Intro: what is a good visualization?

Alexey Zaytsev,

Skoltech, CDISE

14 January

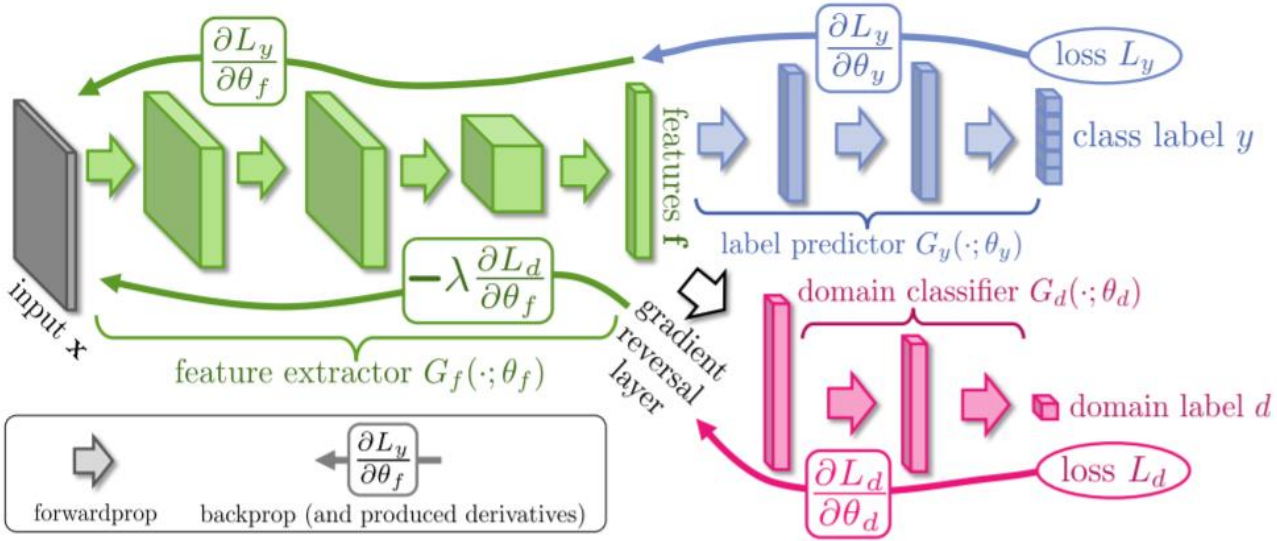
Skoltech
Skolkovo Institute of Science and Technology

Yes, there really is a lot of space junk

https://visme.co/blog/best-data-visualizations/#GhMQDUYUzb1AvoyD.99

https://www.axios.com/yes-there-really-is-a-lot-of-space-junk-930b166d-68c4-4803-9bd6-9e23514eb942.html

How and how well does our domain adaptation work?



Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation." *arXiv preprint arXiv:1409.7495* (2014).

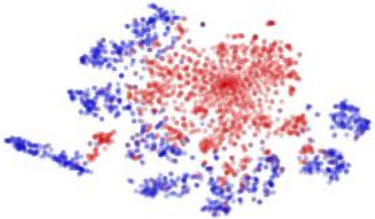
MSCVS2011: V. Lempitsky, Skoltech - Writing a computer vision paper https://youtu.be/-Oq8AHIC92A

How to Write a Great Research Paper, S.P. Jones, Microsoft Research. https://youtu.be/WP-FkUaOcOM

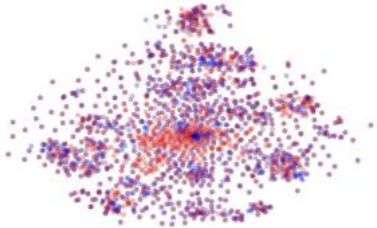
Skoltech
Skolkovo Institute of Science and Technology

How and **how well** does our domain adaptation work?

MNIST → MNIST-M: top feature extractor layer

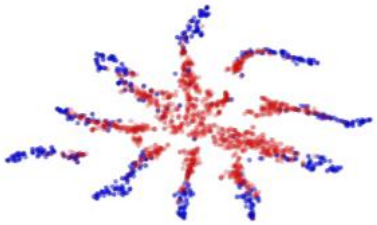


(a) Non-adapted

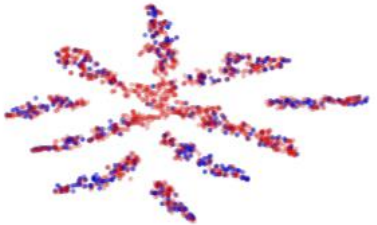


(b) Adapted

SYN NUMBERS → SVHN: last hidden layer of the label predictor



(a) Non-adapted



(b) Adapted

Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation." *arXiv preprint arXiv:1409.7495* (2014).

MSCVS2011: V. Lempitsky, Skoltech - Writing a computer vision paper https://youtu.be/-Oq8AHIC92A

How to Write a Great Research Paper, S.P. Jones, Microsoft Research. https://youtu.be/WP-FkUaOcOM

How and **how well** does our domain adaptation work?

Figure 2. Examples of domain pairs used in the experiments. See Section 4.1 for details.

METHOD	SOURCE	MNIST	SYN NUMBERS	SVHN	SYN SIGNS
	TARGET	MNIST-M	SVHN	MNIST	GTSRB
SOURCE ONLY		.5749	.8665	.5919	.7400
SA (FERNANDO ET AL., 2013)		.6078 (7.9%)	.8672 (1.3%)	.6157 (5.9%)	.7635 (9.1%)
PROPOSED APPROACH		**.8149** (57.9%)	**.9048** (66.1%)	**.7107** (29.3%)	**.8866** (56.7%)
TRAIN ON TARGET		.9891	.9244	.9951	.9987

Table 1. Classification accuracies for digit image classifications for different source and target domains. MNIST-M corresponds to difference-blended digits over non-uniform background. The first row corresponds to the lower performance bound (i.e. if no adaptation is performed). The last row corresponds to training on the target domain data with known class labels (upper bound on the DA performance). For each of the two DA methods (ours and (Fernando et al., 2013)) we show how much of the gap between the lower and the upper bounds was covered (in brackets). For all five cases, our approach outperforms (Fernando et al., 2013) considerably, and covers a big portion of the gap.

Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation." *arXiv preprint arXiv:1409.7495* (2014).

MSCVS2011: V. Lempitsky, Skoltech - Writing a computer vision paper https://youtu.be/-Oq8AHIC92A

How to Write a Great Research Paper, S.P. Jones, Microsoft Research. https://youtu.be/WP-FkUaOcOM

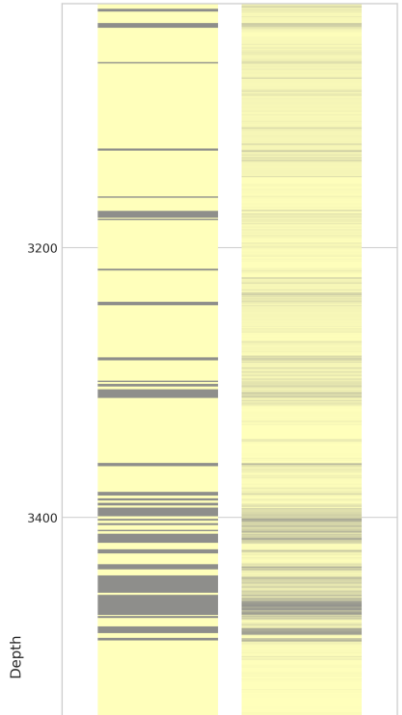
A quick intro on storytelling with data

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

Knaflic, C.N. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, 2015. p. 254

1. Understand the context

1. **Who**
   - → Your audience
   - → You



Klyuchnikov, N., Zaytsev, A et al., 2018. Data-driven model for the identification of the rock type at a drilling bit. *arXiv preprint arXiv:1806.03218*.

