INTRODUCTION TO DATA ANALYTICS

Data Visualization

Dr. Rathachai Chawuthai

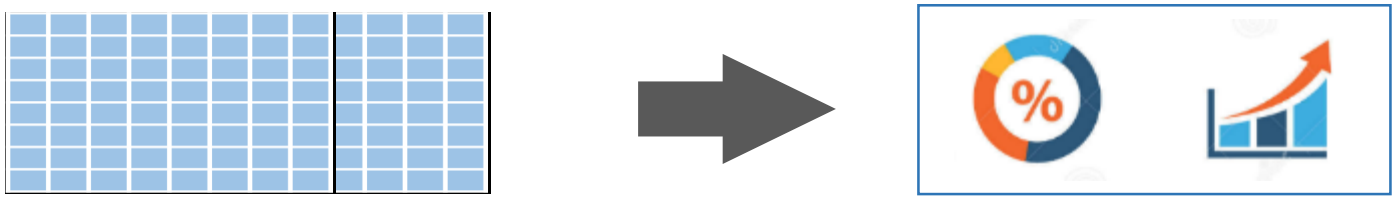
Department of Computer Engineering
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang

Agenda

- Introduction
- Pre-Attractive Attributes
- Visual Encoding
- Charts
- Dashboard

Introduction

Recall



Data

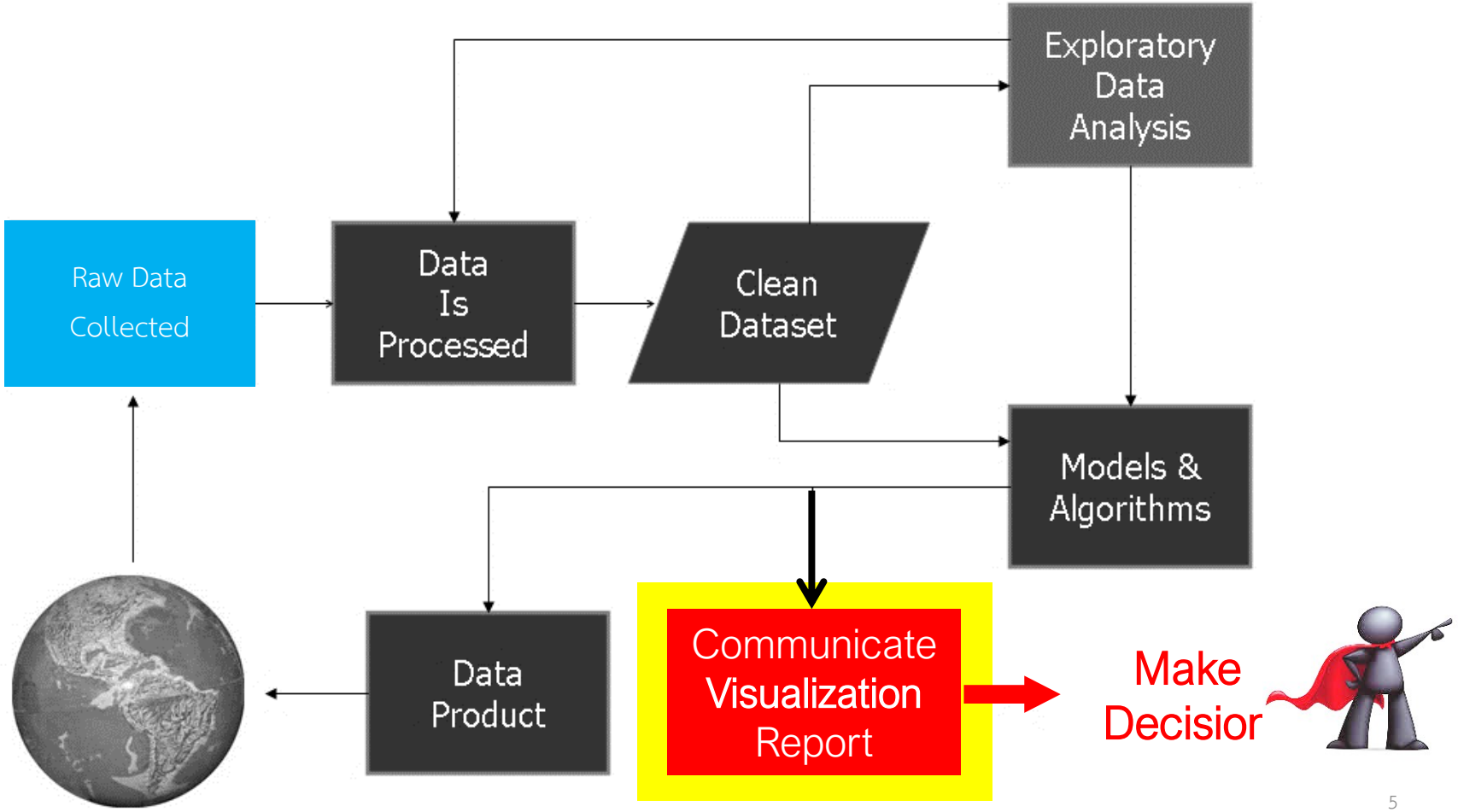
Visualization

"Help Us Consume a lot of Information Quickly"

- Good Data Visualization
  - Helps us think
  - Reduces load on working memory
  - Offloads cognition
  - Uses the power of human perception

Scenario

- Data Science Process

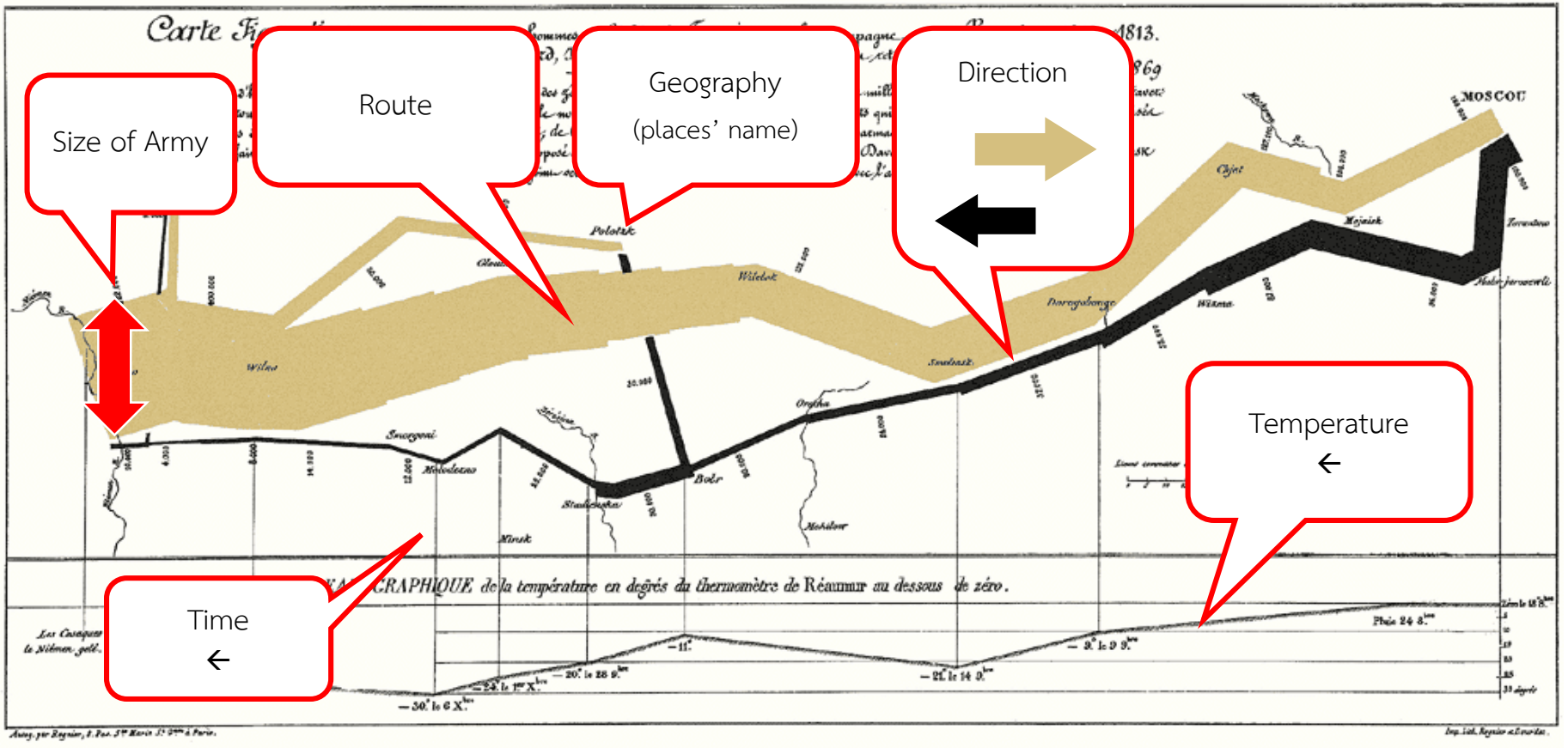


Scenario

- Quick Action

- Score Report *(which better?)*

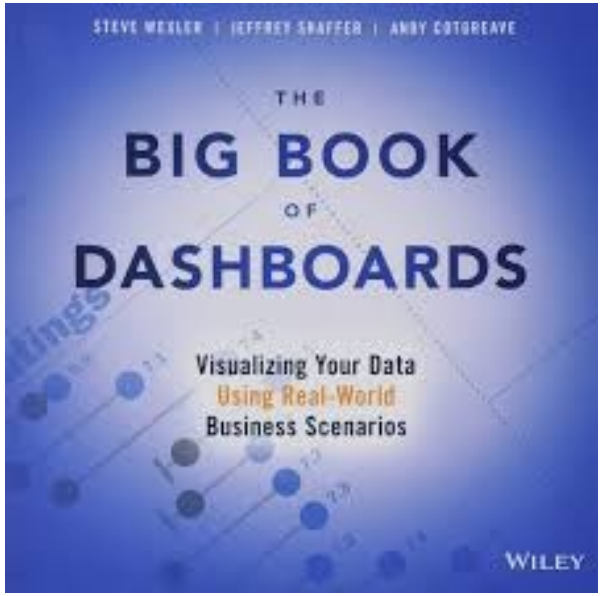
Scenario



Charles Joseph Minard's 1869 diagram of Napoleon's March
(an early example of an information graphic)

Pre-Attentive Attributes

Reference



- **The big book of dashboards: visualizing your data using real-world business scenarios.**
  - Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave.
    John Wiley & Sons, 2017.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Why Do We Visualize Data?

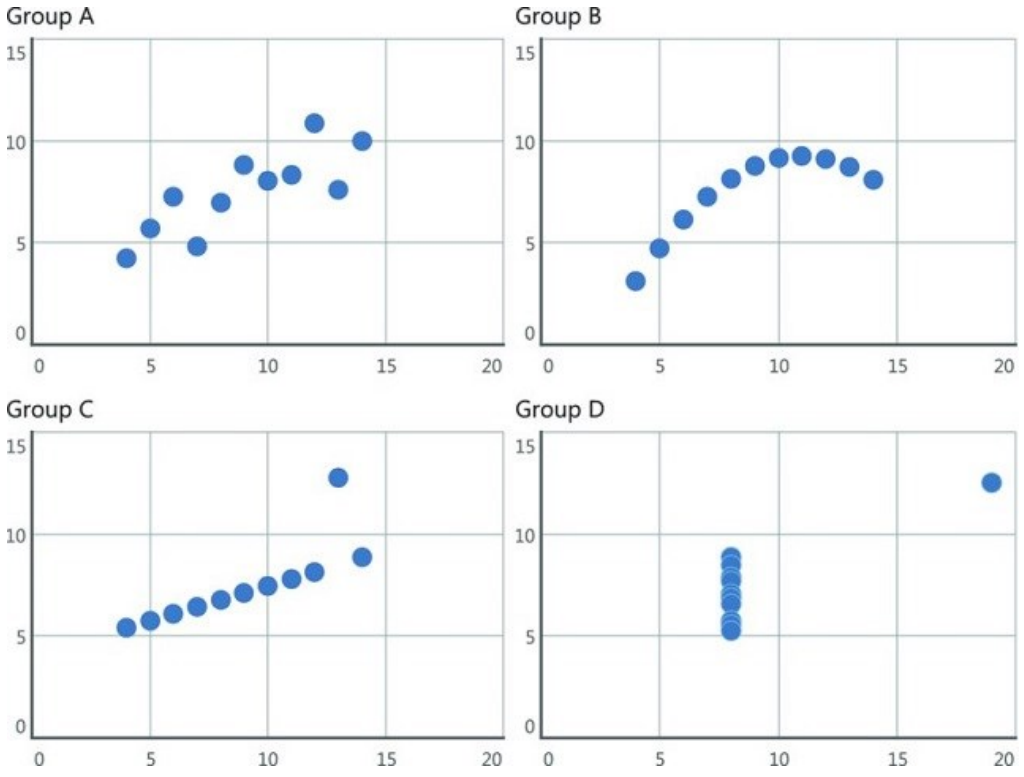
- Let's see why it's vital to visualize numbers by beginning with the table. There are four groups of numbers, each with 11 pairs.
- In a moment, we will create a chart from them, but before we do, take a look at the numbers.
- What can you see?
- Are there any discernible differences in the patterns or trends among them?

Group A		Group B		Group C		Group D	
x	y	x	y	x	y	x	y
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Why Do We Visualize Data?

- Let me guess: You don't really see anything clearly. It's too hard.
- Before we put the numbers in a chart, we might consider their statistical properties. Were we to do that, we'd find that the statistical properties of each group of numbers are very similar. If the table doesn't show anything and statistics don't reveal much, what happens when we plot the numbers? Take a look at



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

What trends can you see?

- Another reason to visualize numbers is to help our memory. Consider Table 1.2, which shows sales numbers for three categories, by quarter, over a four-year period. What trends can you see?

Category	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2	2014 Q3	2014 Q4
Furniture	$463,988	$352,779	$338,169	$317,735	$320,875	$287,934	$319,537	$324,319
Office Supplies	$232,558	$290,055	$265,083	$246,946	$219,514	$202,412	$198,268	$279,679
Technology	$563,866	$244,045	$432,299	$461,616	$285,527	$353,237	$338,360	$420,018
Category	**2015 Q1**	**2015 Q2**	**2015 Q3**	**2015 Q4**	**2016 Q1**	**2016 Q2**	**2016 Q3**	**2016 Q4**
Furniture	$307,028	$273,836	$290,886	$397,912	$337,299	$245,445	$286,972	$313,878
Office Supplies	$207,363	$183,631	$191,405	$217,950	$241,281	$286,548	$217,198	$272,870
Technology	$333,002	$291,116	$356,243	$386,445	$386,387	$397,201	$359,656	$375,229

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

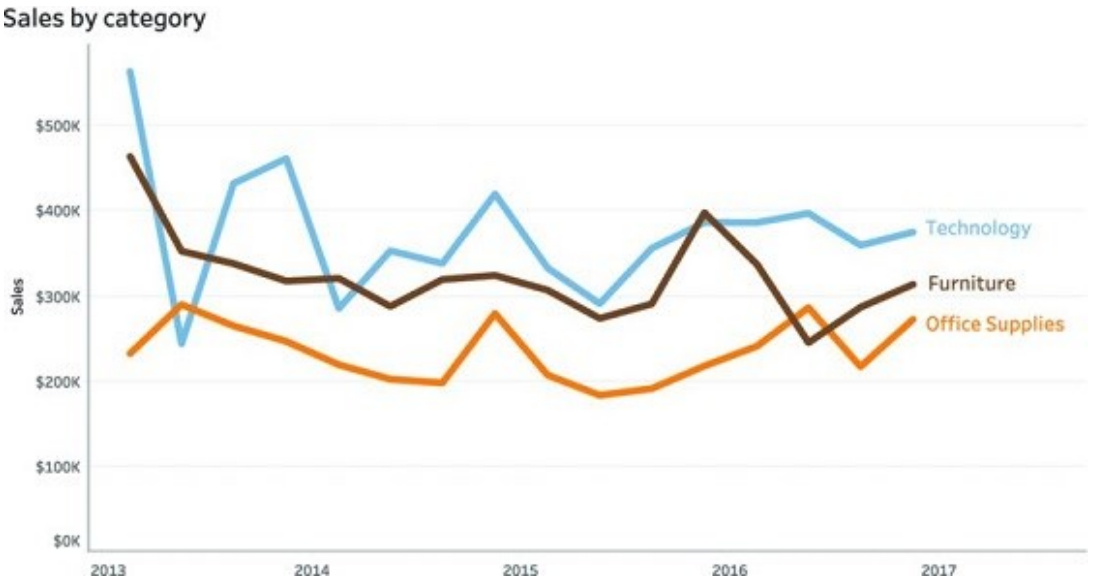
What trends can you see?

- Our short-term memories aren't designed to store many pieces of information.
- By the time we've reached the fourth or fifth number, we will have forgotten the first one we looked at.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Now can you see the trends?

Category	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2	2014 Q3	2014 Q4
Furniture	$463,988	$352,779	$338,169	$317,735	$320,875	$287,934	$319,537	$324,319
Office Supplies	$232,558	$290,055	$265,083	$246,946	$219,514	$202,412	$198,268	$279,679
Technology	$563,866	$244,045	$432,299	$461,616	$285,527	$353,237	$338,360	$420,018
Category	**2015 Q1**	**2015 Q2**	**2015 Q3**	**2015 Q4**	**2016 Q1**	**2016 Q2**	**2016 Q3**	**2016 Q4**
Furniture	$307,028	$273,836	$290,886	$397,912	$337,299	$245,445	$286,972	$313,878
Office Supplies	$207,363	$183,631	$191,405	$217,950	$241,281	$286,548	$217,198	$272,870
Technology	$333,002	$291,116	$356,243	$386,445	$386,387	$397,201	$359,656	$375,229



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Now can you see the trends?

- Now we have much better insight into the trends. Office supplies has been the lowest-selling product category in all but two quarters. Furniture trends have been dropping slowly over the time period, except for a bump in sales in 2015 Q4 and a rise in the last two quarters. Technology sales have mostly been the highest but were particularly volatile at the start of the time period.

- The table and the line chart each visualized the same 48 data points, but only the line chart lets us see the trends. The line chart turned 48 data points into three chunks of data, each containing 16 data points. Visualizing the data hacks our short-term memory; it allows us to interpret large volumes of data instantly.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

How Do We Visualize Data?

- We've just looked at some examples of the power of visualizing data.
- Now we need to move on to how we build the visualizations.
- To do that, we first need to look at two things:
  - pre-attentive attributes
  - types of data.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Pre-Attentive Attributes

- Visualizing data requires us to turn data into marks on a canvas. What kind of marks make the most sense? One answer lies in what are called "pre-attentive attributes." These are things that our brain processes in milliseconds, before we pay attention to everything else. There are many different types. Let's look at an example.

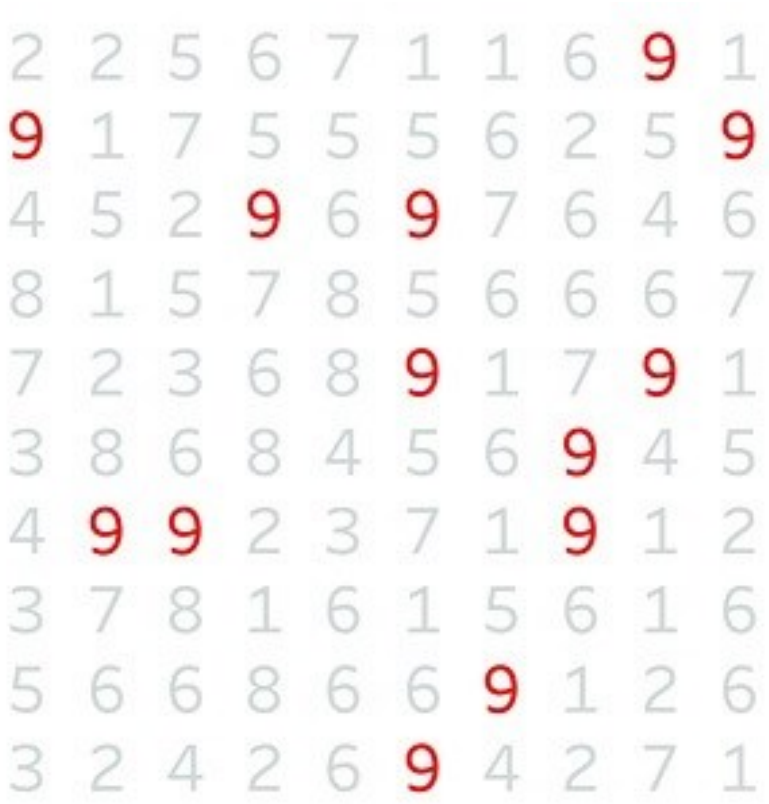
Look at the numbers in the figure.

How many 9s are there?

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Now it's easy to count the 9s.

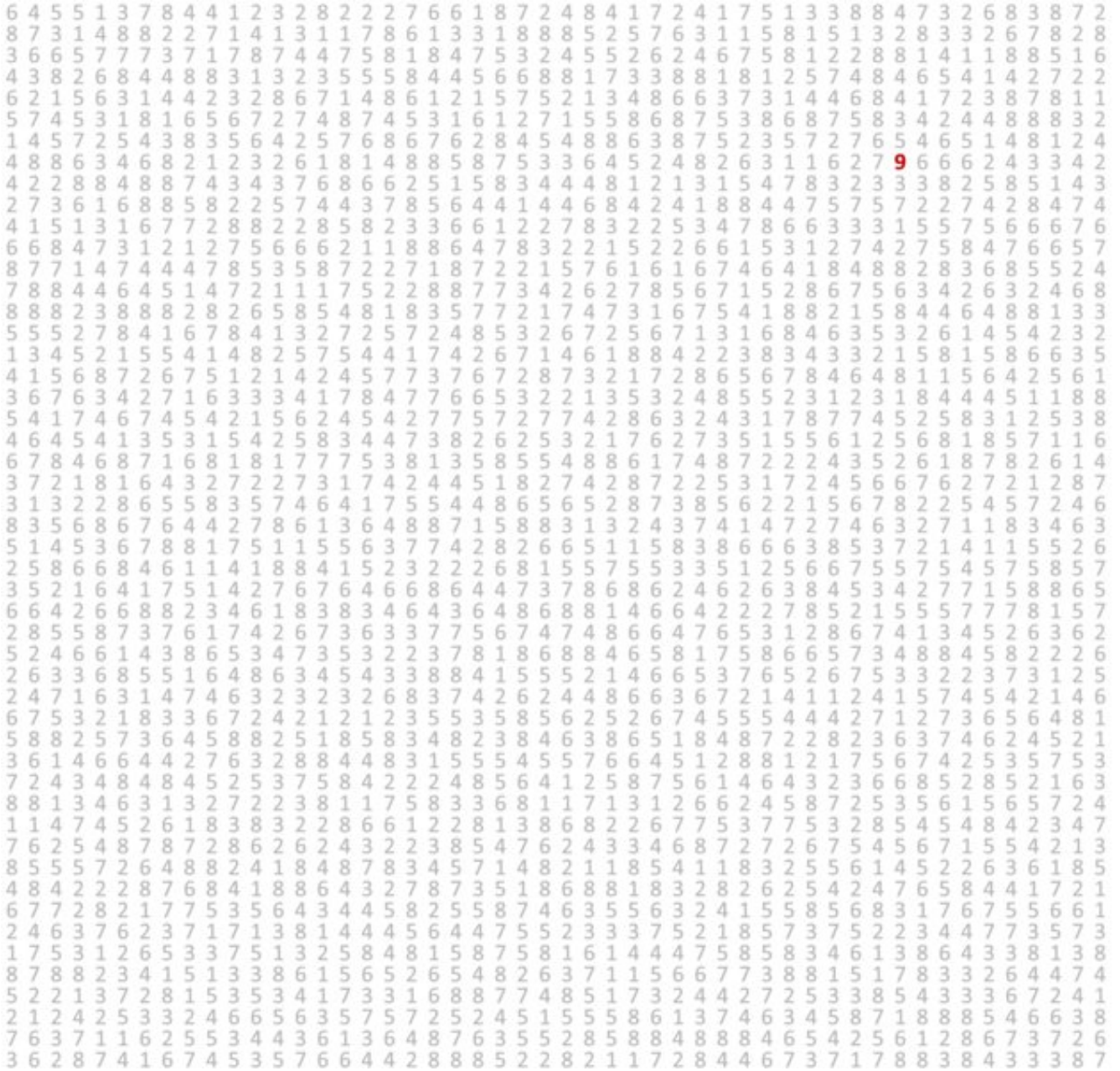
- Now the task is easy. Why? Because we changed the color: 9s are red, and all the other numbers are light gray.
- Color differences pop out. It's as easy to find one red 9 on a table of hundreds of digits as it is on a 10-by-10 grid.
- Your brain registers the red 9s before you consciously addressed the grid to count them.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

9

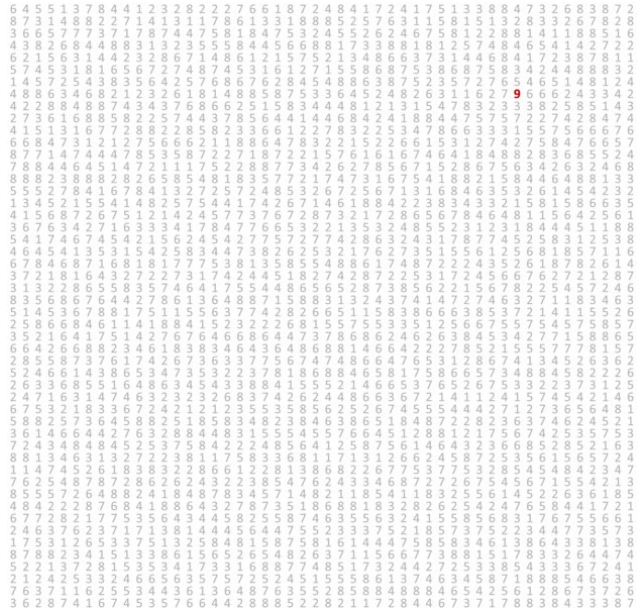
9 in this grid of 2,500 numbers



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

9 in this grid of 2,500 numbers.

