Data Visualization with R

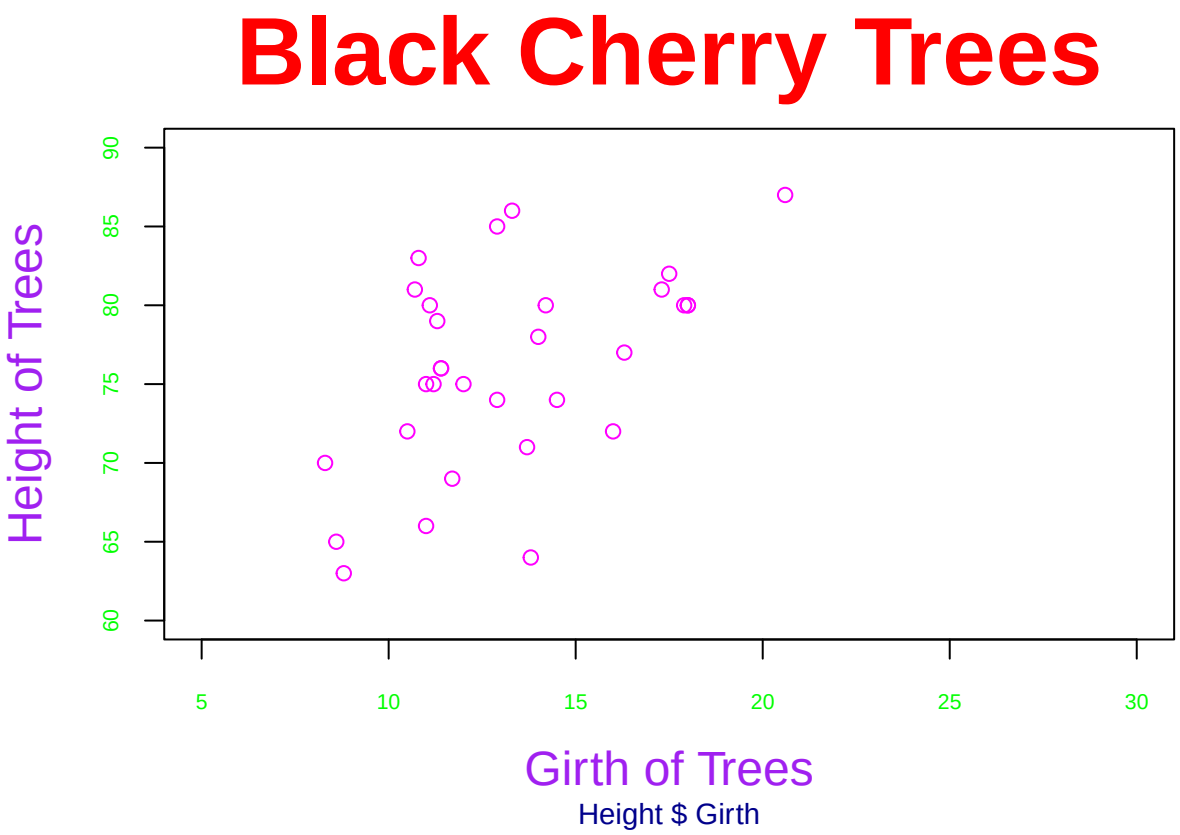
Hazim Fitri

Contents

Plot

```
plot(x = trees$Girth,
     y = trees$Height,
     main='Black Cherry Trees',
     sub='Height $ Girth',
     xlab = 'Girth of Trees',
     ylab = 'Height of Trees',
     cex.main = 3,
     cex.sub = 0.9,
     cex.lab = 1.5,
     cex.axis = 0.7,
     col.main = 'red',
     col.sub = 'darkblue',
     col.axis = 'green',
     col.lab = 'purple',
```

```
     col = 'magenta',
     xlim = c(5,30),
     ylim = c(60,90))
```



Options for parameter 'type'

p = point (default)

l = line

b = both point & line (not overlapping)

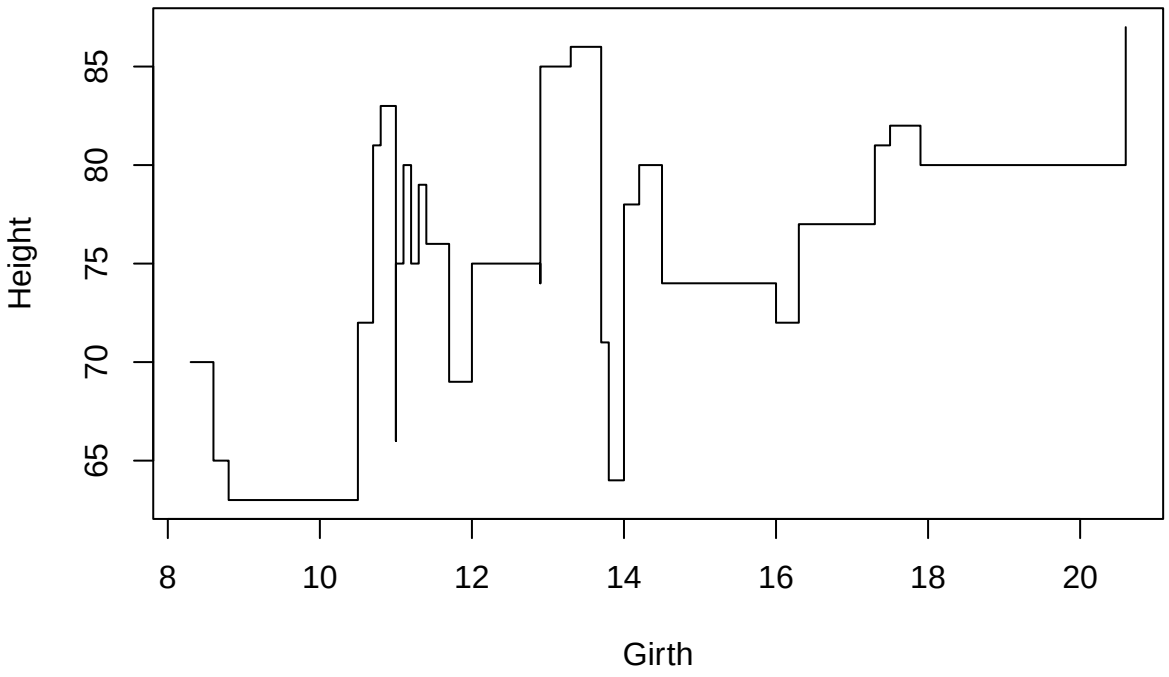
o = both point & line (overlapping)

