STATISTICAL RETHINKING WINTER 2020/2021
HOMEWORK, WEEK 1 SOLUTIONS

1. Really all you need is to modify the grid approximation code in Chapter 2.

```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prior <- rep( 1 , 1000 )
prob_data <- dbinom( 4 , size=15 , prob=p_grid )
posterior <- prob_data * prior
posterior <- posterior / sum(posterior)
set.seed(100)
samples <- sample( p_grid , prob=posterior , size=1e4 , replace=TRUE )
```

When you plot the result, the posterior will have much more mass over values below 0.5, since that is what the sample indicates. The posterior mean is about 0.30.

2. Modifying only the prior:

```
p_grid <- seq( from=0 , to=1 , length.out=1000 )
prior <- c( rep( 0 , 500 ) , rep( 1 , 500 ) )
prob_data <- dbinom( 4 , size=15 , prob=p_grid )
posterior <- prob_data * prior
posterior <- posterior / sum(posterior)
set.seed(100)
samples2 <- sample( p_grid , prob=posterior , size=1e4 , replace=TRUE )
```

The posterior mean should be about 0.55. This prior yields a posterior with more mass around the true value of 0.7. With the impossible values less than 0.5 ruled out, the second model piles up more plausibility on the higher values near the true value. The data are still misleading it to think that values just above 0.5 are the most plausible. But the posterior mean of 0.55 is much better than 0.30 from the previous problem.

Informative priors, when based on real scientific information, help. Here, the informative prior helps because there isn't much data. That is common in a lot of fields, ranging from astronomy to paleontology.

3. The code is very simple, but the interpretation maybe less so:

```
set.seed(100)
PI( samples2 )
HPDI( samples2 )
```

```
       5%       94%
0.5035035 0.6326326
> HPDI( samples2 )
    |0.89     0.89|
0.5005005 0.6046046
```

The percentile interval (the top one) is wider. The lower bound is above 0.5 (a little bit) and the upper bound is 0.63. The HPDI (bottom interval) is narrower. The lower bound is right at 0.50, but the upper bound is 0.60 instead of 0.63. You should expect the HPDI to be narrower and the include the point with highest posterior probability. This is just like the example on page 57 in the book. But since the boundaries of these intervals aren't really informative—nothing special happens at the boundary—when these intervals are very different, the best thing is not to report intervals at all. Just draw the posterior distribution, so your colleagues can see what is going on.