

Tutorial: Up and down the PyData Stats Stack

Or Lies damned lies and statistics



Peadar Coyle

Who am I?



I work as a Senior Data Scientist at Channel 4
We're a media company, and we leverage data for targeted advertising, customer analytics and recommendation engines
We're hiring, so have a chat with me if interested.

What else do I do?

- Open Source Contributor to PyMC3
- Mathematics and Physics background
- Fellow of the Royal Statistical Society and member of NumFOCUS
- Author of interviews with data scientists book :)



Stats is everywhere...

Just 1% Of Men Take Shared Parental Leave

The policy was meant to allow women to return to work sooner after giving birth, but most say they don't want to share leave.

14:44, UK,
Tuesday 05 April 2016



[Video: Progress Report On Shared Leave](#)

Sky TV customers get
Sky Broadband Unlimited
£5 a month for 12 months

Monthly line rental £17.40 applies

[Find out more](#)

An advertisement for Sky TV. It features a white panda with black ears and eyes, looking slightly to the side. The text above the panda reads "Sky TV customers get Sky Broadband Unlimited £5 a month for 12 months". Below the text is a small note about monthly line rental. There is a pink "Find out more" button and the Sky logo at the bottom right.

Top Stories



Police Dig At
Former Home Of
Wests' Friends

What was wrong
with this?



Fathers and Babies

[http://www.bbc.co.uk/programmes/b075th
gl](http://www.bbc.co.uk/programmes/b075thgl)

More or Less debunked this - the sample was wrong!

