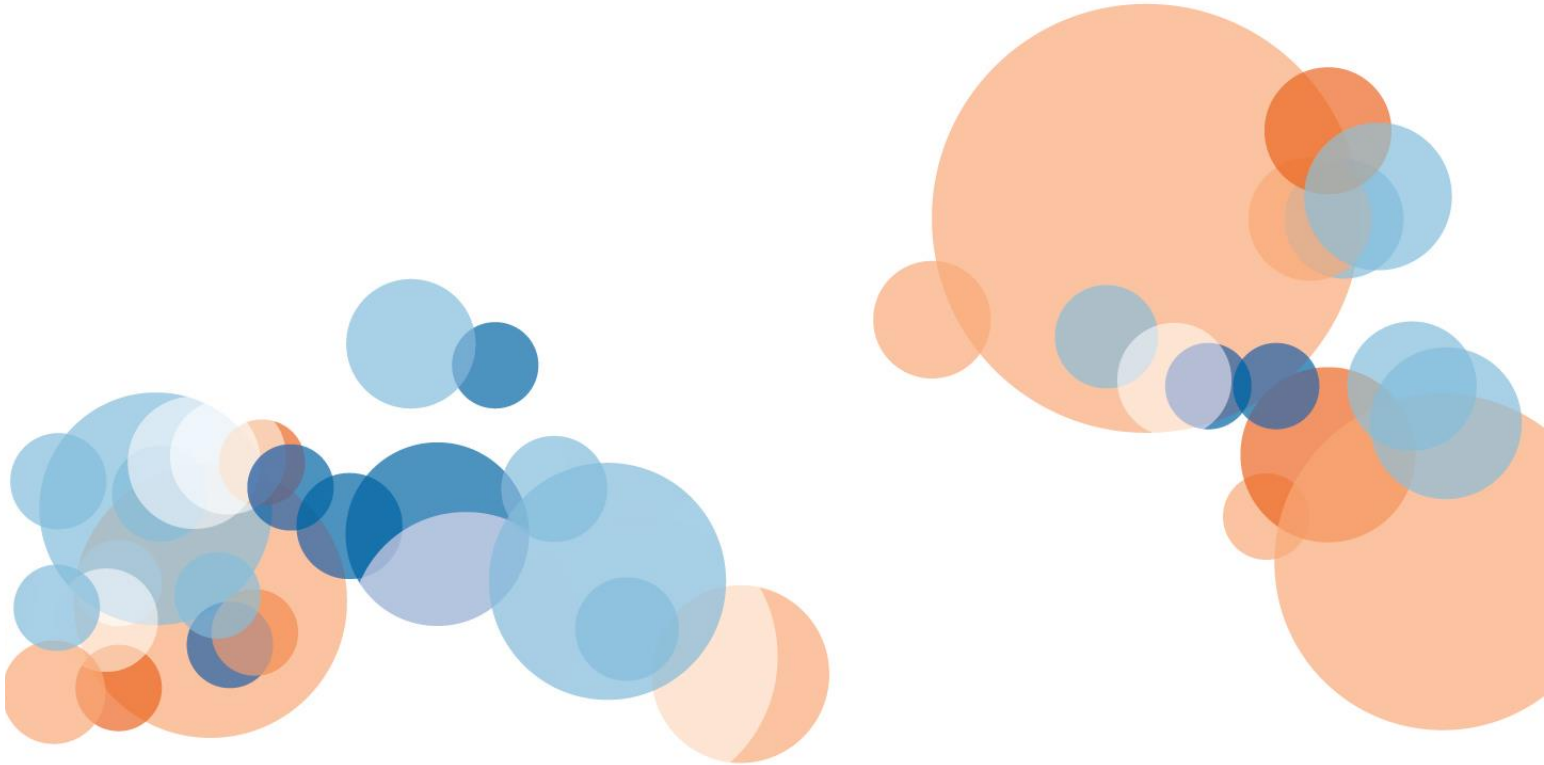
Data Visualization

Tatyana Yakushev

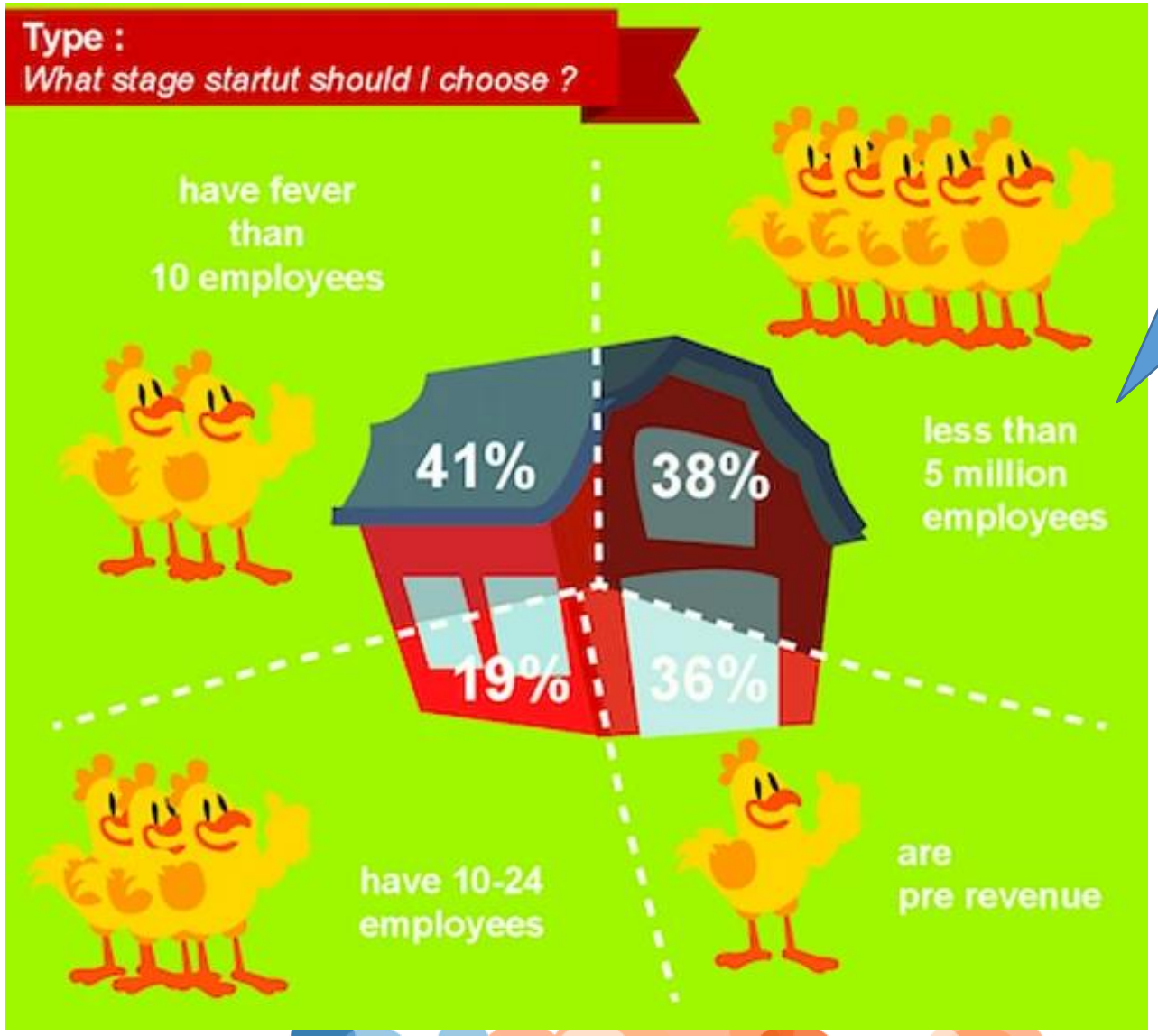
Senior Software Engineer

Tableau

5/16/2016



Why study visualization?



To create visualizations better than this one

Try to determine as quickly as you can the number of time the number five appears in the list

98735123489243218863218112134898432134268346213176132156401268984613216284988462132168366132168465132132068464213189198413651684162416841635678

That was hard, wasn't it?

How hard is it now?

98735123489243218863218112134898432134268346213176132156401268984613216284988462132168366132168465132132068464213189198413651684162416841635678

Why Visualize?

- “A picture is worth a thousand words”
- Visualization is the highest bandwidth channel into the human brain [Palmer 99]
- The visual cortex is the largest system in the human brain; it’s wasteful not to make use of it.
- As data volumes grow, visualization becomes a necessity rather than a luxury.

Graphs Reveal Data that Statistics May Not

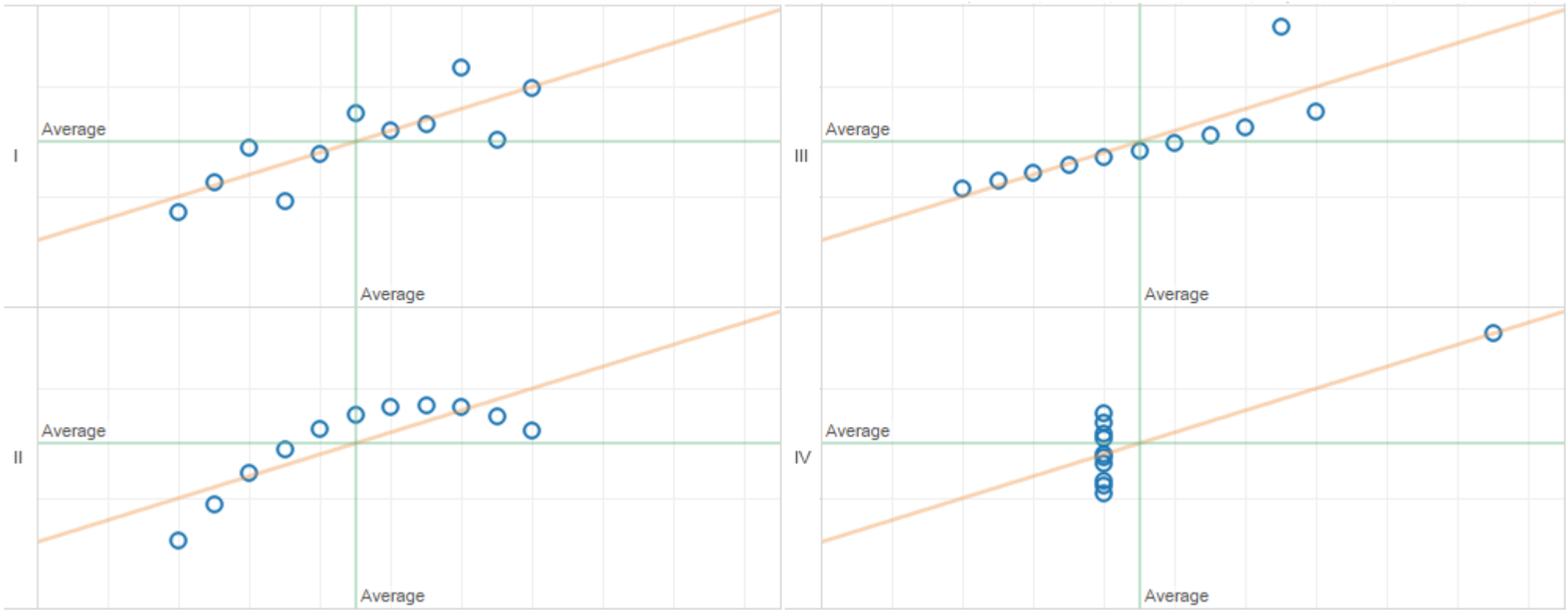
Four Data Sets with Identical Linear Model (Anscombe's Quartet)

