STATISTICS LAB
ASSIGNMENT 1

Name: Shreya Maheshwari

Reg. no.: 18BCE0167

Q1) Given matrices A and B

a) Find transpose of A+B

```
> A=matrix(c(2,4,4,7,4,5,5,3,1), nrow=3, ncol=3, byrow=TRUE)
> A
     [,1] [,2] [,3]
[1,]    2    4    4
[2,]    7    4    5
[3,]    5    3    1
> B=matrix(c(9,3,5,8,6,24,1,3,2), nrow=3, ncol=3, byrow=TRUE)
> B
     [,1] [,2] [,3]
[1,]    9    3    5
[2,]    8    6   24
[3,]    1    3    2
> C=A+B
> C
     [,1] [,2] [,3]
[1,]   11    7    9
[2,]   15   10   29
[3,]    6    6    3
> t(C)
```

```
     [,1] [,2] [,3]
[1,]   11   15    6
[2,]    7   10    6
[3,]    9   29    3
```

b) Find A*B

```
> D=A*B
> D
     [,1] [,2] [,3]
[1,]   18   12   20
[2,]   56   24  120
[3,]    5    9    2
```

c) Access 2nd row and 3rd column element

```
> print(D[2,3])
[1] 120
```

d) Access 3rd row of A*A

```
> print(D[3, ])
[1] 5 9 2
```

Q2. Form the table for the following data.

X=sequence from 1 to 10 with difference 2

Y= letter from a to e

Z= any five names

a)

```
> X=seq(1,10,by=2)
> X
```

```
[1] 1 3 5 7 9
```

b)

```
> Y=head(letters,5)
> Y
[1] "a" "b" "c" "d" "e"
```

c)

```
> Z <- c("iishi", "garima", "hrishita", "shreya", "yukta")
> Z
[1] "iishi"    "garima"   "hrishita" "shreya"   "yukta"
> data.frame(X,Y,Z)
  X Y        Z
1 1 a    iishi
2 3 b   garima
3 5 c hrishita
4 7 d   shreya
5 9 e    yukta
```

Q3. Solve x+y+z=1, x-3y+z=3, 5x+4y+2z=8.

```
> E=matrix(c(1,1,1,1,-3,1,5,4,2), nrow=3, ncol=3, byrow=TRUE)
> E
     [,1] [,2] [,3]
[1,]    1    1    1
[2,]    1   -3    1
[3,]    5    4    2
> F=matrix(c(1,3,8), nrow=3, ncol=1)
> F
     [,1]
[1,]    1
```

```
[2,]    3
[3,]    8
> EX=F
> solve(E,F)
           [,1]
[1,]  2.3333333
[2,] -0.5000000
[3,] -0.8333333
```

Q4. Write the code to find mean for the following data.
2,4,5,2,7,9,4,5,2,8,5,7,3,1

```
> x<-c(2,4,5,2,7,9,4,5,2,8,5,7,3,1)
> x
 [1] 2 4 5 2 7 9 4 5 2 8 5 7 3 1
> mean(x)
[1] 4.571429
```

Q5. Find median and mode for the following distribution

x	0-10	10-20	20-30	30-40	40-50
f	12	8	17	3	10

Median:

```
> x=c("0-10", "10-20", "20-30", "30-40", "40-50")
> f=c(12, 8, 17, 3, 10)
> cf=cumsum(f)
> data.frame(x,f,cf)
      x  f cf
1  0-10 12 12
```

```
2 10-20  8 20
3 20-30 17 37
4 30-40  3 40
5 40-50 10 50
> N=50
> mp=min(which(cf>((N+1)/2)))
> mp
[1] 3
> medianclass=x[mp]
> medianclass
[1] "20-30"
> mf=f[mp]
> mf
[1] 17
> mc=cf[mp-1]
> mc
[1] 20
> l=20
> h=10
> median=l+(h/mf)*(N/2-mc)
> median
[1] 22.94118
```

Mode:

```
> x=c("0-10", "10-20", "20-30", "30-40", "40-50")
> f=c(12, 8, 17, 3, 10)
> cf=cumsum(f)
```

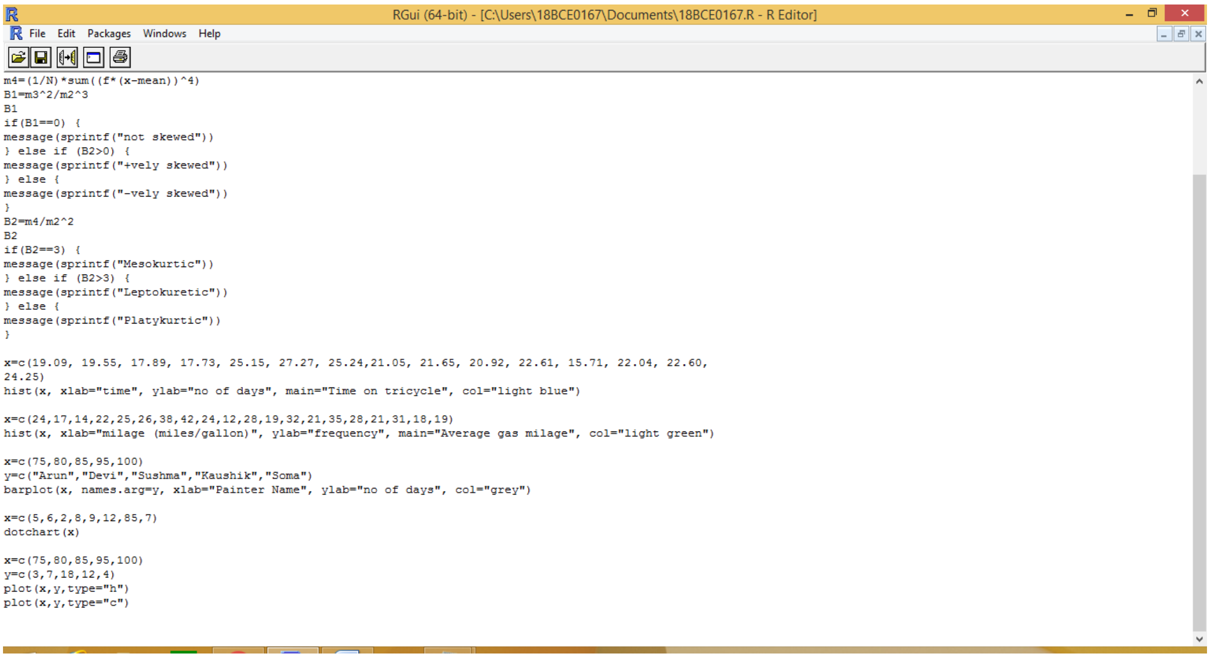
```
> data.frame(x,f,cf)
      x  f cf
1  0-10 12 12
2 10-20  8 20
3 20-30 17 37
4 30-40  3 40
5 40-50 10 50
> mp=which(f==max(f))
> mp
[1] 3
> modalclass=x[mp]
> modalclass
[1] "20-30"
> l=20
> h=10
> f0=f[mp-1]
> f1=f[mp]
> f2=f[mp+1]
> mode=l+(h*(f1-f0)/(2*f1-f0-f2))
> mode
[1] 23.91304
```

<u>STATS LAB ASSIGNMENT 2</u>

SHREYA MAHESHWARI

18BCE0167

Q1)



CODE:

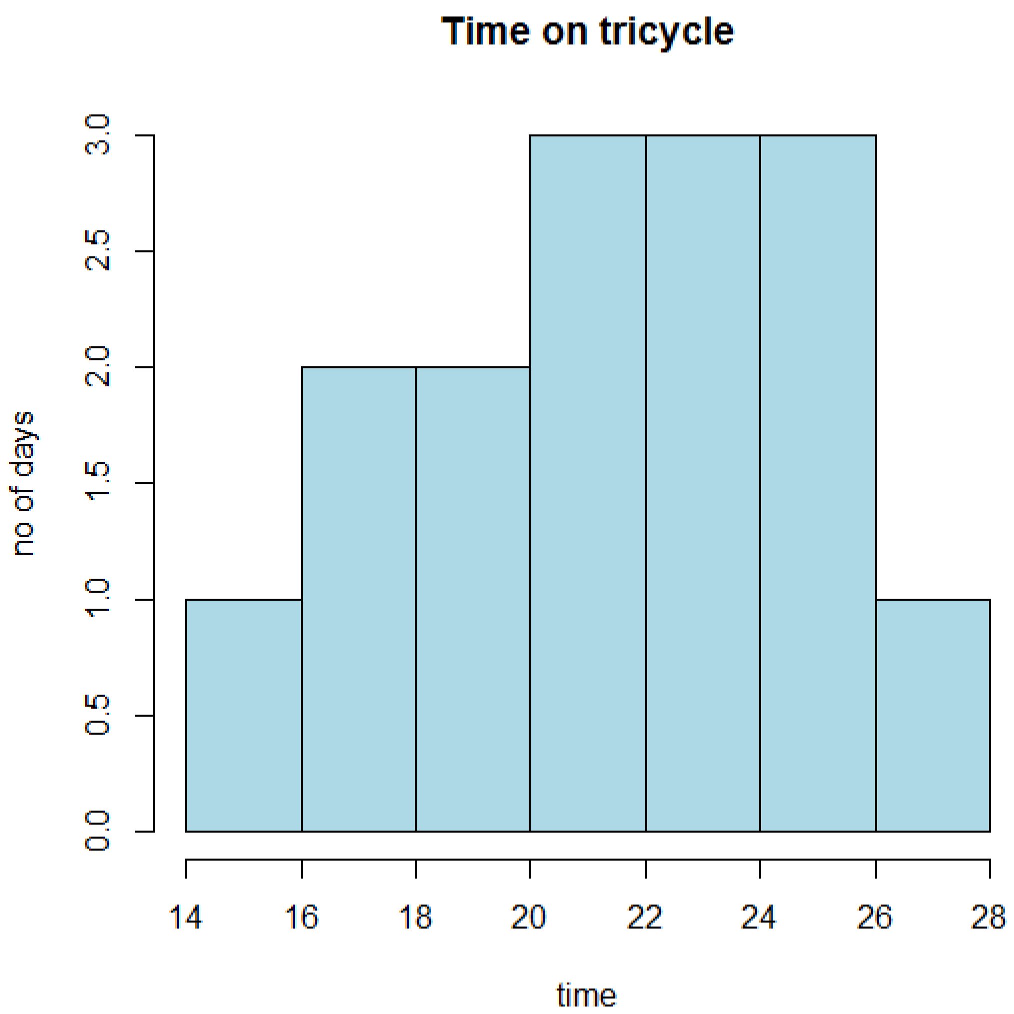
```
> f=c(3,7,18,12,4)

> data.frame(x,f)

    x  f

1  75  3

2  80  7

3  85 18

4  95 12

5 100  4

> N=sum(f)

> mean = sum(x*f)/N

> sd = sqrt(sum(f*x^2)/N - mean^2)

> cv = sd*100/mean
```

```
> m2=(1/N)*sum((f*(x-mean))^2)

> m3=(1/N)*sum((f*(x-mean))^3)

> m4=(1/N)*sum((f*(x-mean))^4)

> B1=m3^2/m2^3

> B1

[1] 2.398862

> if(B1==0) {

+ message(sprintf("not skewed"))

+ } else if (B2>0) {

+ message(sprintf("+vely skewed"))

+ } else {

+ message(sprintf("-vely skewed"))

+

+vely skewed

> B2=m4/m2^2

> B2

[1] 12.9078

> if(B2==3) {

+ message(sprintf("Mesokurtic"))

+ } else if (B2>3) {

+ message(sprintf("Leptokuretic"))

+ } else {

+ message(sprintf("Platykurtic"))

+ }

Leptokuretic

>
```

Q2)

CODE:

```
> x=c(19.09, 19.55, 17.89, 17.73, 25.15, 27.27, 25.24,21.05, 21.65, 20.92, 22.61, 15.71, 22.04, 22.60,
+ 24.25)
> hist(x, xlab="time", ylab="no of days", main="Time on tricycle", col="light blue")
>
```

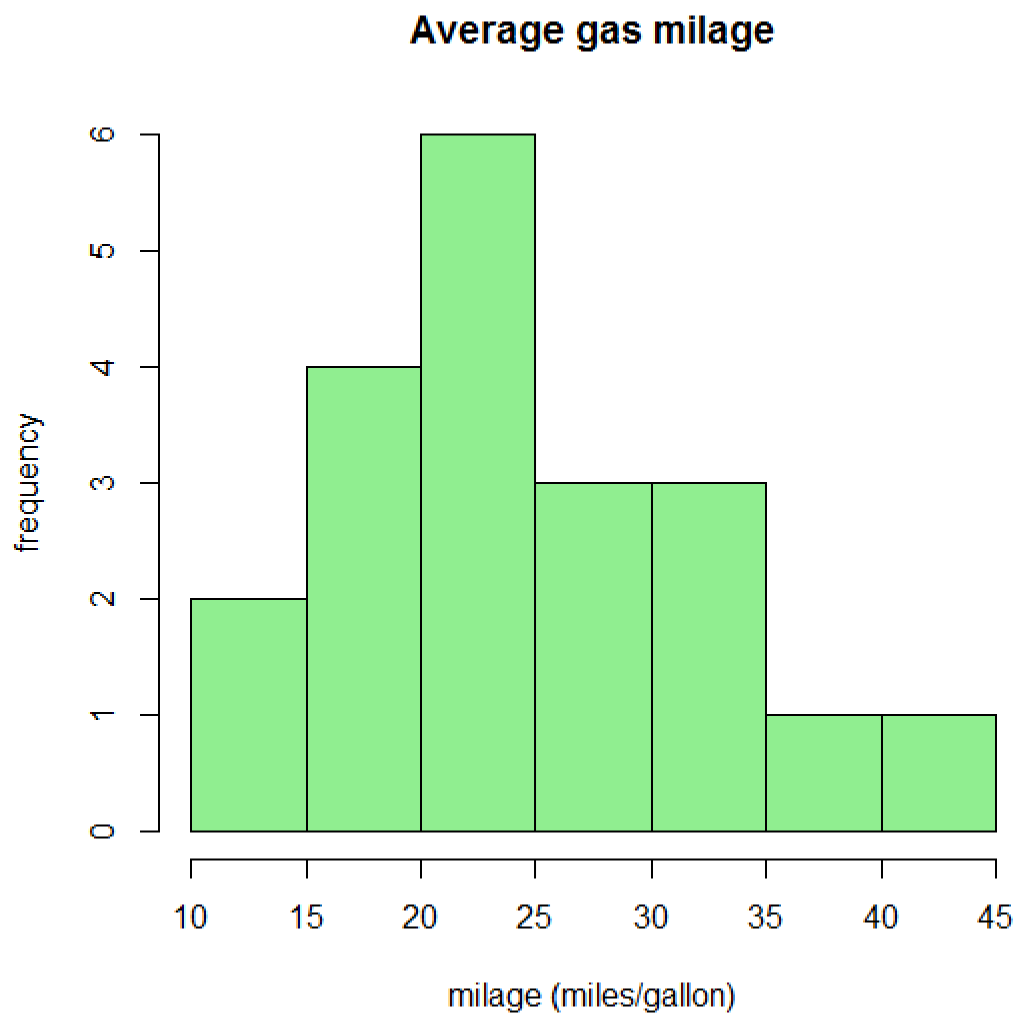


Q3)

CODE:

```
> x=c(24,17,14,22,25,26,38,42,24,12,28,19,32,21,35,28,21,31,18,19)
```

