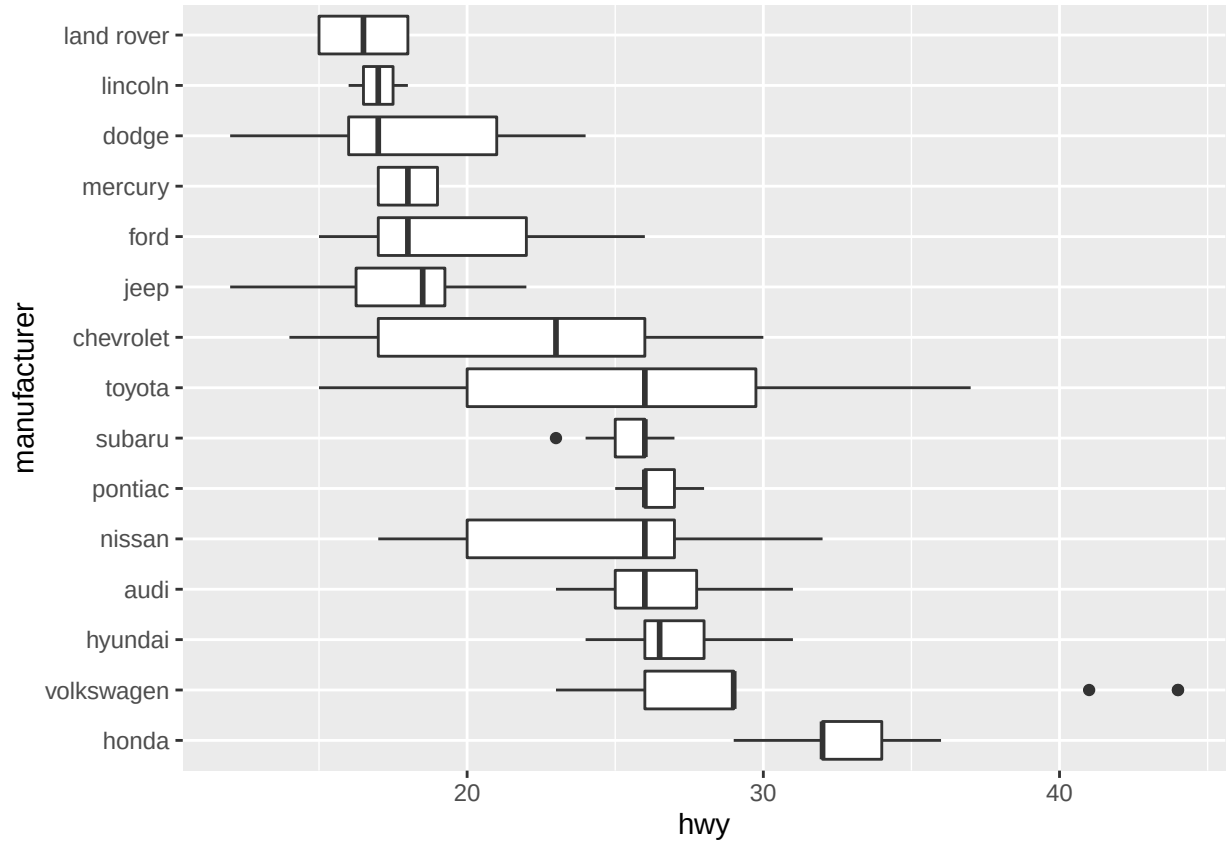
Question 1

An ideal way to described the relationship between highway mpg and car manufacturer, is to graph a box plot corresponding to changing mpg values with respect to the car manufacturer as shown in the figure below:

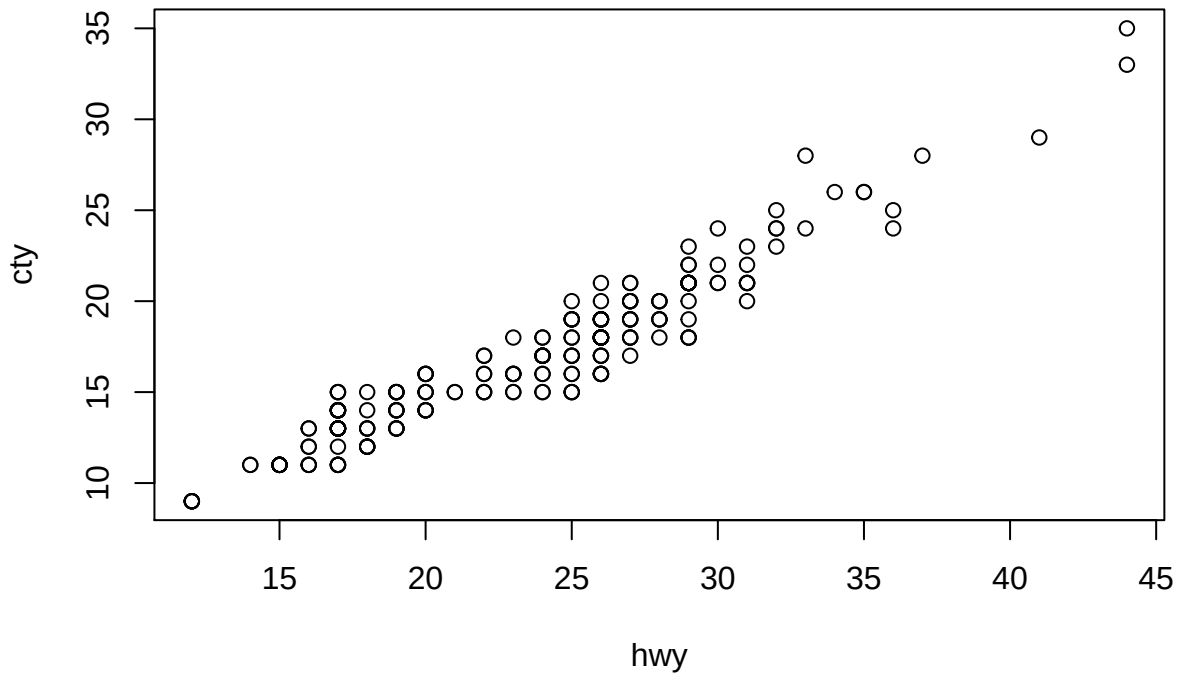
```
library(ggplot2)
ggplot(mpg ,aes(reorder(manufacturer, -hwy, median), hwy)) +
  geom_boxplot() + coord_flip() + scale_x_discrete("manufacturer")
```



From the graph above, we deduce the following fuel efficiency order among car manufactures: 1)honda , 2)volkswagen with two outliers 3) hyundai 4) both audi, nissan,pontaic and toyota are tied.Toyota have a much larger spread from this group followed by nissan. subaru and pontaic have the shortest spread in this group with subaru having one outlier.5)Chevrolet 6)jeep 7)ford and mecury are tied though ford has the larger spread. 8) dodge and lincoln are tied but lincoln has a much more larger spead. 9) lad rover. Land rover is the least fuel efficient among all car maufacturer, honda is the most fuel effiecient and toyota have the largest spread.

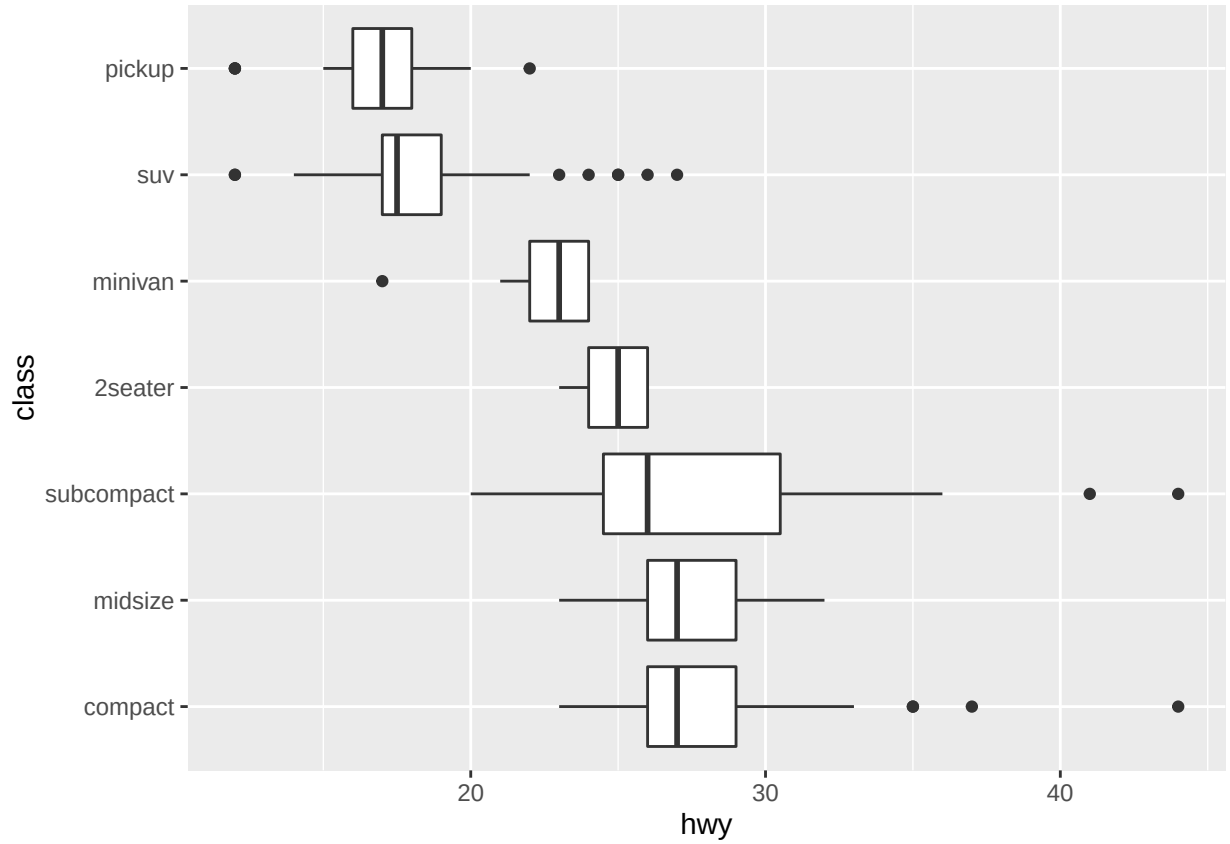
Question 2 To describe the relationship between highway mpg , city mpg and model class can be done with the aid of a box plot.In this case we have two numeric variable and factor variable model class.Lets use a scatter plot to see the relationship between the two numeric variable, as shown below:

```
plot(mpg$hwy,mpg$cty, xlab = "hwy", ylab="cty")
```