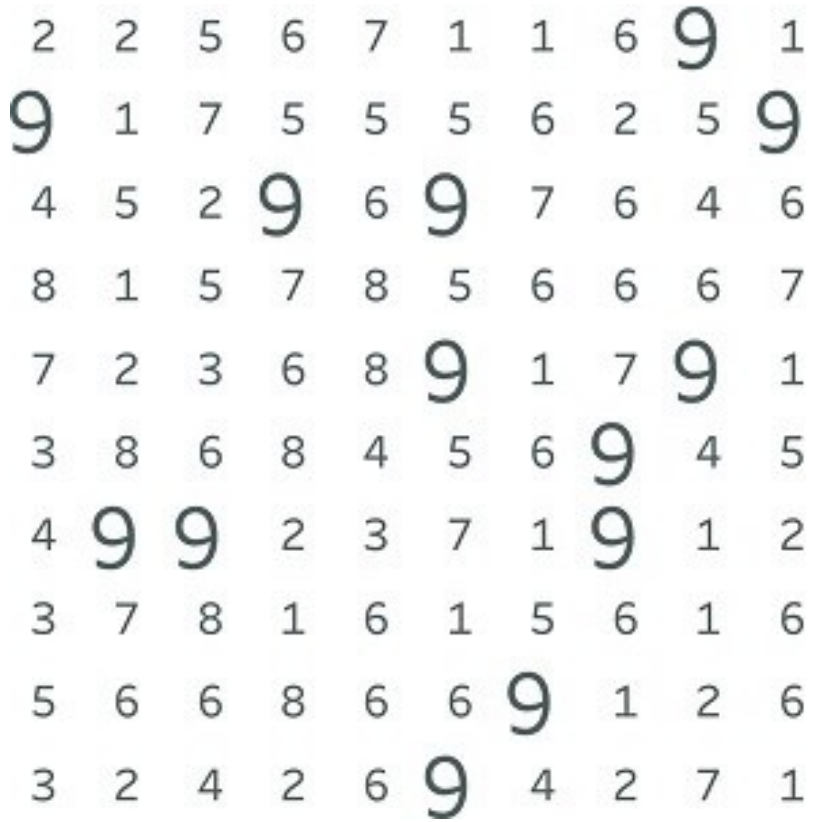
- It's easy to spot the 9. Our eyes are amazing at spotting things like this.
- Color (in this case, hue) is one of several pre-attentive attributes. When we look at a scene in front of us, or a chart, we process these attributes in under 250 milliseconds.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Differences in Size

- Differences in size are easy to see too



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

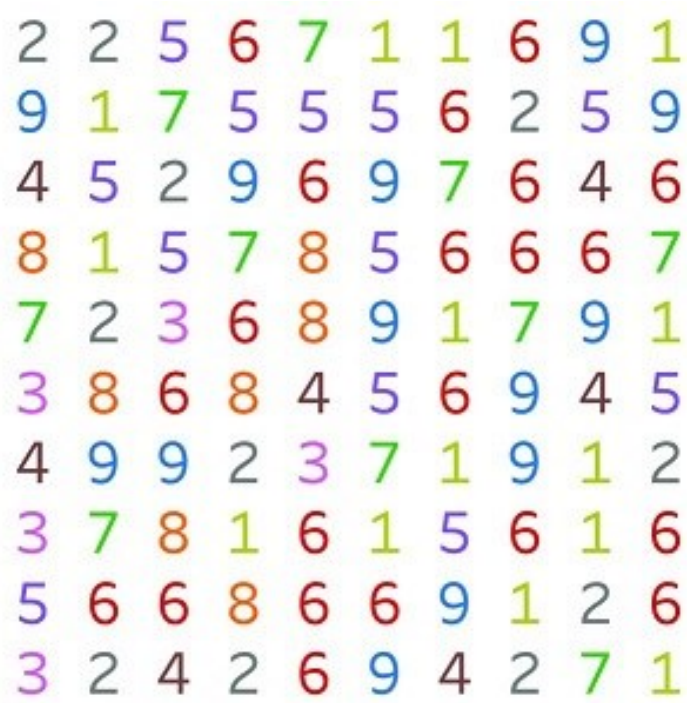
Size and Hue

- Size and Hue:
- Aren't they amazing?
- That's all very well when counting the 9s.
- What if our task is to count the frequency of each digit?
- That's a slightly more realistic task, but we can't just use a different color or size for each digit.
  That would defeat the **pre-attentive nature of the single color**.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Coloring every Digit

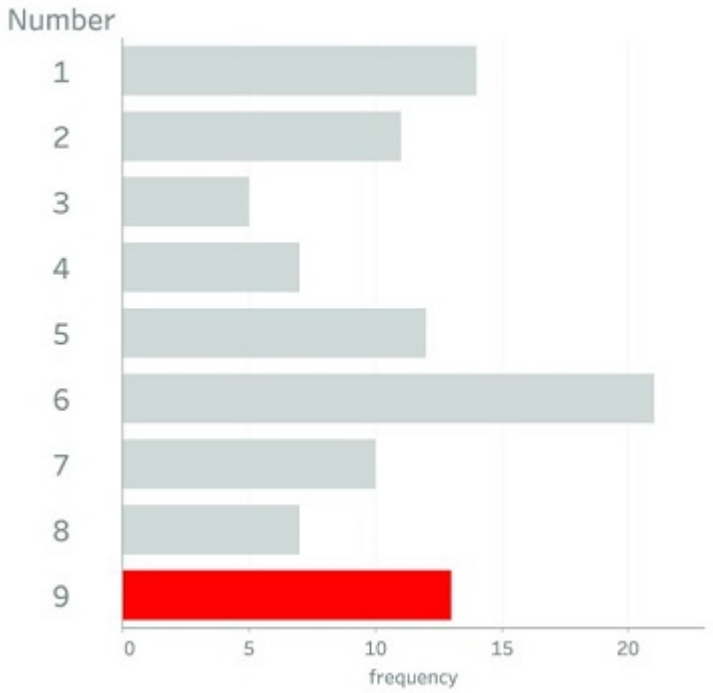
- Coloring every digit is nearly as bad as having no color.
- It's not a complete disaster: If you're looking for the 6s, you just need to work out that they are red and then scan quickly for those. Using one color on a visualization is highly effective to make one category stand out. Using a few colors, as we did in the figure to distinguish a small number of categories, is fine too. Once you're up to around eight to ten categories, however, there are too many colors to easily distinguish one from another.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

To count each digit

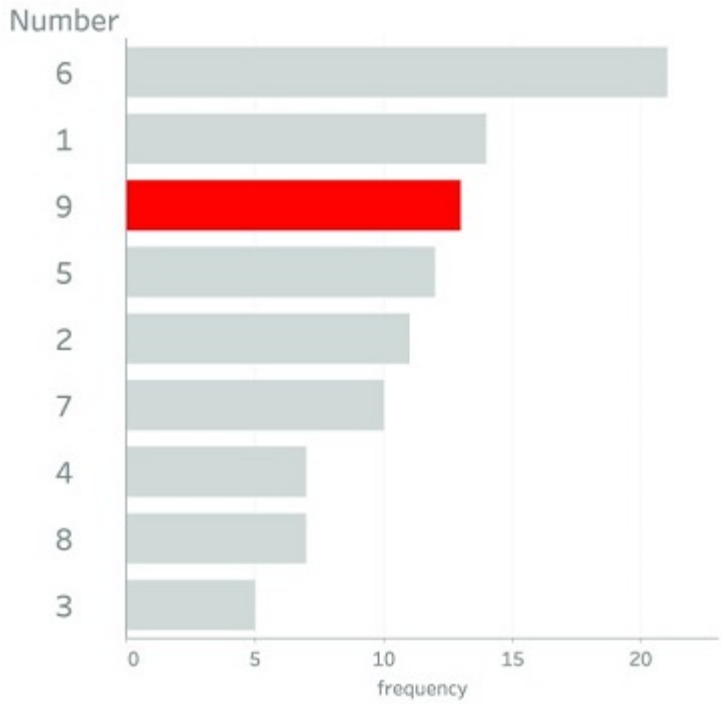
- To count each digit, we need to aggregate.
- Visualization is, at its core, about encoding aggregations, such as frequency, in order to gain insight. We need to move away from the table entirely and encode the frequency of each digit. The most effective way is to **use length**, which we can do in a **bar chart**. The figure shows the frequency of each digit. We've also colored the bar showing the number 9.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

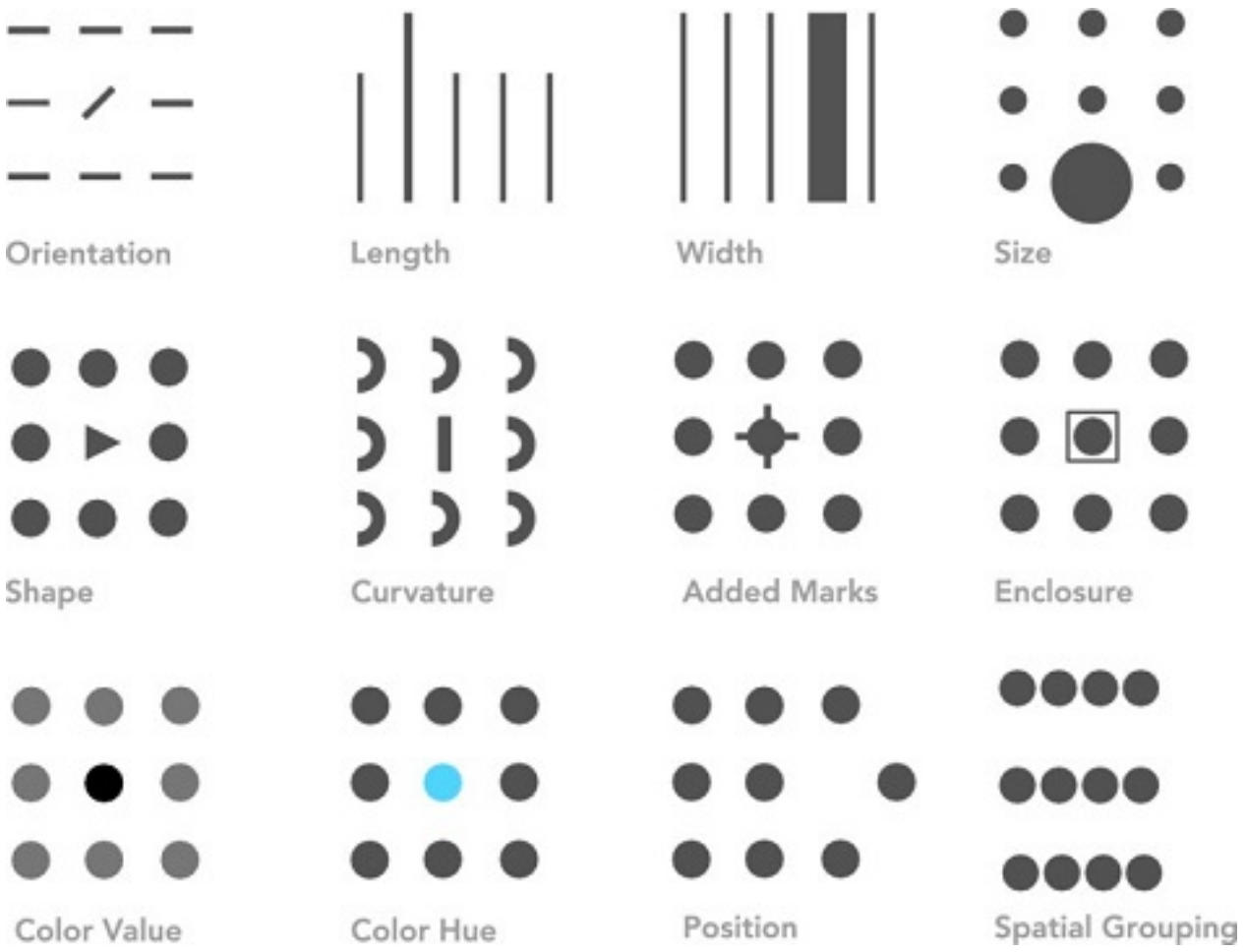
There are 13 9s

- Sorted bar chart using color and length to show how many 9s are in our table.
- Since the task is to count the 9s in the data source, the bar chart is one of the best ways to see the results.
- This is because length and position are best for quantitative comparisons.
- If we extend the example one final time and consider which numbers are most common, we could sort the bars, as shown in the figure



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

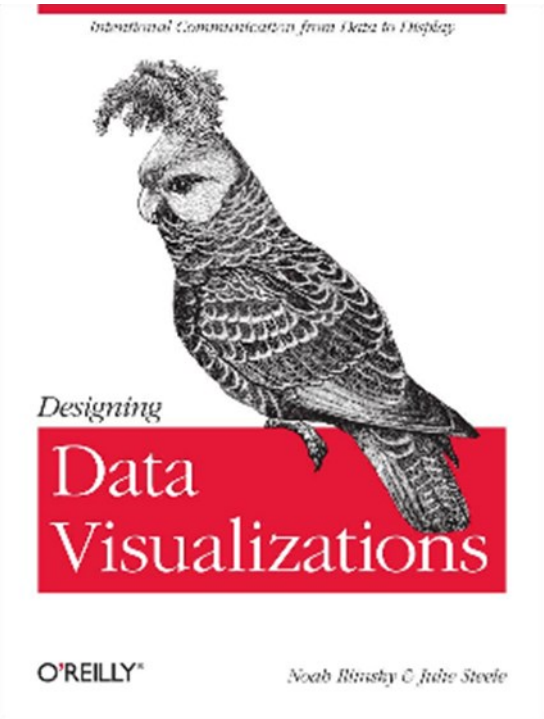
Pre-Attentive Features



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Visual Encodings

Reading



- **Designing Data Visualizations: Representing Informational Relationships**
  - Julie Steele and Noah Iliinsky
  - Chapter 4 Choose Appropriate Visual Encodings

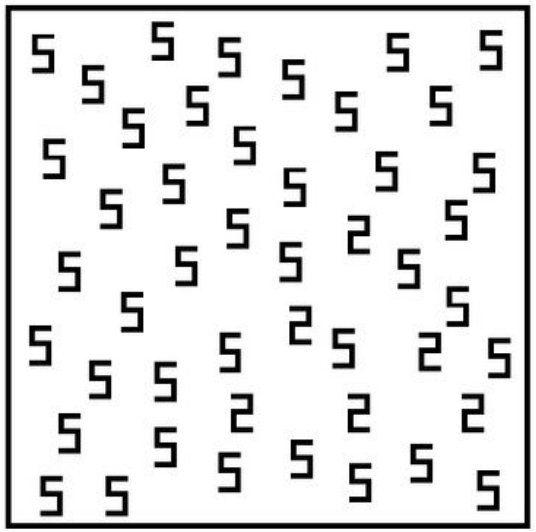
Ref: • Julie Steele and Noah Iliinsky: Designing Data Visualizations: Representing Informational Relationships.

Types of Data

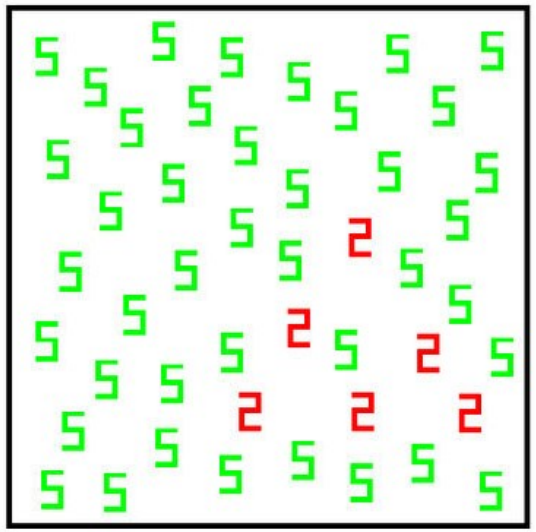
- **Quantitative Data**
  - Information that can be measured and written down with numbers. *(27, 33, 41)*
  - e.g. number of things, temperature, height, age, time, …
- **Ordinal Data**
  - Data type consisting of numerical scores that exist on an ordinal scale. *(small, medium, large, enormous)*
  - e.g. ranking, rating, ordering, …
- **Categorical Data**
  - Types of data which may be divided into groups. *(A, B, C)*
  - e.g. gender, classification, educational level, …

Visual Perception

- A human can distinguish differences size, position, and color (hue) readily without significant processing effort
- How many 2 ?

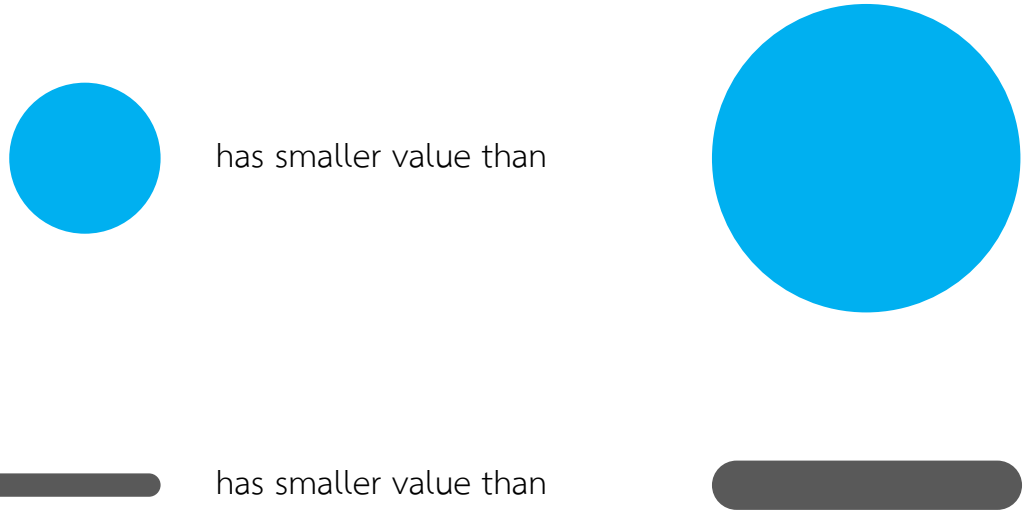


vs



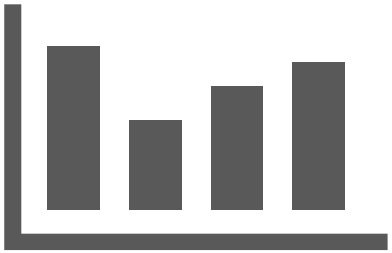
Size

- Size can refer to length, width, height, depth, area, and thickness.
- Size has a natural ordering.
- For example,

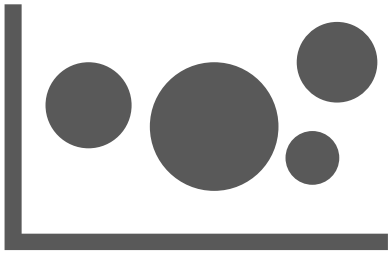


Size

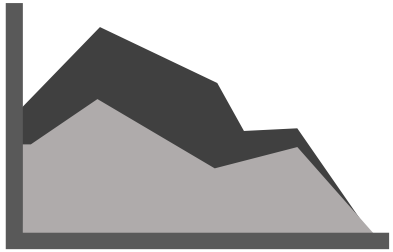
- Diagrams that use Size to encode value



Bar chart



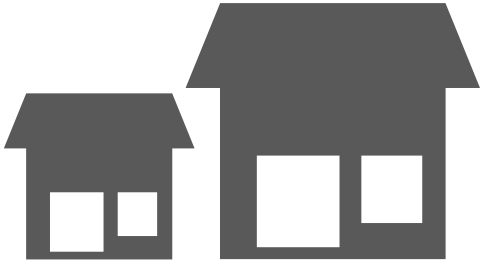
Bubble chart



Area chart

Pie chart

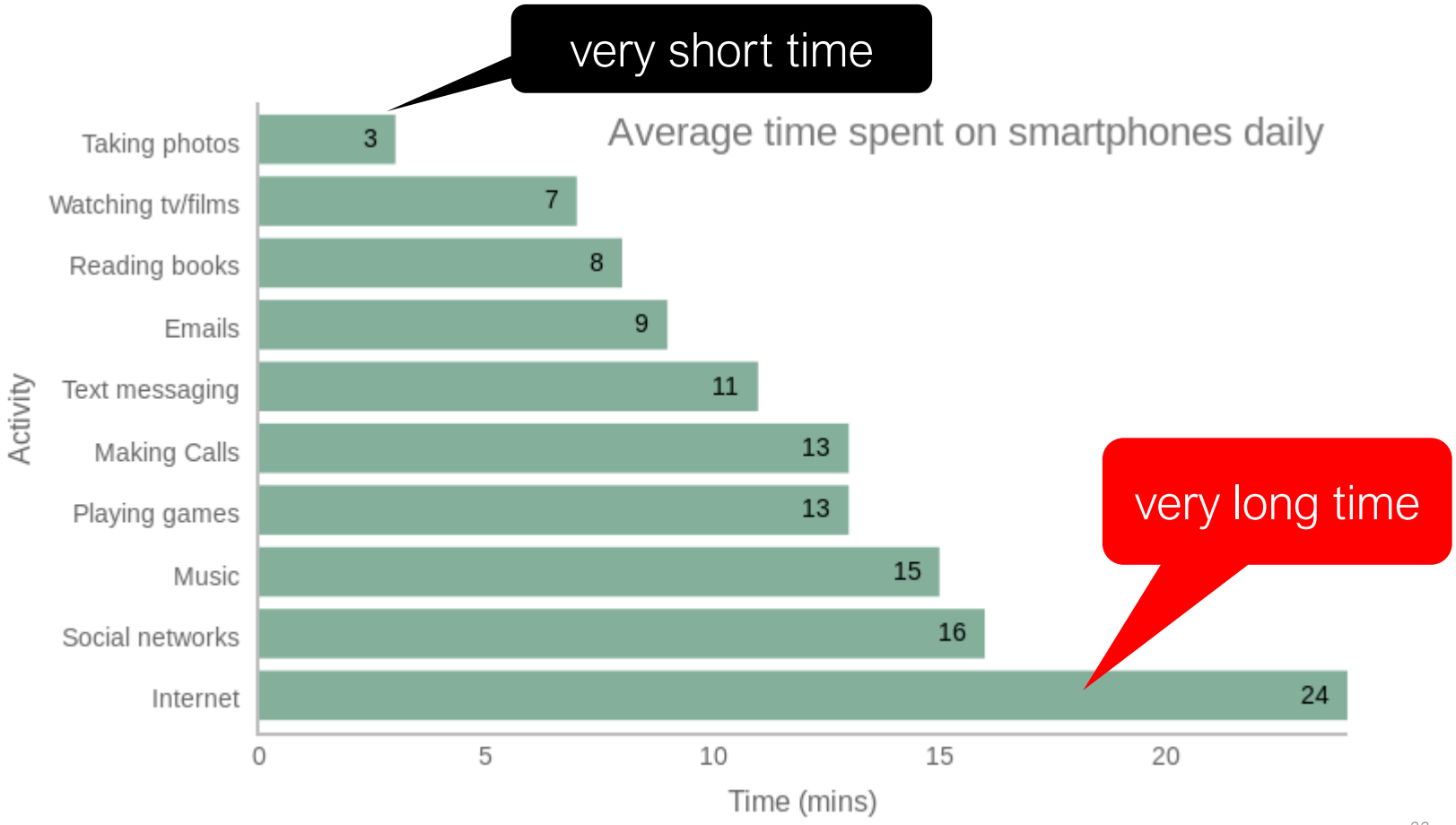
Text cloud



Any graphics

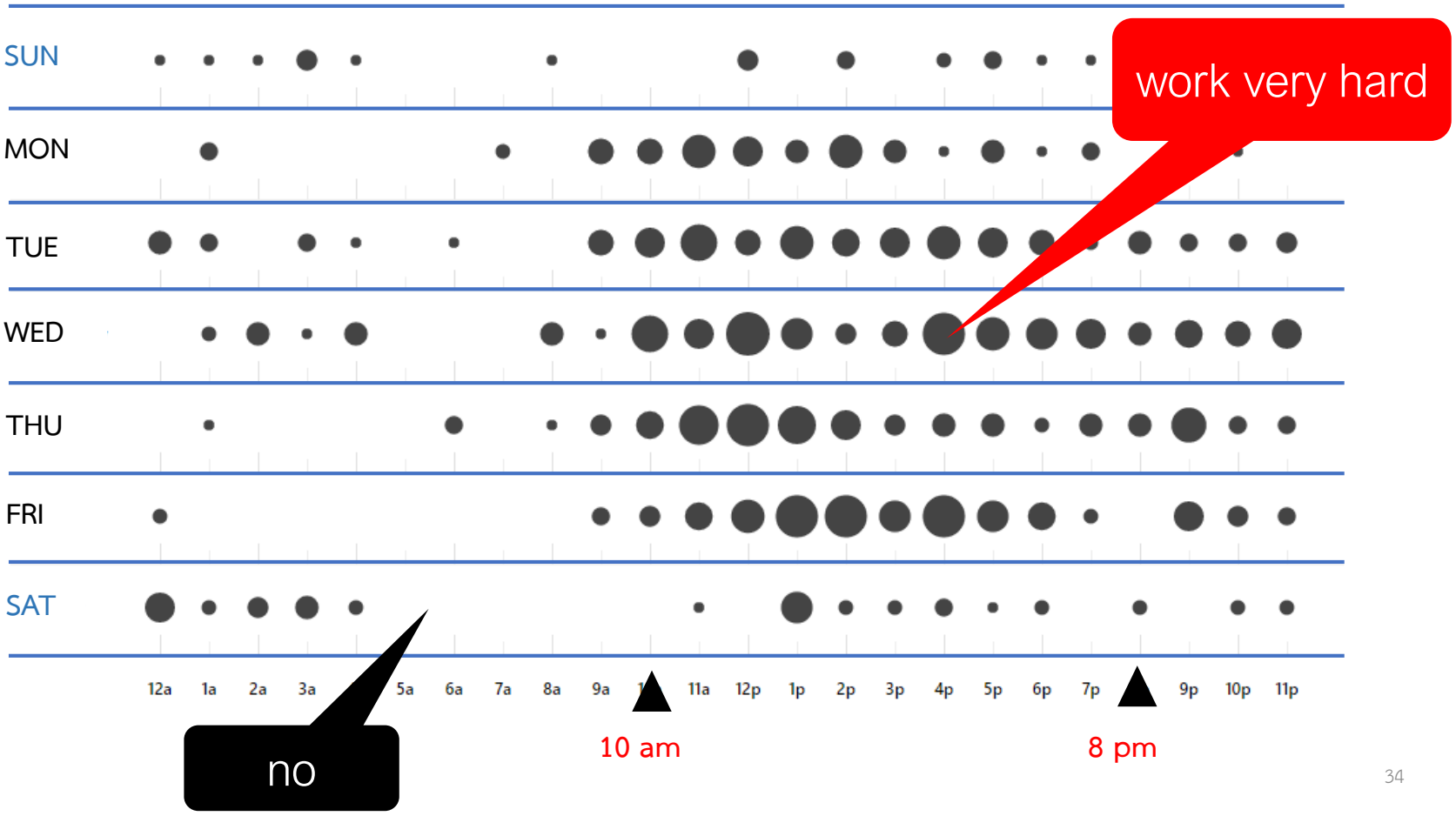
Size

- Example: average time spent on smartphones daily



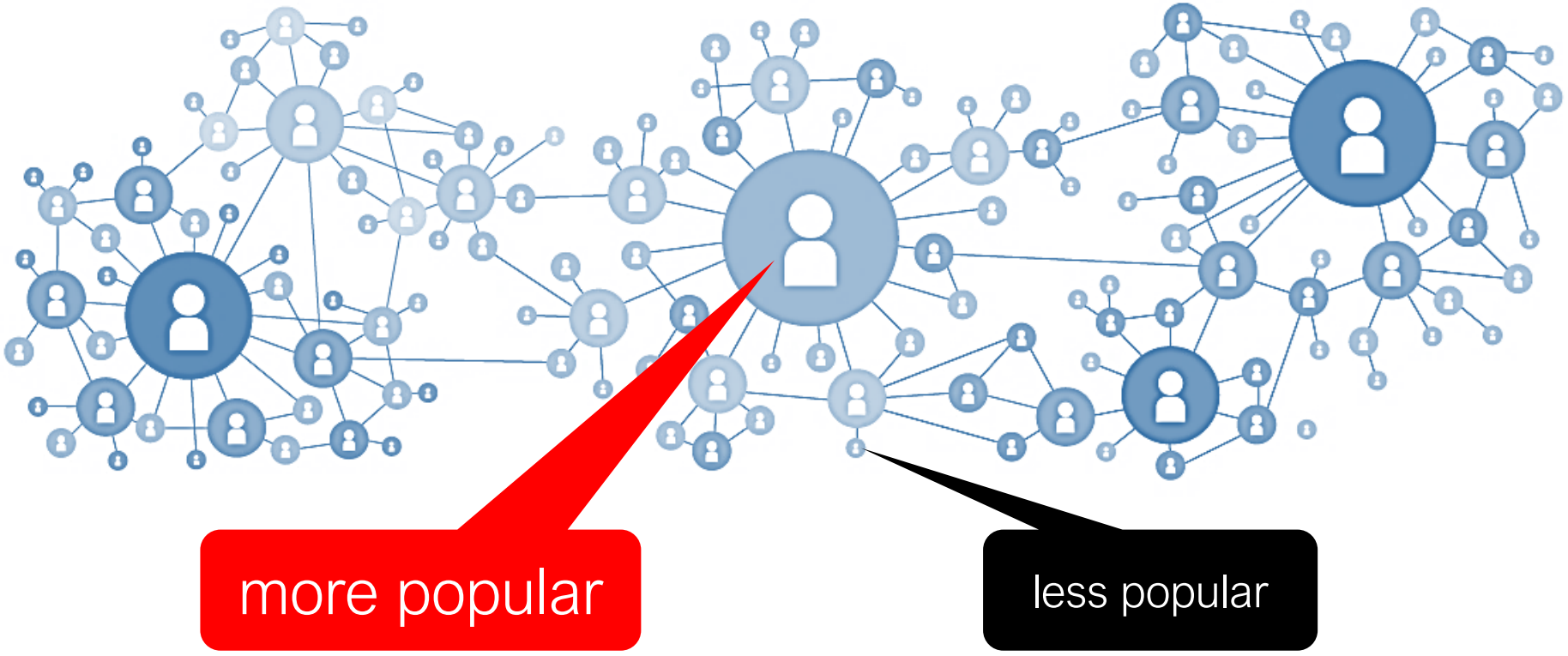
Size

- Example: how often our team commit source code



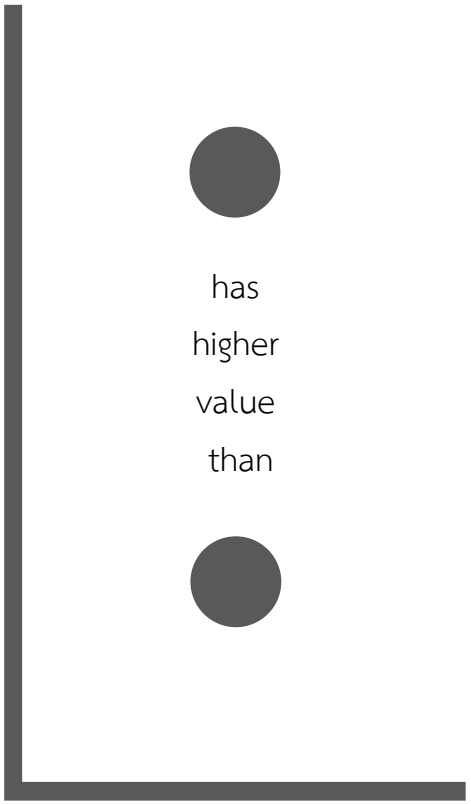
Size

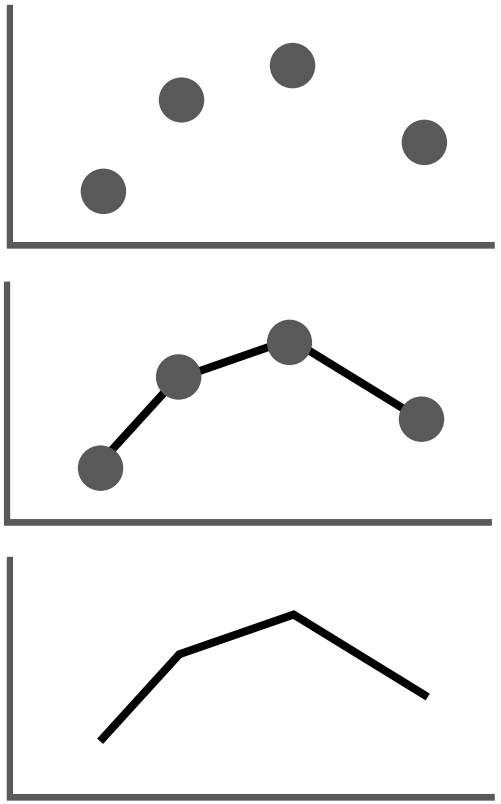
- Example: Social Network



Position

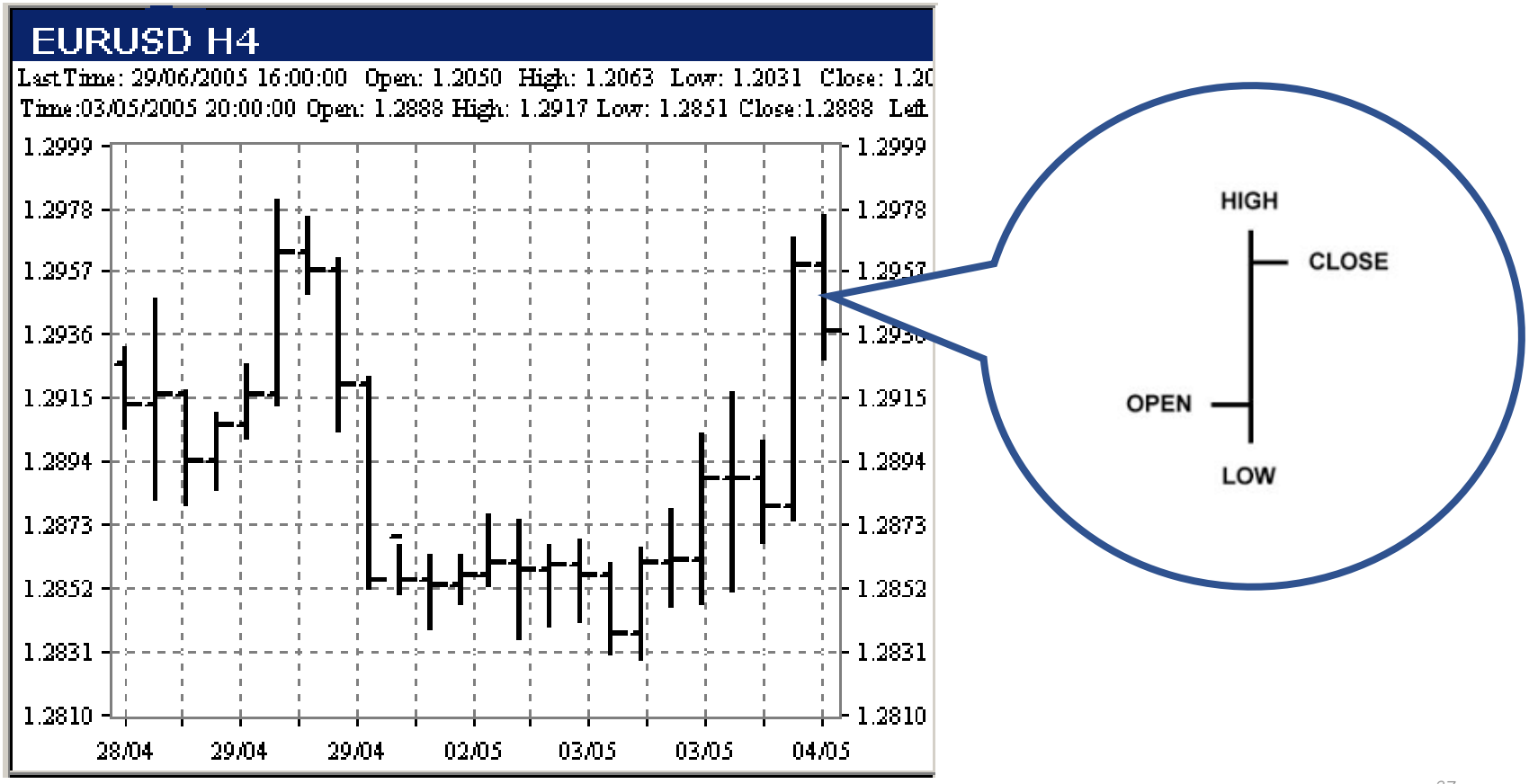
- Position has a natural ordering.