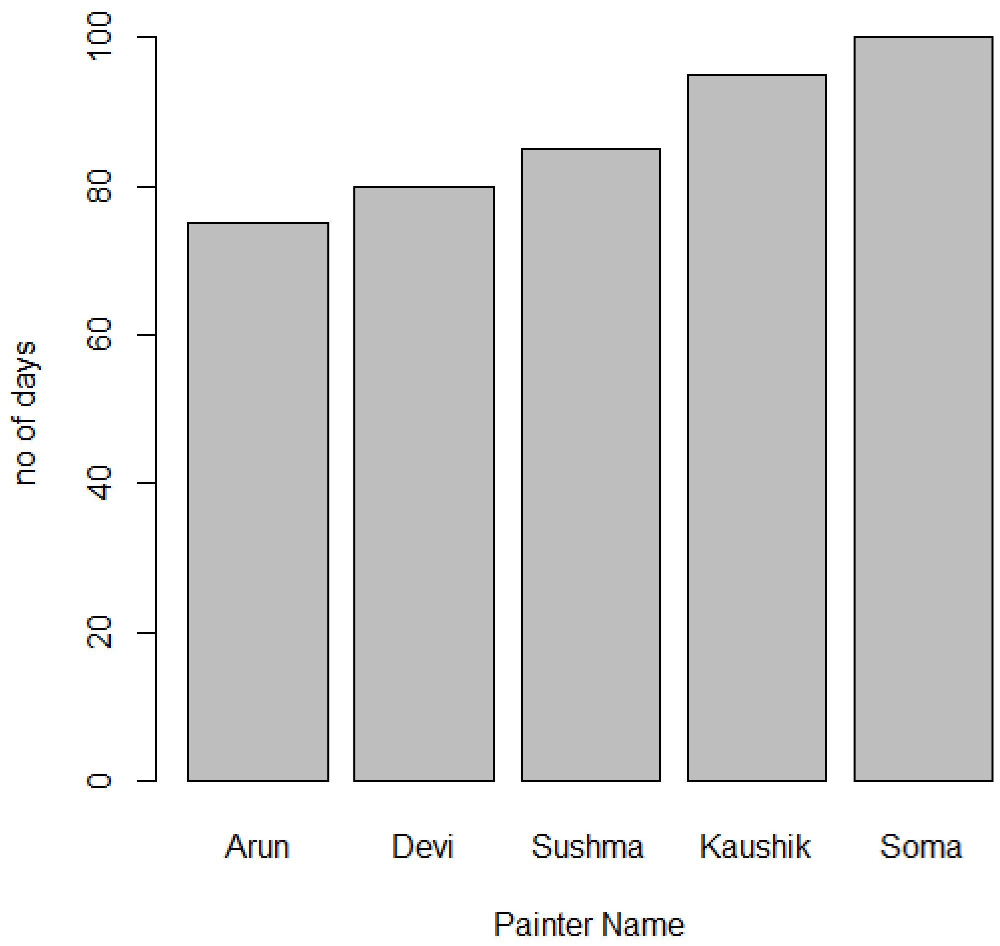
```
> hist(x, xlab="milage (miles/gallon)", ylab="frequency", main="Average gas milage", col="light green")

>
```



Q4)

CODE:

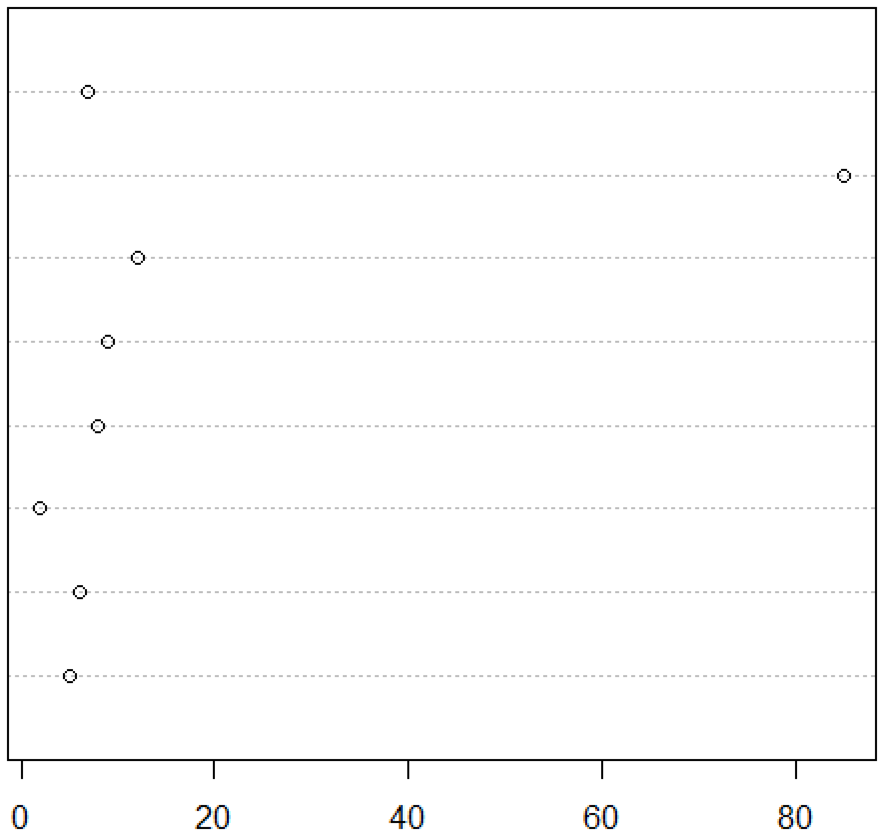
```
> x=c(75,80,85,95,100)

> y=c("Arun","Devi","Sushma","Kaushik","Soma")

> barplot(x, names.arg=y, xlab="Painter Name", ylab="no of days", col="grey")

>
```



Q5)

CODE:

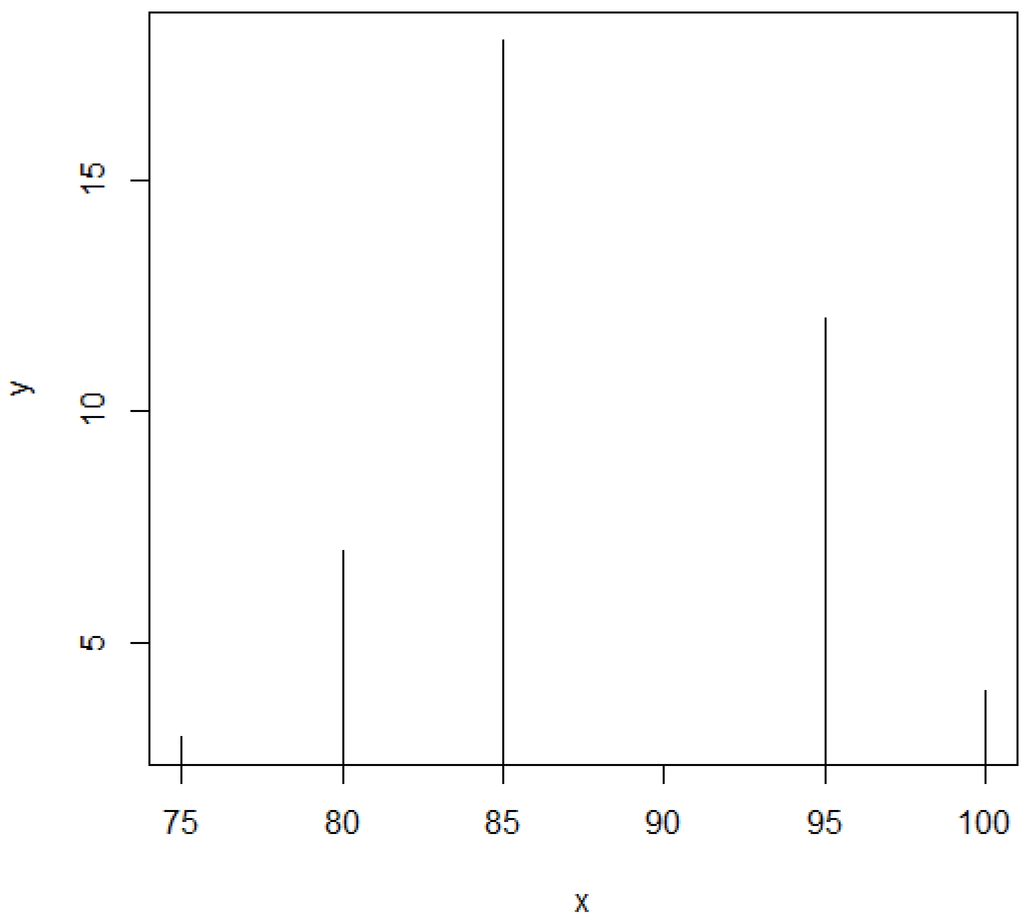
```
> x=c(5,6,2,8,9,12,85,7)
> dotchart(x)
>
```



