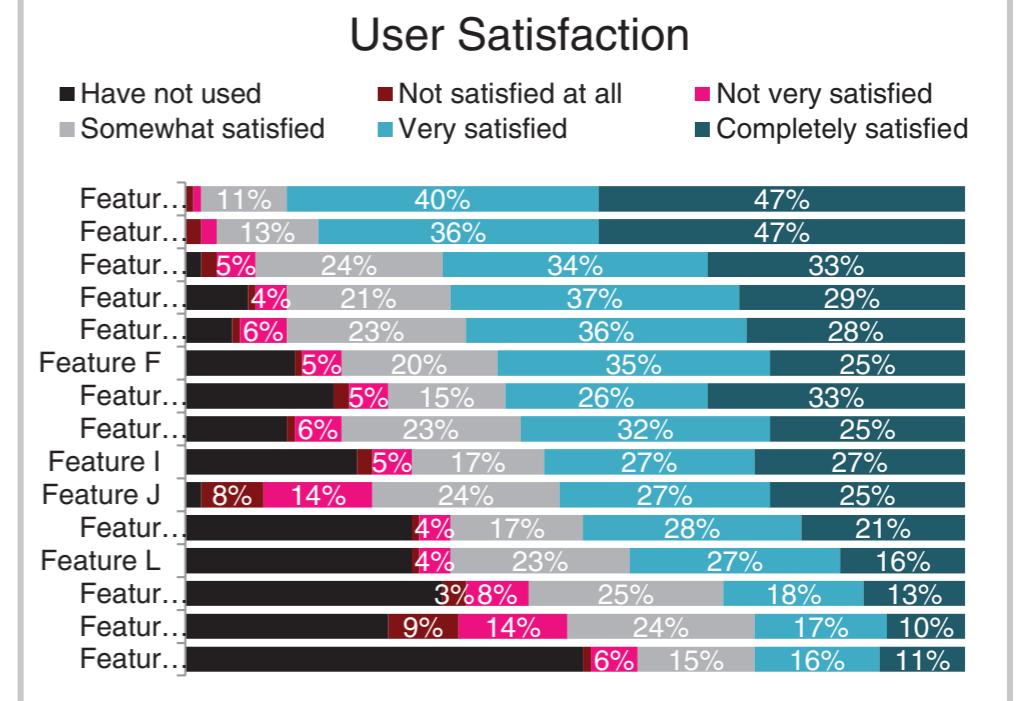
Why important ...

- typically the only part of the analytical/experimental process that your audience ever sees
- **difference between success or failure** by communicating the findings of your study (raising money, or getting your point across to your audience)

problem

naturally not good at storytelling with data

- learned about language and math in school
- **nobody teaches us** how to make good visualisations or tell stories with data
- poorly prepared for important task
- **Result:** plots impossible to understand, meaningless color
- No intention. But it happens. Again and again.



explain data with good visualizations

By

1. **Understand the context:** Who is audience? What do you need them to know or do?
2. **Choose an appropriate visual display:** Best way to show the data?
3. **Eliminate clutter:** clutter is your enemy!
4. **Focus attention where you want it:** Use preattentive attributes strategically!
5. **Think like a designer:** Form follows function
6. **Tell a story:** Use the magic of stories!

1. the importance of context

Exploratory (1st) vs. explanatory analysis (2nd)

- **exploratory data analysis** is hunting for pearls in oysters (open 100 oysters to find 2 pearls)
- communicating our analysis, **explanatory space**: a specific story you want to tell (probably about those two pearls)
- Here, we focus on **explanatory analysis!**
(Don't make your audience reopen all of the oysters!)



Who, what, and how?

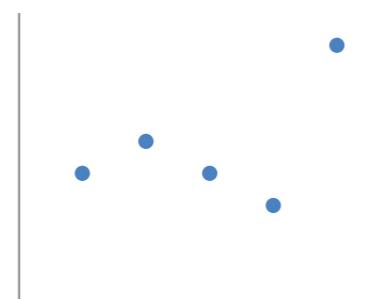
- to whom are you communicating? who your **audience**? how they perceive you?
- What do you want your audience to know or do? how you want audience to act?
- **How can you use data to help make your point?**

2. choosing an effective visual

What is the right graph for my situation?

91%

Simple text



Scatterplot

	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%

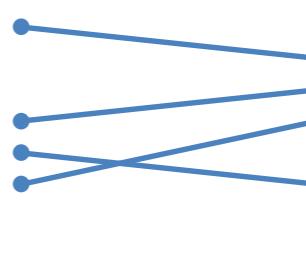
Table



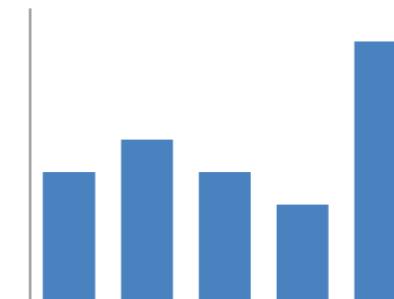
Line

	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%

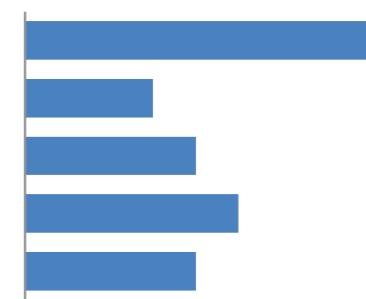
Heatmap



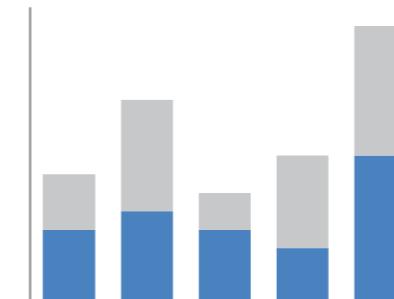
Slopegraph



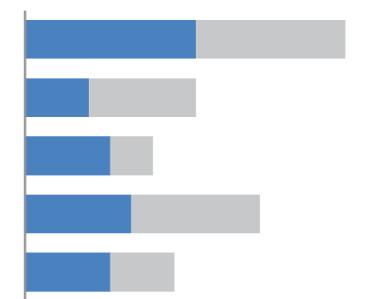
Vertical bar



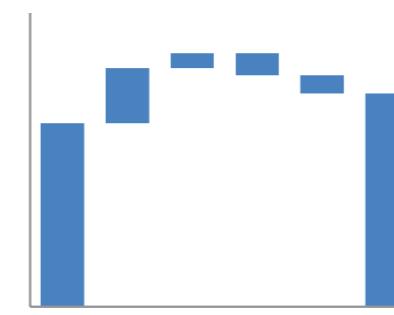
Horizontal bar



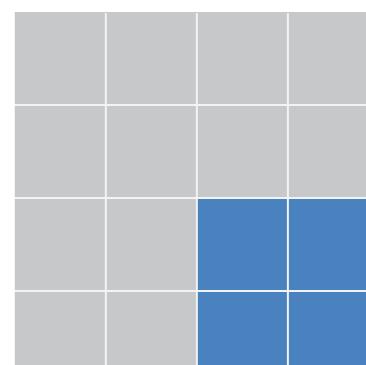
Stacked vertical bar



Stacked horizontal bar



Waterfall



Square area

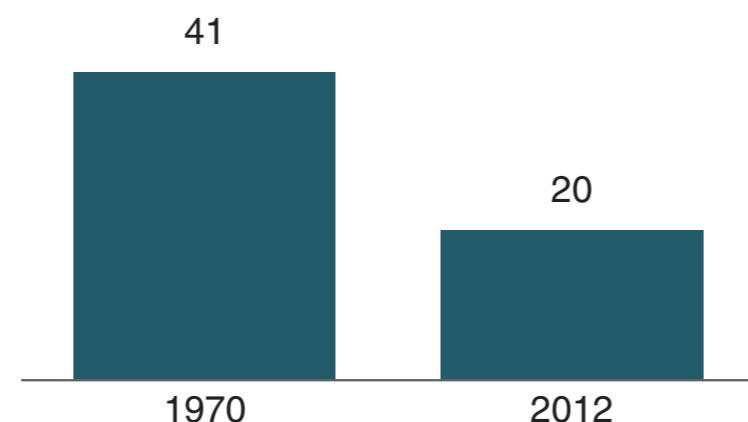
2. choosing an effective visual

Simple text or number

- just a number or two to share: **simple text** can be a great way to communicate

Children with a "Traditional" Stay-at- Home Mother

*% of children with a married
stay-at-home mother with a
working husband*



Note: Based on children younger than 18.
Their mothers are categorized based on
employment status in 1970 and 2012.

Source: Pew Research Center analysis of
March Current Population Surveys
Integrated Public Use Microdata Series
(IPUMS-CPS), 1971 and 2013

Adapted from PEW RESEARCH CENTER

2. choosing an effective visual

Simple text or number

- just a number or two to share: **simple text** can be a great way to communicate

20%

of children had a
traditional stay-at-home mom
in 2012, compared to 41% in 1970

2. choosing an effective visual

Tables

- want design to fade into background, letting the **data take center stage**
- don't let heavy borders compete for attention
- interact with verbal system (**read** them across rows and down columns)

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	\$X.X	Y%	Z,ZZZ
Group 2	\$X.X	Y%	Z,ZZZ
Group 3	\$X.X	Y%	Z,ZZZ
Group 4	\$X.X	Y%	Z,ZZZ
Group 5	\$X.X	Y%	Z,ZZZ

2. choosing an effective visual

Tables

- want design to fade into background, letting the **data take center stage**
- don't let heavy borders compete for attention
- interact with verbal system (**read** them across rows and down columns)

Table

	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%

2. choosing an effective visual

Tables

- want design to fade into background, letting the **data take center stage**
- don't let heavy borders compete for attention
- interact with verbal system (**read** them across rows and down columns)

Table

	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

LOW-HIGH

	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

- tables with leveraged colored cells
- here we use color saturation to reduce mental processing
- helping your eyes and brains more **quickly to target**

2. choosing an effective visual

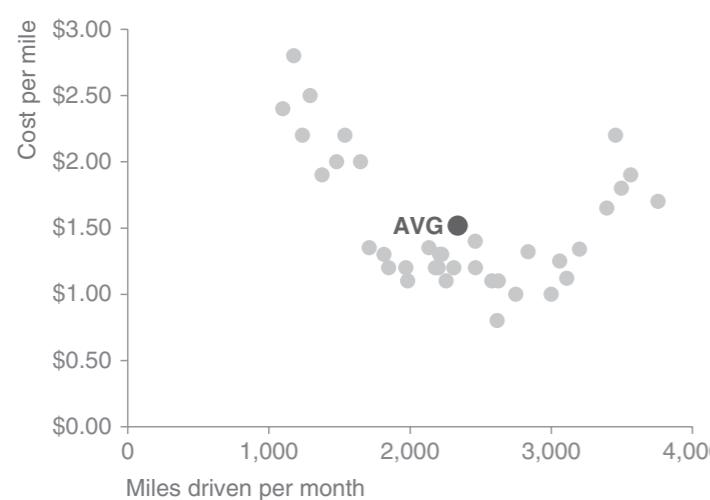
Graphs

- graphs interact with our visual system: **faster** at processing information
- a well-designed graph will typically get the information across more quickly

Scatterplots

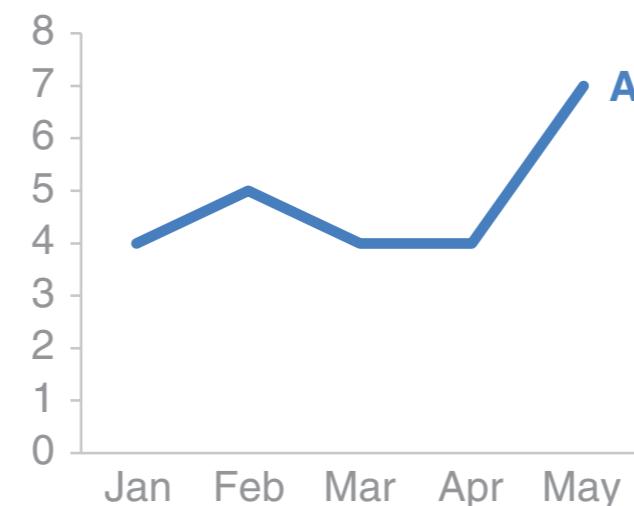
useful for showing relationships

Cost per mile by miles driven



Line graph

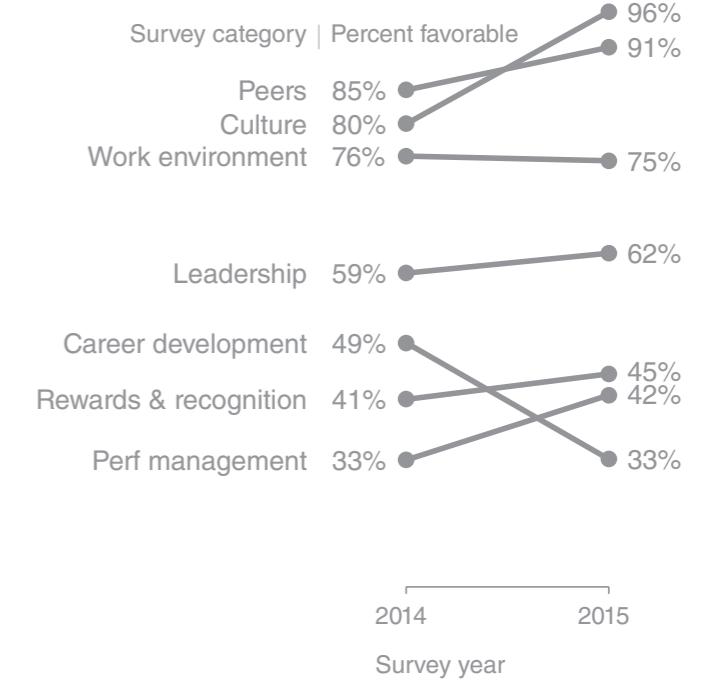
used to plot continuous data



Slopegraph

quickly shows relative differences across various categories

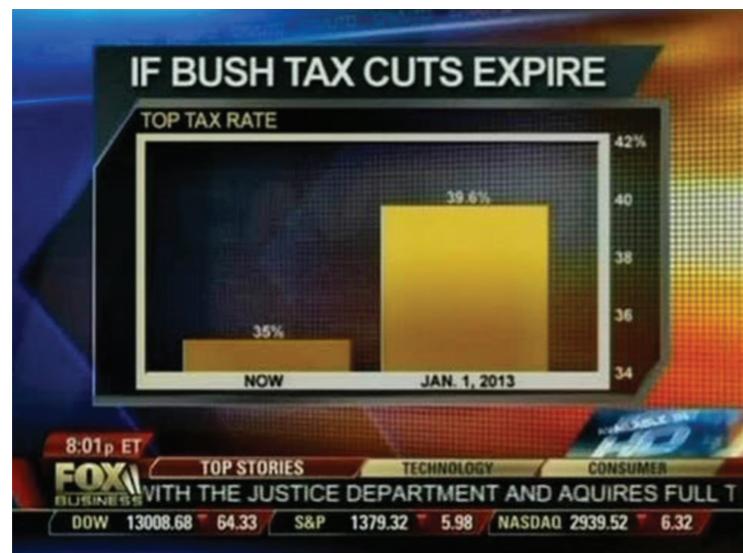
Employee feedback over time



2. choosing an effective visual

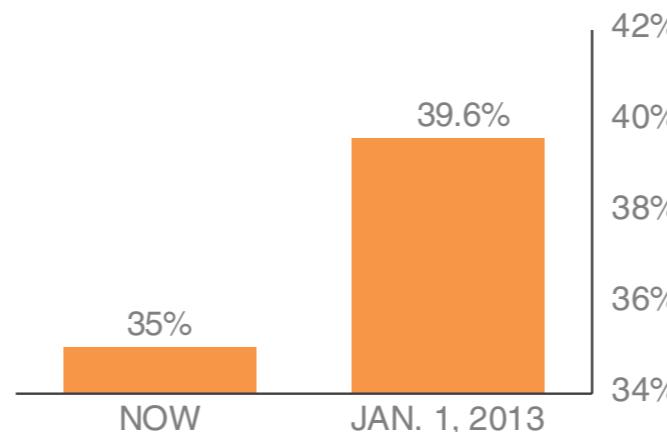
Barplots

- are easy for our eyes to read (our eyes compare the end points of the bars)
- need zero baseline



