



Data Visualization

Eman Raslan 

Course RoadMap

Session 1

Session 2



Lab 1

Lab 2



Lab 3

Lab 4



Lab 5

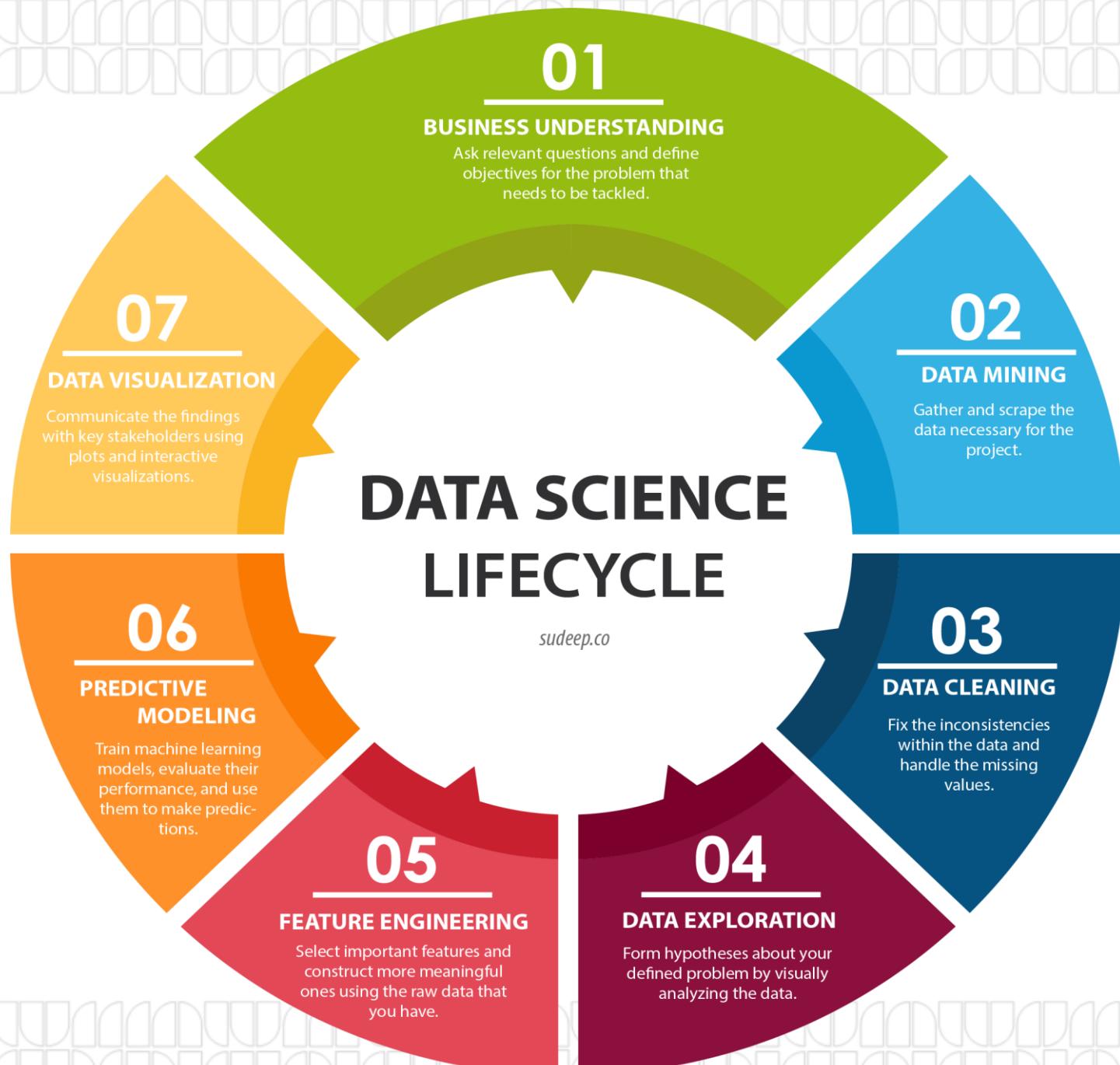
Lab 6



Project Deliver

What Is DATA VISUALIZATION?





Data Visualization in Data Science



The First Mile ...

Data Exploration, Exploratory Data Analysis

Making many interactive graphics quickly for:

- ‘Getting a sense’ of the data.
- Looking for patterns & outliers.
- Validating data quality.

... and the Last Mile.

Communication, Explanation, Operationalization

Making fewer, more polished interactive graphics for:

- Publication in articles or on the web.
- Non-data- scientists to interact with outputs.
- Making models actionable.



Let's **PLAY** Together

Which **MONTH** had the **HIGHEST** sales?

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

\$44,703	\$20,301	\$58,872	\$36,522	\$44,261	\$52,982	\$45,264	\$63,121	\$87,867	\$77,777	\$118,448	\$83,829
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	----------

3...2...1...

NOOOOO! NOOOOOOO!

I WASN'T READYYYYYY!!!!

quickmeme.com

Can you name the **TOP 3** performing months?

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

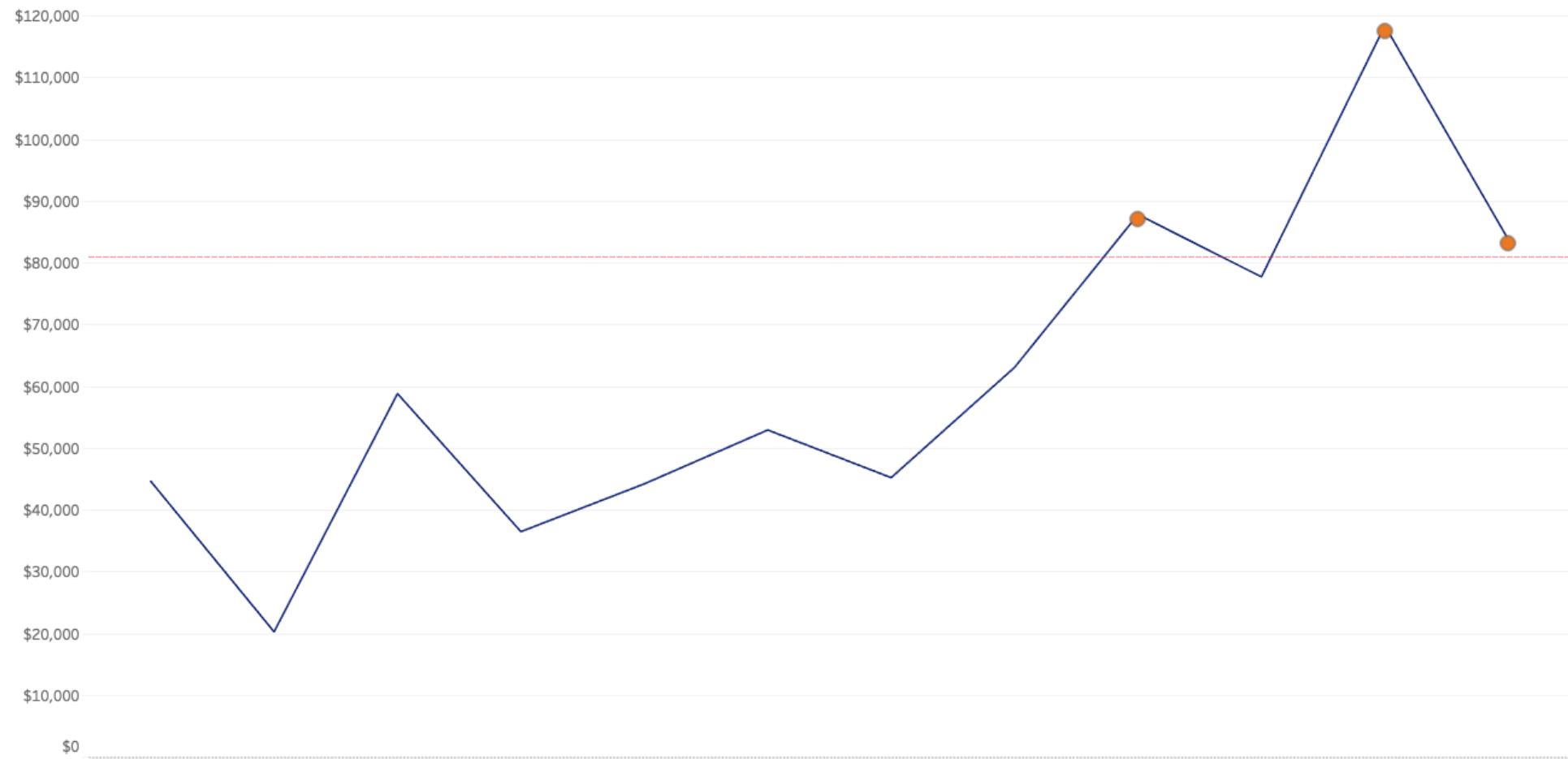
\$44,703	\$20,301	\$58,872	\$36,522	\$44,261	\$52,982	\$45,264	\$63,121	\$87,867	\$77,777	\$118,448	\$83,829
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------	-----------	----------

3...2...1...

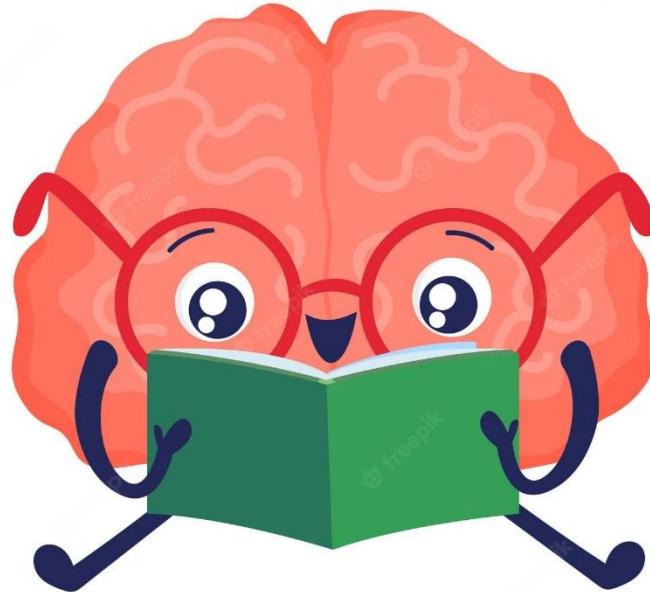
**THAT MOMENT WHEN YOUR
TEACHER CALLS ON YOU**

AND YOUR NOT PAYING ATTENTION

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
\$44,703	\$20,301	\$58,872	\$36,522	\$44,261	\$52,982	\$45,264	\$63,121	\$87,867	\$77,777	\$118,448	\$83,829



THE



HAS
LIMITS!!!

Memory Limits

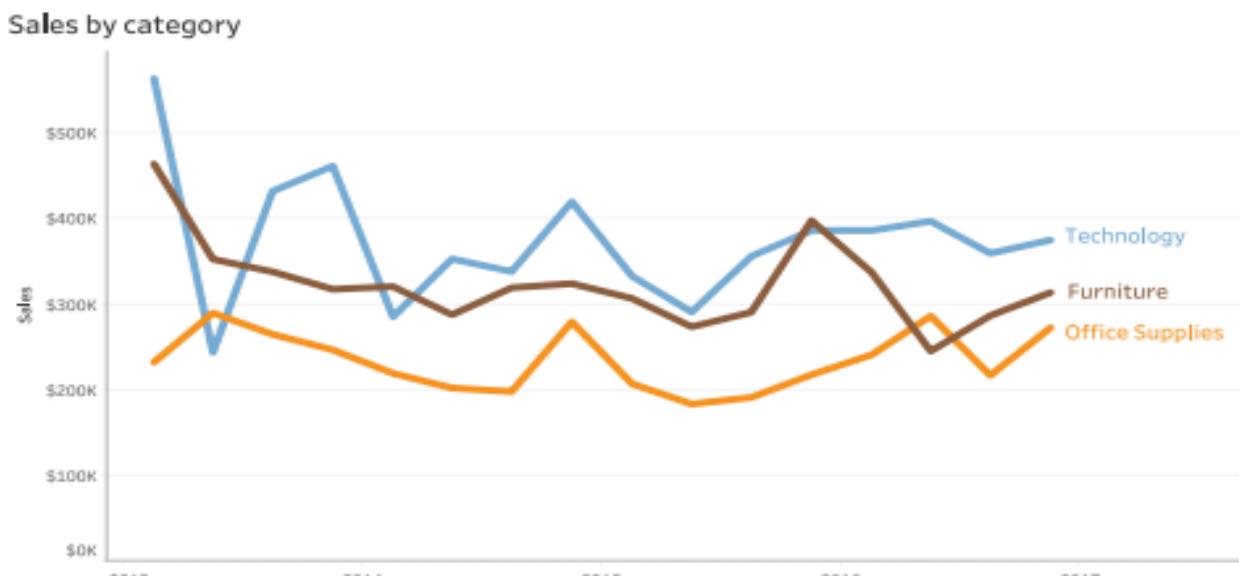
48 numbers



TABLE 1.2 What are the trends in sales?

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

3 Patterns



Why Do We Visualize Data?

- Making data engaging and easily digestible.
- Identifying **relationships**, **trends** and **outliers** within a set of data.
- Telling a **story** found within the data.
- Reinforcing an argument or opinion.
- Highlighting the important parts of a set of data.

How Do We Visualize Data?

Thinking is Slow. Seeing is FAST



How many 9s are there?

2 2 5 6 7 1 1 6 9 1
9 1 7 5 5 5 6 2 5 9
4 5 2 9 6 9 7 6 4 6
8 1 5 7 8 5 6 6 6 7
7 2 3 6 8 9 1 7 9 1
3 8 6 8 4 5 6 9 4 5
4 9 9 2 3 7 1 9 1 2
3 7 8 1 6 1 5 6 1 6
5 6 6 8 6 6 9 1 2 6
3 2 4 2 6 9 4 2 7 1

2 2 5 6 7 1 1 6 9 1
9 1 7 5 5 5 6 2 5 9
4 5 2 9 6 9 7 6 4 6
8 1 5 7 8 5 6 6 6 7
7 2 3 6 8 9 1 7 9 1
3 8 6 8 4 5 6 9 4 5
4 9 9 2 3 7 1 9 1 2
3 7 8 1 6 1 5 6 1 6
5 6 6 8 6 6 9 1 2 6
3 2 4 2 6 9 4 2 7 1

2 2 5 6 7 1 1 6 9 1
9 1 7 5 5 5 6 2 5 9
4 5 2 9 6 9 7 6 4 6
8 1 5 7 8 5 6 6 6 7
7 2 3 6 8 9 1 7 9 1
3 8 6 8 4 5 6 9 4 5
4 9 9 2 3 7 1 9 1 2
3 7 8 1 6 1 5 6 1 6
5 6 6 8 6 6 9 1 2 6
3 2 4 2 6 9 4 2 7 1

What If Our Task Is To Count The Frequency Of Each Digit?

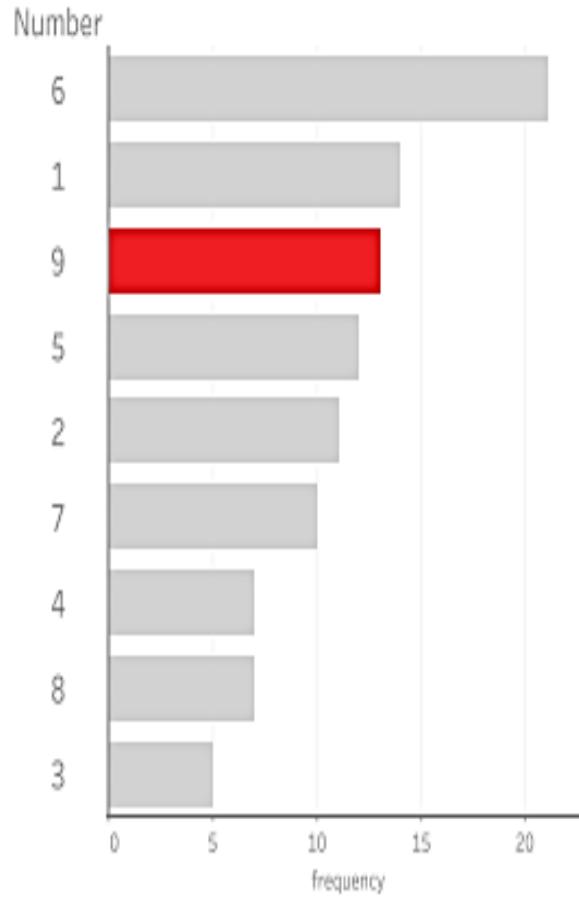
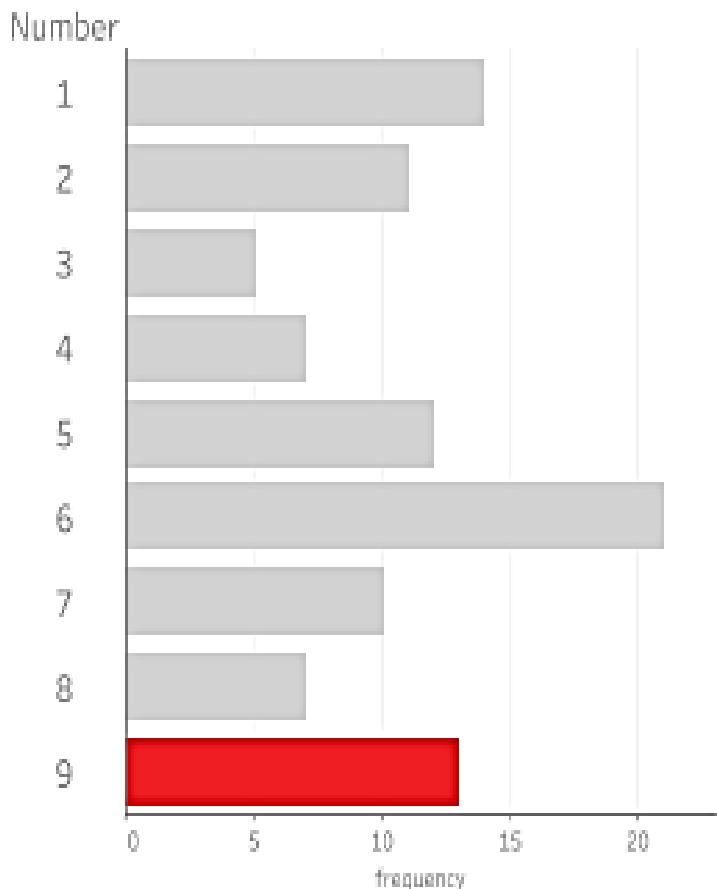
2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1



2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	9	7	6
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

What If Our Task Is To Count The Frequency Of Each Digit?

2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

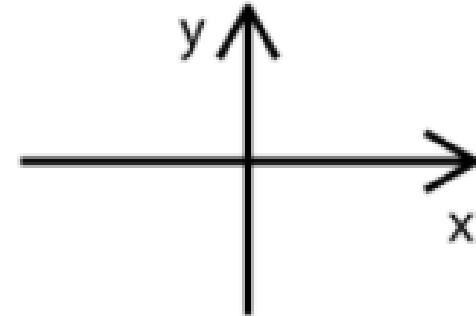


Pre-Attentive Attributes

- 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.
- Pre-attentive attributes provide us with ways to **encode** our **data** in **charts**.

Pre-Attentive Attributes Types

position

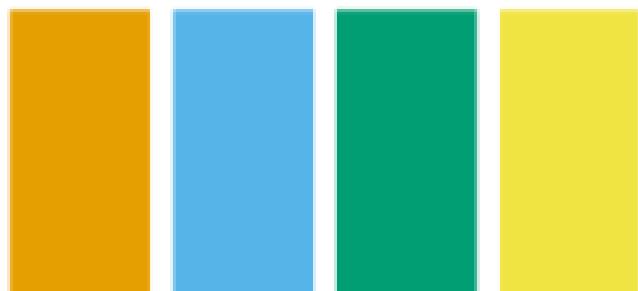


shape



size

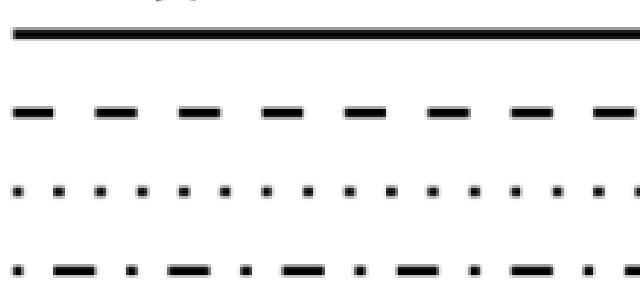
color



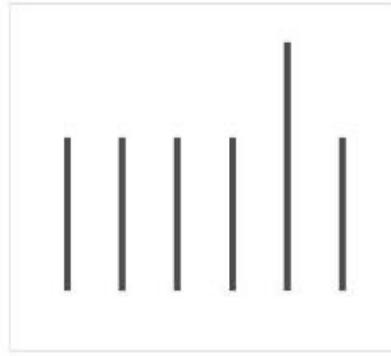
line width



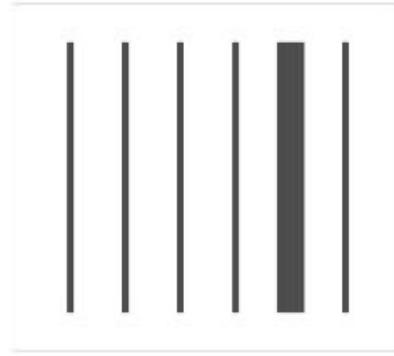
line type



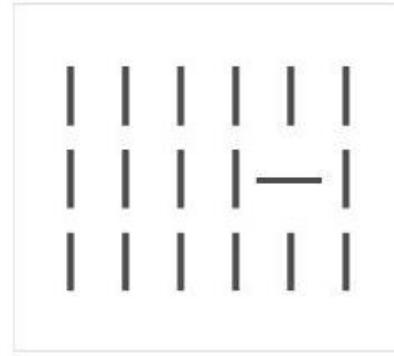
Pre-Attentive Attributes Types



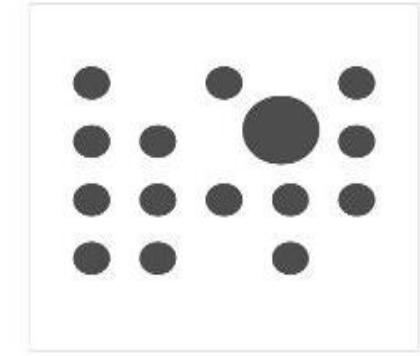
Length



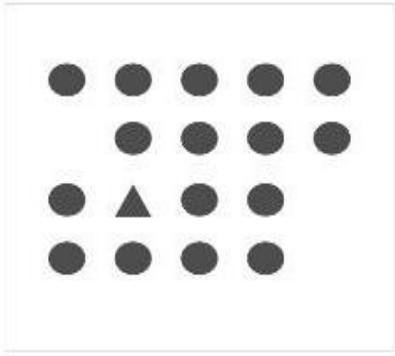
Width



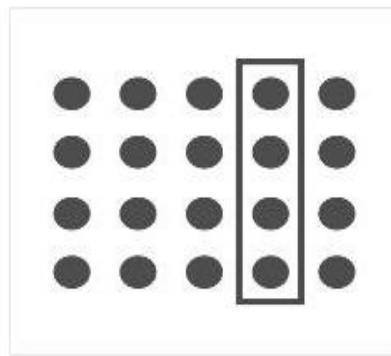
Orientation



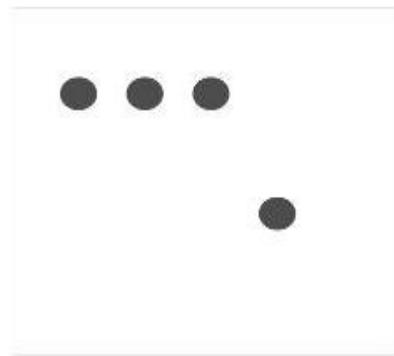
Size



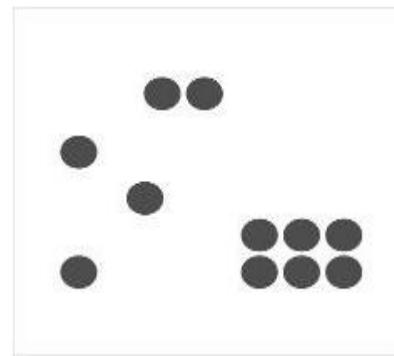
Shape



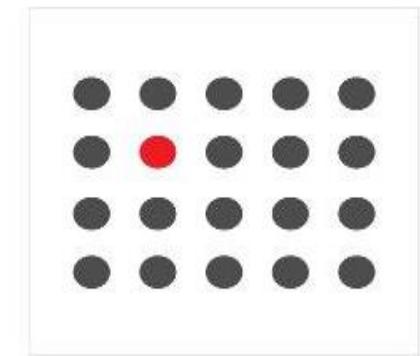
Enclosure



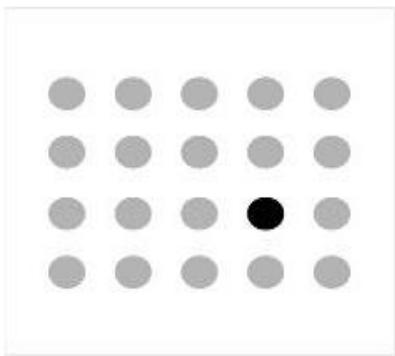
2D Position



Grouping



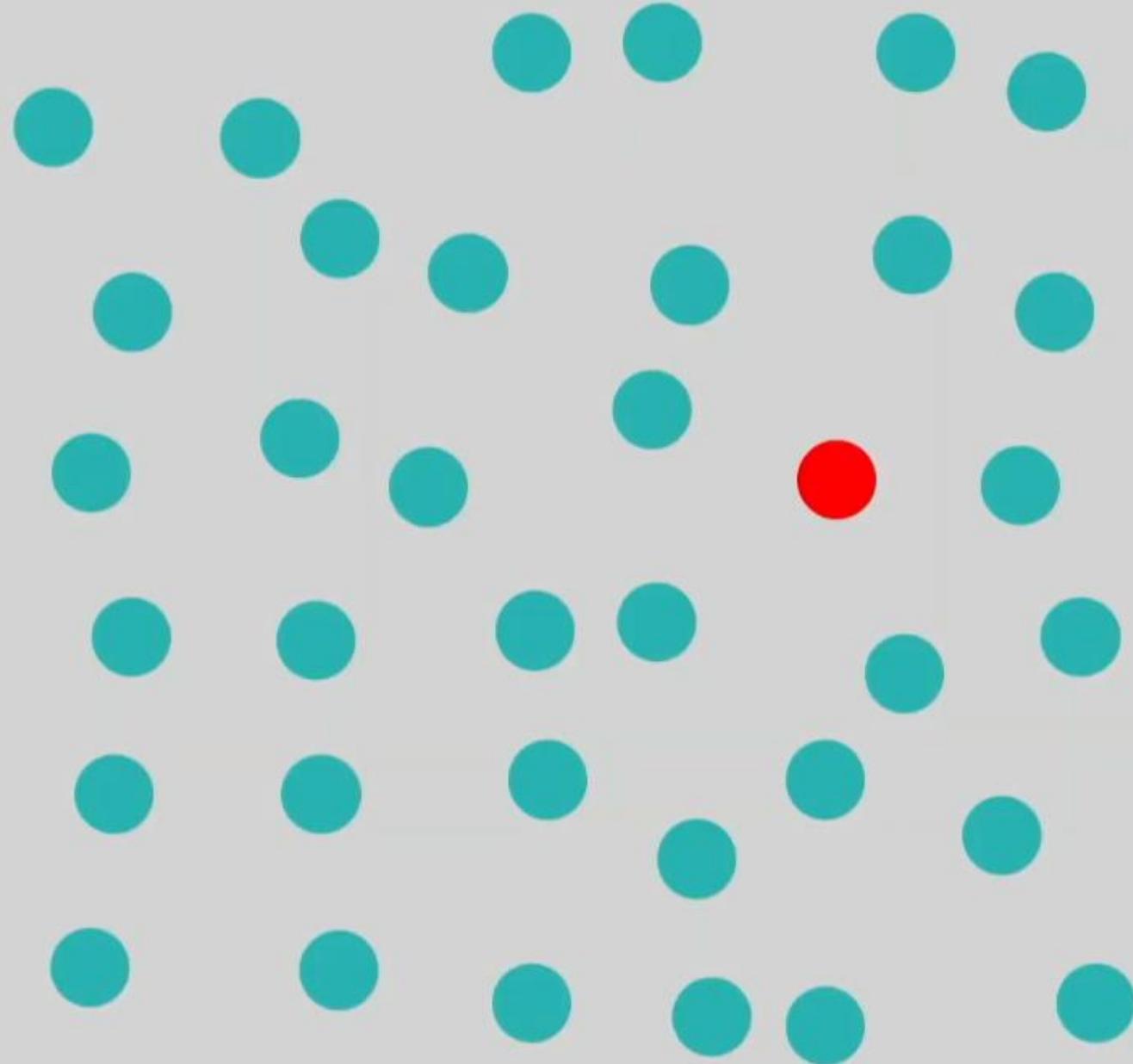
Color (Hue)



Color (Intensity)

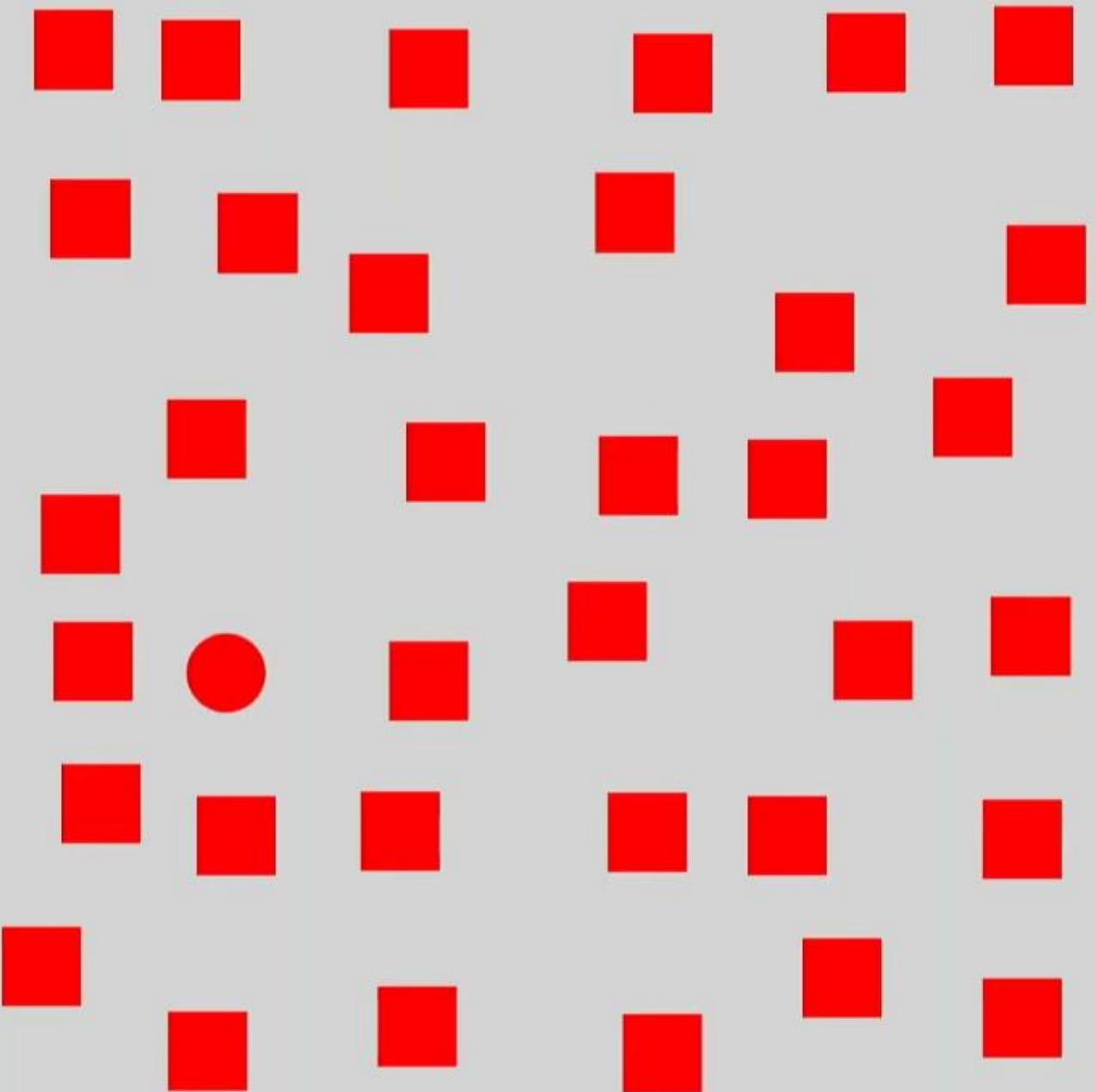
What Is Different?

Let's **P** **L** **A** **Y** Together



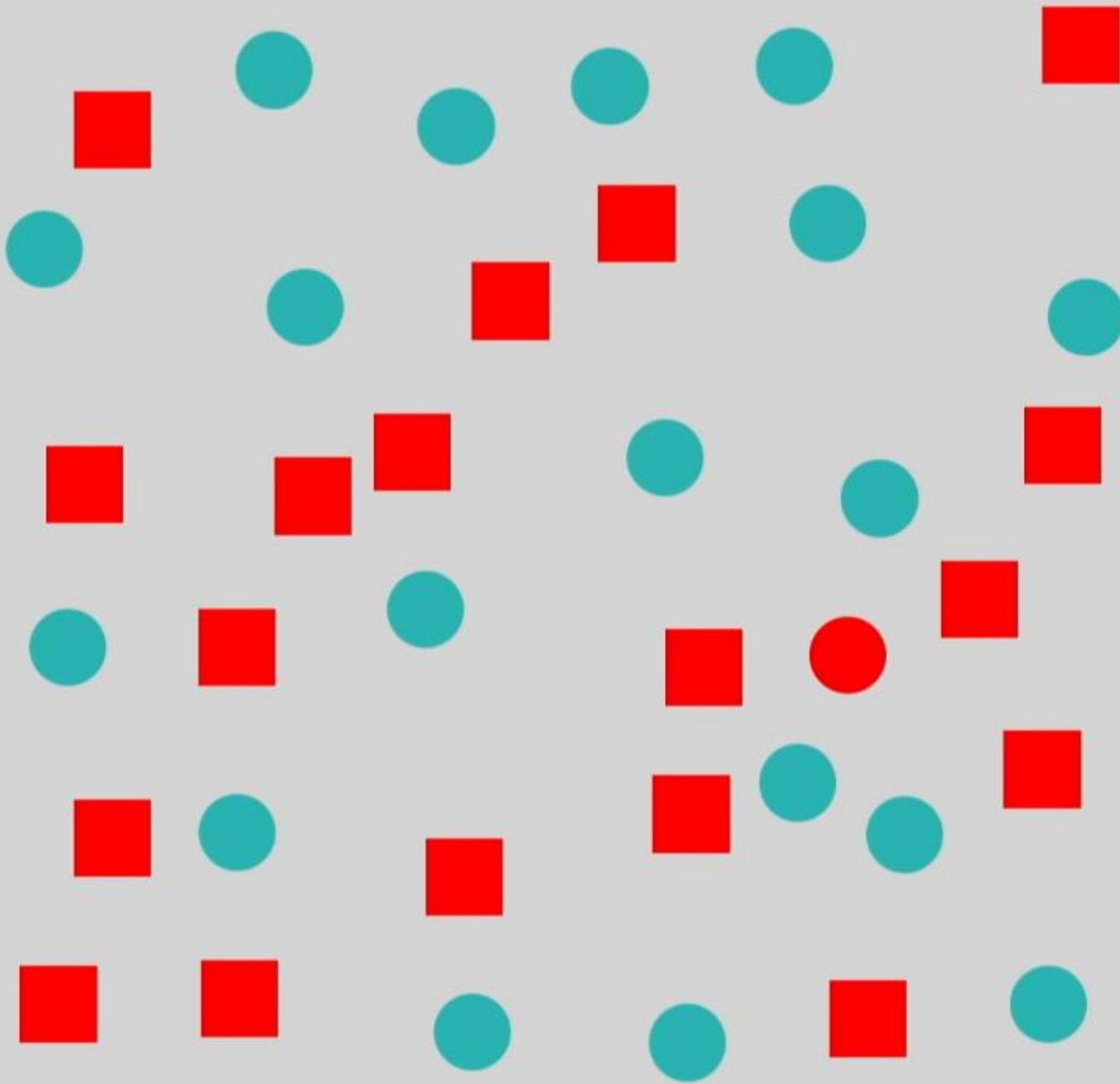
What Is Different?

Let's **P** **L** **A** **Y** Together



What Is Different?

Let's **P** **L** **A** **Y** Together



How Do Humans Like Their Data?

POSITION
COLOR
SIZE
SHAPE

More important



Less important

Data Types

Let's **PLAY** Together

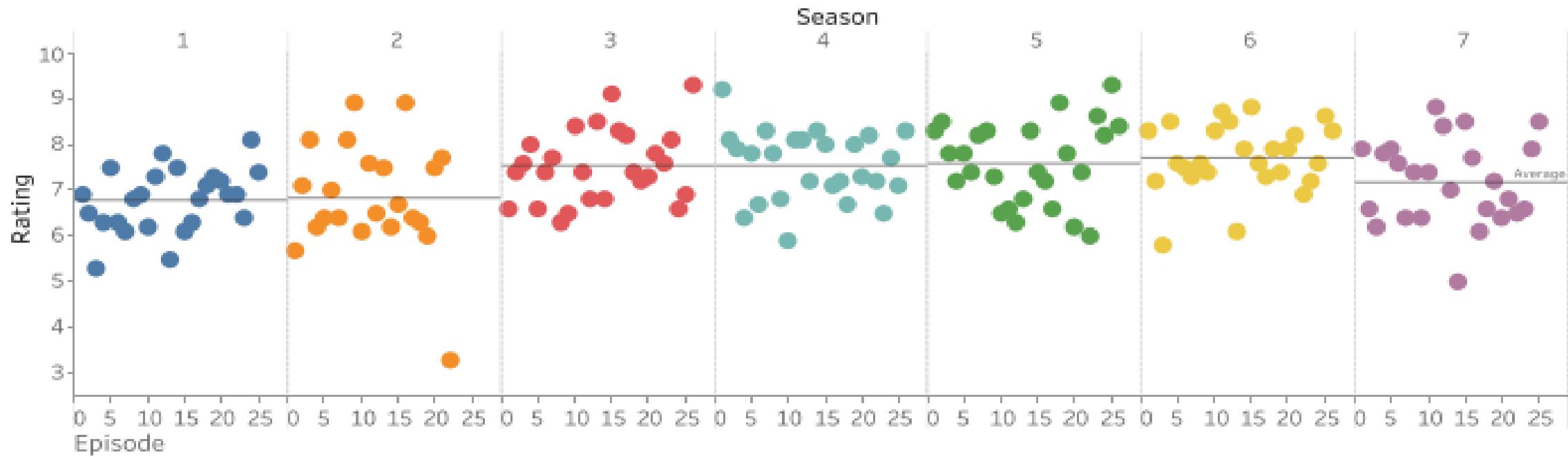
Type of variable	Examples	Appropriate scale	Description
quantitative/numerical continuous	1.3, 5.7, 83, 1.5×10^{-2}	continuous	Arbitrary numerical values. These can be integers, rational numbers, or real numbers.
quantitative/numerical discrete	1, 2, 3, 4	discrete	Numbers in discrete units. These are most commonly but not necessarily integers. For example, the numbers 0.5, 1.0, 1.5 could also be treated as discrete if intermediate values cannot exist in the given dataset.
qualitative/categorical unordered	dog, cat, fish	discrete	Categories without order. These are discrete and unique categories that have no inherent order. These variables are also called <i>factors</i> .
qualitative/categorical ordered	good, fair, poor	discrete	Categories with order. These are discrete and unique categories with an order. For example, "fair" always lies between "good" and "poor". These variables are also called <i>ordered factors</i> .
date or time	Jan. 5 2018, 8:03am	continuous or discrete	Specific days and/or times. Also generic dates, such as July 4 or Dec. 25 (without year).
text	The quick brown fox jumps over the lazy dog.	none, or discrete	Free-form text. Can be treated as categorical if needed.

TV Series Dataset

- 7 season.
- Each season has 25 episode.
- Each episode has a rating.



