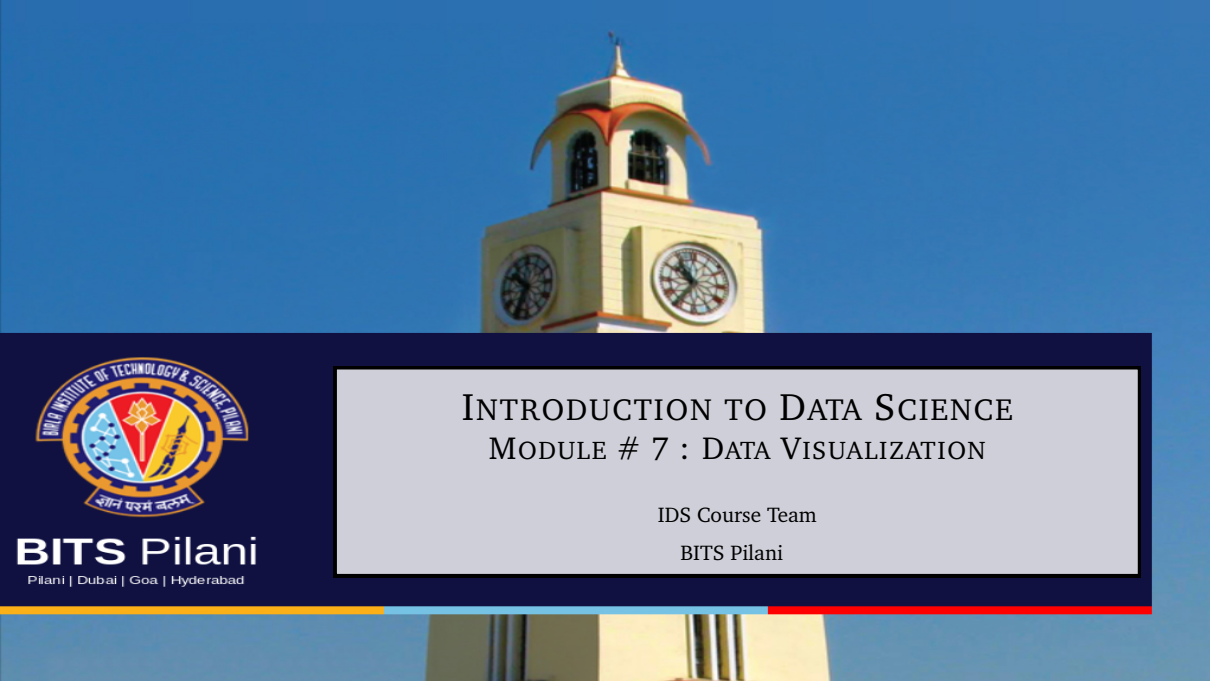


Introduction to Data Science
Module # 7 : Data Visualization

IDS Course Team

BITS Pilani

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

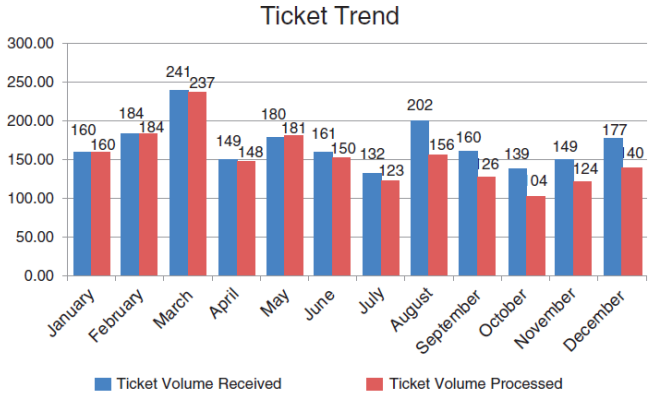
The instructor is gratefully acknowledging the authors who made their course materials freely available online.

Table of Contents

Types of Data Representations

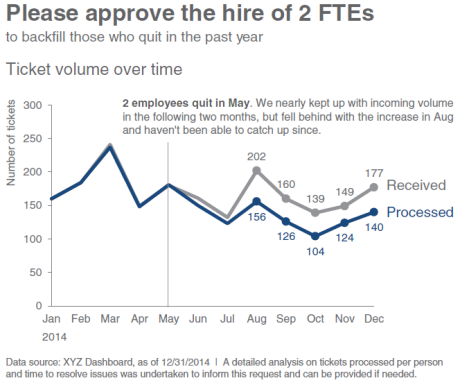
- Presentation
  - Uses data visuals to communicate
  - Two roles : Presenter and Audience
  - Communicate and Persuade
- Visualization
  - Use visuals to think.
  - Involves people trying to answer questions.

