

Bài 1

Bài 1: Một bảng mạch điện tử sẽ được sản xuất quy mô lớn với một quy trình sản xuất mới được đề xuất. Quy trình này được kiểm tra bằng cách sản xuất 20 bảng mạch, và số lỗi được

24

đếm trên mỗi bảng mạch. Số lỗi sau đây được quan sát trên những bảng mạch này: 0, 1, 3, 2, 1, 2, 2, 2, 3, 4, 1, 0, 0, 1, 0, 2, 0, 0, 2, 0.

- Hãy lập bảng phân phối tần số và tần suất của số lỗi.
- Vẽ biểu đồ cột tần số và đa giác tần số tương ứng với bảng phân phối tần số trên.
- Vẽ biểu đồ cột cho tần suất của số lỗi.
- Tính các số đặc trưng của dữ liệu.

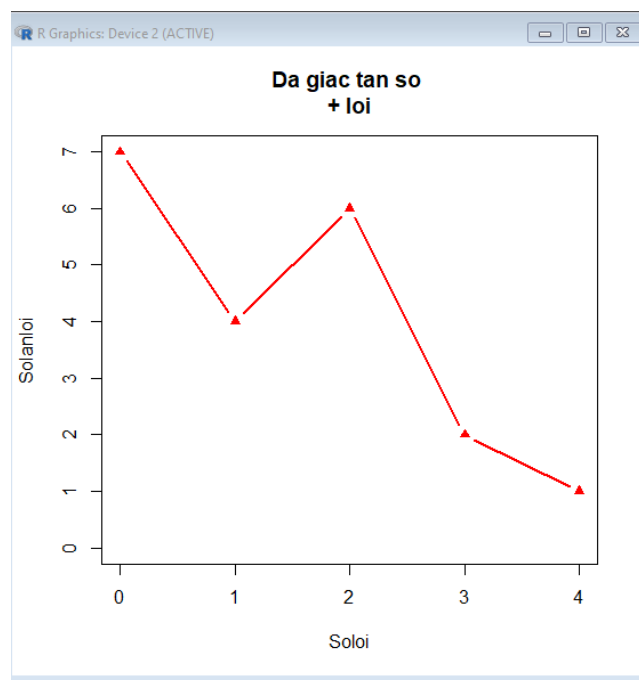
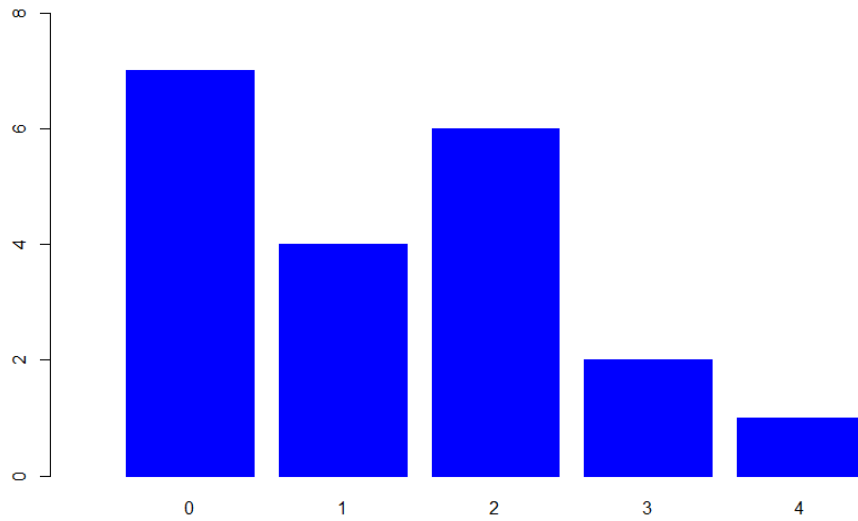
a.

```
> setwd("E:/thuc hanh R")
> #Nhập dữ liệu dưới dạng véc tơ
> soloi=c(0,1,3,2,1,2,2,2,3,4,1,0,0,1,0,2,0,0,2,0)
> #tính tần số
> #Chọn thư mục để lưu dữ liệu
> setwd("F:/thuc hanh R")
Error in setwd("F:/thuc hanh R") : cannot change working directory
> setwd("E:/thuc hanh R")
> #Lưu dữ liệu vừa nhập vào thư mục này
> save(soloi,file="soloi.rda")
> #Tính tần số của các giá trị
> table(soloi)
soloi
0 1 2 3 4
7 4 6 2 1
> #Tính tần suất các giá trị
> prop.table(table(soloi))
soloi
 0    1    2    3    4
0.35 0.20 0.30 0.10 0.05
```

b.

