

Accenture Applied Intelligence

THE BRIEF ON DATA POST WORKSHOP PACK

**A crash course for student consultants
in improving their data literacy**

May 2020



Workshop Takeaways

01

That you will have an **understanding** of the conceptual data landscape

02

An **appreciation** of the different approaches and tools that can be used on data

03

That you have increased **confidence** in creating meaningful and accessible insights from data

04

That you know where you lie on your journey and you know **what is next for you and data**

**CHANCES ARE,
THERE IS SOMEONE
BETTER THAN YOU**

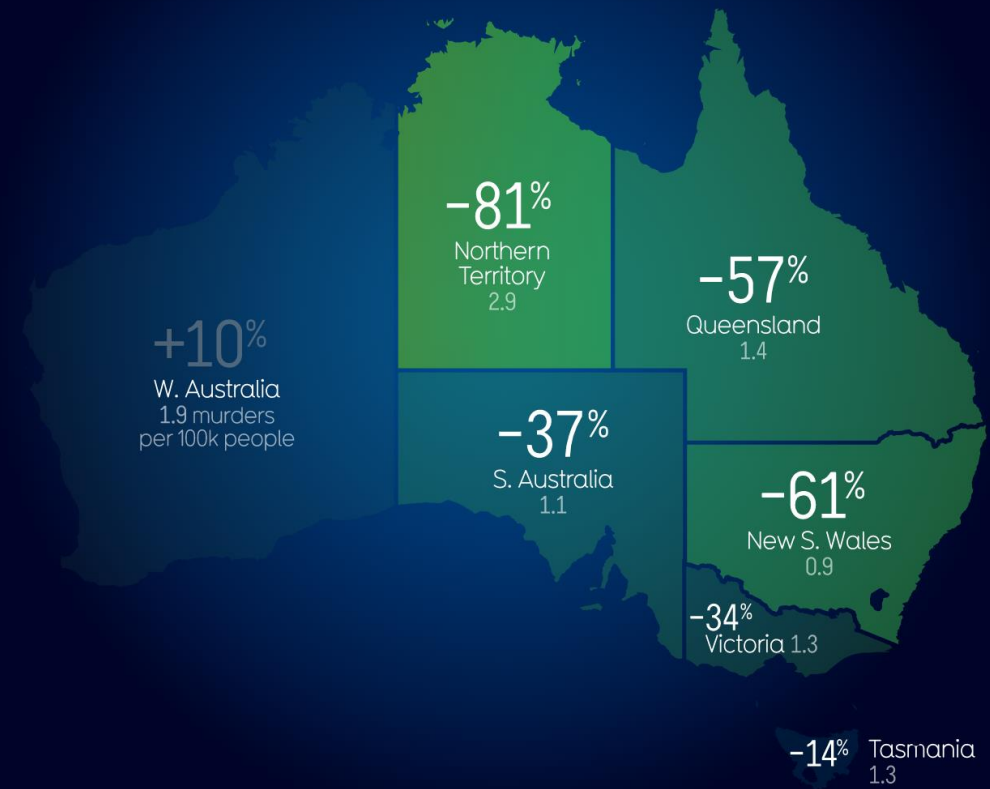
What inspires me to keep working with data (1)

I get excited when data is presented in **immediately engaging ways** that also are **thought provoking before first glance**.

This is the careful combination of **human centred design** with **storytelling through data**.

Australia's Homicide Rate Has Never Been Lower

% variation 1990-2016



beautifulnews

source: World Bank

What inspires me to keep working with data (2)

The battle of Midway in June 1942 was one of the most important Naval battles in history.

The United States were outnumbered by the Japanese yet prevailed through the use of military intelligence, led by Edwin T. Layton, to partially decode Japanese messages and detect patterns in these messages to create meaning.

This was dramatized in the 2019 movie titled 'Midway'. **Data collection, pattern finding and presentation was key for success.** A detailed analysis can be found [here](#).



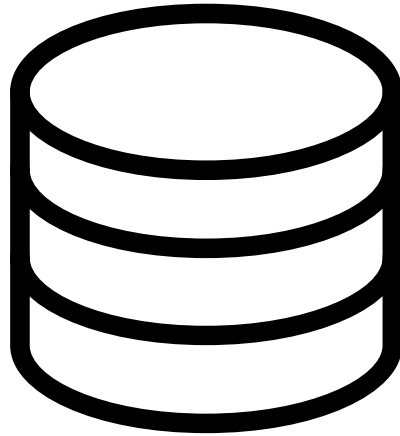
What inspires me to keep working with data (3)

People



Data has the ability to
liberate people. Not
enslave people to it.

Data

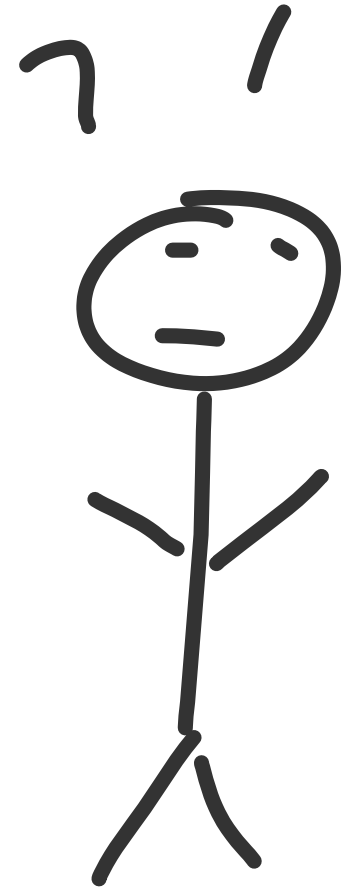


Has the capability to provide
guidance and truth in some
cases where there is none.

**BUT WHERE
DO WE BEGIN?**

Imagine a day in your life without data

1. You wake up – only to discover that the alarm on your phone is not working. More to that, you don't know what time it is.
2. You realise that your phone is seemingly not working at all, you decide that this is a problem that you'll have to fix later today.
3. After looking at the sun to discover that it is still morning (hopefully) you decide to wait for the bus to get to university. You wait an extraordinarily long time, it really is one of those days.
4. To make it worse, your Smartrider doesn't work. Guess cash wasn't pointless at all.
5. You get into university finally after what seems like an eternity after not being able to listen to music from your phone.
6. You get to class, only to discover the doors don't unlock.
7. You manage to find someone and you ask them what is going on. To your amazement, the world has shutdown because of a large-scale cybersecurity attack.





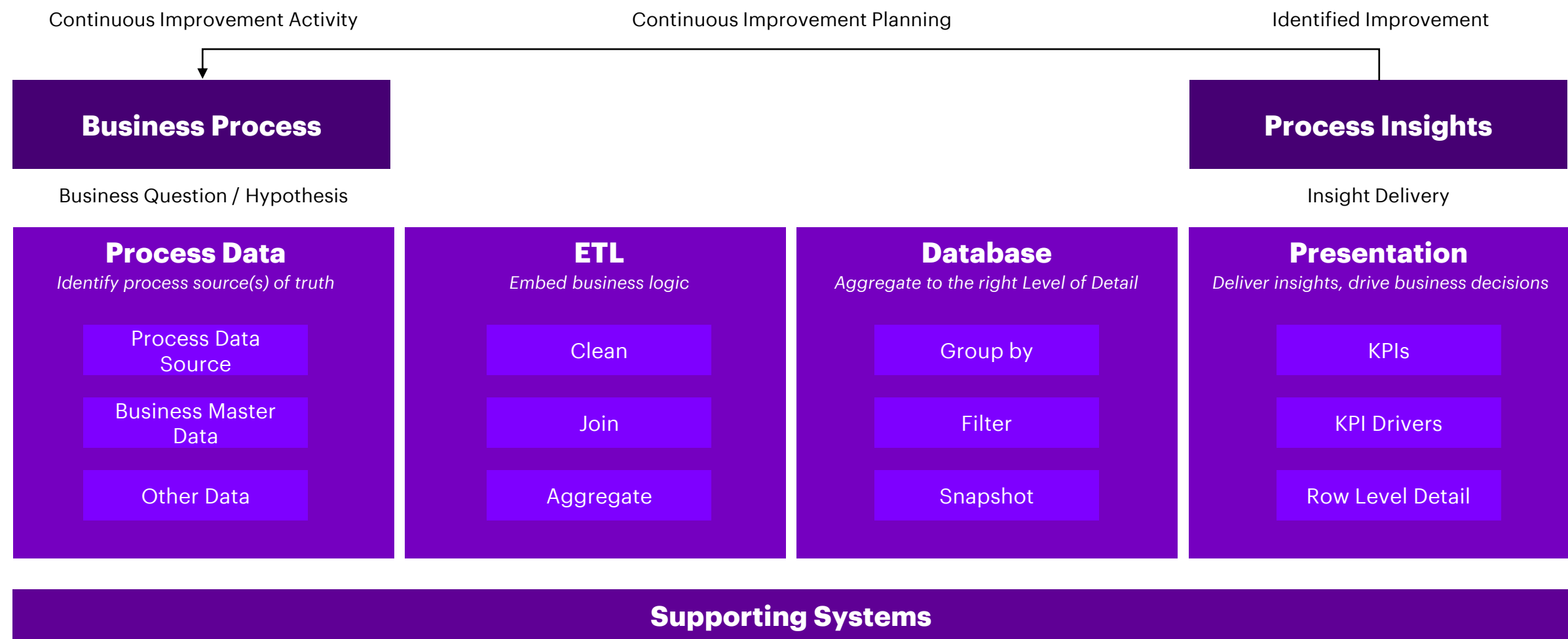
**Information is the oil of the 21st century,
and analytics is the combustion engine**

