



Modeling

Chris Piech

CS109, Stanford University

Midterm Location by Last Name



CEMEX:
A through K

Hewlett 200:
L through S

Hewlett 201:
T through Z

Look out for your assignment! Let us know if you are at risk

No Class on Monday



This is how I
imagine students
are when they are
studying 😊

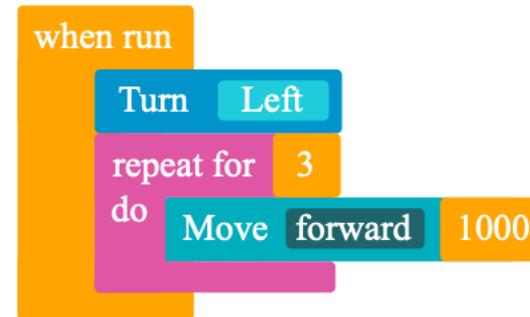
(recall why we are
here. Learning is a joy.
You are growing)

Computers Couldn't Understand Student Code

60,000 students attempted this problem
37,000 unique solutions



Challenge

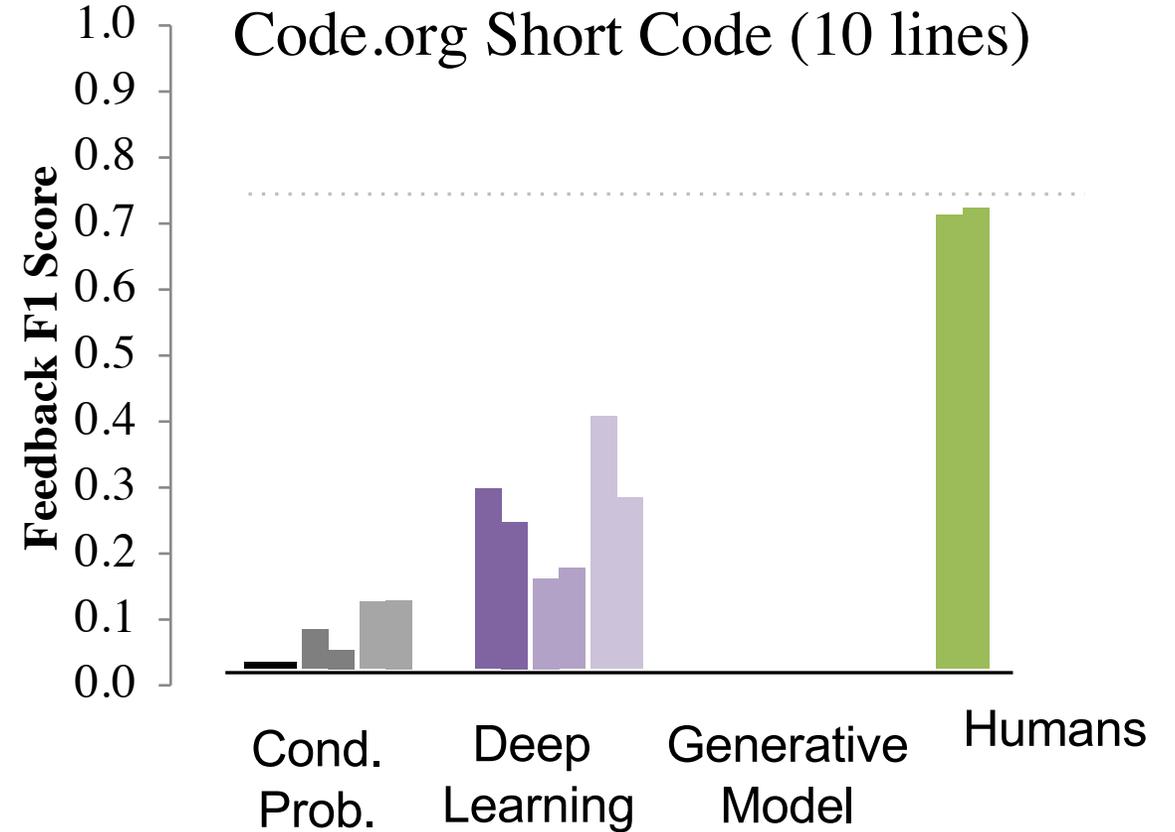


Student Code

You need to
move and
turn in your
loop

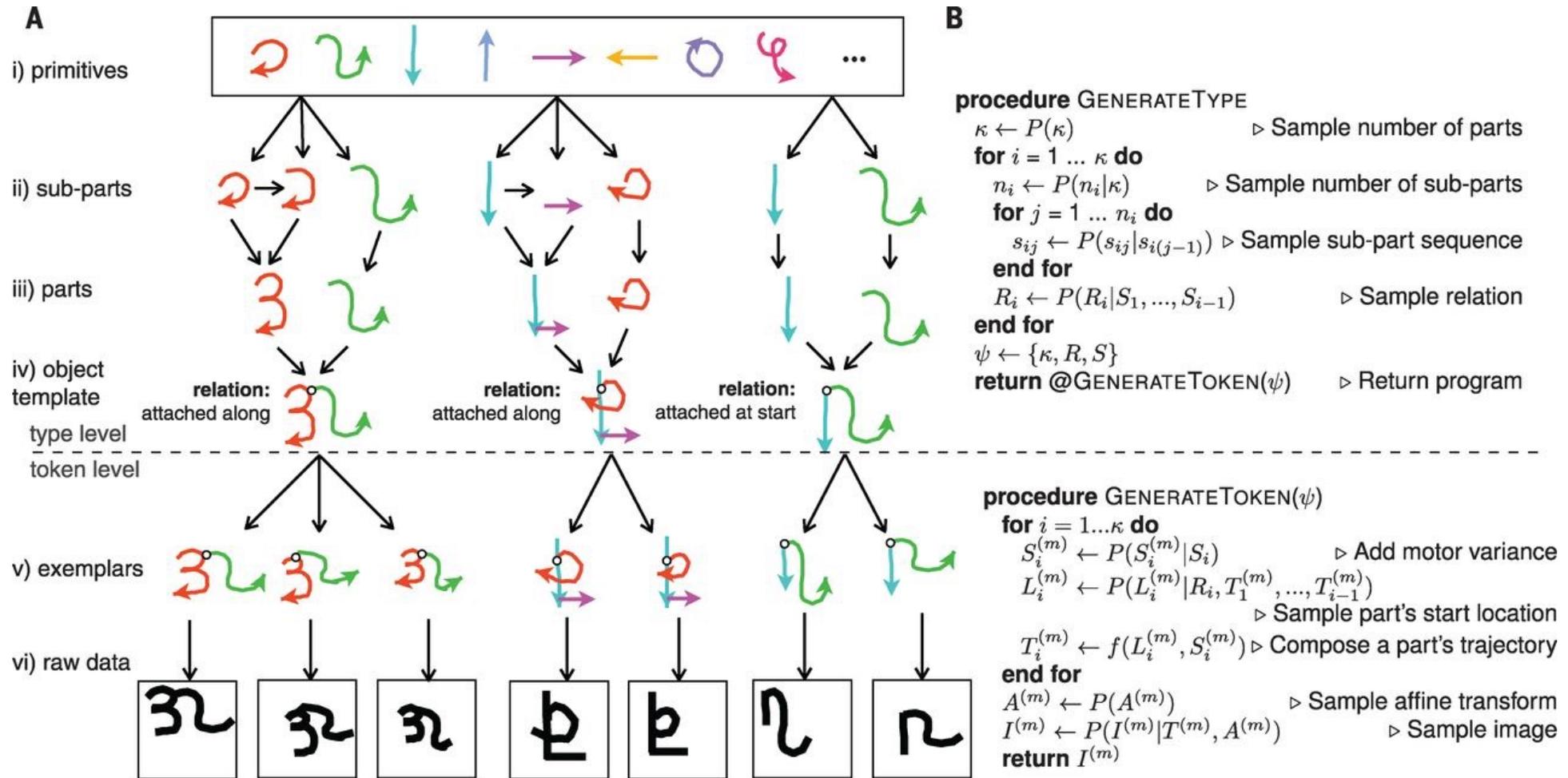
Insight

Computers Couldn't Understand Student Code



Generative Model of Characters

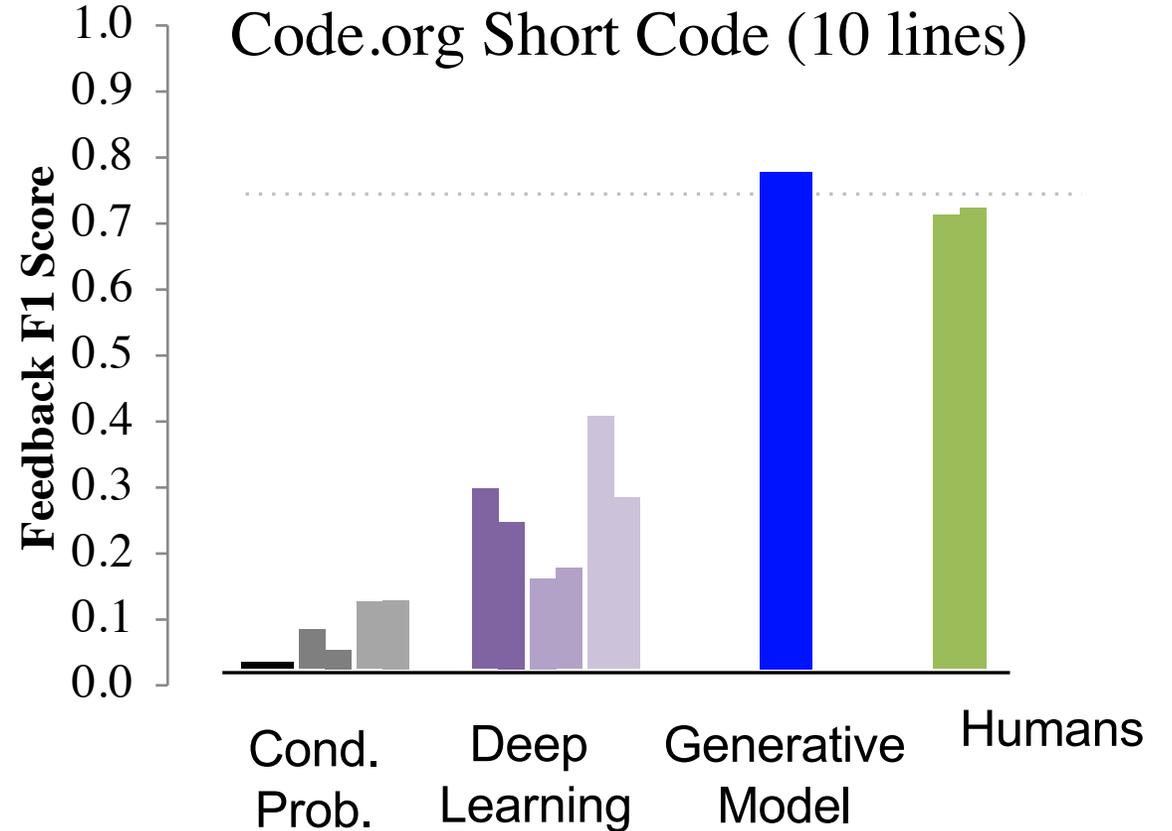
Lake et al, 2015



Computers Couldn't Understand Student Code



*Outstanding Student
paper award, AAAI 2019* ²⁰²²



Review

Last Week: Joint Distributions

Joint Distribution *noun*

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

Eg:

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

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

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

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.

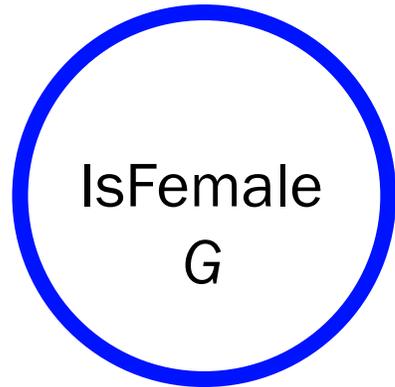
TLDR: conditional probability with random variables.

Inference with Continuous

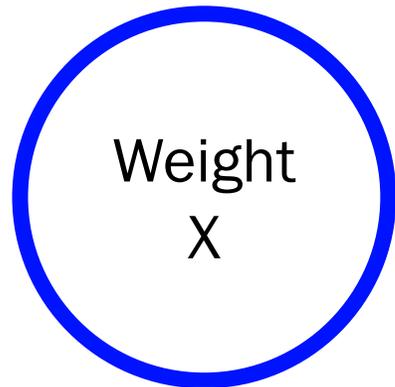
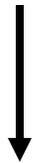
Q: At birth, girl elephant weights are distributed as a Gaussian with mean = 160kg, std = 7kg. At birth, boy elephant weights are distributed as a Gaussian with mean = 165kg, std = 3kg. All you know about a newborn elephant is that it is 163kg. What is the probability that it is a girl?



Model Shown Graphically



$G = 1$ is $\text{Bern}(p = 0.5)$



$X \mid G = 1$ 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?

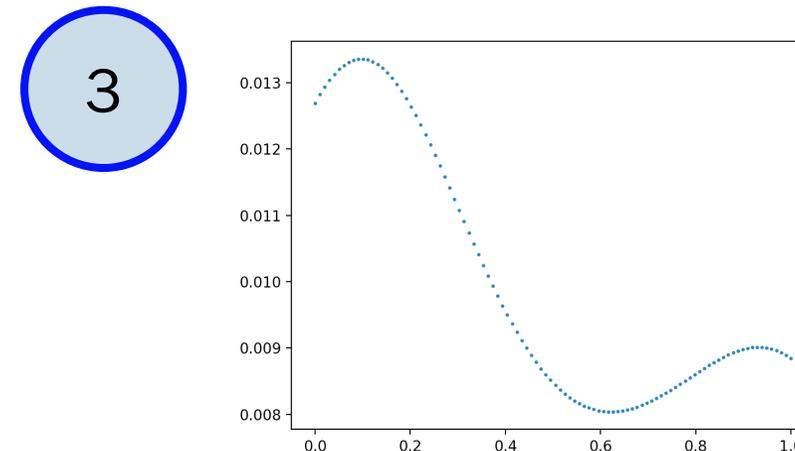
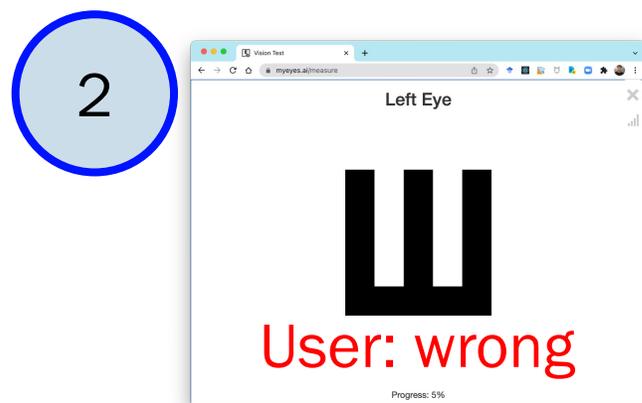
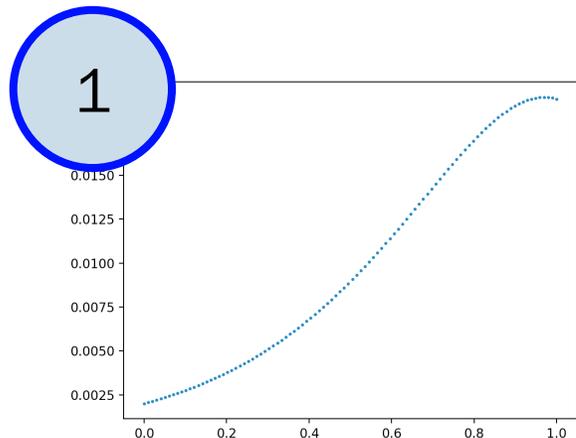
$$f(G = g, X = x)$$

$$= f(X = x \mid G = g)P(G = g)$$

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

Harder when inferred random variable
is not a Bernoulli

Inference on a non-bernoulli random variable



$$P(A = a)$$

Observation $Y = 0$

$$P(A = a|E)$$

Known function. Also needs font size (constant)

$$P(A = a|Y = 0) = \frac{P(Y = 0|A = a)P(A = a)}{P(Y = 0)}$$

Number or Dictionary?

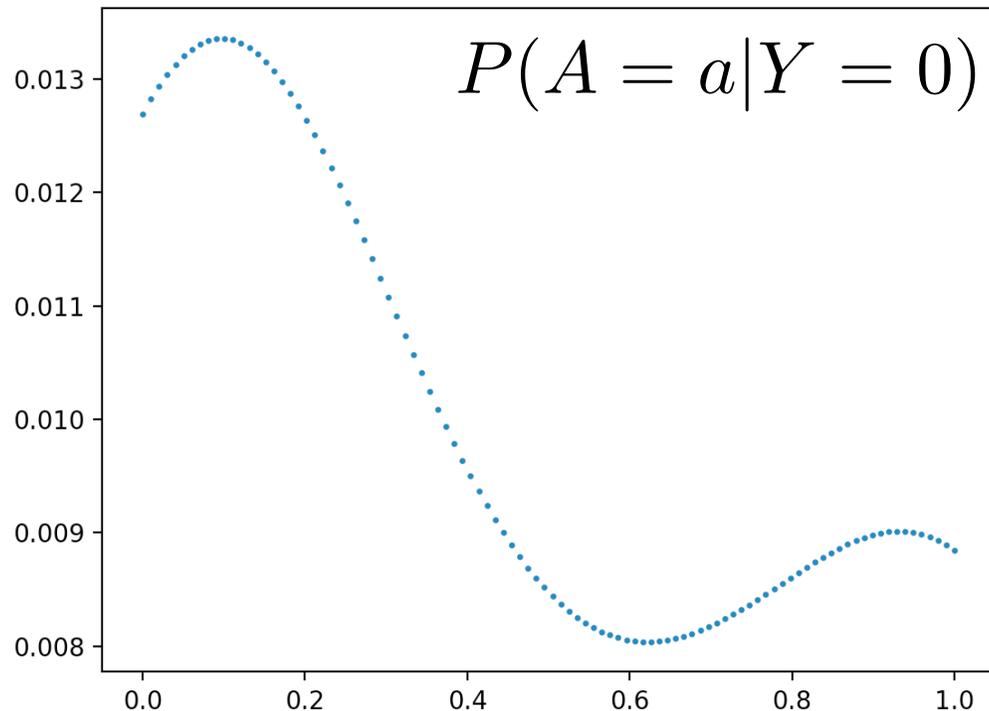
belief

$$P(A = a | Y = 0) = \frac{P(Y = 0 | A = a)P(A = a)}{P(Y = 0)}$$

belief[a] = 0.001

Inference on a non-bernoulli random variable

In plain English: run bayes for each value of a



RV bayes as code

```
def update(belief, obs):  
    for a in support:  
        prior_a = belief[a]  
        likelihood = calc_likelihood(a, obs)  
        belief[a] = prior_a * likelihood  
    normalize(belief)
```

likelihood

$$P(A = a | Y = 0) = \frac{P(Y = 0 | A = a)P(A = a)}{P(Y = 0)}$$

Normalize???

RV bayes as code

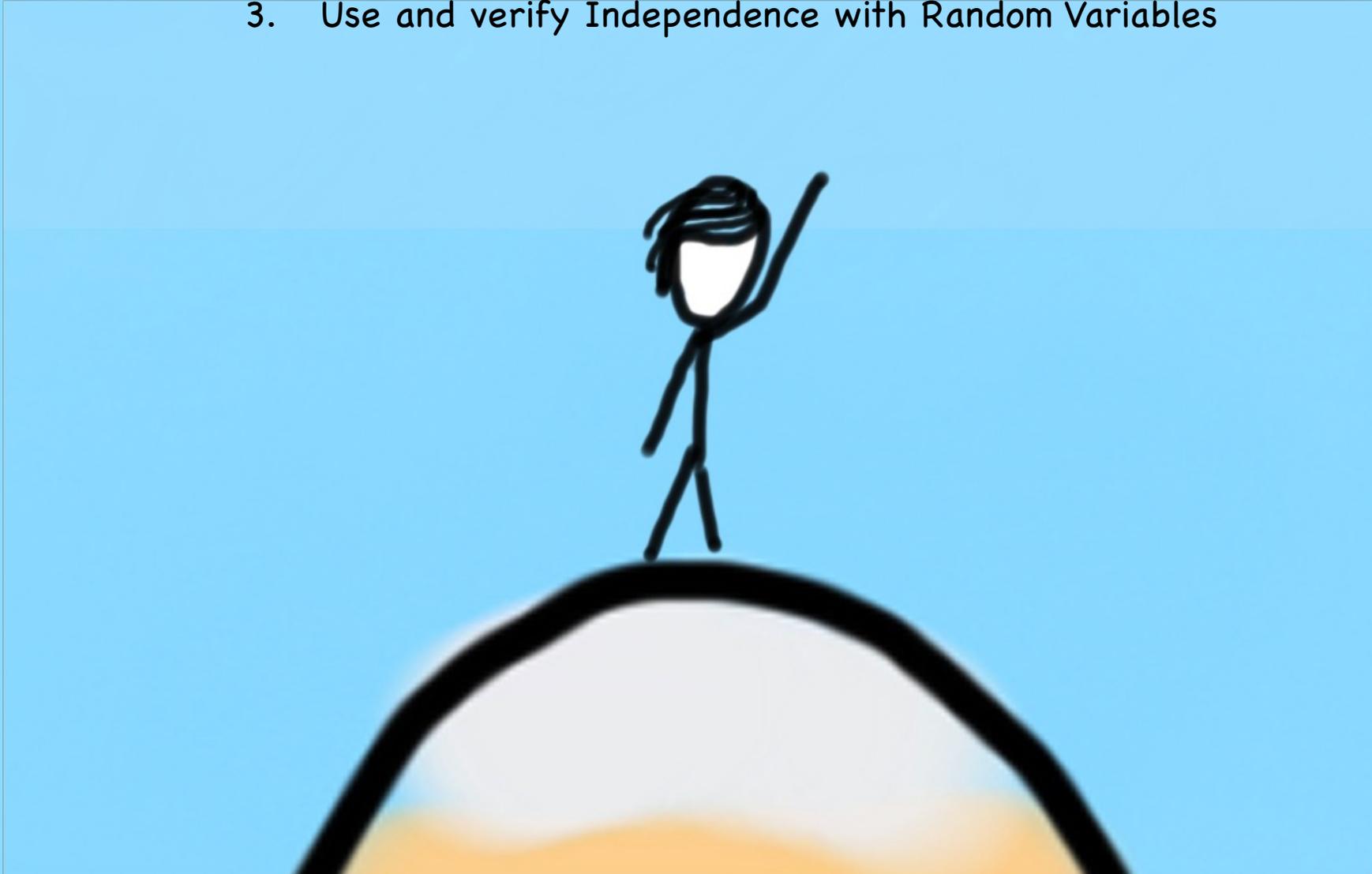
```
def update(belief, obs):  
    for a in support:  
        prior_a = belief[a]  
        likelihood = calc_likelihood(a, obs)  
        belief[a] = prior_a * likelihood  
normalize(belief)
```

In plain English: this
is the numerator,
summed over all
values of A

$$\begin{aligned}P(A = a|Y = 0) &= \frac{P(Y = 0|A = a)P(A = a)}{P(Y = 0)} \\ &= \frac{P(Y = 0|A = a)P(A = a)}{\sum_{x \in A} P(Y = 0, A = x)} \\ &= \frac{P(Y = 0|A = a)P(A = a)}{\sum_{x \in A} P(Y = 0|A = x)P(A = x)}\end{aligned}$$

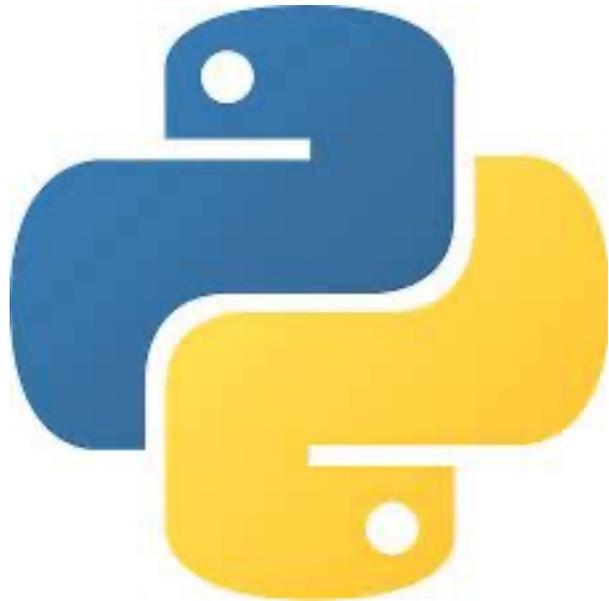
Learning Goals

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

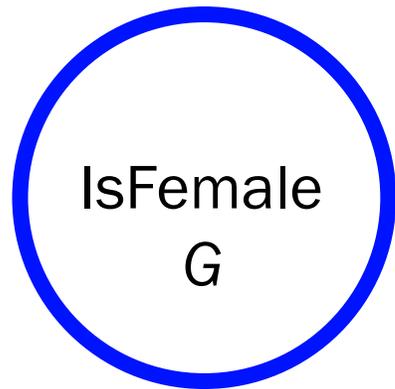
Model Version #1: Python That Outputs a **Joint** Sample



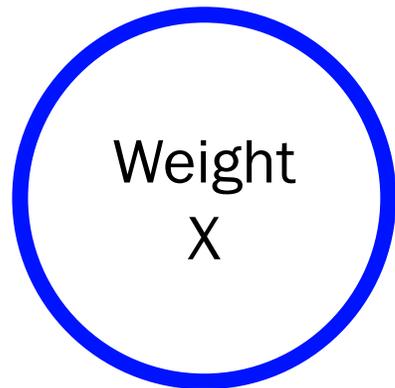
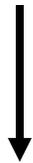
Sample Baby
Elephant

Sex: Female
Weight: 161kg

Model Version #2: Bayesian Network



$G = 1$ is $\text{Bern}(p = 0.5)$



$X | G = 1$ is $\text{N}(\mu = 160, \sigma^2 = 7^2)$

$X | G = 0$ is $\text{N}(\mu = 165, \sigma^2 = 3^2)$

Does this define the joint?

