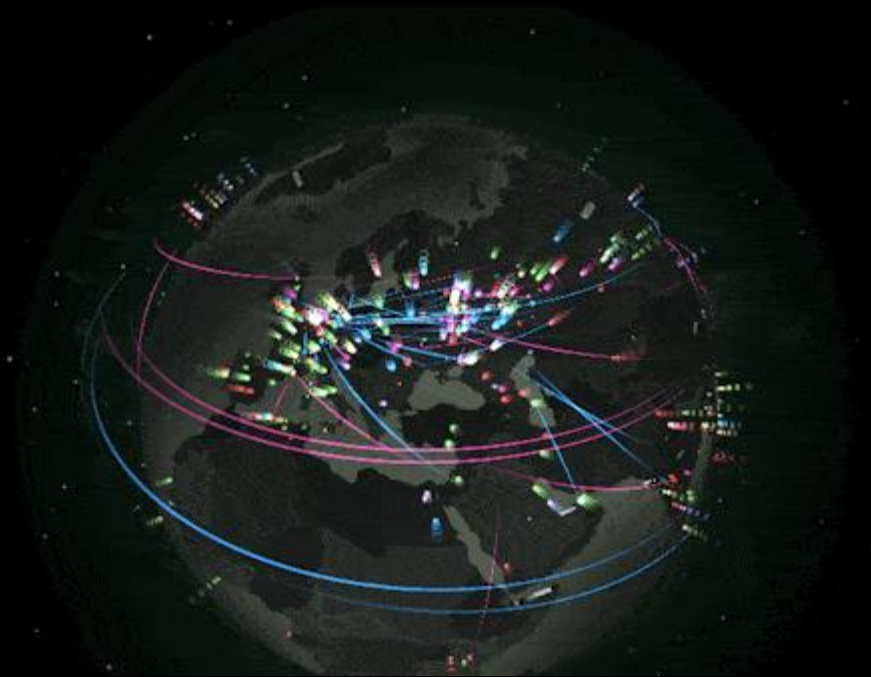


Data Visualisation

Simon Temby | Data and AI

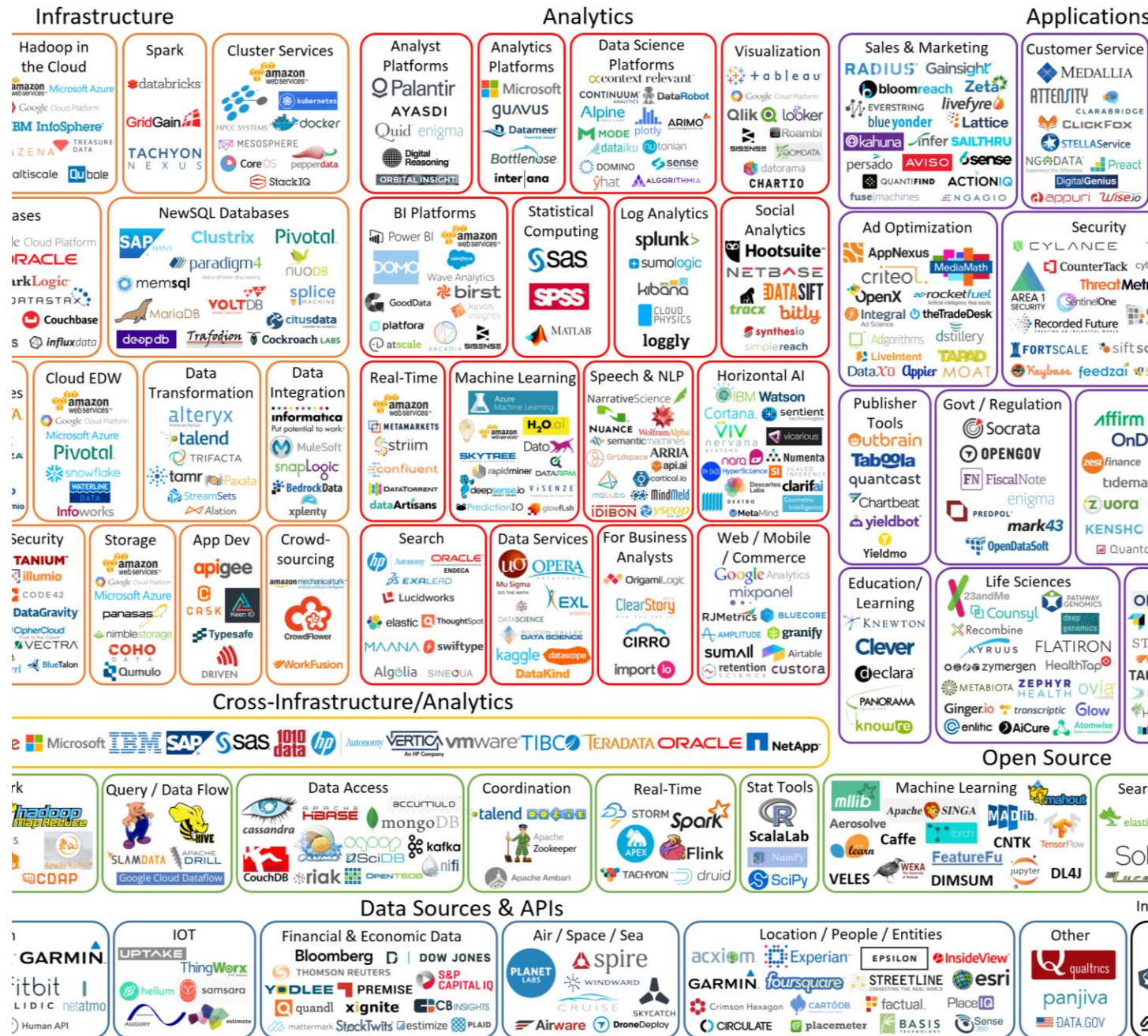


Agenda

-
- Data landscape
 - What is Data Visualisation, and Why
 - Terminologies
 - Types of visualisations
 - Comparisons (statistics vs visualisation)
 - Examples
 - Useful Resources

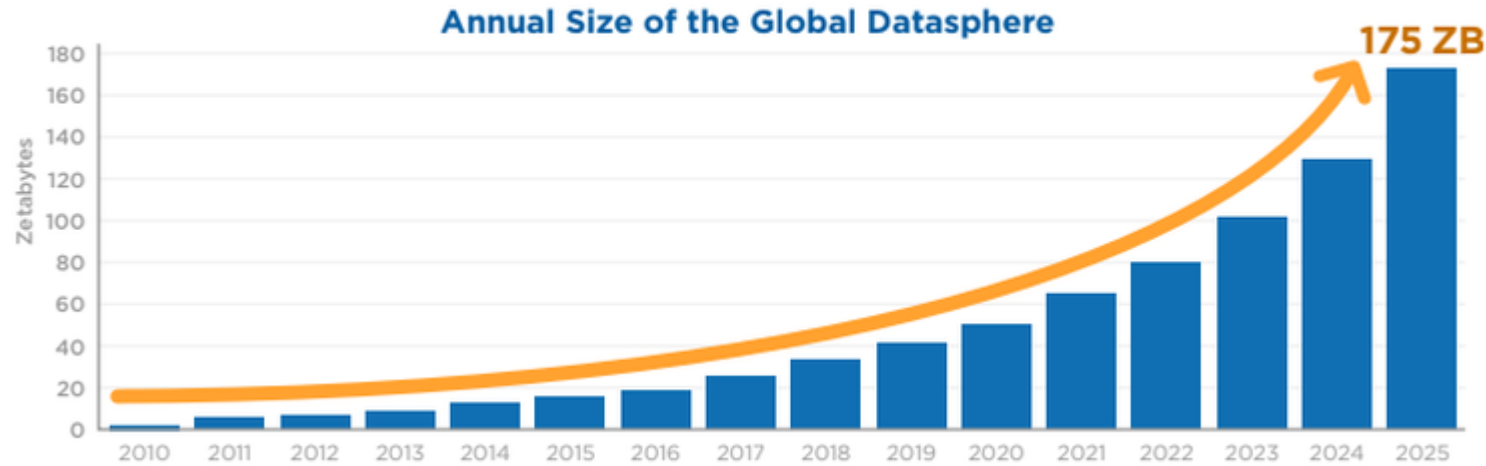
Data Landscape

Big Data Landscape 2016 (Version 3.0)



Data Landscape

Figure 1 - Annual Size of the Global Datasphere



Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

data.gov.au statistics

5.4k

discoverable
datasets

8.6k

API enabled
resources

3.3k

openly licenced
datasets

22

unpublished
datasets

Big Data

Big data describes that volume and variety of data sets that can be used for analysis.

The Five V's of Big Data



Scale of Data

This refers to the sheer volume of data being generated every second.

6 Billion People have cell phones



40 Zettabytes of data will be created by 2020 and increase of 300 times from 2005



Most companies in the U.S. have at least **100 Terabytes** of data stored.



1 in 3 Business leaders don't trust the information they use to make decisions



Uncertainty Of Data

This refers to the discrepancies found in the data.

Poor data quality costs the US economy around **\$ 3.1 Trillion a year**



The New York Stock Exchange capture **1 TB of Trade Information**

Analysis of Streaming Data

Denotes the speed at which data is emanating and changes are occurring between the diverse data sets.



By 2016 it is projected there will be **18.9 Billion** network connections

Modern cars have close to **100 Sensors**



4 Billion+ hours of video are watched on You Tube each month



30 Billion pieces of content are shared on facebook every month



400 Million tweets are sent per day by about 200 million monthly active users

Different forms of data

As more and more data is being digitized.



