Statistics for Empowerment: opportunities and challenges

Jim Ridgway

www.procivicstat.org

Erasmus+

Durham


North Sea
Edinburgh
Newcastle
Durham
Stockton
Irish Sea
Manchester
Birmingham
Cardiff
London
English Channel

Durham University

- 17,500 students (20% non-UK from 156 countries)
- About 30% post grad
- 3000 staff (30% non-UK)
- Students: collegial, charitable fundraising; sports
- About 70th in world rankings

Erasmus+

Statistics for Empowerment: opportunities and challenges

Jim Ridgway

www.procivicstat.org

Erasmus+

Early Statisticians

- Florence Nightingale (1820-1910) – health, infographics
- Enid Charles (1894-1972) – demography
- Elizabeth Scott (1917-1988) - weather, astronomy, equal opportunities
- Dorothy Adkins (1912-1975) - psychometrics

Royal Statistical Society Founders

- Henry Lansdowne - politician (Chancellor of the Exchequer)
- Charles Babbage - mathematician, engineer, astronomer and inventor of the computer
- John Elliot Drinkwater - administrator and champion of girls' education in India
- Henry Hallam - historian and political activist
- Richard Jones – economist

Pullinger (2014)

Statistics: Roots and Raison D'etre

- Tackle real problems using evidence
- Invent some mathematics
- Change the world
- Collaborate with sympathetic people – whatever their background

Statistics: Roots and Raison D'etre

- Tackle real problems using evidence
- Invent some mathematics
- Change the world
- Collaborate with sympathetic people – whatever their background

STATISTICS – still does

STATISTICS EDUCATION – has lost the plot??

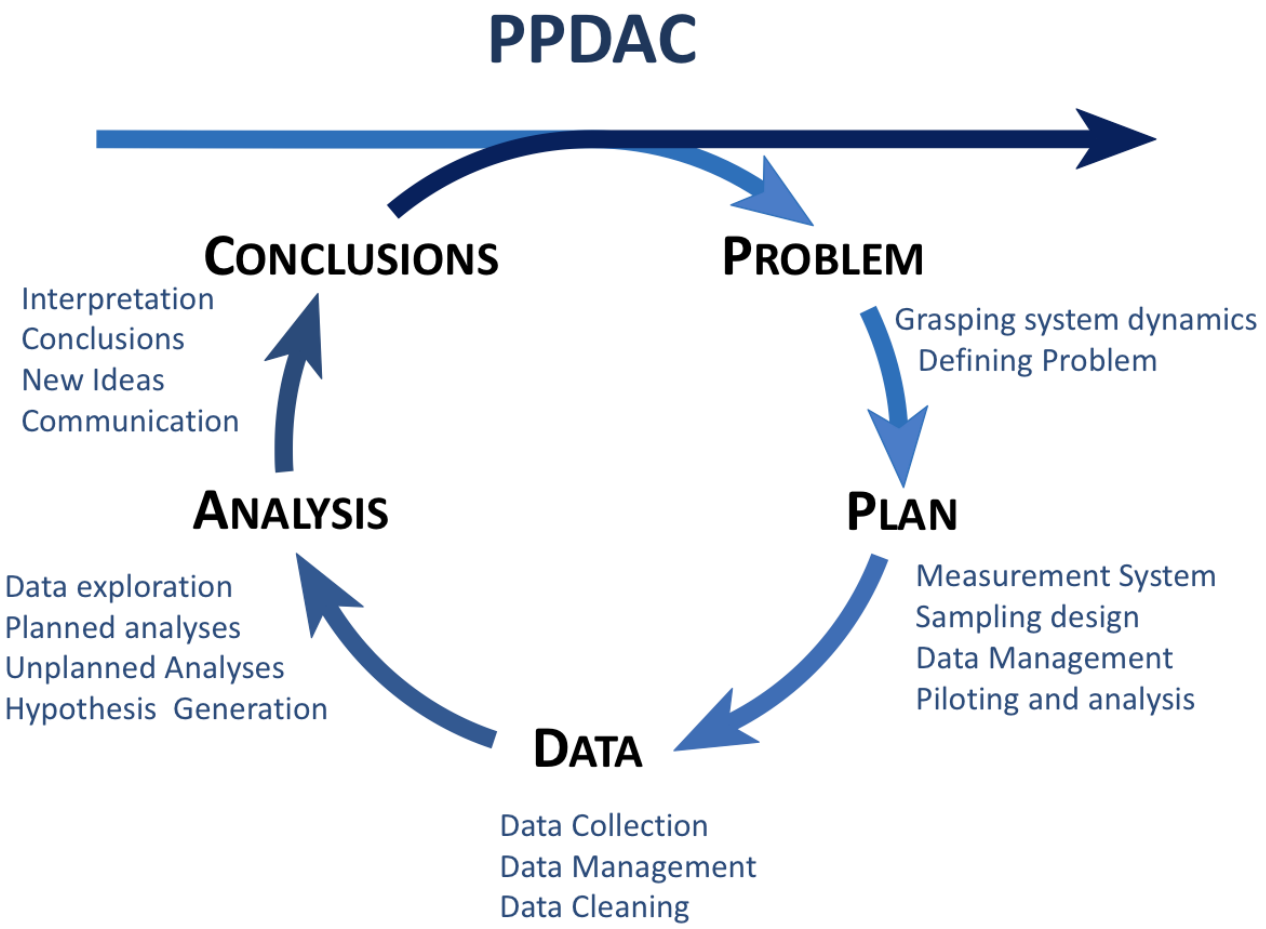
George Cobb…

Mere renovation is too late: we need to rethink our undergraduate curriculum from the ground up

American Statistician

doi:10.1080/00031305.2015.1093029

Wild & Pfannkuch (1999)



Critique of Statistics Education

- Doesn't address enough of the PPDA Cycle
- Stuck in the 1900s
  - Over-value tractable mathematical models
  - Believe in 'simple to complex' pedagogy
- Conservative
  - Focussed on what *we* know, not on student interests or needs
- Resists algorithmic thinking

