Python初級數據分析員證書

（六）數據分析 及可視化專案

14. Storytelling for Data Analysis

Recap

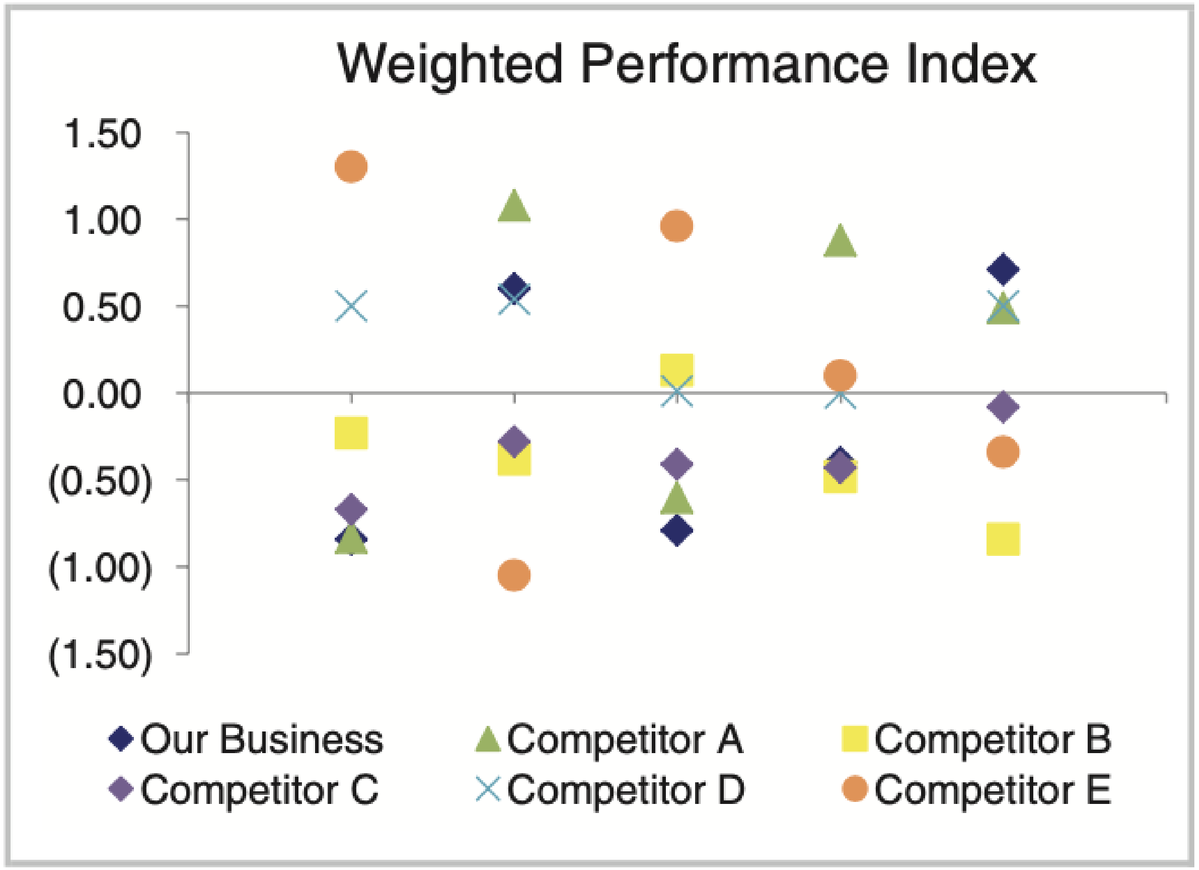
In previous chapters we had leant:

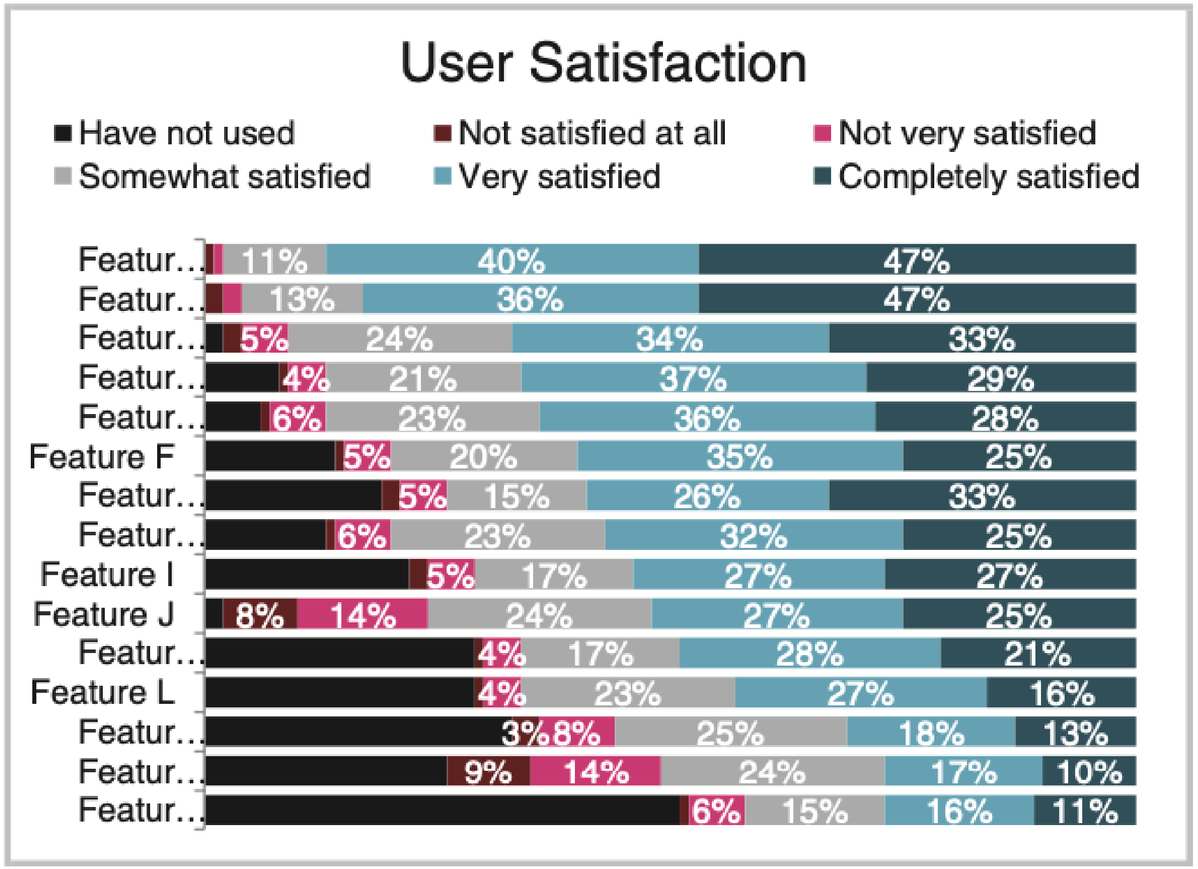
- Statistics
- Algebra
- Python
- R
- Pandas, NumPy, SciPy
- Data Analysis
- Data Visualization, Matplotlib, Seaborn, Plotly
- Dashboard Visualization, Business Intelligence