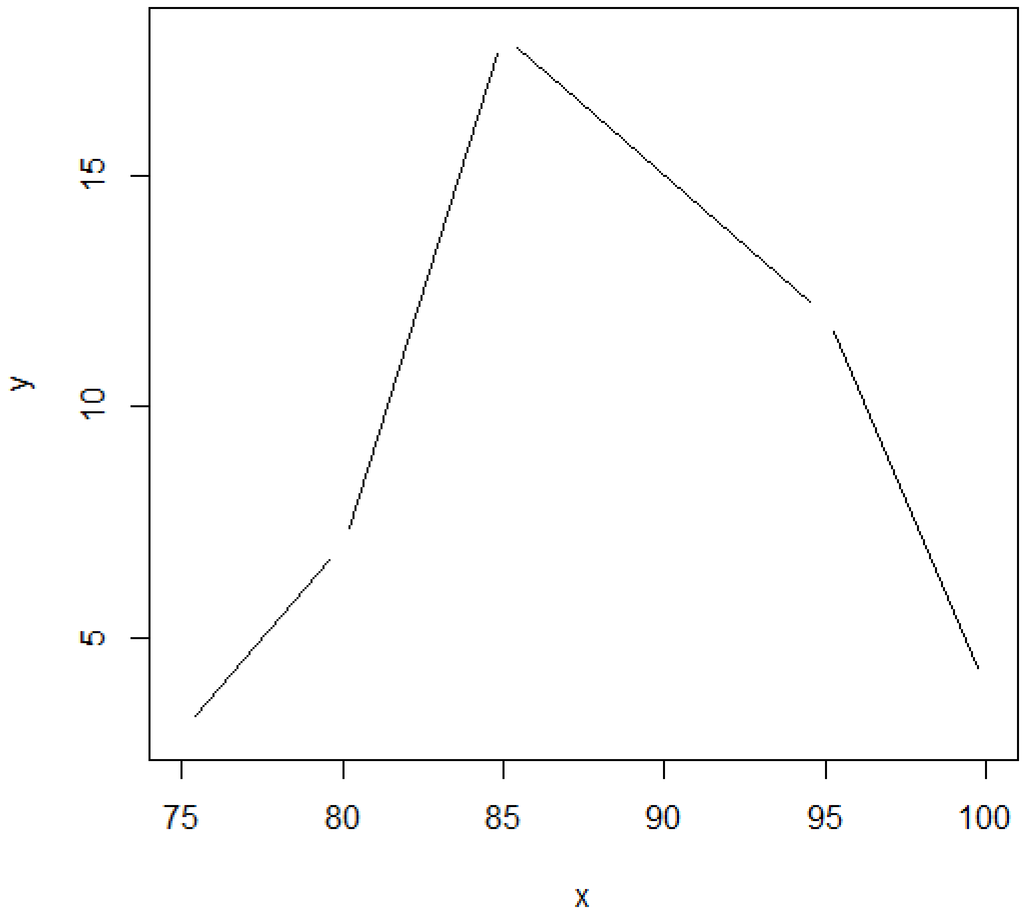
Q6)

CODE:

```
> x=c(75,80,85,95,100)

> y=c(3,7,18,12,4)

> plot(x,y,type="h")

> plot(x,y,type="c")

>
```



y
5
10
15
75
80
85
90
95
100
x

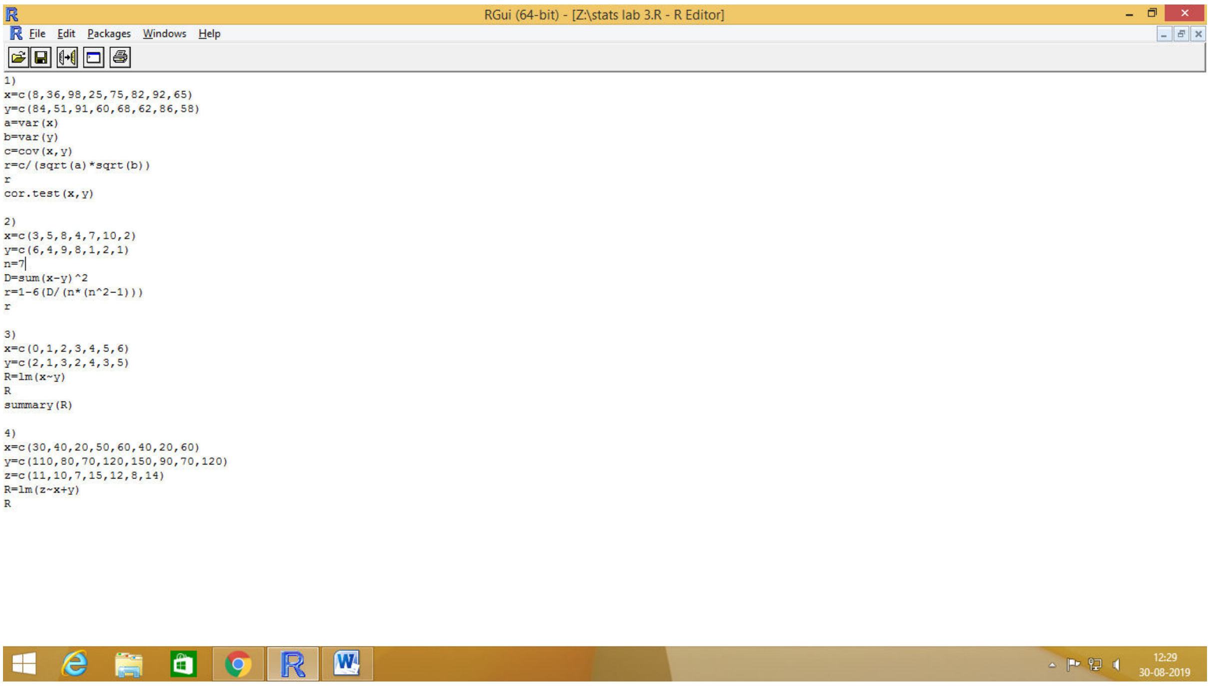


y
5
10
15
75
80
85
90
95
100
x

STATS LAB ASSIGNMENT 3

Name: Shreya Maheshwari

Reg. No. 18BCE0167



Q1. Ten students got the following percentage of marks in Mathematics and Physics.

Maths	8	36	98	25	75	82	92	65
Physics	84	51	91	60	68	62	86	58

Find the coefficient of correlation.

```
> x=c(8,36,98,25,75,82,92,65)

> y=c(84,51,91,60,68,62,86,58)

> a=var(x)

> b=var(y)

> c=cov(x,y)

> r=c/(sqrt(a)*sqrt(b))

> r

[1] 0.3229668
```

```
> cor.test(x,y)


        Pearson's product-moment correlation


data:  x and y

t = 0.8359, df = 6, p-value = 0.4352

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 -0.4941730  0.8371224

sample estimates:

      cor

0.3229668
```

Q2. The rankings of ten students in two subjects A and B are as follows

A	3	5	8	4	7	10	2
B	6	4	9	8	1	2	1

Find the rank correlation coefficient.

```
> x=c(3,5,8,4,7,10,2)

> y=c(6,4,9,8,1,2,1)

> n=7

> D=sum(x-y)^2

> r=1-6(D/(n*(n^2-1)))

Error: attempt to apply non-function

> r

[1] 0.3229668

>
```

Q3) Fit a straight line of Y on X from the following data.

X	0	1	2	3	4	5	6
Y	2	1	3	2	4	3	5

```
> x=c(0,1,2,3,4,5,6)
> y=c(2,1,3,2,4,3,5)
> R=lm(x~y)
> R

Call:
lm(formula = x ~ y)

Coefficients:
(Intercept)            y
    -0.6842       1.2895

> summary(R)

Call:
lm(formula = x ~ y)

Residuals:
      1       2       3       4       5       6       7
-1.8947  0.3947 -1.1842  1.1053 -0.4737  1.8158  0.2368

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.6842     1.3342  -0.513   0.6299
y             1.2895     0.4281   3.012   0.0297 *
---
```

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.41 on 5 degrees of freedom

Multiple R-squared: 0.6447,   Adjusted R-squared: 0.5737

F-statistic: 9.074 on 1 and 5 DF, p-value: 0.02968

Q4) Find the regression equation for the following data

Y	110	80	70	120	150	90	70	120
X	30	40	20	50	60	40	20	60
Z	11	10	7	15	19	12	8	14

```
> x=c(30,40,20,50,60,40,20,60)

> y=c(110,80,70,120,150,90,70,120)

> z=c(11,10,7,15,12,8,14)

> R=lm(z~x+y)

Error in model.frame.default(formula = z ~ x + y, drop.unused.levels = TRUE) :

  variable lengths differ (found for 'x')

> R


Call:

lm(formula = x ~ y)


Coefficients:

(Intercept)            y

    -0.6842       1.2895
```

Q5) Six independent space missions to the moon are planned. The estimated probability of success on each mission is 0.95.

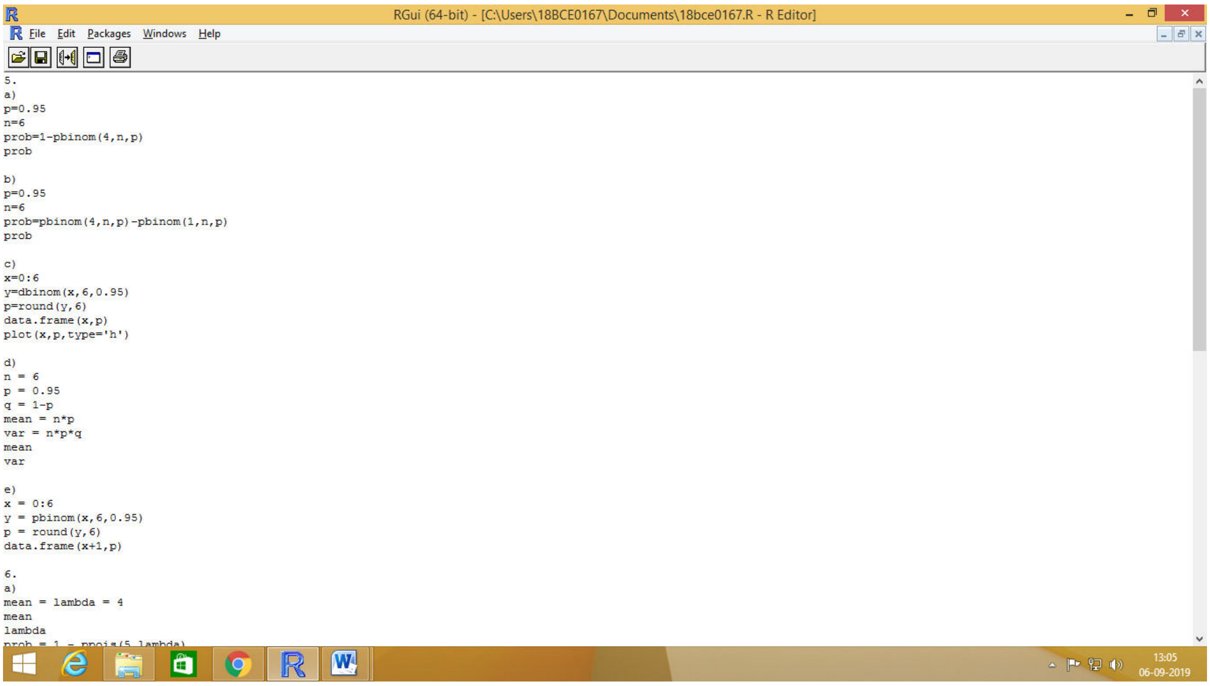
a. What is the probability that at least five of the planned missions will be successful?

b. What is the probability that from two to four missions will be successful?

c. From the table and plot the figure with h type for the distribution

d. Find mean and variance for the above distribution

e. Find cumulative distributive function



a)

```
> p=0.95

> n=6

> prob=1-pbinom(4,n,p)

> prob

[1] 0.9672262

>
```

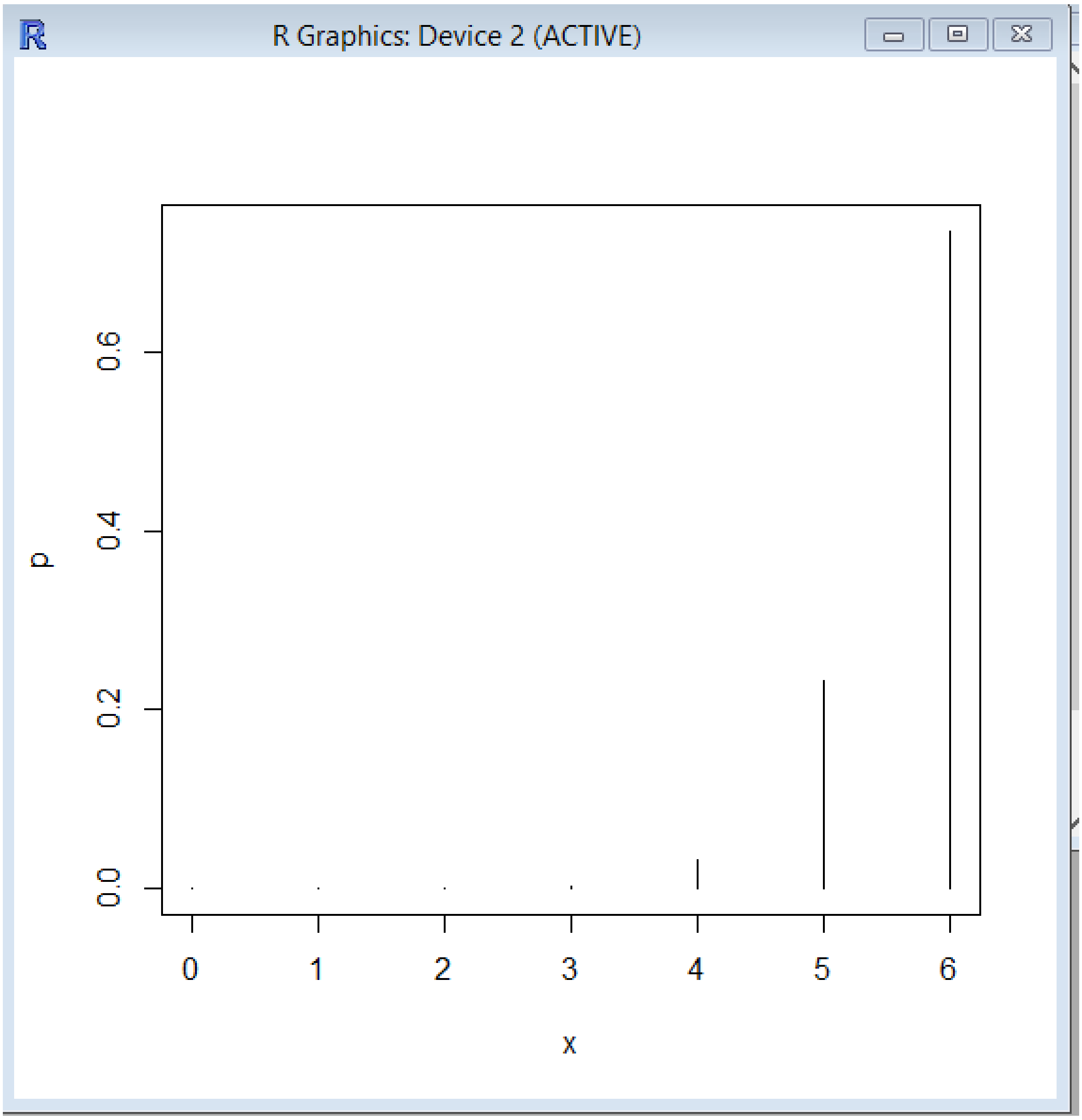
b)

```
> p=0.95

> n=6

> prob=pbinom(4,n,p)-pbinom(1,n,p)

> prob
```

```
[1] 0.03277203
```

c)

```
> x=0:6
> y=dbinom(x,6,0.95)
> p=round(y,6)
> data.frame(x,p)
  x        p
1 0 0.000000
2 1 0.000002
3 2 0.000085
4 3 0.002143
5 4 0.030544
6 5 0.232134
7 6 0.735092
> plot(x,p,type='h')
>
```



d)

> n = 6

> p = 0.95

> q = 1-p

> mean = n*p

> var = n*p*q

> mean

[1] 5.7

> var

```
[1] 0.285
>
```

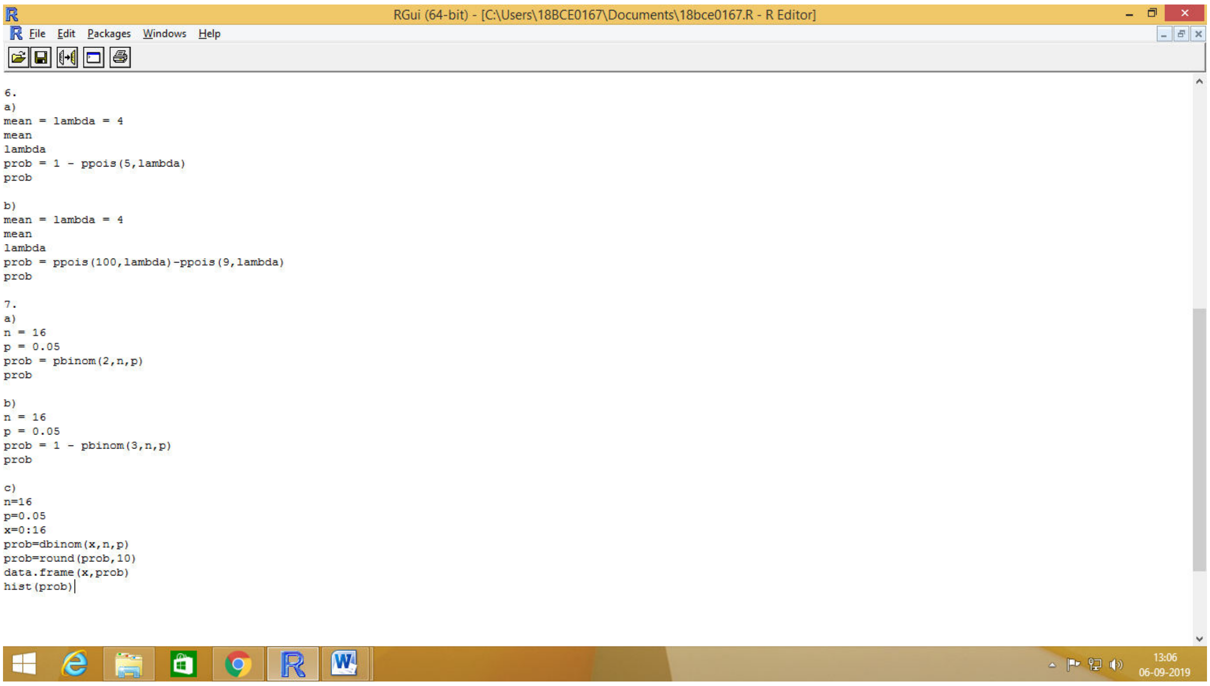
e)

```
> x = 0:6
> y = pbinom(x,6,0.95)
> p = round(y,6)
> data.frame(x+1,p)
  x...1        p
1     1 0.000000
2     2 0.000002
3     3 0.000086
4     4 0.002230
5     5 0.032774
6     6 0.264908
7     7 1.000000
>
```