Anscombe's quartet							
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

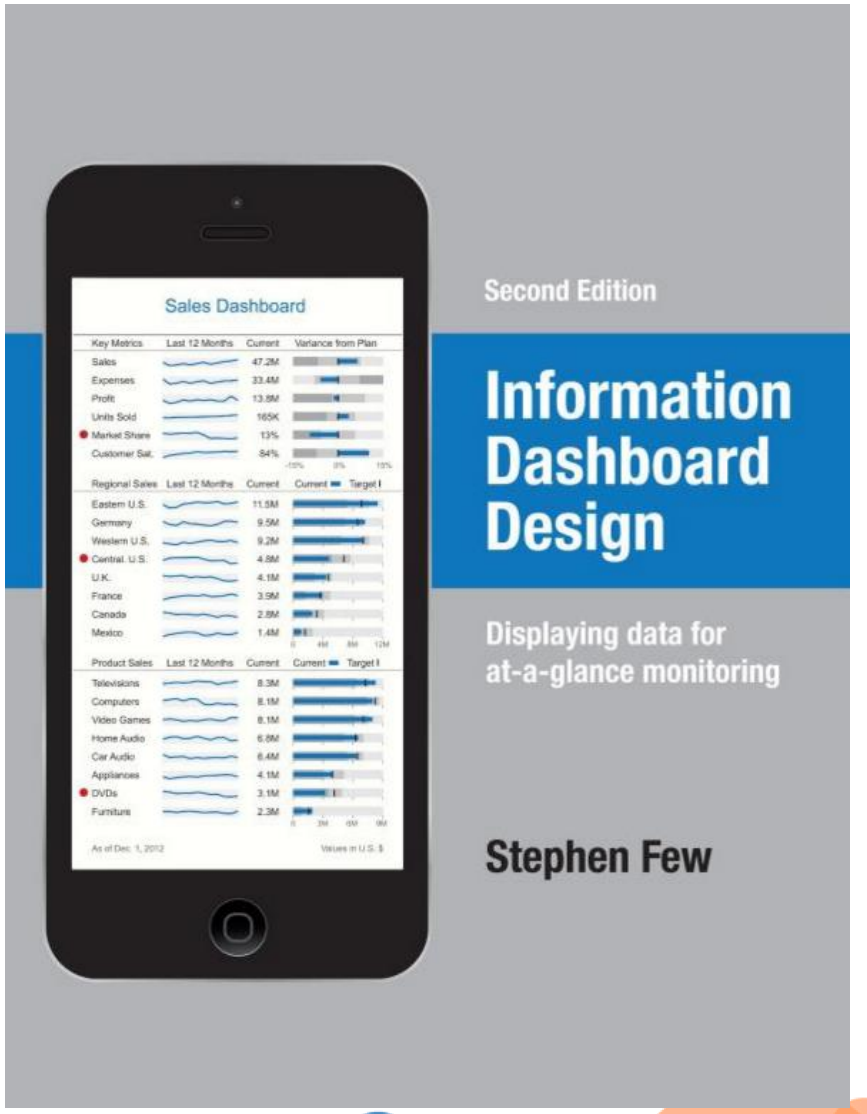
Graphs Reveal Data that Statistics May Not

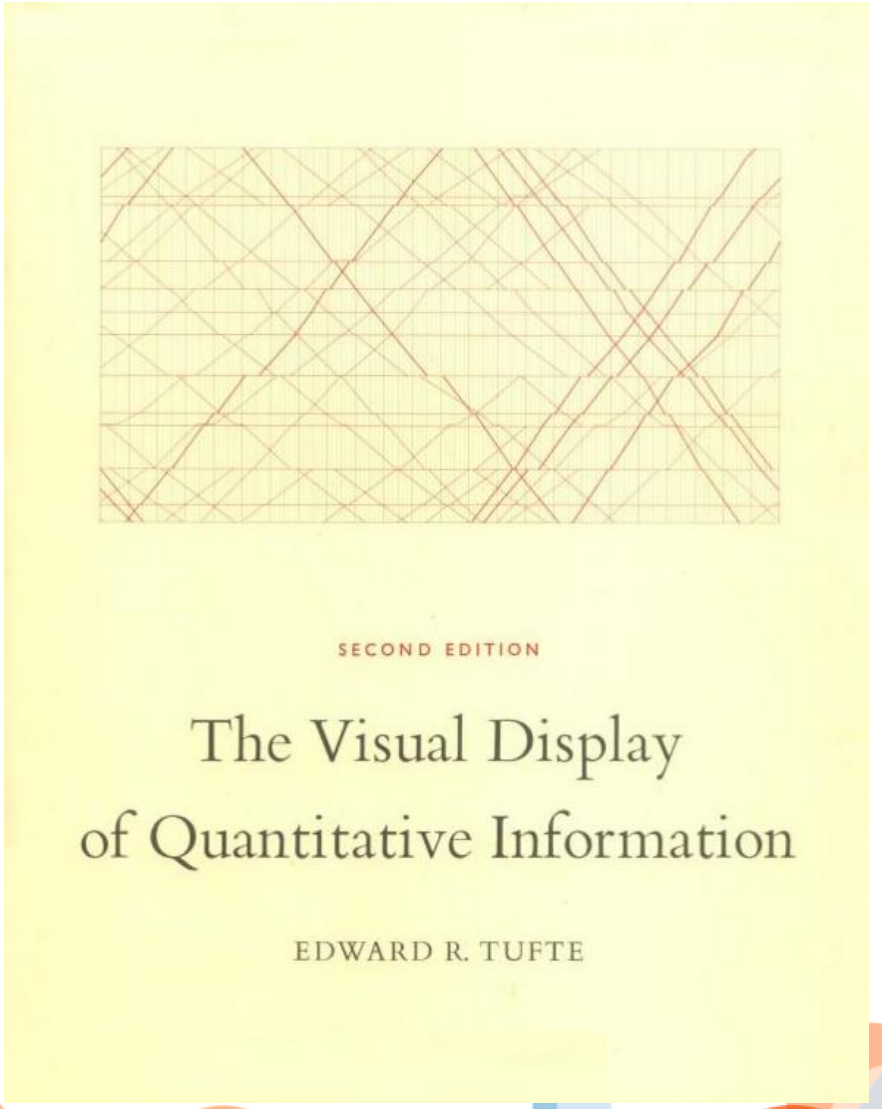
- mean of the x values = 9.0
- mean of the y values = 7.5
- equation of the least-squared regression line: y = 3 + 0.5x
- sums of squared errors (about the mean) = 110.0
- regression sums of squared errors (variance accounted for by x) = 27.5
- residual sums of squared errors (about the regression line) = 13.75
- correlation coefficient = 0.82
- coefficient of determination = 0.67

Graphs Reveal Data that Statistics May Not



Resources


Sales Dashboard
Second Edition
Information Dashboard Design
Displaying data for at-a-glance monitoring
Stephen Few


SECOND EDITION
The Visual Display of Quantitative Information
EDWARD R. TUFTE

13 Common Mistakes

1. Exceeding the boundaries of a single screen

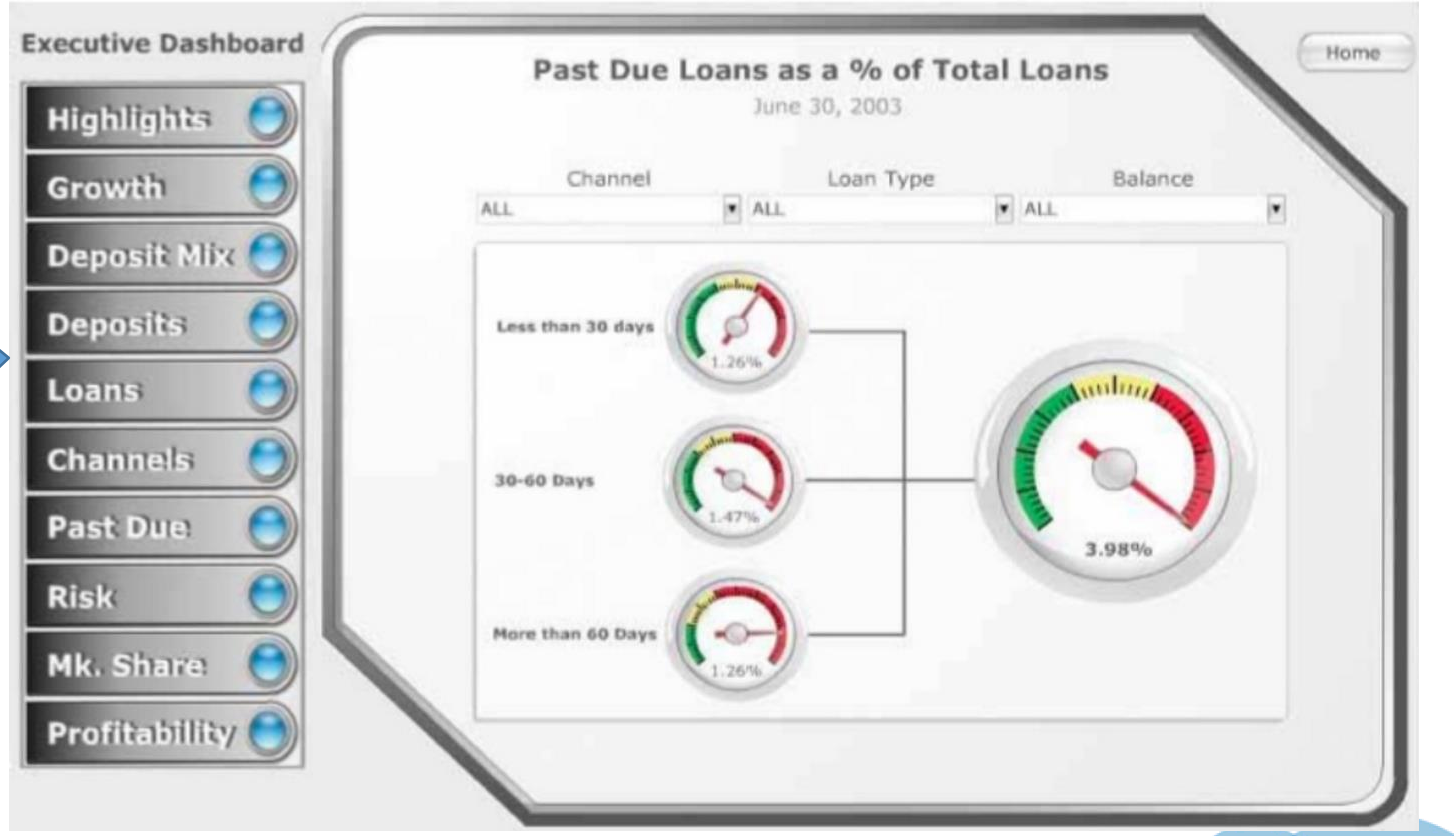
Information that appears on dashboards is often fragmented in one of two ways:

- Separated into discrete screens to which one must navigate
- Separated into different instances of a single screen that are accessed through some form of interaction

13 Common Mistakes

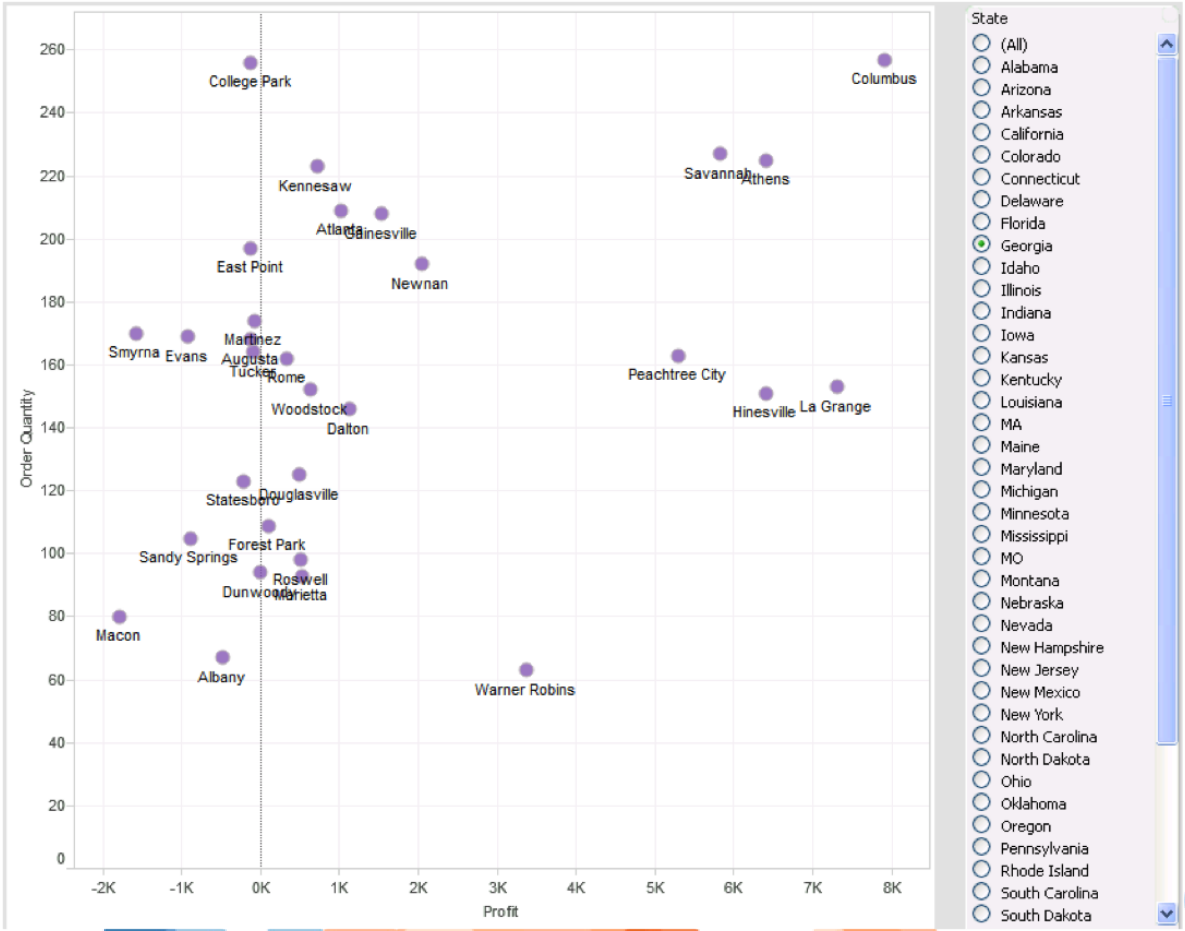
1. Exceeding the boundaries of a single screen

This dashboard fragments the data in a way that undermines the viewer's ability to see meaningful relationships.



