

# STATISTICS

Statistics is concerned with scientific methods for collecting, organizing, summarizing, presenting, and analyzing data as well as with drawing valid conclusions and making reasonable decisions on the basis of such analysis.

(Murray Spiegel, Statistics)

# STATISTICS

Statistics is the field of study concerned with the collection, analysis, and interpretation of uncertain data.

The methods of statistics allow scientists and engineers to design valid experiments and to draw reliable conclusions from the data they produce.

(William Naviti, Statistics for Engineers and Scientists)

# STATISTICS

The field of **statistics** deals with the collection, presentation, analysis, and use of data to make decisions, solve problems, and design products and processes.

Statistical techniques can be a powerful aid in designing new products and systems, improving existing designs, and designing, developing, and improving production processes.

(Douglas Montgomery, Applied Statistics and Probability for Engineers)

# Countries Ranked by Population Size: 2011, 2025, and 2050

(in millions)

Rank	2011		2025		2050	
	Country	Population	Country	Population	Country	Population
1	China	1,337	India	1,396	India	1,657
2	India	1,189	China	1,395	China	1,304
3	United States	311	United States	351	United States	423
4	Indonesia	246	Indonesia	279	Nigeria	402
5	Brazil	203	Nigeria	234	Indonesia	313
6	Pakistan	187	Brazil	232	Pakistan	291
7	Nigeria	166	Pakistan	228	Ethiopia	278
8	Bangladesh	159	Bangladesh	198	Brazil	261
9	Russia	139	Ethiopia	140	Bangladesh	250
10	Japan	127	Mexico	130	Philippines	172
	Mexico (11)	114	Philippines (11)	129	Mexico (11)	148
	Philippines (12)	102	Russia (12)	128	Russia (16)	109
	Ethiopia (13)	91	Japan (13)	123	Japan (17)	107

Note - Table shows all countries projected to be in top 10 during the interval 2011-2050 and their ranking in each year.

