

The Role of Open Data, Open Source Software and Data Visualisation in Developing Quantitative Citizenship

Professor Di Cook, Econometrics and Business Statistics

Auckland Statistics Teachers' Day, Nov 25, 2016

H. G. Wells (1903) Mankind in the Making

“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write!”

- Data is available everywhere today, publicly, free
- Software, very powerful software, for analysis of data is available publicly, free
- Combined with a knowledge of mathematics and statistics empowers each of us to contribute to understand and improve our world
-
- I'm going to show you a few projects that I worked on using open data and open software, and how it helped me understand the world a little better: gender gap, climate change and politics. Oh, and where people walk in Melbourne city, too.

Math Gender Gap

The Sydney Morning Herald
National

work relax eat sleep

Investigations Interactives Health Education Public Service News World War 1 Centenary Clique Photos Photo Galleries

You are here: Home > National > Education >

Search here... National Search

Numbers point to maths 'gap'

May 2, 2011

Read later

Caroline Milburn

[Twitter](#) [Facebook](#) [G+](#) [Share](#) [in](#) [Submit](#)

[Email article](#) [Print](#) [Reprints & permissions](#)



Australia is one of the few countries with a maths gap in favour of boys.

GIRLS are performing at much lower levels in maths than boys — and the gap is widening as students progress through school.

A study of NAPLAN numeracy test results for students in years 3, 5, 7 and 9, to be presented

"Girls are performing at much lower levels in maths than boys - and the gap is widening as students progress through school."
SMH, 2011

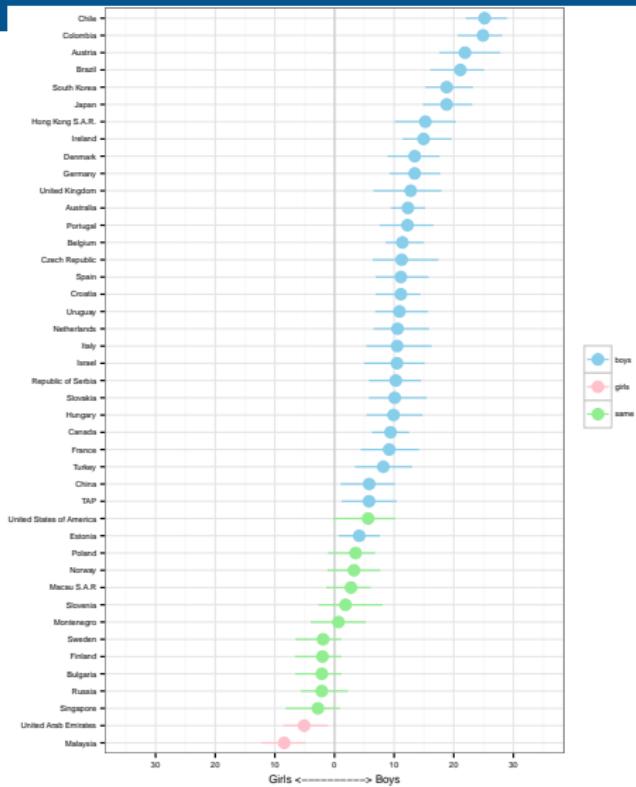
- OECD PISA survey “the world's global metric for quality, equity and efficiency in school education”.
- Workforce readiness of 15-year old students
- 500,000 students were tested across 65 countries and 18,000 schools
- Math, reading and science
- Data available from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm>

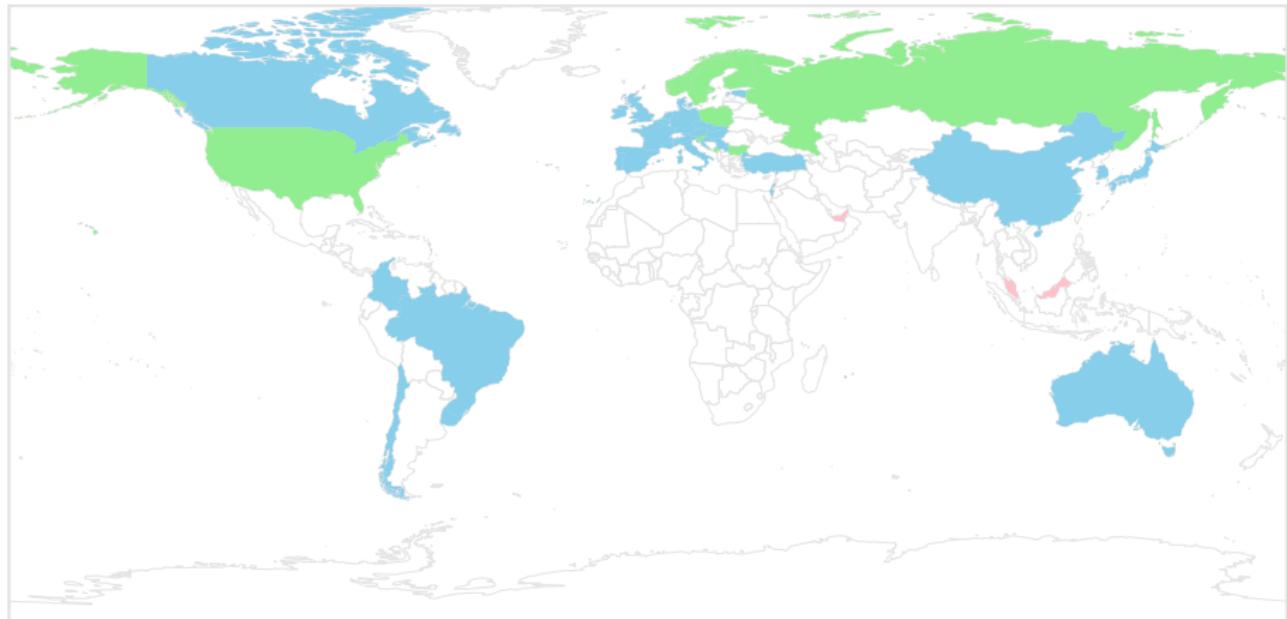
Measuring the difference

- 1** Compute the mean by country and gender
- 2** Difference the means: boys - girls

Bootstrap is used to produce the confidence intervals

- 1** Sample the sample, with replacement, separately by country
- 2** Compute the difference
- 3** Repeat 1 and 2, 1000 times
- 4** Keep the 2.5 and 97.5 percentiles, to yield 95% confidence interval

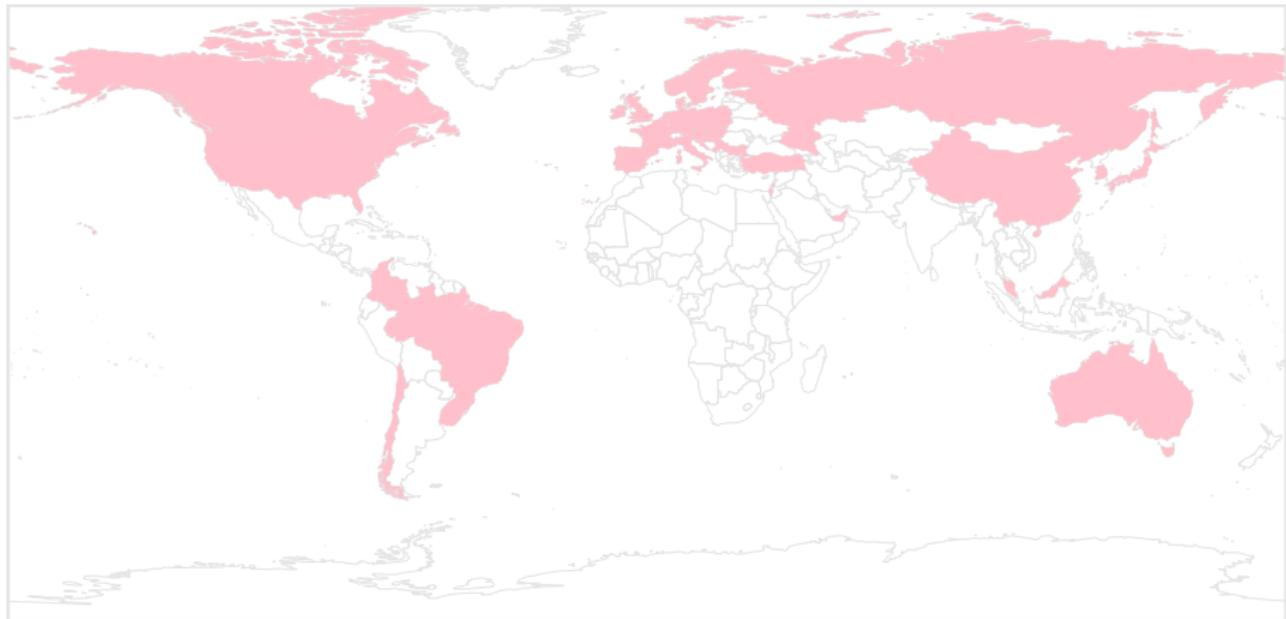




Reading Gap

M





Climate change: What is it about carbon dioxide?



UN Climate Conference Weather Climate Change Whale Watch Animals Conservation Energy

You are here: Home > Environment >

CSIRO team's study erodes credibility of key soil carbon model

November 1, 2015

Read later



Peter Hannam

Environment Editor, The Sydney Morning Herald

View more articles from Peter Hannam

Follow Peter on Twitter Follow Peter on Google+ Email Peter



459



11



in Share 1 submit



Reprints & permissions



Sydney Barrister Peter King argues a royal commission into banks' treatment of farmers is needed. Photo: Jessica Shapiro

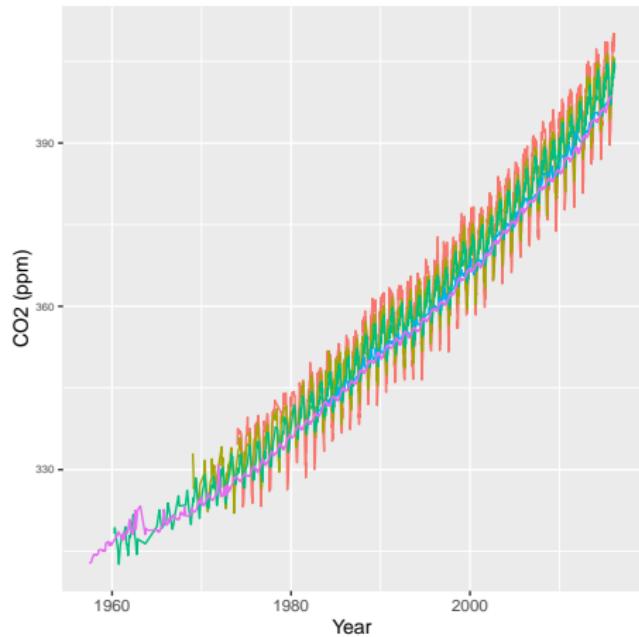
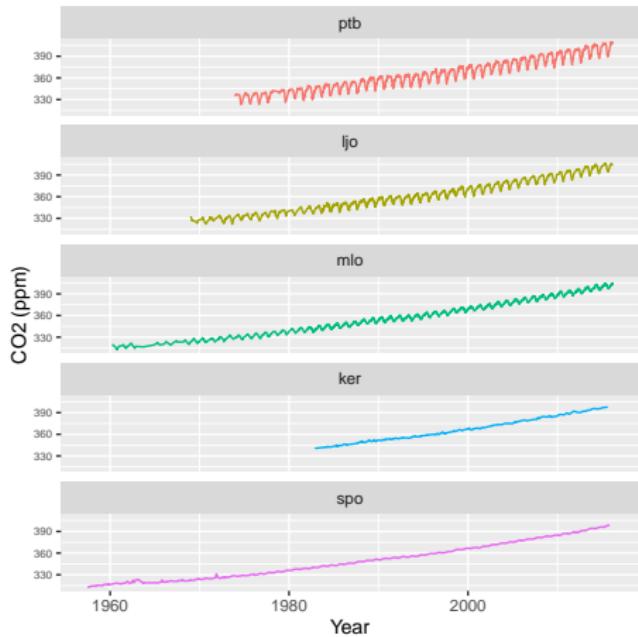
Australia's method of measuring how much carbon is being stored in its soil is flawed, undermining the credibility of government programs to pay farmers to sequester the climate change inducing element, a new study by CSIRO researchers has found.

