

# The Sceptical Data Scientist

Dr Linda McIver  
Australian Data Science Education  
Institute



# Consider the t-shirt

---



Data Science is, indeed, a  
superpower

---



But it's actually one  
everyone could have

---



“Without seeing the data, we can't really be aware the problem exists and if we're solving it.” -- Larene Le Gassick

---



Data enables Change -- Also Larene

---



Hence...

---

Australian Data Science Education  
Institute - [adsei.org](http://adsei.org)



# Why this?

---

- What is science?
- What is data science?
- Why do we need more scepticism?
- How do we get more scepticism?
- What can you do about it?





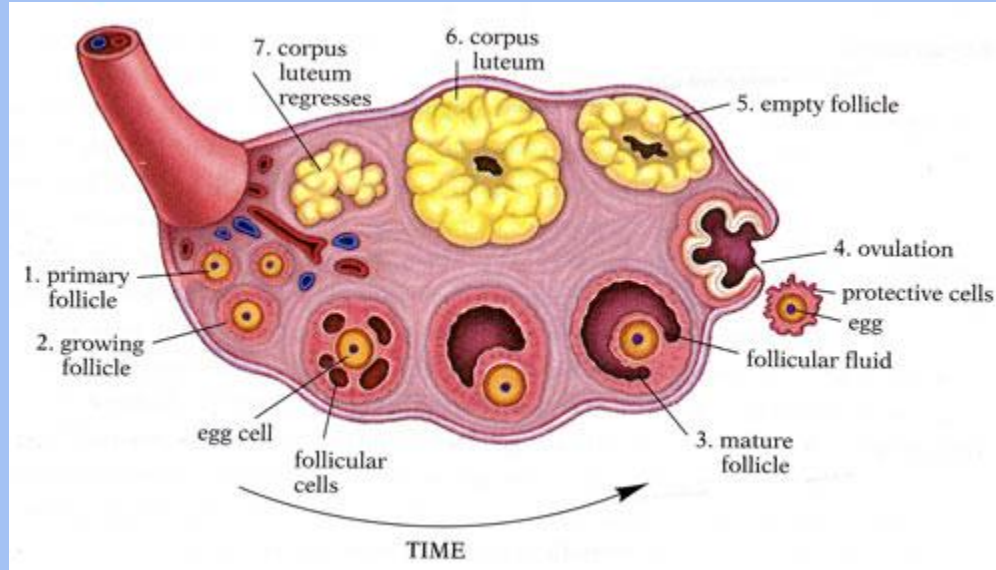


Image credit: [wikipedia](https://www.wikipedia)  
Ref: <https://www.sciencedaily.com/releases/2012/07/120726180259.htm>

Antibiotics + alcohol = ???



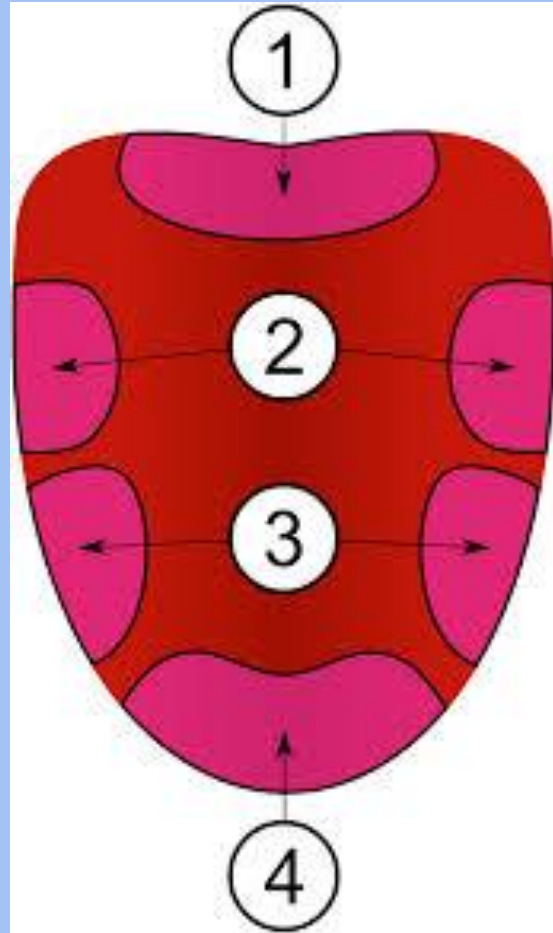


Image credit: [wikipedia](https://en.wikipedia.org)



Debunking these myths required scepticism

---



# How is science done?

---



# Is it by proving theories?

---



Science is (or should be) done by  
**DISPROVING** theories

---




# We have the same problem in Data Science

---





 Tesla's autopilot is 40%  
safer than a human driver!!