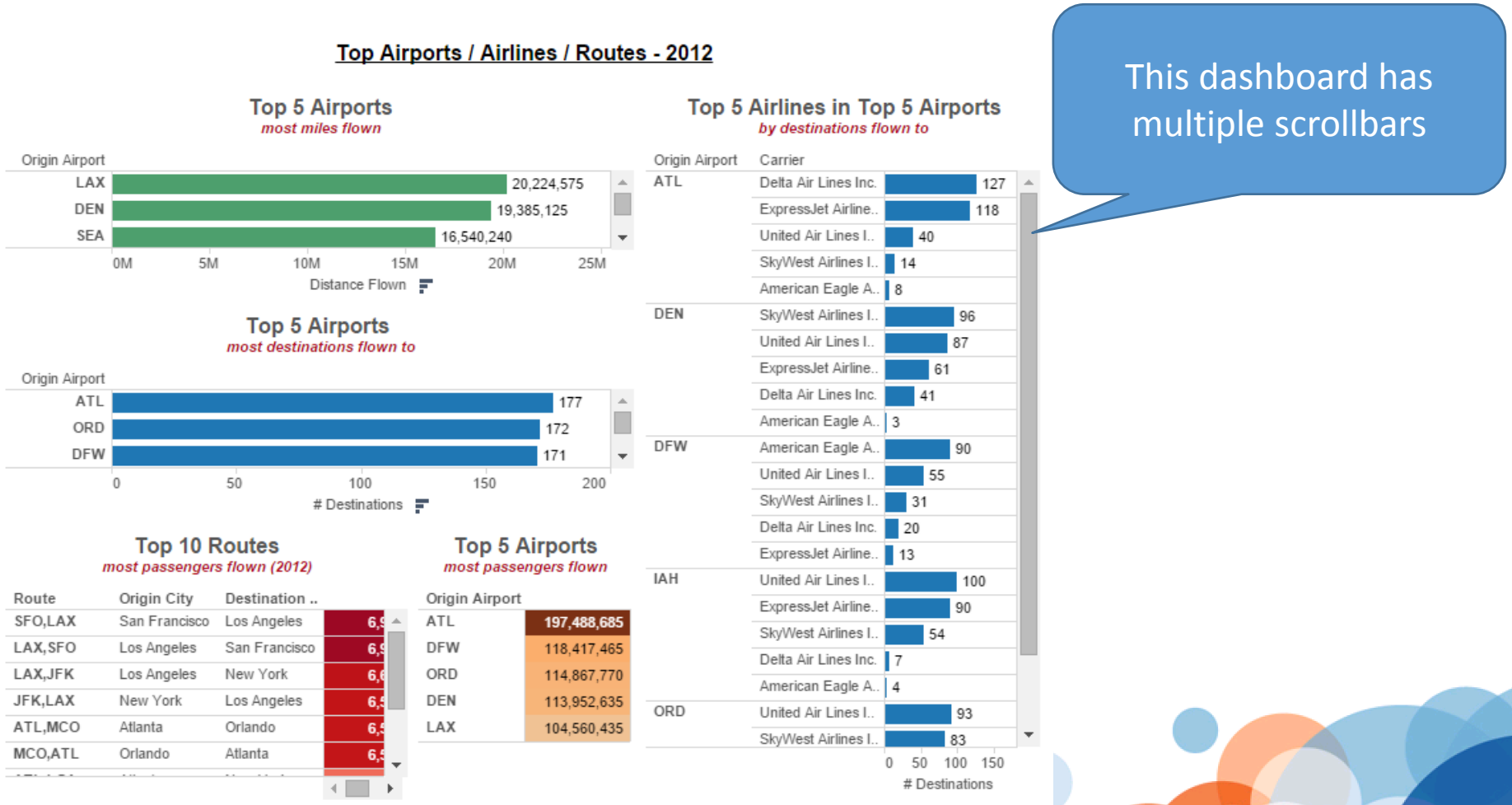
13 Common Mistakes

1. Exceeding the boundaries of a single screen



13 Common Mistakes

1. Exceeding the boundaries of a single screen



13 Common Mistakes

2. Supplying inadequate context for the data

or

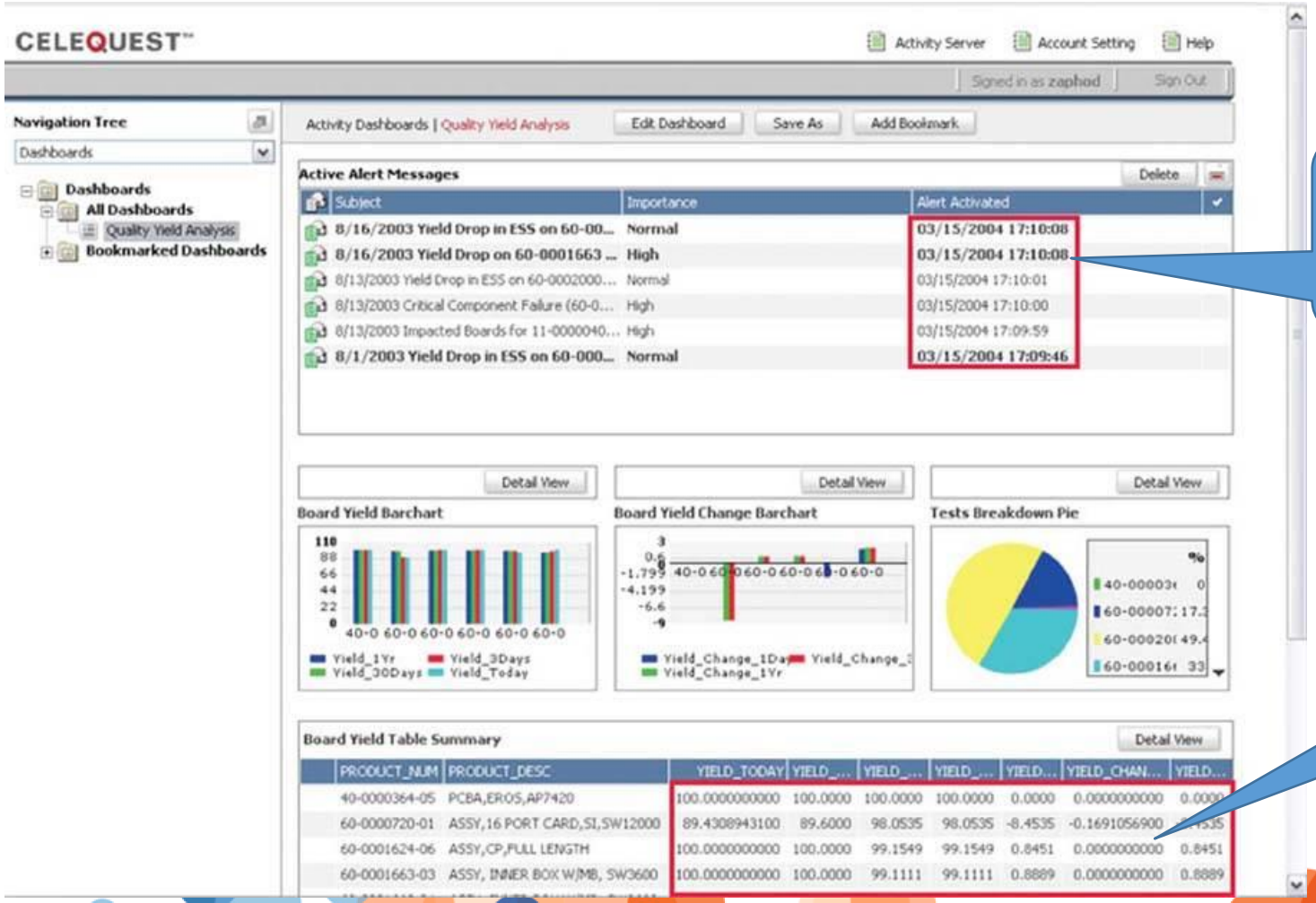
October Units	869
YTD Units	7,822
Returns Rate	0.26%