As seen in the graph above , highway mpg are generally higher than city mpg and the relationship between the two is linear.Now lets see the relationship between highway mpg and model class as whown in the plot below:

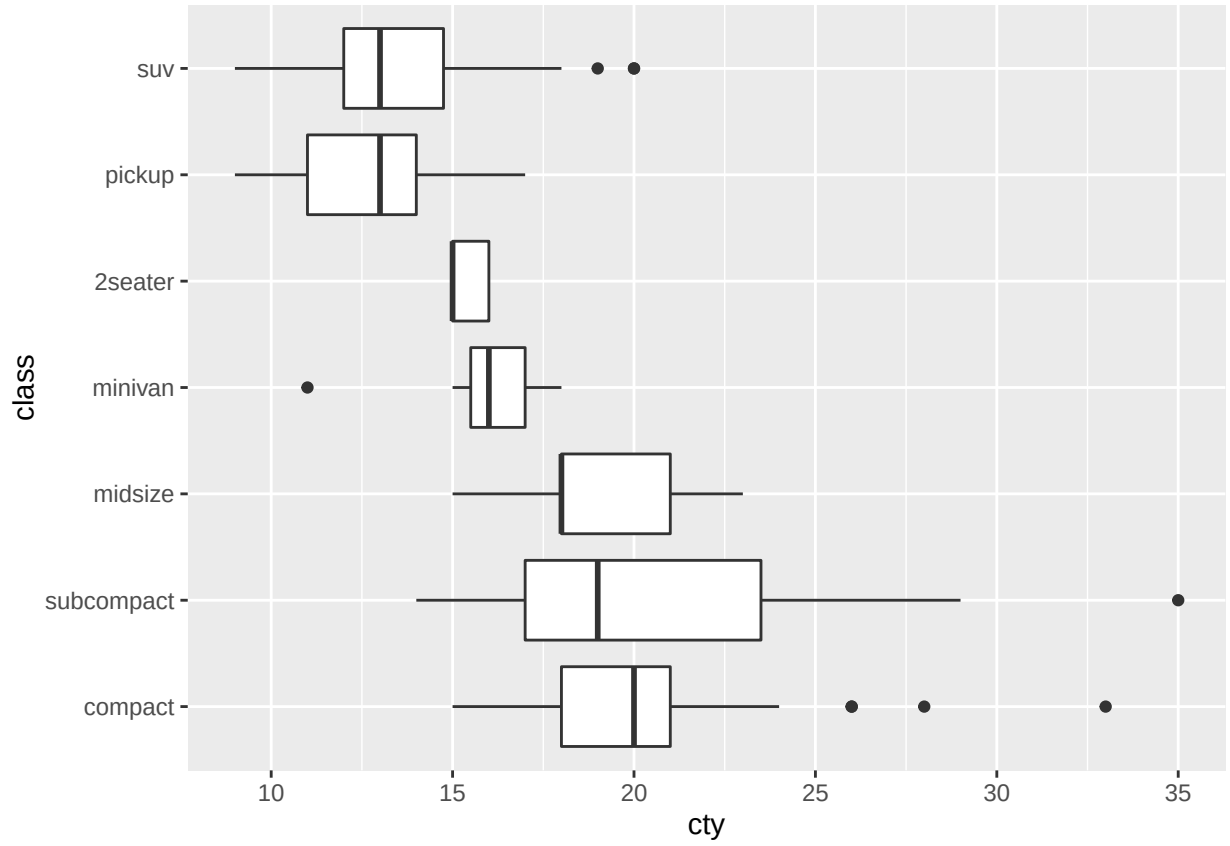
```
ggplot(mpg ,aes(reorder(class, -hwy, median), hwy)) +
  geom_boxplot() + coord_flip() + scale_x_discrete("class")
```



As seen from the plot above the order of fuel efficiency in highway with respect to model class are compact, midsize , subcompact, 2seater, minivan , suv and pickup .Subcompact has a much wider spread with two outliers. The box for both compact and midsize are identical with compact having three outliers.

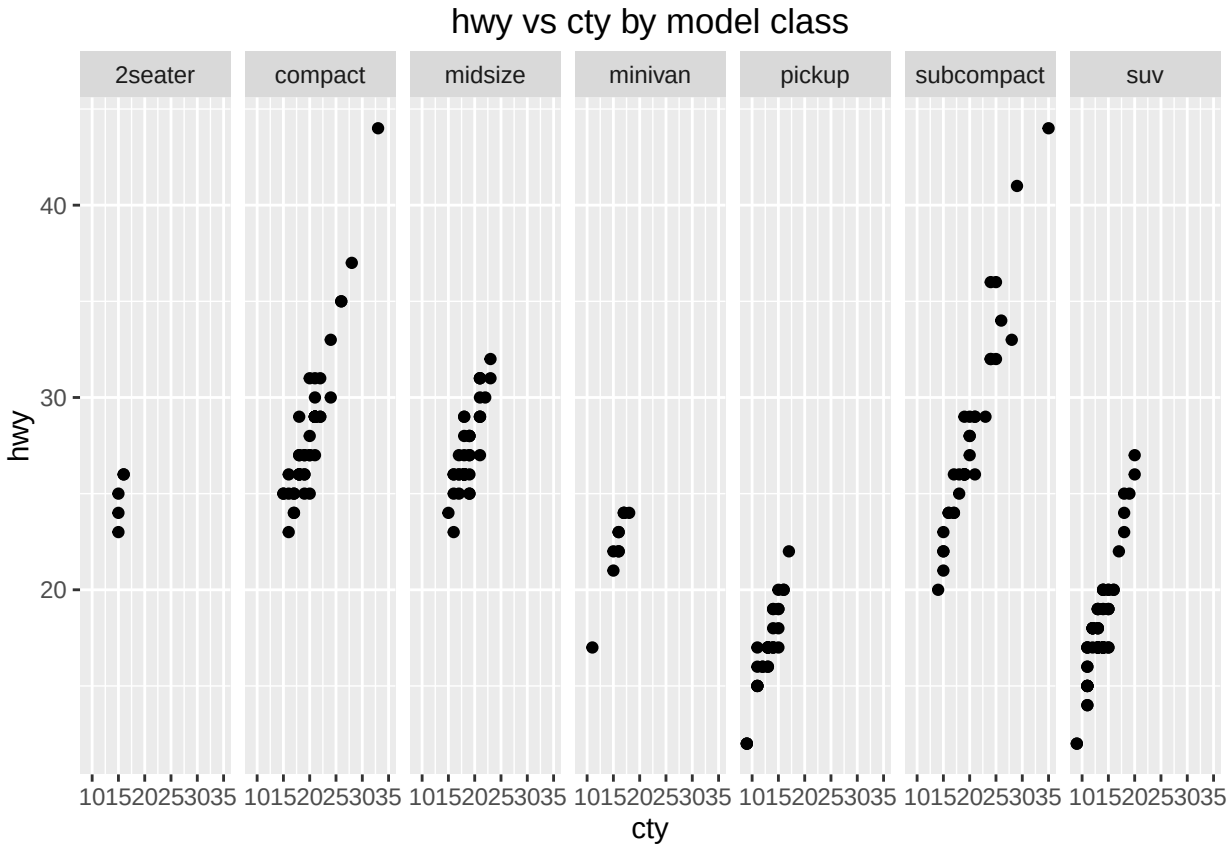
Now lets look at the relationship between city mpg and model class as shwon in the plot below:

```
ggplot(mpg ,aes(reorder(class, -cty, median), cty)) +
  geom_boxplot() + coord_flip() + scale_x_discrete("class")
```



As seen from the above plot , the fuel effieciency order are : compact , subcompact, midsize , minivan, 2seater, pickup and suv.Now lets observe the city and highway mpg side by side using the model class as the facets. This is shown in the graph below:

```
qplot(x=cty,y=hwy,facets = .~class,data=mpg, main= "hwy vs cty by model class")
```



As seen from the graph above , highway mpg are generally higher than city mpg for the same model class.

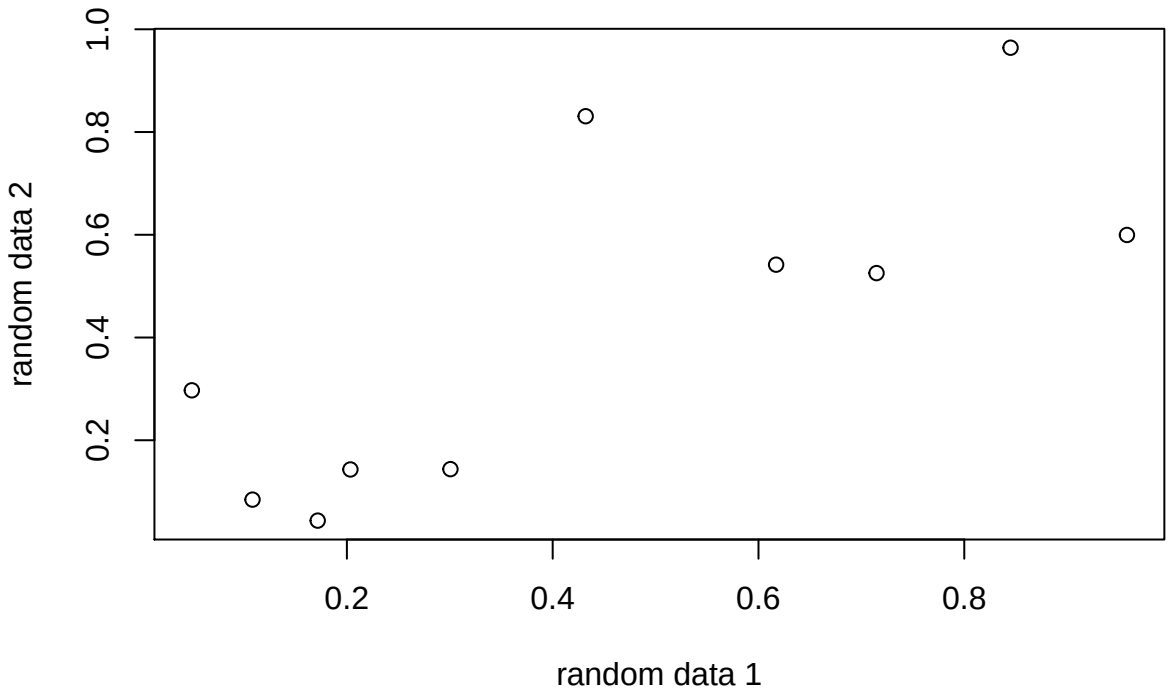
Question 3: 1. Histogram can be used to graph continous, discrete and unordered data 2.Mean and mode can easily be determine from an histogram plot 2.Better suited for small data set 3. we cannot display several histogram at the same time for the purpose of comparison

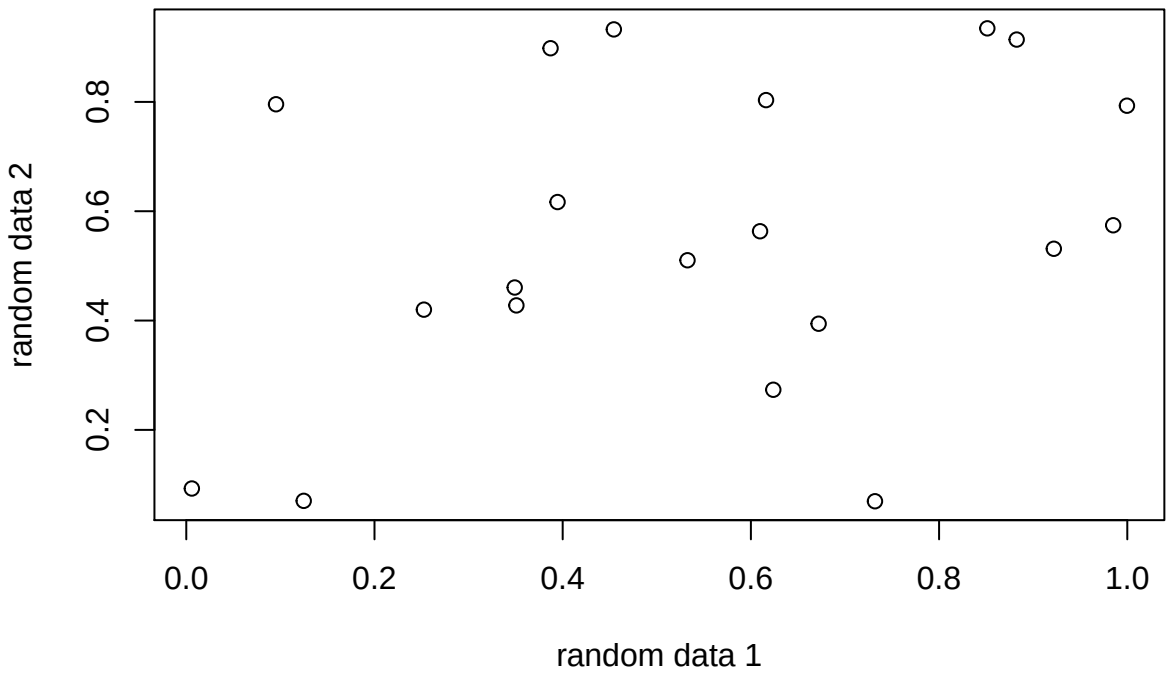
Box plot: 1.Well suited for large data set 2.Box plots provide some indication of the data's symmetry and skew-ness 3.We can see outliers from the box plot 4.We cannot determine mean and mode from box plot 5.Not suitable for small number of data point

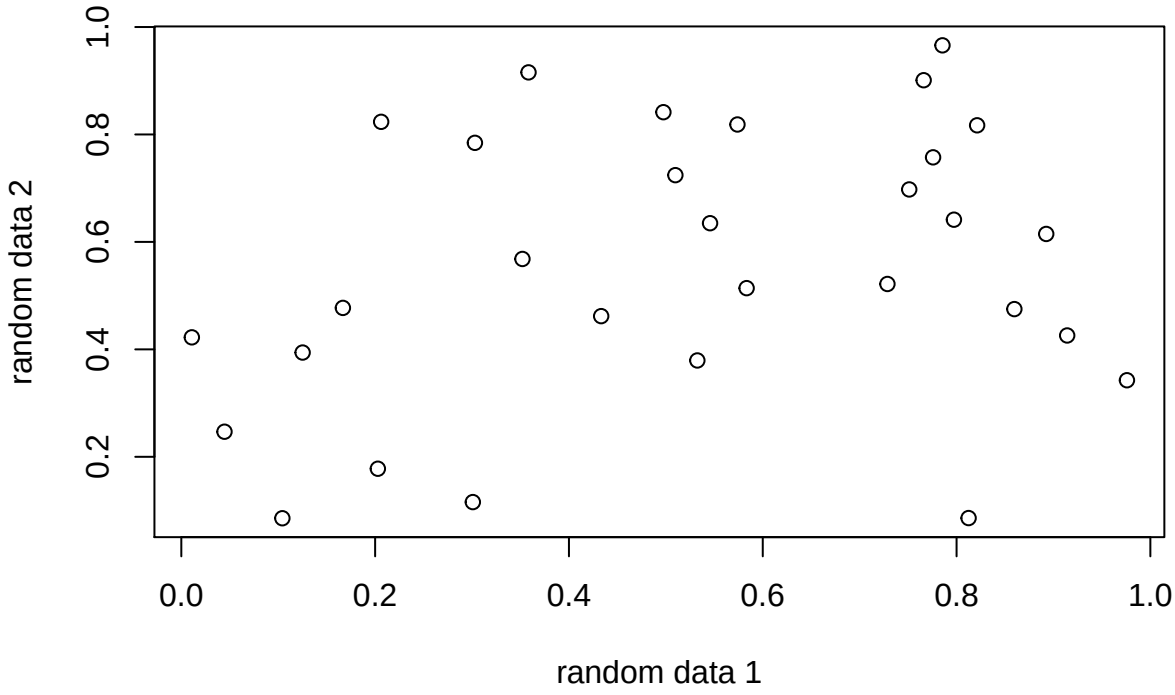
Question 4:

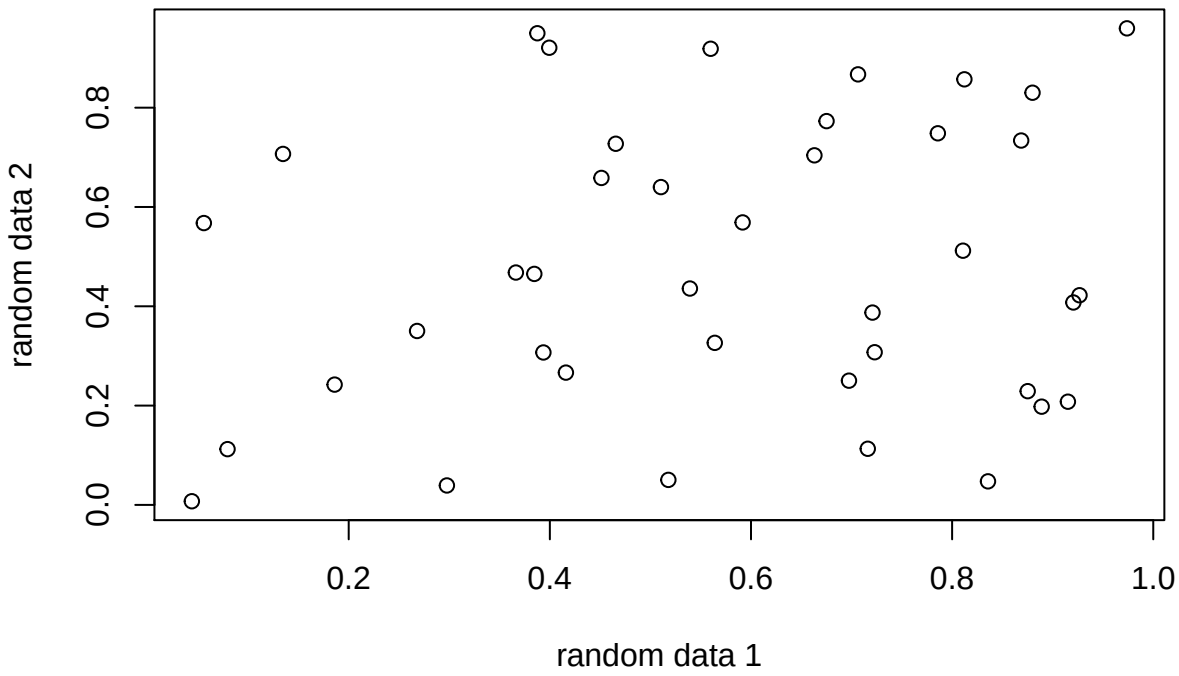
```
list_ps = vector(mode="numeric", length = 20);
list_jpg= vector(mode="numeric", length = 20);
list_pdf = vector(mode="numeric", length = 20);
list_png = vector(mode="numeric", length = 20);
for (num in seq(10, 200 , by=10)){
  index = num/10;
  data1 = runif(num);
  data2 = runif(num);

plot(data1,data2, xlab="random data 1" , ylab = "random data 2")
postscript("mydata.ps")
list_ps[index] = file.info("mydata.ps")$size
dev.off()
jpeg("mydata.jpg")
list_jpg[index] = file.info("mydata.jpg")$size
dev.off()
```

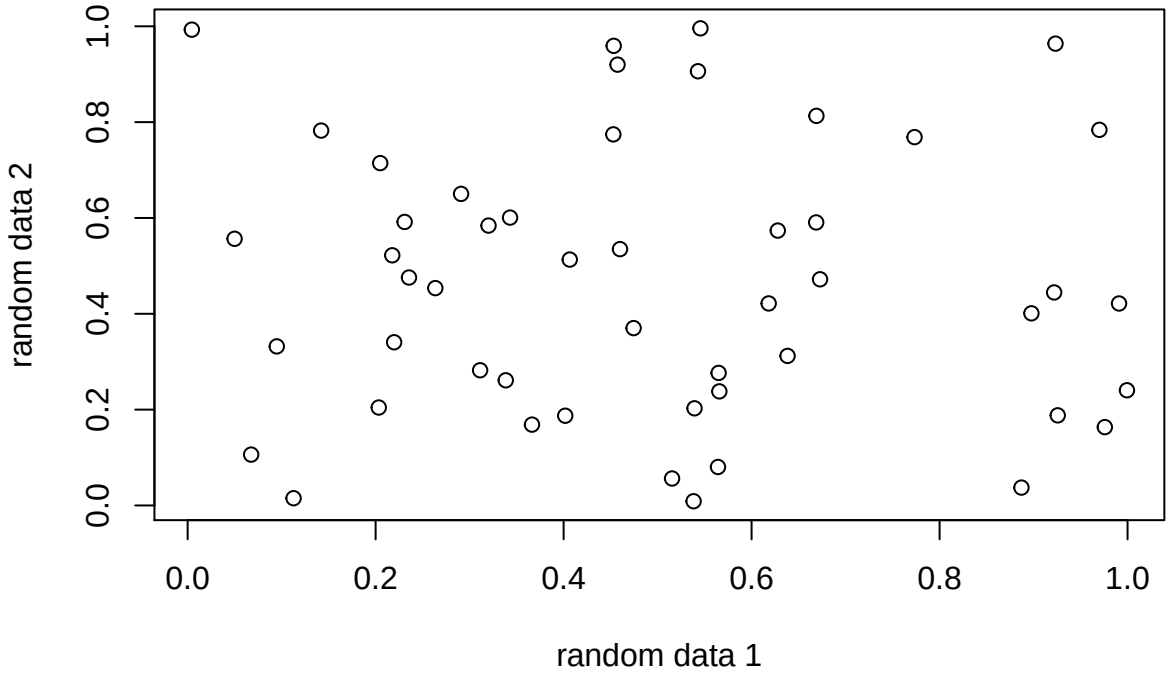
```
png("mydata.png")
list_png[index] = file.info("mydata.png")$size
dev.off()
pdf("mydata.pdf")
list_pdf[index] = file.info("mydata.pdf")$size
dev.off()
}
```

