# Assignment-1

Ahmed Ashraf Mohamed ID:2103134

## Importing the libraries and the data

Importing the data and printing it.

```
data("ames")
print(ames)
```

```
## # A tibble: 2,930 x 82
##    Order       PID  area  price MS.SubClass MS.Zo~1 Lot.F~2 Lot.A~3 Street Alley
##    <int>     <int> <int>  <int>       <int> <fct>     <int>   <int> <fct>  <fct>
##  1     1 526301100  1656 215000          20 RL          141   31770 Pave   <NA>
##  2     2 526350040   896 105000          20 RH           80   11622 Pave   <NA>
##  3     3 526351010  1329 172000          20 RL           81   14267 Pave   <NA>
##  4     4 526353030  2110 244000          20 RL           93   11160 Pave   <NA>
##  5     5 527105010  1629 189900          60 RL           74   13830 Pave   <NA>
##  6     6 527105030  1604 195500          60 RL           78    9978 Pave   <NA>
##  7     7 527127150  1338 213500         120 RL           41    4920 Pave   <NA>
##  8     8 527145080  1280 191500         120 RL           43    5005 Pave   <NA>
##  9     9 527146030  1616 236500         120 RL           39    5389 Pave   <NA>
## 10    10 527162130  1804 189000          60 RL           60    7500 Pave   <NA>
## # ... with 2,920 more rows, 72 more variables: Lot.Shape <fct>,
## #   Land.Contour <fct>, Utilities <fct>, Lot.Config <fct>, Land.Slope <fct>,
## #   Neighborhood <fct>, Condition.1 <fct>, Condition.2 <fct>, Bldg.Type <fct>,
## #   House.Style <fct>, Overall.Qual <int>, Overall.Cond <int>,
## #   Year.Built <int>, Year.Remod.Add <int>, Roof.Style <fct>, Roof.Matl <fct>,
## #   Exterior.1st <fct>, Exterior.2nd <fct>, Mas.Vnr.Type <fct>,
## #   Mas.Vnr.Area <int>, Exter.Qual <fct>, Exter.Cond <fct>, ...
```

## Exercise 1

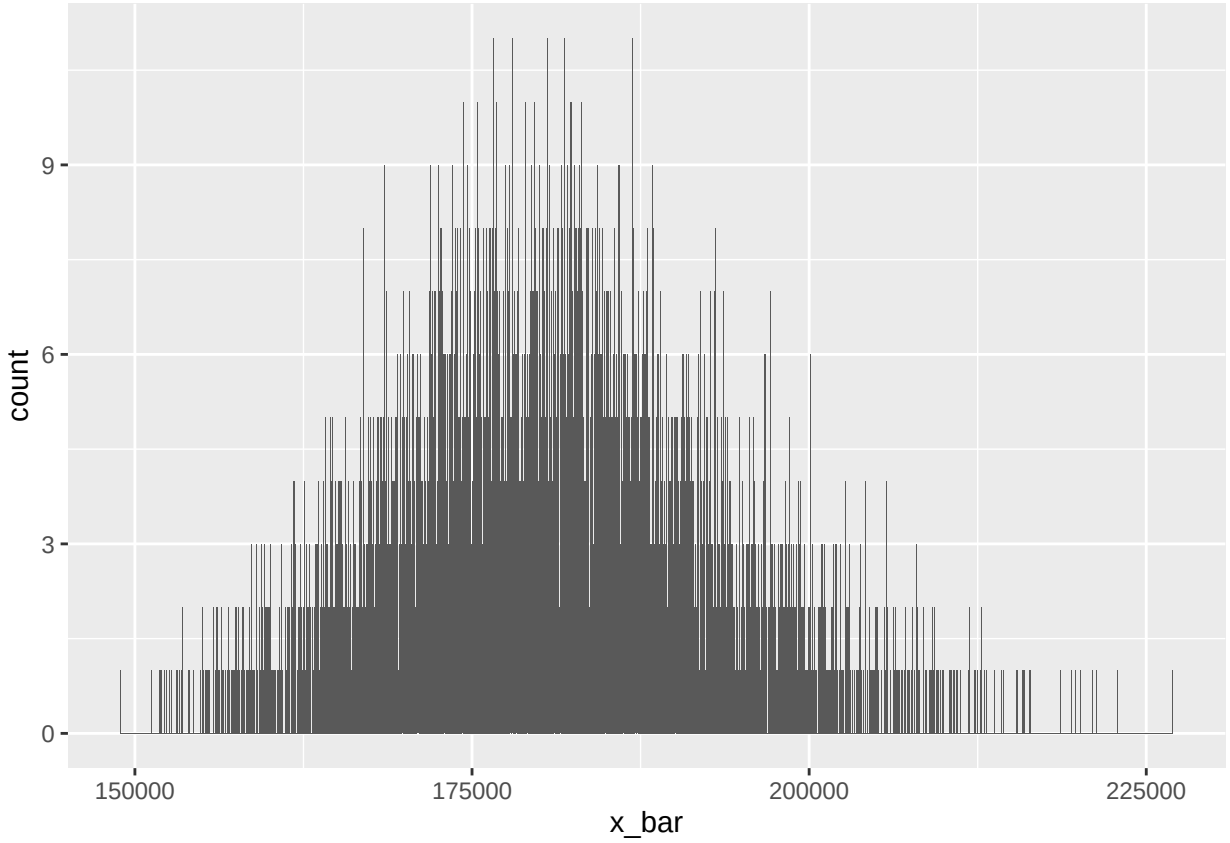
```
price_sample <- ames %>% sample_n(size = 50)
price_sample %>%
summarise(mu = mean(price),
          pop_med = median(price),
          sigma = sd(price), pop_iqr = IQR(price),
          pop_min = min(price),pop_max = max(price),
          pop_q1 = quantile(area, 0.25),
          pop_q3 = quantile(area, 0.75))
```

```
## # A tibble: 1 x 8
##        mu pop_med  sigma pop_iqr pop_min pop_max pop_q1 pop_q3
##     <dbl>   <dbl>  <dbl>   <dbl>   <int>   <int>  <dbl>  <dbl>
## 1 191230.  176895 69773.  101250   68400  395000   1161  1859.
```

## Exercise 2

```
sample_means50 <- ames %>%
rep_sample_n(size = 50, reps = 5000, replace = TRUE) %>%
summarise(x_bar = mean(price))

ggplot(data = sample_means50, aes(x = x_bar)) + geom_histogram(binwidth = 30)
```




Based on this plot, it seem that the distribution is a normal distribution. Eyeing this the means seems around 180,000 mark, to make sure we are going to print the mean of our sample.

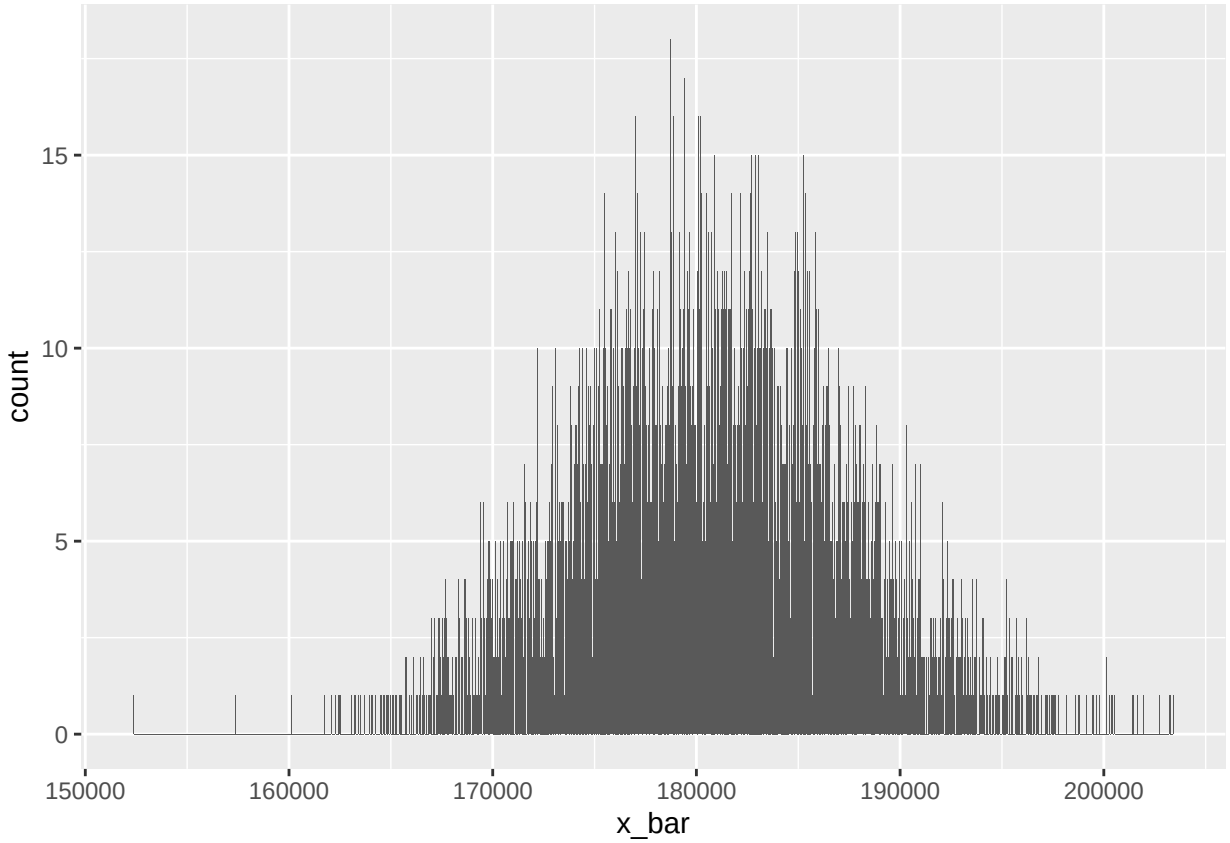
```
print(paste("Mean : ",mean(sample_means50$x_bar)))
```

```
## [1] "Mean :  180951.489664"
```

## Exercise 3

```
sample_means150 <- ames %>%
rep_sample_n(size = 150, reps = 5000, replace = TRUE) %>%
summarise(x_bar = mean(price))

ggplot(data = sample_means150, aes(x = x_bar)) + geom_histogram(binwidth = 30)
```




Looking at this plot and comparing it to the previous plot, it seems clear that the mean is just a bit higher than 180,000, also noting that both plots are normal distributions.

```
print(paste("Mean:",mean(sample_means150$x_bar)))
```

```
## [1] "Mean: 180734.234645333"
```

## Exercise 4

```
price_sample15 <- ames %>% sample_n(size = 15)
print(as_tibble(price_sample15$price))
```

```
## # A tibble: 15 x 1
##     value
##     <int>
##  1 405000
##  2 184750
##  3 222000
##  4 172500
##  5 163500
##  6 175500
##  7 154000
##  8 119750
##  9 164500
## 10 158000
## 11 239000
## 12 155000
```

```
## 13 148000
## 14 148400
## 15 162000
```

