V½ dö **BÀI TẬP THỐNG KÊ ỨNG DỤNG**

**Buổi 1**

1. **Cài phần mềm R và sử dụng các chức năng cơ bản.**

+ Gõ từ khóa tìm kiếm “cran r” và “rstudio” trên google. Cài đặt cả 2 phần mềm R và RStudio (bản RStudio Desktop) tải về từ các website

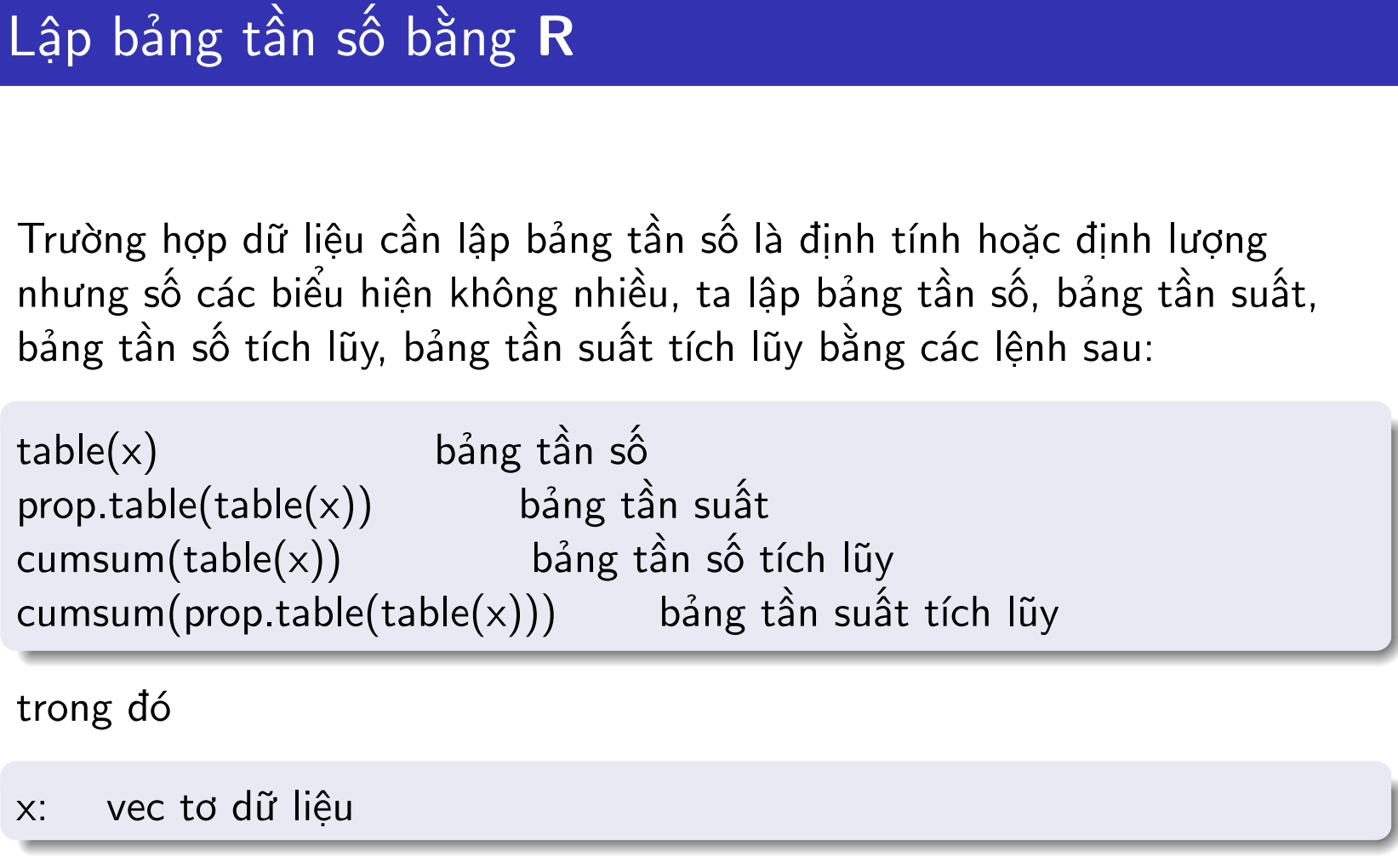
<https://cran.r-project.org/> và <https://www.rstudio.com/products/rstudio/download/>

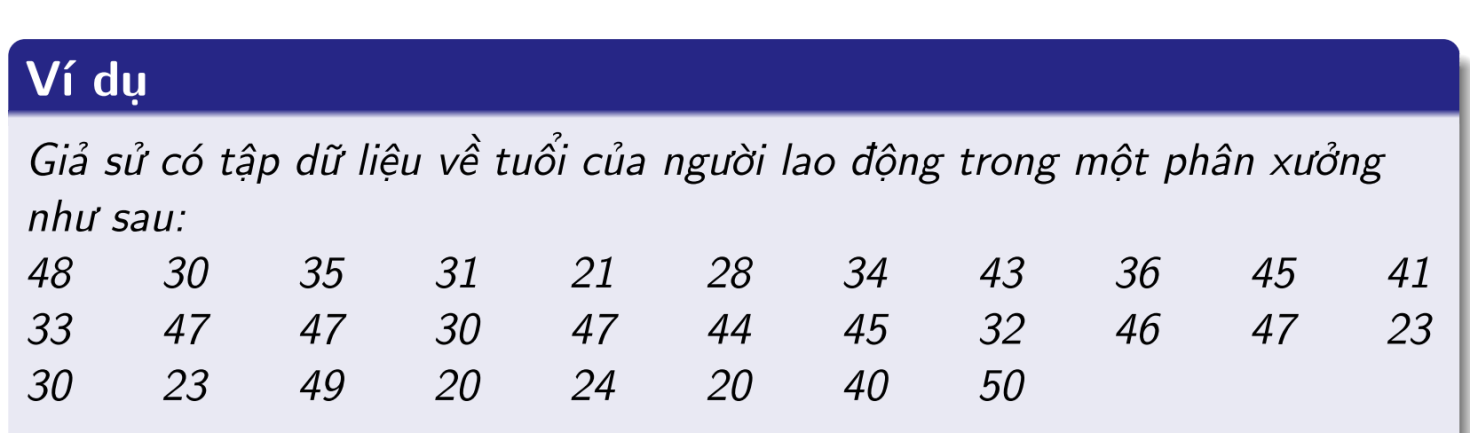
+ Cài đặt thư viện BSDA: Kết nối Internet, mở cửa sổ giao diện Rstudio gõ lệnh

> install.packages("BSDA")

Bấm Enter

1. **Lập các bảng tần số bằng R.**
2. **Vẽ các loại đồ thị bằng R.**





***(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)***

***Yêu cầu:***

***1) Lập bảng tần số, bảng tần suất, bảng tần số tích lũy, tần suất tích lũy. Vẽ biểu đồ thanh(biểu đồ tần số), biểu đồ tròn(biểu đồ tần suất), biểu đồ thân lá của dữ liệu.***