Position

- Example: Stock bar chart



Position

- Example: Geographical map

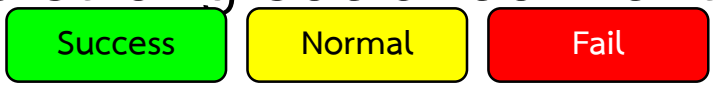
Note: Geographical Map is so powerful when working with other visual encodings.

Color

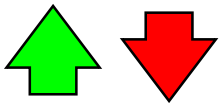
- Color (hue) is not naturally ordered in our brains.

- Brightness is naturally ordered.

- We have strong social conventions about color.

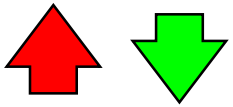


Stock exchange

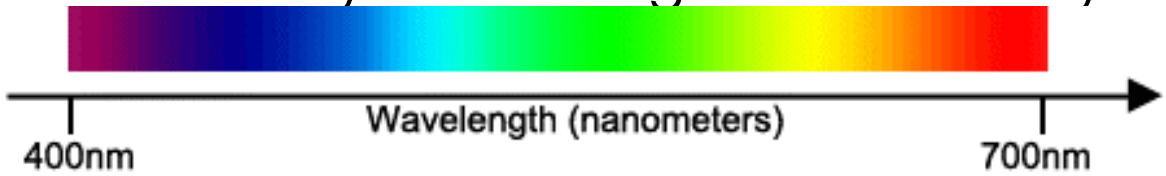


  - But not all people agree.

Stock exchange in  shows

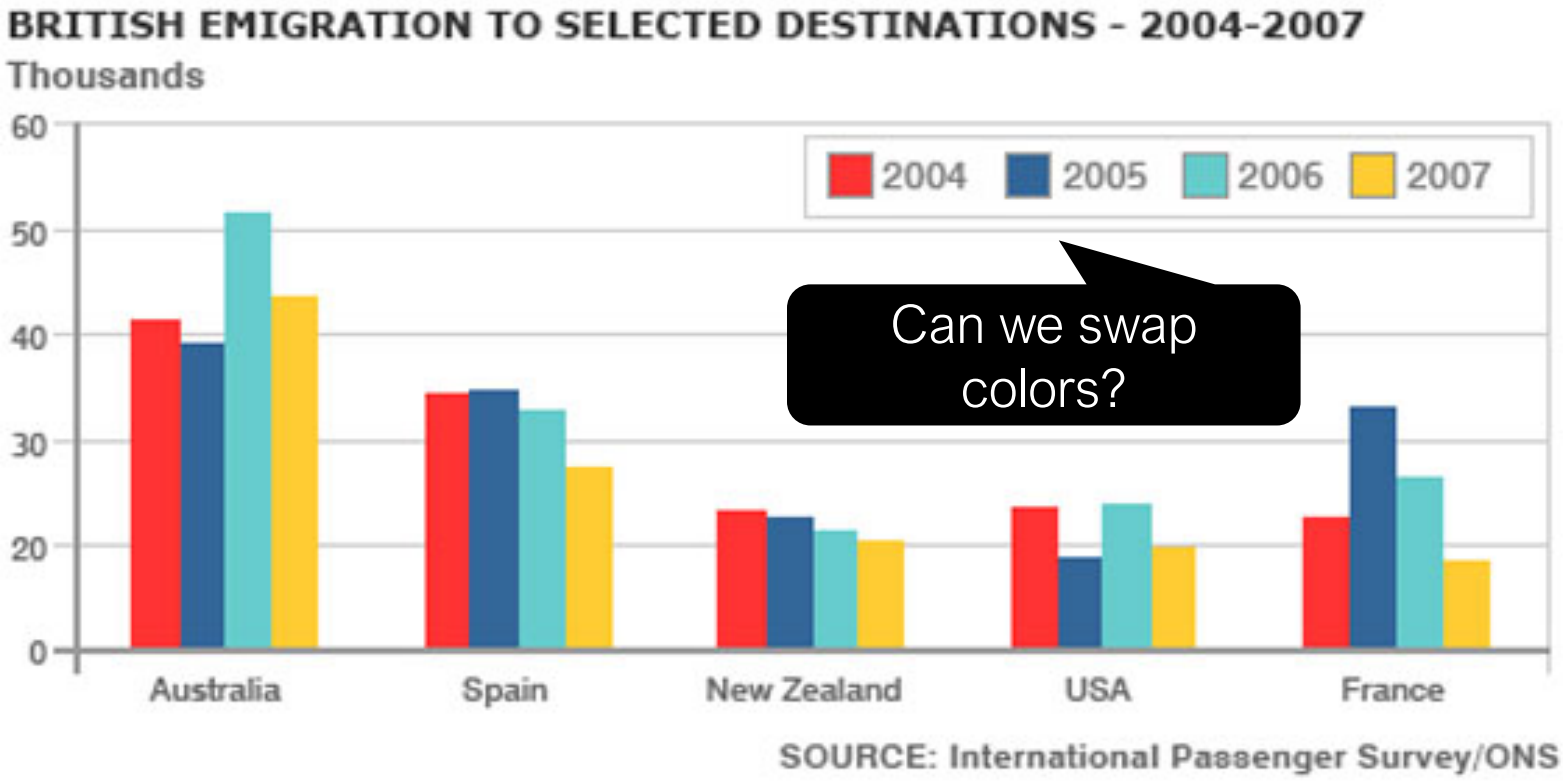


- Colors ordered by wavelength is not easy to read.



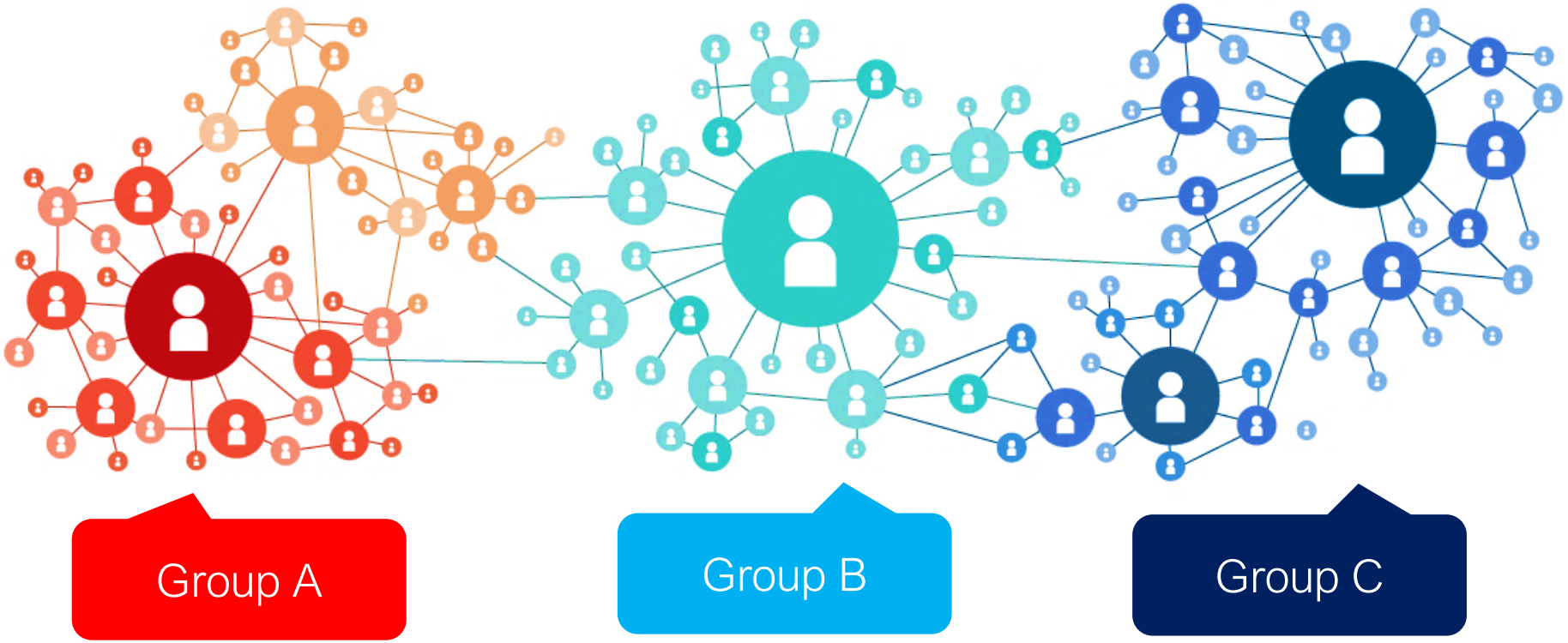
Color

- Example: bar chart of multiple data series



Color

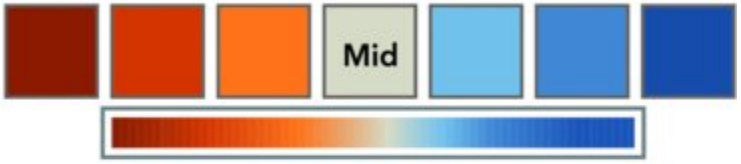
- Example: network clustering



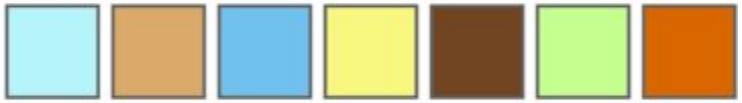
Use of Color in Data Visualization

SEQUENTIAL
color is ordered from low to high

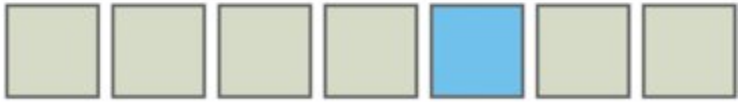
DIVERGING
two sequential colors with a neutral midpoint



CATEGORICAL
contrasting colors for individual comparison



HIGHLIGHT
color used to highlight something

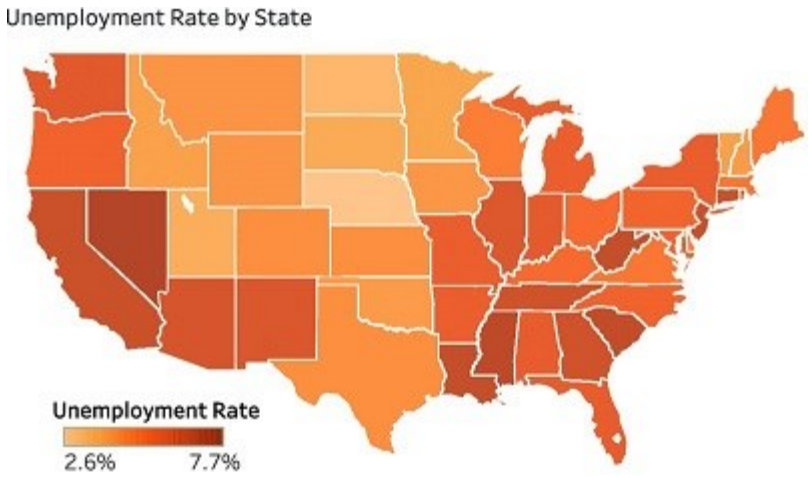


ALERT
color used to alert or warn reader

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Sequential color

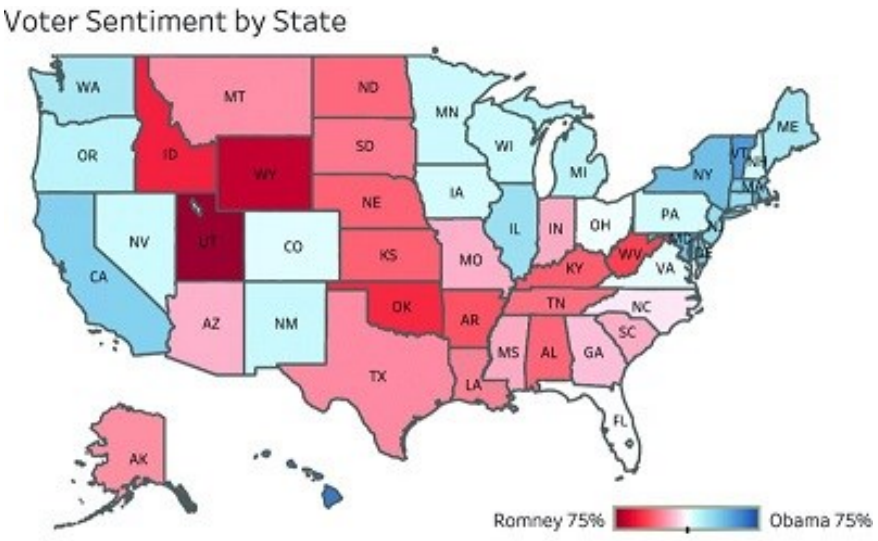
- Sequential color is the use of a single color from light to dark.
- An example is encoding the total amount of sales by state in blue, where the darker blue shows higher sales and a lighter blue shows lower sales. The figure shows the unemployment rate by state using a sequential color scheme.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Diverging color

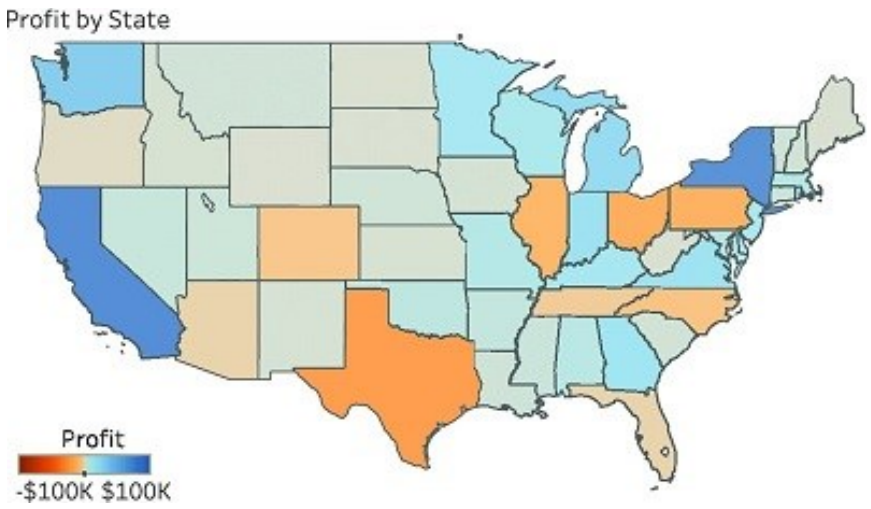
- Diverging color is used to show a range diverging from a midpoint.
- This color can be used in the same manner as the sequential color scheme but can encode two different ranges of a measure (positive and negative) or a range of a measure between two categories. An example is the degree to which electorates may vote Democratic or Republican in each state



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Diverging color

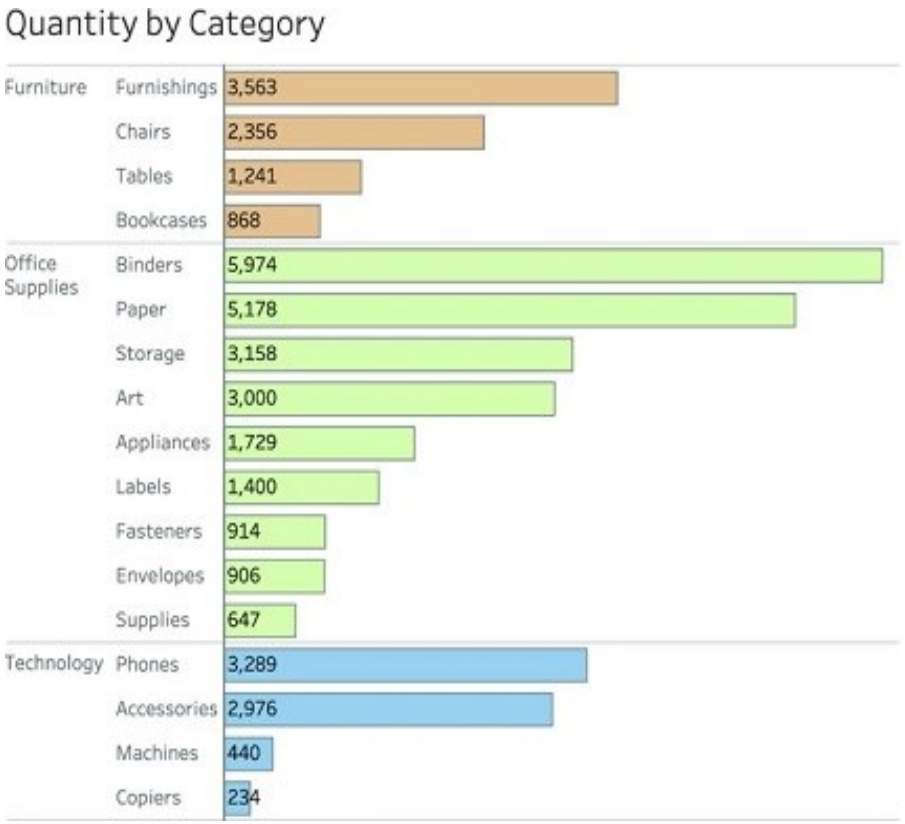
- Diverging color can also be used to show the weather, with blue showing the cooler temperatures and red showing the hotter temperatures.
- The midpoint can be the average, the target, or zero in cases where there are positive and negative numbers. The figure shows an example with profit by state, where profit (positive number) is shown in blue and loss (negative number) is shown in orange



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Categorical color

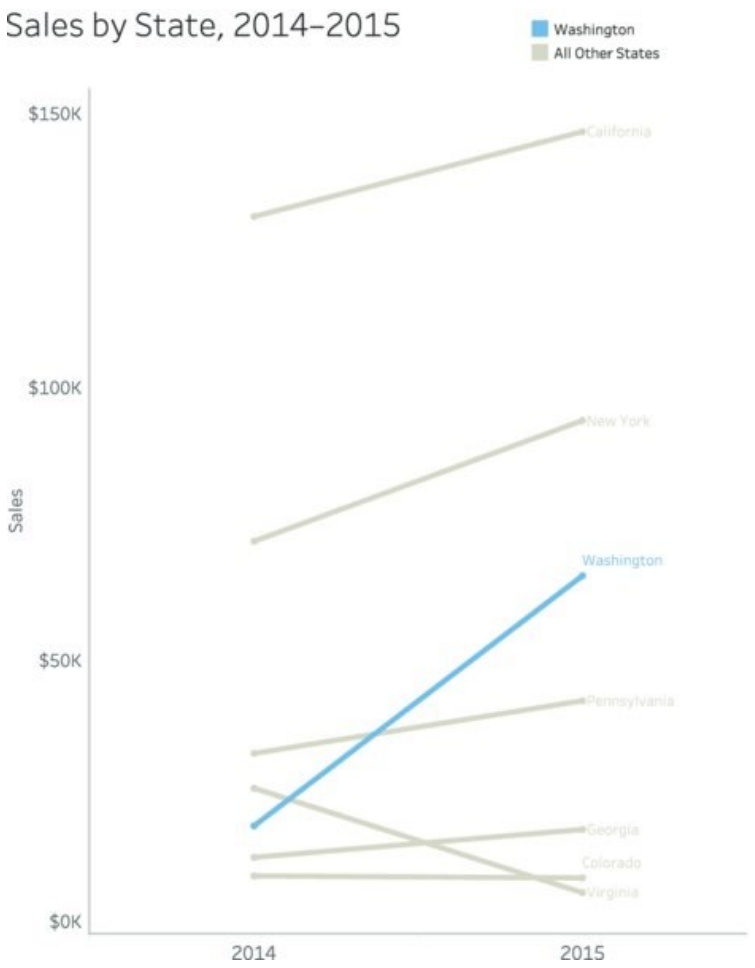
- Categorical color uses different color hues to distinguish between different categories.
- For example, we can establish categories involving apparel (e.g., shoes, socks, shirts, hats, and coats) or vehicle types (e.g., cars, minivans, sport utility vehicles, and motorcycles). Figure 1.20 shows quantity of office supplies in three categories.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Highlight color

- Highlight color is used when there is something that needs to stand out to the reader, but not alert or alarm them.
- Highlights can be used in a number of ways, as in highlighting a certain data point, text in a table, a certain line on a line chart, or a specific bar in a bar chart. The figure shows a slopegraph with a single state highlighted in blue.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

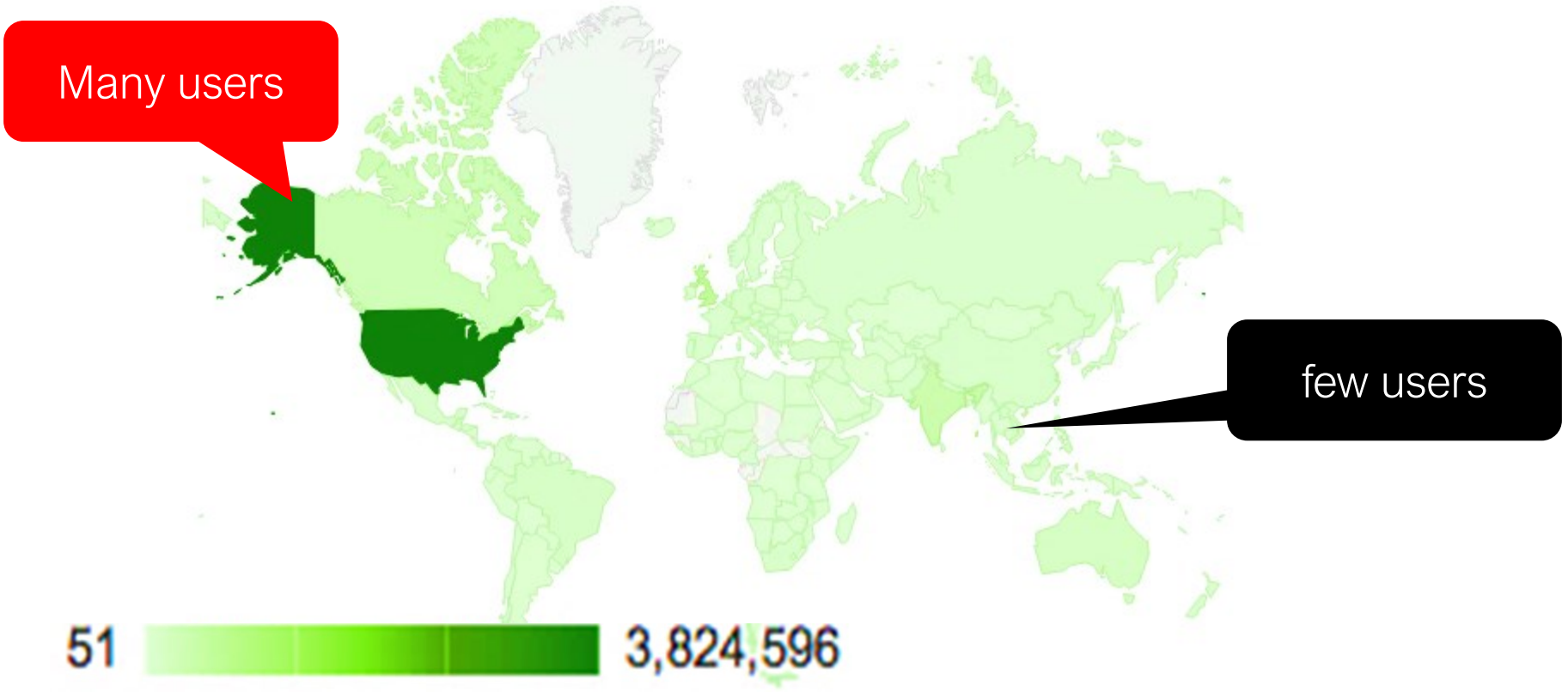
Highlight color

- Alerting color is used when there is a need to draw attention to something for the reader. In this case, it's often best to use bright, alarming colors, which will quickly draw the reader's attention, as in the figure.
- Red and orange indicators to alert the reader that something on the dashboard needs attention.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

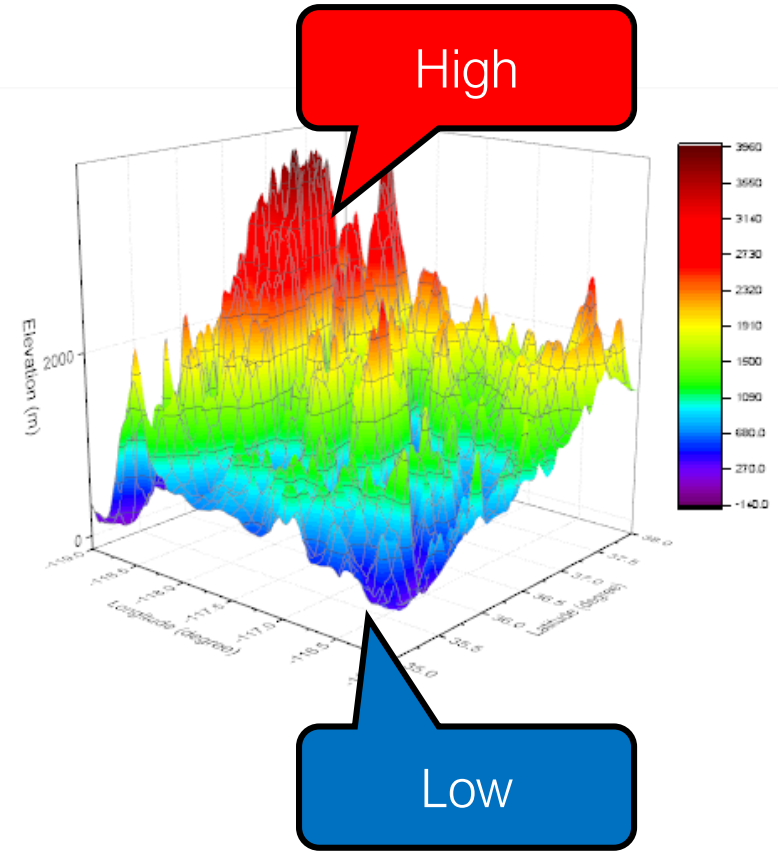
Color

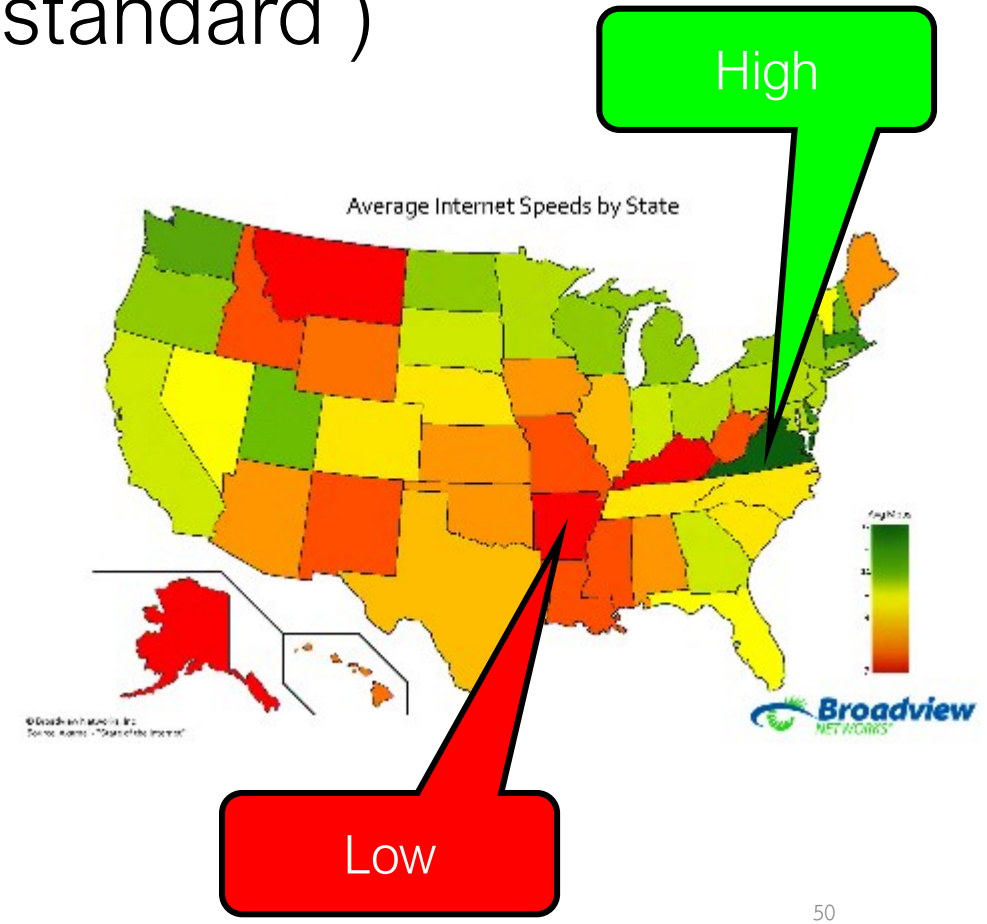
- Example: number of webpage's visitors from different countries



