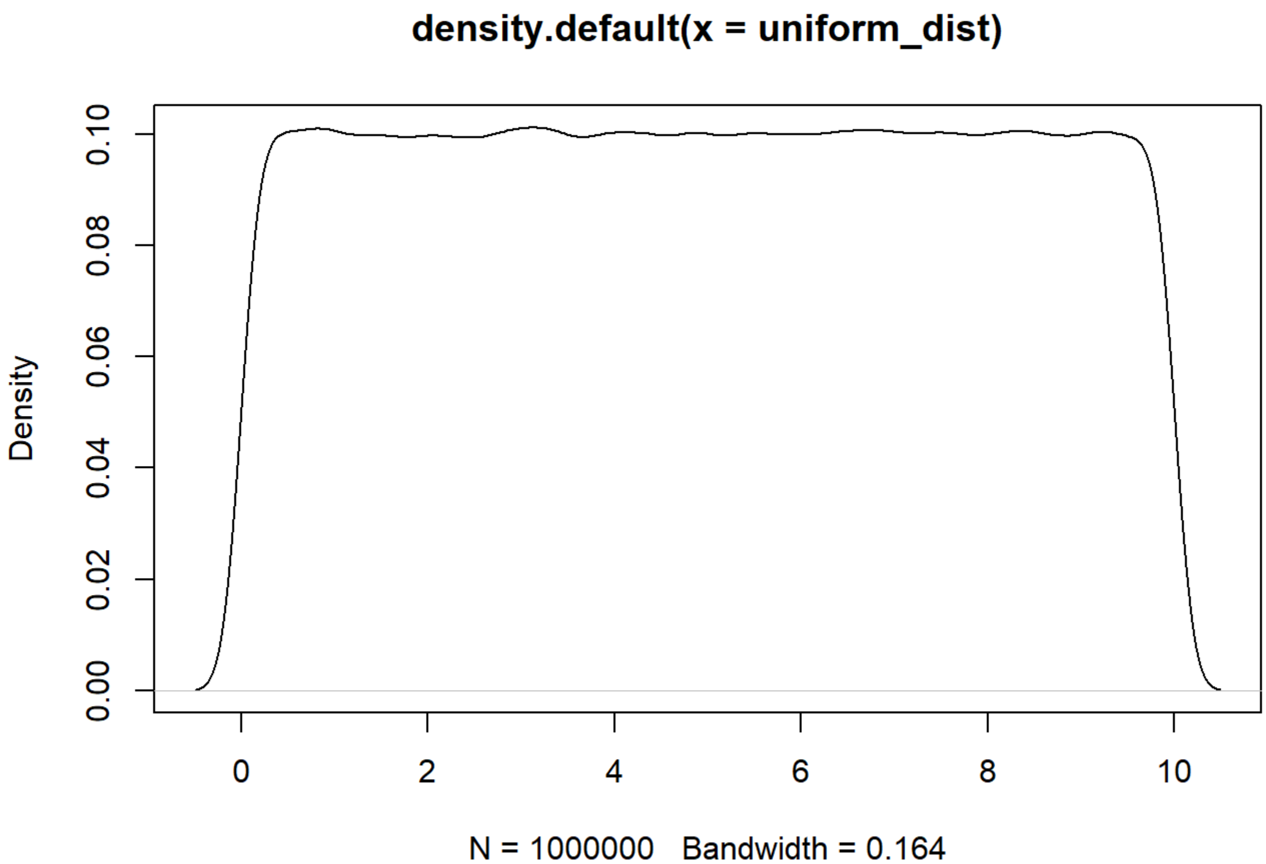
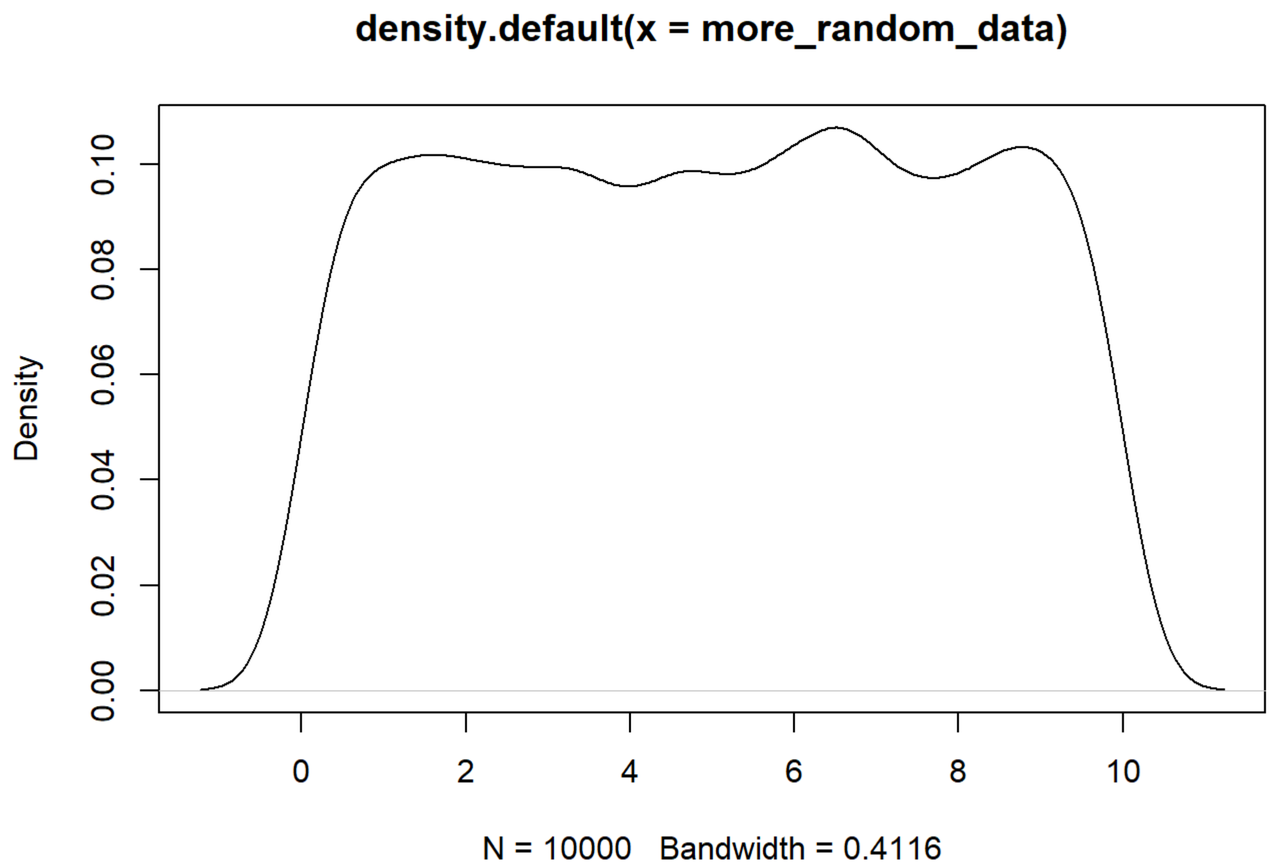
```
#it has the property that for any set B of real numbers

#P{X in B}=integrate(f(x)dx) lim tending to B

#This function is called probability density function






#####R nomenclature for distributions

#For every distribution there are four commands. The commands for each distribution are prepende
d with
#a letter to indicate the functionality:

  #"d"  returns the height of the probability density function
  #"p"  returns the cumulative density function
  #"q"  returns the inverse cumulative density function (quantiles)
 # "r"  returns randomly generated numbers








######THE UNIFORM DISTRIBUTION##########

#The uniform distribution is a probability distribution where each value within a certain range
#is equally liely to occur and values outside the range never occur.If we make a density plot
#of a uniform distribution,it appears flat because no value is more likely(hence same density) t
han any other value

##creating data for a uniform distribution
uniform_dist<-runif(1000000,0,10)
#plotting the distribution
plot(density(uniform_dist))
```




```
more_random_data<-runif(n=10000,min=0,max=10)
plot(density(more_random_data))
```




```
#The prefix "p" is used to determine the
