Part A

```
set.seed(1)
data=rbinom(1000000,1,0.9)
dim(data)=c(100000,10)
sumofrows=apply(data,1,sum)
data2=rbinom(100000,10,0.9)
par=(mfrow=c(2,1))
vectorofbreaks=seq(from=-0.5,to=10.5,by=1)
hist(sumofrows,breaks=vectorofbreaks)
hist(data2,breaks=vectorofbreaks)
```

First 5 rows of `data`:

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	1	1	1	1	1	1	1	1	1	1
[2,]	1	1	1	1	1	1	1	0	1	1
[3,]	1	1	1	1	1	1	1	1	1	0
[4,]	0	1	1	1	1	1	1	0	1	1
[5,]	1	1	1	1	1	1	1	1	1	1

First 5 entries of `sumofrows`

[1] 10 9 9 8 10

First 5 entries in `data2`

[1] 10 9 9 10 10

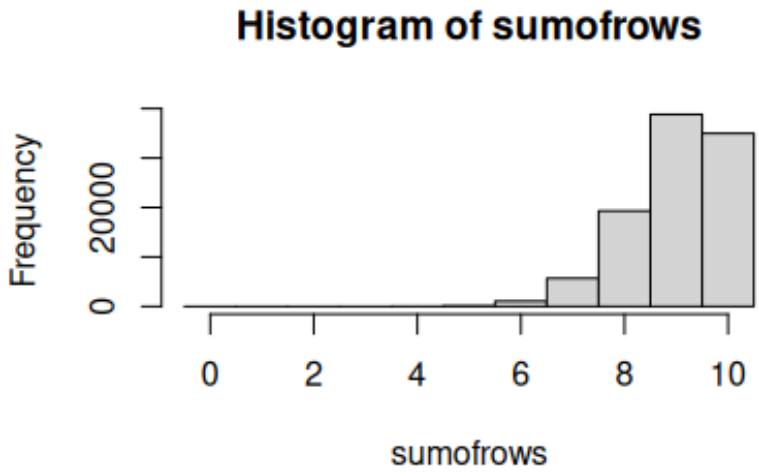
Do the Histograms look the same?

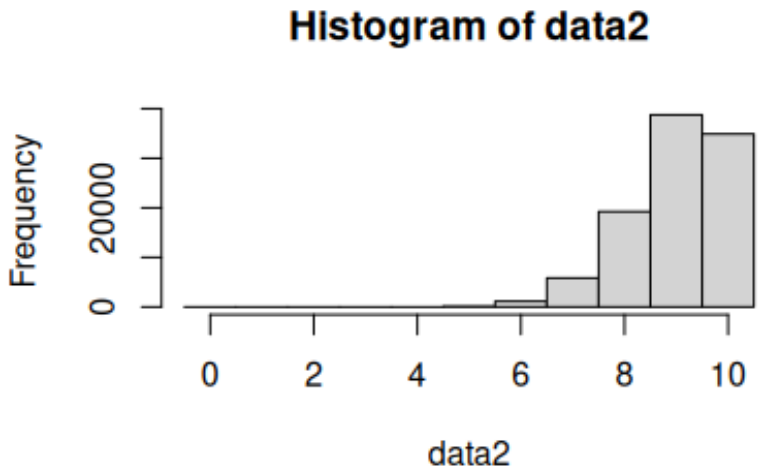
- Yes.

Why, and should they?

- The two histograms look almost identical because of the way we average data to create `sumofrows`. `data` begins as a matrix with 100,000 rows and 10 columns. `sumofrows` averages each row to find the marginal distribution, resulting in a 100,000 element vector. When we create `data2`, we are making a vector of 100,000 elements that follows the same probability of success as `data`, and by finding the marginal distribution of `data`, we have ensured that it maintains the same general probability. This means that `sumofrows` and `data2` should be nearly identical from a general perspective.

  –> We're essentially "averaging" `data` so that it will maintain the same slope/shape, but have fewer elements.


Histogram of sumofrows
Frequency
0
20000
0
2
4
6
8
10
sumofrows


Histogram of data2
Frequency
0
20000
0
2
4
6
8
10
data2

Part B

Summary(tr)

mode	FALSE	TRUE
logical	10027	89973

Calcualated probability of first success on first trial: $0.9^1 = 0.9$

Summary(tr2)

mode	FALSE	TRUE
logical	90949	9051

Calculated probability of first success on second trial: $\binom{1}{0} * 0.9^1 * 0.1^1 = 0.09$

Explanation of data summaries:

As explained in the lab document, `tr` finds every instance of 1 in the first column, registering `1` as a logical `true` and `0` as `false`. `tr2` finds every instance of 0 in the first column, then given that, every instance of `1` in the second column where the first column is 0.

This gives us likelihood of the first success on first trial (`tr`) and likelihood of first success on second trial (`tr2`). When we calculate the mathematical probability of sucess on first trial, we get 0.9, which is similar to what we got from our R code (Since $\frac{89973}{89973+10027} = 0.8997$).

Similarly, when we calculate the mathematical likelihood of first success on second trial, we get 0.09, which is also similar to what we got from our R code (since $\frac{9051}{9051+90949} = 0.09053$)