Case 1

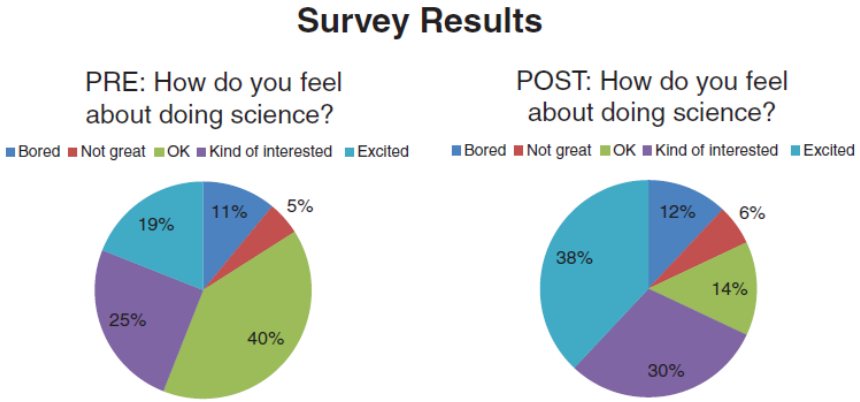

Ticket Trend
300.00
250.00
200.00
150.00
100.00
50.00
0.00
160
160
184
184
241
237
149
148
180
181
161
150
132
123
202
156
160
126
139
04
149
124
177
140
January
February
March
April
May
June
July
August
September
October
November
December
Ticket Volume Received
Ticket Volume Processed

Case 1

Better Visualization

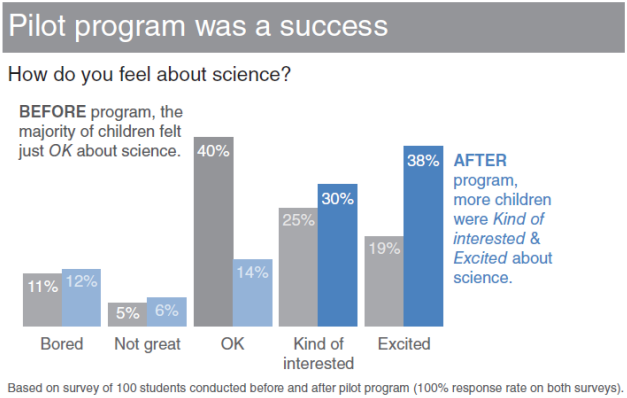


Case 2

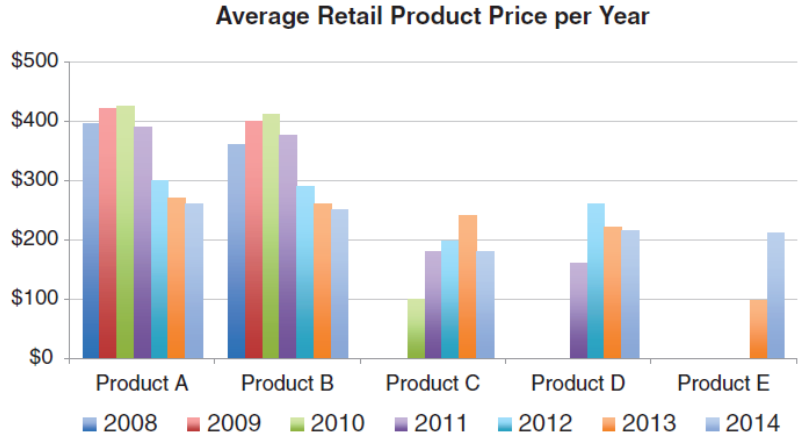

Survey Results
PRE: How do you feel about doing science?
Bored
Not great
OK
Kind of interested
Excited
11%
5%
40%
25%
19%
POST: How do you feel about doing science?
Bored
Not great
OK
Kind of interested
Excited
12%
6%
14%
30%
38%

Case 2

Better Visualization



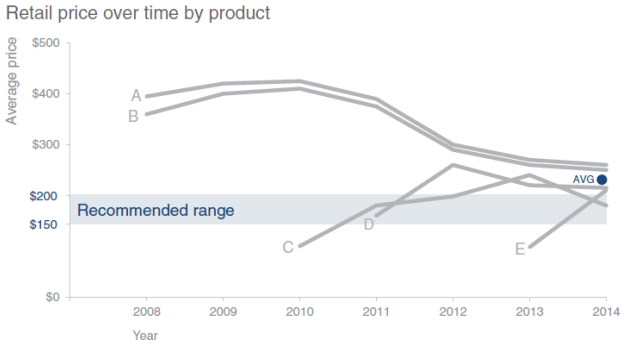
Case 3


Average Retail Product Price per Year
$500
$400
$300
$200
$100
$0
Product A
Product B
Product C
Product D
Product E
2008
2009
2010
2011
2012
2013
2014

Case 3

Better Visualization

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



Data Visualization

- Tell stories with numbers.
- Communicate something to someone using data.

Ben Shneiderman

"The purpose of visualization is insight, not just picture."

Need for Data Visualization

- What?
  - Graphical / Visual representation of data.
- Why?
  - Way to identify patterns, trends and outliers in data.
  - Helps is making data-driven decisions.

Stephen Few

"Data visualization is the use of visual representations to explore, make sense of, and communicate data."

Data Visualization Goals

We visualize quantitative data to perform three fundamental tasks in an effort to achieve three essential goals:

Task	Description	Goal
Exploration	Searching for significant facts	Discovery
Sense-making	Examining and making sense of data	Understanding
Communication	Conveying information to the audience	Informed decisions

Why do we really visualize data?