Color

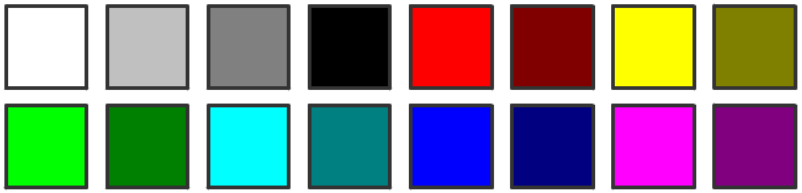
- Example: heat map
  (can be used, but no standard)





Color

~~4-bit (16 colors) Palette~~

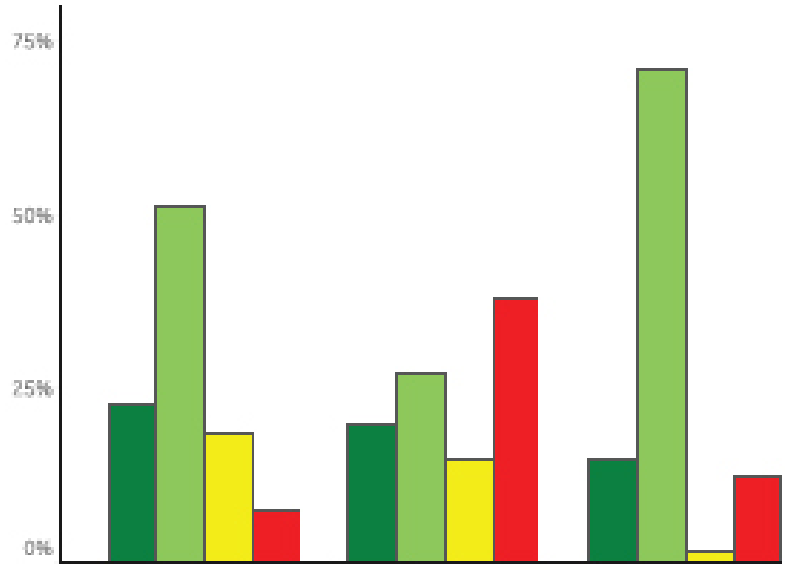


- Hint

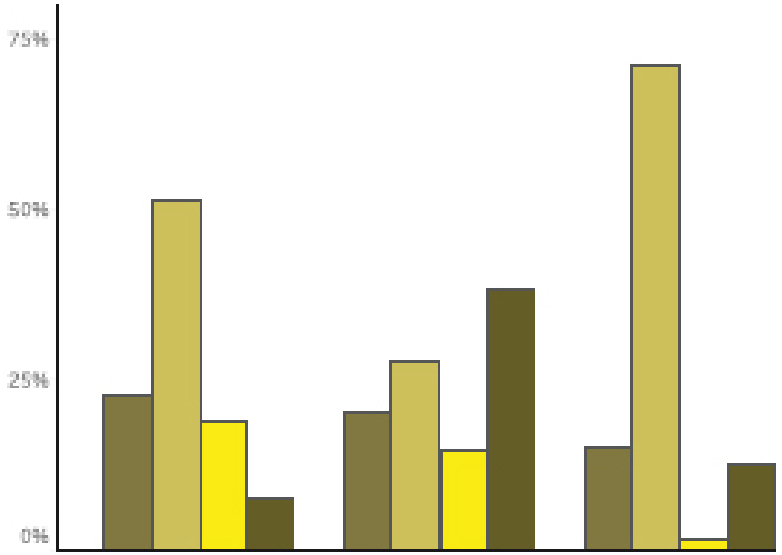
Color Blindness

- Seeing the Problem for Yourself

Traffic Light Colors



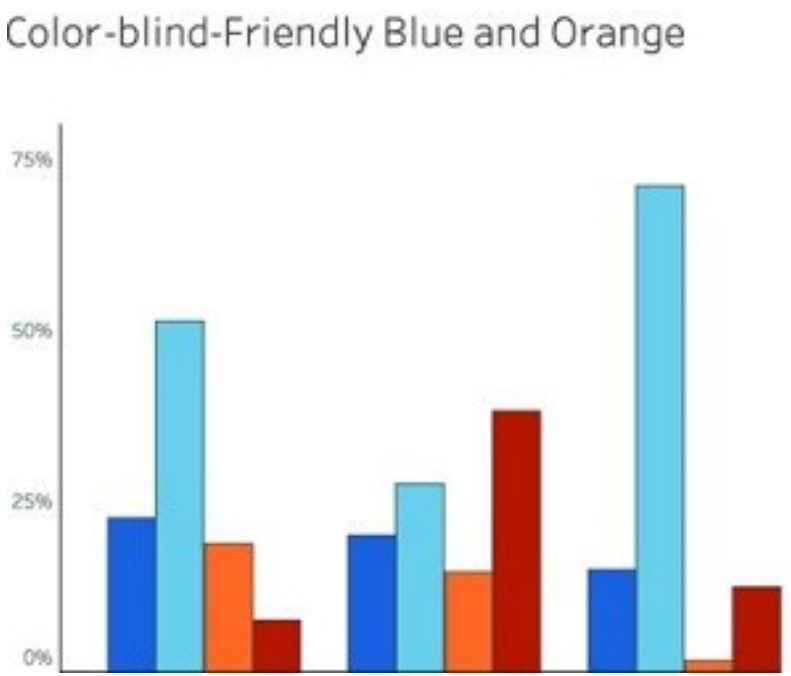
Protanopia Simulation

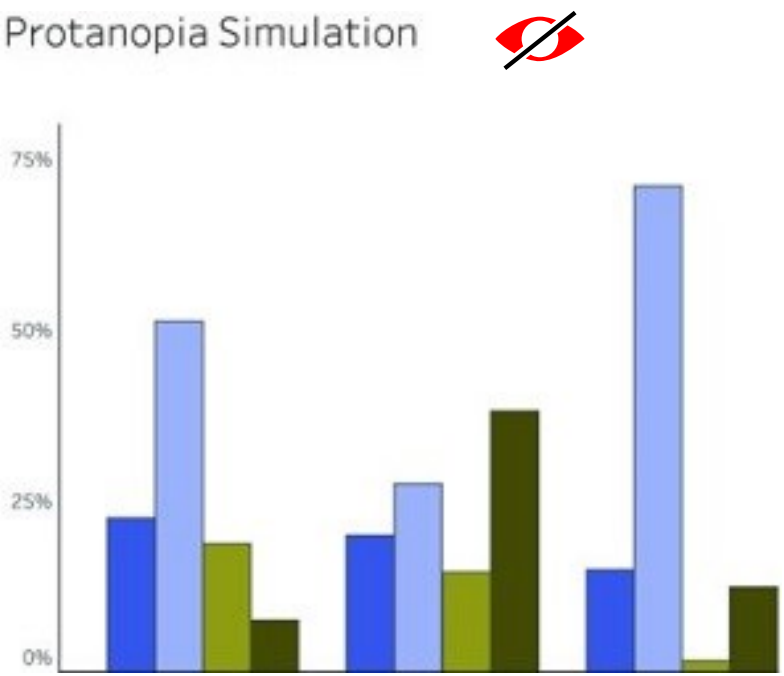


Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Color Blindness

- Seeing the Problem for Yourself

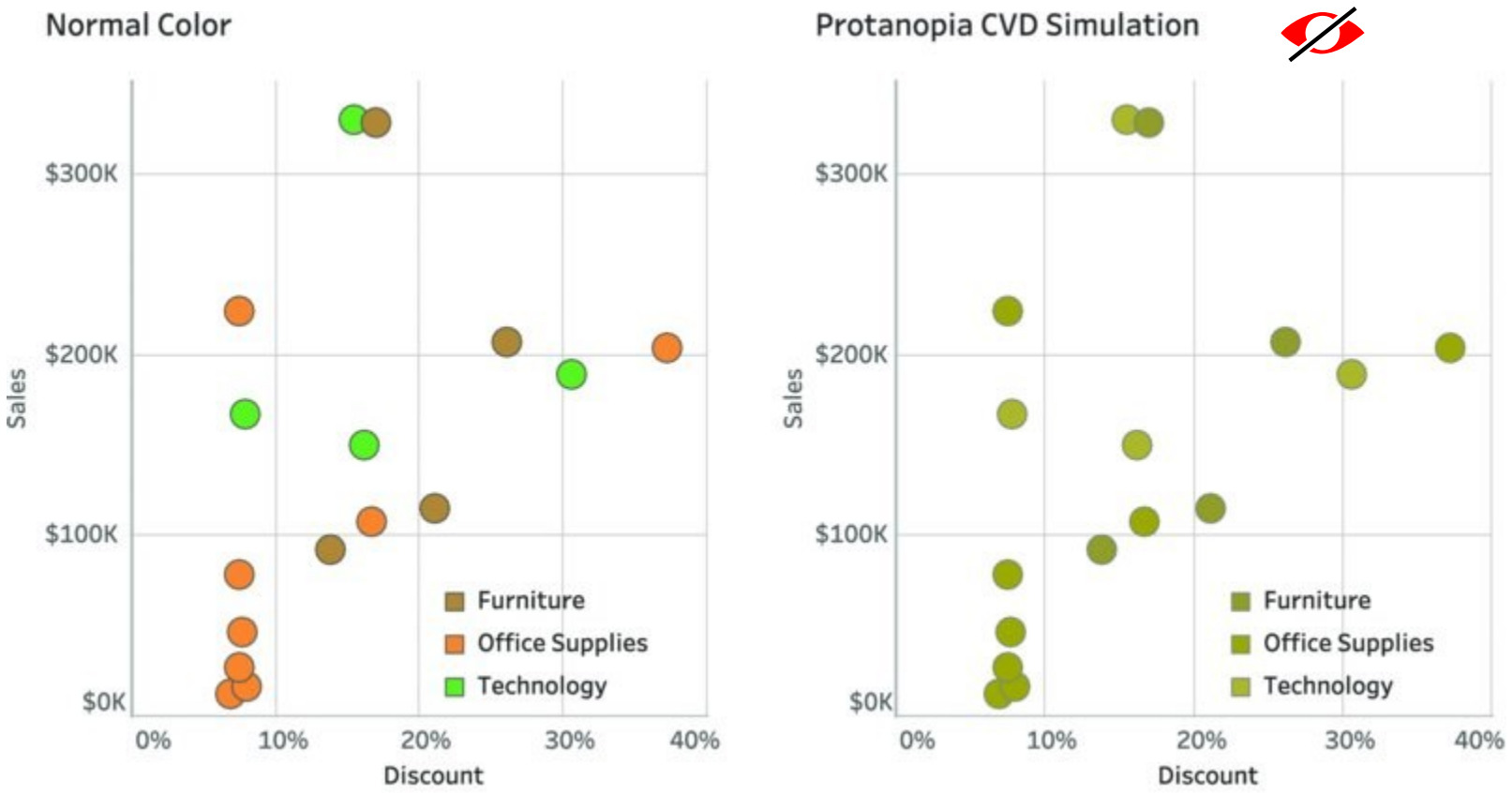




Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Color Blindness

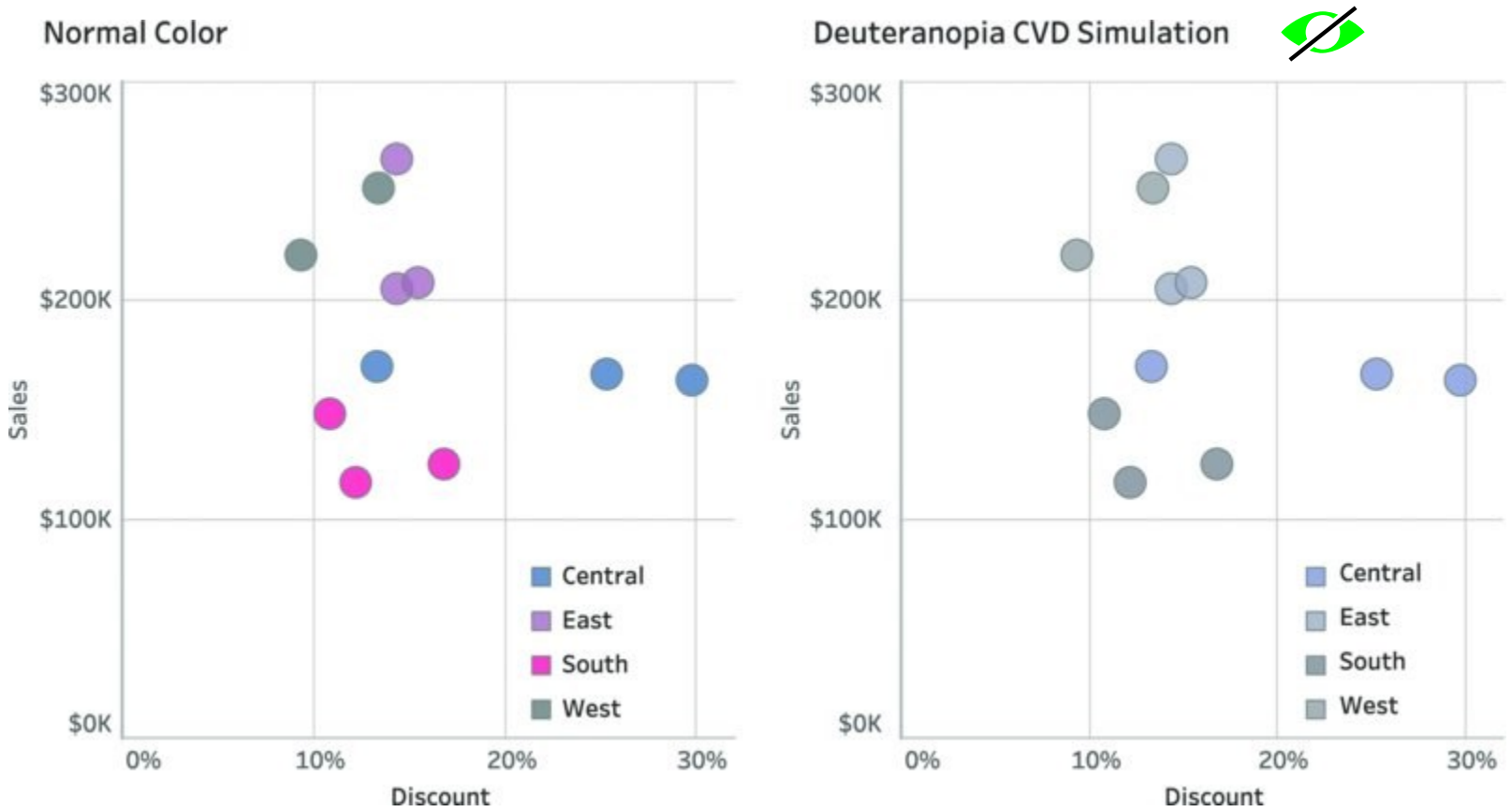
- Seeing the Problem for Yourself



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Color Blindness

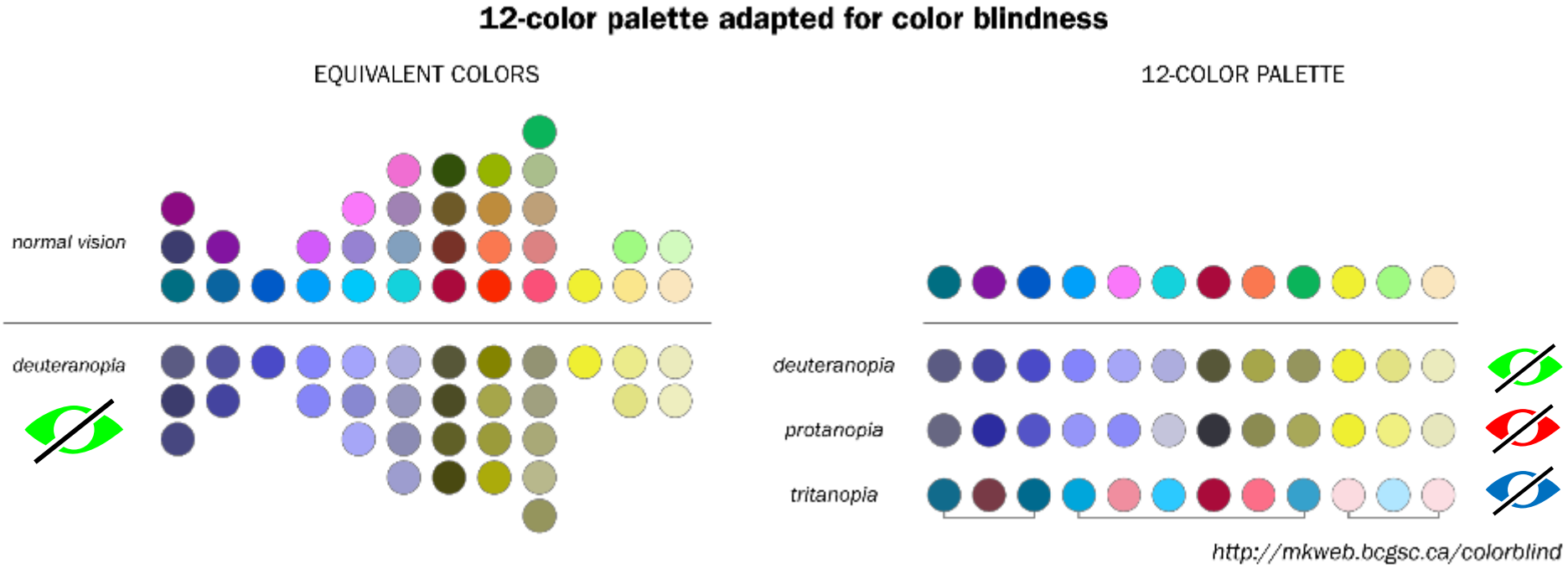
- Seeing the Problem for Yourself



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Color Blindness

- Please concern about Color blindness

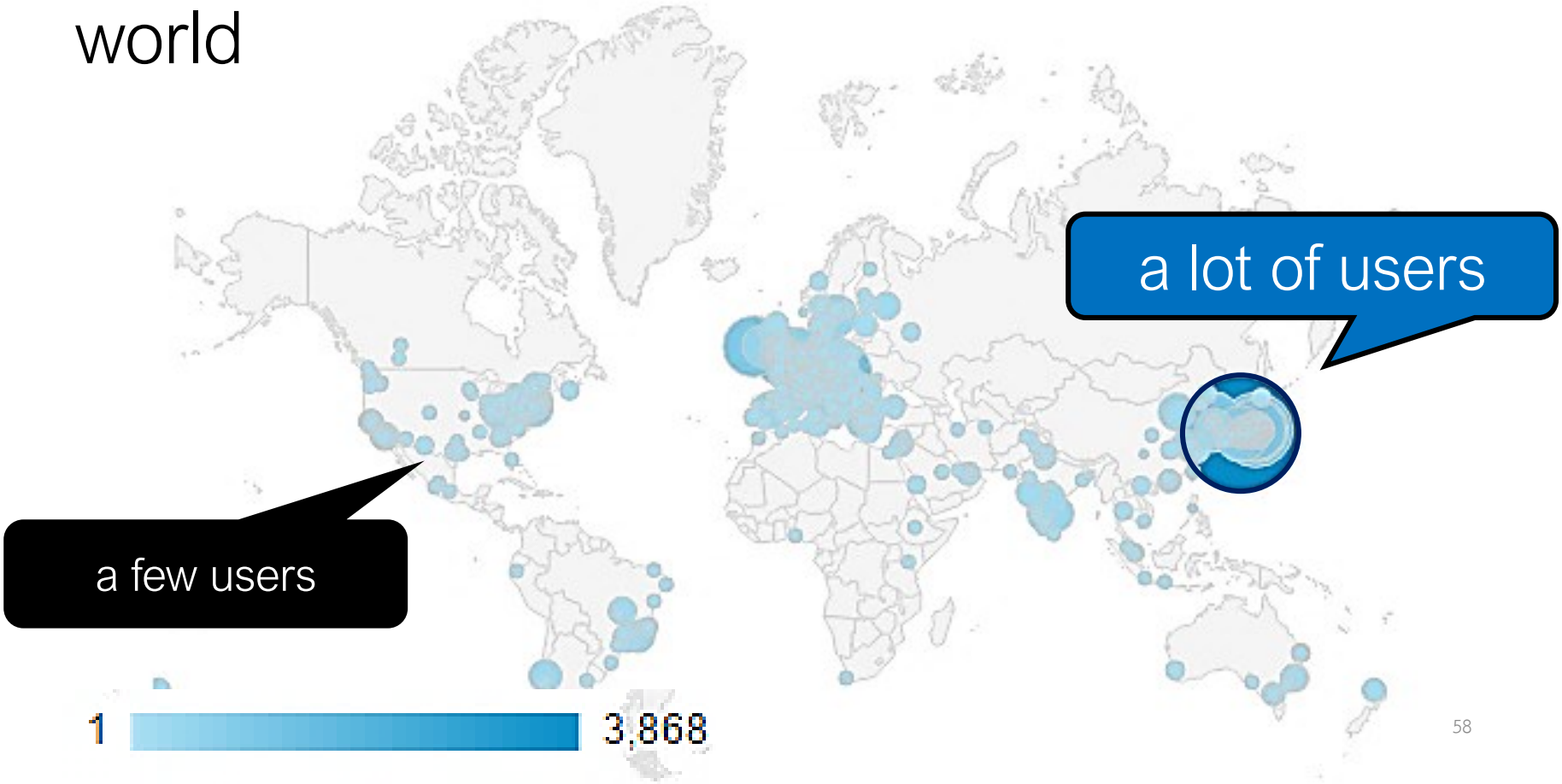


Common Visual Properties

Graphic	Encoding	Ordered	Number of Values	Quantitative	Ordinal	Categorical
	Position	✓	infinity	✓	✓	✓
	Length	✓	infinity	✓	✓	
	Area	✓	many	✓	✓	
	Thickness	✓	few	✓ (not recommend)	✓	
	Brightness	✓	few	✓ (not recommend)	✓	
	Color	✗	(<20)			✓

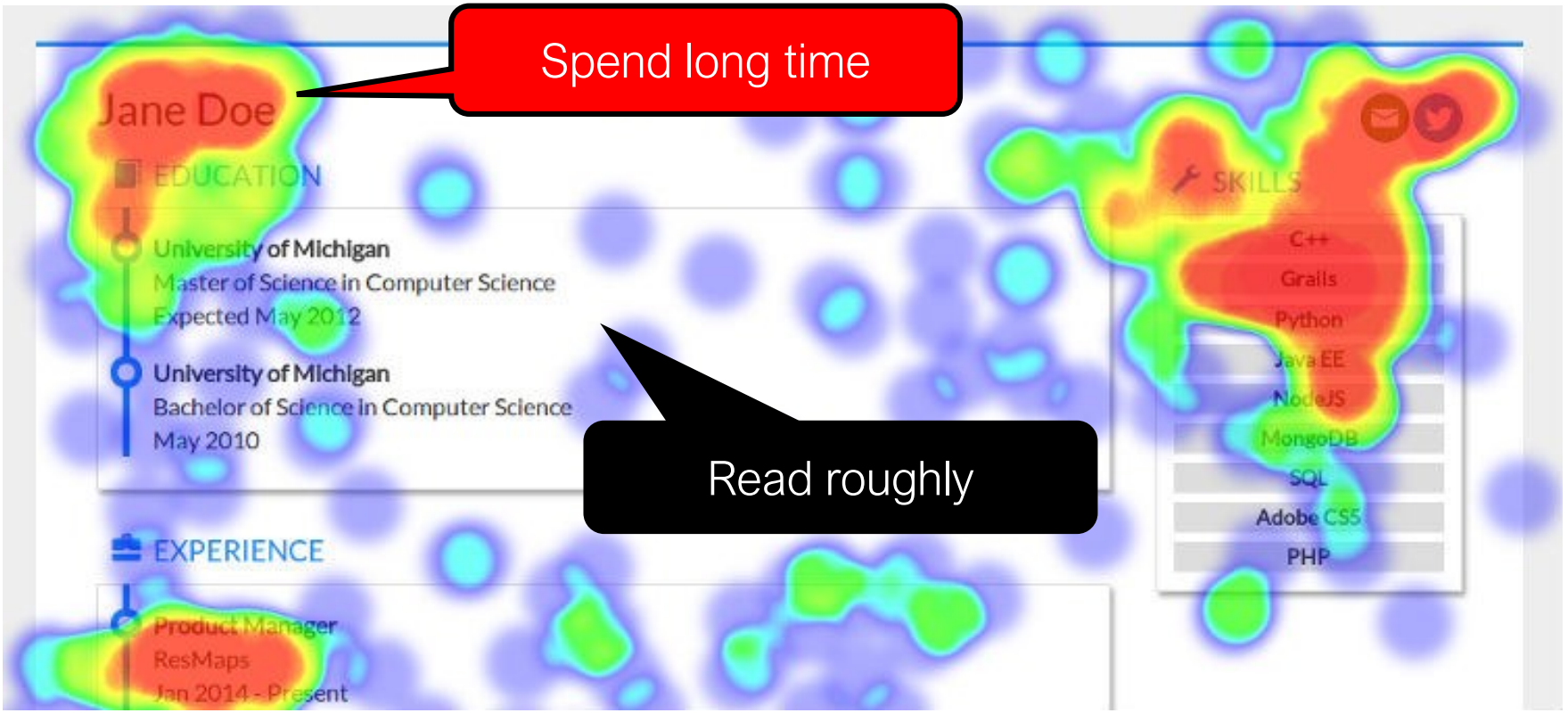
Data Visualization

- Bubble chart on a map showing number of website's visitors from different areas in the world



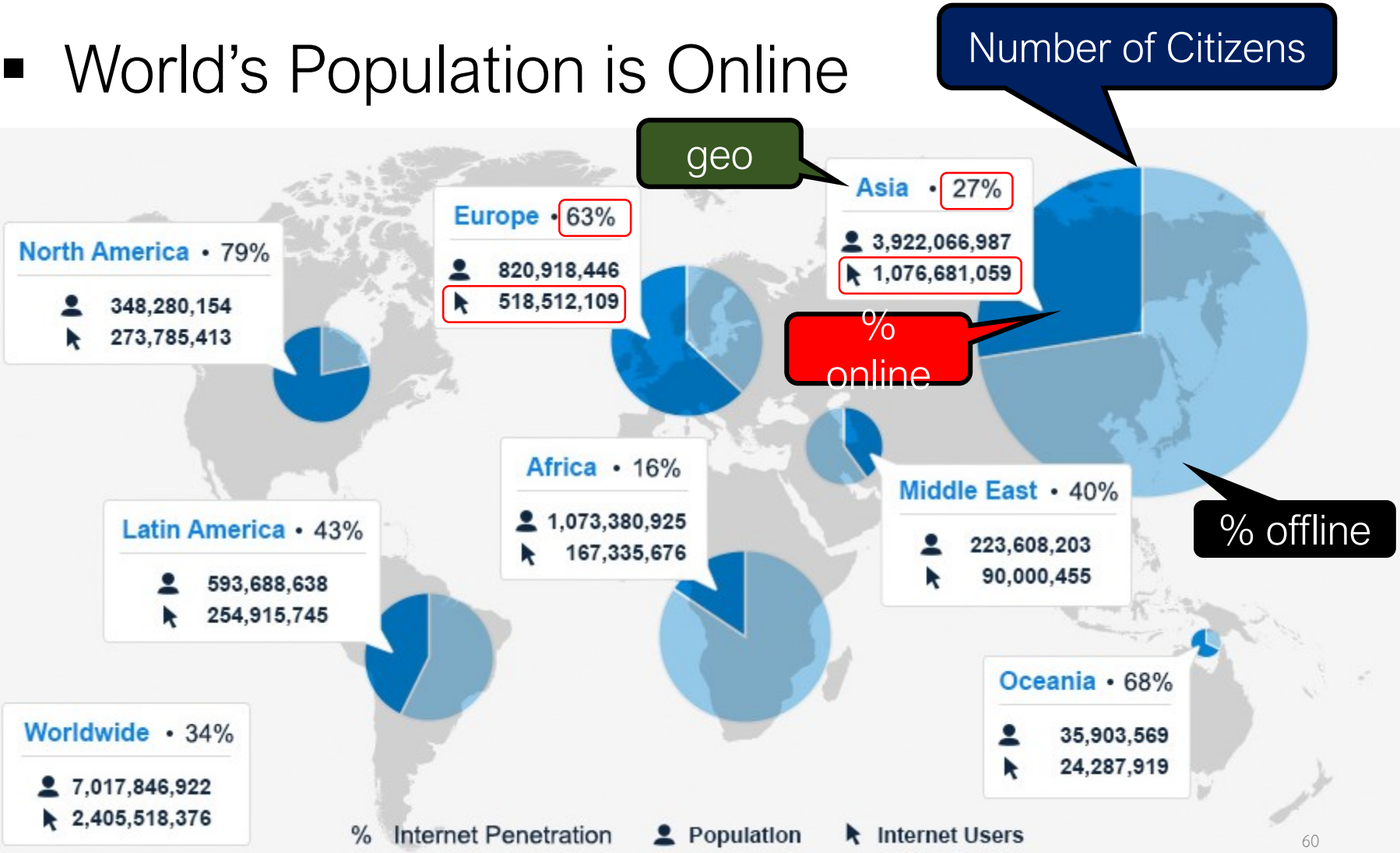
Data Visualization

- In a short time, where do they read your CV?



