

Using Bayesian Methods to Compare Weather in April & October

Samuel Greeman

1 Introduction

Our goal here is as simple as can be, but using Bayesian methods does make the process a bit more difficult. We simply want to find out if weather in April is similar to weather in October. We will use three different measurements to test this theory, namely maximum temperature, minimum temperature, and precipitation. This theory of mine comes from the bigger theory that weather in months is symmetrical. More specifically: the months of December and January are similarly cold and snowy. The months of February and November are relatively cold. The months of March, April, and October are relatively fair with high amounts of precipitation. May and September are warm and a bit dry. June, July, and August are all hot and humid. For this test we will use the data from Boston in April and October 2011 and 2012.

2 Methodology

In order to do a posterior prediction about these parameters, we must establish prior distributions for them.

-For High Temps in April and October, we will say they follow a $N(64, 15)$ distribution, meaning average 64 degrees.

-For Low Temps in April and October, we will say they follow a $N(41, 11)$ distribution.

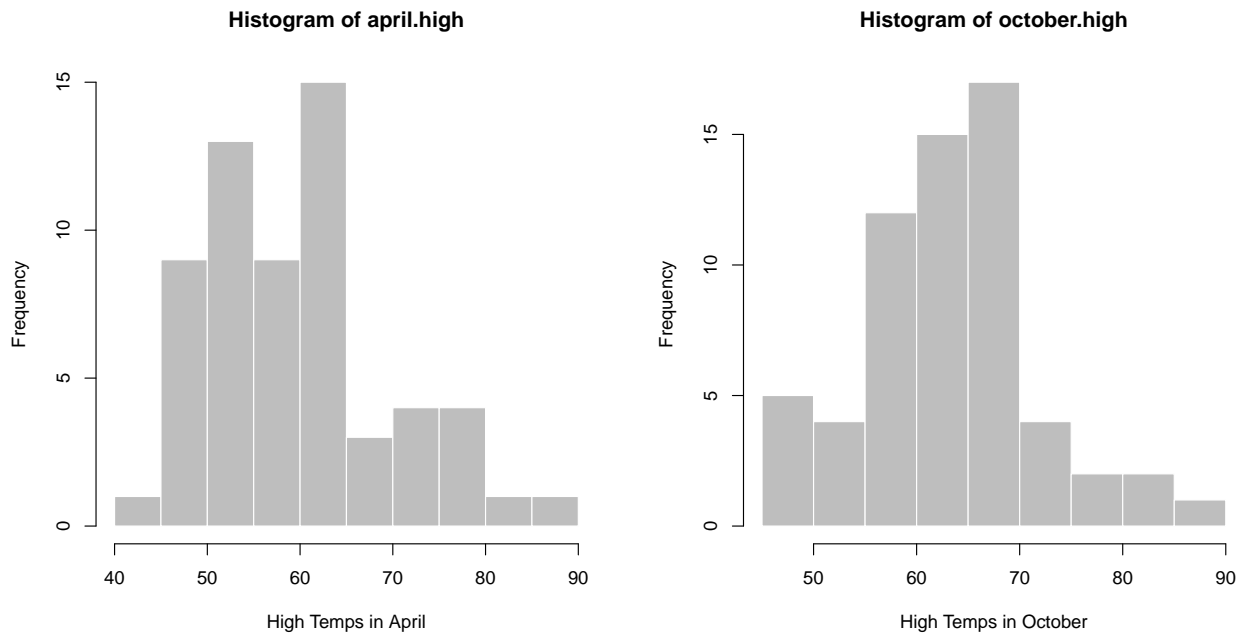
-For Precipitation in April and October, we will say they follow a $N(0.1, 0.03)$ distribution.

For all of our posterior predictions, these are the prior distributions we will use.

