Hard Data in a world of Truthiness…

Richardh.young@beca.com

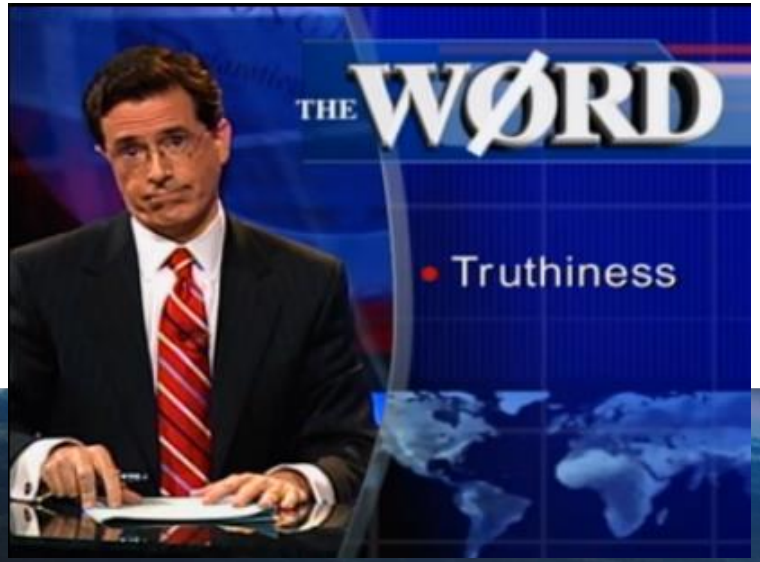
IGNITE
your thinking

Truthiness…..

…the belief or assertion that a particular statement is true based on the intuition or perceptions of some individual or individuals, without regard to evidence, logic, intellectual examination, or facts.

Dr Stephen Colbert, October 17, 2005

- *Full name, Maestro Professor Field Marshal The Rev. Sir Dr. Stephen Tiberius Mos Def Colbert D.F.A., Esquire Heavyweight Champion of the World.*
- *Source: The Colbert Report, a CBS satirical show.*



Truthiness…. and us

Ethical Obligations (Engineering New Zealand, 2016)

- Possess sound engineering knowledge, apply with skill, diligence and care.
- Keep engineering knowledge up to date.
- Understand our limits of competence.
- Accept personal responsibility for the work we do.
- Ensure that we do not misrepresent our areas or levels of experience or competence.

Hard Data in a world of Truthiness

Hard Data

- Verified,
- Specific numbers,
- Auditable,
- Often Independent.

27% of traffic opted for Route A.

The worst affected journeys were delayed by 17 minutes.

Truthiness

- Rough guesses,
- Numbers said with conviction,
- Repeated numbers,
- Anecdotal impressions.

The queue is always about 3km.

Felt like an hour's queuing.

Our traffic data is always correct.

So why is Hard Data important?

- Data driven decisions
- Objective assessments
- Needed for evidential processes
- Performance-based Contracts
- Pain/ Gain calculations
- Responding to statements based on perceptions, truthiness.

The ideal Traffic Data set..

- The duration, route, occupancy, mode, purpose and timing of every vehicle on the road, plus a multitude of other factors.
- Cover the whole of New Zealand and be continually up to date and 100% accurate.
- Whilst being a perfect data set, it is guaranteed to fail to gain a social licence to exist and additionally fall foul of the Privacy Act.
- There are few, if any democratic countries that would aspire to gather this level of insight into their population.

Concept 1 - Don’t measure every vehicle…

- High sample rates may appear to be ideal, but the time and cost to gather it is seldom justified.
- Remember Margin of Error and Confidence Level ? (Stats 101, Gaussian Distribution)

Population 10,000	Margin of Error				
Confidence Level Size	1%	2%	3%	4%	5%
95%	4,900	1,937	965	567	370
99%	6,240	2,932	1,557 (16%)	940	623

Concept 2 – Avoid Sample Bias

- Repeat the sampling at a range of locations and times of day.
- For example. if you get similar sample rates on
  - Freight dominated routes
  - Hire car routes
  - Busy commuter routes
- This is a reasonable indication that the sampling technique is capturing a Representative Sample of all traffic.

Big Data – bigger challenges?

- Collecting data is often expensive and companies that can exploit the commercial value of this data often will do so.
- Check for specific statements and warrants on data quality.
- Commercial organisations will often place stringent limitations around how data can be used and how long it can be stored.
- Read the small print, especially for large providers.

