

Marketing Analytics Course 1 module 4

First, we upload our times series data.

```
realdata.df <- read.csv("edx_datavis_time.csv", header=TRUE)
```

We check our upload by doing a summary of our data

```
summary(realdata.df)
```

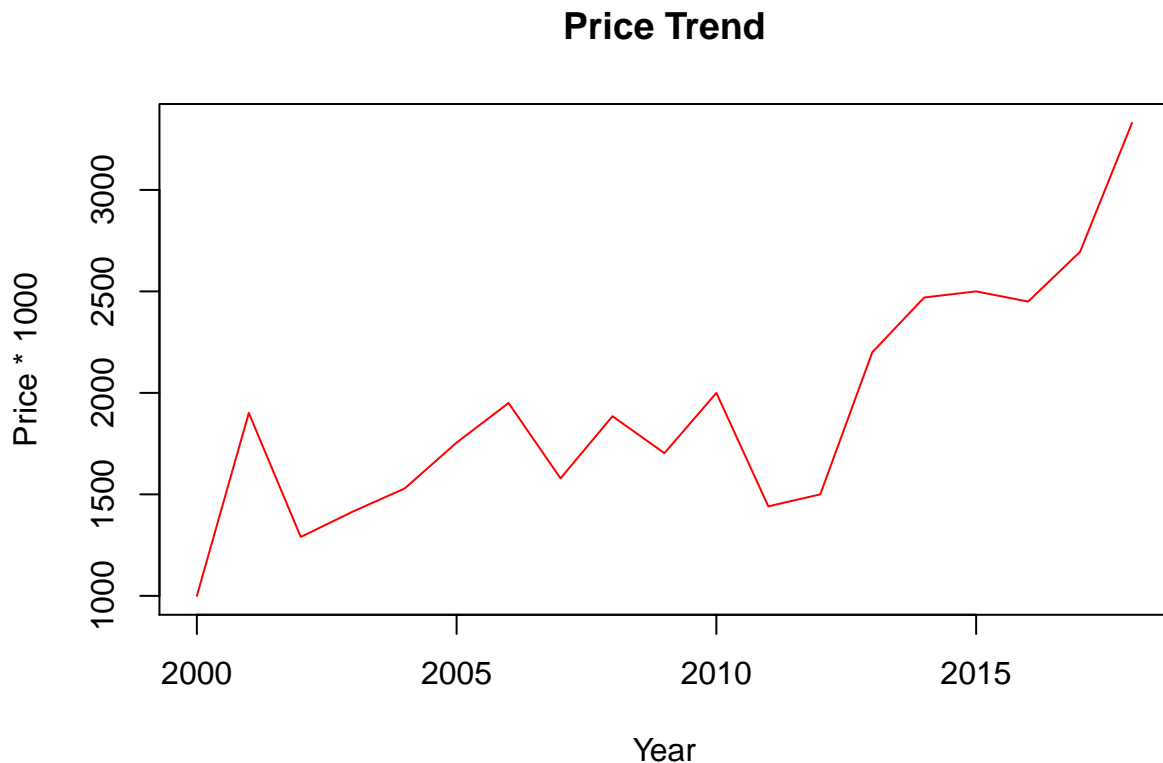
```
##      Price1000      Year
##  Min.   :1000   Min.   :2000
##  1st Qu.:1514   1st Qu.:2004
##  Median :1885   Median :2009
##  Mean   :1926   Mean   :2009
##  3rd Qu.:2325   3rd Qu.:2014
##  Max.   :3330   Max.   :2018
```

We create a variables year and price and assign the year and price columns to them.

```
year <- realdata.df$Year
price <- realdata.df$Price1000
```

Here we do a simple line plot in red of year and data

```
plot(year, price, type="l", col="red", main="Price Trend", xlab="Year", ylab="Price * 1000")
```



Now we do a bar plot. Bar plots are useful for comparing statistics of one variable.

```
data.df <- read.csv ("edx_rbasics__1_.csv", header =TRUE)
```