Peter Sindergaard
Senior Vice President, Gartner Research

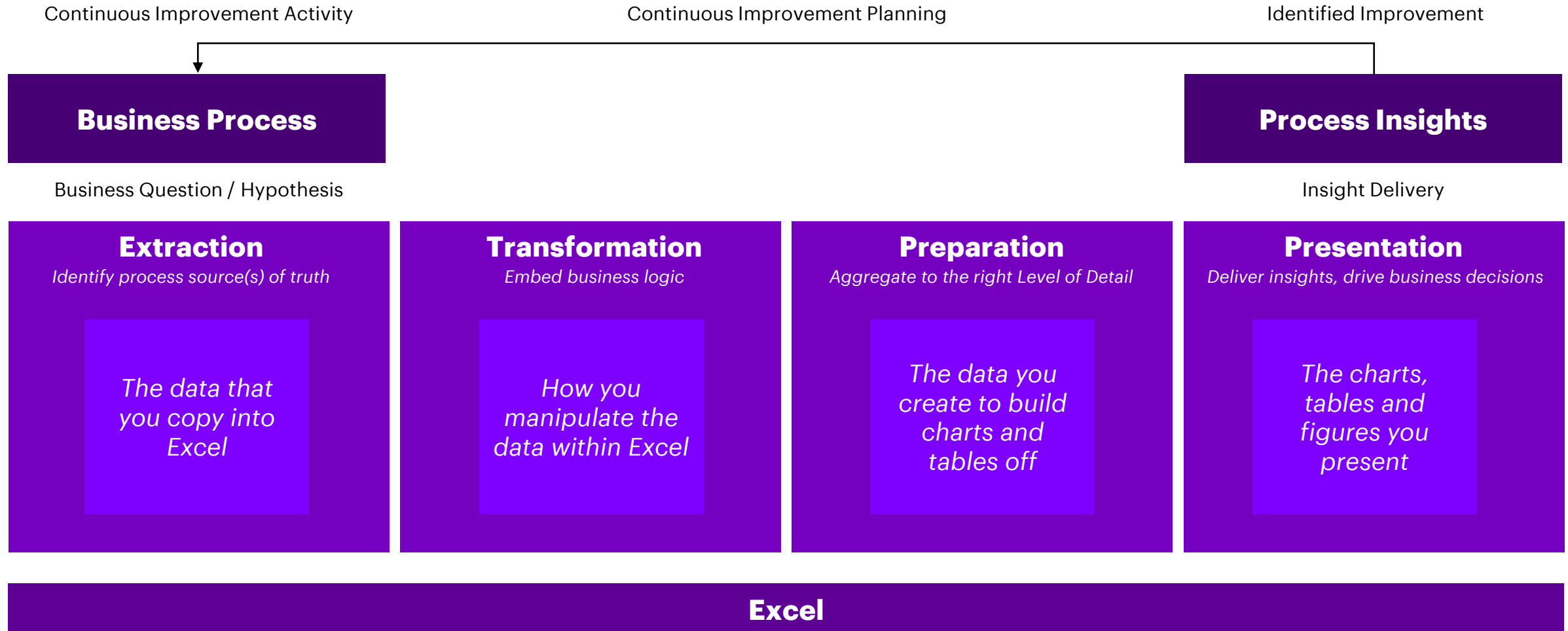
**EVEN THOUGH IT
MAY NOT SEEM IT,
DATA IMPROVES
OUR LIVES.**

**WHERE DOES DATA
COME FROM?**

The Data Pipeline



The Data Pipeline (for Excel)



What tools should you use?

Aim to use tools that are free (open source) or are widely used. Look not to specialise in tools but to understand the approaches that are used agnostic of tools.

Essential:

Excel – for Simple Data Storage, Simple Transformation and Simple Presentation

Recommended next steps to grow your skills:

SQL – Data Storage and Simple Transformation

Python – Data Transformation and Simple Presentation

R – Data Transformation and Simple Presentation

PowerBI – Simple Data Transformation and Presentation

There are many more tools you can choose from... **There is always going to be a cost to learn a new tool but question is the benefit of this really worth the effort?**

CASE STUDY

COVID-19 - Case Study

COVID-19 (coronavirus) is a virus that impacts the respiratory system. The virus was first detected in Wuhan, China in December 2019. The virus subsequently spread internationally. Following this the World Health Organisation (WHO) called the virus a pandemic on the 12th March 2020. Information is difficult to source, but the death rate appears to be between 2-10% for those infected depending on local circumstances. The virus has an incubation period of up to two weeks meaning it is difficult to track new infections due to transmission in the asymptomatic stage.

There are two agreed methods to handle virus outbreak to minimise its impact, herd immunity and curve flattening. Australia and its states have decided to pursue curve flattening in its approach to allow for time for vaccines and other treatments to be developed to reduce the impact of the virus. In doing so the hope is that the healthcare system will not be overloaded which may lead to additional deaths.

In order to achieve curve flattening, the Australian and state governments have implemented social distancing measures implemented in increasing severity from March 14th through to April 1st 2020. **The Australian and Western Australian CMO (Chief Medical Officer) have tasked you with performing an analysis to see if the measures have been effective based on current data available.**

Hospital capacity may be of concern, in WA this is estimated to be at 70,000 beds. For simplification, assume the virus impacts everyone the same (i.e. everyone with the virus must be hospitalised).

So what is the question that our client is seeking to answer and what is this logically?

Key Question

How can we minimise the impact of COVID-19 on the population?

