Diamonds.R

preetha.varma

Thu Feb 18 21:31:46 2016

# Load Diamonds  
  
  
library(ggplot2)  
data(diamonds)  
summary(diamonds)

## carat cut color clarity   
## Min. :0.2000 Fair : 1610 D: 6775 SI1 :13065   
## 1st Qu.:0.4000 Good : 4906 E: 9797 VS2 :12258   
## Median :0.7000 Very Good:12082 F: 9542 SI2 : 9194   
## Mean :0.7979 Premium :13791 G:11292 VS1 : 8171   
## 3rd Qu.:1.0400 Ideal :21551 H: 8304 VVS2 : 5066   
## Max. :5.0100 I: 5422 VVS1 : 3655   
## J: 2808 (Other): 2531   
## depth table price x   
## Min. :43.00 Min. :43.00 Min. : 326 Min. : 0.000   
## 1st Qu.:61.00 1st Qu.:56.00 1st Qu.: 950 1st Qu.: 4.710   
## Median :61.80 Median :57.00 Median : 2401 Median : 5.700   
## Mean :61.75 Mean :57.46 Mean : 3933 Mean : 5.731   
## 3rd Qu.:62.50 3rd Qu.:59.00 3rd Qu.: 5324 3rd Qu.: 6.540   
## Max. :79.00 Max. :95.00 Max. :18823 Max. :10.740   
##   
## y z   
## Min. : 0.000 Min. : 0.000   
## 1st Qu.: 4.720 1st Qu.: 2.910   
## Median : 5.710 Median : 3.530   
## Mean : 5.735 Mean : 3.539   
## 3rd Qu.: 6.540 3rd Qu.: 4.040   
## Max. :58.900 Max. :31.800   
##

#Data grouped by price vs color and price vs cut.  
  
  
by(diamonds$price,diamonds$color,sum)

## diamonds$color: D  
## [1] 21476439  
## --------------------------------------------------------   
## diamonds$color: E  
## [1] 30142944  
## --------------------------------------------------------   
## diamonds$color: F  
## [1] 35542866  
## --------------------------------------------------------   
## diamonds$color: G  
## [1] 45158240  
## --------------------------------------------------------   
## diamonds$color: H  
## [1] 37257301  
## --------------------------------------------------------   
## diamonds$color: I  
## [1] 27608146  
## --------------------------------------------------------   
## diamonds$color: J  
## [1] 14949281

by(diamonds$price,diamonds$color,mean)

## diamonds$color: D  
## [1] 3169.954  
## --------------------------------------------------------   
## diamonds$color: E  
## [1] 3076.752  
## --------------------------------------------------------   
## diamonds$color: F  
## [1] 3724.886  
## --------------------------------------------------------   
## diamonds$color: G  
## [1] 3999.136  
## --------------------------------------------------------   
## diamonds$color: H  
## [1] 4486.669  
## --------------------------------------------------------   
## diamonds$color: I  
## [1] 5091.875  
## --------------------------------------------------------   
## diamonds$color: J  
## [1] 5323.818

by(diamonds$price,diamonds$cut,sum)

## diamonds$cut: Fair  
## [1] 7017600  
## --------------------------------------------------------   
## diamonds$cut: Good  
## [1] 19275009  
## --------------------------------------------------------   
## diamonds$cut: Very Good  
## [1] 48107623  
## --------------------------------------------------------   
## diamonds$cut: Premium  
## [1] 63221498  
## --------------------------------------------------------   
## diamonds$cut: Ideal  
## [1] 74513487

by(diamonds$price,diamonds$cut,mean)

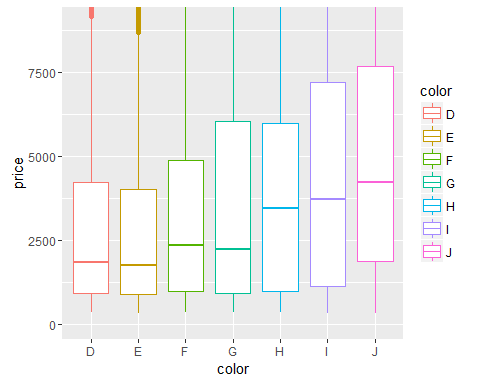
## diamonds$cut: Fair  
## [1] 4358.758  
## --------------------------------------------------------   
## diamonds$cut: Good  
## [1] 3928.864  
## --------------------------------------------------------   
## diamonds$cut: Very Good  
## [1] 3981.76  
## --------------------------------------------------------   
## diamonds$cut: Premium  
## [1] 4584.258  
## --------------------------------------------------------   
## diamonds$cut: Ideal  
## [1] 3457.542

#Summary view of diamonds data based on carat and color variables  
  
  
summary(diamonds$color,diamonds$price)

## Warning in if (length(ll) > maxsum) {: the condition has length > 1 and  
## only the first element will be used

## D E F G H I J   
## 6775 9797 9542 11292 8304 5422 2808

#Boxplot view of color vs price. Its observed that the mean of the price is going higher in the alphabetic order of color names or color category.  
  
  
qplot(x = color, y=price,data = diamonds,  
 geom = 'boxplot',color=color)+  
 coord\_cartesian(ylim = c(0,9000))



#Summary view of diamonds data based on carat and color variables  
  
  
summary(diamonds$cut,diamonds$price)

## Warning in if (length(ll) > maxsum) {: the condition has length > 1 and  
## only the first element will be used

## Fair Good Very Good Premium Ideal   
## 1610 4906 12082 13791 21551

#Boxplot view of color vs price. Its observed that the median price of fair and premium cut diamonds are very close to each other though their prices vary drastically.  
  
  
  
  
qplot(x = cut, y=price,data = diamonds,  
 geom = 'boxplot',color = cut)+  
 coord\_cartesian(ylim = c(0,7000))

