Modeling

Chris Piech
CS109, Stanford University

Midterm **Tuesday July 23rd**, 7pm

Where to Go

Building 300, room 300. Building 300 is part of main quad (the older buildings in the center of campus).

5 pages front and back of notes, closed book, closed calculator

Review

Last Week: Joint Distributions

Joint Distribution *noun*

The probability of a simultaneous assignment to ***all*** the random variables in a probabilistic model.

Eg:

$$P(X = x, Y = y)$$

$$f(X = x, Y = y)$$

$$P(X = x, Y = y, \cdots, Z = z)$$

Monday, Wednesday: Inference

Inference *noun*

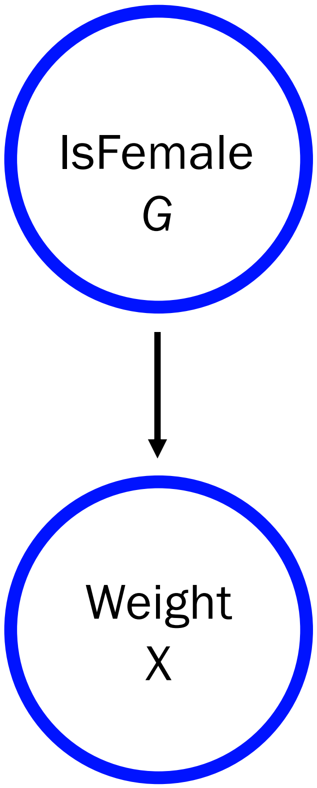
An updated belief about a random variable (or multiple) based on conditional knowledge regarding another random variable (or multiple) in a probabilistic model.

TLDR: conditional probability with random variables.

Inference with Continuous

Q: At birth, girl elephant weights are distributed as a Gaussian with mean = 160kg, std = 7kg. At birth, boy elephant weights are distributed as a Gaussian with mean = 165kg, std = 3kg. All you know about a newborn elephant is that it is 163kg. What is the probability that it is a girl?

Model Shown Graphically



G = 1 is Bern(p = 0.5)

$X \mid G = 1$ is $N(\mu = 160, \sigma^2 = 7^2)$

$X \mid G = 0$ is $N(\mu = 165, \sigma^2 = 3^2)$

Q: What is $P(G = 1 \mid X = 163)$

Harder when inferred random variable is not a Bernoulli

Learning Goals

1. Perspective on the artform of how to design probabilistic models
2. How to calculate Correlations
3. Use and verify Independence with Random Variables

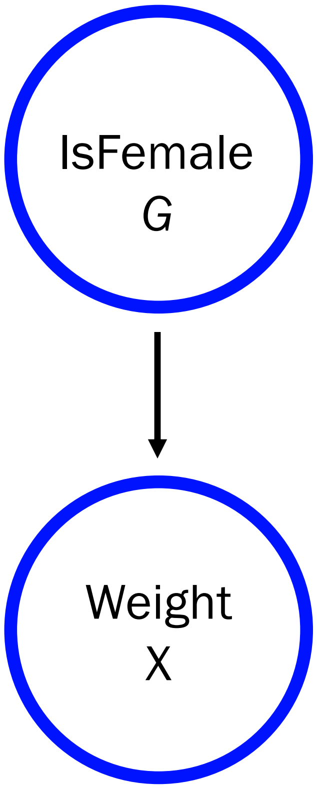
Lets talk about how to make a model

Model Version #1: Python That Outputs a **Joint** Sample

Sample Baby Elephant

Sex: Female
Weight: 161kg

Model Version #2: Bayesian Network



G = 1 is Bern(p = 0.5)

$X \mid G = 1$ is $N(\mu = 160, \sigma^2 = 7^2)$
$X \mid G = 0$ is $N(\mu = 165, \sigma^2 = 3^2)$

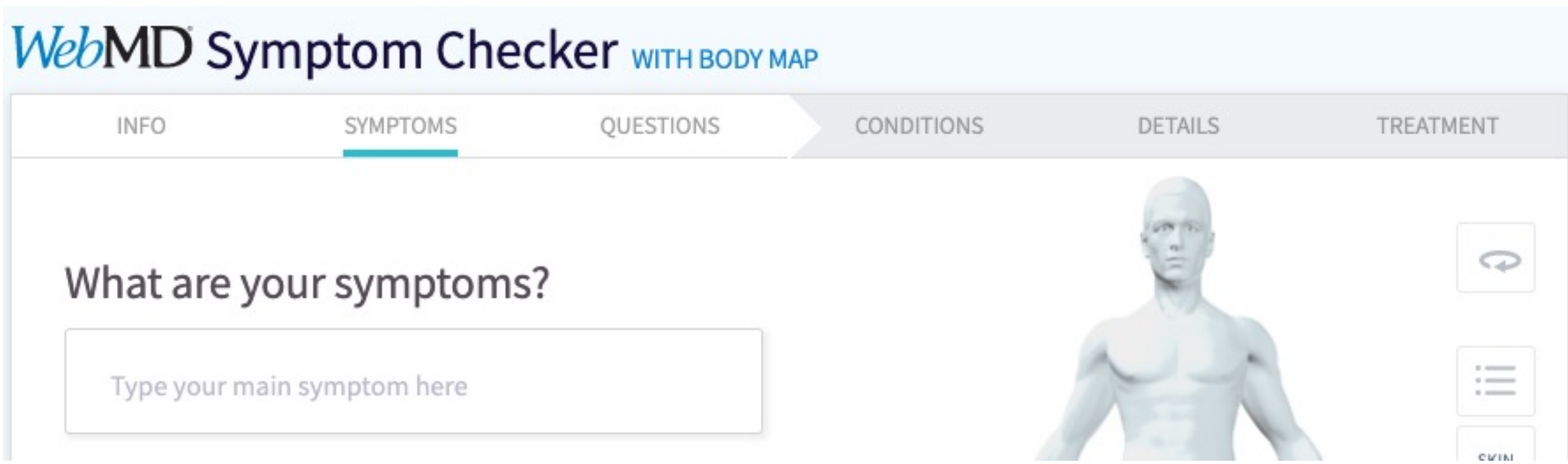
Does this define the joint?

$$f(G = g, X = x)$$
$$= f(X = x | G = g) P(G = g)$$

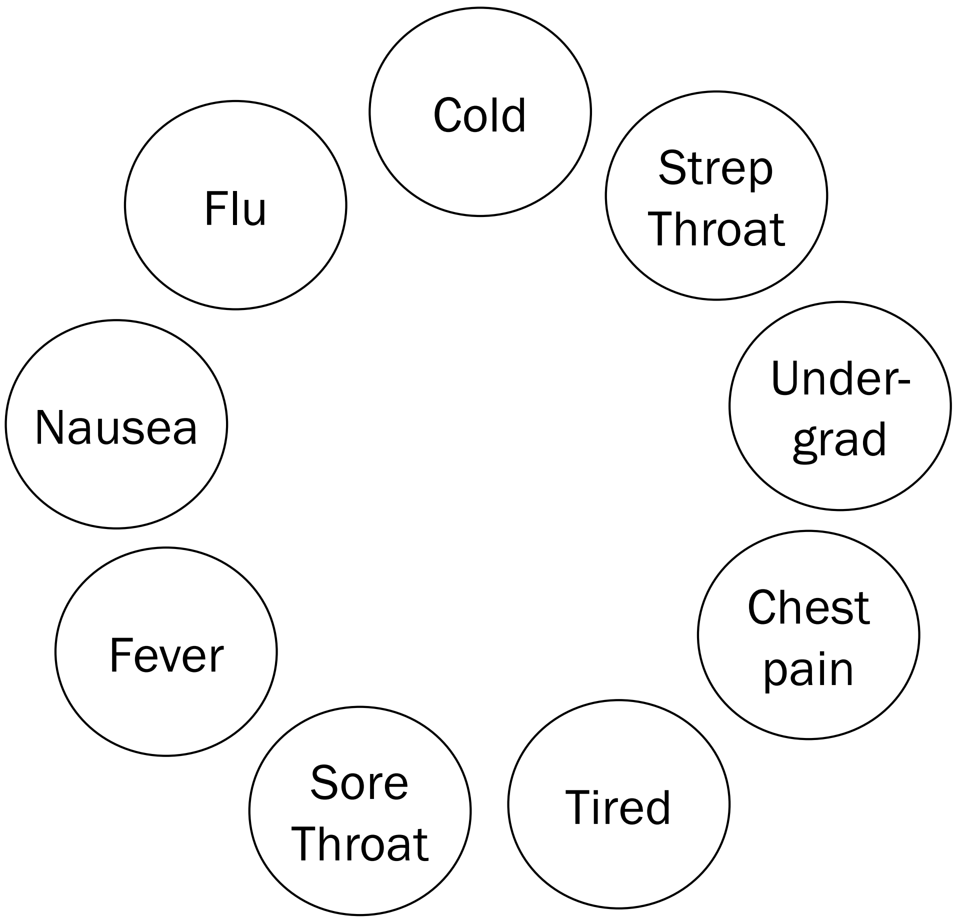
Why You Need a Model

WebMD®

Inference


WebMD Symptom Checker WITH BODY MAP
INFO
SYMPTOMS
QUESTIONS
CONDITIONS
DETAILS
TREATMENT
What are your symptoms?
Type your main symptom here

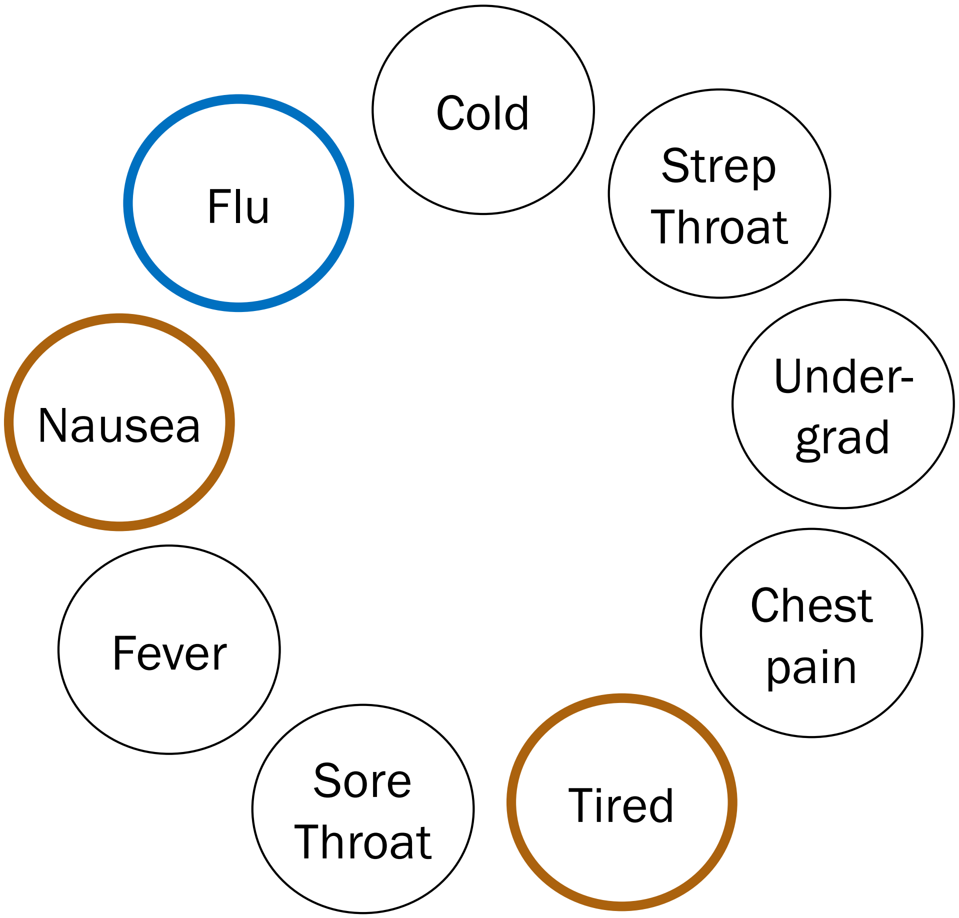
Inference



Inference question:

Given the values of some random variables, what are the conditional distributions of some other random variables?

