Breaking tables apart and putting them back together

Remy Wang, 4/3/2025

name	age	bday	kind
casa	8	2017-01-01	cat
kira	6	2019-09-16	cat
toby	17	2008-02-03	dog
maya	10	2015-11-21	dog

name	age	bday	kind
casa	8	2017-01-01	cat
kira	6	2019-09-16	cat
maya	10	2015-11-21	dog

```
SELECT kind, avg(age*7)
  FROM pets
 WHERE year(bday)>2010;
GROUP BY kind
```

age	kind
49	cat
70	dog

pets

name	breed	age	origin	kind
casa	tabby	8	seatte	cat
kira	tuxedo	6	hawaii	cat
toby	border collie	17	seattle	dog
maya	husky	10	LA	dog

people

name	pet	addr.	phone	job
remy	casa	LA	###	UCLA

places

name	addr.	type
UCLA	LA	edu.

everything

pet	breed	age	origin	kind	person	addr.	phone	job
casa	tabby	8	seattle	cat	remy	LA	###	UCLA
kira	tuxedo	6	hawaii	cat	remy	LA	###	UCLA
toby	border collie	17	seattle	dog	remy	LA	###	UCLA
maya	husky	10	LA	dog	remy	LA	###	UCLA
...	...	...	...	...	...	...	...	...

everything

pet	breed	age	origin	kind	person	addr.	phone	job	addr.	type
casa	tabby	8	seattle	cat	remy	LA	###	UCLA	LA	edu
kira	tuxedo	6	hawaii	cat	remy	LA	###	UCLA	LA	edu
toby	border collie	17	seattle	dog	remy	LA	###	UCLA	LA	edu
maya	husky	10	LA	dog	remy	LA	###	UCLA	LA	edu
...	...	...	...	...	...	...	...	...	...	...

everything

pet	breed	age	origin	kind	person	addr.	phone	job
casa	tabby	8	seattle	cat	remy	LA	###	UCLA
kira	tuxedo	6	hawaii	cat	remy	LA	###	UCLA
toby	border collie	17	seattle	dog	remy	LA	###	UCLA
maya	husky	10	LA	dog	remy	LA	###	UCLA
...	...	...	...	...	...	...	...	...

everything

pet	breed	age	origin	kind	person	addr.	phone	job
casa	tabby	8	seattle	cat	remy	LA	###	UCLA
kira	tuxedo	6	hawaii	cat	remy	LA	###	UCLA
toby	border collie	17	seattle	dog	remy	LA	###	UCLA
maya	husky	10	LA	dog	remy	LA	###	UCLA
casa	tabby	8	seattle	cat	remy	seattle	###	UW
kira	tuxedo	6	hawaii	cat	remy	seattle	###	UW
toby	border collie	17	seattle	dog	remy	seattle	###	UW
maya	husky	10	LA	dog	remy	seattle	###	UW

Problem: job **independent** from pet!

But what does that mean?

(a detour to probabilities)

c
a
b
c
d

n
1
2
3
4

$$P(n = 1) =$$

$$P(c = a) =$$

$$P(n = 1 \wedge c = a) =$$

$$P(n = 1 \mid c = a) =$$

$$P(A \mid B) = P(A)$$

A independent from B

$$P(AB) = P(A)P(B) \qquad A \perp B$$

everything

pet	breed	age	origin	kind	person	addr.	phone	job
casa	tabby	8	seattle	cat	remy	LA	###	UCLA
kira	tuxedo	6	hawaii	cat	remy	LA	###	UCLA
toby	border collie	17	seattle	dog	remy	LA	###	UCLA
maya	husky	10	LA	dog	remy	LA	###	UCLA
casa	tabby	8	seattle	cat	remy	seattle	###	UW
kira	tuxedo	6	hawaii	cat	remy	seattle	###	UW
toby	border collie	17	seattle	dog	remy	seattle	###	UW
maya	husky	10	LA	dog	remy	seattle	###	UW

everything

$$P(\text{kind} = \text{cat}) =$$

$$P(\text{job} = \text{UCLA}) =$$

$$P(\text{cat} \wedge \text{UCLA}) =$$

$$\text{kind} \perp \text{job}$$

kind
cat
cat
dog
dog
cat
cat
dog
dog

job
UCLA
UCLA
UCLA
UCLA
UW
UW
UW
UW

kind	person	job
cat	remy	UCLA
dog	remy	UCLA
cat	remy	UW
dog	remy	UW
cat	vincent	UCLA

$$P(\text{cat}) =$$

$$P(\text{cat} \mid \text{UCLA}) =$$

Not independent!

