



ACADGILD

SESSION 7: Basic Statistics

Assignment 1

PROBLEM STATEMENT

1. Histogram for all variables in a dataset **mtcars**. Write a program to create histograms for all columns
2. Check the probability distribution of all variables in **mtcars**.
3. Write a program to create boxplot for all variables.

SOLUTION

1. **Histogram for all variables in a dataset mtcars. Write a program to create histograms for all columns**

The R-script for the given problem is as follows:

```
library(readr)
library(ggplot2)
mtcars <- read_csv("F:/ACADGILD - Online Course/1. DATA
SETS/mtcars.csv")
View(mtcars)
mtcars
str(mtcars)

par(mfrow=c(3,4))      # set the graph area
lapply(mtcars[2:12], hist)
# apply histogram plot function to all column of mtcars
```

The output of the R-Script (from Console window) is given as follows:

```
> library(readr)
> library(ggplot2)
> mtcars <- read_csv("F:/ACADGILD - Online Course/1. DATA
SETS/mtcars.csv")
Parsed with column specification:
cols(
  x1 = col_character(),
  mpg = col_double(),
```

```

cyl = col_double(),
disp = col_double(),
hp = col_double(),
drat = col_double(),
wt = col_double(),
qsec = col_double(),
vs = col_double(),
am = col_double(),
gear = col_double(),
carb = col_double()
)
> View(mtcars)

```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Assignment 7.1.R x mtcars x

Filter

| | X1 | mpg | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb |
|----|---------------------|------|-----|-------|-----|------|-------|-------|----|----|------|------|
| 1 | Mazda RX4 | 21.0 | 6 | 160.0 | 110 | 3.90 | 2.620 | 16.46 | 0 | 1 | 4 | 4 |
| 2 | Mazda RX4 Wag | 21.0 | 6 | 160.0 | 110 | 3.90 | 2.875 | 17.02 | 0 | 1 | 4 | 4 |
| 3 | Datsun 710 | 22.8 | 4 | 108.0 | 93 | 3.85 | 2.320 | 18.61 | 1 | 1 | 4 | 1 |
| 4 | Hornet 4 Drive | 21.4 | 6 | 258.0 | 110 | 3.08 | 3.215 | 19.44 | 1 | 0 | 3 | 1 |
| 5 | Hornet Sportabout | 18.7 | 8 | 360.0 | 175 | 3.15 | 3.440 | 17.02 | 0 | 0 | 3 | 2 |
| 6 | Valiant | 18.1 | 6 | 225.0 | 105 | 2.76 | 3.460 | 20.22 | 1 | 0 | 3 | 1 |
| 7 | Duster 360 | 14.3 | 8 | 360.0 | 245 | 3.21 | 3.570 | 15.84 | 0 | 0 | 3 | 4 |
| 8 | Merc 240D | 24.4 | 4 | 146.7 | 62 | 3.69 | 3.190 | 20.00 | 1 | 0 | 4 | 2 |
| 9 | Merc 230 | 22.8 | 4 | 140.8 | 95 | 3.92 | 3.150 | 22.90 | 1 | 0 | 4 | 2 |
| 10 | Merc 280 | 19.2 | 6 | 167.6 | 123 | 3.92 | 3.440 | 18.30 | 1 | 0 | 4 | 4 |
| 11 | Merc 280C | 17.8 | 6 | 167.6 | 123 | 3.92 | 3.440 | 18.90 | 1 | 0 | 4 | 4 |
| 12 | Merc 450SE | 16.4 | 8 | 275.8 | 180 | 3.07 | 4.070 | 17.40 | 0 | 0 | 3 | 3 |
| 13 | Merc 450SL | 17.3 | 8 | 275.8 | 180 | 3.07 | 3.730 | 17.60 | 0 | 0 | 3 | 3 |
| 14 | Merc 450SLC | 15.2 | 8 | 275.8 | 180 | 3.07 | 3.780 | 18.00 | 0 | 0 | 3 | 3 |
| 15 | Cadillac Fleetwood | 10.4 | 8 | 472.0 | 205 | 2.93 | 5.250 | 17.98 | 0 | 0 | 3 | 4 |
| 16 | Lincoln Continental | 10.4 | 8 | 460.0 | 215 | 3.00 | 5.424 | 17.82 | 0 | 0 | 3 | 4 |
| 17 | Chrysler Imperial | 14.7 | 8 | 440.0 | 230 | 3.23 | 5.345 | 17.42 | 0 | 0 | 3 | 4 |
| 18 | Fiat 128 | 32.4 | 4 | 78.7 | 66 | 4.08 | 2.200 | 19.47 | 1 | 1 | 4 | 1 |
| 19 | Honda Civic | 30.4 | 4 | 75.7 | 52 | 4.93 | 1.615 | 18.52 | 1 | 1 | 4 | 2 |
| 20 | Toyota Corolla | 33.9 | 4 | 71.1 | 65 | 4.22 | 1.835 | 19.90 | 1 | 1 | 4 | 1 |
| 21 | Toyota Corona | 21.5 | 4 | 120.1 | 97 | 3.70 | 2.465 | 20.01 | 1 | 0 | 3 | 1 |
| 22 | Dodge Challenger | 15.5 | 8 | 318.0 | 150 | 2.76 | 3.520 | 16.87 | 0 | 0 | 3 | 2 |
| 23 | AMC Javelin | 15.2 | 8 | 304.0 | 150 | 2.76 | 3.435 | 17.30 | 0 | 0 | 3 | 2 |

Showing 1 to 23 of 32 entries

Console

```

> mtcars
# A tibble: 32 x 12
  x1      mpg  cyl  disp  hp  drat   wt  qsec   vs
am
  <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
<dbl>
1 Mazda RX4    21      6  160   110  3.9   2.62  16.5    0
1
2 Mazda RX4~   21      6  160   110  3.9   2.88  17.0    0
1
3 Datsun 710   22.8     4  108    93  3.85  2.32  18.6    1
1
4 Hornet 4 ~   21.4     6  258   110  3.08  3.22  19.4    1
0
5 Hornet Sp~   18.7     8  360   175  3.15  3.44  17.0    0
0
6 Valiant     18.1     6  225   105  2.76  3.46  20.2    1
0
7 Duster 360   14.3     8  360   245  3.21  3.57  15.8    0
0
8 Merc 240D    24.4     4  147.    62  3.69  3.19  20      1
0
9 Merc 230     22.8     4  141.    95  3.92  3.15  22.9    1
0
10 Merc 280    19.2     6  168.   123  3.92  3.44  18.3    1
0
# ... with 22 more rows, and 2 more variables: gear <dbl>,
#   carb <dbl>
> str(mtcars)
Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 32 obs.
of 12 variables:
 $ x1 : chr  "Mazda RX4" "Mazda RX4 wag" "Datsun 710" "Hornet 4
Drive" ...
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs : num  0 0 1 1 0 1 0 1 1 1 ...

```

```

$ am : num 1 1 1 0 0 0 0 0 0 0 ...
$ gear: num 4 4 4 3 3 3 3 4 4 4 ...
$ carb: num 4 4 1 1 2 1 4 2 2 4 ...
- attr(*, "spec")=
.. cols(
..   x1 = col_character(),
..   mpg = col_double(),
..   cyl = col_double(),
..   disp = col_double(),
..   hp = col_double(),
..   drat = col_double(),
..   wt = col_double(),
..   qsec = col_double(),
..   vs = col_double(),
..   am = col_double(),
..   gear = col_double(),
..   carb = col_double()
.. )
> par(mfrow=c(3,4))           # set the graph area
> lapply(mtcars[2:12], hist)  # apply histogram plot function to
all column of mtcars
$mpg
$breaks
[1] 10 15 20 25 30 35

$counts
[1] 6 12 8 2 4

$density
[1] 0.0375 0.0750 0.0500 0.0125 0.0250

$mids
[1] 12.5 17.5 22.5 27.5 32.5

$xname
[1] "x[[i]]"

$equidist
[1] TRUE

attr("class")

```

```
[1] "histogram"
```

```
$cyl
```

```
$breaks
```

```
[1] 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0
```

```
$counts
```

```
[1] 11  0  0  7  0  0  0 14
```

```
$density
```

```
[1] 0.6875 0.0000 0.0000 0.4375 0.0000 0.0000 0.0000 0.8750
```

```
$mids
```

```
[1] 4.25 4.75 5.25 5.75 6.25 6.75 7.25 7.75
```

```
$xname
```

```
[1] "x[[i]]"
```

```
$equidist
```

```
[1] TRUE
```

```
attr("class")
```

```
[1] "histogram"
```

```
$disp
```

```
$breaks
```

```
[1]  50 100 150 200 250 300 350 400 450 500
```

```
$counts
```

```
[1] 5 7 4 1 4 4 4 1 2
```

```
$density
```

```
[1] 0.003125 0.004375 0.002500 0.000625 0.002500 0.002500  
0.002500
```

```
[8] 0.000625 0.001250
```

```
$mids
```

```
[1]  75 125 175 225 275 325 375 425 475
```

```
$xname
```

```
[1] "x[[i]]"
```

```
$equidist
```

```
[1] TRUE
```

```
attr("class")
```

```
[1] "histogram"
```

```
$hp
```

```
$breaks
```

```
[1] 50 100 150 200 250 300 350
```

```
$counts
```

```
[1] 9 10 6 5 1 1
```

```
$density
```

```
[1] 0.005625 0.006250 0.003750 0.003125 0.000625 0.000625
```

```
$mids
```

```
[1] 75 125 175 225 275 325
```

```
$xname
```

```
[1] "x[[i]]"
```

```
$equidist
```

```
[1] TRUE
```

```
attr("class")
```

```
[1] "histogram"
```

```
$drat
```

```
$breaks
```

```
[1] 2.5 3.0 3.5 4.0 4.5 5.0
```

```
$counts
```

```
[1] 4 9 12 6 1
```

```
$density
```

```
[1] 0.2500 0.5625 0.7500 0.3750 0.0625
```

```
$mids
```

```
[1] 2.75 3.25 3.75 4.25 4.75
```

```
$xname  
[1] "x[[i]]"
```

```
$equidist  
[1] TRUE
```

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

```
$wt  
$breaks  
[1] 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5
```

```
$counts  
[1] 4 4 4 9 7 1 0 3
```

```
$density  
[1] 0.2500 0.2500 0.2500 0.5625 0.4375 0.0625 0.0000 0.1875
```

```
$mids  
[1] 1.75 2.25 2.75 3.25 3.75 4.25 4.75 5.25
```

```
$xname  
[1] "x[[i]]"
```

```
$equidist  
[1] TRUE
```

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

```
$qsec  
$breaks  
[1] 14 15 16 17 18 19 20 21 22 23
```

```
$counts  
[1] 2 3 4 10 6 4 2 0 1
```

```
$density  
[1] 0.06250 0.09375 0.12500 0.31250 0.18750 0.12500 0.06250  
0.00000
```



```
[9] 0.03125
```

```
$mids
```

```
[1] 14.5 15.5 16.5 17.5 18.5 19.5 20.5 21.5 22.5
```

```
$xname
```

```
[1] "x[[i]]"
```

```
$equidist
```

```
[1] TRUE
```

```
attr("class")
```

```
[1] "histogram"
```

```
$vs
```

```
$breaks
```

```
[1] 0.0 0.2 0.4 0.6 0.8 1.0
```

```
$counts
```

```
[1] 18  0  0  0 14
```

```
$density
```

```
[1] 2.8125 0.0000 0.0000 0.0000 2.1875
```

```
$mids
```

```
[1] 0.1 0.3 0.5 0.7 0.9
```

```
$xname
```

```
[1] "x[[i]]"
```

```
$equidist
```

```
[1] TRUE
```

```
attr("class")
```

```
[1] "histogram"
```

```
$am
```

```
$breaks
```

```
[1] 0.0 0.2 0.4 0.6 0.8 1.0
```

\$counts

[1] 19 0 0 0 13

\$density

[1] 2.96875 0.00000 0.00000 0.00000 2.03125

\$mids

[1] 0.1 0.3 0.5 0.7 0.9

\$xname

[1] "x[[i]]"

\$equidist

[1] TRUE

attr("class")

[1] "histogram"

\$gear

\$breaks

[1] 3.0 3.5 4.0 4.5 5.0

\$counts

[1] 15 12 0 5

\$density

[1] 0.9375 0.7500 0.0000 0.3125

\$mids

[1] 3.25 3.75 4.25 4.75

\$xname

[1] "x[[i]]"

\$equidist

[1] TRUE

attr("class")

[1] "histogram"

\$carb

\$breaks

[1] 1 2 3 4 5 6 7 8

\$counts

[1] 17 3 10 0 1 0 1

\$density

[1] 0.53125 0.09375 0.31250 0.00000 0.03125 0.00000 0.03125

\$mids

[1] 1.5 2.5 3.5 4.5 5.5 6.5 7.5

\$xname

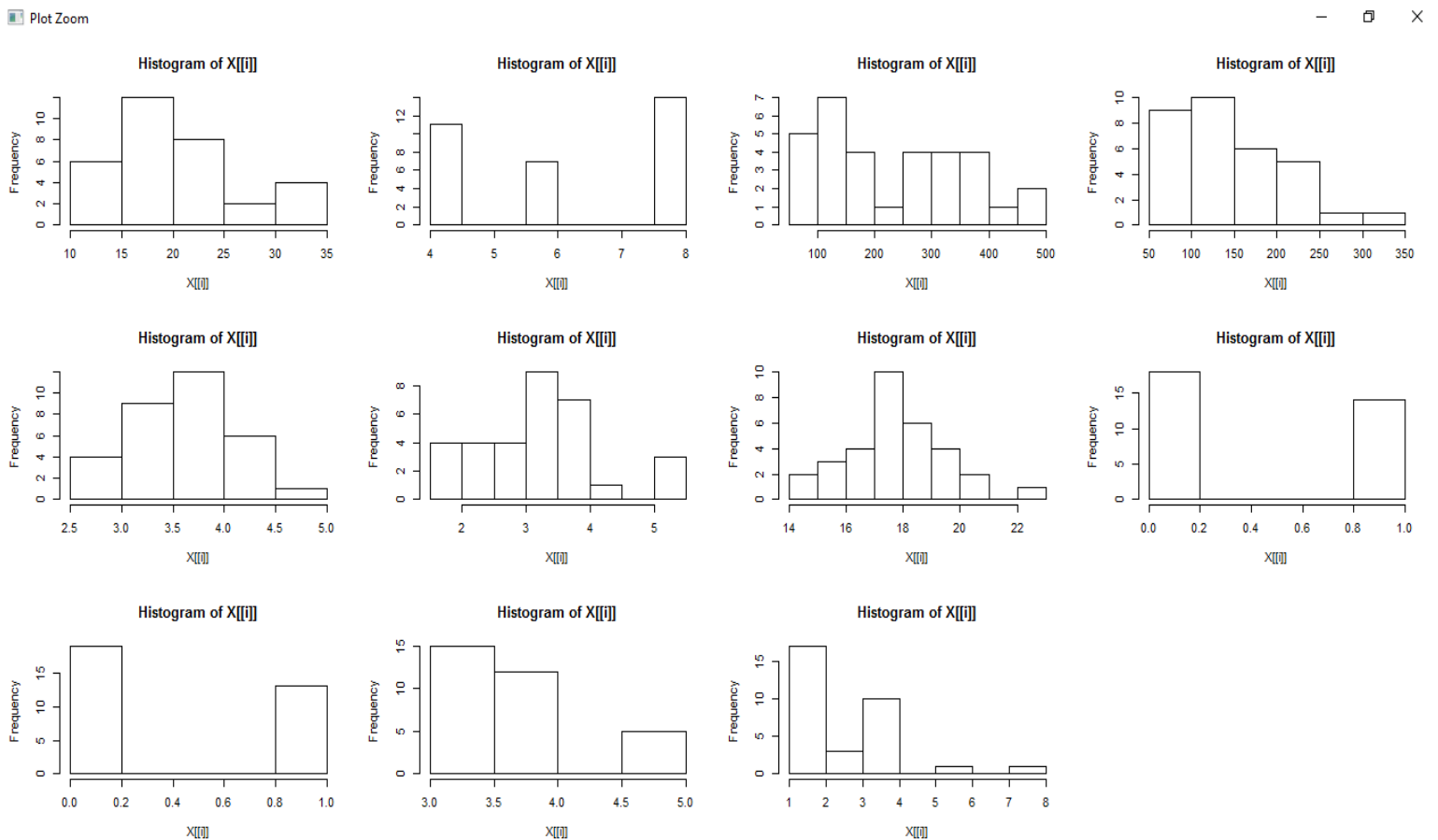
[1] "X[[i]]"

\$equidist

[1] TRUE

attr("class")

[1] "histogram"



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

The R-script for the given problem is as follows:

```
par(mfrow=c(3,4)) # set the graph area
# writing a function to plot probability
prob <- function(prob){
  x <- sort(prob)
  hx <- dnorm(prob)
  p <- plot(x, hx, type="l")
}
lapply(mtcars[2:12], prob) # applying the function to all the columns
```

The output of the R-Script (from Console window/Plot window) is given as follows:

```
> # writing a function to plot probability
> prob <- function(prob){
+   x <- sort(prob)
+   hx <- dnorm(prob)
+   p <- plot(x, hx, type="l")
+ }
> lapply(mtcars[2:12], prob) # applying the function to all the
columns
$mpg
NULL

$cyl
NULL

$disp
NULL

$hp
NULL

$drat
NULL

$wt
NULL
```

\$qsec

NULL

\$vs

NULL

\$am

NULL

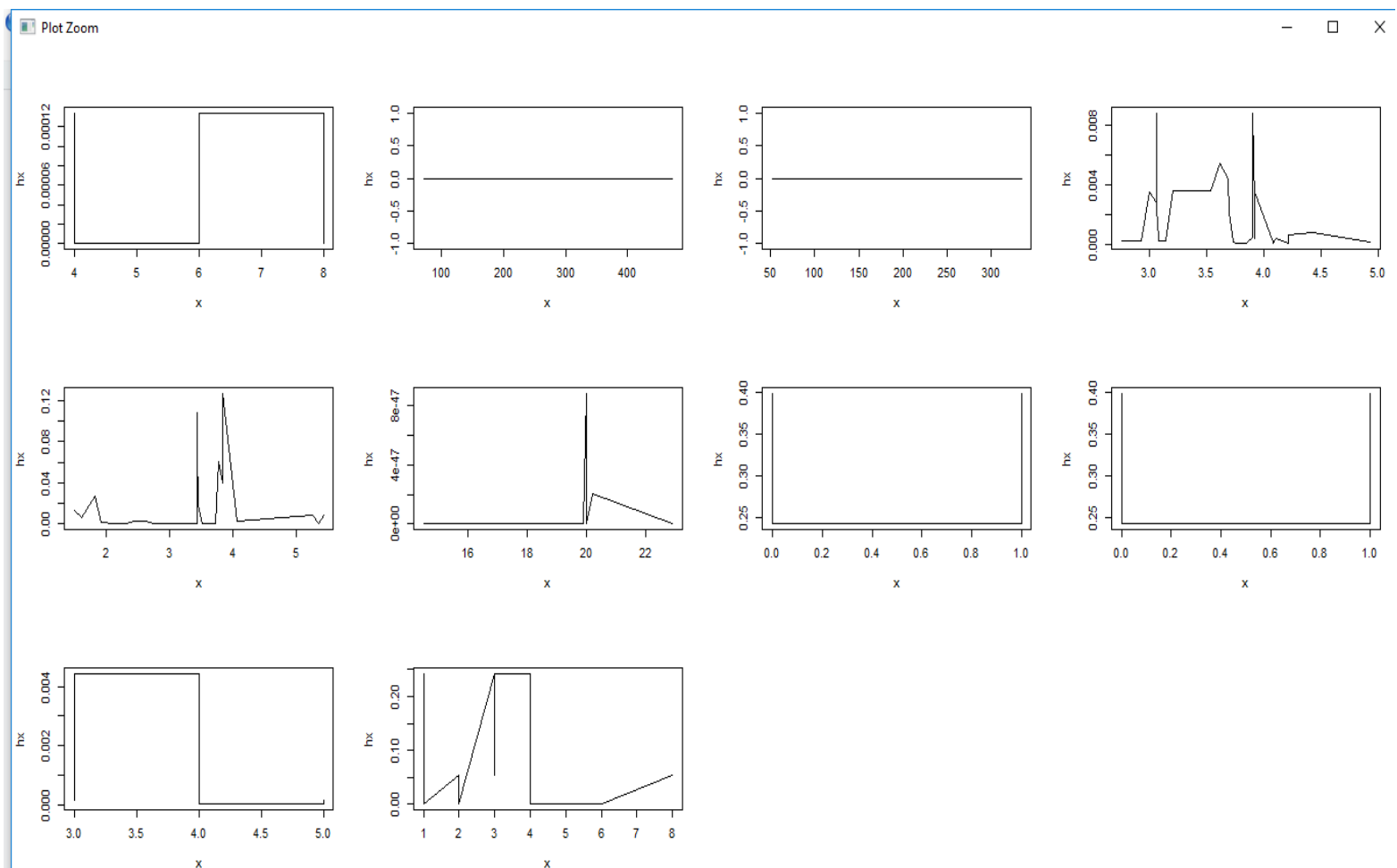
\$gear

NULL

\$carb

NULL

The output of the R-Script (from Plot window) is given as follows:



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

The R-script for the given problem is as follows:

```
par(mfrow=c(3,4))  
lapply(mtcars[2:12], boxplot) # applying the function to all the columns
```

