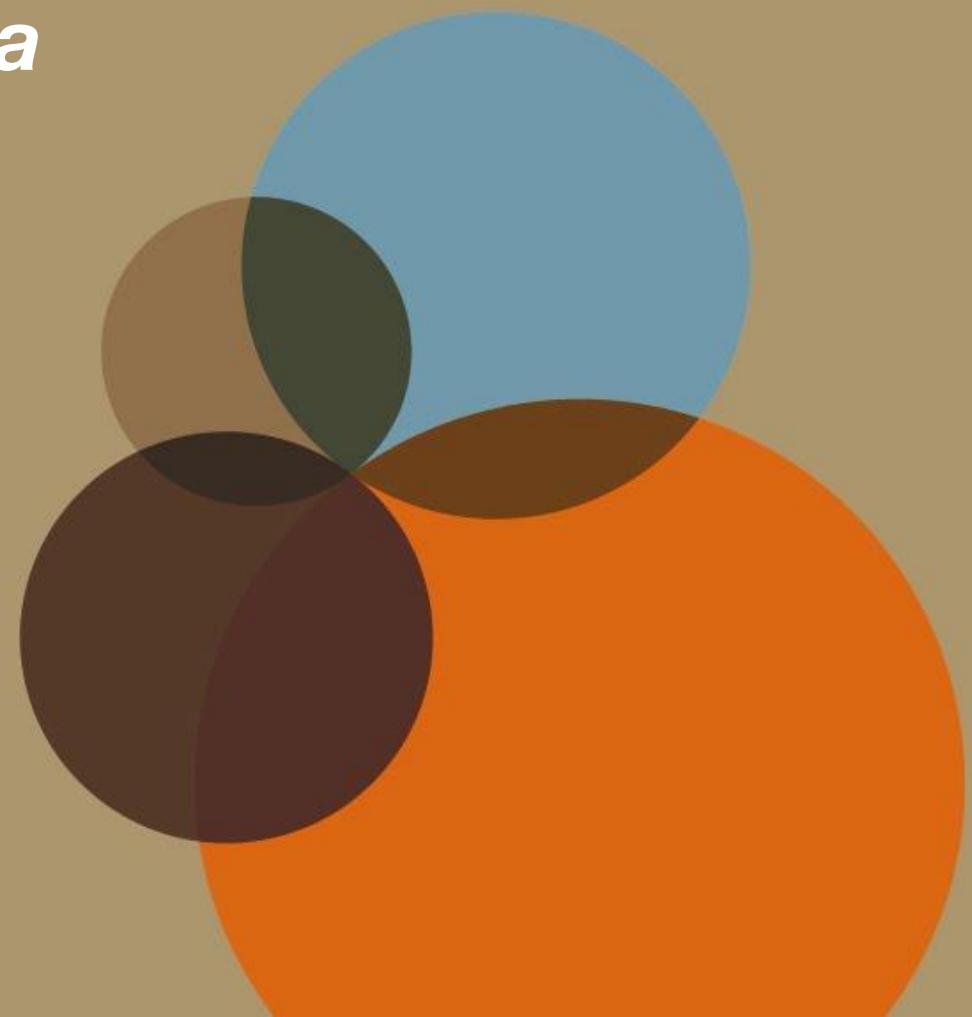


Data Management for Big Data

Storytelling with data

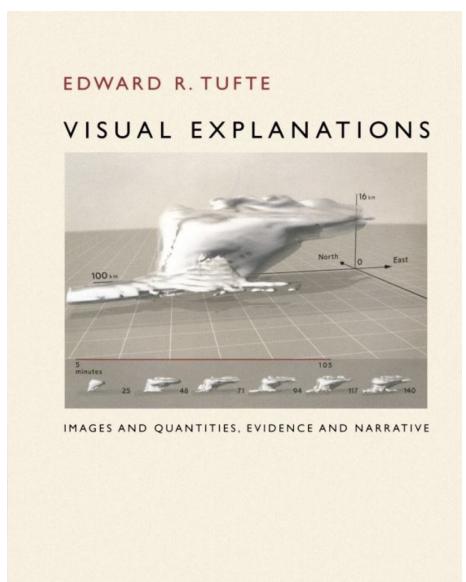
Andrea Brunello

andrea.brunello@uniud.it



“Power Corrupts. PowerPoint Corrupts Absolutely.”

— *Edward Tufte, Yale Professor Emeritus*



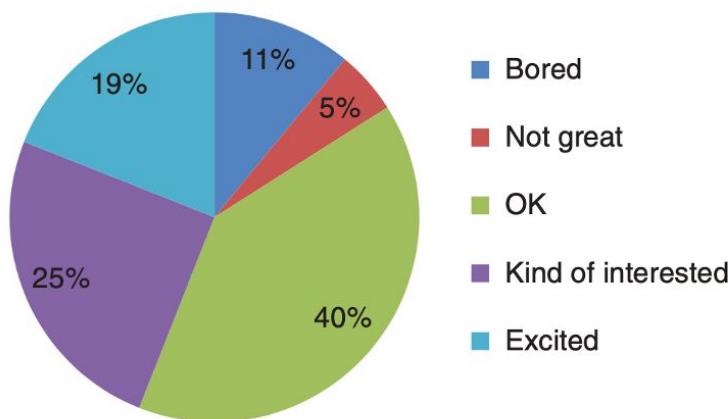
Material mainly taken from:

Storytelling with Data: A Data Visualization Guide for Business Professionals
Cole Nussbaumer Knaflic, Wiley, 2015

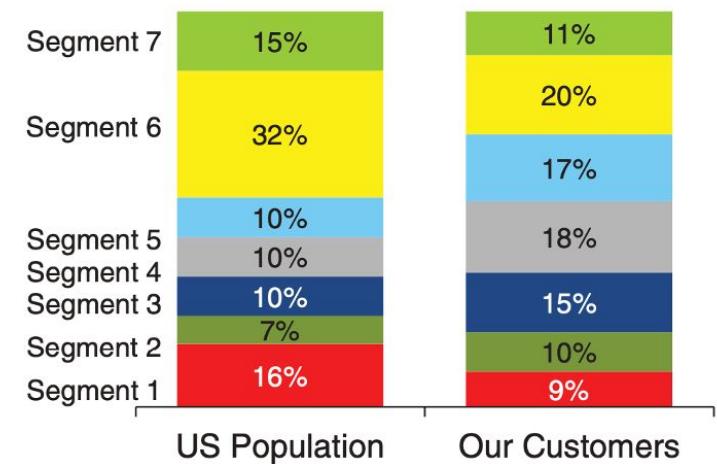
- Having all the information in the world at your fingertips doesn't make it easier to communicate: **it makes it harder**
- The more information you're dealing with, the more difficult it is to filter down to the most important bits
- It's **too easy** today to generate tables, charts, graphs
- This also generalizes to the broader realm of **presentations**

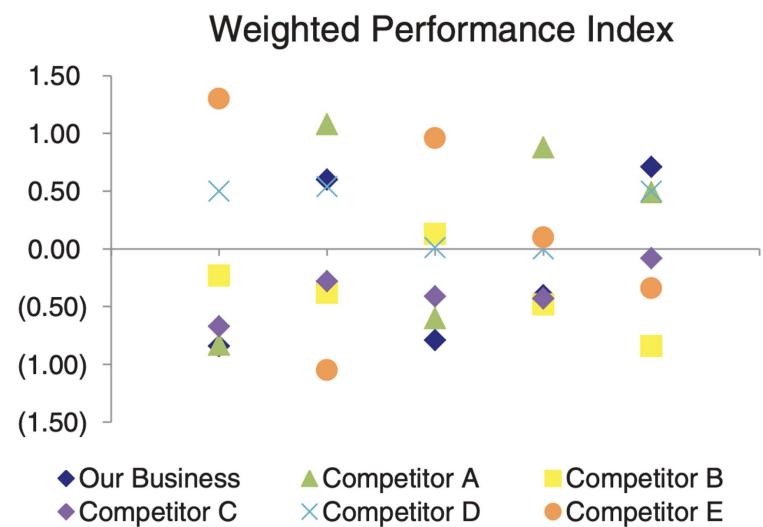
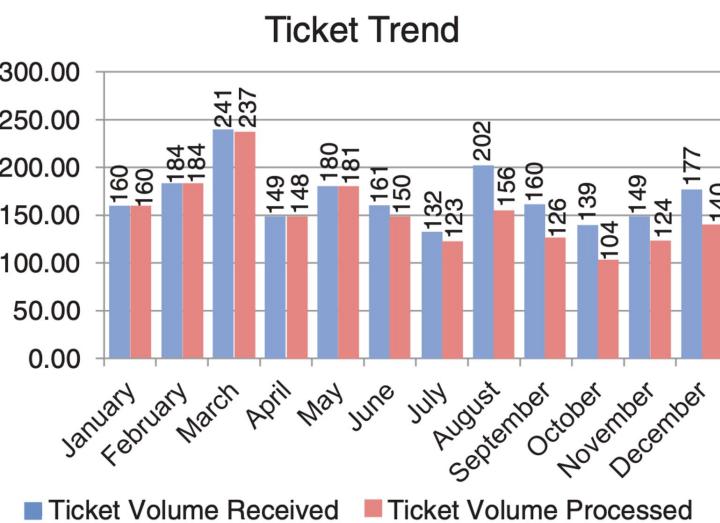


Survey Results



Our Customers





- In school, we learn a great deal about language and math
 - ◆ **On the language side**, we learn how to put words together into sentences and stories
 - ◆ **With math**, we learn to make sense of numbers
- These two sides are rarely paired...
- ... however, being able to visualize data and tell stories with them is key in many presentations, in **all domains** (not just business ones!)

An effective data visualization can mean the **difference between success and failure** when it comes to communicating the findings of your thesis, presenting to your supervisors at work, convince someone of the validity of your idea, or simply getting your point across to your audience during a seminar



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

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

Before even beginning to plan your presentation, or design your graphs, ask yourself:

- ◆ **[WHO]:** To whom am I communicating?
- ◆ **[WHAT]:** What do I want my audience to know or do?
- ◆ **[HOW]:** How can I use data to help make my point?

→ **Avoid general presentations**

- By trying to communicate to too many different people with disparate needs at once, you put yourself in a position where you cannot communicate to any one of them effectively
- Other questions:
- ◆ Does the audience already know you?
 - ◆ Do they already trust you?
 - ◆ Is there any common ground that you can leverage?

- You should always want your audience to know or do something
- If you can't concisely articulate that, you should revisit *whether you need to communicate* in the first place
- Better than just presenting data is *recommending an action*. When it really isn't appropriate to recommend an action explicitly, encourage discussion toward one
- This elicits a more productive reaction from your audience, which can lead to a more productive conversation

- What data is available that will help make my point?
- Which kinds of visuals should I employ?
- Are there also non-supporting data?
 - ◆ NEVER hide them!
- Remember: **live presentation ≠ offline report**

US Wireless Market – Q2 2010 Update

Executive Summary

The US wireless data market grew 6% Q/Q and 22% Y/Y to exceed \$13.2B in mobile data service revenues in Q2 2010 - on track so far to meet our initial estimate of \$54B for the year.

Having narrowly edged NTT DoCoMo last quarter for the first time, Verizon Wireless continued to maintain its number one ranking for the 1H 2010 in terms of the operator with the most mobile data revenues (though the difference was thinner than the amoeba membrane). The total wireless connections for Verizon were almost 100M with 92.1M being the traditional subscriber base. Rest of the 3 top US operators also maintained leading positions amongst the top 10 global mobile data operators.

Sprint had the first positive netadd quarter in 3 years and has been slowly and steadily turning the ship around. T-Mobile did better on the postpaid netadds but overall additions declined again. The larger question for the market is if 4 large players can stay competitive. Generally, the answer is no. But these are different times and there are a number of permutations and combinations that are possible.

The US subscription penetration crossed 95% at the end of Q2 2010. If we take out the demographics of 5 yrs and younger, the mobile penetration is now past 100%. While the traditional net-adds have been slowing, the "connected device" segment is picking up so much that both AT&T and Verizon added more connected devices than postpaid subs in Q2 2010. Given the slow postpaid growth, operators are fiercely competing in prepaid, enterprise, connected devices, and M2M segments.

Data traffic continued to increase across all networks. By 1H 2010, the average US consumer was consuming approximately 230 MB/mo up 50% in 6 months. US has become ground zero for mobile broadband consumption and data traffic management evolution. While it lags Japan and Korea in 3G penetration by a distance, due to higher penetration of smartphones and datacards, the consumption is much higher than its Asian counterparts. Given that it is also becoming the largest deployment base for HSPA+ and LTE, most of the cutting edge research in areas of data management and experimentation with policy, regulations, strategy, and business models is taking place in the networks of the US operators and keenly watched by players across the global ecosystem.

As we had forecasted, the tiered pricing structure for mobile broadband touched the US shores with AT&T becoming the first major operator to change its pricing plan based on consumer consumption. We will see the pricing evolve over the next 2-4 quarters as the US mobile ecosystem adjusts to the new realities and strategies for mobile data consumption.