Encoding Data in Charts



Data	Encoding
Episode Number	Position
Season	Color
	Position
IMDB rating	Position
Average season rating	Position

Terrorism Dataset

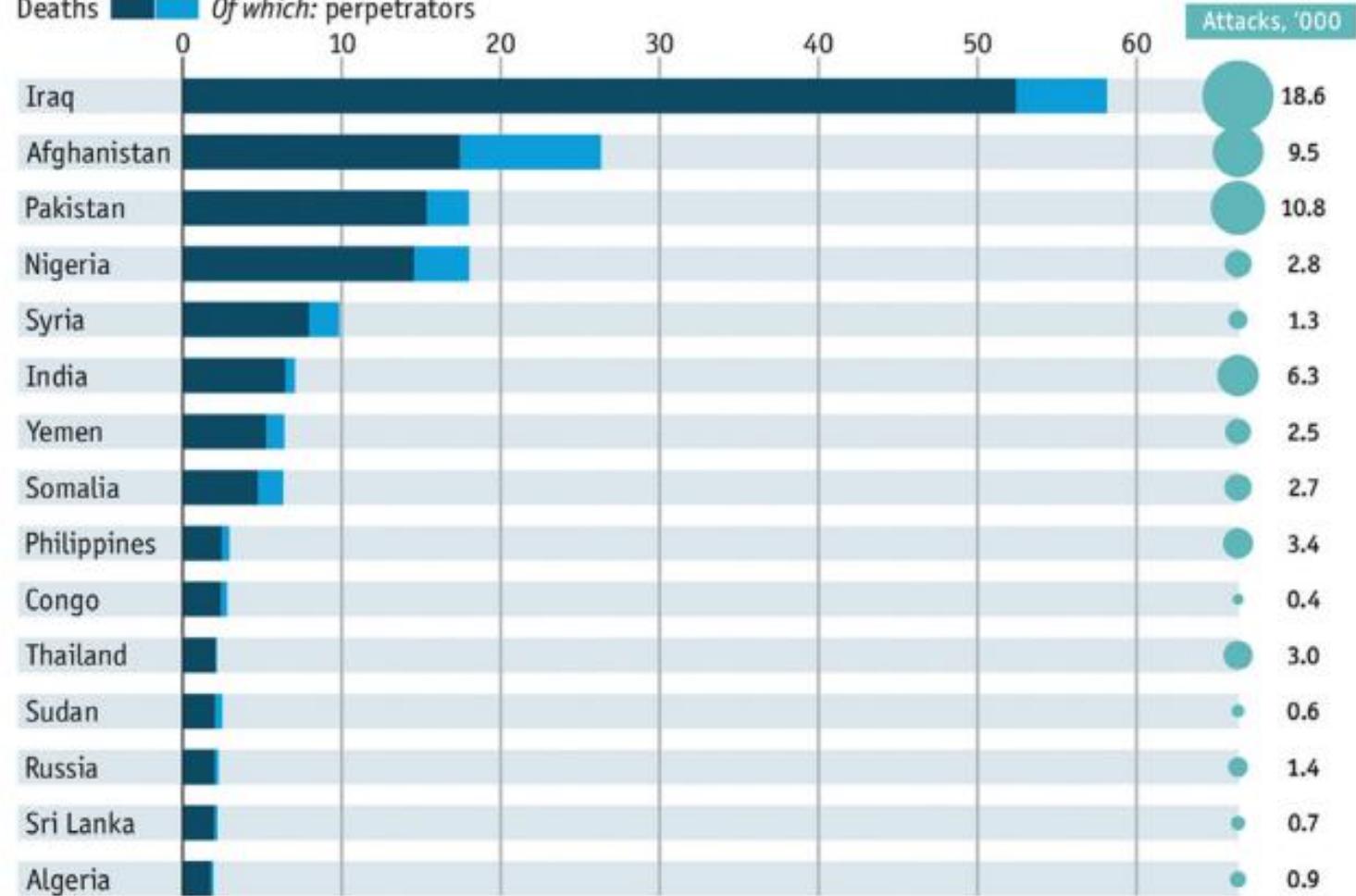


Encoding Data in Charts

A terrible record

Countries with highest number of deaths from terrorism, 2003-15, '000

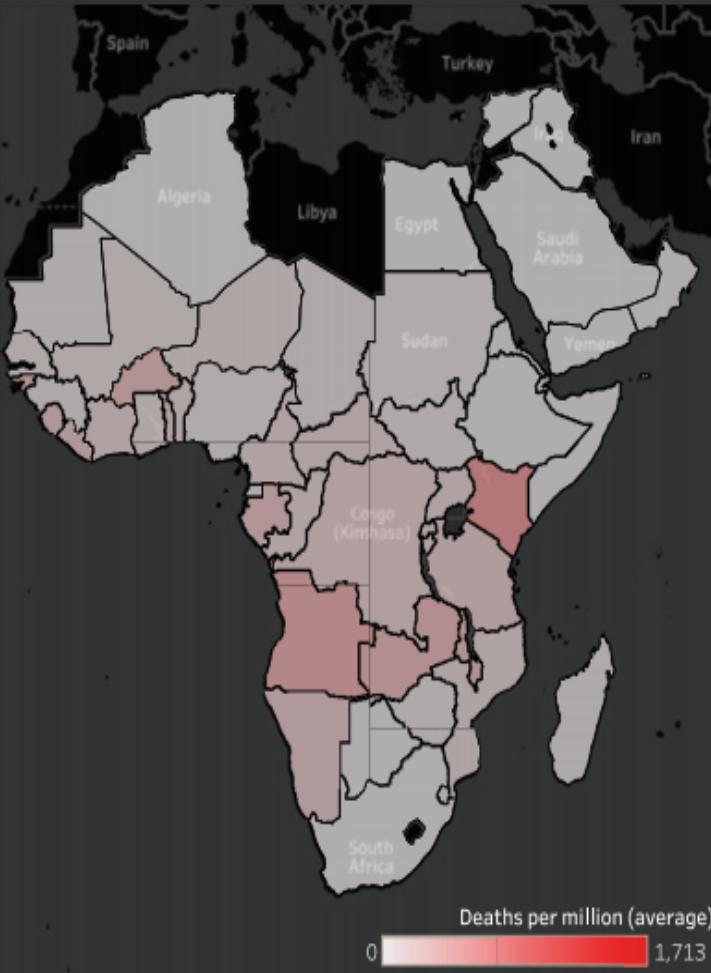
Deaths — Of which: perpetrators



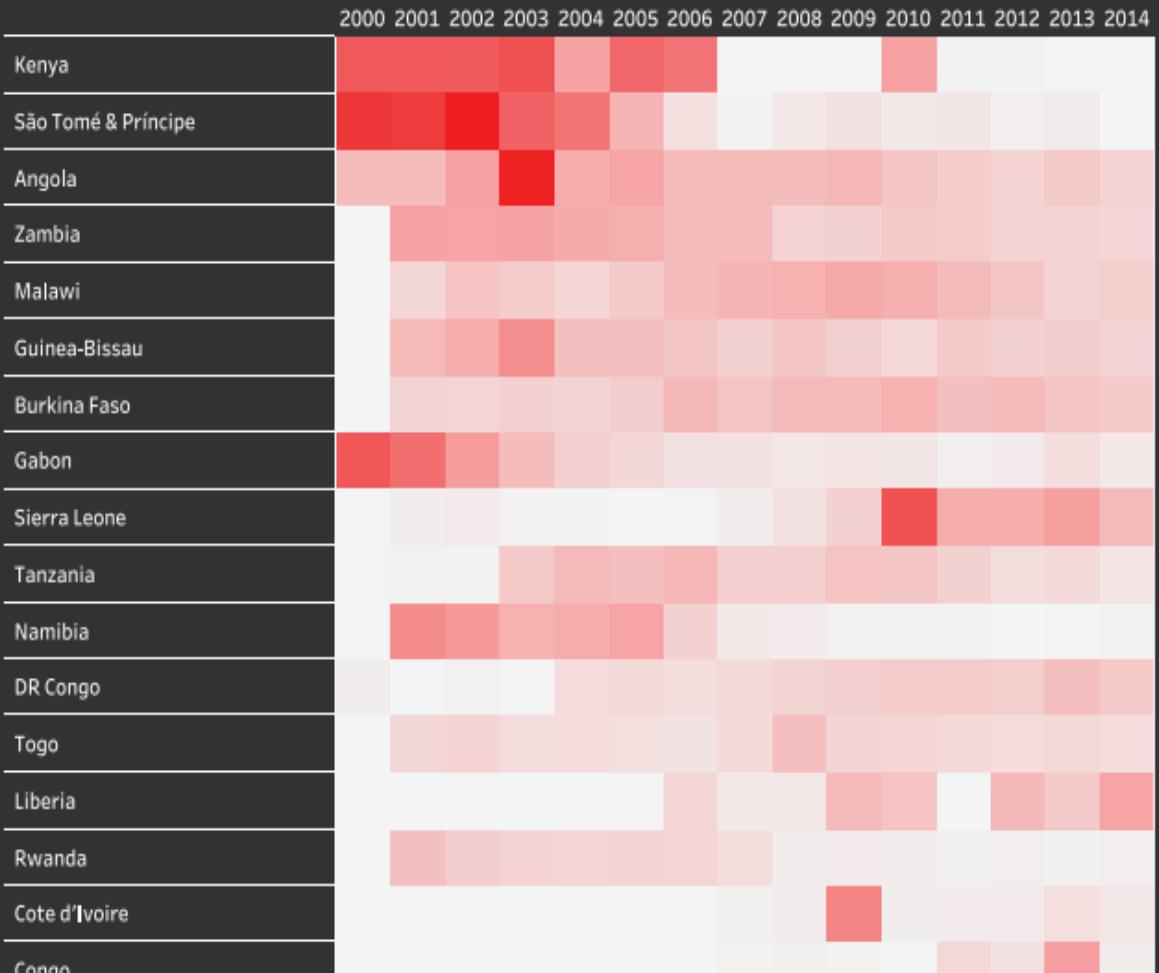
Data	Encoding
Country	Position
Deaths	Length
Death type	Color
Attacks	Size

Encoding Data in Charts

Malaria deaths in Africa, 2000-2014



The 17 most affected countries



Data	Encoding
Country	Position
Deaths per million	Color
Year	Position

Use

Color

of



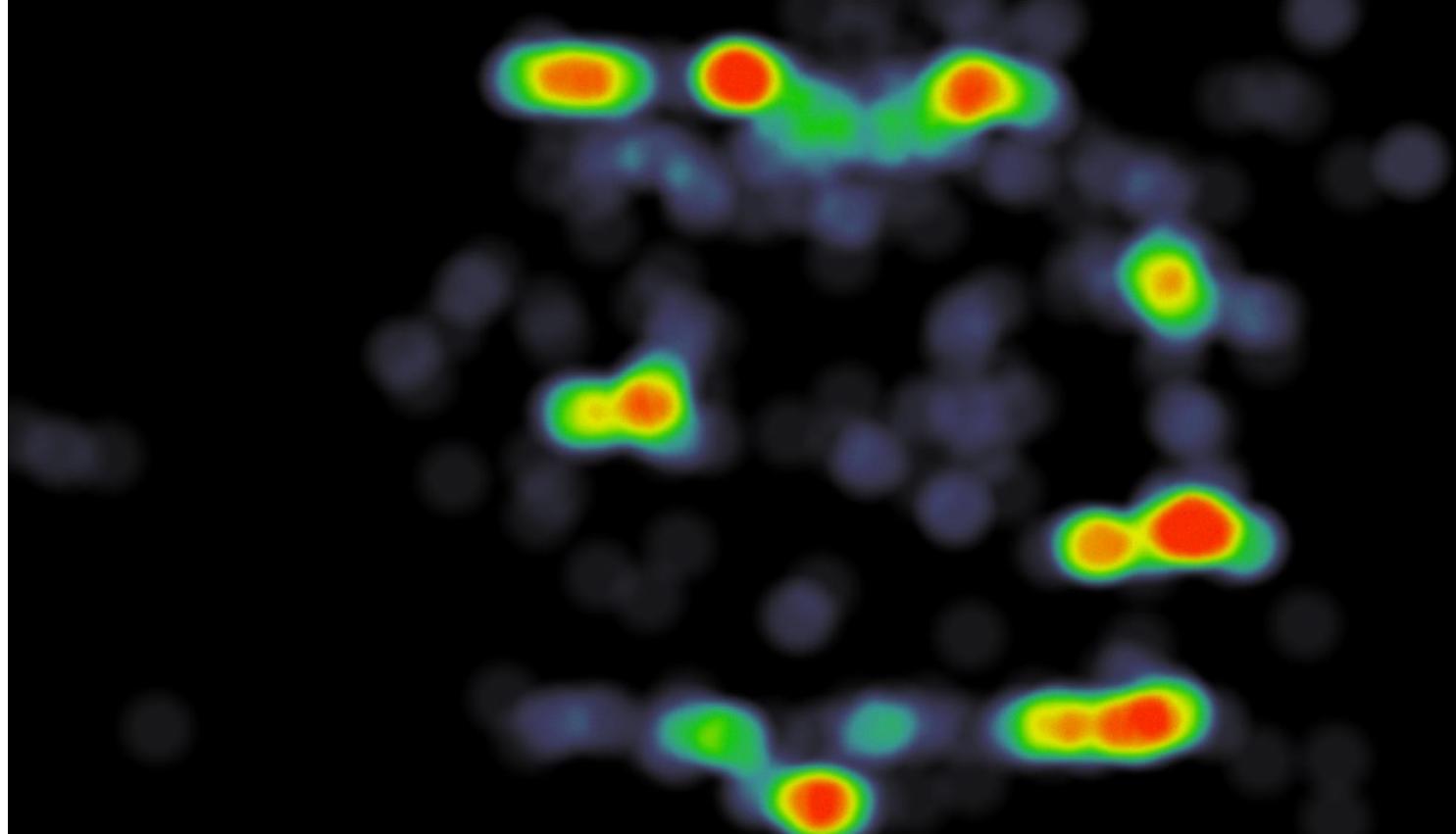


Colors in Data Visualizations



- ❖ There are three fundamental use cases for color in data visualizations:
 - **Distinguish** groups of data from each other
 - **Represent** data values
 - **Highlight**.

Colors in Data Visualizations



Color Scales

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

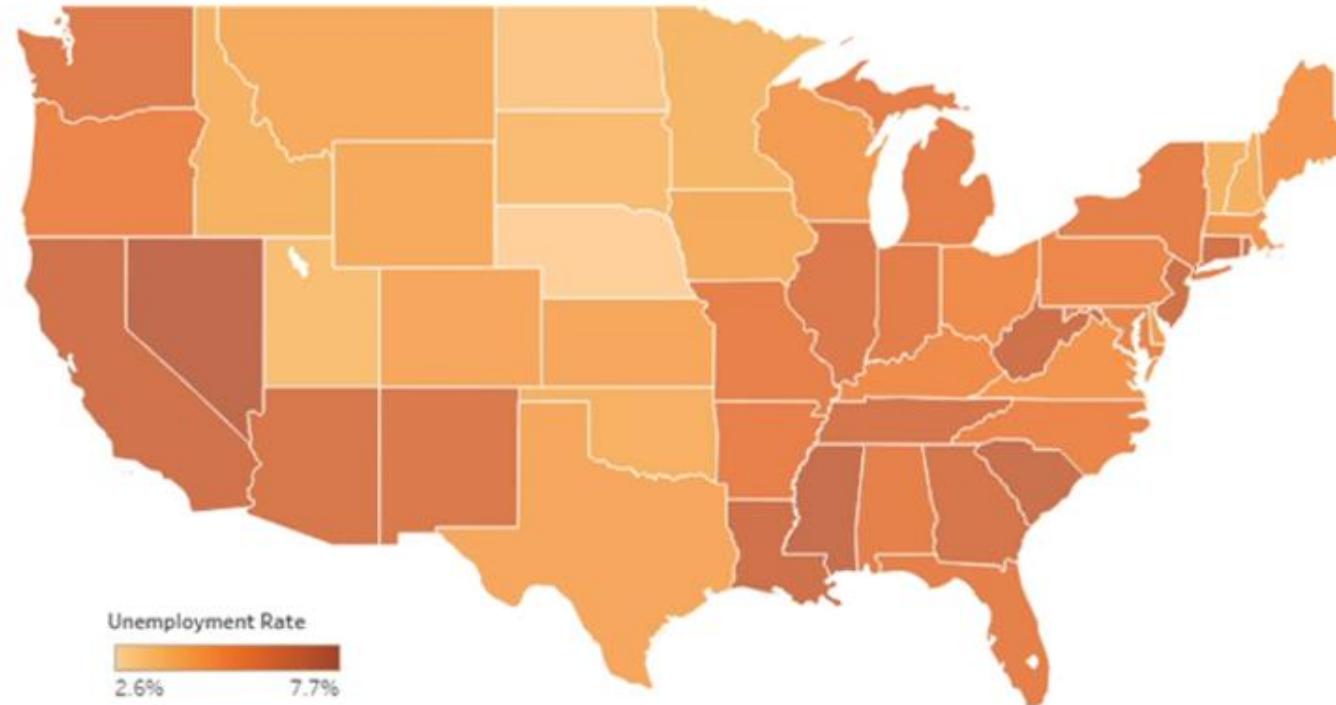


SEQUENTIAL

color is ordered from low to high



Unemployment Rate by State



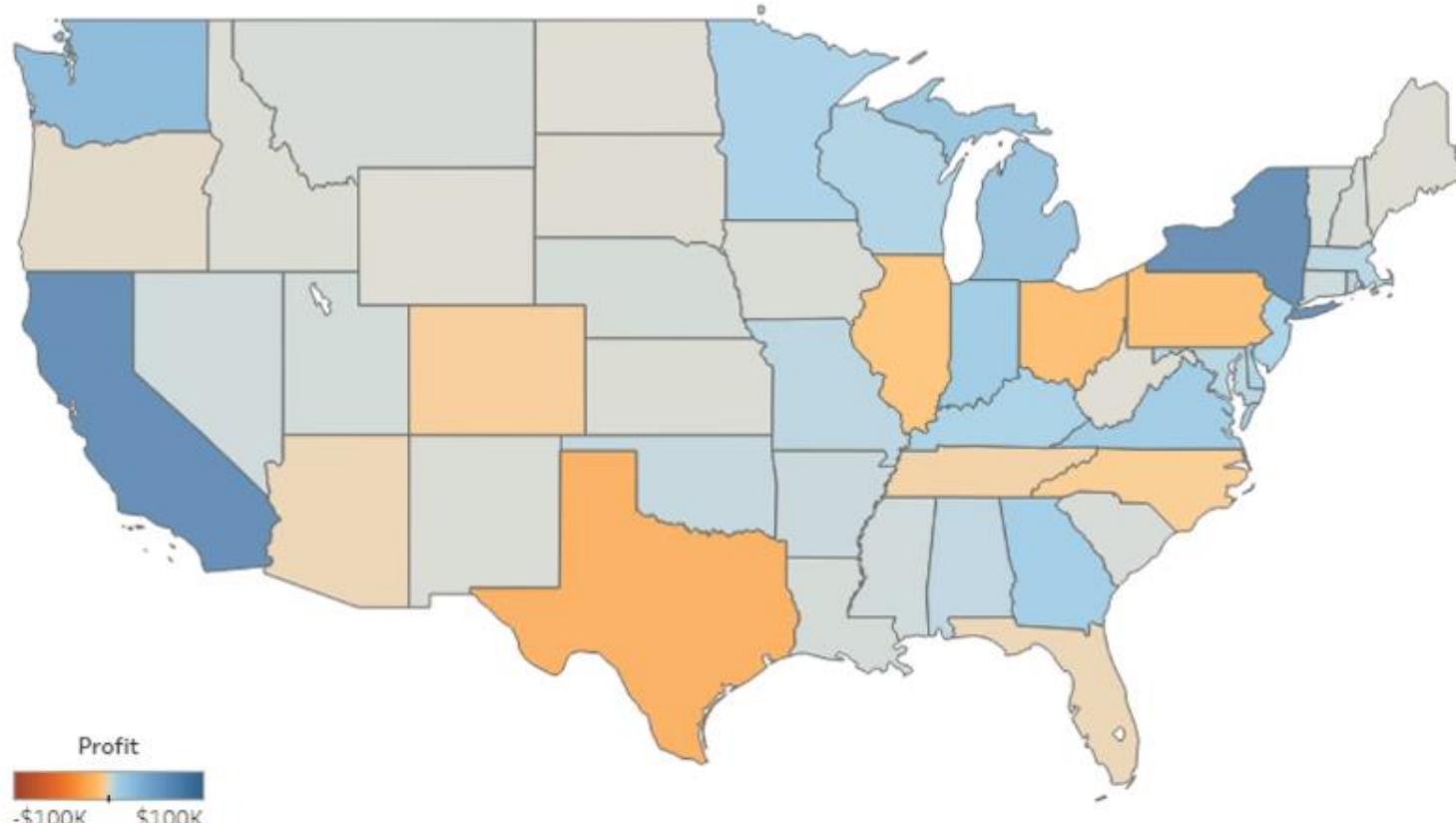
Source: *The Big Book of Dashboards* (Figure 1.17)

DIVERGING

two sequential colors with a neutral midpoint



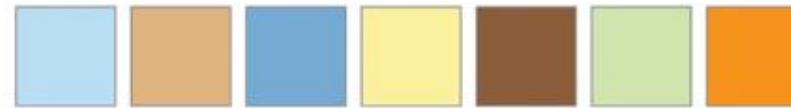
Profit by State



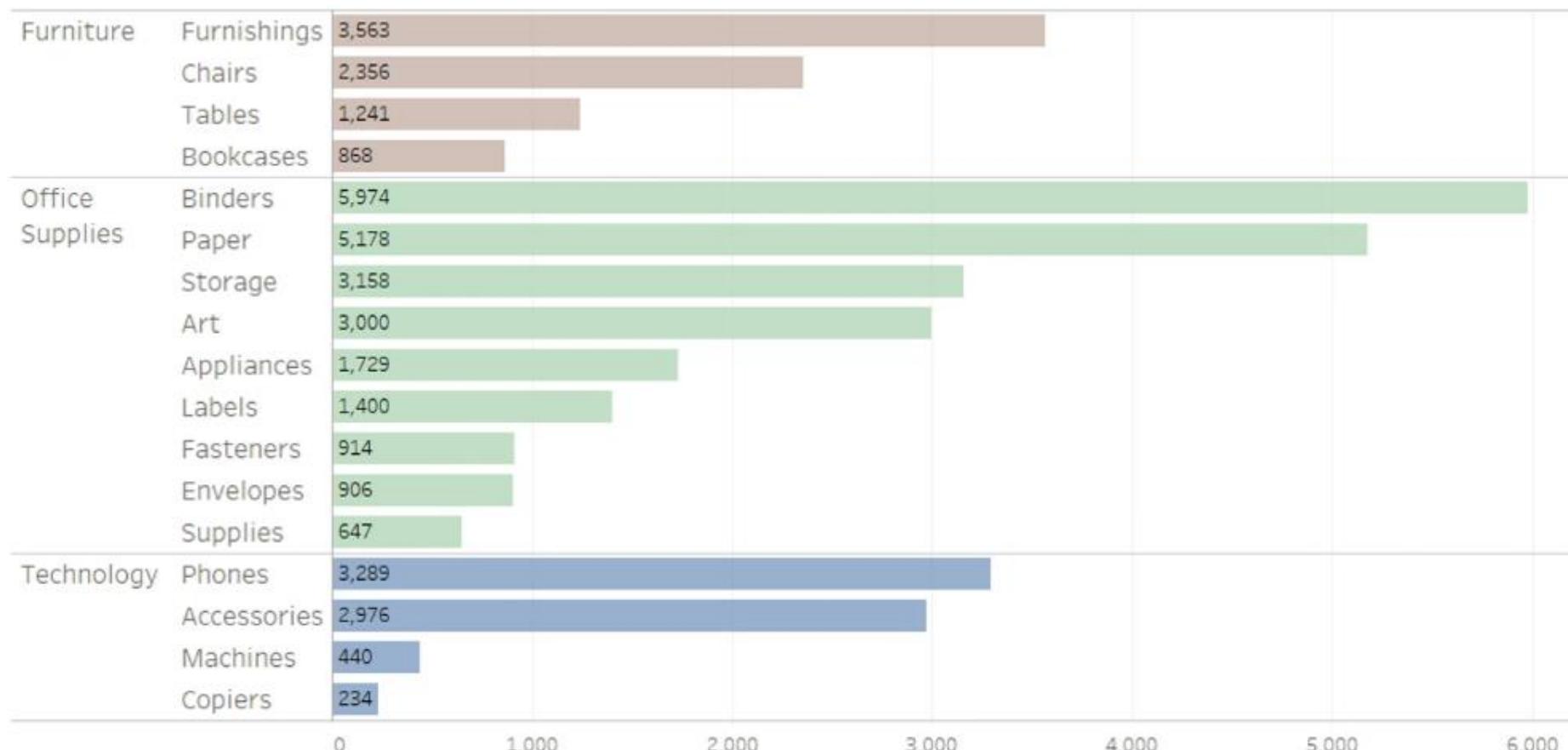
Source: *The Big Book of Dashboards* (Figure 1.19)

CATEGORICAL

contrasting colors for individual comparison

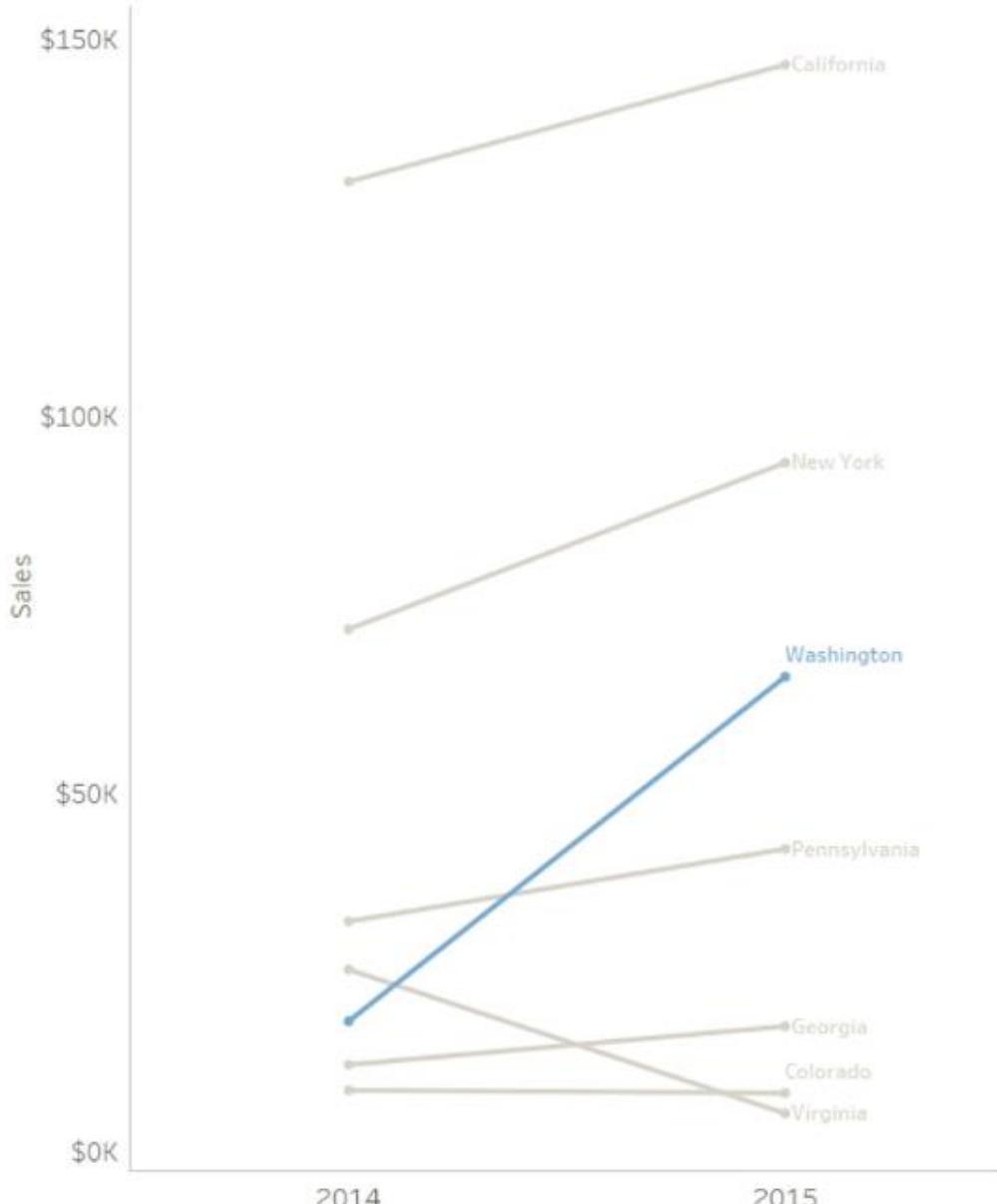


Quantity by Category and Subcategory



Source: *The Big Book of Dashboards* (Figure 1.20)

Sales by State, 2014-2015



HIGHLIGHT

color used to highlight something



Source: *The Big Book of Dashboards (Figure 1.21)*

ALERT

color used to get reader's attention



ACTUAL VS. EXPECTED



83 %



86 %



90 %



92 %



97 %

14-DAY TREND



LAST INSPECTION

3 days



19 days

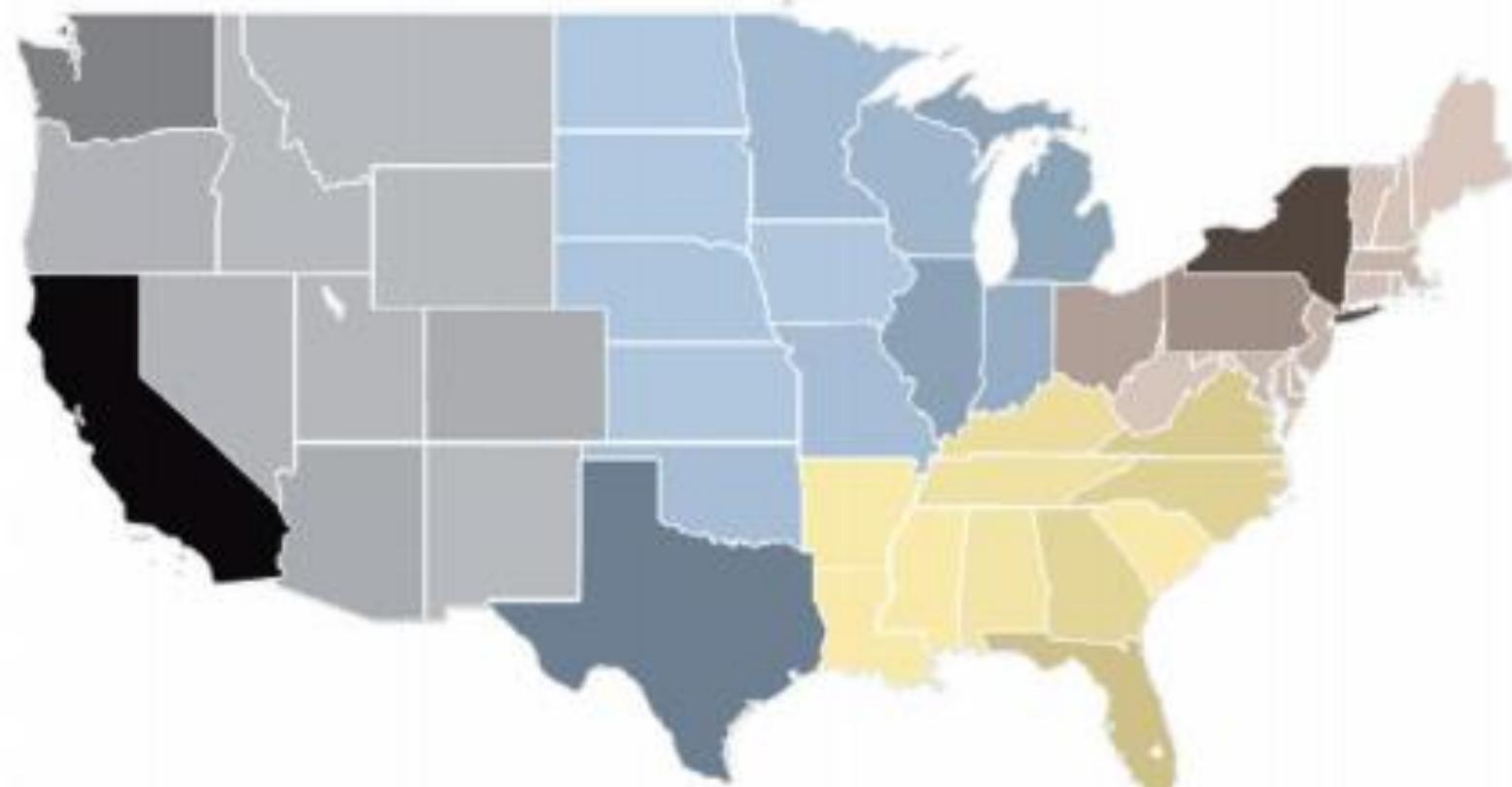
2 days

7 days

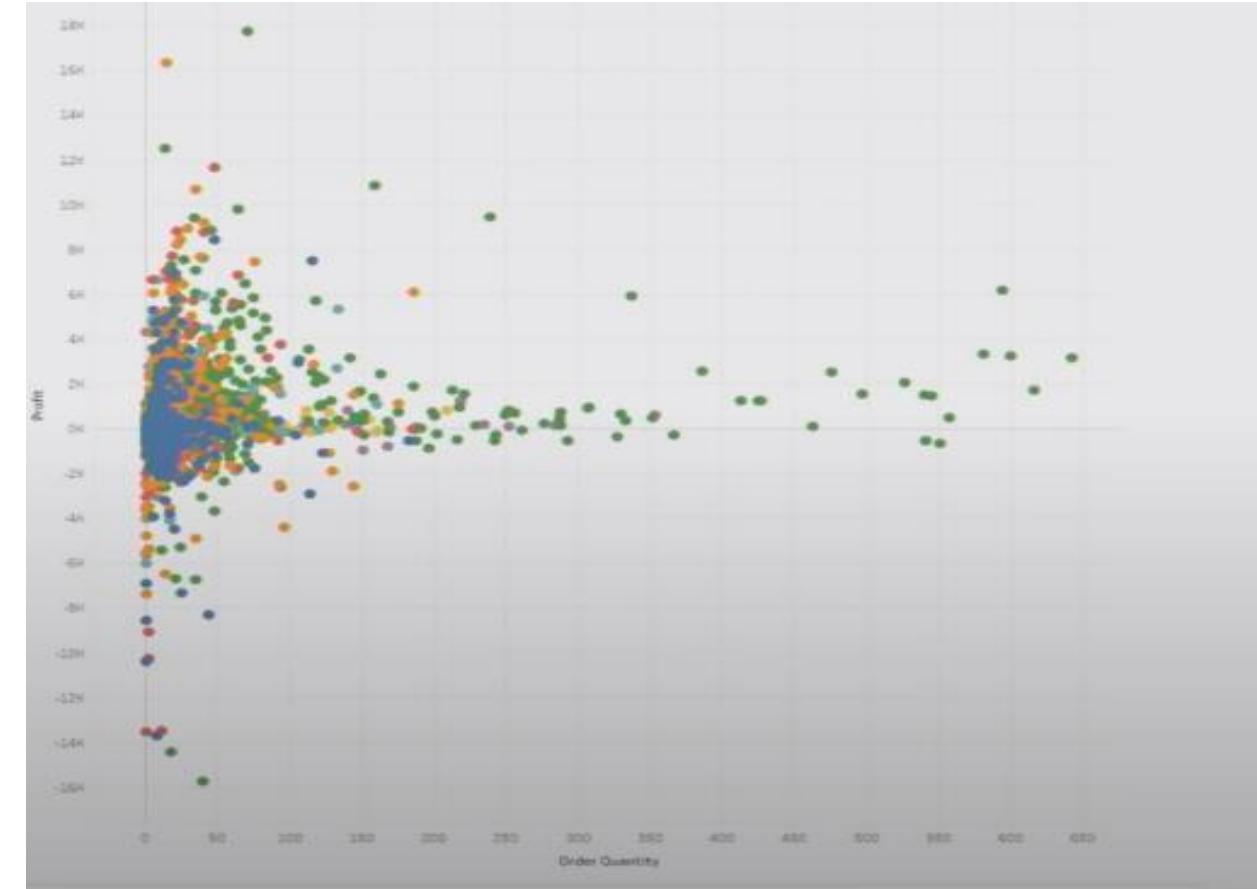


Categorical-Sequential Color Scheme

Sales by Region

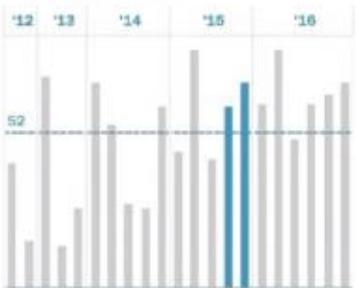


Too Many Colors



Course Metrics

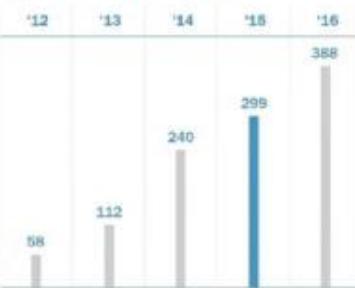
Students



1097

Total students in five years

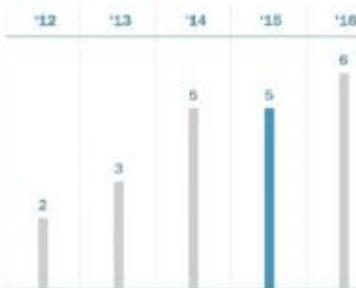
Enrollments



687

Total students in 2015-2016

Classes



21

Total classes in five years

Ratings



7.7 of 8

Most Recent Instructor Rating (out of 8.0)

Semesters

2015 Fall Semester 001

Questions

I developed specific skills and competencies

Overall, this was an excellent course

The instructor communicated clearly

The instructor graded fairly

The instructor was well organized

The instructor interacted well with students

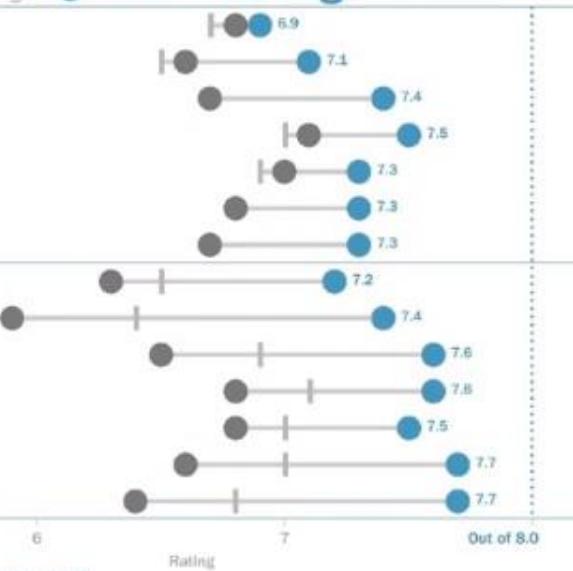
Overall, this instructor was excellent

BANA

|College

Shaffer

Ratings



2015 Fall Semester 002

Course Metrics Dashboard created by Jeffrey A. Shaffer. Data from University of Cincinnati Course Evaluations. Blue indicates the 2 most recent rating periods.

Source: *The Big Book of Dashboards* (BigBookofDashboards.com)

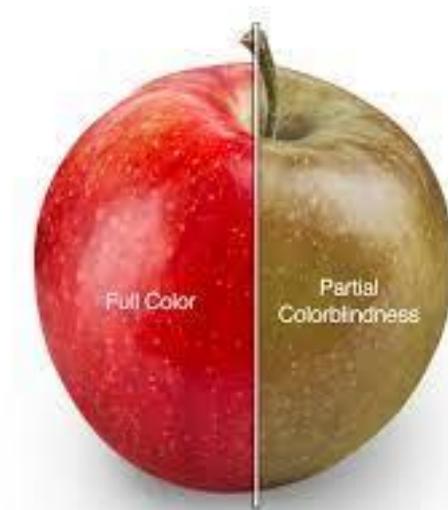


Color
Vision
Deficiency,
aka Color Blind

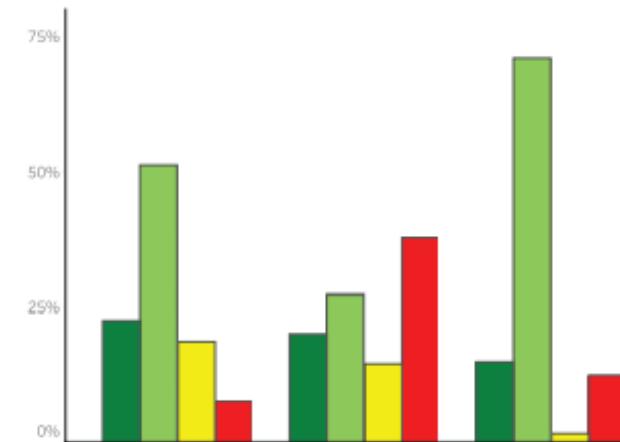
Color Blindness



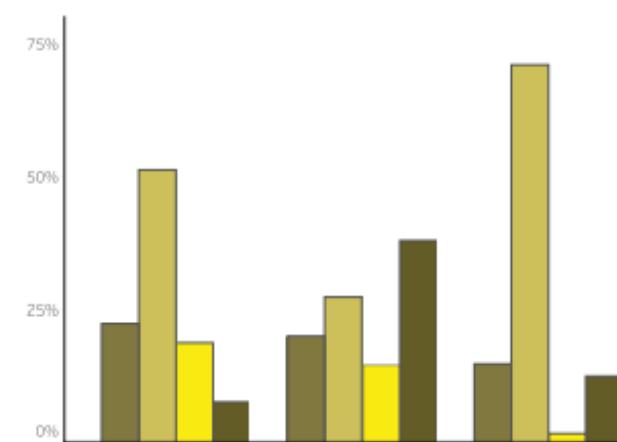
- The primary problem among people with CVD is with the colors **red** and **green**.
- This is why it is best to avoid using red and green together and, in general, to avoid the commonly used **traffic light colors**.



