R Assingment 2 - MTCars

Surya Bhosale

29/03/2020

Downloading Packages:

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(explore)
```

Loading Data:

```
data(mtcars)
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

Correspondence of the 11 numeric variables:

mpg : Miles/(US) gallon cyl : Number of cylinders disp : Displacement (cu.in.) hp : Gross horsepower drat : Rear axle ratio wt : Weight (1000 lbs) qsec : 1/4 mile time vs : Engine (0 = V-shaped, 1 = straight) am : Transmission (0 = automatic, 1 = manual) gear : Number of forward gears carb : Number of carburetors

Displaying Strucutre And Contents:

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ 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 ...
```

```
structure(mtcars)
```

```
##                      mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4           21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag       21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710          22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive      21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout   18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
## Valiant             18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Merc 240D           24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
## Merc 230            22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
## Merc 280            19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
## Merc 280C           17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
## Merc 450SE          16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
## Merc 450SL          17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
## Merc 450SLC         15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Toyota Corona       21.5   4 120.1  97 3.70 2.465 20.01  1  0    3    1
## Dodge Challenger    15.5   8 318.0 150 2.76 3.520 16.87  0  0    3    2
## AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
## Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
```

Making Carnames a variable in new data set called mtcars1:

```
mtcars1<- add_rownames(mtcars, "Carnames")
```

```
## Warning: Deprecated, use tibble::rownames_to_column() instead.
```

```
structure(mtcars1)
```

```
## # A tibble: 32 x 12
##    Carnames   mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear
##    <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 Mazda R~  21       6  160    110  3.9   2.62  16.5     0     1     4
##  2 Mazda R~  21       6  160    110  3.9   2.88  17.0     0     1     4
##  3 Datsun ~  22.8     4  108     93  3.85  2.32  18.6     1     1     4
##  4 Hornet ~  21.4     6  258    110  3.08  3.22  19.4     1     0     3
##  5 Hornet ~  18.7     8  360    175  3.15  3.44  17.0     0     0     3
##  6 Valiant   18.1     6  225    105  2.76  3.46  20.2     1     0     3
##  7 Duster ~  14.3     8  360    245  3.21  3.57  15.8     0     0     3
##  8 Merc 24~  24.4     4  147.    62  3.69  3.19  20       1     0     4
##  9 Merc 230  22.8     4  141.    95  3.92  3.15  22.9     1     0     4
```

```
## 10 Merc 280  19.2     6  168.   123  3.92  3.44  18.3     1     0     4
## # ... with 22 more rows, and 1 more variable: carb <dbl>
```

```
Transmission = rep("Manual", nrow(mtcars1))
Transmission[mtcars1$am == 0] = "Automatic"
mtcars1$Transmission = Transmission
```

The dataset mtcars1 is initalised as a sister dataset in case we would like to use carnames as a variable is our EDA.

Initialising EDA:

```
summary(mtcars)
```

```
##       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
```

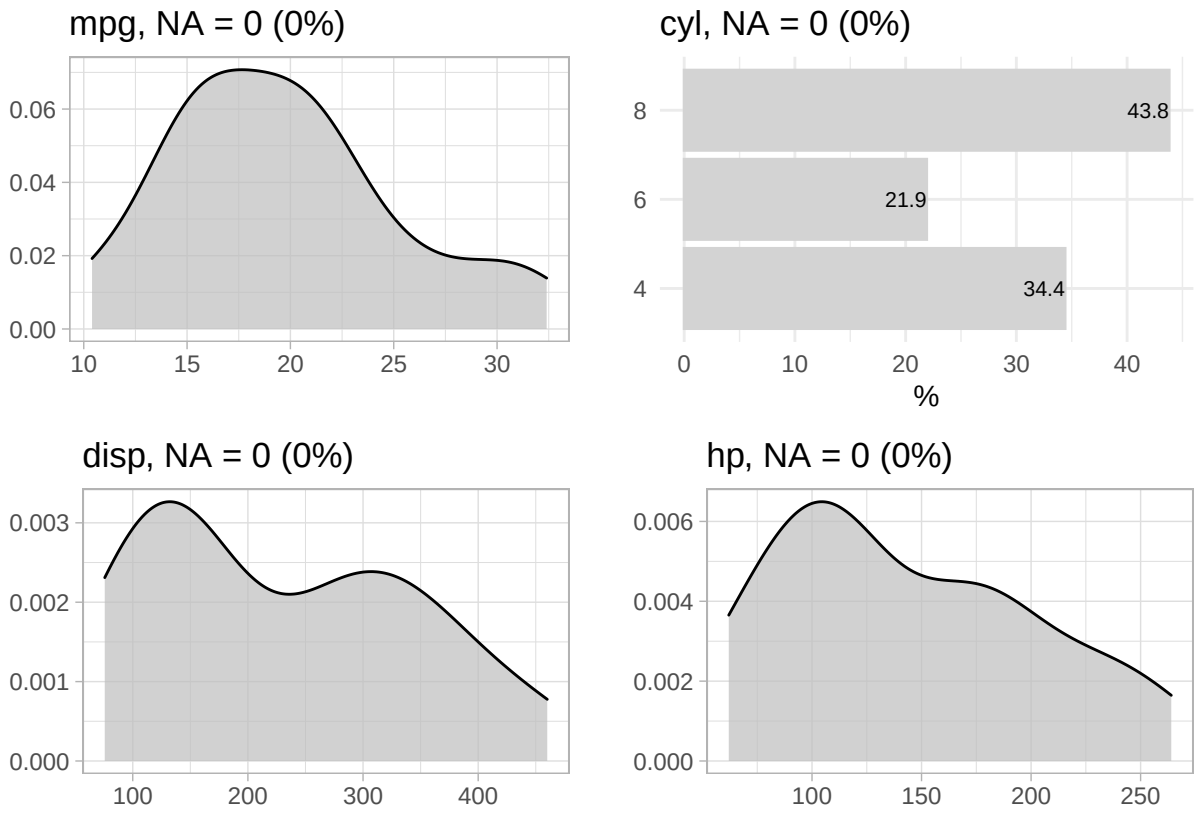
```
mtcars%>%
  describe()
```

```
## # A tibble: 11 x 8
##    variable type     na na_pct unique   min   mean    max
##    <chr>    <chr> <int>  <dbl>  <int> <dbl>  <dbl>  <dbl>
##  1 mpg      dbl       0      0     25 10.4   20.1   33.9
##  2 cyl      dbl       0      0      3  4      6.19   8
##  3 disp     dbl       0      0     27 71.1  231.   472
##  4 hp       dbl       0      0     22 52    147.   335
##  5 drat     dbl       0      0     22  2.76   3.6    4.93
##  6 wt       dbl       0      0     29  1.51   3.22   5.42
##  7 qsec     dbl       0      0     30 14.5   17.8   22.9
##  8 vs       dbl       0      0      2  0      0.44   1
##  9 am       dbl       0      0      2  0      0.41   1
## 10 gear     dbl       0      0      3  3      3.69   5
## 11 carb     dbl       0      0      6  1      2.81   8
```

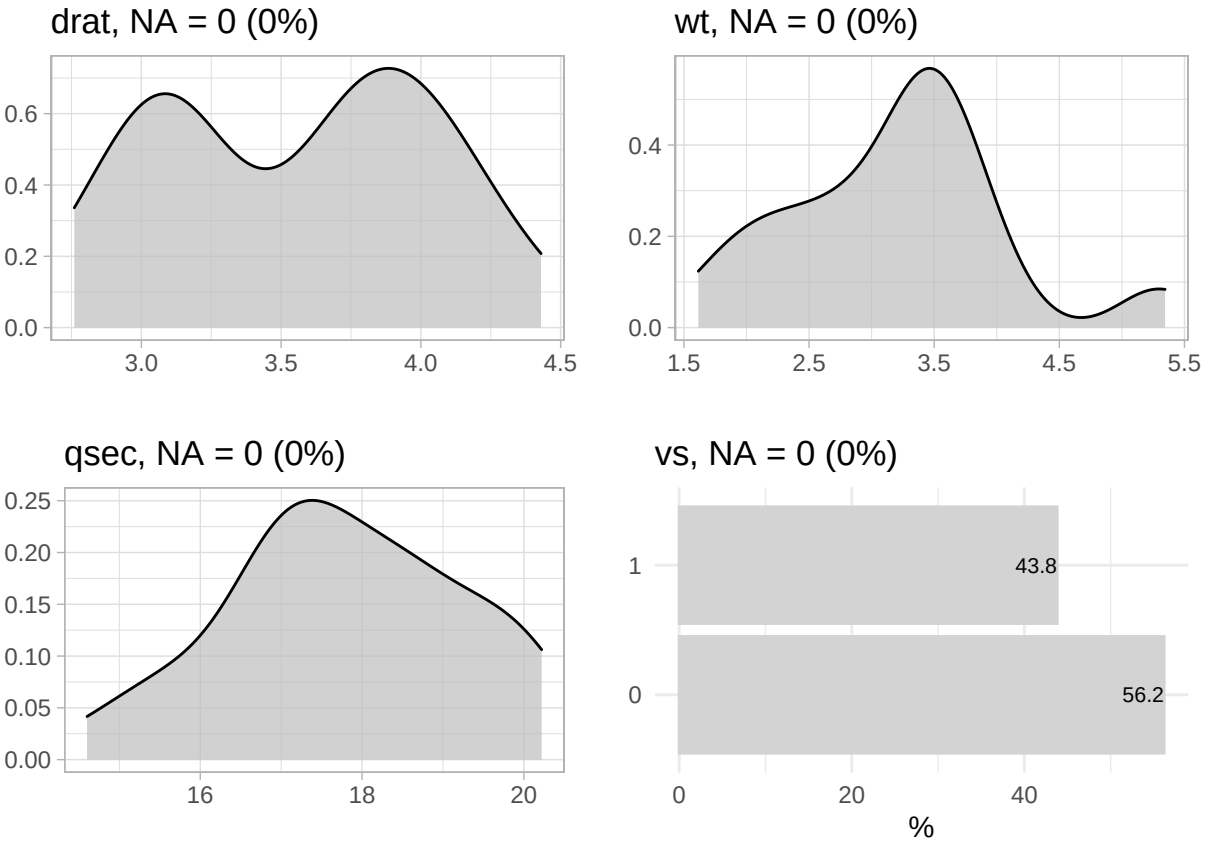
Figure 1 :

