Michael Grossberg

Intro to Data Science

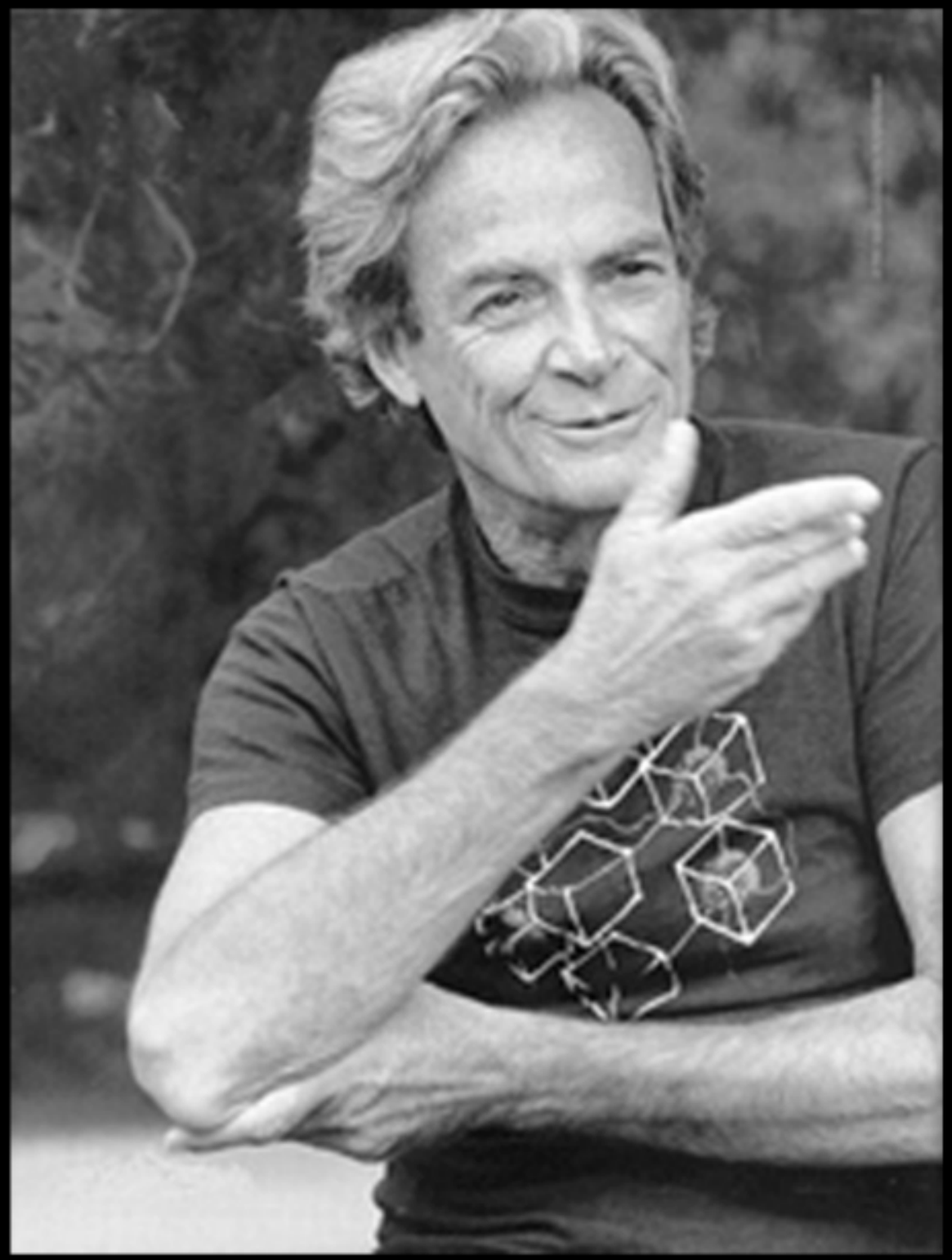
CS59969

Some Stats

Why care about stats?

