

Causal diagrams

Katherine Muller

Objectives

- Use scientific knowledge to guide your data analysis
- Learn a diagramming technique to model causation in a system

Useful resources:

YouTube playlist from Leslie Myint about causal inference and diagramming:

<https://www.youtube.com/playlist?list=PLtjTgbl6JvXZ-rrZ9FOLG37IWwoyR1GcF>

Rethinking lectures related to causal inference:

<https://www.youtube.com/watch?v=FdnMWdICdRs&list=PLDcUM9US4XdPz-KxHM4XHt7uUVGWWVSus&index=1>

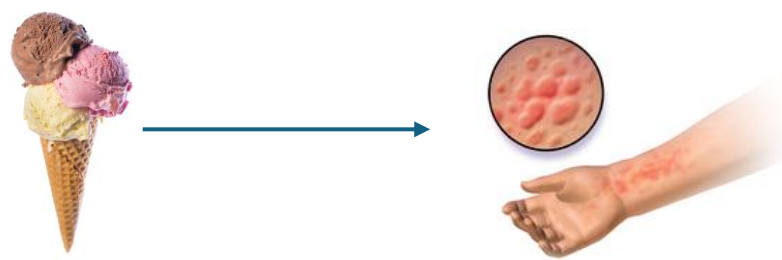
<https://www.youtube.com/watch?v=mBEA7PKDmiY&list=PLDcUM9US4XdPz-KxHM4XHt7uUVGWWVSus&index=5>

<https://www.youtube.com/watch?v=uanZZLlzKHw&list=PLDcUM9US4XdPz-KxHM4XHt7uUVGWWVSus&index=6>

What is causal inference?

- Process of determining cause-and-effect relationships in complex systems
- Principal used to design and analyze experimental and observational studies
- Approaches draw from information and graph theory