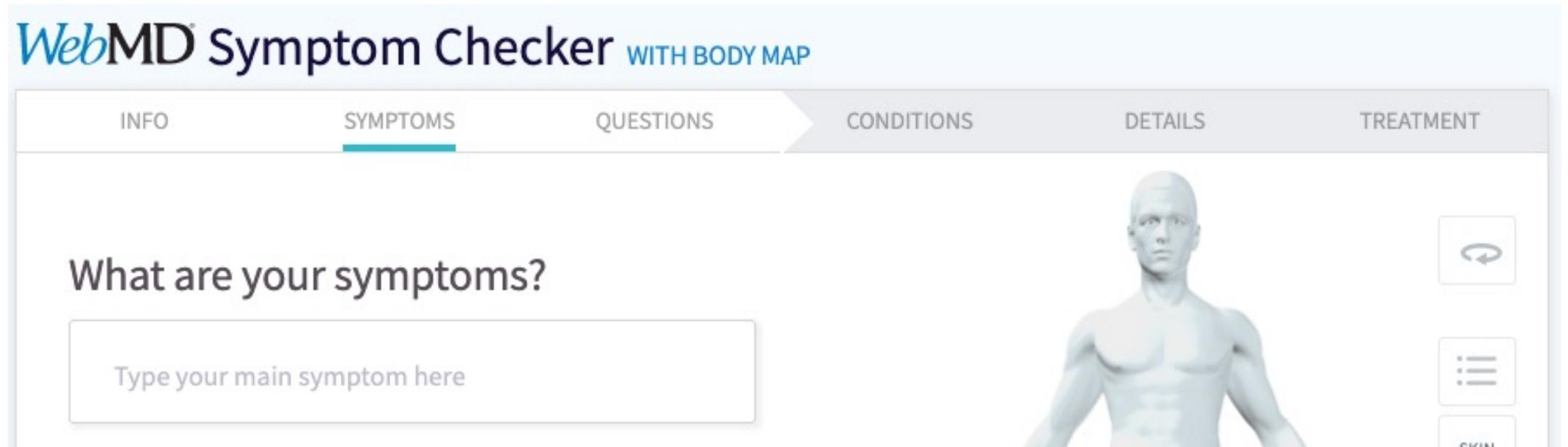
$$f(G = g, X = x)$$

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

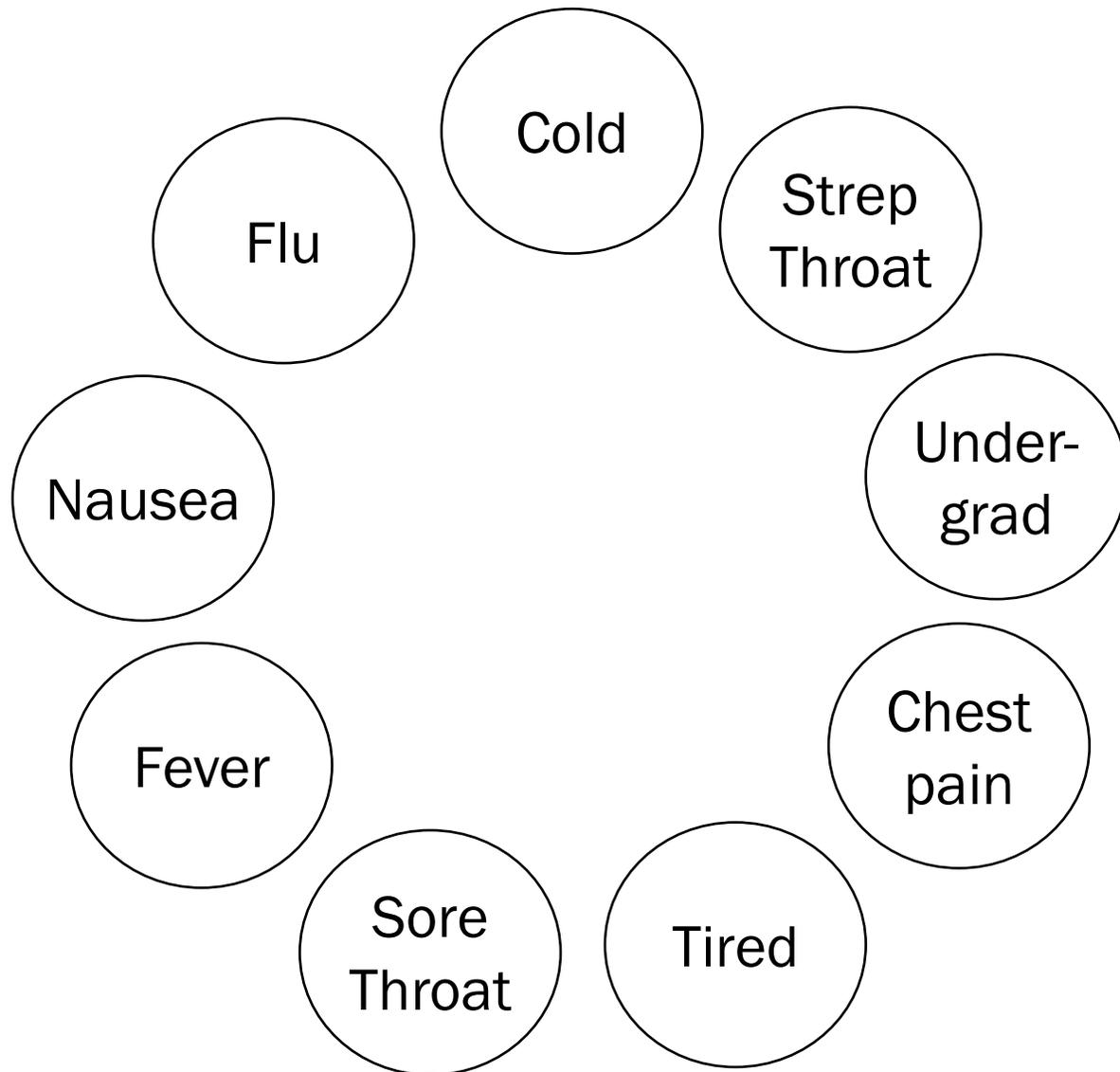
Why You Need a Model

*Web*MD[®]

Inference



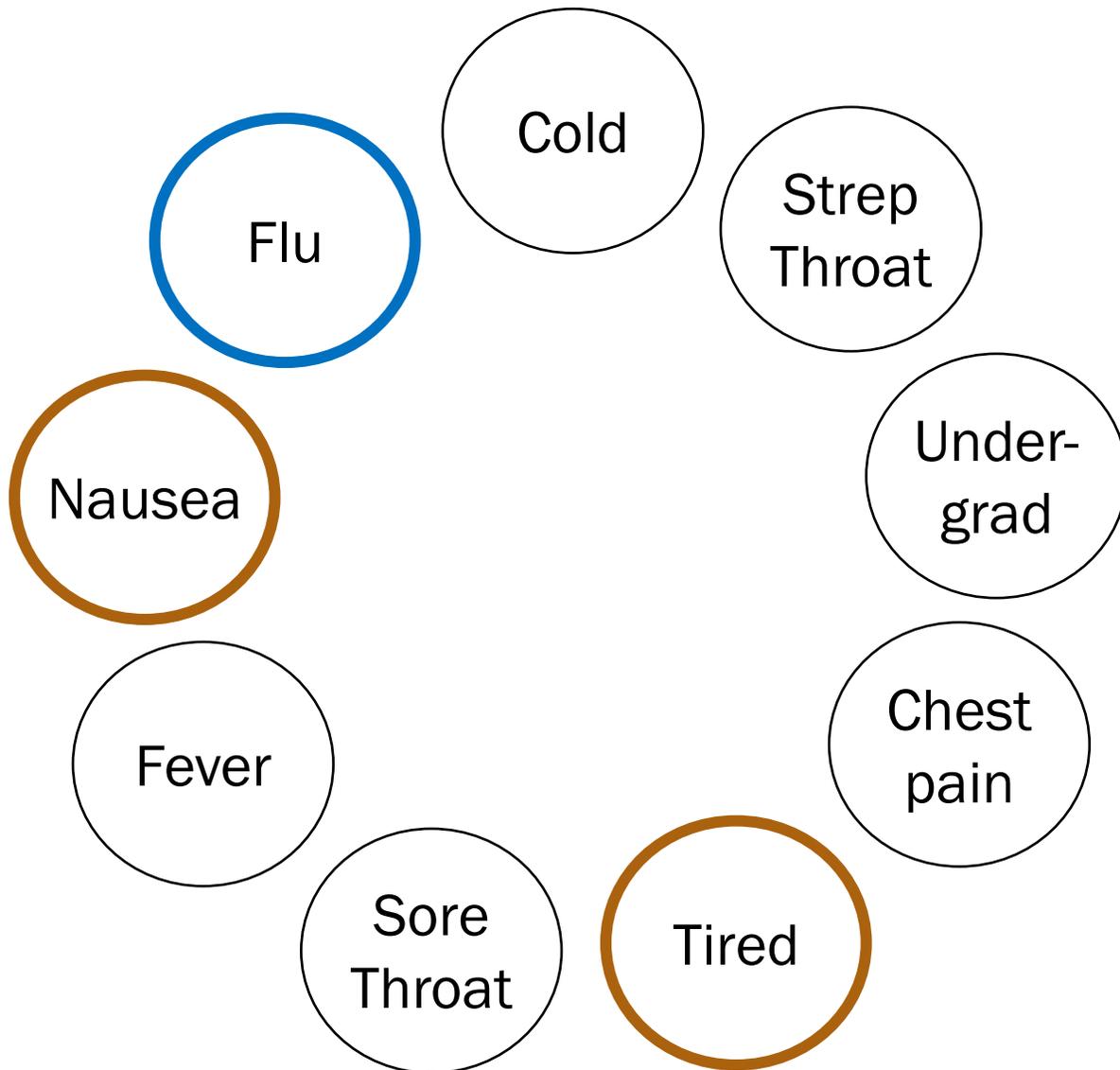
Inference



Inference question:

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

Inference

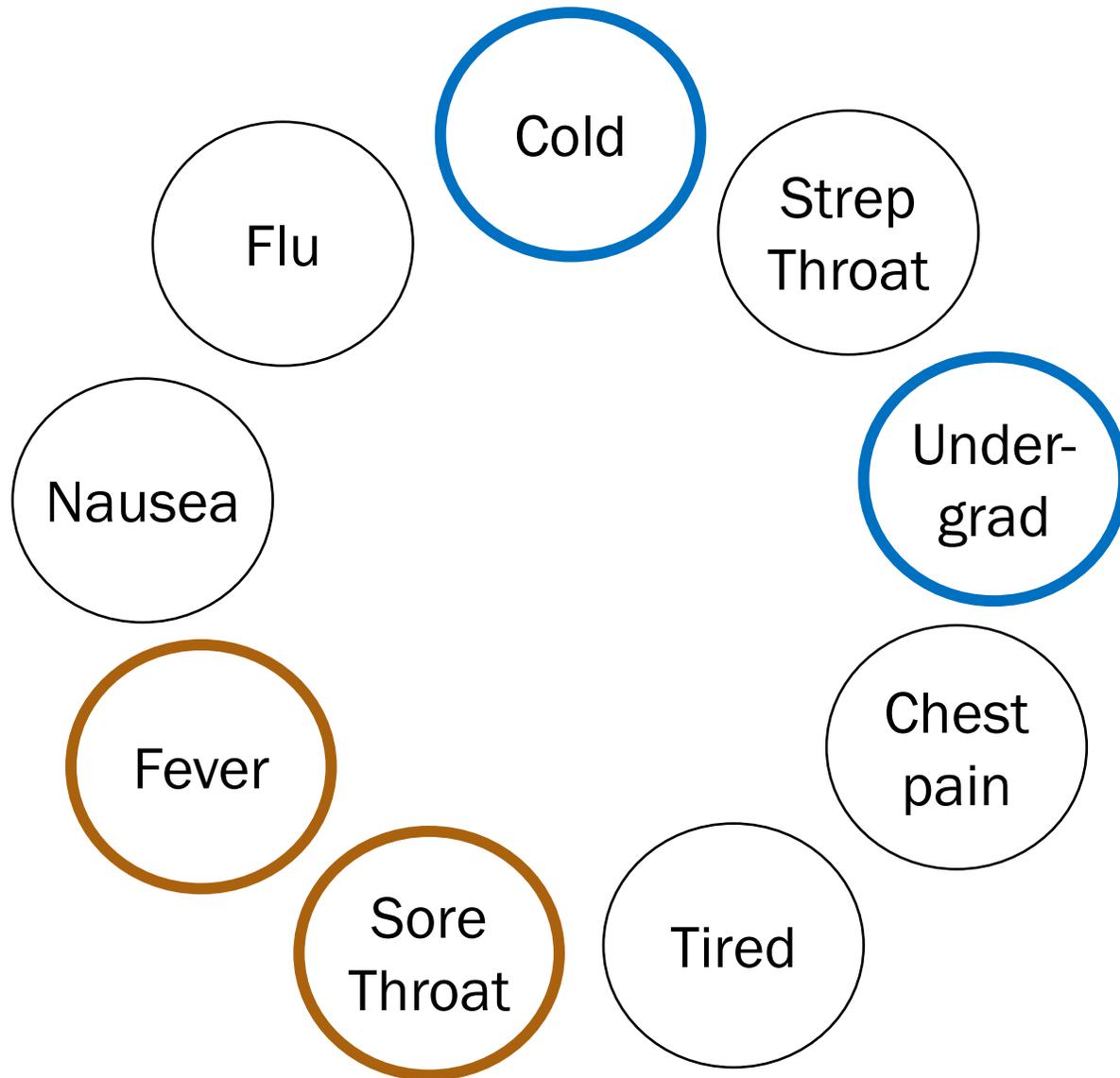


One inference question:

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

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

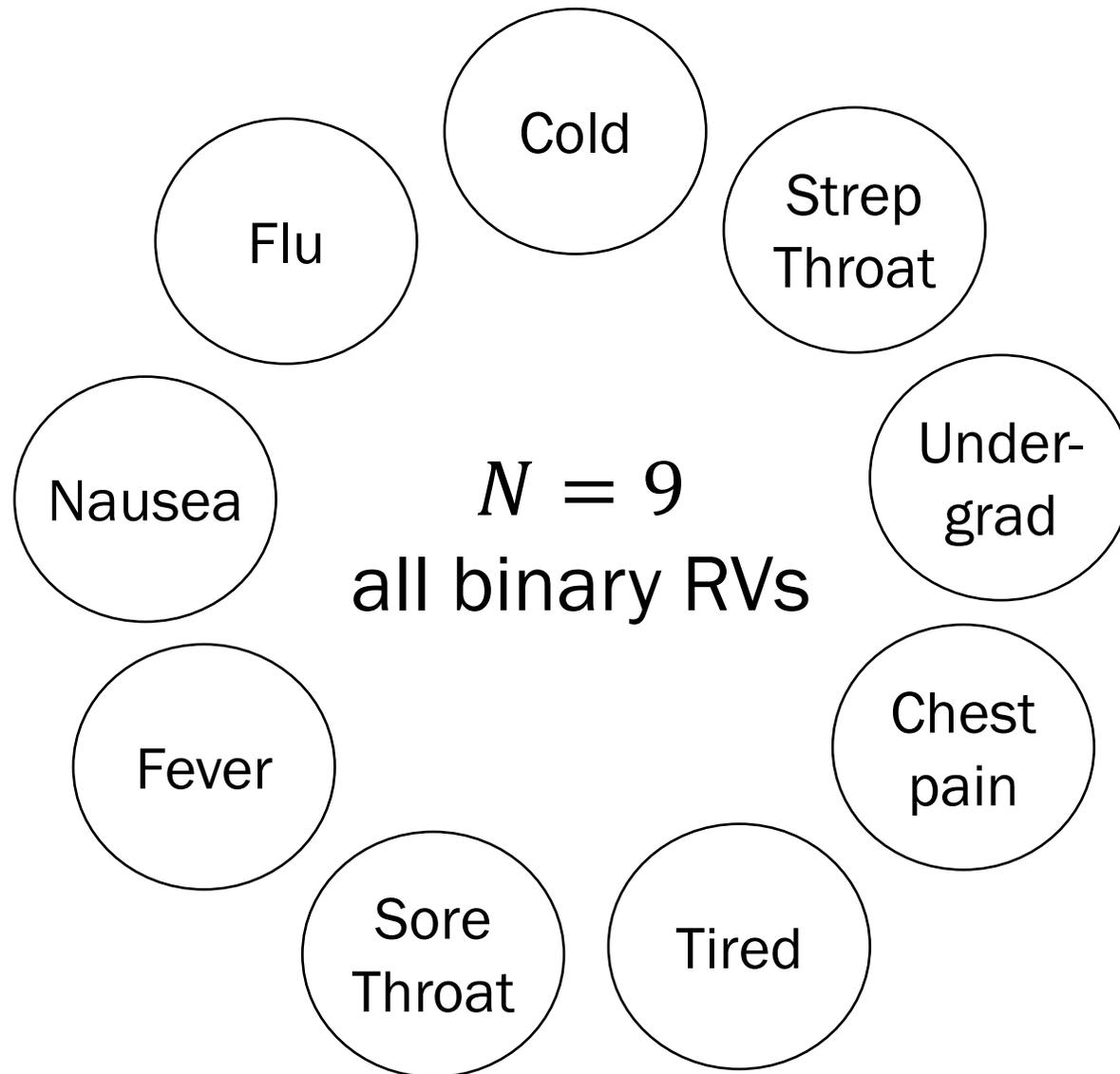
Inference



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



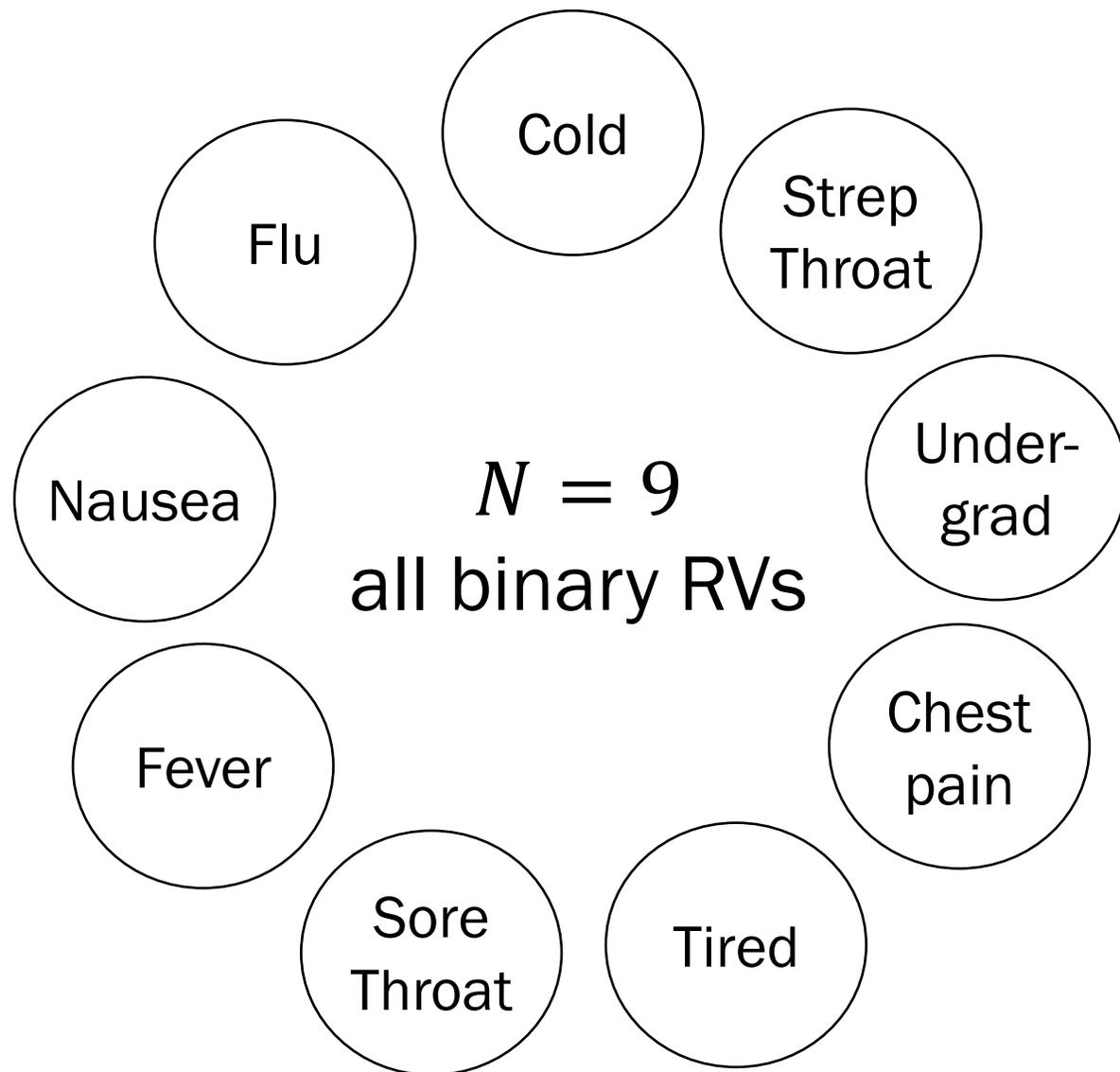
If we knew the **joint distribution**, we can answer all probabilistic inference questions.

