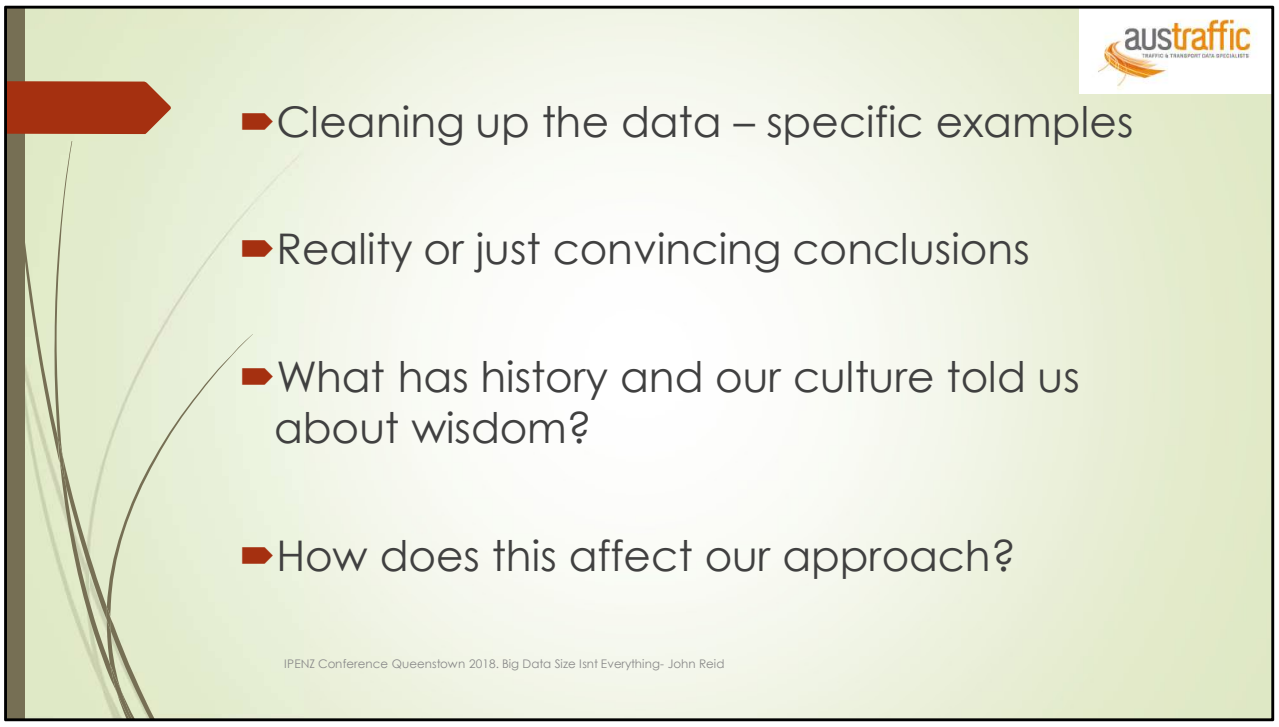
G'day

In binary code
The zeros and ones

Who noted that I actually dropped the capitalisation of the G ?

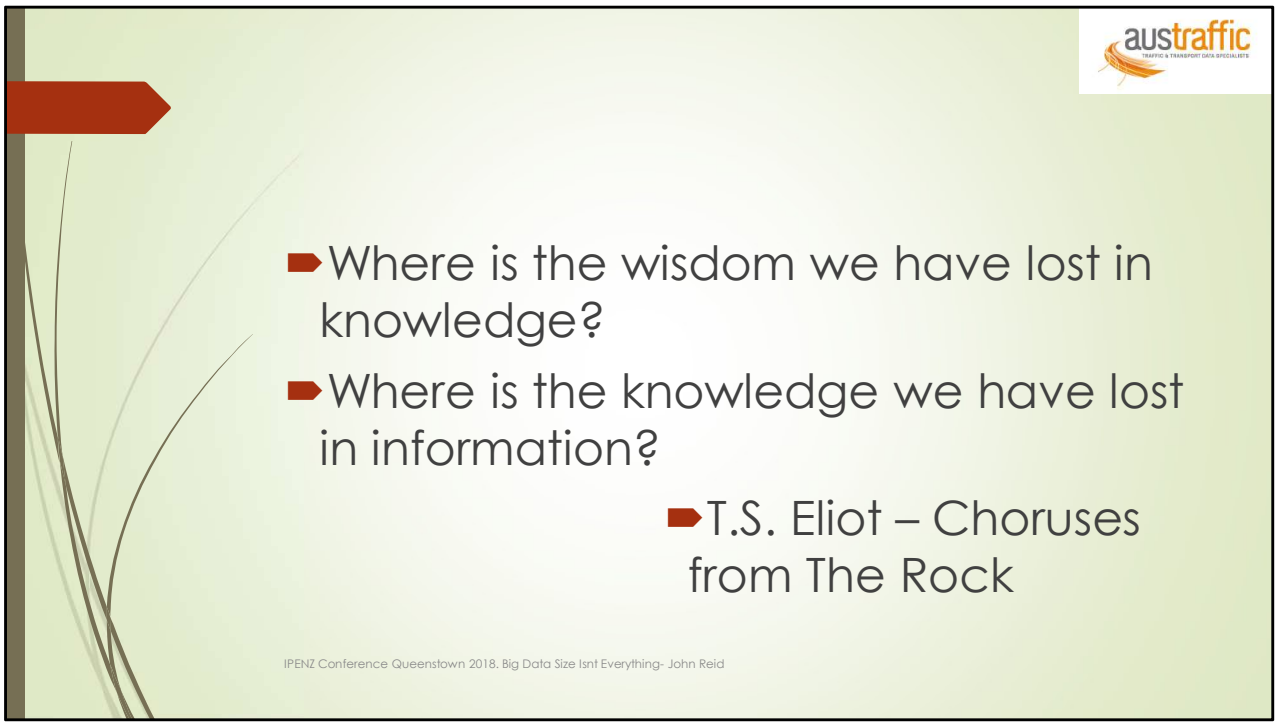
Does it matter, just a different sequence of one and zeros ?
Or is that zeros and ones ?



In this presentation and in the paper I touch on four issues:

- Cleaning up the data – specific examples of data errors
- Data Reality or just convincing conclusions
- What has history and our culture told us about wisdom?
- How does this affect our approach?

