

Diamonds in R

Yimmy Eman

2022-07-07

Análisis de los diamantes

```
diamonds <- read.csv("../data/diamonds.csv")  
dim(diamonds)
```

```
## [1] 53940    10
```

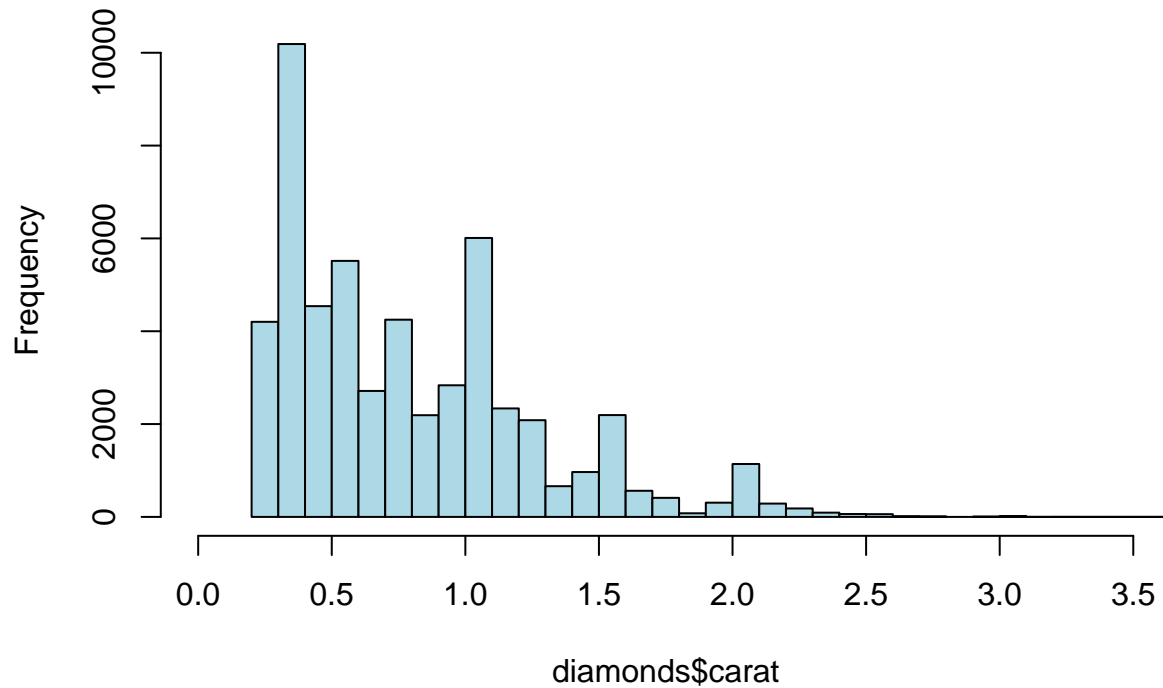
```
head(diamonds)
```

```
##   carat      cut color clarity depth table price     x     y     z  
## 1  0.23    Ideal     E    SI2   61.5    55   326  3.95  3.98  2.43  
## 2  0.21  Premium     E    SI1   59.8    61   326  3.89  3.84  2.31  
## 3  0.23     Good     E    VS1   56.9    65   327  4.05  4.07  2.31  
## 4  0.29  Premium     I    VS2   62.4    58   334  4.20  4.23  2.63  
## 5  0.31     Good     J    SI2   63.3    58   335  4.34  4.35  2.75  
## 6  0.24 Very Good     J   VVS2   62.8    57   336  3.94  3.96  2.48
```

Histograma

```
hist(diamonds$carat, breaks = 50, col = "lightblue", xlim = c(0,3.5))
```