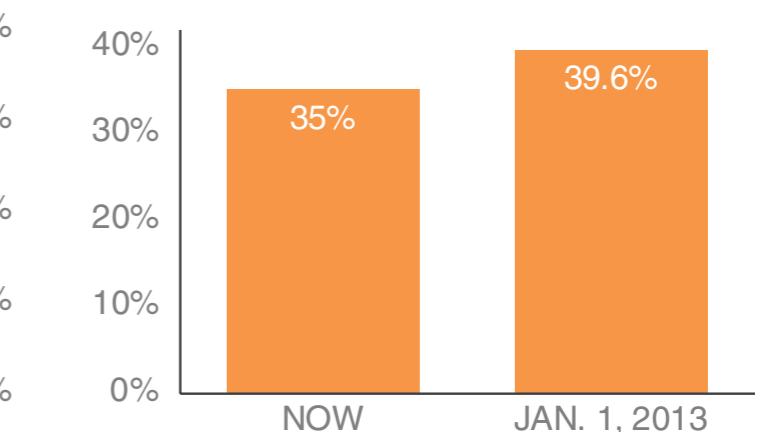
Non-zero baseline: as originally graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE



Zero baseline: as it should be graphed

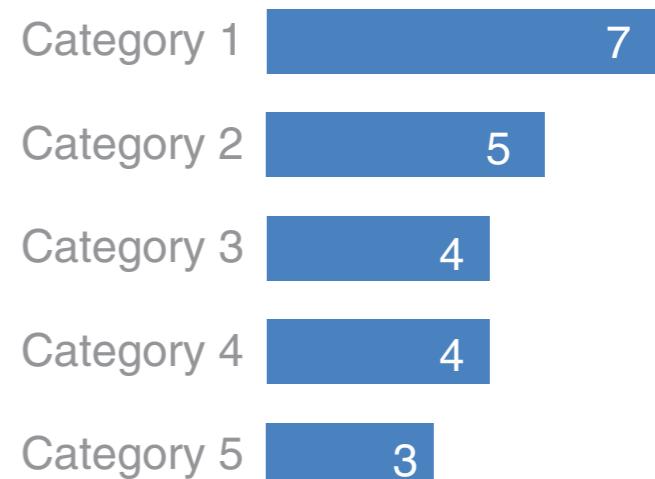
IF BUSH TAX CUTS EXPIRE
TOP TAX RATE



Horizontal bar chart

- especially useful if category names are long
- way we typically process information — starting at top left

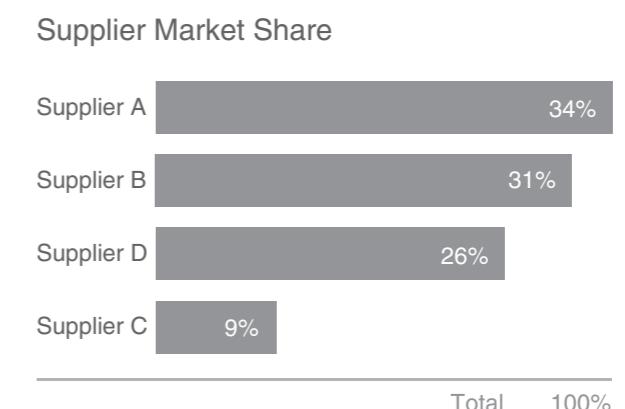
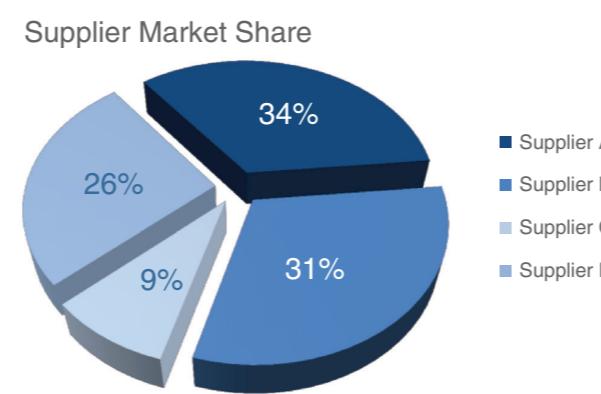
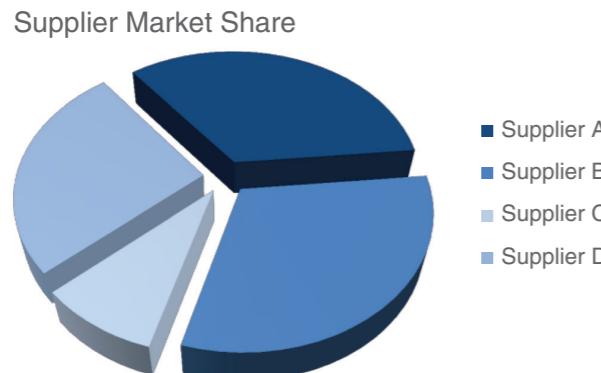
Single series



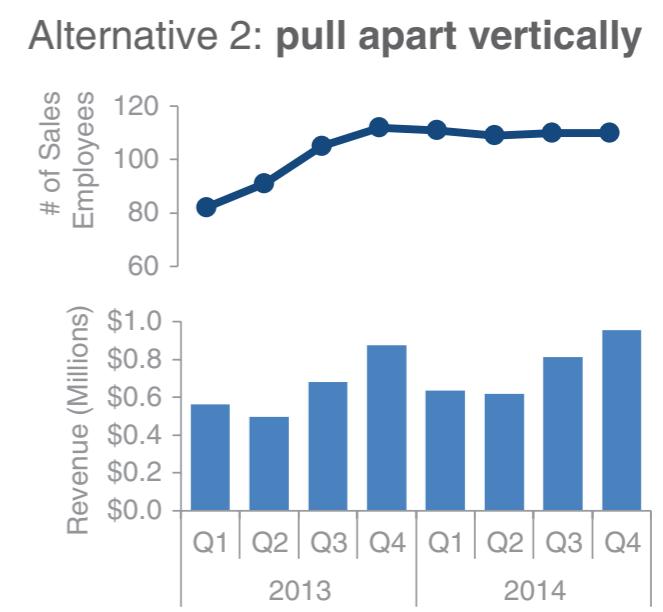
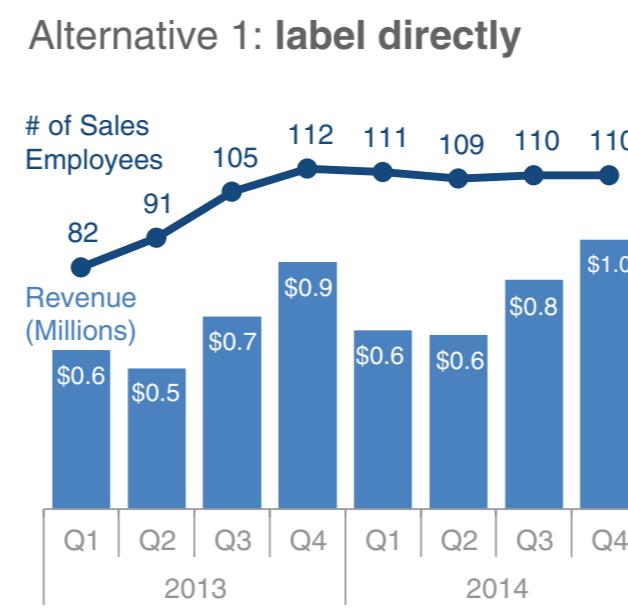
2. choosing an effective visual

To be avoided

- Pie charts are evil



- secondary y-axis: generally not a good idea



2. choosing an effective visual

Closing

- What is the right graph for my situation?
- Answer: **Whatever will be easiest for your audience to read.**

Until now and next...

1. Understand the context
2. Choose an appropriate visual display
3. **Eliminate clutter:** clutter is your enemy!

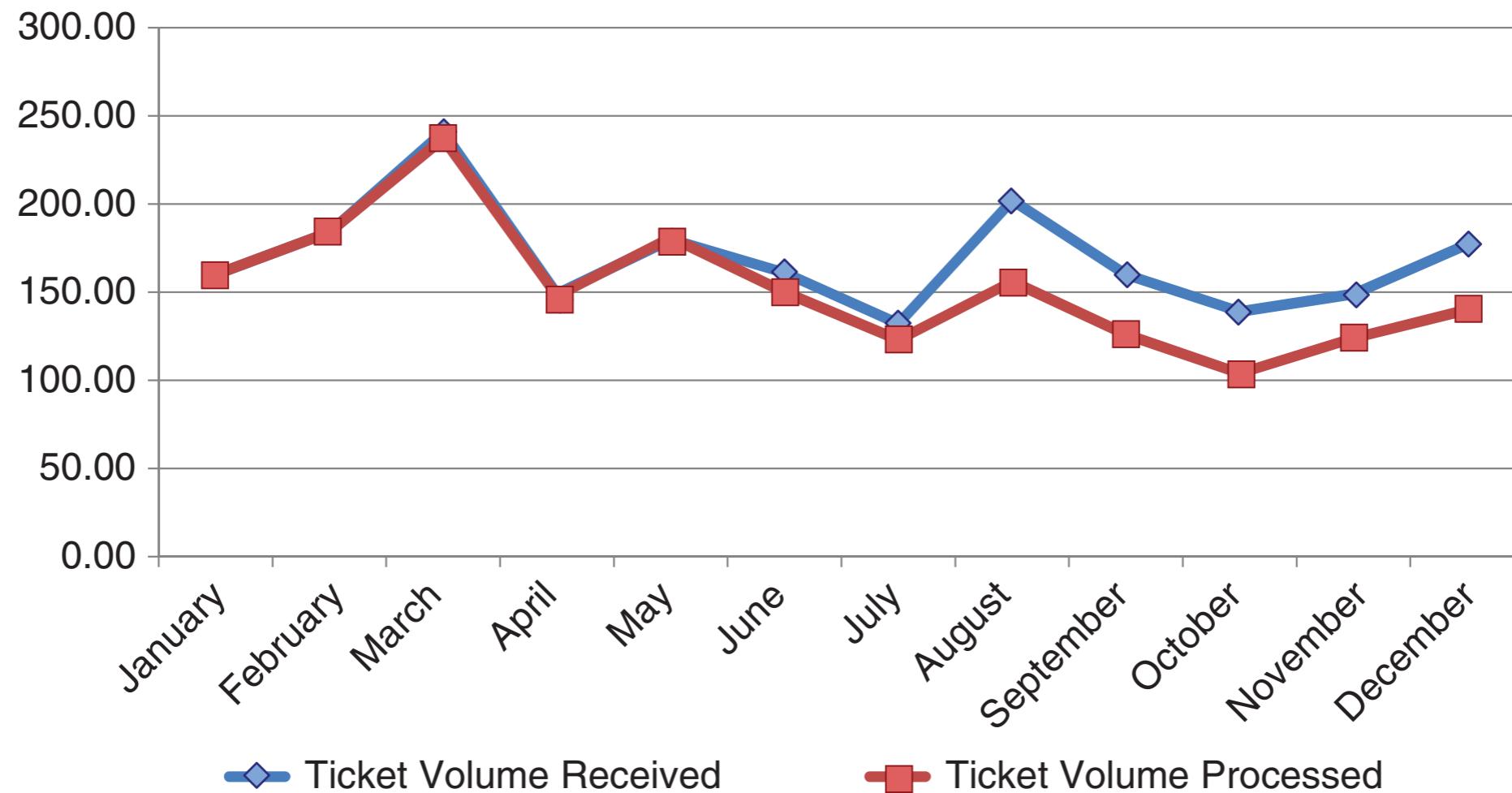
3. clutter is your enemy!

- every element added to **blank page** takes up **cognitive load**
(audience **brains' power to process**)
- **clutter:** visual elements taking up space but don't increase understanding
- **avoid:** using mental resources without providing information
- **perceived cognitive load:** how hard they believe they have to work to understand us
(audience may decide not to give us their attention)



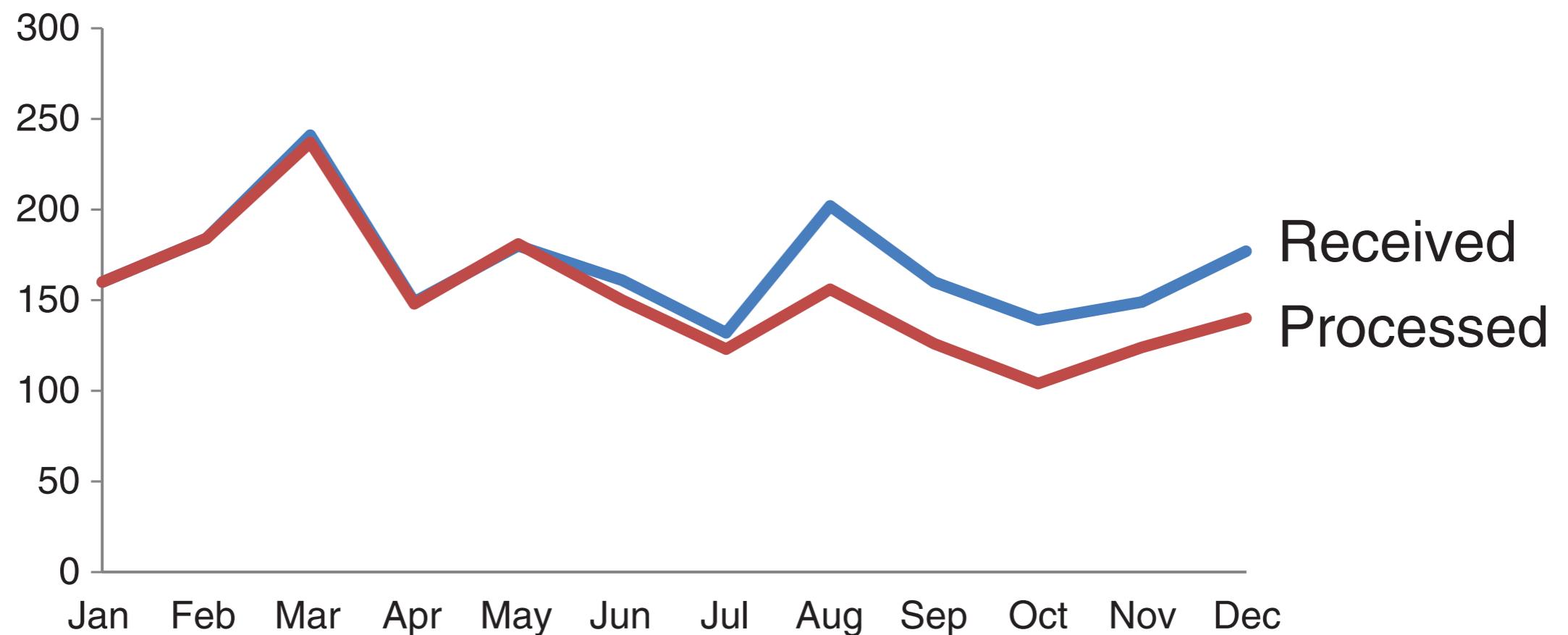
3. clutter is your enemy!

