Data Visualization and Storytelling

TABLEAU

Data Visualization and storytelling

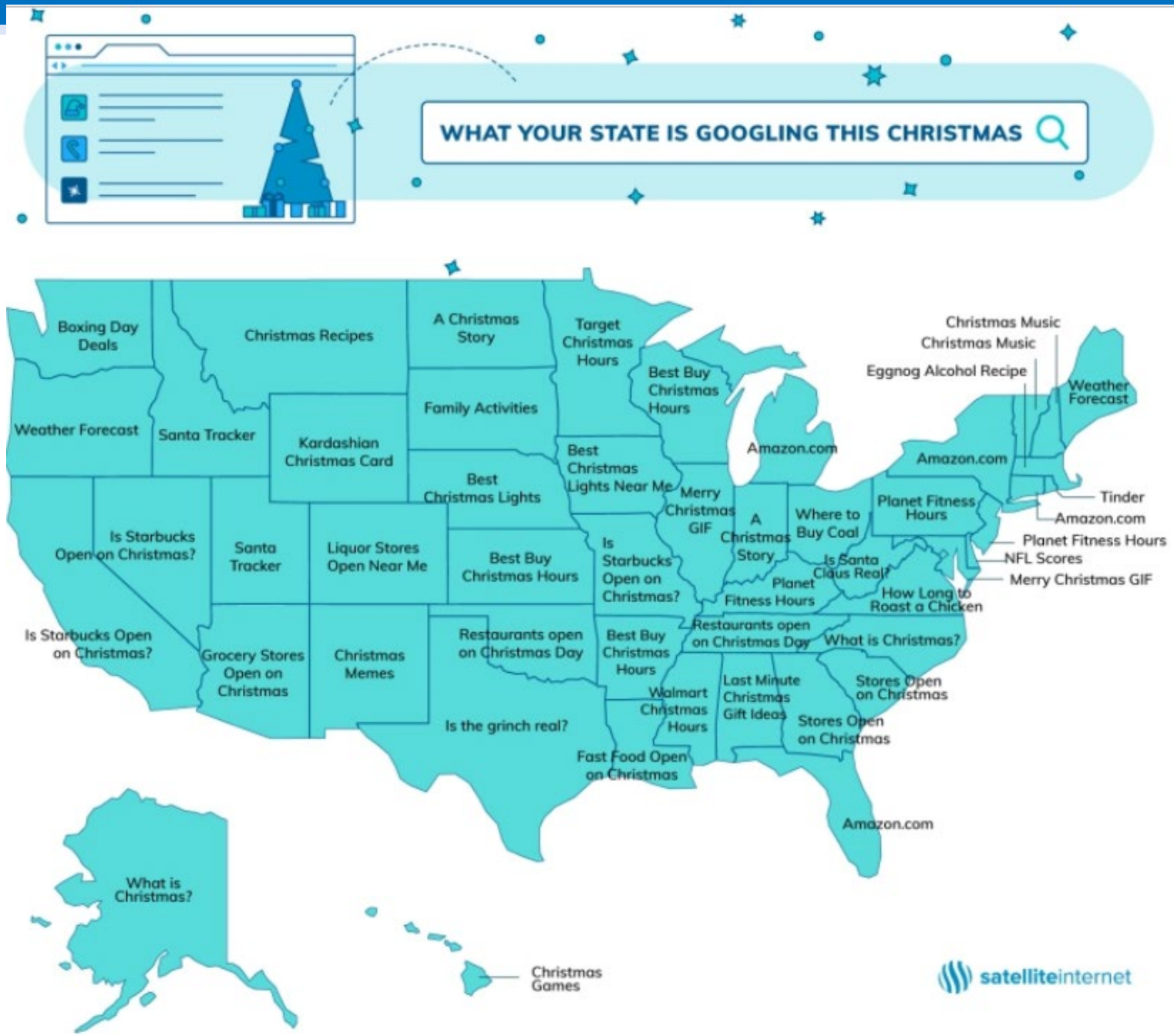
Data Visualization

- Visualization is an important step in presenting/motivating business ideas and solutions.
- Tableau Desktop, Microsoft Power BI, SPSS, Excel, and other software can generate many different kinds of charts.
- While technology has increased access to and proficiency in tools to work with data, there remain gaps in capabilities. Your tools don't know what that story is
- A Picture Is Worth a Thousand Words.
- Being able to visualize data and tell stories with it is key to turning it into *information* that can be used to drive better decision making.

Data Visualization

- A Picture Is Worth a Thousand Words

https://www.youtube.com/watch?v=jbkSRLYSojo

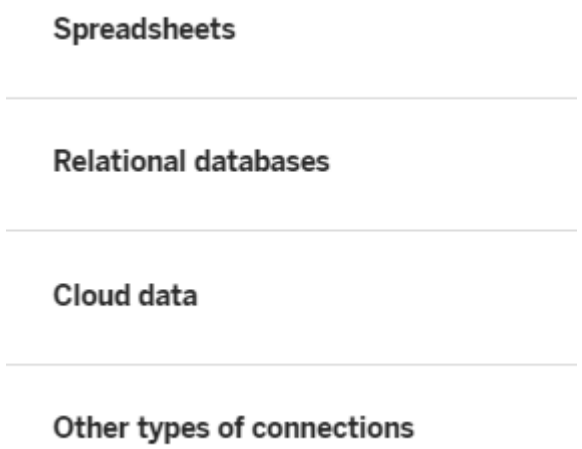


Source:
https://www.satelliteinternet.com/resources/what-your-state-googles-on-christmas/

BBC FOUR
lifespan
75 years
50 years
25 years
927
income
$400
$4 000
$40 000

What is a data set?

A data set (sometimes referred to as data source, or database) in the context of Tableau, contains the data used to build visualizations. Every bar chart, scatter plot, or line chart you see in Tableau has a connected database or spreadsheet that supplies the data.
You can learn more about different data source types below.



Why use data visualizations?

- **Visualizations help us understand complex data**
- The best reason to use a visualization to understand your data is that most data sets are far too large to consume in their raw format. Humans are limited in what information we can process and compare in our heads, especially if that information resides in a million row data set, but we are good at quickly processing visual information.
- Even if you are new to reading data in a chart, you already have the built-in capabilities to spot light and dark colors, large and small shapes, groups and orientations of objects. These are referred to as pre-attentive attributes.

Visual analytics

- Visual analytics leverages our *pre-attentive attributes* – visual cues humans process automatically with sensory memory.

- We can notice and interpret these kinds of attributes quickly and without special effort.

The Tableau Universe

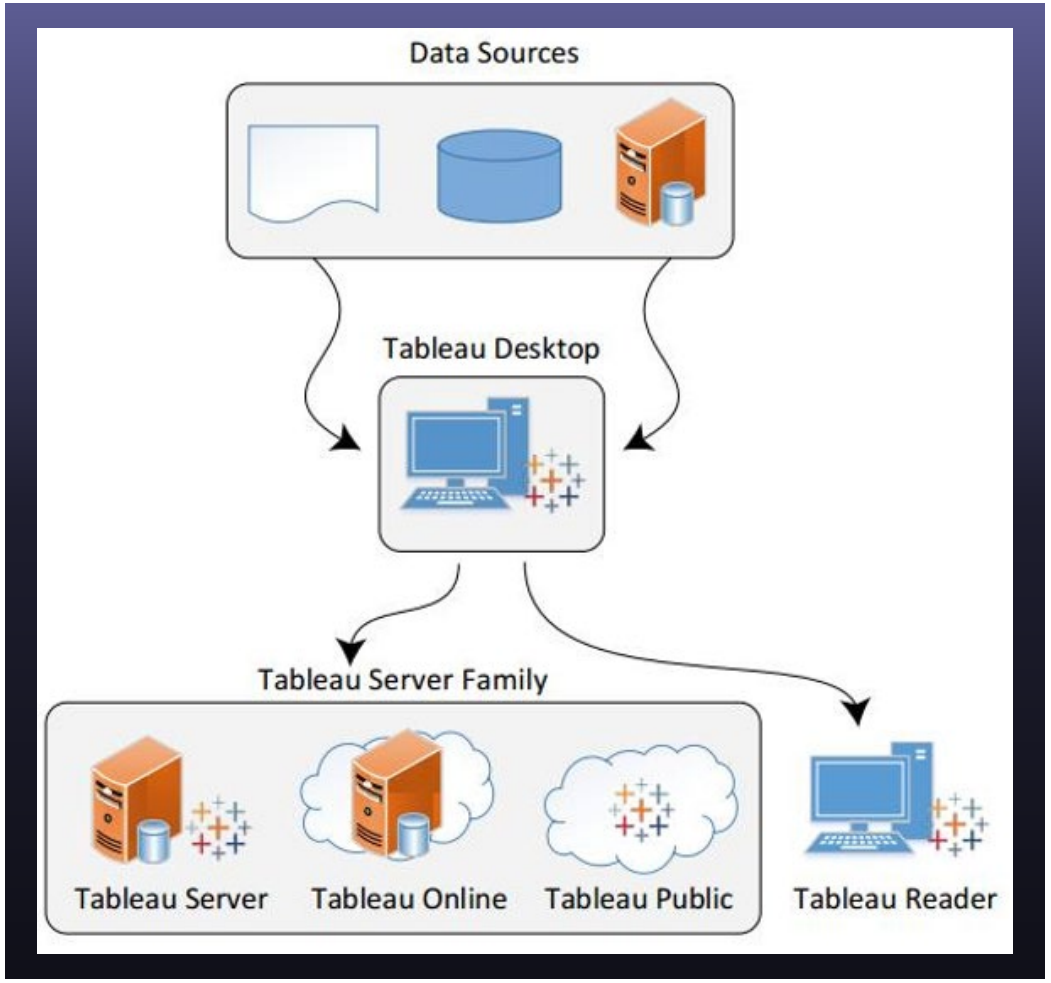


Tableau Desktop: where visualizations are created

Tableau Server: secure web based environment where end users can access visualizations

Tableau Online: cloud based version of Tableau Server hosted by Tableau Software

The Tableau Universe

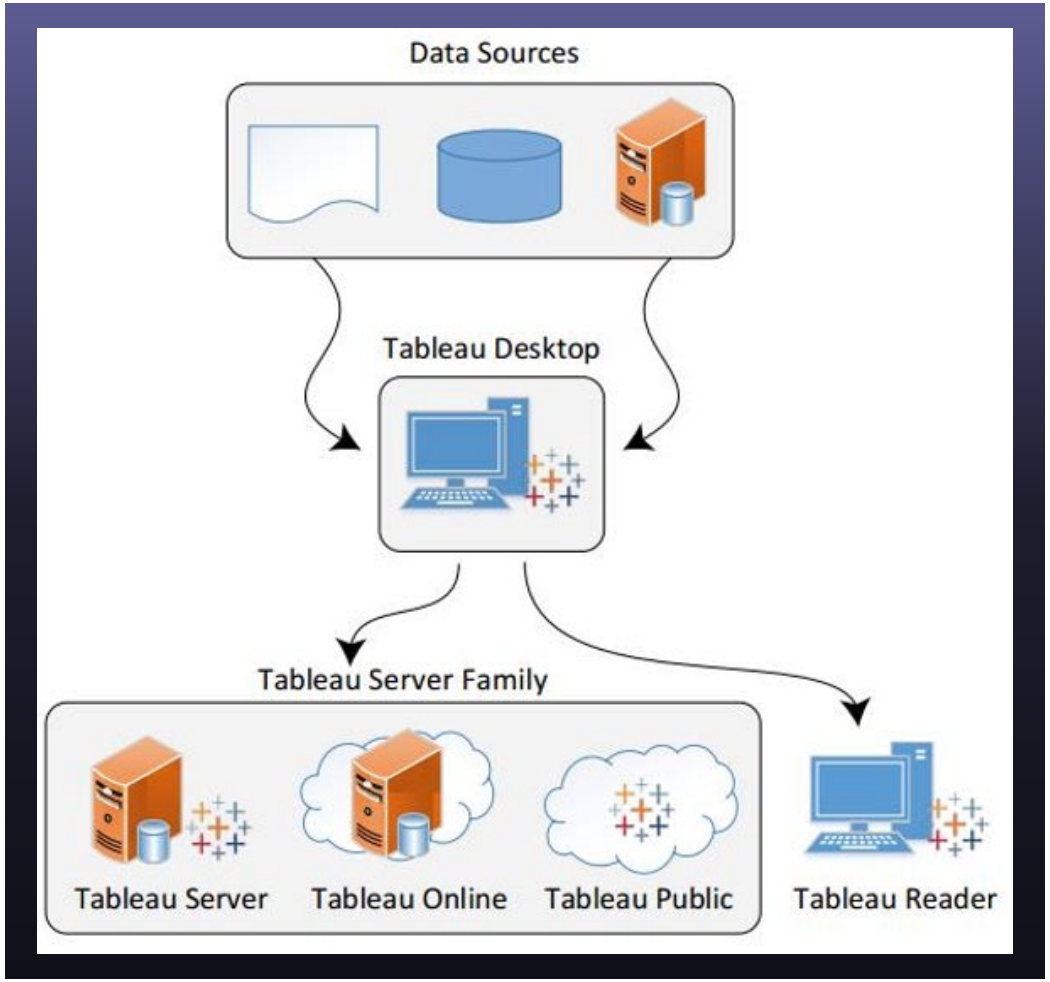


Tableau Public
1. Public Client
2. Public-facing Tableau Server

Tableau Client : like Tableau desktop but
- Data sources are limited to excel, access and text files
- You can only publish to tableau Public
- Limited to 15 million rows

Tableau Server (**available to the public**): secure web based environment where end uses can access visualizations

The Tableau Universe

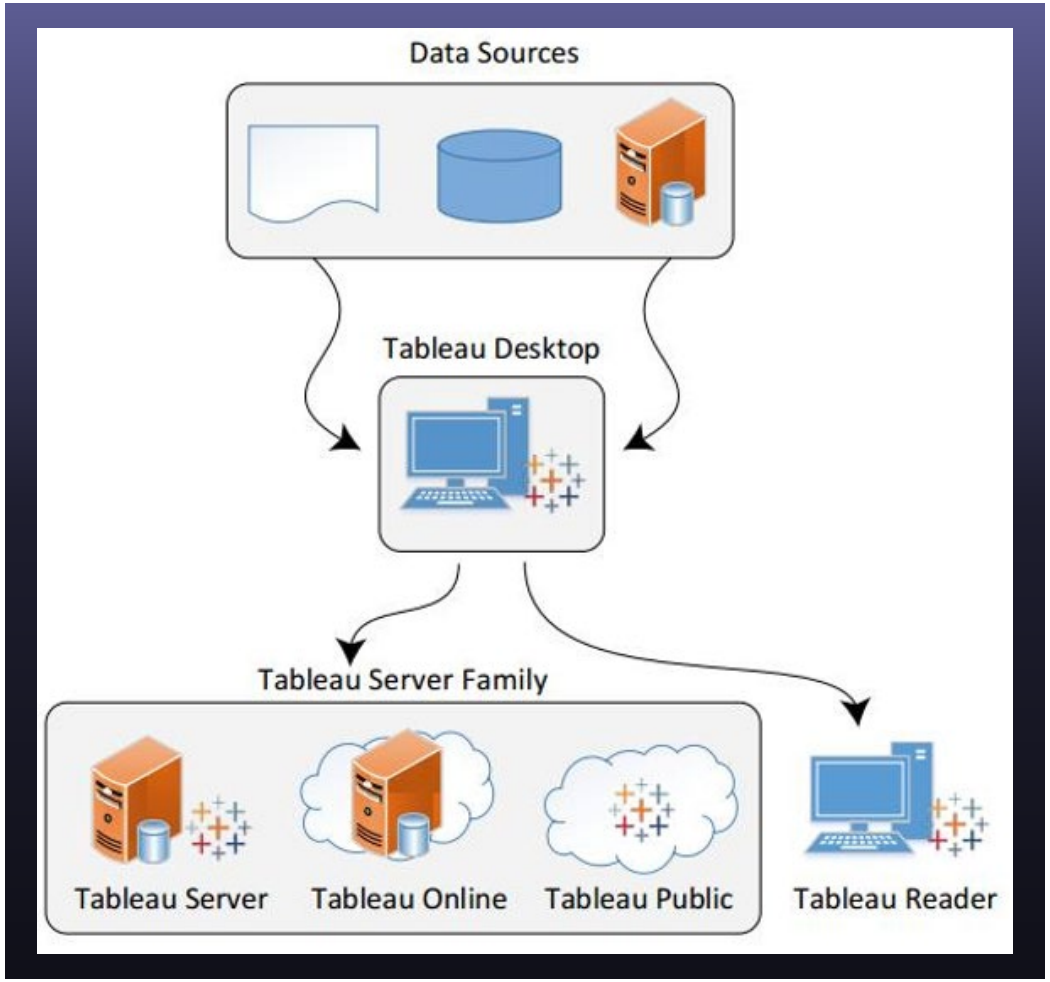


Tableau Reader: used for viewing tableau workbook packages

FREE

Tableau Reader :

- Authoring is not enabled
- Filters, animation, dashboard actions are accessible
- Useful for presentations

Integration: PYTHON, R, AND MATLAB

THE Tableau Ecosystem

Tableau Desktop Tableau Desktop is an application for Windows and Mac, appreciated by both analysts and business users.
You can connect to flat files (such as Excel and CSV files) and save your workbooks to your local hard drive. To tap into an organization's IT infrastructure, you can also use Tableau Desktop to connect to a host of different database solutions, and you can share your workbooks via Tableau Server or the cloud-based Tableau Online.

