-A binary welcome

01100111,
00100111,01100100,
01100001, 01111001

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 ?

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   1

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

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   2

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

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?

3

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

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?"

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   4

Big Data
We are being bombarded, swamped with data.
What do the numbers mean ? Are they credible numbers ?
Any 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.
34 years ago, I specifically moved into the area of actually collecting the data.
In the early days intersection counts were often collected perhaps in the morning and afternoon peak periods on one day of the week, by people standing on corners with clicker boards.
Simplistic – yes. Yet we should not belittle this.
The important point was that we pursued an understanding of the specific situation and hopefully some aspects of the broader picture.
We can now collect a lot more data at one spot but even more, we are getting data from a much wider number of locations.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   5

Fed into a computer, these numbers will lead to a conclusion.   But is it reality?
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?

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   5

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,

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   6

Big Data is a Big Task

- The first step is clean data

- Where are the data outliers

Cleaning up data

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

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?

7
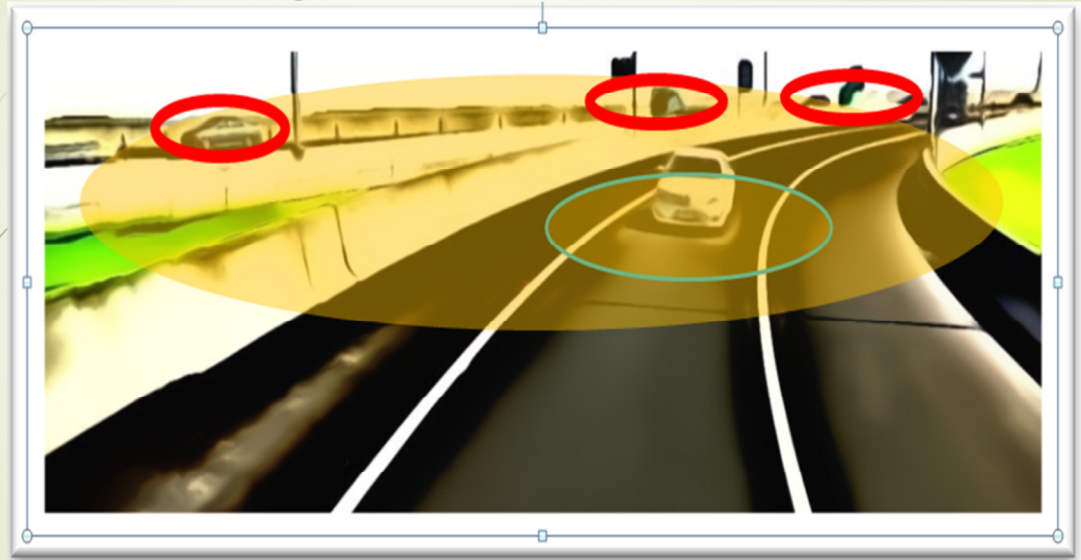
At the AITPM National Conference in 2013 Scott Benjamin, presented research on the initial use of recording Bluetooth devices as a way of getting the origin and destination of trips. Where motorway roads and ramps were near each other, there were specific interference issues.
The picture is of a freeway interchange and four vehicles shown, 3 highlighted red and our target circled green, yet the BT loggers recorded all 4 events. Producing Origin destination trip assignments distorted by factors of 2 & 3.

Bluetooth counting is giving us much more data.

But we have to remember it is still a sample and it is measuring people trips not vehicle trips (which is actually a very interesting and a valuable measure).

We have to consider what a bus load of people means, especially if the bus is stopping regularly.

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

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   8

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.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   8

