

Fundamentals of Data Wrangling

29 & 30 June 2022

Wong Hau Shian

School of Infocomm, Republic Polytechnic

• REC

OFFICIAL (CLOSED) \ NON-SENSITIVE

Administrative matters

Please change your log in screen name to resemble the registered name for the course



[Attendance link](#)

The instructor is **busy** taking
attendance ☺

Perform software installation (if haven't done so):

- Tableau Desktop

Set your microphone to muted when not speaking

Course Material Link: <https://tinyurl.com/2j437n5>



Trainer for the course



Wong Hau Shian

(wong_hau_shian@rp.edu.sg)



Learning Outcomes

Upon completion of this course, participants will have learned the following:

- Apply data cleansing and data transformation techniques
- Apply mathematical and statistical functions for deriving insights from data
- Create PivotTables and PivotCharts to calculate, summarise, analyze and visualise data
- Apply visualisation principles for creating effective visualisations
- Create and interpret visualisation charts
- Build interactive dashboards for monitoring key metrics
- Create data-driven stories for communicating key insights



Overview

Day 3	Morning	Data Visualisation Principles Install Tableau desktop
	Afternoon	Create and interpret Data Visualisation charts
Day 4	Morning	Build interactive Data Visualisation dashboard Tell Data-driven stories
	Afternoon	Quiz Mini Project & Presentation

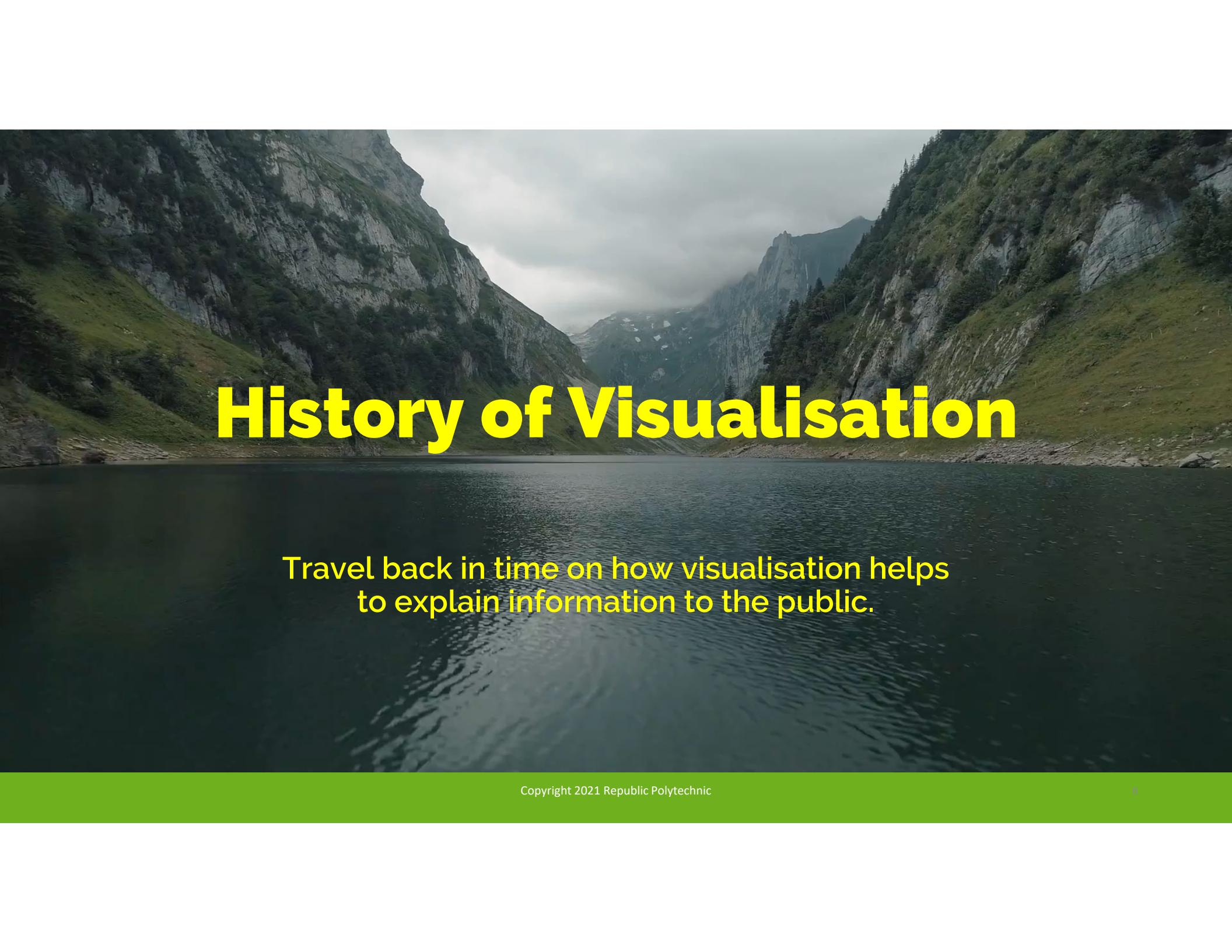


Outline for the Day 3

Time	Agenda
9.00am	Welcome and admin matters
9.15am – 10.30am	
10.30am – 10.45am	Break
10.45am – 12.30pm	
12.30pm – 1.30pm	Lunch
1.30pm – 3.15pm	
3.15pm – 3.30pm	Break
3.30pm – 4.45pm	
4.45pm – 5.00pm	Wrap up, Q&A

Recall the terminology

- Data
 - Either raw quantitative (numeric) or qualitative (non-numeric) collected to answer questions. Eg. Number of sales for each dish in the restaurant per month.
- Information
 - Organization of data presented in a meaningful context. Eg. The most popular and least popular dish in the restaurant every month.
- Insights
 - Gained by analyzing data and information to understand what is going on with a particular situation or phenomena. Eg, the reason behind why a dish is so popular in that month.
- Data Preparation
 - The process of cleaning and transforming raw data in a form suitable for processing, analysis and visualisation.
- Data Analytics
 - Statistical analysis of the collected data to reveals patterns, correlations, and cause-and-effect relationships between factors. The goal is to support data driven decision making.
- Data Visualisation
 - Both an art and a science. It is a graphical representation of data or information. The goal is to communicate the data/information clearly and efficiently to users in a visual manner.

The background image shows a wide, dark lake in the foreground, leading to a deep valley between two massive, rocky mountains. The mountains are covered in patches of green vegetation and exposed rock faces. The sky above is overcast with heavy, grey clouds.

History of Visualisation

Travel back in time on how visualisation helps
to explain information to the public.



Claudius Ptolemaeus

Almagest (Astronomy Treatise)



A copy of the Almagest from the 9th century, in Greek, on parchment

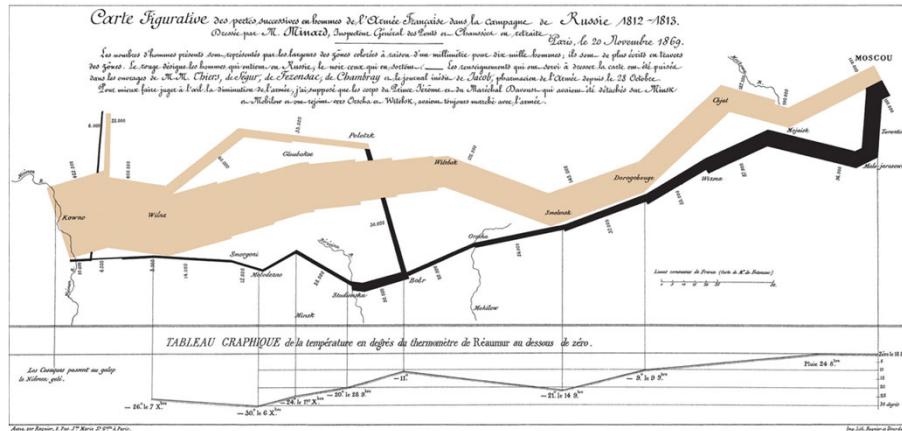
- Published the Almagest in the 2nd-century Greek-language mathematical and astronomical treatise on the motions of the stars and planetary paths.
 - Earliest preserved use of a table with detailed astronomical information
 - Much of medieval astronomy and geography were built on his ideas
- His world map, published as part of his treatise Geography, was the first to use longitudinal and latitudinal lines.

Source: <https://en.wikipedia.org/wiki/Almagest#:~:text=The%20Almagest%20%2F%CB%88%C3%A6l,170>
<https://www.newscientist.com/people/claudius-ptolemy/>.



Charles Minard

Map of Napoleon's Disastrous Russian Campaign of 1812



Source: https://en.wikipedia.org/wiki/Charles_Joseph_Minard

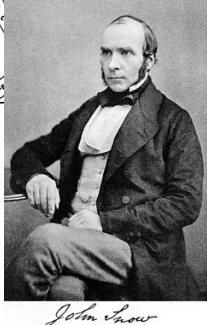


Source: <http://www.datavis.ca/papers/iebs.pdf>

- Charles Joseph Minard (1781 - 1870), a French civil engineer produced an array of graphics that combined many data points into a compelling visual story
- Most cited example: Charles Minard mapped Napoleon's invasion of Russia which depicted the advance (tan) of his army from the Polish-Russian border, and the disastrous retreat (black) of his army through Russia using 6 types of data in 2 dimensions:
 - Size of Napoleon's troops
 - Distance travelled
 - Temperature
 - Latitude & Longitude
 - Direction of travel
 - Location relative to specific dates



Dr John Snow Cholera Map of London (1854)



- In 1854, London was gripped by cholera; many thousands were to die in the ensuing epidemic. Most doctors at the time believed that the disease was caused by foul smelling mist "miasmas" a view contested by Dr John Snow who suspected that contaminated drinking water was the cause.
- Dr Snow drew a map of Soho marking each case of cholera with a black dot. He also marked each public water pump in the area with an "X". Dr Snow was able to track the origin of outbreak in Soho to Broad Street.

Source: <http://johnsnowbicentenary.lshtm.ac.uk/about-john-snow/>

OFFICIAL (CLOSED) \ NON-SENSITIVE

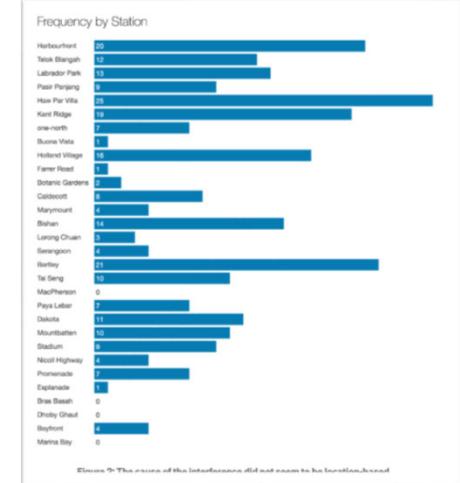
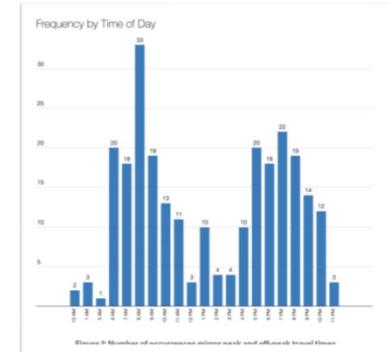


Visualisation Today



How the circle line rogue train was caught with data (2016)

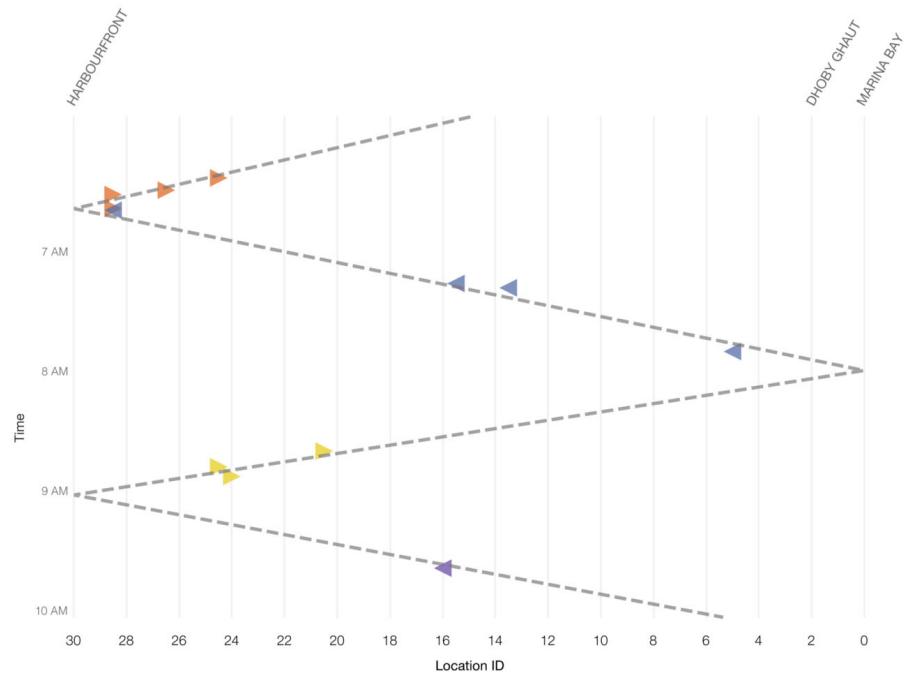
- Background:
 - Singapore's MRT circle line was hit by a spate of mysterious disruption.
 - Govtech was given the tasks to solve the problem
- Data collected:
 - Date and time of each incident
 - Location of incident
 - IDs of trains involved
 - Direction of trains





How the circle line rogue train was caught with data cont'd

- GovtTech drew a Marey Chart to visualize time, location and direction all on the same plot
 - A pattern was observed
 - When a train was hit by interference, another train behind, moving in the same direction got hit soon after
 - Found the culprit
 - PV46 is the rogue train





What is Data Visualisation?

- Graphical representation of information and data
- Uses visual elements such as charts, graphs and maps to:
 - See and understand trends
 - Identify outliers
 - Visualise patterns in the data
- Essential tool to analyse massive amounts of data
 - Big Data



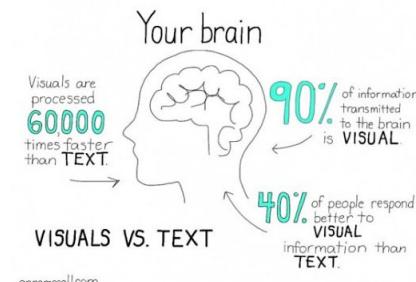
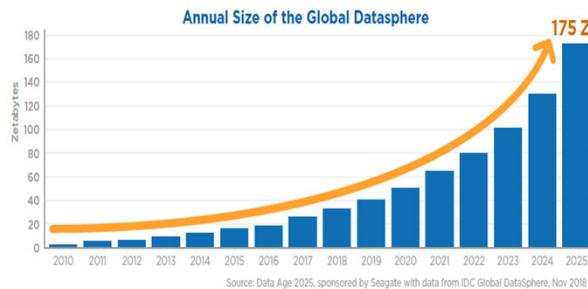
“Giving Life to Data”





Impetus for Data Visualisation

- **Data is growing at an astronomical rate**
 - How to manage and present the data , without losing relevant information, becomes a big challenge.
- **Human Minds work well with Visuals**
 - Reading charts is easier than going through pages of rows and columns of numbers
 - Patterns, trends and correlations undetected in text-based data can be easily exposed and recognized when presented in a visual format, helping decision-makers to grasp the meaning and insights.





Why is Visualisation Important?

Example: Anscombe's quartet comprises four data sets that have identical simple descriptive statistics

Anscombe's quartet									
I		II		III		IV			
x	y	x	y	x	y	x	y		
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58		
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76		
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71		
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84		
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47		
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04		
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25		
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50		
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56		
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91		
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89		

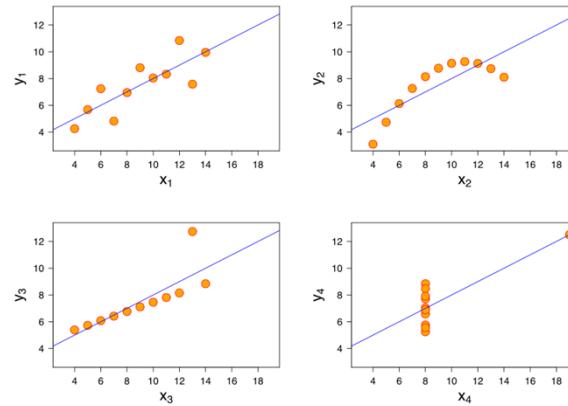
So, Why is Visualisation
Important?

Answer: Graphs reveal data
that statistics may not

Summary statistics are all identical for all the four datasets:

Property	Value
Mean of x in each case	9 (exact)
Sample variance of x in each case	11 (exact)
Mean of y in each case	7.50 (to 2 decimal places)
Sample variance of y in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y in each case	0.816 (to 3 decimal places)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)

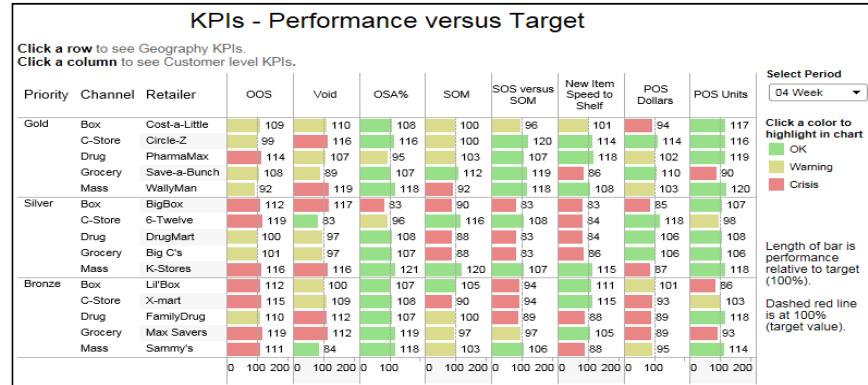
However, the four datasets vary considerably when graphed even though their statistics are identical





Benefits of Visualisation

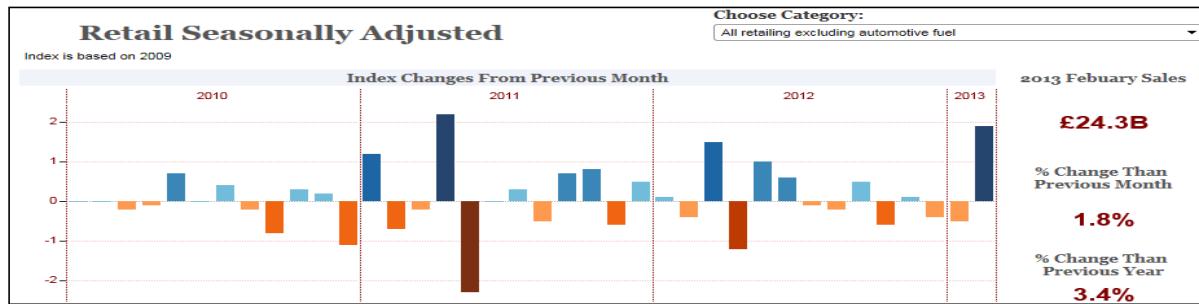
This dashboard shows the overall health of an organization in a single view while allowing for customized views for different levels of management.