Opportunities and Challenges

- Newfangled models are available
  - Most of the stats in the introductory curriculum was invented before 1900 (t-test 1908)
- New contexts
  - Social upheaval – we need more evidence informed decision making
- New resources
  - Open data, big data, social media, data visualisation
- New players
  - Data science, data driven journalism, fact checkers, FAKE NEWS
- New audiences
  - Citizens, politicians, social scientists

Opportunities and Challenges

- Newfangled models
  - Most of the stats in the introductory curriculum was invented before 1900 (t-test 1908)
  - Embrace models that everyone uses every day – decision trees, pattern recognition... BUT NOT TODAY!
- New contexts
  - Social upheaval
  - Use data about social phenomena; multivariate, non-linear, with interactions

Erasmus+

Opportunities and Challenges

- New resources
  - *Open data*, big data, social media, *data visualisation*
  - Use them

Sample Open Data Sources

- OECD *https://data.oecd.org/ World Bank*
- European Union Open Data Portal http://open-data.europa.eu/en/data/
- The CIA World Factbook https://www.cia.gov/library/publications/the-world-factbook/

Big Data Sources

Topsy http://topsy.com/
A searchable database of public tweets going back to 2006 as well as several tools to analyze the conversations

Likebutton http://likebutton.com/
Mines Facebook's public data – globally and from your own network

Big Data Sources

New York Times http://developer.nytimes.com/docs
Searchable, indexed archive of news articles going back to 1851

Freebase http://www.freebase.com/
A community-compiled database of structured data about people, places and things, with over 45 million entries

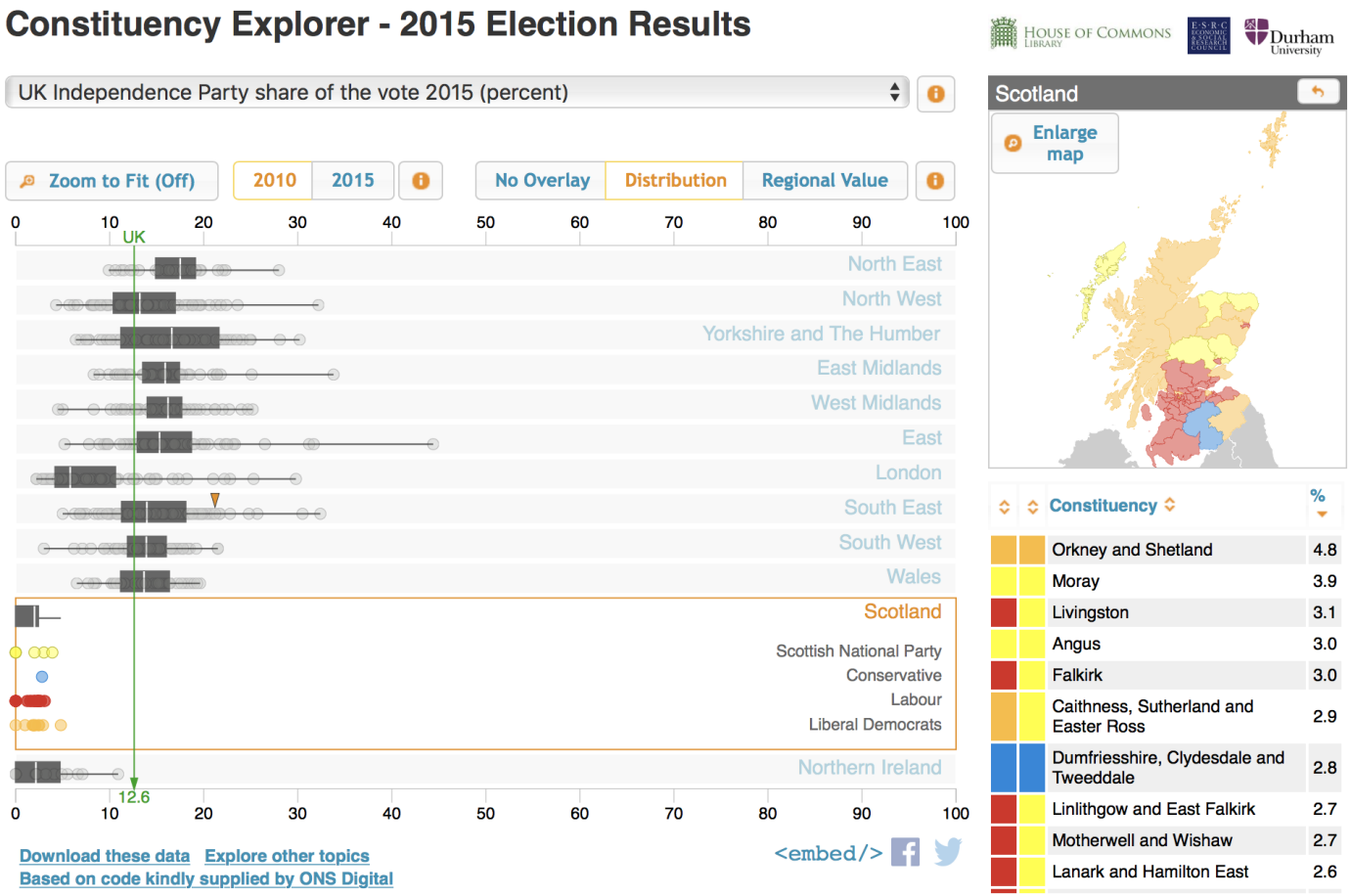
Million Song Data Set http://aws.amazon.com/datasets/6468931156960467
Metadata on over a million songs and pieces of music. Part of Amazon Web Services

Opportunities and Challenges

- New players
  - Data science, data driven journalism, *fact checkers*
  - Use them
  - FAKE NEWS
  - Teach statistical *habits of mind*
- New audiences
  - Citizens, *politicians*, social scientists
  - Teach students what they need to become informed citizens

Erasmus+

Interactive Open Data



Erasmus+

Statistical Ideas

- *Work with people who influence decision makers*
- *Heuristic: assemble lots of data into an accessible display*