Forecasts vs Hard Data (News) – I blame the weather presenters

How bad is the traffic now?
How does this compare to last week?

Now

What is the weather likely to do?
How bad is traffic likely to be?

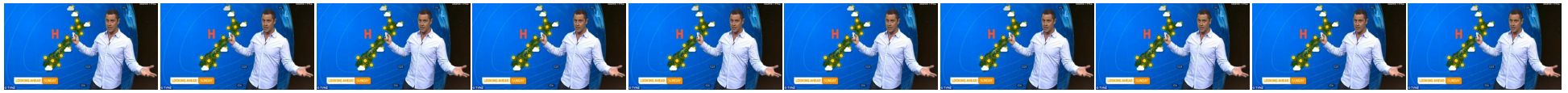
Then → **Future Prediction**

- The News

- The Forecast

Methods of capturing Hard Numbers

Storing a series of Forecasts into the Future



07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00

Storing a series of Records of what happened

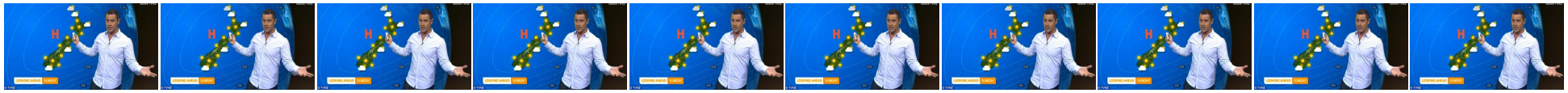
07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00

Forecasts are NOT Hard Numbers

Google Maps Platform

Web Services › Directions API

Storing a series of Forecasts into the Future



07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00

- `traffic_model` (defaults to `best_guess`) — Specifies the assumptions to use when calculating time in traffic. This setting affects the value returned in the `duration_in_traffic` field in the response, which contains the predicted time in traffic based on historical averages. The `traffic_model` parameter may only be specified for driving directions where the request includes a `departure_time`, and only if the request includes an API key or a Google Maps APIs Premium Plan client ID. The available values for this parameter are:
  - `best_guess` (default) indicates that the returned `duration_in_traffic` should be the best estimate of travel time given what is known about both historical traffic conditions and live traffic. Live traffic becomes more important the closer the `departure_time` is to now.

