18CSE490T - BIG DATA VISUALIZATION

Lecture by

Dr. K. Rajkumar
Department of Data Science and Business Systems
SRM Institute of Science and technology

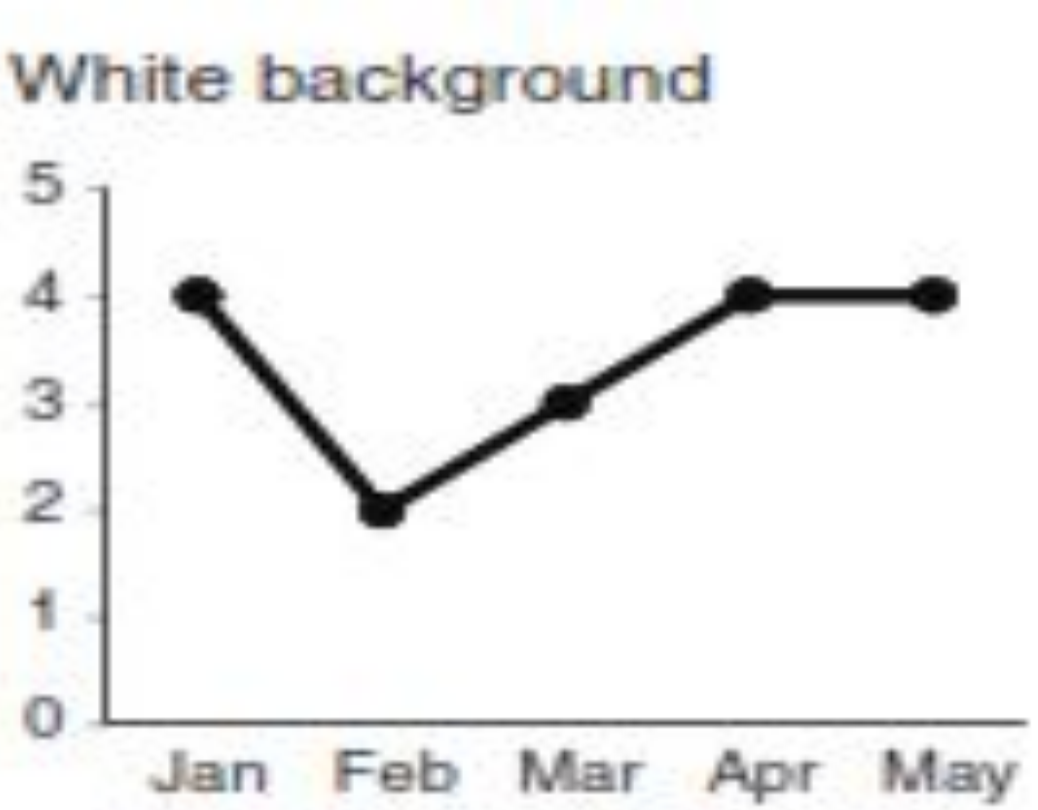
Case Studies

- To have a solid foundation for communicating effectively with data
- explore strategies for tackling common challenges faced when communicating with data through a number of case studies

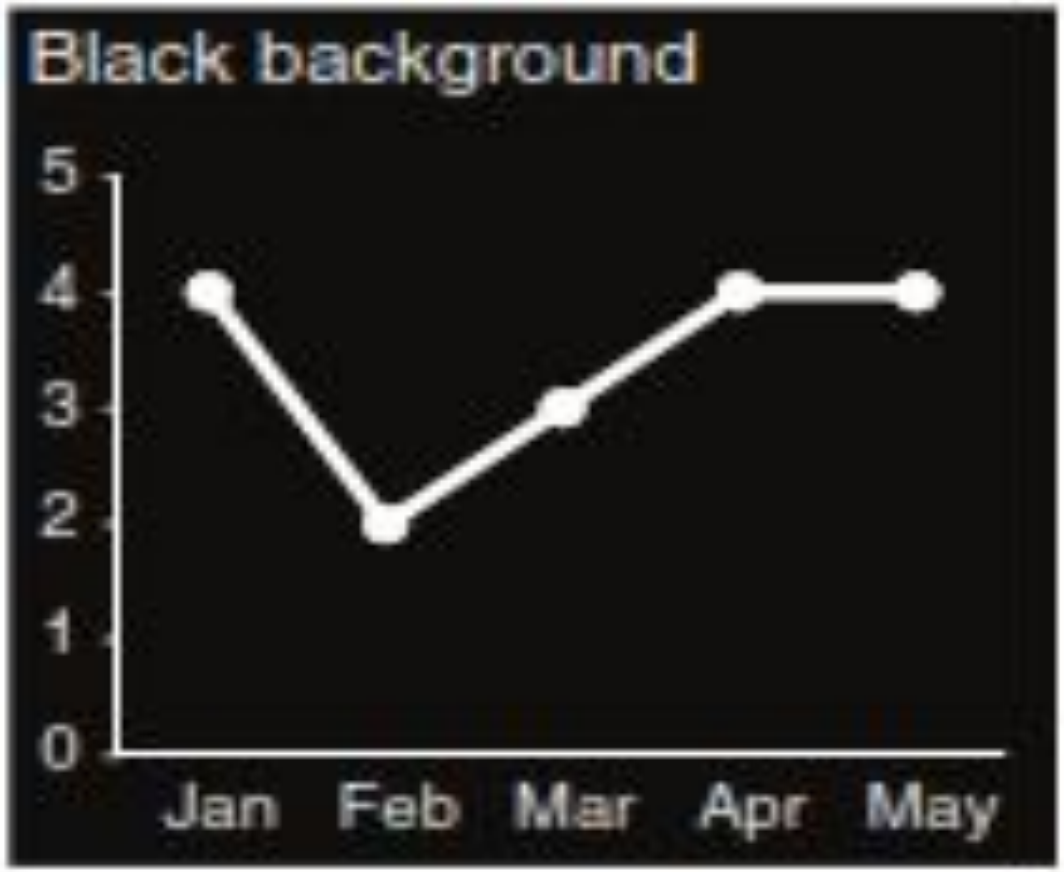
To be discussed

- Color considerations with a dark background
- Leveraging animation in the visuals you present
- Establishing logic in order
- Strategies for avoiding the spaghetti graph
- Alternatives to pie charts

CASE STUDY 1: Color considerations with a dark background




Blue background
5
4
3
2
1
0
Jan
Feb
Mar
Apr
May


Black background
5
4
3
2
1
0
Jan
Feb
Mar
Apr
May

- White background – Easy to focus on data
- Dark background – Focus more on background; Focus away from data
- Light Element on Dark Background – Create stronger contrast;

  Harder to read
- Avoid dark colored backgrounds
- sometimes there are considerations outside of the ideal scenario for communicating with data that must be taken into account,
- such as your company or client’s brand and corresponding standard template.

Good illustration of how color can impact the overall tone of a visualization

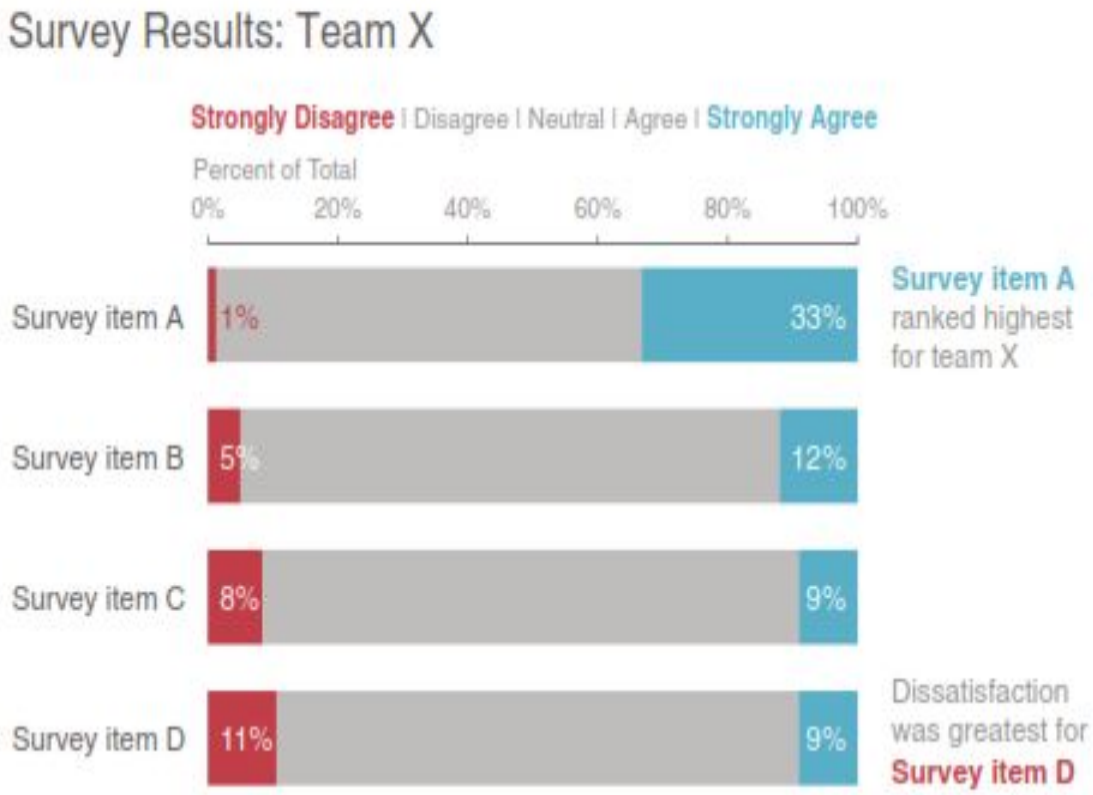
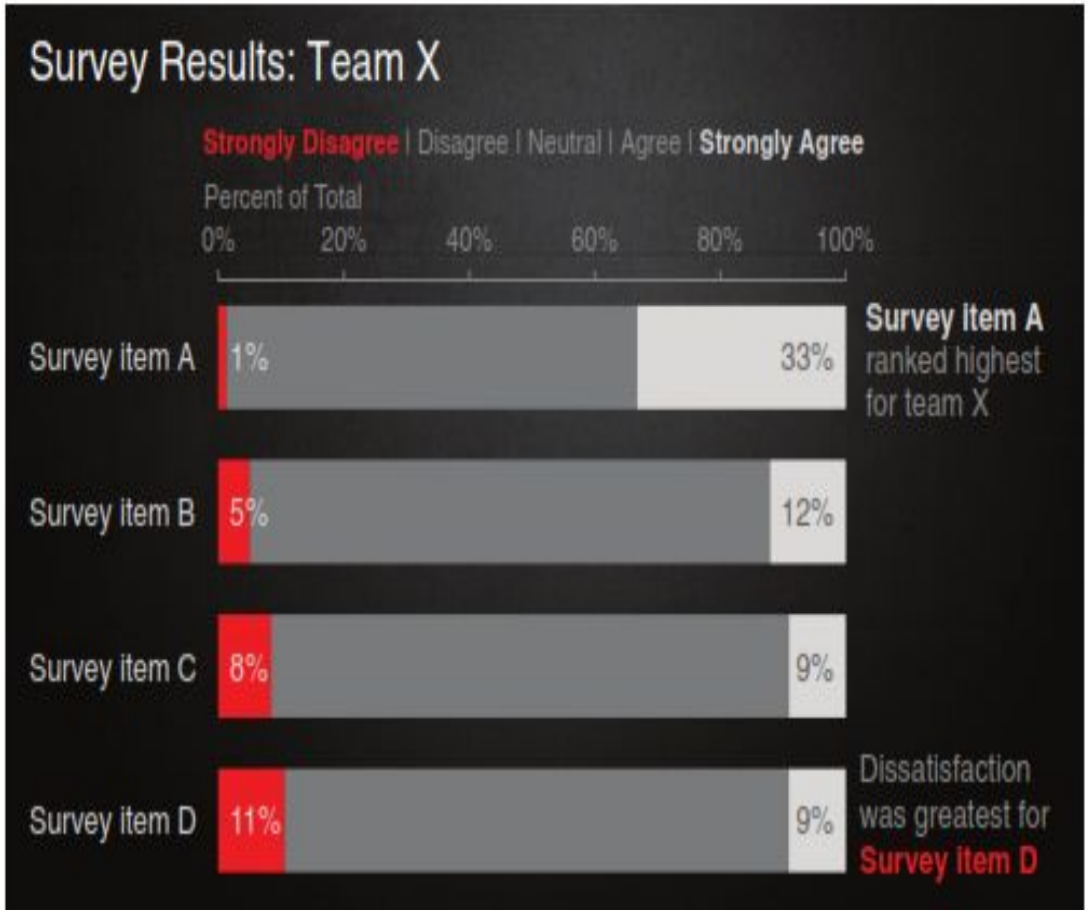