- **Visualize relationships and patterns between operational and business activities**
 - Enable users to more effectively see connections between operating conditions and business performance from multi-dimensional data sets
- **Absorb information in new and more constructive ways**
 - Provide effective ways to interpret data through the use of heat maps, fever charts, and other rich graphical representations



Benefits of Visualisation



- **Identify and act on emerging trends & anomalies faster**
 - Able to grasp shifts in data behaviors and various conditions across multiple data sets quickly
 - **Manipulate and interact directly with data**
 - Enable users to interact with data (e.g. drill down, filter, prompt, pivot) through visualisation tools
 - **Powerful tool to support story-telling**
 - Support data-driven storytelling through the graphical depiction of statistical information

OFFICIAL (CLOSED) \ NON-SENSITIVE

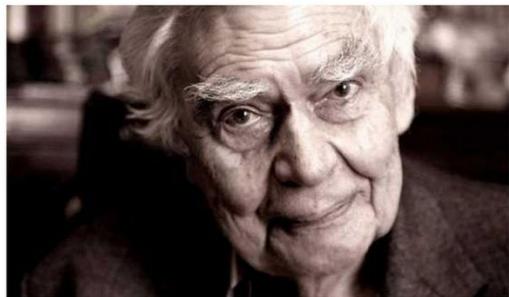


The art of data visualisation

How human read and view things



Human Visualisation and Perception



Gregory's Theory of
Constructive Perception

- Past knowledge and experience is the most important thing in making sense of what is around us.
- Our perception works because our brain makes reasonable guesses about what we see on the basis of what it is most likely to be – these are known as **perceptual hypotheses**.
- Visual illusions provide evidence to support this theory

Psychologist Richard Gregory (Bottom up, Top-down Processing)

- 90% of information reaching the eye is lost by the time it reaches the brain.
- The brain has to guess what a person sees based on expectations, belief and past experiences.

➔ It is important to understand how the visual perception works in human in order to design effective visuals to communicate information effectively.



A Young or An Old Woman?



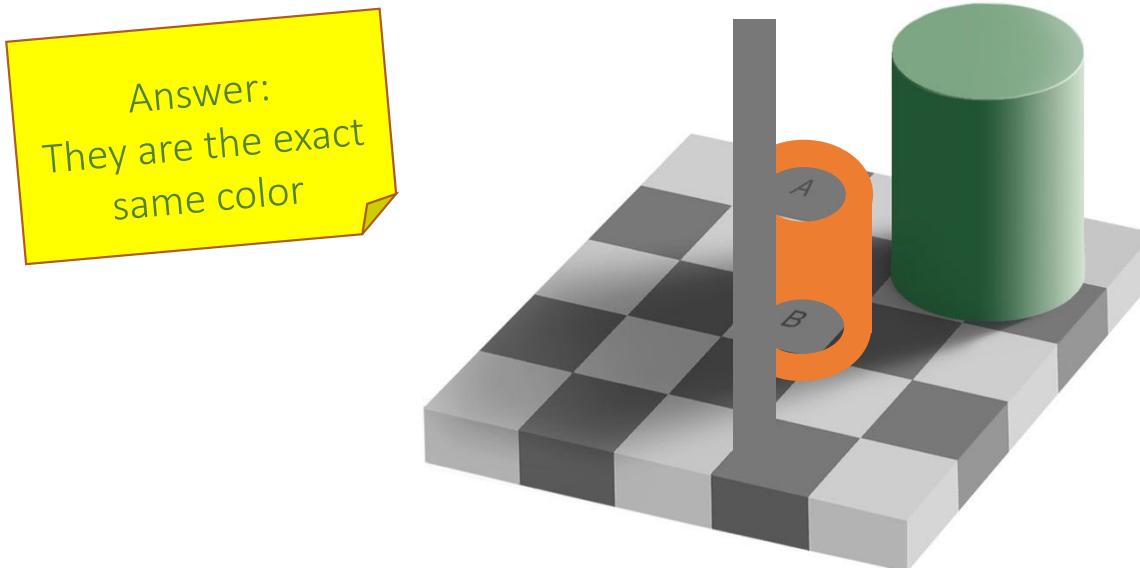
Credit: Wikimedia Commons

Source: <https://www.thesun.co.uk/news/7307450/optical-illusion-young-or-old-woman-depends-on-age/>



Do We Really Perceive Reality?

- Edward H. Adelson Checker shadow illusion.
- Based on the diagram below, which square is darker, square A or square B?



- Watch this video if you are not convinced: <https://youtu.be/zgSen1HTu5o>

Key Message

Keep in mind the intended audiences of your Data Visualisation as the viz is subject to interpretation

Let's do a simple exercise to visualise a given set of raw data.



How many times does the digit 7 appear?

Raw Data

5	2	8	3	6	1	9	3	6	2	5	3	7	4	3	8	3
8	5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9
1	0	2	7	5	2	8	3	6	1	6	2	9	3	8	3	8
5	8	4	7	2	0	3	7	3	5	4	7	1	8	2	0	1
2	5	3	6	4	3	9	1	0	8	9	5	7	3	4	5	3
2	7	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	7	2	0	3	7	3	5	4	7	1	8	2	0	1	9	6
2	1	4	4	3	9	3	6	5	2	4	9	1	0	2	7	5
2	8	3	6	1	6	2	9	3	8	3	8	5	8	4	7	2
0	3	7	3	5	4	7	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	7	3	4	5	3	2	7	5	2	8
3	6	1	6	2	4	6	2	7	5	9	1	5	2	6	3	6



We could colour it

5	2	8	3	6	1	9	3	6	2	5	3	7	4	3	8	3
8	5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9
1	0	2	7	5	2	8	3	6	1	6	2	9	3	8	3	8
5	8	4	7	2	0	3	7	3	5	4	7	1	8	2	0	1
2	5	3	6	4	3	9	1	0	8	9	5	7	3	4	5	3
2	7	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	7	2	0	3	7	3	5	4	7	1	8	2	0	1	9	6
2	1	4	4	3	9	3	6	5	2	4	9	1	0	2	7	5
2	8	3	6	1	6	2	9	3	8	3	8	5	8	4	7	2
0	3	7	3	5	4	7	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	7	3	4	5	3	2	7	5	2	8
3	6	1	6	2	4	6	2	7	5	9	1	5	2	6	3	6



We could colour and bold it

5	2	8	3	6	1	9	3	6	2	5	3	7	4	3	8	3
8	5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9
1	0	2	7	5	2	8	3	6	1	6	2	9	3	8	3	8
5	8	4	7	2	0	3	7	3	5	4	7	1	8	2	0	1
2	5	3	6	4	3	9	1	0	8	9	5	7	3	4	5	3
2	7	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	7	2	0	3	7	3	5	4	7	1	8	2	0	1	9	6
2	1	4	4	3	9	3	6	5	2	4	9	1	0	2	7	5
2	8	3	6	1	6	2	9	3	8	3	8	5	8	4	7	2
0	3	7	3	5	4	7	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	7	3	4	5	3	2	7	5	2	8
3	6	1	6	2	4	6	2	7	5	9	1	5	2	6	3	6



We could colour, bold and underline it.

5	2	8	3	6	1	9	3	6	2	5	3	7	4	3	8	3
8	5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9
1	0	2	7	5	2	8	3	6	1	6	2	9	3	8	3	8
5	8	4	7	2	0	3	7	3	5	4	7	1	8	2	0	1
2	5	3	6	4	3	9	1	0	8	9	5	7	3	4	5	3
2	7	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	7	2	0	3	7	3	5	4	7	1	8	2	0	1	9	6
2	1	4	4	3	9	3	6	5	2	4	9	1	0	2	7	5
2	8	3	6	1	6	2	9	3	8	3	8	5	8	4	7	2
0	3	7	3	5	4	7	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	7	3	4	5	3	2	7	5	2	8
3	6	1	6	2	4	6	2	7	5	9	1	5	2	6	3	6



What have we done thus far?

We make it easier to see the digit seven embedded in the given raw data

- ✓ Colour
 - ✓ Colour & Bold
 - ✓ Colour, Bold and Underline

7 7 7 7 7 7 7 7 7 7 7 7 7 7 7

7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7

7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7

7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7

7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7

77777777777777777777

More can be done!



We could re-order it

7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
5	2	8	3	6	1	9	3	6	2	5	3	4	3	8	3	8		
5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9	1		
0	2	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8		
4	2	0	3	3	5	4	1	8	2	0	1	2	5	3	6	4		
3	9	1	0	8	9	5	3	4	5	3	2	5	2	8	3	6		
1	6	2	9	3	8	3	8	5	8	4	2	0	3	3	5	4		
1	8	2	0	1	9	6	2	1	4	4	3	9	3	6	5	2		
4	9	1	0	2	5	2	8	3	6	1	6	2	9	3	8	3		
8	5	4	8	2	0	3	3	5	4	1	8	2	0	1	2	5		
3	6	4	3	9	1	0	8	9	5	3	4	5	3	2	5	2		
8	3	6	1	6	2	4	6	2	5	9	1	5	2	6	3	6		



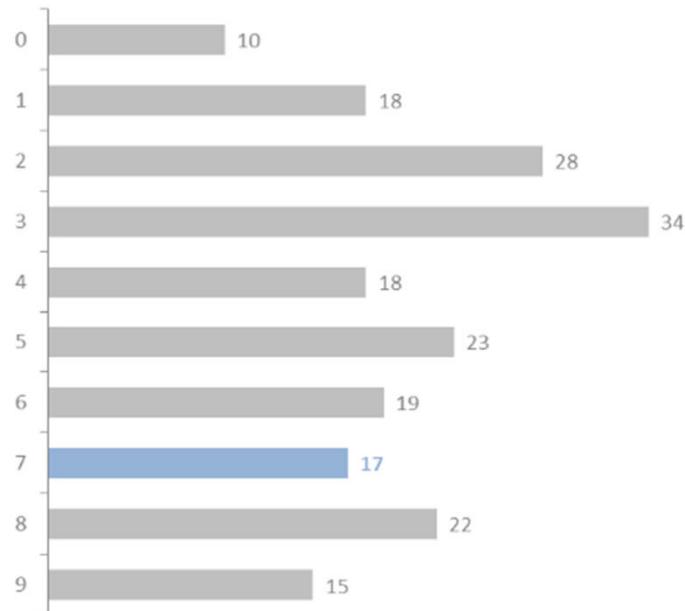
We could de-emphasizing the non-relevant data

7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
5	2	8	3	6	1	9	3	6	2	5	3	4	3	8	3	8		
5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9	1		
0	2	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8		
4	2	0	3	3	5	4	1	8	2	0	1	2	5	3	6	4		
3	9	1	0	8	9	5	3	4	5	3	2	5	2	8	3	6		
1	6	2	9	3	8	3	8	5	8	4	2	0	3	3	5	4		
1	8	2	0	1	9	6	2	1	4	4	3	9	3	6	5	2		
4	9	1	0	2	5	2	8	3	6	1	6	2	9	3	8	3		
8	5	4	8	2	0	3	3	5	4	1	8	2	0	1	2	5		
3	6	4	3	9	1	0	8	9	5	3	4	5	3	2	5	2		
8	3	6	1	6	2	4	6	2	5	9	1	5	2	6	3	6		

We could count the occurrences of each number and present in a bar chart



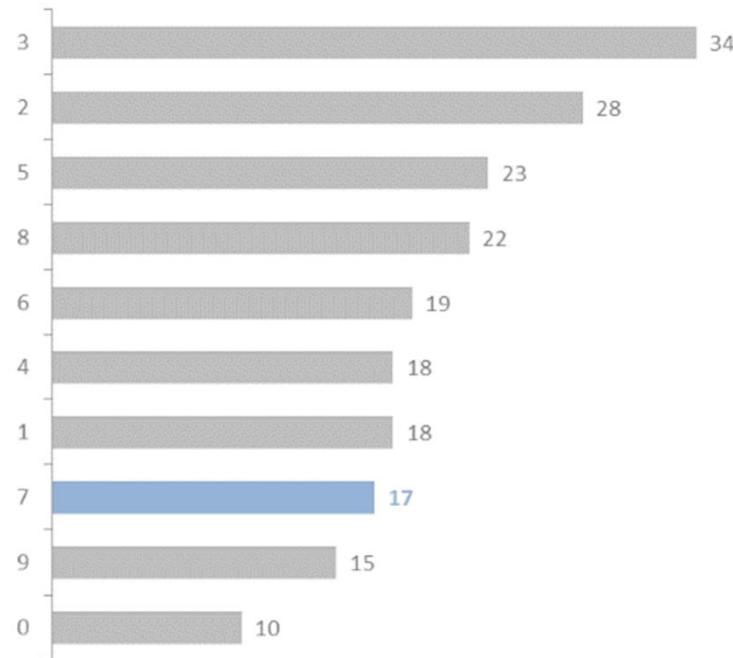
Number of times digit 7 appears: 17



7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
5 2 8 3 6 1 9 3 6 2 5 3 4 3 8 3 8
5 8 9 6 2 1 4 4 3 9 3 6 5 2 4 9 1
0 2 5 2 8 3 6 1 6 2 9 3 8 3 8 5 8
4 2 0 3 3 5 4 1 8 2 0 1 2 5 3 6 4
3 9 1 0 8 9 5 3 4 5 3 2 5 2 8 3 6
1 6 2 9 3 8 3 8 5 8 4 2 0 3 3 5 4
1 8 2 0 1 9 6 2 1 4 4 3 9 3 6 5 2
4 9 1 0 2 5 2 8 3 6 1 6 2 9 3 8 3
8 5 4 8 2 0 3 3 5 4 1 8 2 0 1 2 5
3 6 4 3 9 1 0 8 9 5 3 4 5 3 2 5 2
8 3 6 1 6 2 4 6 2 5 9 1 5 2 6 3 6



Finally, we could reorder the bar chart



7 7 7 7 7 7 7 7 7 7 7 7 7 7
5 2 8 3 6 1 9 3 6 2 5 3 4 3 8 3 8
5 8 9 6 2 1 4 4 3 9 3 6 5 2 4 9 1
0 2 5 2 8 3 6 1 6 2 9 3 8 3 8 5 8
4 2 0 3 3 5 4 1 8 2 0 1 2 5 3 6 4
3 9 1 0 8 9 5 3 4 5 3 2 5 2 8 3 6
1 6 2 9 3 8 3 8 5 8 4 2 0 3 3 5 4
1 8 2 0 1 9 6 2 1 4 4 3 9 3 6 5 2
4 9 1 0 2 5 2 8 3 6 1 6 2 9 3 8 3
8 5 4 8 2 0 3 3 5 4 1 8 2 0 1 2 5
3 6 4 3 9 1 0 8 9 5 3 4 5 3 2 5 2
8 3 6 1 6 2 4 6 2 5 9 1 5 2 6 3 6

of times digit 7 appears: 17



Understanding Human Perception

- Visualisations are subject to interpretation
- Therefore, it is important to understand some aspects of human perception in order to design effective visuals to communicate information effectively.
- Knowledge in the following can help us to create good visualisations.
 1. Preattentive Attributes
 2. Analytical Patterns
 3. Gestalt Principles





1. Preattentive Attributes

- **Preattentive** attributes, such as color, size and orientation, help to capture human's attention.
- Preattentive attributes are visual properties that human notice **without using conscious effort** to do so. These attributes are:
 - Perceived in less than 10 milliseconds
 - Unconsciously processed
- If we can harness preattentive attributes, we can craft intuitive and easily understood visualisations.

56789

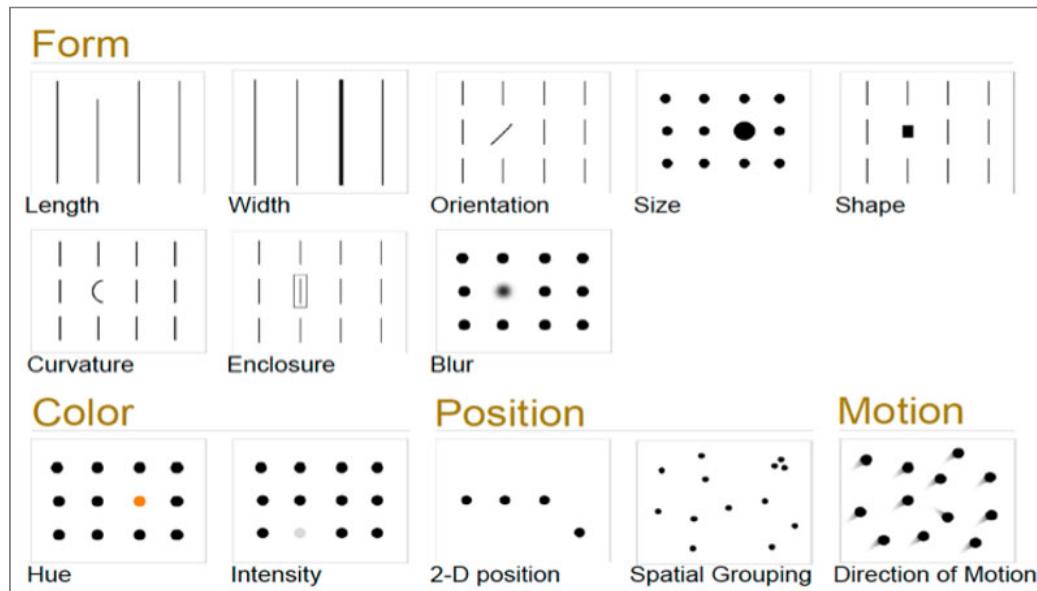
56⁷89 color

56⁷89 size

56⁷89 orientation



Preattentive Attributes of Visual Perception



- Only Spatial and Length can be used to communicate quantitative information with precision.
- The remaining attributes enable us to communicate:
 - Categorical information
 - Relational information
 - Quantitative information
- Combinations of attributes allow us to communicate several dimensions in one graphic.