h = macam histogram

s = step

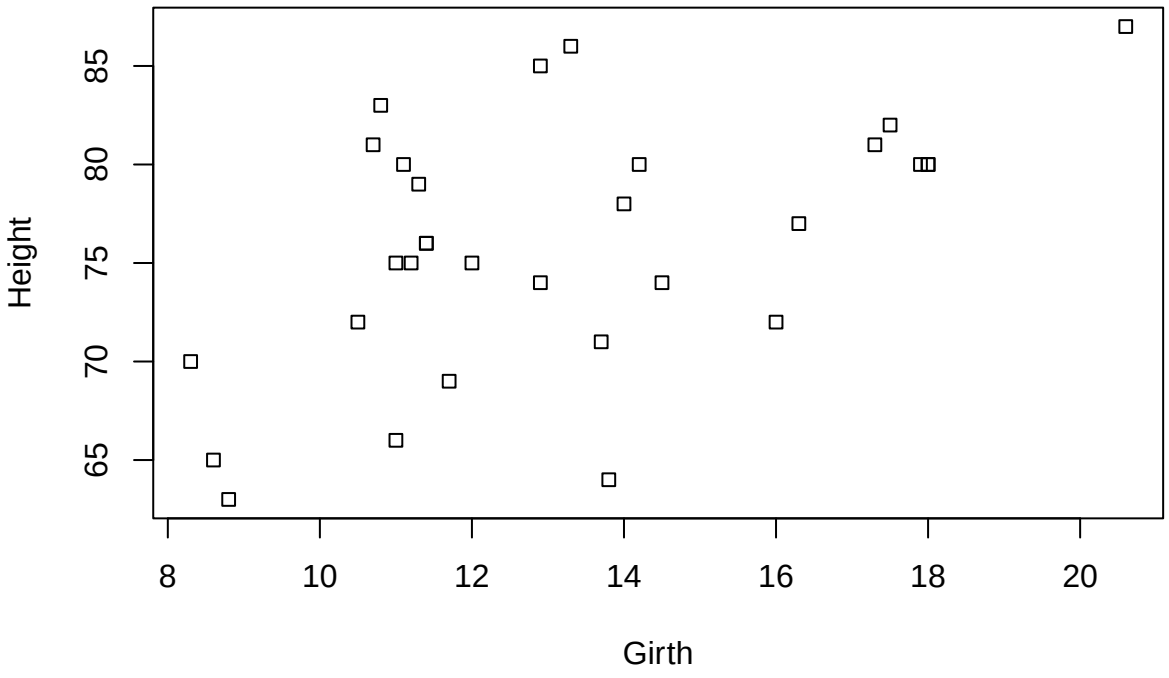
n = none

```
plot(Height~Girth,
     data = trees,
     type = 's')
```



Options for parameter 'pch'

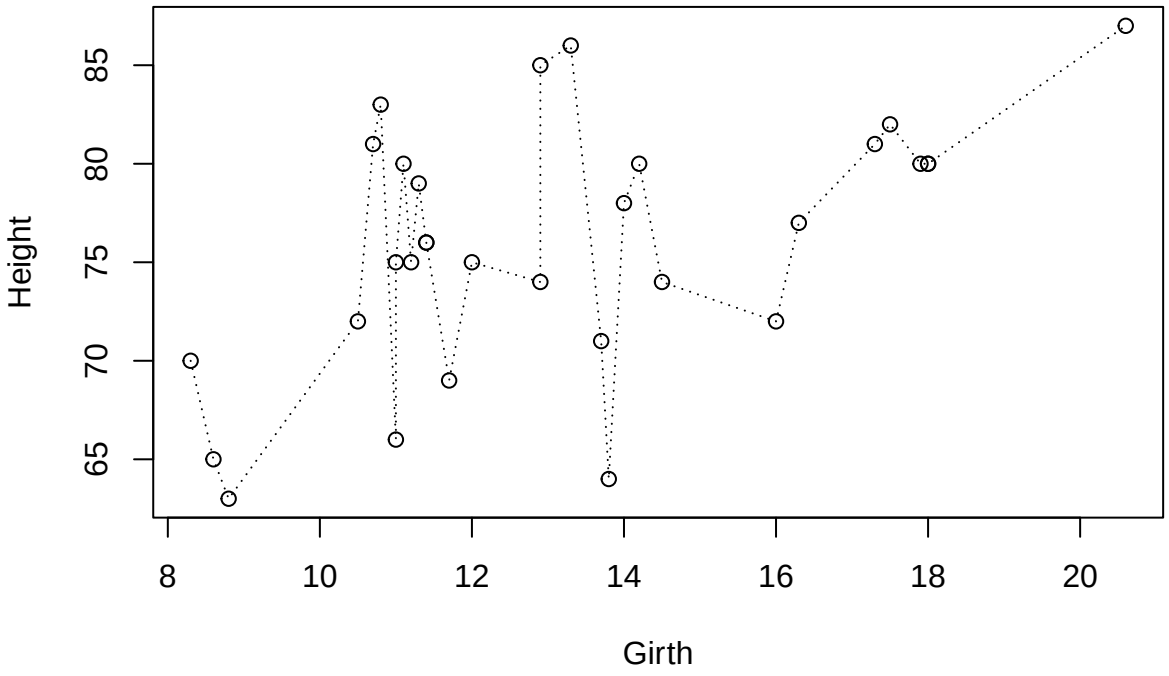
```
plot(Height~Girth,
     data = trees,
     pch = 22)
```