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



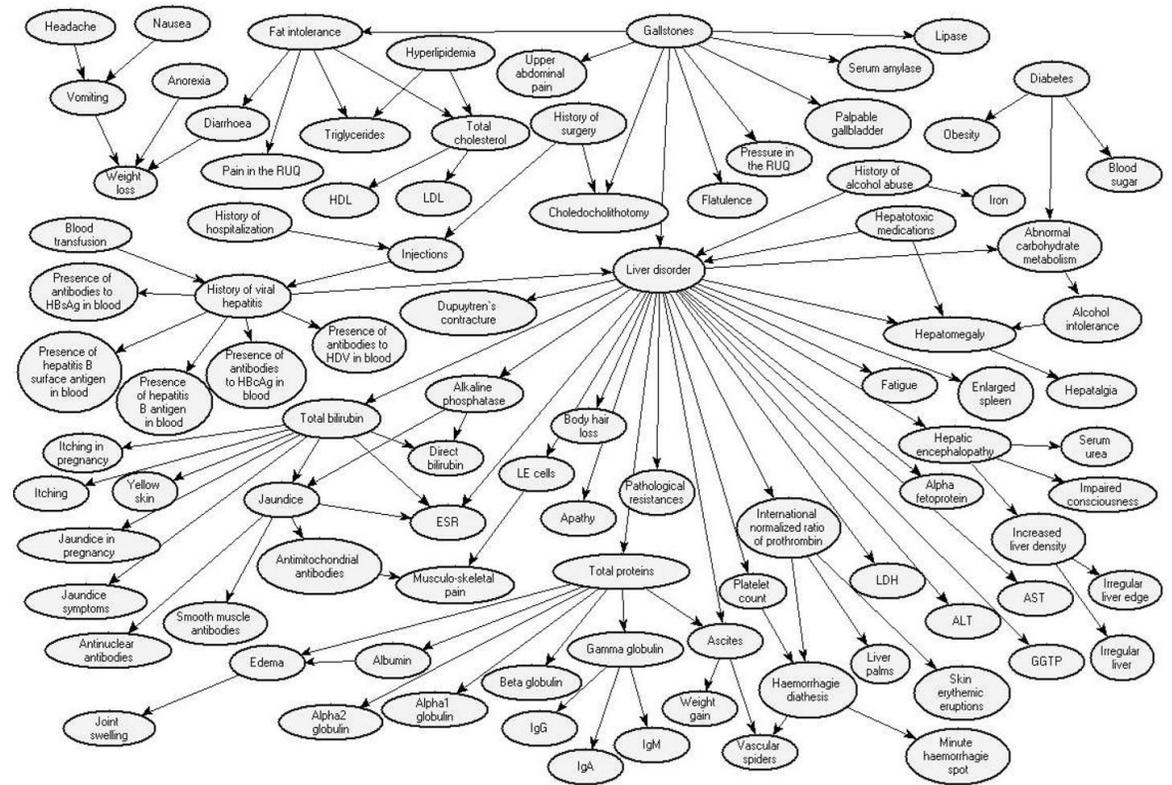
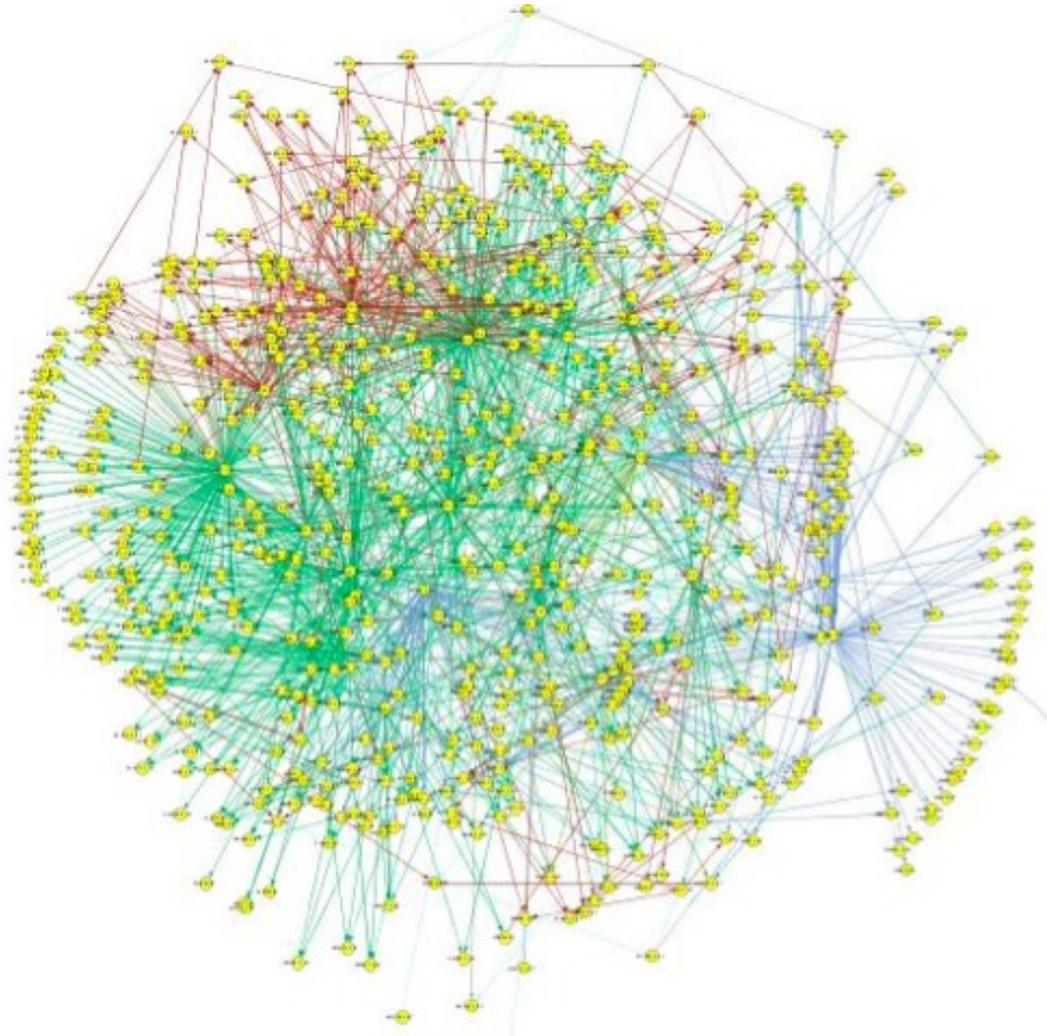
If we knew the **joint distribution**, we can answer all probabilistic inference questions.

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

Flu

Under-
grad

Fever

Tired

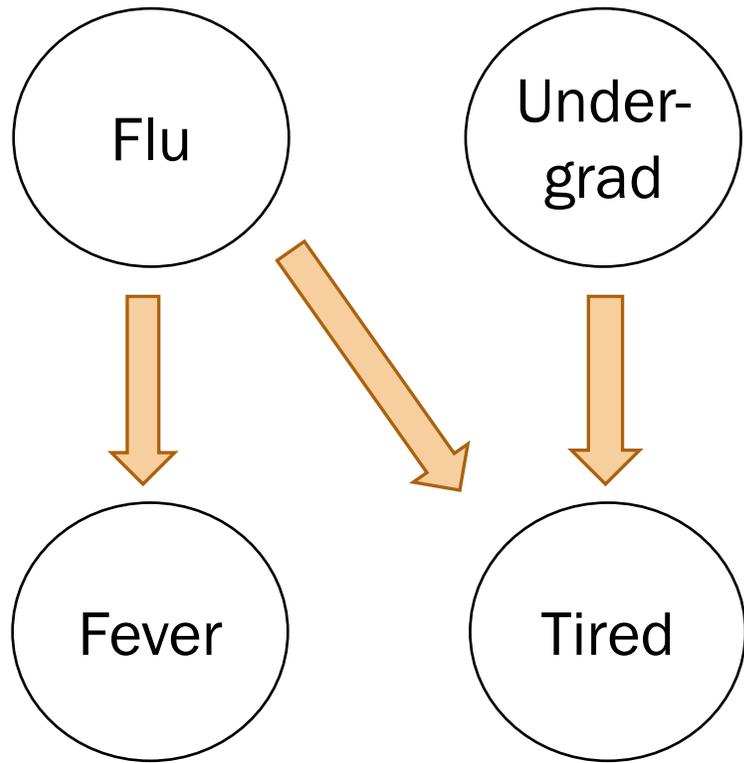
Great! Just specify $2^4 = 16$ joint probabilities...?

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

What would a Stanford flu expert do?

Describe the joint distribution using causality!

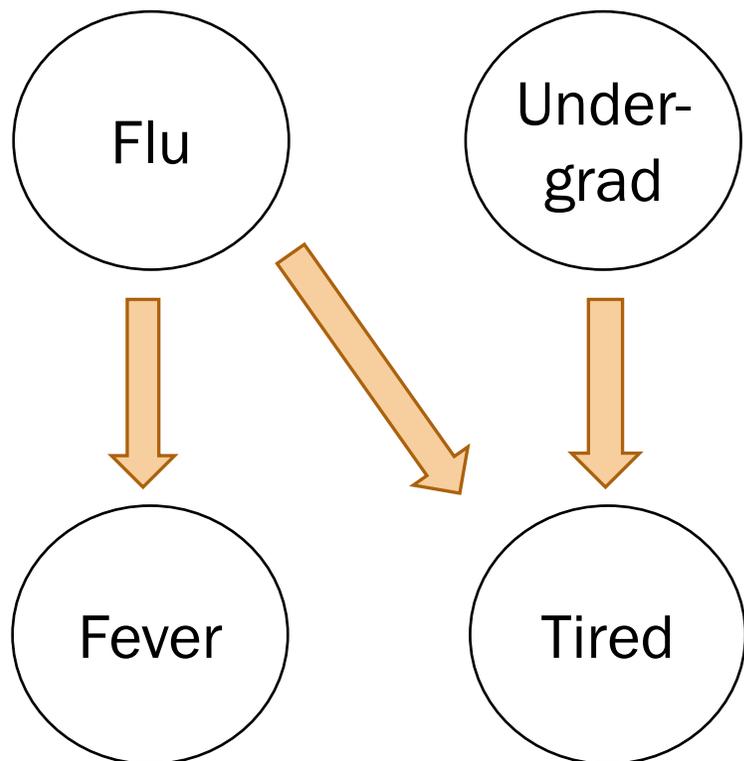
Constructing a Bayesian Network



What would a Stanford flu expert do?

- ✓ 1. Describe the joint distribution using causality.
2. Provide $P(\text{values}|\text{parents})$ for each random variable
3. Implicitly assumes independences.

Constructing a Bayesian Network



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

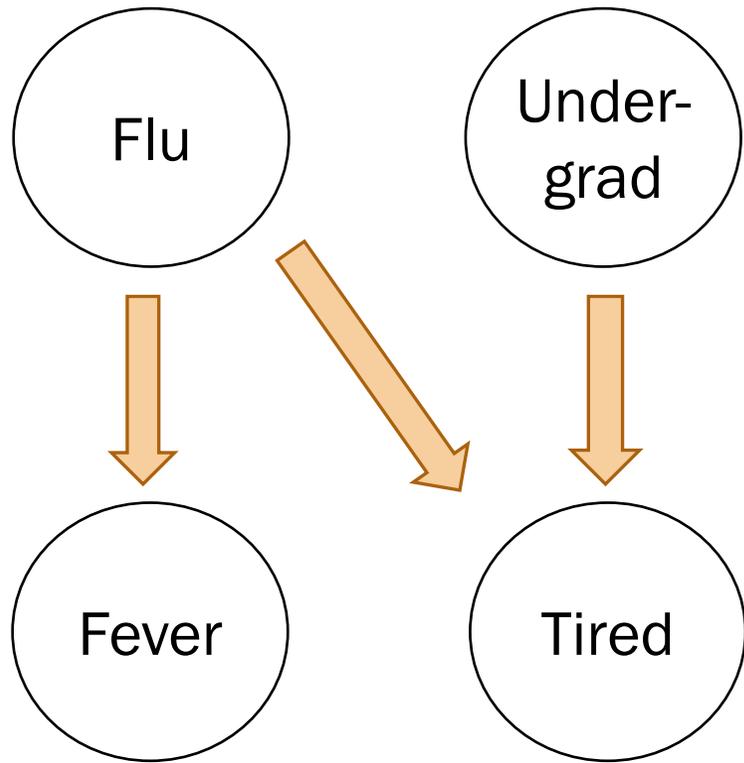
In a Bayesian Network,
Each random variable is caused by
its **parents**. Def $P(\text{node} \mid \text{parents})$

- Node: random variable
- Directed edge: causality

Examples:

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

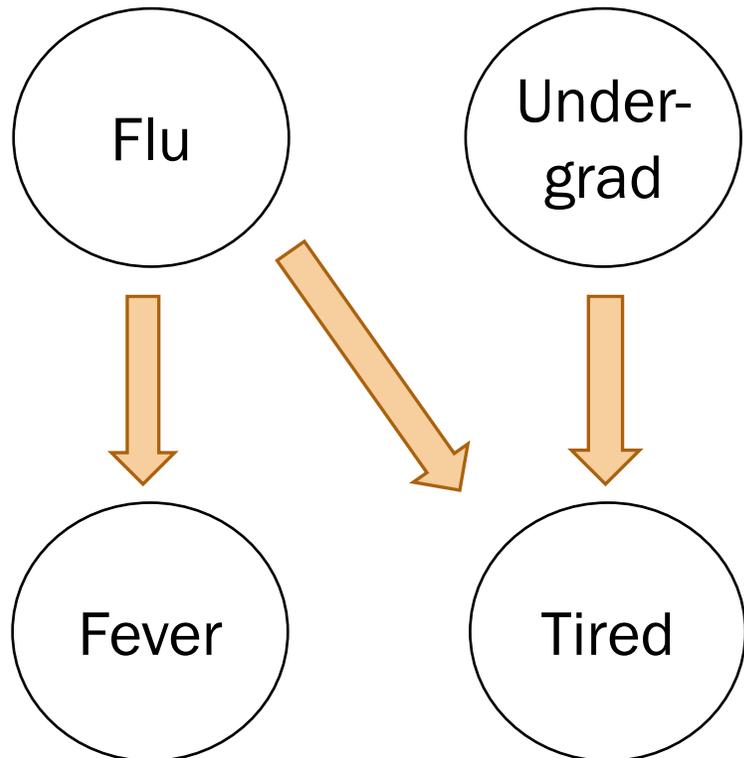
Constructing a Bayesian Network



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

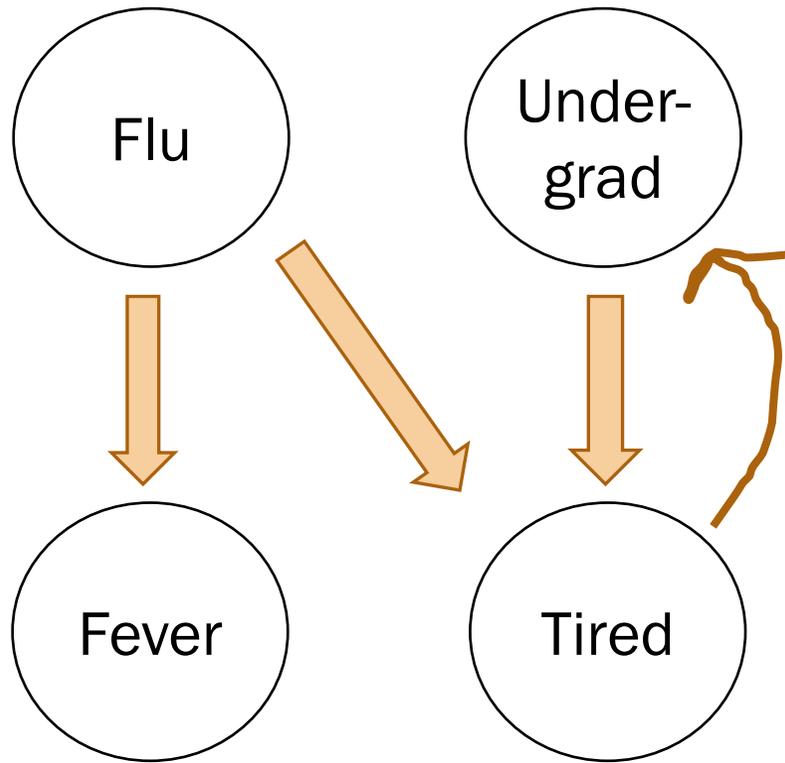


This model assumes that Flu and being an Undergraduate are independent.

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

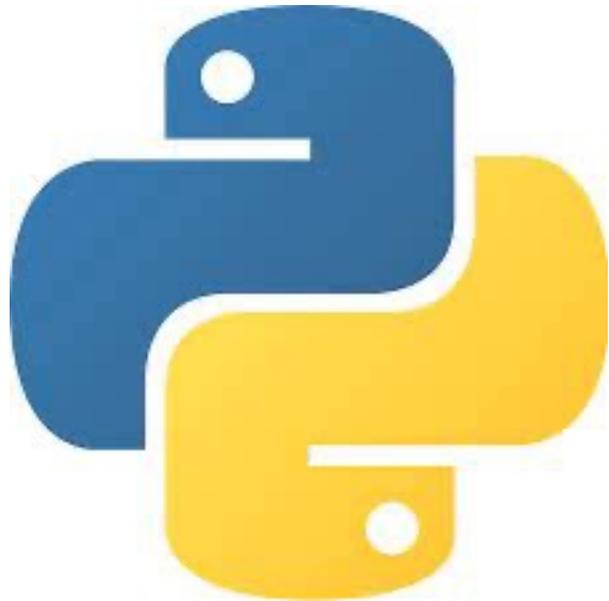
You need to tell a generative story. You do **not** need to be able to reason about all the implied independencies

Bug: Constructing a Bayesian Network



Must be acyclic!

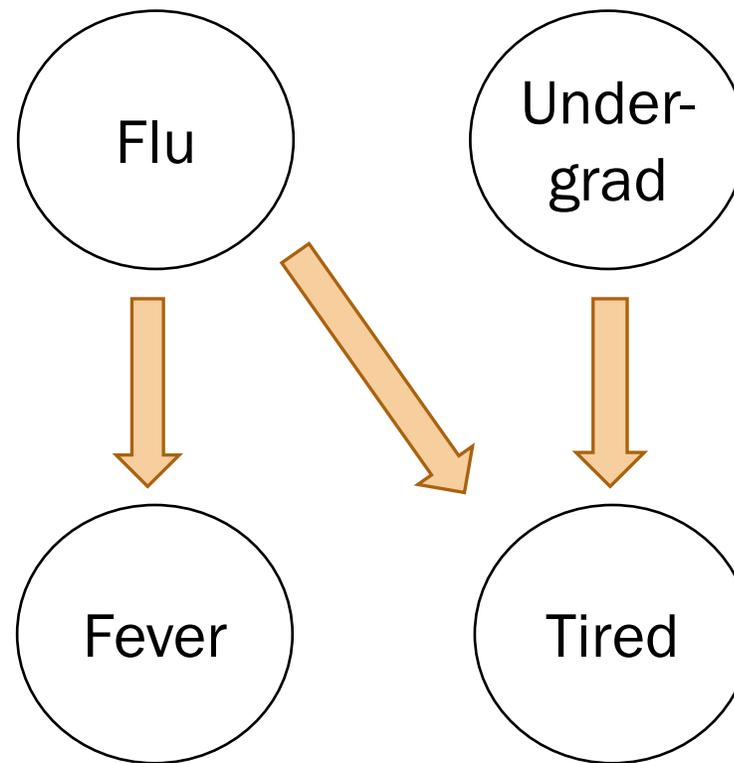
Model Version #1: Python That Outputs a **Joint** Sample



Sample Baby
Elephant

Sex: Female
Weight: 161kg

Bayesian Network Assumption:



$$\begin{aligned} P(\text{Joint}) &= P(X_1 = x_1 \dots X_n = x_n) \\ &= \prod_i P(X_i = x_i | \text{parents of } X_i) \end{aligned}$$



Bayes Nets tell a generative story.

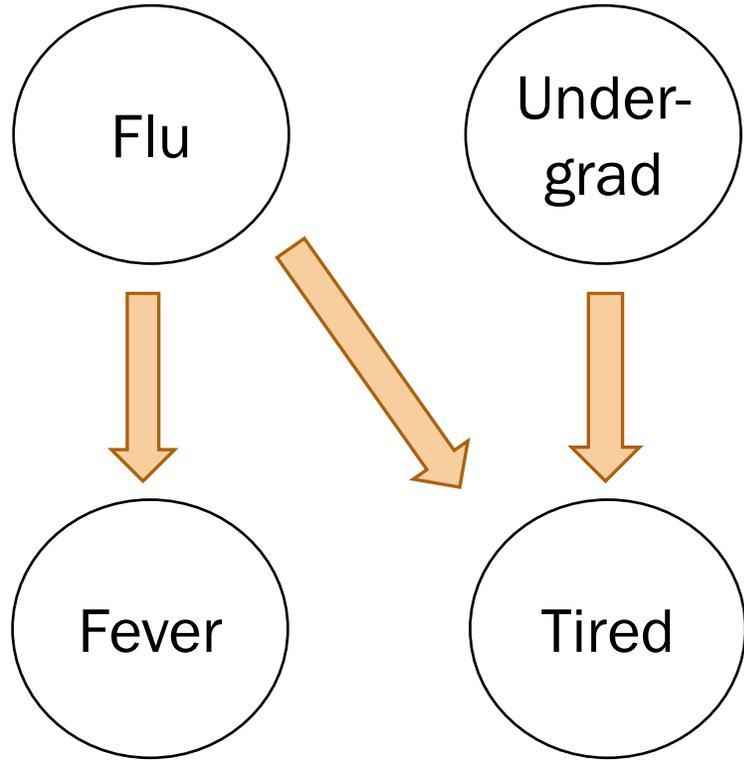
This leads to many independence assumptions

Makes it **tractable** to represent the joint

Challenge: Exact Inference in a Bayes Net

$$P(Fl = x)$$

$$P(U = u)$$



$$P(Fl = 0 | Fe = 1)$$

$$= \frac{P(Fl = 0, Fe = 1)}{P(Fe = 1)}$$

$$= \frac{P(Fe = 1 | Fl = 0)P(Fl = 0)}{\sum_i P(Fe = 1 | Fl = i)P(Fl = i)}$$

$$P(F = f | Fl = x) P(T = t | U = u, Fl = x)$$

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:

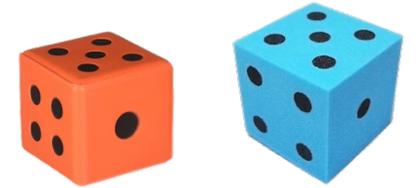
$$p(x, y) = p(x)p(y)$$

- 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**.

Dice (after all this time, still our friends)

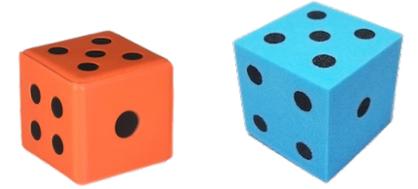
Let: D_1 and D_2 be the outcomes of two rolls
 $S = D_1 + D_2$, the sum of two rolls

- Each roll of a fair, 6-sided die is an independent trial.
 - Random variables D_1 and D_2 are independent.
1. Are events $(D_1 = 1)$ and $(S = 7)$ independent?
 2. Are events $(D_1 = 1)$ and $(S = 5)$ independent?
 3. Are random variables D_1 and S independent?



Dice (after all this time, still our friends)

Let: D_1 and D_2 be the outcomes of two rolls
 $S = D_1 + D_2$, the sum of two rolls



- Each roll of a 6-sided die is an independent trial.
- Random variables D_1 and D_2 are independent.

1. Are events $(D_1 = 1)$ and $(S = 7)$ independent? 

2. Are events $(D_1 = 1)$ and $(S = 5)$ independent? 

3. Are random variables D_1 and S independent? 

All events $(X = x, Y = y)$ must be independent for X, Y to be independent RVs.

Can I discover independence from
data?

ROCK

The Sound: Vigorous, defiant, energetic, inventive

The Roots: Rhythm & blues, country

The Pioneers: Bill Haley, Chuck Berry, Fats Domino, Little Richard, Buddy Holly, Elvis Presley

The Places: Cleveland, New Orleans, Detroit, New York City

The Ensemble: Electric guitar, bass, drums, keyboard, vocals

"We're a rock group. We're noisy, raucous, emotional and wild."

— Angus Young (c. 1960)
Lead guitarist of the band AC/DC

HIP-HOP R&B

The Sound: Rhythmic, unvarnished, adaptable, streetwise

The Roots: Rhythm & blues, soul, funk, reggae

The Pioneers: Afrika Bambaataa, Kool Herc, DJ Hollywood, Grandmaster Flash, Kurtis Blow, Grandmaster Caz

The Places: New York City (South Bronx)

The Ensemble: Vinyl, turntable, vocals

"The beautiful thing about hip-hop is it's like an audio collage. You can take any form of music and do it in a hip-hop way and it'll be a hip-hop song."

— Tom Mchale (1971)
Hip-hop artist

LATIN American

The Sound: Syncopated, enthusiastic, diverse, vibrant

The Roots: Spain, Africa, Caribbean, South America

The Pioneers: Arsenio Rodriguez, Machito, Pérez Prado, Tito Puente, Celia Cruz, Johnny Pacheco

The Places: Cuba, Puerto Rico, Mexico, Miami, New York

The Ensemble: Congas, bongos, maracas, güiro, guitar, vocals

"The emphasis was dancing and rhythm. I came in with an emphasis on lyrics... telling stories that were familiar to people in Latin America—and everybody identified with the songs."

— Rubén Blades (c. 1960)
Salsa singer and composer

Folk

The Sound: Grassroots, narrative, sincere, lyrical

The Roots: Ballads, immigrant folklore, spirituals, cowboy songs

The Pioneers: Lead Belly, Odetta, Woody Guthrie, Pete Seeger, Bob Dylan, Joan Baez

The Places: Appalachia, Deep South, Western frontier

The Ensemble: Guitar, banjo, fiddle, accordion, vocals

"I find the rhythms [of folk music]. I find the melodies, time-tested by generations of singers. Above all, I find the words... they seemed punchy, straightforward, honest."

— Peter Dinklage (c. 1960)
Folk musician

COUNTRY Western

The Sound: Genuine, uncomplicated, nostalgic, informal

The Roots: European ballads, folk and gospel songs

The Pioneers: Uncle Dave Macon, the Carter Family, Jimmie Rodgers, Roy Acuff, Gene Autry, Bill Monroe

The Places: Appalachia, Nashville, Chicago, Western U.S.

The Ensemble: Fiddle, banjo, guitar, harmonica, accordion, vocals

"Country music is three chords and the truth."

— Hank Williams (1917–1953)
Country music singer

CLASSICAL

The Sound: Intricate, polished, structured, harmonious

The Roots: Sacred music, choral chants, madrigals, dance rhythms

The Pioneers: J.S. Bach, Handel, Haydn, Mozart, Beethoven, Brahms

The Places: Austria, Germany, France, Italy

The Ensemble: Strings, woodwinds, brass, percussion, vocals

"I carry my thoughts about with me a long time... before writing them down. I change many things, discard others, and try again and again until I am satisfied."

— Ludwig van Beethoven (1770–1827)
Classical music composer

Home Insert Page Layout Formulas Data >> Share >

Clipboard Font Alignment Number Conditional Formatting > Format as Table > Cell Styles > Cells Editing

C15 fx 3

	A	B	C	D	E	F	G	H	I
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	

Ready music + 100%

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

	A	B	C	D	E	F	G	H	Me
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	

Dance of the Covariance

Recall our Ebola Bats



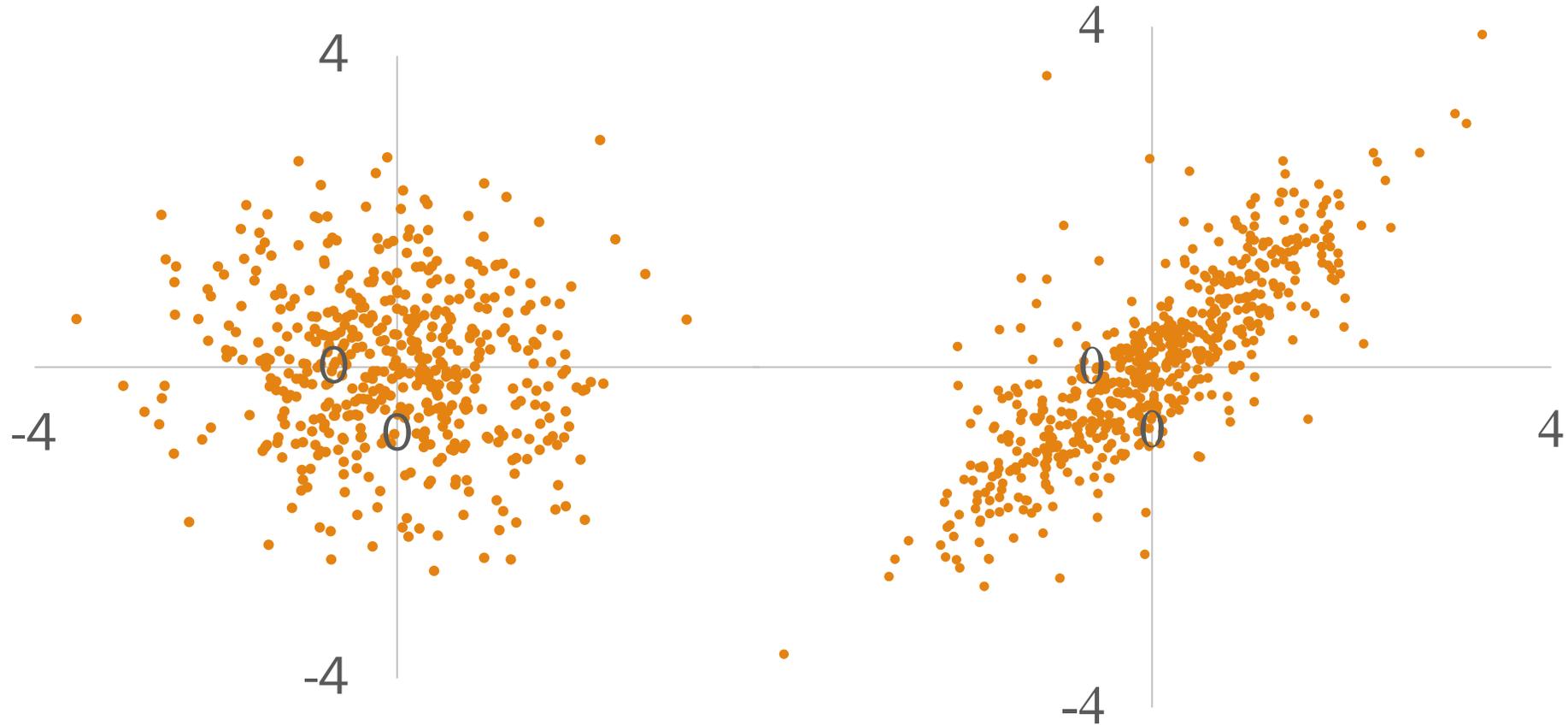
Bat Data

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	...	FALSE	FALSE
TRUE	FALSE	FALSE	TRUE	FALSE	FALSE

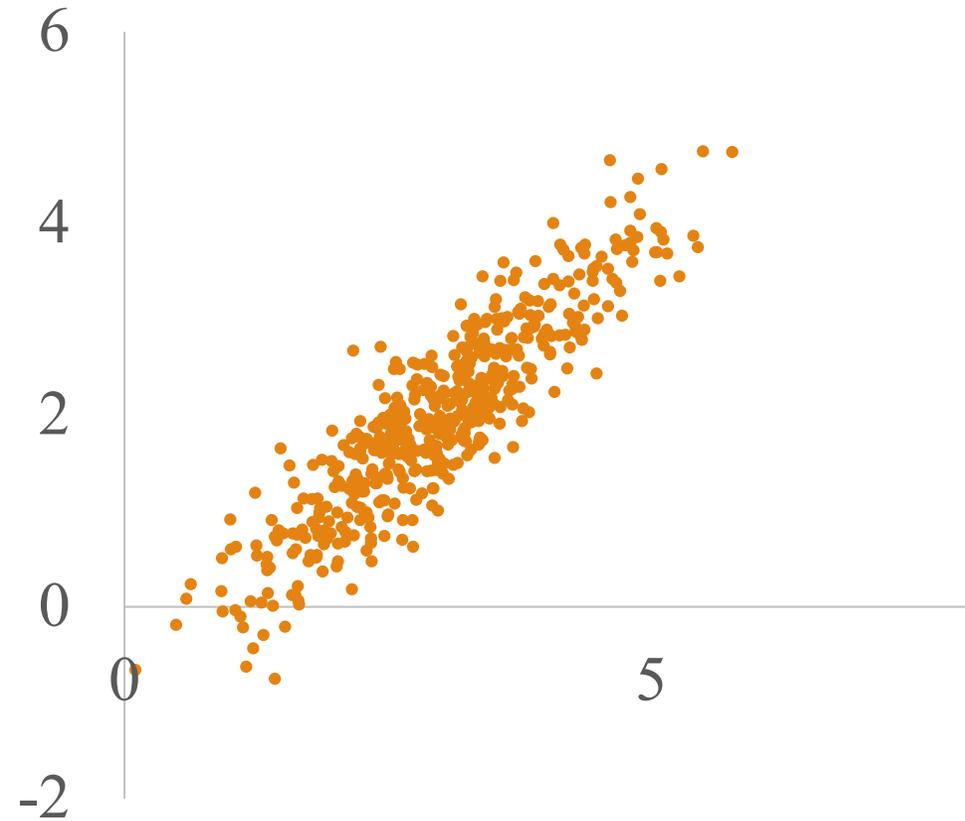
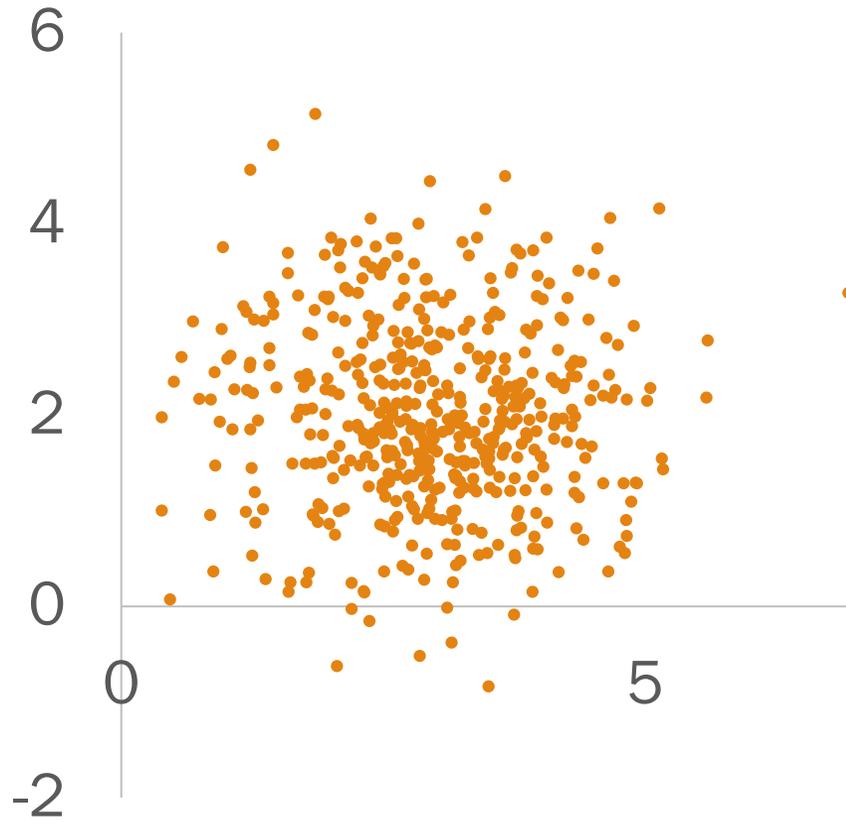
Expression Amount

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

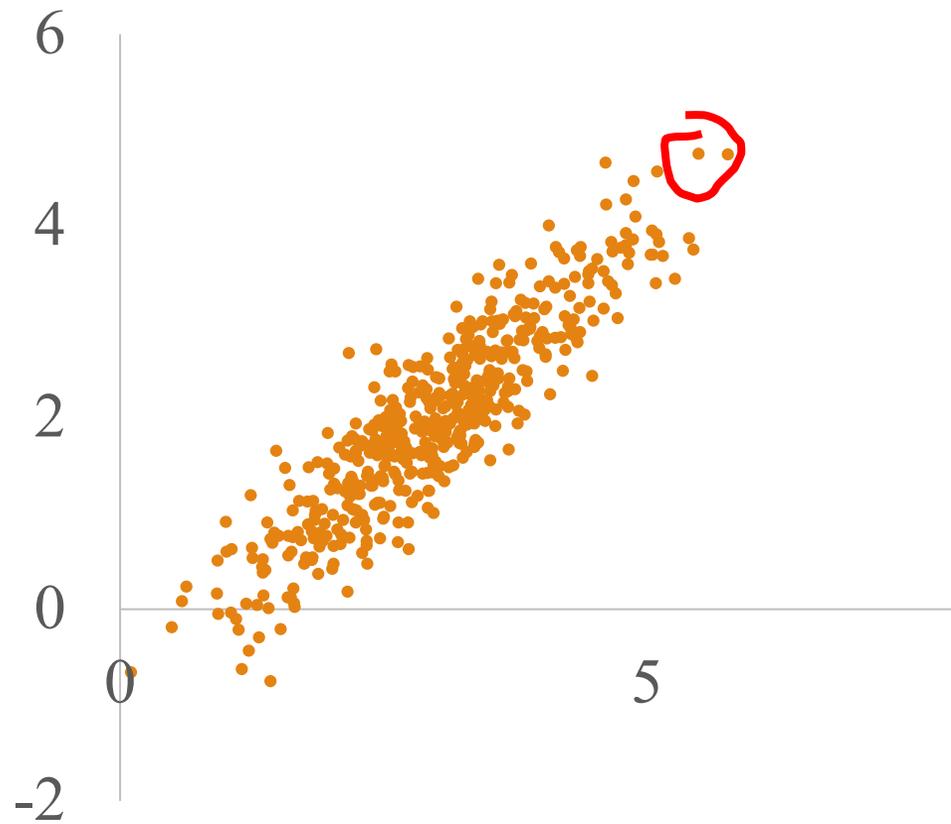
Spot The Difference



Spot The Difference



Vary Together

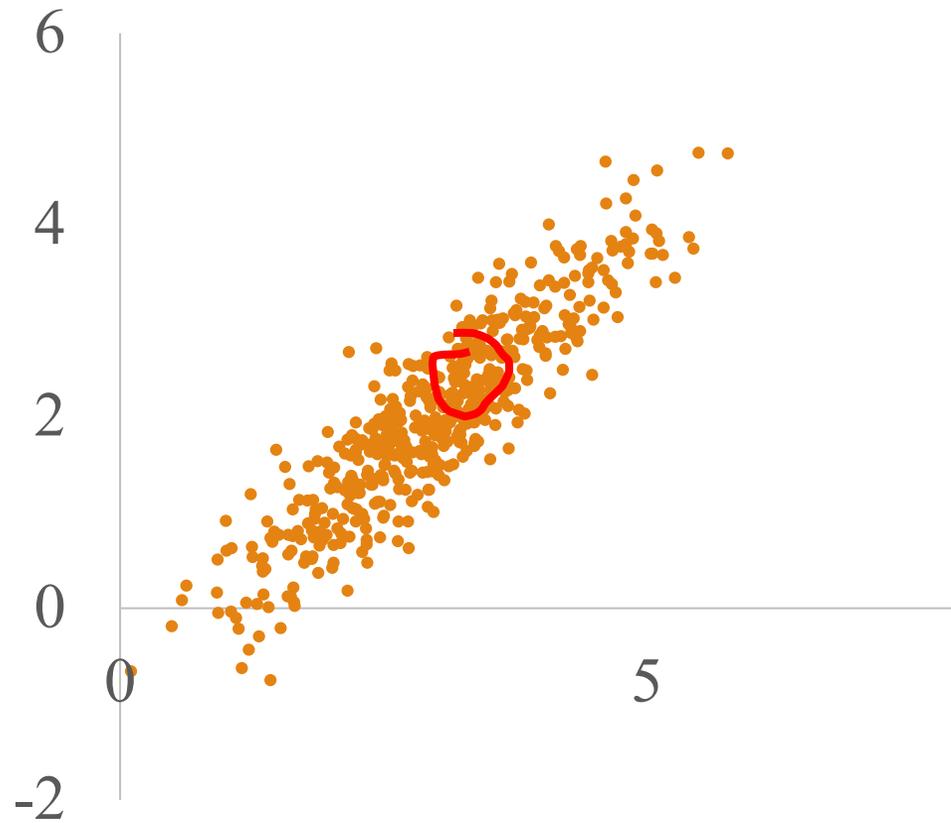


$$x - E[x] = 3$$

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

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

Vary Together

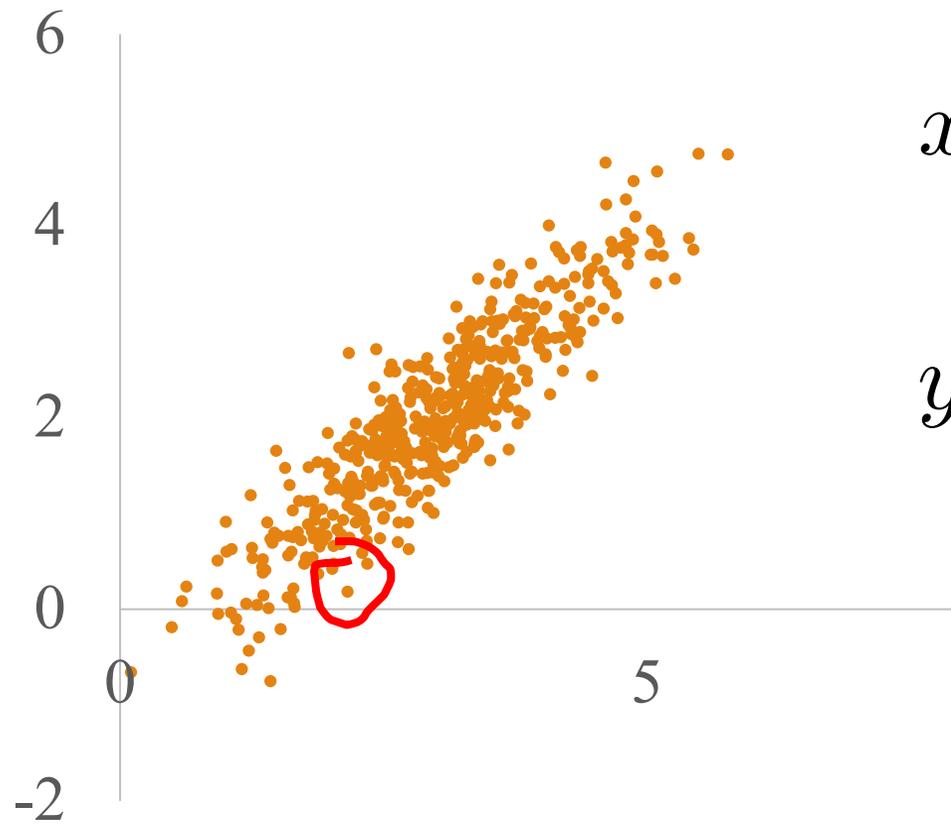


$$x - E[x] \approx 0$$

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

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

Vary Together

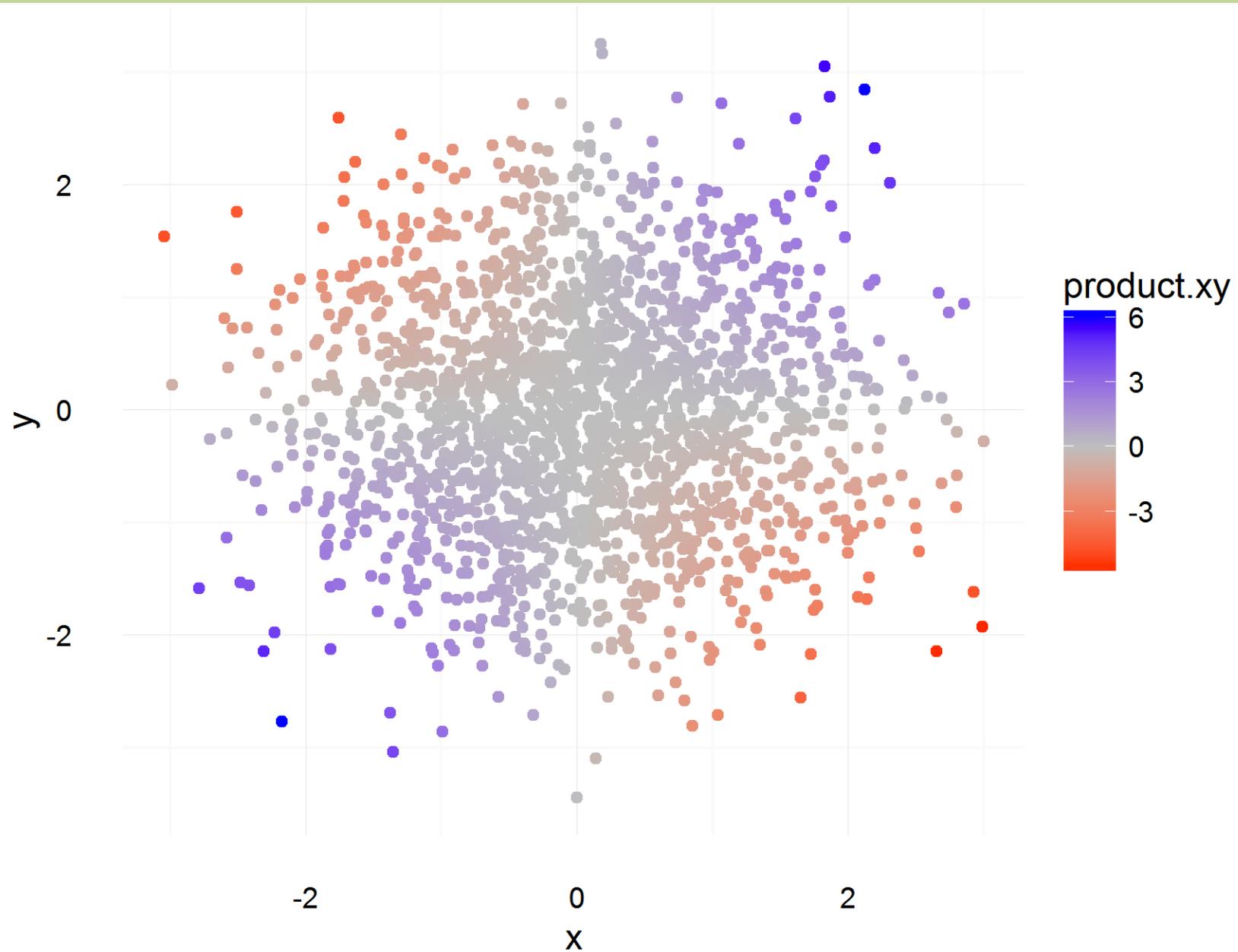


$$x - E[x] = -1.1$$

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

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

Understanding Covariance



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, $E[XY] = E[X]E[Y] \rightarrow \text{Cov}(X, Y) = 0$
- But $\text{Cov}(X, Y) = 0$ does **not** imply X and Y independent!

