# Constructing appropriate graphs using GGPLOT2

## A Project submitted by Saba Akram

### To Moonis Shakil Sir.

# Defination of GGPLOT

“ggplot2 is a plotting system for R, based on the grammar of graphics, which tries to take the good parts of base and lattice graphics and none of the bad parts. It takes care of many of the fiddly details that make plotting a hassle (like drawing legends) as well as providing a powerful model of graphics that makes it easy to produce complex multi-layered graphics.” - [<http://ggplot2.org>)

* Adding it to R:
* install.packages(“ggplot2”)
* Or use the menu option
* We’ll also use dplyr (a package to help manipulate data)
* install.packages(dplyr)

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.3

##   
## 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(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.3

What is the Need for the ggplot2?

There are two major issues with the base graphics.

* Base graphics are not of print quality, if one wants to publish in an international journal or presentation, it will create abad image.
* These are just images and the layers cannot be added.It is not possible with the base graphics. Hence ggplot2 package required.

gg in gglot stands for grammar of graphics.

# Illustrating GGPLOT2

Lets start the work Using mtcars dataset.

data(mtcars)  
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 ...

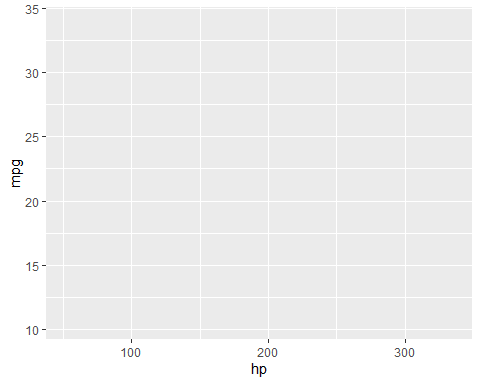
summary(mtcars)

## 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

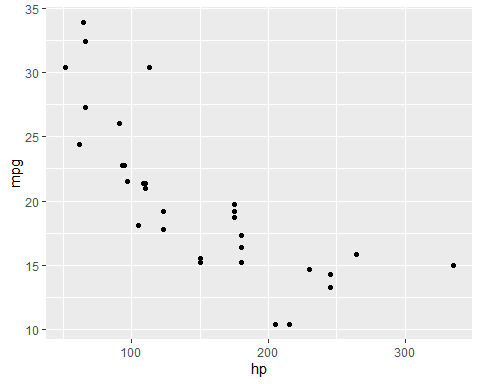
# SCATTER PLOT

Main code used in scattering plot diagram is geom\_pont() and later various codes are added which are mentioned in the examples below.

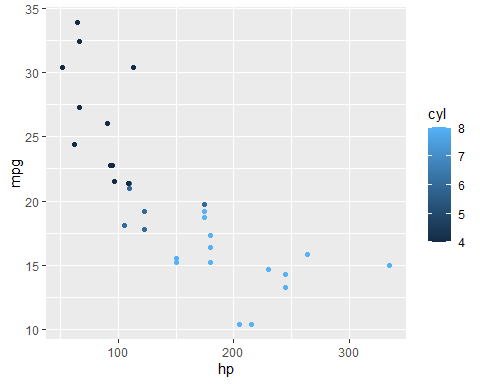
ggplot(data = mtcars,aes(y=mpg,x=hp))



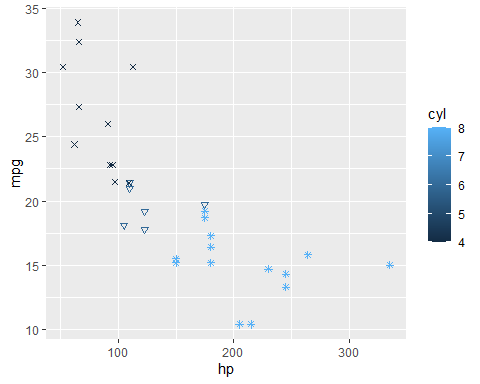
#Introducing grom\_point  
ggplot(data = mtcars,aes(y=mpg,x=hp))+geom\_point()