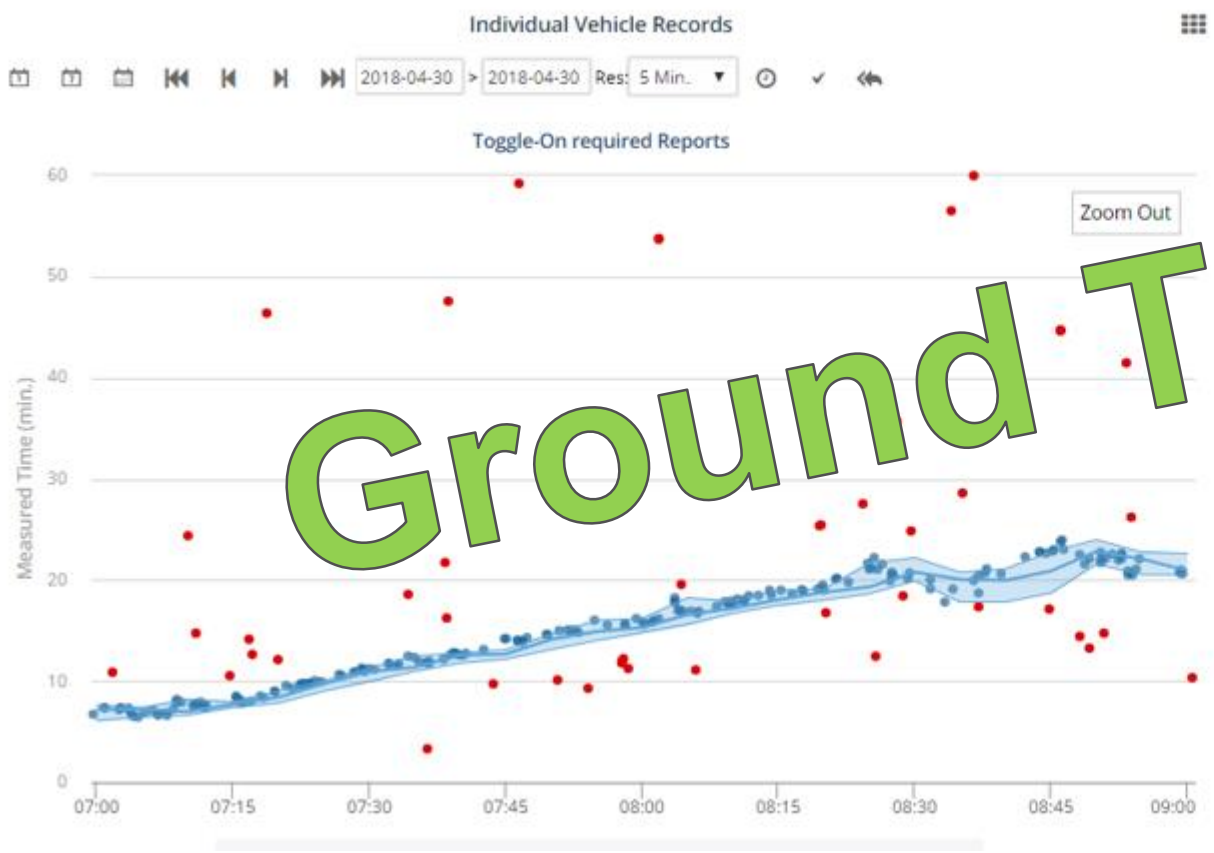
The search for geniune Hard Numbers

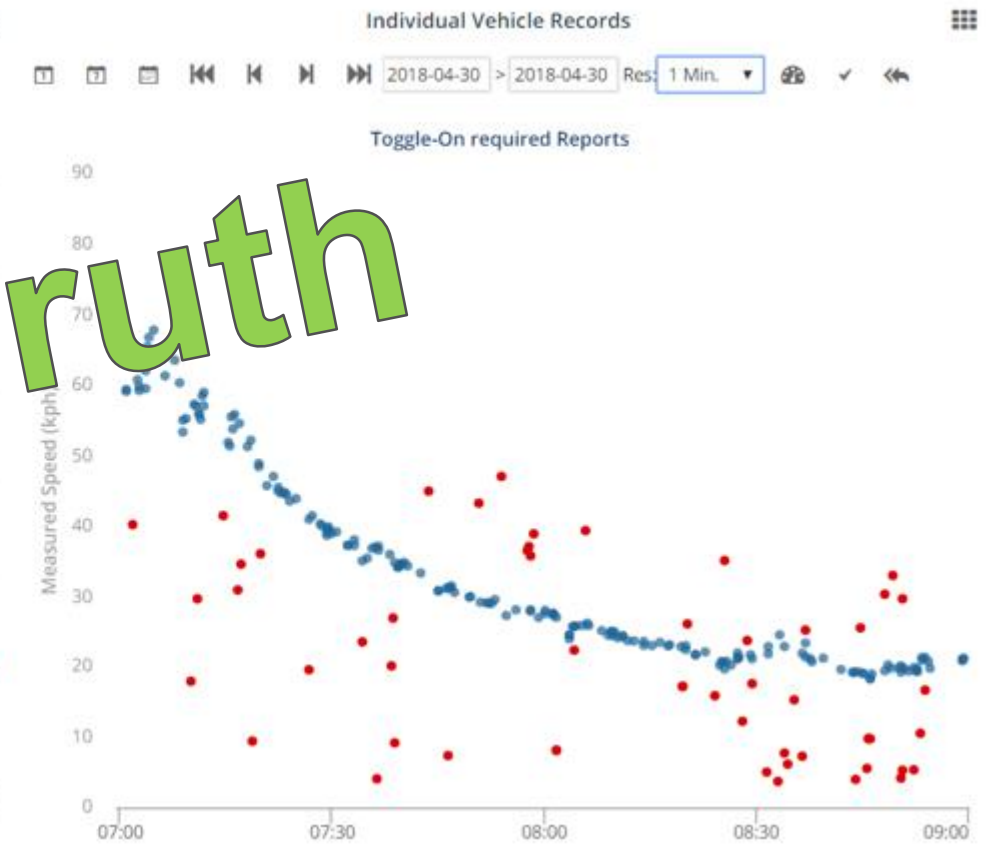
Storing a series of Forecasts into the Future

07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00

Forecasts play a useful function, but would you like your client's $100M investment measured against a series of future looking predictions or against a record of what actually happened?