Association ≠ Causation

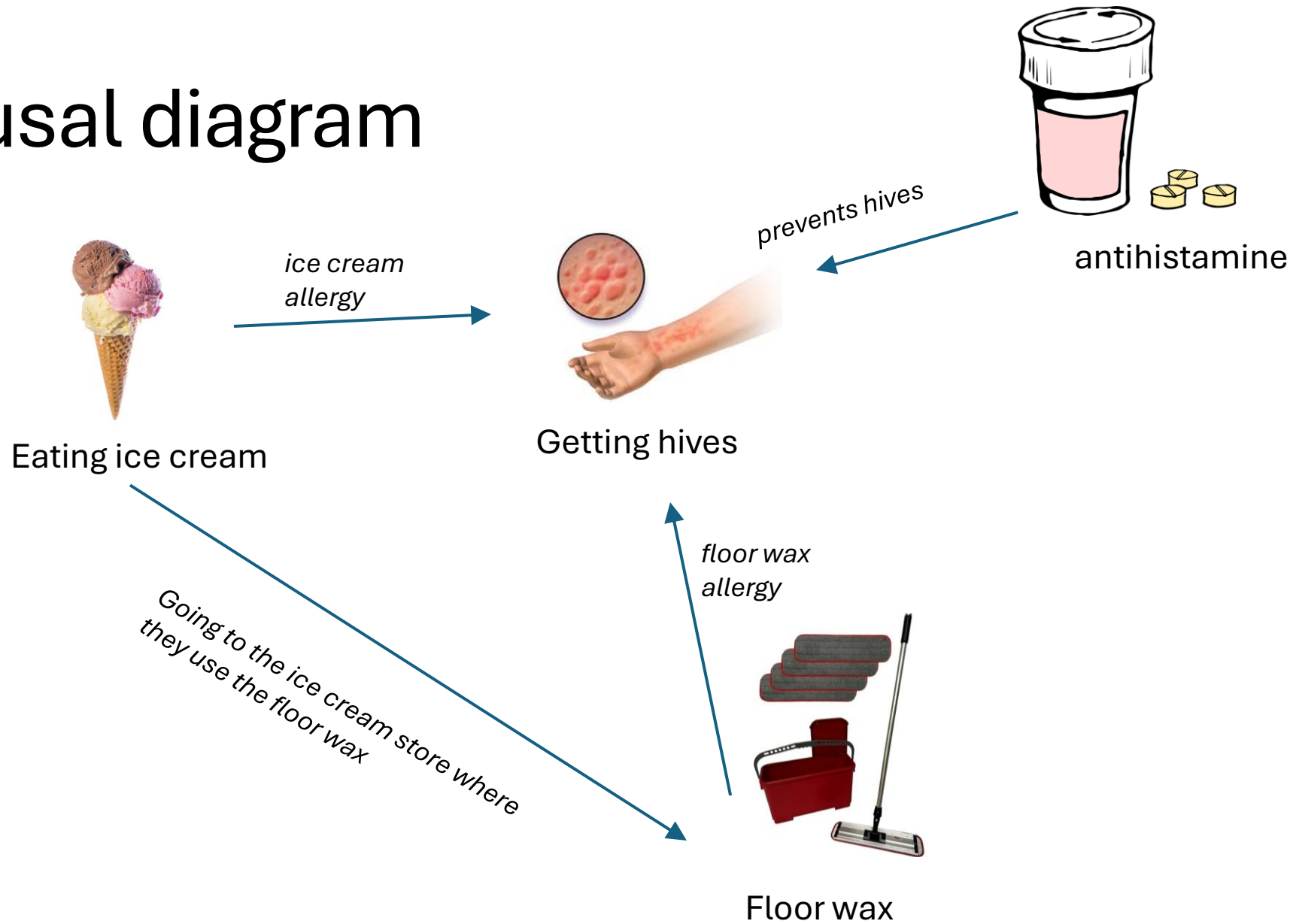


Eating ice cream

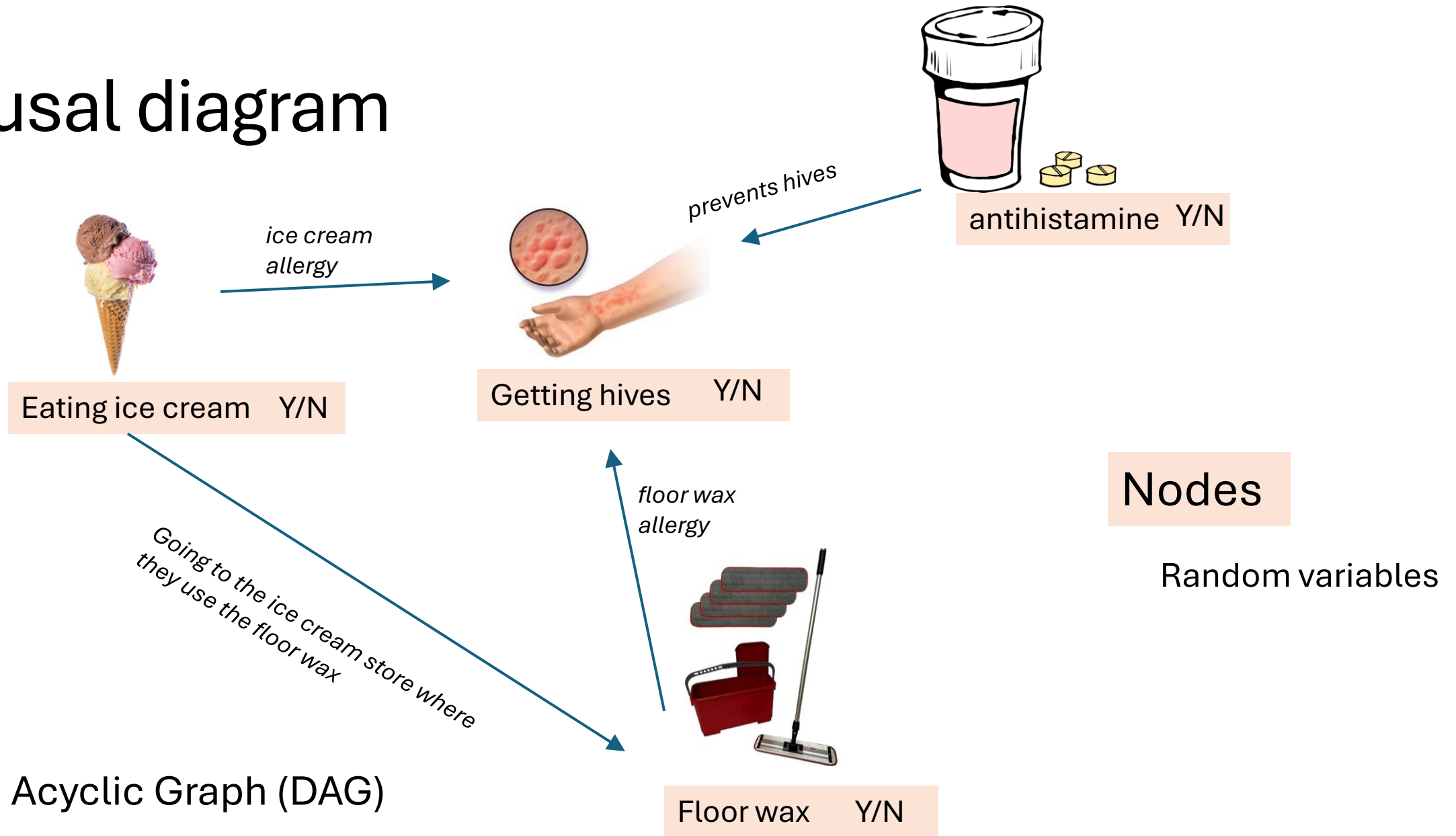
Getting hives

Association		Causation	
		Yes <i>I'm allergic to ice cream</i>	No <i>I'm not allergic to ice cream</i>
		Eating ice cream gives me hives.	I'm allergic to the floor wax in the ice cream shop.
		I am taking an antihistamine that prevents hives.	Eating ice cream doesn't give me hives.