Skoltech
Skolkovo Institute of Science and Technology

1. Understand the context

1. **Who**

   → Your audience

   → You

2. **What**

   → Action

   What do you need your audience to know or do?

   → Mechanism

   How will you communicate to your audience?

1. Understand the context

1. **Who**

   → Your audience

   → You

2. **What**

   → Action

   What do you need your audience to know or do?

   → Mechanism

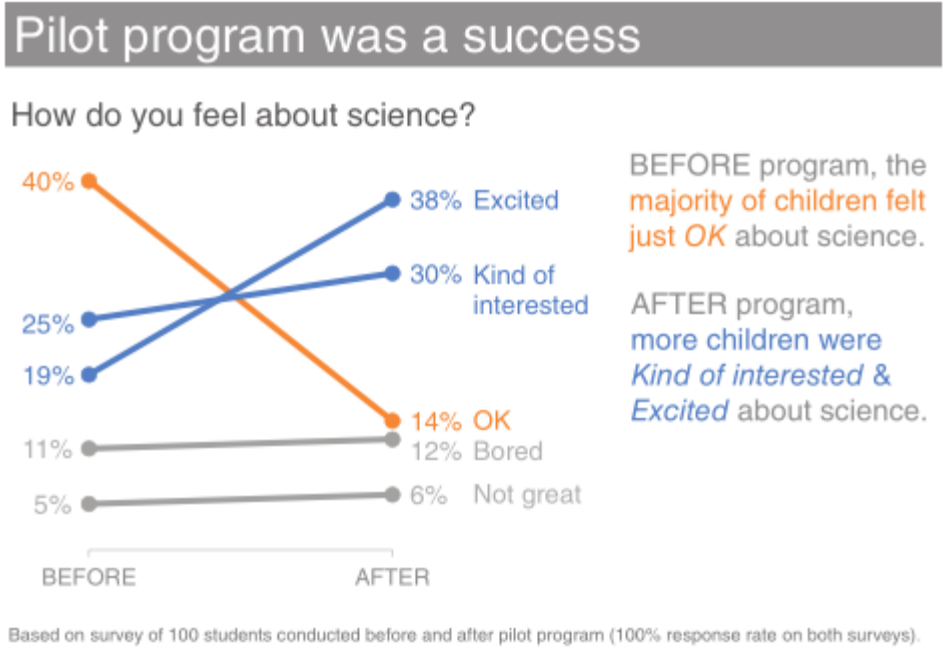
   How will you communicate to your audience?

3. **How**

   What data is available that will help make my point?

1. Understand the context: example

1. **Who:** The budget committee that can approve funding for continuation of the summer learning program.
2. **What:** The summer learning program on science was a success; please approve budget of $X to continue.
3. **How:** Illustrate success with data collected through the survey conducted before and after the pilot program.



Skoltech
Skolkovo Institute of Science and Technology

2. Choose an appropriate visual display

FIGURE 2.1 The visuals I use most

Knaflic, C.N. *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, 2015. p. 37

Skoltech
Skolkovo Institute of Science and Technology

Main display types

- Simple text
- Tables
- Graphs

91%

Simple text

	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

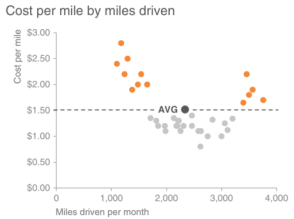


FIGURE 2.7 Modified scatterplot

Skoltech
Skolkovo Institute of Science and Technology

Graph types

- Scatter plot
- Lines
- Parallel coordinates
- Bars

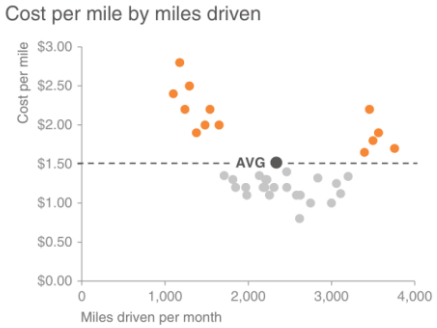


FIGURE 2.7 Modified scatterplot

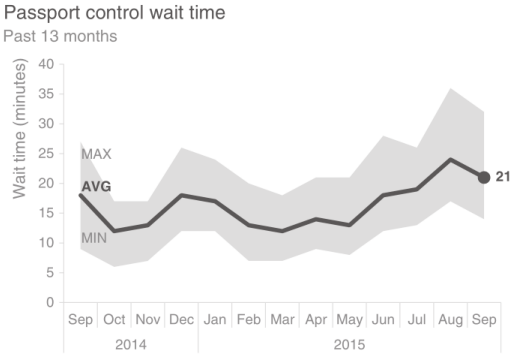
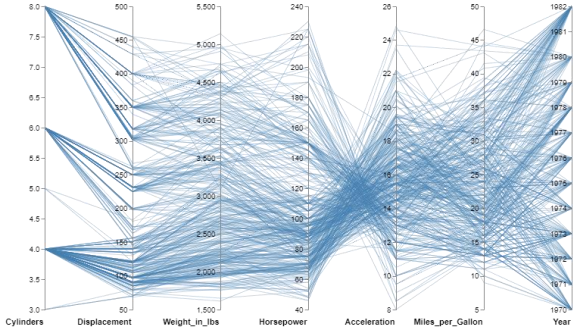


FIGURE 2.9 Showing average within a range in a line graph



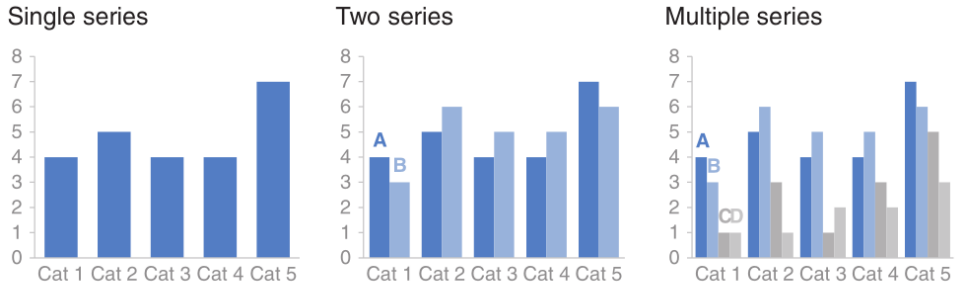


FIGURE 2.15 Bar charts

Skoltech
Skolkovo Institute of Science and Technology

Graph types to avoid

- Pie charts are evil
- Never use 3D
- Secondary y-axis: generally not a good idea