Inference

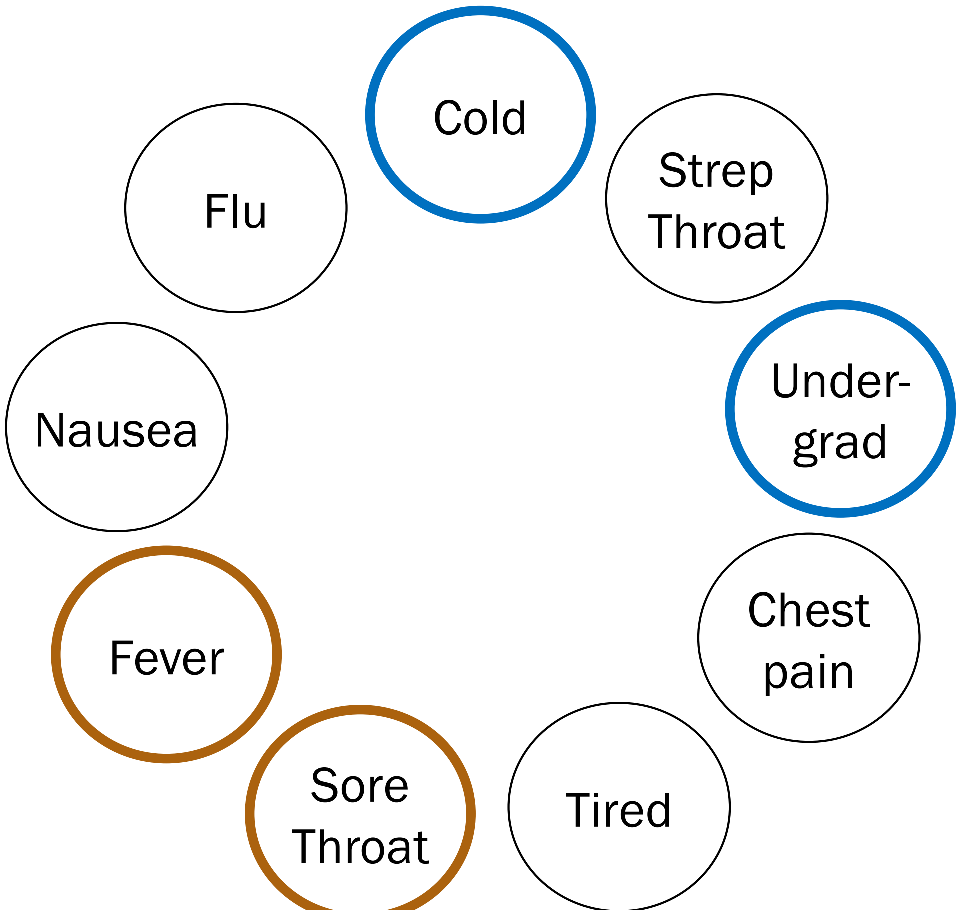


One inference question:

$$P(F = 1|N = 1, T = 1)$$

$$= \frac{P(F = 1, N = 1, T = 1)}{P(N = 1, T = 1)}$$

Inference

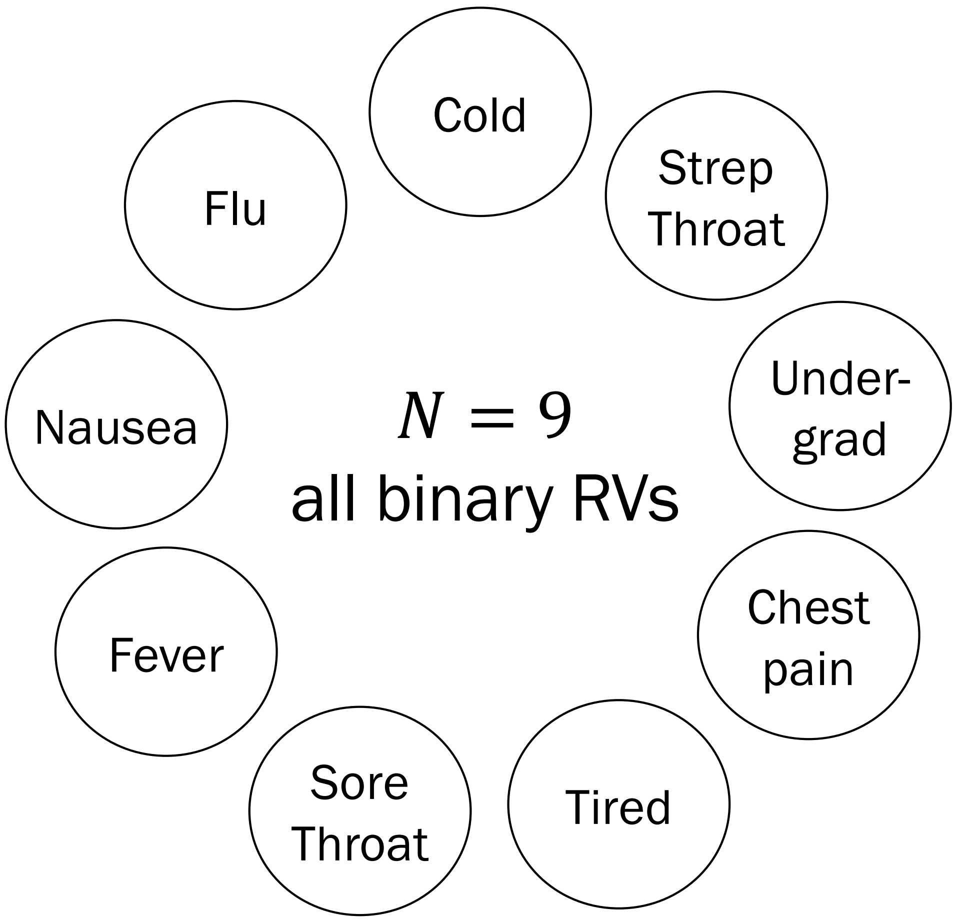


Another inference question:

$$P(C_o = 1, U = 1 | S = 0, F_e = 0)$$

$$= \frac{P(C_o = 1, U = 1, S = 0, F_e = 0)}{P(S = 0, F_e = 0)}$$

Inference

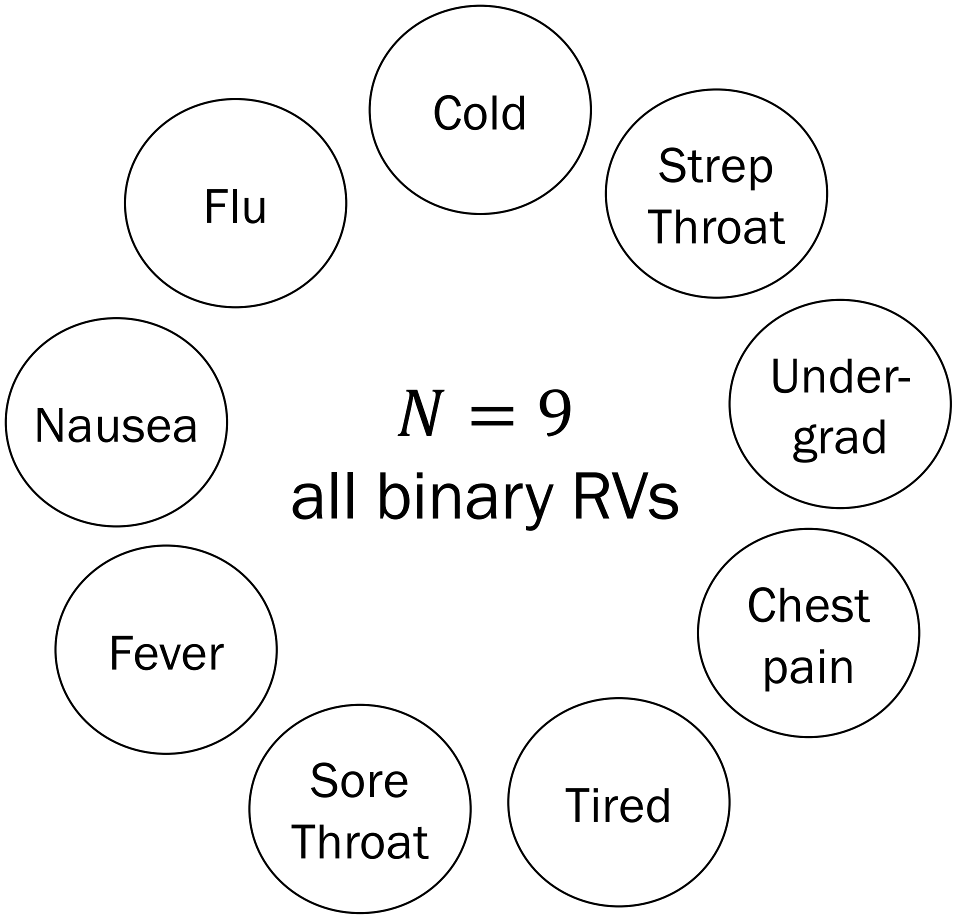


If we knew the **joint distribution**, we can answer all probabilistic inference questions.

What is the size of the joint probability table?

A. 2^{N-1} entries
B. N^2 entries
C. 2^N entries
D. None/other/don't know

Inference



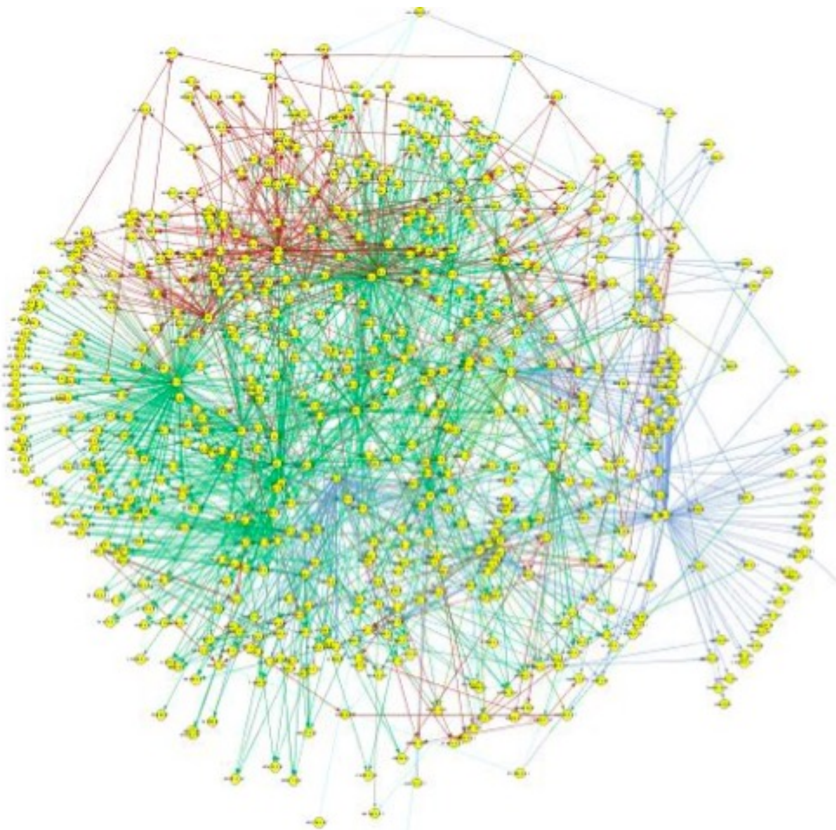
If we knew the **joint distribution**, we can answer all probabilistic inference questions.

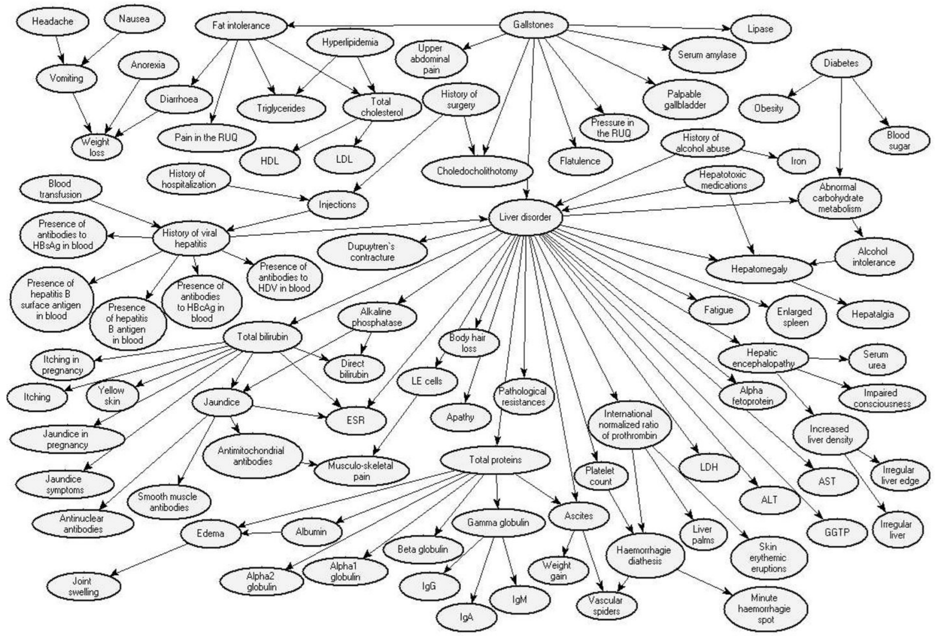
What is the size of the joint probability table?

A. 2^{N-1} entries
B. N^2 entries
C. 2^N entries
D. None/other/don't know

Naively specifying a joint distribution is, in general, intractable.

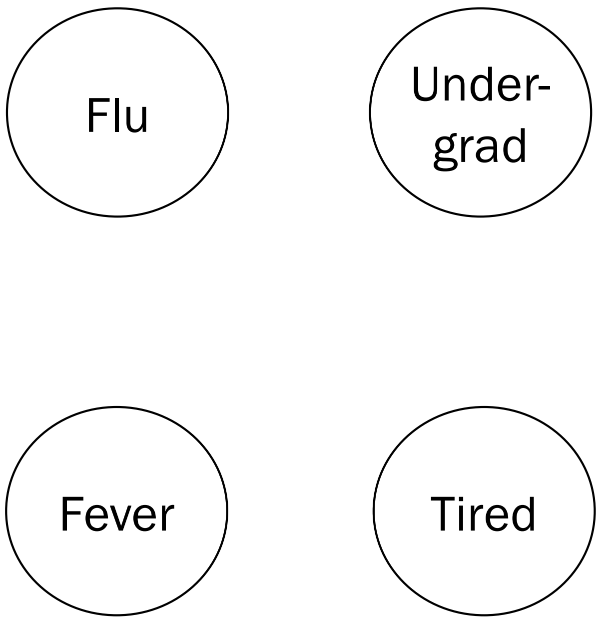
N can be large…





Bayesian Networks

A simpler WebMD



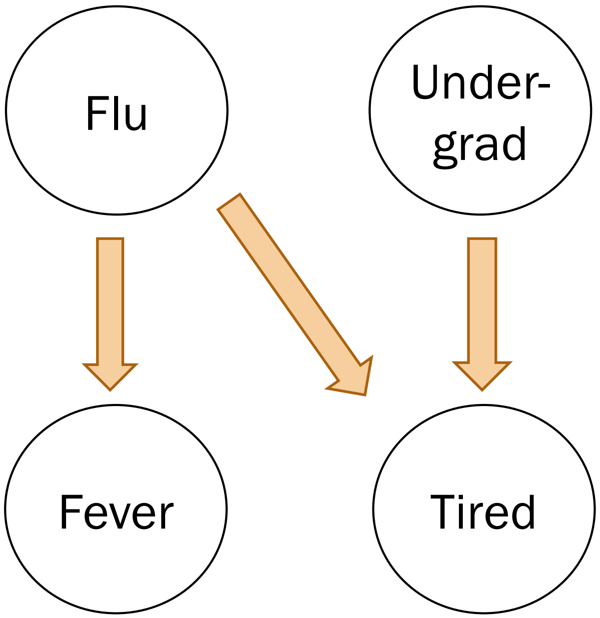
Great! Just specify $2^4 = 16$ joint probabilities…?

$$P(F_{lu} = a, F_{ev} = b, U = c, T = d)$$

What would a Stanford flu expert do?