***2) Phân thành 6 tổ, lập bảng tần số dữ liệu phân tổ(đưa giá trị min vào khoảng đầu), vẽ biểu đồ tần số dữ liệu phân tổ, vẽ đa giác tần số dữ liệu phân tổ, vẽ đồ thị phân phối xác suất dữ liệu phân tổ***

**Hướng dẫn:Copy paste các lệnh màu vàng vào RStudio và bấm Enter.**

**1)**

> library(BSDA)

**Nhập dữ liệu**

**Cách 1:**

> 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)

**Cách 2:** Với mẫu dữ liệu không có dấu phẩy: 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

Copy paste và dùng lệnh scan() như sau:

|  |
| --- |
| Bấm 2 lần Enter |
|  |
| |  | | --- | |  | |

> x=scan()

1: 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

31:

Read 30 items

Bang tan so cua 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 1 2 1 4 1 1 1

---------------------------------------

Bang tan suat

> 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

44 45 46 47 48 49 50

0.03333333 0.06666667 0.03333333 0.13333333 0.03333333 0.03333333 0.03333333

---------------------------------------

Bang tan so tich 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 tich 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

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 (tan so), bieu do tron(tan suat)

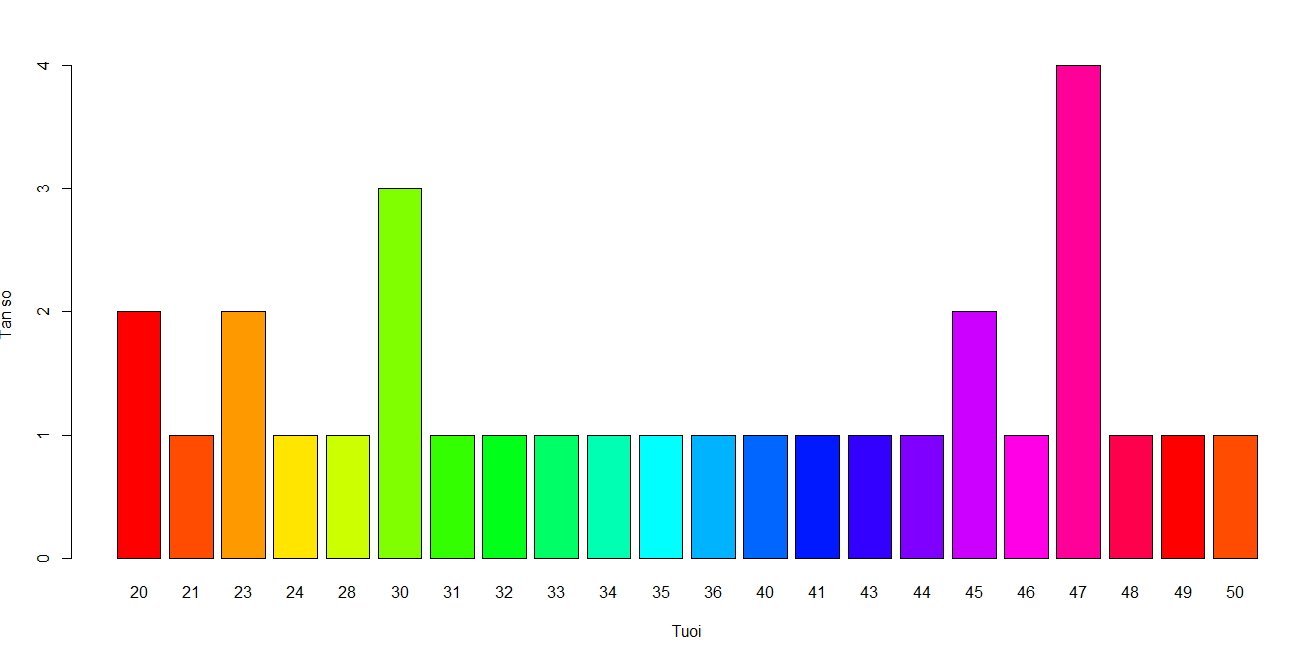
> table(x)

x

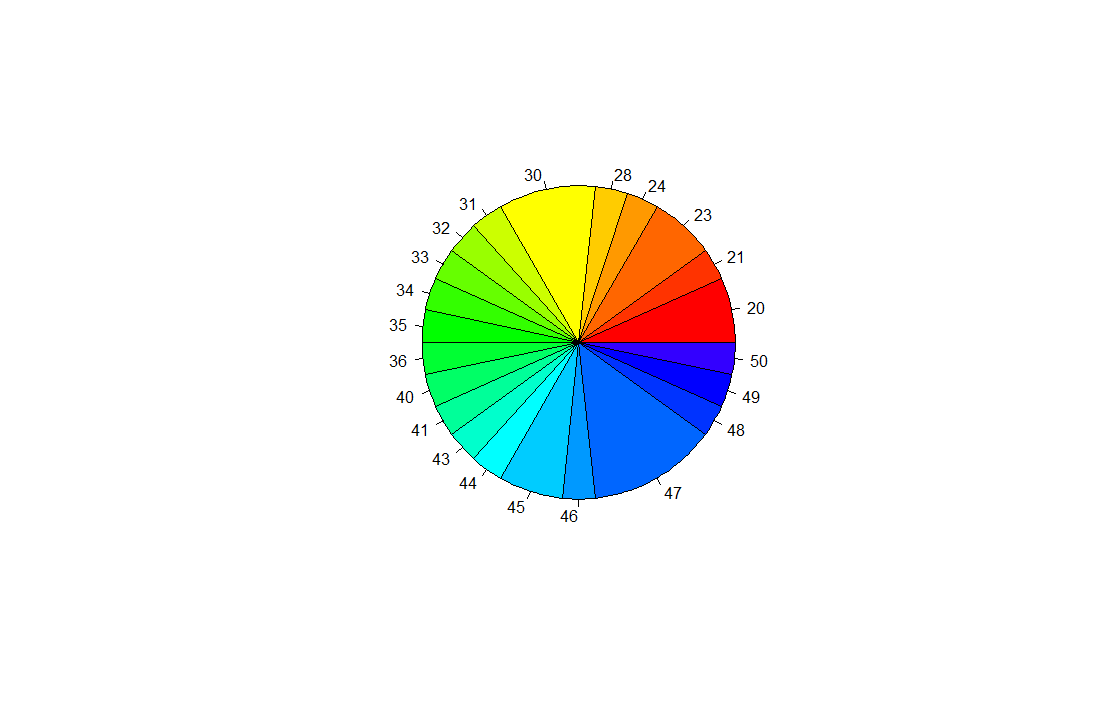
20 21 23 24 28 30 31 32 33 34 35 36 40 41 43 44 45 46 47 48 49 50

2 1 2 1 1 3 1 1 1 1 1 1 1 1 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, ...)