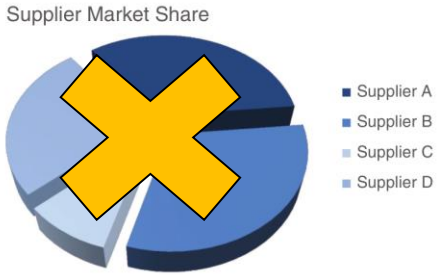


FIGURE 2.21 Pie chart

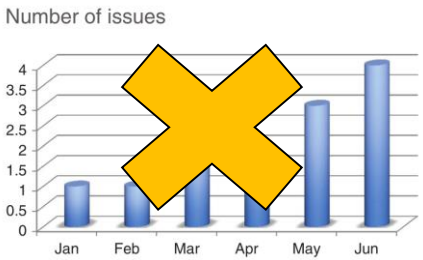


FIGURE 2.25 3D column chart

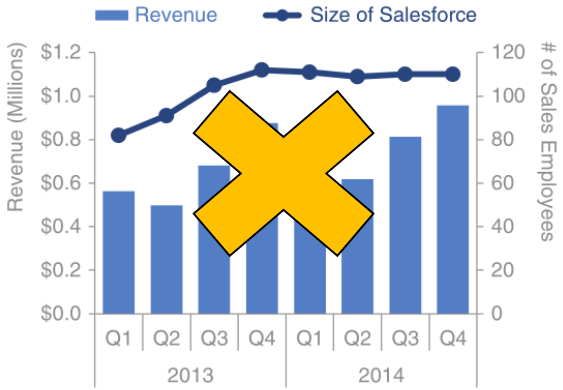


FIGURE 2.26 Secondary y-axis

Skoltech
Skolkovo Institute of Science and Technology

3. Eliminate clutter

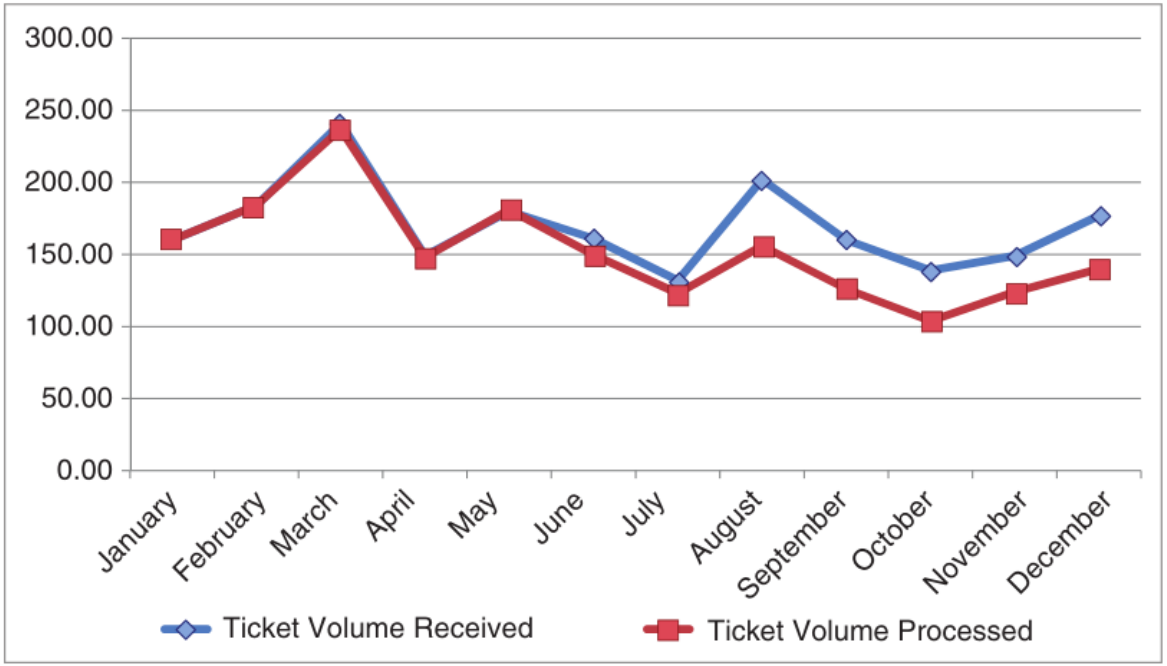
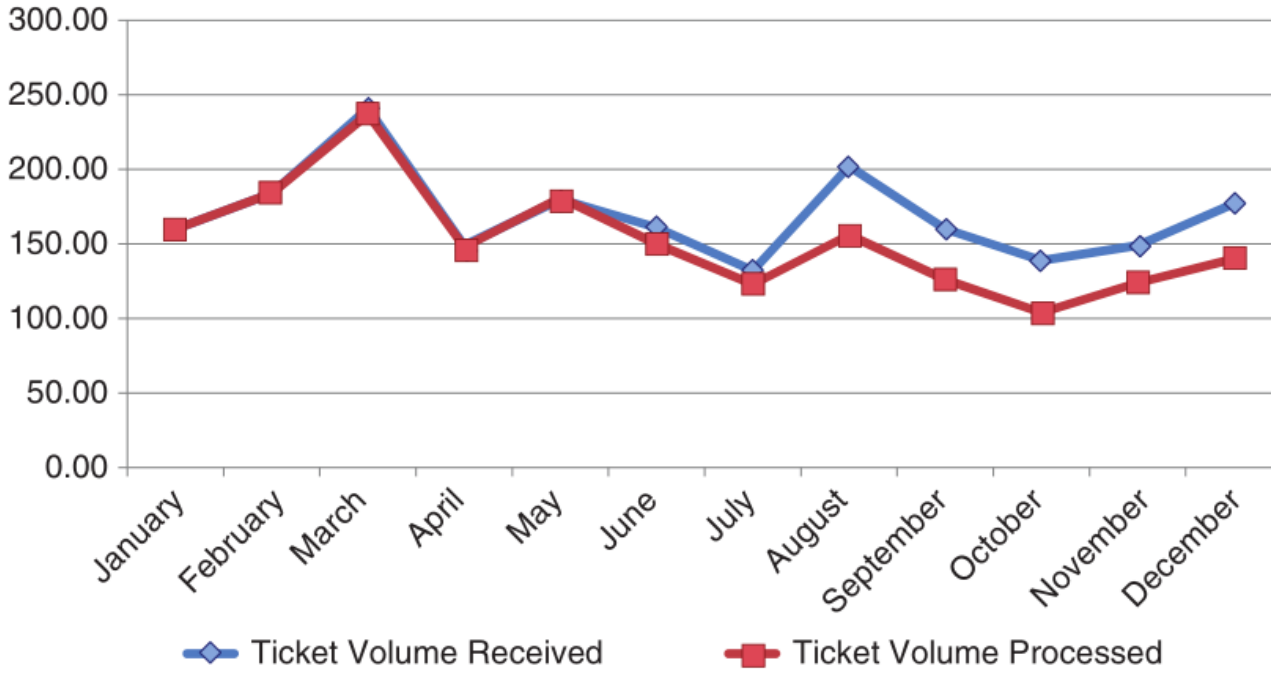


