Descriptive Statistics

Ryan

September 12, 2015

Descriptive Statistics

You now know vectors...

Lets generate some data in vector form

```
set.seed(123)
ranks <- as.factor(sample(10,40,replace = T ))
ranks
```

```
##  [1] 3  8  5  9  10 1  6  9  6  5  10 5  7  6  2  9  3  1  4  10 9  7  7
## [24] 10 7  8  6  6  3  2  10 10 7  8  1  5  8  3  4  3
## Levels: 1 2 3 4 5 6 7 8 9 10
```

```
set.seed(234)
heights <- rnorm(40, 171, 5)
heights
```

```
##  [1] 174.3038 160.7351 163.5040 178.3562 178.2957 171.7007 172.0459
##  [8] 155.8196 168.5653 165.5607 171.2893 176.5199 170.8719 173.5742
## [15] 175.9503 172.5173 166.3496 171.4202 173.6339 171.0793 172.0268
## [22] 176.0814 173.0450 167.4738 172.1781 172.7257 165.9010 163.9409
## [29] 164.1808 175.3655 169.3166 167.7037 170.4962 174.5195 173.1903
## [36] 178.4857 172.0812 163.5233 171.6250 165.8095
```

Descriptive Statistics for Categorical data

Frequency Distribution Table

```
ft <- table(ranks)
ft
```

```
## ranks
##  1  2  3  4  5  6  7  8  9 10
##  3  2  5  2  4  5  5  4  4  6
```

Relative Frequency Distribution

```
ft/length(ranks)
```

```
## ranks
##     1     2     3     4     5     6     7     8     9    10
## 0.075 0.050 0.125 0.050 0.100 0.125 0.125 0.100 0.100 0.150
```

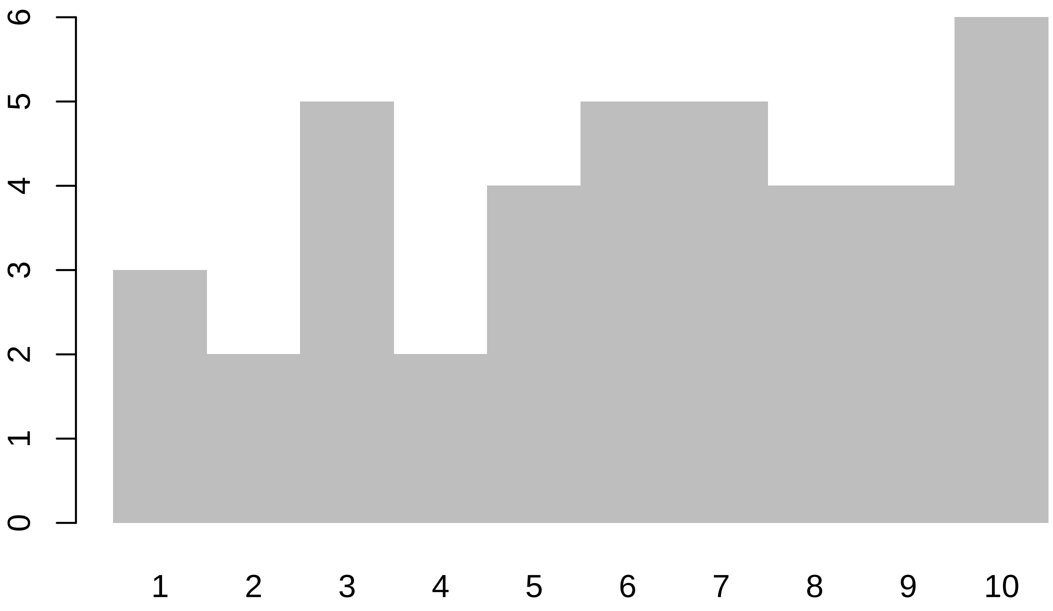
Percentage Relative Frequency Distribution

```
ft/length(ranks)*100
```

```
## ranks
##    1    2    3    4    5    6    7    8    9   10
##  7.5  5.0 12.5  5.0 10.0 12.5 12.5 10.0 10.0 15.0
```