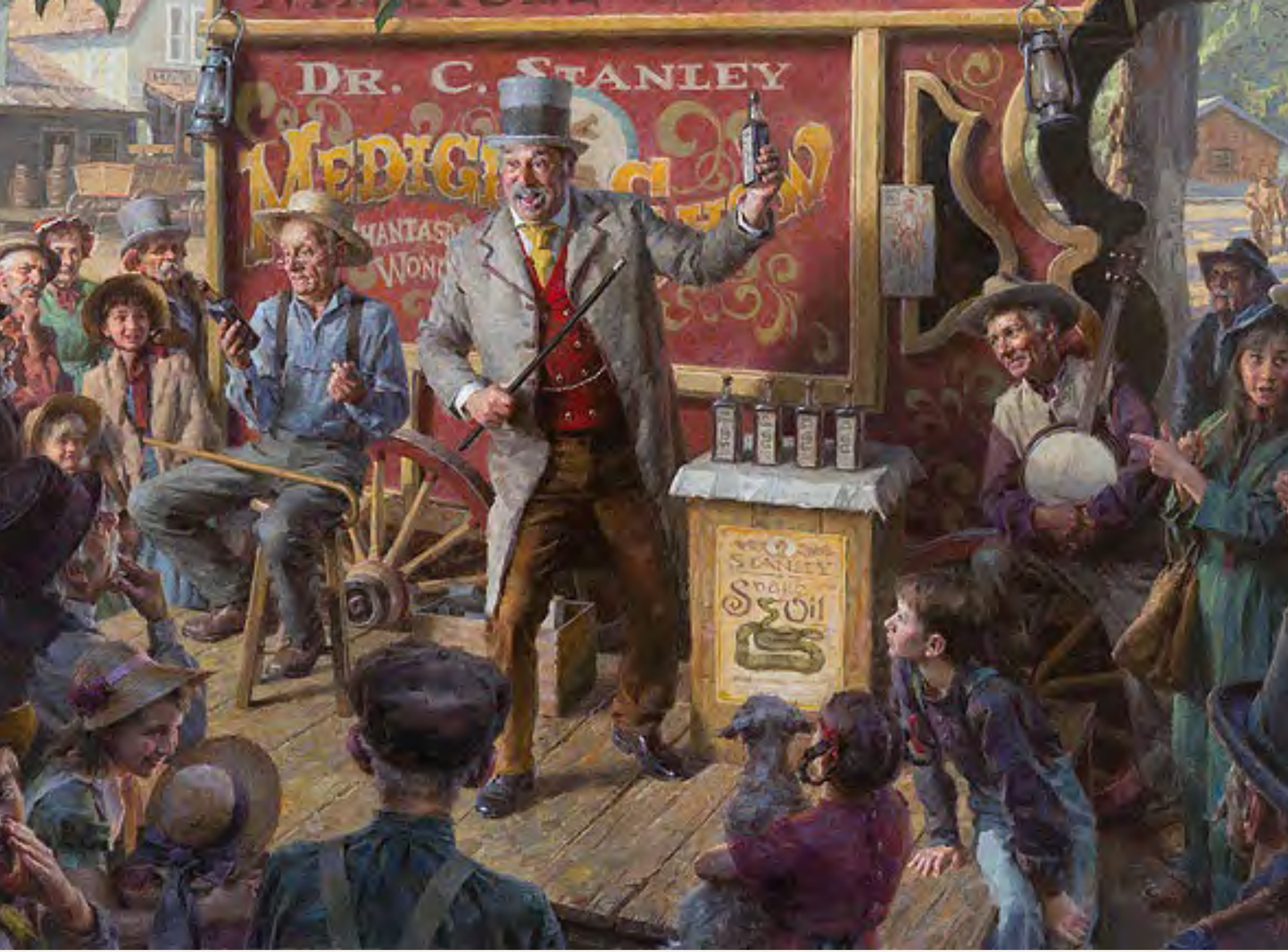
Stats = Poop Detector

Why?



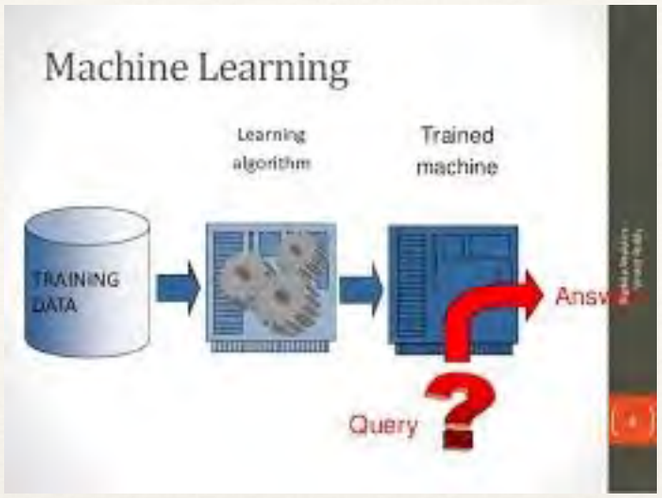
The first principle is that you must not fool yourself - and you are the easiest person to fool.

- Richard Feynman


DR. C. STANLEY
STANLEY
Snake
Oil

What does this have to do with Data Science?

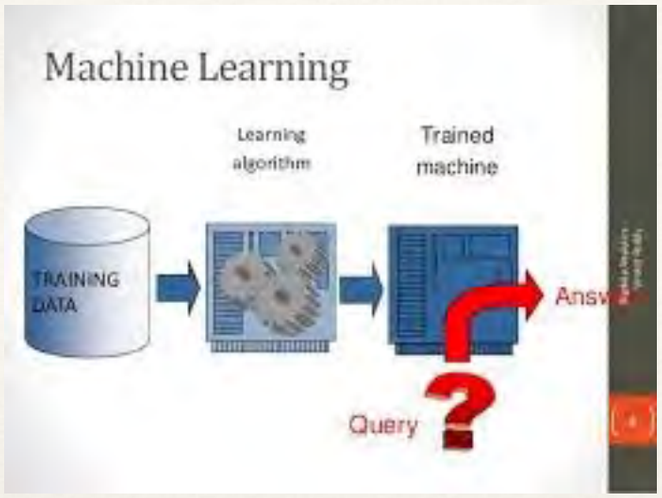
Data



Algorithm

Pattern
Prediction
Insight

Data



Algorithm

Pattern
Prediction
Insight

Or maybe?