Some Curriculum Approaches

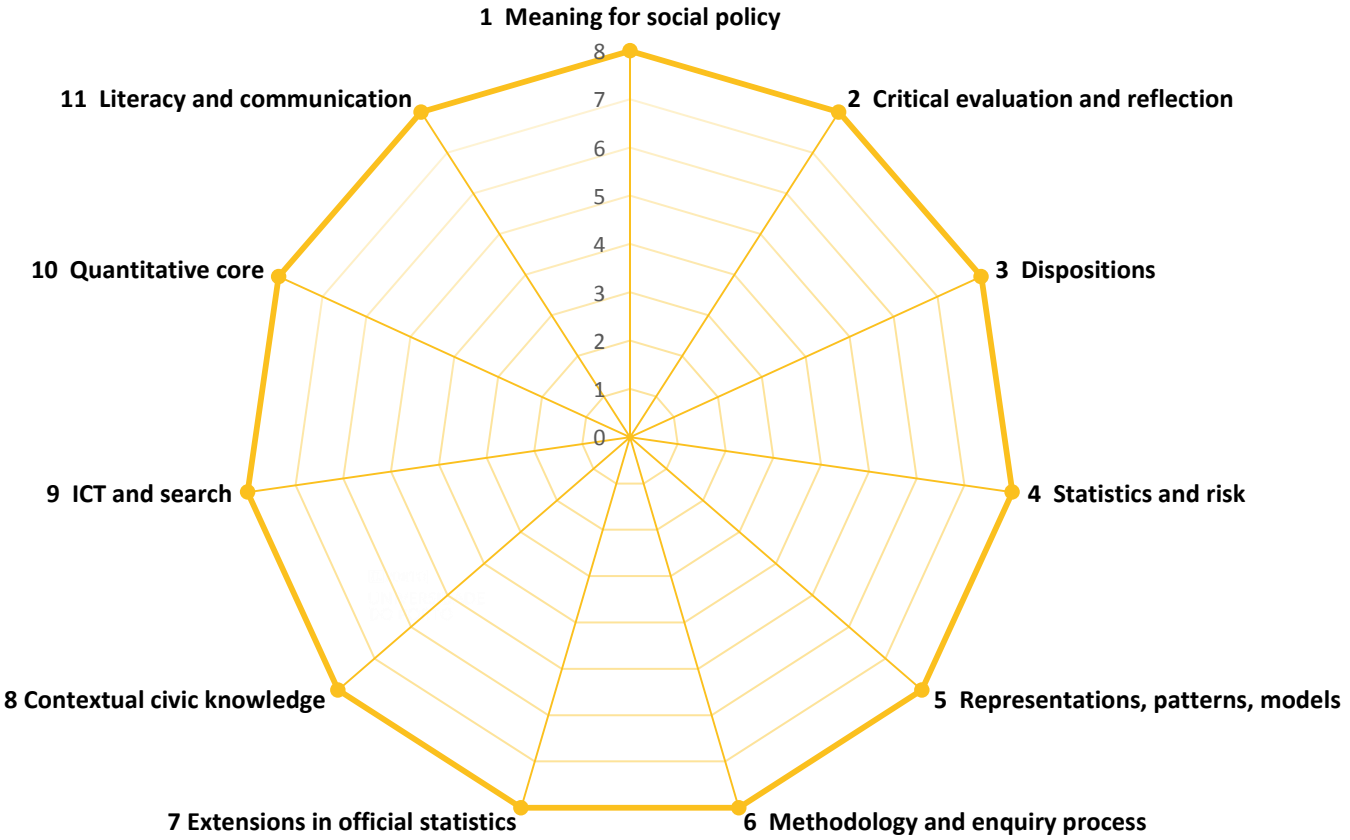
- Set Statistics in the context of *problem solving*
- More data visualisation, fewer equations
  - Formalise *after* understanding
- More large scale authentic data sets
  - To teach *basics* as well as *big ideas*
- Introduce multivariate thinking *early*
  - *Effect size, non-linearity, interaction*
- Show powerful, counter-intuitive results
  - *Birthday problem; when will you die?...*

Erasmus+

Thinking WITH and ABOUT Data

- Critical thinking
  - Exploring and Challenging
- Problem exploration
  - What evidence is out there? Can we get it, clean it, explore it?
- Decision making
  - OK, so what do we DO?
- BIG Ideas
  - What MUST we teach?