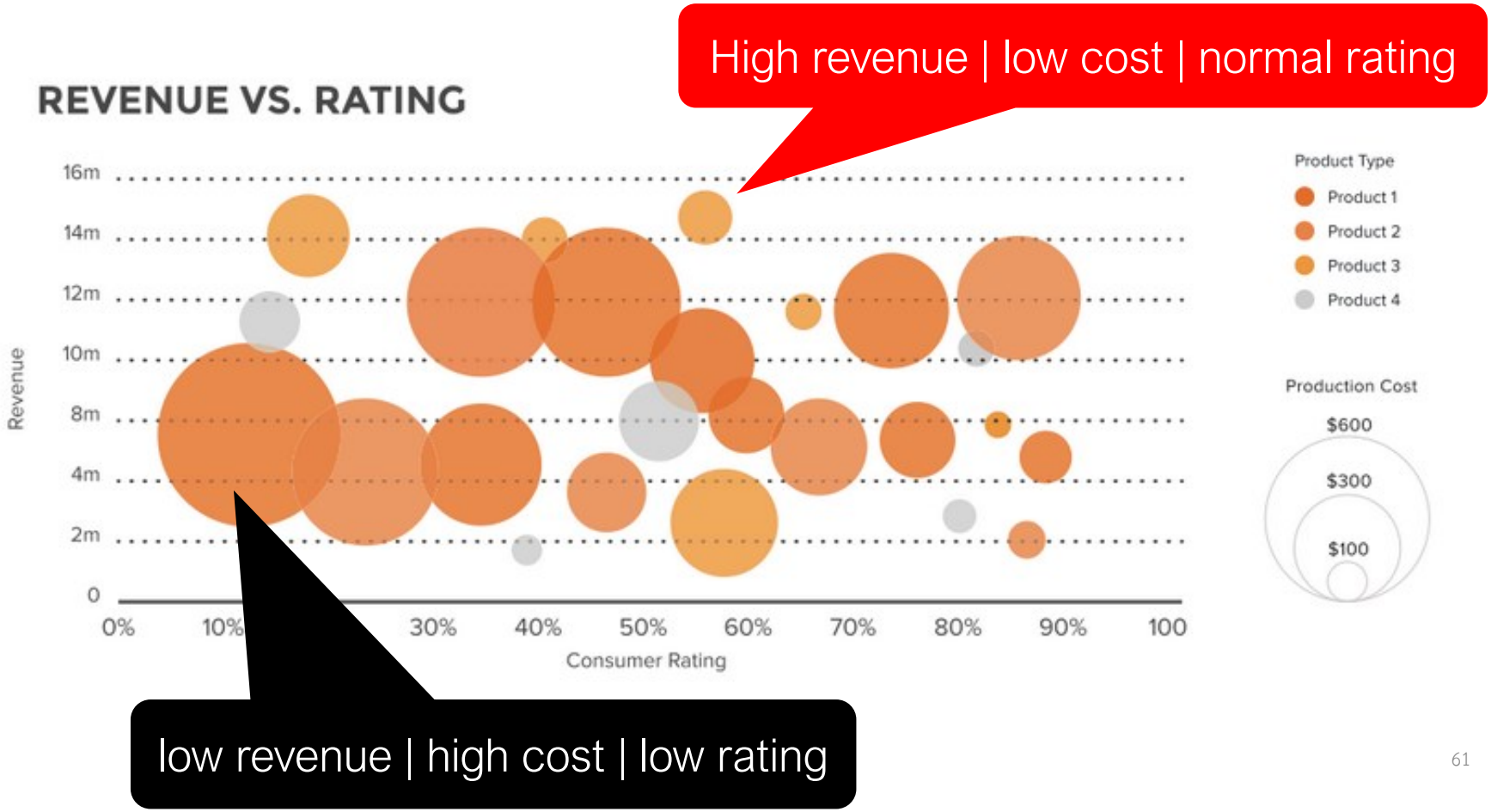
Data Visualization

- World's Population is Online



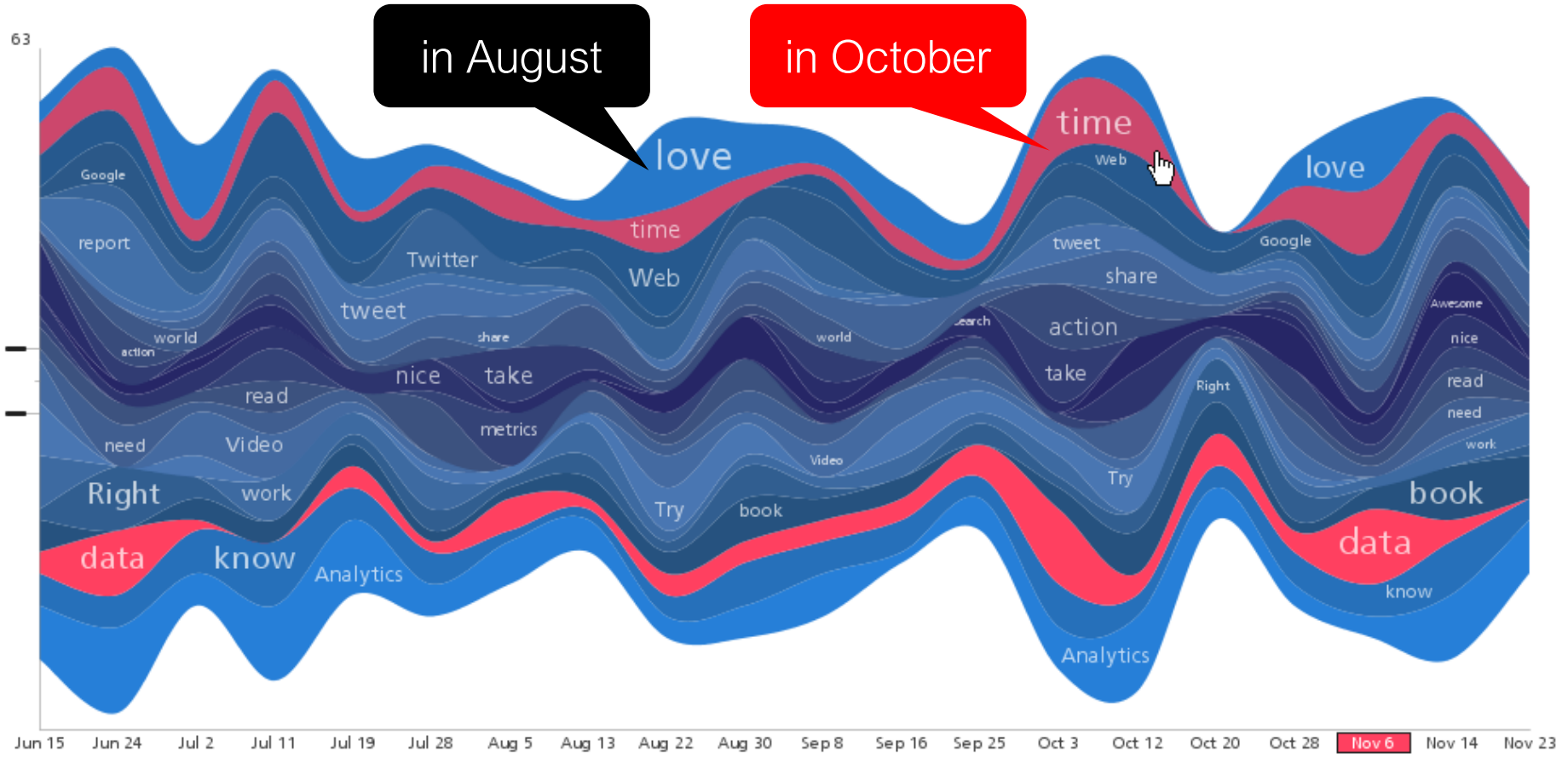
Data Visualization

- Revenue Customer Rating
  Product Cost Product Types



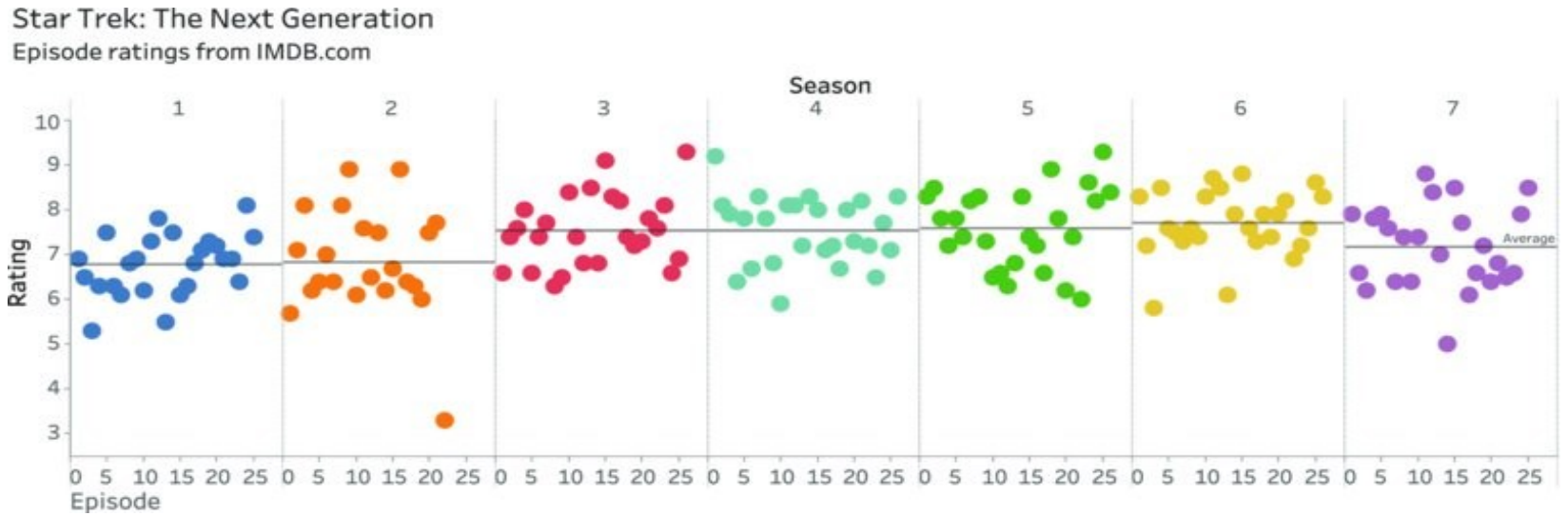
Data Visualization

- Keywords of tweets of an example account



Data in Chart

Every episode of Star Trek: The Next Generation

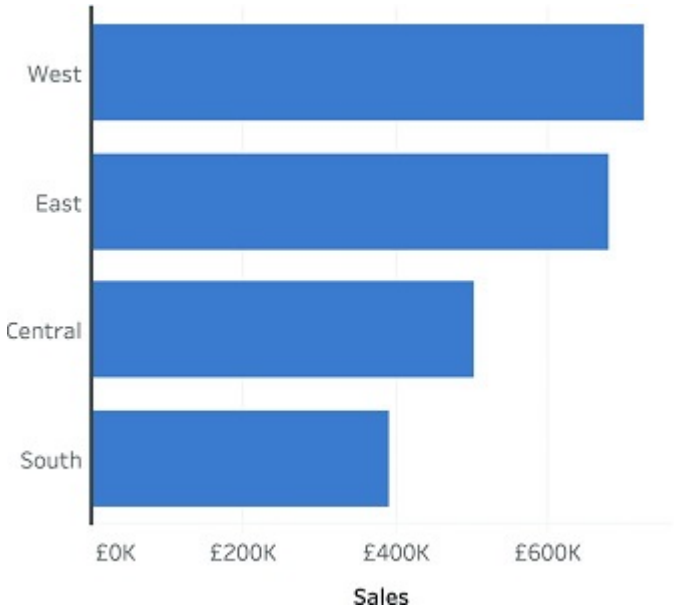


Data	Data Type	Encoding	Note
Episode	Categorical	Position	Each episode is represented by a dot. Each dot has its own position on the canvas.
Episode Number	Ordinal	Position	The x-axis shows the number of each episode in each season.
Season	Ordinal	Color	Each season is represented by a different color (hue).
		Position	Each season also has its own section on the chart.
IMDB rating	Ordinal	Position	The better the episode, the higher it is on the y-axis.
Average season rating	Quantitative	Position	The horizontal bar in each pane shows the average rating of the episodes in each season. There is some controversy over whether you should average ordinal ratings. We believe that the practice is so common with ratings it is acceptable.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Bar Chart

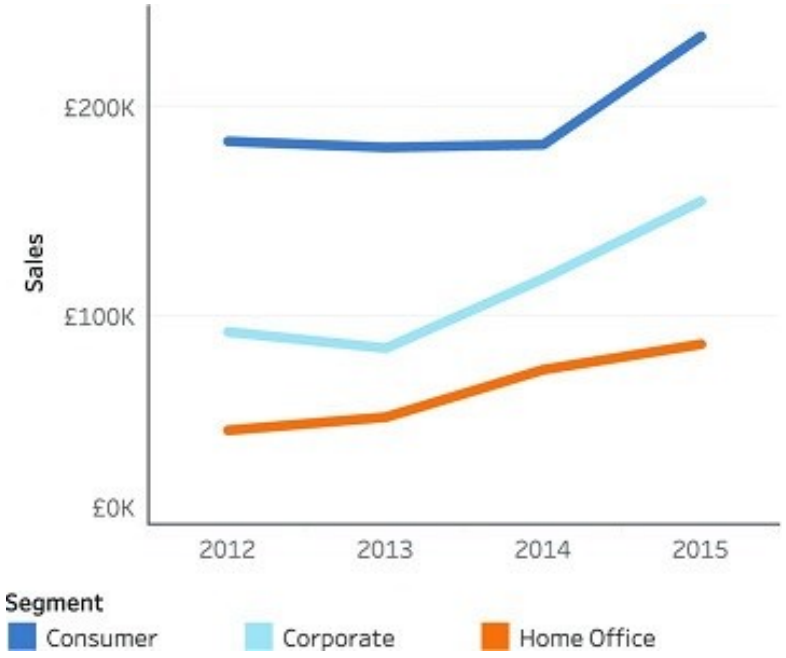
- A bar chart uses length to represent a measure.
- Human beings are extremely good at seeing even small differences in length from a common baseline. Bars are widely used in data visualization because they are often the most effective way to compare categories. Bars can be oriented horizontally or vertically. Sorting them can be very helpful because the most common task when bar charts are used is to spot the biggest/smallest items.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Time-Series Line Chart

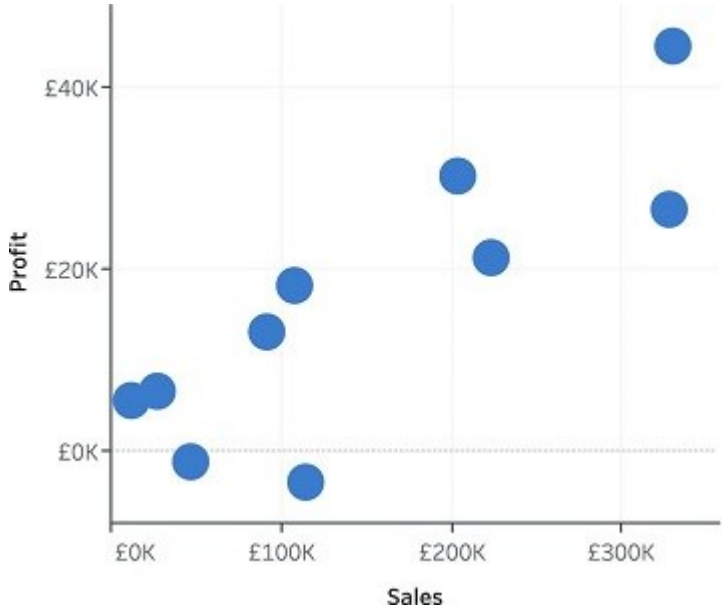
- Line charts usually show change over time. Time is represented by position on the horizontal x-axis. The measures are shown on the vertical y-axis. The height and slopes of the line let us see trends.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Scatterplot

- A scatterplot lets you compare two different measures. Each measure is encoded using position on the horizontal and vertical axes. Scatterplots are useful when looking for relationships between two variables.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

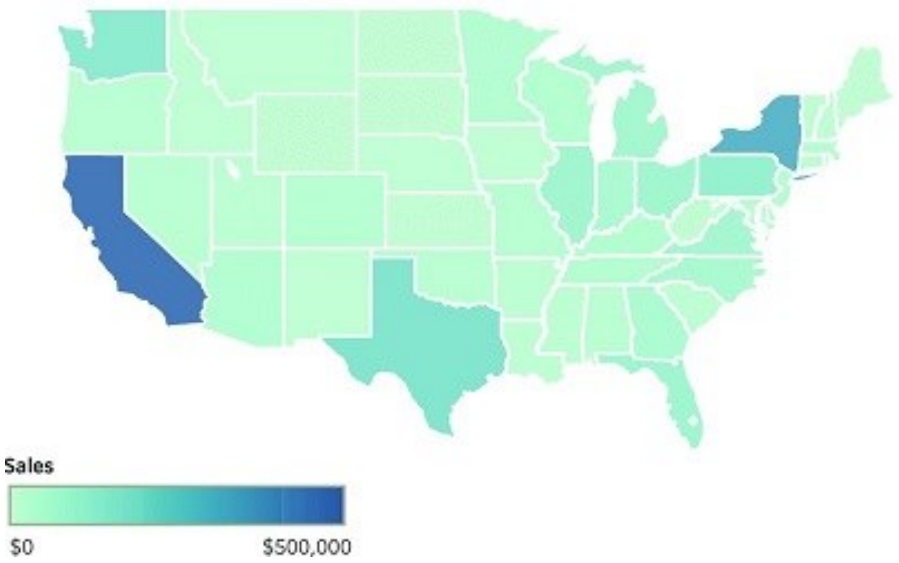
Dot Plot

- A dot plot allows you to compare values across two dimensions. In our example, each row shows sales by ship mode. The dots show sales for each ship mode, broken down by each segment. In the example, you can see that corporate sales are highest with standard class ship mode.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Choropleth Map

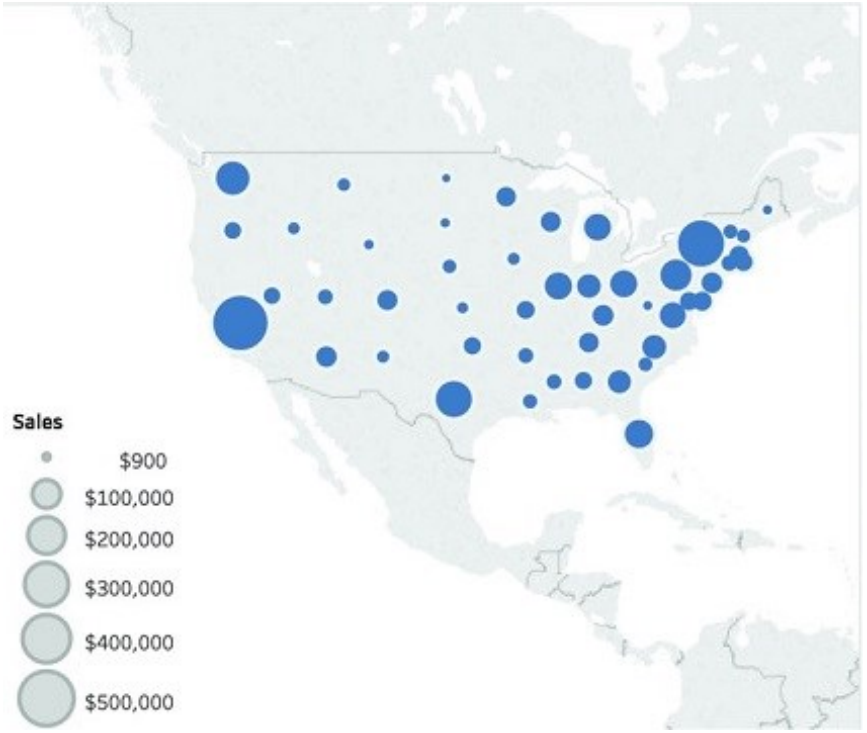
- A choropleth (also known as a filled) map uses differences in shading or coloring within predefined areas to indicate the values or categories in those areas.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Symbol Map

- A symbol map shows values in specific places. These could be the center points of large regions (e.g., the center of each U.S. state) or specific locations determined by an exact latitude/longitude measurement.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Table

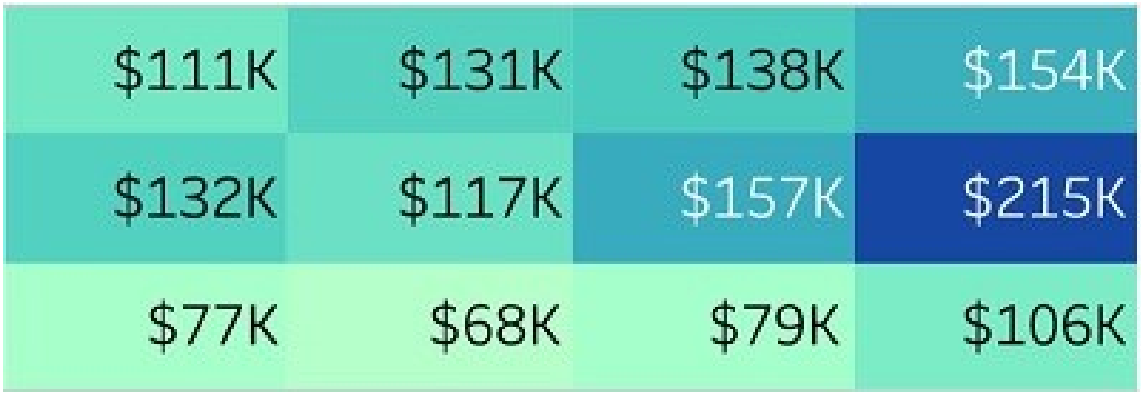
- Sometimes you do need to be able to look up exact values. A table is an acceptable way to show data in that situation. On most dashboards, a table shows details alongside summary charts.

$111K	$131K	$138K	$154K
$132K	$117K	$157K	$215K
$77K	$68K	$79K	$106K

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Highlight Table

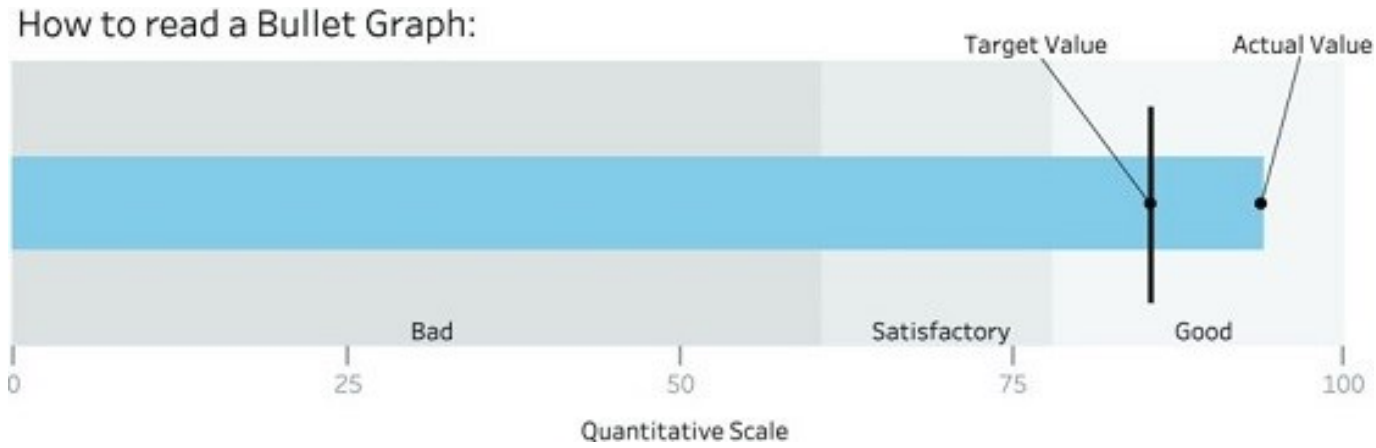
- Adding a color encoding to your tables can transform them into highly visual views that also enable exact lookup of any value.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Bullet Graph

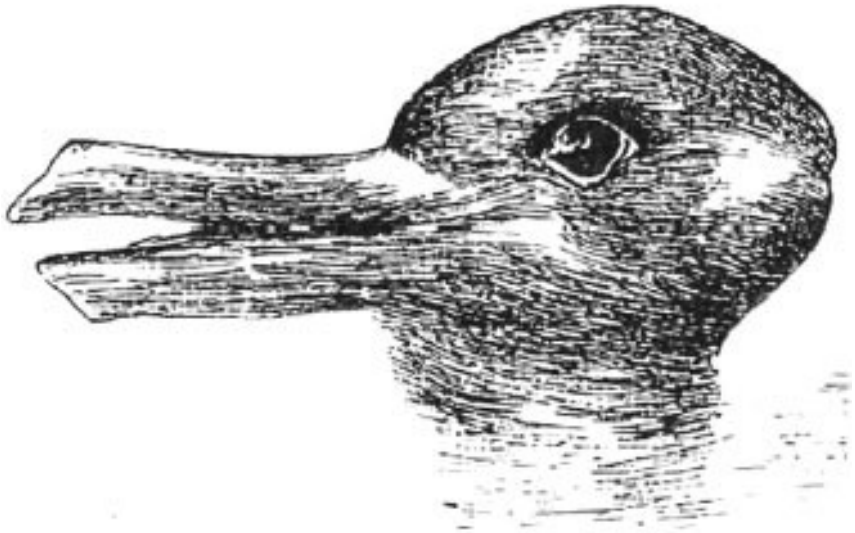
- A bullet graph is one of the best ways to show actual versus target comparisons. The blue bar represents the actual value, the black line shows the target value, and the areas of gray shading are performance bands.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

When Our Visual Processing System Betrays Us

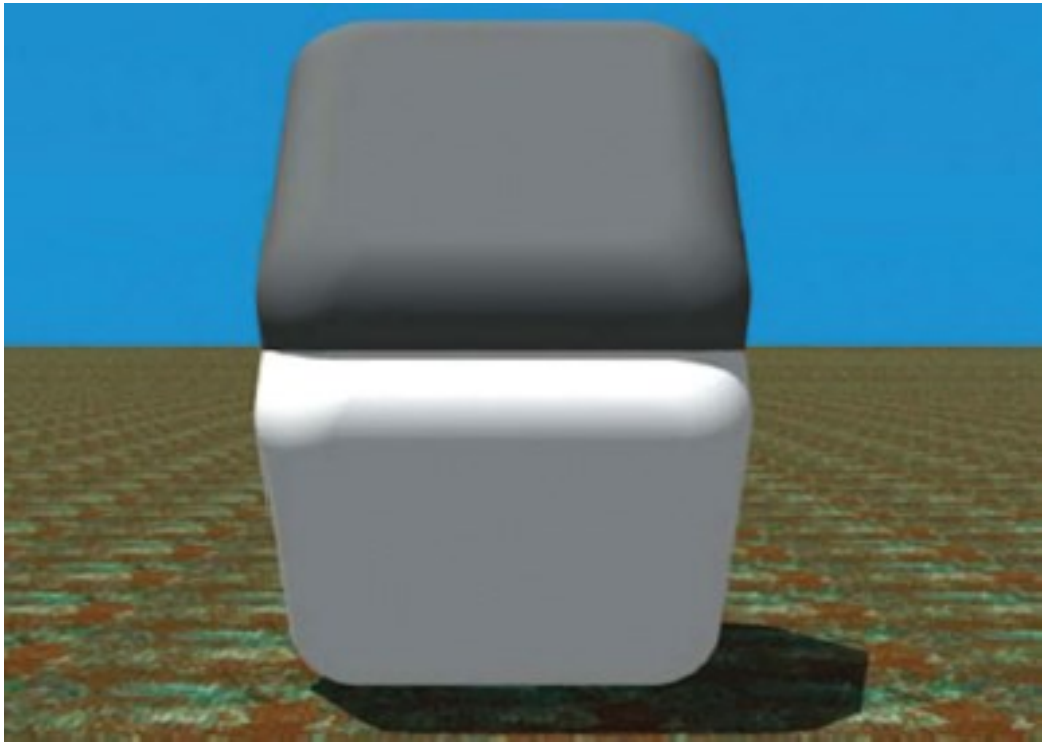
- We have talked about how to use pre-attentive attributes to craft good data visualizations. By using those attributes, we can use the power of our visual system to our advantage. Unfortunately, our visual system also can be confused easily.
- Duck of Rabbit?



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

When Our Visual Processing System Betrays Us

- Does the top appear darker than the bottom?

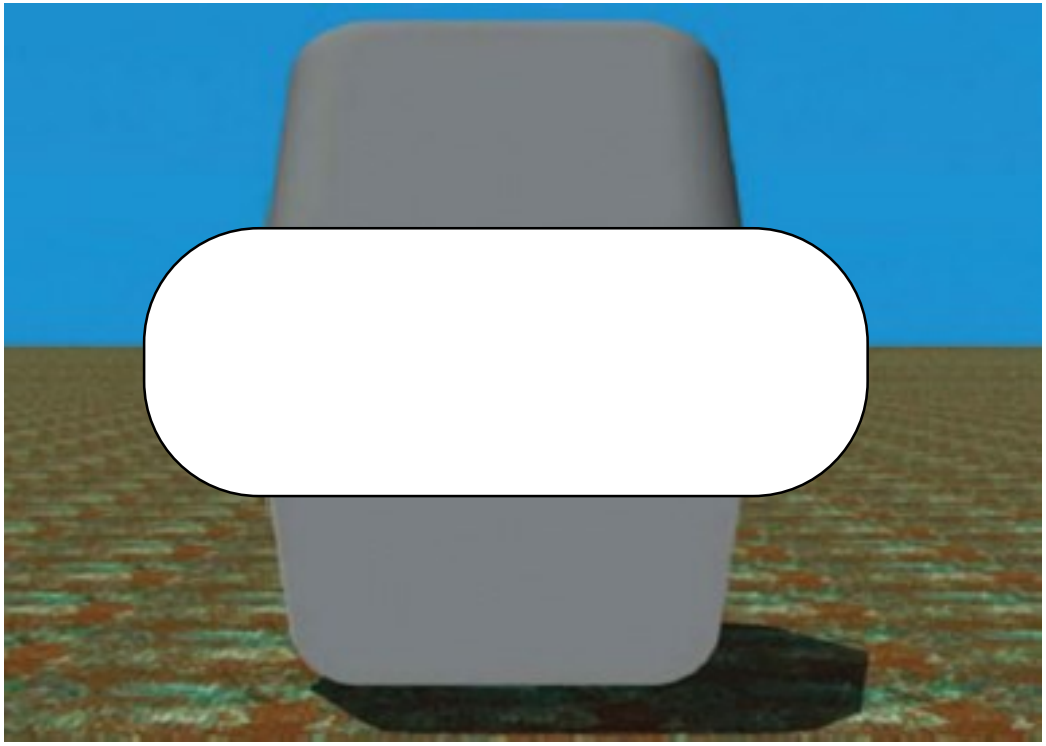


Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

When Our Visual Processing System Betrays Us

- Does the top appear darker than the bottom?