Causal diagram

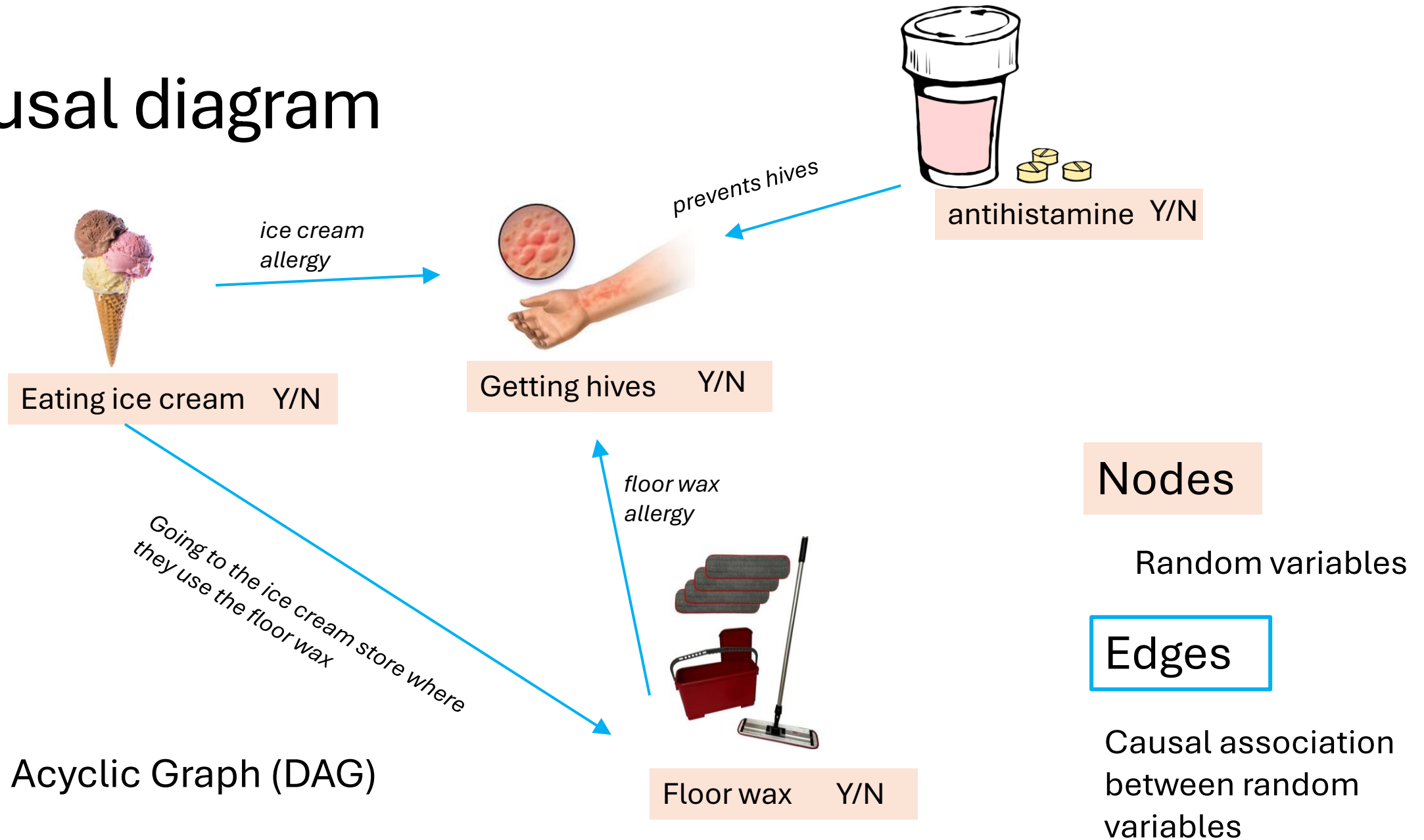


Causal diagram



Directed Acyclic Graph (DAG)

Causal diagram

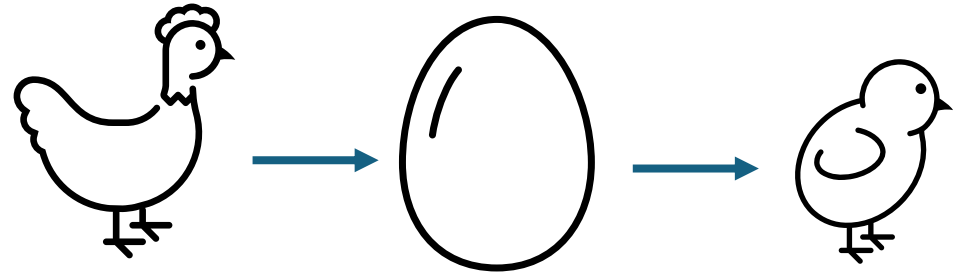
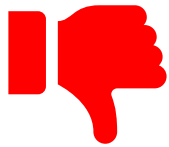


Directed Acyclic Graph (DAG)

= one direction of cause and effect

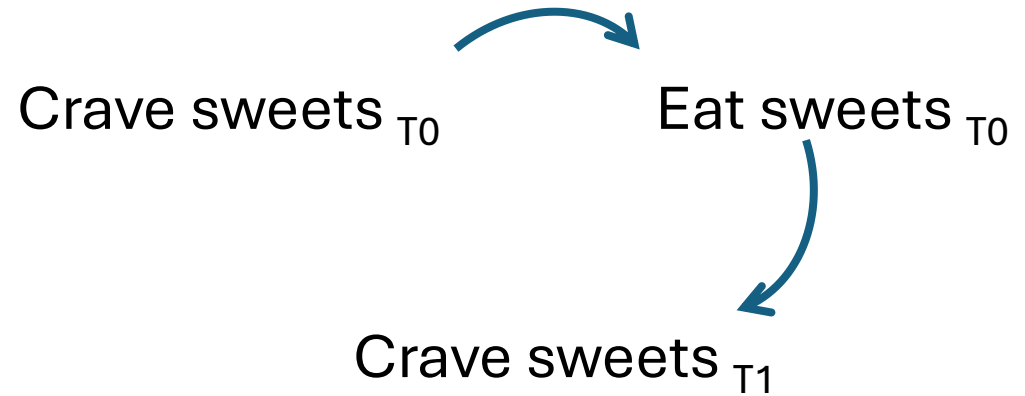
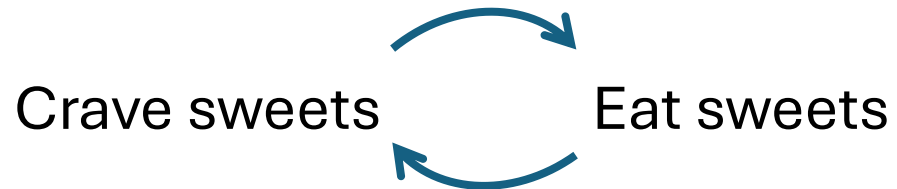


Undirected = two-way relationship

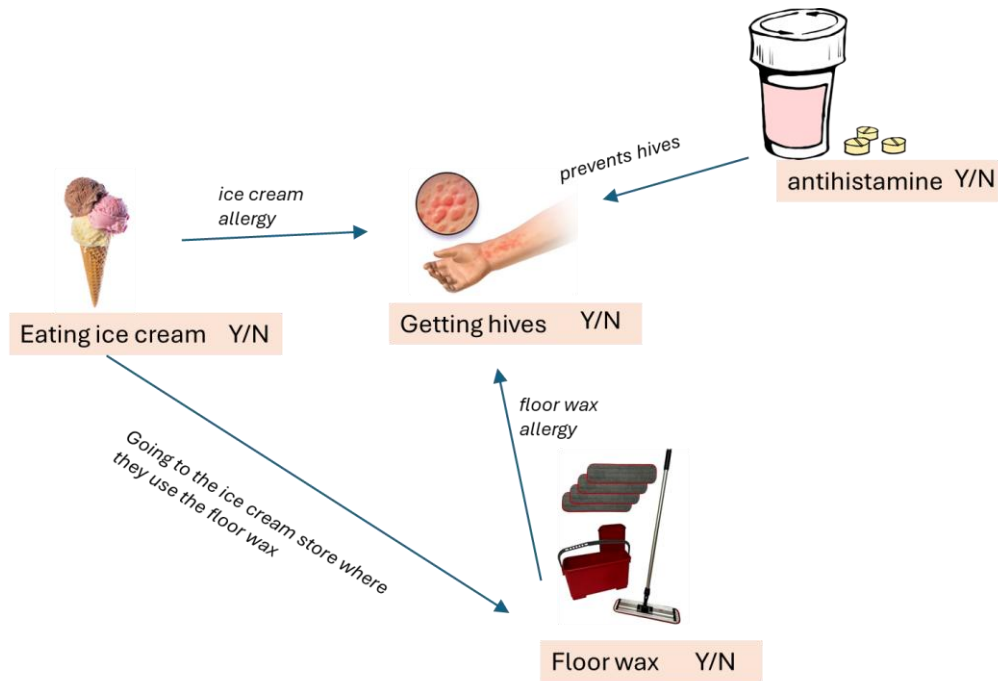


Directed Acyclic Graph (DAG)

= no feedback loops



Causal diagrams can help build statistical models



$\text{Prob}(\text{Getting hives}) = \text{Function}(\text{Eating ice cream}, \text{Floor wax}, \text{Antihistamine})$

Causal diagramming is helpful for designing models for complex systems

Birth weight and weaning weight of Dorper x Red Maasi lambs

ilri.sheep dataset

agridat package

How does genotype
influence birth weight,
weaning weight, and
weaning age?



