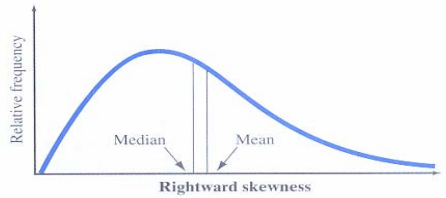
Chapter 3: Statistics for Describing, Exploring, and Comparing Data

Common summary descriptions

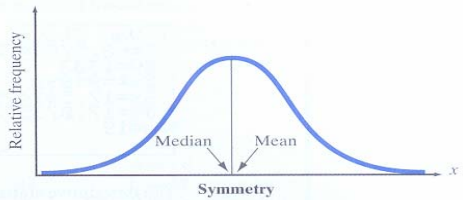
- Center (location)
  - Mean ($\bar{x}$ and μ)
  - Median
  - Mode
  - Midrange
- Distribution (modes and skew)
- Variation (spread)
  - Range
  - Variance (s^2 and σ^2)
  - Standard deviation (s and σ)
    - Interpretation: range rule and empirical rule
- Measures of relative standing
  - Z-scores
  - Quartiles and percentiles
- Outliers
  - 1.5(IQR) criterion
  - Boxplot
- Changes over time

Comparing the Mean and the Median

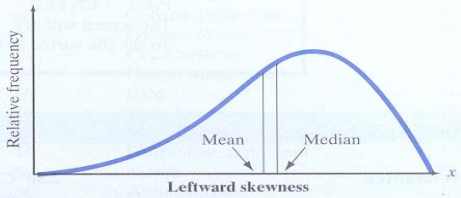
If the data set is skewed to the right, the mean is greater than (to the right of) the median.



If the data set is symmetric, the mean equals the median.



If the data set is skewed to the left, the mean is less than (to the left of) the median.



Example

Sample of 5000 data values drawn from a normal population with $\mu = 350$ and $\sigma = 25$.

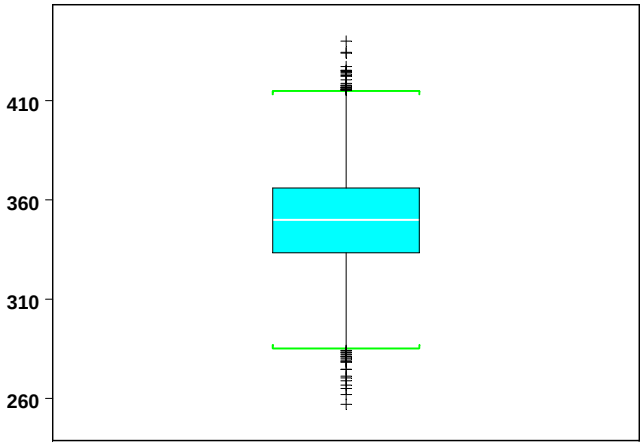
```
R summary output
   Min.:257.0
 1st Qu.:333.3
  Median:350.0
   Mean:349.8
 3rd Qu.:366.0
   Max.:439.9
```

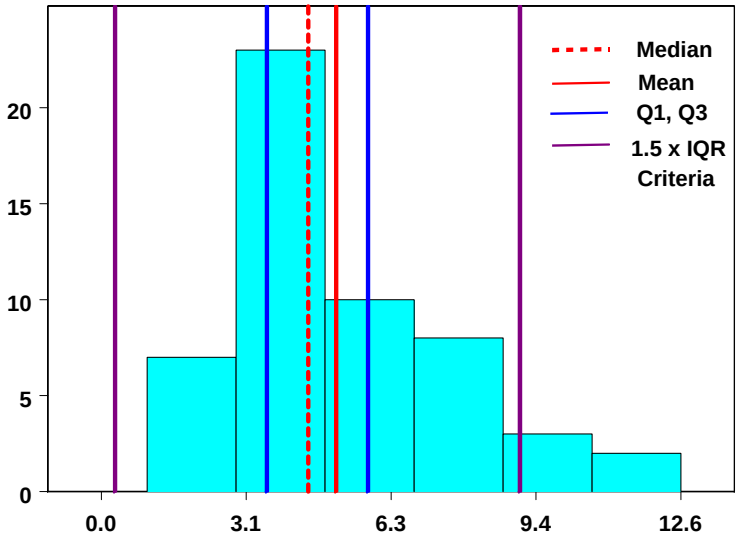
$IQR = 366 - 333.3 = 32.7$

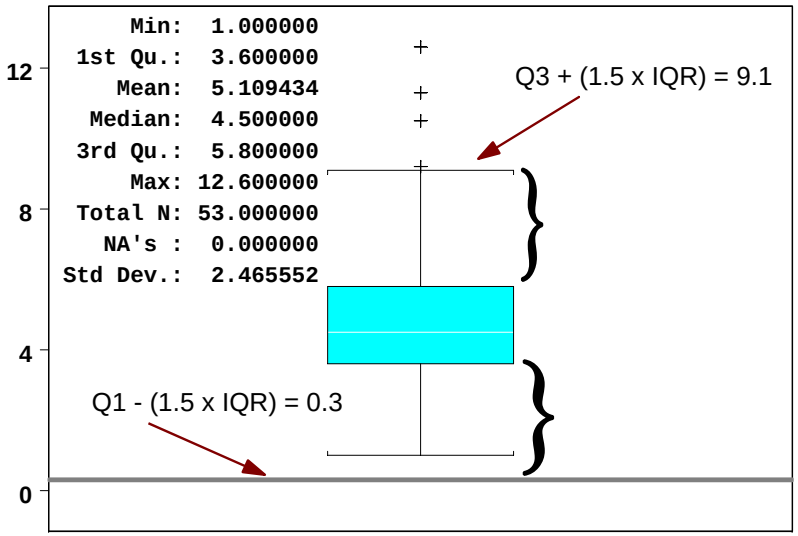
$1.5 \times IQR = 49.05$

$366.0 + 49.05 = 415.05$

$333.3 - 49.05 = 284.25$




Median
Mean
Q1, Q3
1.5 x IQR
Criteria
0
5
10
15
20
0.0
3.1
6.3
9.4
12.6


Min: 1.000000
1st Qu.: 3.600000
Mean: 5.109434
Median: 4.500000
3rd Qu.: 5.800000
Max: 12.600000
Total N: 53.000000
NA's : 0.000000
Std Dev.: 2.465552
Q3 + (1.5 x IQR) = 9.1
Q1 - (1.5 x IQR) = 0.3
0
4
8
12