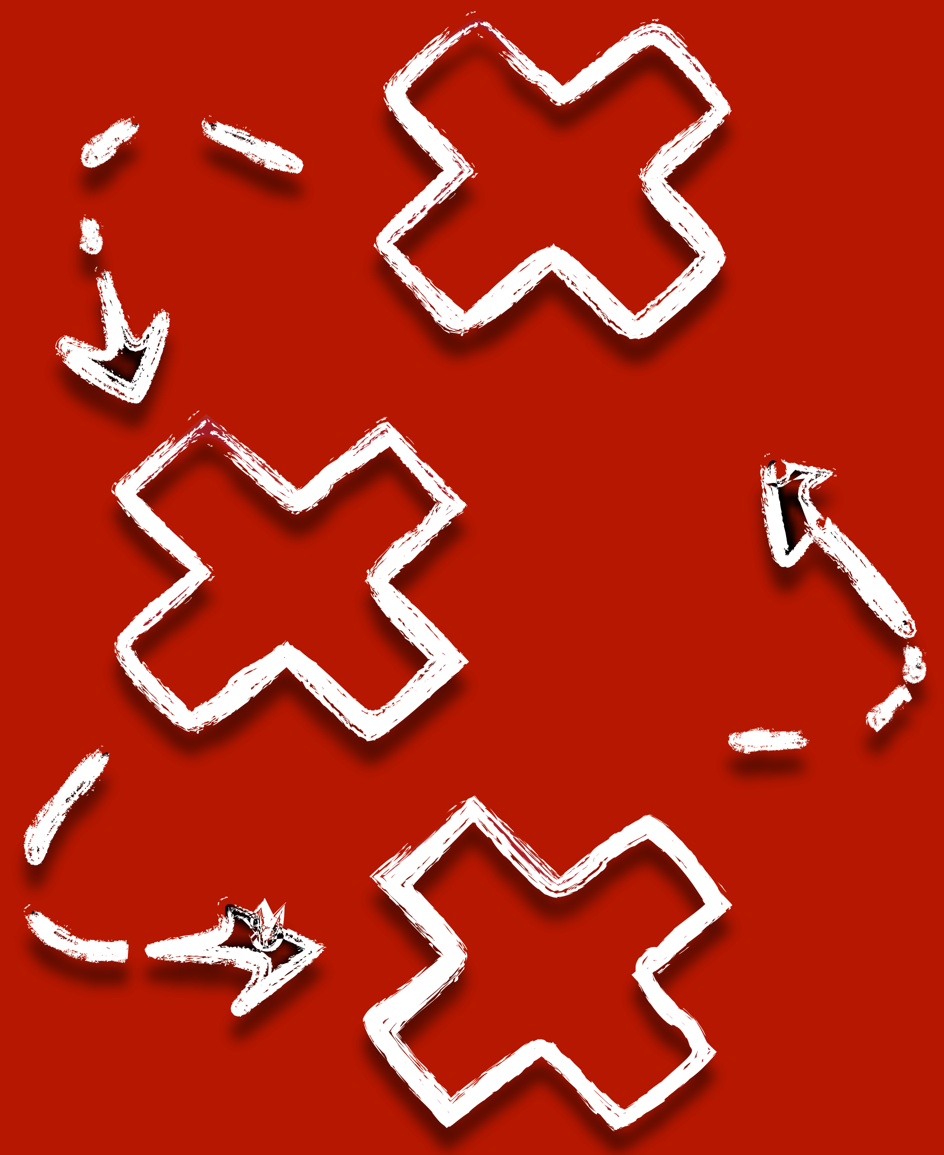


Causal Data Science

Lecture 4:2 Structural causal models

Lecturer: Sara Magliacane

UvA - Spring 2023

Truncated factorisation formula [Pearl 2009]

- If for any $\mathbf{W} \subset \mathbf{V}$:

$$p(X_{\mathbf{V}} \mid do(X_{\mathbf{W}} = x_{\mathbf{W}})) = \underbrace{\prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)})}_{\text{doesn't change from the observational distr.}} \cdot \mathbb{1}(X_{\bar{W}} = x_{\bar{W}})$$

Truncated factorisation formula [Pearl 2009]

- If for any $\mathbf{W} \subset \mathbf{V}$:

$$p(X_{\mathbf{V}} \mid \mathrm{do}(X_{\mathbf{W}} = x_{\mathbf{W}})) = \underbrace{\prod_{i \in \bar{V} \setminus \bar{W}} p(X_i \mid X_{pa(i)})}_{\text{doesn't change from the observational distr.}} \cdot \mathbb{1}(X_{\bar{W}} = x_{\bar{W}})$$

MODULARITY ASSUMPTION

Causal mechanisms and Modularity

- In a causal BN (G, p), each $p(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$ is the **causal mechanism of X_i**
- **Modularity assumption:** intervening on X_j will not change any causal mechanism $p(X_i \mid \mathbf{X}_{\mathrm{pa}(i)})$ for any $i \neq j$