# SUMMARY OF COVID-19 CASES IN THE PHILIPPINES

As of: Jul 10, 2020

GMA NEWS  
ONLINE

TOTAL CASES

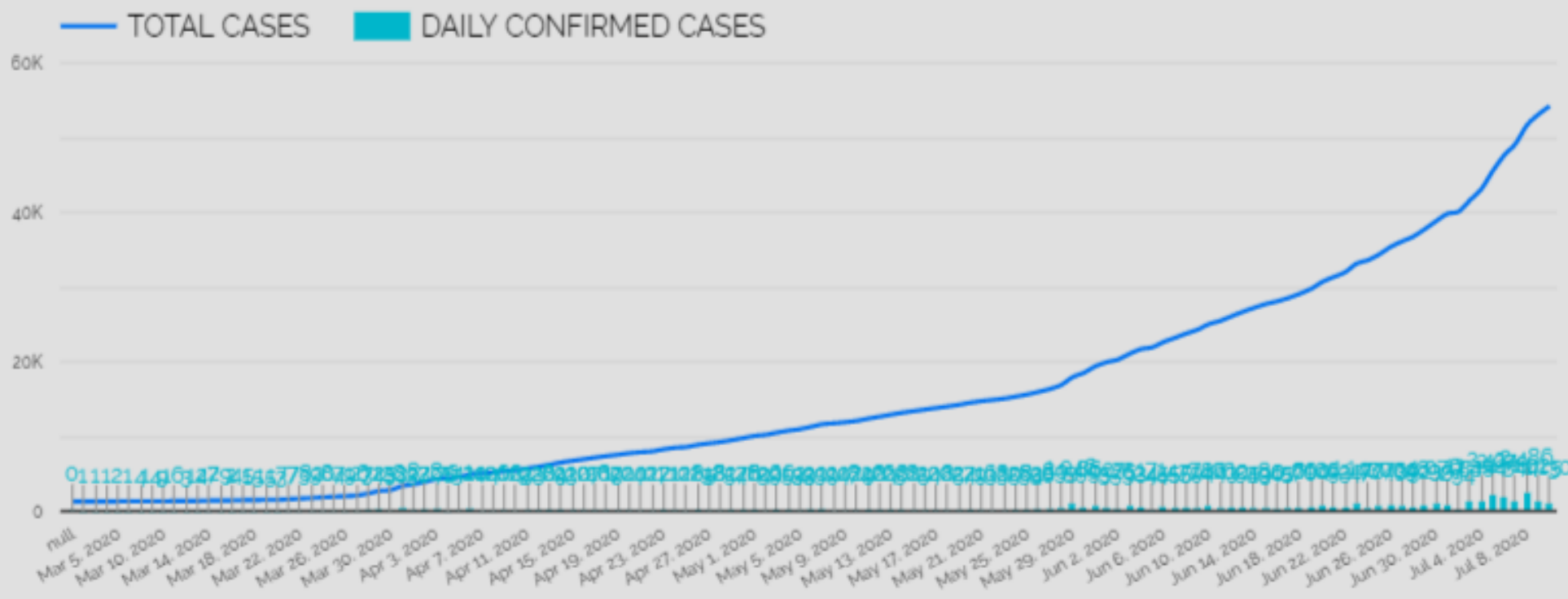
54,222

RECOVERIES

13,055

DEATHS

1,356



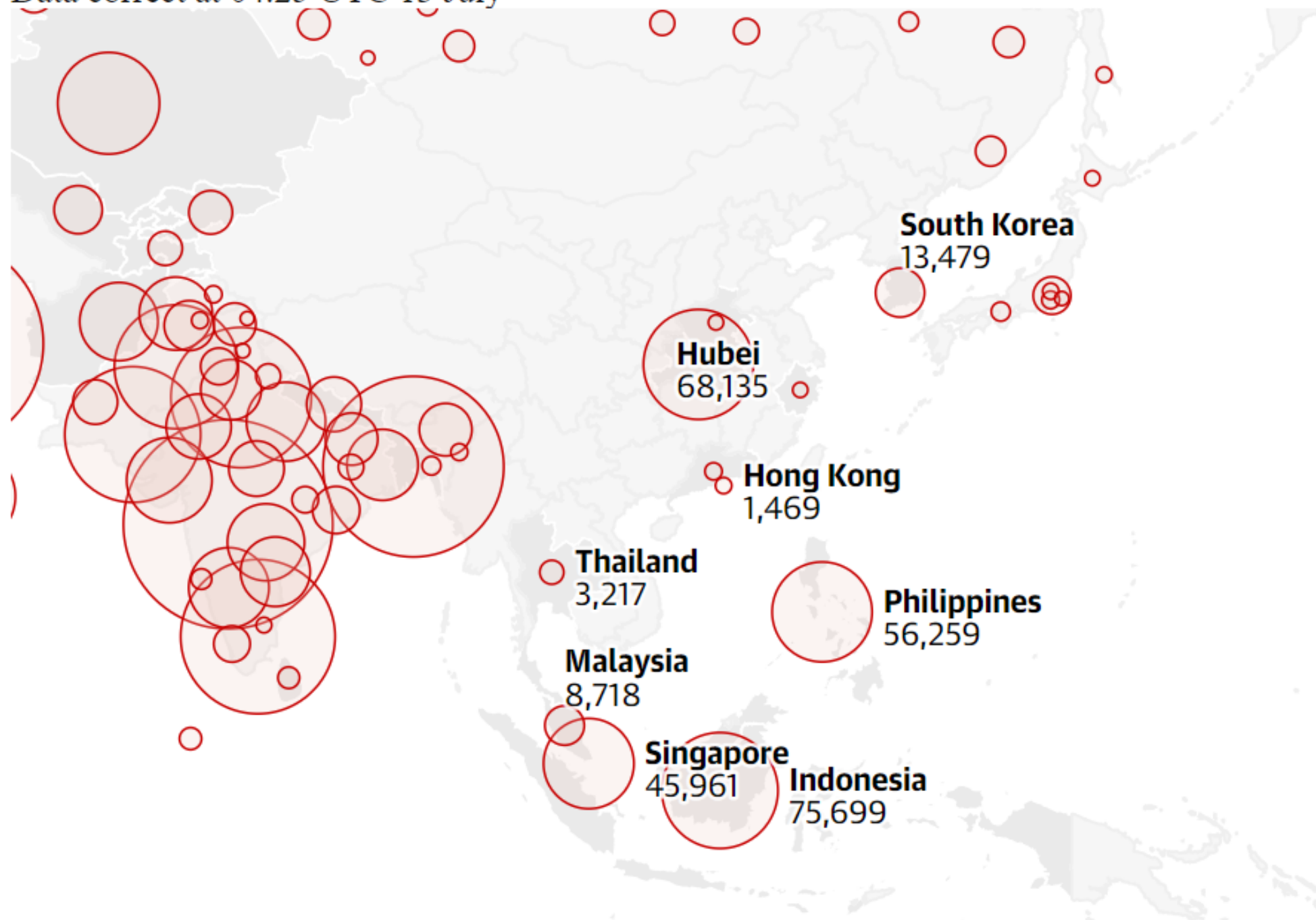
## Number of Cases per Location

(Click on any location below to update the graphics specific to the selected location)

RESIDENCE IN THE PHILIPPINES		CONFIRMED CASES
1.	For validation	20,708
2.	CEBU CITY	5,701
3.	QUEZON CITY	3,807
4.	CITY OF MANILA	2,730
5.	No data	0
6.	CALOOCAN CITY	1,210
7.	CITY OF MAKATI	1,104
8.	CITY OF PARAÑAQUE	993
9.	TAGUIG CITY	924
10.	CITY OF PASIG	875
		1 - 10 / 843 < >

# Confirmed cases of Covid-19 in Asia

Data correct at 04.25 UTC 13 July



# Population and Sample

**Population** – consists of the totality or the entire collection of objects or outcomes about which information is sought

**Sample** – a subset of a population, containing the objects or outcomes that are actually observed



# Variable

- A **variable** is a symbol, such as  $X$ ,  $Y$ ,  $H$ ,  $x$ , or  $B$ , that can assume any of a prescribed set of values, called the domain of the variable.
- If the variable can assume only one value, it is called a **constant**.
- A variable is the characteristic that is being studied.

# Variable

- Discrete variable and continuous variable
- Qualitative variable and quantitative variable
- Dependent variable and independent variable

# Types of Data (Measurement Scales)

- Nominal
- Ordinal
- Interval
- Ratio

# Nominal

- Nominal scales are used to name or label a series of values.

**What is your hair color?**

- ☒ 1 – Brown
- ☐ 2 – Black
- ☐ 3 – Blonde
- ☐ 4 – Gray
- ☐ 5 – Other

# Ordinal

- Ordinal scales provide good information about the order of values.

**How do you feel today?**

- ☒ 1 – Very Unhappy
- ☐ 2 – Unhappy
- ☐ 3 – OK
- ☐ 4 – Happy
- ☐ 5 – Very Happy

**How satisfied are you with our service?**

- ☒ 1 – Very Unsatisfied
- ☐ 2 – Somewhat Unsatisfied
- ☐ 3 – Neutral
- ☐ 4 – Somewhat Satisfied
- ☐ 5 – Very Satisfied

# Interval

- Interval scales give the order of values and the difference between each one.



# Ratio

- Ratio scales give the order, the interval values and the ability to calculate ratios since true zero can be defined.
- Ex: height and weight

# Categories of Statistics

**Descriptive statistics** are simply numerical or graphical summaries of data, including charts, graphs, and simple summary statistics such as means and standard deviations, used to describe characteristics of a population sample.

**Inferential statistics** are statistical techniques (e.g., chi-square test, the t-test, the one-way ANOVA) that allow conclusions to be drawn about the relationships found among different variables in a population sample.



# Importance of Statistics

- Economy
- Business
- Astronomy
- Psychology
- Engineering
- Mathematics
- Science

# Statistical Presentation of Data

- Textual
- Graphical
- Tabular

# Textual

- It is a method of presenting statistical data using paragraphs of text.
- It is usually employed in official reports, where the activities, plans or programmes are described in words, inserting relevant figures in between them.

# Textual

- Example:

“...2,652 men and 1,726 women participated in an opinion poll about a certain government measure. 1,460 persons, of whom, 1,096 were male, voted against the measure. In all 2,225 persons voted for the measure, while 356 women were indifferent...”

