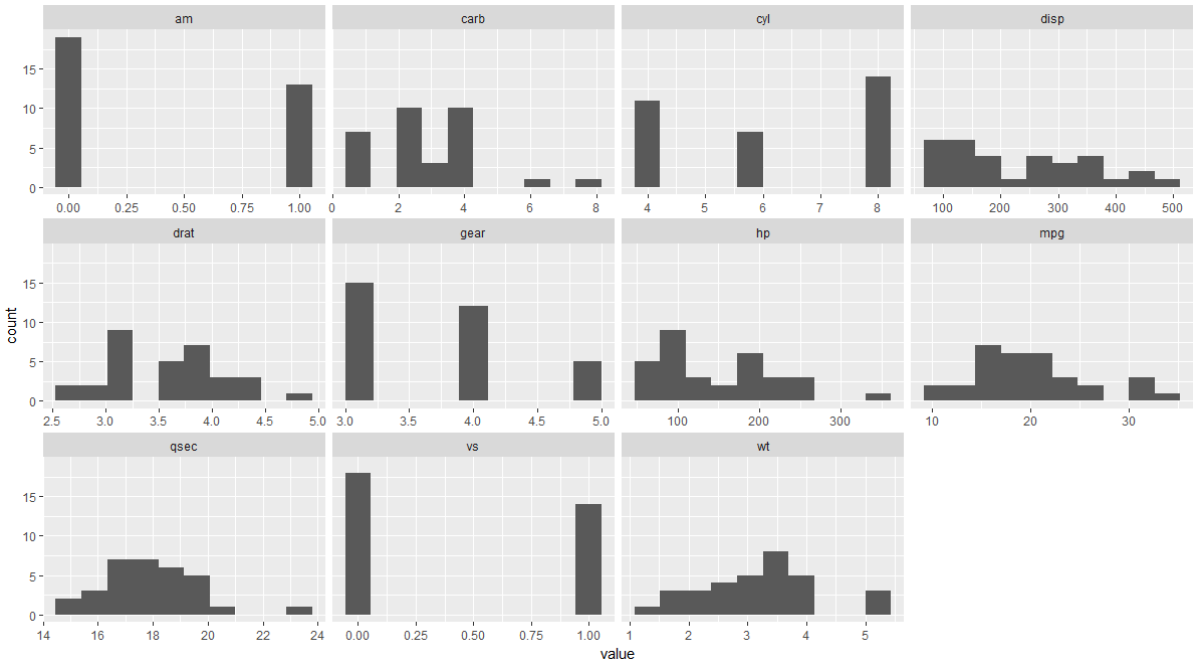
Assignment - 1

Session 7 – Basic Statistics

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

```
Ans: > library(tidyr)
> library(ggplot2)
> mtcars %>% gather() %>% head()
  key value
1 mpg  21.0
2 mpg  21.0
3 mpg  22.8
4 mpg  21.4
5 mpg  18.7
6 mpg  18.1
> ggplot(gather(mtcars), aes(value)) +
+     geom_histogram(bins = 10) +
+     facet_wrap(~key, scales = 'free_x')
```



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

Ans:

```
> cars_auto = subset(mtcars, am == 0)
> cars_gear = subset(mtcars, am == 1)
> dim(mtcars)
[1] 32 11
> > dim(cars_auto) ; dim(cars_gear)
Error: unexpected '>' in ">"
> dim(cars_auto) ; dim(cars_gear)
```

```
[1] 19 11
[1] 13 11
> mean(cars_auto$mpg); mean(cars_gear$mpg)
[1] 17.14737
[1] 24.39231
> sd(cars_auto$mpg) ; sd(cars_gear$mpg)
[1] 3.833966
[1] 6.166504
> t.test(cars_gear$mpg, cars_auto$mpg, paired = F, var.equal = F)

        Welch Two Sample t-test

data:  cars_gear$mpg and cars_auto$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
```

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

Ans: