Statistics for Engineers (MAT2001)- Lab
Experiment-I: Descriptive Statistics

1 Measure of central tendency

1.1 Ungrouped data

Problem:
Twenty students, graduates and undergraduates, were enrolled in a statistics course. Their ages were 18,19,19,19,19,20,20,20,20,20,21,21,21,21,22,23,24,27,30,36. a) Find Mean and Median of all students b) Find median age of all students under 25 years. c) Find modal age of all students

```
x=c(18,19,19,19,19,20,20,20,20,20,21,21,21,21,22,23,24,27,30,36)
print(x)

##  [1] 18 19 19 19 19 20 20 20 20 20 21 21 21 21 22 23 24 27 30 36

print(mean(x))

## [1] 22

print(median(x) )

## [1] 20.5

y=x[x<25]
md=median(y)
print(md)

## [1] 20

#mode
xr=table(x)
xr

## x
## 18 19 20 21 22 23 24 27 30 36
##  1  4  5  4  1  1  1  1  1  1
```

```r
mode=which(xr==max(xr))
print(mode)

## 20
##  3
```

1.2 Grouped data

Problem: A survey of 25 faculty members is taken in a college to study their vocational mobility.They were asked the question "In addition to your present position,at how many educational instistutes have served on the faculty?.Following is the frequency distribution of their responses. Find mean and median of the distribution

X	0	1	2	3
f	8	11	5	1

```r
x=c(0,1,2,3)
f=c(8,11,5,1)
y=rep(x,f)
y

##  [1] 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 3

mean=(sum(y))/(length(y)) #mean
mean

## [1] 0.96

median(y) #median

## [1] 1
```

Problem: Compute mean, median and mode of for the following frequency distribution

Height in cm	145-150	150-155	155-160	160-165	165-170	170-175	175-180	180-185
No. of Adult men	4	6	28	58	64	30	5	5

```r
mid=seq(147.5,182.5,5)
mid

## [1] 147.5 152.5 157.5 162.5 167.5 172.5 177.5 182.5
```

```r
f=c(4,6,28,58,64,30,5,5)
fr_dist=data.frame(mid,f)
fr_dist
```

```
##     mid  f
## 1 147.5  4
## 2 152.5  6
## 3 157.5 28
## 4 162.5 58
## 5 167.5 64
## 6 172.5 30
## 7 177.5  5
## 8 182.5  5
```

```r
mean=(sum(mid*f))/sum(f)
mean
```

```
## [1] 165.175
```

```r
midx=seq(147.5,182.5,5)
frequency=c(4,6,28,58,64,30,5,5)
fr_dist<-data.frame(midx,frequency)
fr_dist
```

```
##    midx frequency
## 1 147.5         4
## 2 152.5         6
## 3 157.5        28
## 4 162.5        58
## 5 167.5        64
## 6 172.5        30
## 7 177.5         5
## 8 182.5         5
```

```r
cl=cumsum(frequency)
cl
```

```
## [1]   4  10  38  96 160 190 195 200
```

```r
fr_dist$cf = cl
fr_dist
```

```
##    midx frequency  cf
## 1 147.5         4   4
## 2 152.5         6  10
## 3 157.5        28  38
## 4 162.5        58  96
```

```
## 5 167.5        64 160
## 6 172.5        30 190
## 7 177.5         5 195
## 8 182.5         5 200
```

```
n=sum(frequency)
n
```

```
## [1] 200
```

```
ml=min(which(cl>=n/2))# The serial number of the median class
ml
```

```
## [1] 5
```

```
h=5
h
```

```
## [1] 5
```

```
f=frequency[ml] #frequency of the median class
f
```

```
## [1] 64
```

```
c=cl[ml-1] # cumulative frequency of the median class
c
```

```
## [1] 96
```

```
l=mid[ml]-h/2
l
```

```
## [1] 165
```

```
median=l+(((n/2)-c)/f)*h #median
median
```

```
## [1] 165.3125
```

```
m=which(frequency==max(frequency)) #serial number of the median class
m
```

```
## [1] 5
```

```
fm=frequency[m] # frequency of the modal class
fm
```

```
## [1] 64
```

```
f1=frequency[m-1] # frequency of the pre modal class
f2=frequency[m+1] # frequency of the post modal class
f1
```

```
## [1] 58
```

```
f2
```

```
## [1] 30
```

```
l=midx[m]-h/2
l
```

```
## [1] 165
```

```
mode=l+((fm-f1)/(2*fm-f1-f2))*h
mode
```

```
## [1] 165.75
```