BT Read Rate & Repeatability
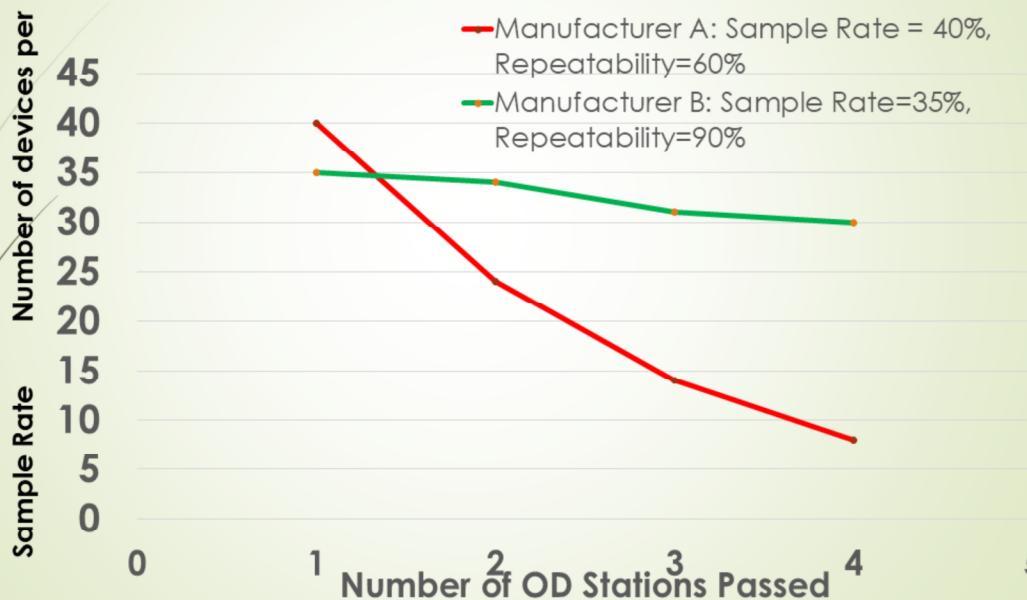
Austraffic began reviewing a range of BT Solutions 5 years ago. We reviewed a range of technologies, each for between 3 to 6 months on a range of standard test sites, which included local roads, urban arterial roads, rural arterial road, and metropolitan freeways.

Validation of BT equipment has often been undertaken by driving a vehicle with multiple active BT devices passing a number of BT monitoring devices.
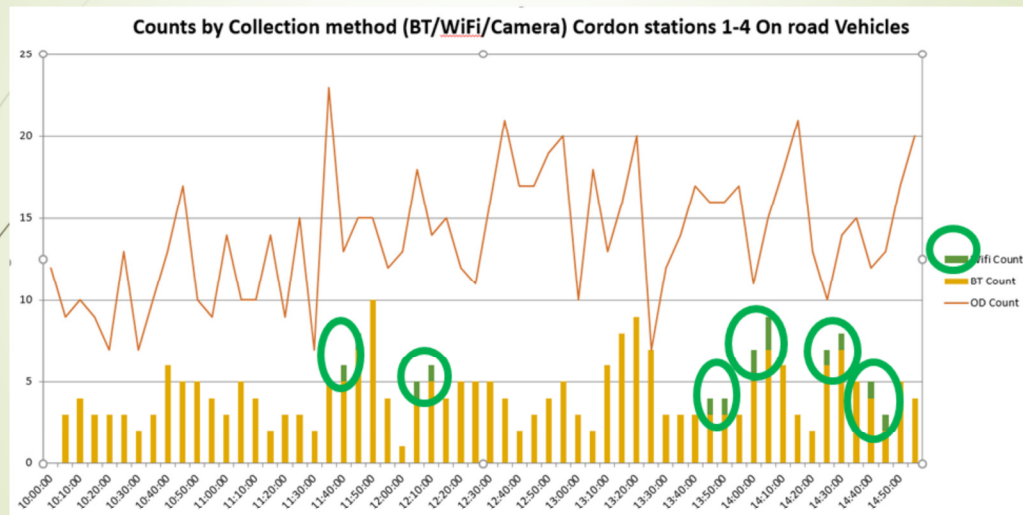
Whilst this may provide some indication of the reliability or likelihood of a certain type of device being monitored, it does not assist in understanding how many devices are present in the vehicle stream as a whole.

We sought to bench mark each system against a common set of measurements by validating BT device read rates 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

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   9

Sample Rates & Matched Events

Counts by Collection method (BT/WiFi/Camera) Cordon stations 1-4 On road Vehicles

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.

Time increments shown are 10 minute and vertical scaling is units of 10 minutes.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   10

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.