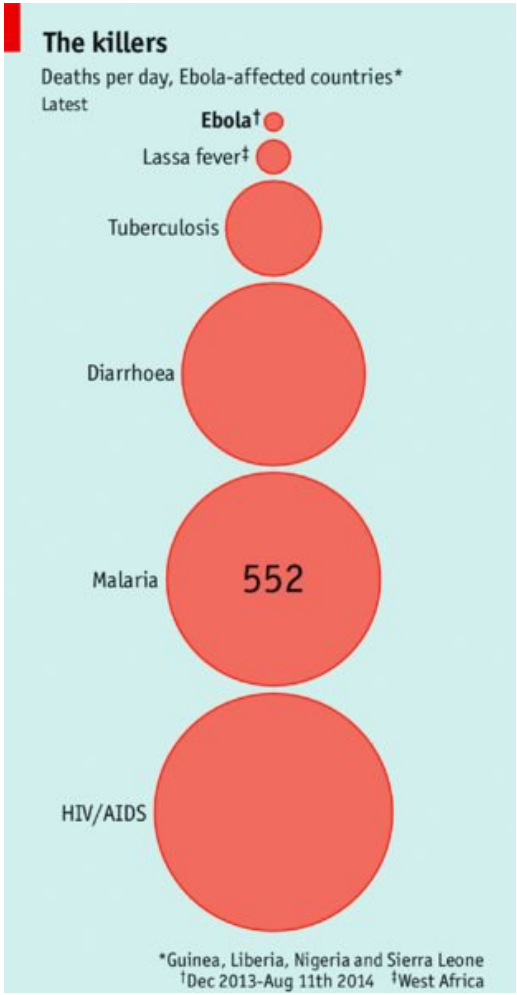
  Put your thumb or finger over the center line and then try again.

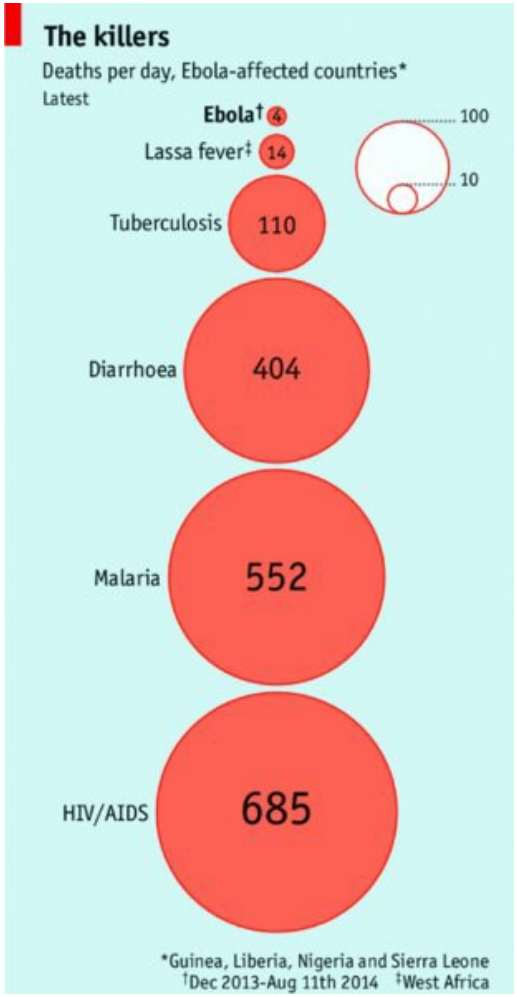


Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Ambiguity in Images

- Each circle is sized according to the number of deaths.
- We have removed all the labels except the one for malaria (552 deaths/day).
- How many deaths per day are there from diarrhea? How much bigger is the HIV/AIDS circle than the diarrhea circle?

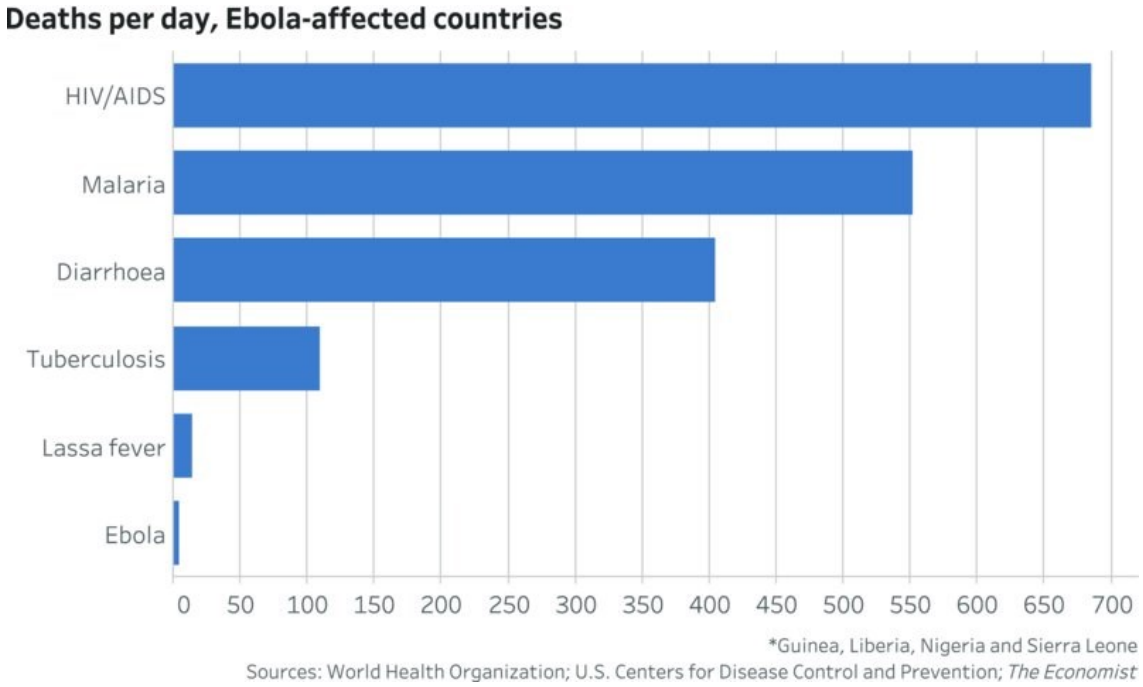




Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Bar chart version of the circle charts.

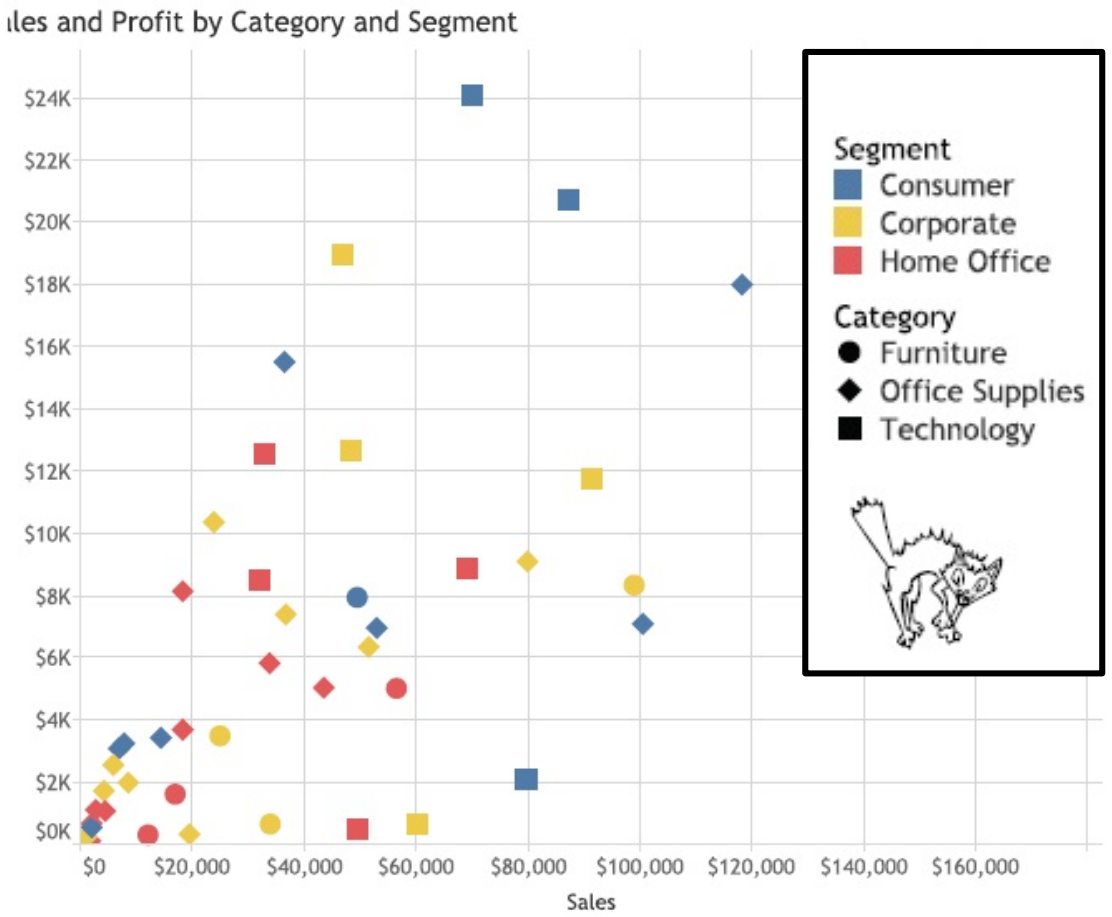
- Most people underestimate the size of the bigger circles. The point is that while size is pre-attentive, we're not able to tell the differences with any accuracy. Consider the same data shown as a bar chart in



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

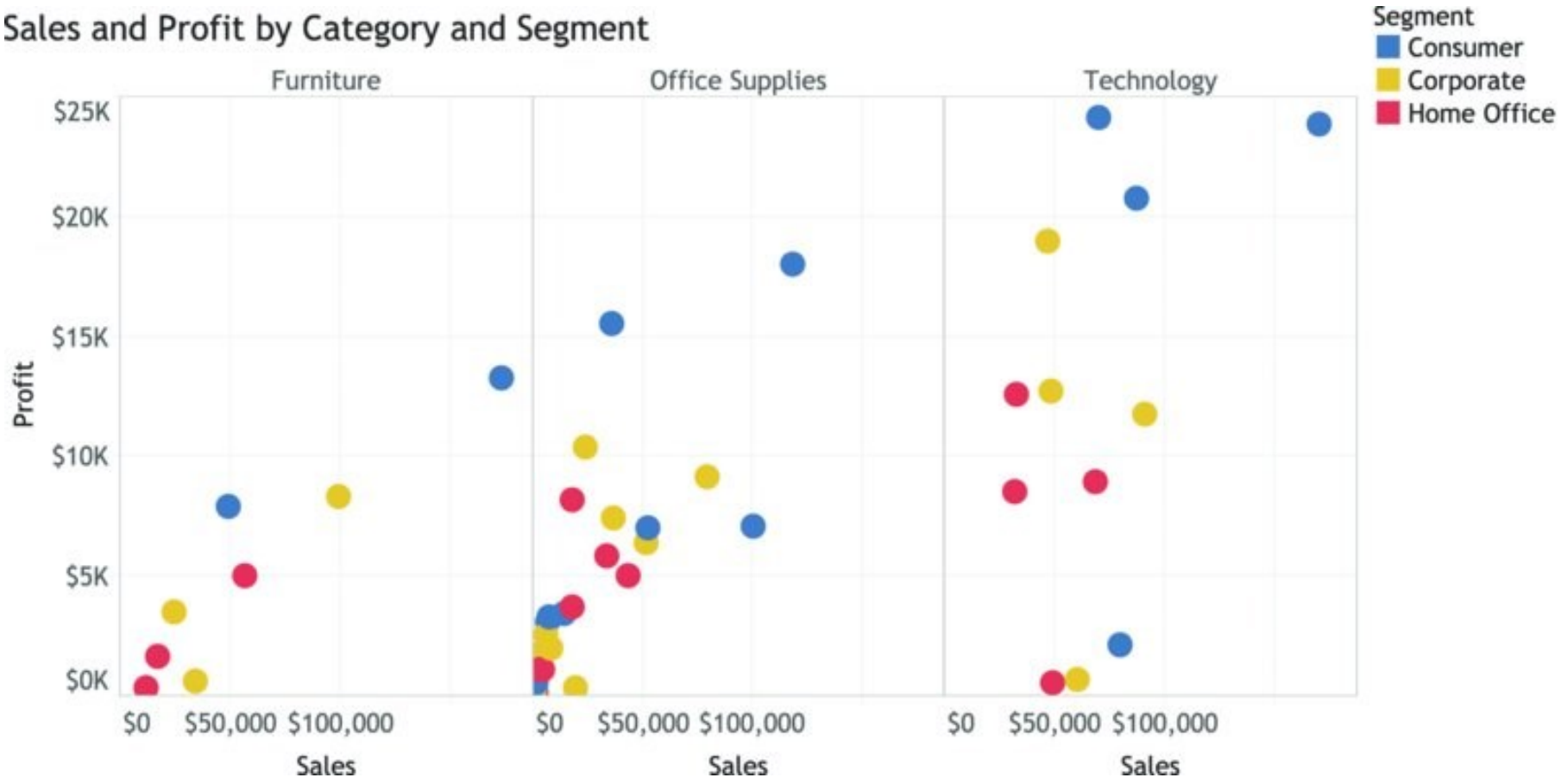
Using multiple pre-attentive attributes in one chart can lead to problems

- shows a scatterplot of sales and profit for a fictional sales company.
- Position is used for sales (x-axis) and profit (y-axis).
- Color shows different segments.
- Shape shows the categories of products.
- Which category has, on average, the highest profits?



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

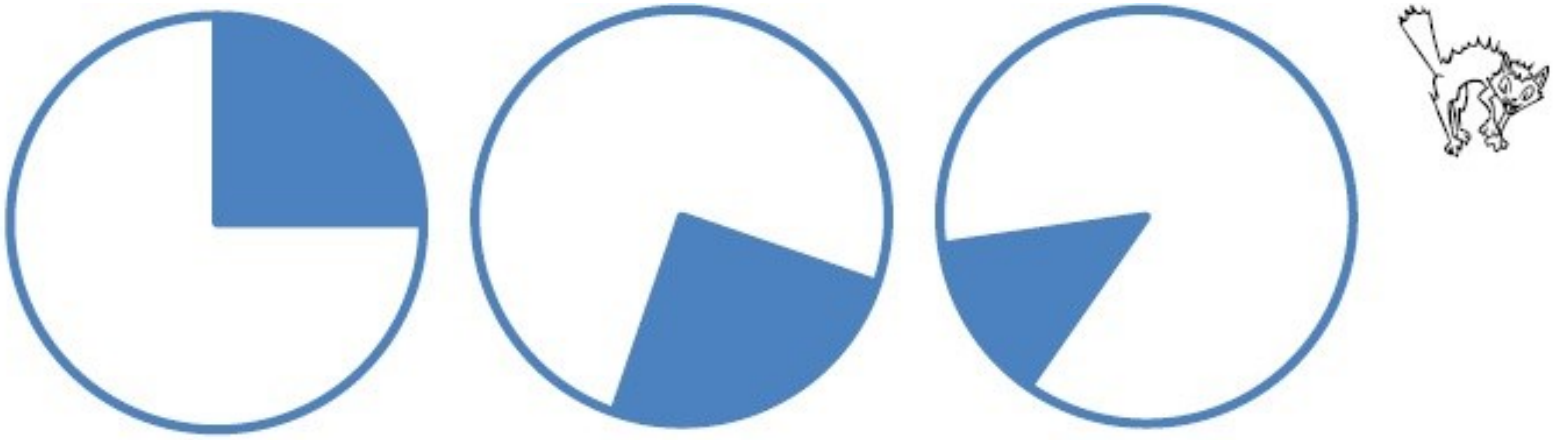
Breaking the single scatterplot into three panels



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Pie charts don't play well with visual systems

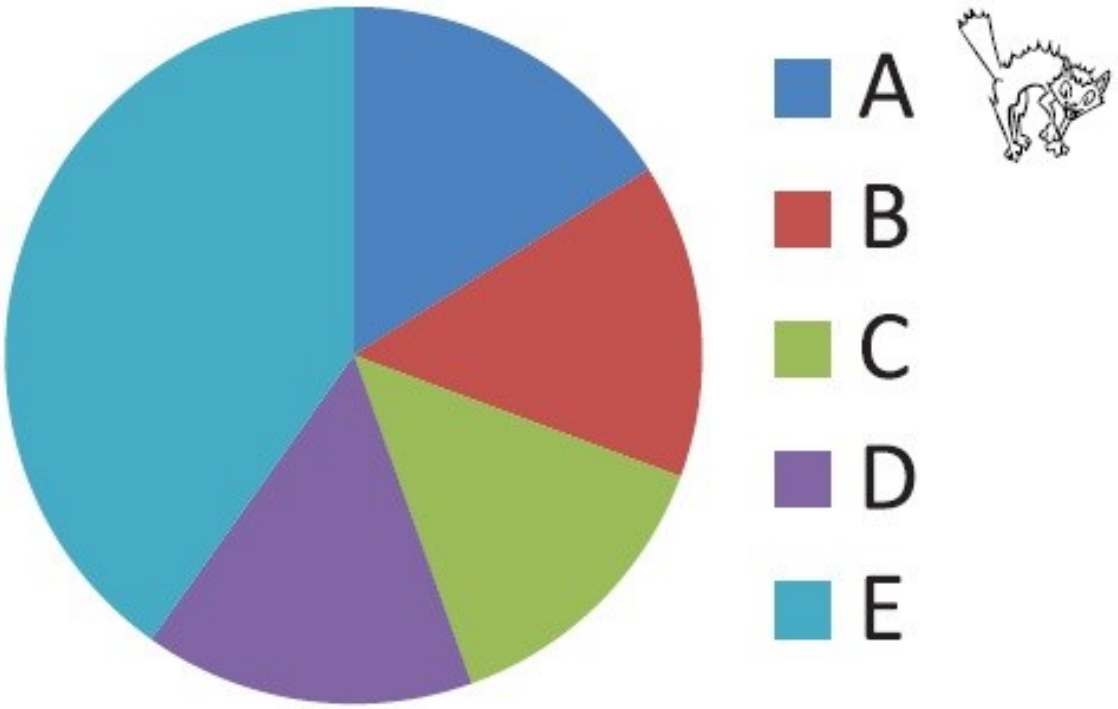
- What percentage of each circle is covered by the blue segment?



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Pie charts don't play well with visual systems

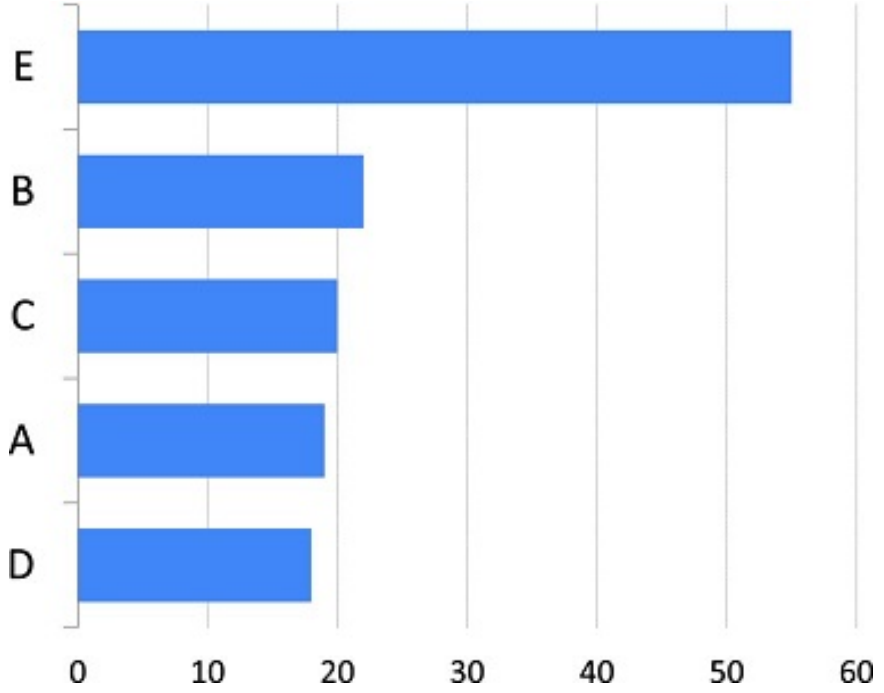
- What percentage of each pie does the blue segment represent?



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Bars make it very easy to see small differences in size

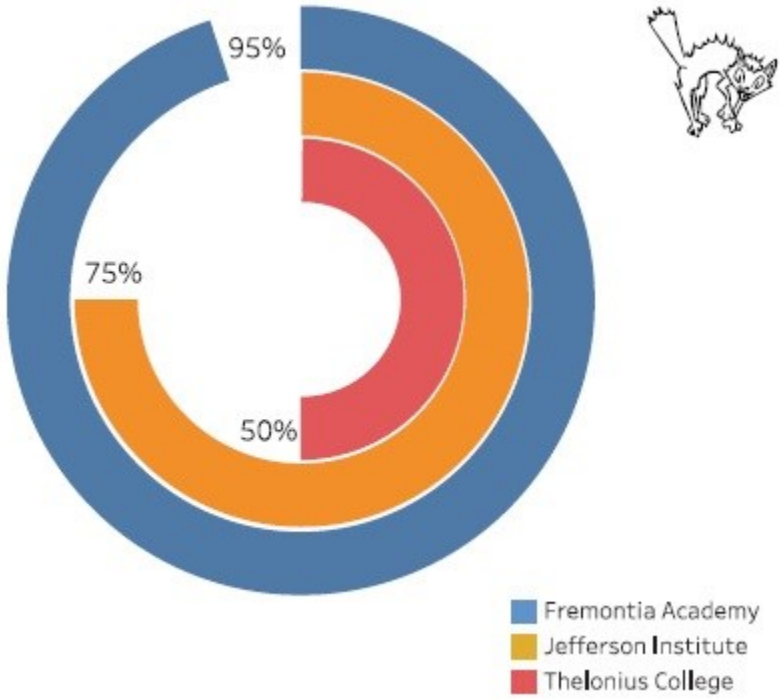
- The sorted bar chart made it very easy to distinguish size differences: Length is such an effective visual attribute, we can see very small differences with ease.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Keep you away from the lure of the circles

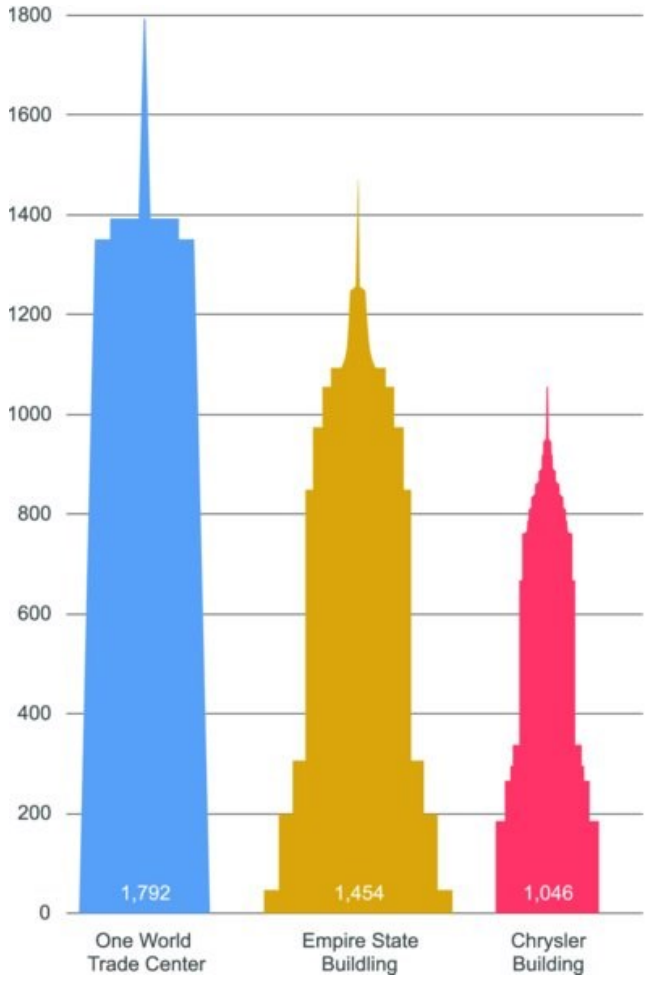
- To make effective dashboards, you must resist the temptation to use purely decorative chart types.
- Let's look at one more example in order to keep you away from the lure of the circles. Sometimes people acknowledge the power of bars but then get tempted to put them in a circle, fashioning what is known as a donut chart.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

3 famous buildings

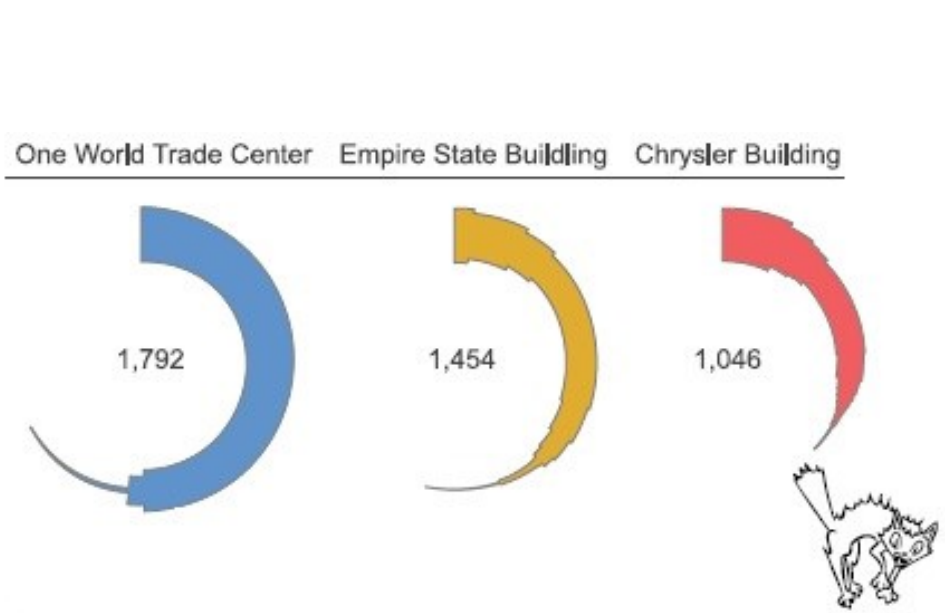
- Let's suppose you wanted to compare the heights of three famous buildings:
  - One World Trade Center,
  - the Empire State Building,
  - the Chrysler Building.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Difficult Comparisons

- Now let's see how easy the comparison is with donuts
- Can you tell the difference in heights of the buildings in the second chart?



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

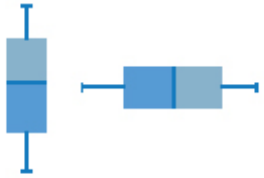
Glossary of Chart Types

Area Chart

encodes data using position and height and shows trend/volume over time.

Bar Chart

encodes data using height/length of bar and shows categorical comparisons.



Box Plot

encodes data using position and height/length to show the distribution of the data.

Bullet Graph

encodes data using length/height, position, and color to show actual compared to target and performance bands.

Choropleth Map (Shaded Map)

encodes data using color and position to show data geographically.

Diverging Bar Chart

encodes data using height/length of bar diverging from a midpoint to show categorical comparisons.

Dot Plot

encodes data using position to show the comparisons.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Glossary of Chart Types

Dot Plot with Jitter (Jitterplot)

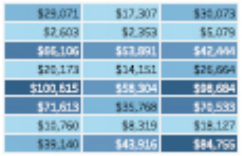
encodes data using position to show comparisons but offsets points randomly to reduce overlap of dots.

Gantt Chart

encodes data using length and position to show amount of work completed in segments of time.

Heat Map

encodes a data table using color to highlight the differences in the table without numbers.



Highlight Table

encodes a data table using color to highlight the differences in the table numbers.

Histogram

encodes data using height and shows a distribution.

Line Chart

encodes data using position and often shows trend over time.

Lollipop Chart

encodes data using height/length of bar and shows categorical comparisons.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Glossary of Chart Types

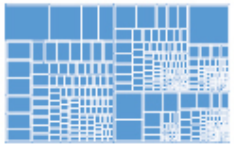
Scatter Plot
encodes data using position to show the relationship between two variables. Size can also be used to show a secondary comparison.

Slopegraph
encodes data using position to show quantitative comparison or rank, typically between two time periods.

Sparkline/Sparkbar
encodes data using position (line) or height/length (bar) in a small, word-sized graphic.

Stacked Bar Chart *
encodes data using height/length of bar and color by segment and shows categorical and part-to-whole comparisons.

Symbol Map (Dot Map)
encodes data using position to show data geographically and can also use size to show quantitative data.



Treemap
encodes data using size and color and is useful for hierarchical data or when there is a very large number of categories to compare.

Waterfall Chart
encodes data using height and often color to show increase and decrease between time periods or categories.

* Caution
Be careful not to slice stacked charts into too many segments.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Glossary of Chart Types

Bubble Chart

encodes data using size of circle to show comparisons, which is difficult for making precise quantitative comparisons.

Concentric Circles (Radial Bar Chart)

encodes data using arc and area to show comparisons but problematic for many reasons.

Donut Chart

encodes data using arc and area to show a part-to-whole comparison but problematic for many reasons.

Pie Chart

encodes data using angle, area, and arc to show a part-to-whole comparison but problematic for many reasons.

Word Cloud

encodes data using size of word to show comparisons, which is difficult for making precise quantitative comparisons.