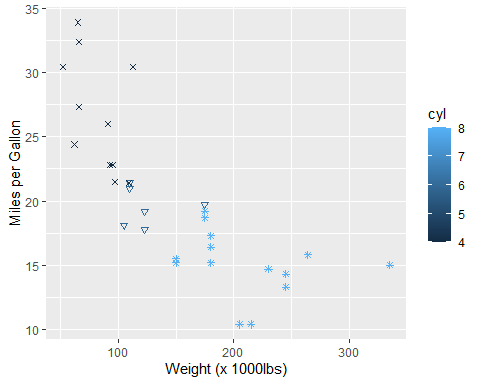
#Adding colour  
ggplot(data = mtcars,aes(y= mpg,x=hp,col=cyl))+geom\_point()



#Adding Shape  
ggplot(data = mtcars,aes(y=mpg,x=hp,col=cyl,shape=cyl))+geom\_point()+scale\_shape\_identity()

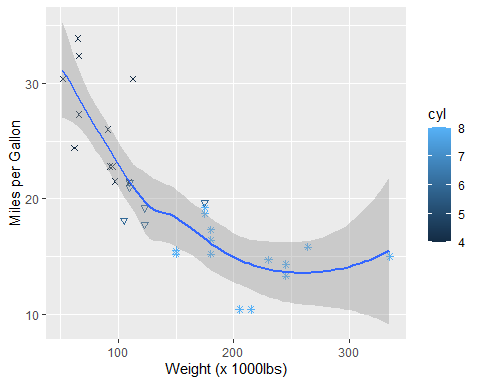


#labelling both the axes  
ggplot(data = mtcars,aes(y=mpg,x=hp,col=cyl,shape=cyl))+geom\_point()+scale\_shape\_identity() + xlab('Weight (x 1000lbs)') + ylab('Miles per Gallon')

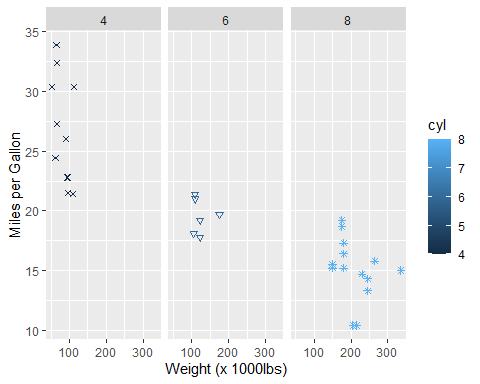


#Introducing geon\_smooth()  
ggplot(data = mtcars,aes(y=mpg,x=hp,col=cyl,shape=cyl))+geom\_point()+scale\_shape\_identity() + xlab('Weight (x 1000lbs)') + ylab('Miles per Gallon') + geom\_smooth()

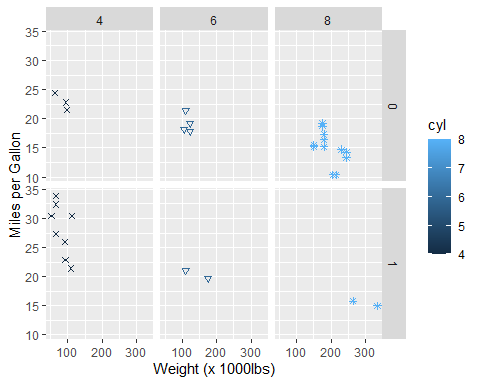
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



#Introducing facets(Small multiples)  
ggplot(data = mtcars,aes(y=mpg,x=hp,col=cyl,shape=cyl))+geom\_point()+scale\_shape\_identity() + xlab('Weight (x 1000lbs)') + ylab('Miles per Gallon') + facet\_grid(~cyl)



ggplot(data = mtcars,aes(y=mpg,x=hp,col=cyl,shape=cyl))+geom\_point()+scale\_shape\_identity() + xlab('Weight (x 1000lbs)') + ylab('Miles per Gallon') + facet\_grid(am~cyl)



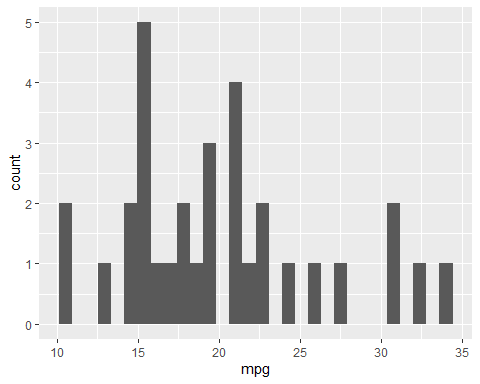
Basic reason to draw the scatter plot between mpg(miles per gallon) and hp(horse power) to show the relationshp between the two an also to learn the new concepts in the ggplot2 such as geom\_point(), geom\_smooth(), filing colours in them and also giving various shapes to them.