We summarize to check if that data was read in correctly.

```
summary(data.df)
```

```
##      Price1000      SizeSqFt      LotAcre      Year
## Min.   : 1925   Min.   :1660   Min.   : 0.200   Min.   :1950
## 1st Qu.: 2870   1st Qu.:2640   1st Qu.: 0.340   1st Qu.:1956
## Median : 4025   Median :3150   Median : 1.020   Median :1959
## Mean   : 5069   Mean   :3446   Mean   : 1.922   Mean   :1969
## 3rd Qu.: 5695   3rd Qu.:3900   3rd Qu.: 2.060   3rd Qu.:1972
## Max.   :16988   Max.   :8333   Max.   :17.990   Max.   :2015
##      Bedrooms      Bathrooms      NeighCentralPV      NeighLadera
## Min.   :2.000   Min.   :2.00   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:3.000   1st Qu.:3.00   1st Qu.:0.0000   1st Qu.:0.0000
## Median :4.000   Median :3.00   Median :0.0000   Median :0.0000
## Mean   :3.918   Mean   :3.51   Mean   :0.3878   Mean   :0.3265
## 3rd Qu.:4.000   3rd Qu.:4.00   3rd Qu.:1.0000   3rd Qu.:1.0000
## Max.   :6.000   Max.   :8.00   Max.   :1.0000   Max.   :1.0000
```

We are going to use an aggregate function to find the mean for all the prices of different bedrooms.

```
chartdata <- aggregate (data.df$Price1000, by=list (data.df$Bedrooms), FUN=mean)
```

We print the chartdata to see the mean of all the different houses by the number of bedrooms that they have.

```
chartdata
```

```
##      Group.1      x
## 1          2 4194.000
## 2          3 3892.733
## 3          4 4343.952
## 4          5 8026.455
## 5          6 6300.000
```

Now we create names for our chartdata chart.

```
names (chartdata) <- c("Bedrooms", "MeanPrice")
```

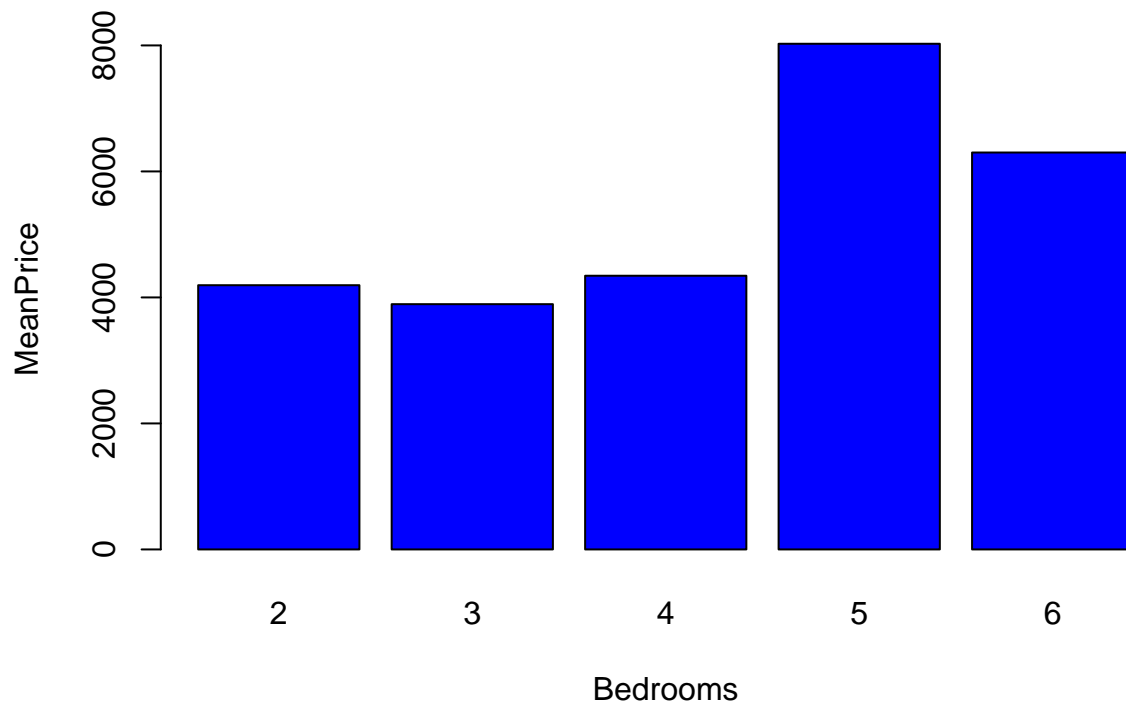
We print with our labeled columns.

```
chartdata
```

```
##      Bedrooms MeanPrice
## 1          2 4194.000
## 2          3 3892.733
## 3          4 4343.952
## 4          5 8026.455
## 5          6 6300.000
```

We then create a bar plot with our chartdata.

```
barplot (chartdata$MeanPrice, names.arg=chartdata$Bedrooms, xlab="Bedrooms", ylab="MeanPrice", col="blue")
```



Scatterplots are useful for finding relationships between two variables.