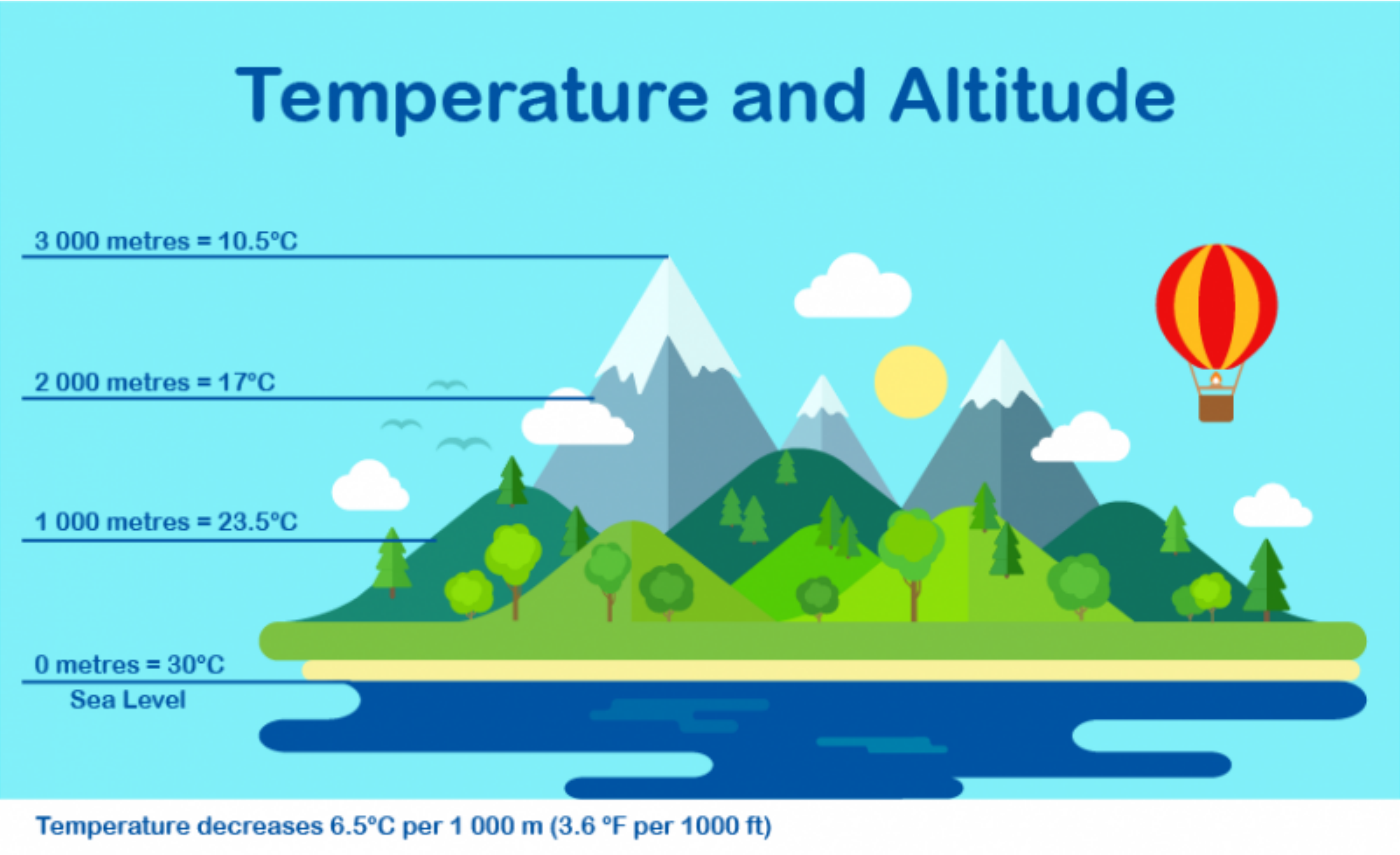
Causal mechanisms and Modularity

- In a causal BN (G, p), each $p(X_i \mid \mathbf{X}_{\mathrm{Pa}(i)})$ is the **causal mechanism of X_i**
- **Modularity assumption:** intervening on X_j will not change any causal mechanism $p(X_i \mid \mathbf{X}_{\mathrm{Pa}(i)})$ for any $i \neq j$
- **Independent Causal Mechanism Principle:** the generative process is composed of **autonomous models** that do **not inform or influence** each other

Knowing **Changing**	$p(X_j \mid \mathbf{X}_{\mathrm{Pa}(j)})$	Does not give info **Does not change**	$p(X_i \mid \mathbf{X}_{\mathrm{Pa}(i)})$ $i \neq j$

Section 2.1 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)