kind	person	job
cat	remy	UCLA
dog	remy	UCLA
cat	remy	UW
dog	remy	UW
cat	vincent	UCLA

Break down by **person**

$$P(\text{cat}) = P(\text{cat} \mid \text{UCLA}) =$$

$$\text{kind} \perp \text{job} \mid \text{person}$$

$$P(\text{cat}) = P(\text{cat} \mid \text{UCLA}) =$$

kind	person	job
cat	remy	UCLA
dog	remy	UCLA
cat	remy	UW
dog	remy	UW
cat	vincent	UCLA

no probabilities? no prob

kind = cat, job = { }

kind = dog, job = { }

check for each **person**:

every kind has same set of jobs?

kind	person	job
cat	remy	UCLA
dog	remy	UCLA
cat	remy	UW
dog	remy	UW
cat	vincent	UCLA

```
#{ } = COUNT DISTINCT

#{(k,j)} = #{k} * #{j}
```

$$\frac{1}{P(KJ)} = \frac{1}{P(K)}\frac{1}{P(J)}$$

$$K \perp J$$

Dependence is bad, now what?

pet	breed	age	origin	kind	person	addr.	phone	job
casa	tabby	8	seattle	cat	remy	LA	###	UCLA
kira	tuxedo	6	hawaii	cat	remy	LA	###	UCLA
toby	border collie	17	seattle	dog	remy	LA	###	UCLA
maya	husky	10	LA	dog	remy	LA	###	UCLA
casa	tabby	8	seattle	cat	remy	seattle	###	UW
kira	tuxedo	6	hawaii	cat	remy	seattle	###	UW
toby	border collie	17	seattle	dog	remy	seattle	###	UW
maya	husky	10	LA	dog	remy	seattle	###	UW

pet	breed	age	origin	kind	person	addr.	phone	job
casa	tabby	8	seattle	cat	remy	LA	###	UCLA
kira	tuxedo	6	hawaii	cat	remy	LA	###	UCLA
toby	border collie	17	seattle	dog	remy	LA	###	UCLA
maya	husky	10	LA	dog	remy	LA	###	UCLA
casa	tabby	8	seattle	cat	remy	seattle	###	UW
kira	tuxedo	6	hawaii	cat	remy	seattle	###	UW
toby	border collie	17	seattle	dog	remy	seattle	###	UW
maya	husky	10	LA	dog	remy	seattle	###	UW

SELECT DISTINCT ..., person
FROM everything

How?

SELECT DISTINCT **...**
FROM everything

pet	breed	age	origin	kind	person
casa	tabby	8	seattle	cat	remy
kira	tuxedo	6	hawaii	cat	remy
toby	border collie	17	seattle	dog	remy
maya	husky	10	LA	dog	remy

person	addr.	phone	job
remy	LA	###	UCLA
remy	seattle	###	UW

person	addr.	phone	job	job addr.
remy	...	123	UCLA	LA
zifan	...	234	UCLA	LA
vincent	...	345	UCLA	LA
remy	...	123	UW	seattle
dan	...	456	UW	seattle
magda	...	567	UW	seattle

job **determines** job addr.

$$P(\text{LA} \mid \text{UCLA}) = 1$$

person	addr.	phone	job
remy	...	123	UCLA
zifan	...	234	UCLA
vincent	...	345	UCLA
remy	...	123	UW
dan	...	456	UW
magda	...	567	UW

job	job addr.
UCLA	LA
UW	seattle

person	addr.	phone	job
remy	...	123	UCLA
zifan	...	234	UCLA
vincent	...	345	UCLA
remy	...	123	UW
dan	...	456	UW
magda	...	567	UW

job	job addr.
UCLA	LA
UW	seattle

Where's Remy's office?

pet	breed	age	origin	kind	person
casa	tabby	8	seattle	cat	remy
kira	tuxedo	6	hawaii	cat	remy
toby	border collie	17	seattle	dog	remy
maya	husky	10	LA	dog	remy

name	addr.	phone	job
remy	LA	###	UCLA
remy	seattle	###	UW

Where do cat people work?

pet	breed	age	origin	kind	person	addr.	phone	job
casa	tabby	8	seattle	cat	remy	LA	###	UCLA
kira	tuxedo	6	hawaii	cat	remy	LA	###	UCLA
toby	border collie	17	seattle	dog	remy	LA	###	UCLA
maya	husky	10	LA	dog	remy	LA	###	UCLA
casa	tabby	8	seattle	cat	remy	seattle	###	UW
kira	tuxedo	6	hawaii	cat	remy	seattle	###	UW
toby	border collie	17	seattle	dog	remy	seattle	###	UW
maya	husky	10	LA	dog	remy	seattle	###	UW