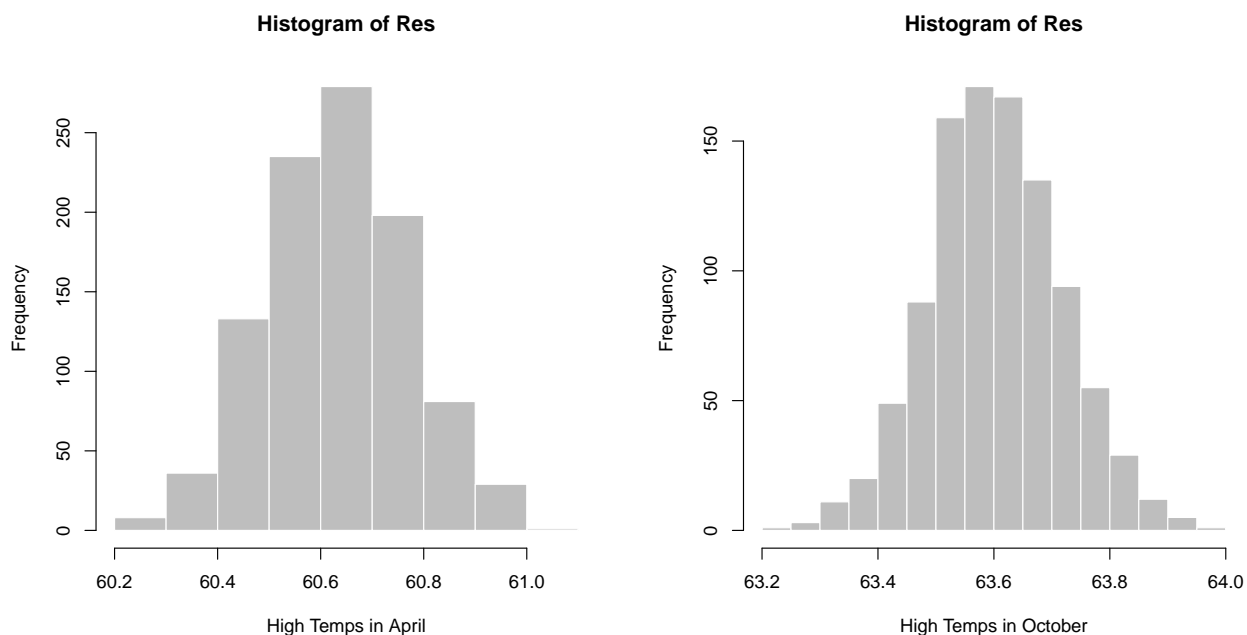
3 Results

For our actual high temps in October and April, we obtain the following histograms:

High Temps



We can see that while it may be a bit hotter in October, the histograms seem pretty normal, which is good. The mean backs this fact up: the mean high temp in April was 60.07 while it was 63.55 for October. Now let's look at the plots generated by our posterior:

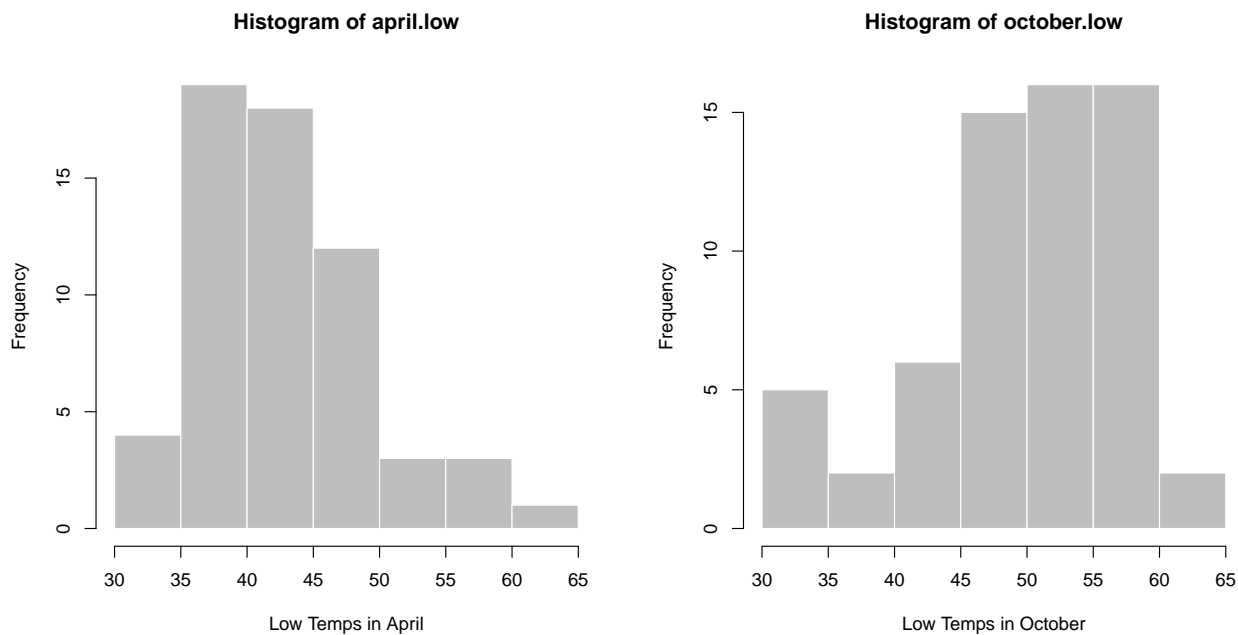


These plots look fantastic. Our posterior mean for April is 60.63, and for October, it is 63.60. This means that even with the prior assumption that April and October have the same high

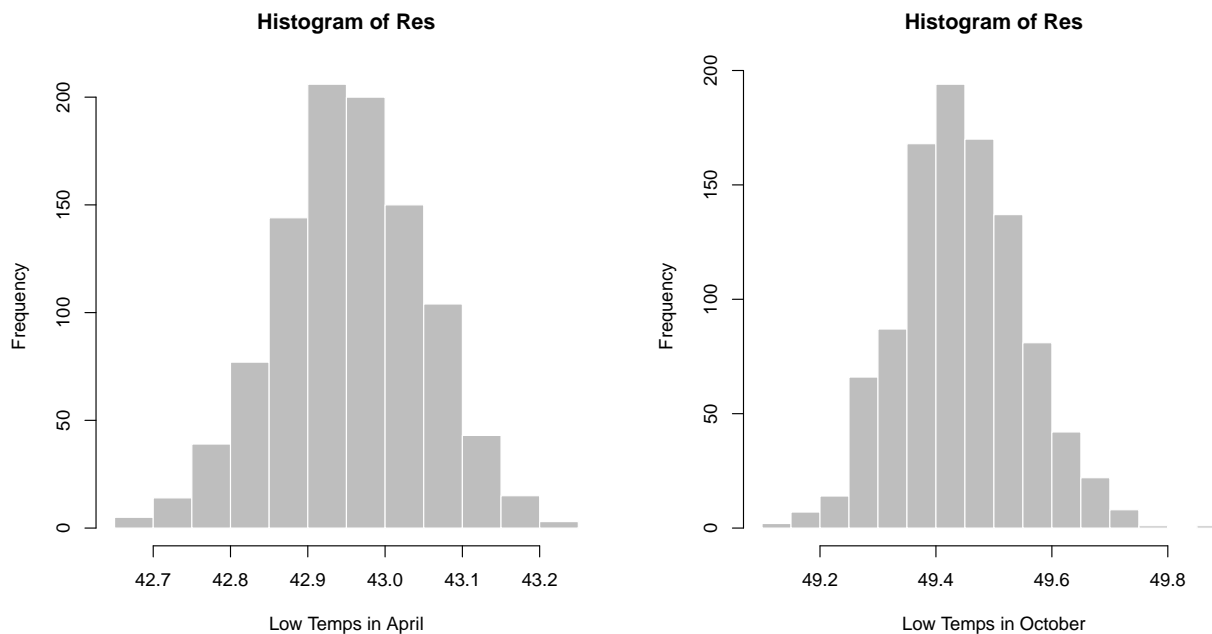
temperatures, October is still predicted to be 3 degrees hotter on average than April.

