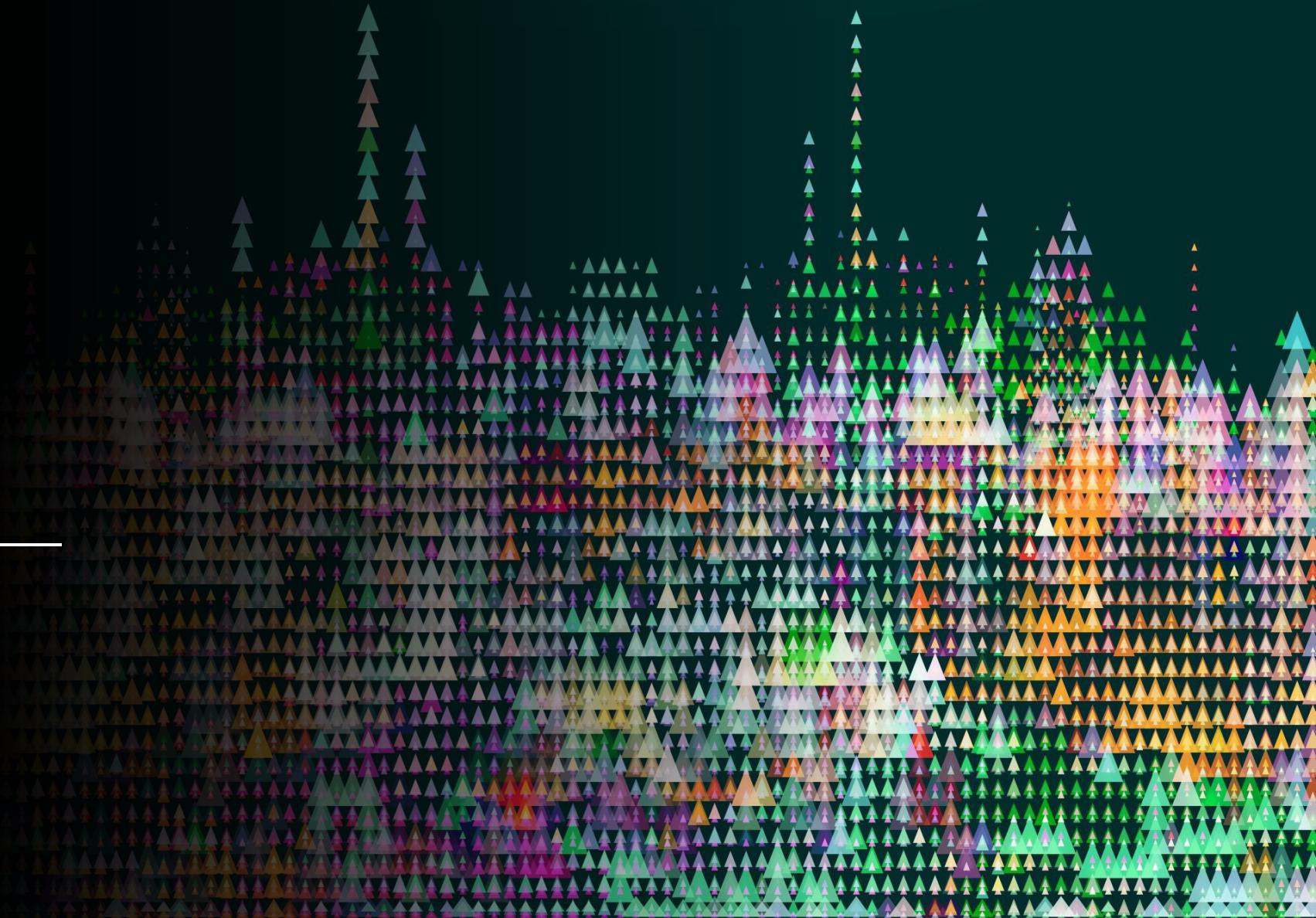




Descriptive Analytics through Data Storytelling

Raras Tyasnurita



Motivation



Bad graphs are everywhere, including in places where you would expect people to know better



Proficiency in dashboard tools is not enough, your tools do not know what is the story within your data

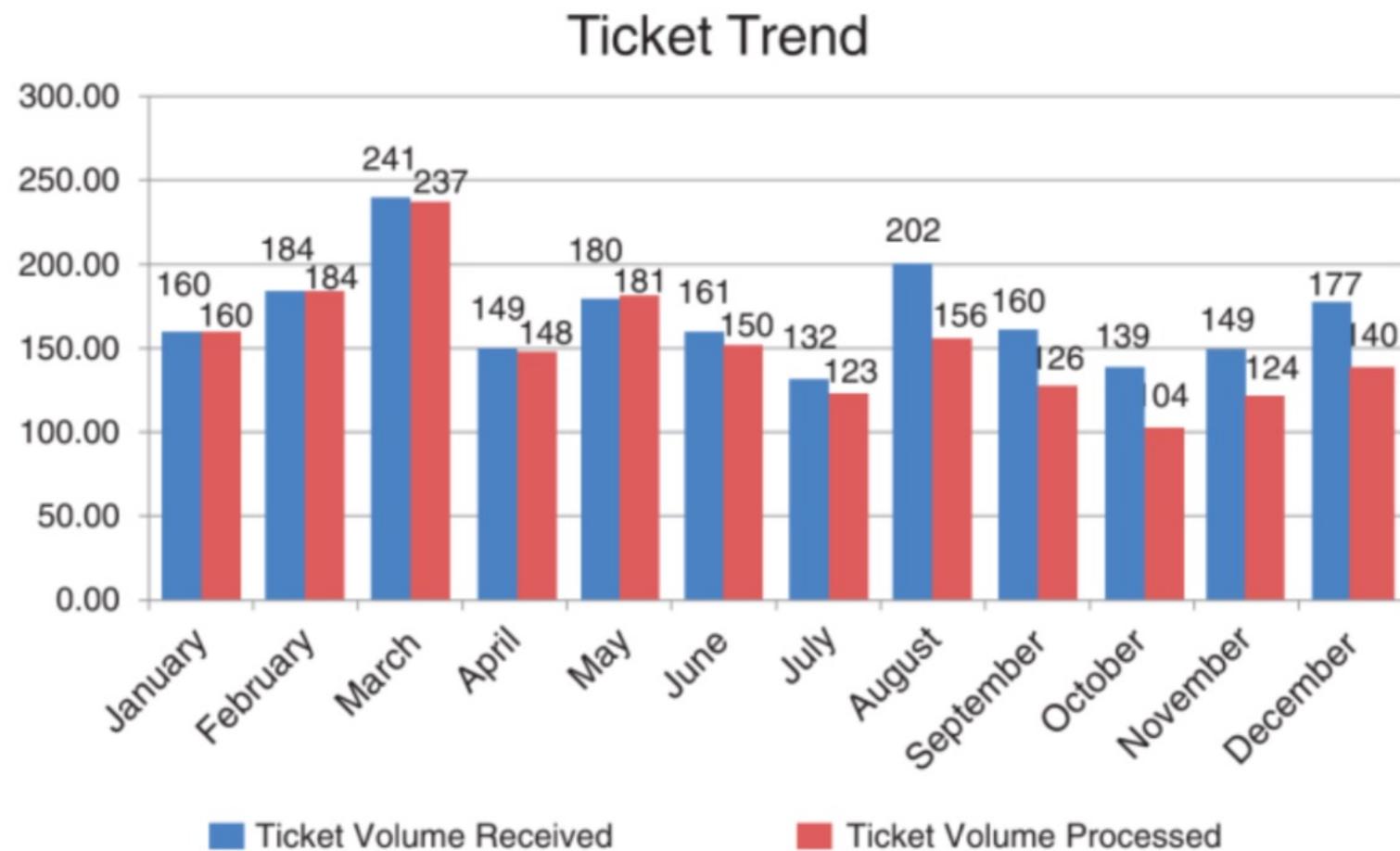


We learn a lot about language and math, but we are not naturally good in pairing them (tell stories with numbers)



We need an analyst or communicator of the information to bring the story visually and contextually to life

Example 1



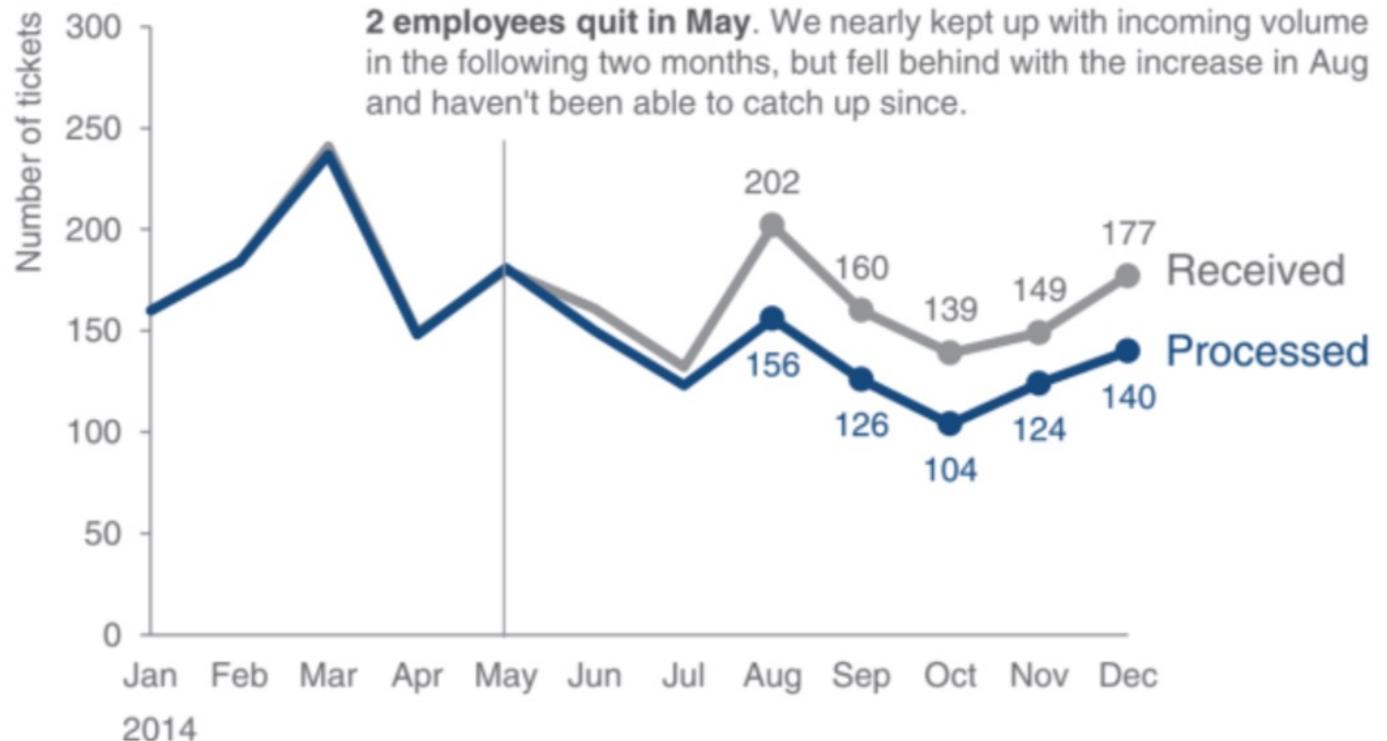
Example 1

*storytelling with
data instead of
showing data*

Please approve the hire of 2 FTEs

to backfill those who quit in the past year

Ticket volume over time

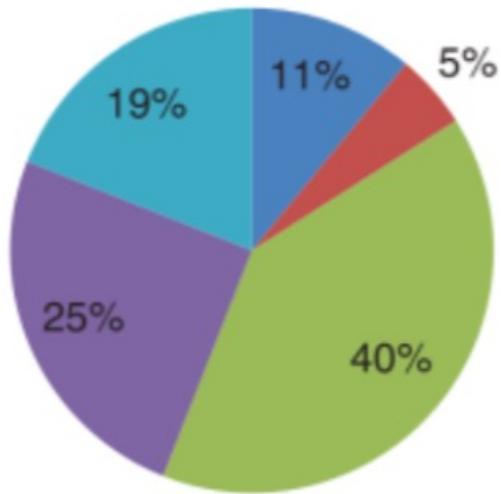


Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

Survey Results

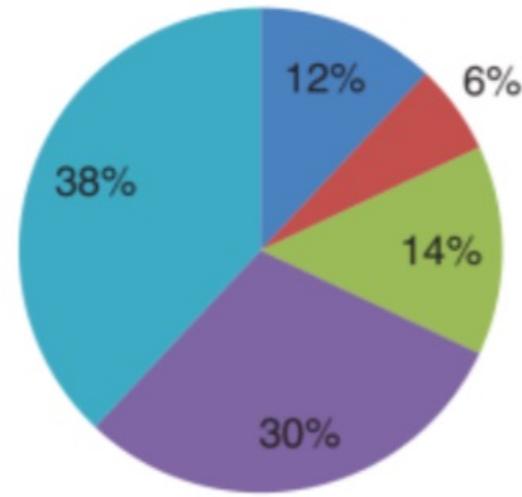
PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



Example 2

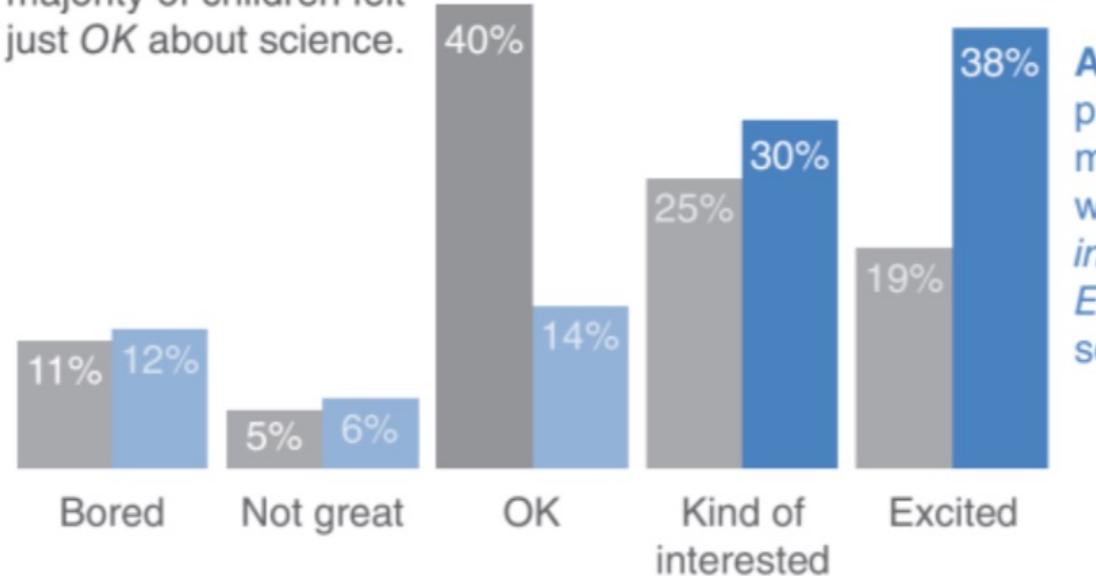
Example 2

*storytelling with
data instead of
showing data*

Pilot program was a success

How do you feel about science?

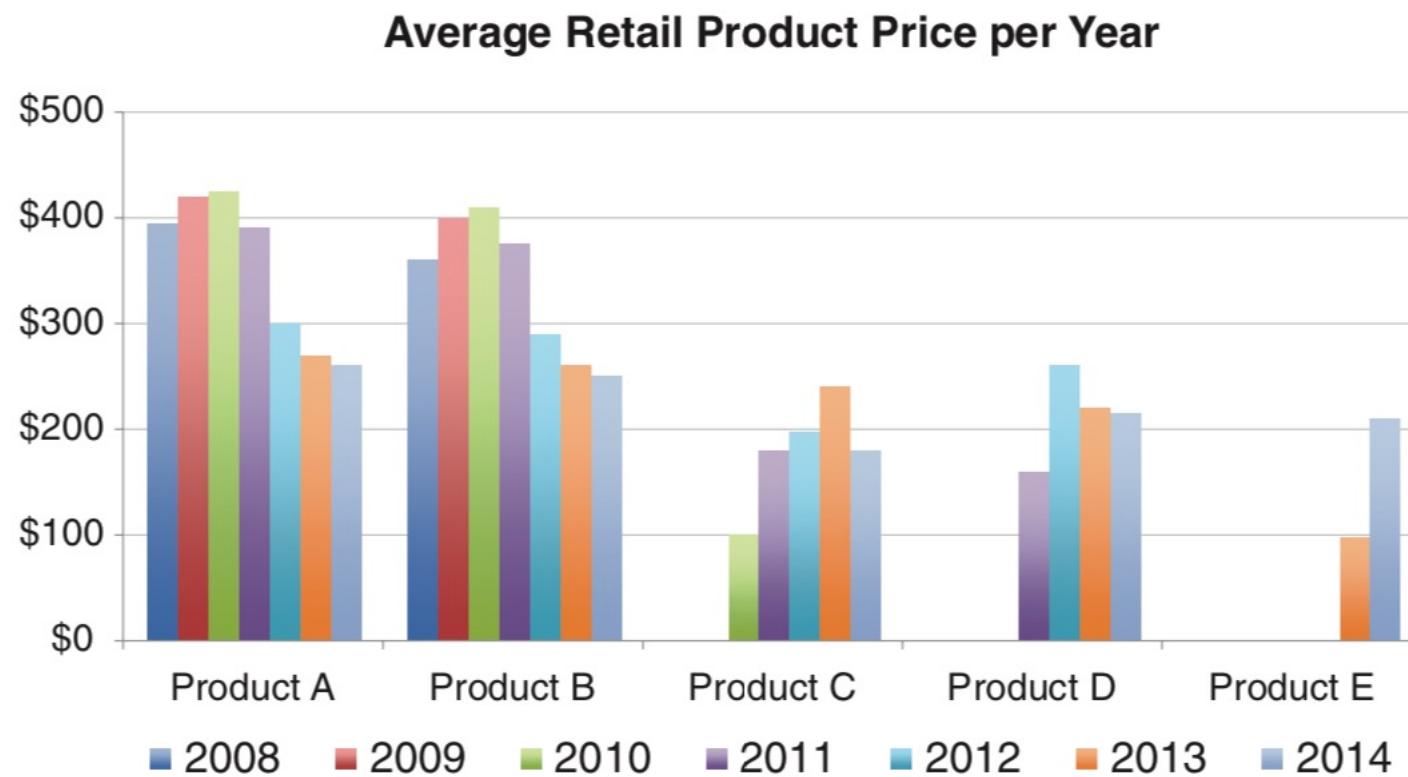
BEFORE program, the majority of children felt just *OK* about science.



AFTER program,
more children
were *Kind of
interested &
Excited* about
science.

Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

Example 3

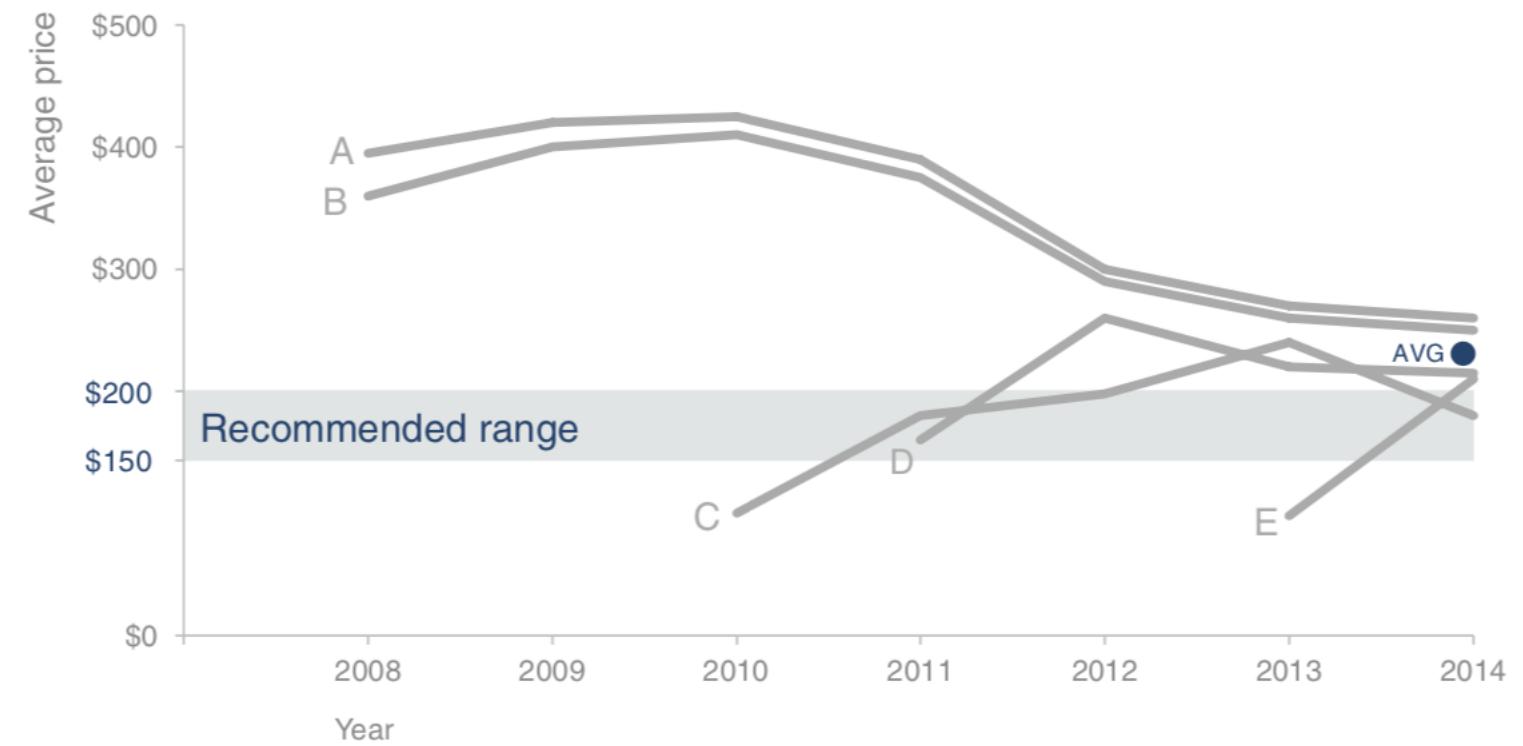


Example 3

*storytelling with
data instead of
showing data*

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

Retail price over time by product



How to tell stories with data: 6 lessons

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

Choose an effective visual

Simple text

Tables Heatmap

Points (Scatterplot)

Lines

Slopegraph

Bars Vertical & Horizontal bar chart Stacked bar chart Waterfall chart

Area

Tips: Specific graph types and elements you should avoid

Simple Text

The fact that you have some numbers does not mean that you need a graph!

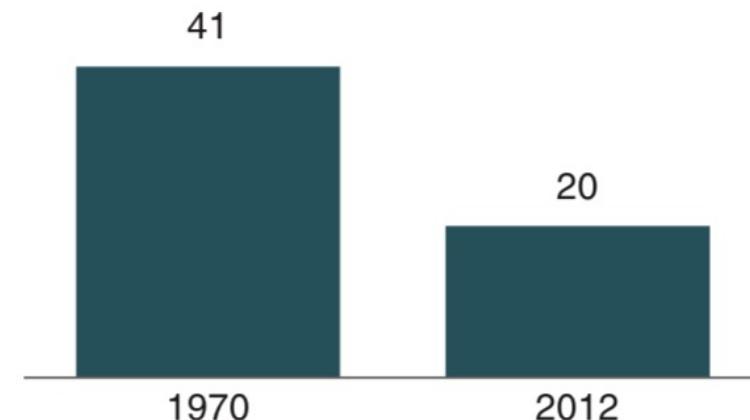
When you have a number or two that you want to communicate, think about solely using the number—making it as prominent as possible—and a few supporting words to clearly make your point.

20%

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

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

Tables

The data should be what stands out, not the borders.

Using a table in a live presentation is rarely a good idea ,
consider including the full table in the appendix and a link or reference.

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

Heatmap

A heatmap is a way to visualize data in tabular format, where in place of (or in addition to) the numbers, you leverage colored cells that convey the relative magnitude of the numbers.

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

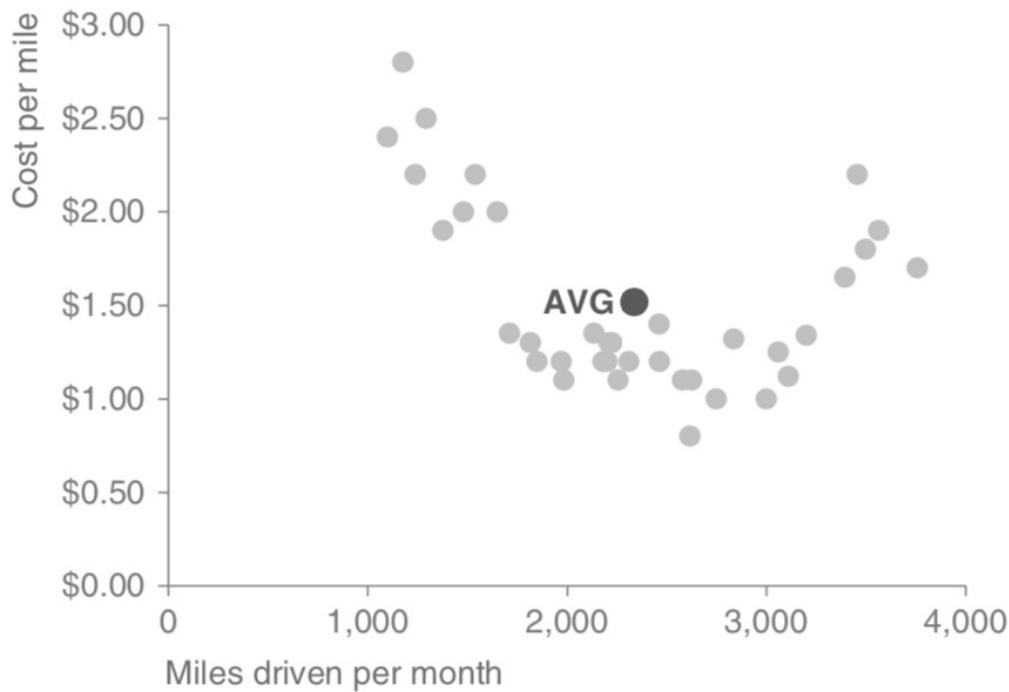
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

the lowest number (11%) and highest number (58%)

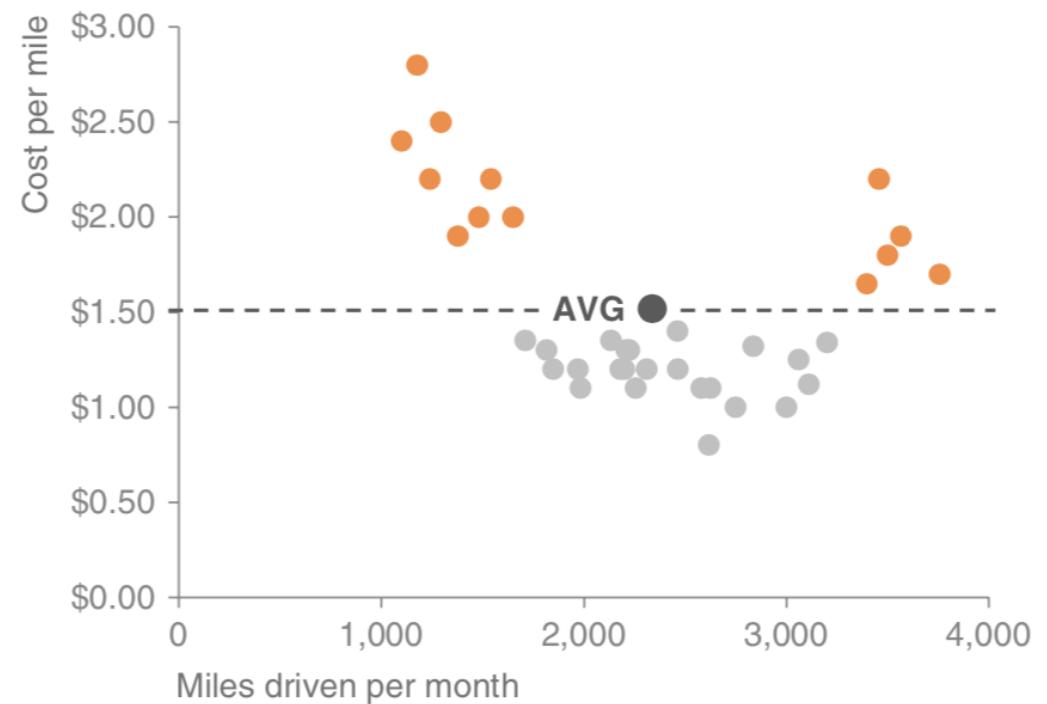
Scatterplot

“chart” is the broader category, with “graphs” being one of the subtypes (other chart types include maps and diagrams)

Cost per mile by miles driven

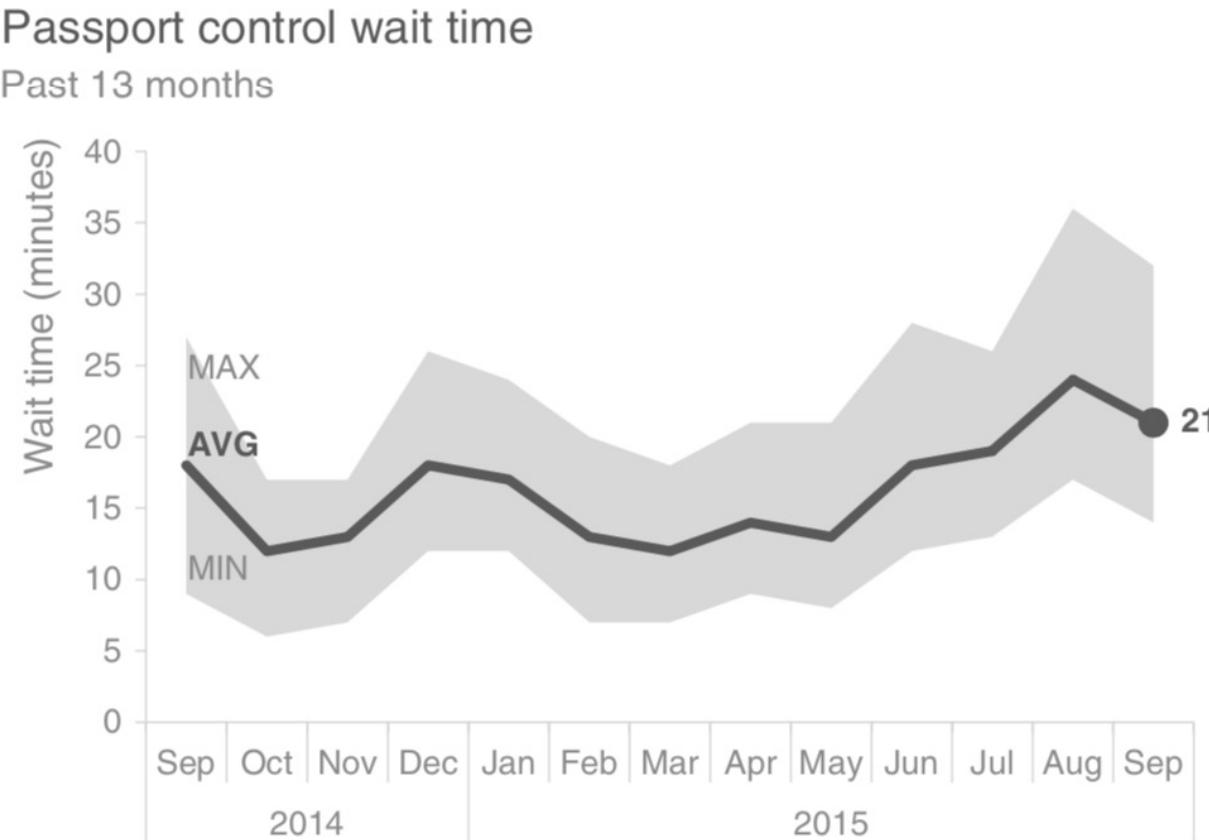


Cost per mile by miles driven



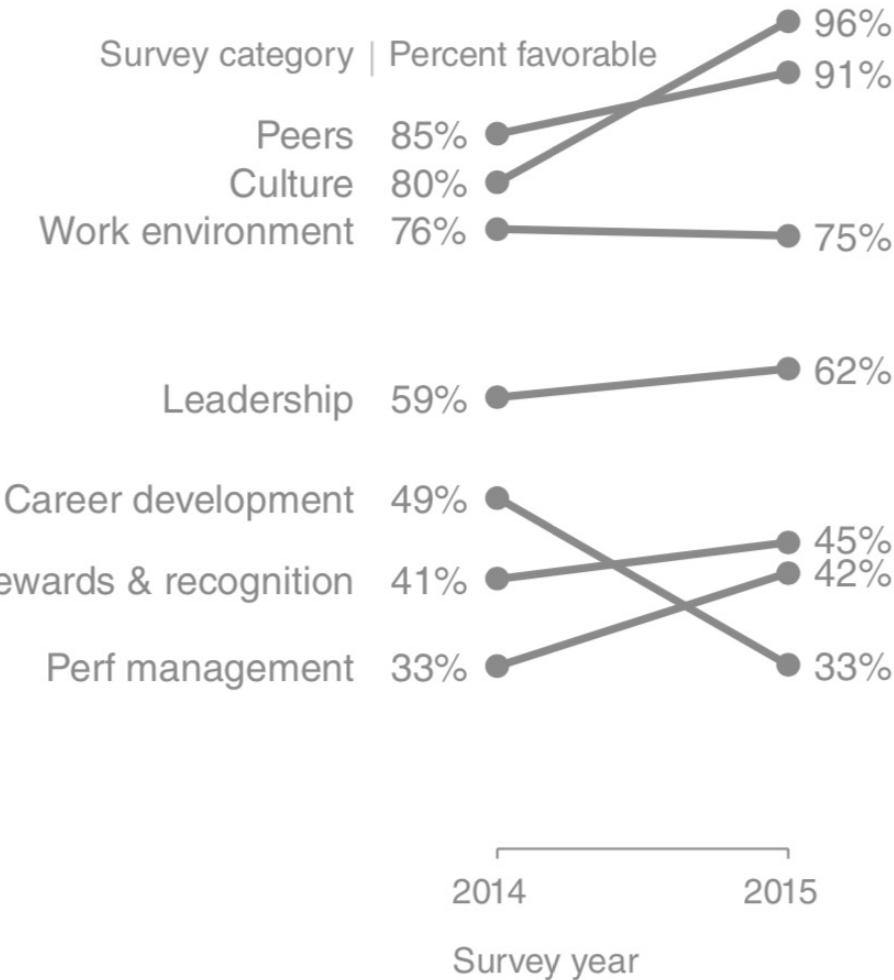
Lines

The line graph can show a single series of data, two series of data, or multiple series. Note that when you're graphing time on the horizontal x-axis of a line graph, the data plotted must be in consistent intervals (daily, weekly, monthly, yearly).

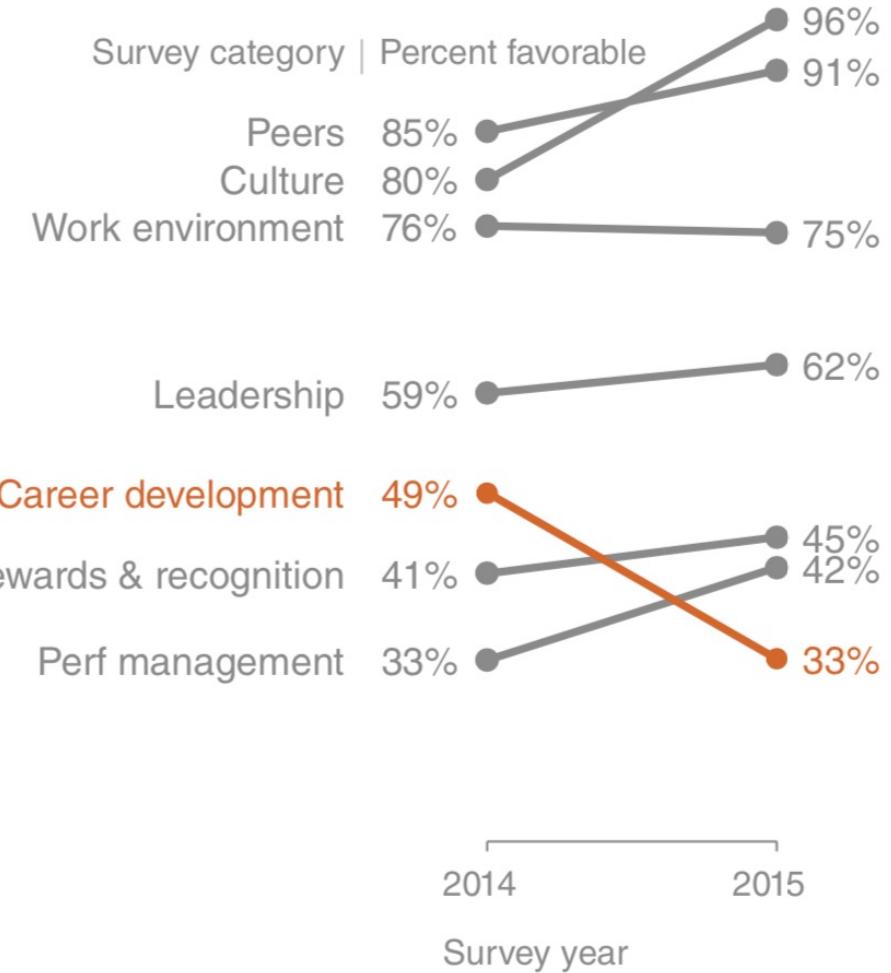


Slopegraph

Employee feedback over time



Employee feedback over time



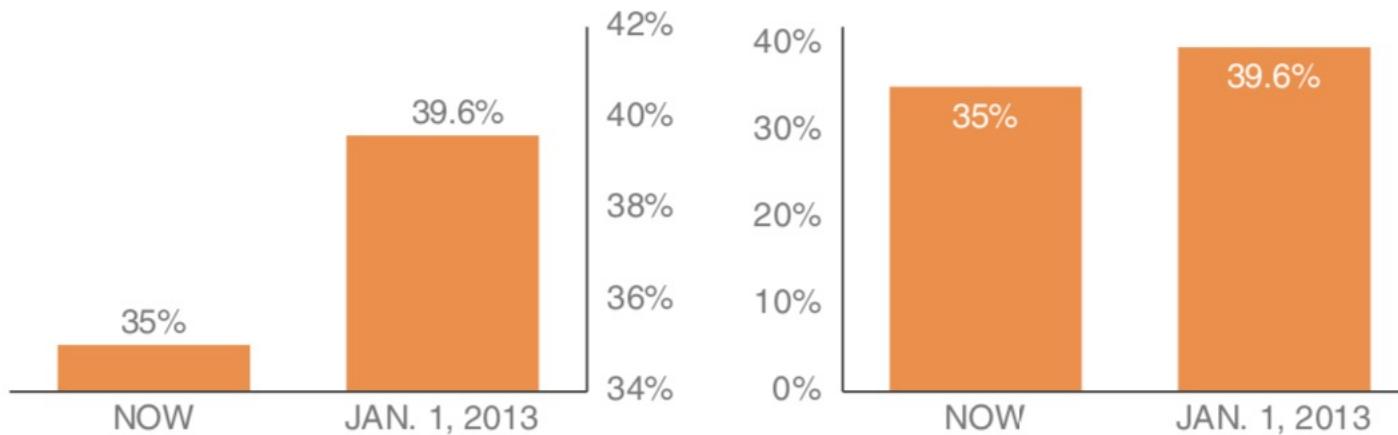
Bars

Sometimes bar charts are avoided *because they are common*. This is a mistake. Rather, bar charts should be leveraged, as this means less of a learning curve for your audience.