Describe the joint distribution using causality!

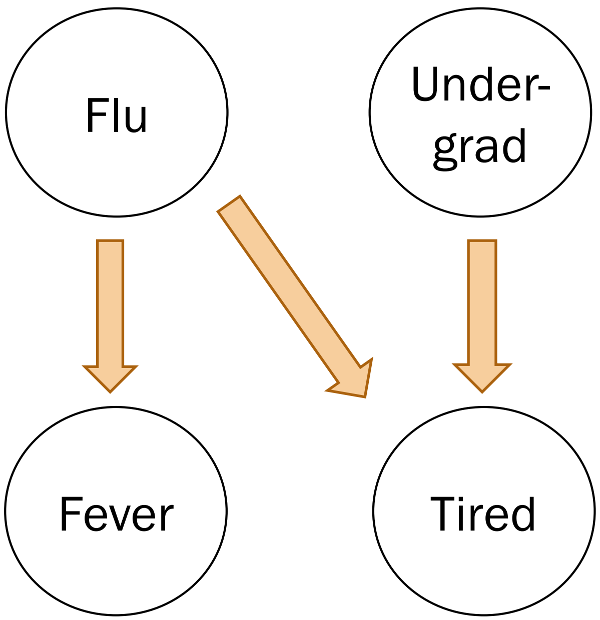
Constructing a Bayesian Network



What would a Stanford flu expert do?

✅ 1. Describe the joint distribution using causality.
2. Provide P(values|parents) for each random variable
3. Implicitly assumes independences.

Constructing a Bayesian Network



$P(T = 1|F_{lu} = 0, U = 0)$
$P(T = 1|F_{lu} = 0, U = 1)$
$P(T = 1|F_{lu} = 1, U = 0)$
$P(T = 1|F_{lu} = 1, U = 1)$

In a Bayesian Network,

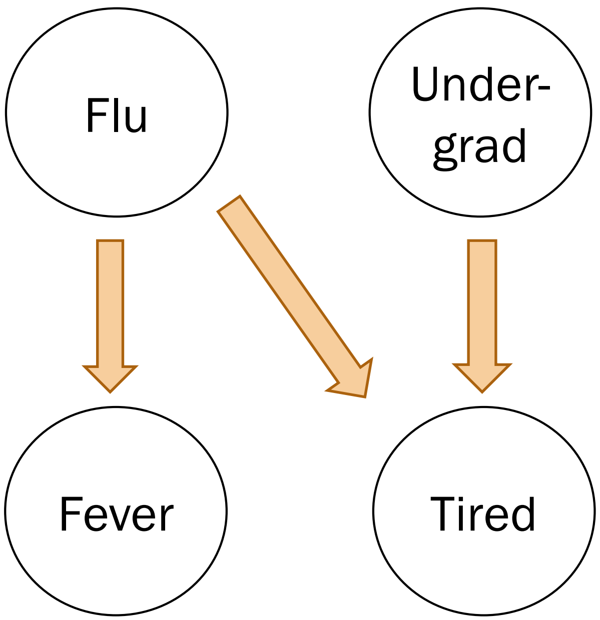
Each random variable is caused by its parents. Def P(node | parents)

- Node: random variable
- Directed edge: causality

Examples:

- $P(F_{lu} = 1)$
- $P(U = 0)$
- $P(F_{ev} = 1|F_{lu} = 1)$, $P(F_{ev} = 1|F_{lu} = 0)$
- $P(T = 1|F_{lu} = 0, U = 0)$ …

Constructing a Bayesian Network



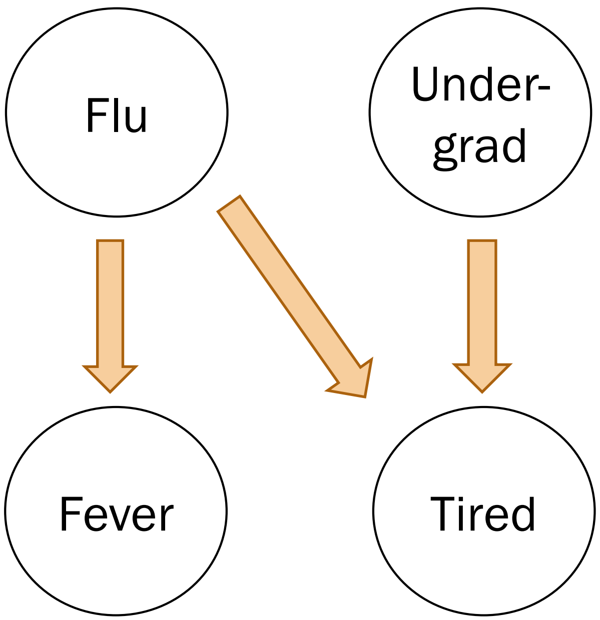
What would a Stanford flu expert do?

✅ 1. Describe the joint distribution using causality.

✅ 2. Provide $P(\text{values}|\text{parents})$ for each random variable

✅ 3. Implicitly assumes independences.

Constructing a Bayesian Network

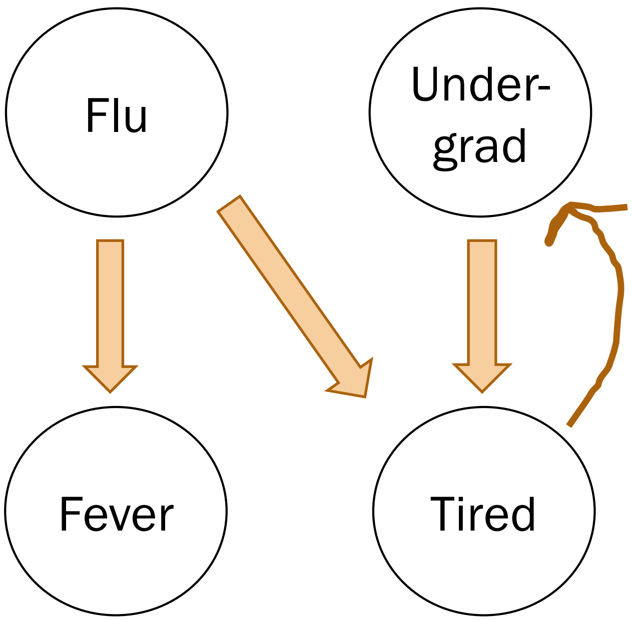


This model assumes that Flu and being an Undergraduate are independent.

Neat trick: it also assumes that fever and tired are conditionally independent given Flu.

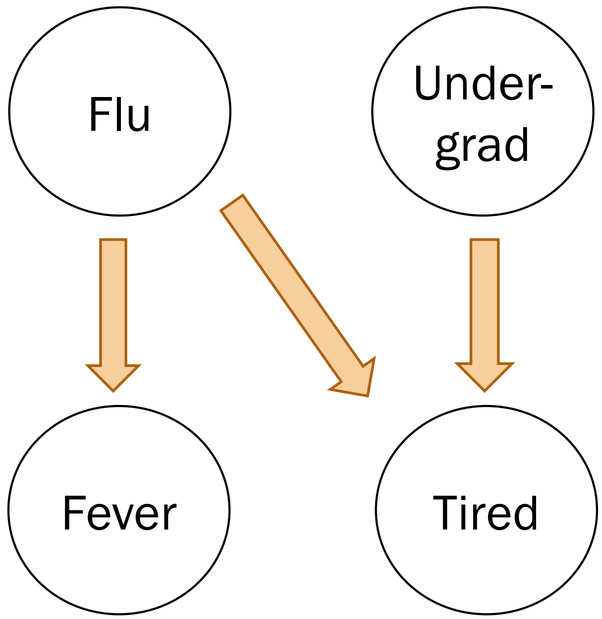
You need to tell a generative story. You do not need to be able to reason about all the implied independencies

Bug: Constructing a Bayesian Network



Must by acyclic!

Bayesian Network Assumption:



$$P(\text{Joint}) = P(X_1 = x_1 \ldots X_n = x_n)$$
$$= \prod_i P(X_i = x_i | \text{parents of } X_i)$$

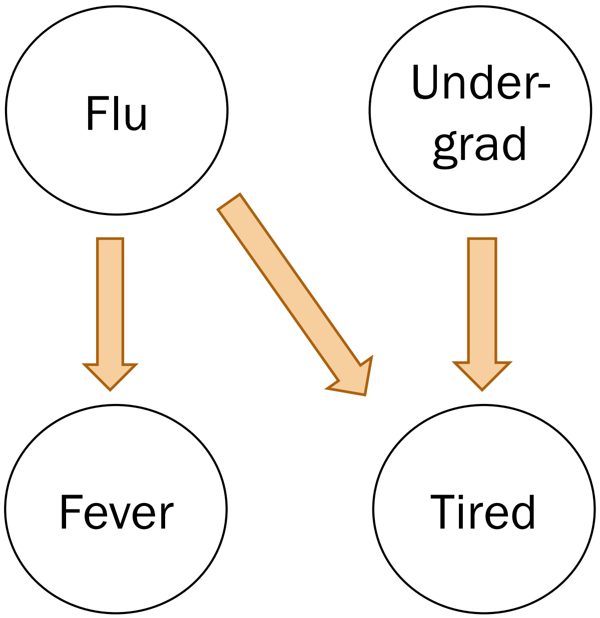
Bayes Nets tell a generative story.

This leads to many independence assumptions

Makes it **tractable** to represent the joint

Inference via math

$P(F_{lu} = 1) = 0.1$ $\quad P(U = 1) = 0.8$



$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

1. $P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1)$?

Compute joint probabilities using chain rule.

Independence of RVs

Independent discrete RVs

Recall the definition of independent events E and F:

$$P(EF) = P(E)P(F)$$

Two discrete random variables X and Y are **independent** if:

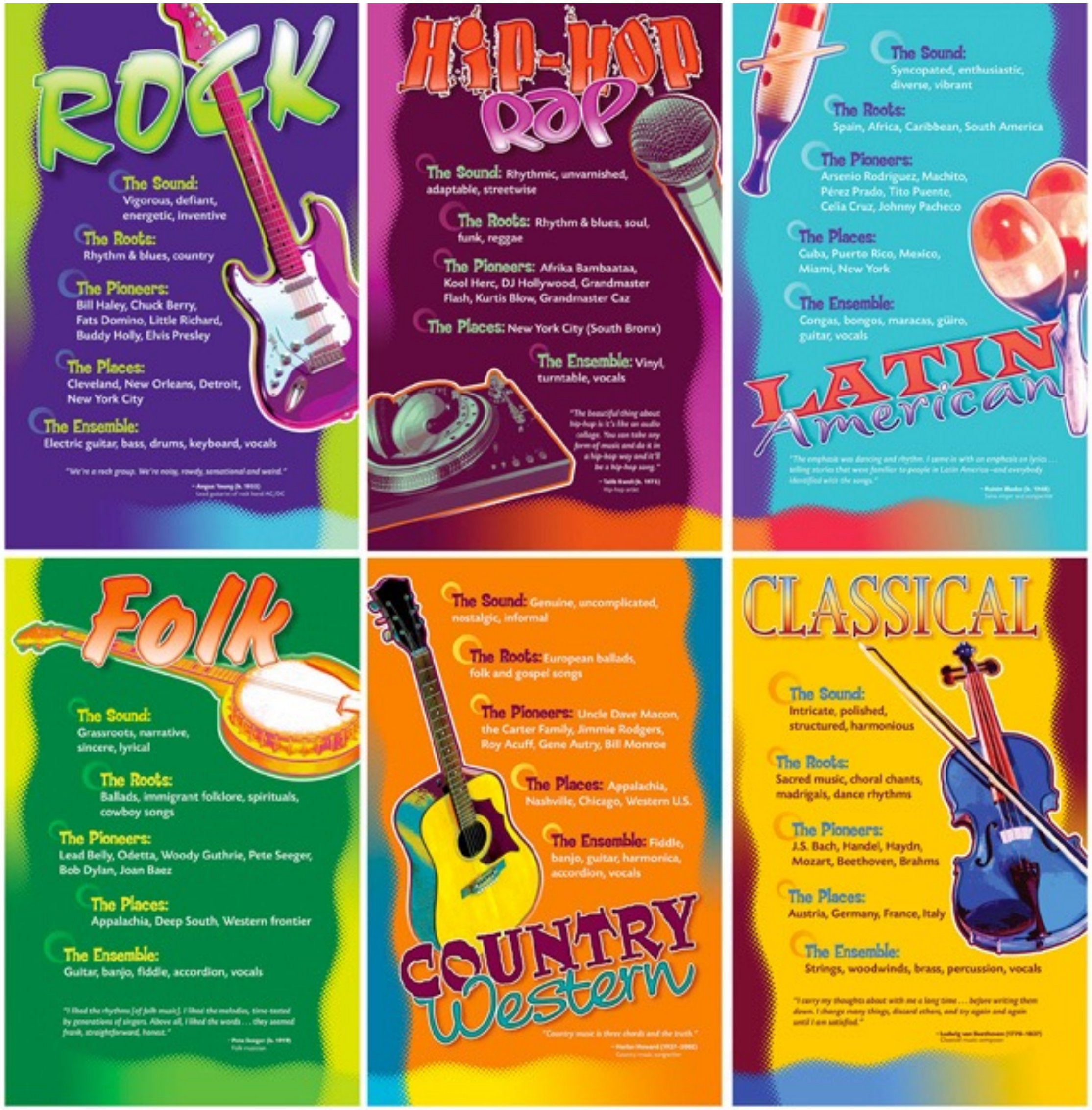
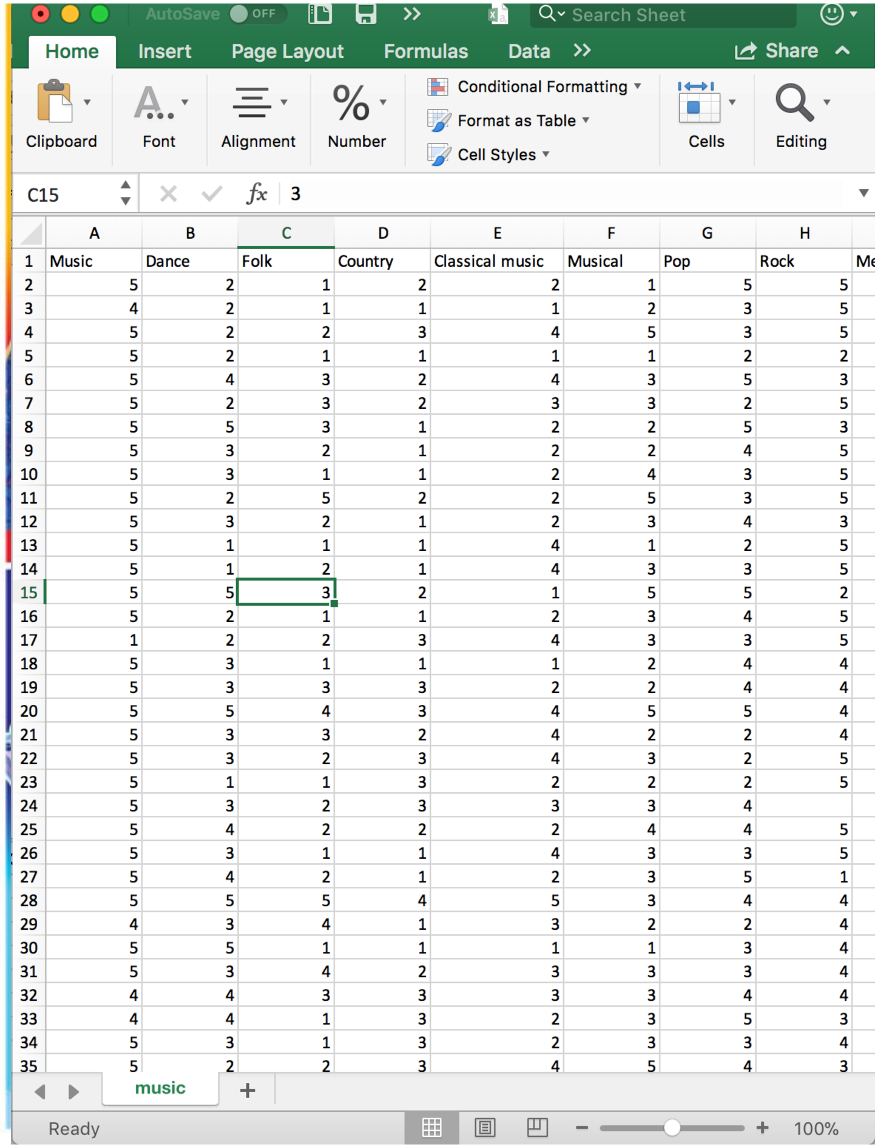
for all x, y:

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

Different notation, same idea:

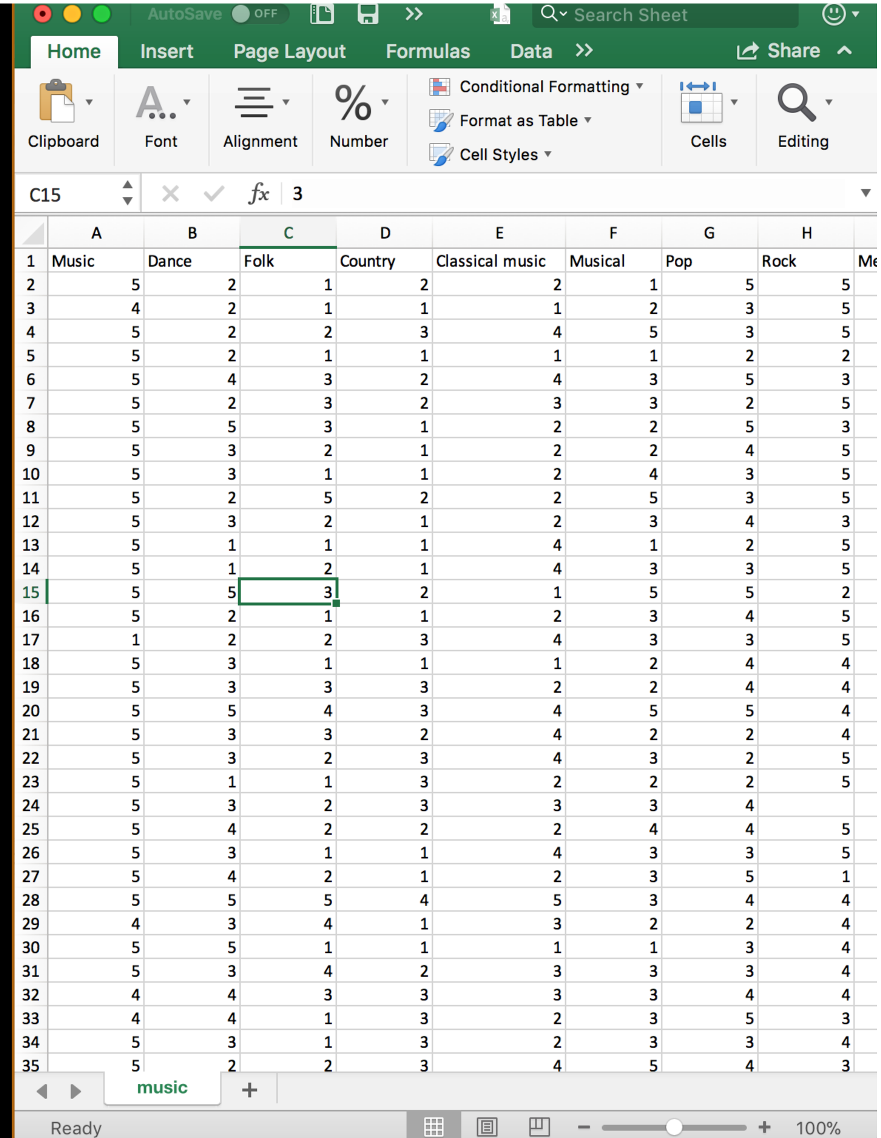
$$p(x, y) = p(x)p(y)$$

- Intuitively: knowing value of X tells us nothing about the distribution of Y (and vice versa)
- If two variables are not independent, they are called **dependent**.

Can I discover independence from data?



	A	B	C	D	E	F	G	H
1	Music	Dance	Folk	Country	Classical music	Musical	Pop	Rock
2	5	2	1	2	2	1	5	5
3	4	2	1	1	1	2	3	5
4	5	2	2	3	4	5	3	5
5	5	2	1	1	1	1	2	2
6	5	4	3	2	4	3	5	3
7	5	2	3	2	3	3	2	5
8	5	5	3	1	2	2	5	3
9	5	3	2	1	2	2	4	5
10	5	3	1	1	2	4	3	5
11	5	2	5	2	2	5	3	5
12	5	3	2	1	2	3	4	3
13	5	1	1	1	4	1	2	5
14	5	1	2	1	4	3	3	5
15	5	5	3	2	1	5	5	2
16	5	2	1	1	2	3	4	5
17	1	2	2	3	4	3	3	5
18	5	3	1	1	1	2	4	4
19	5	3	3	3	2	2	4	4
20	5	5	4	3	4	5	5	4
21	5	3	3	2	4	2	2	4
22	5	3	2	3	4	3	2	5
23	5	1	1	3	2	2	2	5
24	5	3	2	3	3	3	4	
25	5	4	2	2	2	4	4	5
26	5	3	1	1	4	3	3	5
27	5	4	2	1	2	3	5	1
28	5	5	5	4	5	3	4	4
29	4	3	4	1	3	2	2	4
30	5	5	1	1	1	1	3	4
31	5	3	4	2	3	3	3	4
32	4	4	3	3	3	3	4	4
33	4	4	1	3	2	3	5	3
34	5	3	1	3	2	3	3	4
35	5	2	2	3	4	5	4	3

Why is it harder to find independences here than for bat DNA expression?



	A	B	C	D	E	F	G	H
1	Music	Dance	Folk	Country	Classical music	Musical	Pop	Rock
2	5	2	1	2	2	1	5	5
3	4	2	1	1	1	2	3	5
4	5	2	2	3	4	5	3	5
5	5	2	1	1	1	1	2	2
6	5	4	3	2	4	3	5	3
7	5	2	3	2	3	3	2	5
8	5	5	3	1	2	2	5	3
9	5	3	2	1	2	2	4	5
10	5	3	1	1	2	4	3	5
11	5	2	5	2	2	5	3	5
12	5	3	2	1	2	3	4	3
13	5	1	1	1	4	1	2	5
14	5	1	2	1	4	3	3	5
15	5	5	3	2	1	5	5	2
16	5	2	1	1	2	3	4	5
17	1	2	2	3	4	3	3	5
18	5	3	1	1	1	2	4	4
19	5	3	3	3	2	2	4	4
20	5	5	4	3	4	5	5	4
21	5	3	3	2	4	2	2	4
22	5	3	2	3	4	3	2	5
23	5	1	1	3	2	2	2	5
24	5	3	2	3	3	3	4	
25	5	4	2	2	2	4	4	5
26	5	3	1	1	4	3	3	5
27	5	4	2	1	2	3	5	1
28	5	5	5	4	5	3	4	4
29	4	3	4	1	3	2	2	4
30	5	5	1	1	1	1	3	4
31	5	3	4	2	3	3	3	4
32	4	4	3	3	3	3	4	4
33	4	4	1	3	2	3	5	3
34	5	3	1	3	2	3	3	4
35	5	2	2	3	4	5	4	3

Dance of the Covariance

Recall our Ebola Bats

niversity

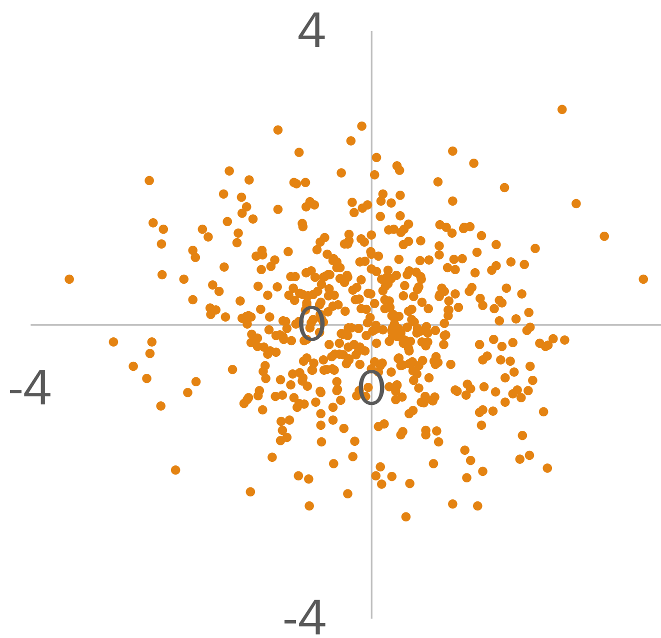
Bat Data

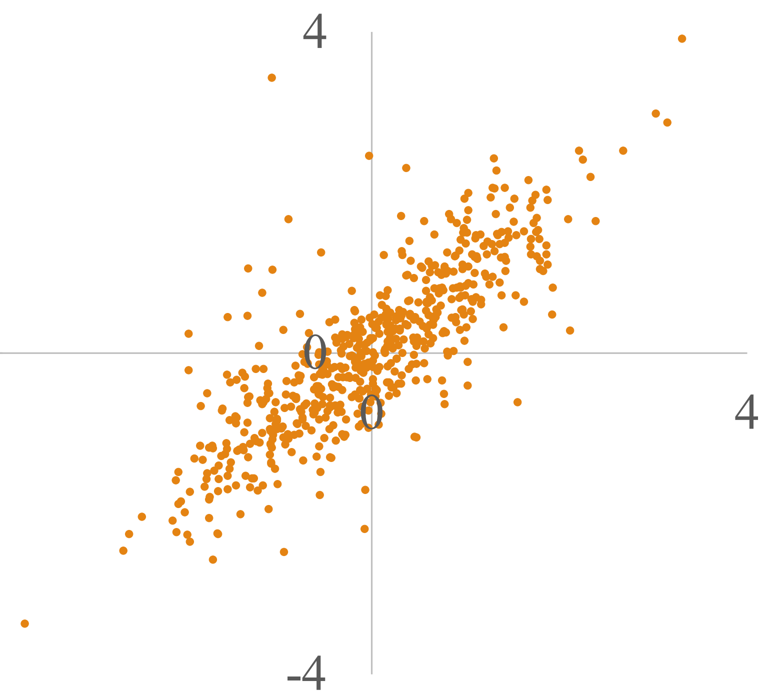
Gene1	Gene2	Gene3	Gene4	Gene5	Trait
TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
FALSE	FALSE	TRUE	TRUE	TRUE	TRUE
TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
FALSE	TRUE	FALSE	TRUE	FALSE	FALSE
TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
TRUE	FALSE	TRUE	TRUE	TRUE	FALSE
FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE
			...		
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE

Expression Amount

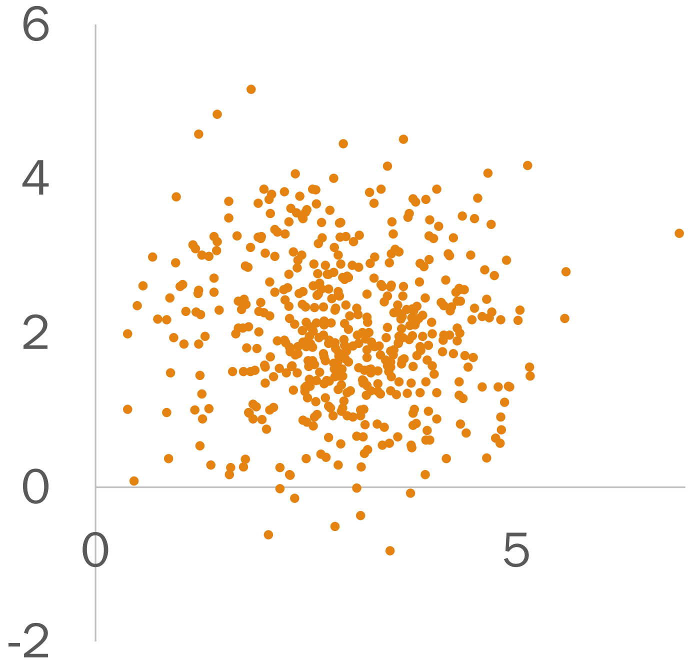
Gene5	Trait
0.76	0.83
0.94	0.85
0.82	0.03
0.94	0.32
0.50	0.10
0.40	0.53
0.90	0.67
0.29	0.71
0.72	0.25
0.15	0.24
0.79	0.98
0.68	0.77
0.71	0.37
0.36	0.18
0.62	0.08
0.59	0.38
0.82	0.76