Modularity example

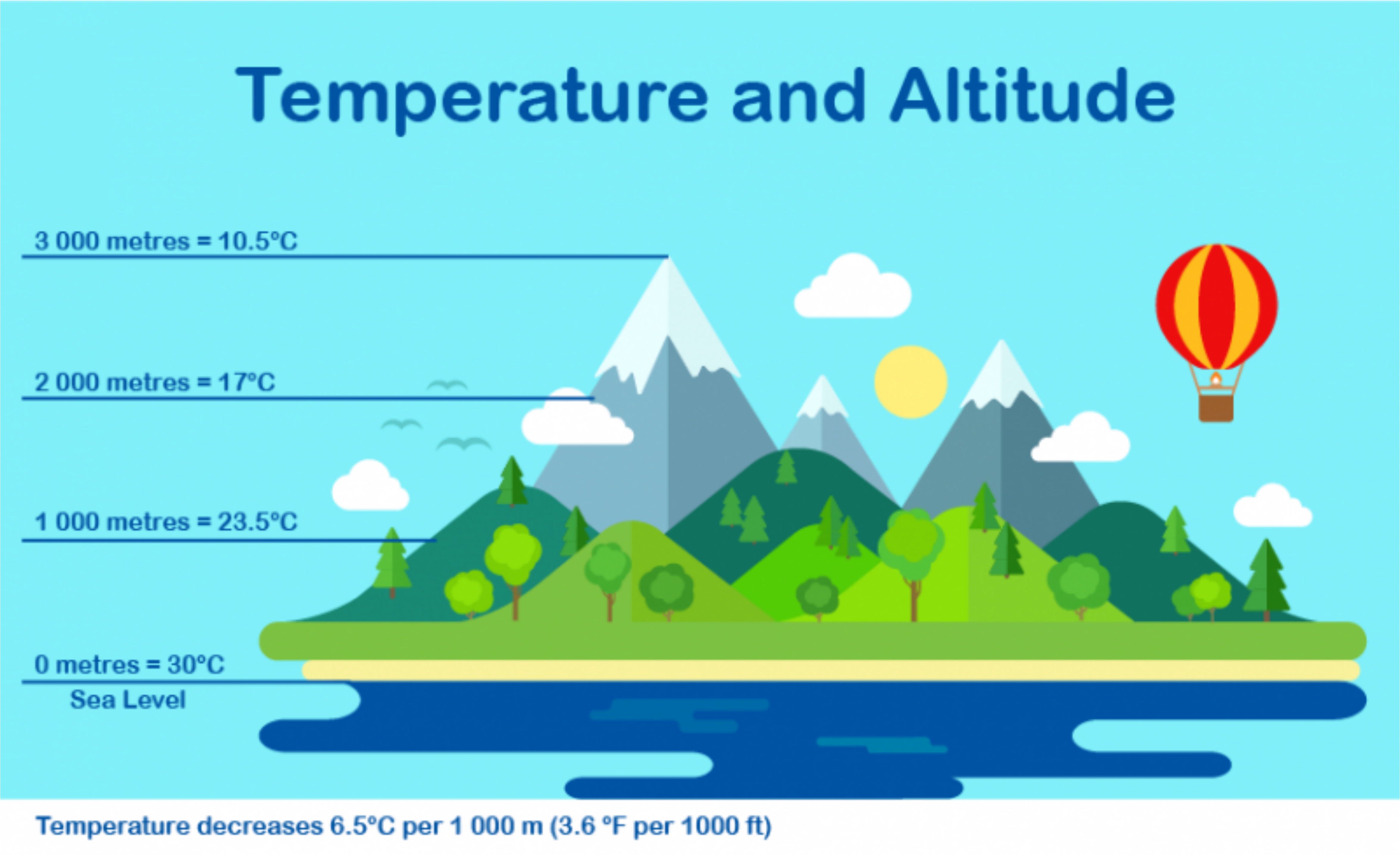


$$P(A,T) = P(T\,|\,A)P(A)$$

$$P(A,T) = P(A\,|\,T)P(T)$$

Example from Section 2.1 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)
Image from https://lisbdnet.com/how-altitude-affects-climate/

Modularity example



$$P(A, T) = P(T \mid A)P(A)$$

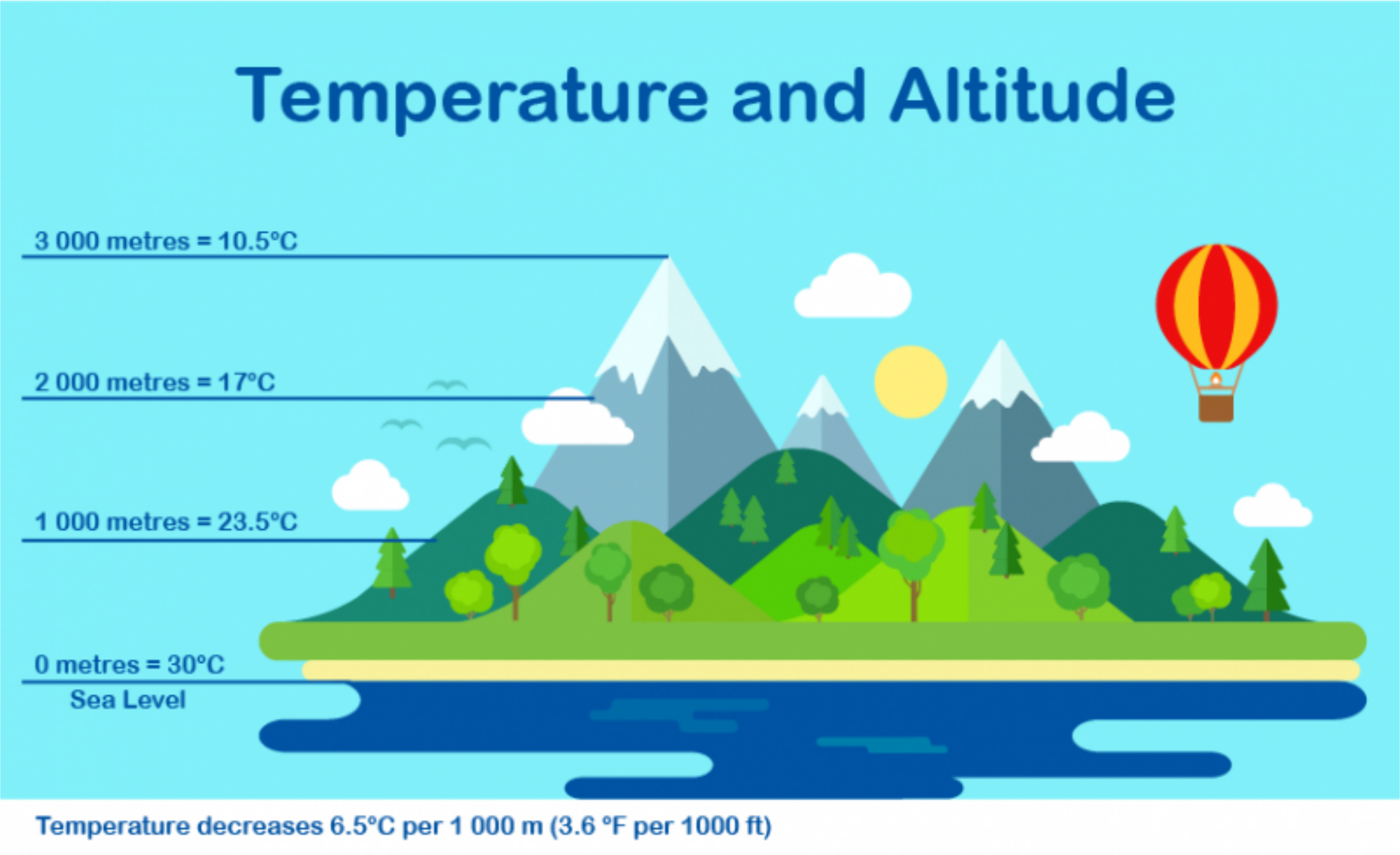
changing P(A) does not change P(T | A)

$$P(A, T) = P(A \mid T)P(T)$$

Changing P(T) might change P(A | T)

Example from Section 2.1 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)
Image from https://lisbdnet.com/how-altitude-affects-climate/