#probability that an observation drawn from a distribution falls below a specified value (known
 as the
#cumulative distribution function.). In essence, "p" gives you the area under the distribution's

#density curve to the left of a certain value on the x axis. For example, in the uniform distrib
ution
#above, there is a 25% chance that an observation will be in the range 0 to 2.5 and a 75% chance
 it will
#fall in the range 2.5 to 10.


#The p function gives the area covered in the pdf

punif(q=2.5,min=0,max=10)
```

```
## [1] 0.25
```

```
#The prefix "q" is the inverse of the prefix "p": it
#returns the cutoff value (quantile) associated with a given probability.

qunif(p=0.4,min=0,max=10)
```

```
## [1] 4
```

```
#Finally the "d" prefix gives you the density (height of the density curve)
#at a given point. Our plot of randomly generated uniform data was flat at
#0.1 so the density of the distribution should be constant at 0.1 throughout the range 0 to 10:


#checking the height of the curve at diff values
dunif(x=0,min=0,max=10)
```

```
## [1] 0.1
```

```
dunif(5,0,10)
```

```
## [1] 0.1
```

```
dunif(7.5,0,10)
```

```
## [1] 0.1
```

```
dunif(10,0,10)
```

```
## [1] 0.1
```

```
#####NORMAL DISTRIBUTION###########################


