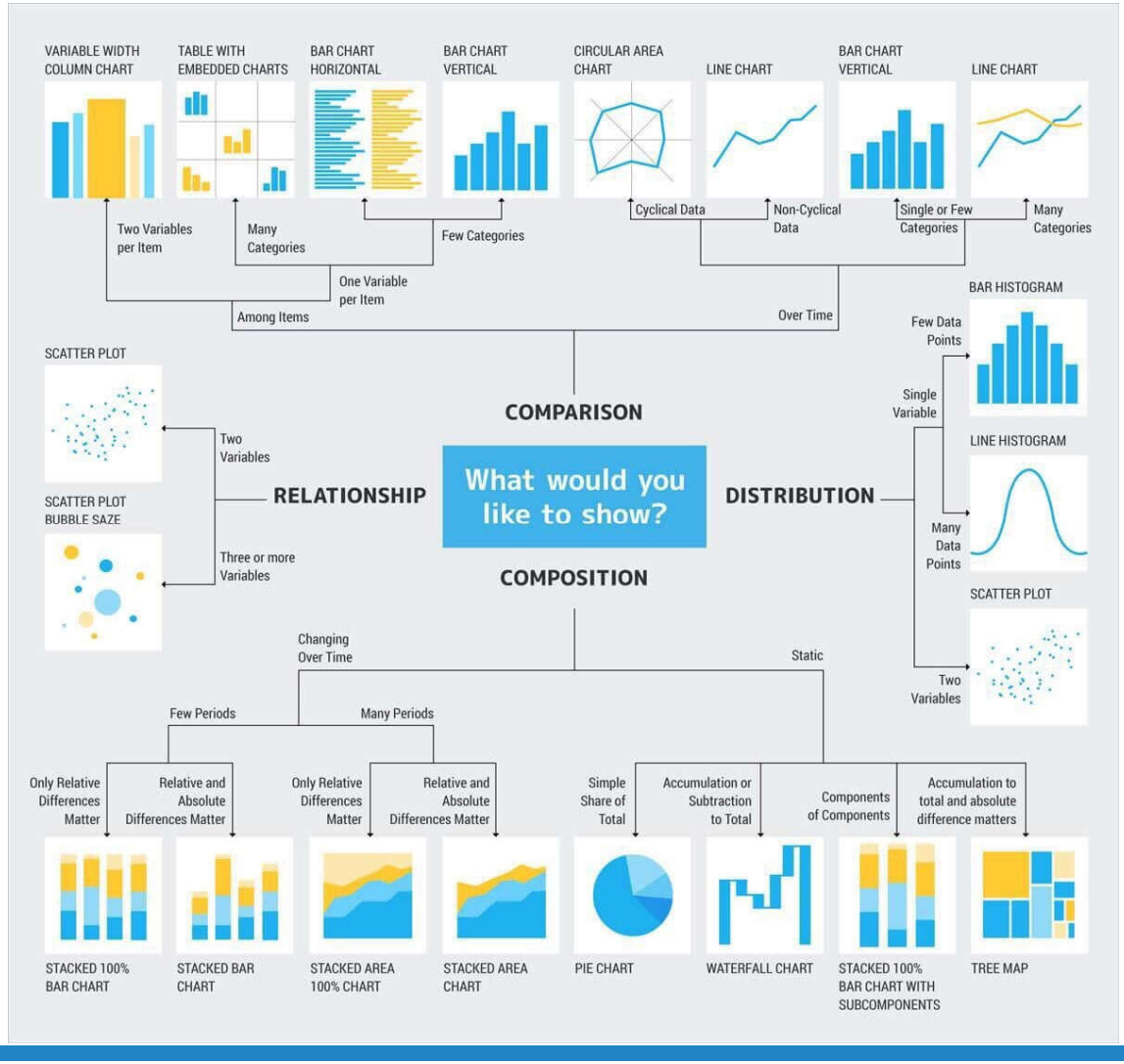
Introduction to Data Science

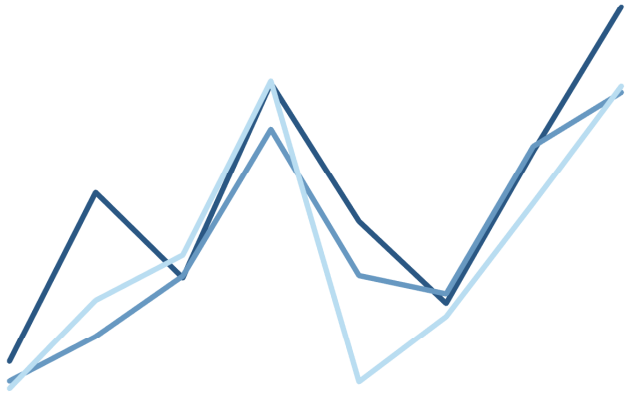
Dashboards and Storytelling

Do I have the right chart?


What would you like to show?
COMPARISON
RELATIONSHIP
DISTRIBUTION
COMPOSITION
VARIABLE WIDTH COLUMN CHART
TABLE WITH EMBEDDED CHARTS
BAR CHART HORIZONTAL
BAR CHART VERTICAL
CIRCULAR AREA CHART
LINE CHART
BAR CHART VERTICAL
LINE CHART
Two Variables per Item
Many Categories
Few Categories
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
One Variable per Item
Among Items
Over Time
BAR HISTOGRAM
Few Data Points
Single Variable
LINE HISTOGRAM
Many Data Points
SCATTER PLOT
Two Variables
SCATTER PLOT BUBBLE SAZE
Three or more Variables
SCATTER PLOT
Two Variables
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Accumulation to total and absolute difference matters
STACKED 100% BAR CHART
STACKED BAR CHART
STACKED AREA 100% CHART
STACKED AREA CHART
PIE CHART
WATERFALL CHART
STACKED 100% BAR CHART WITH SUBCOMPONENTS
TREE MAP