FIGURE 3.17 Original graph

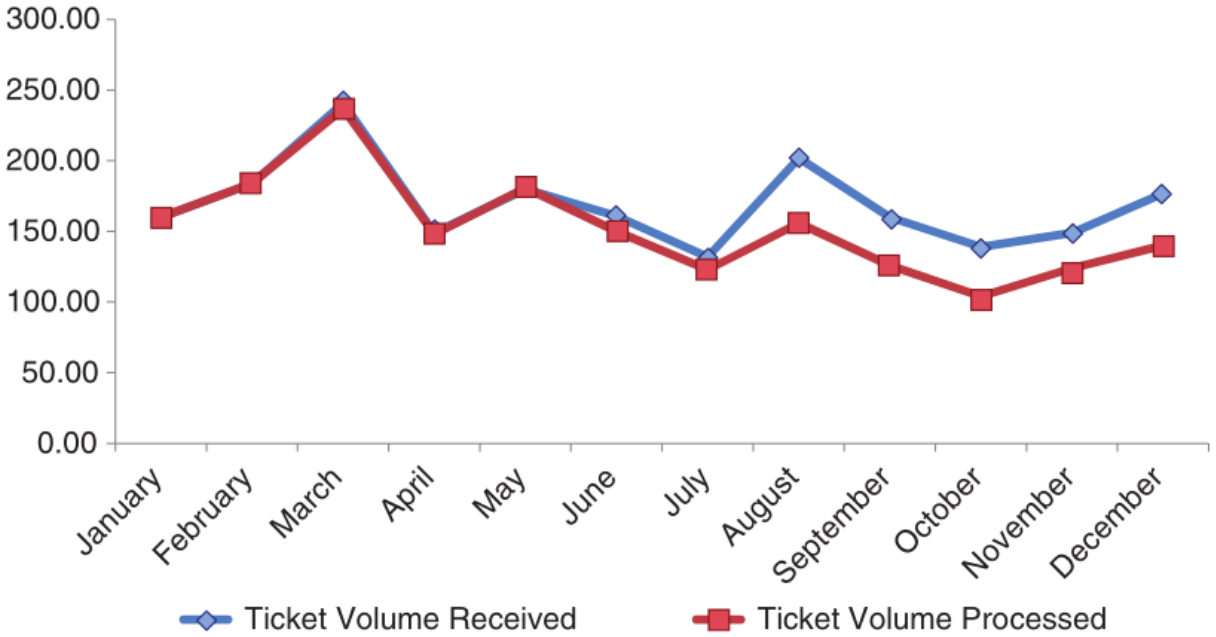
Skoltech
Skolkovo Institute of Science and Technology

