Lesson 02: Charts and Graphs

References

- Black, Chapter 2 Charts and Graphs (pp. 15-34)
- Kabakoff, Chapter 2.5 Useful Functions (pp. 43-44), Chapter 6 Basic Graphs (pp. 117-136)
- Davies, Chapter 14 Basic Data Visualization (pp. 298-308)
- Stowell, Chapter 8 Creating Plots (pp. 99-117)

Data set: home_prices.csv

Description: This data file is derived from a random sample of home resale records maintained by realtors. There are 117 observations and eight variables:

1. PRICE = Selling price ($hundreds)
2. SQFT = Square feet of living space
3. YEAR = Year of construction (year)
4. BATHS = Number of bathrooms
5. FEATS = Number out of 11 features (dishwasher, refrigerator, microwave, disposal, washer, intercom, skylight(s), compactor, dryer, handicap fit, cable TV access)
6. NBR = Located in northeast sector of city (YES) or not (NO)
7. CORNER = Corner location (YES) or not (NO)
8. TAX = Annual taxes ($)

```
# Read the comma-delimited text file creating a data frame object in R
# as we did in Lesson 1, then examine its structure:

houses <- read.csv("home_prices.csv")
str(houses)
```

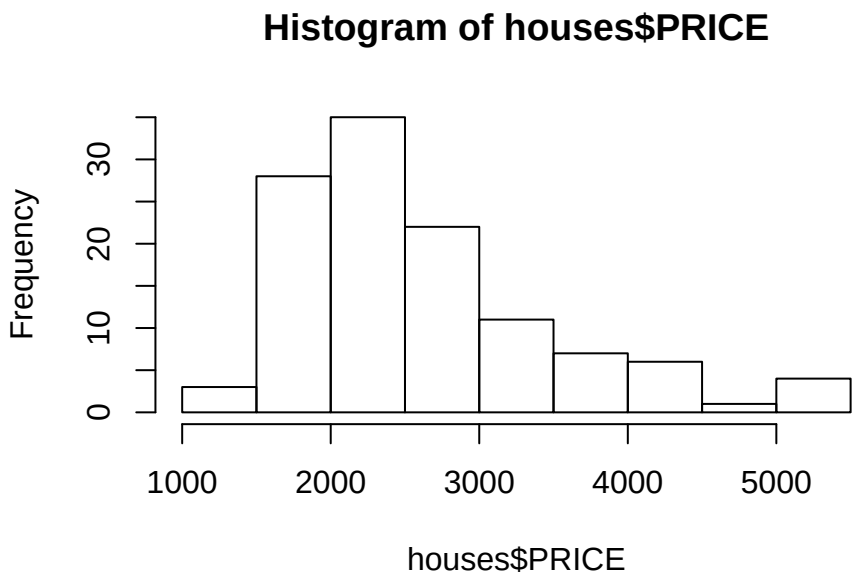
```
## 'data.frame':    117 obs. of  8 variables:
##  $ PRICE : num  1350 2550 1550 1828 1800 ...
##  $ SQFT  : int  1142 1478 1480 1299 1121 1400 1505 1050 900 1215 ...
##  $ YEAR  : int  1959 1961 1965 1967 1968 1969 1969 1970 1971 1971 ...
##  $ BATHS : num  1.5 2 1.5 1 1.5 1.5 1.5 1 1 1.5 ...
##  $ FEATS : int  0 3 4 6 4 1 2 1 3 3 ...
##  $ NBR   : Factor w/ 2 levels "NO","YES": 1 2 1 2 2 1 1 2 1 2 ...
##  $ CORNER: Factor w/ 2 levels "NO","YES": 1 2 1 1 1 2 2 1 1 1 ...
##  $ TAX   : num  558 1565 1275 1462 995 ...
```

Exercises:

1) For the following exercises use hist(), plot(), boxplot() and par() functions supplied by R.

   a) Construct a histogram for PRICE. Describe the distribution shape.

```
hist(houses$PRICE)  # looks positively skewed, mean > median
```




```
# or use with() function
# with(houses, hist(PRICE))
```

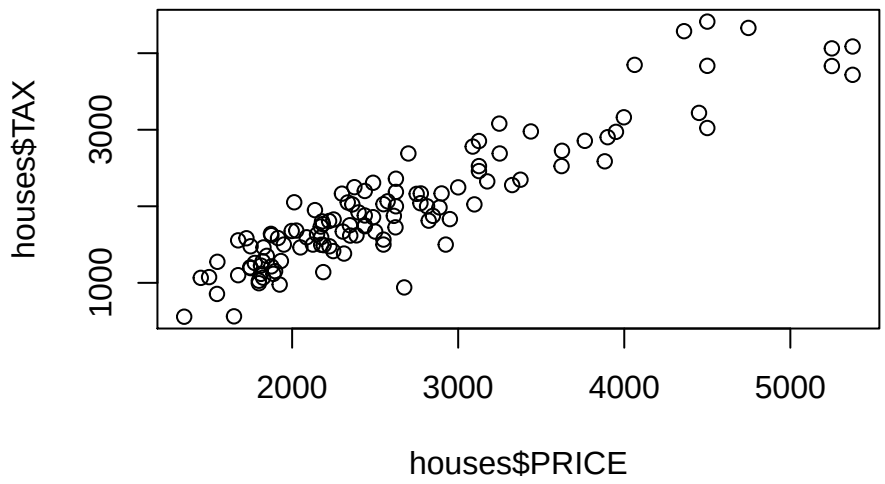
b) Construct a histogram for TAX. Describe the distribution shape.

```
hist(houses$TAX)  # also looks positively skewed, mean > median
```


```
# or use with() function
# with(houses, hist(TAX))
```

c) Construct a scatterplot displaying TAX versus PRICE. Is there a relationship?

```r
plot(houses$PRICE, houses$TAX)
```




```r
# or use with() function
# with(houses, plot(PRICE, TAX))
```

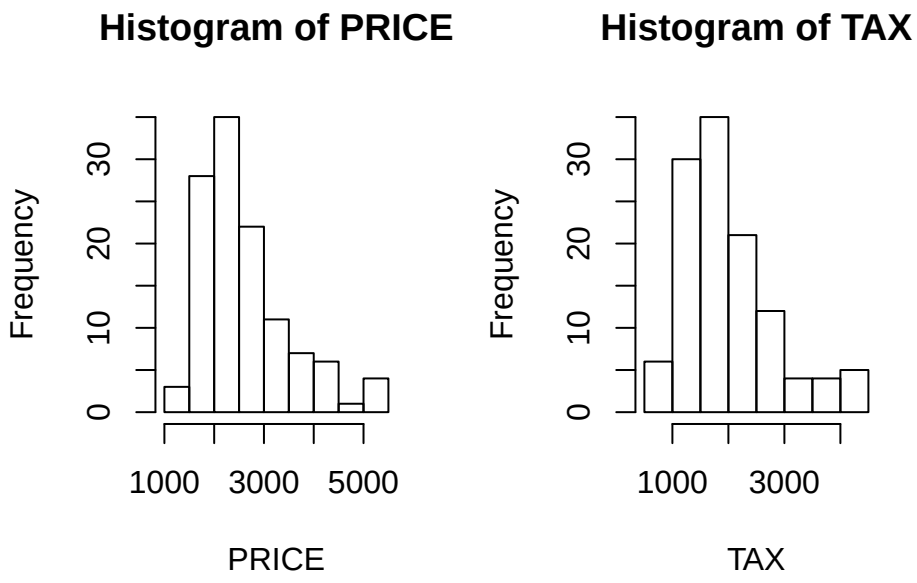
d) Construct a stem-and-leaf plot for TAX using stem().

```r
stem(houses$TAX)
```

```
##
##   The decimal point is 2 digit(s) to the right of the |
##
##    4 | 66
##    6 |
##    8 | 648
##   10 | 03778022459
##   12 | 023688958
##   14 | 266889000006789
##   16 | 00222347788344557
##   18 | 011336888259
##   20 | 00333455767779
##   22 | 05581356
##   24 | 6339
##   26 | 9938
##   28 | 56078
##   30 | 286
##   32 | 2
##   34 |
##   36 | 2
##   38 | 345
##   40 | 69
##   42 | 93
##   44 | 1
```

e) Use the par() and mfrow() or mfcol() functions to construct a multi-plot figure of two rows and one column showing the histograms for PRICE and TAX.

```
par(mfrow=c(1,2))
with(houses, hist(PRICE))
with(houses, hist(TAX))
```




```
par(mfrow=c(1,1))
```

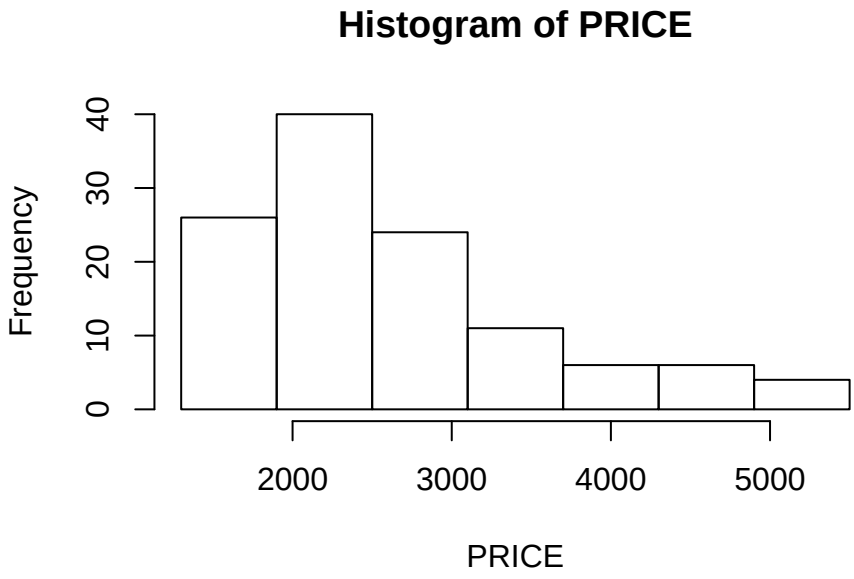
2) For the following exercises use hist() and, within hist(), *breaks*.

a) Construct a histogram for PRICE starting the first class at 1300 with a class width of 600.

```
max(houses$PRICE) # this will let us know how many breaks we need
```

```
## [1] 5375
```

```
with(houses, hist(PRICE, breaks = c(1300, 1900, 2500, 3100, 3700, 4300, 4900,
    5500)))
```



b) Construct a histogram for TAX starting the first class at 500 with a class width of 500.

```
max(houses$TAX)
```

```
## [1] 4412.5
```

```
with(houses, hist(TAX, breaks = c(500, 1000, 1500, 2000, 2500, 3000, 3500,
    4000, 4500)))
```