- Because humans can perceive particular quantitative features and perform particular quantitative tasks most effectively when the data is expressed graphically.
- Visual data processing provides optimal support for the following:
  - Seeing the big picture
  - Easily and rapidly comparing values
  - Seeing patterns among values
  - Comparing patterns

Visual Business Intelligence by Stephen Few

Types of Data Visualization

- Based on Interactivity
  - Static Visualizations
    - Used in presentations, documents, info-graphics.
    - Requires careful design as it is meant for offline viewing.
    - User cannot adjust the views.
    - Usually focused on a specific data story, users can't go beyond a single view to explore additional stories beyond what's in front of them.
    - The story is specifically captured in an engaging single page layout.
  - Dynamic Visualizations
    - Used for exploratory data analysis.
    - Meant for live or online interactions.
    - Users have more viewing options.
    - User gains more control over the display.
    - Users can select specific data points to build a visualized story of their choosing.
    - These visualizations allow the user to be part of the data visualization process by building a story of their choosing.

Types of Data Visualization

- Based on Application
  - Common/ General Types of Data Visualization
    - * Charts * Tables * Graphs * Plots
    - * Maps * Info-graphics * Dashboards
  - Specific methods
    - * Area chart * Heat Map * Scatter plot
    - * Box and Whisker plots * Highlight table * Stream graph
    - * Bubble cloud * Histogram * Text tables
    - * Cartogram * Matrix * Time-line
    - * Circle view * Network * Tree map
    - * Dot distribution map * Polar Area * Wedge Stack graph
    - * Gantt Chart * Radial Tree * Word cloud

Usage of Data Visualization

- Exploratory Analysis
  - Purpose is to understand the data and figure out what might be noteworthy or interesting to highlight to others.
  - Explore or dig through the data to become familiar with data.
  - Find trends and relationships w.r.t. specific goals.
  - Helps to determine the analyses to apply to data.
  - Eg: hunting for pearls in oysters.
- Explanatory Analysis
  - Explain outcomes or results of analysis. It's about communicating our findings.
  - Tell a story with data.
  - We need to be in the explanatory space when we want to communicate the results of our analysis to our audience.
  - At this stage, we should have a specific story to tell or a specific thing we want to explain.
  - Eg: probably about those two specific pearls.

Table of Contents

Visualization Context

Cole Knaflic

Success in data visualization does not start with data visualization itself.
. . . understanding the context sets solid foundation for data visualization creation.

- Understand the context for the need to communicate.
- W^2H – Who, What, How.

Visualization Context

- Who
  - To whom are you communicating?
  - Who is your audience?
    - Knowing them place you in better position for communication.
    - Be specific while identifying audience.
    - Different content for different set of audience.
    - Avoid general audiences, such as ”internal and external stakeholders” or ”anyone who might be interested” .
- What
  - What do you want your audience to know or do?
- How
  - How can you use data to help make your point?
  - What data is available?
  - Supporting evidence of the story.

Audience to Know or Act

- What
  - Action – What do you want your audience to know or do?
    - Make sure audience care about what you say.
    - You are subject matter expert – unique position to interpret the data and help lead people to understanding and take action.
    - If no action recommendation possible / feasible, then encourage discussion towards one.
  - Mechanism – How you will communicate to your audience?
    - Determines level of control and level of detail.
    - Communication mechanism can be placed on a continuum of live presentation on one end to a written document/email on the other.
    - Three levels of communication mechanisms - Live presentation or Written document or Slideument
  - Tone – How you want your audience to act?
    - Overall tone that you want to set for your communication.
    - Celebrating success? Is topic light-hearted or serious?
    - Lighting a fire to drive action?

Communication Mechanism

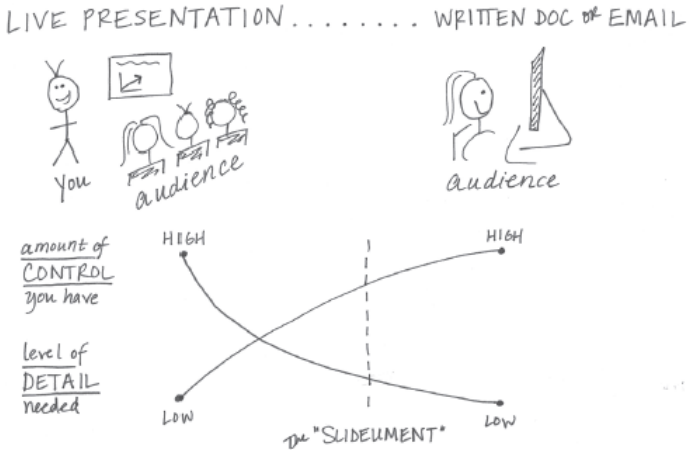


FIGURE 1.1 Communication mechanism continuum

Table of Contents

Storytelling Strategies

- The 3-minute story
- Big Idea
- Storyboarding

The 3-Minute Story

- Within 3 minutes you have to tell audience, what they need to know .
- This is a great way to ensure we are clear on and can articulate the story we want to tell.
- Removes dependence on supporting material like visuals , presentations etc.
- Helps in a situation where we have to give elevator speech, or if our half hour agenda is cut short to ten minutes, or to five.
- Need to know exactly what data is saying.
- Being concise is more challenging than being verbose.

