

Data Visualization Introduction

Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects (points, lines or bars, graphs) contained in graphics.

VISUAL REPRESENTATION OF THE DATA

Bar chart

A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart.

Line graph

A line chart displays information as a series of data points connected by straight line segments.

Pie chart

A pie chart is a circular statistical graphic which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice is proportional to the quantity it represents.

Pictogram

In the pictogram, the symbols are used instead of bars as in a bar chart represent to frequencies or other values of interest.

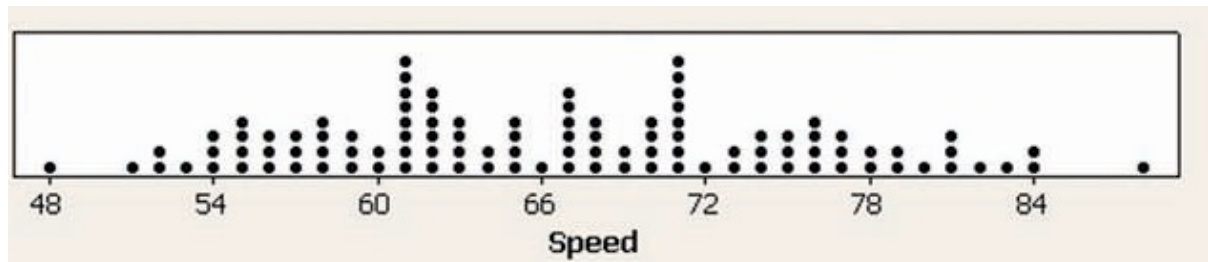
Histogram

The histogram describes a frequency distribution by using a series of adjacent rectangles, each of which has a length proportional to either the frequency or the relative frequency of the class it represents.

Dotplot

The **dotplot** displays each data value as a dot and allows us to readily see the shape of the distribution as well as the high and low values.

A. Raw Data									
54.2	58.7	71.2	69.7	51.7	75.6	53.7	57.0	82.5	76.8
62.1	67.4	64.6	70.5	48.4	69.0	65.2	64.1	65.8	56.9
82.0	62.0	54.3	73.2	74.5	73.6	76.2	55.1	66.9	62.4
70.7	68.3	62.8	83.5	54.8	68.3	69.4	59.2	60.9	60.4
60.2	75.4	55.4	56.3	77.1	61.2	50.8	67.1	70.6	60.7
56.0	80.7	80.1	56.2	70.1	63.7	70.9	54.6	61.6	63.2
72.2	84.0	76.8	61.6	61.1	80.8	58.3	52.8	74.3	71.4
63.2	57.8	61.9	75.8	80.8	57.3	62.6	75.1	67.6	78.0
52.2	57.6	61.1	66.9	88.5	66.7	61.2	73.9	79.1	70.2
65.2	61.3	69.5	72.8	57.5	71.4	64.7	78.4	67.6	76.3
78.6	66.8	71.1	58.9	61.1					