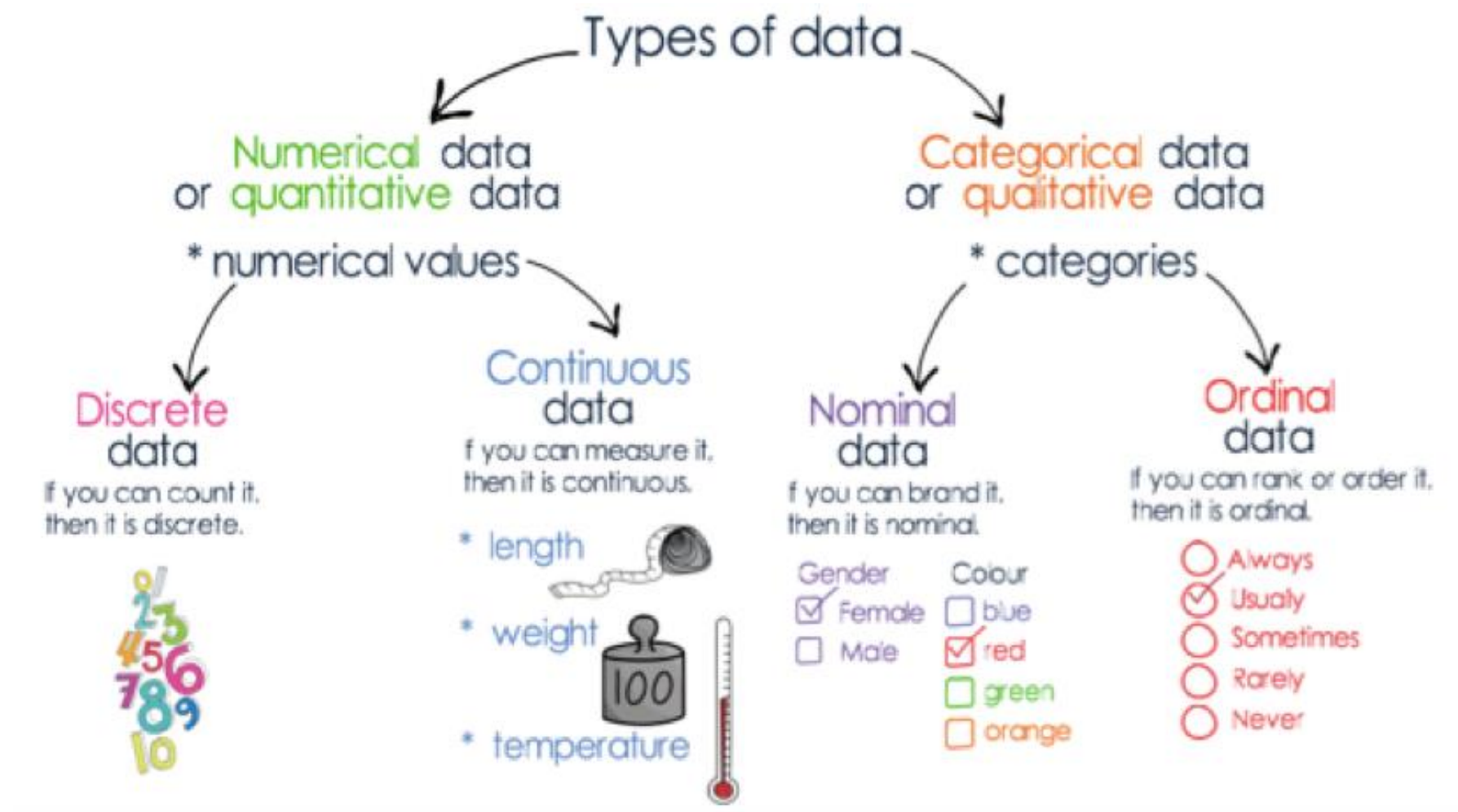
5V of Big Data

Value Of Data

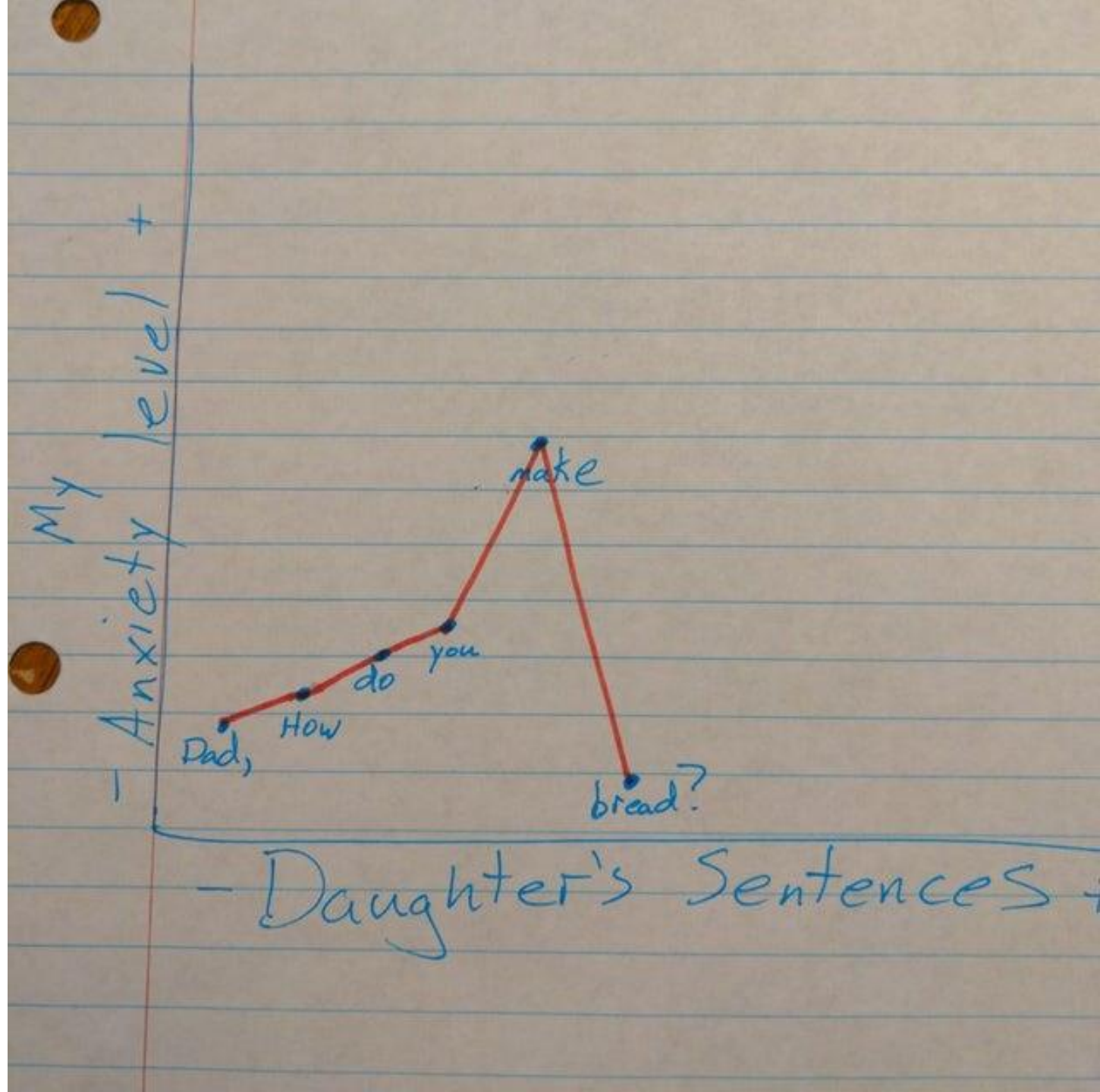
Having access to big data is all well and good but that's only useful if we can turn it into a value.



Types of Data



Data Visualisation



What is Data Visualisation and Why

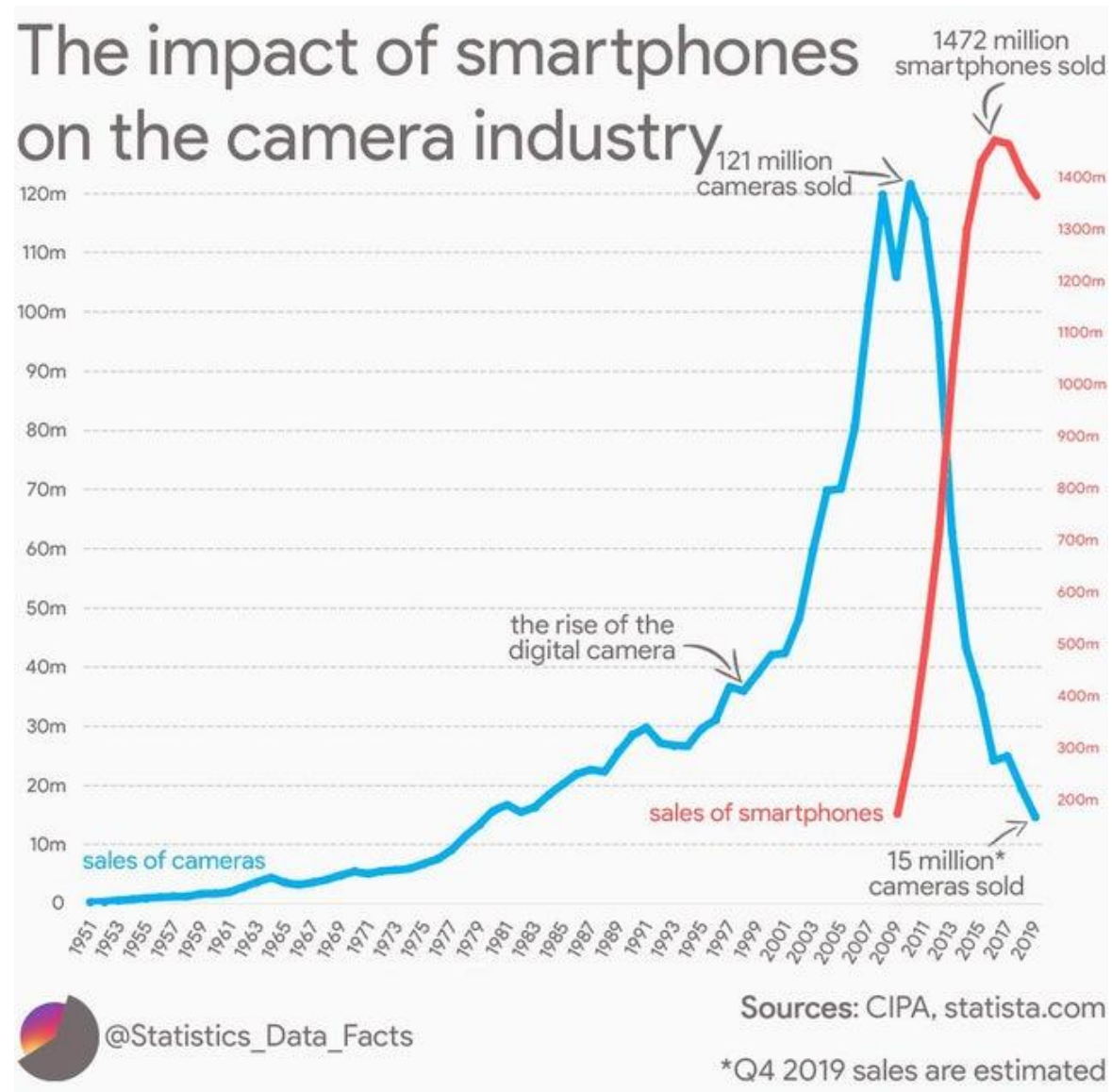
Our brains value visuals over any other type of information.