#The normal or Gaussian distribution is a continuous probability distribution
#characterized by a symmetric bell-shaped curve. A normal distribution is defined by its
#center (mean) and spread (standard deviation.). The bulk of the observations generated from
#a normal distribution lie near the mean, which lies at the exact center of the distribution:
# as a rule of thumb, about 68% of the data lies within 1 standard deviation of the mean,
#95% lies within 2 standard deviations and 99.7% lies within 3 standard deviations.


#a standard normal curvel has mena=0 and  sd =1

normal_dist<-rnorm(1000000,mean=0,sd=1)
plot(density(normal_dist))


##geting prob value at a particular point

pnorm(q=0.5,mean=0,sd=1)
```

```
## [1] 0.6914625
```

```
pnorm(q=1,mean=0,sd=1)
```

```
## [1] 0.8413447
```

```
##this value should be equal to 0.5+0.34
##84.13% area under the curve at point q=1
##this indiscates that there is 84.13% prob of getting a value <=1

#prob of values lesser than -1

prob_less_minusone<-pnorm(q=-1,mean=0,sd=1)
prob_less_minusone
```

```
## [1] 0.1586553
```

```
##getting prob of observing a value greater than 1

1-pnorm(q=1,mean=0,sd=1)
```

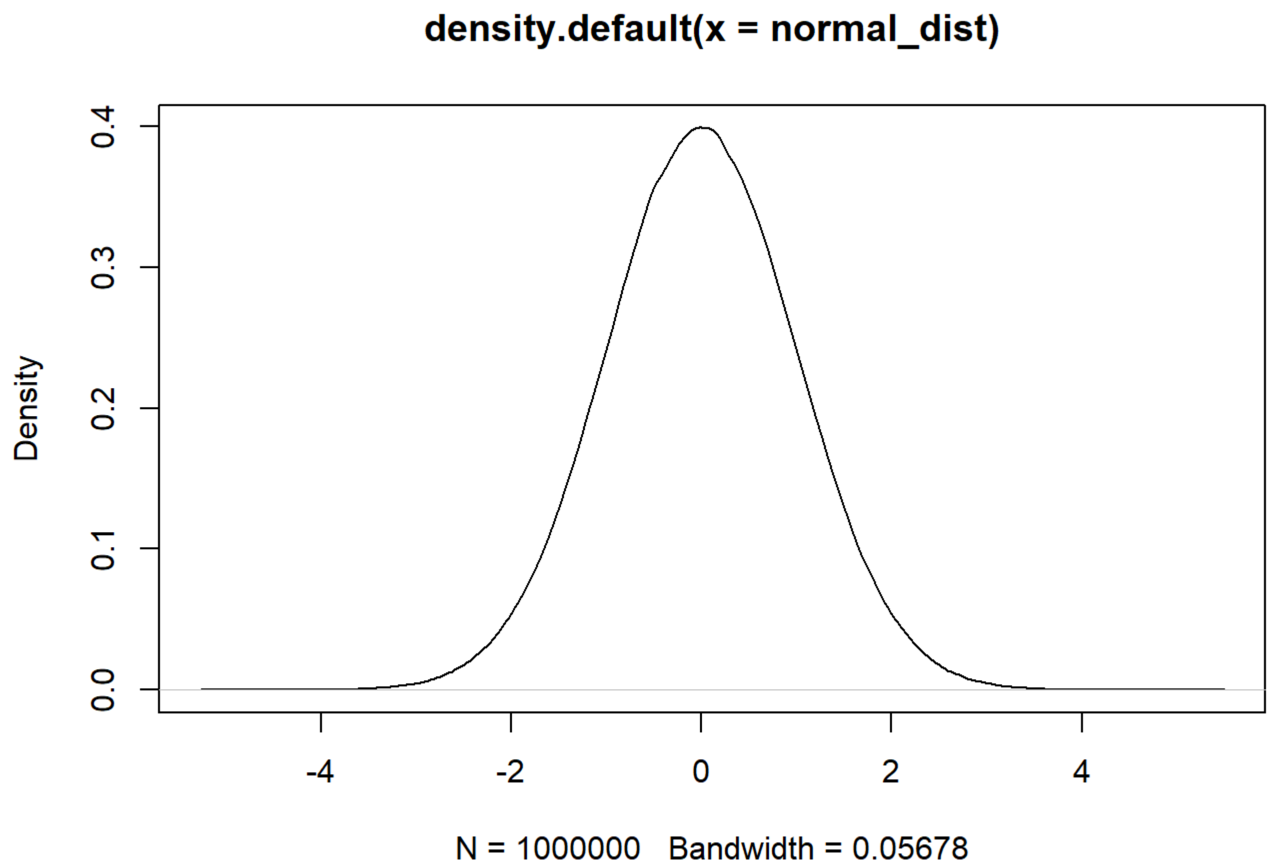
```
## [1] 0.1586553
```

```