Bar charts must have a zero baseline

Non-zero baseline: as originally graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

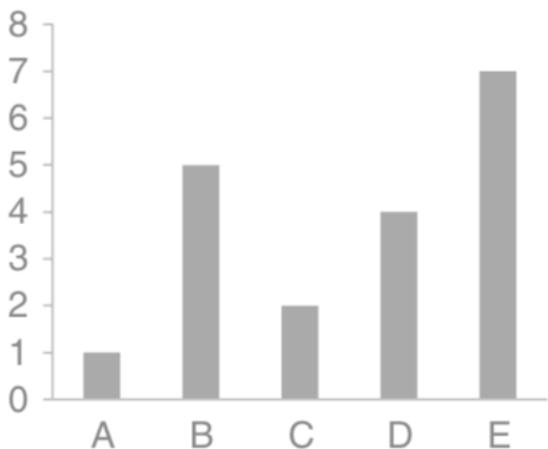


the visual increase is 460% (the heights of the bars are $35 - 34 = 1$ and $39.6 - 34 = 5.6$, so $(5.6 - 1)/1 = 460\%$)

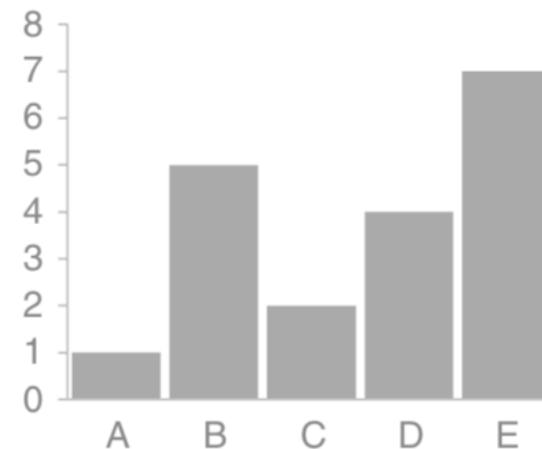
an actual visual increase is 13% from $((39.6 - 35) / 35)$

Width of the bar

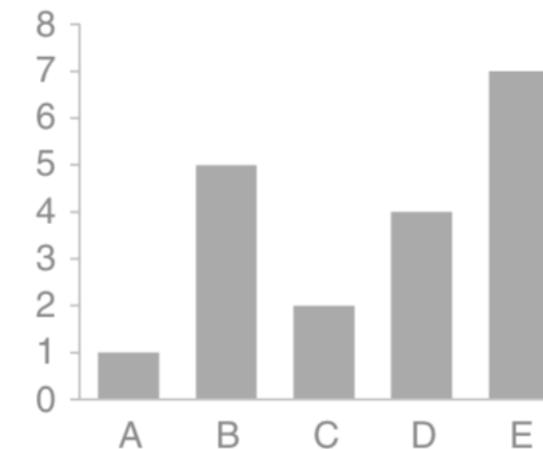
Too thin



Too thick



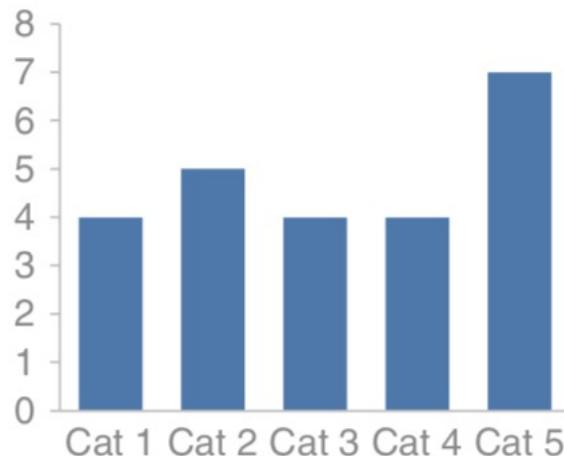
Just right



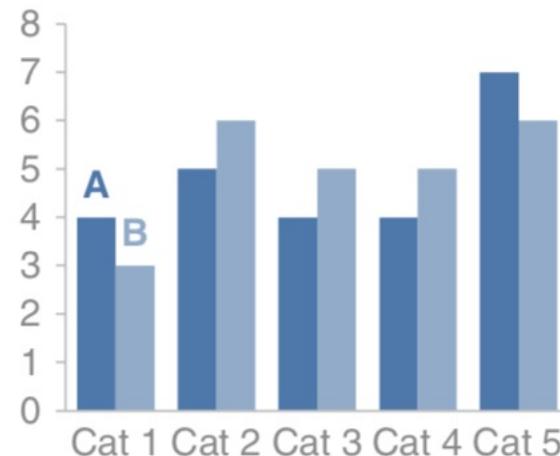
Vertical bar chart

Note that as you add more series of data, it becomes more difficult to focus on one at a time and pull out insight, so use multiple series bar charts with caution.

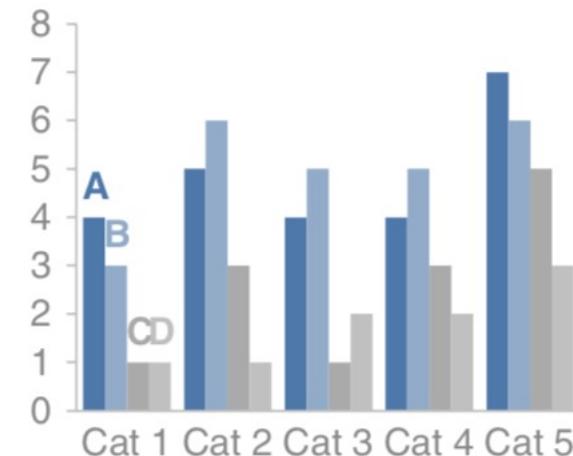
Single series



Two series



Multiple series



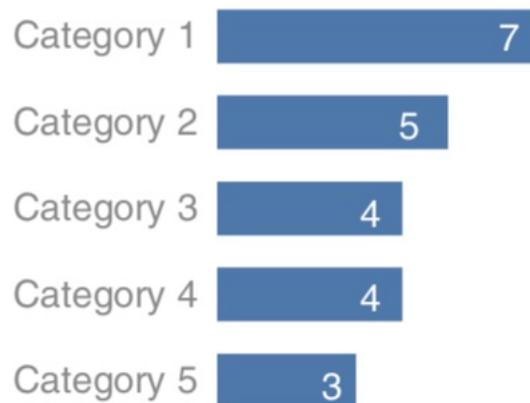
Horizontal bar chart

If we had to pick a single go-to graph for categorical data, it would be the horizontal bar chart.

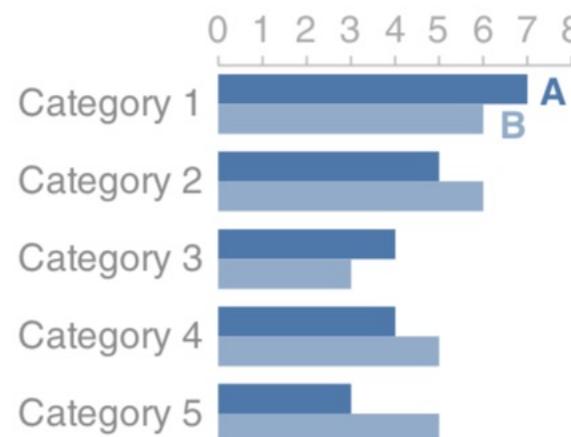
The horizontal bar chart is especially useful if your **category names are long**,

The structure of the horizontal bar chart is such that our eyes hit the category names before the actual data. This means by the time we get to the data, we already know what it represents.

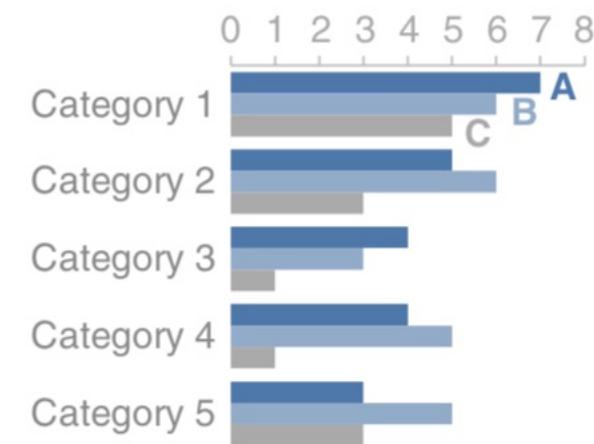
Single series



Two series

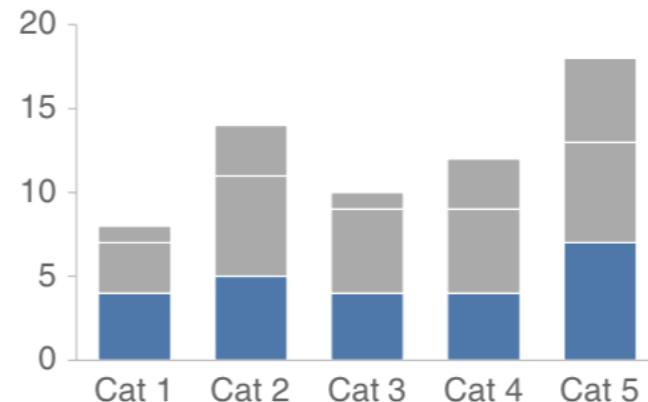


Multiple series

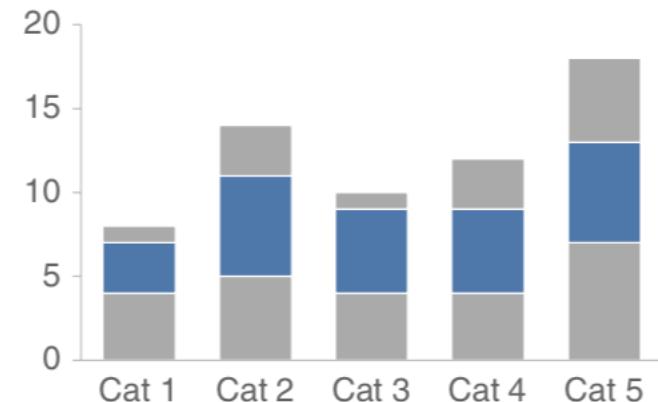


Stacked vertical bar chart

Comparing **these** is easy

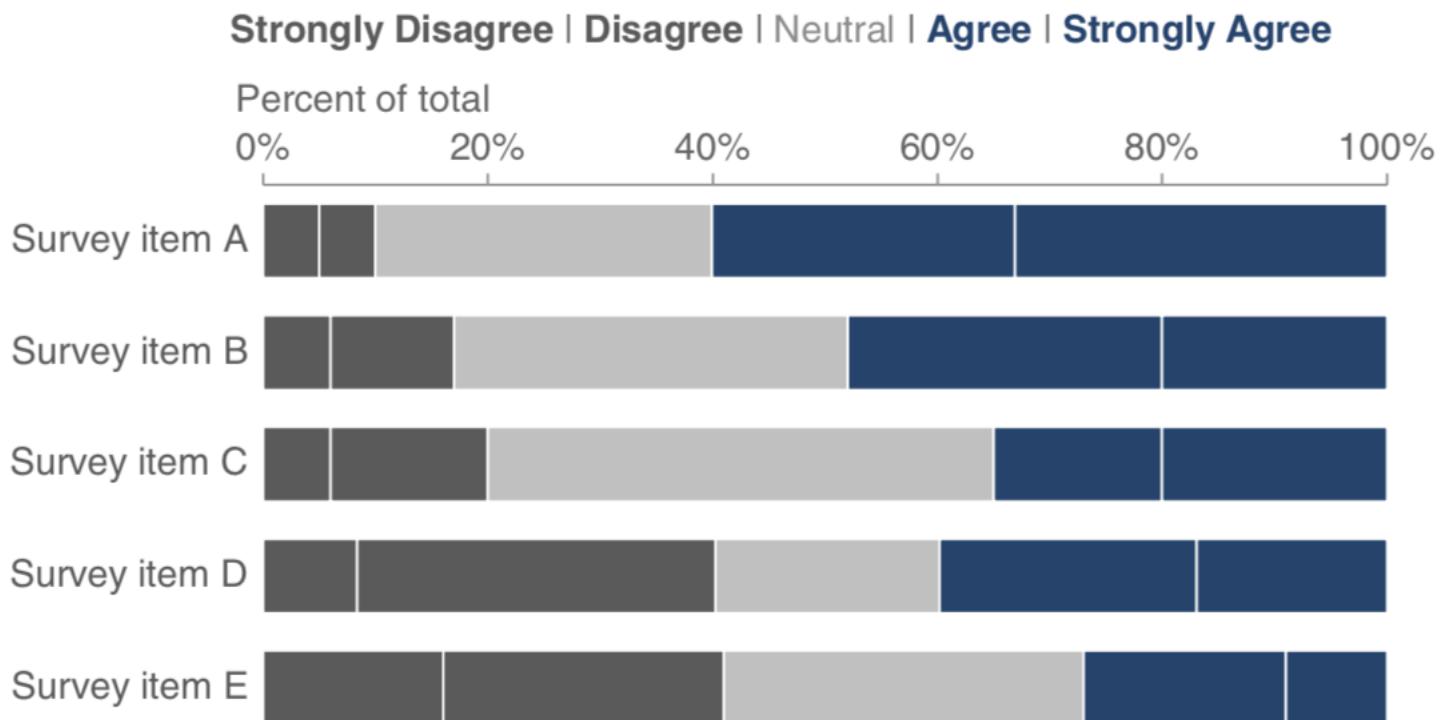


Comparing **these** is hard



Stacked horizontal bar chart

Survey results

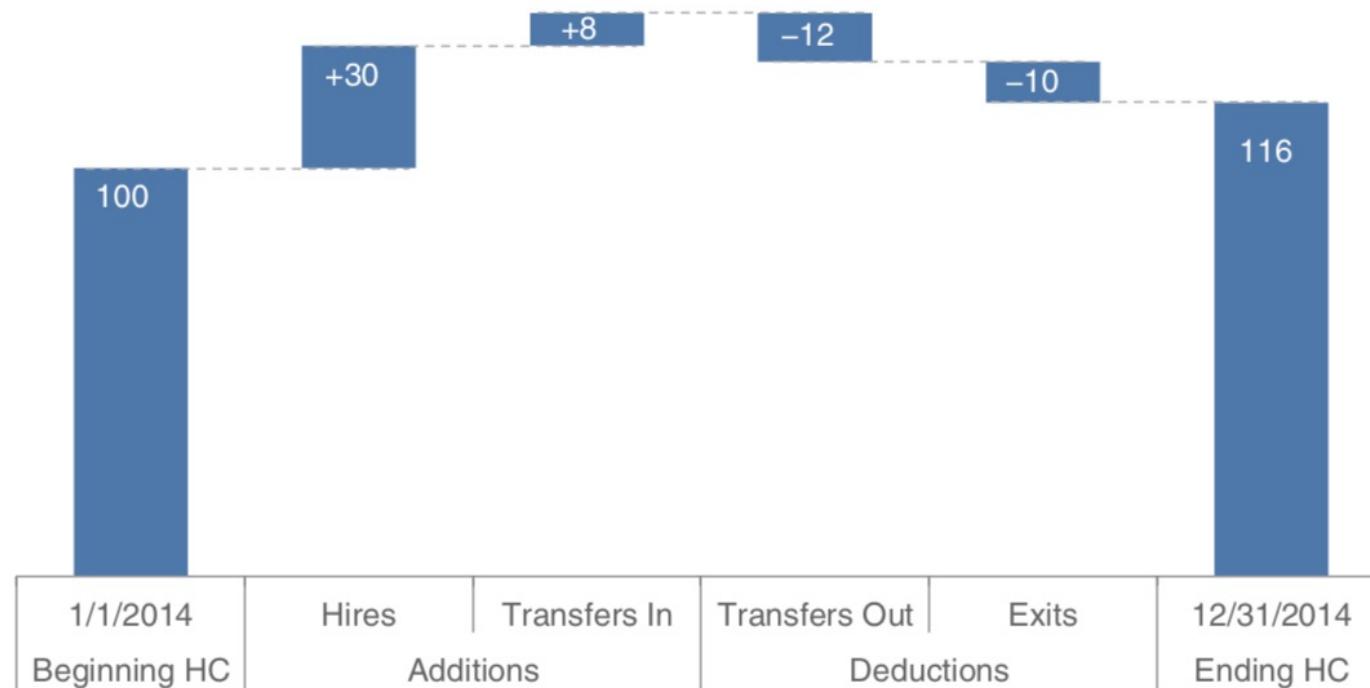


Waterfall chart

The waterfall chart can be used to pull apart the pieces of a stacked bar chart to focus on one at a time, or to show a starting point, increases and decreases, and the resulting ending point.

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.

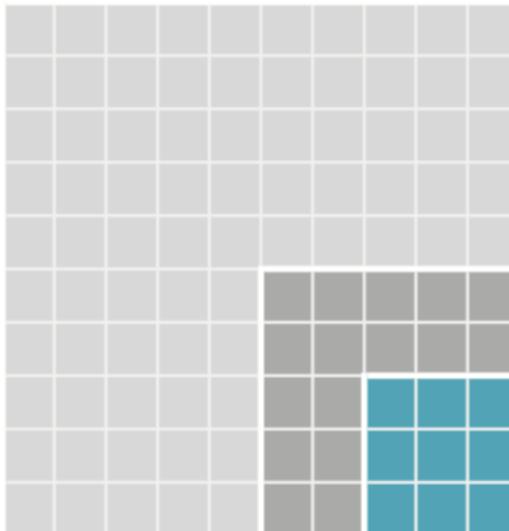


Area

Use it when you need to visualize numbers of vastly different magnitudes.

Typically avoid. Humans' eyes don't do a great job of attributing quantitative value to two-dimensional space, which can render area graphs harder to read than some of the other types of visual displays.

Interview breakdown



Out of every 100
phone screens...

we bring **25**
candidates onsite
for interviews...