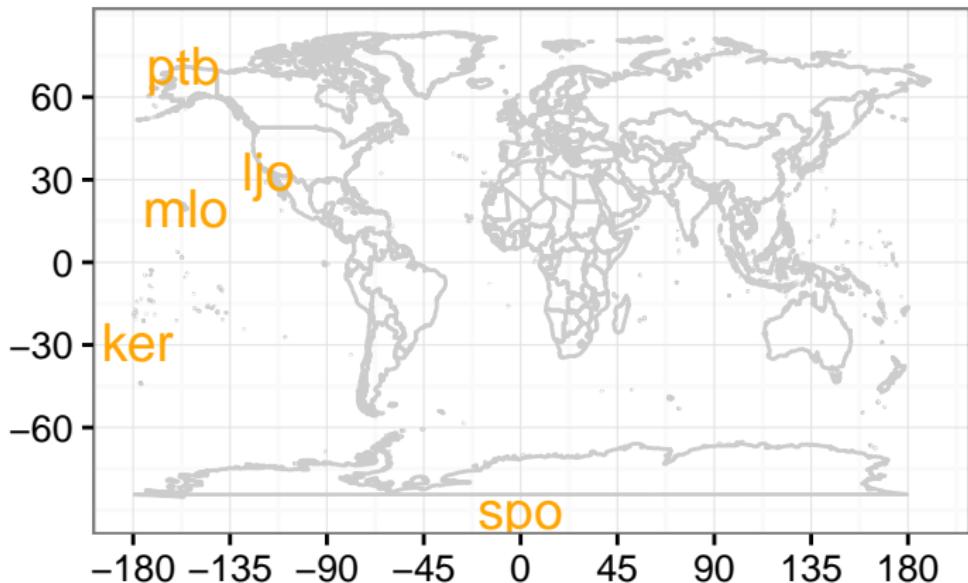
"Australia's method of measuring how much carbon is being stored in its soil is flawed, undermining the credibility of government programs to pay farmers to SEQUESTER the climate change inducing element."

- “Scientific consensus states that carbon emissions must be reduced by 80% by 2050 to avoid temperature rise of more than 2°C.” Carbon Neutral
- Carbon offsets: Carbon offsetting is the use of carbon credits to enable businesses to compensate for their emissions.
- Kyoto protocol in 1992, attempt to get international cooperation to reduce emissions.

Carbon dioxide data

- Data is collected at a number of locations world wide.
- See Scripps Inst. of Oceanography
- Let's pull the data from the web and take a look . . .
-
- Recordings from South Pole (SPO), Kermadec Islands (KER), Mauna Loa Hawaii (MLO), La Jolla Pier, California (LJO), Point Barrow, Alaska (PTB).