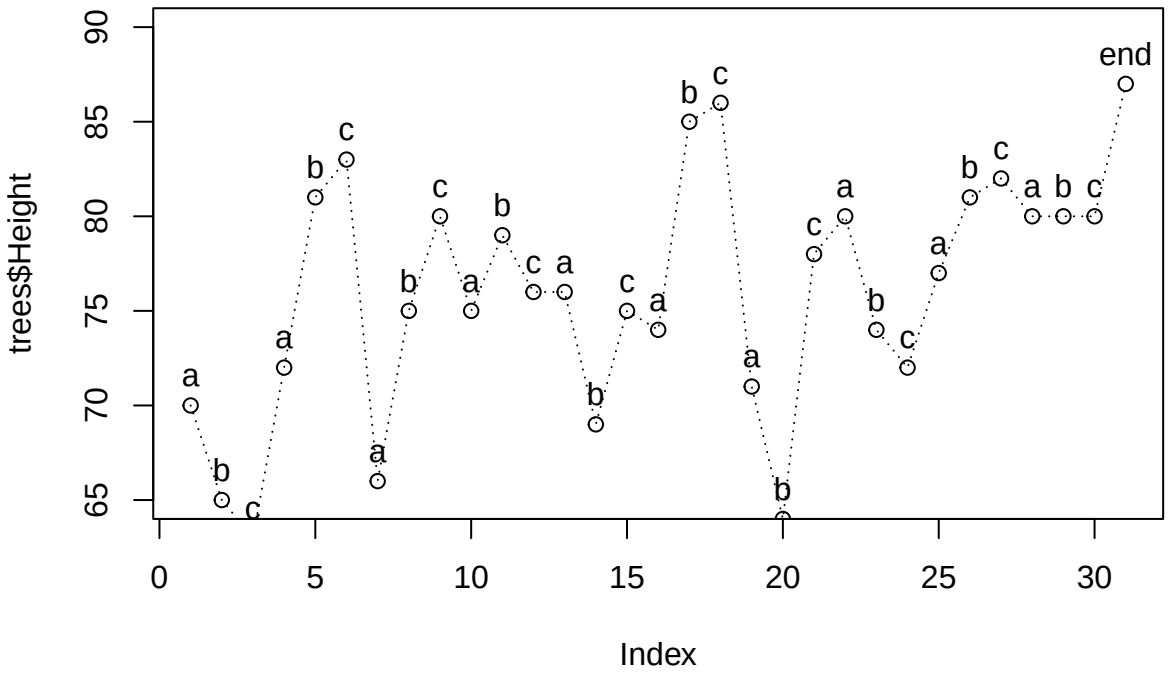
Options for parameter 'lty'

3 = dashed line

```
plot(Height~Girth,
     data = trees,
     type = 'o',
     lty = 3)
```



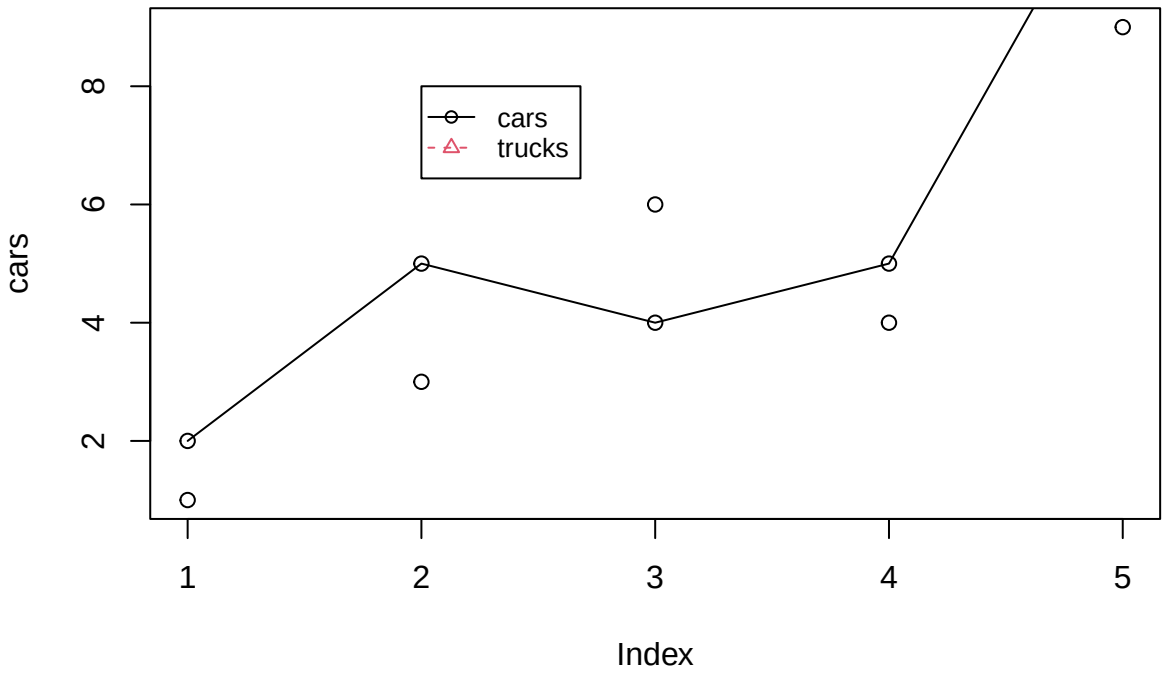

```
plot(trees$Height,
     type = 'o',
     lty = 3,
     ylim = c(65,90))

tplot = c(rep(c('a', 'b', 'c'),10),'end')
text(trees$Height, tplot, pos = 3)
```



y = 15)