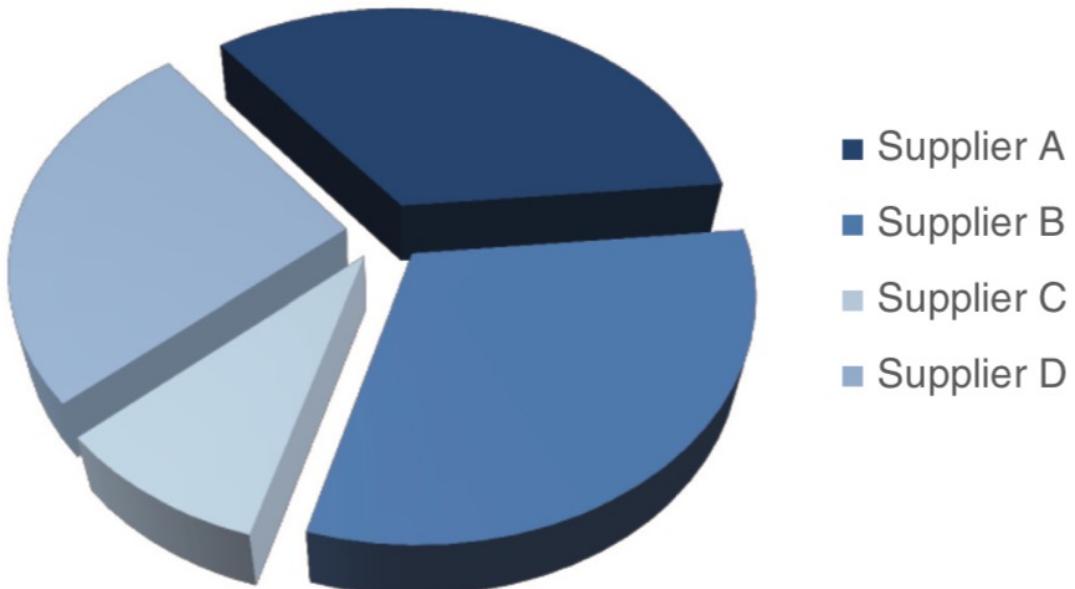
and
extend 9 offers.

TO BE
AVOIDED

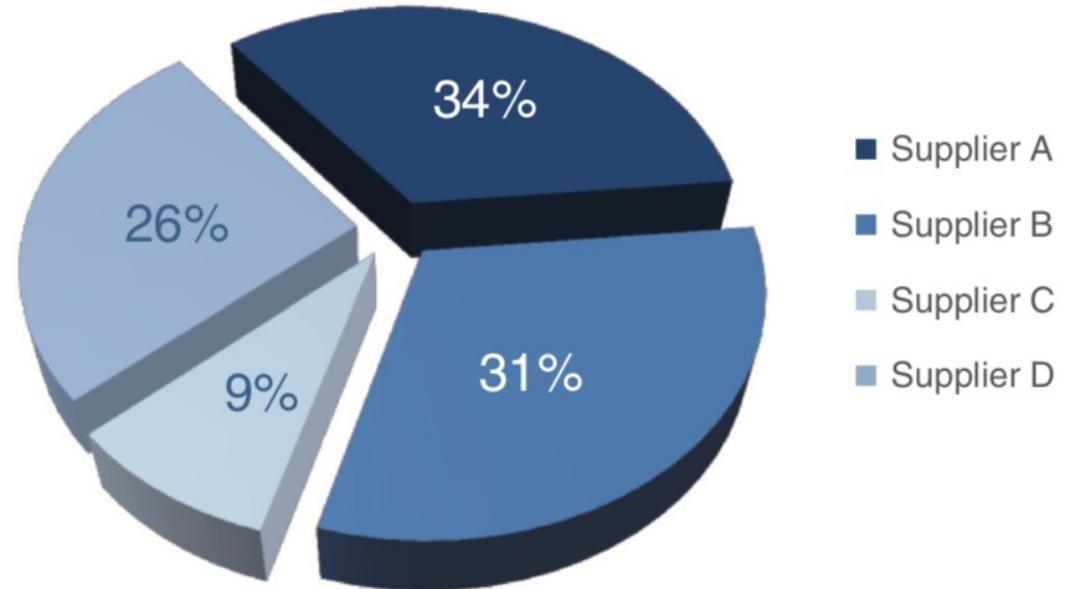


Pie charts

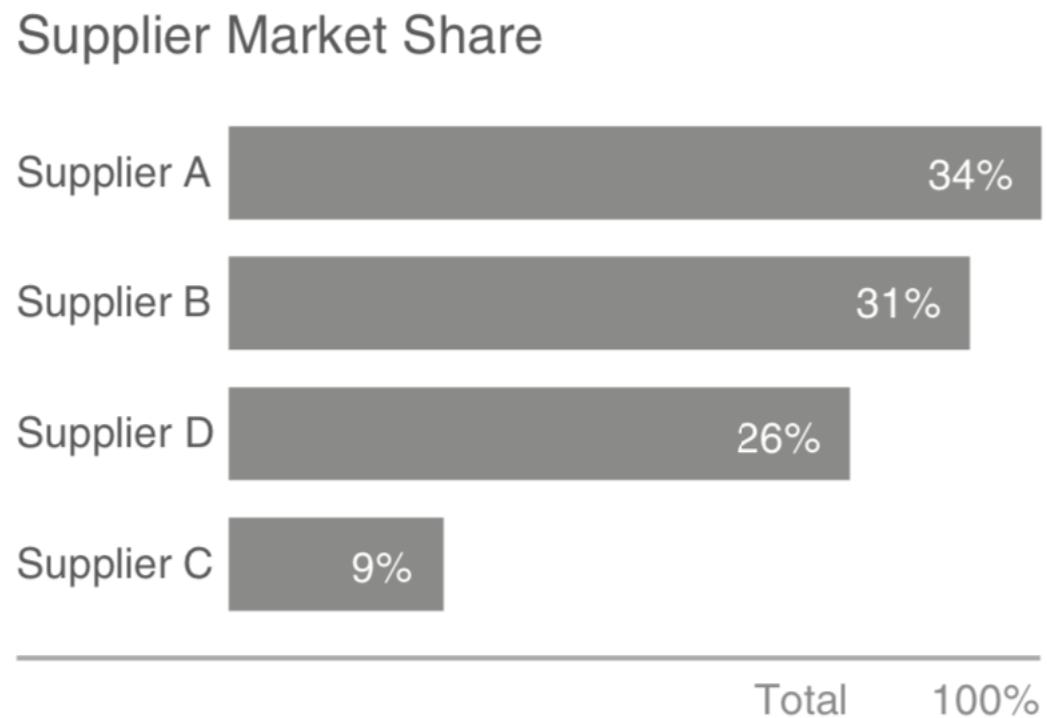
Supplier Market Share



Supplier Market Share

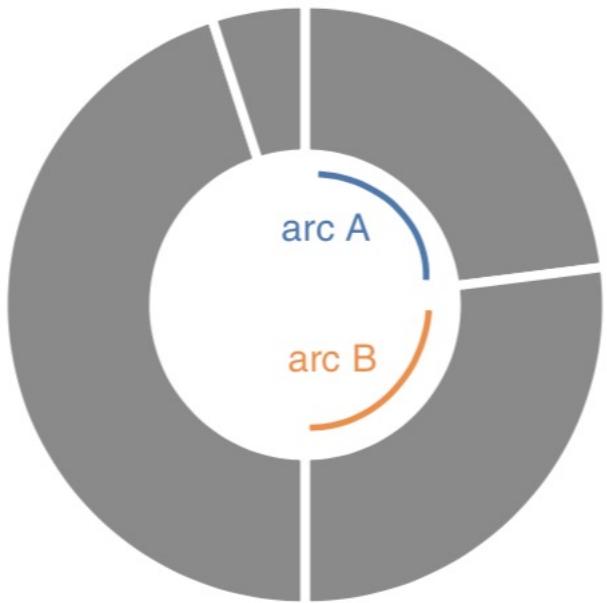


Alternative to pie chart

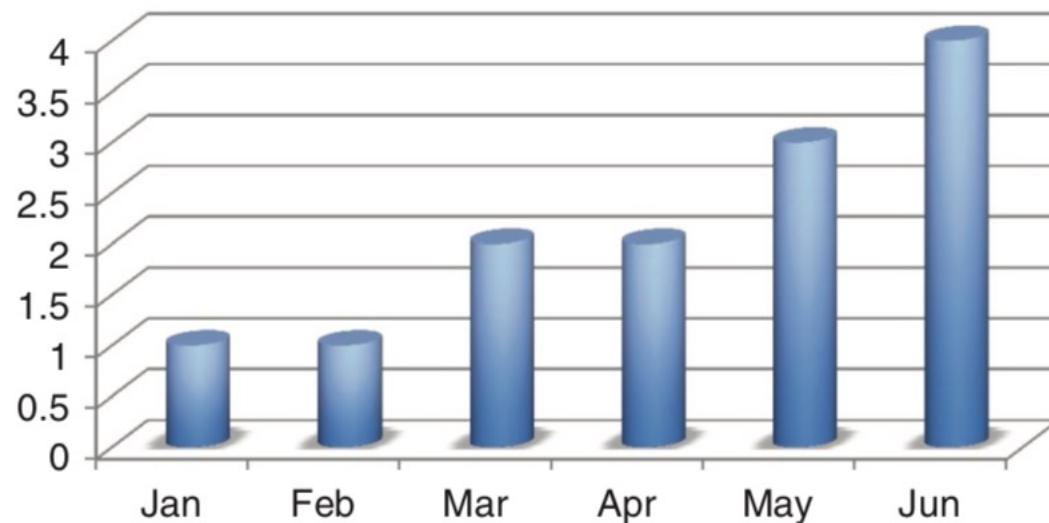


Donut and 3D chart

The donut chart



Number of issues



Secondary y-axis

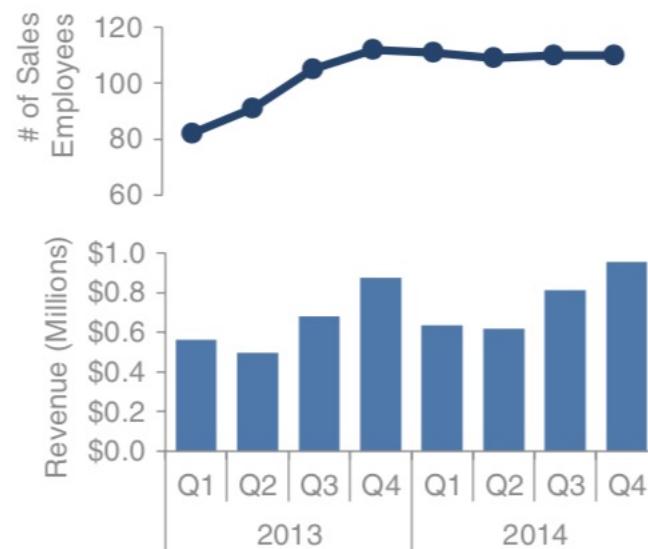


Alternative to secondary y-axis

Alternative 1: label directly



Alternative 2: pull apart vertically

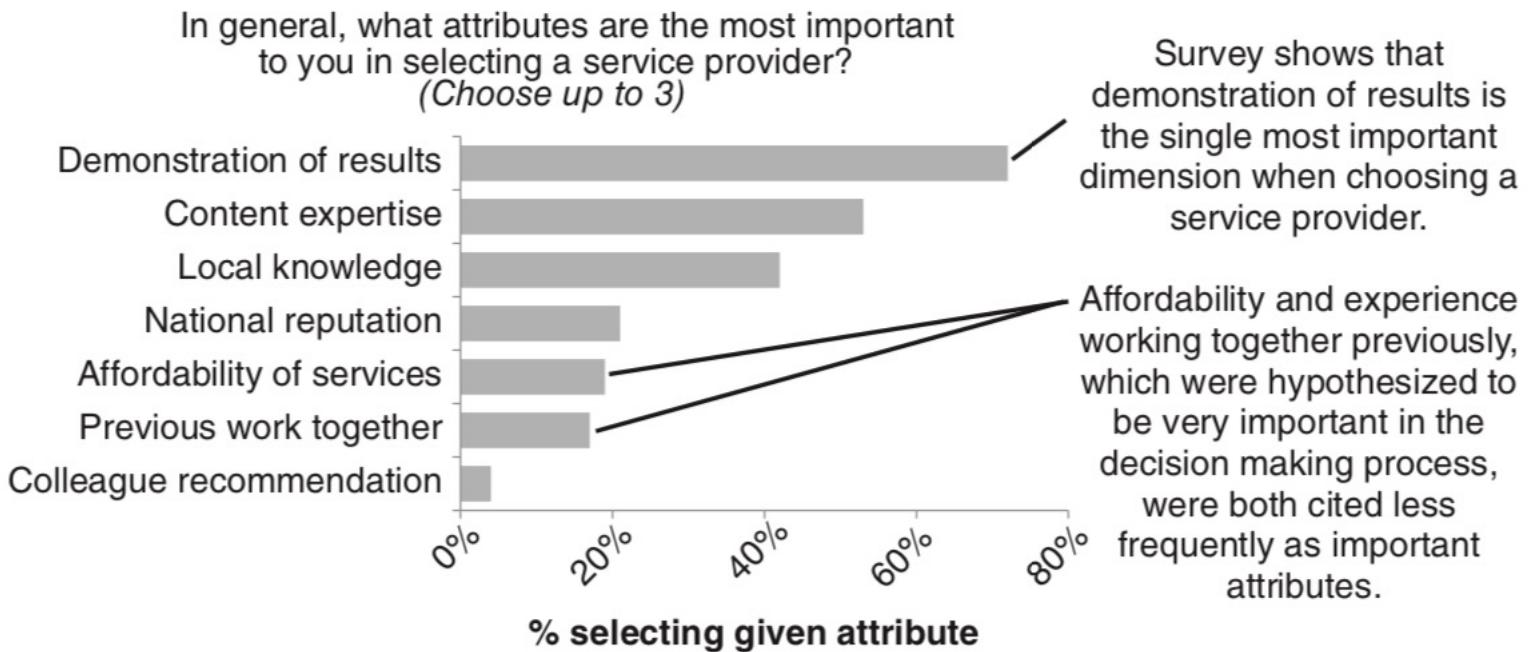




Clutter is an enemy

Lack of visual order

Demonstrating effectiveness is most important consideration when selecting a provider



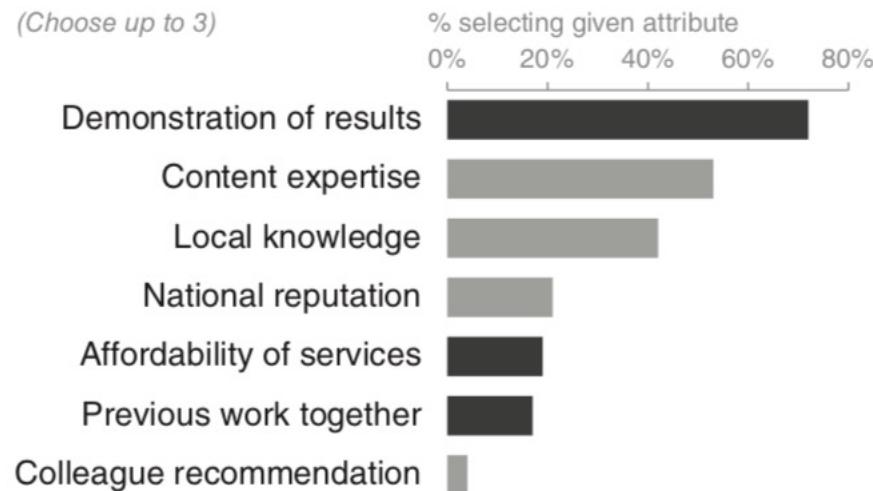
Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

Revamped summary of survey feedback

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.