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.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   11
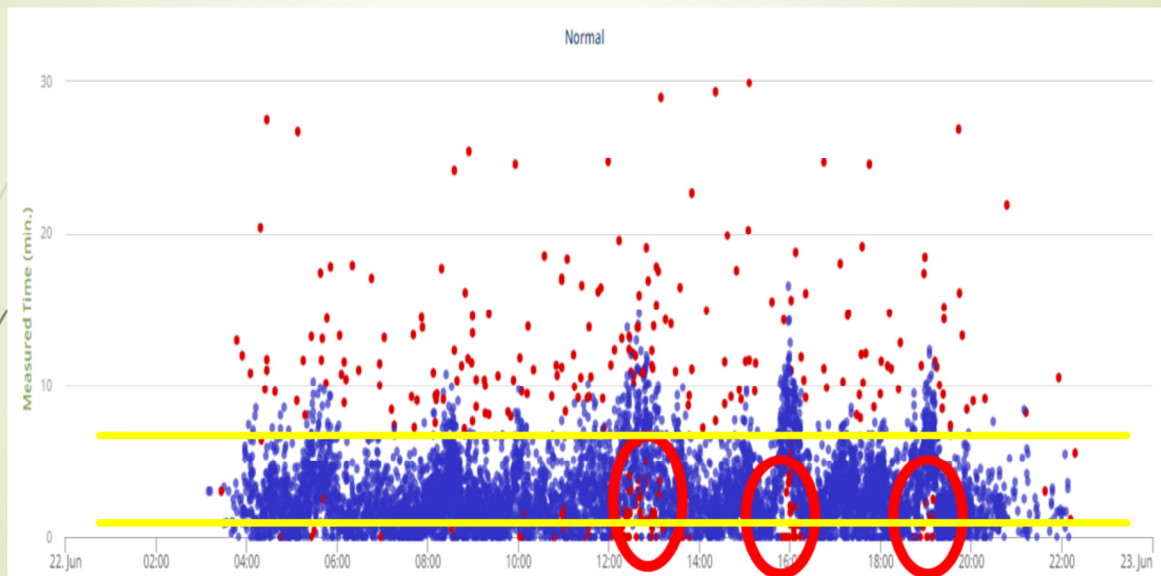
Next is an example of outliers (data records that are beyond the expected range) with probe sensed data, it could be the same graphic for any probe sourced data set be it Cellular, BT, WiFi,or navigational

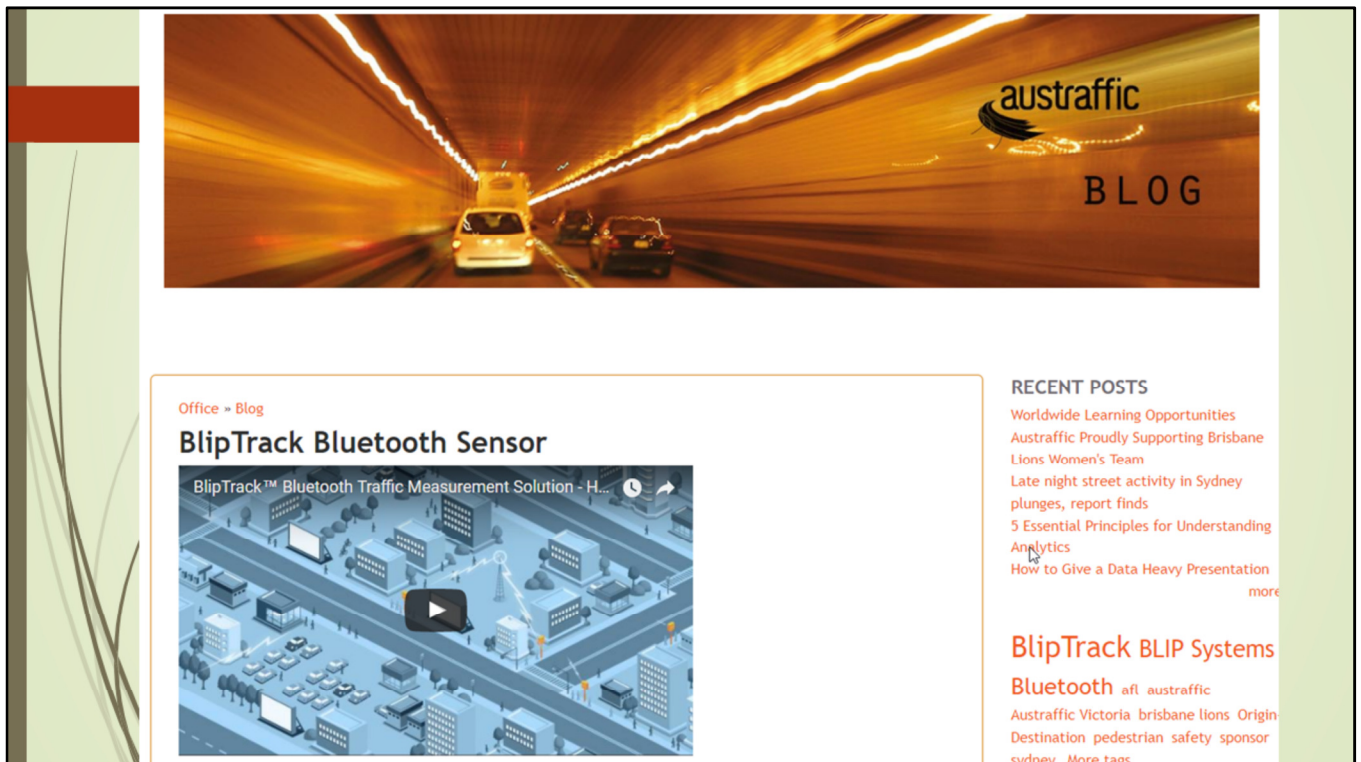This data is from static BT sensors, pairs of sensors that form to provide bidirectional zone based data along the corridor of interest.