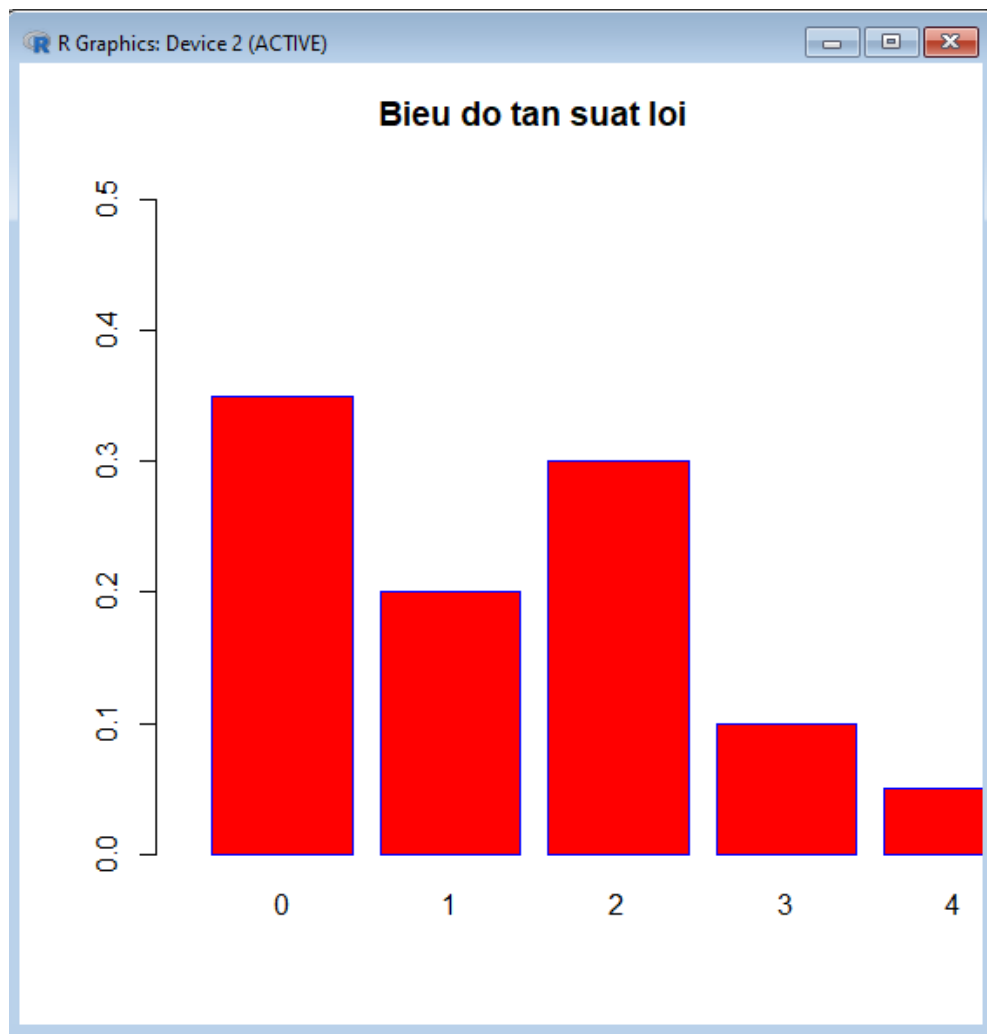
```
> #b, Vẽ biểu đồ cột tần số của số lỗi
> barplot(table(soloi), main = "Bieu do tan so so loi", col
+ = "blue1", border = "blue1", xlim = c(0,10), ylim = c(0, 9))
> # Đa giác tần số tương ứng
> plot(table(soloi), type="b",main="Da giac tan so
+ loi",col="blue1",pch=16,xlab="Soloi",ylab="Solanloi")
> plot(table(soloi), type="b",main="Da giac tan so
+ + loi",col="red",pch=17,xlab="Soloi",ylab="Solanloi")
```

Bieu do tan so so loi



C

```
> # Biểu đồ cột cho tần suất  
> barplot(prop.table(table(soloi)),main="Bieu do tan suat loi  
+ ", col= "red",border= "blue",xlim=c(0,5),ylim=c(0,0.5))  
..
```



D,

```
> #d  
> #Tinh giá trị trung bình  
> mean(soloi)  
[1] 1.3  
> #Tinh trung vị  
> median(soloi)  
[1] 1  
> #Tinh mode  
> which(table(soloi) == max(table(soloi)))  
0  
1  
> #Tìm giá trị nhỏ nhất, giá trị lớn nhất  
> range(soloi)  
[1] 0 4  
> # Tính độ rộng của dữ liệu  
> dorong = 4-0  
> dorong  
[1] 4  
> #Tinh phương sai mẫu  
> var(soloi)  
[1] 1.484211  
> #Tinh độ lệch chuẩn mẫu  
> sd(soloi)  
[1] 1.218282
```

Bài 2;

"Van tai, kho bai", "Tai chinh, ngan hang va bao hiem", "Kinh doanh bat dong san", "Giao duc va dao tao"))

Bài 2: Năng suất lao động (đơn vị: triệu đồng/người) được thống kê theo ngành kinh tế và theo năm được cho trong bảng dưới đây

Thành phần kinh tế	Năm 2005	Năm 2010	Năm 2015
Vận tải, kho bãi	21,7	43,8	71,9
Tài chính, ngân hàng và bảo hiểm	257,3	457,8	631,1
Kinh doanh bất động sản	3232,2	1300	1284,7
Giáo dục và đào tạo	21,4	30	72,1

Hãy vẽ biểu đồ cột biểu diễn năng suất lao động theo thành phần kinh tế và theo năm và cho nhận xét.

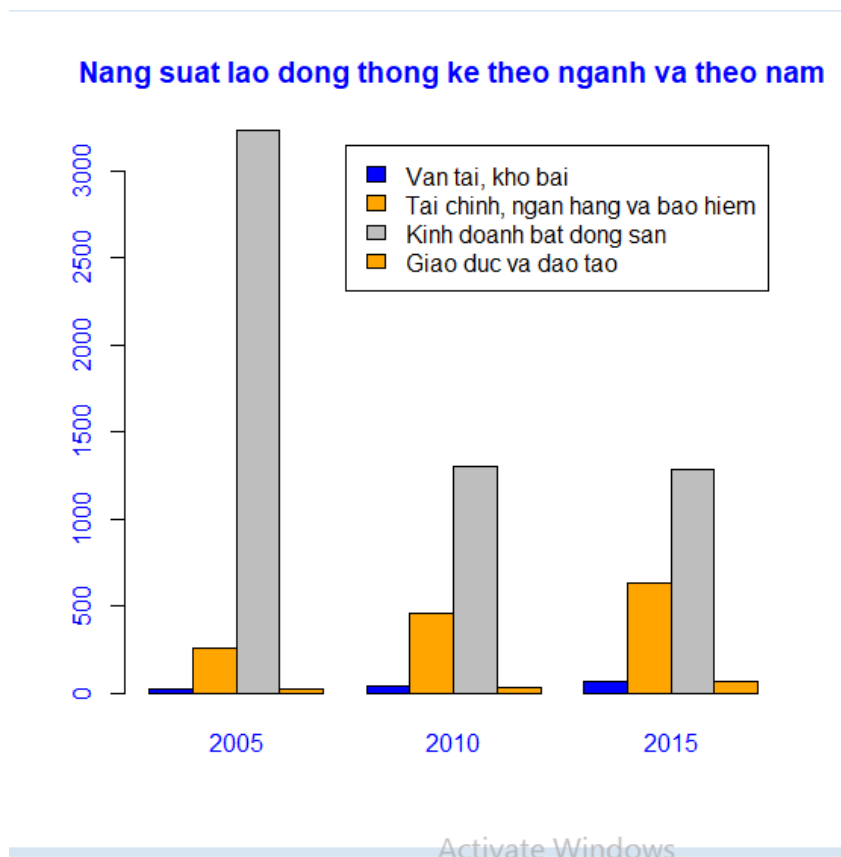
> # Nhập dữ liệu dưới dạng matrix với 4 hàng