Driver Level 1

Minimise Loss of Life

Minimise Economic Loss

Minimise Loss to Social Freedoms

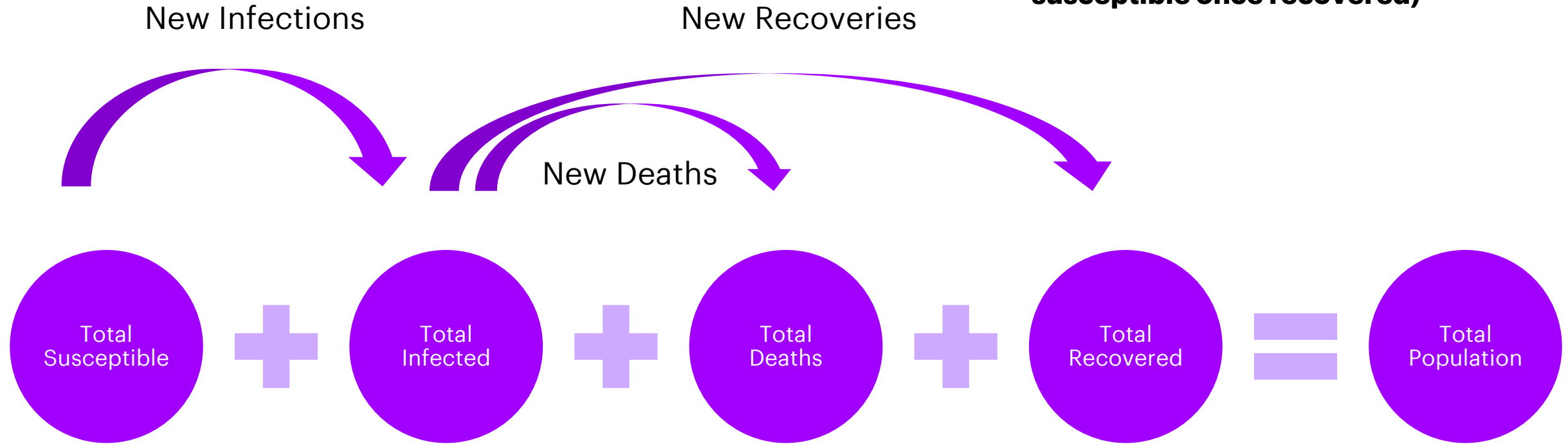
Driver Level 2

Reduce the death rate once infected

Reduce the number of infections

Domain Knowledge – SIR Modelling

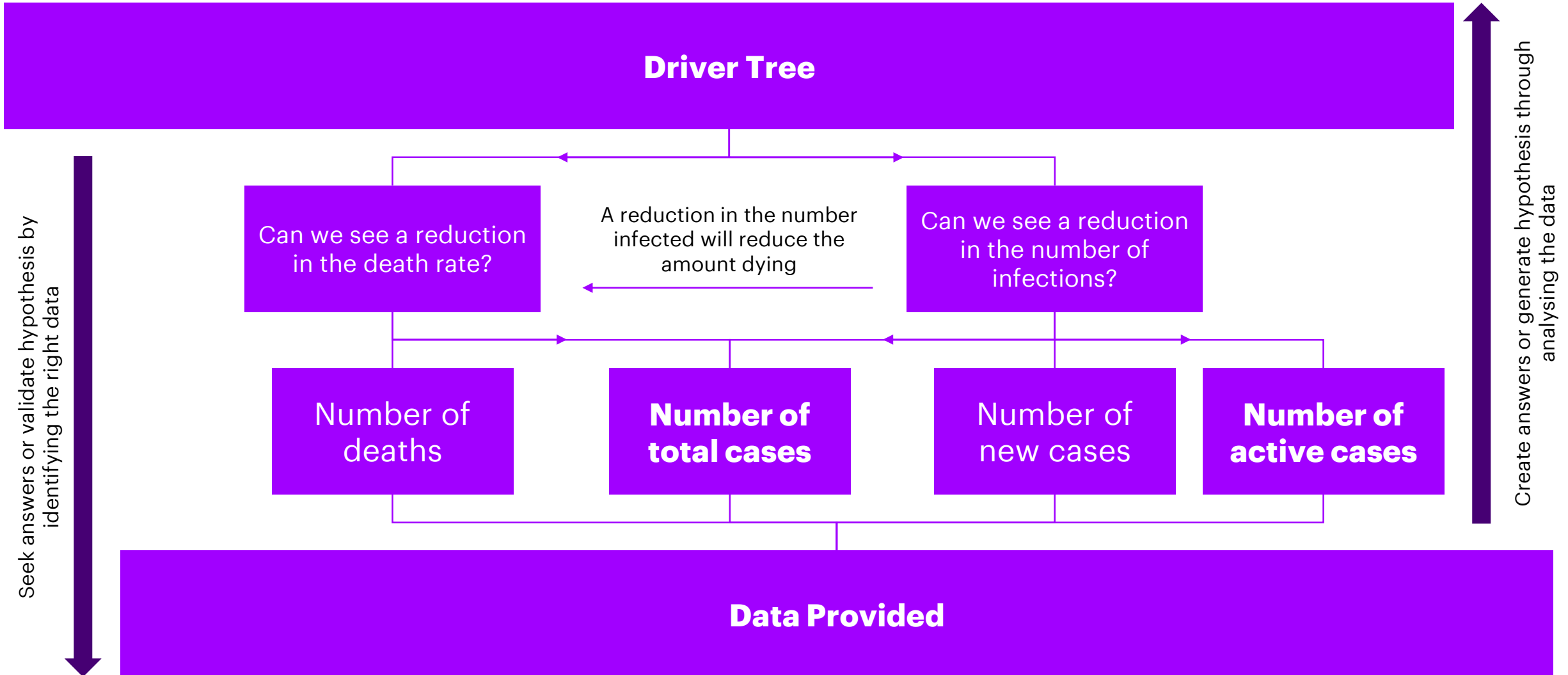
For simplification, please assume that those who have recovered can not get the virus again (i.e. they are not susceptible once recovered)



New Infection Rate (i.e. % of People who are likely to be infected) is dependent on the amount of people who are susceptible, the number of people who are currently infected and the success of measures in minimising the spread of the virus (i.e. social distancing between susceptible and infected people)

Recovery Rate (i.e. % of People who will recover if infected) is dependent on environmental conditions including the quality of healthcare available to treat those with the virus. **Death Rate** is the inverse of this (i.e. as Recovery Rate increases Death Rate decreases)

Top Down and Bottom up Analysis



**WHAT CHART
WOULD WORK
BEST TO PROVIDE
INSIGHT?**

So you've got the data you want – how do you present it? (1)

Deviation

Emphasize variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (continuous/magnitude).

Example FT uses

Trade surplus/deficit, climate change

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar

Perfect for presenting survey results which include sentiment (eg disagreement/agree).

Spine

Splits a single value into two contrasting components (eg confidence).

Surplus/deficit lined line

The standard way of these charts allows a means to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationship you show them to be causal is one causes the other.

Example FT uses

Inflation and unemployment, income and life expectancy

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line timeline

A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot

Usually used to show how the relationship between 2 variables has changed over time.

Bubble

Like a scatterplot but adds additional detail by coding the bubbles according to a third variable.

XY heatmap

A good way of showing the patterns between 2 categories of data. Use effective at showing the difference in amounts.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses

Health, education, league tables, constituency election results

Ordered bar

Standard bar charts display the rank of values much more easily when ordered into order.

Ordered column

See above.

Ordered proportional symbol

Use when there are big variations between values and/or seeking to show where data is not so important.

Dot strip plot

Dots placed in order on a strip are a space-efficient method of laying out data across multiple categories.

Slope

Perfect for showing how rates have changed over time or size between categories.

Lollipop

Lollipop charts draw more attention to the data value than standard bar/columns and can also show rank and value effectively.

Bump

Effective for showing changing rankings across multiple dates. For large datasets, consider grouping lines using colour.

Distribution

Show values in a dataset and how often they occur. The shape can 'show' if a distribution is skewed or has a bimodal shape. Highlighting the lack of uniformity or equality in the data.

Example FT uses

Income distribution, population, Olympic medal counts, revealing inequality

Histogram

The standard way to show a statistical distribution – using the gaps between columns used to highlight the shape of the data.

Dot plot

A simple way of showing the range (minimum) of data across multiple categories.

Dot strip plot

Good for showing the distribution, can be a problem when many data have the same value.

Barcode plot

Like dot strip plots, good for displaying all the data in a single, they work best when highlighting individual values.

Boxplot

Summarise multiple distributions by showing the median, quartiles and range of the data.

Violin plot

Similar to a box plot but more effective with complex distributions. Good for comparing low and high-low points of each data.

Population pyramid

A standard way for showing the age and sex distribution of a population. Distribution effectively laid to back histograms.

Cumulative curve

A good way of showing a cumulative distribution. A plot is almost cumulative. Represents a value is shown a measure.

Frequency polygons

For displaying multiple distributions of data. Like a regular line chart but instead of a maximum of 2 or 4 datasets.

Beeswarm

Use to emphasise individual points in a distribution. Points can be coded to an additional variable. Best with medium-sized datasets.

Change over Time

Give emphasis to changing trends. These can be short (one-day) movements or slower trends over time. Highlighting the trend or direction. Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses

Home price movements, economic time series, national changes in market

Line

The standard way to showing change over time. A line chart is best for showing a range of data points.