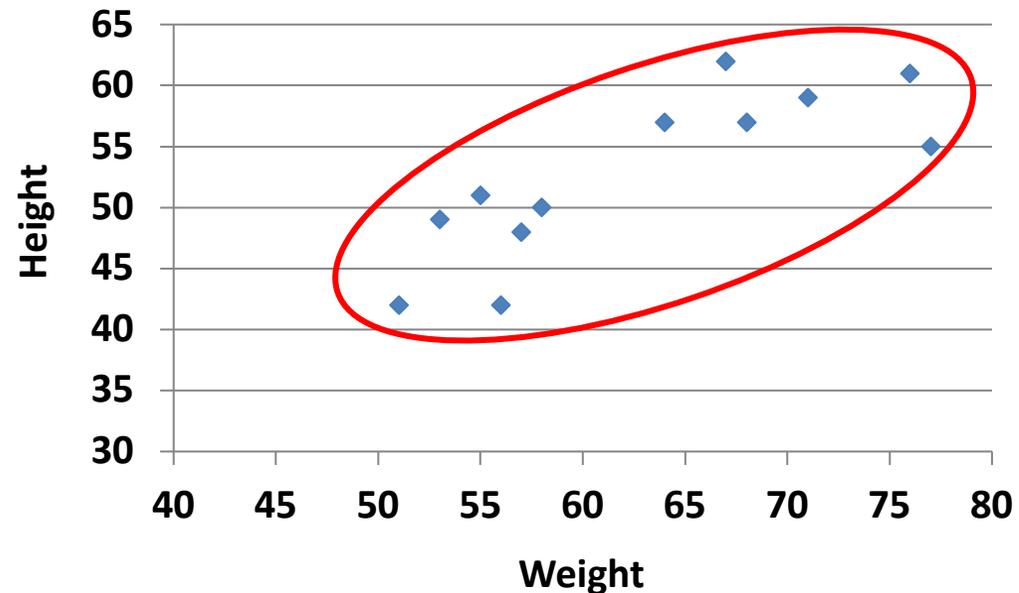
Covariance and Data

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

$$\begin{aligned} E[W] &= 62.75 \\ E[H] &= 52.75 \end{aligned}$$

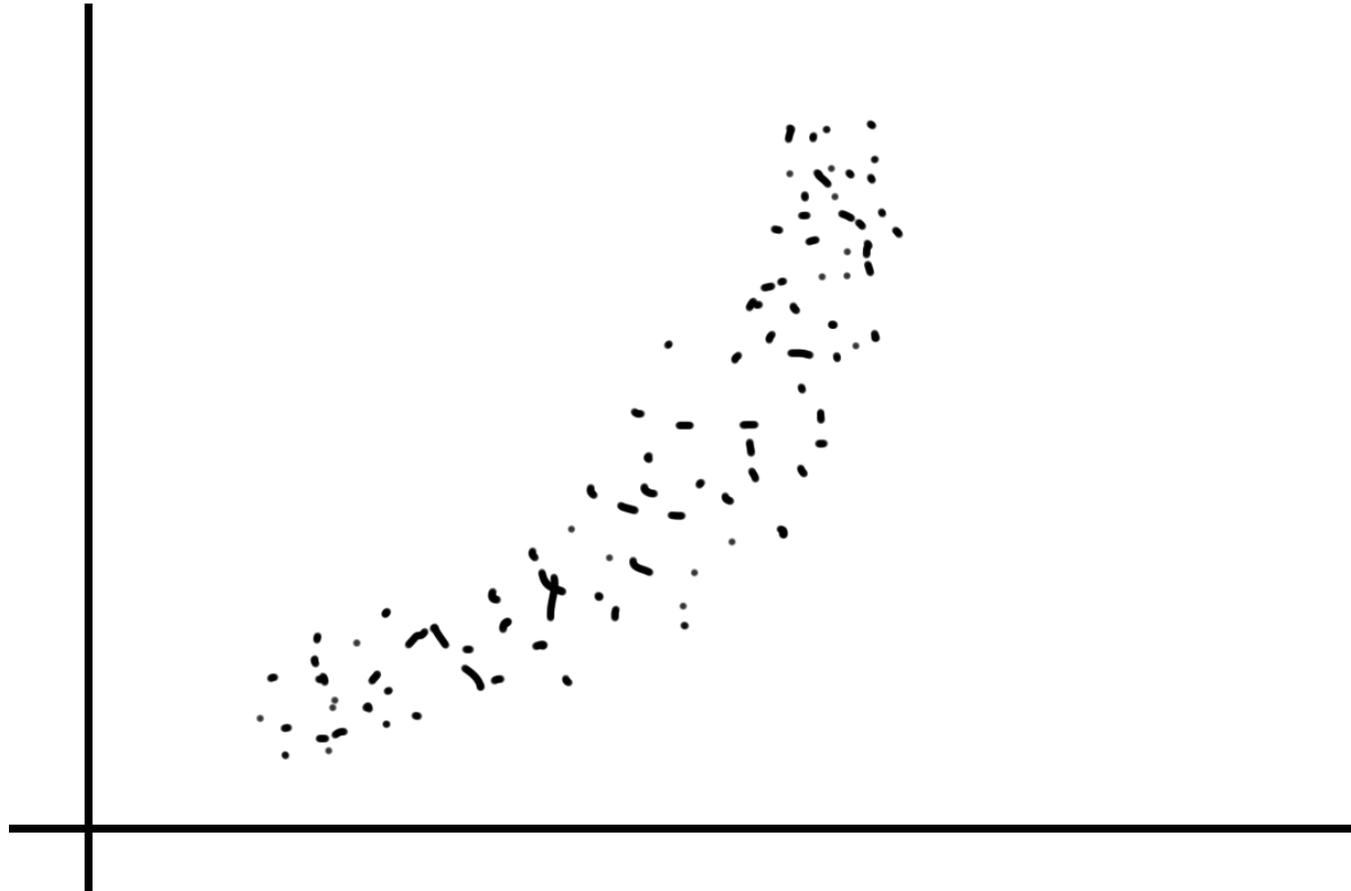
$$E[W*H] = 3355.83$$



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

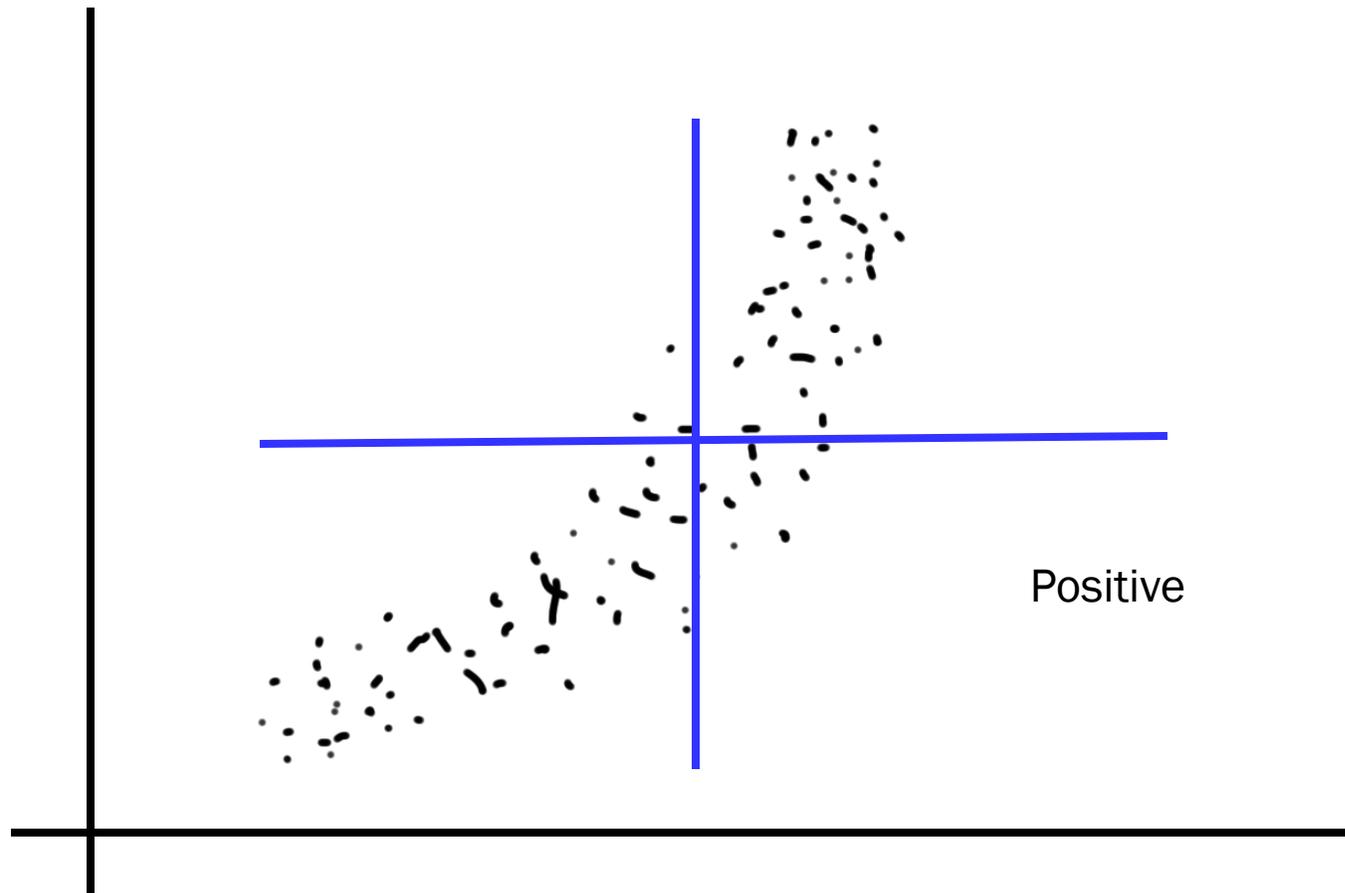
Covariance

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



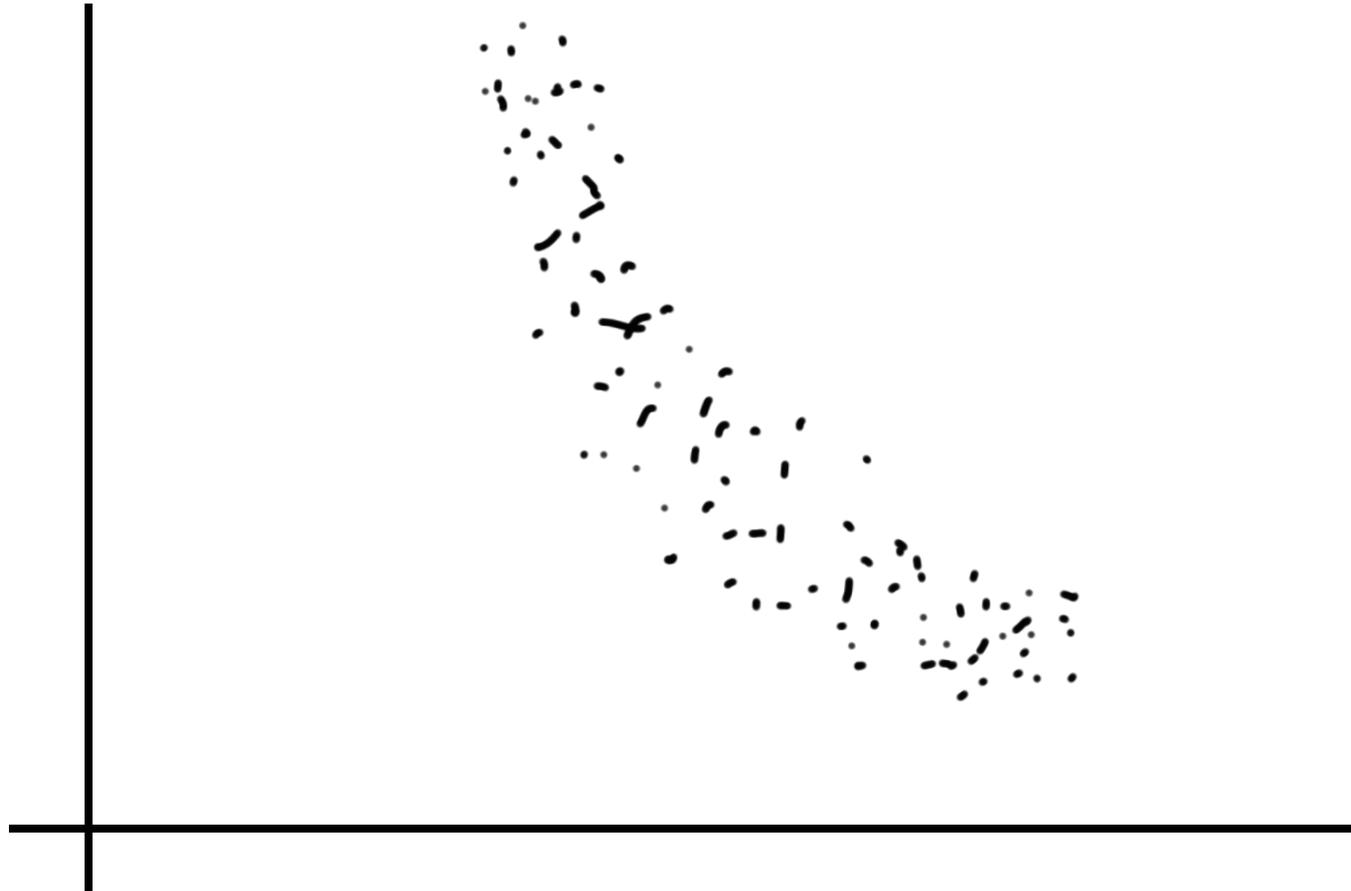
Covariance

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



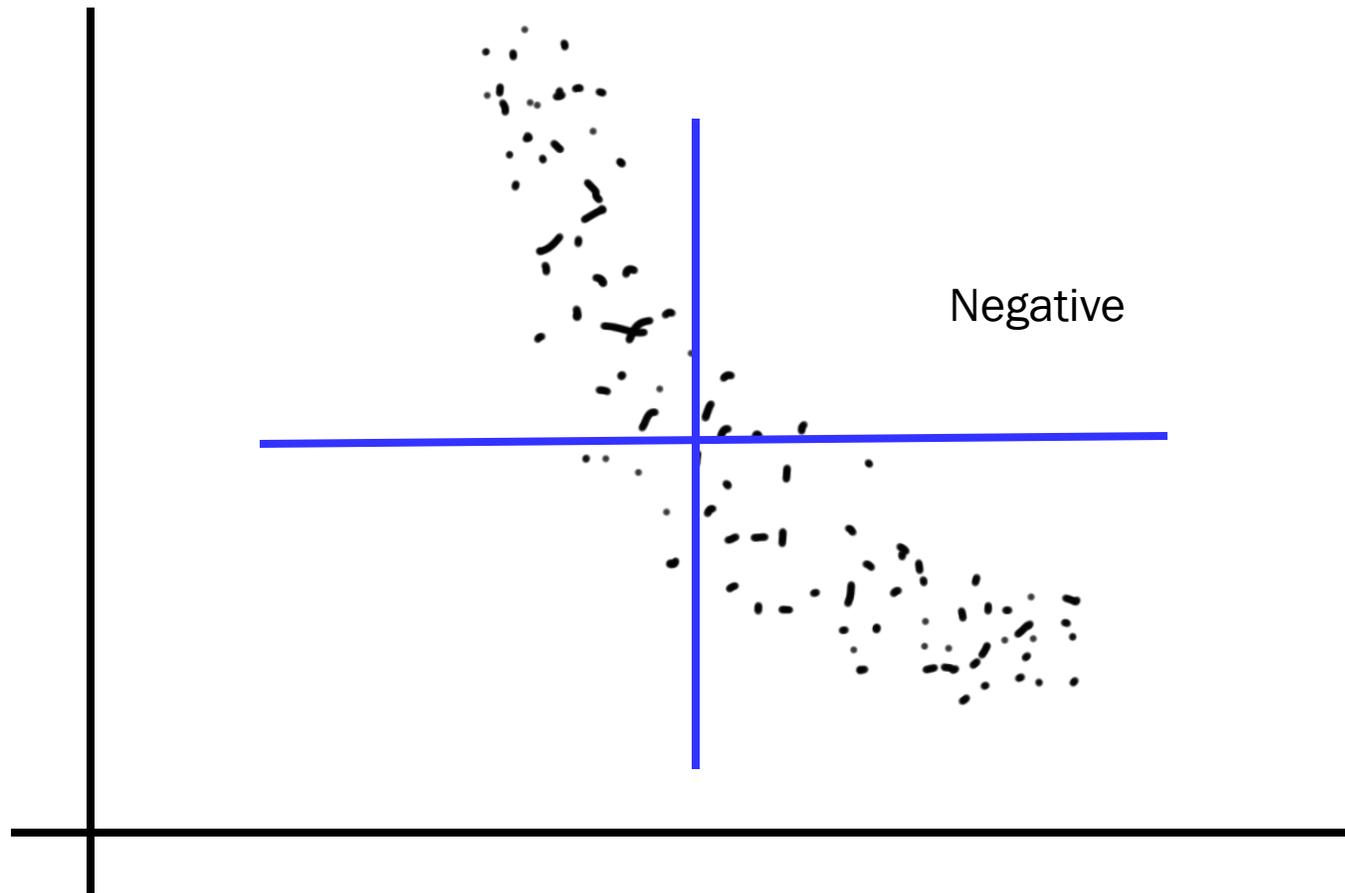
Covariance

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



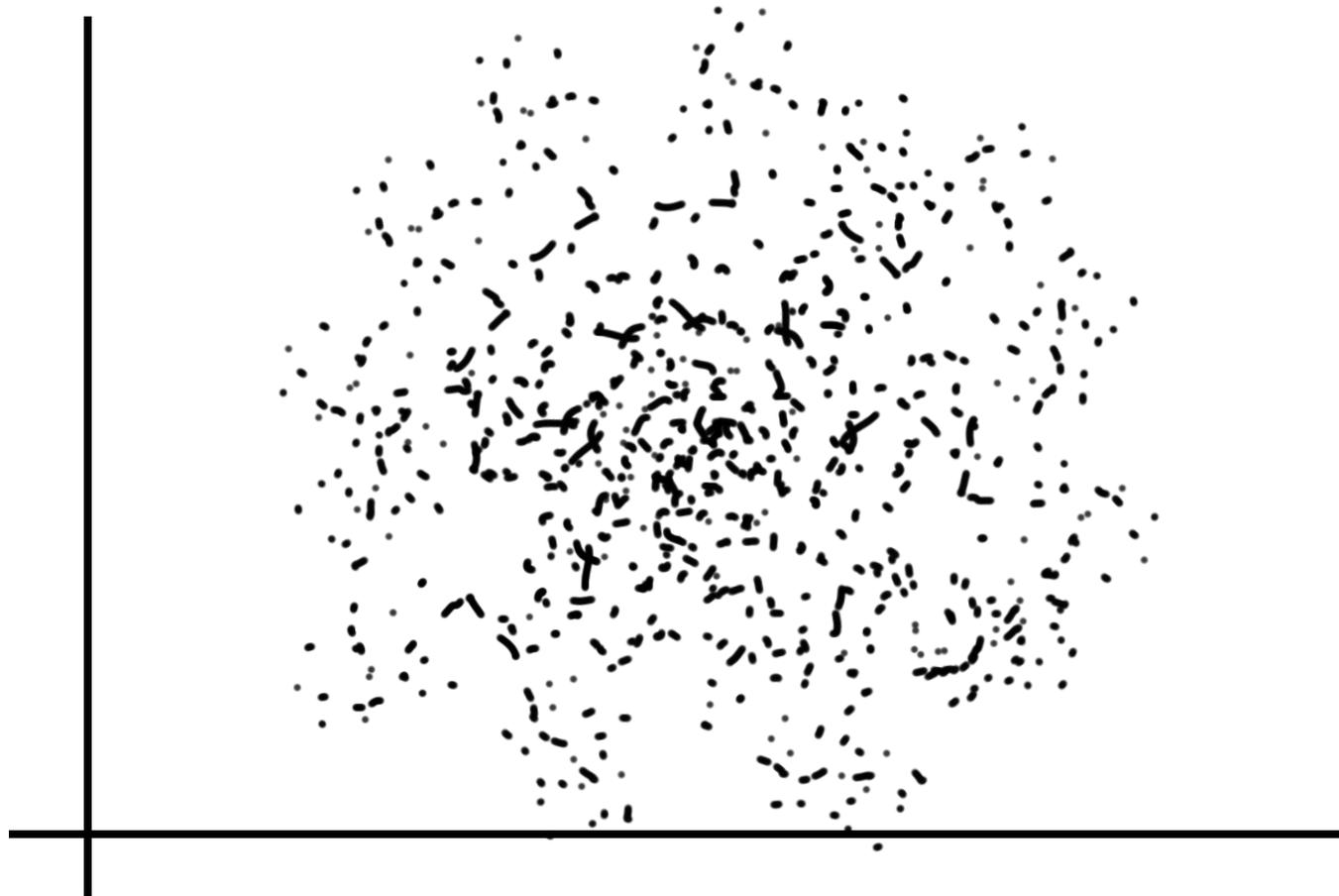
Covariance

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



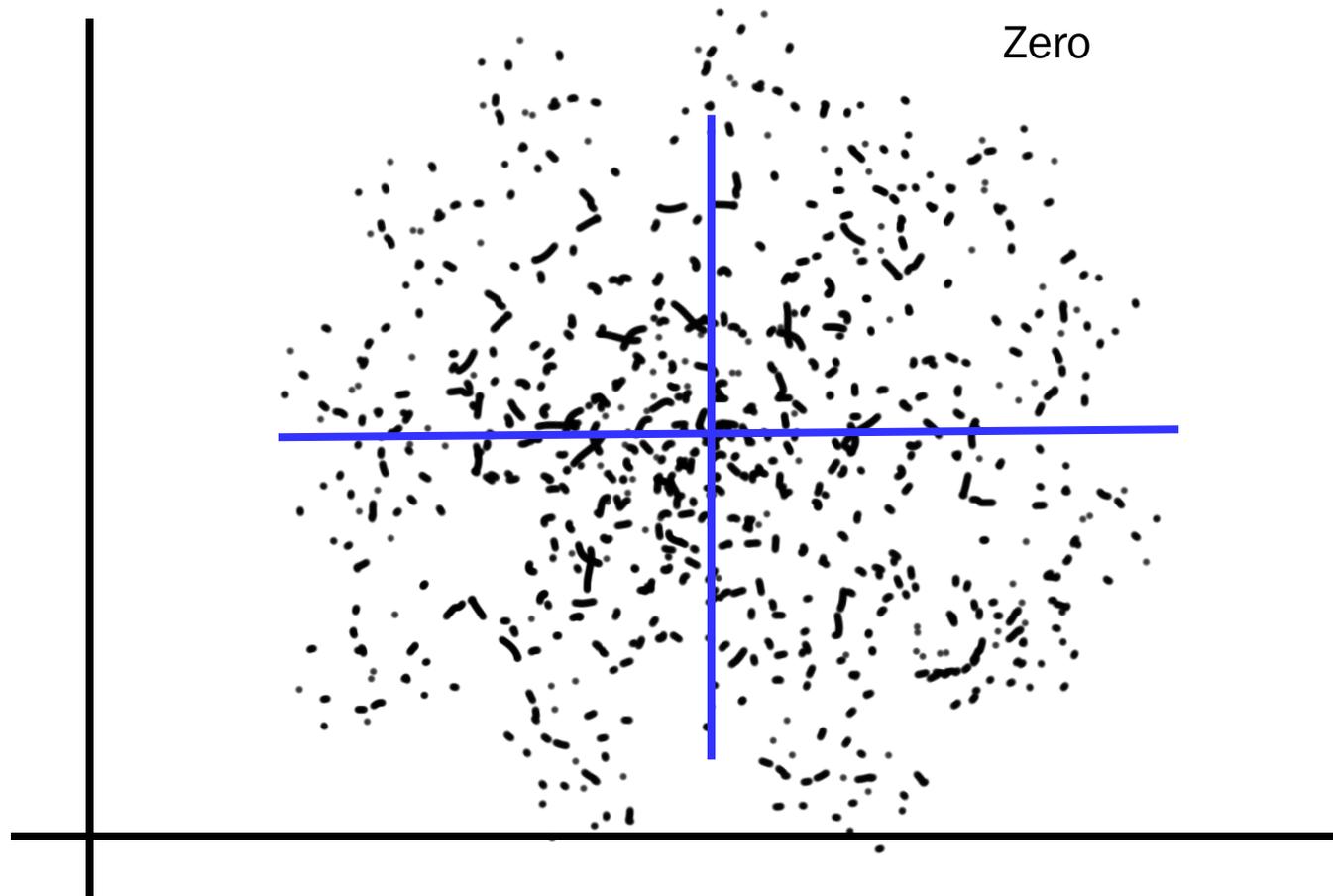
Covariance

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:

	X	-1	0	1	$p_Y(y)$
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 $XY = 0$, $E[XY] = 0$
- $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0 - 0 = 0$

But, X and Y are clearly dependent!

Properties of Covariance

Say X and Y are arbitrary random variables

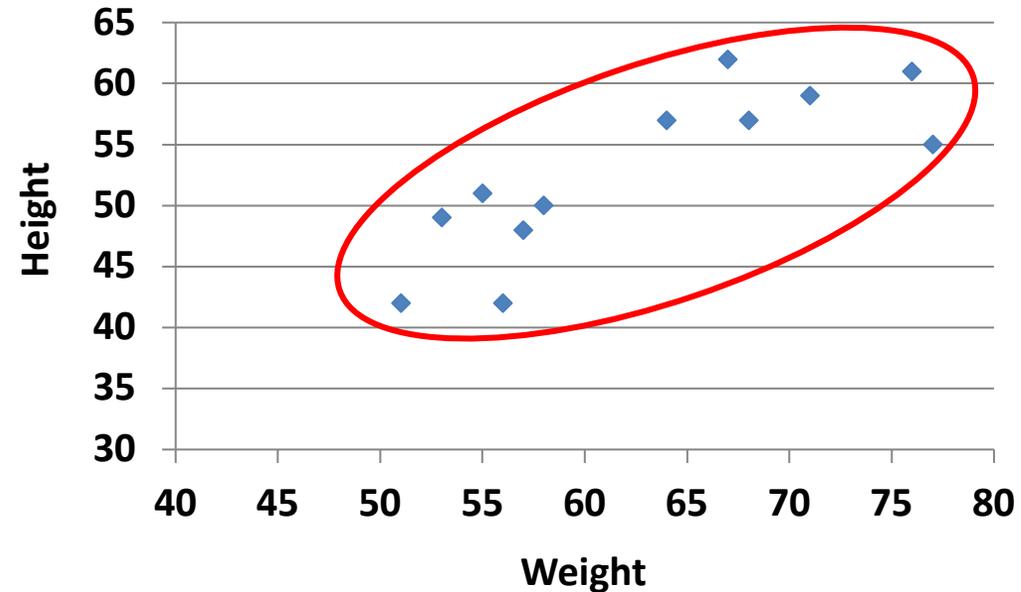
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = E[X^2] - E[X]E[X] = \text{Var}(X)$
- $\text{Cov}(aX + b, Y) = a\text{Cov}(X, Y)$

Correlation

What is Wrong With This?