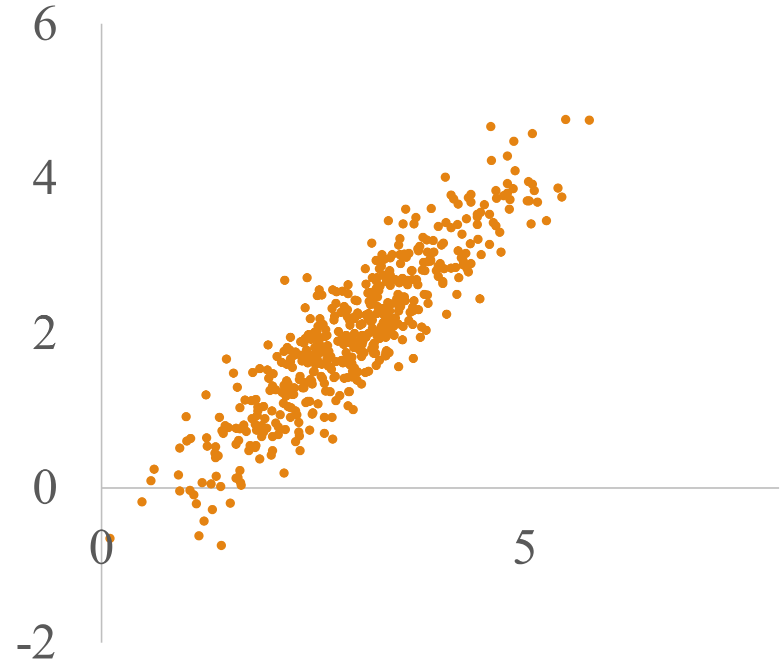
Spot The Difference



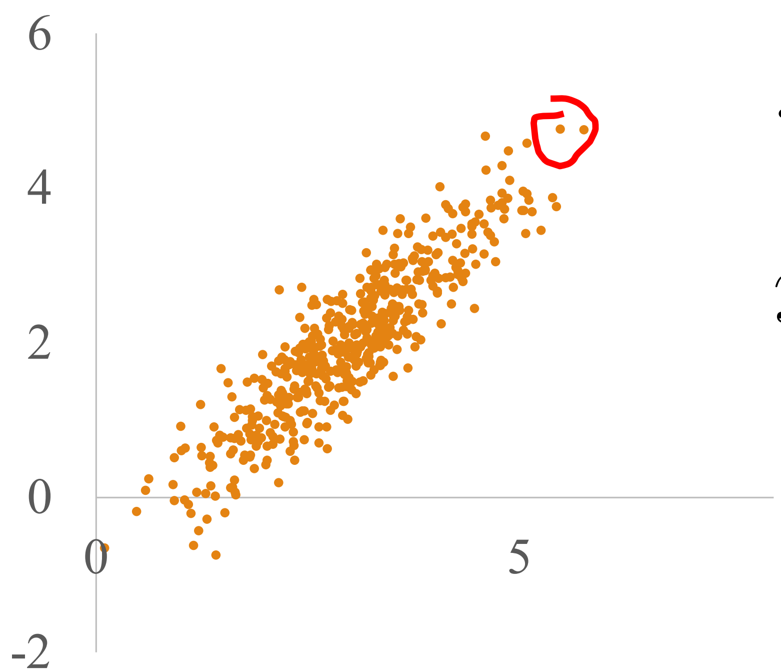


Spot The Difference





Vary Together

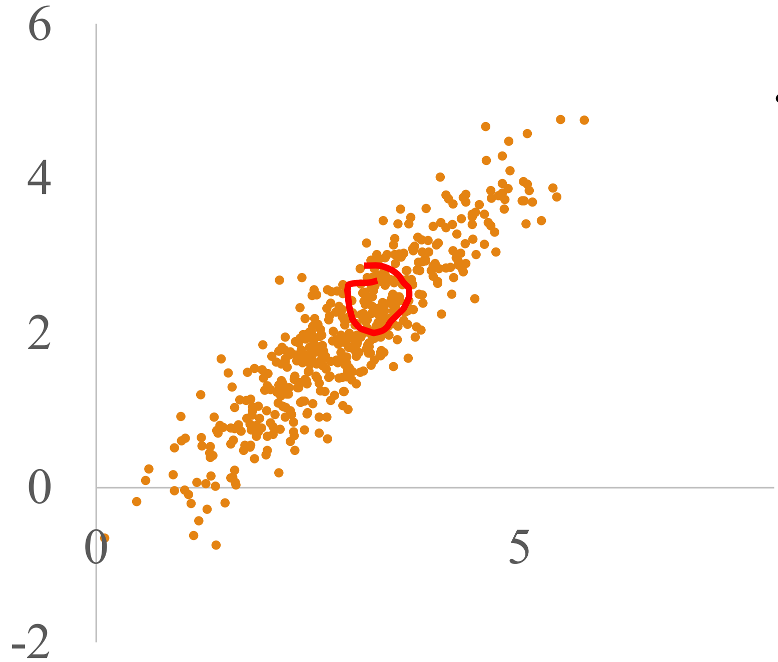


$$x - E[x] = 3$$

$$y - E[y] = 2.6$$

$$(x - E[x])(y - E[y]) = 7.8$$

Vary Together

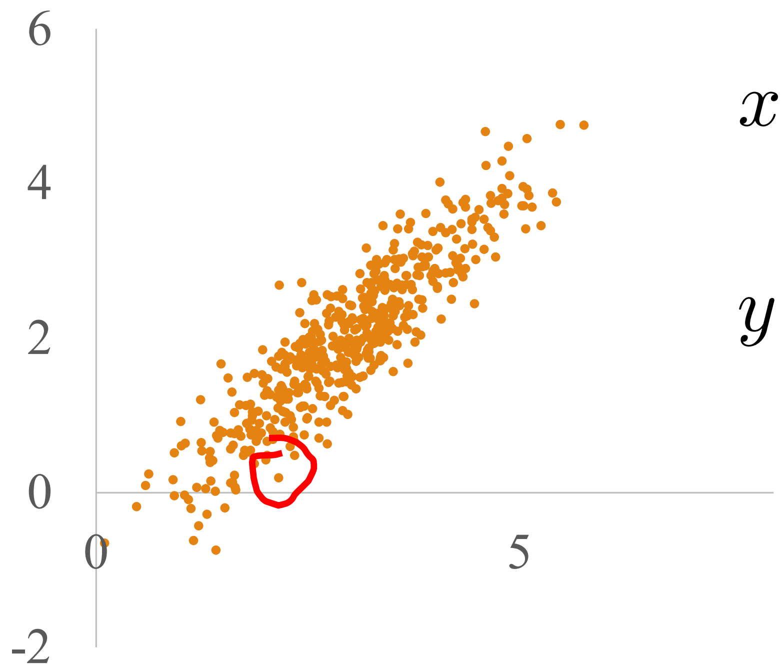


$$x - E[x] \approx 0$$

$$y - E[y] \approx 0$$

$$(x - E[x])(y - E[y]) = 0$$

Vary Together

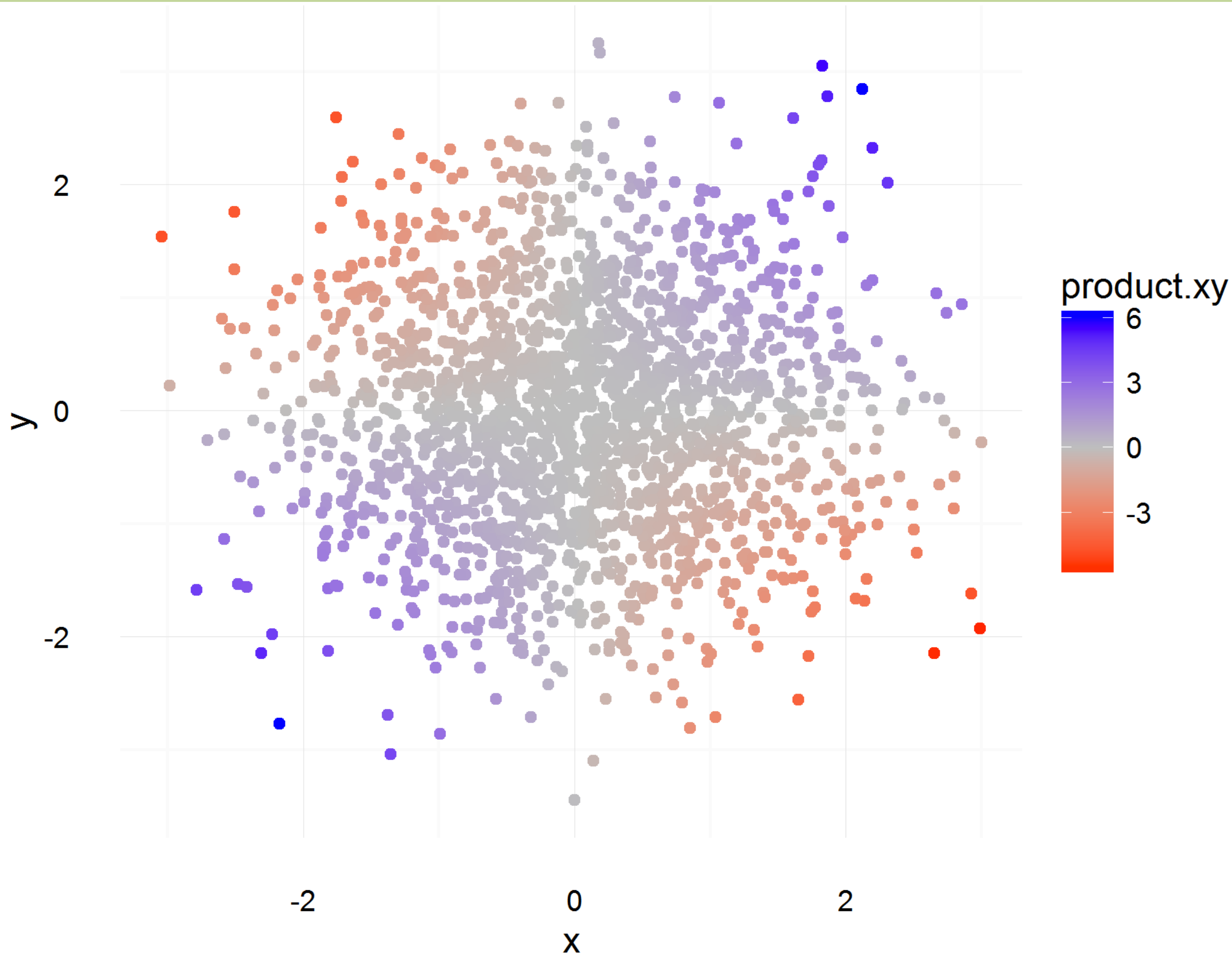


$$x - E[x] = -1.1$$

$$y - E[y] = -2.8$$

$$(x - E[x])(y - E[y]) \approx 3.1$$

Understanding Covariance



nford University

The Dance of the Covariance

Say X and Y are arbitrary random variables

Covariance of X and Y:

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

x	y	(x – E[X])(y – E[Y])p(x,y)
Above mean	Above mean	Positive
Bellow mean	Bellow mean	Positive
Bellow mean	Above mean	Negative
Above mean	Bellow mean	Negative

The Dance of the Covariance

Say X and Y are arbitrary random variables

Covariance of X and Y:

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

Equivalently:

$$\begin{aligned}\text{Cov}(X,Y) &= E[XY - E[X]Y - XE[Y] + E[Y]E[X]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

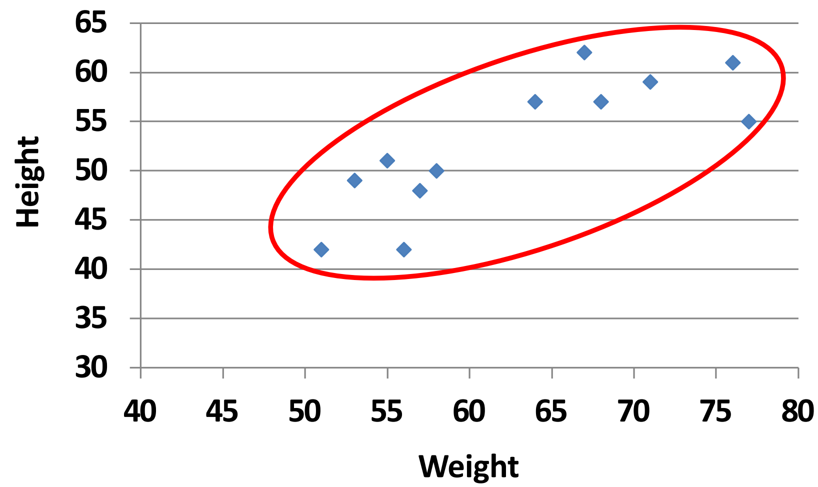
- X and Y independent, E[XY] = E[X]E[Y] → Cov(X,Y) = 0
- But Cov(X,Y) = 0 does **not** imply X and Y independent!