Column

Columns work well for showing change over time – but usually best with only one series of data at a time.

Column + line timeline

A good way of showing the relationship over time between an amount (columns) and a rate (line).

Slope

Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart

Use with care – these are good at showing changes in total, but showing change in composition can be very difficult.

Candlestick

Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fun chart (quasiplot)

Use to show the uncertainty in future predictions – usually this grows the further forward in projection.

Connected scatterplot

A good way of showing change over time for two variables when there is a relatively clear pattern of progression.

Calendar heatmap

A great way of showing temporal patterns (data weekly/monthly) – if the x-axis is showing prediction in quantity.

Priority lines

Great when date and duration are an important part of the story in the data.

Circle timeline

Good for showing discrete values of varying sizes across multiple categories (eg attributed to by comments).

Vertical timeline

Presents time on the Y axis. Good for displaying detailed time series that work especially well when scrolling on mobile.

Salamander

Another alternative to the circle timeline for showing time series where there are big variations in the data.

Streamgraph

A type of area chart, use when seeing trends in proportions over time is more important than individual values.

Magnitude

Show size comparisons. These can be relative (not being able to see large/small) or absolute (need to use the differences). Usually these show a 'counted' number (for example, items, data or people) rather than a calculated rate or per cent.

Example FT uses

Commodity production, market capitalisation, volume in general

Column

The standard way to showing the size of things. Start always at 0 on the axis.

Bar

See above. Good when the data are not those series and labels have long category names.

Paired column

As per standard column but allows for multiple series. Can be used to compare the size of 2 series.

Paired bar

See above.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Proportional symbol

Use when there are big variations between values and/or using the difference between data is not so important.

Isotype (quasiplot)

Isotype solution is some instances – use only with whole numbers. Do not allow off an axis to represent a decimal.

Lollipop

Lollipop charts draw more attention to the data value than standard bar/columns – does not have to start at zero (if preferred).

Radar

A space-efficient way of showing value of multiple variables but makes sure they are organised in a way that makes sense to reader.

Parallel coordinates

An alternative to radar charts – again, the arrangement of the variables is important. Usually benefit from highlighting values.

Barlet

Good for showing a measurement against the context of a large set or performance range.

Grouped symbol

An alternative to bar/column charts when being able to count data or highlight individual elements is useful.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is in the size of the components, consider a magnitude-type chart instead.

Example FT uses

Political budgets, company structures, national election results

Stacked column/bar

A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Marimekko

A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Pie

A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Donut

Similar to a pie chart – but the centre can be a good way of adding more information about the data (eg title).

Tree map

Use for hierarchical data. For large datasets, it can be difficult to read when there are many small segments.

Venn

A Venn diagram, often used for visualising set theory, is a composition for number of data.

Gridset

Good for showing % information that can be split into whole numbers and work well in most multiple layout forms.

Venn

Generally only used for schematic representation.

Waterfall

Can be useful for showing part-to-whole relationships where some of the components are negative.

Spatial

Aside from location maps only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses

Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (rank/ratio)

The standard approach for showing data on a map – should always be ratio rather than totals and use a perceptible base geography.

Proportional symbol (count/magnitude)

Use for kinds where there are many small differences in data will be hard to see.

Flow map

For showing movement across a map.

Contour map

For showing areas of equal value on a map. Can use deviation colour schemes for showing + or - values.

Equalized cartogram

Counting each unit on a map to a regular and equal-sized shape – good for representing cities with equal value.

Scaled cartogram (value)

Stretching and distorting a map so that each area is sized according to a particular value.

Dot density

Used to show the location of individual events/locations – make sure to include any gathering the reader should see.

Heat map

Grid-based data values mapped with an intensity colour scale. As choropleth map – but not dependent on administrative unit.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses

Movement of funds, trade, migrants, beauty, information relationship graphs

Sankey

Show changes in flows from one condition to all learn one other good for tracing the eventual outcome of a complex process.

Waterfall

Designed to show the sequencing of data through a flow process. Typically budgets. Can include +/- components.

Chord

A complex but powerful diagram which can display 2-way flows (not just 1-way) in a matrix.

Network

Used for showing the strength and inter-connections of relationships of varying types.

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith, Chris Goodwin, Ben Hill, Lyle Francis, Thomas Pridmore, Billy Dunningham, Sharon Pugh, Phil Collins, Martin Smith. Inspired by the Graphic: Contributions by Jon Cookson and Swanton Edwards.

ft.com/vocabulary

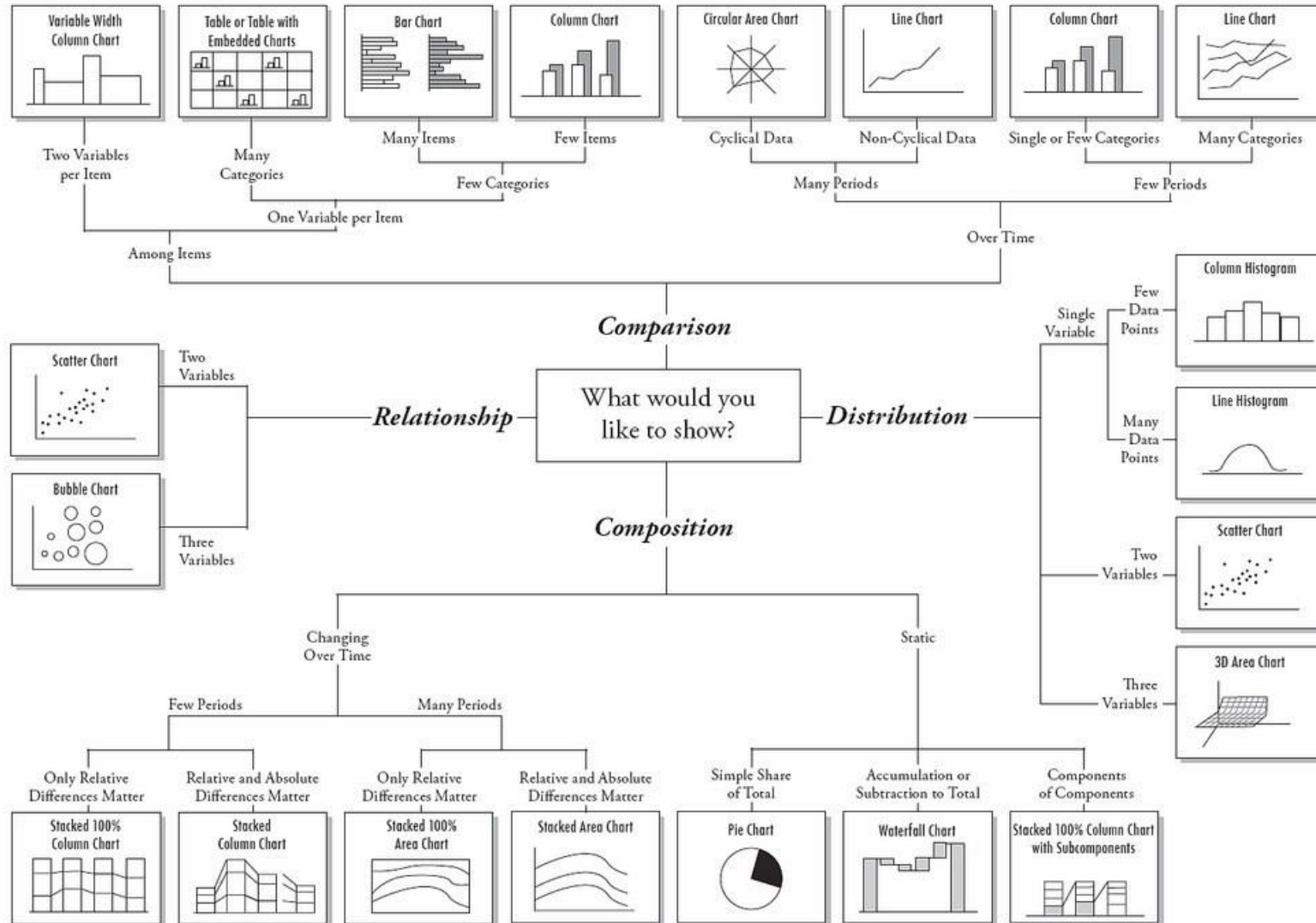
© Copyright Times 2014-2020
The work is licensed under a Creative Commons
Attribution-ShareAlike 4.0 International License.