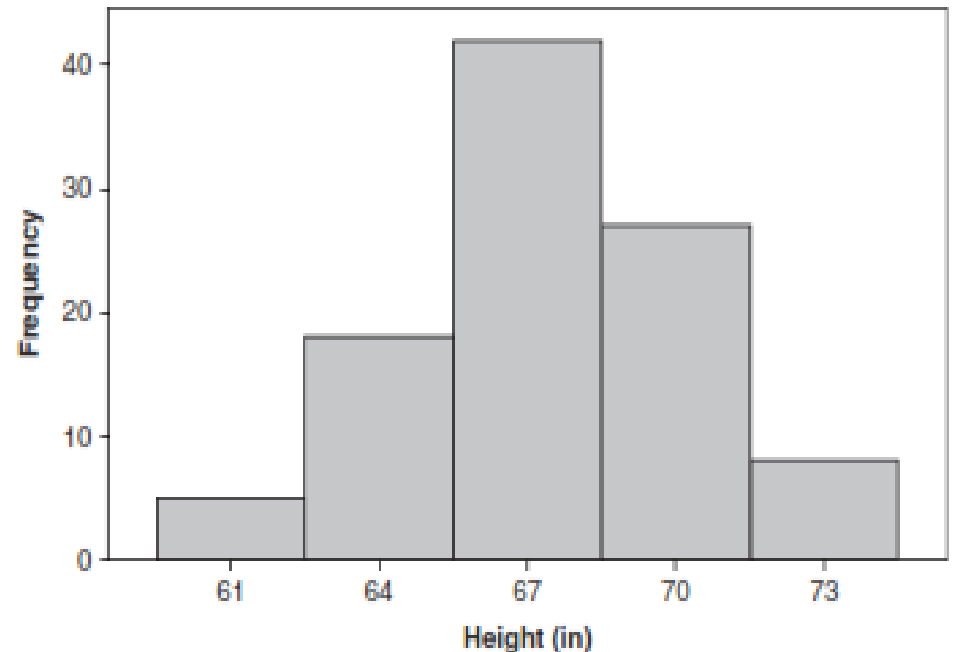
# Graphical Description

- A graph is a pictorial presentation of the relationship between variables.
- Examples: bar graphs, pie graphs, pictographs, histogram or frequency histogram, frequency polygon, etc.

# Graphical Description

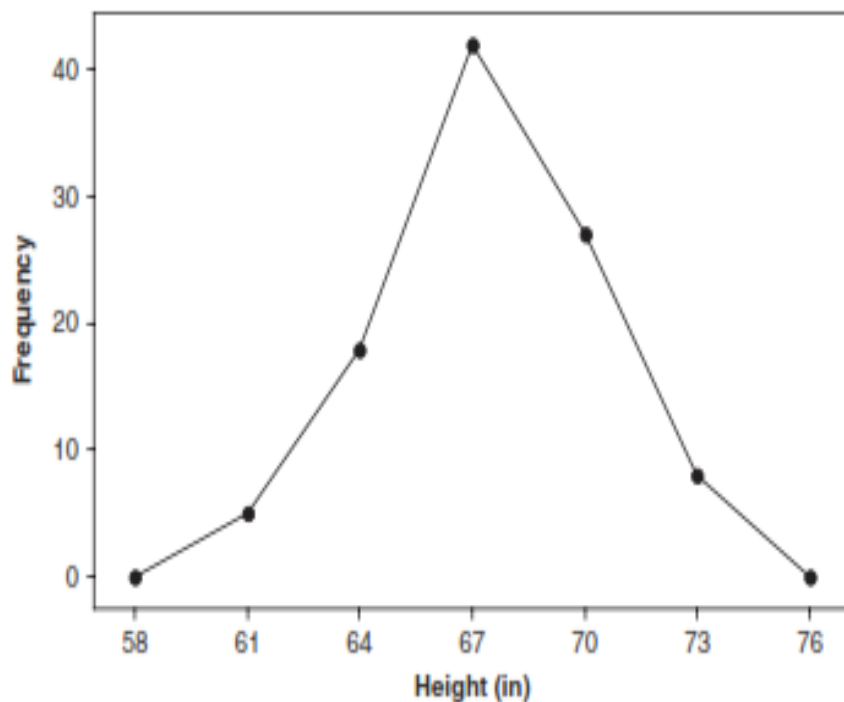
## Histogram

A **histogram or frequency histogram**, consists of a set of rectangles having (a) bases on a horizontal axis, with centers at the class marks and lengths equal to the class interval sizes, and (b) areas proportional to the class frequencies.