- **90%** of the information transmitted to the brain is visual - (Source: [MIT](#))
- The human brain can process an image in just **13 milliseconds** - (Source: [MIT](#))
- **50%** of the brain is active in visual processing - (Source: [Piktochart](#))
- Human brains process visuals **60,000 times faster** than they do text - (Source: [University of Minnesota](#))
- **93%** of communication is nonverbal - (Source: [Ubiquity](#))
- We are exposed to **5x more information today** than we were in 1986 - (Source: [Telegraph](#))

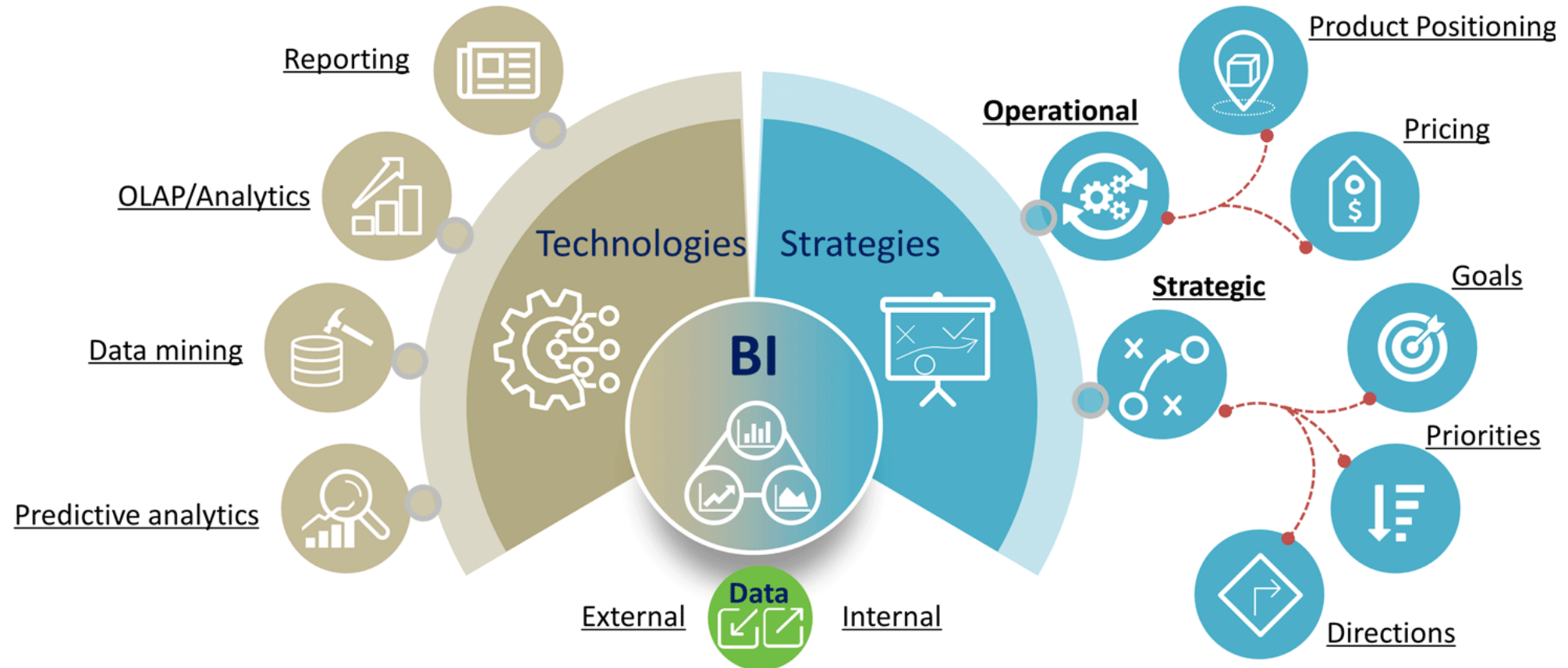
カメラ・交換レンズの総出荷① 出典：一般社団法人カメラ映像機器工業会
Total Shipments of Cameras and Interchangeable Lenses Source: CIPA

総出荷数量 Quantity of Total Shipments				単位:千個 Unit:1000pcs				
機種 Item	カメラ合計		デジタルカメラ		銀塩カメラ		カメラ用交換レンズ	
	Cameras Total		Digital Still Cameras		Film Cameras		Interchangeable Lenses	
	前年比(%)	前年比(%)	前年比(%)	前年比(%)	前年比(%)	前年比(%)		
年 Current Year	Comparison with the previous year		Comparison with the previous year		Comparison with the previous year		Comparison with the previous year	
1951	258				258			
1952	376	145.7			376	145.7		
1953	586	155.9			586	155.9		
1954	787	134.3			787	134.3		
1955	949	120.6			949	120.6	55	
1956	1,178	124.1			1,178	124.1	76	138.2
1957	1,292	109.7			1,292	109.7	69	90.8
1958	1,296	100.3			1,296	100.3	87	126.1
1959	1,725	133.1			1,725	133.1	104	119.5
1960	1,746	101.2			1,746	101.2	108	103.8
1961	1,997	114.4			1,997	114.4	99	91.7
1962	2,833	141.9			2,833	141.9	141	142.4
1963	3,702	130.7			3,702	130.7	221	156.7
1964	4,419	119.4			4,419	119.4	402	181.9
1965	3,637	82.3			3,637	82.3	468	116.4
1966	3,256	89.5			3,256	89.5	558	119.2
1967	3,650	112.1			3,650	112.1	671	120.3
1968	4,103	112.4			4,103	112.4	867	129.2
1969	4,825	117.6			4,825	117.6	1,180	136.1
1970	5,480	113.6			5,480	113.6	1,486	125.9
1971	5,143	93.9			5,143	93.9	1,697	114.2
1972	5,554	108.0			5,554	108.0	1,754	103.4
1973	5,749	103.5			5,749	103.5	1,954	111.4
1974	5,995	104.3			5,995	104.3	2,140	109.5
1975	6,808	113.6			6,808	113.6	2,071	96.8
1976	7,577	111.3			7,577	111.3	2,398	115.8
1977	9,166	121.0			9,166	121.0	2,914	121.5
1978	11,478	125.2			11,478	125.2	3,156	108.3
1979	13,391	116.7			13,391	116.7	3,856	122.2
1980	15,732	117.5			15,732	117.5	4,602	119.3
1981	16,730	106.3			16,730	106.3	5,102	110.9
1982	15,531	92.8			15,531	92.8	4,535	88.9
1983	16,358	105.3			16,358	105.3	4,719	104.1
1984	18,444	112.8			18,444	112.8	4,480	94.9
1985	20,246	109.8			20,246	109.8	4,855	108.4
1986	21,994	108.6			21,994	108.6	5,367	110.5
1987	22,671	103.1			22,671	103.1	5,236	97.6
1988	22,350	98.6			22,350	98.6	4,805	91.8
1989	25,675	114.9			25,675	114.9	5,761	119.9
1990	28,525	111.1			28,525	111.1	6,740	117.0
1991	29,764	104.3			29,764	104.3	6,601	97.9
1992	27,188	91.3			27,188	91.3	5,103	77.3
1993	26,765	98.4			26,765	98.4	4,658	91.3
1994	26,681	99.7			26,681	99.7	4,624	99.3
1995	29,571	110.8			29,571	110.8	4,456	96.4
1996	31,115	105.2			31,115	105.2	4,771	107.1
1997	36,671	117.9			36,671	117.9	5,423	113.7
1998	36,000	98.2			36,000	98.2	6,466	119.2
1999	38,967	108.2	5,088		33,879	94.1	5,954	92.1

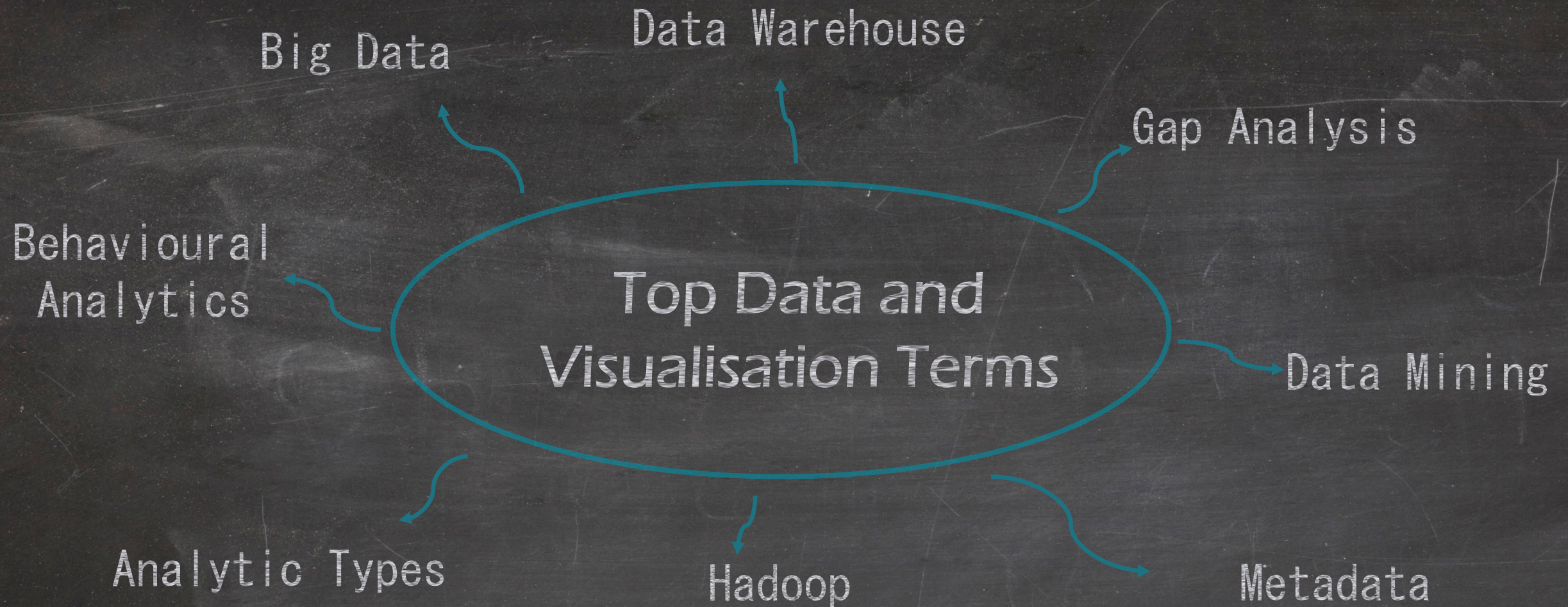
What's easier?



