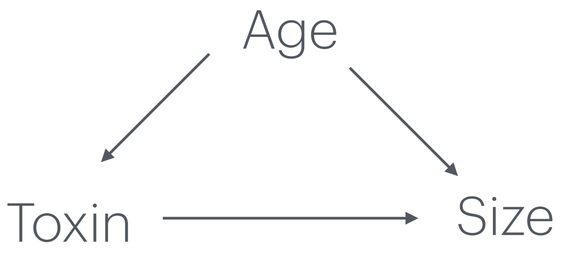
Using simulations to understand multiple regression

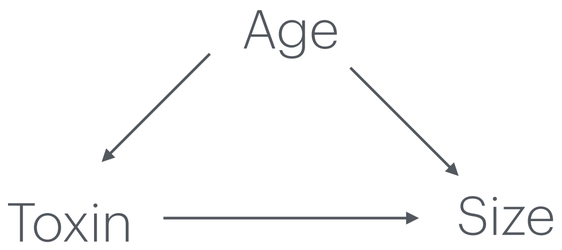
- Simulations are a powerful tool to understand how these models behave
- **Use them liberally! Use them often!**
- Ask:
  - What is the data generating process of system?
  - What model represents this process?
- Simulate data under this model, try to fit the data

- Example:
  - **Question**: What is the effect of toxin exposure on the size of individuals in a population?
  - Age affects both toxin exposure and size



Using simulations to understand multiple regression

- Question: What is the effect of toxin exposure on the size of a population?
- Age affects both toxin exposure and size



Possible model:

$$age_i = Uniform(0, 50)$$
$$toxin_i = Normal(10 + 0.4 \times age_i, 1)$$
$$size_i = Normal(30 - 1 \times toxin_i + 0.5 \times age_i, 5)$$

```
set.seed(1)
age <- runif(200, 0, 50)
toxin <- rnorm(200, 10 + 0.4*age, 1)
size = 30 - 1*toxin + 0.5*age + rnorm(200, 0, 5)
df = data.frame(age = age,
                toxin = toxin,
                size = size)
```