Chapter Summary

- Storytelling - Context
- Choosing effective visuals
- Tidiness everything
- Focus your audience’s attention
- Use colour wisely
- Storytelling
- Case Study: MacroMicro

Sample of ineffective charts





Storytelling - Context

Before you begin down the path of creating a data visualization or communication, attention and time should be paid to understanding the **context** for the need to communicate.

There are two kinds of analysis:

- Exploratory – technical research on data for peers in same field
- Explanatory – understandable data information consumed by general audience

In our course we mainly focus on explanatory and communication.

Who, What, How

When it comes to explanatory analysis, there are a few things to think about and be extremely clear on: who, what, and how

Who

- internal and external stakeholders?
- anyone who might be interested?
- knowledge level to the data and related area?

Who, What, How

What

- What do you need your audience to know or do?
- What is their needs for analysing the data?
- What is the current pain point?
- What is their expectation of the analysis?

Who, What, How

How

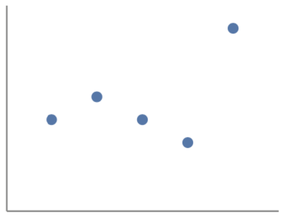
- How will you communicate to your audience(presentation, docs)?
- How detail the information shall be given?
- Suggesting possible next steps can be a great way to get the conversation going

Live presentation tips

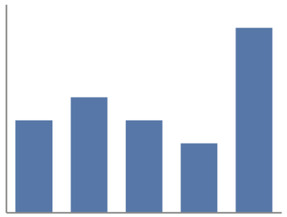
- Write out speaking notes with the important points you want to make with each slide.
- Practice what you want to say out loud to yourself: this ignites a different part of the brain to help you remember your talking points. It also forces you to articulate the transitions between slides that sometimes trip up presenters.
- Give a mock presentation to a friend or colleague.
- Record your mock presentation and review for modification.

Choosing effective visual

Simple text



Scatterplot



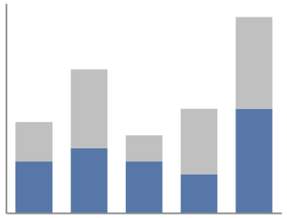
Vertical bar

Horizontal bar

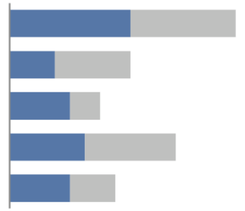
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Table

Line



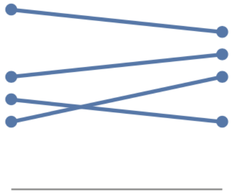
Stacked vertical bar



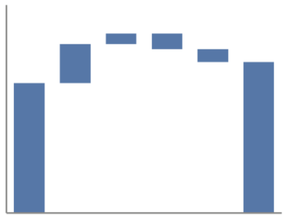
Stacked horizontal bar

	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%

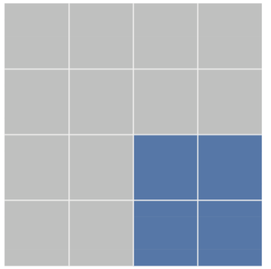
Heatmap



Slopegraph



Waterfall



Square area

Choosing effective visual - Simple text

When you have just a number or two to share, simple text can be a great way to communicate. You may enlarge the critical number for deep impact.

Choosing effective visual - Table

Tables are great for just that—communicating to a mixed audience whose members will each look for their particular row of interest.

- When audience read the table, you lose their ears and attention.
- If you must show a table to audience, ask a question like “what interesting thing do you find in the table?” and give them 20 seconds.
- When I have a table in front of me, I typically have my index finger out: I’m reading across rows and down columns or I’m comparing values.

Choosing effective visual – Heat Map

One approach for mixing the detail you can include in a table while also making use of visual cues is via a heatmap.