Tableau Prep Tableau Prep is the latest addition to the Tableau product suite and is designed to help you prepare your data before you analyze it in Tableau Desktop. The visual interface allows you to quickly merge differently formatted datasets, clean the data, and unify the level of aggregation. Tableau Prep fits seamlessly into your analysis workflow.

Tableau Server Tableau Server is a platform for data analysis and is used by small family-run businesses and large Fortune 500 companies alike. It is intended for the organization-wide provision of data visualizations and dashboards that can be viewed in a browser and are frequently embedded into the organization's intranet.

Tableau Online Tableau Online is a Tableau-hosted solution for storing and deploying dashboards. It provides similar functionality to Tableau Server but is a cloud-based service. No purchase and maintenance of server hardware is necessary here.

Tableau Public Tableau Public is a hosting service for the publication of data visualizations to the web. It is used by newsrooms and bloggers but also by companies, research institutes, governmental bodies, and non-governmental organizations that aim to get their data stories into the public eye. The interactive visualizations can be viewed in the browser directly on the Tableau Public platform, or they can be embedded into blogs and websites.

Tableau Reader Tableau Reader is a free desktop application that allows you to open and interact with Tableau workbook files that have been created in Tableau Desktop. However, it is not possible to make any changes to the visualizations in Tableau Reader.

Fee: $100

Suggested Training: Tableau Desktop Specialist Exam Readiness

Required Prerequisites: None

Suggested Product Experience: 3+ Months

Product Version: Currently testing on 2020.1

Title Valid For: No Expiration

Exam Prep Guide: Download Here

REGISTER NOW

Skills Measured

Connecting to & Preparing Data

- Create and save data connections
  - Create a live connection to a data source
  - Explain the differences between using live connections versus extracts
  - Create an extract
  - Save metadata properties in a . TDS
- Modify data connections
  - Add a join
  - Add a blend
  - Add a union
- Manage data properties
  - Rename a data field
  - Assign an alias to a data value
  - Assign a geographic role to a data field
  - Change data type for a data field (number, date, string, Boolean, etc.)
  - Change default properties for a data field (number format, aggregation, color, date format, etc.)

Exploring & Analyzing Data

- Create basic charts
  - Create a bar chart
  - Create a line chart
  - Create a scatterplot
  - Create a map using geographic data
  - Create a combined axis chart
  - Create a dual axis chart
  - Create a stacked bar
  - Create a chart to show specific values (crosstab, highlight table)
- Organize data and apply filters
  - Create a visual group
  - Create a group using labels
  - Create a set
  - Organize dimensions into a hierarchy
  - Add a filter to the view
  - Add a context filter
  - Add a date filter
- Apply analytics to a worksheet
  - Add a manual or a computed sort
  - Add a reference line or trend line
  - Use a table calculation
  - Use bins and histograms
  - Create a calculated field (e.g. string, date, simple arithmetic)
  - Add a parameter

Sharing Insights

- Format view for presentation
  - Use color
  - Use bolding
  - Use shapes
  - Use viz animations
  - Change size of marks
  - Select fonts
- Create and modify a dashboard
  - Create a dashboard layout
  - Add interactive or explanatory elements
  - Add dashboard actions
  - Modify existing dashboard layout for mobile devices
  - Create a story using dashboards or views
  - Share a twbx as a PDF
  - Share a twbx as an image

Continued on next page.

Skills Measured

Understanding Tableau Concepts

- Dimensions and measures
  - Explain what kind of information dimensions usually contain
  - Explain what kind of information measures usually contain
- Discrete and continuous fields
  - Explain how discrete fields are displayed in Tableau
  - Explain how continuous fields are displayed in Tableau
  - Explain the difference between discrete date parts and continuous date values in Tableau
- Aggregation
  - Explain why Tableau aggregates measures
  - Describe how an aggregated measure changes when dimensions are added to the view

Timeliness

Completing a task effectively and efficiently has become a standard that organizations expect from employees. This exam is timed because we view time as a critical competency needed to be successful.

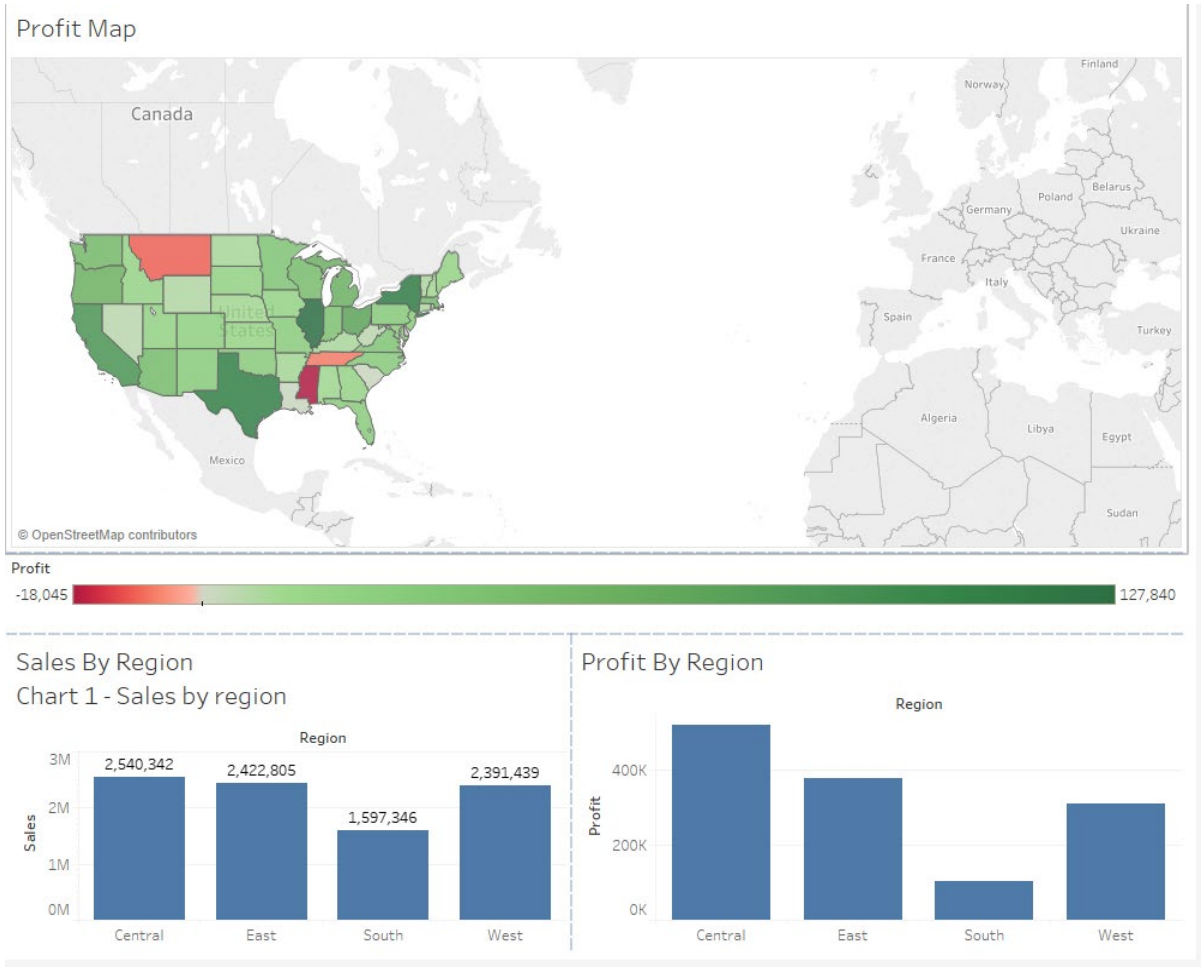
Connect

Analyze

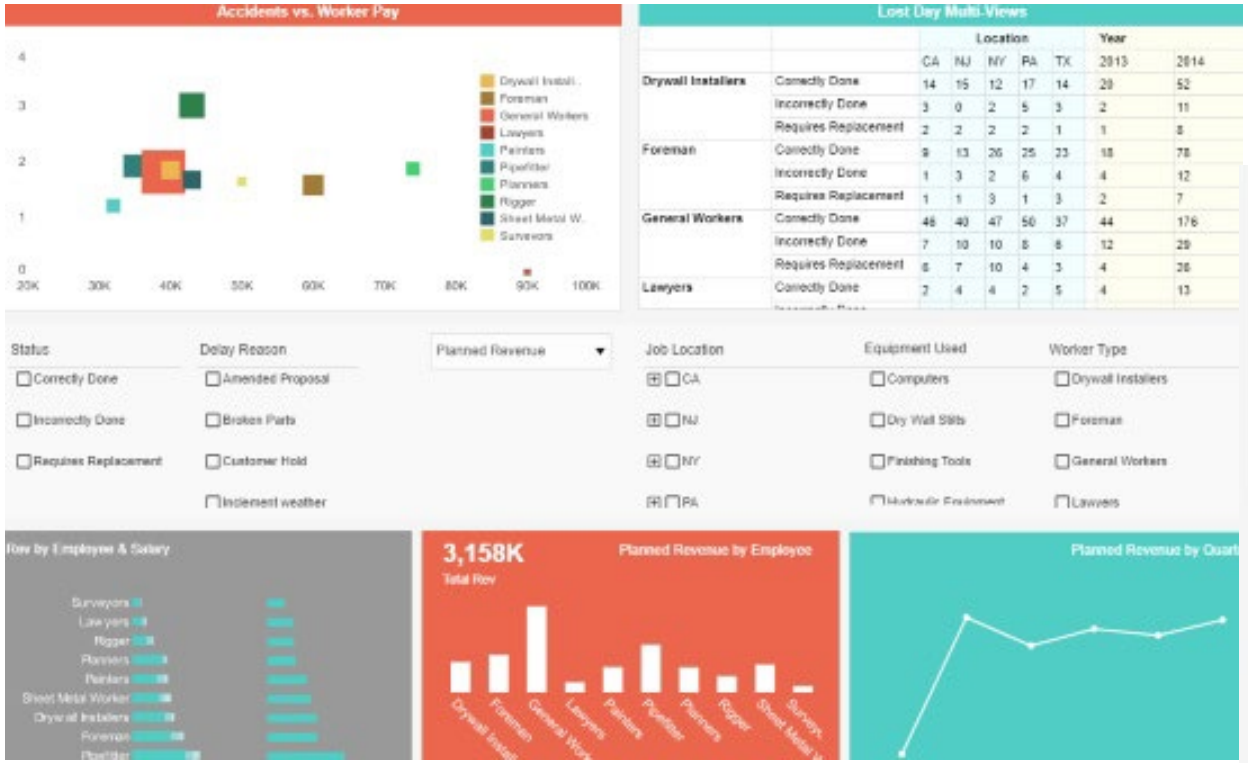
Share

Data Dashboards

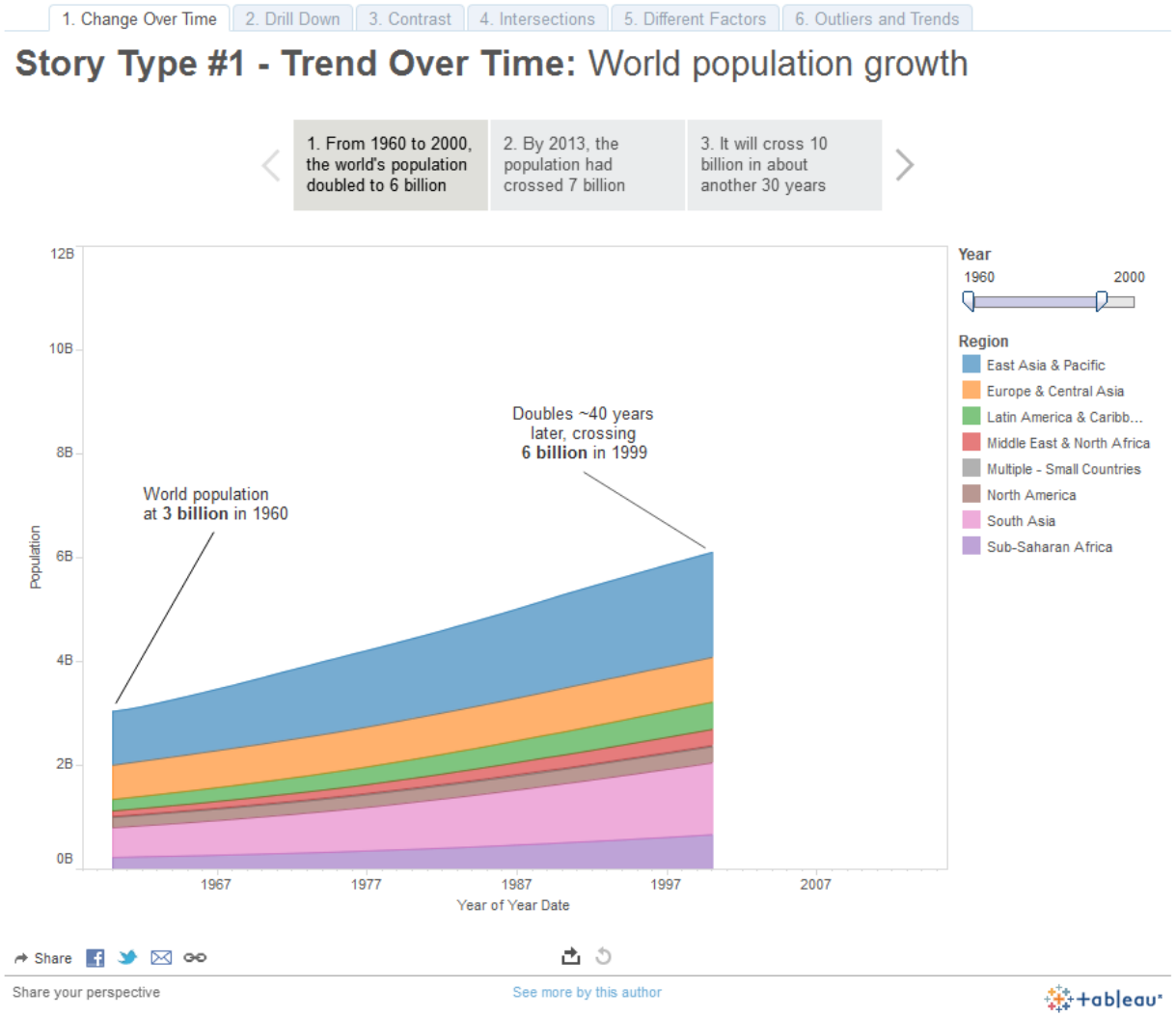
- Data dashboards combine several visualizations and are often updated in real time