Caution

These chart types are not recommended.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

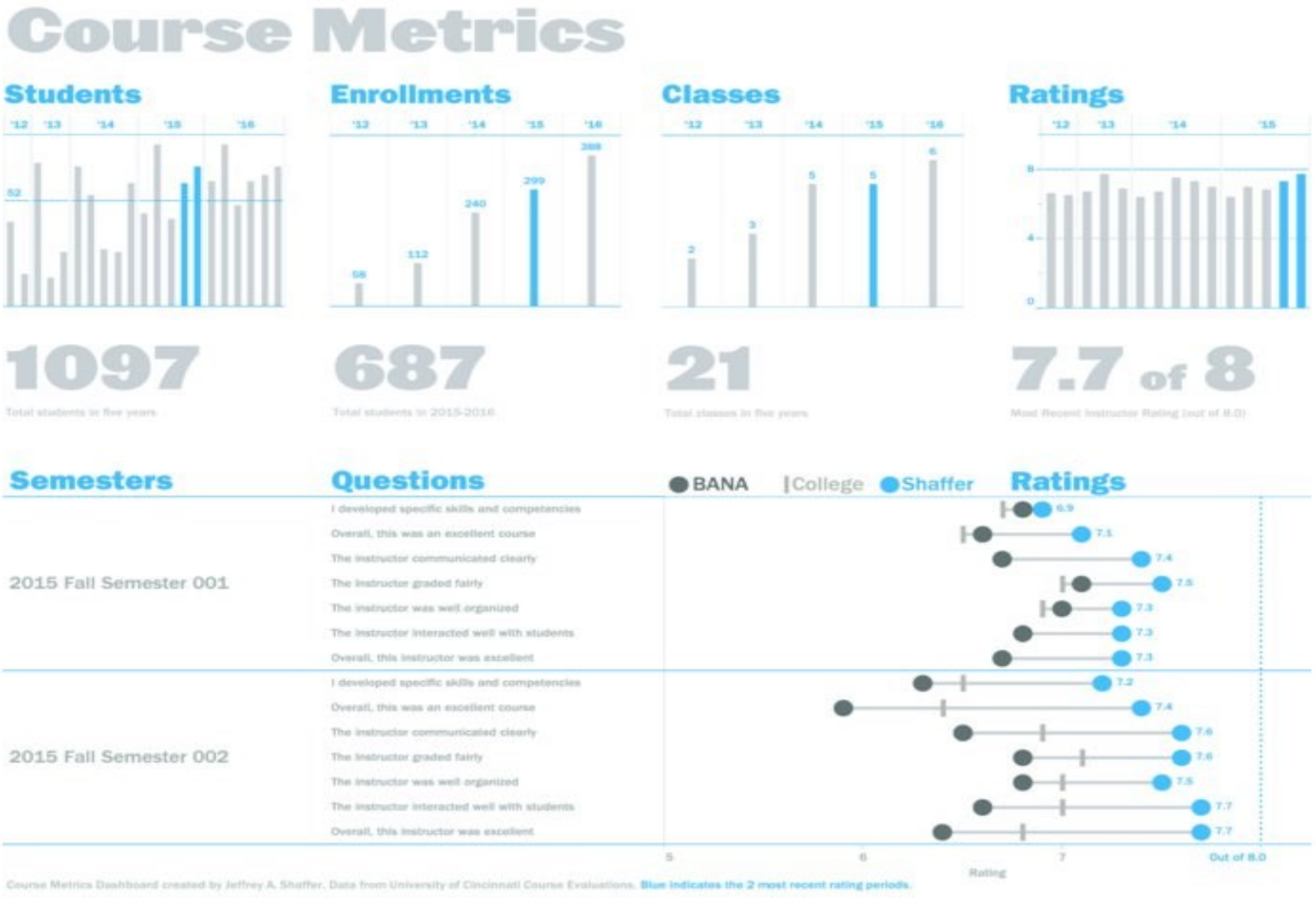
Dashboard

Course Metrics Dashboard

Course Metrics Dashboard

- Big Picture
  - You are a professor at a university or the head of an academic department. You want to know how a particular professor's course is rated compared to others in the department and the college. You want to know the overall course load, the number of students, and the overall growth or decline of a particular course. You also want to be able to see the rating comparing one specific course against all of the courses in the department and in the college.
- Specifics
  - You need to see how many courses an instructor has been teaching over time.
  - You need to see how many students are registering for the classes.
  - You want to see the trend over a specific period of time. This might be the entire span of the course (as shown in the overview dashboard), or it may be a rolling period of time, for example, the last five years.
  - You would like to see the detailed ratings of the most recent course and instructor feedback.
  - You need to be able to quickly compare this course and instructor to other courses in the department and the college.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

How People Use the Dashboard

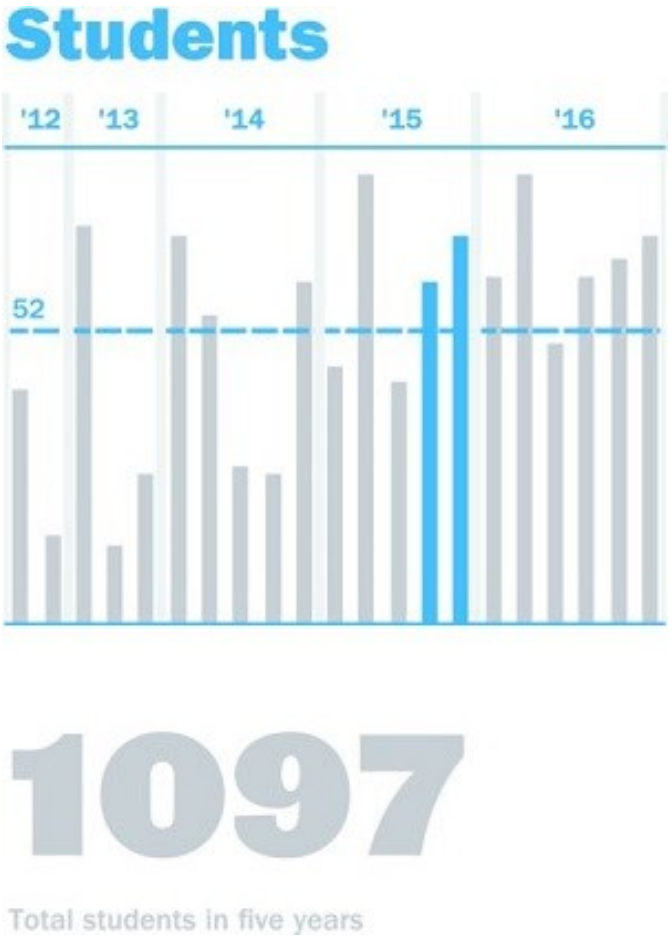
- This dashboard shows the course history for the Data Visualization class at the University of Cincinnati. The course is taught by Jeffrey Shaffer in the Carl H. Lindner College of Business and is part of the operations, business analytics, and information systems department. Courses are registered under their discipline within the department.
  - **OM** Operations Management
  - **BANA** Business Analytics
  - **IS** Information Systems

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Number of students

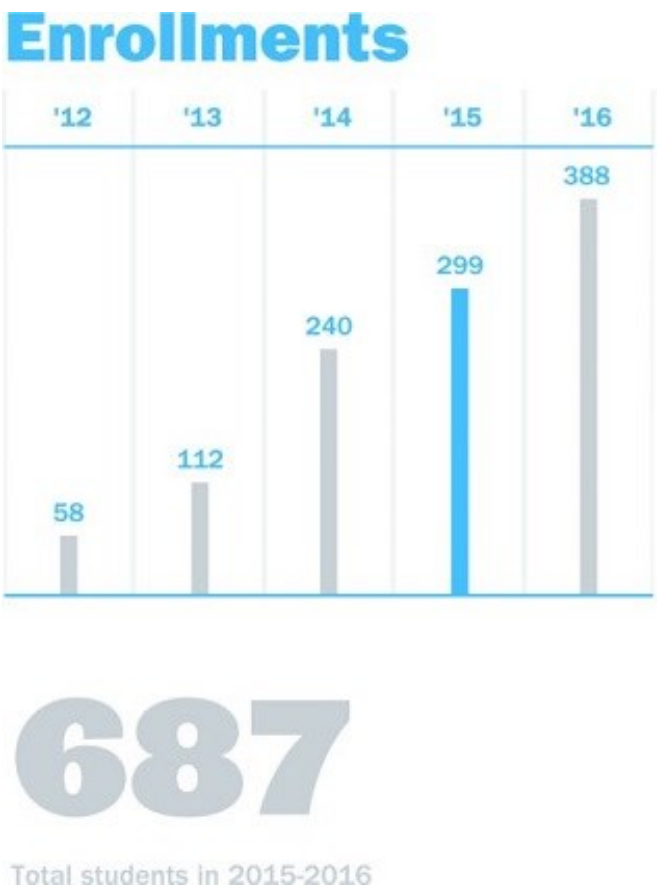
- Number of students in each course by year with two recent rating periods highlighted in blue.

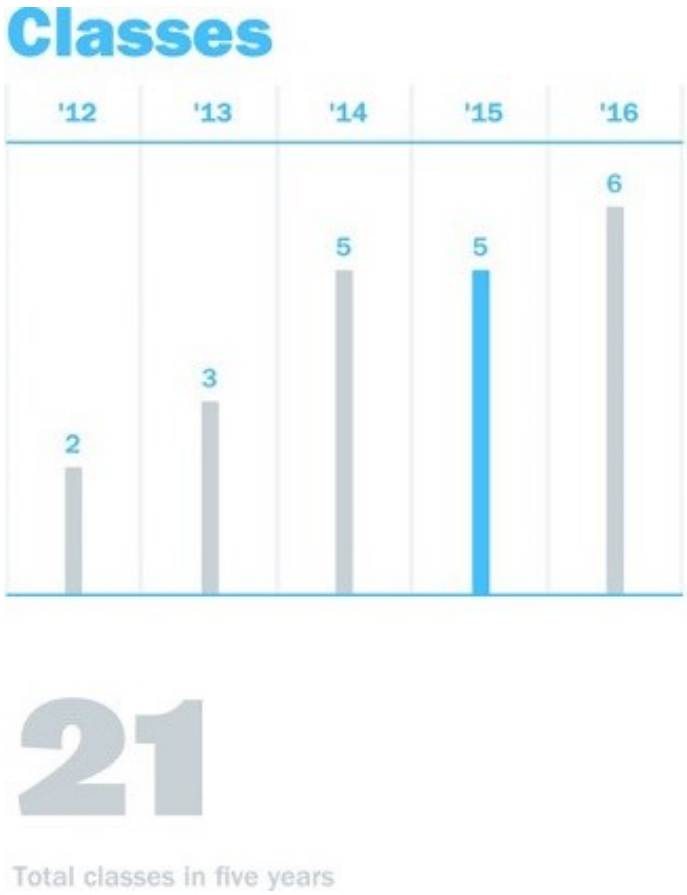
- (2016 has not been rated yet.)



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Enrollments & Classes by Year

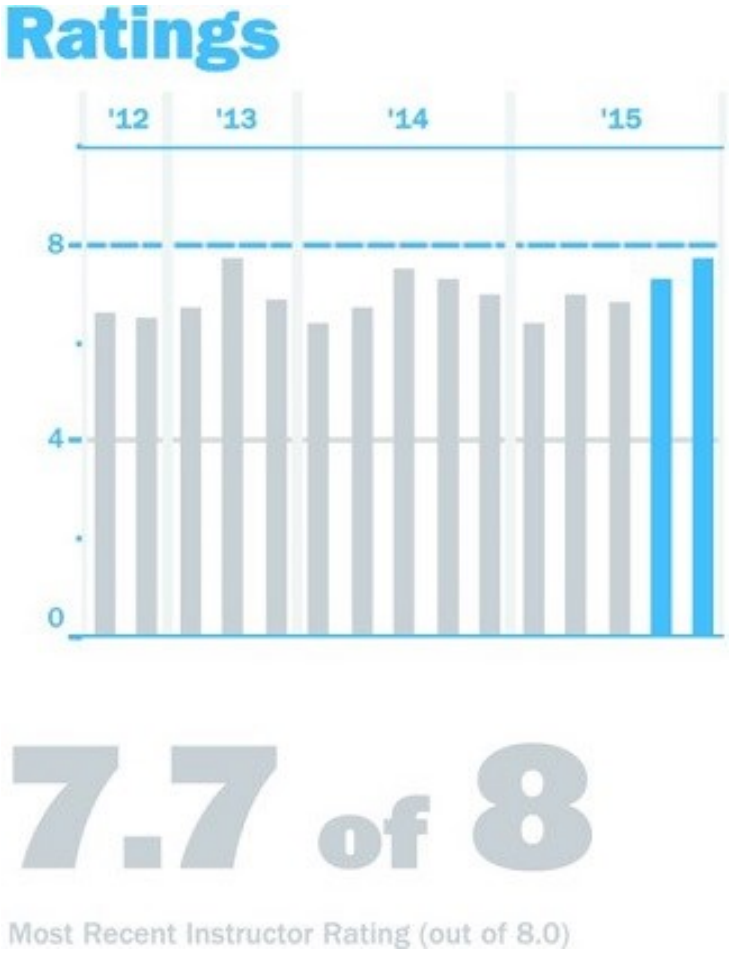




Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Rating

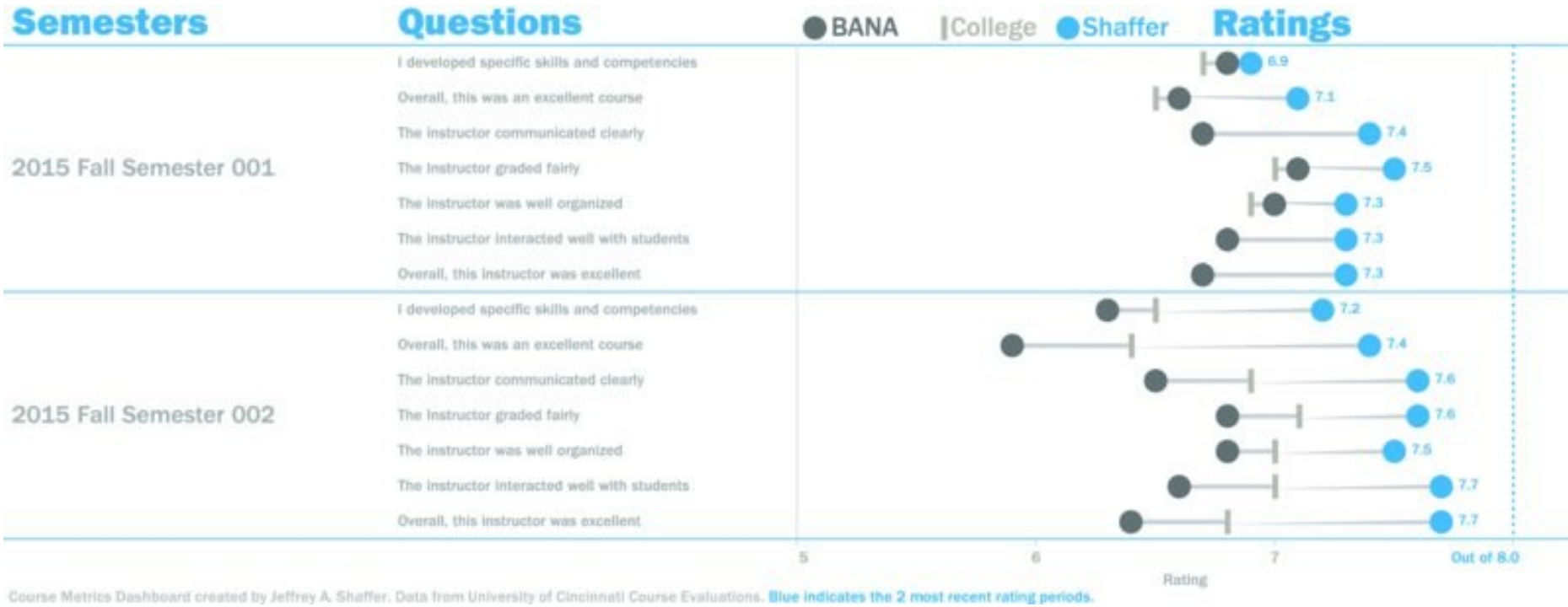
- the overall course rating for each course semester by semester



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Rating Comparison

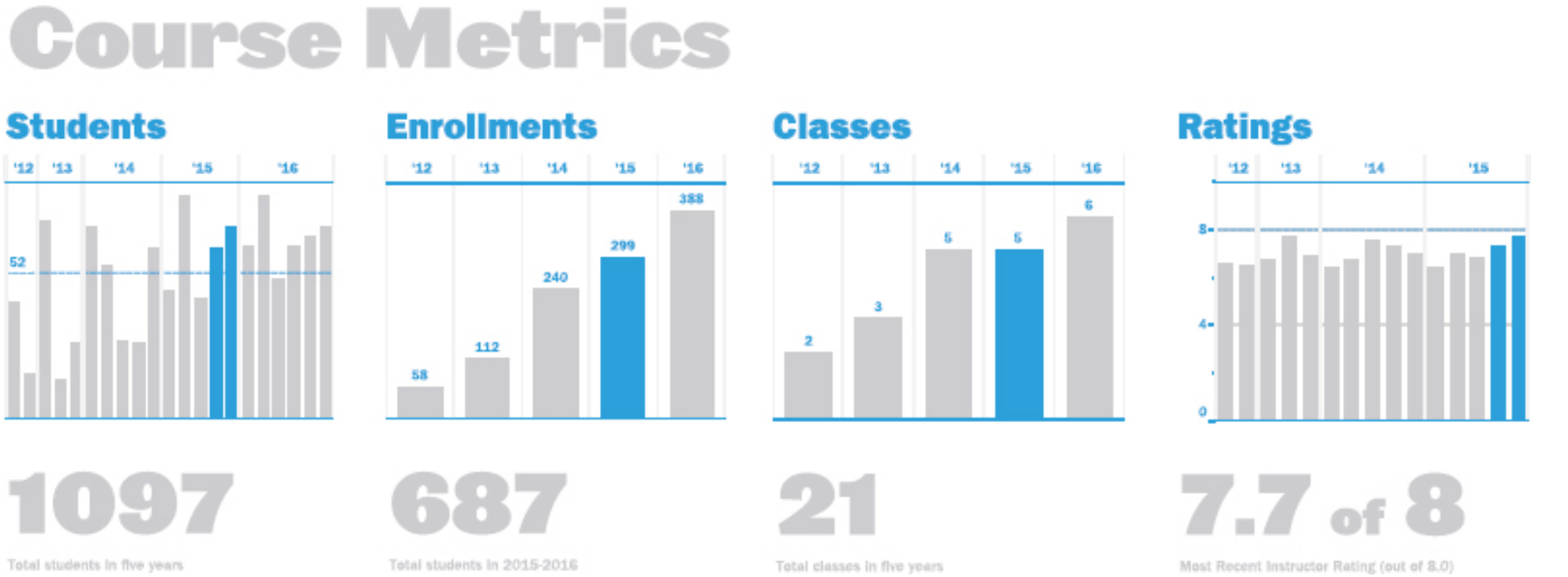
- Dot plot of course ratings comparing the Data Visualization course (blue) to the **BANA courses (dark gray)** and the college average (light gray line)



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Course Matrics

- This figure might look like with wider bars. This displays as a bar chart on the two inside charts and more of a lollipop-style chart on the two outside bar charts.
- Both styles encode the data using length and do not distort the data in any way, so thinner bars as a style choice for the overall design is used.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Why This Works

- **Easy-to-See Key Metrics**
  - This dashboard provides a quick overview across the top of key metrics about this course. It shows the number of classes each semester and by year, the number of students, and the overall course rating. It provides quick information at a glance but also allows for a deeper analysis by comparing the metrics over time.
- **Simple Color Scheme**
  - Only three colors are used in this dashboard: blue, light gray, and dark gray. Blue is used to highlight the two most recent rating periods. Since course feedback isn't available until weeks after a course is complete, registration for the next semester is already complete, and often a new course has begun. Therefore, blue highlights each portion of the chart that corresponds to the specified course rating period. In the first and last chart on the top row, the two specific courses are highlighted. In the second and third chart, the two courses are part of the five classes that are summarized in the year. In the dot plot, the blue represents those two courses versus the department and college averages.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

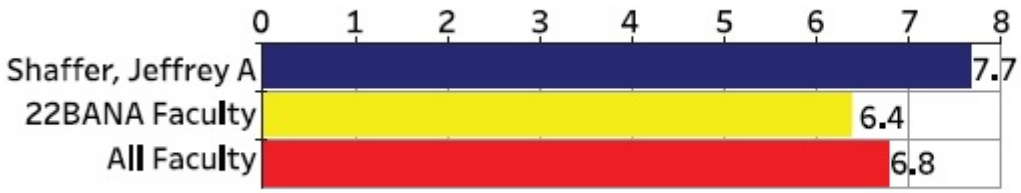
Why This Works

- **Potential to Be Static or Interactive**
  - This dashboard can work as a static dashboard that can be emailed as a PDF or printed, but it can also be interactive. The dashboard could connect to a database of all of the available courses. With a simple drop-down box with a list of courses, a department head or professor could easily select a course to generate this report.
- **Both Overview and Details Are Clear**
  - This dashboard offers both an overview and details. The overview shows four key metrics across the top along with those details over time. The section on the bottom offers a very detailed look at each survey question. It provides a quick comparison for each survey question rating the course compared to the department and the entire college.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

The Traditional Approach and Why You Should Avoid It

- The current system requires a distinct report to be run for each individual semester. There's no easy way to compare one period with the previous period or to see trends over time.
- The reporting template lists each survey question on a separate page along with the feedback. This does not allow for a quick review of the scores or a comparison across questions.



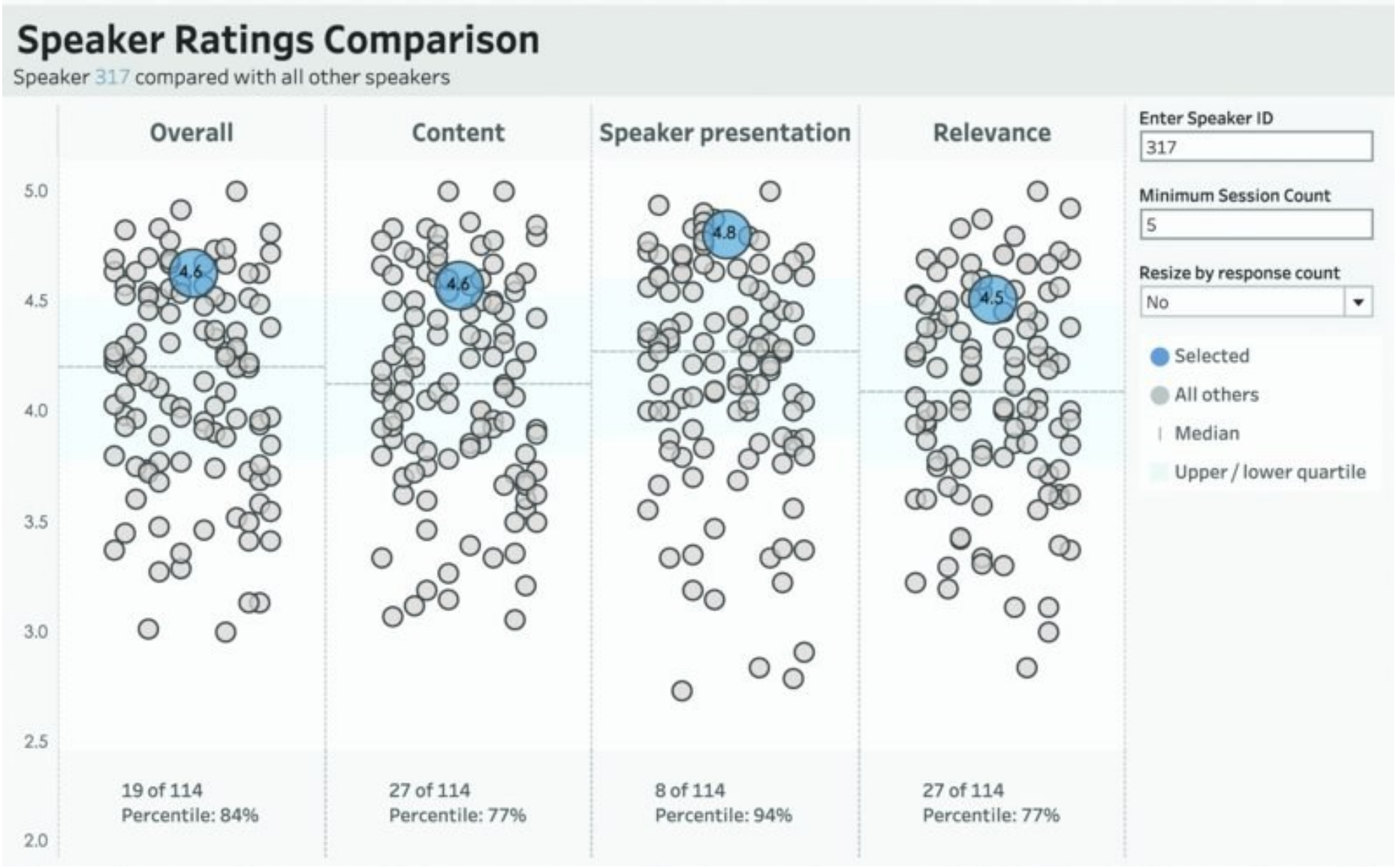
Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Speaker Rating Comparison

Speaker Rating Comparison

- **Big Picture**
  - Your company just had its annual three-day conference where hundreds of presenters educated, enlightened, and entertained over 10,000 participants. You know the conference was a big success, but you need to know which sessions were highly rated to begin planning for the next event.
  - This is not the first year the conference was offered. As in other years, the speakers themselves want to know if their sessions were well-received.
  - You create a dashboard that makes it easy for you, others in your organizations, and all the presenters to see how well or poorly their sessions were rated.
- **Specifics**
  - You hosted a conference and want to see how a certain speaker rated compared to the other speakers so you can plan for future events.
  - You need to fashion a dashboard that allows individual presenters to see how they performed at the event.
  - You need to make it easy to see how well or poorly a speaker performed compared to other presenters.
  - You want to be able to see how many people rated a session—that is, did dozens of people rate a session or only a handful?

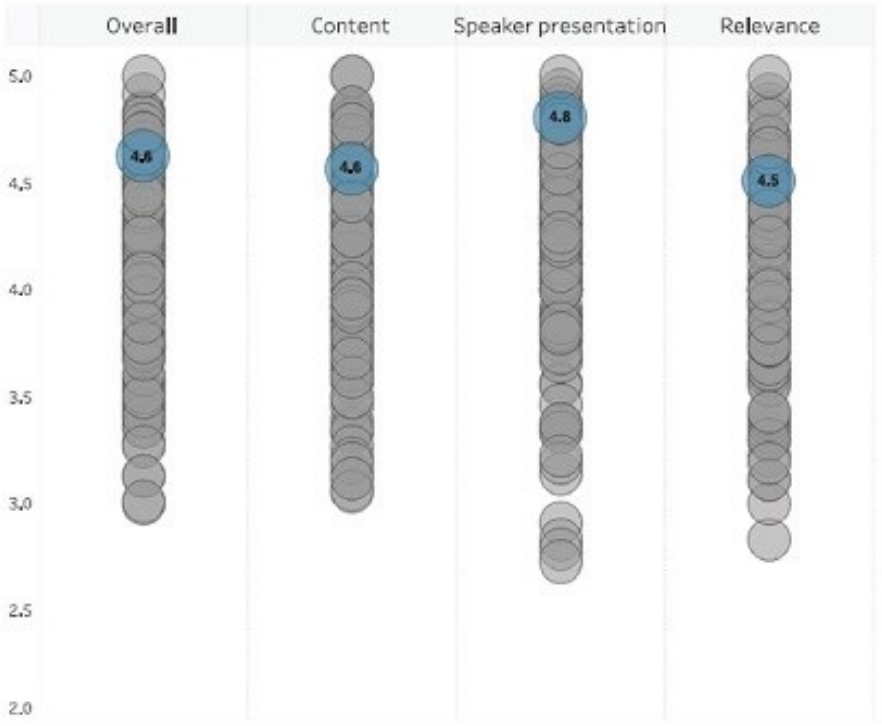
Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Dot plot here?

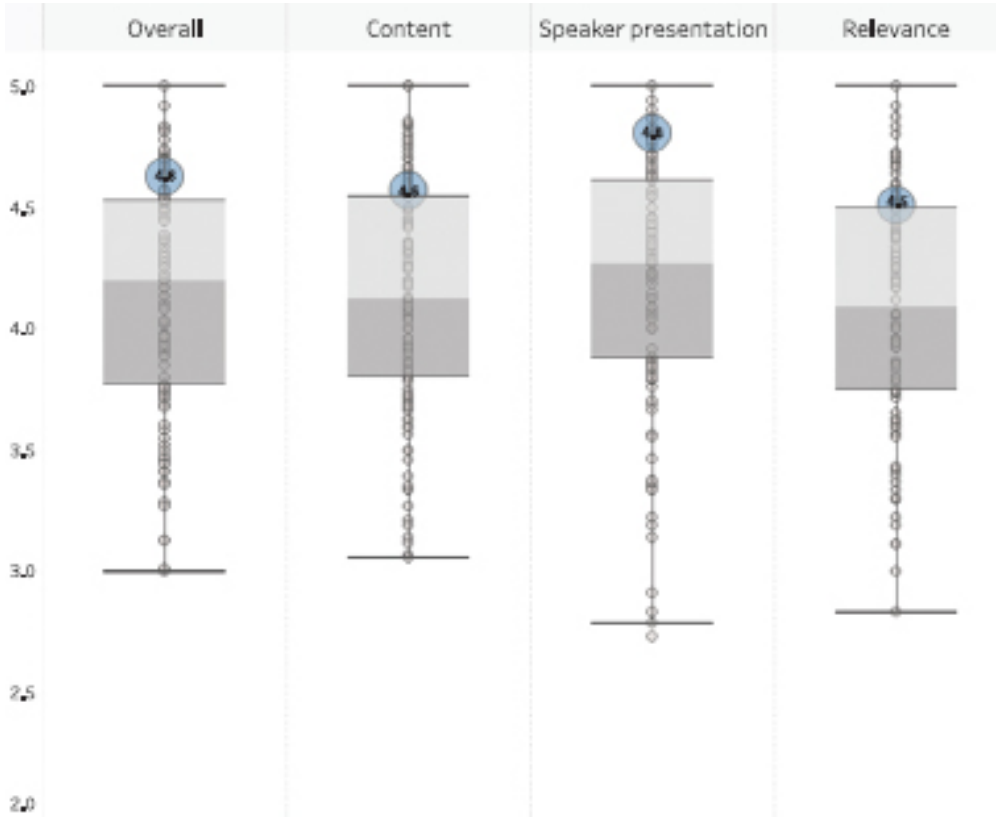
- Dot plot with all ratings overlapping. By placing an "ugly cat" here we don't mean to suggest that a dot plot is a bad chart type.
- It's just a bad choice here because there are so many overlapping dots.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Boxplot is fine

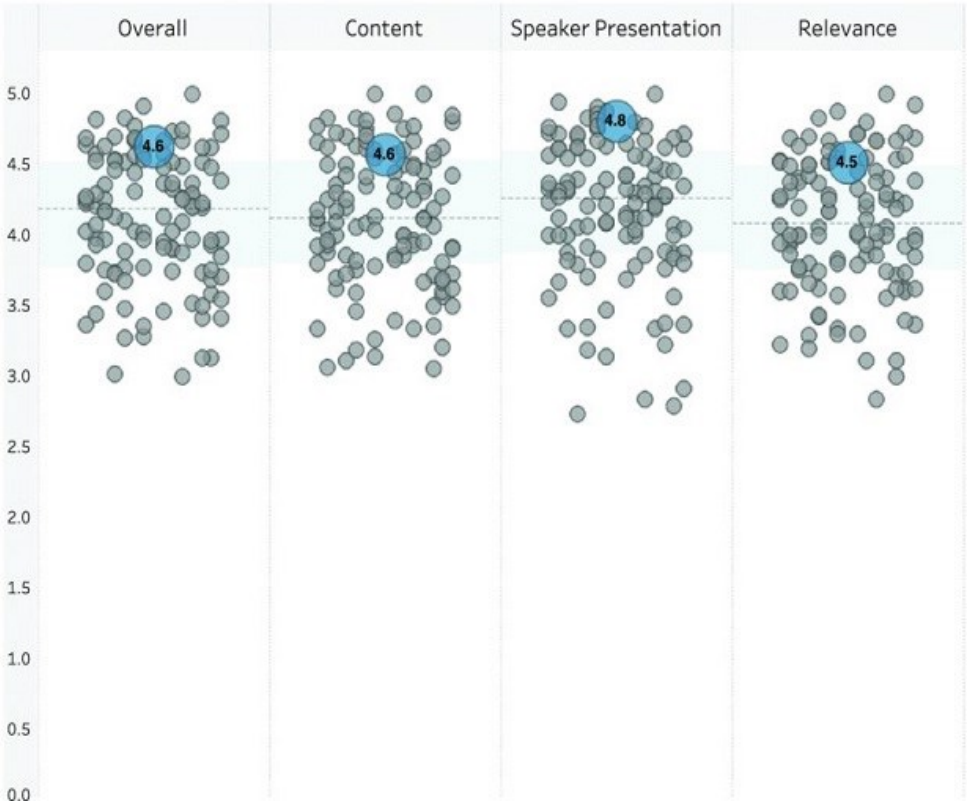
- Box and whisker plot showing quartiles, median, and outliers.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Doesn't the y-Axis Start at Zero?