More details [here](#)

Copyright © 2020 Accenture. All rights reserved. 21

So you've got the data you want – how do you present it? (2)

Chart Suggestions—A Thought-Starter



**WHAT IS THE BEST
WAY TO PRESENT
OUR FINDINGS?**

But what does the user want to see?

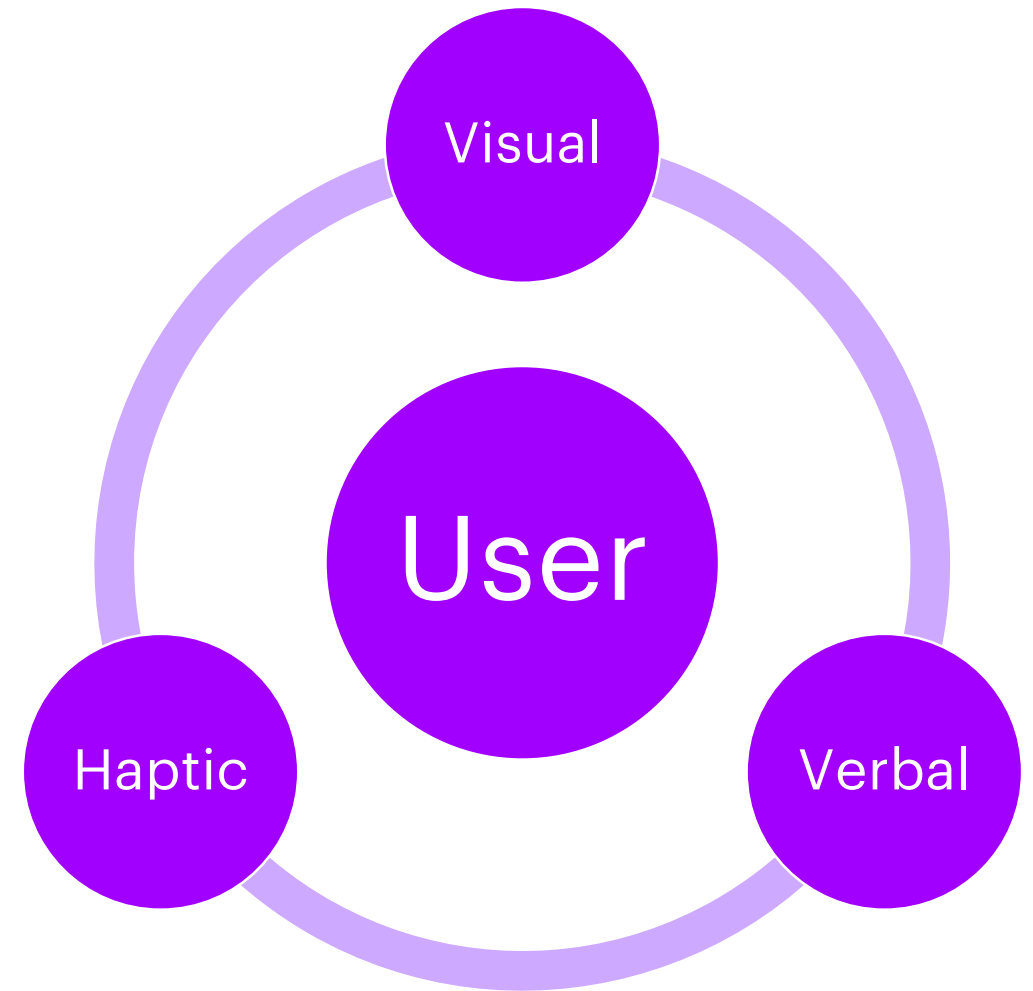
Augmented Insights is the field of analytics that focuses on the presentation of data in the best way possible for consumption for the end user

Remember when you were in primary school, you learnt things in one of the following ways:

- Visually – learning through seeing
- Verbal – learning through communicating
- Haptic – learning through interacting

Think about this, you likely have a preference for how you learn. It may not just be one of these ways but a combination.

In the same way, when we present data we must consider how our end user(s) best learn from the data and build the best insight delivery mechanism for them

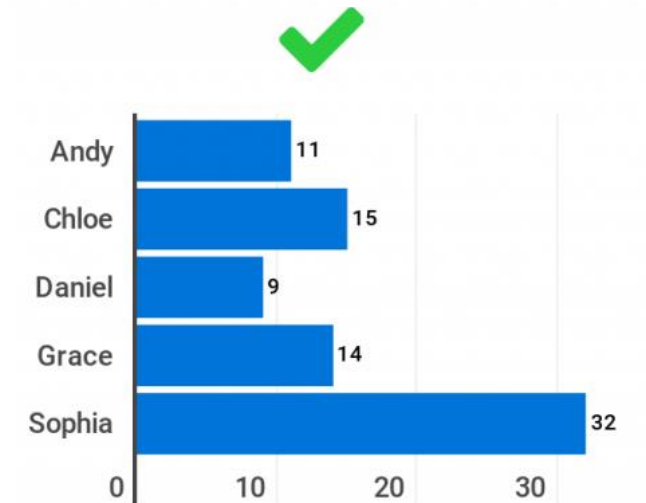
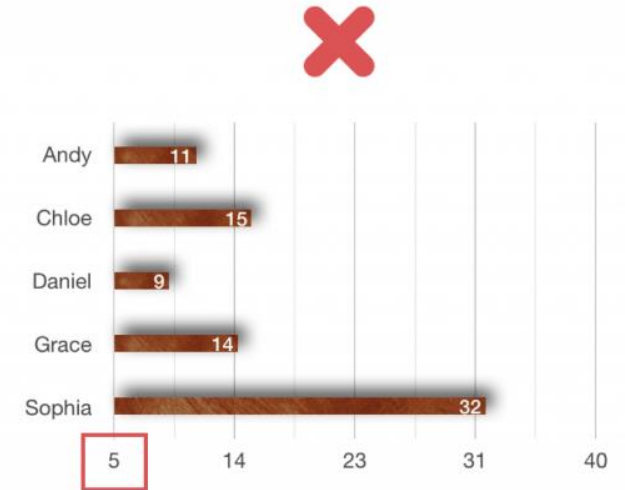