Both of the views are missing useful context

1. How bad or good are these numbers?
2. Are we on track?
3. Are we doing better or worse than in the past?

13 Common Mistakes

3. Displaying excessive detail or precision



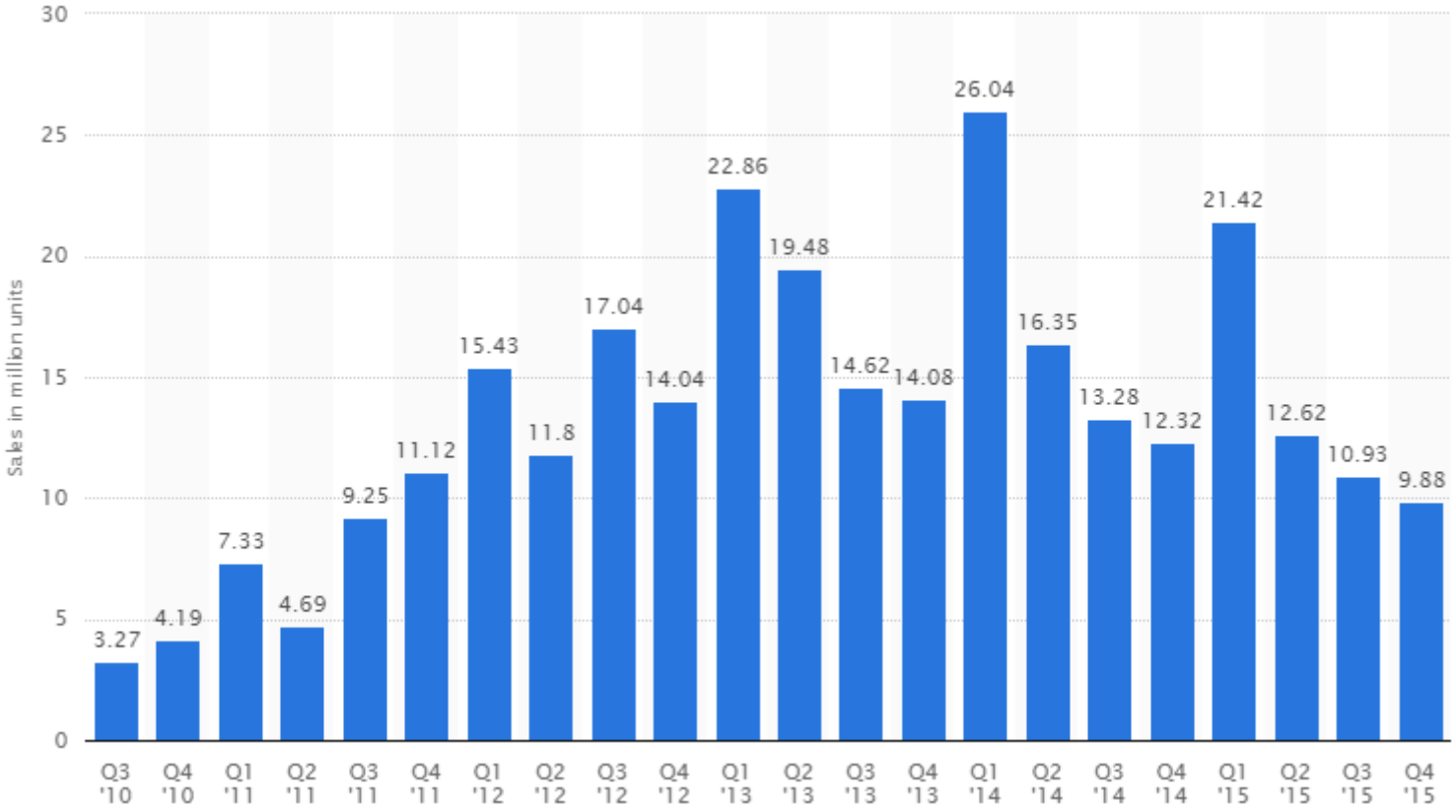
13 Common Mistakes

4. Choosing a deficient measure

13 Common Mistakes

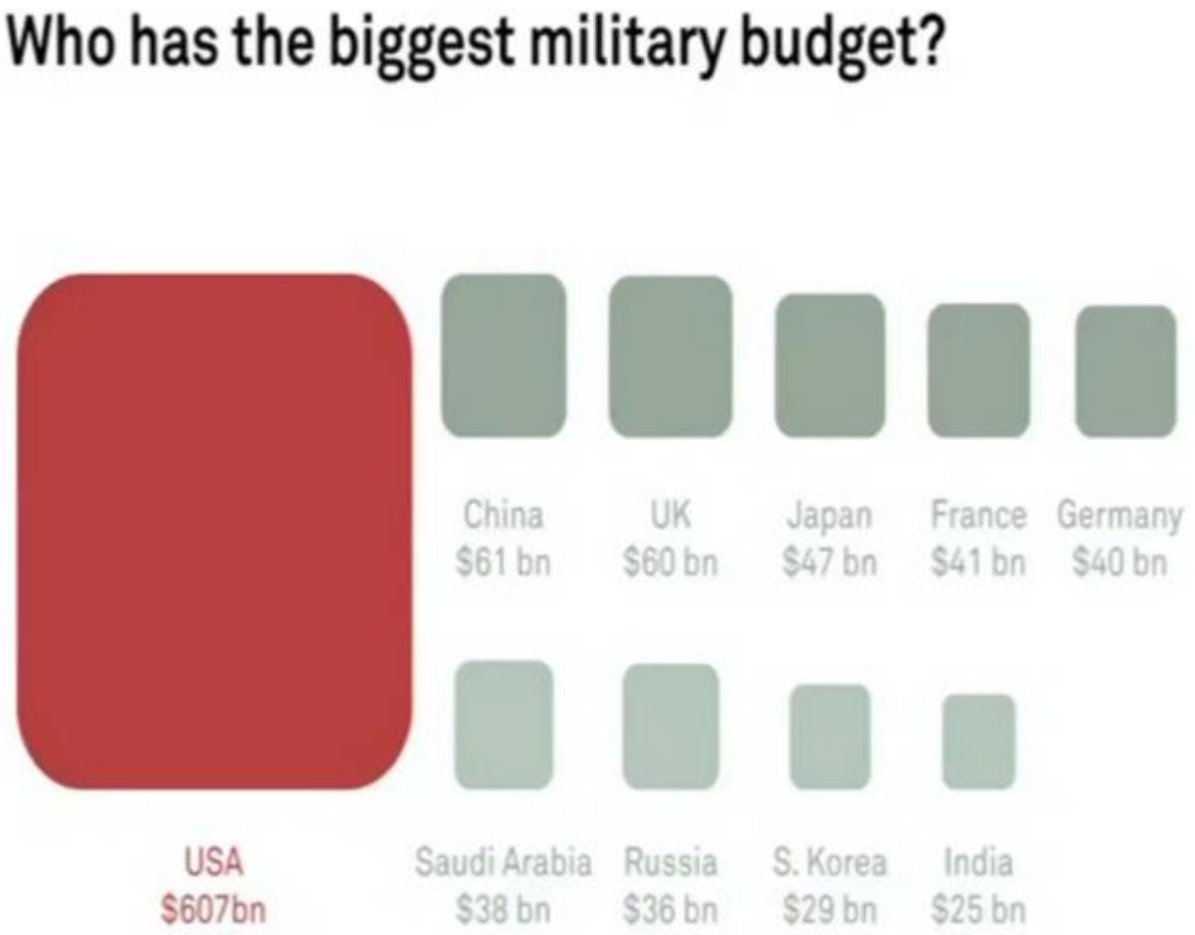
4. Choosing a deficient measure

IPad sales



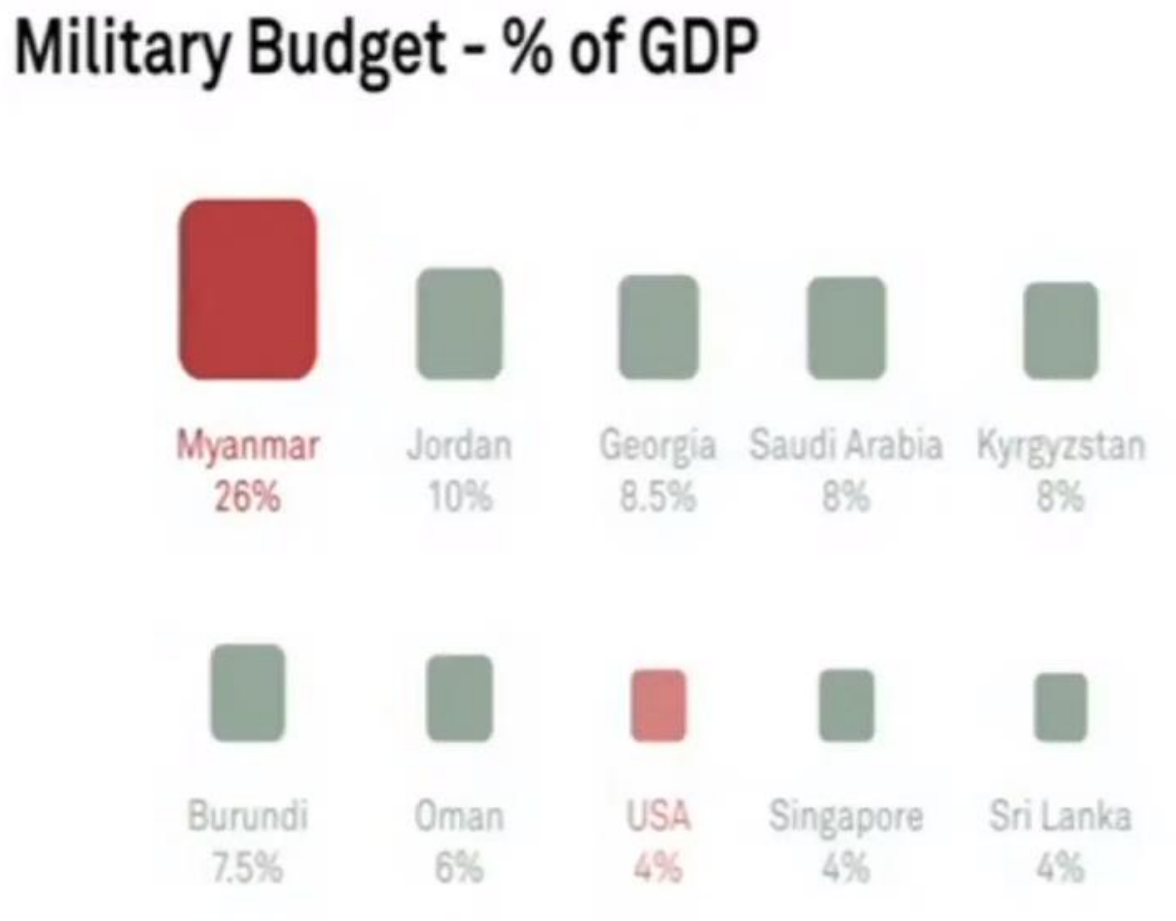
13 Common Mistakes

4. Choosing a deficient measure



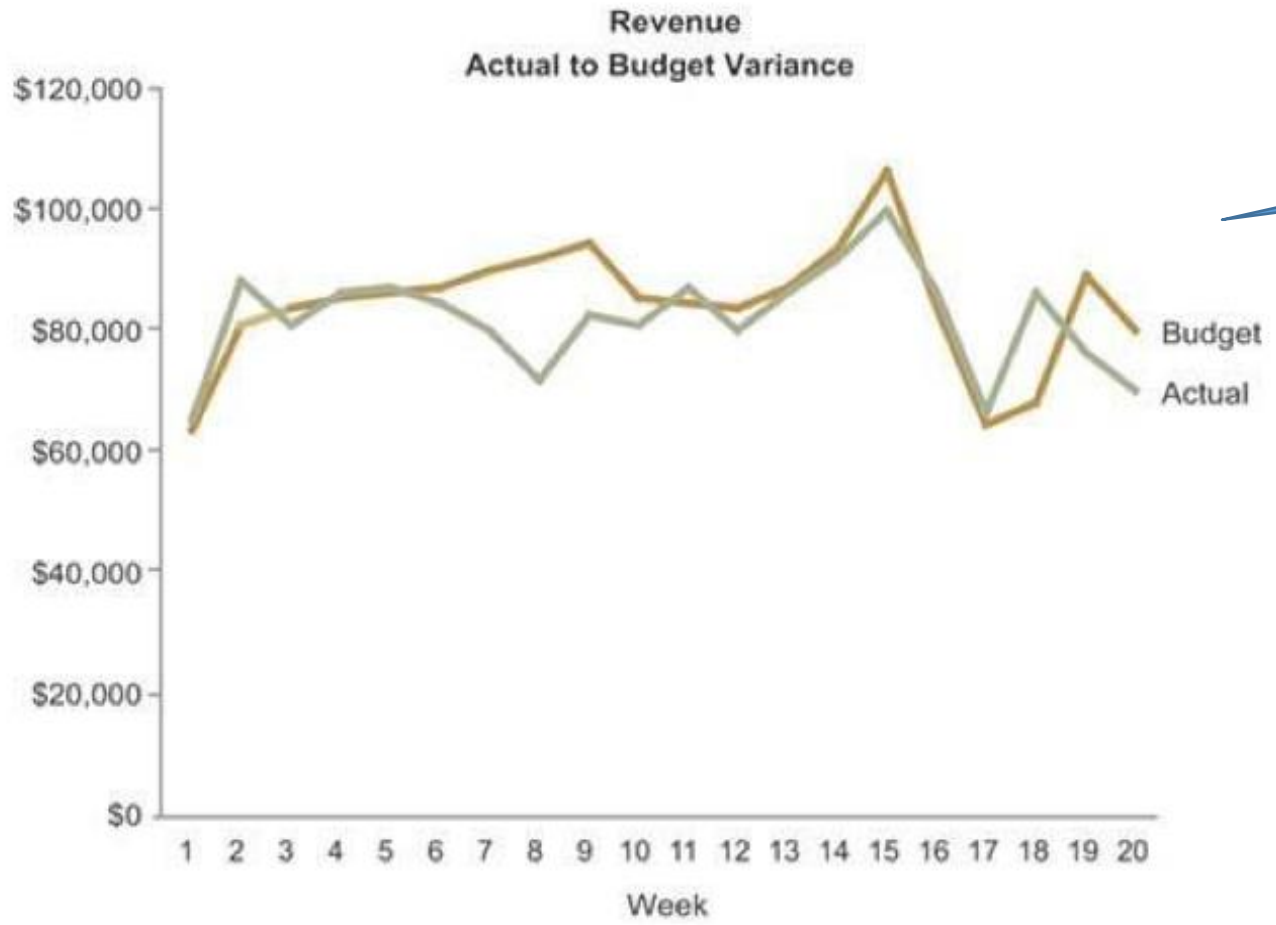
13 Common Mistakes

4. Choosing a deficient measure



13 Common Mistakes

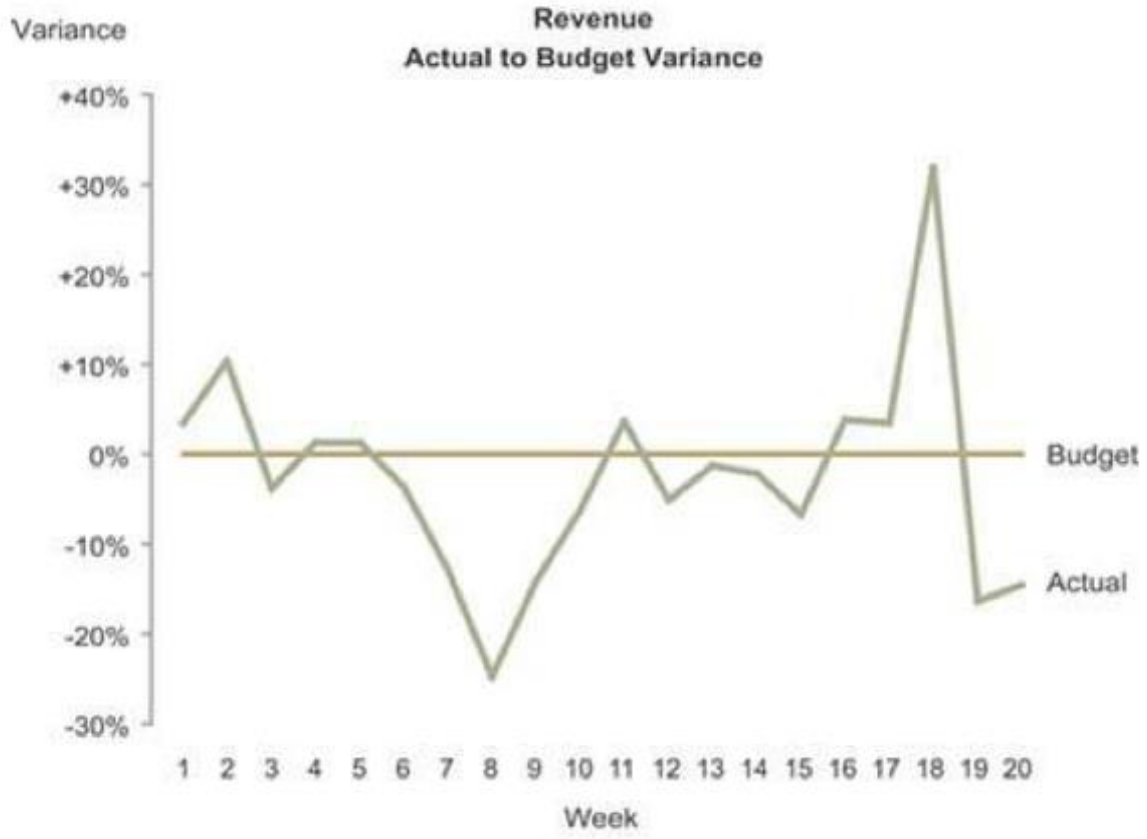