Decluttering: What would you count as clutter and remove?



3. clutter is your enemy!

Decluttering: What would you count as clutter and remove?



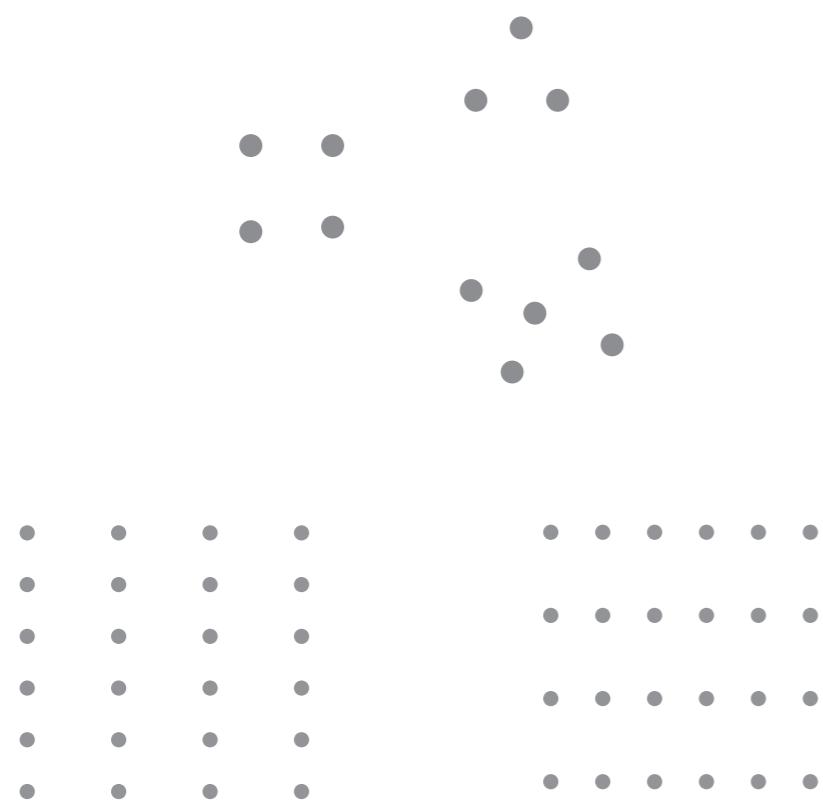
3. clutter is your enemy!

Gestalt principles of visual perception

- remove all unnecessary details and then use your weapons wisely
- help the audience to understand your message easily by using the following principles:

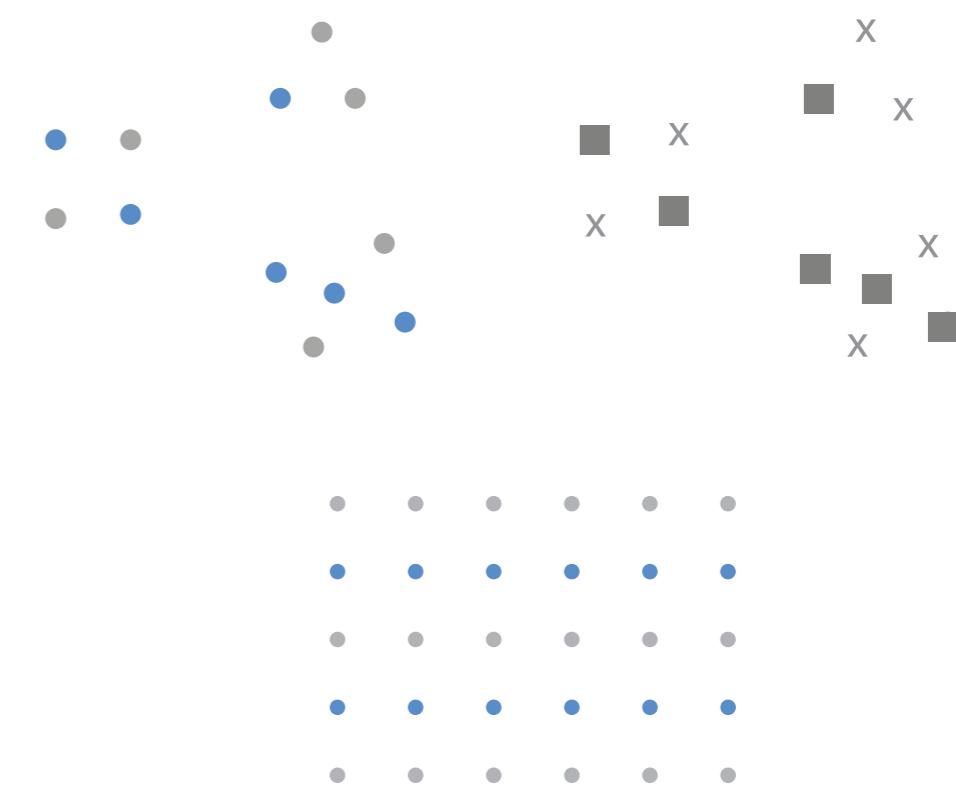
Proximity

- objects that are physically close together belonging to a group



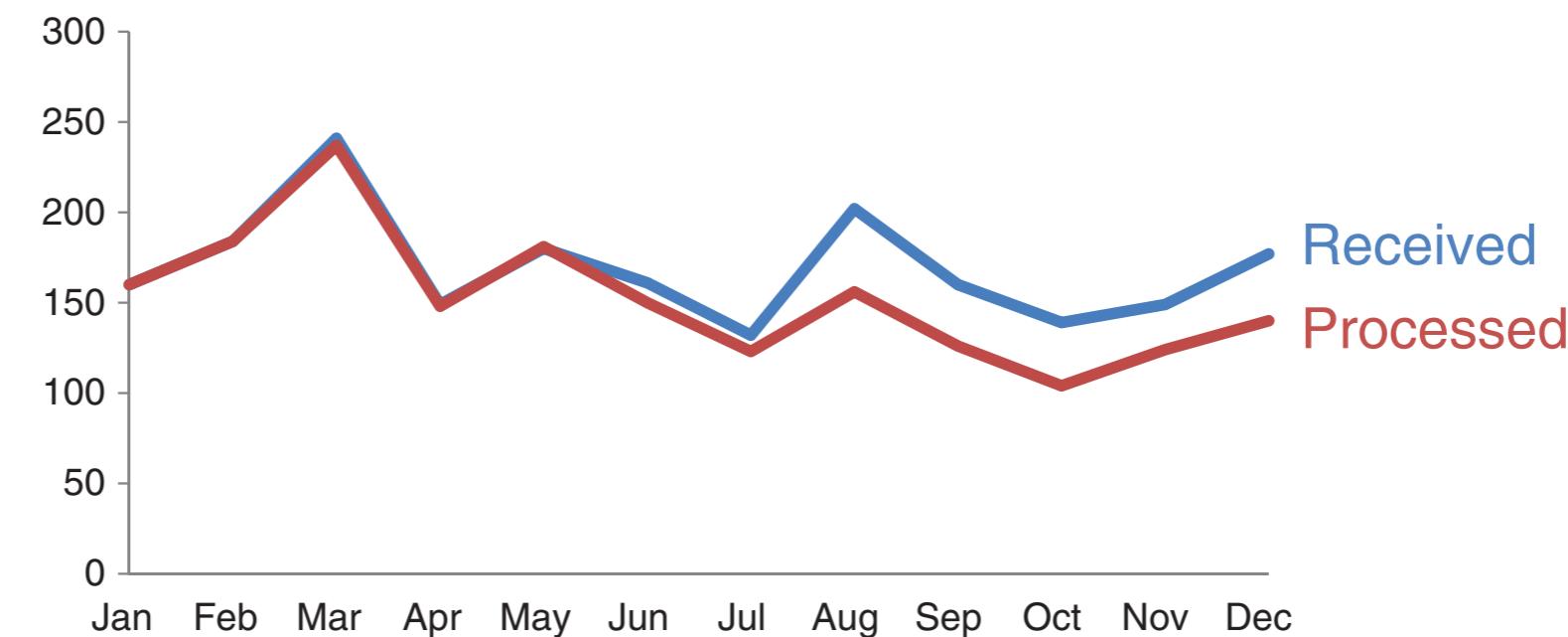
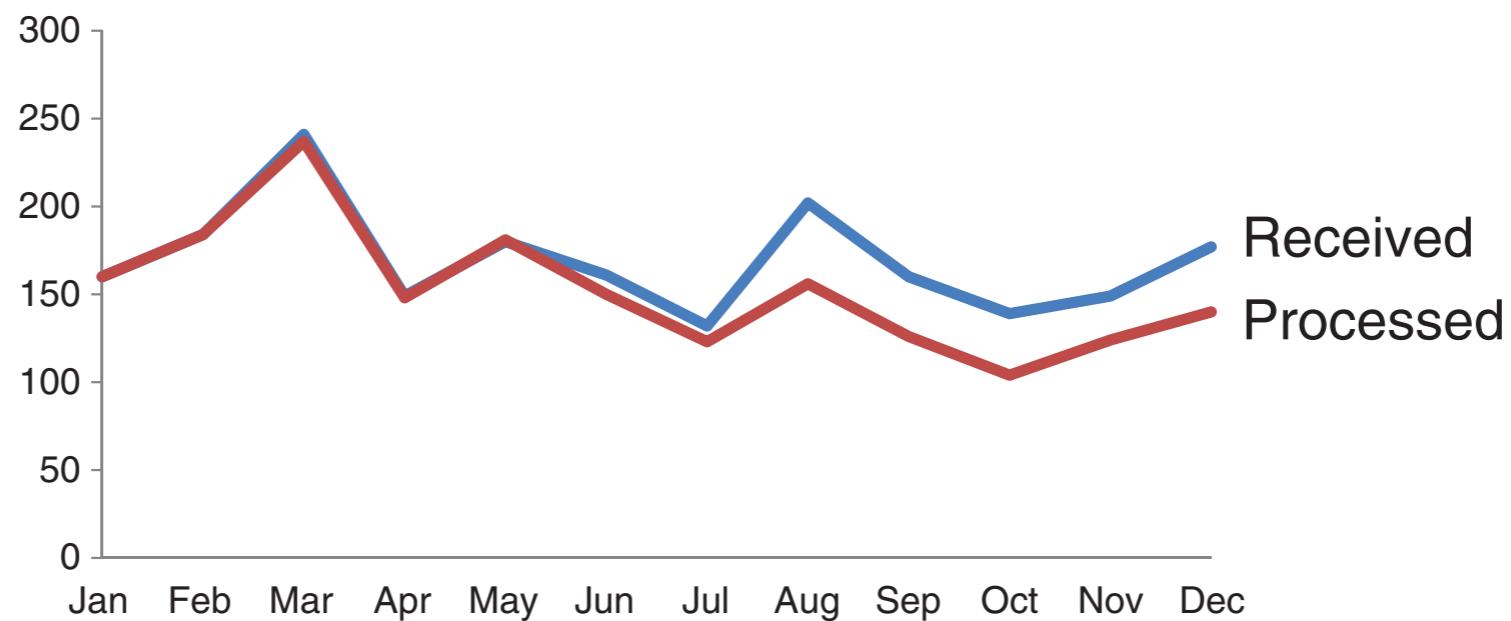
Similarity

- objects of similar color, shape, size, or orientation form a group



3. clutter is your enemy!

Gestalt principles of visual perception

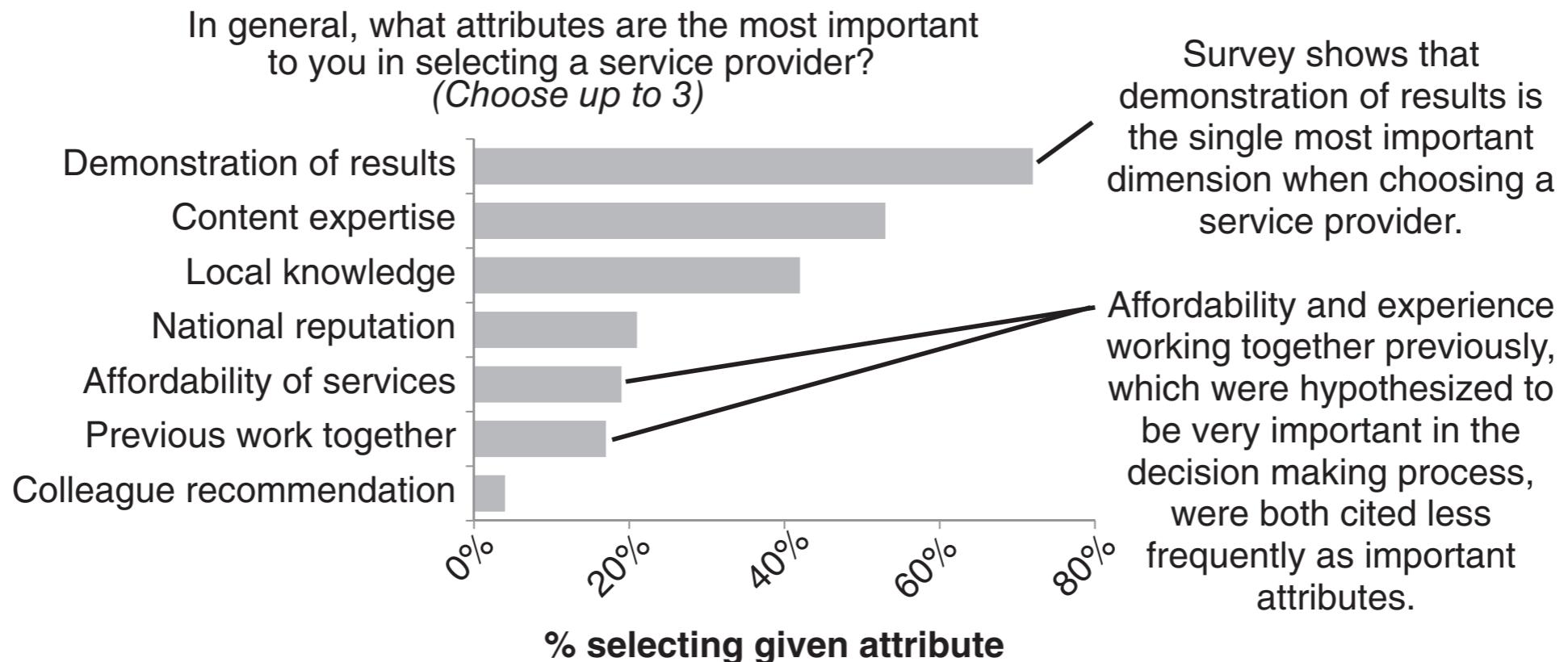


3. clutter is your enemy!

lack of visual order

- thoughtful design fades into the background and audience doesn't notice it
- if not, however, your audience **feels the burden** and lack of order
- any observations you have regarding the following figure?