WHO:

The advertising department board

WHAT:

The springtime promotional campaign regarding life insurances has been a success. Please approve budget \$X to continue with another campaign next year

HOW:

Illustrate success of the campaign with data collected by the data science team

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

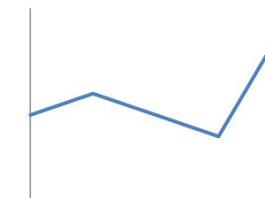
Although there are many different graphs, these are the (widely) most common visuals:

91%

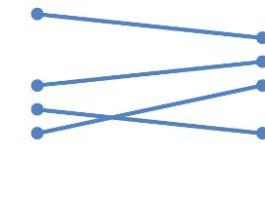
Simple text



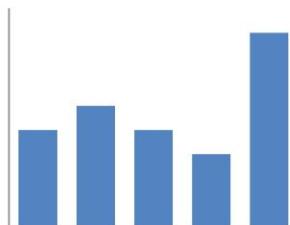
Scatterplot



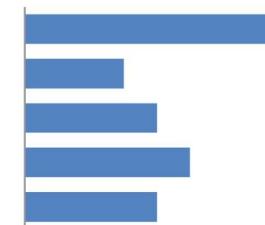
Line



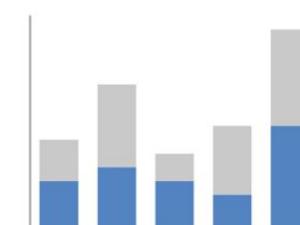
Slopegraph



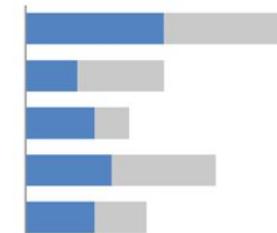
Vertical bar



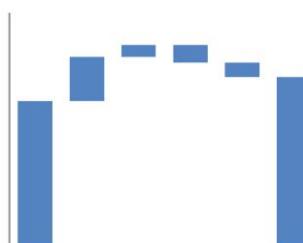
Horizontal bar



Stacked vertical bar



Stacked horizontal bar



Waterfall

	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

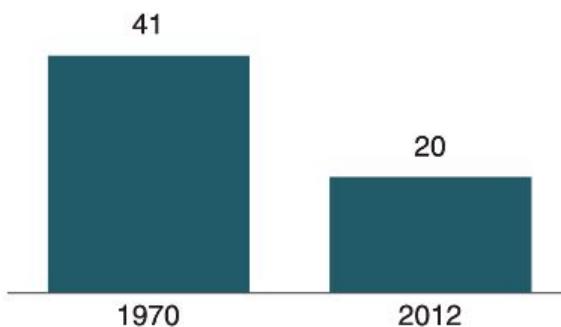
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

- When you have just a number or two to share, simple text can be a great way to communicate
- Think about solely using the number, making it as prominent as possible, and add a few supporting words to clearly make your point
- In other words: **when you have just a number or two that you want to communicate, use the numbers directly**

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

20%

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

- Using tables in a live presentation is **rarely a good idea**
- In any case, make sure that the table structure fades back into the background, leaving numbers as the protagonists

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

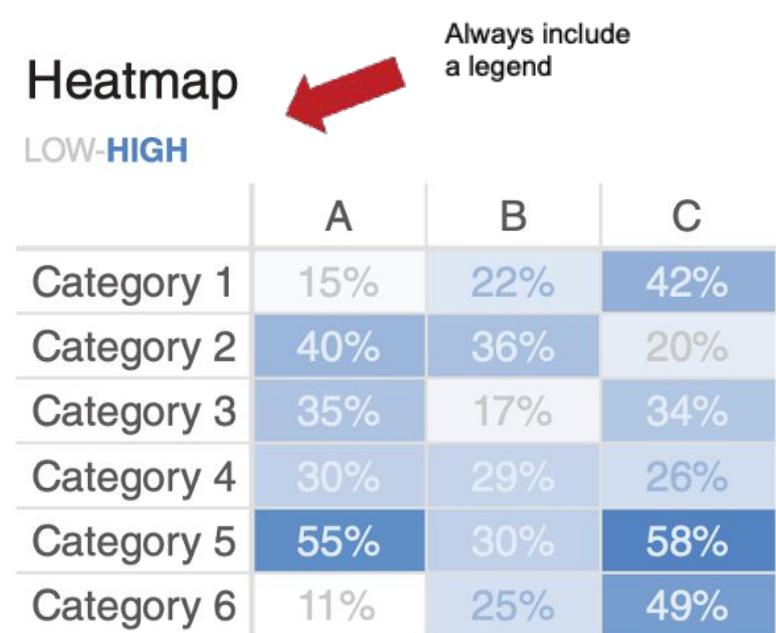
Minimal borders

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

To reduce the mental processing inherent with reading a table, you can use **color saturation** to provide visual clues

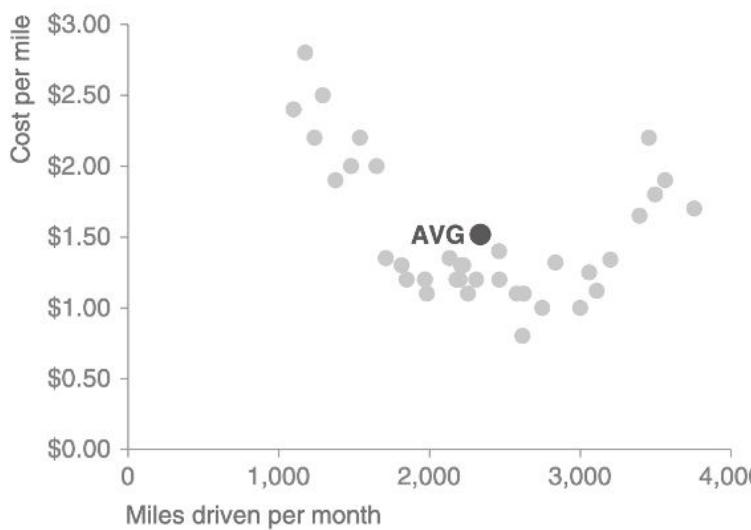
Table

	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%

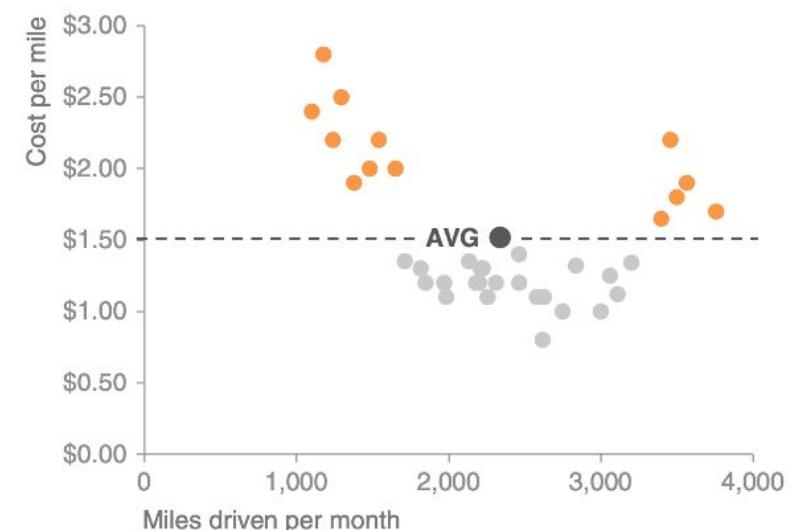


Useful for showing the relationship between two things

Cost per mile by miles driven



Cost per mile by miles driven



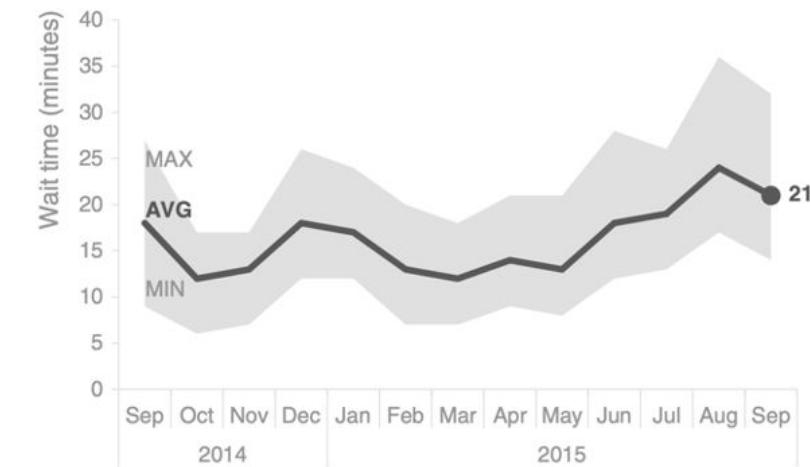
Here: focus is on those cases where the cost per mile is above average

- Here points are physically connected via a line, implying a **connection between them**
- Use them to plot **continuous** data (e.g., time), DO NOT use them for categorical data

Single series

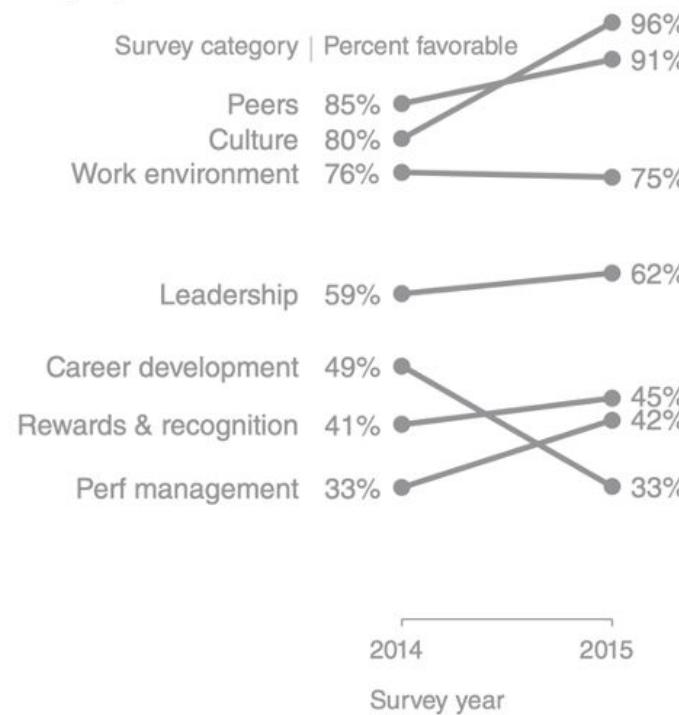


Passport control wait time
Past 13 months

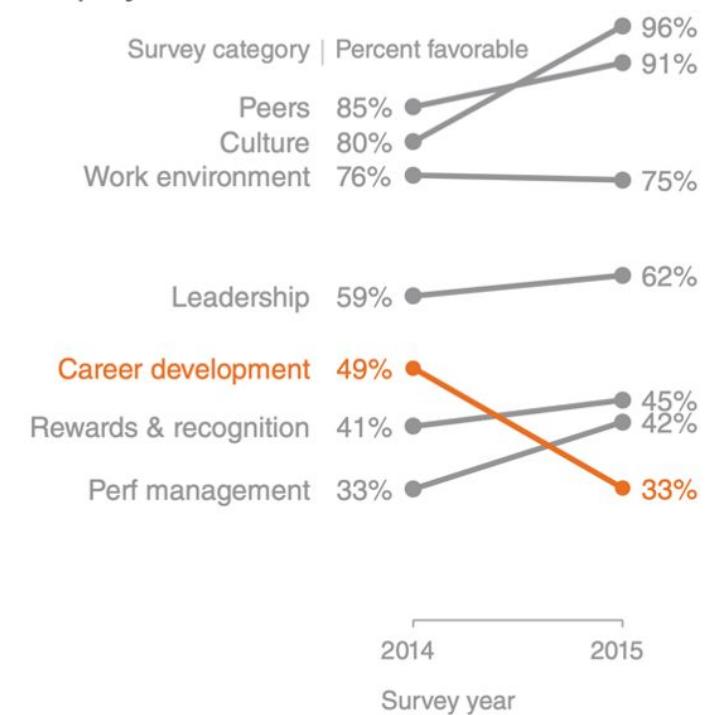


There are two time periods or points of comparison, and you want to show relative increases and decreases or **differences across various categories between the two data points**

Employee feedback over time

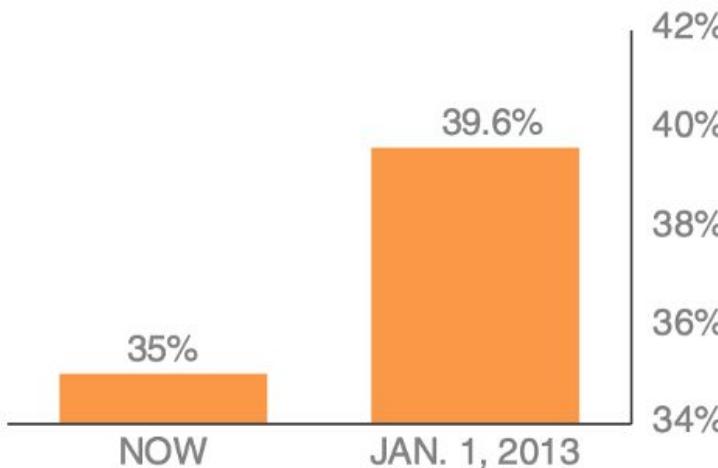


Employee feedback over time

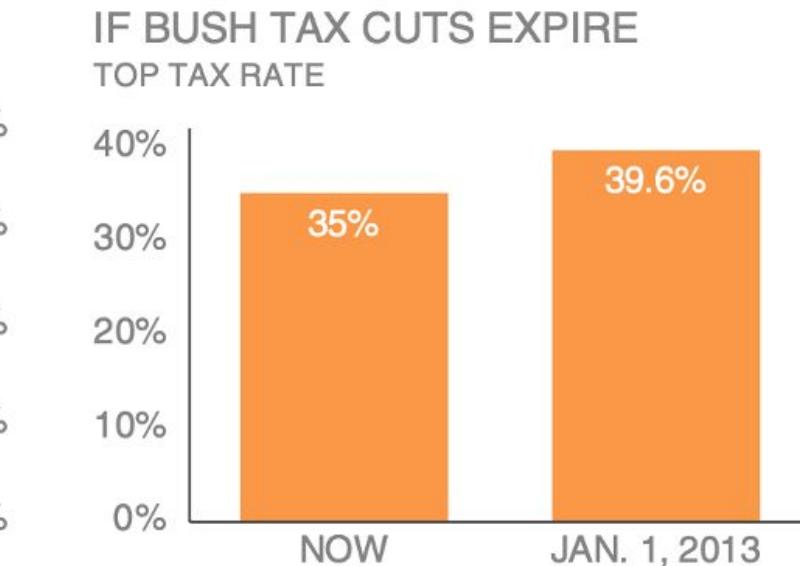


- Very useful because they are **common**, thus, easy to read
- Bar charts should always have a zero baseline!