Modularity example



P(T|A) is an invariant physical mechanism

$$P(A, T) = P(T \mid A)P(A)$$

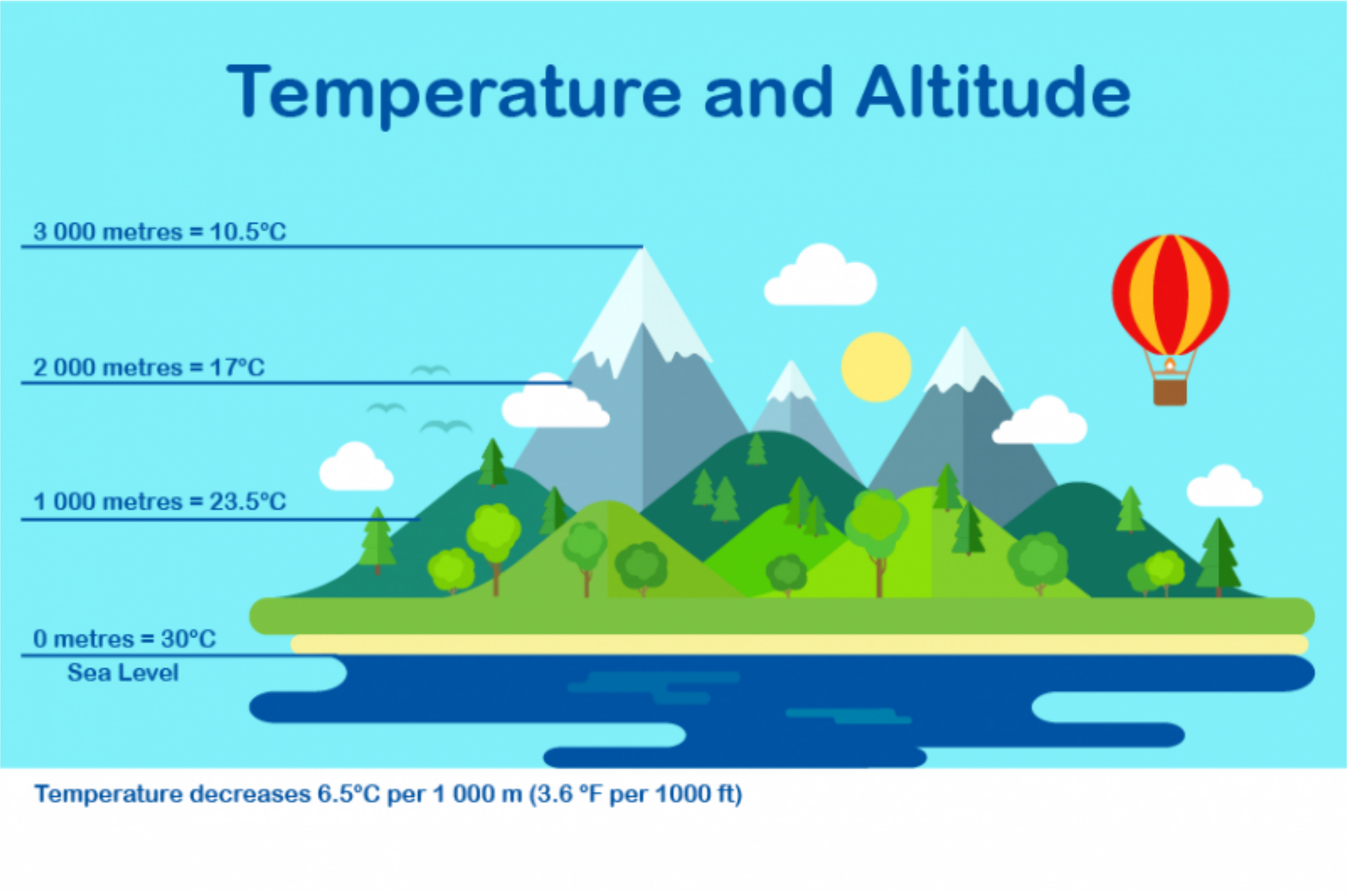
changing P(A) does not change P(T|A)

$$P(A, T) = P(A \mid T)P(T)$$

Changing P(T) might change P(A|T)

Example from Section 2.1 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)
Image from https://lisbdnet.com/how-altitude-affects-climate/

Modularity example



$$P(A, T) = P(T \mid A)P(A)$$

changing P(A) does not change P(T | A)

The **causal factorisation** allows for **localised/sparse** interventions

Example from Section 2.1 in Elements of Causal Inference (http://web.math.ku.dk/~peters/jonas_files/ElementsOfCausalInference.pdf)
Image from https://lisbdnet.com/how-altitude-affects-climate/

Structural equation models (SEMs)

- Let (G, p) be a Bayesian network

- We can write each variable X_i for $i \in \mathbf{V}$ as a **function of its parents** in G and a **noise term** ϵ_i in a **structural equation:**

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i)$$

- We assume all noises are **independent of each other** $\forall i \neq j : \epsilon_i \perp\!\!\!\perp \epsilon_j$

Structural equation models (SEMs)

- Let (G, p) be a Bayesian network
- We can write each variable X_i for $i \in \mathbf{V}$ as a **function of its parents** in G and a **noise term** ϵ_i in a **structural equation:**

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i)$$

often linear often Gaussian

- We assume all noises are **independent of each other** $\forall i \neq j : \epsilon_i \perp\!\!\!\perp \epsilon_j$

Structural causal models (SCMs)

- Let (G, p) be a **causal** Bayesian network
- We can write each variable X_i for $i \in \mathbf{V}$ as a **function of its parents** in G and a **noise term** ϵ_i in a **structural equation:**

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i)$$

- We assume all noises are **independent of each other** $\forall i \neq j : \epsilon_i \perp\!\!\!\perp \epsilon_j$

Structural causal models (SCMs)

- Let (G, p) be a causal Bayesian network

- We can write each variable X_i for $i \in \mathbf{V}$ as a **function of its parents** in G and a **noise term** ϵ_i in a **structural equation:**

