HW 4

Diego Valdes

January 31, 2019

```
library(moments)
```

```
## Warning: package 'moments' was built under R version 3.5.2
```

```
#rm(list=ls()) # clear work space
#dev.off(dev.list()["RStudioGD"]) # clear plots

# function to give vector info
printVecInfo = function(v_data) {

  print(paste0("Mean: ", mean(v_data)))
  print(paste0("Median: ", median(v_data)))
  print(paste0("min: ", min(v_data)))
  print(paste0("min: ", max(v_data)))
  print(paste0("sd: ", sd(v_data)))
  print(paste0("quantile: ", quantile(v_data, probs = c(0.05,0.95))))
  print(paste0("skewness: ", skewness(v_data)))

}

jar = c(rep('red',50), rep('blue',50)) # jar of 50 red/50 blue marbles
table(jar) # verify counts
```

```
## jar
## blue  red
##   50   50
```

```
jarSample = sample(jar, 10, replace = TRUE) # sample 10 marbles
a = table(jarSample) # assign results
a[2] # number red
```

```
## red
##   5
```

```
a[2]/sum(a) # percentage red
```

```
## red
## 0.5
```

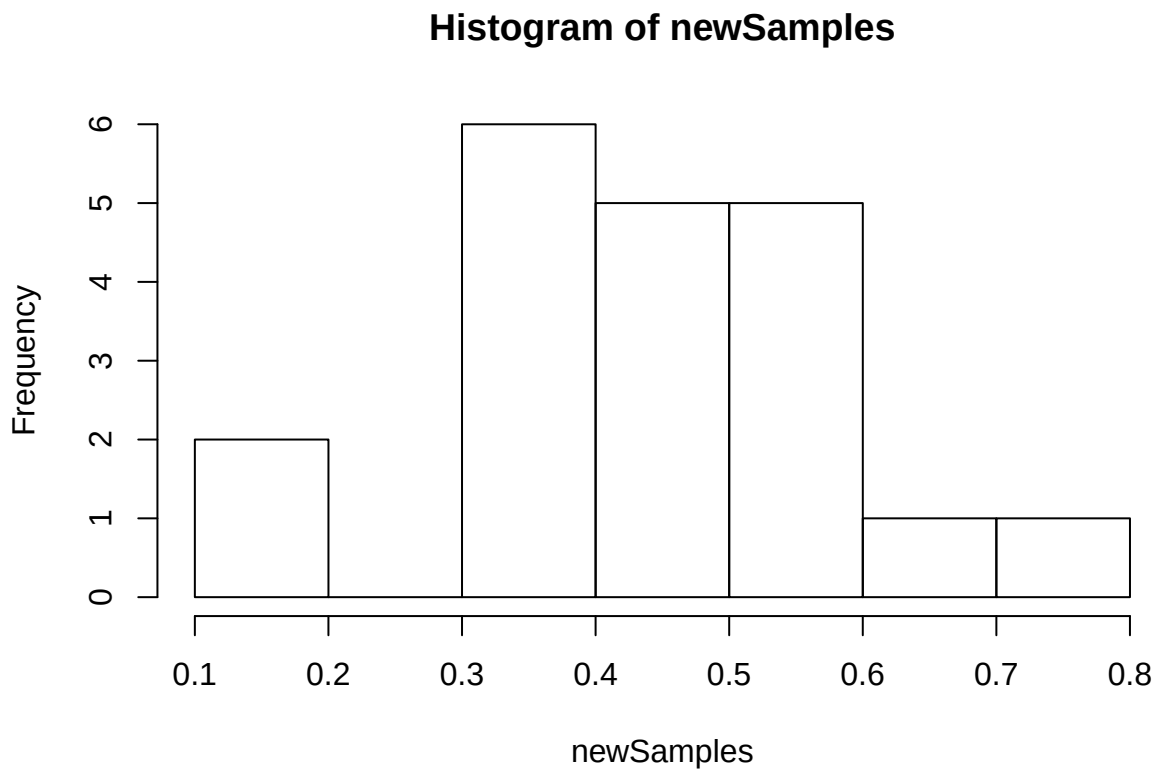
```
# test stuff
#mean(sample(jar, 10, replace = TRUE))
#replicate(20, sample(jar, 10, replace = TRUE))
#mean(table(sample(jar, 10, replace = TRUE))[2]/10)

# replicate sample 20x of 10 samples
newSamples = replicate(20, mean(table(sample(jar, 10, replace = TRUE))[2]/10))
printVecInfo(newSamples)
```

```
## [1] "Mean: 0.485"
## [1] "Median: 0.5"
```

```
## [1] "min: 0.1"
## [1] "min: 0.8"
## [1] "sd: 0.159851905146443"
## [1] "quantile: 0.195" "quantile: 0.705"
## [1] "skewness: -0.463823370327956"
```

