Homework 5

A bulb company company produces high efficient bulbs. The probability a bulb to be defective is 0.01.

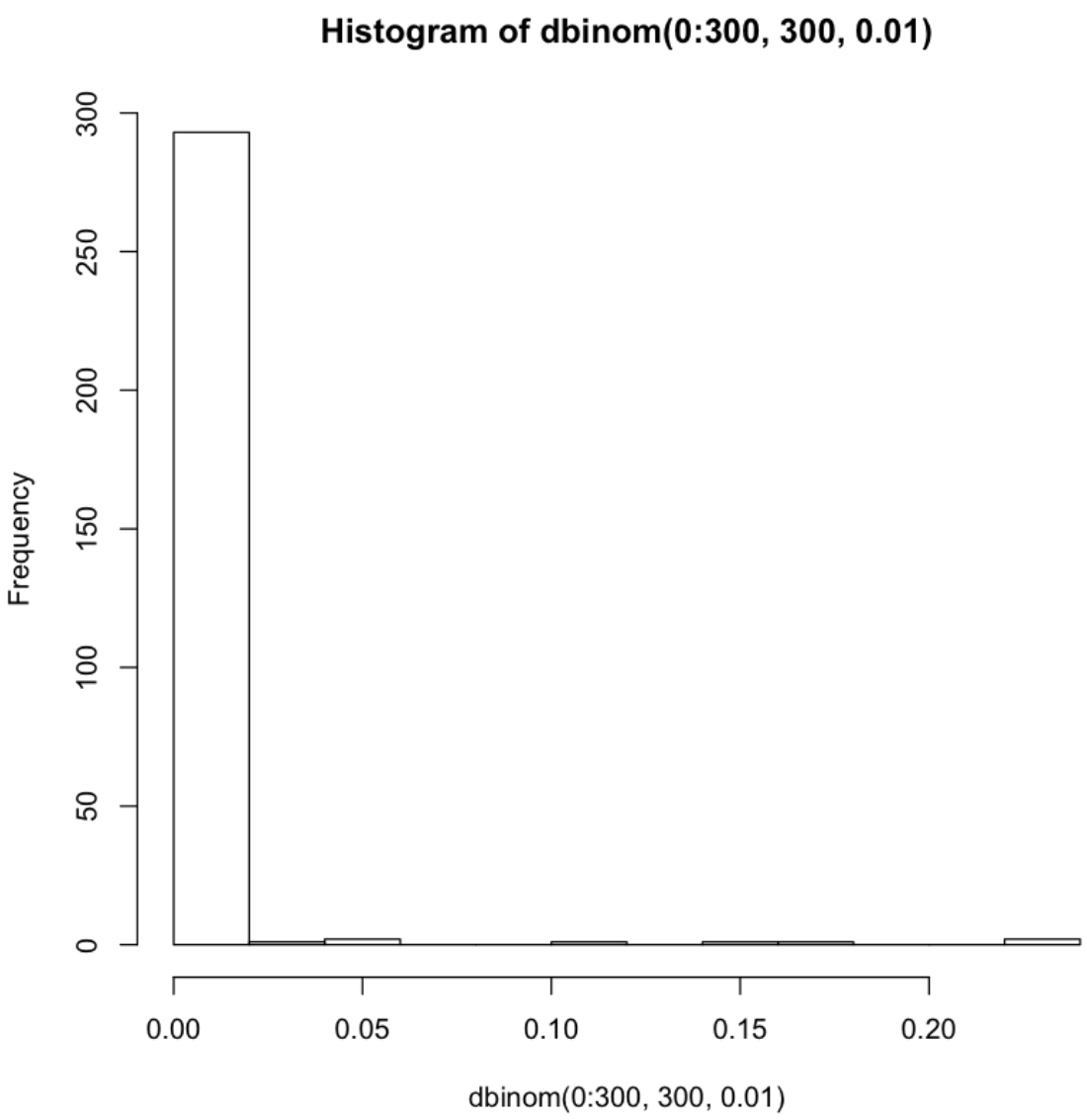
a. What is the probability that a sample of 300 bulbs will contain exactly 5 defective?