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

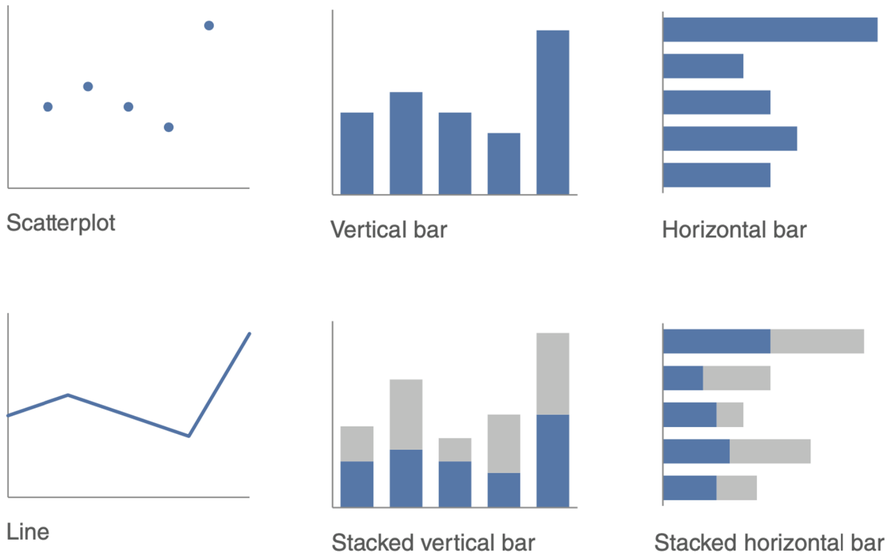
LOW-**HIGH**

	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%

Choosing effective visual – Graph

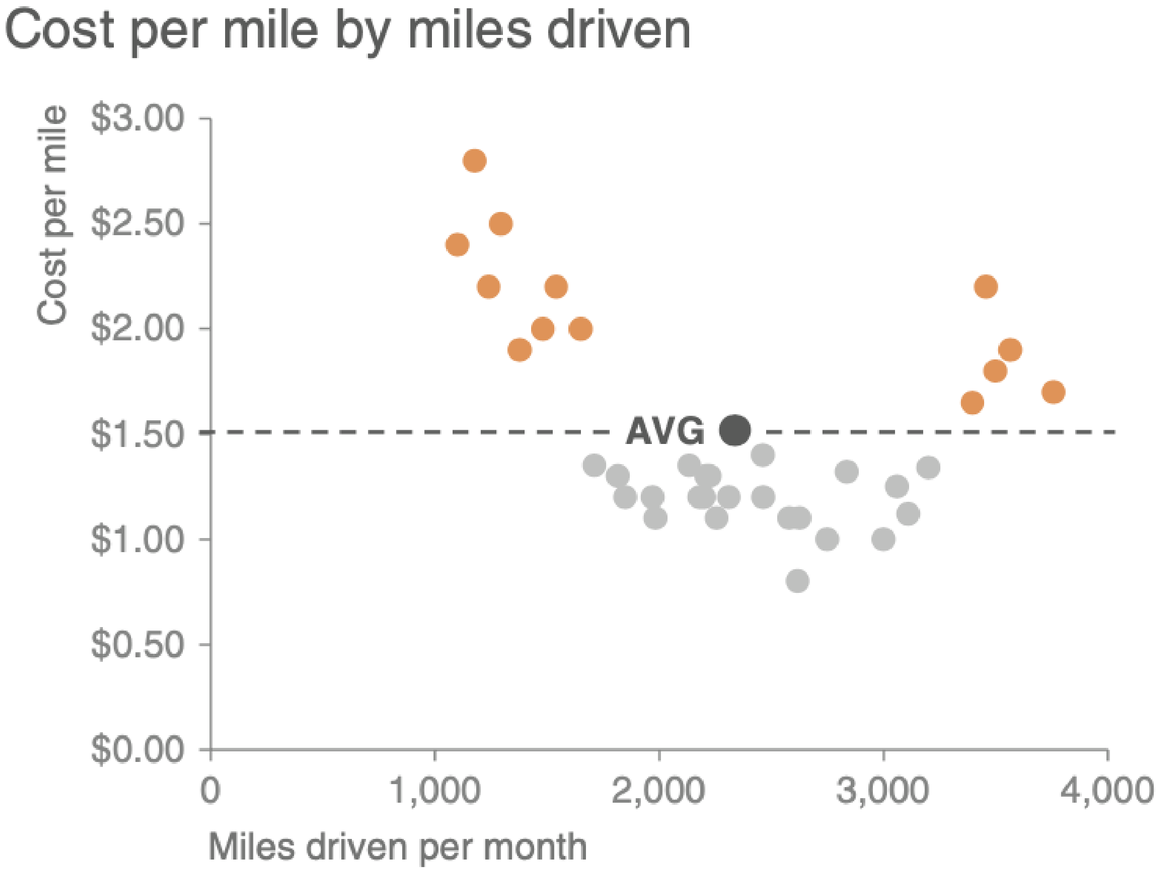
A well-designed graph will typically get the information across more quickly than a well-designed table.

While tables interact with our verbal system, graphs interact with our visual system, which is faster at processing information.



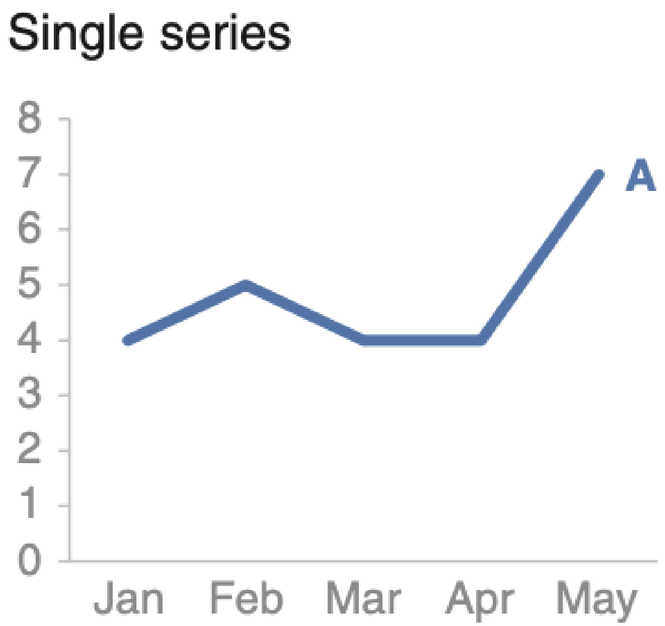
Choosing effective visual – Scatterplot

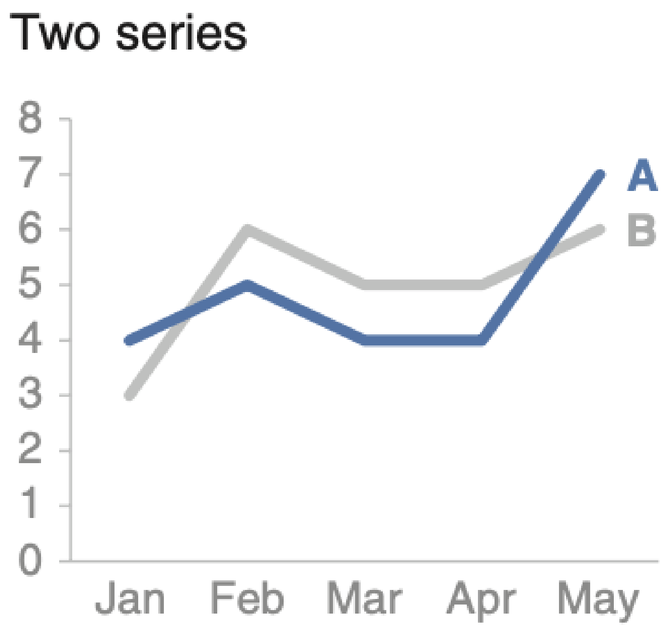
Scatterplots can be useful for showing the relationship between two things, because they allow you to encode data simultaneously on a horizontal *x*-axis and vertical *y*-axis to see whether and what relationship exists.