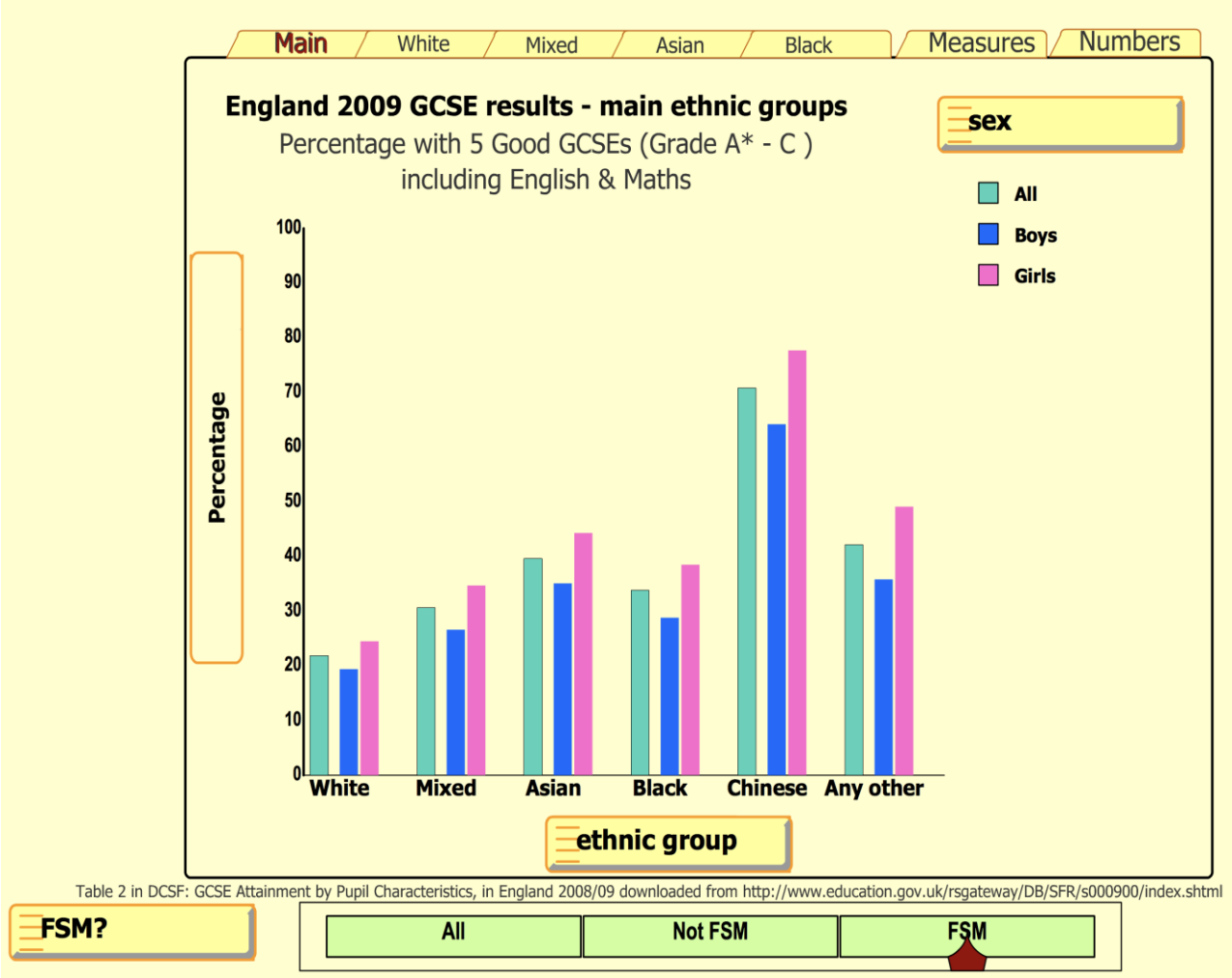
Some Dimensions of Literacy



Problem Exploration

- Explaining educational attainment
  - Intelligence?
  - Social class?
  - Poverty?

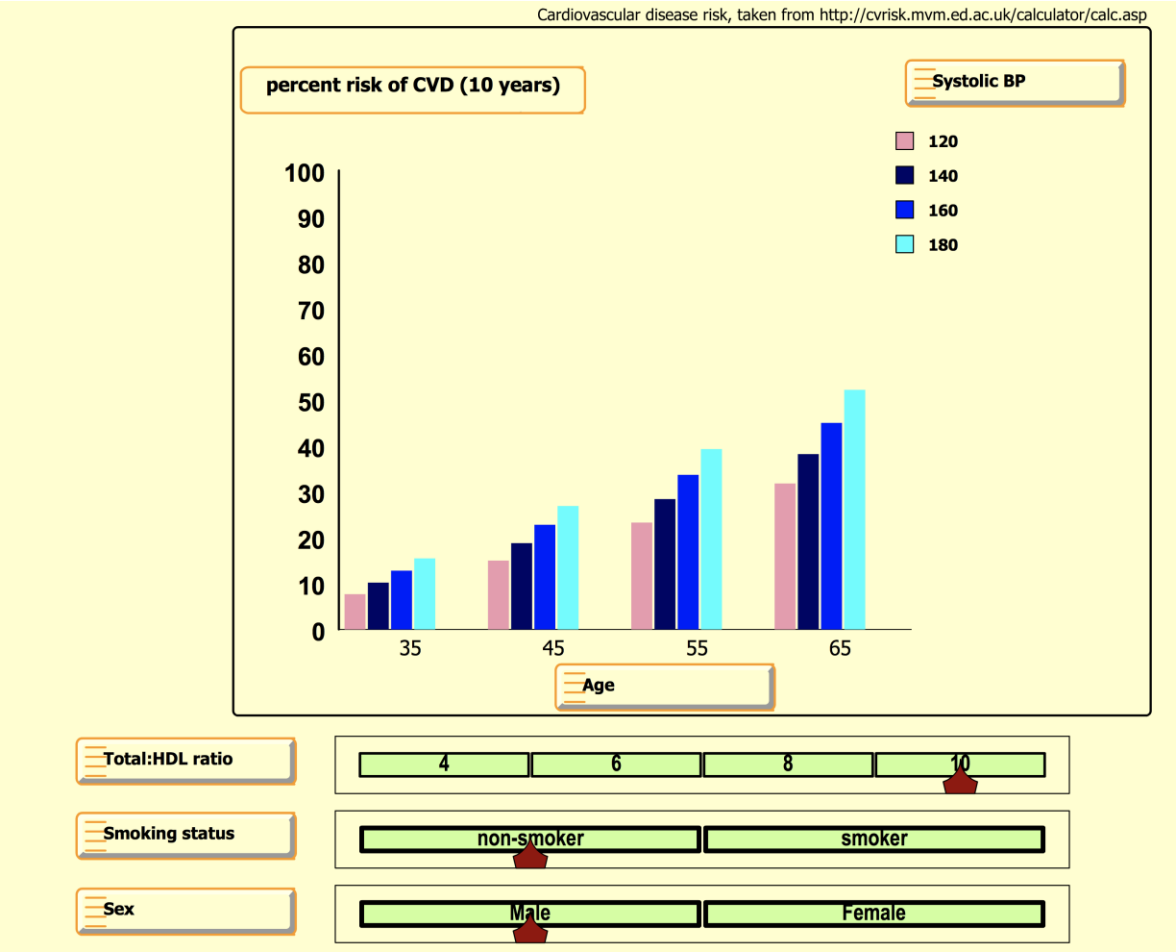
Problem Exploration



Statistical Ideas

- *Heuristic – look at the size of different effects*
- *Heuristic – pay attention to interactions*
- *Heuristic – look carefully at measures*