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

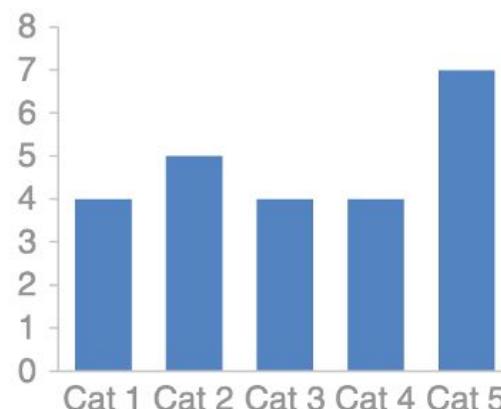


IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

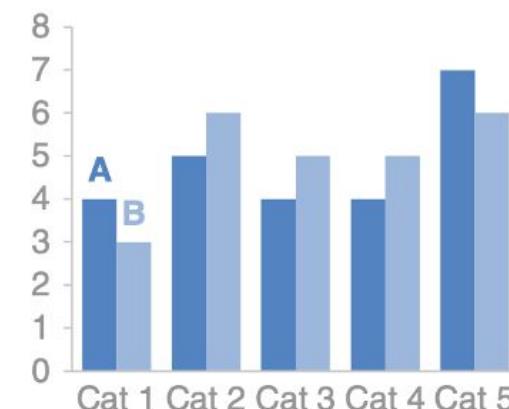


Here data on the horizontal axis is **categorical**, thus, a line plot would make no sense

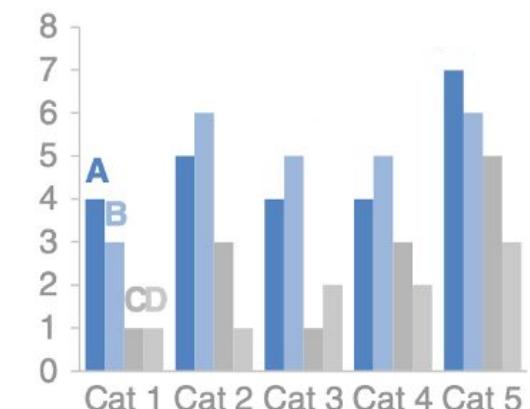
Single series



Two series



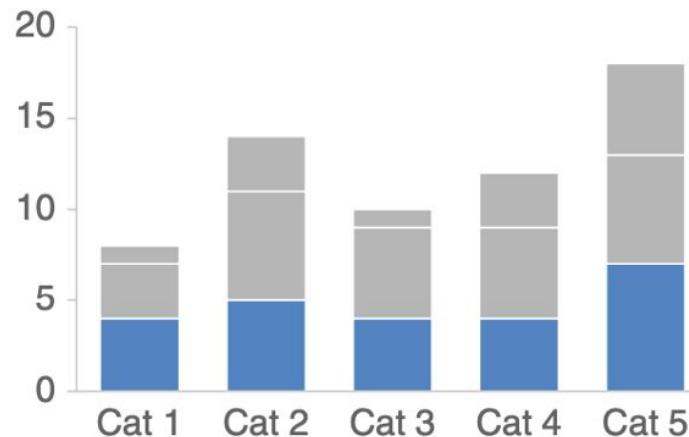
Multiple series



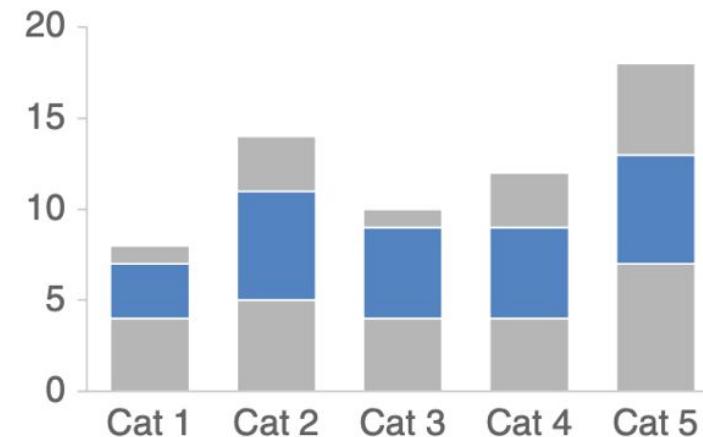
They also allow you to compare totals across categories

- E.g., for different exhibitions, you can display both the total number of visitors, and its decomposition into age range

Comparing **these** is easy

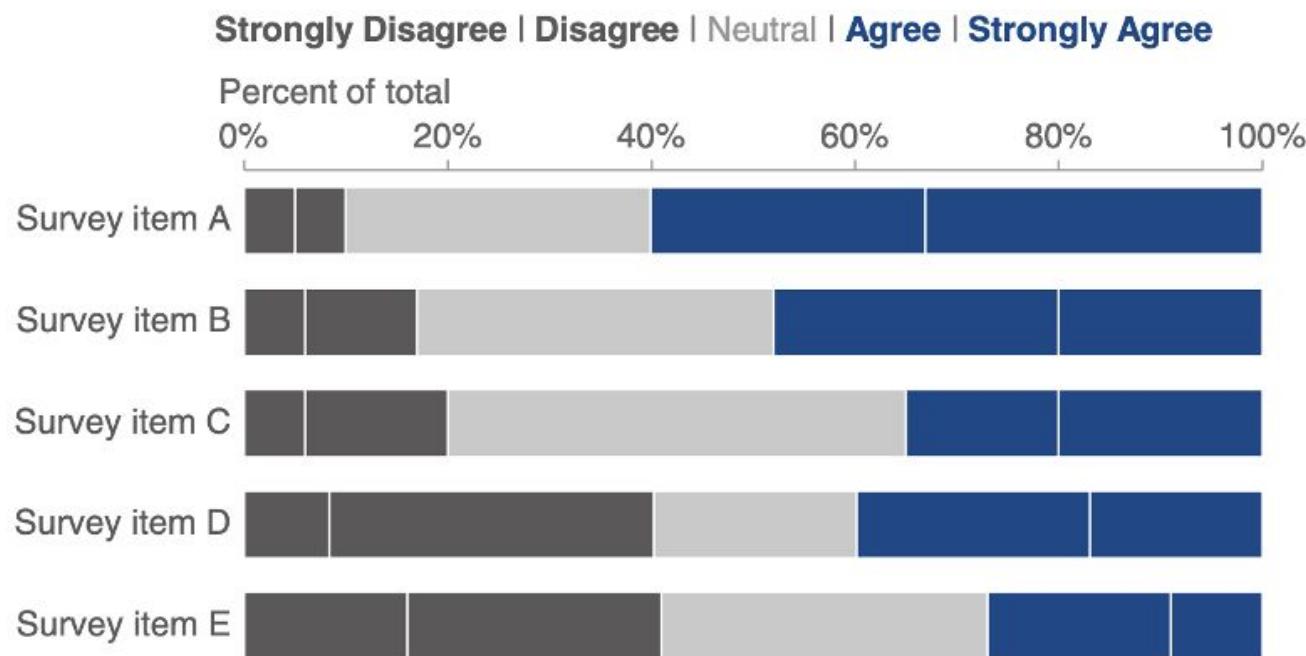


Comparing **these** is hard



Here, with bars that sum to 100%, it is easy to compare the extreme left/right bars

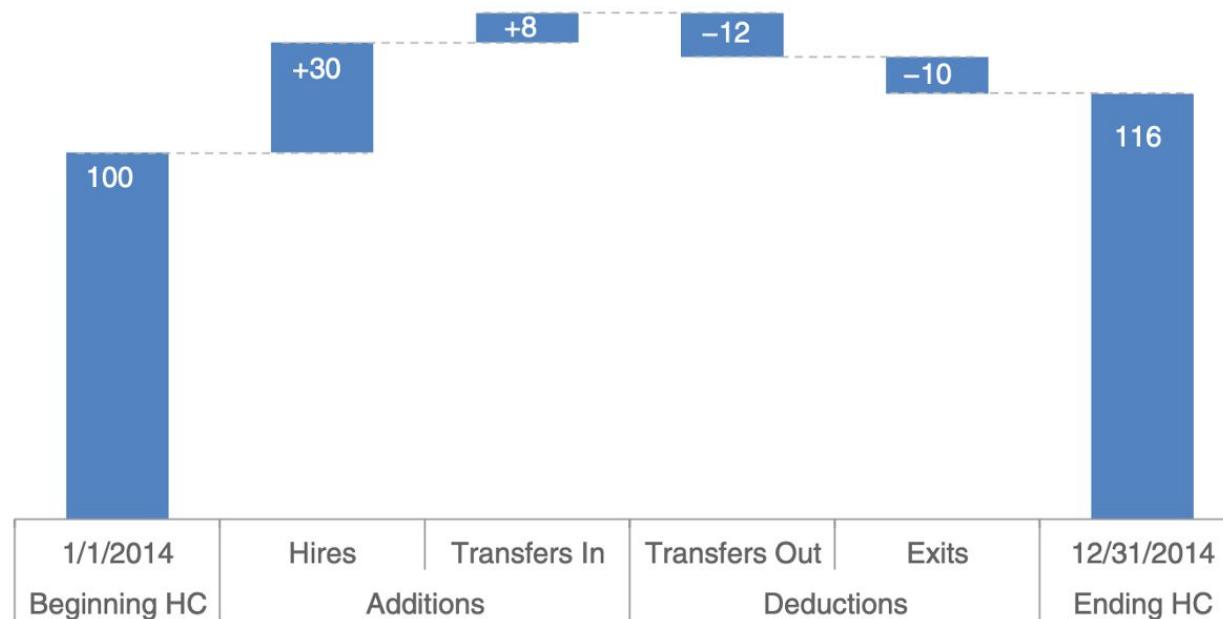
Survey results



They can be used to show how, from a starting point, we reach an ending point through a series of increases and decreases

2014 Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.



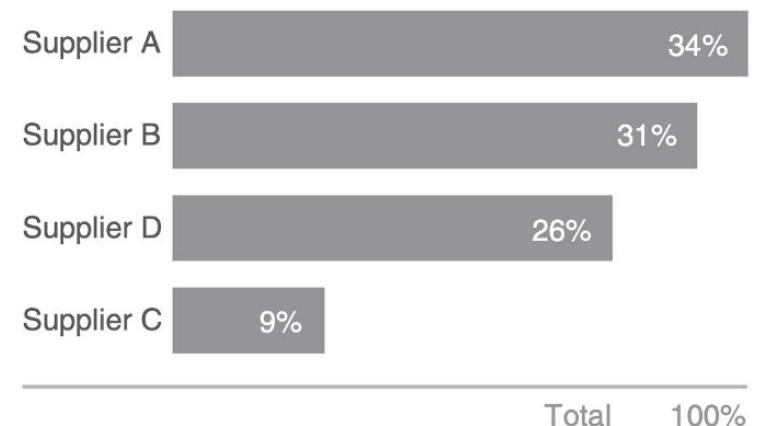
AVOID using graphs that use areas to convey information, at all costs

- E.g., pie charts, donut charts, and square area graphs
- They have a relative usefulness only when comparing numbers of vastly different magnitudes

Supplier Market Share



Supplier Market Share



Number of issues

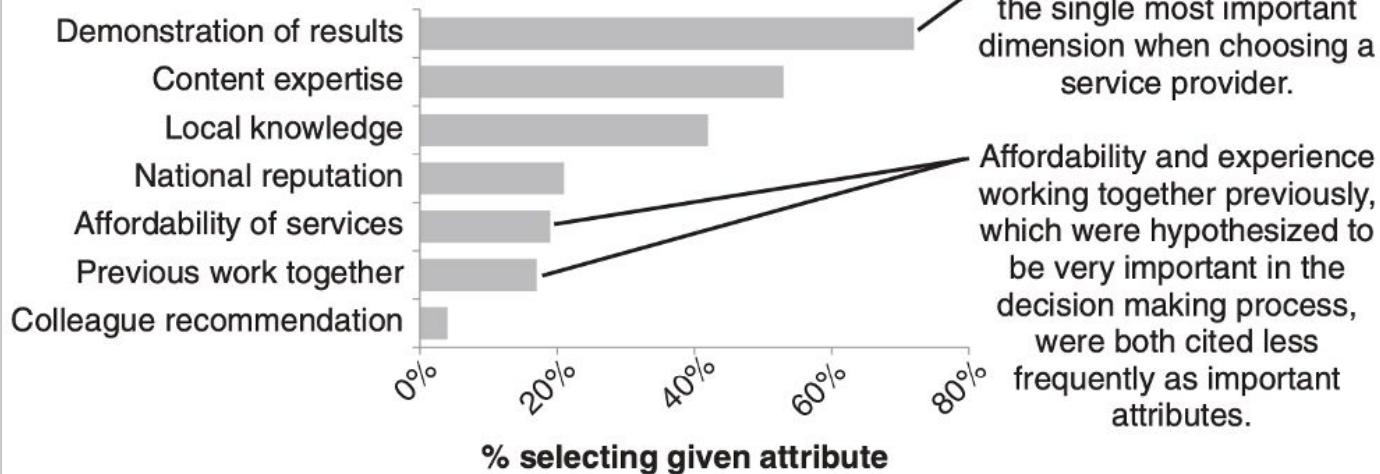


1. Understand the context
2. Choose an appropriate visual display
- 3. Eliminate clutter**
4. Focus attention where you want it
5. Think like a designer
6. Tell a story

- **Clutter:** visual elements that take up space but do not increase understanding
- Picture a blank page or a blank screen: every single element you add to that page or screen takes up cognitive load on the part of your audience
- Thus, try to make your graphs/slides **as simple as possible**
- Identify anything that **isn't adding (enough) informative value** and remove those things!
 - ◆ This is a subjective/domain-dependent concept!

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)



Survey shows that demonstration of results is the single most important dimension when choosing a service provider.

Affordability and experience working together previously, which were hypothesized to be very important in the decision making process, were both cited less frequently as important attributes.

