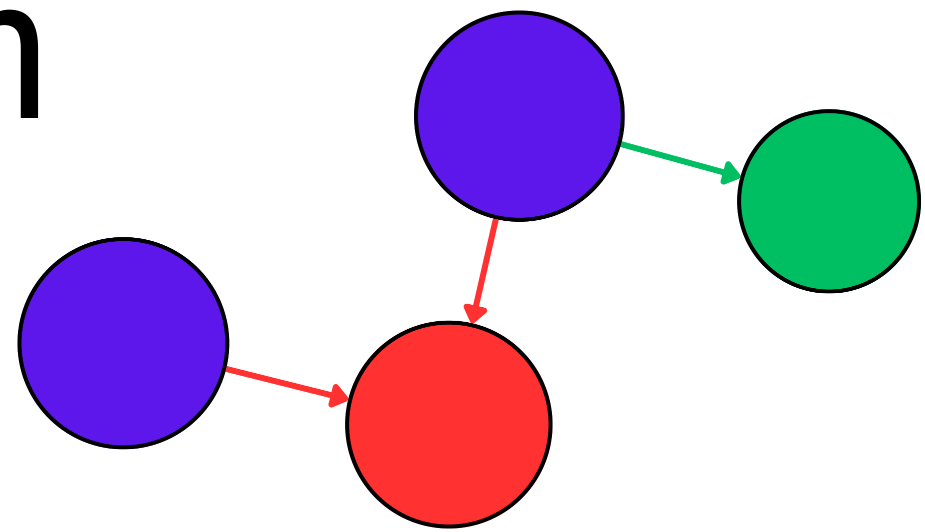
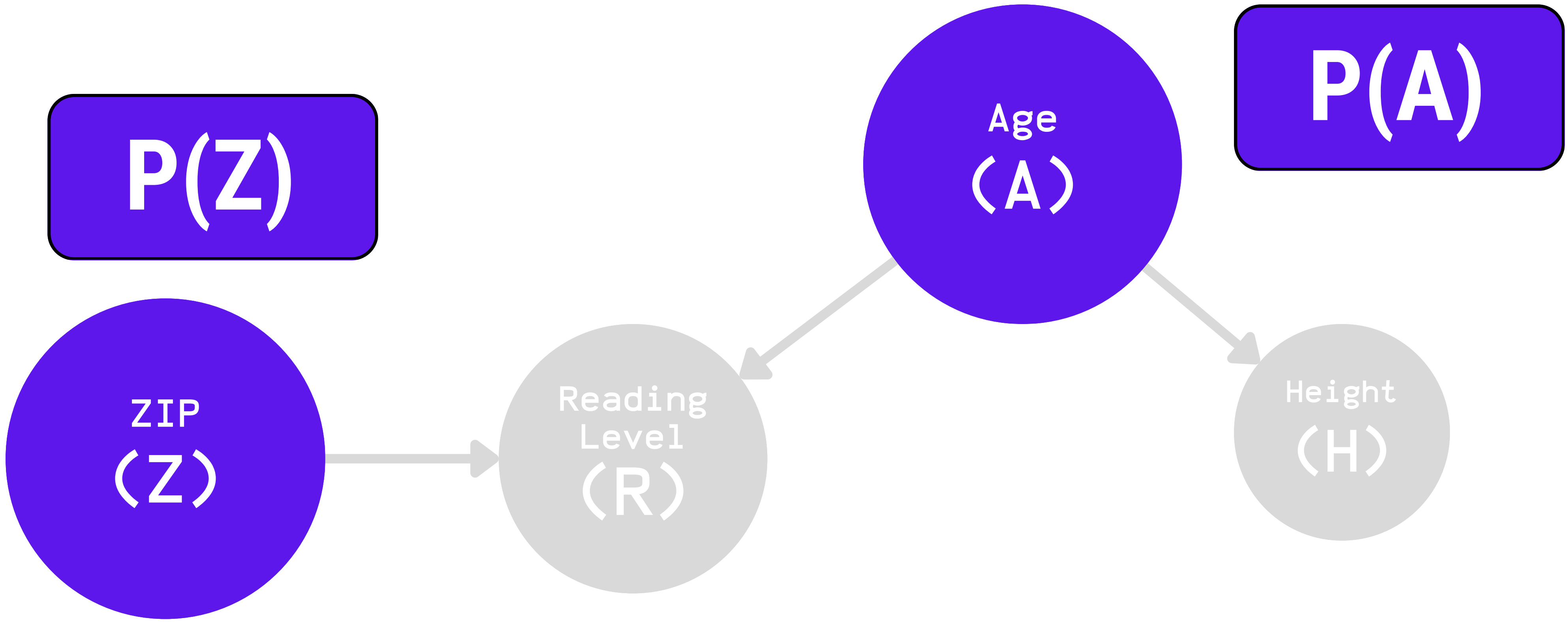
Going from a DAG to a factorization:

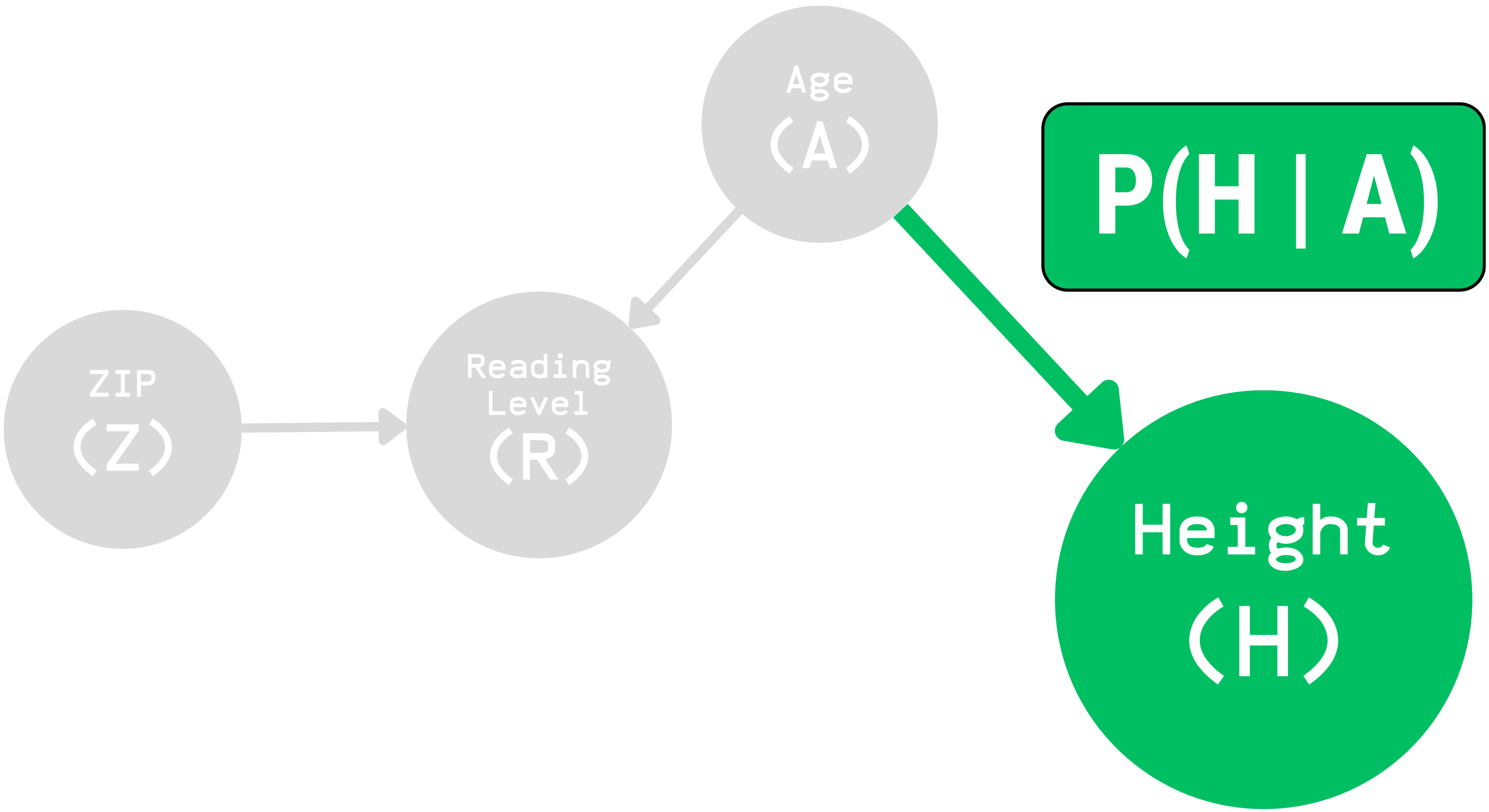
the “code” in
3 Rules





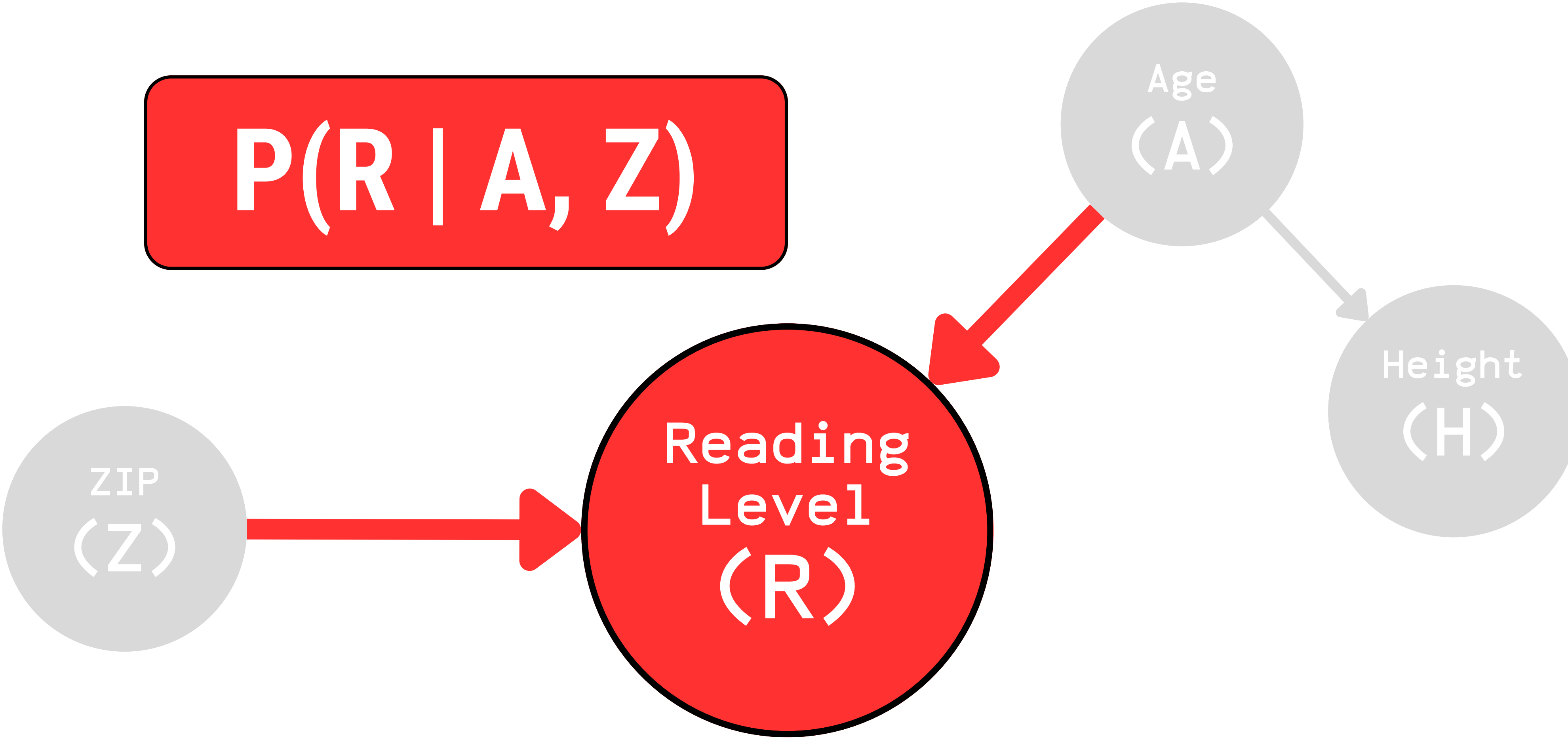
When a node has ***NO arrows*** *pointing into it... we must include each*