Comparison

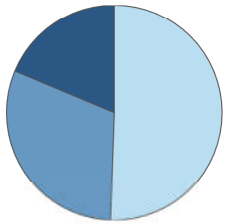
Among Items		Over Time
Bar	Column	Line



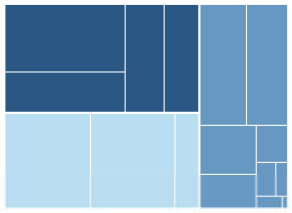
Composition

Static	Over Time

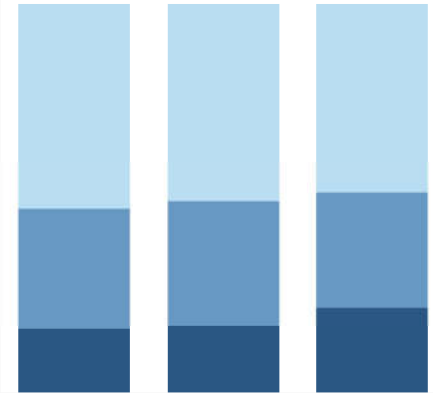
Pie



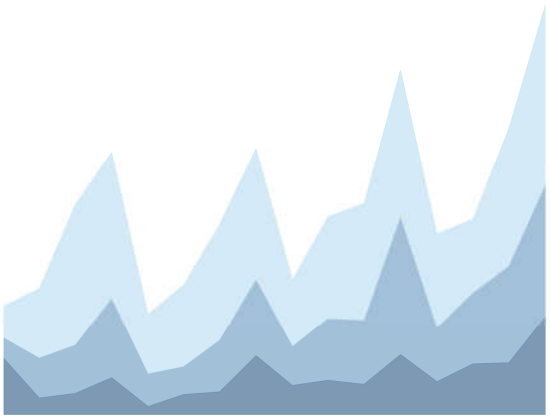
Tree



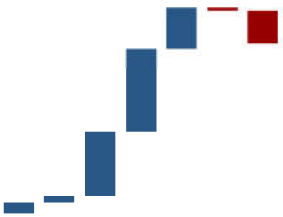
Column Stacked



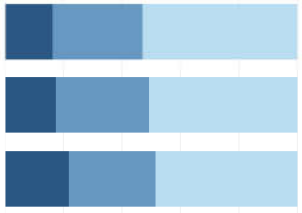
Stacked Area



Waterfall

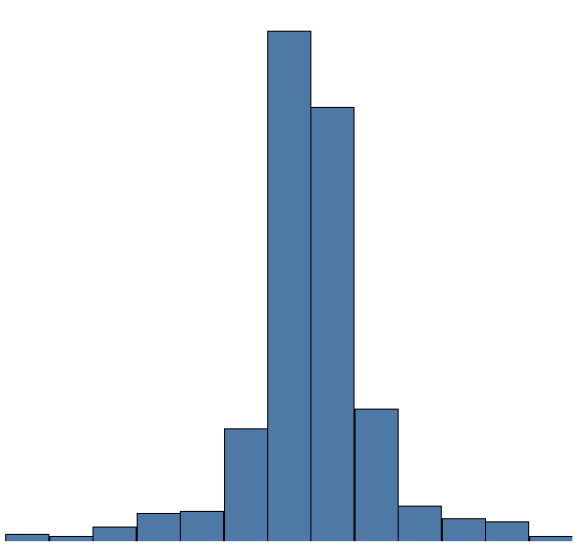


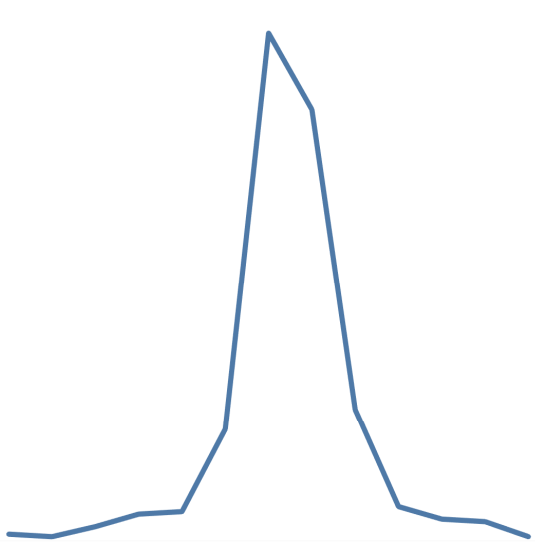
Stacked Bar

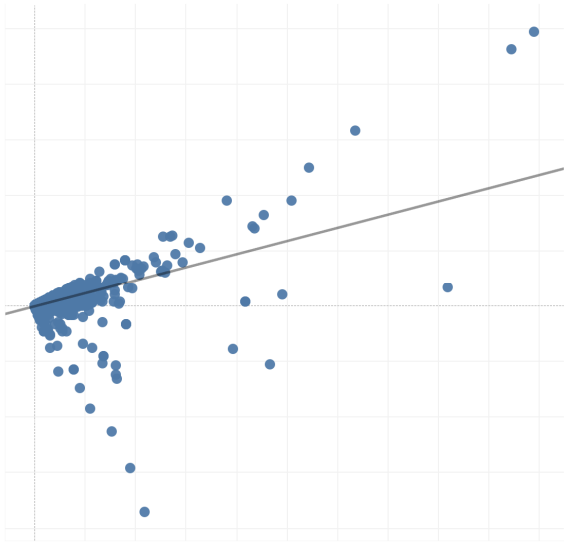


Distribution

Single Variable		Two Variables
Few Data Points	Many Data Points	
Column Histogram	Line Histogram	Scatter



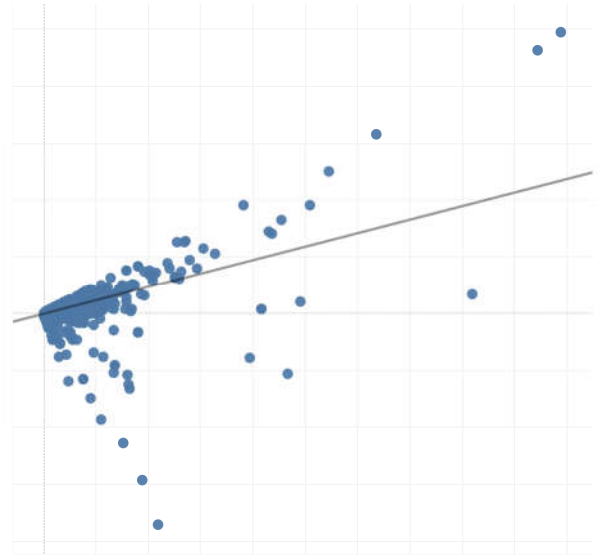




Relationship

Two Variables	Three+ Variables

Scatter

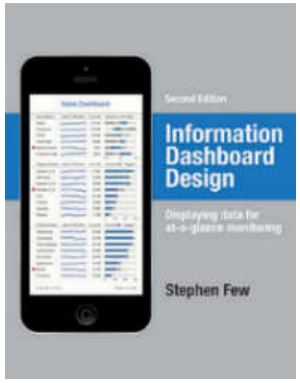


Bubble

Dashboard Flow

What is a dashboard?

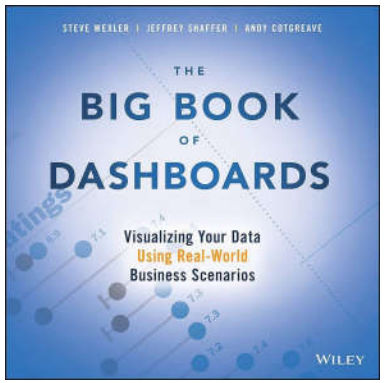
"A dashboard is a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance."

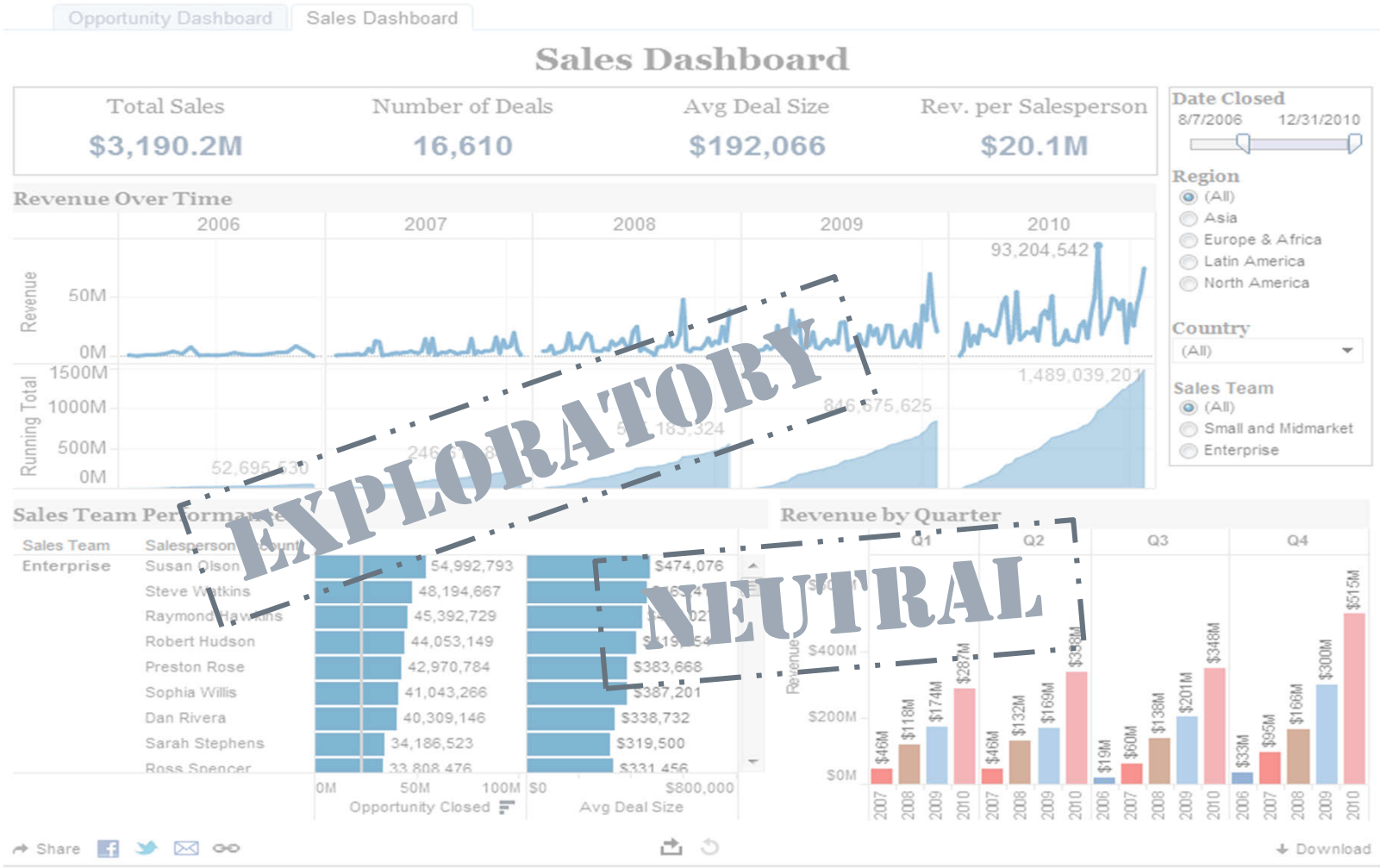


Stephen Few (2004)

"A dashboard is a visual display of data used to monitor conditions and/or facilitate **understanding**"

Big Book of Dashboards (2017)




Opportunity Dashboard
Sales Dashboard
Sales Dashboard
Total Sales
$3,190.2M
Number of Deals
16,610
Avg Deal Size
$192,066
Rev. per Salesperson
$20.1M
Date Closed
8/7/2006
12/31/2010
Region
(All)
Asia
Europe & Africa
Latin America
North America
Country
(All)
Sales Team
(All)
Small and Midmarket
Enterprise
Revenue Over Time
2006
2007
2008
2009
2010
Revenue
50M
0M
93,204,542
Running Total
1500M
1000M
500M
0M
52,695,530
846,675,625
1,489,039,201
EXPLORATORY
NEUTRAL
Sales Team Performance
Sales Team
Salesperson
Enterprise
Susan Olson
Steve Watkins
Raymond Havens
Robert Hudson
Preston Rose
Sophia Willis
Dan Rivera
Sarah Stephens
Ross Spencer
54,992,793
48,194,667
45,392,729
44,053,149
42,970,784
41,043,266
40,309,146
34,186,523
$474,076
$383,668
$387,201
$338,732
$319,500
$331,456
0M
50M
100M
Opportunity Closed
$0
$800,000
Avg Deal Size
Revenue by Quarter
Q1
Q2
Q3
Q4
Revenue
$400M
$200M
$0M
$46M
$118M
$174M
$287M
$46M
$132M
$169M
$358M
$19M
$60M
$138M
$201M
$348M
$33M
$95M
$166M
$300M
$515M
2007
2008
2009
2010
2006
Share
Download

PHILADELPHIA CRIME SCENE

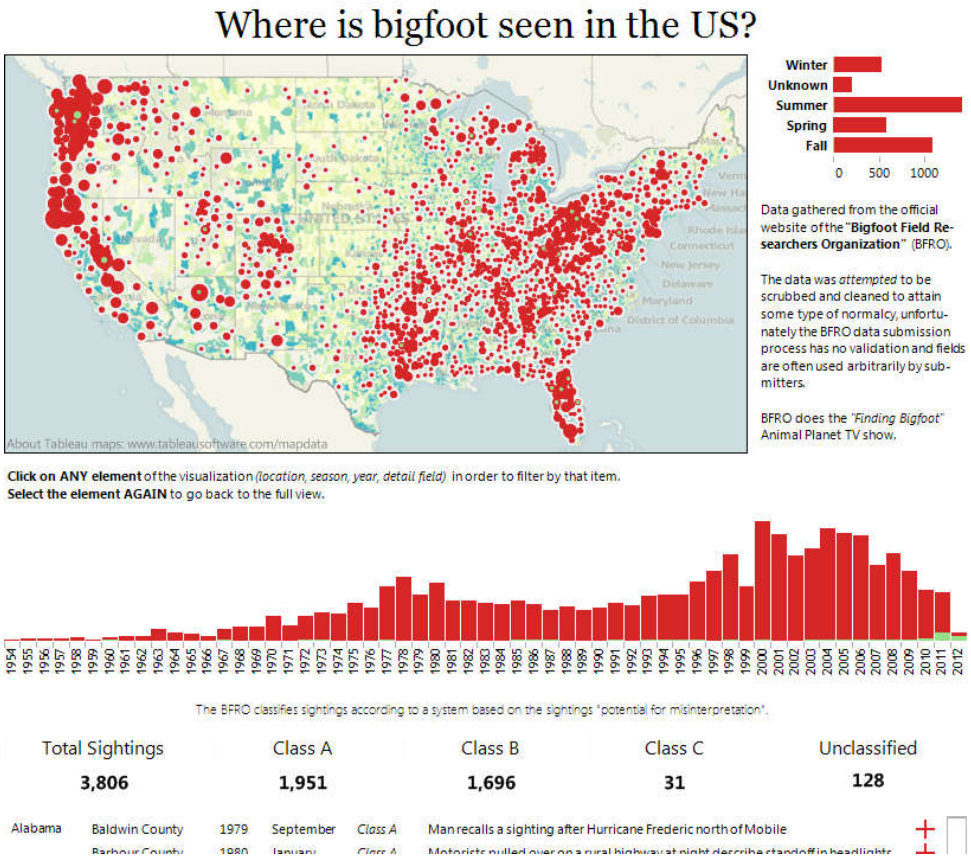
What makes a good dashboard?

- Answers a set of questions
- Follows a flow and invites interactivity
- Condensed; primarily in the form of summaries and exceptions
- Specific to and customized for the dashboard's audience and objectives
- Provides appropriate text for clarity and direction, if needed
- Makes strategic use of color

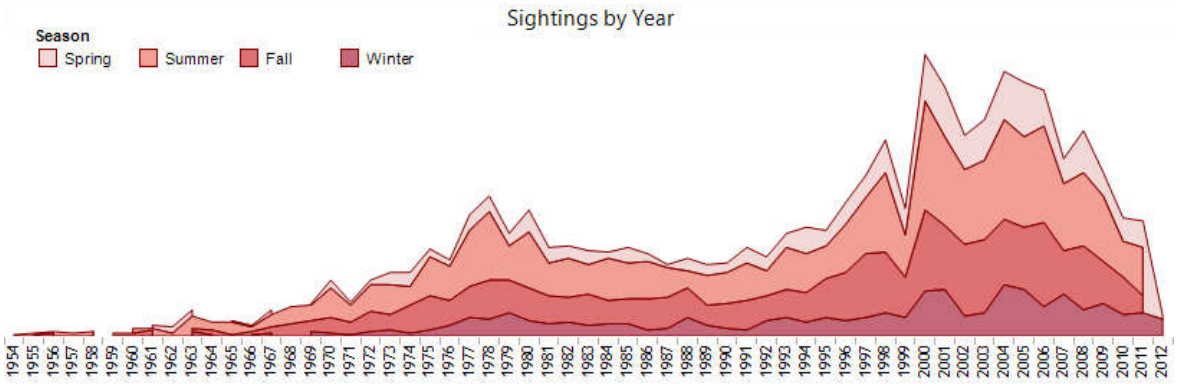
Layout

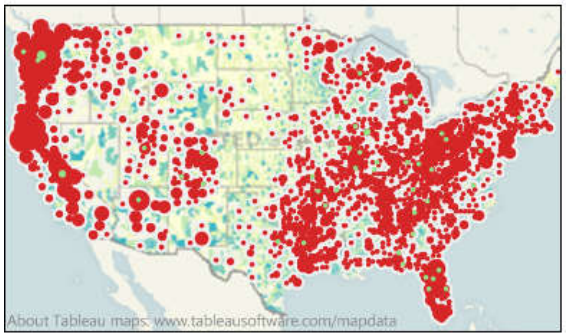
www.useit.com

Layout Top Left



Bigfoot sightings are in decline





Data gathered from the official website of the **"Bigfoot Field Researchers Organization"** (BFRO).

