# Chapter 11: Visualization

## Exercise 1: Painters school – PIE

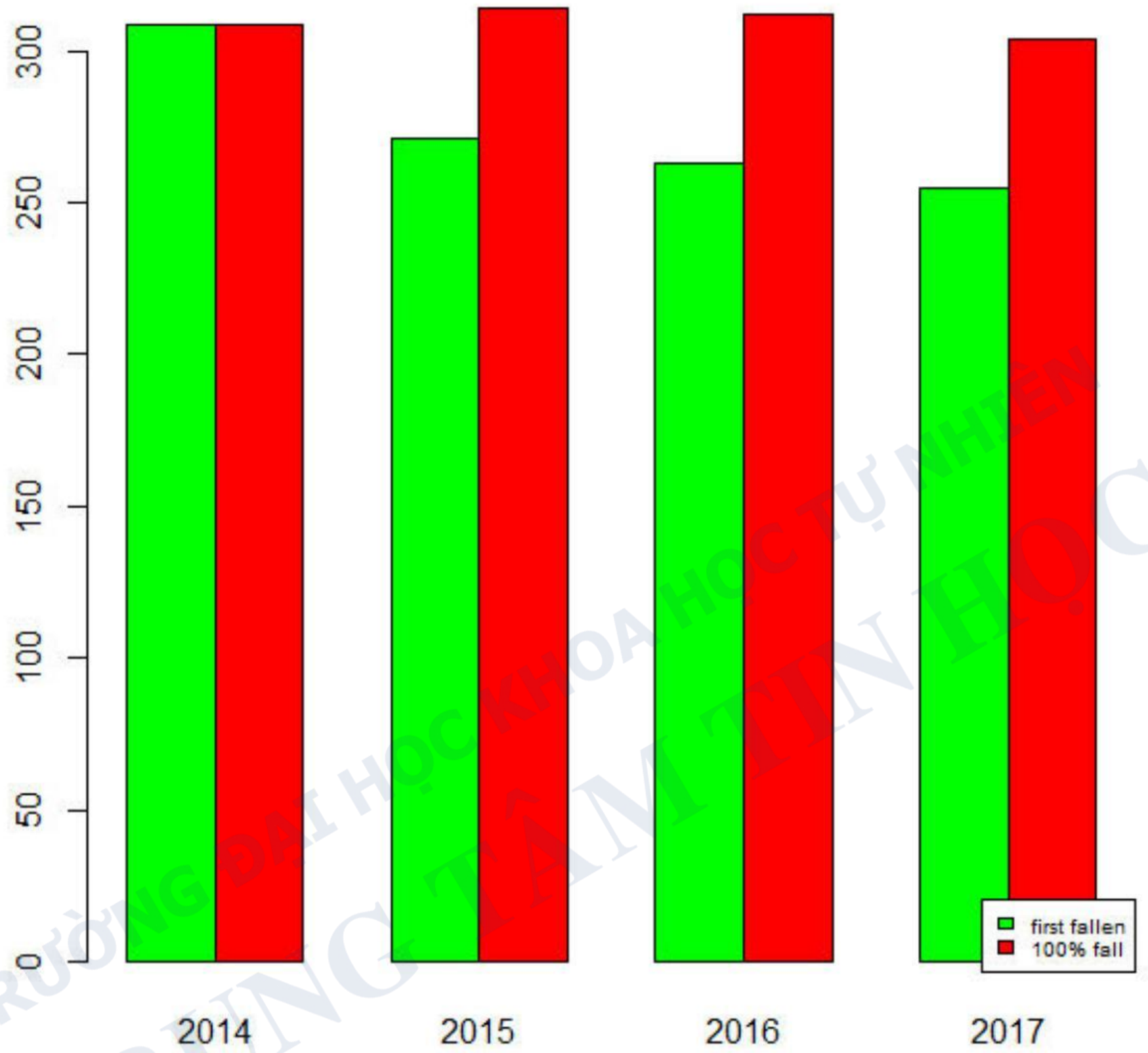
- Sử dụng thư viện MASS
- Trong MASS có dữ liệu painters
- Đọc dữ liệu painters => cho biết kiểu dữ liệu, số dòng, số cột
- Lấy dữ liệu school trong painters => cho biết kiểu dữ liệu school (Factor)
- Chuyển school thành table (dùng phương thức table(school)). In kết quả
- Vẽ pie chart dựa trên kết quả này
- Vẽ pie chart với vector màu tùy chọn

## Exercise 2: PIE 3D

- Sử dụng thư viện library("plotrix") để vẽ pie 3D
- Tạo dữ liệu như sau: Tree_species <- c("Beech", "Chestnut", "Hawthorn", "Red maple", "Witch Hazel", "Yellow Birch"), number_of_trees <- c(2, 2, 1, 4, 1, 1)
- Tính pie percent
- Vẽ pie 3D dựa trên dữ liệu phía trên
- Lấy nhãn hiển thị từng phần là số cây, màu cầu vồng, tên là "Number of Trees of each Species"
- Tạo ghi chú cho pie 3D ở topright, lấy tên cây, màu cầu vồng
- Vẽ barchart cho dữ liệu trên với thông tin trên trục hoành là Tree_species, trên trục tung là number_of_trees

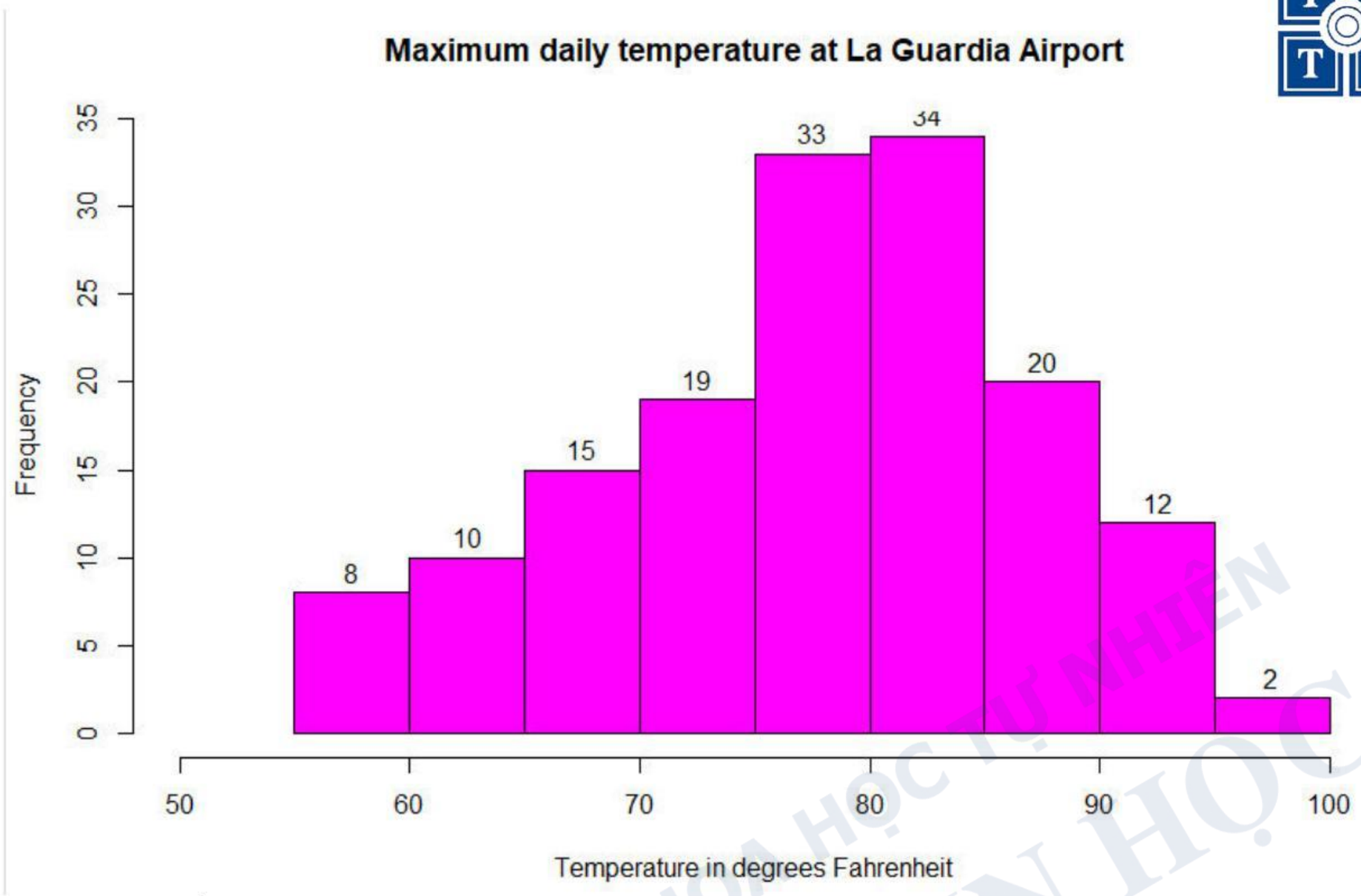
## Exercise 3: BAR CHART nhóm cột

- Cho dữ liệu: year <- c(2014, 2015, 2016, 2017), first_fallen <- c(309, 271, 263, 255), one_hundred_fallen <- c(309,314,312,304)
- Hãy vẽ biểu đồ nhóm cột như hình sau:




## Exercise 4: Airquality - Histogram

- Sử dụng dữ liệu airquality
- In thông tin về dữ liệu này
- Tạo biến chứa dữ liệu Temp của airquality
- Vẽ histogram liên quan đến biến chứa Temp này với số cột phân chia dữ liệu là 5.
- Thay vì trục tung là Frequency => hãy chuyển thành đồ thị có trục tung là Density
- Hãy vẽ một đồ thị có số lượng mẫu trên mỗi cột (có 9 cột) như hình sau:




- Hãy vẽ một đồ thị chỉ có các cột giá trị 55, 60, 70, 75, 80, 90, 100

# Exercise 5: AirPassengers - Histogram

- Sử dụng dữ liệu AirPassengers
- In thông tin về dữ liệu này
- Vẽ histogram thể hiện thông tin AirPassengers với 10 khoảng (cột), màu cầu vồng, dùng density

# Exercise 6: Chol - Histogram

- Cung cấp tập tin chol.txt
- Đọc nội dung tập tin vào
- Cho biết kiểu của nội dung
- In thông tin nội dung
- Vẽ histogram thể hiện thông tin AGE, có gắn thông tin số lượng trên mỗi cột

# Exercise 7: Leaves had fallen – Line graph

- Sử dụng dữ liệu: years <- c(2014, 2015, 2016, 2017), first_fallen <- c(309, 271, 263, 255) one_hundred_fallen <- c(309,314,312,304)
- Vẽ biểu đồ hiển thị cả first_fallen và one_hundred_fallen

# Exercise 8: Tech_com – line graph

- Cung cấp tập tin Tech_comp.csv
- Đọc và hiển thị nội dung tập tin
- Vẽ line graph thể hiện sự so sánh giữa men_tech và men_non_tech, women_tech và women_non_tech

## Exercise 9: Tech_com – Scatter plot

- Sử dụng dữ liệu có sẵn faithful
- Vẽ biểu đồ thể hiện mối quan hệ giữa eruptions và waiting
- Vẽ biểu đồ thể hiện mối quan hệ giữa eruptions và waiting, có thêm regression line

---

# Gợi ý

## Exercise 1: Painters school – PIE

In [1]:
```
library(MASS)
```

In [2]:
```
print(paste("Data type:", class(painters)))
print(paste("Nrows:", nrow(painters), ", ncols:", ncol(painters)))
```

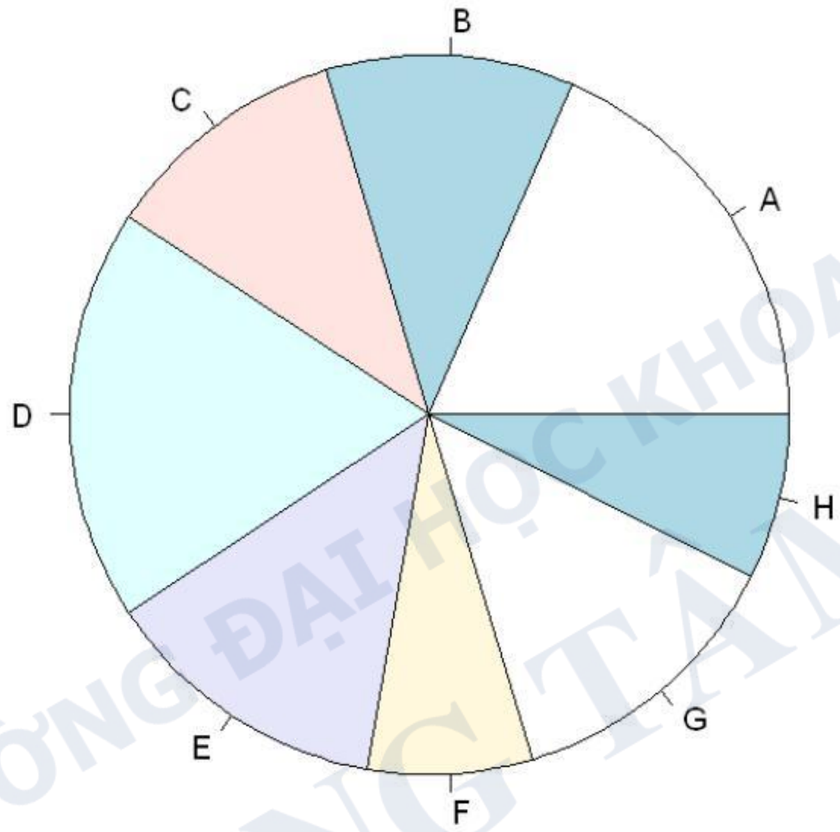
```
[1] "Data type: data.frame"
[1] "Nrows: 54 , ncols: 5"
```

In [3]:
```
school <- painters$School
print(paste("School type:", class(school)))
print("School:")
print(school)
```

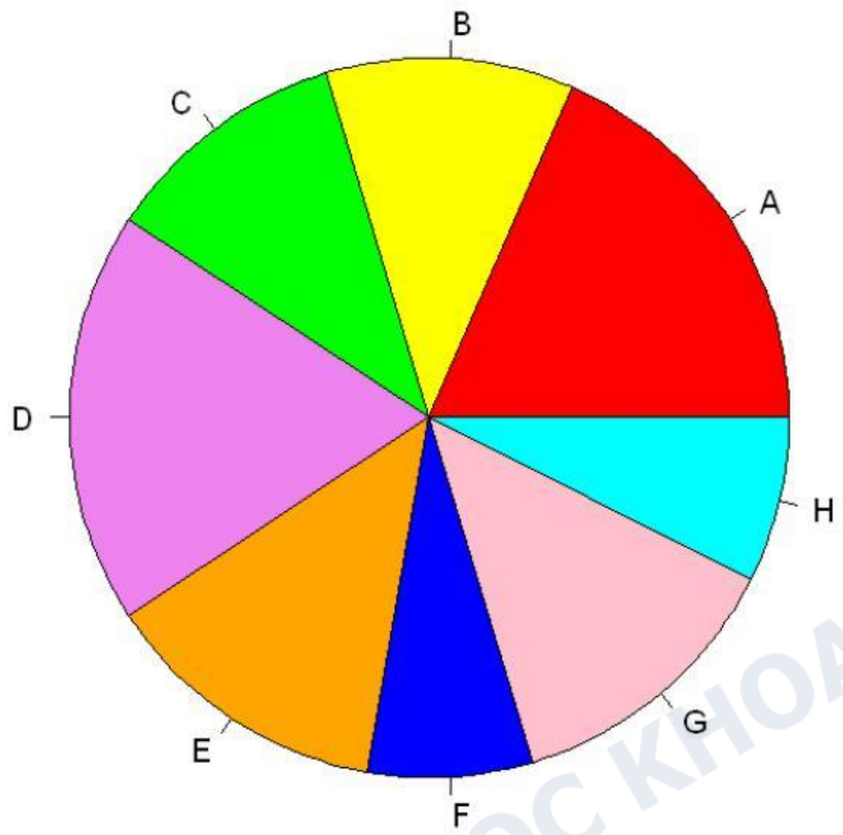
```
[1] "School type: factor"
[1] "School:"
 [1] A A A A A A A A A A B B B B B B C C C C C C D D D D D D D D D D E E E E E
E
[39] E F F F F G G G G G G G H H H H
Levels: A B C D E F G H
```

In [4]:
```
# chuyen factor thanh table
school.freq <- table(school)
print("School table:")
print(school.freq)
#ve
pie(school.freq)
```

```
[1] "School table:"
school
 A  B  C  D  E  F  G  H
10  6  6 10  7  4  7  4
```



In [5]:
```
# cho cac mau
colors <- c("red", "yellow", "green", "violet", "orange", "blue", "pink", "cyan")
#ve bieu do co mau
pie(school.freq,col=colors)
```



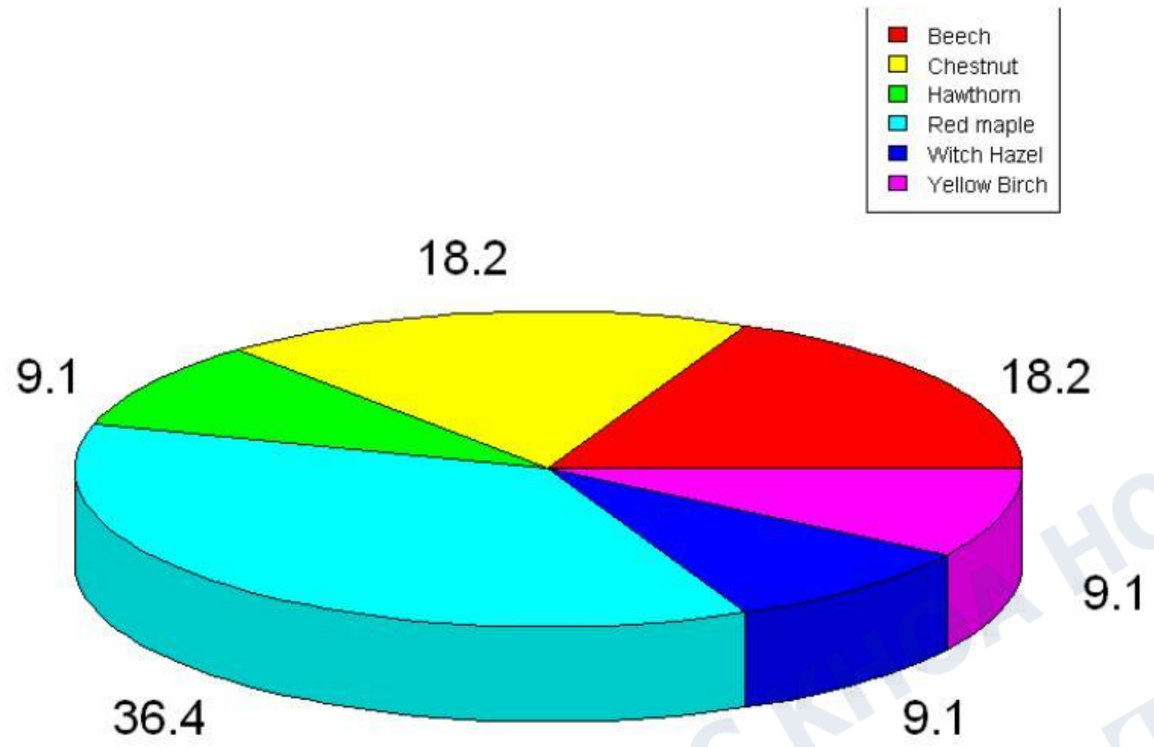



# Exercise 2: PIE 3D

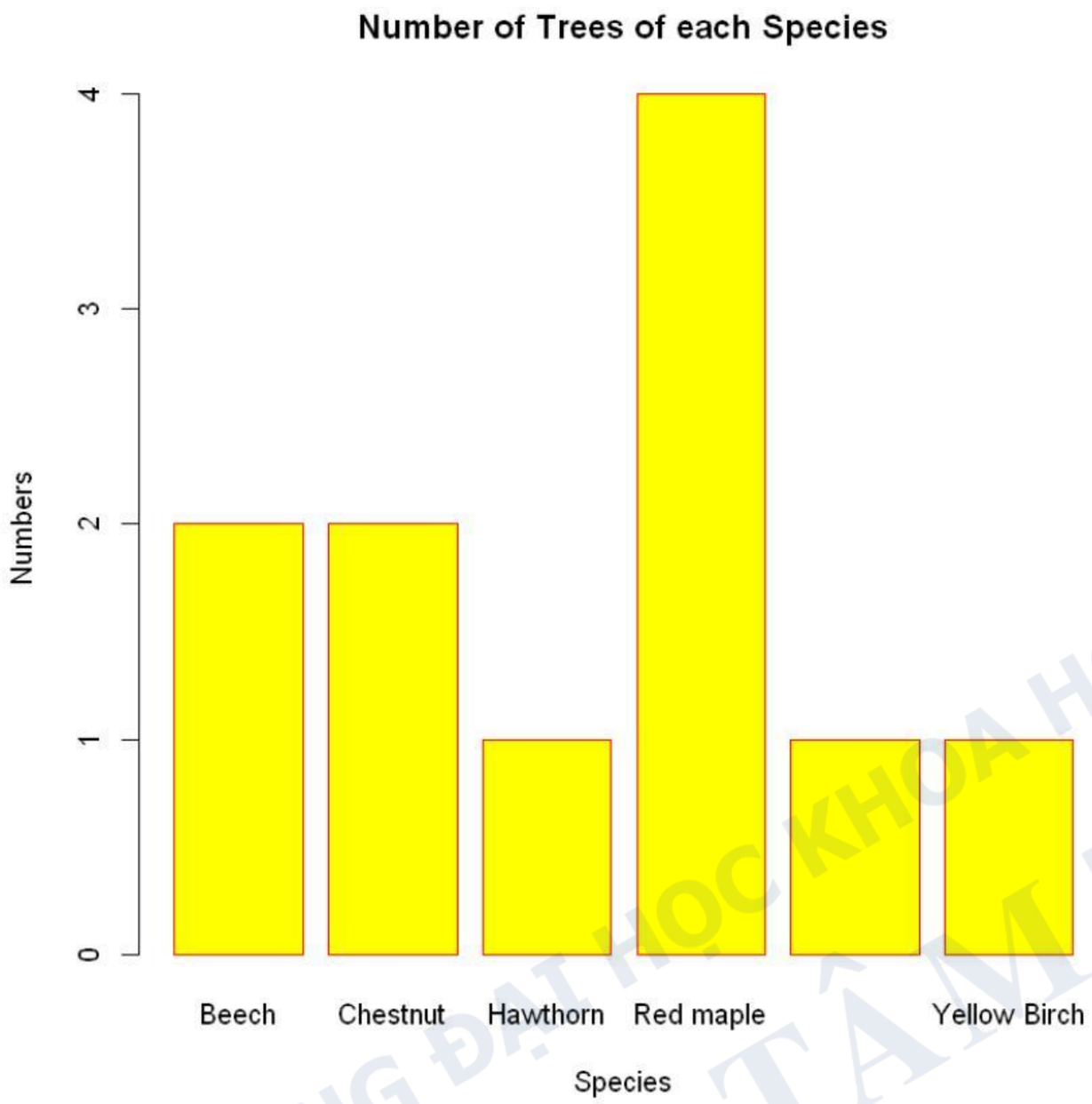
In [7]:
```
library("plotrix")
```

In [11]:
```
Tree_specices <- c("Beech", "Chestnut", "Hawthorn",
                   "Red maple", "Witch Hazel", "Yellow Birch")
number_of_trees <- c(2, 2, 1, 4, 1, 1)

piepercent<- round(100*number_of_trees/sum(number_of_trees), 1)
```

In [12]:
```
pie3D(number_of_trees, labels = piepercent ,
      col = rainbow(length(number_of_trees)),
      main = "Number of Trees of each Species")

legend("topright", Tree_specices, cex =0.7,
       fill = rainbow(length(number_of_trees)))
```



In [13]:
```
# Plot the bar chart.
barplot(number_of_trees,names.arg = Tree_spices,
        xlab = "Species",ylab = "Numbers", col = "yellow",
        main = "Number of Trees of each Species",border = "red")
```



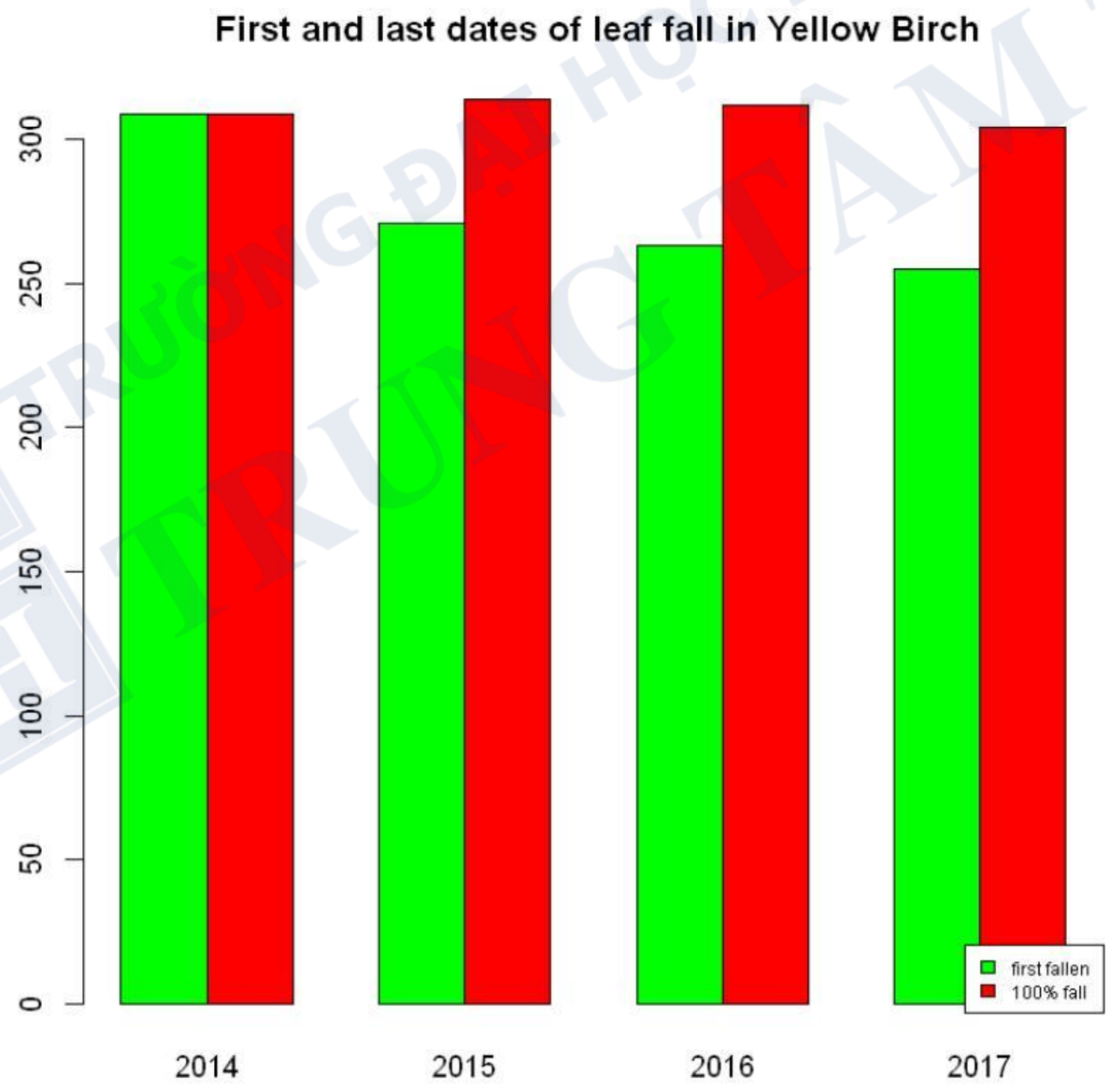



## Exercise 3: BAR CHART nhóm cột

In [14]:
```
#ghep cot
year <- c(2014, 2015, 2016, 2017)
first_fallen <- c(309, 271, 263, 255)
one_hundred_fallen <- c(309,314,312,304)

data_tree <- data.frame(row.names = year,
                        first_fallen,one_hundred_fallen )
print(data_tree)
```

```
     first_fallen one_hundred_fallen
2014          309                309
2015          271                314
2016          263                312
2017          255                304
```

In [15]:
```
# Plot the bar chart.
barplot(t(as.matrix(data_tree)),beside=TRUE,
        main = " First and last dates of leaf fall in Yellow Birch",
        col = c("green","red"))
legend("bottomright", c("first fallen", "100% fall"),
       cex =0.6, fill = c("green","red"))
```




# Exercise 4: Airquality - Histogram

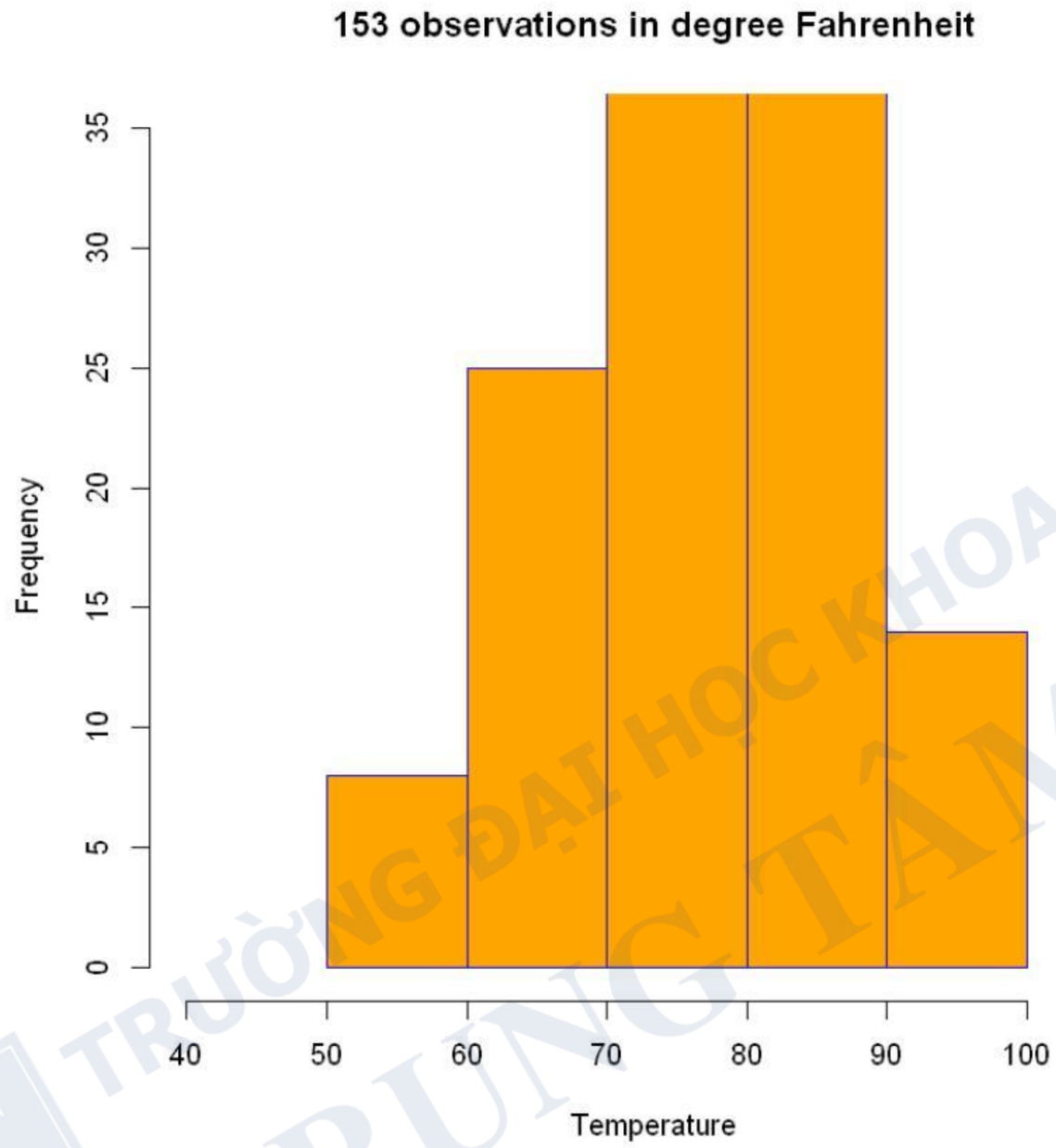
In [16]:
```
# Create data for the graph.
print("Daily air quality measurements in New York, May to September 1973.")
str(airquality)
```

```
[1] "Daily air quality measurements in New York, May to September 1973."
'data.frame':   153 obs. of  6 variables:
 $ Ozone  : int  41 36 12 18 NA 28 23 19 8 NA ...
 $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
 $ Wind   : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
 $ Temp   : int  67 72 74 62 56 66 65 59 61 69 ...
 $ Month  : int  5 5 5 5 5 5 5 5 5 5 ...
 $ Day    : int  1 2 3 4 5 6 7 8 9 10 ...
```

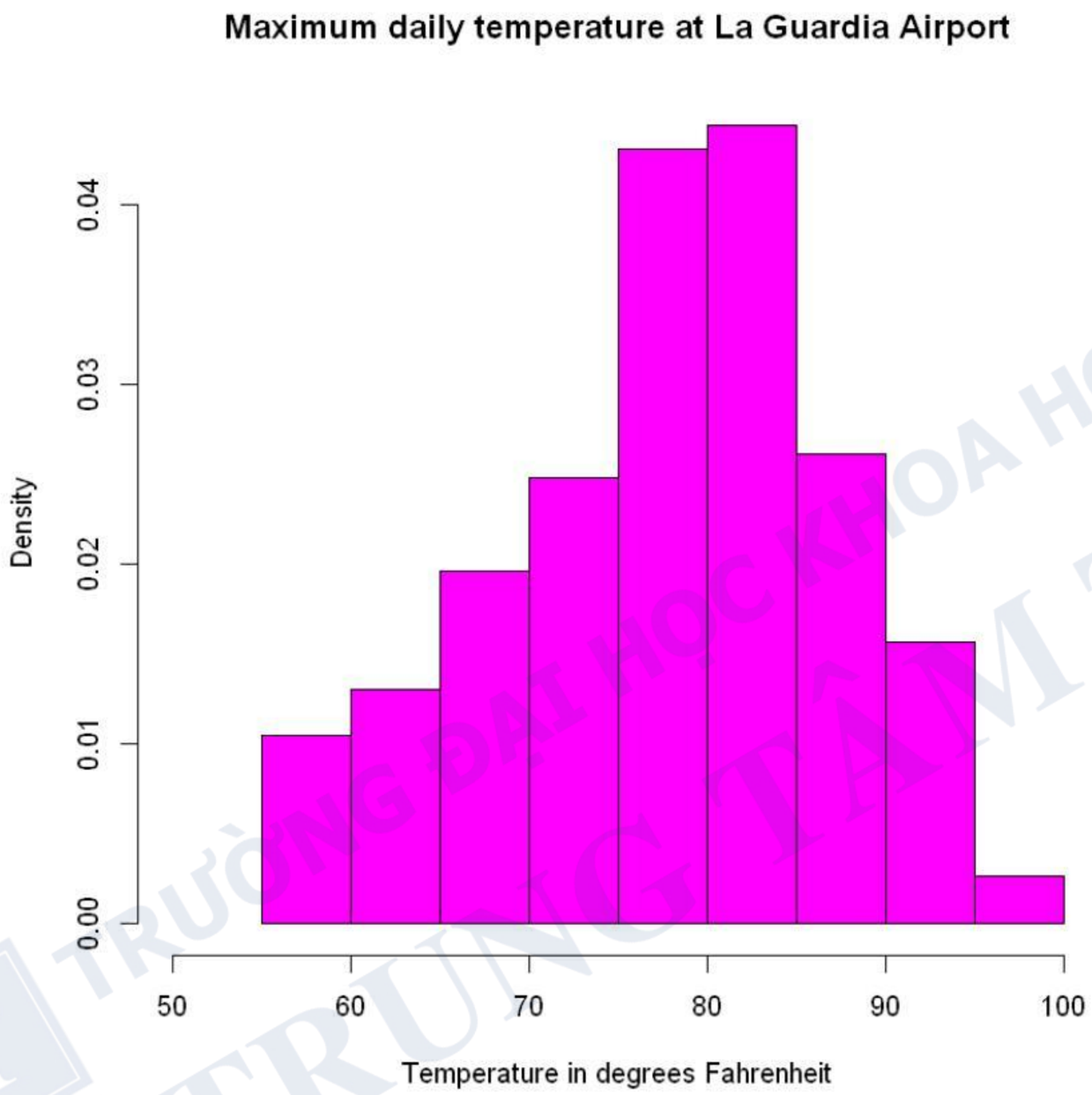
In [17]:
```
v <-  airquality$Temp
print(v)
```

```
  [1] 67 72 74 62 56 66 65 59 61 69 74 69 66 68 58 64 66 57 68 62 59 73 61 61 5
7
 [26] 58 57 67 81 79 76 78 74 67 84 85 79 82 87 90 87 93 92 82 80 79 77 72 65 7
3
 [51] 76 77 76 76 76 75 78 73 80 77 83 84 85 81 84 83 83 88 92 92 89 82 73 81 9
1
 [76] 80 81 82 84 87 85 74 81 82 86 85 82 86 88 86 83 81 81 81 82 86 85 87 89 9
0
[101] 90 92 86 86 82 80 79 77 79 76 78 78 77 72 75 79 81 86 88 97 94 96 94 91 9
2
[126] 93 93 87 84 80 78 75 73 81 76 77 71 71 78 67 76 68 82 64 71 81 69 63 70 7
7
[151] 75 76 68
```

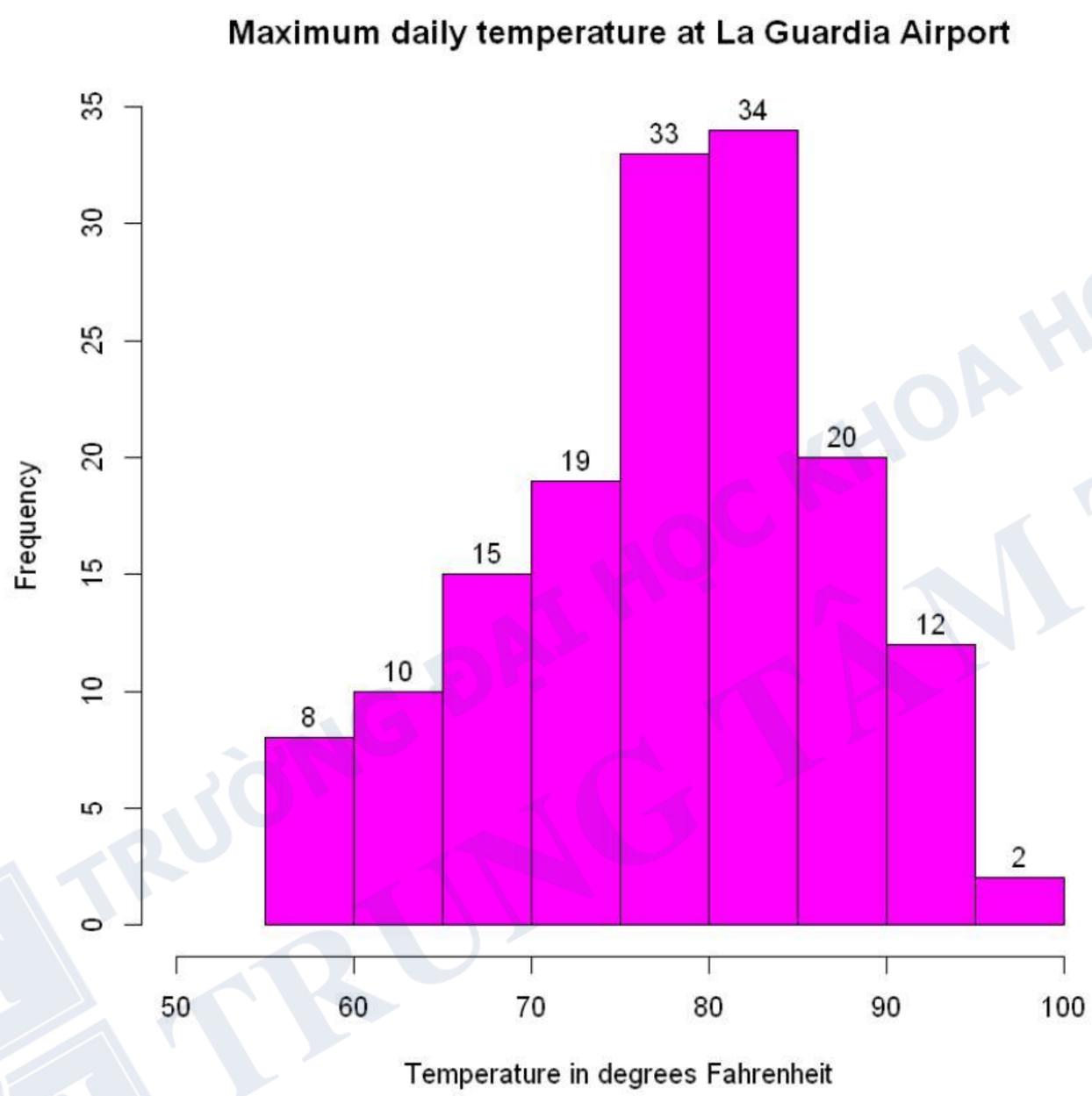
In [18]:
```
# Give the chart file a name.
#png(file = "Hinh/ages_histogram.png")

# Create the histogram.
# chia lam 5 cot
hist(v, main = "153 observations in degree Fahrenheit",
     xlab = "Temperature",
     xlim = c(40, max(v)+10), ylim = c(0, 35), col = "orange",
     border = "blue", breaks = 5)
```



In [21]:
```
# Create the histogram.
h <- hist(v,
     main="Maximum daily temperature at La Guardia Airport",
     xlab="Temperature in degrees Fahrenheit",
     xlim=c(50,100),
     col="magenta",
     freq=FALSE
)

#print(h)
```

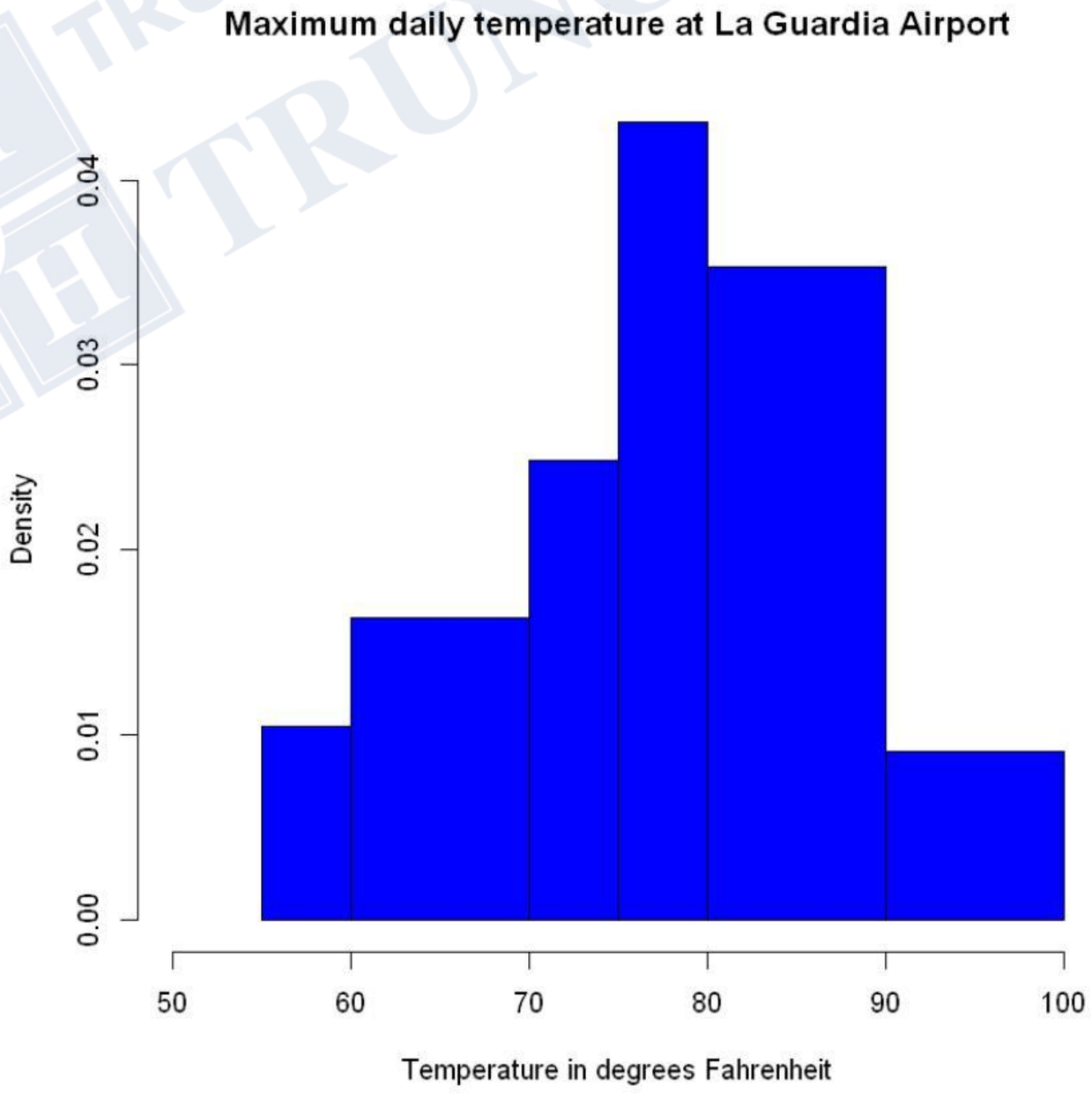


In [22]:
```
h <- hist(v,
     main="Maximum daily temperature at La Guardia Airport",
     xlab="Temperature in degrees Fahrenheit",
     xlim=c(50,100),
     col="magenta",
     breaks = 10
)

# su dung them hien thi text va slxh
text(h$mids,h$counts,labels=h$counts, adj=c(0.5, -0.5))
#print(h)
```



In [23]:
```
h <- hist(v,
     main="Maximum daily temperature at La Guardia Airport",
     xlab="Temperature in degrees Fahrenheit",
     xlim=c(50,100),
     col="blue",
     breaks=c(55,60,70,75,80,90, 100)
)
print(h)
```

```
$breaks
[1]  55  60  70  75  80  90 100

$counts
[1]  8 25 19 33 54 14

$density
[1] 0.010457516 0.016339869 0.024836601 0.043137255 0.035294118 0.009150327

$mids
[1] 57.5 65.0 72.5 77.5 85.0 95.0

$xname
[1] "v"

$equidist
[1] FALSE

attr(,"class")
[1] "histogram"
```

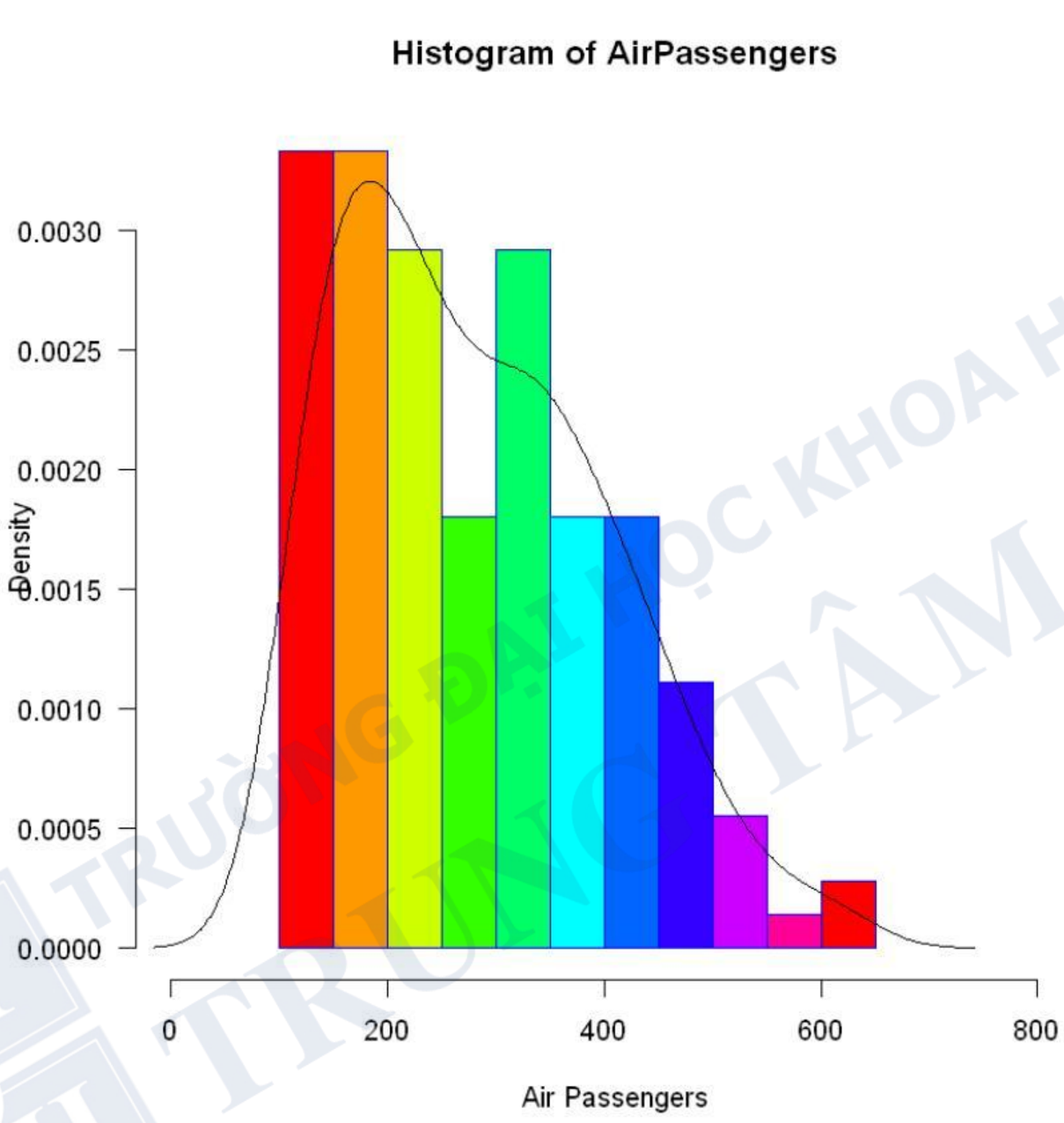


# Exercise 5: AirPassengers - Histogram

In [24]:
```
# set AirPassengers
print("AirPassengers")
str(AirPassengers)
```

```
[1] "AirPassengers"
 Time-Series [1:144] from 1949 to 1961: 112 118 132 129 121 135 148 148 136 119
...
```

In [25]:
```
# Create the histogram.
hist(AirPassengers, main = "Histogram of AirPassengers",
     xlab = "Air Passengers",
     xlim = c(0, max(AirPassengers)+200),
     col = rainbow(10),
     breaks = 10,
     border = "blue",
     # so lieu tren y theo hang ngang
     las = 1,
     freq=FALSE
     )
lines(density(AirPassengers))
```

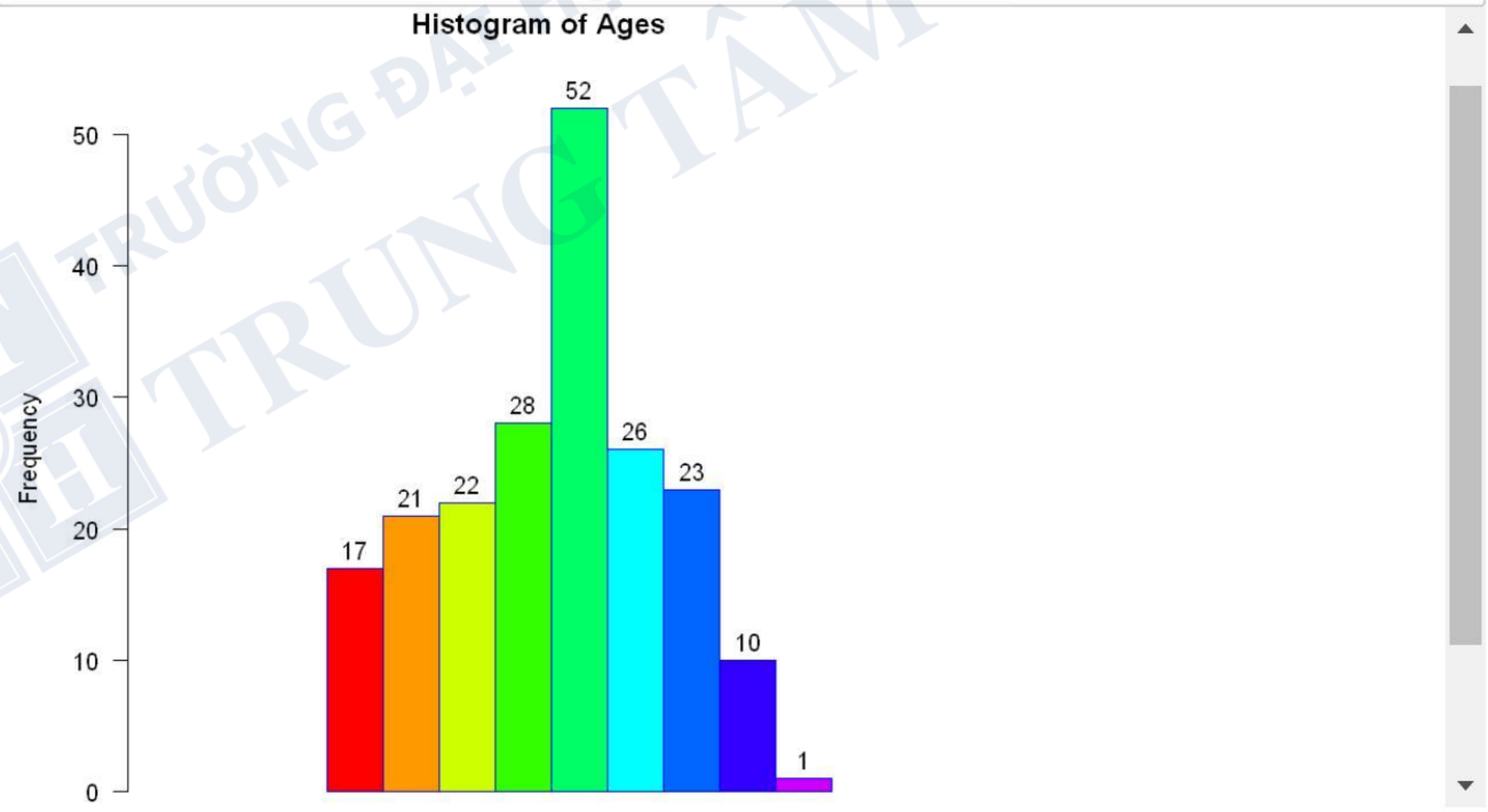



# Exercise 6: Chol - Histogram

In [26]:
```
data_chol <- read.table(file = "chol.txt",
                        header = TRUE)
print(class(data_chol))
str(data_chol)
```

```
[1] "data.frame"
'data.frame':   200 obs. of  7 variables:
 $ AGE   : int  20 53 44 37 26 41 39 28 33 39 ...
 $ HEIGHT: int  176 167 170 173 170 165 174 171 180 166 ...
 $ WEIGHT: int  77 56 80 89 71 62 75 68 100 74 ...
 $ CHOL  : int  195 250 304 178 206 284 232 152 209 150 ...
 $ SMOKE : Factor w/ 3 levels "nonsmo","pipe",..: 1 3 3 1 3 3 3 2 3 3 ...
 $ BLOOD : Factor w/ 4 levels "a","ab","b","o": 3 4 1 4 4 4 4 1 1 1 ...
 $ MORT  : Factor w/ 2 levels "alive","dead": 1 2 2 1 1 1 1 1 1 1 ...
```

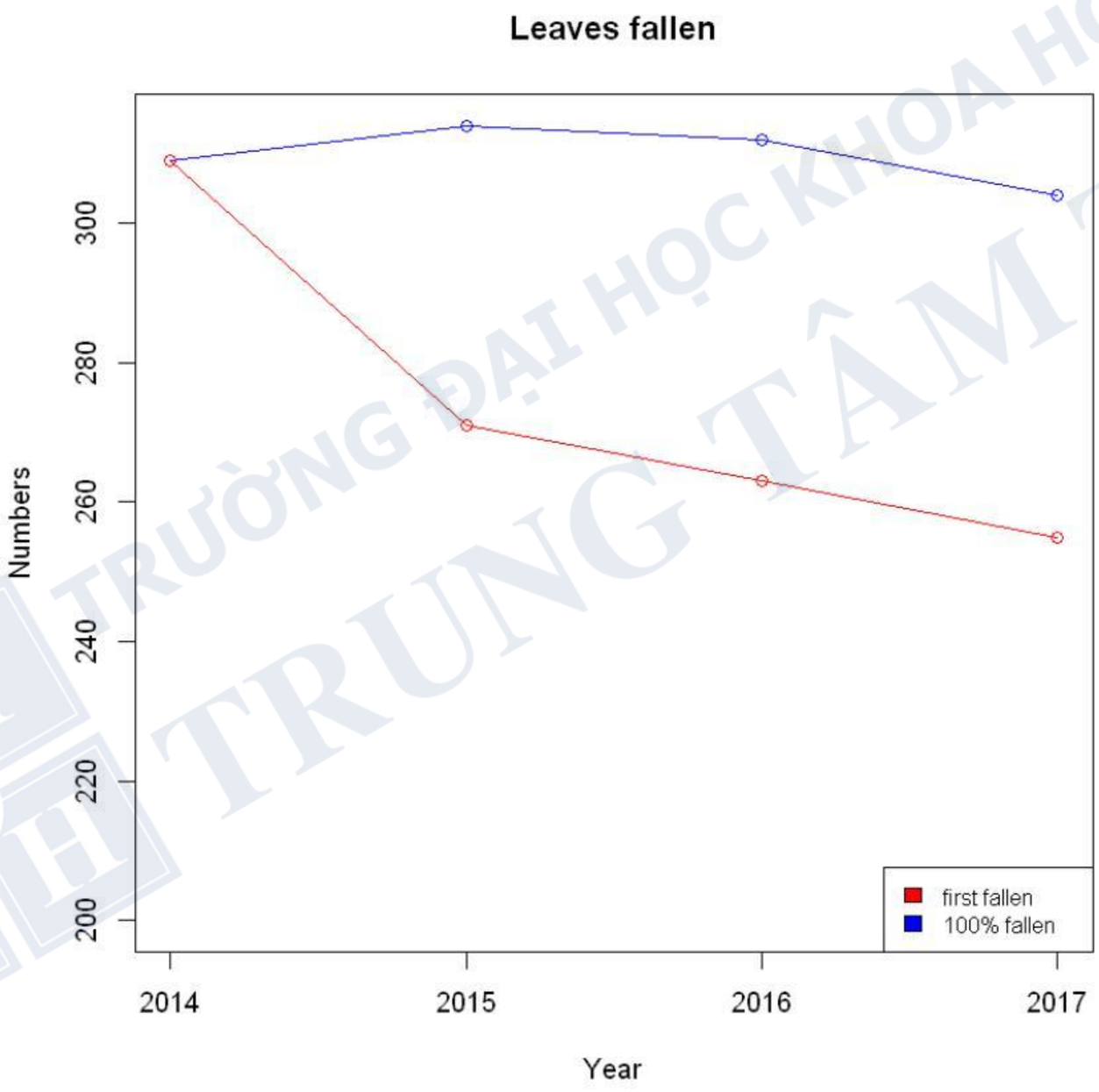
In [27]:
```
# Create the histogram.
h<-hist(data_chol$AGE, main = "Histogram of Ages",
      xlab = "Ages",
      xlim = c(0, max(data_chol$AGE)+10),
      col = rainbow(10),
      border = "blue",
      # so lieu tren y theo hang ngang
      las = 1
      )
text(h$mids,h$counts,labels=h$counts, adj=c(0.5, -0.5))
```




# Exercise 7: Leaves had fallen – Line graph

In [29]:
```
# Create the data for the chart.
years <- c(2014, 2015, 2016, 2017)
first_fallen <- c(309, 271, 263, 255)
one_hundred_fallen <- c(309,314,312,304)
```

In [30]:
```
# ve nhieu line tren cung mot graph

# Create the data for the chart.
# Calculate range from 0 to max values
g_range <- range(200,one_hundred_fallen)

# Plot the line chart.
plot(one_hundred_fallen, type="o", col="blue", ylim = g_range,
     xlab = "Year", ylab = "Numbers",
     main = "Leaves fallen", xaxt='n')
axis(1, at=1:4, labels = years)

lines(first_fallen, type = "o", col = "red")

# insert legend for chart
legend("bottomright", c("first fallen","100% fallen"), cex = 0.8,
        fill = c("red", "blue"))
```






# Exercise 8: Tech_com – line graph

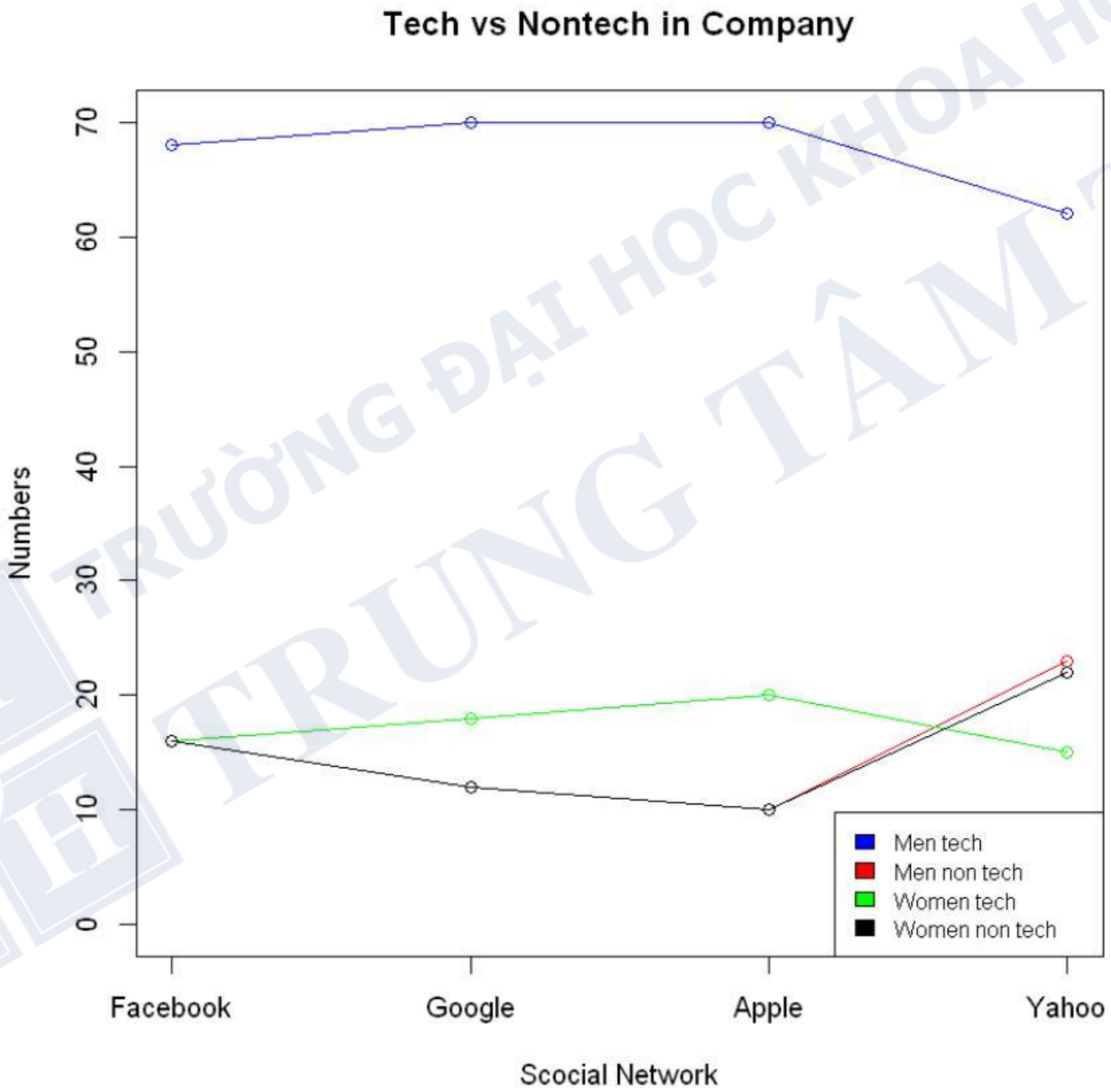
In [31]:
```
data_tech_com <- read.csv(file= "Tech_comp.csv")
print(data_tech_com)
```

```
   Company Women..all Women..tech. Men..all. Men..tech.
1 Facebook         32           16        84         68
2   Google         30           18        82         70
3    Apple         30           20        80         70
4    Yahoo         37           15        85         62
```

In [32]:
```
Men..non..tech <- data_tech_com$Men..all. - data_tech_com$Men..tech.
Women..non..tech <- data_tech_com$Women..all - data_tech_com$Women..tech.
```

In [33]:
```
g_range <- range(0,data_tech_com$Men..tech., data_tech_com$Women..tech.,
                 Men..non..tech, Women..non..tech)
```

In [37]:
```
# Plot the line chart.

plot(data_tech_com$Men..tech., type="o", col="blue", ylim = g_range,
     xlab = "Scocial Network", ylab = "Numbers",
     main = "Tech vs Nontech in Company",
     xaxt='n')
axis(side=1, at=seq(1,4,1), labels = data_tech_com$Company)
# Create box around plot
box()
# lines
lines(Men..non..tech, type = "o", col = "red")
lines(data_tech_com$Women..tech., type = "o", col = "green")
lines(Women..non..tech, type = "o", col = "black")

# insert legend for chart
legend("bottomright",legend = c("Men tech", "Men non tech", "Women tech", "Women
       cex = 0.8,
       fill = c("blue","red", "green","black"))
```



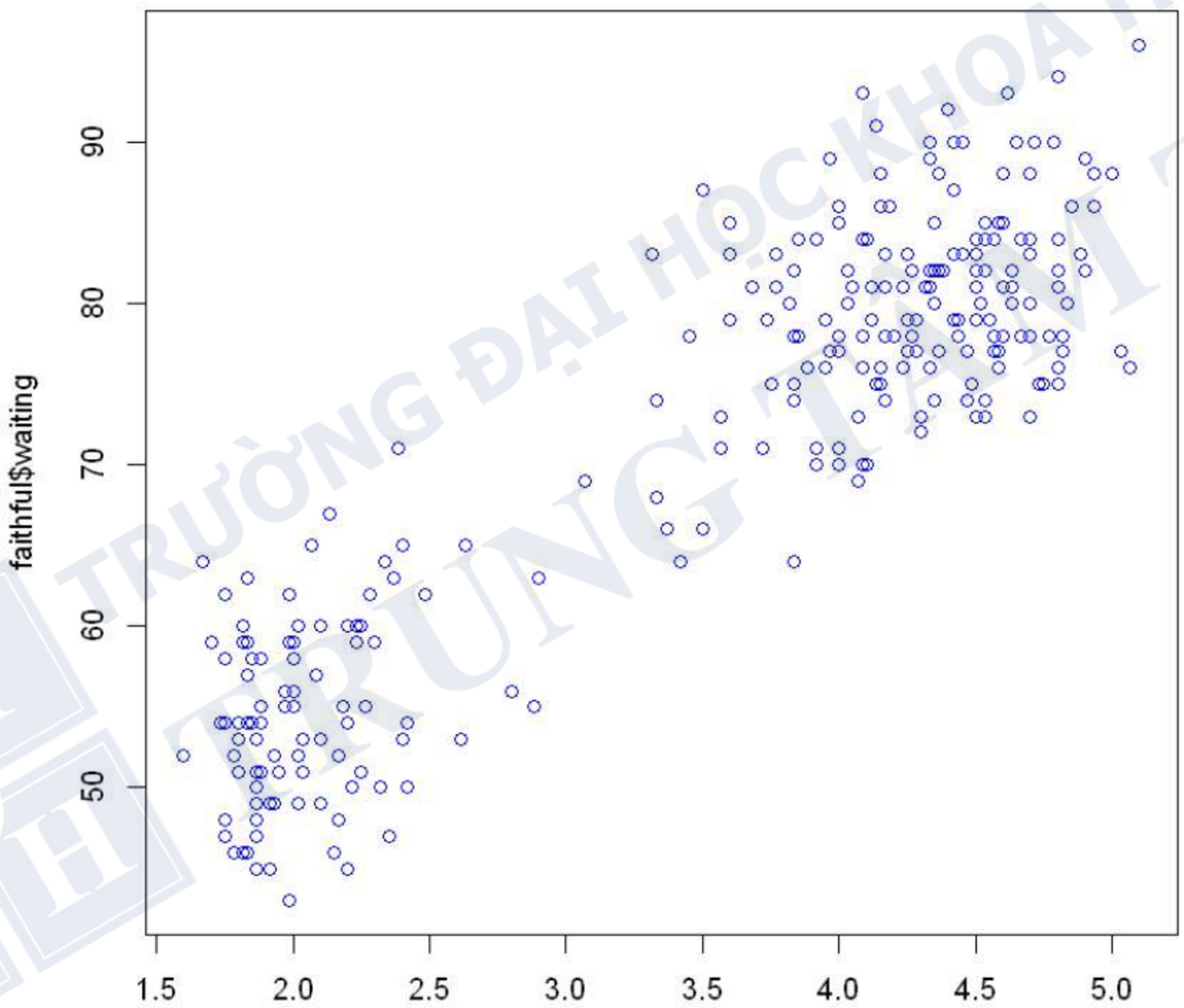

# Exercise 9: Scatter plot

In [39]:
```
print(head(faithful))
```

```
  eruptions waiting
1     3.600      79
2     1.800      54
3     3.333      74
4     2.283      62
5     4.533      85
6     2.883      55
```

In [41]:
```
# Finding the Correlation
cor(faithful$eruptions, faithful$waiting)

# Drawing Scatter Plot
plot(faithful$eruptions, faithful$waiting, col = 'blue')
```

```
0.900811168321813
```



In [51]:
```
# Drawing Scatter Plot
plot(faithful$eruptions, faithful$waiting,
     main = "eruptions vs waiting",
     xlab = "Eruptions",
     ylab = "Waiting",
     col = "chocolate",
     cex = 1.2,
     pch = 8,
     xlim = c(1.0, 6.0),
     ylim = c(40, 100))

#Linear Regression Line
abline(lm(faithful$waiting~faithful$eruptions), col = "red", lwd = 3)
```