>

```
ns=matrix(c(21.7,257.3,3232.2,21.4,43.8,457.8,1300,30,71.9,631.1,1284.7,72.1), nrow=4)
```

> # Vẽ biểu đồ cột năng suất lao động thống kê theo ngành và theo năm

```
> barplot(ns, main = "Nang suat lao dong thong ke theo nganh va theo nam", col = c("blue1", "orange", "gray", "orange"), names.arg = c(2005, 2010, 2015), beside=T, col.main = "blue", col.axis = "blue", legend.text = c("Van tai, kho bai", "Tai chinh, ngan hang va bao hiem", "Kinh doanh bat dong san", "Giao duc va dao tao"))
```



Bài 3

Bài 3: Dữ liệu sau đây là về chất thải rắn (đơn vị: triệu tấn) được thải ra môi trường hàng năm của một quốc gia:

Rác thải đô thị	150
Công nghiệp	350
Khai mỏ	1700
Nông nghiệp	2300

- Hãy vẽ biểu đồ cột biểu diễn dữ liệu trên và cho nhận xét.
- Hãy vẽ biểu đồ hình tròn biểu diễn dữ liệu trên và cho nhận xét.
- So sánh thông tin có được khi dùng biểu đồ cột và biểu đồ hình tròn biểu diễn dữ liệu trên.

A,

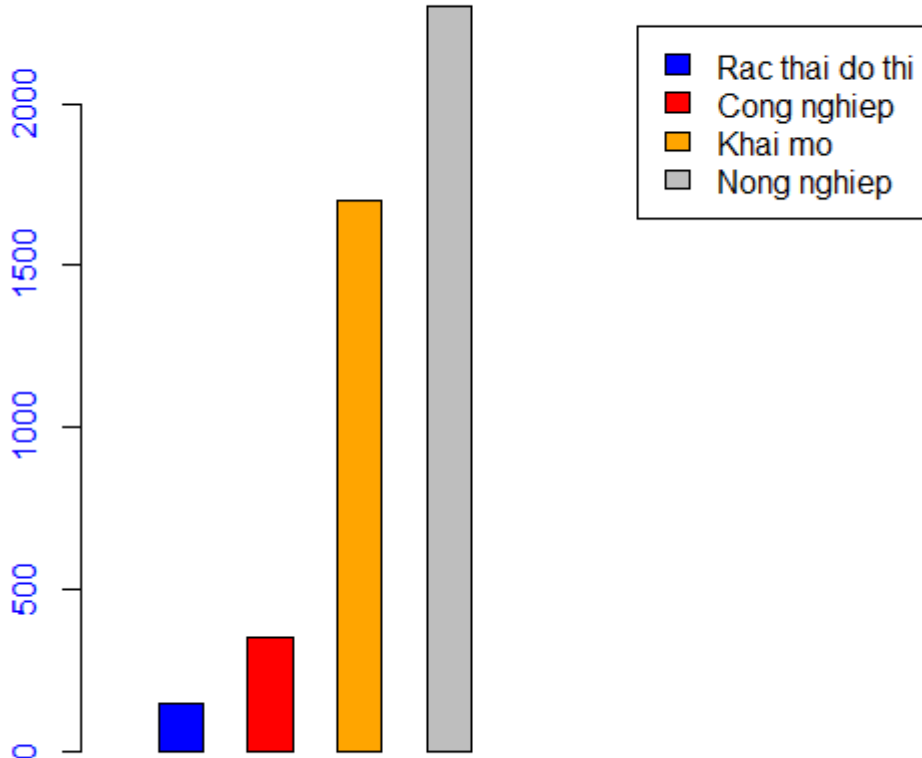
>#Nhập dữ liệu dưới dạng matrix với 1 hàng

```
>chatthai = matrix(c(150,350,1700,2300),nrow=1)
```

># Biểu đồ cột lượng chất thải môi trường hàng năm

```
> barplot(chatthai,main="Luong chat thai ra moi  
truong",col=c("blue1","red","orange","gray"),beside=T,col.main="black",col.a  
xis="blue" , xlim=c(0,18),legend.text=c("Rac thai do thi","Cong nghiep","Khai  
mo","Nong nghiep"))
```

Luong chat thai ra moi truong



B,

Nhập dữ liệu dạng bảng

```
> chatthai=c(150,350,1700,2300)
```

```
> loaichatthai=c("Rac thai do thi","Cong nghiep","Khai mo","Nong nghiep")
```

```
> Dulieuchatthai=data.frame(chatthai,loaichatthai)
```

