Yuchen Shi

6962247146

Computer Project 1 (Using R)

#1

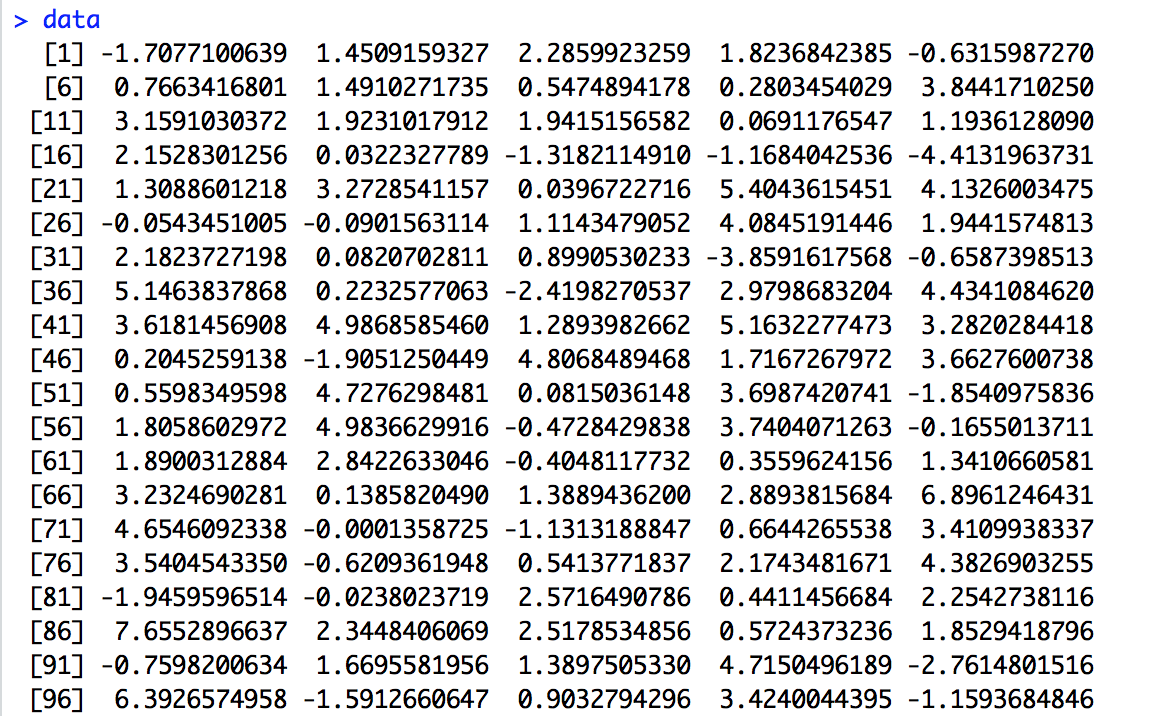
data<-rnorm(100, 1.5, 2)

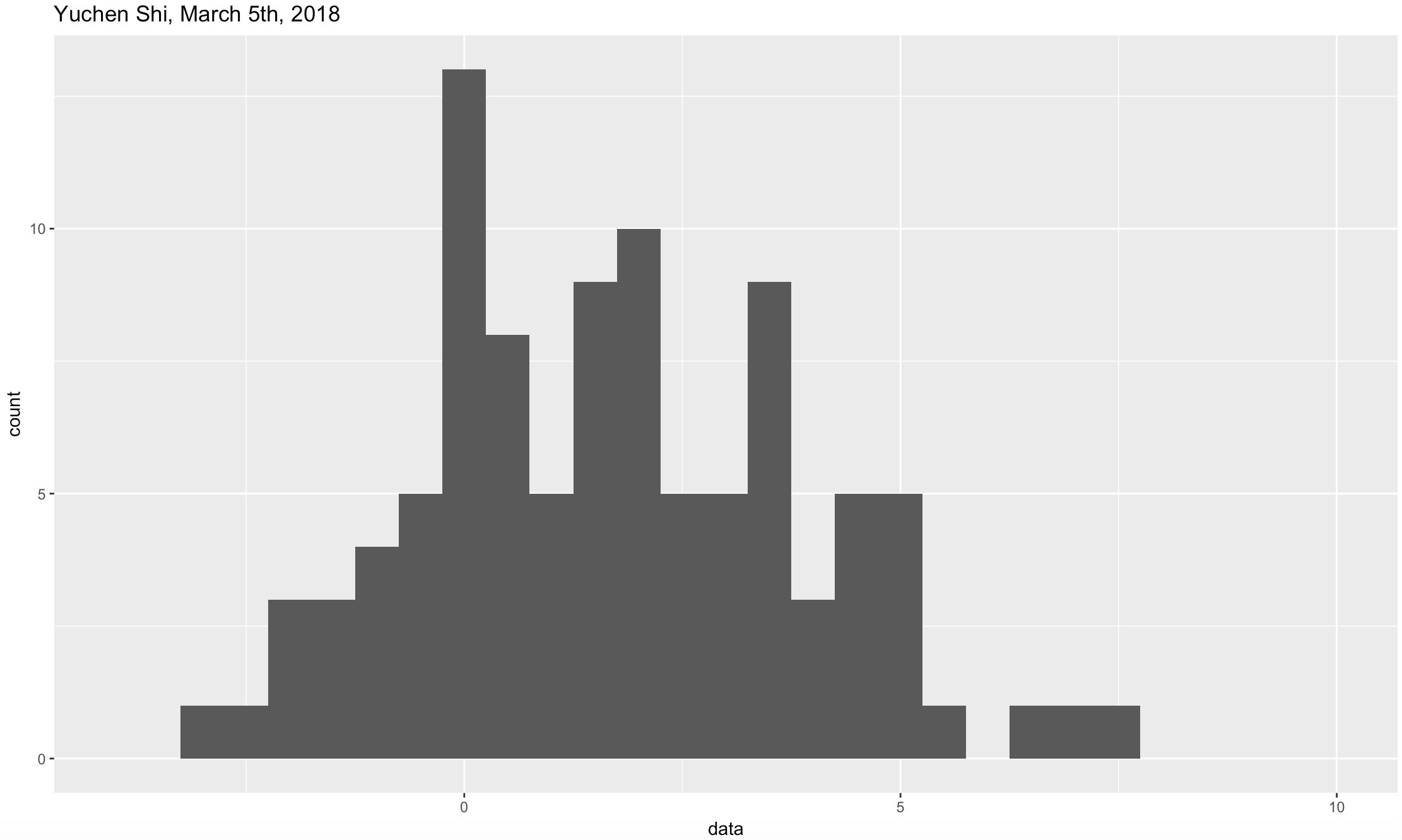
dataframe<-data.frame(data)

install.packages("ggplot2")

library(ggplot2)

ggplot(dataframe, aes(data))+geom\_histogram(binwidth = 0.5)+xlim(-4, 10)+ggtitle("Yuchen Shi, March 5th, 2018")





#2

install.packages("dplyr")

library(dplyr)

dataframe%>%filter(data>0)%>%count()

# A tibble: 1 x 1

n

<int>

1 76

76 variables are bigger than 0.

Estimated probability that a single normal variable with mean 1.5 and standard deviation 2 is above 0: 76/100=0.76

Theoretical value:

1-pnorm(0, 1.5, 2)

[1] 0.7733726

#3

sampleMeans=c()

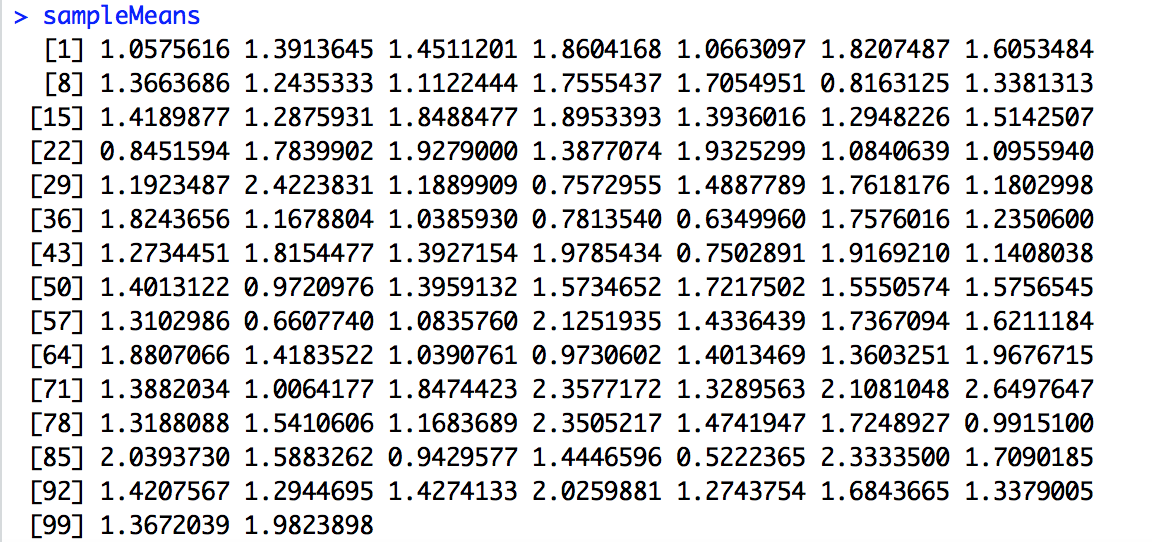
for (i in 1:100){

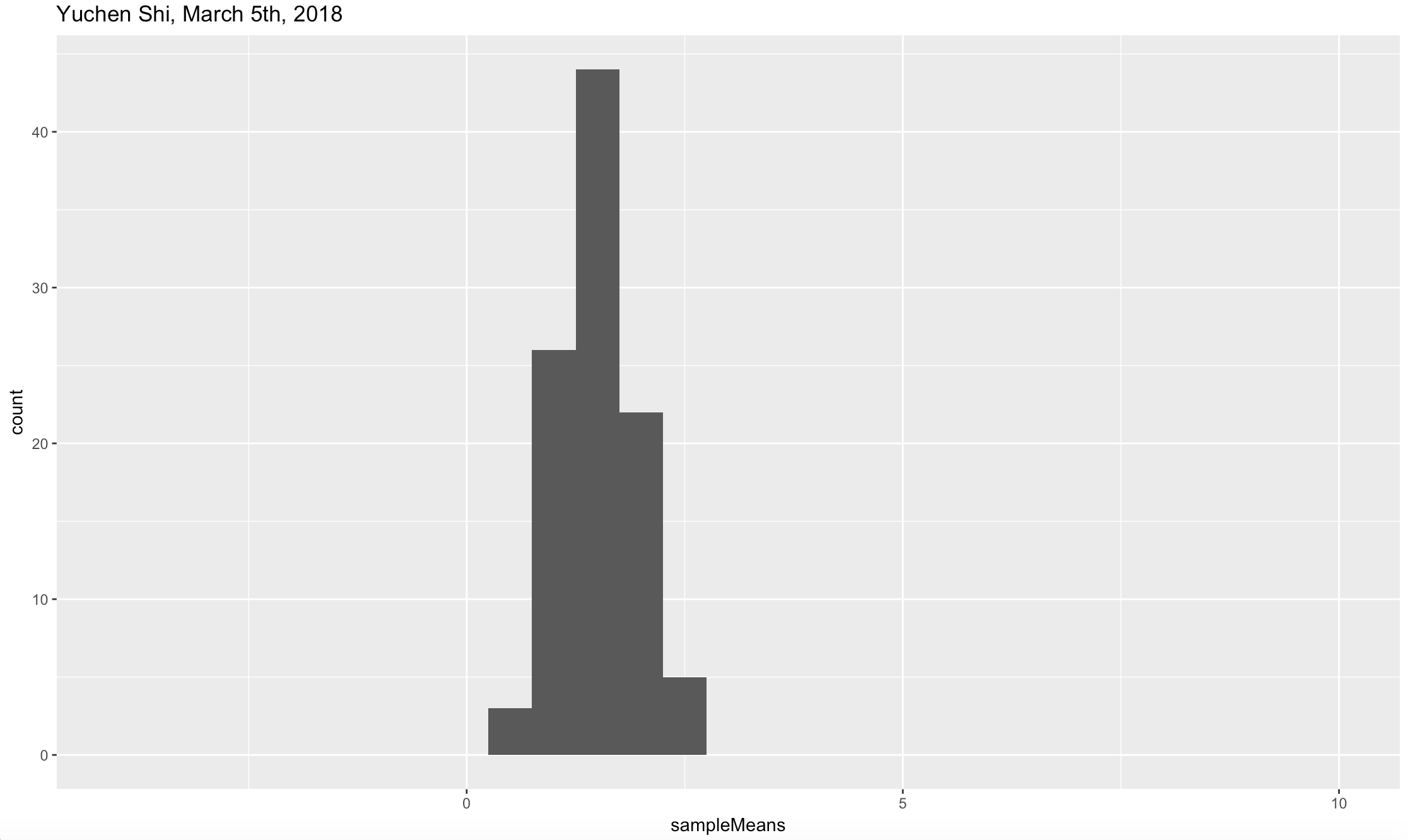
sampleMeans[i]=mean(rnorm(20, 1.5, 2))