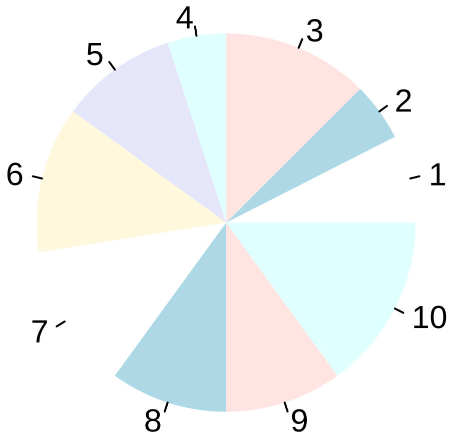
Bar Chart

```
barplot(ft, space = 0, border = 0)
```



Pie Chart

```r
pie(table(ranks), border = 0)
```



Descriptive Statistics for Numerical data

Frequency Distribution Table

```r
b <- seq(154, 181, 3)
heights.cut <- cut(heights, breaks = b, right = F)
cft <- table(heights.cut)
cft
```

```
## heights.cut
## [154,157) [157,160) [160,163) [163,166) [166,169) [169,172) [172,175)
##         1         0         1         7         4         8        12
## [175,178) [178,181)
##         4         3
```

Relative Frequency Distribution

```r
cft/length(heights)
```

```
## heights.cut
## [154,157) [157,160) [160,163) [163,166) [166,169) [169,172) [172,175) 
##     0.025     0.000     0.025     0.175     0.100     0.200     0.300 
## [175,178) [178,181) 
##     0.100     0.075
```

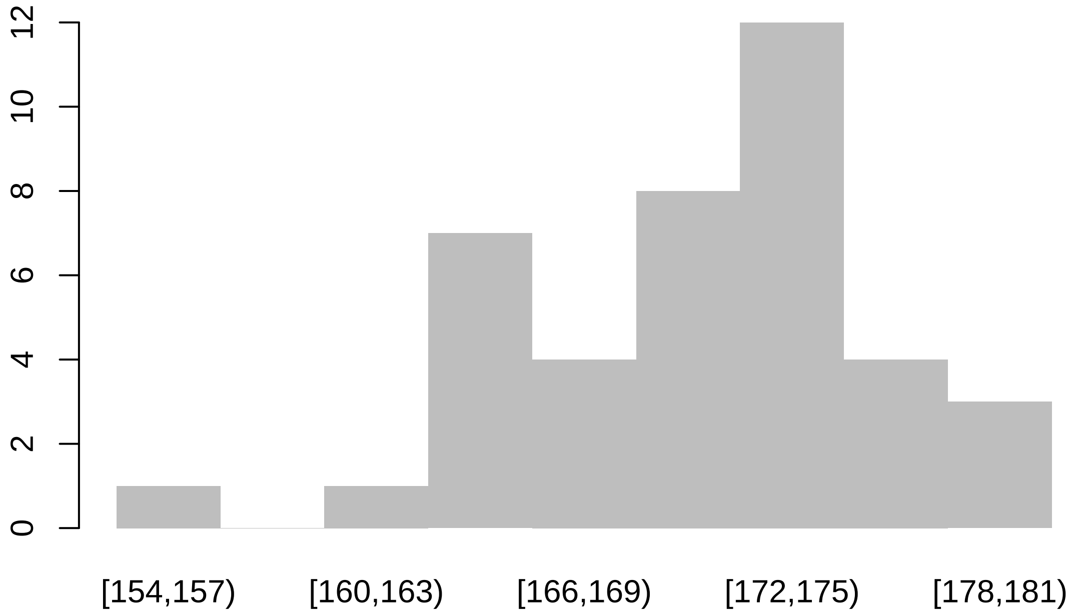
Percentage Relative Frequency Distribution

```r
cft/length(heights) * 100
```