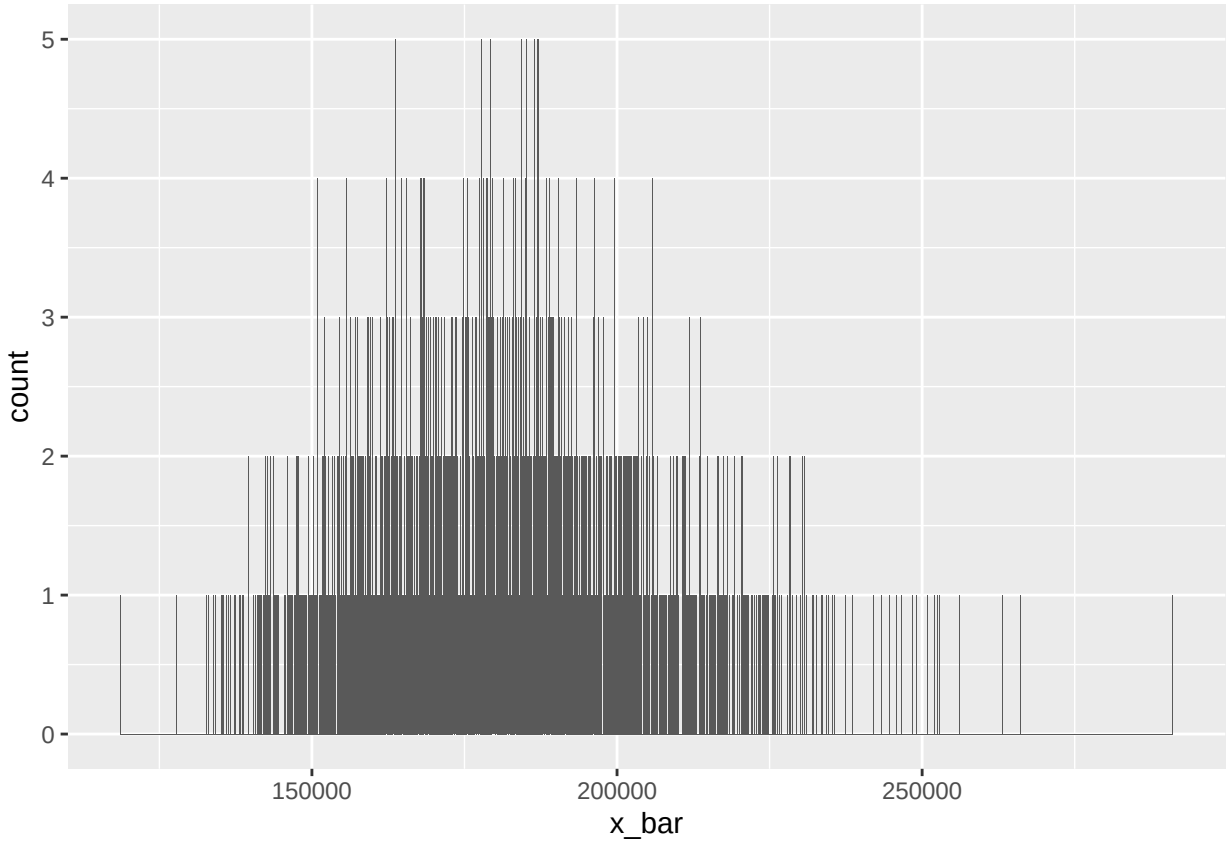
```
print(mean(price_sample15$price))
```

```
## [1] 184793.3
```

## Exercise 5

```
sample_means15 <-ames %>%
  rep_sample_n(size = 15,reps = 2000,replace = TRUE) %>%
  summarise(x_bar = mean(price))

ggplot(data = sample_means15, aes(x = x_bar)) + geom_histogram(binwidth = 30)
```




```
print(mean(sample_means15$x_bar))
```

```
## [1] 181339.5
```

### Population Summary

```
ames %>%
summarise(mu = mean(price),
          pop_med = median(price),
          sigma = sd(price), pop_iqr = IQR(price),
```