Preattentive Attributes – Visual Variables

- Research has shown that certain visual attributes can more accurately represent certain data to the human brain
- Cleveland and McGill* studied what people are able to decode most accurately and ranked them in the following list.

	Quantitative	Ordinal	Nominal
More Accurate ↑	Position •••	Position •••	Position •••
	Length ==	Density ••••	Hue ••••
	Angle <	Saturation •••••	Density ••••
	Slope //	Hue ••••	Saturation ••••
	Area •••	Length ==	Shape •▲■
	Density ••••	Angle <	Length ==
	Saturation •••••	Slope //	Angle <
	Hue ••••	Area •••	Slope //
↓ Less Accurate	Shape •▲■	Shape •▲■	Area •••

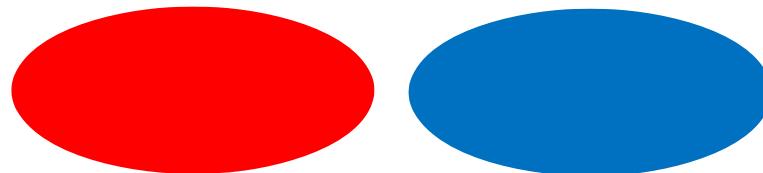
* William S. Cleveland and Robert McGill's paper *Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods*



Preattentive Attributes – Visual Variables

	Quantitative	Ordinal	Nominal
More Accurate			
Position	••	Position	••
Length	==	Density	•••
Angle	<	Saturation	•••
Slope	/\	Hue	•••
Area	••	Length	==
Density	•••	Angle	<
Saturation	•••	Slope	/\
Hue	•••	Area	••
Shape	•▲■	Shape	•▲■
Less Accurate			

- **Example:** Let's test using hue to represent Quantitative data.
Can you tell which color is greater than the other?



- Hue is less accurate for Quantitative data, but more accurate for Nominal data (e.g. Female, Male)



2. Analytical Patterns

- We can use Preattentive Attributes to form **Analytical Patterns**.
- Analytical Patterns provide a visual language to communicate information visually.
- From the pattern, we may use it as a model for analysis, such as prediction (eg linear regression model)

Pattern	Example	Pattern	Example
High, low and in between		Non-intersecting and intersecting	
Going up, going down and remaining flat		Symmetrical and skewed	
Steep and gradual		Wide and narrow	
Steady and fluctuating		Clusters and gaps	
Random and repeating		Tightly and loosely distributed	
Straight and curved		Normal and abnormal	

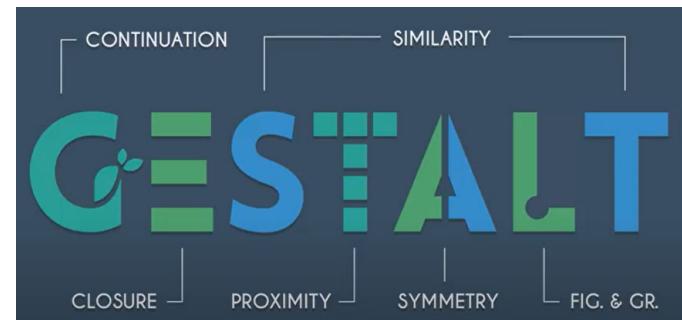
Source: <http://is2000.weebly.com/uploads/8/6/5/9/8659576/principles-of-data-visualization.pdf>



3. Gestalt Principles

Gestalt Principles of Perception

- A set of visual perception theories developed by German psychologist to describe how humans organize and perceive groups of objects.
- Our brains organise and group visual elements into **groups** or **unified wholes** when certain principles are applied.
 - Objects will be perceived in their simplest form
 - Humans naturally follow lines or curves
 - The human mind will attempt to fill in "missing" details
- We can use these principles to **highlight** patterns that are important, and downplay other patterns.
- Focus on the following 6 Gestalt Principles:
 - Continuity
 - Similarity
 - Proximity
 - Symmetry
 - Closure
 - Figure & Ground





Gestalt Law of Continuity

Continuity

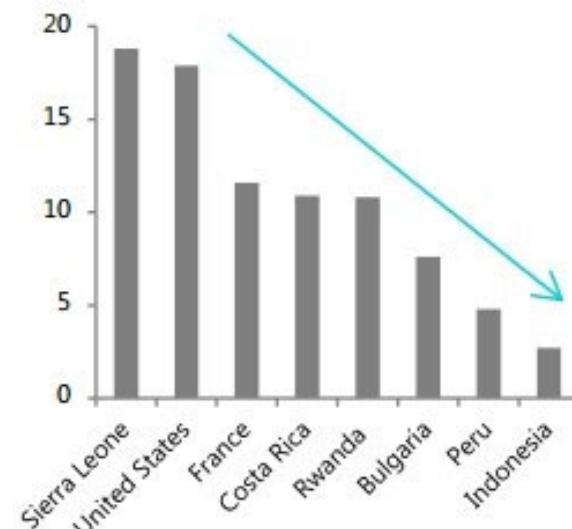
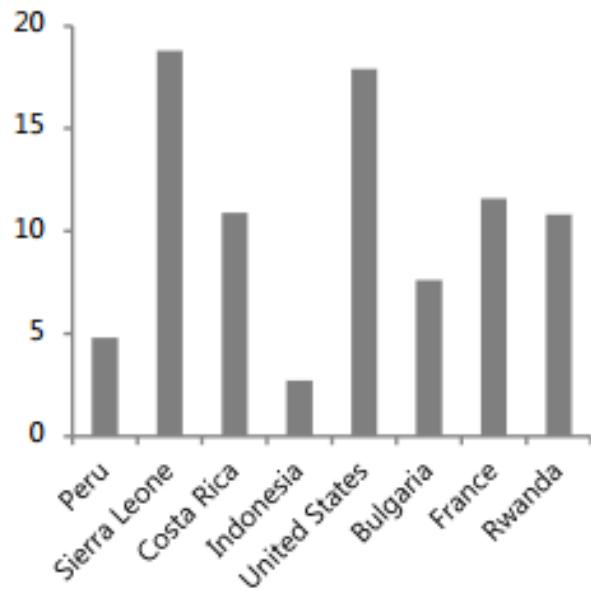
- Continuation occurs when the eye is compelled to move through one object and continue to another object.
- Continuity occurs in the example below, because the viewer's eye will naturally follow a line or curve. The smooth flowing crossbar of the "H" leads the eye directly to the maple leaf.



Source:<https://www.youtube.com/watch?v=FryaH599ec0>



Gestalt Law of Continuity





Gestalt Law of Similarity

Similarity

- Similarity occurs when objects look similar to one another. People often perceive them as a group or pattern.
- The example below (containing 11 distinct objects) appears as a single unit because all of the shapes have similarity. Unity occurs because the triangular shapes at the bottom of the eagle symbol look similar to the shapes that form the sunburst.

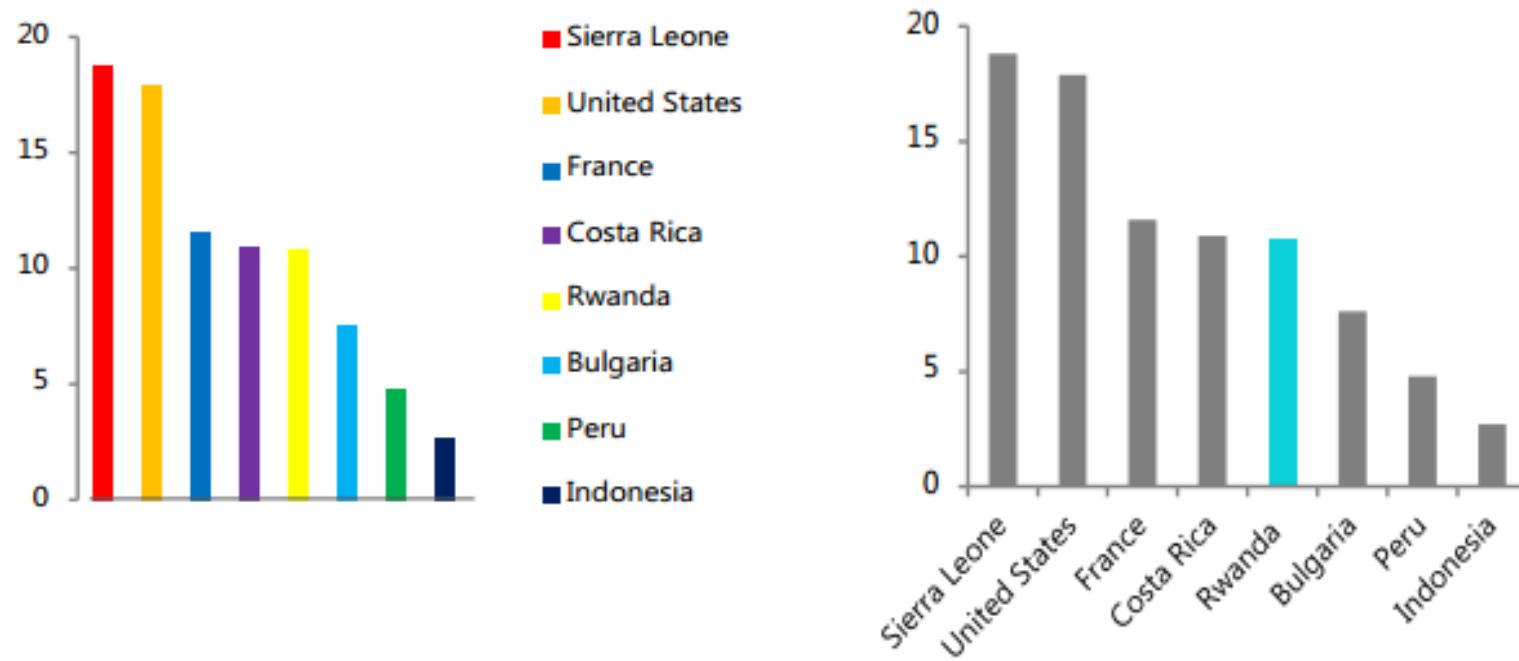


- When similarity occurs, an object can be emphasised if it is dissimilar to the others. This is called **anomaly**.
- The figure on the far right becomes a focal point because it is dissimilar to the other shapes.





Gestalt Law of Similarity

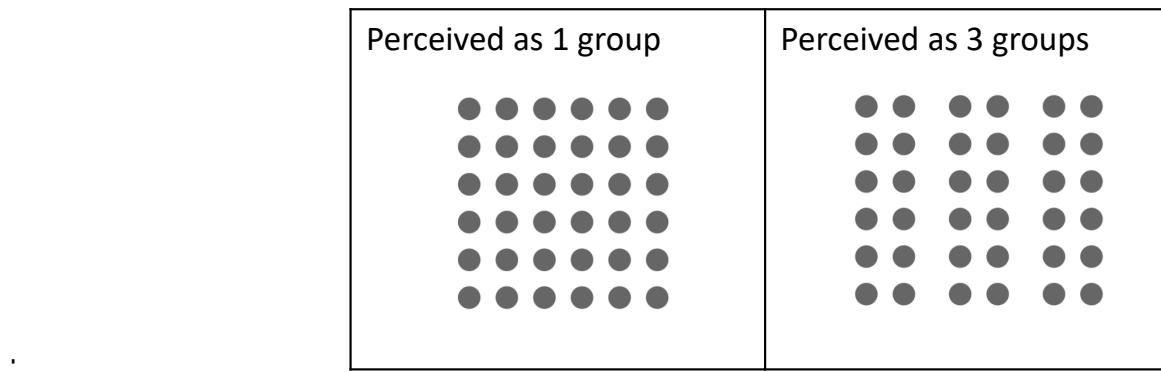




Gestalt Law of Proximity

Proximity

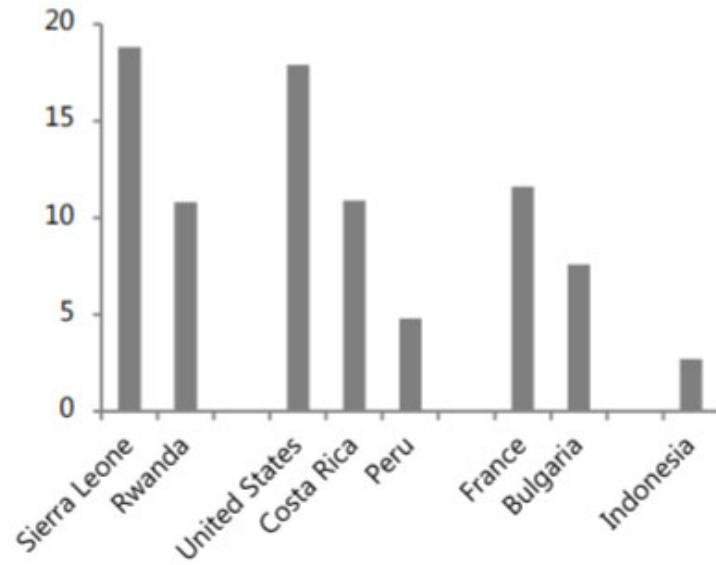
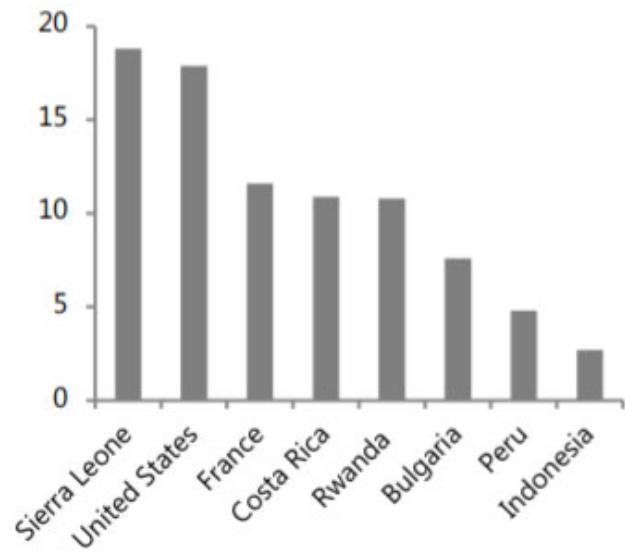
- Proximity occurs when elements are placed close together. They tend to be perceived as a group or related to one another.



Source: <https://www.toptal.com/designers/ui/gestalt-principles-of-design>



Gestalt Law of Proximity





Gestalt Law of Symmetry

Symmetry

- Symmetry occurs when symmetrical elements tend to be perceived as a unified group
- Conversely, a lack of symmetry tends to be perceived as a lack of relationship



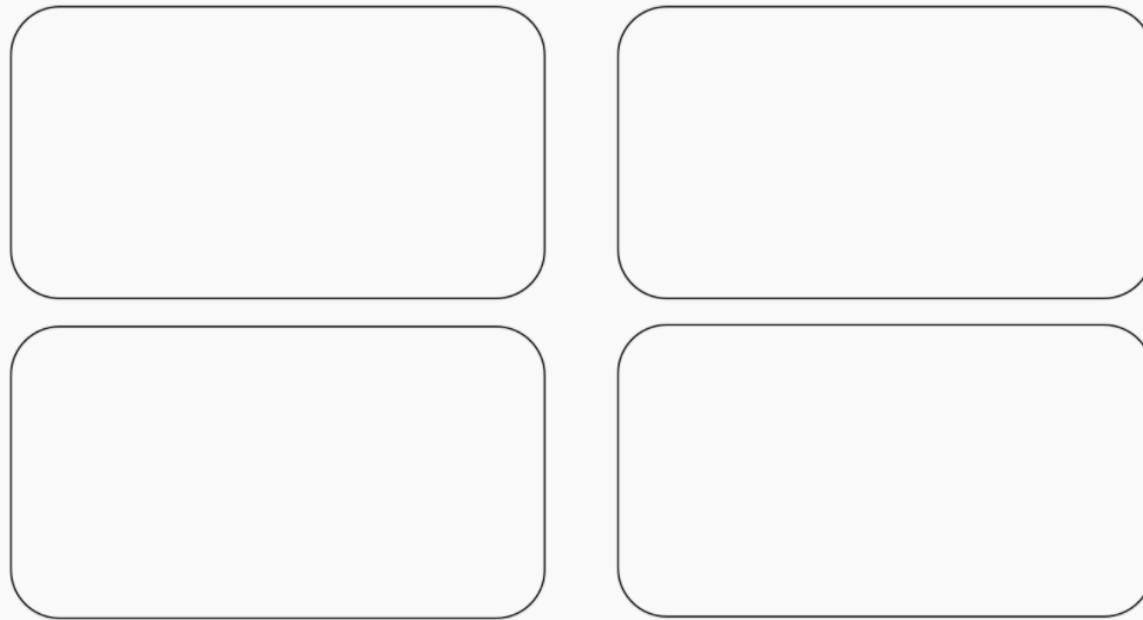
Source:<https://www.youtube.com/watch?v=FryaH599ec0>



Gestalt Law of Symmetry

Dashboard Title

Subtext explaining the different elements of the dashboard as well as other info that you might want to know as an end user. It is helpful to provide a little context here to help users understand what they are looking at. Have I typed enough in here yet?



Source: <https://www.tessellationtech.io/dashboard-design-essentials-dashboard-layout-and-formatting/>

Copyright 2021 Republic Polytechnic



Gestalt Law of Closure

Closure

- Closure occurs when an object is incomplete or a space is not completely enclosed. If enough of the shape is indicated, people perceive the whole by filling in the missing information.
- Although the panda above is not complete, enough is present for the eye to complete the shape. When the viewer's perception completes a shape, closure occurs.