**HOW MIGHT WE
MAKE THIS EVEN
MORE ENGAGING?**

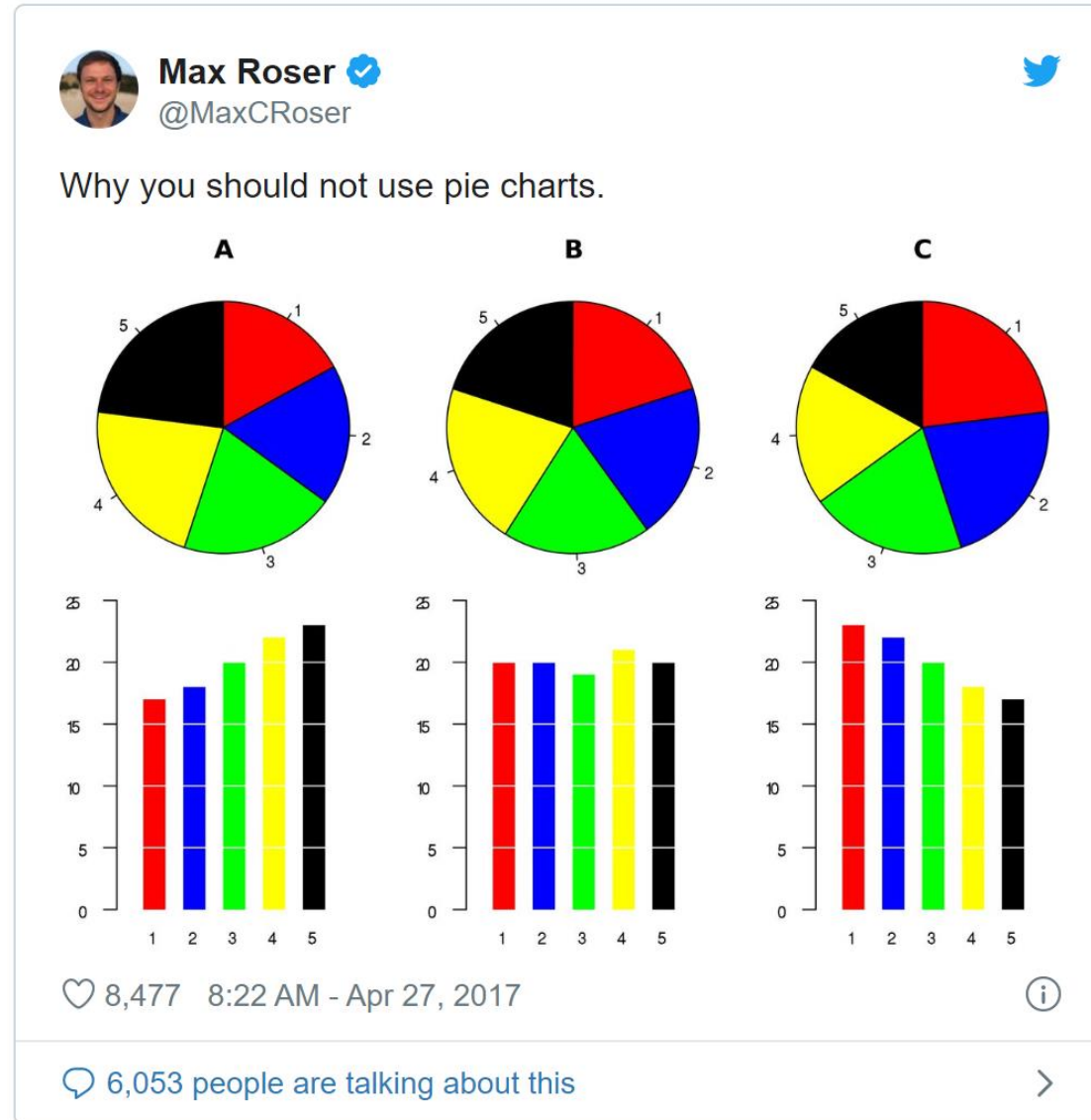
All the small things...

When designing your chart consider the following:

- Is the point your trying to make from your chart self-evident (i.e. would the client be able to understand the point your trying to make without you having to explain it)
- Try to include a chart title to help explain what is shown on the chart
- Try to include a axes titles to help explain what is shown on the chart
- Make sure that the axes are easy to read
- Always start your axes at zero and use set intervals between your major and minor axes
- Minimise how distracting your axes are (remove minor axes and use lighter colours)
- Use data labels to explicitly call out figures, make sure data labels are clear to read
- Don't use 3D elements (i.e. shapes or shadows)
- Avoid pie charts (see the next slide for evidence)
- Make sure that the axes is ordered in a way that helps to demonstrate your point (i.e. sort from largest to smaller or alphabetically)
- Does the chart align to the branding of the client (i.e. colours and text formatting)?
- Use bold colours sparingly to highlight the important features of your chart (consider if your audience may be colour blind though – if so check to see if your chart works in grayscale)



Max is smart. Be more like Max.



**WHAT IF WE WERE
ASKED TO REPEAT
THIS ANALYSIS?**

You've done it! Now do it again...

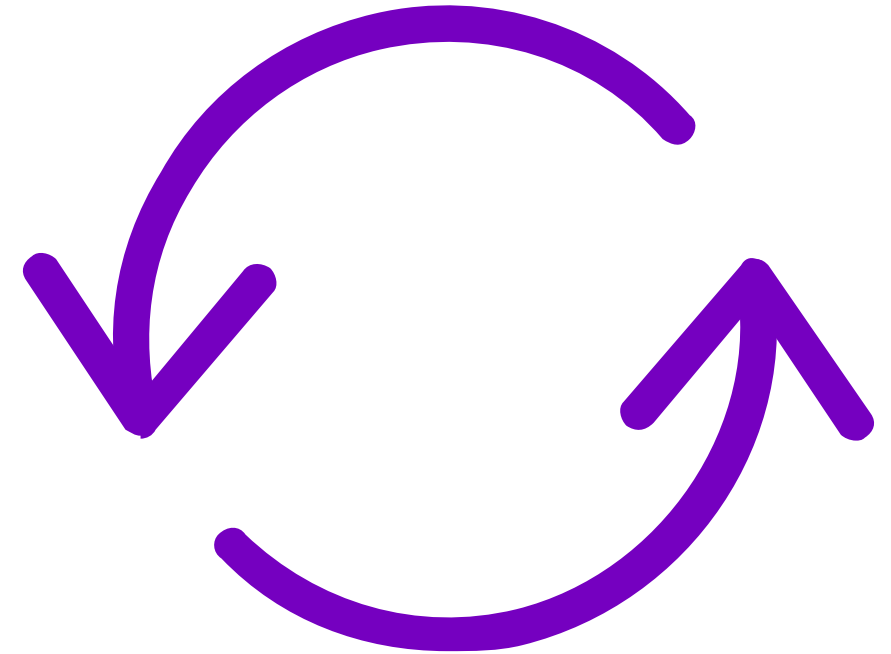
If your solution is good and is relevant for more than a one time analysis, expect to be asked by your client to produce your analysis again!

In preparation for this consider the following:

- Can you add new data (more row, but the same columns) where you extracted your original data set to?
- Can your model (transformation) work again if new data?
- Does the preparation of your data work as expected with this new data?
- Is this new data presented in a way that acceptable (i.e. still satisfies all the presentation rules)

Some tips for this:

- Leverage relative references or absolute references that stretch beyond the number of rows in your current data set
- Try not to 'hard code' parameters into your model or your presentation, leave dynamic if possible



**WHAT COULD OUR
USER WANT FROM
US IN THE FUTURE?**

Making your analysis modular and ready for change

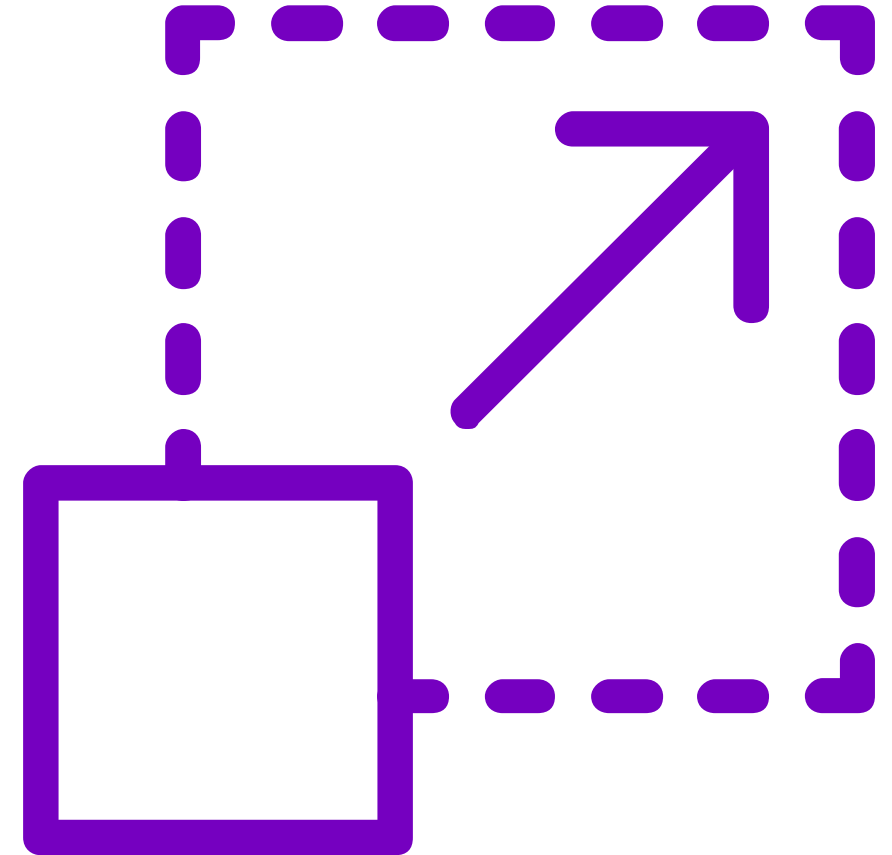
Some things in life never change... Business requests and Data are some one of those things. Either your user will request new views which can hopefully use the same data or may require new data or the data itself will change but the user will want the same view.