Non-strategic use of contrast



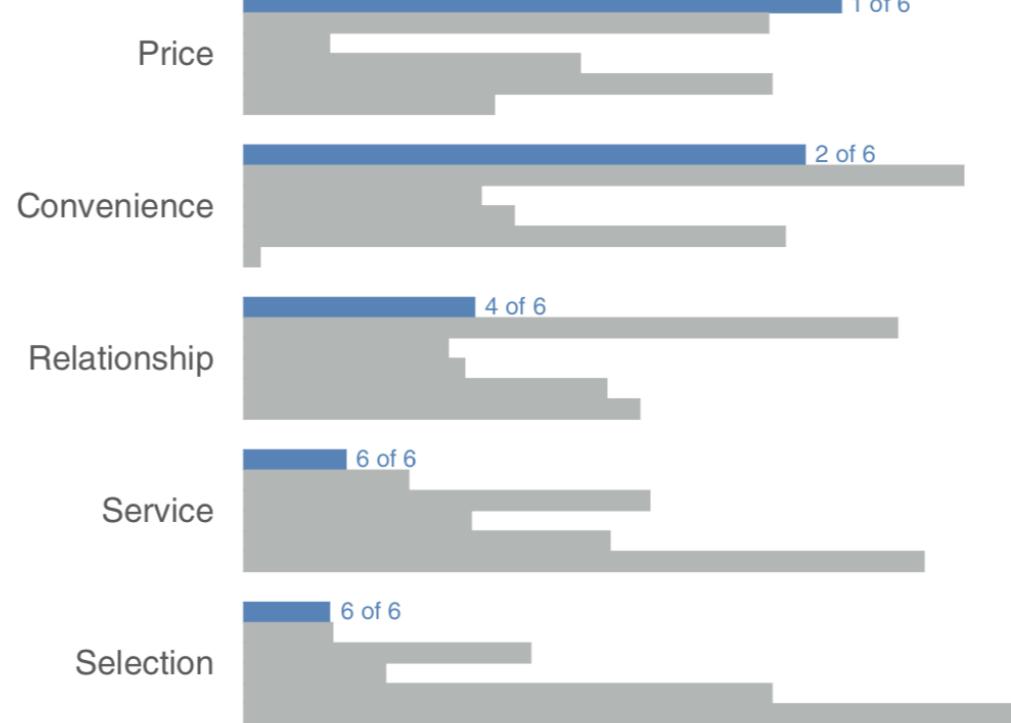
Using contrast strategically

Performance overview

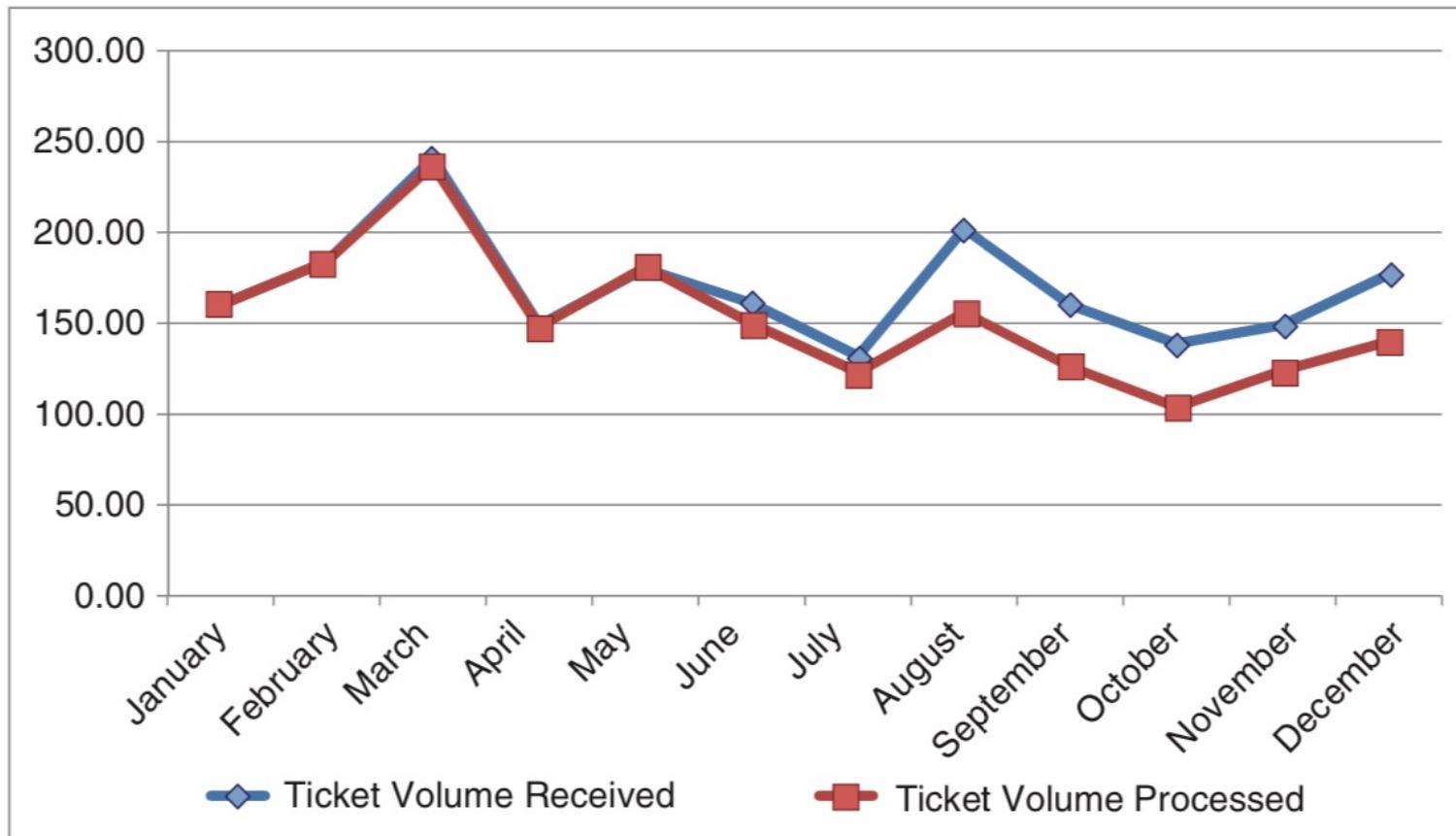
■ Our business

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

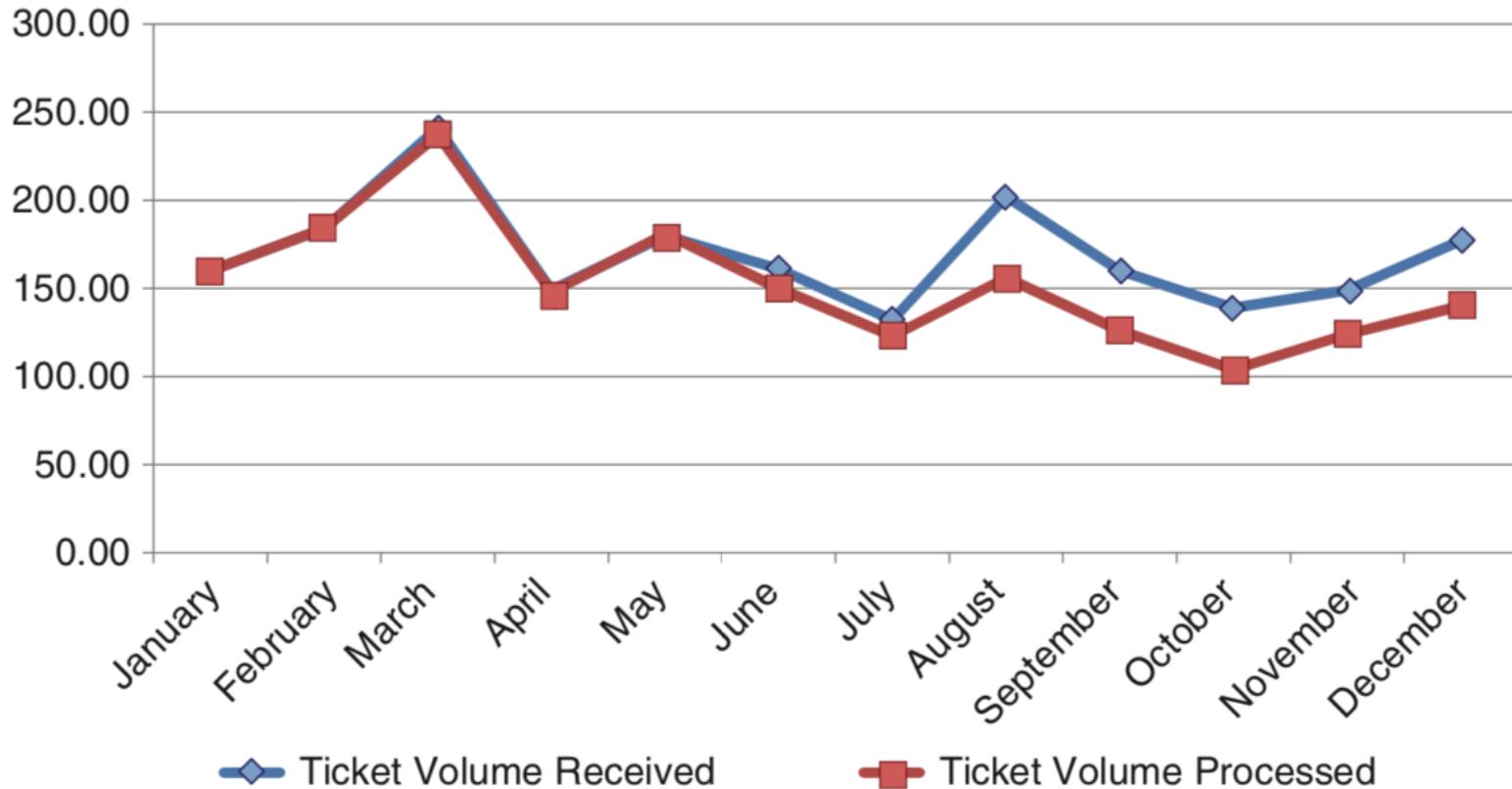
Weighted performance index | relative rank



