

#VIVEK PAKALAPATI
#U49253220

1) R script

```
rm(list = ls())  
library(rio)  
getwd()
```

#File imported in text format

```
house.p=import("HousePrices.csv")  
  
colnames(house.p)=tolower(make.names(colnames(house.p)))  
attach(house.p)
```

#Histogram for distribution of prices

```
hist(price,col = "pink",main = "distribution of prices")  
summary(price)  
sd(price)
```

#Histogram for distribution of sqft

```
hist(sqft,col="red",main="distribution of sqft")  
summary(sqft)  
sd(sqft)  
skewness(sqft)
```

#Corrplot between prices and sqft

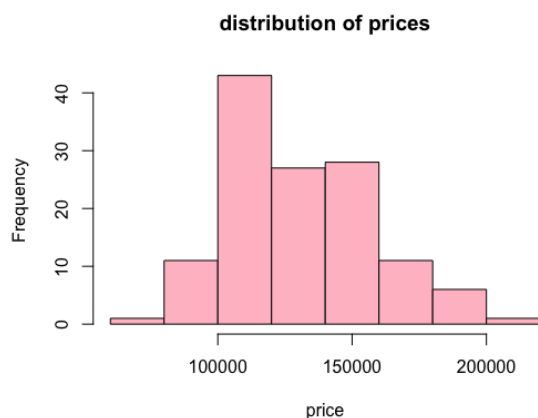
```
library(corrplot)  
a1=cor(house.p[,c(1,2)])  
corrplot(a1)
```

#Standard deviation of sqft

```
set.seed(3220)  
mean.sd=replicate(1000,sd(sample(sqft,size=30,replace = TRUE)))  
sd(sample.sd)  
sd(sqft)
```

2) The results of the script's execution.

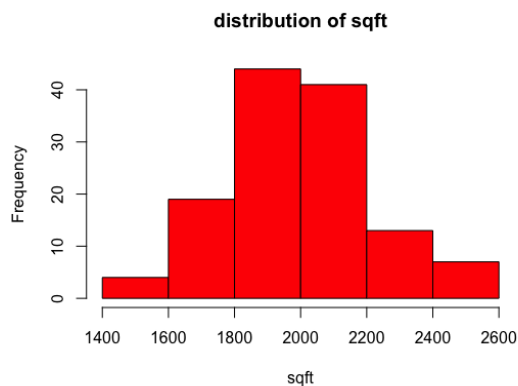
```
rm(list = ls())
> library(rio)
> getwd()
[1] "/Users/vivekvarma"
> library(readxl)
> house.p=import("HousePrices.csv")
> colnames(house.p)=tolower(make.names(colnames(house.p)))
> attach(house.p)
> hist(price,col = "pink",main = "distribution of prices")
```



```
> summary(price)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 69100 111325 125950 130427 148250 211200

> sd(price)
[1] 26868.77

> hist(sqft,col="red",main="distribution of sqft")
hist(sqft,col="red",main="distribution of sqft")
```

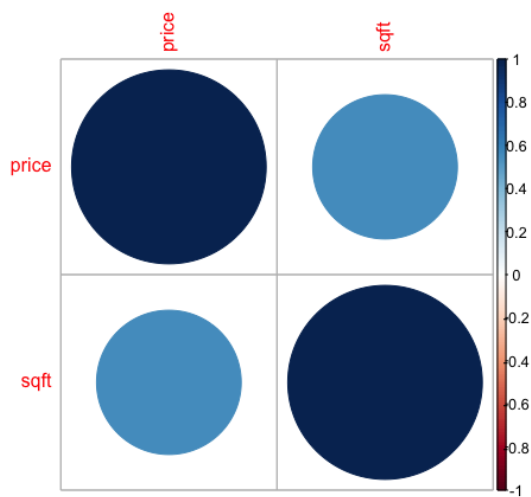


```
> summary(sqft)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  1450   1880   2000   2001   2140   2590
```

```
> sd(sqft)
[1] 211.5724
```

```
> skewness(sqft)
[1] 0.07755647
```

```
> library(corrplot)
> a1=cor(house.p[,c(1,2)])
> corrplot(a1)
```



```
set.seed(3220)
> mean.sd=replicate(1000,sd(sample(sqft,size=30,replace = TRUE)))
> sd(sample.sd)
Error in is.data.frame(x) : object 'sample.sd' not found
> sd(sqft)
[1] 211.5724
```