Data Dashboards



Story


1. Change Over Time
2. Drill Down
3. Contrast
4. Intersections
5. Different Factors
6. Outliers and Trends
Story Type #1 - Trend Over Time: World population growth
1. From 1960 to 2000, the world's population doubled to 6 billion
2. By 2013, the population had crossed 7 billion
3. It will cross 10 billion in about another 30 years
12B
10B
8B
6B
4B
2B
0B
Population
World population at 3 billion in 1960
Doubles ~40 years later, crossing 6 billion in 1999
1967
1977
1987
1997
2007
Year of Year Date
Year
1960
2000
Region
East Asia & Pacific
Europe & Central Asia
Latin America & Caribb...
Middle East & North Africa
Multiple - Small Countries
North America
South Asia
Sub-Saharan Africa
Share
Share your perspective
See more by this author
+ableau

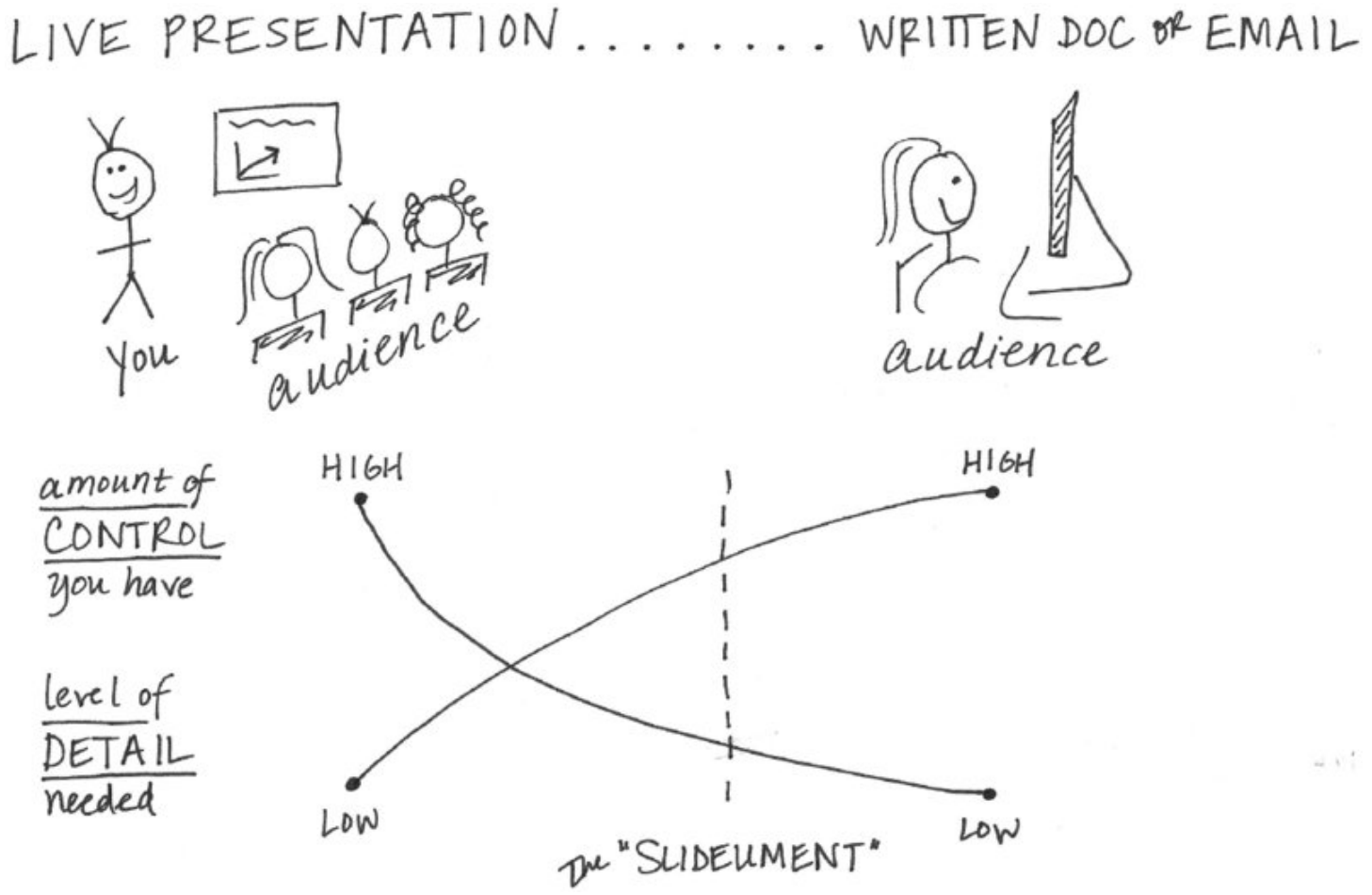
Viewer

- A **Viewer** is primarily a consumer of content created by others. They may simply view existing workbooks and dashboards, or perform basic interactions such as filtering or selecting items of interest.

- Viewers can:
- See published and custom views others have created
- Explore the data in a view using filters and legends, sorting, and tooltips
- Share, comment on, and download content (options vary based on permissions)

Know your audience

- Be specific
  - Sometimes this means creating different communications for different audiences
- Mechanism
  - Control vs Details



Know your purpose

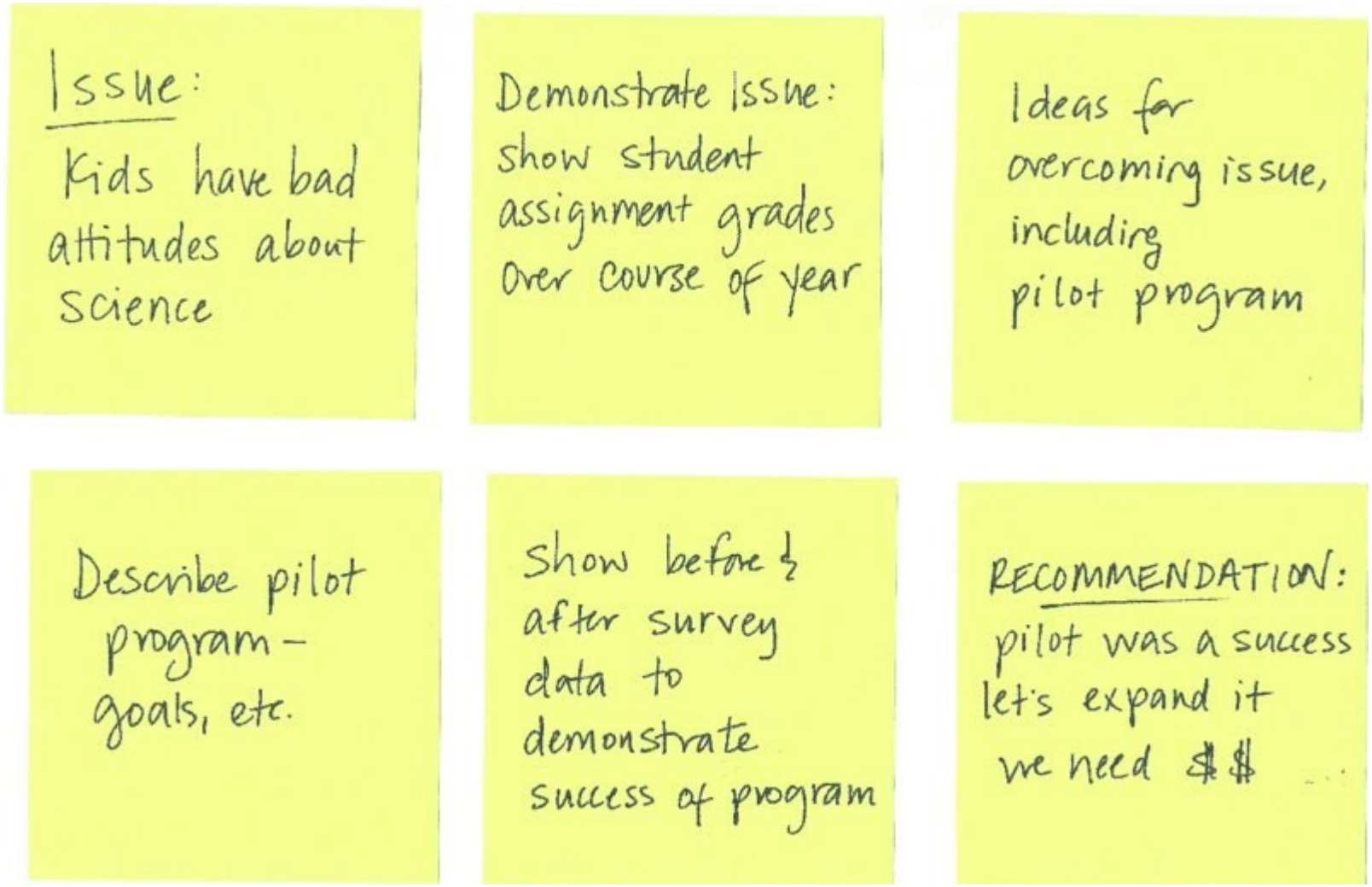
An executive audience might be interested in:	An analyst audience might prefer:
Percent of totals or year-over-year percentages	Histograms showing distribution
Detailed maps	Highlight tables with data
Icons	Pareto charts
Key performance indicators (KPIs)	Box and whisker plots
Summary views, monthly reports, high-level figures	Having control, the ability to drill down into the data

Questions to ask

- What background information is relevant or essential?
- Who is the audience or decision maker? What do we know about them?
- What biases does our audience have that might make them supportive of or resistant to our message?
- What data is available that would strengthen our case? Is our audience familiar with this data, or is it new?
- Where are the risks: what factors could weaken our case and do we need to proactively address them?
- What would a successful outcome look like?
- If you only had a limited amount of time or a single sentence to tell your audience what they need to know, what would you say?

What data is available that will help make my point?

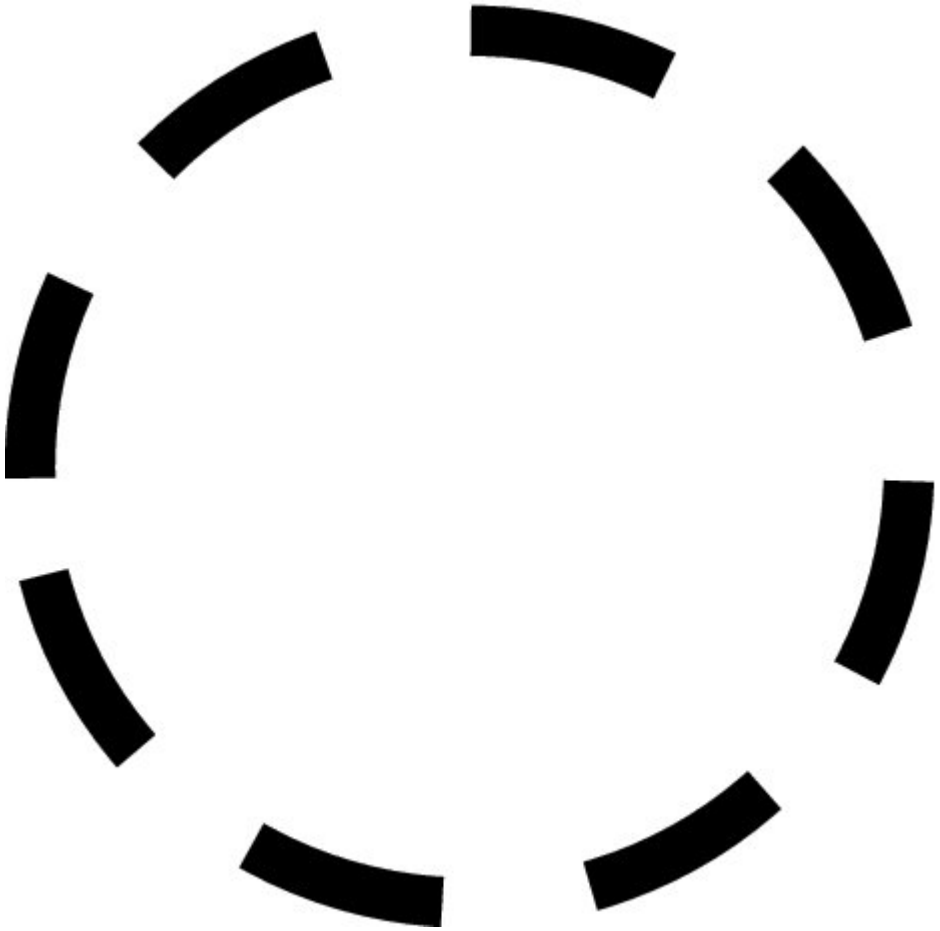
Storyboarding



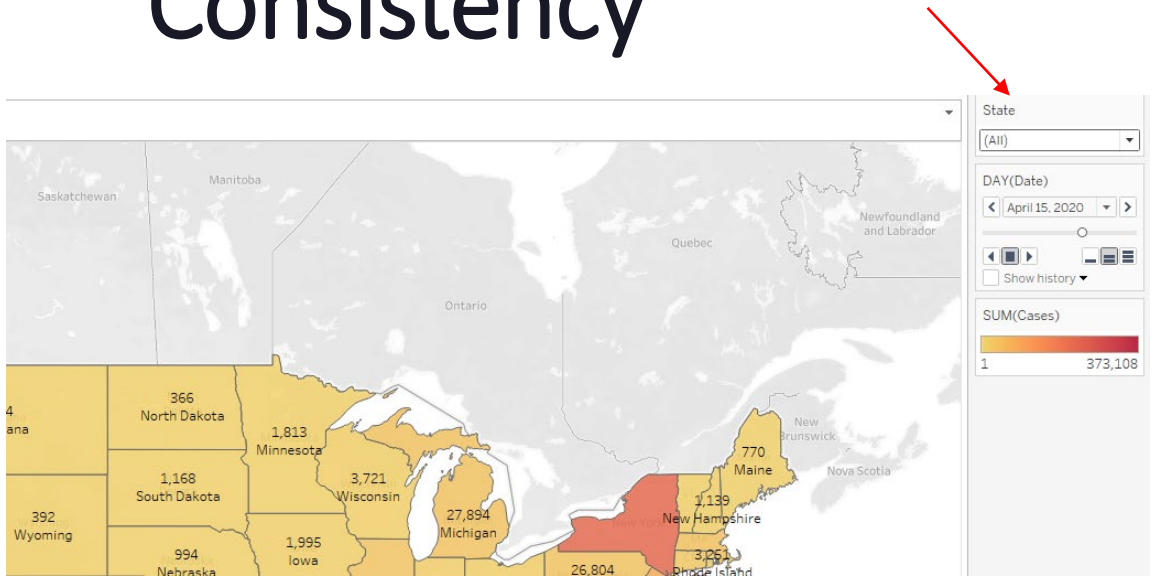
Creating Stories with Multiple Pages

- Stories can contain multiple pages to help you successfully tell your data story. When creating a multi-page report, each page should
  - contain a limited number of objects
  - focus on a single idea
  - stand on its own
  - communicate one point that advances the data story.

Closure



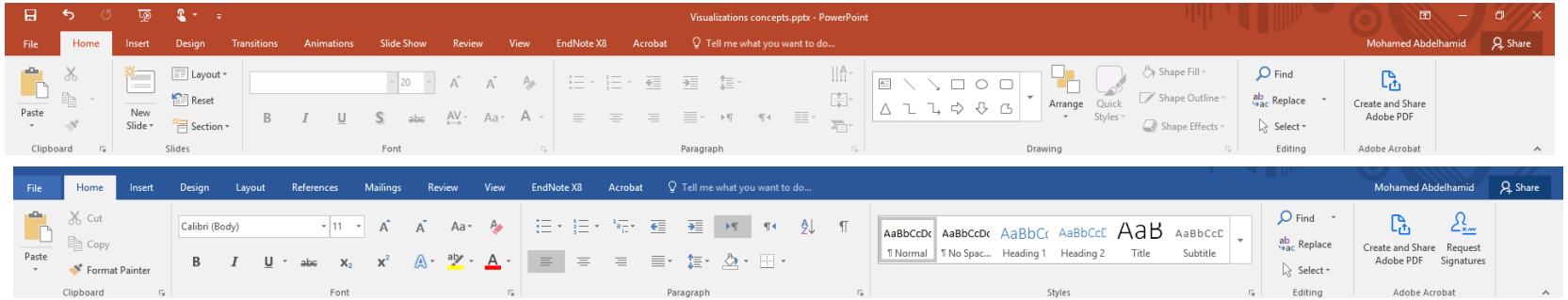
Consistency



Inconsistent use of filters layout

Think like a designer

- **Affordance**
- **Consistency**
- **Don’t overcomplicate**
  - **Make it legible:** use a *consistent*, easy-to-read font (consider both typeface and size).
  - **Keep it clean:** make your data visualization approachable by leveraging visual affordances.
  - **Use straightforward language:** choose simple language over complex, choose fewer words over more words, define any specialized language with which your audience may not be familiar, and spell out acronyms (at minimum, the first time you use them or in a footnote).
  - **Remove unnecessary complexity:** when making a choice between simple and complicated, favor simple.