># Tính tỉ lệ phần trăm từng loại đường

```
>Tile=round(prop.table(chatthai),2)*100
```

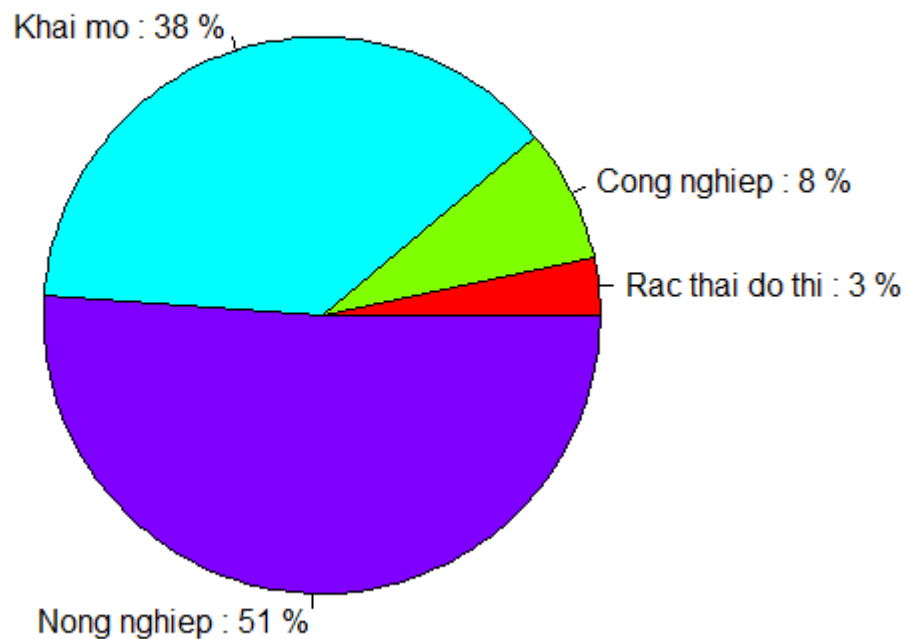
```
>Tile
```

```
[1] 3 8 38 51
```

>#Vẽ biểu đồ hình tròn lượng chất thải ra môi trường hàng năm

```
>pie(chatthai,labels=paste(loaichatthai,":",Tile,"%"),col=rainbow(4),lty=1,mai  
n="Luong chat thai ra moi truong ",sub="So lieu hang nam")
```

Luong chat thai ra moi truong



C, So sánh

Biểu đồ tròn thể hiện rõ hơn về tỷ trọng phần trăm của từng loại rác

Biểu đồ cột thể hiện rõ hơn về chi tiết khối lượng từng loại rác

Bài 4

Bài 4: Một nhà máy xử lý nước cung cấp cho một khu vực dân cư được xây dựng với công suất thiết kế 17000 mét khối một ngày. Khi nhu cầu dùng nước vượt quá khả năng cung cấp, các hệ thống tưới tiêu công cộng sẽ bị dừng hoạt động. Người ta khảo sát nhu cầu nước (đơn vị: nghìn mét khối) trong một giai đoạn và kết quả được cho dưới đây:

8,7	12,1	12,6	13,7	14,8	15,1	15,4	15,9	16,2	16,5
16,8	16,8	17,1	17,2	17,3	17,4	17,6	17,7	17,7	17,9
18,0	18,1	18,2	18,2	18,4	18,5	18,6	18,6	18,6	18,7
18,9	19,1	19,1	19,1	19,5	19,5	19,5	20,2	21,0	16,6
17,9	18,9								

a) Xác định các tứ phân vị, các đặc trưng số về tâm, các đặc trưng số về sự phân tán của dữ liệu này. Giải thích ý nghĩa của các giá trị.

b) Vẽ biểu đồ hộp và râu để minh họa sự phân bố của tập dữ liệu.

c) Xác định số khoảng chia theo công thức Sturge, các điểm chia và vẽ biểu đồ histogram tần số. Tính tỉ lệ quan sát ở đó nhu cầu vượt quá khả năng cung cấp.

d) Vẽ biểu đồ histogram tần số, sử dụng 8 khoảng chia.

># Nhập dữ liệu:

```
>nuoc=c(8.7,12.1,12.6,13.7,14.8,15.1,15.4,15.9,16.2,16.5,16.8,16.8,17.1,17.2,17.3,17.4,17.6,17.7,17.7,17.9,18.0,18.1,18.2,18.2,18.4,18.5,18.6,18.6,18.6,18.7,18.9,19.1,19.1,19.1,19.5,19.5,19.5,20.2,21.0,16.6,17.9,18.9)
```

```
> xulynuoc=data.frame(nuoc)
```

```
>#tứ phân vị
```

```
> summary(nuoc)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
8.70 16.65 17.90 17.33 18.68 21.00
```

```
>#đặc trưng số về tâm
```

```
> n=length(nuoc)
```

```
> n
```

```
[1] 42
```

```
># trung bình
```

```
> mean(nuoc)
```

```
[1] 17.32619
```

```
>#trung vị
```

```
> median(nuoc)
```

```
[1] 17.9
```

```
>#min max
```

```
> min(nuoc)
```

