Bayes Theorem

* Know the theorem and how to calculate around it
  + **P(Y|X) = P(X|Y)\*P(Y) / P(X)**
* Given all the number but how do we calculate those numbers
* Given a document, how can we calculate each of the different pieces of bayes theorem
* Joint Probability Distribution Examples
  + P(Cavity v Toothache) = .1
  + P(Cavity) = all toothache + all cavity without overlap = .06+.01+.04
  + P(Cavity|Toothache) = probability of cavity out of people with toothaches = .04/(.04+.01)
* Using Bayes theorem for Classification
  + Identify whether this email is spam or not, want to find which of the two below are higher (spam, or not spam)
  + P(s|e) -> p(s|x\_1, x\_2, x\_2…) ->(P(x1..xd)|s)\*p(s)) / p(x1…xd)
  + P(!s|e) -> p(!s|x\_1, x\_2, x\_2…) -> (p(x1…xd)\*p(!s))/p(x1…xd)
    - x1…xd represents different words
  + Estimating is very very difficult if not impossible. Would need to find an identical email which is basically impossible. Because of this, we want to convert this problem to bayes format.
* Conditional independence
  + z = age
  + y = arm length / reading skills (doesn’t matter)
  + x = arm length / reading skills (doesn’t matter)