---



Except... oops... turns out they  
actually increased by 59%

---



# *The result*

---

- We think in terms of *THE* result
- But how many datasets only have one possible result/analysis?



# The dangers of screen time!

---

“their many variables and observations are often explored with an **analytical flexibility** that marks small effects as statistically significant, thereby leading to potential false positives and conflicting results”

<https://www.nature.com/articles/s41562-018-0506-1>



# The dangers of screen time!

---

*“This makes it necessary to employ transparent and robust analytic practices”*



... as opposed to....???

| Small Unit Installation Postcode | Previous Years (2001-2016) - Installations Quantity | Previous Years (2001-2016) - SGU Rated Output In kW | Jan 2017 - Installations Quantity |
|----------------------------------|---|---|-----------------------------------|
| 4670                             | 10,598  | 34,298.31   | 67                                |
| 6210                             | 10,101  | 25,929.69   | 77                                |
| 4655                             | 9,346   | 27,713.17   | 55                                |
| 4551                             | 8,611   | 23,956.86   | 31                                |
| 4350                             | 7,766   | 25,187.09   | 31                                |
| 6065                             | 7,174   | 22,991.62   | 99                                |
| 4211                             | 6,990   | 23,588.38   | 71                                |
| 4305                             | 6,773   | 19,862.10   | 26                                |
| 4740                             | 6,443   | 26,016.07   | 34                                |
| 4207                             | 6,399   | 20,911.23   | 37                                |
| 6155                             | 6,341   | 18,627.15   | 55                                |
| 4570                             | 6,248   | 20,379.99   | 40                                |
| 3029                             | 6,203   | 18,688.43   | 38                                |
| 3977                             | 6,146   | 18,643.95   | 82                                |
| 6164                             | 5,986   | 17,725.29   | 72                                |
| 4550                             | 5,800   | 17,804.00   | 00                                |



average solar installations by state between 2001 and 2016

QLD 170.03

WA 113.0

Vic 212.13

NSW 246.05

TAS 202.39

NT 91.5







# What is Data Science?

---



Less Science...  
More “Ta da!”

---



# How we teach Science

---

- We teach Science with experiments where we know the outcome
- And if the kids don't get the expected outcome?
  - They fake it
  - They certainly don't ask why
  - Or consider that what they expect might not be true



# How we teach Science

---

- So we encourage confirmation bias
- We *discourage* critical thinking
- And we encourage reporting the *expected* finding



**EVEN IF WE DIDN'T  
ACTUALLY FIND IT!**



# What if

---

- We teach science differently
- We do experiments in class where we don't know the outcome?
- We encourage and reward kids for asking difficult questions?
- We spend time exploring WHY we got those results



# What if

---

- We use science in schools to tackle the replication crisis?
- What if science classes try to DISPROVE existing studies?



# What if

---

- We teach data science the same way
- We encourage kids to be sceptical of the data stories they are presented with
- And even more sceptical of their own results





# What if

---

The first question we ask of a result is

“How many ways could this be wrong?”

And the second is

“How many ways could this be biased?”



# What if

---

The first question we ask of a dataset is

“How many ways could this dataset be flawed?”

And the second is

“What’s missing from this dataset?”



But who has time??



Step 1: Data

Step 2: ???

Step 3: Profit



Free labour!



Imagine if the default state of data were open



# Imagine if...

---

- Students could learn data science by testing other people's results
- their first encounter with data is to be taught to question it



# What would Data Science look like?

---

- We could produce Facial Recognition software that doesn't categorise people of colour as gorillas
- Self-driving cars that are NOT more likely to hit people of colour
- Official government studies that do NOT show that Tesla's autopilot is 40% safer when it's actually significantly less safe



# Imagine if...

---

We could do this in schools!





# Good News!

---

We can!

That's what The Australian Data Science Education  
Institute is all about



We teach teachers to teach kids to be

*Data Scientists*

Not Data Cowboys



# Even Better News!

---

- You can help!
- Give us your datasets
  - And annotate them so we can make sense of them
- Take our projects into schools
- Join our volunteer army
  - [bit.do/adseiSlack](https://bit.do/adseiSlack)
- Sponsor a Workshop
- Talk to me about other ways to help...



# Thank you for your scepticism!

---

@lindamciver

[linda@adsei.org](mailto:linda@adsei.org)

<https://adsei.org>

[bit.do/adseiSlack](https://bit.do/adseiSlack)