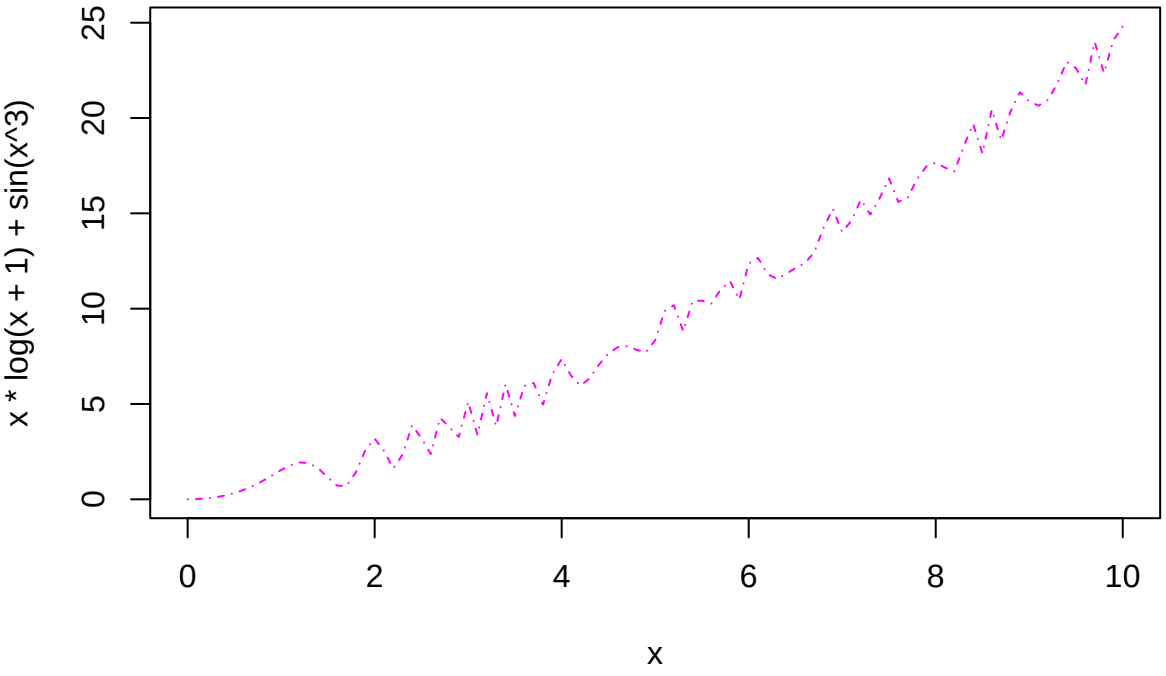
Legend

```
cars = c(1,3,6,4,9)
trucks = c(2,5,4,5,12)
suvs = c(4,4,6,6,15)

plot(cars)
points(trucks)
lines(trucks)
legend(2,
       8,
       cex = 0.8,
       pch = 1:3, # also can use c(1,3)
       lty = 1:2,
       col = 1:2,
       c('cars', 'trucks'))
```



Curve function

```
curve(x*log(x+1)+sin(x^3),from=0,
to=10,col="magenta",lty=4)
```

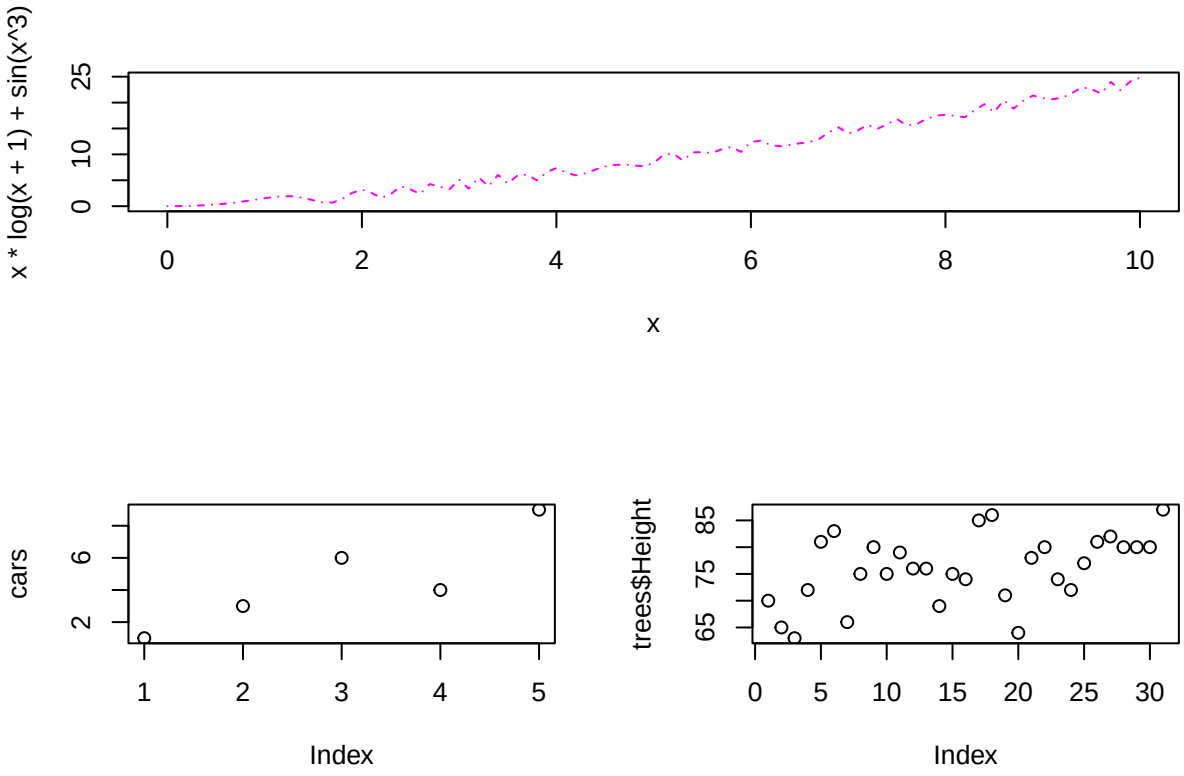


Layout function

```
matrix(c(1,1,2,3),nrow=2,
 byrow=T)
```

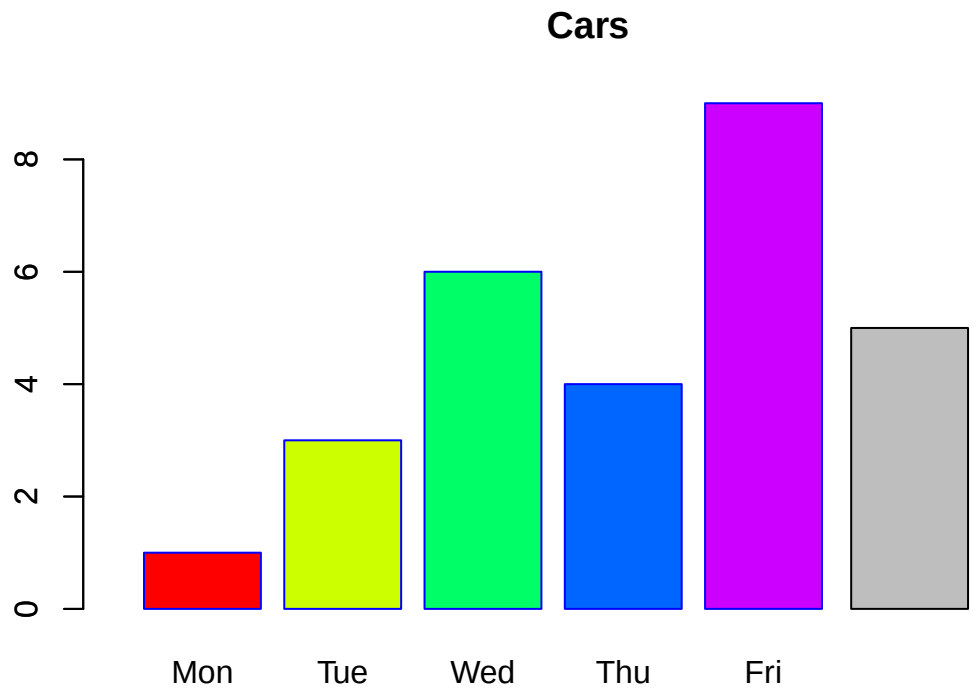
```
##      [,1] [,2]
## [1,]    1    1
## [2,]    2    3
```

```
layout(matrix(c(1,1,2,3),nrow=2,
 byrow=T))

 curve(x*log(x+1)+sin(x^3),from=0,
 to=10,col="magenta",lty=4)

 plot(cars)

 plot(trees$Height)
```