b. Draw the histogram of probabilities

In [81]:
```
dbinom(5,300,0.01)
```

0.100985269246102

In [79]:
```
hist(dbinom(0:300,300,0.01))
```

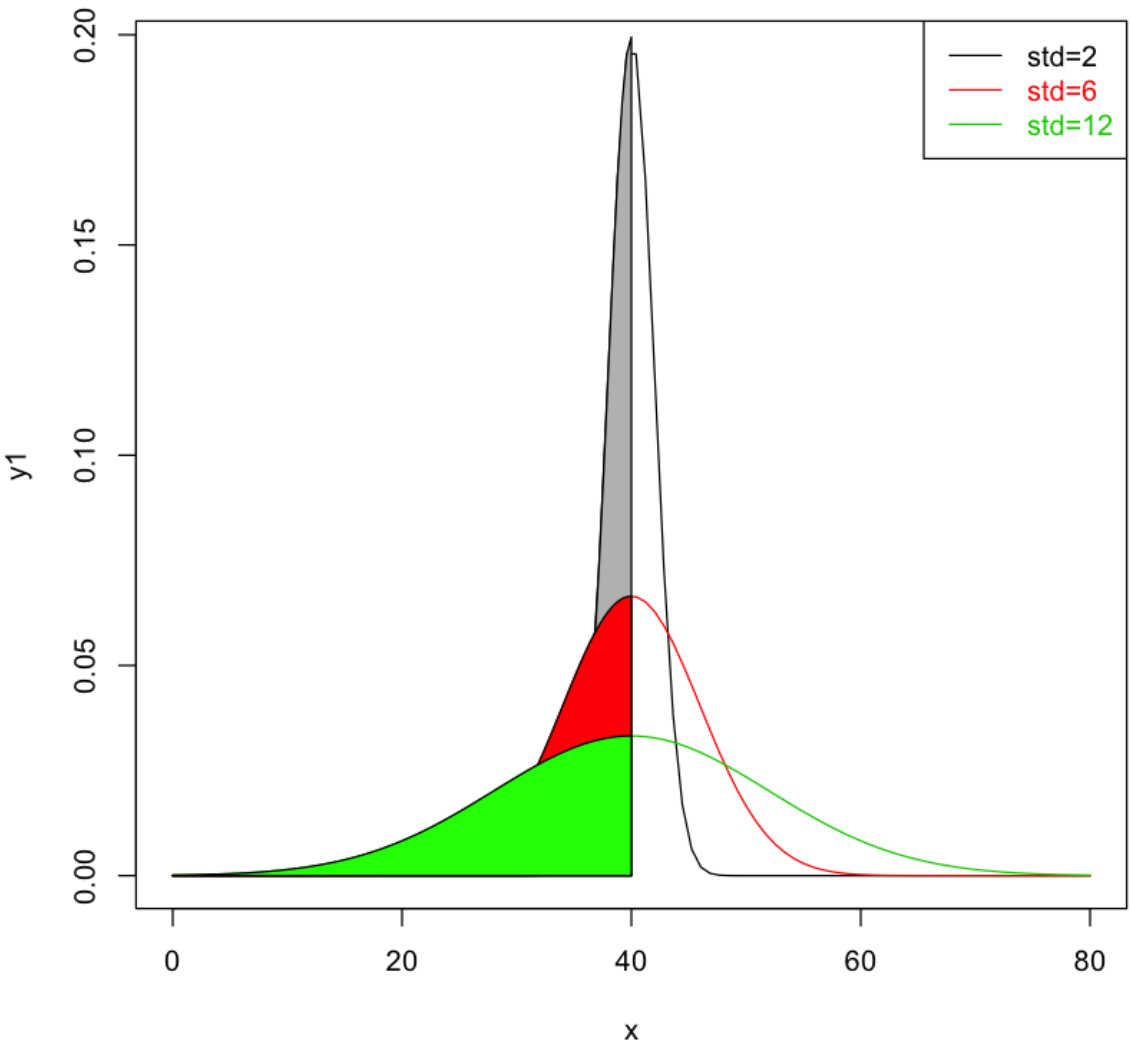


B. X is with mean μ = 40, Find and plot the density distribution with three different standard deviations (2,6,12) (Please draw the curves on the same figure with different colors and a legend) for:

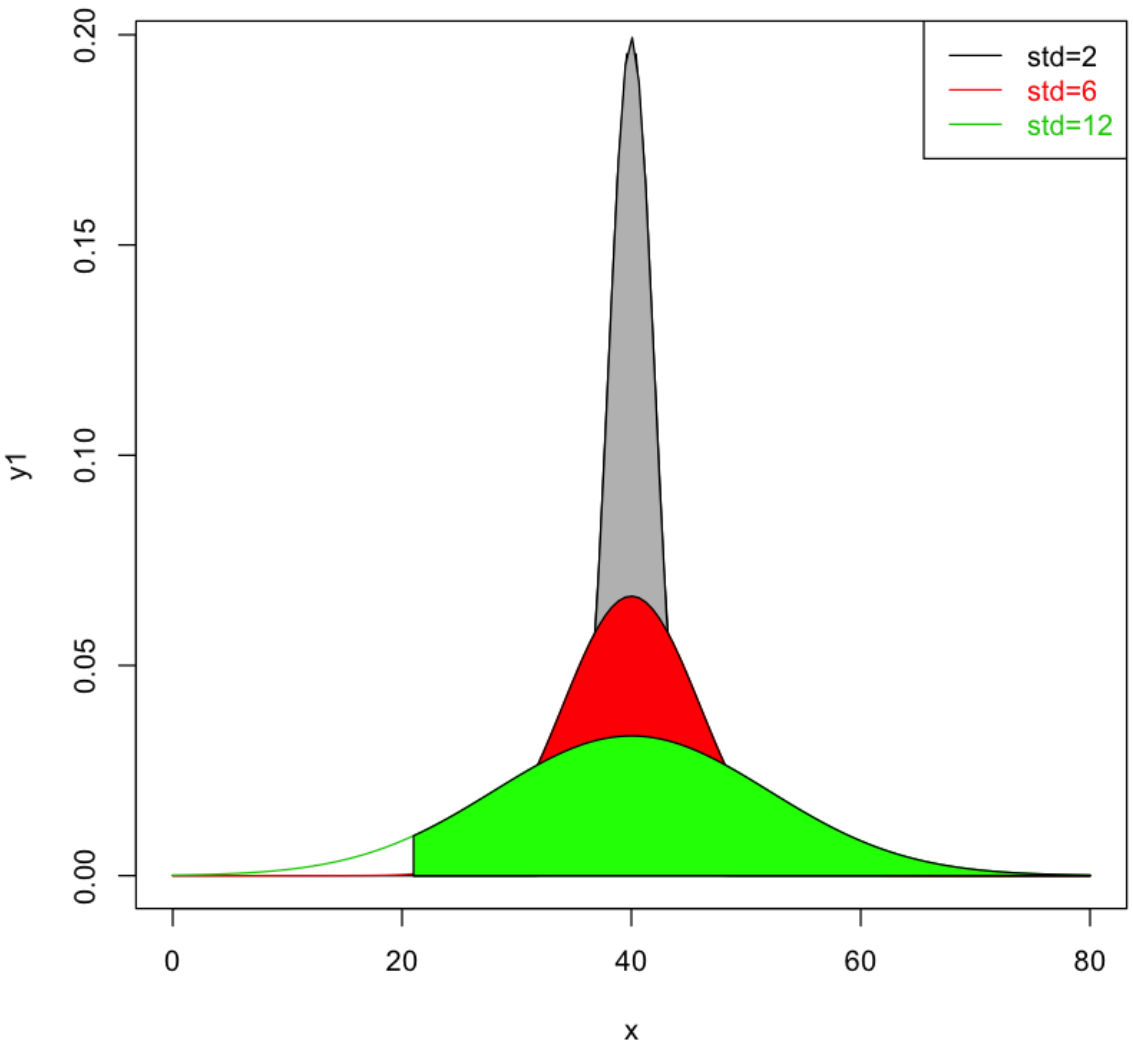
a. P(x < 40)