FIGURE 9.2 Initial makeover on white background



CASE STUDY 2: Leveraging animation in the visuals you present

- Let’s assume that you work for a company that makes online social games.
- You are interested in telling the story around how active users for a given game—let’s call it Moonville—have grown over time

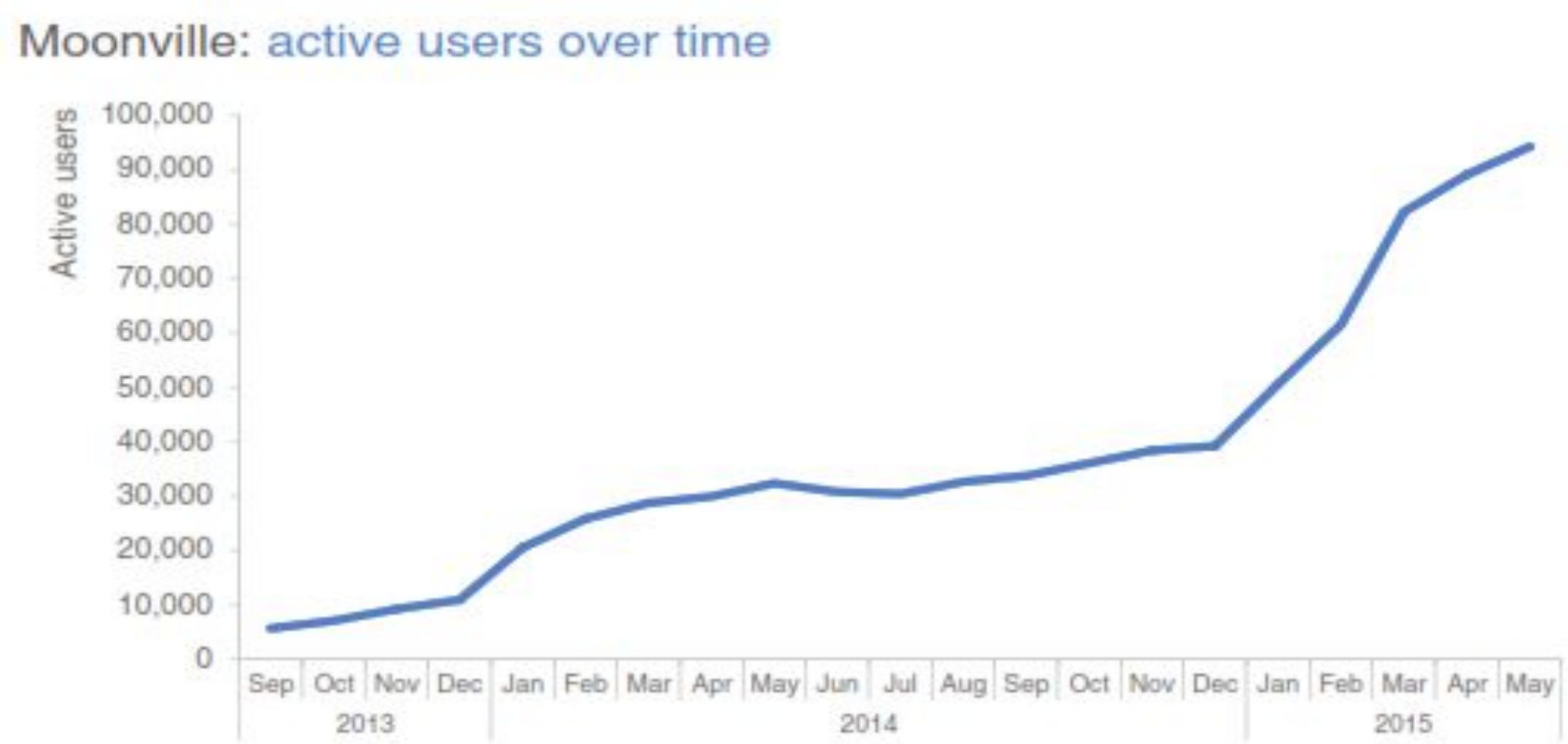


FIGURE 9.4 Original graph

The challenge, is that when you put this much data in front of your audience,

- you lose control over their attention
- You might be talking about one part of the data while they are focusing somewhere else entirely
- you want to tell the story chronologically,
- but your audience may jump immediately to the sharp increase in 2015 and wonder what drove that.
- When they do so, they stop listening to you

Moonville: active users over time

Data source: ABC Report. For purpose of analysis "active user" is defined as the number of unique users in the past 30 days.

FIGURE 9.5

Moonville: active users over time

Data source: ABC Report. For purpose of analysis "active user" is defined as the number of unique users in the past 30 days.

FIGURE 9.6

Moonville: active users over time

Data source: ABC Report. For purpose of analysis "active user" is defined as the number of unique users in the past 30 days.

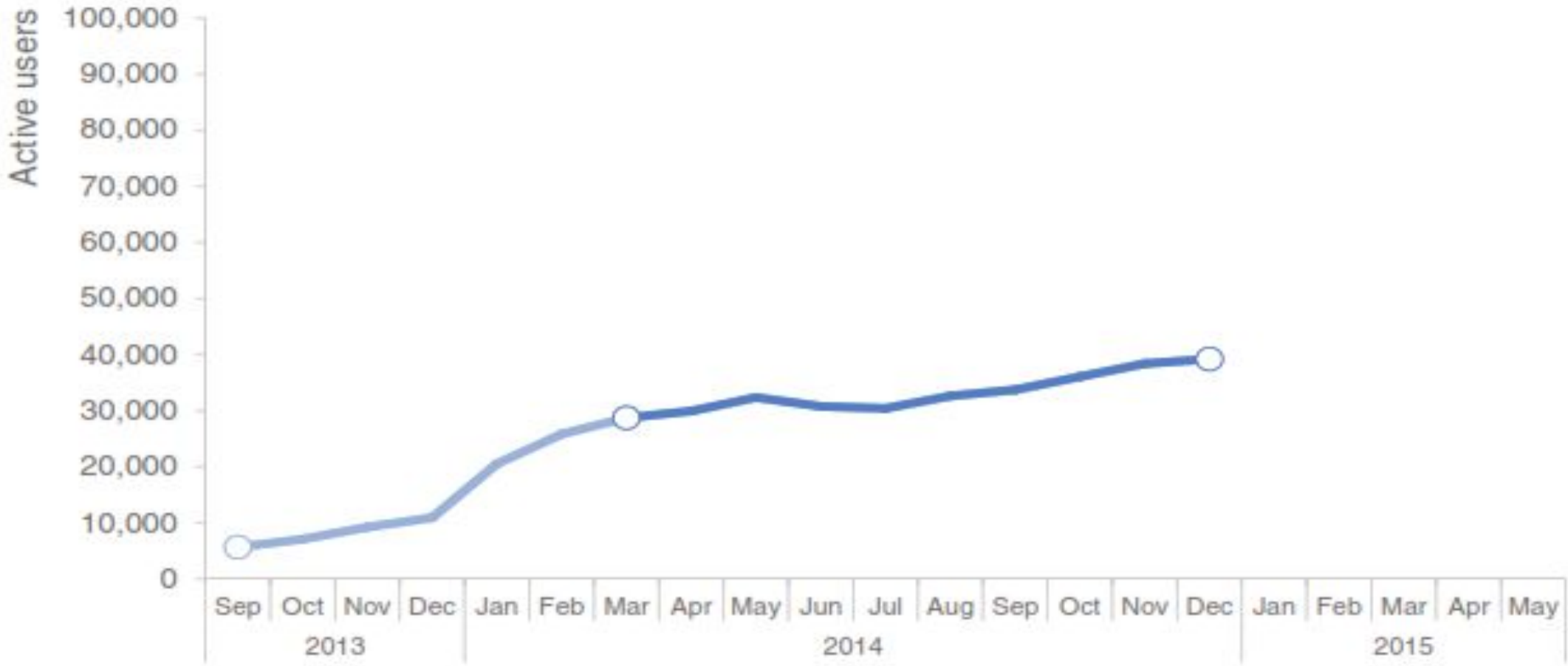
FIGURE 9.7

Moonville: active users over time

Data source: ABC Report. For purpose of analysis "active user" is defined as the number of unique users in the past 30 days.

FIGURE 9.8

Moonville: active users over time



Data source: ABC Report. For purpose of analysis "active user" is defined as the number of unique users in the past 30 days.

FIGURE 9.9

Moonville: active users over time

Data source: ABC Report. For purpose of analysis "active user" is defined as the number of unique users in the past 30 days.

FIGURE 9.10

Moonville: active users over time

Data source: ABC Report. For purpose of analysis "active user" is defined as the number of unique users in the past 30 days.

FIGURE 9.11

CASE STUDY 3: Logic in order

- Let’s say you work at a company that sells a product that has various features
- You’ve recently surveyed your users to understand whether they are using each of the features and
- How satisfied they’ve been with them and want to put that data to use.
- The initial graph you create might look something like Figure 9.12.

