Using data science principles for developing business students critical thinking skills

Capital thinking. Globally minded.

TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI VICTORIA UNIVERSITY OF WELLINGTON

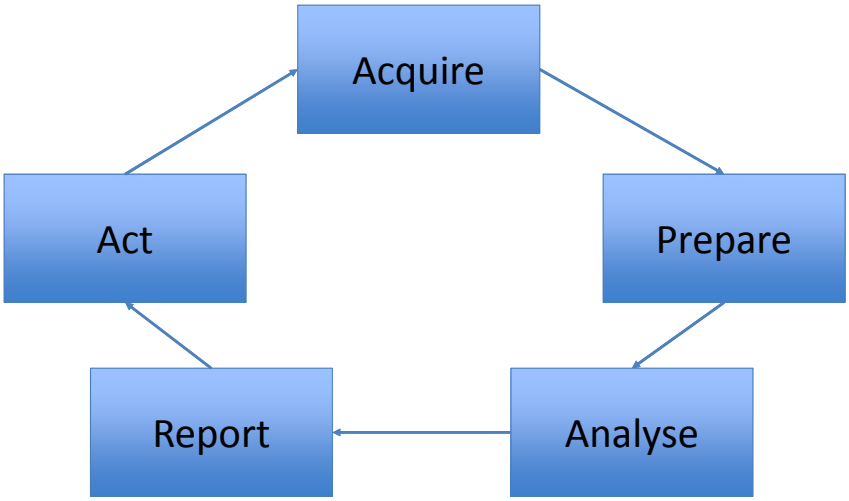
Data, it's all Greek to me

- Business today is over supplied with data creating the need for a new skill; the tāuru kaitātari (the data translator) that person who can combine seemingly unlikely sources of data and generate new insights and knowledge from those data

Objective analysis of facts

- Sometimes the tāuru kaitātari is a data scientist but can also be a well-trained manager or similar data specialist.
- The goal is to identify, access and authenticate data that can be combined to address critical questions

The data science lifecycle



Acquire

- Raw data
- Internal sources – systems, logs, reports etc.
- External sources – supply chain, government, industry, 3rd Party
- Open data – IDI, Open gov't and many others

Students can access many open data sources online e.g. data.govt.nz

- *5,400 data sets*
- *145 organisations*
- *25 groups*

Formulating questions

- Is the first skill tāuru kaitātari need to be able to master
- The process is far from tidy and linear when you get into it – data translation has the familiar challenge that people don't know what they don't know but we have to begin somewhere.

Examples

- Can sales data and contact centre call data provide product a meaningful product evaluation?
- Will sensor data from a machine plus weather observation data provide a meaningful picture of system reliability?
- Can student lateness and attendance combined with traffic congestion data inform our timetabling decisions?

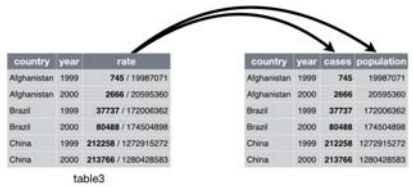
Asking the right question

- Helping students develop the skills to formulate questions that will produce results that will, in-turn, lead to meaningful insights
- To ask questions about business, students need an opportunity to live business

how can we do that? (TPS?)

Prepare

- Data often does not come in tidy packages suitable for the kind of analysis we want to do
- For example, operational data from sales or accounting systems has redundancy reduced by design
- The tāuru kaitātari is also a detective, locating data in unlikely places and imagining combinations that might not have occurred to anyone before



country	year	rate
Afghanistan	1999	745 / 19987071
Afghanistan	2000	2666 / 20595360
Brazil	1999	37737 / 172006362
Brazil	2000	80488 / 174504898
China	1999	212258 / 1272915272
China	2000	213766 / 1280428583

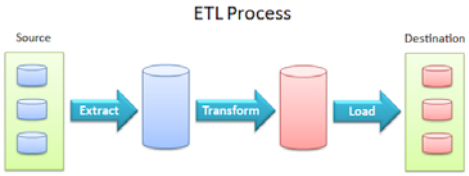
table3

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Extract, transform, load

- Often, we need to extract data from its source systems, translate it into a new form by cleaning, sub-setting, or filtering it to get data into a form that lends itself to analysis.
- The extract-transform-load (ETL) process is a mature branch of programming that have specialist skills and tools of their own