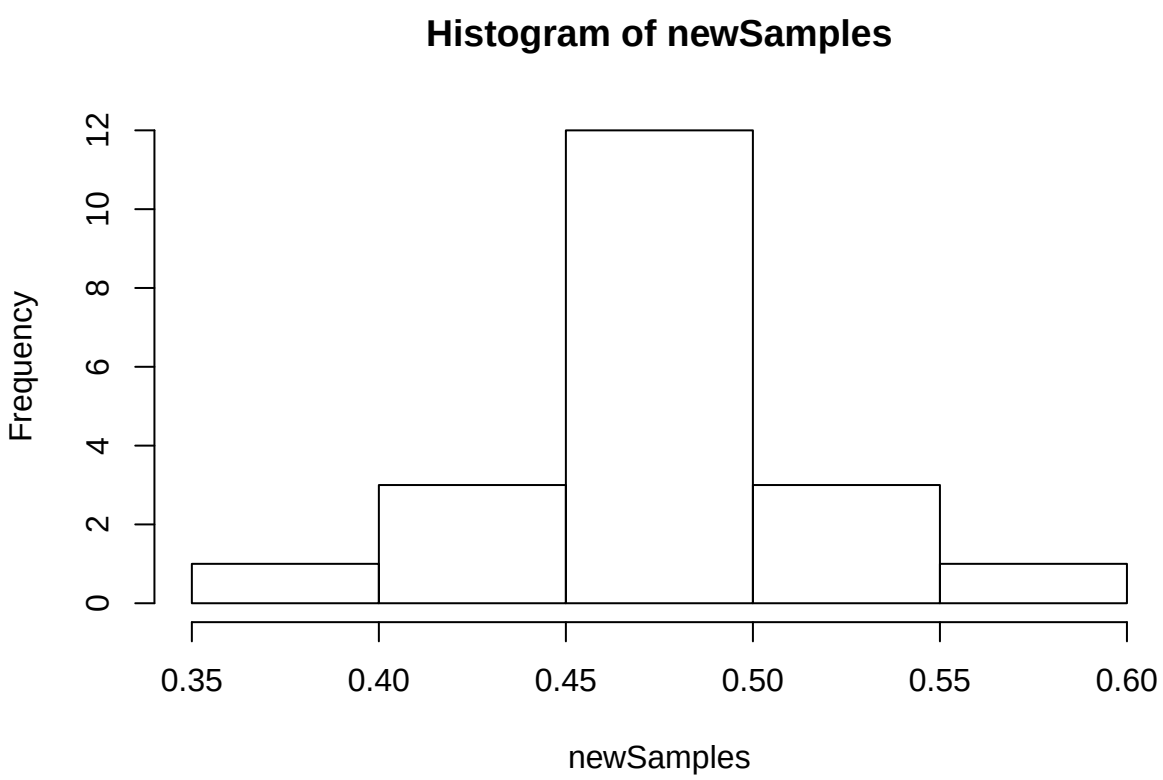
```
hist(newSamples)
```




```
# replicate sample 20x of 100 samples
newSamples = replicate(20, mean(table(sample(jar, 100, replace = TRUE))[2]/100))
printVecInfo(newSamples)
```

```
## [1] "Mean: 0.478"
## [1] "Median: 0.48"
## [1] "min: 0.38"
## [1] "min: 0.56"
## [1] "sd: 0.042624430849727"
## [1] "quantile: 0.4275" "quantile: 0.5505"
## [1] "skewness: -0.121829212875411"
```