Traffic Light Colors

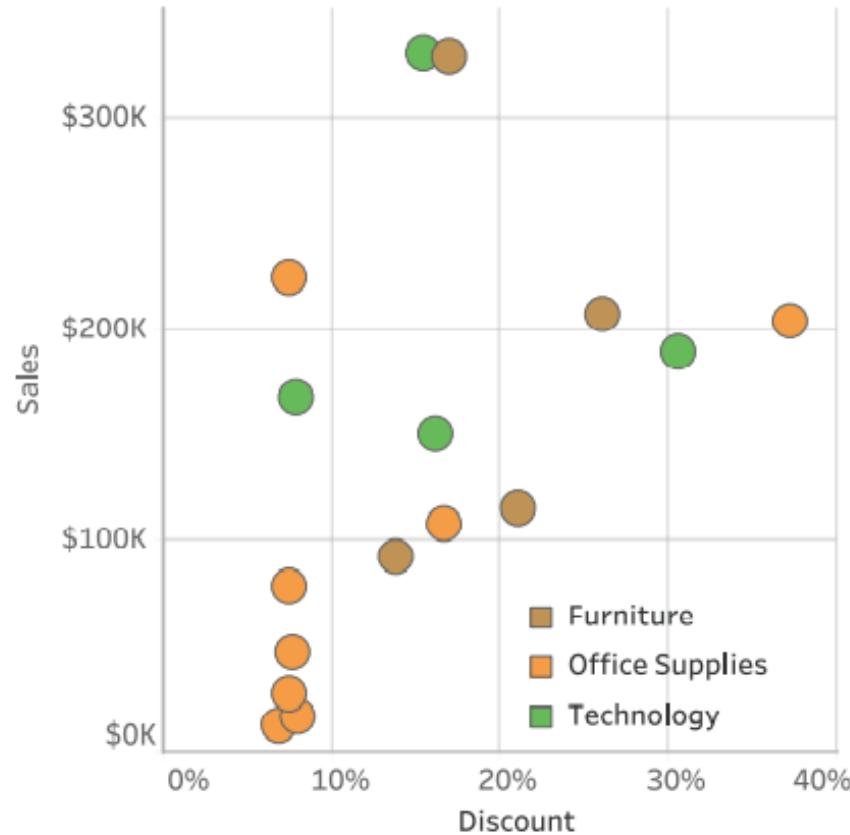


Protanopia Simulation

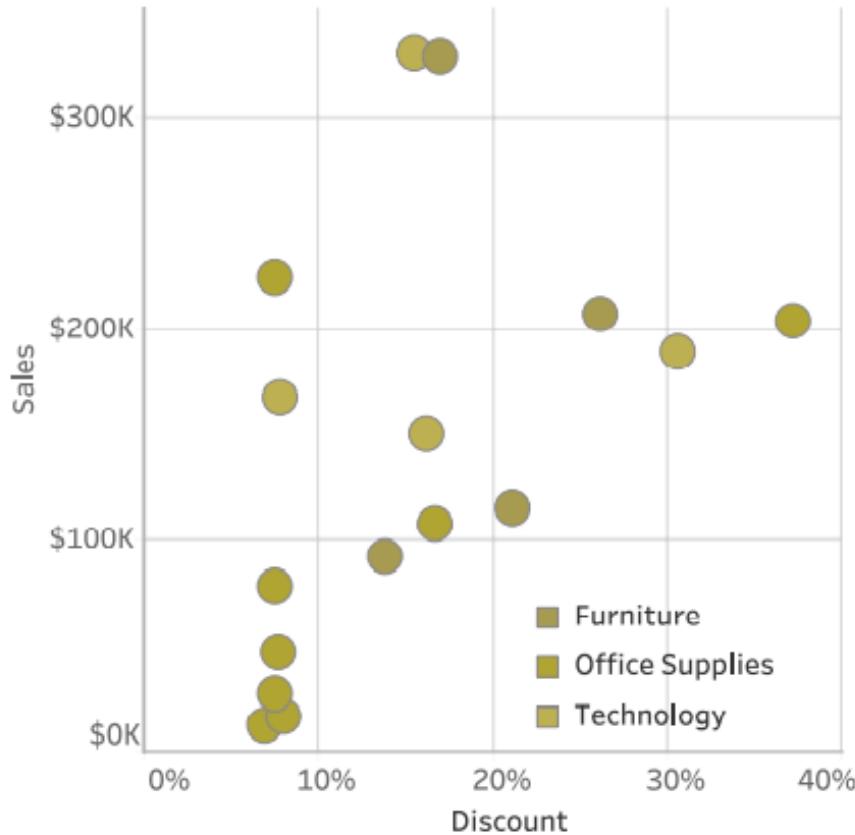


Color Blindness

Normal Color

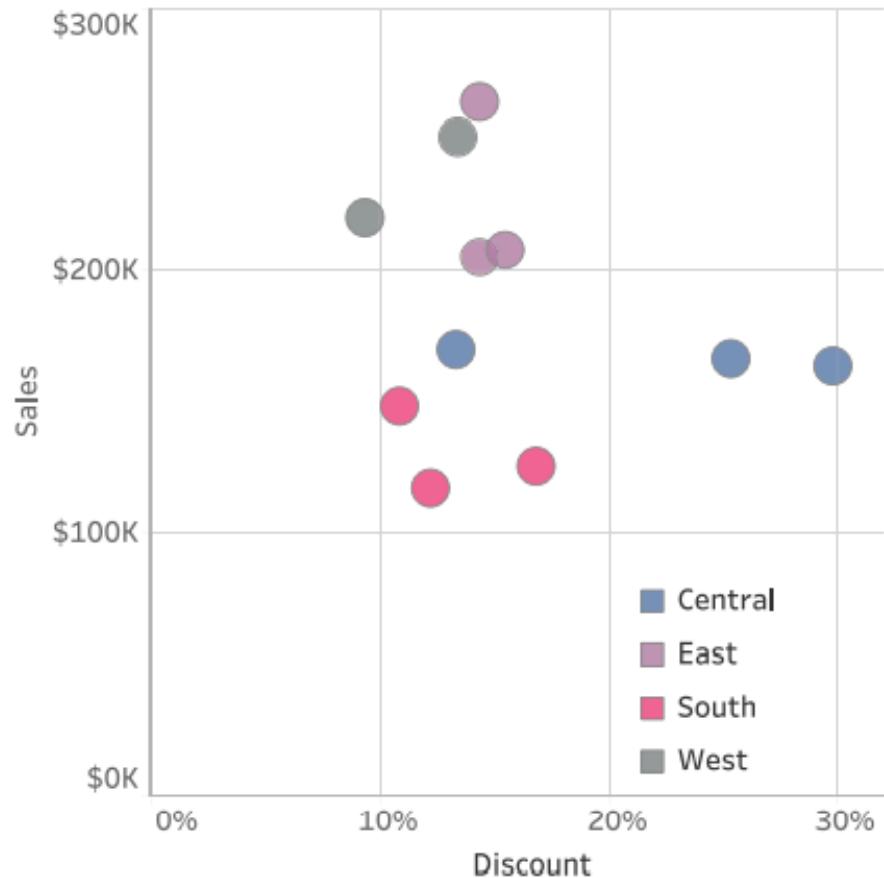


Protanopia CVD Simulation

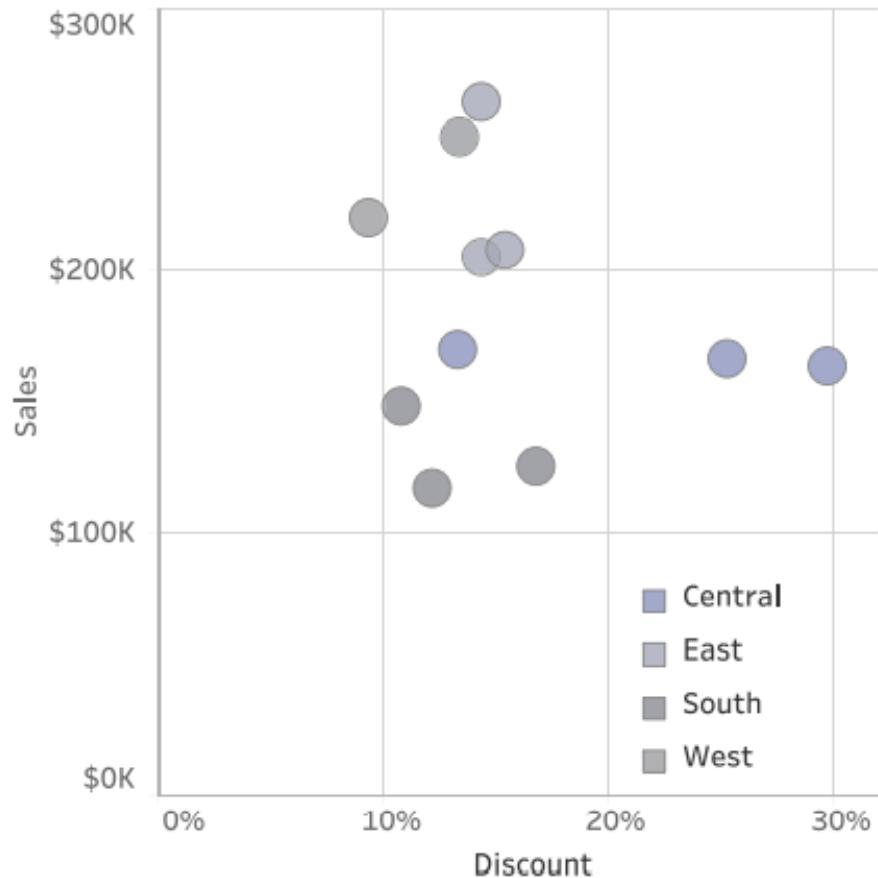


Color Blindness

Normal Color

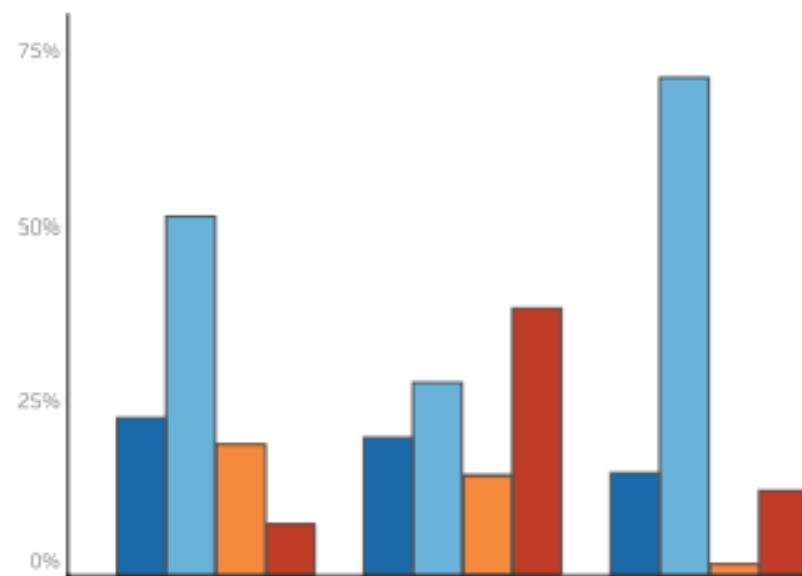


Deuteranopia CVD Simulation

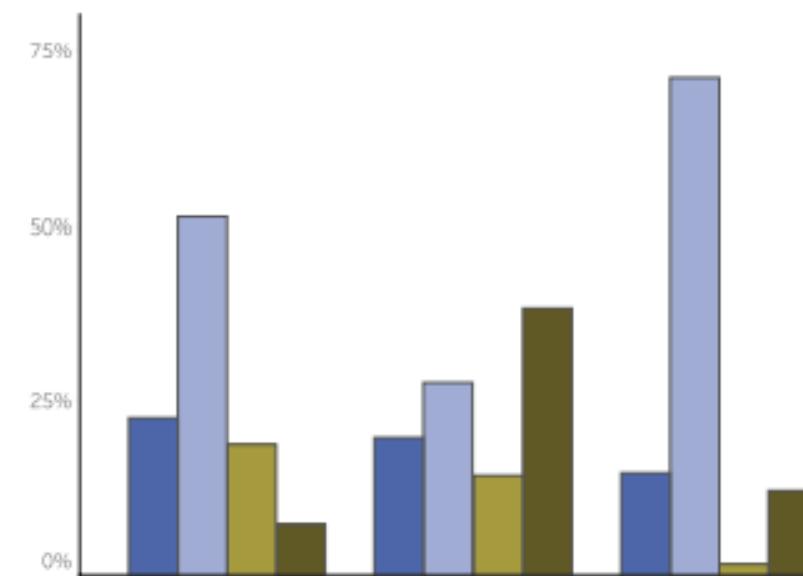


Color Blindness

→ Color-blind-Friendly Blue and Orange



Protanopia Simulation

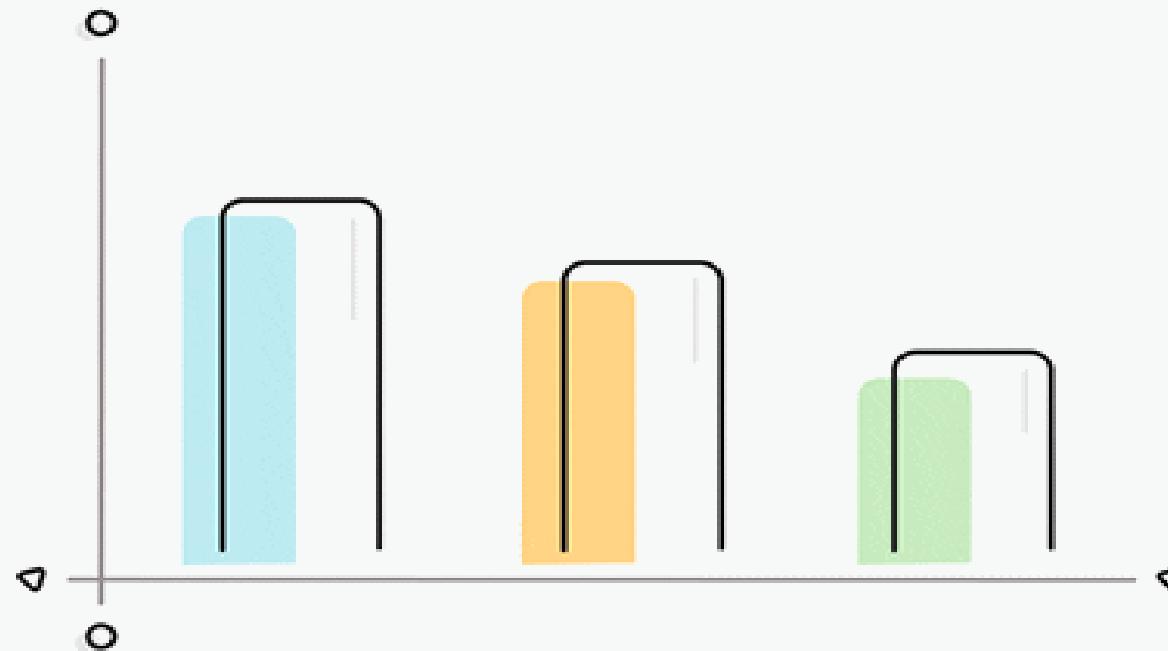




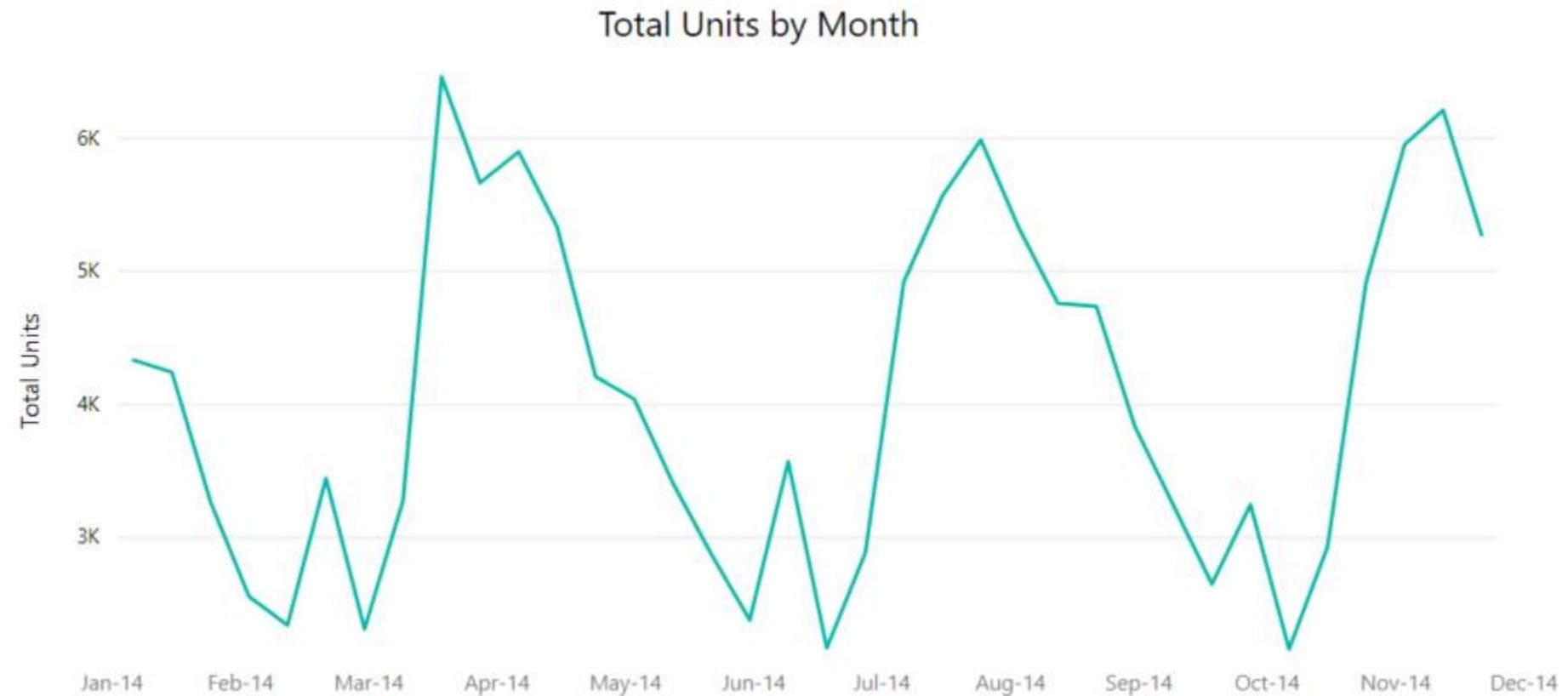
<https://coolors.co/>



How to choose the right **chart**?!



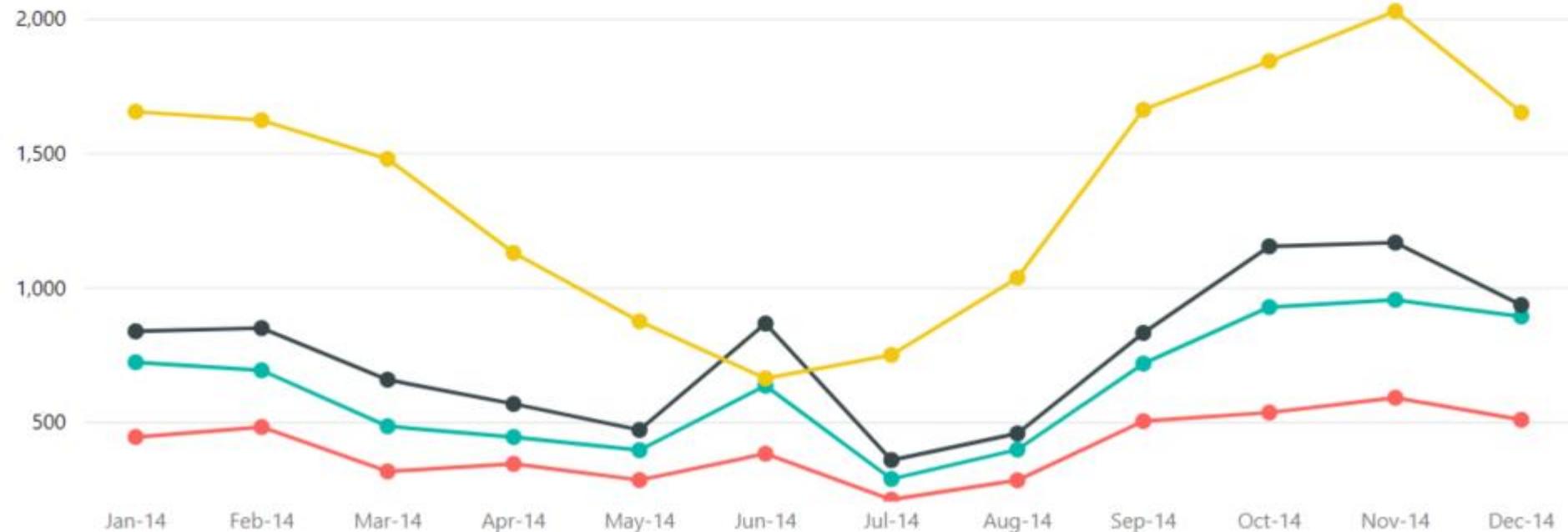
LINE CHART



LINE CHART

Total Units by Month and Manufacturer

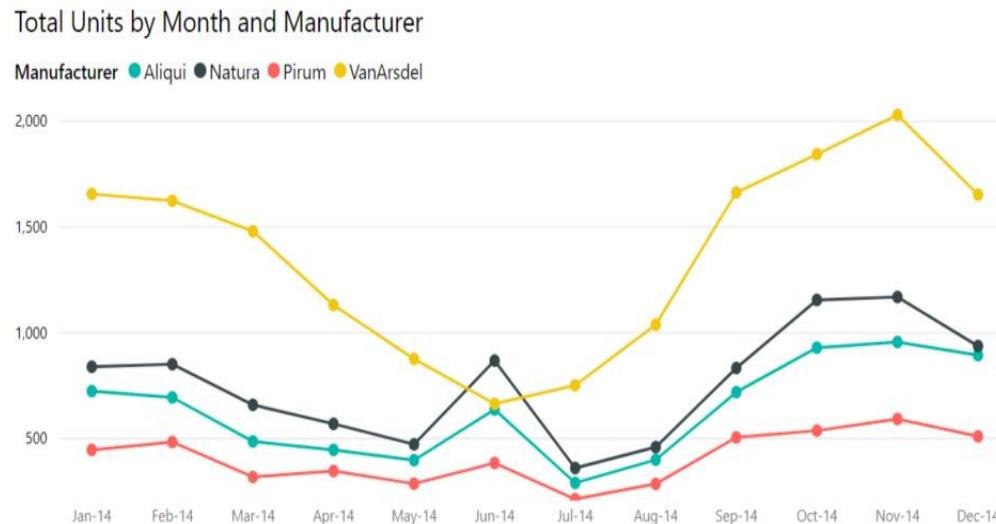
Manufacturer ● Aliqui ● Natura ● Pirum ● VanArsdel



LINE CHART

When to use a LINE chart ?

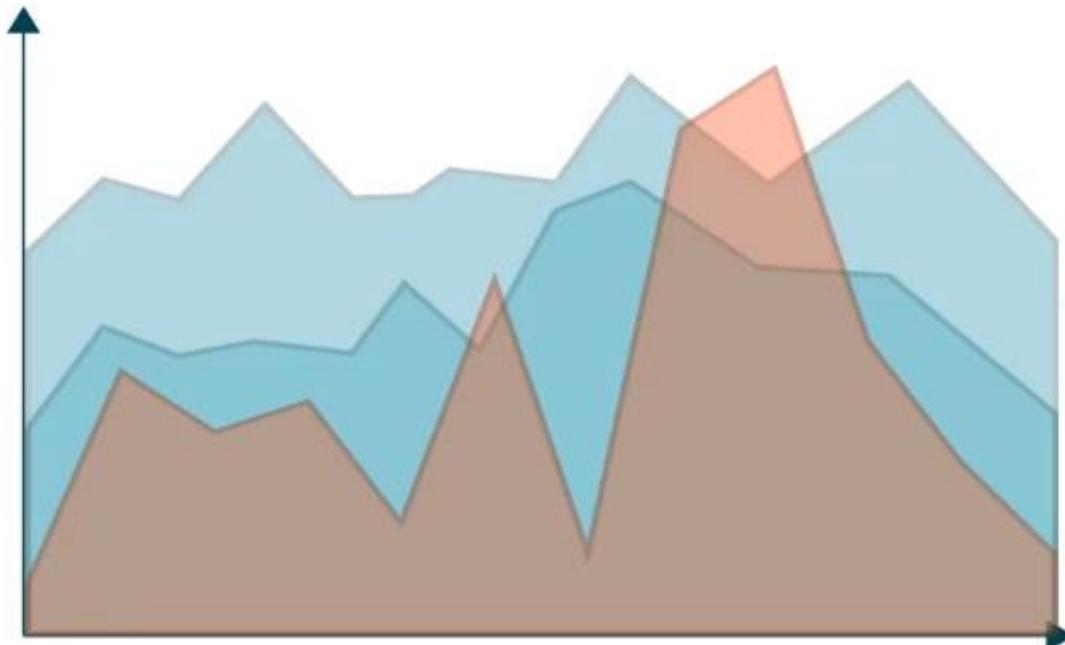
- To see and compare the volume trend across a time series.
- For an individual series representing a physically countable set.



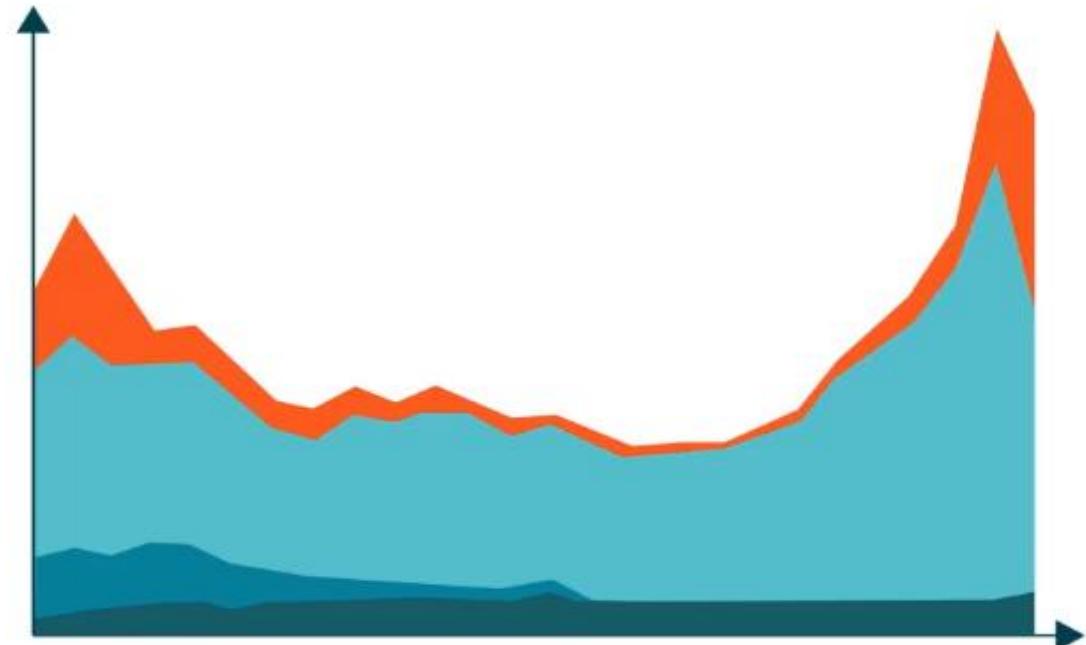
LINE CHART



AREA CHARTS: BASIC (LAYERED) AND STACKED

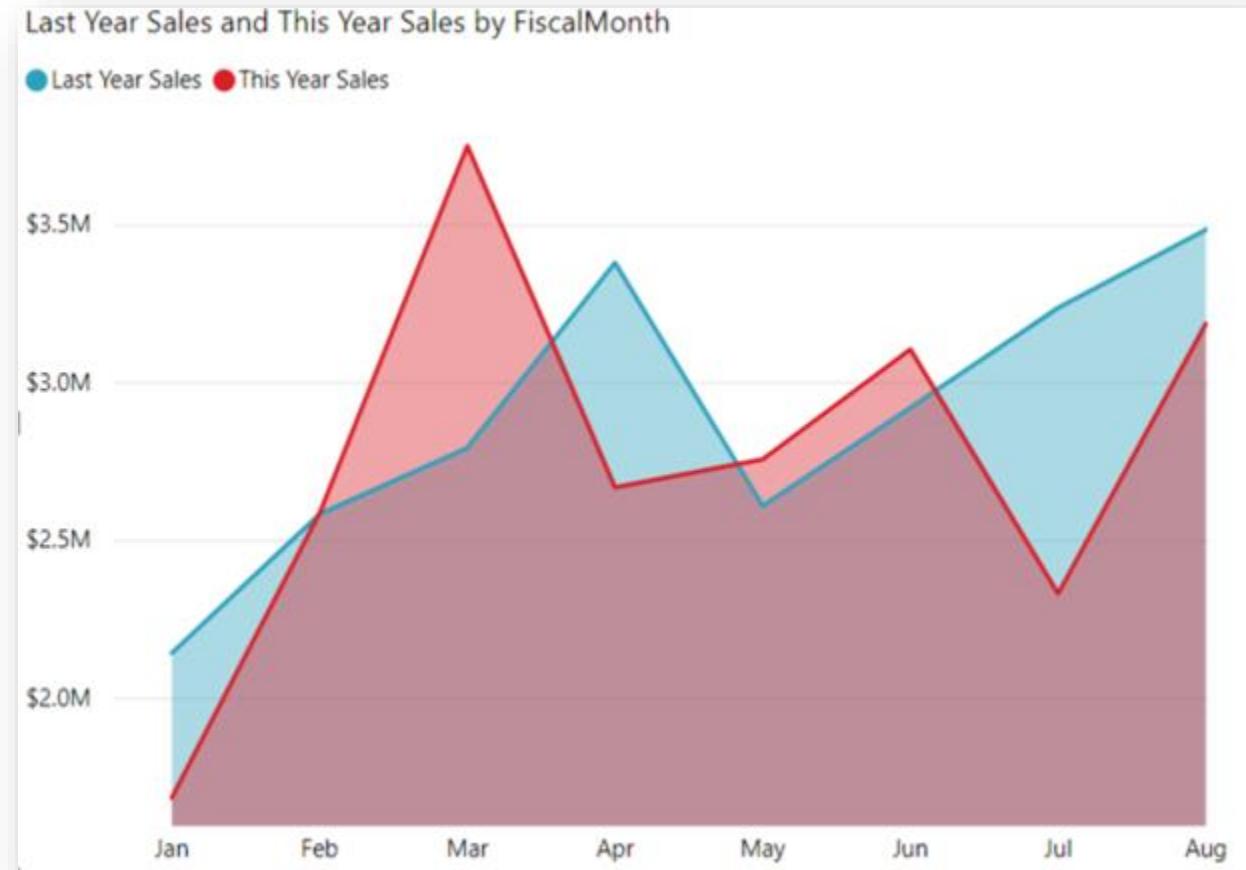


Classic area chart
- Overlap among categories

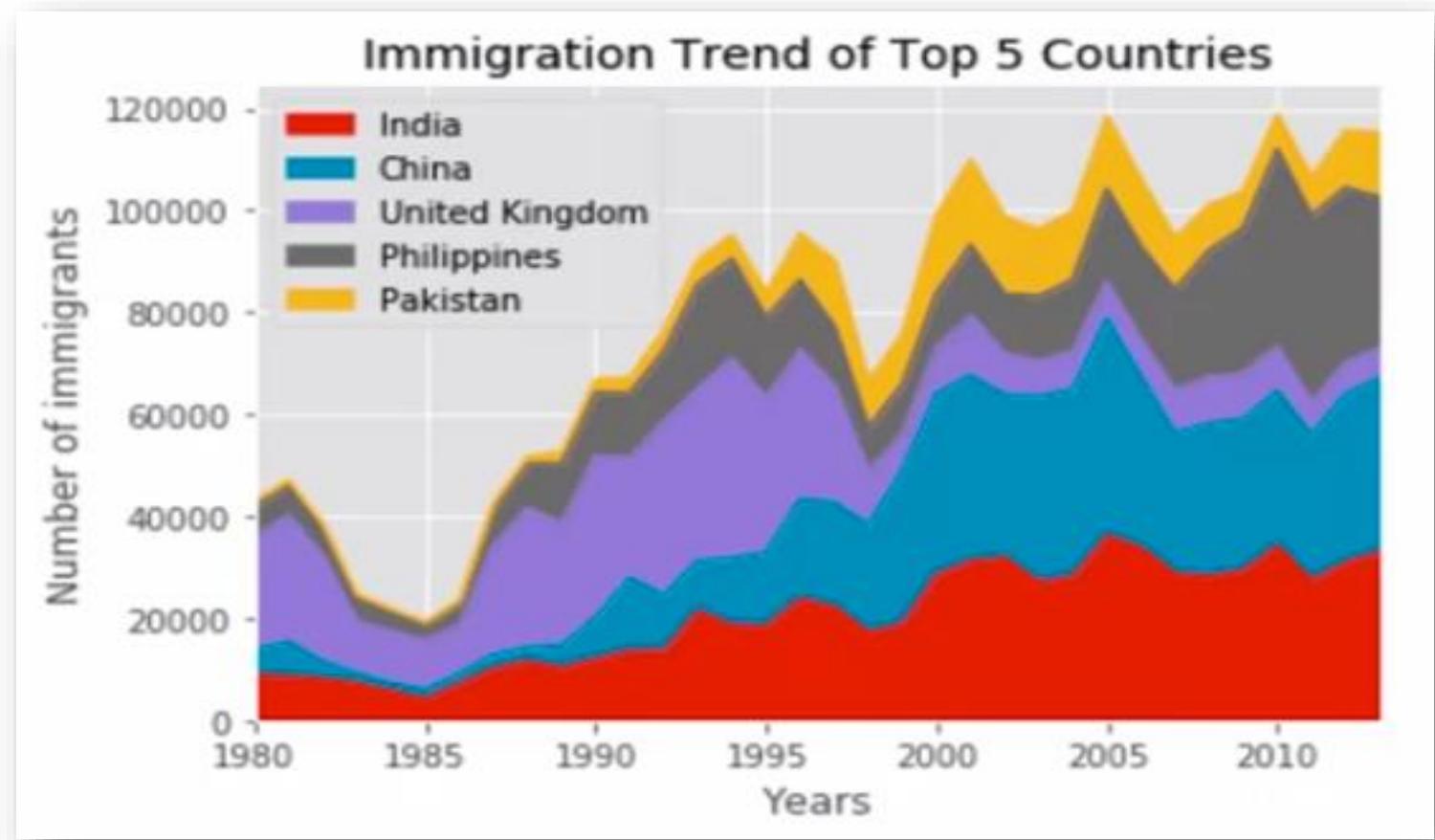


Stacked area chart
- No overlap among categories
- The entire area is the sum of all categories

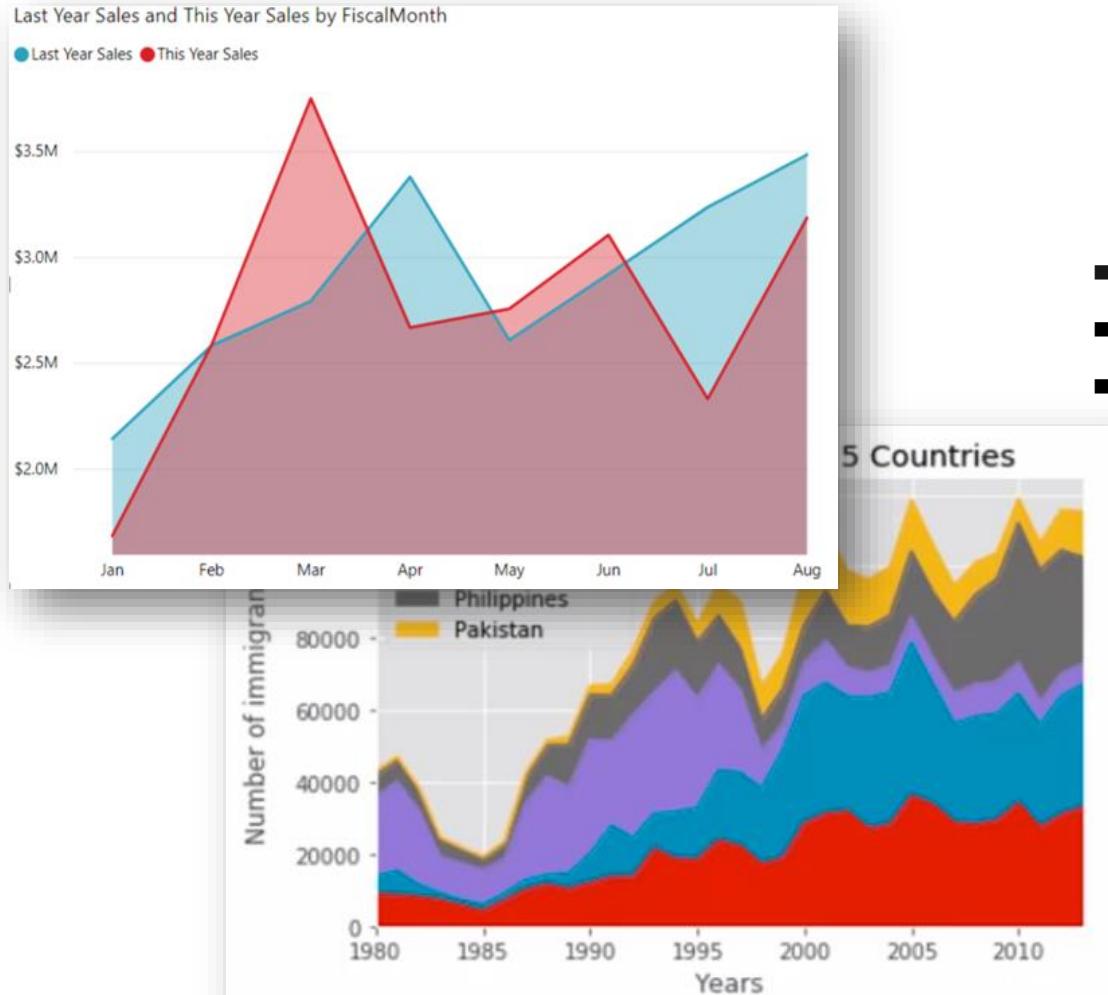
AREA CHARTS: BASIC (LAYERED) AND STACKED



AREA CHARTS: BASIC (LAYERED) AND STACKED



AREA CHARTS: BASIC (LAYERED) AND STACKED



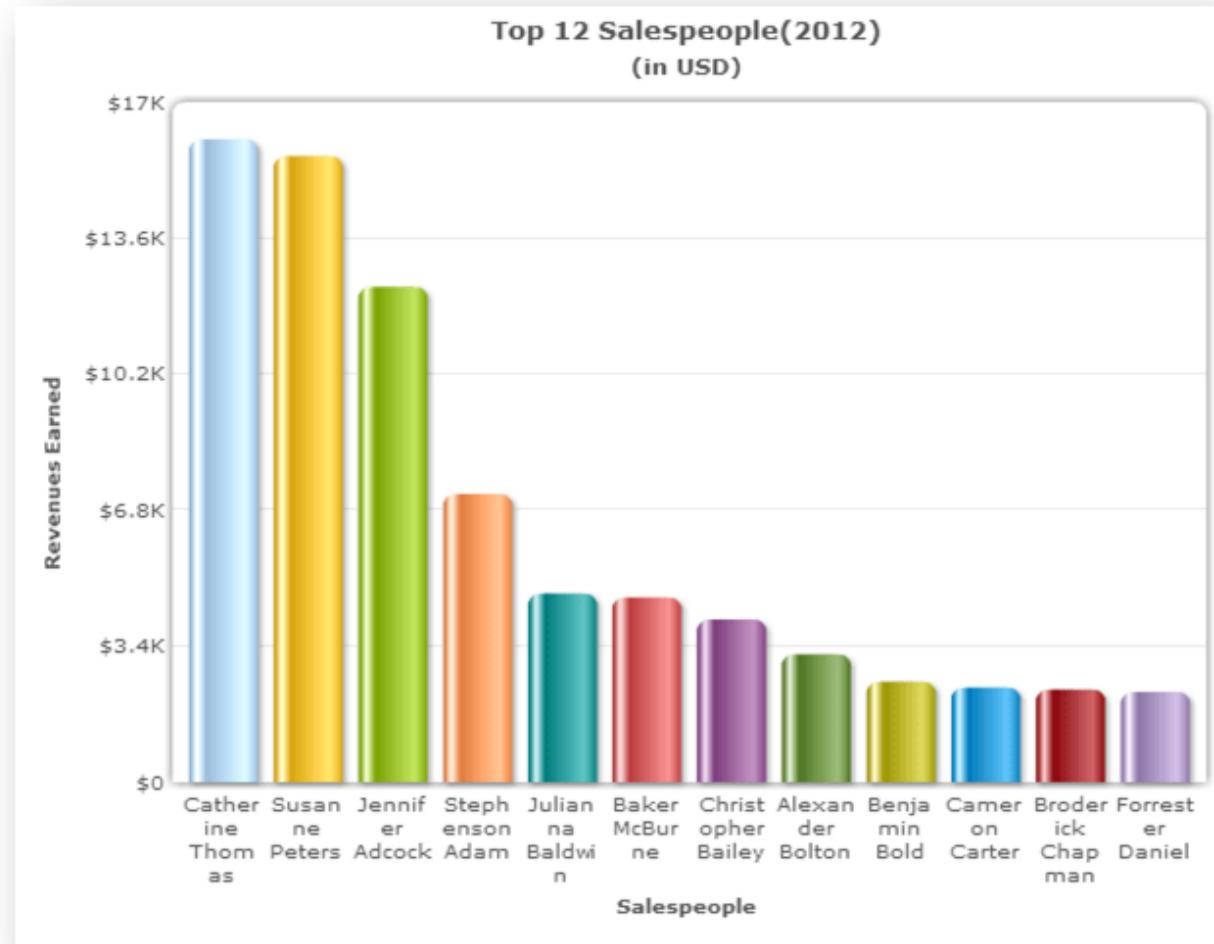
When to use area chart ?

- To see and compare the volume trend across a time series.
- There is data expressed as a total
- There are multiple data series with part-to-whole relationships, or a cumulative series of values.

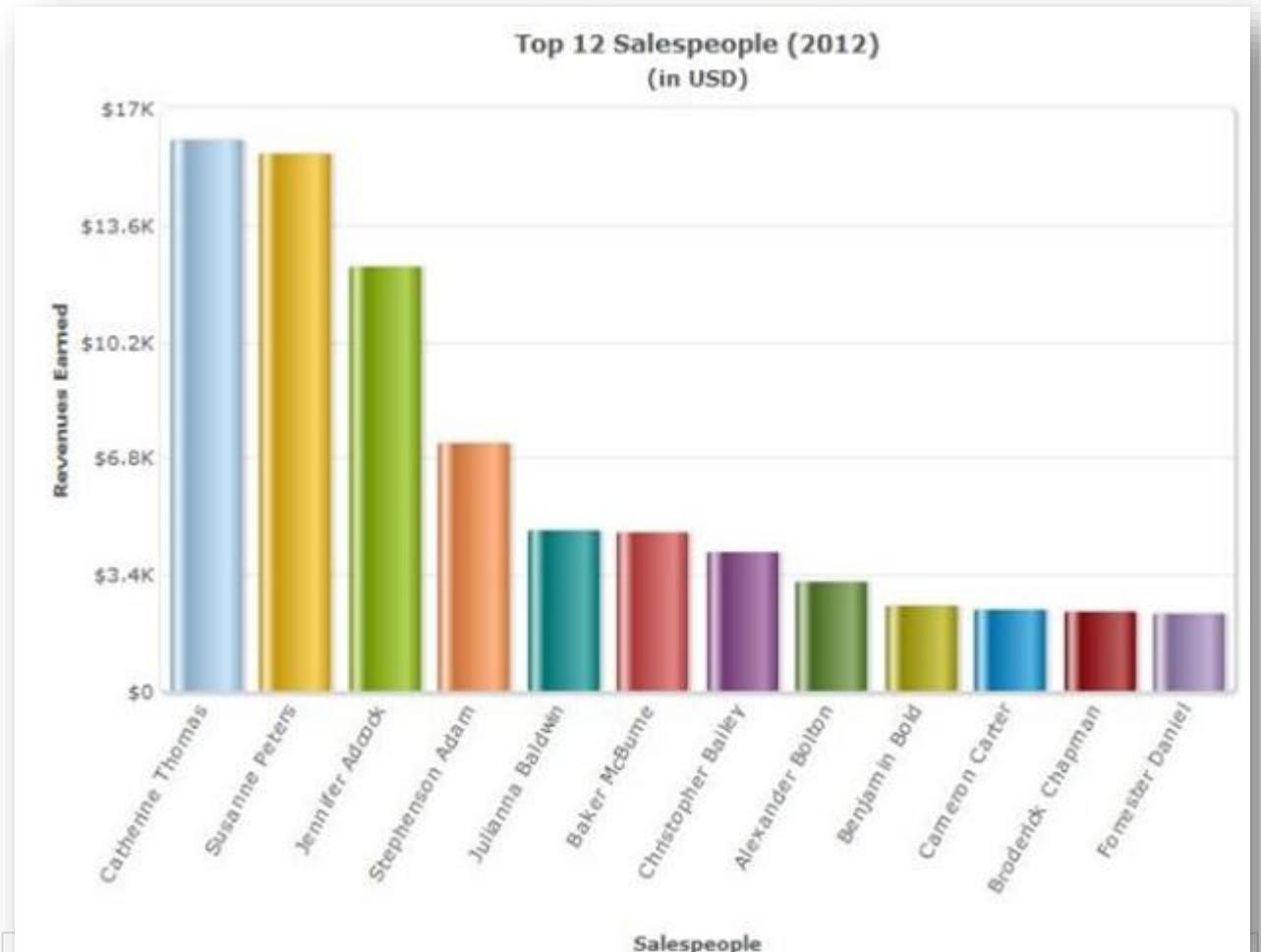
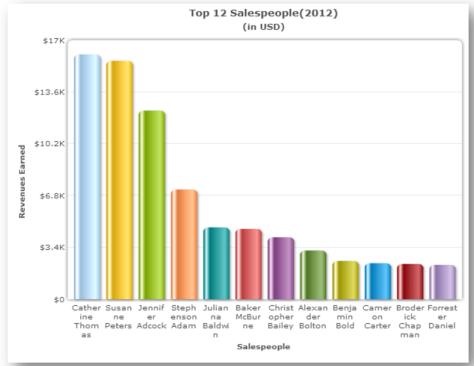
BAR AND COLUMN CHARTS



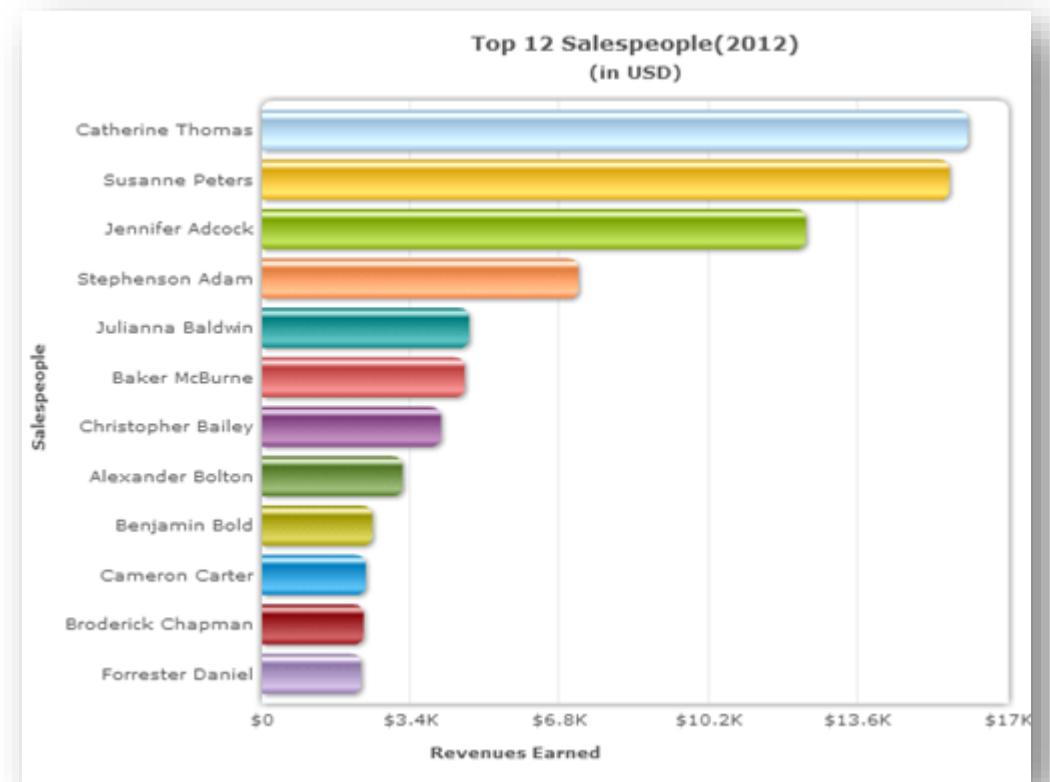
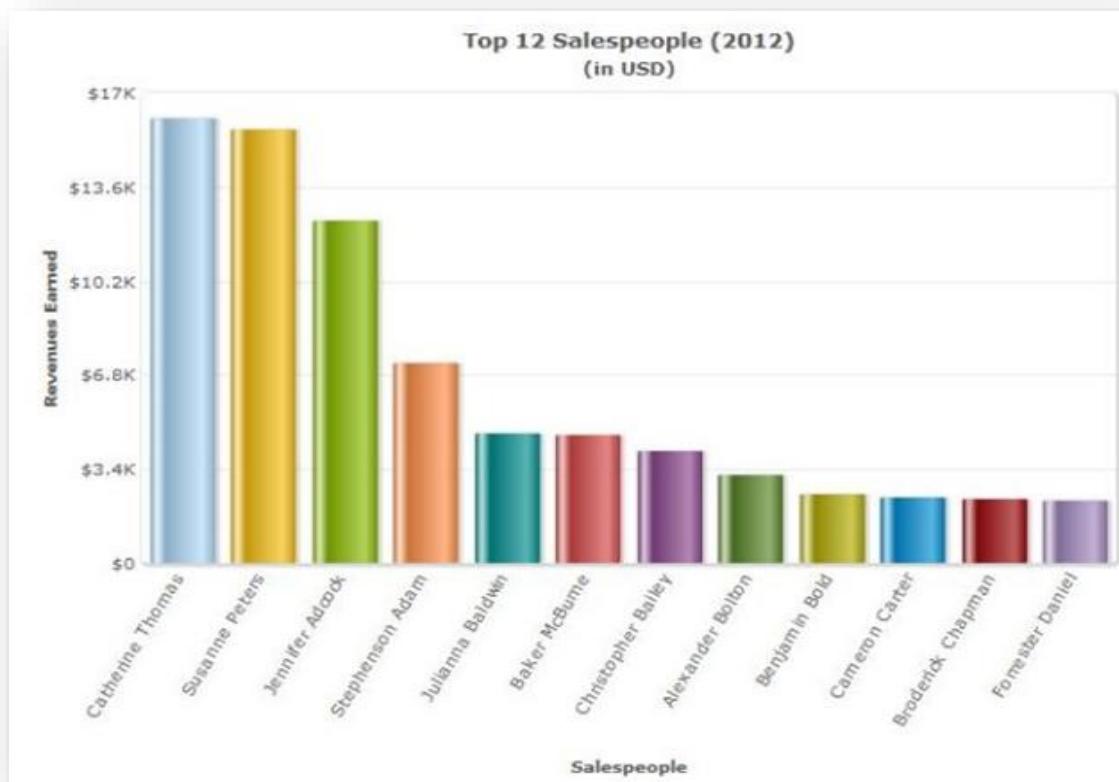
BAR AND COLUMN CHARTS