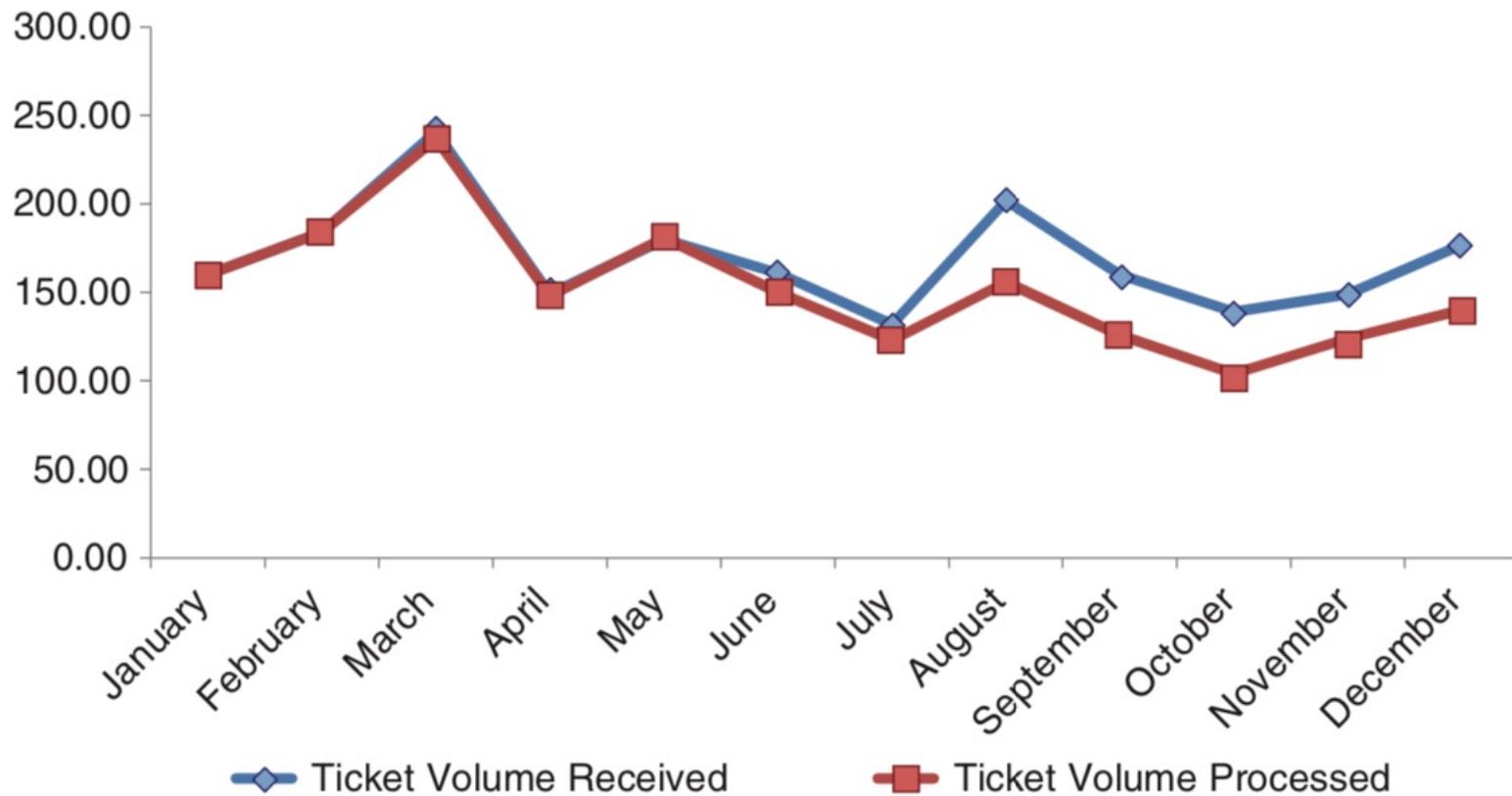
Decluttering step by step



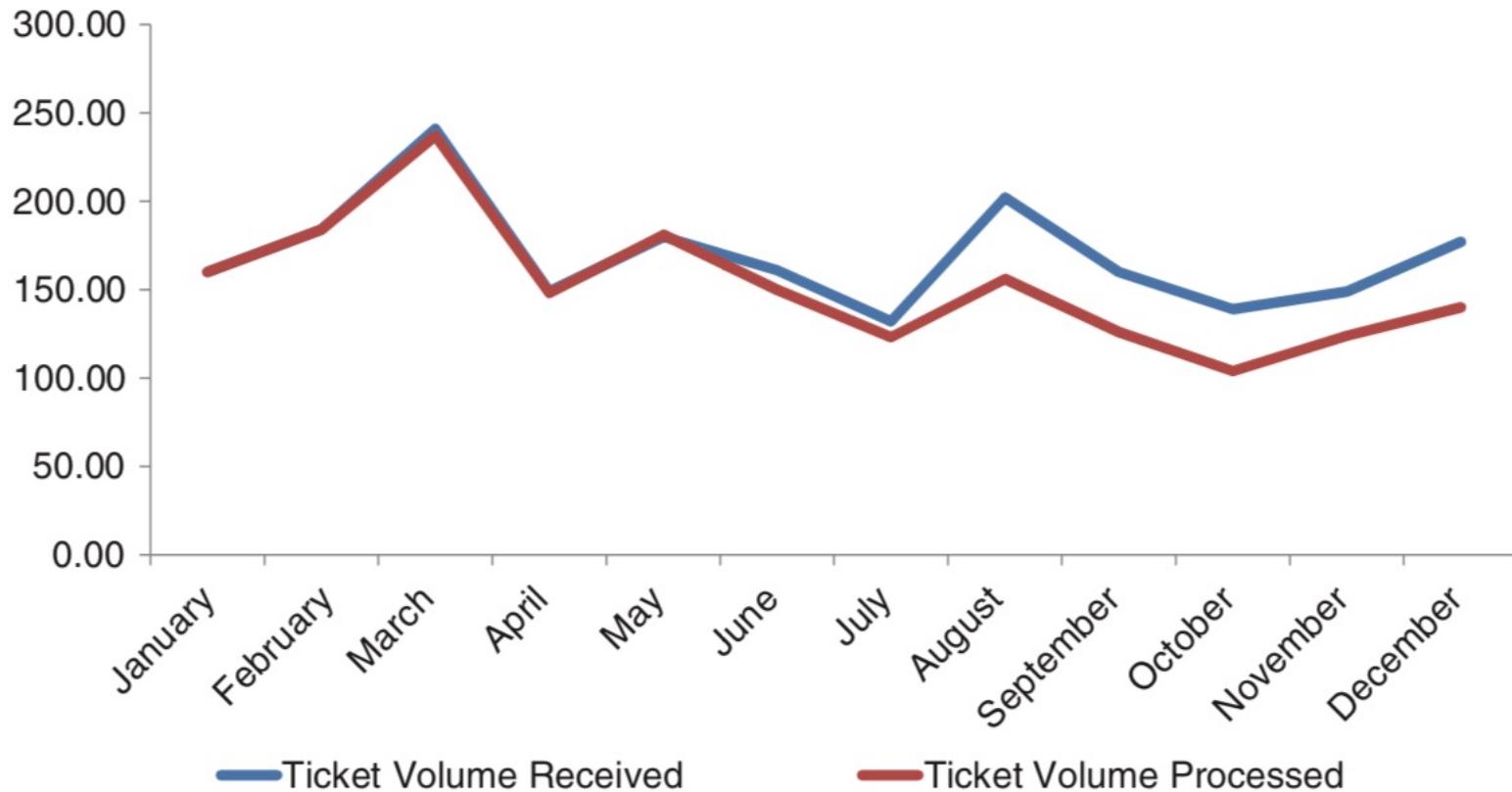
Remove chart border



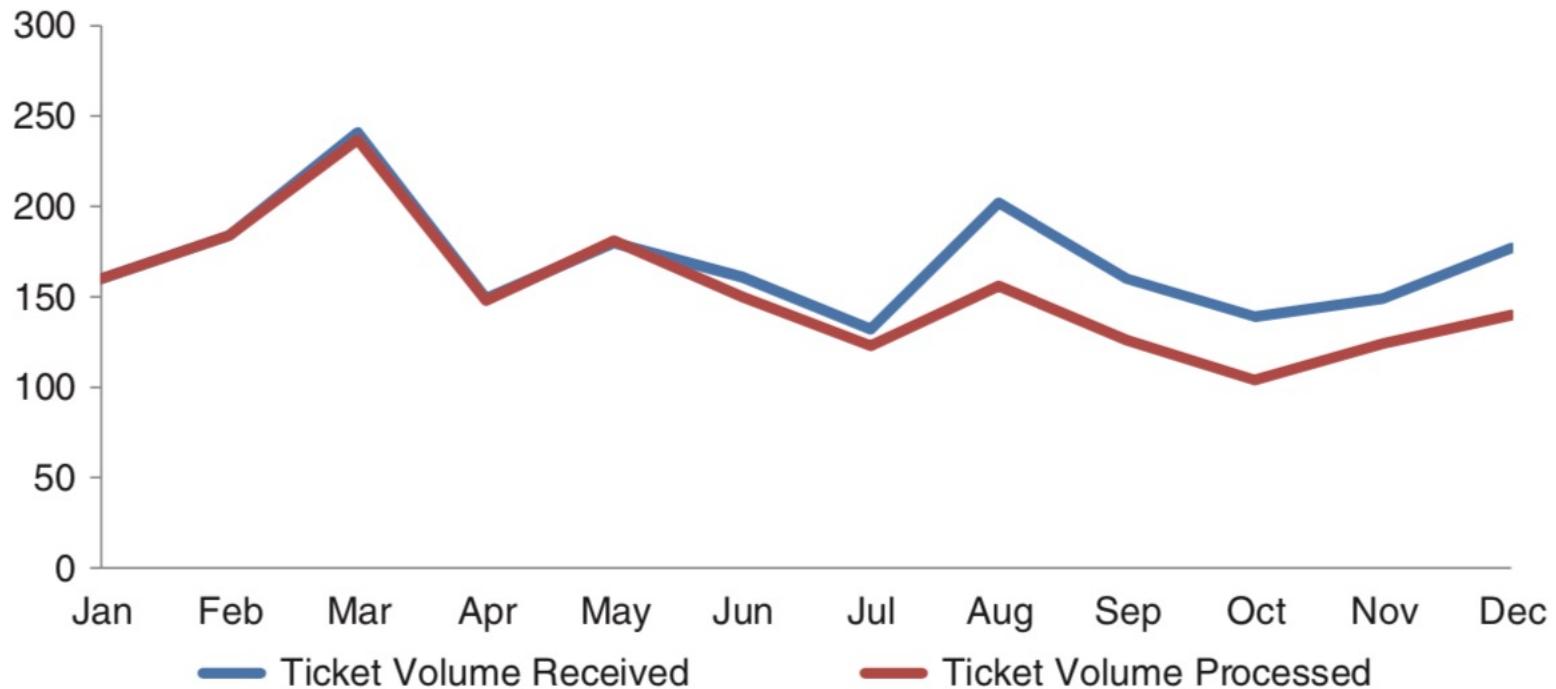
Remove gridlines



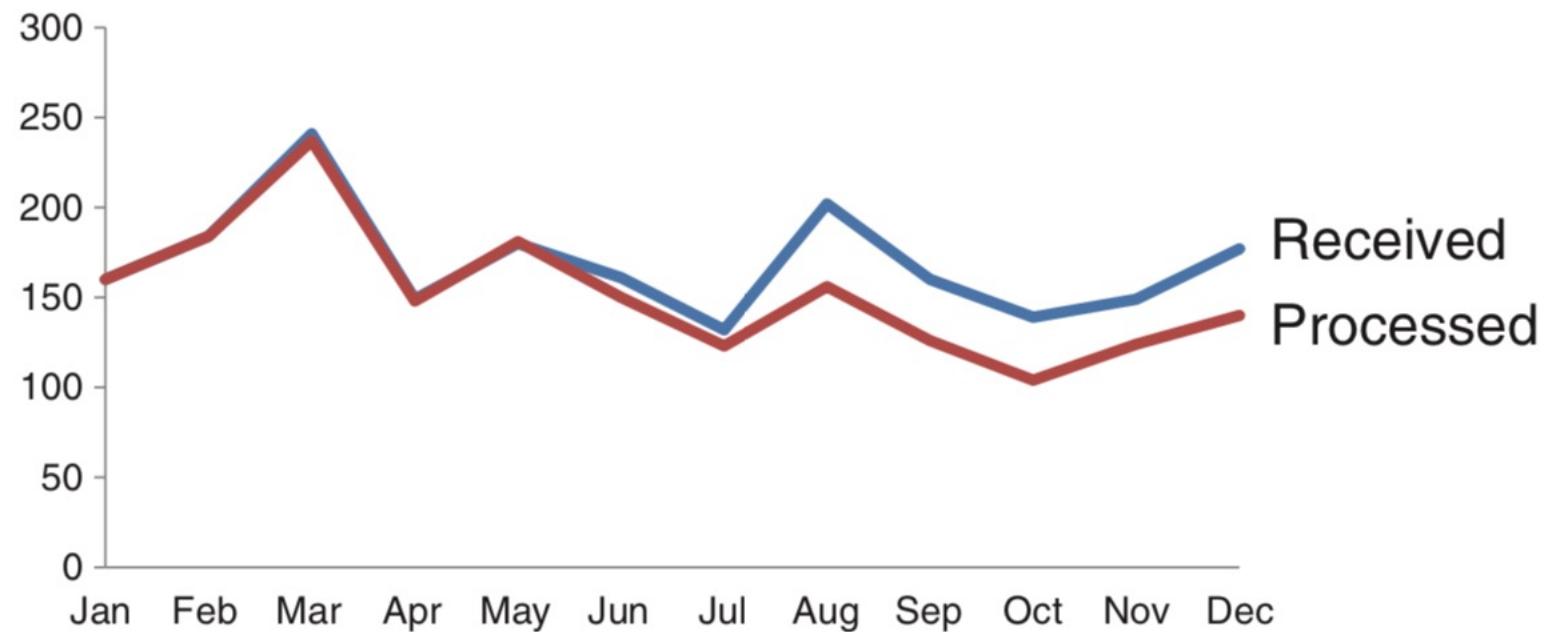
Remove data markers



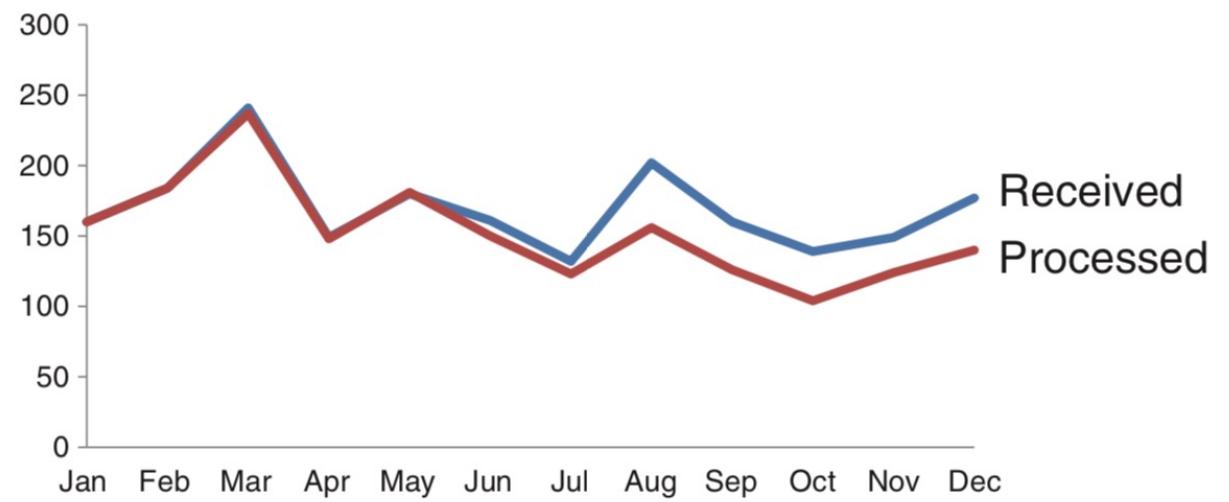
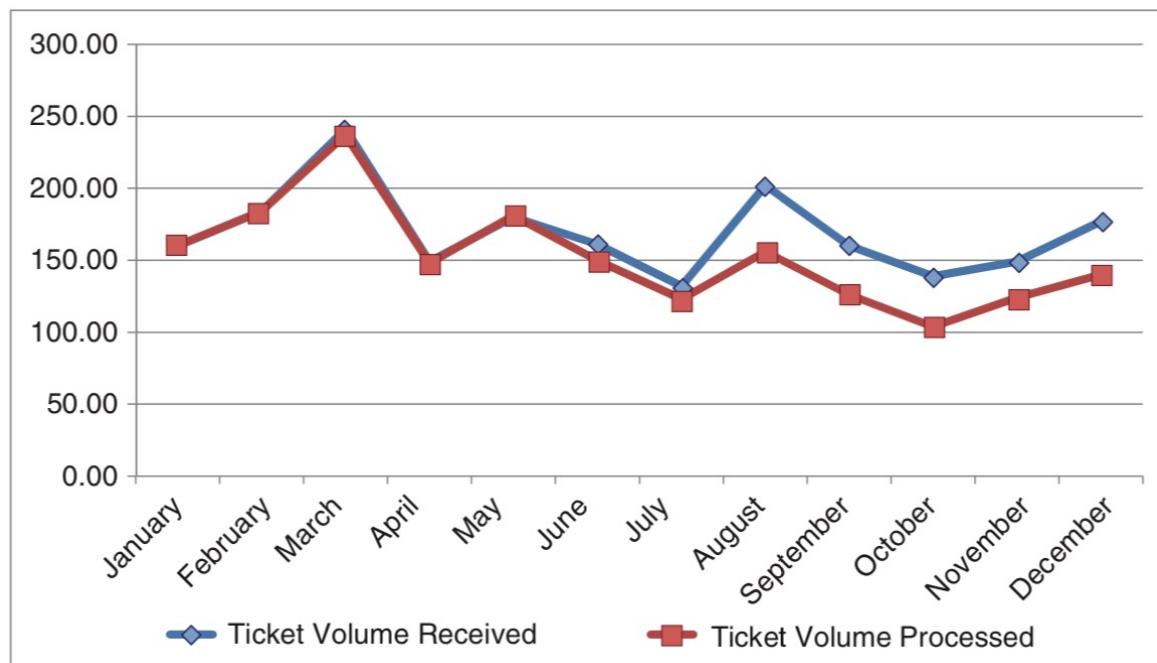
Clean up axis labels



Labels data directly



Before and after

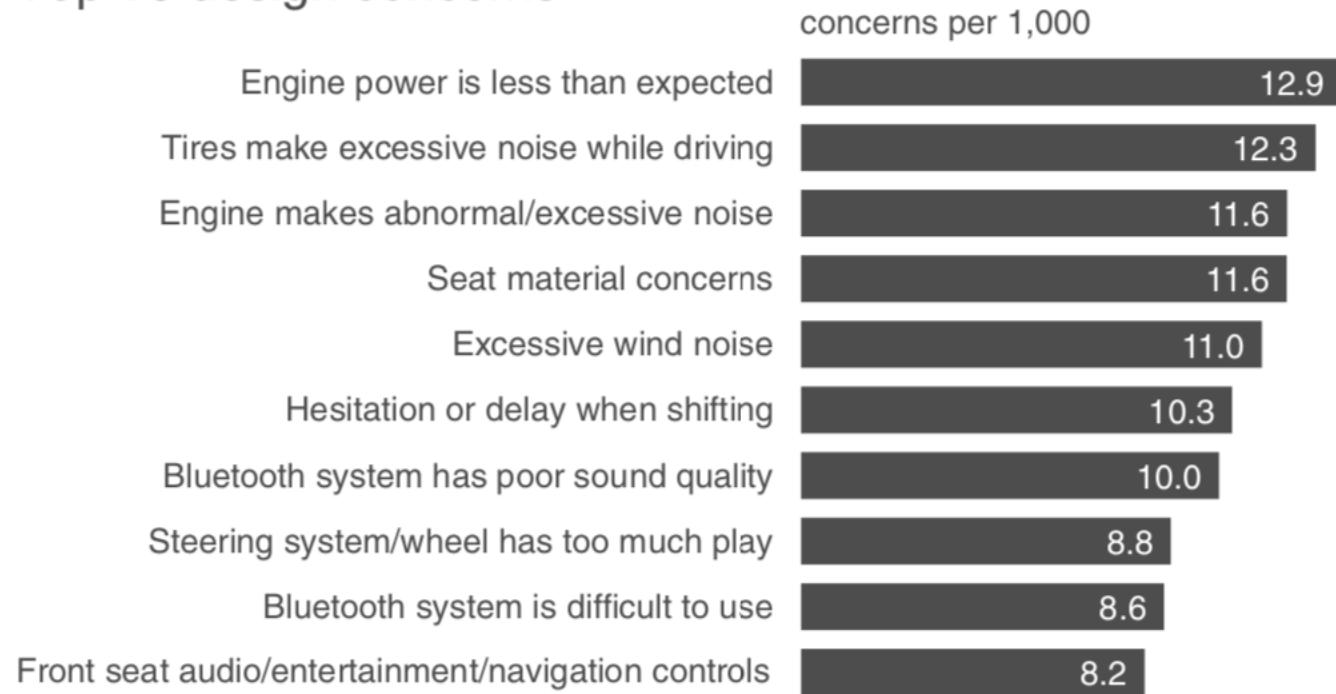


A white origami boat is positioned in the upper left corner of the slide. It has a simple, triangular sail-like shape on its side. The rest of the slide is blank white space.

Focus your audience's
attention

No preattentive attributes

Top 10 design concerns

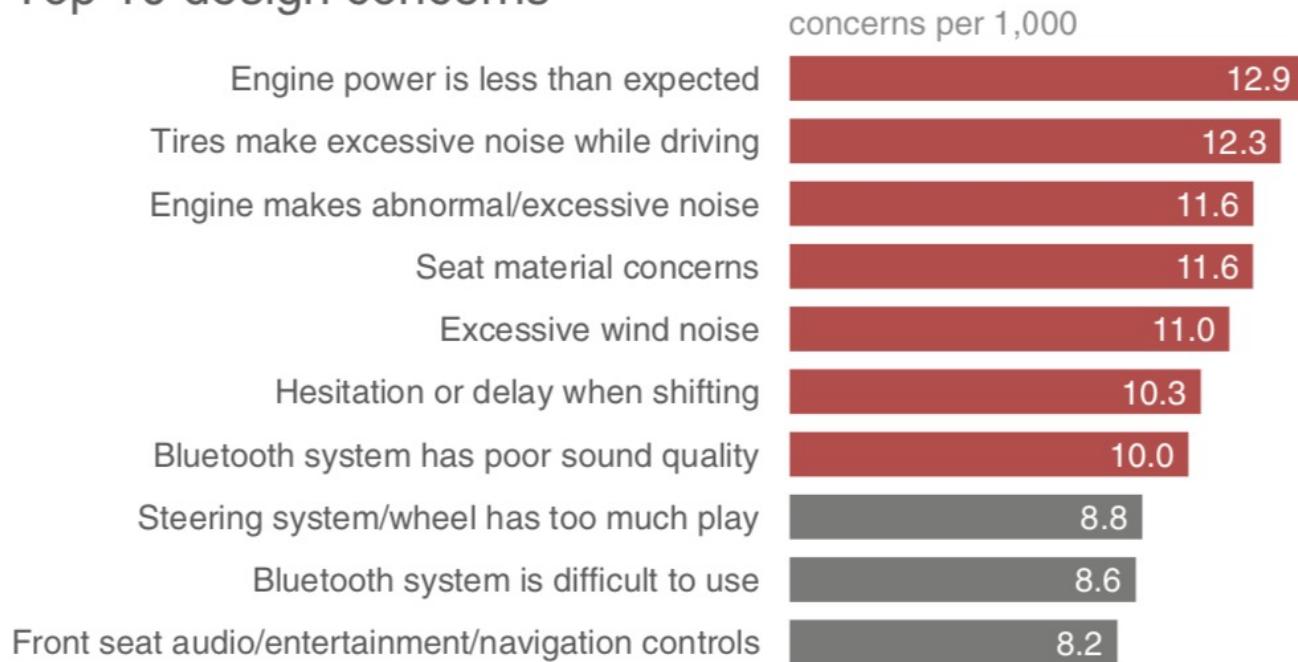


Leverage color to draw attention

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

Discussion: is this an acceptable default rate?