4. Choosing a deficient measure



This is not the best way to show variance of actual revenues from the budget

13 Common Mistakes

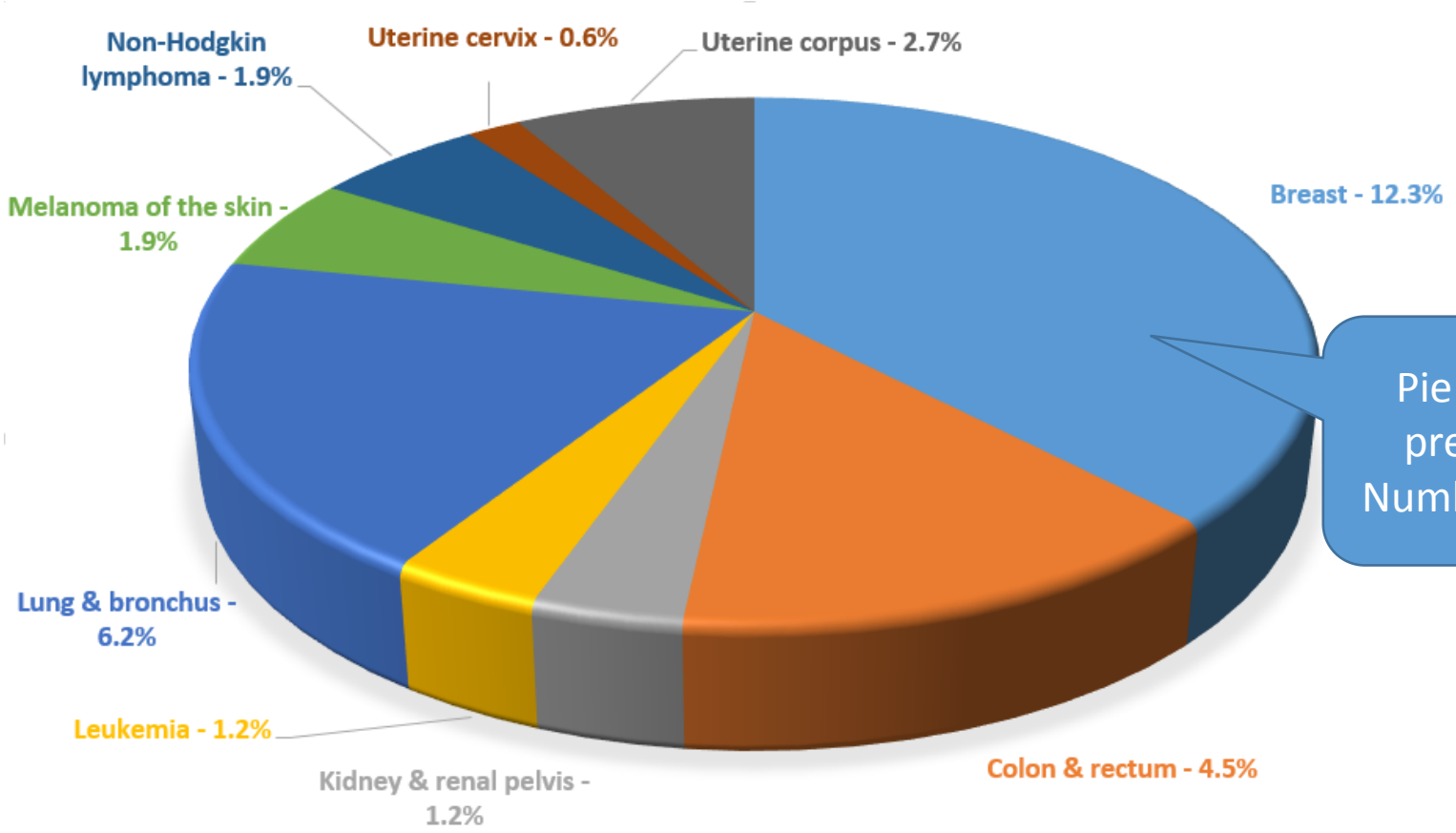
4. Choosing a deficient measure



This is much better

13 Common Mistakes

5. Choosing inappropriate media of display

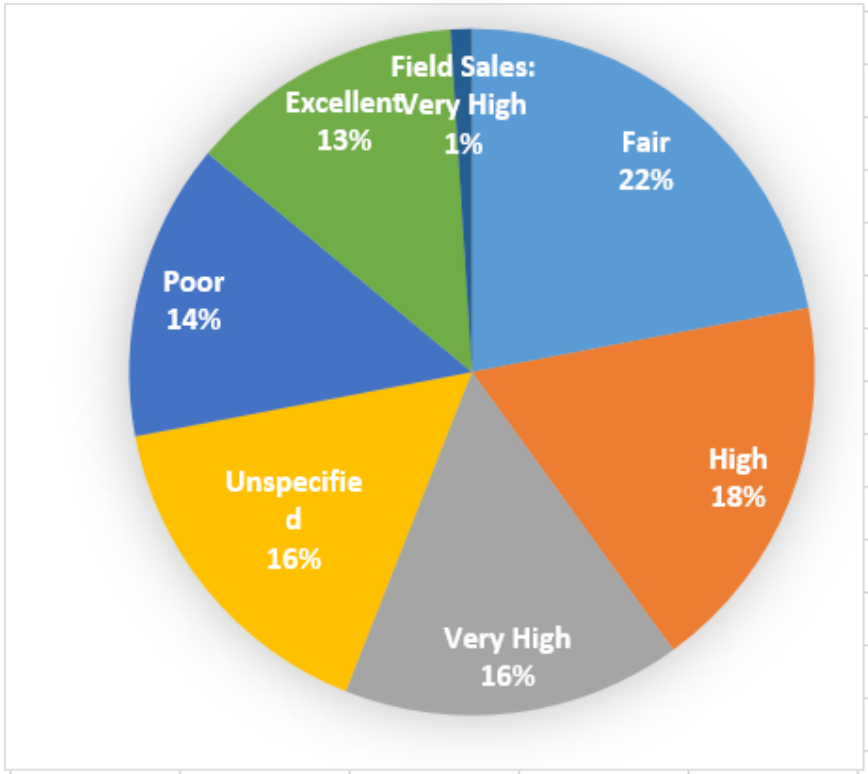


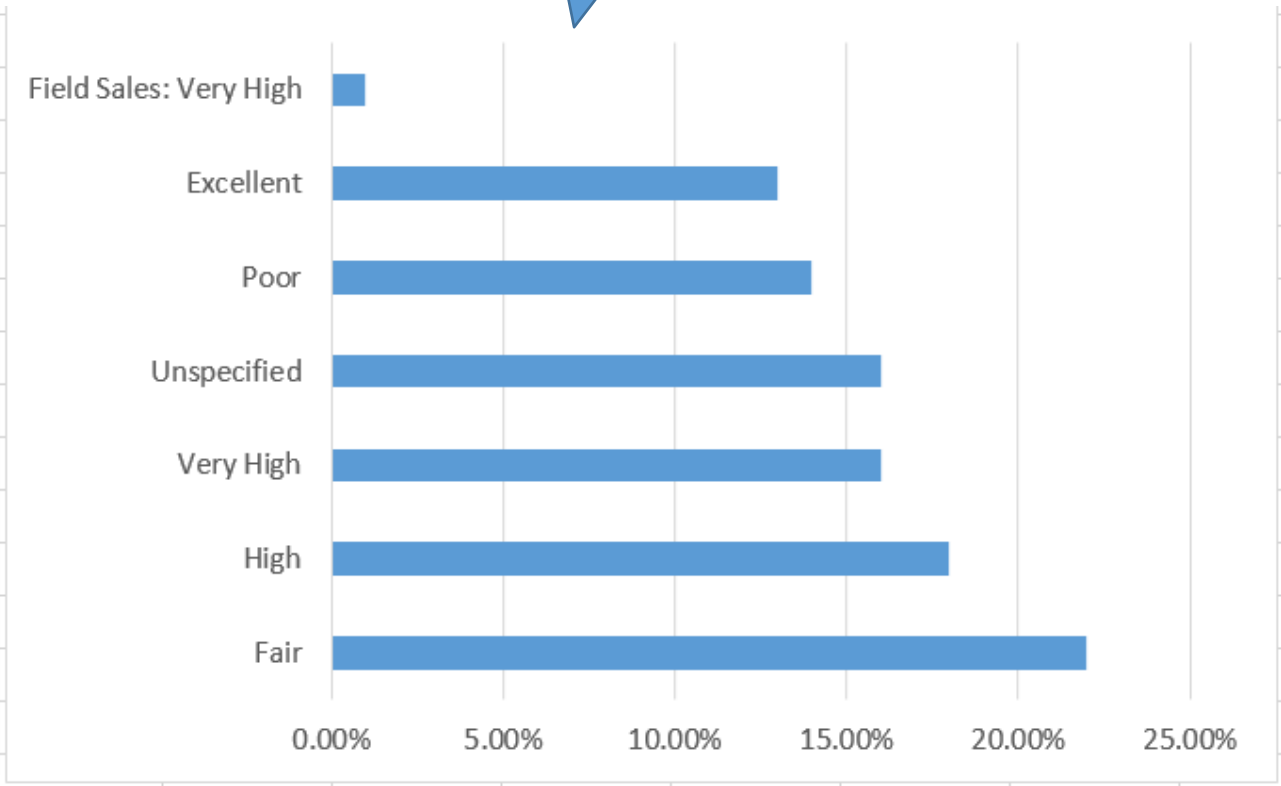
Pie charts are designed to present parts of a whole. Numbers should add to 100%

13 Common Mistakes

5. Choosing inappropriate media of display

Even when used correctly, pie charts are difficult to interpret accurately. Visualization on the right is much better.

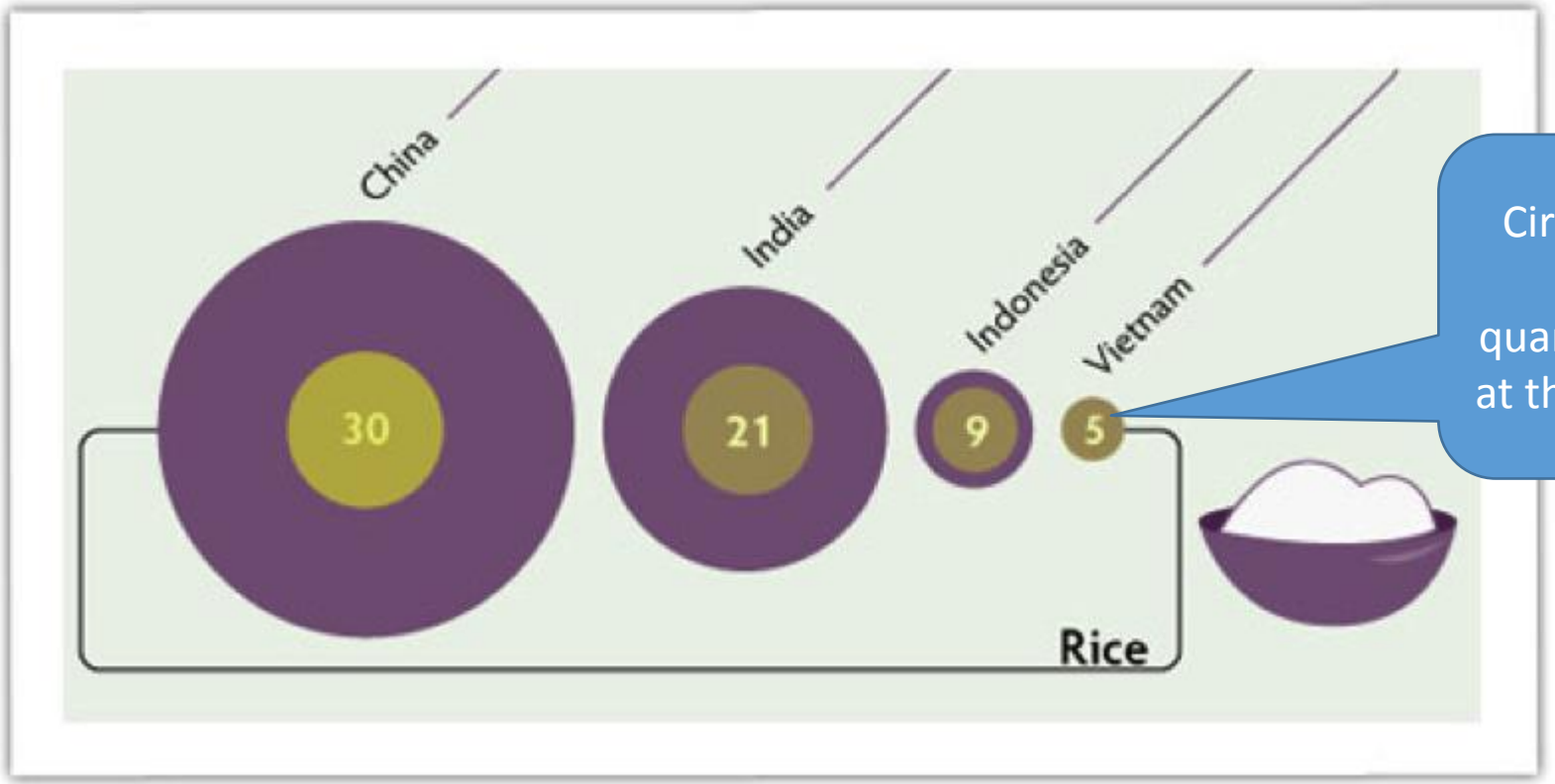




13 Common Mistakes

5. Choosing inappropriate media of display

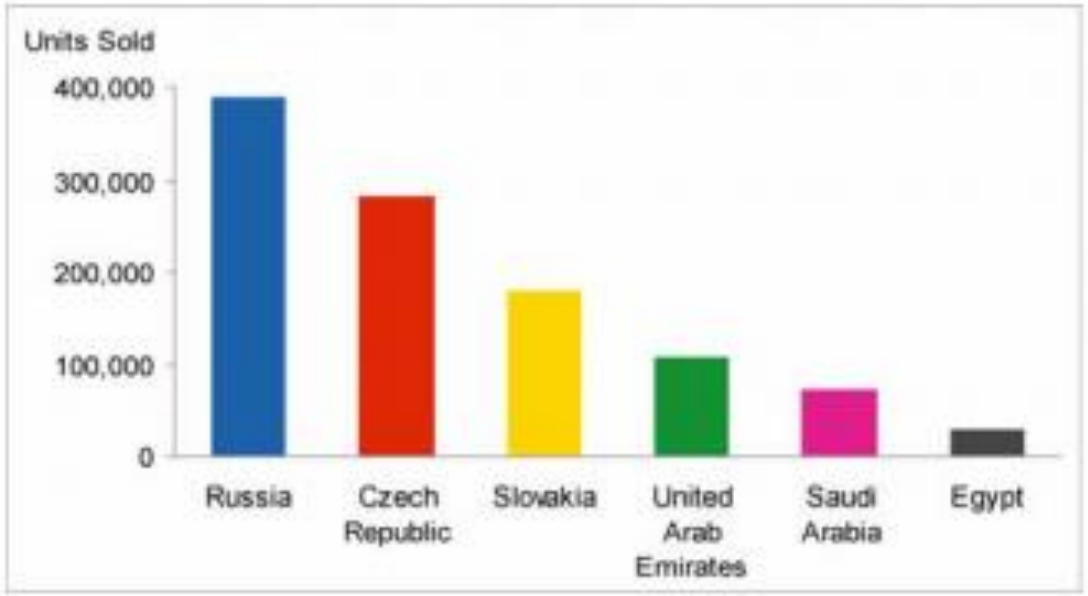
Other types of graphs can be equally ineffective

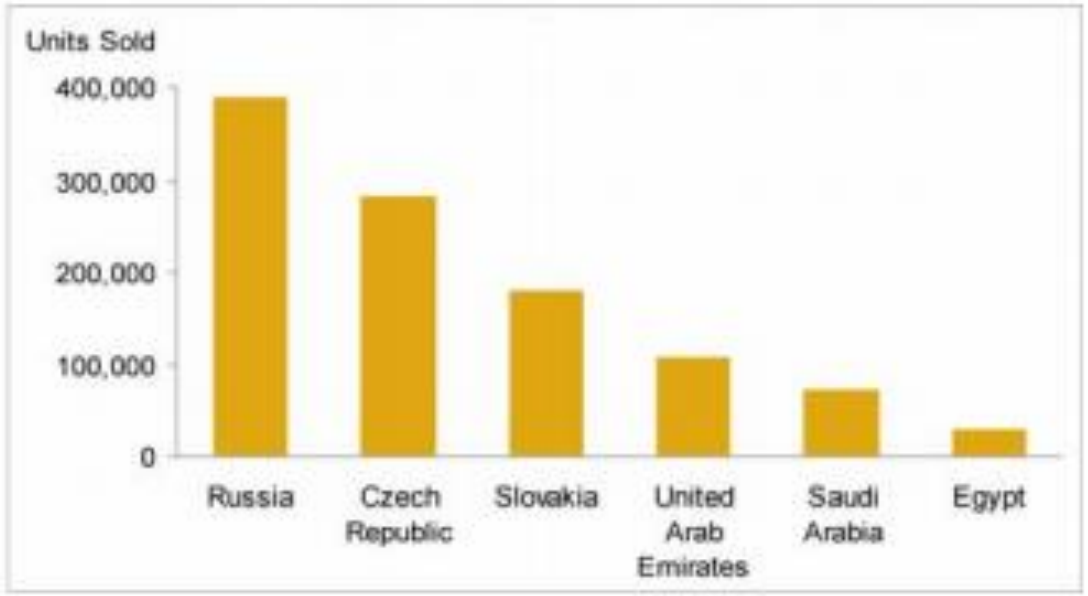


Circles are not a good way to encode quantities. Should I look at the diameter or area?

13 Common Mistakes

6. Introducing meaningless variety

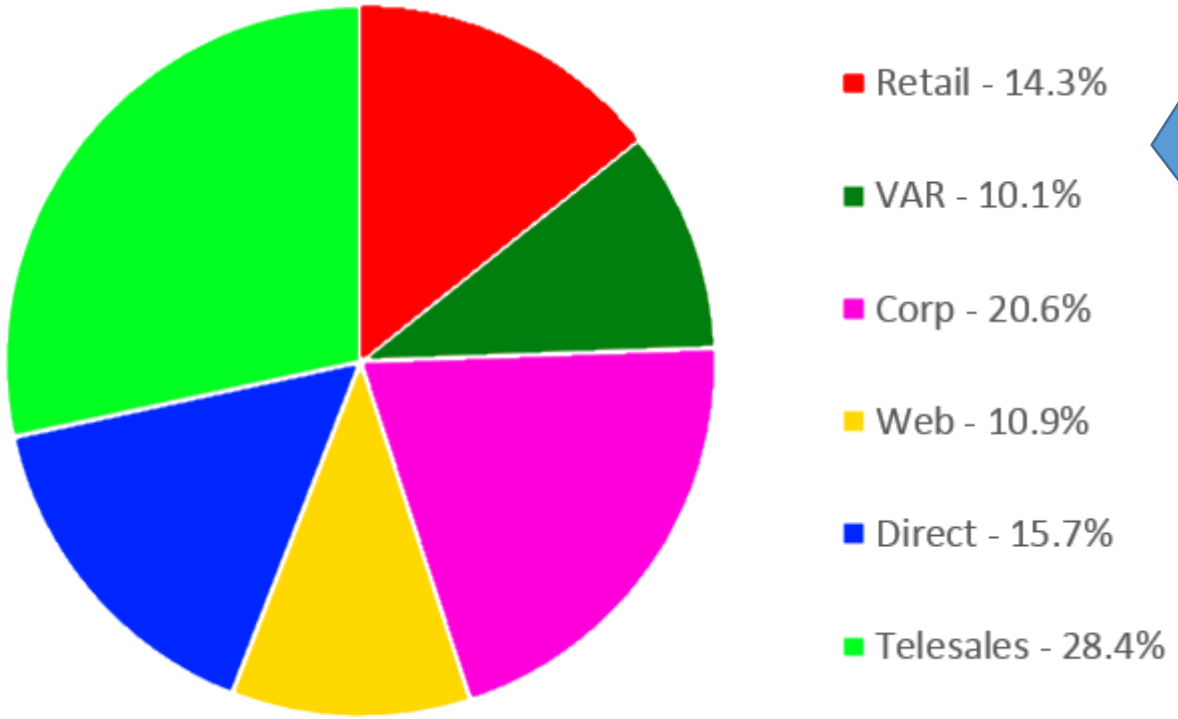




The bars on the left vary in color for no meaningful reason.

13 Common Mistakes

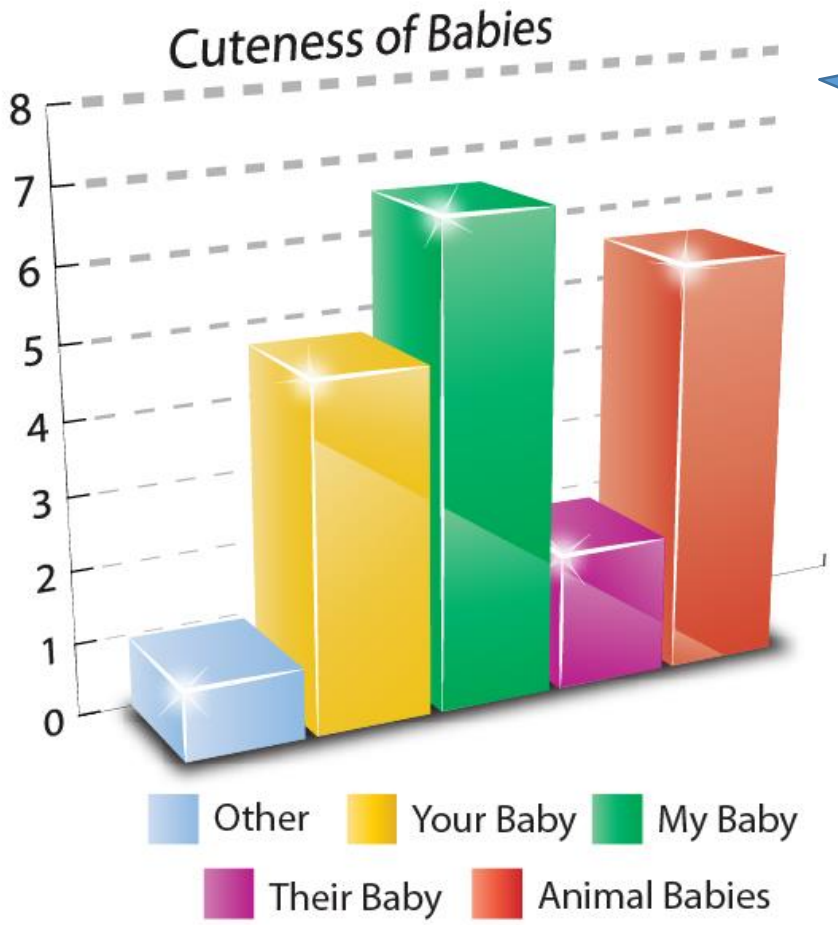
7. Using poorly designed display media



1. A legend was used to label and assign values to the slices of the pie. This forces our eyes to bounce back and forth between the graph and the legend to glean meaning, which is a waste of time and effort when the slices could have been labeled directly.
2. The order of the slices and the corresponding labels appears random. Ordering them by size would have provided useful information that could have been assimilated instantly.
3. The bright colors of the pie slices produce sensory overkill. Bright colors ought to be reserved for specific data that should stand out from the rest.