b. P(x > 21)

In [60]:
```
x = seq(0,80,length=100)
y1=dnorm(x, mean = 40, sd = 2)

plot(x,y1,type = "l",col=1)

y1 = dnorm(x, mean = 40, sd = 6)
lines(x,y1, col = 2)

y1 = dnorm(x, mean = 40, sd = 12)
lines(x,y1, col = 3)

legend("topright",
        legend = c("std=2",
                   "std=6",
                   "std=12"),
      col = c(1,2,3),
      text.col = c(1,2,3),
       lty = c(1, 1, 1),
        lwd = c(1, 1, 1))

x=seq(0,40,length=100)
y=dnorm(x, mean = 40, sd = 2)
polygon(c(0,x,40),c(0,y,0),col="gray")

y=dnorm(x, mean = 40, sd = 6)
polygon(c(0,x,40),c(0,y,0),col="red")

y=dnorm(x, mean = 40, sd = 12)
polygon(c(0,x,40),c(0,y,0),col="green")
```


std=2
std=6
std=12
y1
0.00
0.05
0.10
0.15
0.20
0
20
40
60
80
x

In [61]:
```
x = seq(0,80,length=100)
y1=dnorm(x, mean = 40, sd = 2)

plot(x,y1,type = "l",col=1)

y1 = dnorm(x, mean = 40, sd = 6)
lines(x,y1, col = 2)

y1 = dnorm(x, mean = 40, sd = 12)
lines(x,y1, col = 3)

legend("topright",
        legend = c("std=2",
                           "std=6",
                           "std=12"),
      col = c(1,2,3),
      text.col = c(1,2,3),
       lty = c(1, 1, 1),
        lwd = c(1, 1, 1))

x=seq(21,80,length=100)
y=dnorm(x, mean = 40, sd = 2)
polygon(c(21,x,80),c(0,y,0),col="gray")

y=dnorm(x, mean = 40, sd = 6)
polygon(c(21,x,80),c(0,y,0),col="red")

y=dnorm(x, mean = 40, sd = 12)
polygon(c(21,x,80),c(0,y,0),col="green")
```