BAR AND COLUMN CHARTS



BAR AND COLUMN CHARTS

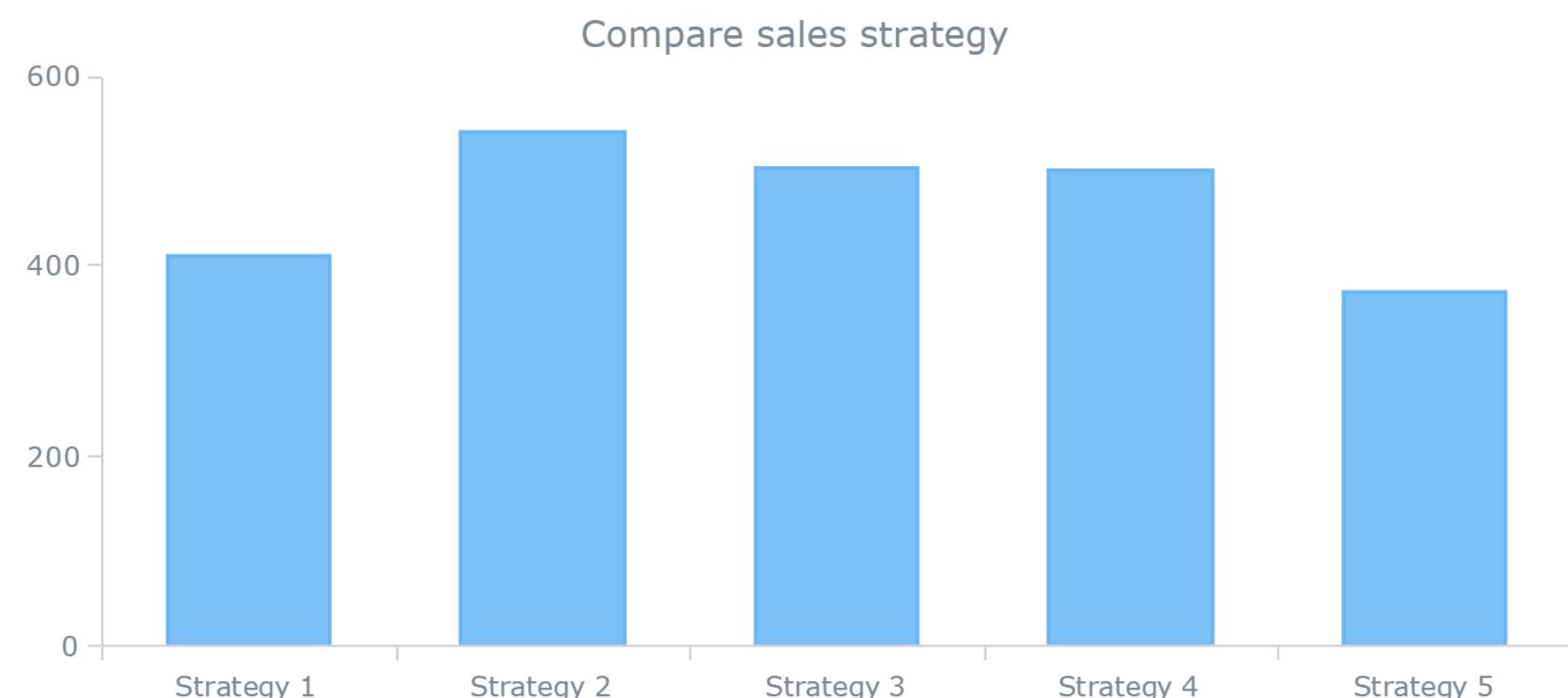


BAR AND COLUMN CHARTS

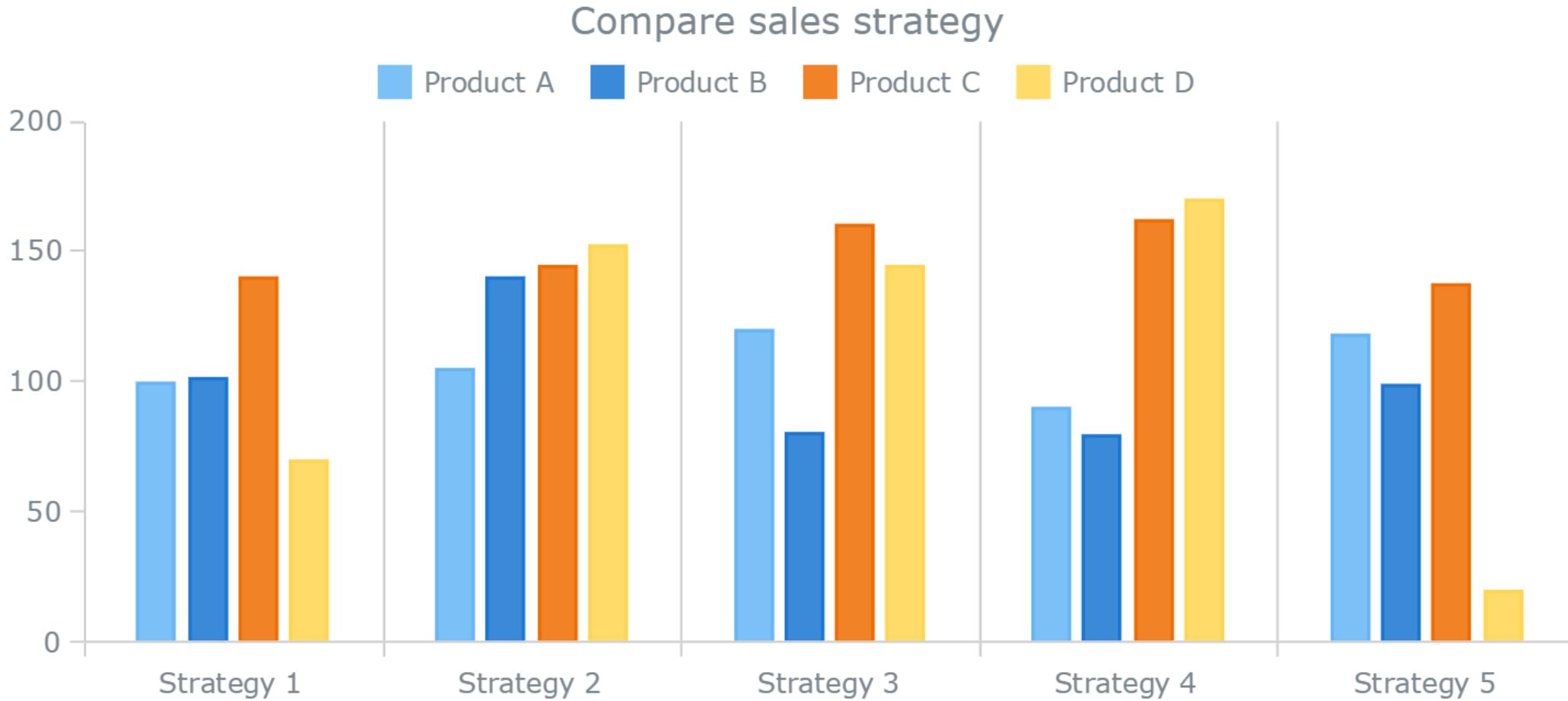
Product ▾	Strategy 1 ▾	Strategy 2 ▾	Strategy 3 ▾	Strategy 4 ▾	Strategy 5 ▾
Product A	100	110	120	80	110
Product B	100	140	70	70	100
Product C	140	145	155	160	140
Product D	60	150	145	180	30

BAR AND COLUMN CHARTS

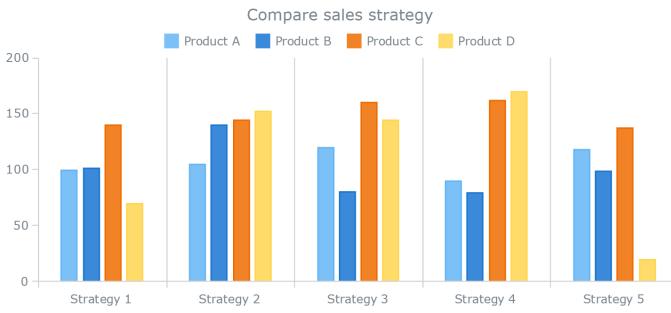
Product	Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5
Product A	100	110	120	80	110
Product B	100	140	70	70	100
Product C	140	145	155	160	140
Product D	60	150	145	180	30



CLUSTERED BAR & COLUMN CHARTS

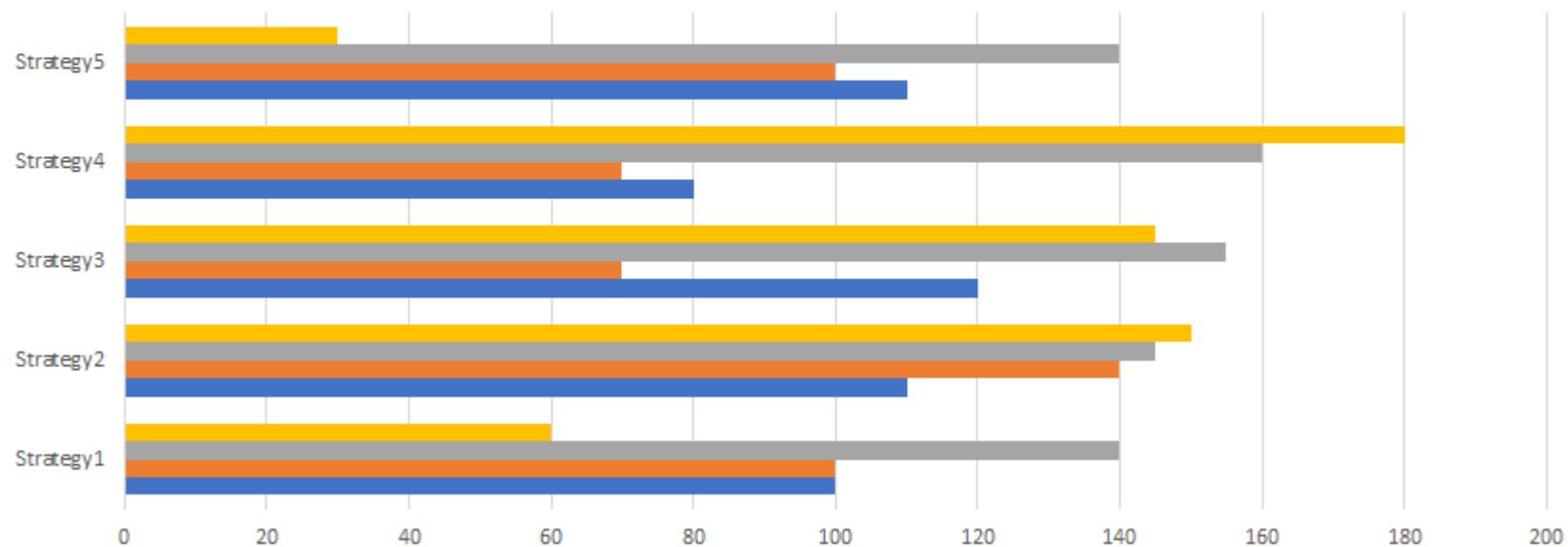


CLUSTERED BAR & COLUMN CHARTS

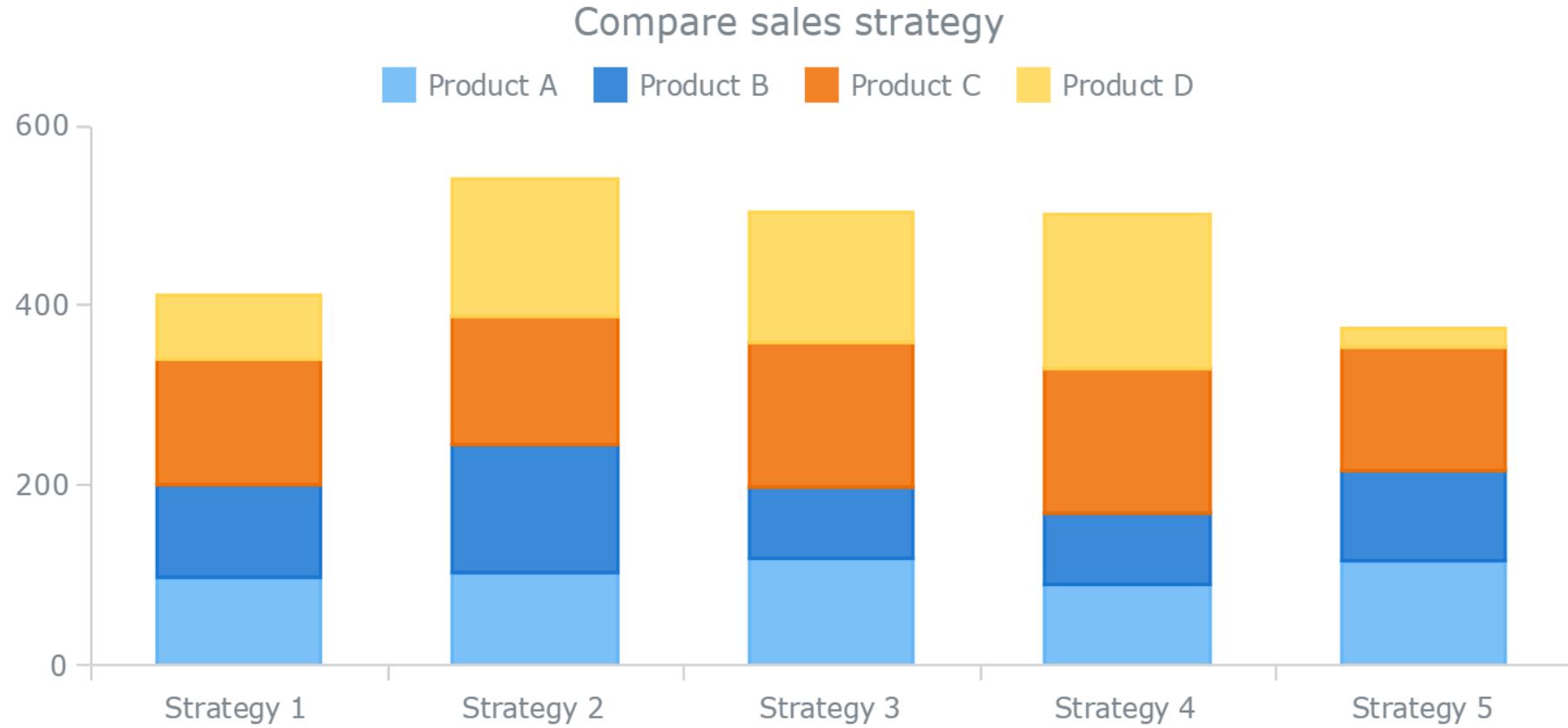


Compare sales strategy

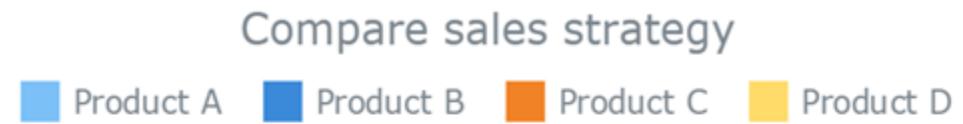
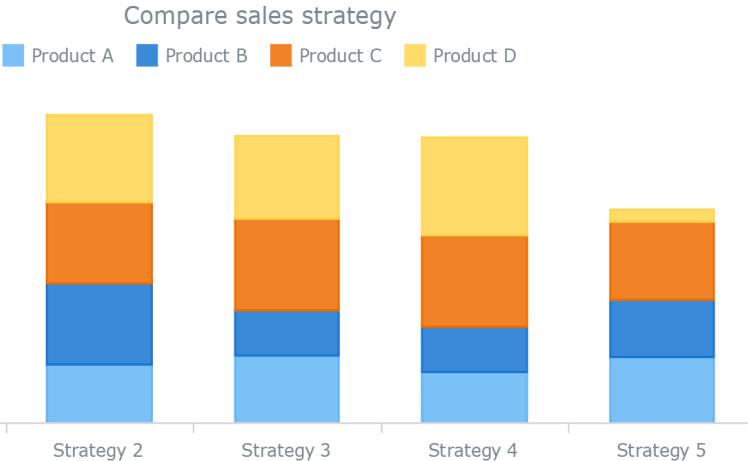
Product A Product B Product C Product D



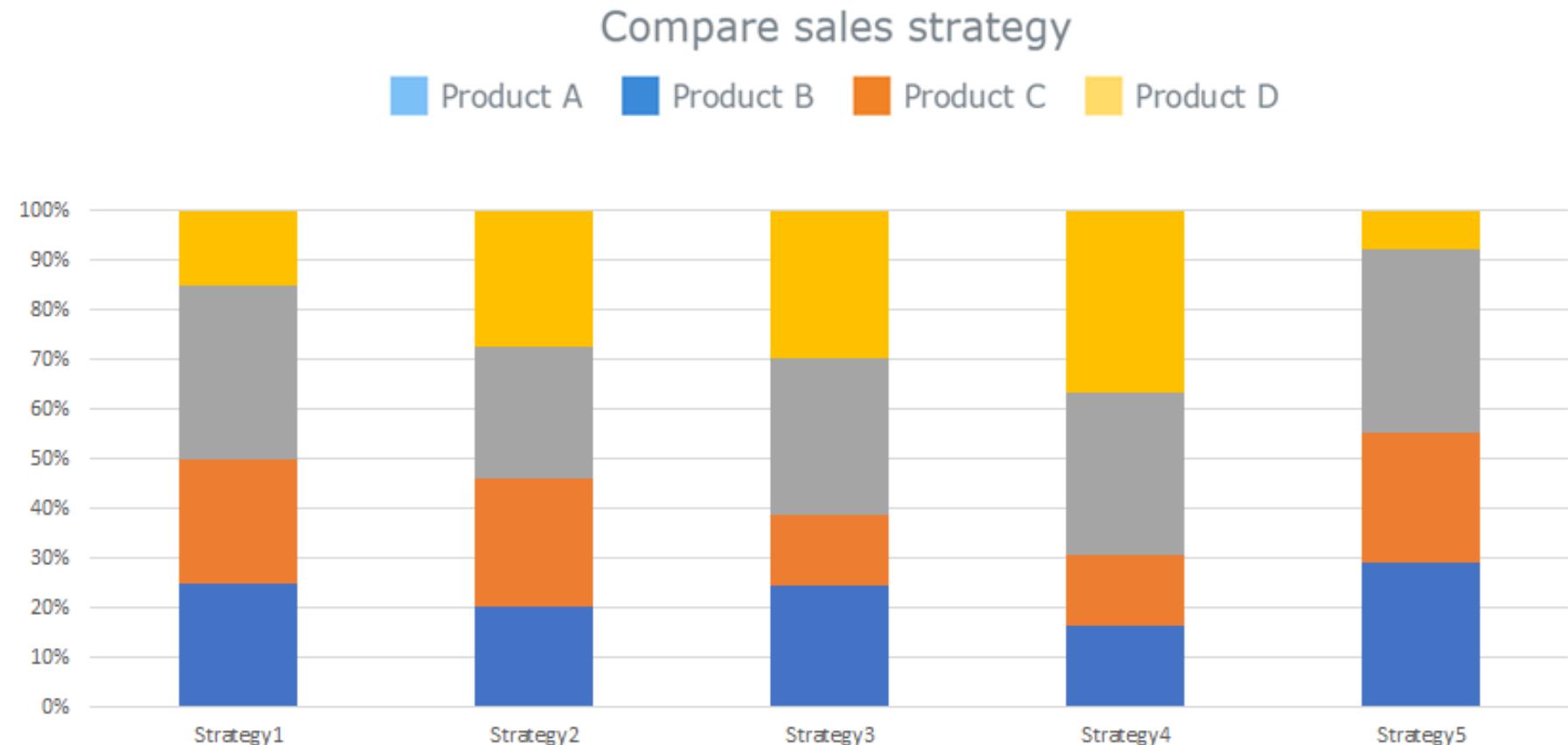
STACKED BAR & COLUMN CHART



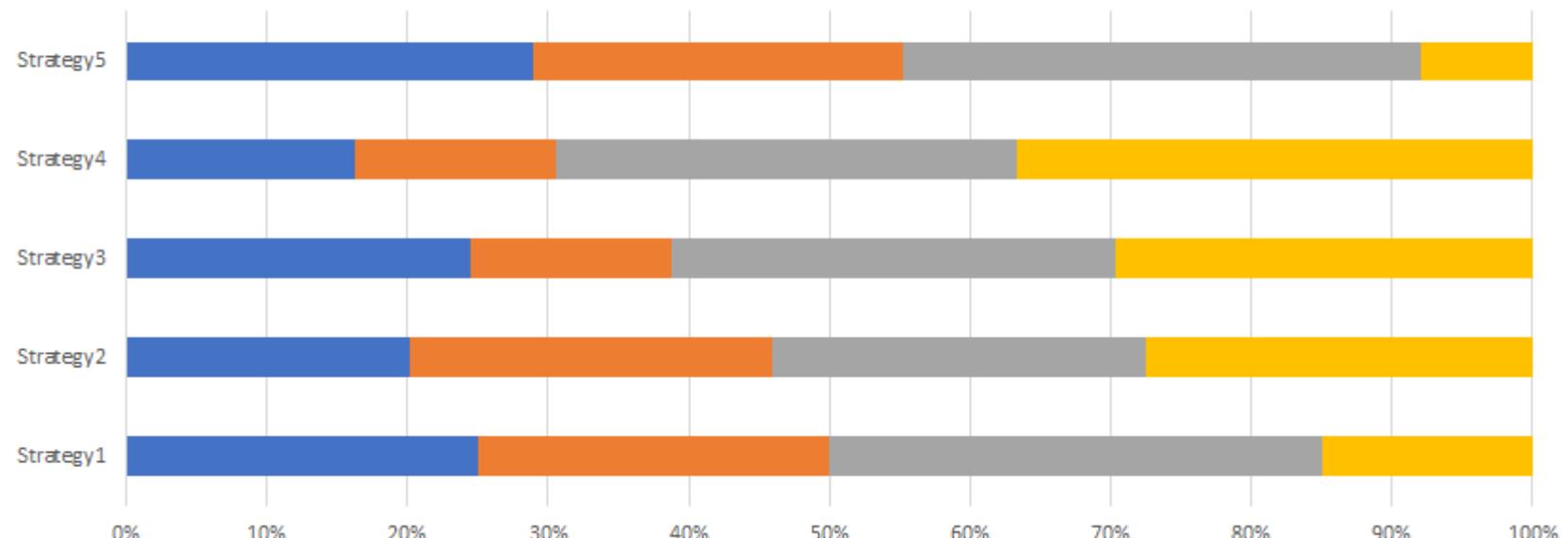
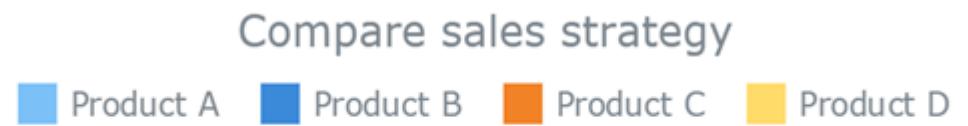
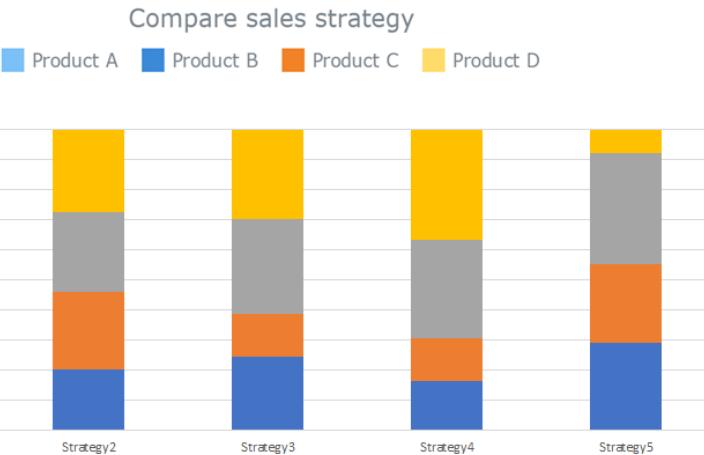
STACKED BAR & COLUMN CHART



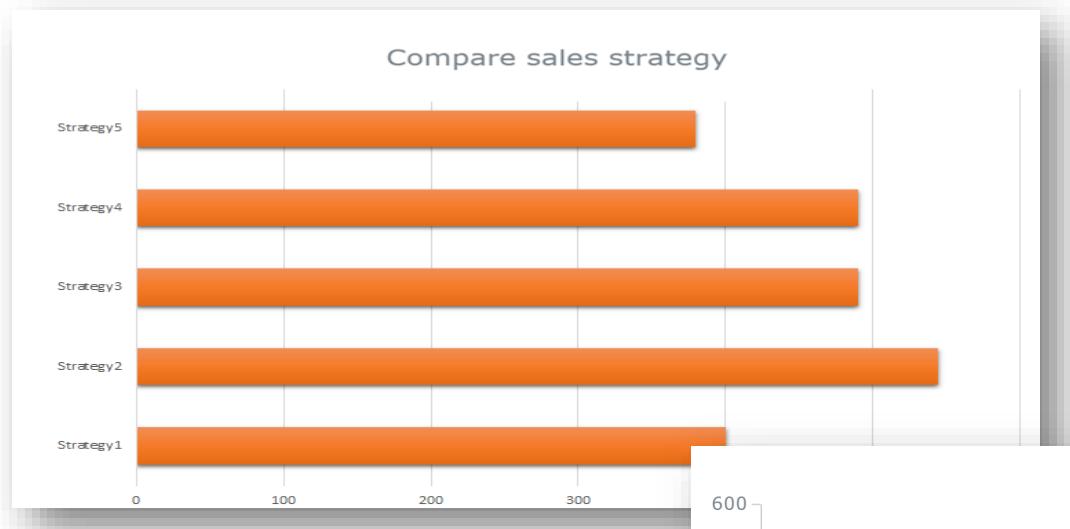
100% STACKED BAR & COLUMN CHART



100% STACKED BAR & COLUMN CHART



BAR AND COLUMN CHARTS

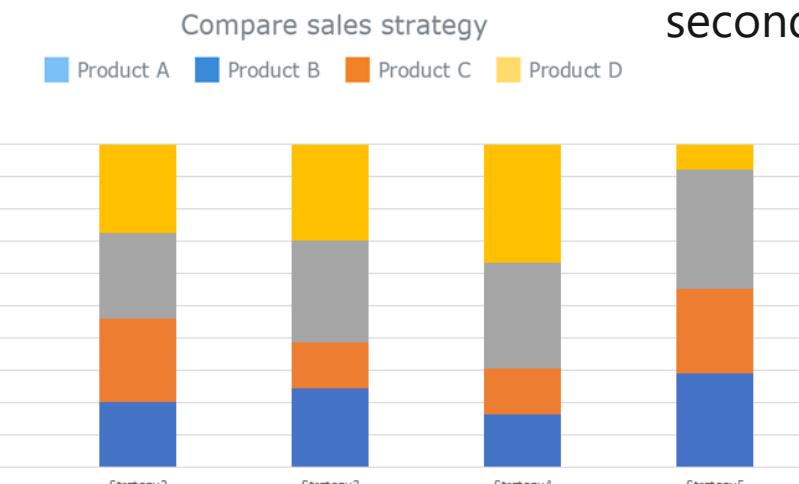
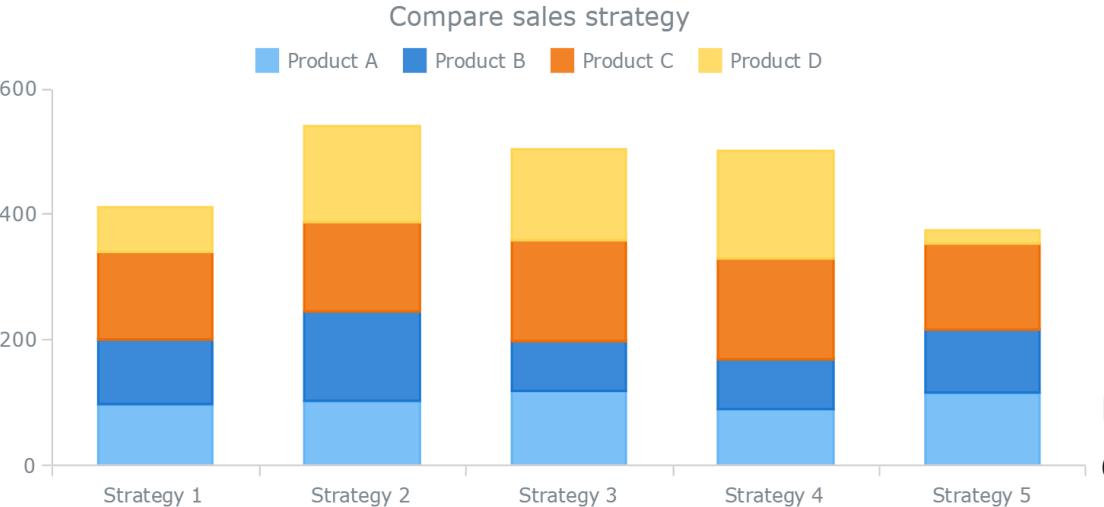


When to use a BAR AND COLUMN chart ?

- presents the value of each category intuitively
- visually for making a comparison of different categories.



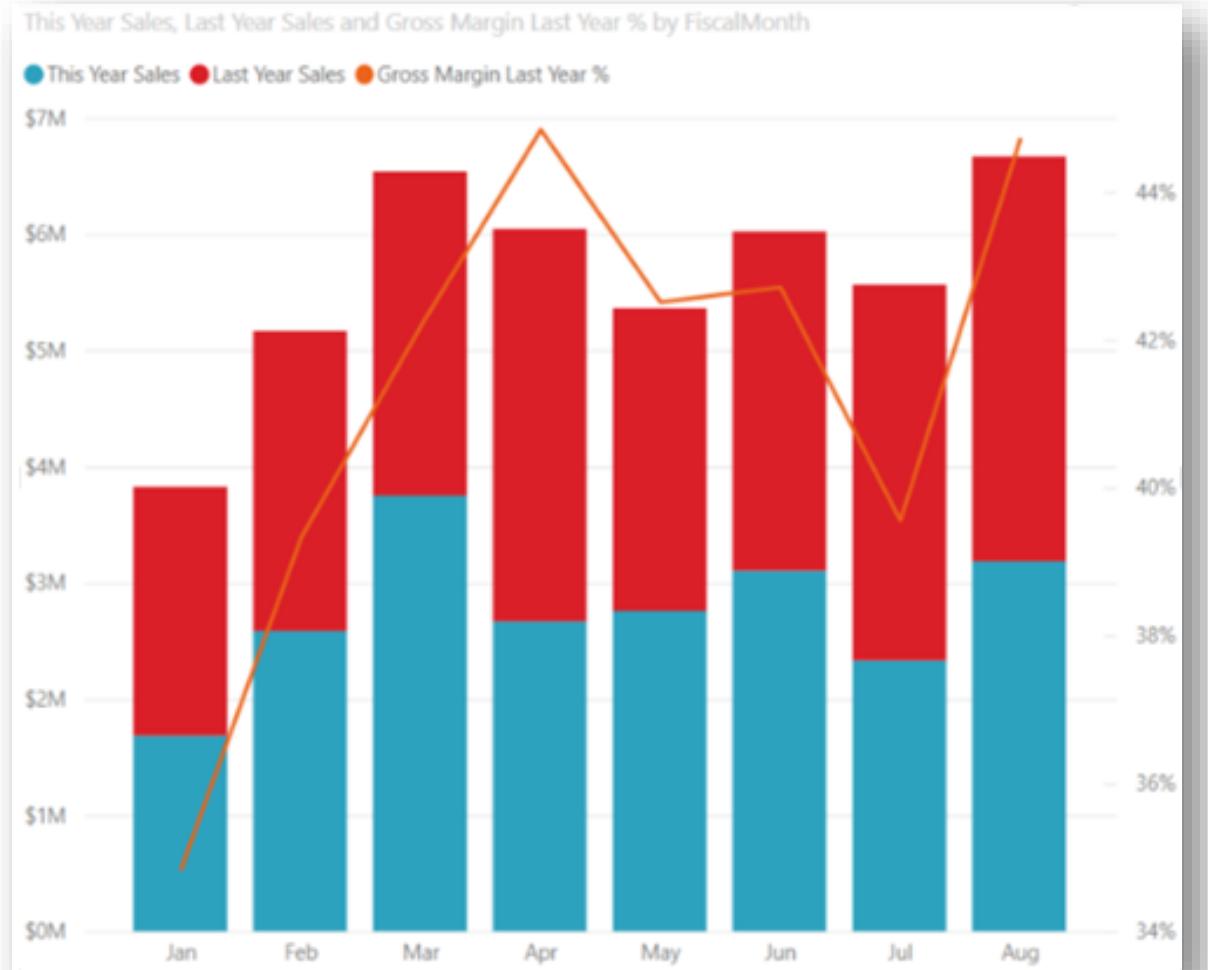
STACKED BAR & COLUMN CHART



When to use a Stacked Bar chart ?

move to a stacked bar chart when we care about the relative decomposition of each primary bar based on the levels of a second categorical variable. Each bar is now comprised of a number of sub-bars, each one corresponding with a level of a secondary categorical variable.

COMBO CHARTS



PIE CHARTS



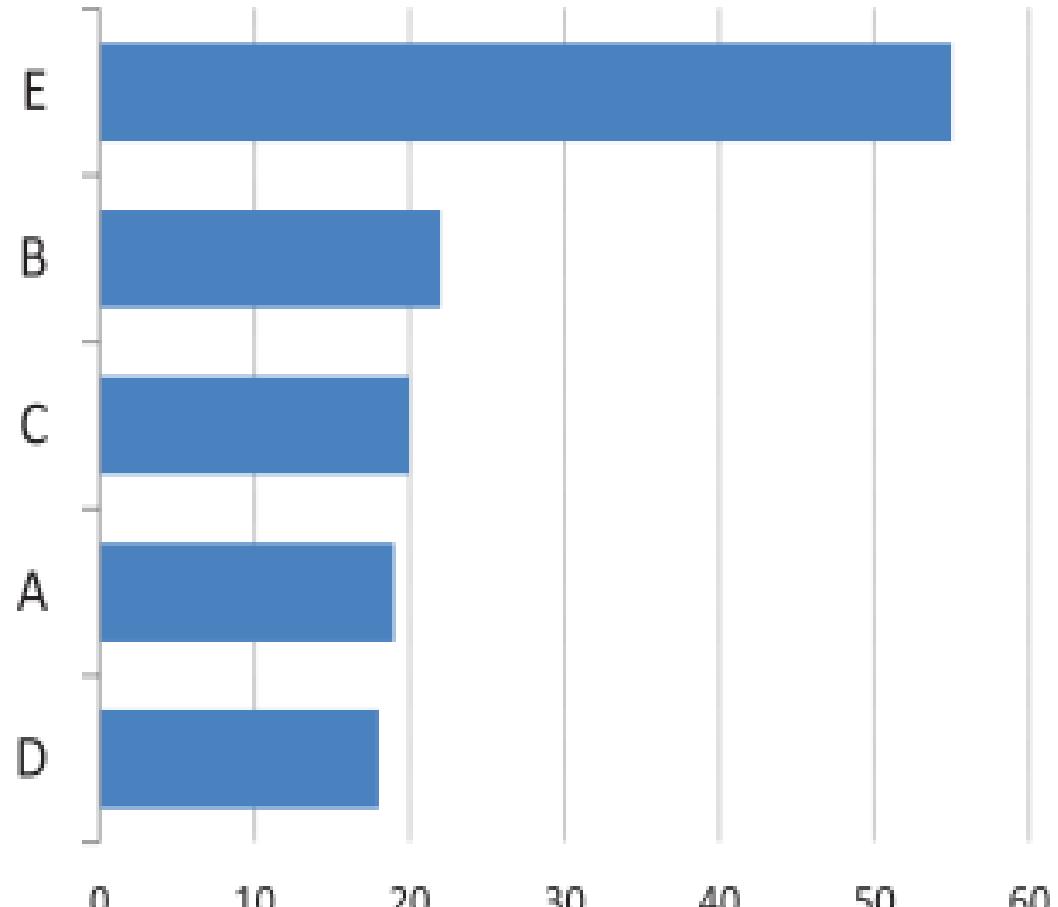
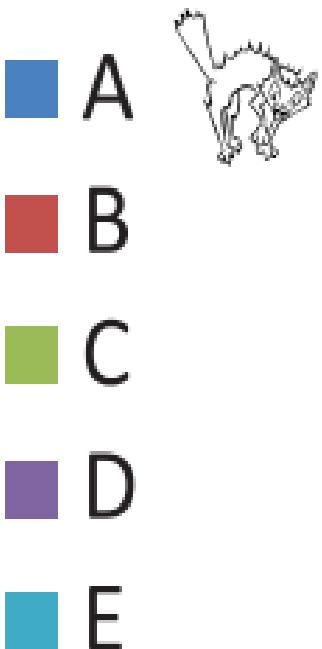
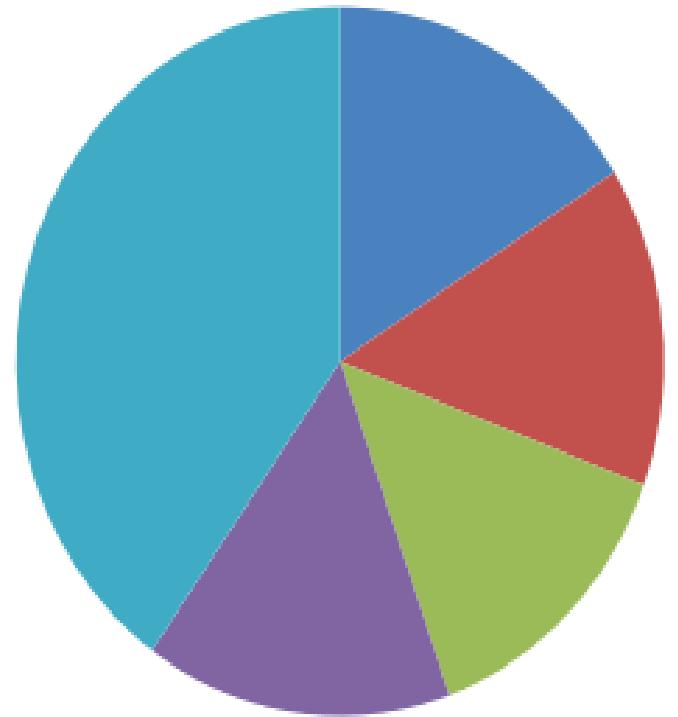
This Year Sales by Chain



Chain ● Fashions Direct ● Lindseys

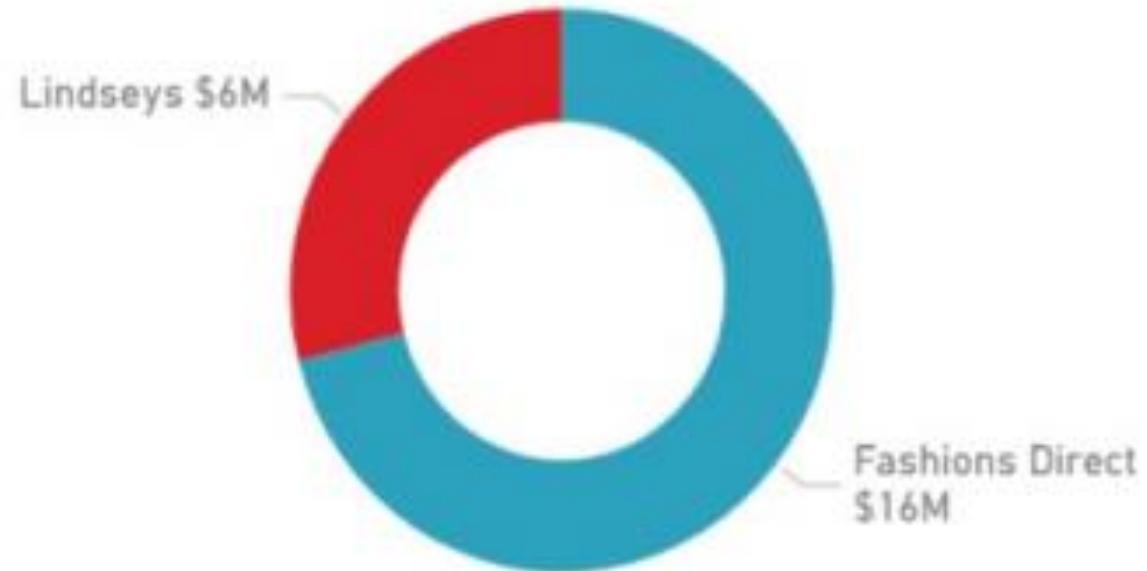
DON'T USE IT

Can you order the slices from biggest to smallest?

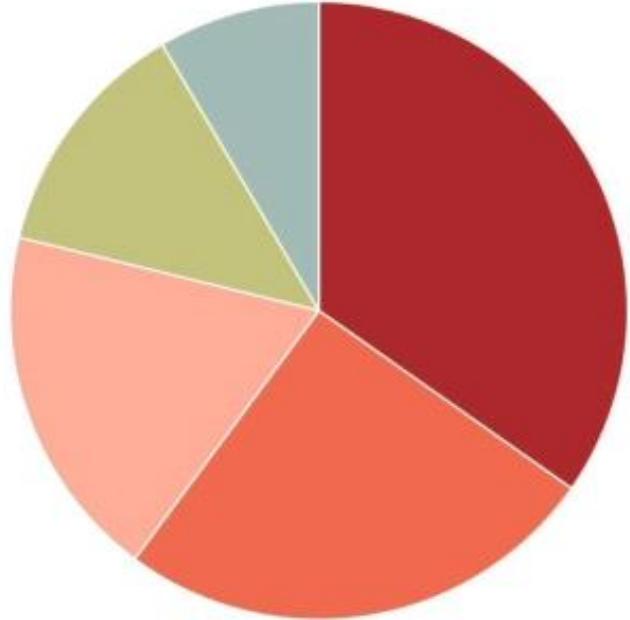


DOUGHNUT CHARTS

This Year Sales by Chain



PIE AND DOUGHNUT CHARTS



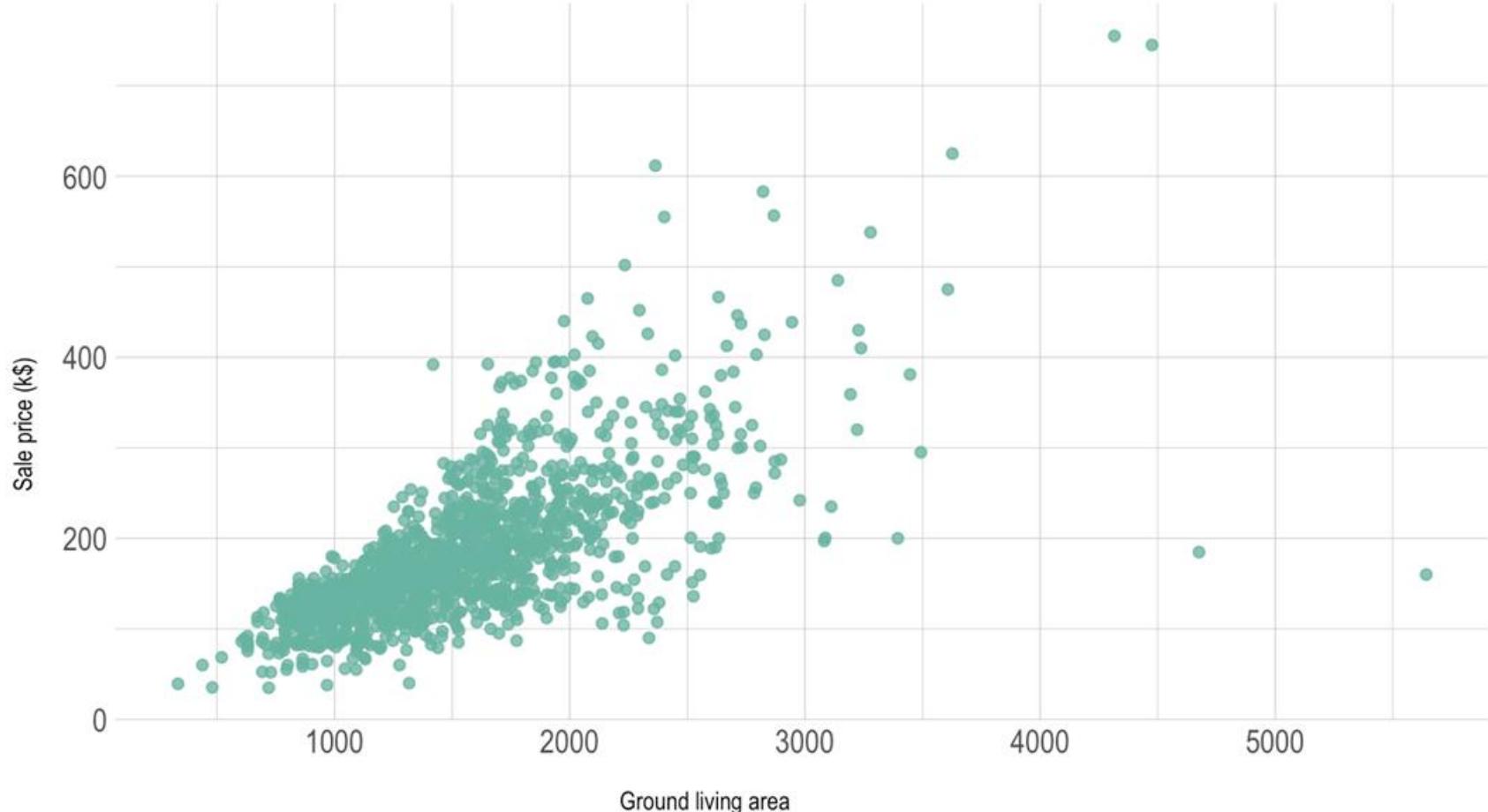
PIE  DONUT

When to use a PIE AND DOUGHNUT charts ?

