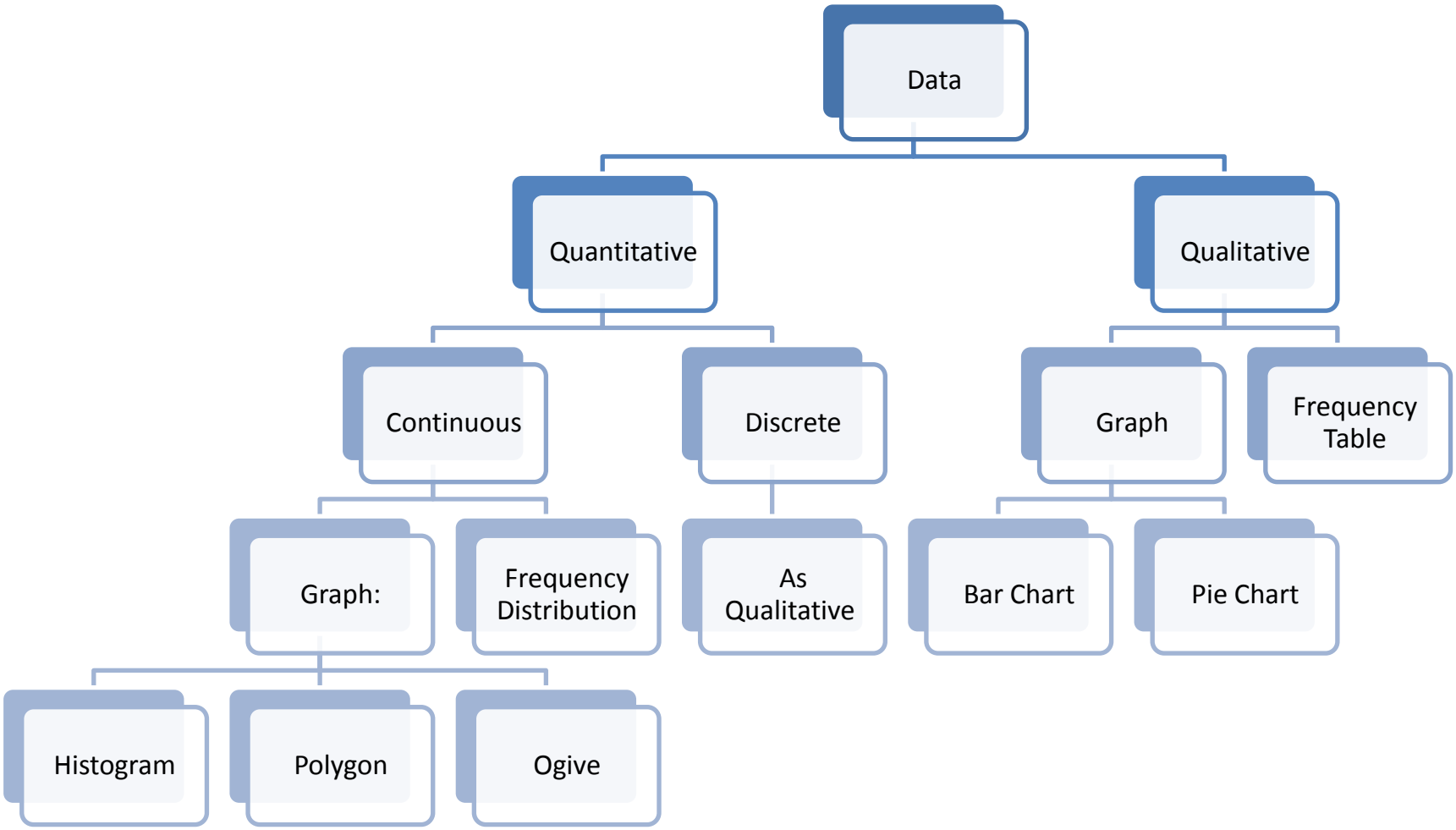
Lecture#4

STAT 110 Revision

Describing Data



Frequency Table

FREQUENCY TABLE A grouping of qualitative data into mutually exclusive classes showing the number of observations in each class.