In preparedness for this, consider the following:

- Can you add new data (more columns and rows) where you extracted your original data set to?
- What do these new columns do to your transformation and data preparation? (expect some #errors)
- Does the choice of presentation of the data still work with this new data?

Flexibility of your data solutions are key to handle these changes, in Excel please consider using:

- Data Tables (*Ctrl + T* on your selected data)
- Be sure to reference field/column names not row and column IDs
- PivotCharts and PivotTables
- The Refresh Data button (*Ctrl + Alt + F5*)



So what have we achieved?

If you have followed the steps above, we have created a meaningful presentation of insights that:

- **Explicitly addresses the key question or validates a hypothesis in a clear manner**
- **Is repeatable**
- **Is scalable**

What about the case though?

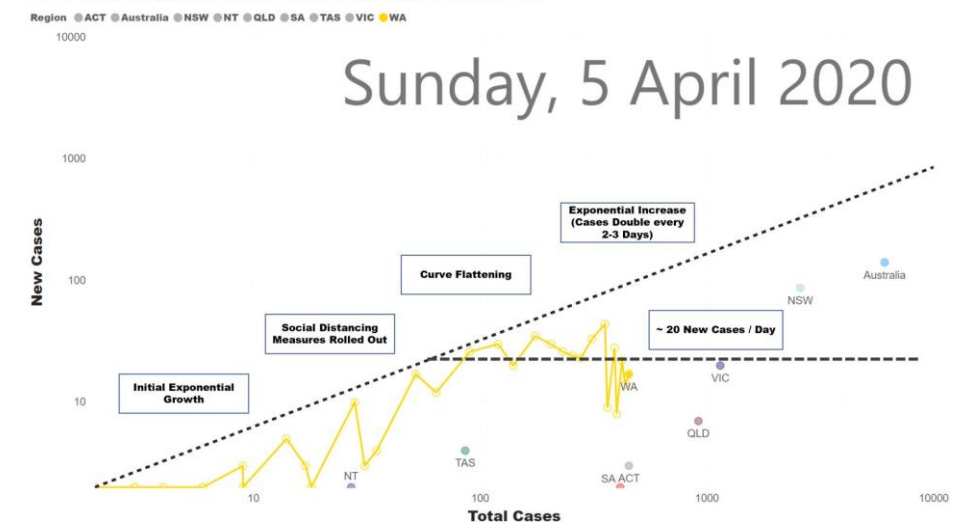
The pre-prepared graphs are an advanced view of the data. The chart plots (each day is a point) the trajectory of the ratio of New Cases against the Total Cases on a set of logarithmic axes.

The charts show that initial growth in WA was exponential in nature but social distancing has likely led to curve flattening and therefore a reduction of risk that the hospital system will be overwhelmed.

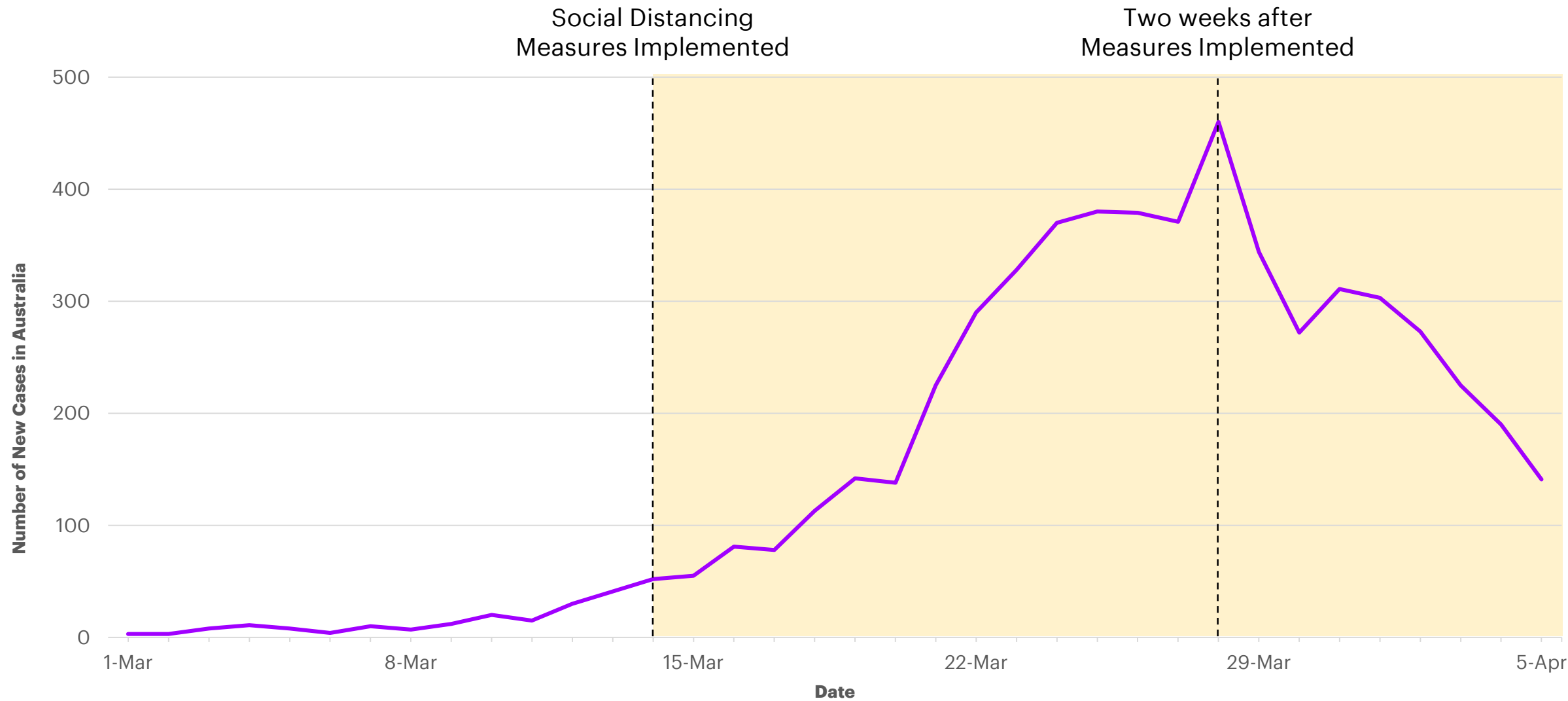
New Cases against Total Cases by Region and Date



