

Ví dụ

Giả sử có tập dữ liệu về tuổi của người lao động trong một phân xưởng như sau:

| | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|
| 48 | 30 | 35 | 31 | 21 | 28 | 34 | 43 | 36 | 45 | 41 |
| 33 | 47 | 47 | 30 | 47 | 44 | 45 | 32 | 46 | 47 | 23 |
| 30 | 23 | 49 | 20 | 24 | 20 | 40 | 50 | | | |

Ta có $X_{\min} = 20$, $X_{\max} = 50$, $k = \log_2(30) + 1 \approx 6$, $h = (50 - 20)/6 = 5$.
Các tổ là:

$[20, 25]; (25, 30]; (30, 35]; (35, 40]; (40, 45]; (45, 50]$

```
> library(BSDA)
```

```
> x=c(48, 30, 35, 31, 21, 28, 34, 43, 36, 45, 41, 33, 47, 47, 30, 47, 44, 45, 32, 46, 47, 23, 30, 23, 49, 20, 24, 20, 40, 50)
```

Bang tần số của x

```
> table(x)
```

x

20 21 23 24 28 30 31 32 33 34 35 36 40 41 43 44 45 46 47 48 49 50

2 1 2 1 1 3 1 1 1 1 1 1 1 1 1 2 1 4 1 1 1

Bang tần suất

```
> prop.table(table(x))
```

x

| | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|
| 20 | 21 | 23 | 24 | 28 | 30 | 31 | 32 | 33 | 34 | 35 | 36 |
| 40 | 41 | 43 | | | | | | | | | |

| | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|--|--|--|--|
| 0.06666667 | 0.03333333 | 0.06666667 | 0.03333333 | 0.03333333 | 0.10000000 | 0.03333333 | | | | | |
| 0.03333333 | 0.03333333 | 0.03333333 | 0.03333333 | 0.03333333 | 0.03333333 | 0.03333333 | 0.03333333 | | | | |
| 0.03333333 | | | | | | | | | | | |

| | | | | | | |
|----|----|----|----|----|----|----|
| 44 | 45 | 46 | 47 | 48 | 49 | 50 |
|----|----|----|----|----|----|----|

```
0.03333333 0.06666667 0.03333333 0.13333333 0.03333333 0.03333333 0.03333333
```

Bang tan so tích luy

```
> cumsum(table(x))
```

```
20 21 23 24 28 30 31 32 33 34 35 36 40 41 43 44 45 46 47 48 49 50
```

```
2 3 5 6 7 10 11 12 13 14 15 16 17 18 19 20 22 23 27 28 29 30
```

Bang tan suat tích luy

```
> cumsum(prop.table(table(x)))
```

```
20    21    23    24    28    30    31    32    33    34    35    36
40    41    43
```

```
0.06666667 0.10000000 0.16666667 0.20000000 0.23333333 0.33333333 0.36666667
0.40000000 0.43333333 0.46666667 0.50000000 0.53333333 0.56666667 0.60000000
0.63333333
```

```
44    45    46    47    48    49    50
```

```
0.66666667 0.73333333 0.76666667 0.90000000 0.93333333 0.96666667 1.00000000
```

Bang tan so phan to

```
> x=c(48, 30, 35, 31, 21, 28, 34, 43, 36, 45, 41, 33, 47, 47, 30, 47, 44, 45, 32, 46, 47, 23, 30, 23,
49, 20, 24, 20, 40, 50)
```

```
> table(cut(x, breaks = 6))
```

```
(20,25] (25,30] (30,35] (35,40] (40,45] (45,50]
```

```
6    4    5    2    5    8
```

Dua gia tri min vao khoang dau

```
> table(cut(x, breaks =6,include.lowest=TRUE))
```