STEM-AND-LEAF DISPLAY

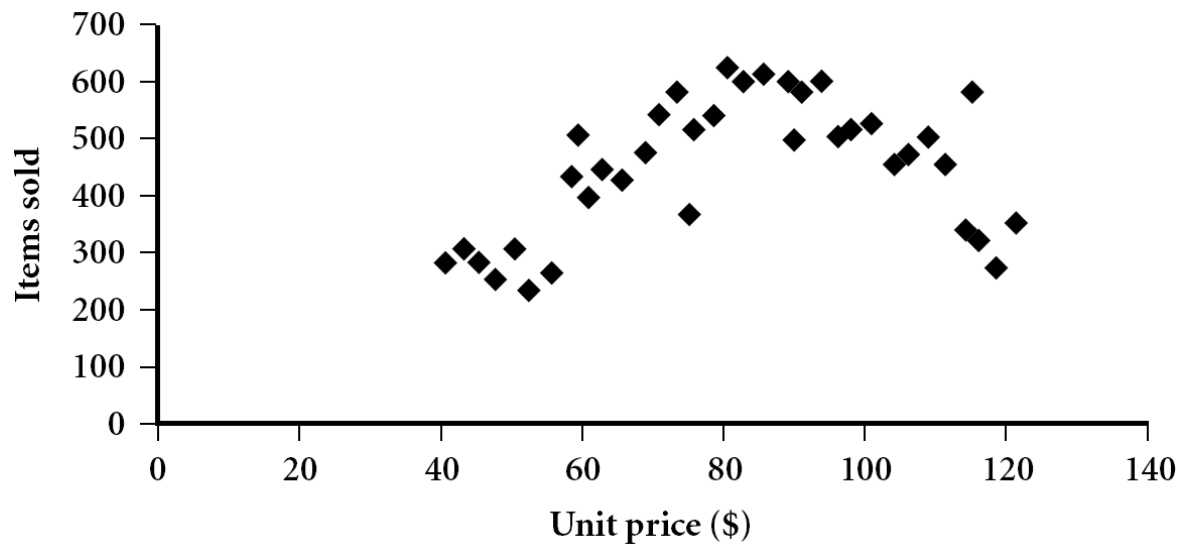
- The **stem-and-leaf display**, a variant of the frequency distribution, uses a subset of the original digits as class descriptors.
- The figure to the left of the divider (|) is the **stem**, and the digits to the right are referred to as **leaves**.

Johnson	Nixon	Ford	Carter	Reagan	Bush	Clinton
30	43	66	31	78	44	38

Stem (10's Digit)	Leaf (1's Digit)
3 018	(represents 30, 31, and 38)
4 34	(represents 43 and 44)
5	(no data values in the 50s)
6 6	(represents 66)
7 8	(represents 78)

THE SCATTER DIAGRAM

- Relationship between two quantitative variables
- Each point in the diagram represents a pair of known or observed values of two variables, generally referred to as y and x , with y represented along the vertical axis and x represented along the horizontal axis.
- The two variables are referred to as the **dependent** (y) and **independent** (x) variables, since a typical purpose for this type of analysis is to estimate or predict what y will be for a given value of x .



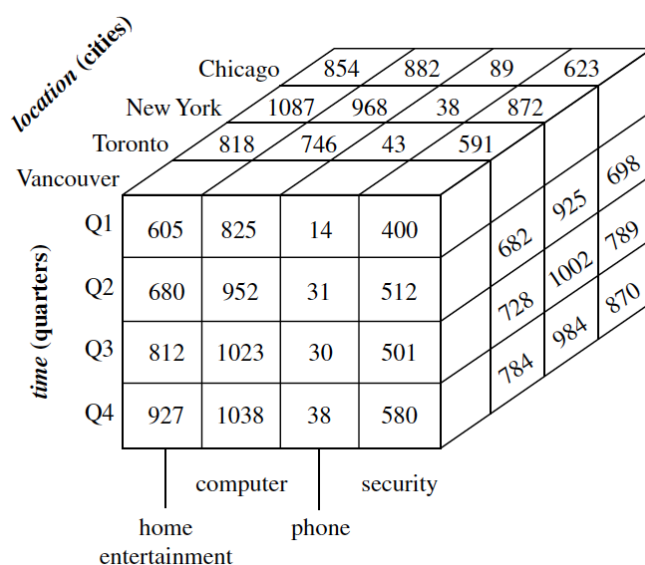
CONTINGENCY TABLES

- The **cross-tabulation**, also known as the **crosstab** or the **contingency table**, summarizes data in *combinations* of categories.

C. Cross-Tabulation or Contingency Table, by Age Category and Gender

Age Category:	Gender:		Total
	1 = Male	2 = Female	
1 = under 30 yr.	7	10	17
2 = 30–60 yr.	13	8	21
3 = over 60 yr.	8	4	12
Total	28	22	50

Multidimensional data model -A 3-D data cube representation



Dimensions: Item, Time, Location

Dimensions are qualitative values that categorize, segment, and reveal the details in data. They define the level of granularity of data.

Measures are quantitative values formed by performing aggregation on raw data with respect to the dimensions defined.