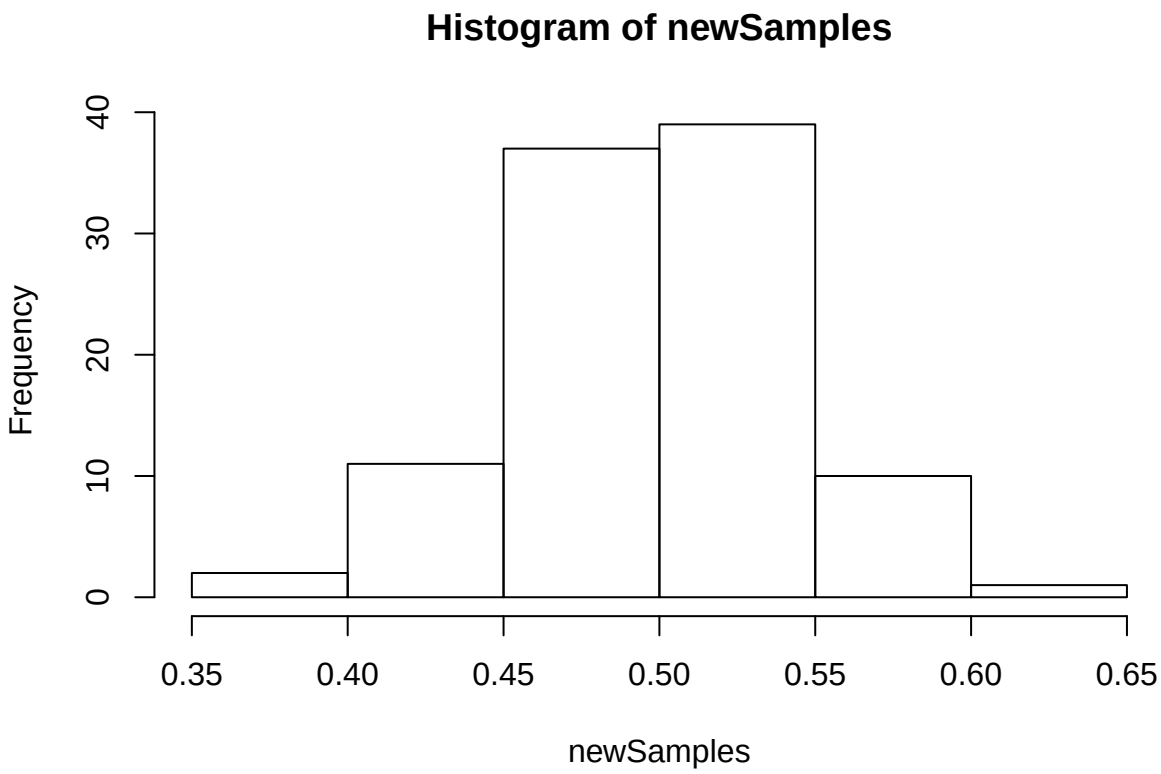
```
hist(newSamples)
```




```r
# replicate sample 100x of 10 samples
newSamples = replicate(100, mean(table(sample(jar, 100, replace = TRUE))[2]/100))
printVecInfo(newSamples)
```

```
## [1] "Mean: 0.5031"
## [1] "Median: 0.505"
## [1] "min: 0.38"
## [1] "min: 0.61"
## [1] "sd: 0.0446466486124461"
## [1] "quantile: 0.4295" "quantile: 0.5605"
## [1] "skewness: -0.300607241205251"
```

```r
hist(newSamples)
```