- when You need to do a part-to-whole data analysis

SCATTER & BUBBLE CHARTS

Ground living area partially explains sale price of apartments



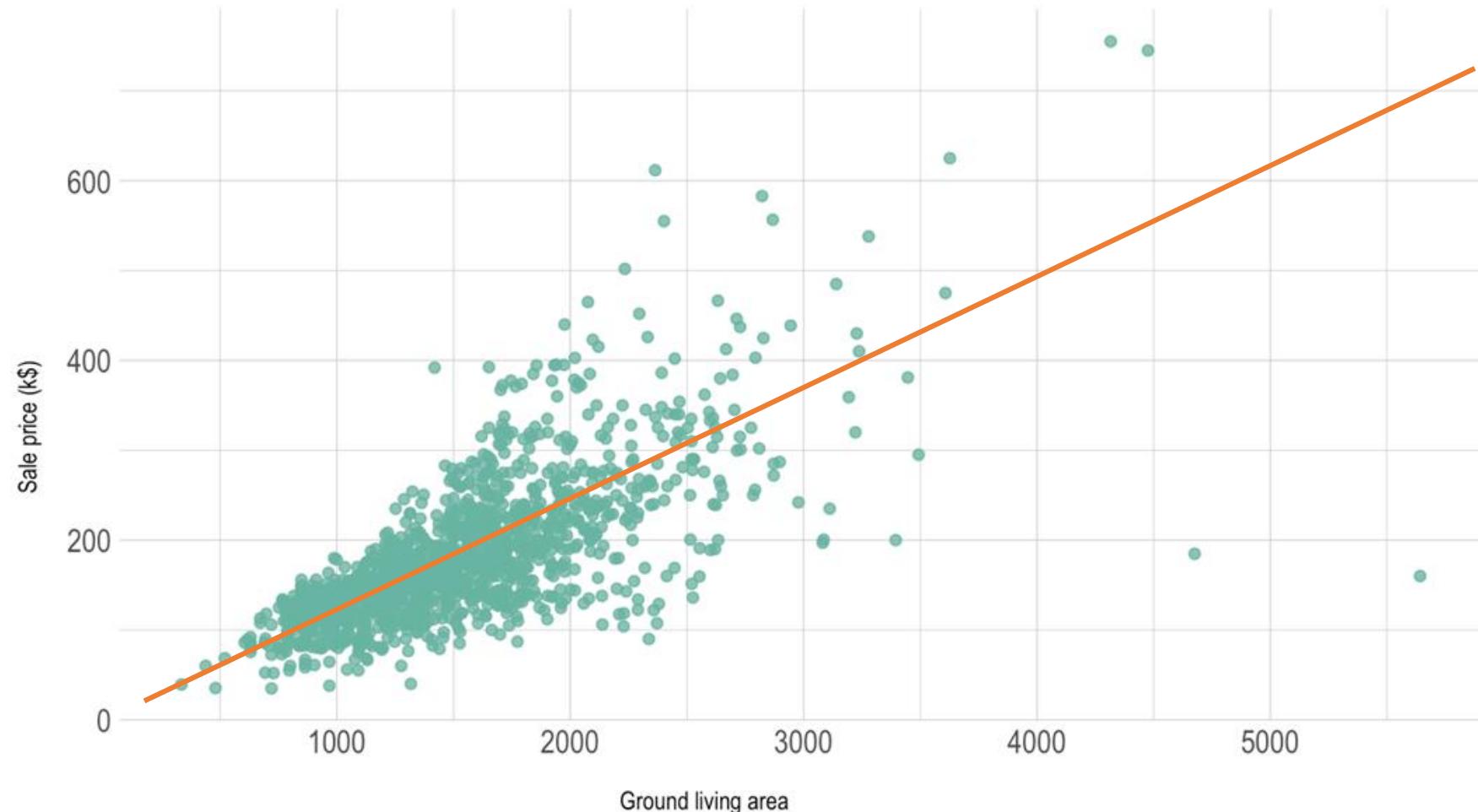
SCATTER & BUBBLE CHARTS

CORRELATION

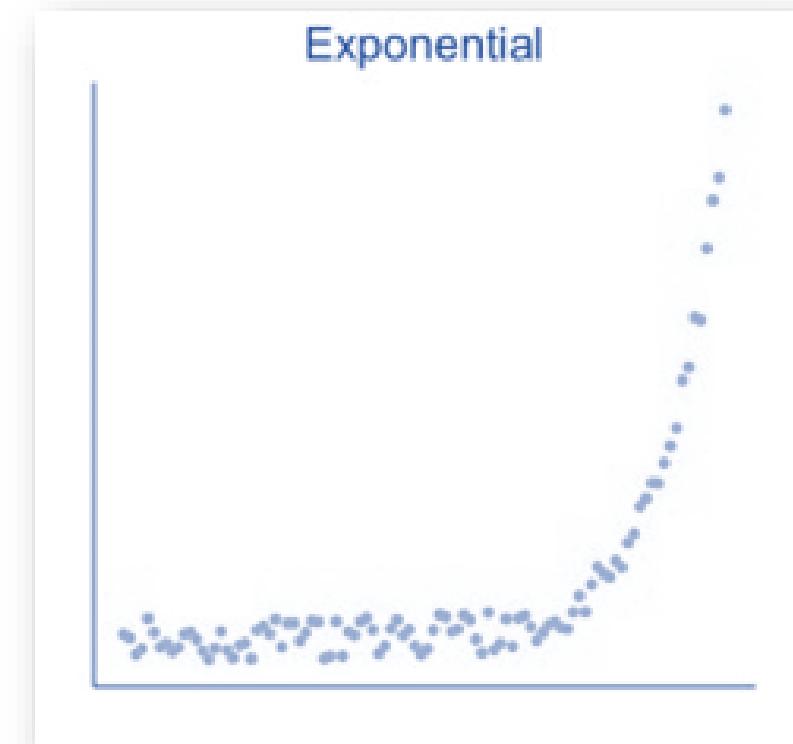
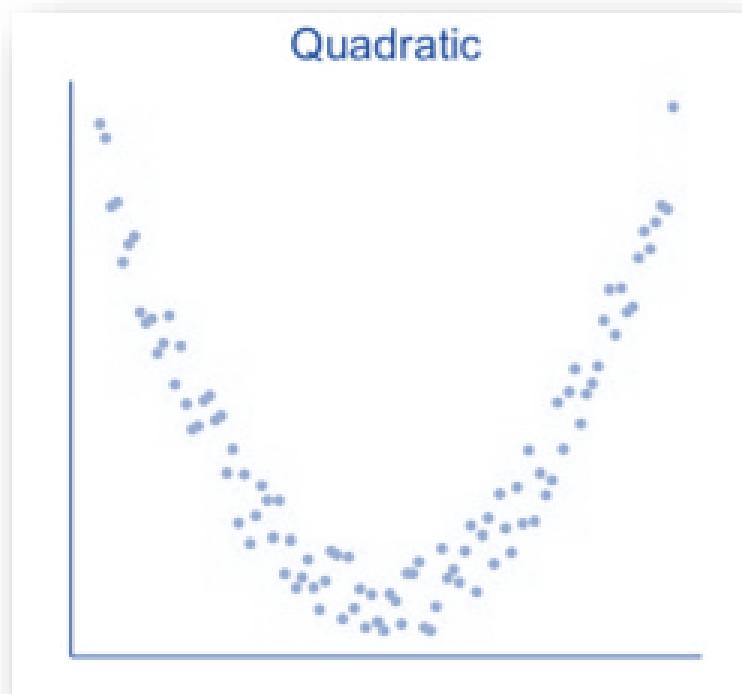
X & Y MOVE TOGETHER



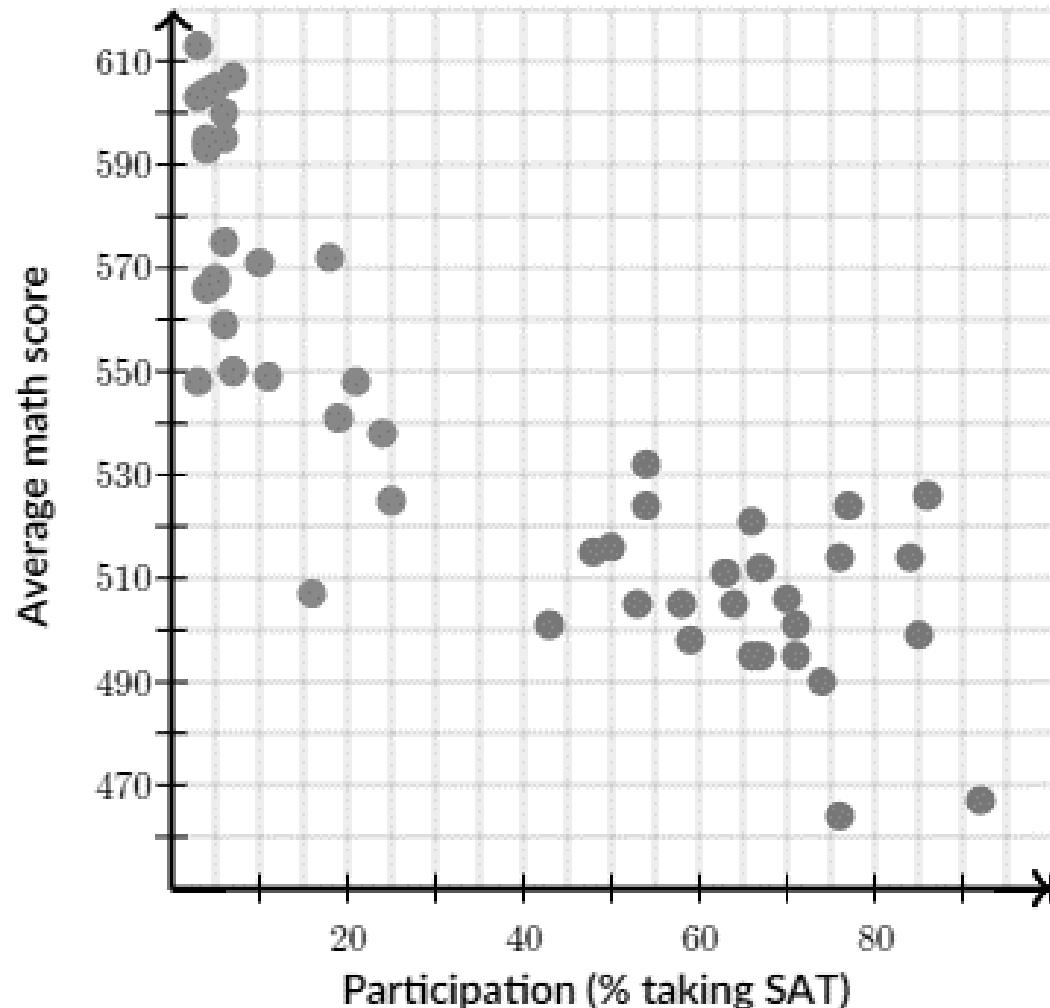
Ground living area partially explains sale price of apartments



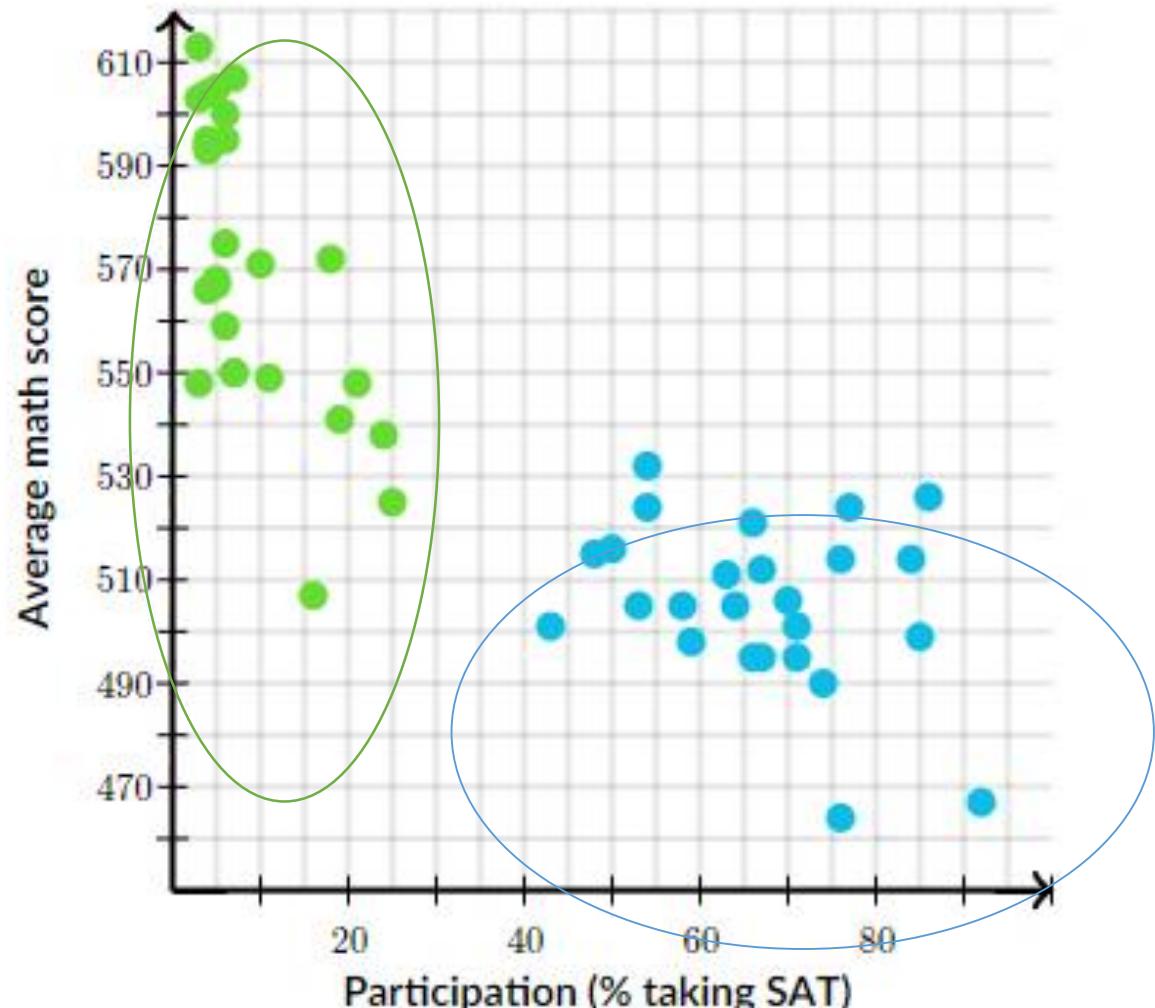
SCATTER & BUBBLE CHARTS



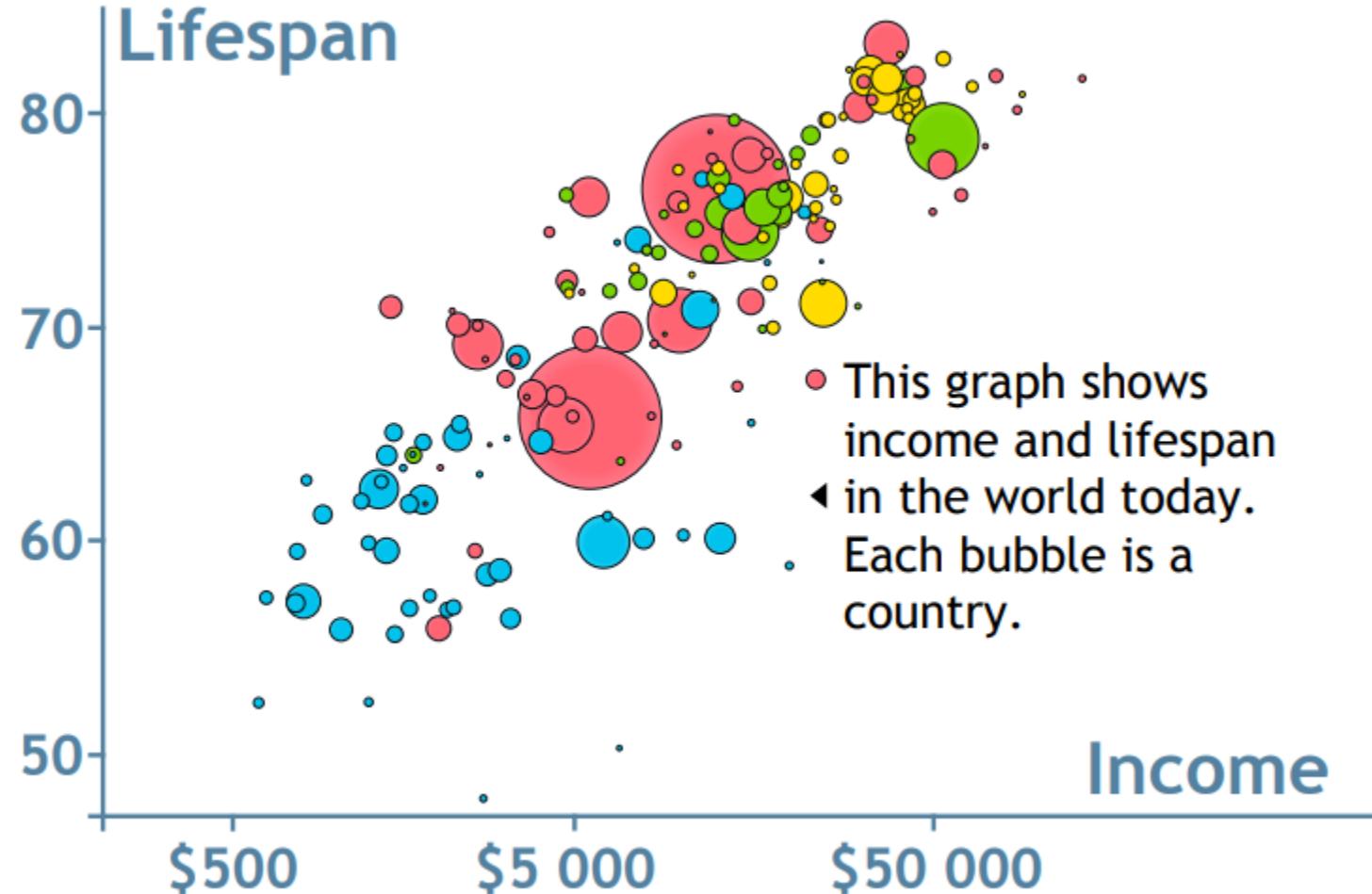
SCATTER & BUBBLE CHARTS



SCATTER & BUBBLE CHARTS

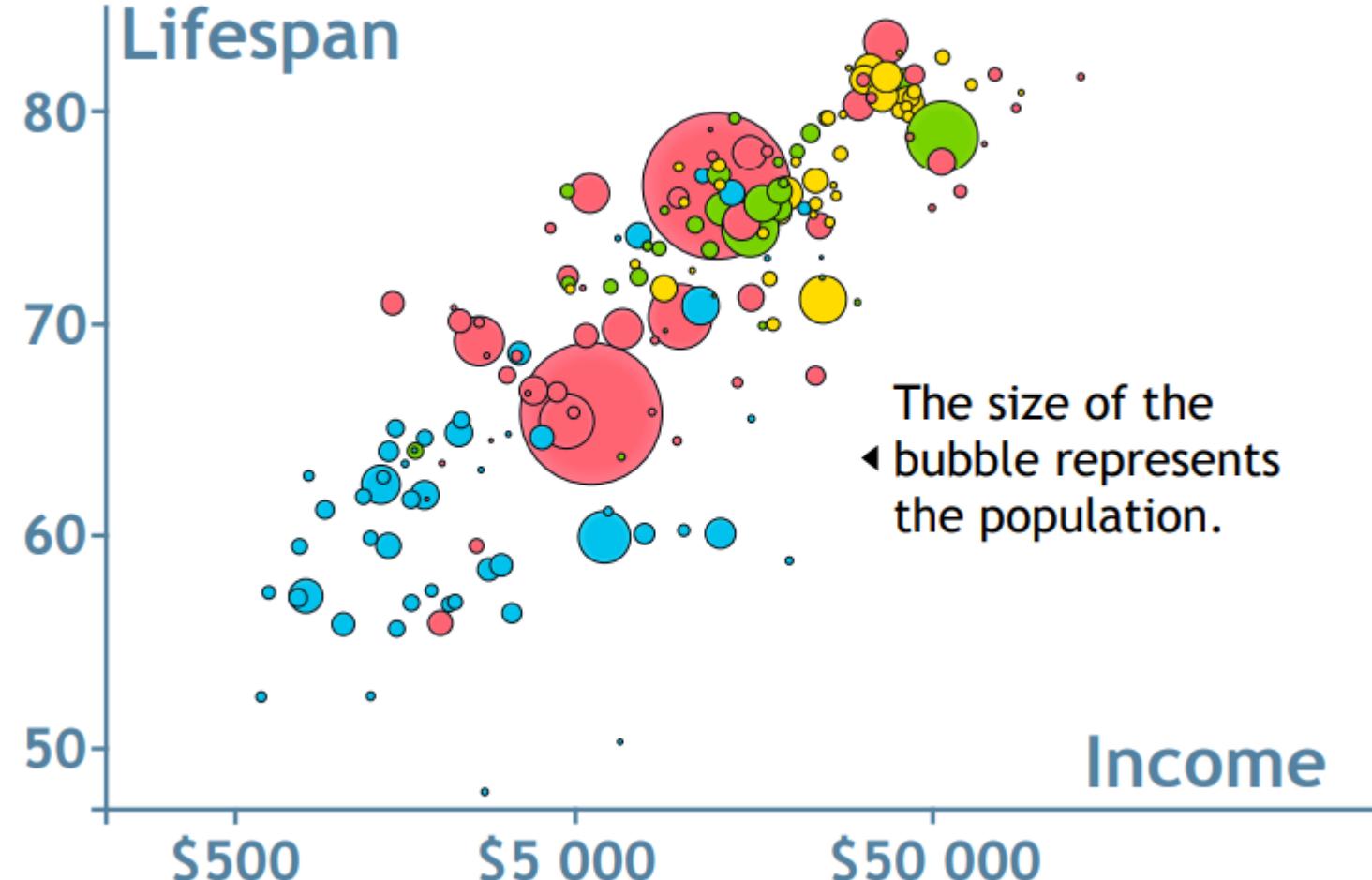


SCATTER & BUBBLE CHARTS

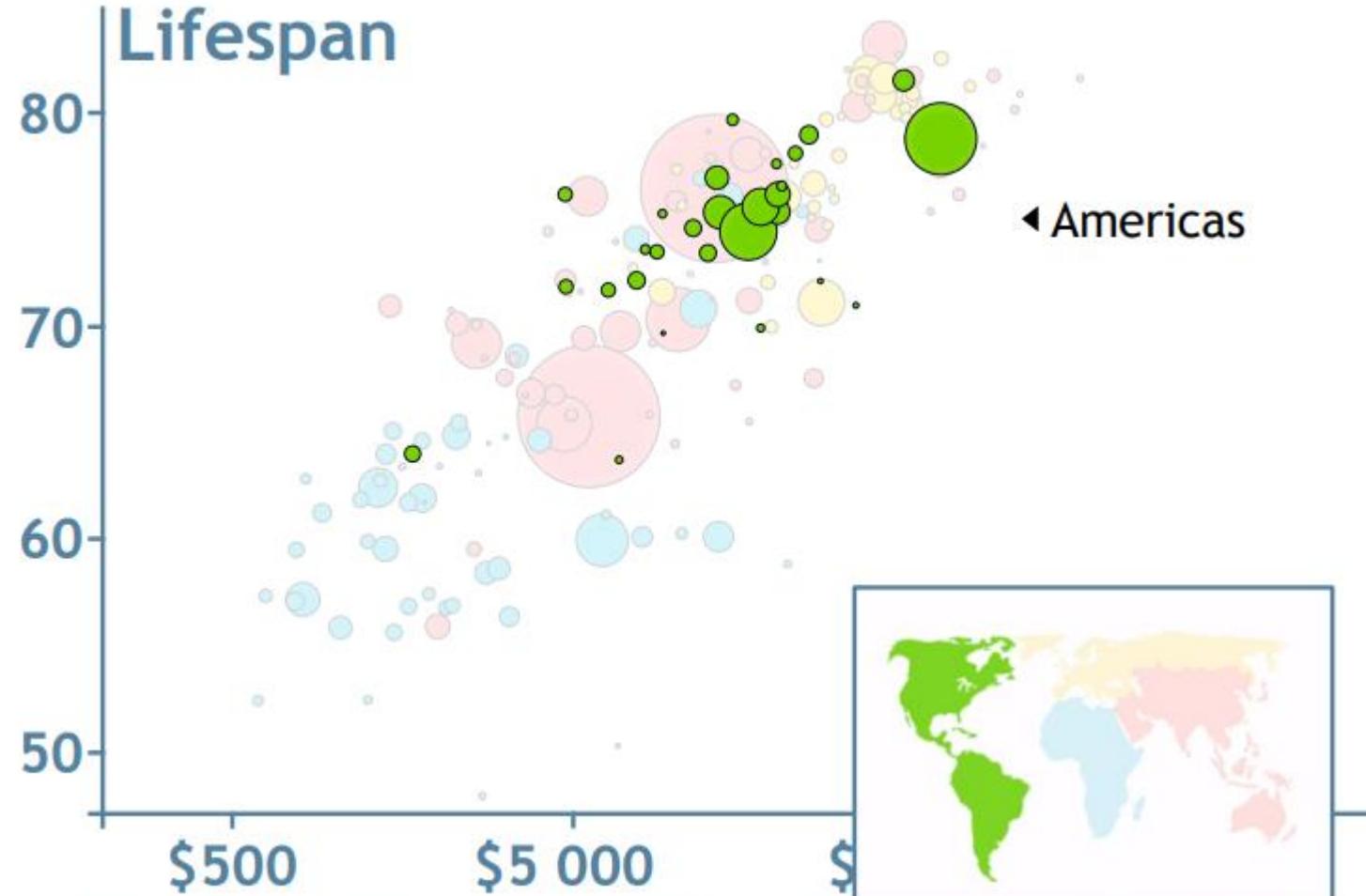


Sources: After 1950: UN World Pop. Prospr. 2012. Before 1950: hundreds of sources combined by Gapminder.

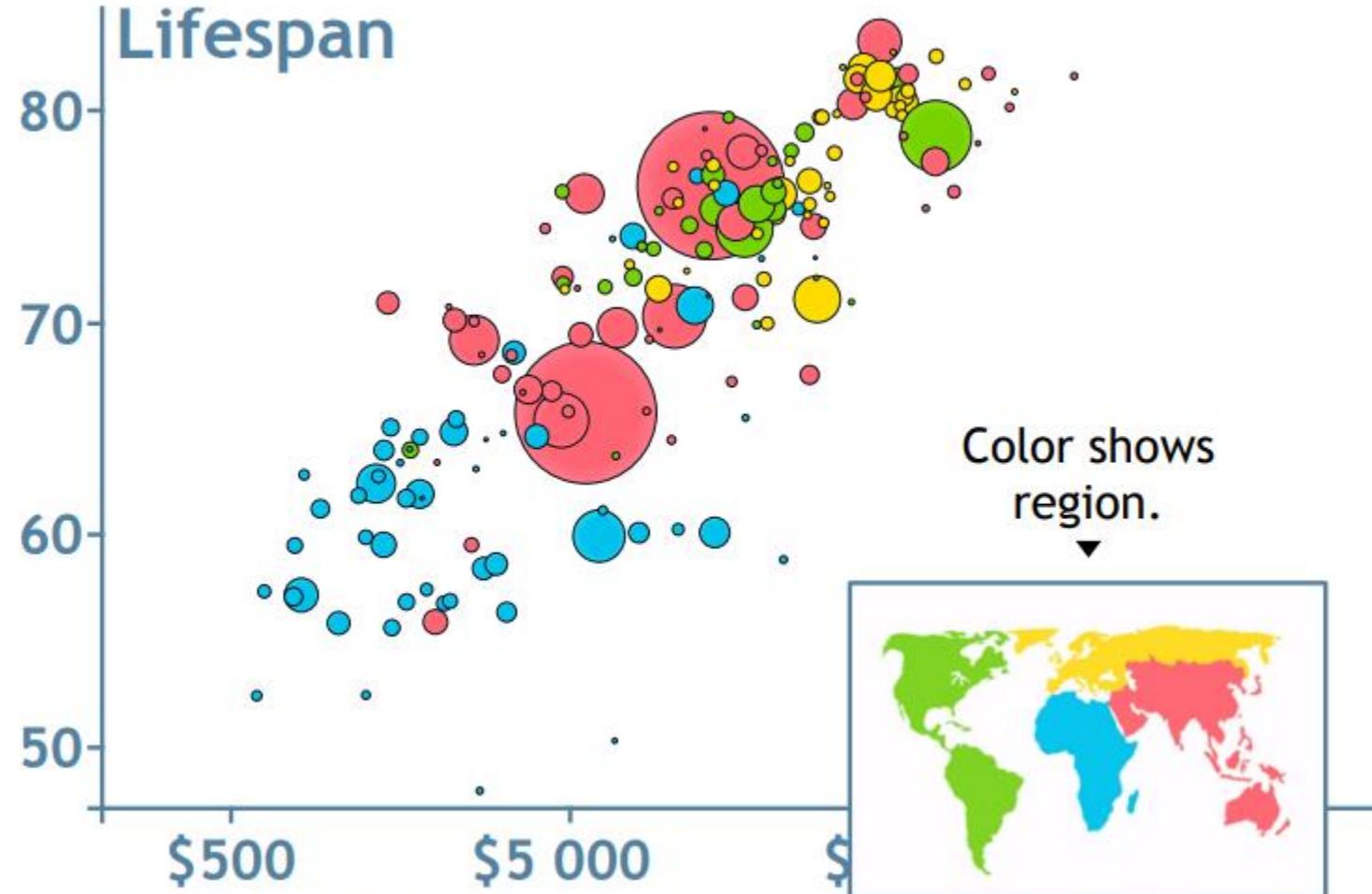
SCATTER & BUBBLE CHARTS



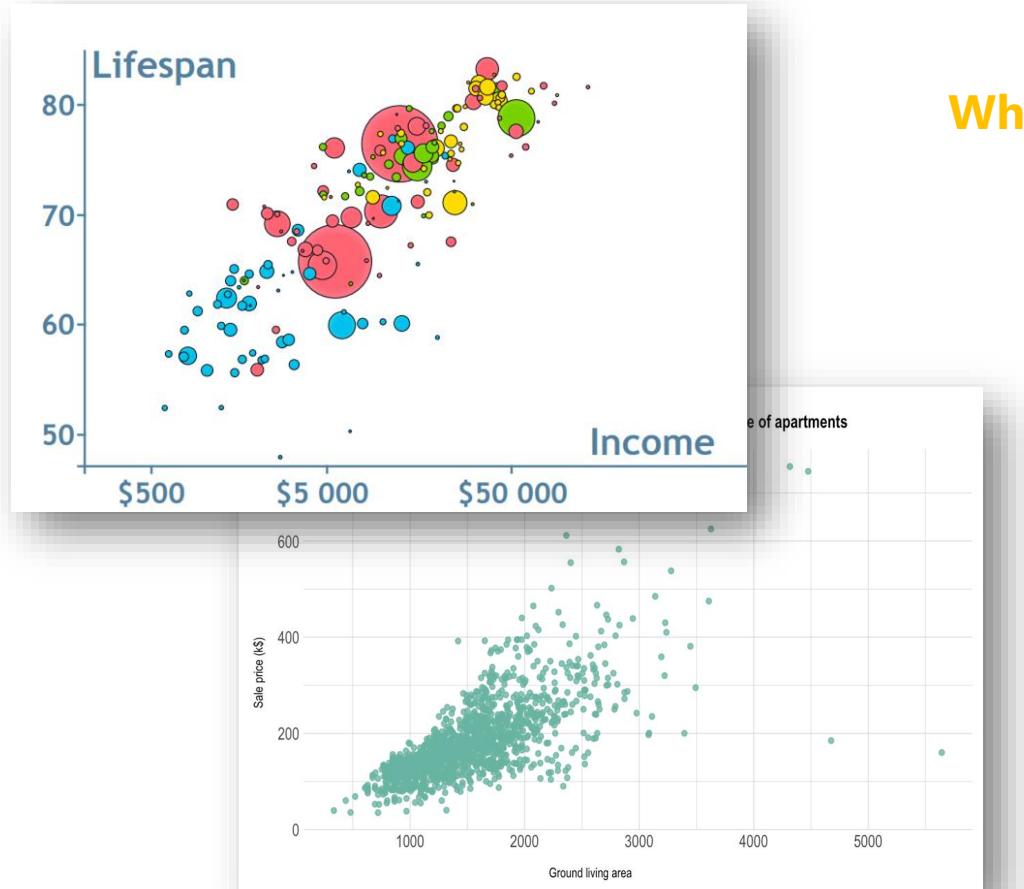
SCATTER & BUBBLE CHARTS



SCATTER & BUBBLE CHARTS



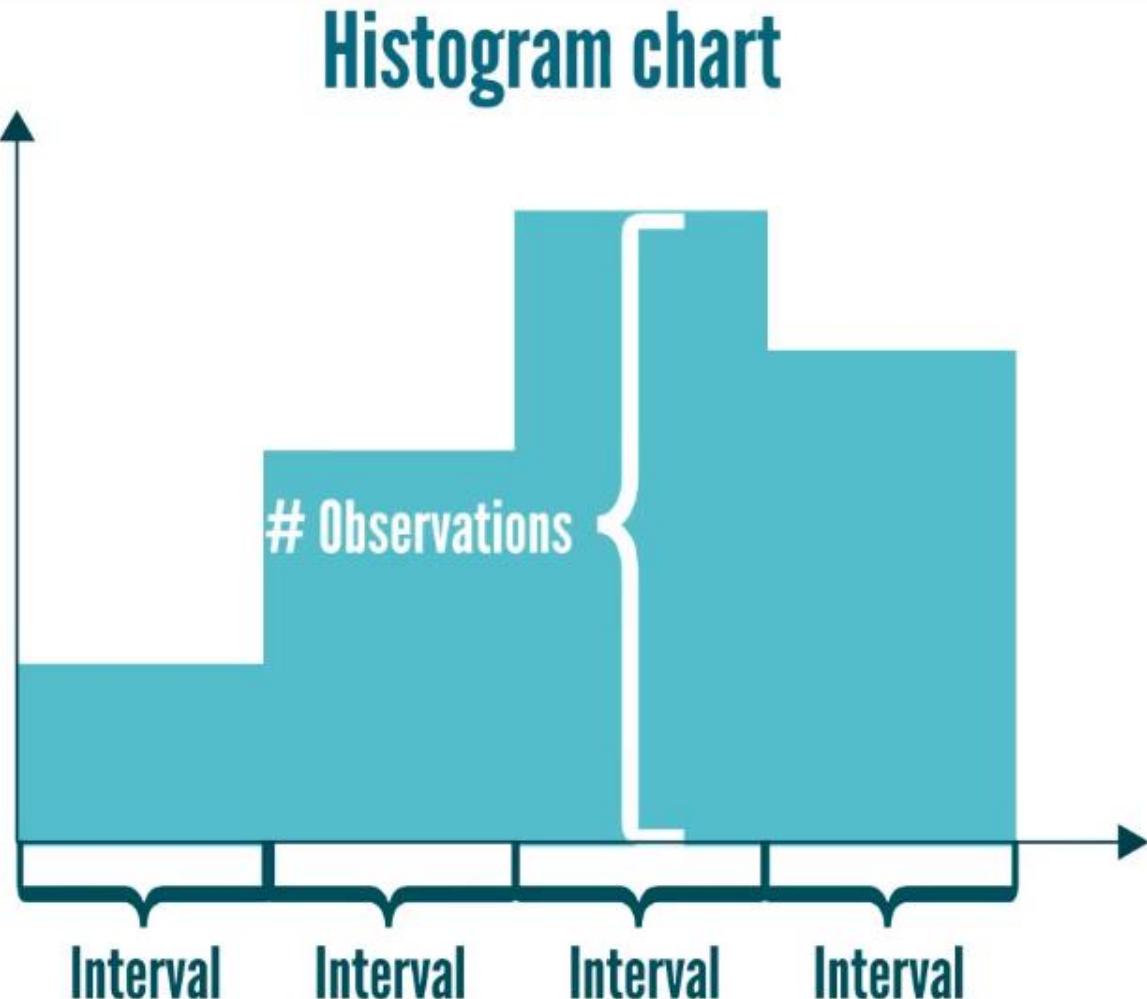
SCATTER & BUBBLE CHARTS



When to use a scatter chart, bubble chart charts?

- A scatter chart shows the relationship between two numerical values
- If your data has three data series that each contains a set of values.
- To present financial data. Different bubble sizes are useful to visually emphasize specific values.
- To show patterns in large sets of data, for example by showing linear or non-linear trends, clusters, and outliers.

Histograms



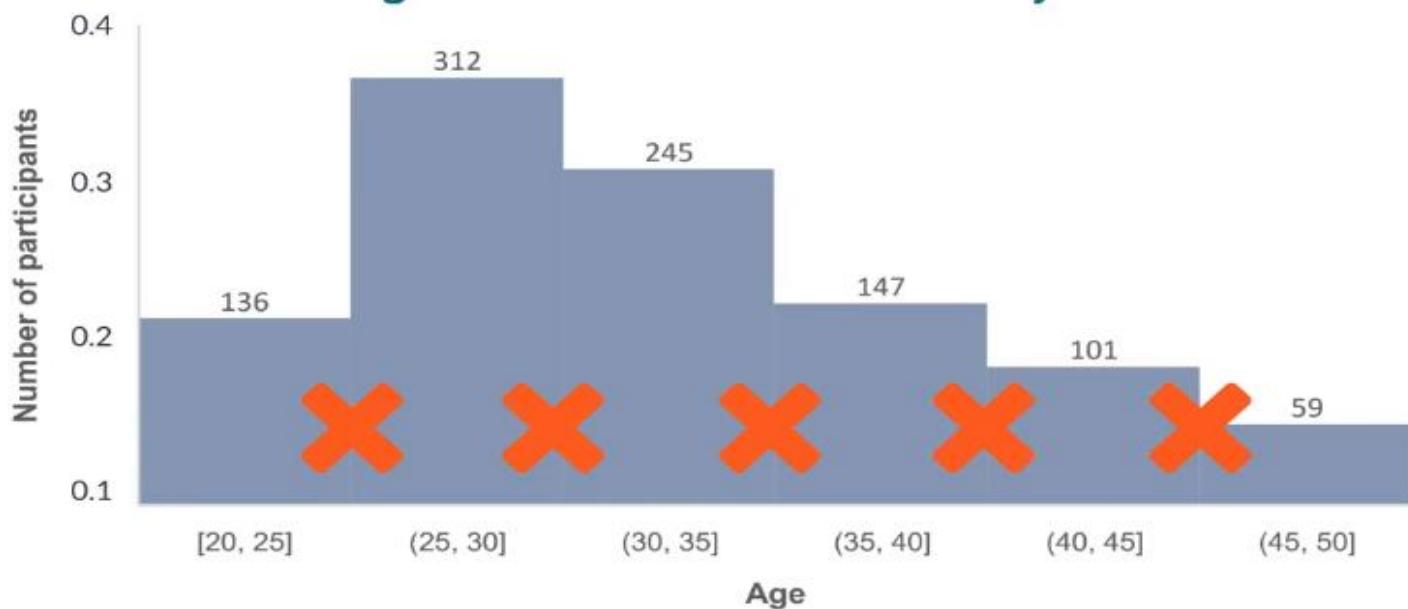
- Shows the distribution of a numeric variable
- The variable's range of values is split into intervals, represented by different bins
- The height of the bins shows the number of observations within an interval

Histograms

Histogram chart

- Represents bins (intervals)
- Bars are not separated

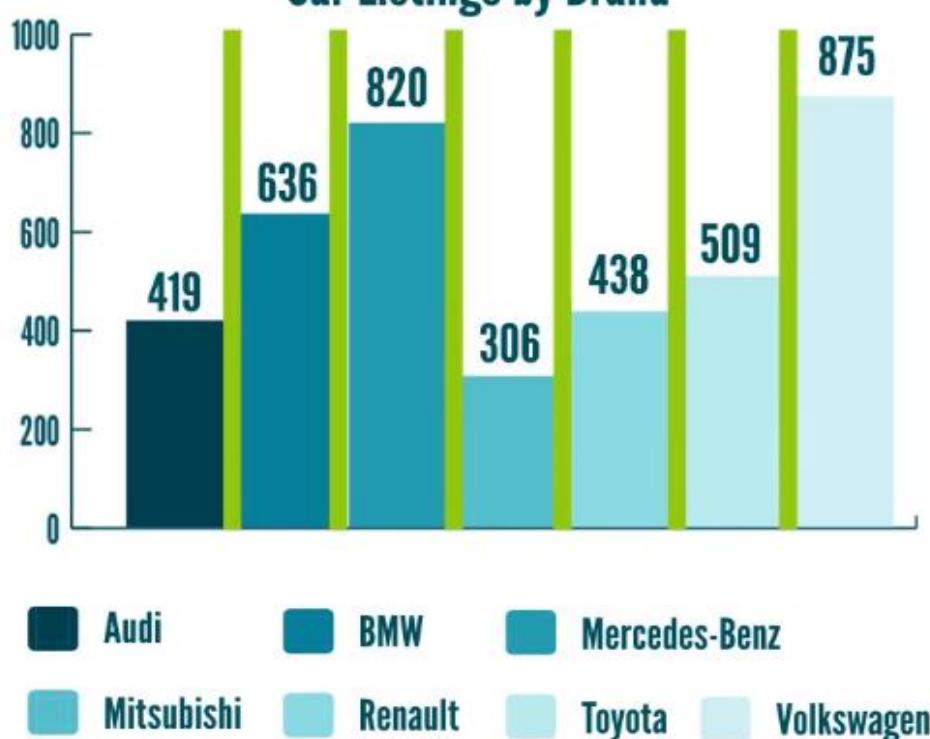
Age Distribution in Customers' Survey



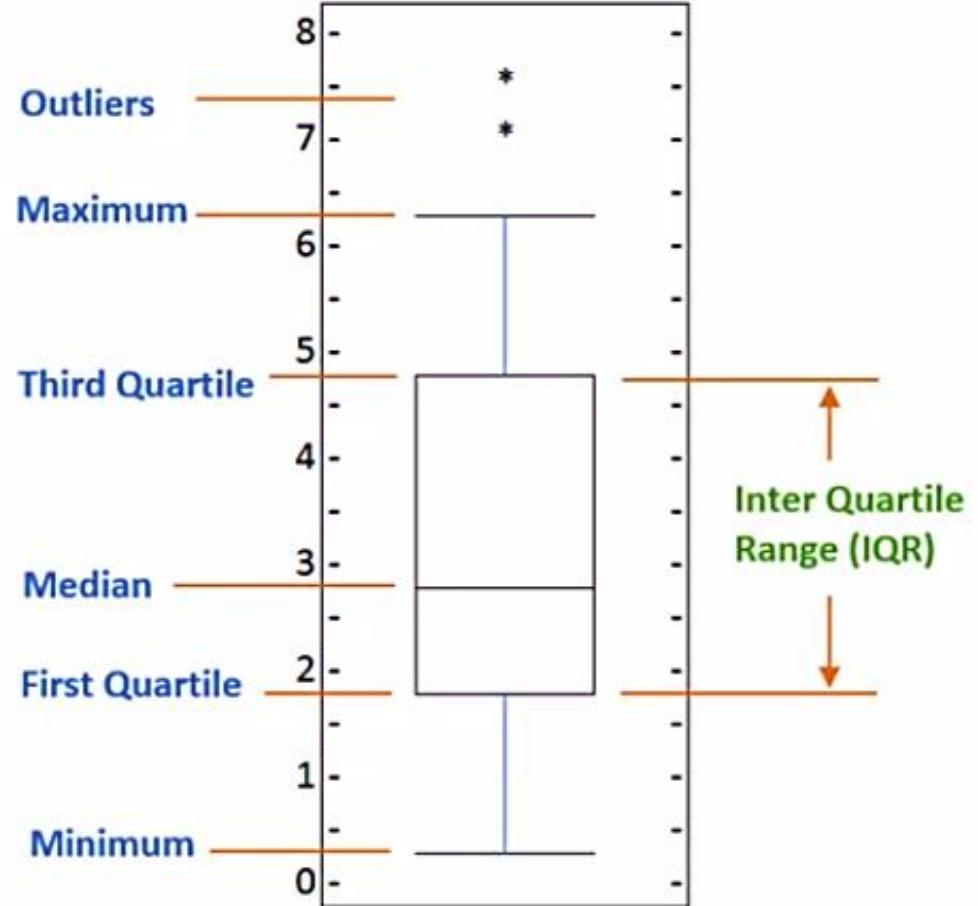
Bar chart

- Represents different categories
- Bars are separated

Car Listings by Brand

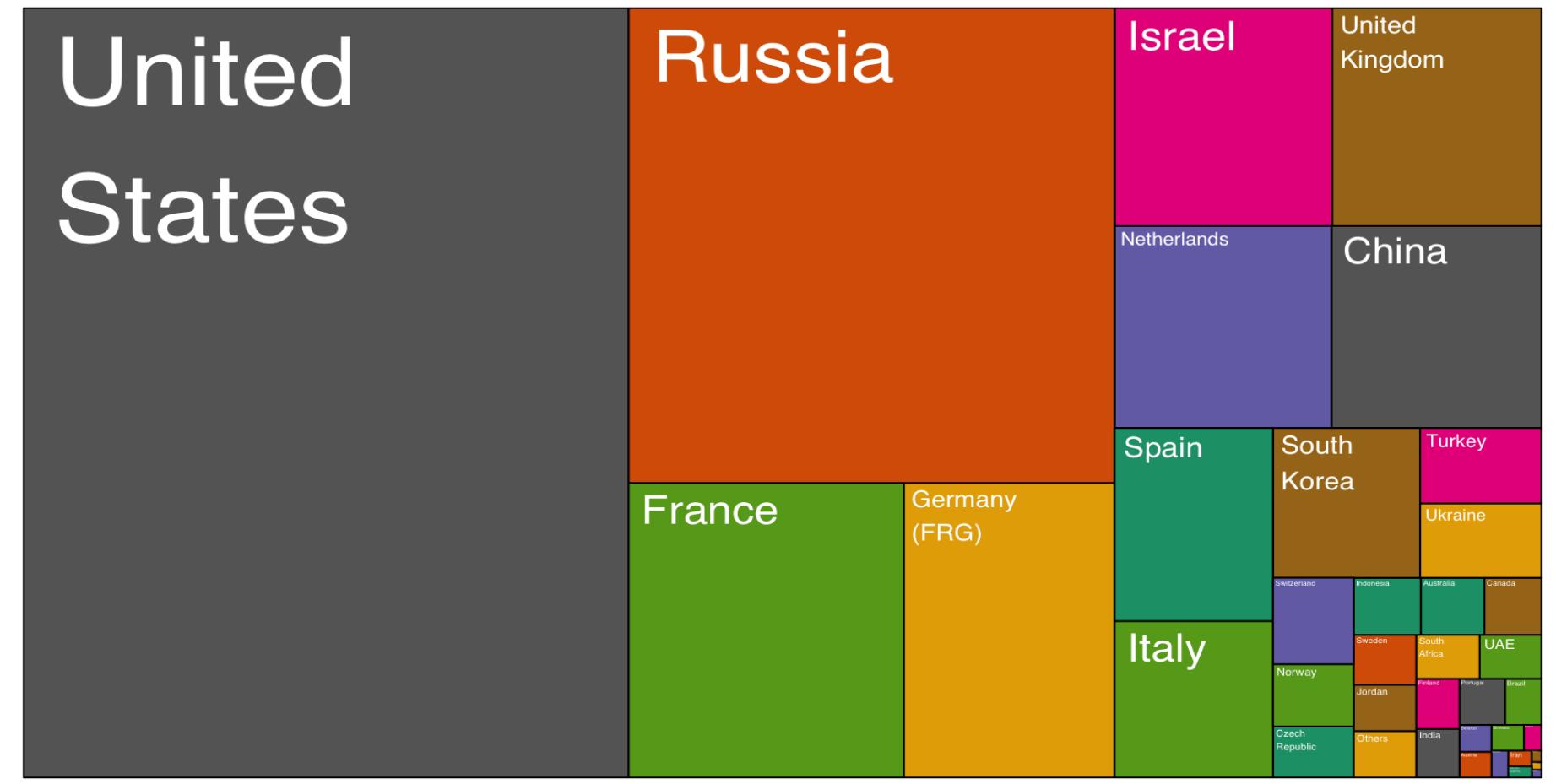


Box Plots

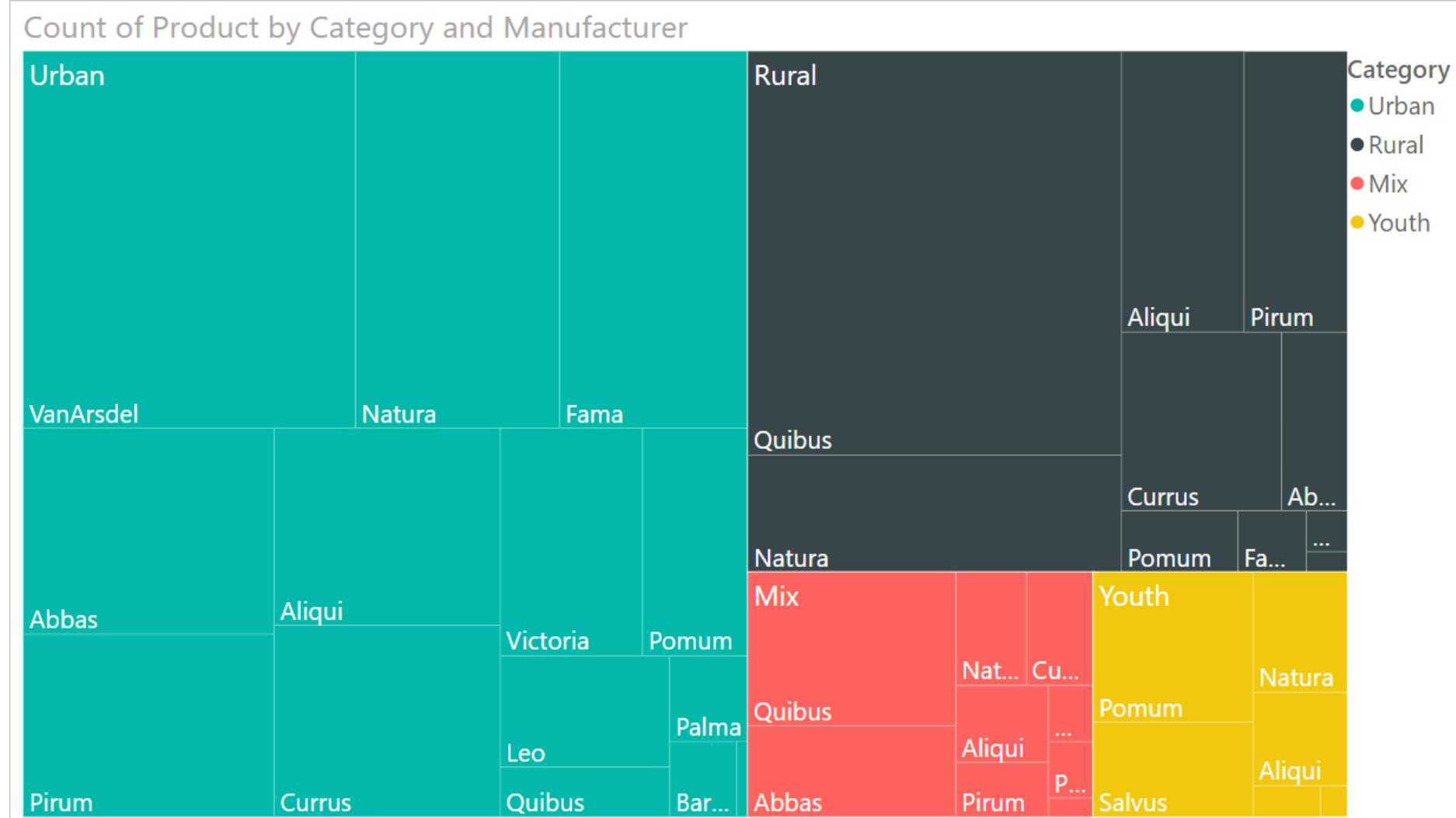


TreeMaps

WHO SELLS MORE WEAPONS?



