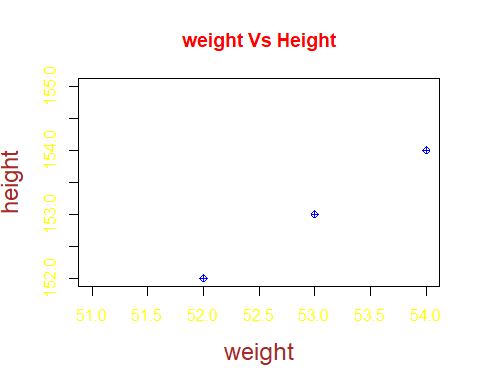
scatterplots.R

R Roseline Mary

2021-01-28

#Simple scatter plot  
x<-50:55  
y<-150:155  
plot(x,y,xlab="weight",ylab="height",xlim=c(51,54),ylim=c(152,155),main="weight Vs Height",col.main="red",cex.lab="1.5",col="blue",pch=10,col.axis="yellow",col.lab="brown")



?mtcars

## starting httpd help server ... done

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 ...

head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

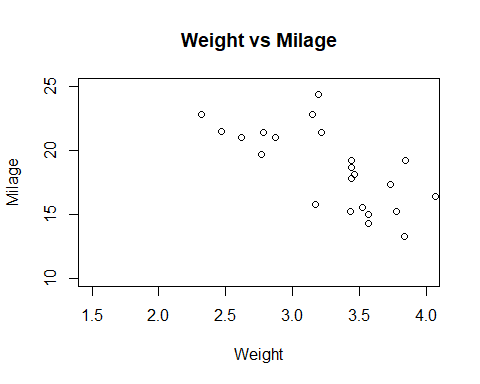
tail(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Porsche 914-2 26.0 4 120.3 91 4.43 2.140 16.7 0 1 5 2  
## Lotus Europa 30.4 4 95.1 113 3.77 1.513 16.9 1 1 5 2  
## Ford Pantera L 15.8 8 351.0 264 4.22 3.170 14.5 0 1 5 4  
## Ferrari Dino 19.7 6 145.0 175 3.62 2.770 15.5 0 1 5 6  
## Maserati Bora 15.0 8 301.0 335 3.54 3.570 14.6 0 1 5 8  
## Volvo 142E 21.4 4 121.0 109 4.11 2.780 18.6 1 1 4 2

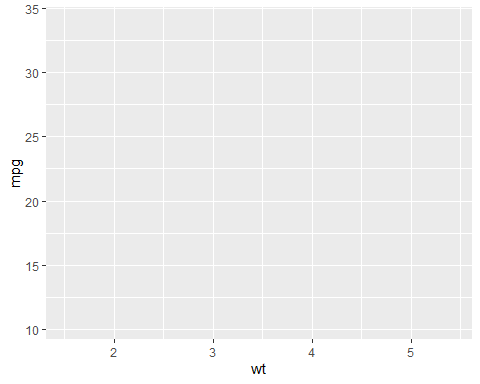
input<-mtcars[,c('wt','mpg')]  
plot(x = input$wt, y = input$mpg,   
 xlab = "Weight",   
 ylab = "Milage",   
 xlim = c(1.5, 4),   
 ylim = c(10, 25),   
 main = "Weight vs Milage"  
)   
  
  
head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

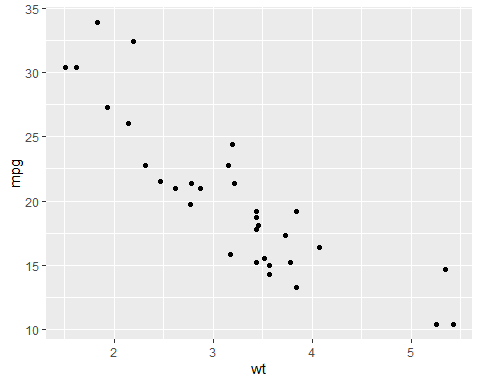
library(ggplot2)



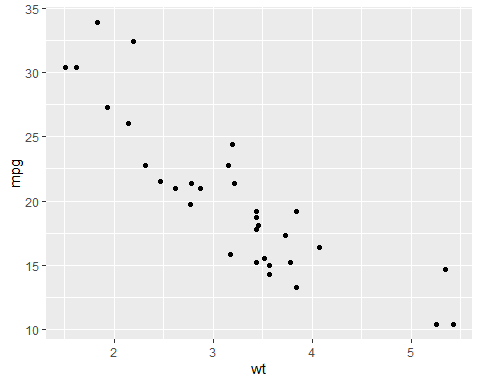
p<-ggplot(mtcars,aes(wt,y=mpg))  
p



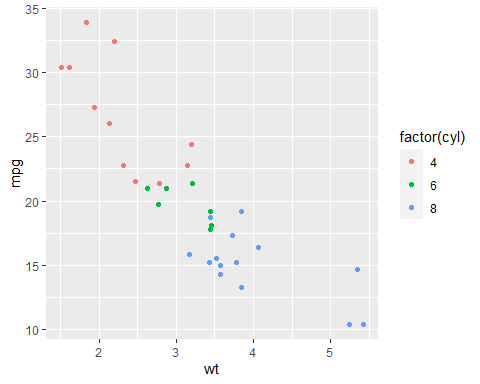
p+geom\_point()