Eliminate distractions

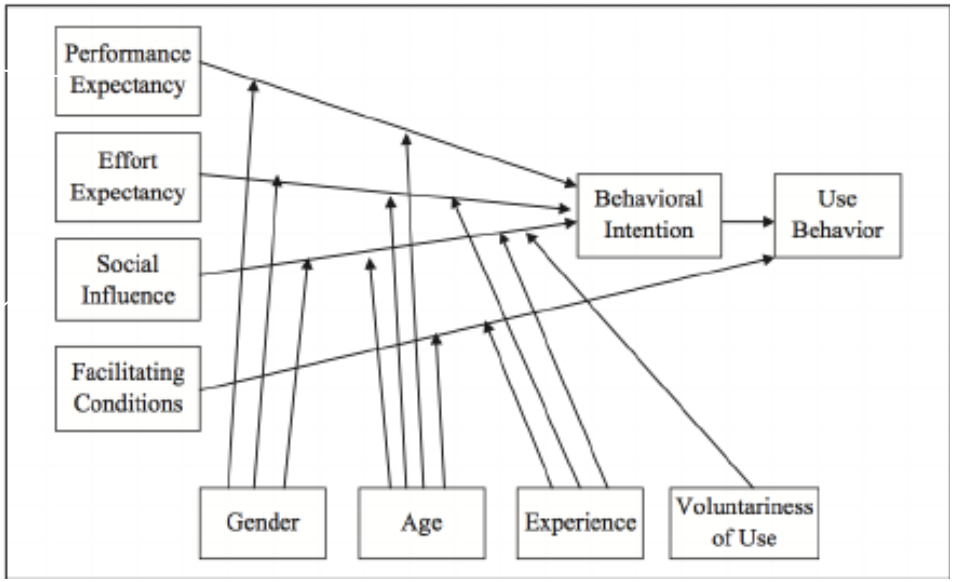
- **Not all data are equally important.**
  - Use your space and audience's attention wisely by getting rid of noncritical data or components.
- **When detail isn't needed, summarize.**
  - You should be familiar with the detail, but that doesn't mean your audience needs to be.
- **Ask yourself: would eliminating this change anything?**
  - No? Take it out! Resist the temptation to keep things because they are cute or because you worked hard to create them
- **Push necessary, but non-message-impacting items to the background**.
  - Use your knowledge of preattentive attributes to de-emphasize. Light grey works well for this.

Unified Theory of Acceptance and Use of Technology (UTAUT)

Benefits

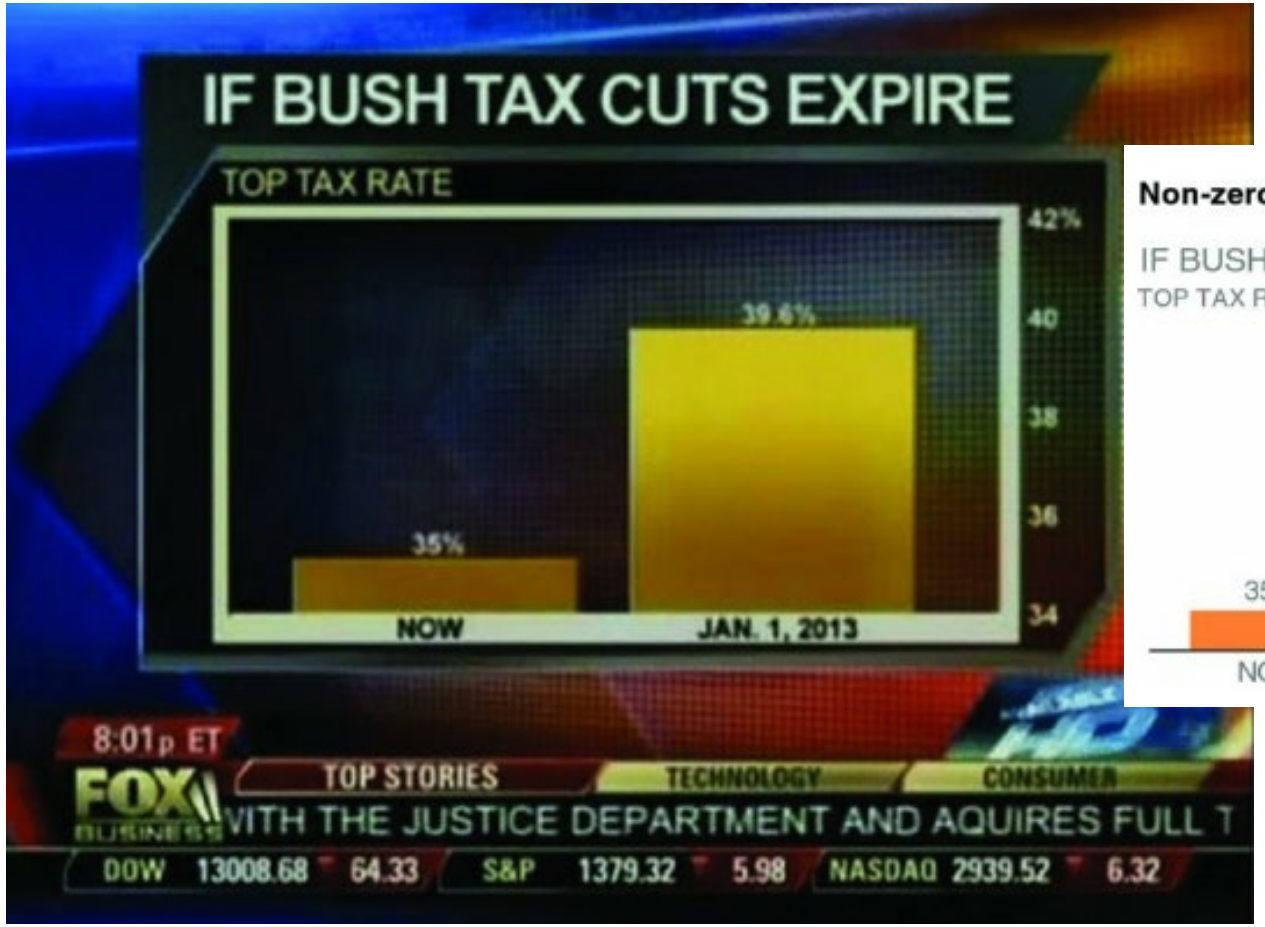
Easy to learn

Easy to use

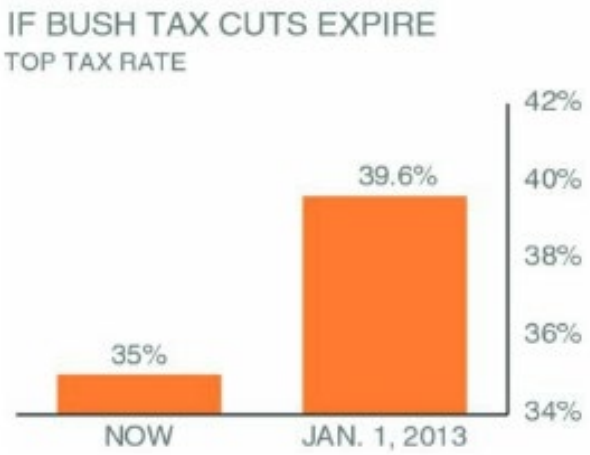


VISUALS TO BE AVOIDED

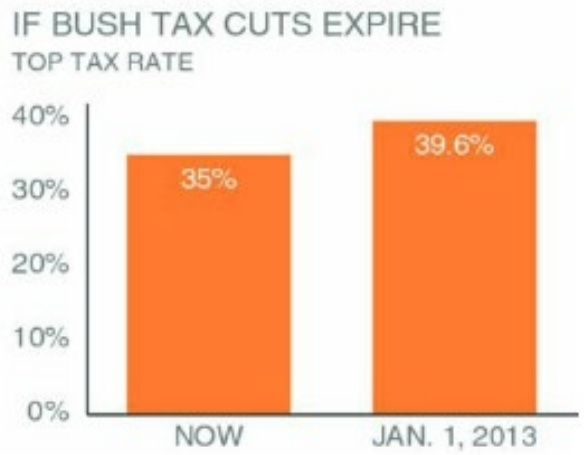
What are your thoughts on this graph?



Non-zero baseline: as originally graphed



Zero baseline: as it should be graphed



Tables

- Your data should stand out not the borders
- Your audience will determine the table format

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

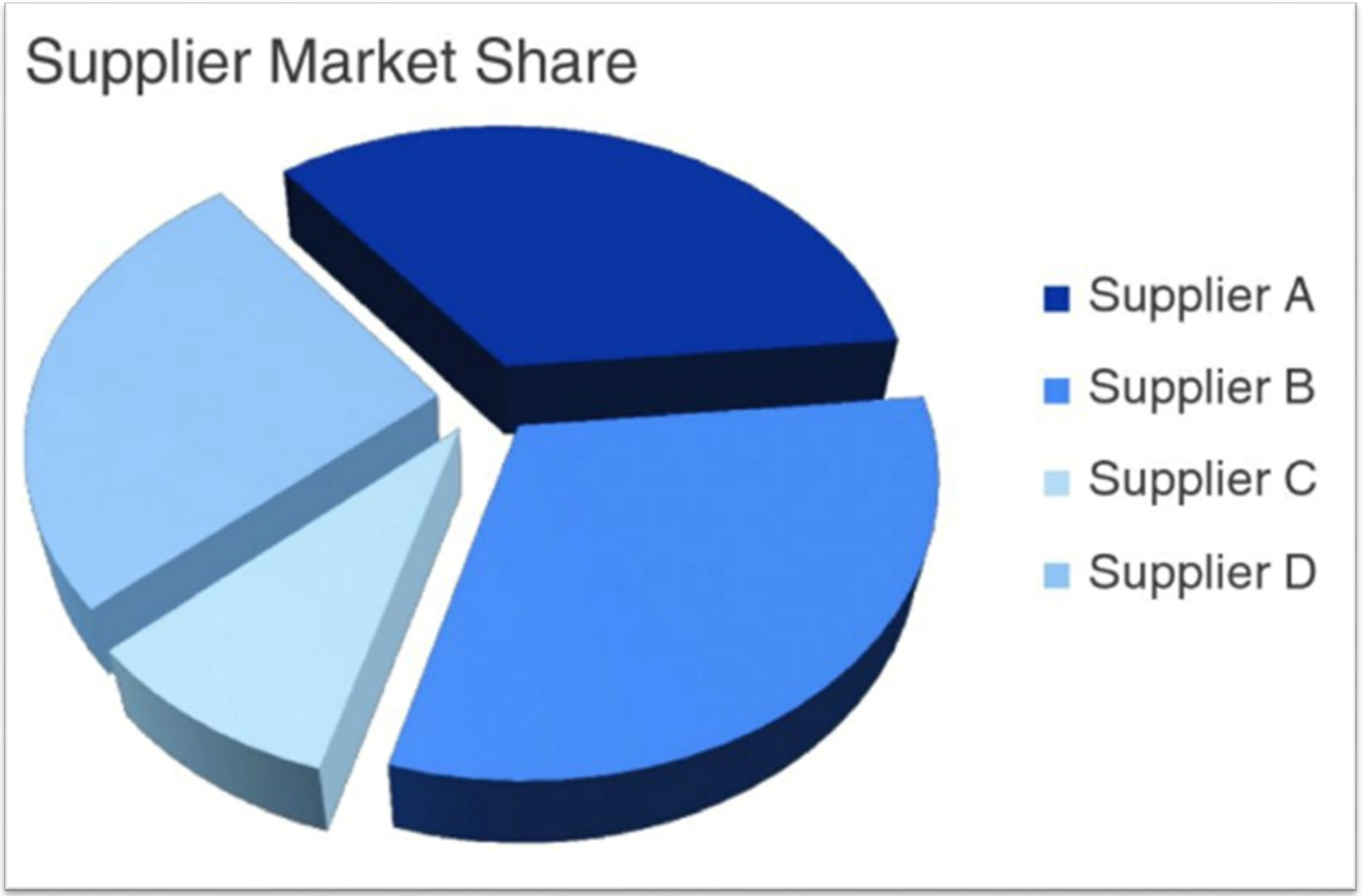
Light borders

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

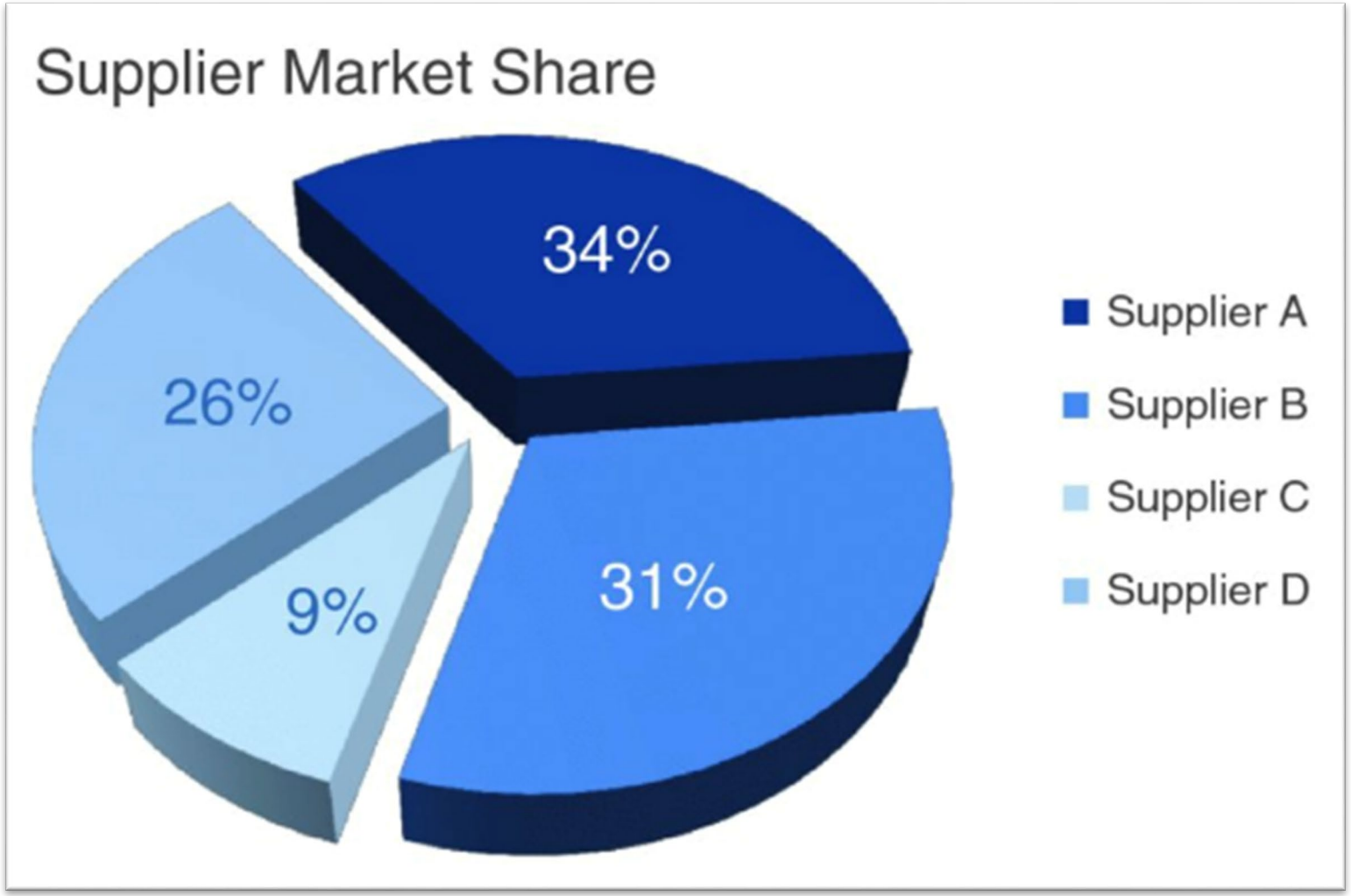
Minimal borders

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

Avoid 3D –
What supplier has the largest share?