ILRI/Paul Karaimu

```
library(agridat)
?ilri.sheep
```

ilri.sheep {agridat}

Birth weight and weaning weight of Dorper x Red Maasi lambs

Description

Birth weight and weaning weight of 882 lambs from a partial diallel cross of Dorper and Red Maasi breeds.

Format

A data frame with 882 observations on the following 12 variables.

year

year of lamb birth, 1991-1996

lamb

lamb id

sex

sex of lamb, M=Male/F=Female

gen

genotype, DD, DR, RD, RR

birthwt

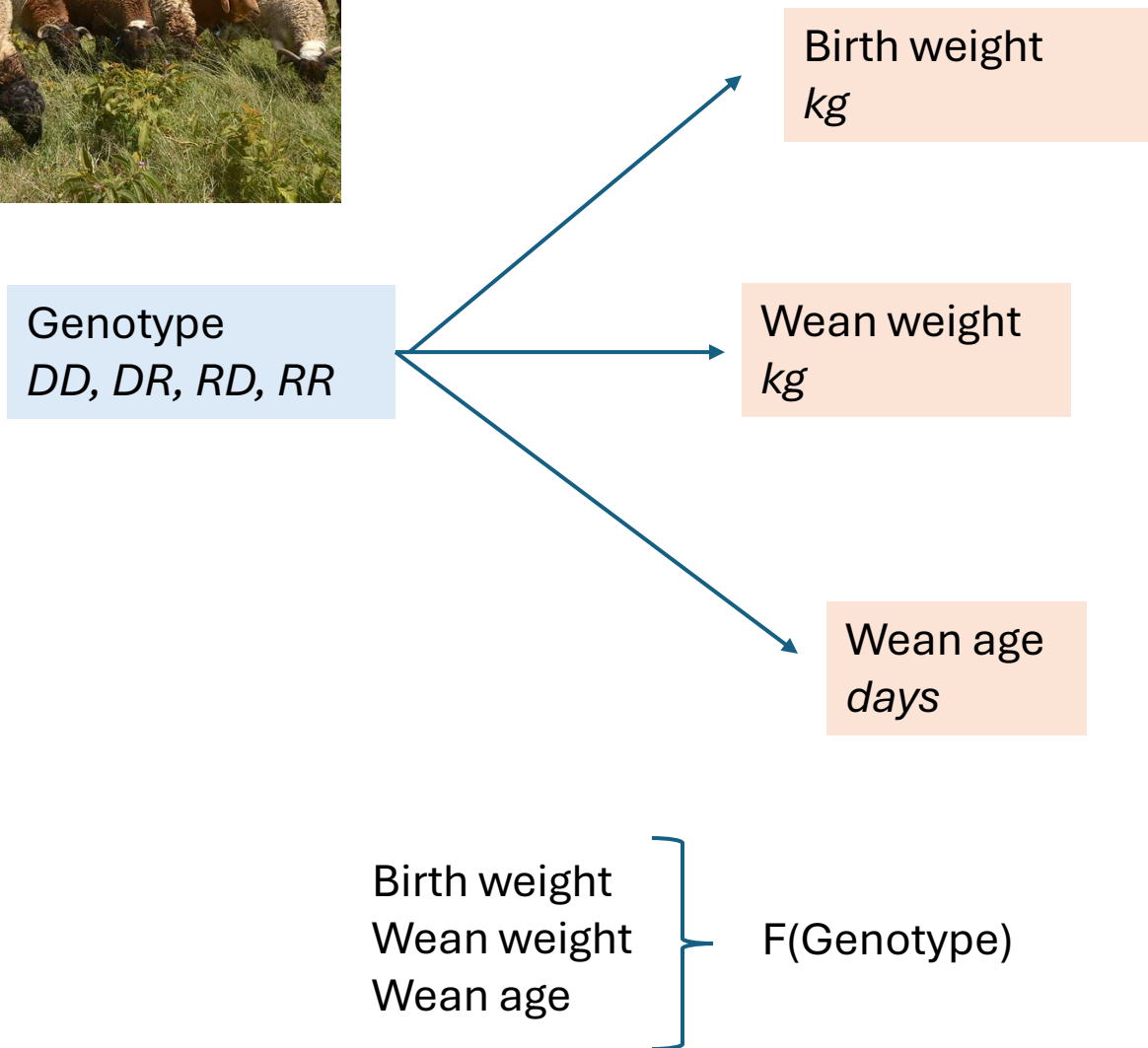
weight of lamb at birth, kg

weanwt

<https://www.rdocumentation.org/packages/agridat/versions/1.23/topics/ilri.sheep>



How does genotype influence birth weight, weaning weight, and weaning age?