Consider the following data:

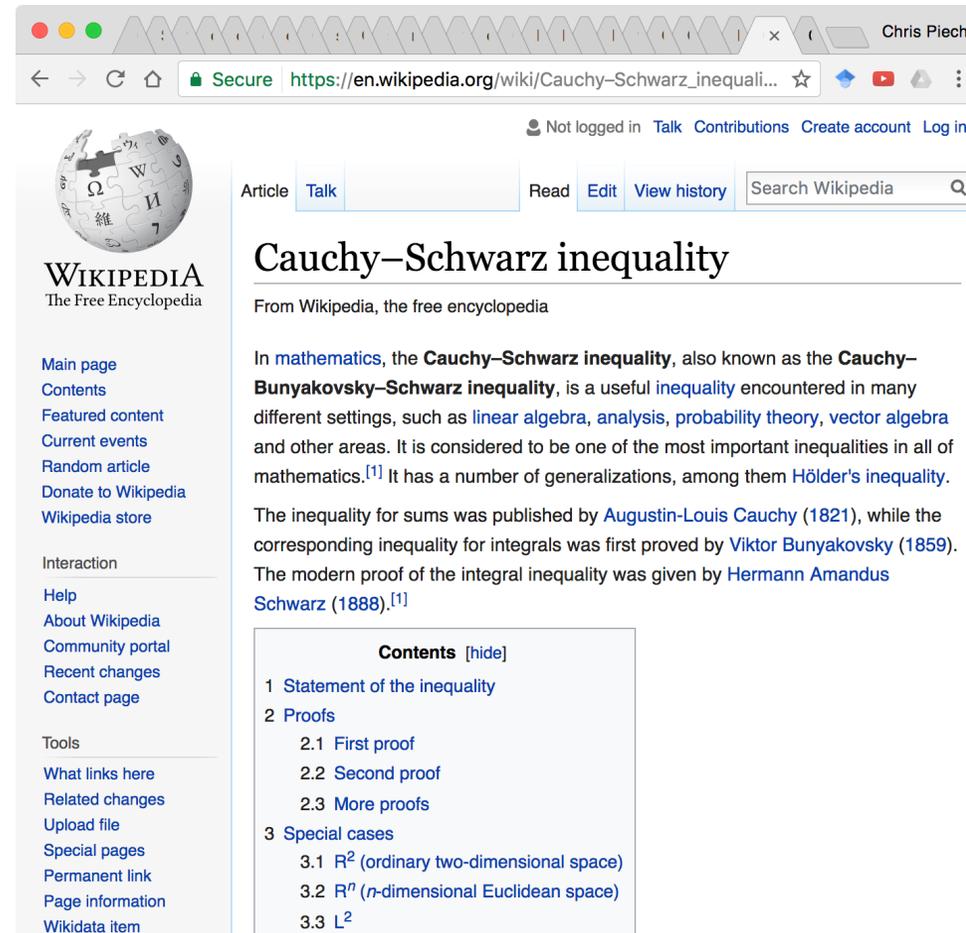
Weight	Height	Weight * Height
64	57	3648
71	59	4189
53	49	2597
67	62	4154
55	51	2805
58	50	2900
77	55	4235
57	48	2736
56	42	2352
51	42	2142
76	61	4636
68	57	3876



$$\begin{array}{lll} E[W] & E[H] & E[W*H] \\ = 62.75 & = 52.75 & = 3355.83 \end{array}$$

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

Cauchy Schwarz, a great way to normalize!



The screenshot shows a web browser window displaying the Wikipedia article for "Cauchy–Schwarz inequality". The browser's address bar shows the URL "https://en.wikipedia.org/wiki/Cauchy–Schwarz_inequali...". The page title is "Cauchy–Schwarz inequality". The article text states: "In **mathematics**, the **Cauchy–Schwarz inequality**, also known as the **Cauchy–Bunyakovsky–Schwarz inequality**, is a useful **inequality** encountered in many different settings, such as **linear algebra**, **analysis**, **probability theory**, **vector algebra** and other areas. It is considered to be one of the most important inequalities in all of mathematics.^[1] It has a number of generalizations, among them **Hölder's inequality**. The inequality for sums was published by **Augustin-Louis Cauchy** (1821), while the corresponding inequality for integrals was first proved by **Viktor Bunyakovsky** (1859). The modern proof of the integral inequality was given by **Hermann Amandus Schwarz** (1888).^[1]" Below the text is a "Contents" section with the following items: 1 Statement of the inequality, 2 Proofs (with sub-items 2.1 First proof, 2.2 Second proof, 2.3 More proofs), and 3 Special cases (with sub-items 3.1 R² (ordinary two-dimensional space), 3.2 Rⁿ (n-dimensional Euclidean space), 3.3 L²).

$$-\text{Std}(X)\text{Std}(Y) \leq \text{Cov}(X, Y) \leq \text{Std}(X)\text{Std}(Y)$$

Viva La Correlación

Say X and Y are arbitrary random variables

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

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

- Note: $-1 \leq \rho(X, Y) \leq 1$
- Correlation measures linearity between X and Y
- $\rho(X, Y) = 1 \quad \Rightarrow \quad Y = aX + b \quad \text{where } a = \sigma_y/\sigma_x$
- $\rho(X, Y) = -1 \quad \Rightarrow \quad Y = aX + b \quad \text{where } a = -\sigma_y/\sigma_x$
- $\rho(X, Y) = 0 \quad \Rightarrow \quad \text{absence of linear relationship}$
 - 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 Correlación

Say X and Y are arbitrary random variables

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

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

Say $Y = cX$. Correlation should be 1.

ROCK

The Sound: Vigorous, defiant, energetic, inventive

The Roots: Rhythm & blues, country

The Pioneers: Bill Haley, Chuck Berry, Fats Domino, Little Richard, Buddy Holly, Elvis Presley

The Places: Cleveland, New Orleans, Detroit, New York City

The Ensemble: Electric guitar, bass, drums, keyboard, vocals

"We're a rock group. We're noisy, raucous, emotional and wild."

— Angus Young (c. 1960)
Lead guitarist of the band AC/DC

HIP-HOP R&B

The Sound: Rhythmic, unvarnished, adaptable, streetwise

The Roots: Rhythm & blues, soul, funk, reggae

The Pioneers: Afrika Bambaataa, Kool Herc, DJ Hollywood, Grandmaster Flash, Kurtis Blow, Grandmaster Caz

The Places: New York City (South Bronx)

The Ensemble: Vinyl, turntable, vocals

"The beautiful thing about hip-hop is it's like an audio collage. You can take any form of music and do it in a hip-hop way and it'll be a hip-hop song."

— Tom Mchale (1971)
Hip-hop artist

LATIN American

The Sound: Syncopated, enthusiastic, diverse, vibrant

The Roots: Spain, Africa, Caribbean, South America

The Pioneers: Arsenio Rodriguez, Machito, Pérez Prado, Tito Puente, Celia Cruz, Johnny Pacheco

The Places: Cuba, Puerto Rico, Mexico, Miami, New York

The Ensemble: Congas, bongos, maracas, güiro, guitar, vocals

"The emphasis was dancing and rhythm. I came in with an emphasis on lyrics... telling stories that were familiar to people in Latin America—and everybody identified with the songs."

— Rubén Blades (c. 1960)
Salsa singer and composer

Folk

The Sound: Grassroots, narrative, sincere, lyrical

The Roots: Ballads, immigrant folklore, spirituals, cowboy songs

The Pioneers: Lead Belly, Odetta, Woody Guthrie, Pete Seeger, Bob Dylan, Joan Baez

The Places: Appalachia, Deep South, Western frontier

The Ensemble: Guitar, banjo, fiddle, accordion, vocals

"I find the rhythms [of folk music]. I find the melodies, time-tested by generations of singers. Above all, I find the words... they seemed punchy, straightforward, honest."

— Peter Dinklage (c. 1960)
Folk musician

COUNTRY Western

The Sound: Genuine, uncomplicated, nostalgic, informal

The Roots: European ballads, folk and gospel songs

The Pioneers: Uncle Dave Macon, the Carter Family, Jimmie Rodgers, Roy Acuff, Gene Autry, Bill Monroe

The Places: Appalachia, Nashville, Chicago, Western U.S.

The Ensemble: Fiddle, banjo, guitar, harmonica, accordion, vocals

"Country music is three chords and the truth."

— Hank Williams (1917–1953)
Country music singer

CLASSICAL

The Sound: Intricate, polished, structured, harmonious

The Roots: Sacred music, choral chants, madrigals, dance rhythms

The Pioneers: J.S. Bach, Handel, Haydn, Mozart, Beethoven, Brahms

The Places: Austria, Germany, France, Italy

The Ensemble: Strings, woodwinds, brass, percussion, vocals

"I carry my thoughts about with me a long time... before writing them down. I change many things, discard others, and try again and again until I am satisfied."

— Ludwig van Beethoven (1770–1827)
Classical music composer

AutoSave OFF Search Sheet

Home Insert Page Layout Formulas Data >> Share

Clipboard Font Alignment Number Conditional Formatting Format as Table Cell Styles

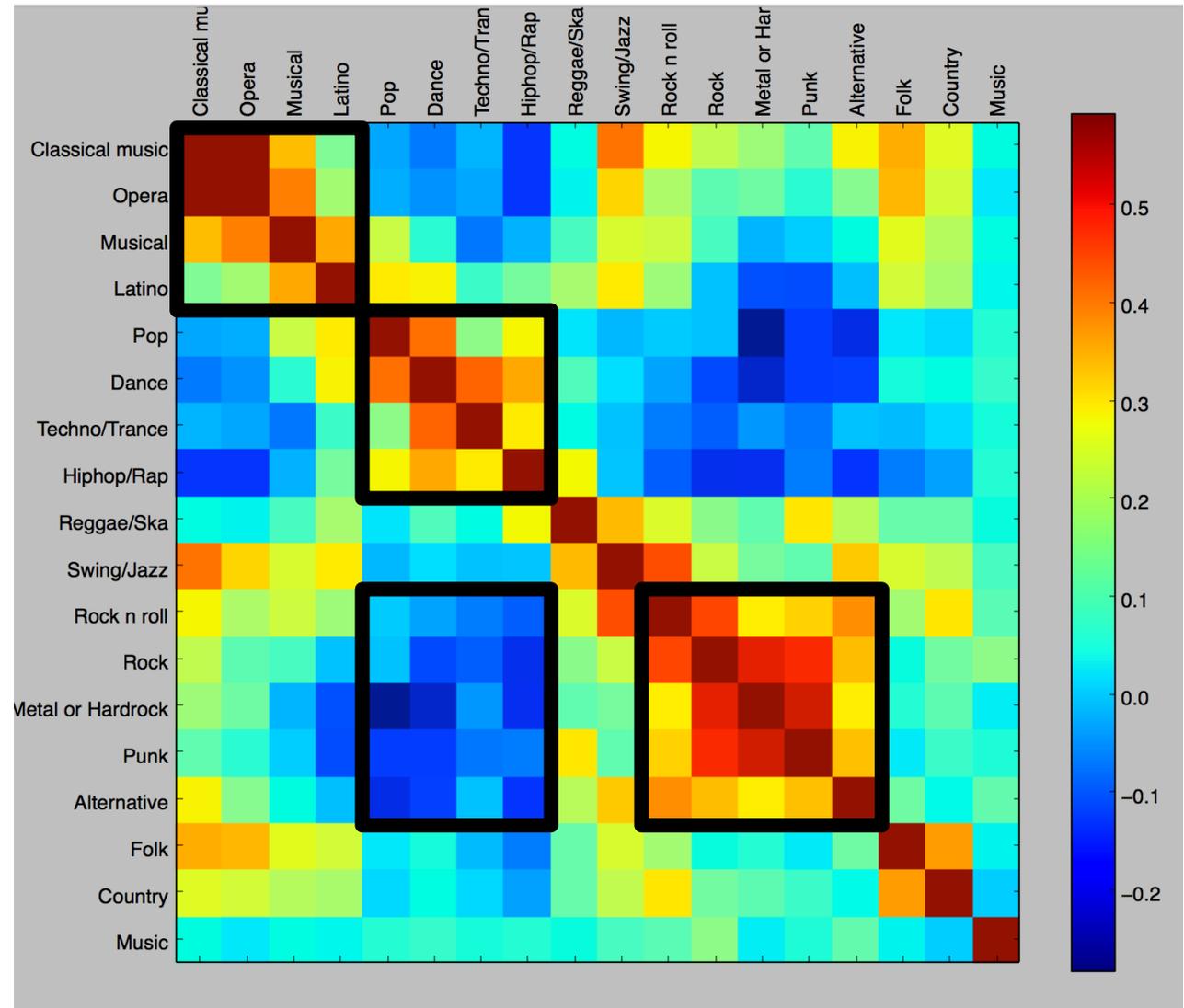
C15 fx 3

	A	B	C	D	E	F	G	H	I
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	

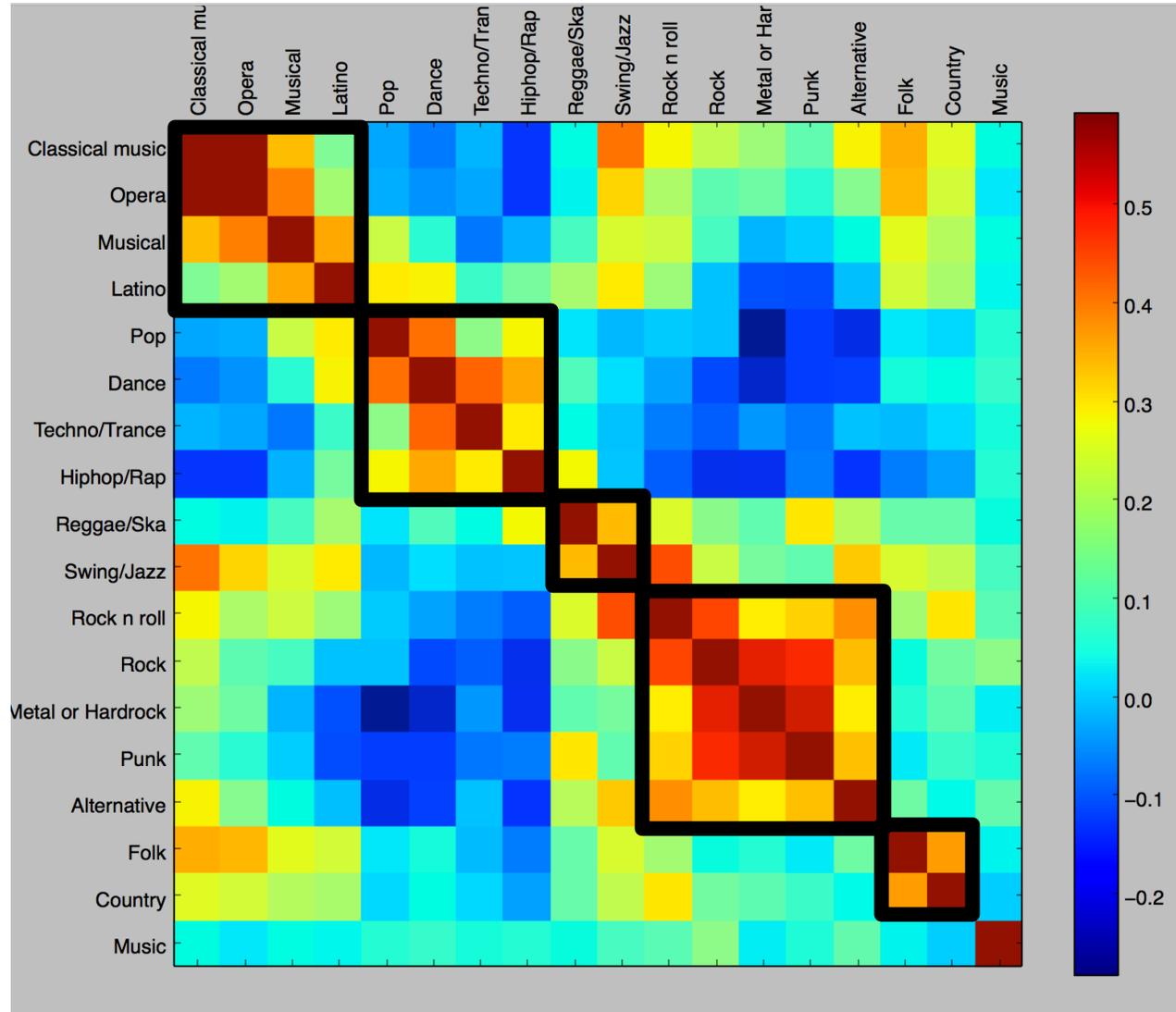
music +

Ready 100%

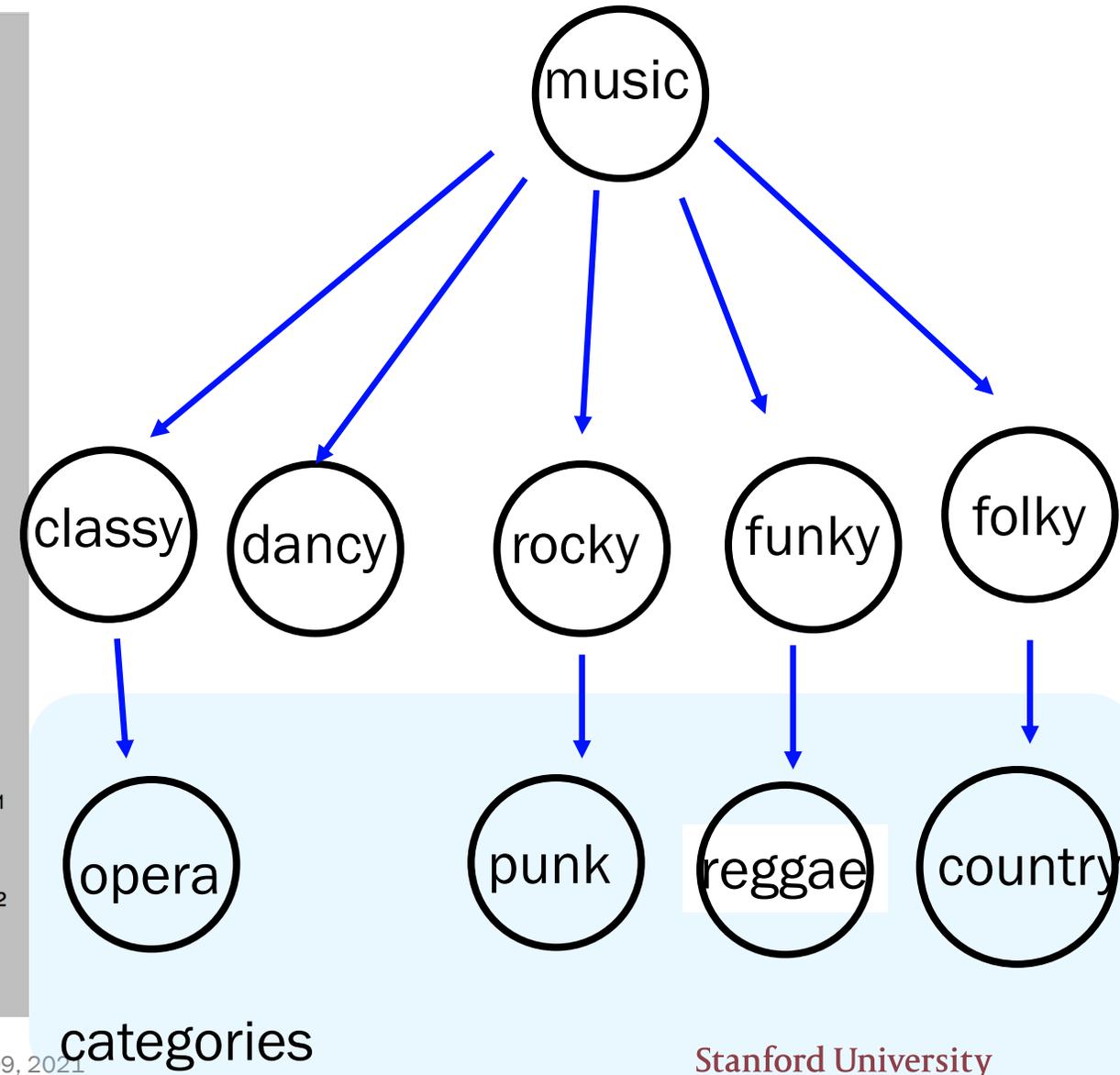
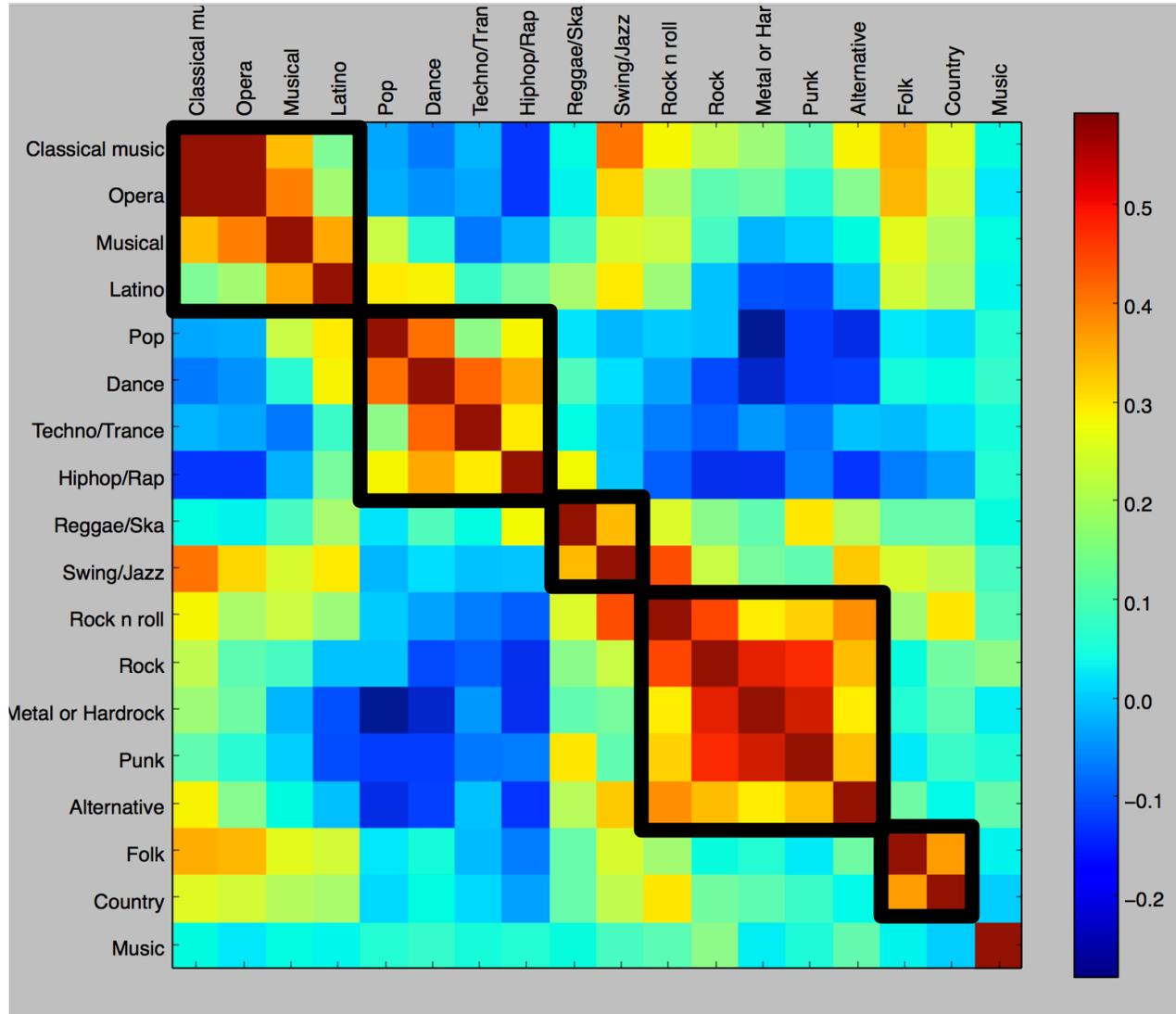
Correlation of Music Tastes