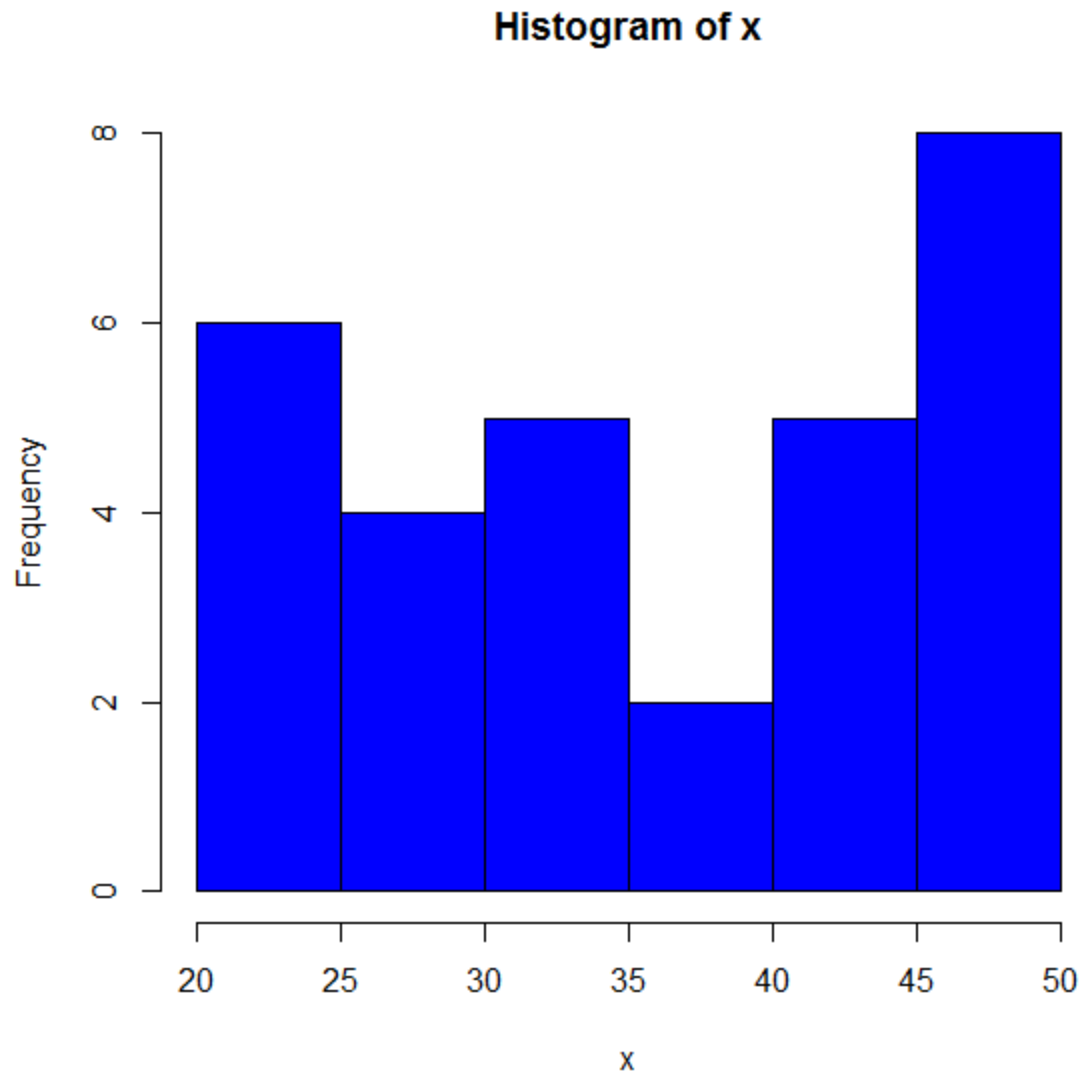
[20,25] (25,30] (30,35] (35,40] (40,45] (45,50]

6 4 5 2 5 8

Bieu do tan so

```
hist(x, breaks = "Sturges",  
     freq = NULL, probability = !freq,  
     include.lowest = TRUE, right = TRUE,  
     density = NULL, angle = 45, col = NULL, border = NULL,  
     main = paste("Histogram of" , xname),  
     xlim = range(breaks), ylim = NULL,  
     xlab = xname, ylab,  
     axes = TRUE, plot = TRUE, labels = FALSE,  
     nclass = NULL, warn.unused = TRUE, ...)
```

```
> hist(x, breaks = 6, include.lowest = TRUE,col="blue")
```



Da giac tan so

```
> BieuDo= hist(x, breaks = 6, include.lowest = TRUE,col="blue")
```

```
> HoanhDo=c(min(BieuDo$breaks),BieuDo$mids,max(BieuDo$breaks))
```

```
> TungDo=c(0,BieuDo$counts,0)
```