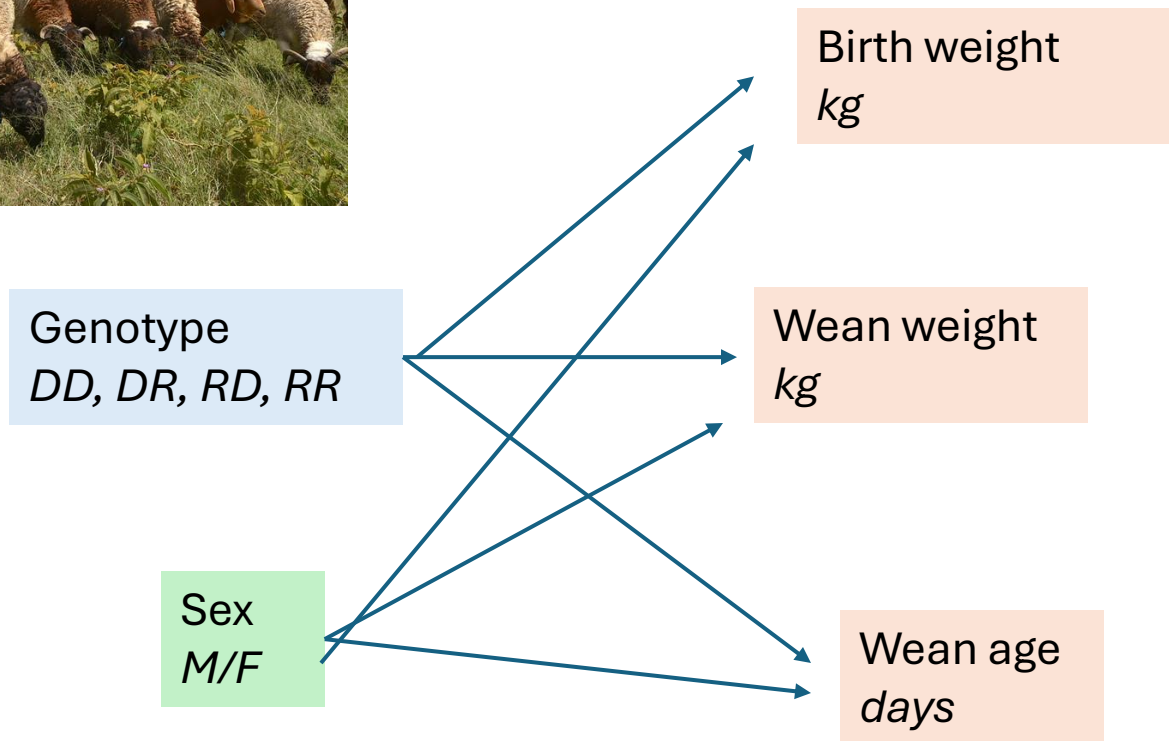


Treatment

Outcome



How does genotype influence birth weight, weaning weight, and weaning age?



Birth weight
Wean weight
Wean age

} F(Genotype, Sex)

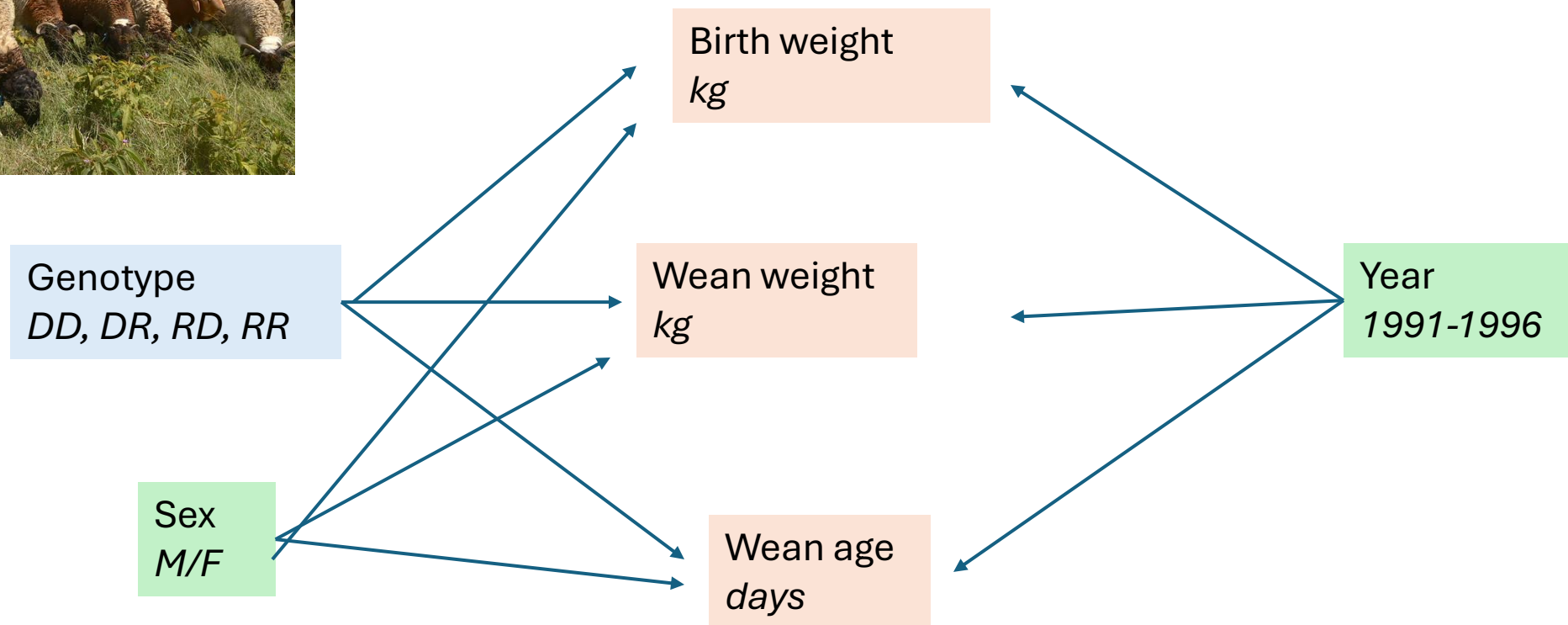
Treatment

Outcome

Covariate



How does genotype influence birth weight, weaning weight, and weaning age?