The data was *attempted* to be scrubbed and cleaned to attain some type of normalcy, unfortunately the BFRO data submission process has no validation and fields are often used arbitrarily by submitters.

BFRO does the *"Finding Bigfoot"* Animal Planet TV show.

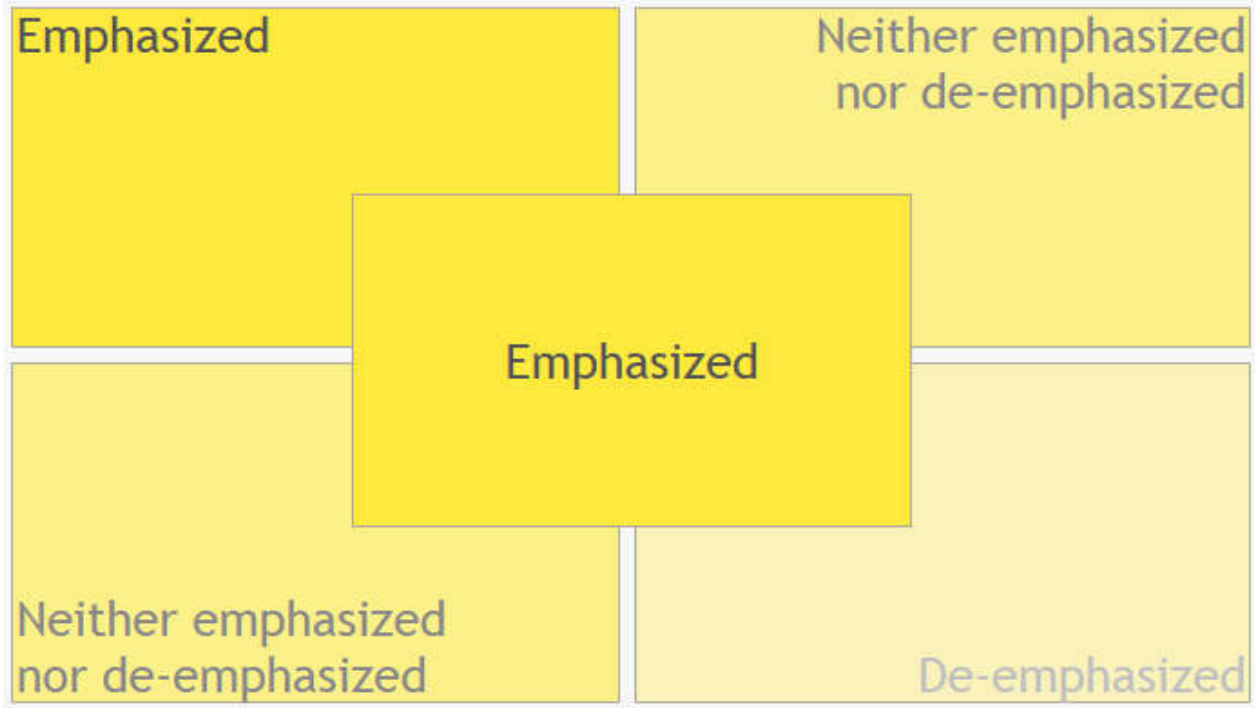
Click on ANY element of the visualization *(location, season, year, detail field)* in order to filter by that item.
Select the element AGAIN to go back to the full view.

The BFRO classifies sightings according to a system based on the sightings "potential for misinterpretation".

Total Sightings	Class A	Class B	Class C	Unclassified
3,806	**1,951**	**1,696**	**31**	**128**

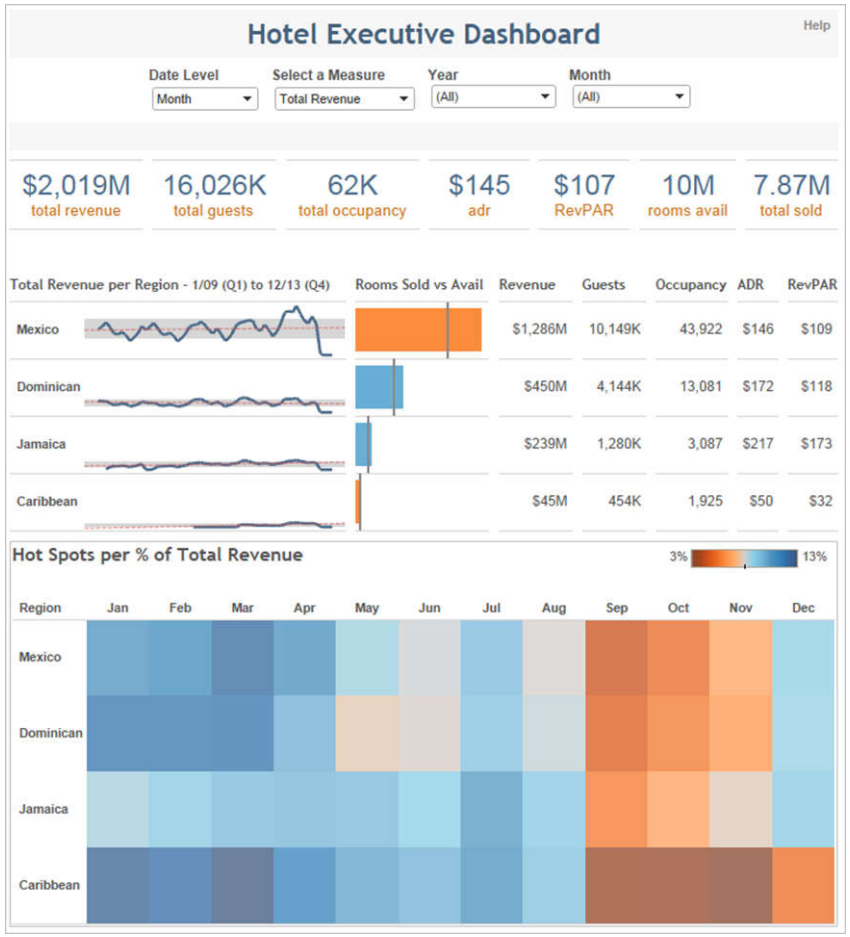
Alabama	Baldwin County	1979	September	*Class A*	Man recalls a sighting after Hurricane Frederic north of Mobile

What are you emphasizing?