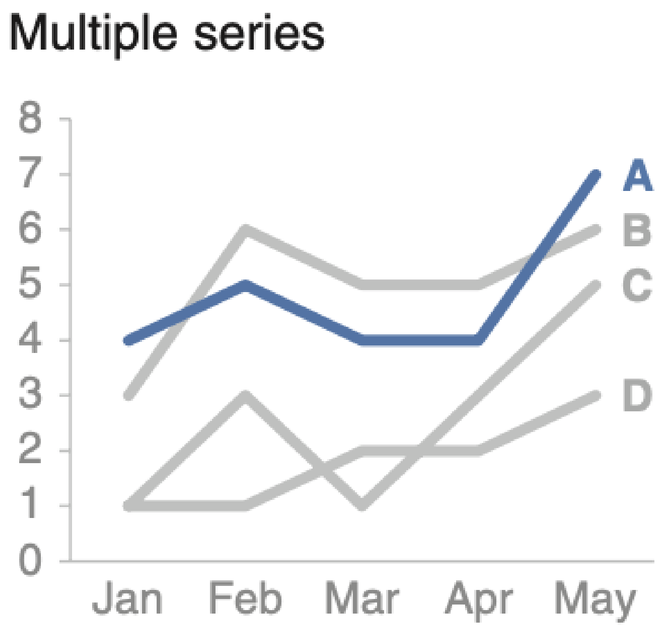


Choosing effective visual – Line graph

Line graphs are most commonly used to plot continuous data. Because the points are physically connected via the line, it implies a connection between the points that may not make sense for categorical data



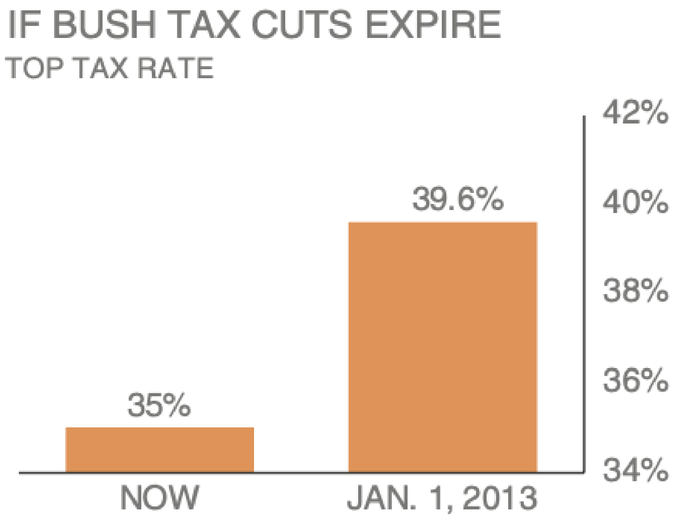




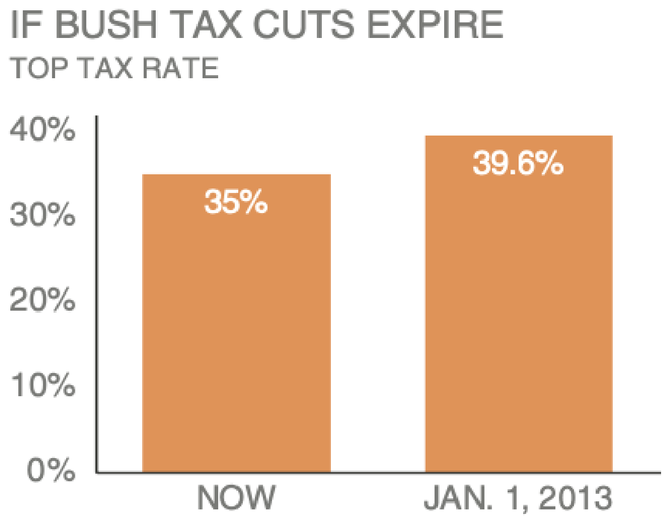
Choosing effective visual – Bar Chart

Bar charts are easy for our eyes to read. Our eyes compare the end points of the bars, so it is easy to see quickly which category is the biggest, which is the smallest, and also the incremental difference between categories.