# Graphical Description

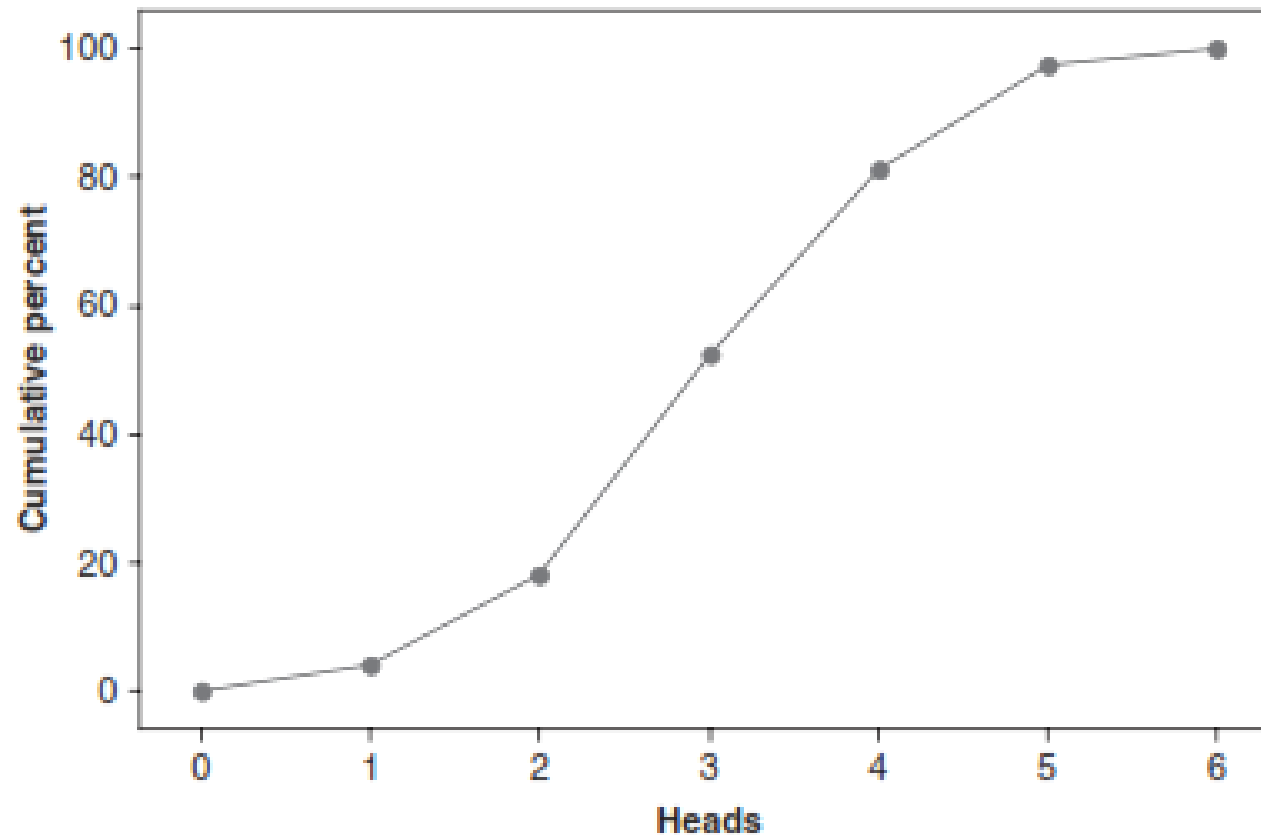
## Frequency Polygon



A **frequency polygon** is a line graph of the class frequencies plotted against class marks. It can be obtained by connecting the midpoints of the tops of the rectangles in the histogram.

# Graphical Description

## Cumulative Frequency Polygon





# Graphical Description

## Stem and leaf Plot

**Stem and leaf plot** presents a graphical display of the data using the actual numerical values of each data point.

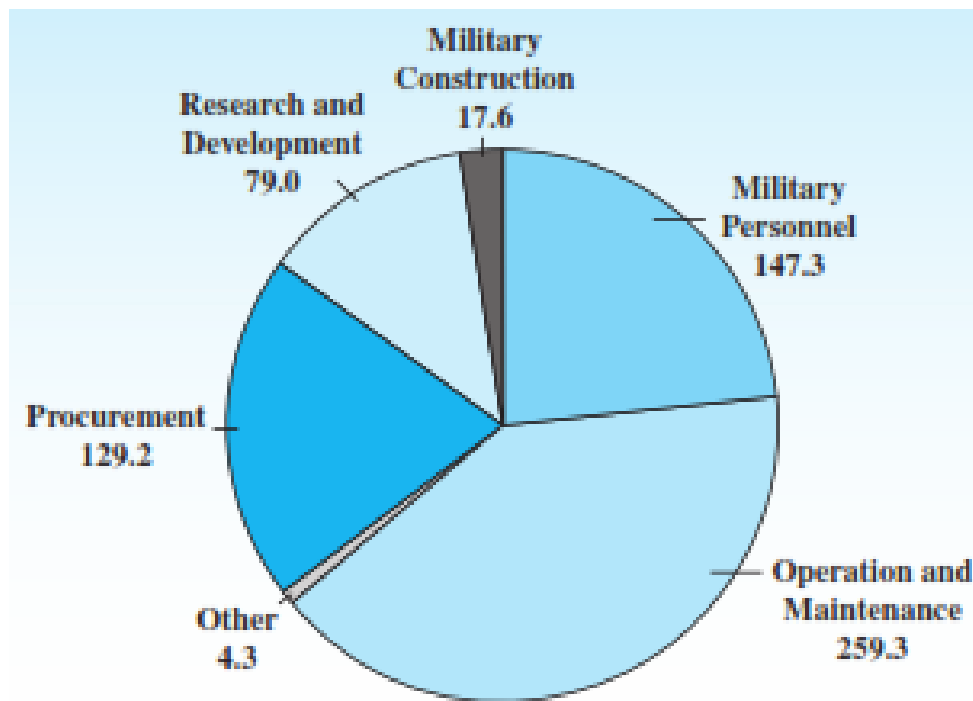
### Prices of Walking Shoes

90	70	70	70	75	70
65	68	60	74	70	95
75	70	68	65	40	65
70					

4		0
5		
6		5 8 0 8 5 5
7		0 0 0 5 0 4 0 5 0 0
8		
9		0 5

# Graphical Description

## Pie Chart

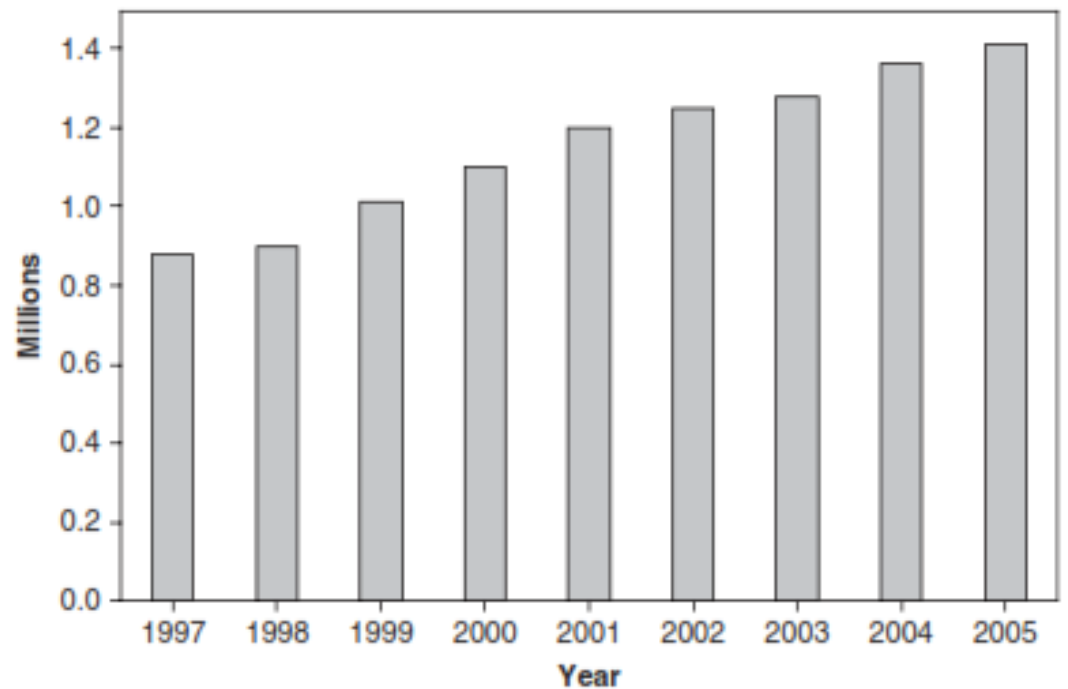


A **pie chart** is a circular graph that shows how the measurements are distributed among the categories.

# Graphical Description

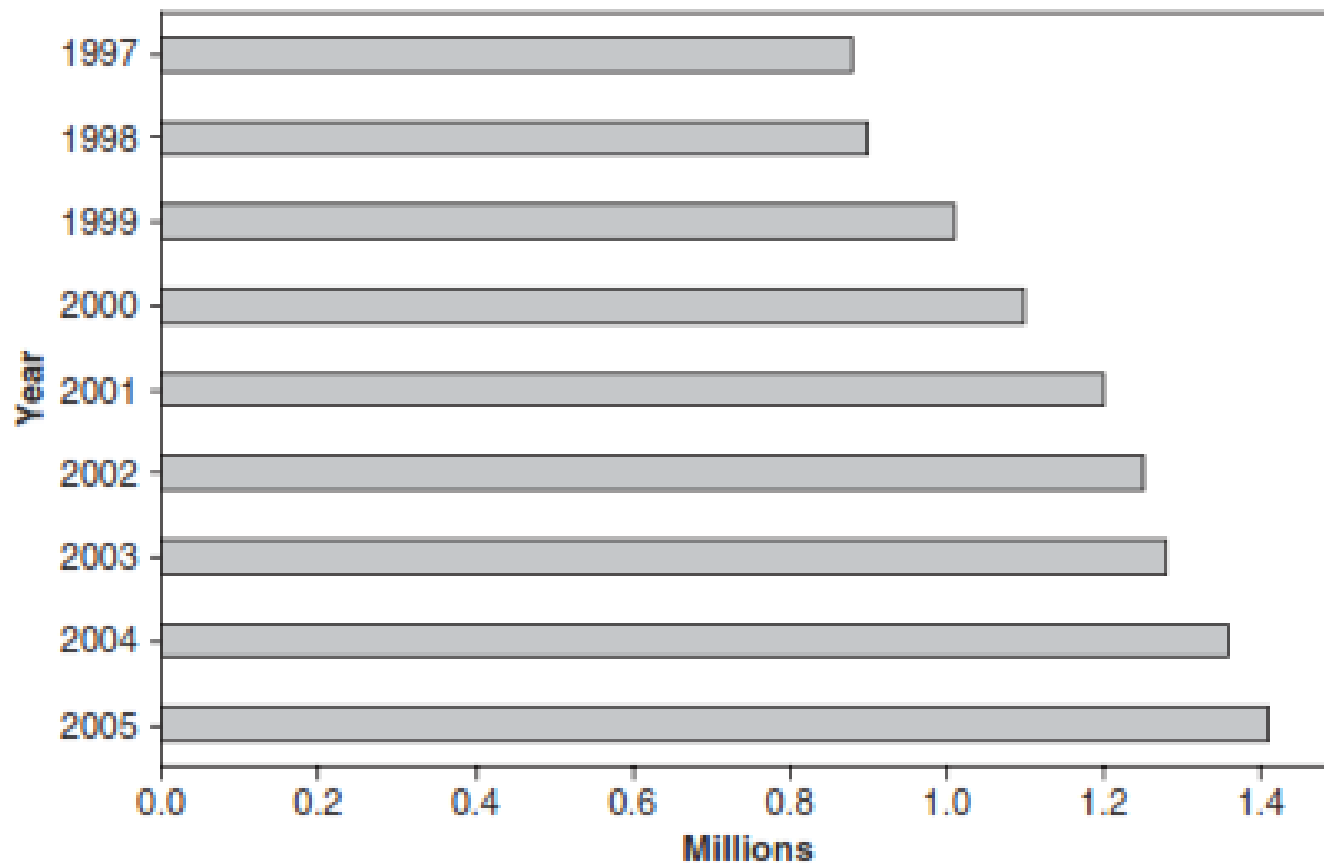
## Bar Graph

A **bar chart/bar graph** shows how the measurements are distributed among the categories, with the height of the bar measuring how often a particular category was observed.