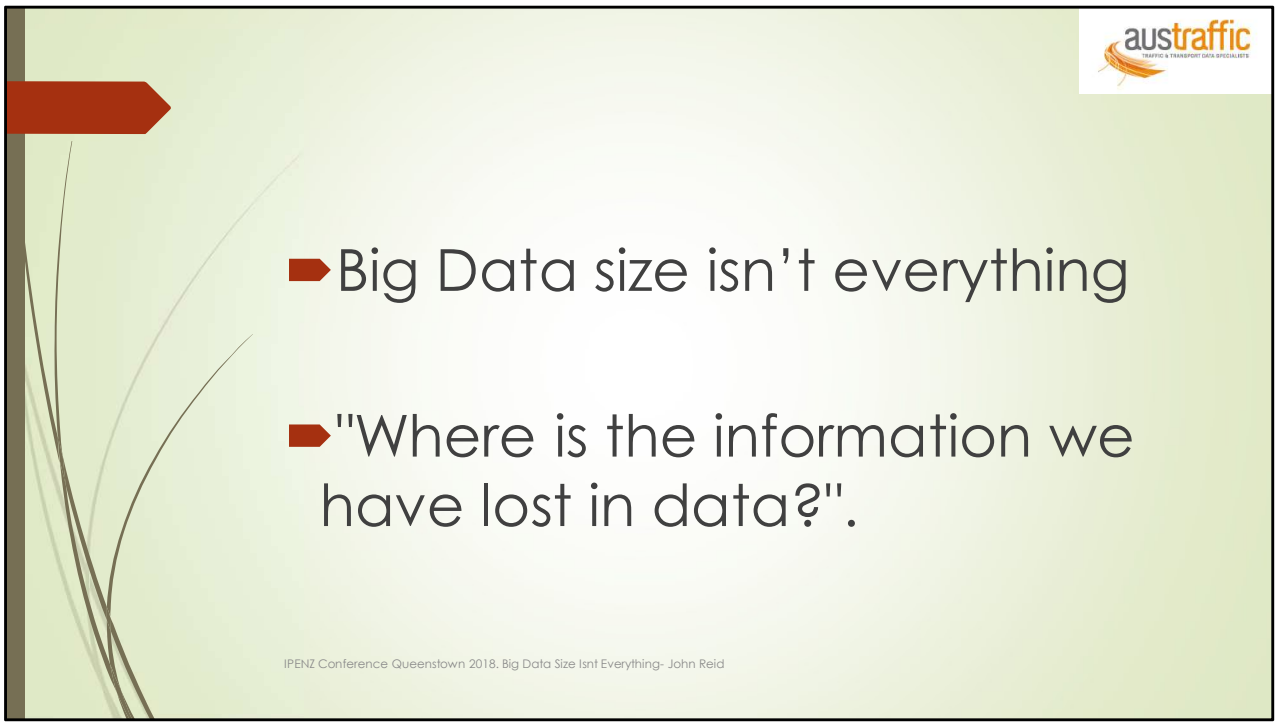
The paper was prompted by a quote of TS Eliot from 1934



Where is the wisdom we have lost in knowledge?

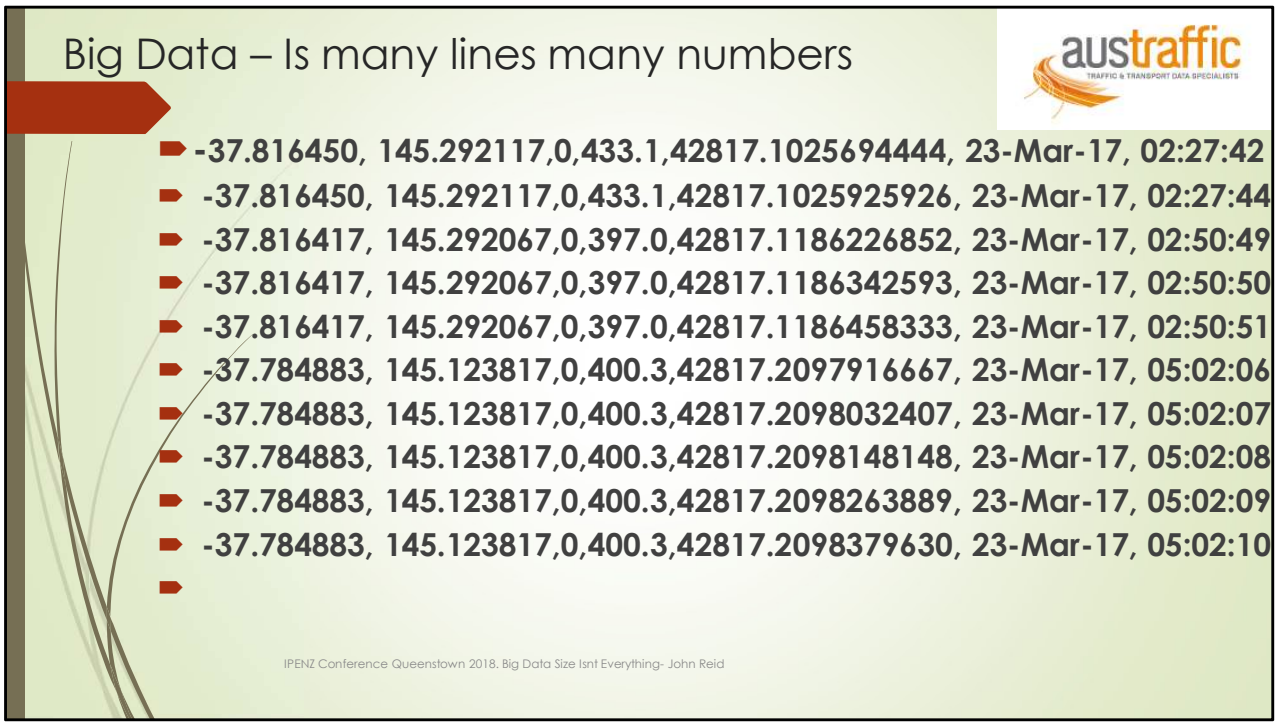
Where is the knowledge we have lost in information?

T.S. Eliot – Choruses from The Rock



With a great deal of diffidence may I add a line to TS Eliot's reflection

"Where is the information we have lost in data?"



Big Data
We are being bombarded, swamped with data.
What do the numbers mean ? Are they credible numbers ?
Any IPENZ & AITPM conference will undoubtedly be full of numbers and results.
This is not necessarily a bad thing.
But I wish to take you on a quick journey of caution and application of wisdom to big data.
Big Data has a big role in the design and management of our future.
Big data sets alone do not provide us wisdom
Big data on first flush, may reflect a sewer.
We can slice it , we can dice it.

Are we spending enough time to know what we are getting ?
to ensure it is what we really want and to blend that into the best community outcomes with all the wisdom we have?

I am not saying that big data is automatically wrong.

Far from it, big data is the present and the future.

But big data is rarely a set of correct or “perfect” numbers.

We need to know what we are getting and we need to get rid of the imperfections.

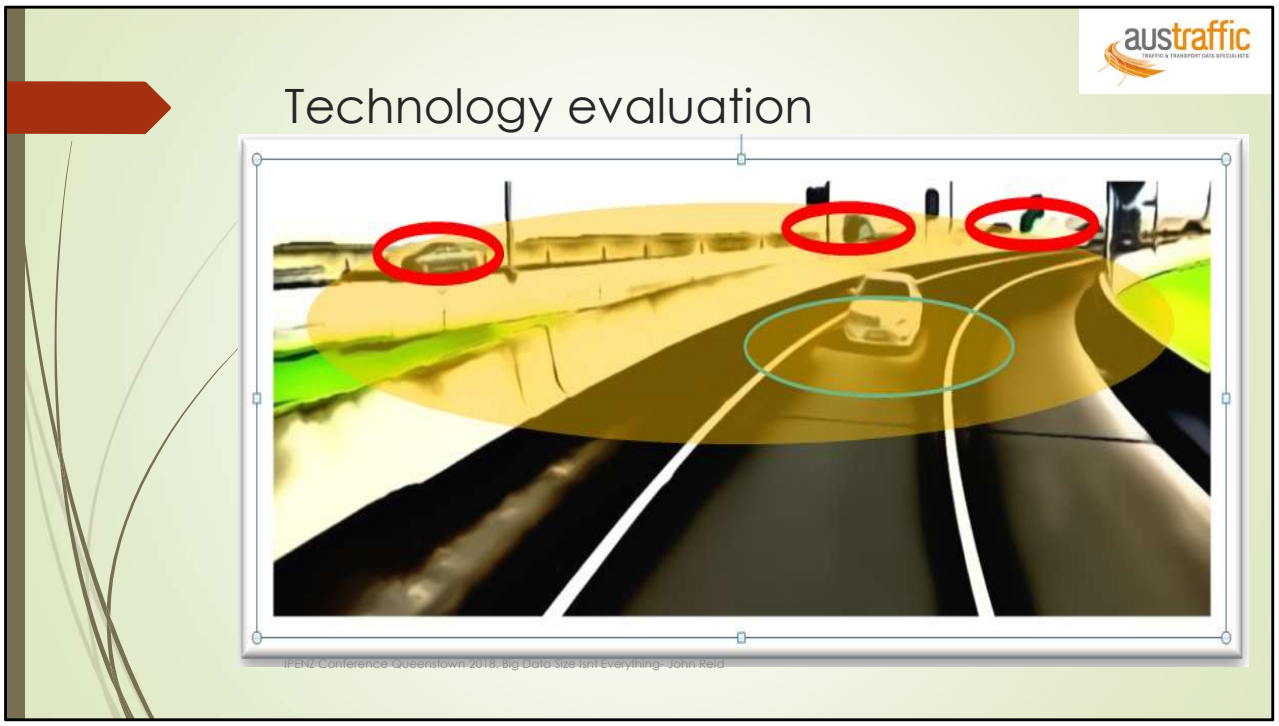
For those with skills to read binary, you may smile, for the rest just interpret the emoticon,

Having a huge volume of numbers does not necessarily iron out the problems. It can bury them.

