* Calculate the probability of flipping a balanced coin four times and getting each pattern: HTTH, HHHH and TTHH.

P(HTTH) = ½ \* ½ \* ½ \* ½ = 1/16

P(HHHH) = ½ \* ½ \* ½ \* ½ = 1/16

P(TTHH) = ½ \* ½ \* ½ \* ½ = 1/16

* If a list of people has 24 women and 21 men, then the probability of choosing a man from the list is 21/45. What is the probability of not choosing a man?

P(not man) = 24/45

* The probability that Bernice will travel by plane sometime in the next year is 10%. The probability of a plane crash at any time is .005%. What is the probability that Bernice will be in a plane crash sometime in the next year?

P(plane crash) = .10 \* .00005 = .0005 = .05%

* A data scientist wants to study the behavior of users on the company website. Each time a user clicks on a link on the website, there is a 5% chance that the user will be asked to complete a short survey about their behavior on the website. The data scientist uses the survey data to conclude that, on average, users spend 15 minutes surfing the company website before moving on to other things. What is wrong with this conclusion?

They are only getting data from people who filled out the survey, so their data will be skewed. People who don’t fill out the survey may spend more or less time surfing the website

Now it's time to use Bayes' rule to compute some conditional probabilities. First, look over the numbers and estimate each of the four probabilities, using your intuition. Then, calculate the probabilities using Bayes' rule. Keep track of your work in a Google document or markdown file that you can share with your mentor.

A diagnostic test has a 98% probability of giving a positive result when applied to a person suffering from Thripshaw's Disease, and 10% probability of giving a (false) positive when applied to a non-sufferer. It is estimated that 0.5 % of the population are sufferers. Suppose that the test is now administered to a person whose disease status is unknown. Calculate the probability that the test will:

1. Be positive

P(+|d) = 0.98

P(+|h) = 0.10

P(d) = 0.005

P(d | +) = P(+|d) \* P(d) / [P(d)\*P(+|d) + P(d~)\*P(+|d~)]

= 0.98 + 0.005 / (0.005 \* 0.98 + 0.995 \* 0.10)

= 0.0049 / (0.0049 + 0.0995)

= 0.0049 / 0.1044

= 0.0469

P(d|+) = P(+|d) \* P(+) / P(d)

.98 \* .005 + .1 \* .995

(0.005 \* 0.98 + 0.995 \* 0.10) = P(d) \* P(+|d) + P(d~)\*P(+|d~) = 0.1044

P(+) = P(d|+) / P(+|d) \* P(d)

= 0.0469 / 0.98 \* 0.005

= 0.0469 / 0.0049

= 9.57

1. Correctly diagnose a sufferer of Thripshaw's

P(+|d) = 0.98

1. Correctly identify a non-sufferer of Thripshaw's

P(-|h) = 0.90

1. Misclassify the person

P(+|h) P(-|d) = .10\*.995 + .005\*.02

Were your intuitions on the mark, or way off? If your statistical intuition is leading you astray, you aren't alone. According to Nobel-prize winning Daniel Kahneman, [humans simply are not good intuitive statisticians](http://www.burns-stat.com/review-thinking-fast-slow-daniel-kahneman/). That fact has two strong implications for you as you prepare for a career in data science:

1. Just because your statistical intuition is wrong does not mean you're a bad statistician. We're all in this boat. Don't give up on yourself. And,
2. Just because you easily intuit the answer to a statistical question does not mean you're right. Don't trust your intuition. Check your work.