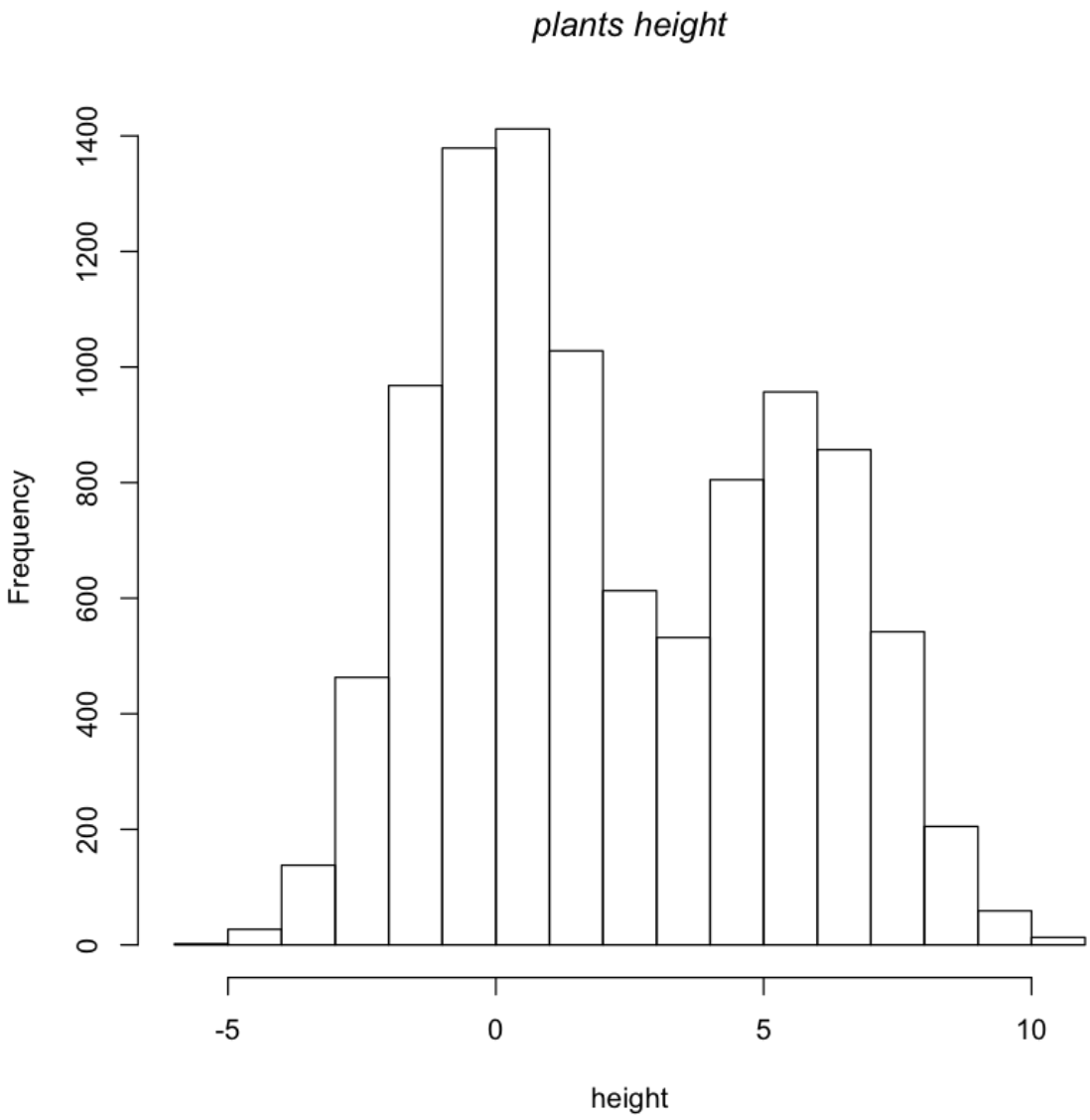


C. The file plant_heights.csv contains the heights for 1000 plants chosen along a altitudinal gradient (from low to high elevation) in a natural park, the aim of the study is to understand canopy density across a gradient.