Big data can soon lead to a conclusion which becomes a slogan that ends up being conventional wisdom. And the conclusion can be wrong.

Before you start to make use of data we need to know what has been recorded, that it is measuring and reporting what is actually happening, and what is the local “noise”, the data outliers.

Big data is a big task

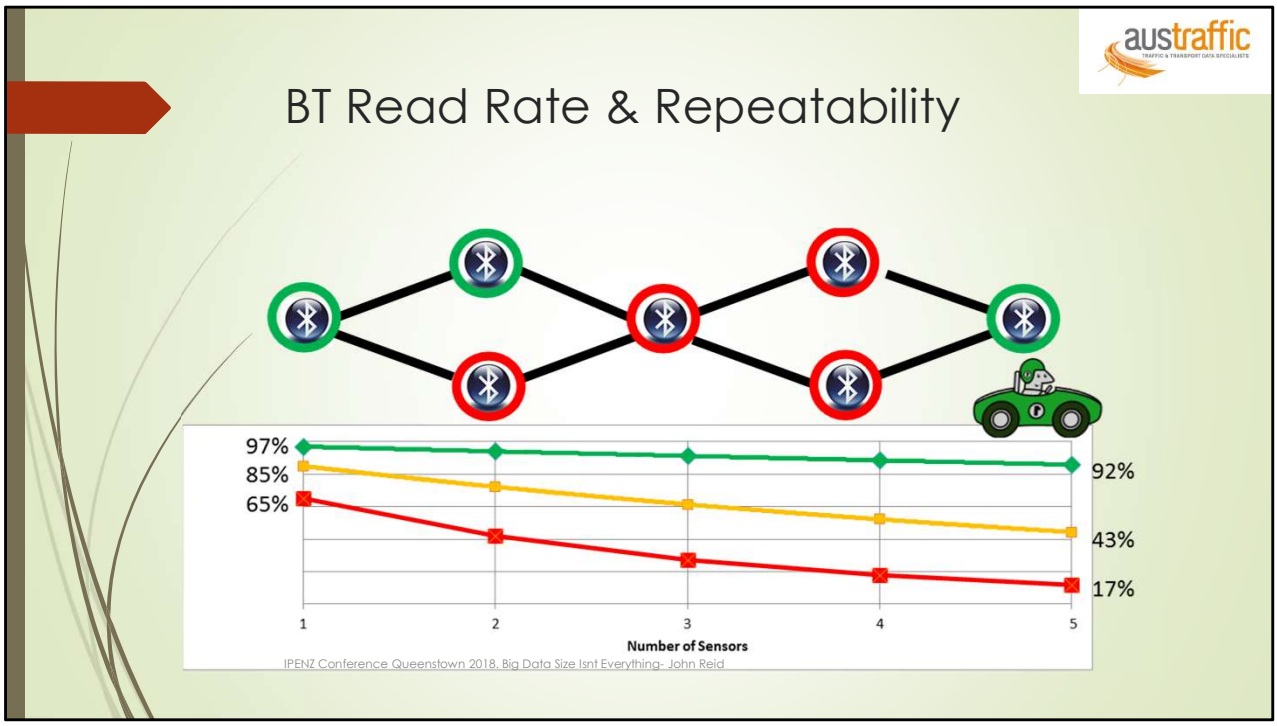


At the AITPM National Conference in 2013 Scott Benjamin , presented research on the initial use of using Bluetooth devices to measure the origin and destination of vehicles by recording the presence of travelers' devices as they passed various points.
Where motorway roads and ramps were near each other, there were specific interference issues. Origin/destination trip assignments could be distorted by vehicles travelling on nearby facilities.
The picture is of a freeway interchange highlighting in red the *Extraneous sensed probes not in the corridor of interest (Green),*

Bluetooth counting is giving us much more data., but is it an accurate reflection.

So, the results can be more revealing if we spend time to understand what they mean.

So is Bluetooth always much better than the old method of a few cars collecting travel times. Or is it? Could both offer some different insights.

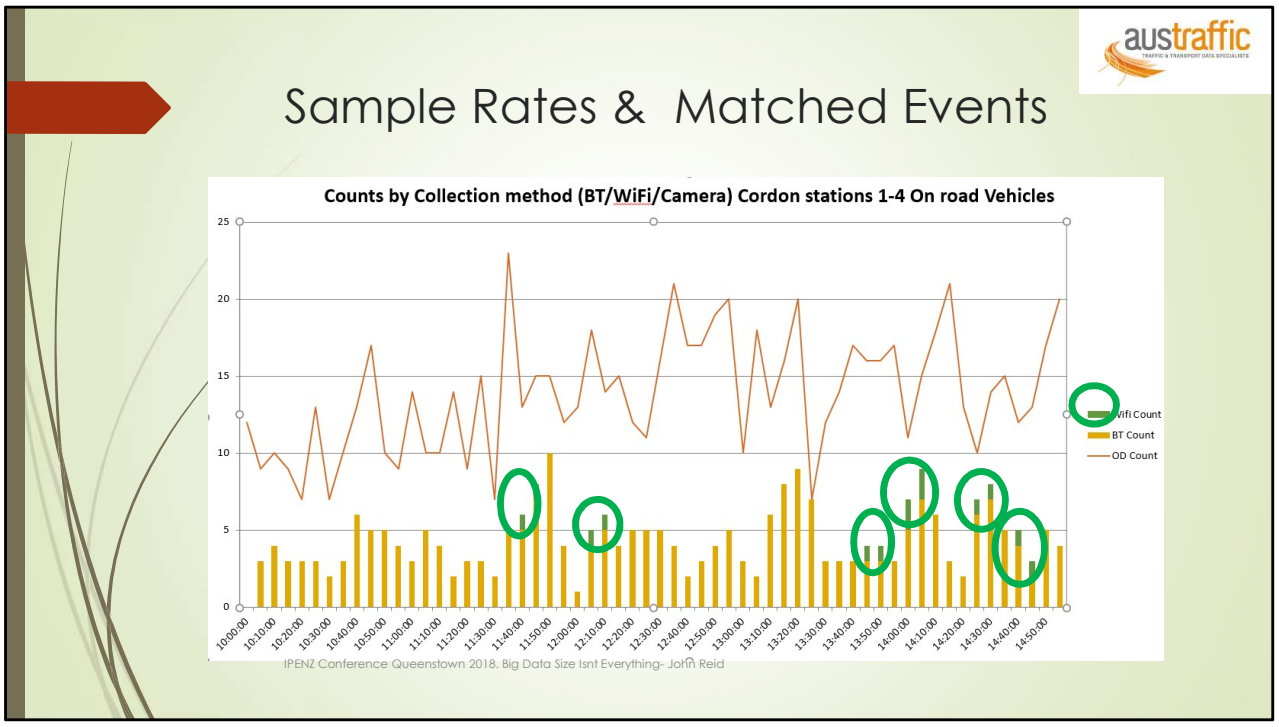


Austraffic began reviewing a range of BT Solutions 6 years ago. We reviewed a range of technologies, each for between 3 to 6 months on a range of standard test sites, which included local roads, urban arterial roads, rural arterial road, and metropolitan freeways.