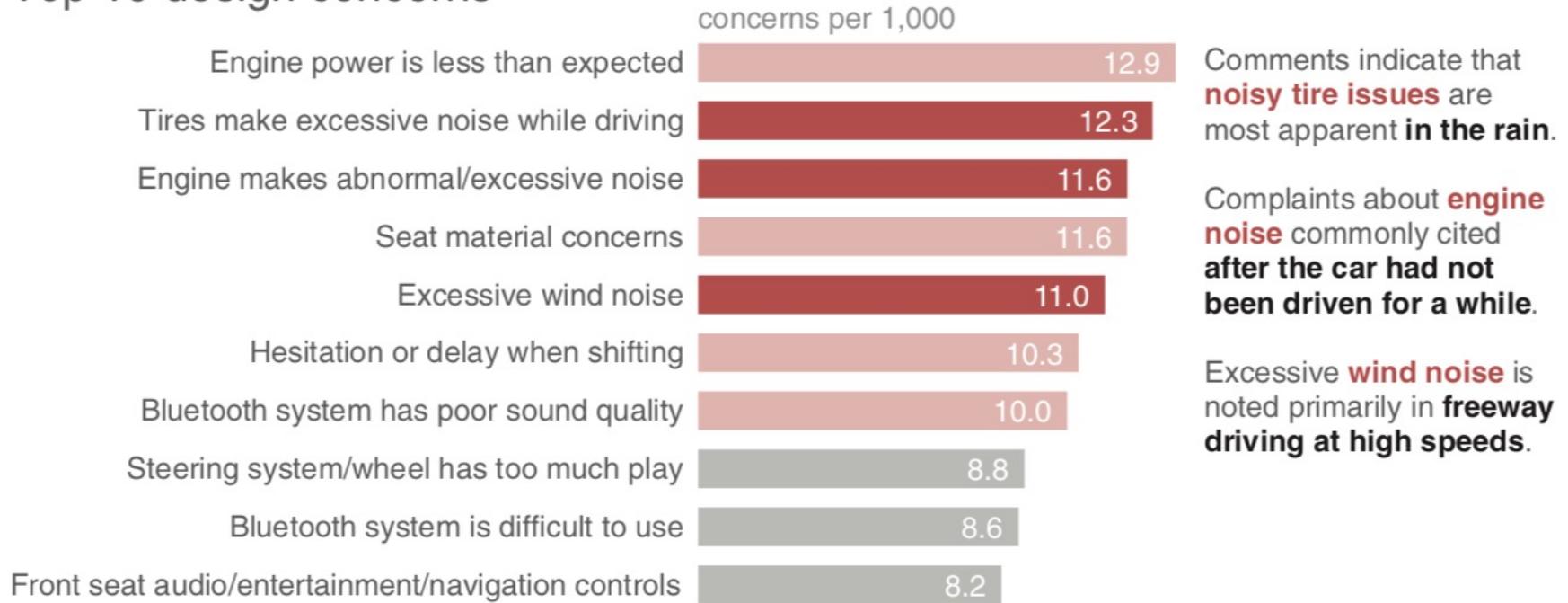
Top 10 design concerns



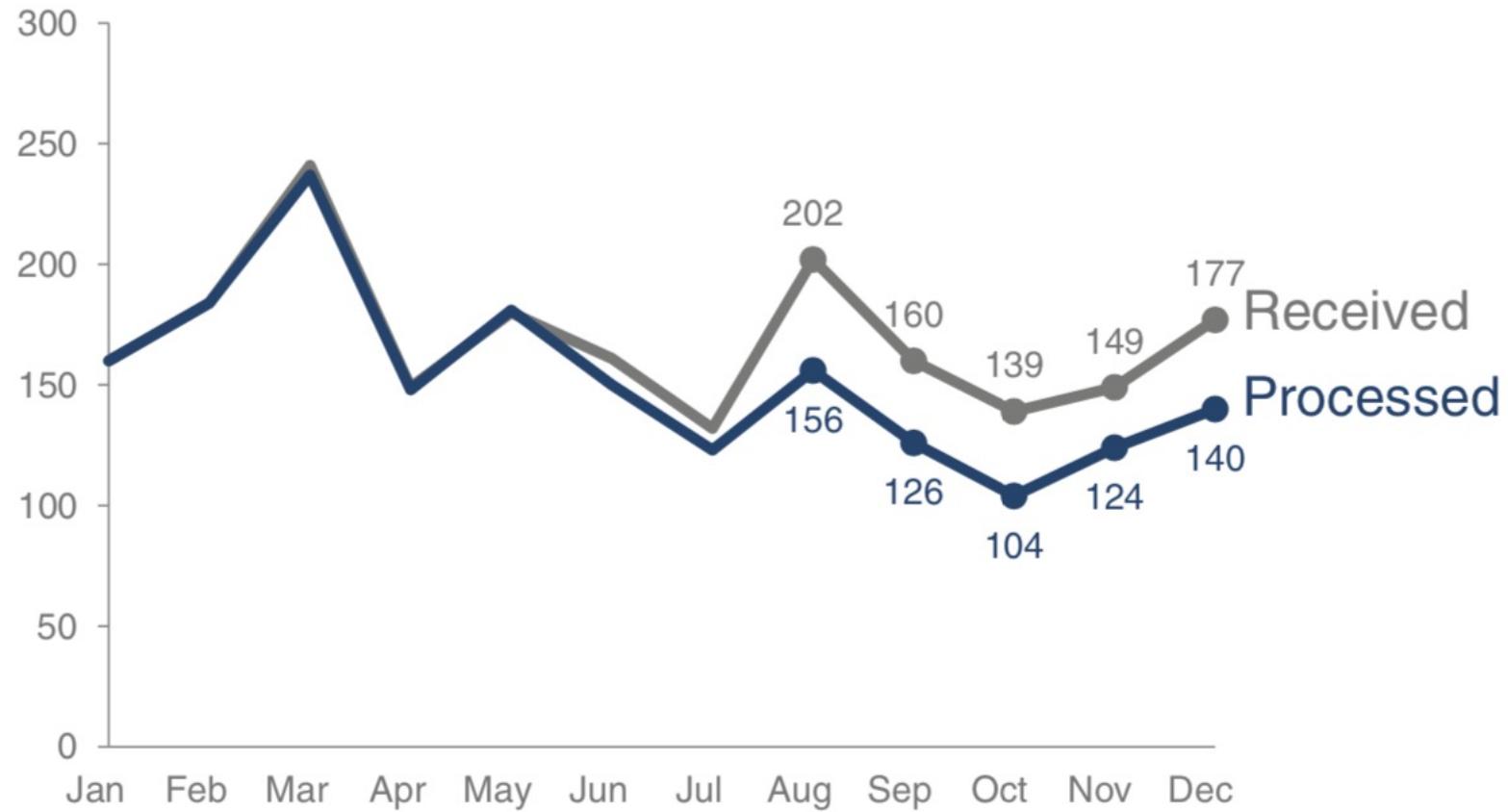
Create visual hierarchy of information

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

Top 10 design concerns



Example revisited



Use color sparingly

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

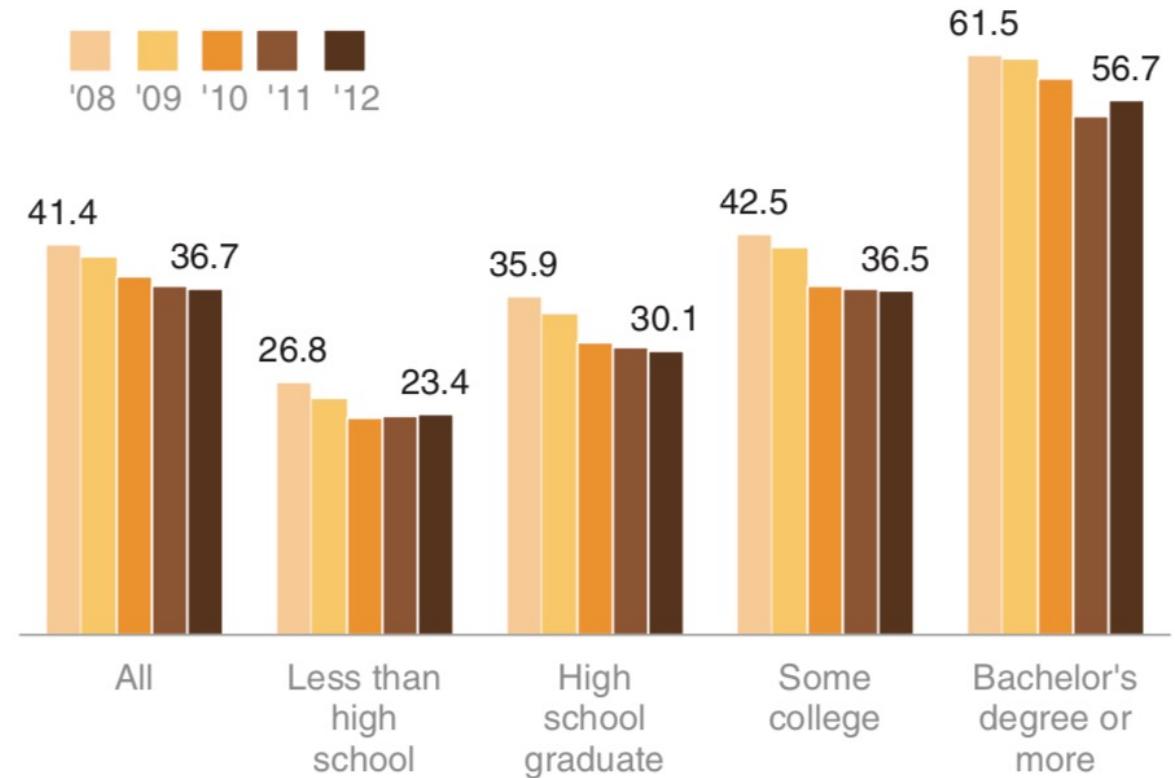


Think like a designer

Original

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



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

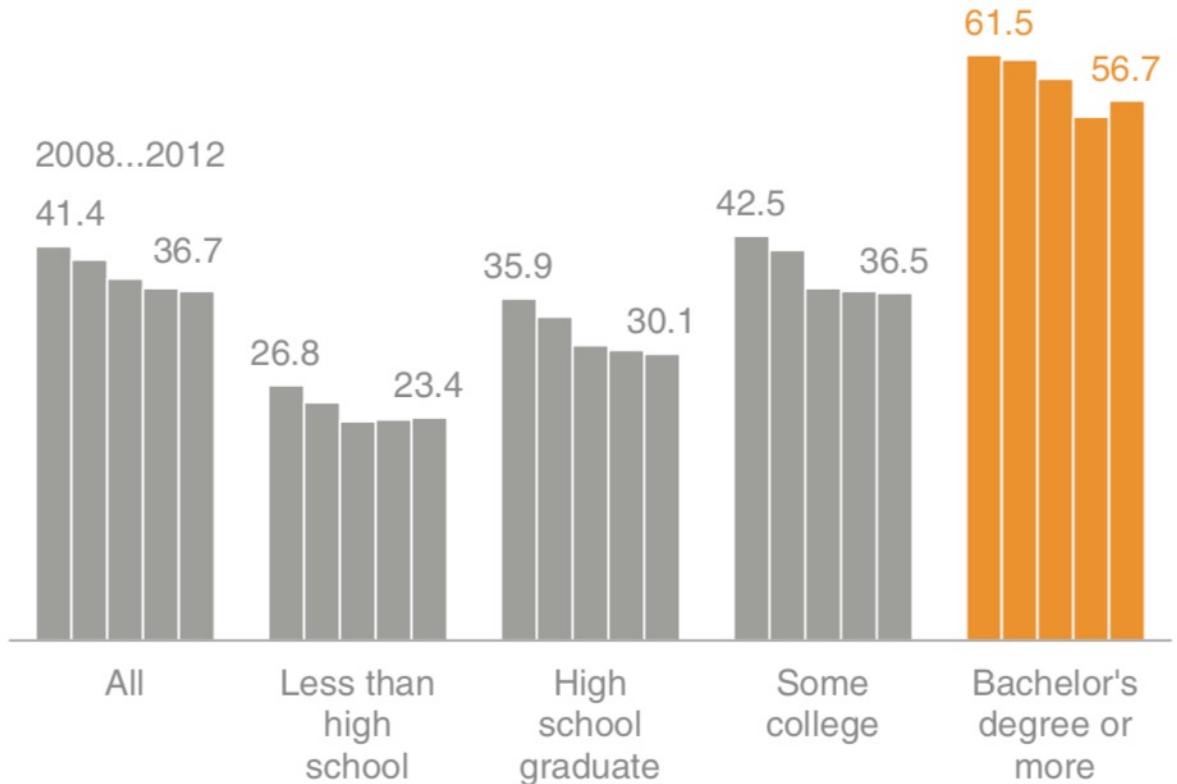
Source: U.S. Census

Adapted from PEW RESEARCH CENTER

Highlight the important stuff

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



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

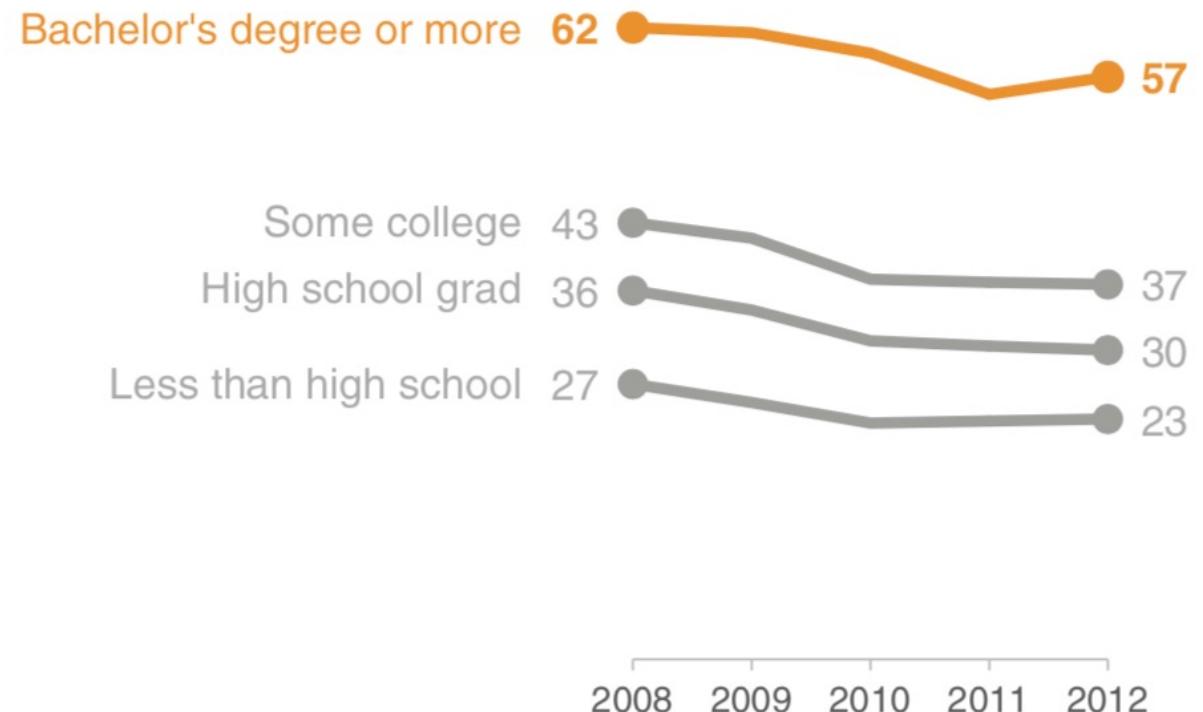
Source: U.S. Census

Adapted from PEW RESEARCH CENTER

Eliminate Distractions

New marriage rate by education

Number of newly married adults per 1,000 marriage eligible adults



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

Source: U.S. Census

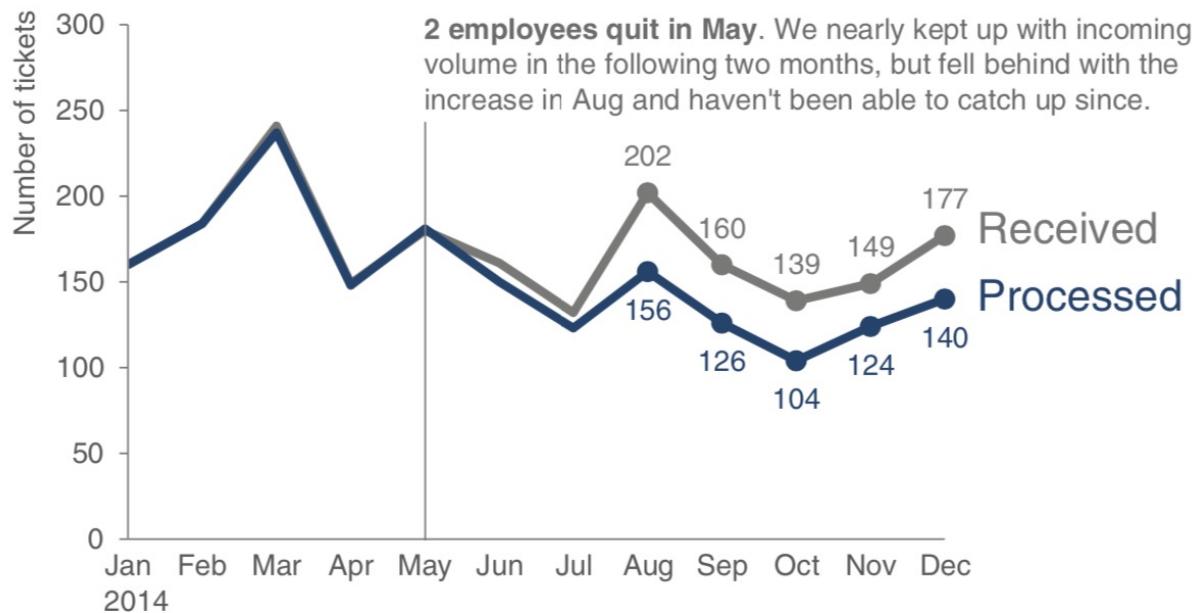
Adapted from PEW RESEARCH CENTER

Use words to make the graph accessible (action title, annotation)

Please approve the hire of 2 FTEs

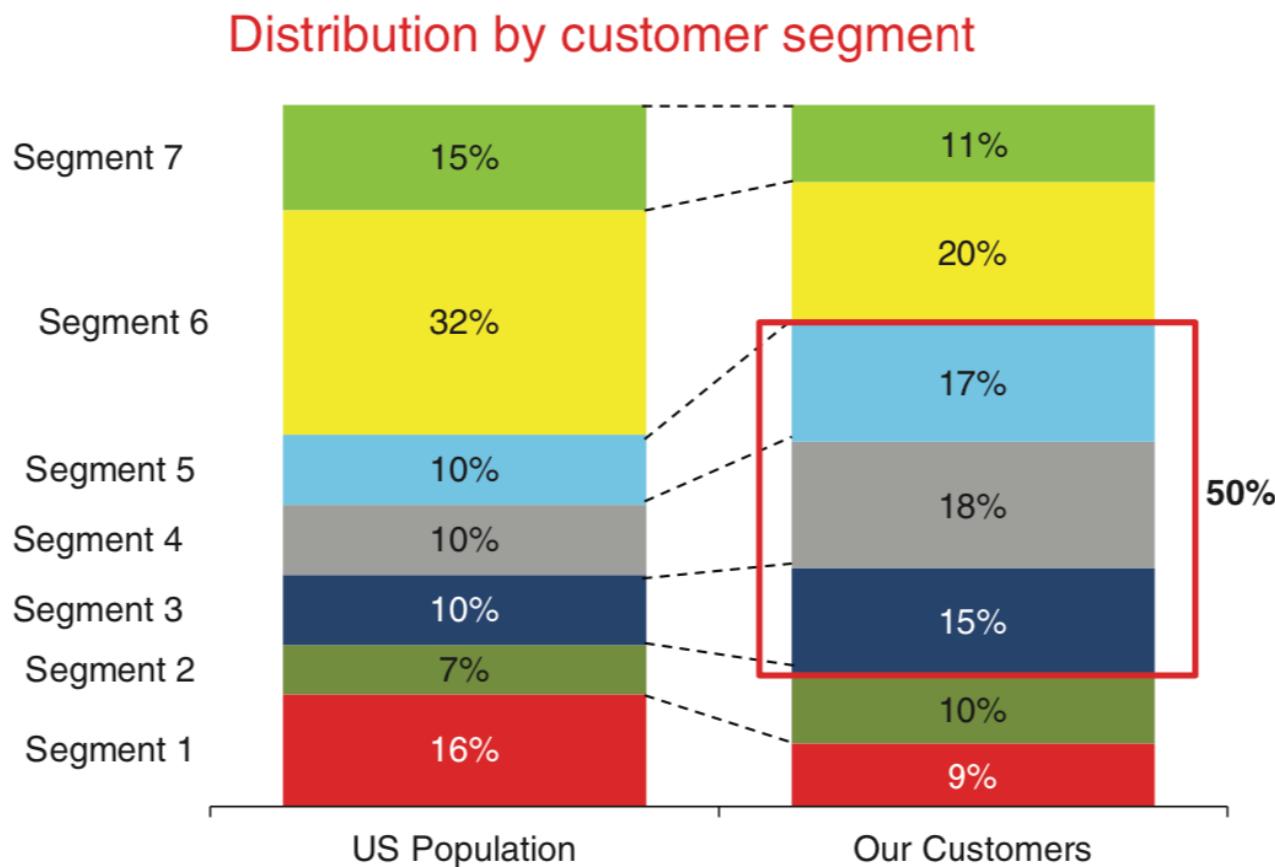
to backfill those who quit in the past year

Ticket volume over time



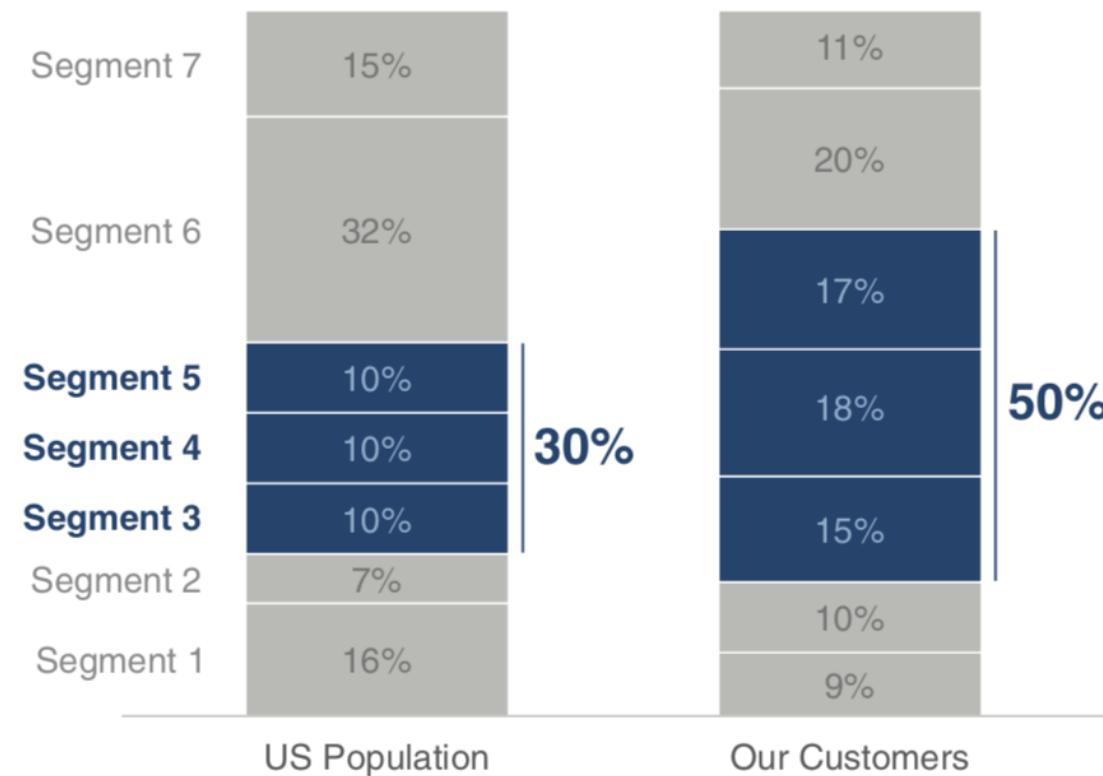
Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