**2)**

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),

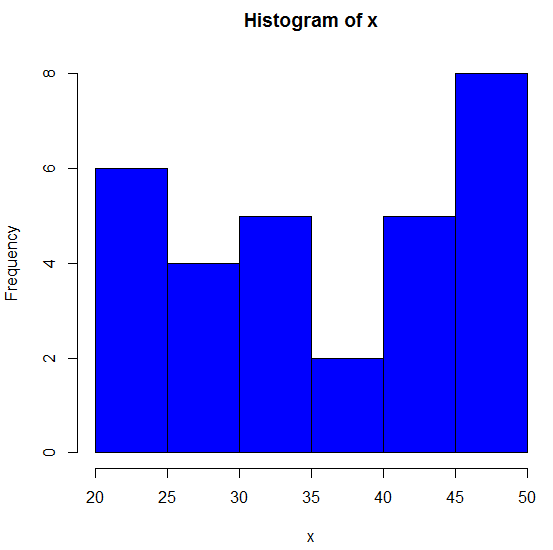
xlim = range(breaks), ylim = NULL,

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

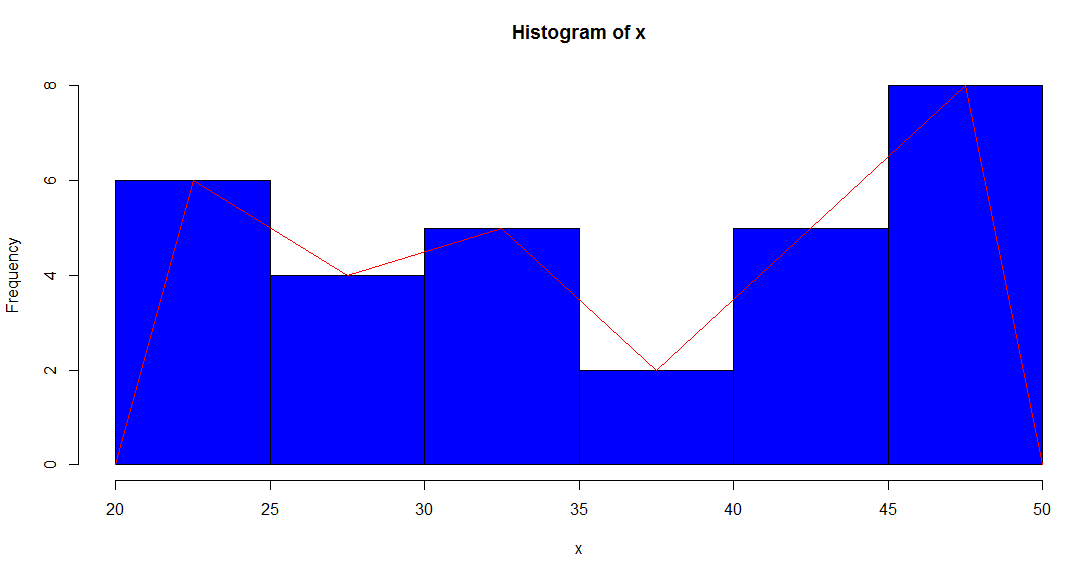
> HoanhDo

[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

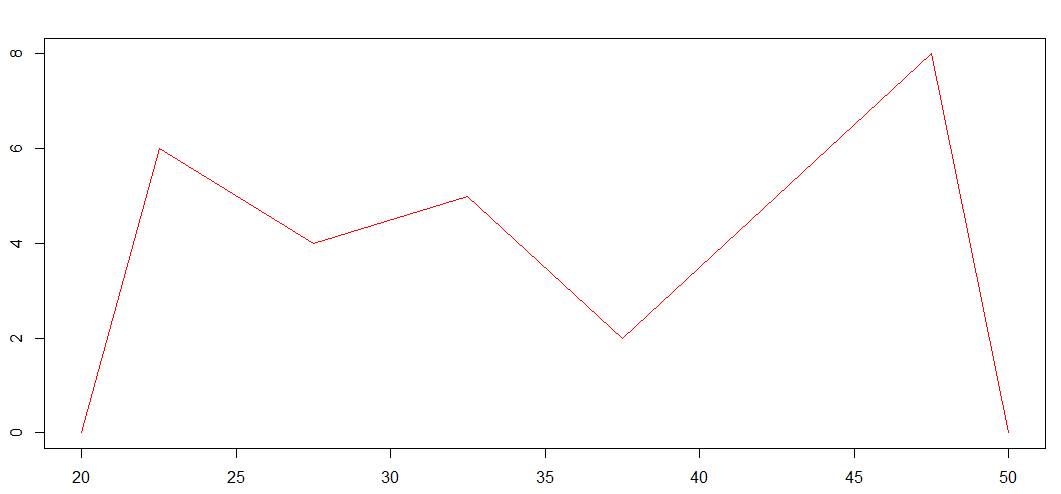
> lines(HoanhDo,TungDo,col="red")



---------------------

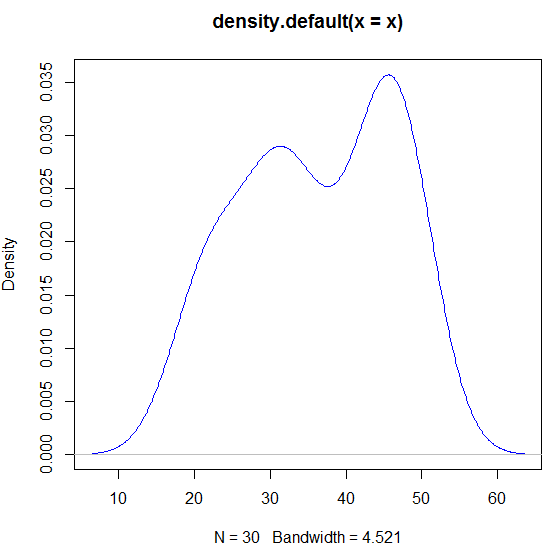
Ve rieng da giac tan so

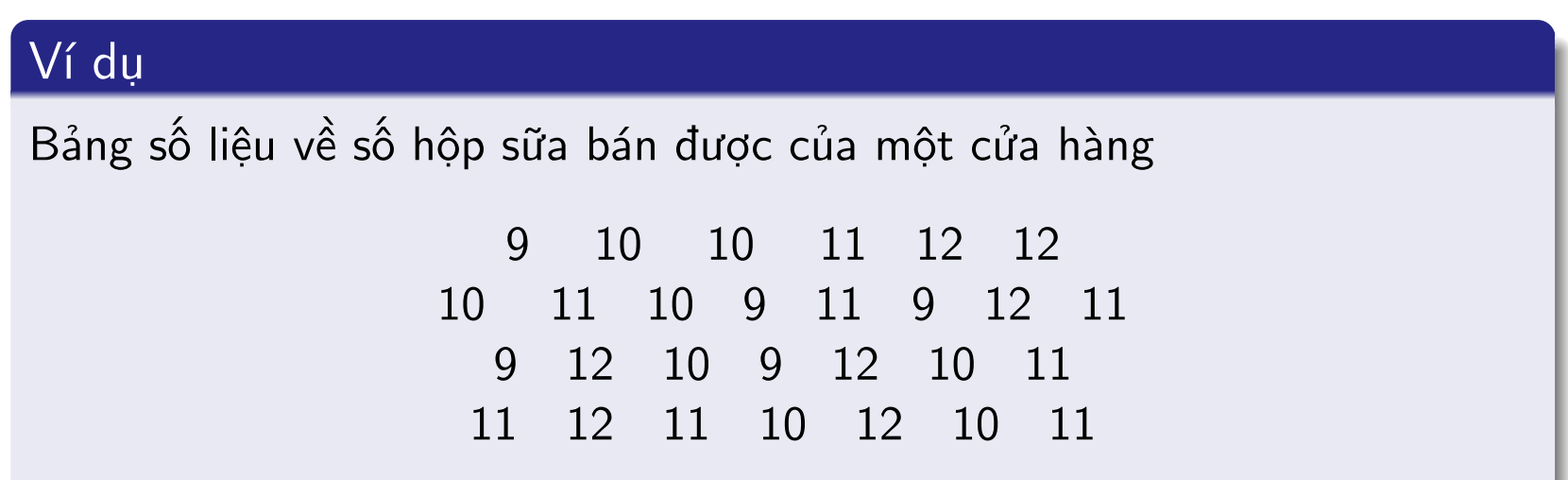
> plot(HoanhDo,TungDo,type="l",col="red")

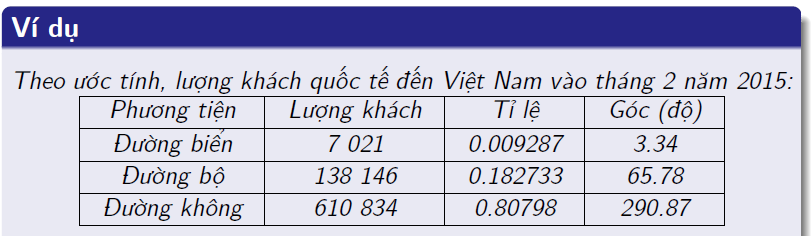


Do thi phan phoi xac suat

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



***Yêu cầu: câu hỏi như ví dụ trên.***

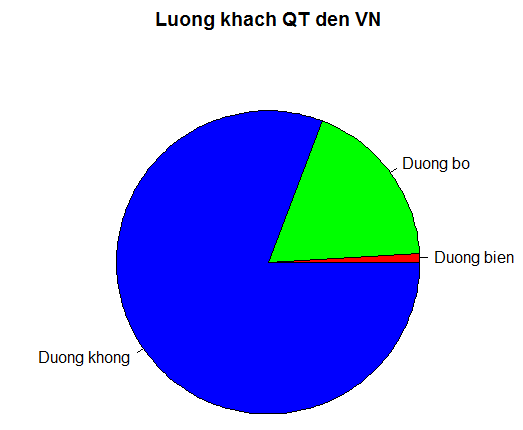


***Yêu cầu: Vẽ biểu đồ hình tròn + biểu đồ thanh***

> TyLe=c(0.009287,0.182733,0.80798)

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

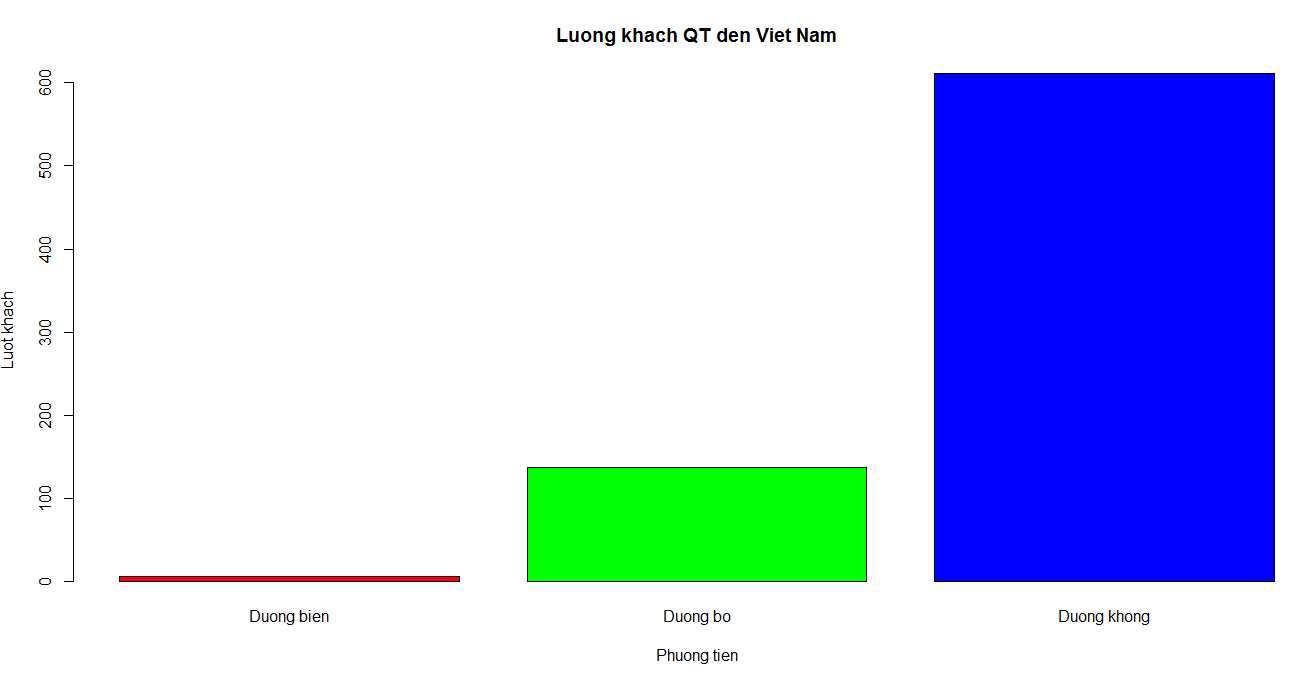
>pie(TyLe,names,col=rainbow(3),main="Luong khach QT den VN")