13 Common Mistakes

7. Using poorly designed display media

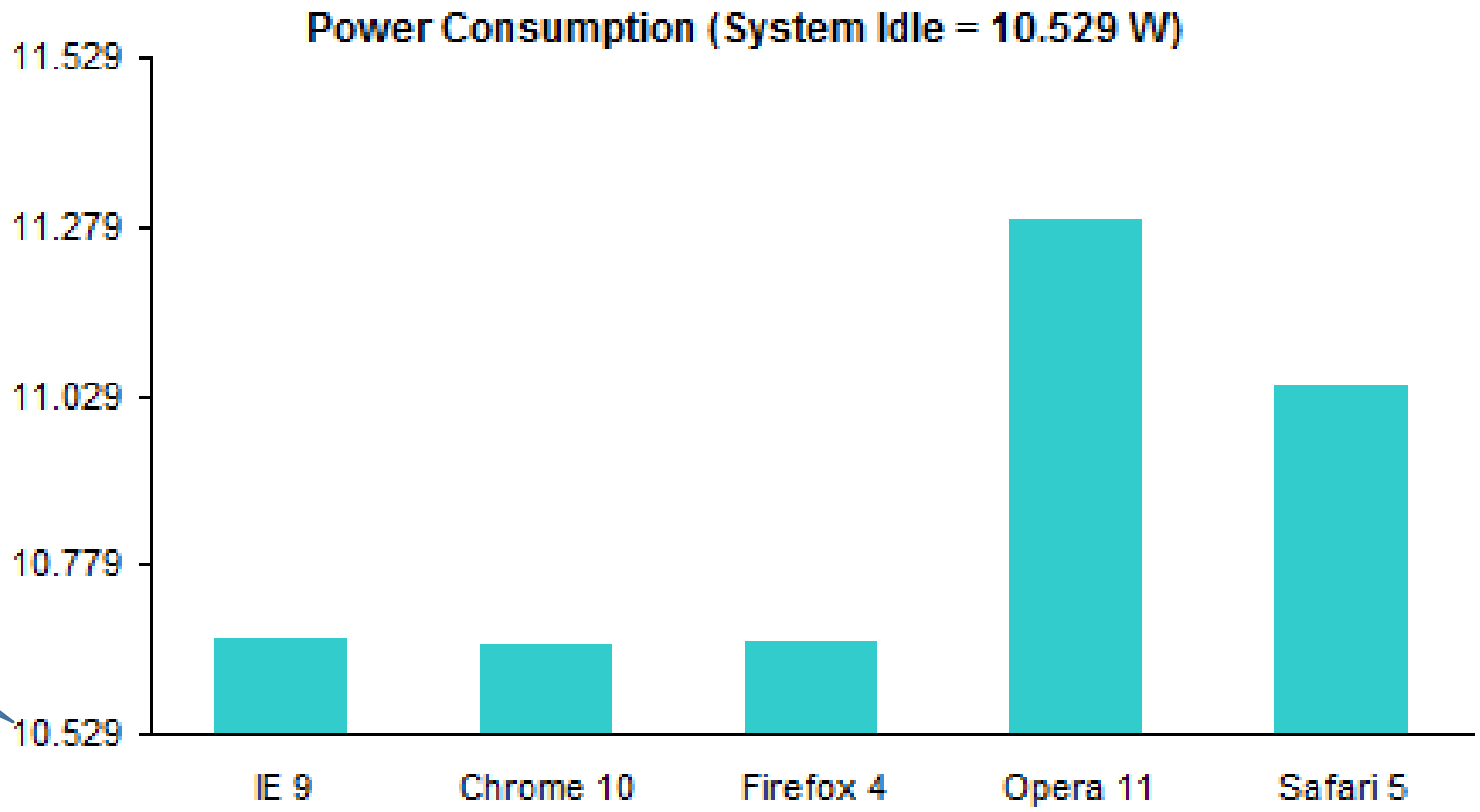


1. 3-D effect makes values encoded by the bars harder to interpret.
2. Perspective makes it more difficult to compare numbers

13 Common Mistakes

8. Encoding quantitative data inaccurately

By failing to begin the scale at zero, these bars inaccurately encode the quantitative values.



13 Common Mistakes

9. Arranging the data poorly

The most important data ought to be prominent.

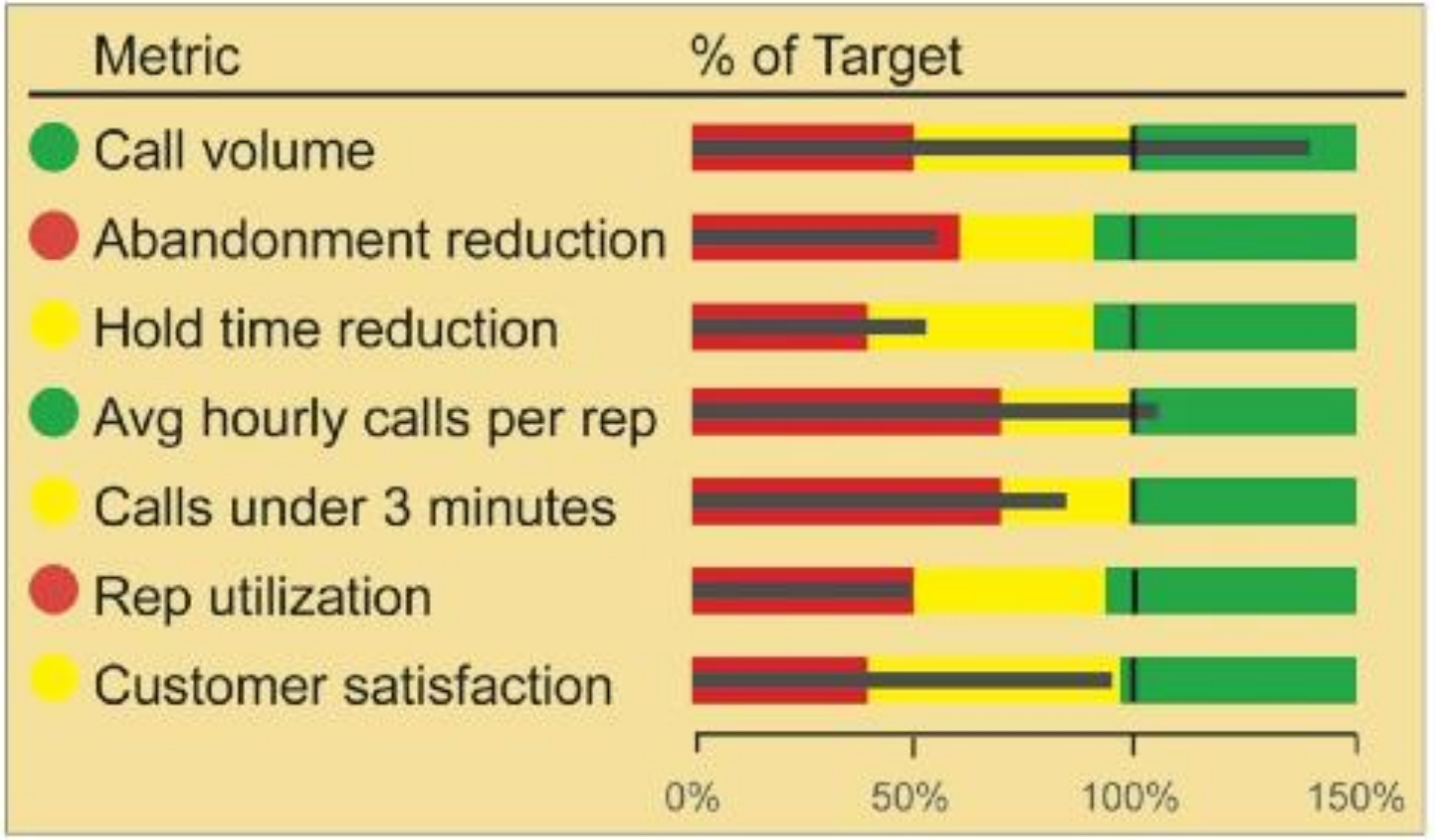
- Data that require immediate attention ought to stand out.
- Data that should be compared ought to be arranged and visually designed to encourage comparisons.

13 Common Mistakes

9. Arranging the data poorly

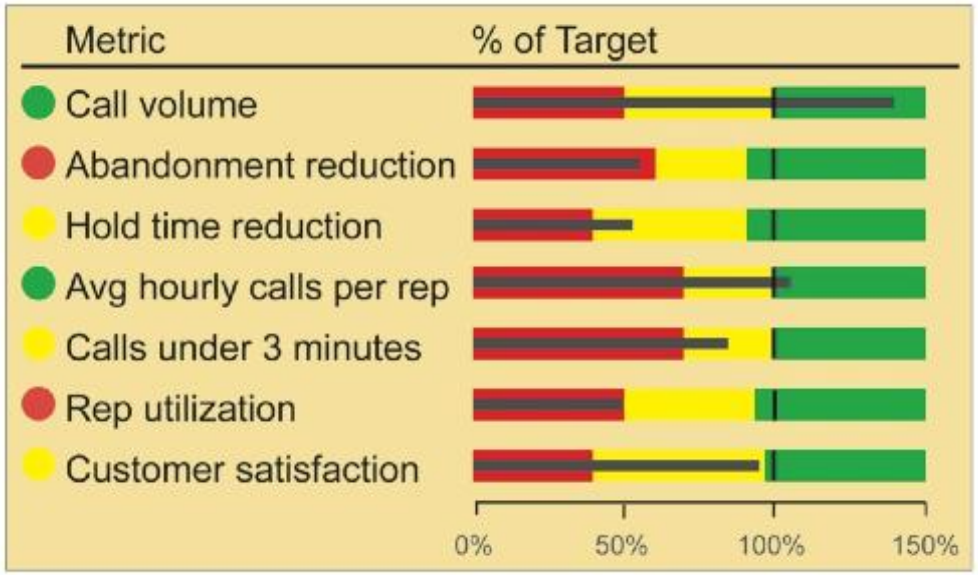
13 Common Mistakes

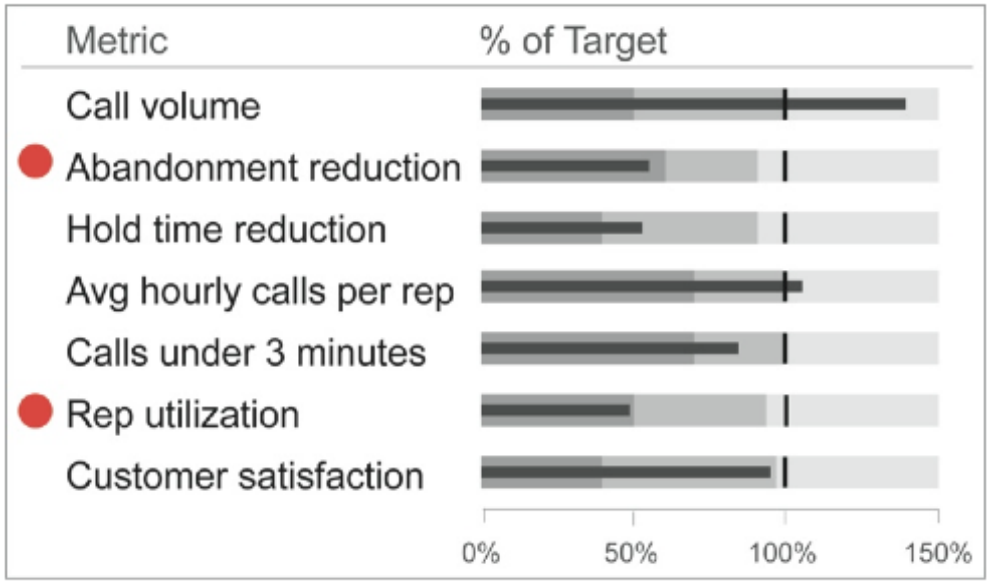
10. Ineffectively highlighting what's important