#Kiem tra lai

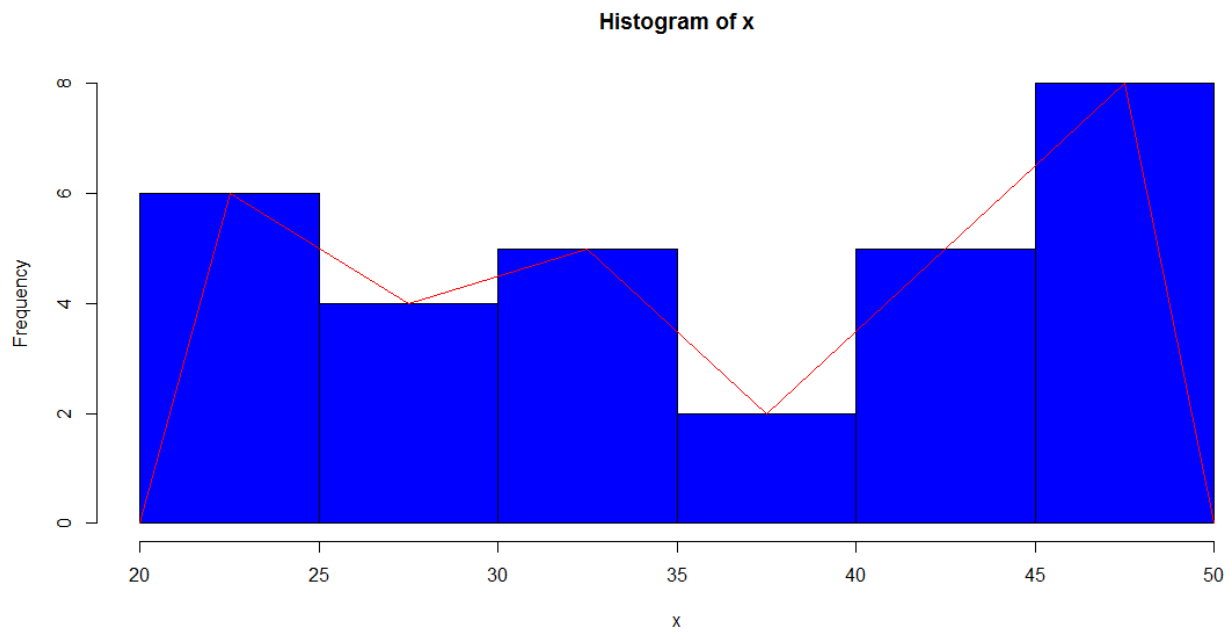
```
> HoanhDo
```

```
[1] 20.0 22.5 27.5 32.5 37.5 42.5 47.5 50.0
```

```
> TungDo
```

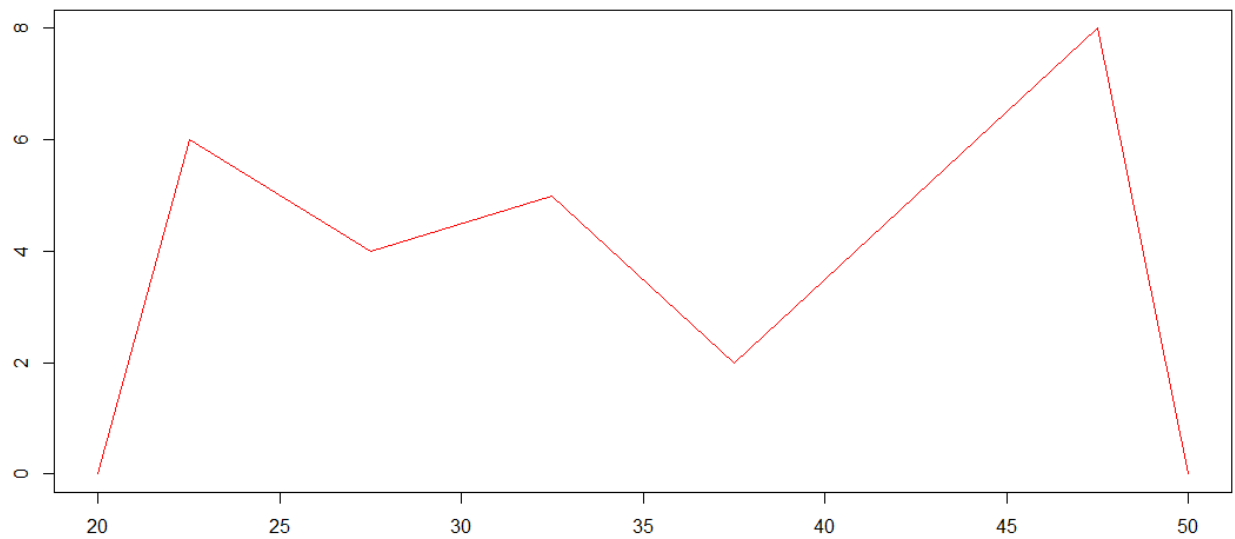
```
[1] 0 6 4 5 2 5 8 0
```

```
> lines(HoanhDo,TungDo,col="red")
```