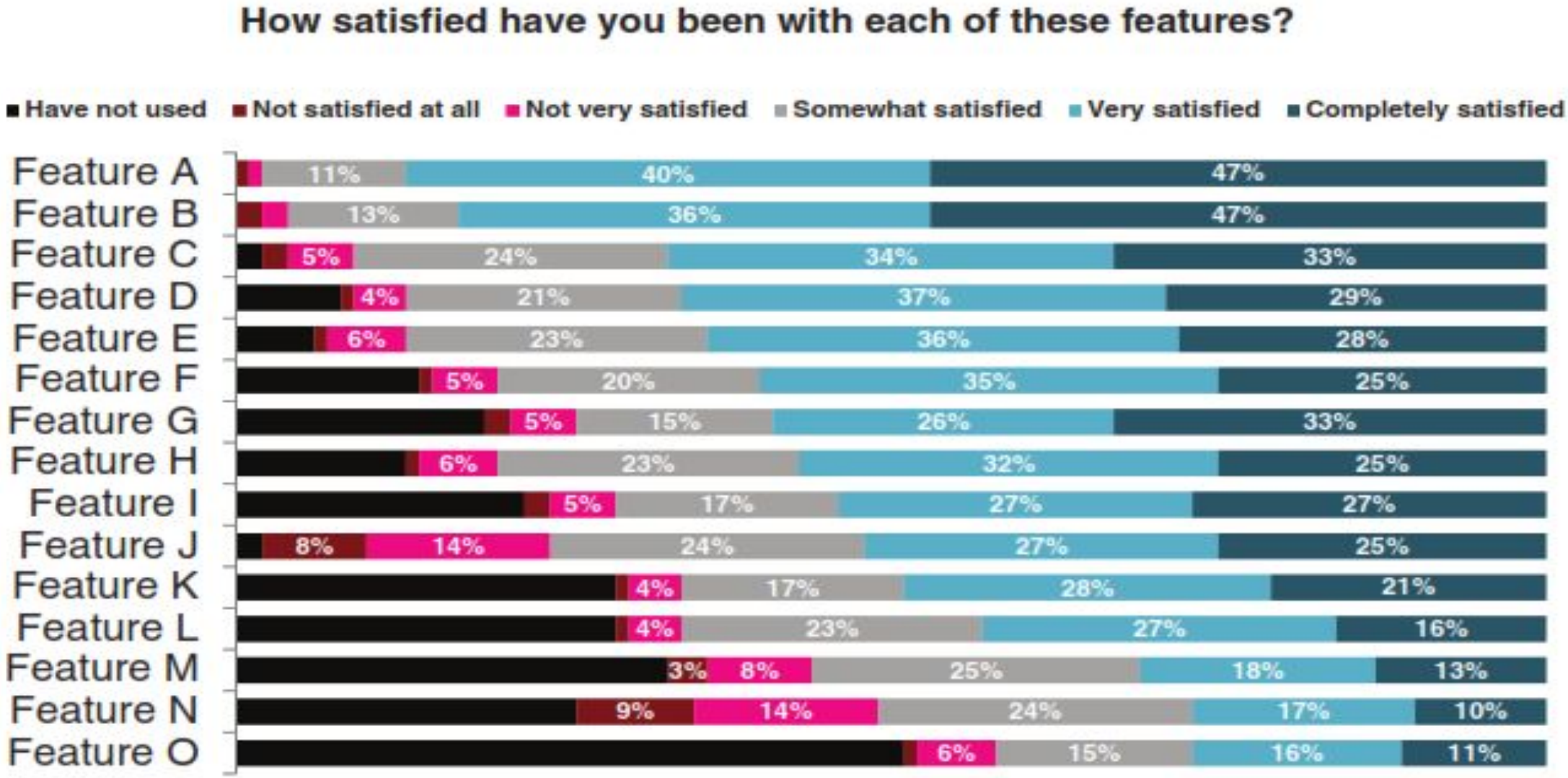
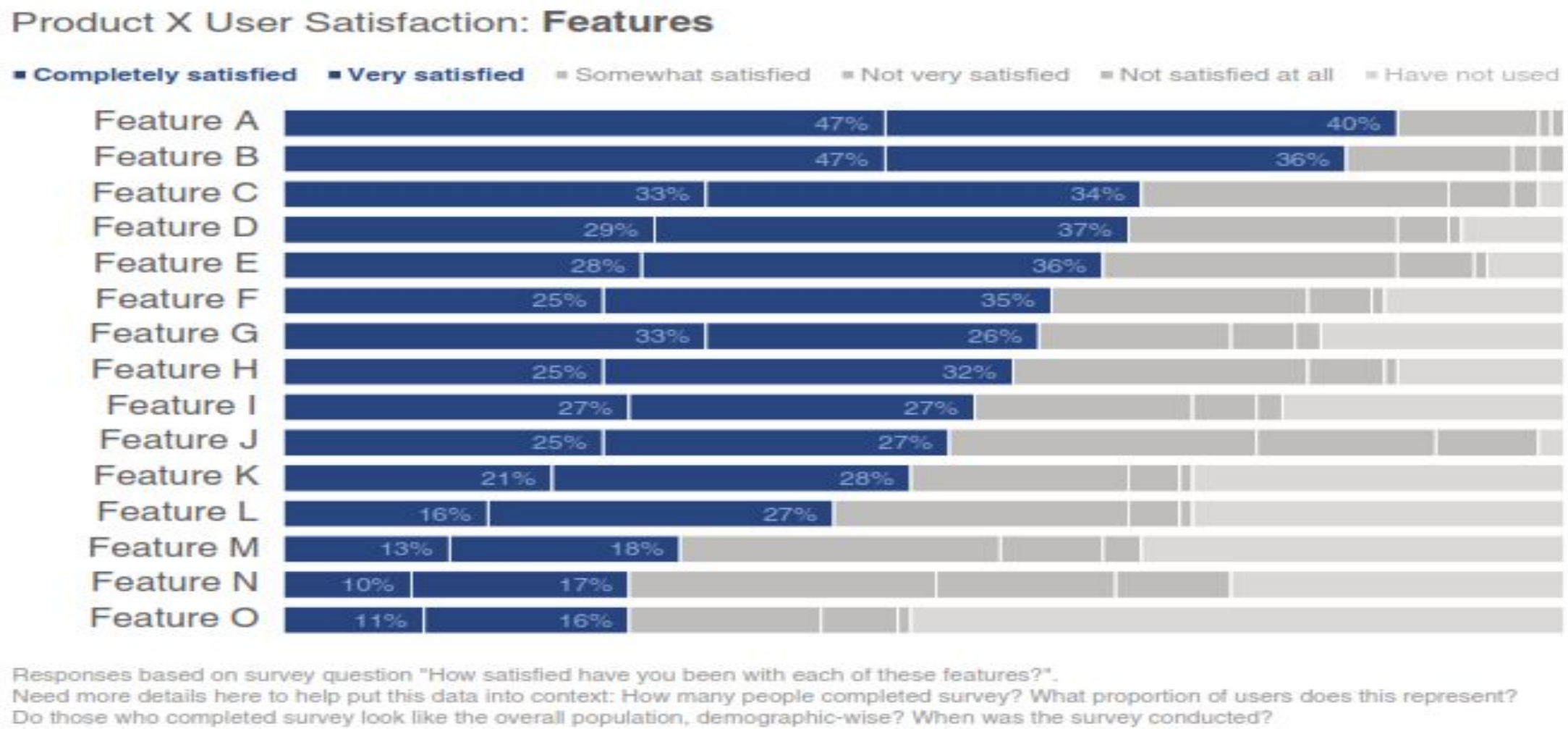


FIGURE 9.12 User satisfaction, original graph

Features A & B top user satisfaction



Responses based on survey question "How satisfied have you been with each of these features?".
Need more details here to help put this data into context: How many people completed survey? What proportion of users does this represent?
Do those who completed survey look like the overall population, demographic-wise? When was the survey conducted?

FIGURE 9.13 Highlight the positive story

Users least satisfied with Features N & J

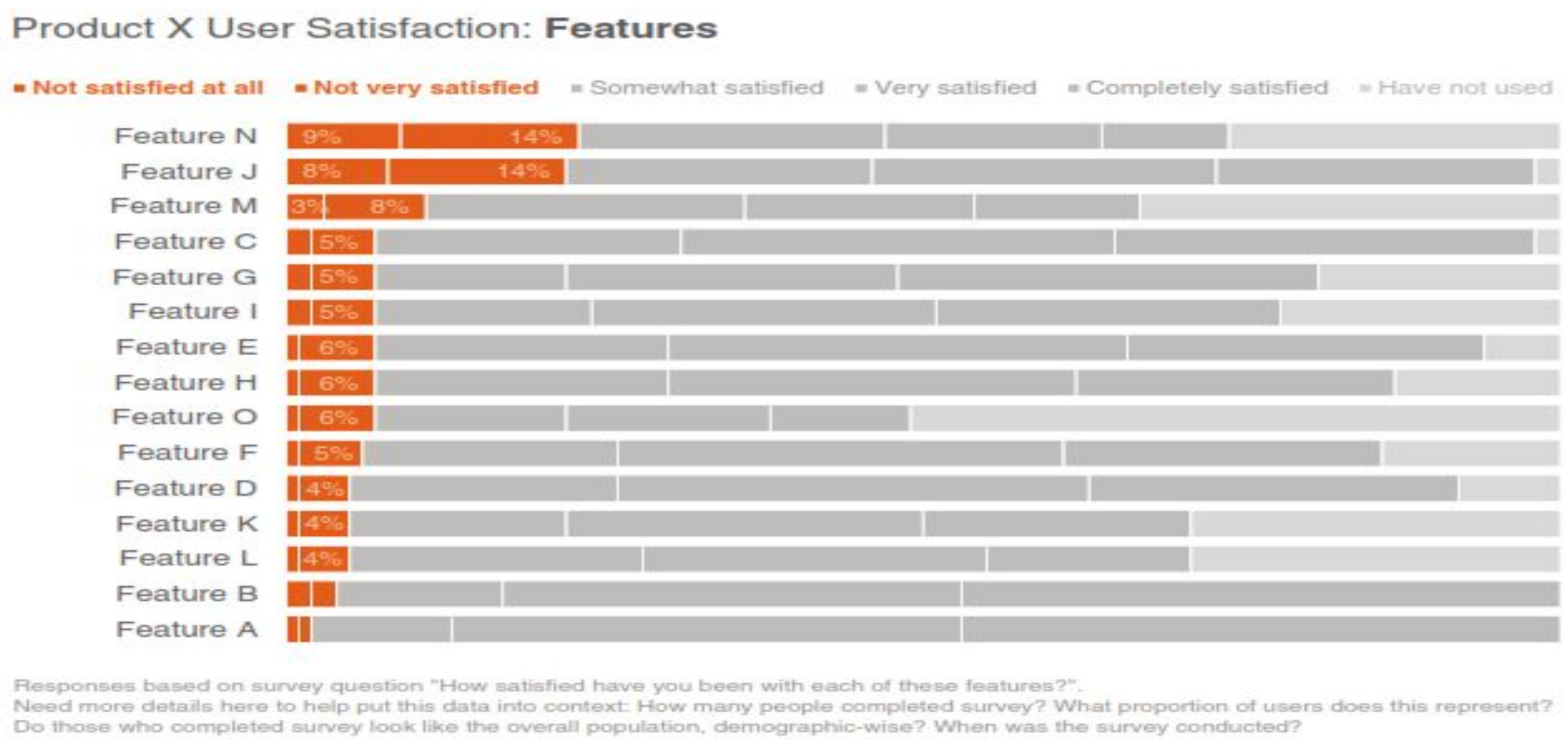


FIGURE 9.14 Highlight dissatisfaction

Feature O is least used

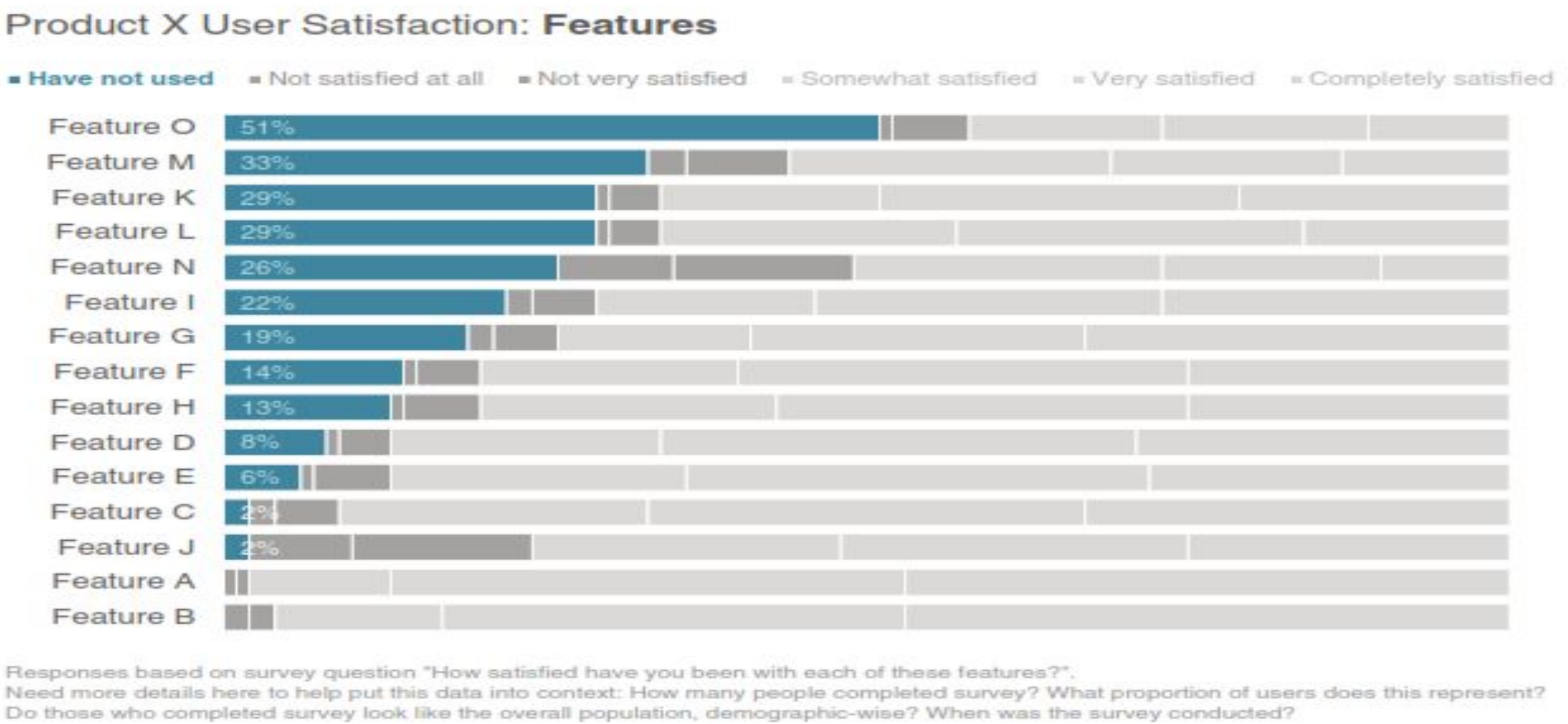


FIGURE 9.15 Focus on unused features

- If we want to tell one of the above stories,
- we can leverage order, color, position, and words
- As shown to draw our audience's attention to where we want them to pay it in the data.
- If we want to tell all three stories, however, I'd recommend a slightly different approach.