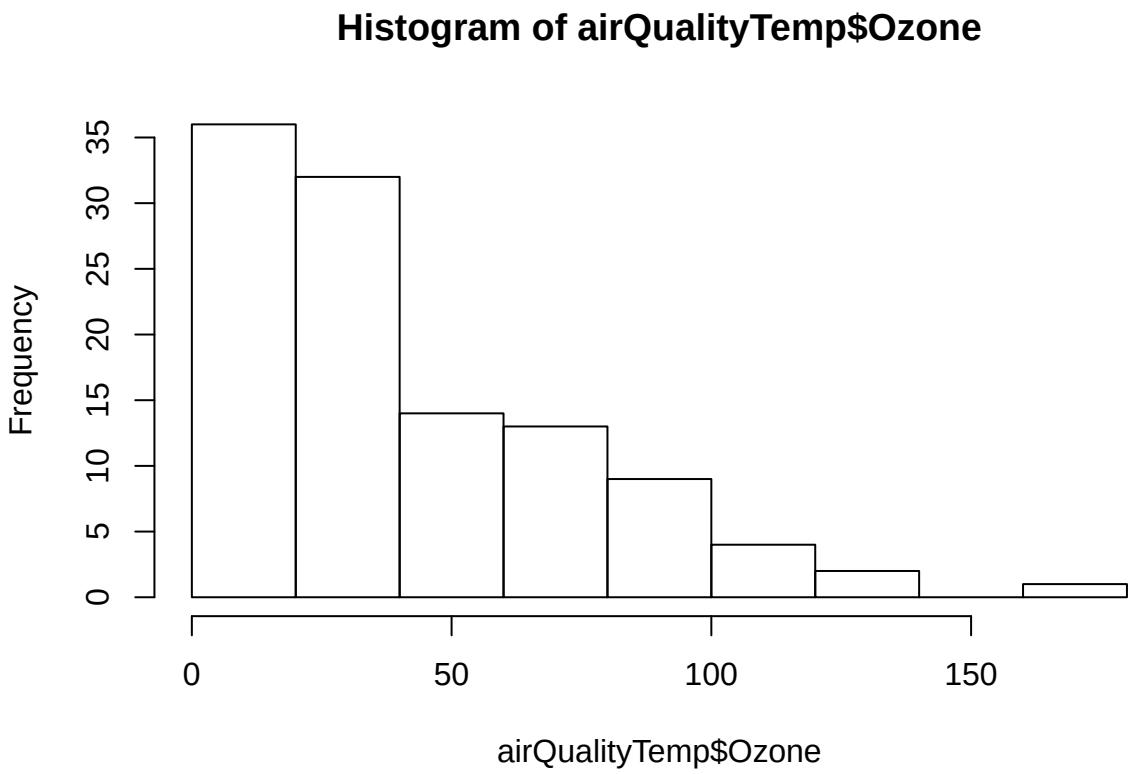

```
# airquality dataset
airQualityTemp = airquality

airQualityTemp = na.omit(airQualityTemp) # remove na
# airQualityTemp = airQualityTemp[complete.cases(airQualityTemp), ]

# explore ozone
printVecInfo(airQualityTemp$Ozone)
```

```
## [1] "Mean: 42.0990990990991"
## [1] "Median: 31"
## [1] "min: 1"
## [1] "min: 168"
## [1] "sd: 33.2759686574274"
## [1] "quantile: 8.5" "quantile: 109"
## [1] "skewness: 1.24810370040404"
```

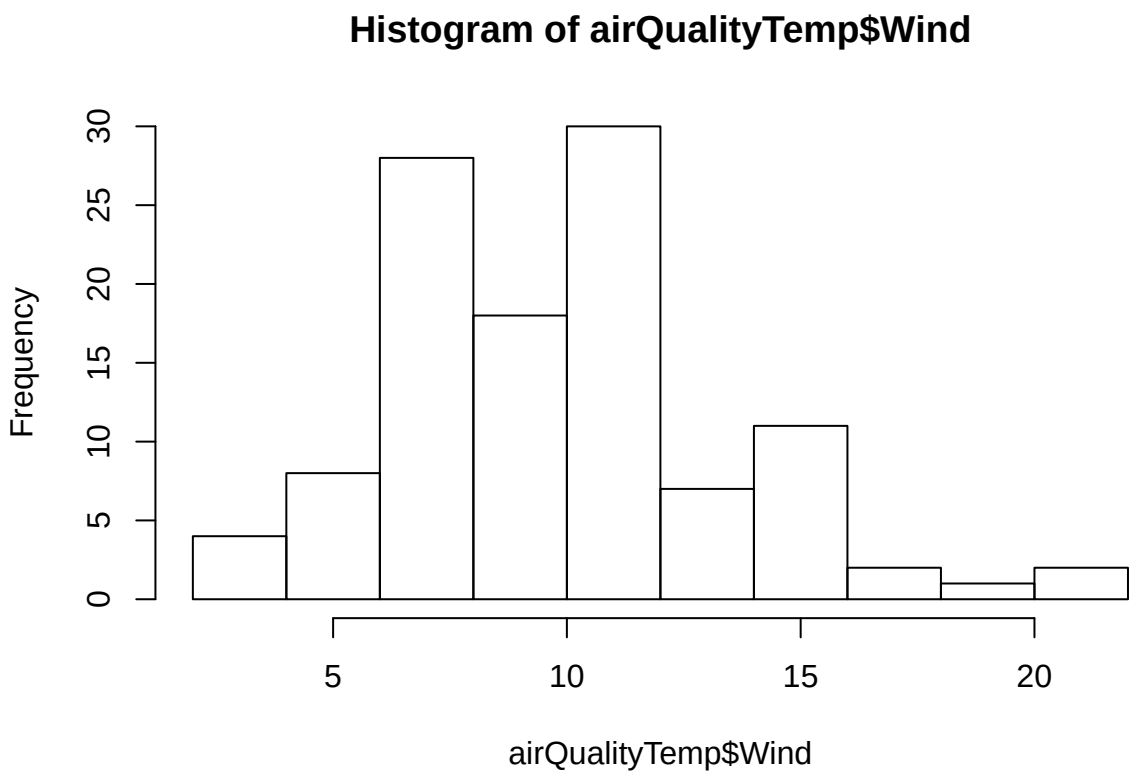
```
hist(airQualityTemp$Ozone)
```




```
# explore wind
printVecInfo(airQualityTemp$Wind)
```

```
## [1] "Mean: 9.93963963963964"
## [1] "Median: 9.7"
## [1] "min: 2.3"
## [1] "min: 20.7"
## [1] "sd: 3.55771324101922"
## [1] "quantile: 4.6"  "quantile: 15.5"
## [1] "skewness: 0.455641432036776"
```