TreeMaps



TreeMaps



When to use a TREEMAP ?

- To display large amounts of **hierarchical** data.
 - When a bar chart can't effectively handle the large number of values.
 - To show the **proportions** between each part and the whole.



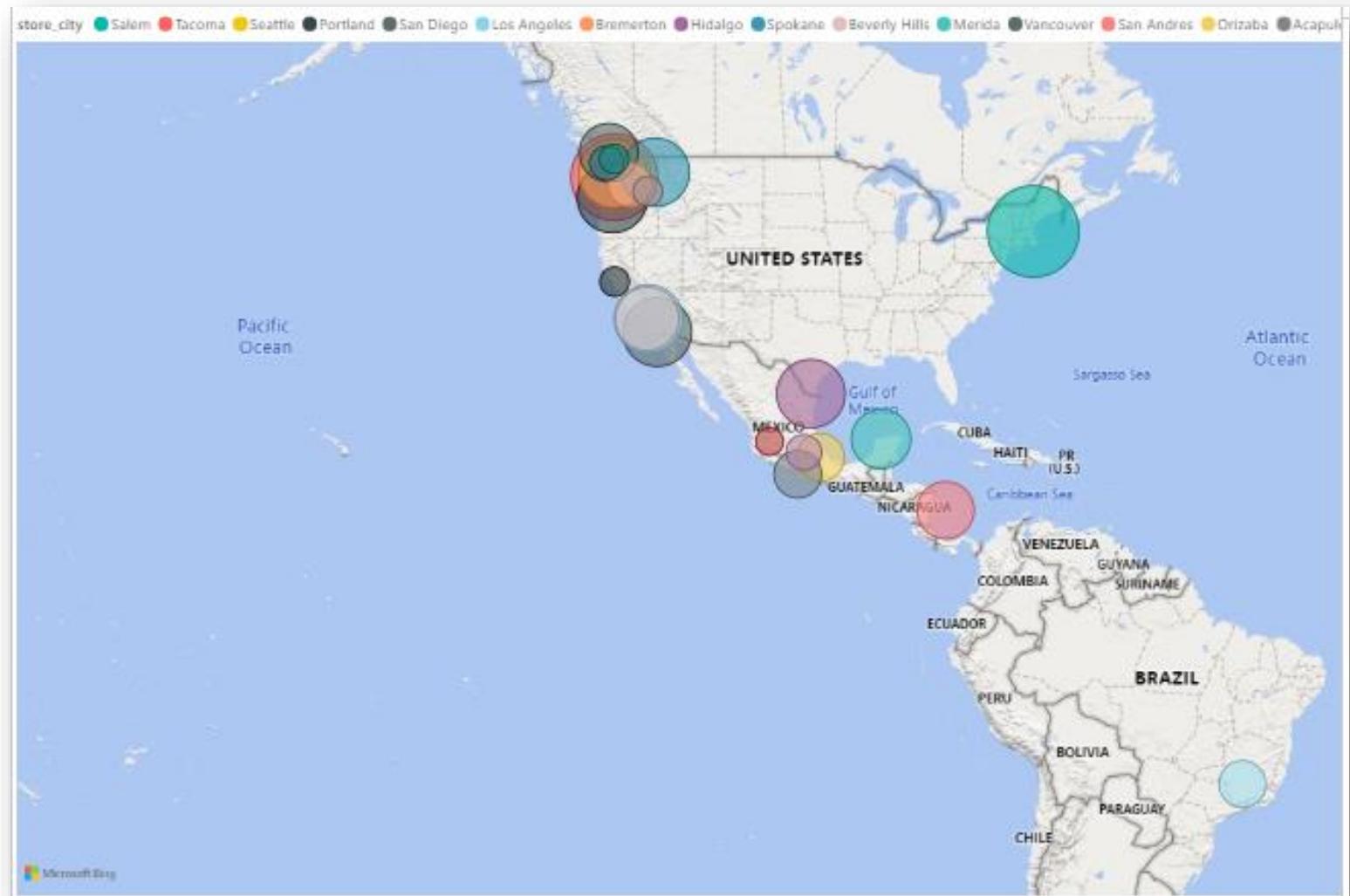
Word Cloud



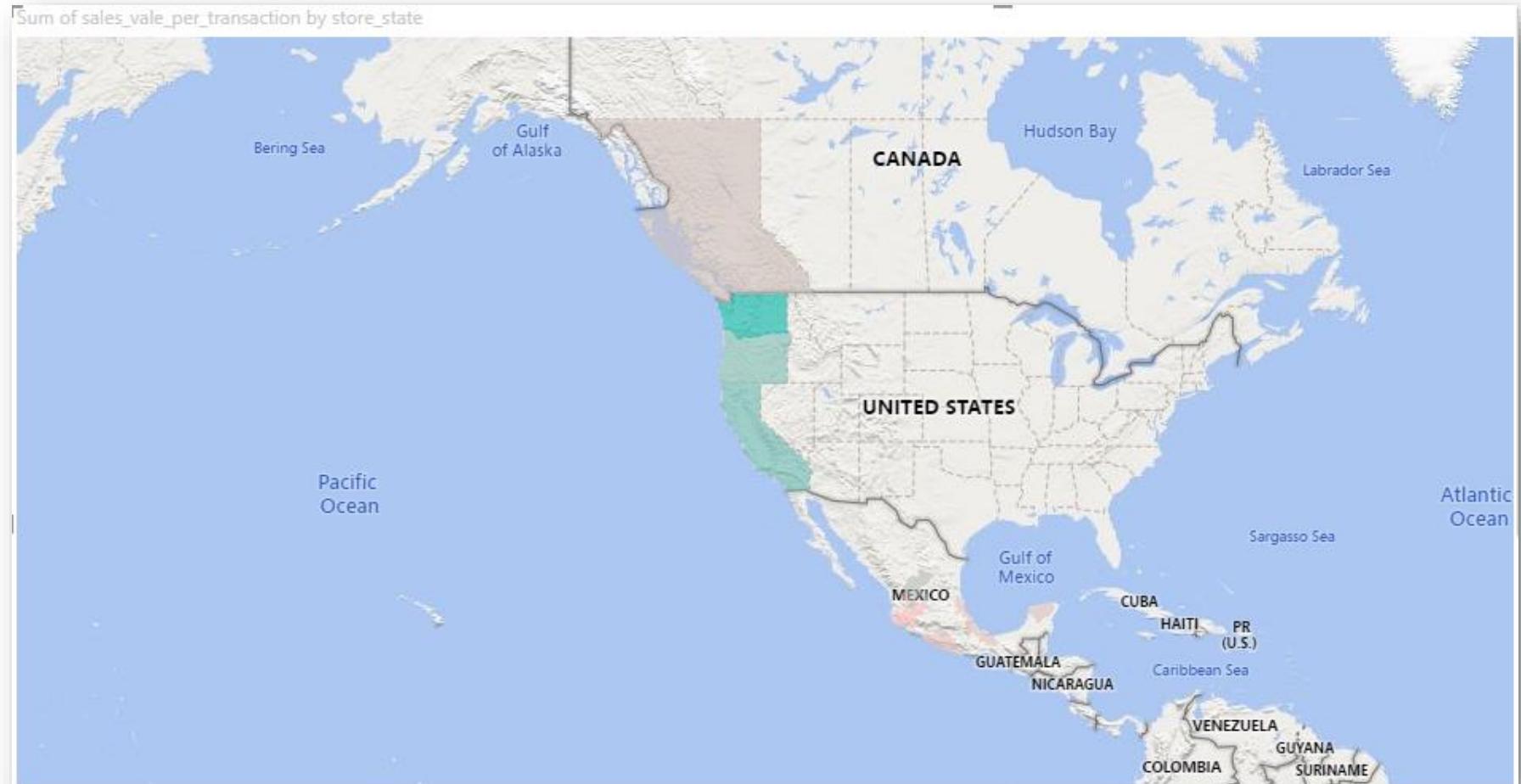
A word cloud centered around the words COVID-19, CORONAVIRUS, EPIDEMIC, and PNEUMONIA. The words are arranged in a cluster, with larger words like COVID-19 and CORONAVIRUS being the most prominent. Other visible words include MEDICAL, HEALTH, SAFETY, ATTENTION, PNEUMONIA, RISK, SPREADING, HUMAN, HAZARDOUS, FEAR, VIRUS, DISTANCING, FEVER, DOCTOR, ANALYZING, MEDICINE, HAZARD, PEOPLE, CAUTION, DEATH, PREVENTION, VIROLOGY, OUTBREAK, IMMUNITY, PATHOLOGY, SCIENCE, INFECTIOUS, FEVER, SICK, CARE, WORLD, PATHOGEN, VACCINE, WARNING, SYMPTOM, CONTAGIOUS, LABORATORY, WUHAN, and ALERT.

INFECTED INFECTION
DIAGNOSTIC
CONTAMINATION
BIOHAZARD NATION
HEALTH CARE
PANDEMIC
HOSPITAL
CONTROL
REPORT
TERM
LOSE
OFFICIAL
SITUATION
TESTING
CHEST
REPORT
CONFIRMED
RISK
SPREADING
HUMAN
HAZARDOUS
FEAR
VIRUS
DISTANCING
FEVER
DOCTOR ANALYZING
MEDICINE
HAZARD
PEOPLE
CAUTION
DEATH
PREVENTION
VIROLOGY
OUTBREAK
IMMUNITY
PATHOLOGY
SCIENCE
INFECTIOUS
FEVER
SICK
CARE
WORLD
PATHOGEN
VACCINE
WARNING
SYMPTOM
CONTAGIOUS
LABORATORY
WUHAN
ALERT
PATIENT

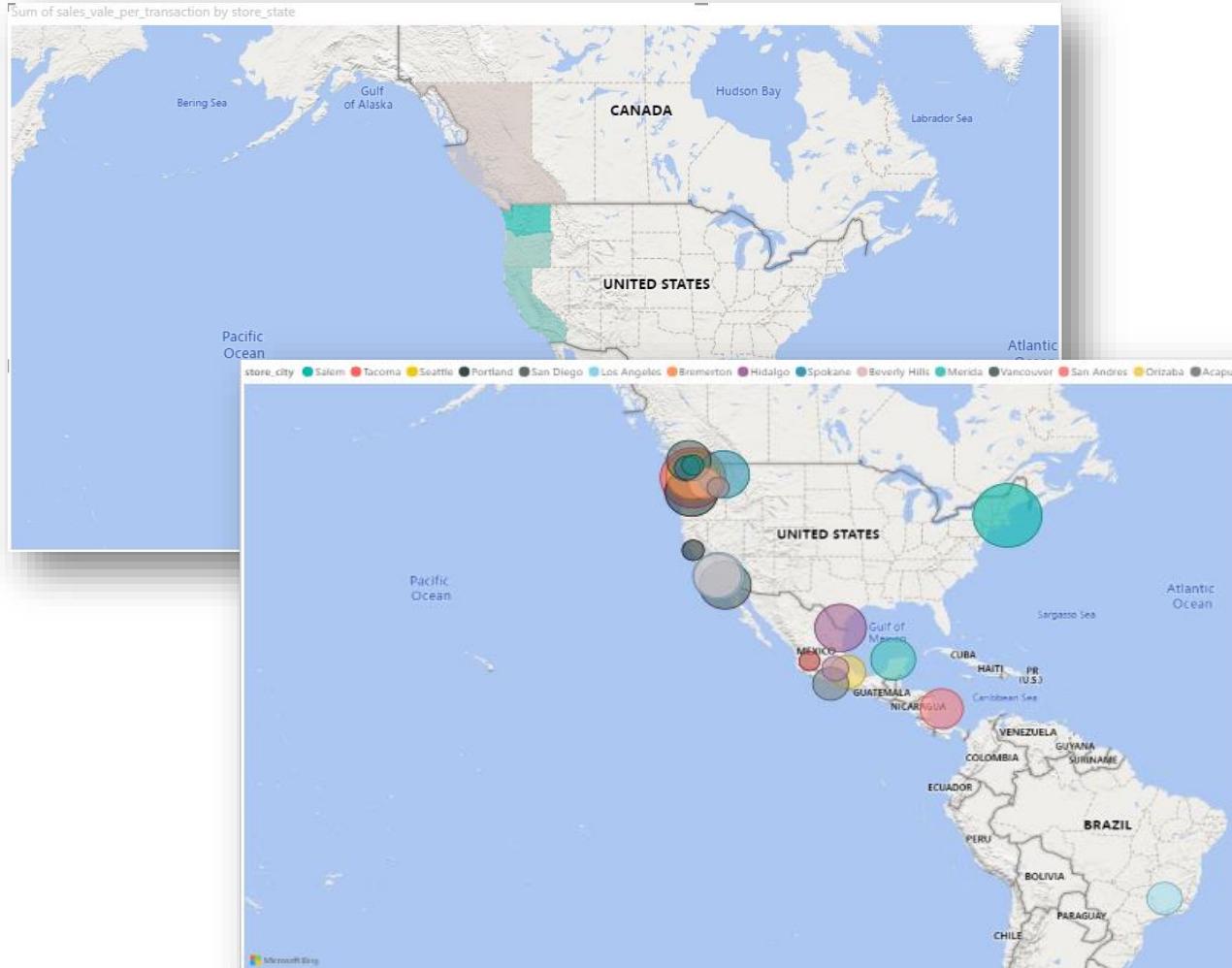
MAP VISUALIZATIONS



CHOROPLETH MAPS

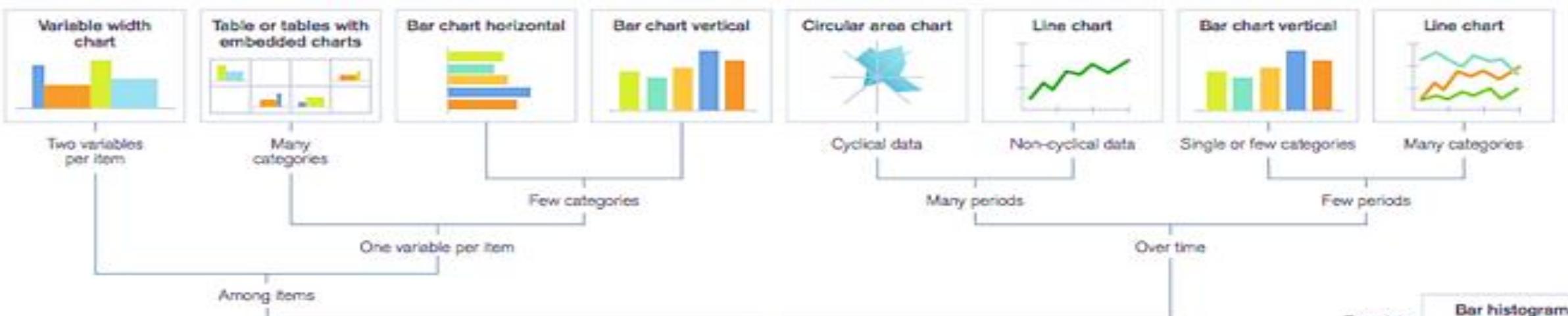


MAPS



When to use MAPS ?

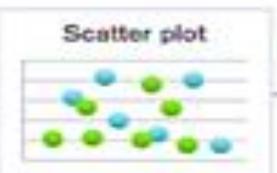
- to display quantitative information on a map.
- to show spatial patterns and relationships.
- when working with socioeconomic data.
- to get an overview of the distribution across the geographic locations.



Comparison

What would you like to show?

Relationship



Two variables



Three variables

Distribution



Few data points



Many data points



Two variables

Composition

Changing over time

Few periods

Many periods

Static

Only relative differences matter

Stacked 100% bar chart



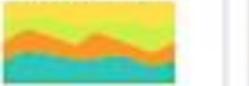
Relative and absolute differences matter

Stacked bar chart



Only relative differences matter

Stacked 100% area chart



Relative and absolute differences matter

Stacked area chart



Simple share of total

Pie chart



Accumulation or subtraction to total

Waterfall chart



Components of components

Stacked 100% bar chart w/subcomponents



Accumulation to total & absolute difference matters

Tree map



- [From data to Viz | Find the graphic you need \(data-to-viz.com\)](http://data-to-viz.com)
- [List View: The Data Visualisation Catalogue \(datavizcatalogue.com\)](http://datavizcatalogue.com)
- [charts and graphs - a complete guide — storytelling with data](#)
- [Datawrapper Blog](#)

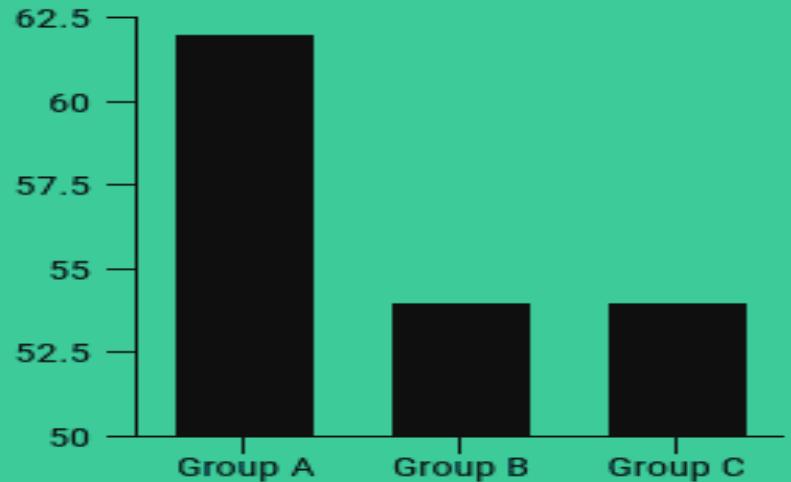
Avoid Distorting The Data



1

OMITTING THE BASELINE

In most cases, the baseline for a graph is 0. But writers can skew how data is perceived by making the baseline a different number. This is known as a “truncated graph”.



MISLEADING

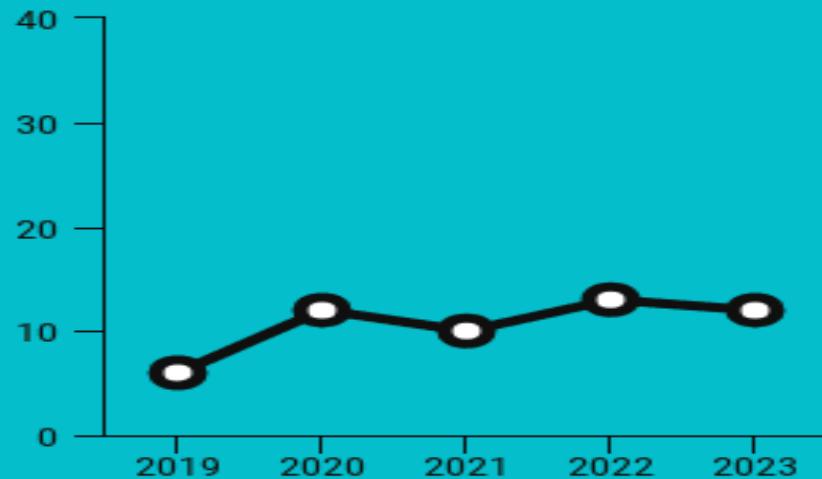
- Starting the vertical axis at 50 makes a small difference between groups seem massive
- Group A looks much larger than Groups B and C

VS

2

MANIPULATING THE Y-AXIS

Expanding or compressing the scale on a graph can make changes in data seem more or less significant than they actually are.



MISLEADING

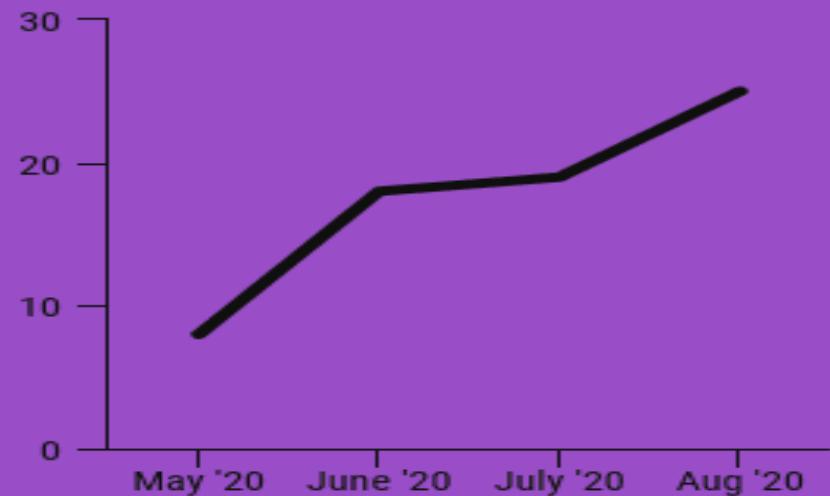
- The scale is disproportionate to the data, making the change over time seem small

VS

3

CHERRY PICKING DATA

Writers may only include certain data points on their graphs to reinforce their narratives. This can create a false impression of the data.



MISLEADING

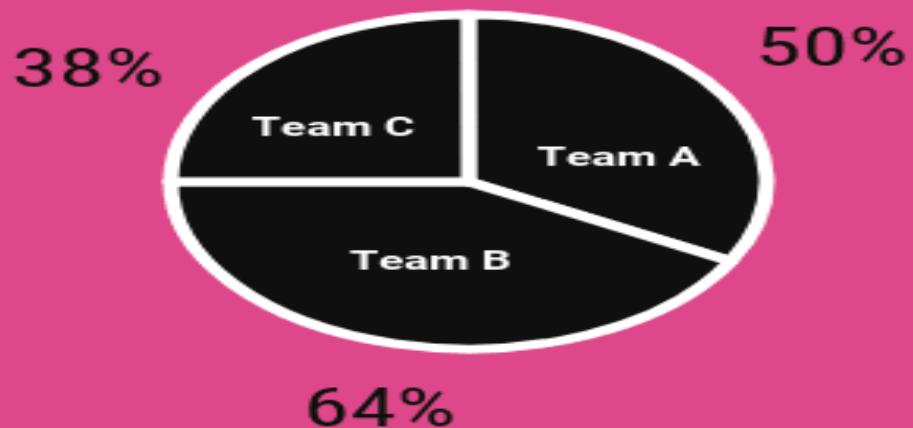
- Only a few months out of the year are graphed, depicting an upward trend

VS

4

USING THE WRONG GRAPH

The type of graph you use should depend on the type of data you want to visualize.
Using the wrong type of graph can skew the data. Writers will sometimes use
the wrong type of graph on purpose.



MISLEADING

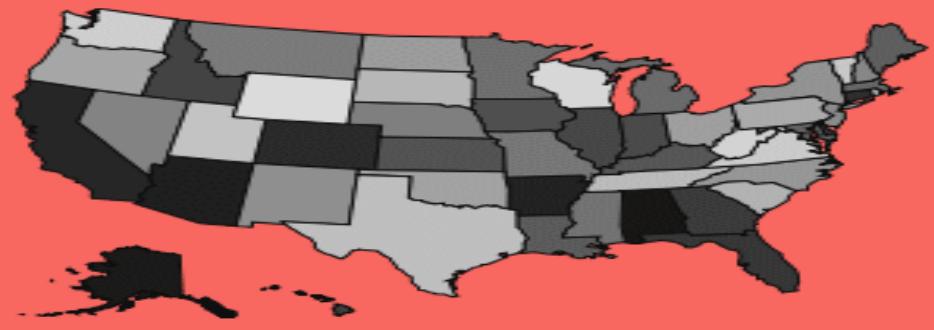
- Pie charts are used to compare parts of a whole, not the difference between groups
- A different type of graph should be used to compare the three teams

VS

5

GOING AGAINST CONVENTIONS

Over time, we have developed standards for how data is visualized. Flipping those conventions can make a graph confusing or misleading to readers.



Individuals per km

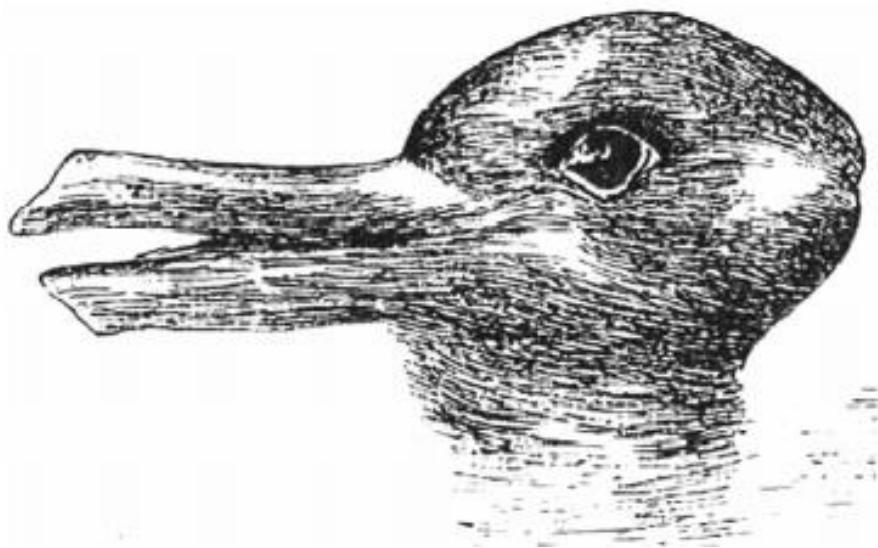
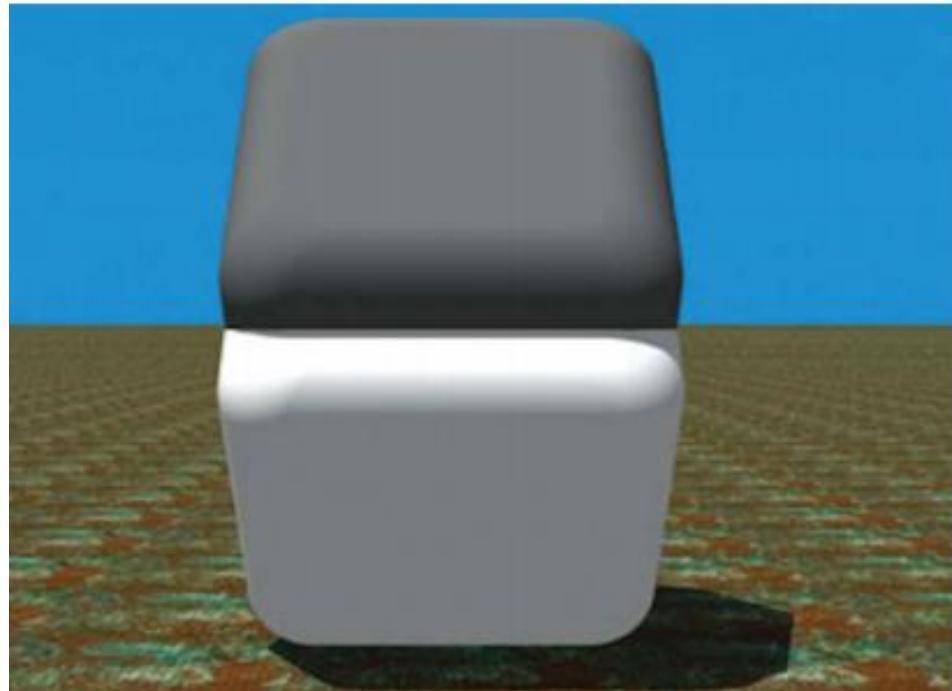


MISLEADING

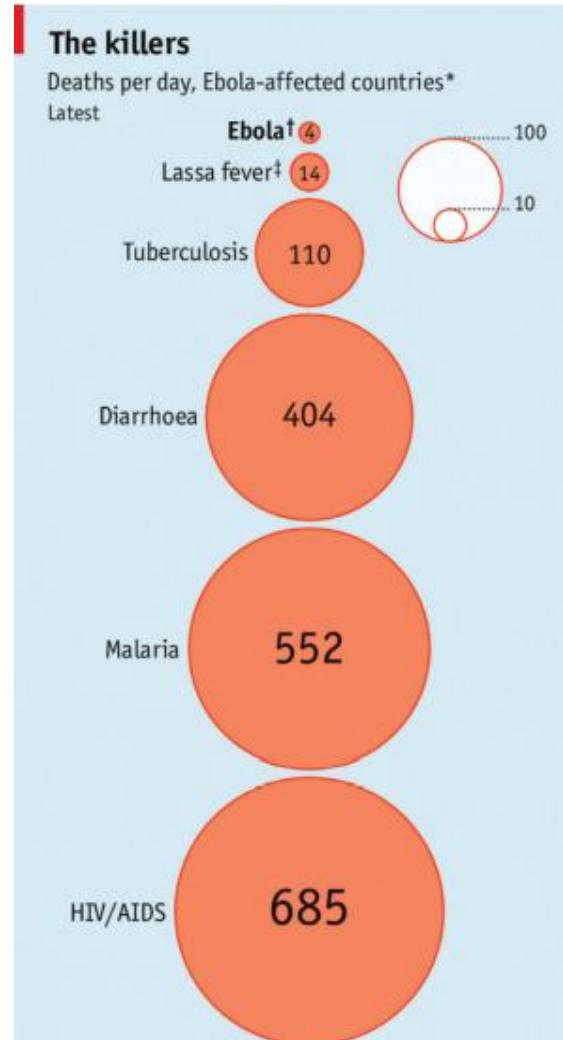
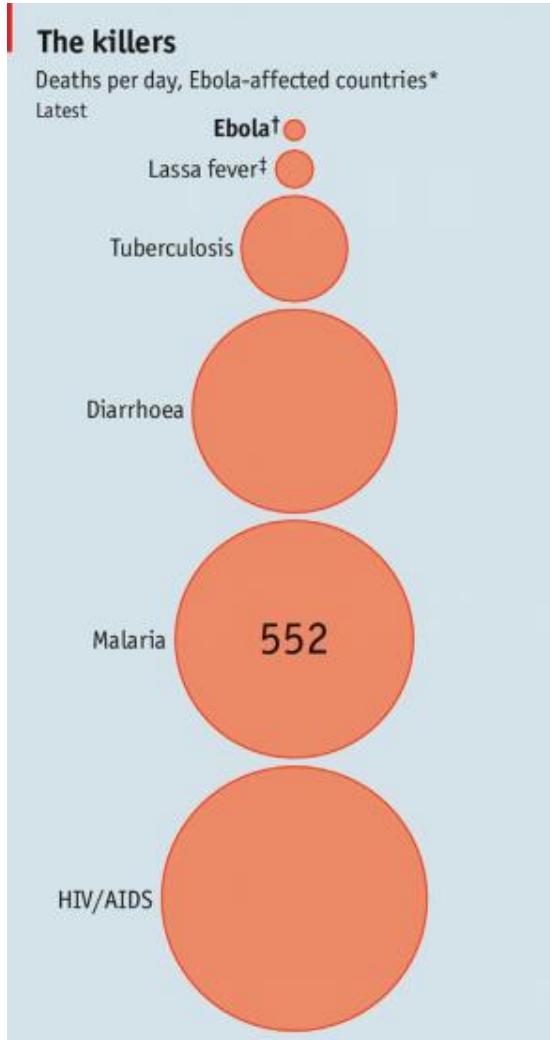
VS

- Normally, darker shades are associated with density on a map but here, dark has been used to depict lower population density
- This graph can confuse and mislead readers, who expect dark to represent a higher population density

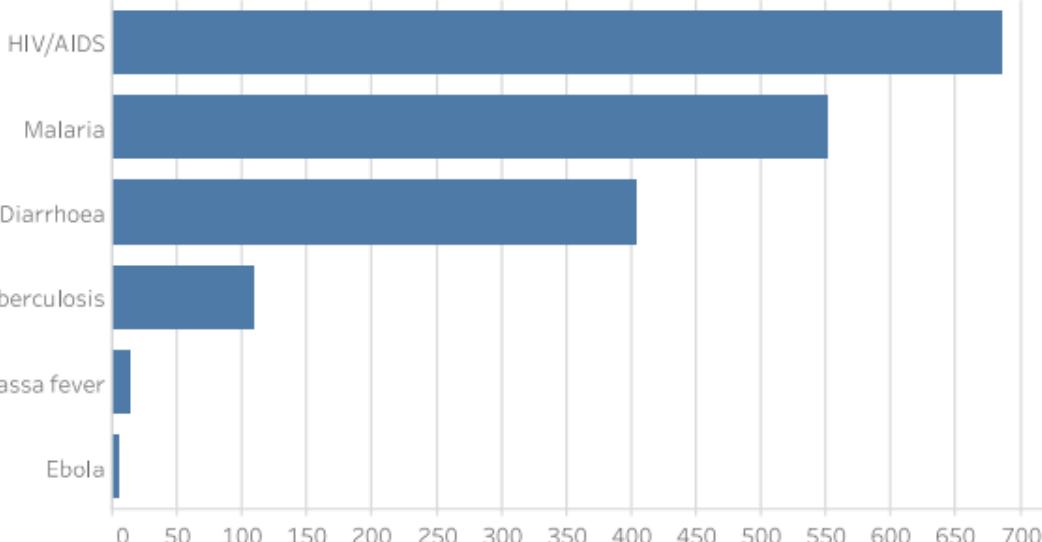
When our visual Processing System Betrays Us?



When our visual Processing System Betrays Us?



Deaths per day, Ebola-affected countries

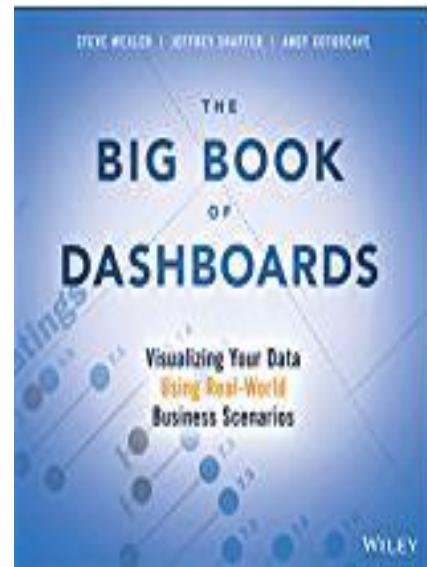


*Guinea, Liberia, Nigeria and Sierra Leone
Sources: World Health Organization; U.S. Centers for Disease Control and Prevention; *The Economist*

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.

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



Are All Dashboards the Same?

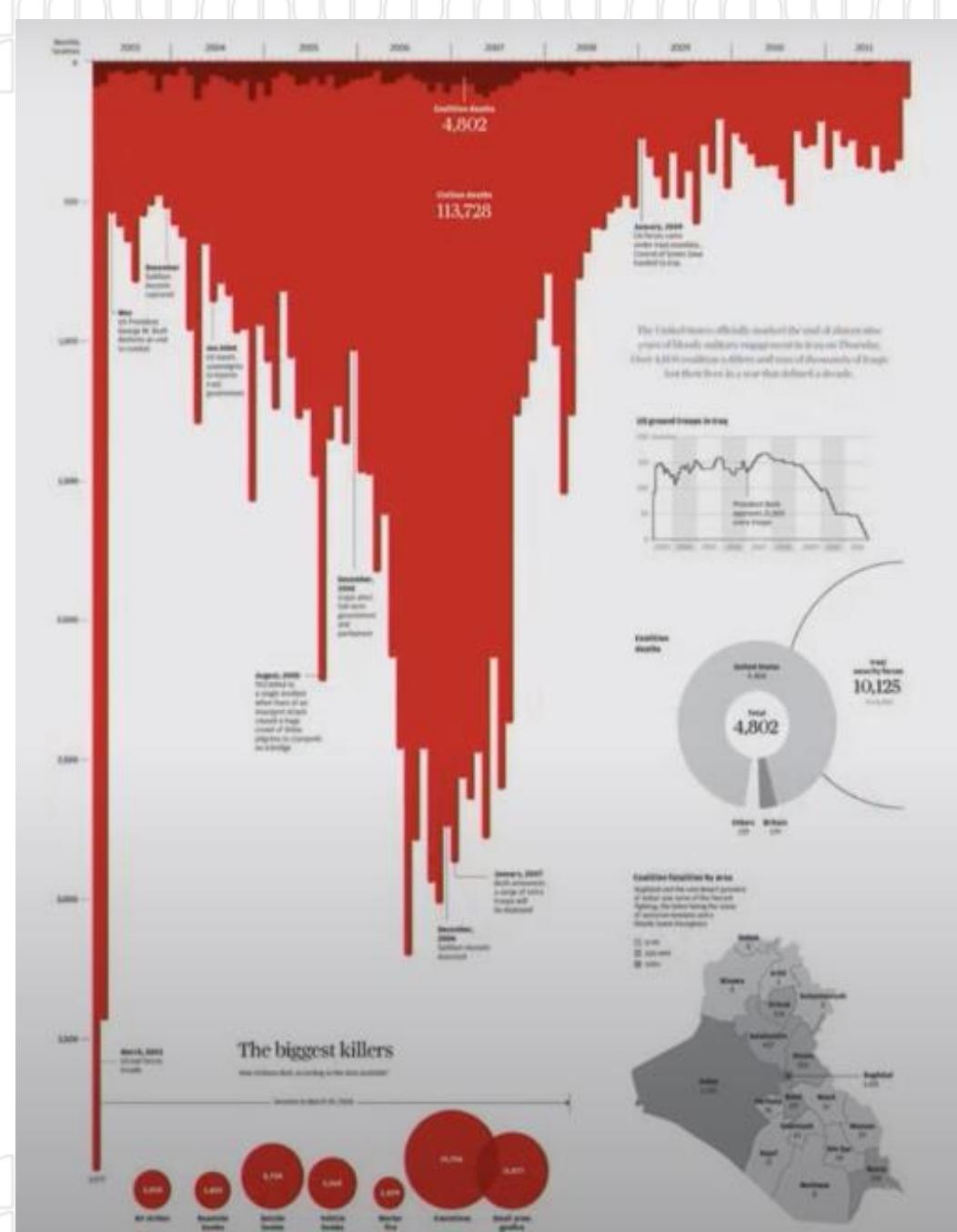


Are All Dashboards the Same?

