

Diamonds

Raju Adhikari & Mario Alonso Rodriguez

October 6th, 2016

Data

First, we have to select a dataset from the ones already preloaded in R

```
### Listing the in-built dataset from R
data()
```

```
### Loading the dataset diamonds and obtaining its description
data(diamonds)
? diamonds
```

```
### Listing all the variables that the dataset contains
names(diamonds)
```

```
## [1] "carat" "cut" "color" "clarity" "depth" "table" "price"
## [8] "x" "y" "z"
```

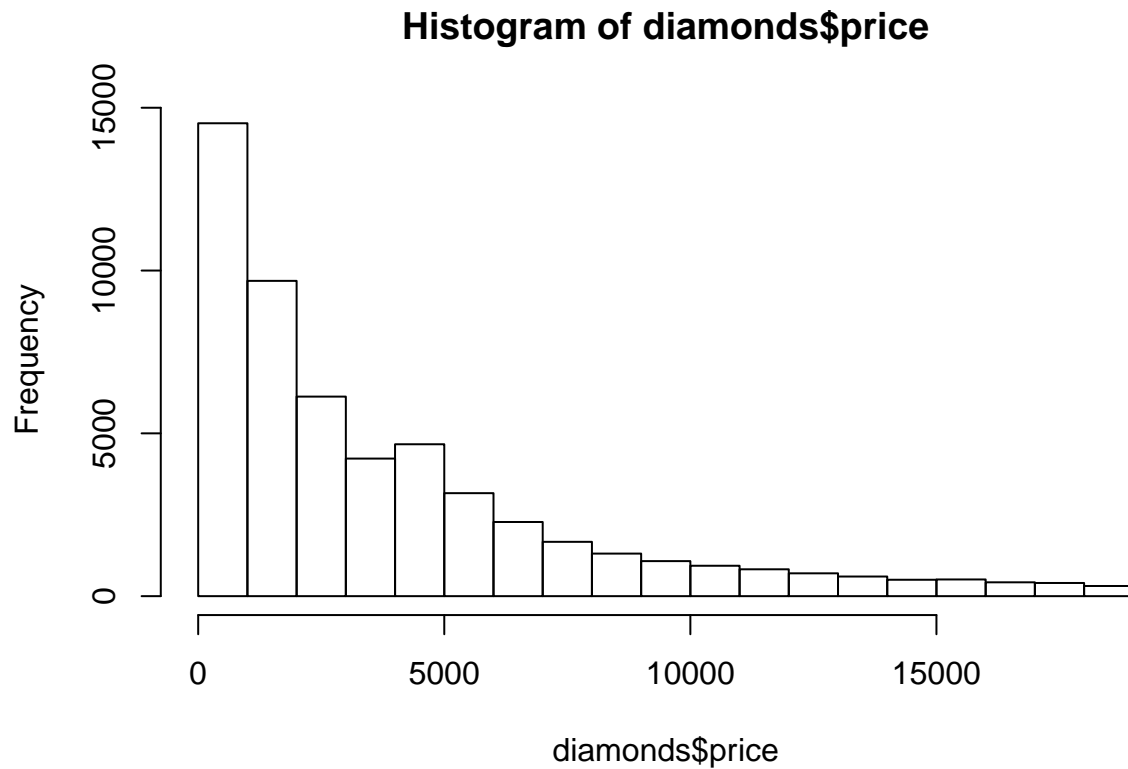
```
### Getting a descriptive summary of the variables in the dataset
summary (diamonds)
```

```
##      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
##
```