Low Temps

Moving on to low temps, we first look at the histograms for our actual data:



Mean low temp for April is 43.17, while it is significantly higher for October, at 50.45. Using the prior distribution for low temps, we get the following histograms for our posterior:

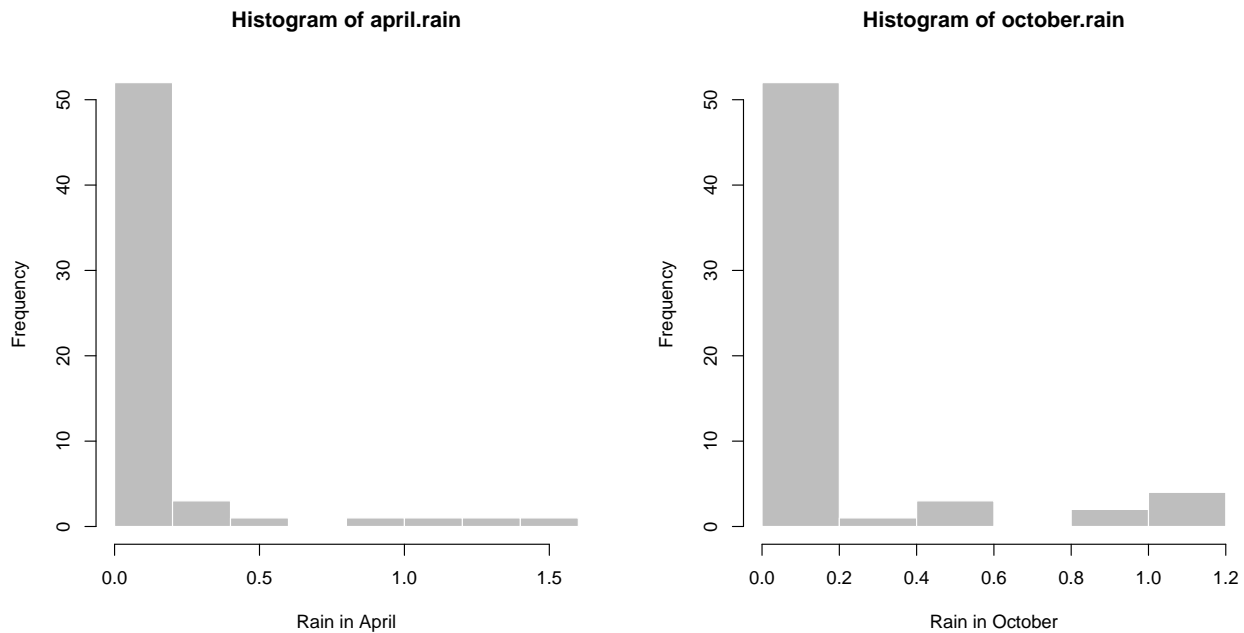


Our histograms once again indicate to us that it is warmer on average in October than in April. Our posterior mean for low temp in April is 42.95, and October posterior mean is 49.45. Both the posterior and actual means indicate that the low temp is on average around 7 degrees hotter