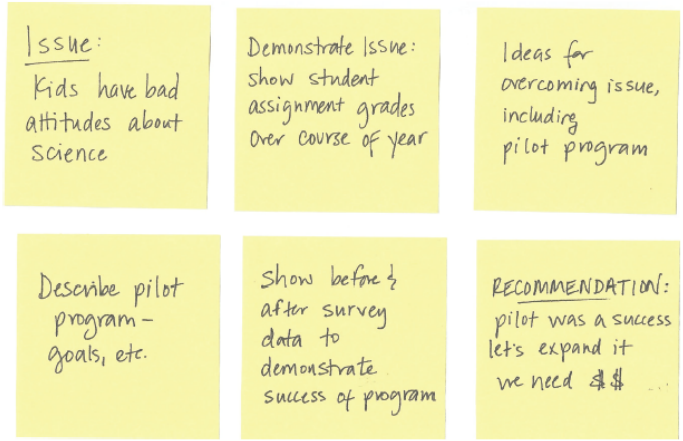
Big Idea

- Boils down to a single most important sentence.
- As per Nancy Duarte (in her book Resonate, 2010), big Idea has three components:
  - It must articulate your unique point of view.
  - It must convey what's at stake.
  - It must be complete sentence.

Story Boarding

- Storyboarding helps ensure the communication we craft is on point.
- A storyboard establishes a structure for our communication.
- It is a visual outline of the content we plan to create .
- It can be subject to change as we work through the details.
- Avoid using presentation computer-aided tools in beginning.
- Use whiteboard, post-it notes, paper.
- When possible, get acceptance from your client or stakeholder at this step. It will help ensure that what you're planning is in line with the need.

Story Boarding Example



Tools for Data Visualization

1. Simple Text
2. Table
3. Heatmap
4. Scatterplot
5. Line
6. Slopegraph
7. Bar graphs
8. Stacked bar graphs
9. Waterfall
10. Square area

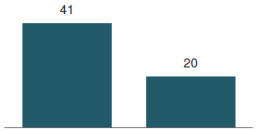
Simple Text

- Use a number or two directly.
- Use the number solely to make it prominent.
- Add a few supporting words to improve calrity.

Simple Text

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

20%

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

FIGURE 2.3 Stay-at-home moms simple text makeover

Table

- When we have more numbers to show.
- Interact with our verbal system.
- Used to compare values, communicate to a mixed audience, or multiple different units.
- Design of the table should be at minimum.
- No heavy borders or shading.
- White space or light/gray borders as separator.
- Data should be highlighted.

Table

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

FIGURE 2.4 Table borders

Heatmap

- Using visual cues in a heatmap.
- A heatmap is a way to visualize data in tabular format, where in place of the numbers, you leverage colored cells that convey the relative magnitude of the numbers.
- Use color saturation to provide visual cues to quickly target the potential points of interest.
- Always include a legend as a subtitle on the heatmap with color corresponding to the conditional formatting color.

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%

FIGURE 2.5 Two views of the same data

Scatterplot

- Show relationship between data.
- Better for categorical data.

Scatterplot

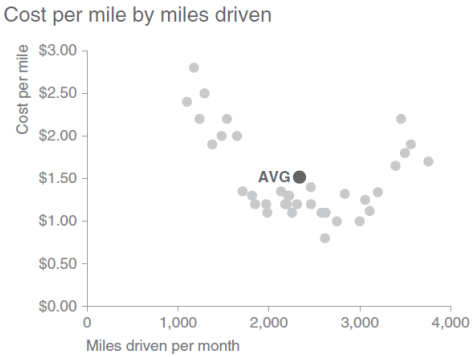


FIGURE 2.6 Scatterplot

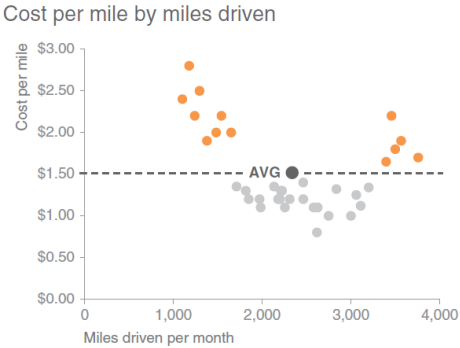


FIGURE 2.7 Modified scatterplot

Graph

- When we have more numbers to show.
- Interact with our visual system.
- Faster at processing information.
- 4 categories
  - Points
  - Lines
  - Bars
  - Area

Lines

- Plot continuous data.
- The line graph can show a single series of data, two series of data, or multiple series.
- Be consistent in the time points you plot.

Lines

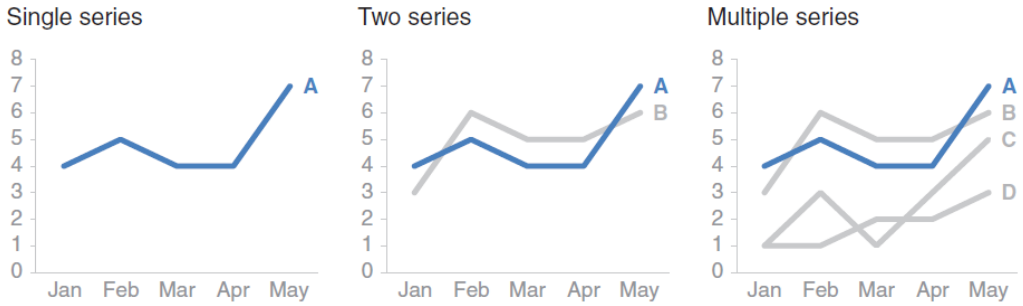


FIGURE 2.8 Line graphs

Lines

FIGURE 2.9 Showing average within a range in a line graph

Slopegraph

- Slopegraphs can be useful when you have two time periods or points of comparison and want to quickly show relative increases and decreases or differences across various categories between the two data points.
- Less overlapping lines.
- Emphasize a single series for highlighting.