Business Intelligence



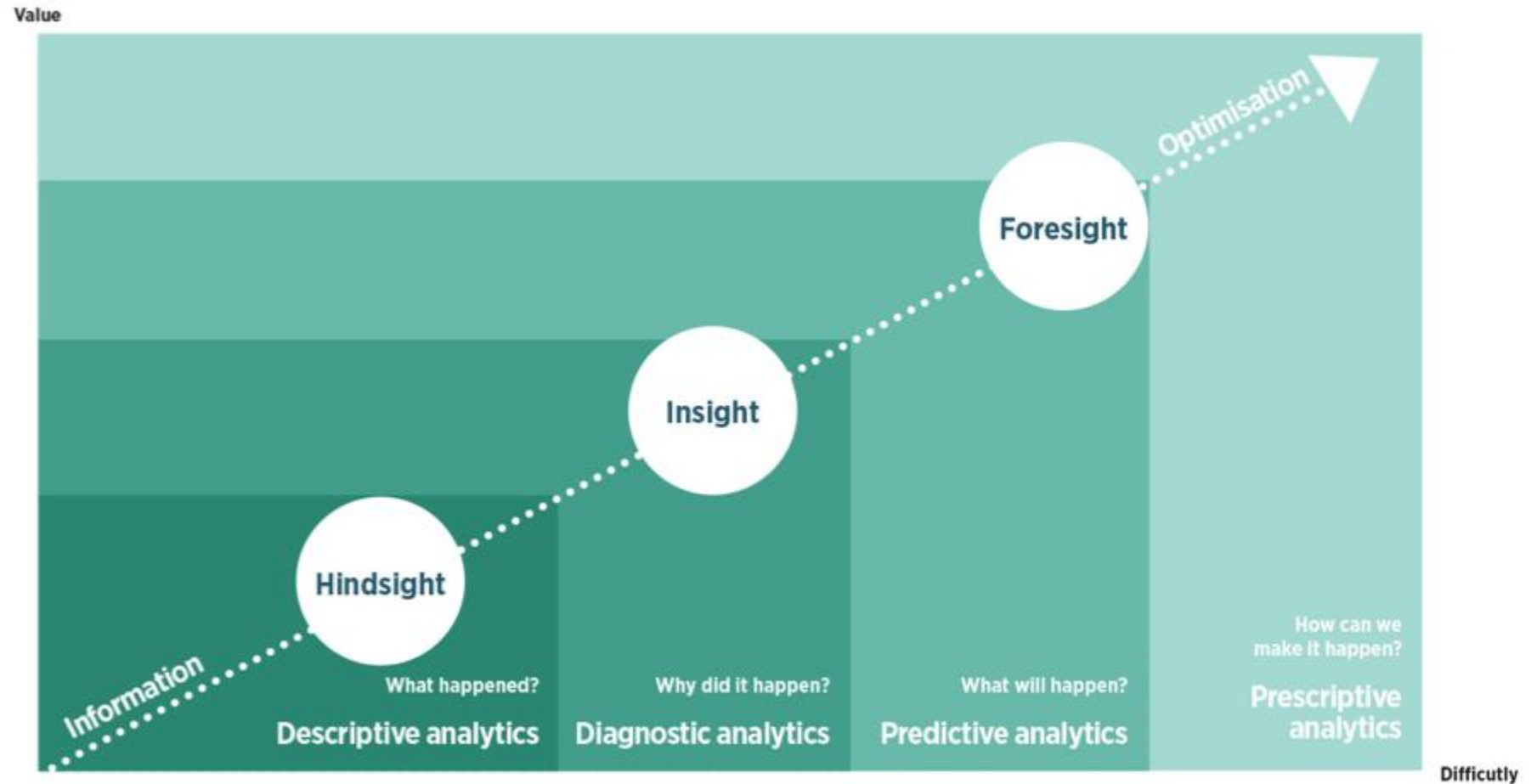




Terminologies

Word/Phrase	Description
Behavioural Analytics	Using data about people's behaviour to understand intent and predict future actions.
Big Data	Large volume, wide variety and high volume of data created or ingested.
Cube	A multi-dimensional collection of data that is created by an OLAP system with each sector organized into a hierarchy.
Data Cleansing	Transforming data in its native state to a pre-defined standardized format using vendor software.
Data Warehouse	A data repository that deals with multiple subject areas (or data marts).
Gap Analysis	A study of whether the data that a company has can meet the business expectations that the company has set for its reporting and BI, and where possible data gaps or missing data might exist.
Hadoop	A programming framework that supports the processing of large data sets in a distributed computing environment.
Metadata	Data that gives information about what the primary data is about (e.g., if a photo is the primary data, its metadata might consist of what its resolution is, when the photo was taken, etc.).
OLAP	A technology solution that is used to organize the databases of large businesses, supporting Business Intelligence.
Descriptive Analytics	This is generally point in time, current or real time information on data.
Diagnostic Analytics	Analytics on data that helps determine a cause of an event.
Predictive Analytics:	BI solutions that help the user discover patterns in large data sets in order to predict future behaviour.
Prescriptive Analytics:	The area of Business Intelligence dedicated to finding the best course of action for a given situation.
Real-Time Analytics:	The ability to use all available enterprise data as needed and usually involves streaming data that allows users to make business decisions on the fly.

Terminologies – Analytic Types

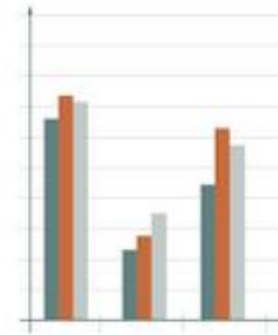


Source: Gartner

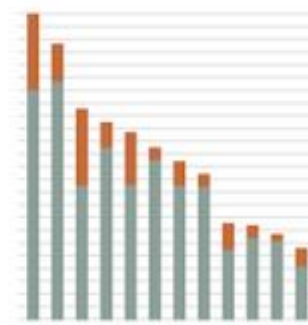
Types of Visualisations and Tools

Reading charts

Bar chart



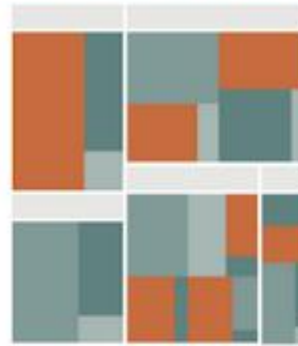
Stacked bar chart



Radar Chart



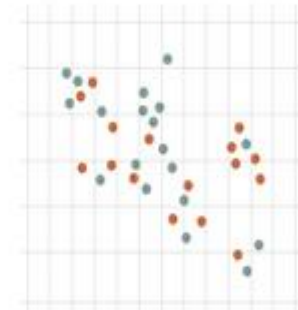
Tree map



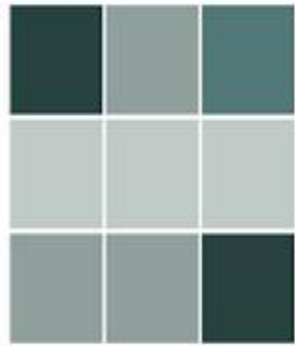
Sankey Diagram



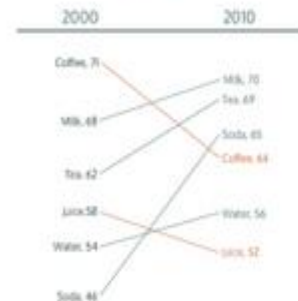
Scatter plot



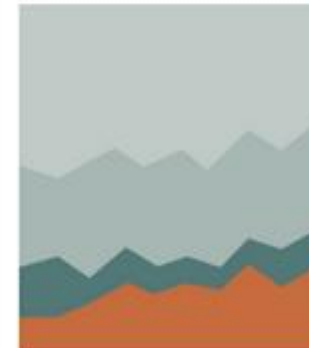
Heat map



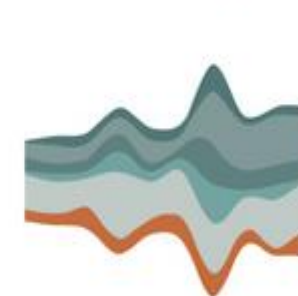
Slope graph



Stacked area chart



Stream Graph



Choropleth map



Types of Tools

There are more and more tools in the market than organisations know what to do with.

In government, we see Power BI, Tableau and Qlik leading the way, with SAP and SAS having similar offerings.