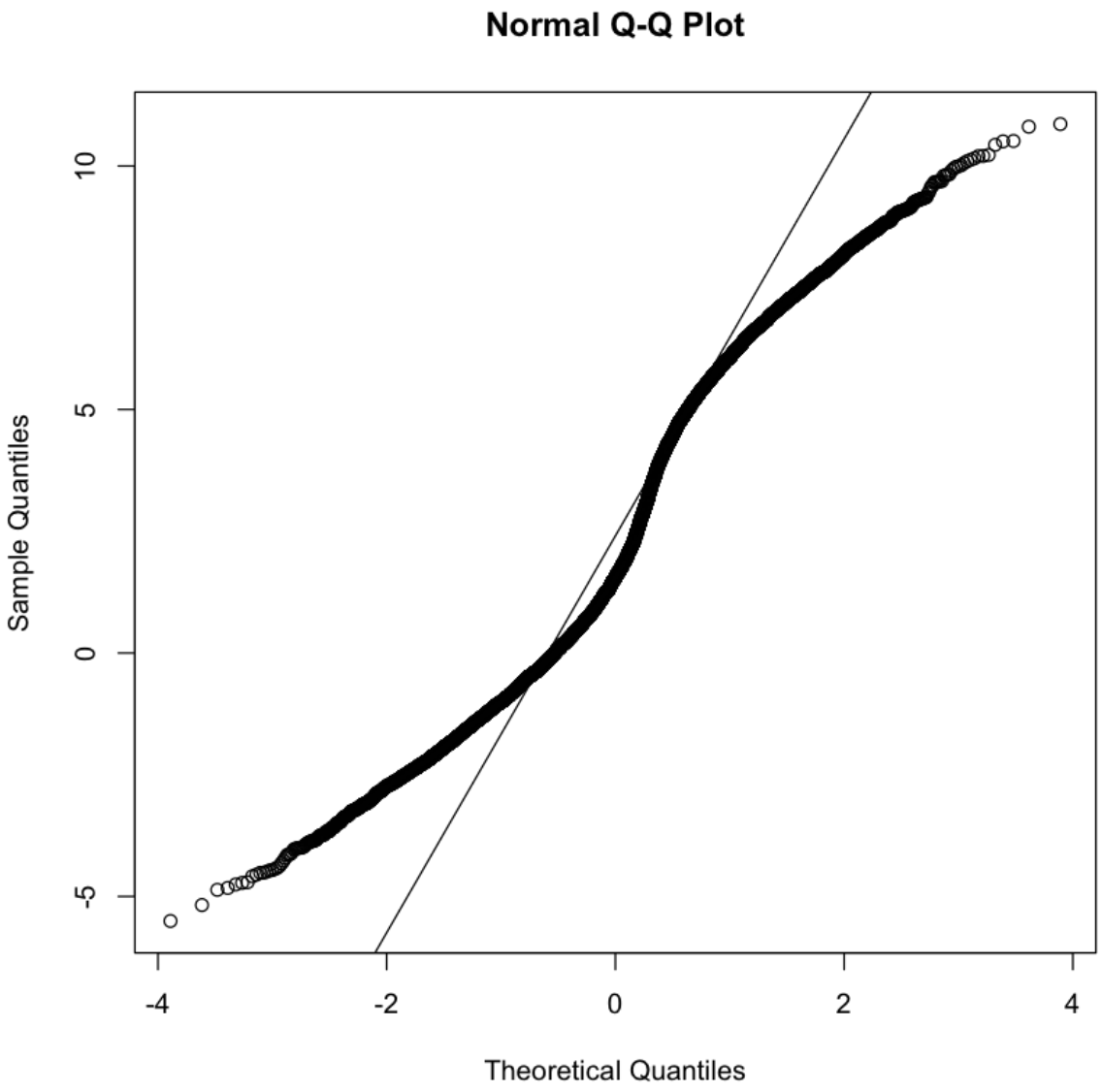
Apply some of the descriptive statistics we have seem so far (histograms, density plots, qqplots) to explain the behavior of the data, write a summary paragraph that gives an initial overview of what you see in this data (are the plants randomly distributed across the gradient? what kind of distribution does the data have and how can you explain it in a biological sense.

hint(try to plot the data using the log transformation, which will make the patterns more evident)

In [70]:
```
heights = read.table(file = "plant_heights.csv", header = TRUE)
hist(log(heights$x,2), xlab = "height", main = expression(italic("plants
 height")))
```



In [72]:
```
x=log(heights$x,2)
qqnorm(x)
qqline(x)
```



In [ ]:
```
From the hist graph, we can see most of plants heights are in the range
(2^-2,2^8). Its log transforamtion almost statisfies with the normal
distribution. Therefore, heights are not randomly distributed along the
gradient. Most of them has the smaller heights and at the lower gradien
t.
It has logarithmic normal distribution. From biological case, at the low
er
gradient zones, they have more water and sunshine for plants.
```