TABLE 2–1 Frequency Table for Vehicles Sold at Whitner Autoplex Last Month

Car Type	Number of Cars
Domestic	50
Foreign	30

Bar Charts

BAR CHART A graph in which the classes are reported on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are proportional to the heights of the bars.

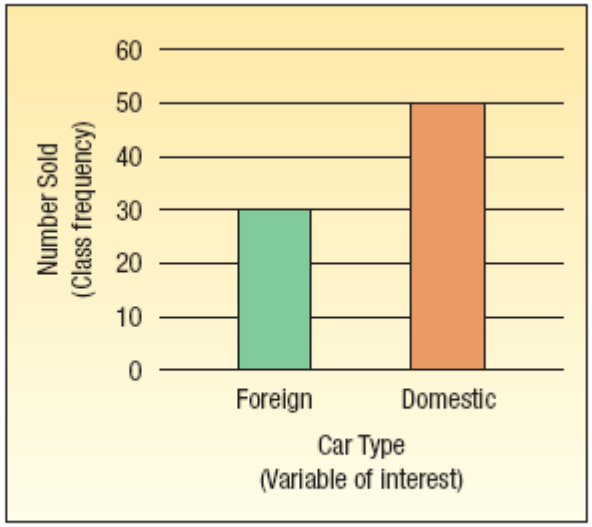
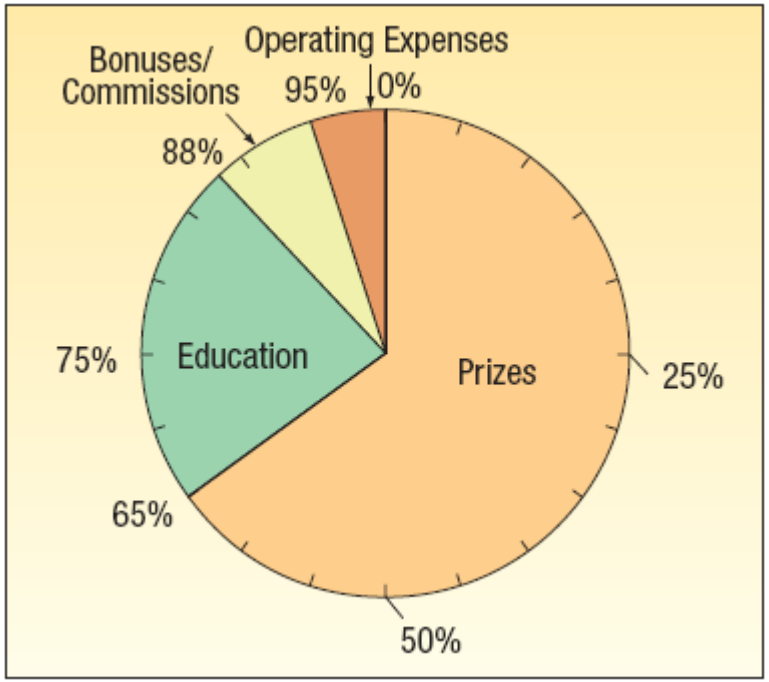


CHART 2–1 Vehicle Sold by Type Last Month At Whitner Autoplex

Pie Charts

PIE CHART **A chart that shows the proportion or percent that each class** represents of the total number of frequencies.

Use of Sales	Amount ($ million)	Percent of Share
Prizes	1,311.1	65
Education	464.3	23
Bonuses	139.8	7
Expenses	109.8	5
Total	2,025.0	100



Frequency Distribution

FREQUENCY DISTRIBUTION A grouping of data into mutually exclusive classes showing the number of observations in each class.