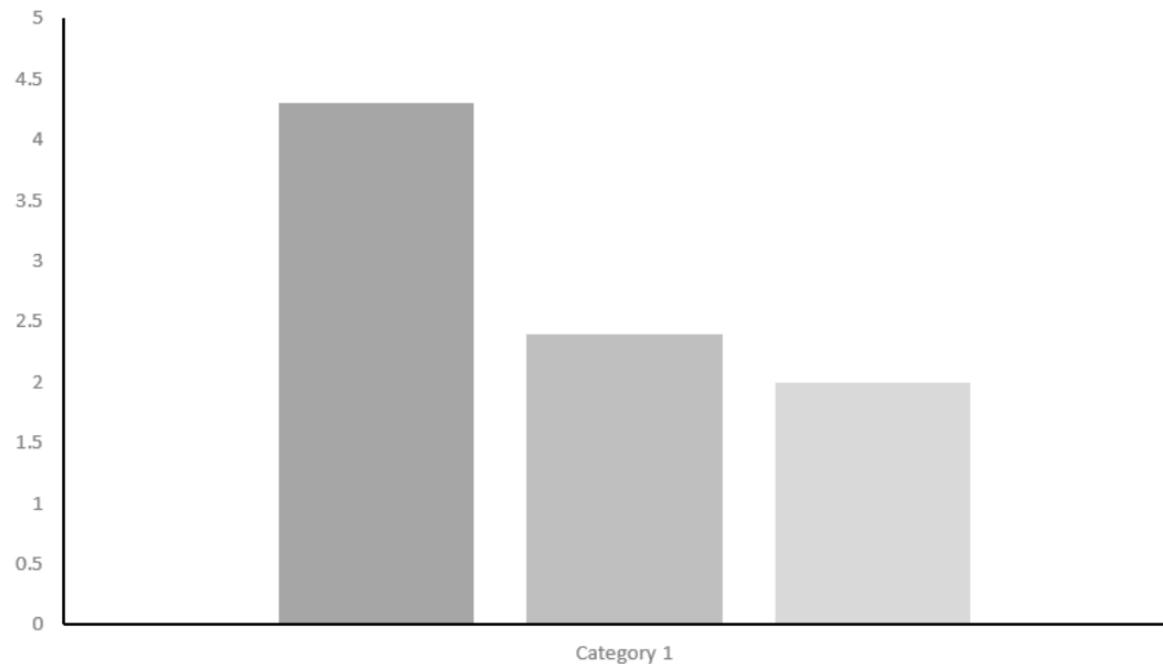


Sources: <https://www.youtube.com/watch?v=FryaH599ec0>

<https://vizzendata.com/2020/07/06/utilizing-gestalt-principles-to-improve-your-data-visualization-design/>



Gestalt Law of Closure



Source:

<https://www.topcoder.com/blog/gestalt-principles-for-data-visualization/#:~:text=When%20faced%20with%20ambiguous%20objects,we%20can%20reasonably%20do%20so.>



Gestalt Law of Figure & Ground

Figure & Ground

- The eye differentiates an object from its surrounding area. A form, silhouette, or shape is naturally perceived as **figure** (object), while the surrounding area is perceived as **ground** (background).
- Balancing figure and ground can make the perceived image more clear. Using unusual figure/ground relationships can add interest and subtlety to an image.
- In the following image, the figure and ground relationships change as the eye perceives the form of a tree or that of an ape and lion facing each other.

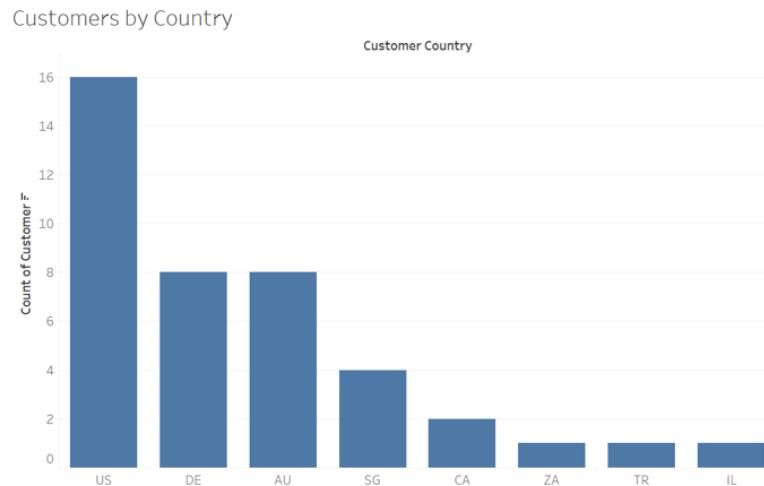


Source:<https://www.youtube.com/watch?v=FryaH599ec0>

Copyright 2021 Republic Polytechnic



Gestalt Law of Figure & Ground



Source:<https://www.youtube.com/watch?v=FryaH599ec0>

Copyright 2021 Republic Polytechnic



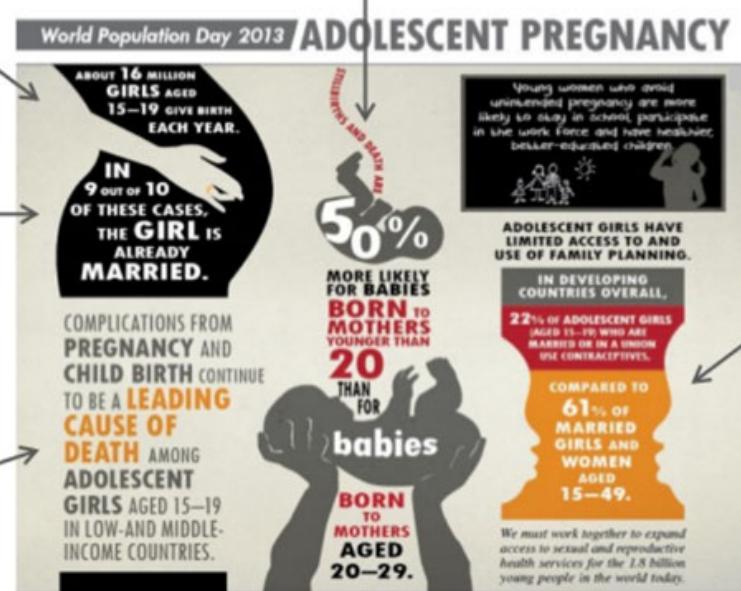
Putting What We Have Learnt Together

Gestalt law of closure allows you to recognize this incomplete image as a pregnant woman

Preattentive attributes for orientation and curvature cause this “umbilical chord” of text to catch your eye. Gestalt law of continuity brings your eye down the chord and through the center column of content.

Preattentive attribute for size allows some of these words and figures to stand out

preattentive attribute for hue cause this orange text to pop and grab your attention



Gestalt law of figure/ground: because it is colorful and packed with text, the center of this image is the figure even though the face outlines provide a dynamic ground.

For the rest of the graphic the background is muted to allow content to stand out.



Key points covered

Data Visualisation is the graphical representation of data and information.

The main purpose is to provide insights in a visually efficient and effective manner for the purpose of addressing pertinent questions.

Data Visualisation can be subject to interpretation

To allow for clearer and effective representation, leverage:

1. Pre-attentive Attributes
2. Analytical Patterns
3. Gestalt Principles

OFFICIAL (CLOSED) \ NON-SENSITIVE



Let's take a break



OFFICIAL (CLOSED) \ NON-SENSITIVE



The art of Data Visualisation

Graphical Excellence & General Guidelines

Graphical Excellence



Edward Tufte, a data visualisation expert, says “Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.” This trait of visualisation is what makes them vital to businesses.



Elements of Graphical Excellence

We will cover the following three elements of Graphical Excellence:

1. Data:Ink ratio & Chartjunk
2. Graphical Integrity
3. Data Density



1. Data:Ink ratio and Chartjunk

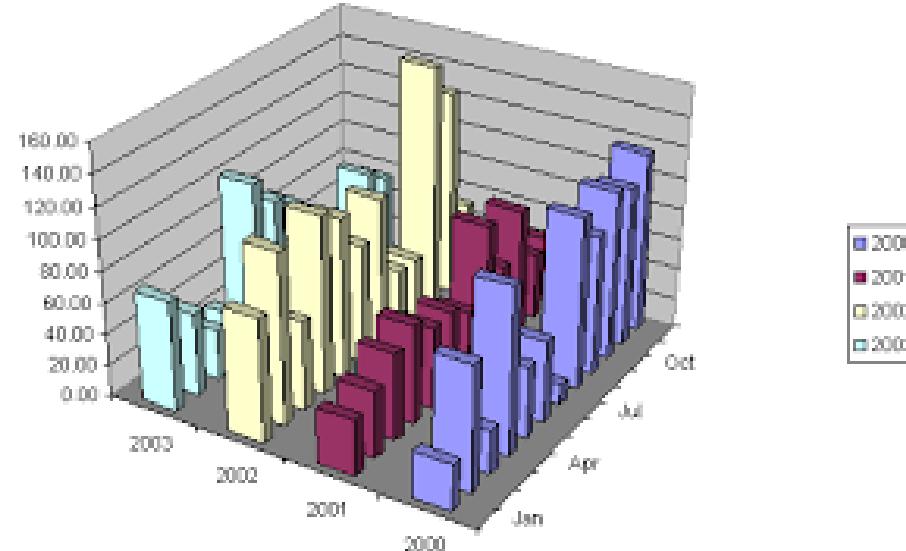
- Graphical Excellence is high when Data:Ink ratio is high.
- Good graphical representations maximize the Data:Ink ratio and erase "non-data-ink", "redundant data-ink" and chartjunk.
 - Data ink means the ink on a graph used to represent data
 - Chartjunk means excessive decorative or unnecessary use of graphical effect that seems to demonstrate the ability of designer rather than displaying the data
 - To minimize Chartjunk, we must remove irrelevant and superfluous (unnecessary) graphics



Data:Ink ratio and Chartjunk

(Example of A Bad Chart – Low Graphical Excellence)

- Redundant Data Ink
 - Is the gray background necessary? Is it required to show the data?
- Avoid Chartjunk - Superfluous graphics
 - Are the 3D bars superfluous graphics or helping showing the data more effectively?

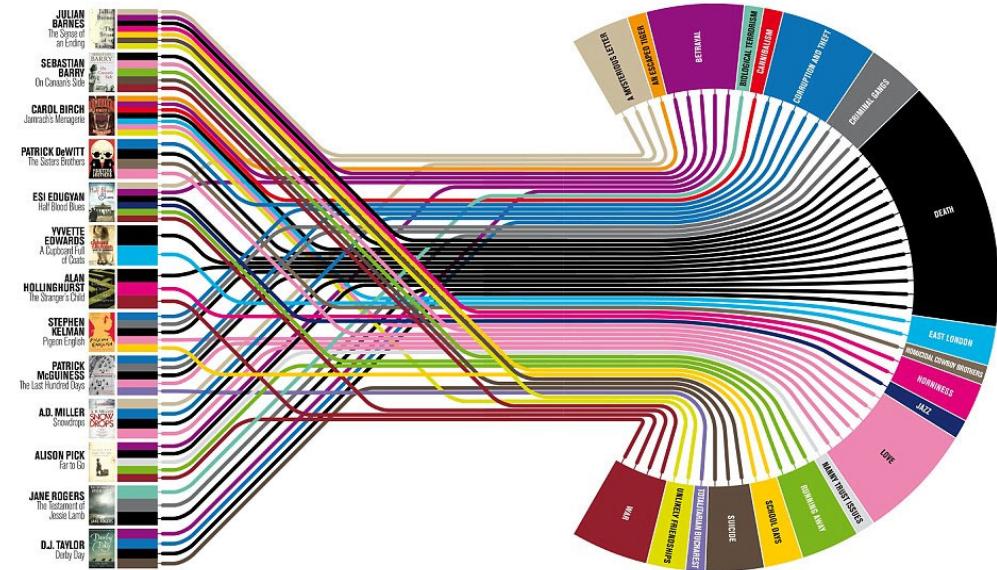




Data:Ink ratio and Chartjunk

(Example of a Bad Chart – Low Graphical Excellence)

- Redundant Data Ink
- Superfluous graphics



Plot lines

What makes a prize-winning novel? As Julian Barnes wins the Booker Prize, Delayed Gratification's Johanna Kamradt charts the themes of this year's longlisters.



2. Graphical Integrity

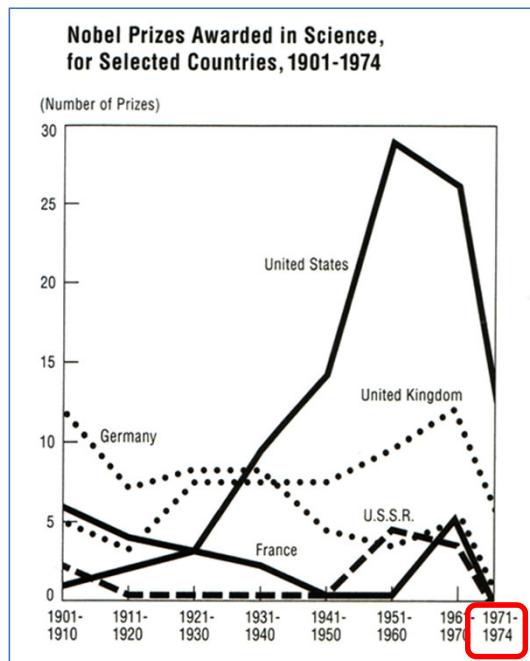
- Graphical Excellence is high when Graphical Integrity is high.
- Graphical Integrity requires an absence of misrepresentation, persuasion, masking, or lies. Specifically:
 - Consistency in baseline and labels
 - Consistency in compared scales
 - Partial data to be highlighted
 - Compared data to be normalised
 - Context to be represented



Low Graphical Integrity

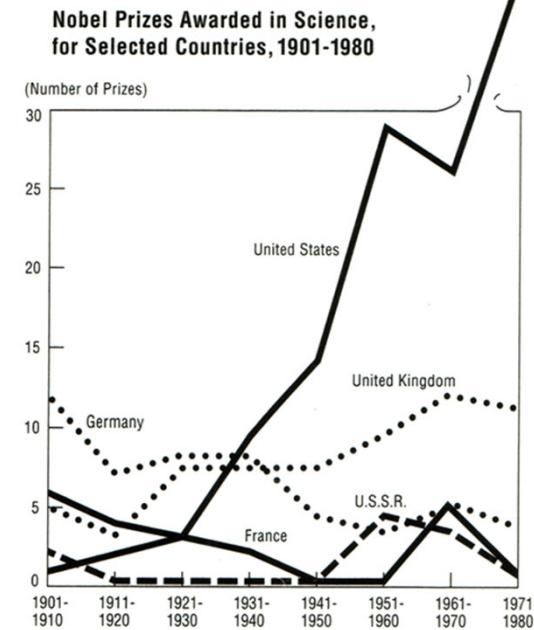
(Example of a Bad Chart – Low Graphical Excellence)

- Partial data not highlighted
 - The data interval was 10 years, except the last data point but this was not highlighted
- Not consistent in scales



Copyright 2021 Republic Polytechnic

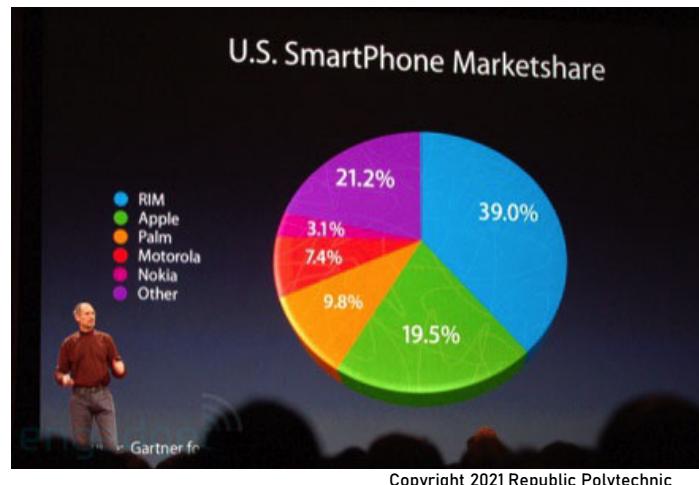
Using a consistent scale, the chart should be as below:





Speaking of the Pie Chart...

- There's a multitude of articles written against using pie chart
 - Pie Charts are poor at communicating data
 - Pie Charts take up more space and are harder to read than the alternatives
 - The brain's not very good at comparing the size of angles and because there's no scale, reading accurate values can be challenging
 - When there are more segments and colors, the problem gets worse. Labels can be hard to fit, especially for smaller segments, so often legends are required.
- With the right perspective you can even trick people, like what Apple did:

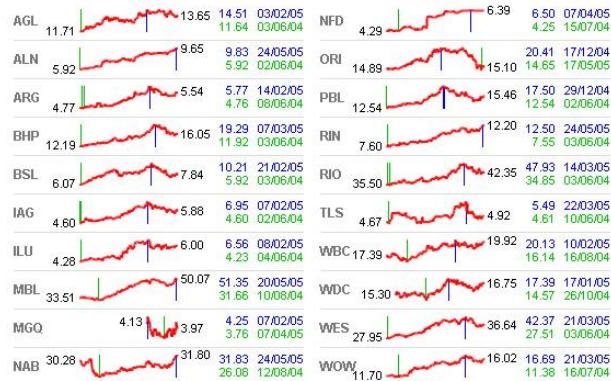


Perspective Trick
makes 19.5% (for
Apple) look BIGGER
than 21.2% (for Other)



3. Data Density

- Graphical Excellence is high when Data Density is high
- The data density of a graph is the proportion of the total size of the graph that is dedicated to display data
- Maximize data density and the size of the data matrix within reason
 - One way of achieving this is through the Shrink Principle
 - Most graphs can be shrunk way down without losing legibility or information



Source: <https://thedoublethink.com/tuftes-principles-for-visualizing-quantitative-information/#:~:text=One%20way%20of%20achieving%20this,graph%20repeated%20in%20one%20visual>



Guidelines to Data Visualisation

The following are guidelines on what you need to do before you start to visualise a given set of data:

1. Understand the Context
2. Choose the Appropriate Visual
3. Eliminate Clutter
4. Be a Designer
5. Tell the Story

According to Edward Tufte, an excellent visualisation expresses “complex ideas communicated with clarity, precision and efficiency”.



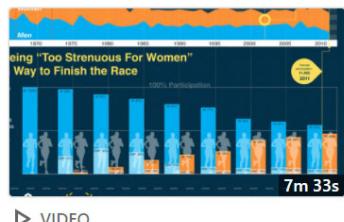
1. Understand the Context

- The success of Data visualisation is to use the data to communicate effectively
- It is important to understand the requirement/context that needs to be communicated
- Turn your data into information that can be consumed by an audience:
 - **Who** am I communicating to?
 - **What** do I want my audience to know or do?
 - **How** can I use data to help make my point?



2. Choose the Appropriate Visual

- Different graph types and visual displays have different impact in the terms of :
 - The communication to the audience
 - Presentation and control of the information to the audience
 - Level of explicit details to be presented at the right time
- Useful links:
 - <https://www.klipfolio.com/resources/articles/what-is-data-visualization#commonviztypes>
 - [LinkedIn Video Link](#)



► VIDEO

So what is data visualization?

By: Bill Shander

Copyright 2021 Republic Polytechnic



2. Choose the Appropriate Visual

Charts and graphs are typically used to convey one of the following:

- **Comparison**

- To compare one set of values with another. A comparison tries to set one set of variables apart from another, and display how those variables interact.

- **Relationship**

- To show a correlation or pattern between two or more variables.

- **Distribution**

- To list or show all the possible values (or intervals) of the data and how often they occur.