Decision Making



Evidence should INFLUNCE Decisions

- *Heuristic: externalise your beliefs and theories*
- *Heuristic: think about RISKS and COSTS when making decisions*

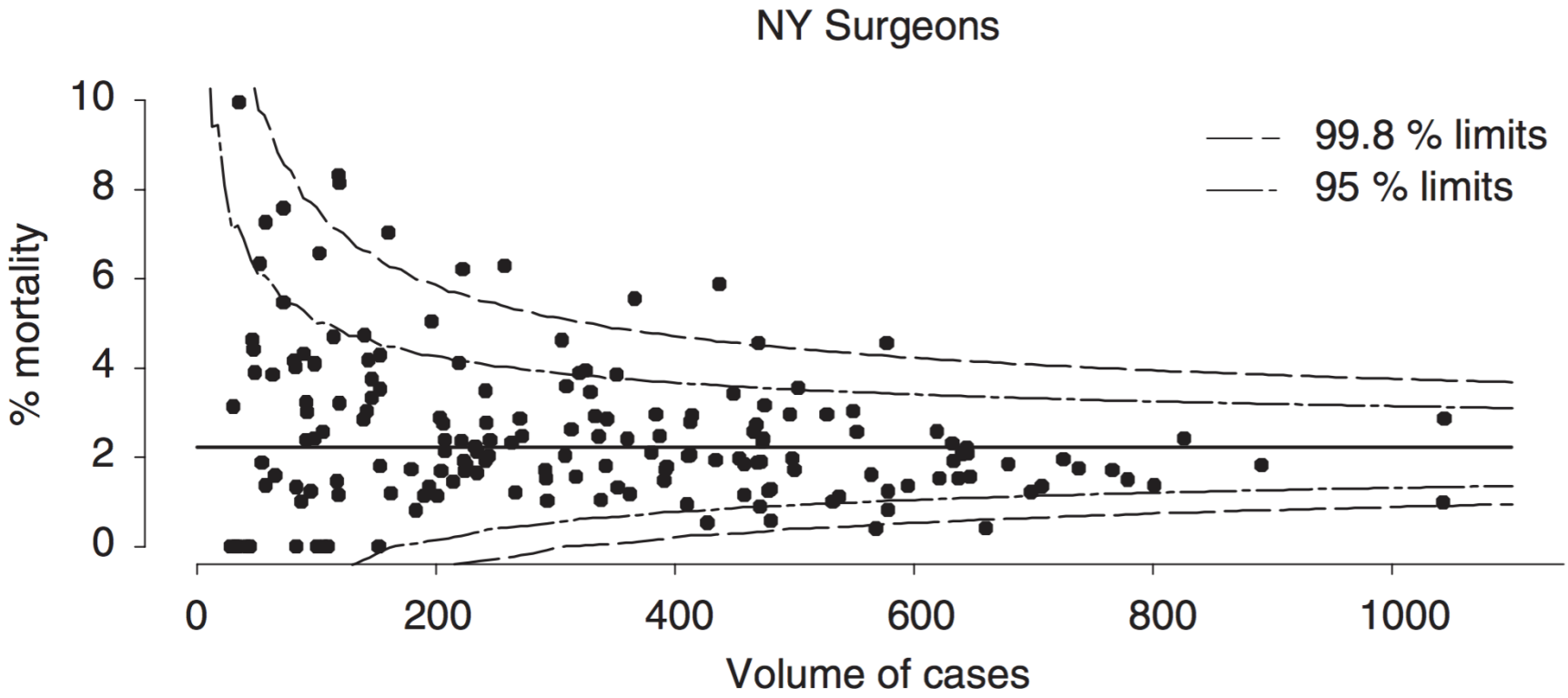
From Spiegelhalter (2005)

- Risk-adjusted 30-day mortality rates for coronary artery bypass grafts performed in New York State (1997–99) for 175 surgeons
- Funnel plots for comparing surgical performance, showing 95% and 99.8% control limits

- Who would YOU choose?

Erasmus+

Decision Making



Erasmus+

Statistical Ideas

- *Heuristic: look for robust evidence – sample size IS important*

Opportunities to Introduce Core Ideas

Measurement
Central tendency
Probability
Indices
Time series

Measurement: measures have different properties

- The international standard of *extreme poverty* - an income of less than 1$ a day (UNESCO, 2016)
- UK *relative poverty* - if household income is below 60% of the median household income (Department for Work and Pensions, 2014)

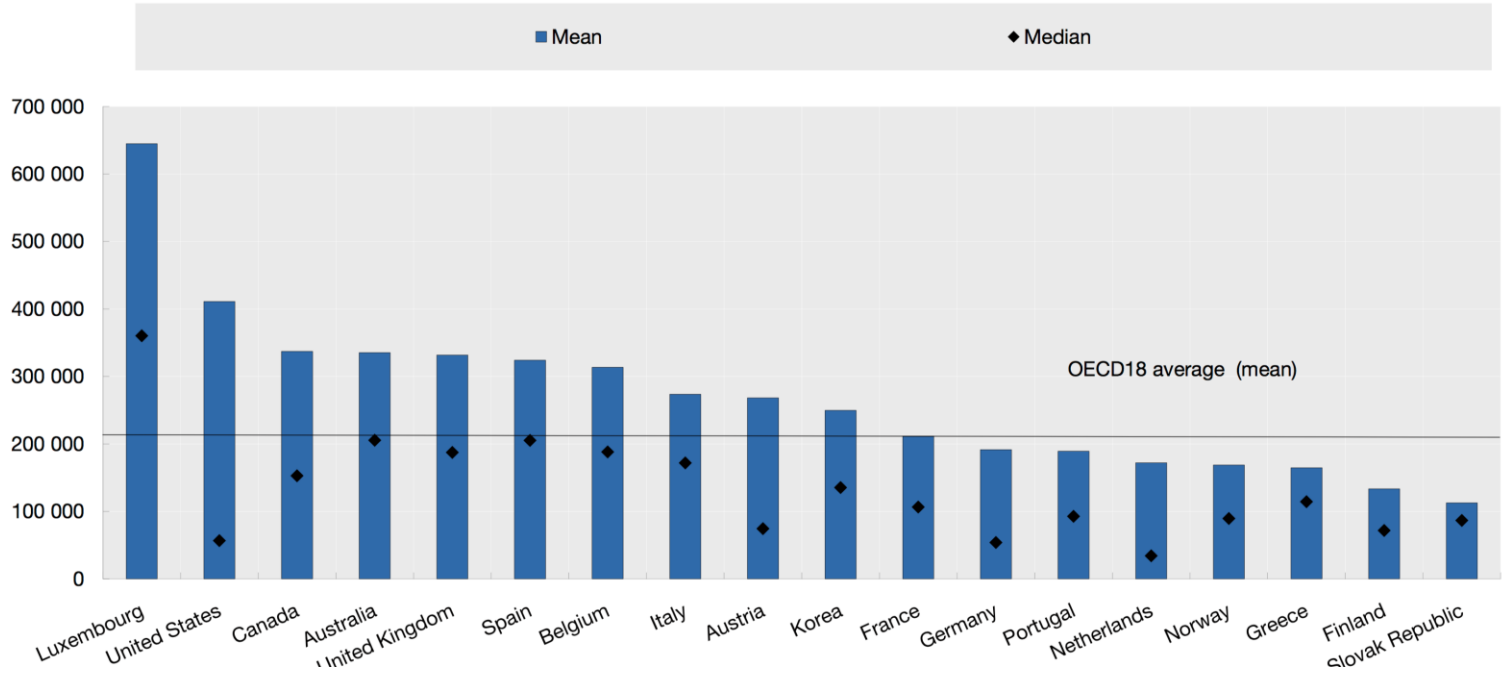
- ???? So if a county's economy is destroyed…