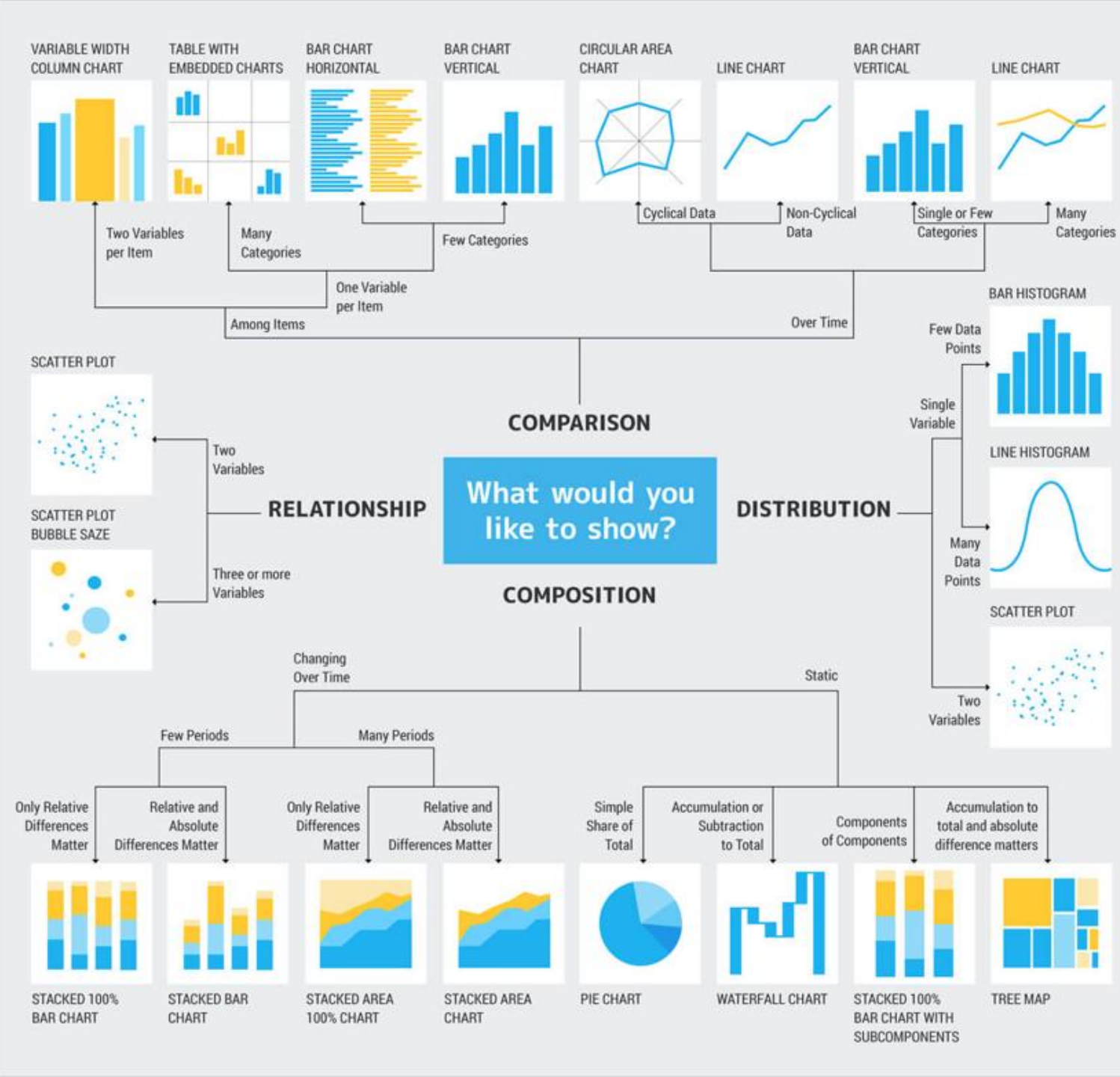
Power BI



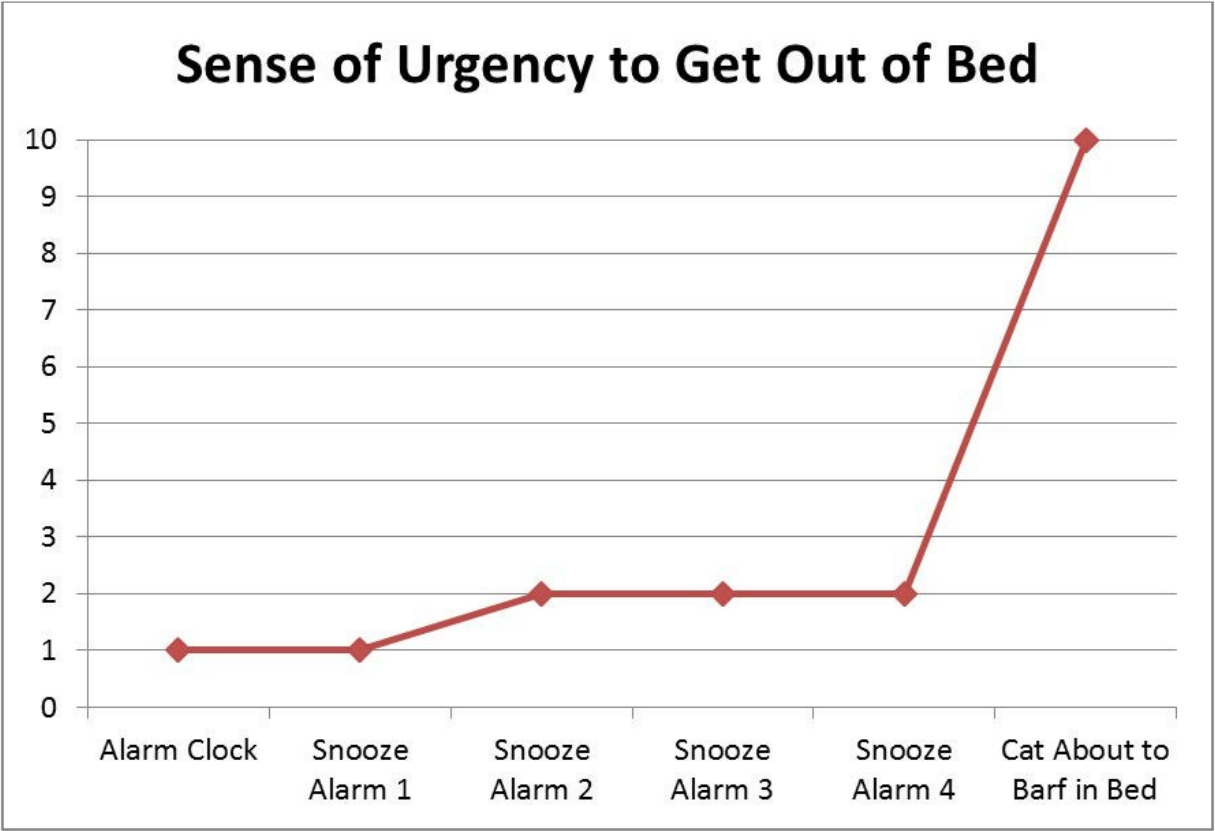
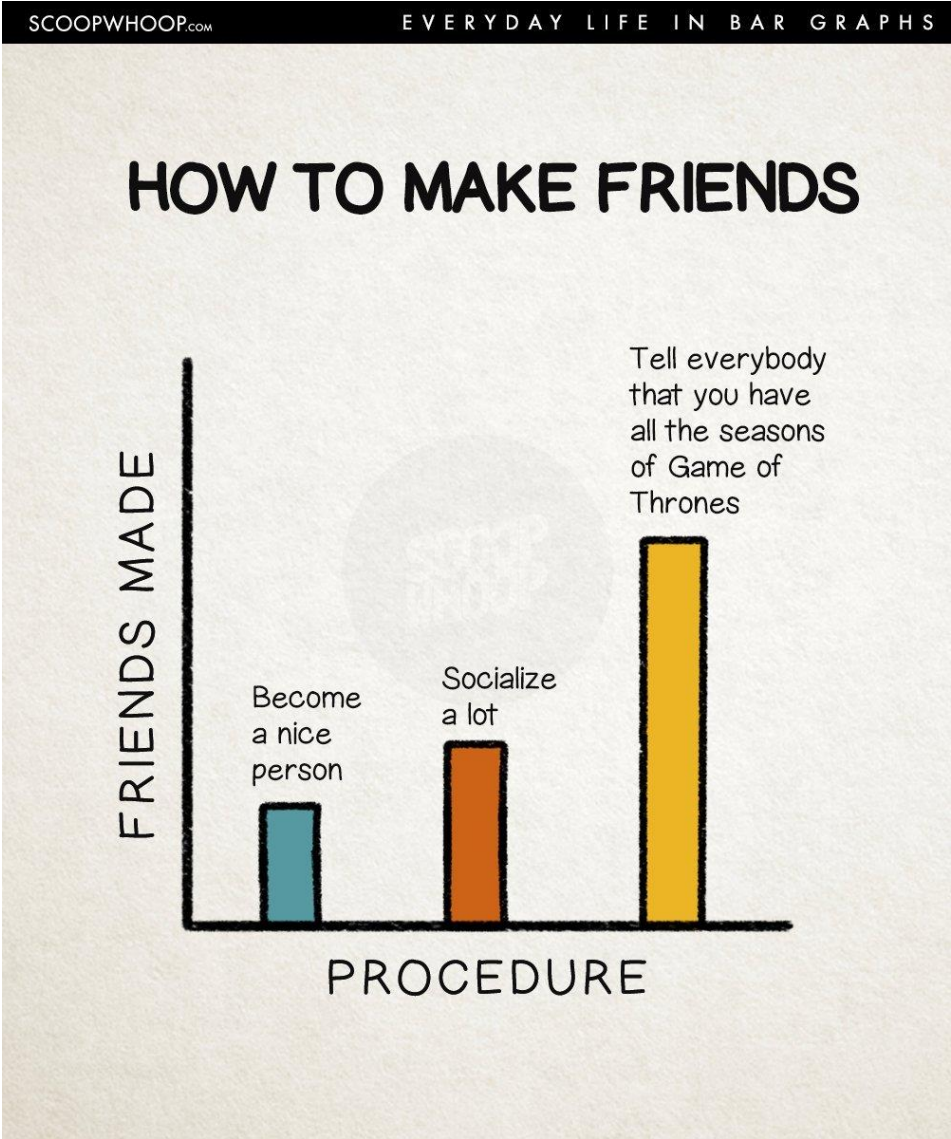
Types of Visualisation

There are multiple types and even variants of data visualisation options.

How do you know what to use?