MARGINAL PROBABILITY



*When a node has **ONLY ONE arrow** pointing into it... we must include that*

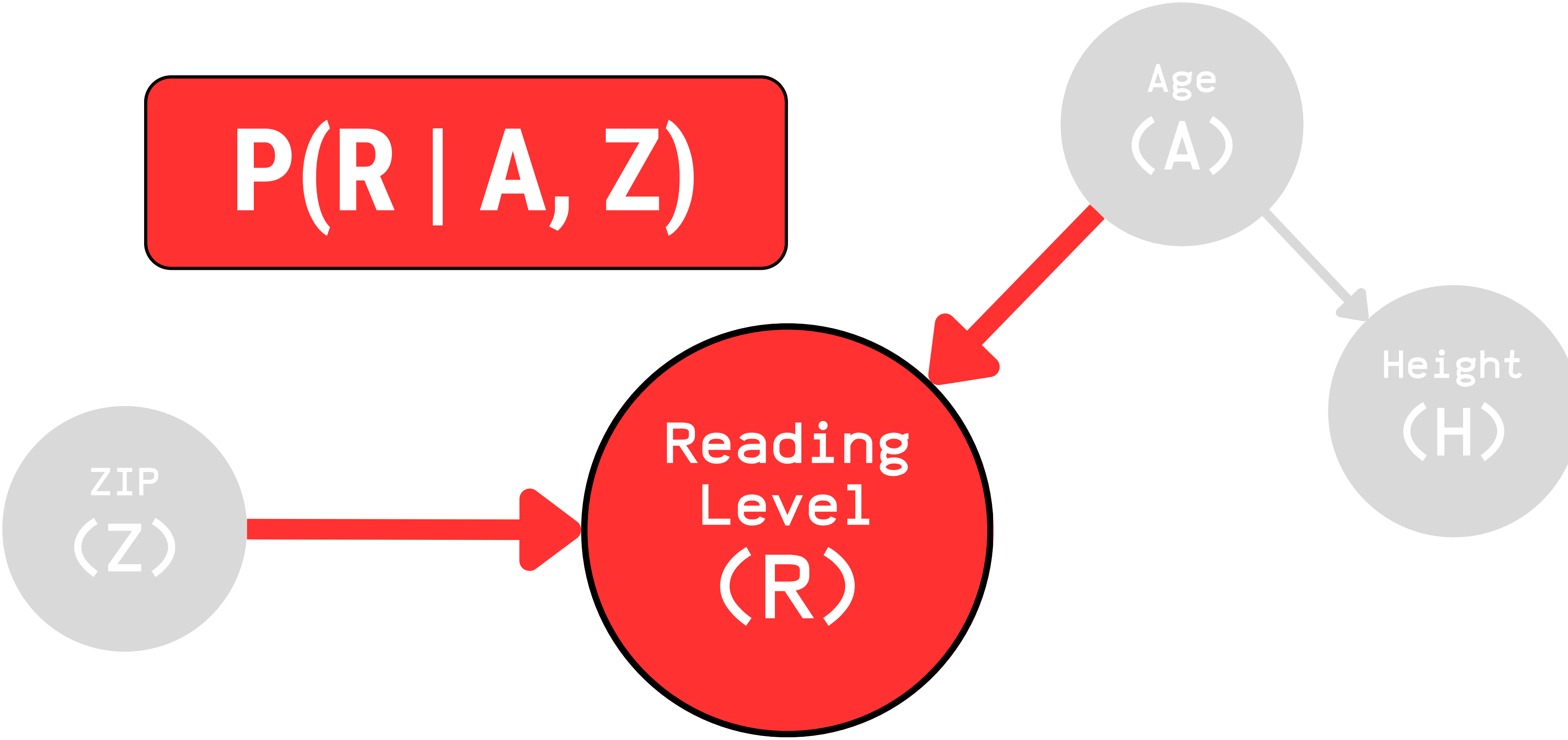
CONDITIONAL PROBABILITY



*When a node has **MULTIPLE arrows** pointing into it... we must include its*