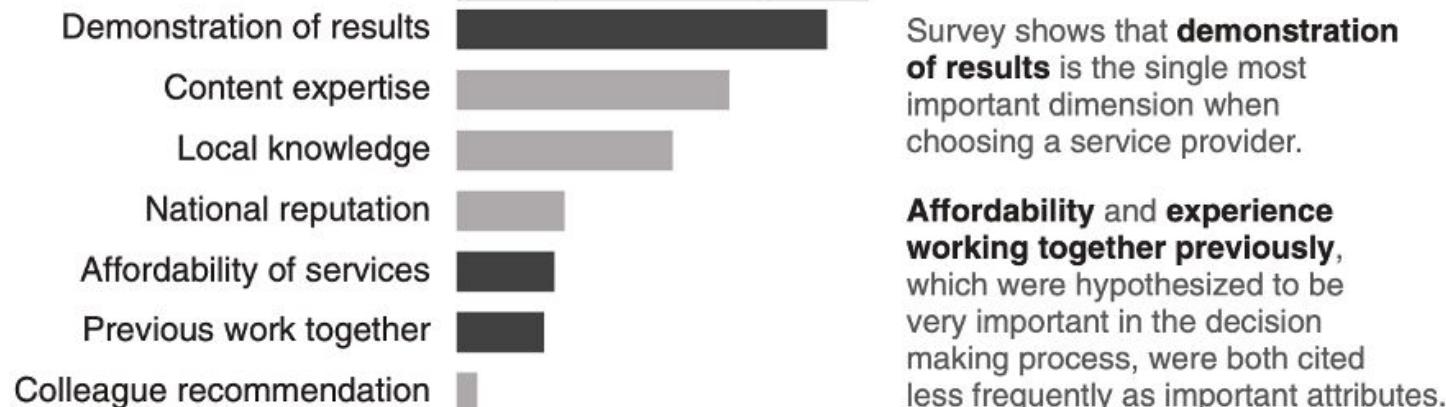
Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

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)

% selecting given attribute
0% 20% 40% 60% 80%



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



Performance overview

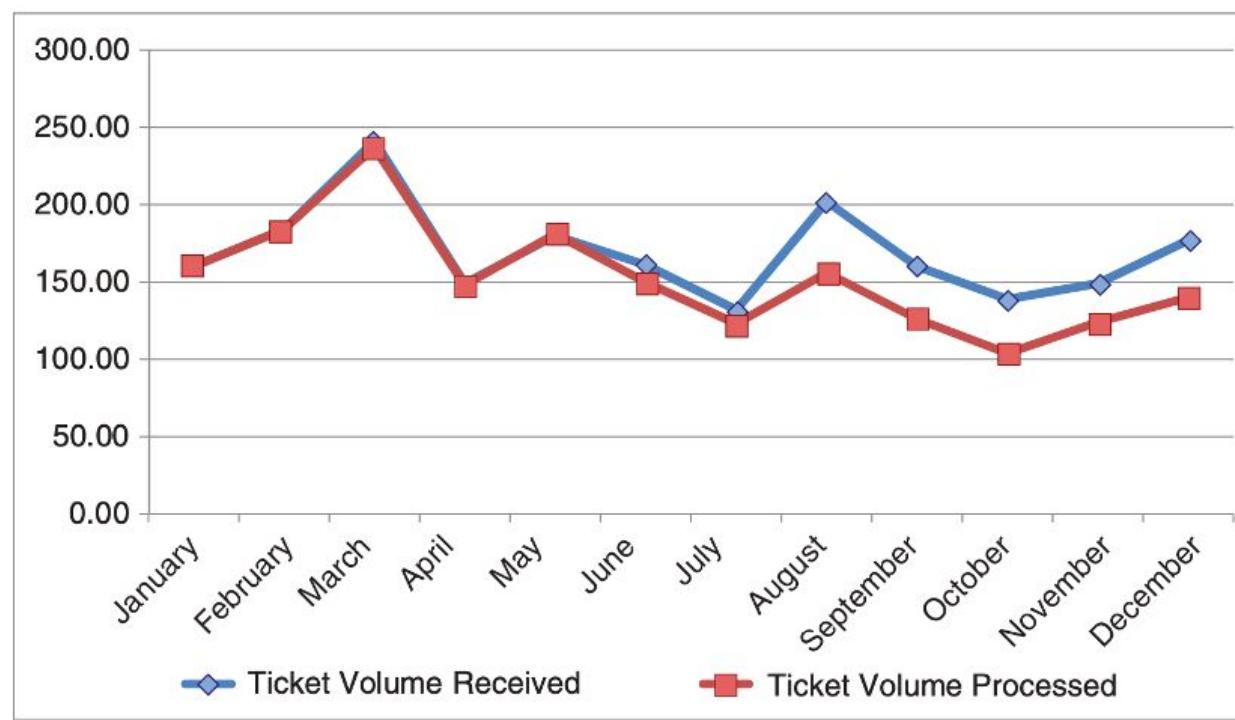
■ Our business

- Competitor A
- Competitor B
- Competitor C
- Competitor D
- Competitor E

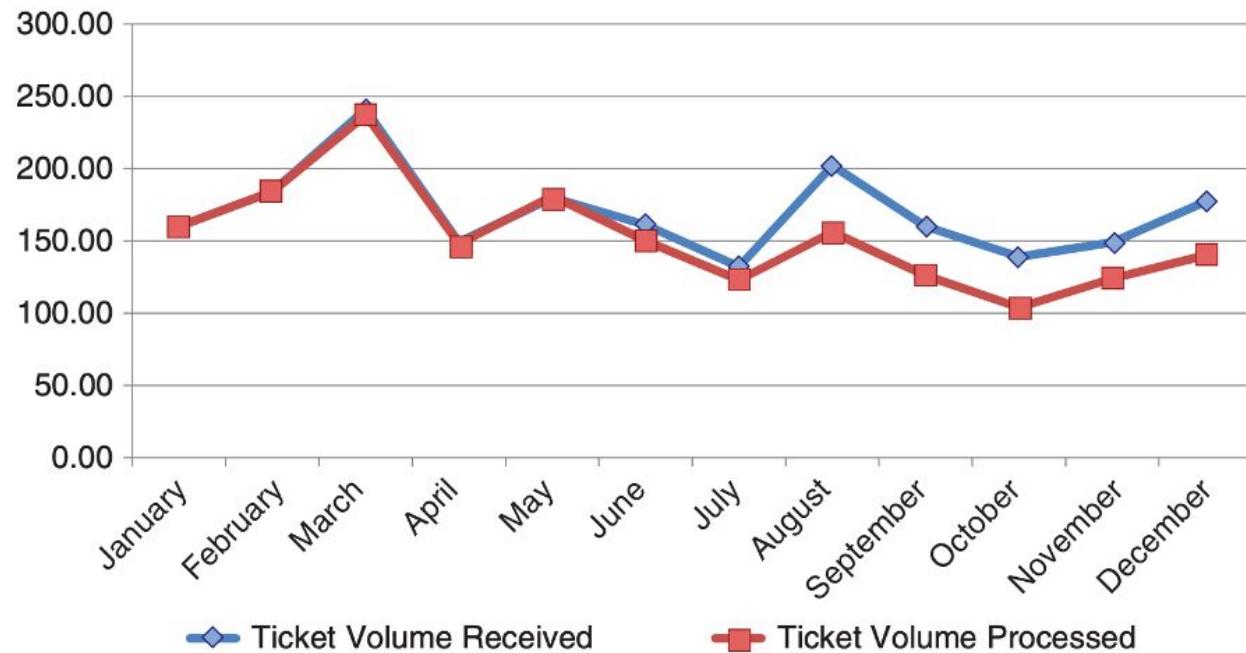
Weighted performance index | relative rank



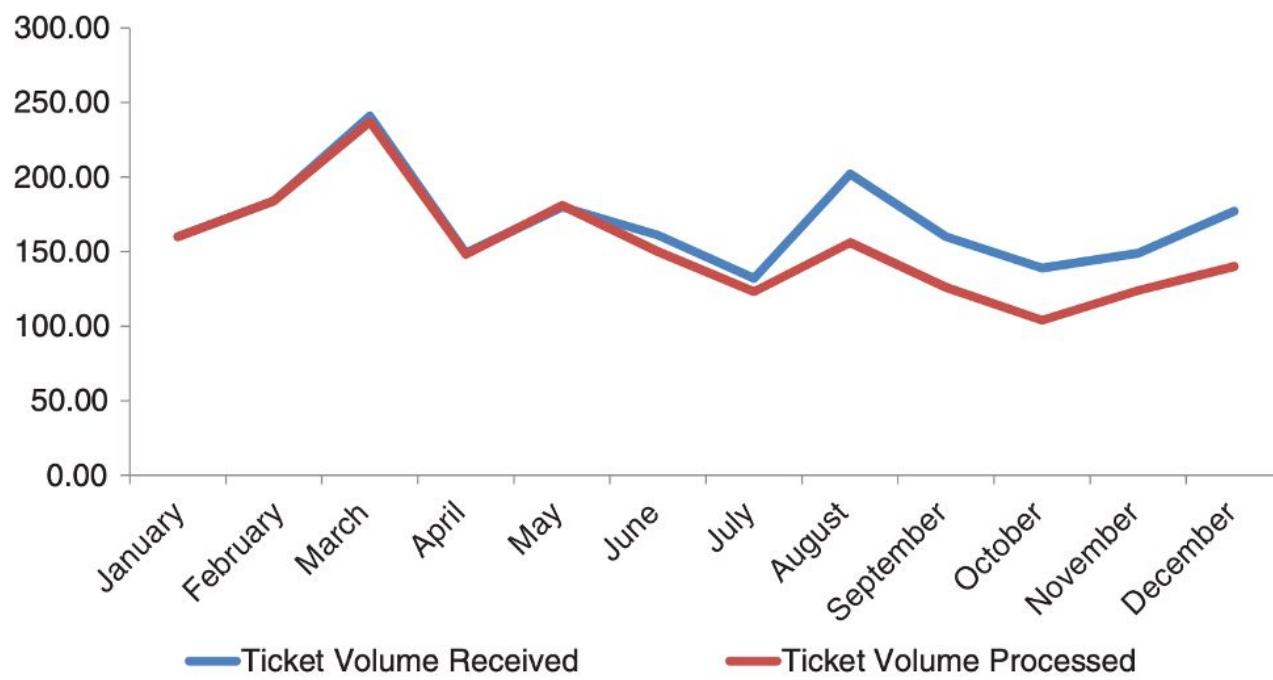
Scenario: you manage an IT team, that receives and addresses tickets from other employees. In the past year, a couple of people left the team in May and, since then, you hear complaints from the remaining ones about their workload. You have just been asked about your hiring needs for the next year. Thus, you want to investigate the matter further.



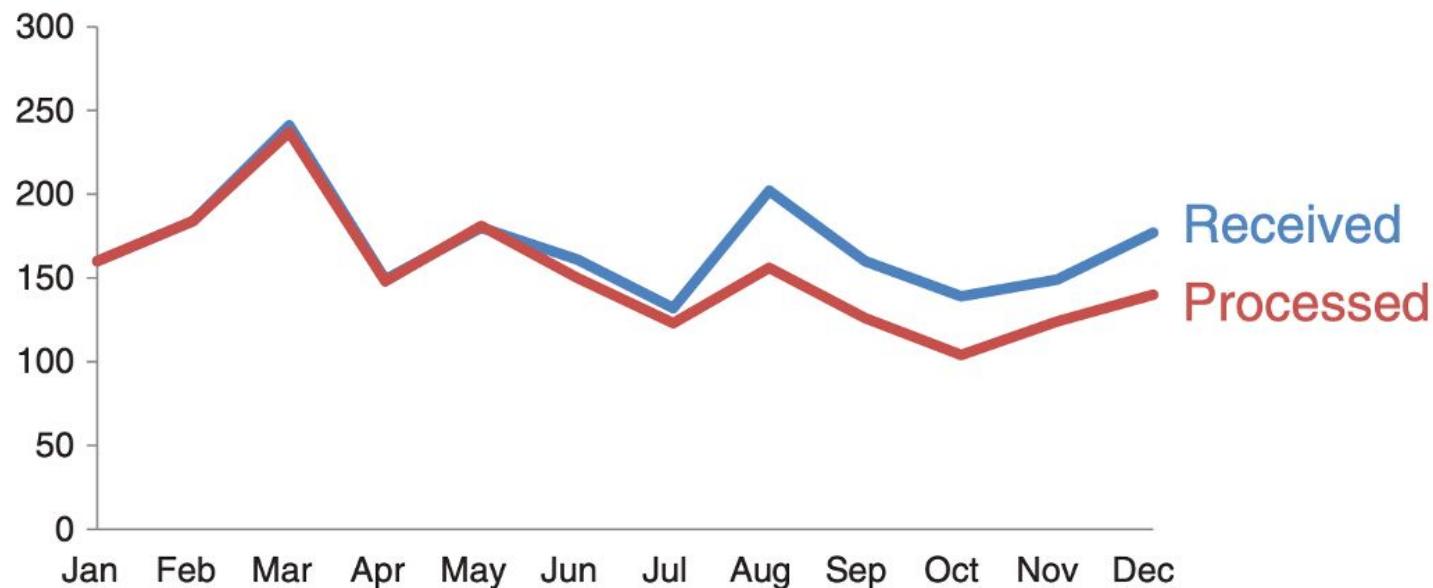
Instead of using the graph border, you can rely on white space to differentiate the visual from other elements of the page

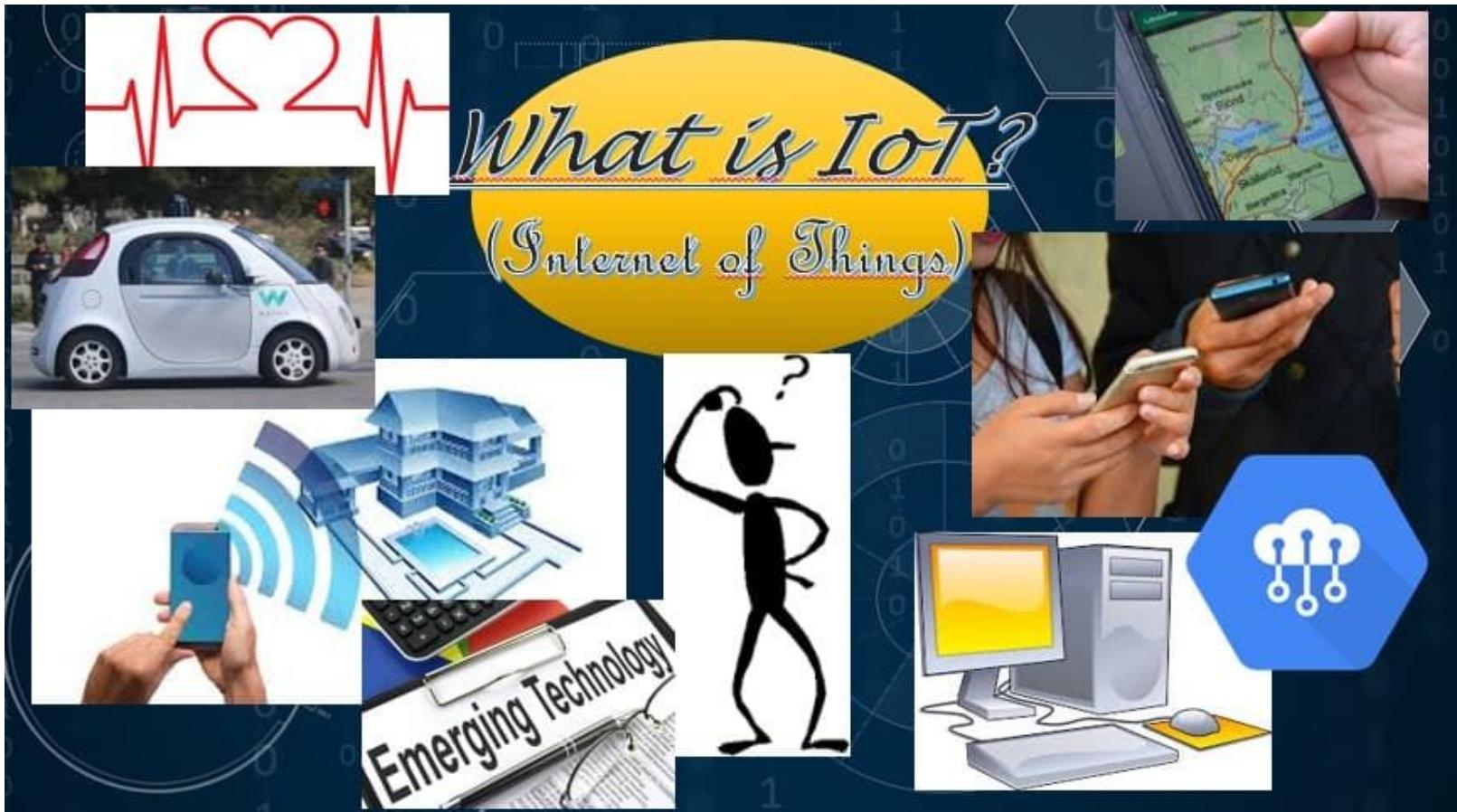


If relative trends are more important than precise numbers, remove gridlines; same for data markers that do not highlight specific events of interest



Override default editor choices if they do not seem fit to your purpose





Note also the font, which is *difficult to read...*

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

Compare this...

