SESSION 7: Basic Statistics

Assignment 1

Table of Contents

1. Introduction

This assignment will help you understand the concepts learnt in the session.

2. Objective

This assignment will test your skills on basic statistics.

3. Prerequisites

Not applicable.

4. Associated Data Files

Not applicable.

5. Problem Statement
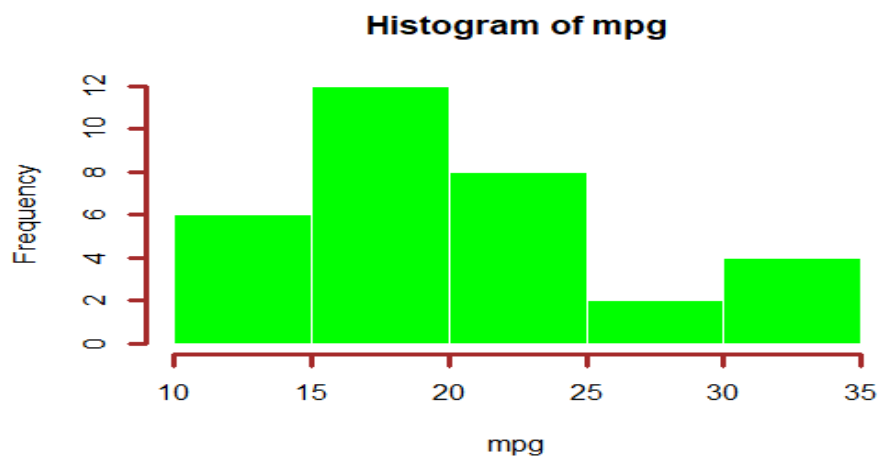
1. **Histogram for all variables in a dataset mtcars.**

   ANS: attach(mtcars)

   names(mtcars)

   View(mtcars)

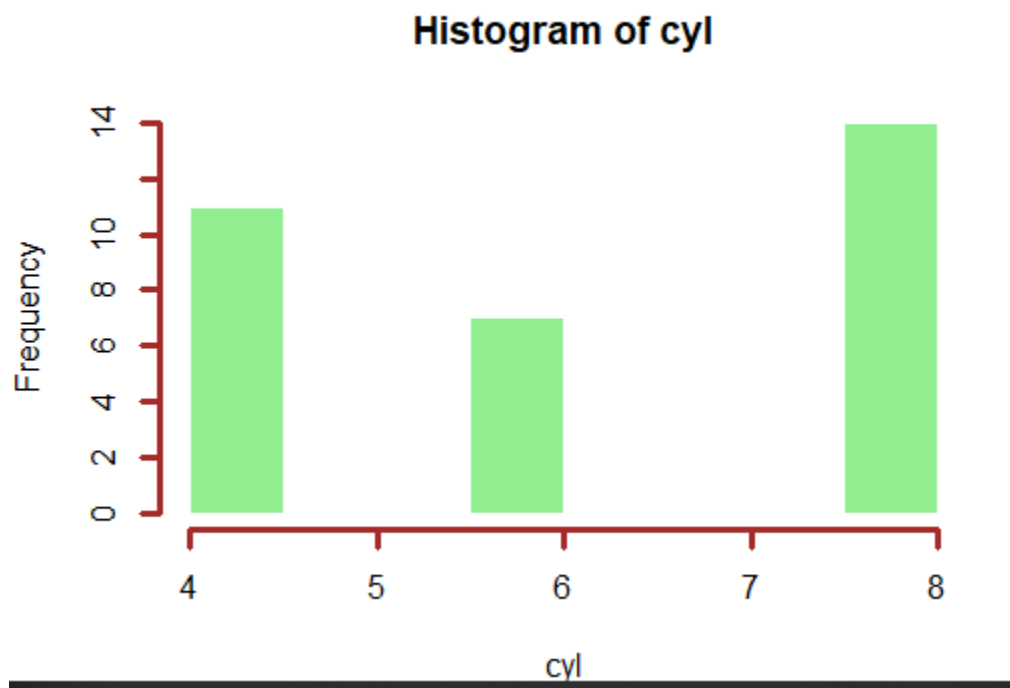   #Histogram of mpg

   hist(mpg, col = 'green', fg='brown', lwd= 3 , border = F)