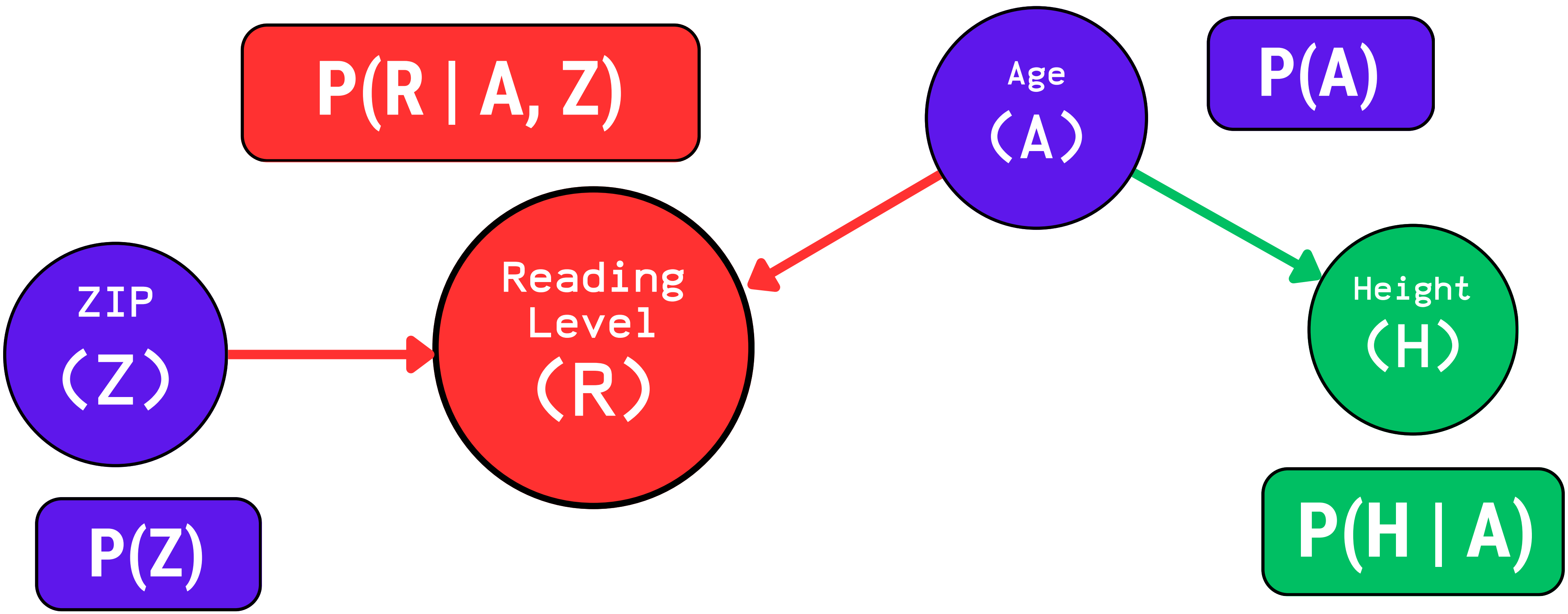
CONDITIONAL PROBABILITY

$P(R \mid A, Z)$



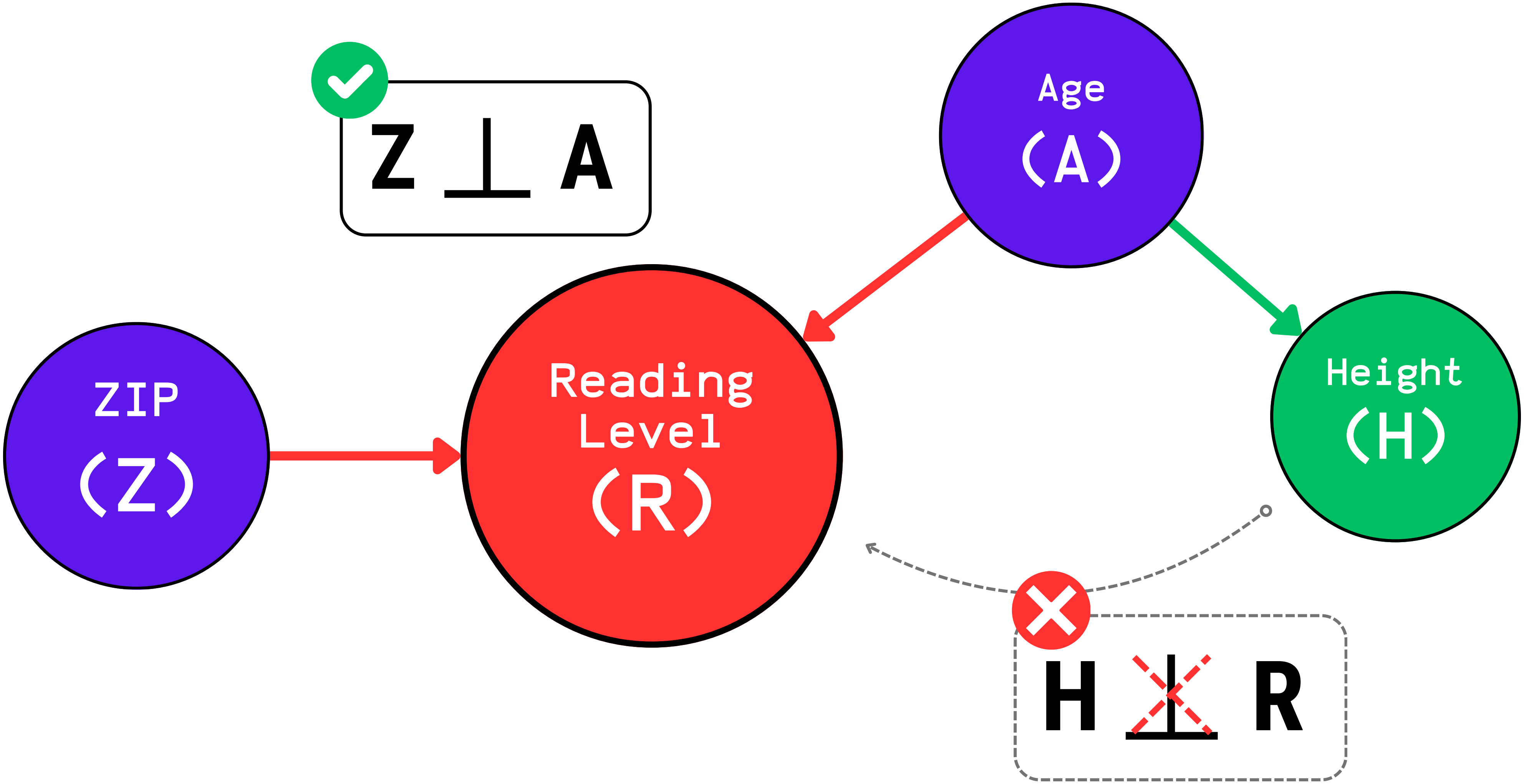
*When a node has **MULTIPLE arrows** pointing into it... we must include its*