JOIN ⋈

pet	breed	age	origin	kind	person
casa	tabby	8	seattle	cat	remy
kira	tuxedo	6	hawaii	cat	remy
toby	border collie	17	seattle	dog	remy
maya	husky	10	LA	dog	remy

name	addr.	phone	job
remy	LA	###	UCLA
remy	seattle	###	UW

```
SELECT job FROM
 pets JOIN people
      ON pets.person = people.name
```

Join key

pet	breed	age	origin	kind	person
casa	tabby	8	seattle	cat	remy
kira	tuxedo	6	hawaii	cat	remy
toby	border collie	17	seattle	dog	remy
maya	husky	10	LA	dog	remy

name	addr.	phone	job
remy	LA	###	UCLA
remy	seattle	###	UW

```
for pet in pets:
  for person in people:
    if pet.person = person.name:
      print(job)
```

Join key

pet	breed	age	origin	kind	person
casa	tabby	8	seattle	cat	remy
kira	tuxedo	6	hawaii	cat	remy
toby	border collie	17	seattle	dog	remy
maya	husky	10	LA	dog	remy

name	addr.	phone	job
remy	LA	###	UCLA
remy	seattle	###	UW

```
for pet in pets:
  for person in people:

        print(job)
```

Join key

pet	breed	age	origin	kind	person
casa	tabby	8	seattle	cat	remy
kira	tuxedo	6	hawaii	cat	remy
toby	border collie	17	seattle	dog	remy
maya	husky	10	LA	dog	remy

name	addr.	phone	job
remy	LA	###	UCLA
remy	seattle	###	UW

$$T_1 \bowtie_p T_2 = \sigma_p(T_1 \times T_2)$$

```sql
SELECT job FROM
 pets JOIN people
      ON pets.person = people.name
```

```sql
SELECT job
 FROM pets, people
WHERE pets.person = people.name
```

