Methods for Describing Sets of Data

I. Describing Qualitative Data

Example: Data on 22 Adult Aphasiacs

APHASIA.MTP ***

↓	C1-T TYPE
1	Brocas
2	Anomic
3	Anomic
4	Conduction
5	Brocas
6	Conduction
7	Conduction
8	Anomic
9	Conduction
10	Anomic
11	Conduction
12	Brocas
13	Anomic
14	Brocas
15	Anomic
16	Anomic
17	Anomic
18	Conduction
19	Brocas
20	Anomic
21	Conduction
22	Anomic
23	

A *class* is one of the categories into which qualitative data can be classified.

Identify the classes (categories) in this example.

A *class frequency* is the number of observations in the data set falling in a particular class.

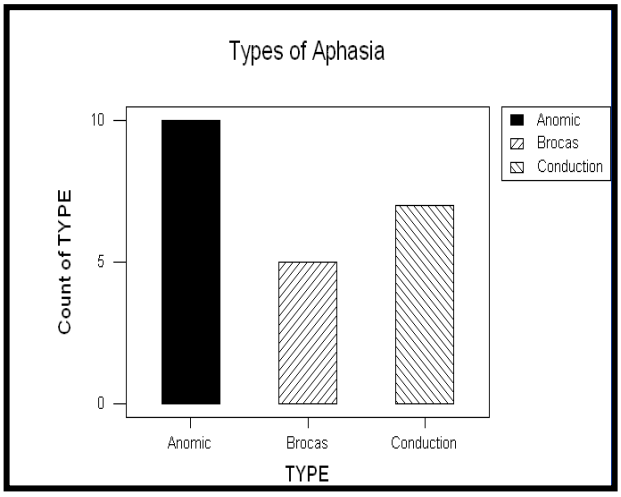
The *class relative frequency* is the class frequency divided by the total number of observations in the data set.

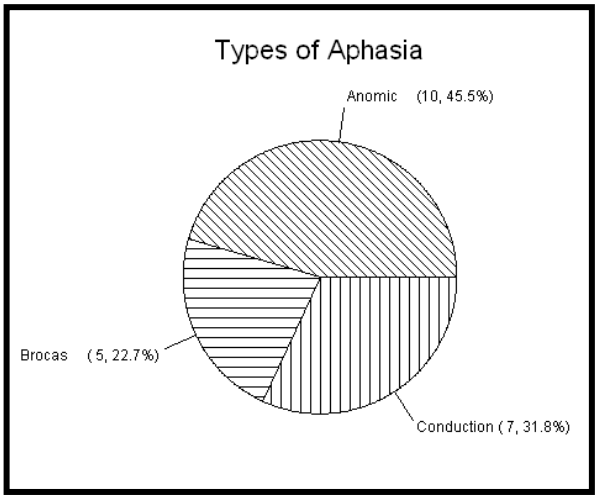
The *class percentage* is the class relative frequency multiplied by 100.

Complete the summary table below.

Type	Frequency	Relative Frequency	Percent	Cumulative Percent
Anomic				
Brocas				
Conduction				
Total				

Graphing Qualitative Data

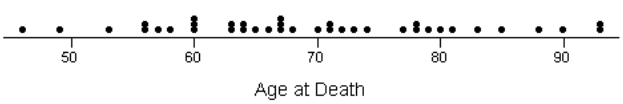




II. Describing Quantitative Data - Graphical

President	Age at Death
Washington	67
Adams	60
Jefferson	83
Madison	85
Monroe	73
Adams	80
Jackson	78
Van Buren	79
Harrison	68
Tyler	71
Polk	53
Taylor	65
Fillmore	74
Pierce	64
Buchanan	77
Lincoln	56
Johnson	66
Grant	63
Hayes	70
Garfield	49
Arthur	56
Cleveland	71
Harrison	67
McKinley	58
Roosevelt	60
Taft	72
Wilson	67
Harding	57
Coolidge	60
Hoover	90
Roosevelt	63
Truman	88
Eisenhower	78
Kennedy	46
Johnson	64
Nixon	81
Ford	93
Reagan	93

Dot Plot of Presidents' Ages at Death

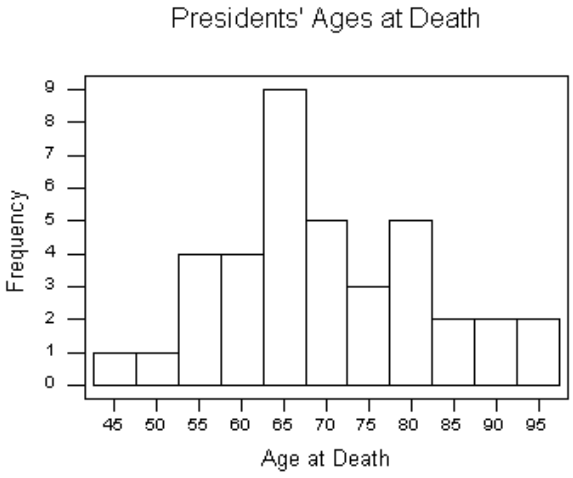


Character Stem-and-Leaf Display

```
Stem-and-leaf of Age at D  N  = 38
Leaf Unit = 1.0

  2     4 69
  3     5 3
  7     5 6678
 14     6 0003344
 (6)    6 567778
 18     7 011234
 12     7 7889
  8     8 013
  5     8 58
  3     9 033
```

Histogram



III. Numerical Measures of Central Tendency

The *central tendency* of a set of measurements is the tendency of the data to cluster, or center, about certain numerical values.

The *mean* of a set of quantitative data is the sum of the measurements divided by the number of measurements contained in the data set.

Minitab calculation of mean death age:

Column Mean

```
Mean of Age at Death = 69.605
```

The *median* of a quantitative data set is the middle number when the measurements are arranged in ascending (or descending) order. If the number of measurements is even, then the median is the mean of the two measurements in the middle.

Minitab calculation of the median death age:

Column Median

```
Median of Age at Death = 67.500
```

The *mode* is the measurement that occurs most frequently in the data set.

In the case of the death ages of presidents, 60 and 67 occur most frequently.

If we consider the histogram of the presidents' ages at death shown on the previous page, the measurement class containing the largest relative frequency is called the *modal class*. In that example the modal class is the interval 62.5-67.5. We can call the mode the midpoint of the interval or 65.

IV. Numerical Measures of Variability

The *variability* of a set of measurements is the spread of the data.

The *range* of a set of quantitative data is the difference between that largest and smallest measurement.

Minitab calculation of the range of the presidents' death ages:

Column Range

```
Range of Age at Death = 47.000
```

The *sample variance* for a sample of n measurements is equal to the sum of the squared distances from the mean divided by $(n - 1)$. The symbol s^2 is used to represent the sample variance.

In the case of the presidents ages at death the sample variance is **141.434566**.

The *sample standard deviation*, s, is defined to be the positive square root of the sample variance, s^2. That is $s = \sqrt{s^2}$.

Minitab Calculation of the Standard Deviation

Column Standard Deviation

```
Standard deviation of Age at Death = 11.893
```

In-Class Example (EPAGASS)