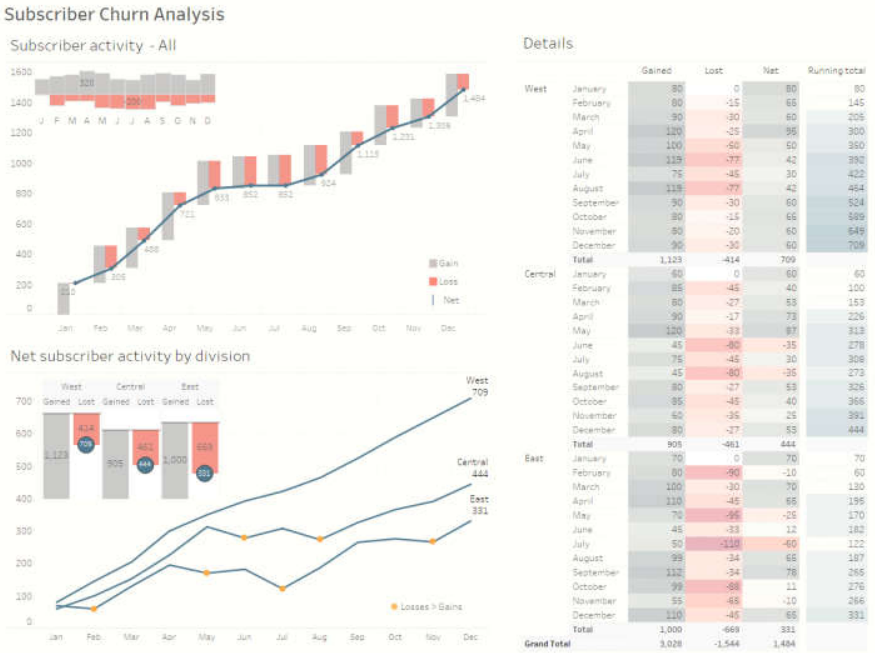


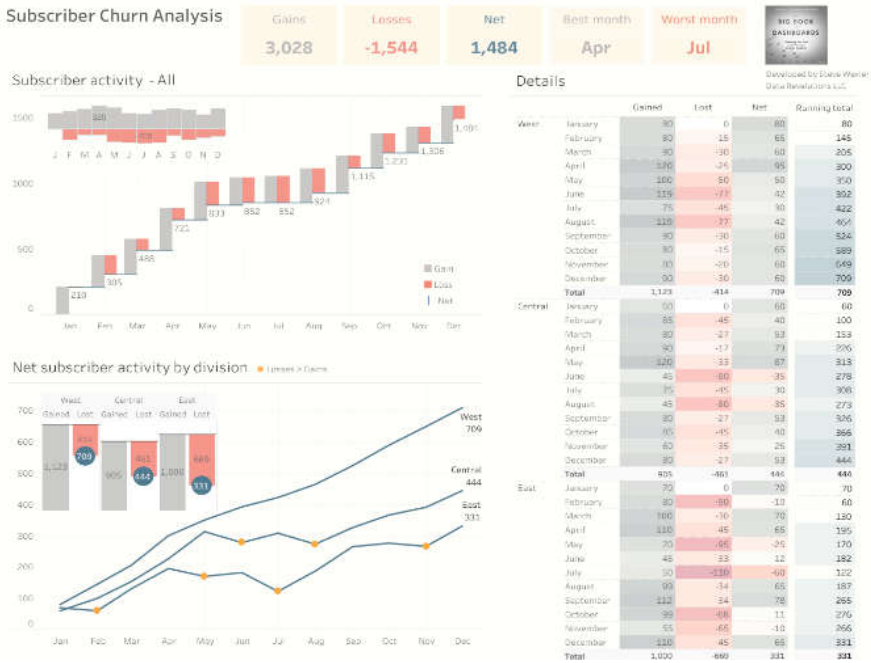
Dashboard Flow

- Filters at the top, so user knows when and how to start interacting
- Most important information at the top and centered
- Starts with the highest level and works down to more precise details
- No more than 4-5 worksheets



Using BANs


Subscriber Churn Analysis
Subscriber activity - All
Net subscriber activity by division
Details
Gained
Lost
Net
Running total
Gain
Loss
Net
West
709
Central
444
East
331
Losses > Gains


Subscriber Churn Analysis
Gains
3,028
Losses
-1,544
Net
1,484
Best month
Apr
Worst month
Jul
Subscriber activity - All
Net subscriber activity by division
Details
Gained
Lost
Net
Running total
Gain
Loss
Net
West
709
Central
444
East
331

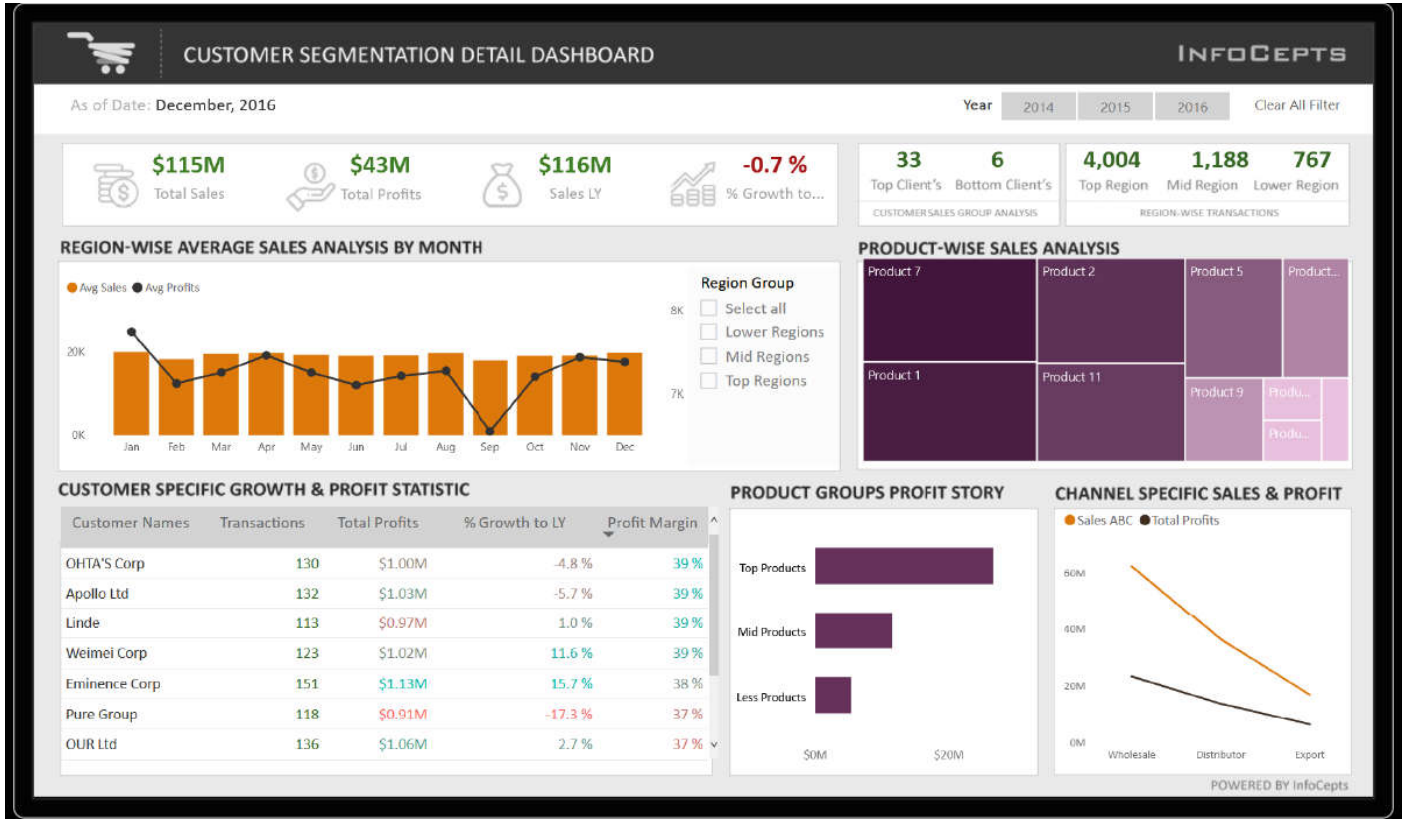
Does the dashboard pass the 5 second test?

- Most important view is on top or top left
- Legends are near their views
- Avoid multiple color schemes
- Uses 5 or fewer views (charts)

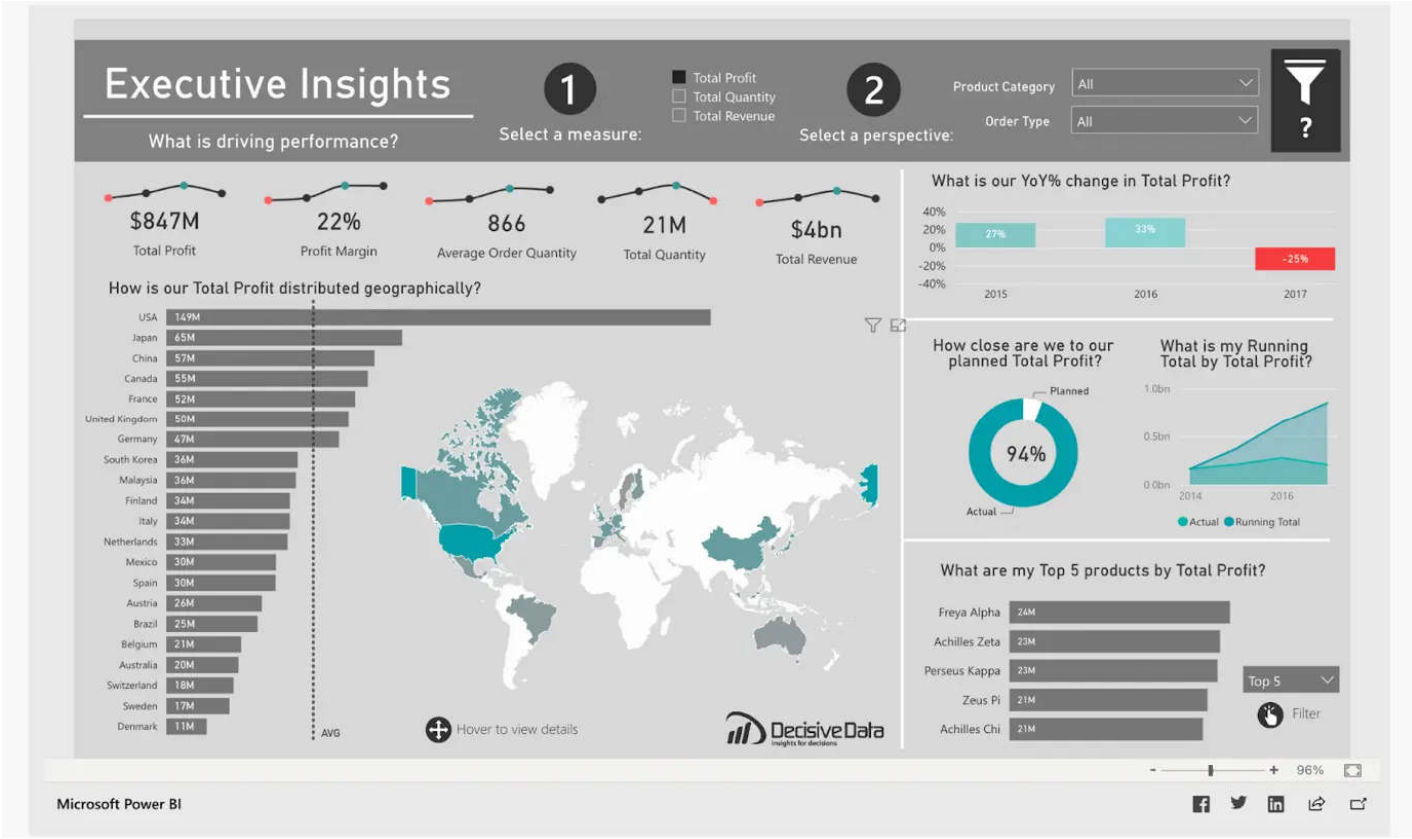
Dashboard Mistakes

- Not creating a clear goal for the dashboard
- Presenting too much information in one dashboard
- Including unnecessary information
- Representing data in the wrong type of graph or chart
- Using a cluttered, confusing, or inaccessible design
- Including too many colors
- Including decimals

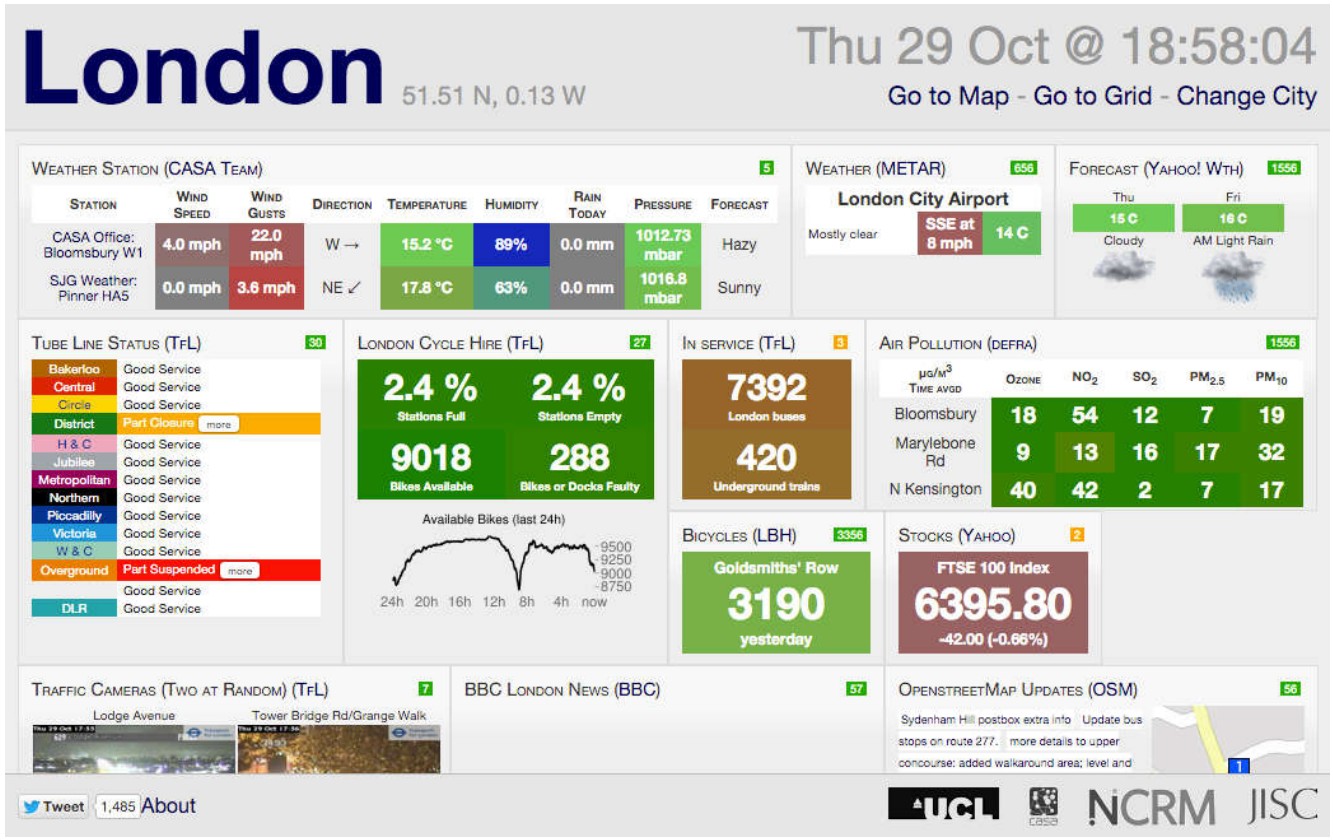
Example: Sales Dashboard



Example: Financial Dashboard



Example: Bad Dashboard Design

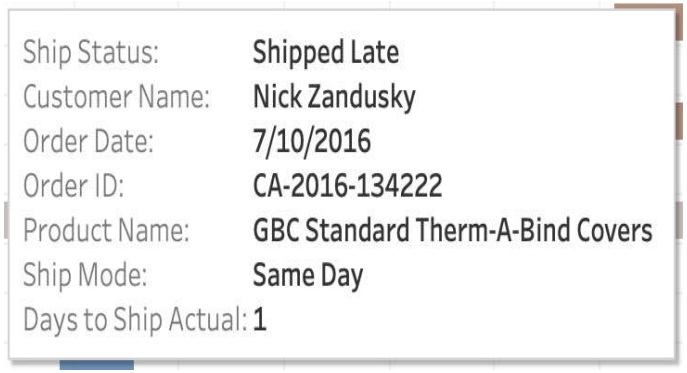


Using Tooltips

Tooltips can make the difference between a user loving your visualization and not understanding it.