```
mtcars%>%
  select(mpg, cyl, disp, hp)%>%
```

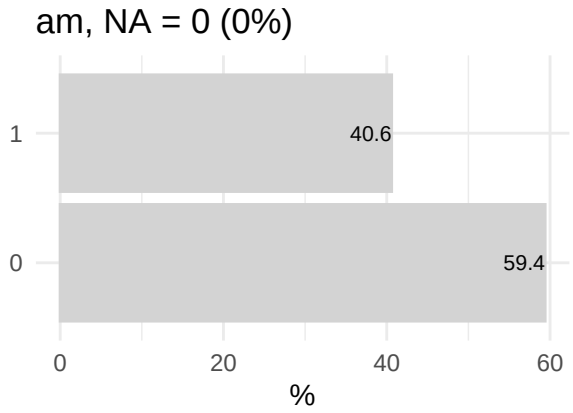
```
  explore_all()
```

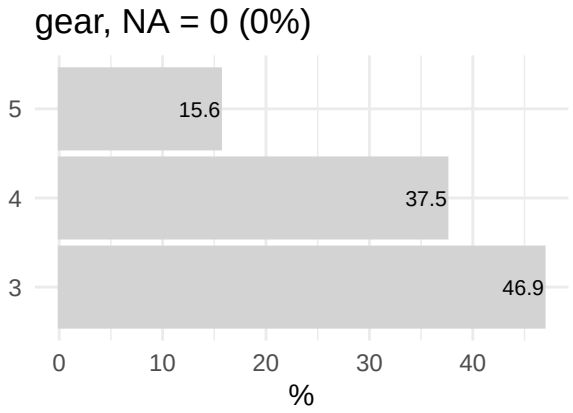


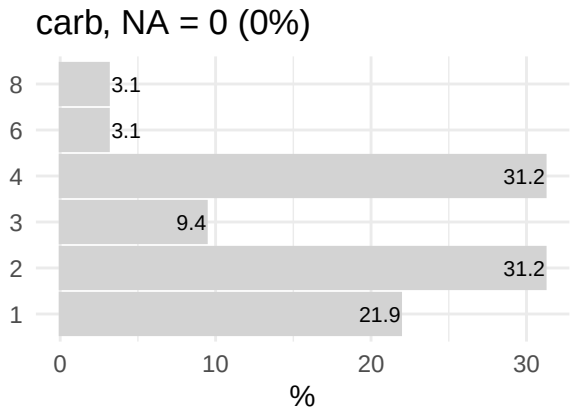

```
mtcars%>%
  select(drat, wt, qsec,vs)%>%
  explore_all()
```



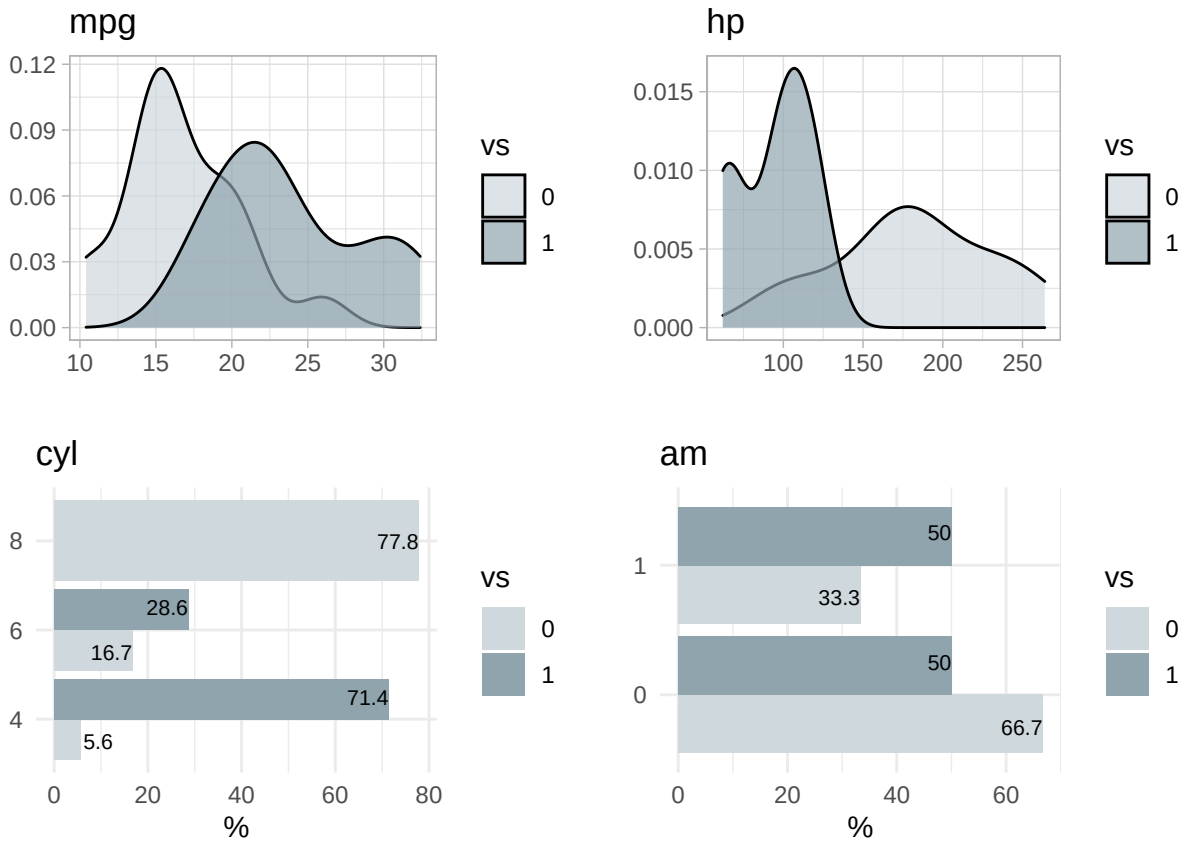