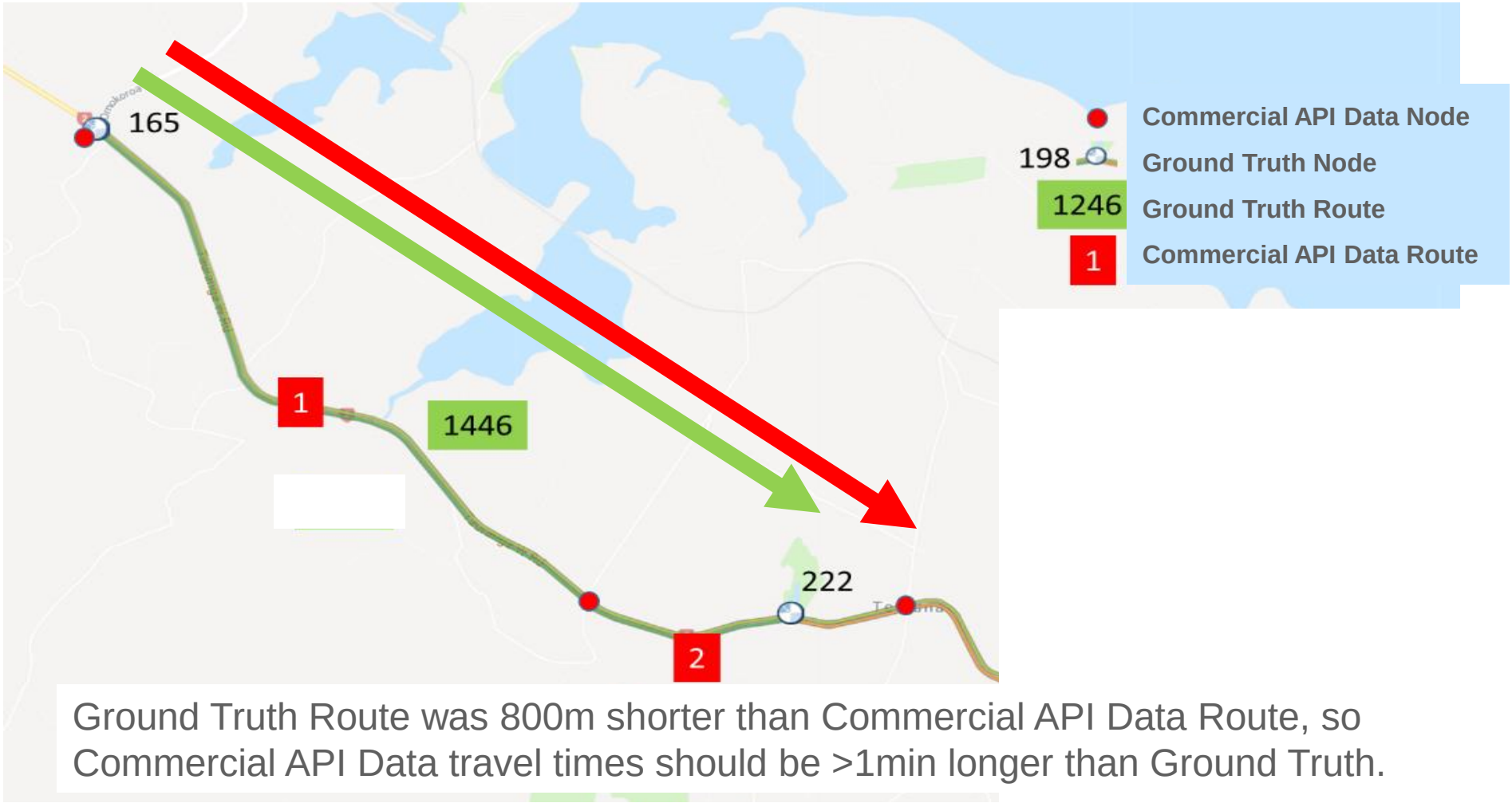
Really boring. Point-to-point traffic tracking…