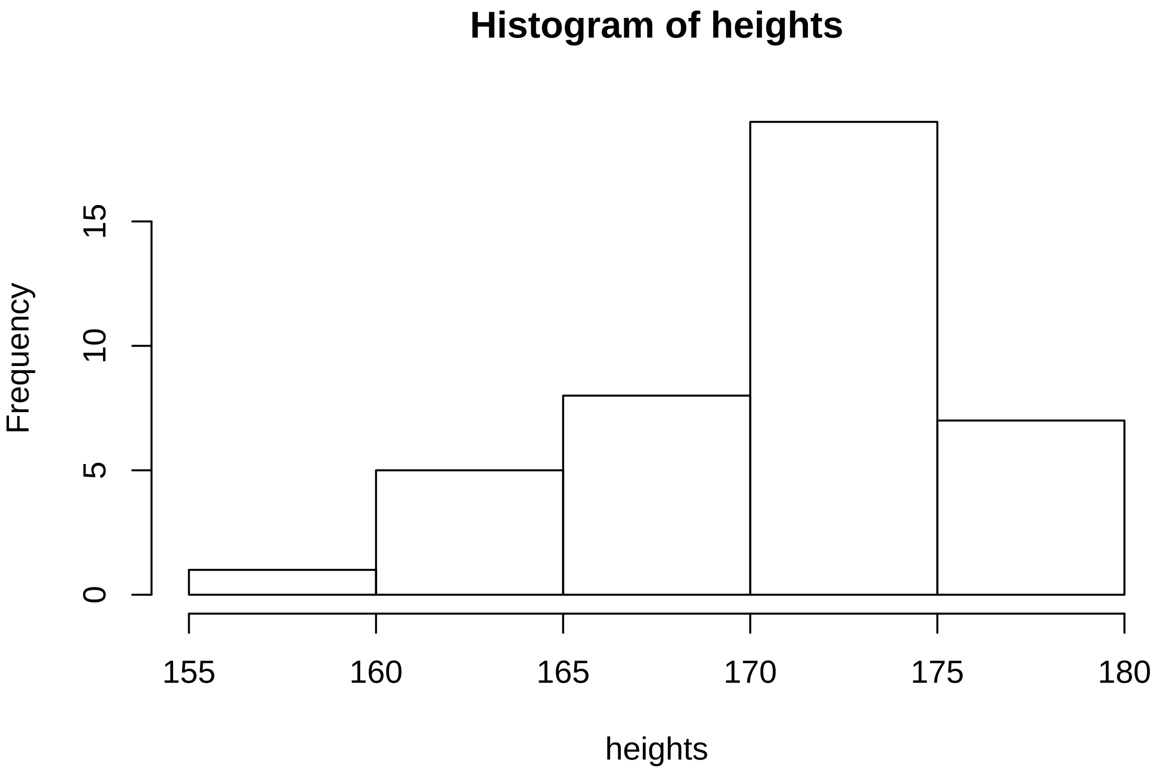
```
## heights.cut
## [154,157) [157,160) [160,163) [163,166) [166,169) [169,172) [172,175) 
##       2.5       0.0       2.5      17.5      10.0      20.0      30.0 
## [175,178) [178,181) 
##      10.0       7.5
```