We sought to bench mark each system against a common set of measurements by validating BT device read rates and comparing to automatic tube counters and video count and number plate surveys; we compared tens of thousands of records during the course of the validation of the following:

- Read rate
- Repeatability

The attempt to produce OD trip matrices underscores the virtue or constraints of those BT sensors that have poor Read rates and Repeatability

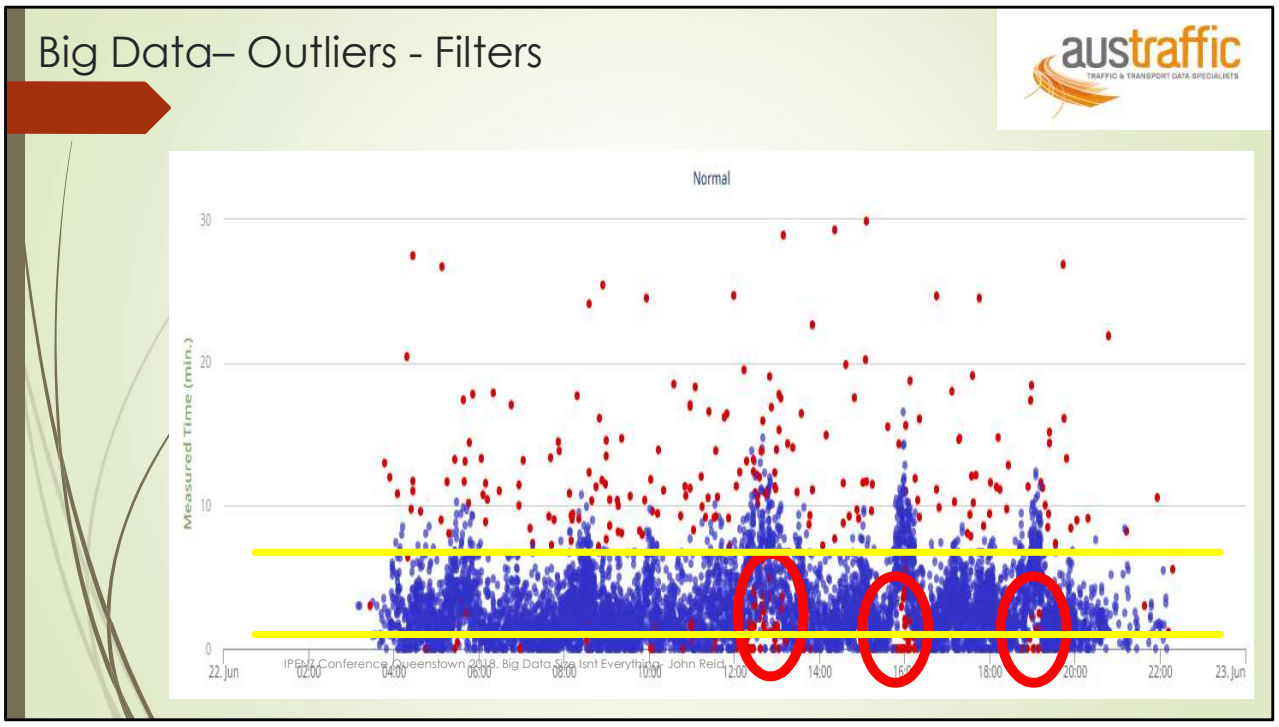


Austraffic's on road experience also highlights the vagaries of sample rates between probe point sensors
In this instance to match events along a freeway corridor and it's arterial branches.
The green circles are to highlight the few wi-fi events.
This relatively low volume match rate is shown to highlight the variations between a camera derived count versus that from Blue Tooth and WiFi.
Wi fi for on road data collection is of dubious consistency as compared with either camera or BT.

In the July 2017 AITPM newsletter I commented on how we often set up Bluetooth recorders at key nodes and measure the average travel time between the two points, covering traffic across multiple lanes.

However the virtue of a GPS based survey, is that we can look at individual lanes in second-by-second detail.

This is a key issue if you are looking at how cars from a turning bay might queue out into the main stream of traffic blocking that lane. This is critical information for designing the signal phasing and the turning bay capacity.
This is data, that gets lost in a big data probe point data set.



Next is an example of outliers (data records that are beyond the expected range) with probe sensed data, it could be the same graphic for any probe sourced data set be it Cellular, BT, WiFi, or navigational

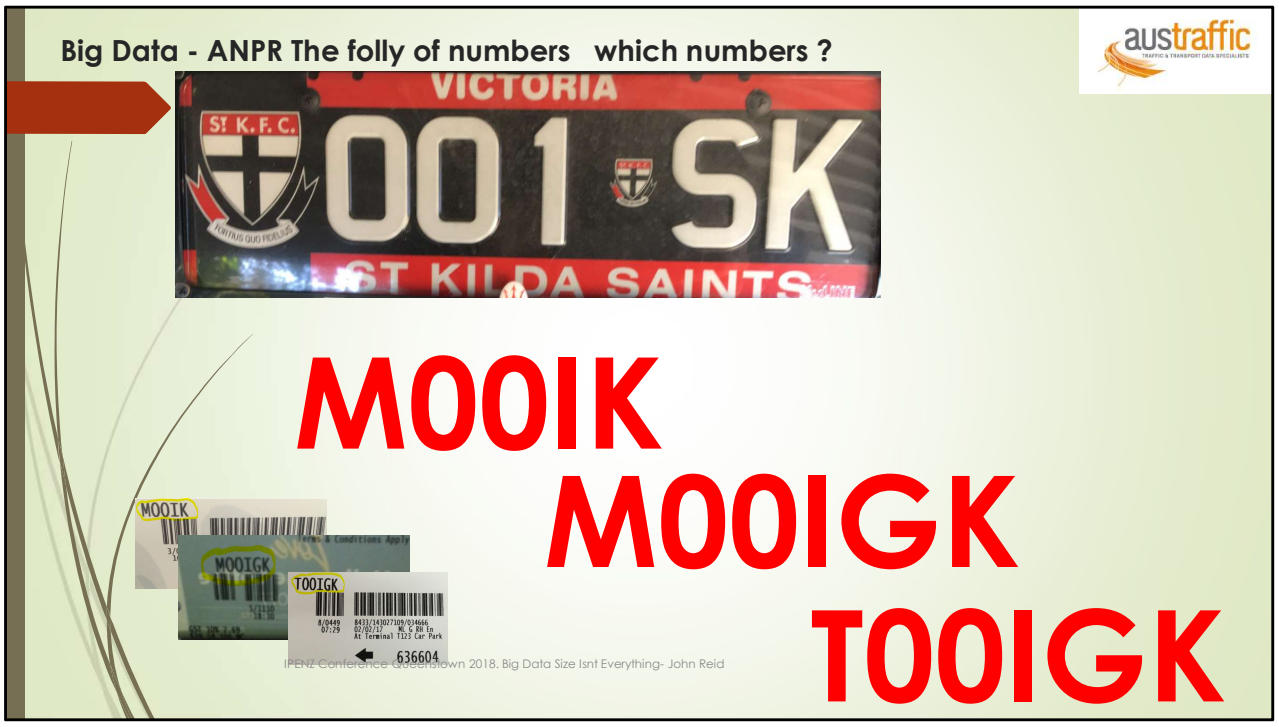
This data is from static BT sensors, pairs of sensors that form to provide bidirectional zone based data along the corridor of interest.

In this case the experts at Blip Track have built a very sophisticated filter engine that recognises the outliers for the task at hand. Note outliers that exist high and low on the plot chart are the RED dots.