causal mechanism

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i)$$

endogenous variable

V

using ← instead of = to emphasize asymmetry

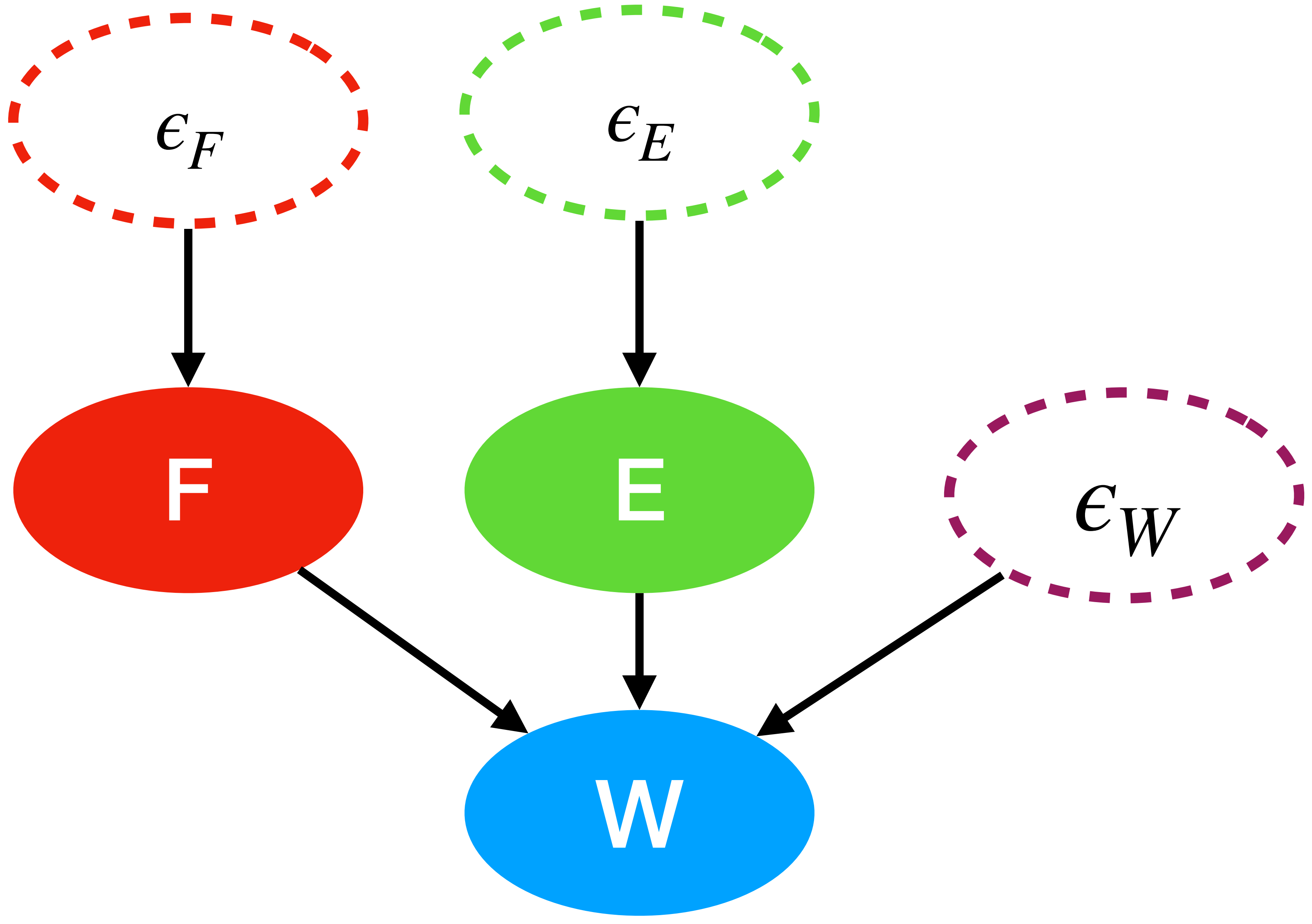
exogenous variable

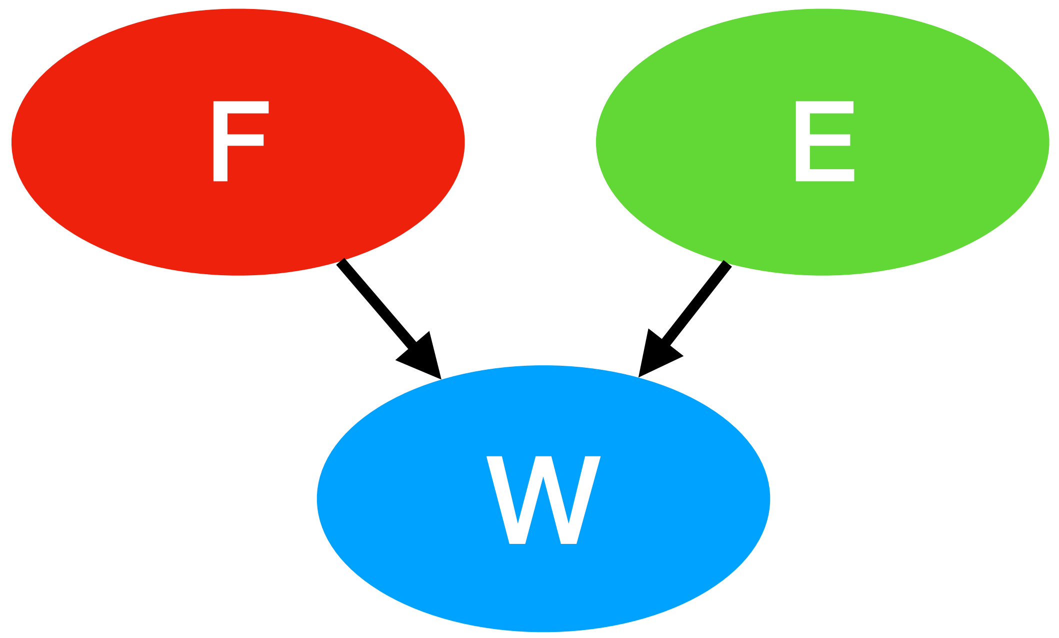
U

Example SCM

$$\begin{cases} F \leftarrow 2000 + \epsilon_F \\ E \leftarrow 500 + \epsilon_E \\ W \leftarrow \dfrac{F - E - 1500}{7700} + 0.1\epsilon_W \end{cases}$$