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 fifth of my slides with hundreds of thousand or millions of data points, how would you recognise the crap ?

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   12

The austraffic website contains a video of BlipTrack for those seek more information.
Blip is one of a number of probe sensing devices available in the market. Our field experience suggests it is a market leader.
Hence our take up.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   13

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.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   14

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.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   14
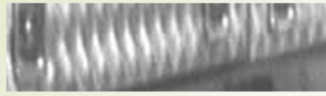
Now this slide is one of a substantial collection of mine

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.

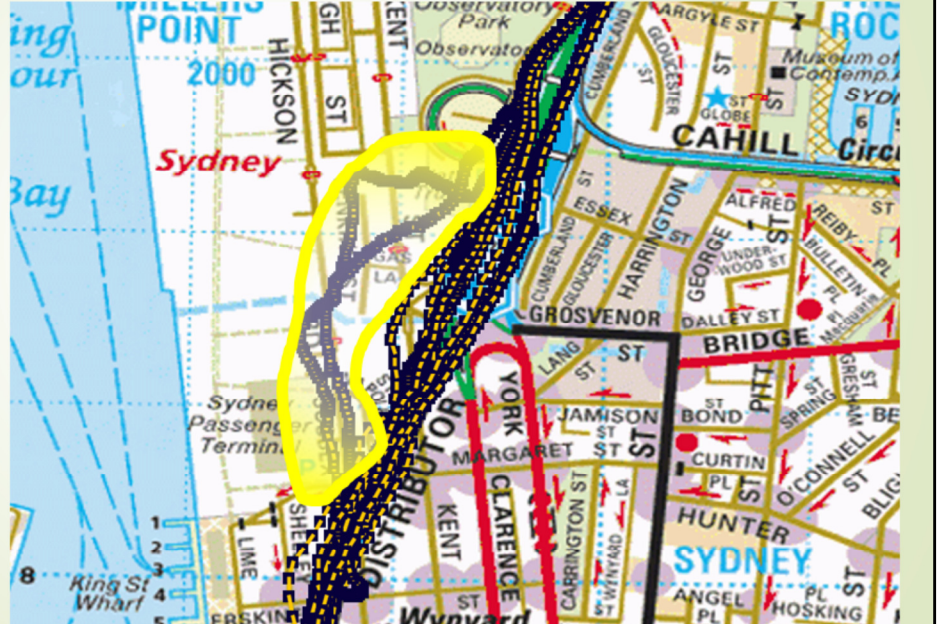Better ANPR is the short form answer.

But when you receive a data file how do you know the calibre and credibility of the

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?

15

ANPR algorithm ?

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   15

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 ?