Histogram

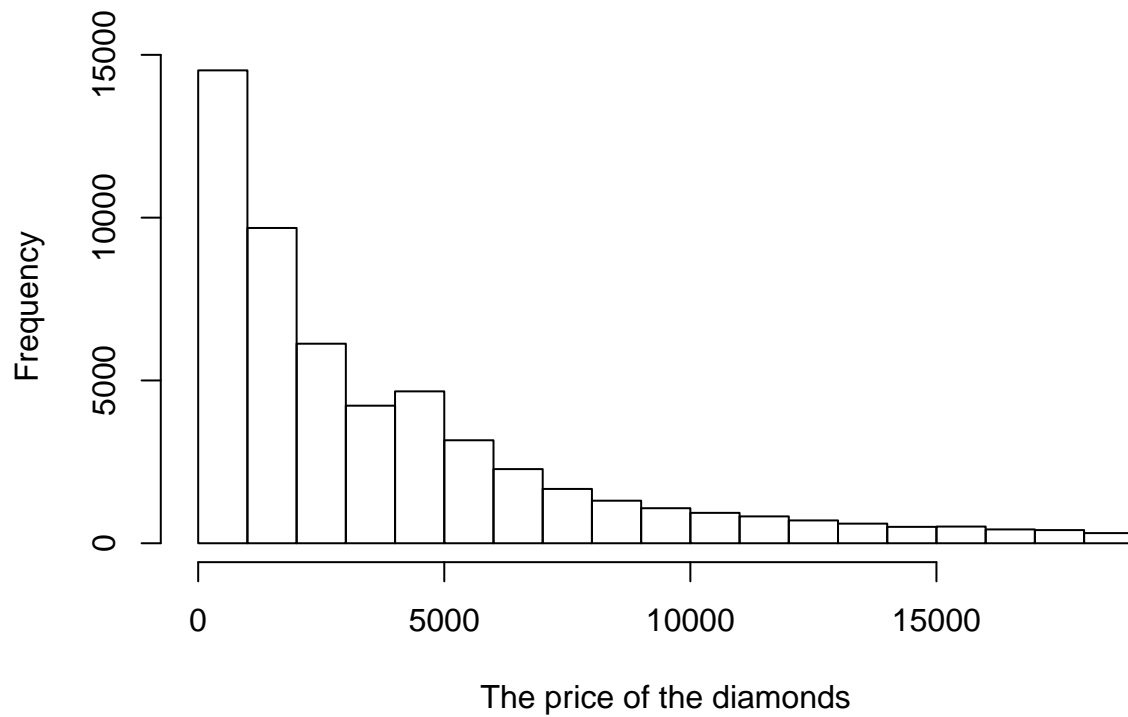
A histogram of price variable shows the number of diamonds at different price level

```
#### Create the histogram  
hist(diamonds$price)
```



```
hist(diamonds$price,  
     main = 'Diamonds at Different Price Levels',  
     xlab = 'The price of the diamonds')
```

Diamonds at Different Price Levels



Statistics

A series of commands will show us the different descriptive statistics to analyze the data

```
### Finding the mean, one of the main measures of central tendency  
mean(diamonds$price, na.rm = TRUE)
```

```
## [1] 3932.8
```

```
mean(diamonds$depth, na.rm = TRUE)
```

```
## [1] 61.7494
```

```
mean(diamonds$carat, na.rm = TRUE)
```

```
## [1] 0.7979397
```

```
mean(diamonds$table, na.rm = TRUE)
```

```
## [1] 57.45718
```

```
mean(diamonds$x, na.rm = TRUE)
```

```
## [1] 5.731157
```

```
mean(diamonds$y, na.rm = TRUE)
```

```
## [1] 5.734526
```

```
mean(diamonds$z, na.rm = TRUE)
```

```
## [1] 3.538734
```

```
### Finding the standard deviation for price variable  
sd(diamonds$price)
```

```
## [1] 3989.44
```

Tables

A Crosstab shows how two variables are related

```
### Crosstab shows that color and cut are roughly equally distributed  
table(diamonds$cut, diamonds$color)
```

```
##  
##           D      E      F      G      H      I      J  
## Fair      163   224   312   314   303   175   119  
## Good      662   933   909   871   702   522   307  
## Very Good 1513  2400  2164  2299  1824  1204   678  
## Premium   1603  2337  2331  2924  2360  1428   808  
## Ideal     2834  3903  3826  4884  3115  2093   896
```

Please find the other part of this assignment on this link

Drinks