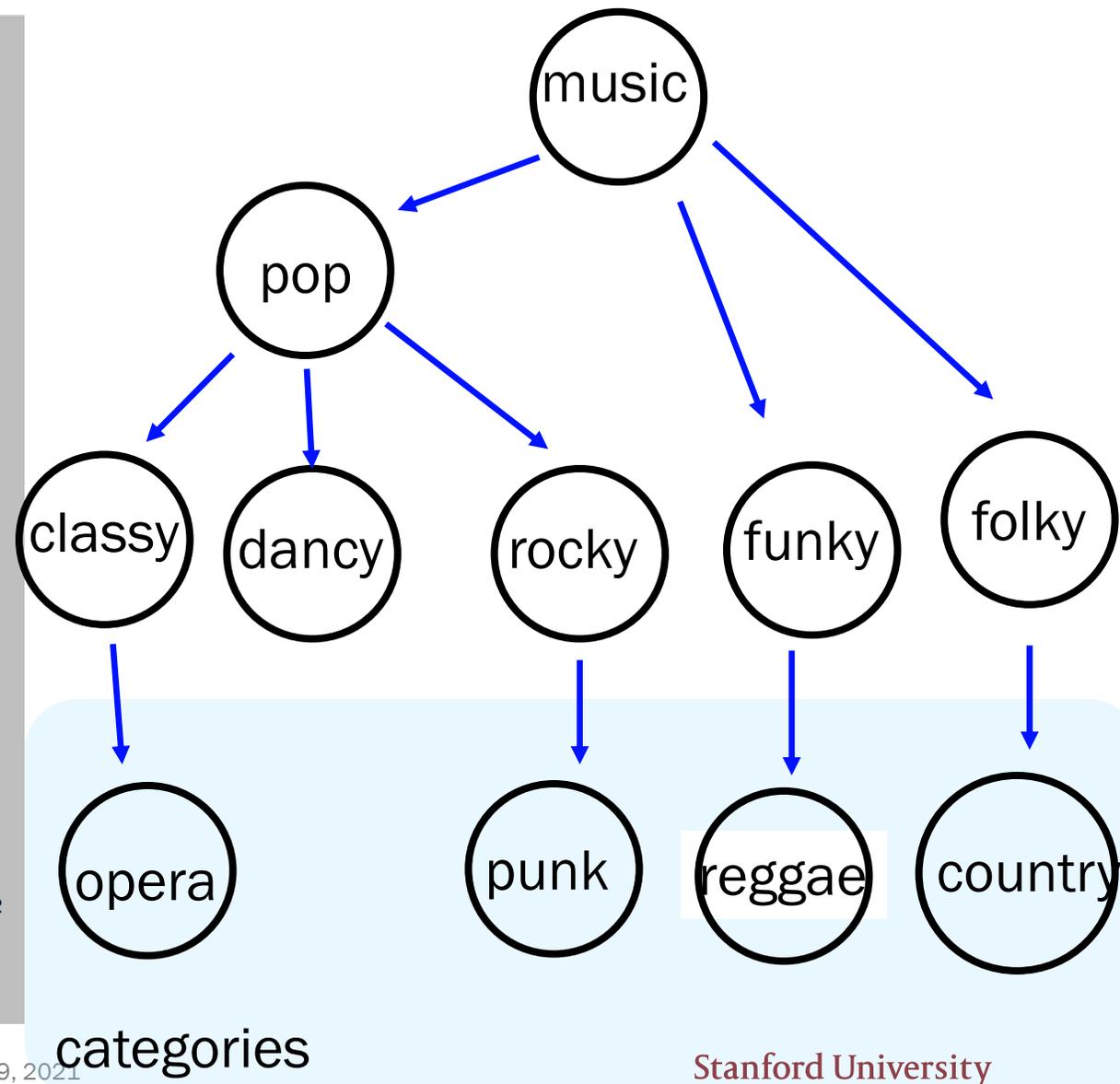
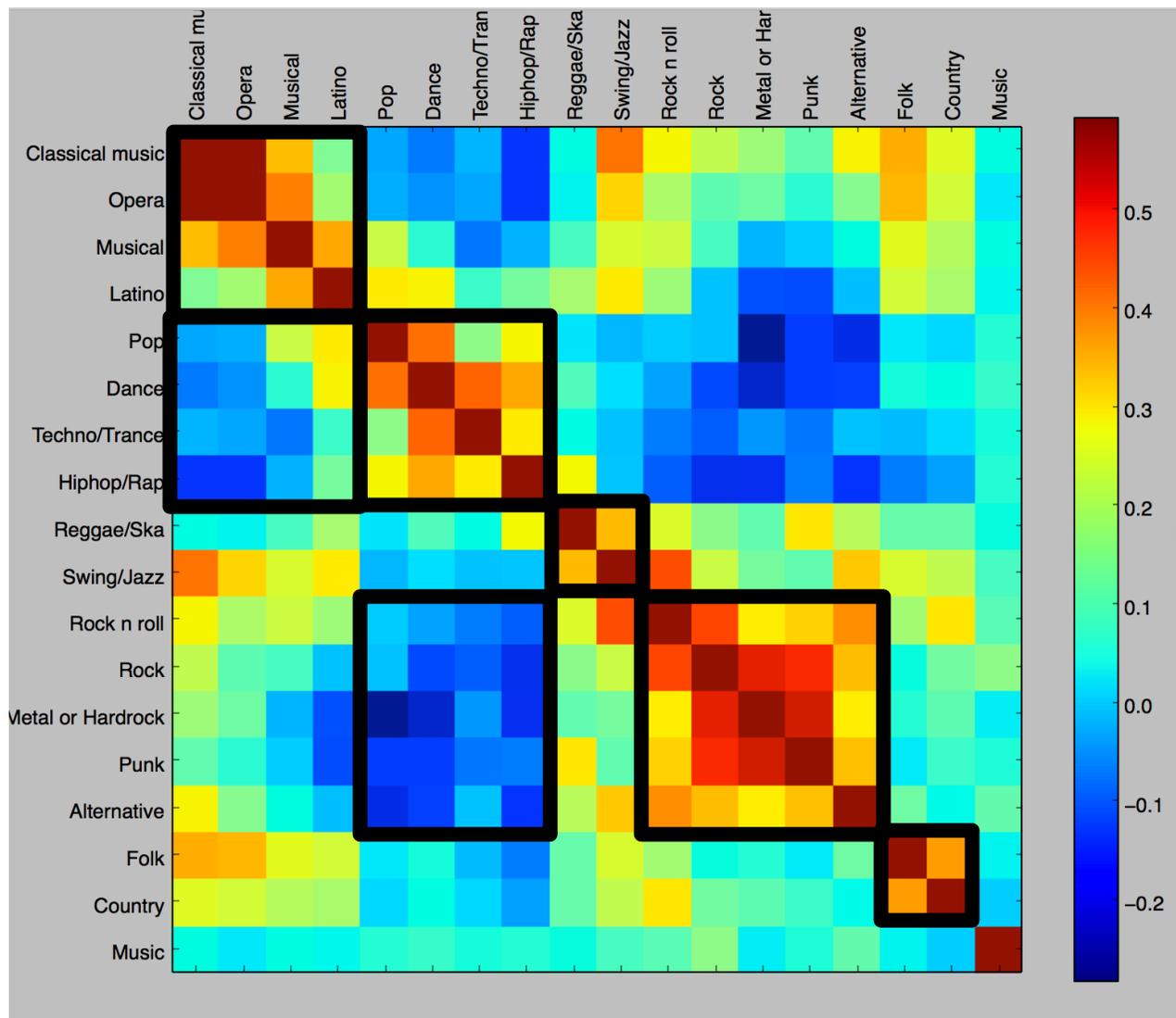
From Correlation to Bayes Net. Alternative!



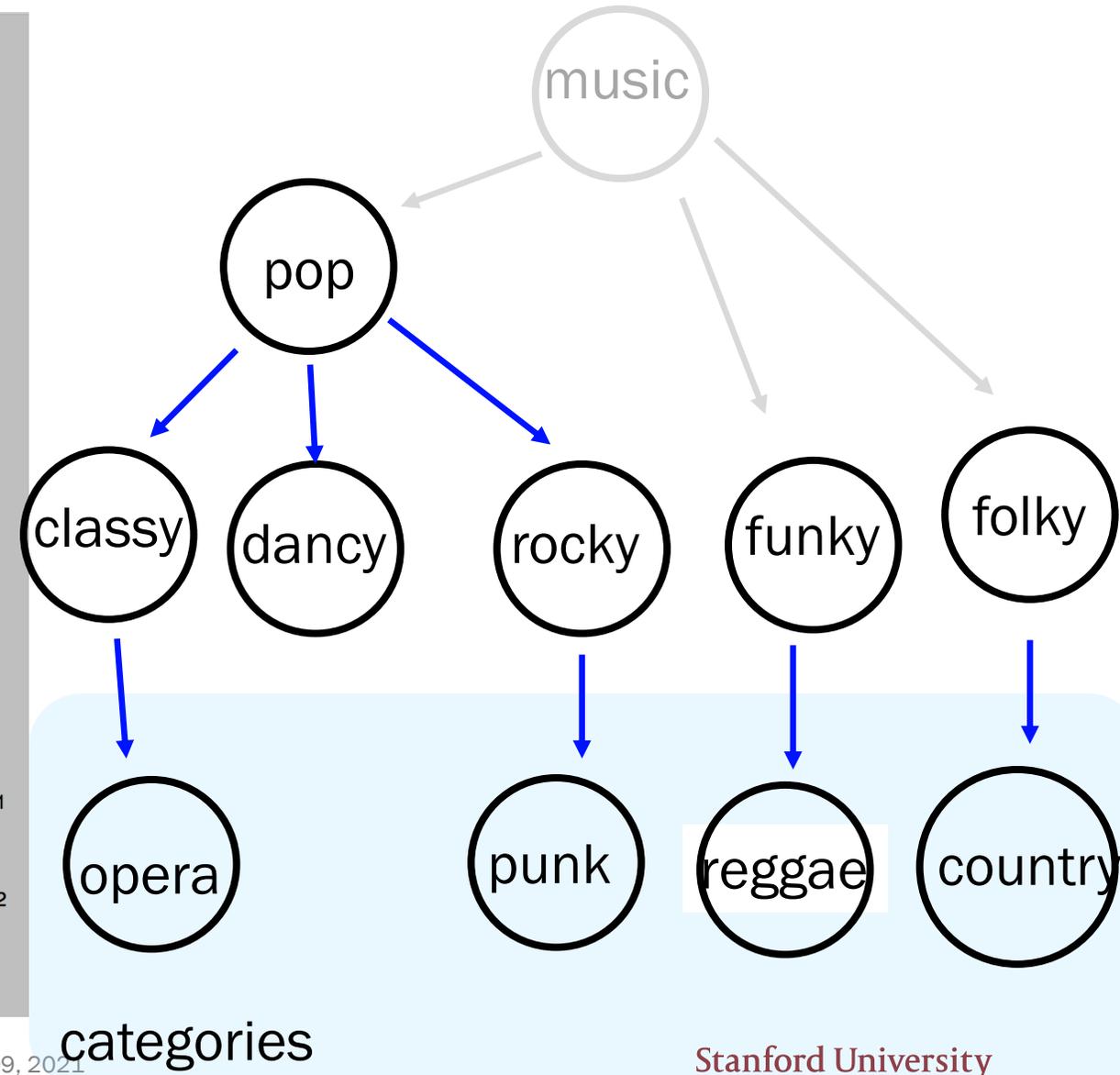
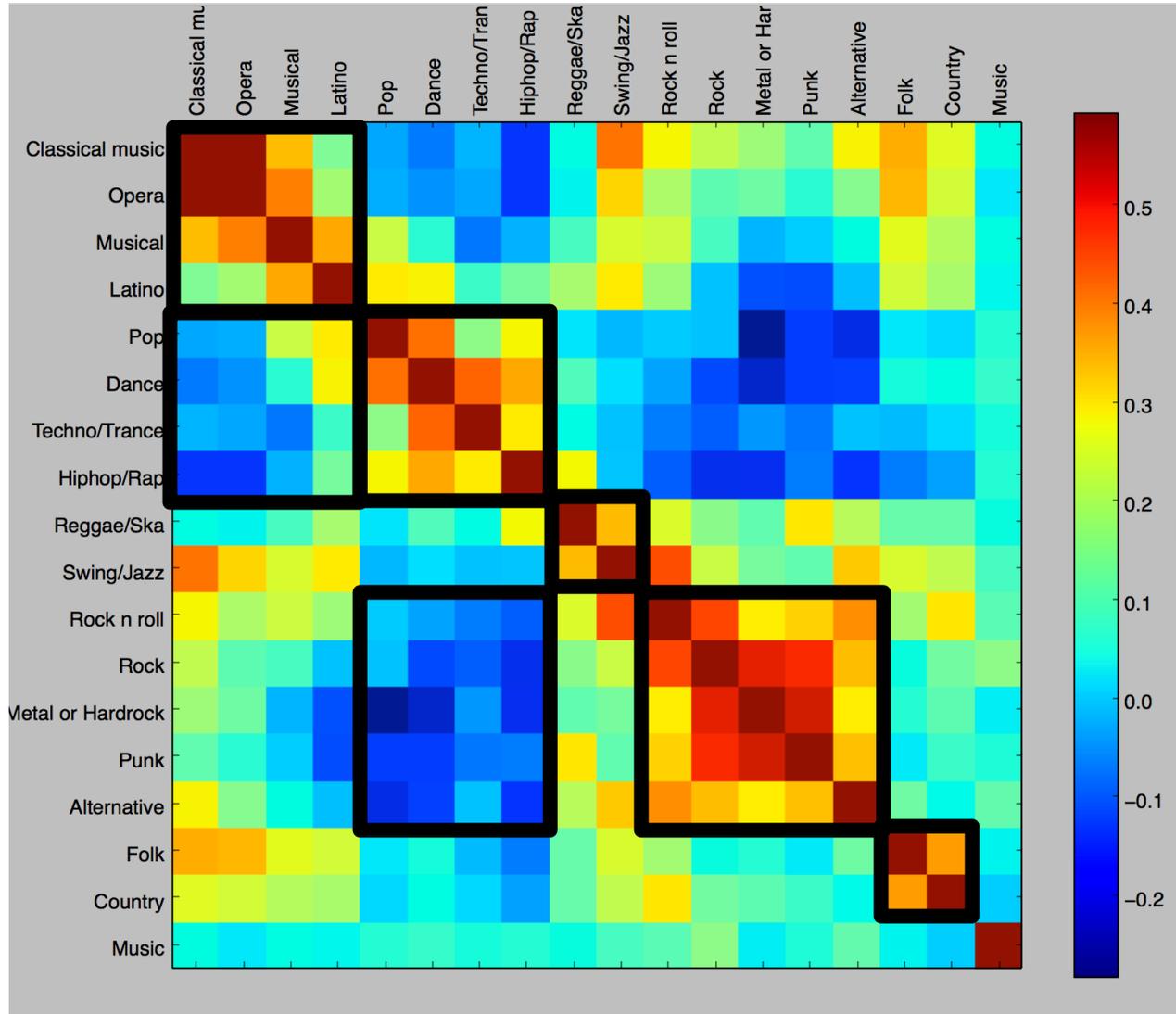
From Correlation to Bayes Net. Alternative!



From Correlation to Bayes Net. Alternative!



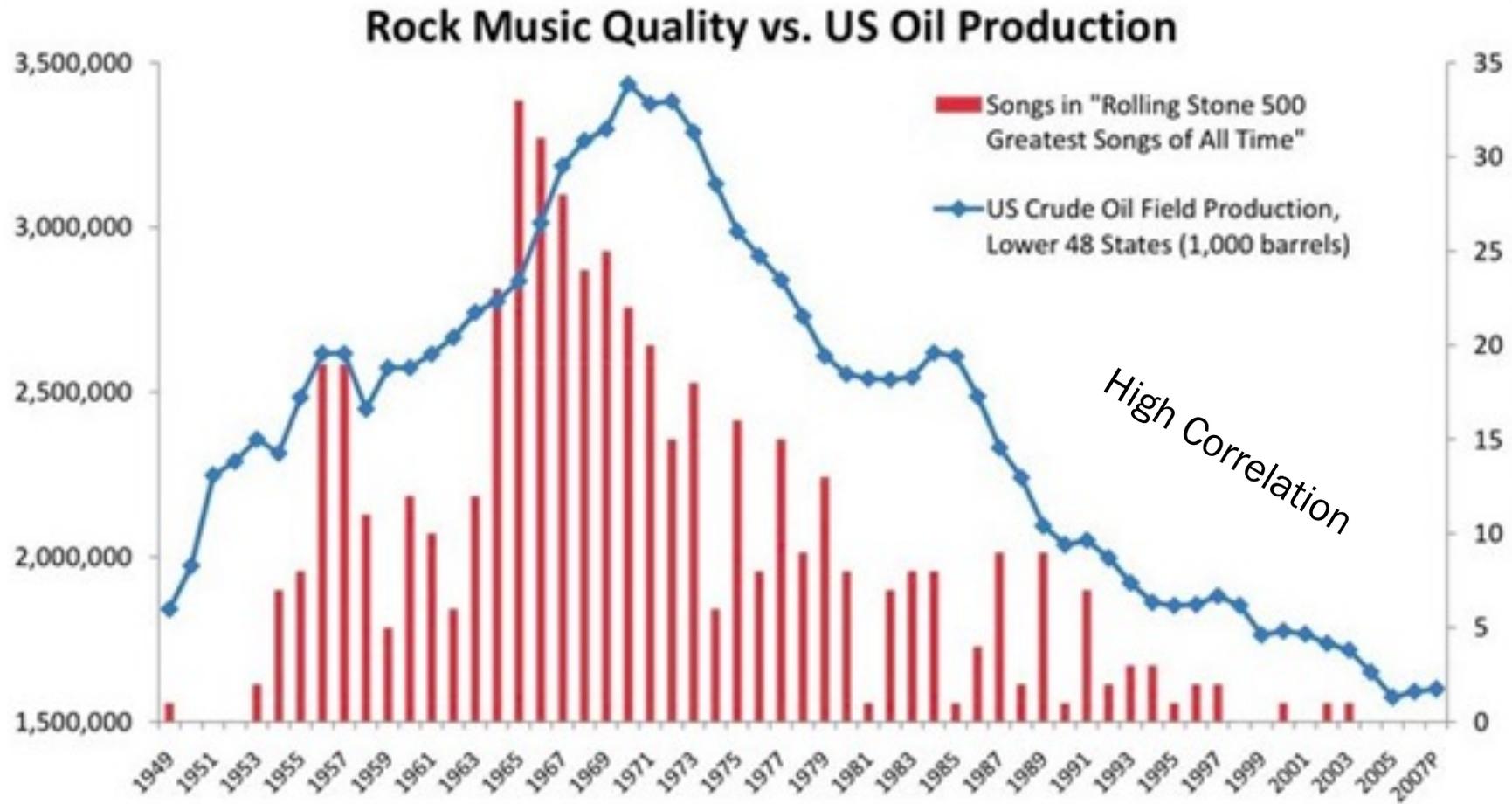
From Correlation to Bayes Net. Alternative!



How do you know if your model is good?

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

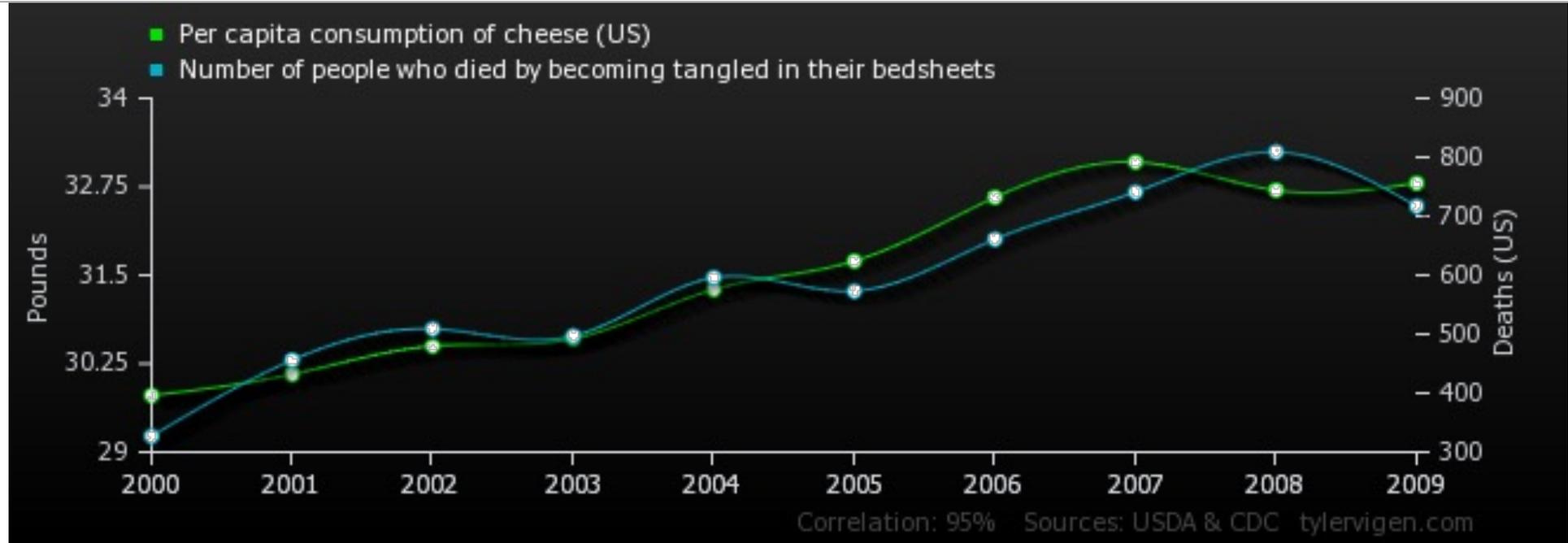
Rock Music Vs Oil?



Hubbert Peak Theory

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

Tell your friends!



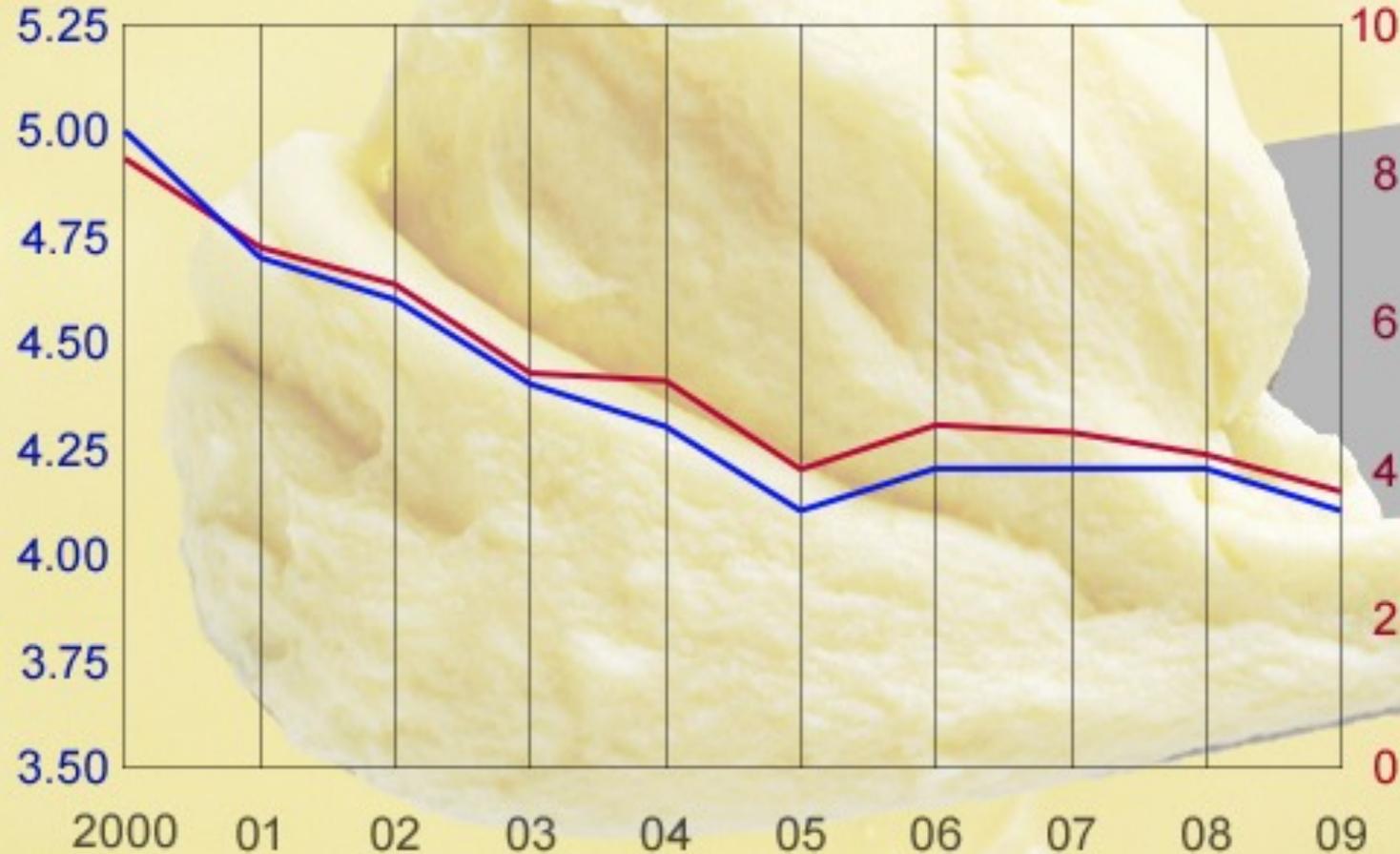
	<u>2000</u>	<u>2001</u>	<u>2002</u>	<u>2003</u>	<u>2004</u>	<u>2005</u>	<u>2006</u>	<u>2007</u>	<u>2008</u>	<u>2009</u>
<i>Per capita consumption of cheese (US) Pounds (USDA)</i>	29.8	30.1	30.5	30.6	31.3	31.7	32.6	33.1	32.7	32.8
<i>Number of people who died by becoming tangled in their bedsheets Deaths (US) (CDC)</i>	327	456	509	497	596	573	661	741	809	717
Correlation: 0.947091										

Divorce Vs Butter?