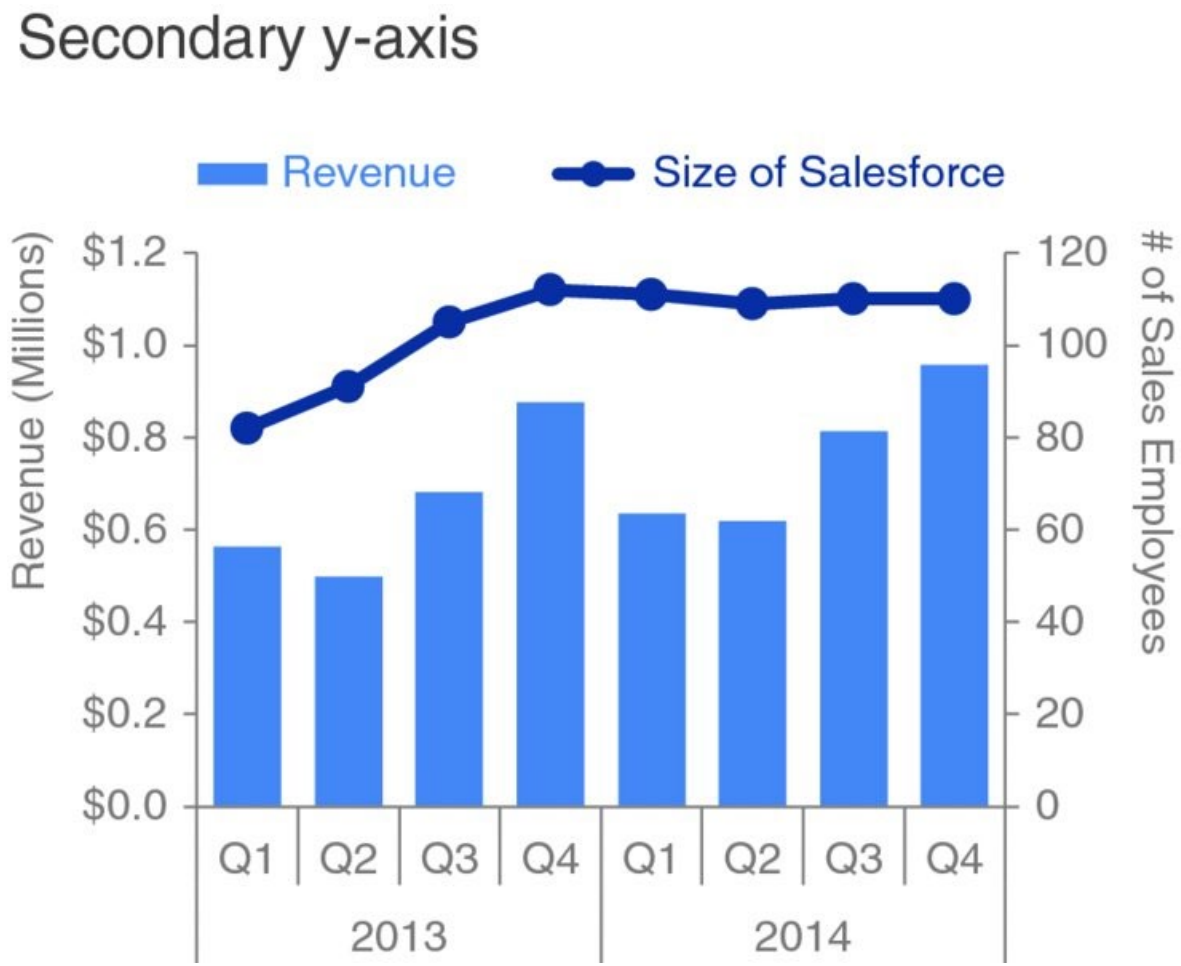


Pie charts are evil

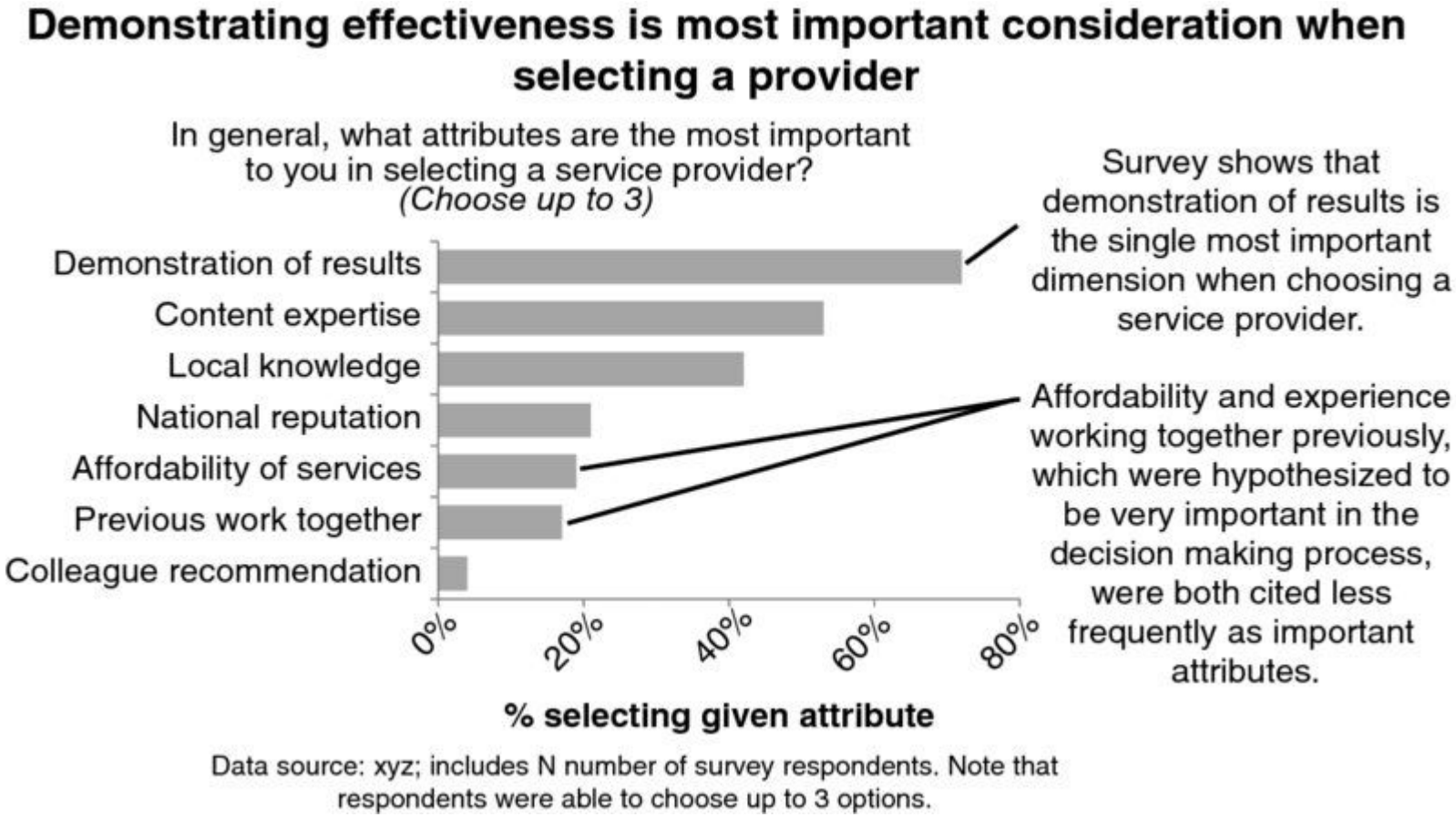


What Problems can you spot ?

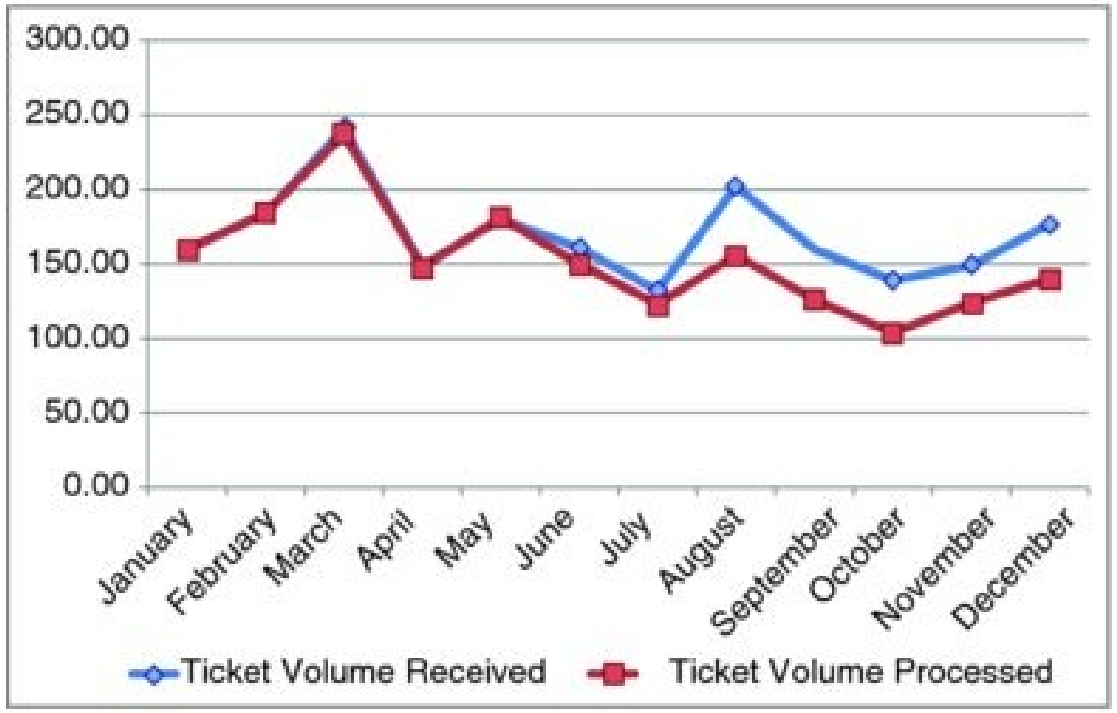
Secondary *y*-axis: generally not a good idea



Clutter is your enemy!



Clutter is your enemy!



Use color sparingly

Limit number of colors

Country Level Sales Rank Top 5 Drugs

Rainbow distribution in color indicates sales rank in given country from #1 (red) to #10 or higher (dark purple)

Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5

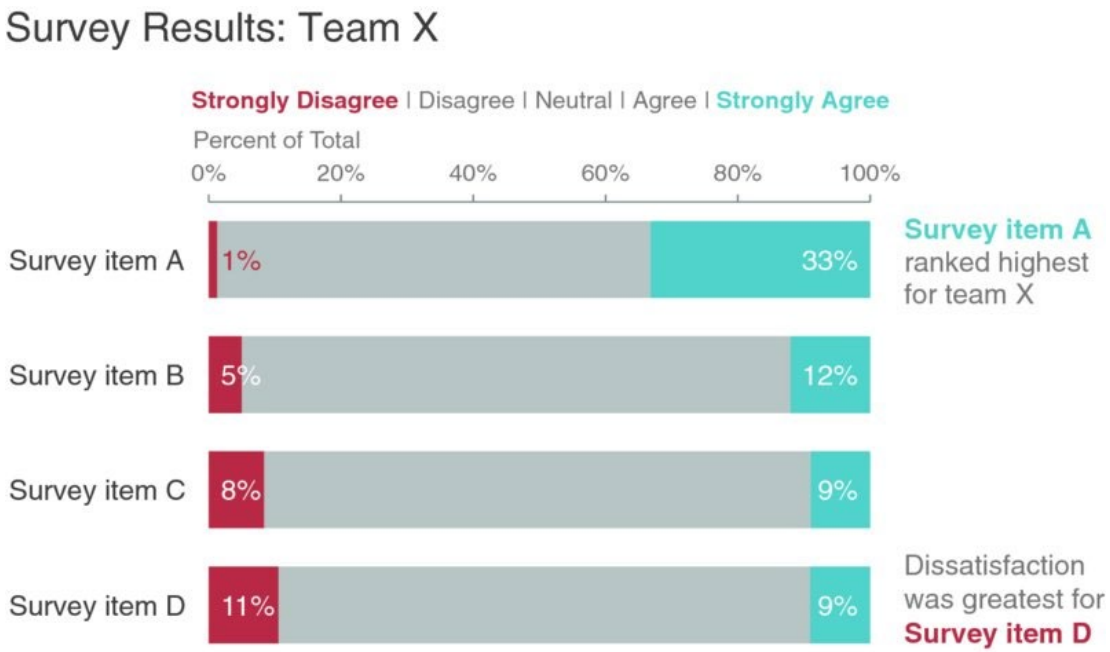
Top 5 drugs: country-level sales rank

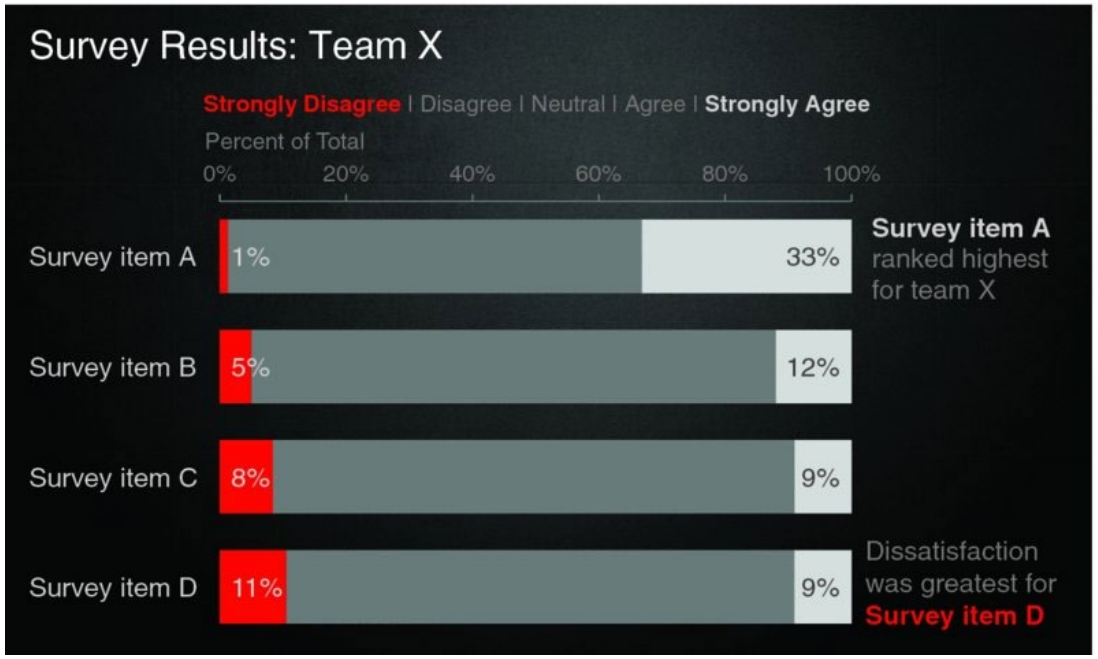
RANK 1 2 3 4 5+

COUNTRY | DRUG

	A	B	C	D	E
Australia	1	2	3	6	7
Brazil	1	3	4	5	6
Canada	2	3	6	12	8
China	1	2	8	4	7
France	3	2	4	8	10
Germany	3	1	6	5	4
India	4	1	8	10	5
Italy	2	4	10	9	8
Mexico	1	5	4	6	3
Russia	4	3	7	9	12
Spain	2	3	4	5	11
Turkey	7	2	3	4	8
United Kingdom	1	2	3	6	7
United States	1	2	4	3	5

Try to avoid dark backgrounds





Where to go from here

1. **learn your tools well**
2. **Iterate and seek feedback**

- Understand the context.
- Know your audience
- Choose an appropriate visual display.
- Eliminate clutter.
- Focus attention where you want it.
- Think like a designer. Tell a story.