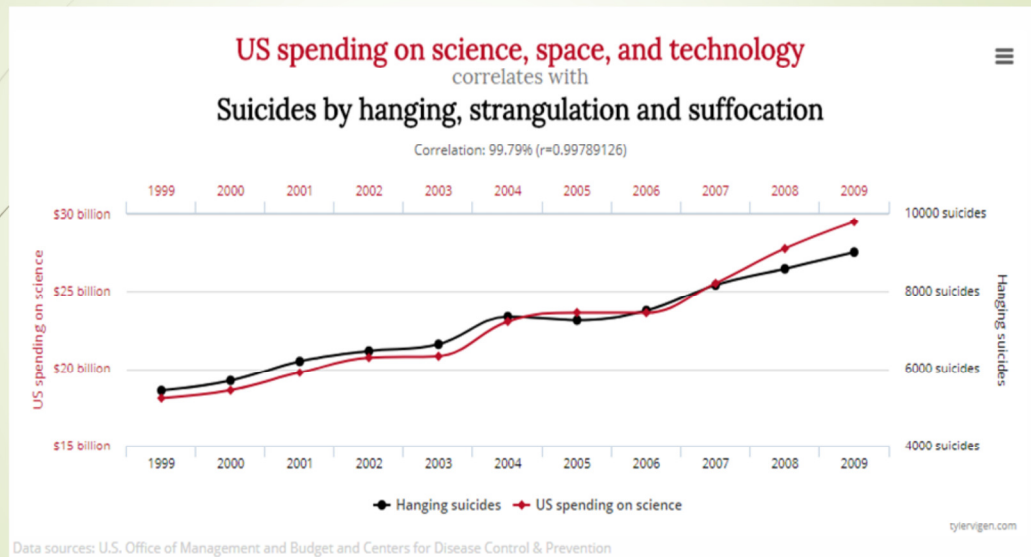My point is not to distract your thinking as to why this occurs, that's for another place and day.

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 ?

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   16

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

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

Data and analytics impacts outcomes, Just as missing the contribution of swimming as a transport mode in Fiji as referenced in the paper.

Then we have the issue of Spurious Correlations as shown here and the misuse of data which can be problematic, the contrivance commences with the lack of perspective as neither of the y axis commence with zero or a discernible base. Let alone question the suitability of the comparables.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   17

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.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?

18

TEACHING

The Fine Art of Sniffing Out Crappy Science

By Steve Kolowich | JANUARY 16, 2017

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

We owe it to our profession to make sure <u>we</u> sniff out crappy science.

The world won't believe us, just because we follow rigorous technical procedures.
How can we support political pressures and community best interests, if we produce poor science.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   19

Calling Bullshit in the Age of Big Data

Logistics

Course: INFO 198 / BIOL 106B. University of Washington
To be offered: Spring Quarter 2017
Credit: 1 credit, C/NC
Enrollment: 160 students
Instructors: Carl T. Bergstrom and Jevin West
Synopsis: Our world is saturated with bullshit. Learn to detect and defuse it.

The course will be offered as a 1-credit seminar this spring through the Information School at the University of Washington. We aim to expand it to a 3

Regarding the course, Carl Bergstrom, a professor in the university's biology department, and his colleague professor Jevin West, got the idea for the course when they noticed a trend in the last few years: 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.

Basically, you can have an algorithm that so specifically matches a particular data set, meaning it reflects even errors or noise, it fails when applied to another data set where you would otherwise expect it to work. You would normally want an algorithm that is sufficiently general to fit more than one data set.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   20

Big Data – Smart Cities
Perfect Practice   Perfect Wisdom

Banjo Paterson

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.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?

21

Can we go from good data to full blown wisdom?

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.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   22

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 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

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   23

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

The wisdom of Solomon who suggested a solution, not as a real alternative but to see two people's responses. Do you cut a child in half to prove your point.

The Maori principle of not just judging an offense but bring the offender and the victim together to bring about understanding.

- 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

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   24

people who are affected, own the solution.

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   24

A binary farewell

01110100, 01101000, 01100001, 01101110, 01101011, 01111001, 01101111, 01110101

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

AITPM Melb 2017 Conference, John Reid
Where are we on the scale of data,
information, knowledge and wisdom?   25