$$\epsilon_E, \epsilon_F, \epsilon_W \sim \mathcal{N}(0,100)$$





Interventions in SCMs

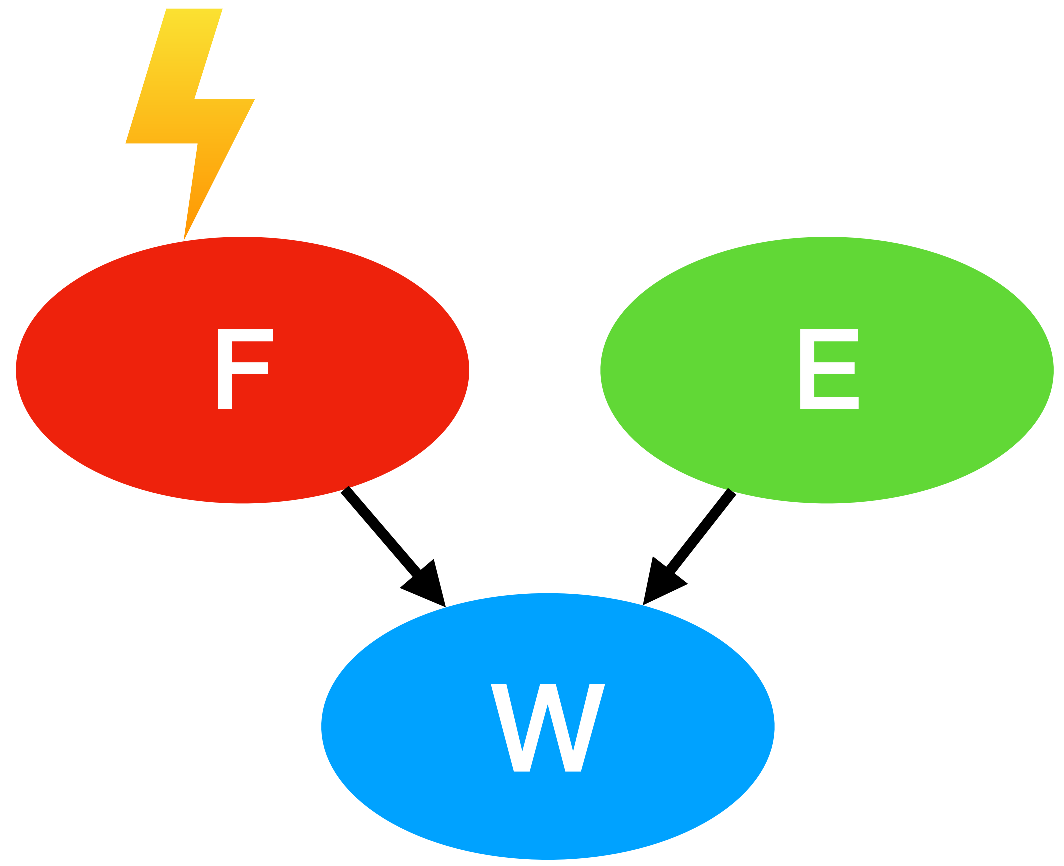
- An intervention $\mathrm{do}(X_i = x_i)$ can be modelled by replacing

$$X_i \leftarrow h_i(X_{\mathrm{Pa}(i)}, \epsilon_i) \text{ with } X_i \leftarrow x_i$$

$\mathrm{do}(F=1200)$

$$\begin{cases} F \leftarrow \cancel{2000} + \epsilon_F \\ E \leftarrow 500 + \epsilon_E \\ W \leftarrow \dfrac{F - E - 1500}{7700} + 0.1\epsilon_W \end{cases}$$

$F \leftarrow 1200$



$$\epsilon_E, \epsilon_F, \epsilon_W \sim \mathcal{N}(0,100)$$

Example 3.2 in Elements of Causal Inference

$$\begin{cases} X \leftarrow \epsilon_X \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

Example 3.2 in Elements of Causal Inference

$$\begin{cases} X \leftarrow \epsilon_X \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$P(X) = \mathcal{N}(0,1)$$

Example 3.2 in Elements of Causal Inference

$$\begin{cases} X \leftarrow \epsilon_X \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

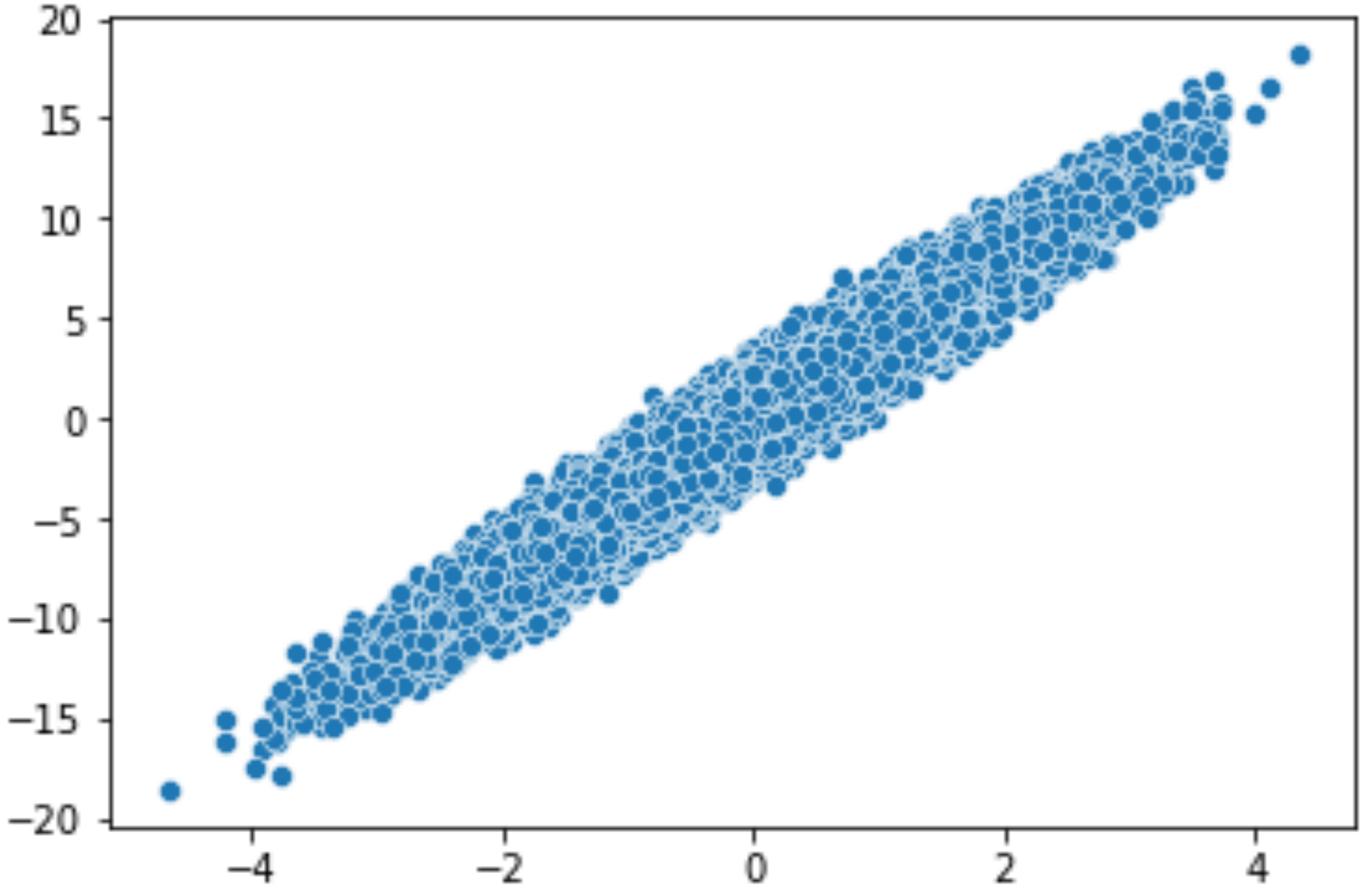
$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$P(X) = \mathcal{N}(0,1)$$