# HISTOGRAM

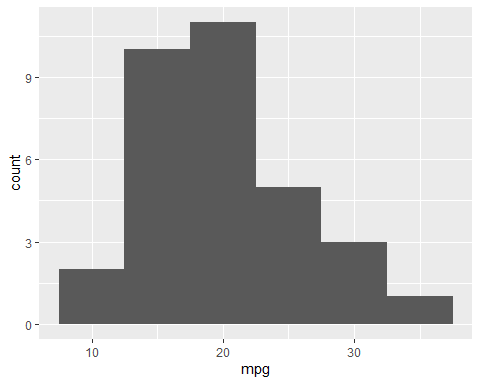
Main code used in the histogram is geom\_histogram and later various codes are added which are mentioned in the examples below.

ggplot(mtcars,aes(x=mpg)) + geom\_histogram()

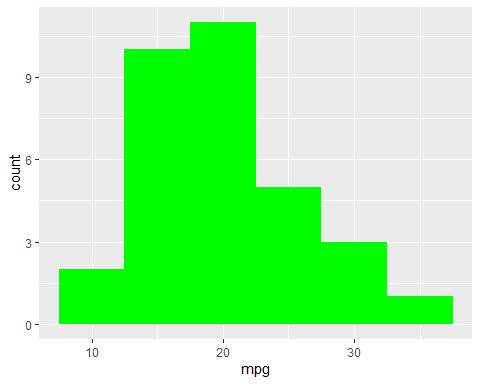
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



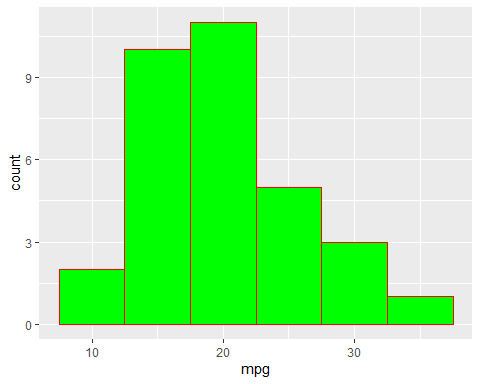
# Now changing the bandwidth  
ggplot(mtcars,aes(x=mpg)) + geom\_histogram(binwidth=5)



#filling colours  
ggplot(data = mtcars,aes(x=mpg))+geom\_histogram(binwidth =5,fill="green")

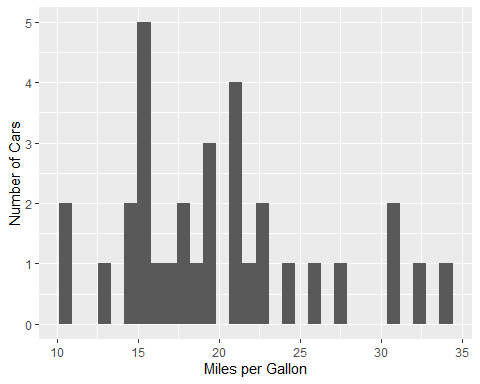


#Now colouring the borderline  
ggplot(data = mtcars,aes(x=mpg))+geom\_histogram(binwidth =5,fill="green",col="red")



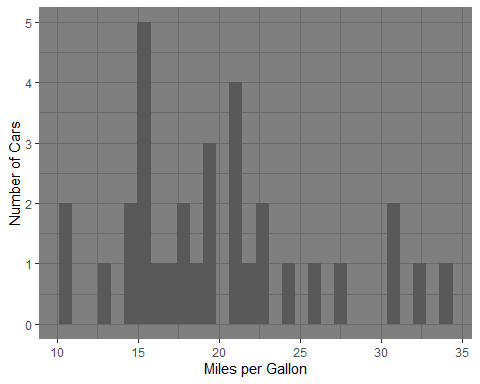
#Now mentioning names of each axis  
ggplot(mtcars,aes(x=mpg)) + geom\_histogram() + xlab("Miles per Gallon")+ylab("Number of Cars")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Now changing theme  
ggplot(mtcars,aes(x=mpg)) + geom\_histogram() + xlab("Miles per Gallon")+ylab("Number of Cars") + theme\_dark()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