```
mtcars%>%
  select(am, gear, carb)%>%
  explore_all()
```



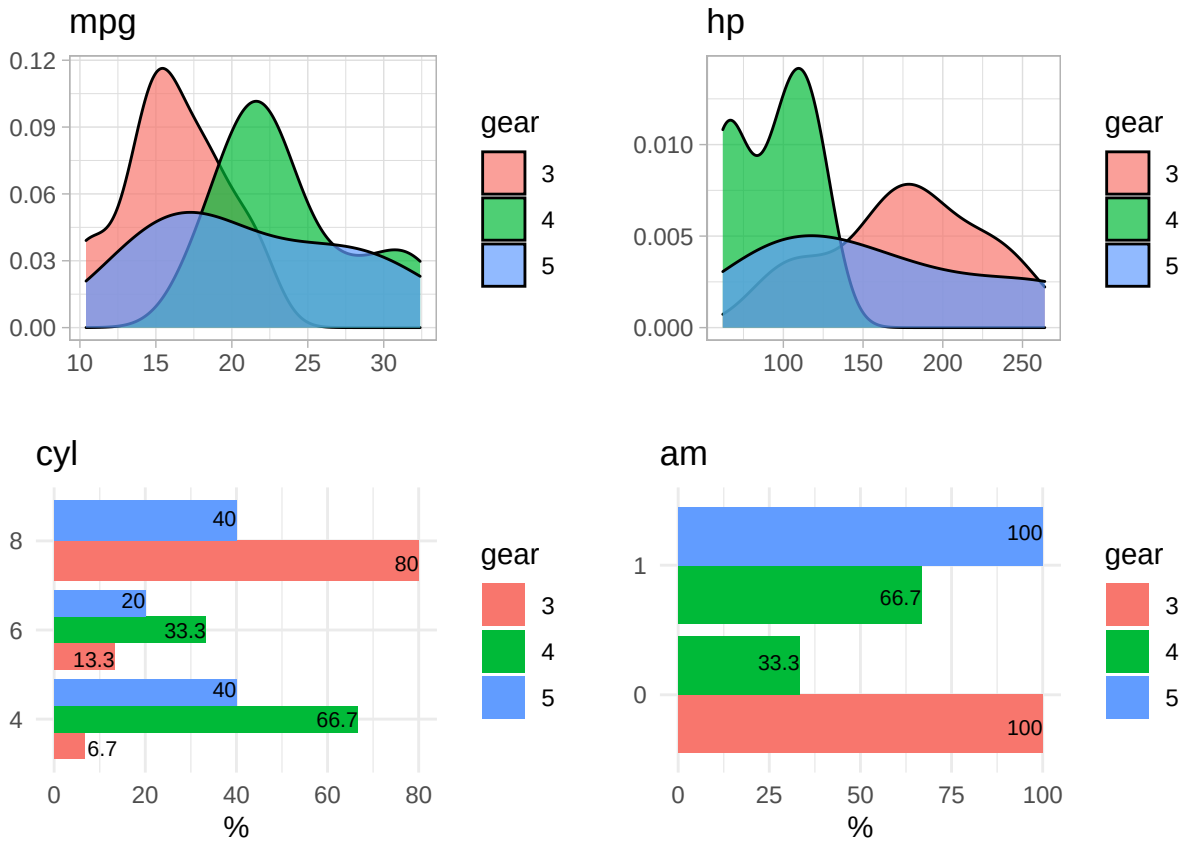




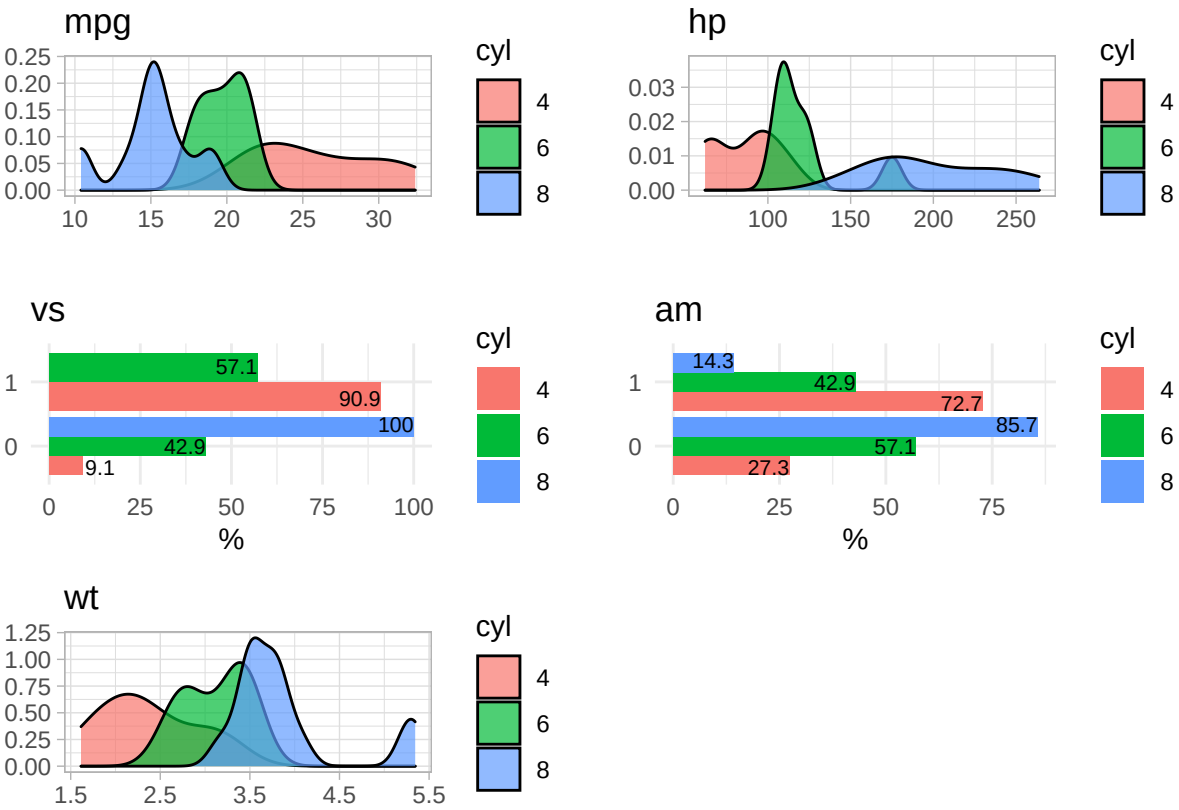

```
mtcars %>%
  select(vs, mpg, hp, cyl, am) %>%
  explore_all(target = vs)
```




```
mtcars %>%
  select(gear, mpg, hp, cyl, am) %>%
  explore_all(target = gear)
```




```
mtcars %>%
  select(cyl, mpg, hp, vs, am, wt) %>%
  explore_all(target = cyl)
```



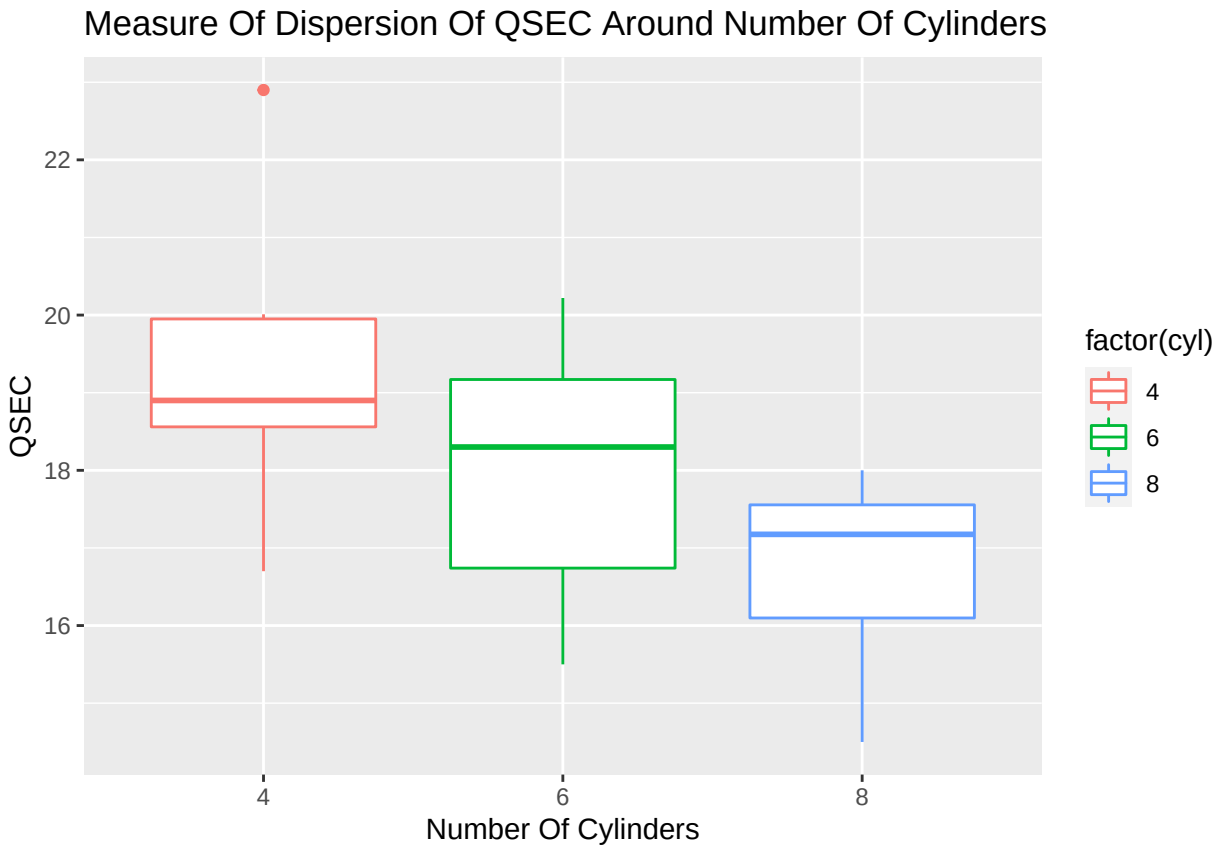
Comment: 1. Around 43.8% of the cars have 8 cylinders. Of which 85.7% are automatic vehicles. Futhermore, 100% of the 8 cylinder cars have a v-shaped engine, where as none of the straight engine cars cars have 8 cylinders. However, 94% of 4 cylinder cars have a straight engine. Moving on, 4 cylinder cars have the highest average mpg, whereas 8 cylinders cars have the lowest average mpg, whilst 8 cylinder cars have a higher horse power on average than 4 and 6 cylinder vehicles.

2. Out of the 32 cars 52.6% of the cars have a v-shaped engine, 66.7% are automatic cars. Whereas, exactly 50% of straight engine cars are automatic. 77.8% of the cars that have a v-shaped engine also have 8 cylinders. Straight engines on average have a higher mpg than v-shaped cars, which corresponds to v-shaped engine cars having higher cylinder count on average. V-shaped engines also on average have a higher horsepower.

3.)Out of the 32 cars 46.9% of the cars have three gears followed by 4 gears at 37.5%. Out of the three gear cars, 100% of then are automatic, where as a 100% of the 5 gear cars are manual. Furthermore, 80% of the 3 gear cars have 8 cylinder engines, whereas 66.7 of 4 gear cars have 4 cylinders. Out of the 5 gear cars, 80% of the cars have 8 and 6 cylinders, and the 20% remaining percent belong to 6 cylinder engines. Three gear cars which are mostly automatic and high-perforance vehicles have the lowest mpg and the highest hp on average.

Figure 2 :

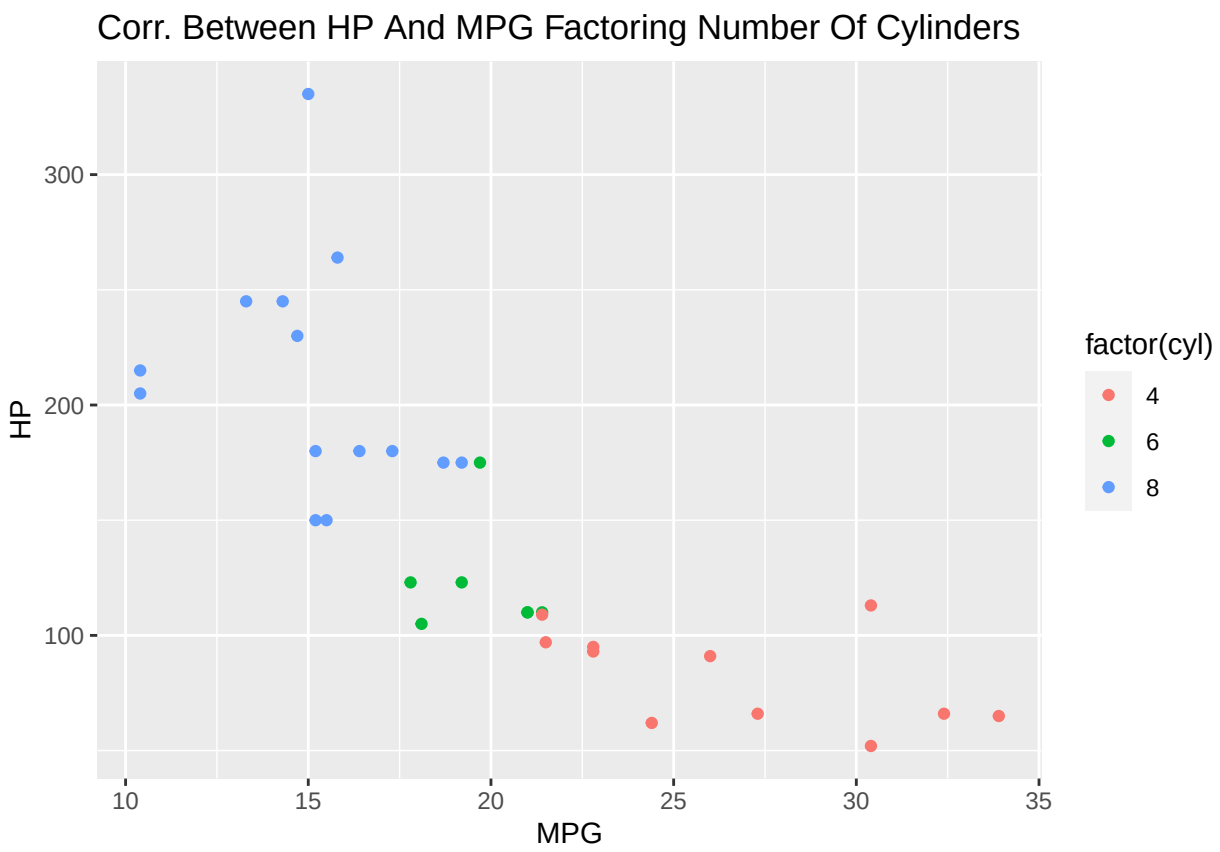
```
library(ggplot2)
ggplot(mtcars, aes(factor(cyl), qsec)) + geom_boxplot(aes(color = factor(cyl))) + ggtitle("Measure Of D
```



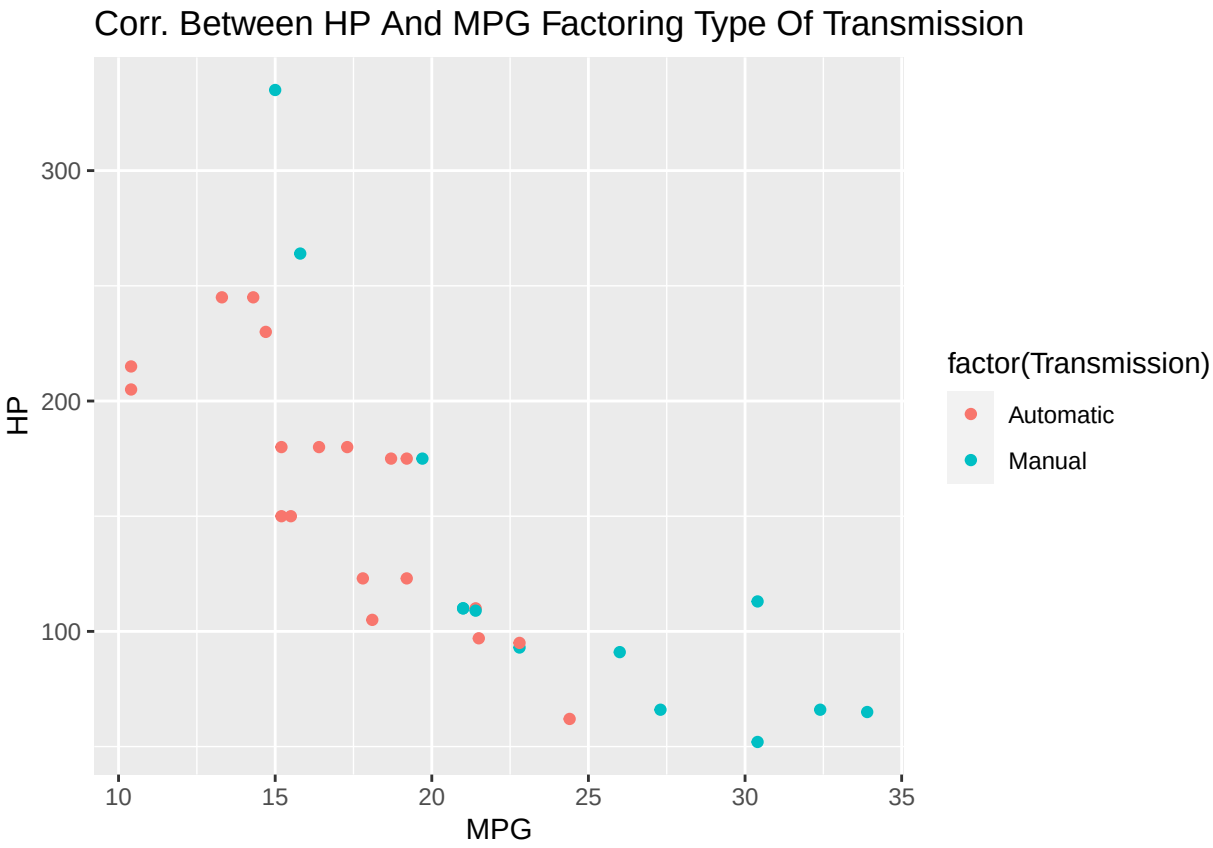
The data correspnds to 4 cylinder vehicles seem to be the quickest in 1/4th of a mile. With its central values being significantly higher than the two cylinder counts. The 8 cylinder high performance and mostly automatic cars seem to be the slowest. This may be attributed the these cars having the highest weight on average of 4000 lbs *(this can also be visualised in the plots in Figure 1)*.

Figure 3 :

```
ggplot(mtcars, aes(mpg, hp)) + geom_point(aes(colour = factor(cyl)))+ ggtitle("Corr. Between HP And MPG
```



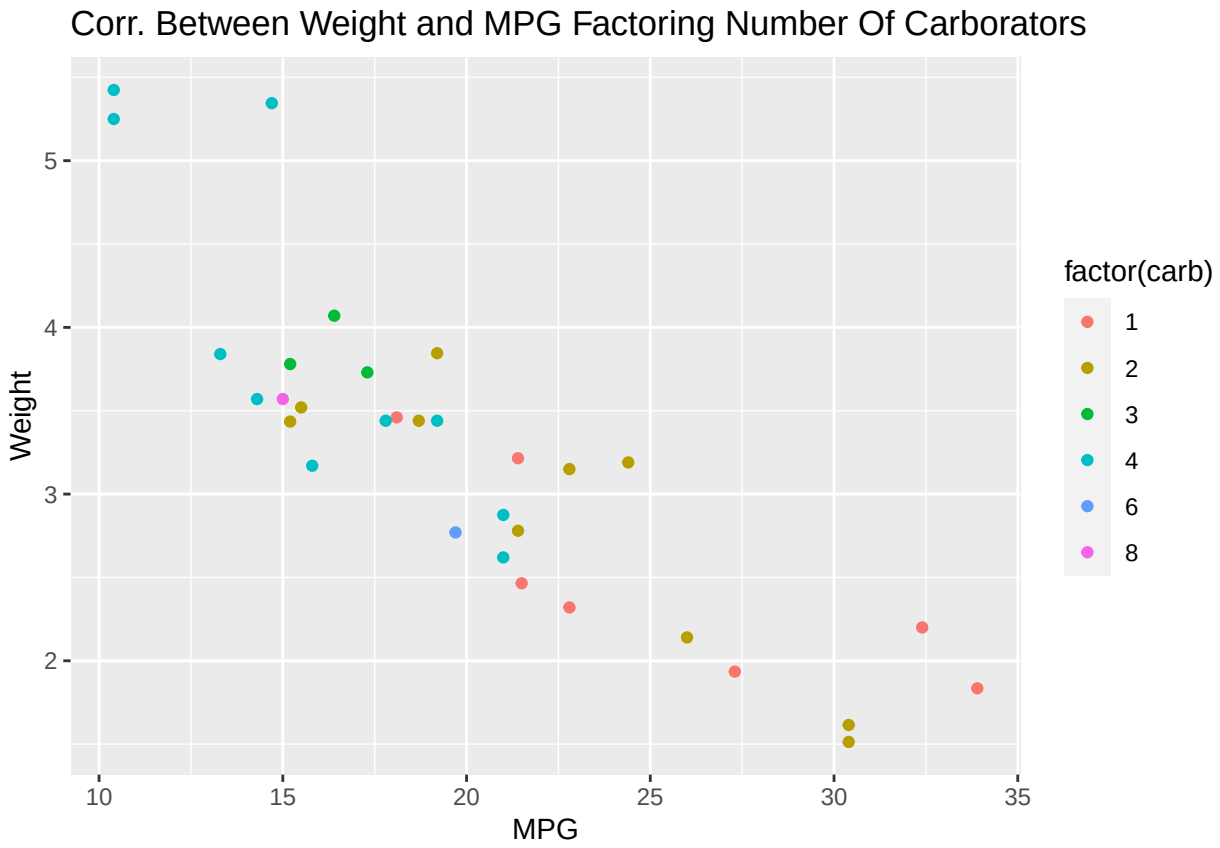

```
ggplot(mtcars1, aes(mpg, hp)) + geom_point(aes(colour = factor(Transmission)))+ ggtitle("Corr. Between
```



Comment: Horsepower and MPG appear to have a negatice correlationship with cars that have higher counts of cylinders (8) being towards to the lower side of MPG but higher up on the horsepower axis. The exact opposite for cars with lower count of cylinders (4) having healthy MPG values and lower HP values. The 7 cars with the 6 cylinders seem to be positioned between the two extremes in the range of 17.5 to 22.5 mpg. As expected automatic cars apear to the least efficient, whereas low HP manual cars have higher MPGs. The outlier with regards to transmission type is Masserati Bora and Ford Pantera which are low MPG high HP manual cars.

Figure 4 :

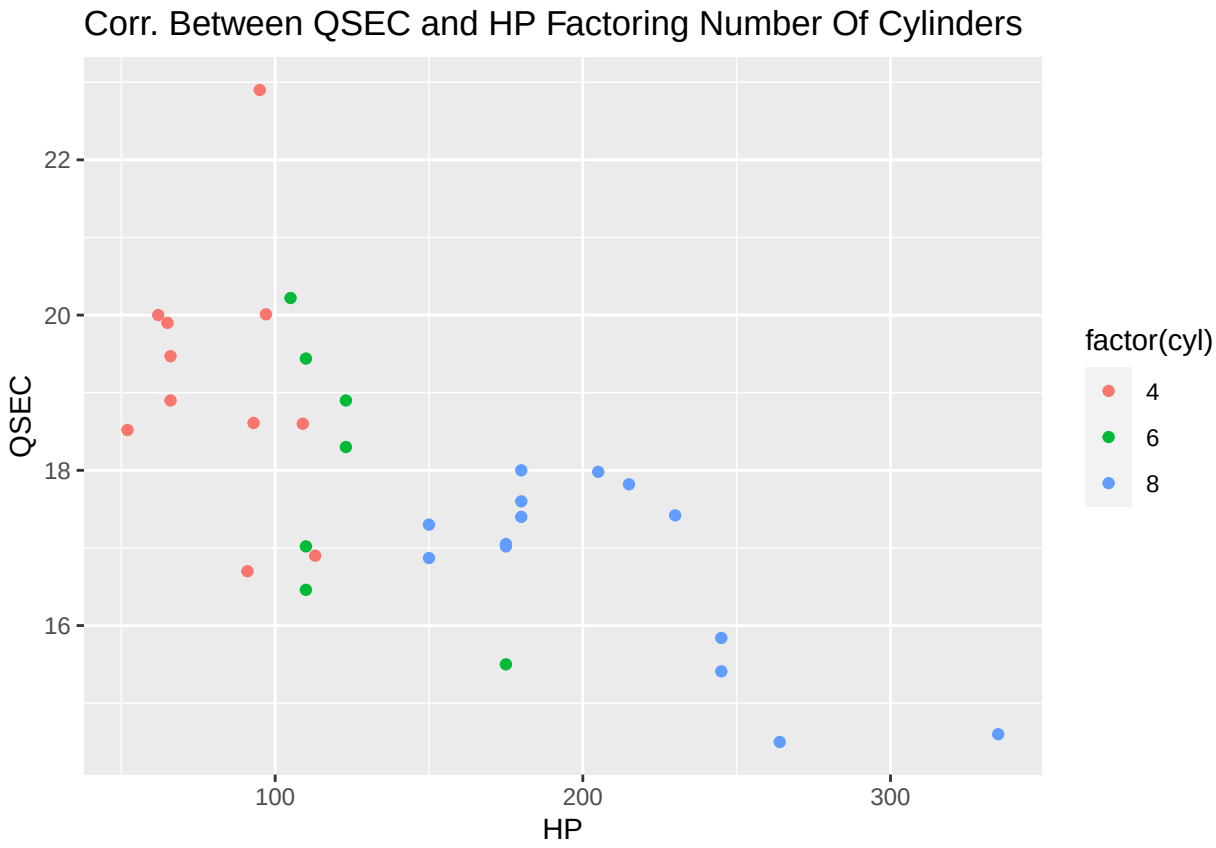
```
ggplot(mtcars, aes(mpg, wt)) + geom_point(aes(colour = factor(carb)))+ ggtitle("Corr. Between Weight an
```



Comment: Weight and MPG appear to have a negative correlation. It is practical to assume that cars that weigh a lot have are fuel inefficient. The plot is factored by the number of carborators in the car. Cars with 1 to 2 carborators appear to have the highest MPG and the lowest weight. Whereas the range of 15-20 MPG is dominated by mix of 2,3,4, and 6 caroborators. The heaviest cars Lincoln Continental, Chrysler Imperial and Cadillac Fleetwood each have 4 corboratos. The only car with 8 carborators is Maserati Bora and the only car with 6 carborators is Ferrari Dino.

Figure 5 :

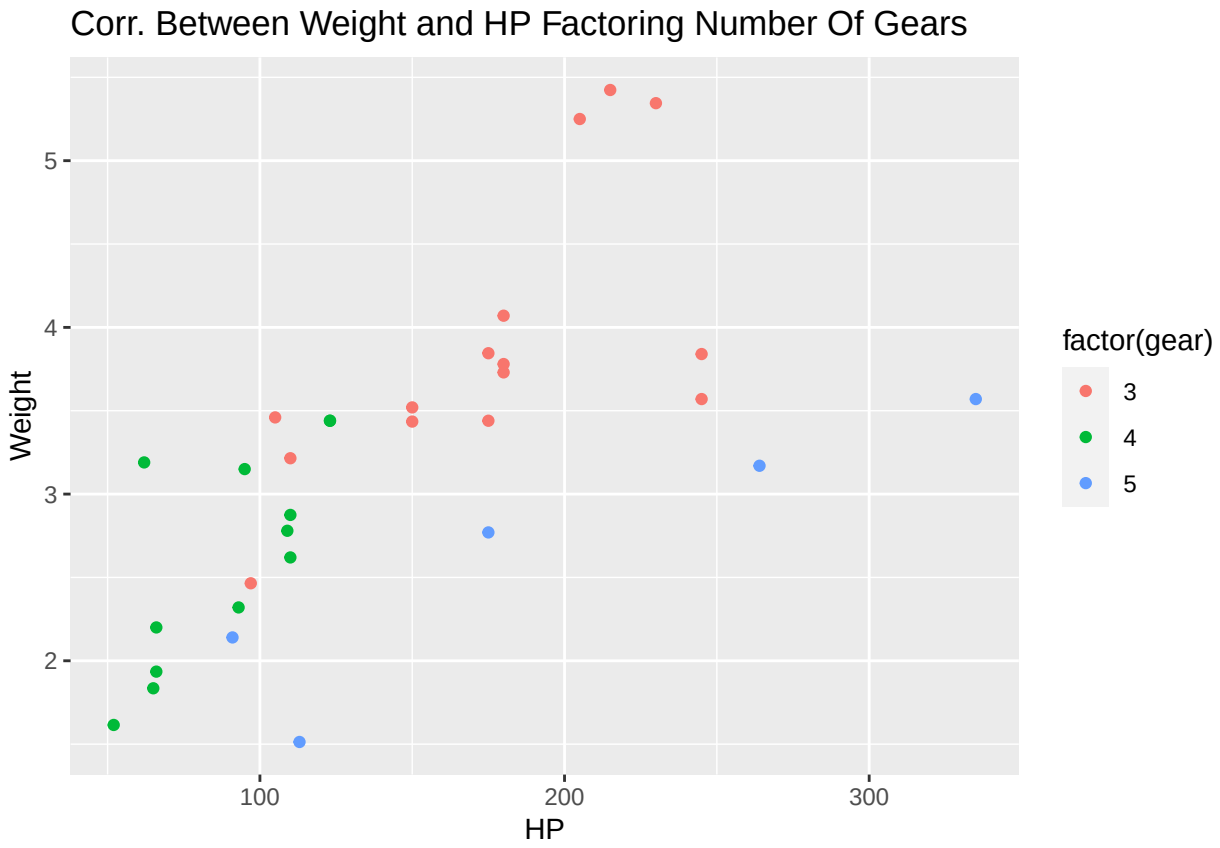
```
ggplot(mtcars, aes(hp, qsec)) + geom_point(aes(colour = factor(cyl)))+ ggtitle("Corr. Between QSEC and
```



Comment: There appears to be a negative correlation between HP and QSEC. Higher the HP lower the time in seconds in 1/4th of mile. High performace cars with higher number of cylinders have the lowest QSEC. The lowest QSEC is of Ford Pantera of 14.5 seconds where as the low performance 4 cylinder cars dominate the region of 0-100 HP and 18 to 23 QSEC. The slowest car is the Merc 230. The 6 cylinder cars are located between 100-150 HP and 16-21 QSEC. Apart from Ferrari Dino which is the fastest 6 cylinder car with a HP of just 175.

Figure 6 :

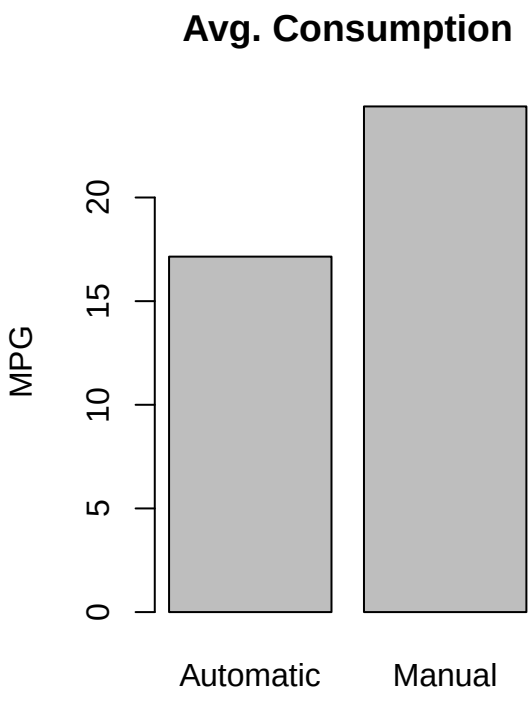
```
ggplot(mtcars, aes(hp, wt)) + geom_point(aes(colour = factor(gear)))+ ggtitle("Corr. Between Weight and
```

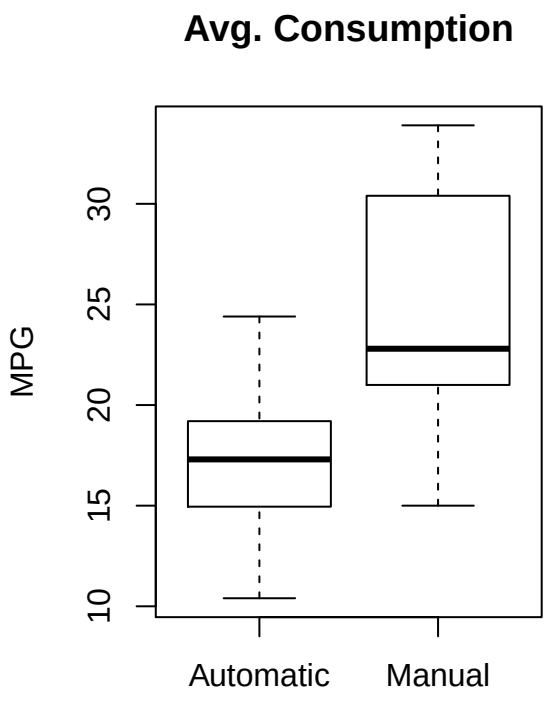


Comment: There appears to be a positive relationship between weight and HP. This can be assumed as heavy cars need stronger engines to mitigate for the weight. The three gear automatic vehiles are amongst the heaviest cars. These include the three heaviest Lincoln Continental, Chrysler Imperial and Cadillac Fleetwood which weigh 5420lbs, 5320lbs, 5250lbs respectively, along with a HP of 215,230 and 205. The car with the highest with HP is the Masserati Bora and the car with lowest weight and HP is Honda Civic.

Figure 7 :

```
par(mfrow = c(1,2))

mpg.geartype = tapply(mtcars$mpg, mtcars$am, mean)
barplot(mpg.geartype, names.arg = c("Automatic", "Manual"), main = "Avg. Consumption", xlab = "Type Of

boxplot(mtcars1$mpg ~ mtcars1$Transmission, data = mtcars1, xlab = "Type Of Transmission", ylab = "MPG"
```



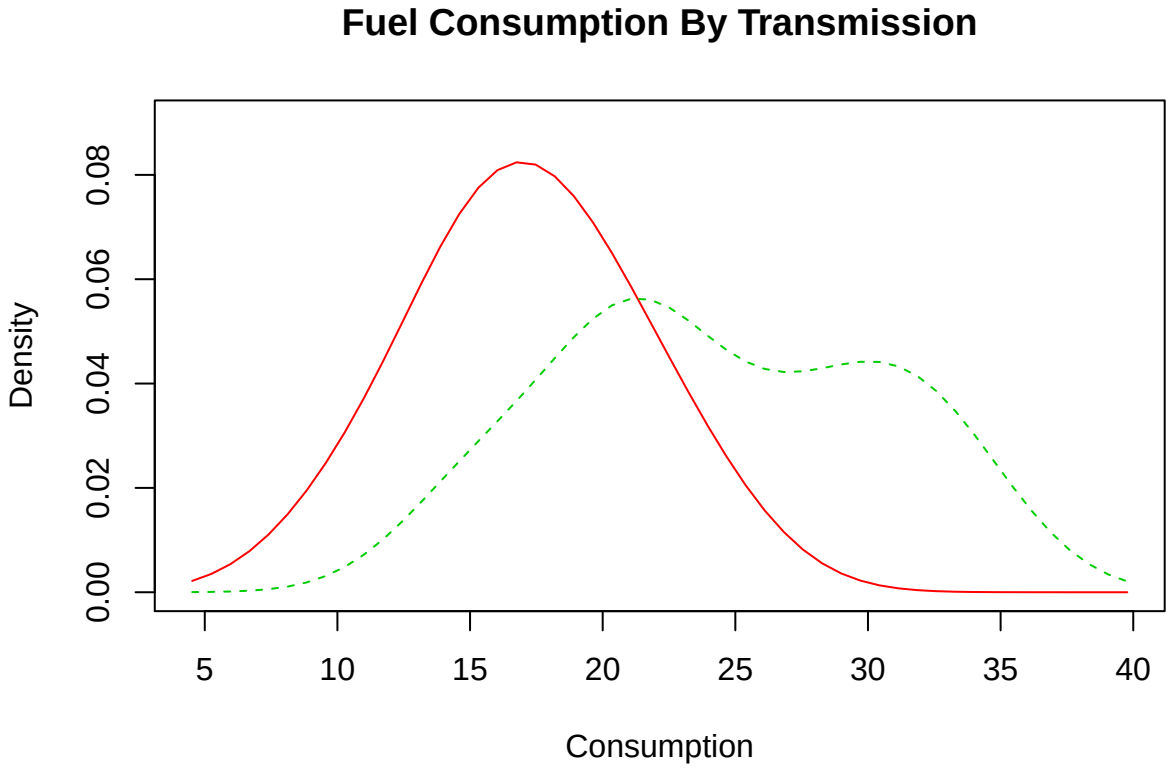



```
par(mfrow = c(1,1))

library(sm)
```

```
## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE):
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/
## Resources/library/tcltk/libs//tcltk.so'' had status 69
```

```
## Package 'sm', version 2.2-5.6: type help(sm) for summary information
```

```
sm.density.compare(mtcars1$mpg, mtcars1$am, xlab = "Consumption")
title(main = "Fuel Consumption By Transmission")
```



Comment: 1. The density plot aids is to recognise that the distribution of mpg is normally distrubuted. 2. The outliers is not skewing the plot.

Conducting Hypothesis Test:

H0: Type of tranmission do not have any effect on average fuel consumption. **H1:** Type of tranmission does have an effect on average fuel consumption. The average mpg of automatic cars is lower than manual cars.

```
auto.rows = mtcars1[mtcars1$Transmission == "Automatic",]
manual.rows = mtcars1[mtcars1$Transmission == "Manual",]
t.test(manual.rows$mpg, auto.rows$mpg)
```

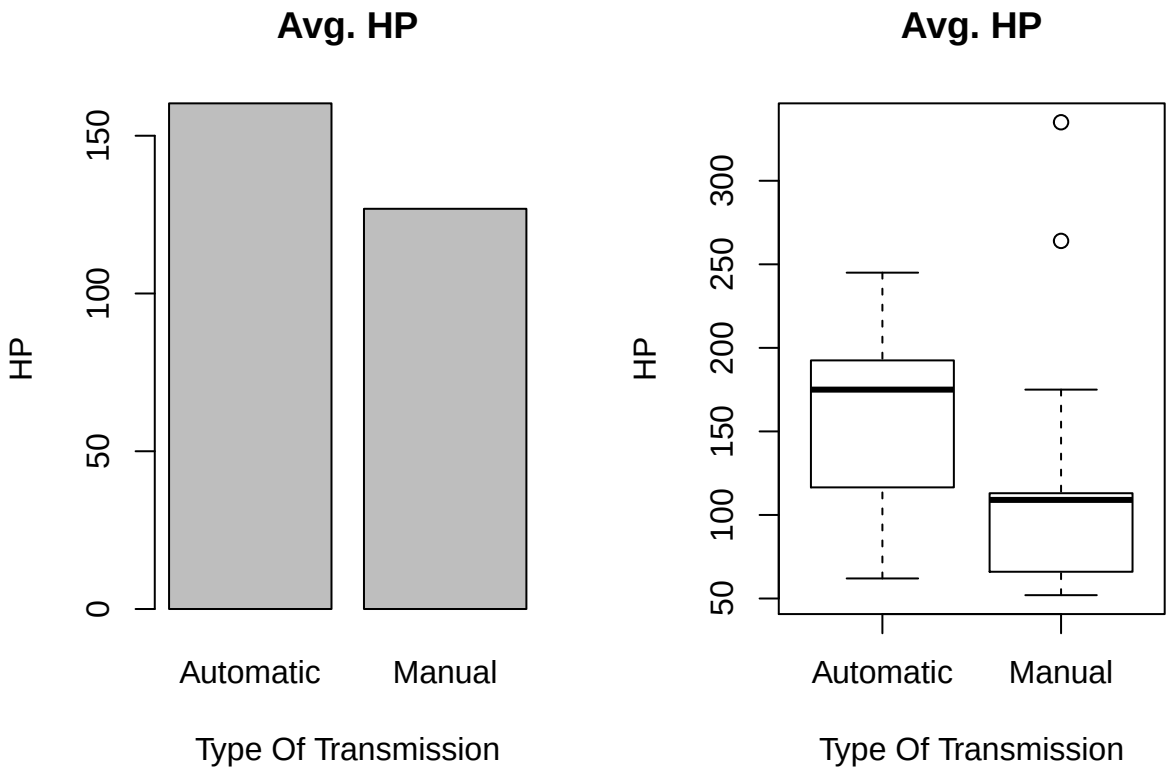
```
##
##  Welch Two Sample t-test
##
## data:  manual.rows$mpg and auto.rows$mpg
## t = 3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

Conclusion:

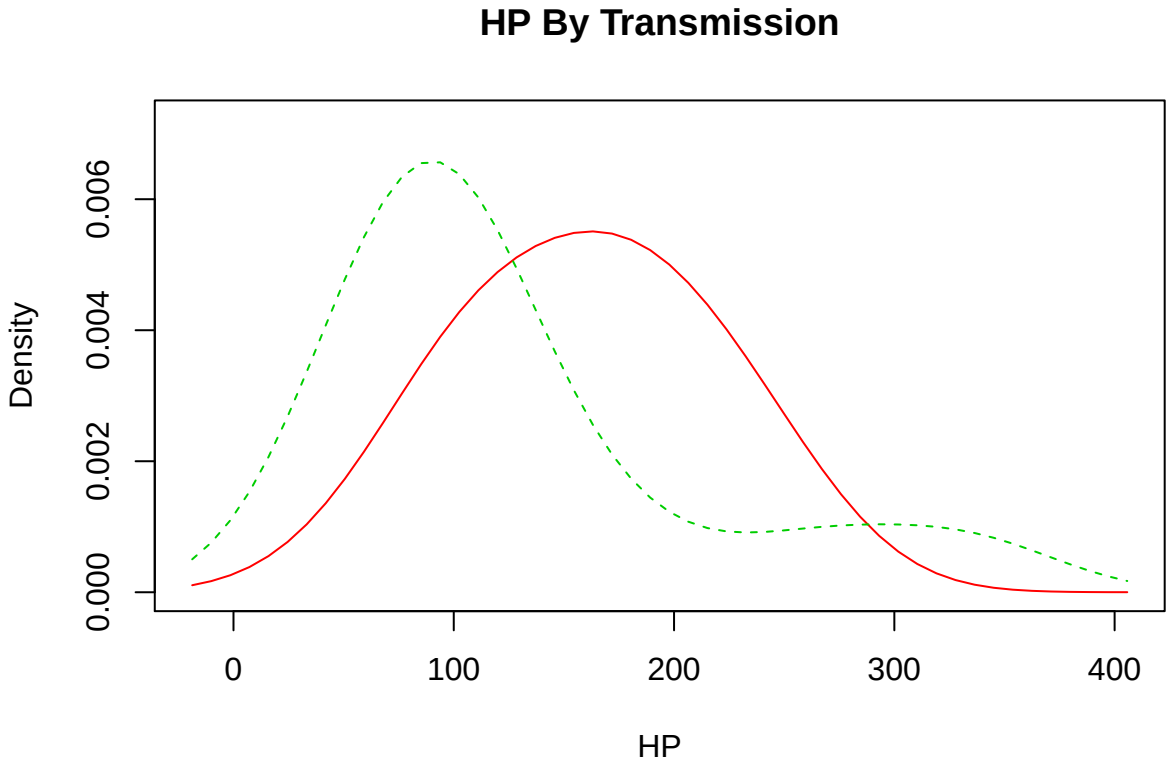
The pvalue is 0.0014, which is less than the default l.o.s of 0.015. Given the the test we reject the null hypothesis and accept the alternative hypothesis that automatic cars on an average have lower mpg than manual cars. This supports the graphical evidence of the barplots and boxplots above.

Figure 8 :

```r
par(mfrow = c(1,2))

hp.geartype = tapply(mtcars$hp, mtcars$am, mean)
barplot(hp.geartype, names.arg = c("Automatic", "Manual"), main = "Avg. HP", xlab = "Type Of Transmissi

boxplot(mtcars1$hp ~ mtcars1$Transmission, data = mtcars1, xlab = "Type Of Transmission", ylab = "HP", m
```




```r
par(mfrow = c(1,1))

library(sm)
sm.density.compare(mtcars1$hp, mtcars1$am, xlab = "HP")
title(main = "HP By Transmission")
```



Conducting Hypothesis Test:

H0: Type of tranmission do not have a significant effect on average HP. **H1:** Type of tranmission does have a significant effect on average QSEC. The average HP of automatic cars is higher than manual cars.

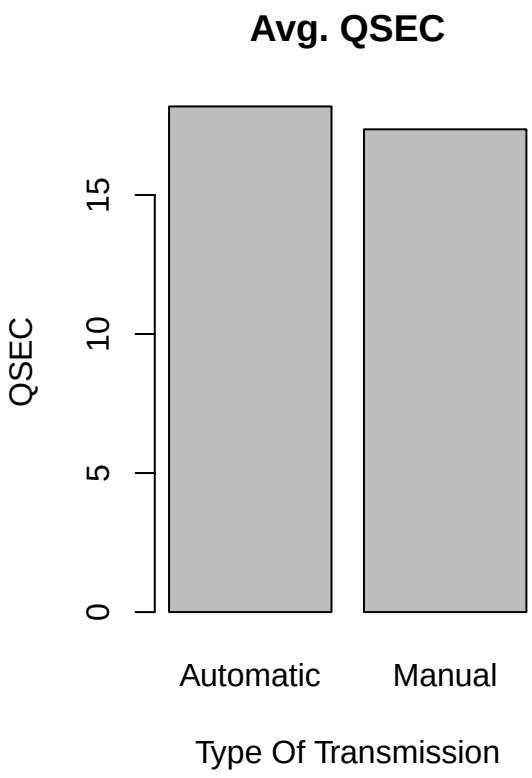
```
t.test(manual.rows$hp, auto.rows$hp)
```

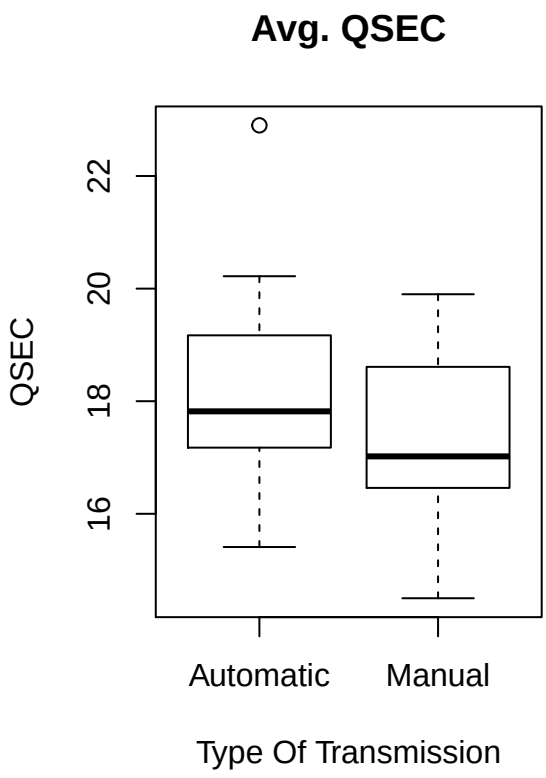
```
##
##  Welch Two Sample t-test
##
## data:  manual.rows$hp and auto.rows$hp
## t = -1.2662, df = 18.715, p-value = 0.221
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -88.71259  21.87858
## sample estimates:
## mean of x mean of y
##  126.8462  160.2632
```

Conclusion: The p-vlaue of the two sample t-test is 0.221 which is higher than the l.o.s of 0.05. With this we conlude that we fail to reject the null hypotheses which states that type of tranmission does not have a significant effect on HP.