Histogram of mpg
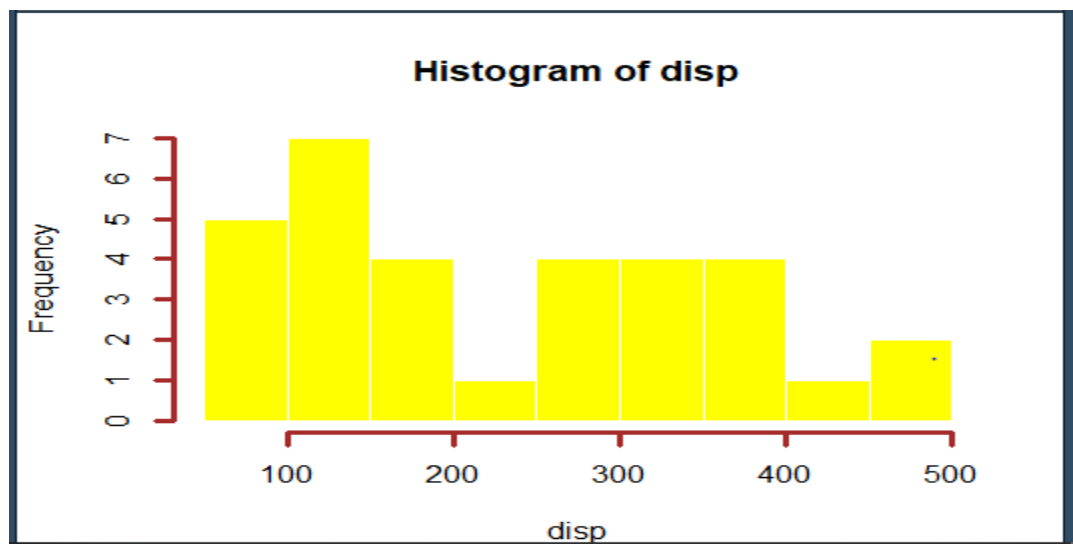
#Histogram of cyl

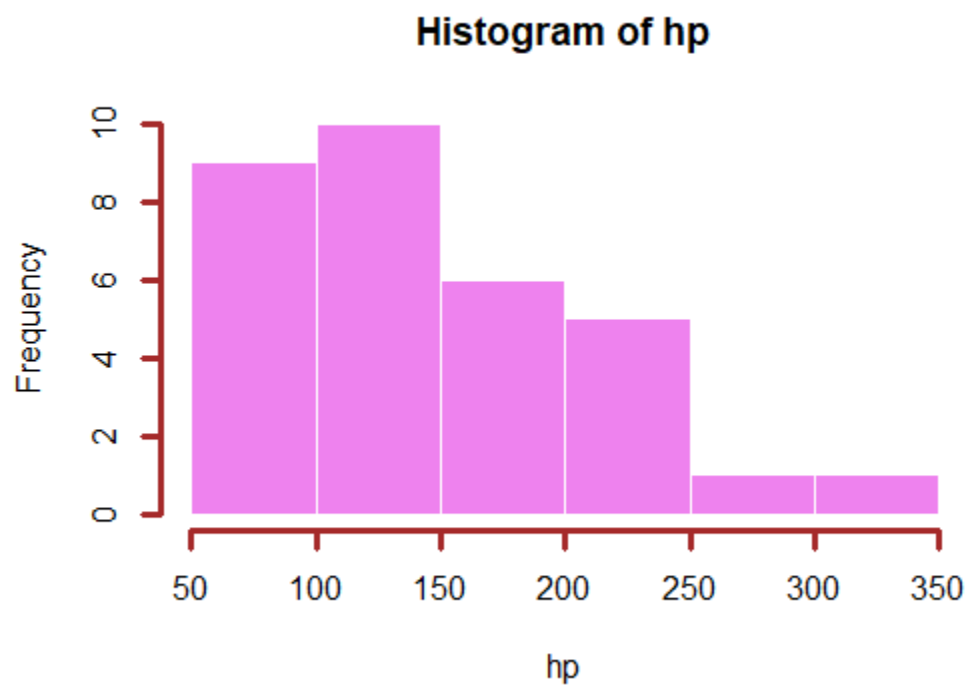hist(cyl, col = 'lightgreen', fg='brown', lwd= 3, border