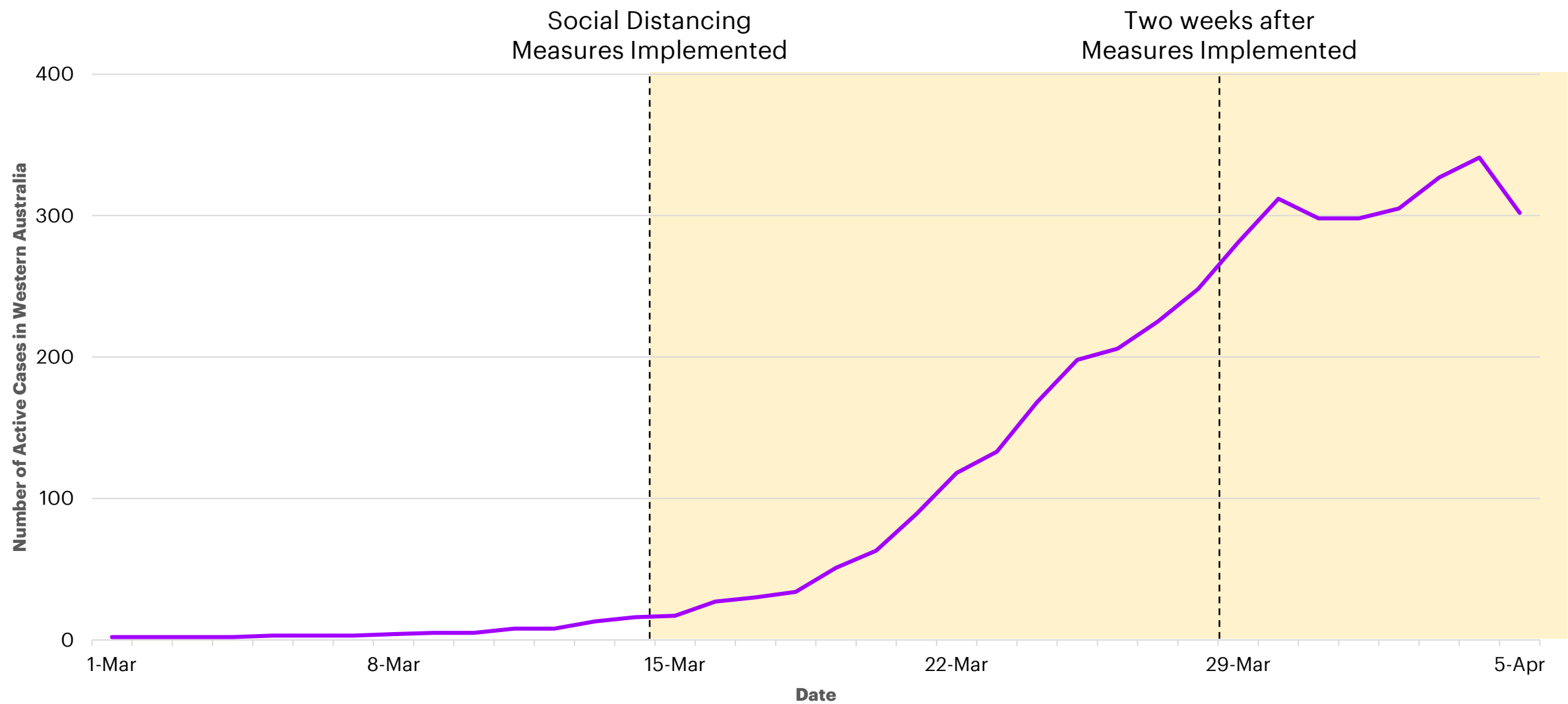
New Cases against Total Cases by Region and Date



The number of new cases for COVID-19 appears to be decreasing in Australia as a consequences of social distancing measures being implemented



The number of active cases in Western Australia appears to be stabilising. Showing that number of infections is reducing and number of recoveries is increasing.



Summary of approach

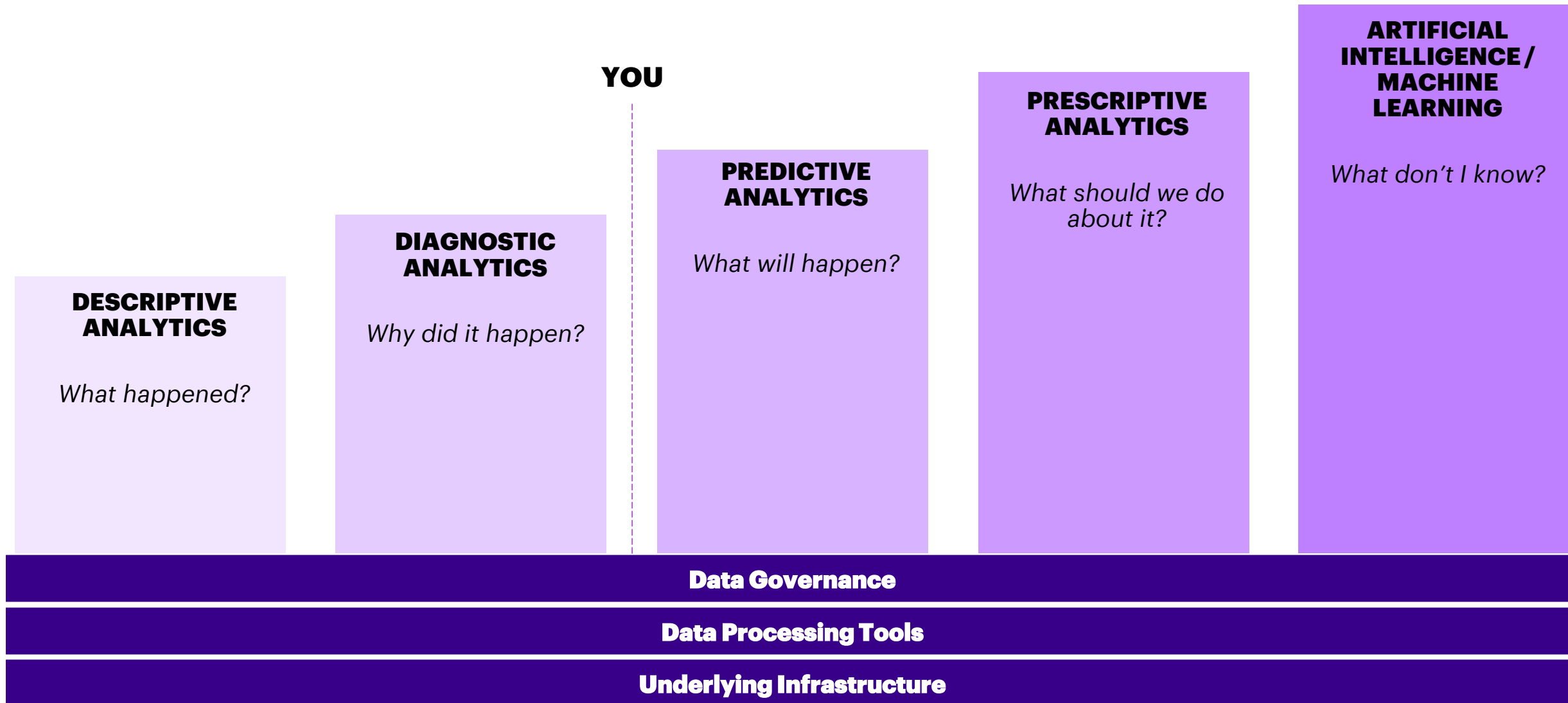
1. Define the problem you are trying to solve or identify a hypothesis you wish to test
2. Think about how the consumer of your findings may want the solution presented
3. Identify data (be it one data set or many) and validate through mapping to the question/hypothesis that may help in answering the question or proving the hypothesis
4. Perform data transformation and preparation
5. Create your initial presentation of the solution
6. Answer the question or validate the hypothesis
7. Test this with your user and iterate if needs be
8. As required – improve the presentation of the solution and build avenues for the solution to be repeatable and scalable

Words of wisdom

- “**Focus on Simplicity**, ability to see the details (easy and quickly), ability see information along different dimensions.”
- “**The view must be co-created with the end user** - but be prepared and bring your own thinking to challenge conventions. Some users need more details than others on a single screen (and depending on device). E.g. field worker on tablet vs clinician at workstation.”
- “**Try to understand the end-user**, walk in their shoes, live their pains and frustrations, and think about where this information adds value. And if it doesn't, then the job wasn't done right.”

**WHERE ARE WE
NOW AND WHERE
TO NEXT?**

So where are we on our journey?



What happens when data grows up?

Expect more mature use cases driven by improvements in underlying technologies, reducing the cost required to implement solutions to find value and improve human experiences. Some key use cases in some relevant areas include:

Finance / Commerce – Fraud Detection

Marketing – Personalised Marketing

HR – Recruitment Analysis and Automation

STEM – Energy Optimisation

Health – Early Diagnosis

Law – Paralegal Automation

Arts – Self-Learning Composition and Creation

Operations – Robotic Process Automation (RPAs)

There are current and emerging ethical issues around the use of data, not limited to the sharing of private information (i.e. GDPR) and ethical artificial intelligence.



Why does this all matter to you?

01

Imagine a world without data again...

02

What career are you heading towards – how will technology and data impact this in the short term, mid term or long term?

03

How will data impact (and hopefully) improve your life?

**INTERESTED IN
DOING MORE?**

Your Graduate Journey



Thanks.

PERTH
INNOV
ATION
HUB ●

accenture