756395068473
658663037576
860372658602
846589107830

... with this

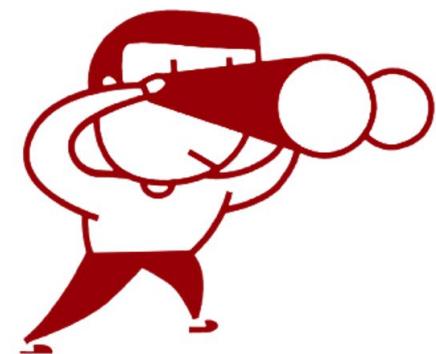
756**3**9506847**3**

65866**3**037576

860**3**72658602

8465891078**3**0

- The difference is the use of **preattentive attributes**, that leverage iconic memory directly
- The brain responds automatically and instantaneously to preattentive attributes, without any form of reasoning, which takes instead much more time
- Preattentive attributes enable our audience to see what we want them to see before they even know they are seeing it!



- Your eye is immediately drawn to the element within each group that is **different** from the rest
- Thus, preattentive attributes are extremely powerful to **focus the attention** of our public



Orientation



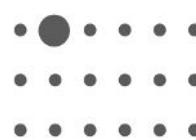
Shape



Line length



Line width



Size



Curvature



Added marks



Enclosure



Hue



Intensity



Spatial position



Motion

- Different preattentive attributes have different capabilities of drawing attention
- For example, **bold** text is “more powerful” than *italic*

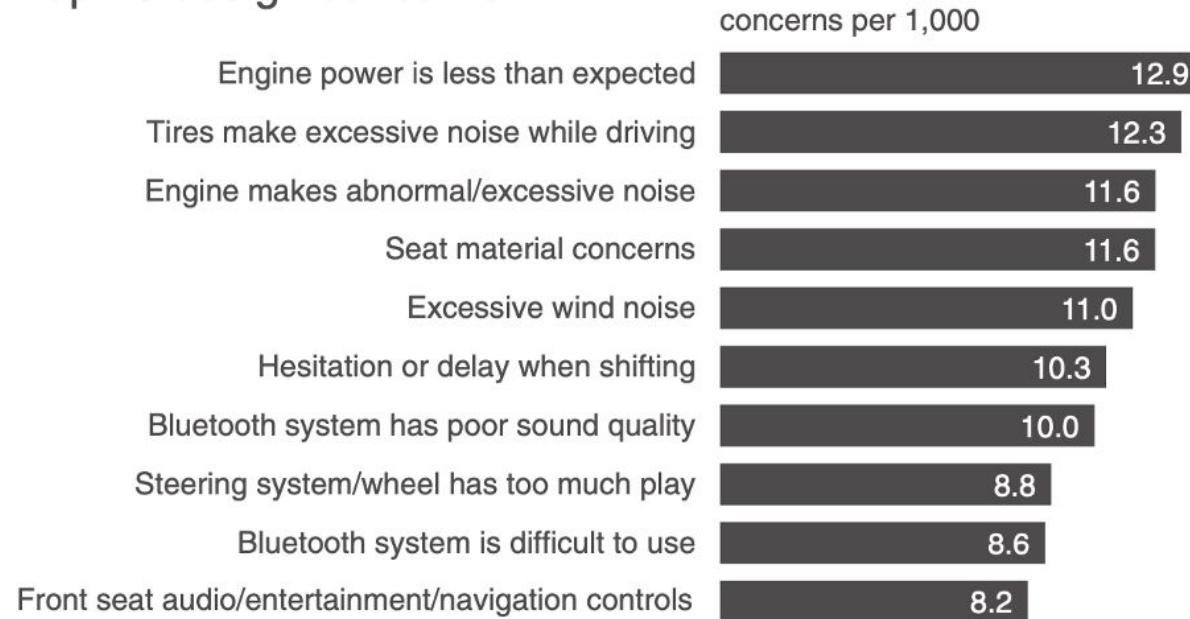
What are we doing well?

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!"

No preattentive attributes

Top 10 design concerns

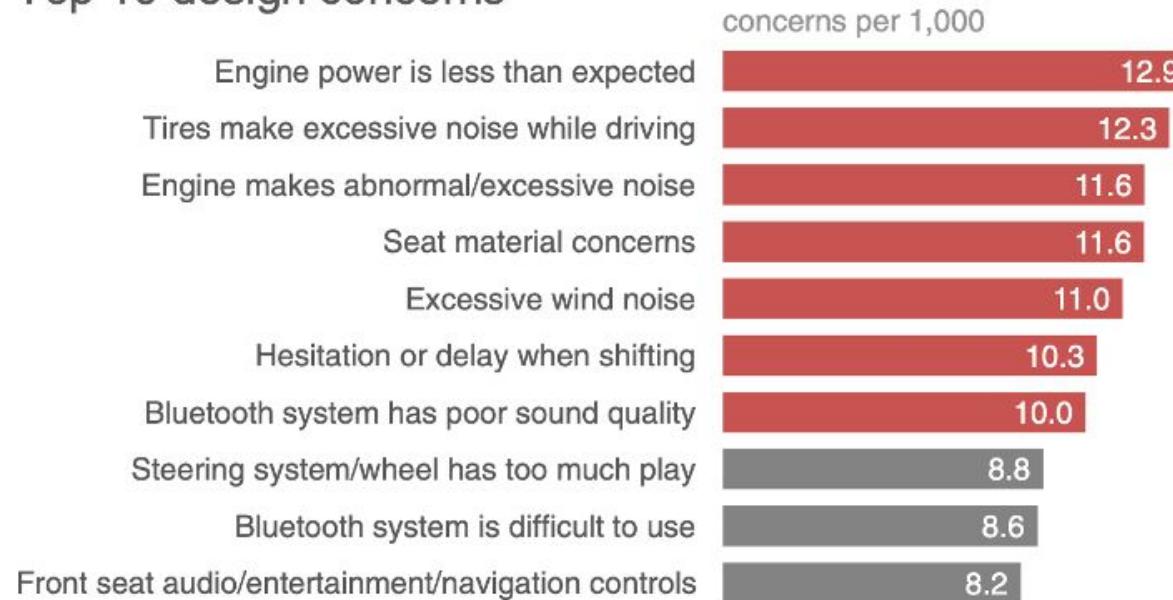


With preattentive attributes

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

Discussion: is this an acceptable default rate?

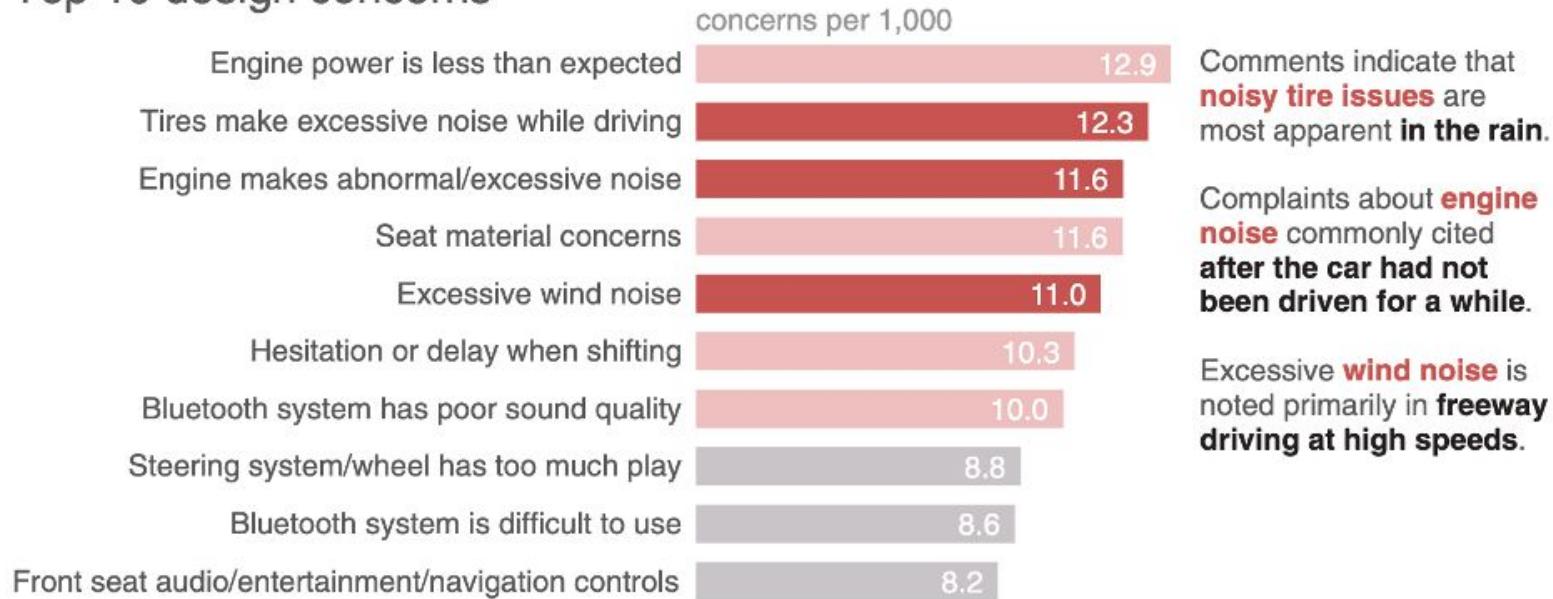
Top 10 design concerns



Even more preattentive attributes... one or more stories develop

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

Top 10 design concerns



An over-usage preattentive attributes leads to the opposite result

Country Level Sales Rank Top 5 Drugs

Rainbow distribution in color indicates sales rank in given country from #1 (red) to #10 or higher (dark purple)

Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5

Top 5 drugs: country-level sales rank

RANK	1	2	3	4	5+
COUNTRY DRUG	A	B	C	D	E
Australia	1	2	3	6	7
Brazil	1	3	4	5	6
Canada	2	3	6	12	8
China	1	2	8	4	7
France	3	2	4	8	10
Germany	3	1	6	5	4
India	4	1	8	10	5
Italy	2	4	10	9	8
Mexico	1	5	4	6	3
Russia	4	3	7	9	12
Spain	2	3	4	5	11
Turkey	7	2	3	4	8
United Kingdom	1	2	3	6	7
United States	1	2	4	3	5

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

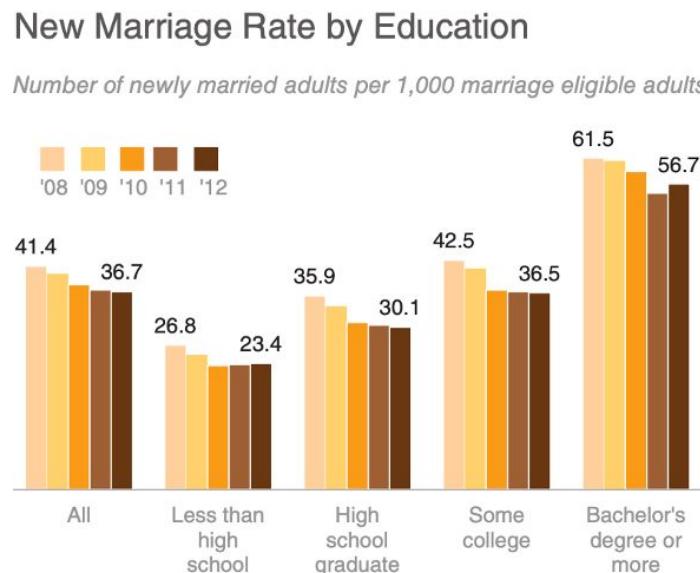
Form

First, think about what it is that you want your audience to be able to do with the data

Function

Then, create a visualization that will allow for that with ease

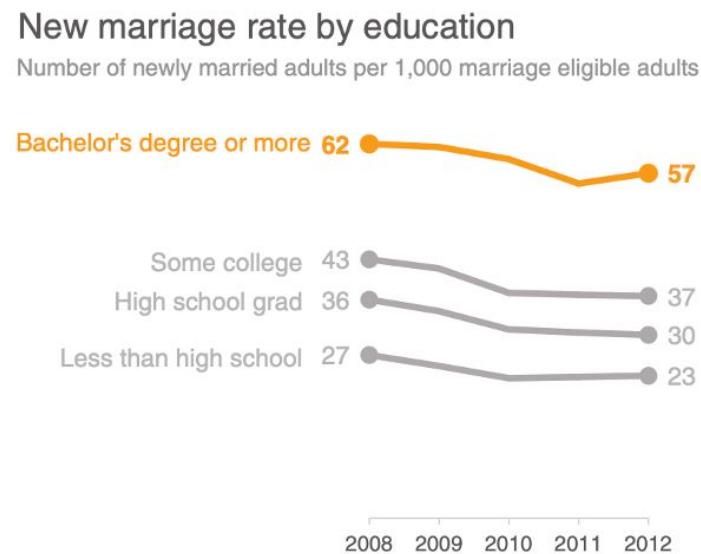
1. Choose the **right kind of visual** to convey your message
2. Highlight the important stuff using **preattentive attributes**
3. Eliminate distractions **removing clutter**
4. Make the visual **accessible** with text and careful labelling



Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

Adapted from PEW RESEARCH CENTER



Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

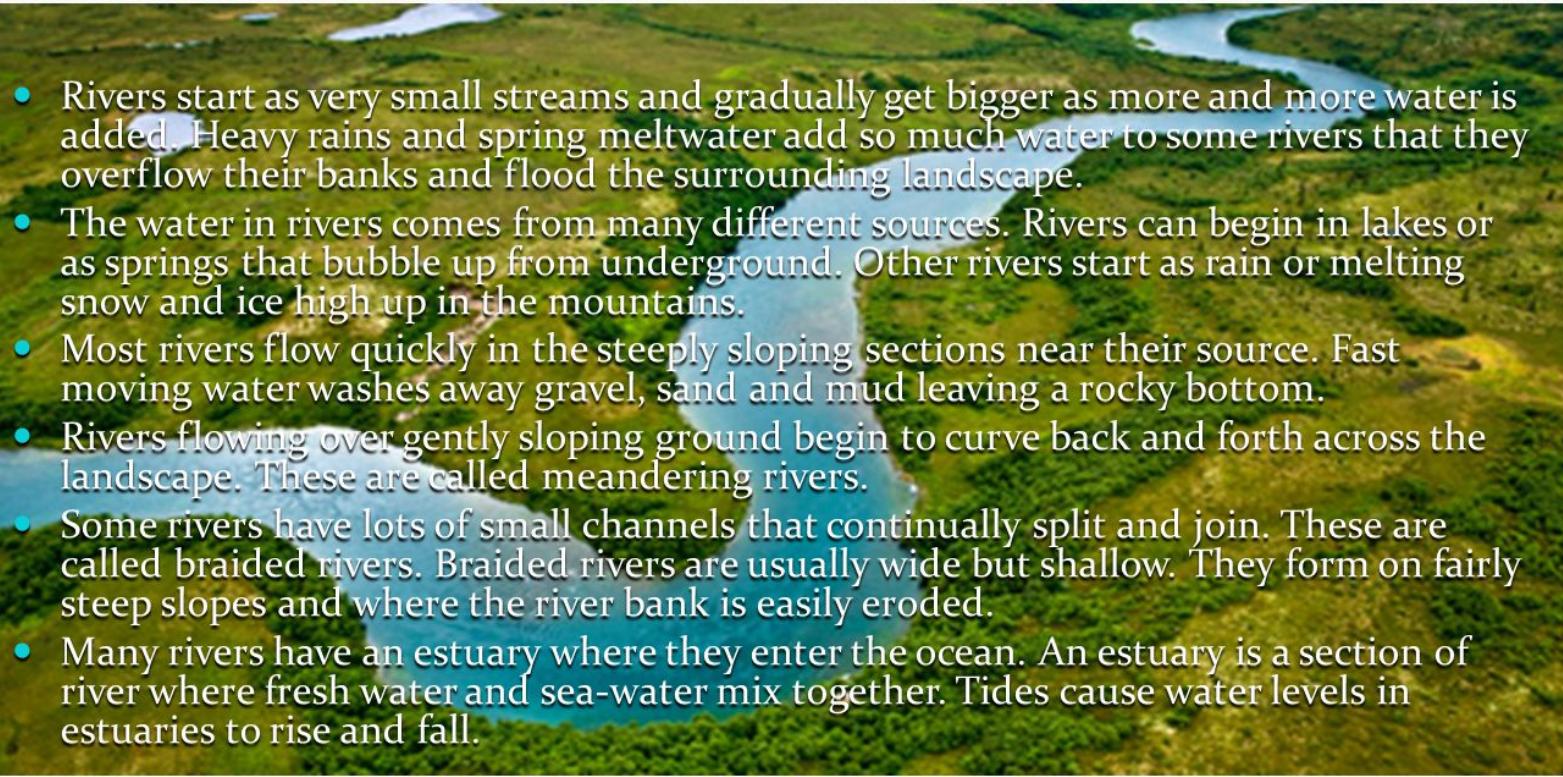
Adapted from PEW RESEARCH CENTER



*“You know you’ve achieved perfection, not when you have nothing more to add,
but when you have nothing to take away.”*

— Saint-Exupery, 1943

How Rivers Are Formed

- 
- An aerial photograph showing a river flowing through a lush, green landscape. The river's path is winding and meandering, creating a series of curves and loops. The surrounding terrain appears to be a mix of fields and possibly some forested areas. The water of the river is a light blue-grey, contrasting with the green of the land.
- Rivers start as very small streams and gradually get bigger as more and more water is added. Heavy rains and spring meltwater add so much water to some rivers that they overflow their banks and flood the surrounding landscape.
 - The water in rivers comes from many different sources. Rivers can begin in lakes or as springs that bubble up from underground. Other rivers start as rain or melting snow and ice high up in the mountains.
 - Most rivers flow quickly in the steeply sloping sections near their source. Fast moving water washes away gravel, sand and mud leaving a rocky bottom.
 - Rivers flowing over gently sloping ground begin to curve back and forth across the landscape. These are called meandering rivers.
 - Some rivers have lots of small channels that continually split and join. These are called braided rivers. Braided rivers are usually wide but shallow. They form on fairly steep slopes and where the river bank is easily eroded.
 - Many rivers have an estuary where they enter the ocean. An estuary is a section of river where fresh water and sea-water mix together. Tides cause water levels in estuaries to rise and fall.

5 Great Productivity Apps



[ToDoist](#)

Ultimate “to-do” list app



[Slack](#)

communication app for collaboration



[Toggl](#)

Time tracker app



[Evernote](#)

Organizing and IdeThoughtsas

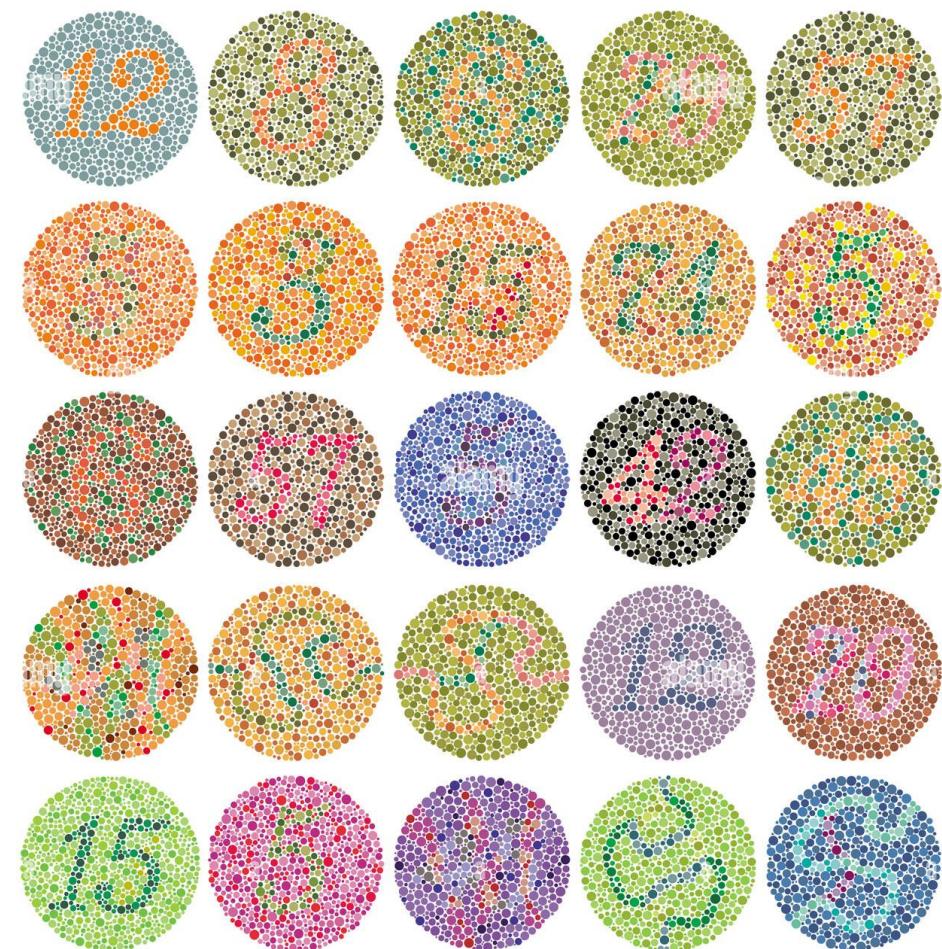


[Trello](#)

Kanban Project Management App



In individuals of Northern European ancestry, color blindness affects 8% of men and 0.4% of women



Ishihara color perception test

[https://www.color-blindness.com/
coblis-color-blindness-simulator/](https://www.color-blindness.com/coblis-color-blindness-simulator/)



- Fonts can be grouped into two categories:
 - ◆ Serif
 - ◆ Sans-serif
- Serifs are those **elongations** at the ends of letters
- They may derive from the marks left by scribes at the end of each brushstroke, a habit which then evolved into the intentional addition of small regular strokes which became an integral part of the letters
- In small-print text, serif fonts help you distinguish the letters and not lose your reading thread

Times New Roman

Aa Ee Rr **a**
Aa Ee Rr
Publisher

abcdefghijklmnopqrstuvwxyz
0123456789

Traditional,
established,
trustworthy

Futura

Aa Qq Rr
Aa Qq Rr
Zuführung

abcdefghijklmnopqrstuvwxyz
0123456789

Modern,
accessible,
clean

Helvetica

Futura

Bodoni

Gotham

FF DIN

Baskerville

Brandon Grotesque

Circular

Titillium Web

Roboto

Comic Sans

Lobster

Bradley Hand

Papyrus

Courier New

Kristen ITC

Impact

Arial

STENCIL

Myriad Pro

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

A good story grabs the attention and takes you on a journey, evoking an emotional response

Fortunately, stories have a time-tested structure; humans have been communicating with stories throughout history

We can leverage this powerful tool for **our communications**



The notion of narrative structure was first described in ancient times by Greek philosophers such as Aristotle and Plato; **three acts**:

1. **Sets up the story**, the main characters, their relationships, and the world around them; an incident occurs, leading to a dramatic situation and rising a question
2. Describes the main character's (unsuccessful) **attempts to resolve the problem**; the main characters goes through major changes
3. **Resolves the story** and all subplots; it includes a climax, where the tensions reach the highest point of intensity; the question introduced in the first act is answered

In business stories terms:

1. Get everyone on **common ground**, building the context for your audience; elicit the problem, and the difference between “what is” and “what could be”
2. Further develop “what could be”, retaining the audience’s attention by addressing **how they can solve the problem** you introduced, possibly proposing several options. Provide examples. What will happen if no action or decision is taken?
3. End with a **call to action**: make clear to your audience how they should leverage the new understanding or knowledge you have imparted to them

Summarize your message with a **single sentence**, with three characteristics:

1. It must articulate your unique point of view
2. It must convey what is at stake
3. It must be a complete, short sentence

“Since we lost two employees in May it has become increasingly difficult to cope with the workload, thus, please, consider hiring some new staff.”

“The springtime promotional campaign on life insurances was successful in increasing sales and, because of this, we recommend to continue with another campaign next year.”

- Storyboarding is perhaps the single most important thing you can do up front to ensure the communication you craft is on point
- The storyboard establishes a structure for your communication; it is a (physical) visual outline of the content you plan to create
- **Don't start** with presentation software right away:
 - ◆ Promotes slide-generating rampage
 - ◆ Fosters attachment to your creation



Scenario:

- Assume you work for a company that has designed a **new product** for its customers
- You have been asked by the management **how to price** the product
- One of the considerations in the decision-making process is how competitors' retail prices for products in this marketplace have changed over time
- Let us cover the 6 main points of storytelling with this example

WHO:

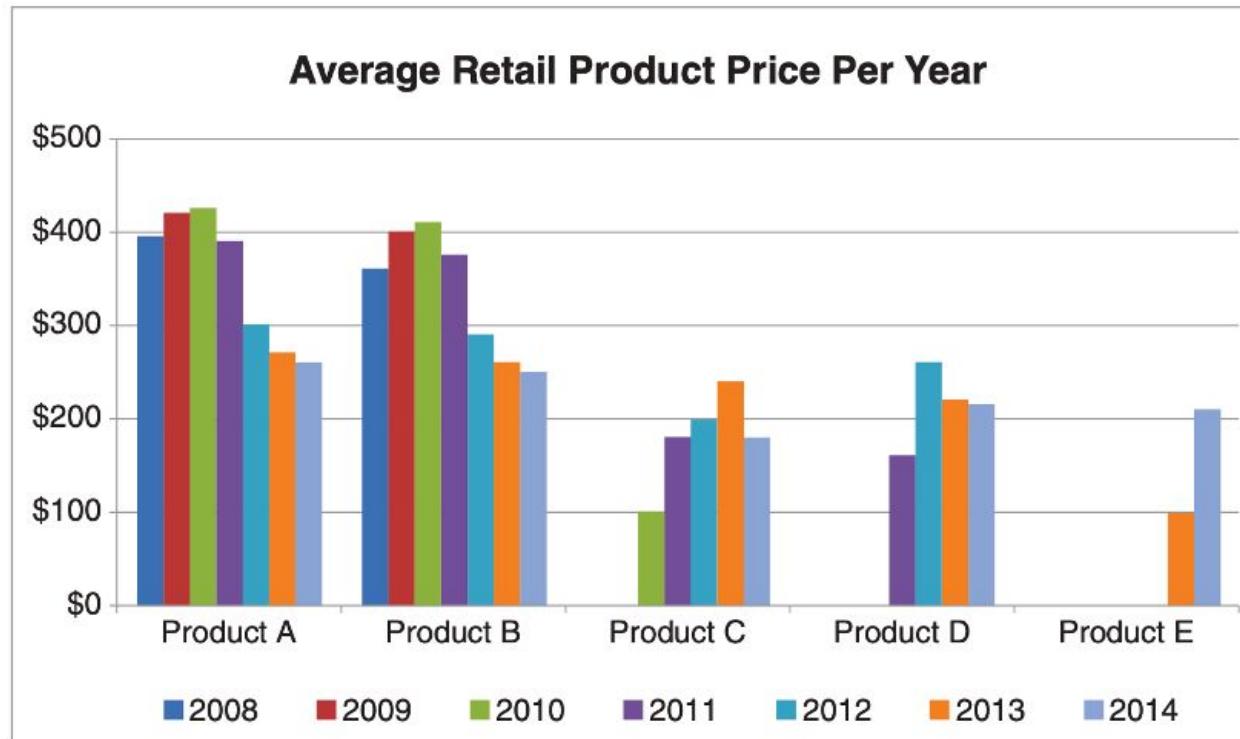
Sales department board, the primary decision maker in establishing the new product's price