#this indicates that there is 15.86% prob of getting a value greater than 1


##getting a prob value between -1 and +1

pnorm(q=1,mean=0,sd=1)-pnorm(q=-1,mean=0,sd=1)
```

```
## [1] 0.6826895
```

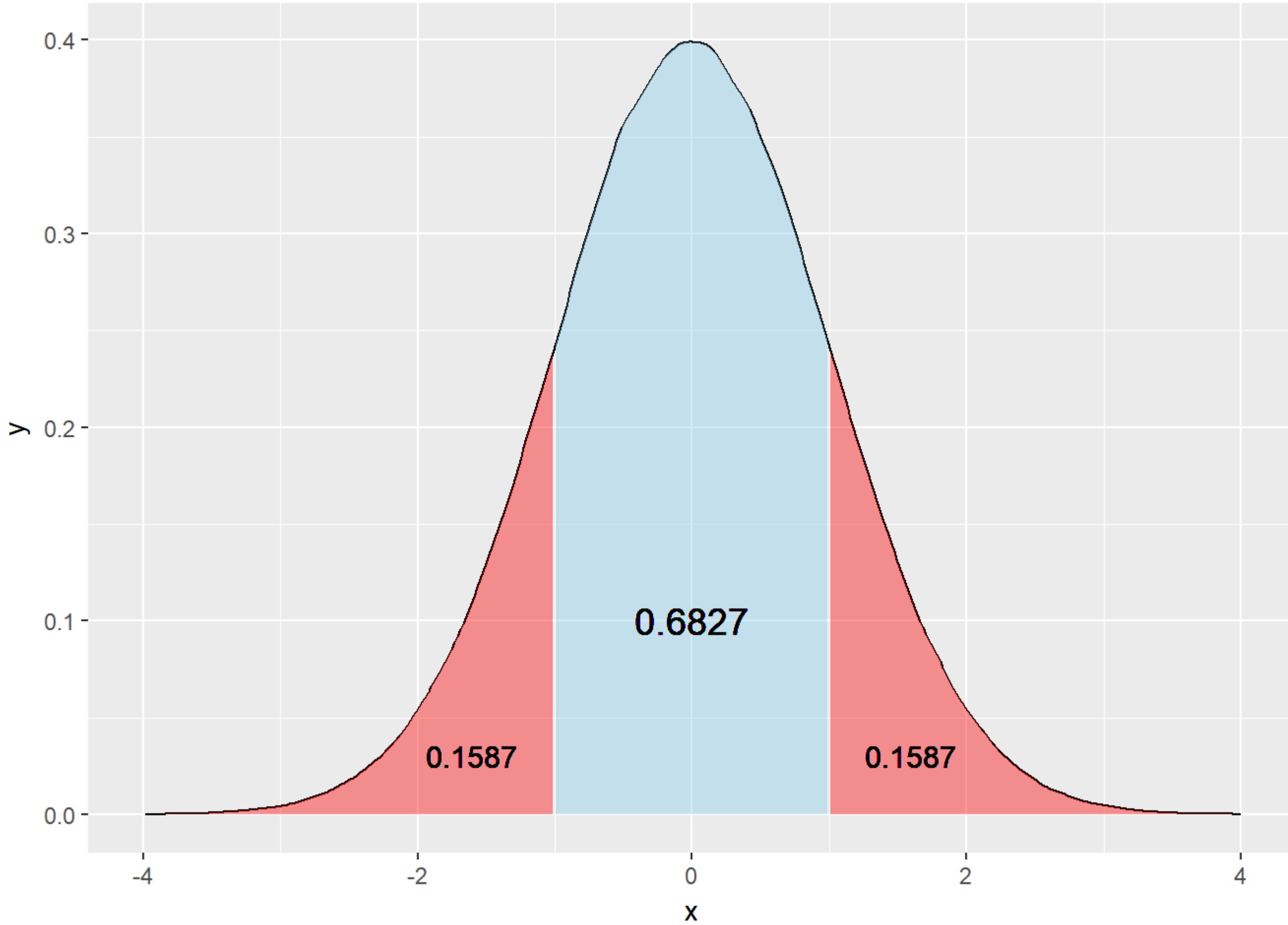
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```


density.default(x = normal_dist)
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
N = 1000000 Bandwidth = 0.05678

```
# Plot the density curve with the cutoff areas
norm_frame = with(density(normal_dist),  # Create data frame density values
                  data.frame(x,y))

myplot <- ggplot(data = norm_frame, aes(x = x, y = y)) +   # Create the plot
  geom_line() +
  geom_ribbon(data=subset(norm_frame,x < -1),
              aes(ymax=y),
              ymin=0,
              fill="red",
              alpha=0.4) +
  geom_ribbon(data=subset(norm_frame,x > 1),
              aes(ymax=y),
              ymin=0,
              fill="red",
              alpha=0.4) +
  geom_ribbon(data=subset(norm_frame,x > -1 & x < 1),
              aes(ymax=y),
              ymin=0,
              fill="skyblue",
              alpha=0.4) +
  geom_text(x=-1.6,y=0.03,label=round(prob_less_minusone,4),size=4) +
  geom_text(x=1.6,y=0.03,label=round(prob_less_minusone,4),size=4) +