Histogram of diamonds\$carat



##

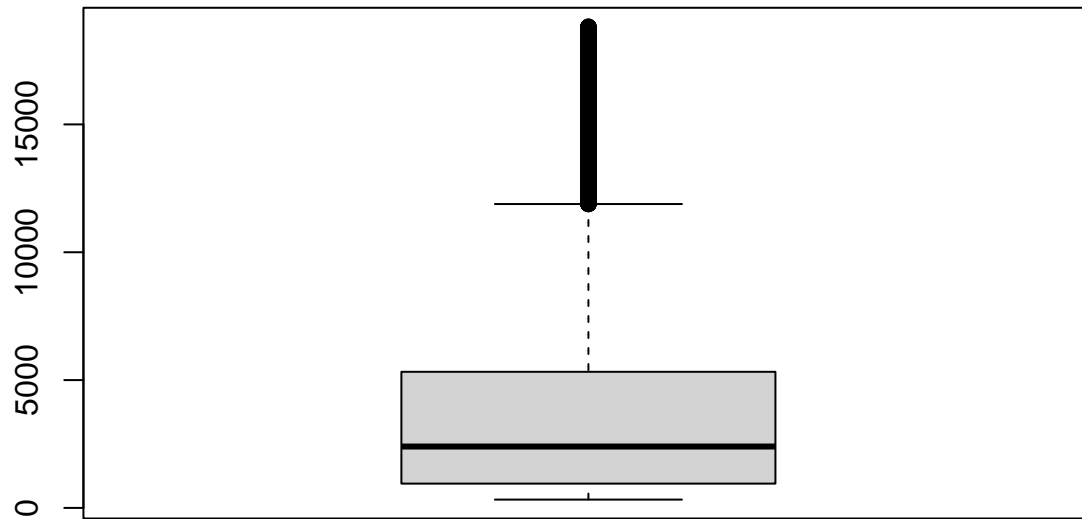
Filtro de outliers

```
subset(diamonds, diamonds$carat > 3.5)
```

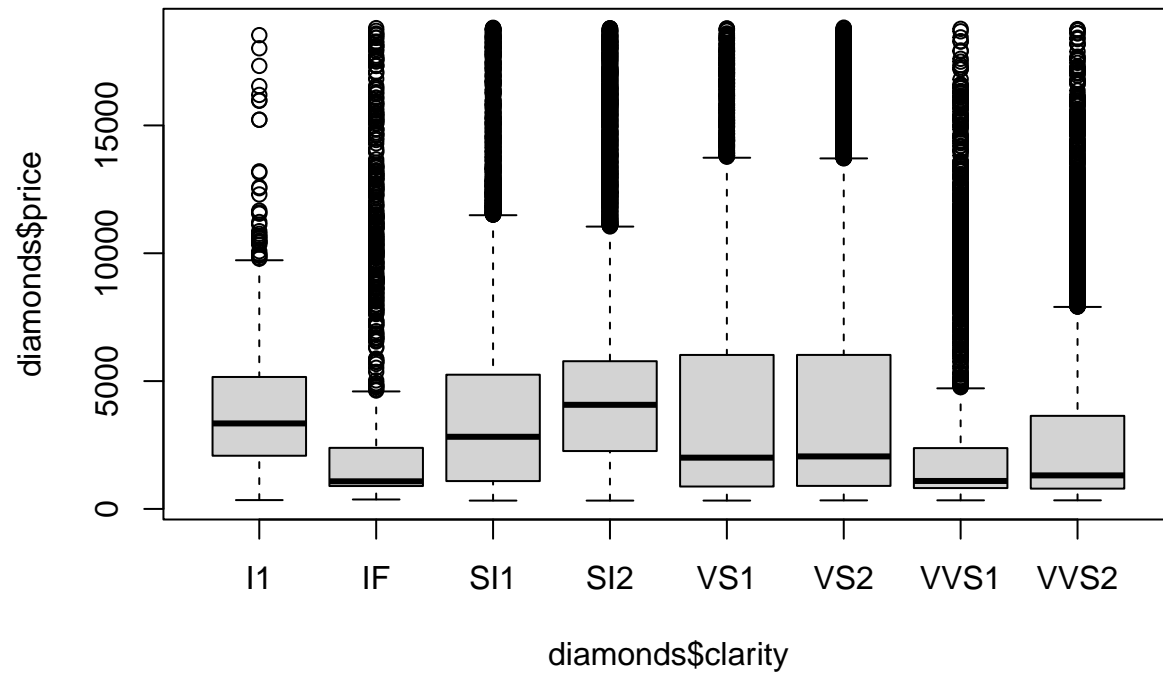
##	carat	cut	color	clarity	depth	table	price	x	y	z
## 23645	3.65	Fair	H	I1	67.1	53	11668	9.53	9.48	6.38
## 25999	4.01	Premium	I	I1	61.0	61	15223	10.14	10.10	6.17
## 26000	4.01	Premium	J	I1	62.5	62	15223	10.02	9.94	6.24
## 26445	4.00	Very Good	I	I1	63.3	58	15984	10.01	9.94	6.31
## 26535	3.67	Premium	I	I1	62.4	56	16193	9.86	9.81	6.13
## 27131	4.13	Fair	H	I1	64.8	61	17329	10.00	9.85	6.43
## 27416	5.01	Fair	J	I1	65.5	59	18018	10.74	10.54	6.98
## 27631	4.50	Fair	J	I1	65.8	58	18531	10.23	10.16	6.72
## 27680	3.51	Premium	J	VS2	62.5	59	18701	9.66	9.63	6.03