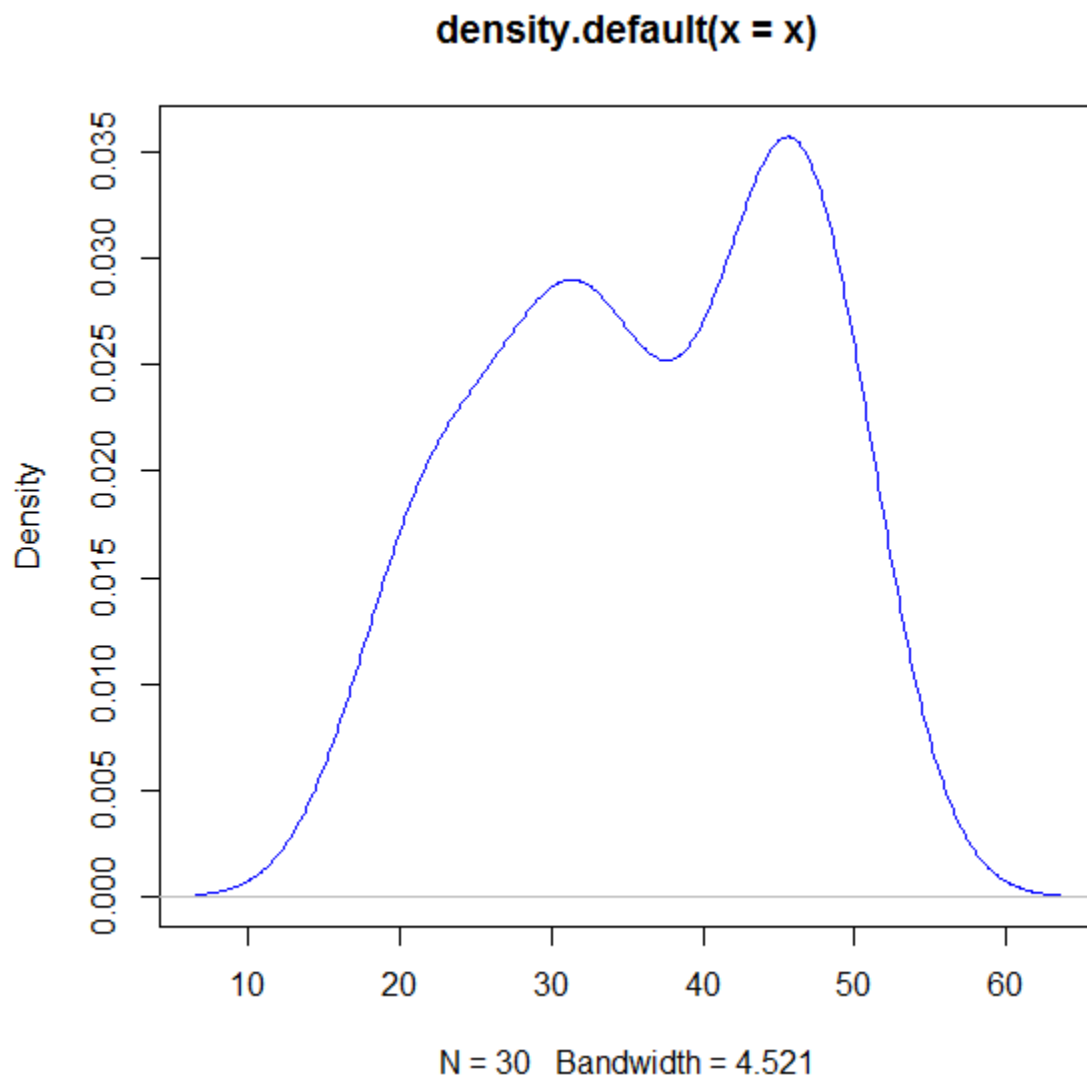
Ve rieng da giac tan so

```
> plot(HoanhDo,TungDo,type="l",col="red")
```



Do thi phan phoi xac suat

```
> plot(density(x), breaks = 6, include.lowest = TRUE,col="blue")
```



Bieu do than la

```
> stem(x,scale=1)
```

The decimal point is 1 digit(s) to the right of the |

2 | 001334

```

2 | 8
3 | 0001234
3 | 56
4 | 0134
4 | 556777789

```

```

5 | 0

```

Bieu do thanh, bieu do tron(tan suat)