}

dataframe2<-data.frame(sampleMeans)

ggplot(dataframe2, aes(sampleMeans))+geom\_histogram(binwidth=0.5)+xlim(-4, 10)+ggtitle("Yuchen Shi, March 5th, 2018")





dataframe2%>%filter(sampleMeans>0)%>%count()

# A tibble: 1 x 1

n

<int>

1 100

100 variables are bigger than 0.

Estimated probability that a sample mean is bigger than 0: 100/100=1

Theoretical value:

1-pnorm(0, 1.5, 2/sqrt(20))

[1] 0.9996019

#4

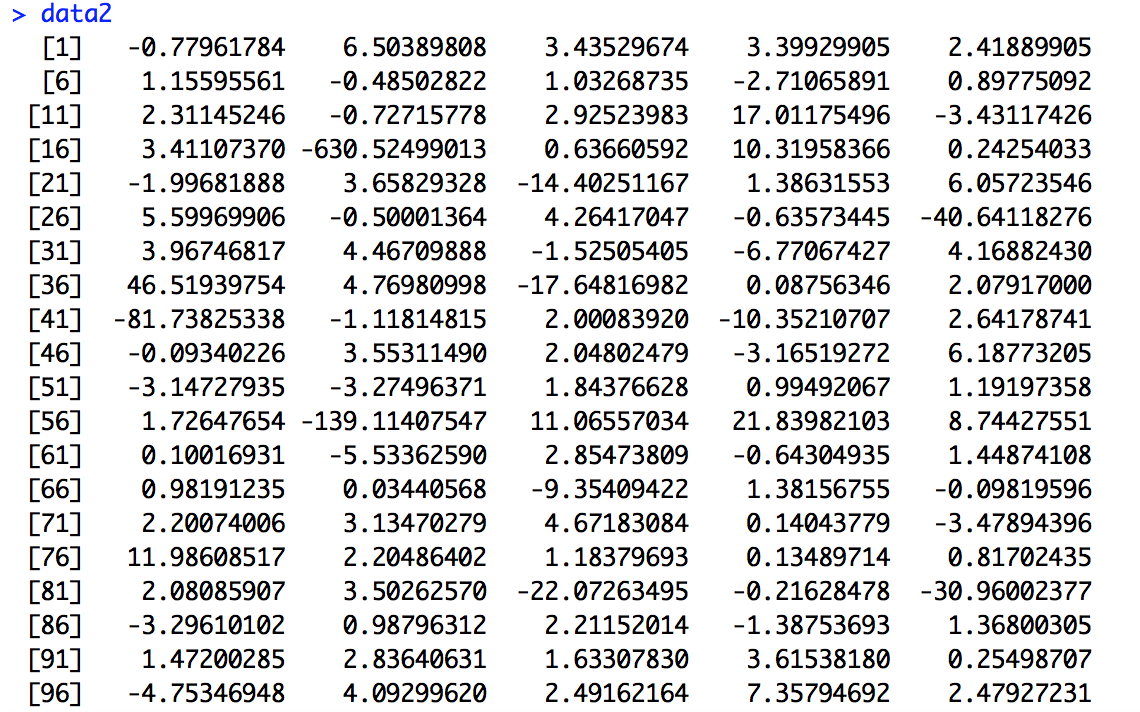
The histogram of the sample means is very narrow, while the histogram of 100 random normal variables is very wide, since the standard deviation of the data is smaller for the sample means (sd/sqrt(n)). As a result, the probability of getting above 0 is very high for the 100 sample means, compared to that of the 100 random normal variables.