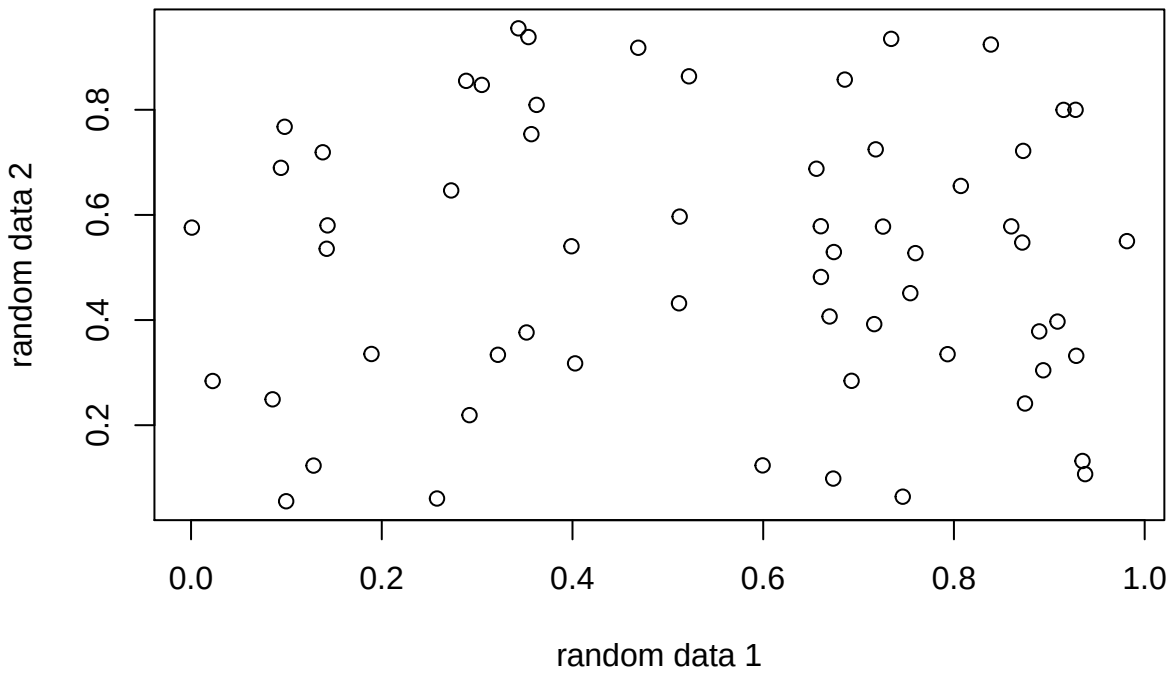


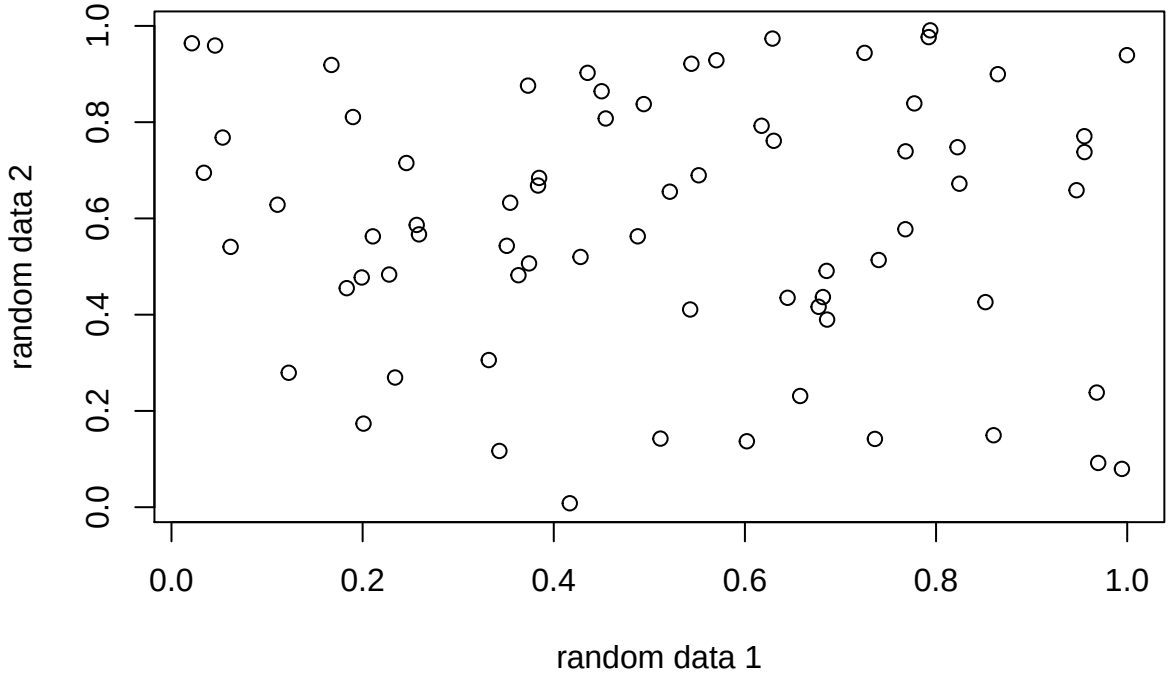

random data 2
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

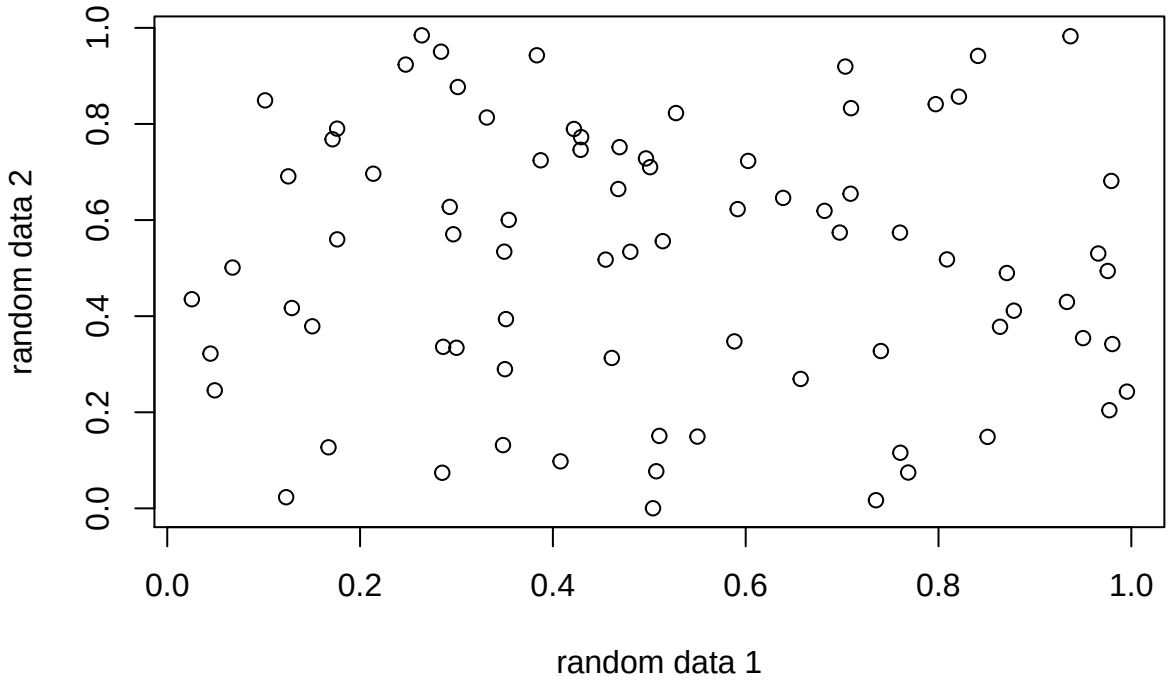

random data 2
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

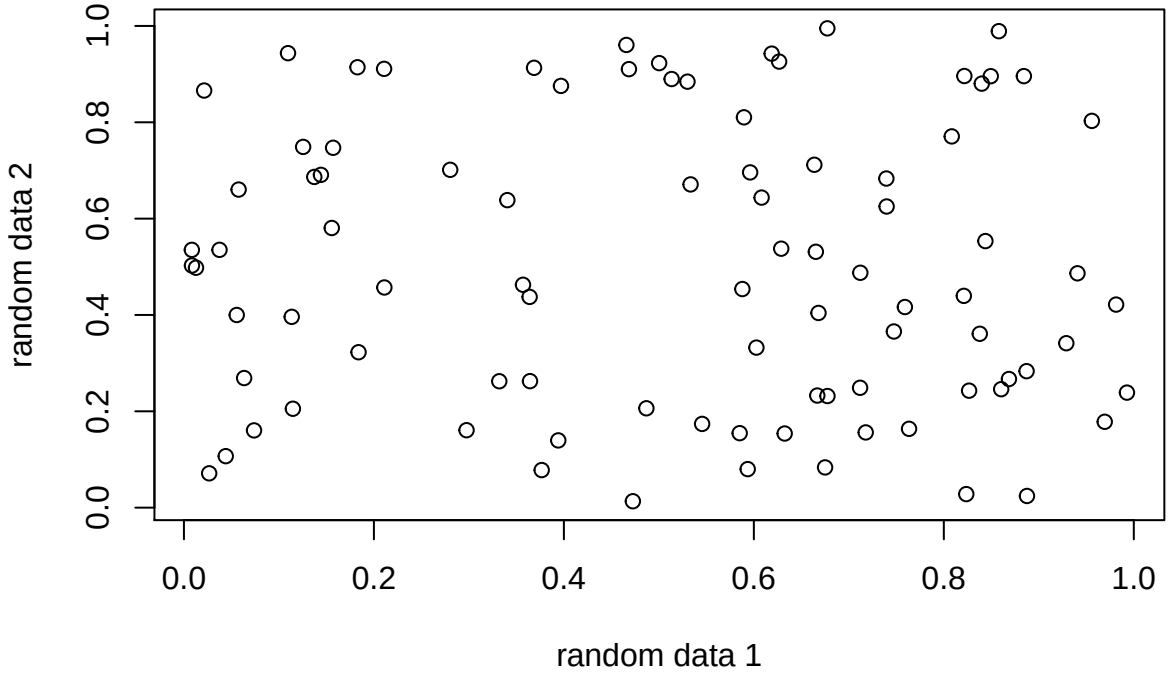

random data 2
0.0
0.2
0.4
0.6
0.8
0.2
0.4
0.6
0.8
1.0
random data 1

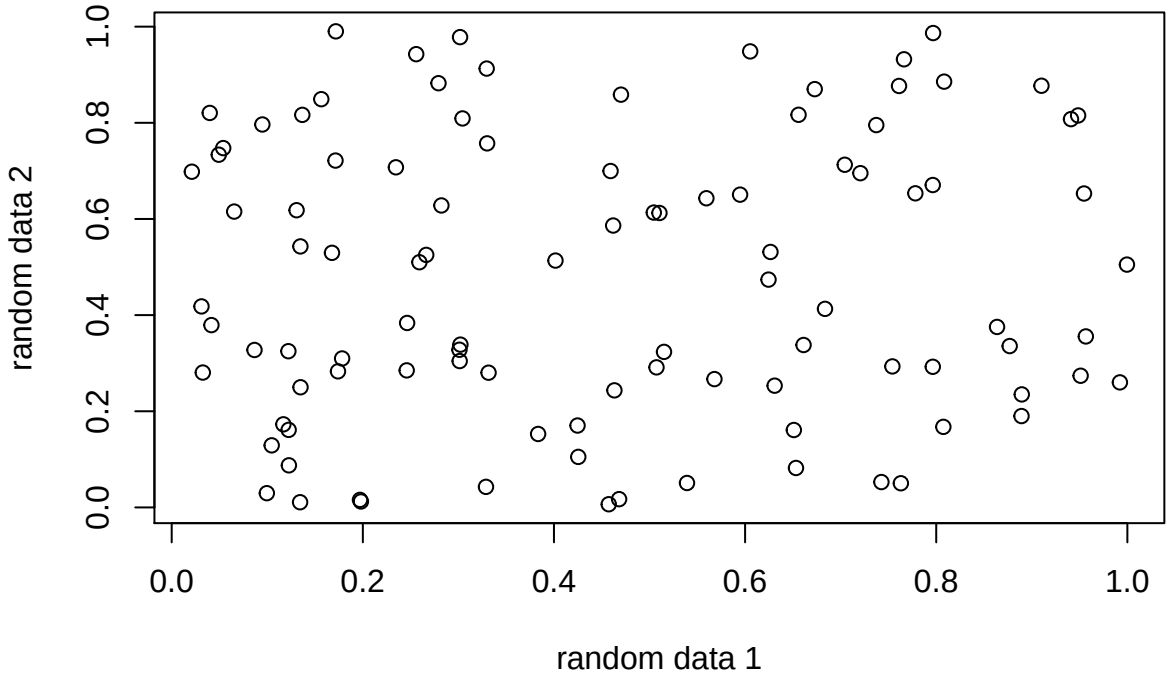