Slopegraph

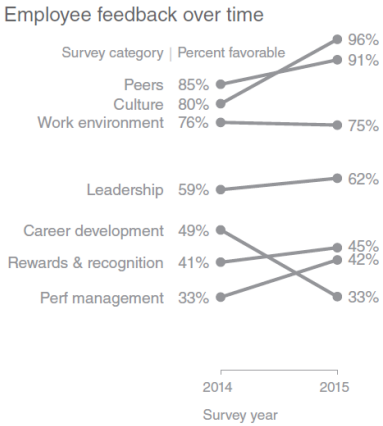


FIGURE 2.10 Slopegraph

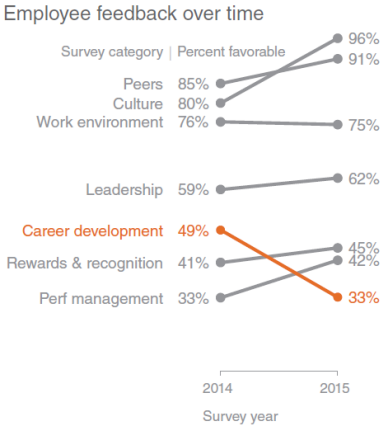


FIGURE 2.11 Modified slopegraph

Bar Charts

- 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.
- Bar charts should always have a zero baseline.
- Less learning curve for the audience.
- Categorical data or data is organised into groups.

Bar Charts

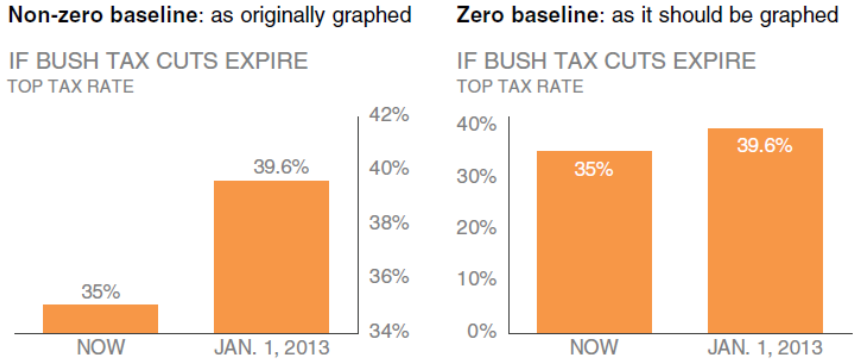


FIGURE 2.13 Bar charts must have a zero baseline

Bar Charts

FIGURE 2.14 Bar width

Vertical Bar Charts

- Ensure that there is visual grouping when using multiple series of vertical bars.
- The relative order of the categorization is important.

Vertical Bar Charts

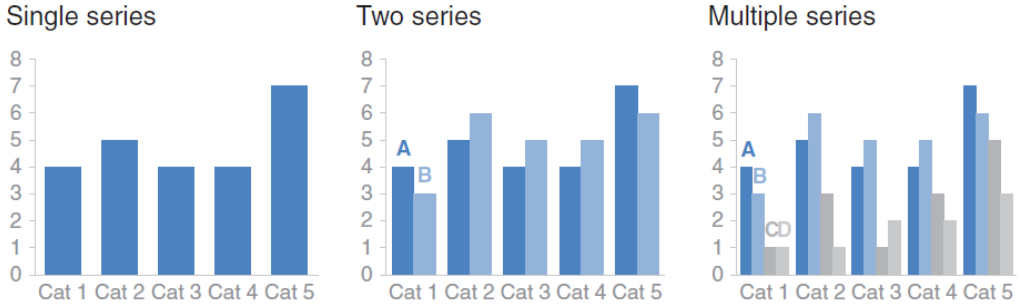


FIGURE 2.15 Bar charts

Stacked Vertical Bar Charts

- Compare totals across categories and also compare the subcomponent pieces within a given category.
- Visually overwhelming
- Less use of the varied default color schemes.
- The stacked vertical bar chart can be absolute numbers or with each column summing to 100%.
- With 100% stacked bar, include the absolute numbers for each category total for better interpretation of the data.

Stacked Vertical Bar Charts

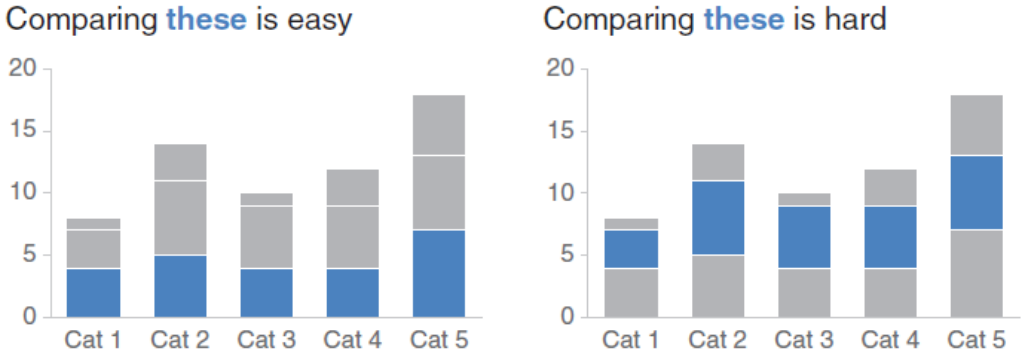


FIGURE 2.16 Comparing series with stacked bar charts

Waterfall Charts

- The waterfall chart can be used to pull apart the pieces of a stacked bar chart to focus on one at a time.
- Show a starting point, incremental additions and deductions, and the resulting ending point.