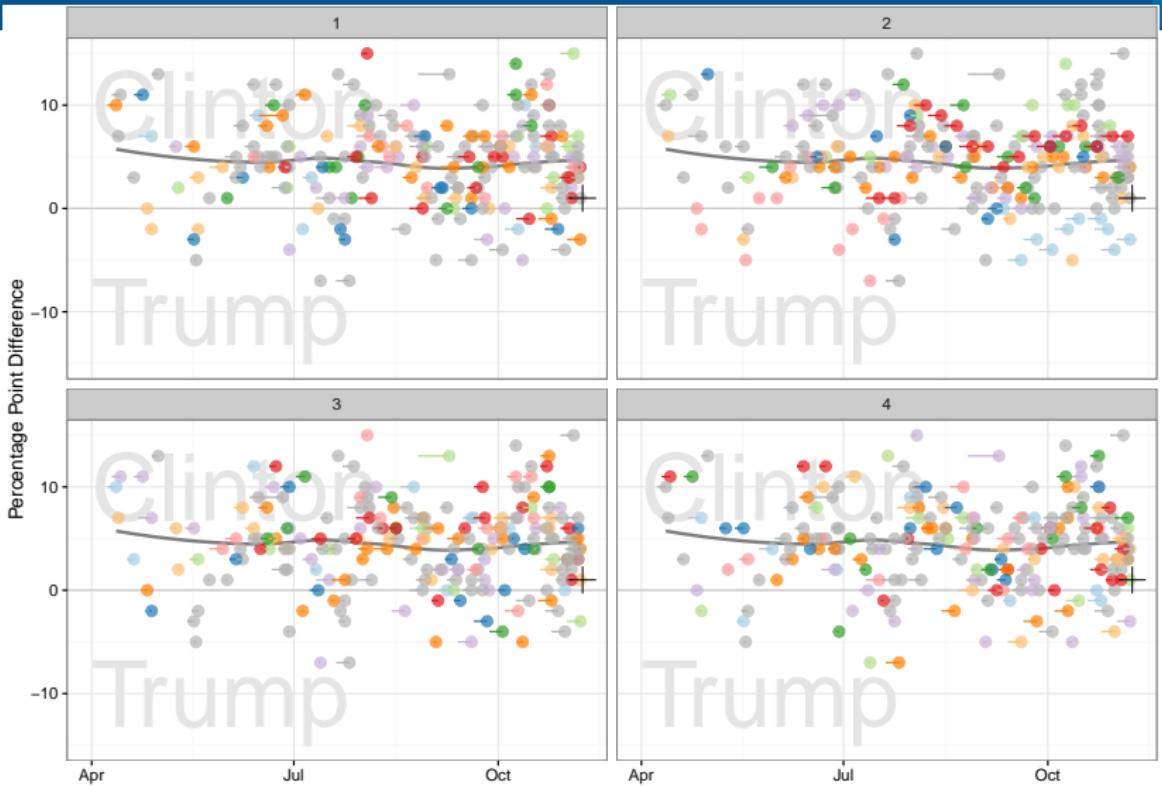
- CO₂ is increasing, and it looks like it is exponential increase. **I really expected that the concentration would have flattened out with all of the efforts to reduce carbon emissions.**
- The same trend is seen at every location - physical mixing of our atmosphere. **I was suspicious of the data on first seeing this because it looks too perfect.**
- Some stations show seasonal pattern - actually the more north the more seasonality - population centres and types of trees.

- US election process is complicated - depends on the electoral votes for each state in a winner takes all approach
- Polls released on a regular basis (Real Clear Politics)
- Monitoring the Election Visually
- Can You Buy a President? Politics After the Tillman Act

Play a game



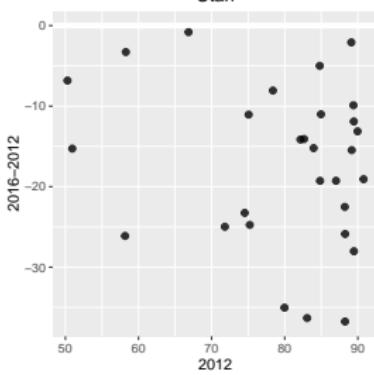
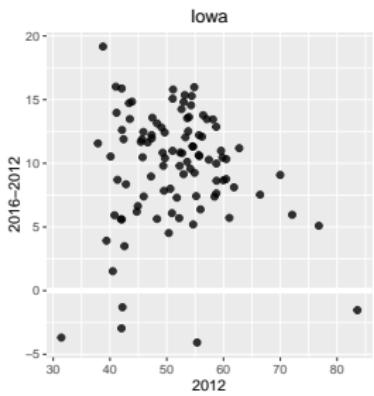
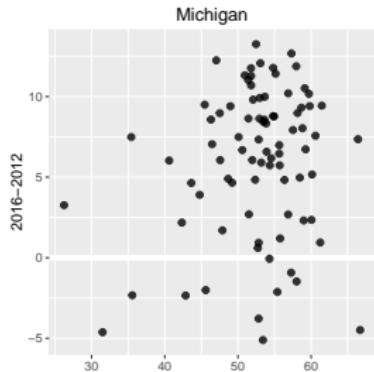
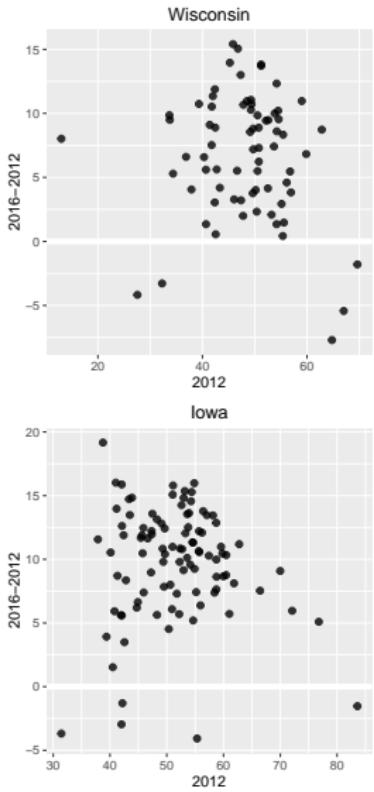
- 1 Scramble the association between pollster and result
- 2 Put data plot in a random location among the plots of permuted data samples
- 3 Ask your reader to pick the plot that is different