Covariance and Data

Consider the following data:

Weight	Height	Weight * Height
64	57	3648
71	59	4189
53	49	2597
67	62	4154
55	51	2805
58	50	2900
77	55	4235
57	48	2736
56	42	2352
51	42	2142
76	61	4636
68	57	3876

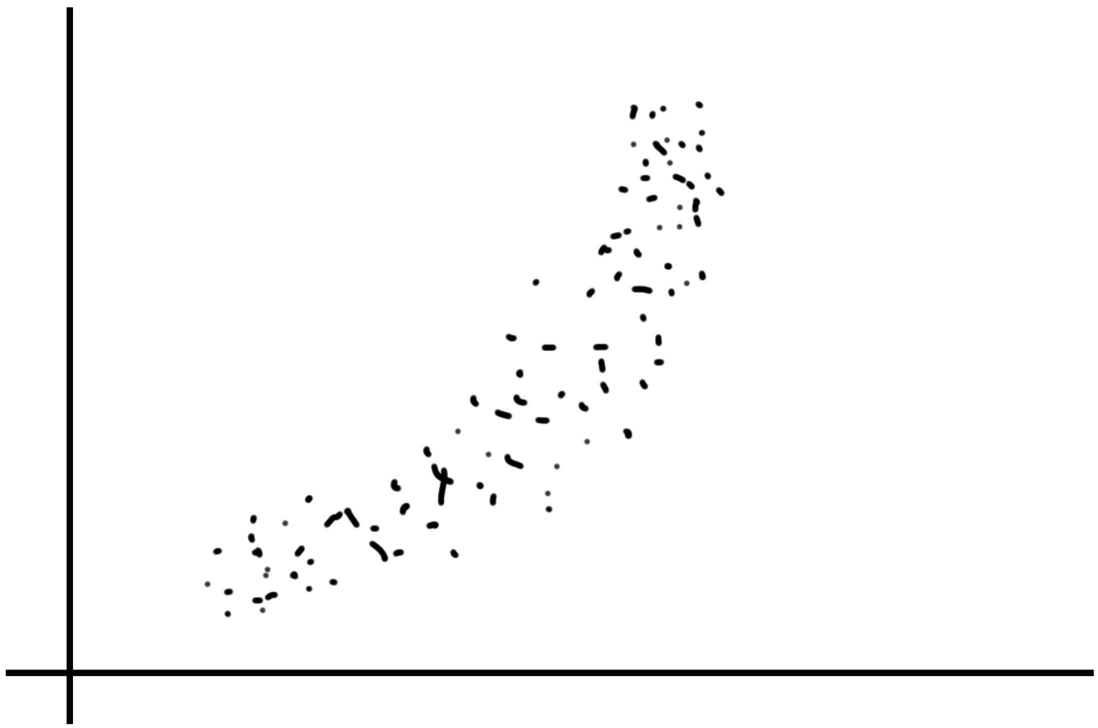
E[W] = 62.75 **E[H] = 52.75** **E[W*H] = 3355.83**



$$\text{Cov}(W, H) = E[W*H] - E[W]E[H]$$
$$= 3355.83 - (62.75)(52.75)$$
$$= 45.77$$

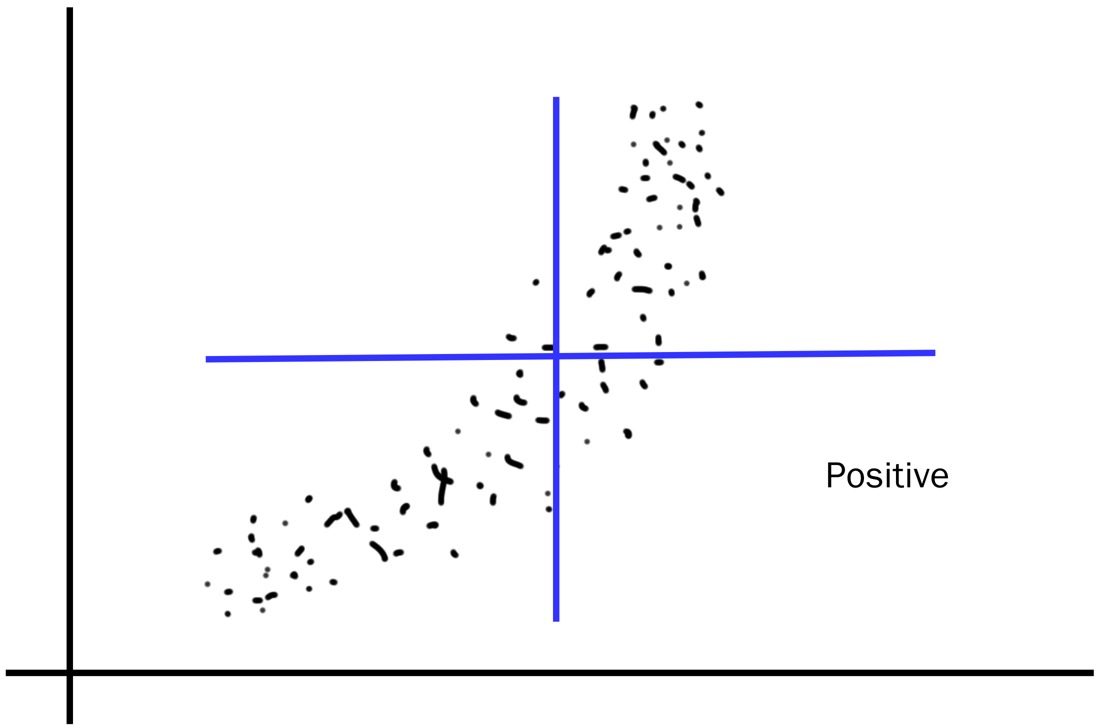
Covariance

Poll: (a) positive, (b) negative, (c) zero



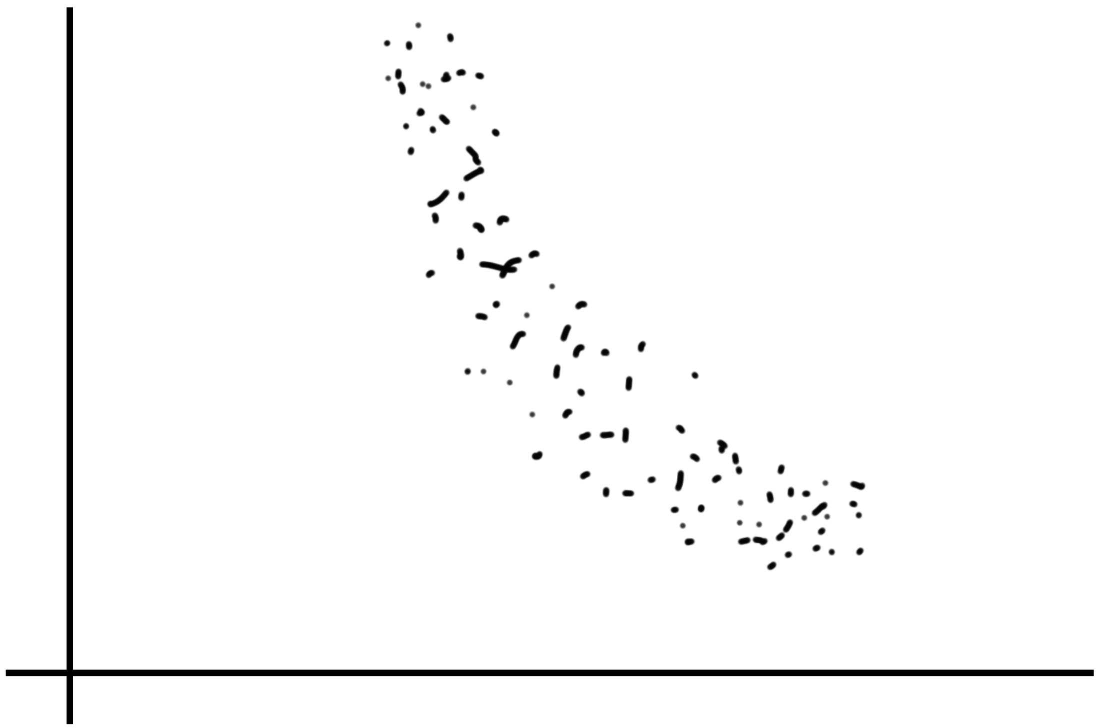
Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero



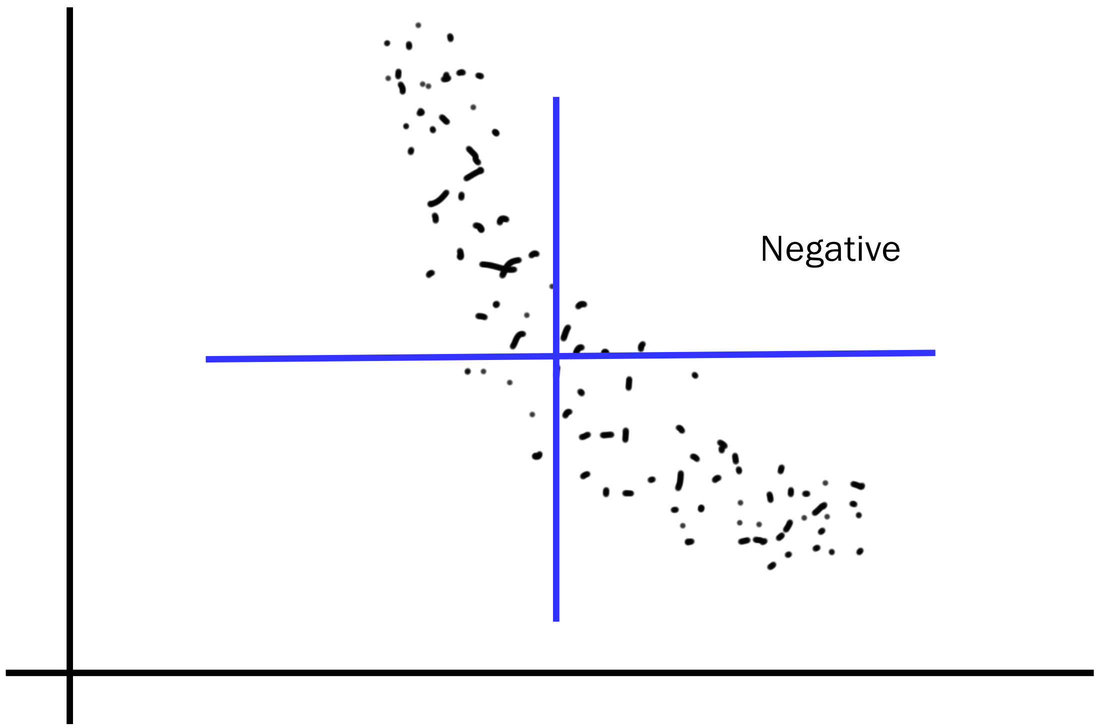
Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero



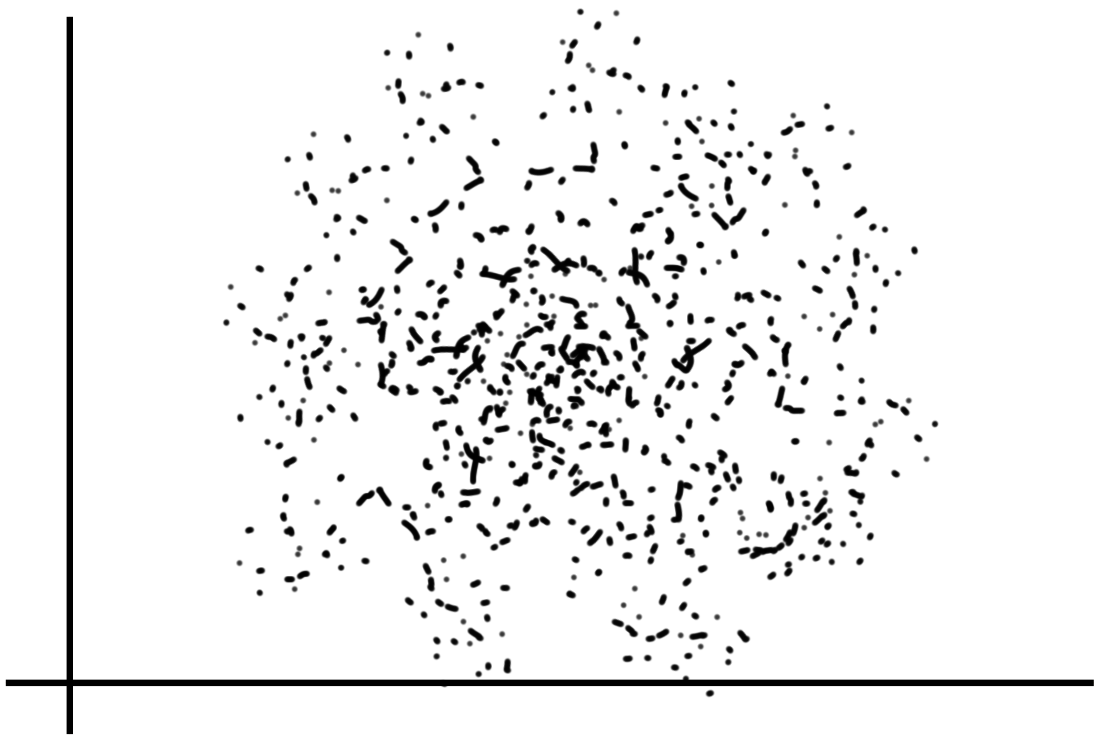
Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero



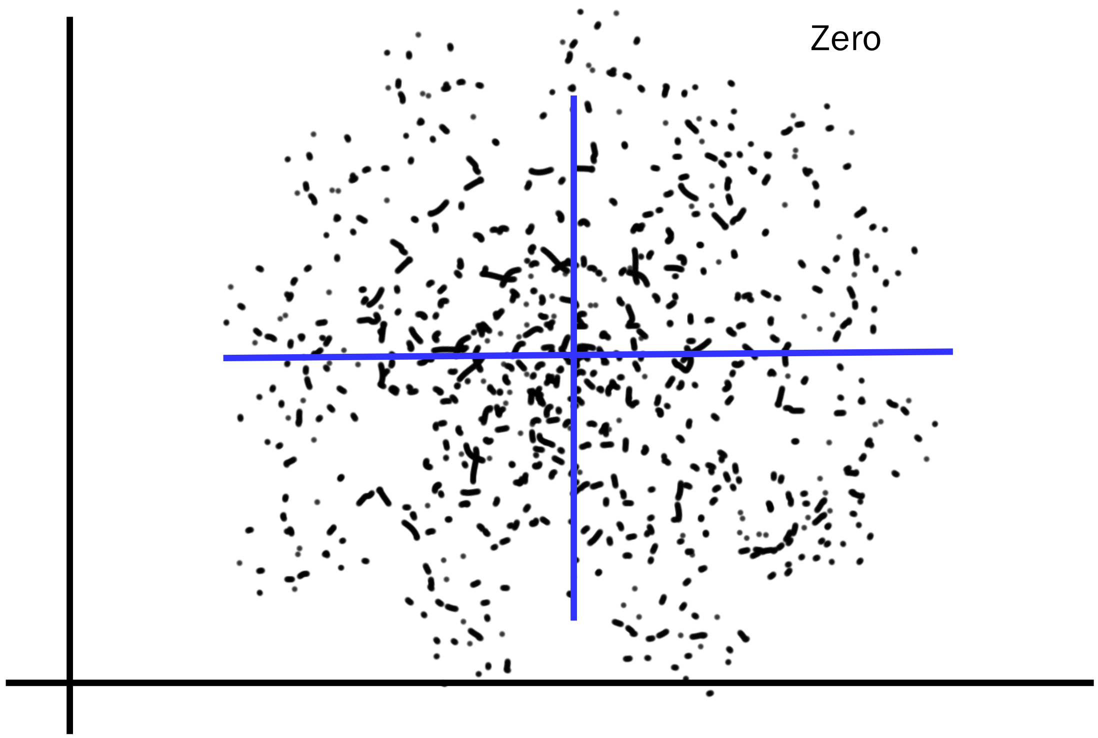
Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero



Covariance

Is the Covariance: (a) positive, (b) negative, (c) zero



Independence and Covariance

X and Y are random variables with PMF:

Y \ X	-1	0	1	$p_Y(y)$
0	1/3	0	1/3	2/3
1	0	1/3	0	1/3
$p_X(x)$	1/3	1/3	1/3	1

$$Y = \begin{cases} 0 & \text{if } X \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

- E[X] = -1(1/3) + 0(1/3) + 1(1/3) = 0
- E[Y] = 0(2/3) + 1(1/3) = 1/3
- Since XY = 0, E[XY] = 0
- Cov(X, Y) = E[XY] – E[X]E[Y] = 0 – 0 = 0

But, X and Y are clearly dependent!

