

Diamonds

```
library(readxl)
diamond=read_excel("diamonds.xlsx")
```

```
dim(iris)
```

```
## [1] 150 5
```

```
diamond[50000,3]
```

```
## # A tibble: 1 x 1
##   color
##   <chr>
## 1     H
```

now im going to see the first 6 rows using the head command

```
head(diamond)
```

```
## # A tibble: 6 x 10
##   carat      cut color clarity depth table price      x      y      z
##   <dbl>    <chr> <chr>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 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
```

```
head(diamond)
```

```
## # A tibble: 6 x 10
##   carat      cut color clarity depth table price      x      y      z
##   <dbl>    <chr> <chr>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 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
```

```
tail(diamond) # last 6 rows
```

```
## # A tibble: 6 x 10
##   carat      cut color clarity depth table price      x      y      z
##   <dbl>    <chr> <chr>   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1  0.72  Premium     D     SI1  62.7   59  2757  5.69  5.73  3.58
## 2  0.72    Ideal     D     SI1  60.8   57  2757  5.75  5.76  3.50
## 3  0.72     Good     D     SI1  63.1   55  2757  5.69  5.75  3.61
## 4  0.70 Very Good     D     SI1  62.8   60  2757  5.66  5.68  3.56
## 5  0.86  Premium     H     SI2  61.0   58  2757  6.15  6.12  3.74
## 6  0.75    Ideal     D     SI2  62.2   55  2757  5.83  5.87  3.64
```

summarize my data

```
summary(diamond)
```

```
##      carat      cut      color      clarity
## Min.   :0.2000 Length:53940 Length:53940 Length:53940
## 1st Qu.:0.4000 Class :character Class :character Class :character
## Median :0.7000 Mode  :character Mode  :character Mode  :character
## Mean   :0.7979
## 3rd Qu.:1.0400
## Max.   :5.0100
##      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
```

Making the cut variable a factor variable

```
diamond$cut=as.factor(diamond$cut)
summary(diamond)
```

```
##      carat      cut      color      clarity
## Min.   :0.2000 Fair      : 1610 Length:53940 Length:53940
## 1st Qu.:0.4000 Good      : 4906 Class :character Class :character
## Median :0.7000 Ideal     :21551 Mode  :character Mode  :character
## Mean   :0.7979 Premium   :13791
## 3rd Qu.:1.0400 Very Good:12082
## Max.   :5.0100
##      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
```

```
diamond$color=as.factor(diamond$color)
summary(diamond)
```

```
##      carat      cut      color      clarity
## Min.   :0.2000 Fair      : 1610 D: 6775 Length:53940
```

```

## 1st Qu.:0.4000    Good      : 4906    E: 9797    Class :character
## Median :0.7000    Ideal      :21551    F: 9542    Mode  :character
## Mean   :0.7979    Premium    :13791    G:11292
## 3rd Qu.:1.0400    Very Good:12082    H: 8304
## Max.   :5.0100                                I: 5422
##                                              J: 2808
##      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
##

```