While we use and teach process this with quite sophisticated tools, a lot can be achieved with Excel



Create models

- Model building is a helpful and potentially dangerous activity
- You have to explore your data first – usual statistical thinking
- Then propose a model, then challenge that model with the data and revise as many times as it takes

Kinds of analysis

- Classification – data into classes, e.g. sunny, windy etc.
- Regression – prediction, numeric target
- Clustering – e.g. segmenting a market
- Graphing – e.g. infographics, geospatial representations etc – typhoid Mary!
- Association analysis – e.g. the apocryphal nappies and beer Walmart example

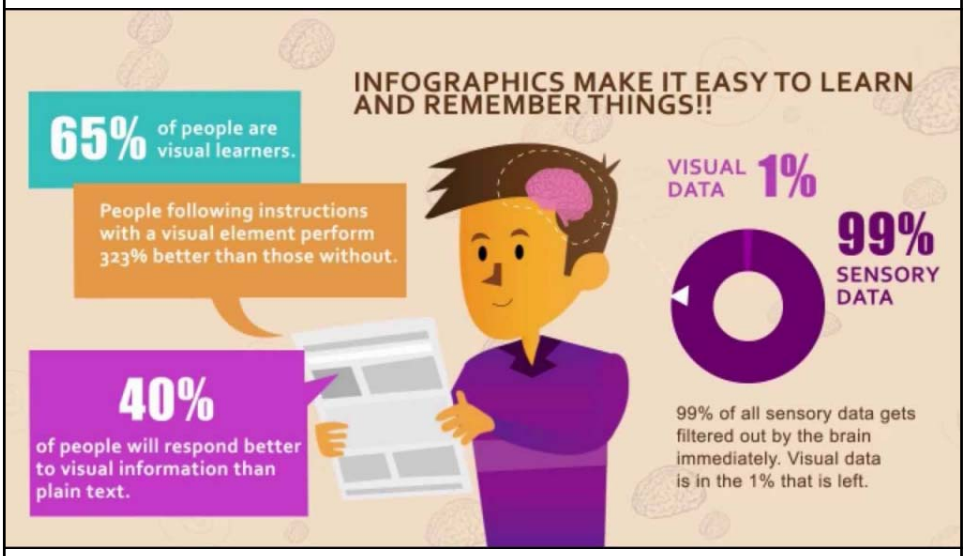
Kinds of models

- Spatial (co-variant)
- Time series (freq/time)
- Survival analysis
- Market segmentation
- Recommendations
- Association rules
- Attribution
- Scoring
- Indexing
- Predictive
- Clusters
- Supervised classification
- Extreme value
- Simulations
- Churn analysis
- Inventory movement
- Bidding optimisation
- Queueing

And there are many more

Actionable insights

- The very big challenge of communicating results
- Choosing what and how to present
- Answering the question
- Scientific integrity – a good story isn't helpful if it isn't backed up by the data
- Contrary results can still be interesting



An example

TEC – the bank of tertiary education

- Where does data come from
- Schools
- Population dynamics
- Transport
- Employment data

A desire to better understand the secondary-tertiary transition boundary led to some surprising insights

Skills of the tāuru kaitātari

In some part:

- Statistician – do these numbers make sense?
- Programmer – How can I make this do what I want?
- Detective – where can I get that data from?
- Scientist – what problem are we solving?
- Engineer – how can we do this?

Perhaps we are seeing the renaissance of the polymath?

Tools of the tāuru kaitatari

- Excel
- Python
- R
- Qlik-sense
- Tableau
- Matlab
- Javascript
- JSON
- D3
- XML/XSLT
- Google charts
- And many more

Attributes of the tāuru kaitatari

- Curious (aka nosey)
- Asking unexpected questions
- Numerate thinking
- Some programming

Our courses in 2018

- QUAN102 – the statistical thinking intro
- INFO151 – the tools and vocab of data
- INFO264 – data analytics
- INFO281 – machine learning

Capital thinking. Globally minded.