Problem Set 3.5

Exercise 3.18

a)

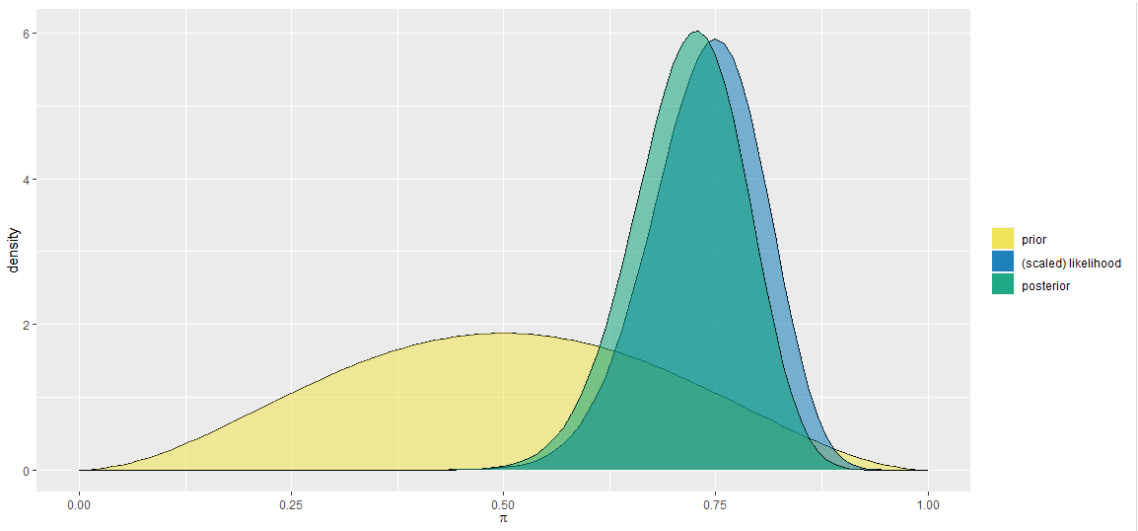
The prior model is:

$$\pi \sim Beta(3,3)$$

The data from the survey is: n=40 and y=30

That makes the posterior model be:

$$Beta(3+30, 3+40-30) = Beta(33, 13)$$




```
> plot_beta_binomial(alpha = 3, beta = 3, y = 30, n = 40)
> summarize_beta_binomial(alpha = 3, beta = 3, y = 30, n = 40)
      model alpha beta      mean      mode         var        sd
1     prior     3    3 0.5000000 0.5000000 0.035714286 0.1889822
2 posterior    33   13 0.7173913 0.7272727 0.004313639 0.0656783
>
```

We can observe a huge difference between the prior model and the posterior model this is because the prior model doesn't have much information, and the survey data represents a really big number compared to the prior one, that is why the posterior model is more closely related to the data than to the prior. But the posterior is not exactly equal to the survey data, since the prior still have some importance, even if its small.

b)

The prior is the same:

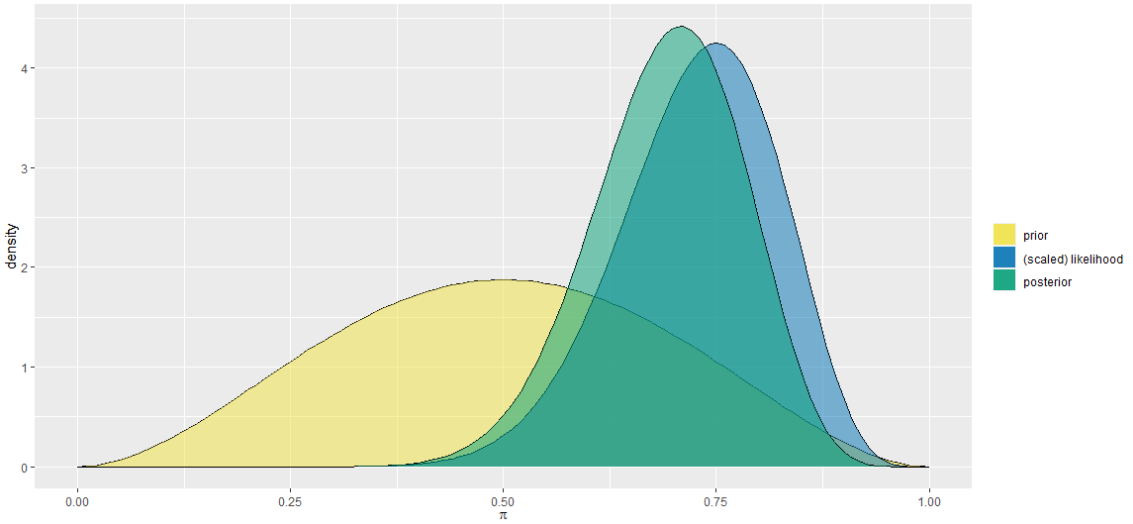
Alejandro C Parra Garcia

$$\pi \sim Beta(3,3)$$

The data from the survey is: n=20 and y=15

That makes the posterior model be:

$$Beta(3+15, 3+20-15) = Beta(18,8)$$




```
> summarize_beta_binomial(alpha = 3, beta = 3, y = 15, n = 20)
      model alpha beta      mean      mode         var         sd
1     prior     3    3 0.5000000 0.5000000 0.035714286 0.18898224
2 posterior    18    8 0.6923077 0.7083333 0.007889546 0.08882312
```

This is a similar case as before, where the survey data dominates the posterior because it has a larger sample size than the prior.

c)

The difference in the two models can be easily seen. In Harold model we can see that the posterior is closer to the prior than in Patrick model. This is easily explained as in the case of Harold the number of people survey is 20 while Patrick's is 40. This difference makes a difference, because since surveying 40 people gives more information that surveying 20 the posterior of Patrick's case is going to be more closely related to the survey data than in the case of Harold. That's the main difference because the percentage of people answering the surveying saying that they are going to the protest is the same in both cases 30/40 = 15/20 = 0.75. The only difference is in the number of samples. As one survey have more data, we can we more certain that is going to be more correct.