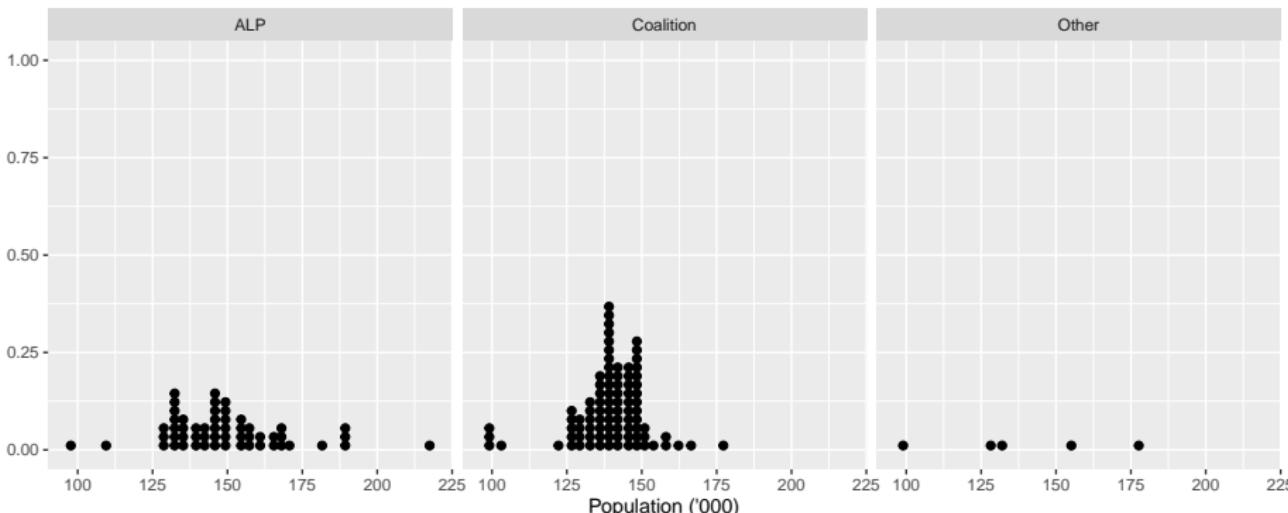
- Many pollsters are not operating objectively
- News stories that lead with a poll result may not be accurately reflecting the potential vote outcome
- This year's polls missed the mark