  geom_text(x=0,y=0.1,label=round(1-(prob_less_minusone*2),4),size=5) +
  xlim(-4,4)

myplot
```

```
## Warning: Removed 132 rows containing missing values (geom_path).
```




```
###finding quantile values

qnorm(p=0.025,mean=0,sd=1)
```

```
## [1] -1.959964
```

```
qnorm(p=0.975,mean=0,sd=1)
```

```
## [1] 1.959964
```

```
###checking the height of the curve
##density at that particular point in the pdf

dnorm(x=0,mean=0,sd=1)
```

```
## [1] 0.3989423
```

```
dnorm(x=1.959964,mean=0,sd=1)
```

```
## [1] 0.05844507
```

```
##thus we see when set cut of value of 0.05 for p value the cut-off value
##for z is 1.96



############Exponential distribution#################

#A continuous random variable whose probability function is given as
    #f(x)=lambda*e^(-lambda*x)   if x>=0
    #    =0            if x<0
#is called an exponentail random variable with parameter lambda.

#Finally the cumulative value of this function is given as follows
    #F(a)=1-e^(lambda*a)

#Mean or expected value=1/lambda
#variance=1/lambda^2

#Use of exponential distribution is given by
#In practice, the exponential distribution often arises as the distribution of the
#amount of time until some specific event occurs. For instance, the amount of time