- **Composition**

- To show how various parts of the data make up the whole by collecting different types of information that make up a whole and display them together in one chart.



2. Choose the Appropriate Visual





An Influential Data Visualisation Guru



In a 2007 TED Talk, Hans Rosling (1948-2017), a Swedish physician, academic, statistician, and public speaker, used statistics and visual storytelling effectively to debunk myths about the so-called "developing world" – the often overlooked relationship between income and life expectancy.

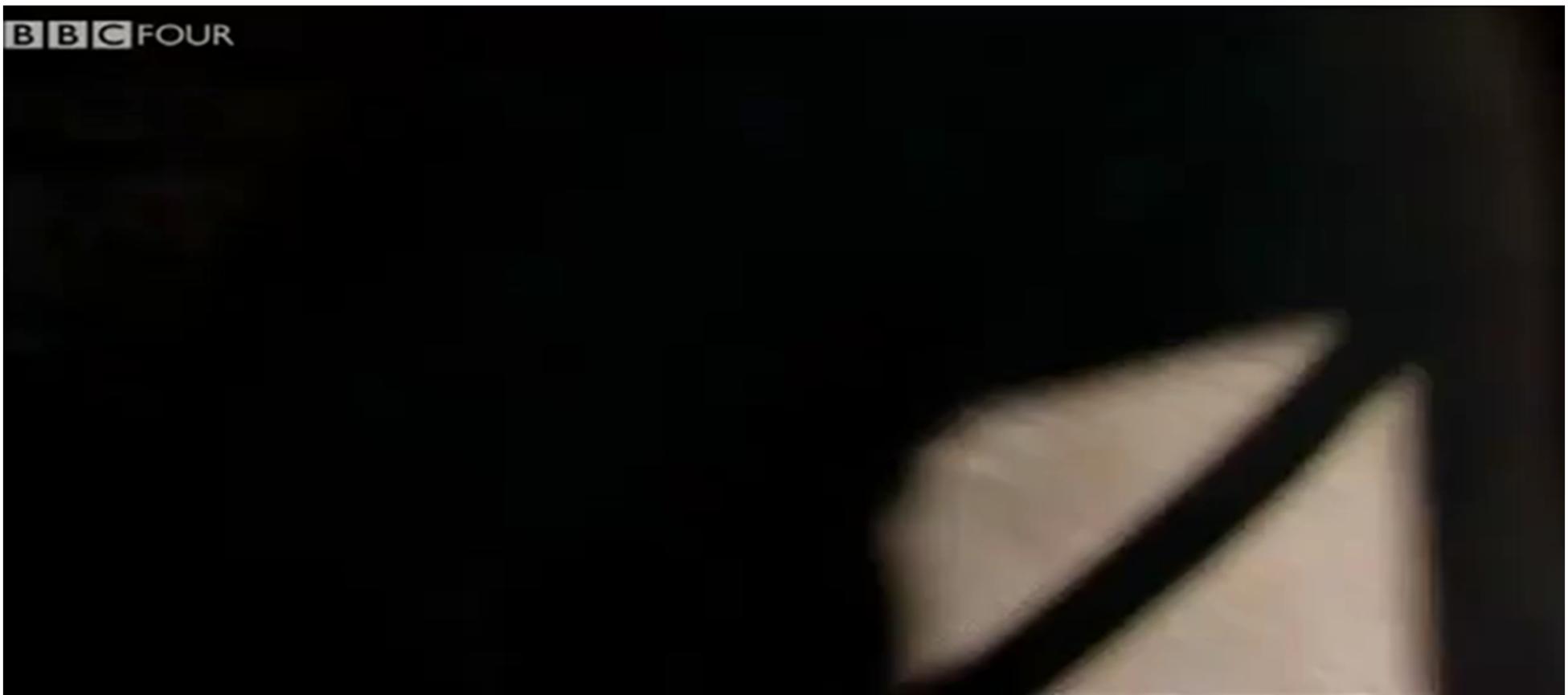
OFFICIAL (CLOSED) \ NON-SENSITIVE

Hans Rosling

200 Countries, 200 Year in 4 Min



BBC FOUR

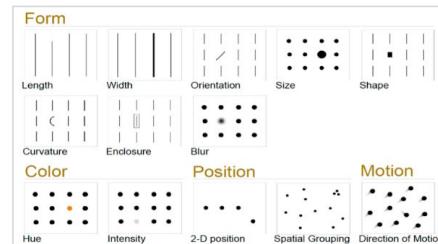




3. Be a Designer

- **Highlight important stuff** - Use preattentive attributes to draw our audience's attention to where we want them to focus

- Form
- Color
- Position
- Motion



- **Aesthetics** - Studies have shown that more aesthetic designs are perceived as easier to use and more readily accepted and used over time
 - Be smart with color
 - ✓ Use color sparingly and strategically to highlight the important parts of your visual
 - ✓ Consider using colourblind-friendly palette, when appropriate
 - Pay attention to alignment
 - ✓ Organize elements on the page to create clean vertical and horizontal lines to establish a sense of unity and cohesion
 - Leverage white space



Creating Colourblind-Friendly Viz

- **Red/Green** Colourblindness

- Avoid using the following combinations:

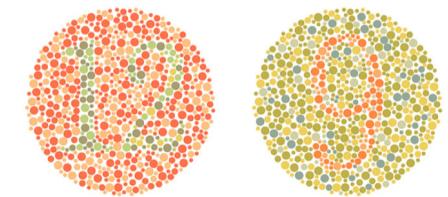
- Red, Green, Brown, Orange
- Blue, Purple
- Pink, Gray
- Brown, Gray
- Blue, Yellow

- If must use colourblind-unfriendly combinations:

- Leveraging colour value (light vs dark)
- Include alternative indicators to distinguishing data (e.g. icon, arrows, labels, annotations)

- Colourblind-safe combinations:

- Blue, Orange
- Blue, Red
- Blue, Brown
- Colourblind-friendly palette of Viz software (e.g. Tableau has 10 such palettes)



Original Image

Select Color Palette: Color Blind 10

Blue	Orange
Dark Gray	Light Gray
Medium Gray	Light Blue
Dark Gray	Light Orange
Blue	Light Gray

Deutanope Simulation

Select Color Palette: Color Blind 10

Blue	Yellow
Dark Gray	Light Gray
Medium Gray	Light Blue
Dark Gray	Light Yellow
Blue	Light Gray

Original Image

Select Color Palette: Color Blind 10

Blue	Orange
Dark Gray	Light Gray
Medium Gray	Light Blue
Dark Gray	Light Orange
Blue	Light Gray

Protanope Simulation

Select Color Palette: Color Blind 10

Blue	Green
Dark Gray	Light Gray
Medium Gray	Light Blue
Dark Gray	Light Green
Blue	Light Gray



3. Be a Designer – Some Tips

10 DATA DESIGN DOS AND DON'TS



1 | DO USE ONE COLOR TO REPRESENT EACH CATEGORY.



6 | DON'T USE HIGH CONTRAST COLOR COMBINATIONS SUCH AS RED/GREEN OR BLUE/YELLOW.



2 | DO ORDER DATA SETS USING LOGICAL HIERARCHY.



7 | DON'T USE 3D CHARTS. THEY CAN SKEW PERCEPTION OF THE VISUALIZATION.



3 | DO USE CALLOUTS TO HIGHLIGHT IMPORTANT OR INTERESTING INFORMATION.



8 | DON'T ADD CHART JUNK. UNNECESSARY ILLUSTRATIONS, DROP SHADOWS, OR ORNAMENTS DISTRACT FROM THE DATA.



4 | DO VISUALIZE DATA IN A WAY THAT IS EASY FOR READERS TO COMPARE VALUES.



9 | DON'T USE MORE THAN 6 COLORS IN A SINGLE LAYOUT.



5 | DO USE ICONS TO ENHANCE COMPREHENSION AND REDUCE UNNECESSARY LABELING.



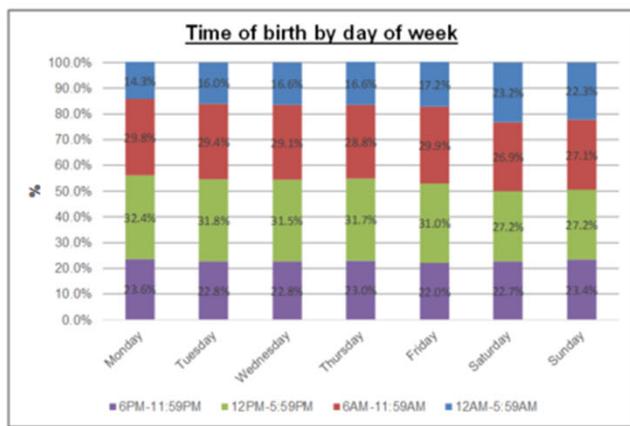
10 | DON'T USE DISTRACTING FONTS OR ELEMENTS (SUCH AS BOLD, ITALIC, OR UNDERLINED TEXT).



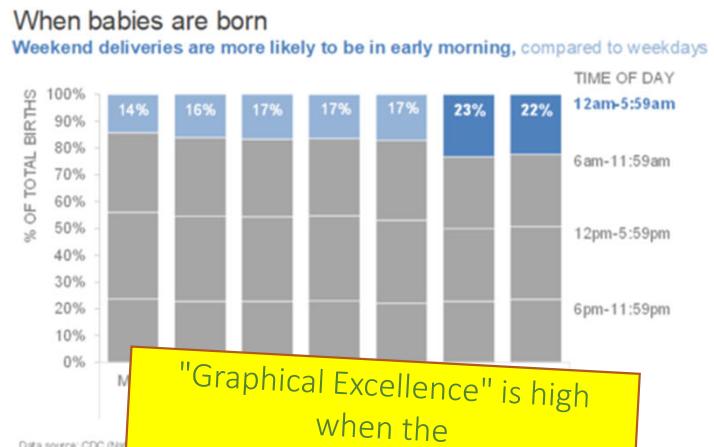
4. Eliminate Clutter

- There is a tendency that we may like to add too much information into the presentation. We may create fanciful chart but it doesn't serve good communication.
- Challenge yourself to examine what specific elements that aren't adding information and remove them.

BEFORE



AFTER

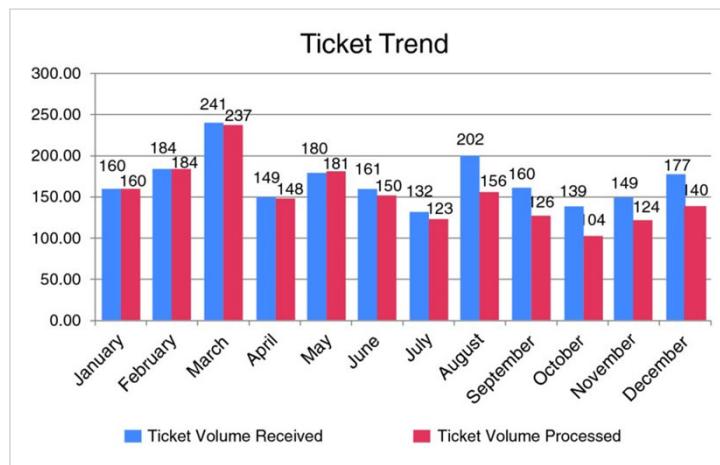


"Graphical Excellence" is high
when the
Data:Ink ratio is high and
Chartjunk is low.

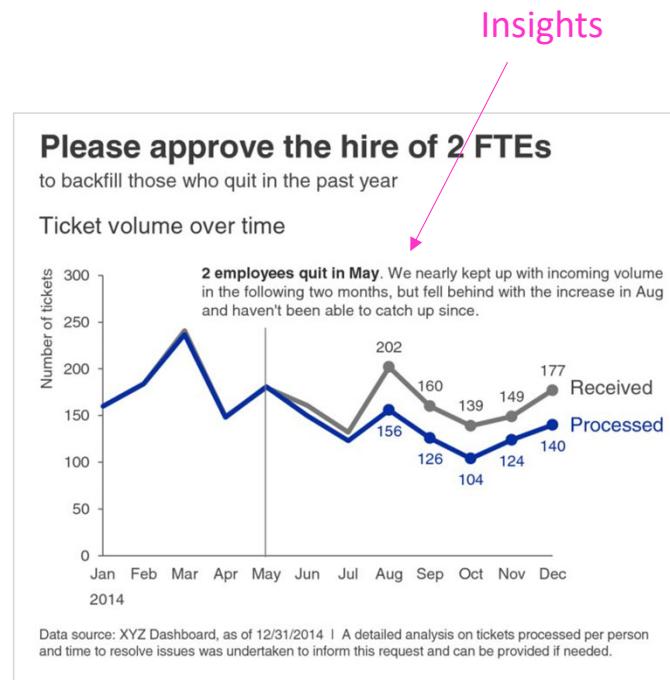


5. Tell the Story – Example

Instead of showing data, **tell the story** with data



BEFORE: Showing Data



AFTER: Storytelling with Data

OFFICIAL (CLOSED) \ NON-SENSITIVE



Introduction to Tableau



Popular Platforms



Source: Gartner (February 2021)

Copyright 2021 Republic Polytechnic



Benefits of Tableau

- **Easily connect to almost any data source:**
 - Corporate Data Warehouse
 - Microsoft Excel or CSV, JSON, Text files
 - Other web-based data

- **Can use real-time data feeds:**
 - Microsoft SQL Server
 - MySQL
 - Amazon Redshift, etc.





Other Benefits of Tableau

- **Tableau is a smart tool:**

- Recommends which visualisation tool to use
- 'Show Me' feature to show different types of graphs and charts
- Supports map function
- Interactive UI to design and develop data visualisation charts.
- Offers numerous fields and chart options:
 - Heat map
 - Scatter Plot
 - Packed Bubble
 - and more....





Tableau Product Suite

Tableau Product Suite

	Desktop		Reader	Server	Public	Online
Details	Personal	Professional				
Details	<ul style="list-style-type: none"> - Local client for building dashboards - Limited data sources, no ability to connect to Tableau Server 	<ul style="list-style-type: none"> - Local client for building dashboards - Full enterprise capabilities 	<ul style="list-style-type: none"> - Local client to view and interact with local files - Unable to modify workbooks or connect to server 	<ul style="list-style-type: none"> - Privately managed Tableau Server (may be on premise or service hosted) - Users may directly interact with dashboards via browser 	<ul style="list-style-type: none"> - Essentially a massive, public non-commercial Tableau server - All data published is public - Free client available to create dashboards 	<ul style="list-style-type: none"> - Private version of Tableau Public eliminates need for infrastructure - Live connections currently only possible with Google BigQuery and Amazon Redshift
OS	 	 	 		N/A	N/A
License	\$999	\$1,999	Free	Named User or Core Licensing	Free	\$500/user per year

John Mathis



Key Differences

Products	Tableau Desktop	Tableau Reader	Tableau Public	Tableau Online
Features:	Free trial for 14 days	Read-only version.	Absolutely free. Workbooks are stored in the cloud.	Private version of Tableau Public
Licensing:	Request for free license for students and educators for 1 year. US\$999/user	Free	(Not recommended for use in CET FDV course) Free	US\$500/user per year
Limitations:	Has basic data pre-processing abilities.	Cannot edit or modify workbooks. Cannot download and save soft copy.	User's data and workbooks are public domain. Cannot download and save soft copy.	Live connections to only Google BigQuery and Amazon RedShift

OFFICIAL (CLOSED) \ NON-SENSITIVE



Working with Tableau



Tools used for this training

- Tableau Ver. 2021



Data Types

- Sometimes Tableau may identify a field with a data type that is incorrect.
- For example, a field that contains dates may be identified as an integer rather than a date.
- You can change the data type in Tableau by right-clicking the field in the Data window, selecting Change Data Type, and then selecting the appropriate data type.

A screenshot of the Tableau Data window. On the left, there's a sidebar with tables like 'Country_Lookup' and 'CUST_ORDERS'. In the center, under 'Marks', there's a 'Pie' icon. Below it, a context menu is open for the 'Employee_ID' field, which is highlighted with a pink border. The menu options include 'Number (decimal)', 'Number (whole)', 'Date & Time', 'Date', 'String', 'Boolean', 'Default', and 'Geographic Role'. A pink arrow points from the text 'Click for drop-down to change the data type.' to the top of the context menu. The status bar at the bottom says 'Employee_ID 1000 rows'.

Data Type	Icon
Text (string) values	Abc
Date Values	📅
Date & Time Values	🕒
Numerical Values	#
Boolean Values	T/F
Geographic Values	🌐
Cluster Group	⌚



Type of Reports

- Standard Report
 - report with fixed layout and generated at scheduled timing for specific group of users
 - example, daily sales report to be sent to the Sales team at the end of each day.
- Parameter Reports
 - report with fixed layout that requires input of parameters before the report is generated.
 - example, the user may be prompted to input the range of months to be reported

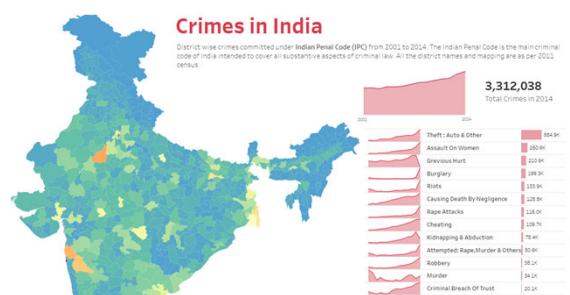


Types of Reports

- Ad hoc Reports
 - Report created by the user on the spot, either starting from an existing standard or parameter report or from scratch.
- OLAP Reports
 - Analysis reports which are generated from pre-calculated data sets which can be drilled down, sliced and diced
- Dashboard Reports
 - Highly aggregated reports showing key KPI's of how the organisation is doing
 - Usually are presented in visual form

Example of Dashboard

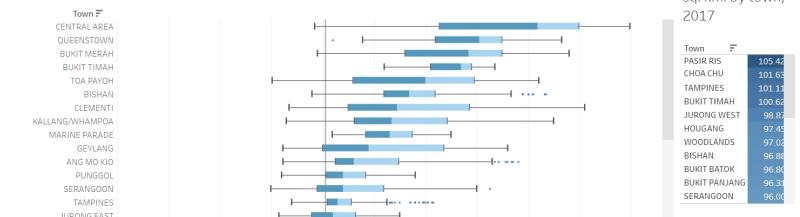
Sales Dashboard, Healthcare Dashboard, Crime Dashboard, Housing resale Dashboard



Distribution of 4-ROOM HDB resale prices by town, 2017

Data source: <https://data.gov.sg/dataset/resale-flat-prices>

There are 9 towns with a median resale price for 4 bedroom HDBs that are lower than the Singapore country median and 17 towns that are more than the national median.