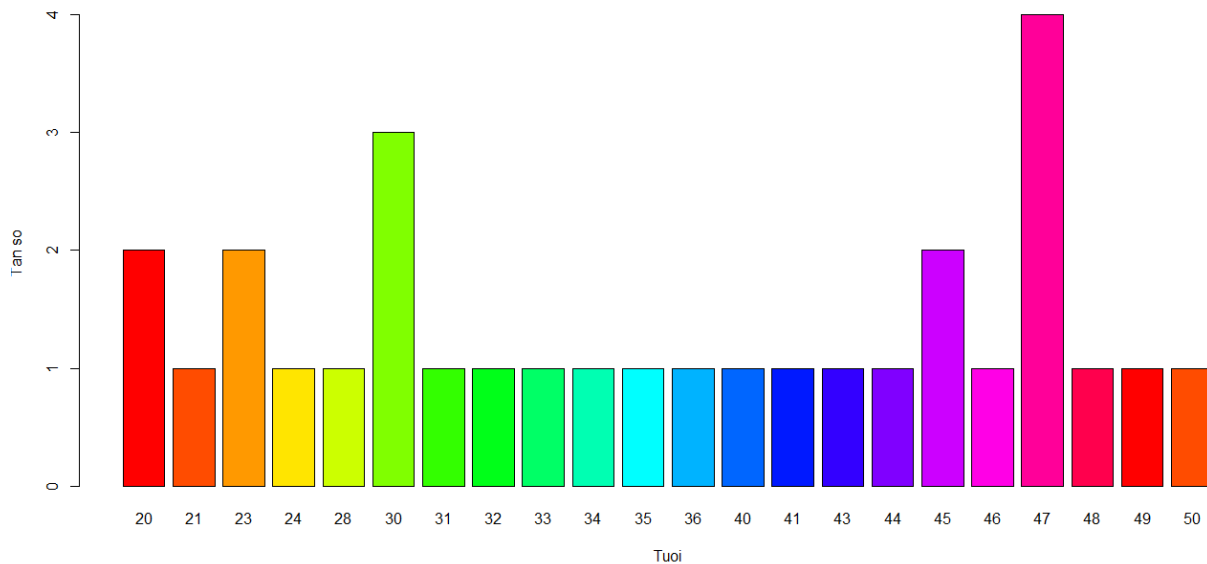
```
> table(x)
```

```
x
```

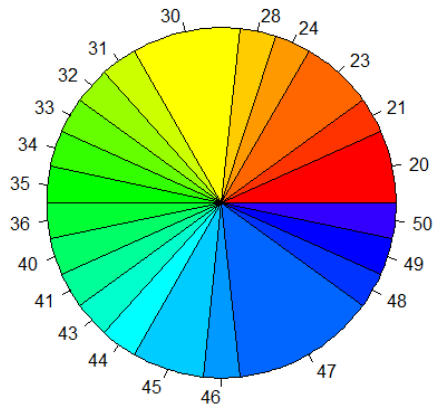
```
20 21 23 24 28 30 31 32 33 34 35 36 40 41 43 44 45 46 47 48 49 50
```

```
2 1 2 1 1 3 1 1 1 1 1 1 1 1 2 1 4 1 1 1
```

```
> barplot(table(x),beside=TRUE,col=rainbow(20),xlab="Tuoi",ylab="Tan so")
```



```
> pie(table(x),col=rainbow(30))
```



Bieu do tron, bieu do thanh

```
pie(x, labels = names(x), edges = 200, radius = 0.8,
    clockwise = FALSE, init.angle = if(clockwise) 90 else 0,
    density = NULL, angle = 45, col = NULL, border = NULL,
    lty = NULL, main = NULL, ...)

barplot(height, width = 1, space = NULL,
    names.arg = NULL, legend.text = NULL, beside = FALSE,
    horiz = FALSE, density = NULL, angle = 45,
    col = NULL, border = par("fg"),
    main = NULL, sub = NULL, xlab = NULL, ylab = NULL,
    xlim = NULL, ylim = NULL, xpd = TRUE, log = "",
    axes = TRUE, axisnames = TRUE,
    cex.axis = par("cex.axis"), cex.names = par("cex.axis"),
    inside = TRUE, plot = TRUE, axis.lty = 0, offset = 0,
    add = FALSE, args.legend = NULL, ...)
```


Ví dụ

Theo ước tính, lượng khách quốc tế đến Việt Nam vào tháng 2 năm 2015:

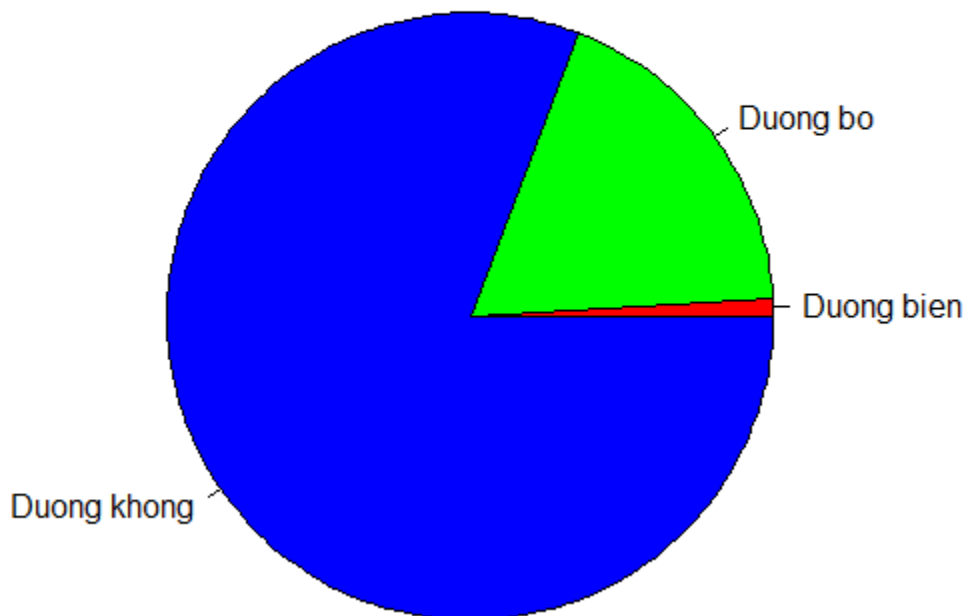
| Phương tiện | Lượng khách | Tỉ lệ | Góc (độ) |
|-------------|-------------|----------|----------|
| Đường biển | 7 021 | 0.009287 | 3.34 |
| Đường bộ | 138 146 | 0.182733 | 65.78 |
| Đường không | 610 834 | 0.80798 | 290.87 |

```
> TyLe=c(0.009287,0.182733,0.80798)
```

```
> names=c("Duong bien","Duong bo","Duong khong")
```

```
> pie(TyLe,names,col=rainbow(3),main="Luong khach QT den VN")
```