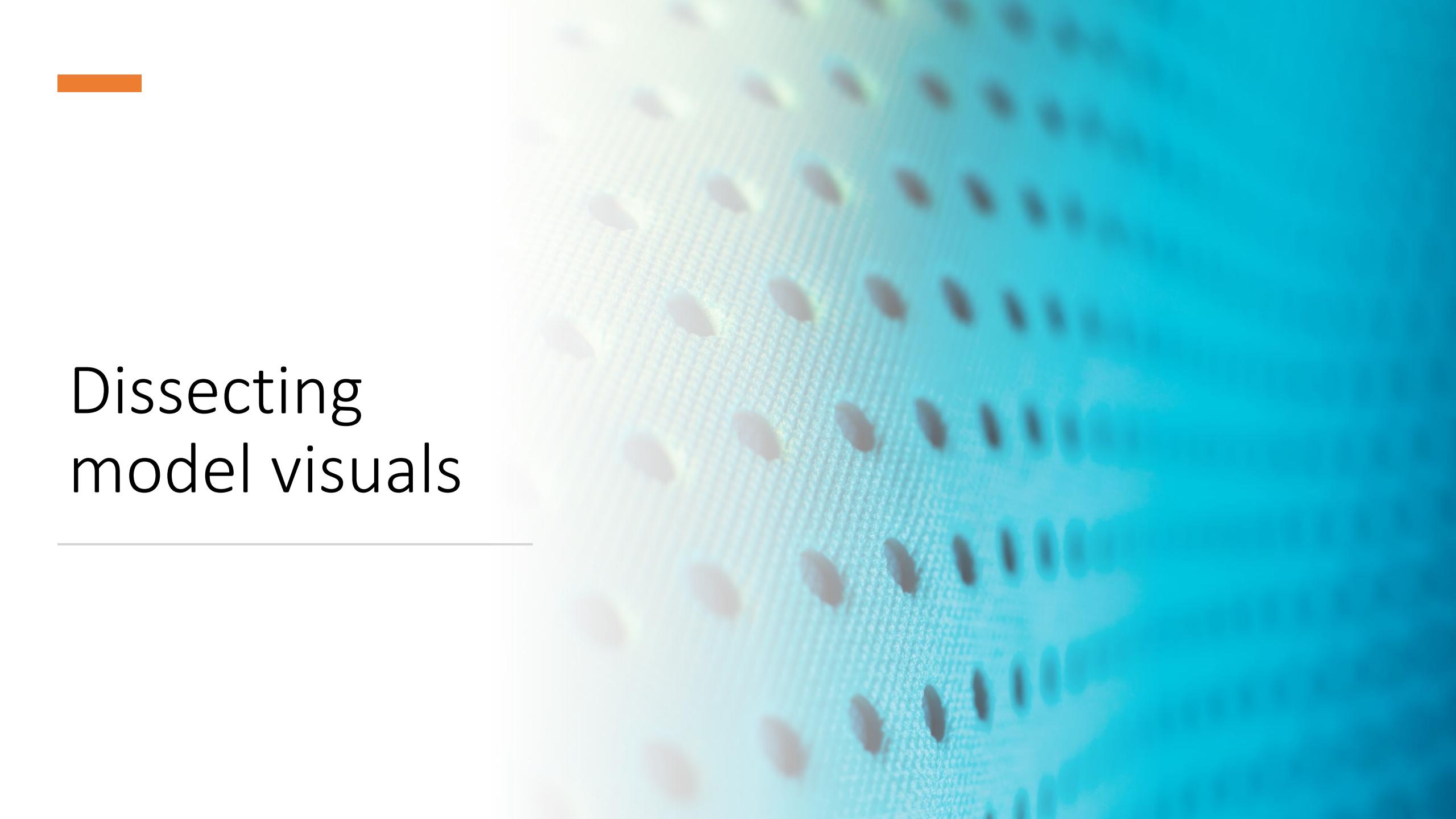
Unaesthetic design



Aesthetic design

Distribution by customer segment

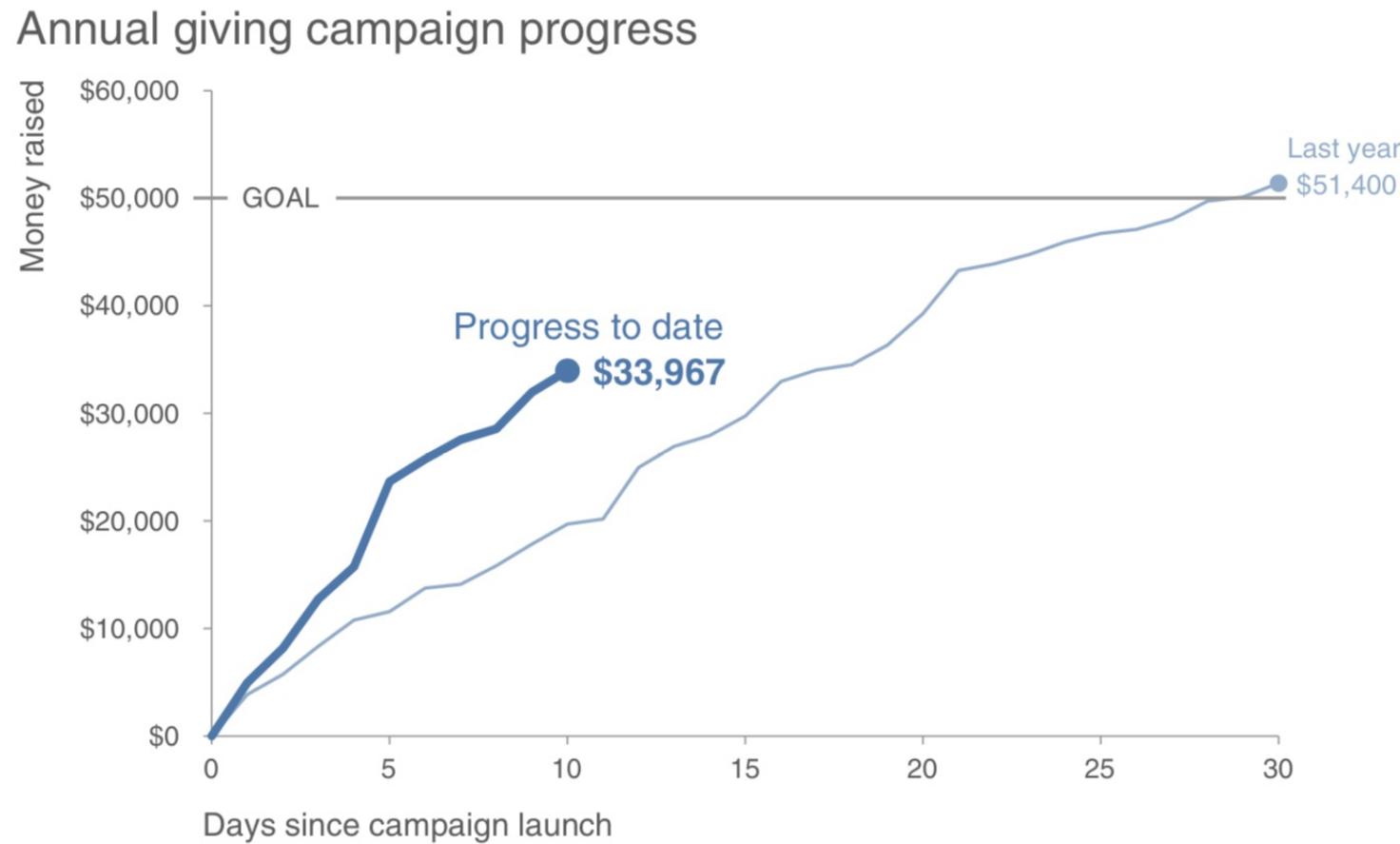




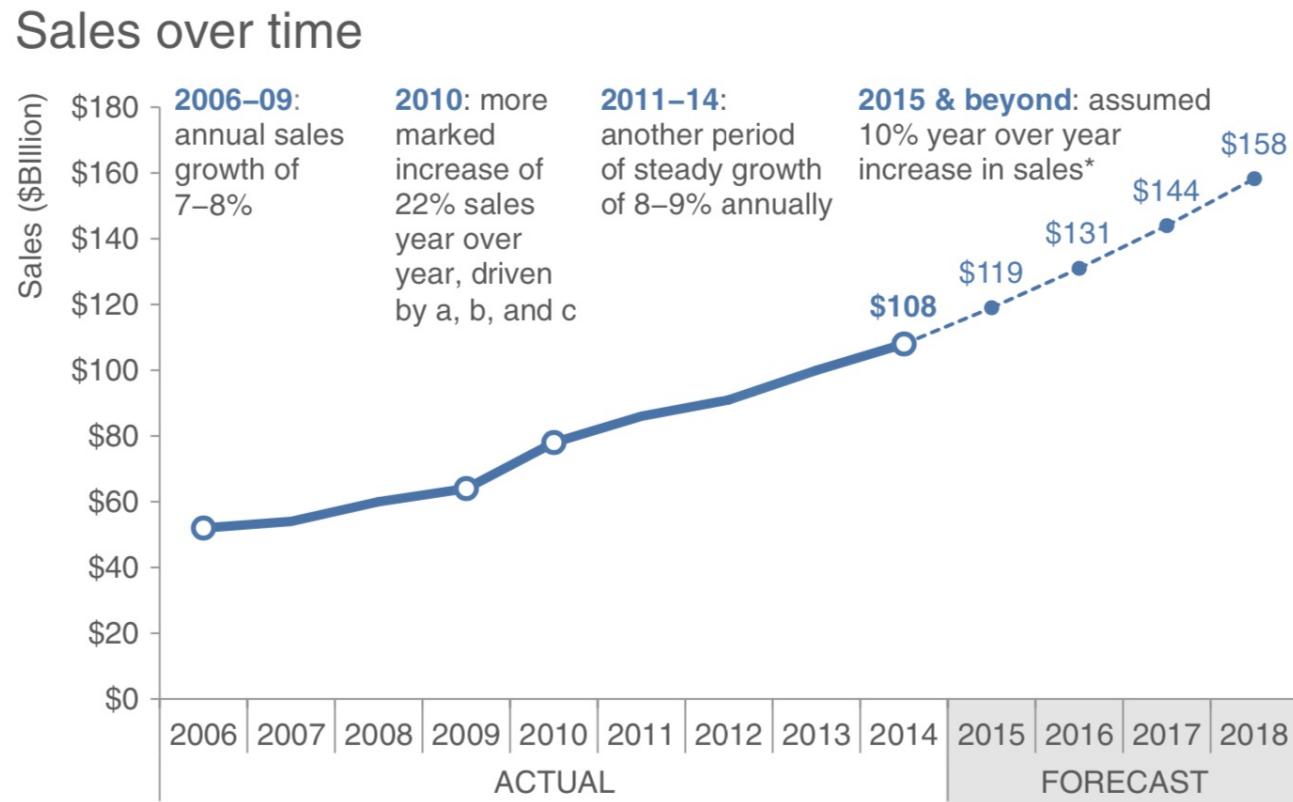
A horizontal orange bar is located at the top left of the slide.

Dissecting model visuals

Line graph



Annotated line graph with forecast



Data source: Sales Dashboard; annual figures are as of 12/31 of the given year.

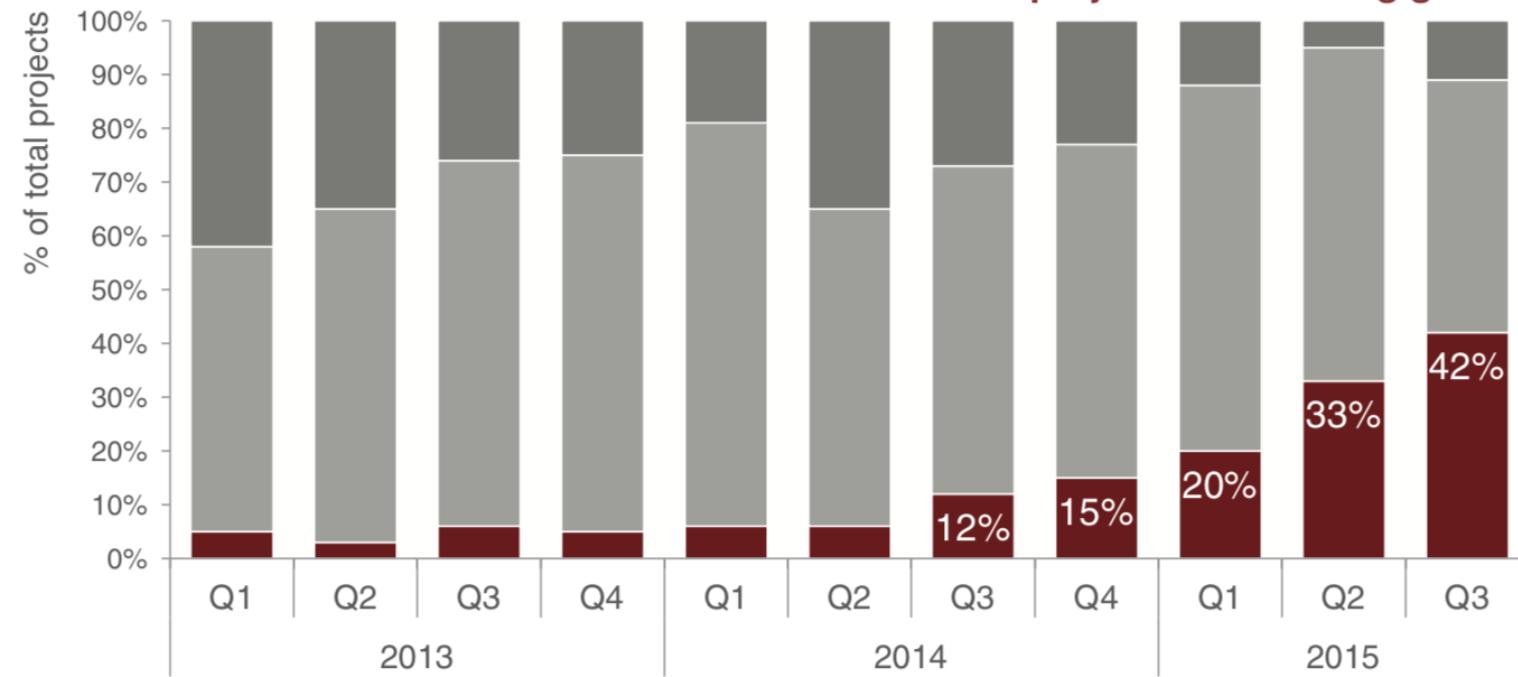
*Use this footnote to explain what is driving the 10% annual growth forecast assumption.

Stacked bars

Goal attainment over time

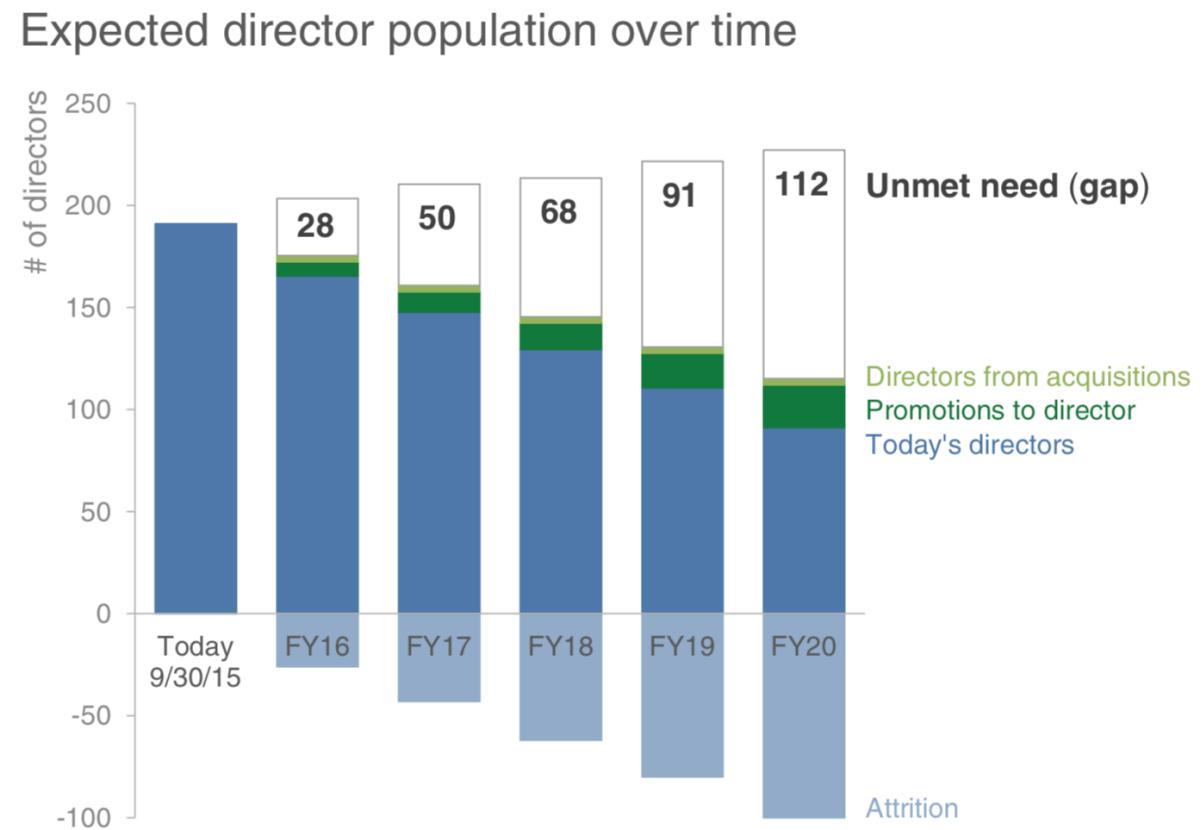
■ Miss ■ Meet ■ Exceed

As of Q3 2015, **more than 1/3 of projects are missing goals**



Data source: XYZ Dashboard; the total number of projects has increased over time from 230 in early 2013 to nearly 270 in Q3 2015.

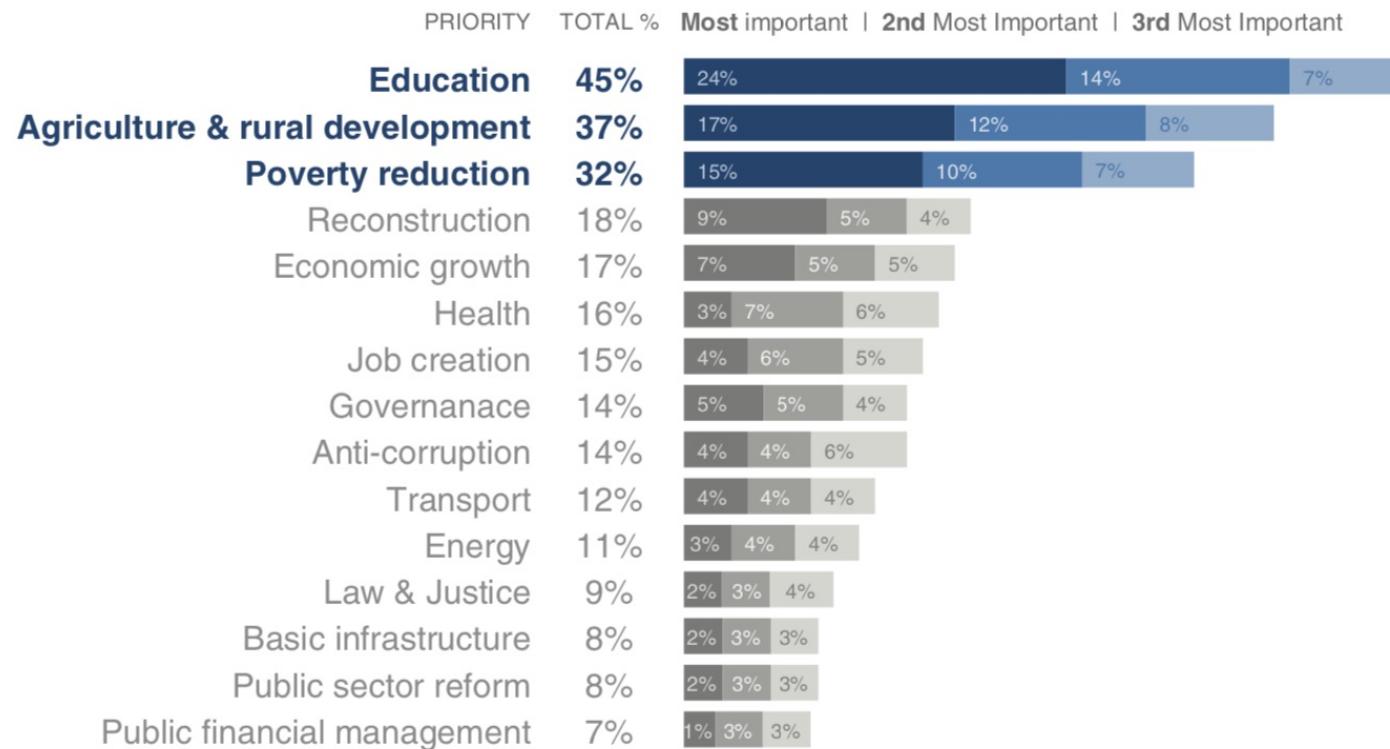
Leveraging positive and negative stacked bars



A footnote explaining relevant forecast assumptions and methodology would go here.

Horizontal stacked bars

Top 15 development priorities, according to survey



N = 4,392. Based on responses to item, *When considering development priorities, which one development priority is the most important? Which one is the second most important priority? Which one is the third most important priority?* Respondents chose from a list. Top 15 shown.

References

Knafllic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

Knafllic, C. N. (2019). *Storytelling with Data: Let's Practice!*. John Wiley & Sons.

<https://www.storytellingwithdata.com/>

<https://www.youtube.com/@storytellingwithdata>