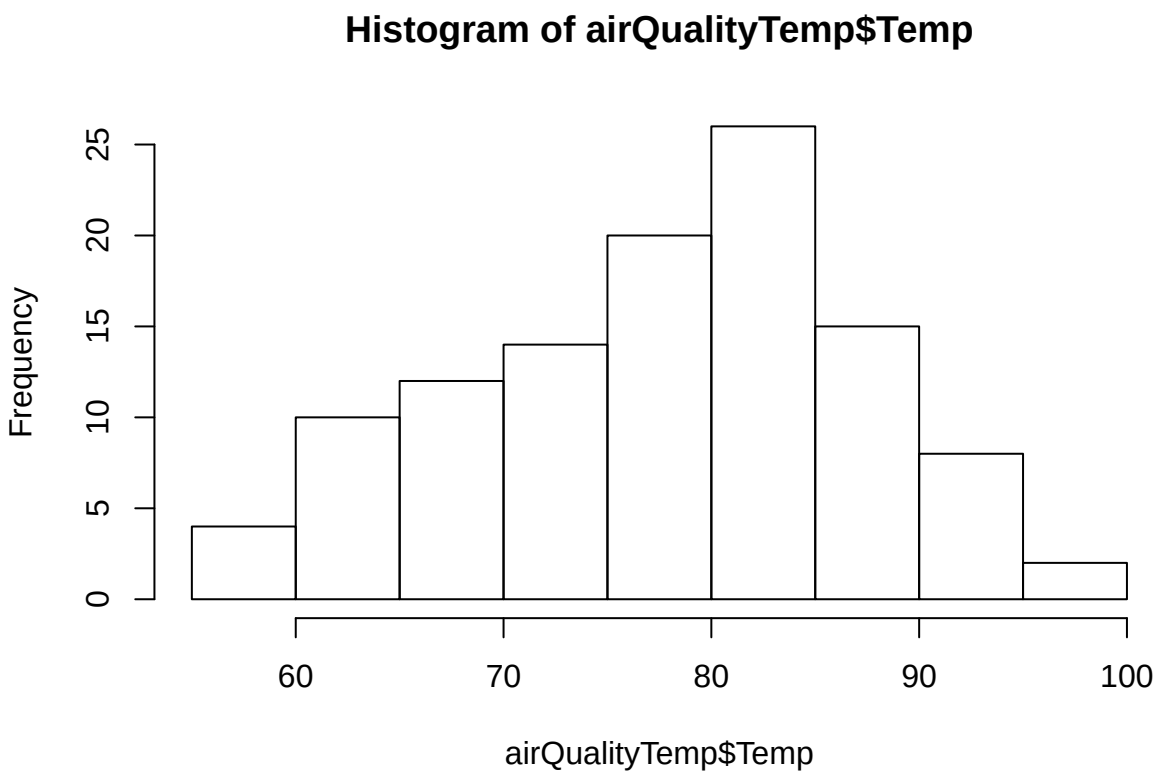
```
hist(airQualityTemp$Wind)
```




```
# explore temp
printVecInfo(airQualityTemp$Temp)
```

```
## [1] "Mean: 77.7927927927928"
## [1] "Median: 79"
## [1] "min: 57"
## [1] "min: 97"
## [1] "sd: 9.52996910909533"
## [1] "quantile: 61"   "quantile: 92.5"
## [1] "skewness: -0.225095889347339"
```

```
hist(airQualityTemp$Temp)
```


Histogram of airQualityTemp$Temp
Frequency
0
5
10
15
20
25
60
70
80
90
100
airQualityTemp$Temp