# Graphical Description

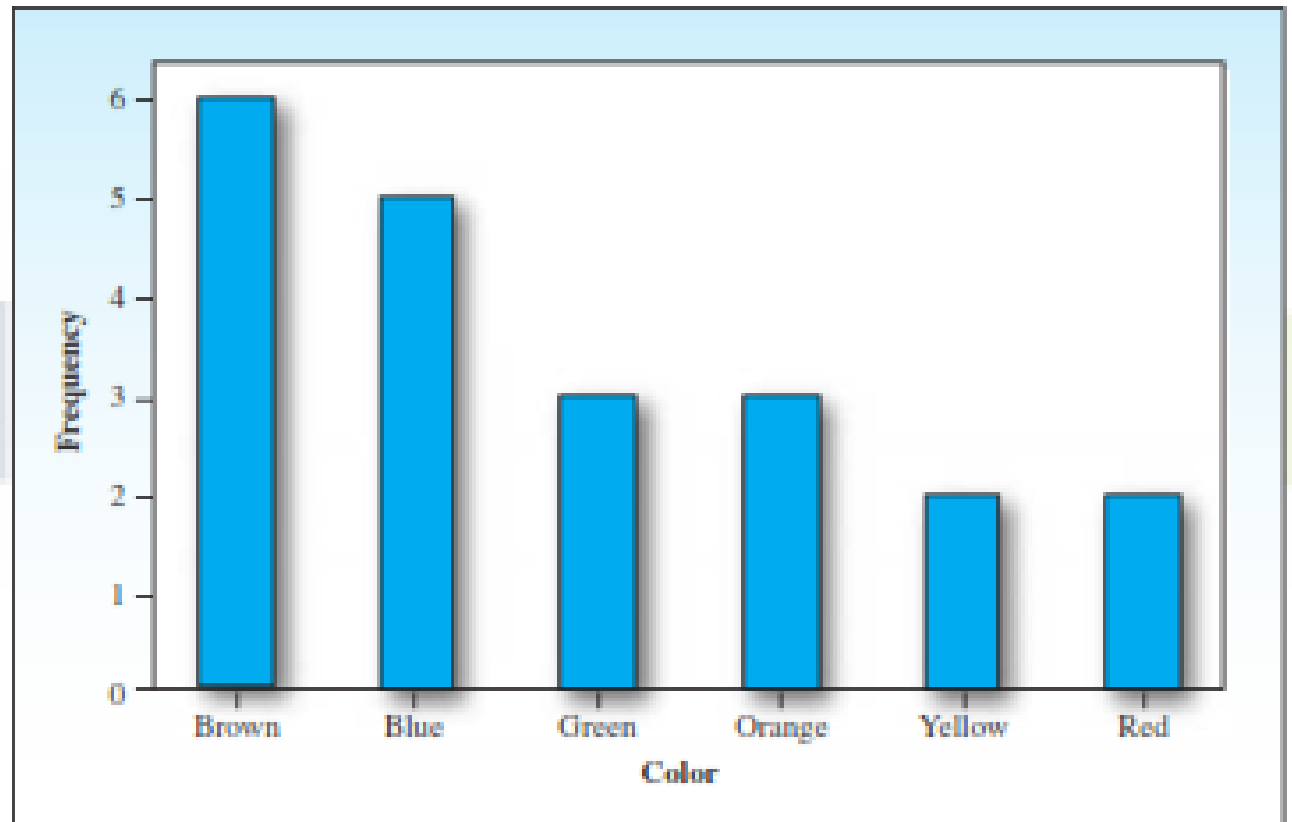
## Bar Graph



# Graphical Description

## Pareto Chart

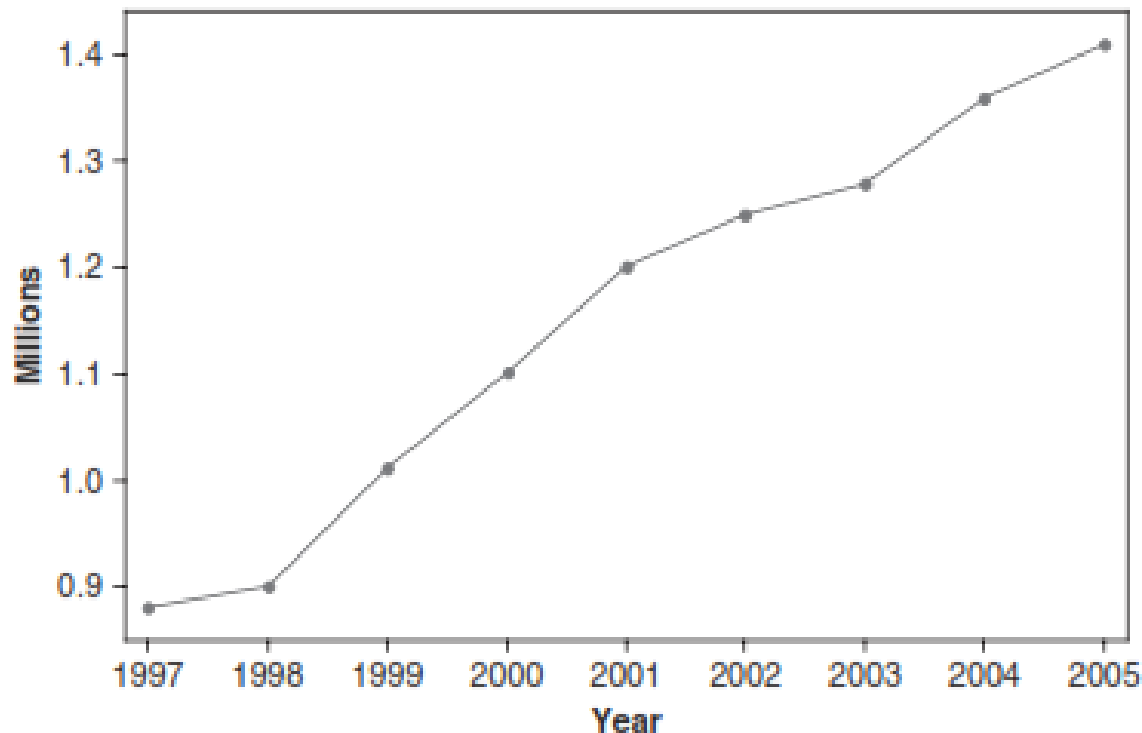
A bar chart in which the bars are ordered from largest to smallest is called a **Pareto chart**.



# Graphical Description

## Time series / Time sequence plot

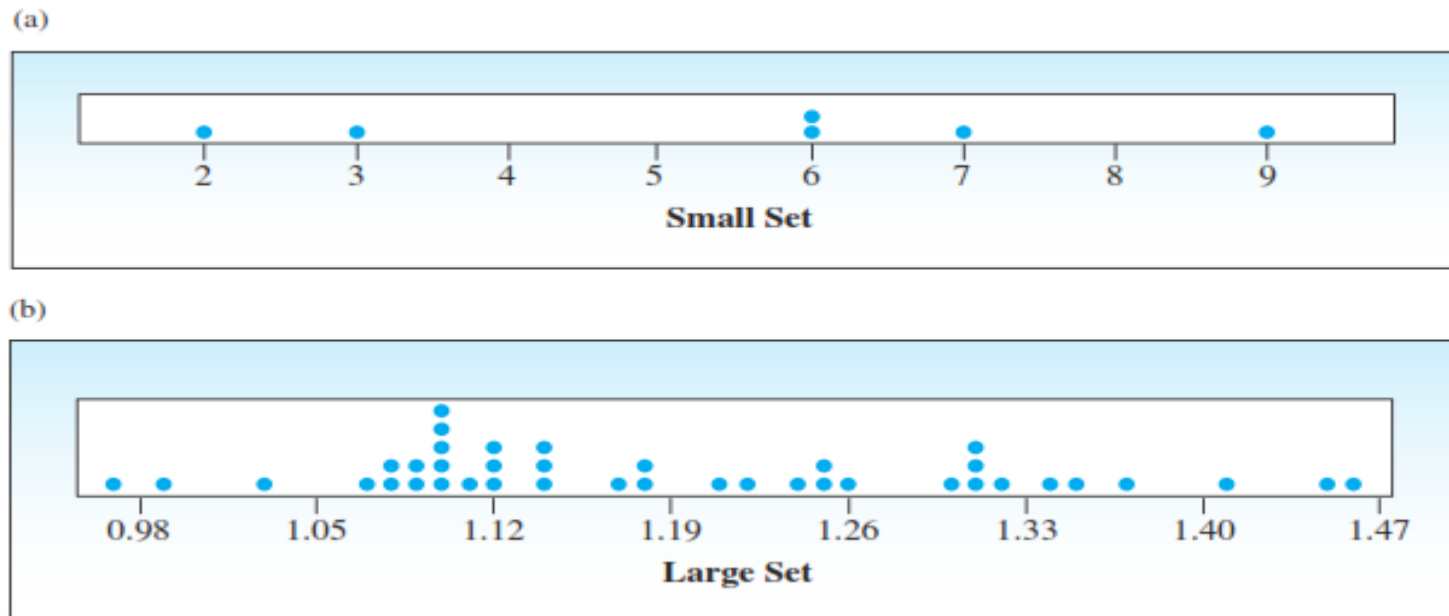
**Time series plot** is a graph with a line chart where the time intervals are marked on the horizontal axis and the projections on the vertical axis.