Q6) . The number of red blood cells per square unit visible under a microscope follows a Poisson distribution with a mean of 4.

a. Find the probability that more than five such blood cells are visible to the observer

b. Find the probability that lies between 10 to 100.



a)

```
> mean = lambda = 4
> mean
[1] 4
> lambda
[1] 4
> prob = 1 - ppois(5,lambda)
> prob
[1] 0.2148696
>
```

b)

```
> mean = lambda = 4
> mean
[1] 4
> lambda
[1] 4
> prob = ppois(100,lambda)-ppois(9,lambda)
```

```
> prob

[1] 0.008132243

>
```

Q7) If the probability is 0.05 that a certain wide flange column will fail under a given axial load, what are the probabilities that among 16 such columns

a. Atmost two will fail

b. Atleast four will fill

c. Plot histogram for the above distribution and determine the skew ness property

a)

```
> n = 16

> p = 0.05

> prob = pbinom(2,n,p)

> prob

[1] 0.9570621
```

b)

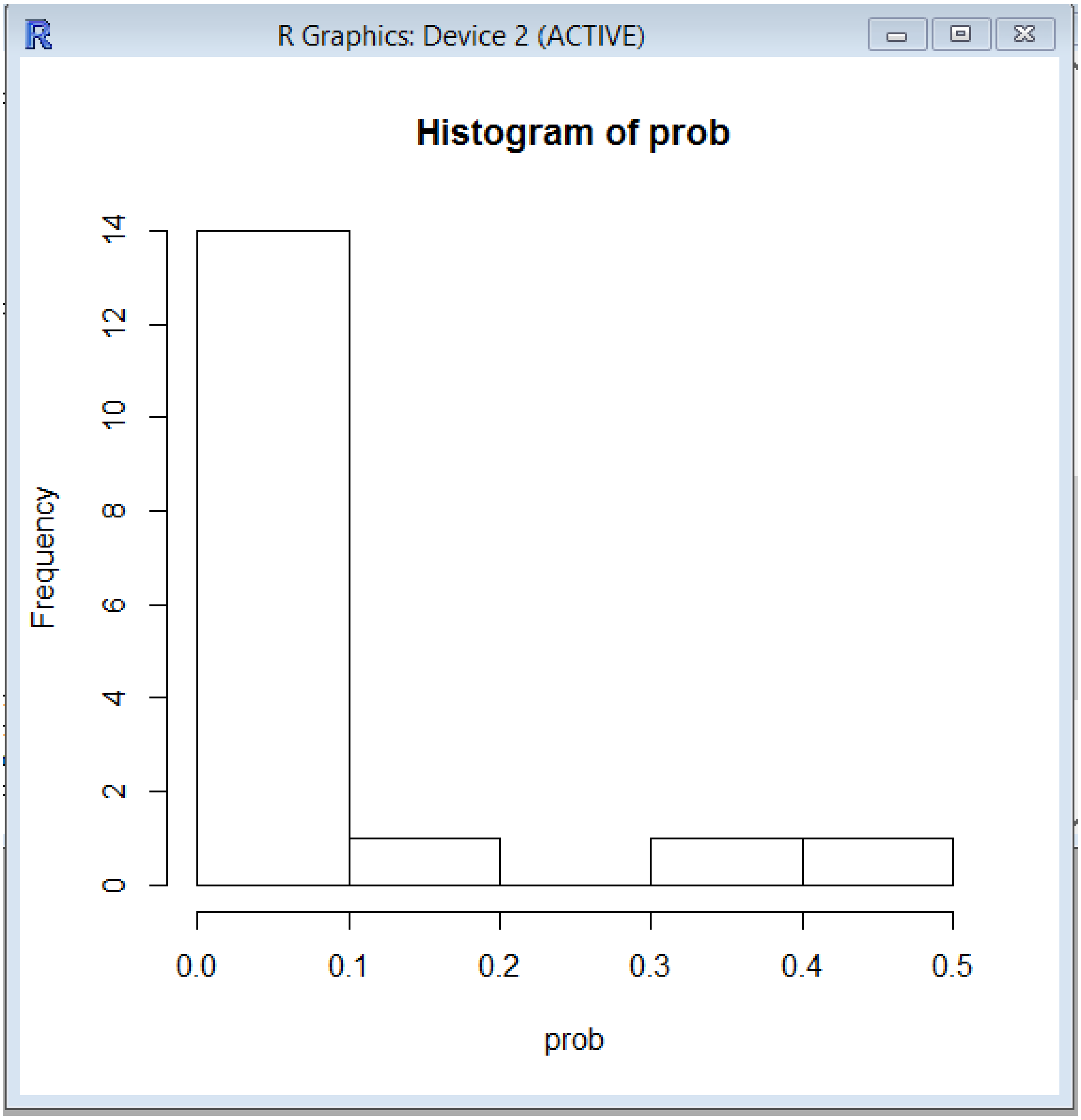
```
> n = 16

> p = 0.05

> prob = 1 - pbinom(3,n,p)

> prob

[1] 0.007003908

>
```

c)

```
> n=16

> p=0.05

> x=0:16

> prob=dbinom(x,n,p)

> prob=round(prob,10)

> data.frame(x,prob)
```

```
    x         prob
1   0 0.4401266687
2   1 0.3706329841
3   2 0.1463024937
4   3 0.0359339458
5   4 0.0061465960
6   5 0.0007764121
7   6 0.0000749170
8   7 0.0000056329
9   8 0.0000003335
10  9 0.0000000156
11 10 0.0000000006
12 11 0.0000000000
13 12 0.0000000000
14 13 0.0000000000
15 14 0.0000000000
16 15 0.0000000000
17 16 0.0000000000
> hist(prob)
>
```

GRAPH:


R Graphics: Device 2 (ACTIVE)
Histogram of prob
Frequency
0
2
4
6
8
10
12
14
0.0
0.1
0.2
0.3
0.4
0.5
prob