Tips for improving the basic tooltip

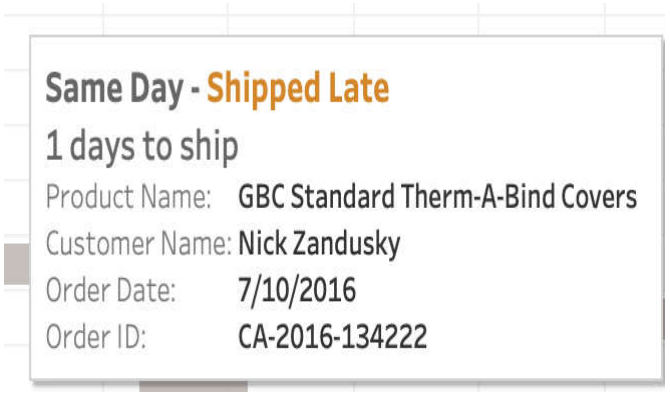
- Use a proper font
- Identify the most important part of the tool tip and make it your title
- Change measure names/values to make them specific and understandable
- Include proper units
- Remove command prompts



Default

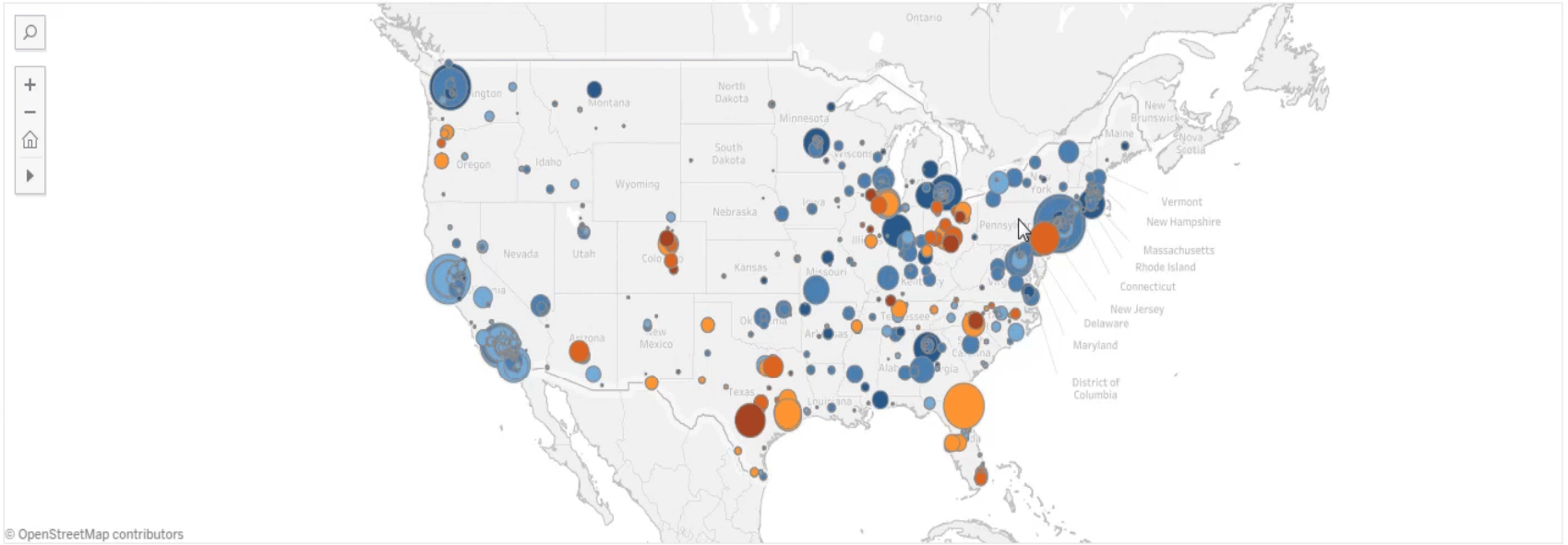
vs

Modified



Executive Overview - Profitability (All)

Sales	Profit	Profit Ratio	Profit per Order	Sales per Customer	Avg. Discount	Quantity
$1,820,895	$231,963	12.7%	$59.04	$2,313.72	15.71%	29,793



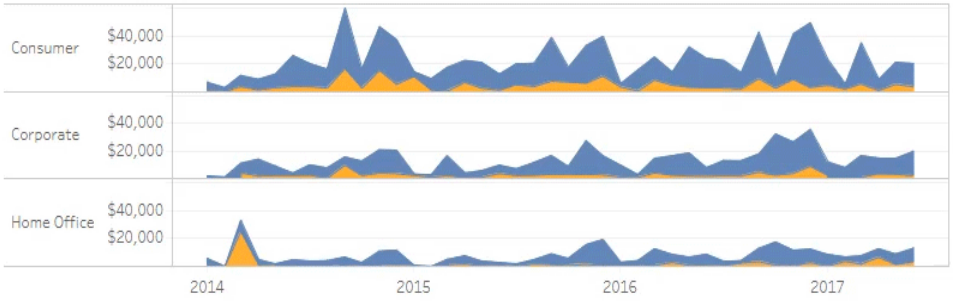
Region
(All)

Order Date
1/3/2014 6/30/2017

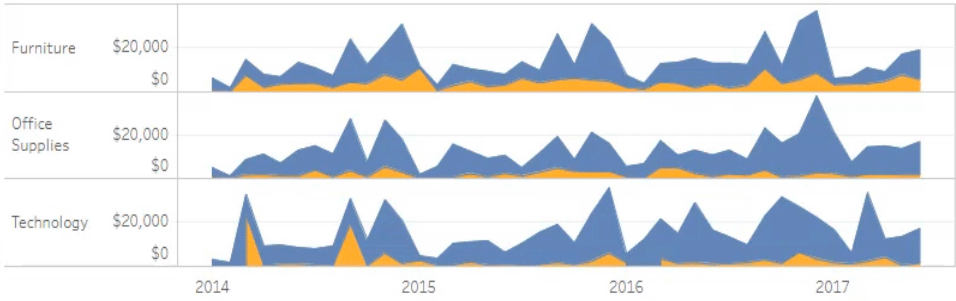
Profit Ratio
-270.0% 50.0%

Profit Ratio
-50.0% 50.0%

Monthly Sales by Segment - States: **All**



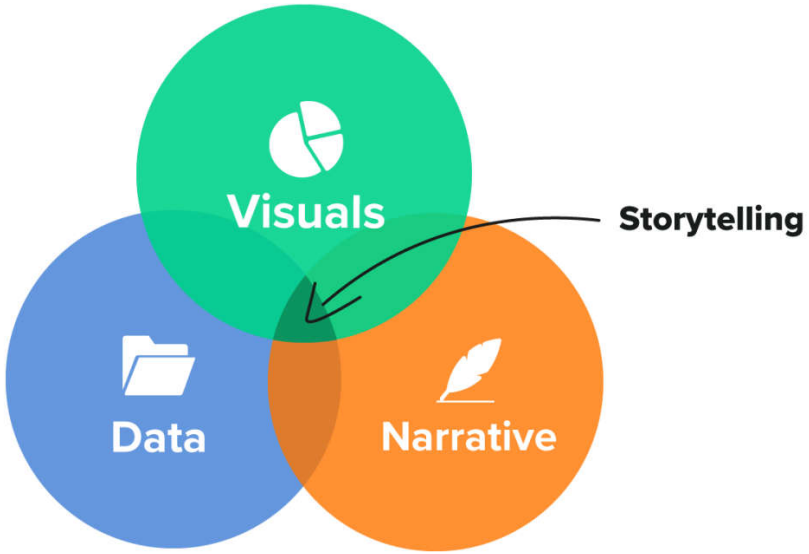
Monthly Sales by Product Category - States: **All**