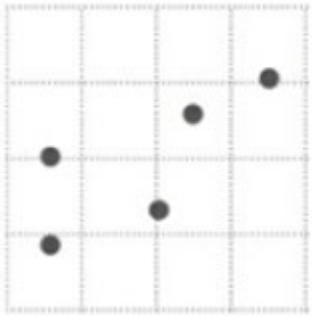
CHOOSING AN EFFECTIVE VISUAL

VISUAL CUES

Visualization involves encoding data with shapes, colors, and sizes. Which cue you choose depends on your data

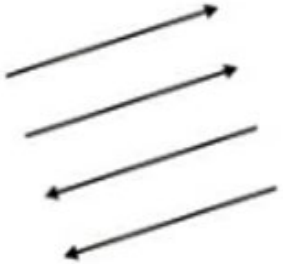
Visual Cues

Position
Where in space the data is



Length
How long the shapes are

Direction
Slope of a vector in space



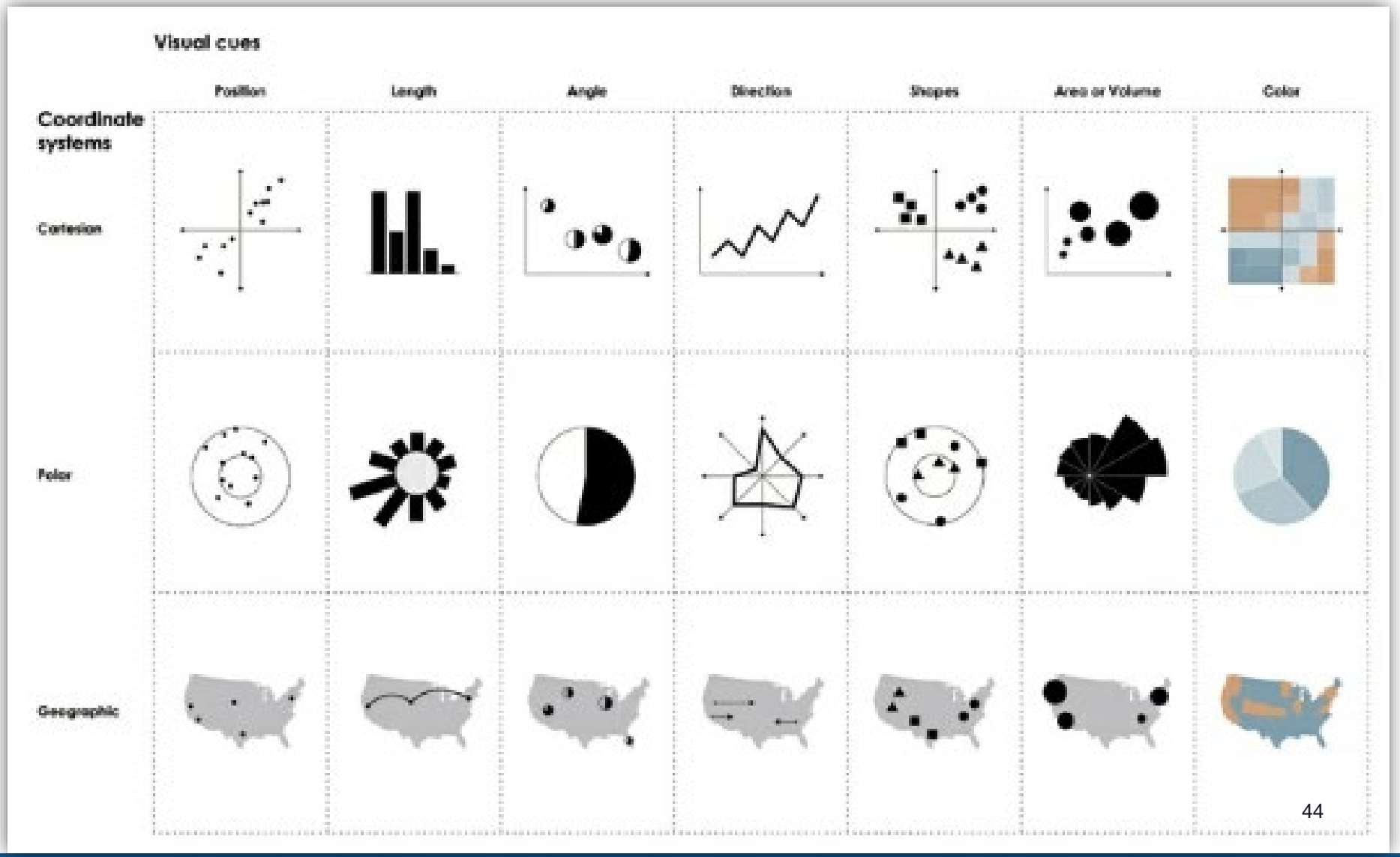
Shapes
Symbols as categories

Area
How much 2-D space

Color saturation
Intensity of a color hue

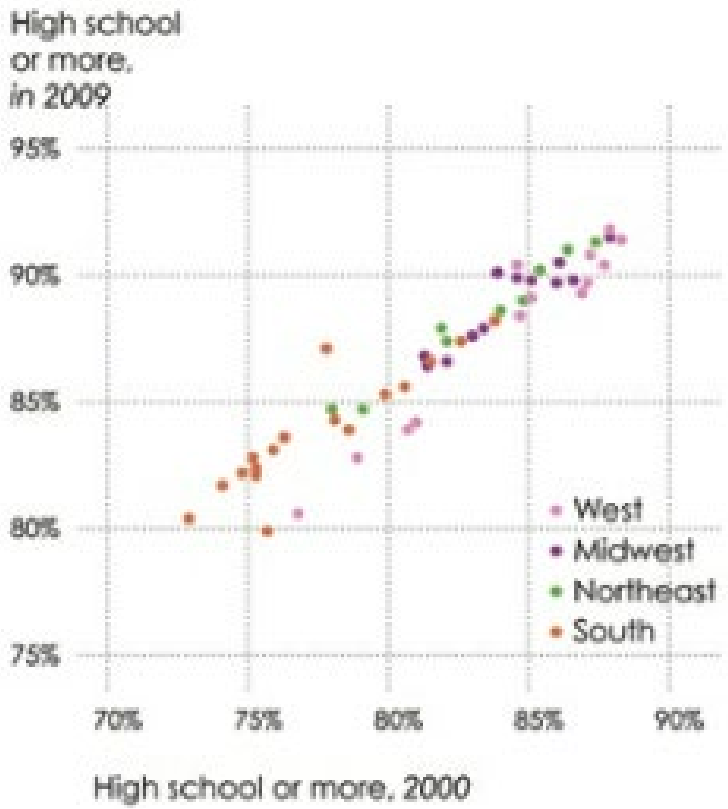
Color hue
Usually referred to as color

Visual Component Combinations

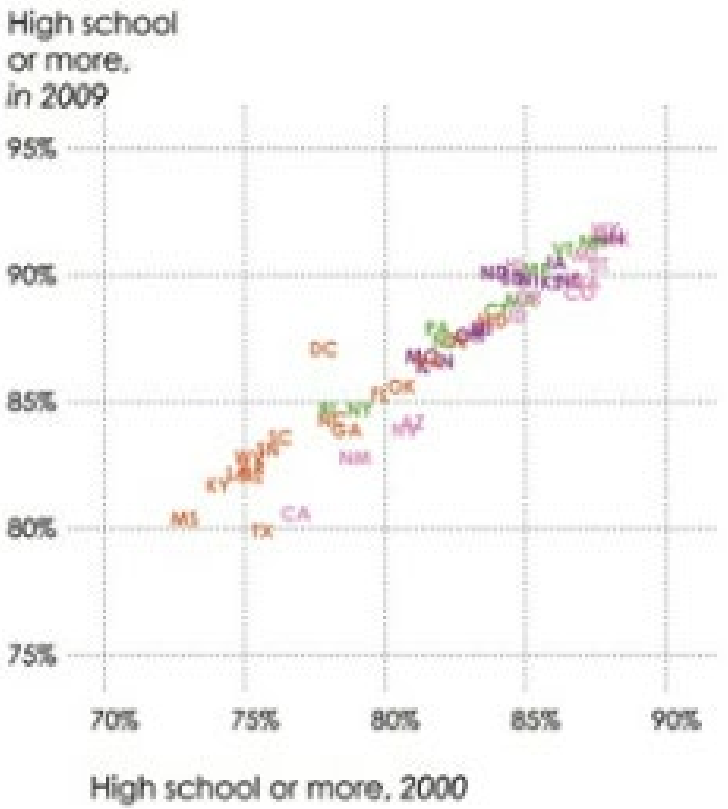


Visual Component Combinations

Position + Color



Position + Symbols + Color



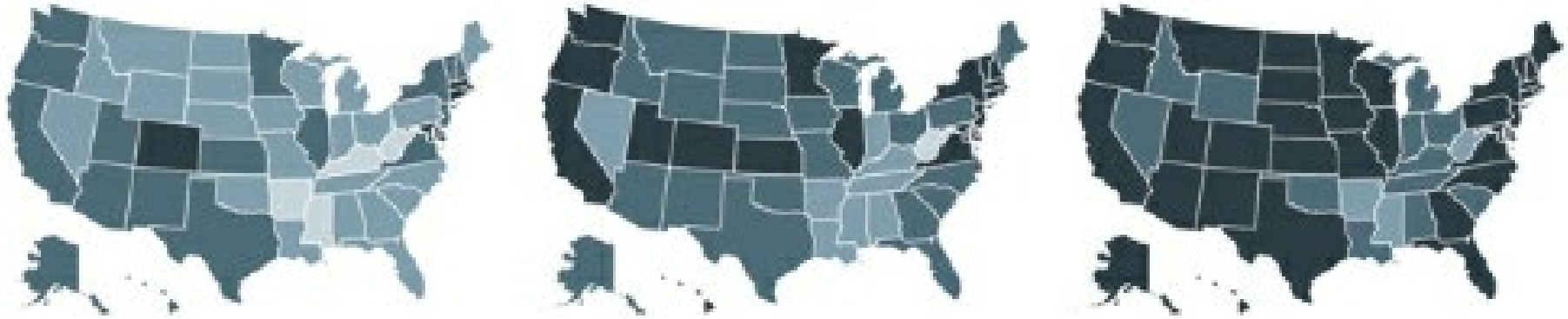
Visual Component Combinations

Geographic coordinates + Time scale + Color

In 1990, about 20 percent of 25+ year olds had at least a bachelor's degree.

In 2000, the percentage was up to 24.

In 2009, the US average was up to 29 percent.

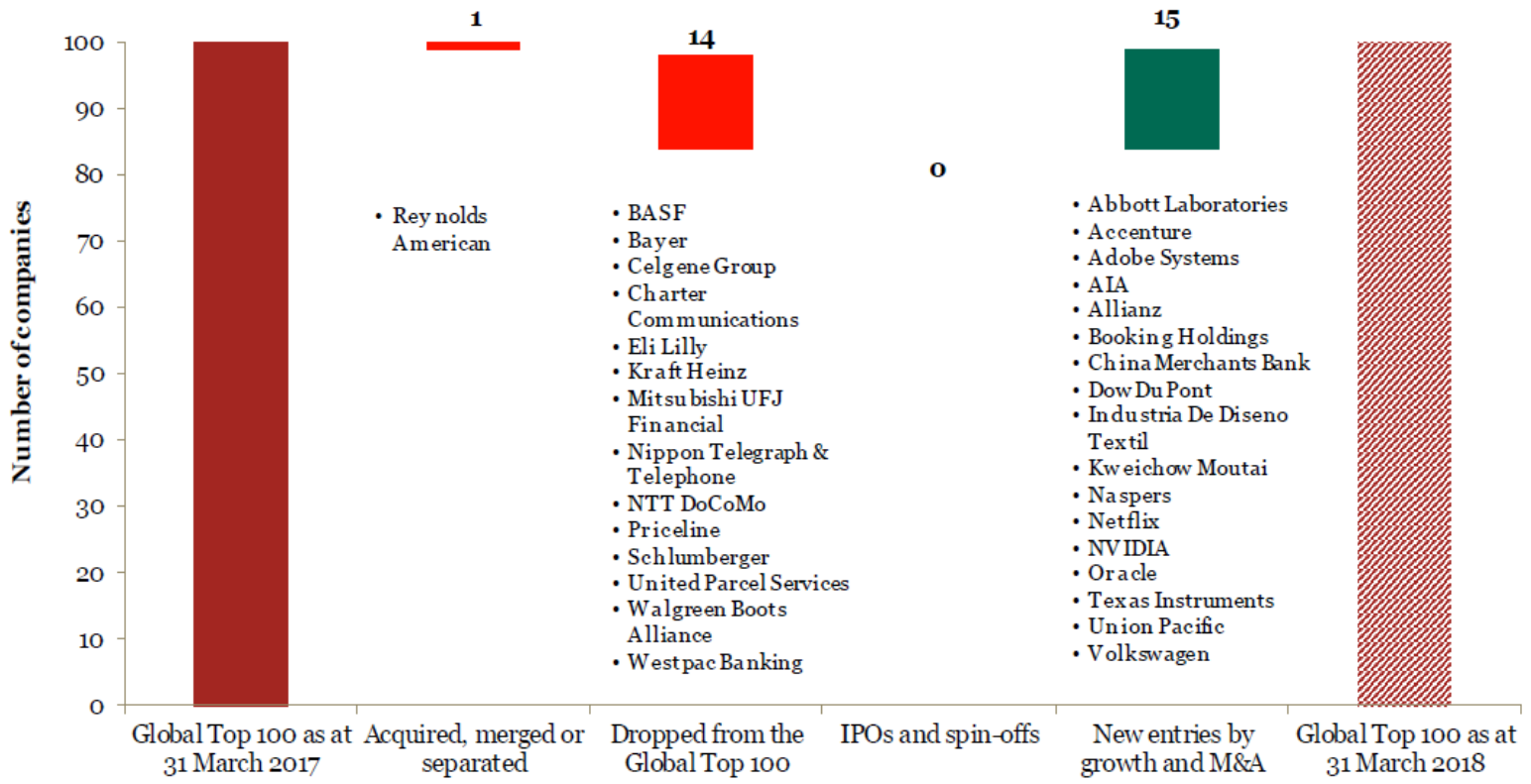


Bachelor's degree or more

15% or less — 20% — More than 25%

Waterfall

The waterfall chart can be used to pull apart the pieces of a stacked bar chart to focus on one at a time, or to show a starting point, increases and decreases, and the resulting ending point.



Source: Bloomberg and PwC analysis

Heat map

- A heatmap is a way to visualize data in tabular format, where in place of (or in addition to) the numbers, you leverage colored cells that convey the relative magnitude of the numbers.