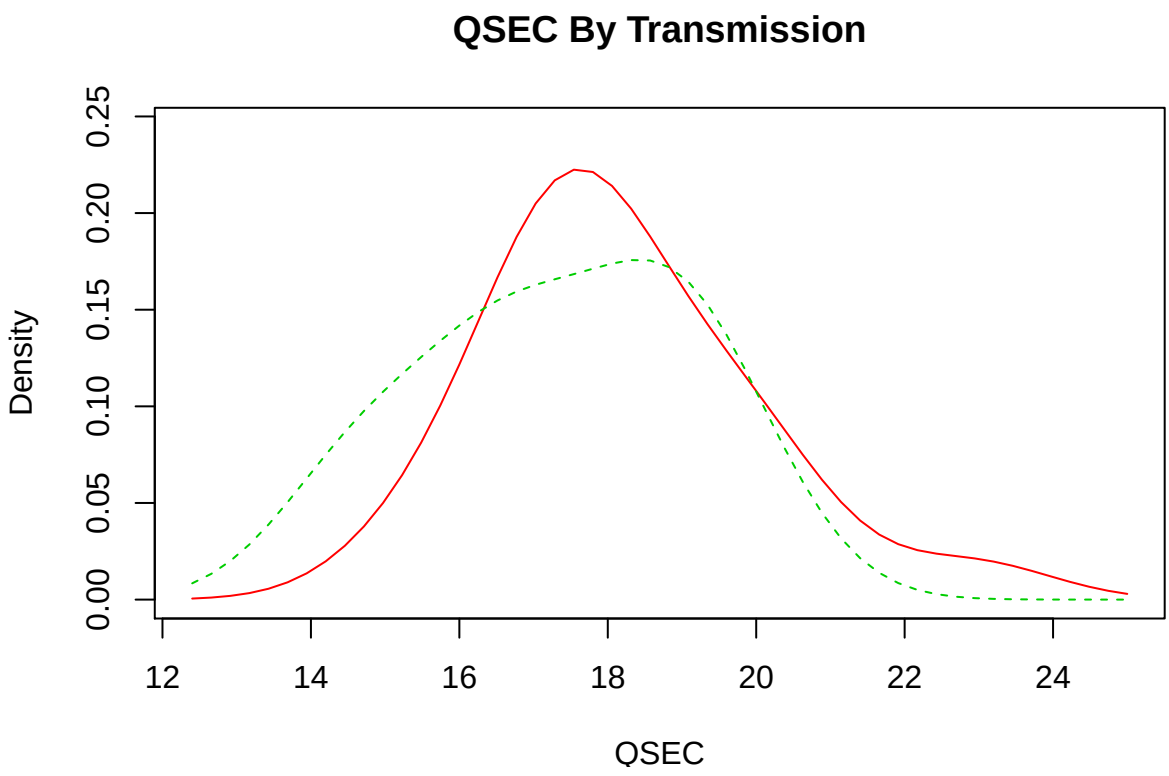
Figure 9 :

```
par(mfrow = c(1,2))

qsec.geartype = tapply(mtcars$qsec, mtcars$am, mean)
barplot(qsec.geartype, names.arg = c("Automatic", "Manual"), main = "Avg. QSEC", xlab = "Type Of Transm

boxplot(mtcars1$qsec ~ mtcars1$Transmission, data = mtcars1, xlab = "Type Of Transmission", ylab = "QSE
```






```r
par(mfrow = c(1,1))

library(sm)
sm.density.compare(mtcars1$qsec, mtcars1$am, xlab = "QSEC")
title(main = "QSEC By Transmission")
```



Conducting Hypothesis Test:

H0: Type of tranmission do not have a significant effect on average QSEC. **H1:** Type of tranmission does have a significant effect on average QSEC. The average QSEC of automatic cars is higher than manual cars.

```
t.test(manual.rows$qsec, auto.rows$qsec)
```

```
##
##  Welch Two Sample t-test
##
## data:  manual.rows$qsec and auto.rows$qsec
## t = -1.2878, df = 25.534, p-value = 0.2093
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.1381679  0.4918522
## sample estimates:
## mean of x mean of y
##  17.36000  18.18316
```

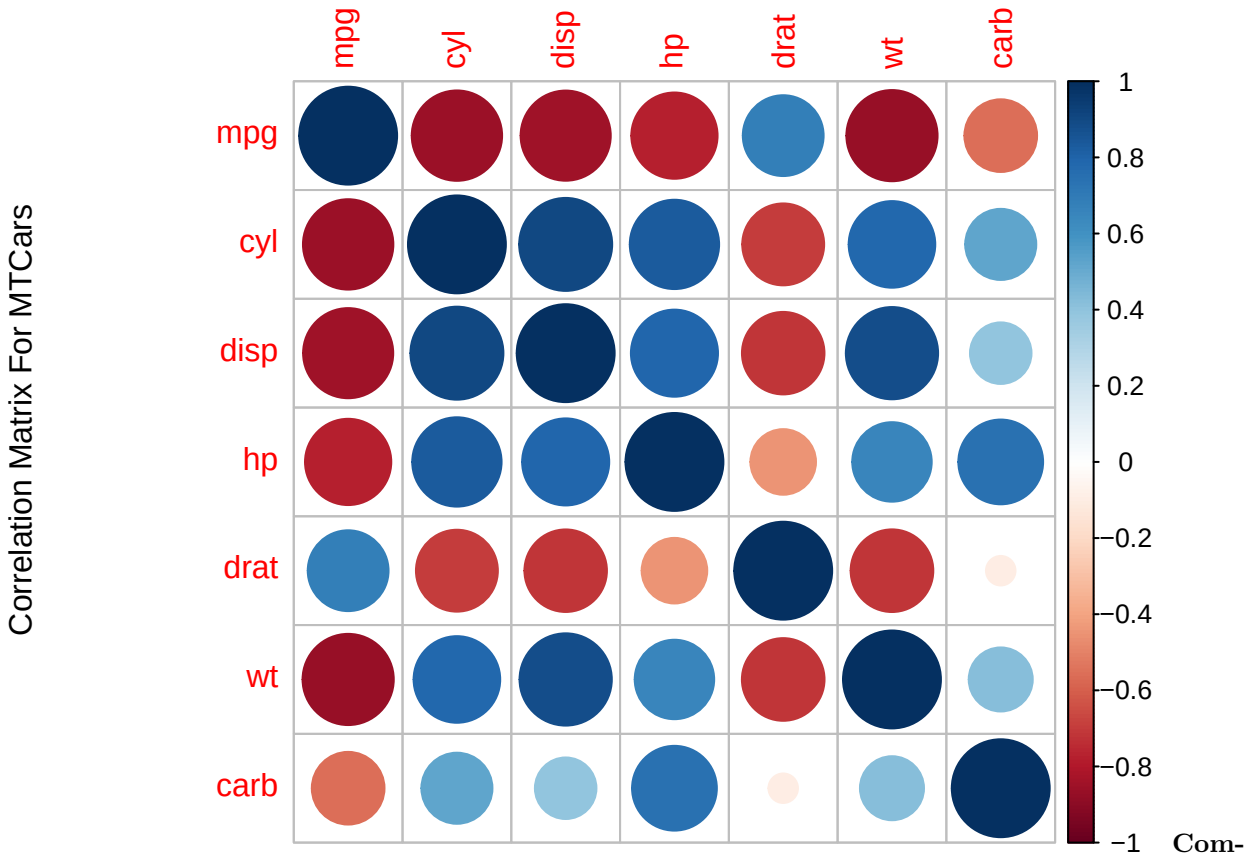
Conclusion: The p-vlaue of the two sample t-test is 0.2093 which is higher than the l.o.s of 0.05. With this we conlude that we fail to reject the null hypotheses which states that type of tranmission does not have a significant effect on the average QSEC.

Figure 10 :

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
correlations = cor(mtcars[,c(1,2,3,4,5,6,11)])
corrplot(correlations, method = "circle")
title(ylab = "Correlation Matrix For MTCars")
```



ment:

The correlation plot above aids in understanding the nature of correlations between the variables in the mtcars dataset. Intersection of the variables in the matrix above and the size and color of the corresponding circles determine the correlation of the two variables. Dark blue corresponds to perfectly positively correlated and dard red corresponds to perfectly negatively correlated, as shown by the color scale on the right of the diagram. We can spot a strong negative relationship between mpg and number of cylinders *(also in Figure 3)*, degree of displacement, horsepower, and weight. While a relatively mild negative relationship between mpg and carborators. This reault could be anticipated prior to plotting this diagram as high-performance cars that have a higher counts of cyl,disp, and hp would have a lower mpg. Variables cylinder, dysp, and hp are positively correalted with one another. The variable weight is also positively correlated with cyl, disp and hp, as cars that weigh more would require a high performance engine with higher counts of cylinders, and horsepower.