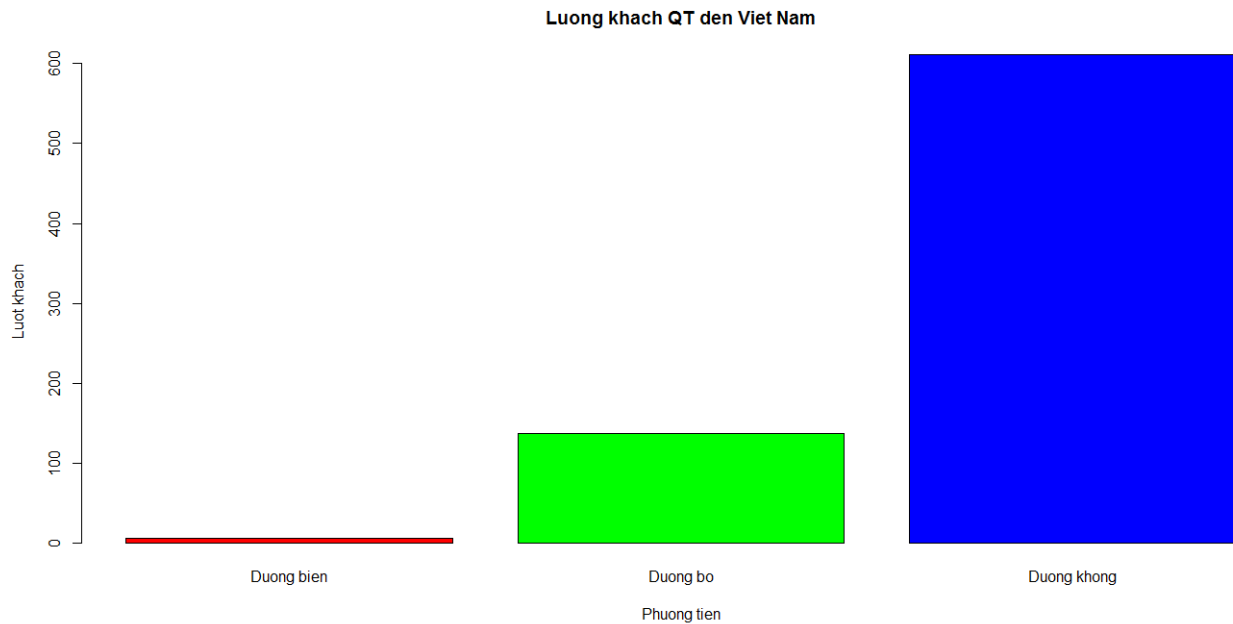
Luong khach QT den VN



Bieu do thanh

```
> LuongKhach=c(7.021,138.146,610.834)
```

```
> barplot(LuongKhach,names.arg=c("Duong bien","Duong bo","Duong khong"),main="Luong  
khach QT den Viet Nam",col=c("red","green","blue"),xlab="Phuong tien",ylab="Luot khach")
```