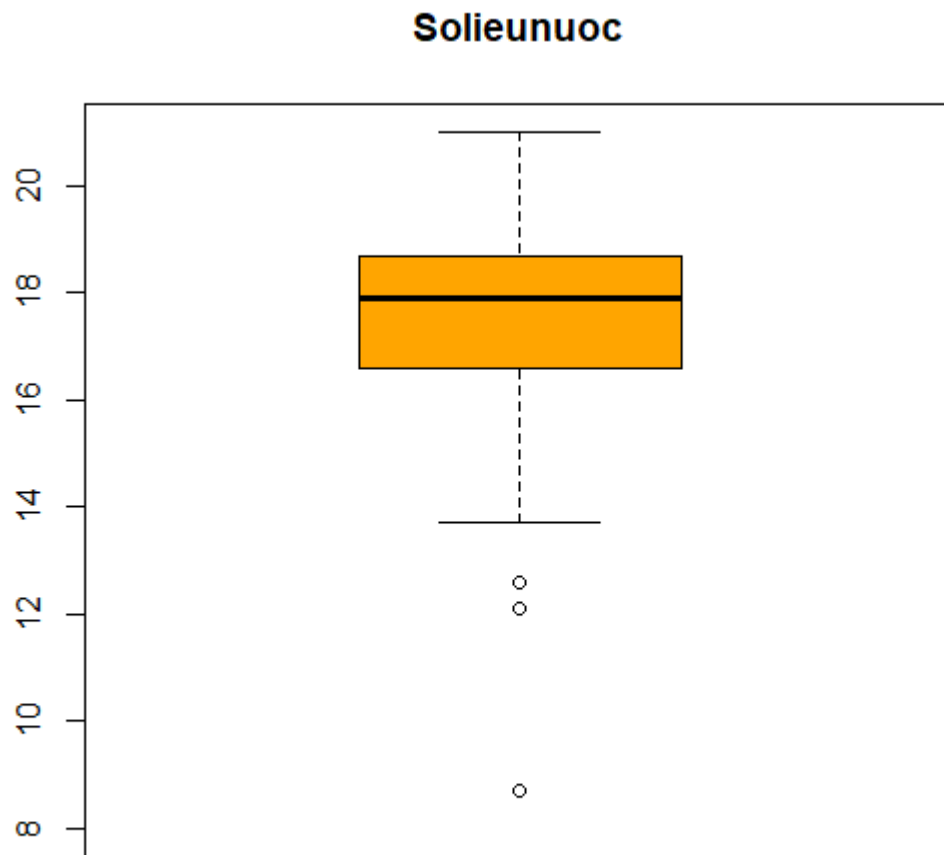
```
[1] 8.7
```

```
> max(nuoc)
```

```
[1] 21
```

```
b.
```

```
>boxplot(nuoc,main="Solieunuoc",boeder="gray",col="orange",horiz=F,ylim=c(8,21))
```

C.

>#xác định khoảng chia

> k=1+3.3*log(42,base=10)

> k

[1] 6.356723

> dorong=(21-8.7)/6

> dorong

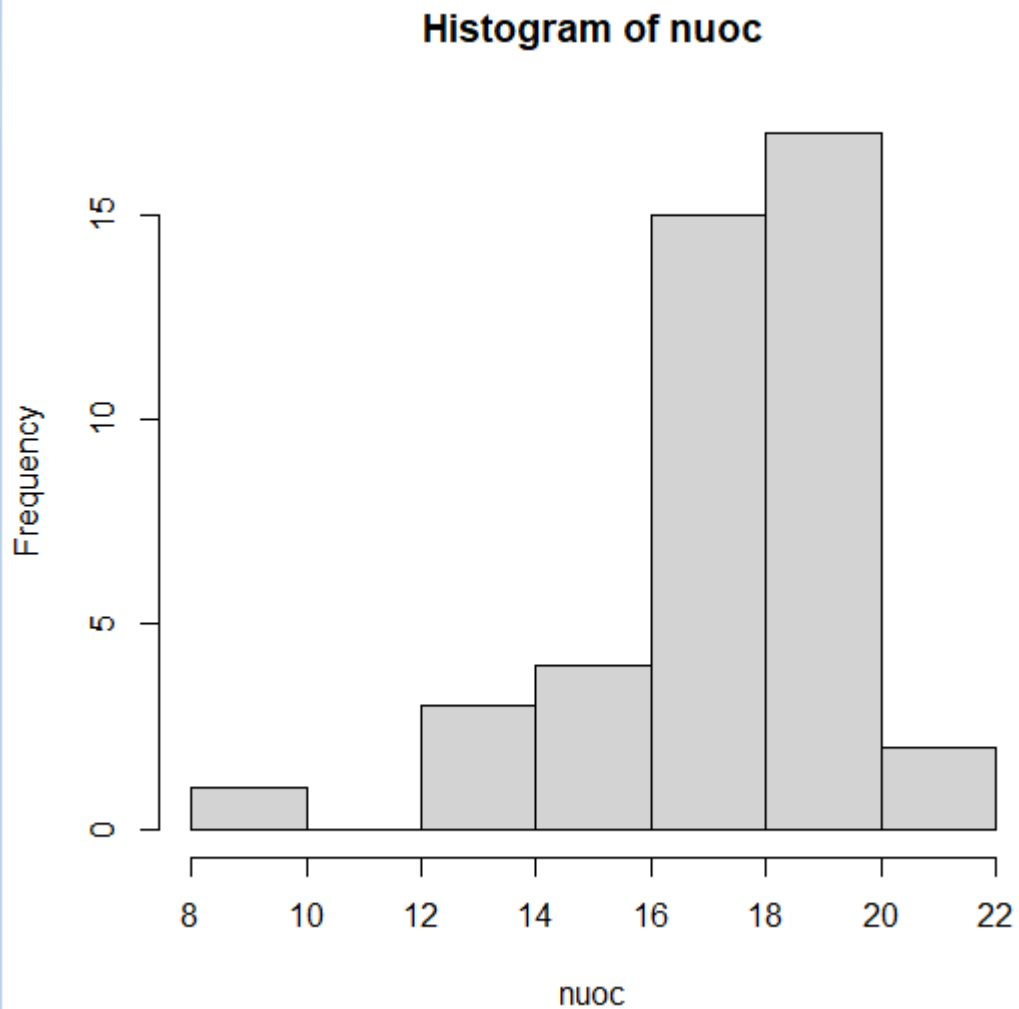
[1] 2.05

> table(cut(nuoc,breaks=6,right=F))