Remove chart border

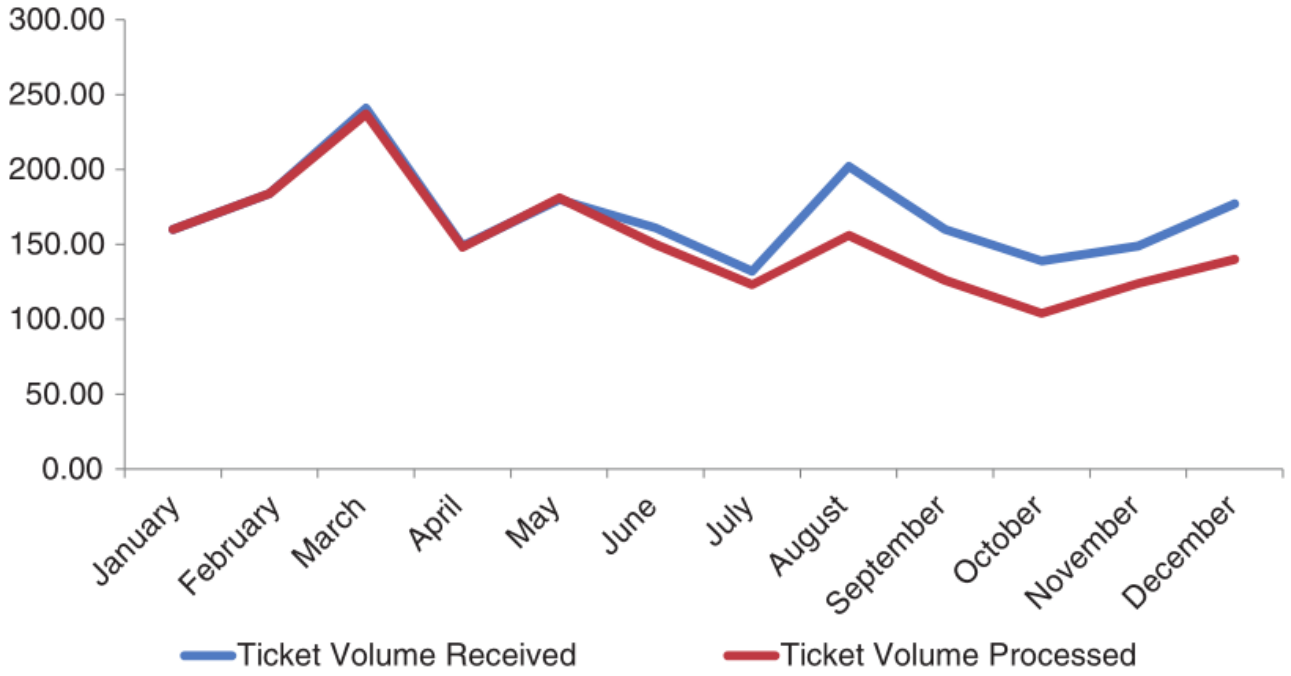


Skoltech
Skolkovo Institute of Science and Technology

Remove gridlines

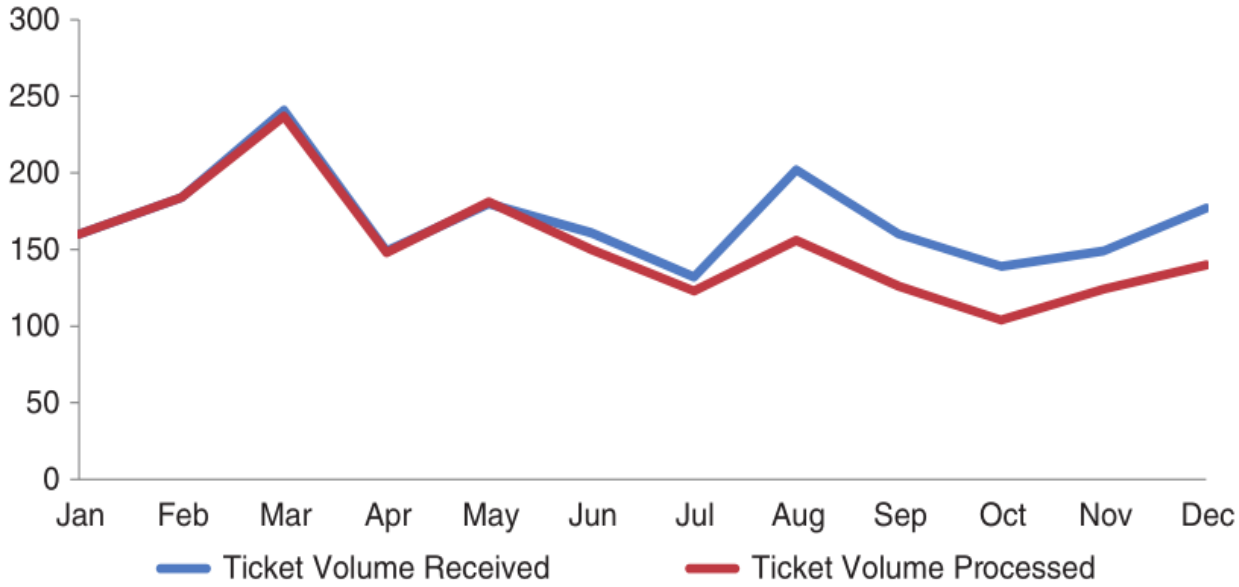


Remove data markers

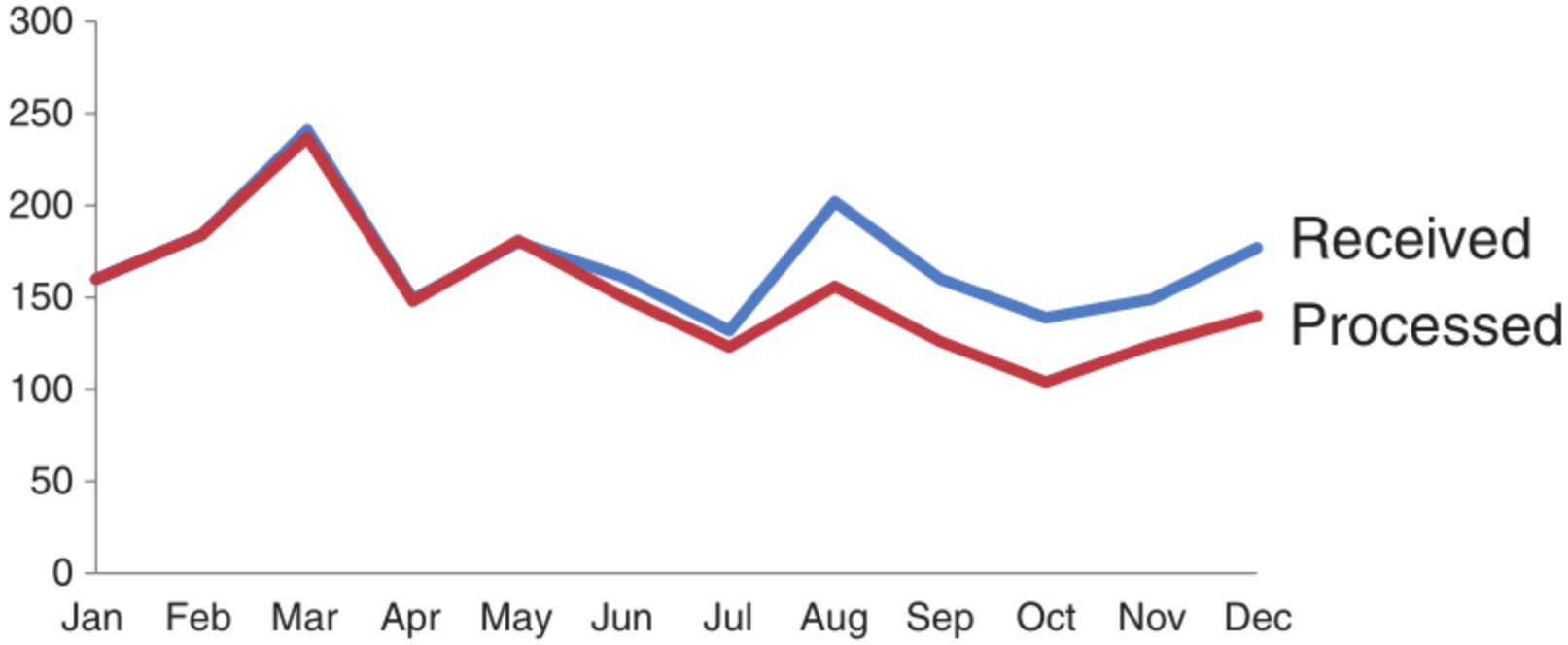


Skoltech
Skolkovo Institute of Science and Technology

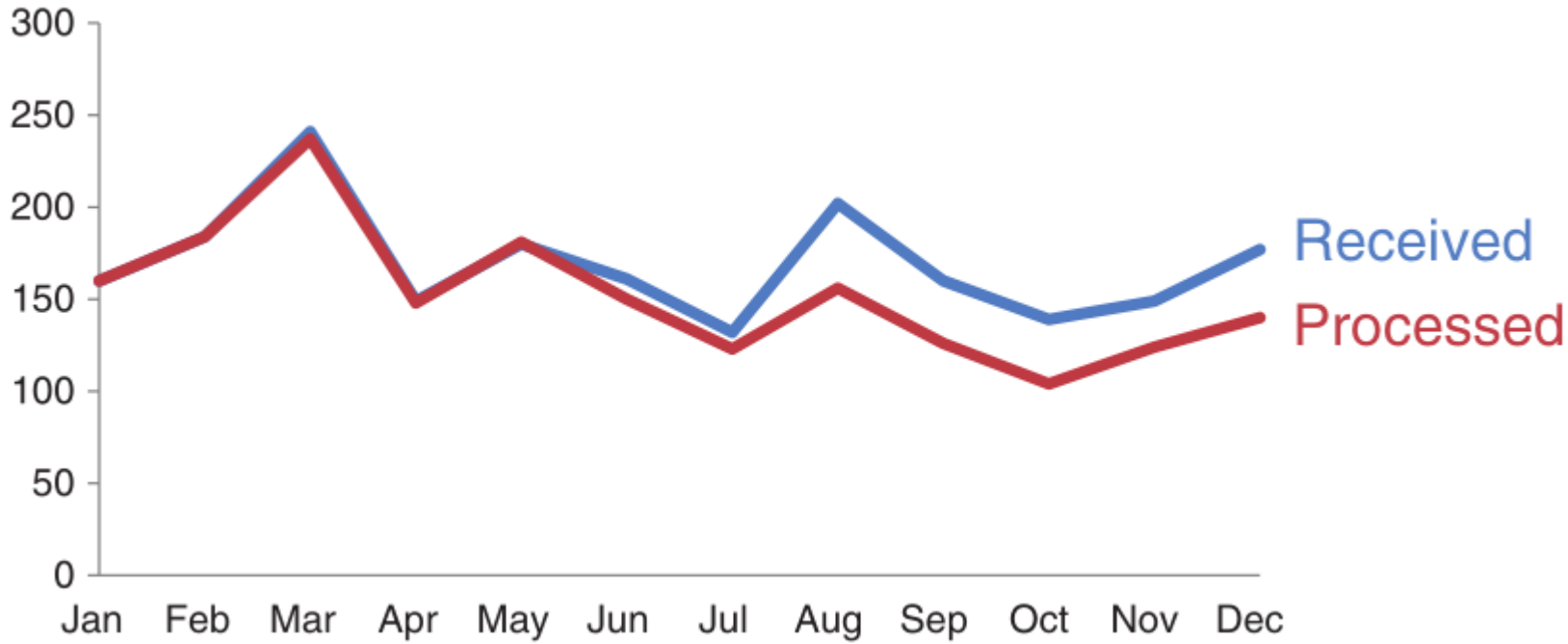
Clean up axis labels



Label data directly

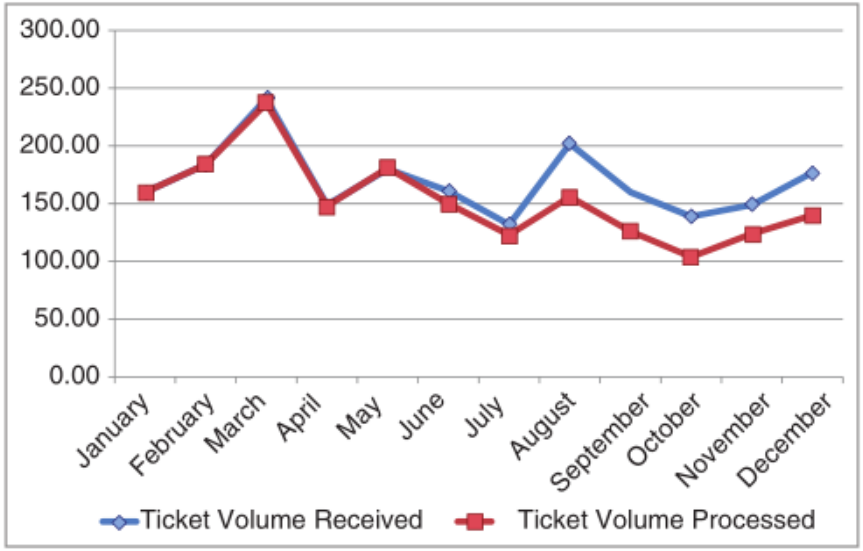


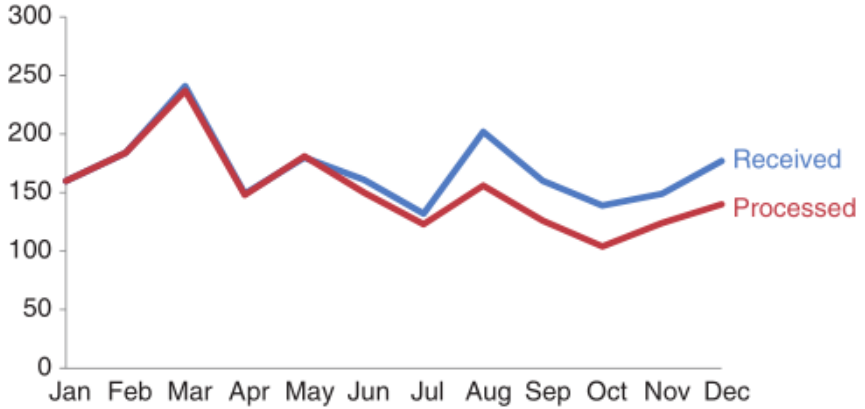
Leverage consistent color



Skoltech
Skolkovo Institute of Science and Technology

Before and after





4. Focus attention where you want it

Count the 3s

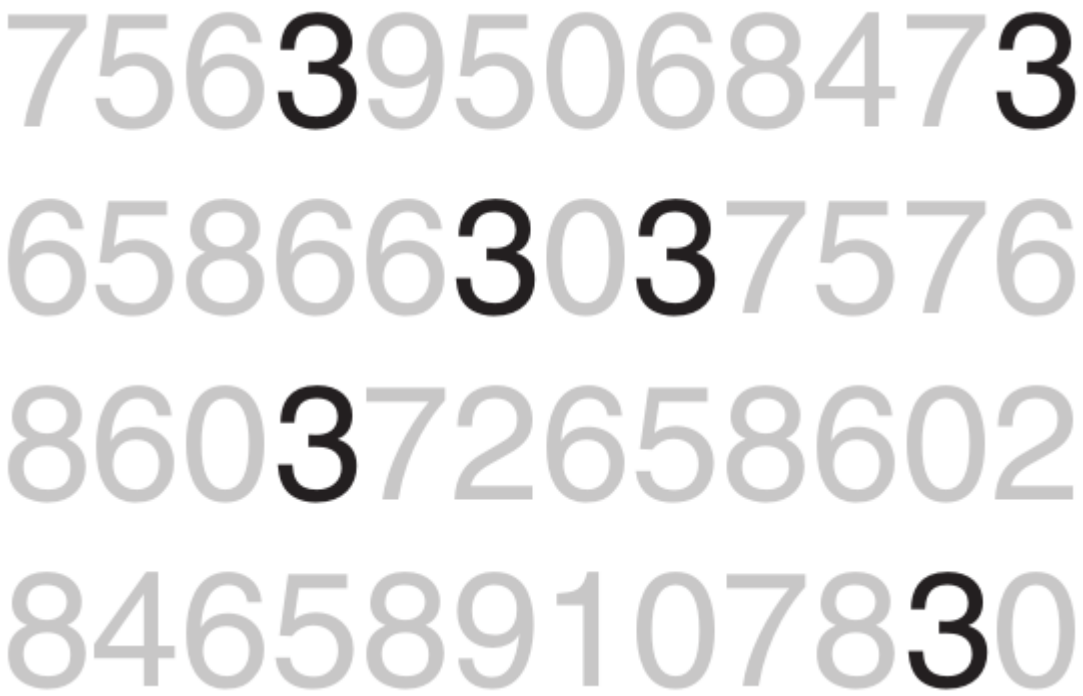
75639506847 3

6586630375 76

8603726586 02

8465891078 30

Count the 3s with preattentive attributes



Skoltech