User satisfaction varies greatly by feature

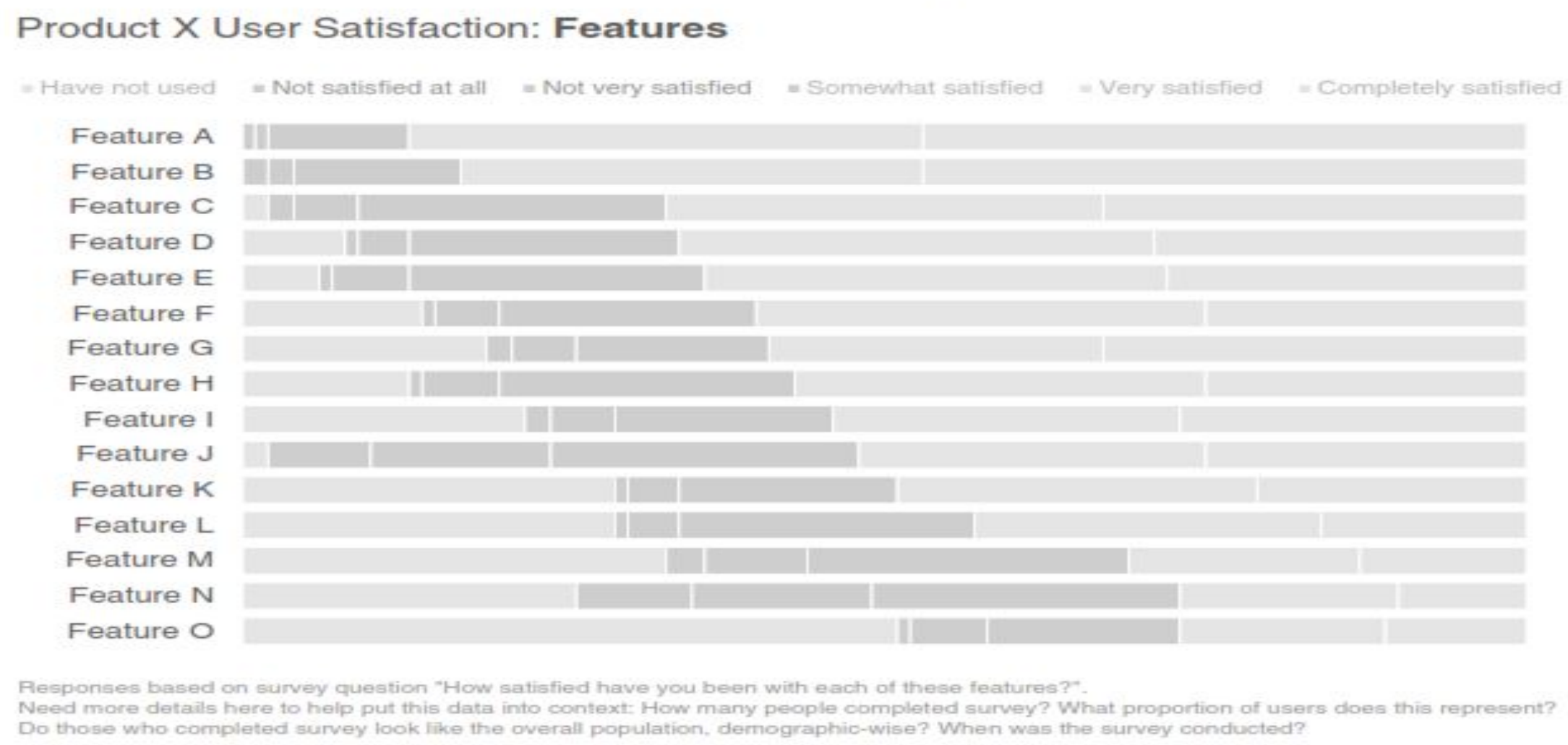


FIGURE 9.16 Set up the graph

User satisfaction varies greatly by feature

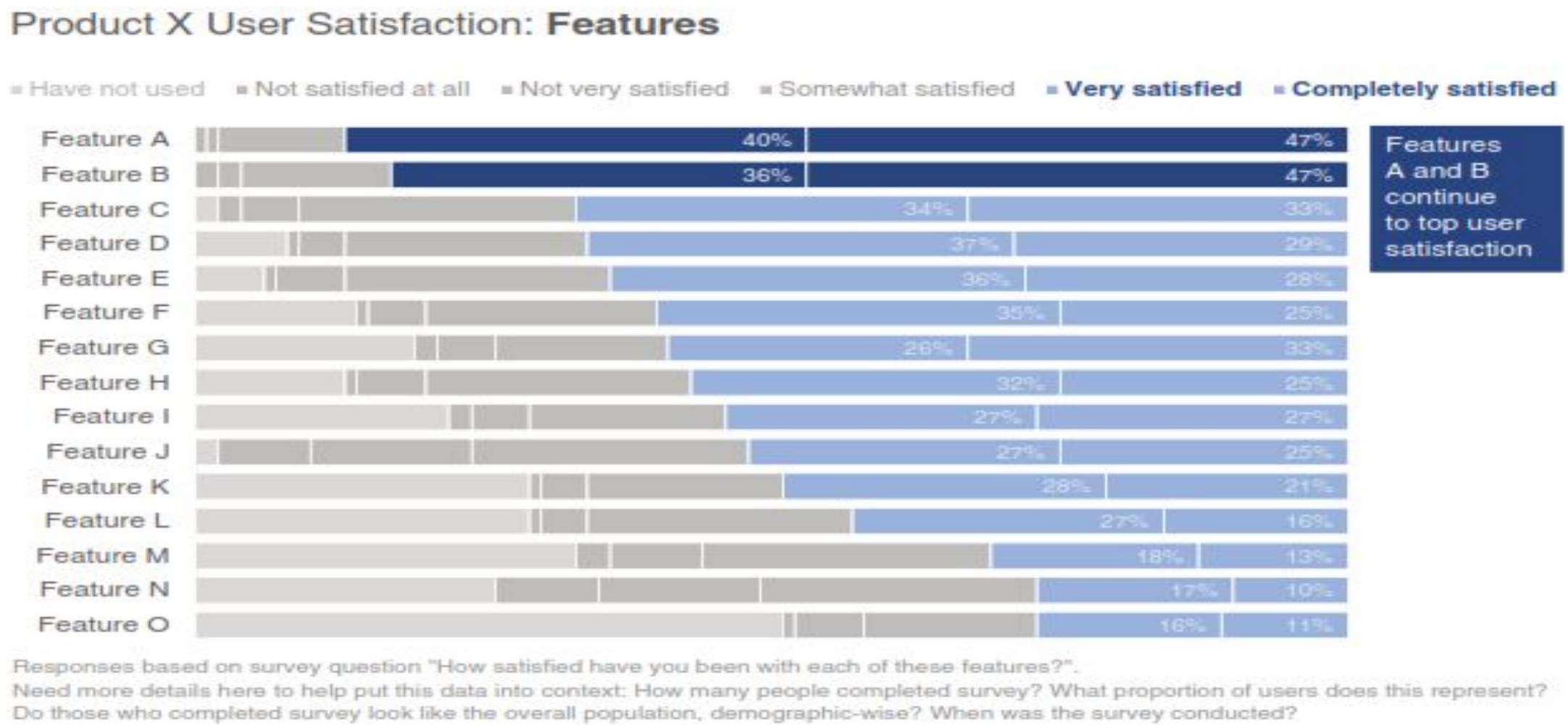


FIGURE 9.17 Satisfaction

User satisfaction varies greatly by feature

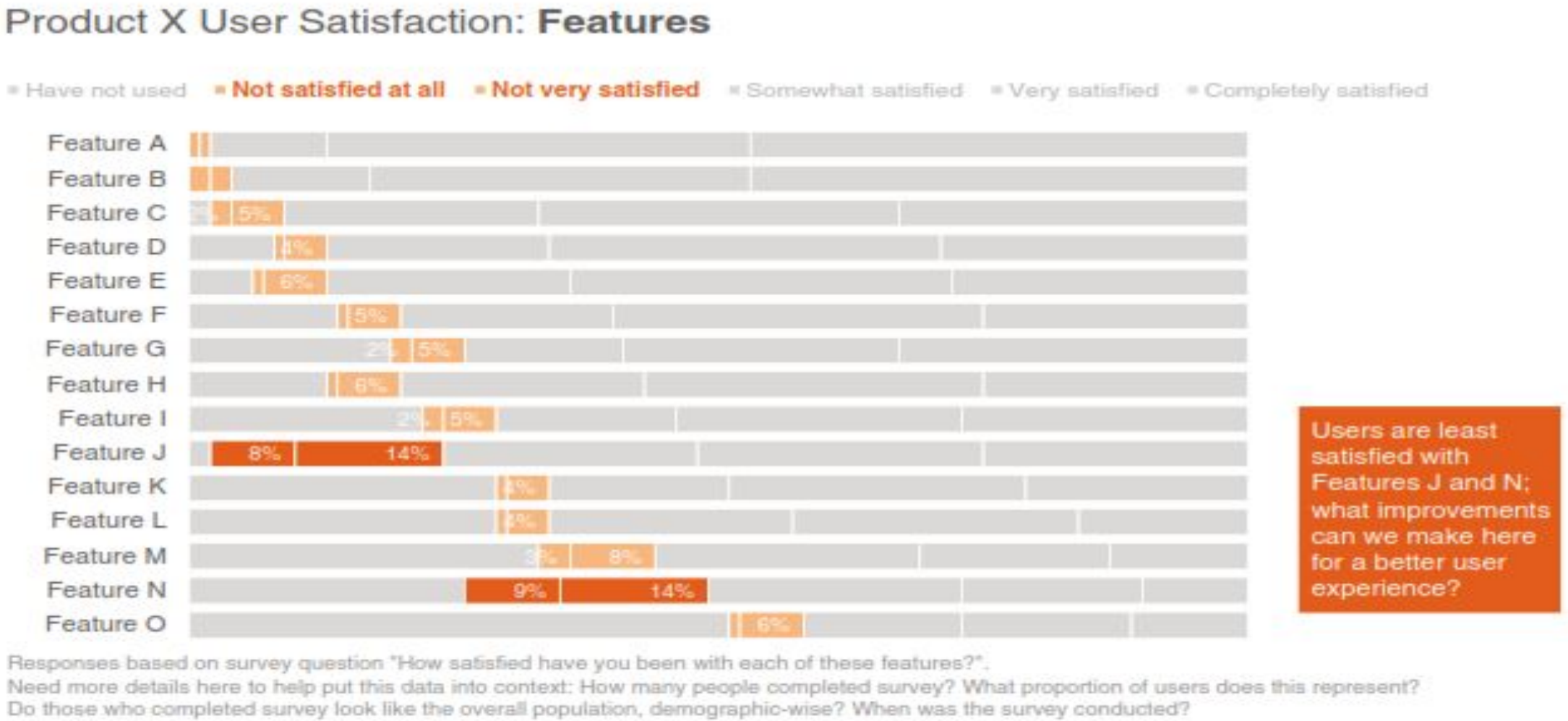


FIGURE 9.18 Dissatisfaction

User satisfaction varies greatly by feature

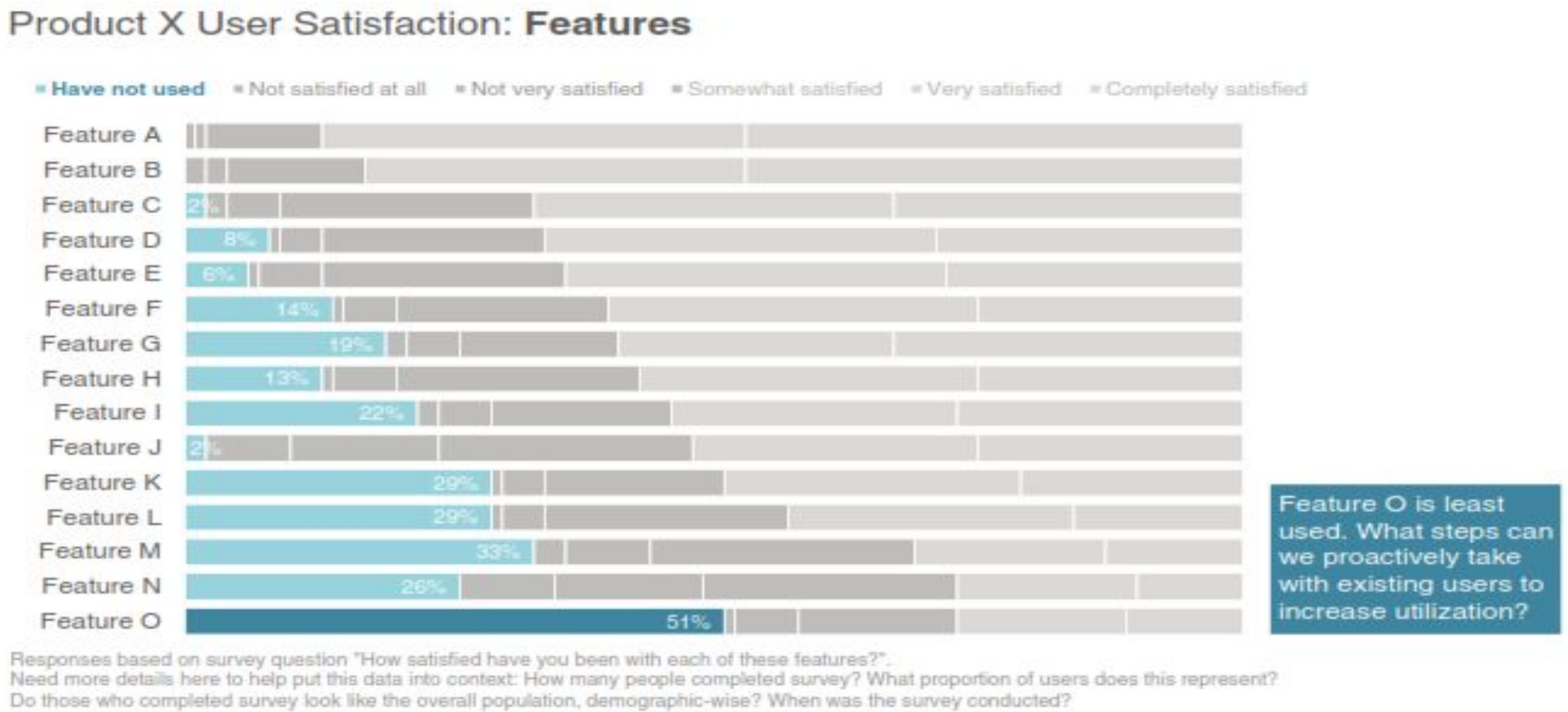


FIGURE 9.19 Unused features

User satisfaction varies greatly by feature

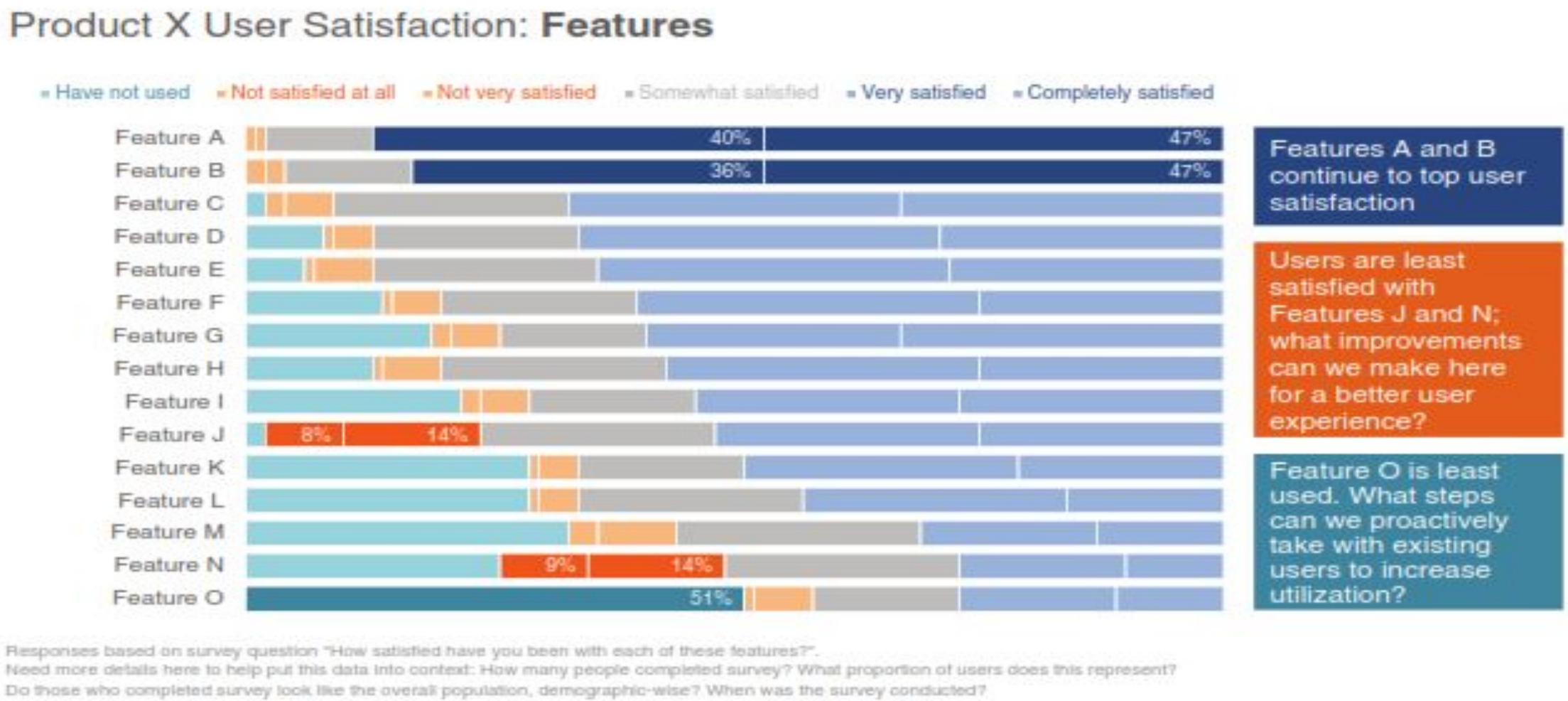


FIGURE 9.20 Comprehensive visual