# Graphical Description

## Dotplot

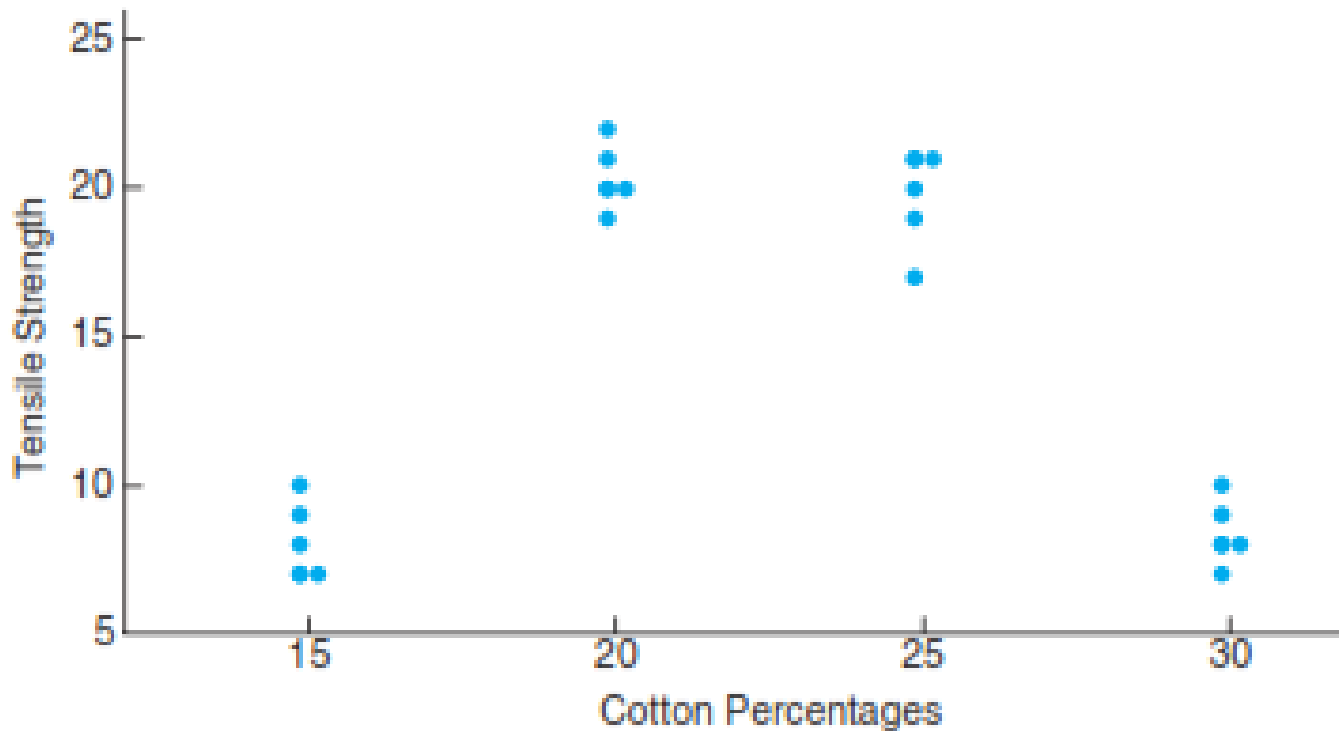
**Dotplot** is the simplest graph of quantitative data and can simply plot the measurements as points on the horizontal axis.



# Graphical Description

## Scatter Plot

Each dot on the **scatter plot** represents one observation from a data set.

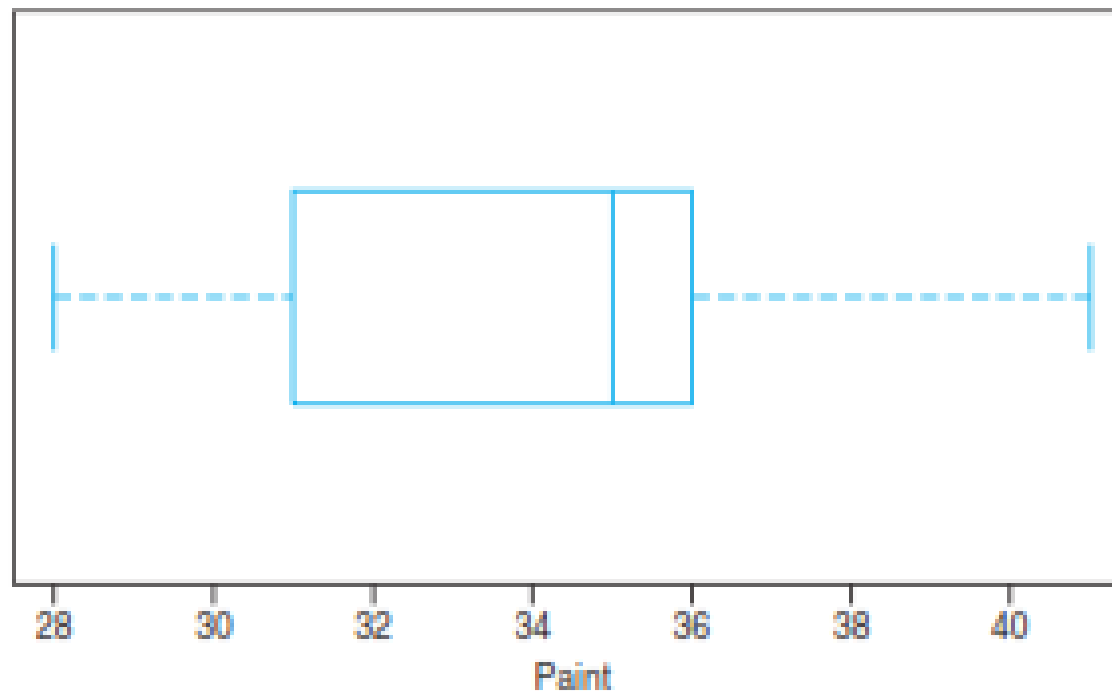




# Graphical Description

## Box-and-Whisker Plot or Box Plot

- ▶ This plot encloses the interquartile range of the data in a box that has the median displayed within.