How can we be sure?

Can quantify (un)certainty

Unlikely P(X) < 1/36 <2.7%

Can quantify (un)certainty

Really Unlikely $P(X) < 1/1296 < 0.078\%$

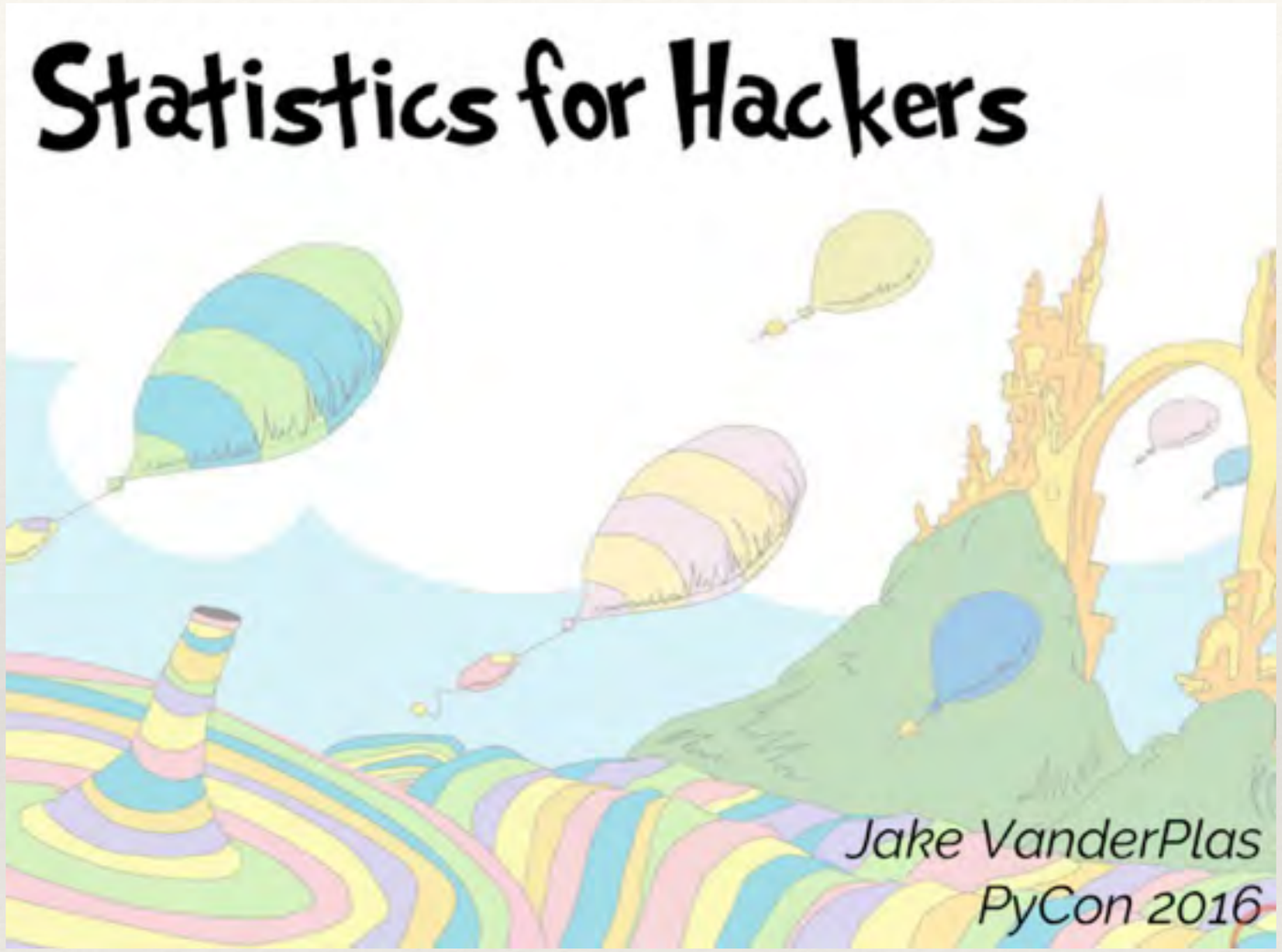
Can quantify (un)certainty

Nearly Impossible $P(X) < 1/46656 < 0.002\%$

Hypothesis Testing

- Happens by pure chance Null Hypothesis
- Probably doesn't happen by pure chance if only 5% chance (why 5%? … or you pick %)

Fair Coin

How often does this happen?

Statistics for hackers