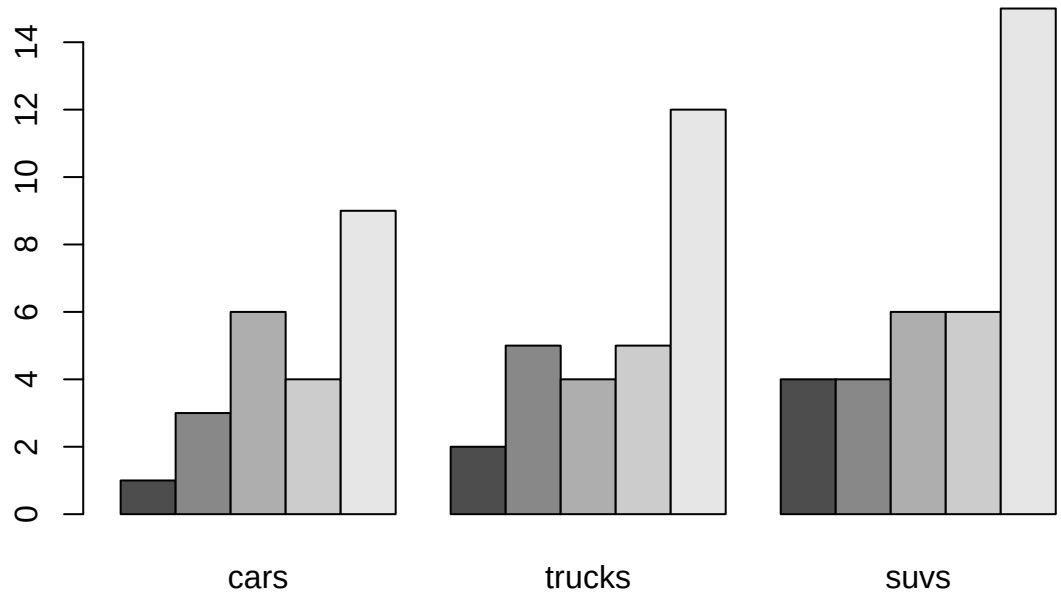


Barplot

```
barplot(cars,
        border = 'blue',
        main = 'Cars',
        col = rainbow(5), # 1:5 # 'skyblue'
        names.arg = c('Mon', 'Tue', 'Wed', 'Thu', 'Fri'),
        xlim = c(0,8))
barplot(5,
        add = T,
        space = 6.25, # position
        )
```

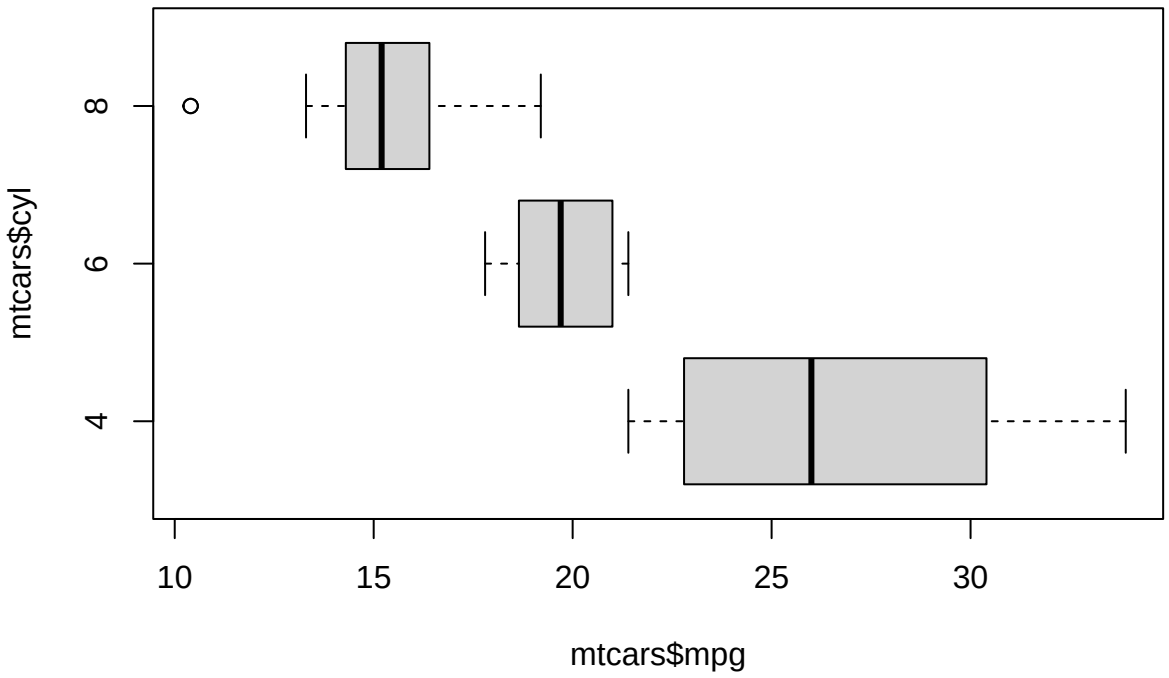



```
autos = cbind(cars, trucks, suvs)
barplot(autos,
        beside = T # by default is stack barplot
        )
```