= F)

Histogram of cyl

#Histogram of disp

hist(disp, col = 'yellow', fg='brown', lwd= 3, border =
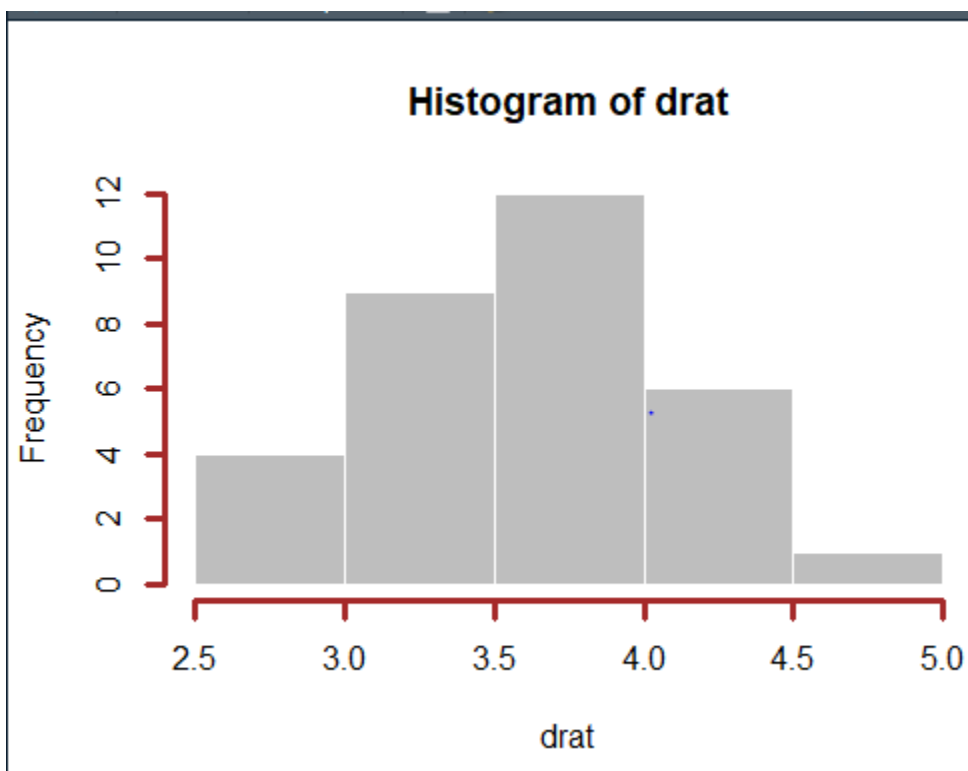
F)

Histogram of disp

#Histogram of hp

hist(hp, col = 'violet', fg='brown', lwd= 3, border = F)
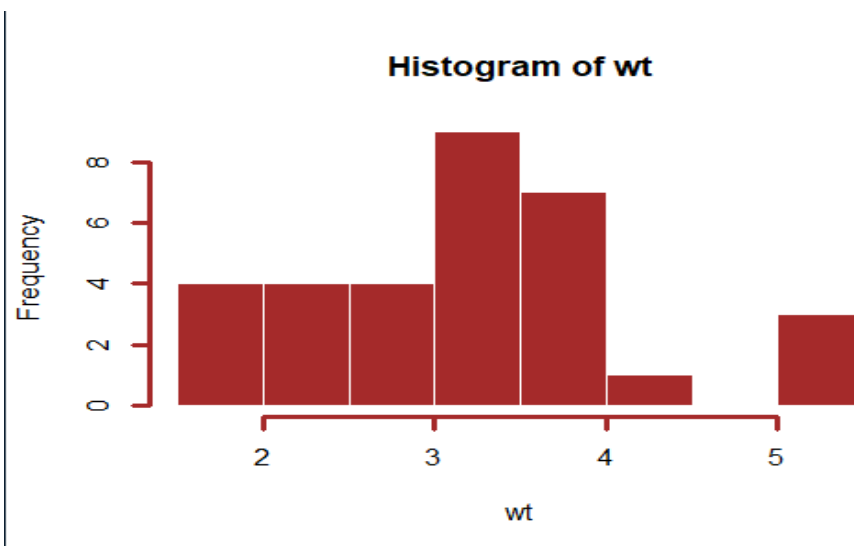
Histogram of hp