- In this case, the lowest speaker score was 2.7 on a scale of 1 through 5 (which is pretty impressive, no?).
- Although we could indeed present the y-axis starting at zero, doing so would just make it harder to distinguish the differences between a relatively poor speaker and a relatively good speaker as the dots will be packed closer together.



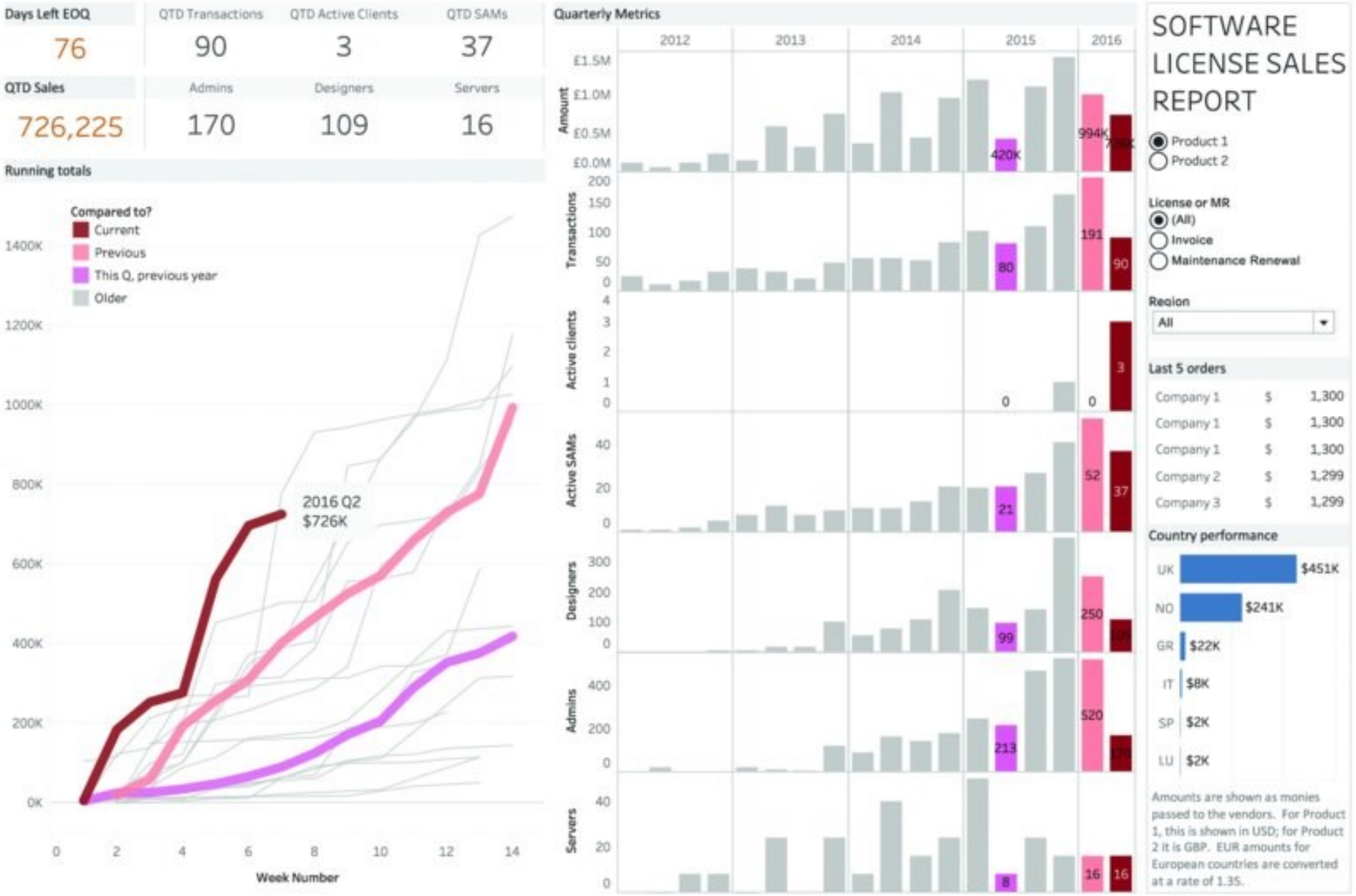
Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Executive Sales Dashboard

Executive Sales Dashboard

- **Big Picture**
  - You are a sales manager and want to know how you and the sales team have been performing this quarter. You want to be able to see, at any time during the quarter, exactly how total sales for that quarter compare to any previous quarter. You need to be able to see the overview for the whole business and also filter down to individual product lines or regions.
- **Specifics**
  - How are we doing this quarter?
  - How is this quarter compared to last quarter and the same quarter last year?
  - Are we on track to beat the previous quarter? Are we on track to beat the same quarter last year?
  - What are the most recent transactions?

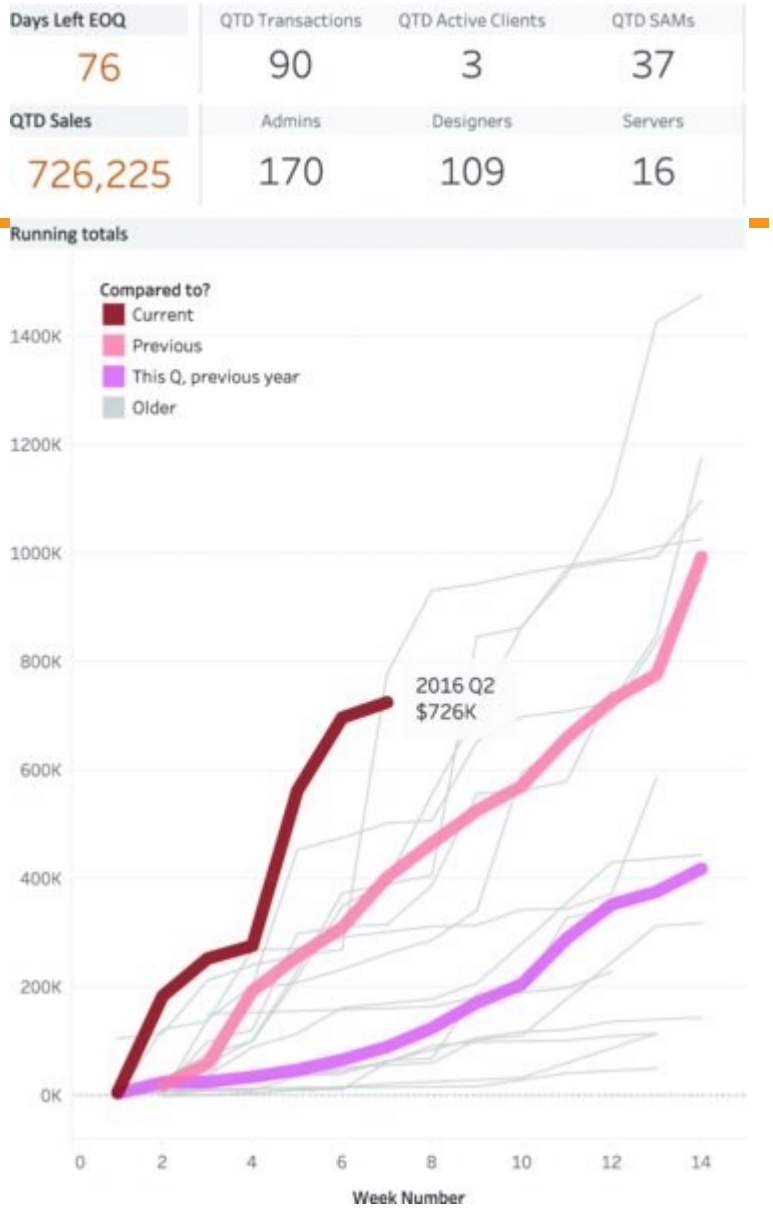
Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Growth Comparisons

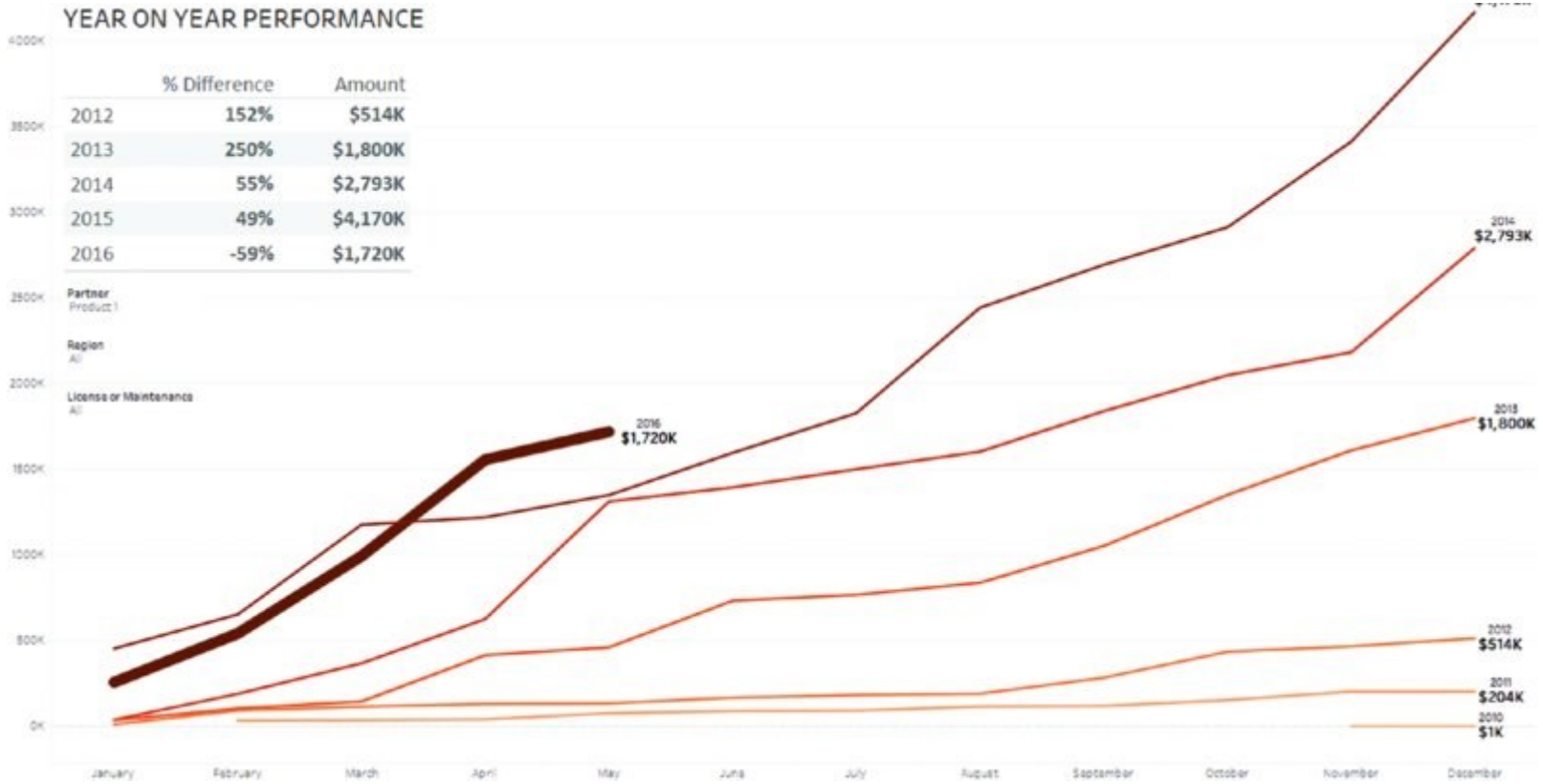
- Focus on Year-over-Year and Quarter-over-Quarter Growth Comparisons.
- In this organization, the most important things to know are: what's happening now, what happened last quarter, and what happened in this quarter last year. This is vital in a sales organization focusing on growth.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Year-over-year performance

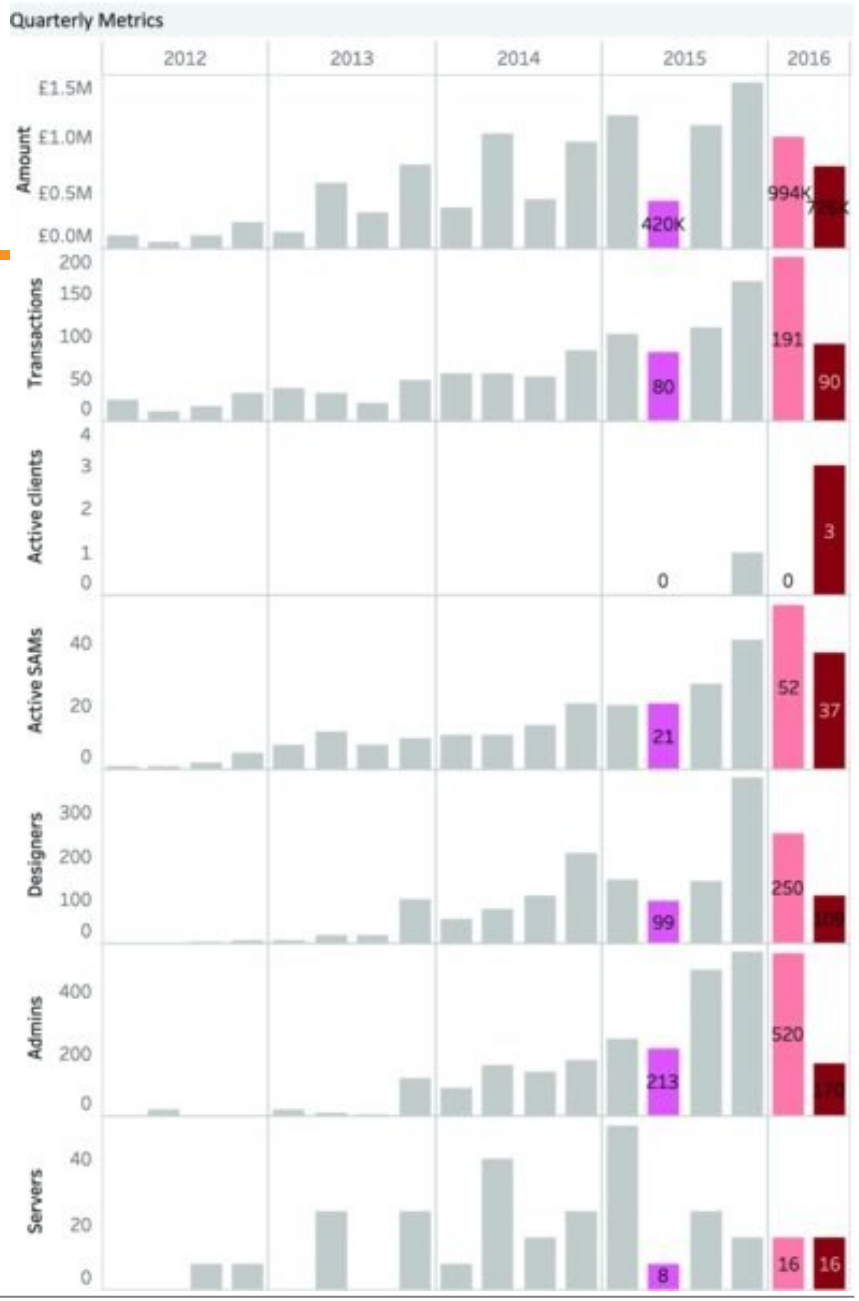
- Visualizing Time, index charts are superb ways to track measures and compare them against different periods



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Growth over time

- the bar chart shows the actual values for each quarter's sales.
- Each row of bars represents one of the seven metrics this business focuses on. Only the labels for the three important bars (this quarter, last quarter, same quarter last year) are shown.
- This reduces clutter on the dashboard while users retain the ability to look up the most important numbers without relying on interactivity.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

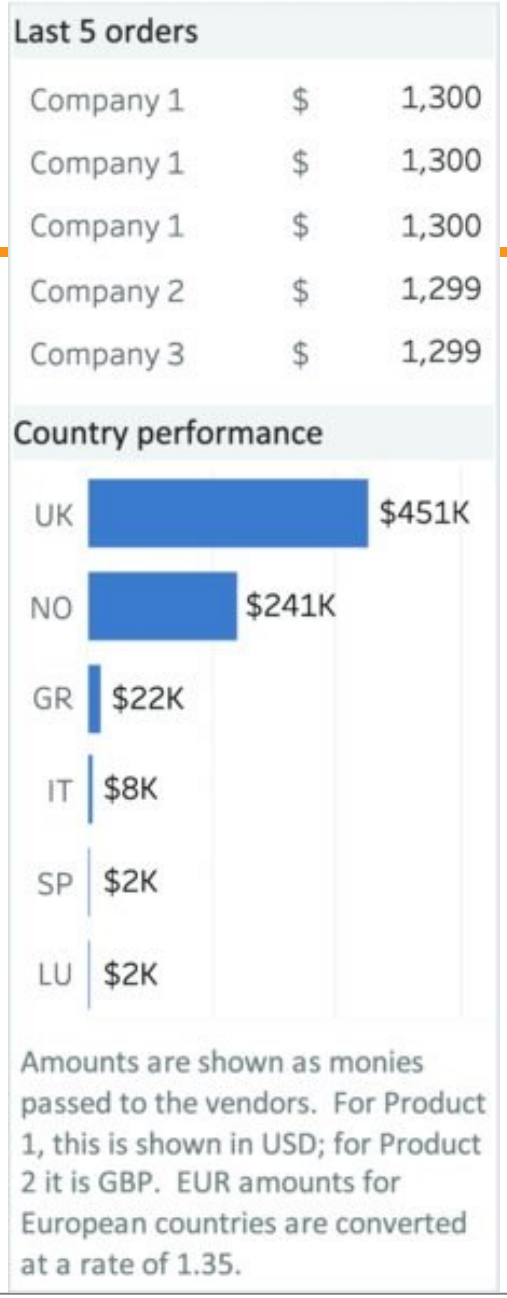
Key Metrics as Text

- At a glance, executives can look to the most prominent part of the dashboard and find exactly the numbers they need to see.
- It can be important to distill the information down to just a few key numbers. If that's the case, just show the numbers. You can show detailed charts elsewhere.
- When looking at the dashboard at the start of a week or ahead of a meeting with a sales team, you can check just these numbers for the main point.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Small Details Where Needed

- This dashboard is dominated by the key metrics, the index chart and the detail bar chart. But the space below the filters is just enough to add extra information needed for a full business snapshot. In the chart, the last five orders are highlighted, as is a geographic snapshot of the countries this company sells to.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Hospitality Dashboard for Hotel Management

Scenario

- **Big Picture**
  - You are a manager for a large chain of hotels. You need a dashboard to monitor some key metrics, including how revenue did last week in comparison to previous weeks. You need to see if the revenue is being influenced by the current daily rate promotions and if the promotions are aligned to occupancy. There are multiple locations around the world, so you also need to know which hotels are the best and worst performers within their respective region. You also need the ability to drill further into the details for any specific weeks as necessary.
- **Specifics**
  - You are tasked with showing the **revenue per available room (RevPAR)** for your hotels and in comparison to a **peer group**.
  - You need to organize hotels by regions and compare those to hotels within the same region.
  - You need to show data over time, showing a selected number of weeks in the past.
  - You want to see the spread for available rooms from the lowest-priced room and the highest-priced room.
  - You need to understand how quickly the business is reacting to these prices and how much alignment exists between the different locations.
  - You need to provide an on-demand (interactive) breakdown to the daily rate by segment and channel.
  - You need to provide on-demand (interactive) access to a specific hotel.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

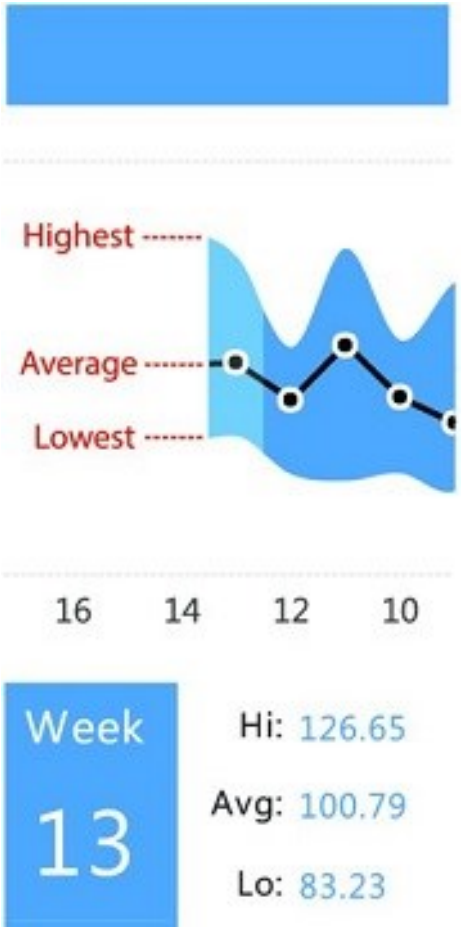
Revenue/Room in Week 13

- Highlighting week 13 in the top chart updates the key metrics on the other charts.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Chart

- The spread from the lowest-priced to the highest-priced available room is shown by the blue band.
- The black line shows the average.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Chart

- The highlighted week is shown as a **light blue band**.
- The data in the text boxes will update showing the details for the highlighted week.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

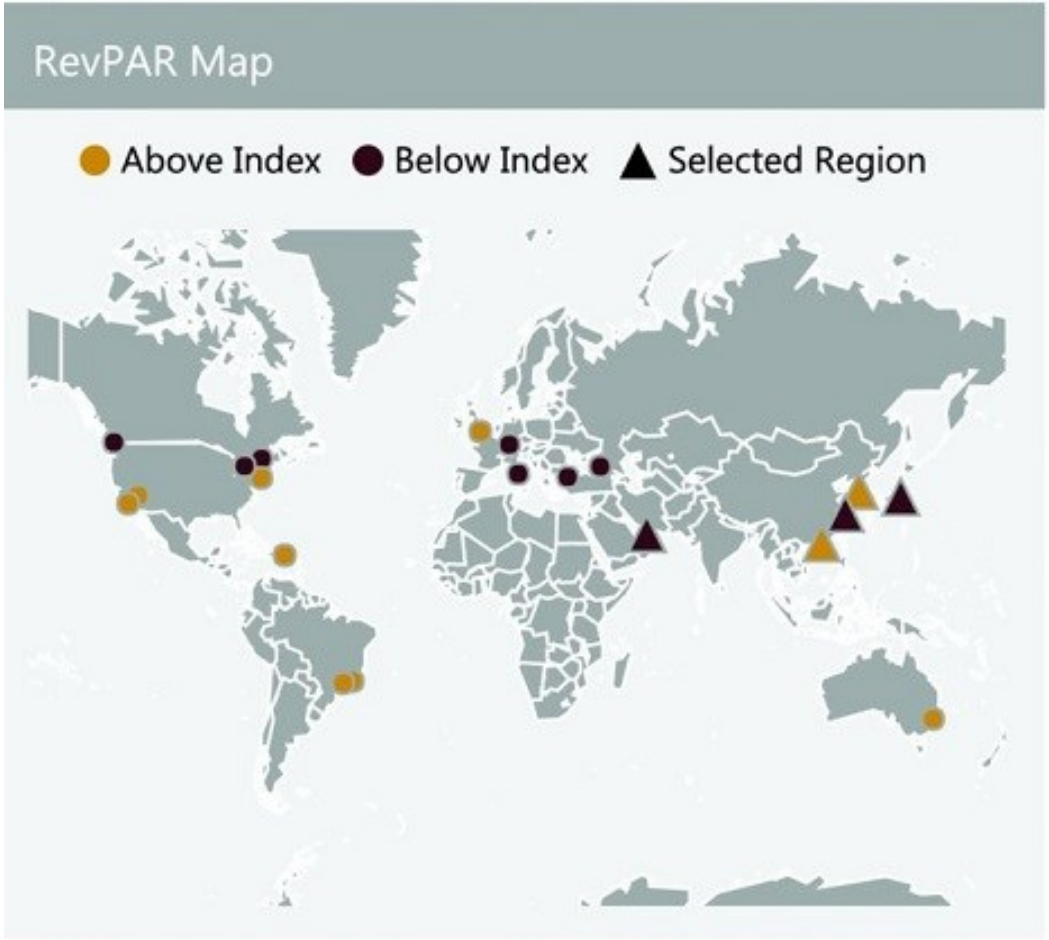
Regions

- Region selector showing Asia selected.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Map

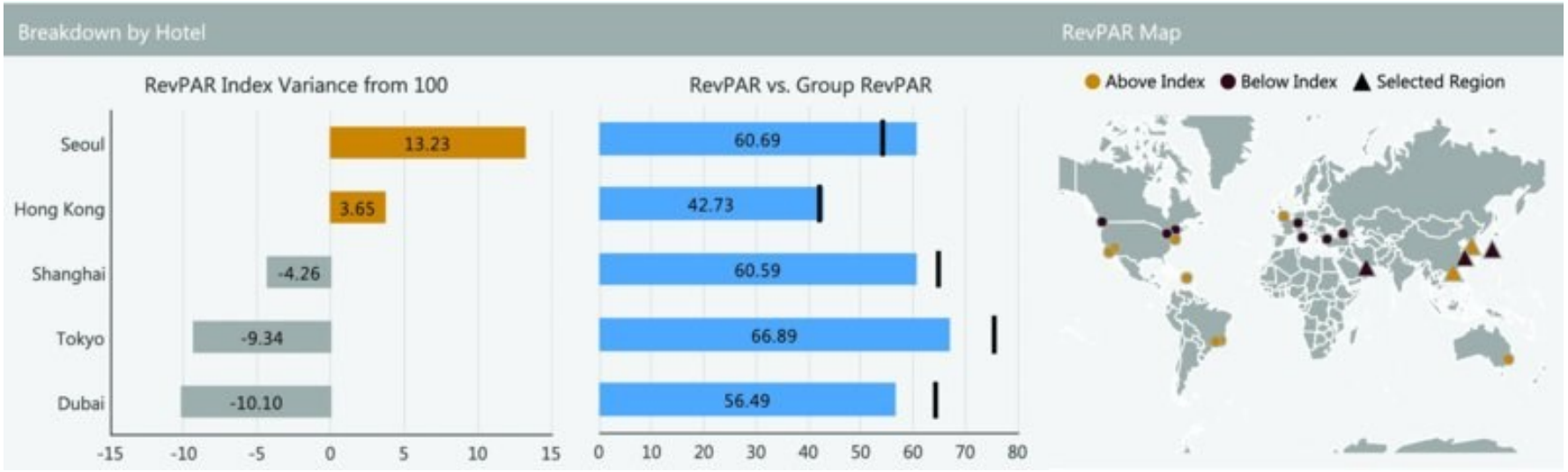
- Map showing the selected region with triangles.
- All points are colored as above or below index.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Breakdown by Hotel

- Bar charts on the bottom of the dashboard display a breakdown by hotel. The blue bars show the RevPAR per hotel. The black line is the Group RevPAR.



Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Slide Control

- The primary view of this dashboard focuses on the current time period and looking back in weeks for some selected period of time. In the top chart, the x-axis of the line chart shows the range of weeks that has been selected. This allows the user to gauge the overall trend as well as the spread of revenue per room within that week.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Why This Works

- **Interactivity That Preserves Context**
  - The dashboard is highly interactive. Nearly every object on the dashboard acts as a highlight or a filter or presents details on demand.
  - When drilling down into more detailed data, the visualizations remain connected to the main dashboard.
  - The portions of the dashboard that are used as selectors are still in view and actionable. This allows the user to immediately change the selection to update those details while still in the details window.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Why This Works

- Clear Focus on Key Measures and Dimensions
  - The dashboard shows only three measures—RevPAR, daily rate, and occupancy—and provides the measures across three dimensions—week, region, and hotel. More dimensions and details are available as details on demand, but the clear focus on these metrics helps the user analyze and understand the data without complication.

Ref: • Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. The big book of dashboards: visualizing your data using real-world business scenarios. John Wiley & Sons, 2017.

Summary

Summary

- We have learned types of data:

  Quantitative, Ordinal, and Categorical data.

- We have learned useful visual encodings:

  Size, Position, and Color.

- We have learned how to use visual encodings for different types of data.

Make your audiences understand
your visualization QUICKLY !!!

“

One Picture Is Worth
Ten Thousand Words.

”

Old Chinese Proverb