Skolkovo Institute of Science and Technology

Types of preattentive attributes

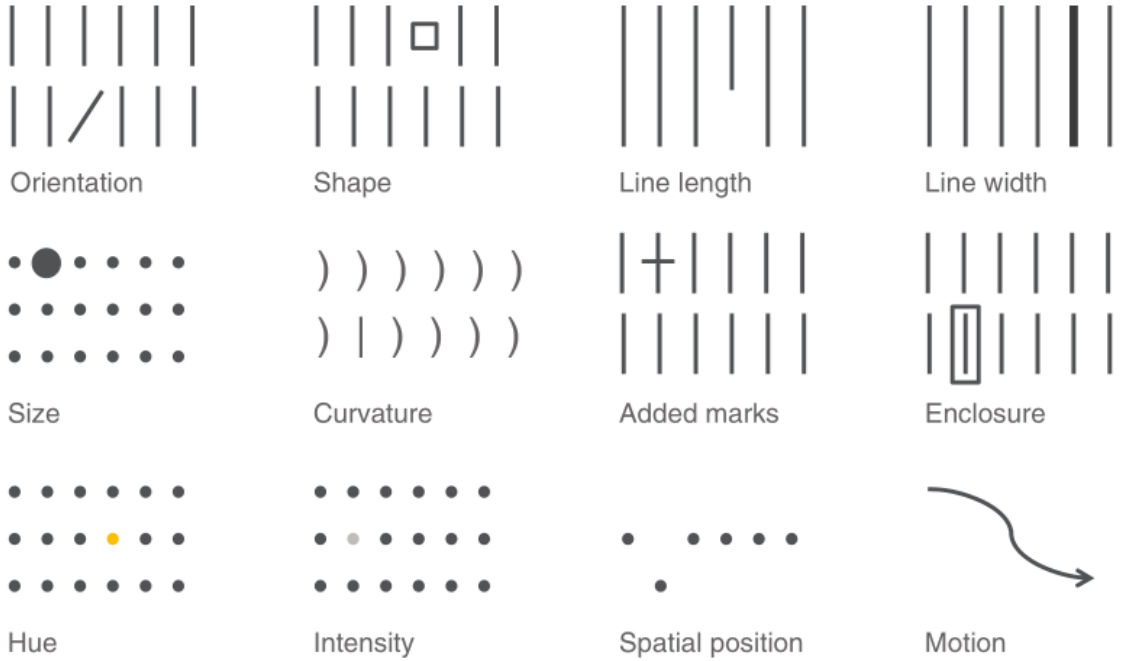


FIGURE 4.4 Preattentive attributes

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

Skoltech
Skolkovo Institute of Science and Technology

Example of preattentive attributes in text

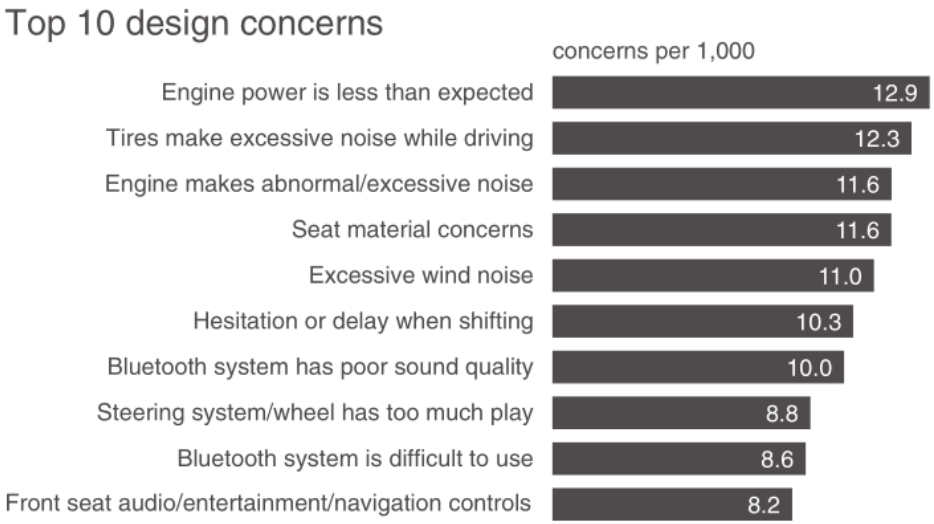
What are we doing well? Great Products. These products are clearly the best in their class. Replacement parts are shipped when needed. You sent me gaskets without me having to ask. Problems are resolved promptly. Bev in the billing office was quick to resolve a billing issue I had. General customer service 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

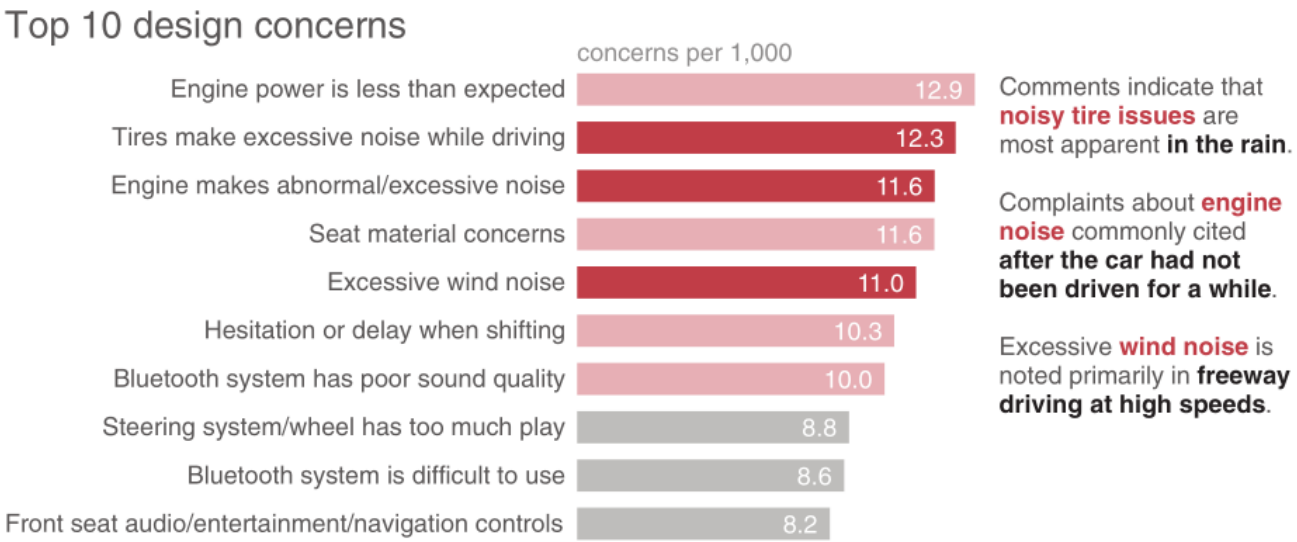
- **Great products**: "These products are clearly the best in class."