The output of the R-Script (from Console window/Plot window) is given as follows:

```
> par(mfrow=c(3,4))  
> lapply(mtcars[2:12], boxplot) # applying the function to all  
the columns  
$mpg  
$mpg$stats  
      [,1]  
[1,] 10.40  
[2,] 15.35  
[3,] 19.20  
[4,] 22.80  
[5,] 33.90  
  
$mpg$n  
[1] 32  
  
$mpg$conf  
      [,1]  
[1,] 17.11916  
[2,] 21.28084  
  
$mpg$out  
numeric(0)  
  
$mpg$group  
numeric(0)  
  
$mpg$names  
[1] "1"  
  
$cyl
```

```
$cyl$stats
```

```
      [,1]
```

```
[1,]      4
```

```
[2,]      4
```

```
[3,]      6
```

```
[4,]      8
```

```
[5,]      8
```

```
$cyl$n
```

```
[1] 32
```

```
$cyl$conf
```

```
      [,1]
```

```
[1,] 4.882771
```

```
[2,] 7.117229
```

```
$cyl$out
```

```
numeric(0)
```

```
$cyl$group
```

```
numeric(0)
```

```
$cyl$names
```

```
[1] "1"
```

```
$disp
```

```
$disp$stats
```

```
      [,1]
```

```
[1,]  71.10
```

```
[2,] 120.65
```

```
[3,] 196.30
```

```
[4,] 334.00
```

```
[5,] 472.00
```

```
$disp$n
```

```
[1] 32
```

```
$disp$conf
```

```
      [,1]
```

```
[1,] 136.7098
```

```
[2,] 255.8902
```

```
$disp$out  
numeric(0)
```

```
$disp$group  
numeric(0)
```

```
$disp$names  
[1] "1"
```

```
$hp  
$hp$stats  
      [,1]  
[1,]    52  
[2,]    96  
[3,]   123  
[4,]   180  
[5,]   264
```

```
$hp$n  
[1] 32
```

```
$hp$conf  
      [,1]  
[1,] 99.5382  
[2,] 146.4618
```

```
$hp$out  
[1] 335
```

```
$hp$group  
[1] 1
```

```
$hp$names  
[1] "1"
```

```
$drat  
$drat$stats  
      [,1]  
[1,] 2.760  
[2,] 3.080
```



```
[3,] 3.695
[4,] 3.920
[5,] 4.930
```

```
$drat$n
```

```
[1] 32
```

```
$drat$conf
```

```
      [,1]
```

```
[1,] 3.460382
```

```
[2,] 3.929618
```

```
$drat$out
```

```
numeric(0)
```

```
$drat$group
```

```
numeric(0)
```

```
$drat$names
```

```
[1] "1"
```

```
$wt
```

```
$wt$stats
```

```
      [,1]
```

```
[1,] 1.5130
```

```
[2,] 2.5425
```

```
[3,] 3.3250
```

```