Demonstrating effectiveness is most important consideration when selecting a provider



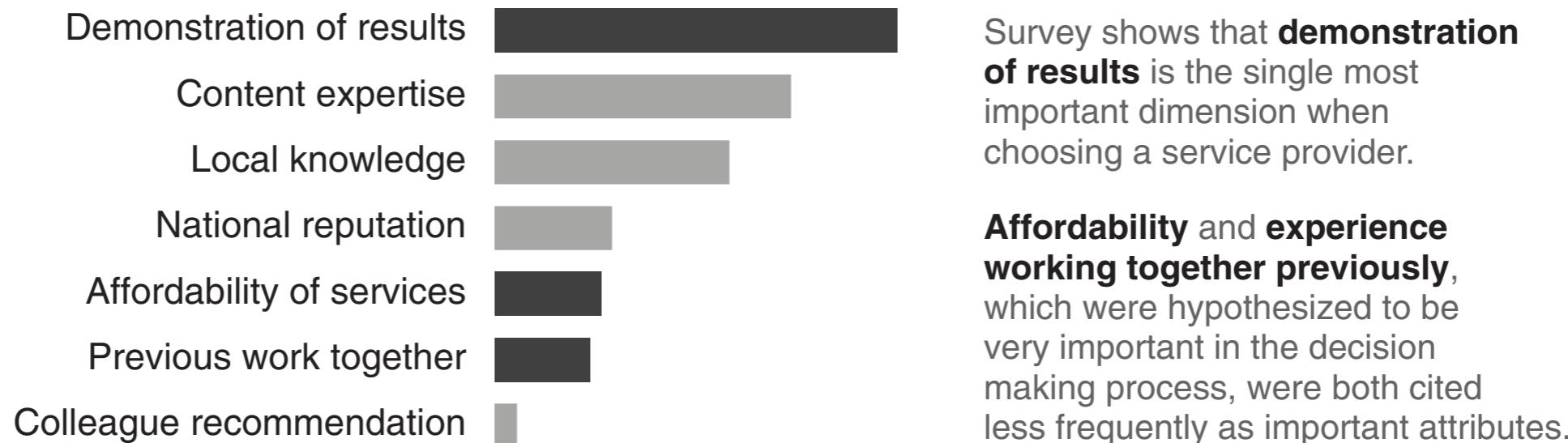
Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

3. clutter is your enemy!

Demonstrating effectiveness is most important consideration when selecting a provider

In general, **what attributes are the most important** to you in selecting a service provider?

(Choose up to 3)



Data source: xyz; includes N number of survey respondents.
Note that respondents were able to choose up to 3 options.

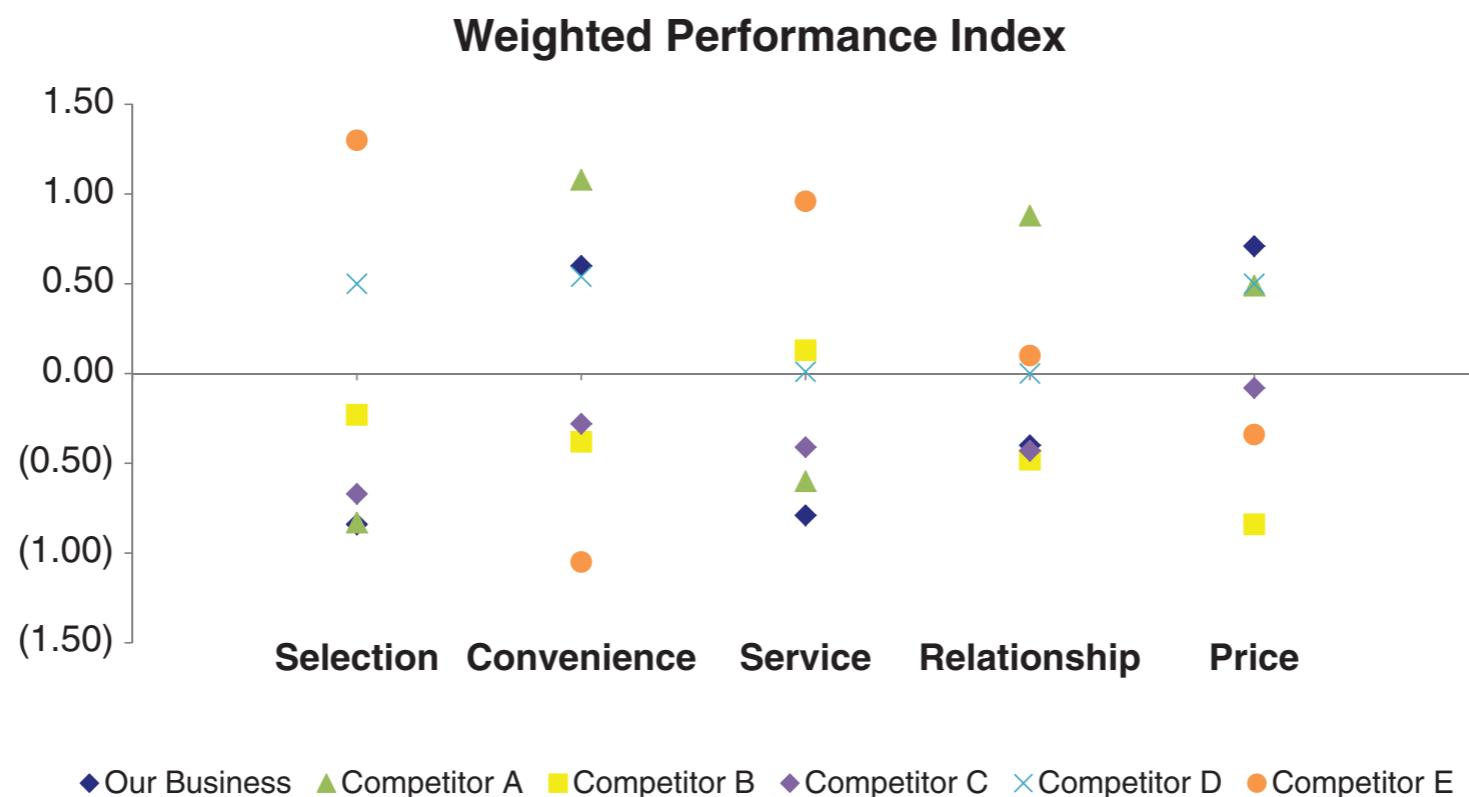
- feels somehow easier: there is **order by alignment**
- goal is to **create clean lines** (both horizontally and vertically)
- without other visual guiding your audience will move eyes in a 'z'-shape (x-label changed)
- avoid diagonal elements, they are harder to read (45° : 52% slower, 90° : 205% slower)

3. clutter is your enemy!

Non-strategic use of contrast/color

- lack of clear contrast can be a form of visual clutter
- the more things are different, the lesser any of them stands out
- "It's easy to spot a hawk in a sky full of pigeons, but as the variety of birds increases, that hawk becomes harder and harder to pick out." (Colin Ware (*Information Visualization: Perception for Design*, 2004))

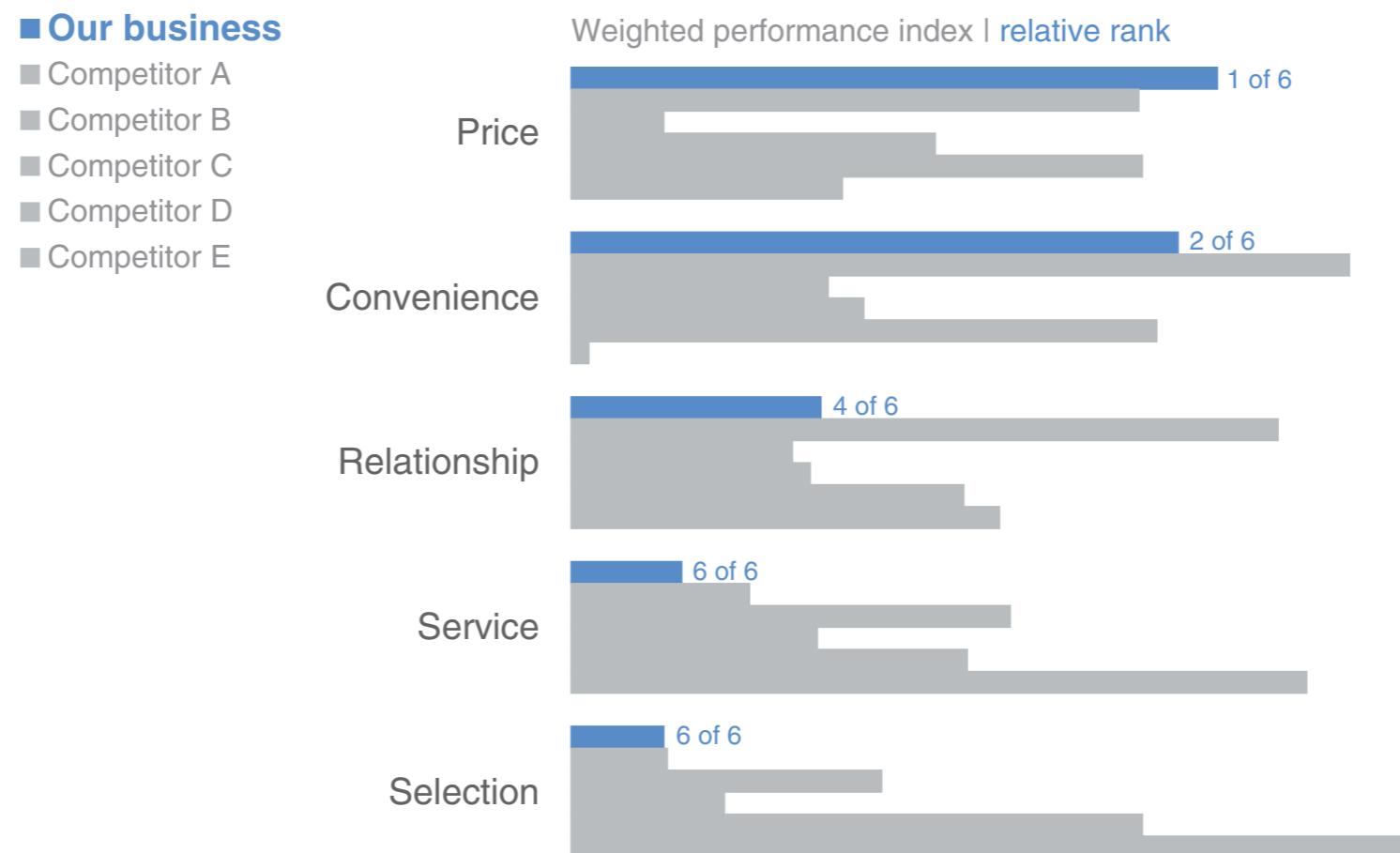
If you need to describe this plot is a single word?



3. clutter is your enemy!

Non-strategic use of contrast/color

Performance overview

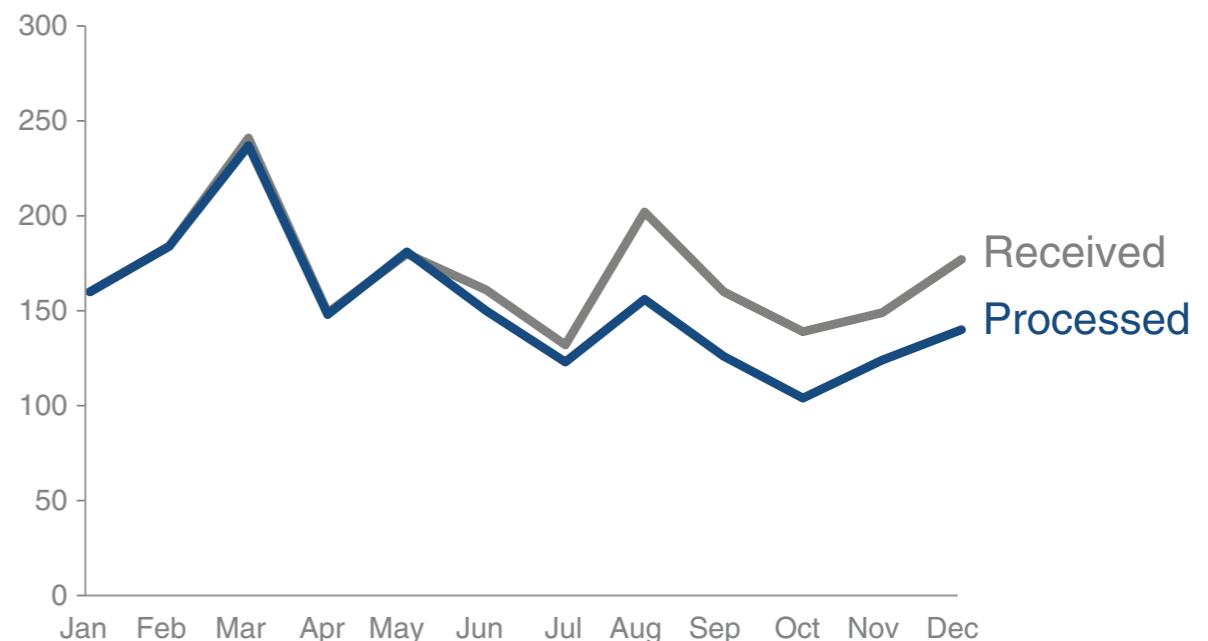
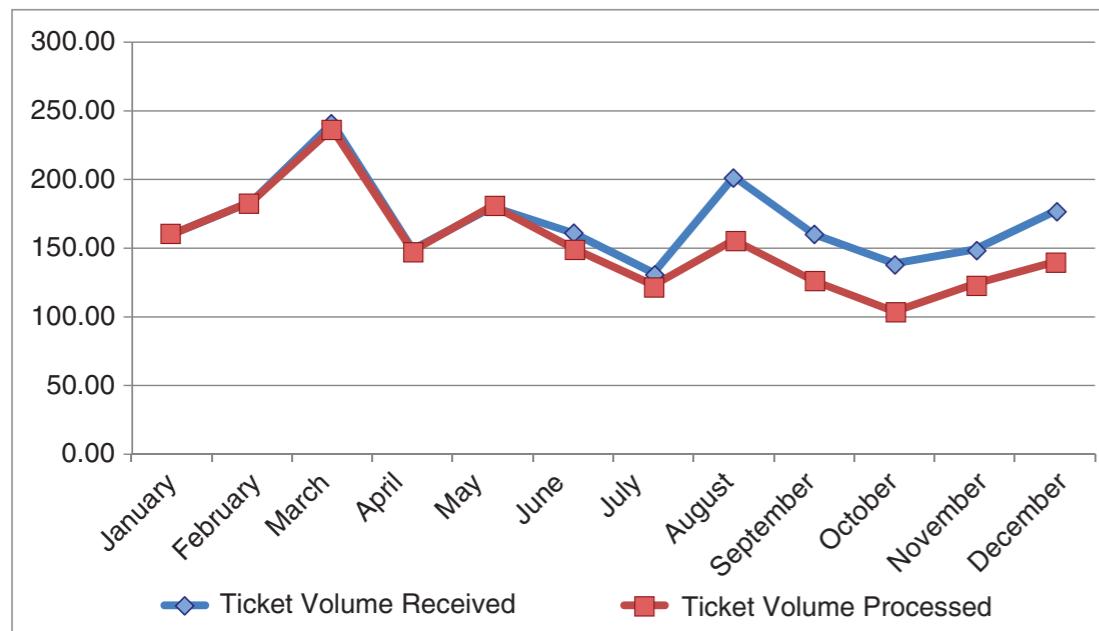


- rescaled all the numbers on a positive scale
- no x-axis scale: forces audience to focus differences
- organized the categories in decreasing weighted performance index for 'Our business'

3. clutter is your enemy!

Closing

- any information in front of your audience is creating **cognitive load** (use their brain power to process that)
- this can hinder the transmission of your message
- use **contrast** strategically and **ban clutter** from your visuals!



now we consider how we want our **audience to interact with our visual** communications.

4. focus your audience's attention

preattentive attributes

- they can help to direct your audience's attention
- best way to prove the power of preattentive attributes is to demonstrate it:
COUNT THE 3s:

756395068473

658663037576

860372658602

846589107830

4. focus your audience's attention

preattentive attributes

- they can help to direct your audience's attention
- best way to prove the power of preattentive attributes is to demonstrate it:

756**3**95068473

658663**0**37576

860**3**72658602

846589107830

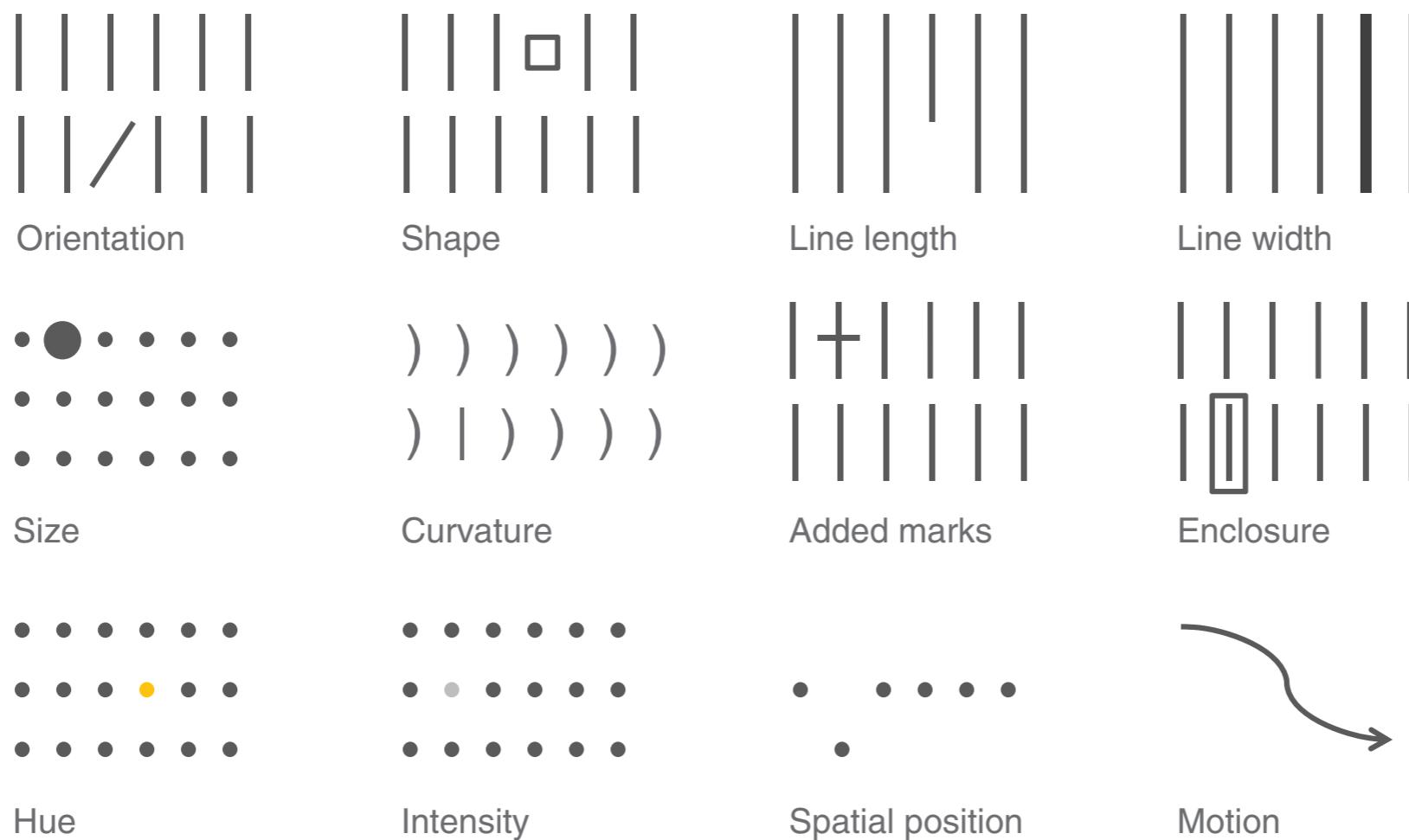
- brain is quick in picking the 3s up without dedicating any conscious thought to it

enable our audience to see what we want them to
before they even know they're seeing it!

4. focus your audience's attention

preattentive attributes

- are visual properties that we notice without using conscious effort
- 4 different preattentive properties: **form**, **color**, **spatial positioning**, **movement**



Source: Adapted from Stephen Few's *Show Me the Numbers*, 2004.

- our brains quickly pick up **differences we see in our environment**

4. focus your audience's attention

preattentive attributes in text

- we have about 3–8 seconds with our audience where they decide whether to continue to look at what we've put in front of them or direct their attention to something else

No preattentive attributes

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Color

What are we doing well? Great Products. **These products are clearly the best in their class.** Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Size

What are we doing well? Great Products. These products are the best in their class. Replacement parts are shipped when needed. You sent gaskets

without me having to ask.

Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

Bold

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Italics

What are we doing well? Great Products. These products are clearly the best in their class. *Replacement parts are shipped when needed.* You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours.
You have a great company – keep up the good work!

Separate spatially

What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask.

Problems are resolved promptly.

Bev in the billing office was quick to resolve a billing issue I had. General customer service exceeds expectations. The account manager even called to check in after normal business hours. You have a great company – keep up the good work!

What are we doing well?

Themes & example comments

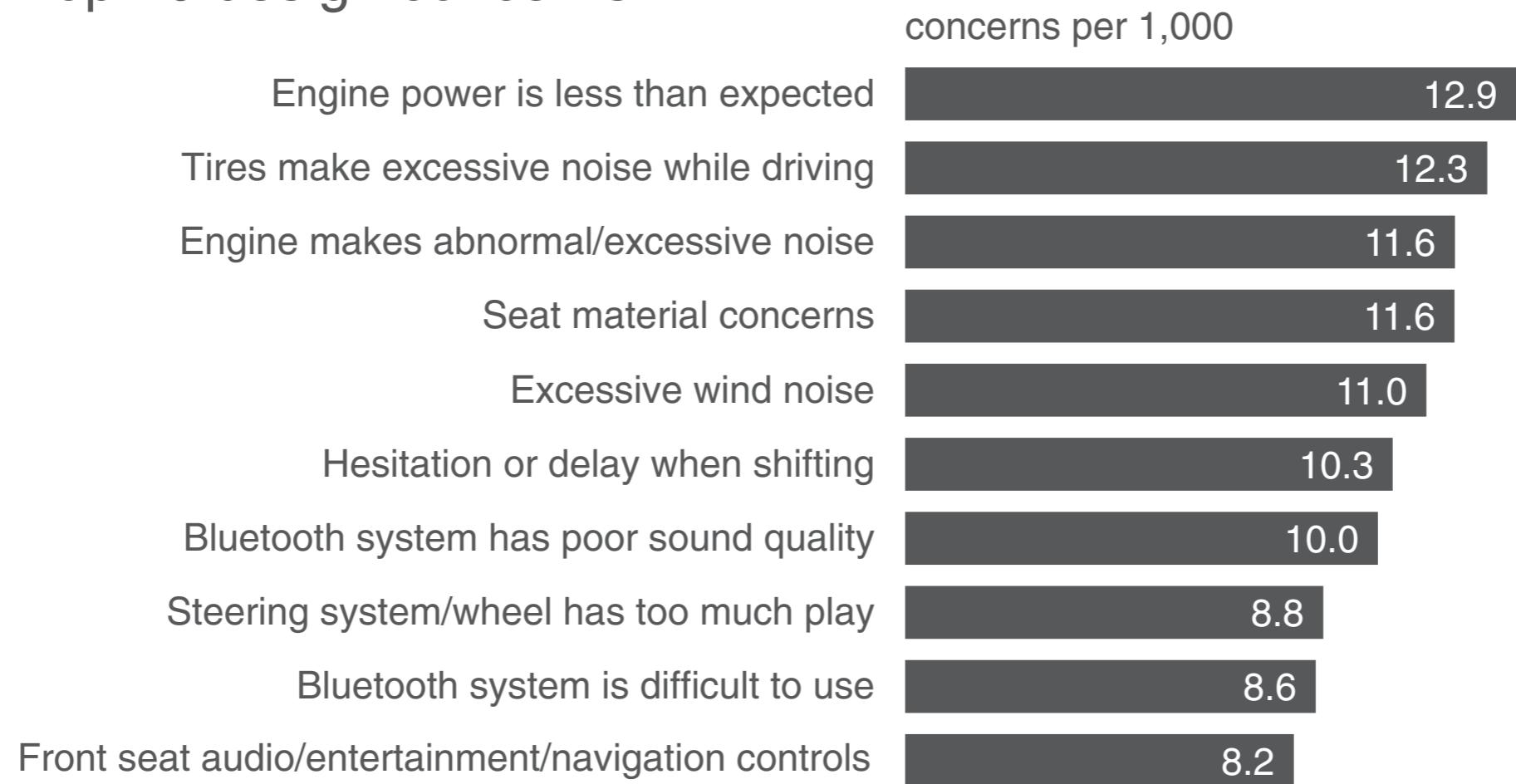
- **Great products:** "These products are clearly the best in class."
- **Replacement parts are shipped when needed:** "You sent me gaskets without me having to ask, and I really needed them, too!"
- **Problems are resolved promptly:** "Bev in the billing office was quick to resolve a billing issue I had."
- **General customer service exceeds expectations:** "The account manager even called after normal business hours.
You have a great company - keep up the good work!"

4. focus your audience's attention

preattentive attributes in graphs

- without other visual cues, you are left to process all of the information (as count 3s)

Top 10 design concerns



4. focus your audience's attention

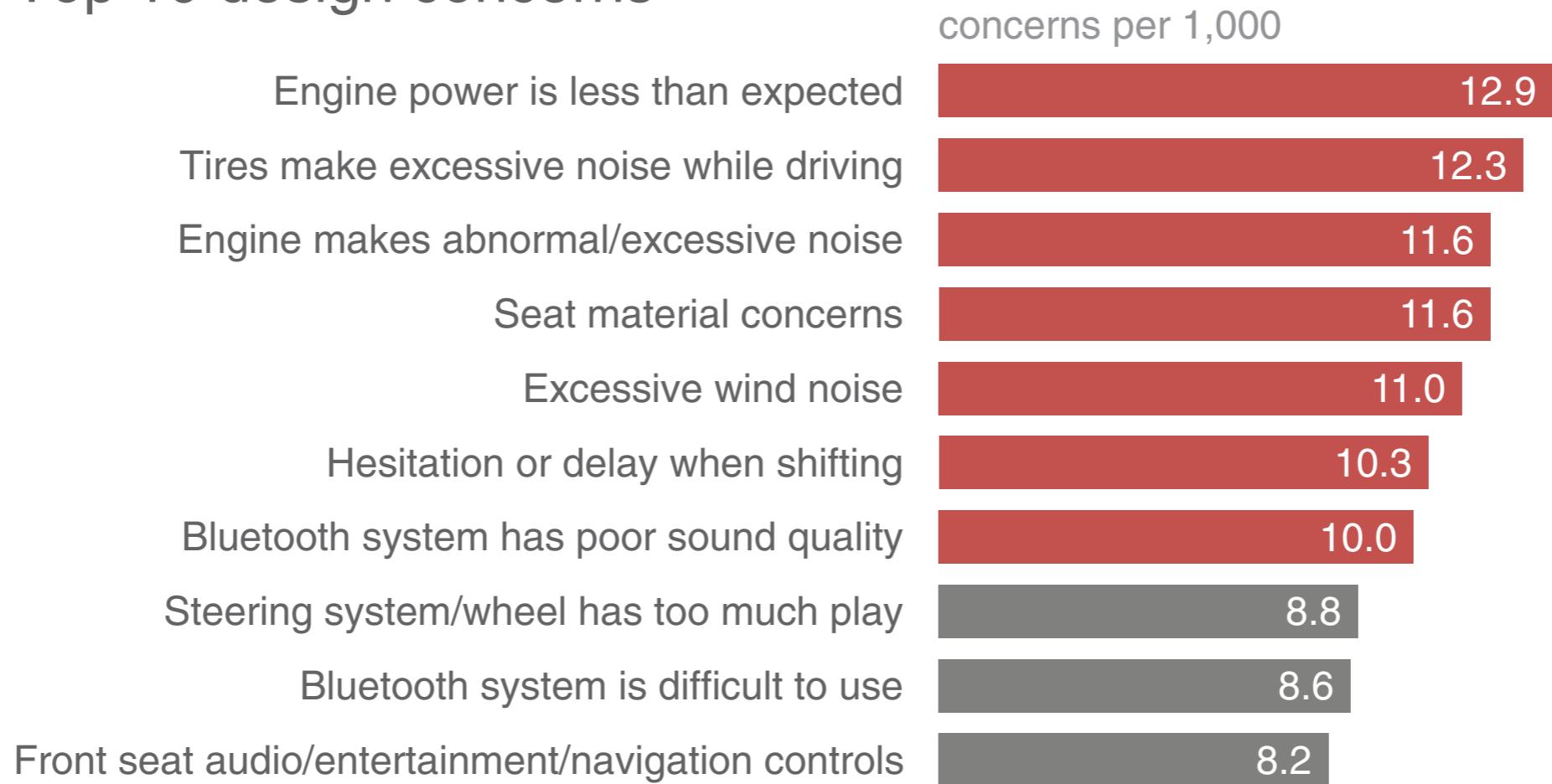
preattentive attributes in graphs

- explanatory analysis: focus on story and use of color and text thoughtfully

7 of the top 10 design concerns have 10 or more concerns per 1,000.

Discussion: is this an acceptable default rate?

Top 10 design concerns

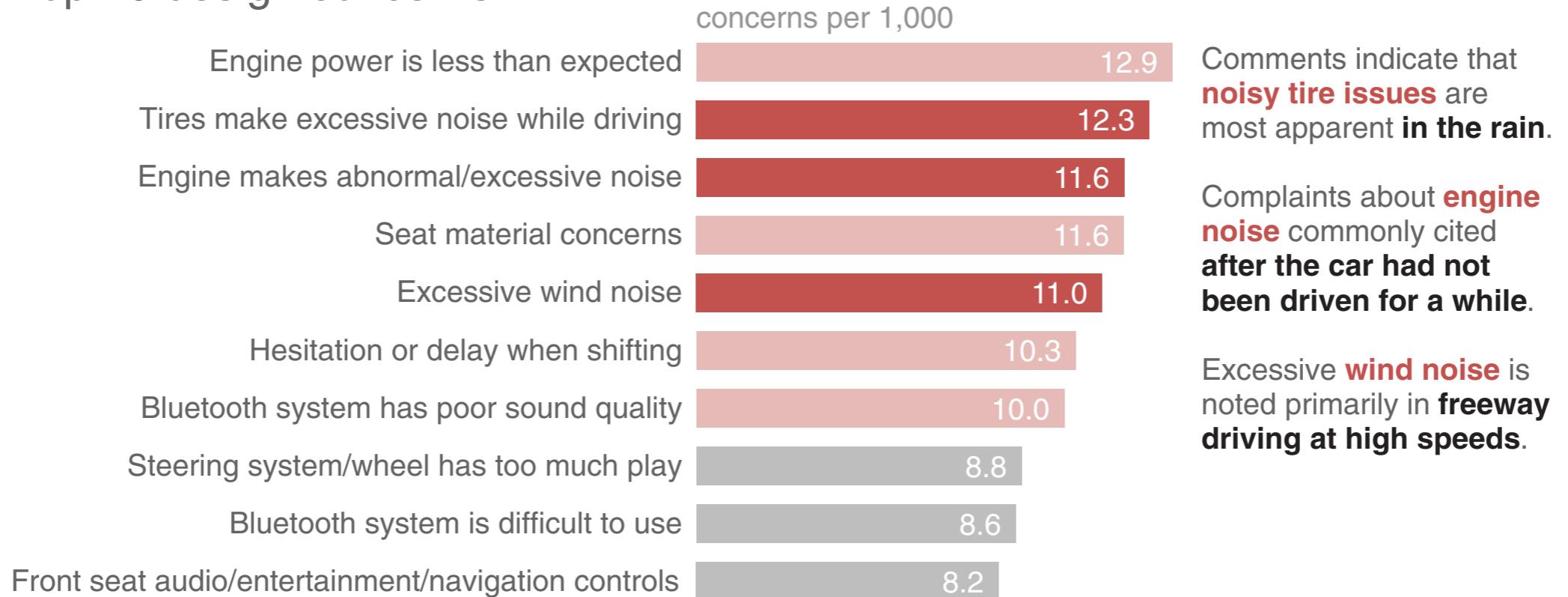


4. focus your audience's attention

preattentive attributes in graphs

Of the top design concerns, three are noise-related.

Top 10 design concerns

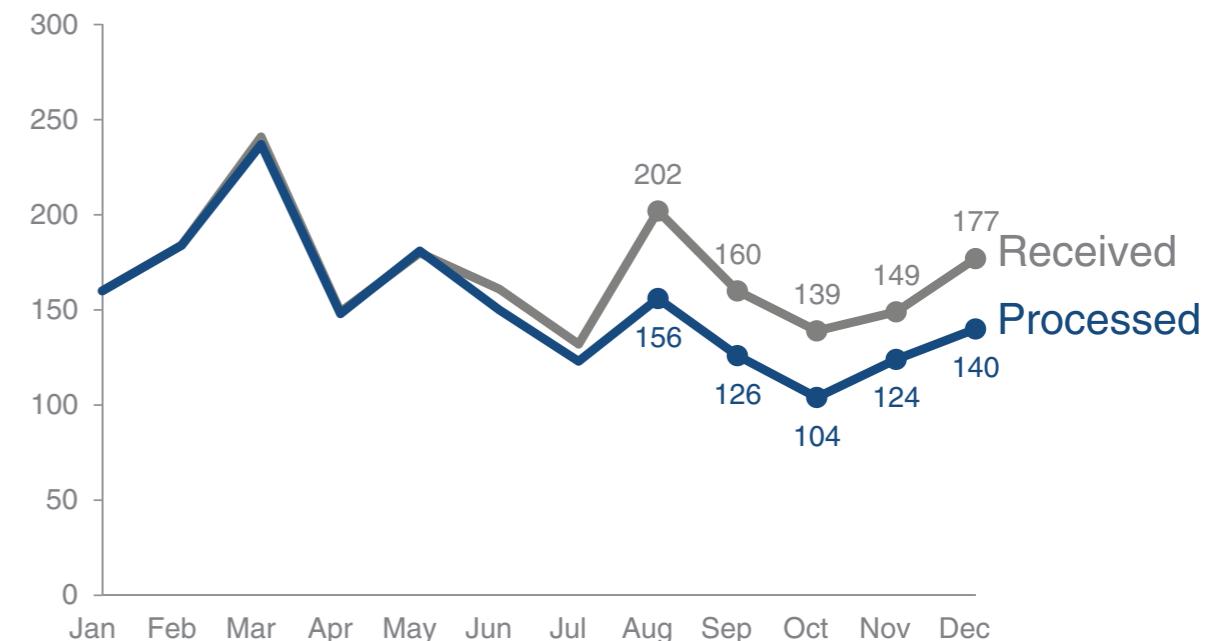
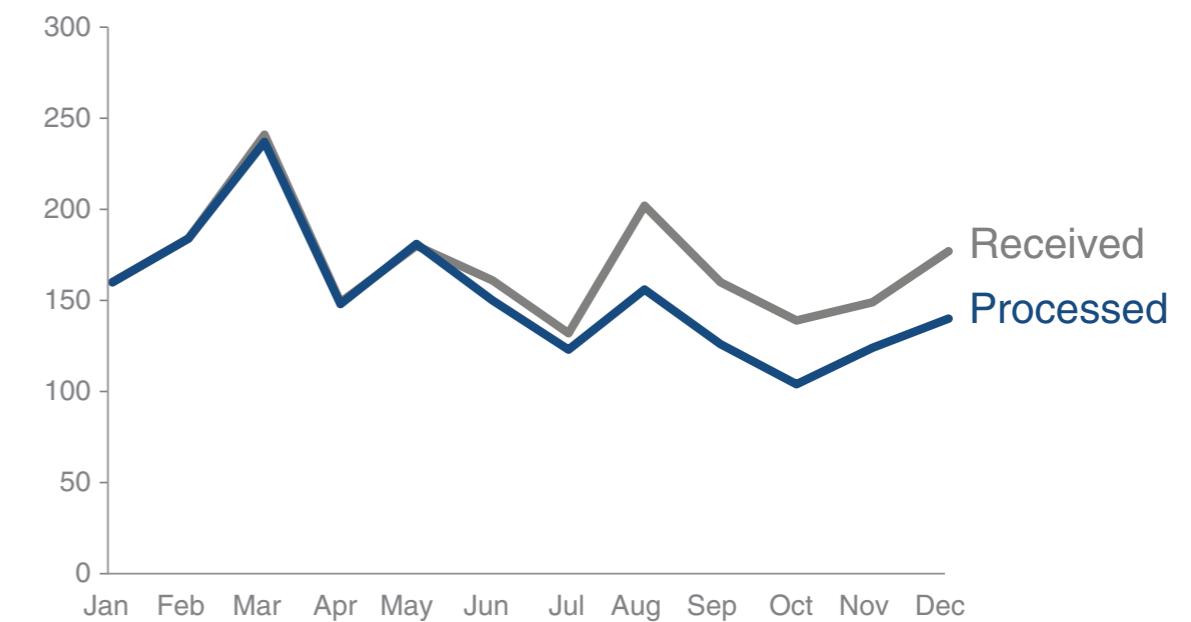
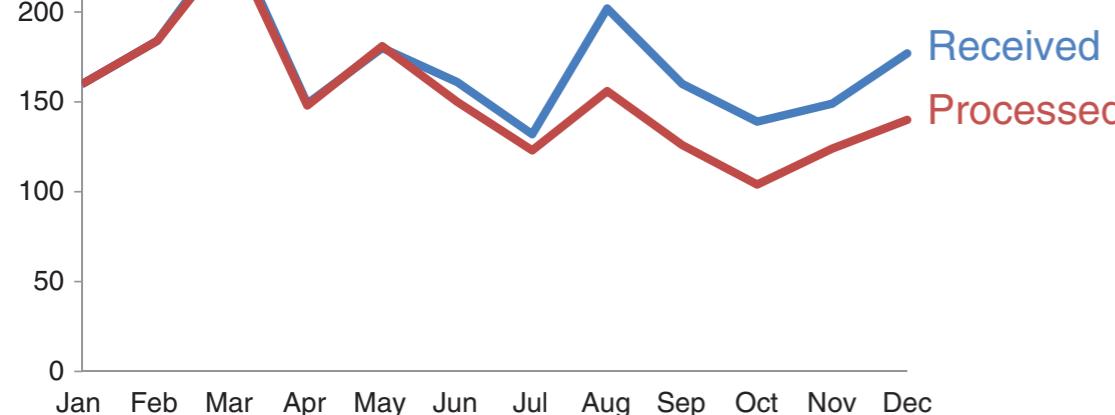


- when **used sparingly**, color most powerful tools for drawing your audience's attention
- too many colors together: we lose their preattentive value
- design with the colorblind in mind: avoid using shades of red and shades of green together (8% of men and 0.5% of women are colorblind)

4. focus your audience's attention

In closing

- preattentive attributes are powerful tools when used sparingly and strategically
- use them to draw **attention** to where you want your audience to look and **create visual hierarchy** to guide your audience



short: (more in book)

5. think like a designer

- form follows functions
- Accessibility (don't overcomplicate and text is your friend)
- Aesthetics (alignment, good use of color)

Peak Break-up Times
According to Facebook status updates

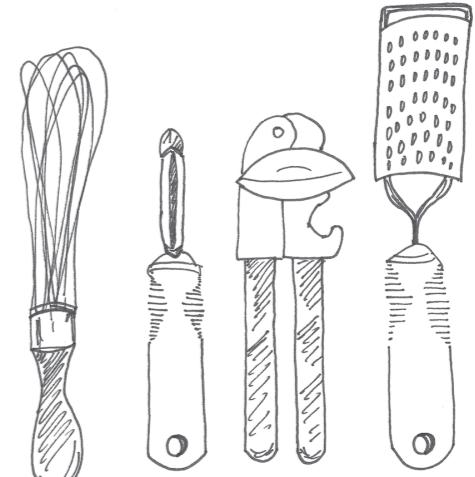
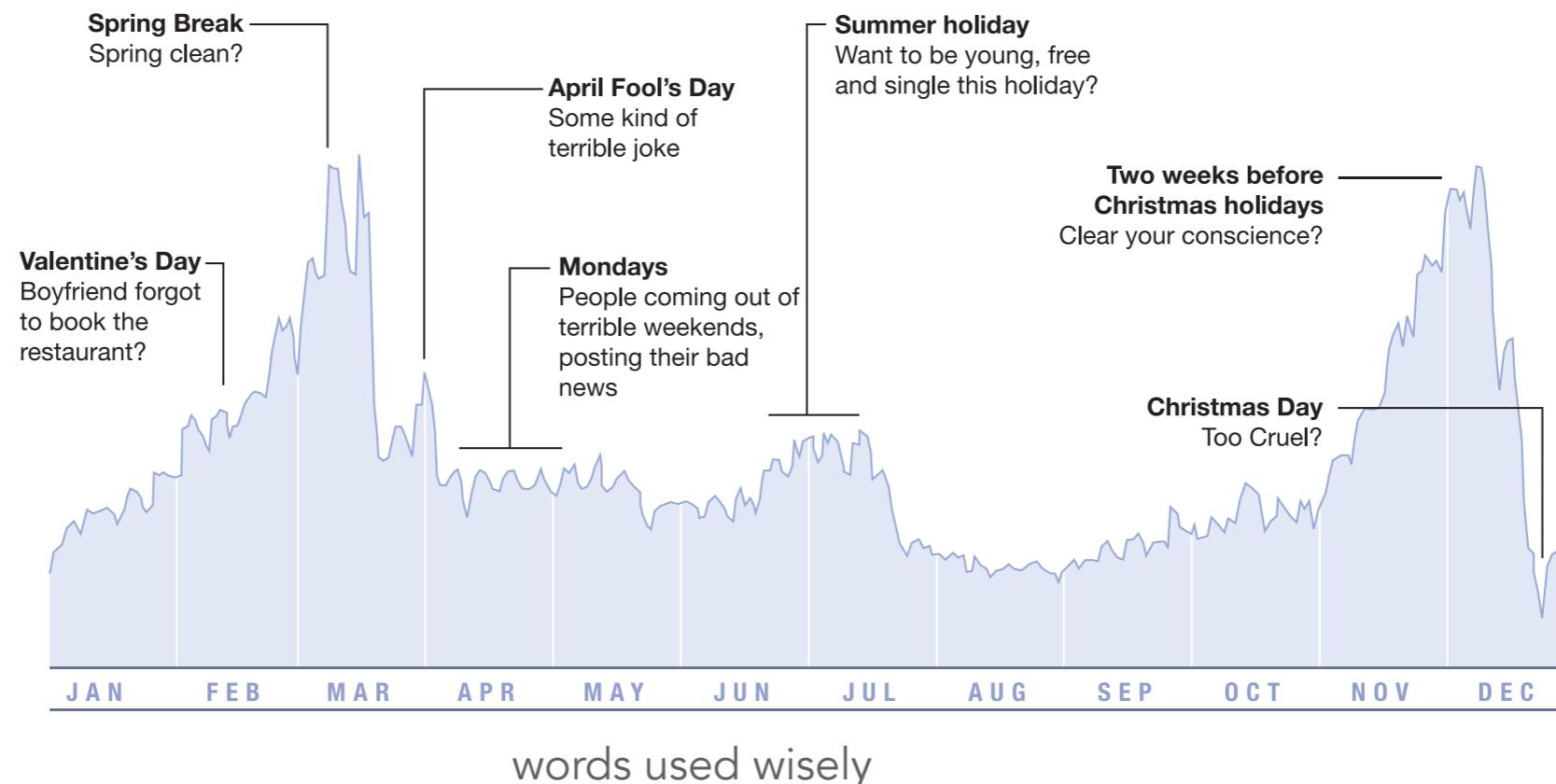


FIGURE 5.1 OXO kitchen gadgets

- few words and phrases make this data more quickly accessible than it otherwise would be

5. think like a designer

- form follows functions
- Accessibility (don't overcomplicate and text is your friend)
- Aesthetics (alignment, good use of color)