#Histogram of drat

hist(drat, col = 'grey', fg='brown', lwd= 3, border = F)

Histogram of drat

#Histogram of wt

hist(wt, col = 'brown', fg='brown', lwd= 3, border = F

)

Histogram of wt

Frequency (y-axis): 0, 2, 4, 6, 8

wt (x-axis): 2, 3, 4, 5

#Histogram of qsec

hist(qsec, col = 'pink', fg='brown', lwd= 3, border = F

)

Histogram of qsec

Frequency (y-axis): 0, 2, 4, 6, 8, 10

qsec (x-axis): 14, 16, 18, 20, 22
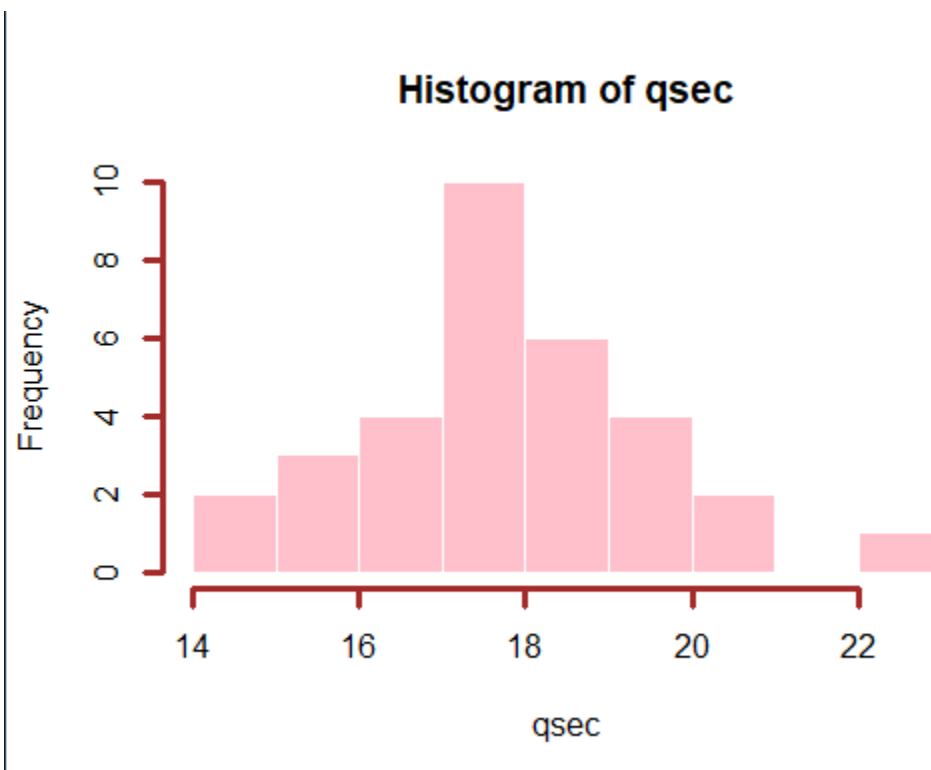
Histogram of vs
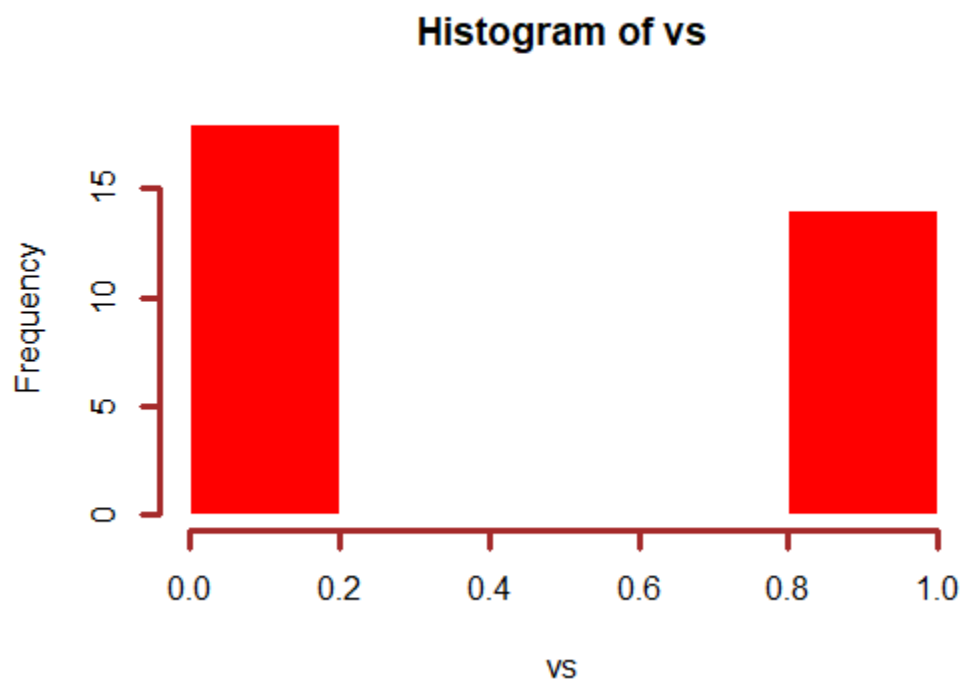
Histogram of am

```
#Histogram of gear

hist(gear, col = 'darkblue', fg='brown', lwd= 3, border

= F )
```
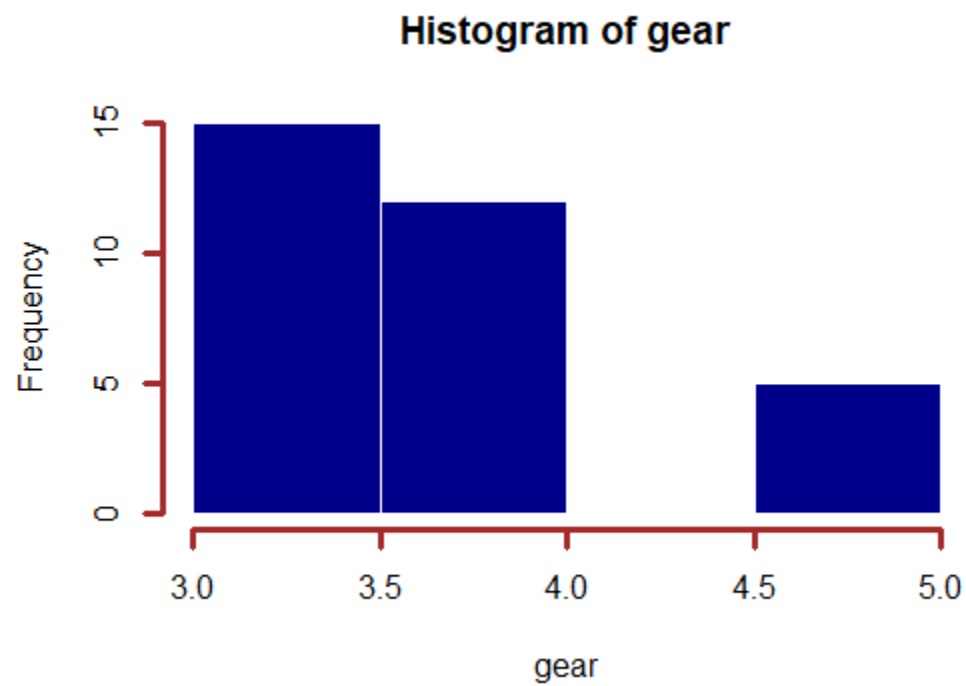
Histogram of gear


```
#Histogram of carb

hist(carb, col = 'orange', fg='brown', lwd= 3, border =

F )
```
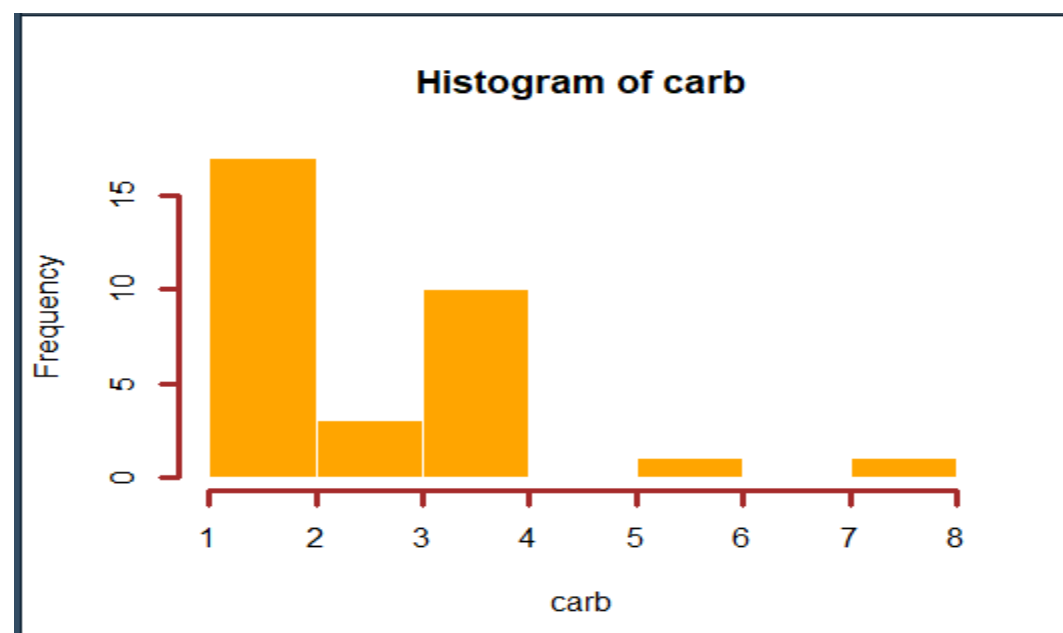
Write a program to create histograms for all columns