Bieu do tan xa

`plot(x, y, type = "p", main = "", xlab = "", ylab = "", bty = "l")`

Example

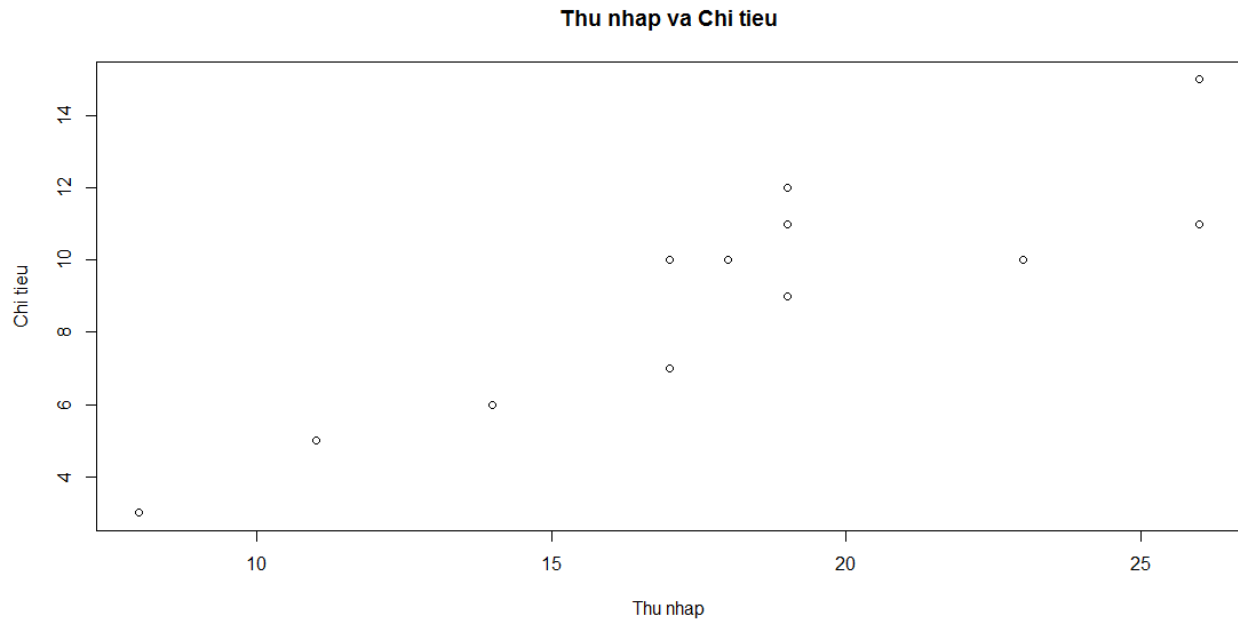
Giả sử ta có thu nhập và chi tiêu của 20 người như sau.

| TN | 19 | 17 | 18 | 23 | 14 | 11 | 17 | 26 | 19 | 19 | 8 | 26 |
|----|----|----|----|----|----|----|----|----|----|----|---|----|
| CT | 12 | 7 | 10 | 10 | 6 | 5 | 10 | 15 | 11 | 9 | 3 | 11 |

```
> a=c(19, 17, 18, 23, 14, 11, 17, 26, 19, 19, 8, 26)
```

```
> b=c(12, 7, 10, 10, 6, 5, 10, 15, 11, 9, 3, 11)
```

```
> plot(a, b, type = "p", main = "Thu nhap va Chi tieu", xlab = "Thu nhap", ylab = "Chi tieu")
```



Về biểu đồ tán so chéo

Ví dụ

Giả sử ta có bảng dữ liệu sau về giới tính và nghề nghiệp của 10 người được điều tra.

| TT | Giới tính | Nghề nghiệp |
|----|-----------|-------------|
| 1 | Nam | CongNhan |
| 2 | Nam | QuanLy |
| 3 | Nam | CongNhan |
| 4 | Nam | NVVP |
| 5 | Nam | NVVP |
| 6 | Nam | NVVP |
| 7 | Nu | QuanLy |
| 8 | Nam | CongNhan |
| 9 | Nu | NVVP |
| 10 | Nam | CongNhan |

```
> DS=edit(data.frame())
```

```
#Nhập bảng dữ liệu như trong Ví dụ vào R
```

```
#Kiểm tra lại bảng DS
```

```
> DS
```

```
TT GioiTinh NgheNghiep
```

```
1 1  Nam  CongNhan
2 2  Nam  QuanLy
3 3  Nam  CongNhan
4 4  Nam  NVVP
5 5  Nam  NVVP
6 6  Nam  NVVP
7 7  Nu   QuanLy
8 8  Nam  CongNhan
9 9  Nu   NVVP
10 10 Nam  CongNhan
```

```
> attach(DS)
```

```
> table(NgheNghiep,GioiTinh)
```

```
GioiTinh
```

```
NgheNghiep Nam Nu
```