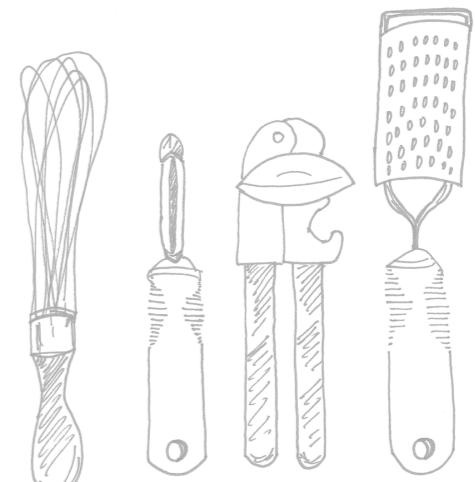
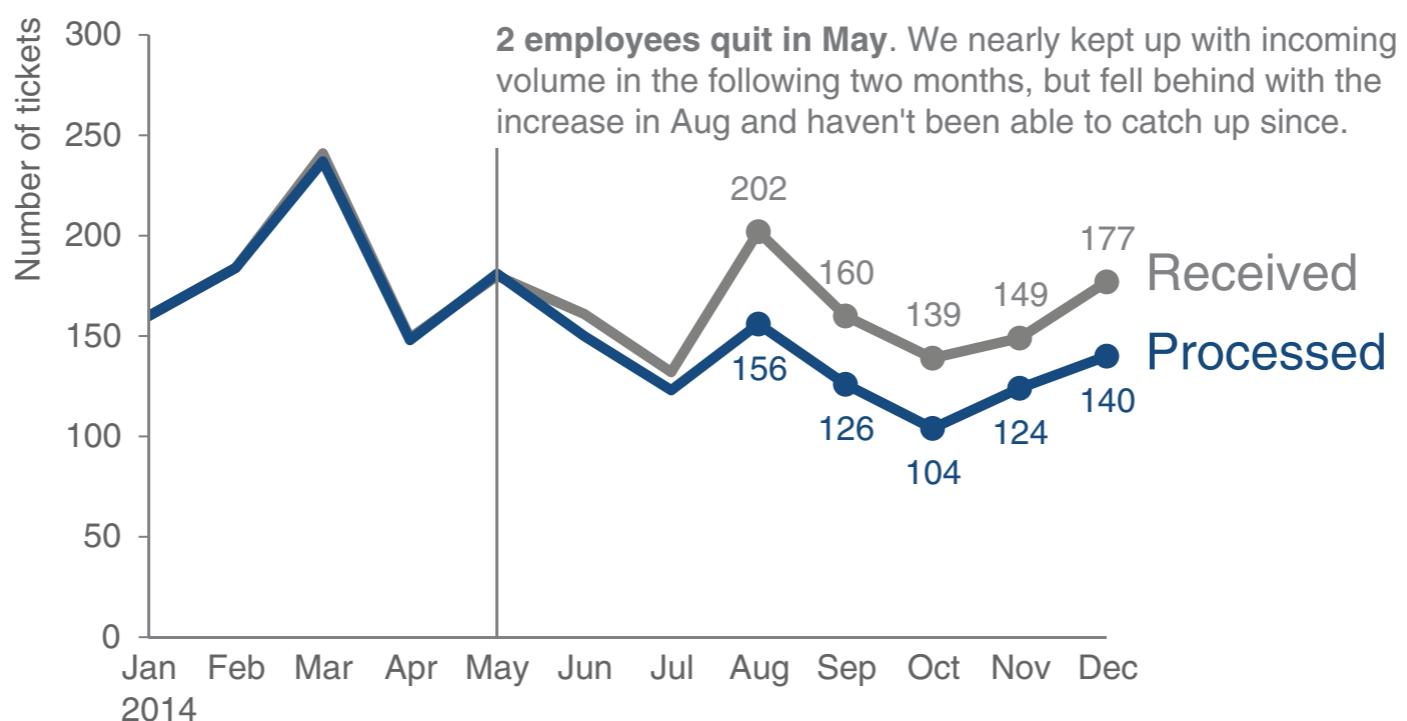


FIGURE 5.1 OXO kitchen gadgets

Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time

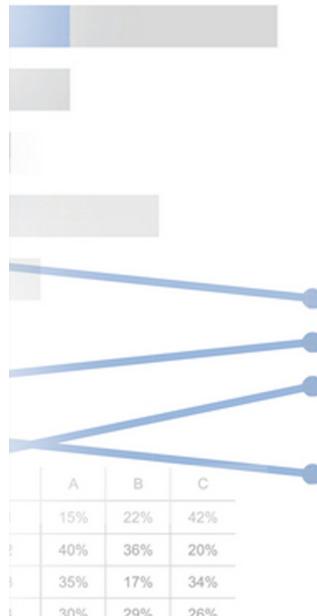


Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

- few words and phrases make this data more quickly accessible than it otherwise would be

6. lessons in storytelling

The magic of story



cole nussbaumer knaflic

storytelling with data



a data
visualization
guide for
business
professionals

	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%

pulling it all together

1. Understand the context
2. Choose an appropriate display
3. Eliminate clutter
4. Draw attention where you want it
5. Think like a designer
6. Tell a story

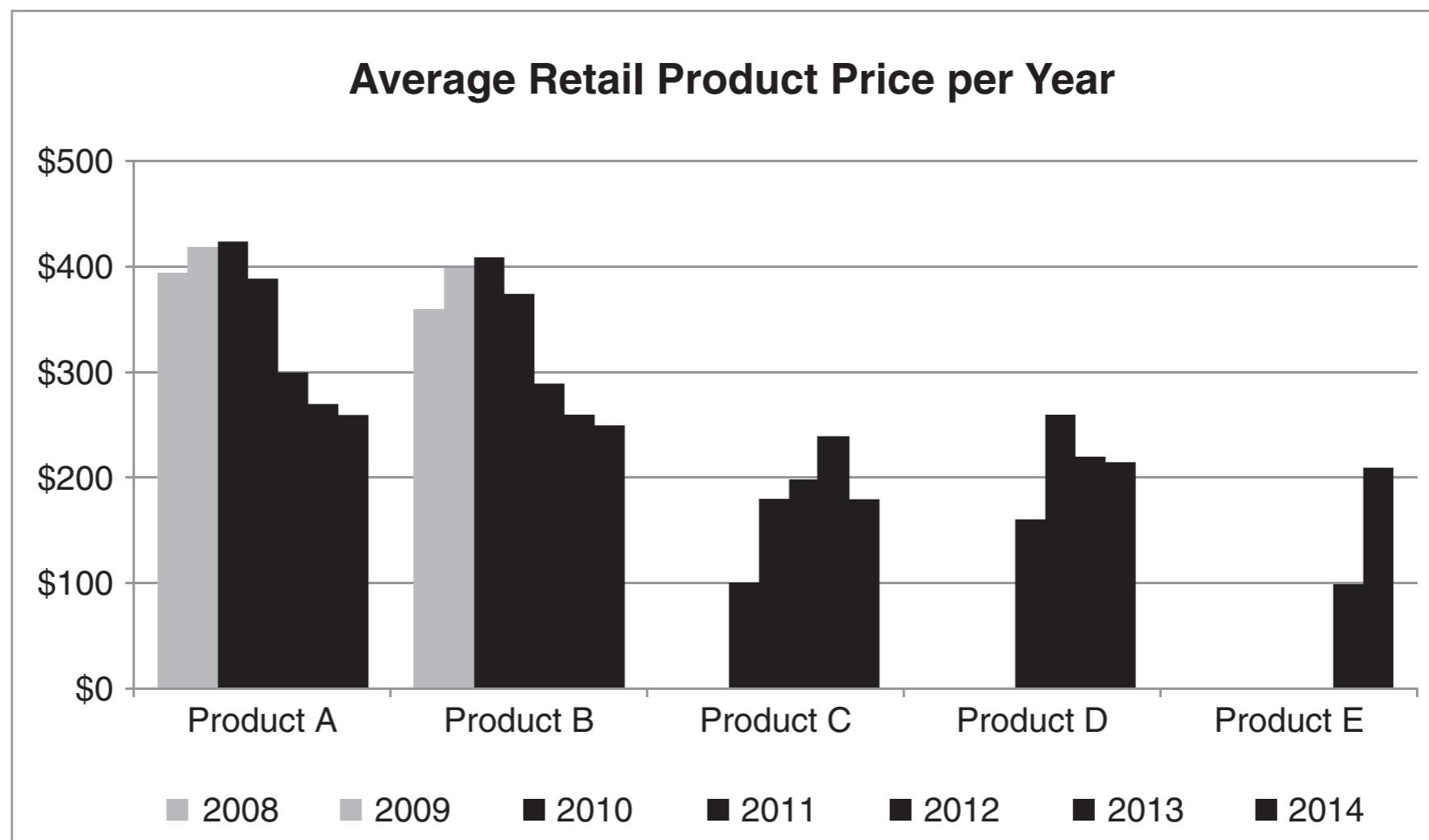
Original visual



pulling it all together

1. **Understand the context**
 2. Choose an appropriate display
 3. Eliminate clutter
 4. Draw attention where you want it
 5. Think like a designer
 6. Tell a story

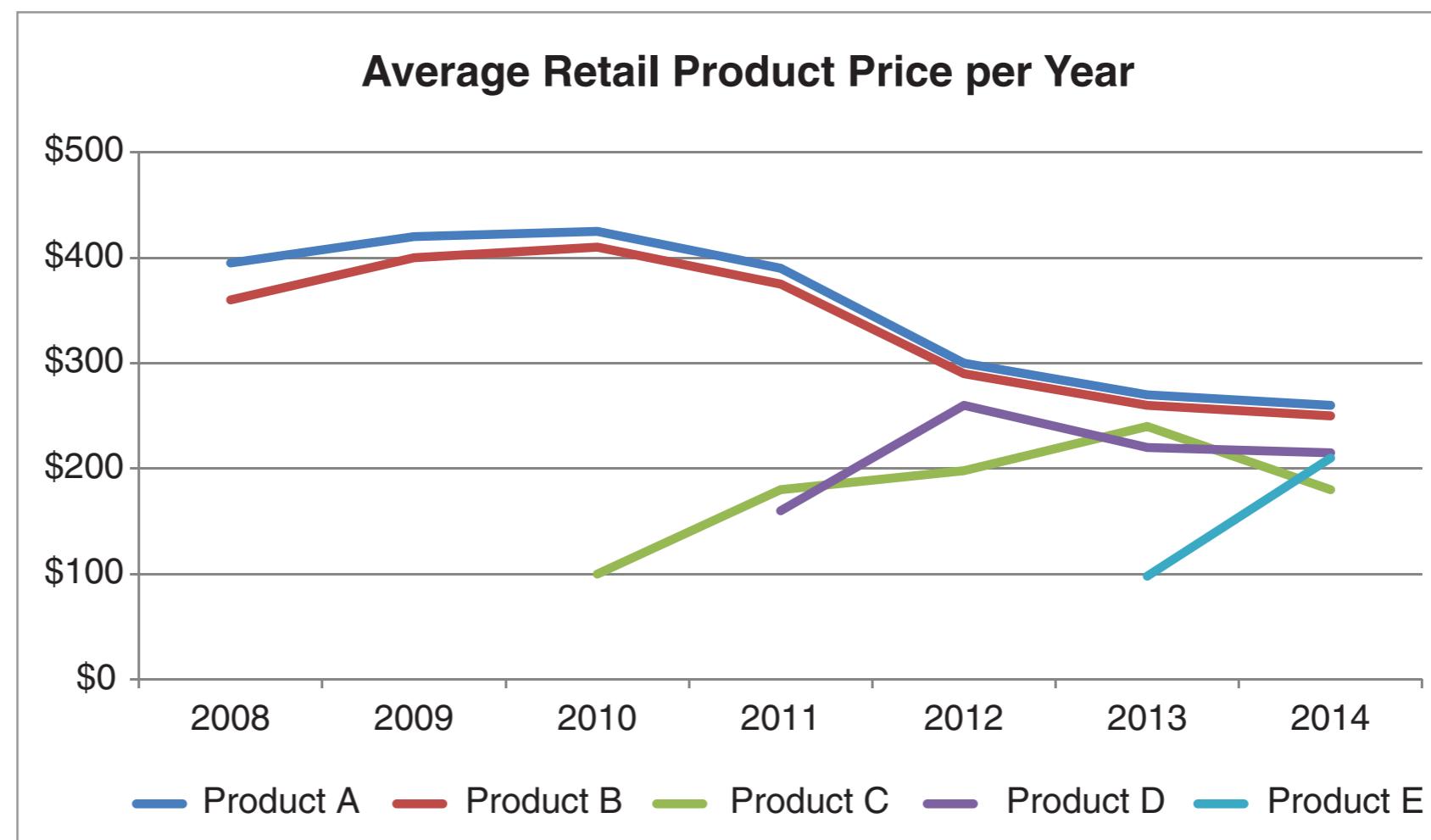
Emphasize 2010 forward



pulling it all together

1. Understand the context
2. **Choose an appropriate display**
3. Eliminate clutter
4. Draw attention where you want it
5. Think like a designer
6. Tell a story

Line graph for all products



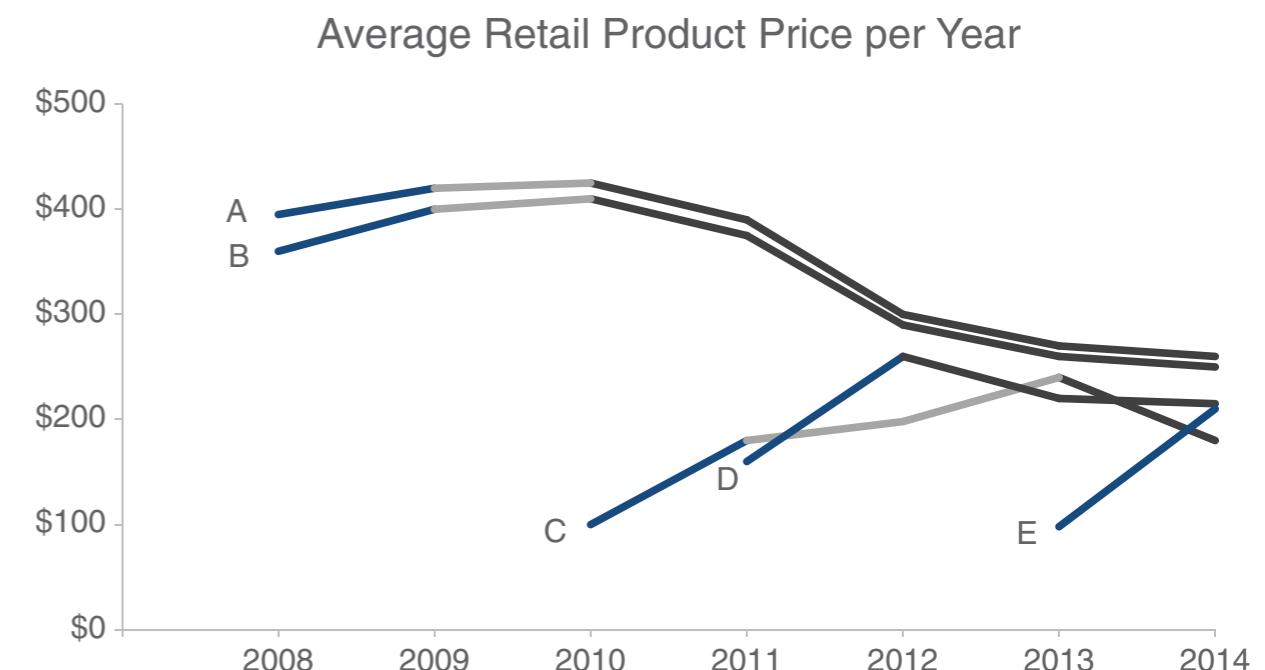
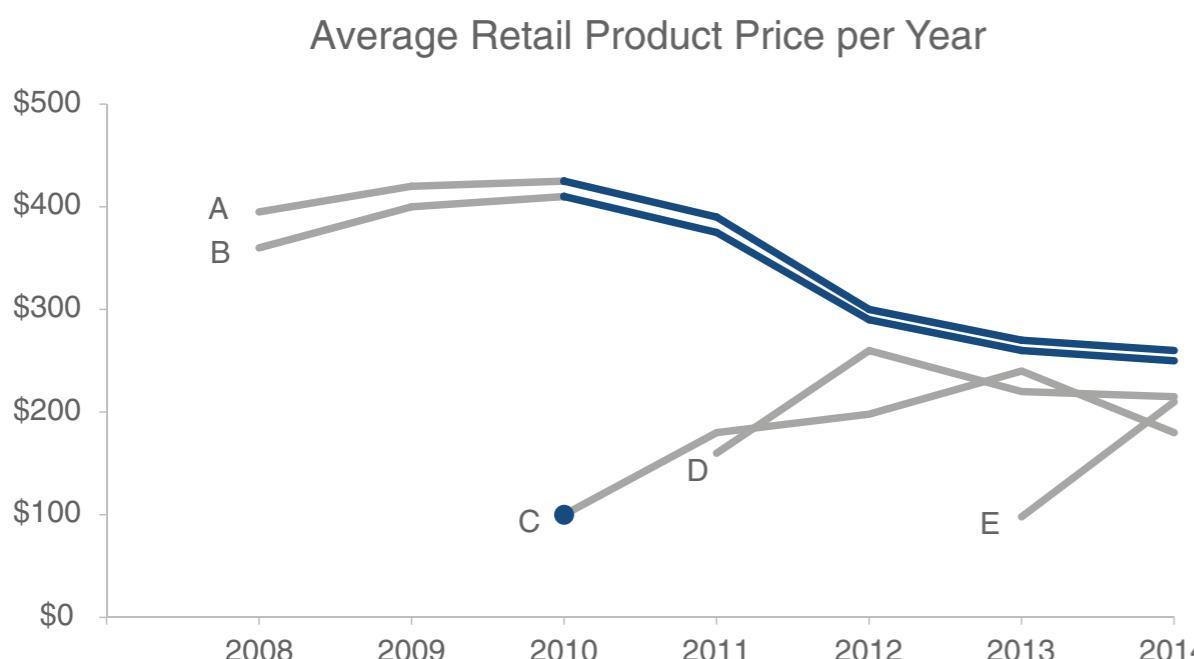
pulling it all together