Non-zero baseline: as originally graphed

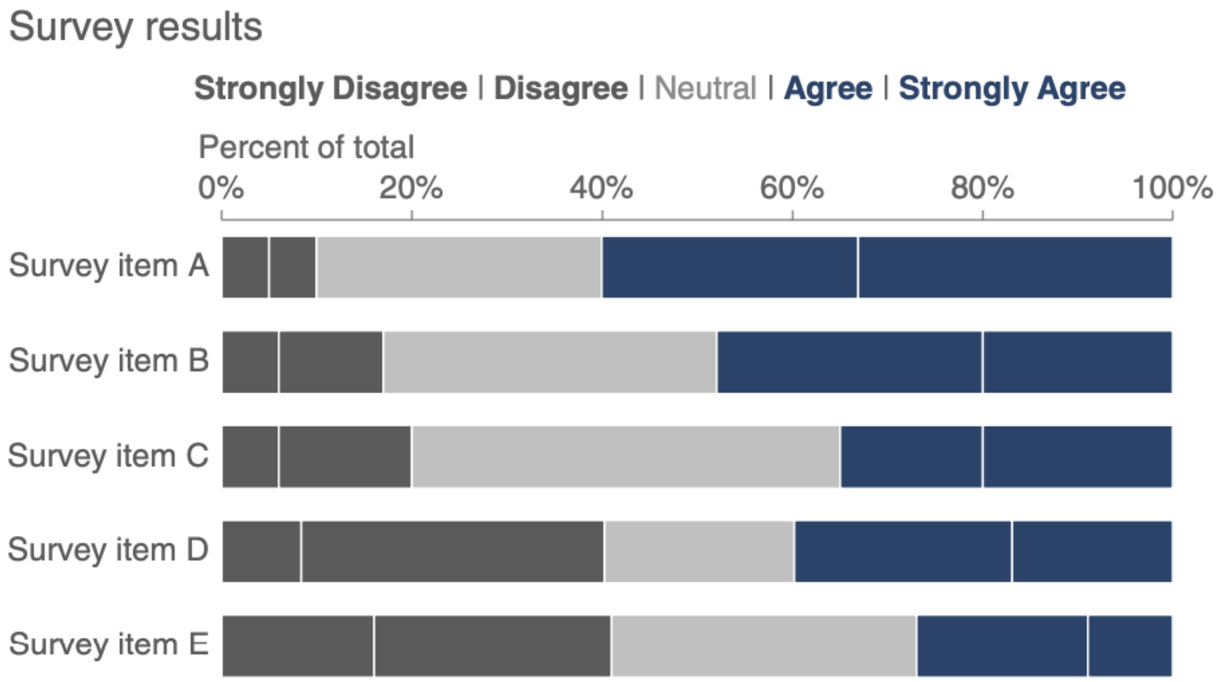


Zero baseline: as it should be graphed



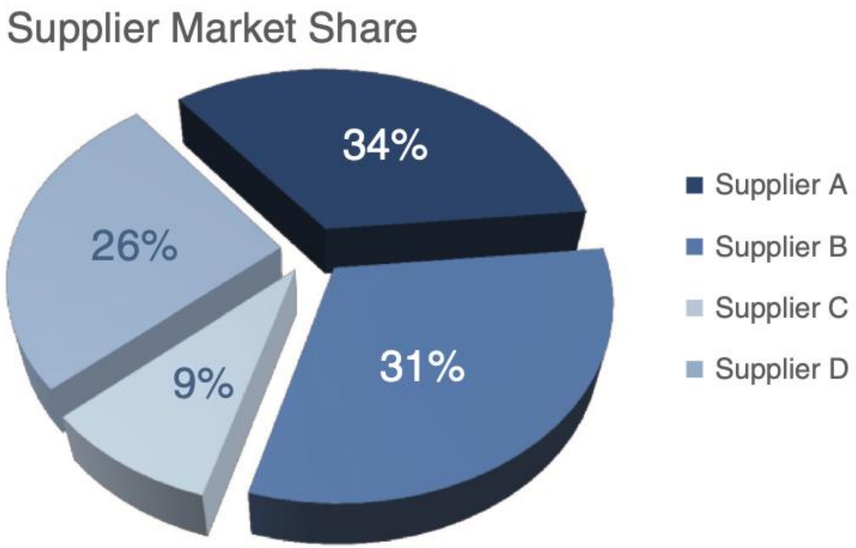
Stacked horizontal bar charts

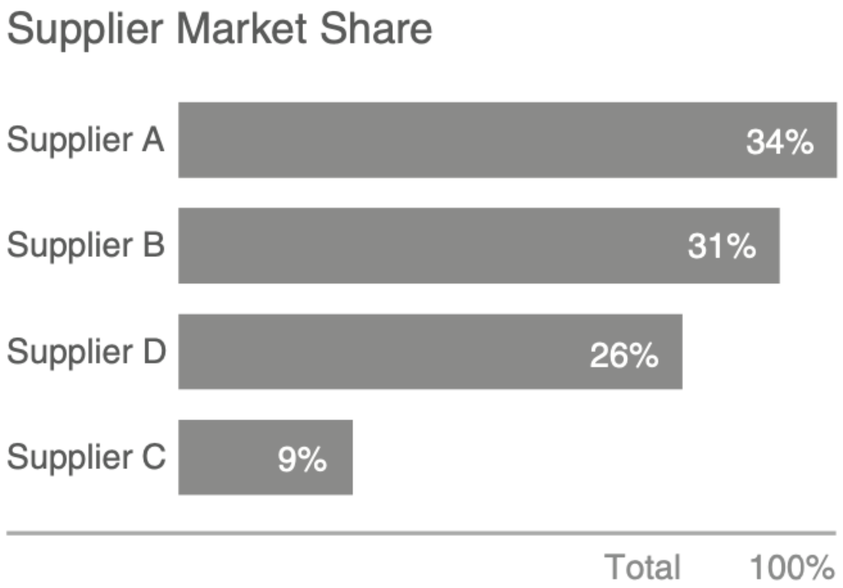
Stacked horizontal bar charts can be used to show the totals across different categories but also give a sense of the subcomponent pieces. They can be structured to show either absolute values or sum to 100%.



Some charts to avoid

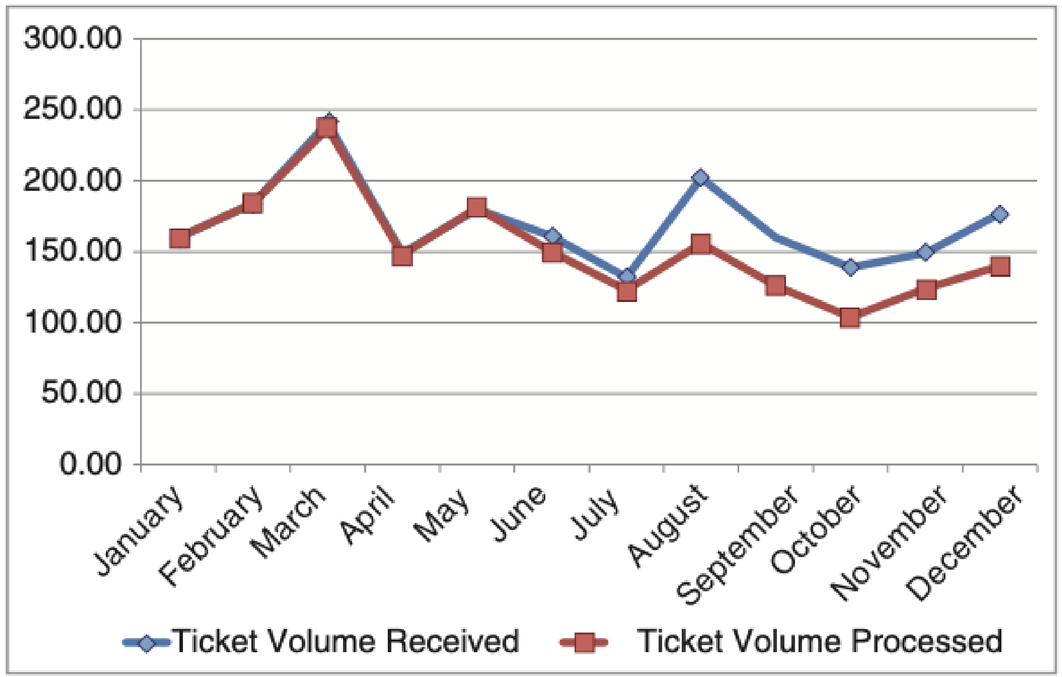
There are also some specific graph types and elements that you should avoid: pie charts, donut charts, 3D, and secondary *y*-axes. As it takes more time to understand them.