Birth weight
Wean weight
Wean age

} F(Genotype, Sex, Year)

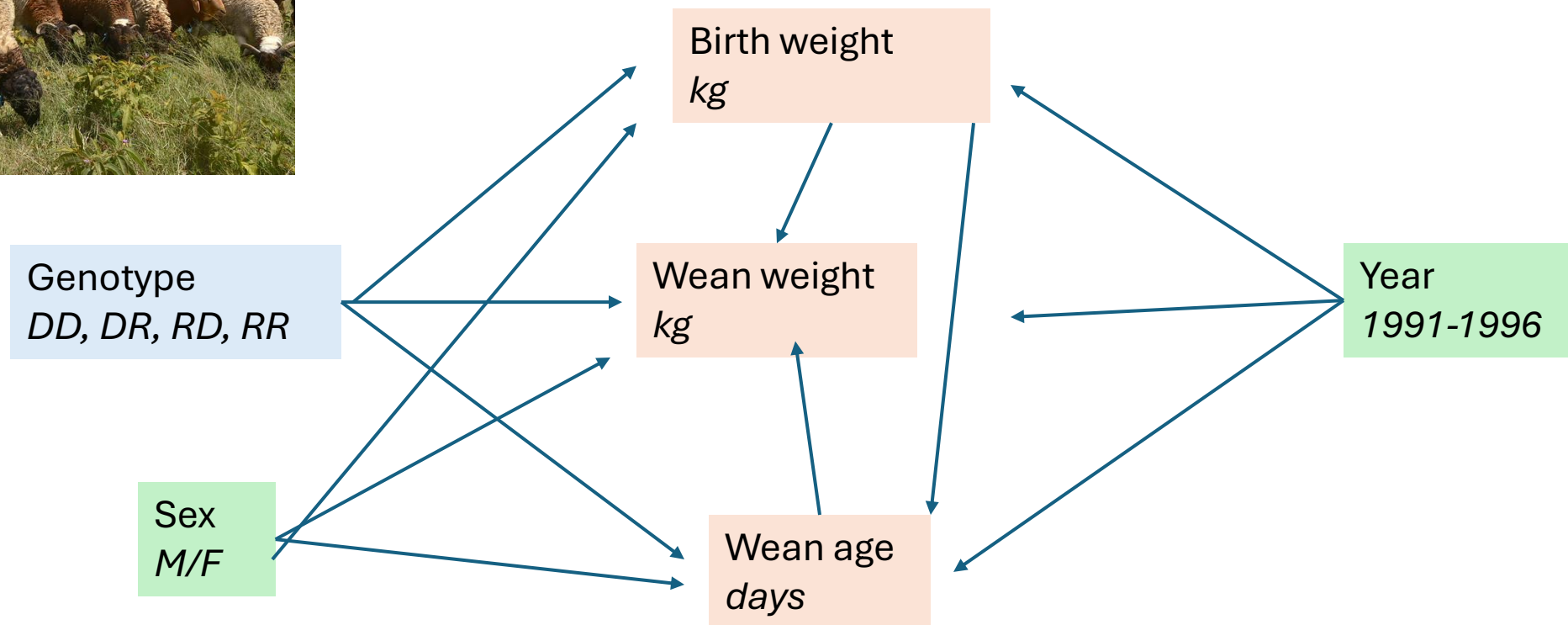
Treatment

Outcome

Covariate



How does genotype influence birth weight, weaning weight, and weaning age?



Birth weight } F(Genotype, Sex, Year)
 Wean weight } F(Genotype, Sex, Year, Birth weight, Wean age)
 Wean age } F(Genotype, Sex, Year, Birth weight)