Central Tendency

- Why Donald Trump Won?

Means and Medians – from OECD

Figure 1. **Mean and median net wealth per household in selected OECD countries**
2010 or latest available year, values in 2005 USD PPPs



Erasmus+

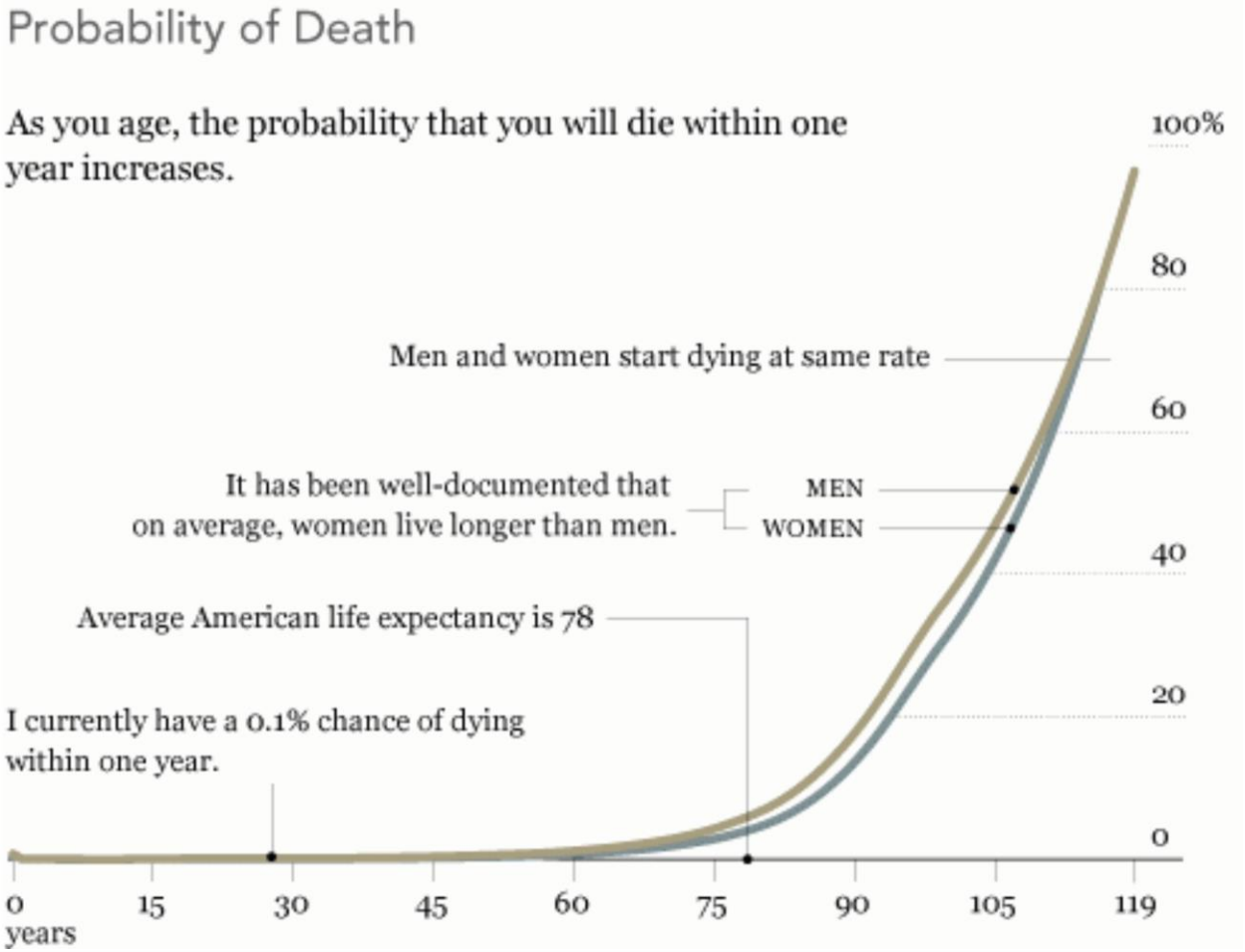
Probability

Jim is an American in average health for his age