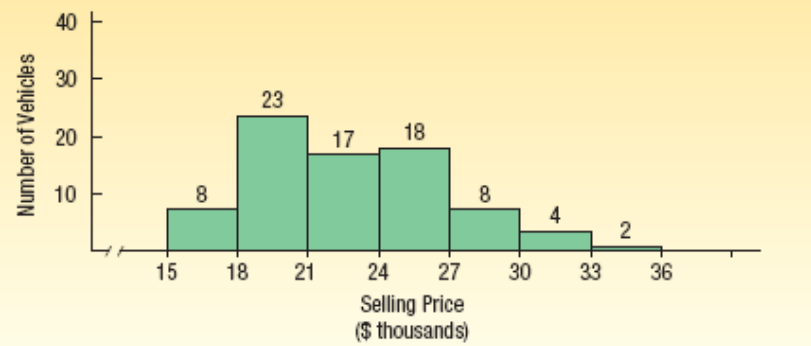
Selling Prices ($ thousands)	Frequency
15 up to 18	8
18 up to 21	23
21 up to 24	17
24 up to 27	18
27 up to 30	8
30 up to 33	4
33 up to 36	2
Total	80

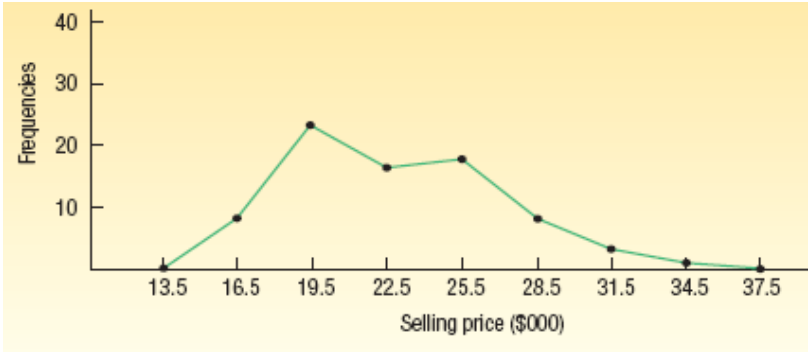
Class midpoint=(Upper limit + Lower limit)/ 2

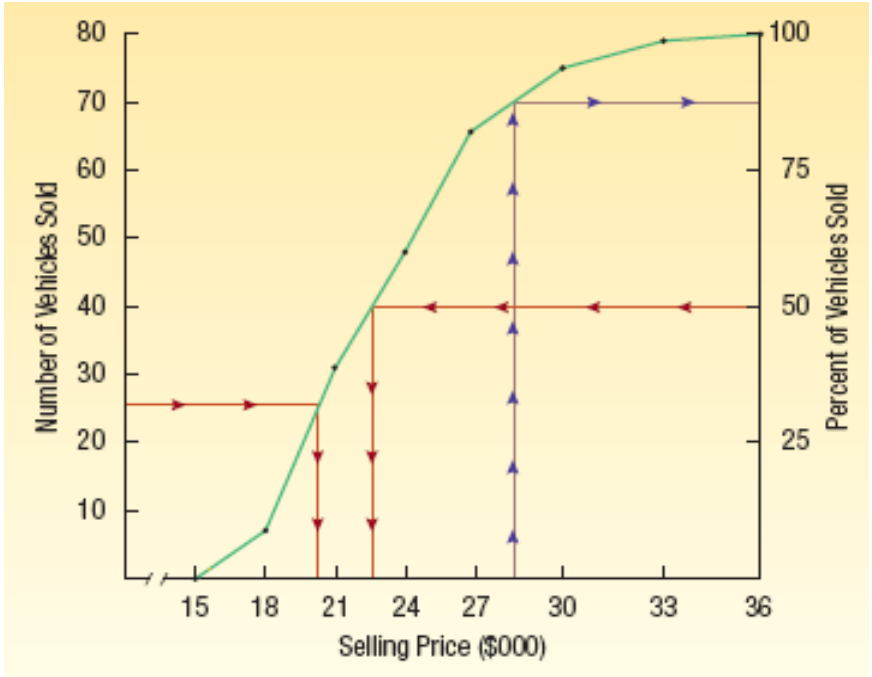
Graphic Presentation of a Frequency Distribution

The three commonly used graphic forms are:

- Histograms
- Frequency polygons (Polygons)
- Cumulative frequency distributions (Ogive)







Measures

Central Tendency	Dispersion	
• Mean • **Median** • Even - Odd • **Mode** • No Mode • Unimodal • Bimodal • Multimodal	• Range • Variance • Standard Deviation • Inter-quartile Range (IQR)	• **Skewness** • $0<Sk\leq 3$ Positive Skewness • $-3\leq Sk<0$ Negative Skewness • SK= 0 Symmetric • **Kurtosis** • $K < 0$ Flat • $K > 0$ Thin • $K = 0$ Symmetric

The Mean

- For ungrouped data, the sample mean is the sum of all the sample values divided by the number of sample values:

SAMPLE MEAN $$\overline{X} = \frac{\Sigma X}{n}$$ **[3–2]**

where:

$\overline{X}$ is the sample mean. It is read "X bar."

n is the number of values in the sample.