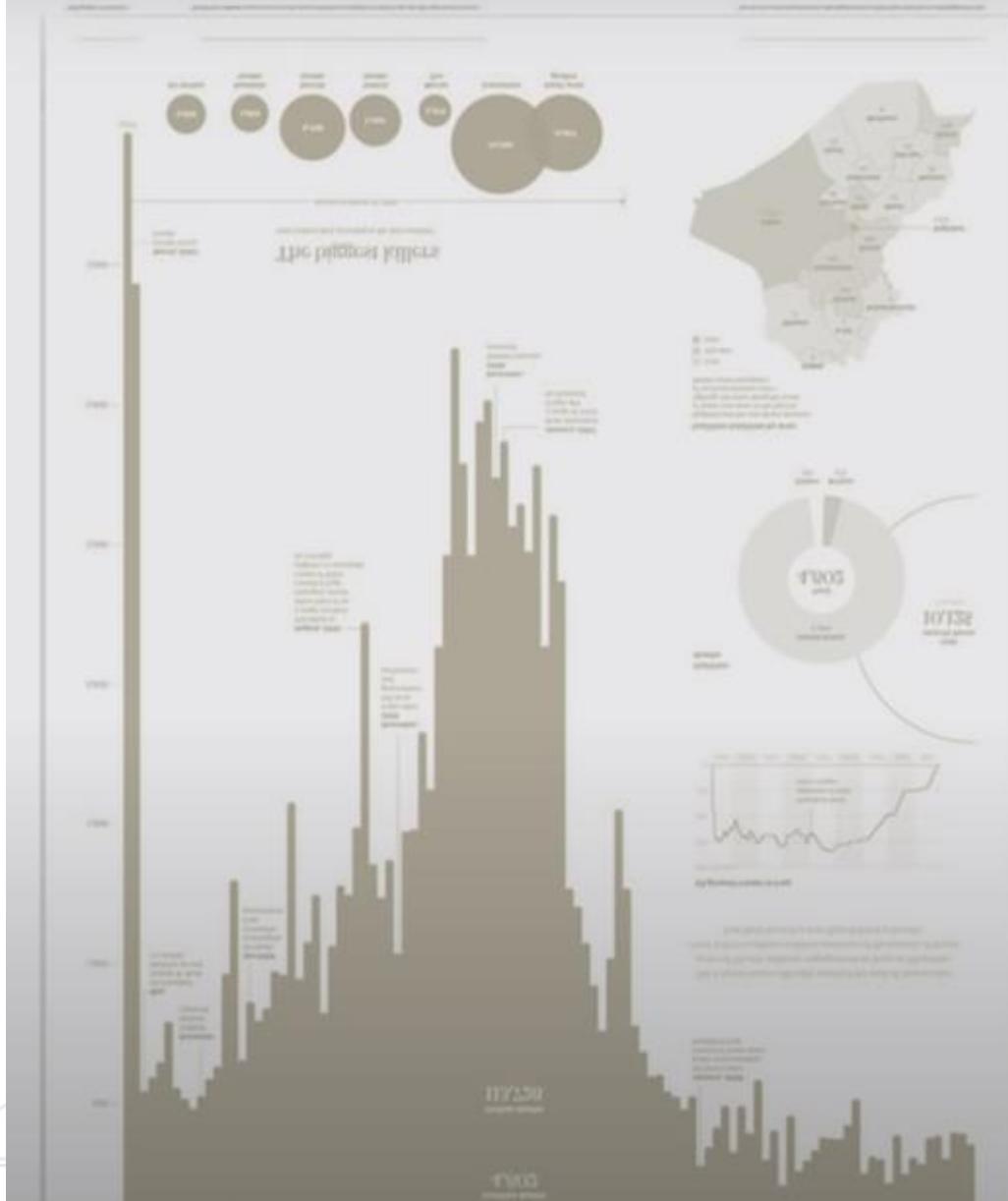


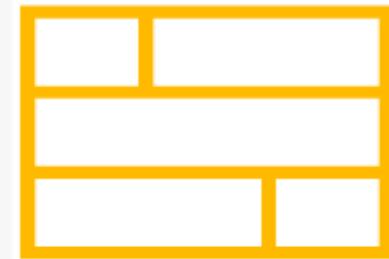
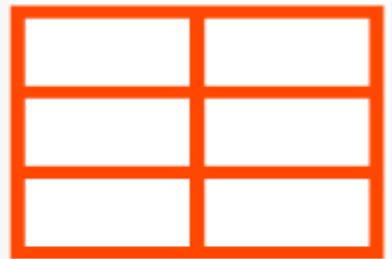
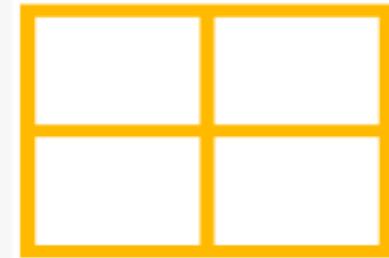
Are All Dashboards the Same?





Iraq: Deaths on the decline





Design To a GRID

162.9k

Last 7 Days Website Visits

10% Increase from Last Week

4.3k

Demos Scheduled

12% Increase from Last Week

2.1k

New Signups

3% Decrease from Last Week

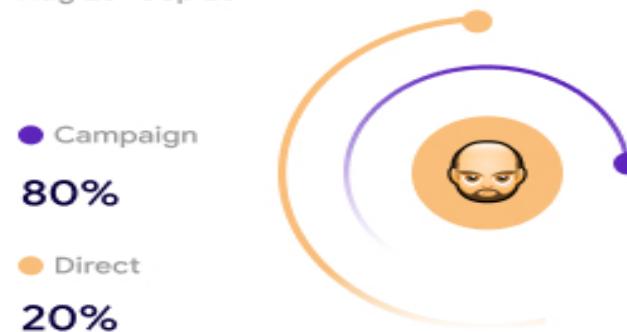
541

Tickets Submitted

16% Decrease from Last Week

Visitors Overview

Aug 25 - Sep 25 ▾

**92**

Conversation Today

+12%

15,980

Page Views

4,324

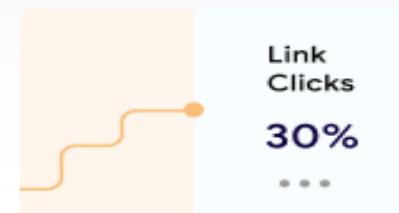
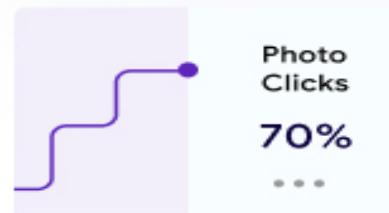
Contacted

114

Advertisement

98**4** Campaigns**+10%****12,499**

Emails Sent



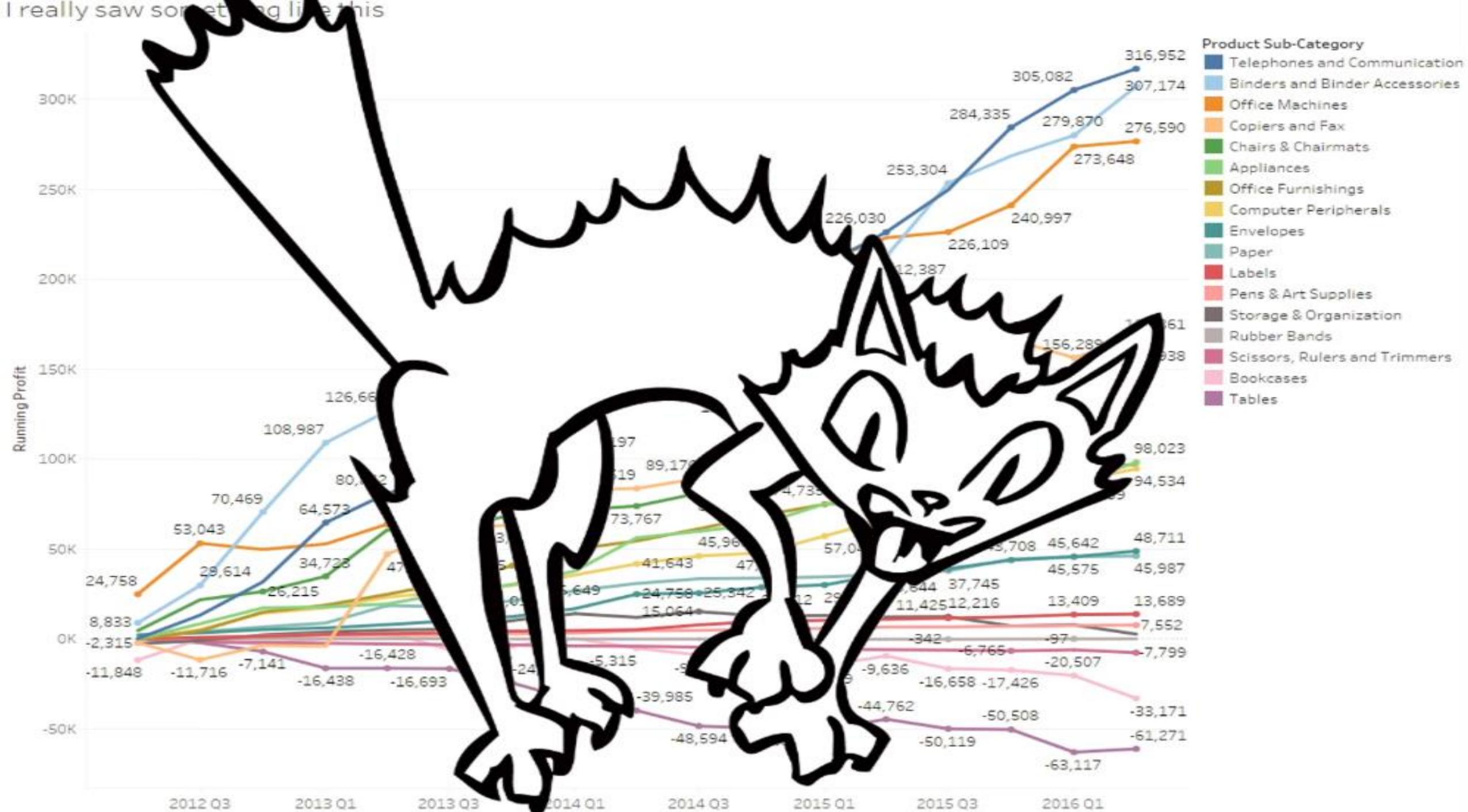
Leads Summary

LinkedIn
1,224Email
2,391Direct
2,418

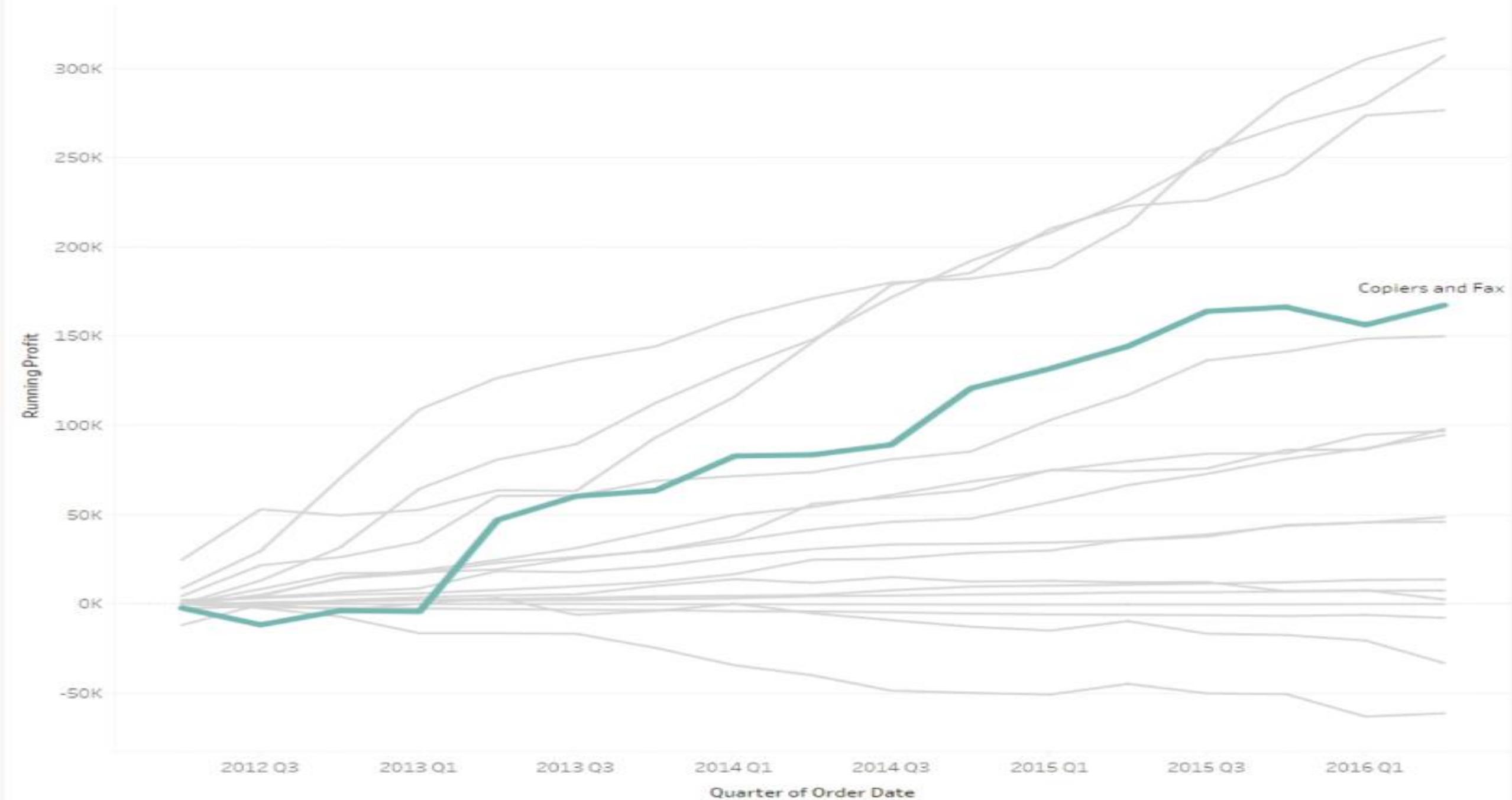
A photograph of a desk covered in a chaotic pile of papers, files, and documents, illustrating the concept of clutter.

Avoid Clutter

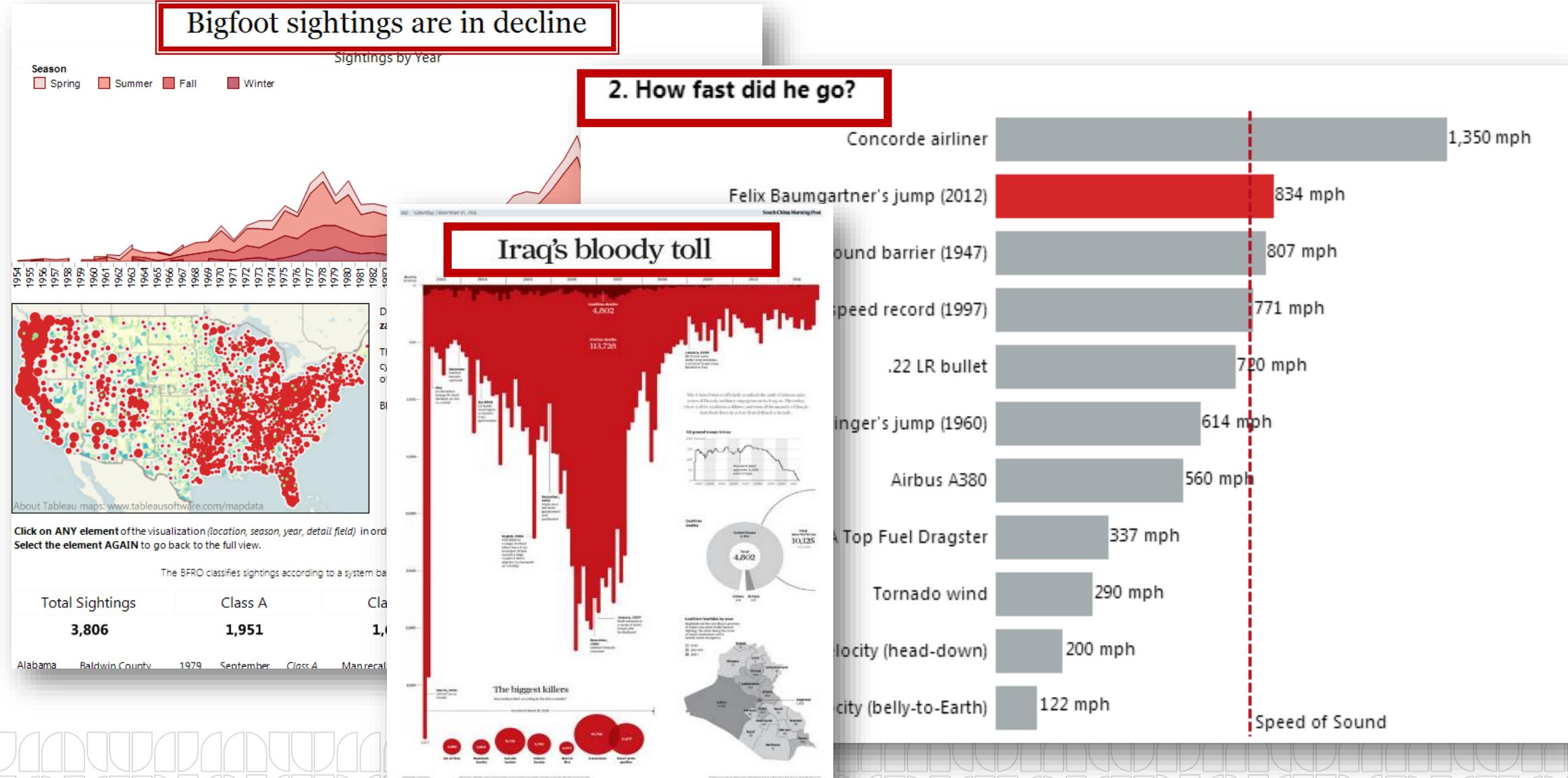
I really saw something like this



Copiers and Fax compared to other sub-categories over time



Do your titles capture attention?

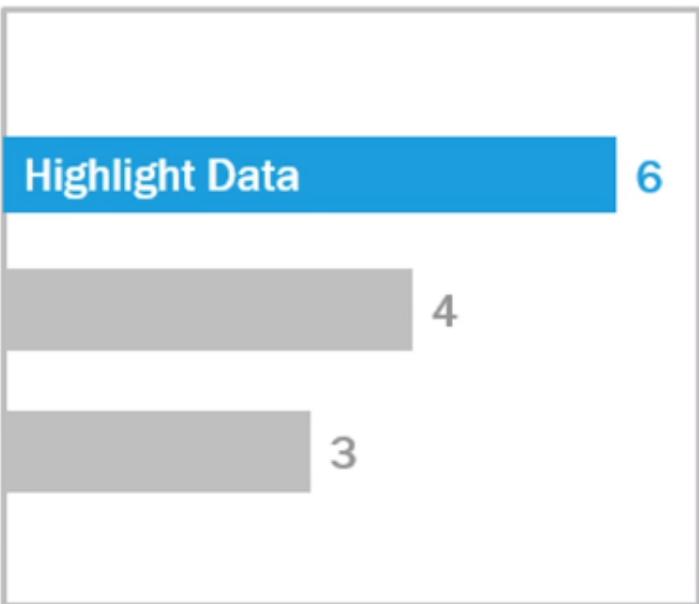




Font Contrast and BANs: Who gives a font?

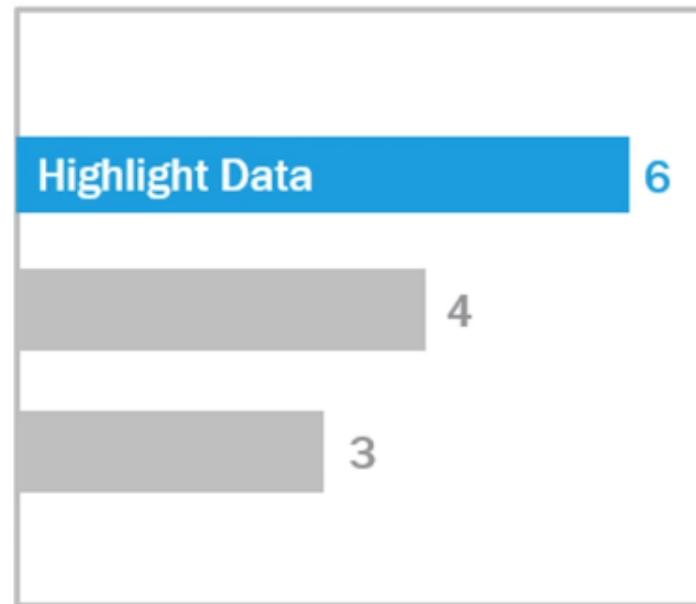
Top Level Font

Mid Level Font



Low Level Font

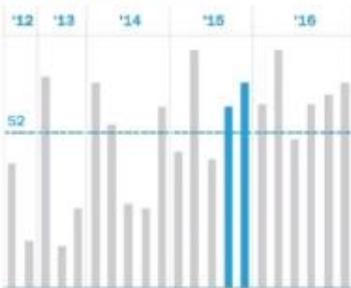
Mid Level Font



Low Level Font

Course Metrics

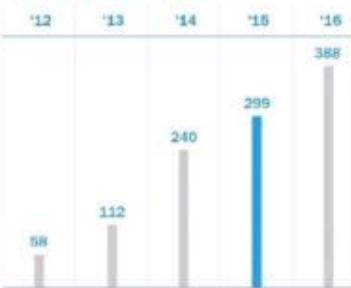
Students



1097

Total students in five years

Enrollments



687

Total students in 2015-2016

Classes



21

Total classes in five years

Ratings



7.7 of 8

Most Recent Instructor Rating (out of 8.0)

Semesters

2015 Fall Semester 001

Questions

I developed specific skills and competencies

Overall, this was an excellent course

The instructor communicated clearly

The instructor graded fairly

The instructor was well organized

The instructor interacted well with students

Overall, this instructor was excellent

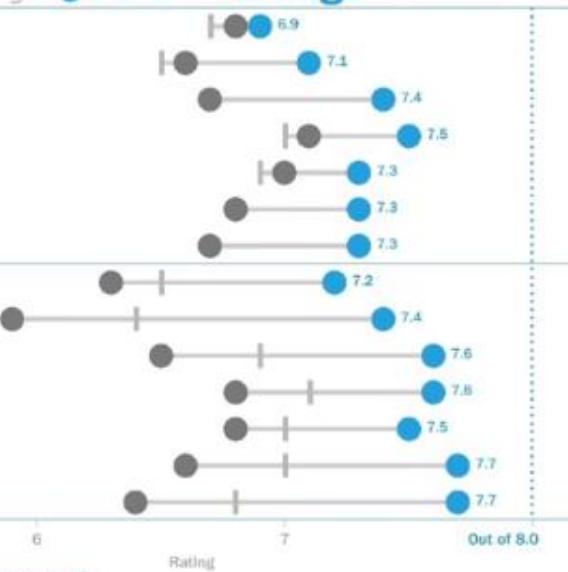
2015 Fall Semester 002

BANA

|College

Shaffer

Ratings



Course Metrics Dashboard created by Jeffrey A. Shaffer. Data from University of Cincinnati Course Evaluations. Blue indicates the 2 most recent rating periods.

Source: *The Big Book of Dashboards* (BigBookofDashboards.com)

Agency Utilization Rollup

BANs

\$3.8M
Fees

• \$3.4M
Potential

\$1.3M
New Biz + Opportunity

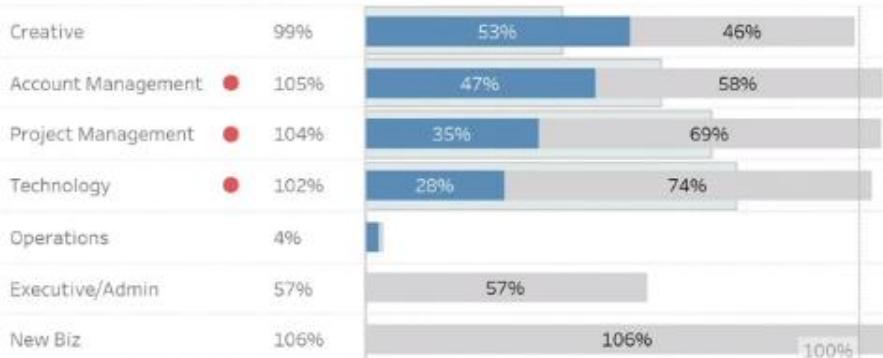
\$2.6M
Internal Projects

• +12.2
FTE Overstaffed

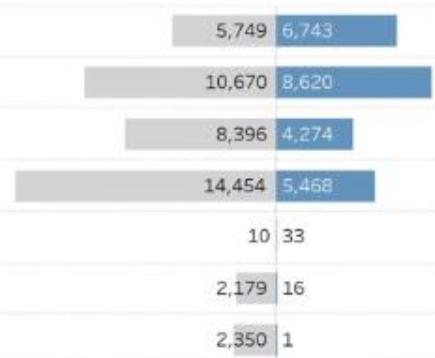
BANs



Target vs Billable vs Non-Billable %



Non-Billable vs Billable Hours



Cost | Fees | Potential



Show Potential at 100% of Target ▾

New Biz + Opp

Category	Hours	%	Hours	%	Hours	%
Creative	1,456 hrs	12%	1,301 hrs	10%	2,992 hrs	24%
Account Management	4,524 hrs	25%	1,579 hrs	9%	4,567 hrs	25%
Project Management	1,455 hrs	12%	2,283 hrs	19%	4,659 hrs	38%
Technology	361 hrs	2%	9,608 hrs	49%	4,485 hrs	23%
Operations	6 hrs	0%	0 hrs	0%	5 hrs	0%
Executive/Admin	0 hrs	0%	4 hrs	0%	2,176 hrs	56%
New Biz	1,764 hrs	80%	40 hrs	2%	546 hrs	25%

Cost \$533K

Cost \$755K

Cost \$1,163K

Internal Projects

Category	Hours	%
Creative	1,301 hrs	10%
Account Management	1,579 hrs	9%
Project Management	2,283 hrs	19%
Technology	9,608 hrs	49%
Operations	0 hrs	0%
Executive/Admin	4 hrs	0%
New Biz	40 hrs	2%

Internal Admin



Utilization Trend



Source: The Big Book of Dashboards (BigBookofDashboards.com)

Agency Utilization Rollup

\$3.8M
Fees

• \$3.4M
Potential

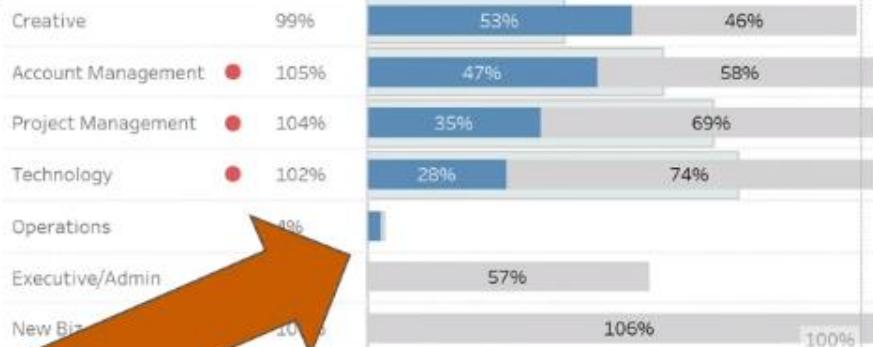
\$1.3M
New Biz + Opportunity

\$2.6M
Internal Projects

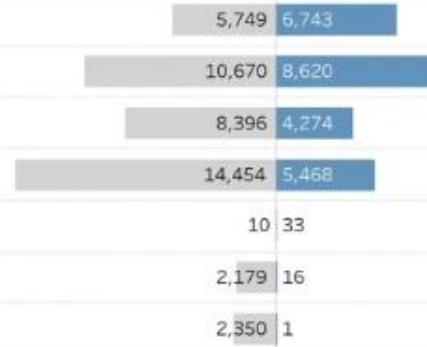
• +12.2
FTE Overstaffed

Headlines

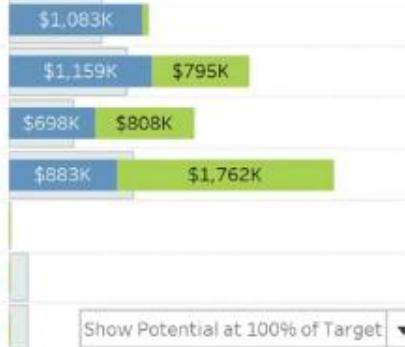
Target vs Billable vs Non-Billable %



Non-Billable vs Billable Hours



Cost | Fees | Potential



Show Potential at 100% of Target ▾

Details

New Biz + Opp

Role	Hours	%	Hours	%	Hours	%
Creative	1,456 hrs	12%	1,301 hrs	10%	2,992 hrs	24%
Account Management	4,524 hrs	25%	1,579 hrs	9%	4,567 hrs	25%
Project Management	1,455 hrs	12%	2,283 hrs	19%	4,659 hrs	38%
Technology	361 hrs	2%	9,608 hrs	49%	4,485 hrs	23%
Operations	6 hrs	0%	0 hrs	0%	5 hrs	0%
Executive/Admin	0 hrs	0%	4 hrs	0%	2,176 hrs	56%
New Biz	1,764 hrs	80%	40 hrs	2%	546 hrs	25%

Internal Projects

Internal Admin



Use as Color Legend

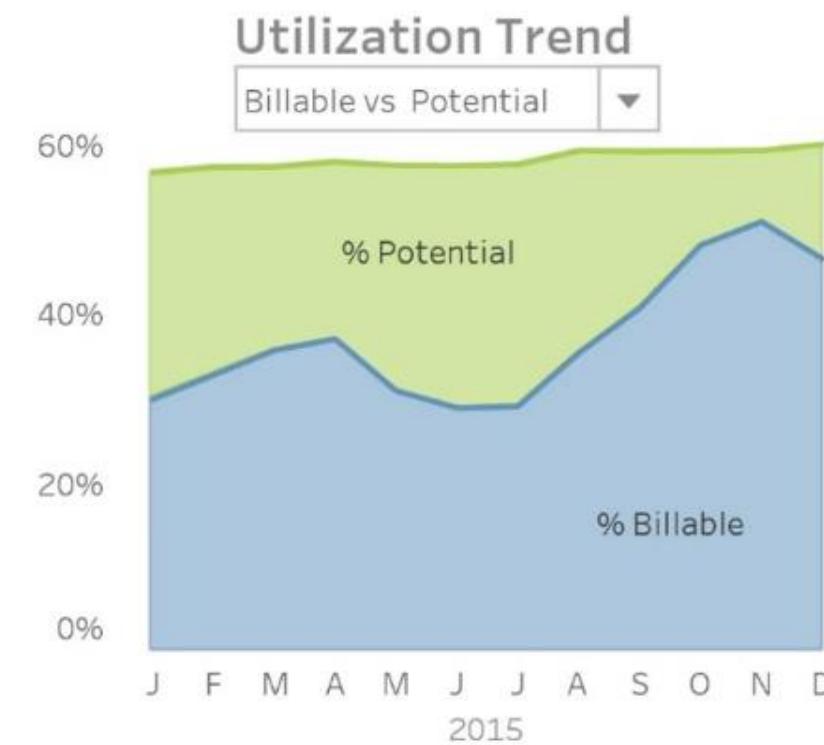
\$3.8M
Fees

• \$3.4M
Potential

\$1.3M
New Biz + Opportunity

\$2.6M
Internal Projects

• +12.2
FTE Overstaffed



Study 1: BANs at TOP

Agency Utilization Rollup

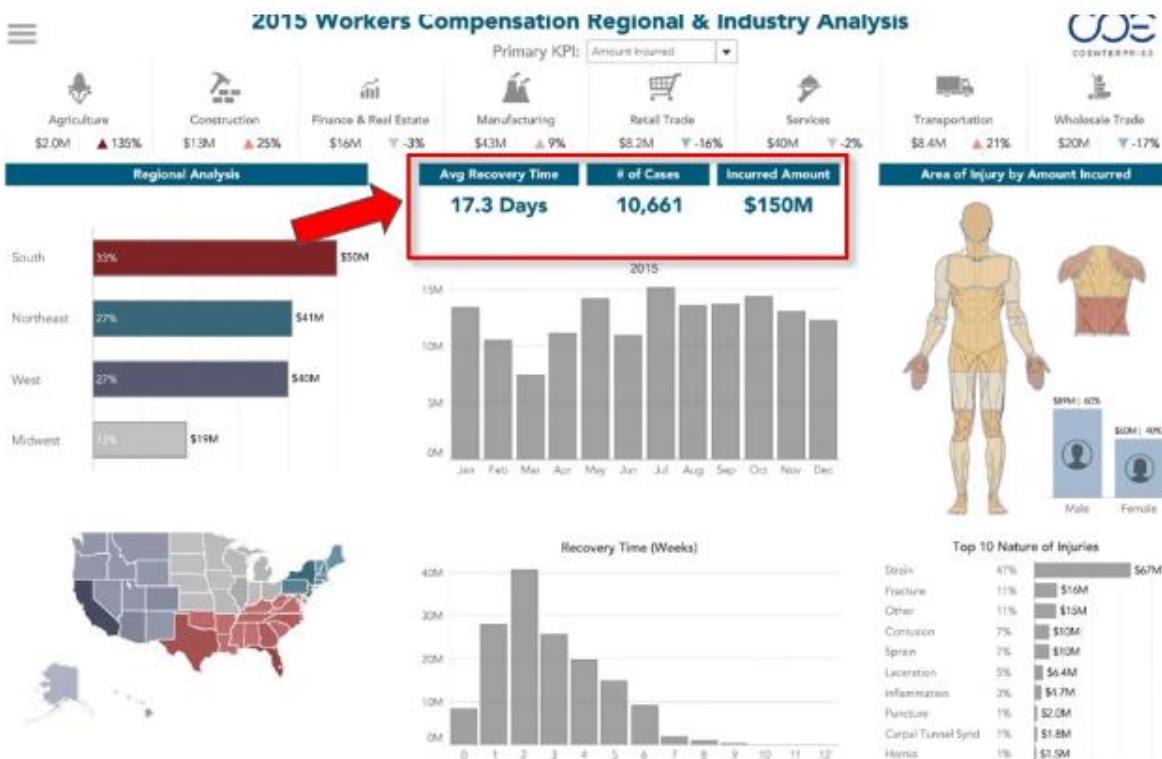


Study 2: BANs at BOTTOM

Agency Utilization Rollup



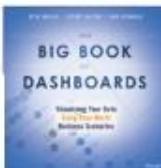
Study 1: Small BANs



Study 2: Big BANs



Dashboards from *Big Book of Dashboards*

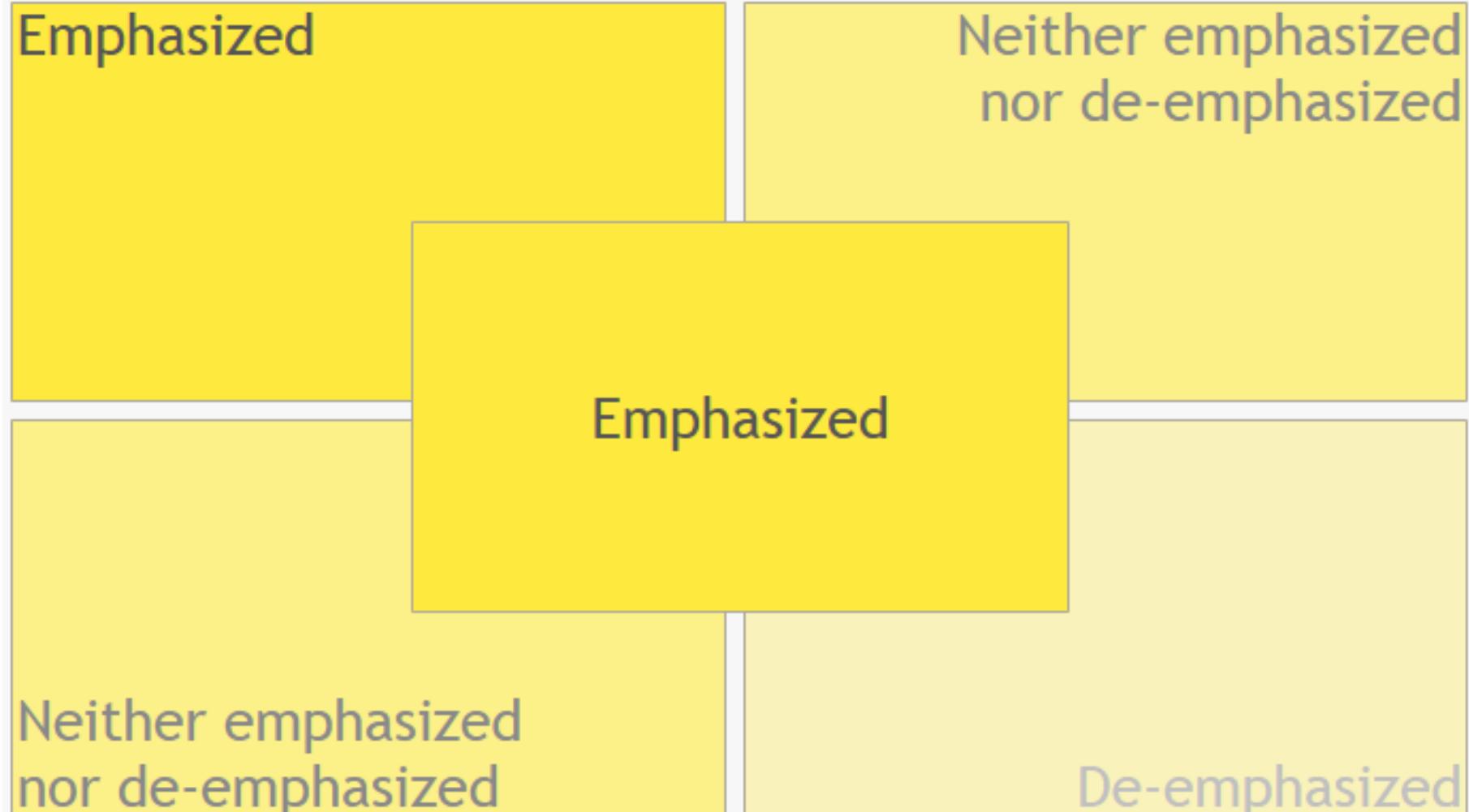


What Makes a Good Dashboard?

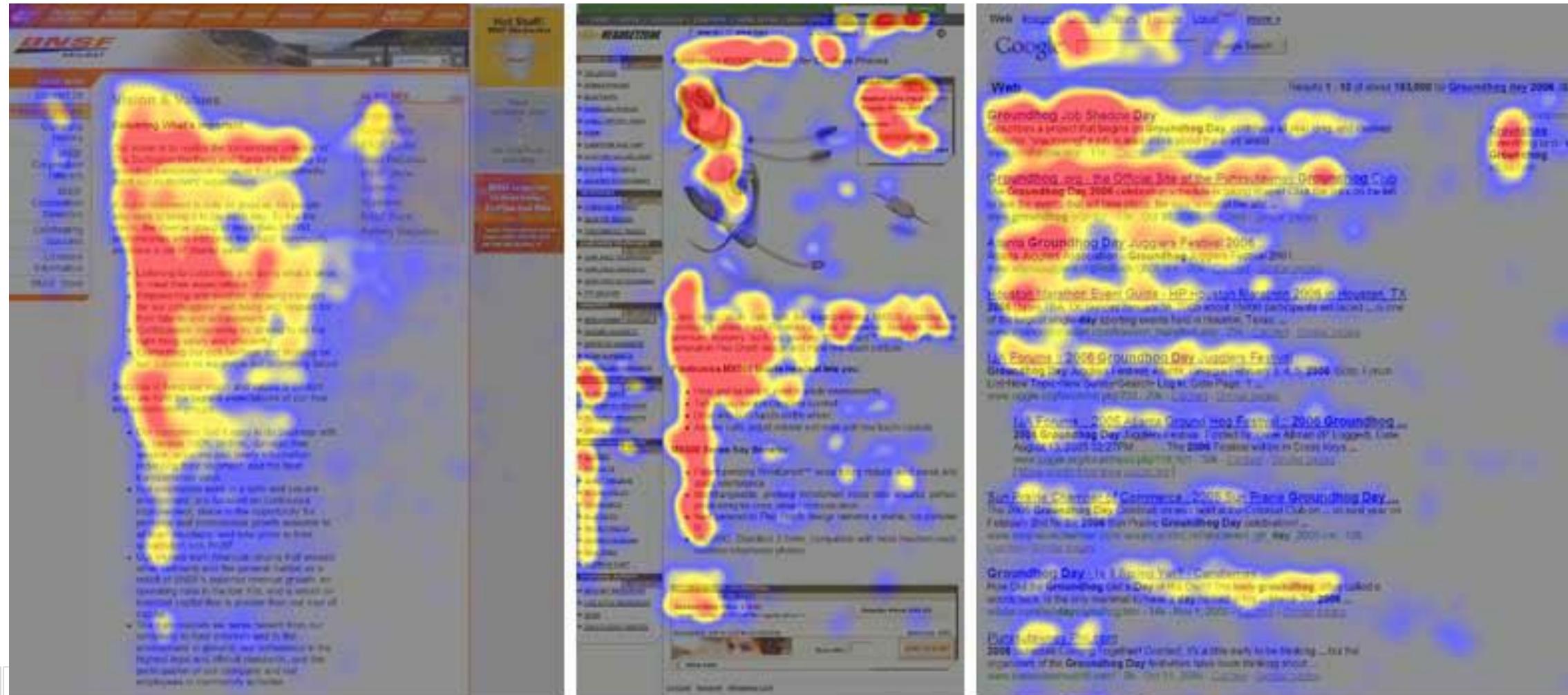
- Answers a set of questions
- Follows a flow and invites interactivity
- Specific to and customized for the dashboard's audience and objectives
- Provides appropriate text for clarity and direction, if needed
- Makes strategic use of color



What Goes Where?



What Goes Where?



Layout



Filters at the top, so user knows when and how to start interacting.



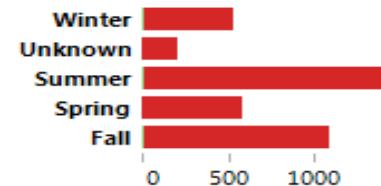
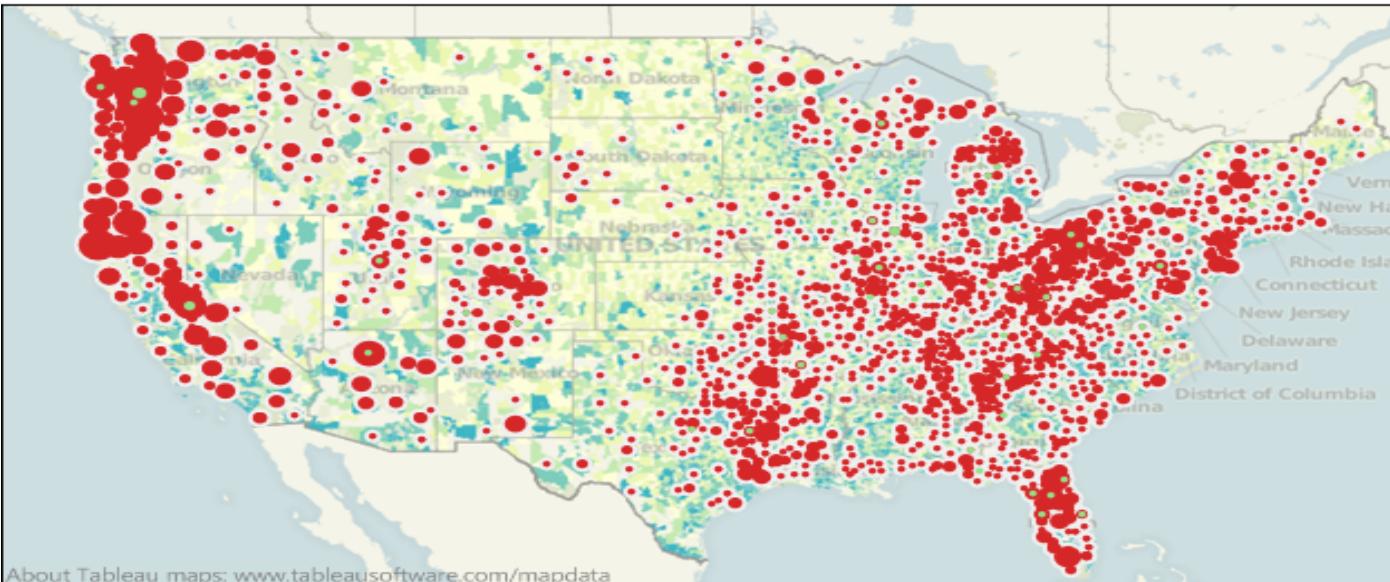
Most important information at the top and centered.



Starts with the highest level and works down to more precise details.



Where is bigfoot seen in the US?



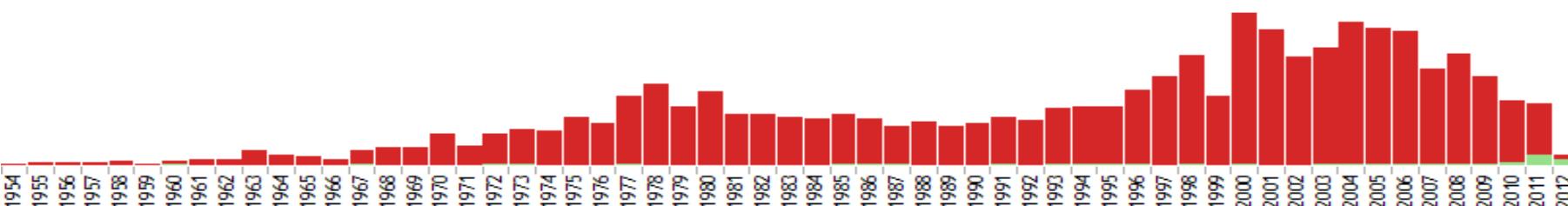
Data gathered from the official website of the "Bigfoot Field Researchers Organization" (BFRO).

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.

