Exercise: Umbrellas, Rain, and Happiness

We observe three binary variables:

- R: Rain (1 = rain, 0 = no rain)

- U: Umbrella (1 = carries umbrella, 0 = no umbrella)

- H: Happiness (1 = happy, 0 = unhappy)

The data are summarized in the tables below.

Given data

Table 1 — Overall (marginal data)

Umbrella U	Happy $H = 1$	Total	$P(H = 1 \mid U)$
1 (Yes)	20	100	0.20
0 (No)	180	200	0.90

Table 2 — Conditional on rain ($R = 1$)

Umbrella U	Happy $H = 1$	Total	$P(H = 1 \mid U, R = 1)$
1 (Yes)	20	80	0.25
0 (No)	2	20	0.10

Question 1 — Marginal association

1. Compute $P(H = 1 \mid U = 1)$ and $P(H = 1 \mid U = 0)$ using Table 1.

2. Is happiness higher or lower among people who carry umbrellas? Interpret the result.
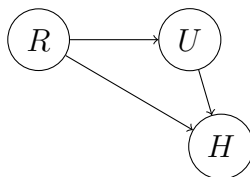
Question 2 — Conditional associations

1. Using Table 2, compare $P(H = 1 \mid U = 1, R = 1)$ and $P(H = 1 \mid U = 0, R = 1)$.

2. Does carrying an umbrella increase or decrease happiness?

Question 3 — Apparent contradiction

1. Explain why the conclusions from Questions 1 and 2 appear to contradict each other.

Question 4 — The causal graph

The data-generating process is described by the DAG below:

1. Identify all paths between U and H.

2. Is there a common cause of U and H? If yes, which variable?

Question 5 — d-separation

1. Are U and H d-separated marginally? Justify using paths.

2. Are U and H d-separated given R? Justify.

Question 6 — Factorization

Write the joint distribution $P(R, U, H)$ as a product of conditional distributions implied by the DAG.

Question 7 — Explaining the "paradox"

1. Using the factorization from Question 6, show that

$$P(H \mid U) = \sum_r P(H \mid U, r)\, P(r \mid U).$$

2. Explain how umbrellas can improve happiness within each rain condition, but appear harmful on average.

Question 8 — Take-home message

In one or two sentences, explain why the Markov properties of a DAG are essential for interpreting associations in data.