[4,] 3.6500
```

```
[5,] 5.2500
```

```
$wt$n
```

```
[1] 32
```

```
$wt$conf
```

```
      [,1]
```

```
[1,] 3.015667
```

```
[2,] 3.634333
```

```
$wt$out
```

```
[1] 5.424 5.345
```

```
$wt$group
```

```
[1] 1 1
```

```
$wt$names
```

```
[1] "1"
```

```
$qsec
```

```
$qsec$stats
```

```
      [,1]
```

```
[1,] 14.500
```

```
[2,] 16.885
```

```
[3,] 17.710
```

```
[4,] 18.900
```

```
[5,] 20.220
```

```
$qsec$n
```

```
[1] 32
```

```
$qsec$conf
```

```
      [,1]
```

```
[1,] 17.1472
```

```
[2,] 18.2728
```

```
$qsec$out
```

```
[1] 22.9
```

```
$qsec$group
```

```
[1] 1
```

```
$qsec$names
```

```
[1] "1"
```

```
$vs
```

```
$vs$stats
```

```
      [,1]
```

```
[1,] 0
```

```
[2,] 0
```

```
[3,] 0
```

```
[4,] 1
```

```
[5,] 1
```

\$vs\$n

[1] 32

\$vs\$conf

[,1]

[1,] -0.2793072

[2,] 0.2793072

\$vs\$out

numeric(0)

\$vs\$group

numeric(0)

\$vs\$names

[1] "1"

\$am

\$am\$stats

[,1]

[1,] 0

[2,] 0

[3,] 0

[4,] 1

[5,] 1

\$am\$n

[1] 32

\$am\$conf

[,1]

[1,] -0.2793072

[2,] 0.2793072

\$am\$out

numeric(0)

\$am\$group

numeric(0)

\$am\$names

```
[1] "1"
```

```
$gear
```

```
$gear$stats
```

```
      [,1]
```

```
[1,]      3
```

```
[2,]      3
```

```
[3,]      4
```

```
[4,]      4
```

```
[5,]      5
```

```
$gear$n
```

```
[1] 32
```

```
$gear$conf
```

```
      [,1]
```

```
[1,] 3.720693
```

```
[2,] 4.279307
```

```
$gear$out
```

```
numeric(0)
```

```
$gear$group
```

```
numeric(0)
```

```
$gear$names
```

```
[1] "1"
```

```
$carb
```

```
$carb$stats
```

```
      [,1]
```

```
[1,]      1
```

```
[2,]      2
```

```
[3,]      2
```

```
[4,]      4
```

```
[5,]      6
```

```
$carb$n
```

```
[1] 32
```

```
$carb$conf
```

```
      [,1]  
[1,] 1.441386  
[2,] 2.558614
```

```
$carb$out  
[1] 8
```

```
$carb$group  
[1] 1
```

```
$carb$names  
[1] "1"
```

The output of the R-Script (from Plot window) is given as follows:

