Basic summaries

Francisco Guzmán

May 12, 2016

In this recipe, you'll learn: - How to look at data summaries (average, standard dev, median) - How to create a basic histogram - How to create a basic annotation line

1 Loading data

```
# Let's load some pre-defined R data
data(mtcars)
help (mtcars)
# Let's peek into the data
head(mtcars)
```

```
>>>                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
>>> Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
>>> Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
>>> Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
>>> Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
>>> Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
>>> Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
# Let's see its structure
str(mtcars)
```

```
>>> 'data.frame':   32 obs. of  11 variables:
>>>  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
>>>  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
>>>  $ disp: num  160 160 108 258 360 ...
>>>  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
>>>  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
>>>  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
>>>  $ qsec: num  16.5 17 18.6 19.4 17 ...
>>>  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
>>>  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
>>>  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
>>>  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

2 Let's calculate summaries of variables!

```
# Let's calculate the average MPG consumption
mean(mtcars$mpg)
```

```
>>> [1] 20.09062
```

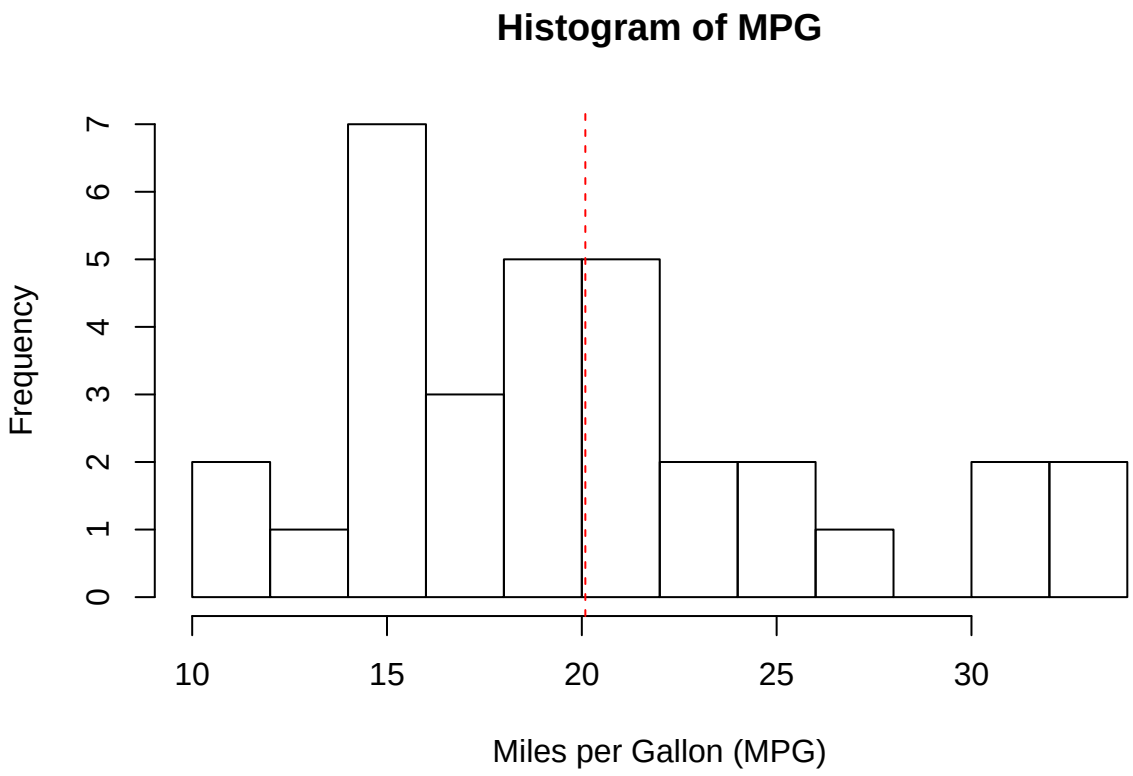
```r
# Let's calculate the standard deviation
sd(mtcars$mpg)
```