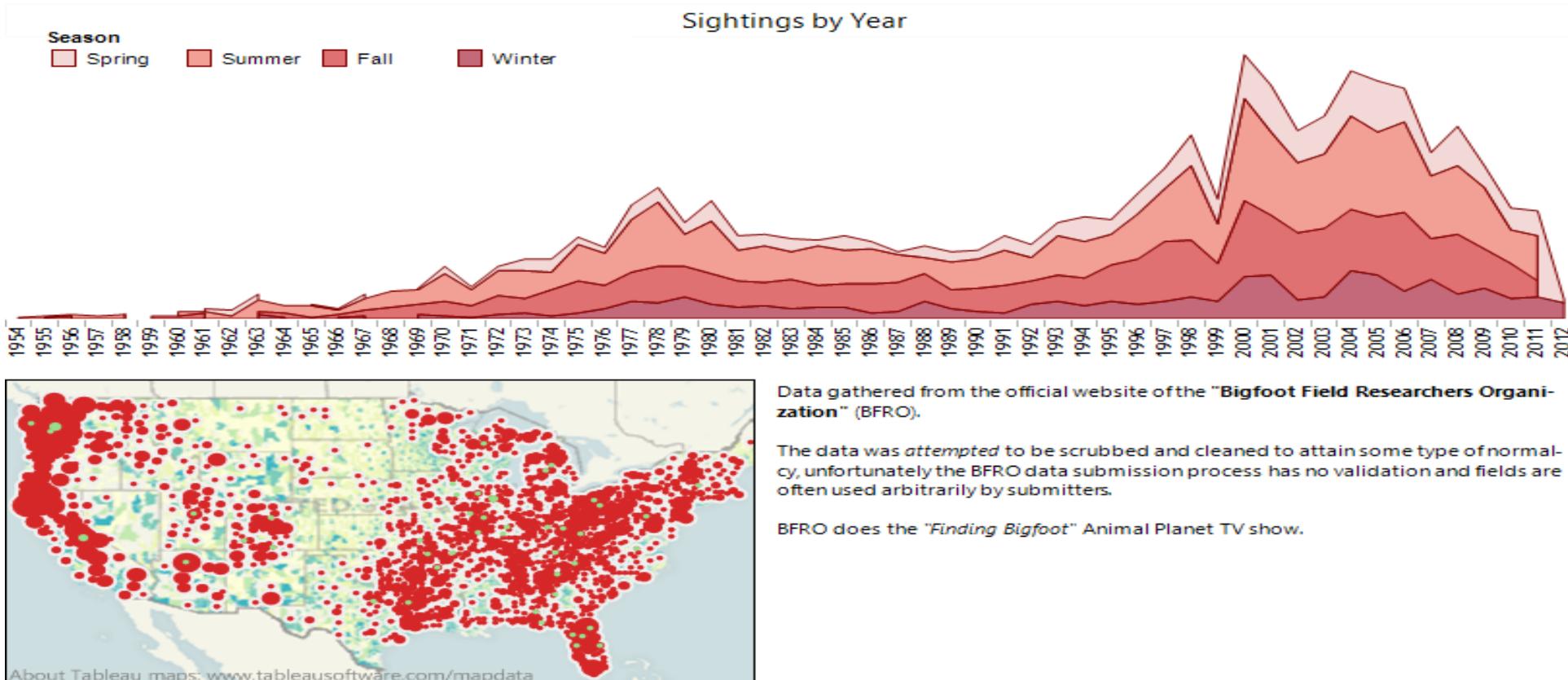
Select the element AGAIN to go back to the full view.



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

Bigfoot sightings are in decline



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

DONT BE AFRAID...
WHITE SPACE

Complaints Dashboard

*NOTE - Timeframe is set at 2014 to present due to changes in recording method.

uniFund
SOURCE TYPE
UniFund

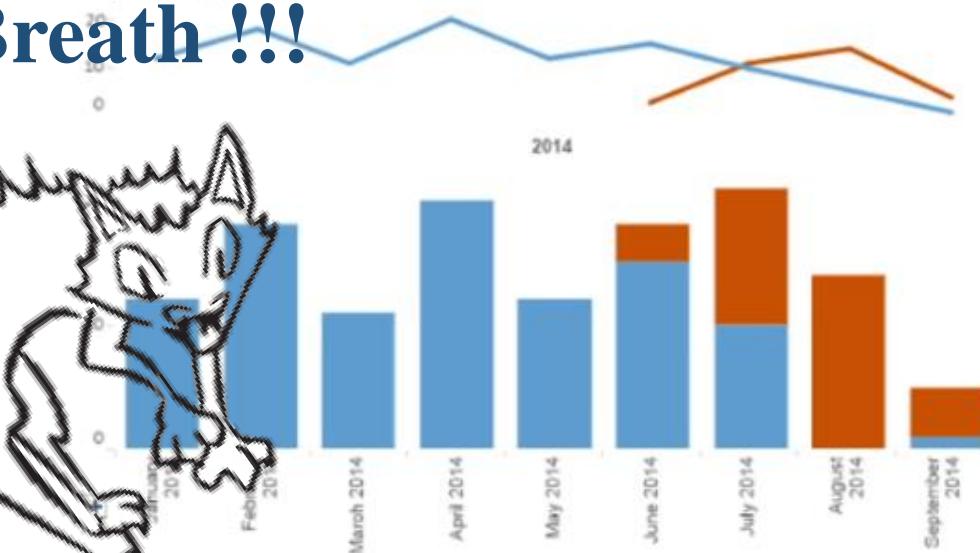
RECEIVED Date
This year

Subject Description
(All)

Color Code by Open/Closed Status
Yes

Show Open/Closed
All

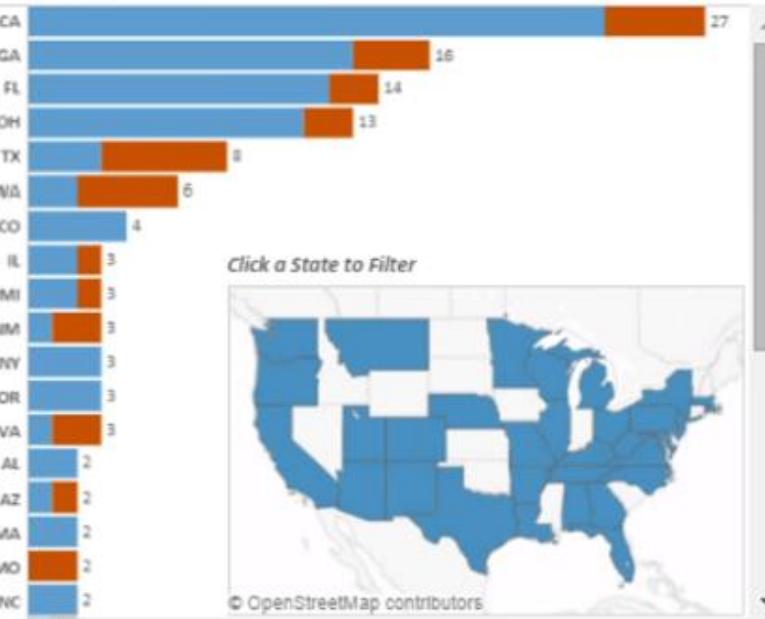
Complaints by Month



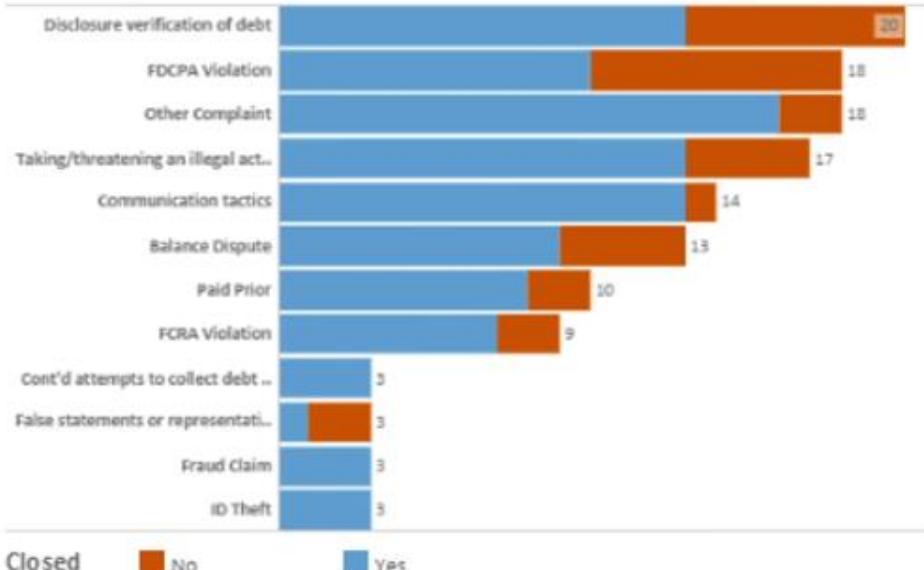
Let it Breath !!!

Total Complaints
131

Complaints by State



Complaints by Reason



Complaints by Party



Before

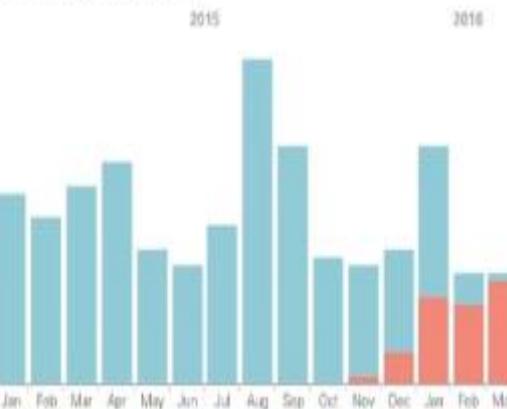


After

Complaints Dashboard

Total Complaints: **288** **Open** **39** **Total** **327**

Complaints by Month



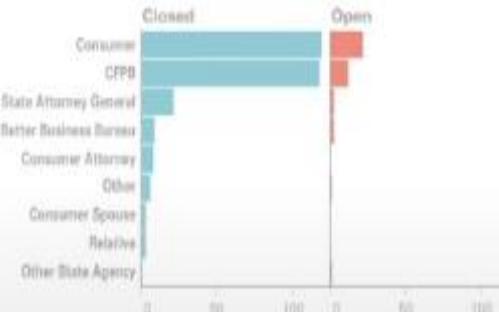
Open Complaints by State (click to filter)



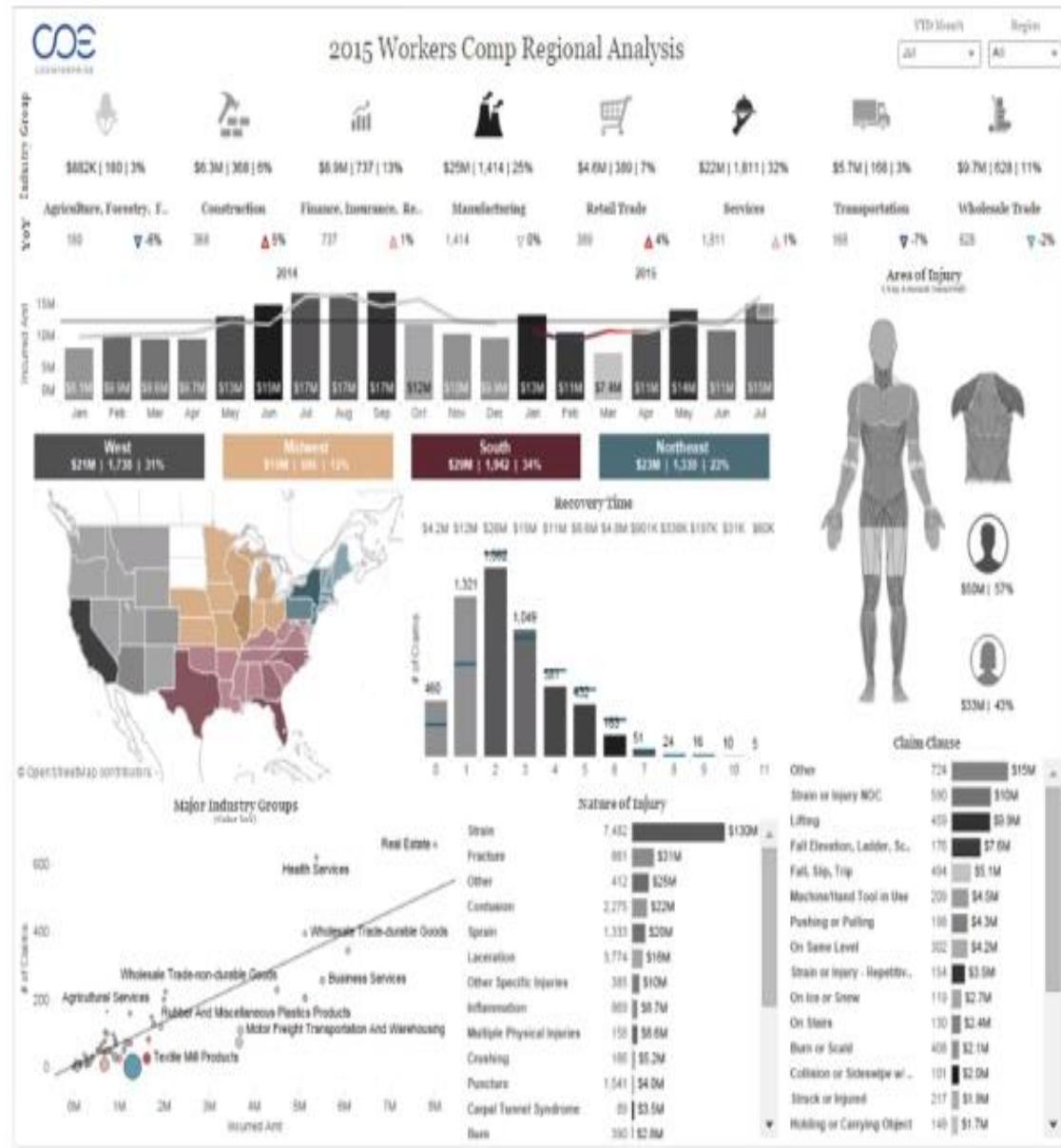
Complaints by Reason



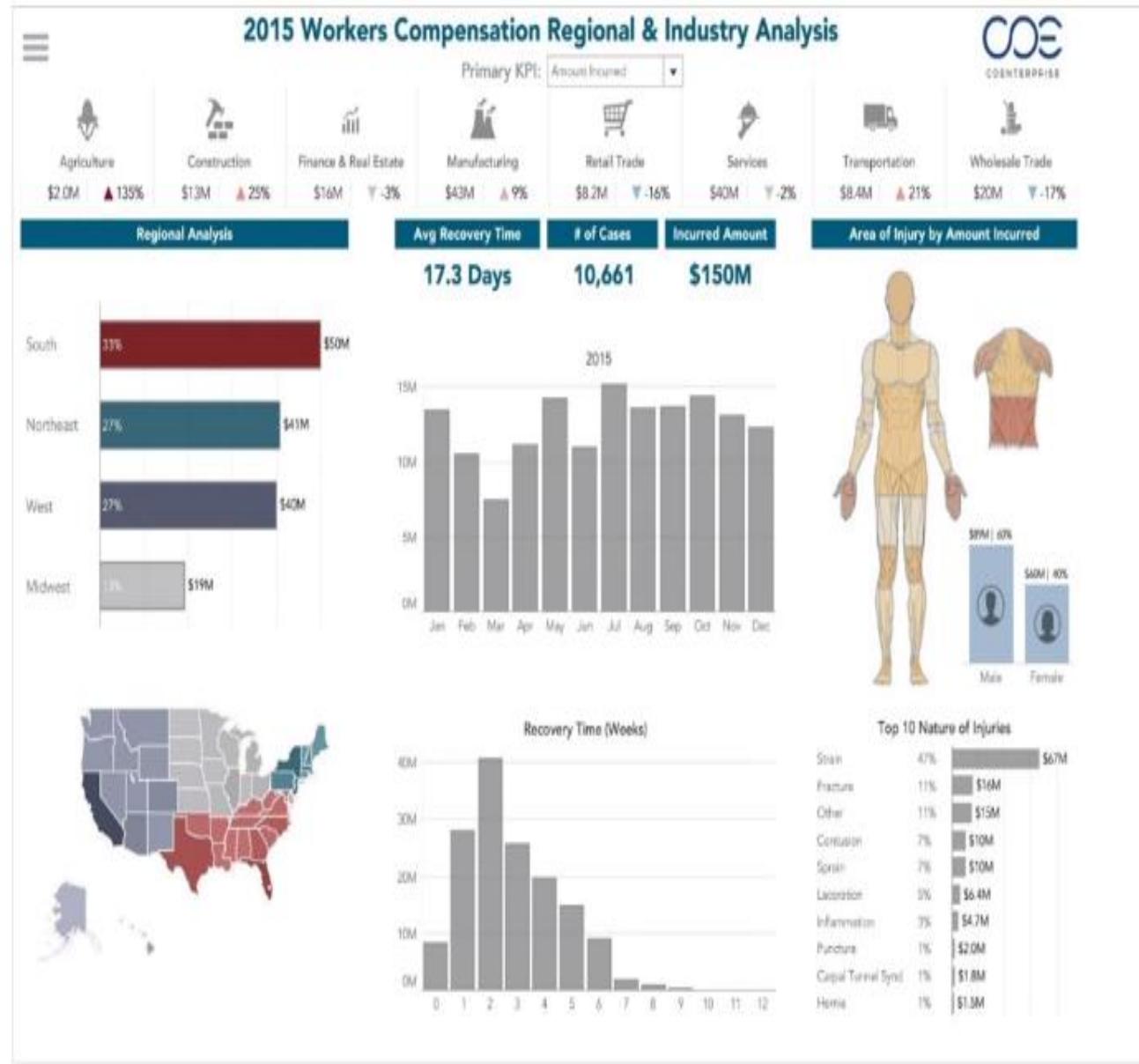
Complaints by Party (click to filter)



Before



After



Does your 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)
- ⊕ Provide interactivity
- ⊕ Titles
- ⊕ Axes& Units
- ⊕ Remove extra digits in numbers
- ⊕ Use familiar chart types.
- ⊕ Don't make audience remember views.
- ⊕ Avoid large legends.
- ⊕ Use intuitive colors and shapes



Collaboration

Our Common Goal

For the largest number of people, provide the greatest degree of understanding, with the least amount of effort.

Find Collaborators...

... and Iterate, Iterate, Iterate!

Scenario

Date	Division	Gained	Lost
1/1/2015	A	70	0
2/1/2015	A	80	-90
3/1/2015	A	100	-30
4/1/2015	A	110	-45
5/1/2015	A	70	-95
6/1/2015	A	45	-33
7/1/2015	A	50	-110
8/1/2015	A	99	-34
9/1/2015	A	112	-34
10/1/2015	A	99	-88
11/1/2015	A	55	-65
12/1/2015	A	110	-45
1/1/2015	B	80	0
2/1/2015	B	80	-15
3/1/2015	B	90	-30
4/1/2015	B	120	-25
5/1/2015	B	100	-50
6/1/2015	B	119	-77
7/1/2015	B	75	-45
8/1/2015	B	119	-77
9/1/2015	B	90	-30



Subscriber Churn Analysis

Subscriber activity - All

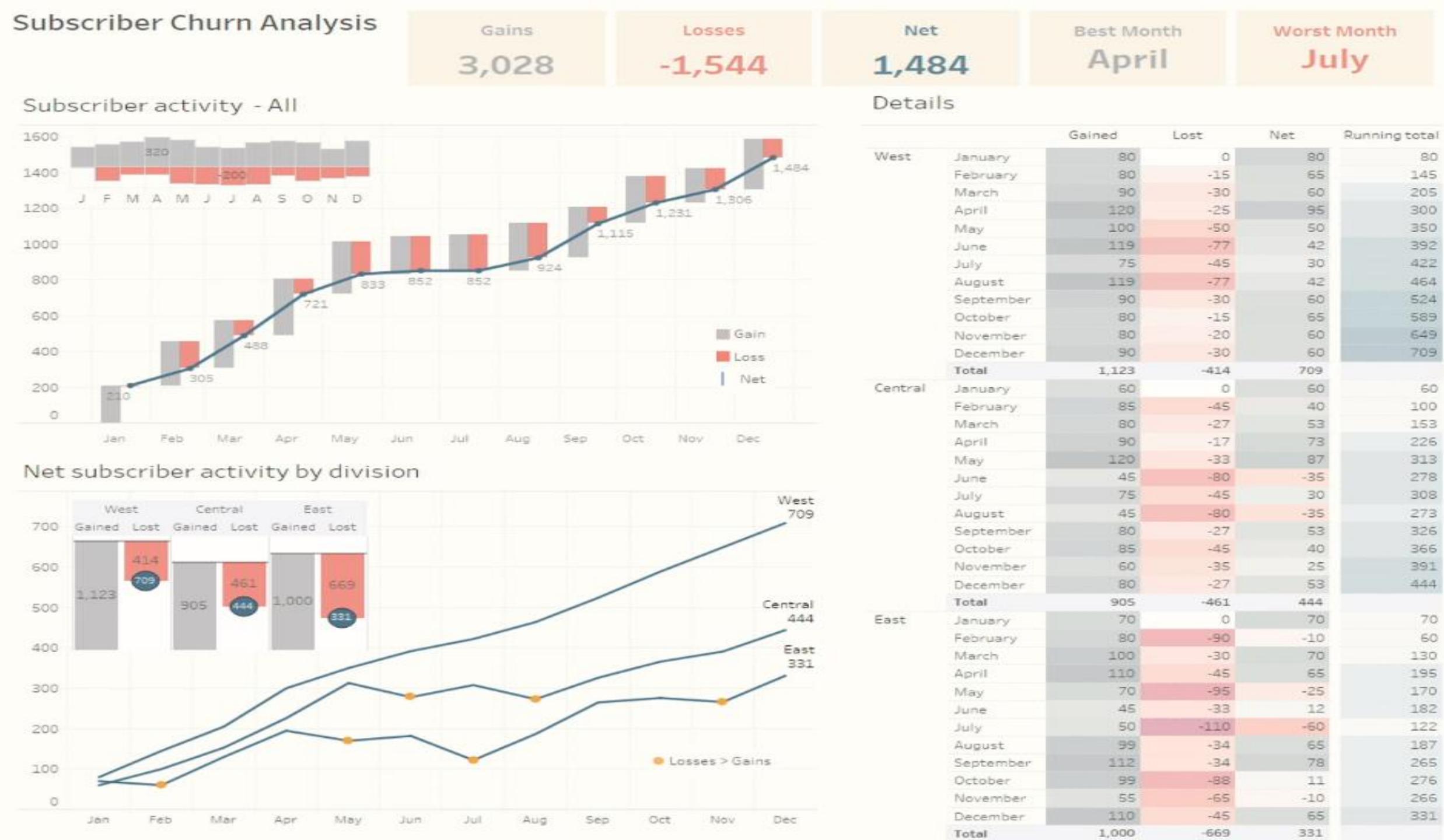


Net subscriber activity by division



Details

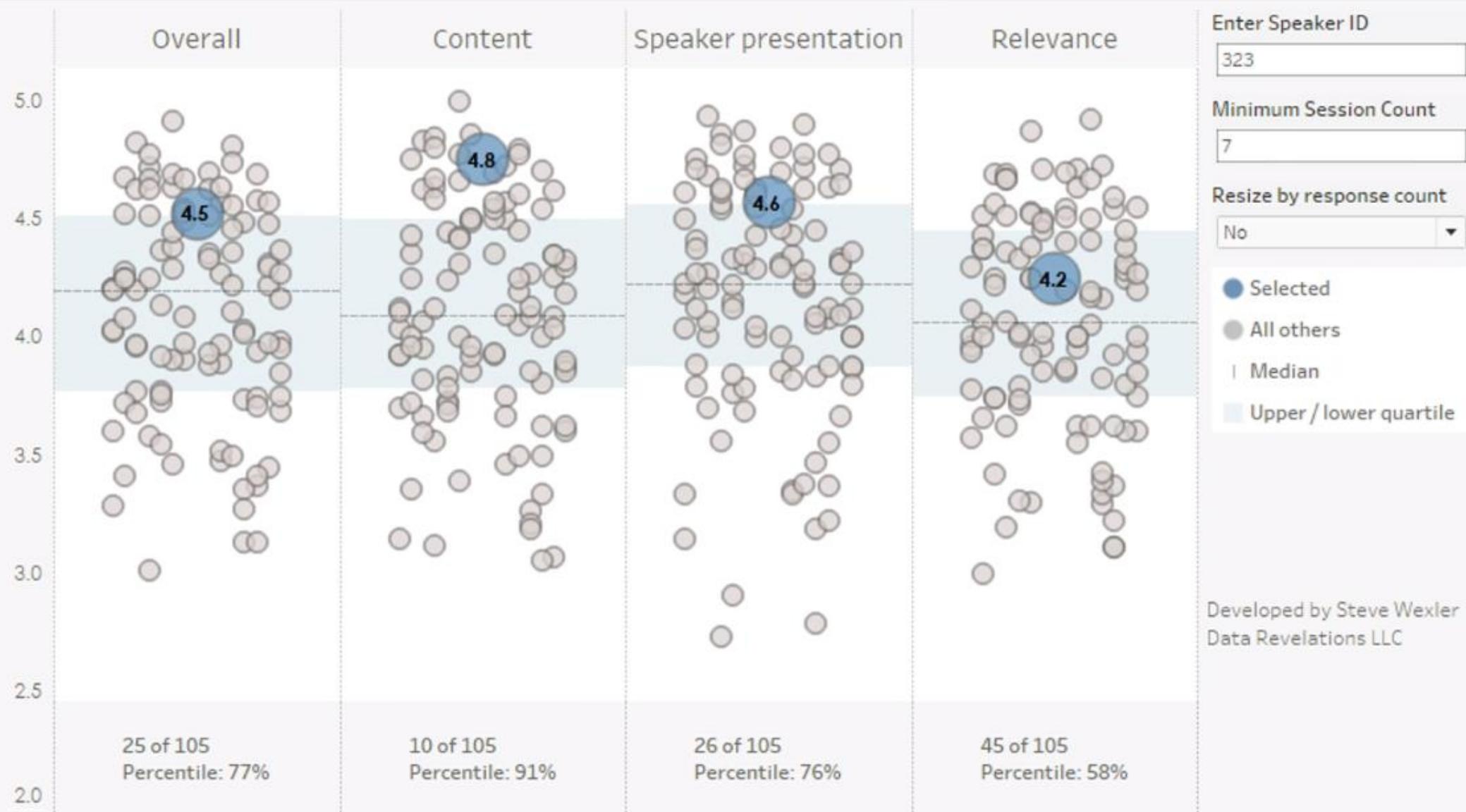
		Gained	Lost	Net	Running total
West	January	80	0	80	80
	February	80	-15	65	145
	March	90	-30	60	205
	April	120	-25	95	300
	May	100	-50	50	350
	June	119	-77	42	392
	July	75	-45	30	422
	August	119	-77	42	464
	September	90	-30	60	524
	October	80	-15	65	589
	November	80	-20	60	649
	December	90	-30	60	709
	Total	1,123	-414	709	709
Central	January	60	0	60	60
	February	85	-45	40	100
	March	80	-27	53	153
	April	90	-17	73	226
	May	120	-33	87	313
	June	45	-80	-35	278
	July	75	-45	30	308
	August	45	-80	-35	273
	September	80	-27	53	326
	October	85	-45	40	366
	November	60	-35	25	391
	December	80	-27	53	444
	Total	905	-461	444	444
East	January	70	0	70	70
	February	80	-90	-10	60
	March	100	-30	70	130
	April	110	-45	65	195
	May	70	-95	-25	170
	June	45	-33	12	182
	July	50	-110	-60	122
	August	99	-34	65	187
	September	112	-34	78	265
	October	99	-88	11	276
	November	55	-65	-10	266
	December	110	-45	65	331
	Total	1,000	-669	331	331
	Grand Total	3,028	-1,544	1,484	1,484



**There Should Be
Lots of Debate**

Speaker Ratings Comparison

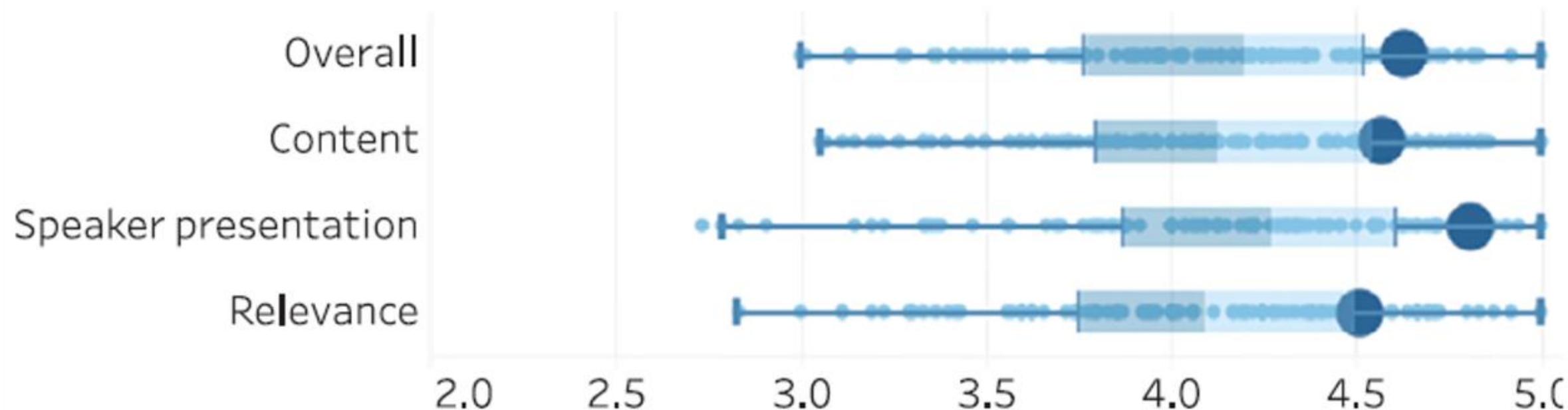
Speaker 323 compared with all other speakers



Source: The Big Book of Dashboards (BigBookofDashboards.com)

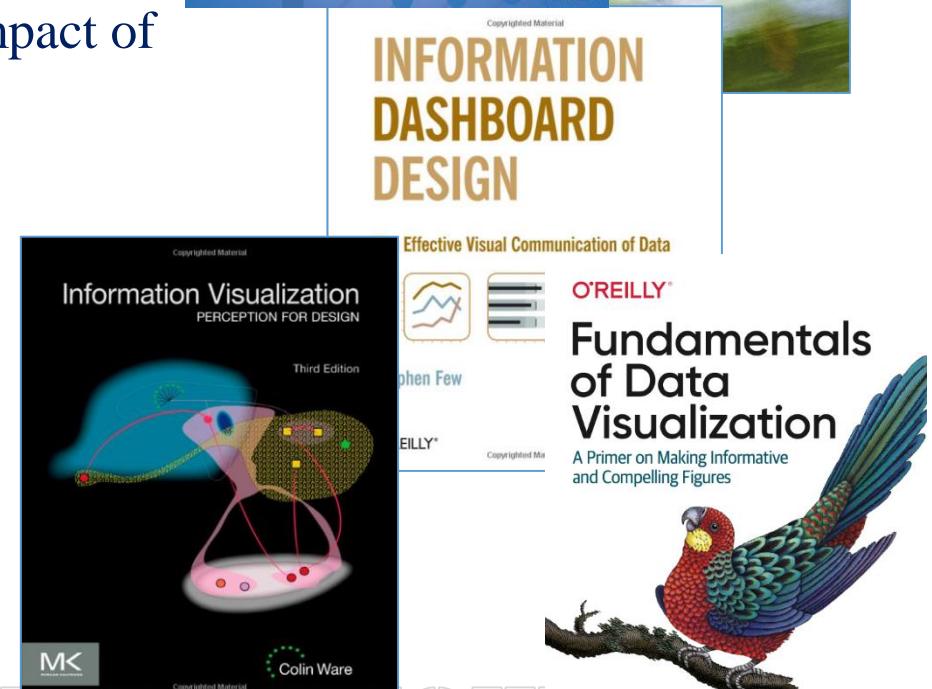
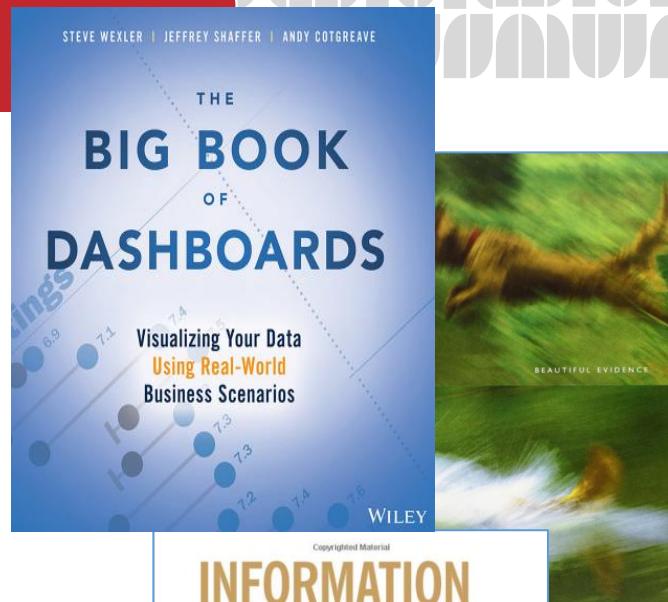
Speaker rating comparison

Speaker 323 compared with all other speakers.



Want to read more about Visualization?

- ⊕ The Big Book of Dashboards
- ⊕ Fundamentals of Data Visualization – Claus O. Wilke
- ⊕ Visual Reporting and Analysis: Seeing is Knowing Whitepaper
- ⊕ Visual Analysis Best Practices: A Guidebook Whitepaper
- ⊕ Data Storytelling: Using visualization to share the human impact of numbers Whitepaper
- ⊕ Beautiful Evidence – Edward Tufte
- ⊕ Information Dashboard Design – Stephen Few
- ⊕ Information Visualization – Colin Ware
- ⊕ Data Visualization: A Handbook for Data Driven Design



Claus O. Wilke

Homework

⊕ The Groups (only 2 for each group)

⊕ The Scenarios

- What is a story?
- Expected layout
- Make a figure for the generals
- Build up towards complex figures
- Make your figures memorable
- Be consistent but don't be repetitive

Example:

- Monitor, report, and predict Egyptian airline performance.

Specifics:

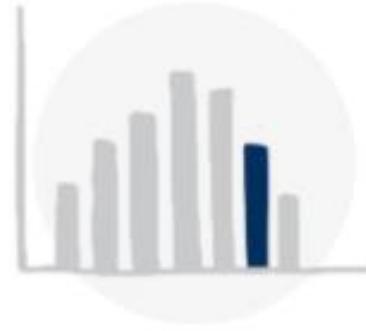
- Top 10 airline carrier in year 2019 in terms of number of flights.
- Number of flights in 2019 split by month.
- Number of travelers from Cairo to other cities.
-



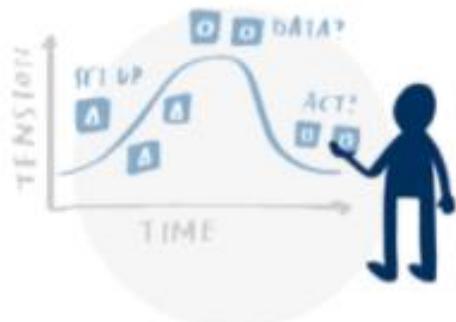
**understand the
context**



**choose an
effective visual**



**focus
attention**



**tell a
story**

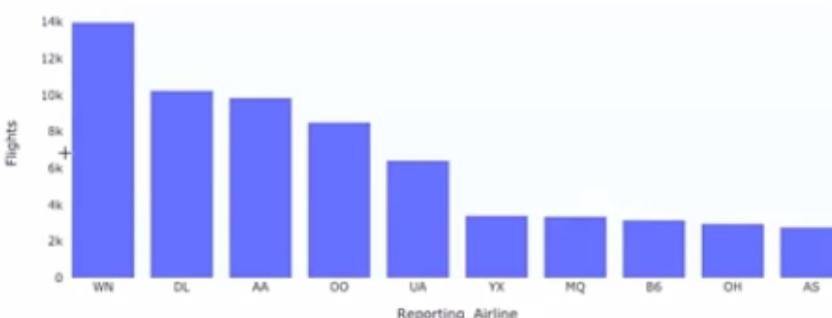
Reporting_Airline	Flights
WN	13972
DL	10241
AA	9853
OO	8515
UA	6430
YX	3416
MQ	3368
B6	3170
OH	2969
AS	2783

Month	Flights
1	6125
2	5578
3	6553
4	6282
5	6441

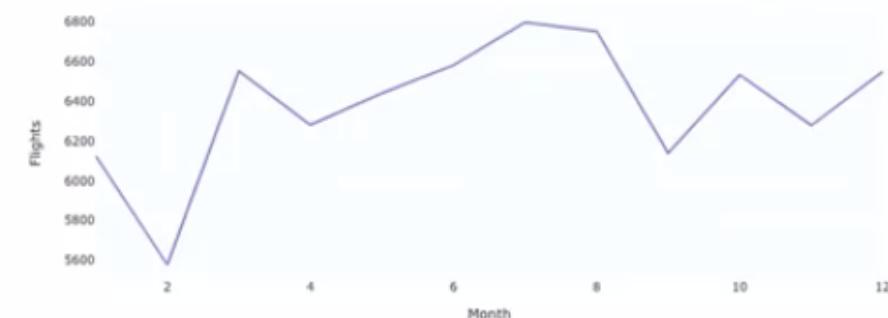
DestState	DistanceGroup	Flights
AK	10	4
AL	8	1
AR	7	1
AZ	2	361
AZ	3	220
AZ	4	9
CA	1	283
CA	2	2256
CA	3	36
CO	3	19

Airline Dashboard

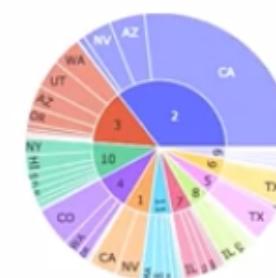
Top 10 airline carrier in year 2019 in terms of number of flights



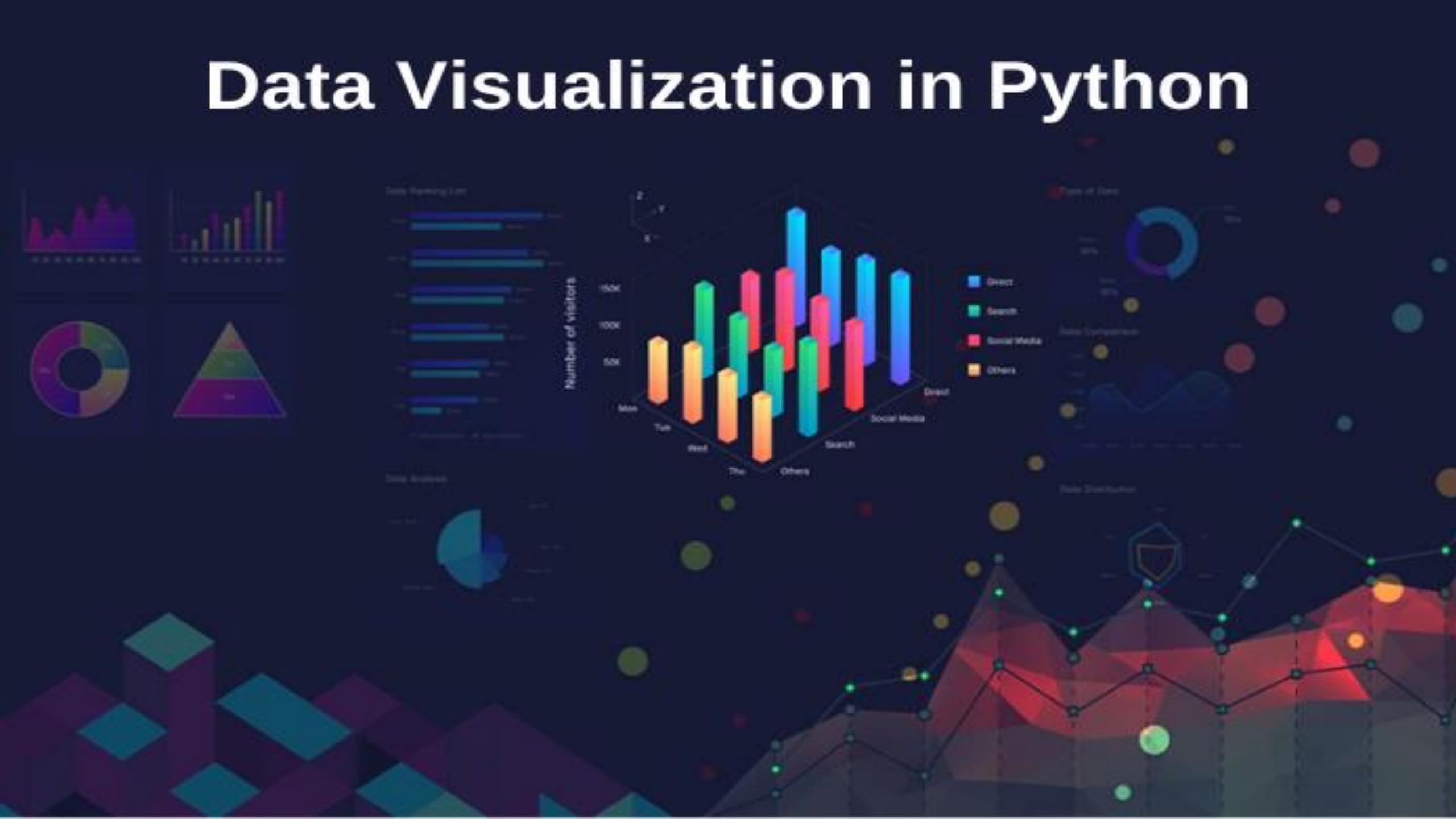
Number of flights by month - 2019



Number of travelers from CA to other states split by distance group



Data Visualization in Python



Plotly and Dash

- What libraries already exist to visualize data with python?

- **Matplotlib**

- “**Grandfather**” of them all.
- Creates **static** image files.
- Can create almost any plot type.

Plotly and Dash

- What libraries already exist to visualize data with python?
 - **Seaborn**
 - Uses **matplotlib** on the background.
 - Designed to create nice looking statistical plots.(**static plots**)
 - Can only create **statistical** plots available in documentation.

Plotly and Dash

- What libraries already exist to visualize data with python?

- **Pandas**

- Main purpose is data analysis
- Uses matplotlib on the background through `.plot()`.
- Limited scope of plot types.
- Can only create **static** plots.

Plotly and Dash

- What is Plotly?

- Plotly the **open source** library is a general data visualization library focused on interactive visualizations.
- Plotly has libraries for **JavaScript**, **React**, **R** and **Python**.
- The most popular version is the Python library.
- Using the Plotly python library by itself creates interactive plot as **.html** files.
- While users can still interact with these plots (zoom in, select, hover over).

Plotly and Dash

- What is Plotly?
 - Once the interactive Plotly plot is generated, the data the plot represents is “locked in” to the exported state of the plot.
 - You would need to re-run the `.py` script and re-generate the `.html` file to see any updates.

Plotly and Dash

- What is Dash?
 - Often users want plots to be able to **interact** with each other, interact with components, or have plots update in **real** time.
 - To accomplish this level of tasks, we need a **dashboard**.

Plotly and Dash

- What is Dash?
 - Dash is an **open-source** library from the Plotly company that allows you to create a full dashboard with **multiple components, interactivity, and multiple plots.**
 - Instead of creating an.html file, Dash will produce a dashboard web application at a local URL (**127.0.0.1:8050**).
 - You can visit and interact with this dashboard in the web application.
 - Since Dash renders a full web app, you can then **deploy** your dashboards online.

Plotly and Dash

- Conclusion

- Python has many visualization libraries(**static**)
- **Plotly** creates interactive images.
- **Dash** creates dashboard web apps.
- To fully understand Dash, we should first get comfortable with Plotly.

Data Visualization in Data Science

The First Mile ...

Plotly Express

- ⊕ Making graphics in a single function call.
- ⊕ Quickly iterate through various visual forms
- ⊕ Easily tailor graphics for publication

Pip install plotly

... and the Last Mile.

Dash

- ⊕ Make web application that link graphics, controls & text
- ⊕ Control every aspect of styling and layout
- ⊕ Easily deploy anywhere Flask can run

Pip install dash
Pip install jupyter_dash

- <https://plotly.com/python-api-reference/generated/plotly.data.html>
- <https://plotly.com/python-api-reference/plotly.express.html#px>
- <https://plotly.com/python/plotly-fundamentals/>
-

Plotly – An Overview

- Interactive, open-source plotting library.
- Support over 40 unique chart types.
- Include chart types like statistical, maps, scientific, and 3-D.
- Visualization can be displayed in Jupyter notebook, saved to Html files, or can be used in developing python- built web applications.

[Plotly Notebook](#)

Plotly Sub-Modules

- Plotly Graph Objects: Low-level interface to figures, traces, and layout

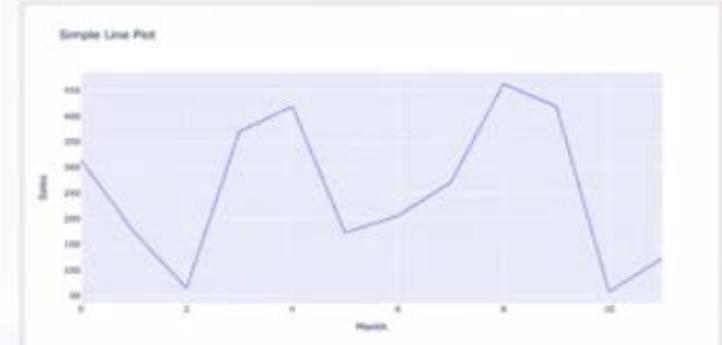
`plotly.graph_objects.Figure`

- Plotly Express: High-level wrapper

```
# Line Plot using graph objects.  
# Plotly.graph contains a JSON object which has a structure of dict.  
# Here, 'go' is the plotly JSON object.  
# Updating values of 'go' object keywords, chart can be plotted.  
# Create figure and add trace (scatter)  
  
fig = go.Figure(data=go.Scatter(x=x, y=y))  
fig.update_layout(title='Simple Line Plot', xaxis_title='Month', yaxis_title='Sales')  
fig.show()
```



```
# Entire line chart can be created in a single command  
fig = px.line(x=x, y=y, title='Simple Line Plot', labels=dict(x='Month', y='Sales'))  
fig.show()
```



**BE CREATIVE
BE UNIQUE
BE YOU!**

Plotly Express Mini Lab

<https://plotly.com/python-api-reference/generated/plotly.data.html>

<https://plotly.com/python-api-reference/plotly.express.html#px>

<https://plotly.com/python/plotly-fundamentals/>

Thank You