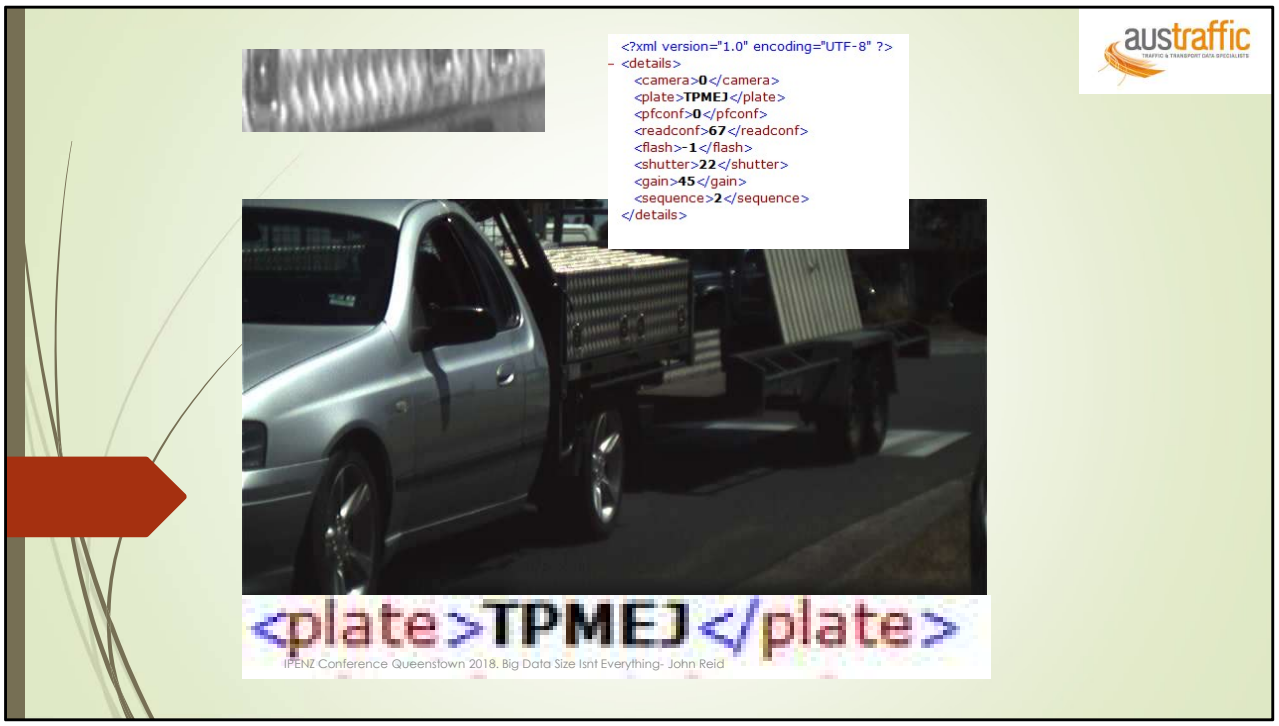
Typical conventional banding techniques eg the yellow lines, may cull most of the high outliers. While culling large travel durations that were real. The higher order skill is making a call on the short duration red dots close to the horizontal axis. As circled in Red

Quality data cleansing is imperative to empowering todays and tomorrows data scientist with credible data sets that will produce useful statistics and trend analysis.

If you were glancing through the my 5th slide with the hundreds of thousand or millions of data points, how would you recognise the crap ?



Big Data crazy outliers, Raw ANPR, to date fails the data scientists
This will change with time, probably sooner than later. I have 2 examples to consider, with the goal being to draw your awareness to what is hidden within big data's zeros and ones.
Number plate recognition and its application, in this instance at car parks.
A nice and controlled environment for plate recording and recognition, as we all stop at a barrier, take a card and have our plate recorded.
If the ANPR does not work properly it leads to bad data and very frustrated drivers.
Here we see the actual rego plate and the 3 examples (in red) of how they were interpreted wrongly by an ANPR based system at an international airport car park.
The evidence is the 3 tickets I collect on my way in during the past year.
The issue is that my attempt to exit is delayed due to having been incorrectly recorded on the way in. I don't exist.
I will leave to your imagine the gross inconvenience this causes me. At least they could get it wrong consistently two or three days apart.
Then ponder what if these data files are to be merged with external data sources or compared with or to track vehicles within their own multi site data sets.
The value of analytics is diminished when you have such errors.
Who has the where with all to review the syntax, audit the image versus it's recorded detail.
Logic checks are imperative on all data.



Now this slide is one of a substantial collection of mine
Represents the very ugly side of bad ANPR, the big data file entry that has no relationship to a plate.

This vehicle passing an expensive commonly used law enforcement camera

Note the top LH picture, it is the cameras ANPR's perspective of a rego plate, in fact it is the checker plate of the side rear portion of the tradies ute

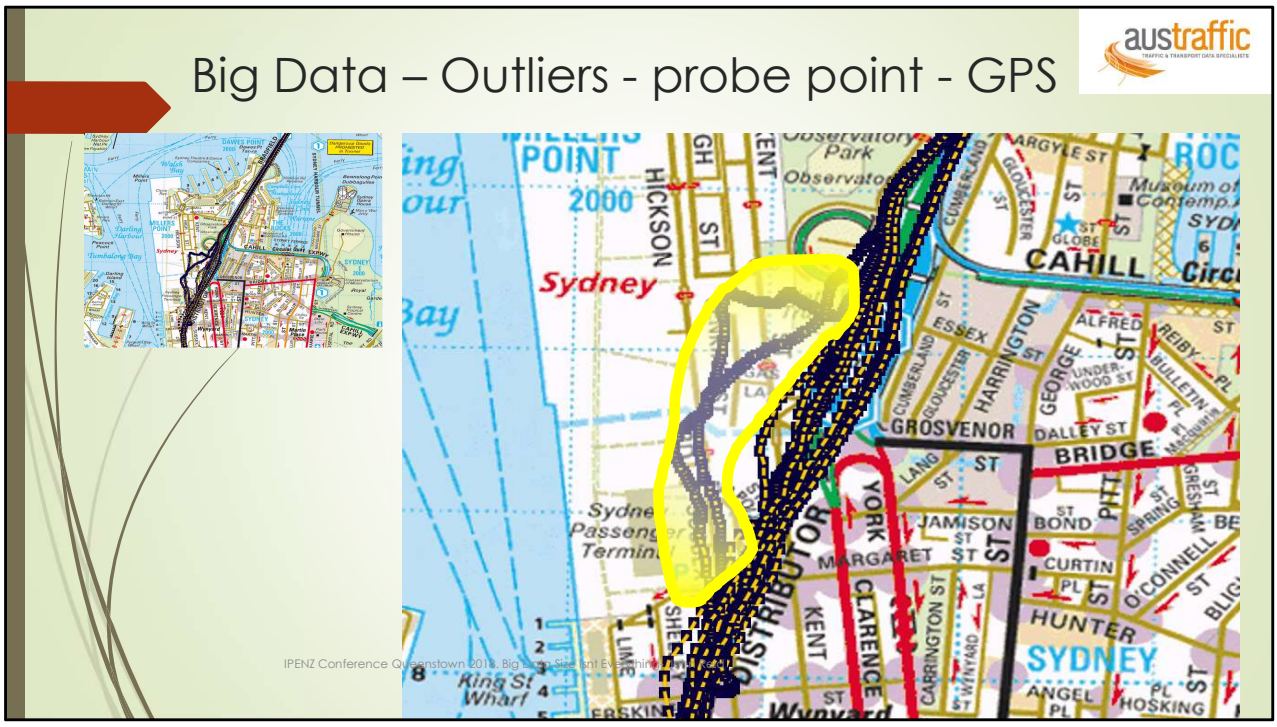
The ute created several recorded events, all dumped into an electronic file.

One record was correct, others were derived from the checker plate, signwriting, fences and grass on the opposite kerb.