**Please ensure
that you have a
working copy of
Tableau Ver.
2021**

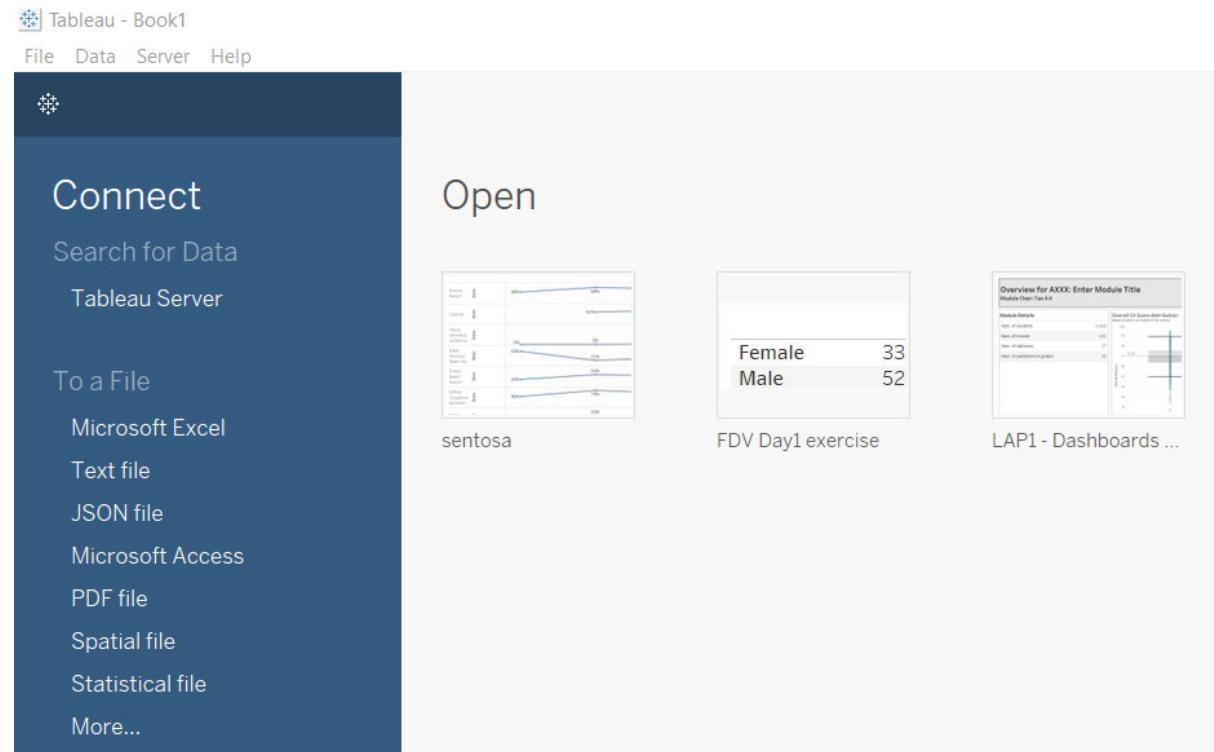




Tableau Product Key

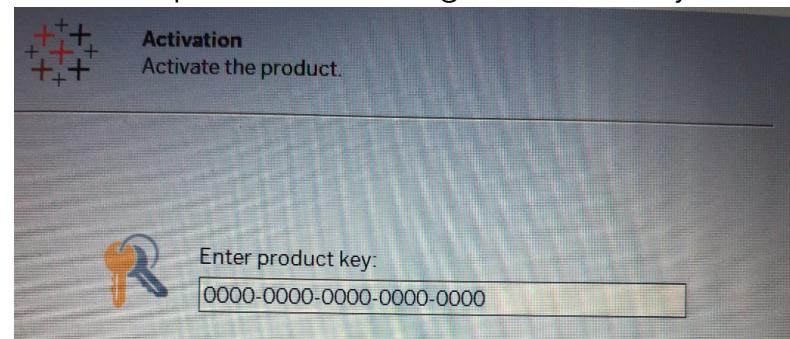
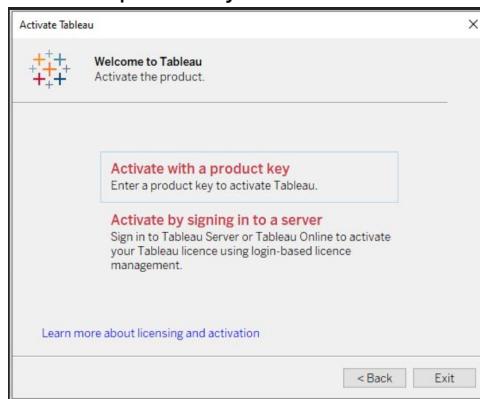
- Download Tableau Desktop into your laptop:
[Click to download the latest version of Tableau Desktop](#)

- Activate with this product key, if required:

TC3I-BC1F-BC40-A200-5F10

Please note that this is acad license, pls use it ONLY for learning purposes

(Or, update your license in Tableau: Help menu -> Manage Product Keys)



Copyright 2021 Republic Polytechnic

Tableau Registration

Almost there
Already purchased? [Activate Tableau](#)

First name	Surname
Business email	Organisation
Department	...
Job role	...
Country/Region	United Kingdom
County	
Postcode	e.g. SE1 0SU
Phone (e.g. 07400 123456)	
Start trial now	
We respect your privacy Having trouble?	



Key points covered

Three elements of graphical Excellence:

1. Data:Ink ratio & Chartjunk
2. Graphical Integrity
3. Data Density

Five Guidelines to data visualization:

1. Understand the Context
2. Choose the Appropriate Visual
3. Eliminate Clutter
4. Be a Designer
5. Tell the Story

Let's have a Lunch Break (11.35 – 1pm)



Demonstration and Exercises

Refer to Tableau Training Exercises

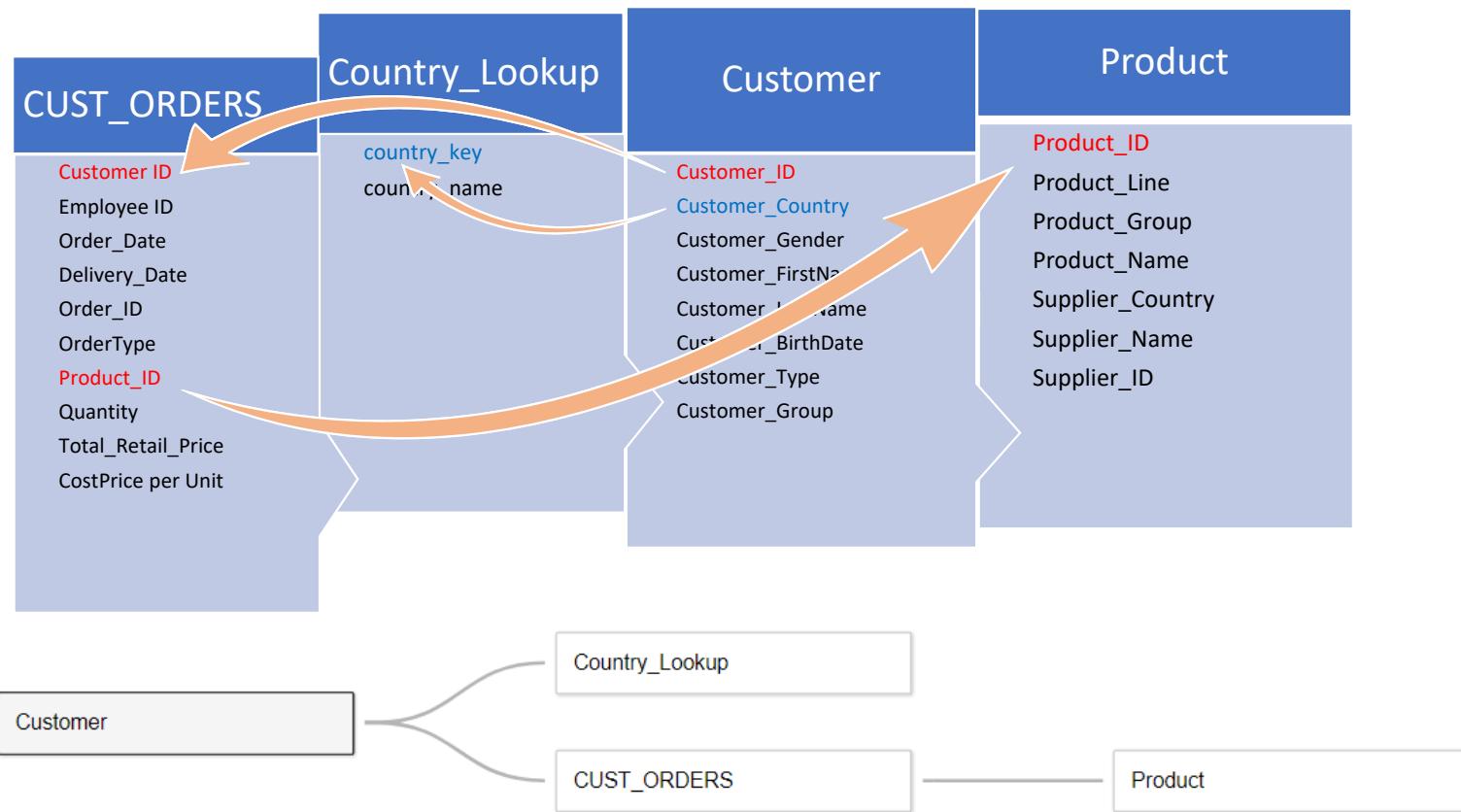


Exercises 1-8

- Exercise 1: Connect to Data Sources
- Exercise 2: Explore the Workspace
- Exercise 3: Tabulate the Data
- Exercise 4: Create Table and World Map
- Exercise 5: Generate a Tabular Report
- Exercise 6: Generate a Visual Report using a Bar Chart
- Exercise 7: Generate other Visual Reports
- Exercise 8: Add New Data Sources via Relationships



Examining the given data source



Orion_Data_V1a.xlsx



Exercise 1 & 2

Tableau - Book1

File Data Worksheet Dashboard Story Analysis Map Format Server Window

Data Analytics < Pages Columns Rows

Orion_Data_V1a

Search

Tables

- Customer BirthDate
- Customer Country
- Customer FirstName
- Customer Gender
- Customer Group
- Customer ID
- Customer LastName
- Customer Name
- Customer Type
- Measure Names
- Customer (Count)
- Latitude (generated)
- Longitude (generated)
- Measure Values

Marks

Automatic

Color Size Text

Detail Tooltip

Drop field here

Sheet 1

Drop field here

0 Data Source Sheet 1

Workbook Name

Main Menu

Shortcut Bar

Explore the Workspace

Analytics Shelf

Data Shelf

Show Me: For Best Practices automation and suggestions

View

Tables Data: Dimensions (Blue) and Measures (Green)

Data Source

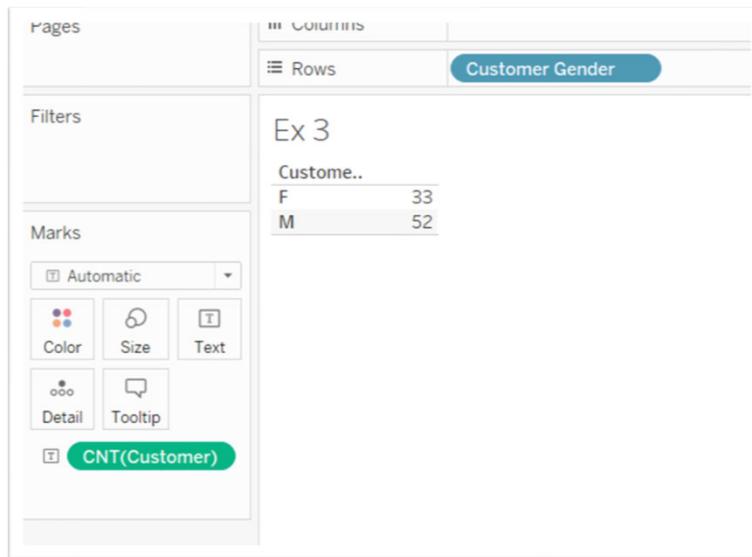
Ex 1: Connect to Data Sources

Ex 2: Explore the Workspace

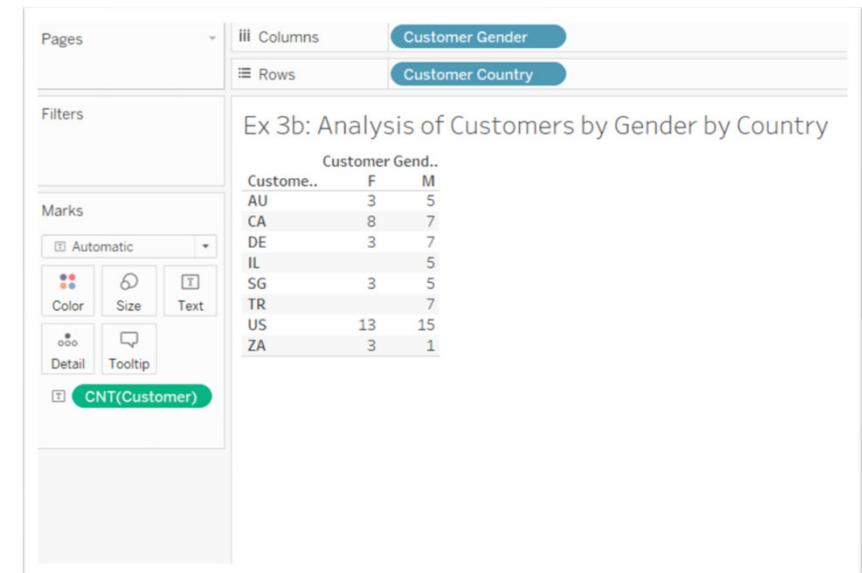


Exercise 3: Tabulate the Data

Ex 3a: Analysis of Customer numbers by Gender using a table



Ex 3b: Analysis of Customer numbers by Gender by Country using a table

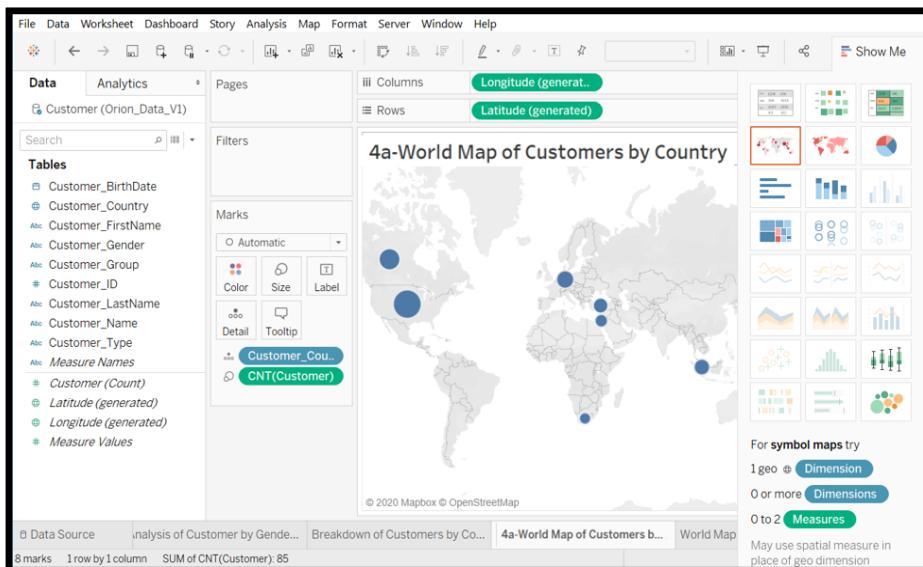


Remember to save your work regularly!

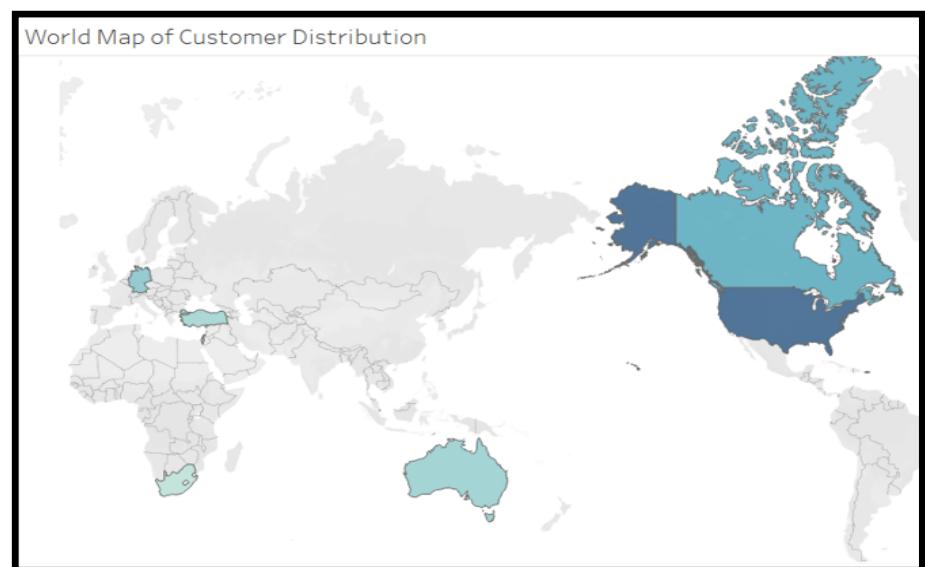


Exercise 4: Create Table and World Map

Ex 4a : Analysis of Customer numbers by country using a country map



Ex 4b: Analysis of Customer numbers by country using a country map



Questions:

What are the advantages and disadvantages of using a World Map to represent the data?
Which version is better?



Exercise 5: Generate a Tabular Report

Ex 5 - Analysis of **Customer numbers** by **Customer Type** and **Country** using a table

Custo..	Customer_Type						
	Internet/ Catalog Customers	Orion Club Gold members high activity	Orion Club Gold members low activity	Orion Club Gold members medium activity	Orion Club members high activity	Orion Club members low activity	Orion Club members medium activity
AU	1		1	2	2		2
CA	1	1	2	2	3	6	
DE	1	3	1	1	1	2	1
IL	1	1		2			1
SG	2	1	1		2		2
TR		1		2	4		
US	4	3	1	4	2	6	8
ZA		1	1				2

Ex. 5a – Analysis of **Customers numbers** by **Customer Type** and **Country** using a heat map



Note that row and column have been amended.

Questions:

What are the advantages and disadvantages of using a Highlight Table to represent the data?
In what situations will this come in useful or not useful?