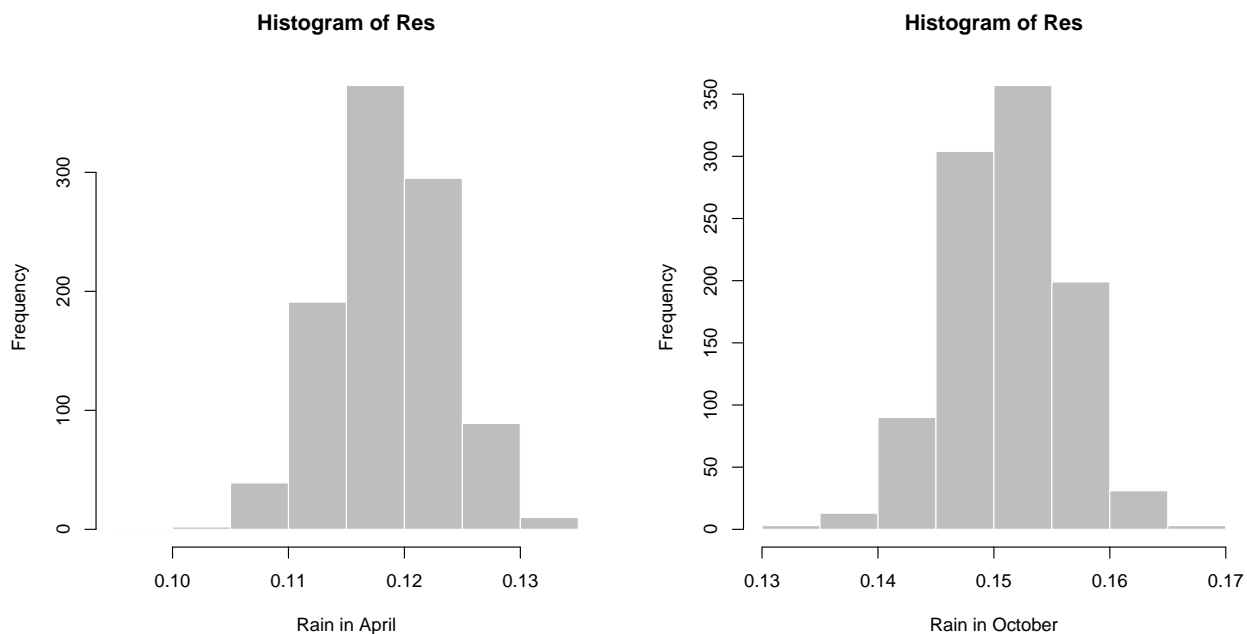
in October than in April.

Precipitation

Now, we investigate precipitation amounts, starting with the histograms of the data observed:



This one will be a bit tricky, and perhaps inaccurate as we see that our normality assumption may not be true. However, we can still think of this as a normal distribution, but we can't see the other (negative) half of the bell curve because, naturally, you can't have negative precipitation. For April, our average precipitation was 0.119 inches, and for October, 0.151 inches. Having a look at our posterior histograms, we see the following:



These, along with our posterior means, 0.119 for April and 0.151 for October, which, amazingly (perhaps due to our small standard deviation) are the same as our observed means, tell us that it definitely rains more in October than in April.

4 Discussion and Final Thoughts

We set out to investigate how similar the weather in April and October in Boston is, with identical prior distributions. We found, using histograms and our posterior means, that, while they are similar, they are certainly not identical. On average, high temps are 3 degrees higher in October than April, low temps are 7 degrees higher, and it rains almost 50 percent more in October. To me, there is an explanation for why the high temps are closer than the low temps. Naturally, more rain means more humidity, and generally, more humidity means less deviation in temperature. Since October has more rain and therefore more humidity, that explains why the difference between low and high temps in October (14 degrees) is smaller than the difference between high and low temps in April (17 degrees). Circling back to the main point, there are definitely a few things that could have been done better. Firstly, our prior distributions were basically educated guesses, as there is no rhyme or reason as to why we chose them, other than the thought that in April and October, it's usually about 64 degree, not much lower than a high of 35 degrees, not much higher than 85 degrees. The precipitation is tricky to estimate because, while on average, the results are pretty accurate, this doesn't mean that every day will yield 0.1 inches of rain because it obviously doesn't rain every day. Other than that, only the normality assumption is questionable, but as we saw with the temperature histograms, they seem roughly normal. All in all, I believe our investigation gave us results that are very plausible.

References

Data:

-National Weather Service Corporate Image Web Team. "National Weather Service Climate."

National Weather Service, 24 Oct. 2005, w2.weather.gov/climate/xmacis.php?wfo=box.