Name: Viren Chhabria
Strand: Data Science
Student ID: 18301780
Course: CS7DS1 (Data Analytics)

Question 1:

Taking 1 million random samples between 1 to 1 million, what is the probability that the numbers are prime to each other?

Answer 1:

Theoretical answer is $6/\pi^2 = 0.60792710185$

Code in R:

```
> start_time <- Sys.time()
> totalCount=1000000
> totalCountDecrementer=1000000
> primeCount=0
> x=sample(1000000,1000000)
> y=sample(1000000,1000000)
>
> gcd <- function(x, y) {
+   while(y) {
+     temp = y
+     y = x %% y
+     x = temp
+   }
+   return(x)
+ }
>
> for(i in 1:totalCount)  {
+   #smaller=min(x,y)
+   #prime=TRUE
+   #for(i in 2:sqrt(smaller)){
+   #  if(x%%i==0 & y%%i==0) {
+   #    prime=FALSE
+   #    break
+   #  }
+   #}
+
+   if(gcd(x[i],y[i]) == 1)
+   {
+     primeCount=primeCount+1
+   }
+ }
> print(paste("Probability that 2 numbers are prime to each other is: ",pr
imeCount/totalCount))
[1] "Probability that 2 numbers are prime to each other is:  0.607472"
> end_time <- Sys.time()
> print(start_time)
[1] "2018-09-22 00:07:30 BST"
> print(end_time)
[1] "2018-09-22 00:07:38 BST"
> print(end_time - start_time)
Time difference of 8.000176 secs
```

Name: Viren Chhabria **Student ID:** 18301780
Strand: Data Science **Course:** CS7DS1 (Data Analytics)

Question 2:

x – number of males arriving in the bank at rate

y – number of females arriving at the bank at rate

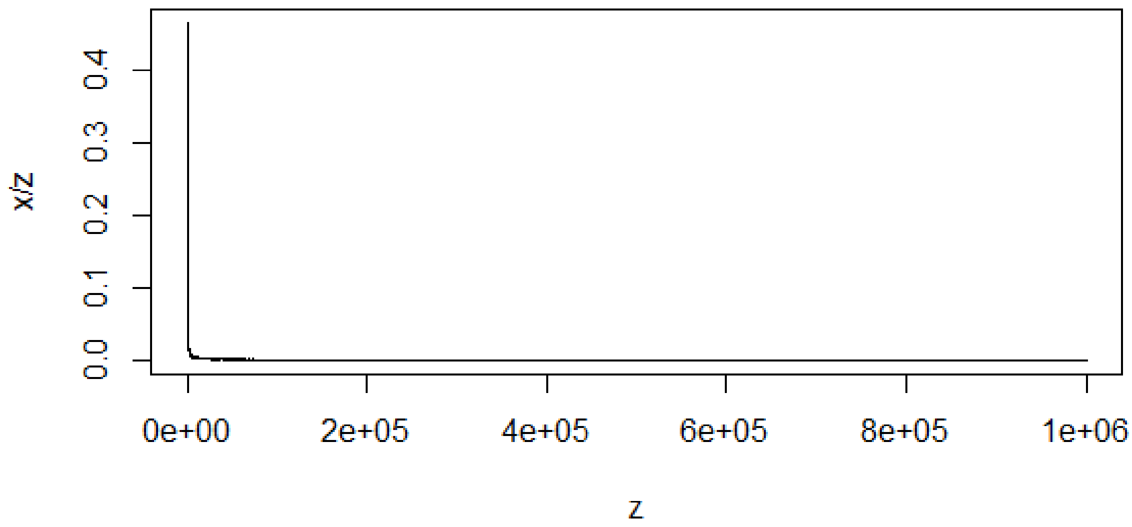
z – total number of people at the bank represented by x + y

E[X | Z=z]