1. Understand the context
2. Choose an appropriate display
3. **Eliminate clutter**
4. Draw attention where you want it
5. Think like a designer
6. Tell a story



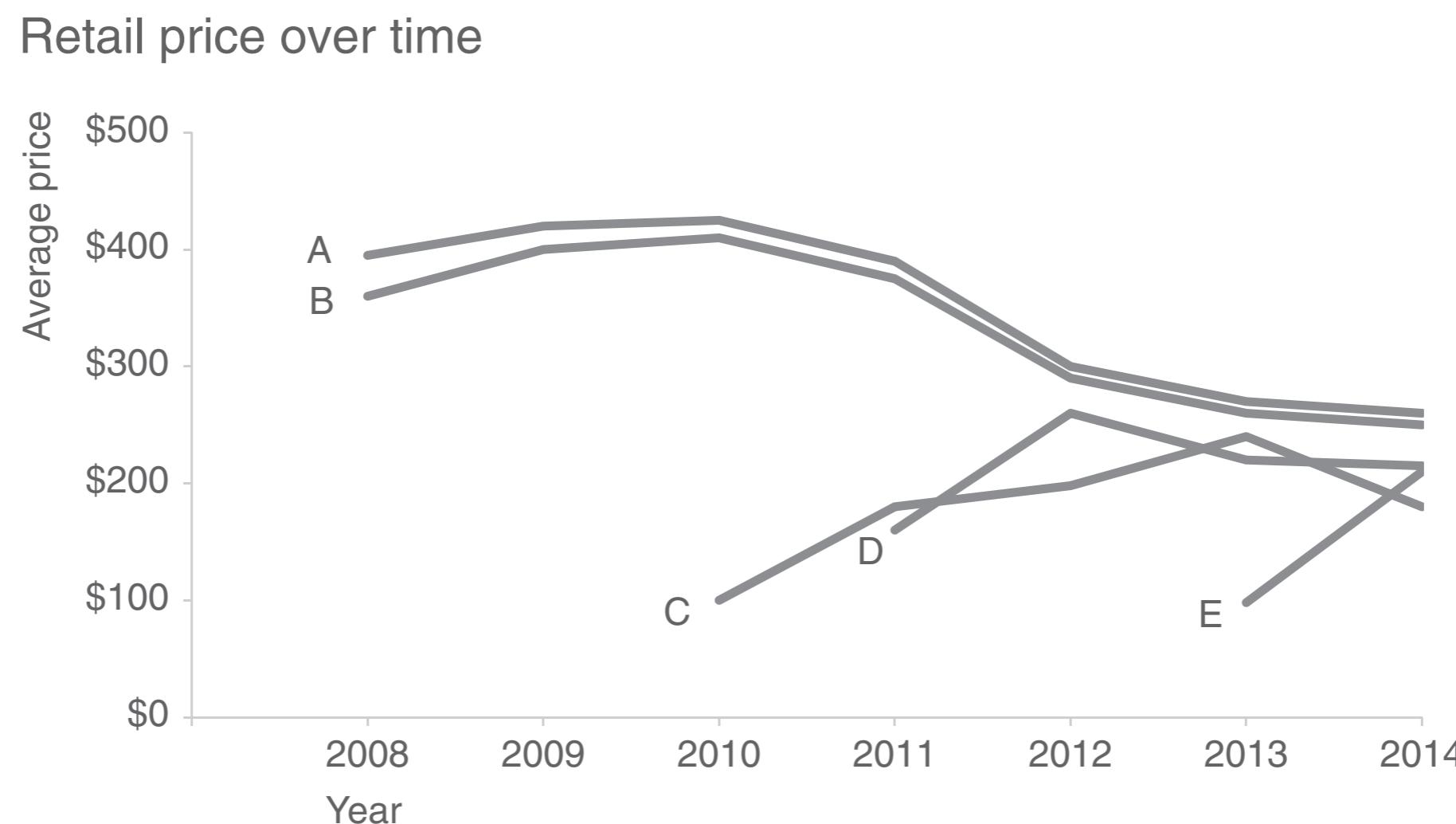
pulling it all together

1. Understand the context
2. Choose an appropriate display
3. Eliminate clutter
4. **Draw attention where you want it**
5. Think like a designer
6. Tell a story



pulling it all together

1. Understand the context
2. Choose an appropriate display
3. Eliminate clutter
4. Draw attention where you want it
5. **Think like a designer**
6. Tell a story



pulling it all together

1. Understand the context
2. Choose an appropriate display
3. Eliminate clutter
4. Draw attention where you want it
5. Think like a designer
6. **Tell a story**

In the next 5 minutes...

OUR GOAL:

- 1** Understand **how prices have changed over time** in the competitive landscape.
- 2** Use this knowledge to **inform the pricing of our product**.

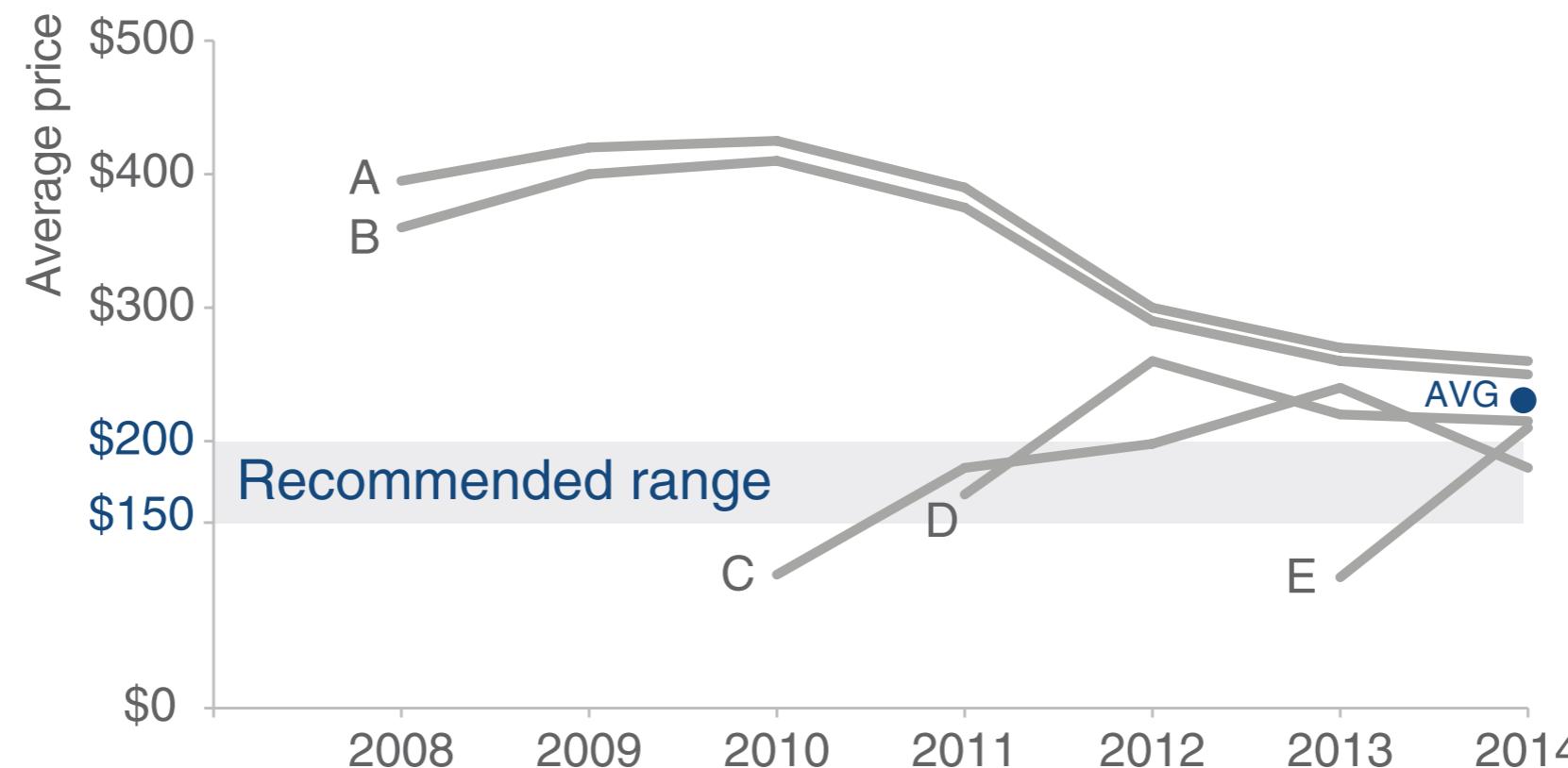
We will end with a **specific recommendation**.

pulling it all together

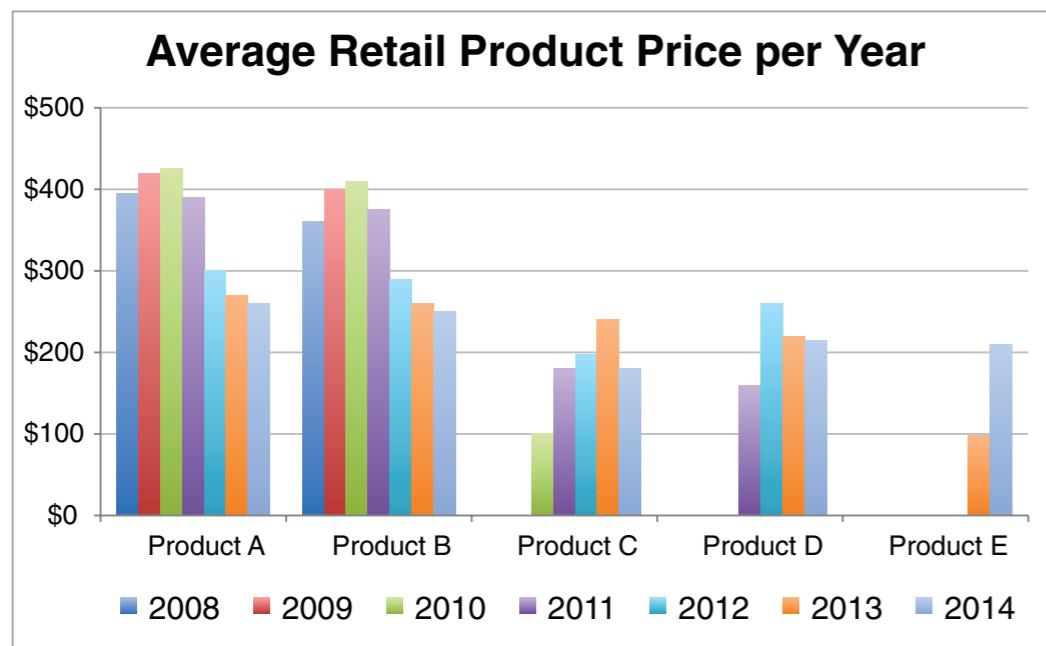
1. Understand the context
2. Choose an appropriate display
3. Eliminate clutter
4. Draw attention where you want it
5. Think like a designer
6. **Tell a story**

To be competitive, we recommend introducing our product *below* the \$223 average price point in the **\$150–\$200 range**

Retail price over time

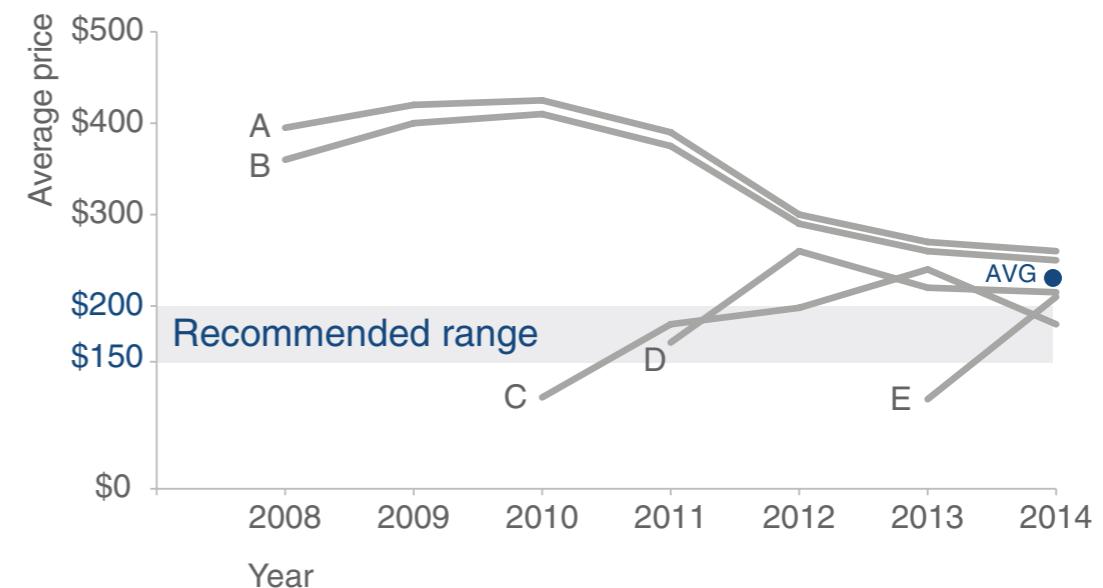


Price has declined for all products on the market since the launch of Product C in 2010



To be competitive, we recommend introducing our product *below* the \$223 average price point in the **\$150–\$200 range**

Retail price over time



I hope this helps us to move from simply showing data to **storytelling with data**

“I would have written a shorter letter, but I did not have the time”

(Mathematician and philosopher Blaise Pascal)

pulling it all together

Annual giving campaign progress