$$P(Y) = 4 \cdot \mathcal{N}(0,1) + \mathcal{N}(0,1) = \mathcal{N}(0,17)$$

```
x = randn(n_samples)
y = 4 * x + randn(n_samples)
# plot P(X,Y)
sns.scatterplot(x=x,y=y)
```

<AxesSubplot:>



Example 3.2 in Elements of Causal Inference

$$\begin{cases} X \leftarrow \epsilon_X \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$P(X) = \mathcal{N}(0,1)$$

$$P(Y) = \mathcal{N}(0,17)$$

do(X=2):

$$\begin{cases} X \leftarrow 2 \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$P(X \mid do(X=2)) = \begin{cases} 1 & X=2 \\ 0 & X \neq 2 \end{cases}$$

Example 3.2 in the SCM Jupyter notebook

do(X=2):

$$\begin{cases} X \leftarrow 2 \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

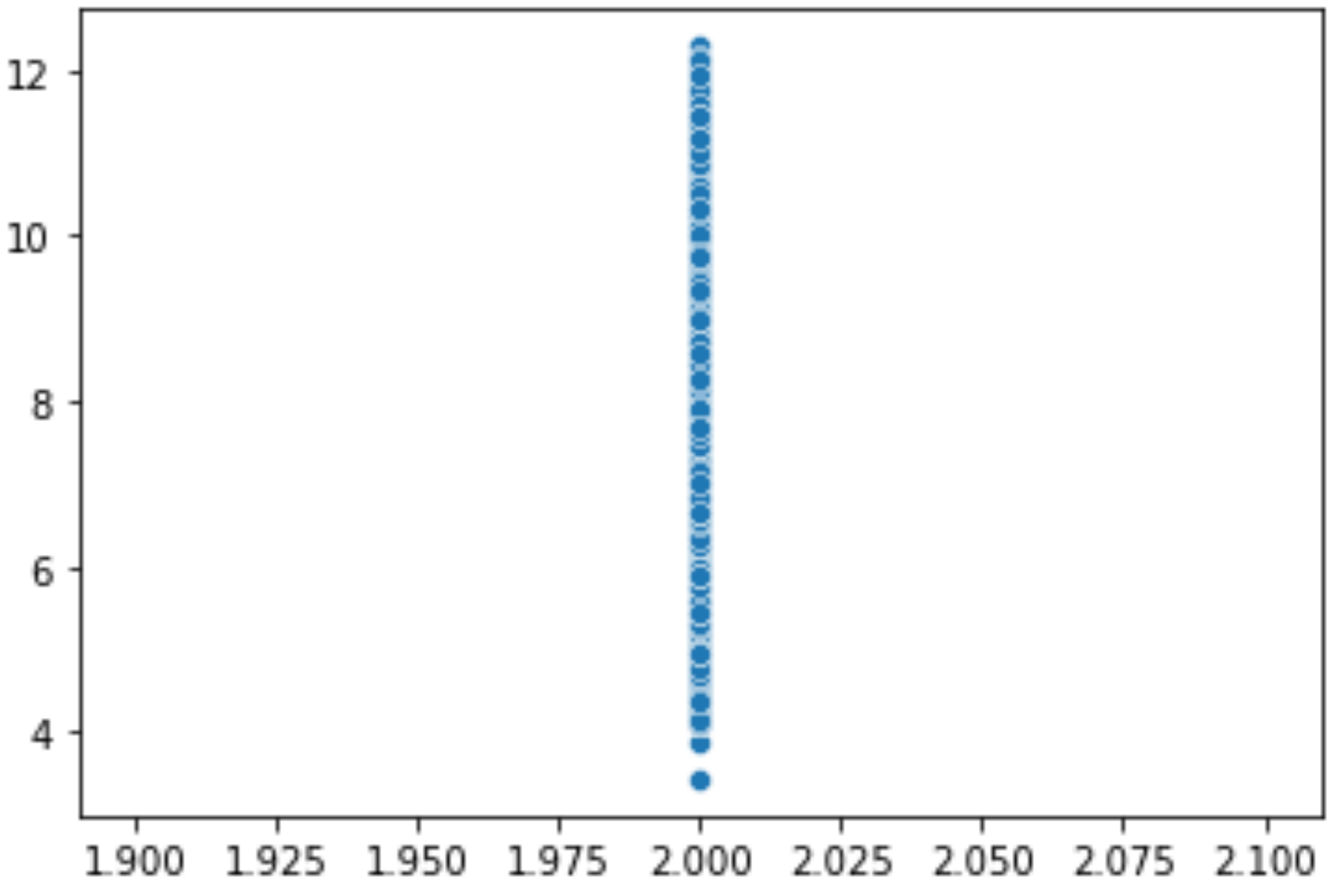
$$P(X \mid do(X=2)) = \begin{cases} 1 & X=2 \\ 0 & X \neq 2 \end{cases}$$

$$P(Y \mid do(X=2)) = N(8,1)$$

```python
x_do_x = np.array([2] * n_samples, dtype="int32")
y_do_x = 4 * x_do_x  + randn(n_samples)

# plot P(X,Y |do(X=2))
sns.scatterplot(x=x_do_x,y=y_do_x)
```

<AxesSubplot:>



A side note: perfect vs soft interventions

- We introduce a new operator that can represent a **hypothetical intervention** on the whole population, i.e. a perturbation of the system:

$$\mathrm{do}(X_i = x_i) \text{ which changes } P(X_i \mid X_{\mathrm{Pa}(i)}) \to \mathbf{1}(X_i = x_i)$$

