

Statistical Data:

- Quantitative: Whose values are numbers.
This can be of four types:
 - Nominal: area codes used in long distance dialing.
 - Ordinal: the ratings of car body design from 1(ugly) to 5(superb)
 - Interval: boiling temperatures of various liquids. \pm operations allowed.
 - Ratio: Computer assembly times. (most common) $\pm, \times, /$ operations allowed.
- Qualitative: Whose values are the names of categories into which the data items have been sorted. e.g. Sex, color of eye, type of occupation etc.

Statistical Population and the Sample:

- A *population* is the complete collection of people, things, or objects which someone wishes to study. e.g. collection of all GPA's of US students in 1997.
- A *Sample* is a subcollection of the population. e.g. a collection of GPA's representing only two scores from each state of US. We are forced to look at samples primarily for the following reasons: *Economy, Timeliness, Large populations, Inaccessibility, Destructiveness of observations and Greater accuracy.*

[see p-.11-13]

Usually samples are chosen using *Random numbers.*

Exploring Statistical Data:

- **Graphical Display of Qualitative Data:**
Sort the data into groups defined by possible variable values. Then count how many data items were sorted into each category.

Frequency: The count for a particular category.

Relative frequency: The decimal value obtained by dividing the frequency by total number of data items. Consider the data on number of professional US women employed in 1986.

Occupation	Freq.	Rel. freq.
EE/CS	347	0.0491
Health Care	1937	0.2739
Education	2833	0.4007
Social/Legal	698	0.0987
Arts/Spotrs	901	0.1274
All others	355	0.0502
<hr/>		
Total	7071	1.0

Common ways to pictorially illustrate the categorical data are:

- (a) Bar graphs and charts. [Slab> barplot]
- (b) Pie graphs or charts. [Slab> pie]

show barplot and piechart.

- **Graphical Display of Quantitative Data:**
When the variable we are studying is numeric, the raw data will arrive in a disorganized stream. Our task will be to organize this data so that we can see any patterns that may exist. This is done using some common graphical displays such as *stem-and leaf* display, *histogram*, *boxplot* etc. We consider the first two displays for a data on 50 college students.

Stem-and-Leaf: To create the stem (that is, the vertical scale of the table) disregard the last digit of the data. The stem consists of all the values between the lowest and highest values. The last digits are then reattached to their respective points on the stem. [Slab> stem]

Stem-and-leaf display of weights (in lbs.) of
50 college students.

9 : 8

10 : 18

11 : 0255688

12 : 00089

13 : 257

14 : 2358

15 : 004478

16 : 122578

17 : 00667

18 : 3468

19 : 01558

20 : 5

21 : 5

Histogram: First create the *frequency distribution*.

Class	Freq.	RFreq.	CFreq.
$80 \leq x < 100$	1	0.02	0.02
$100 \leq x < 120$	9	0.18	0.20
$120 \leq x < 140$	8	0.16	0.36
$140 \leq x < 160$	10	0.20	0.56
$160 \leq x < 180$	11	0.22	0.78
$180 \leq x < 200$	9	0.18	0.96
$200 \leq x < 220$	2	0.04	1.00
Total	50	1.00	

Histogram is a device that restores the visual part of the stem-and-leaf display to the frequency distribution.

[Slab> hist or hplot] See p.33 in the text.

When mid-point (classmark) of each class interval is plotted against the frequency or relative frequency we get **frequency polygon**. When the frequency of each class is added to the sum of the frequencies for the lower classes we obtain **Cumulative frequency distribution**. `show histogram.`