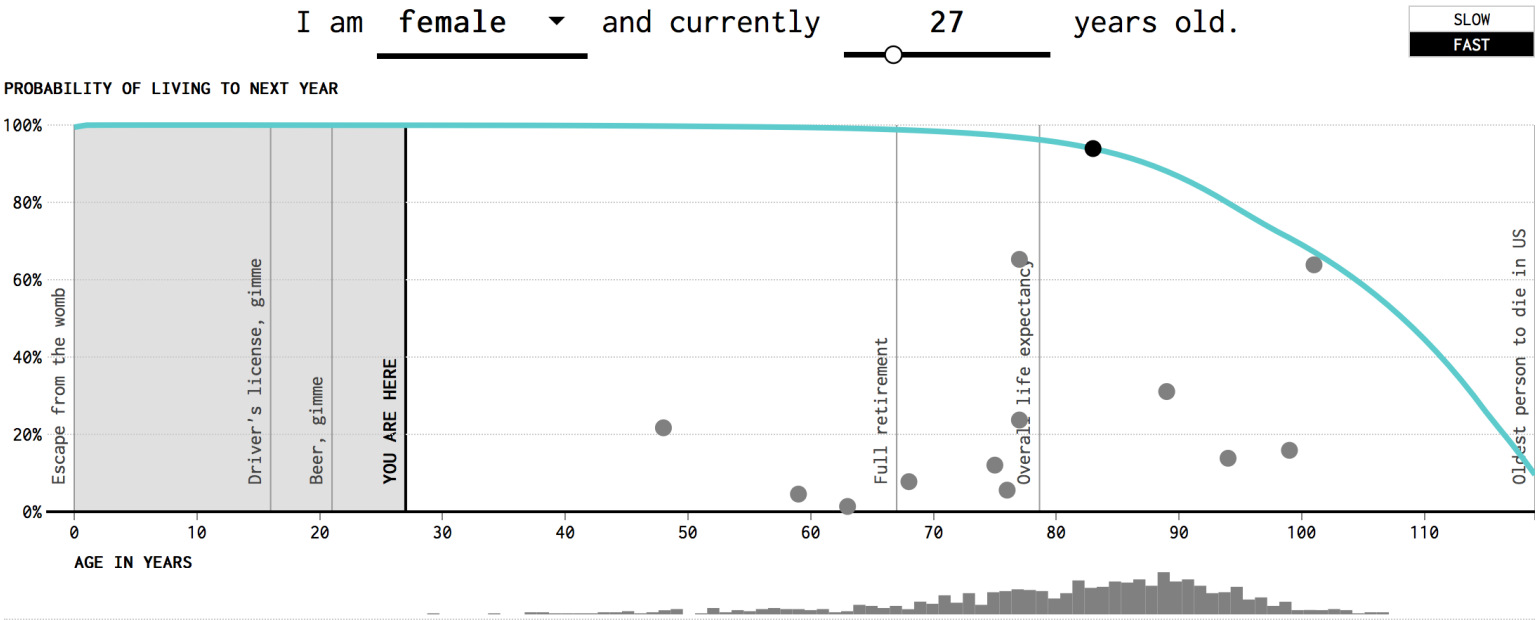
Next year, Jim has a 50% chance of death

How old is Jim?

50% chance of death next year?



Probability Density

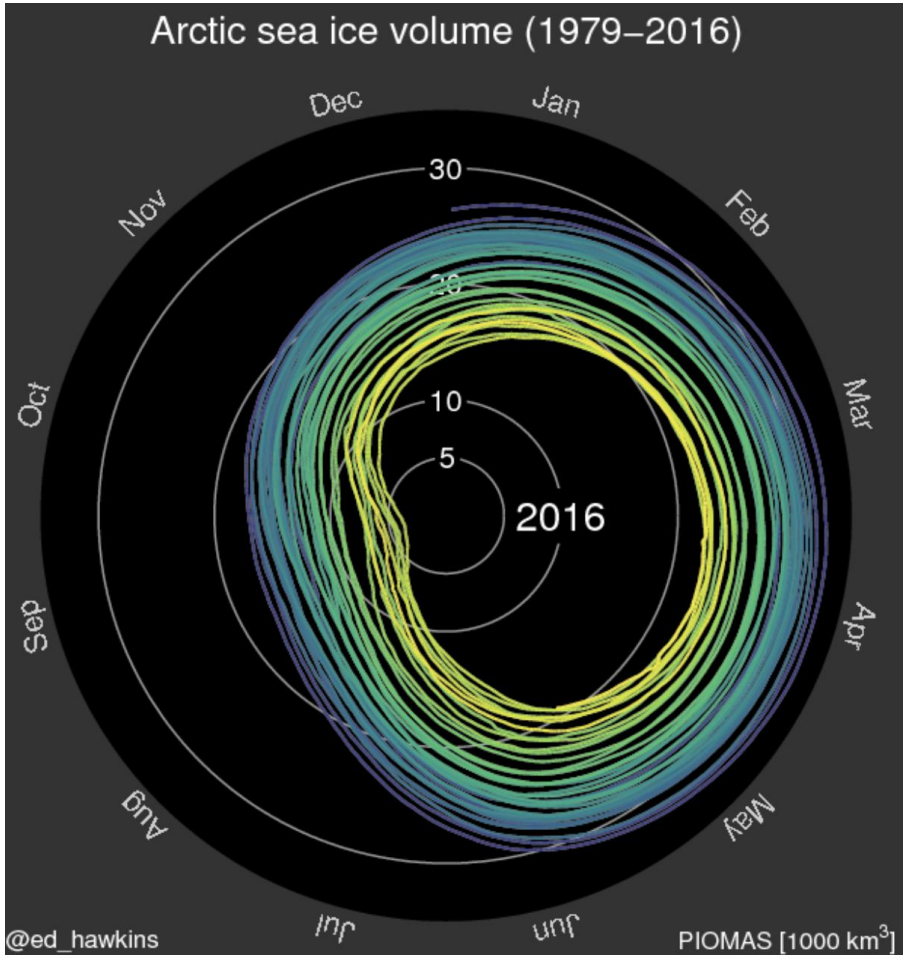


Probabilities For Years Left to Live

0 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 or more
0%	2%	3%	6%	15%	74%
(2)	(15)	(28)	(52)	(120)	(605)