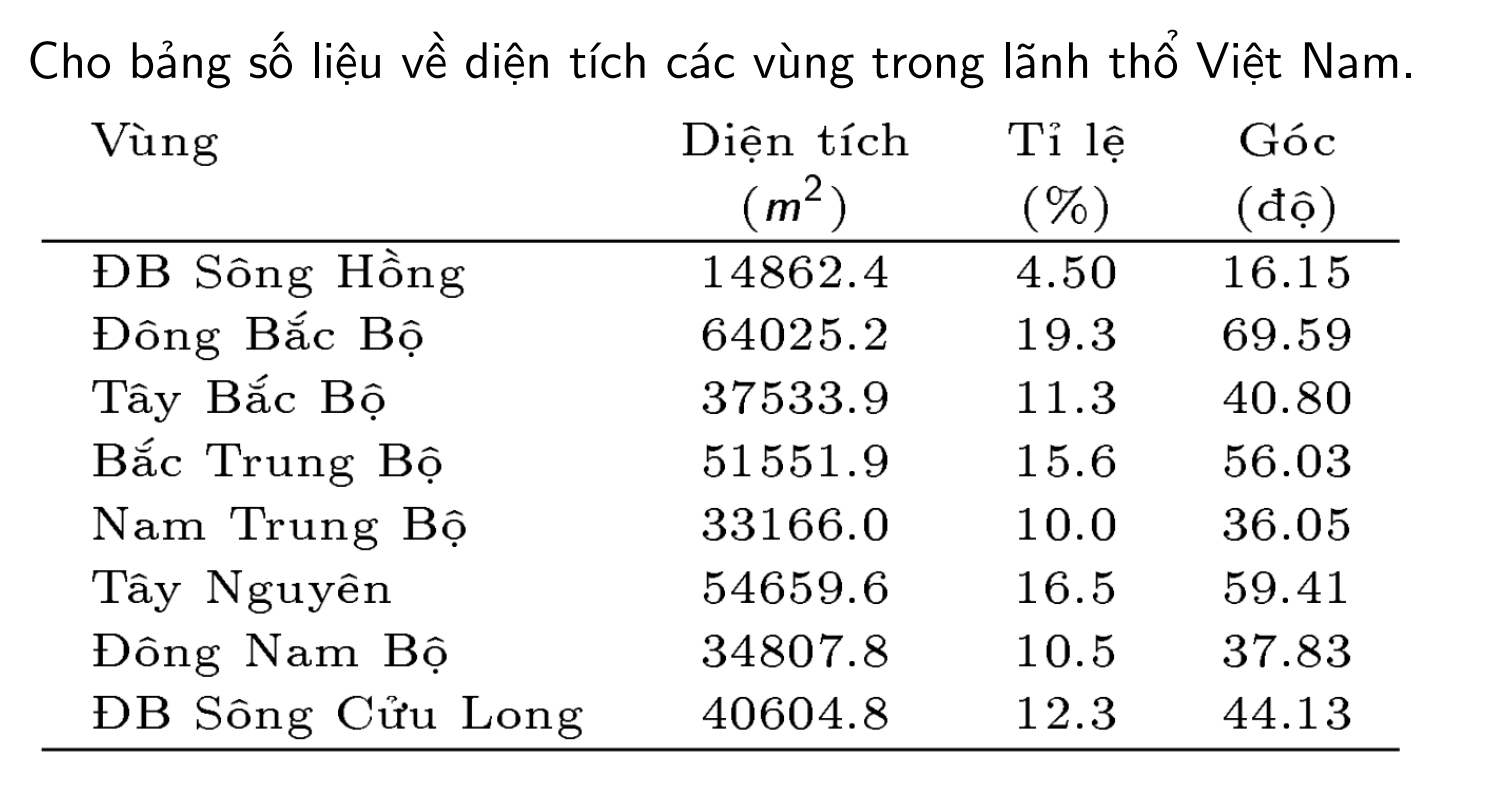


Bieu do thanh

> 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")



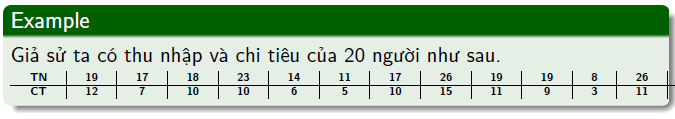
**BT thực hành**

***Yêu cầu: Vẽ biểu đồ hình tròn + biểu đồ thanh***

Bieu do tan xa

H m plot()

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

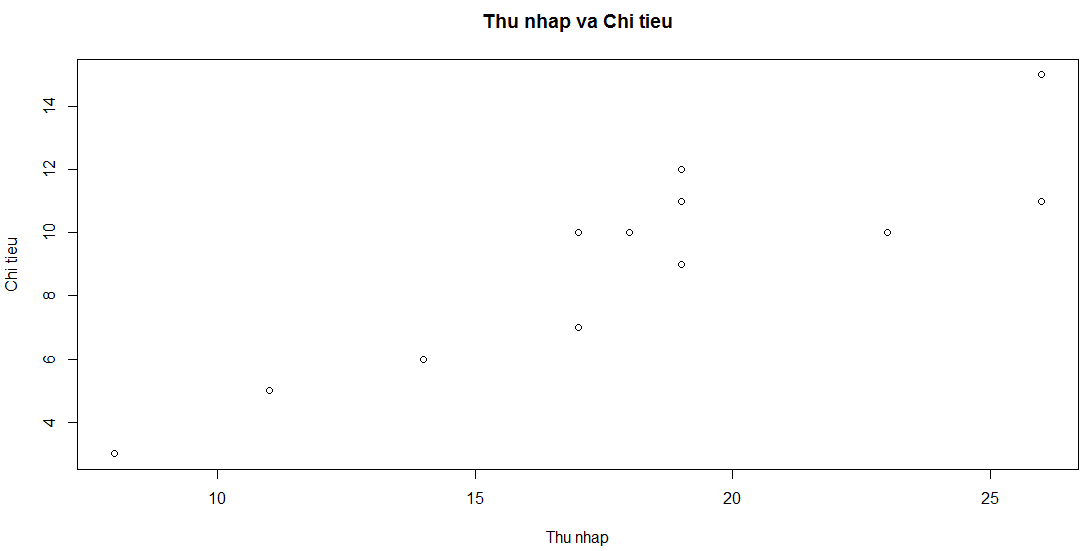


\_É L ***Yêu cầu: Vẽ biểu đồ tán xạ***

> 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")



Ve bieu do tan so cheo



***Yêu cầu: Nhập dữ liệu cho bảng trên và vẽ các biểu đồ tần số chéo***

> DS=edit(data.frame())

#Nhap bang du lieu nhu trong Vi du vao R

#Kiem tra lai bang 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

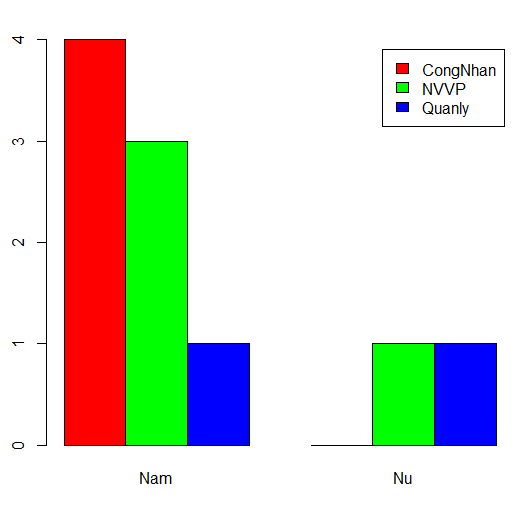
CongNhan 4 0

NVVP 3 1

QuanLy 1 1

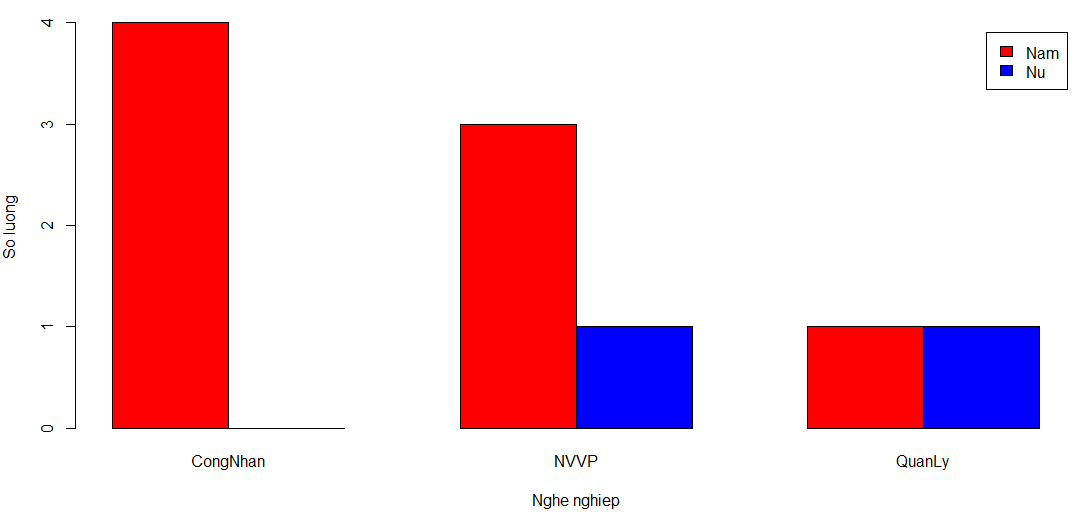
>barplot(table(NgheNghiep,GioiTinh),beside=TRUE,col=rainbow(8))

>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")



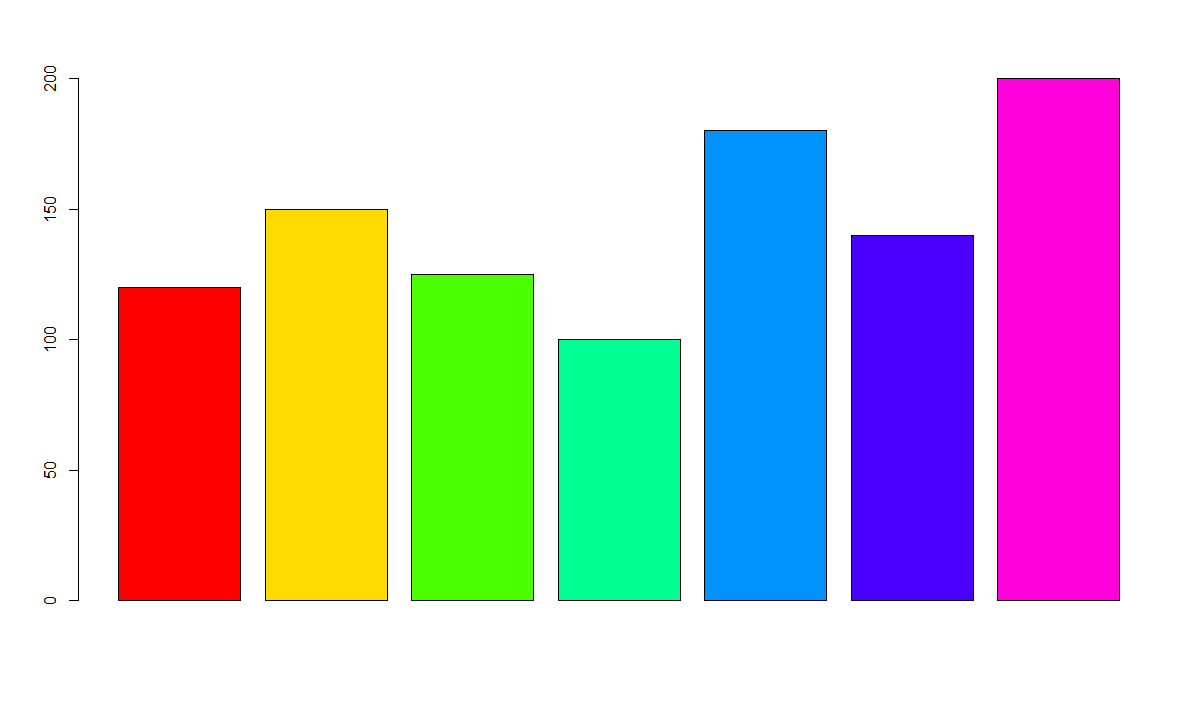
**BT:** Vẽ biểu đồ thanh cho tập dữ liệu

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

HD:

> x=c(120, 150, 125, 100, 180, 140, 200)

> barplot(x,col = rainbow(7))



**BÀI TẬP ÁP DỤNG**

1. **Cho file excel điểm thi của một môn học: “KT1.csv”. Thống kê số lượng sinh viên có ĐQT đạt loại A, B, C, D trong file dữ liệu trên và vẽ các loại biểu đồ minh họa.**

#Cách 1: Đọc dữ liệu từ file excel

Mở file excel KT1.xls, chọn “Save As”, “Save as type” chọn lưu file định dạng đuôi .CSV(comma delimited).

Chạy lệnh

> getwd()

[1] "C:/Users/UserName/Documents"

Copy file KT1.csv vào thư mục R đang làm việc theo đường dẫn trên (mặc định là Documents)

Mở file KT1.csv xem cấu trúc file.

Chạy các lệnh sau

> KT1=read.csv("KT1.csv")

> c=cut(KT1$DQT,breaks = c(0,3.9,5.4,6.9,8.4,10),include.lowest = T,labels = c("F","D","C","B","A"))

> table(c)

c

F D C B A

5 12 37 59 30

**Hoặc có thể dùng lệnh attach như sau**

> Diem=read.csv("KT1.csv")

> attach(Diem)

> x=DQT

> c=cut(x,breaks = c(0,3.9,5.4,6.9,8.4,10),include.lowest = T,labels = c("F","D","C","B","A"))

> table(c)

c

F D C B A

5 12 37 59 30

Số SV đạt điểm A là 30,…

> prop.table(table(c))