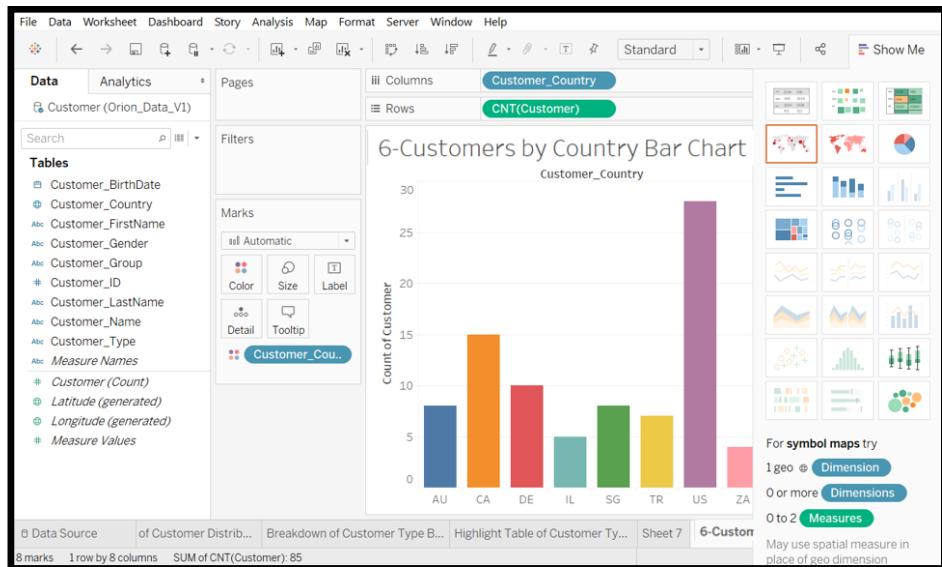
Let's take a break (3pm - 3.30pm)





Exercise 6: Generate a Visual Report using a Bar Chart

Ex 6 - Analysis of Customer numbers by Country using a bar chart



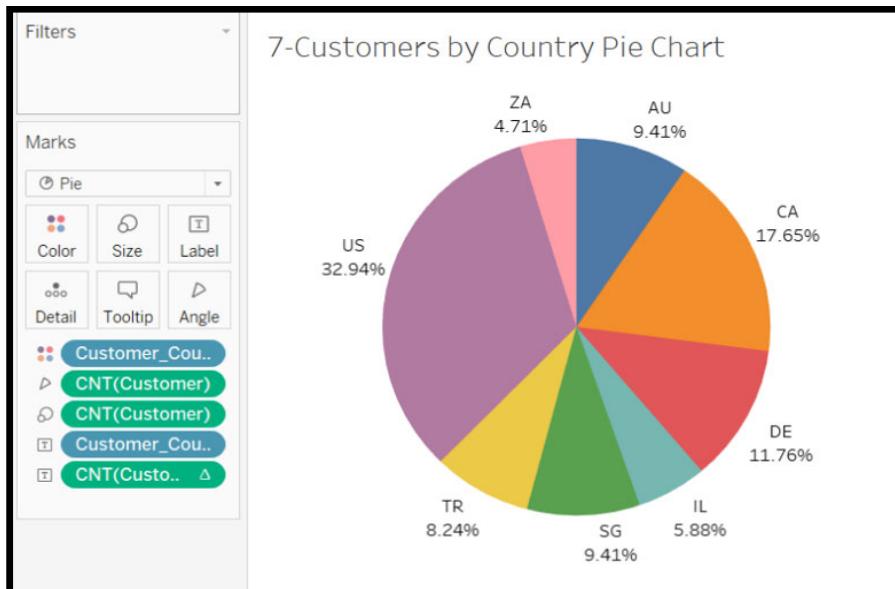
Questions:
Is this a good visualisation? Why or why not

Note the column and row data

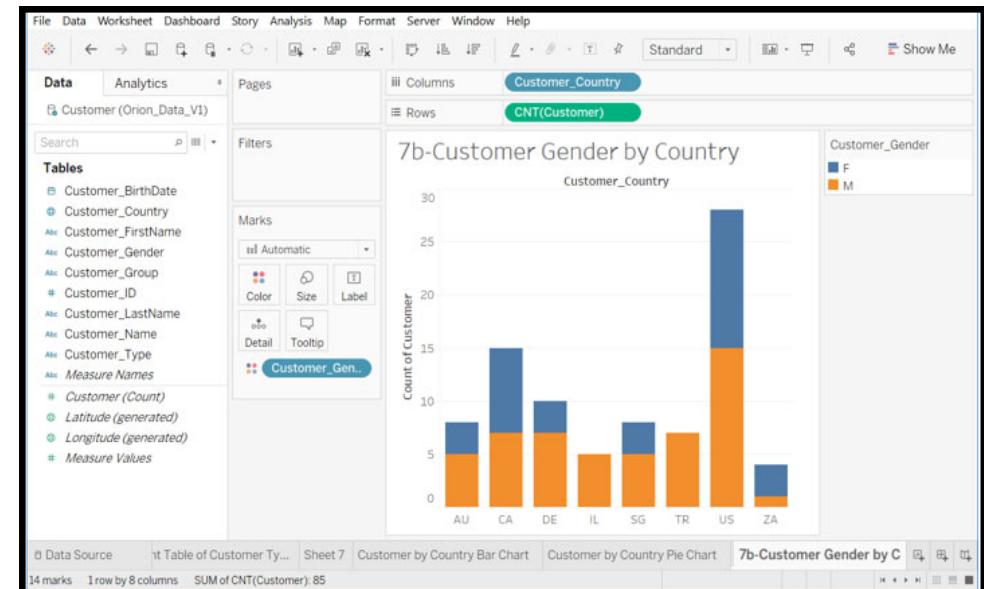


Exercise 7: Generate other Visual Report

Ex 7a – Analysis of Customer numbers by Country using a pie chart



Ex 7a – Analysis of Customer numbers by Gender by Country using a stacked bar chart



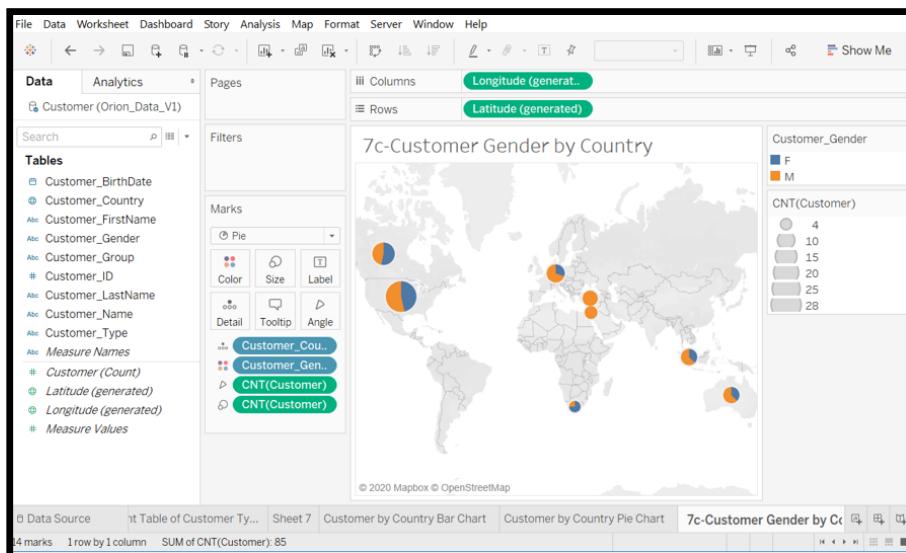
Questions:

What are the advantages and disadvantages of a Stacked Bar Chart?
What immediate inferences can you derive from the chart above?



Exercise 7: Generate a Visual Report

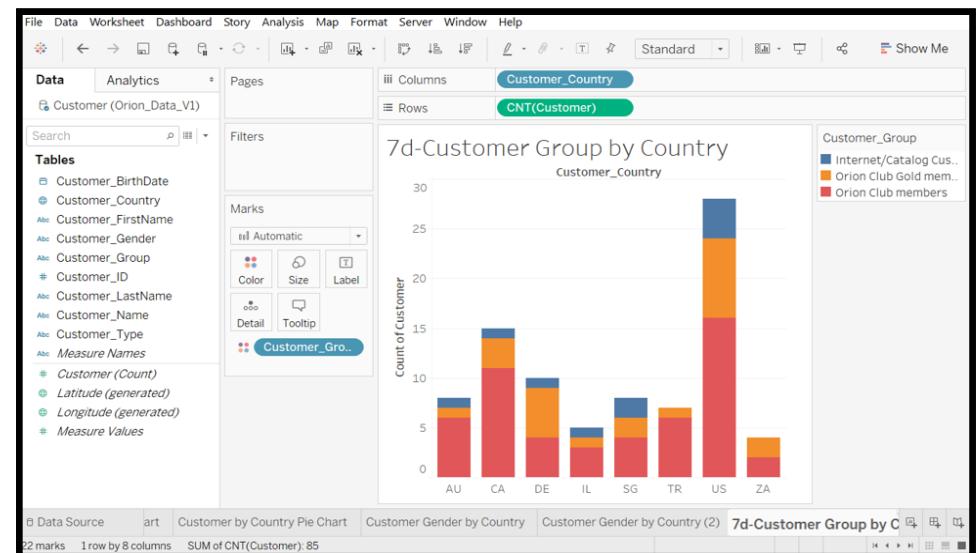
Ex 7c – Analysis of Customer numbers by Gender and Country using a world map



Question:

Can you think of situations where you would want to use a Pie Chart on a World Map?

Ex 7d – Analysis of Customer numbers by Customer group and Country using a stacked bar chart



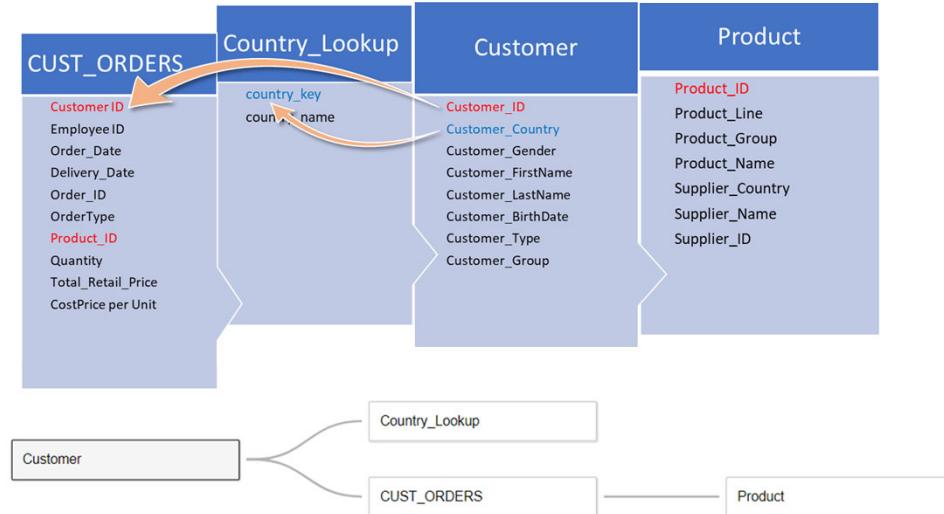
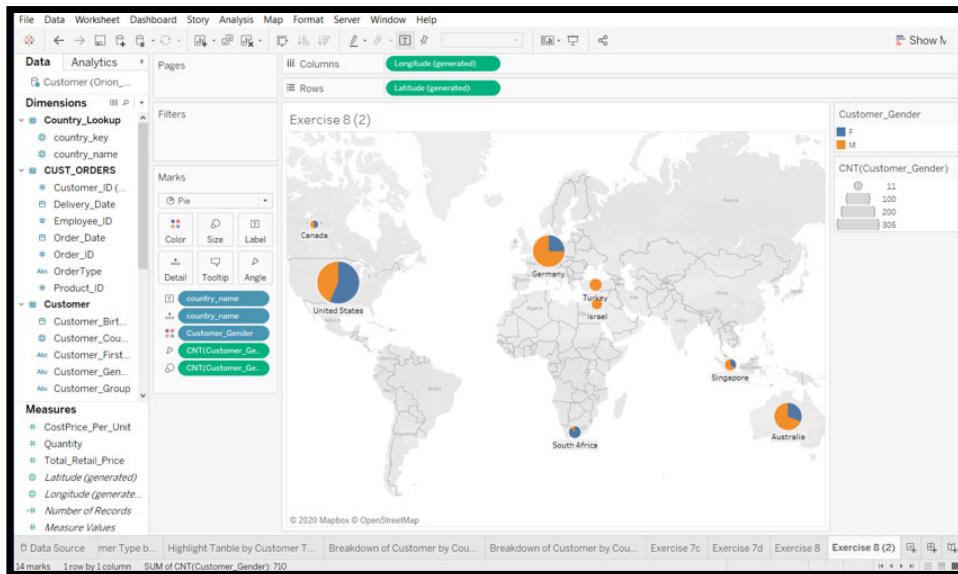
Question:

What immediate inferences can you derive from the chart?



Exercise 8: Add New Data Sources via Relationships

Ex 8a – Analysis of Customer numbers by Gender and Country using a world map. Print out country name.



Orion_Data_V1a.xlsx



Key points covered

Tableau

- Loading data
- The different types of visualisation available
- Dimension vs Measure
- Data type
- Marks:
 - colour, label, size, tool-tip.

OFFICIAL (CLOSED) \ NON-SENSITIVE



End of Day 1

Fundamentals of Data Wrangling

Day 2

• REC

OFFICIAL (CLOSED) \ NON-SENSITIVE

Administrative matters

Please change your log in screen name to resemble the registered name for the course



[Attendance link](#)

The instructor is **busy** taking
attendance ☺

Set your microphone to muted when not speaking



Overview

Day 3	Morning	Data Visualisation Principles Install Tableau desktop
	Afternoon	Create and interpret Data Visualisation charts
Day 4	Morning	Build interactive Data Visualisation dashboard Tell Data-driven stories
	Afternoon	Quiz Mini Project & Presentation



Outline for the Day 2

Time	Agenda
9.00am	Welcome and admin matters
9.15am – 10.30am	
10.30am – 10.45am	Break
10.45am – 12.30pm	
12.30pm – 1.30pm	Lunch
1.30pm – 3.15pm	
3.15pm – 3.30pm	Break
3.30pm – 4.45pm	
4.45pm – 5.00pm	Wrap up, Q&A



Key take-away from day one



Key points covered in day 1

The Art of visualization

- Principles for clearer and effective data representation:
- Three elements of graphical Excellence; and
- Five Guidelines to data visualization.

Tableau

- Loading data
- The different types of visualisation available
- Dimension vs Measure
- Data type
- Marks:
 - colour, label, size, tool-tip.



Day 2 Exercises 9 - 20

- Exercise 9: Use of Filters for interactivity
- Exercise 10: Use the Keep Only Function
- Exercise 11: Assign Alias to Values
- Exercise 12: Create a One-Way Frequency Table with Filter
- Exercise 13: Create a Parameter
- Exercise 14: Add a New Field – Part 1
- Exercise 15: Generate a Dual Line Chart
- Exercise 16: Create a New Field – Part 2
- Exercise 17: Edit Formula
- Exercise 18: Create a Hierarchy
- Exercise 19: Create a Dashboard
- Exercise 20: Create a Story

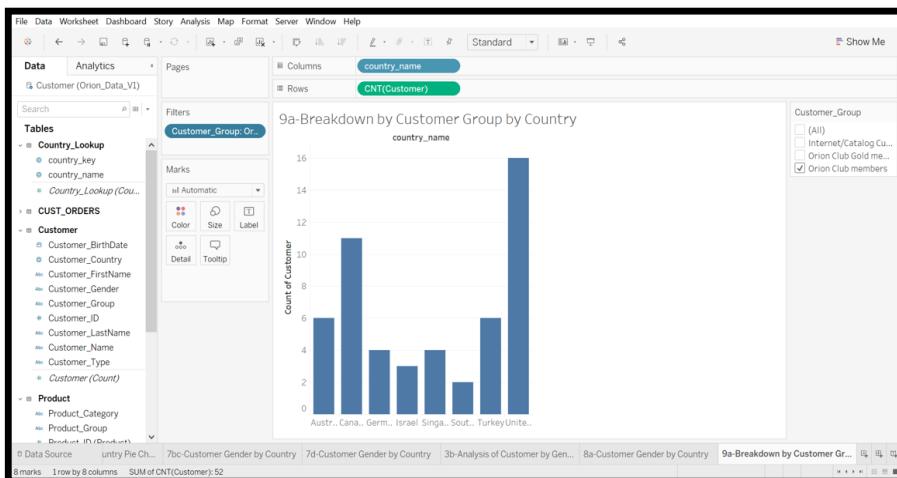
~ Quiz ~

~ Mini Project & Presentation ~

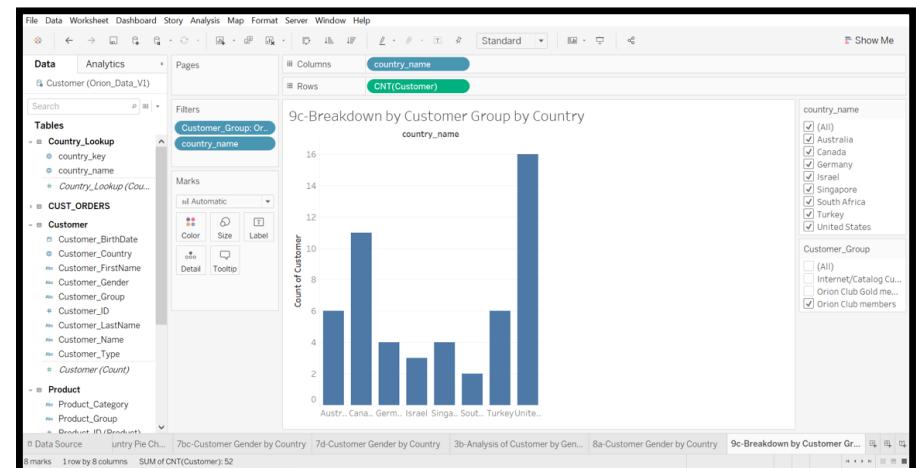


Exercise 9: Use of Filters for Interactivity

Ex 9a – Analysis of Customer numbers by Country using a bar chart. Filter to show Orion Club members only.