CASE STUDY 4: Strategies for avoiding the spaghetti graph

- A spaghetti graph is a line graph where the lines overlap a lot, making it difficult to focus on a single series at a time
- Graphs like Figure 9.21 are known as spaghetti graphs because
- They look like someone took a handful of uncooked spaghetti noodles
- and threw them on the ground

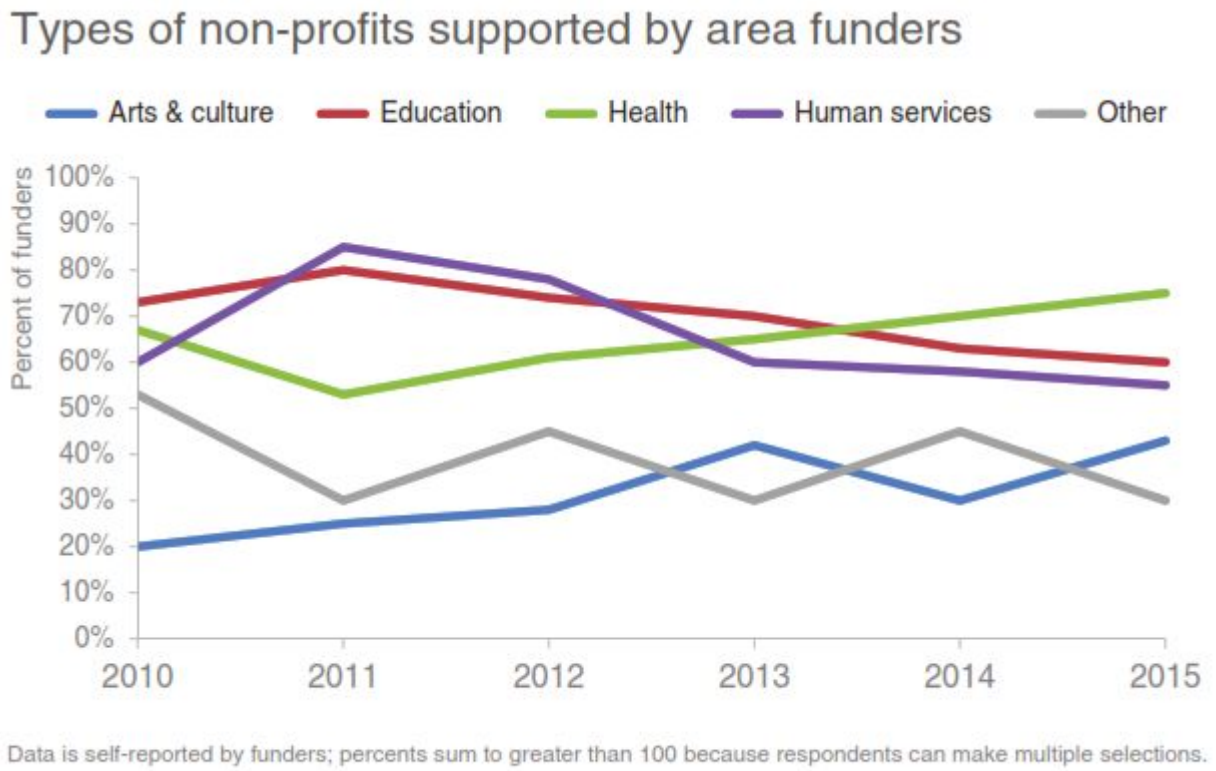


FIGURE 9.21 The spaghetti graph

- How difficult it is to concentrate on a single line…
- Due to all of the crisscrossing and
- because so much is competing for your attention