The Median

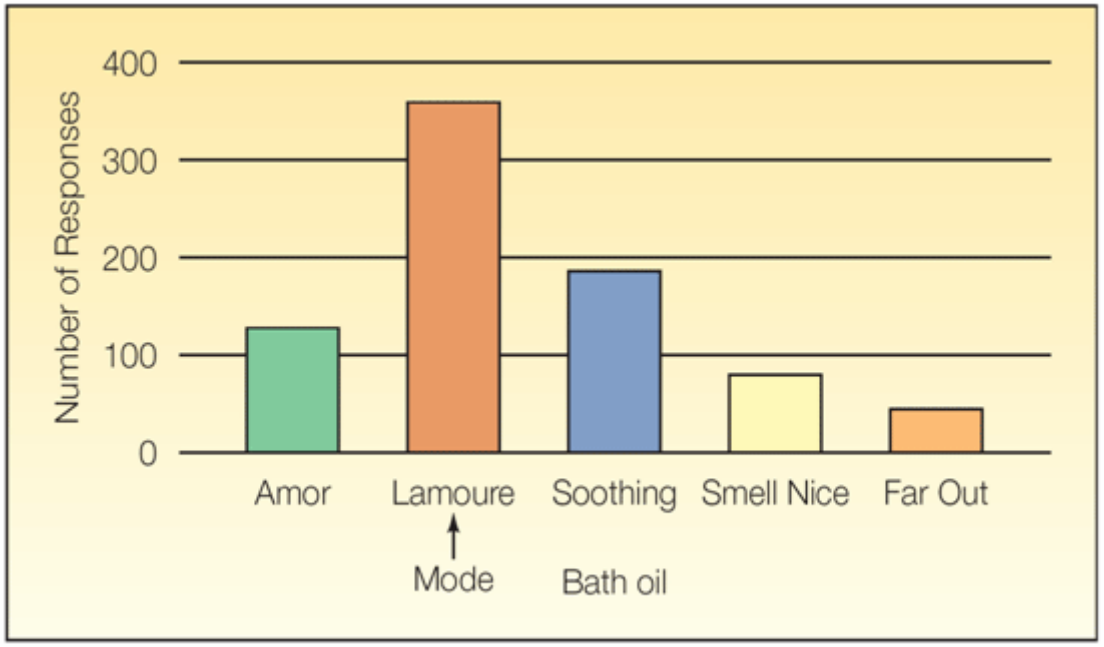
MEDIAN The midpoint of the values after they have been ordered from the smallest to the largest, or the largest to the smallest.

PROPERTIES OF THE MEDIAN

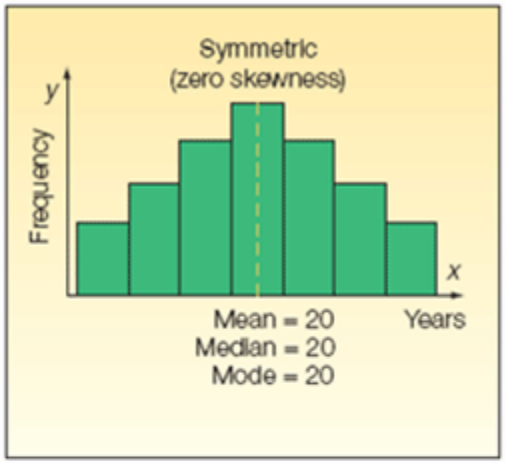
1. There is a unique median for each data set.
2. It is not affected by extremely large or small values and is therefore a valuable measure of central tendency when such values occur.
3. It can be computed for ratio-level, interval-level, and ordinal-level data.

The Mode

MODE The value of the observation that appears most frequently.