Properties of Covariance

Say X and Y are arbitrary random variables

- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- $\mathrm{Cov}(X,X) = E[X^2] - E[X]E[X] = \mathrm{Var}(X)$
- $\mathrm{Cov}(aX+b,Y) = a\mathrm{Cov}(X,Y)$

Correlation

Viva La Correlatión

Say X and Y are arbitrary random variables

- Correlation of X and Y, denoted $\rho(X, Y)$:

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var(X)Var(Y)}}}$$

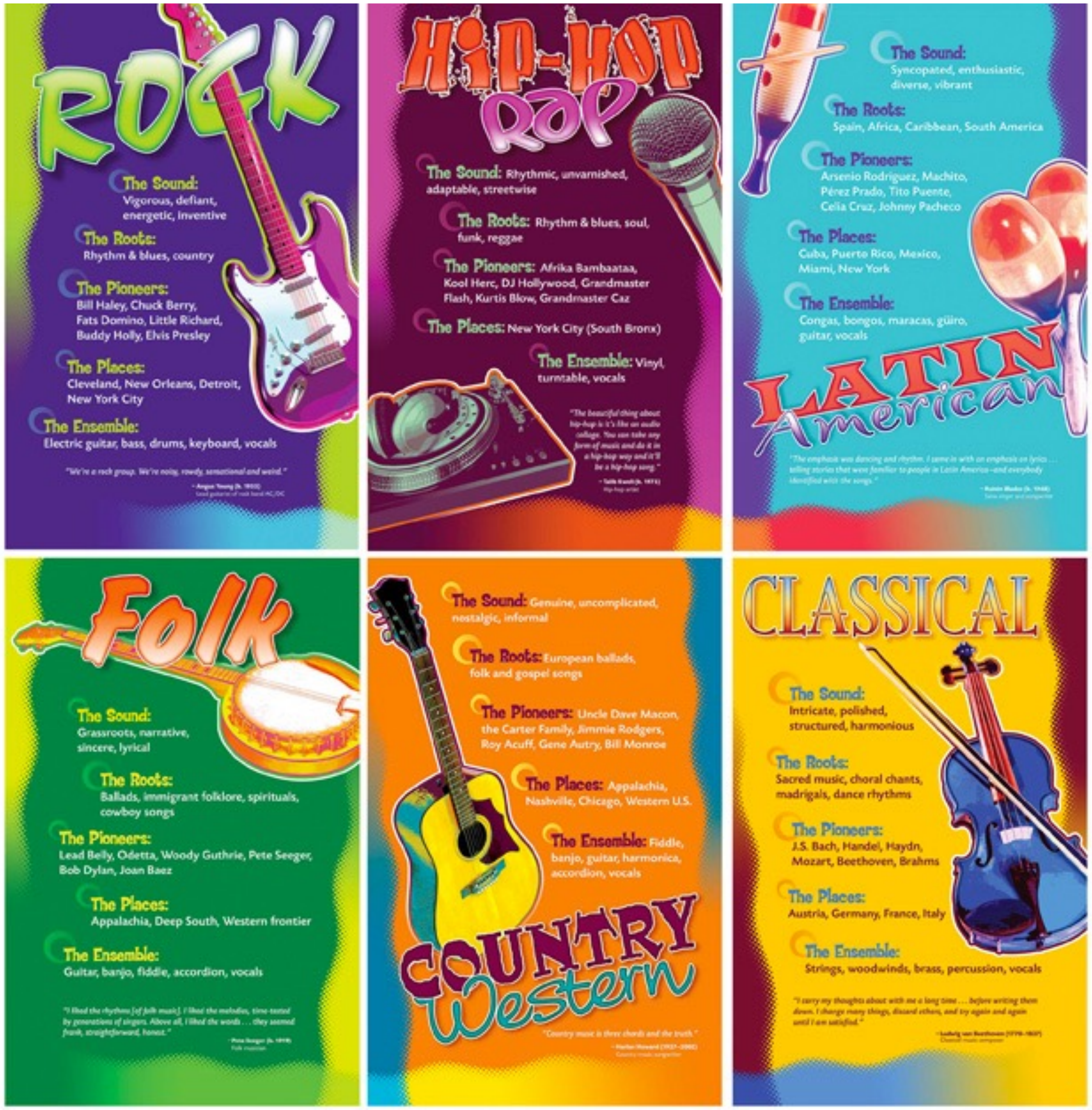
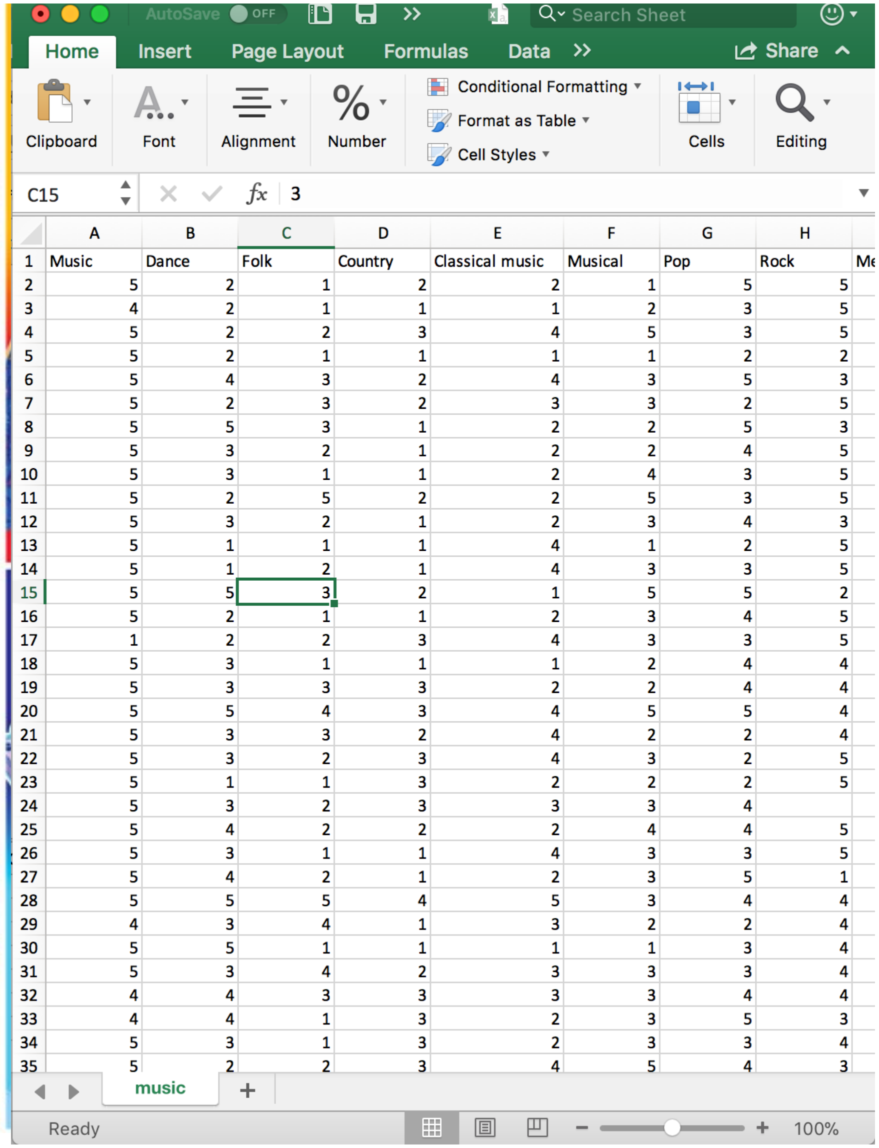
- Note: $-1 \leq \rho(X, Y) \leq 1$
- Correlation measures linearity between X and Y
- $\rho(X, Y) = 1 \quad \Rightarrow \quad Y = aX + b$ where $a = \sigma_y/\sigma_x$
- $\rho(X, Y) = -1 \quad \Rightarrow \quad Y = aX + b$ where $a = -\sigma_y/\sigma_x$
- $\rho(X, Y) = 0 \quad \Rightarrow$ absence of linear relationship
  - But, X and Y can still be related in some other way!
- If $\rho(X, Y) = 0$, we say X and Y are "uncorrelated"
  - Note: Independence implies uncorrelated, but **not** vice versa!

Viva La Correlatión

Say X and Y are arbitrary random variables

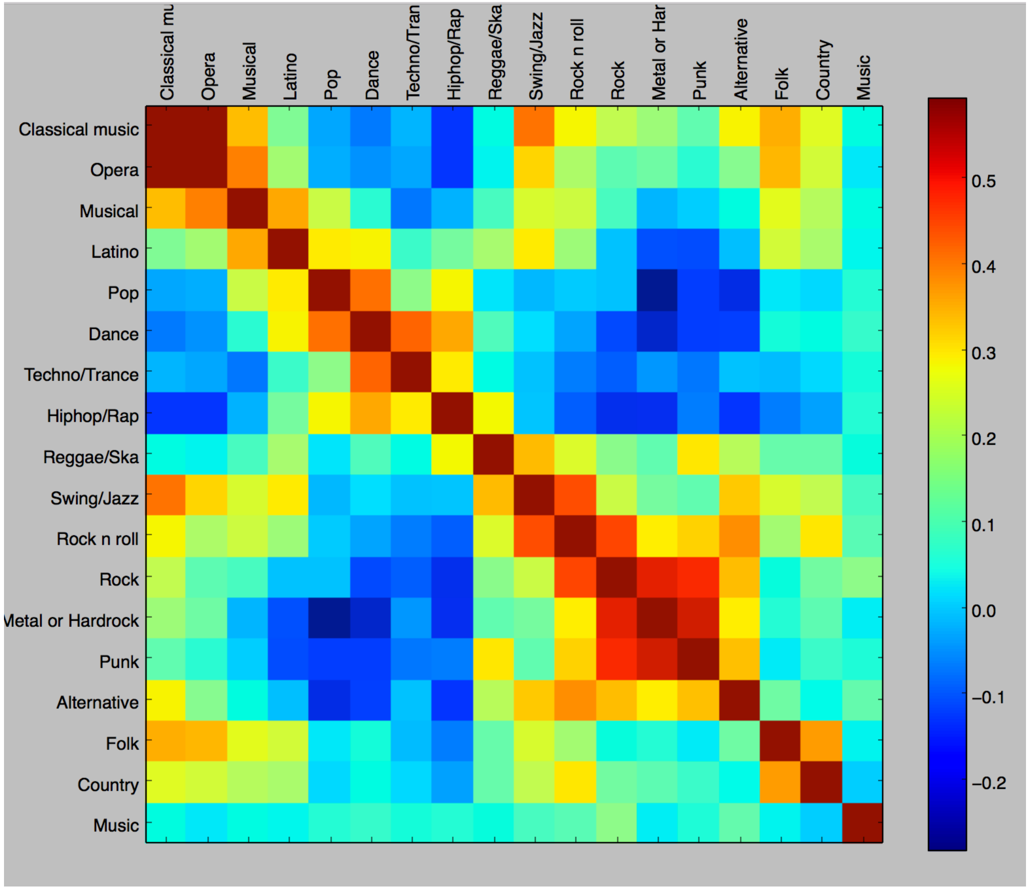
- Correlation of X and Y, denoted $\rho(X, Y)$:

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var(X)Var(Y)}}}$$

Say Y = cX. Correlation should be 1.