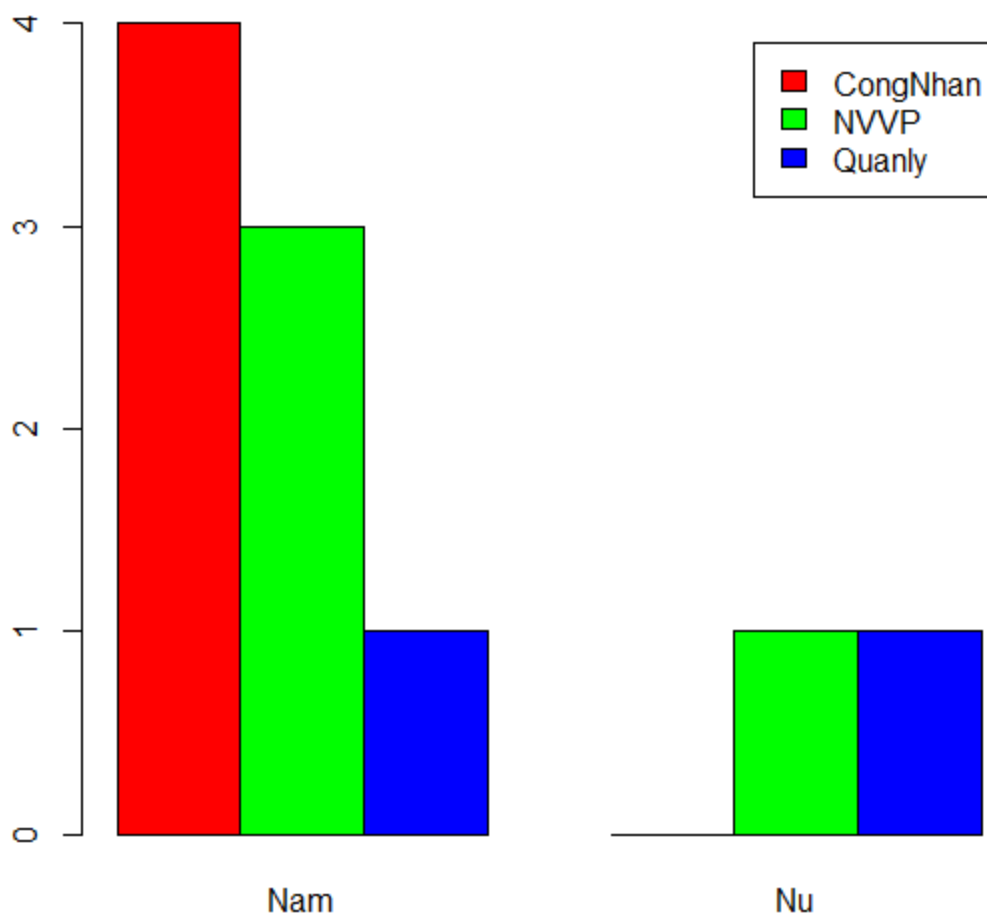
```
CongNhan  4 0
```

```
NVVP      3 1
```

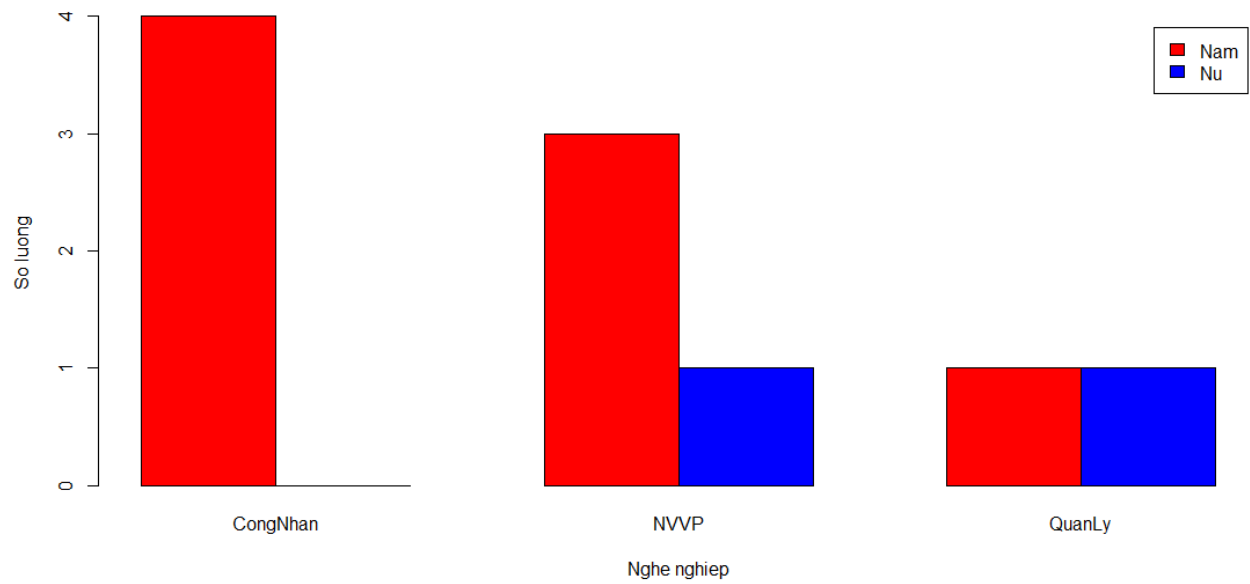
```
QuanLy    1 1
```

```
> barplot(table(NgheNghiep,GioiTinh),beside=TRUE,col=rainbow(8))
```

```
> barplot(table(NgheNghiep,GioiTinh),legend.text=c("CongNhan","NVVP","Quanly"),beside=TRUE,col=c("red","green","blue"))
```



```
>barplot(table(GioiTinh,NgheNghiep),legend.text=c("Nam","Nu"),beside=TRUE,col=c("red","blue"),xlab="Nghe nghiep",ylab="So luong")
```



Ví dụ: Vẽ biểu đồ thanh cho tập dữ liệu

120, 150, 125, 100, 180, 140, 200

HD:

```
> x=c(120, 150, 125, 100, 180, 140, 200)
> barplot(x,col = rainbow(7))
```