- **Replacement parts are shipped when needed**: "You sent me gaskets without me having to ask, and I really needed them, too!"
- **Problems are resolved promptly**: "Bev in the billing office was quick to resolve a billing issue I 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!*"

Example of preattentive attributes in graphs



Example of preattentive attributes in graphs

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



Skoltech
Skolkovo Institute of Science and Technology

Other preattentive attributes

- Size
- Color
- Position on page

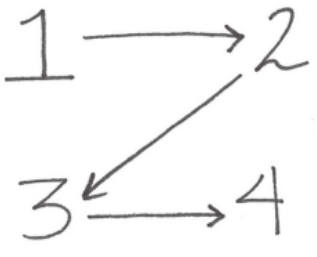
Country Level Sales Rank Top 5 Drugs

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

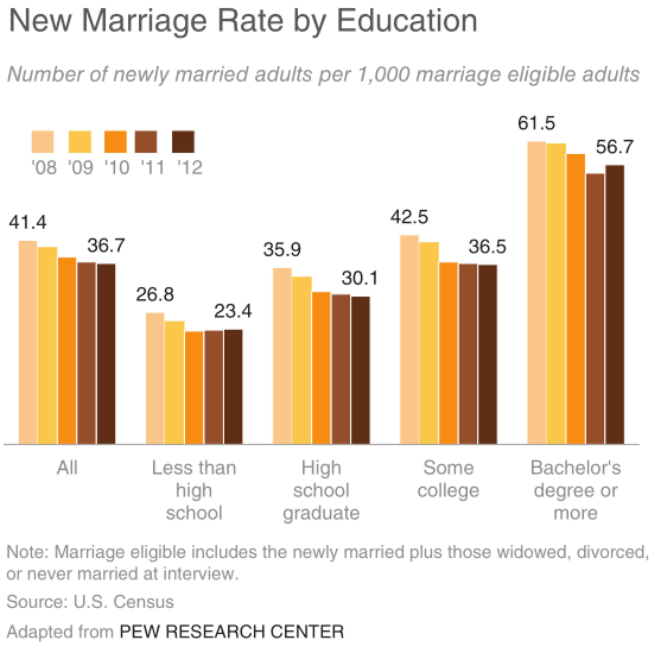
Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5