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

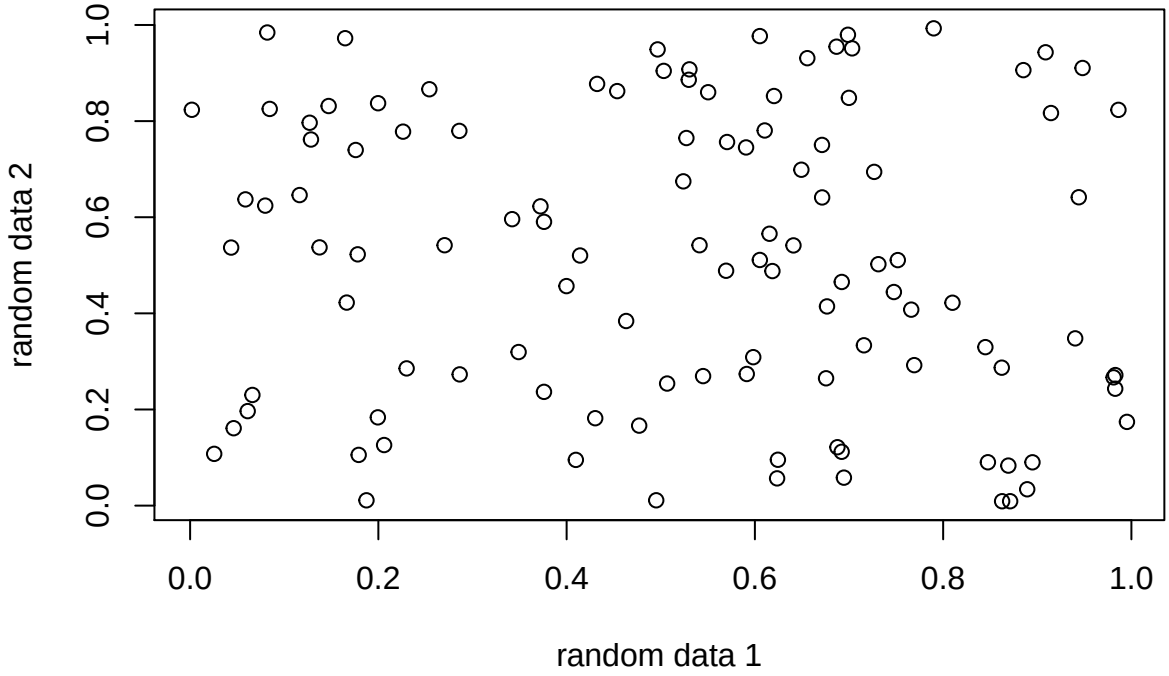

random data 2
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

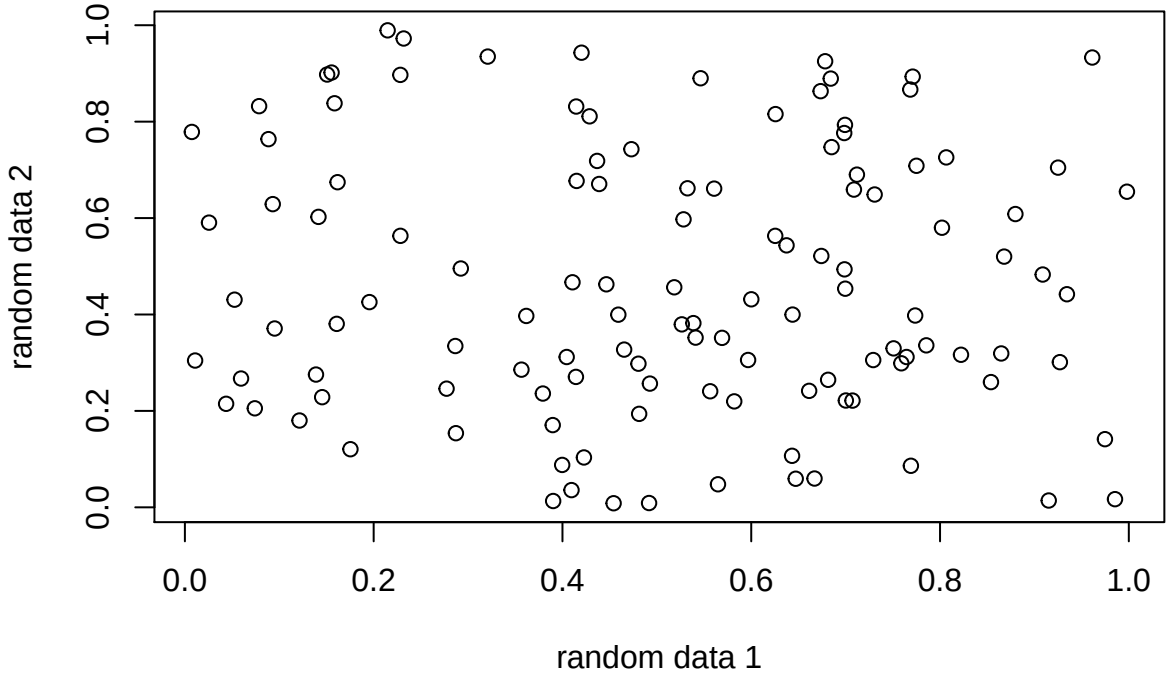

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

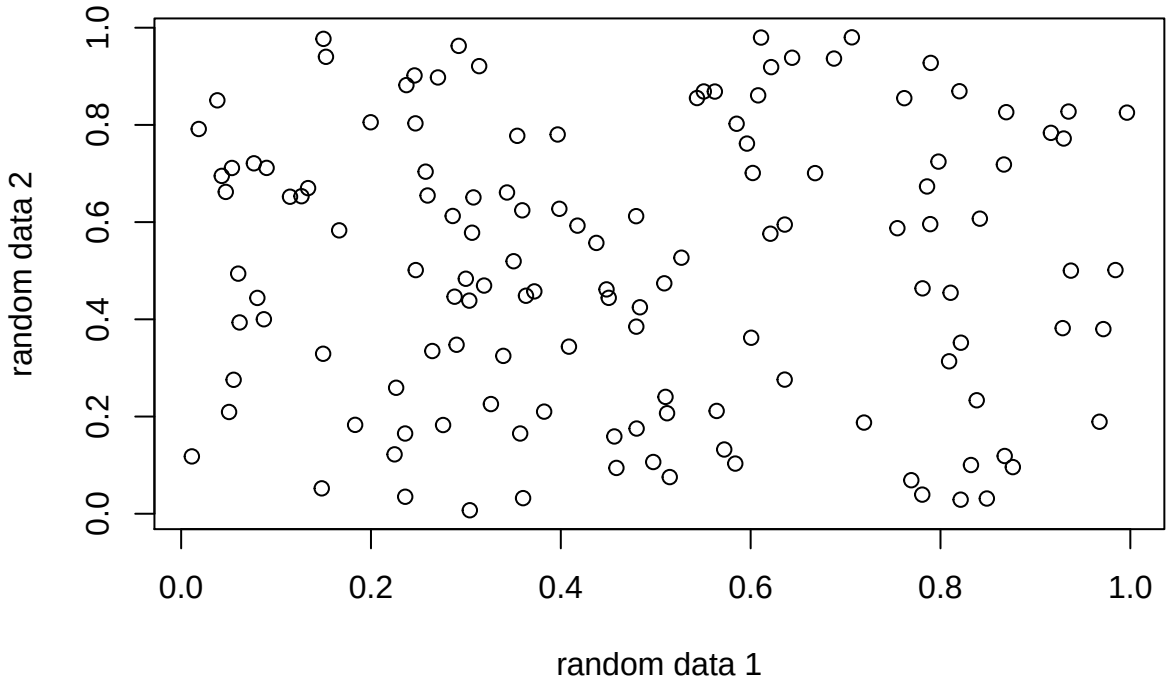

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
random data 1
0.0
0.2
0.4
0.6
0.8
1.0