We assign the two variables we'd like to compare. LotAcre to lot and SizeSqFt to house.

```
lot <- data.df$LotAcre
house <- data.df$SizeSqFt
```

We will now print both lot and house columns

```
lot
```

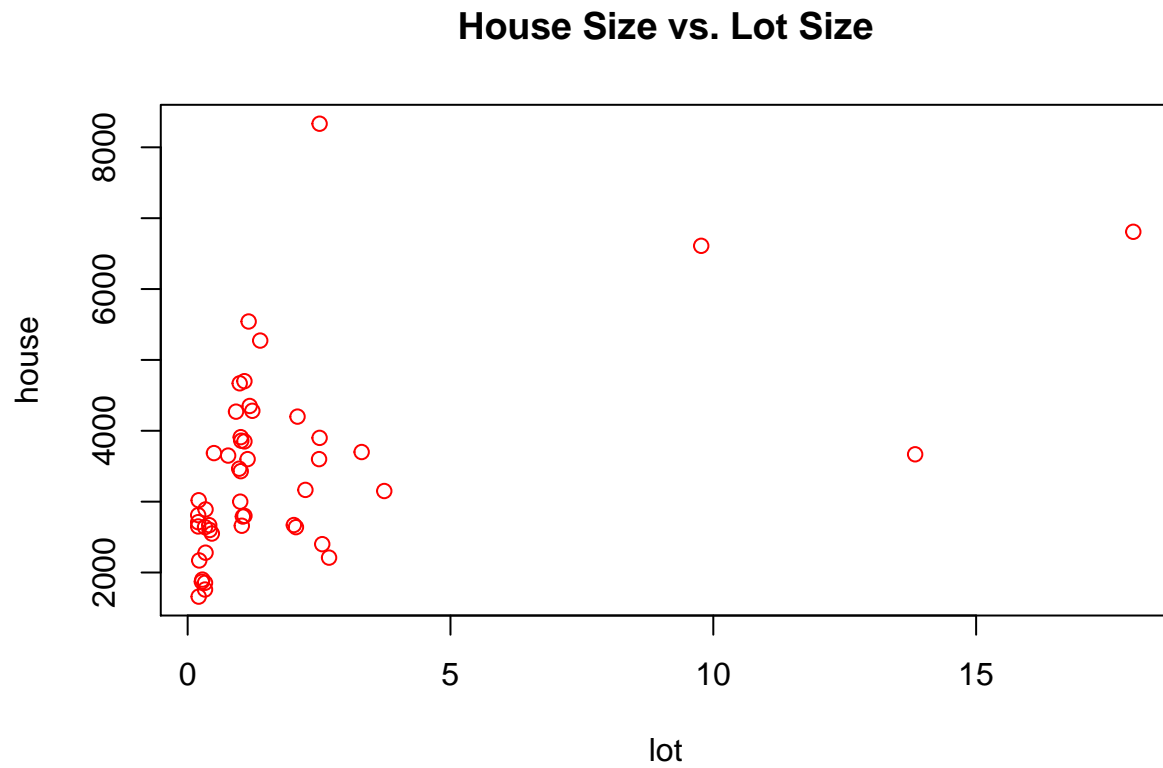
```
## [1] 0.33 0.28 0.33 0.41 0.98 9.77 0.99 17.99 1.16 13.84 1.14
## [12] 2.50 1.08 2.51 0.21 2.69 0.27 0.50 1.38 1.03 2.51 1.18
## [23] 1.05 2.06 0.20 1.23 0.42 2.02 0.20 0.34 1.00 0.20 1.08
## [34] 1.01 0.77 0.46 3.74 2.09 0.33 1.08 2.56 0.92 1.02 1.01
## [45] 0.21 3.31 0.34 2.24 0.22
```

```
house
```

```
## [1] 2640 1900 1760 2668 3465 6610 4670 6808 5542 3668 3600 3600 4700 3900
## [15] 1660 2210 1870 3685 5273 2660 8333 4350 2790 2640 2650 4282 2600 2670
## [29] 2710 2890 3000 2810 2800 3910 3650 2550 3150 4200 1855 3850 2400 4270
## [43] 3860 3430 3020 3700 2280 3166 2170
```

Here is our of house size versus Lot Size scatter plot.

```
plot(lot, house, col="red", main="House Size vs. Lot Size")
```



We interpret the chart and see that there is a linear and clustering relationship that does suggest that there is a small correlation between the house size and lot size. After five thousand feet there are large outliers and the relationship between lot size and house size decreases.