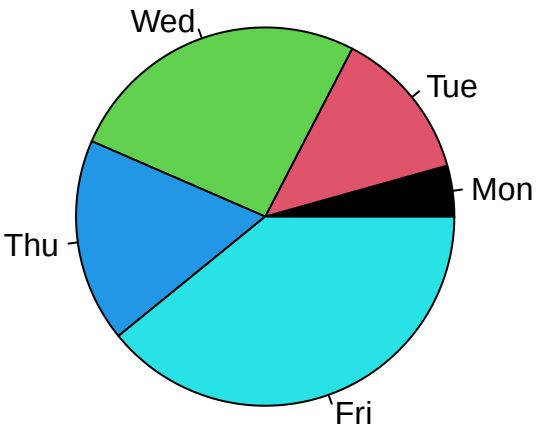
Boxplot

```
boxplot(mtcars$mpg~mtcars$cyl,
        horizontal = T) # by default is vertical
```



Pie chart

```
pie(cars,
    labels = c('Mon', 'Tue', 'Wed', 'Thu', 'Fri'),
    col = 1:5)
```

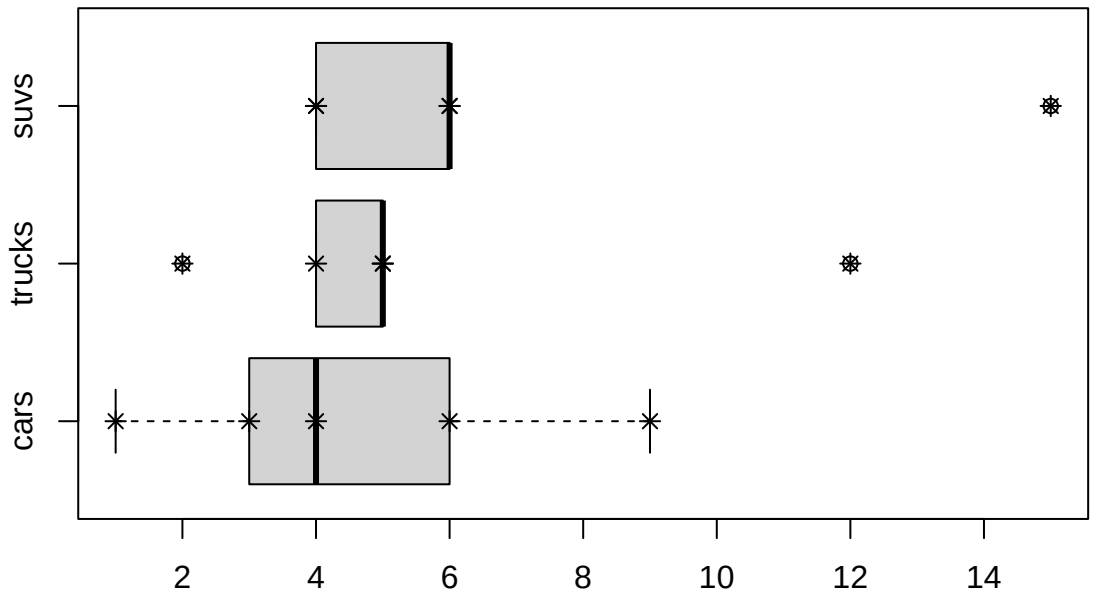


Strip chart

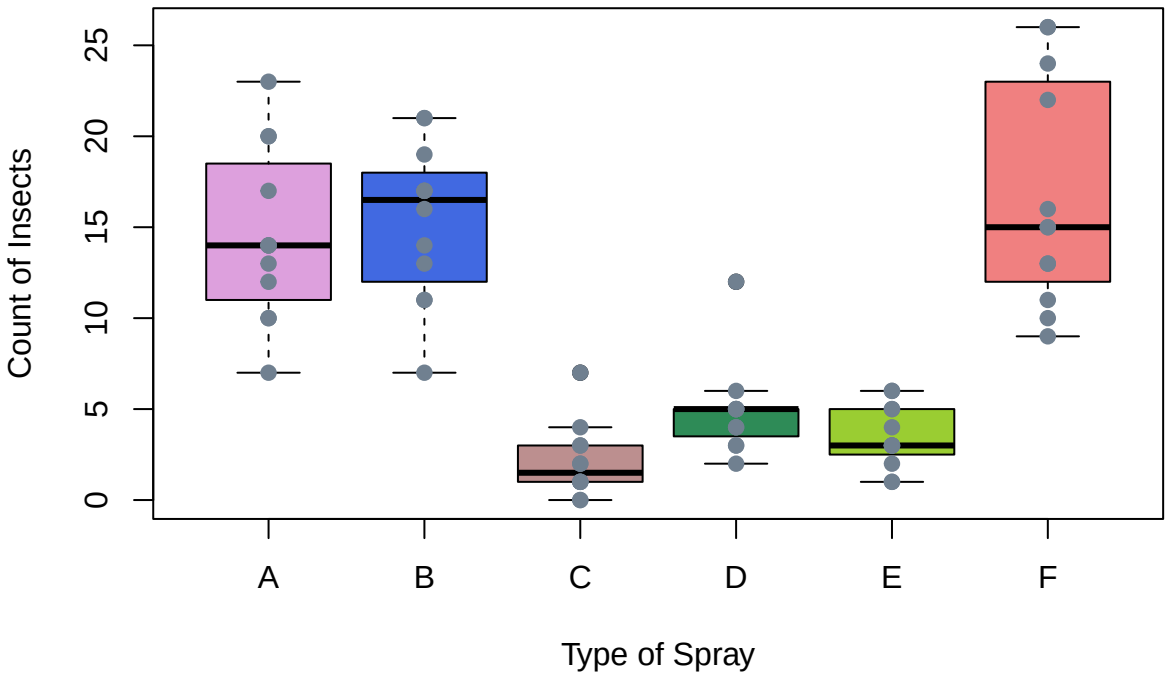
boxplot and stripchart use data frame

barplot use matrix

```
autos = data.frame(autos)

boxplot(autos,
        horizontal = T) # by default is vertical

# default is horizontal, terbalik dengan boxplot
stripchart(autos,
           pch = 8,
           add = T)
```




```
 data(InsectSprays)
 attach(InsectSprays)
 boxplot(count~spray,
col=c("plum","royalblue","rosybrown","seagreen",
 "yellowgreen","lightcoral"),
 xlab="Type of Spray",ylab="Count of Insects")
 stripchart(count~spray,add=T,vertical=T,pch=19,
 col="slategray")
```