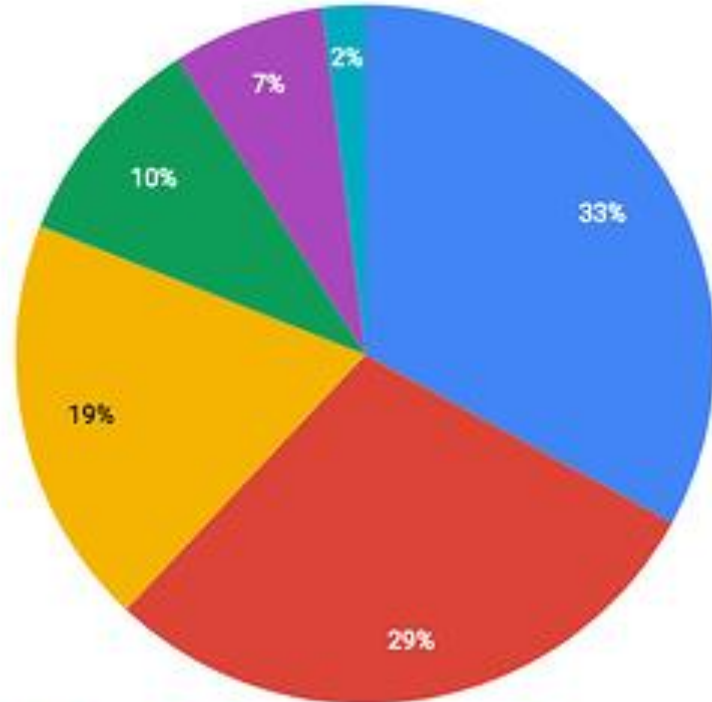


Bar and Line Charts



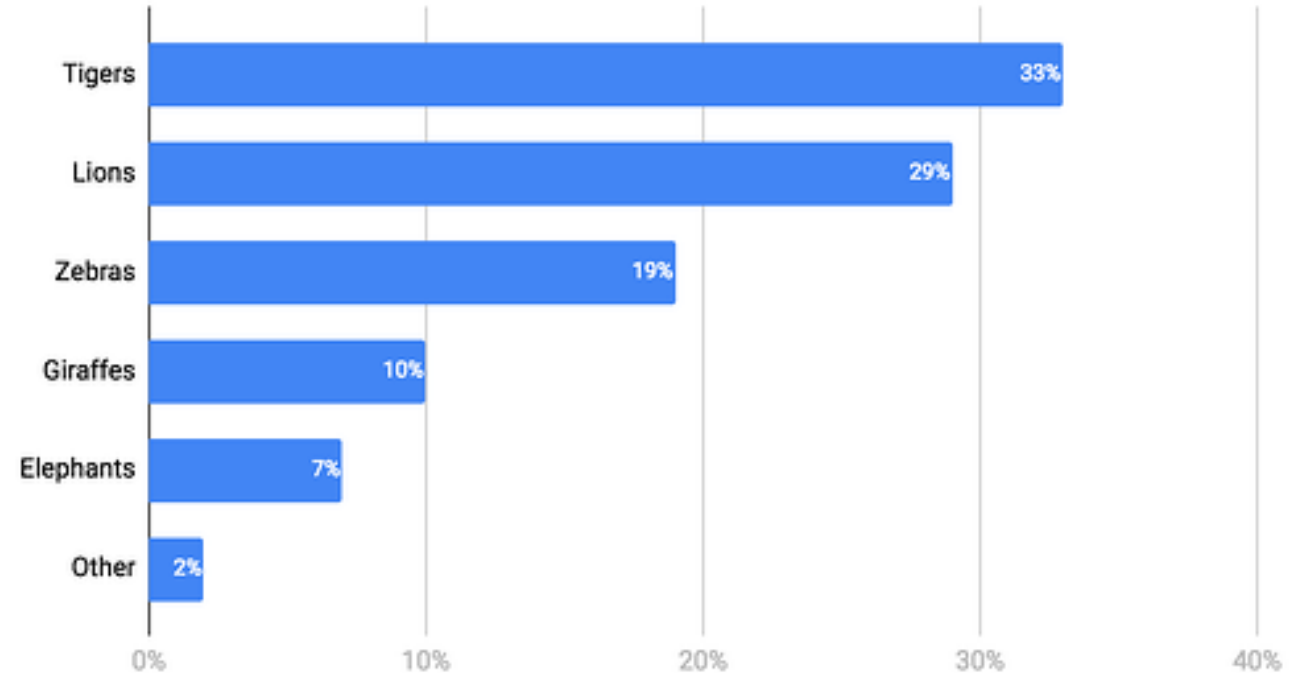
Pie Charts

What is your favourite animal?



● Tigers ● Lions ● Zebras ● Giraffes ● Elephants ● Other

What is your favourite animal?

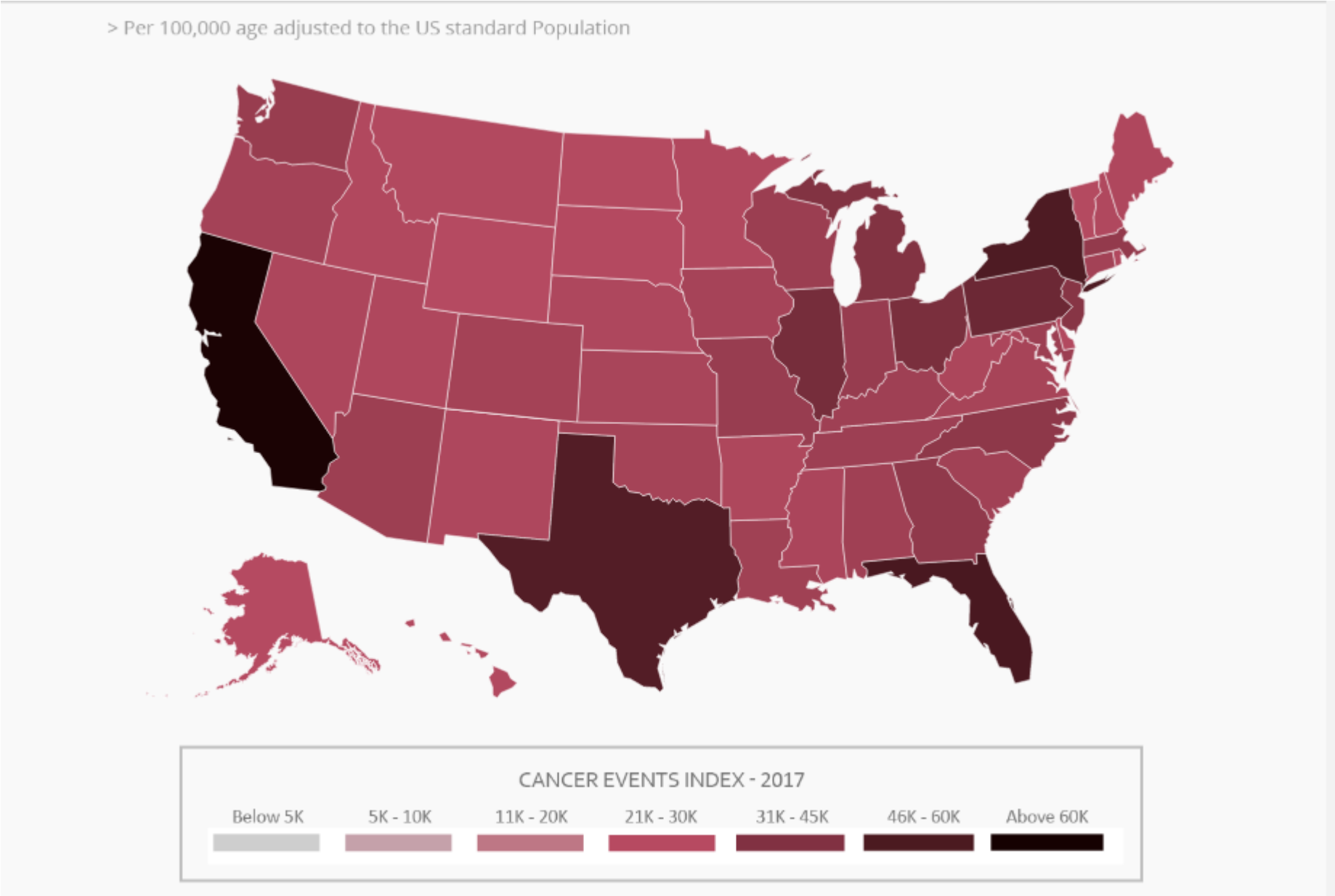


Maps

[Where The Wild Things Glow - Jonni Walker | Tableau Public](#)



Maps



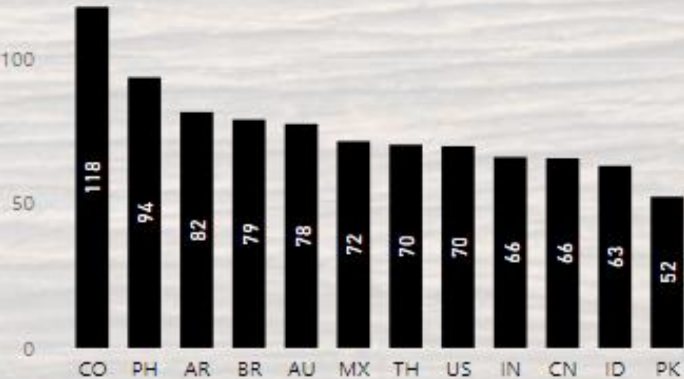
Sugarcane is an example of a renewable and versatile crop that can be used as a source of clean energy and raw material for products. Brazil is today the world's largest producer of sugarcane.

World Sugarcane Harvested (2010)



%Productivity (Country)

Planted x Harvested



Country

Country	Area (ha)	Production (ton)
Brazil	9,081	719,157
Argentina	355	290,000
India	4,200	277,750
China	1,695	111,454
Thailand	978	68,808
Mexico	704	50,423
Pakistan	943	49,373
Philippines	363	34,000
Australia	405	31,457
Indonesia	420	26,500
United States	355	24,821
Colombia	175	20,273
Total	19,674	1,704,016



Compare the crop weight with the weight of a whale.

Production (ton)

1,704K

Whales

11.36K



* Each drawing represents 100 whales.



Compare the harvested area size with soccer fields

Area (ha)

20K

Soccer Fields

23.85K



* Each drawing represents 1000 fields.



Next



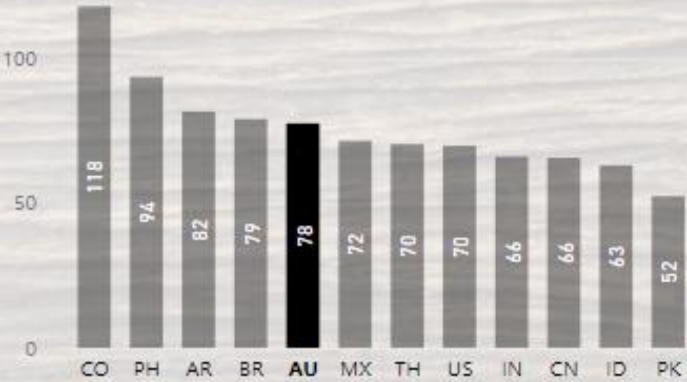
Sugarcane is an example of a renewable and versatile crop that can be used as a source of clean energy and raw material for products. Brazil is today the world's largest producer of sugarcane.

World Sugarcane Harvested (2010)



%Productivity (Country)

Planted x Harvested



Country

Country	Area (ha)	Production (ton)
Australia	405	31,457
Total	405	31,457



Compare the crop weight with the weight of a whale.

Production (ton)

31K

Whales

209.71



* Each drawing represents 100 whales.



Compare the harvested area size with soccer fields

Area (ha)