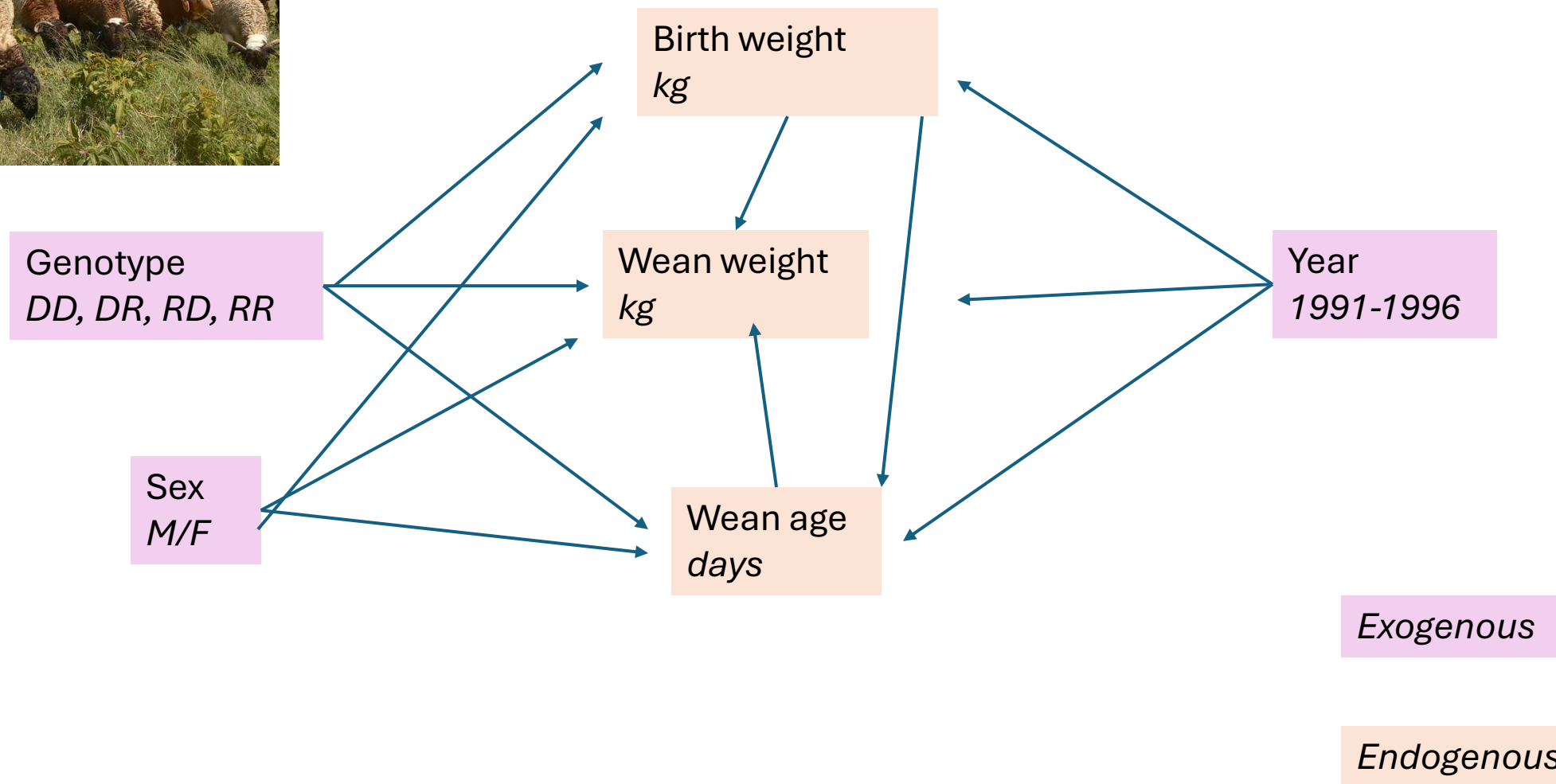
Treatment

Outcome

Covariate

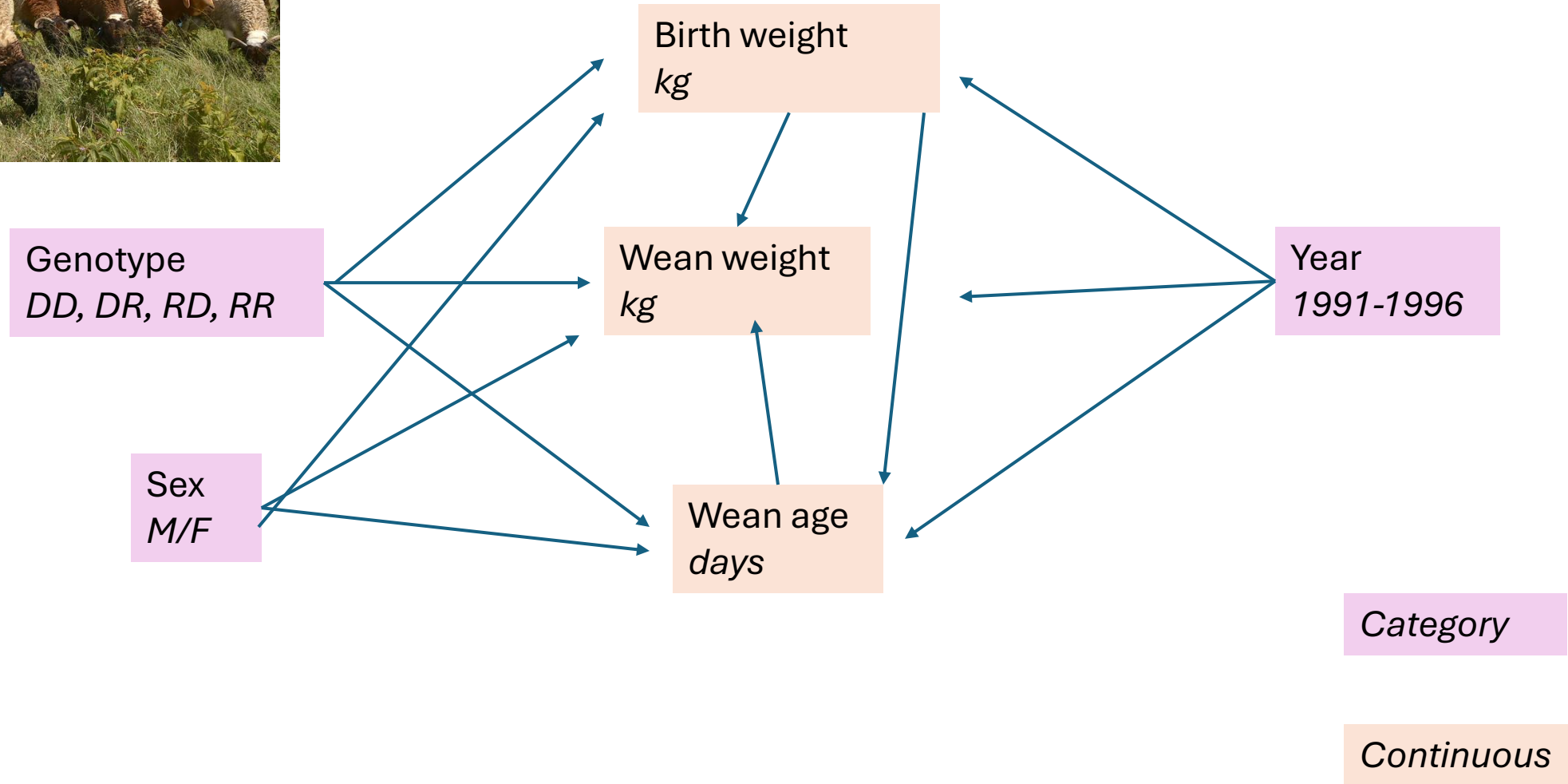


Treatment, covariate, and outcome depend on the question and study design



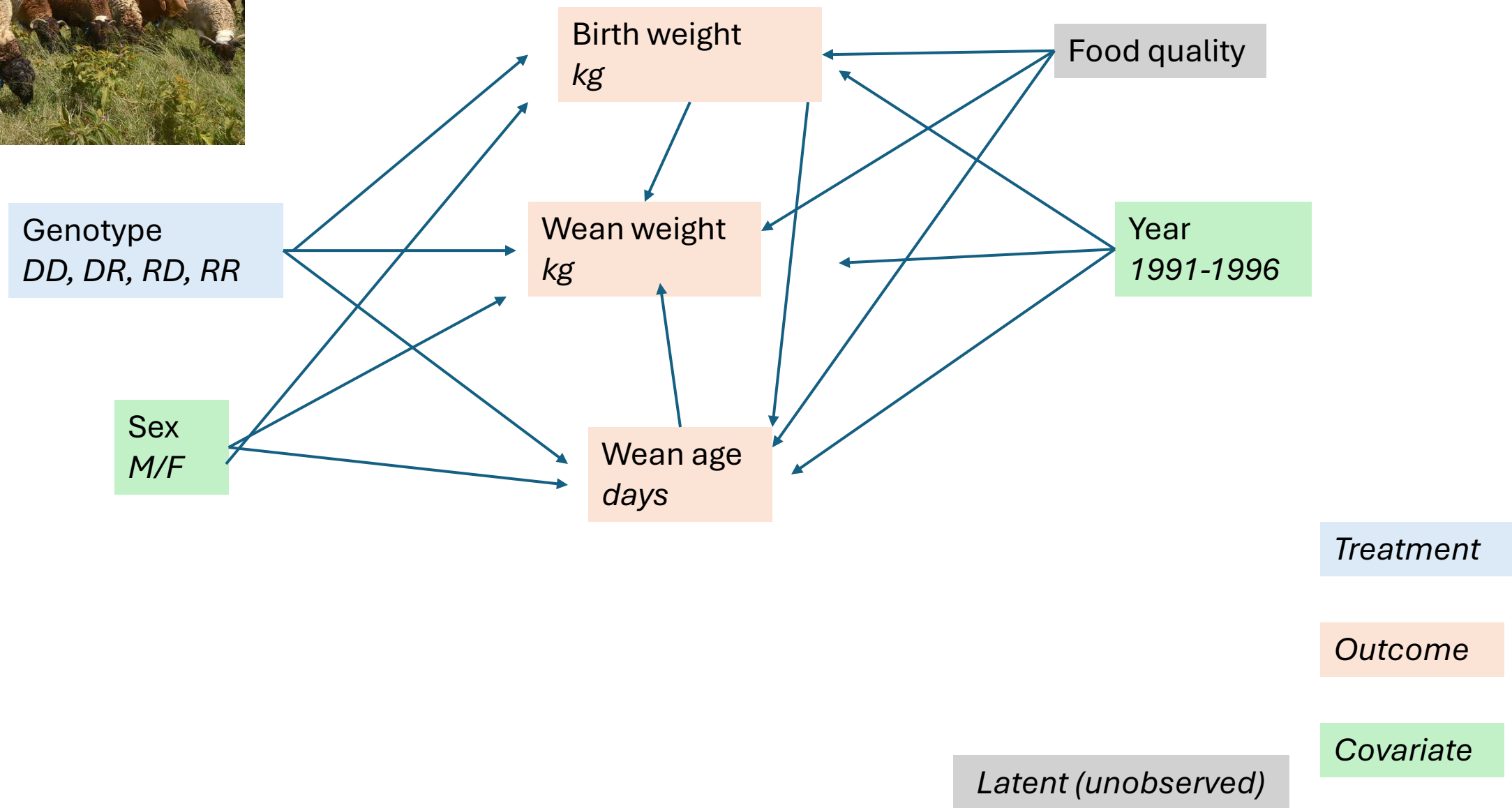


Data type determines how we model a random variable





Some variables of interest may be unobserved



Exercise

1) Choose a different dataset in the agridat package:

<https://kwstat.r-universe.dev/agridat>

2) Read the dataset documentation. Identify one or more research questions and variables of interest.

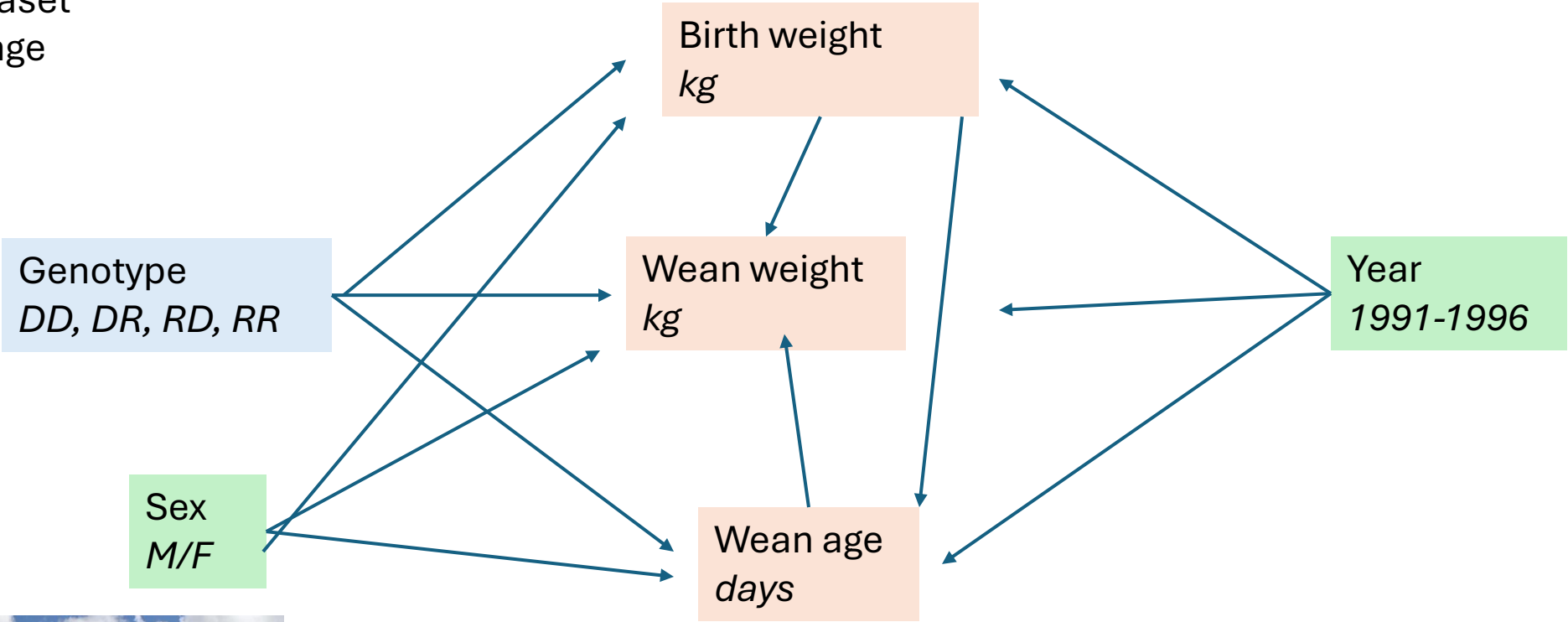
3) Draw a causal diagram (DAG) for the variables of interest in the dataset. Include the units or data type. Distinguish treatment and covariates based on your research question.

4) Write outcomes of interest as general functions of other variables
e.g, Outcome $D = \text{Function}(A, B, C, \dots)$

Birth weight and weaning weight of Dorper x Red Maasi lambs

ilri.sheep dataset
agridat package

How does genotype influence birth weight, weaning weight, and weaning age?



Birth weight } F(Genotype, Sex, Year)
Wean weight } F(Genotype, Sex, Year, Birth weight, Wean age)
Wean age } F(Genotype, Sex, Year, Birth weight)

Treatment

Outcome

Covariate