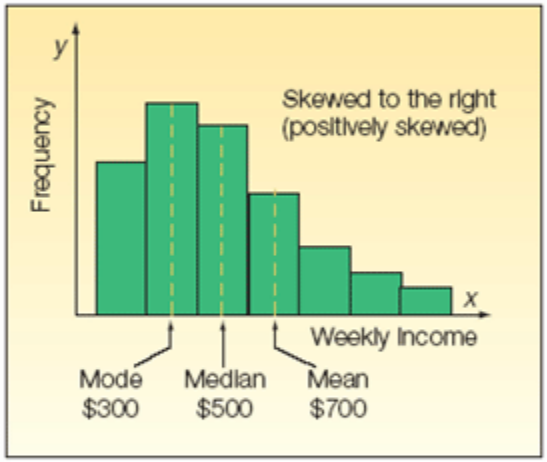


The Relative Positions of the Mean, Median and the Mode



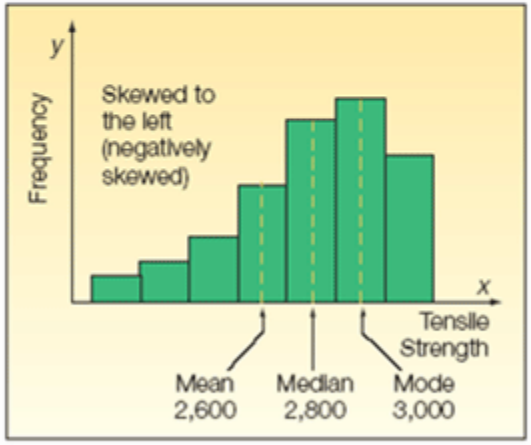
zero skewness

mode = median = mean



positive skewness

mode < median < mean



negative skewness

mode > median > mean

Measures of Dispersion

- Range

RANGE Range = Largest value − Smallest value **[3–6]**

- Variance and Standard Deviation

POPULATION VARIANCE $$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N}$$ **[3–8]**

POPULATION STANDARD DEVIATION $$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$ **[3–9]**

- IQR = Q3 – Q1

Skewness

The coefficient of skewness can range from -3 up to 3.

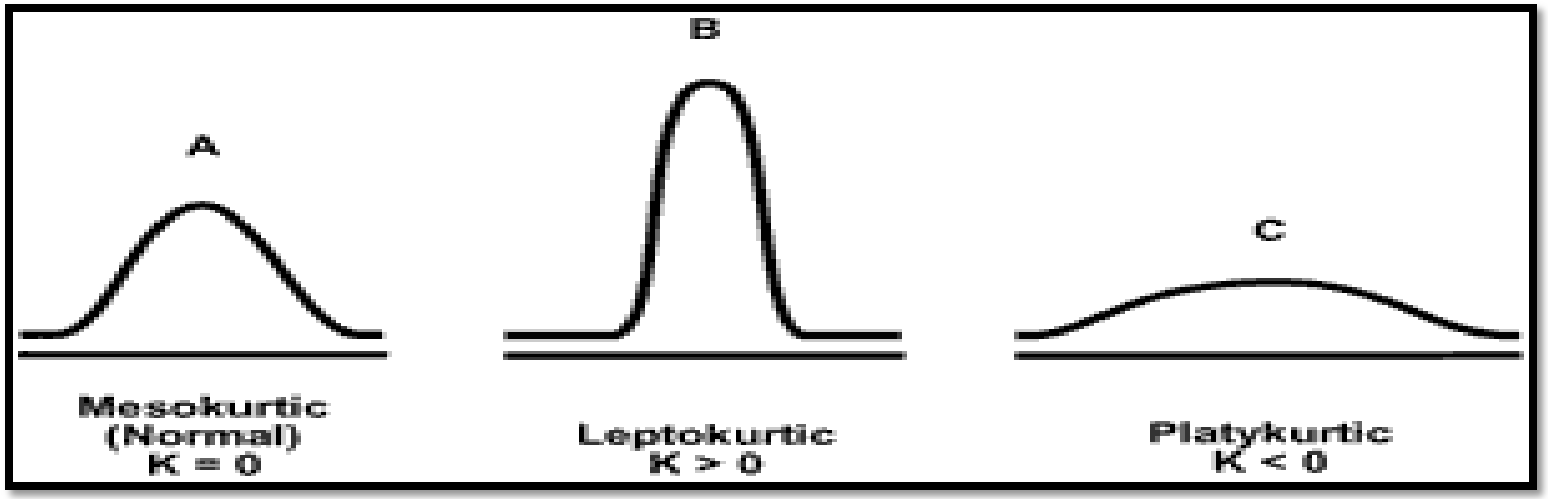
- A value near -3, indicates considerable negative skewness.
- A value such as 1.63 indicates moderate positive skewness.
- A value of 0, which will occur when the mean and median are equal, indicates the distribution is symmetrical and that there is no skewness present.