Histogram

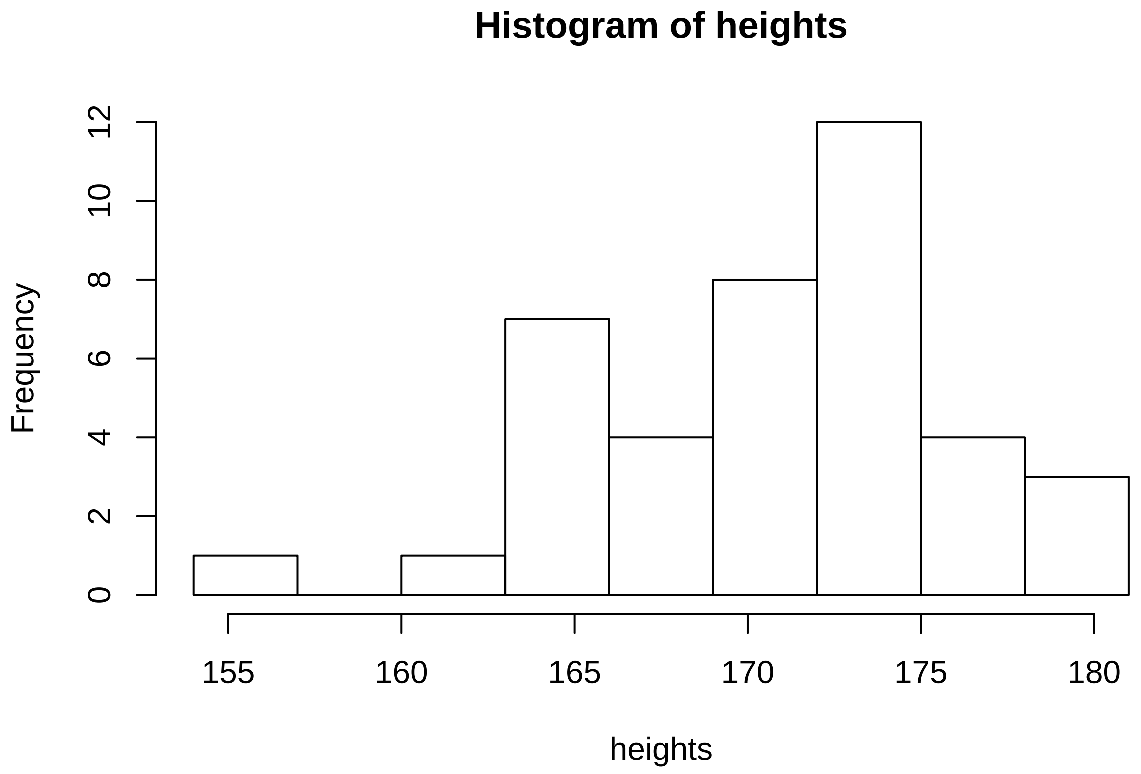
```r
barplot(cft, border = 0, space = 0)
```



```
hist(heights)
```



```r
hist(heights, breaks = b, right = F)
```



Measure of Centrality: Mean and Median

```r
mean(heights)
```

```
## [1] 170.5441
```

```r
median(heights)
```

```
## [1] 171.6628
```

Quantiles

```r
quantile(heights, probs = 0.25, type = 6)
```

```
##      25%
## 166.6307
```

```r
quantile(heights, probs = 0.5, type = 6)
```

```
##      50%
## 171.6628
```

```r
quantile(heights, probs = 0.75, type = 6)
```

```
##     75%
## 173.619
```

```r
quantile(heights, probs = 0.95, type = 6)
```

```
##      95%
## 178.3531
```

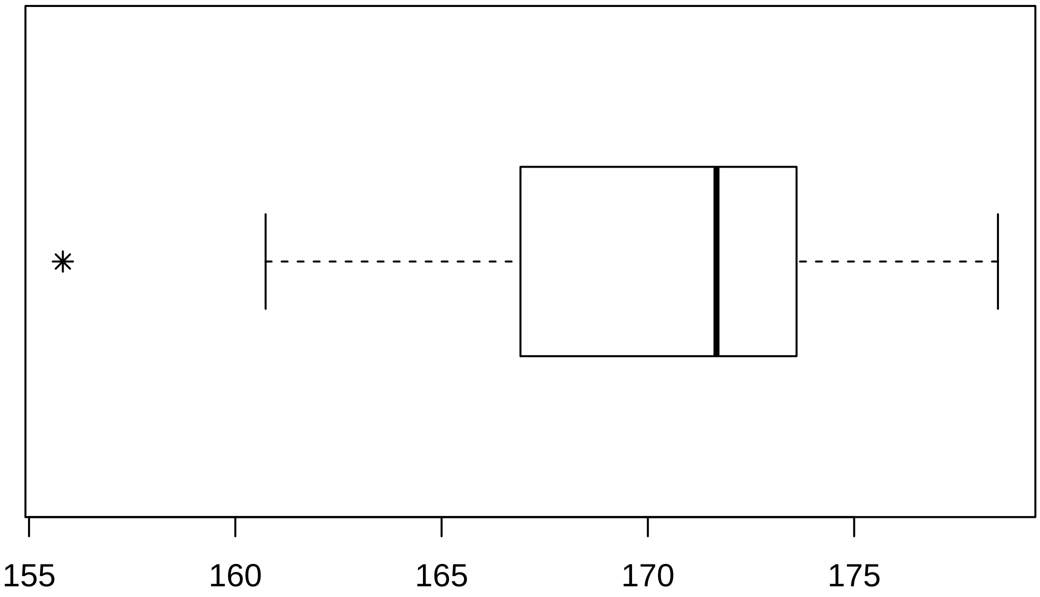
Five Number Summmary

```r
summary(heights)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   155.8   167.2   171.7   170.5   173.6   178.5
```