Boxplots

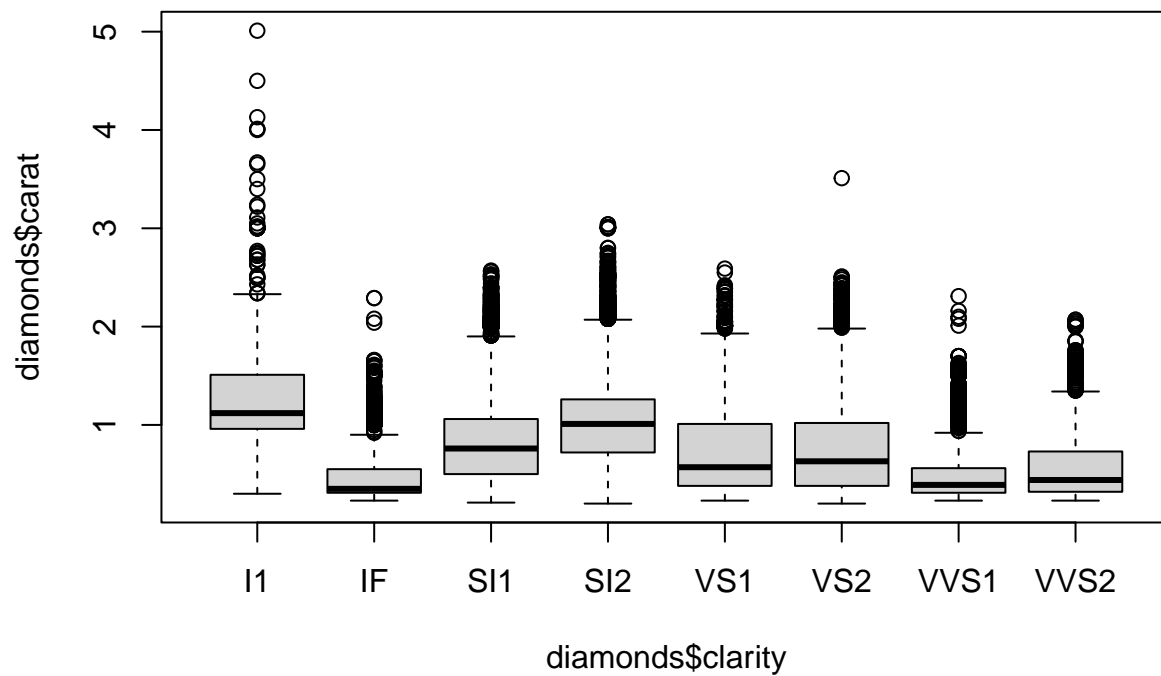
```
boxplot(diamonds$price)
```



```
boxplot(diamonds$price ~ diamonds$clarity)
```



```
boxplot(diamonds$carat ~ diamonds$clarity)
```



Densidades

##

```
plot(density(diamonds$carat), col = "gray")
```

density.default(x = diamonds\$carat)

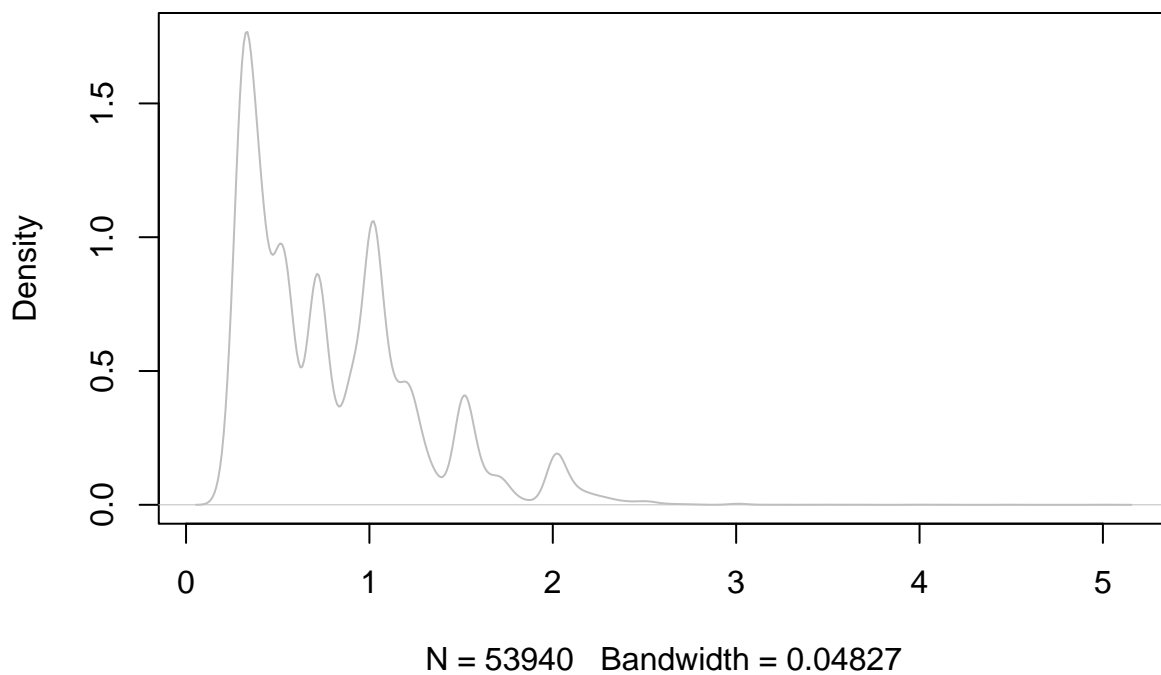
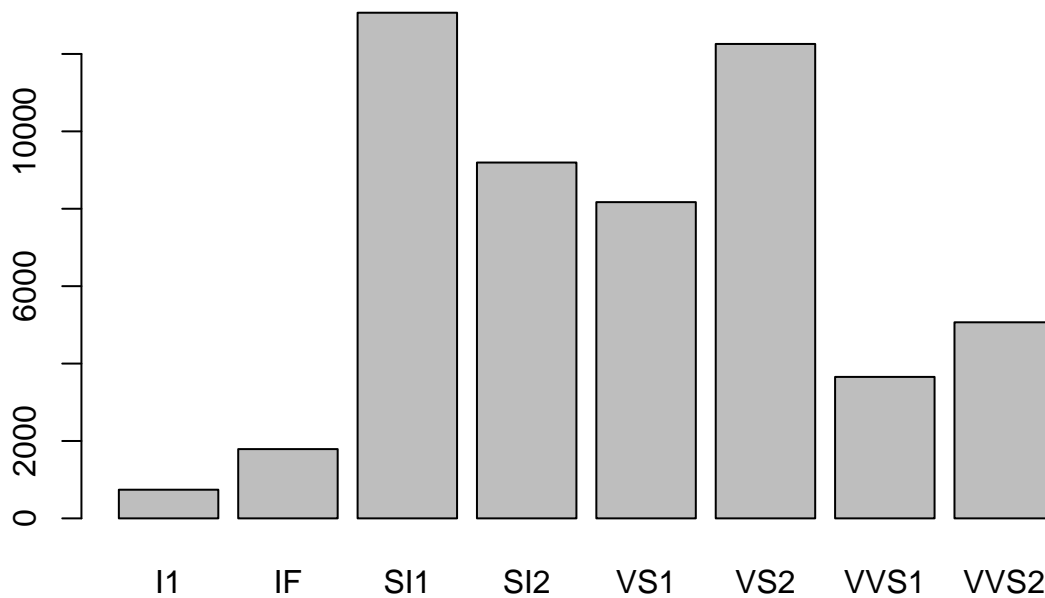


Tabla de frecuencias y Barplot

```
table(diamonds$clarity)
```

```
##  
##      I1      IF     SI1     SI2     VS1     VS2     VVS1     VVS2  
##    741    1790  13065   9194   8171  12258   3655   5066
```