ANS: #codes for histogram for all columns

  library(ggplot2)

  library(tidyr)

  library(purrr)

  mtcars %>% keep(is.numeric) %>% gather() %>%

ggplot(aes(value)) + facet_wrap(~ key, scales = 'free')+

  geom_histogram()

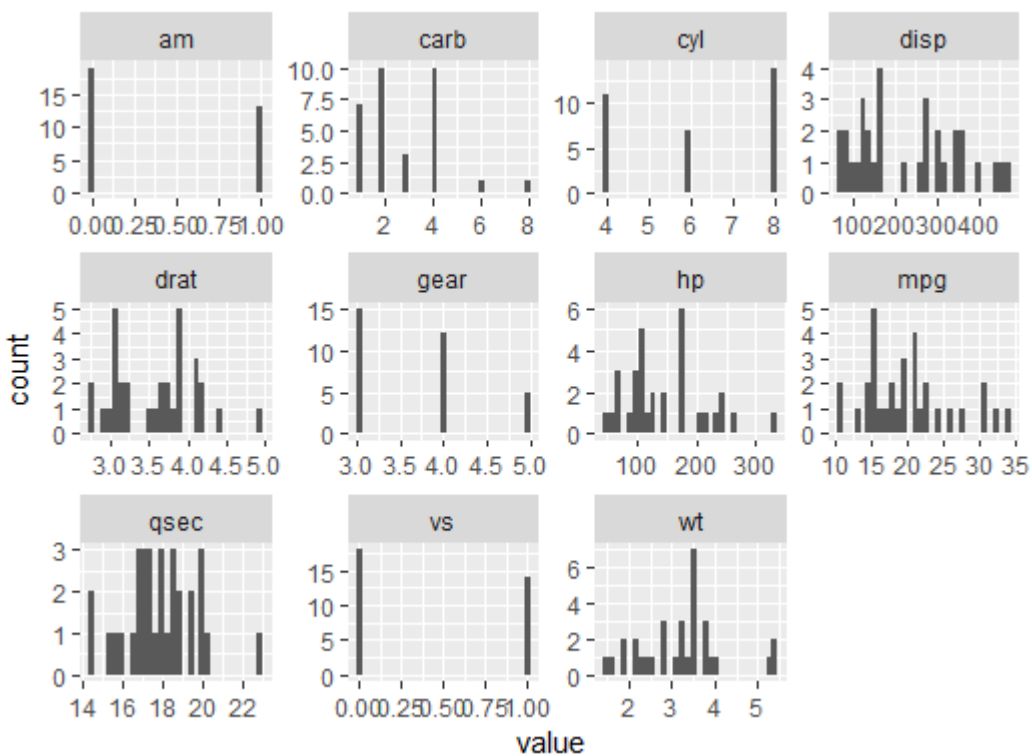## 2. Check the probability distribution of all variables in mtcars.