#(starting from now) until an earthquake occurs, or until a new war breaks out, or
#until a telephone call you receive turns out to be a wrong number are all random
#variables that tend in practice to have exponential distributions.


#example problem

#Suppose that the length of a phone call in minutes is an exponential random variable
#with parameter lambda = 1
#10 . If someone arrives immediately ahead of you at a public
#telephone booth, find the probability that you will have to wait
#more than 10 minutes
#between 10 and 20 minutes
```

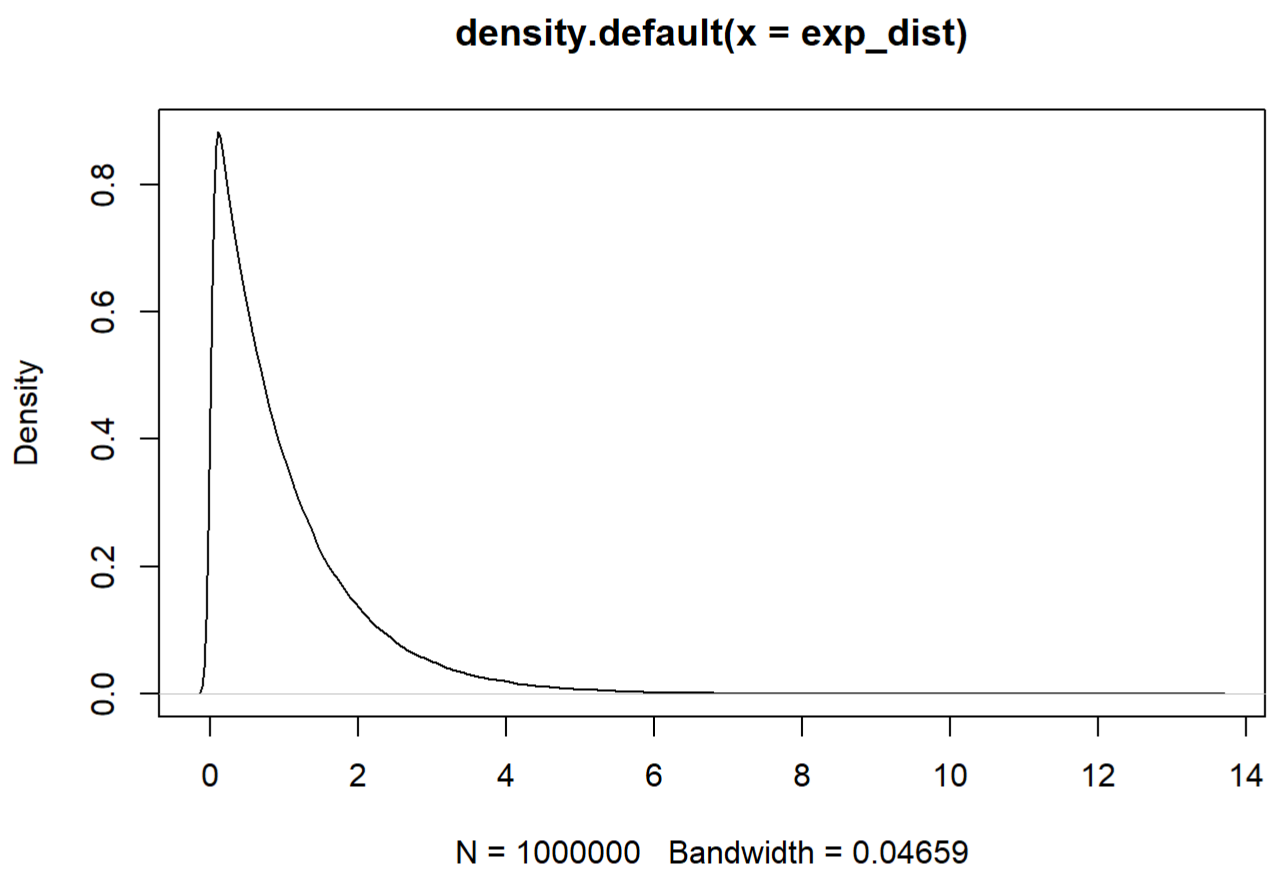
```
1-pexp(q=10,rate=1/10)
```

```
## [1] 0.3678794
```

```
pexp(q=20,rate=1/10)-pexp(q=10,rate=1/10)
```

```
## [1] 0.2325442
```

```
##creating exponential probability distributions

exp_dist<-rexp(n=1000000,rate=1)
plot(density(exp_dist))
```




```
pexp(q=1,rate=1,lower.tail = FALSE)
```

```
## [1] 0.3678794
```

```
1-(pexp(q=1,rate=1))
```

```
## [1] 0.3678794
```

```
p_longer_than_1<-(1-(pexp(q=1,rate=1)))
p_longer_than_1
```

```
## [1] 0.3678794
```

```
##the average arrival time in an exp dist is 1/lambda

##checking particular quantile value

qexp(p=0.6321206,rate=1)
```

```
## [1] 1
```

```
#checking the height of the curve
dexp(x=0,rate=1)
```

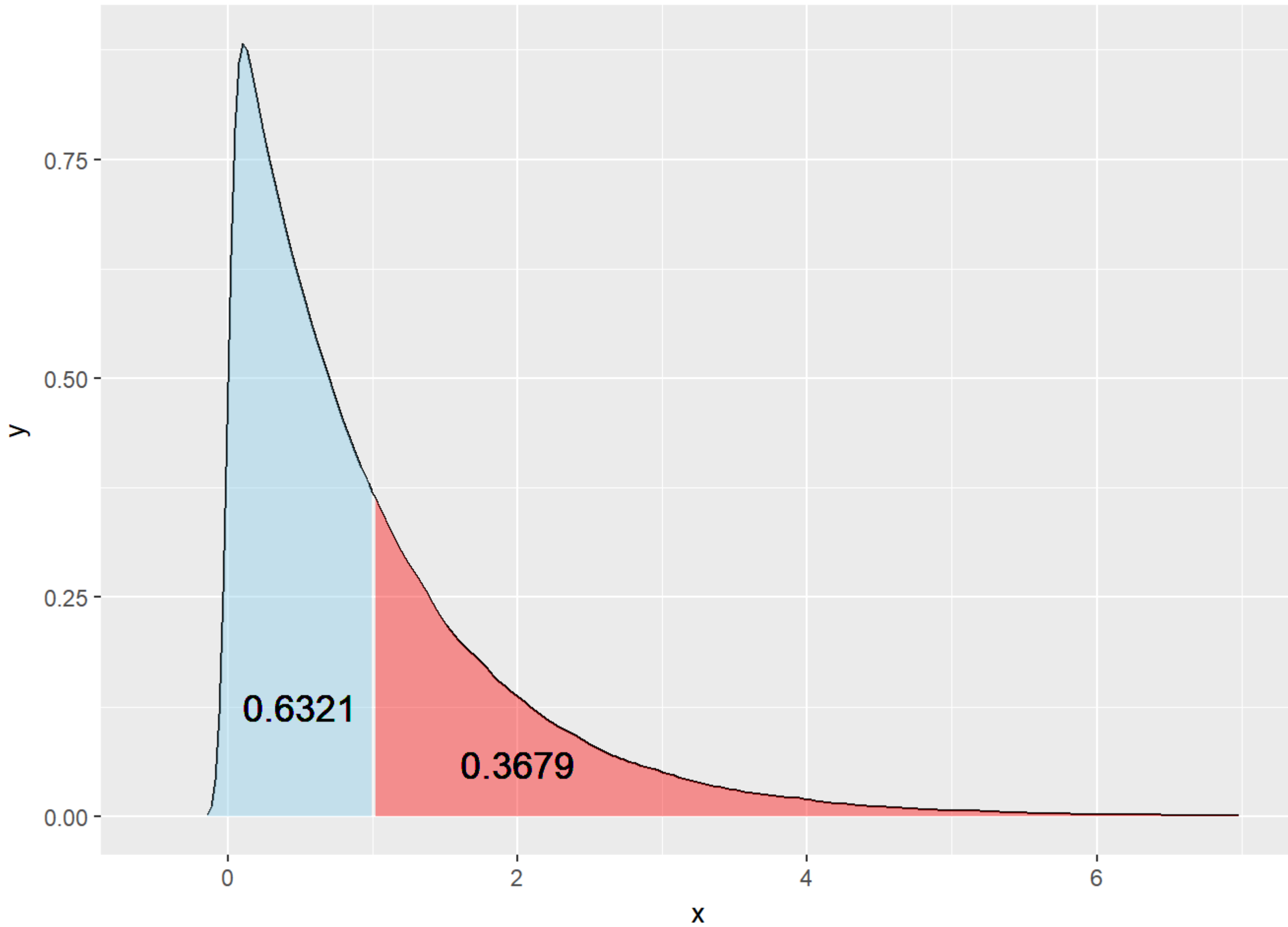