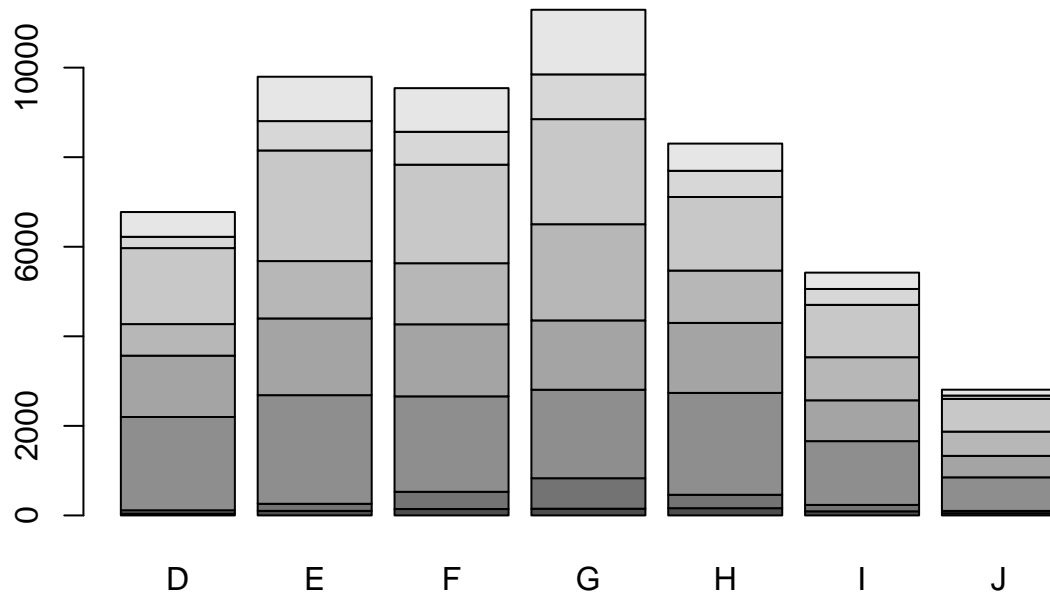
```
barplot(table(diamonds$clarity))
```



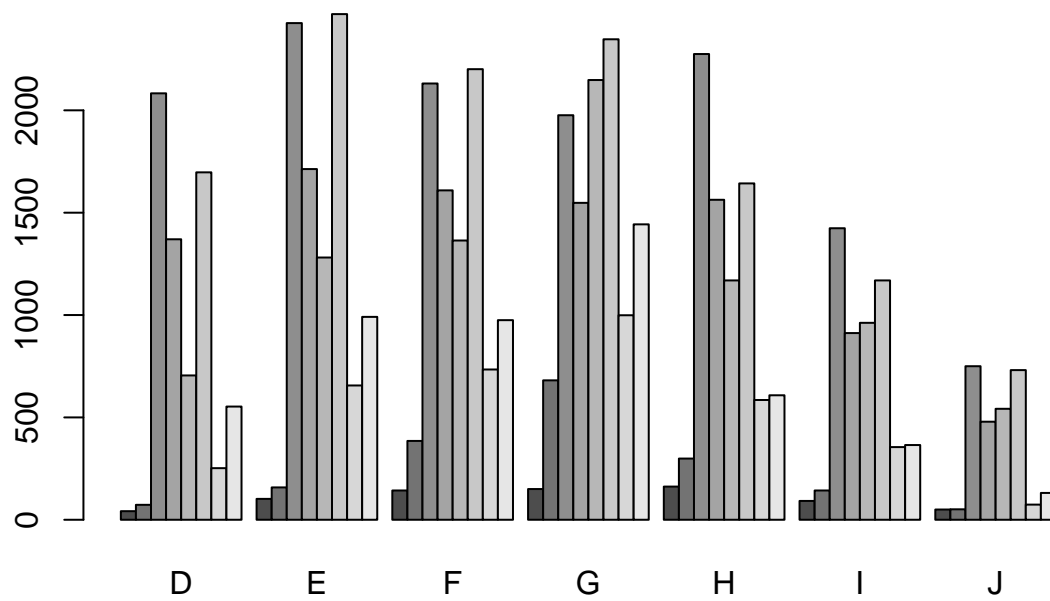
```
table(diamonds$clarity, diamonds$color)
```

```
##  
##      D      E      F      G      H      I      J  
##  I1      42     102    143    150    162     92     50  
##  IF       73     158    385    681    299    143     51  
##  SI1    2083   2426   2131   1976   2275   1424    750  
##  SI2    1370   1713   1609   1548   1563    912    479  
##  VS1     705   1281   1364   2148   1169    962    542  
##  VS2    1697   2470   2201   2347   1643   1169    731  
##  VVS1    252    656    734    999    585    355     74  
##  VVS2    553    991    975   1443    608    365    131
```

```
barplot(table(diamonds$clarity, diamonds$color), beside = FALSE)
```



```
barplot(table(diamonds$clarity, diamonds$color), beside = TRUE)
```



Scatterplot

```
plot(diamonds$carat, diamonds$price, col = "gray", pch = 21, cex = 0.4, ylim = c(0,20000))
```