Waterfall charts

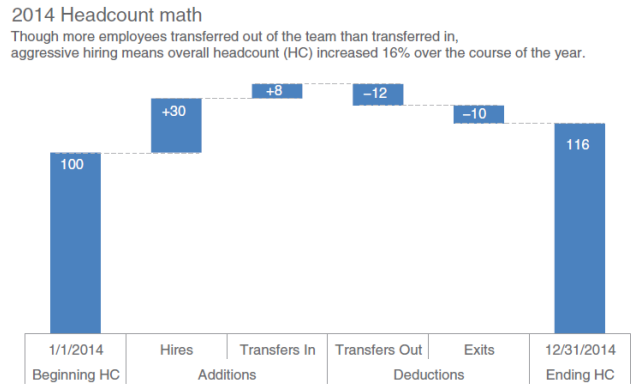


FIGURE 2.17 Waterfall chart

Horizontal Bar Charts

- Categorical data
- Keep categories such that the natural ordering is preserved.
- If the biggest category is the most important, put that first and ordering the rest of the categories in decreasing numerical order.
- If the smallest is most important, put that at the top and order by ascending data values.

Horizontal Bar Charts

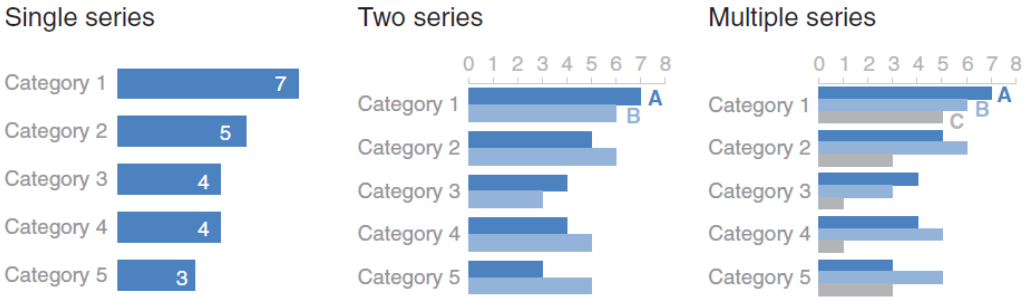


FIGURE 2.18 Horizontal bar charts

Stacked Horizontal Bar Charts

- Compare totals across categories and also compare the subcomponent pieces within a given category.
- The stacked vertical bar chart can be absolute numbers or with each row summing to 100%.
- A consistent baseline is available on both the far left and the far right.
- Use to show a scale from negative to positive.

Stacked Horizontal Bar Charts

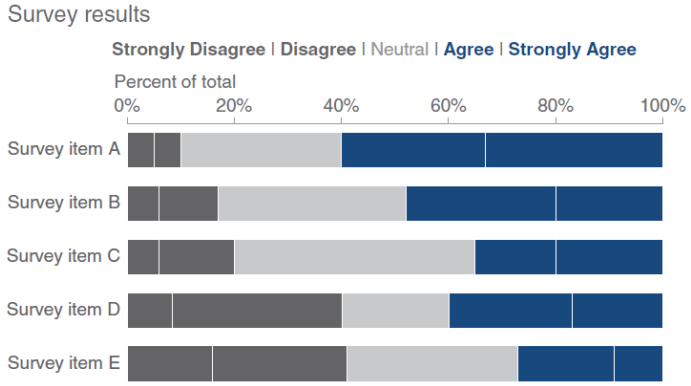


FIGURE 2.19 100% stacked horizontal bar chart

Area Graphs

- 2-dimensional

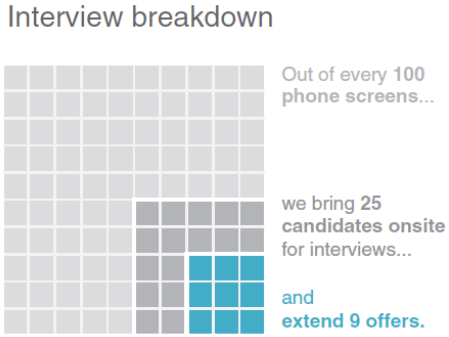


FIGURE 2.20 Square area graph

Avoid Pie Charts

- Pie Charts to be avoided.
- Use bar charts instead.

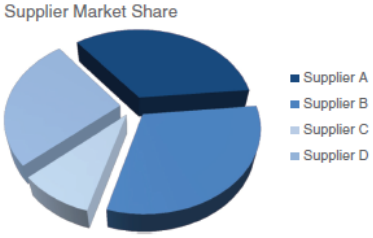


FIGURE 2.21 Pie chart

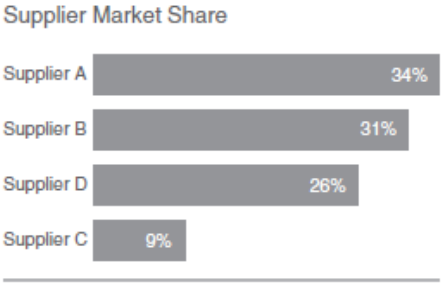


FIGURE 2.23 An alternative to the pie chart

Avoid Secondary y-axis

- Label directly.
- Pull apart vertically.