2 Measure of dispersion

An entomologist studying morphological variation in species of mosquito recorded the following data on body length: 1.2,1.4,1.3,1.6,1.0,1.5,1.7,1.1,1.2,1.3. Compute all the measures of dispersion.

```
x=c(1.2,1.4,1.3,1.6,1.0,1.5,1.7,1.1,1.2,1.3)
x
```

```
##  [1] 1.2 1.4 1.3 1.6 1.0 1.5 1.7 1.1 1.2 1.3
```

```
summary(x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.200   1.300   1.330   1.475   1.700
```

```
range=1.7-1.0 #range
range
```

```
## [1] 0.7
```

```
var(x) #variance
```

```
## [1] 0.049
```

```
sd=sqrt(var(x)) #standard deviation
sd
```

```
## [1] 0.2213594

# There is no separate command for Quartile deviation Mean deviation.
# We have to evaluate the expression coefficient of quartile deviation
qd=1.475-1.2
qd

## [1] 0.275

cqd=(1.475-1.2)/(1.475+1.2)
cqd

## [1] 0.1028037

# Co-efficient of Variation
cv = (sd/mean(x))*100
cv

## [1] 16.64357

# Mean deviation about Mean
y=(x-mean(x))
y

##  [1] -0.13  0.07 -0.03  0.27 -0.33  0.17  0.37 -0.23 -0.13 -0.03

y=abs(y)
y

##  [1] 0.13 0.07 0.03 0.27 0.33 0.17 0.37 0.23 0.13 0.03

mdl=sum(y)/length(y)
mdl

## [1] 0.176

#Mean deviation about Median
z =abs(x-median(x))
md2=sum(z)/length(z)
md2

## [1] 0.17
```