Few strategies

- Emphasize one line at a time
- Separate spatially
- Combined approach

Emphasize one line at a time

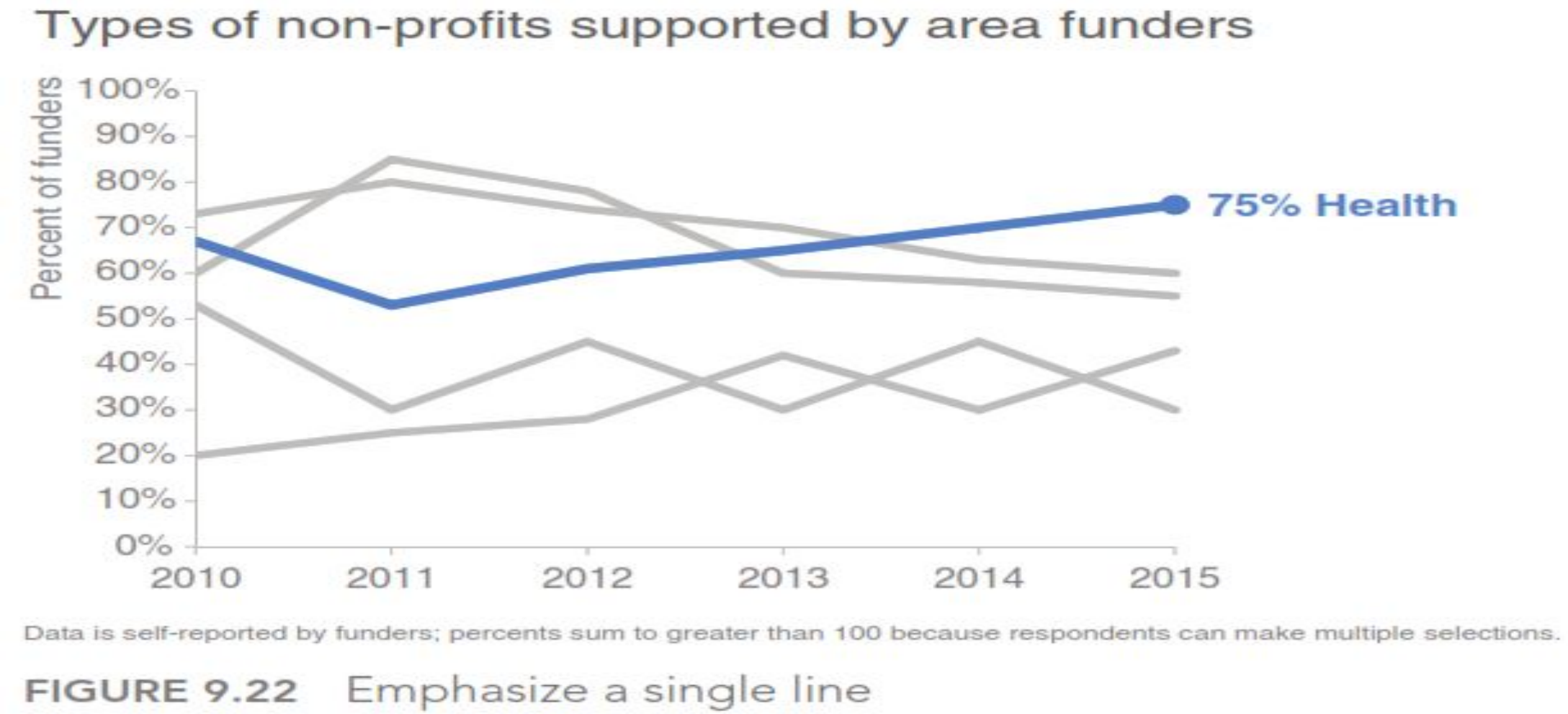


FIGURE 9.22 Emphasize a single line

Emphasize one line at a time

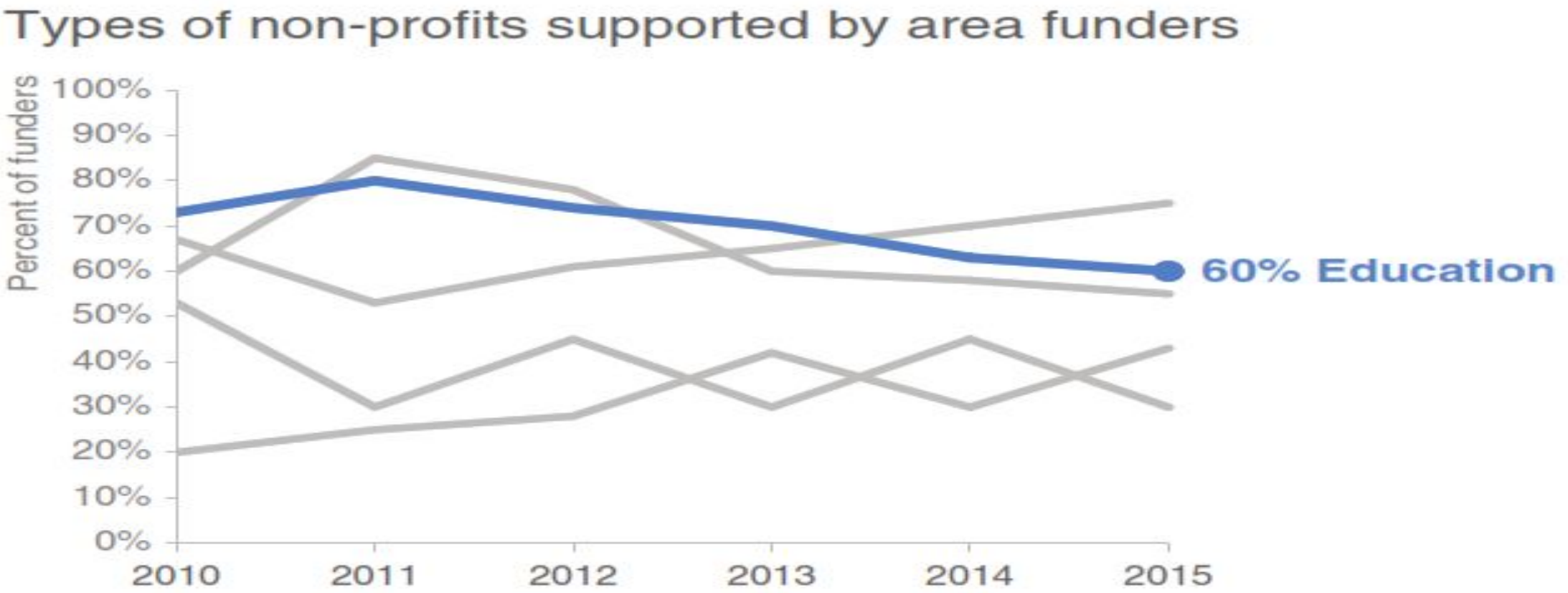


FIGURE 9.23 Emphasize another single line

Separate spatially

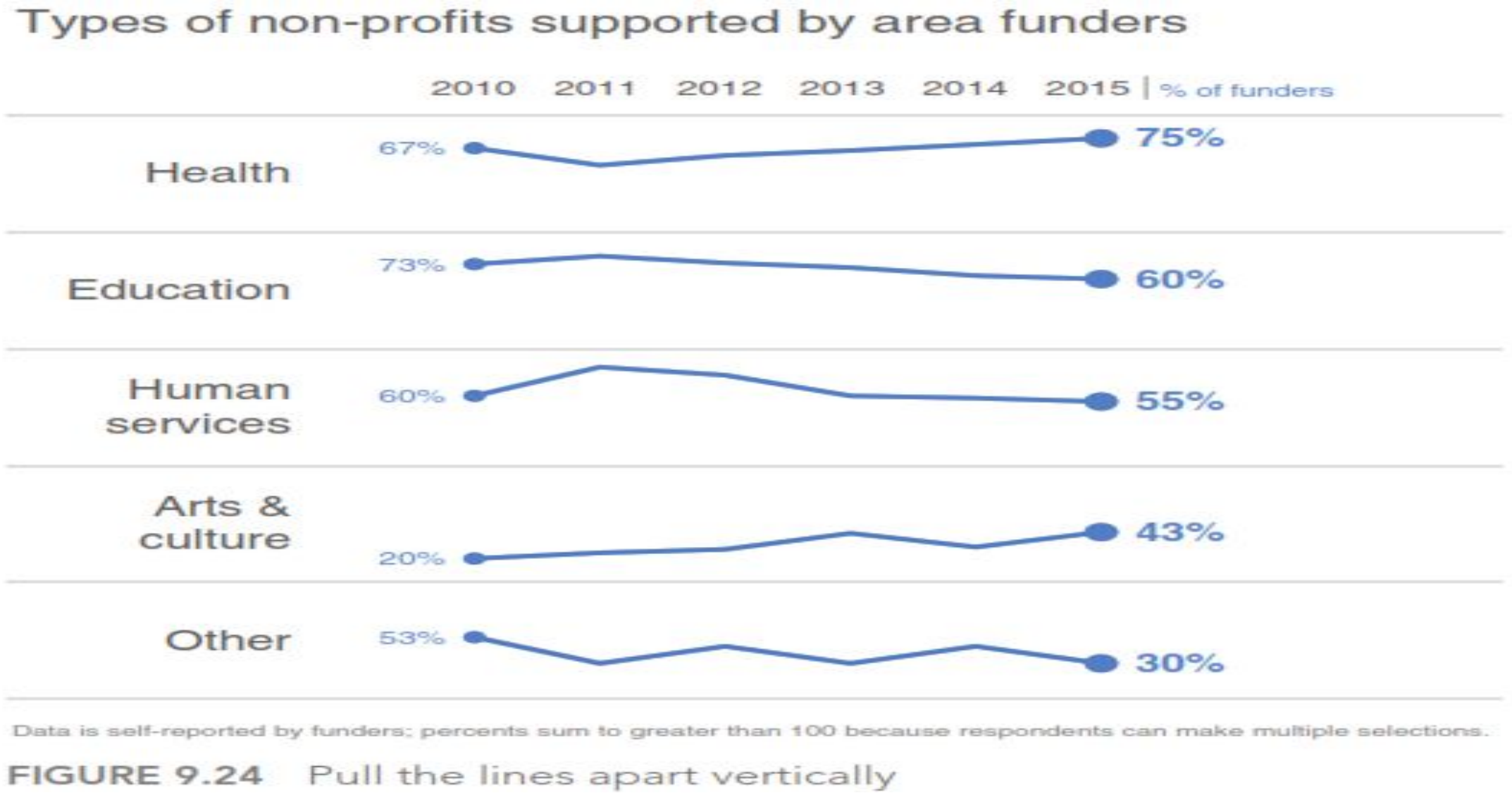


FIGURE 9.24 Pull the lines apart vertically

Separate spatially

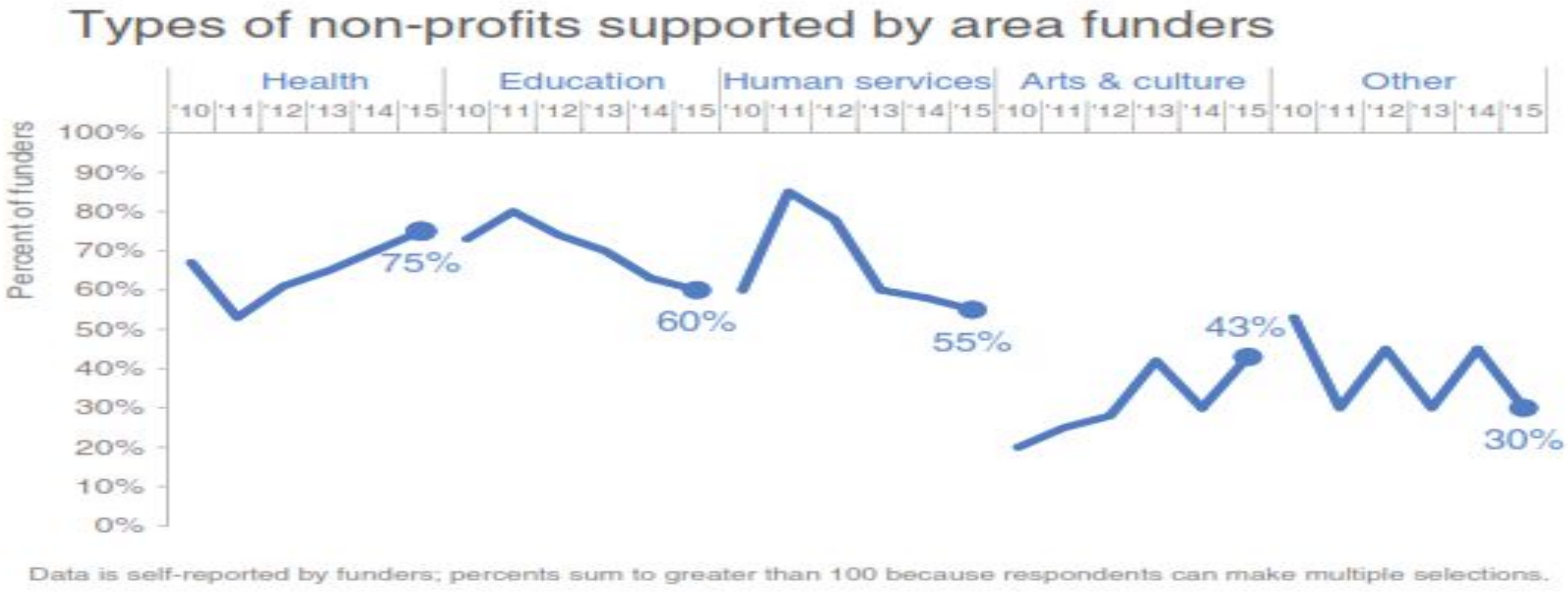


FIGURE 9.25 Pull the lines apart horizontally

