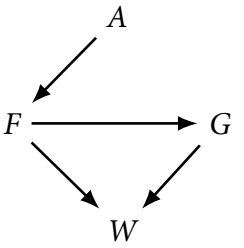
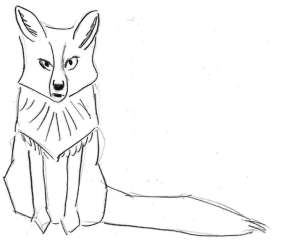
STATISTICAL RETHINKING 2025
WEEK 3

These two problems are based on the same data and (assumed) causal model. The data in `data(foxes)` are 116 foxes from 30 different urban groups in England. These fox groups are like street gangs. We consider four variables. Each group maintains its own (almost exclusive) urban territory. Some territories are larger than others. The `area` variable is the size of each territory. Group size (`groupsize`) varies from 2 to 8 individuals and is the number of foxes in each territory. Some territories also have more `avgfood` than others. And finally food influences the `weight` of each fox. Assume this DAG:





where F is `avgfood`, G is `groupsize`, A is `area`, and W is `weight`.

1. Use the backdoor criterion and estimate the total causal influence of A on F. What effect would increasing the area of a territory have on the amount of food inside it?

2. Suppose your colleague is interested in the causal effect of food F on weight W. They run a linear regression of W on F, G, and A. They report the following coefficients:

```
              mean   sd  5.5% 94.5%
(Intercept)  4.06 0.44  3.37  4.76
F            2.51 1.48  0.15  4.87
G           -0.61 0.16 -0.86 -0.35
A            0.39 0.24  0.00  0.77
```

They interpret the value 2.51 (0.15–4.87) as the causal effect of adding a unit of food to a territory.

What is wrong with your colleague's reasoning? What is a better estimate of the causal effect of food? Why does your estimate and your colleagues' differ so much?