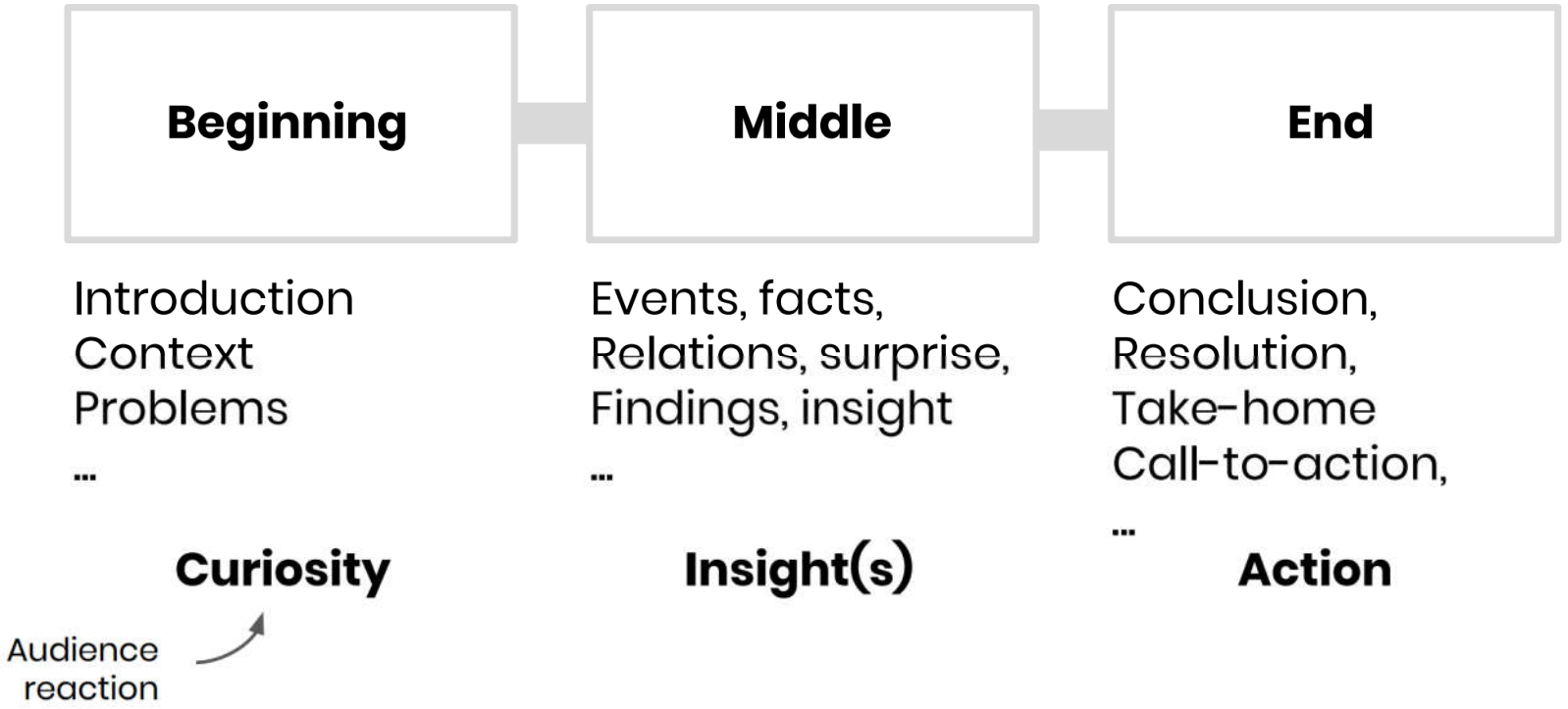
Storytelling

Data Storytelling

- Data visualization expert Stephen Few said, “Data storytelling is a structured approach for communicating data insights, and it involves a combination of three key elements: data, visuals, and narrative”.



Story Structure



The Beginning

- The first thing to do is introduce the plot, building the context for your audience.
- Set up the essential elements of story:
  - the setting
  - main character
  - unresolved state of affairs
  - desired outcome
- Involve your audience, stimulating their interest and answering the questions that are likely on their mind:
  - *Why should I pay attention?*
  - *What is in it for me*?

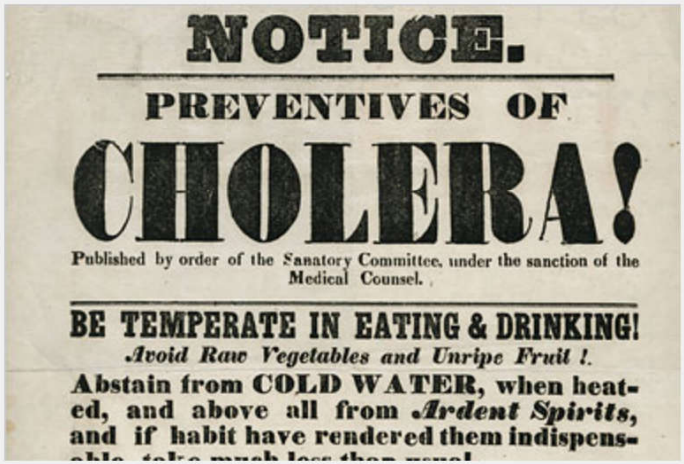
The Middle

- You retain your audience's attention through this part of the story by addressing *how* they can solve the problem you introduced.
  - Further develop the situation or problem by covering relevant background.
  - Incorporate external context or comparison points.
  - Give examples that illustrate the issue.
  - Include data that demonstrates the problem.
  - Articulate what will happen if no action is taken or no change is made.
  - Discuss potential options for addressing the problem.
  - Illustrate the benefits of your recommended solution.
  - Make it clear to your audience why they are in a unique position to make a decision or drive action.

The End

- End the story with a call to action:
  - make it totally clear to your audience what you want them to *do* with the new understanding or knowledge that you've imparted to them.
- One classic way to end a story is to tie it back to the beginning.
  - At the beginning of our story, we set up the plot and introduced the dramatic tension.
  - To wrap up, you can think about recapping this problem and the resulting need for action, reiterating any sense of urgency and sending your audience off ready to act.

London's 1854 Cholera Outbreak:
Data Mapping Halts an Epidemic



NOTICE.

PREVENTIVES OF

CHOLERA!

Published by order of the Sanatory Committee, under the sanction of the Medical Counsel.

BE TEMPERATE IN EATING & DRINKING!

Avoid Raw Vegetables and Unripe Fruit !.

Abstain from COLD WATER, when heated, and above all from *Ardent Spirits*, and if habit have rendered them indispens-

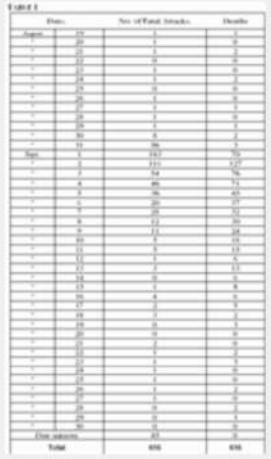
Victorian medical advice was frequently off the mark by modern standards.

In 1854, a Cholera outbreak swept through the Soho neighborhood of London.

616 people died.

Physician John Snow was skeptical of existing theories of disease transmission, which often blamed "miasmas," or bad air. The germ theory of disease circulation had not yet been outlined.

London's 1854 Cholera Outbreak:
Data Mapping Halts an Epidemic



This is Snow's original table showing the chronology of deaths and their total. He notes that not all deaths have recorded addresses; thus, his data set is partially incomplete.

Dr. Snow canvassed the neighborhood. He collected the addresses of those who died, noted the number of deaths at each location, and tabulated the results.

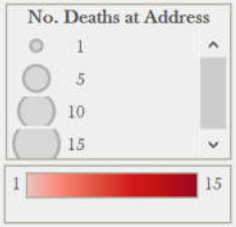
London's 1854 Cholera Outbreak:
Data Mapping Halts an Epidemic

An Outbreak Begins | Collecting the Data | **Mapping the Results** | Snow's Analysis: Focus on Broad St. | Ending an Epidemic

Then, Snow mapped these data points onto a map of the Soho neighborhood.

The results were startling.

NB: The map reprinted here is Snow's original. I geocoded the addresses, introduced the number of deaths as a dimension in the dataset, and uploaded the map as a background image. Then, I conformed the dimensions of the map to the appropriate lat/long. Thus, the locations appear as they would have on Snow's map.



London's 1854 Cholera Outbreak:
Data Mapping Halts an Epidemic

He noted that the bulk of deaths were concentrated in an area that generally used the same municipal water pump: Broad Street.

Snow realized that the Broad Street water pump--and poor water quality--was likely the source of the outbreak.

Armed with this information, he went to city ..

London's 1854 Cholera Outbreak:
Data Mapping Halts an Epidemic

Snow advised that the Broad Street water pump be shut down immediately. City authorities found Snow's evidence persuasive, and they removed the handle from the pump.

The outbreak ended shortly thereafter.

John Snow's pioneering data analysis proved that cholera is a water-borne disease, and paved the way for the field of epidemiology.

It is one of the foundational cases of the systematic use of public health information to save lives.