Combined approach

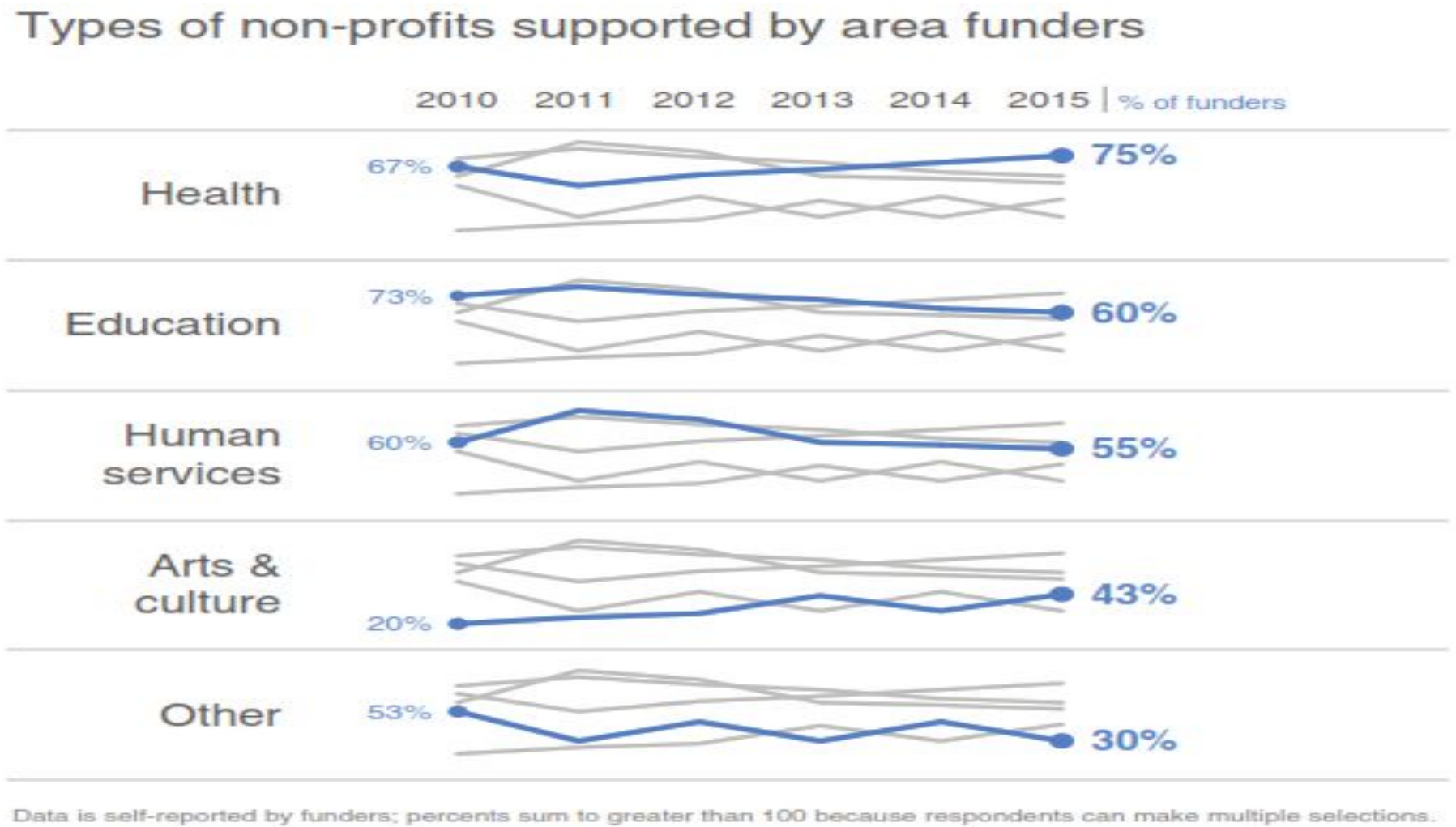


FIGURE 9.26 Combined approach, with vertical separation

Combined approach

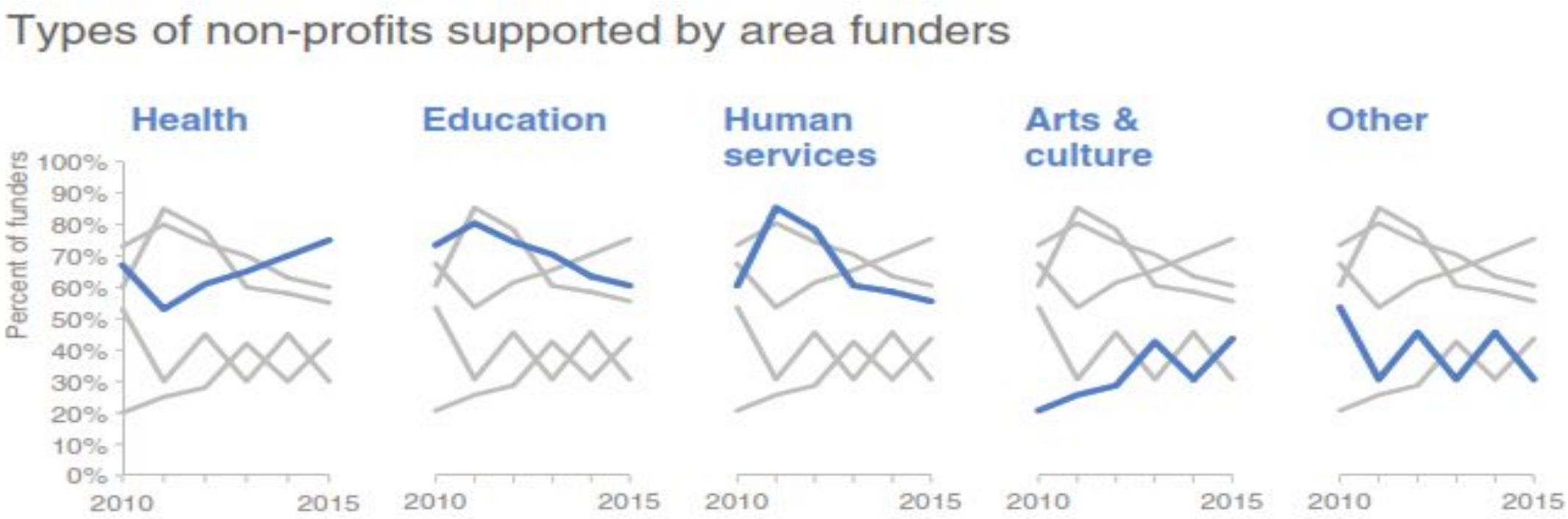


FIGURE 9.27 Combined approach, with horizontal separation

CASE STUDY 5: Alternatives to pies

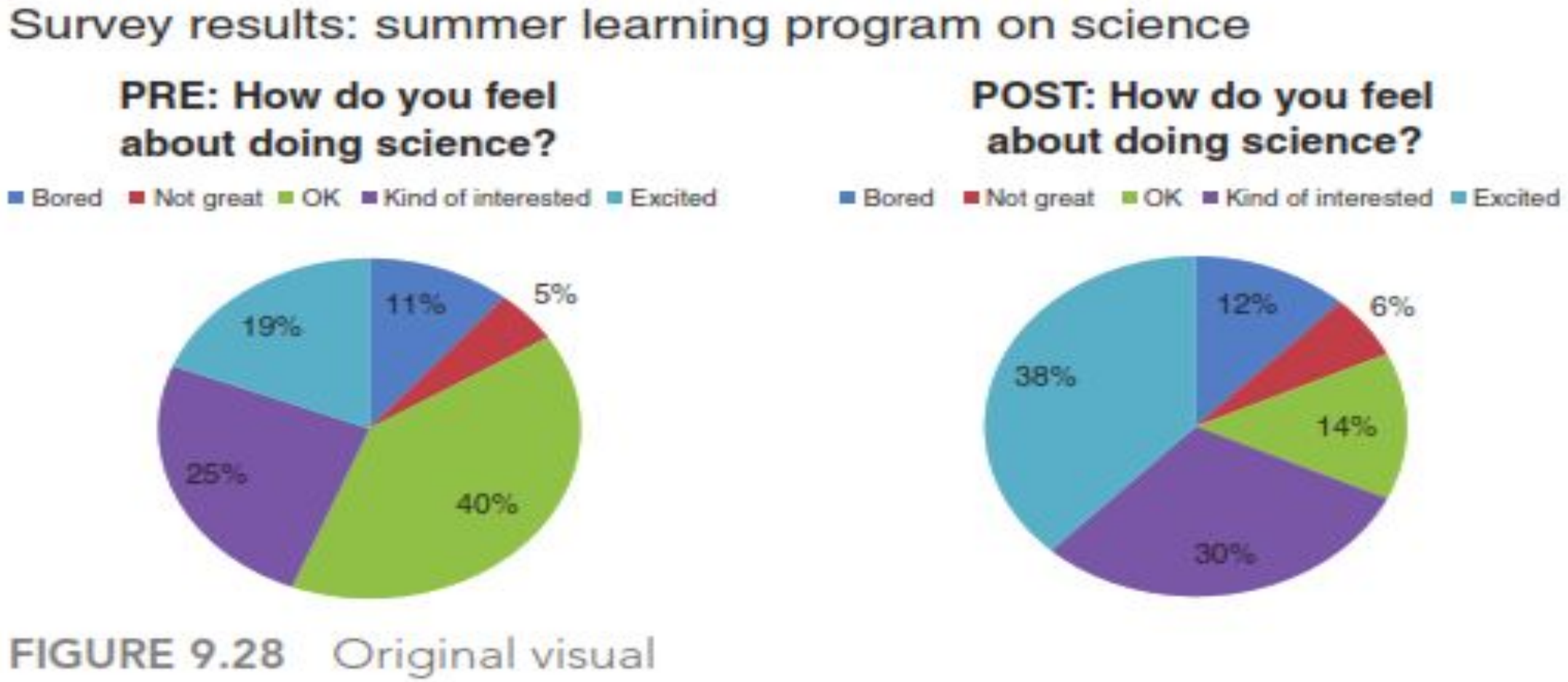


FIGURE 9.28 Original visual

Alternatives

- Alternative #1: show the numbers directly
- Alternative #2: simple bar graph
- Alternative #3: 100% stacked horizontal bar graph
- Alternative #4: slope graph