GOP Swing

M



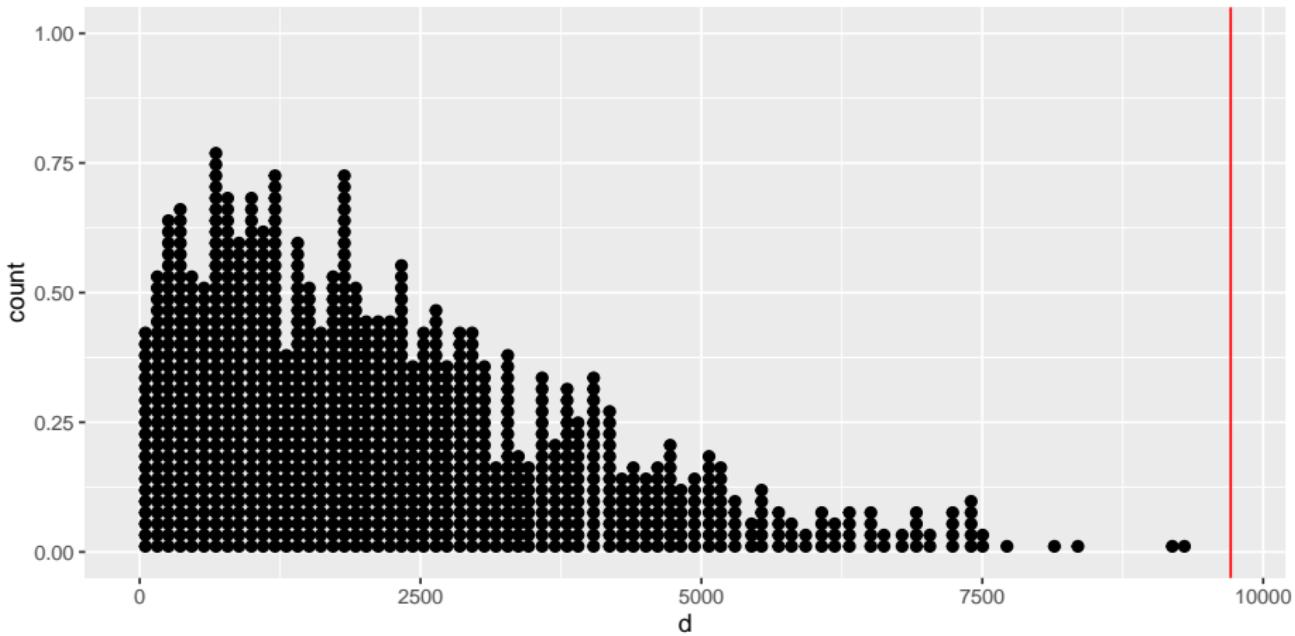
- First ever ROpenSci, Brisbane April 2016
- Project eechidna
- Explore the Australian electorate by voting patterns and demographic makeup
- Is the Australian electorate gerrymandered? Using permutation to assess this.

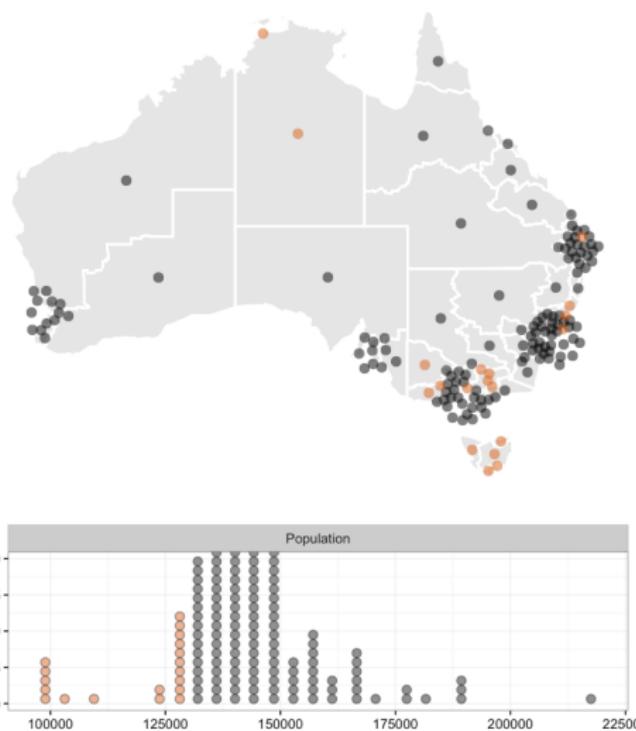


Permutation procedure

- 1 Subset to two major parties, Labor vs Coalition
- 2 Scramble the party label against electorate
- 3 Compute absolute value of difference in mean populations
- 4 Repeat 1-3 10,000 times
- 5 Plot permuted differences, mark observed data value