Indices: OECD Better Life

Global Warming



Opportunities and Challenges

- Newfangled models
  - Most of the stats in the introductory curriculum was invented before 1900 (t-test 1908)
  - Embrace models that everyone uses every day – decision trees, pattern recognition...
- New contexts
  - Social upheaval
  - Use data about social phenomena; multivariate, non-linear, with interactions
- New resources
  - *Open data*, big data, social media, *data visualisation*
  - Use them

Erasmus+

Opportunities and Challenges

- New players
  - Data science, data driven journalism, *fact checkers*
  - Use them
  - FAKE NEWS
  - Teach statistical *habits of mind*
- New audiences
  - Citizens, *politicians*, social scientists
  - Teach students what they need to become informed citizens

Erasmus+

George Cobb (2015) *Am Stat*

- Our territory – thinking with and about data – is too valuable to allow old curricular structures to continue to sit contentedly on their aging assets while more vigorous neighbours take advantage of our latest ideas (p267)
- ...we will need an extended period of ferment, experimentation, and setting out to reach a new concensus... (p273)

Thank you for your attention

תודה על תשומת הלב שלך

Danke für ihre Aufmerksamkeit

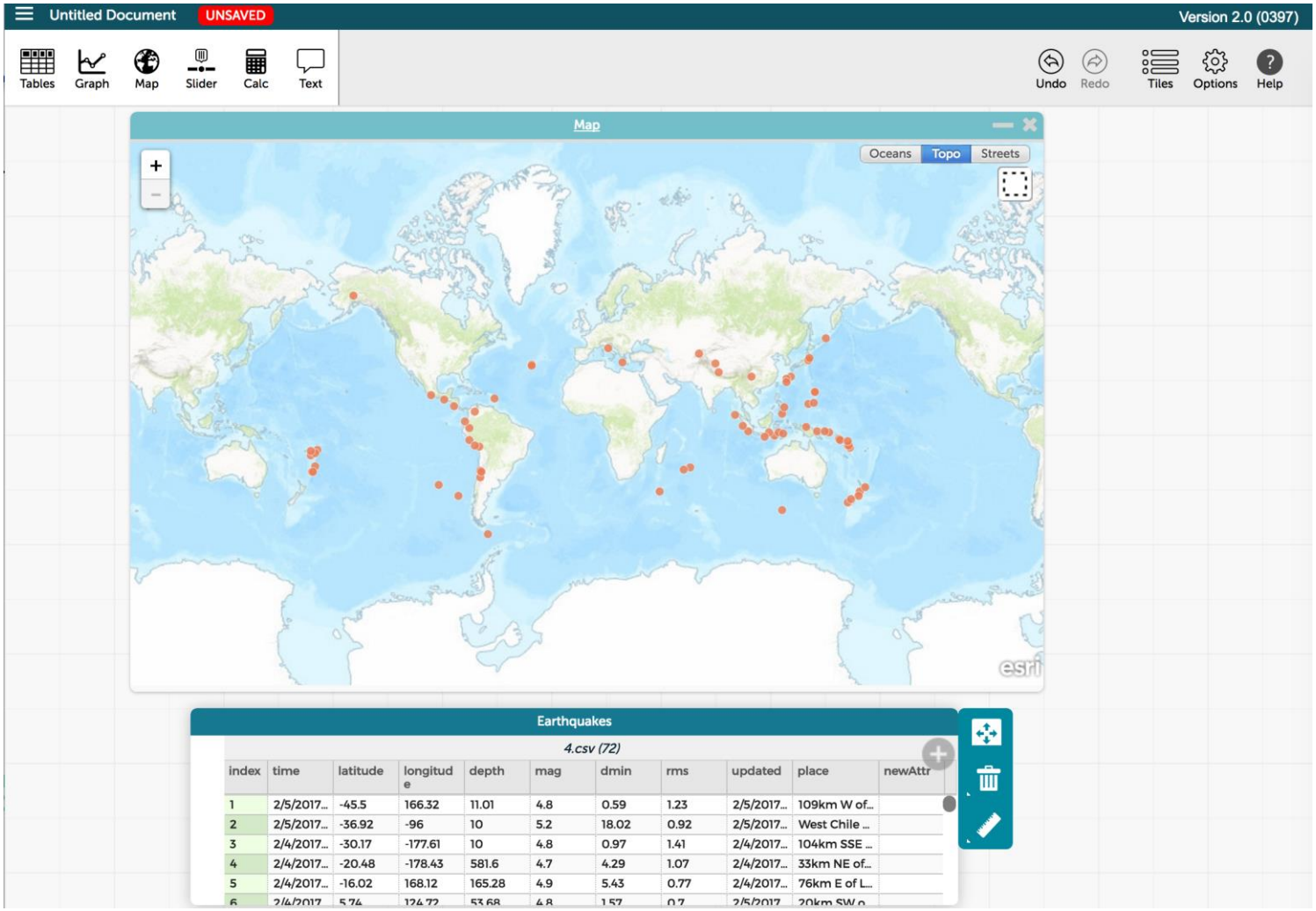
Danke für ihre Aufmerksamkeit

Obrigado pela sua atenção

Köszönöm a figyelmet

The preparation of this presentation has been partially funded with support from the European Commission

Data Science with CODAP



index	time	latitude	longitude	depth	mag	dmin	rms	updated	place	newAttr
1	2/5/2017...	-45.5	166.32	11.01	4.8	0.59	1.23	2/5/2017...	109km W of...	
2	2/5/2017...	-36.92	-96	10	5.2	18.02	0.92	2/5/2017...	West Chile ...	
3	2/4/2017...	-30.17	-177.61	10	4.8	0.97	1.41	2/4/2017...	104km SSE ...	
4	2/4/2017...	-20.48	-178.43	581.6	4.7	4.29	1.07	2/4/2017...	33km NE of...	
5	2/4/2017...	-16.02	168.12	165.28	4.9	5.43	0.77	2/4/2017...	76km E of L...	
6	2/4/2017	5.74	124.72	53.68	4.8	1.57	0.7	2/5/2017	20km SW o	

Erasmus+