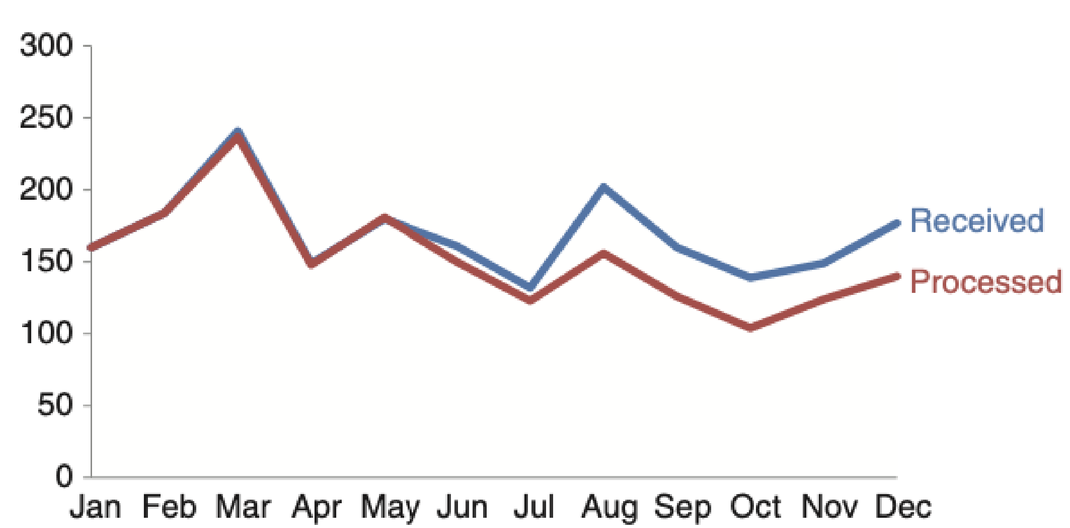


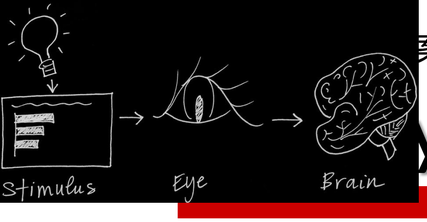


Tidiness everything

Simplicity is always the best. Remove unnecessary details and information to make your audience read quickly. No fancy colour and style.

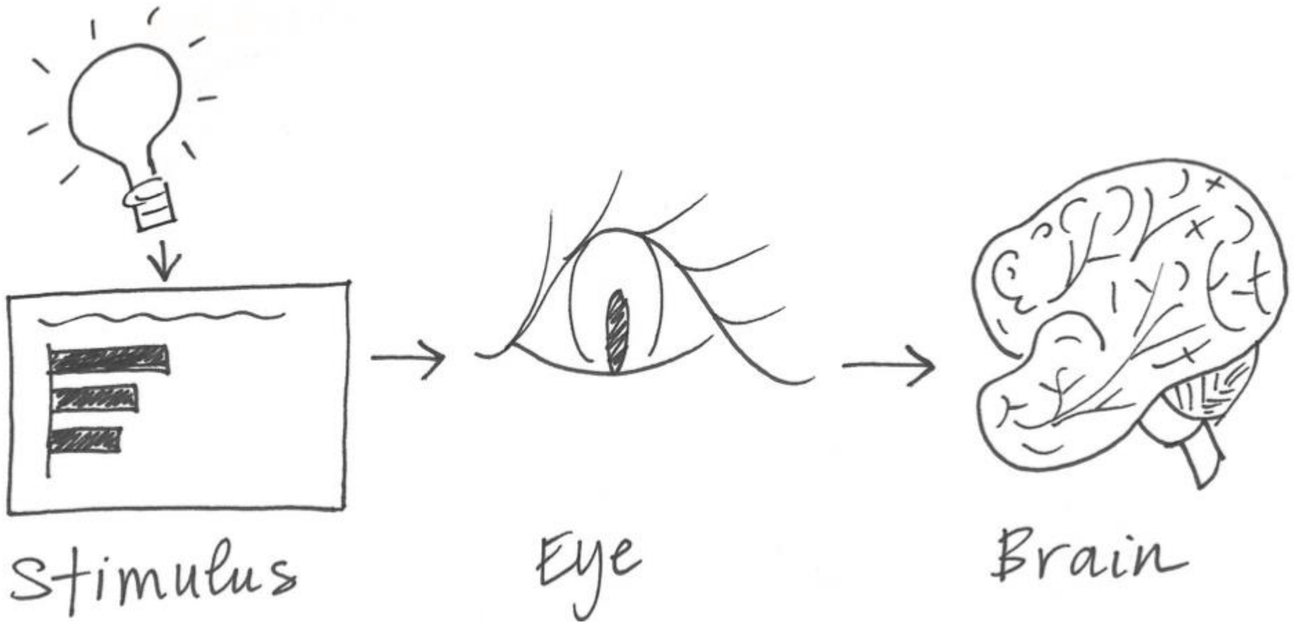






your audience's attention

A simplified picture of how people see. Light reflects off of a stimulus. This gets captured by our eyes. We don't fully see with our eyes; there is some processing that happens there, but mostly it is what happens in our brain.



Focus your audience's attention

Within the brain, there are three types of memory that are important to understand as we design visual communications: iconic memory, short-term memory, and long-term memory.

Iconic memory

Iconic memory is super fast. Information stays in your iconic memory for a fraction of a second before it gets forwarded on to your short-term memory.

Focus your audience's attention

Short-term memory

People can keep about four chunks of visual information in their short-term memory. If we create a graph with ten different data series that are ten different colours with ten different shapes of data markers and a legend off to the side, we're making our audience work very hard going back and forth.

Focus your audience's attention

Long-term memory

When something leaves short-term memory, it either goes into junk and is likely lost forever, or is passed into long-term memory.

Long-term memory is built up over a lifetime and is vitally important for pattern recognition and general cognitive processing.

Let's test our brain in next slide. In next slide, you will see a block of numbers, you are asked to count the number of **3** as quickly as possible. And count your time to do this task.

Focus your audience’s attention

75639506847 3

6586630375 76

8603726586 02

8465891078 30

Focus your audience's attention

The correct answer is six. There were no visual cues to help you reach this conclusion.

Check out what happens when we make a single change to the block of numbers. Turn the next slide and repeat the exercise of counting the 3.