c

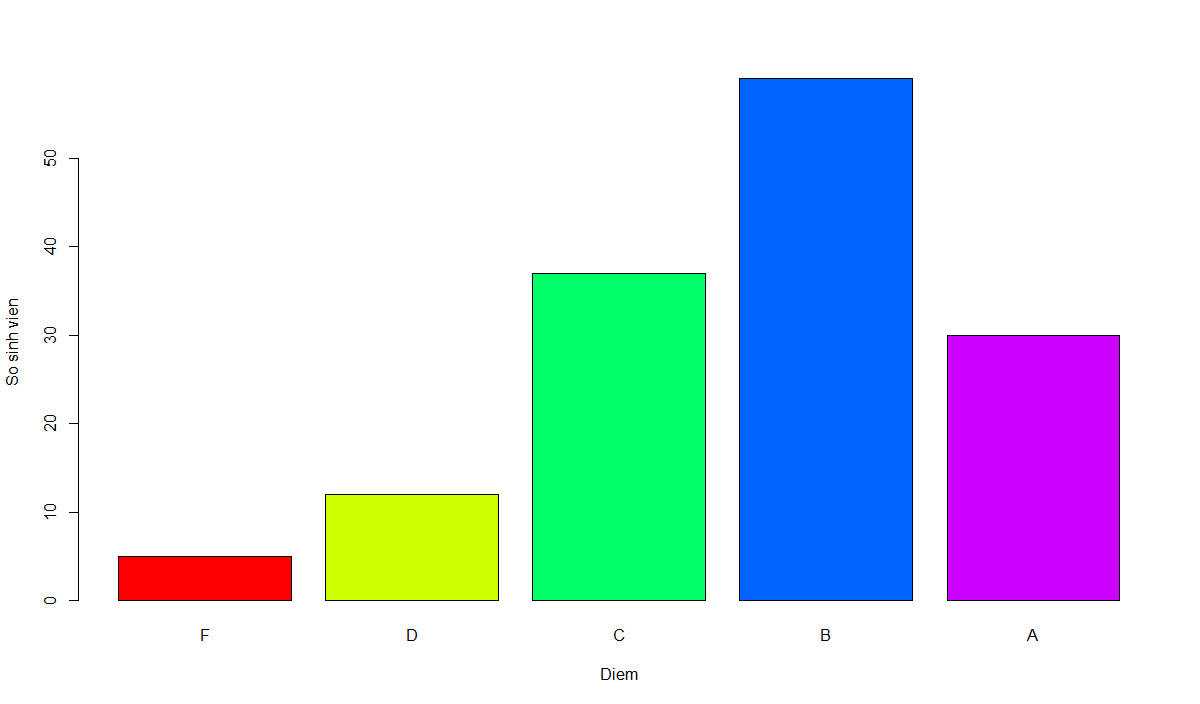
F D C B A

0.03496503 0.08391608 0.25874126 0.41258741 0.20979021

Tỷ lệ điểm A là 0.20979021,...

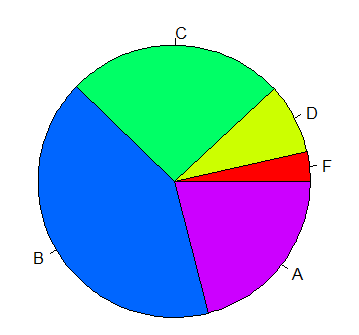
# Biểu đồ cột phổ điểm

> barplot(table(c),beside=TRUE,col=rainbow(5),xlab="Diem",ylab="So sinh vien")



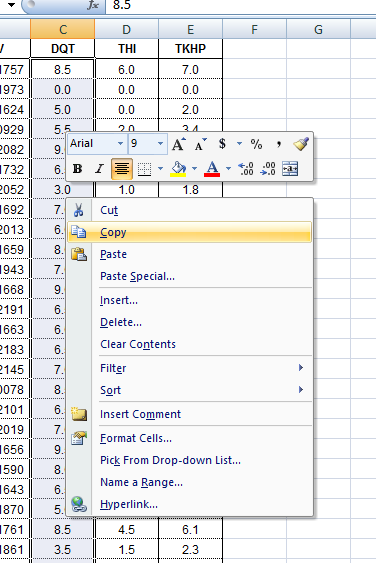
# Biểu đồ tròn tần suất

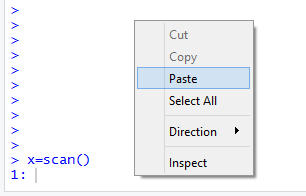
> pie(table(c),col=rainbow(5))



#Cách 2: Copy, Paste trực tiếp điểm quá trình vào R và bấm Enter

> x=scan()





> x=scan()

1: 8.5

2: 0.0

3: 5.0

4: 5.5

5: 9.0

6: 6.5

7: 3.0

8: 7.0

9: 6.0

10: 8.0

11: 7.0

12: 9.0

13: 6.5

14: 6.0

15: 6.5

16: 7.0

17: 8.5

18: 6.5

19: 7.0

20: 9.5

21: 8.0

22: 6.5

23: 5.0

24: 8.5

25: 3.5

26: 7.0

27: 9.5

28: 8.0

29: 7.0

30: 9.0

31: 9.5

32: 9.0

33: 6.0

34: 5.5

35: 6.5

36: 5.5

37: 9.5

38: 4.0

39: 8.5

40: 8.0

41: 7.5

42: 5.5

43: 9.0

44: 7.0

45: 7.0

46: 8.0

47: 7.5

48: 6.0

49: 5.5

50: 7.0

51: 5.5

52: 5.0

53: 0.0

54: 7.5

55: 8.0

56: 6.5

57: 6.0

58: 5.5

59: 5.5

60: 8.5

61: 7.5

62: 7.0

63: 10.0

64: 6.0

65: 9.0

66: 4.0

67: 9.0

68: 8.0

69: 10.0

70: 7.0

71: 6.5

72: 10.0

73: 9.0

74: 8.0

75: 7.0

76: 7.5

77: 6.5

78: 6.5

79: 4.0

80: 8.0

81: 6.0

82: 5.5

83: 8.0

84: 7.0

85: 7.5

86: 6.5

87: 8.5

88: 7.5

89: 8.0

90: 7.5

91: 8.0

92: 7.0

93: 5.0

94: 7.5

95: 6.0

96: 8.0

97: 7.0

98: 7.5

99: 8.0

100: 7.0

101: 7.5

102: 10.0

103: 7.5

104: 6.5

105: 8.0

106: 9.0

107: 8.5

108: 7.0

109: 4.0

110: 5.5

111: 4.5

112: 9.0

113: 6.5

114: 7.0

115: 4.5

116: 7.0

117: 9.5

118: 9.0

119: 7.5

120: 7.5

121: 8.5

122: 4.5

123: 8.0

124: 8.0

125: 6.5

126: 9.0

127: 6.5

128: 5.0

129: 8.5

130: 7.5

131: 6.0

132: 7.0

133: 6.0

134: 5.5

135: 7.5

136: 0.0

137: 7.5

138: 7.0

139: 7.5

140: 6.5

141: 7.0

142: 7.5

143: 7.5

144:

Read 143 items

> c=cut(x,breaks = c(0,3.9,5.4,6.9,8.4,10),include.lowest = T,labels = c("F","D","C","B","A"))