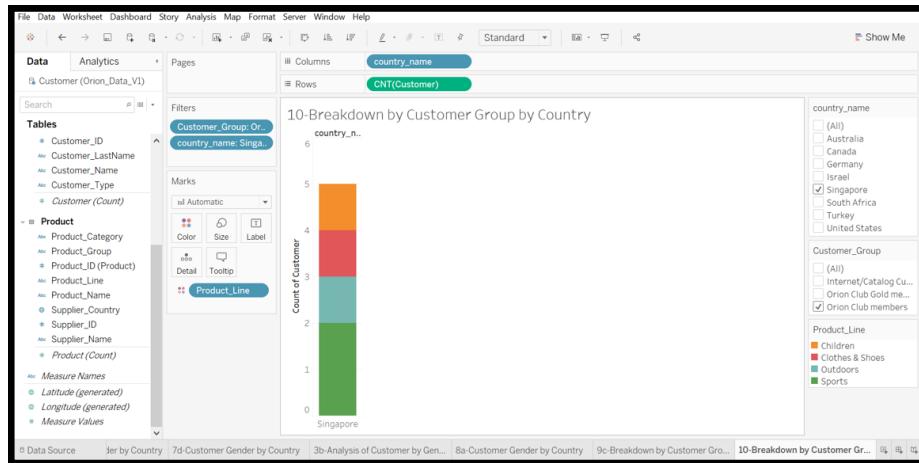
Ex 9b – Analysis of Customer numbers by Country using a bar chart. Configure filter for country_name and customer_group.





Exercise 10: Use the Keep Only Function

Ex 10 – Analysis of Customer numbers by Country using a stacked bar chart. Filter to show Orion Club members only and only for Singapore.



Questions:

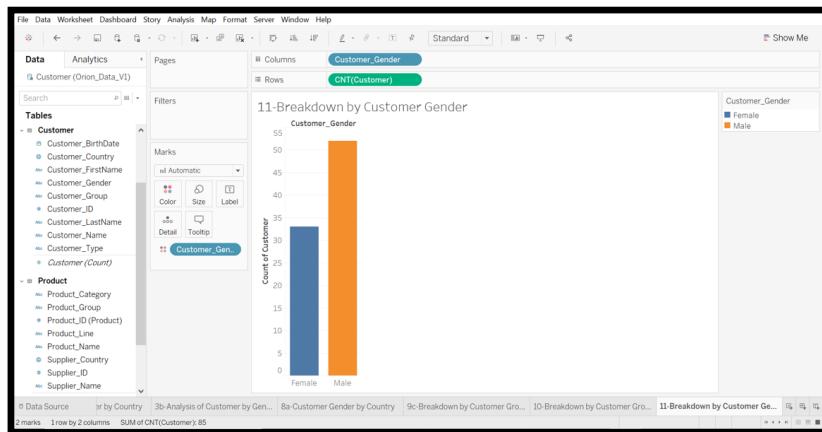
How do you now go back to the original Stacked Bar containing all the countries in the chart or select a new set of countries e.g. Singapore and Australia?

What are the advantages and disadvantages of the 'Keep Only' function?



Exercise 11: Assign Alias to Values

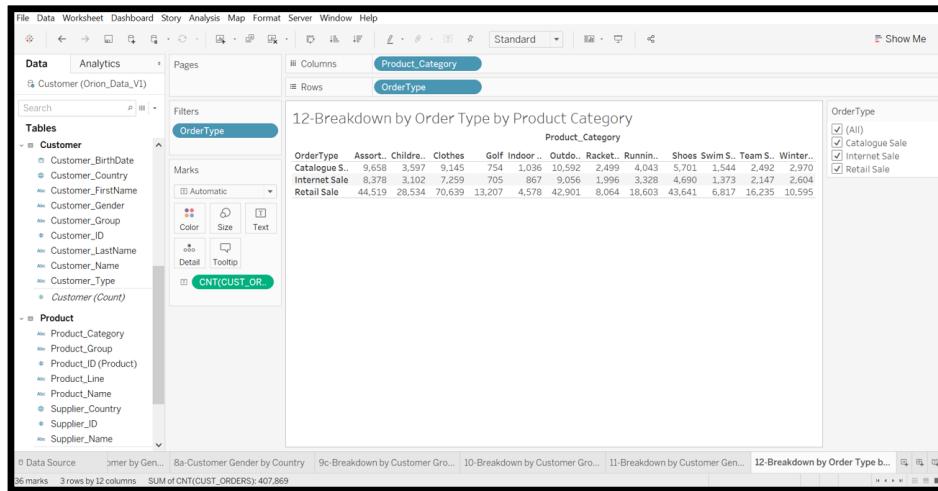
Ex 11 – Analysis of Customer numbers by Gender using a bar chart. Name the gender as “Female” and “Male”.





Exercise 12: Create a One-Way Frequency Table with Filter

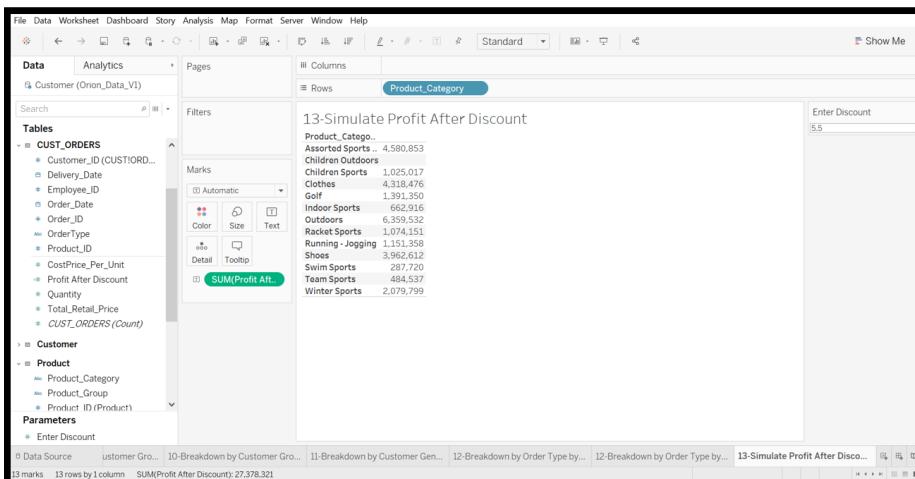
Ex 12 – Analysis of Customer orders by Order type and Product Category, using a table. Allow for filter of order type.





Exercise 13: Create a Parameter

Ex 13 – Analysis of Profit after discount by Product Category, using a table. Allow user to specify the discount amount.



Questions:

What is the maximum discount (in integer) you can give for Swim Sports to avoid being unprofitable?

Let's take a break (10.30am to 10:50am)





Exercise 14: Add a New Field – Part 1

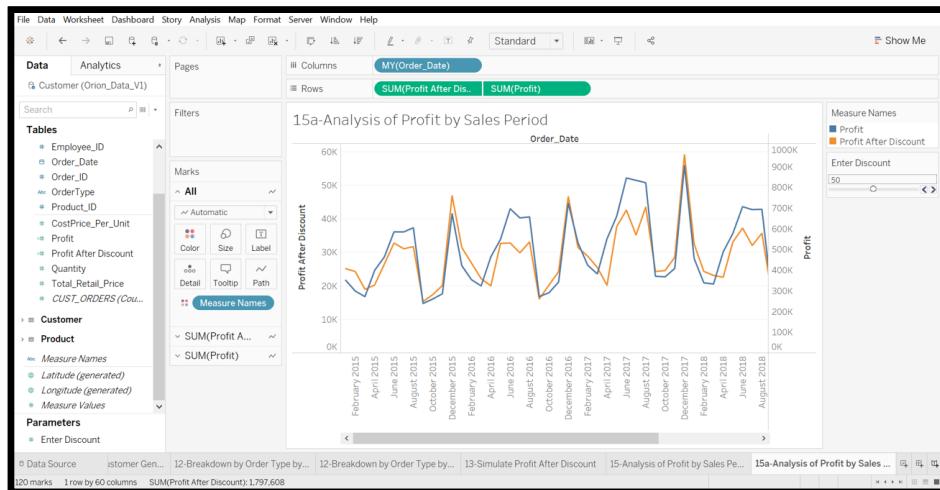
Ex 14 – Create a new calculated field called “profit”.

The screenshot shows a software interface for creating a calculated field. The field name is "Profit" and the formula is "[Total Retail Price] - ([CostPrice Per Unit] * [Quantity])". A tooltip for the "ABS" function is open, stating: "Returns the absolute value of the given number. Example: ABS(-7) = 7".



Exercise 15: Generate a Dual Line Chart

Ex 15 – Analysis of Profit after discount and Profit by order date (month, year) using a dual line. Allow user to specify the discount amount.



Observe:

Input different values in the ‘Enter Discount’ parameter and observe the changes in the chart.



Exercise 16: Edit Formula

Ex 16 – Analysis of average customer age by country using a bar chart.



Hint: You need to create a calculated value to calculate the age of the customers

DATEDIFF(date_part, start_date, end_date, [start_of_week])

Returns the difference between two dates where start_date is subtracted from end_date. The difference is expressed in units of date_part. If start_of_week is omitted, the week start day is determined by the start day configured for the data source.

Example: DATEDIFF('month', #2004-07-15#, #2004-04-03#, 'sunday') = -3

Ex 16a – Create Age Group Using Logical Function

Age Group

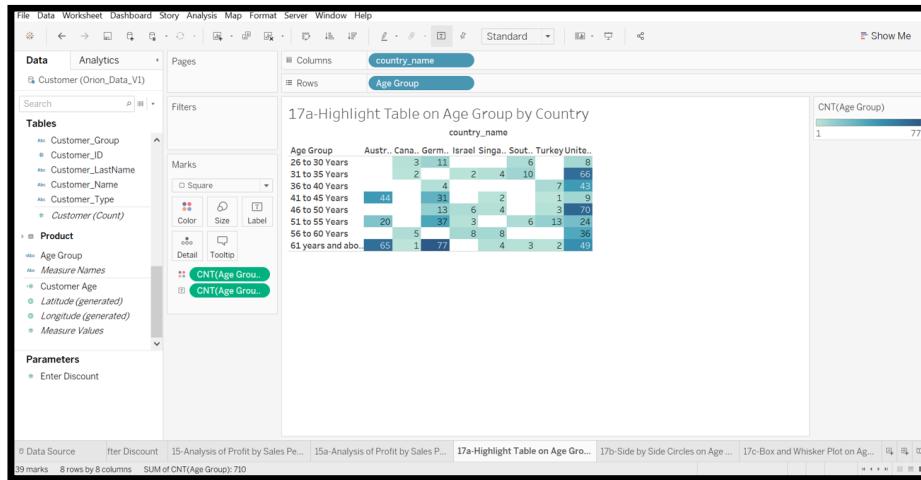
```
If [Customer Age] >= 5 AND [Customer Age] <= 10 then
    '5 to 10 Years'
elseif [Customer Age] >= 11 and [Customer Age] <=15 then
    '11 to 15 Years'
elseif [Customer Age] >= 16 and [Customer Age] <=20 then
    '16 to 20 Years'
elseif [Customer Age] >= 21 and [Customer Age] <=25 then
    '21 to 25 Years'
elseif [Customer Age] >= 26 and [Customer Age] <=30 then
    '26 to 30 Years'
elseif [Customer Age] >= 31 and [Customer Age] <=35 then
    '31 to 35 Years'
elseif [Customer Age] >= 36 and [Customer Age] <=40 then
    '36 to 40 Years'
else
    '41 years and above'
end
```

The calculation is valid.

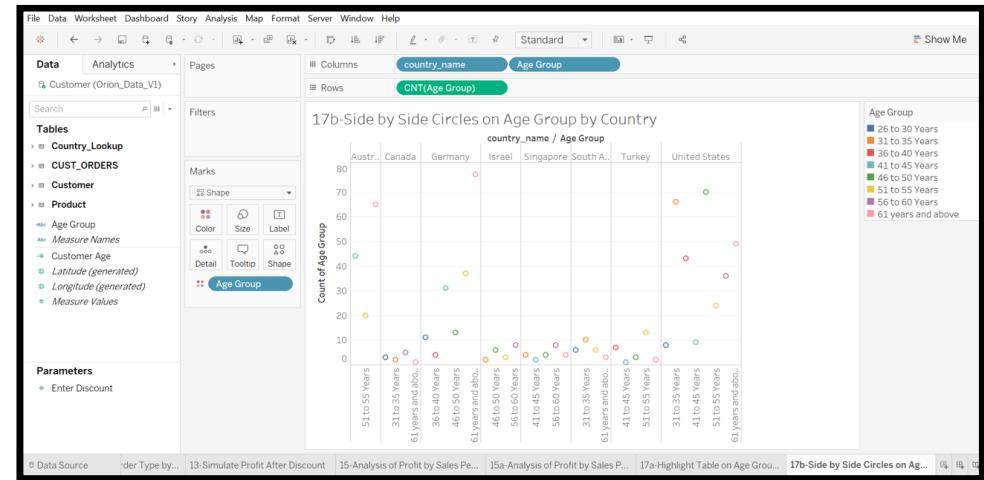


Exercise 17: Edit Formula

Ex 17a – Analysis of customer numbers by age group and country using a heat map.



Ex 17b – Analysis of customer numbers by age group and country using the side-by-side Circles.

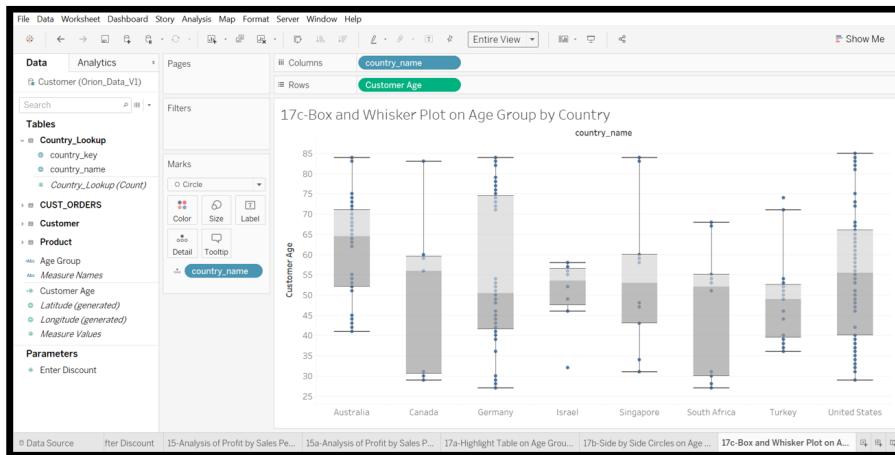


Hint: You will need to edit the formula to increase the resolution of the calculated field: “age group”.

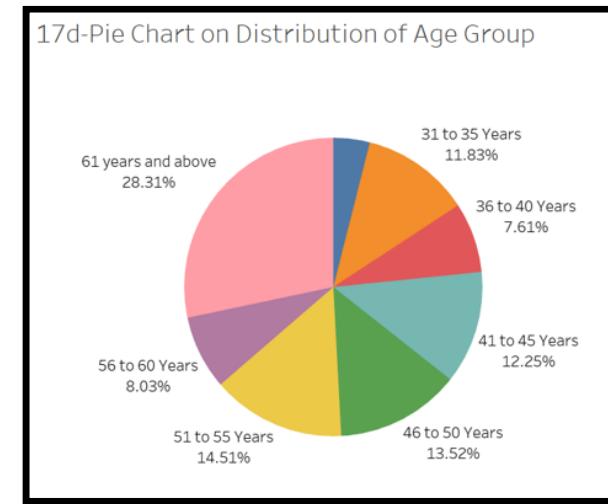


Exercise 17: Edit Formula

Ex 17c – Analysis of customer age by country using a box and whisker plot.



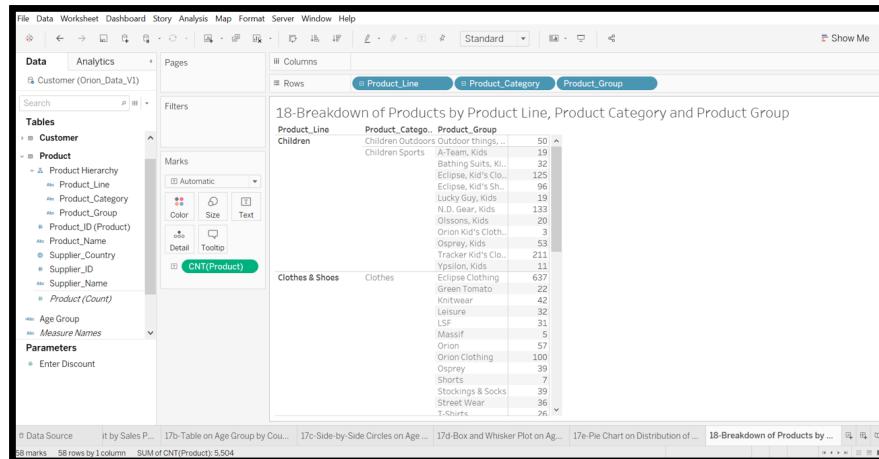
Ex 17d – Analysis of customer numbers by age group using a pie chart.





Exercise 18: Create a Hierarchy

Ex 18 – Analysis of **product line** via a hierarchy **using table**



Let's have a Lunch Break **(12.15pm to 1.30pm)**



Exercise 19: Create a Dashboard

Exercise 20: Create a Story

Let's take a break

()



OFFICIAL (CLOSED) \ NON-SENSITIVE



Day 2 - Quiz

OFFICIAL (CLOSED) \ NON-SENSITIVE



Day 2 – Mini Project & Presentation



What's Next?

Data Visualisation - Beyond the Basics	https://www.rp.edu.sg/ace/short-course/Detail/data-visualisation-beyond-the-basics
Fundamentals of Data Analytics	https://www.rp.edu.sg/ace/short-course/Detail/fundamentals-of-data-analytics
Data Science with Python	https://www.rp.edu.sg/ace/short-course/Detail/data-science-with-python
Text Mining - Making Sense of Unstructured Data	https://www.rp.edu.sg/ace/short-course/Detail/text-mining---making-sense-of-unstructured-data
Getting Started with Text Mining	https://www.rp.edu.sg/ace/short-course/Detail/getting-started-with-text-mining
goData Visualisation	https://www.rp.edu.sg/ace/short-course/Detail/godatavisualisation



Broad Areas Of ICT



Working with data to gather insights



Creating intelligent digital solutions



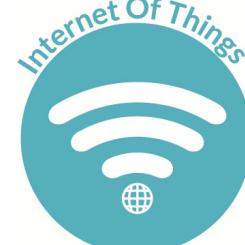
Protecting personal and organizational data assets



Programmatically automating tedious or manual work tasks



Managing outsourced IT infrastructure servers / computers / services



Connecting inter-related computing devices for Smart Nation solutions

***Thank you for attending
Fundamentals of Data Wrangling***