WHAT:

Understand how competitors' pricing has changed over time and recommend a price range for the new product

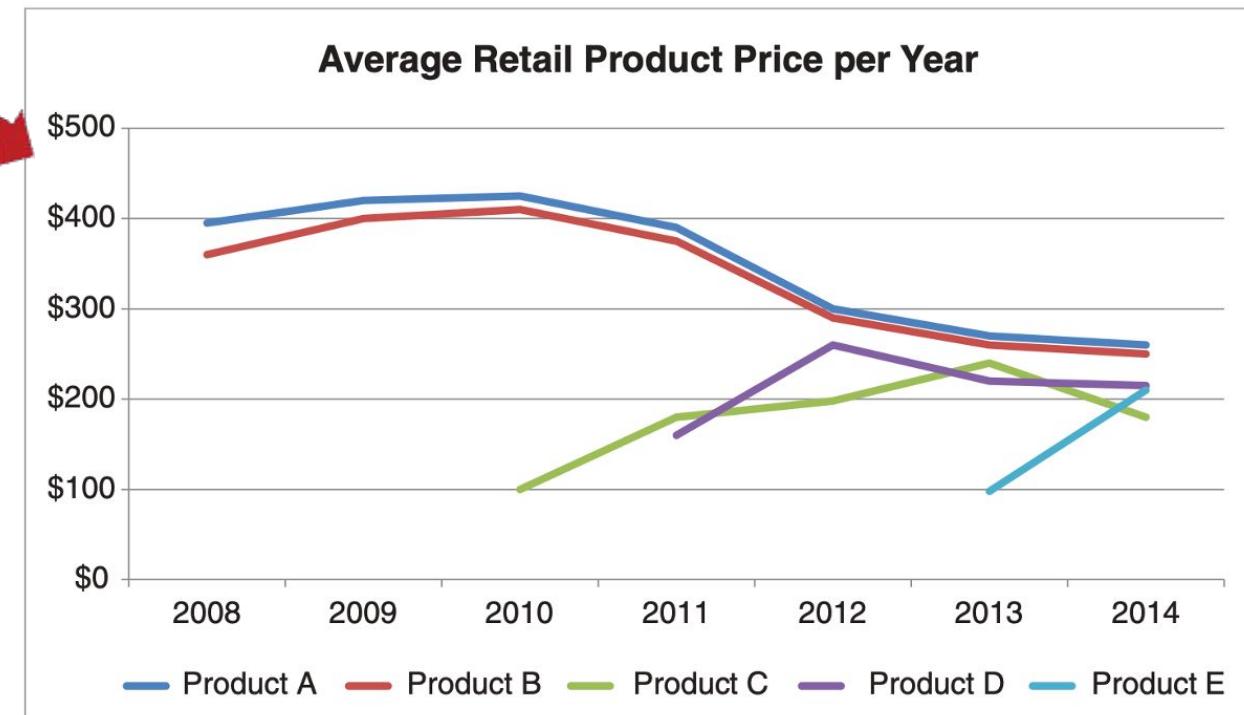
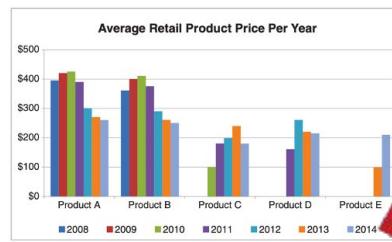
HOW:

Show average retail price over time for Products A, B, C, D, and E, that are those present in the same marketplace



- The graph shows the average retail price over time for five consumer products (A, B, C, D, and E)
- Price has declined for older products on the market since the launch of “Product C” in 2010

- Here, we are interested in the price trend over time for each product
- Thus, let us transform the bar plot into a line plot
- It is much easier now to compare the different price trends over time

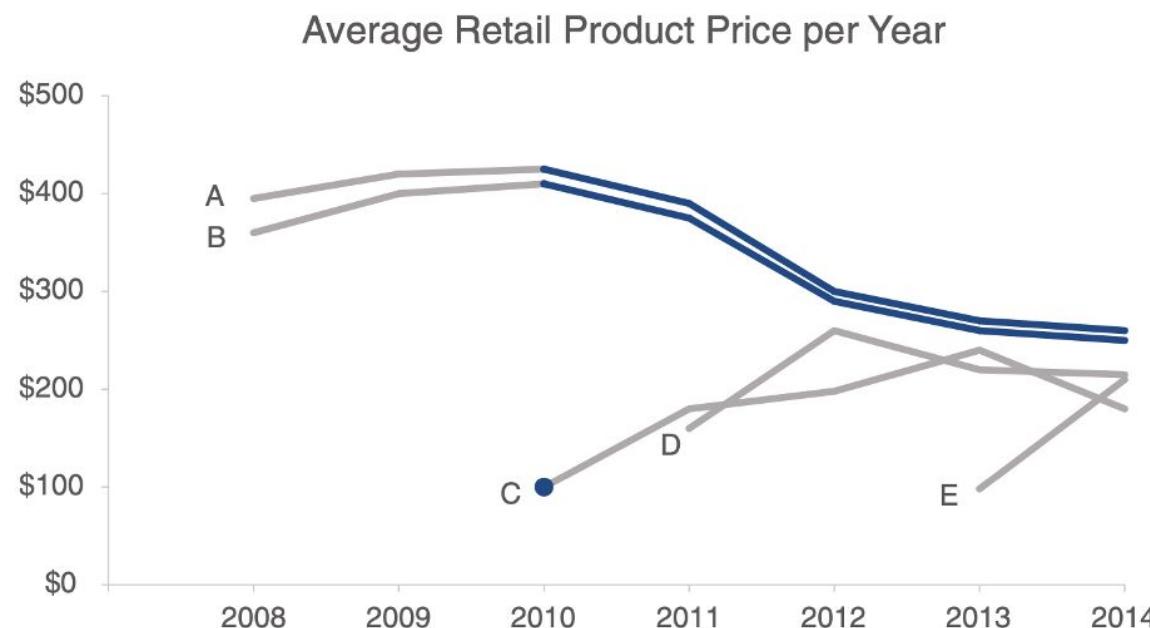


Let us change the default settings of our graphing application, so to better concentrate on what we want to highlight:

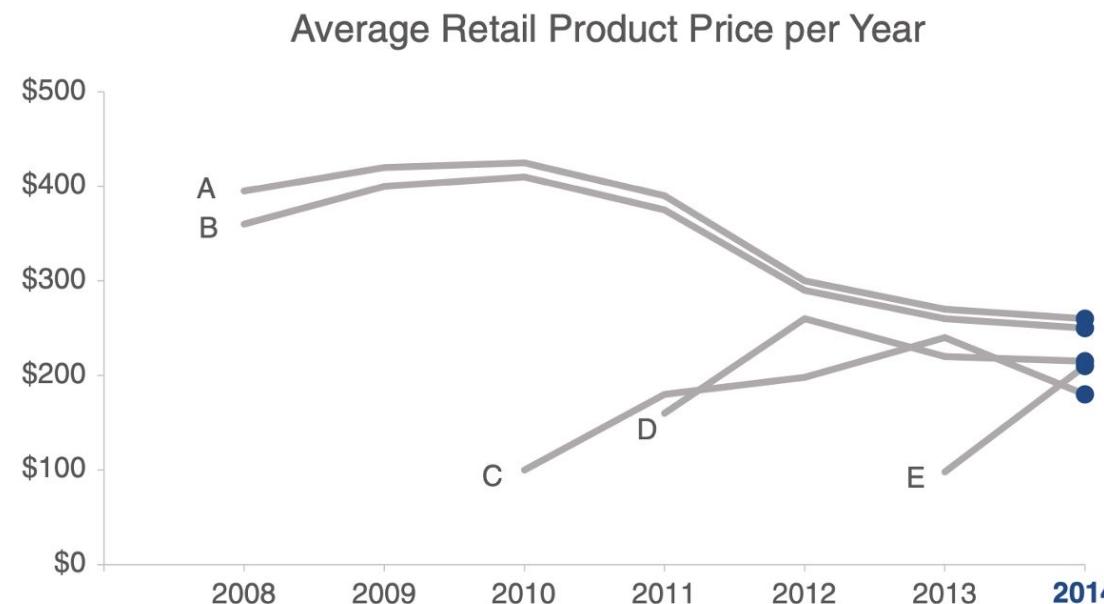
- De-emphasize chart title
- Remove borders and gridlines
- Push back axis lines and labels
- Remove variance in colours
- Label the lines directly



Price has declined for older products on the market since the launch of “Product C” in 2010



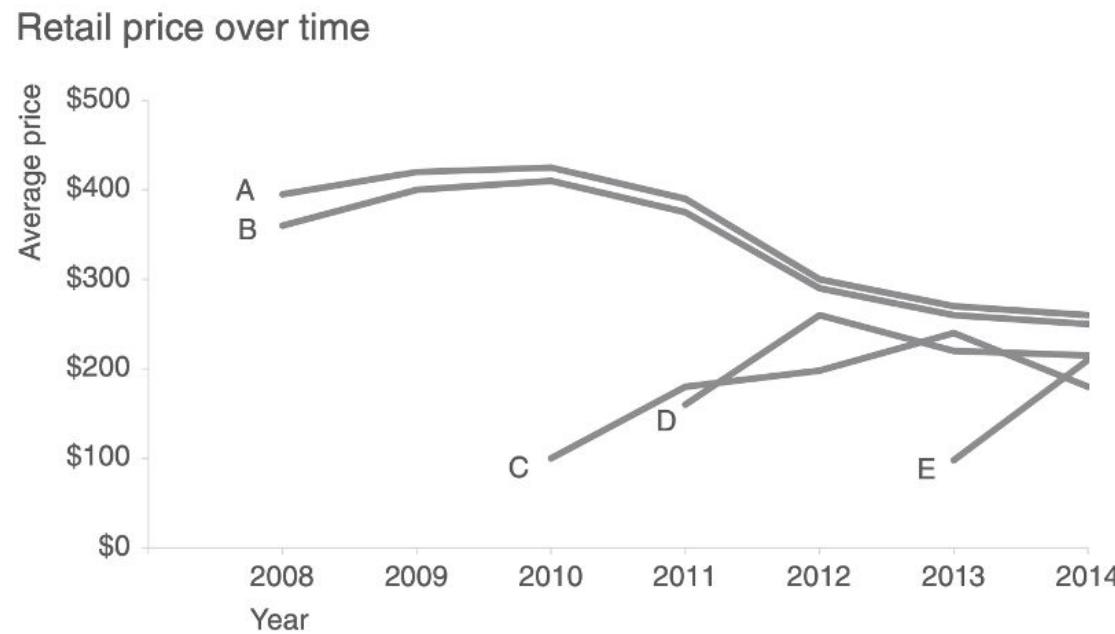
As of 2014, retail prices have converged across products, with an average retail price of \$223, ranging from a low of \$180 (Product C) to a high of \$260 (Product A)



Notice how preattentive attributes allow us to tell **different stories** with the same graph!

We've already been thinking like designers through this process; to complete our design, we now:

- Add axis titles
- Align elements
- Correct casing



In the next **5 minutes...**

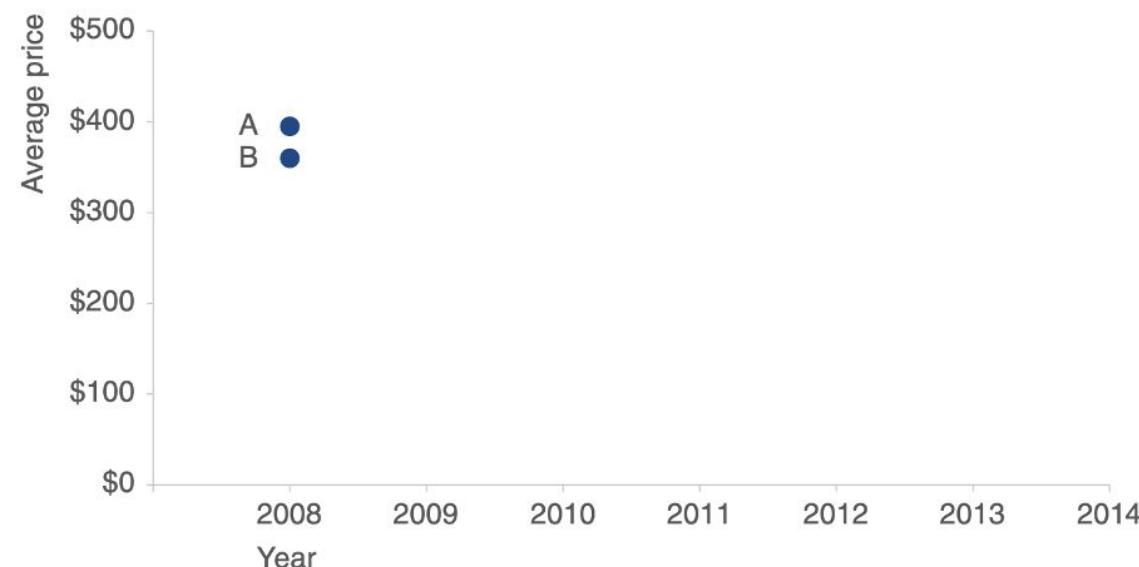
OUR GOAL:

- 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**.