- This is called a **perfect** (or surgical) **intervention**

- There are also other types of intervention, e.g. **soft interventions which change $P(X_i \mid X_{\mathrm{Pa}(i)}) \to P'(X_i \mid X_{\mathrm{Pa}(i)})$**

An example of soft interventions and shift interventions

$$\begin{cases} X \leftarrow \epsilon_X \\ Y \leftarrow 4 \cdot X + \epsilon_Y \end{cases}$$

$$\epsilon_X, \epsilon_Y \sim \mathcal{N}(0,1)$$

$$P(X) = \mathcal{N}(0,1)$$

$$P(Y) = \mathcal{N}(0,17)$$

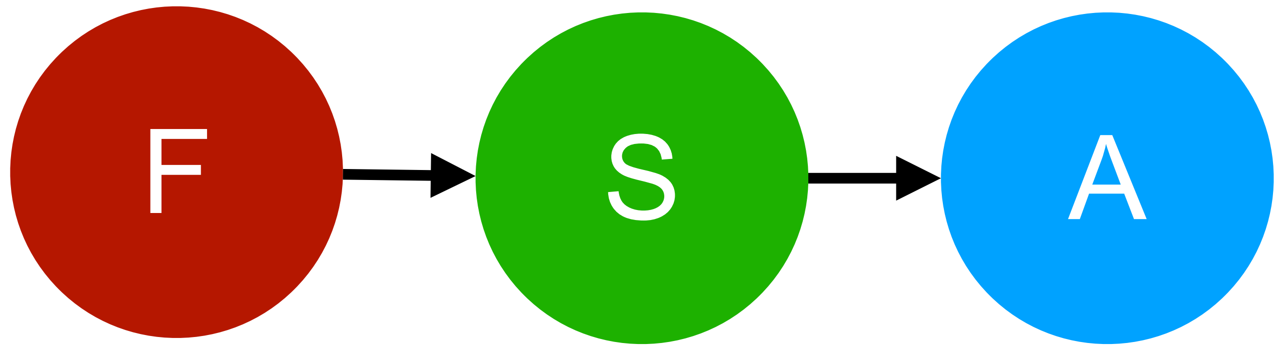
$$\begin{cases} X \leftarrow \epsilon_X \\ Y \leftarrow \cancel{4 \cdot X + \epsilon_Y} \end{cases}$$

$$Y \leftarrow 2 \cdot X + \epsilon_Y$$

$$\begin{cases} X \leftarrow \epsilon_X \\ Y \leftarrow 4 \cdot X + \epsilon_Y + \epsilon' \end{cases}$$

$$Pa_{G^{do(Y)}}(Y) \subseteq Pa_G(Y)$$

Structural causal model for exercise in Canvas



$$\begin{cases} F \leftarrow \varepsilon_F \\ S \leftarrow h_S(F, \varepsilon_S) \\ A \leftarrow h_A(S, \varepsilon_A) \end{cases}$$

$\varepsilon_F, \varepsilon_S, \varepsilon_A$ are all indep.

$P(F=1) = 0.01$

$P(S=1 \mid F=1) = 0.9$

$P(S=1 \mid F=0) = 0.2$

$P(A=1 \mid S=1) = 0.8$

$P(A=1 \mid S=0) = 0.1$

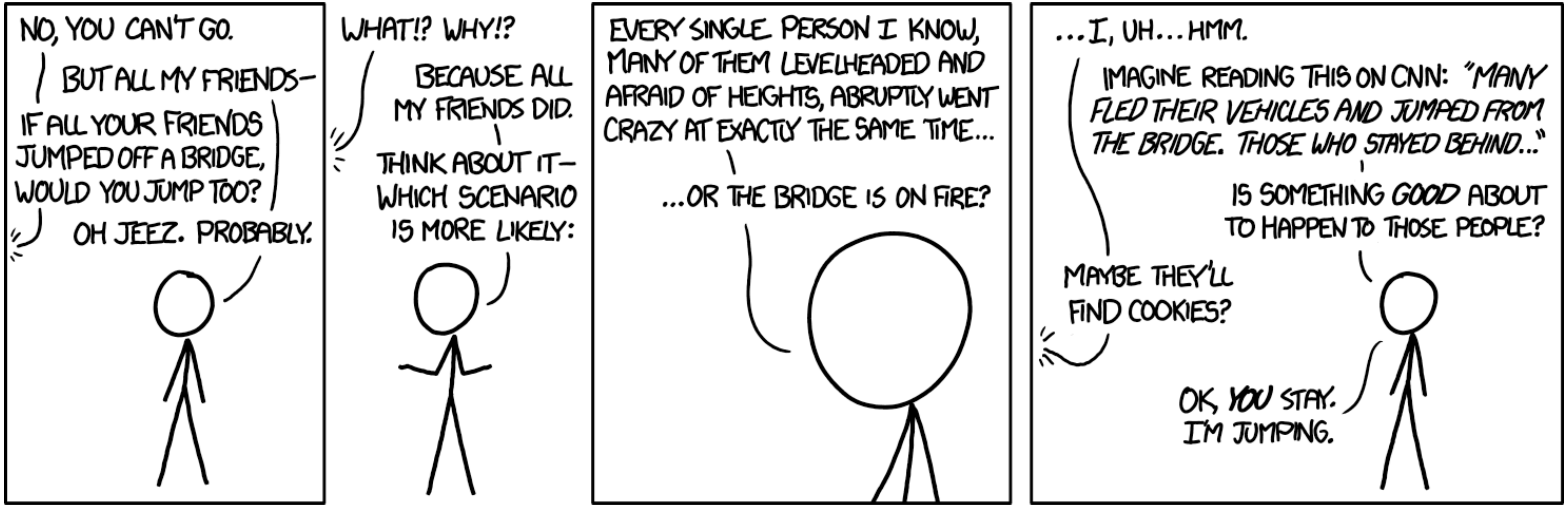
```
n = 1000

F = binomial(1, 0.01, n)
S = F * binomial(1, 0.9, n) + (1-F) * binomial(1, 0.2, n)
A = S * binomial(1, 0.8, n) + (1-S) * binomial(1, 0.1, n)

df = pd.DataFrame({"F":F, "S":S, "A": A})

P_S1 = len(df[S==1])/n
P_F1givenS1 = len(df[(df.F==1) & (df.S==1)])/ len(df[S==1])
```

Questions??



https://xkcd.com/1170/