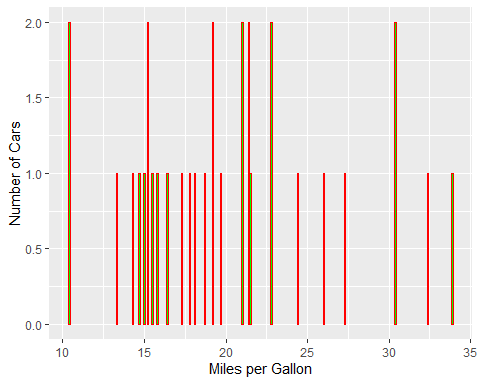


Basic reason to draw the histogram between mpg(miles per gallon) and number of cars to show the number of cars with particular mpg and also to learn the new concepts in the ggplot2 such as geom\_point()' geom\_smooth(), geom\_histogram(), filing colours in them and also giving various shapes to them.

# BARGRAPH

Main code used in scattering plot diagram is geom\_bar() and later various codes are added which are mentioned in the examples below. The basic difference between the histogram and the bargraph is that histogram is continuous whereas bargraph is not.

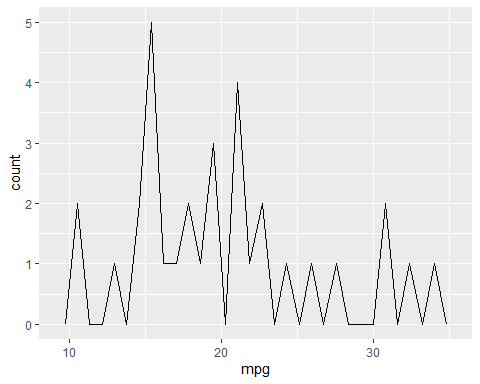
ggplot(data = mtcars,aes(x=mpg))+geom\_bar(fill="green",col="red") + xlab("Miles per Gallon")+ylab("Number of Cars")



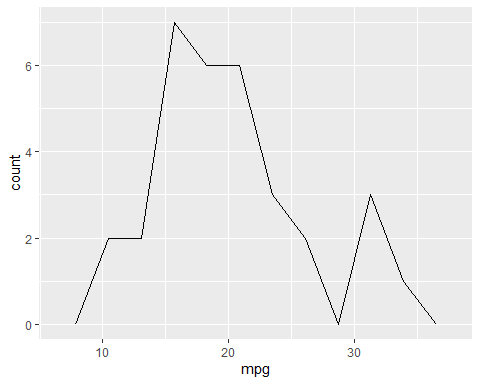
# FREQUENCY POLYGONS

ggplot(data = mtcars,aes(x=mpg))+geom\_freqpoly()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



#Changing bins  
ggplot(data = mtcars,aes(x=mpg))+geom\_freqpoly(bins=10)



# BOXPLOT

Main code used in boxplot diagram is geom\_boxplot() and later various codes are added which are mentioned in the examples below.A box plot or boxplot is a method for graphically depicting groups of numerical data through their quartiles.

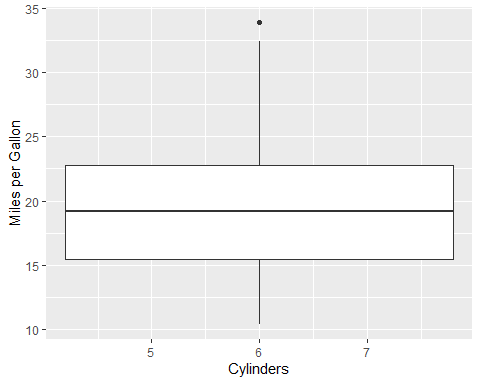
ggplot(mtcars,aes(x=cyl,y=mpg)) + geom\_boxplot()

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?



#labelling x axis and y axis  
ggplot(mtcars,aes(x=cyl,y=mpg)) + geom\_boxplot() + xlab('Cylinders') + ylab('Miles per Gallon')

## Warning: Continuous x aesthetic -- did you forget aes(group=...)?



#THE END