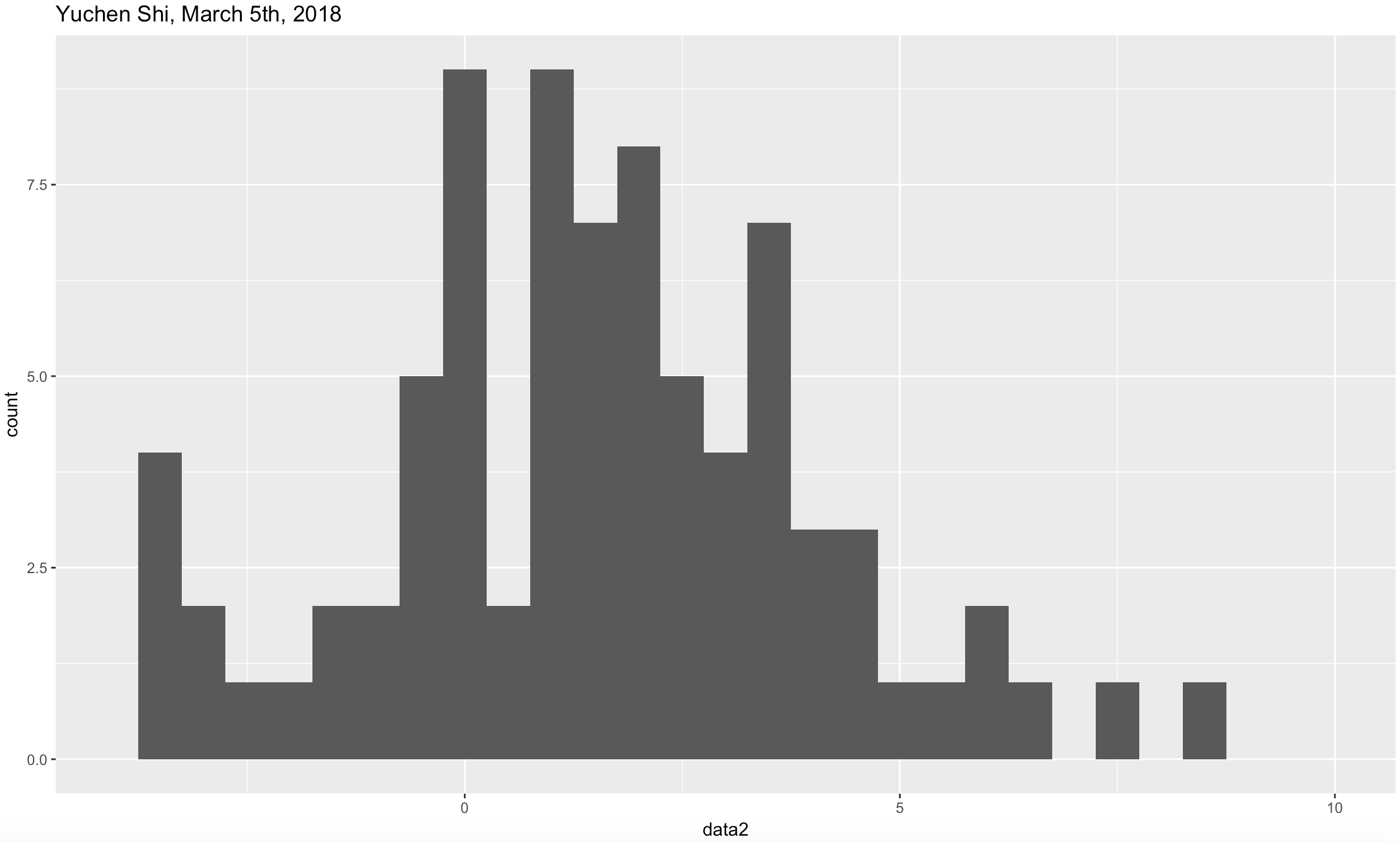
#5

data2<-rcauchy(100, location=1.5, scale=2)

dataframe4<-data.frame(data2)

ggplot(dataframe4, aes(data2))+geom\_histogram(binwidth = 0.5)+xlim(-4, 10)+ggtitle("Yuchen Shi, March 5th, 2018")





dataframe4%>%filter(data2>0)%>%count()

# A tibble: 1 x 1

n

<int>

1 67

67 variables are bigger than 0.

Estimated probability that a single Cauchy variable is greater than 0 is 67/100=0.67.

Theoretical value:

1-pcauchy(0, location=1.5, scale=2)

[1] 0.7048328

sampleMeans2=c()

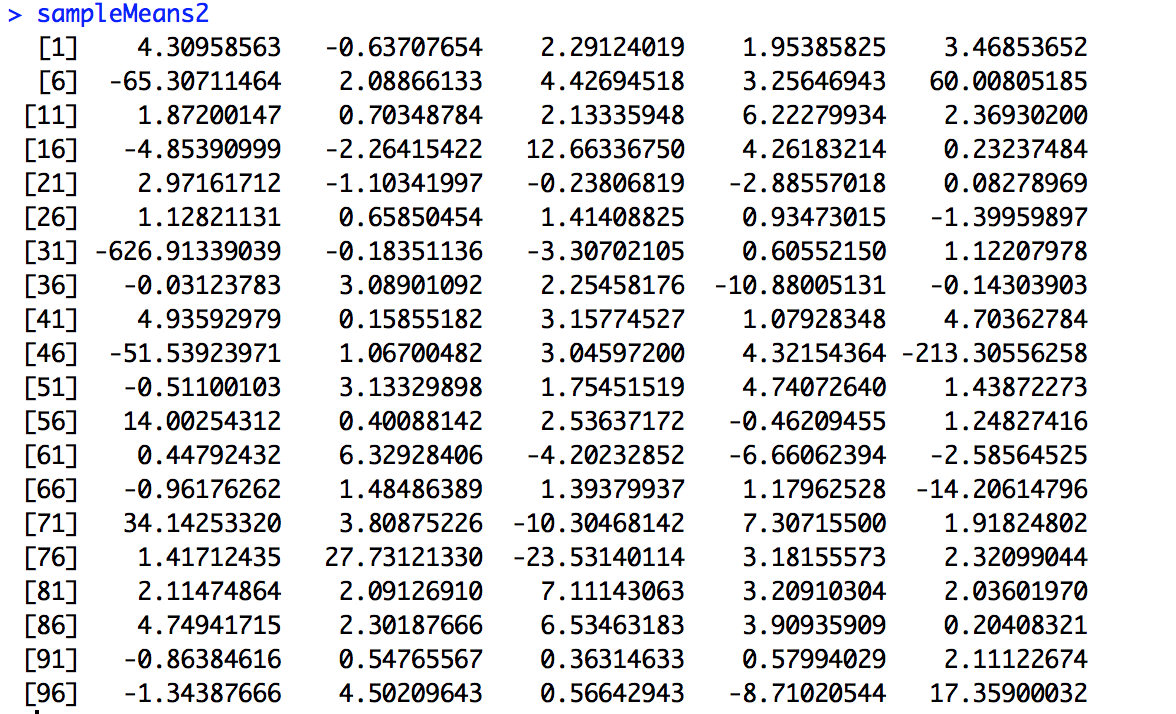
for (i in 1:100){

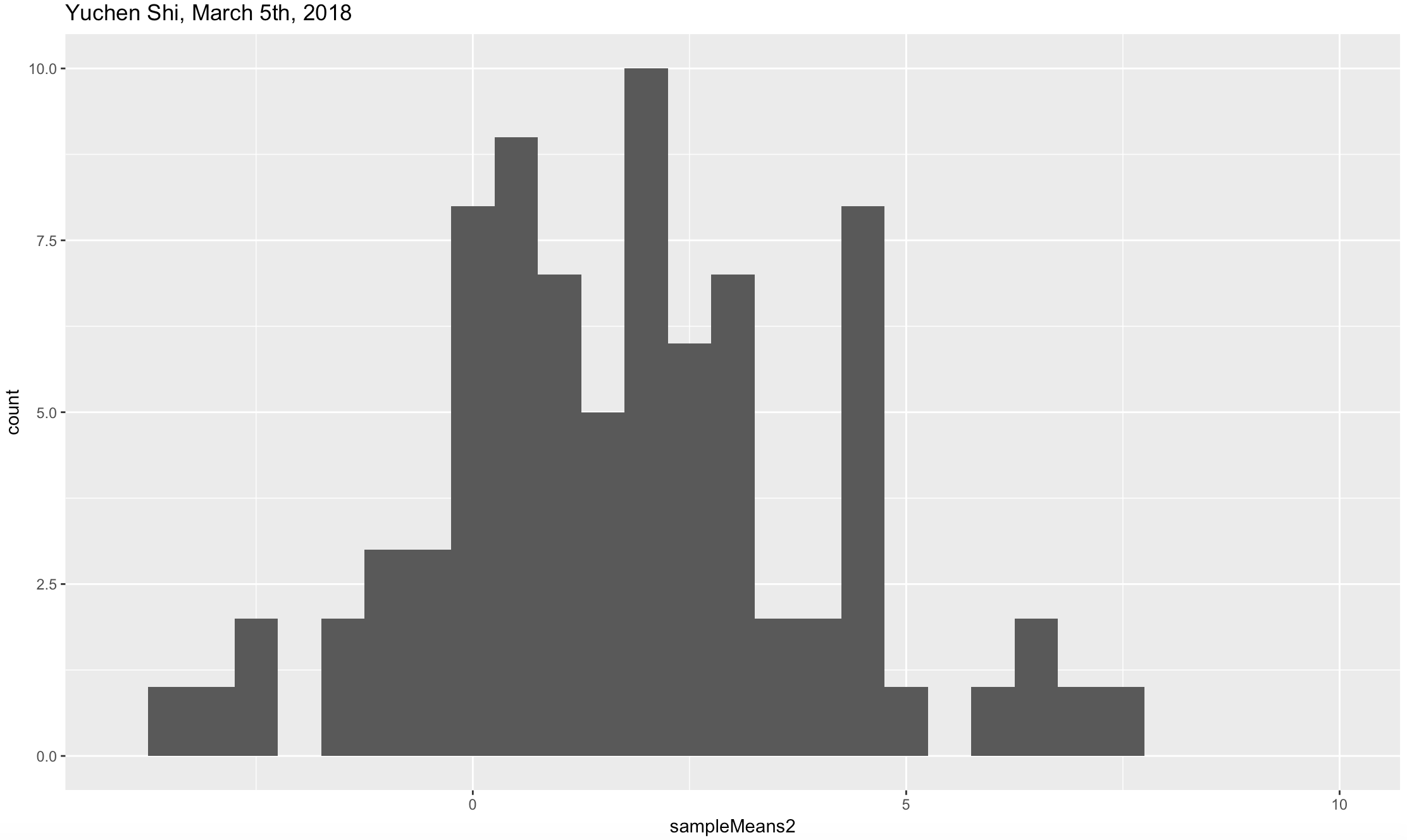
sampleMeans2[i]=mean(rcauchy(20, location=1.5, scale=2))

}

dataframe5<-data.frame(sampleMeans2)

ggplot(dataframe5, aes(sampleMeans2))+geom\_histogram(binwidth = 0.5)+xlim(-4, 10)+ggtitle("Yuchen Shi, March 5th, 2018")





dataframe5%>%filter(sampleMeans2>0)%>%count()

# A tibble: 1 x 1

n

<int>

1 72

72 variables are greater than 0.

Estimated probability that a sample mean is greater than 0: 72/100=0.72.

Theoretical value:

1-pcauchy(0, location=1.5, scale=2)

[1] 0.7048328

Because the distribution of the sample mean of Cauchy variables will be equal to the distribution of the Cauchy variables themselves.

The shape of the histogram of the Cauchy variable is similar to that of the normal variable. However, in the normal case, the histogram of the sample mean is narrower than that of the random normal variables. But in the Cauchy case, they don’t change much, because the two parameters don’t change, so the distribution stays the same. Our theoretical value of probability that is greater than 0 stays the same:0.7048328.