```
## [1] 1
```

```
dexp(x=2,rate=1)
```

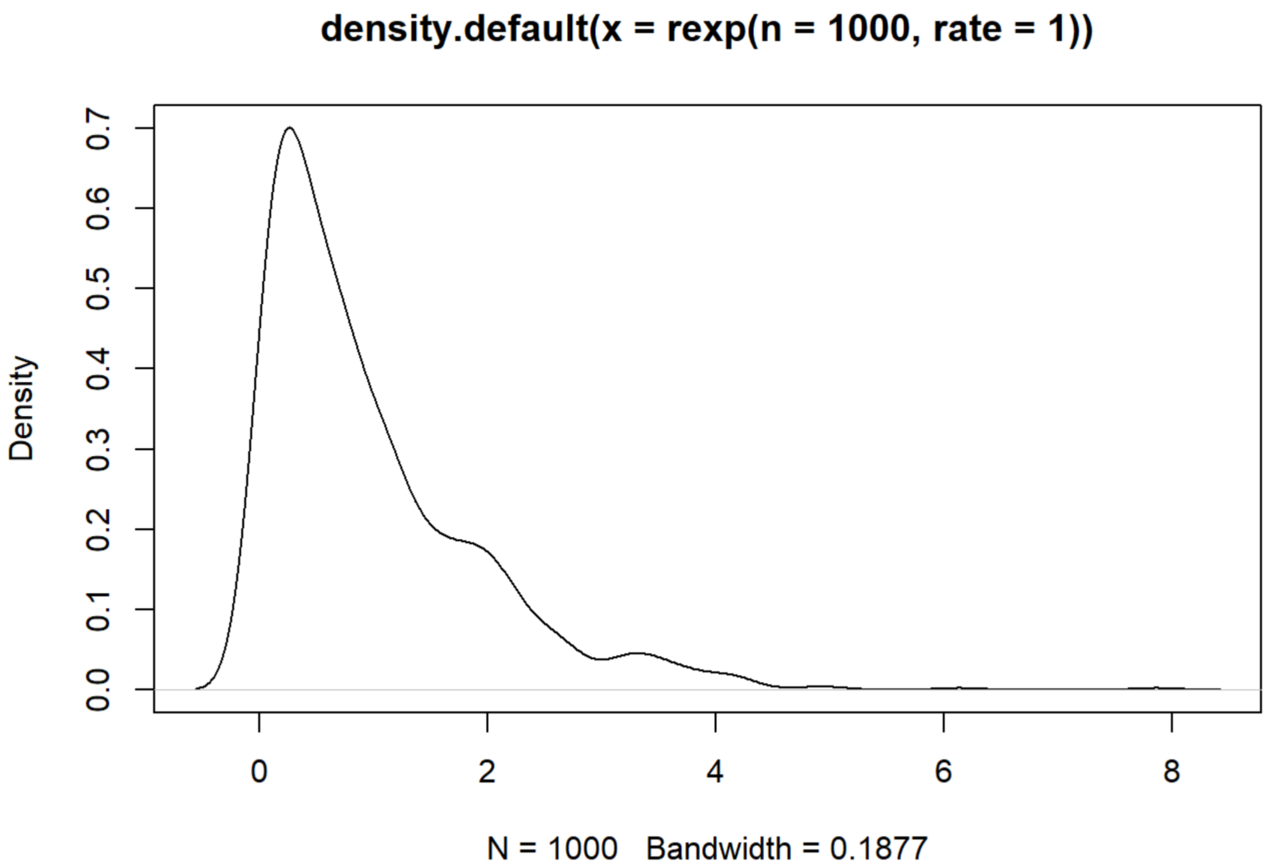
```
## [1] 0.1353353
```

```
#Plot the density curve of the exp dist

exp_frame = with(density(exp_dist),  # Create data frame of x,y density values
                 data.frame(x,y))

myplot2 <- ggplot(data = exp_frame, aes(x = x, y = y)) +   # Create the plot
  geom_line() +
  geom_ribbon(data=subset(exp_frame,x > 1),
              aes(ymax=y),
              ymin=0,
              fill="red",
              alpha=0.4)+
  geom_ribbon(data=subset(exp_frame,x < 1),
              aes(ymax=y),
              ymin=0,
              fill="skyblue",
              alpha=0.4) +
  geom_text(x=2,y=0.06,label=round(p_longer_than_1,4), size=5) +
  geom_text(x=0.5,y=0.125,label=round(1-p_longer_than_1,4), size=5) +
  xlim(-0.5,7)

myplot2
```

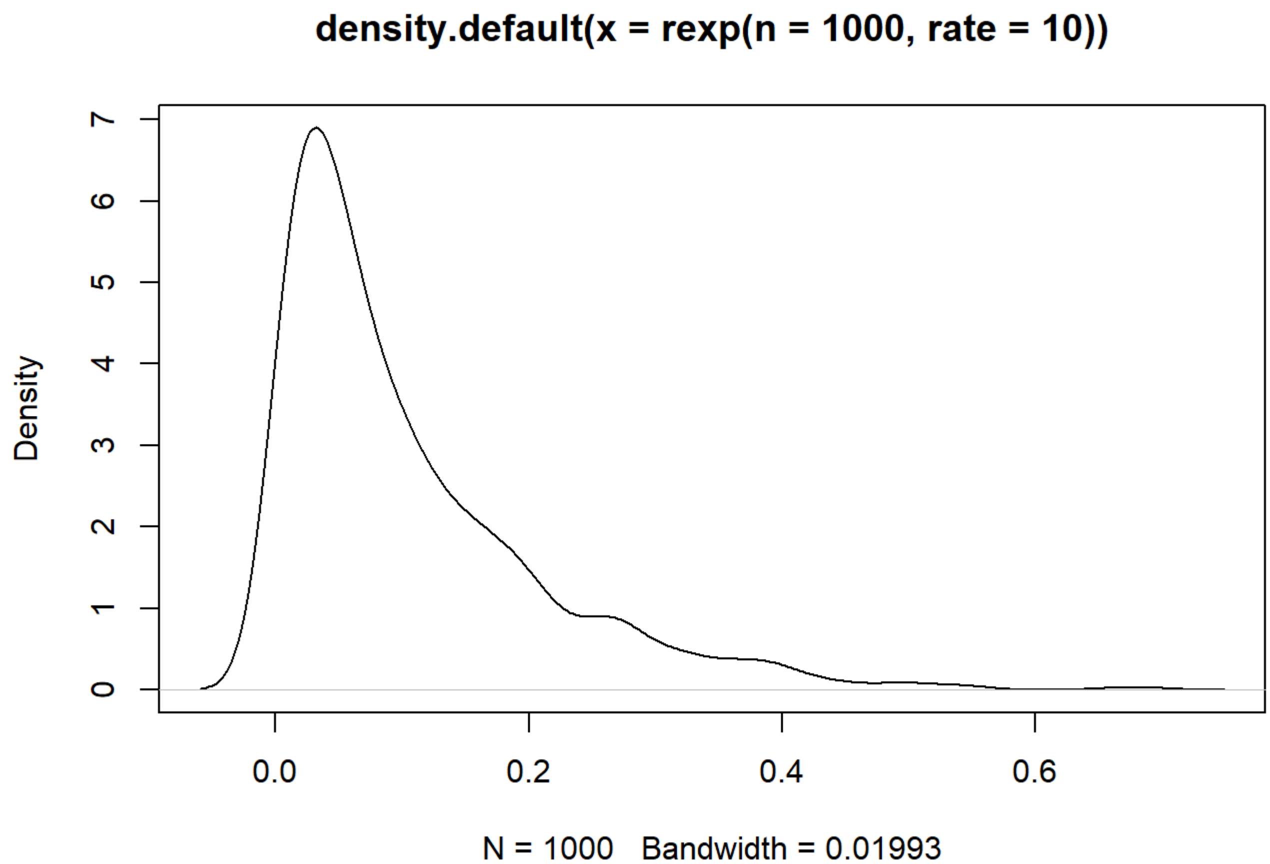
```
## Warning: Removed 248 rows containing missing values (geom_path).
```



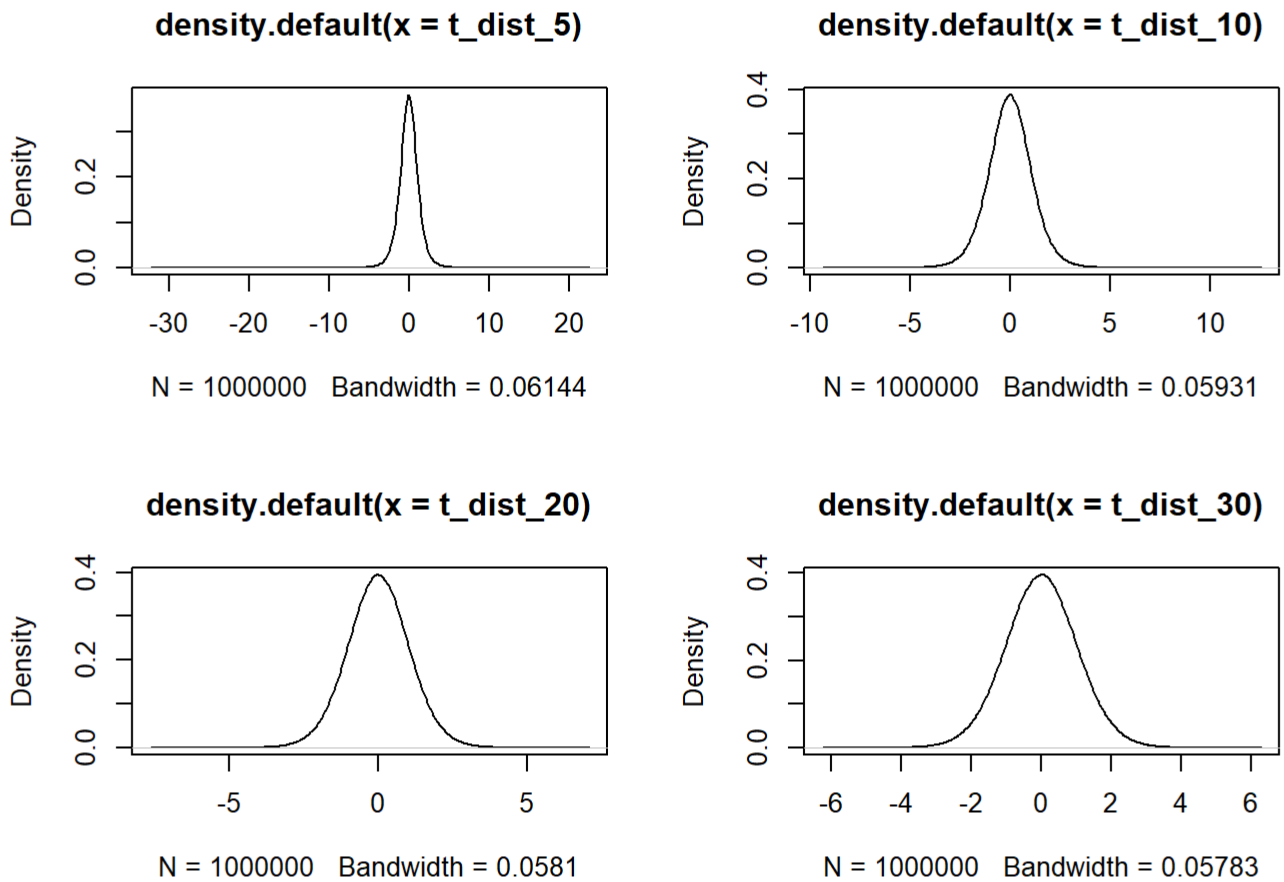

```
plot(density(rexp(n=1000,rate=1)))
```



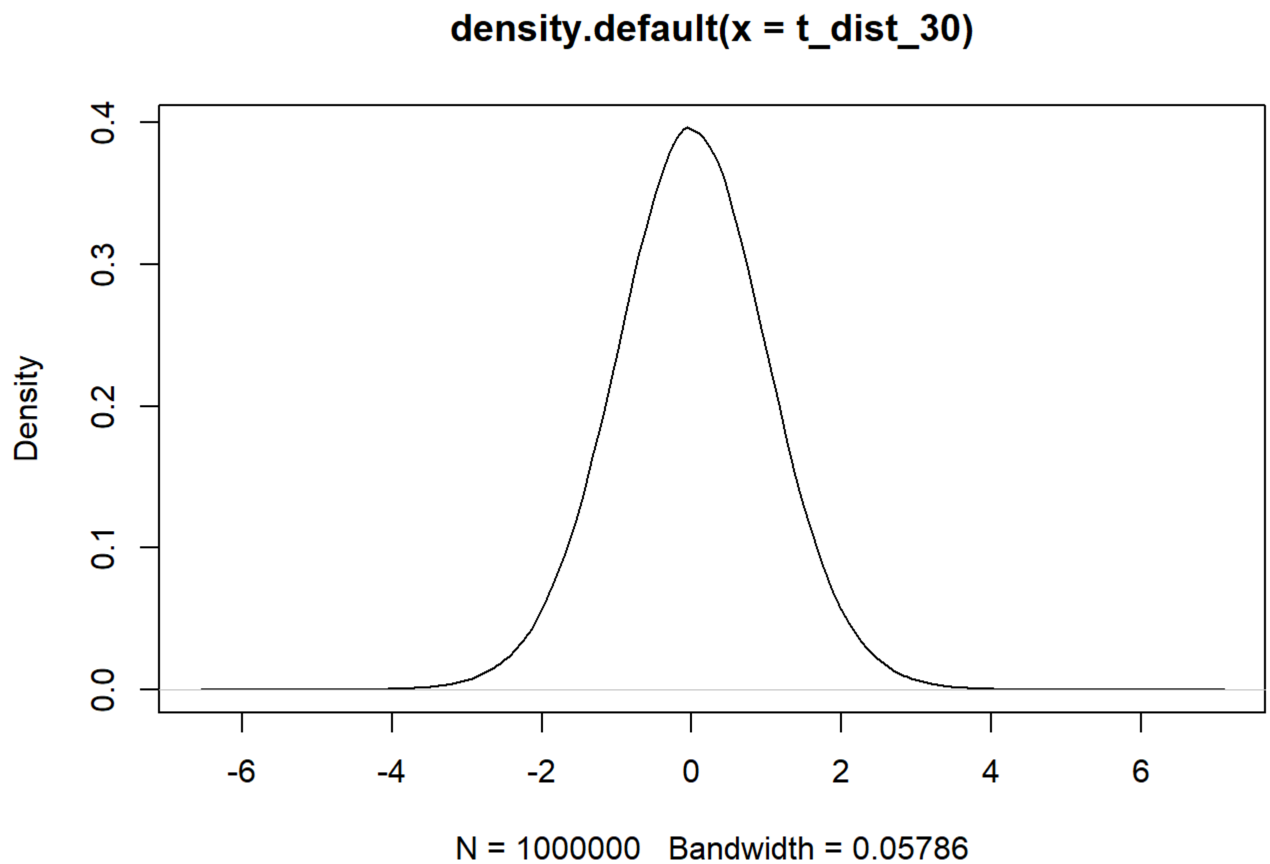

```
plot(density(rexp(n=1000,rate=10)))
```


density.default(x = rexp(n = 1000, rate = 10))
Density
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
N = 1000 Bandwidth = 0.01993

```