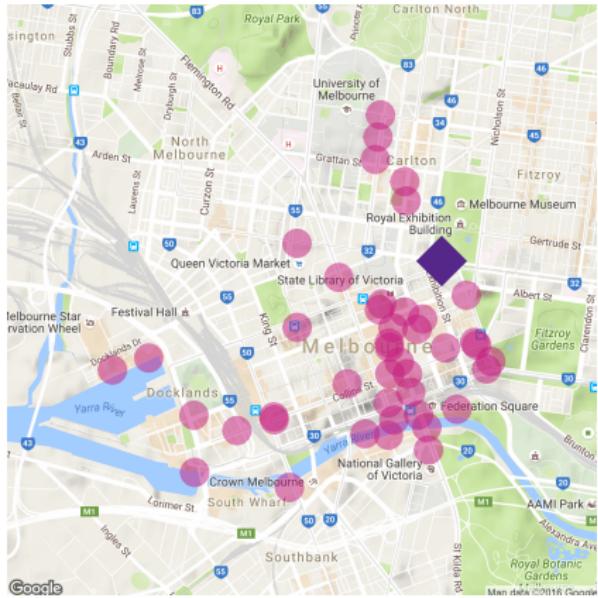
P-value is the number of differences greater than observed data divided by number of permutations

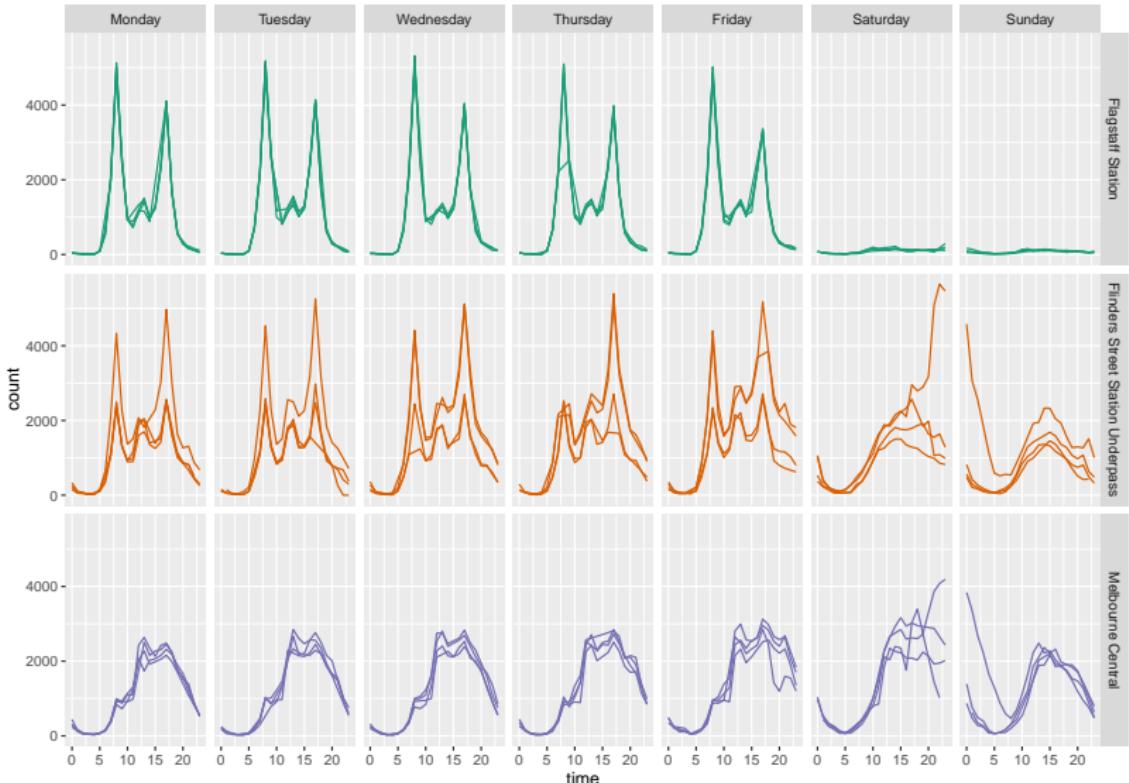




Pedestrians in Melbourne city

M





Big Data day for high schoolers

M

- Anomalous events
- High pedestrian traffic and potential for disasters
- Grand final day effects
- Weekend/week day train station patterns
- Shopping districts

With a laptop loaded with R, and access to open data, the world is your oyster!

- With knowledge of randomisation (bootstrap, permutation, simulation)
- Ability to make data plots (grammar of graphics, appropriate mapping of variables to graphical elements)
- We can learn a lot about our world

Slides and code at

[https://github.com/dicook/Auckland_Teachers_Day_Nov16.](https://github.com/dicook/Auckland_Teachers_Day_Nov16)

Share and share alike

M

This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.