Focus your audience's attention

7563950684733

6586630375576

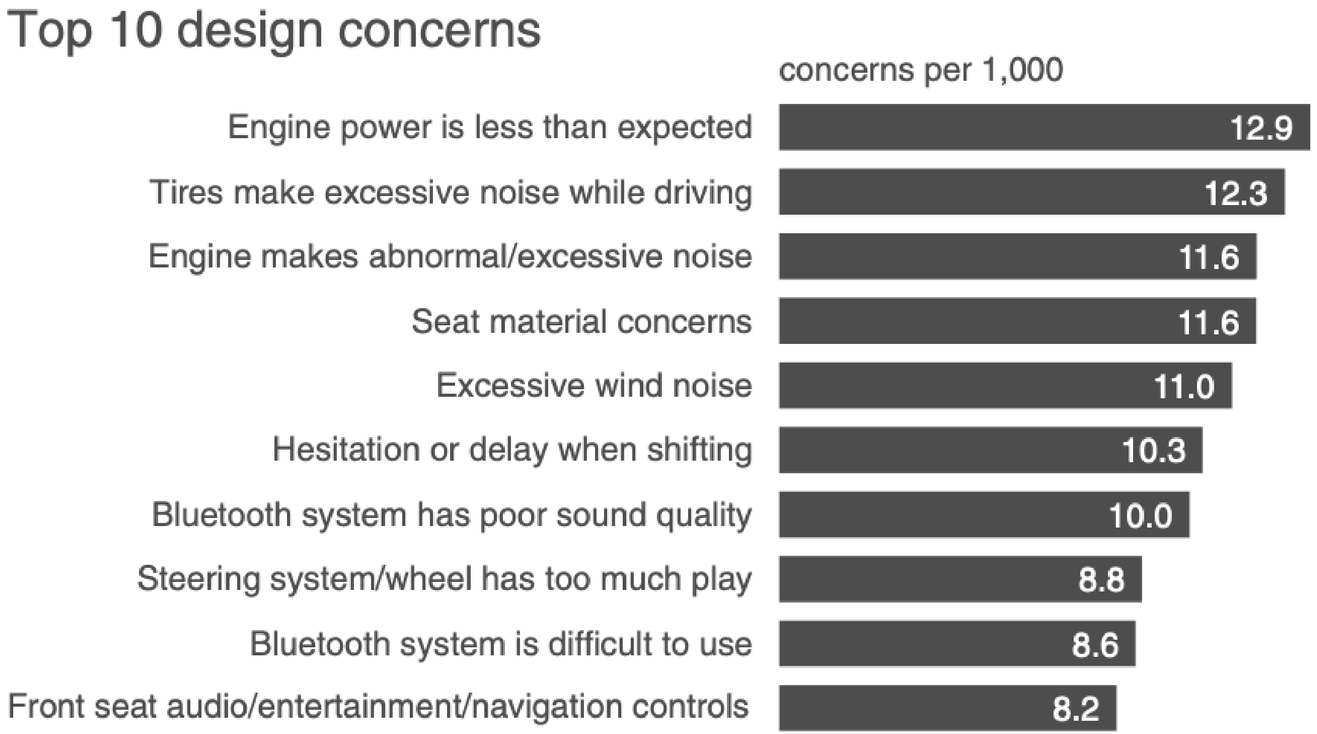
8603726586602

8465891078330

Focus your audience's attention

Note how much easier and faster the same exercise is using the last slide.

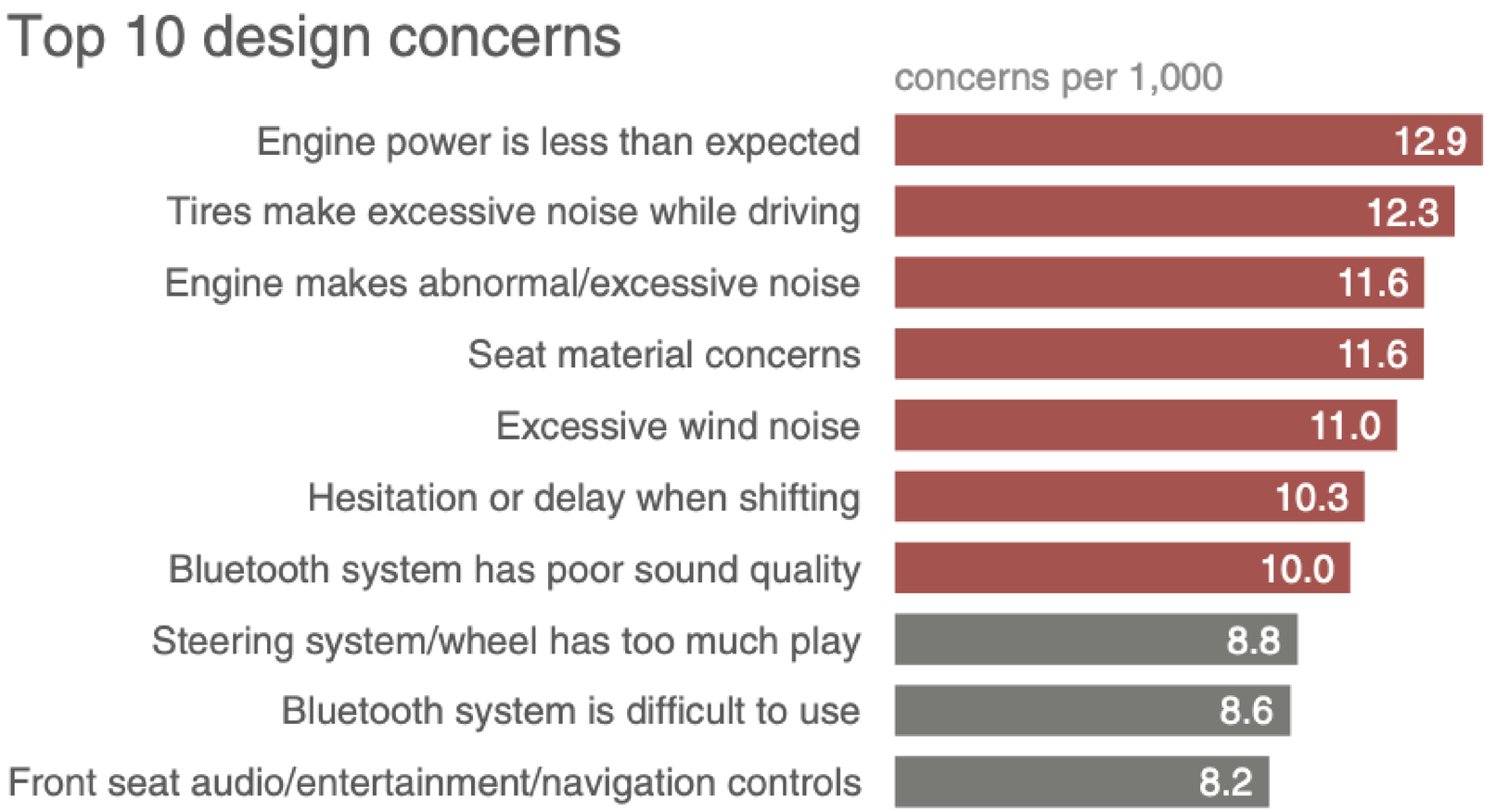
his is so apparent so quickly because in this second iteration, your iconic memory is being leveraged. Let's see an example graph