**WHAT WOULD THE
WORLD LOOK LIKE**



**IF STATISTICS
WASN'T MISUSED?**

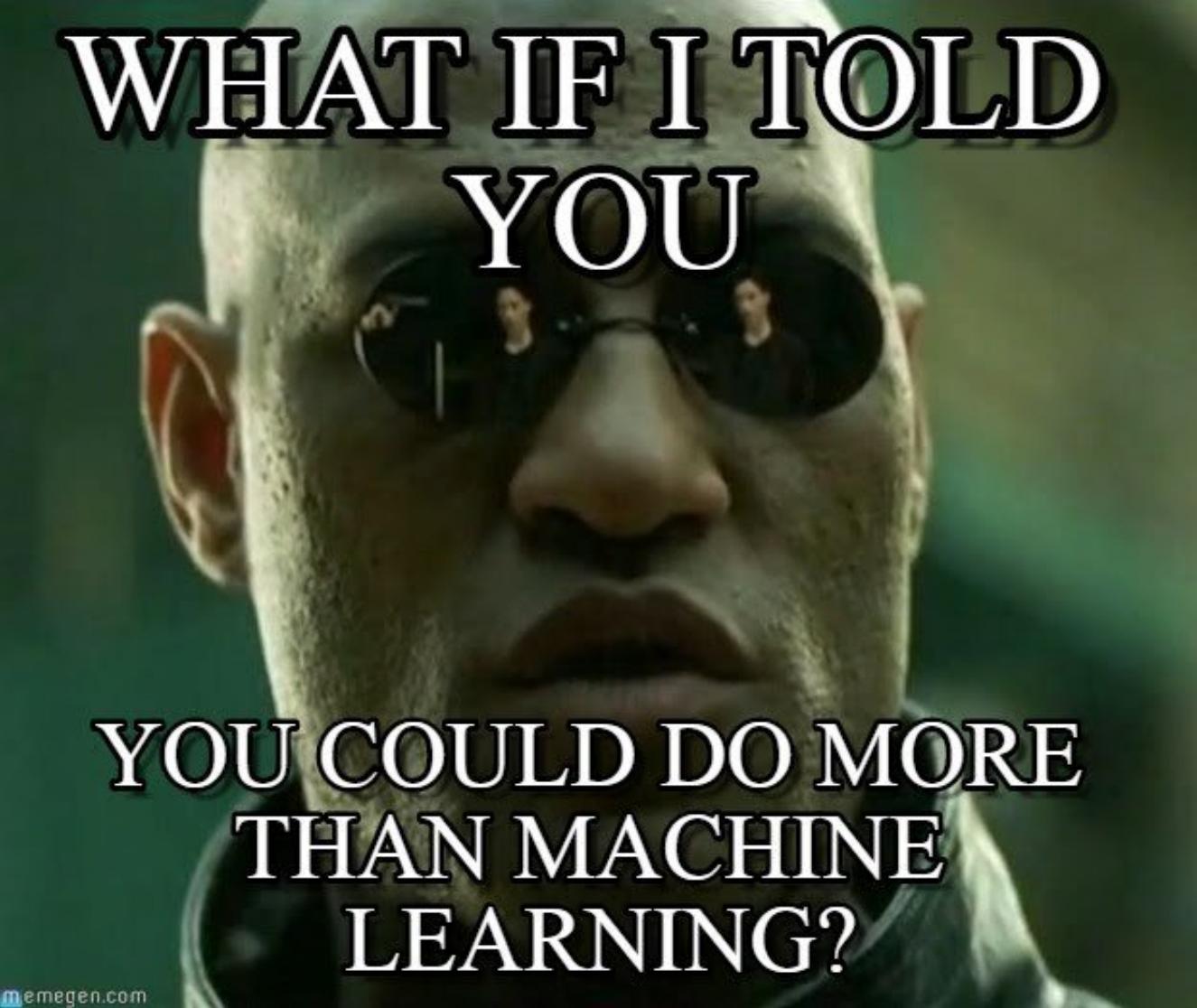
**YOU'RE GOING
TO TEACH US**

**ALL OF
STATS?**

Outcomes of this tutorial

- Have three tools to attack the same problem
- Understand how to interpret and debug these models
- Apply your understanding from a different area to say a new area like Bayesian or Frequentists