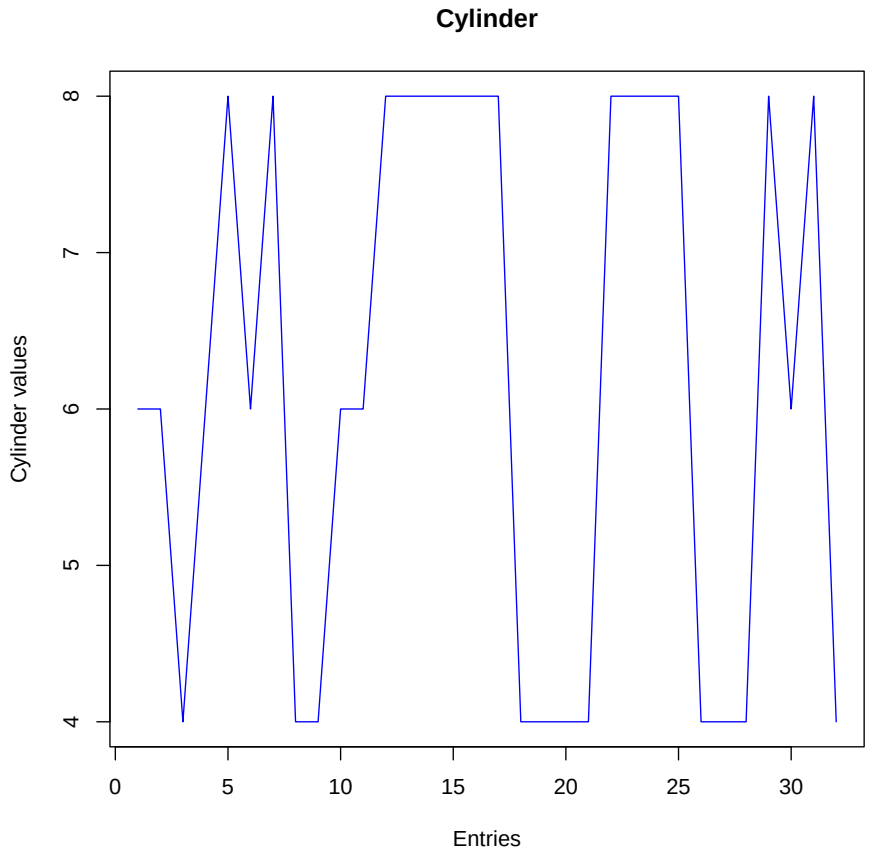
3 Data visualization

```r
data()
```

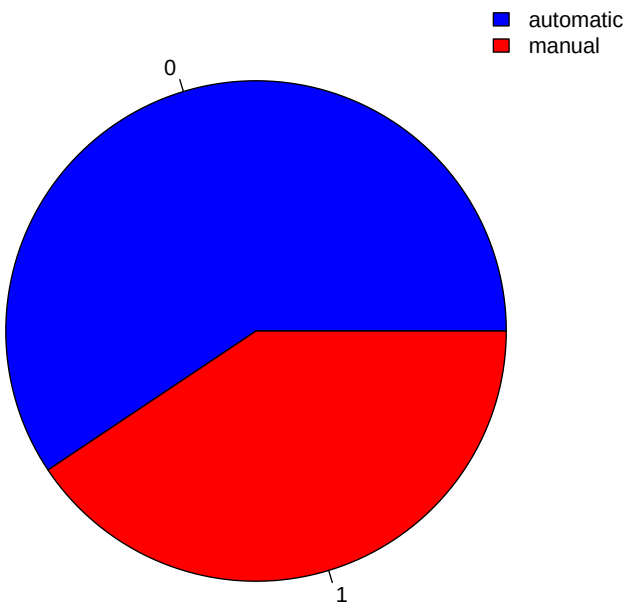
```r
data <- mtcars
summary(data)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

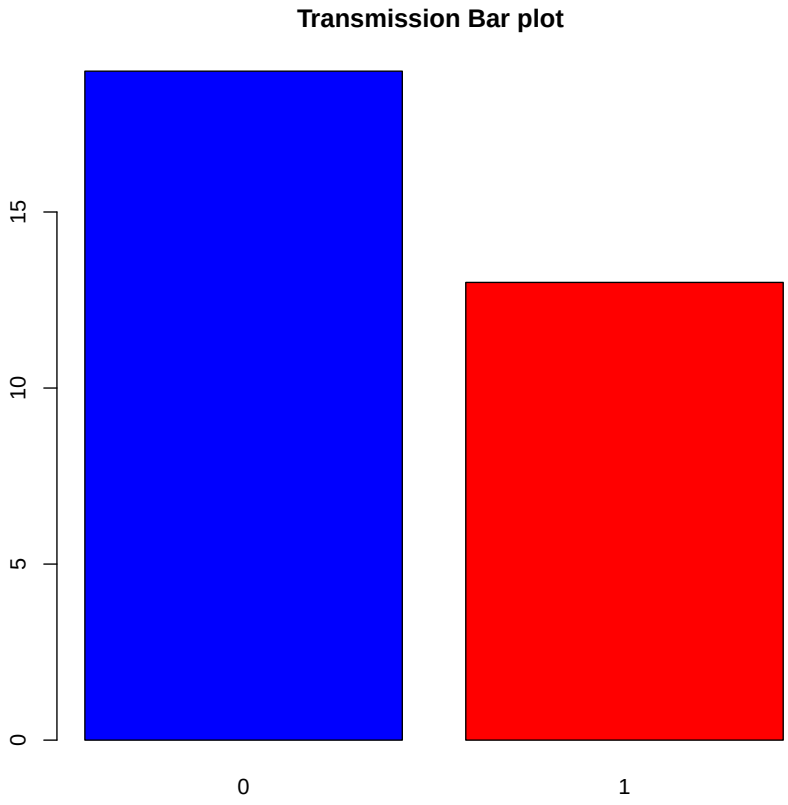
```r
# line plot
plot(data$cyl,type="l",main="Cylinder",xlab="Entries",
     ylab="Cylinder values",col="blue")
```



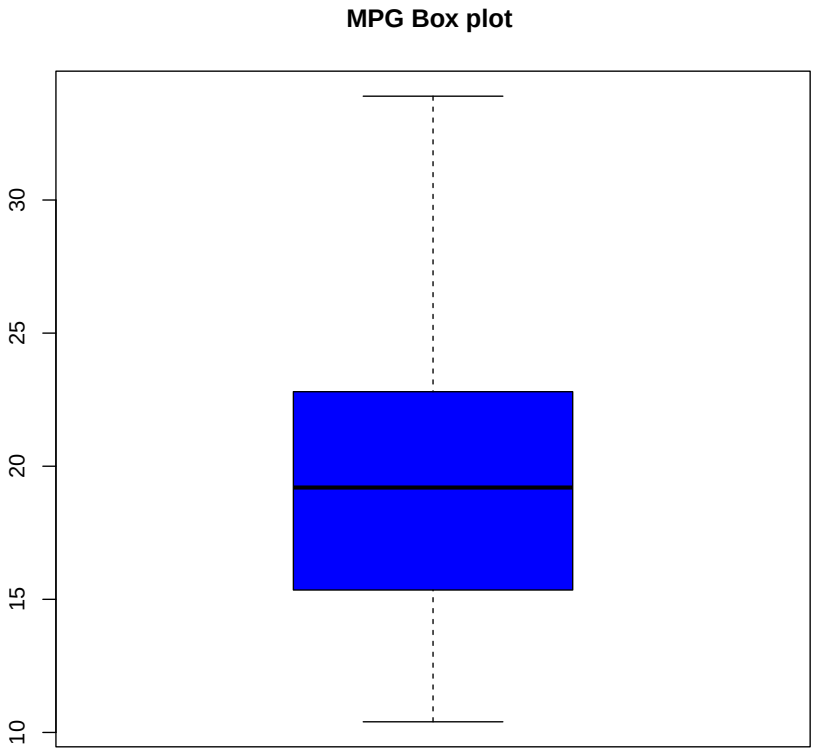