Plot x/z against z

Answer 2:

```
> start_time <- Sys.time()
> avgRateMale=30
> avgRateFemale=40
> totalPersonRequired=1000000
> maleObservationsPerHour=c()
> femaleObservationsPerHour=c()
> totalObservationsPerHour=c()
> totalObservationsPerHourSum=c()
>
> while(totalPersonRequired > 0){
+   maleObservations=rpois(5,lambda = avgRateMale)
+   femaleObservations=rpois(5,lambda = avgRateFemale)
+   totalPeople=sum(maleObservations)+sum(femaleObservations)
+   maleObservationsPerHour=c(maleObservationsPerHour,maleObservations)
+   femaleObservationsPerHour=c(femaleObservationsPerHour,femaleObservatio
ns)
+   totalObservationsPerHour=maleObservationsPerHour+femaleObservationsPer
Hour
+   totalPersonRequired=totalPersonRequired-totalPeople
+ }
> totalWalkIn=sum(totalObservationsPerHour)
> for(i in 1:length(totalObservationsPerHour)){
+   totalObservationsPerHourSum[i]=sum(totalObservationsPerHour[1:i])
+ }
> print(paste("Expected Value = ",(avgRateMale/(avgRateMale+avgRateFemale)
)*totalWalkIn))
[1] "Expected Value =  428730.857142857"
> plot(totalObservationsPerHourSum,maleObservationsPerHour/totalObservatio
nsPerHourSum, type="l")
> end_time <- Sys.time()
> print(start_time)
[1] "2018-09-22 00:24:43 BST"
> print(end_time)
[1] "2018-09-22 00:24:46 BST"
> print(end_time - start_time)
Time difference of 2.854399 secs
> plot(totalObservationsPerHourSum,maleObservationsPerHour/totalObservatio
nsPerHourSum, xlab = "z",ylab = "x/z",type="l")
```


0.0
0.1
0.2
0.3
0.4
x/z
0e+00
2e+05
4e+05
6e+05
8e+05
1e+06
z

Name: Viren Chhabria **Student ID:** 18301780
Strand: Data Science **Course:** CS7DS1 (Data Analytics)

Question 3:

$x_i \sim U(0,1)$

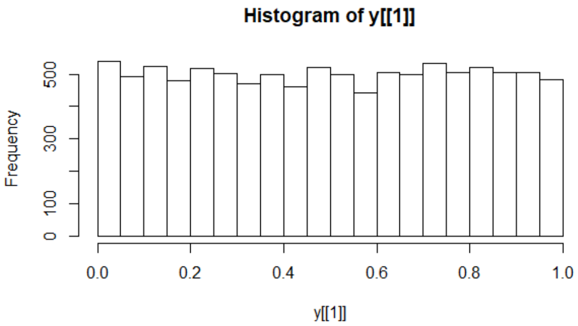
$Y_n = \xi x_i$

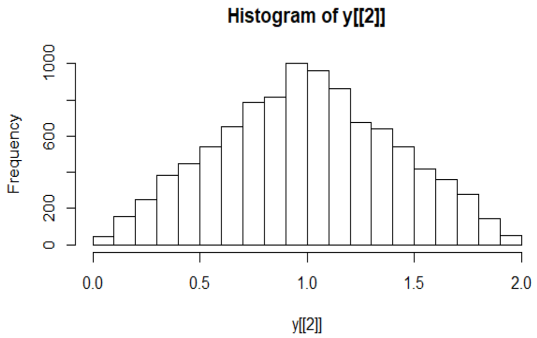
Plot histogram for $y_1, y_2, y_5, y_{10}, y_{30}, y_{100}$

Answer 3:

Code in R:

```
> max=100
> randomSamples=10000
> x=c()
> x=replicate(max,runif(randomSamples),simplify = FALSE)
> y=c()
>
> #for(i in 1:max)
> #{
> #  x[[i]]=runif(10000)
> #}
>
> for(i in 1:max)
+ {
+   sumVector=c()
+   totalVector=x[[1]]
+   if(i>1)
+   {
+     for(j in 2:i)
+     {
+       sumVector=x[[j]]
+       totalVector=totalVector+sumVector
+     }
+   }
+   y[[i]]=totalVector
+ }
>
> hist(y[[1]])
> hist(y[[2]])
> hist(y[[5]])
> hist(y[[10]])
> hist(y[[30]])
> hist(y[[100]])
```





Name: Viren Chhabria
Strand: Data Science
Student ID: 18301780
Course: CS7DS1 (Data Analytics)