<http://bit.ly/pydataldnstats2>

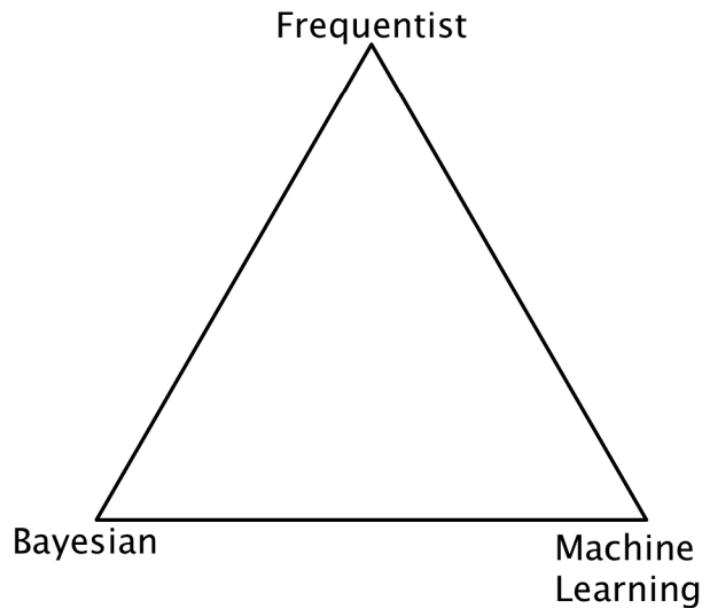


**WHAT IF I TOLD
YOU**

**YOU COULD DO MORE
THAN MACHINE
LEARNING?**

Different schools of data

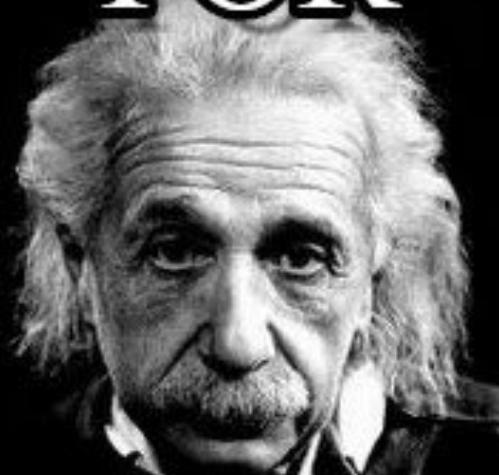
HT: Vincent D. Warmerdam

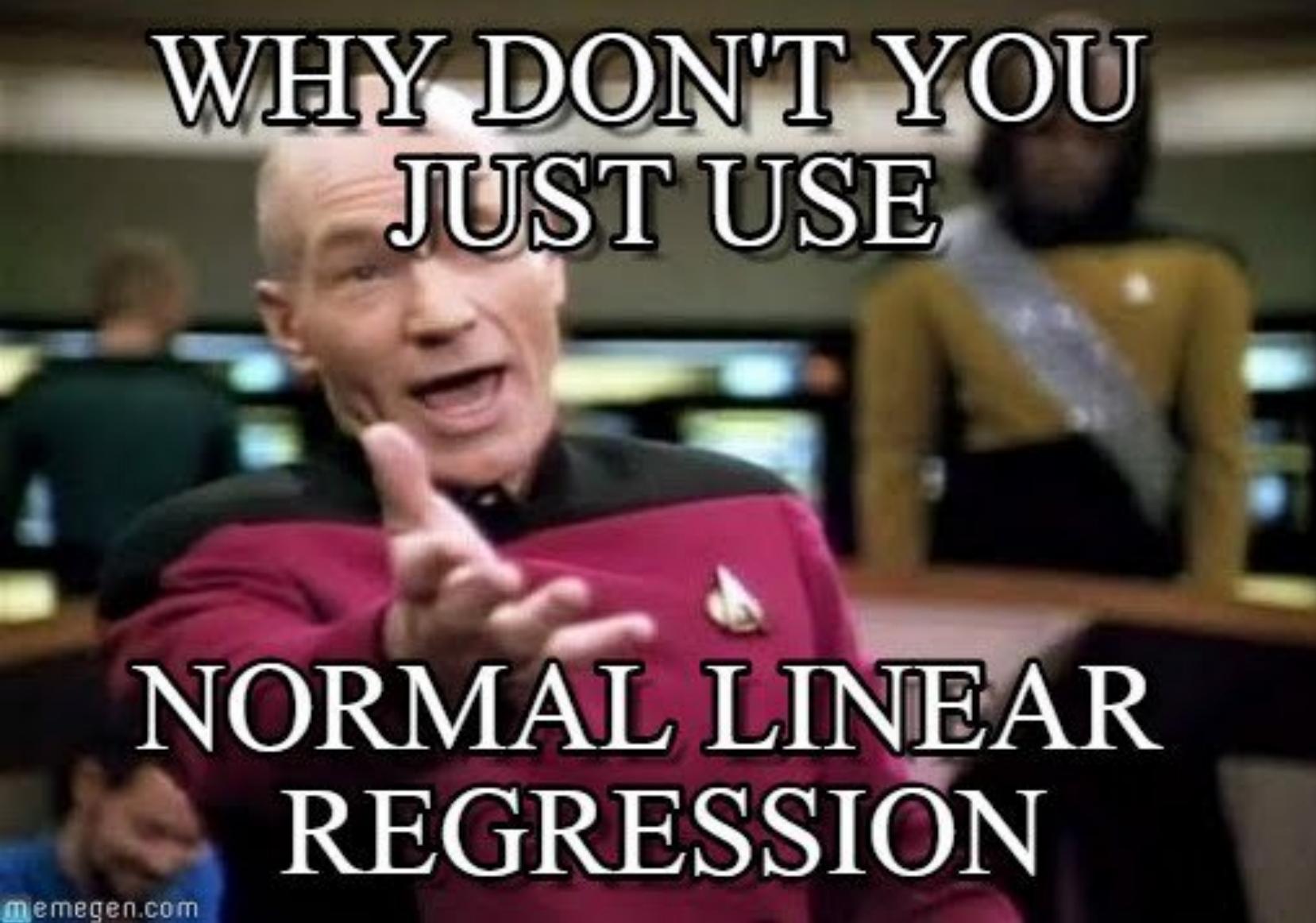


Firstly hypothesis testing

- Or how to do t-tests and all that stuff

IT TOOK AGES
FOR
MY HYPOTHESIS
TO GET TESTED

A black and white portrait of Albert Einstein, showing him from the chest up. He has his characteristic wild, grey hair and a prominent mustache. He is looking slightly to the right of the camera with a thoughtful expression.



