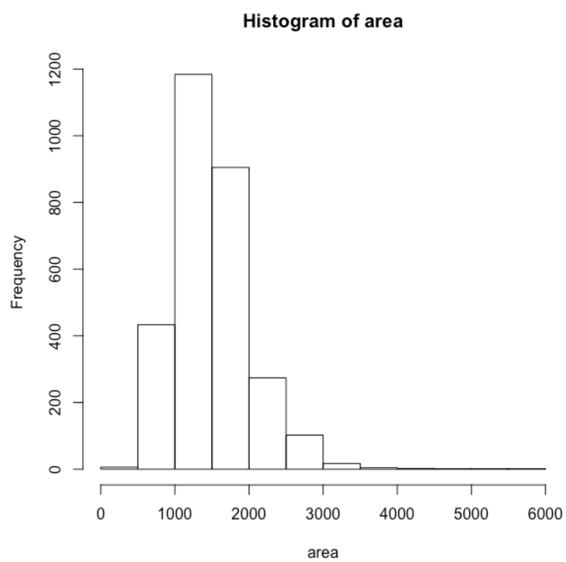
Arsenio Pena
CSC 217
R Assignments
Professor Elahe Vahdani

R Assignment 3

Exercise 1

The distribution is right-skewed. It is not symmetric.



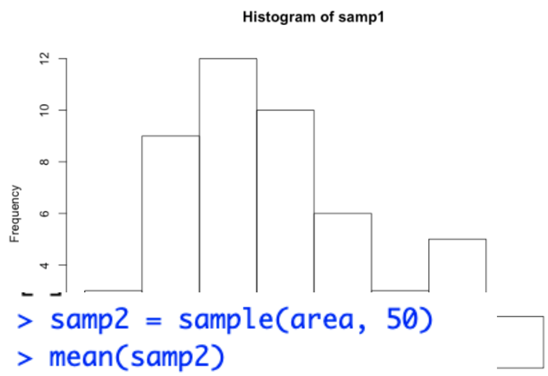
Exercise 2

```
set.seed(13579)
samp1 = sample(area,50)
```

Exercise 3

The distribution for the sample is close to normal/symmetrical. It does seem slightly right-skewed. The distribution for the sample is more normal than the population's distribution.

```
> mean(samp1)
[1] 1473.38
```