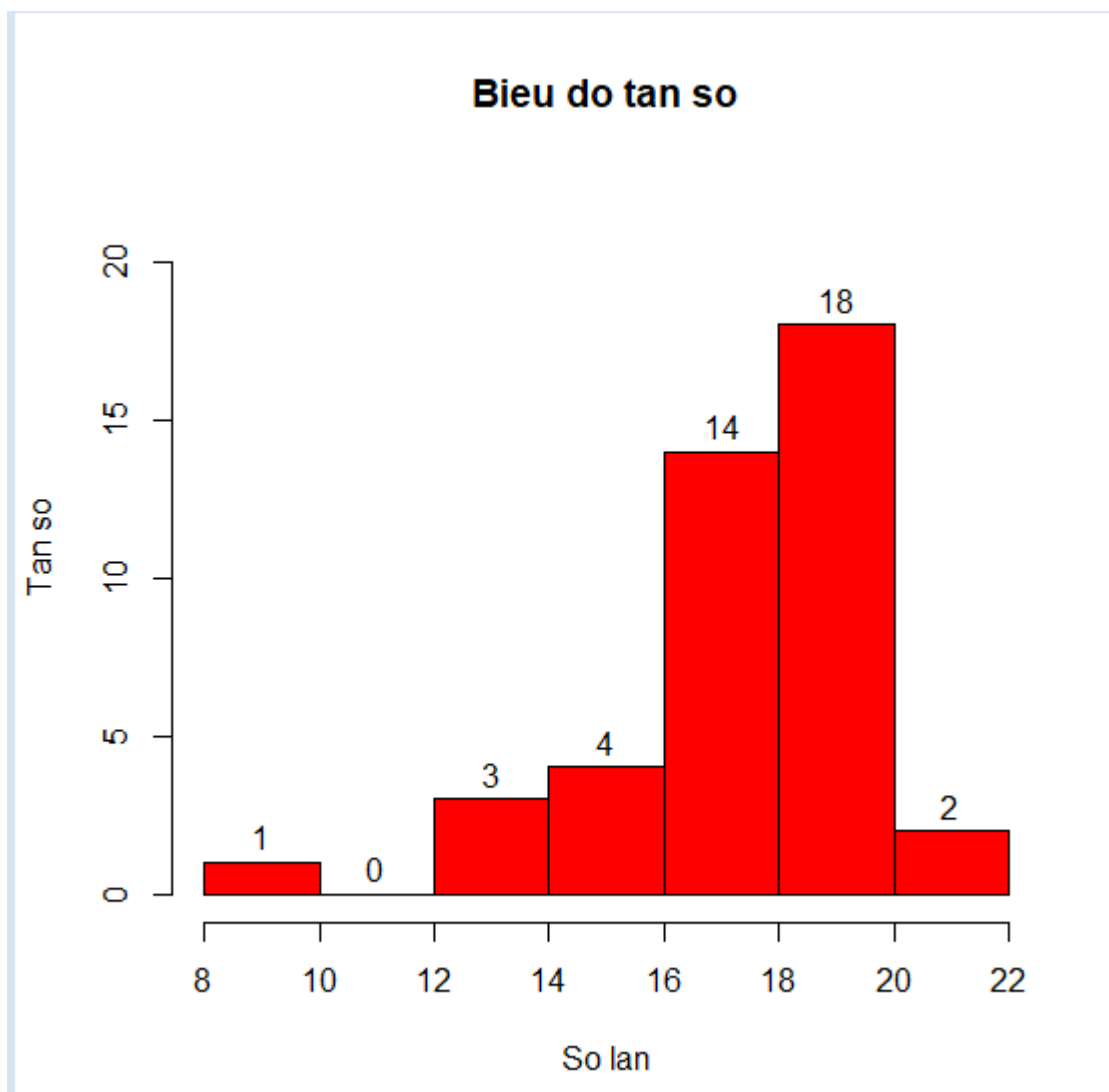
[8.69,10.8) [10.8,12.8) [12.8,14.8) [14.8,16.9) [16.9,19) [19,21)

1 2 2 8 21 8

>hist (nuoc)



```
> hist(nuoc,xlim=c(8,22),ylim=c(0,22),right=F,xlab="So lan",ylab="Tan  
so",labels=T,main="Bieu do tan so",col="red",brder="red")
```



d.

```
>table(cut(nuoc,breaks=8,right=F))
```

```
[8.69,10.2) [10.2,11.8) [11.8,13.3) [13.3,14.8) [14.8,16.4) [16.4,17.9)
```

```
1      0      2      2      4      13
```

```
[17.9,19.5) [19.5,21)
```

```
15      5
```

```
>hist(nuoc, xlim = c(6, 22), ylim = c(0, 20), breaks = seq(6, 22, 8), right = F,
xlab = "so lan", ylab = "Tan so", xaxt = "n", yaxt = "n", labels = T, main =
"Bieu do tan so", col = "lightblue1", border = "blue1")
```

```
>axis(side = 1, c(6,8,10,12,14,16,18,20,22))
```

```
>axis(side = 2, c(0,5,10,15,20))
```

Bài 5

Bài 5: Để chọn một trong hai phương án đầu tư được đề xuất, một người thu thập dữ liệu về lợi nhuận của hai phương án đầu tư như dưới đây