	A	B	C	D	E	F	G	H	
1	Music	Dance	Folk	Country	Classical music	Musical	Pop	Rock	Me
2	5	2	1	2	2	1	5	5	
3	4	2	1	1	1	2	3	5	
4	5	2	2	3	4	5	3	5	
5	5	2	1	1	1	1	2	2	
6	5	4	3	2	4	3	5	3	
7	5	2	3	2	3	3	2	5	
8	5	5	3	1	2	2	5	3	
9	5	3	2	1	2	2	4	5	
10	5	3	1	1	2	4	3	5	
11	5	2	5	2	2	5	3	5	
12	5	3	2	1	2	3	4	3	
13	5	1	1	1	4	1	2	5	
14	5	1	2	1	4	3	3	5	
15	5	5	3	2	1	5	5	2	
16	5	2	1	1	2	3	4	5	
17	1	2	2	3	4	3	3	5	
18	5	3	1	1	1	2	4	4	
19	5	3	3	3	2	2	4	4	
20	5	5	4	3	4	5	5	4	
21	5	3	3	2	4	2	2	4	
22	5	3	2	3	4	3	2	5	
23	5	1	1	3	2	2	2	5	
24	5	3	2	3	3	3	4		
25	5	4	2	2	2	4	4	5	
26	5	3	1	1	4	3	3	5	
27	5	4	2	1	2	3	5	1	
28	5	5	5	4	5	3	4	4	
29	4	3	4	1	3	2	2	4	
30	5	5	1	1	1	1	3	4	
31	5	3	4	2	3	3	3	4	
32	4	4	3	3	3	3	4	4	
33	4	4	1	3	2	3	5	3	
34	5	3	1	3	2	3	3	4	
35	5	2	2	3	4	5	4	3	

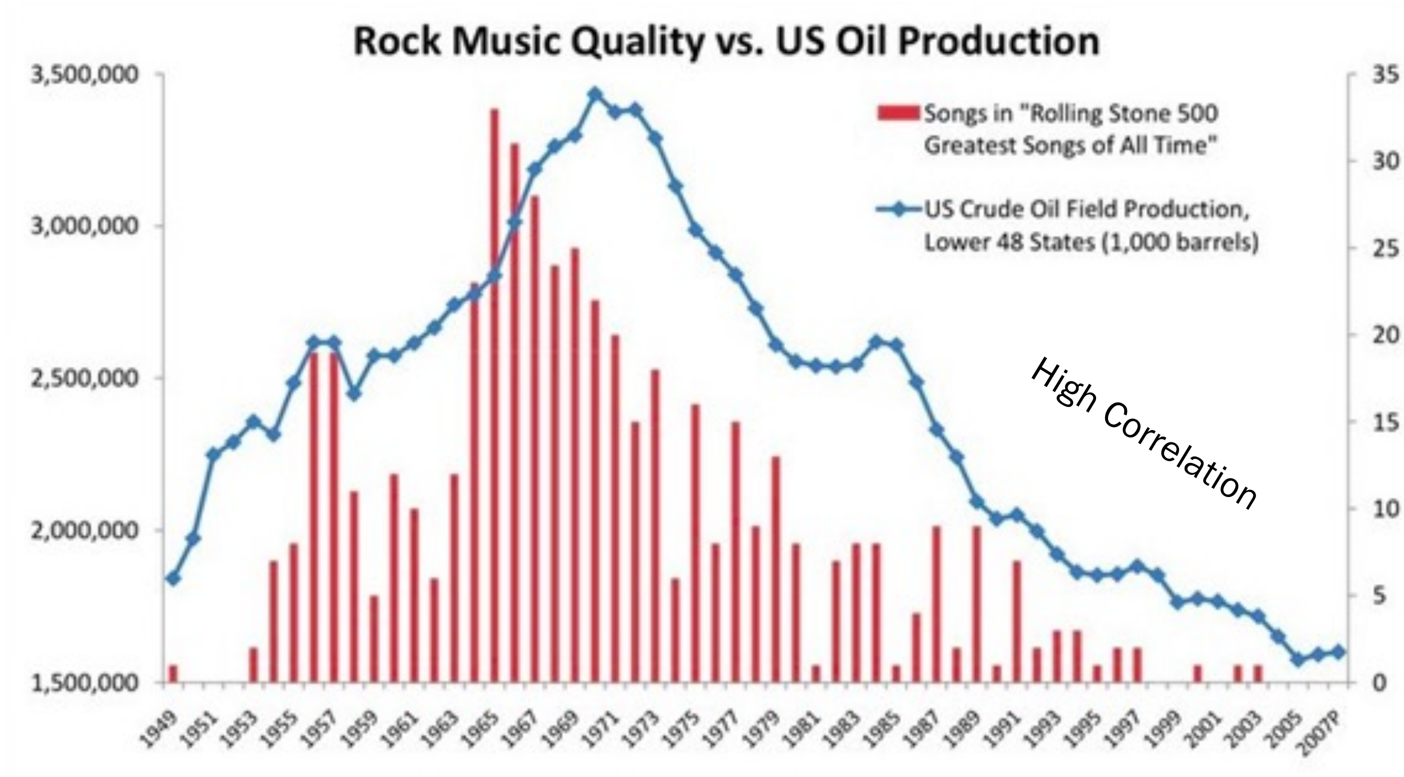
Correlation of Music Tastes



How do you know if your model is good?

Answer: it is accurate at inference (especially tasks you care about)

Rock Music Vs Oil?

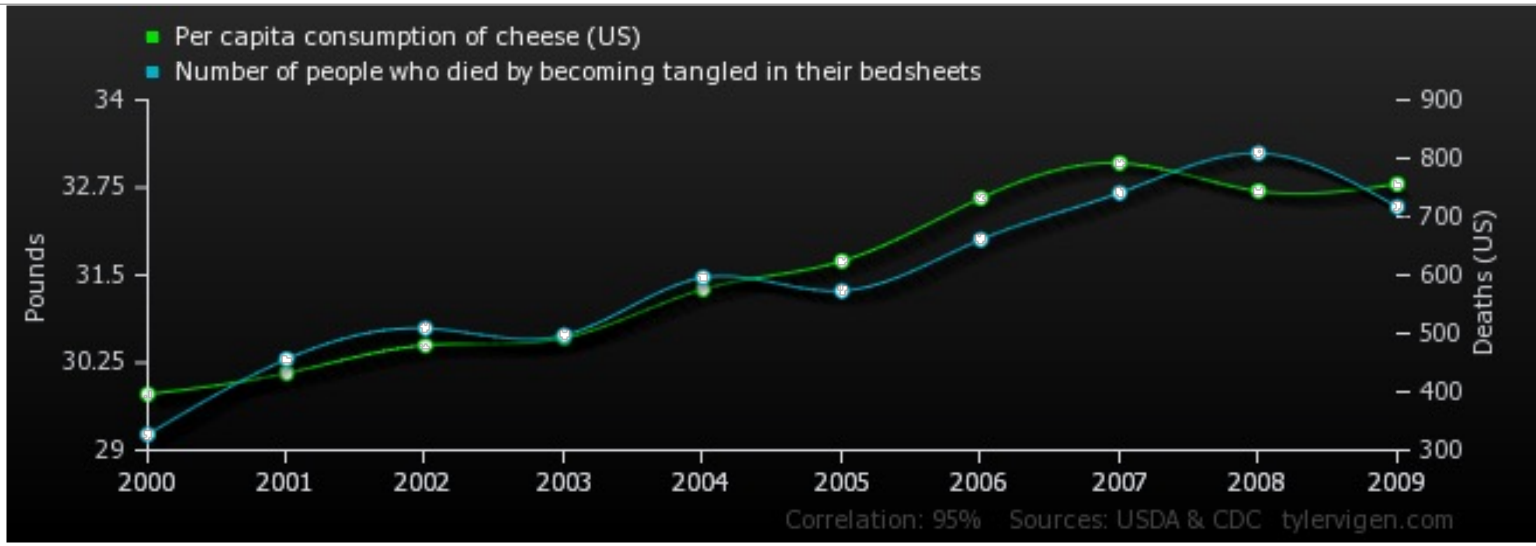


Hubbert Peak Theory

http://www.aei.org/publication/blog/

Stanford University

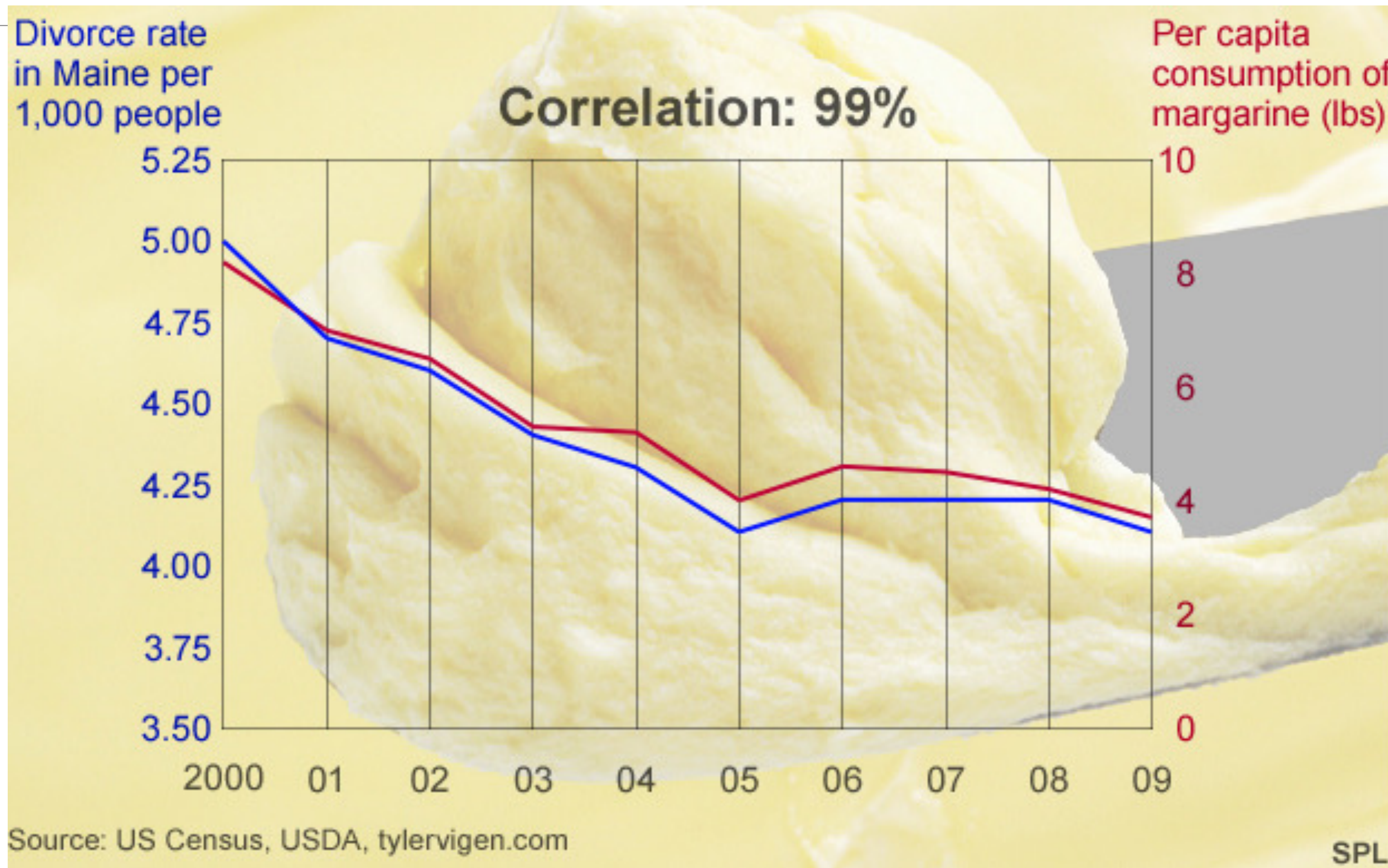
Tell your friends!



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Per capita consumption of cheese (US) *Pounds (USDA)*	29.8	30.1	30.5	30.6	31.3	31.7	32.6	33.1	32.7	32.8
Number of people who died by becoming tangled in their bedsheets *Deaths (US) (CDC)*	327	456	509	497	596	573	661	741	809	717

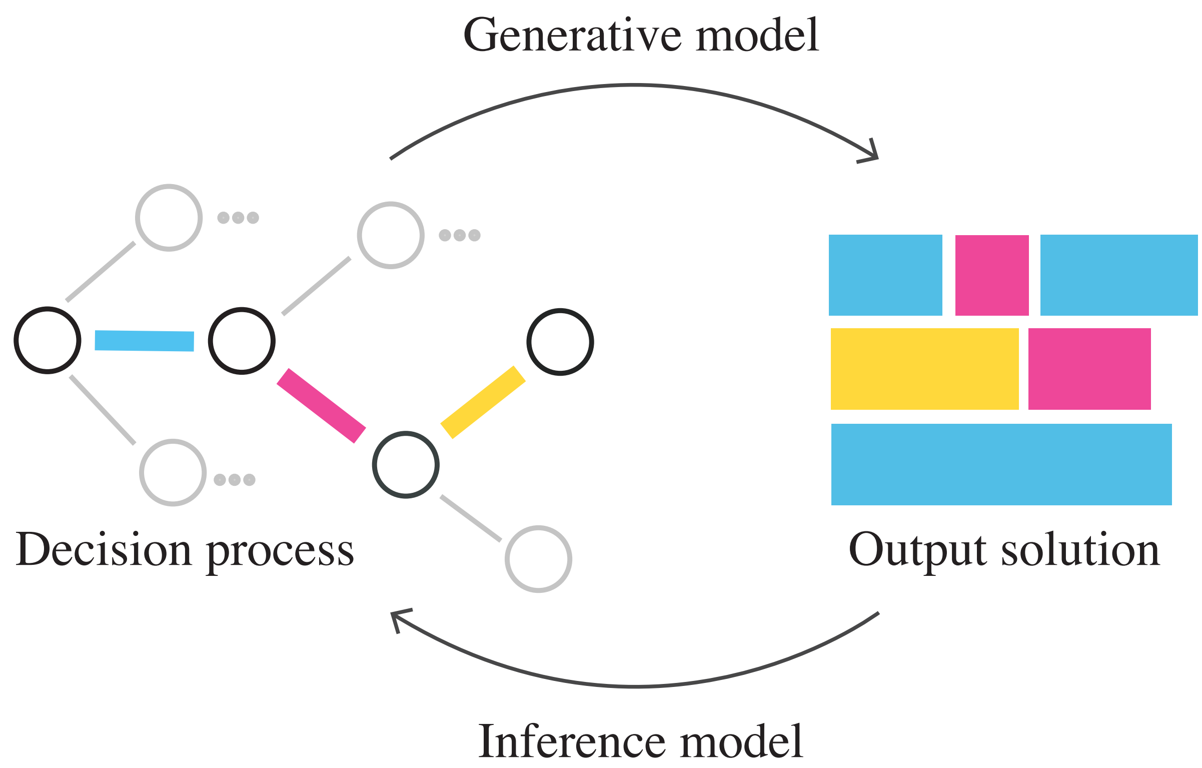
Correlation: 0.947091

Divorce Vs Butter?



http://www.bbc.com/news/magazine-27537142

Wednesday: General Inference



Goodluck on the midterm!!