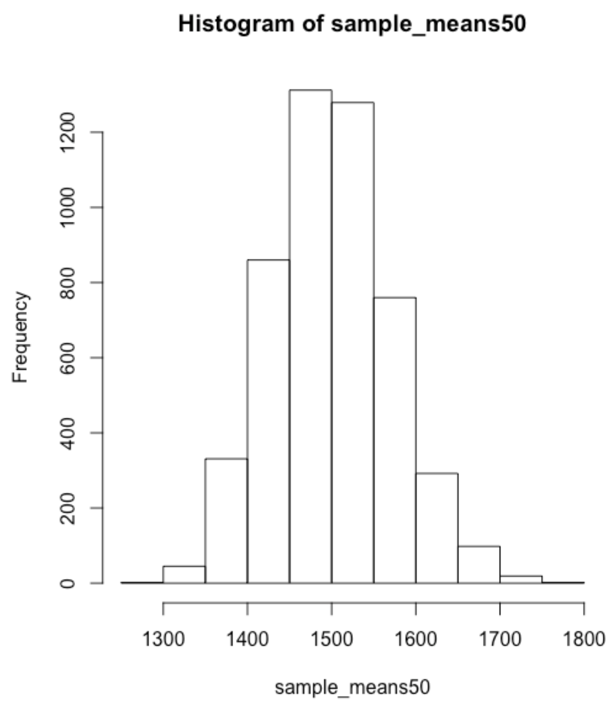


Exercise 4

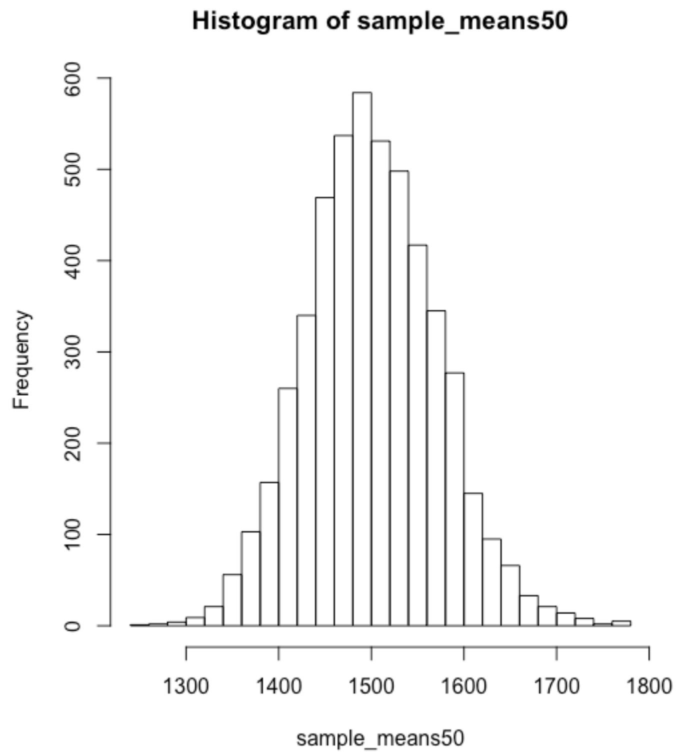
```
> samp2 = sample(area, 50)
> mean(samp2)
[1] 1538.9
```

```
> samp3 = sample(area, 100)
> mean(samp3)
[1] 1603.81
> samp4 = sample(area, 1000)
> mean(samp4)
[1] 1495.898
```

The mean for sample 2 is not close to that of sample 1. Running the test again with 100 and 1000 samples, the mean changed dramatically for both samples when compared to the mean of sample 1. The higher the number for sampling gets us closer to the mean.




```
> sample_means50=rep(0,5000)
> for(i in 1:5000){
+   samp=sample(area,50)
+   sample_means50[i]=mean(samp)
+ }
> hist(sample_means50)
> sample_means50=rep(0,5000)
> for(i in 1:5000){
+   samp=sample(area,50)
+   sample_means50[i]=mean(samp)
+ }
> hist(sample_means50, breaks=25)
```




```
> mean(sample_means50)
[1] 1501.205
```

Exercise 5

There are 5,000 elements in the sample. The distribution is normal. The center is at the peak of the histogram around 1500 which where the mean falls. The mean is 1501.205. The distribution would be different if we had 50,000 samples, since the data would be the same.

Exercise 6

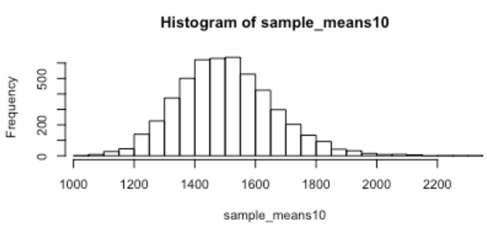
```
sample_mean_small=rep(0,100)
for(i in 1:100){
  samp=sample(area,50)
  sample_mean_small[i]=mean(samp)
}
```

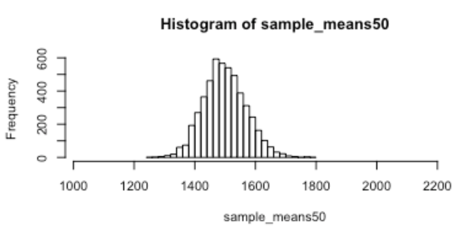
There are 100 elements in the vector. Each element represents the mean of a sample of 50 randomly chosen homes.

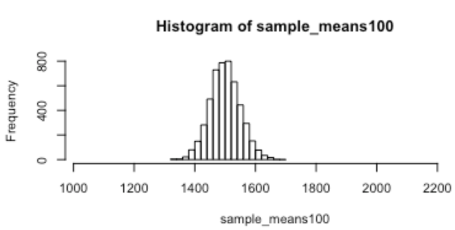
```
  [1] 1395.96 1461.36 1640.62 1428.08 1354.60
  [6] 1508.74 1432.52 1490.98 1479.36 1503.46
 [11] 1438.92 1523.54 1576.02 1392.84 1472.76
 [16] 1589.40 1573.38 1505.00 1452.44 1476.56
 [21] 1526.04 1547.58 1647.96 1678.72 1537.76
 [26] 1500.42 1453.58 1520.86 1378.04 1521.38
 [31] 1440.24 1384.88 1410.96 1541.48 1376.68
 [36] 1581.04 1769.86 1560.50 1497.54 1581.16
 [41] 1458.22 1504.54 1536.50 1456.88 1490.82
 [46] 1461.46 1486.40 1439.06 1487.92 1540.46
 [51] 1417.22 1553.08 1550.30 1561.14 1524.94
 [56] 1536.98 1437.42 1452.58 1558.44 1478.68
 [61] 1639.10 1538.24 1519.76 1568.14 1495.56
 [66] 1440.82 1526.66 1448.78 1578.52 1366.18
 [71] 1500.46 1489.52 1411.86 1564.30 1399.28
 [76] 1471.08 1571.96 1546.28 1484.94 1434.70
 [81] 1607.18 1486.72 1534.06 1494.56 1437.46
 [86] 1519.50 1551.90 1503.72 1393.92 1553.70
 [91] 1516.68 1554.40 1505.58 1574.50 1409.72
 [96] 1496.06 1517.10 1381.14 1646.66 1591.60
```

Exercise 7

The center is steeper when the sample size is larger. The spread is also smaller when the sample size is larger.

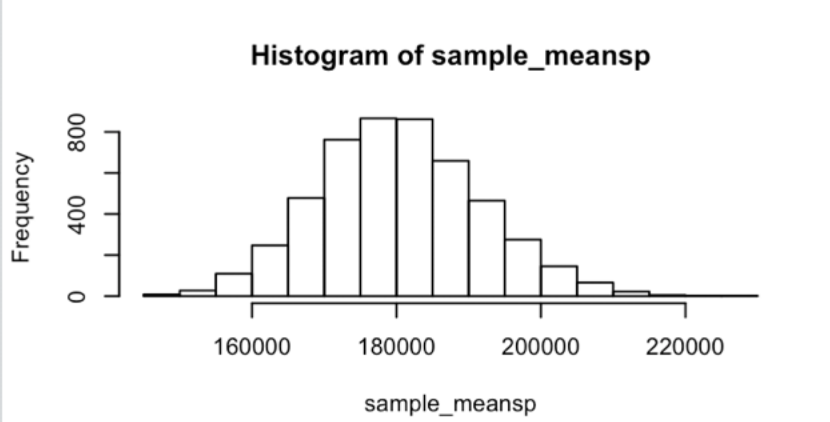






Exercise 8

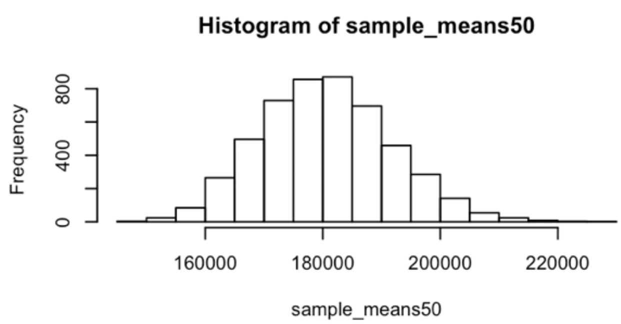
```
> sample_meansp=rep(0,5000)
> for(i in 1:5000){
+   samp=sample(price, 50)
+   sample_meansp[i]=mean(samp)
+ }
> hist(sample_meansp)
> mean(sample_meansp)
[1] 180471.2
```



The best point estimate is around 180,000. Calculating the mean is 180,471.2, and the center of the histogram is around 180,000. The mean would lie in a histogram with data with normal distribution.

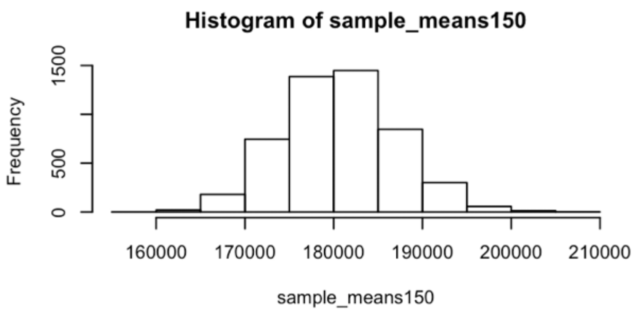
Exercise 9

```
sample_means50=rep(0,5000)
for(i in 1:5000){
  samp=sample(price,50)
  sample_means50[i]=mean(samp)
}
hist(sample_means50)
mean(sample_means50)
> mean(sample_means50)
[1] 180594.7
```



The mean is around 180,000 according to the histogram. The calculated mean for this sample is 180,594.7.

Exercise 10




```
> sample_means150=rep(0,5000)
> for(i in 1:5000){
+   samp=sample(price,150)
+   sample_means150[i]=mean(samp)
+ }
> hist(sample_means150)
> mean(sample_means150)
[1] 180690.6
```

This sampling distribution is symmetric and seems to have normal distribution. The data's distribution for 150 samples is identical to the one of 50 samples. Based on this distribution, the mean is around 180,000.

Exercise 11

The sampling distribution with sample sizes of 50 has a larger spread than the ones of sample size of 150. We want to decrease the spread. Our values are closer to each other and vary less, making them closer to the mean. We can make estimates closer to its true value.