Lợi nhuận của phương án đầu tư A					Lợi nhuận của phương án đầu tư B				
30,00	6,93	13,77	-8,55	-2,13	30,33	-34,75	30,31	24,30	-30,37
-13,24	22,42	-5,29	4,30	-18,95	54,19	6,06	-10,01	-5,61	44
34,40	-7,04	25	9,43	49,87	14,73	35,24	29	-20,23	36,13
-12,11	12,89	1,21	22,92	12,89	40,70	-26,01	4,16	1,53	22,18
-20,24	31,76	20,95	63	1,20	0,46	10,03	17,61	3,24	2,07
11,07	43,71	-19,27	-2,59	8,47	10,51	1,20	25,10	29,44	39,04
-12,83	-9,22	33,00	36,08	0,52	9,94	-24,24	11	24,76	-33,39
-17	14,26	-21,95	61	17,30	-38,47	-25,93	15,28	58,67	13,44
-15,83	10,33	-11,96	52	0,63	8,29	34,21	0,25	68	61
12,68	1,94	38	13,09	28,45	52	5,23	-20,44	-32,17	66

- Tính giá trị trung bình, độ lệch chuẩn của mỗi tập dữ liệu và đưa ra nhận xét.
- Hãy vẽ biểu đồ histogram tần số cho mỗi tập dữ liệu.
- Phân tích biểu đồ và đưa ra kết luận về phương án đầu tư tốt hơn.

A,

># tính dự án a

>#nhập

```
>duana=c(30.00,6.93,13.77,-8.55,-2.13,- 13.24,22.42,-5.29,4.30,-18.95,34.40,-
7.04,25,9.43,49.87,12.11,12.89,1.21,22.92,12.89,-20.24,31.76,20.95,63,1.20,
11.07,43.71,-19.27,-2.59,8.47,-12.83,-9.22,33.00,36.08,0.52,-17,14.26,-
21.95,61,17.30,-15.83,10.33,-11.96,52,0.63,12.68,1.94,38,13.09,28.45)
```

>#độ dài

```
> length(duana)
```

```
[1] 50
```

>#trung bình

```
> mean(duana)
```

```
[1] 10.9454
```

>#Tính độ lệch chuẩn

```
> sd(duana)
```

```
[1] 21.89401
```

>#tính dự án b

>#nhập

```
>duanb=c(30.33,-34.75,30.31,24.30,-30.37,54.19,6.06,- 10.01,-  
5.61,44,14.73,35.24,29,-20.23,36.13,40.70,-  
26.01,4.16,1.53,22.18,0.46,10.03,17.61,3.24,2.07,10.51,1.20,25.10,29.44,39.04  
,9.94,-24.24,11,24.76,-33.39,- 38.47,-  
25.93,15.28,58.67,13.44,8.29,34.21,0.25,68,61,52,5.23,- 20.44,-32.17,66)
```