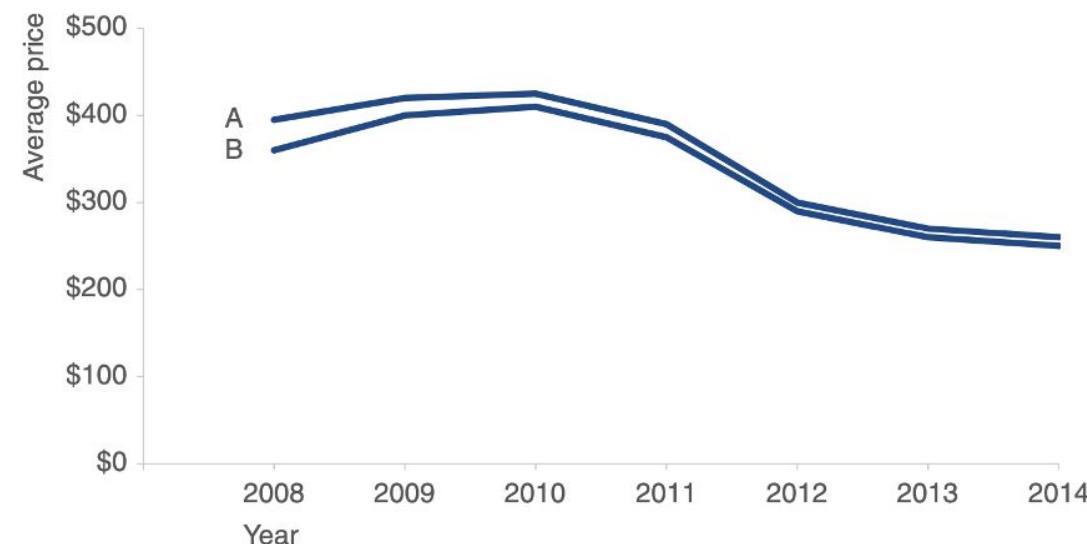
Products A and B were launched in 2008 at price points of **\$360+**

Retail price over time



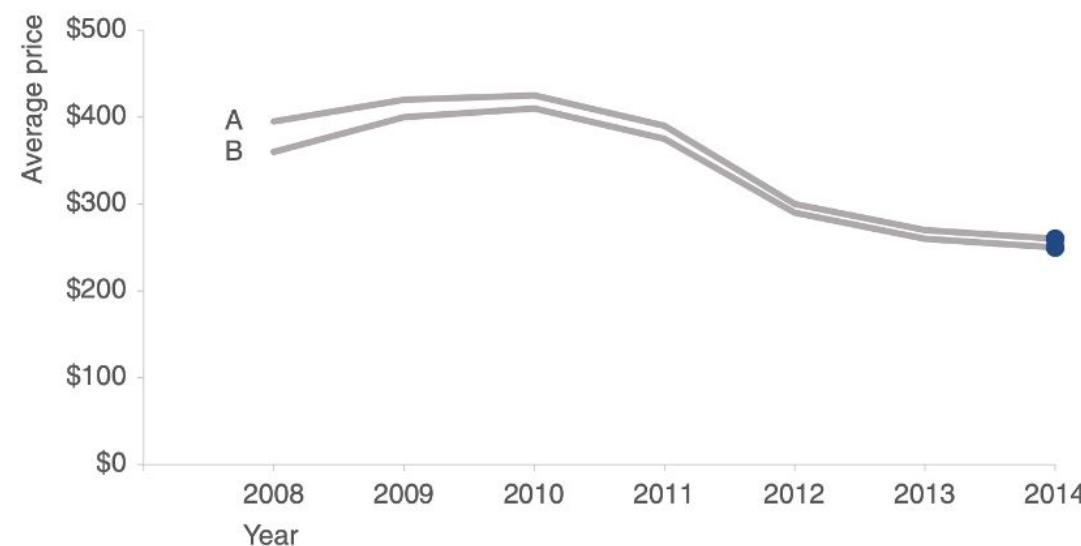
They have been priced similarly over time, with B consistently slightly lower than A

Retail price over time



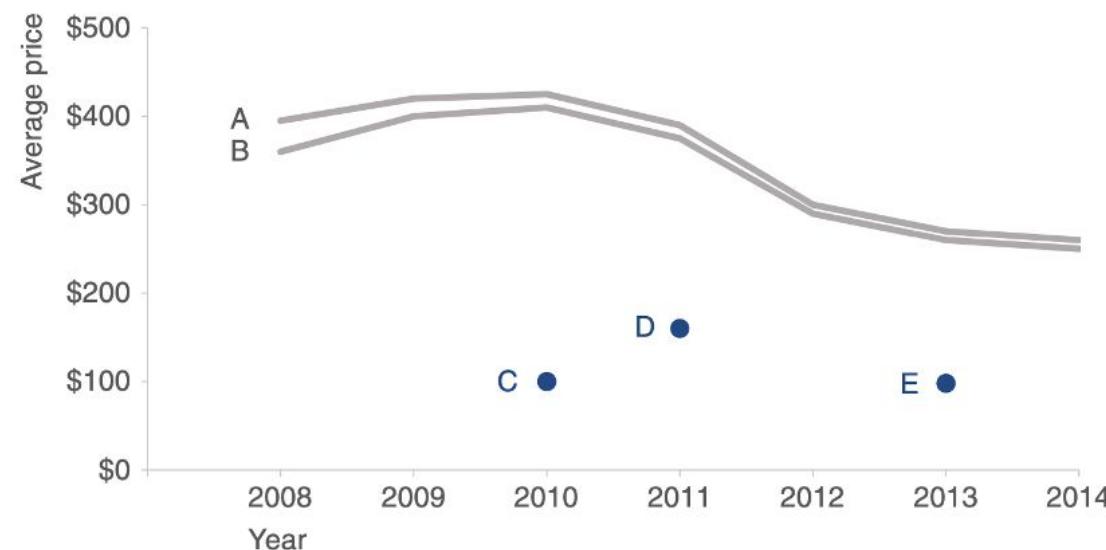
In 2014, Products A and B were priced at **\$260** and **\$250**, respectively

Retail price over time



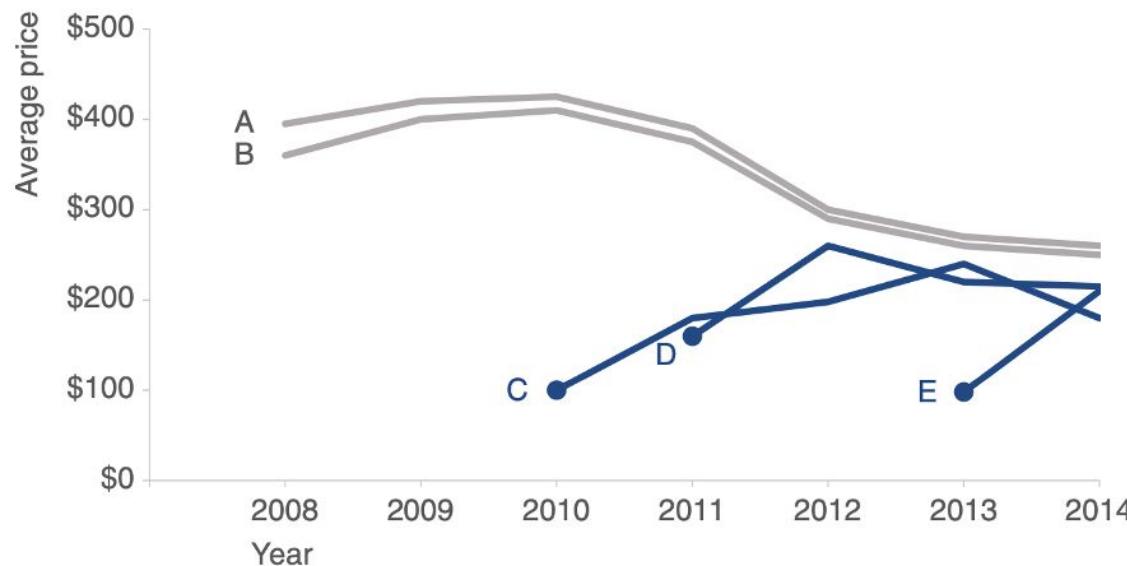
Products C, D, and E were each introduced later at **much lower price points...**

Retail price over time



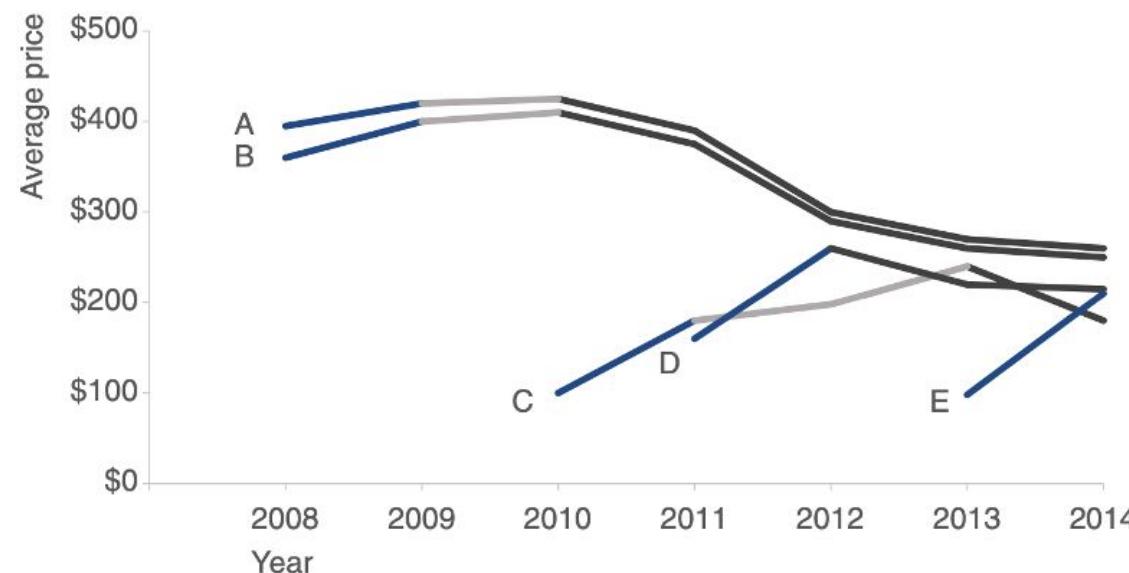
...but all have **increased in price** since their respective launches

Retail price over time



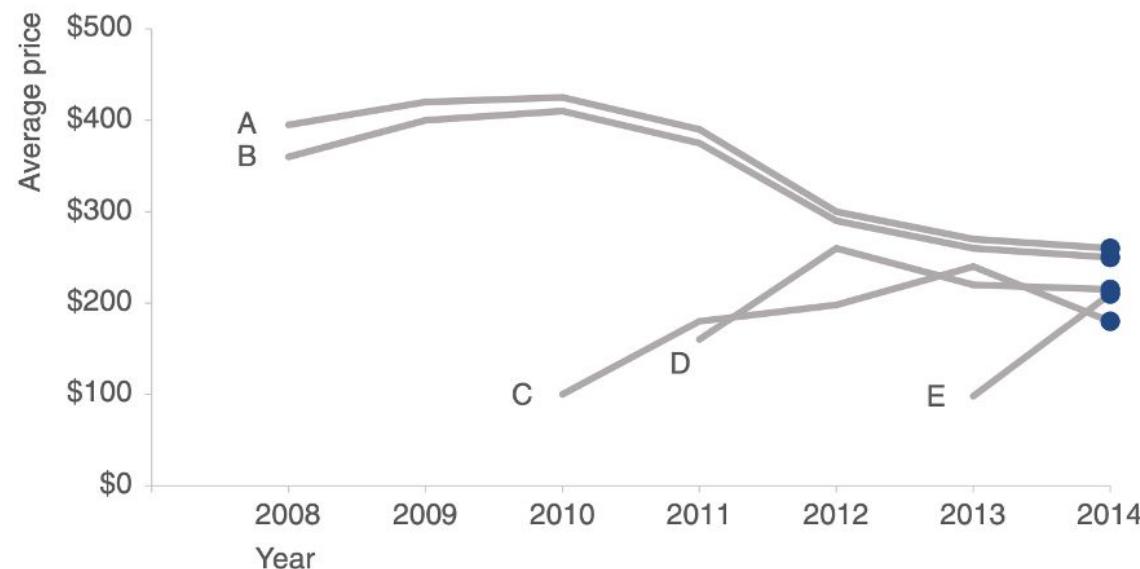
In fact, with the launch of a new product in this space, we tend to see an **initial price increase**, followed by a **decrease** over time

Retail price over time



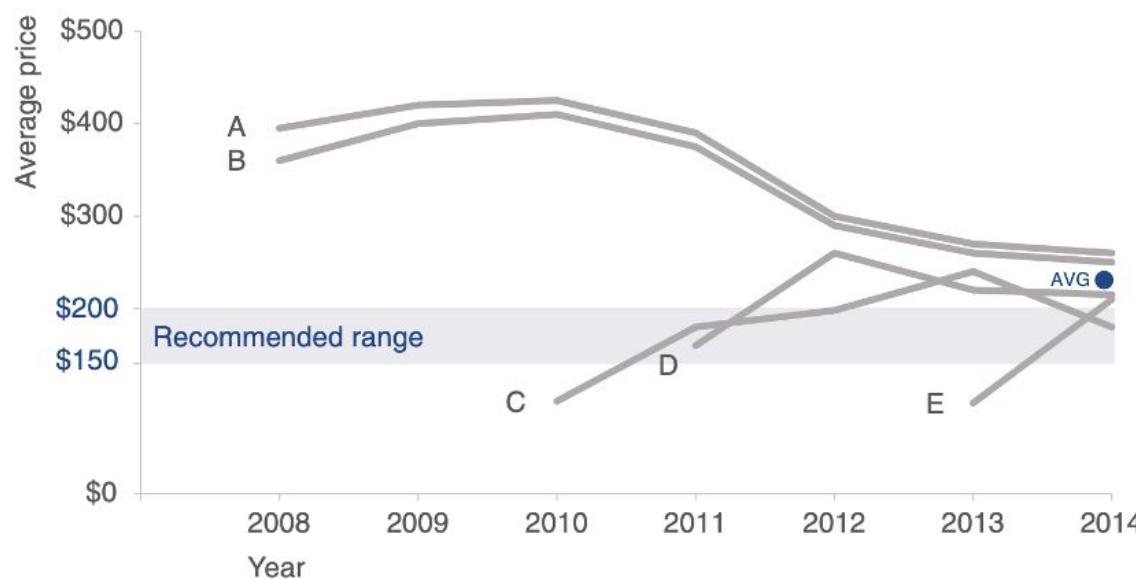
As of 2014, retail prices have converged, with an **average retail price of \$223**, ranging from a low of \$180 (C) to a high of \$260 (A)

Retail price over time



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

Retail price over time



- It is very **hard to generate effective graphs** and diagrams, especially when they are used within a presentation
- Being able to visualize data and tell stories with them is **key** in several domains, not just business ones
- There are **6 main points** that can mean the difference between failure and success when it comes to storytelling with data:
 1. Understand the context
 2. Choose an appropriate visual display
 3. Eliminate clutter
 4. Focus attention where you want it
 5. Think like a designer
 6. Tell a story