**WHY DON'T YOU
JUST USE**

**NORMAL LINEAR
REGRESSION**

Three schools

- ScikitLearn
- Statsmodels
- PyMC3

Bayesian tooling



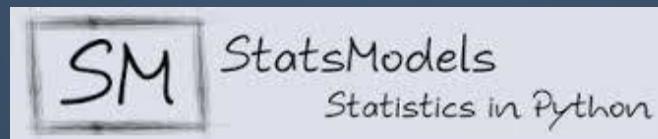
PyMC3

https://pymc-devs.github.io/pymc3/Bayesian_LogReg/

How likely am I to
make more than \$50K?

Statsmodels

Frequentist Logistic Regression





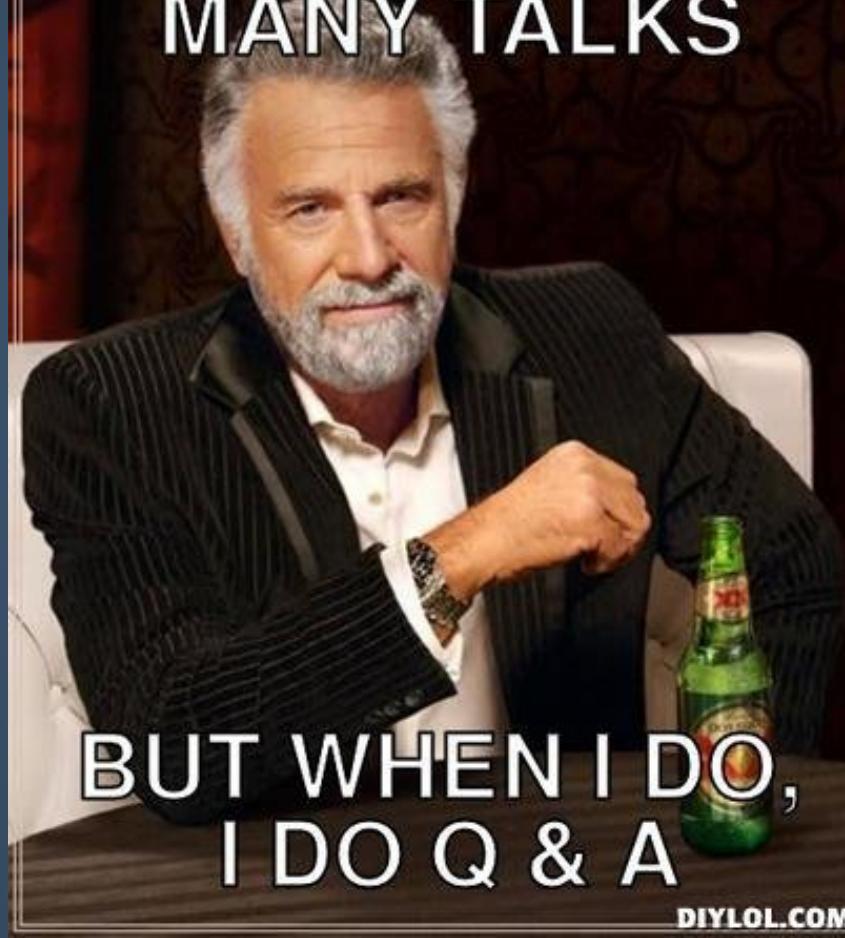
**YEAH, IF YOU COULD JUST SHOW ME
SOME CODE**



THAT'D BE GREAT

memegenerator.net

I DON'T GIVE
MANY TALKS



BUT WHEN I DO,
I DO Q & A

DIYLOL.COM