STATISTICAL RETHINKING WINTER 2020/2021
HOMEWORK, WEEK 10

1. Simulate data from this DAG: $X \rightarrow Y \rightarrow Z$. Now fit a model that predicts Y using both X and Z. What kind of confound arises, in terms of inferring the causal influence of X on Y? Can you think of real variables to help explain what is happening here?

2. Revisit the instrumental variable example from Chapter 14, beginning on page 455. You are going to analyze the same problem using missing data imputation. Modify model m14.6 so that it imputes the unobserved values of U instead of using a covariance matrix to average over them. Since none of the U values are observed, you'll have to define the 500 U values as a vector of parameters. Otherwise the model just encodes the DAG, using one regression equation for each variable. Here is the key bit of the formula to help:

```
W ~ normal( muW , sigmaW ),
E ~ normal( muE , sigmaE ),
muW <- aW + bEW*E + bUW*U[i],
muE <- aE + bQE*Q + bUE*U[i],
vector[500]:U ~ normal(0,1)
```

Complete the formula as necessary and compare the inferences to model 14.6 in the text. (You might notice some parameters have low n_eff or high Rhat values. Can you figure out why?)

3. This is a hard problem, so don't feel bad if you struggle with it. Getting a piece of the solution is good. It might seem arbitrary at first. But it has a similar structure to a lot of problems in biology, from genetics to archaeology to the comparative method. It is likely possible for you to quickly intuit a reasonable answer. Justifying that answer with probability theory is the problem.

Some lad named Andrew made an eight-sided spinner. He wanted to know if it is fair. So he spun it a bunch of times, recording the counts of each value. Then he accidentally spilled coffee over the 4s and 5s. The surviving data are:

Value	1	2	3	4	5	6	7	8
Frequency	18	19	22	NA	NA	19	20	22

Your job is to impute the two missing values (the NA values) in the table above. Andrew doesn't remember how many times he spun the spinner. So you will have to assign a prior distribution for the total number of spins and then marginalize over the unknown total. Andrew is not sure the spinner is fair (every value is equally likely), but he's confident that none of the values is twice as likely as any other. Use a Dirichlet distribution to capture this prior belief. Plot the joint posterior distribution of 4s and 5s.