Exercise

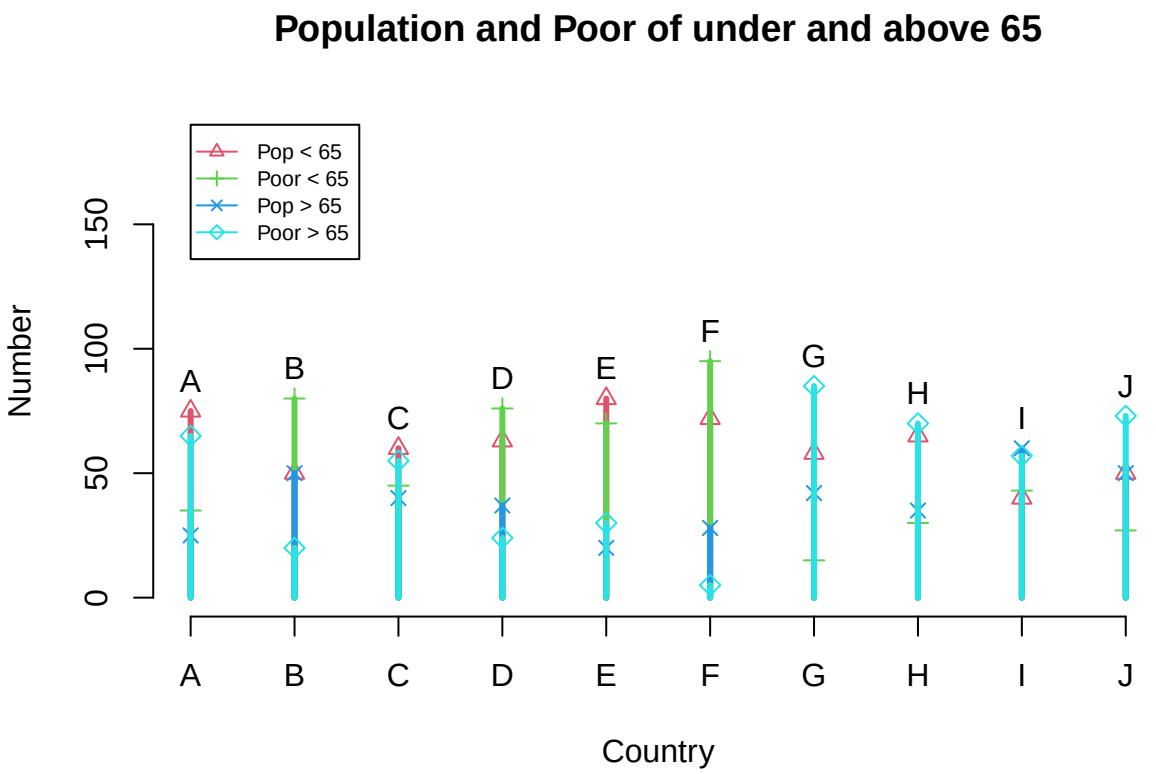
Question 1

```
# Create the data frame
df <- data.frame(
  Country = c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J"),
  Population_under_65 = c(75, 50, 60, 63, 80, 72, 58, 65, 40, 50),
  Poor_under_65 = c(35, 80, 45, 76, 70, 95, 15, 30, 43, 27),
  Population_above_65 = c(25, 50, 40, 37, 20, 28, 42, 35, 60, 50),
  Poor_above_65 = c(65, 20, 55, 24, 30, 5, 85, 70, 57, 73)
)

# Display the data frame
print(df)
```

```
##    Country Population_under_65 Poor_under_65 Population_above_65 Poor_above_65
## 1        A                  75            35                  25            65
## 2        B                  50            80                  50            20
## 3        C                  60            45                  40            55
## 4        D                  63            76                  37            24
## 5        E                  80            70                  20            30
## 6        F                  72            95                  28             5
## 7        G                  58            15                  42            85
## 8        H                  65            30                  35            70
## 9        I                  40            43                  60            57
## 10       J                  50            27                  50            73
```

```
plot(df$Population_under_65, type = 'h', col = 2, ylim = c(0, 190),
     axes = FALSE,
     main = 'Population and Poor of under and above 65',
     ylab = 'Number',
```

```
     xlab = 'Country',
     lwd = 3)
points(df$Population_under_65, type = 'p', col = 2, pch = 2)
points(df$Poor_under_65, type = 'h', col = 3, lwd = 3)
points(df$Poor_under_65, type = 'p', col = 3, pch = 3)
points(df$Population_above_65, type = 'h', col = 4, lwd = 3)
points(df$Population_above_65, type = 'p', col = 4, pch =4)
points(df$Poor_above_65, type = 'h', col = 5, lwd = 3)
points(df$Poor_above_65, type = 'p', col = 5, pch = 5)
text(apply(df[, -1], 1, max),df$Country, pos = 3)

legend(1, 190, c('Pop < 65', 'Poor < 65', 'Pop > 65', 'Poor > 65'), col = 2:5,
       cex = 0.7, lty = 1, pch = 2:5)

axis(1, at = 1:nrow(df), labels = df$Country)
axis(2)
```



Question 2

```
# Data for Male Statistics Students
male_students <- data.frame(
  Hair = c("Black", "Brown", "Red", "Blond"),
  Brown = c(32, 53, 10, 3),
  Blue = c(11, 50, 10, 30),
  Hazel = c(10, 25, 7, 5),
  Green = c(3, 15, 7, 8)
)

# Data for Female Statistics Students
female_students <- data.frame(
  Hair = c("Black", "Brown", "Red", "Blond"),
  Brown = c(36, 66, 16, 4),
  Blue = c(9, 34, 7, 64),
```