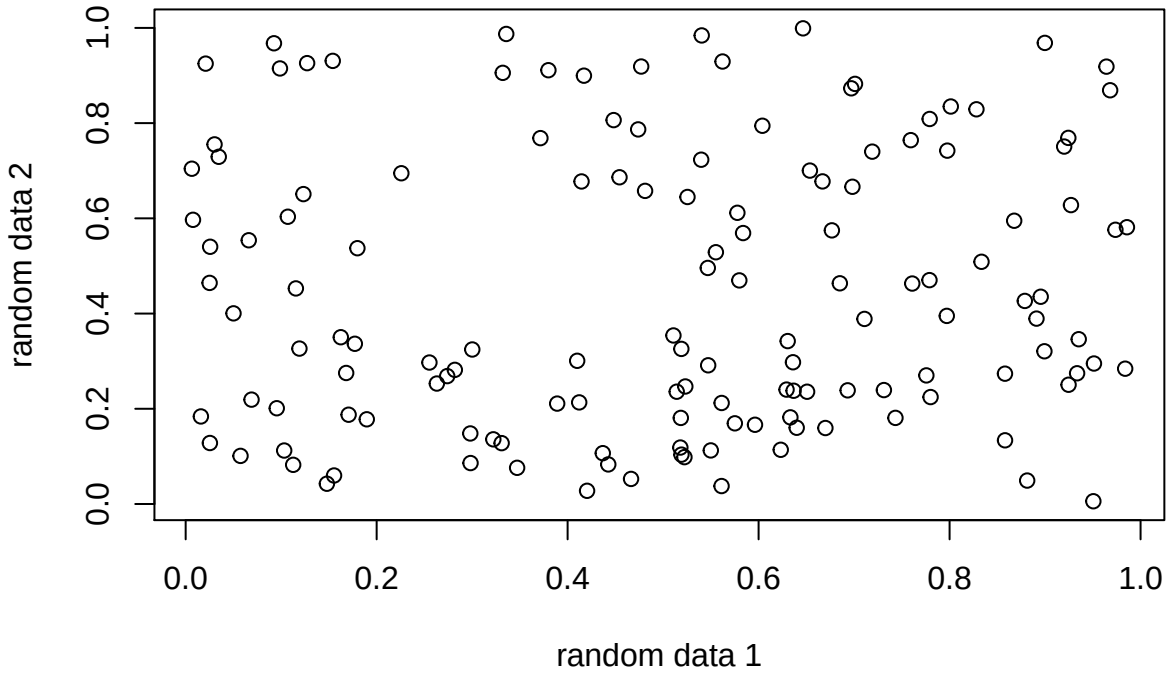

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

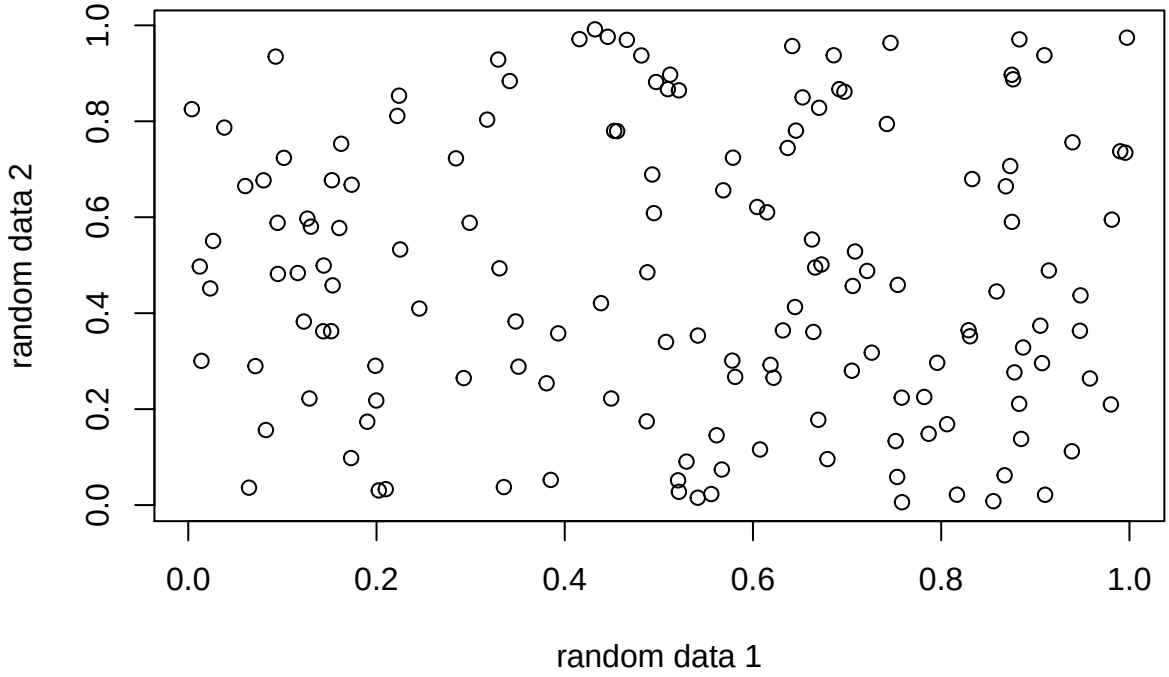

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
random data 1
0.0
0.2
0.4
0.6
0.8
1.0

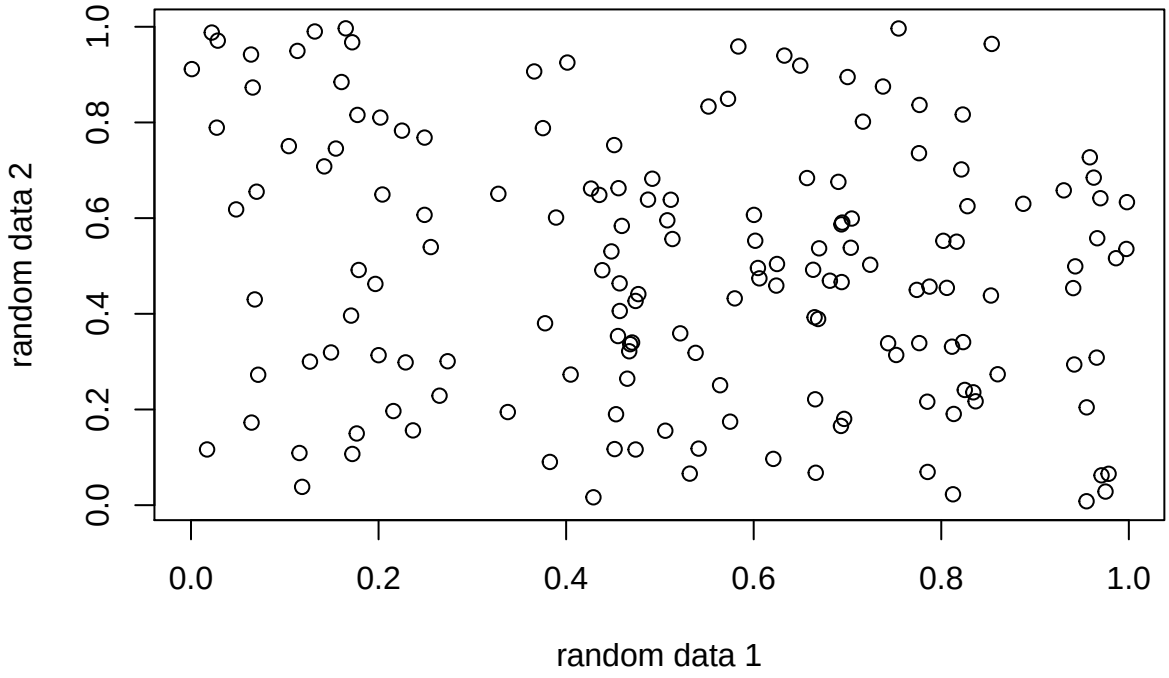

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

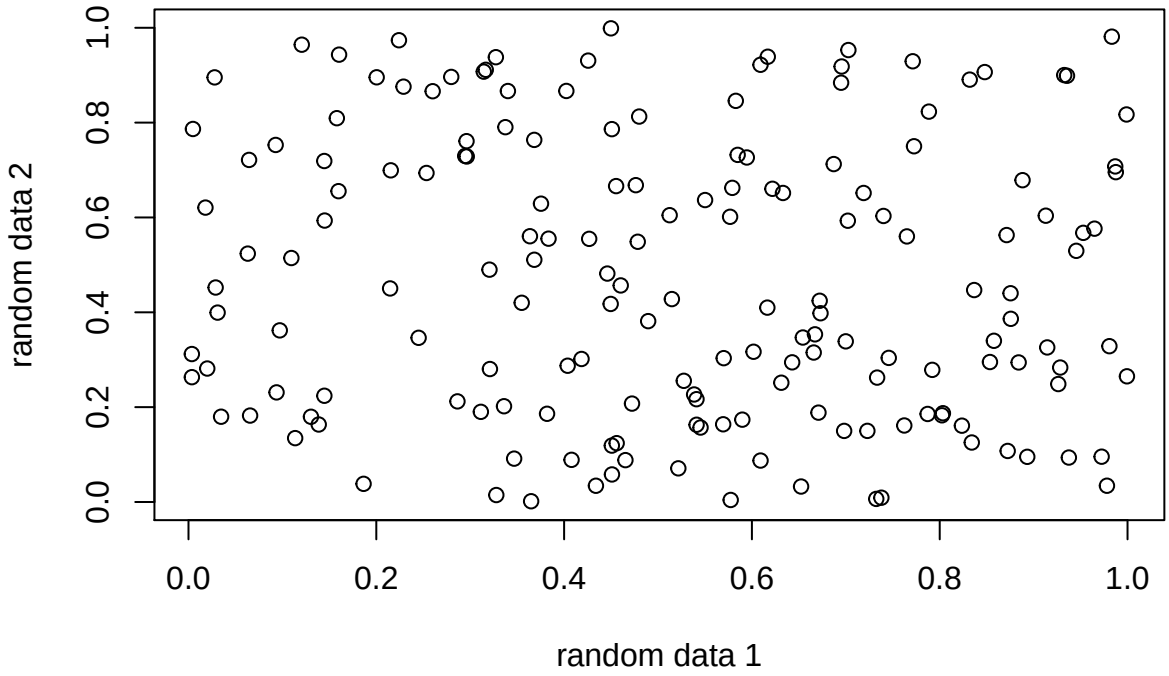

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

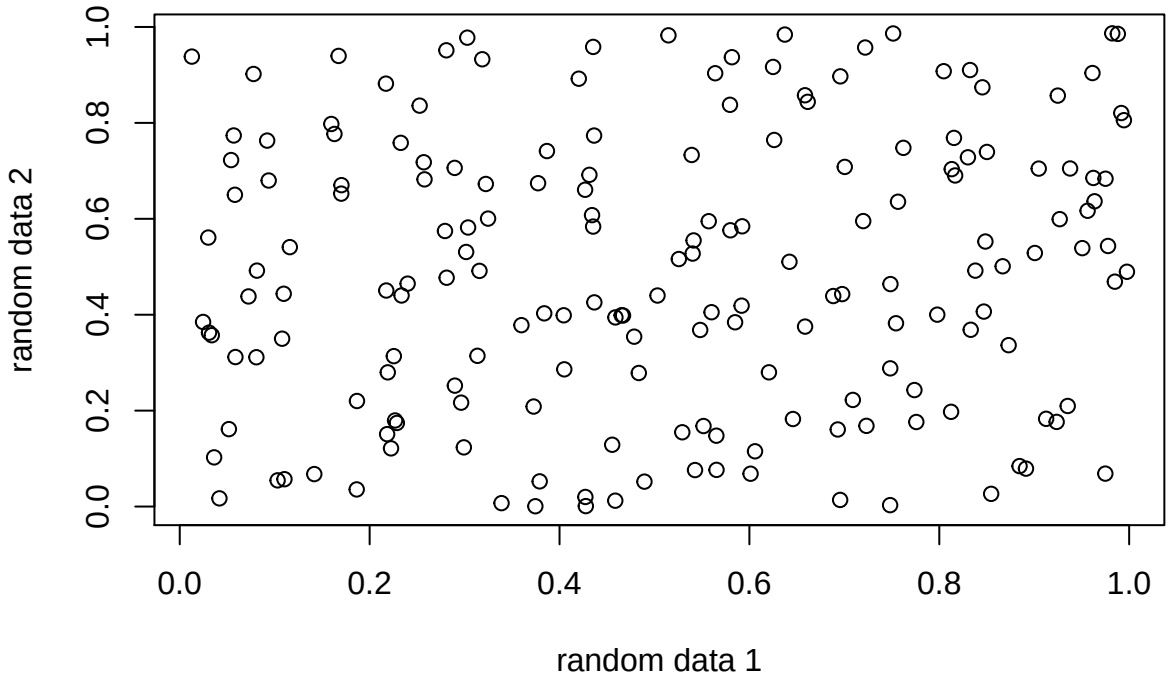

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

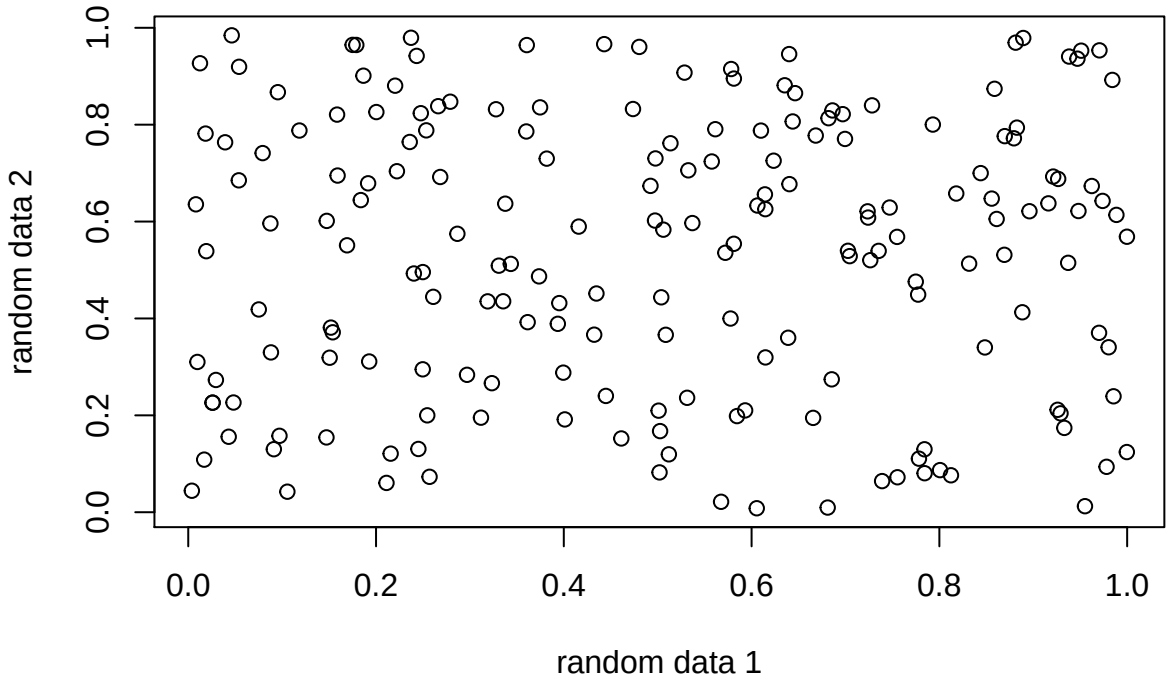

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

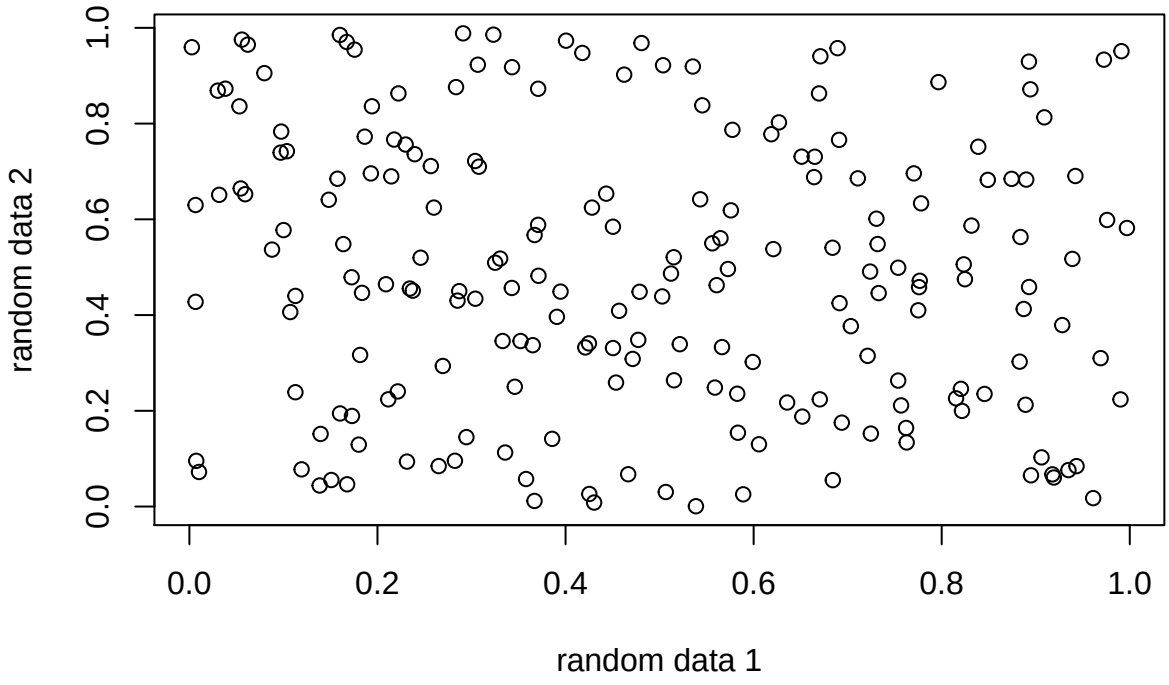

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1


random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1

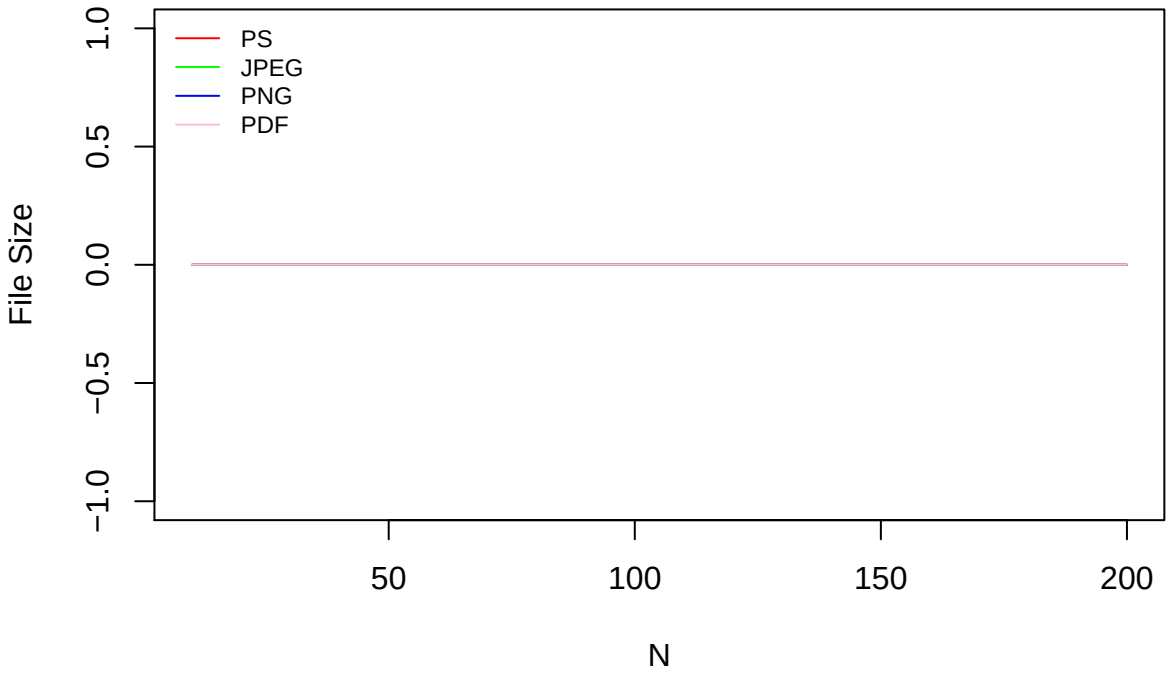

random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1


random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1


random data 2
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
random data 1




```
x = seq(10,200 , by = 10)

g_range =range(0, list_ps,list_jpg,list_png,list_pdf)
plot(x,list_png,type="l",col="red",ylim=g_range,xlab = "N",ylab = "File Size")
lines(x,list_jpg,col="green")
lines(x,list_png,col="blue")
lines(x,list_pdf,col="pink")

legend('topleft', legend = c("PS","JPEG","PNG","PDF") ,
       lty=1, col=c('red', 'green', 'blue','pink'), bty='n', cex=.75)
```


PS
JPEG
PNG
PDF
File Size
1.0
0.5
0.0
−0.5
−1.0
50
100
150
200
N