```
tbl_1 = table(data$am)
pie(tbl_1, main = "Transmission Pie chart", col =c('blue','red') )
legend("topright",legend=c("automatic", "manual"),
       fill=c('blue','red'),bty="n")
```



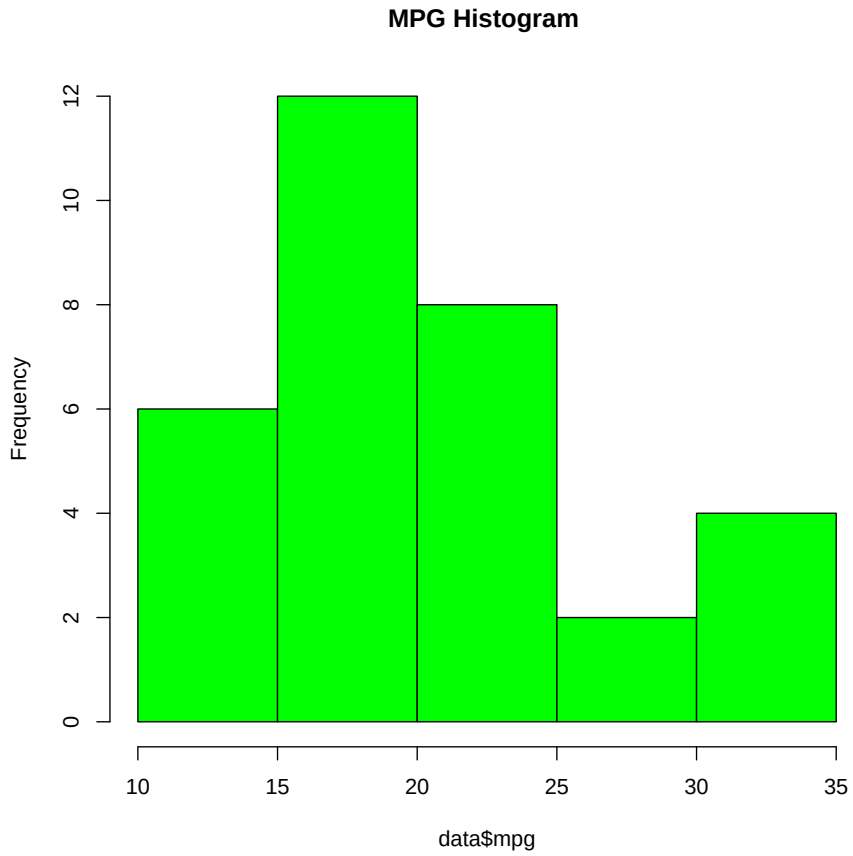

```
# Bar plot
barplot(tbl_1,  main = "Transmission Bar plot ",
        col =c('blue','red'))
```




```
#  Box plot
boxplot(data$mpg,col=c("blue"), main = "MPG Box plot ")
```




```
# Histogram
hist(data$mpg, col=c("green"), main = "MPG Histogram ")
```



4 Exercise problem

1. Calculate the Mean, Median and Mode for the following data. Draw the boxplot with title labels.

x	1	2	3	4	5	6	7	8	9
f	8	10	11	16	20	25	15	9	6

2. Calculate the Mean, Median and Mode for the following data. Draw the histogram with title labels.

Variable	10-13	13-16	16-19	19-22	22-25	25-28	28-31	31-34	34-37
Frequency	8	15	27	51	75	54	36	18	9

3. Import ‘AirPassengers‘ data set. Find the mean and standard deviation for Jan month of every year. Get summary for Sep month of every year. Plot the Feb month of every year