ANS: #prob of mpg
quantile(mtcars$mpg, probs = c(0.5,0.25,0.50,0.95))

```
> quantile(mtcars$mpg, probs = c(0.5,0.25,0.50,0.95))
   50%    25%    50%    95%
19.200 15.425 19.200 31.300
>
```

#prob of cyl
quantile(mtcars$cyl, probs = c(0.5,0.25,0.50,0.95))

```
> #prob of cyl
> quantile(mtcars$cyl, probs = c(0.5,0.25,0.50,0.95))
50% 25% 50% 95%
  6   4   6   8
>
```

#prob of disp
quantile(mtcars$disp, probs = c(0.5,0.25,0.50,0.95))

```
> #prob of disp
> quantile(mtcars$disp, probs = c(0.5,0.25,0.50,0.95))
    50%     25%     50%     95%
196.300 120.825 196.300 449.000
>
```

#prob of hp
quantile(mtcars$hp, probs = c(0.5,0.25,0.50,0.95))

```
> #prob of hp
> quantile(mtcars$hp, probs = c(0.5,0.25,0.50,0.95))
   50%    25%    50%    95%
123.00  96.50 123.00 253.55
>
```

#prob of wt
quantile(mtcars$wt, probs = c(0.5,0.25,0.50,0.95))

```
> #prob of wt
> quantile(mtcars$wt, probs = c(0.5,0.25,0.50,0.95))
    50%     25%     50%     95%
3.32500 2.58125 3.32500 5.29275
```

#prob of qsec
quantile(mtcars$qsec, probs = c(0.5,0.25,0.50,0.95))

```
> #prob of qsec
> quantile(mtcars$qsec, probs = c(0.5,0.25,0.50,0.95))
    50%     25%     50%     95%
17.7100 16.8925 17.7100 20.1045
>
```

#prob of am
quantile(mtcars$am, probs = c(0.5,0.25,0.50,0.95))

```
> #prob of am
> quantile(mtcars$am, probs = c(0.5,0.25,0.50,0.95))
50% 25% 50% 95%
  0   0   0   1
>
```

#prob of vs
quantile(mtcars$vs, probs = c(0.5,0.25,0.50,0.95))

```
> #prob of vs
> quantile(mtcars$vs, probs = c(0.5,0.25,0.50,0.95))
50% 25% 50% 95%
  0   0   0   1
```

#prob of carb
quantile(mtcars$carb, probs = c(0.5,0.25,0.50,0.95))

```
> #prob of carb
> quantile(mtcars$carb, probs = c(0.5,0.25,0.50,0.95))
50% 25% 50% 95%
2.0 2.0 2.0 4.9
>
```

3. Write a program to create boxplot for all variables.

ANS: #plotting boxplot for all variable
library(reshape2)
df=melt(mtcars)
ggplot(df, aes(x= variable,y=value)) +geom_boxplot()