show the numbers directly

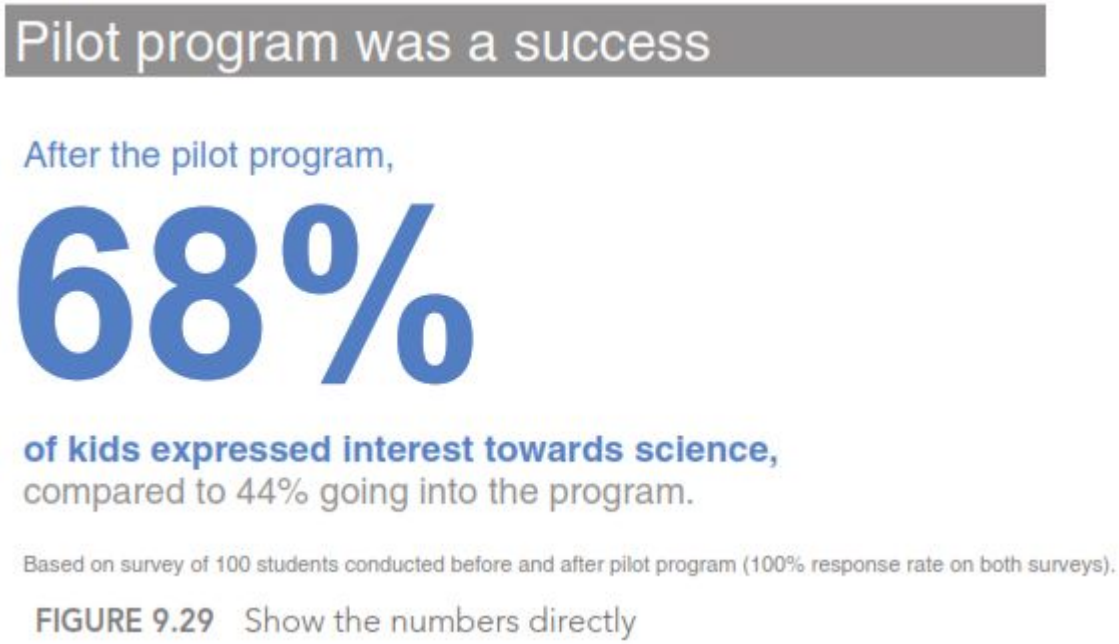


FIGURE 9.29 Show the numbers directly

simple bar graph

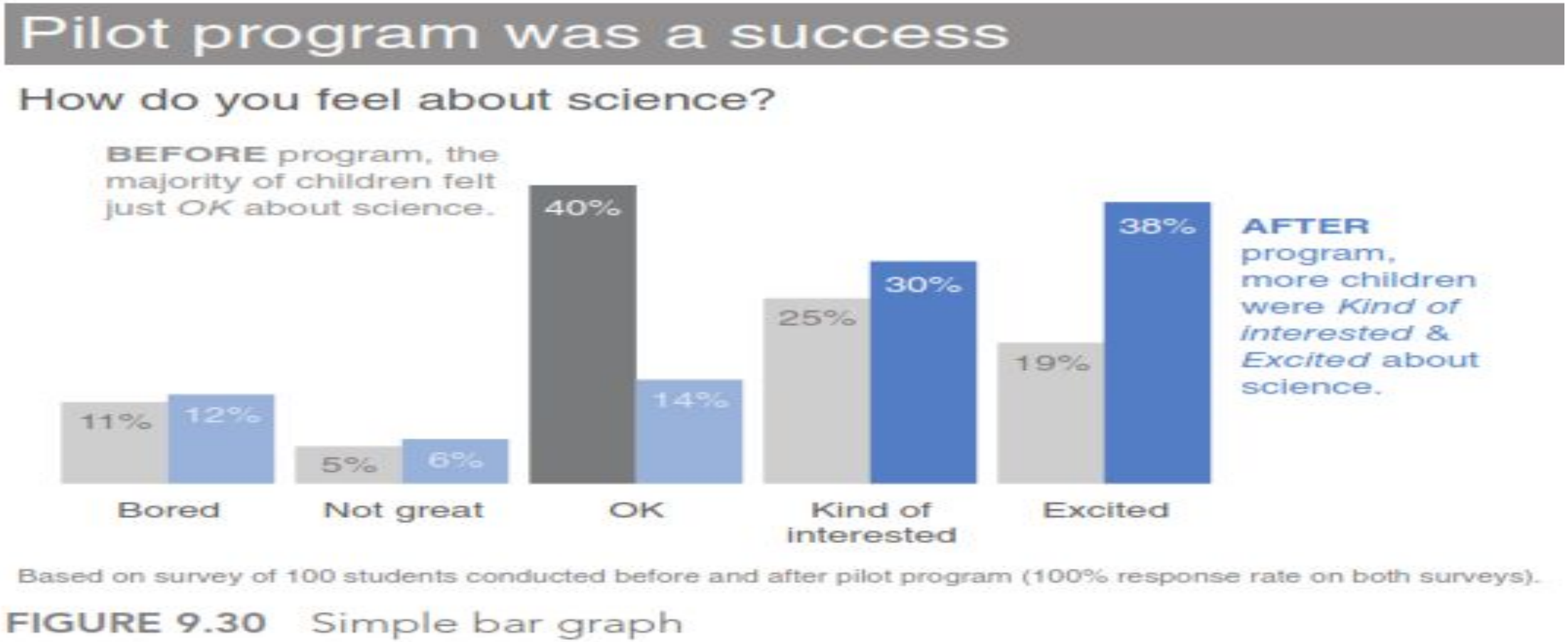


FIGURE 9.30 Simple bar graph

100% stacked horizontal bar graph

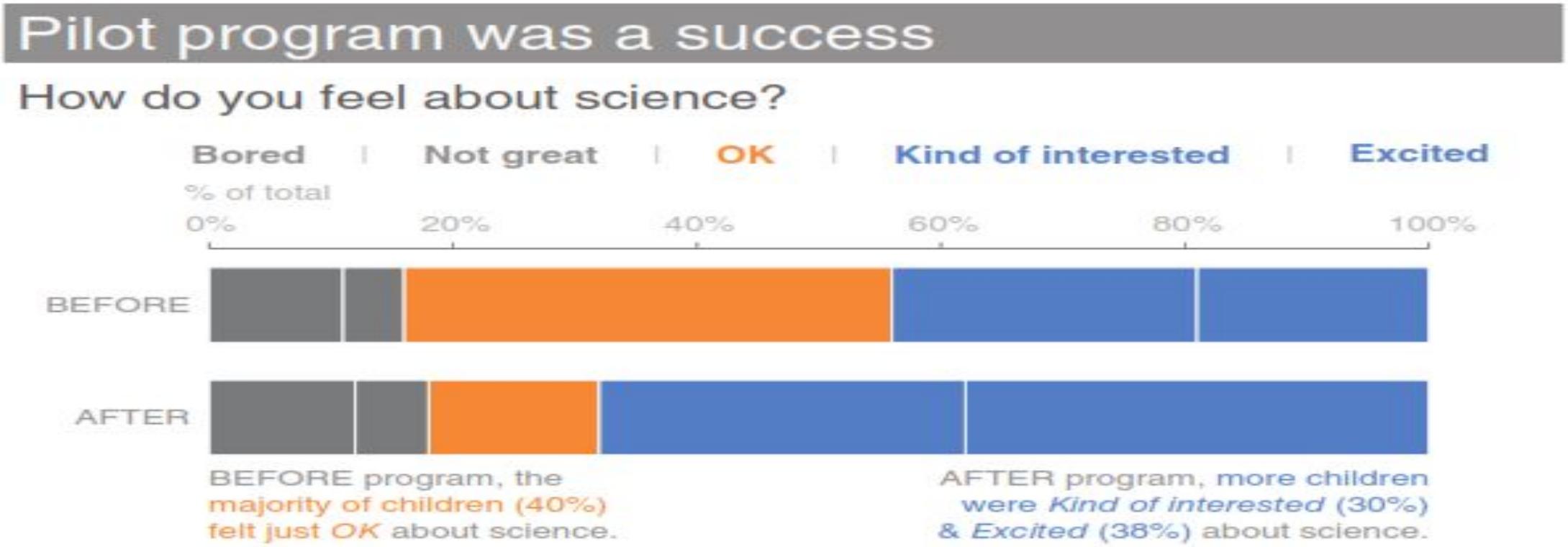


FIGURE 9.31 100% stacked horizontal bar graph

slope graph

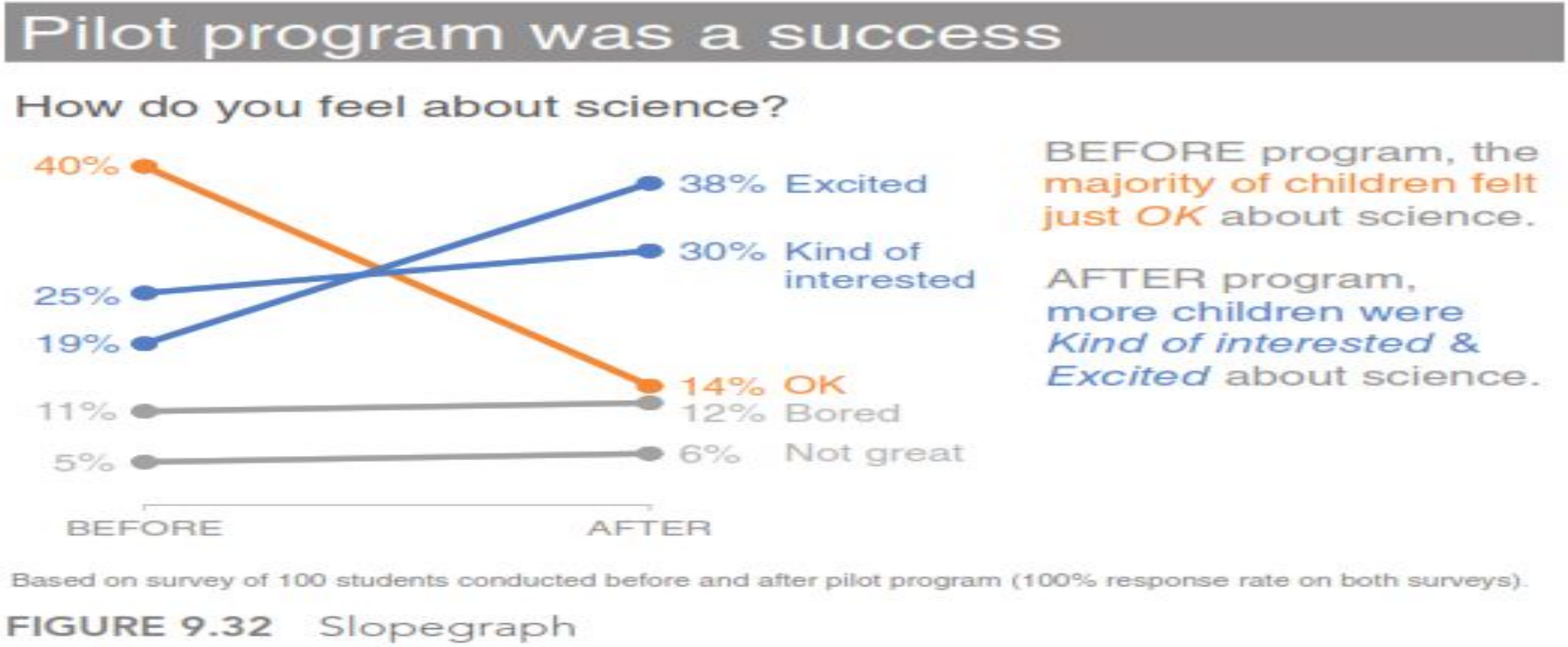


FIGURE 9.32 Slopegraph

- For more case studies like the ones we've considered here,
- check out my blog at storytellingwithdata.com,
- Where you'll find a number of before-and-after examples
- Leveraging the lessons that we've learned.

final thoughts

- Data visualization—and communicating with data in general—sits at the intersection of science and art.
- Different people will approach things in varying ways and come up with distinct solutions to the same data visualization challenge.
- there are often multiple potential paths for communicating effectively with data.
- Simple way to get good at this is to do it: practice, practice, and practice some more

Five final tips:

1. learn your tools well,
2. iterate and seek feedback,
3. devote time to storytelling with data
4. seek inspiration through good examples
5. have fun and find your style

Tip #1: learn your tools well

- Try not to let your tools be a limiting factor when it comes to communicating effectively with data.
- Pick one and get to know it as best you can.
- When you're first starting out, a course to become familiar with the basics may be helpful.
- Best way to learn a tool is to use it.
- When you can't figure out how to do something, don't give up
- Continue to play with the program and search Google for solutions
- Any frustration you encounter will be worth it when you can bend your tool to your will!

Some well known tools

- Google spreadsheets
- Tableau
- Programming languages like R, python, D3
- Some people use Adobe Illustrator- for easier manipulation of graph elements and a professional look and feel.

Tip #2: iterate and seek feedback

- In most cases storytelling with data takes linear path.
- In reality that’s not the case
- start with a blank piece of paper.
- This enables you to brainstorm without the constraints of your tools or what you know how to do in your tools
- determine what will work best for getting your message across to your audience

Tip #3: devote time to storytelling with data

- It takes time to build a robust understanding of the context
- It takes time to understand what motivates our audience
- It takes time to craft the 3-minute story and form the Big Idea
- iterate and seek feedback
- iterate some more to create an effective visual

Example

- Consider the typical analytical process:
- you start with a question or hypothesis,
- then you collect the data,
- then you clean the data,
- and then you analyze the data
- simply throw the data into a graph

Tip #4: seek inspiration through good examples

- Eager Eyes (eagereyes.org, Robert Kosara): Thoughtful content on data visualization and visual storytelling.
- FiveThirtyEight's Data Lab (fivethirtyeight.com/datalab, various

authors): typically minimalist graphing style on a large range of news and current events topics.

- Flowing Data (flowingdata.com)
- The Functional Art (thefunctionalart.com, Alberto Cairo)

Contd…

- The Guardian Data Blog (theguardian.com/data)
- HelpMeViz (HelpMeViz.com)
- Junk Charts (junkcharts.typepad.com)
- VizWiz (vizwiz.blogspot.com)
- storytelling with data (storytellingwithdata.com)

Tip #5: have fun and find your style

- Don’t be afraid to try new approaches and play a little.
- You’ll continue to learn what works and what doesn’t over time

Building storytelling with data competency in your team or organization

- anyone can improve their ability to communicate with data by learning and applying the lessons
- some will have more interest and natural aptitude than others in this space

Few potential strategies to consider:

- upskill everyone,
- invest in an expert, or
- outsource

Upskill everyone

- finding ways to impart foundational knowledge can make everyone better
- Invest in training or use the lessons covered here to generate momentum

some specific ideas:

- **Storytelling with data book club** - read a chapter and discuss it together
- **Do-it-yourself workshop** - After reading books, conduct workshop
- **Makeover Monday** - weekly challenges
- **Feedback loop** - offer feedback to each other
- **And the winner is** - introduce contests (monthly/quarterly)
- Any of these approaches—alone or combined—can create and help ensure continued focus on effective visualization and storytelling with data

Invest in an internal expert or two

- identify an individual or a couple of individuals on your team or in your organization who are interested in data visualization
- and invest in them so they can become your in-house experts
- This can be a great form of recognition and career development for the individual
- Provide time and opportunities to learn and practice.
- As the individual continues to learn, they can share this with others as a way to ensure continued team development as well

Outsource

- In some situations, it may make sense to outsource visual creation to an external expert
- If time or skill constraints are too great to overcome for a specific need, turning to a data visualization or presentation consultant may be worth considering
- The biggest drawback of outsourcing is that you don’t develop the skills and learn in the same way
- look for opportunities to learn from the consultant during the process to overcome this drawback

A combined approach

- TEAMs / Organizations recognize the importance of storytelling with data and invest in training and
- practice to give everyone the foundational knowledge for effective data visualization
- They also identify and support an internal expert, to whom the rest of the team can turn for help overcoming specific challenges
- They bring in external experts to learn from as makes sense.

Recap: a quick look at all we've learned

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