```
>>> [1] 6.026948
```

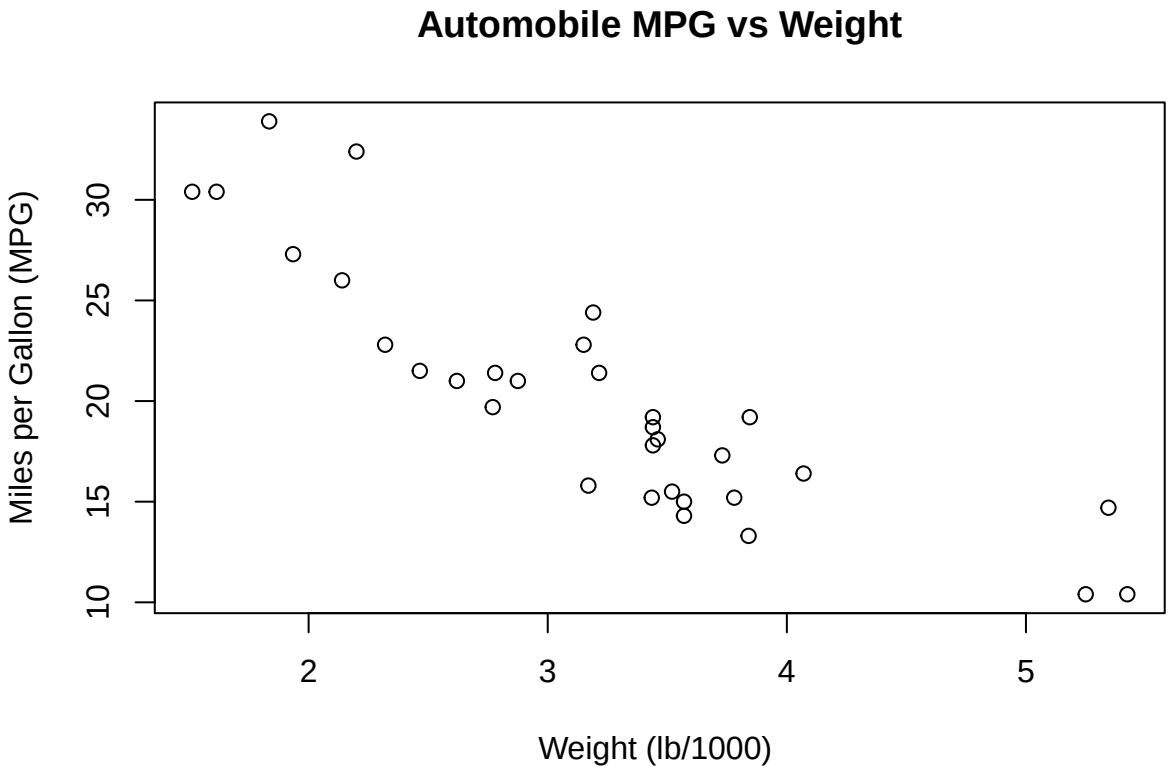
```r
# Shortcut, calculate  summary statistics
summary(mtcars$mpg)
```

```
>>>    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
>>>   10.40   15.42   19.20   20.09   22.80   33.90
```

```r
# let's generate an histogram of the mpg variable with 10 bars
hist(mtcars$mpg,breaks=10,main="Histogram of MPG", xlab="Miles per Gallon (MPG)")
# Let's draw a vertical line at the mean
abline(v=mean(mtcars$mpg),col="red",lty=2)
```




```r
# Calculate some bi-variate stats
#let's plot MPG vs weight
plot(mtcars$wt,mtcars$mpg, main="Automobile MPG vs Weight", ylab="Miles per Gallon (MPG)",xlab="Weight
```




```
#There is a clear (inverse) relationship between both
#Let's calculate correlation
cor(mtcars$mpg,mtcars$wt)
```

```
>>> [1] -0.8676594
```

3 Exercise:

1- Calculate summaries for `HP` and `wt`

2- Plot histogram of `HP`

3- Plot scatterplot of `HP` vs `MPG` and calculate correlation