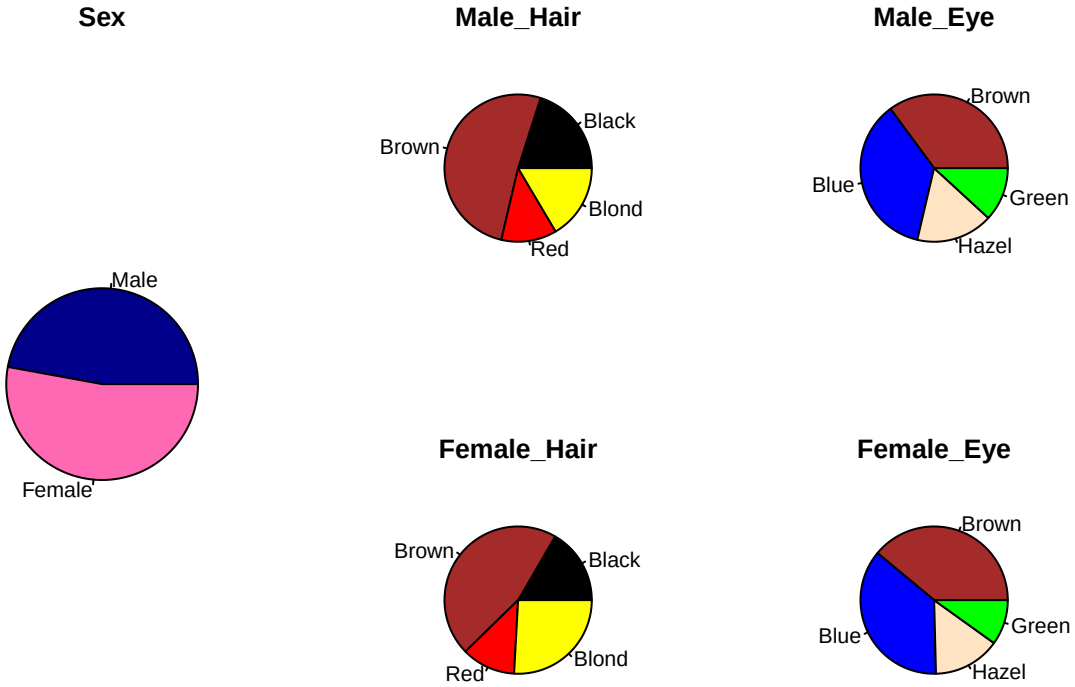
```
  Hazel = c(5, 29, 7, 5),
  Green = c(2, 14, 7, 8)
)

# Print the data frames
print(male_students)
```

```
##    Hair Brown Blue Hazel Green
## 1 Black    32   11    10     3
## 2 Brown    53   50    25    15
## 3   Red    10   10     7     7
## 4 Blond     3   30     5     8
```

```
print(female_students)
```

```
##    Hair Brown Blue Hazel Green
## 1 Black    36    9     5     2
## 2 Brown    66   34    29    14
## 3   Red    16    7     7     7
## 4 Blond     4   64     5     8
```

```
layout(matrix(c(1,1,2,3,4,5), ncol = 3))

pie(c(sum(male_students[,-1]),sum(female_students[,-1])),
    labels = c('Male', 'Female'),
    col = c('darkblue', 'hotpink'),
    main = 'Sex')

pie(rowSums(male_students[,-1]),
    main = 'Male_Hair',
    col = c('black','brown', 'red', 'yellow'),
    labels = c('Black','Brown', 'Red', 'Blond'))

pie(rowSums(female_students[,-1]),
    main = 'Female_Hair',
    col = c('black','brown', 'red', 'yellow'),
    labels = c('Black','Brown', 'Red', 'Blond'))

pie(colSums(male_students[,-1]),
    main = 'Male_Eye',
    col = c('brown','blue', 'bisque', 'green'),
    labels = c('Brown','Blue', 'Hazel', 'Green'))

pie(colSums(female_students[,-1]),
    main = 'Female_Eye',
    col = c('brown','blue', 'bisque', 'green'),
    labels = c('Brown','Blue', 'Hazel', 'Green'))
```



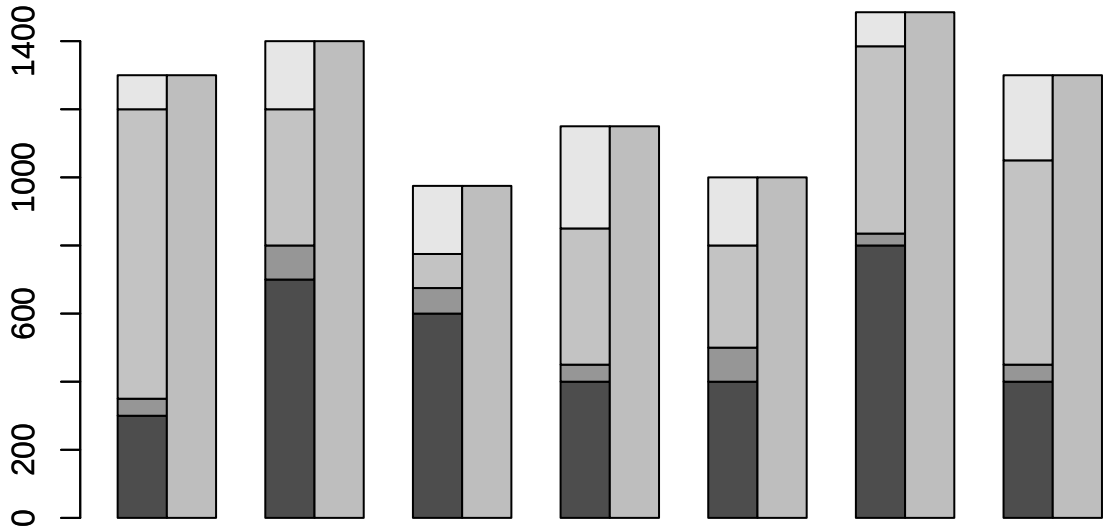
Question 3

```
# Data for Emails
email_data <- data.frame(
  Date = c("1/10", "2/10", "3/10", "4/10", "5/10", "6/10", "7/10"),
  Genuine_emails = c(300, 700, 600, 400, 400, 800, 400),
  Contains_Virus = c(50, 100, 75, 50, 100, 35, 50),
  Spam_emails = c(850, 400, 100, 400, 300, 550, 600),
  Others = c(100, 200, 200, 300, 200, 100, 250)
)

# Print the data frame
print(email_data)
```

```
##   Date Genuine_emails Contains_Virus Spam_emails Others
## 1 1/10            300             50         850    100
## 2 2/10            700            100         400    200
## 3 3/10            600             75         100    200
## 4 4/10            400             50         400    300
## 5 5/10            400            100         300    200
## 6 6/10            800             35         550    100
## 7 7/10            400             50         600    250
```

```
ed = t(as.matrix(email_data[,-1]))
```

```
barplot(ed, space = 2)
barplot(colSums(ed), add = T, space = c(3, rep(2,6)))
```


0
200
600
1000
1400