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%

Examples

Simple text

Scatterplot

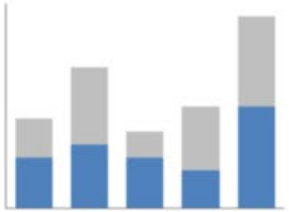
Vertical bar

Horizontal bar

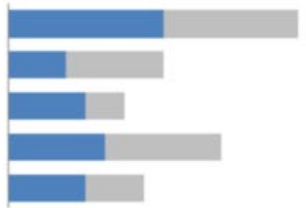
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Table

Line



Stacked vertical bar

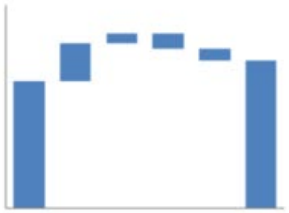


Stacked horizontal bar

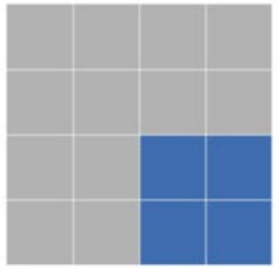
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

Slopegraph



Waterfall

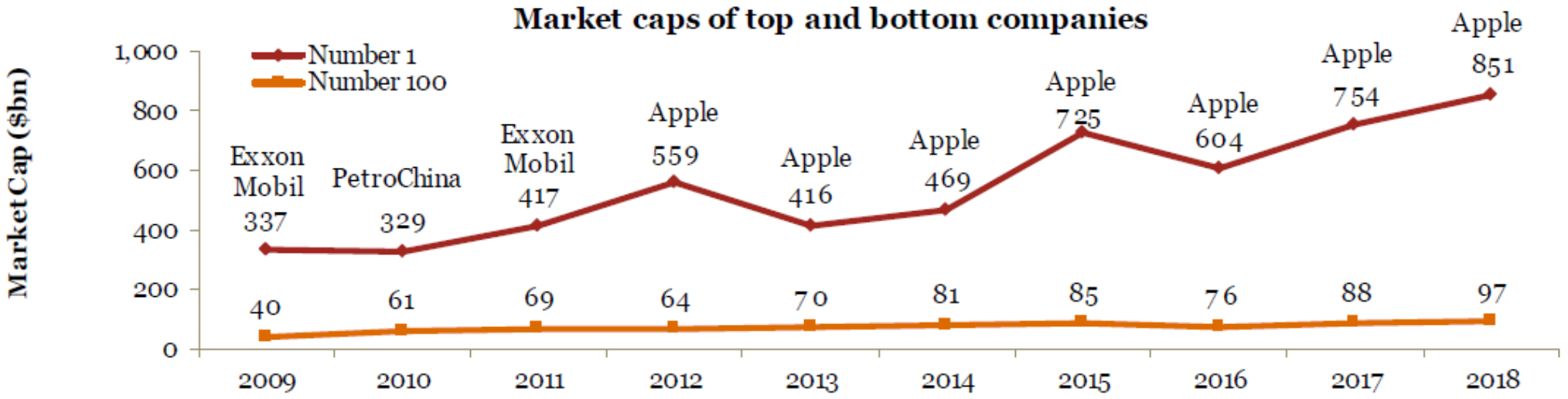


Square area

Examples

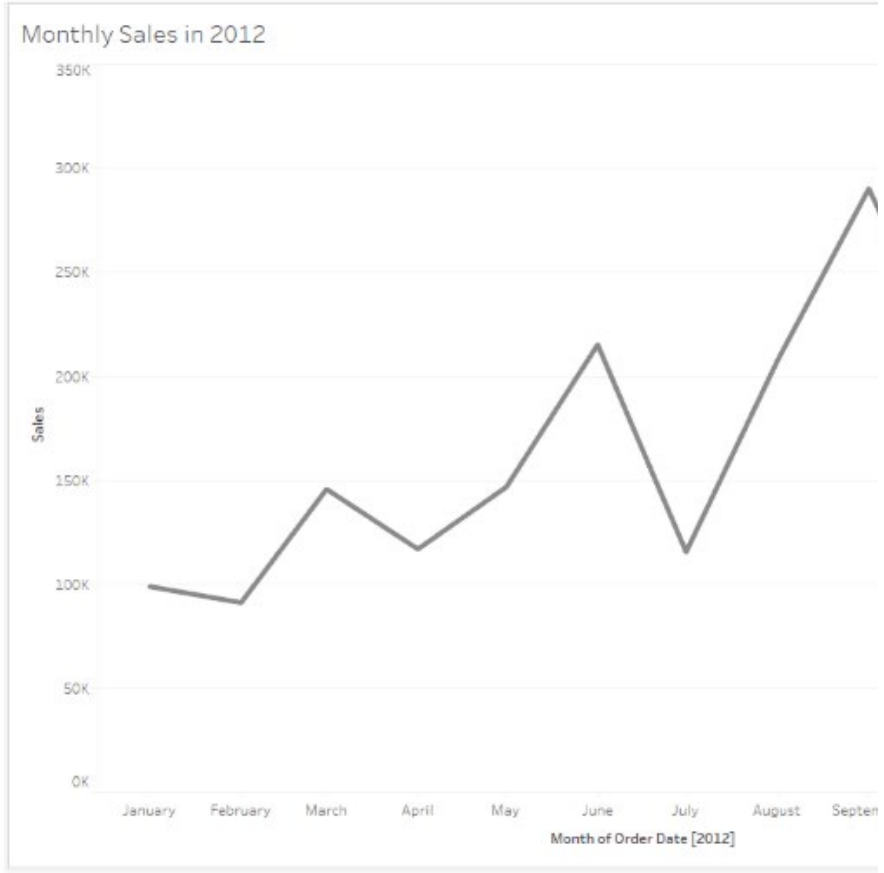
- https://www.pwc.com/gx/en/audit-services/assets/pdf/global-top-100-companies-2018-report.pdf

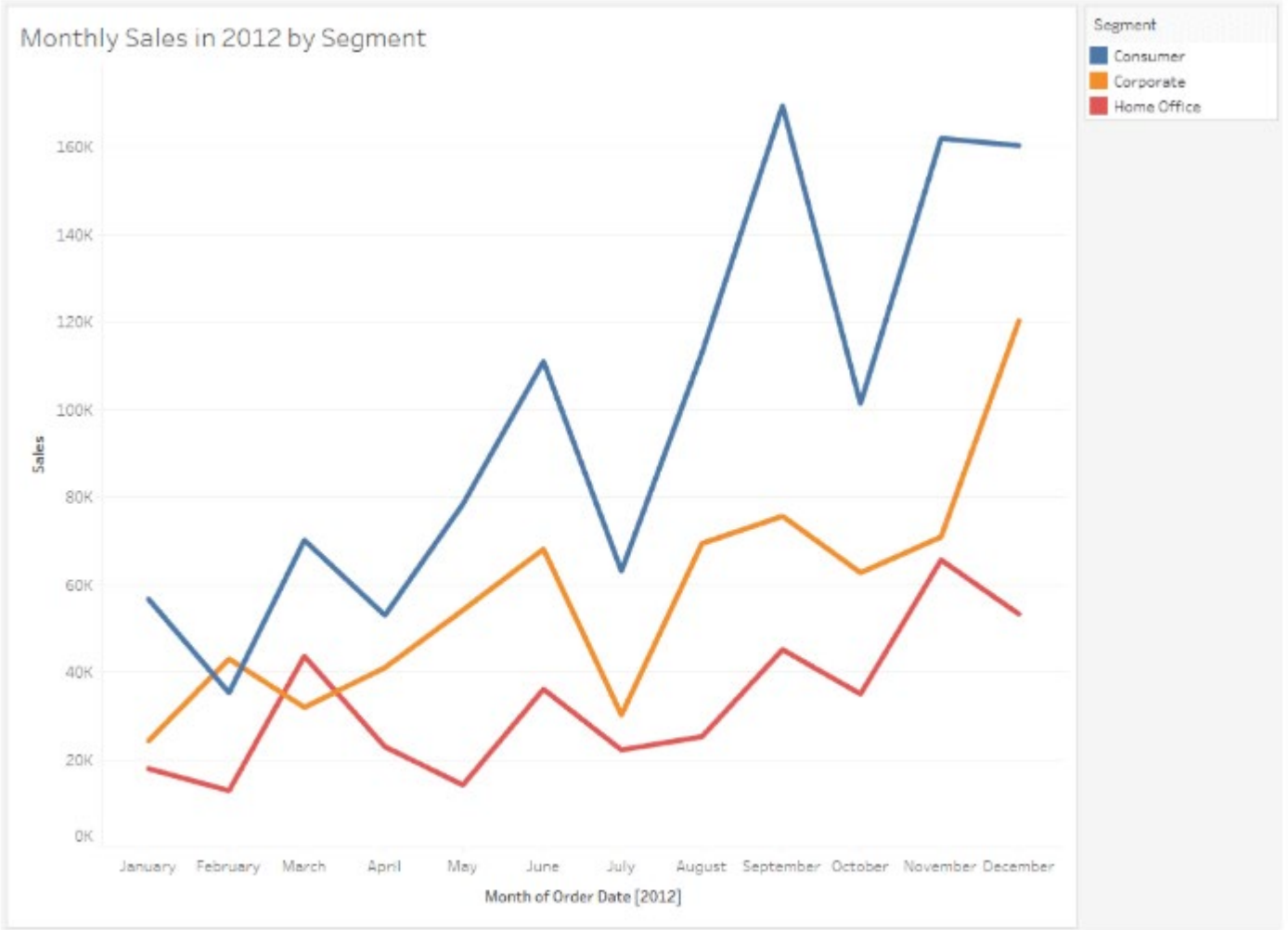
Line Graph



Source: Bloomberg and PwC analysis

Line charts: Viewing data over time

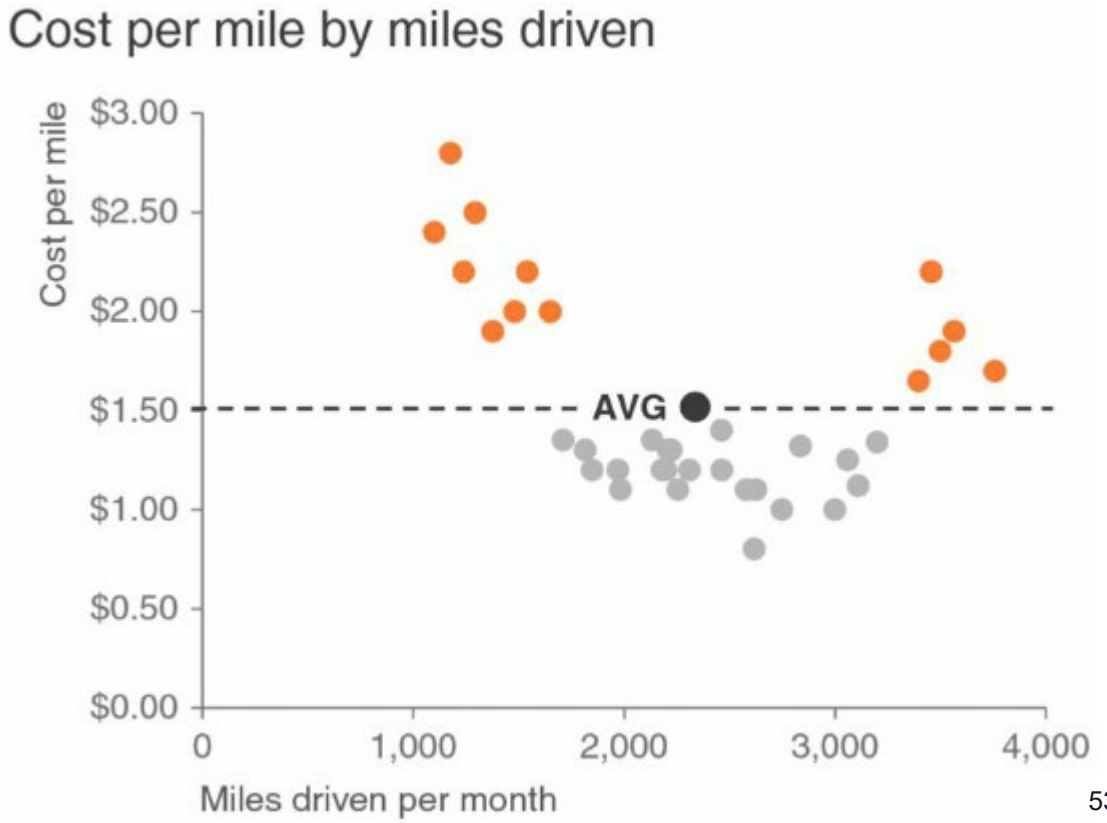




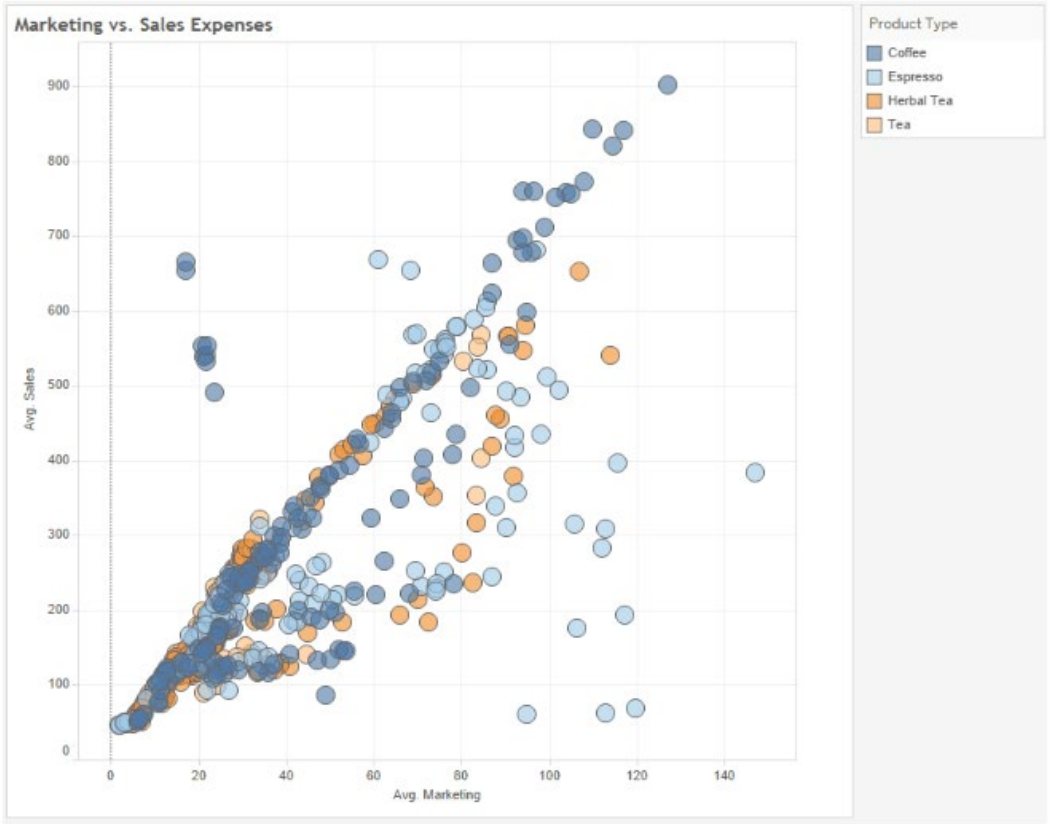
Same data from the earlier chart, now showing detail according to customer segment.

Scatterplot

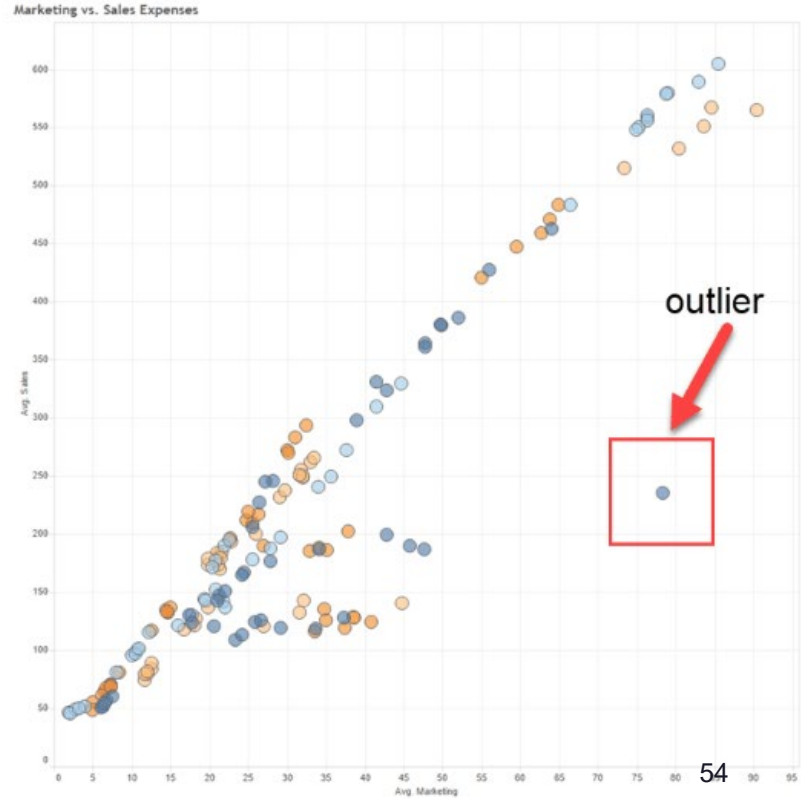
- Scatterplots can be useful for showing the relationship between two things, because they allow you to encode data simultaneously on a horizontal x-axis and vertical y-axis to see whether and what relationships exists.