How would you cull a taxi generated outlier, a vehicle with all its sign writing and roof signage and 4-6 event records in a big data file.

Better ANPR is the short form answer.
But when you receive a data file how do you know the calibre and credibility of the ANPR algorithm ?

Big Data – Outliers - probe point - GPS



Plotting data can help
This map is immediately south of the harbour bridge in Sydney.
There are 2 trips (highlighted with a yellow boundary) that deviated where we know the survey vehicle didn't travel.
They did not drive off the digitised road network nor through buildings.
Because we know what was done then this is easily fixed.
But what if the data set was generic probe tracked data, the probe could have been on foot or cycle or other ?
Two issues : What if the map track was correct, what was the mode of movement ?
How do you know this is an outlier ?
My fifth slide with all the numbers the zeros and ones are from this mapped GPS tracking job.
Where and How in the series of numbers, do you recognise and satisfy the position of good and bad data, or that which is bad for a part of a file only ?
My point is not to distract your thinking as to why this occurs, that's for another place and day.
My question to linger within your head today, How do you recognise and then filter it ?
Big Data, Big Question, how do you manage the bull shit ?

Informed Media are on to it

KGI director Aaron Clarke has joined the leadership program. PHOTO: WAYNE TAYLOR

Gamadji Institute (KGI) at Richmond Football Club, has joined a leadership program championed by the National Australia Bank to accelerate the promotion of Indigenous executives into the corporate world.

The Emerging Indigenous Executive Leaders Program, which kicked off in March, is backed by NAB, Crown Resorts, Richmond Football Club, KPMG, Rio Tinto, Commonwealth Bank, Broadspectrum, CareerTrackers (sponsored by Herbert Smith Freehills), Sodexo, Aboriginal Victoria, Reconciliation Australia and the federal Department of Human Services.

Eleven women and five men will participate in a 10-day course run over

and making critical decisions about how business operates," Mr Clarke said. He said Reconciliation Action Plans had been great stepping stones to expose corporate Australia to Aboriginal Australia, but beyond art, sport and land rights it was now time for Aboriginal Australia to also be better represented in the broader corporate world, and the economy where the doors have been shut for some time.

Each participant is also paired with an executive mentor. NAB has 230 Indigenous employees, up from 10 in 2009, however, most are in entry-level roles.

"Indigenous people share the same aspirations to climb the corporate

an absence of role models."

NAB worked with the Australian Graduate School of Management at the University of NSW to develop the program. Only businesses whose Reconciliation Action Plan has attained "Elevate" status are eligible to send participants on the course.

Mr Brennan is NAB's most senior Indigenous employee. NAB conceived of the program after a round-table discussion with Reconciliation Australia and other businesses that support Indigenous employment and wanted to progress Indigenous careers and leadership. "We didn't want to come up with another course that didn't resonate and wasn't needed," he said.

A masterclass in calling bullshit

Business life

Lucy Kellaway

Students at Washington University were offered a new course last month, entitled Calling Bullshit in the Age of Big Data. For the past couple of decades, week in and week out, I have been calling bullshit for this publication, and so was delighted to hear my favourite pastime had made it into academia.

While this course is limited to spotting bullshit in numbers, there is an equal need for one spotting it in words, especially words used in business. What follows is an outline for a rival course aimed to fill that gap.

It starts with a definition: bullshit means nonsense, usually of a puffed-up variety that pretends to be something it is not. Sharp eyes will spot at once the difficulty in applying this to corporate life – almost everything fits the description. Before I have even got inside my office I have tripped over a yellow plastic sign saying "Caution Wet Floor" – bullshit because usually the floor is not wet and if it were the picture of someone falling spectacularly is wildly overdoing it.

The first rule about calling corporate bullshit is not to do it too assiduously or you will go insane. I have learnt to ignore 95 per cent of it, and of the remainder ask myself two questions: what is the quality? And: how damaging is it? I have gone through dozens of examples of bullshit that have come my way in the past couple of days and picked three that are worth calling.

The first is a branding document produced for a new Pepsi logo in 2008 and resuscitated last week on Twitter. With diagrams comparing the Earth's magnetic fields to "Pepsi energy fields" and text that reads, "The Pepsi DNA finds its origin in the dynamic of perimeter oscillations" – it is grade A, unadulterated BS. But on the second question – whether it was damaging – the answer is no.

Pepsi changed its logo and carried on selling its brown sugar-water around the world willy-nilly. Even so, bullshit like this deserves to be called both for its exceptional quality, and because doing so might encourage its perpetrators to have a dark night of the soul in which they wonder what on earth they are up to.

Exhibit two is a document from Deliveroo on its preferred language for describing the poor sods who cycle round with other people's smelly takeaways on their backs. The memo bans "employees", replacing it with "independent suppliers", and forbids "pay" and "hiring" preferring "invoices" and "onboarding" instead.

On the quality measure this bullshit is tame. "Independent supplier" and

Deliveroo denies "employees" basics like holidays and sick pay. PHOTO: JASON SOUTH

"invoice" are innocuous, and "onboarding", though regrettable as a gerund, especially with no boat in sight, is so common there is little point in protesting. But on the measure of harm, Deliveroo's memo is wicked. It knows that if people used the ordinary words "employee" and "hire", they might make the mistake of thinking they were due ordinary things like holidays and sick pay – which Deliveroo doggedly denies them.

The third example comes from Jim Norton, who has the delightfully bullshitty title of chief business officer, president of revenue at Conde Nast. Last week he outlined his new strategy to all staff in a memo that began "Team" and proceeded with a stream of corporate nonsense about playbooks and journeys and wide arrays of differentiated solutions. It glossed over sackings as "hard personnel decisions", only to declare the new corporate plan, "Conde Nast One".

For companies to claim themselves "one" is standard bullshit – it is a cliche and a lie, given the inevitable number of vested interests in any organisation. If Norton were in the motor trade or banking, I might let this pass. Yet Conde Nast publishes Vanity Fair and The New Yorker, where standards of editing are so exacting that one of the latter's editors has written a whole book based on the correct placement of a comma.

Norton may well bang on about the "heritage of quality journalism", but had he asked his staff to edit his battle cry: "We will all transition this business together", they would surely have told him transition was ugly as a verb, but as a transitive one was a monster.

FINANCIAL TIMES

The informed media is starting to see this issue.

The Australian Financial Review ran a story on the 10 April 2017 with the heading "A Masterclass in calling bullshit".

It made reference to a course at Washington University similarly titled "Calling Bullshit"

It would be easy to conclude that this course was only about the current political environment of fake news and alternate facts. Indeed, the course does address these issues.
But it touches on matters far closer to the heart of our profession.

Although science is not universally popular, it is critically important.

We owe it to our profession to make sure we sniff out crappy science.

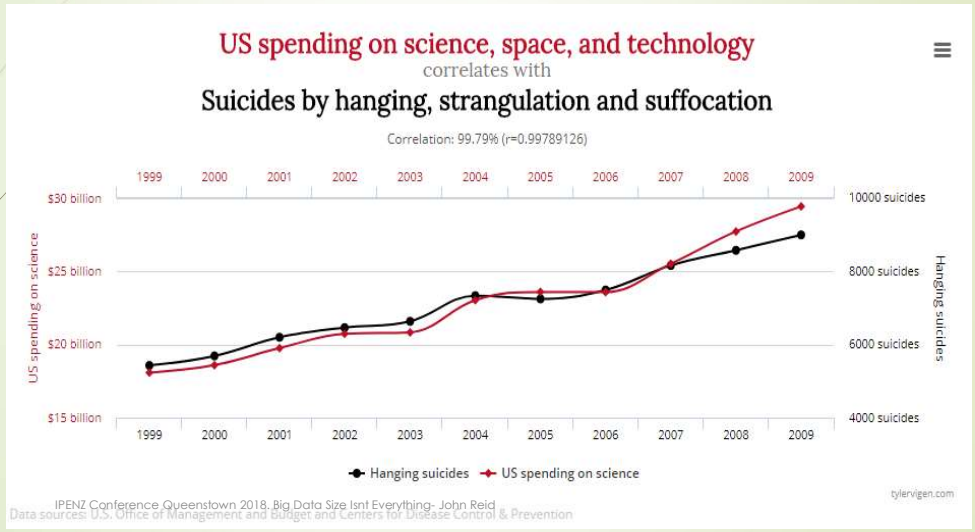
The world won't believe us, just because we follow rigorous technical procedures.