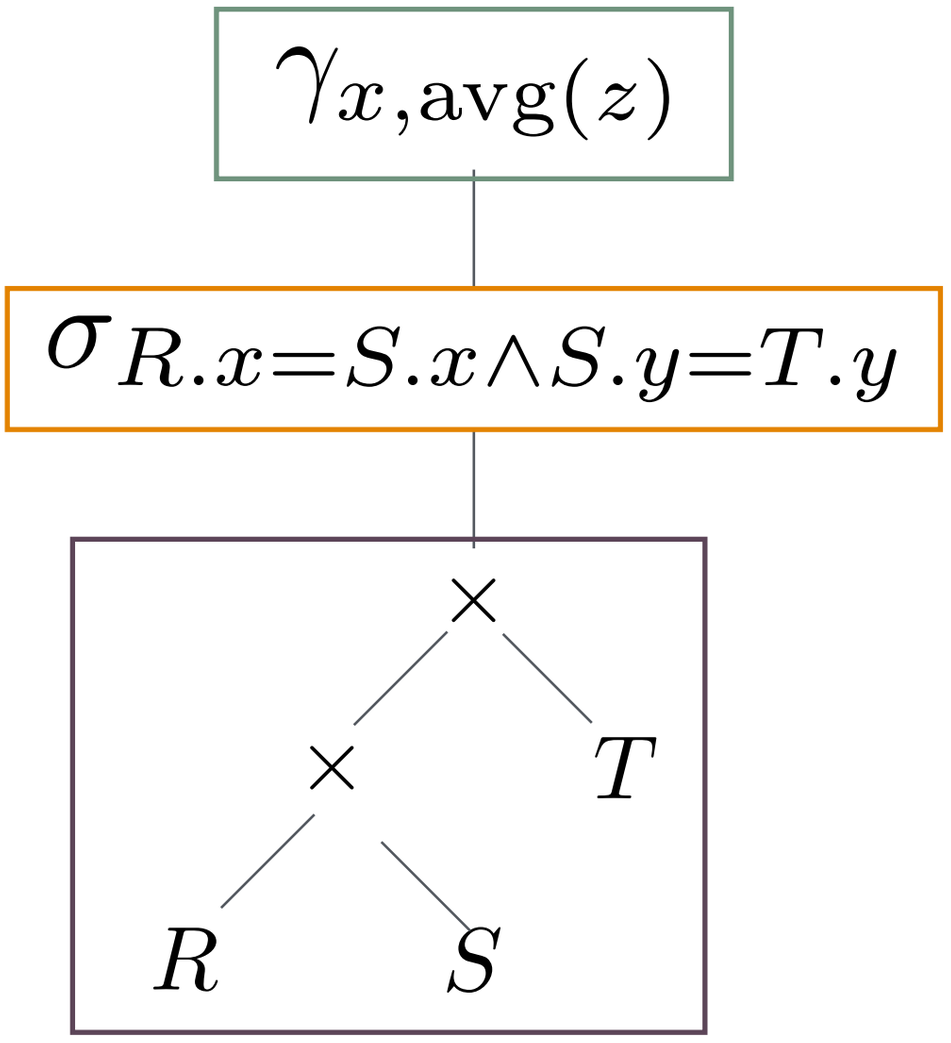
$$\text{job} \in \text{output}$$

$$\updownarrow$$

$$\exists \text{a} \in \text{pets}, \text{p} \in \text{people} : \text{p.job} = \text{job}$$

$$\wedge\ \text{a.person} = \text{p.name}$$

```
SELECT R.x, AVG(T.z)
  FROM R, S, T
 WHERE R.x = S.x
   AND S.y = T.y
GROUP BY R.x
```