#as the value of rate increases the value of x decreases in the exp dist but
#the value of density increases




############T Distribution ##################################

#In probability and statistics, Student's t-distribution (or simply the t-distribution)
#is any member of a family of continuous probability distributions that arises when estimating
#the mean of a normally distributed population in situations where the sample size is small and
#population standard deviation is unknown.

##T_dist considers mean =0 and sd=1

#Generating a random t dist with diff degree of freedom

par(mfrow=c(2,2))


t_dist_5<-rt(n=1000000,df=5)
plot(density(t_dist_5))

t_dist_10<-rt(n=1000000,df=10)
plot(density(t_dist_10))

t_dist_20<-rt(n=1000000,df=20)
plot(density(t_dist_20))

t_dist_30<-rt(n=1000000,df=30)
plot(density(t_dist_30))
```




```
###checking prob values

par(mfrow=c(1,1))

t_dist_30<-rt(n=1000000,df=30)
plot(density(t_dist_30))
```




```
pt(q=0,df=30)
```

```
## [1] 0.5
```

```
pt(q=0,df=10)
```

```
## [1] 0.5
```

```
##we see the prob value is 0.5 since this is the mean value

qt(p=0.5,df=30)
```

```
## [1] 0
```

```
qt(p=0.99,df=30)
```

```
## [1] 2.457262
```

```
##checking the height of the curve

dt(x=0,df=30)
```

```
## [1] 0.3956322
```