0K

Soccer Fields

0.49K

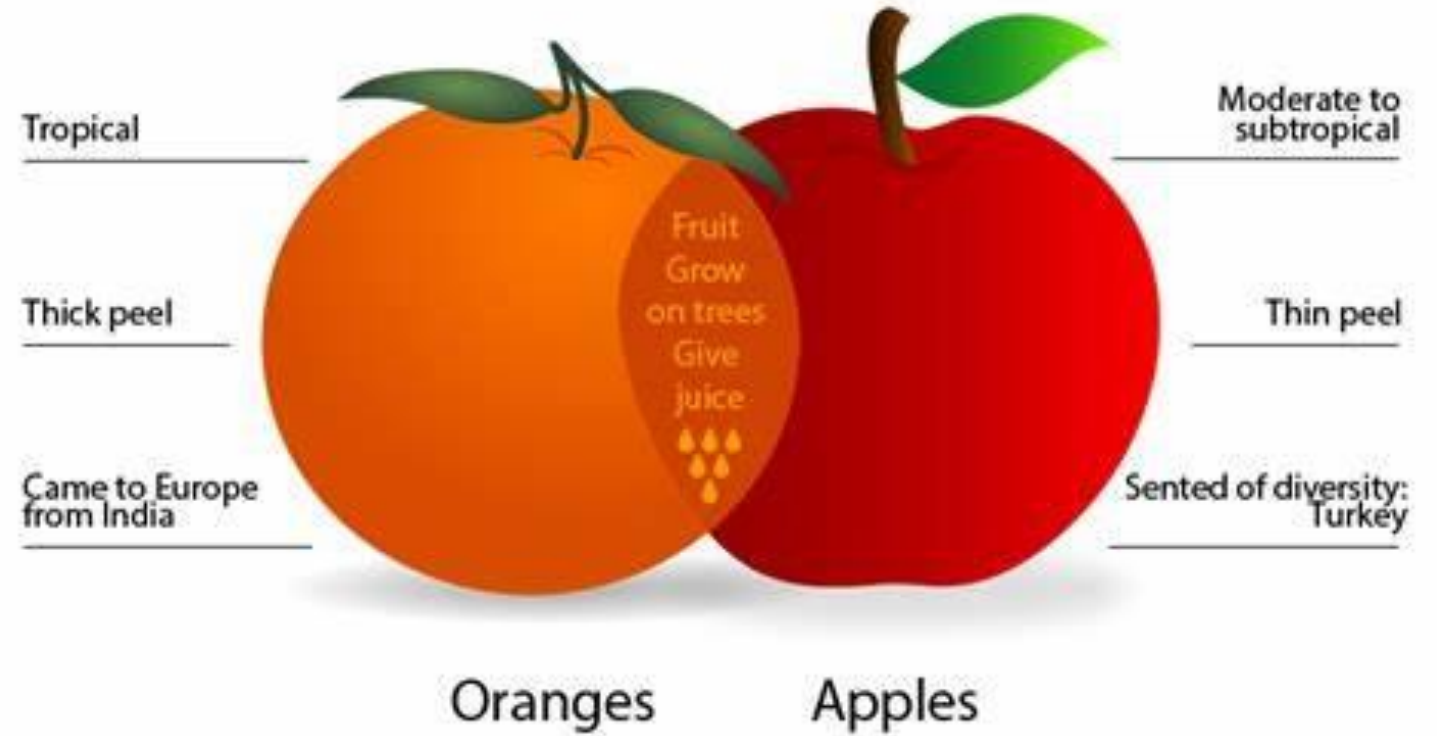


* Each drawing represents 1000 fields.

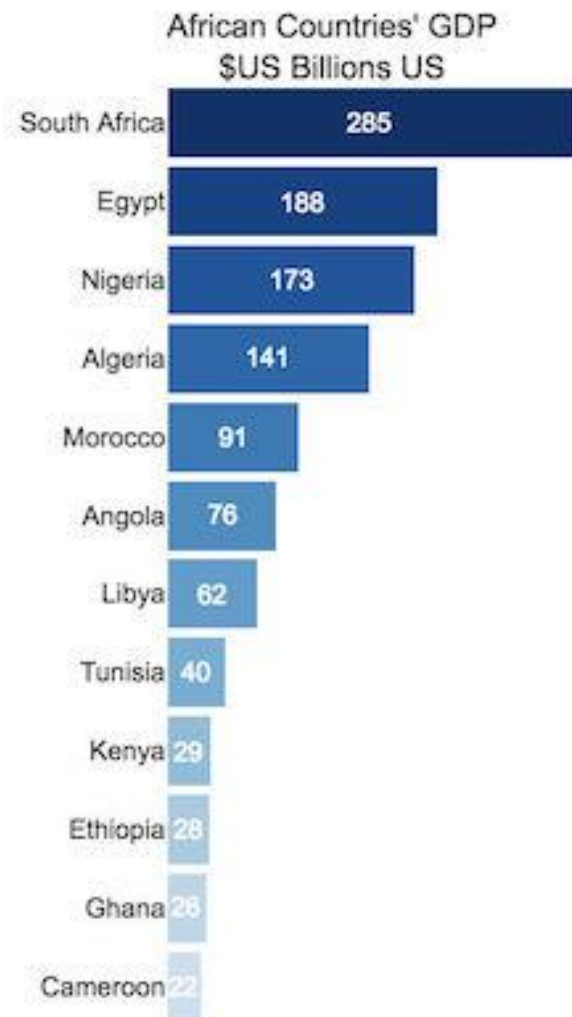
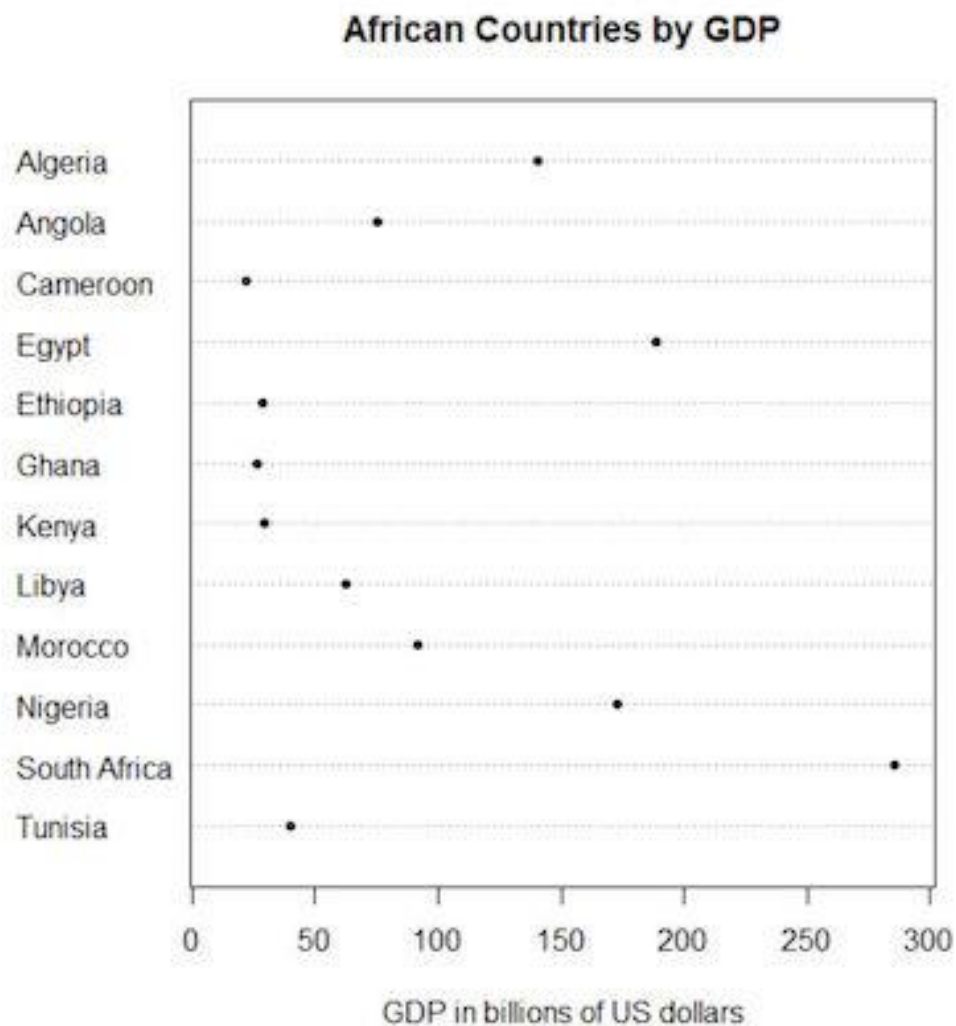


Next →

Compare the Pair

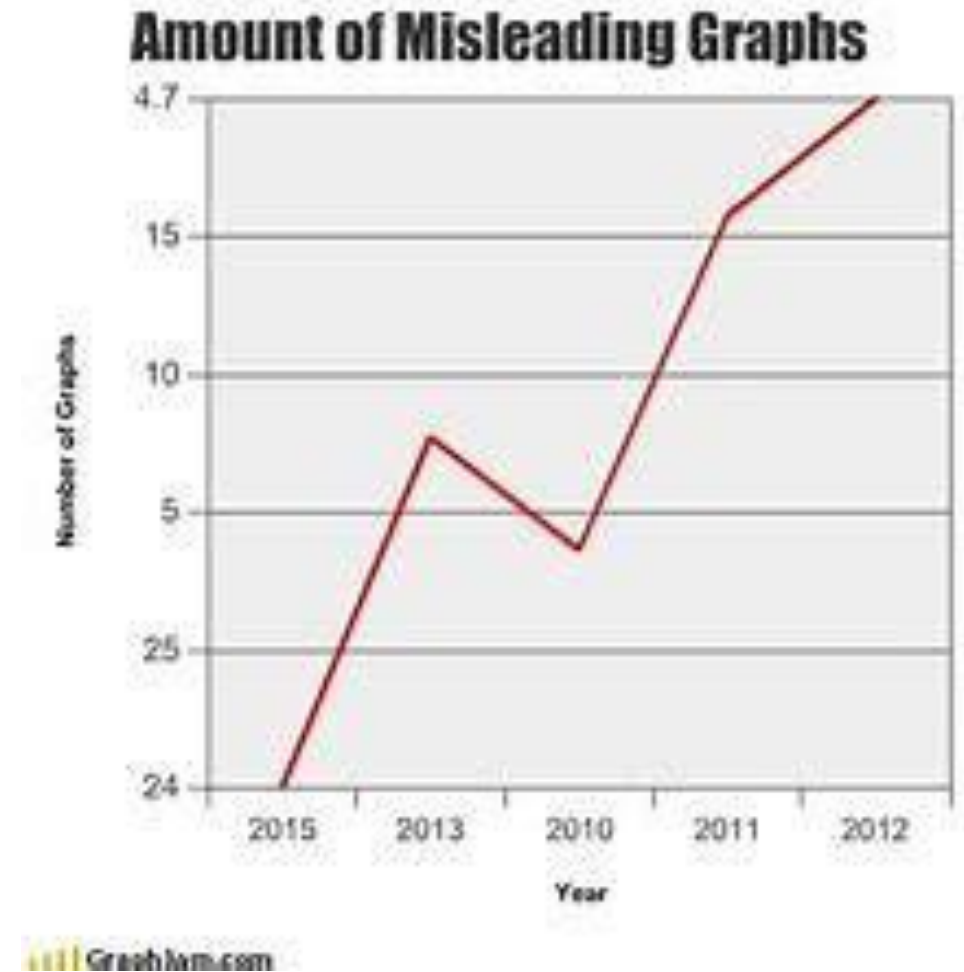


TO BE FINISHED



Example

- Main topic: Segoe UI, size 28pt
 - Segoe UI, size 20pt for second level
 - Segoe UI, size 16pt for third level



Examples

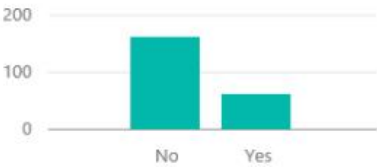
Ask a question about the data on this dashboard

[How to ask](#)