CONDITIONAL PROBABILITY



given this DAG... P(A,H,R,Z) factorizes as:

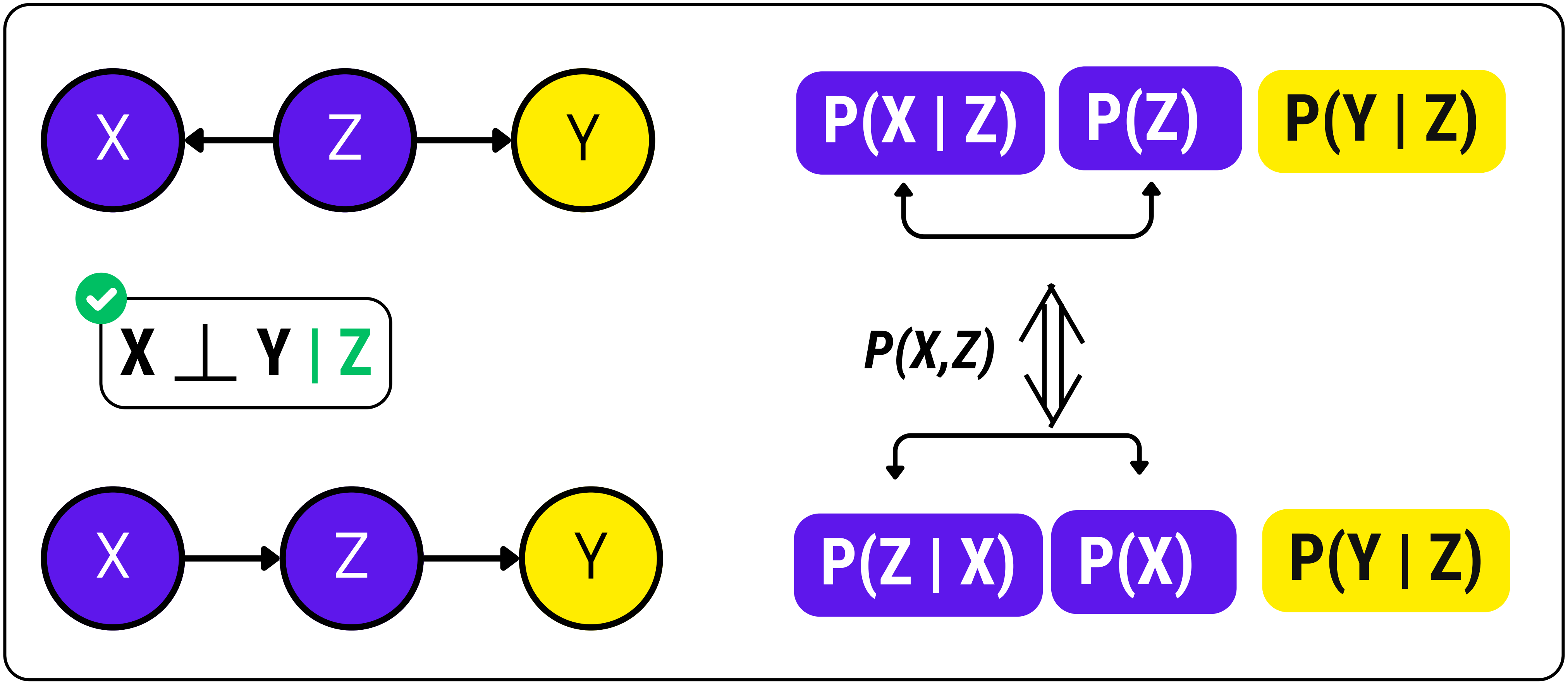
two arrows: ZIP and Age		*one arrow: Age*		*no arrows point at them*		
P(R | A, Z)	*	**P(H | A)**	*	**P(Z)**	*	**P(A)**
conditional probability		***conditional probability***		***marginal probabilities***		