PEARSON'S COEFFICIENT OF SKEWNESS

$$sk = \frac{3(\bar{X} - \text{Median})}{s} \qquad [4\text{–}2]$$

Kurtosis

- **Kurtosis > 0 (Thin)**
- **Kurtosis = 0 (Normal)**
- **Kurtosis < 0 (Flat)**



Boxplot / Stem and Leaf

- **Boxplot:**
- Shape – Q1 – Q2 (Median) – Q3 – IQR – Minimum – Maximum – Range – Outliers

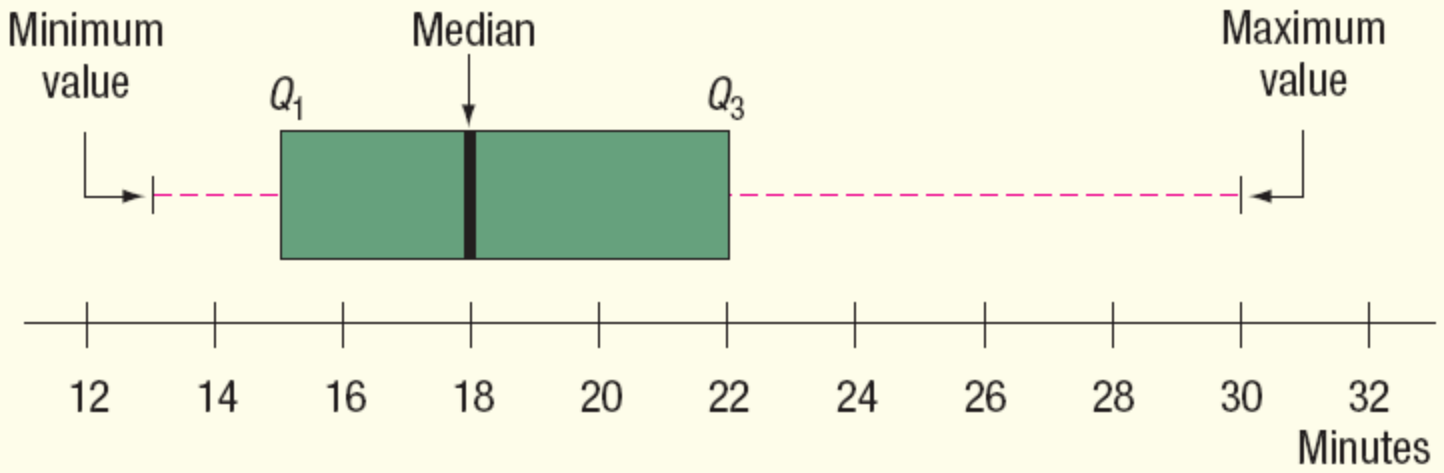
- **Stem and Leaf:**
- Shape - Minimum – Maximum – Range – Mode

Boxplot

Step1: Create an appropriate scale along the horizontal axis.

Step 2: Draw a box that starts at *Q1 (15 minutes) and ends at Q3 (22* minutes). Inside the box we place a vertical line to represent the median (18 minutes).

Step 3: Extend horizontal lines from the box out to the minimum value (13 minutes) and the maximum value (30 minutes).



Stem-and-leaf Plot

Stem	Leaf
8	8 9
9	6 3 5 6 4 4 7
10	8 7 3 4 6 3
11	7 3 2 7 2 1 9 8 3
12	7 5 7 0 5 5 0 4
13	9 5 2 9 4 6 8
14	8 2 3
15	6 5 5