Secondary y-axis

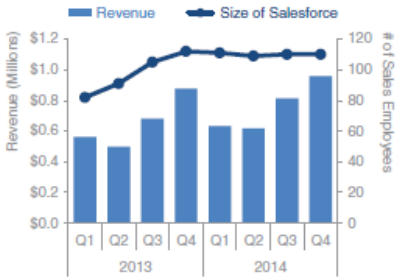


FIGURE 2.26 Secondary y-axis

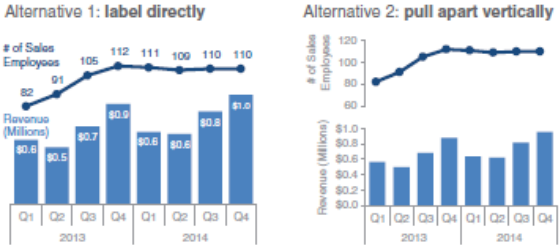


FIGURE 2.27 Strategies for avoiding a secondary y-axis

Better Graphs

- Reduce clutter
- Alignment of text should be made in context of the other elements on the page.
- Upper-left-most justify the text like title, axis titles, legend.
- Diagonal components should be avoided.
- Include lot of white space.
- Margins should remain free of text and visuals.
- The lack of clear contrast can be a visual clutter.

Declutter

- Remove chart border.
- Remove gridlines.
- Remove data markers.
- Clean up axis labels.
- Label data directly.
- Leverage consistent color.

Pre-attentive Attributes

- Employ pre-attentive attributes to create visual hierarchy in the communication.
- **Bold**
- *Italics*
- Color
- Size
- Separate spatially
- Outline
- Underline

Pre-attentive Attributes

- Enable the audience to see what we want them to see before they even know they're seeing it!
- A bright blue will typically draw attention more than a muted blue. Both will draw more attention than a light gray.

Pre-attentive Attributes in Graphs

- Leverage color to draw attention.
- Create a visual hierarchy of information.
- Data labels used sparingly help draw attention.
- Relative size denotes relative importance.
- Use color sparingly.
- Use color consistently.
- Design with colorblind in mind.
- Be thoughtful of tone that color conveys, as color evokes emotion.
- Position on page. Users read in Z pattern.

Make it Better

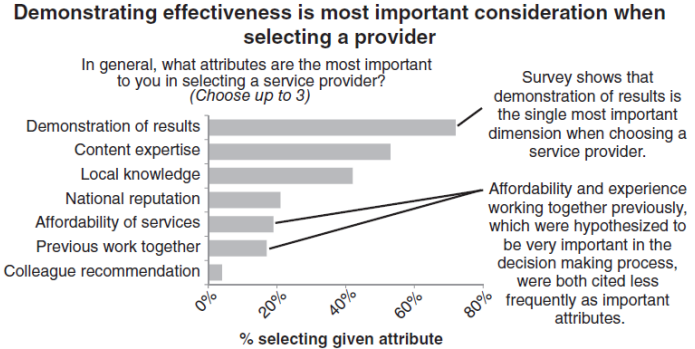


FIGURE 3.13 Summary of survey feedback

This one?

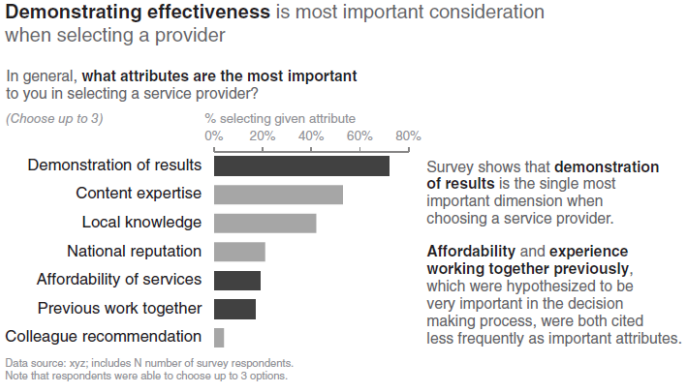


FIGURE 3.14 Revamped summary of survey feedback

Make it Better

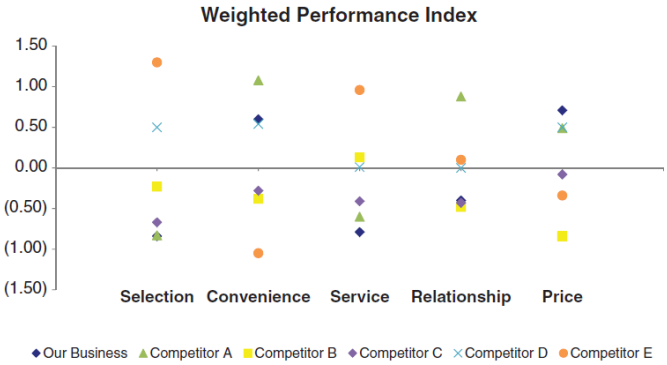


FIGURE 3.15 Original graph

This one?