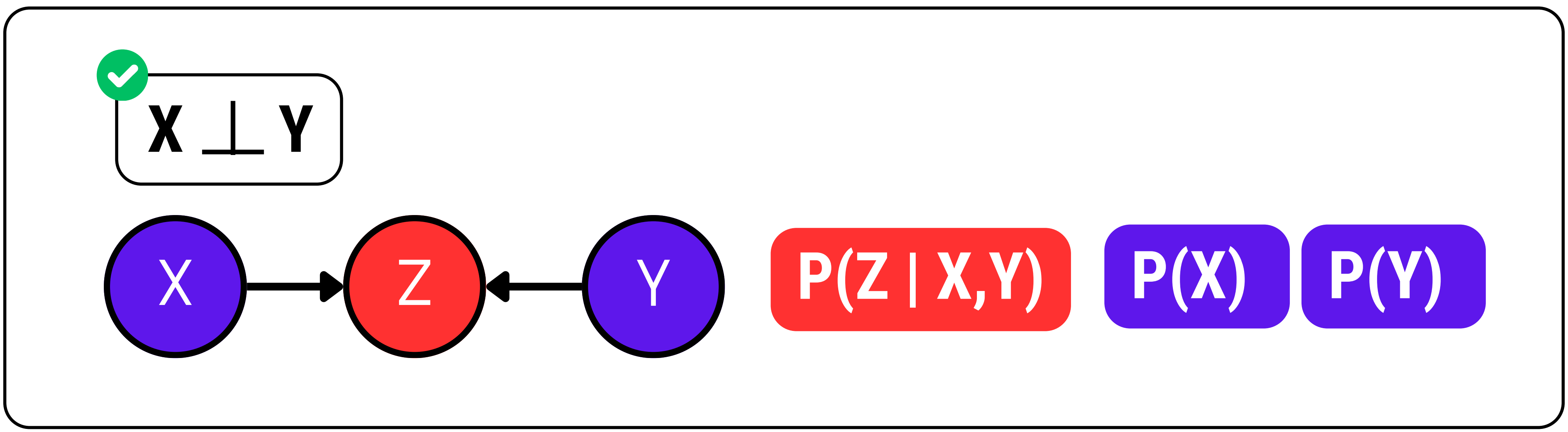


$H \perp R \mid A$

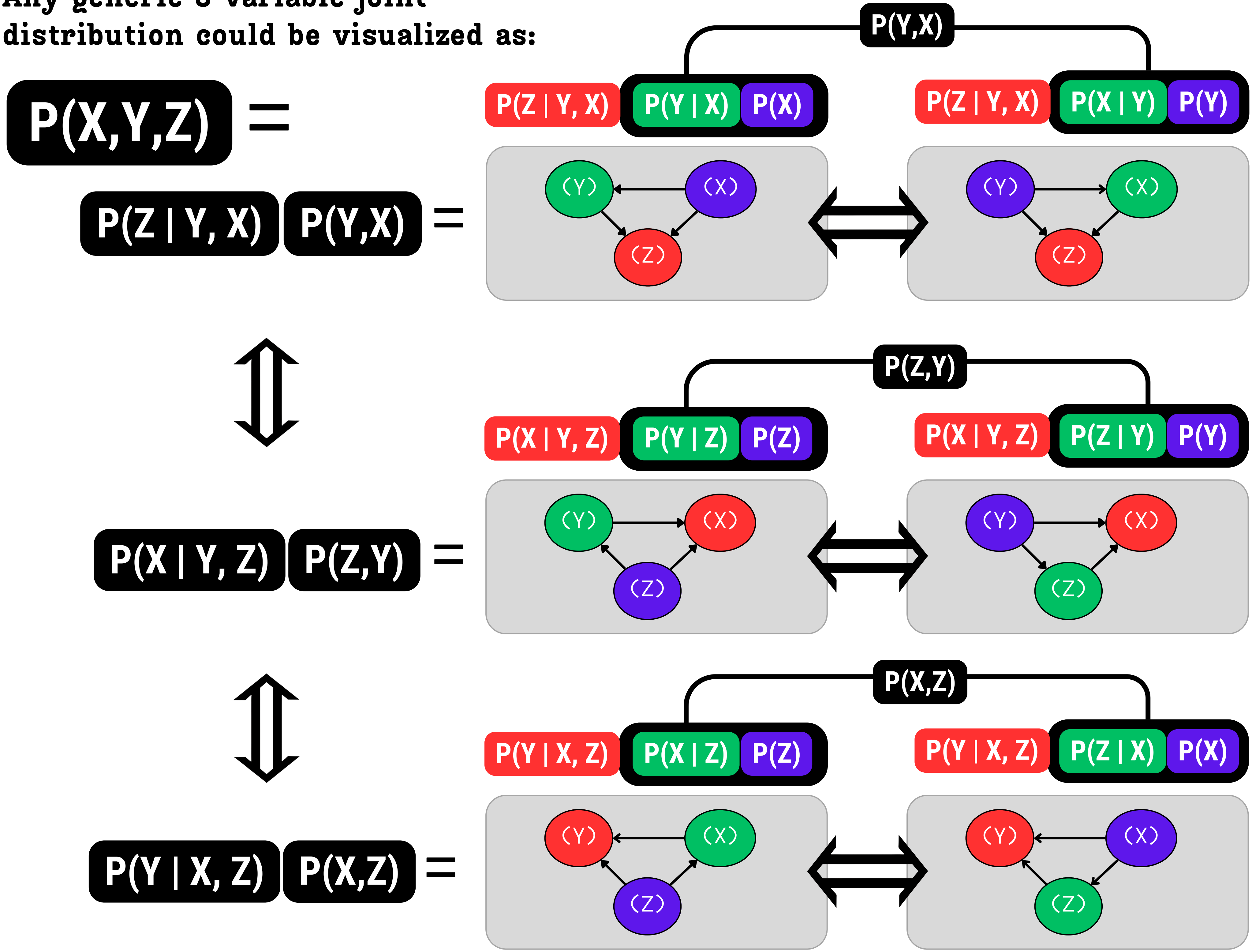
*I think (on average) height **could** reasonably tell us about reading levels...*

*...but **largely** **because of its relationship with age.***


X
Z
Y
P(X | Z)
P(Z)
P(Y | Z)
X ⊥ Y | Z
P(X,Z)
X
Z
Y
P(Z | X)
P(X)
P(Y | Z)


X ⊥ Y
X
Z
Y
P(Z | X,Y)
P(X)
P(Y)

Any generic 3-variable joint distribution could be visualized as:



In certain cases, the factorization can simplify to:

$$P(Z \mid Y, X)\; P(X)\; P(Y)$$

...but ONLY *when the joint distributions of what we are conditioning on are* **INDEPENDENT**

ONLY if this holds

$$P(Y,X) \iff P(X \mid Y)\, P(Y) \overset{?}{\iff} P(X)\, P(Y)$$

$$\Updownarrow$$

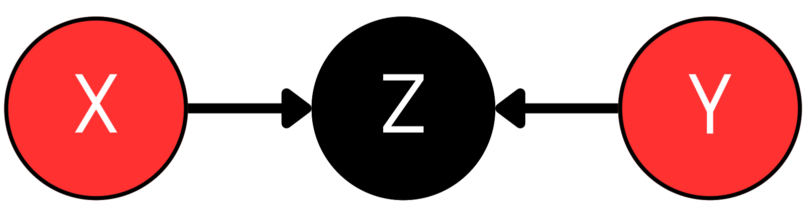
$$P(Y \mid X)\, P(X)$$

this is <u>always</u> true

this is <u>NOT</u> always true

This DAG is telling us that it does:

$$X \perp Y$$



$$= P(Z \mid X, Y)\; P(X)\; P(Y)$$