STATS LAB ASSIGNMENT 4

NAME: SHREYA MAHESHWARI

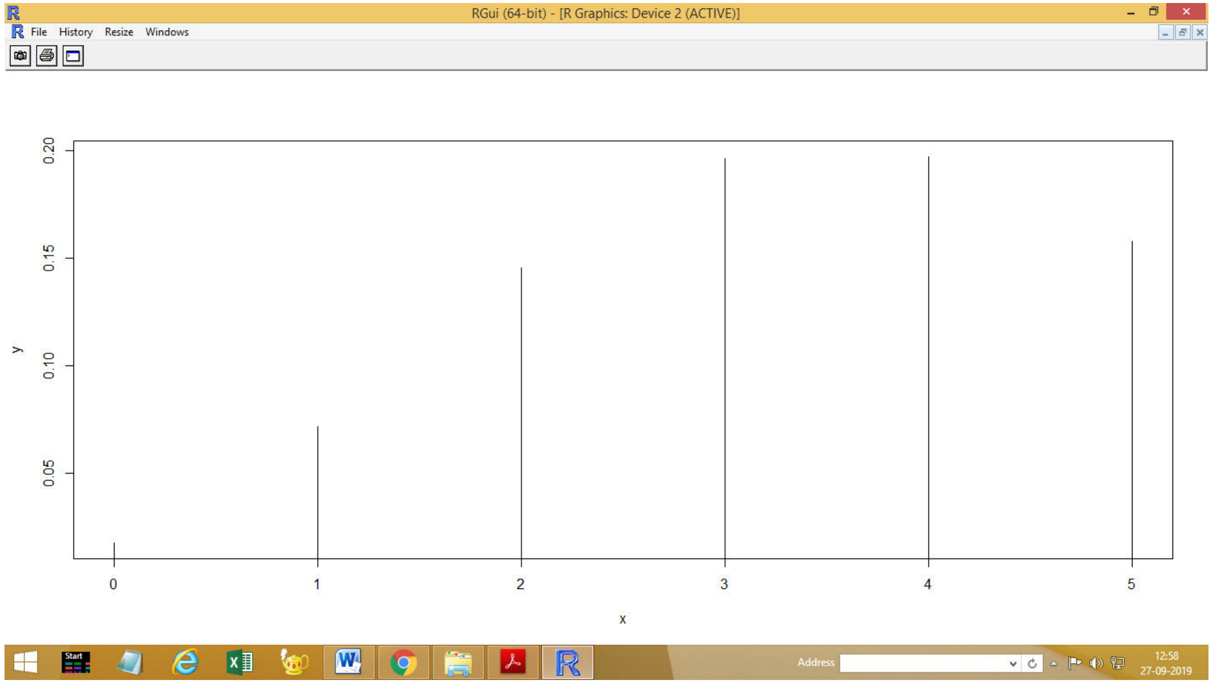
REGISTRATION NUMBER: 18BCE0167

1) Code:

```
x=0:5
y=dbinom(x,200,0.02)
plot(x,y,type='h')
pbinom(5,200,0.02)
```

Output:

```
> x=0:5
> y=dbinom(x,200,0.02)
> plot(x,y,type='h')
> pbinom(5,200,0.02)
[1] 0.7867225
```



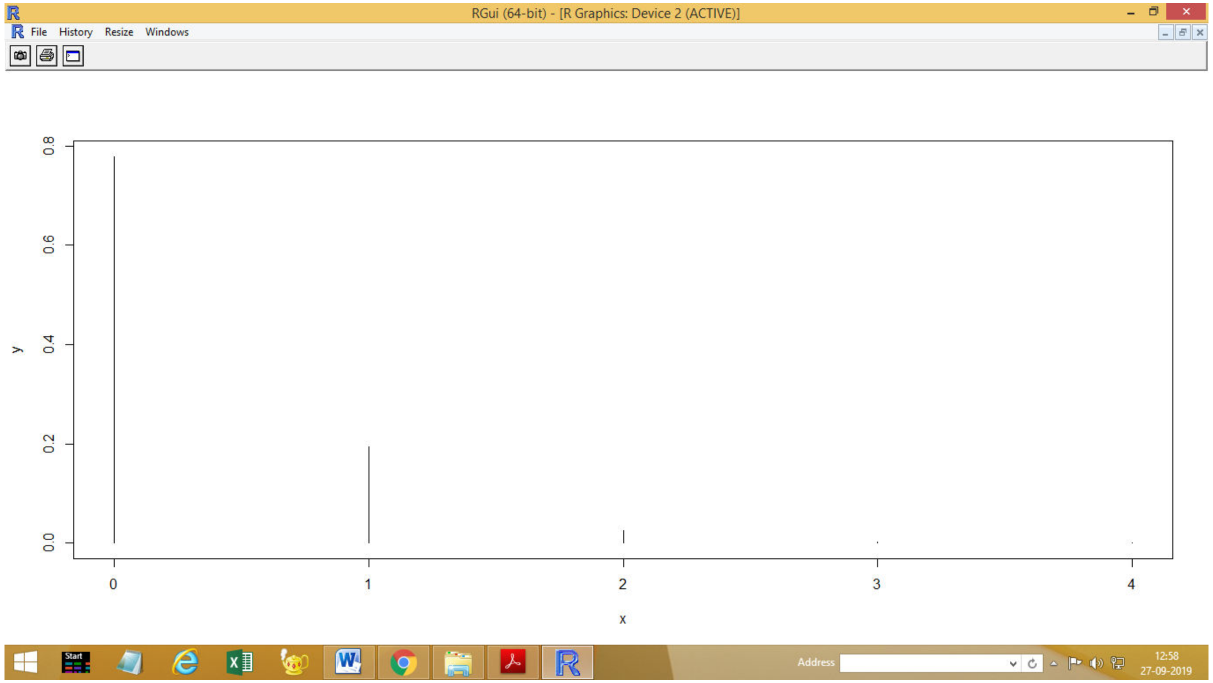
2) Code:

```
n=100
x=0:4
s=c(79,18,2,1,0)
lambda=sum(x*s)/n
y=dpois(x,lambda)
plot(x,y,type='h')
```

Output:

```
> n=100
> x=0:4
> s=c(79,18,2,1,0)
```

```
> lambda=sum(x*s)/n
> y=dpois(x,lambda)
> plot(x,y,type='h')
```

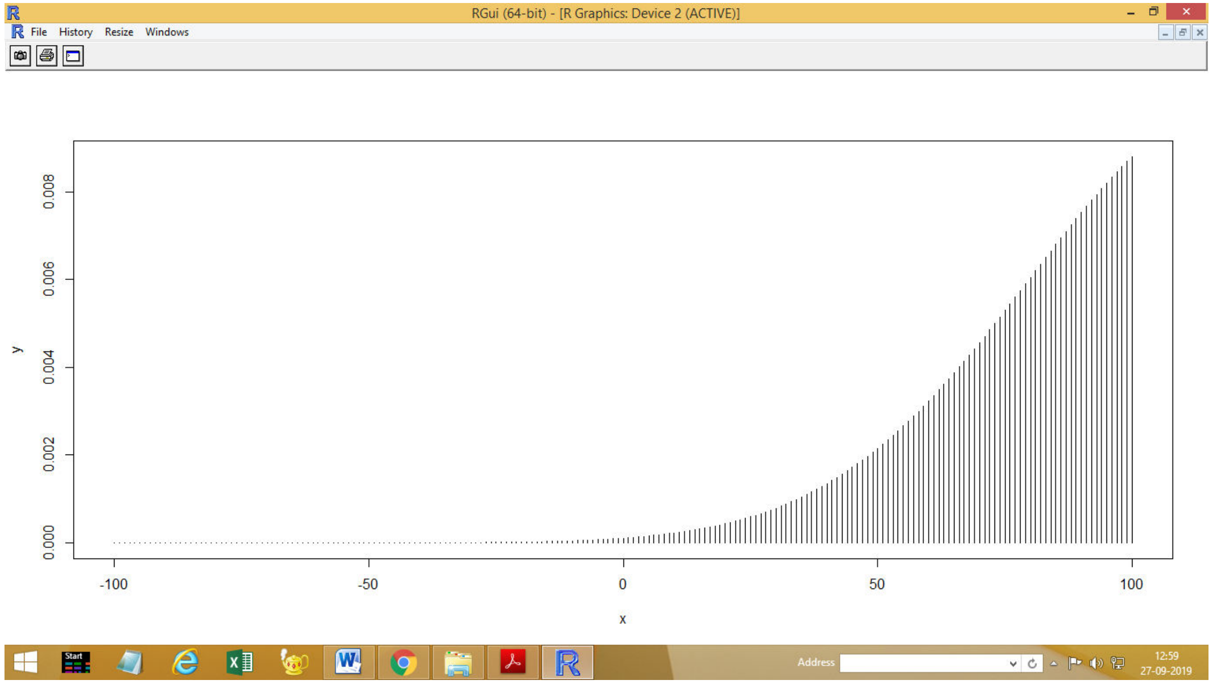


3) Code:

```
1-pnorm(150,120,40)
pnorm(150,120,40)-pnorm(100,120,40)
pnorm(90,120,40)-pnorm(60,120,40)
x=seq(-100,100,by=1)
y=dnorm(x,120,40)
plot(x,y,type='h')
```

Output:

```
> 1-pnorm(150,120,40)
[1] 0.2266274
> pnorm(150,120,40)-pnorm(100,120,40)
[1] 0.4648351
> pnorm(90,120,40)-pnorm(60,120,40)
[1] 0.1598202
> x=seq(-100,100,by=1)
> y=dnorm(x,120,40)
> plot(x,y,type='h')
```



4)

Code:

```
dbinom(4,6,0.5)
dbinom(5,6,0.5)
1-pbinom(3,6,0.5)
```

Output:

```
> dbinom(4,6,0.5)
[1] 0.234375
> dbinom(5,6,0.5)
[1] 0.09375
> 1-pbinom(3,6,0.5)
[1] 0.34375
```

5) Code:

```
n=100
mu0=160
xbar=165
sigma=10
z=(xbar-mu0)/(sigma/sqrt(n))
z
alpha=0.01
z.alpha=qnorm(1-alpha)
z.alpha
```

Output:

```
> n=100
> mu0=160
> xbar=165
> sigma=10
> z=(xbar-mu0)/(sigma/sqrt(n))
> z
[1] 5
```

```
> alpha=0.01
> z.alpha=qnorm(1-alpha)
> z.alpha
[1] 2.326348
```

6) Code:

```
n1 = 32
x1bar = 72
sigma1 = 8
n2 = 36
x2bar = 70
sigma2 = 6
Z = (x1bar - x2bar)/((sigma1/sqrt(n1)) + (sigma2/sqrt(n2)))
Z
alpha = 0.01
z.alpha = qnorm(1 - alpha)
z.alpha
```

Output:

```
> n1 = 32
> x1bar = 72
> sigma1 = 8
> n2 = 36
> x2bar = 70
> sigma2 = 6
> Z = (x1bar - x2bar)/((sigma1/sqrt(n1)) + (sigma2/sqrt(n2)))
> Z
[1] 0.8284271
> alpha = 0.01
> z.alpha = qnorm(1 - alpha)
> z.alpha
```

```
[1] 2.326348
```

7) Code:

```
n1 = 1200
n2 = 900

P1 = 0.3
P2 = 0.25
Q1 = 1 - P1
Q2 = 1 - P2
Z = (P1 - P2)/sqrt(P1*Q1/n1 + P2*Q2/n2)
Z
alpha = 0.05
z.alpha = qnorm(1 - alpha)
z.alpha
```

Output:

```
> n1 = 1200
> n2 = 900
>
> P1 = 0.3
> P2 = 0.25
> Q1 = 1 - P1
> Q2 = 1 - P2
> Z = (P1 - P2)/sqrt(P1*Q1/n1 + P2*Q2/n2)
> Z
[1] 2.55377
> alpha = 0.05
> z.alpha = qnorm(1 - alpha)
> z.alpha
[1] 1.644854
```

8) Code:

```
n1 = 400
n2 = 600
p1 = 200/400
p2 = 325/600
P =(n1*p1 + n2*p2)/(n1+n2)
Q = 1 - P
Z = (p1 - p2)/sqrt(P*Q/(n1+n2))
Z
alpha = 0.05
z.alpha = qnorm(1 - alpha)
z.alpha
```

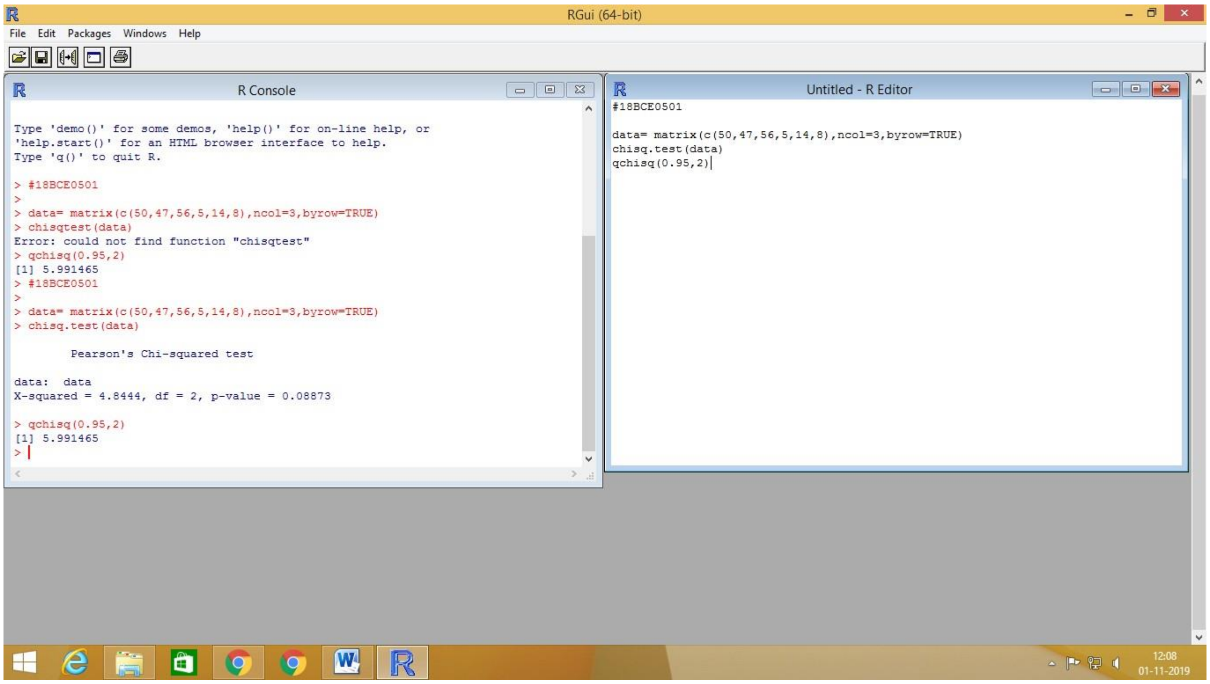
Output:

```
> n1 = 400
> n2 = 600
> p1 = 200/400
> p2 = 325/600
> P =(n1*p1 + n2*p2)/(n1+n2)
> Q = 1 - P
> Z = (p1 - p2)/sqrt(P*Q/(n1+n2))
> Z
[1] -2.638532
> alpha = 0.05
> z.alpha = qnorm(1 - alpha)
> z.alpha
[1] 1.644854
```

STATS LAB ASSIGNMENT 5

NAME: SHREYA MAHESHWARI

REGISTRATION NUMBER: 18BCE0167



1. CODE:

```
data=
matrix(c(50,47,56,5,14,8),ncol=3,byrow=TRUE)
chisq.test(data)
qchisq(0.95,2)
```

OUTPUT:

```
> chisq.test(data)

        Pearson's Chi-squared test

data:  data
X-squared = 4.8444, df = 2, p-value = 0.08873

>
qchisq(0.95,2)
[1] 5.991465
```

SINCE X-SQUARED= 4.84 IS LESS THAN THE GIVEN VALUE OF X-SQUARED FOR 2 DOF, THEREFORE WE ACCEPT THE HYPOTHESIS AND CONCLUDE THAT THE PROPORTIONS ARE STATISTICALLY SAME FOR ALL INSPECTIONS.

2. CODE

```
x=c(5,4,3,2,1,0)
n=5
p=0.5
o=c(14,56,110 ,88 ,40 ,12 )
e=dbinom(x,n,p)*320
xs=sum((o-e)^2/e)
xs
qchisq(0.95,5)
```

Output:

```
> xs
[1] 7.16
>
qchisq(0.95,5)
[1] 11.0705
```

CALCULATED VALUE OF X-SQUARE IS LESS THAN THE TABULATED VALUE, IT IS NOT SIGNIFICANT AT 5% LOS AND HENCE THE NULL HYPOTHESES IS ACCEPTED.

3. CODE:

```
x= c(14,16,8,20,11,9,14)
ef= sum(x)/7
ef
#regrouping as frequencies are less than 10
x=c(14,16,28,11,23)
ef=c(12,12,24,12,24)
cs=sum((x-ef)^2/ef)
cs
qchisq(0.95,4)
```

Output:

```
> cs
[1] 2.458333
>
qchisq(0.95,4)
[1] 9.487729
```

SINCE THE CALCULATED VALUE OF X-SQUARE IS LESS THAN THE TABLE VALUE, THE HYPOTHESIS HOLDS GOOD

4. CODE:

```
x=0:6
f=c(36 ,40 ,19 ,2 ,0 ,2 ,1)
lambda=sum(f*x)/sum(f)
lambda
ef=dpois(x,lambda)*sum(f
) f1=round(ef)
#combining last 5 frequencies
of=c(36,40,24)
ef=c(37,37,26)
cs=sum((of-ef)^2/ef)
cs
qchisq(0.95,2)
```

Output:

```
> cs
[1] 0.4241164
>
qchisq(0.95,2)
[1] 5.991465
```

SINCE THE CALCULATED VALUE OF X-SQUARE IS LESS THAN 5.99, THE HYPOTHESIS HOLDS GOOD, HENCE H0 IS ACCEPTED.