> table(c)

c

F D C B A

5 12 37 59 30

> prop.table(table(c))

c

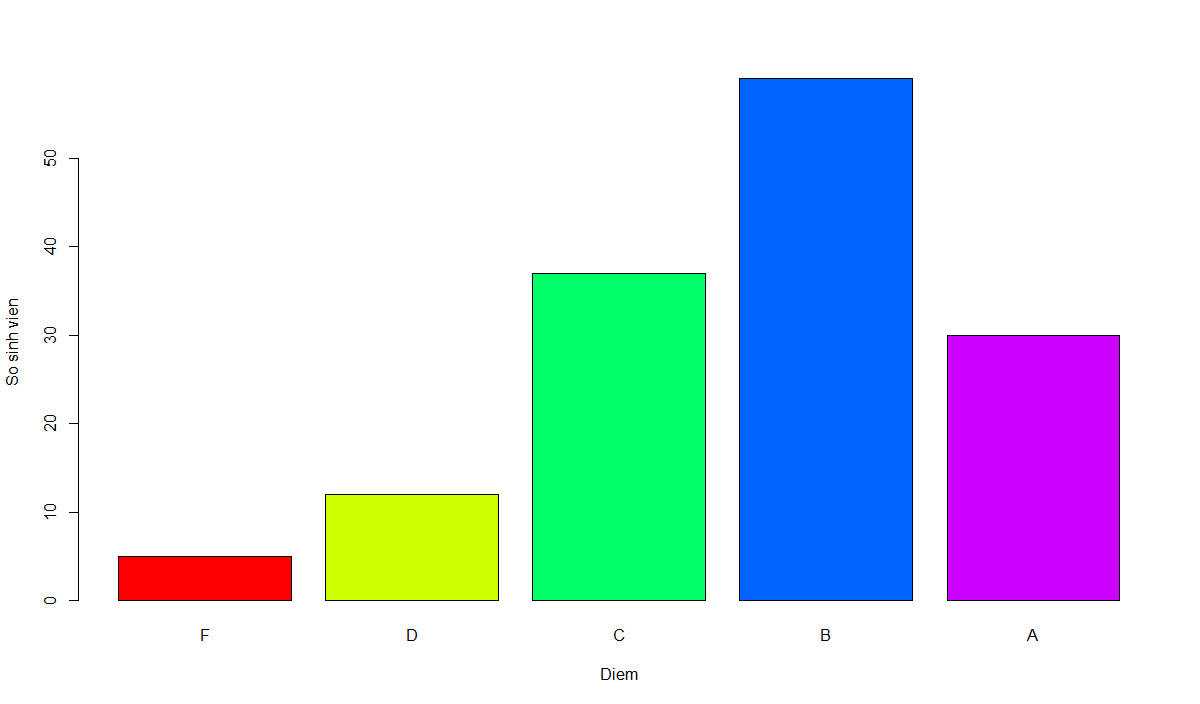
F D C B A

0.03496503 0.08391608 0.25874126 0.41258741 0.20979021

############################################################################

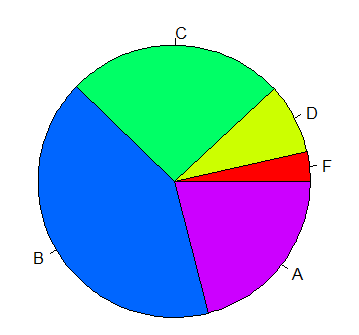
# Biểu đồ cột phổ điểm

> barplot(table(c),beside=TRUE,col=rainbow(5),xlab="Diem",ylab="So sinh vien")



# Biểu đồ tròn tần suất

> pie(table(c),col=rainbow(5))



1. **Cho dữ liệu hoa iris file iris.csv gồm độ dài đài hoa (Sepal.Length), độ rộng đài hoa(Sepal.Width), độ dài cánh hoa (Petal.Length), độ rộng cánh hoa(Petal.Width) chia thành 3 loài(Species).**

**Vẽ đồng thời biểu đồ tần suất histogram 4 yếu tố trên trong cùng một cửa sổ.**

> getwd()

[1] "C:/Users/Phuong/Documents"

#Như vậy R đang làm việc với thư mục "C:/Users/Phuong/Documents"

#Copy file iris.csv vào thư mục có đường dẫn trên (mặc định là Documents)

**Chạy các lệnh sau:**

> iris=read.csv("iris.csv")

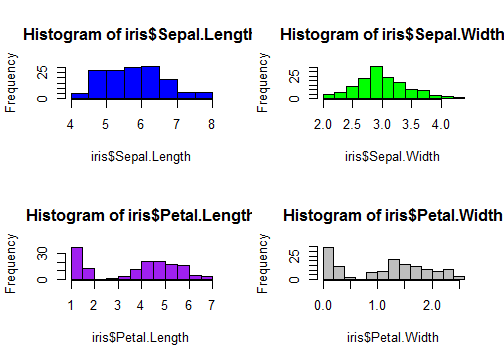
> par(mfrow=c(2,2))

> hist(iris$Sepal.Length, col="blue")

> hist(iris$Sepal.Width, col="green")

> hist(iris$Petal.Length, col="purple")

> hist(iris$Petal.Width, col="grey")



**BÀI TẬP LUYỆN TẬP**

V½ döf

1. ***Cho tập dữ liệu về tuổi người lao động tại một công ty:***

***23 45 23 34 45 57 25 34 32 31 30 45 23 34 45 57 25 34 32 31 34 32 31 30 45 23 34 45 57 25 34 32 23 45 23 34 45 32 31 30 45 23 34 45***

***Yêu cầu:***

***1) Lập bảng tần số, bảng tần suất, bảng tần số tích lũy, tần suất tích lũy. Vẽ biểu đồ thanh(biểu đồ tần số), biểu đồ tròn(biểu đồ tần suất), biểu đồ thân lá của dữ liệu.***

***2) Phân thành 5 tổ, lập bảng tần số dữ liệu phân tổ(đưa giá trị min vào khoảng đầu), vẽ biểu đồ tần số dữ liệu phân tổ, vẽ đa giác tần số dữ liệu phân tổ, vẽ đồ thị phân phối xác suất dữ liệu phân tổ***

***2. Lượng khách quốc tế đến VN trong 1 năm***

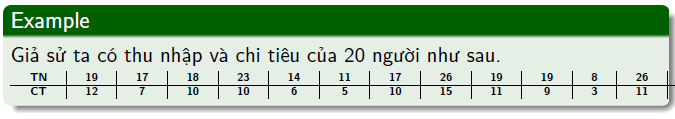
|  |  |  |
| --- | --- | --- |
| ***Phương tiện*** | ***Lượng khách*** | ***Tỷ lệ*** |
| ***Đường biển*** | ***8023*** | **0.009649013** |
| ***Đường bộ*** | ***136245*** | **0.1638576** |
| ***Đường không*** | ***687216*** | **0.8264934** |

***Yêu cầu:***

***Vẽ biểu đồ hình tròn + biểu đồ thanh***

***3.*** > a=c(19, 17, 18, 23, 14, 11, 17, 26, 19, 19, 8, 26, 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, 12, 7, 10, 10, 6, 5, 10, 15, 11, 9, 3, 11)



\_É L ***Yêu cầu: Vẽ biểu đồ tán xạ***

**4. Cho file excel điểm thi của một môn học: “KT.xls”. Thống kê số lượng sinh viên có điểm TKHP đạt loại A, B, C, D trong file dữ liệu trên và vẽ các loại biểu đồ thanh và tròn để minh họa.**

***5.***

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

***Yêu cầu: Nhập dữ liệu cho bảng trên và vẽ các biểu đồ tần số chéo***