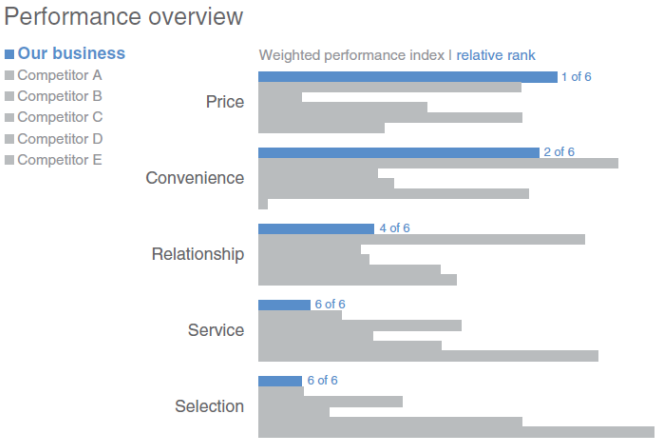


FIGURE 3.16 Revamped graph, using contrast strategically

Make it Better

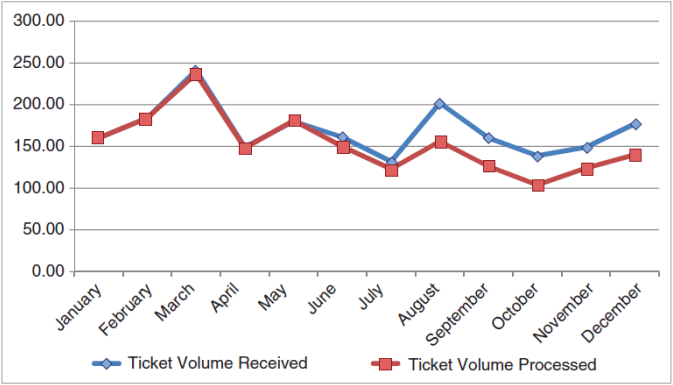


FIGURE 3.17 Original graph

This one?

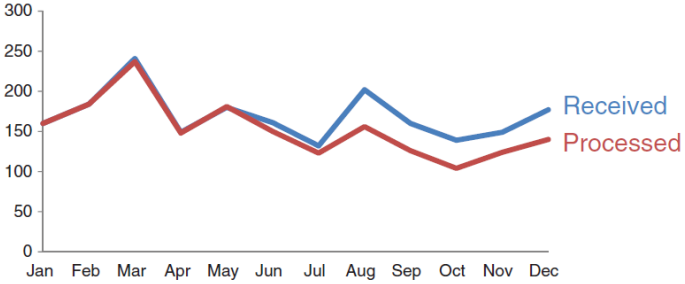


FIGURE 3.23 Leverage consistent color

Make it Better

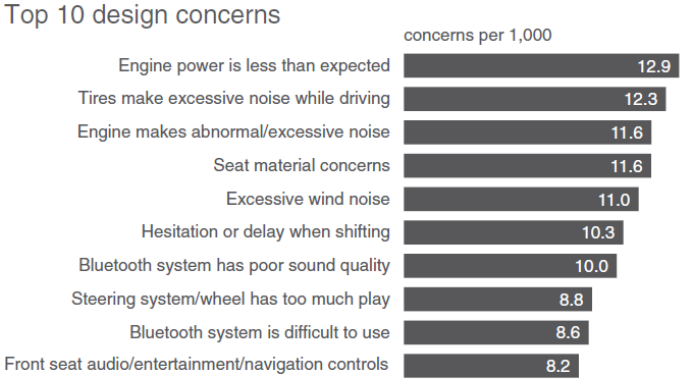


FIGURE 4.7 Original graph, no preattentive attributes

This one?

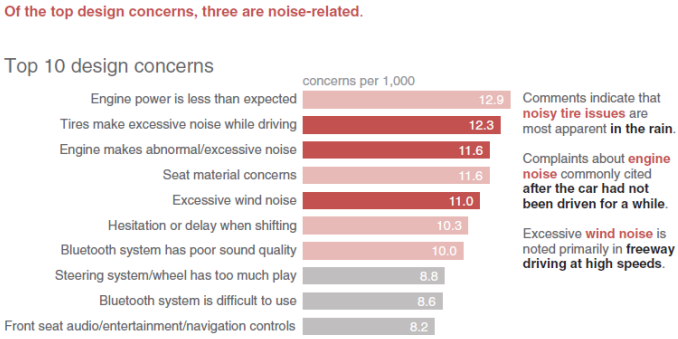
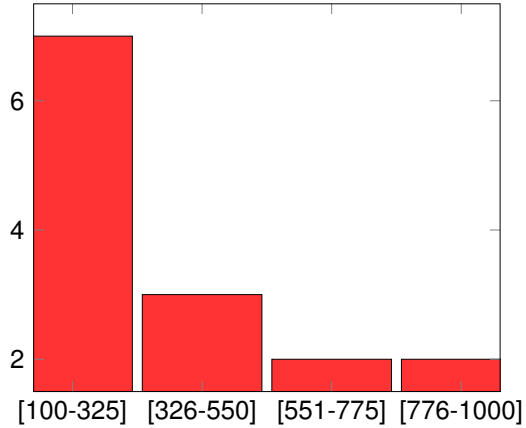


FIGURE 4.9 Create a visual hierarchy of information

Exercise

How can we make this better?



- Storytelling with Data by Cole Nussbaumer Knaflic; Wiley

THANK YOU