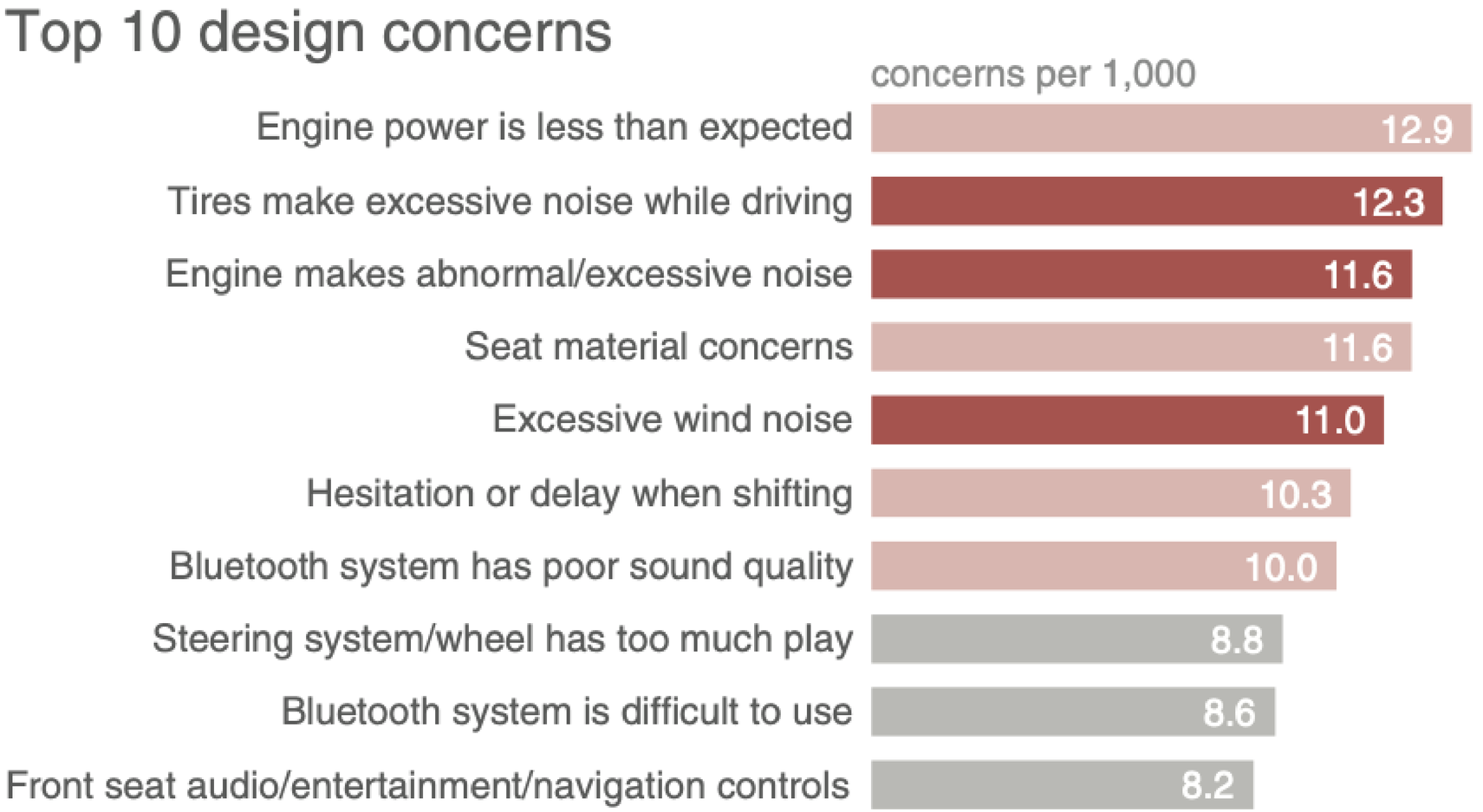
Focus your audience's attention

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

Discussion: is this an acceptable default rate?



Focus your audience's attention



Comments indicate that **noisy tire issues** are most apparent **in the rain**.

Complaints about **engine noise** commonly cited **after the car had not been driven for a while**.

Excessive **wind noise** is noted primarily in **freeway driving at high speeds**.

Use colour wisely

Rainbow colour may not help the audience understand the chart quickly.

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

Storytelling

- **Find a subject you care about. It is this genuine caring, and not your games with language, which will be the most compelling and seductive element in your style.**
- **Do not ramble, though.**
- **Keep it simple. Great masters wrote sentences which were almost childlike when their subjects were most profound. “To be or not to be?” asks Shakespeare’s Hamlet. The longest word is three letters.**

Storytelling

- **Have the guts to cut.** If a sentence, no matter how excellent, does not illuminate your subject in some new and useful way, scratch it out.
- **Sound like yourself.** I myself find that I trust my own writing most, and others seem to trust it most, too, when I sound most like a person from Indianapolis, which is what I am.

Storytelling

- **Say what you meant to say.** If I broke all the rules of punctuation, had words mean whatever I wanted them to mean, and strung them together higgledy-piggledy, I would simply not be understood.
- **Pity the readers.** Our audience requires us to be sympathetic and patient teachers, ever willing to simplify and clarify. Interact with audience.

Final thought

Data visualization and communicating with data in general—sits at the intersection of science and art. There is certainly some science to it: best practices and guidelines to follow. But there is also an artistic component. This is one of the reasons this area is so much fun!

Conclusion

Final tips:

Tip #1: learn your tools well

Tip #2: iterate and seek feedback

Tip #3: devote time to storytelling with data

Tip #4: seek inspiration through good examples

Tip #5: have fun and find your style