# Frequency Curves

1. **Symmetrical or bell-shaped** curves are characterized by the fact that observations equidistant from the central maximum have the same frequency.
2. Curves that have tails to the left are said to be **skewed to the left**.
3. Curves that have tails to the right are said to be **skewed to the right**.
4. Curves that have approximately equal frequencies across their values are said to be **uniformly distributed**.

# Frequency Curves

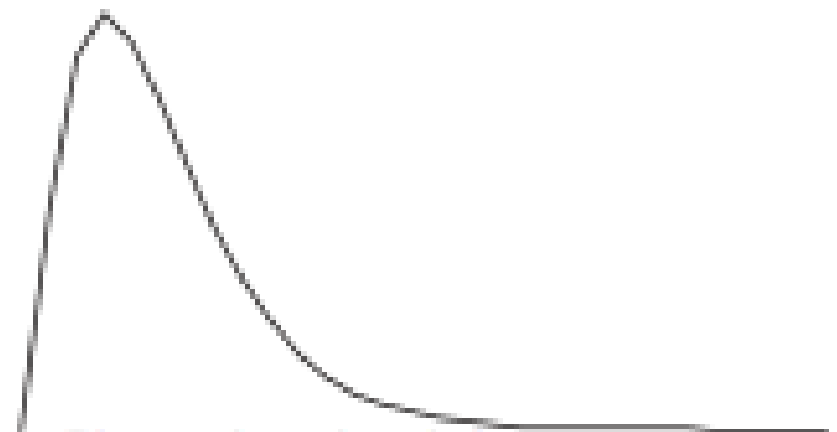
5. In a **J-shaped** or **reverse J-shaped** frequency curve the maximum occurs at one end or the other.
6. A **U-shaped** frequency distribution curve has maxima at both ends and a minimum in between.
7. A **bimodal** frequency curve has two maxima.
8. A **multimodal** frequency curve has more than two maxima.

# Frequency Curves

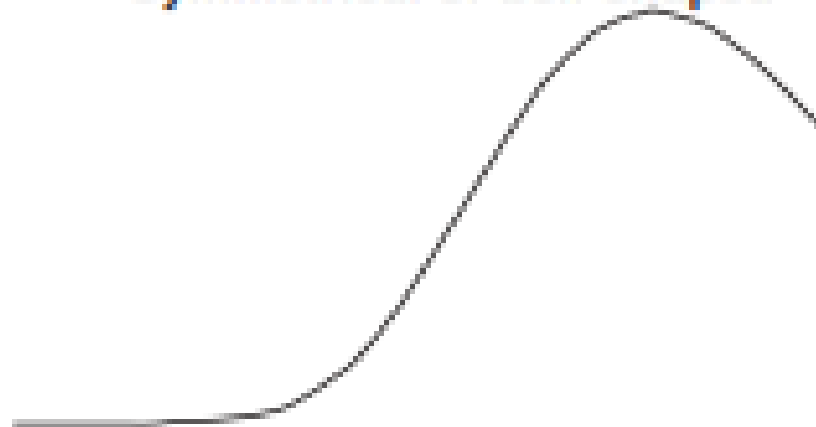
## Four Common Distributions



Symmetrical or bell-shaped



Skewed to the right



Skewed to the left



Uniform

# Frequency Distribution

- It is a tabular arrangement of data by classes together with the corresponding class frequencies.
- Class intervals and class limits
- Class boundaries
- Size or width of a class interval
- Class mark

# Frequency Distribution

- Example: Heights of 100 male students at XYZ University

Height (in)	Number of Students
60–62	5
63–65	18
66–68	42
69–71	27
72–74	8
Total 100	

# Frequency Distribution

- It is a tabular arrangement of data by classes together with the corresponding class frequencies.
- Class intervals and class limits
- Class boundaries
- Size or width of a class interval
- Class mark

# Frequency Distribution

## General Rules for Forming Frequency Distributions:

1. Determine the largest and smallest numbers in the raw data and thus find the range (the difference between the largest and smallest numbers).
2. Divide the range into a convenient number of class intervals having the same size.
3. Determine the number of observations falling into each class interval; that is, find the class frequencies.



# Frequency Distribution

$$\textit{Class Interval} = \frac{\textit{range}}{\textit{number of classes}}$$

$$\textit{Class Interval} = \frac{\textit{highest value} - \textit{lowest value}}{1 + 3.322 \log N}$$

## Example:

Construct a frequency distribution with suitable class interval size of marks obtained by 50 students of a class are given below:

23, 50, 38, 42, 63, 75, 12, 33, 26, 39, 35, 47,  
43, 52, 56, 59, 64, 77, 15, 21, 51, 54, 72, 68,  
36, 65, 52, 60, 27, 34, 47, 48, 55, 58, 59, 62,  
51, 48, 50, 41, 57, 65, 54, 43, 56, 44, 30, 46,  
67, 53

Step 1. Arrange the data from lowest to highest

Step 2. Identify the lowest and highest values

Highest = 77

Lowest = 12

Step 3. Use the formula for Class Interval

$$\textbf{Class Interval} = \frac{\textit{highest value} - \textit{lowest value}}{1 + 3.322 \log N}$$

where N is the total number of data

$$\textbf{Class Interval} = \frac{77 - 12}{1 + 3.322 \log 50} \approx \mathbf{10}$$

Step 4. Create frequency distribution table

Start with the lowest value

Scores		f
12	21	3
22	31	4
32	41	7
42	51	13
52	61	13
62	71	7
72	81	3