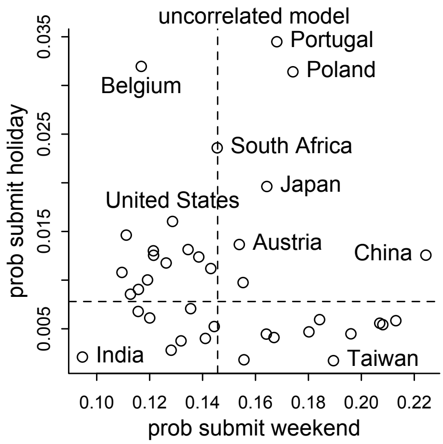
STATISTICAL RETHINKING 2025
WEEK 7 SOLUTIONS

1. The model form is the same as last week, just with a different outcome variable, H instead of W. You could use two separate models and compare their estimates for each country. But I'll instead put both models in the same Markov chain. Like this:

```
library(rethinking)

dat <- read.csv("BMJSubmissions.csv")

str(dat)

m1 <- ulam(
    alist(
        W ~ bernoulli(p),
        logit(p) <- aW + bW[L],
        H ~ bernoulli(q),
        logit(q) <- aH + bH[L],
        aW ~ normal(-2,1),
        aH ~ normal(-2,1),
        bW[L] ~ normal(0,sigmaW),
        bH[L] ~ normal(0,sigmaH),
        c(sigmaW,sigmaH) ~ exponential(1)
    ), data=dat , chains=4 , cores=4 )
```

This is like two shrinkage models in one chain. There are no shared parameters between the W and H models. So they produce independent estimates for each country. Let's plot those estimates against one another:

```
pred1 <- link(m1,data=list(L=1:38))
post1 <- extract.samples(m1)

plot( apply(pred1$p,2,mean) , apply(pred1$q,2,mean) ,
    xlab="prob submit weekend" , ylab="prob submit holiday" )
mtext("uncorrelated model")

abline(v=mean(inv_logit(post1$aW)),lty=2)
abline(h=mean(inv_logit(post1$aH)),lty=2)
```



I've labeled some of the points. And the dashed lines are the means for each outcome. Overall the holiday submission rates are much lower. It looks like there might be a negative correlation overall. But some nations defy that association, like Poland and Portugal.

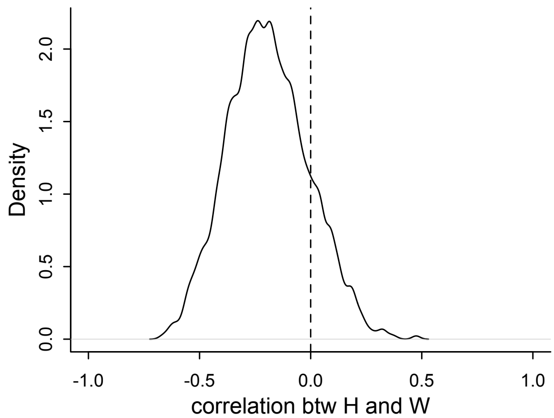
2. To model to correlation between the *W* effects and the *H* effects, we just need to modify the previous model so that the effects come from a common distribution. The code from the lecture and book provides examples. This is how I did it:

```
m2 <- ulam(
    alist(
        W ~ bernoulli(p),
        logit(p) <- aW + b[L,1],
        H ~ bernoulli(q),
        logit(q) <- aH + b[L,2],
        matrix[38,2]:b ~ multi_normal(c(0,0),Rho,sigma),
        aW ~ normal(-2,1),
        aH ~ normal(-2,1),
        corr_matrix[2]:Rho ~ lkj_corr(4),
        vector[2]:sigma ~ exponential(1)
    ) , data=dat , chains=4 , cores=4 )
```

The `b` parameters are now a matrix with 38 rows (one for each country) and 2 columns (W and H). Each row is a draw from a 2D Gaussian prior with correlation matrix Rho and standard deviations sigma (a vector of length two now). The rest of the model remains the same.

Let's look at the posterior for the correlation first:

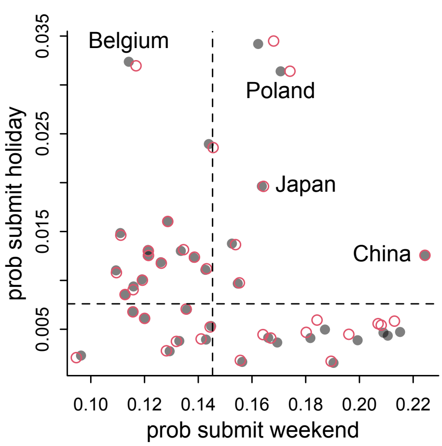
```
post2 <- extract.samples(m2)
dens(post2$Rho[,1,2],xlim=c(-1,1),xlab="correlation btw H and W")
abline(v=0,lty=2)
```



So most of the probability mass is negative, but there isn't much precision here. Both weak and strong negative correlations are plausible, as are weak positive correlations.

What impact does this estimated correlation have on the estimated country effects? Let's plot the new estimates over the previous ones to see.

```
pred2 <- link(m2,data=list(L=1:38))
post2 <- extract.samples(m2)
plot( apply(pred2$p,2,mean) , apply(pred2$q,2,mean) ,
    xlab="prob submit weekend" , ylab="prob submit holiday" ,
    pch=16 , col=grau() )
points( apply(pred1$p,2,mean) , apply(pred1$q,2,mean) , col=2 )
abline(v=mean(inv_logit(post2$aW)),lty=2)
abline(h=mean(inv_logit(post2$aH)),lty=2)
```



The red points are the previous estimates. The filled points are the new estimates. The differences aren't huge, but they are systematic. The nations in the bottom-right have shifted down and to the right, consistent with the negative correlation across nations. These nations have larger-than-average weekend submission probabilities, and often small sample sizes, so the negative correlation has reduced shrinkage of the holiday submission probabilities towards the mean (the dashed line).

The other notable changes are Poland and Portugal in the top-right. These two nations have very high holiday submission probabilities. Given the negative correlation overall between H and W, this suggests that the weekend probabilities are overestimates—nations with very high H rates should have smaller W rates. Notice also Belgium, which is consistent with the negative correlation—it moves even more extreme, because if it has an H probability that high, it should have an even smaller W probability.

The sample sizes here are large in most cases, which means the shrinkage effects are not massive. But the systematic pattern is there, and the point is to understand why estimating the correlation induces this pattern and why it might improve estimates in some cases.