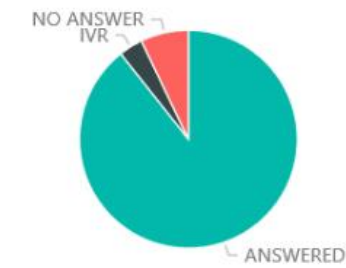
Missed Calls this month

4

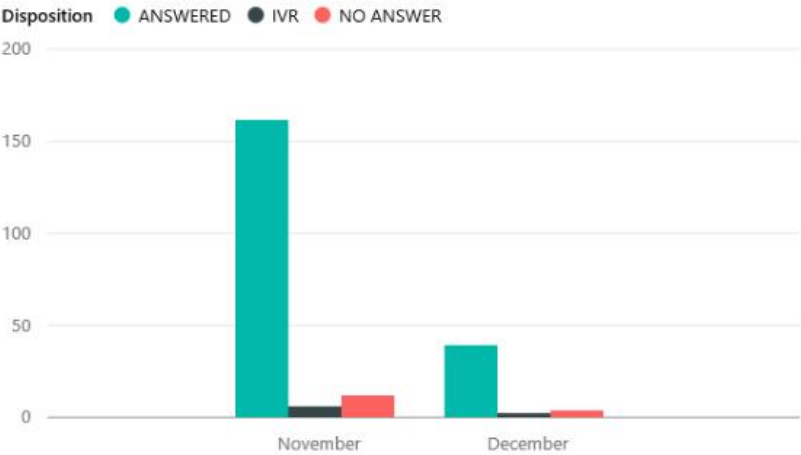
Repeat Callers



Calls by status



Call Volume by Month and Status

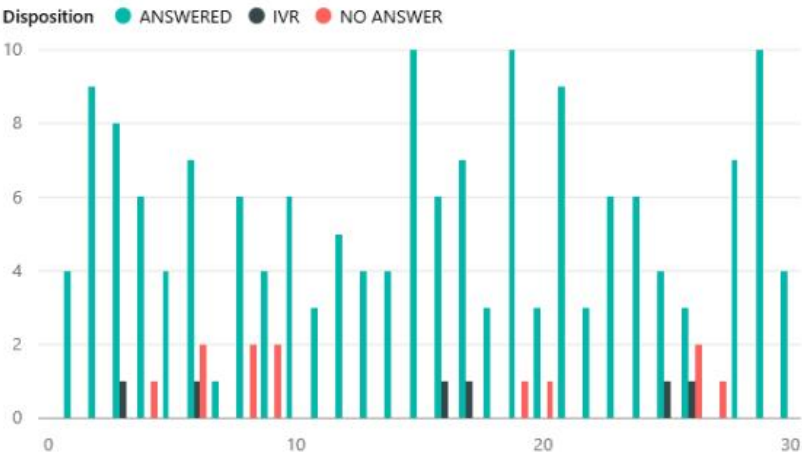


Location



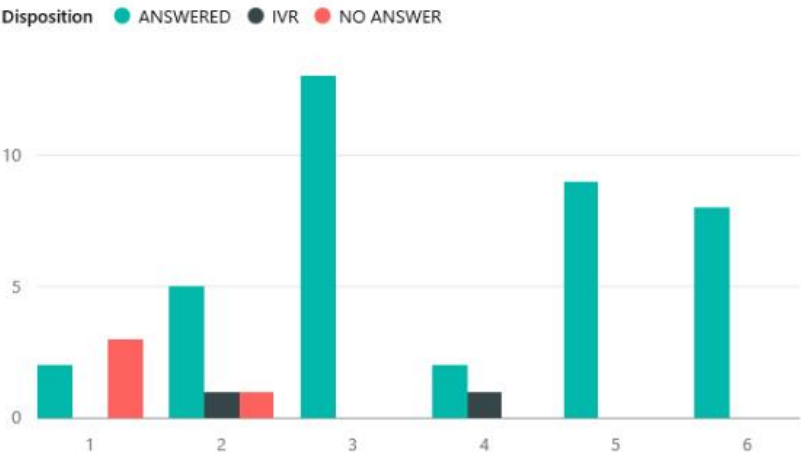
November Call Volume

BY STATUS



December Call Volume

BY STATUS

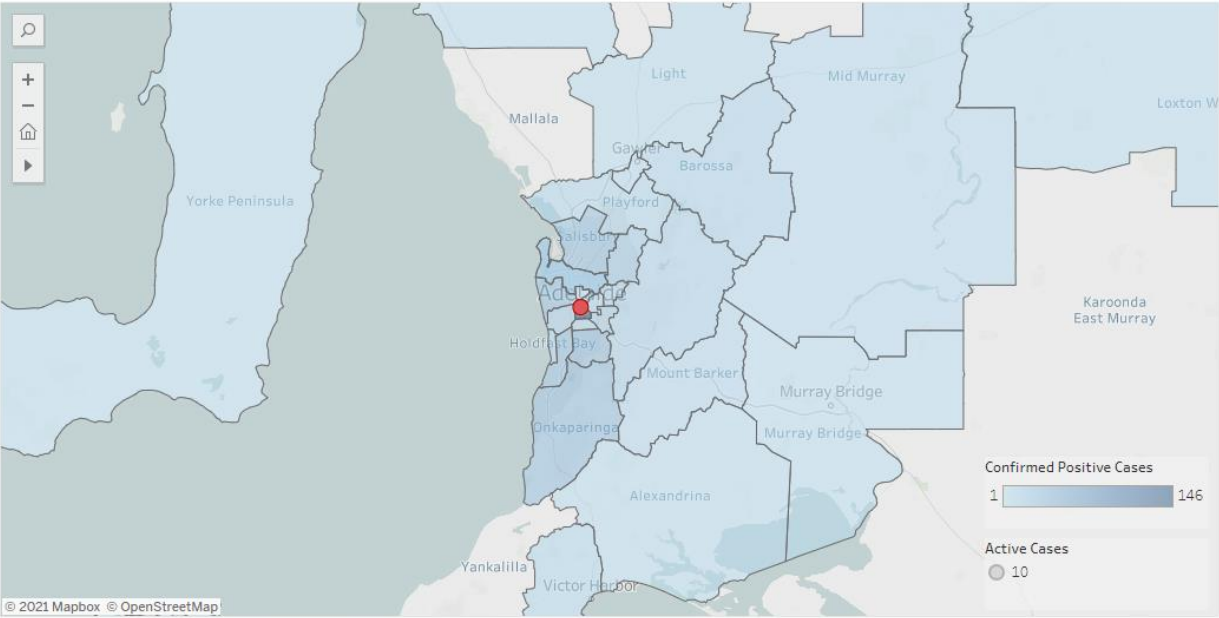


Frequent Callers by duration

0747...	07472...	04325...	0466...	0427...	0413...	041...
Duration	Duration	Duration	Durat...	Dura...	Dura...	Dura...
0885...	03855...	0408...	04781...	040...	04...	04...
Duration	Duration	Duration	Duration	Dur...	Du...	D...
04208...	0884...	04773...	0408...	02686...		
Duration	Duration	Duration	Duration	0417...	04020...	
0358...	07494...	0415...	0420...	04290...		
Duration	Duration	Duration				

Heatmap – SA

COVID-19 cases in South Australia by Local Government Area

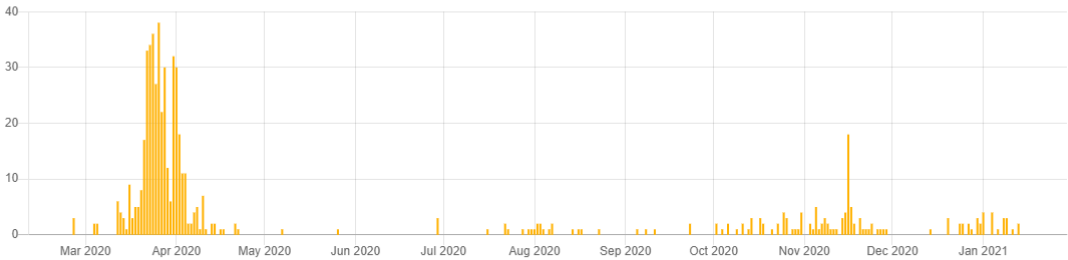


Active and Confirmed Positive COVID-19 cases in South Australia by residential Local Government Area (LGA). Colour shows sum of Confirmed Positive Cases. Size of circles shows sum of Active Cases. Persons that reside outside SA are excluded from this visual.

Daily cases - SA

Last case reported January 13

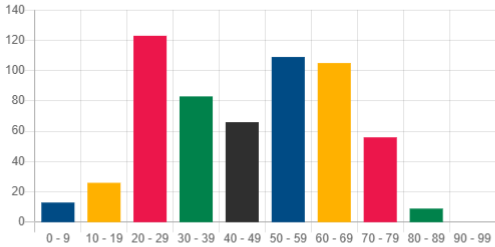
Data current at 0:00 14 January 2021



[Table version of Daily cases - SA graph](#)

Age

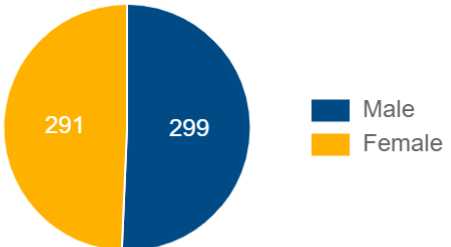
Data current at 0:00 14 January 2021



[Table version of Age graph](#)

Gender

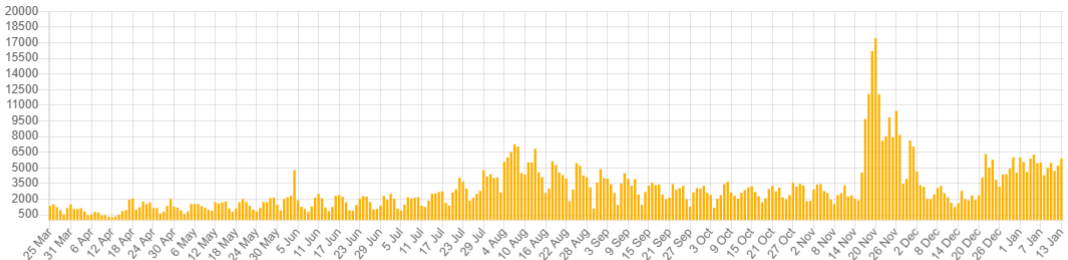
Data current at 0:00 14 January 2021



[Table version of Gender graph](#)

Daily laboratory tests

Data current at 0:00 14 January 2021



[Table version of Daily laboratory tests graph](#)

Resources



Links to useful resources

Power BI Training – [Learning | Microsoft Power BI](#)

What is data Visualisation- [What is Data Visualization? Definitions, Graph Types and How to Use Them \(klipfolio.com\)](#)

15 Statistics That Prove the Power of Data Visualization (csgsolutions.com) - [15 Statistics That Prove the Power of Data Visualization \(csgsolutions.com\)](#)

The Top Trends in Data Visualization for 2018 | CARTO Blog - [The Top Trends in Data Visualization for 2018 | CARTO Blog](#)