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

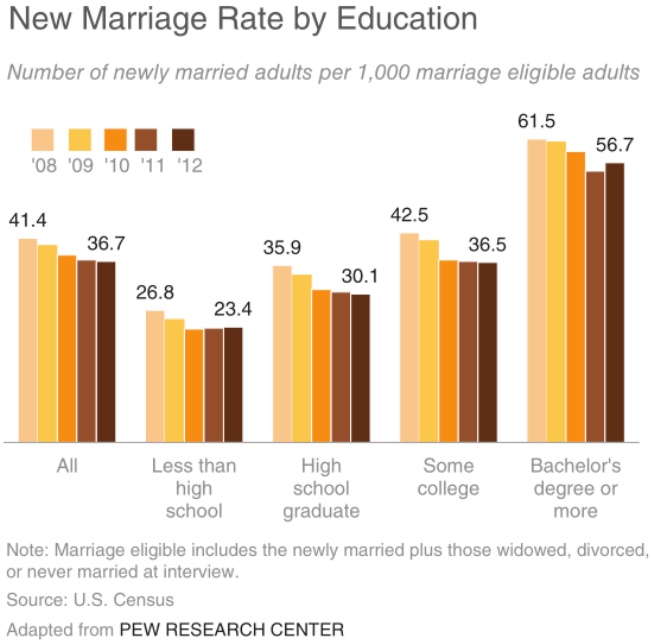


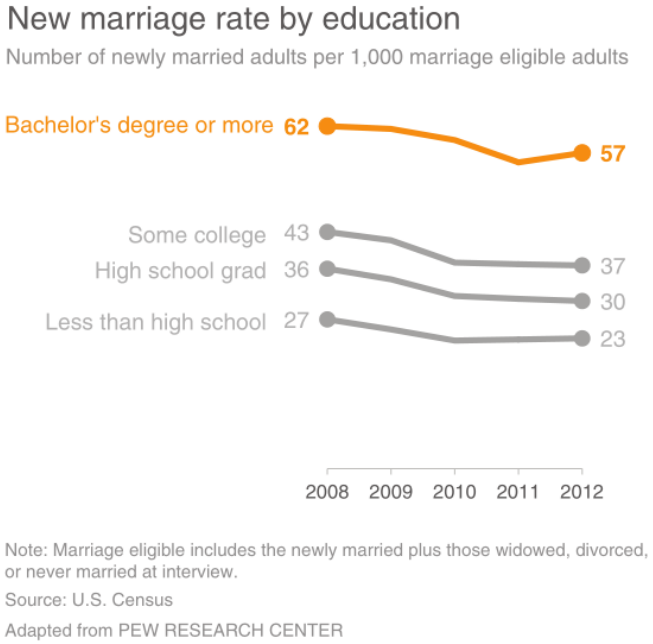
5. Think like a designer



Before

5. Think like a designer





Before

After

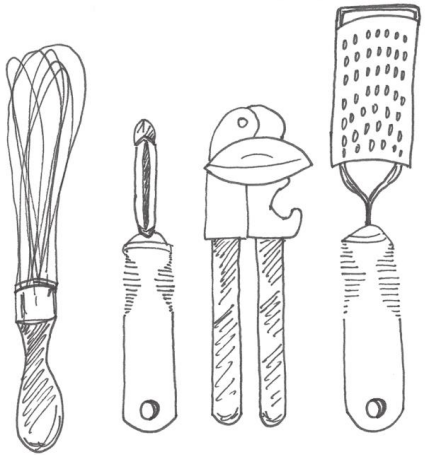
Skoltech
Skolkovo Institute of Science and Technology

5. Think like a designer

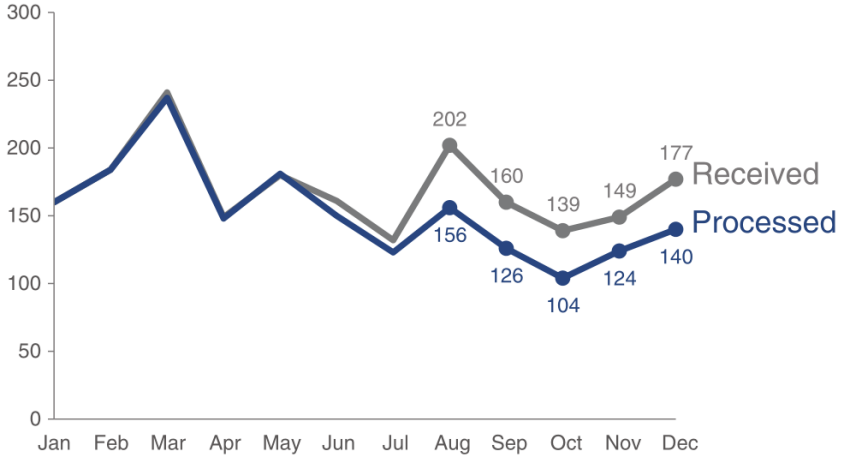
Form follows function. This adage of product design has clear application to communicating with data.

Three A-principles of design

- Affordances
- Accessibility
- Aesthetic

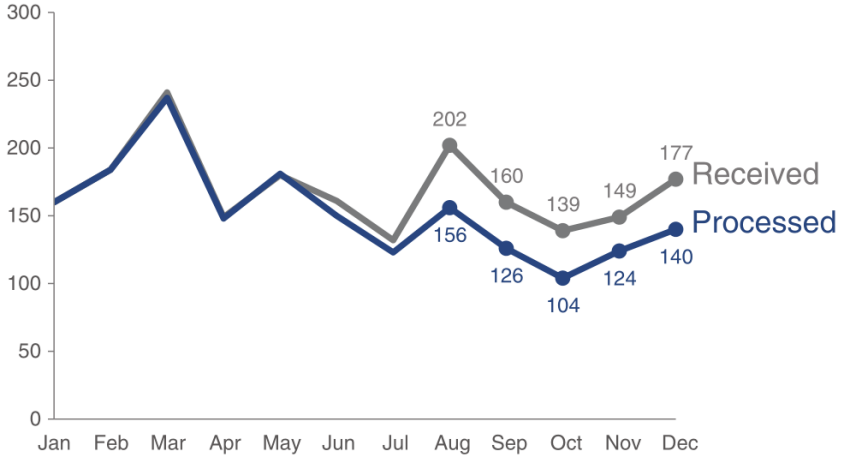


5. Think like a designer

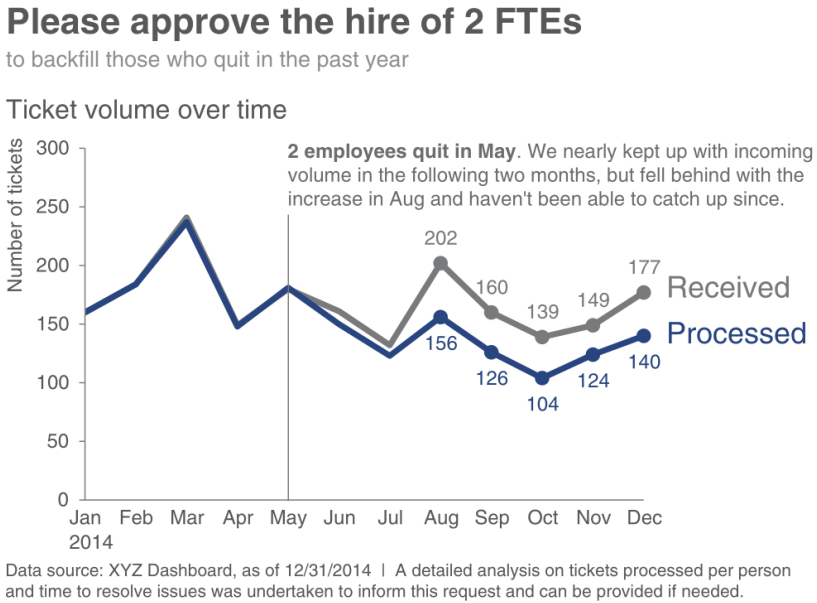


Before

5. Think like a designer



Before



After

Skoltech
Skolkovo Institute of Science and Technology

5. Tell a story

1. The setting: When and where does the story take place?
2. The main character: Who is driving the action? (This should be framed in terms of your audience!)
3. The imbalance: Why is it necessary, what has changed?
4. The balance: What do you want to see happen?
5. The solution: How will you bring about the changes?

Beyond Bullet Points, Cliff Atkinson

Three arcs structure

The beginning:
problem introduced

The middle:
how to solve the problem

The end:
call to action
tie it back to the beginning

A quick intro on 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