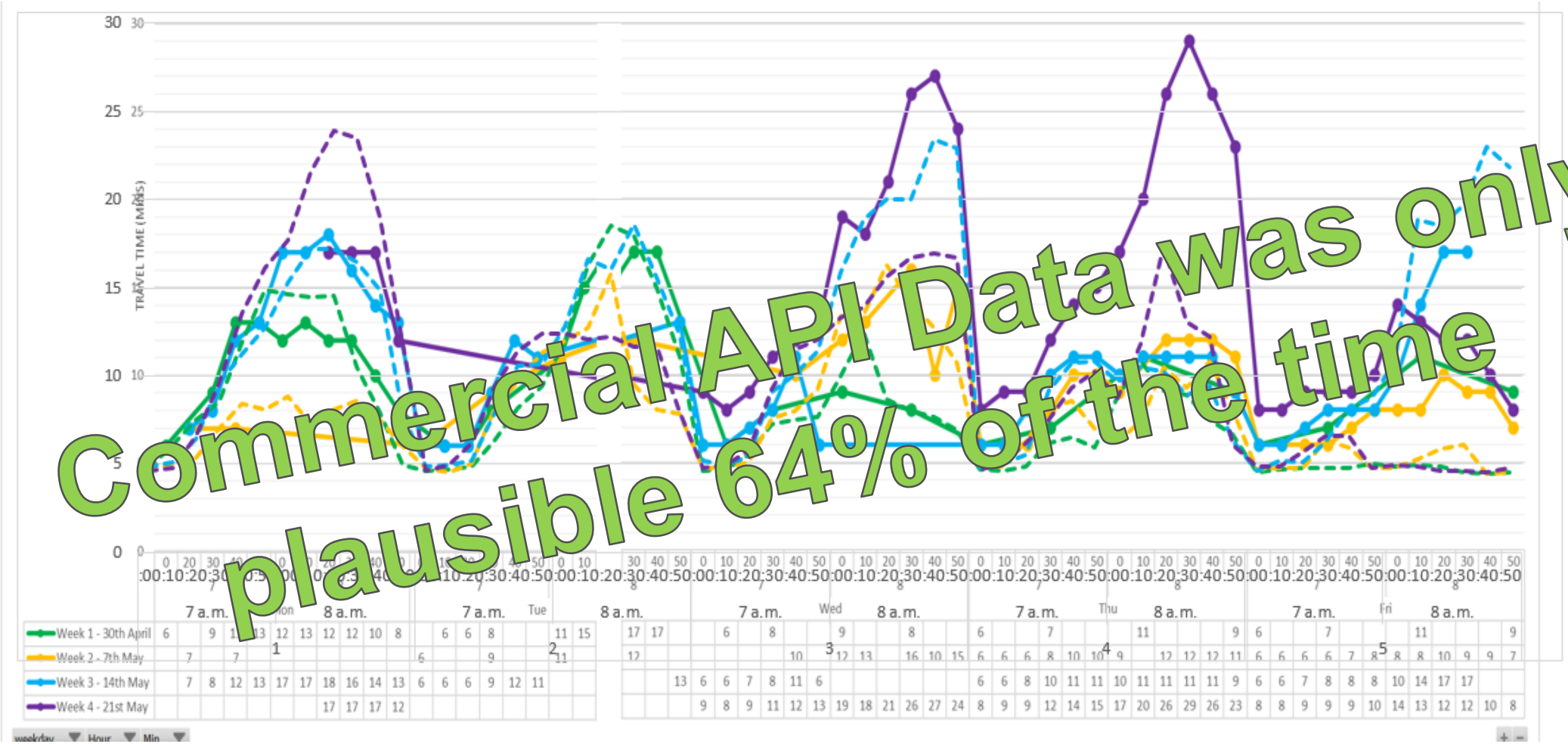
Ground Truth

Ground Truth vs Commercial API Data



Ground Truth Route was 800m shorter than Commercial API Data Route, so Commercial API Data travel times should be >1min longer than Ground Truth.

Ground Truth vs Commercial API data



Truthiness…. and us

Ethical Obligations (Engineering New Zealand, 2016)

- Possess sound engineering knowledge applied with skill, diligence and care.
- Keep engineering knowledge up to date.
- Understand our limits of competence.
- Accept personal responsibility for the work we do.
- Ensure that we do not misrepresent our areas or levels of experience or competence.

Some Best Practice guidelines for Road Traffic Data..

- 1 – Privacy
- 2 – Data Sourcing
- 3 – Fitness for Purpose
- 4 – Client provided Data

1 - Privacy

- Some Traffic Data could contain ‘personal information’.
- The Privacy Act recommends that a Privacy Statement to be produced.
- Transportation Professional should provide this to the Client.