Divorce rate
in Maine per
1,000 people

Per capita
consumption of
margarine (lbs)

Correlation: 99%



Source: US Census, USDA, tylervigen.com

SPL

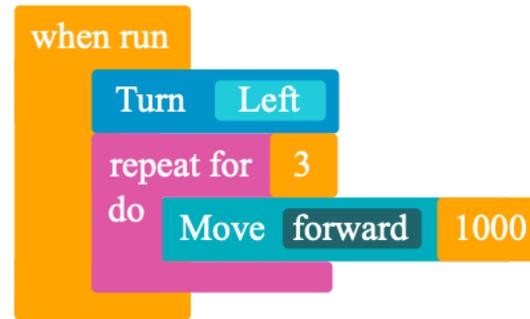
What about Code.org?

Computers Couldn't Understand Code

60,000 students attempted this problem
37,000 unique solutions



Challenge

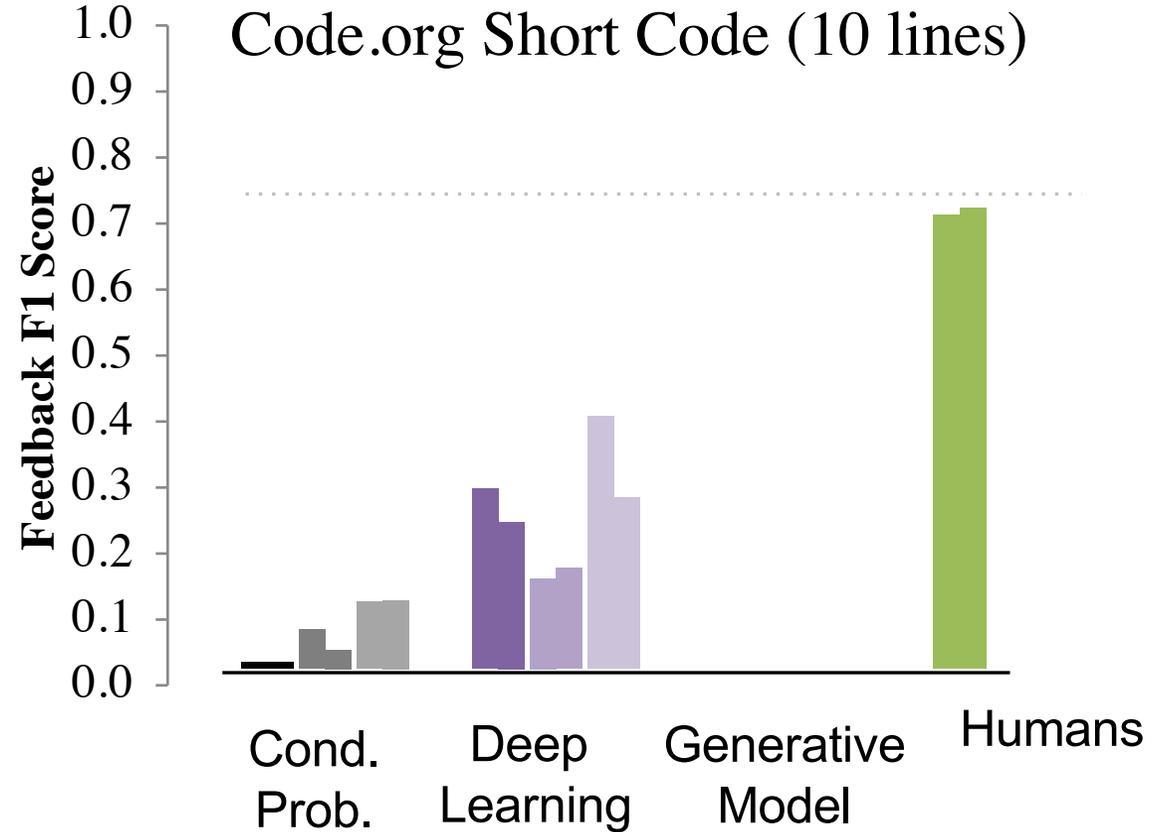


Student Code

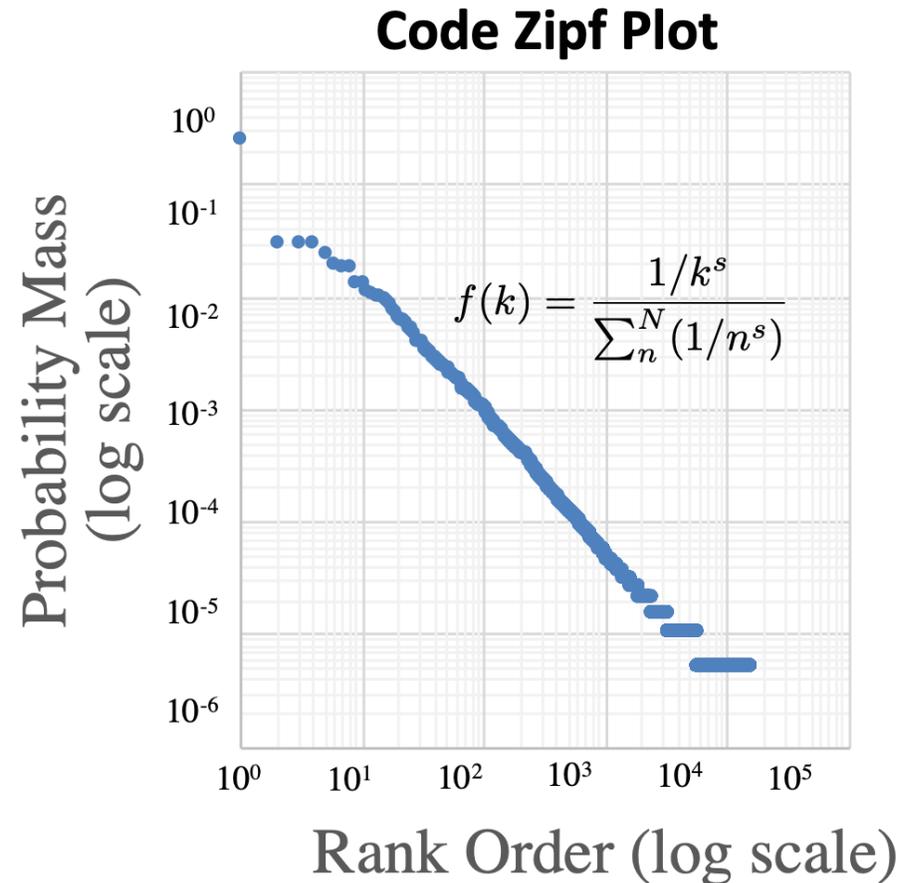
You need to
move and
turn in your
loop

Insight

Computers Couldn't Understand Code

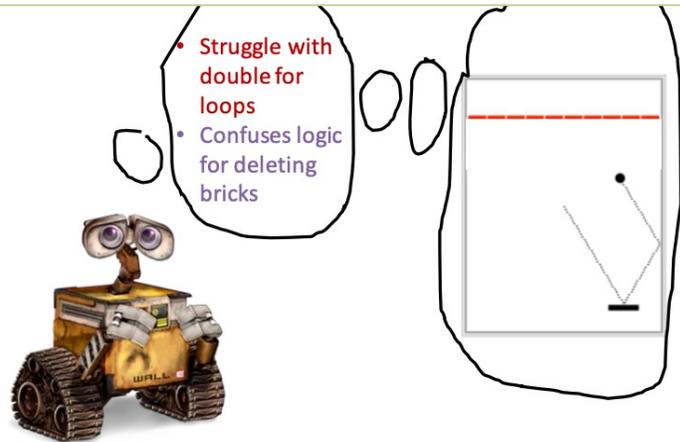


Code is hard to think about...



**Exponential combination of decisions.
Super fat tailed. Everything looks unique**

Generative Model is Easier than Straight Inference



This is easy and exponential



This is hard and linear

A student's
"ability"

$$\Theta \sim \text{pythonSample}$$

A student's
"choices"

$$C \sim \text{pythonSample} | \Theta$$

The resulting
code

$$\Pi \sim \text{pythonSample} | C$$

$$P(\Theta, C | \Pi)$$

Infer ability and
choices from code



Made a probabilistic programming language:

ideaToText

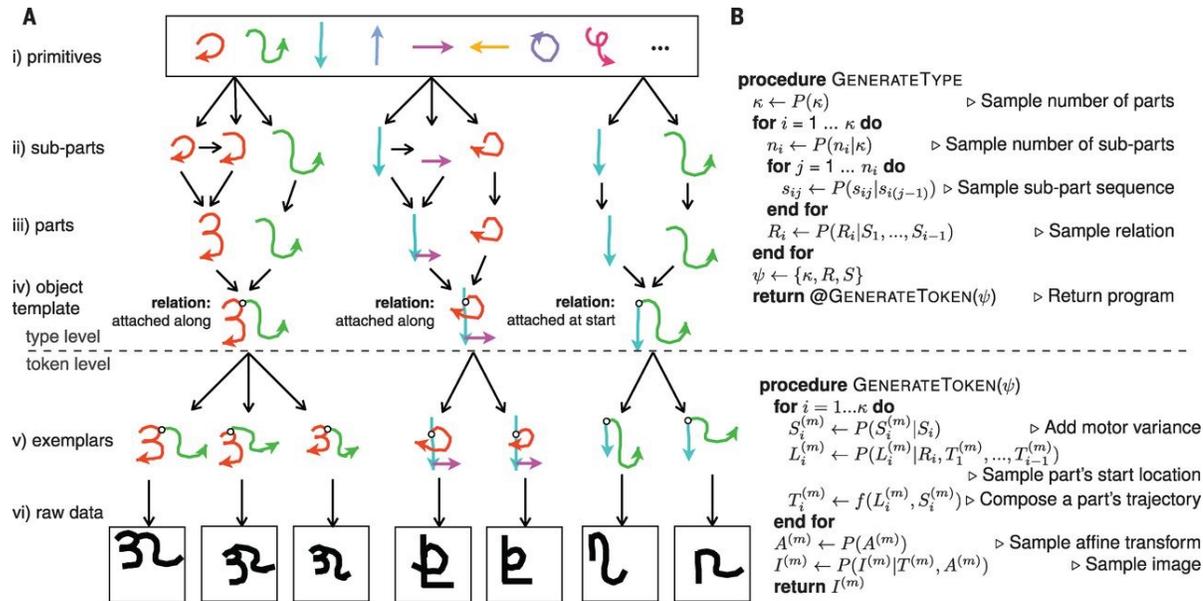
Write a model of each **decision point**, given parents

```
7
8 # This python class is a RubricSampling Decision
9 # it generates programs that print the numbers 10 -> 1
10 class Countdown(Decision):
11
12     def registerChoices(self):
13         # these are the main strategies for printing out a
14         # countdown
15         self.addRubricChoices('loop-style', {
16             'for' : 100,
17             'while' : 40,
18             'none' : 1,
19             'empty' : 1
20         })
21
22     # we can make some grading choices based on which
23     # strategy they chose (did they actually use a loop?)
24     def processChoices(self):
25         style = self.getChoice('loop-style')
26         hasLoop = style != 'none' and style != 'empty'
27         self.addLabel('rubric-hasLoop', hasLoop)
28         self.addLabel('rubric-printsNums', style != 'empty')
29
30     # Based on their strategy render a different decision
31     def renderCode(self):
32         style = self.getChoice('loop-style')
33         if style == 'for': return '{ForSoln}'
34         if style == 'while': return '{WhileSoln}'
35         if style == 'none': return '{NoLoopSoln}'
36         if style == 'empty': return ''
37
```

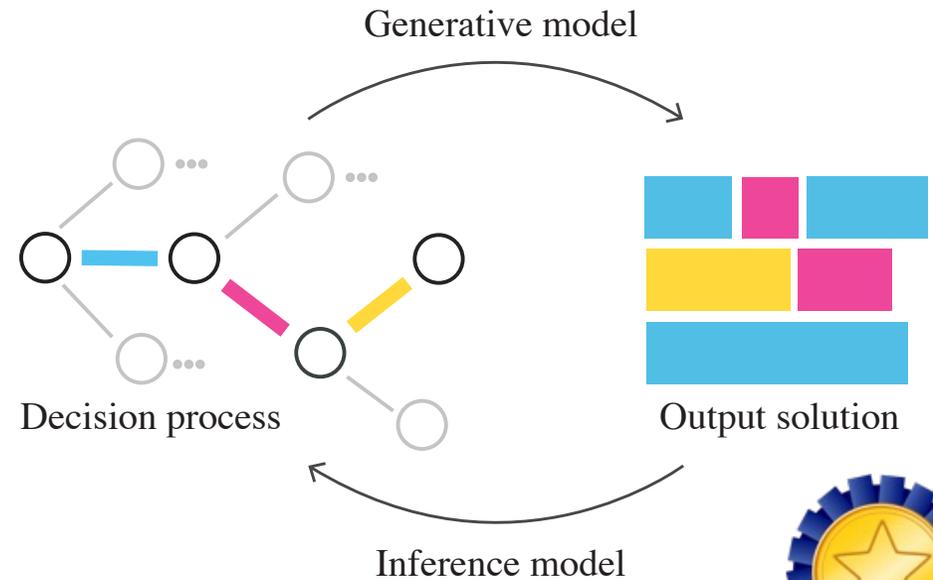
1. This is code for a single decision point
2. Give a name to the choice that the student is making
3. How do those choices translate into grades?
4. What does the code look like? Often evokes other decision points

Generative Model of Grading

Lake et al, 2015

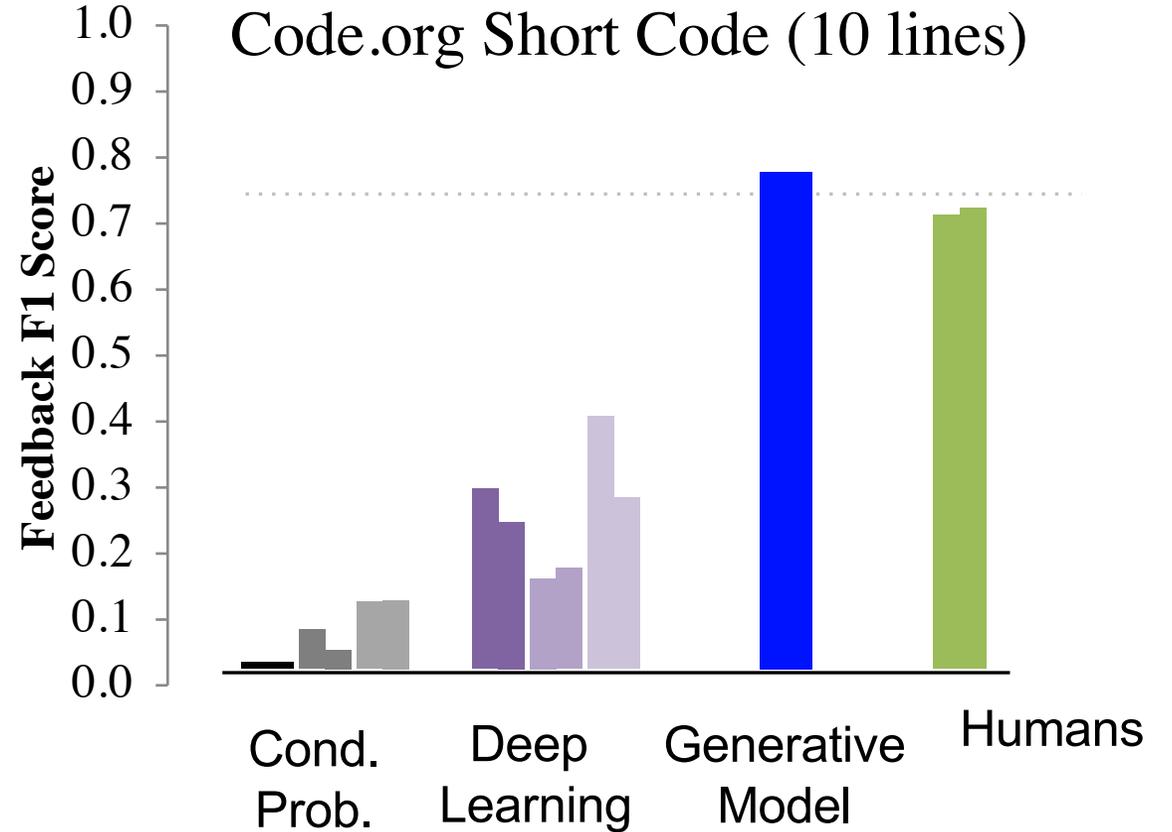


Muke Wu, Ali Malik, Noah Goodman, Chris Piech, 2019



Outstanding Student
paper award, AAI 2019

Computers Couldn't Understand Code



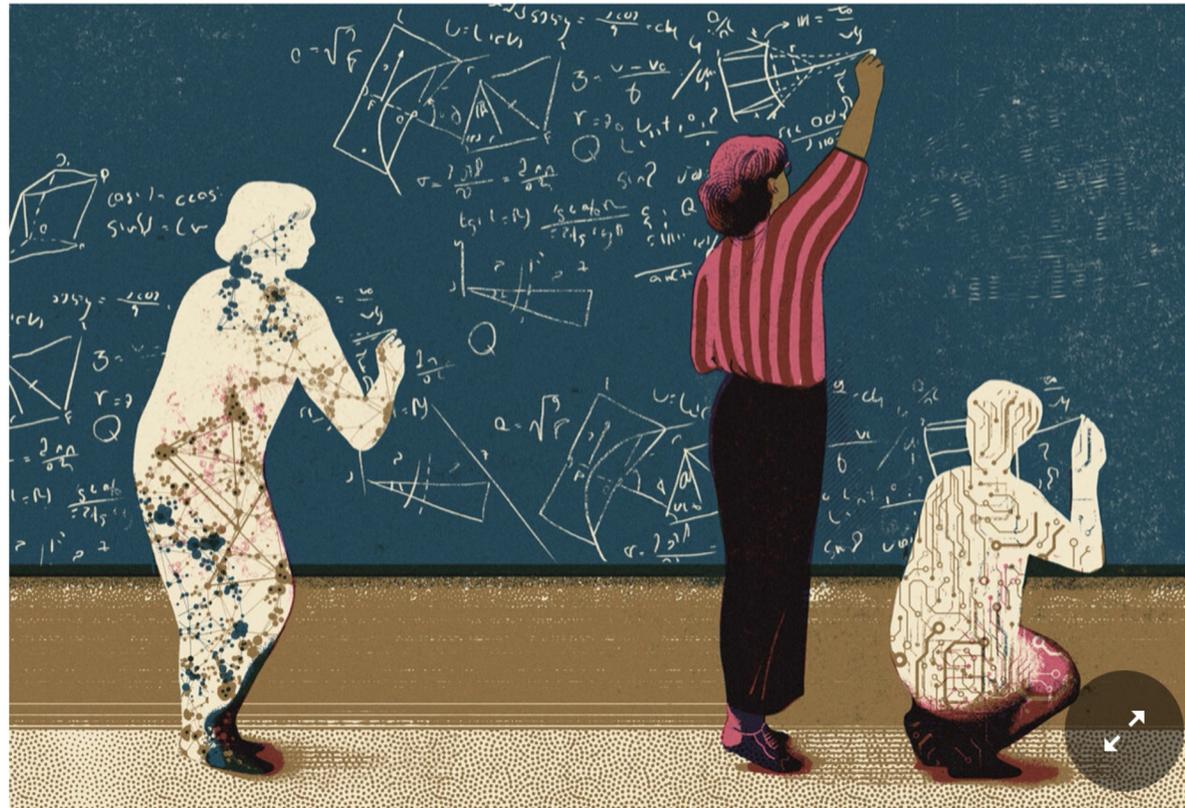
The
New York
Times

Can A.I. Grade Your Next Test?

Neural networks could give online education a boost by providing automated feedback to students.

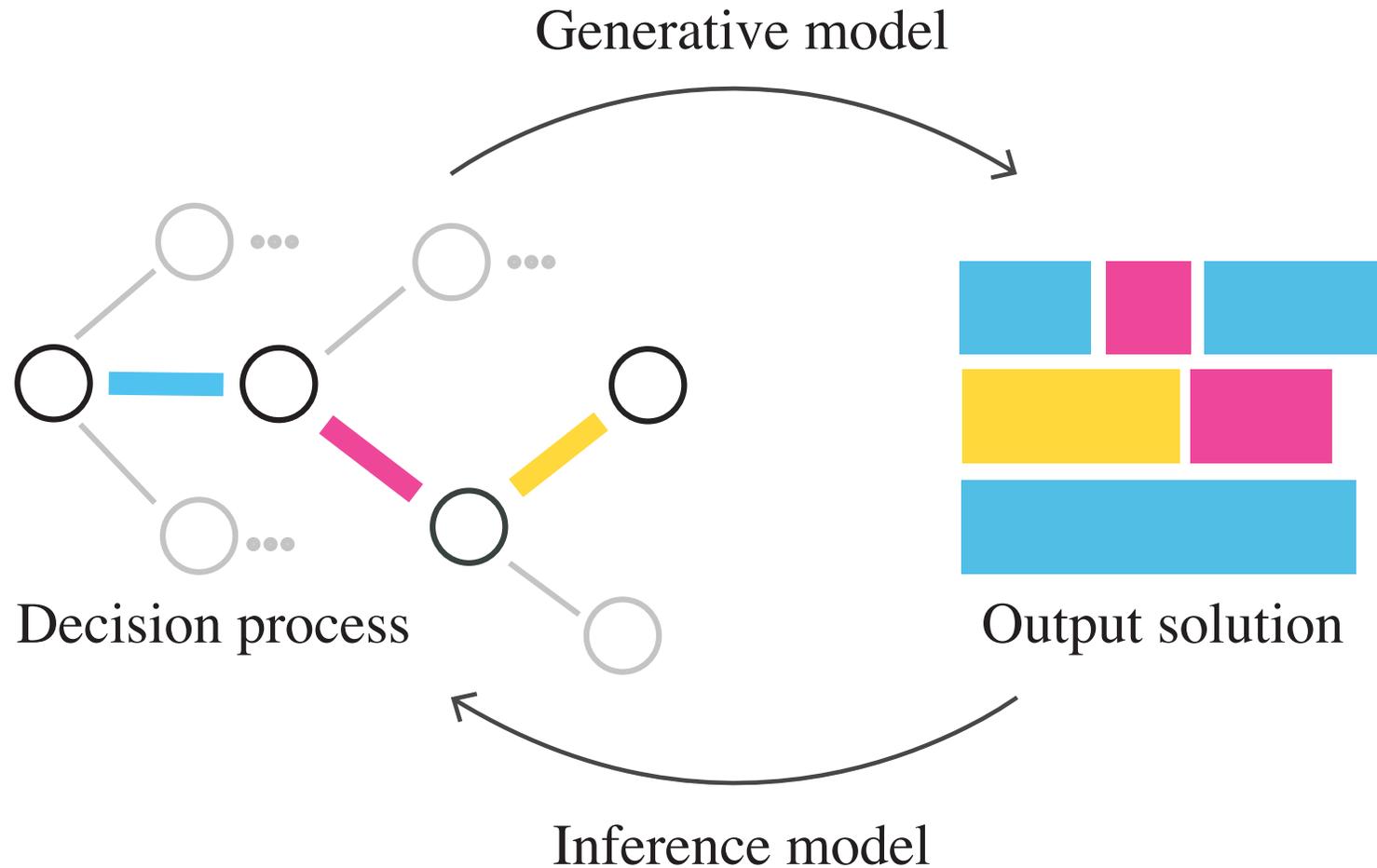


Stanford | News



<https://www.nytimes.com/2021/07/20/technology/ai-education-neural-networks.html>

Friday: General Inference



Que te vayas bien

No Class on Monday



This is how I
imagine students
are when they are
studying 😊

(recall why we are
here. Learning is a joy.
You are growing)