i
1
2
3
4
5
6

⊙

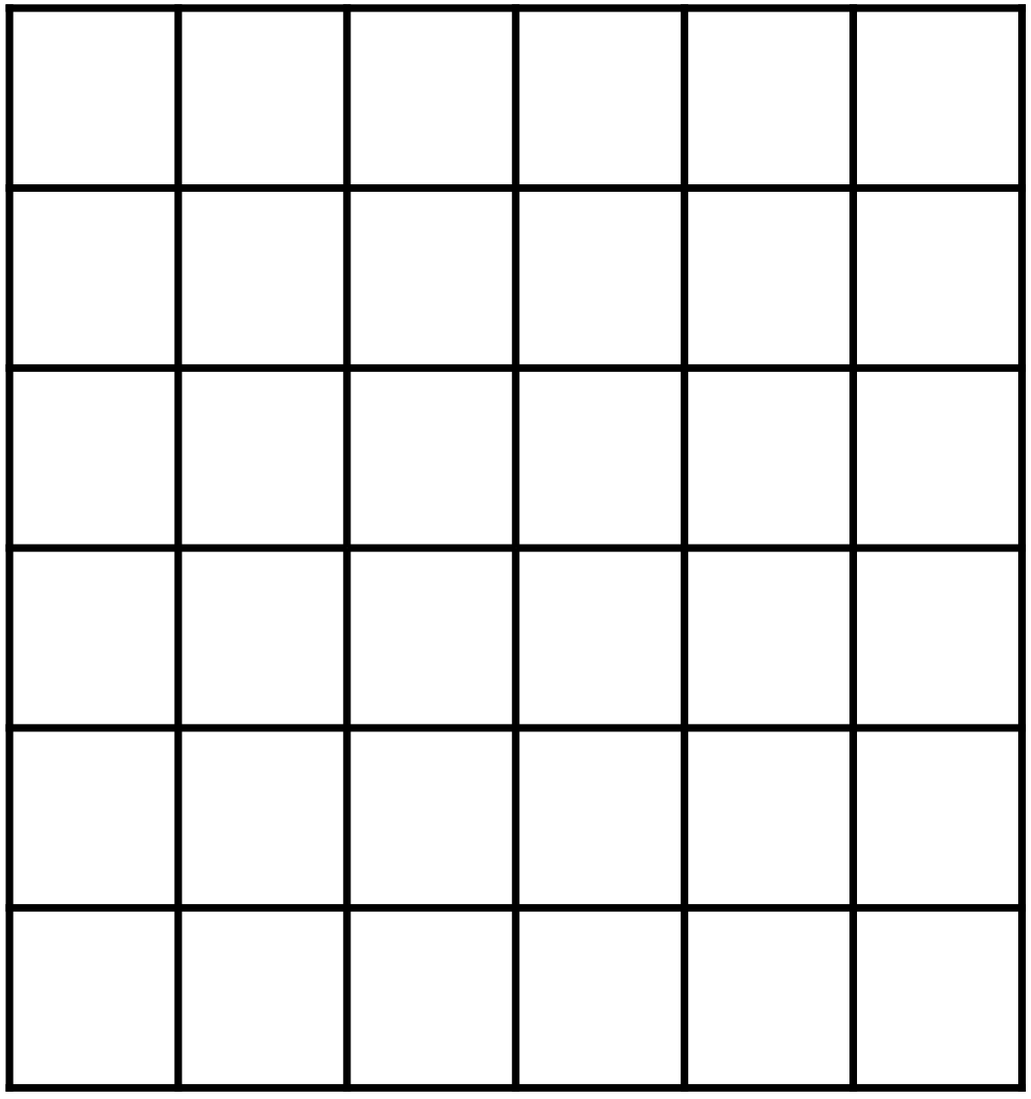
i
1
2
3
4
5
6

i
1
4
9
16
25
36

i
1
2
3
4
5
6

×

i
1
2
3
4
5
6



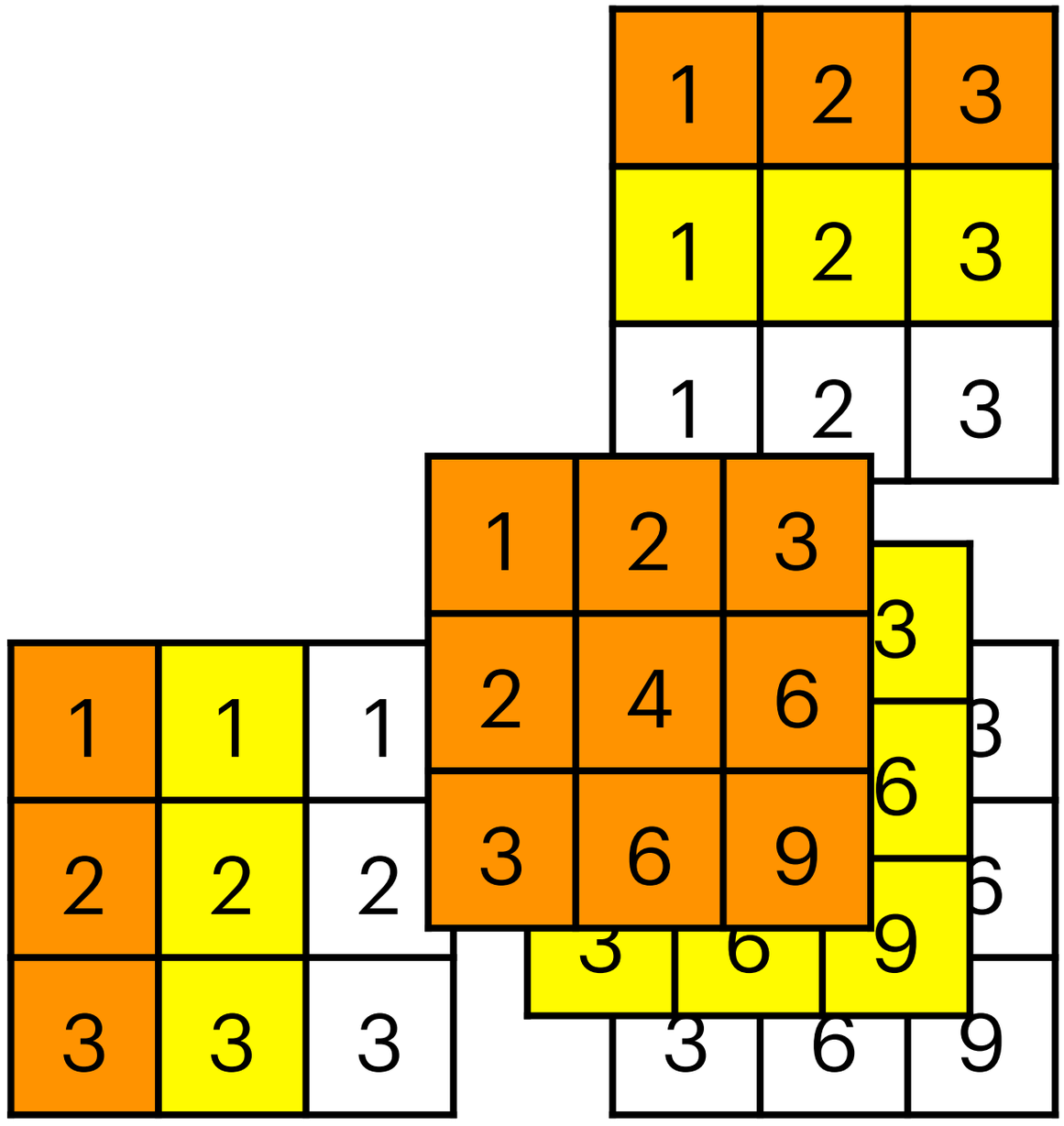
×	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

i
1
2
3
4
5
6

•

i
1
2
3
4
5
6

$$u \cdot v = \sum u \odot v$$



$$AB_{ik} = \sum_j A_{ij} \times B_{jk}$$

f

I	O
1	2
2	3
3	4
4	5

g

I	O
2	4
3	6
4	8
5	10

$g \circ f(x) = g(f(x))$

I	O
1	4
2	6
3	8
4	10