2 - Where Traffic Data needs to be sought for a Client

- Identify several appropriate data sources.
- Provide the Client with options setting out:-
  - reliability,
  - longevity,
  - cost,
  - potential for re-purposing,
  - technical or commercial limitations to using that data.
- In-house data is also covered by same rules.

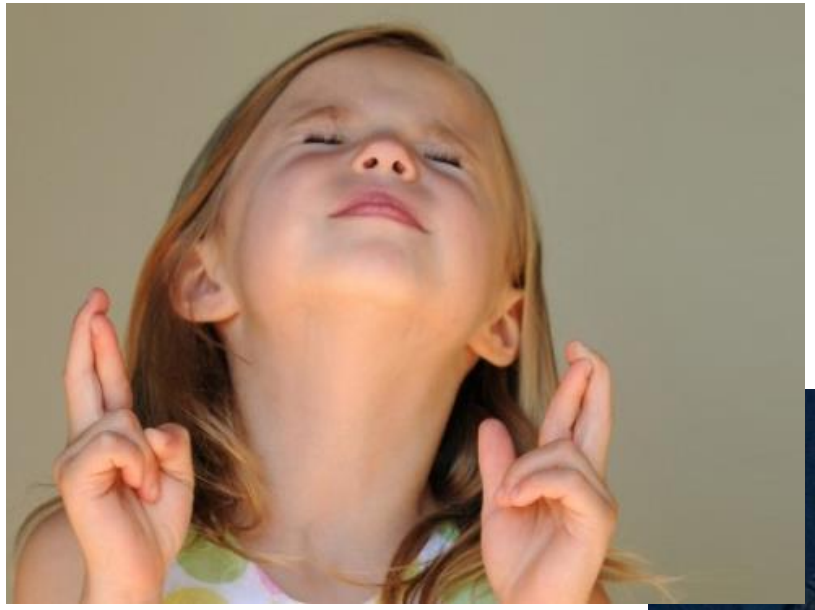
3 - Fitness for Purpose, Ethical / Professional requirements

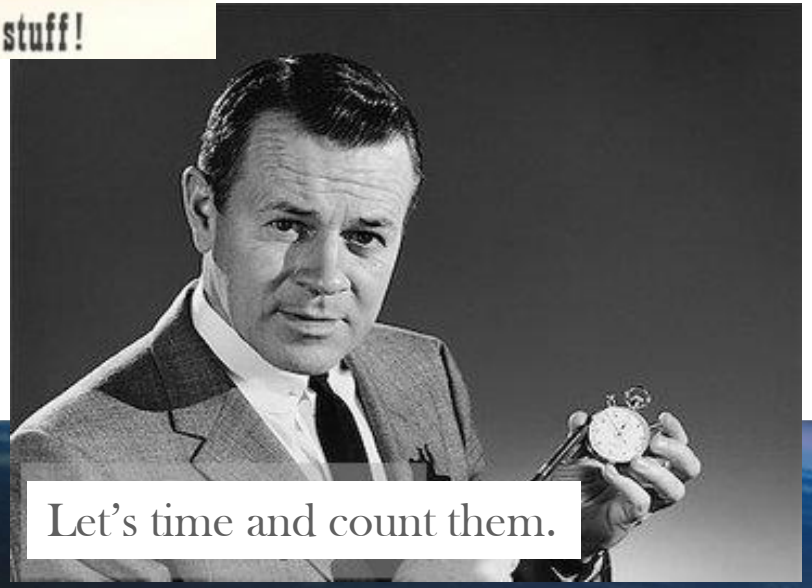
- Transportation Engineer shall have a sound understanding of
  - Collection methodology,
  - Processing of raw data to final data source,
  - strengths and limitations of any data source used.
- Warrant that the data is suitable and appropriate for the Client's needs.

4 - Client Provided / Instructed data sources

- Transportation Engineer is still obligated to provide an impartial professional service.
- Undertake checks on Client provided / instructed data in the same way as they would with data sourced from third parties on in-house.

How do you source and validate your Traffic Data?





IGNITE
your thinking

Hard Data in a world of Truthiness…

Richardh.young@beca.com

IGNITE
your thinking

Tuesday
‘The Autonomous Electric Vehicle Myth…’

Tomorrow 2:30 Oceania.

Richardh.young@beca.com

IGNITE
your thinking