Regarding the course, Carl Bergstrom, a professor in the university's biology department, and his colleague, got the idea for the course when they noticed a trend in the last few years of More bullshit in the articles they were reviewing.

One area of big problems they identified was in Big data. He said he noticed methods of statistics meant for smaller data sets were being applied to "big" data sets with millions or billions of examples, where it's easy to force a correlation that isn't necessarily accurate.

He also observed situations where machine learning algorithms were "overfitting" data.

Reality or Pretty Patterns

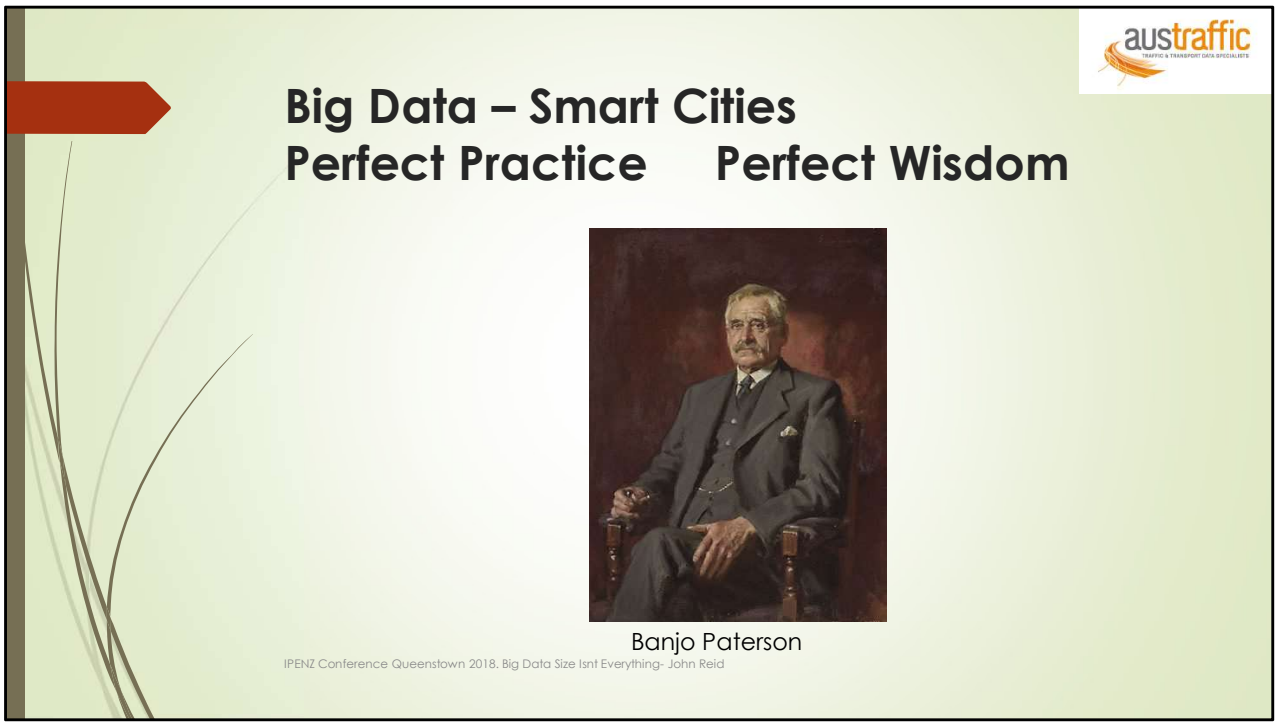


We all know that a colourful, technical graph can look convincing.

But is it true or not, or just a shallow reflection of a deeper issue?

Data and analytics impacts outcomes,

Then we have the issue of Spurious Correlations as shown here and the misuse of data which can be problematic,

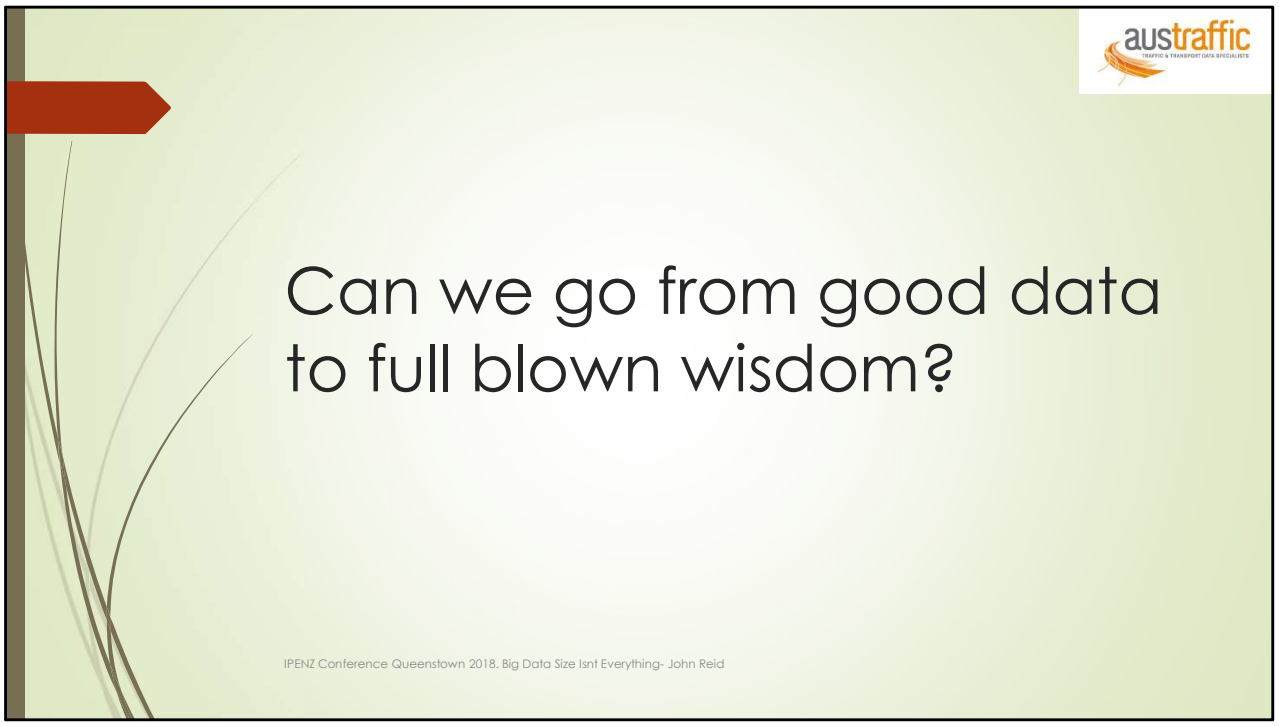


We need time to perfect and practice wisdom.

If we have no time to think and ponder what we are doing and how we are doing it then we will be overpowered by circumstances including new technologies.

At the AITPM Victorian branch end of year function 2016, I quoted a few verses from the classic Australian poem "Clancy of the Overflow" written by Banjo Paterson, first published in *The Bulletin*, an Australian news magazine, on 21 December 1889. . The verse that stands out the most for me is as follows: *And the hurrying people daunt me, and their pallid faces haunt me*
As they shoulder one another in their rush and nervous haste,
With their eager eyes and greedy, and their stunted forms and weedy,
For townsfolk have no time to grow, they have no time to waste.