13 Common Mistakes

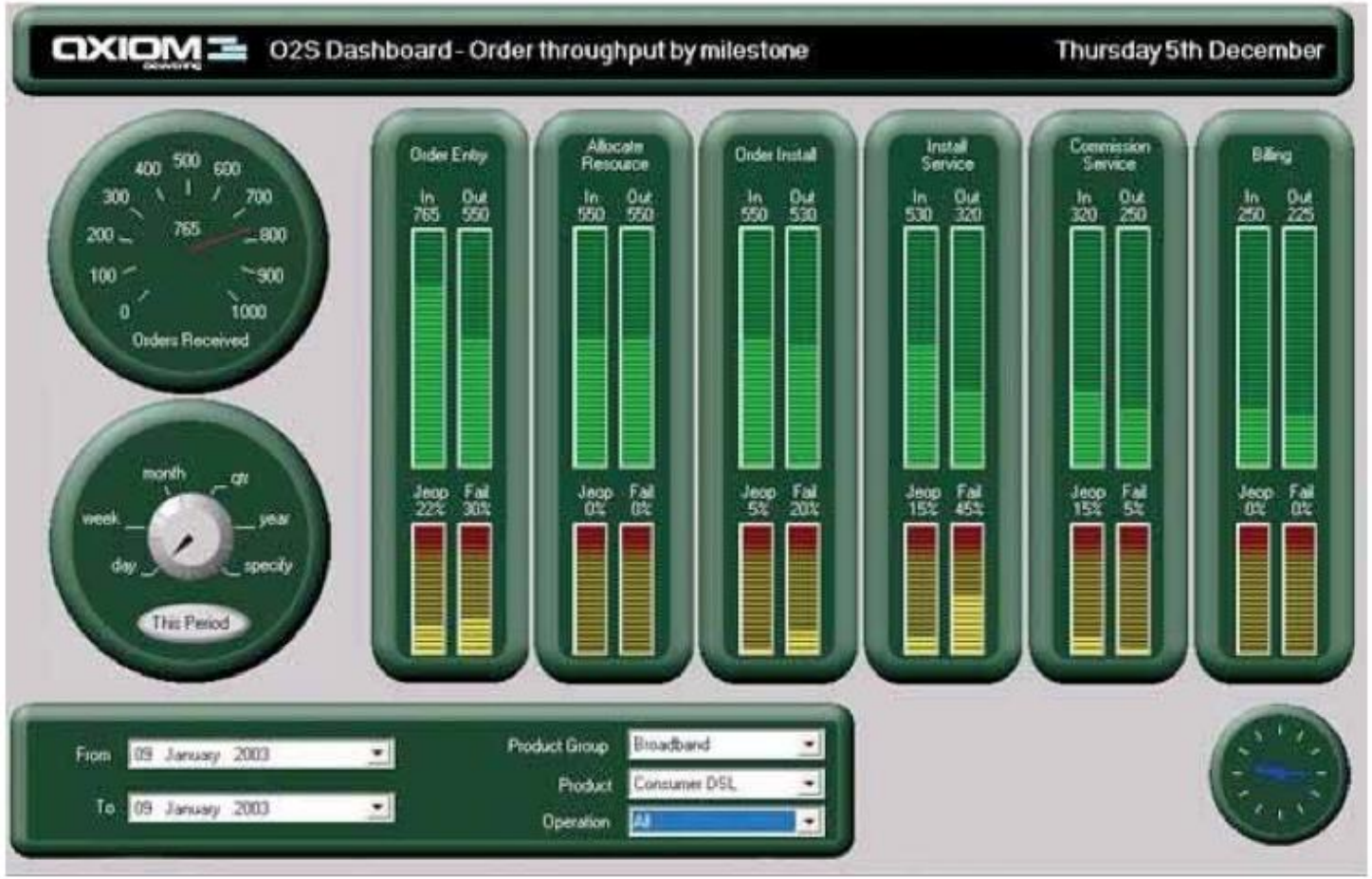
11. Misusing or overusing color





13 Common Mistakes

12. Cluttering the screen with useless decoration

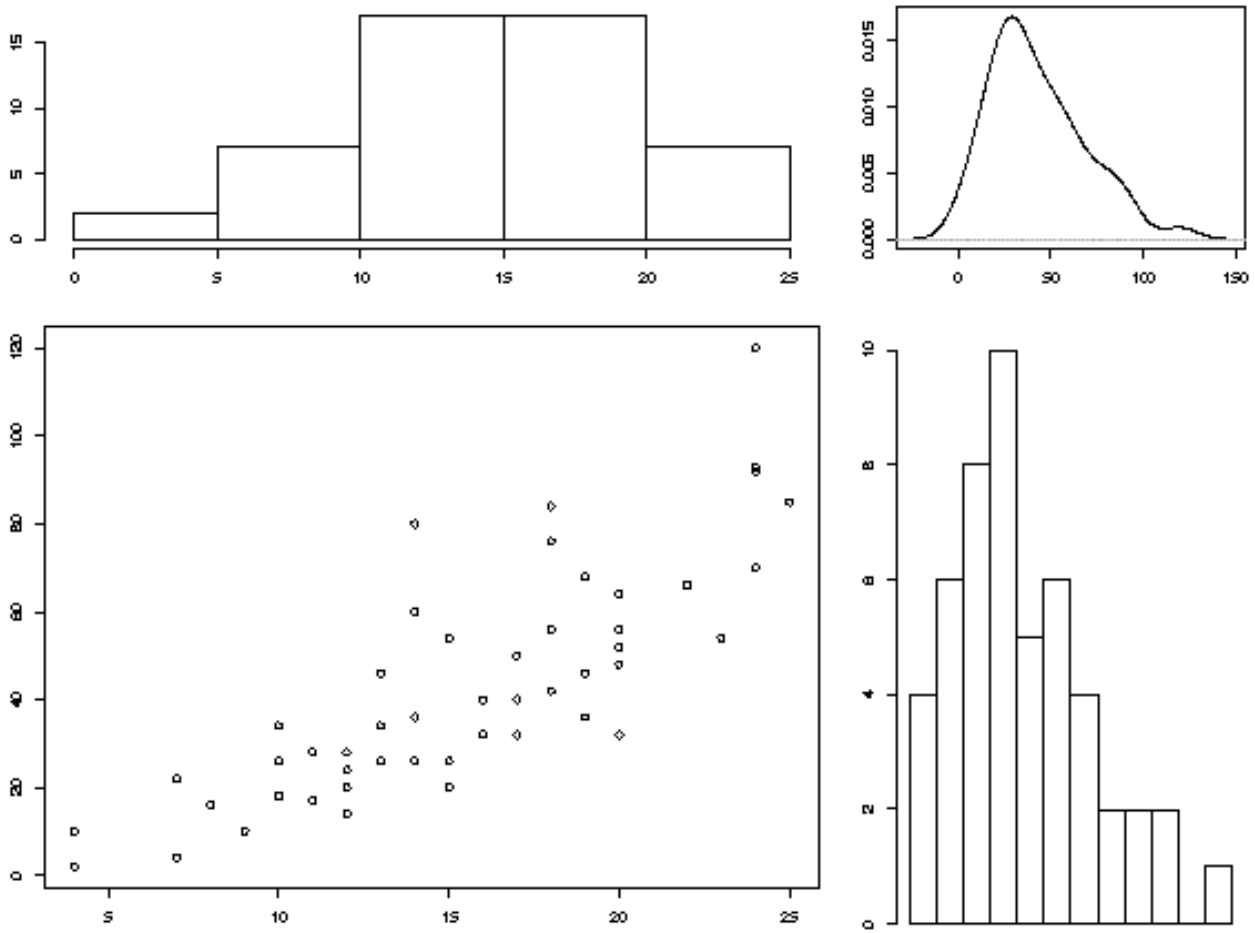


13 Common Mistakes

13. Designing an unappealing visual display

13 Common Mistakes

13. Designing an unappealing visual display



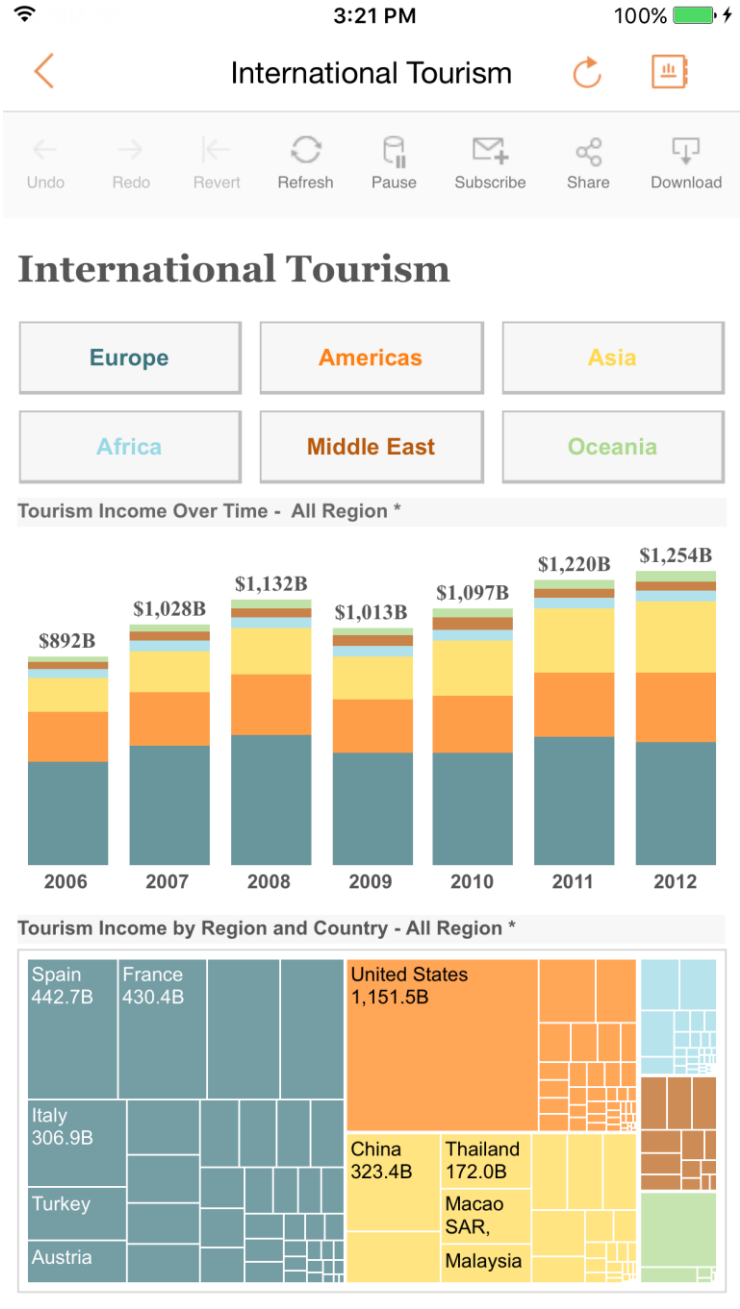
Data Visualization On Mobile Devices

- Small screen size
- Used on the go
- Used for short periods of time

tableau

Start with Your Focus

- What information do you need to present to the users?



Customize the Data for Each Person

- Find ways to personalized the information to each person instead of asking users to filter it themselves

Use Real-Time Data

- People use phones to monitor changes throughout the day

+ableau

Highlight Immediately-Actionable Data

- Current quota performance is likely to be more useful than historical sales data
- Highlight information that requires attention

tableau

Opt for Simple Views

- Simple views are more useful than complicated views and sophisticated dashboards
  - Summary
  - KPIs

Go for

- Line charts
- Bar charts
- Area charts
- Highlight tables
- Simplified field maps
- Dot maps

Avoid

- Scatter plots
- Line charts with many series
- Complex tables