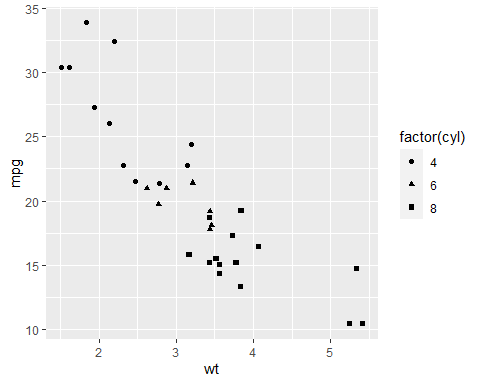
ggplot(mtcars,aes(wt,y=mpg)) +geom\_point()



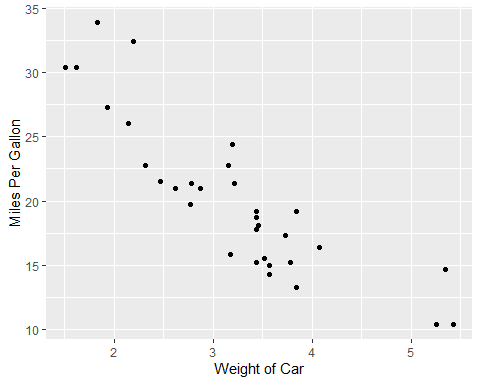
p + geom\_point(aes(colour = factor(cyl)))



p + geom\_point(aes(shape = factor(cyl)))



ggplot(mtcars, aes(x=wt, y=mpg)) + geom\_point() + scale\_x\_continuous("Weight of Car") + scale\_y\_continuous("Miles Per Gallon")

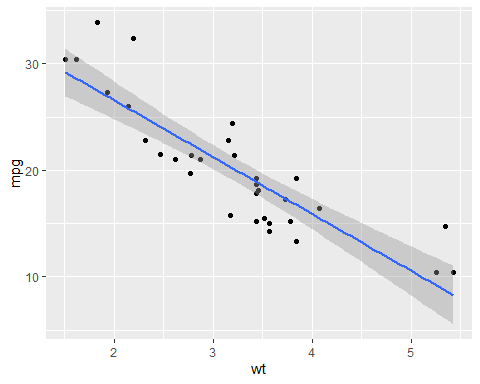


# To find the strength of the linear relationship for the two variables,  
cor(mtcars$wt,mtcars$mpg)

## [1] -0.8676594

#The correlation coefficient is a number between 1 and -1,  
#The magnitude of the number indicates the strength of the linear relationship.  
#The sign corresponds to the direction of that relationship.   
#A correlation coefficient of -1 indicates a near perfect negatively correlated linear   
#relationship.   
#In this case we observe a correlation coefficient of -0.8676594   
#this is an indication of strong, negative linear relationship between the two variables   
#which is what we observed in the scatter plot.  
  
#Adding a regression line   
ggplot(mtcars, aes(x=wt, y=mpg)) + geom\_point()+ geom\_smooth(method=lm)

## `geom\_smooth()` using formula 'y ~ x'



#linear model  
lm(mtcars$mpg~mtcars$wt)

##   
## Call:  
## lm(formula = mtcars$mpg ~ mtcars$wt)  
##   
## Coefficients:  
## (Intercept) mtcars$wt   
## 37.285 -5.344

#to check the statistical significance of the model   
summary(lm(mtcars$mpg~mtcars$wt))

##   
## Call:  
## lm(formula = mtcars$mpg ~ mtcars$wt)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.5432 -2.3647 -0.1252 1.4096 6.8727   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 37.2851 1.8776 19.858 < 2e-16 \*\*\*  
## mtcars$wt -5.3445 0.5591 -9.559 1.29e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.046 on 30 degrees of freedom  
## Multiple R-squared: 0.7528, Adjusted R-squared: 0.7446   
## F-statistic: 91.38 on 1 and 30 DF, p-value: 1.294e-10

#consider a model to statiscally significant when the p value is < 0.05,   
#in this case the p value is significantly less than 0.05