But this is not, as I mentioned at the beginning of this paper, just a sentimental reference to make us feel momentarily reflective. It is a core problem today. Joe Wolfe, an Australian physicist turned comic poet, wrote a parody/homage to the classic Paterson poem. As the narrator sits at his desk trying to answer another correlated big data problem, that of all his emails, he laments: *But the looming deadlines haunt me, And their harrying senders taunt me , That they need response this evening, For tomorrow is too late!, But their texts, too quickly ended, Often can't be comprehended, For their writers have no time to think – They have no time to wait.*



Even if we get much better at collecting data; making sure it is accurate and comprehensive;
even if we get better at using this information to enhance our knowledge;
we have to empower decision makers and other stakeholders and the younger members of our professions with wisdom to be able to understand and adapt to this ever changing world.

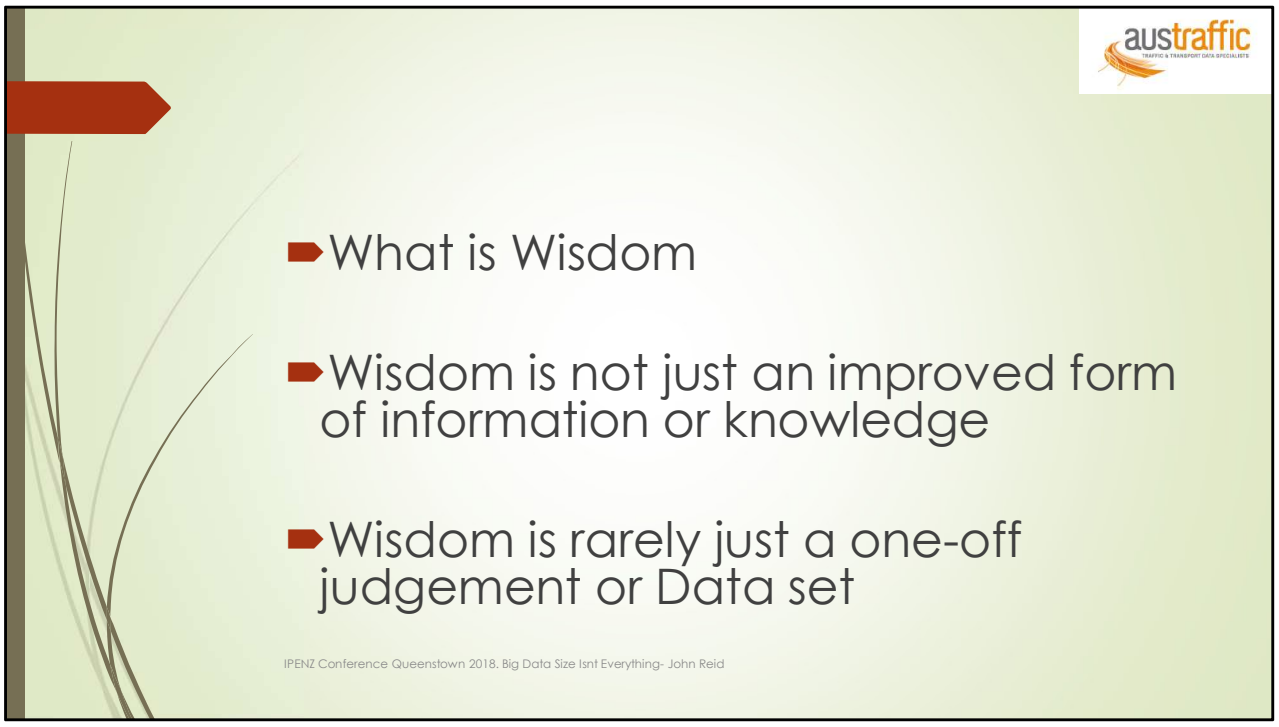
As a profession, we have discussed the need for mentoring.

But this has to be much more than just telling young people how we did it in “my day”.

We are also faced with Aussie government departments that are deskilling their areas and replacing them with generalist managers.

Good management is important but if it is based on poorly developed conclusions, no matter how much data is behind it, then we have lost the plot.

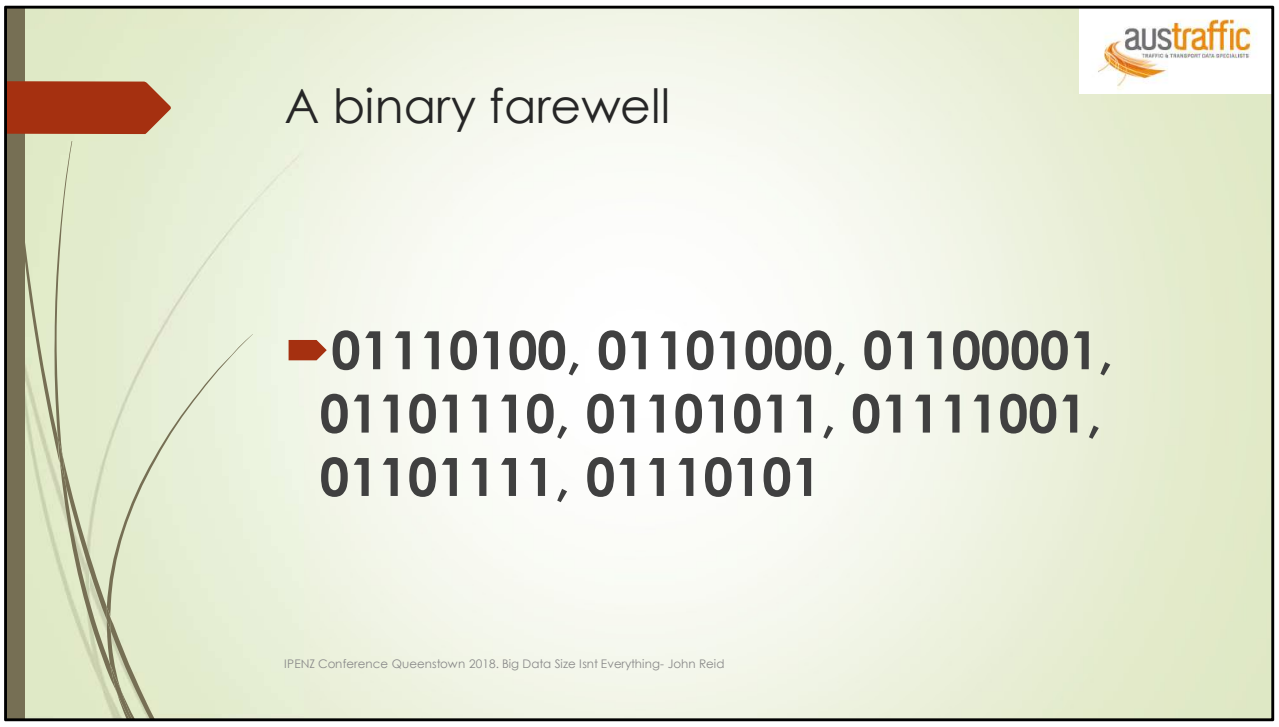
Consultation is much more than informing, rather about engaging and learning



In the paper I reflect on some of the examples of wisdom that are part of our culture.

- Wisdom is more than just a few truisms

- Technology for its own sake is not wisdom
- Wisdom is not just an improved form of information or knowledge
- Wisdom is rarely just a one-off judgement or data set.
- Wisdom includes the ability to know when not to say something or when not to use some data.
- Wisdom can provide guidance and may not give the ultimate answer
- Wisdom is getting to the best answer by the best processes in a way that people who are affected, own the solution.



As noted in the one's and zeros

In this instance are the letters of the word thankyou lower or upper case ?

Upper case would have a different set of ones and zeros, a different outcome

Thank you