Boxplot

```r
boxplot(heights, horizontal = T, pch = 8)
```



Measure of Dispersion : Variance and Standard Deviation

```
var(heights)
```

```
## [1] 25.45171
```

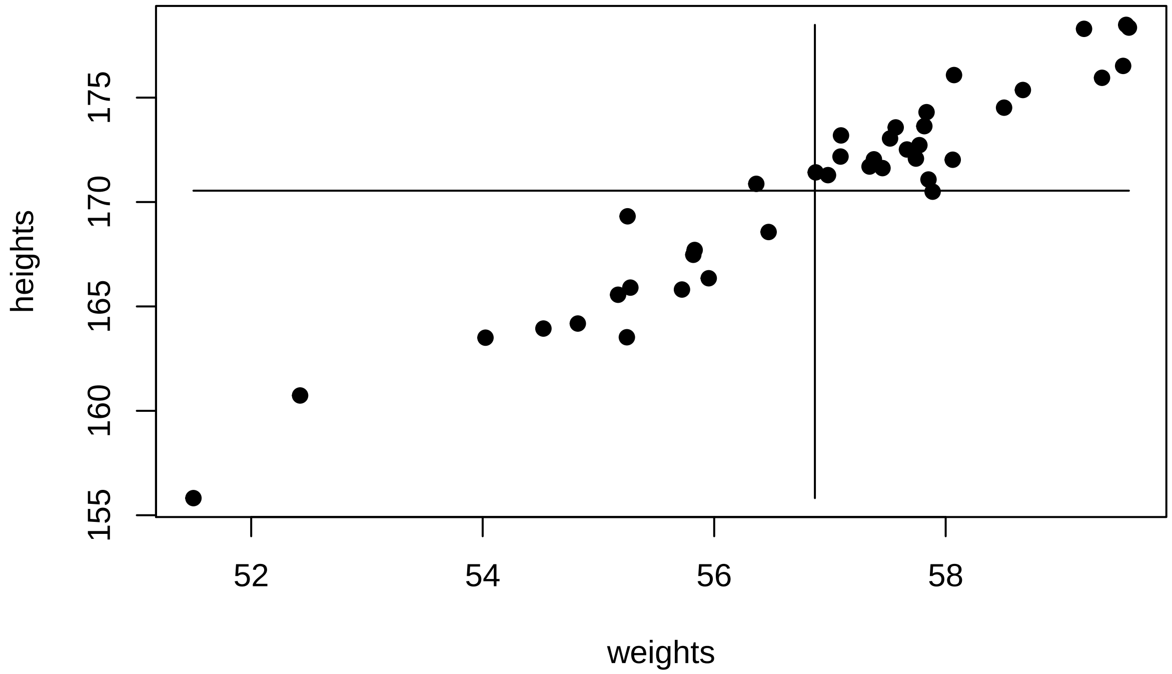
```
sd(heights)
```

```
## [1] 5.044968
```

Descriptive Statistics for Linear Relationship

Scatter plot

```
set.seed(231)
weights <- heights/3+rnorm(40,0,0.5)
plot(weights, heights, pch = 19)
lines(x = rep(mean(weights),2), y = c(min(heights),max(heights)))
lines(y = rep(mean(heights),2), x = c(min(weights),max(weights)))
```



Covariance and Correlation Coefficient

```
cov(heights, weights)
```

```
## [1] 8.941552
```

```
cor(heights, weights)
```

```
## [1] 0.9643189
```