>#độ dài

```
> length(duanb)
```

```
[1] 50
```

>#trung bình

```
> mean(duanb)
```

```
[1] 12.7602
```

>#tính độ lệch chuẩn

```
> sd(duanb)
```

```
[1] 28.04676
```

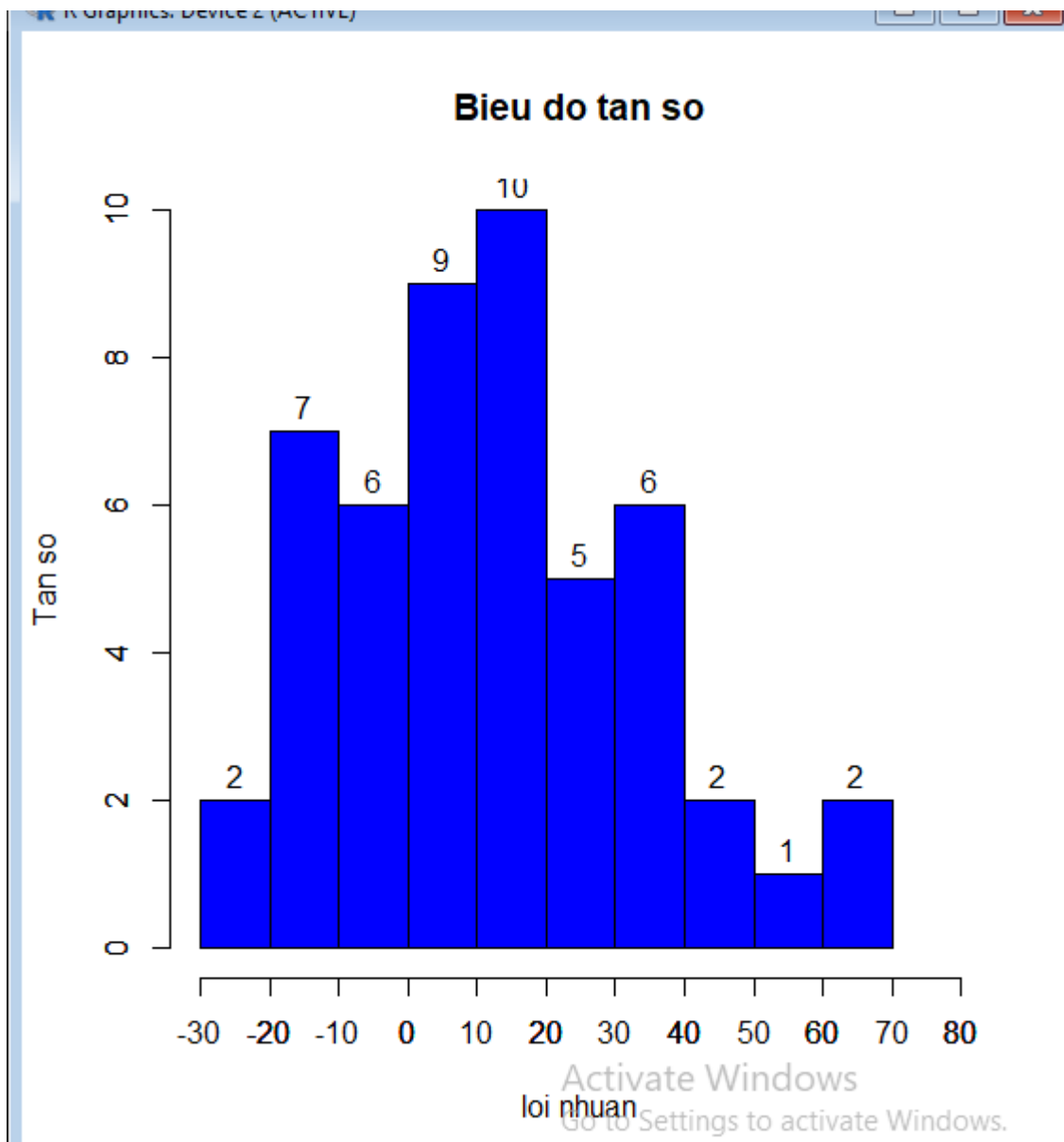
B,

># dự án A

```
>hist(duana,xlim=c(-30,80),ylim=c(0,10),right=F,xlab="loi nhuan",ylab="Tan  
so",labels=T,main="Bieu do tan so",col="blue1",brder="lightblue1")
```

```
> axis(side = 1, c(-30,-20,-10,0,10,20,30,40,50,60,70,80))
```

```
> axis(side = 2, c(0,2,4,6,8,10))
```

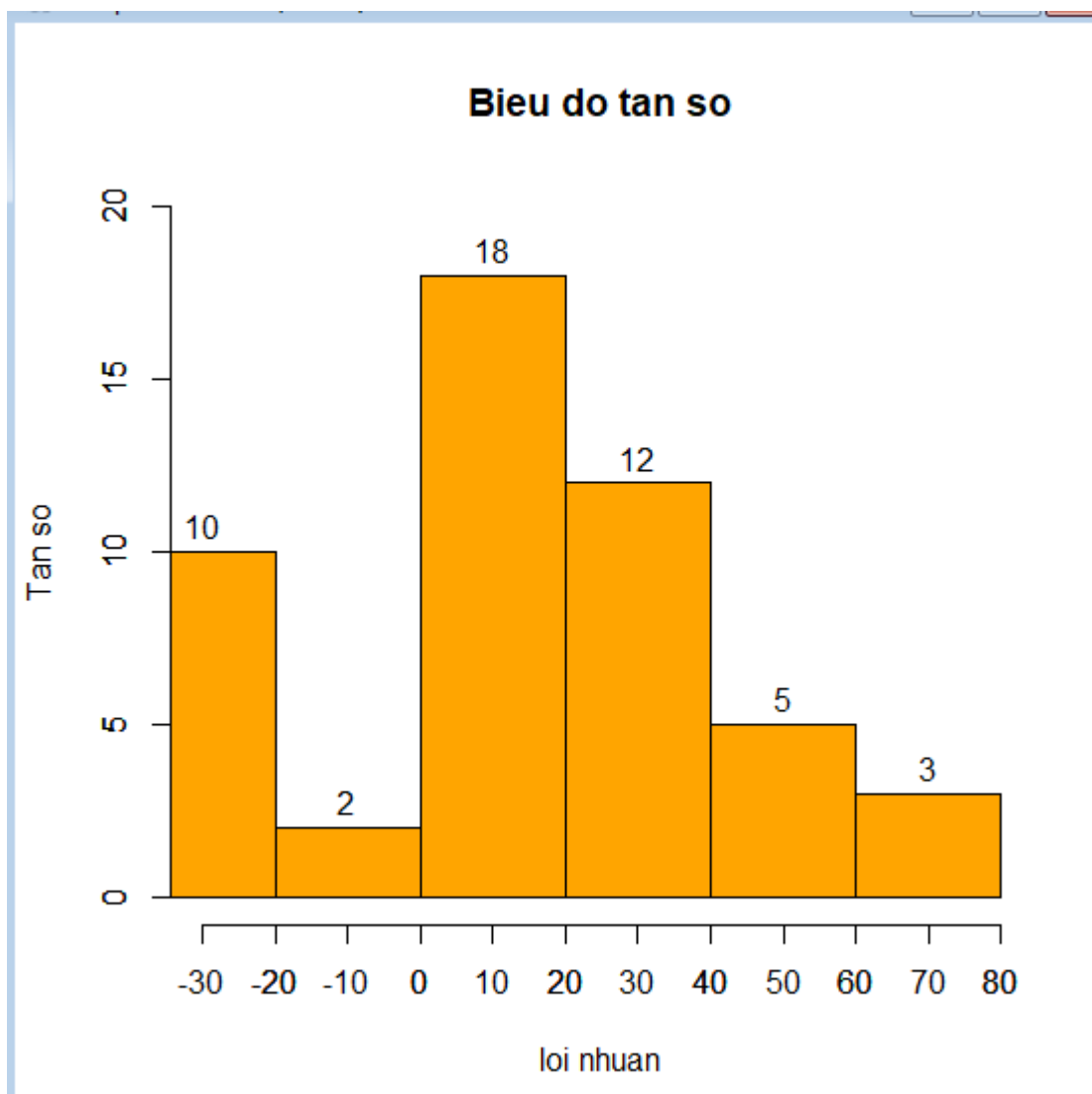


># dự án b

```
>hist(duanb,xlim=c(-30,80),ylim=c(0,20),right=F,xlab="loi nhuan",ylab="Tan so",labels=T,main="Bieu do tan so",col="orange",brder="red")
```

```
> axis(side = 1, c(-30,-20,-10,0,10,20,30,40,50,60,70,80))
```

```
> axis(side = 2, c(0,5,10,15,20))
```



C. dự án A có tần số lợi nhuận rơi ở mức cao hơn dự án b mà mức thua lỗ còn thấp hơn

Vậy,Nên đầu tư vào dự án A