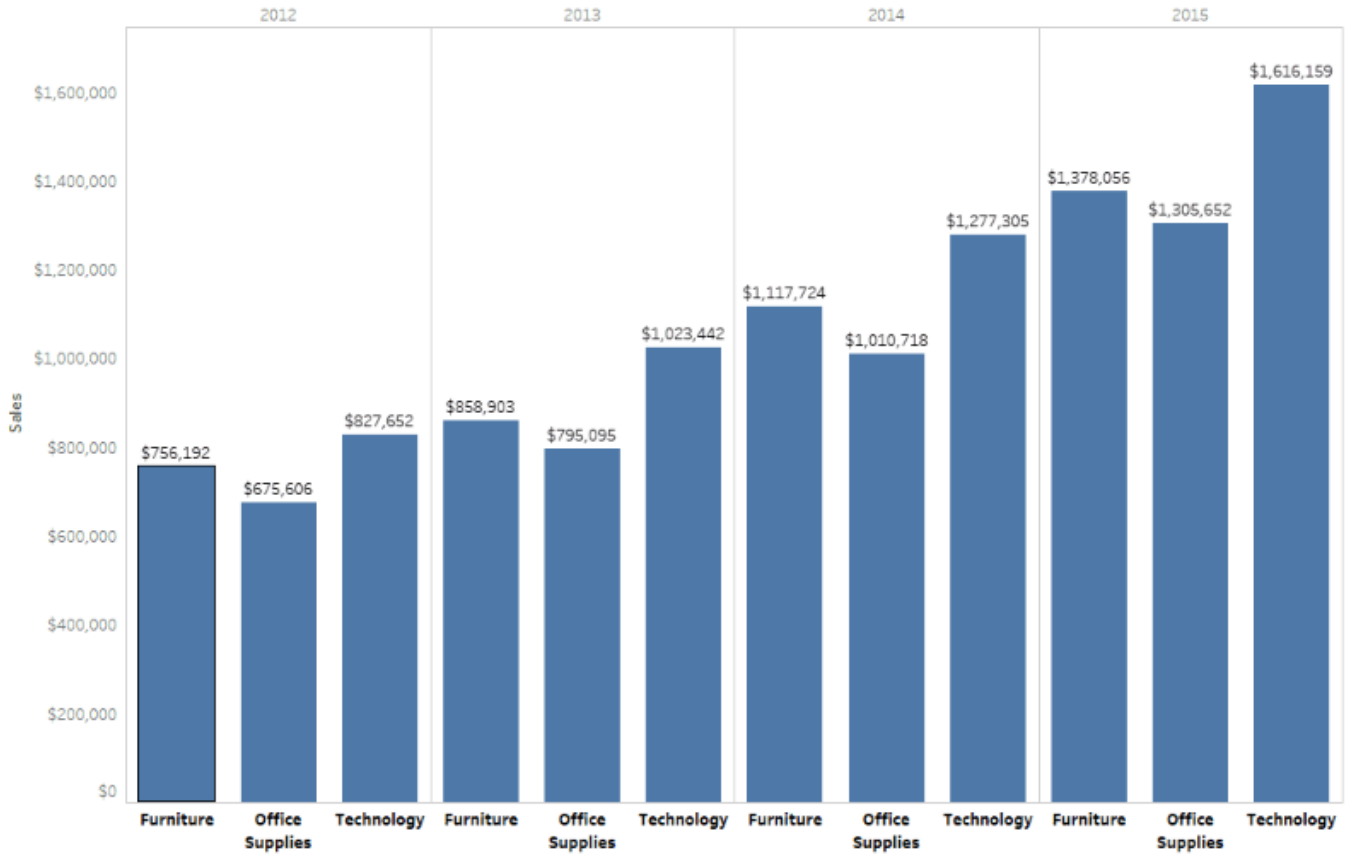
Scatter plot: Viewing data relationships and outliers



A scatter plot showing how marketing expenses correlates to sales of coffee and tea.

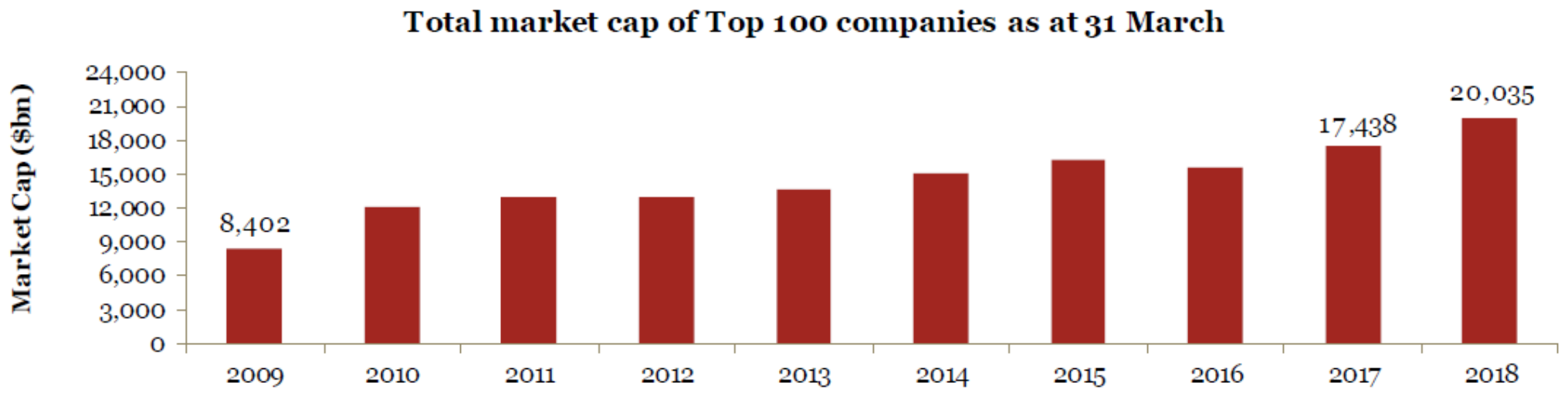


Bar graph: Comparing categories of data



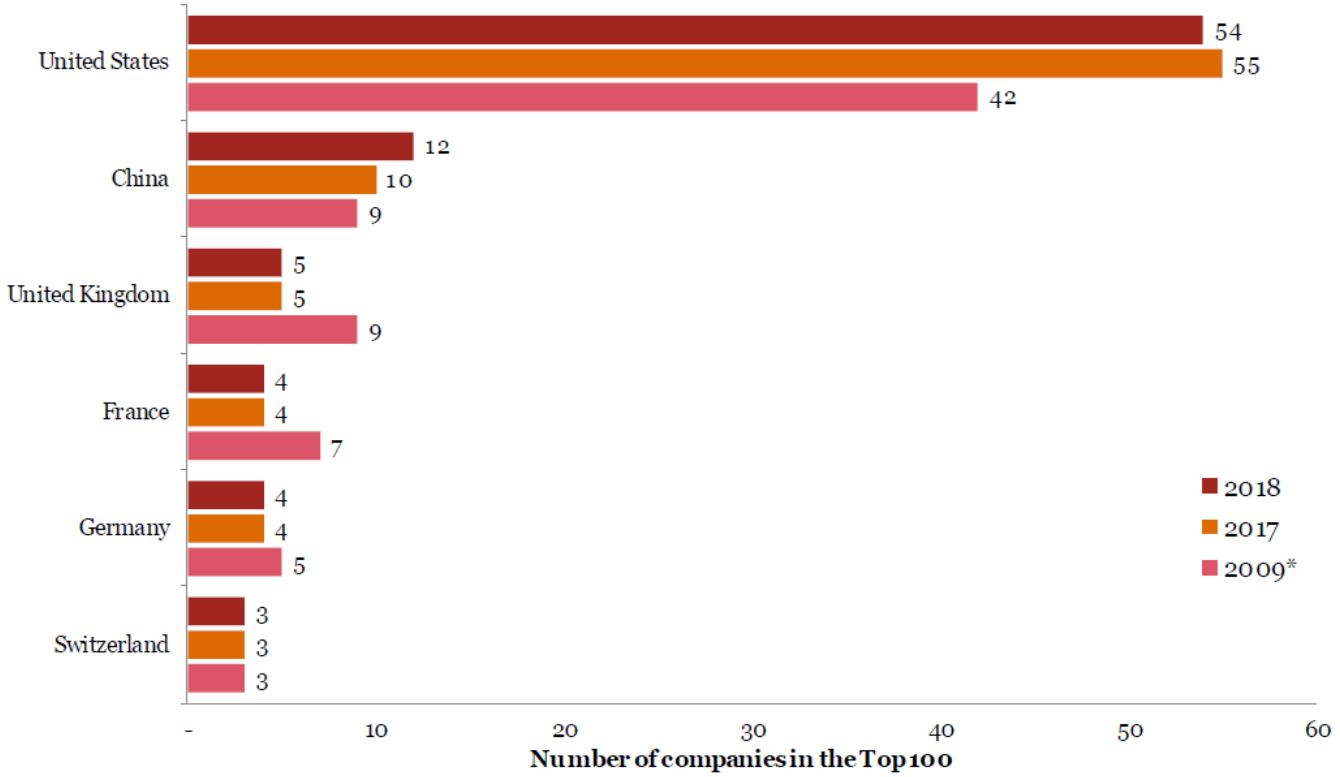
Bar Graph – Vertical

Like line graphs, vertical bar charts can be single series, two series, or multiple series. As you add more series of data, it becomes more difficult to focus on one at a time and pull out insight



Bar Graph – Horizontal

Sometimes bar charts are avoided because they are common. This is a mistake. Rather, bar charts should be leveraged because they are common, as this means less of a learning curve for your audience.



** 2009 figures do not add to 100 due to seven companies in the 2009 Top 100 being in locations of domicile that are no longer in the Global Top 100*

Source: Bloomberg and PwC analysis

Simple text

- When you have just a number or two that you want to communicate: use the numbers directly.

Tables

Location	Change in market cap of the current Top 100 ($bn)	Top 100 2018 Rank	Top 100 2018 Market cap ($bn)	Top 100 2018 # companies	Market cap 2009 ($bn) (1)	Top 100 2009 Rank	Top 100 2009 Market cap ($bn)	Top 100 2009 # companies
United States	8,563	1	12,187	54	3,624	1	3,805	42
China	1,797	2	2,822	12	1,025	2	1,061	9
United Kingdom	343	3	814	5	471	3	715	9
Switzerland	308	4	655	3	347	8	347	3
France	269	5	532	4	263	4	481	7
Germany	207	6	437	4	230	6	299	5
South Korea	237	7	298	1	61	21	61	1
Hong Kong	115	8	290	2	175	9	219	2
Belgium	178	9	222	1	44	13	44	1
Taiwan	181	10	220	1	39	-	-	-
Canada	146	11	216	2	70	12	41	1
Ireland	159	12	212	2	53	-	-	-
Japan	102	13	210	1	108	5	366	6
Australia	55	14	209	2	154	12	118	1
Spain	122	15	202	2	80	11	149	2
Netherlands	107	16	165	1	58	19	59	1
Denmark	93	17	123	1	30	-	-	-
Brazil	87	18	114	1	27	7	234	3
South Africa	100	19	107	1	7	-	-	-
Russia	0	-	0	0	0	10	134	2
Italy	0	-	0	0	0	17	77	1
Norway	0	-	0	0	0	20	56	1
India	0	-	0	0	0	16	47	1
Mexico	0	-	0	0	0	18	45	1
Finland	0	-	0	0	0	15	44	1

(1) Market cap of companies at 31 March 2009 in the Global Top 100 at 31 March 2018

Source: Bloomberg and PwC analysis

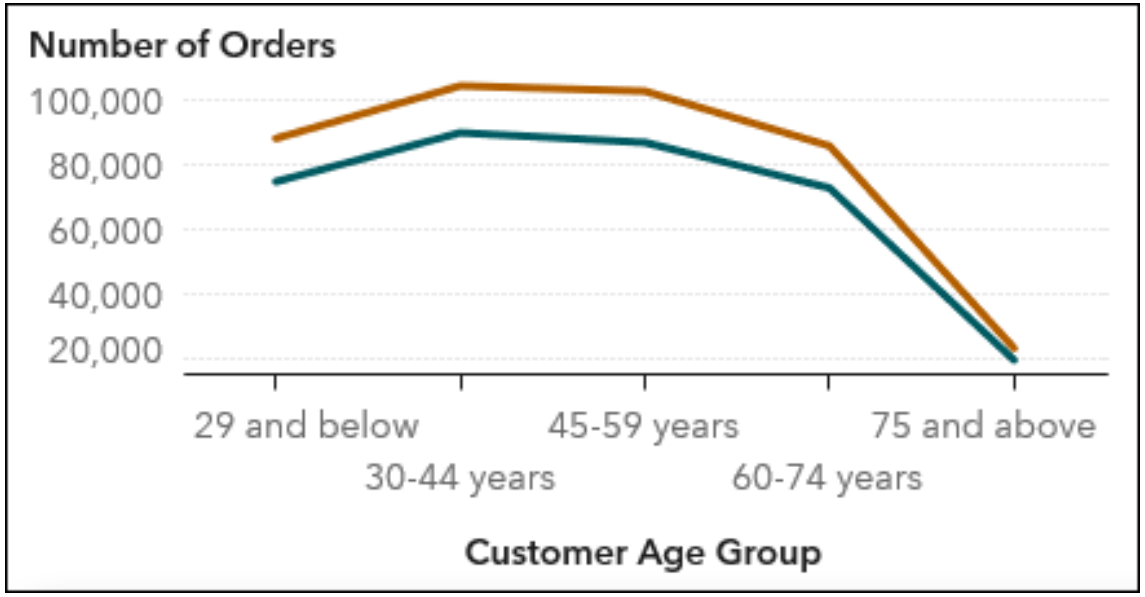
Graphs (Summary)

Use a *heat map* to evaluate the relationship between two high-cardinality measures, between two categories, or between a category and a measure.

Use a *scatter plot* to evaluate the relationship between two measures.

Graphs (Summary)

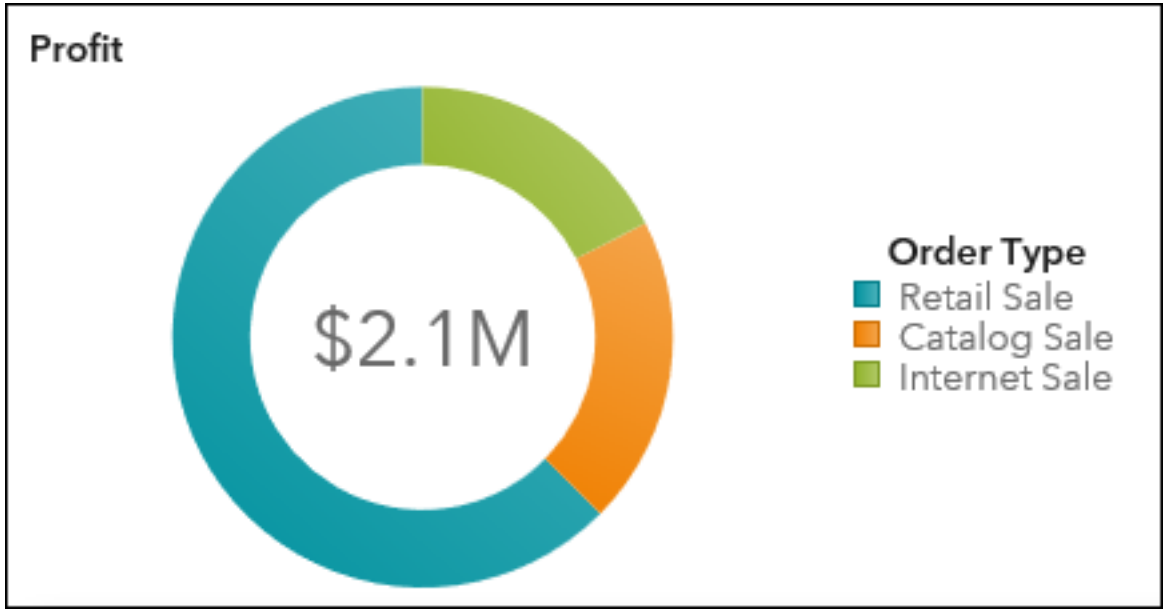
Use a *time series plot* to show trends of measures over time.



Use a *line chart* to show trends over some ordinal variable (time, age group).

Graphs (Summary)

Use a *word cloud* to show summary information in an appealing fashion.



Use a *donut chart (pie chart)* to compare a few groups whose values vary greatly.