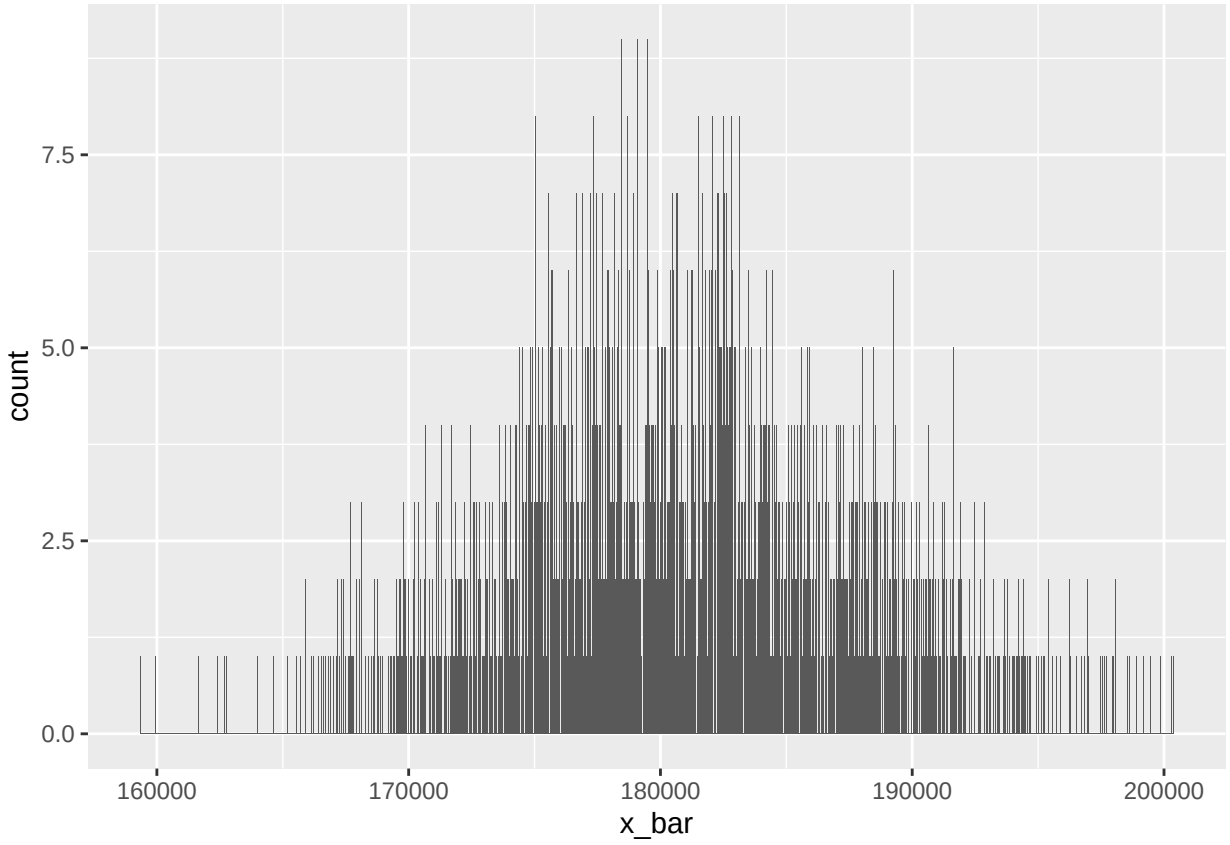
```
          pop_min = min(price),pop_max = max(price),
          pop_q1 = quantile(area, 0.25),
          pop_q3 = quantile(area, 0.75))
```

```
## # A tibble: 1 x 8
##        mu pop_med  sigma pop_iqr pop_min pop_max pop_q1 pop_q3
##     <dbl>   <dbl>  <dbl>   <dbl>   <int>   <int>  <dbl>  <dbl>
## 1 180796.  160000 79887.   84000   12789  755000   1126  1743.
```

## Exercise 6

```
sample_means150 <-ames %>%
  rep_sample_n(size = 150,reps = 2000,replace = TRUE) %>%
  summarise(x_bar = mean(price))

ggplot(data = sample_means150, aes(x = x_bar)) + geom_histogram(binwidth = 30)
```




```
print(mean(sample_means150$x_bar))
```

```
## [1] 180898.6
```