16

Description of Data II

KSU COLLEGE OF MEDICINE
2019 - 2020

ACKNOWLEDGMENTS

TEAM MEMBERS

RAZAN ALZAHRANI

RAHAF ALSHUNAIBER

REVIEWER

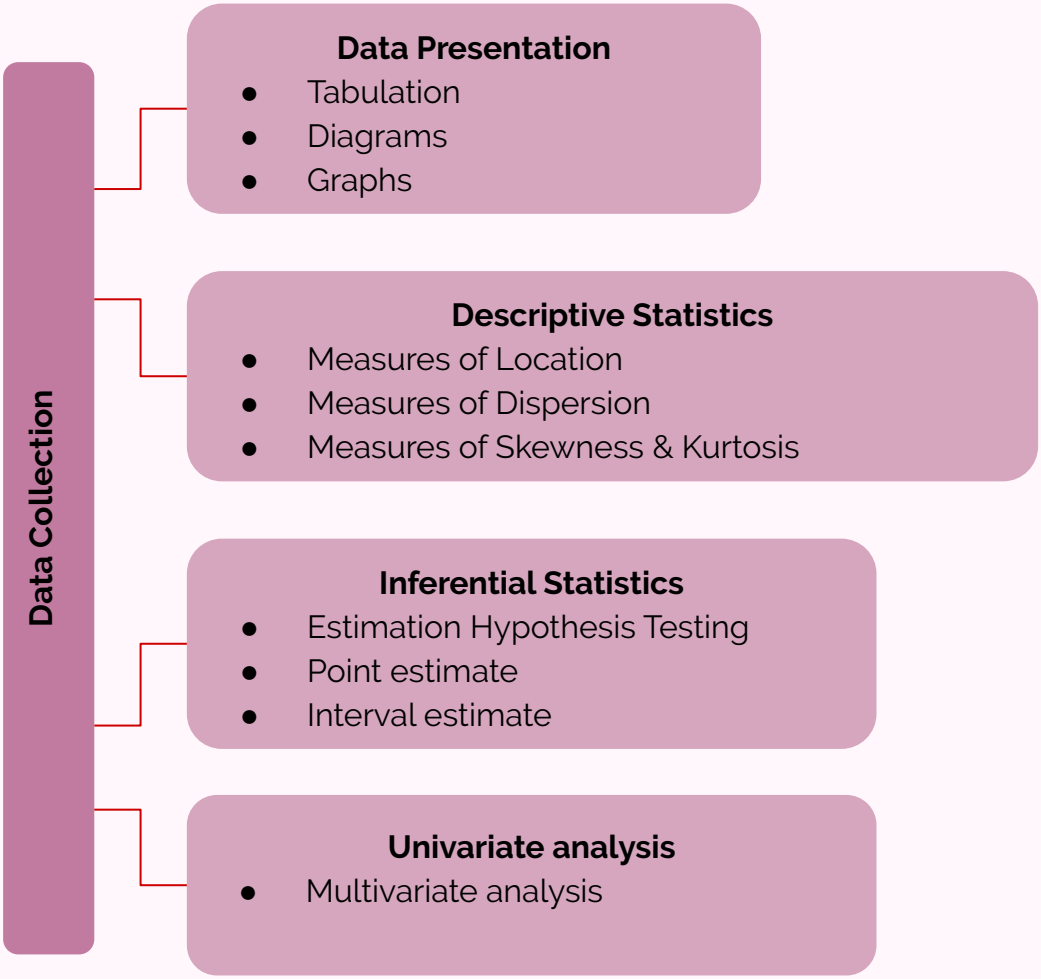
YAZEED AL-DOSSARE

TABLE OF **CONTENTS**

LECTURE **OBJECTIVES**

- To know how to make frequency distributions and its importance
- To know different terminology in frequency distribution table
- To learn different graphs/diagrams for graphical presentation of data.

Investigation



Frequency Distributions

"A Picture is Worth a Thousand Words"

Frequency Distributions

it's the distribution of your data, how much frequency is in the data.

what is the
- Shape
- Range
of the data

- Data distribution
  - pattern of variability.
  - The center of a distribution
  - The range
  - The shapes

2 Types :

- Simple frequency distributions
- Grouped frequency distributions

Example A :
Data of Research midterm results

Simple Frequency Distribution

"Ungrouped data"

- The number of times that score occurs

Ex (A) :
Count how many of the student got 14, 13,...

- Make a table with highest score at top and decreasing for every possible whole number

doesn't matter either in decreasing or increasing order

- N (total number of scores) always equals the sum of the frequency

total frequency = total sample size = number of student (Ex A)

- $Sf = N$

Categorical or Qualitative Frequency Distributions

- What is a categorical frequency distribution?

A categorical frequency distribution represents data that can be placed in specific categories, such as gender, blood group, & hair color, etc.

Categorical or Qualitative Frequency Distributions - Example

The blood types of 25 blood donors are given below. Summarize the data using a frequency distribution.

Ungrouped data →

AB	B	A	O	B
O	B	O	A	O
B	O	B	B	B
A	O	AB	AB	O
A	B	AB	O	A

Assume there is 250 donors
it will be complicated
so we put them in groups

Class (Blood Type)	Frequency, f
A	5
B	8
O	8
AB	4
Total	n = 25

Note: The classes for the distribution are the blood types.

Quantitative Frequency Distributions -- Ungrouped

- What is an ungrouped frequency distribution?

An ungrouped frequency distribution simply lists the data values with the corresponding frequency counts with which each value occurs.

Quantitative Frequency Distributions – Ungrouped -- *Example*

The at-rest pulse rate for 16 athletes at a meet were
57, 57, 56, 57, 58, 56, 54, 64, 53, 54, 54, 55, 57, 55, 60, and 58.
Summarize the information with an ungrouped frequency distribution.

Quantitative Frequency Distributions – Ungrouped -- *Example Continued*

Class (pulse Rate)	Frequency, f
53	1
54	3
55	2
56	2
57	4
58	2
60	1
64	1
Total	n = 16

Note: The (ungrouped) classes are the observed values themselves.

Example of a simple frequency distribution (ungrouped)

- 5 7 8 1 5 9 3 4 2 2 3 4 9 7 1 4 5 6 8 9 4 3 5 2 1
  (No. of children in 25 families)

N.o of children	N.o of families
9	3
8	2
7	2
6	1
5	4
4	4
3	3
2	3
1	3

→ there are 3 families that have 9 children

åf = 25 (No. of families)

Relative Frequency Distribution

- Proportion of the total N
- Divide the frequency of each score by N
- Rel. *f* = *f*/N
- Sum of relative frequencies should equal 1.0
- Gives us a frame of reference

Class (pulse Rate)	Frequency, f	Relative Frequency
53	1	0.0625 1/16
54	3	0.1875 3/16
55	2	0.1250 2/16
56	2	0.1250 2/16
57	4	0.2500 4/16
58	2	0.1250 2/16
60	1	0.0625 1/16
64	1	0.0625 1/16
Total	n = 16	1.0000

Note: The relative frequency for a class is obtained by computing f/n.

Example of a simple frequency distribution

- 5 7 8 1 5 9 3 4 2 2 3 4 9 7 1 4 5 6 8 9 4 3 5 2 1

	f	*rel f* Relative Frequency
9	3	.12
8	2	.08
7	2	.08
6	1	.04
5	4	.16
4	4	.16
3	3	.12
2	3	.12
1	3	.12

å*f* = 25 å rel *f* = 1.0

Cumulative Frequency Distributions

Cumulative: adding subsequent values

- *cf = cumulative frequency: number of scores at or below a particular score*
- A score's standing relative to other scores
- Count from lower scores and add the simple frequencies for all scores below that score

Example of a simple frequency distribution

5 7 8 1 5 9 3 4 2 2 3 4 9 7 1 4 5 6 8 9 4 3 5 2 1

N.o of children	*f*	*rel f*	*cf*	
9	3	.12	3	
8	2	.08	5	(3+2)
7	2	.08	7	(5+2)
6	1	.04	8	(7+1)
5	4	.16	12	(8+4)
4	4	.16	16	(12+4)
3	3	.12	19	(16+3)
2	3	.12	22	(19+3)
1	3	.12	25	(22+3)
	åf = 25	*å rel f = 1.0*		

Advantage : answer questions such as ; how many families have 5 and less children?
25 - 8 = 17 families

Example of a simple frequency distribution (ungrouped)

5 7 8 1 5 9 3 4 2 2 3 4 9 7 1 4 5 6 8 9 4 3 5 2 1

N.o of children	f	cf	rel f	rel. cf
9	3	3	.12	.12
8	2	5	.08	.20
7	2	7	.08	.28
6	1	8	.04	.32
5	4	12	.16	.48
4	4	16	.16	.64
3	3	19	.12	.76
2	3	22	.12	.88
1	3	25	.12	1.0
			åf = 25	å rel f = 1.0

Same as cf but with Rel f

Quantitative Frequency Distributions -- Grouped

- What is a grouped frequency distribution?

A grouped frequency distribution is obtained by constructing classes (or intervals) for the data, and then listing the corresponding number of values (frequency counts) in each interval.

Tabulate the hemoglobin values of 30 adult male patients listed below

Patient No	Hb (g/dl)	Patient No	Hb (g/dl)	Patient No	Hb (g/dl)
1	12.0	11	11.2	21	14.9
2	11.9	12	13.6	22	12.2
3	11.5	13	10.8	23	12.2
4	14.2	14	12.3	24	11.4
5	12.3	15	12.3	25	10.7
6	13.0	16	15.7	26	12.5
7	10.5	17	12.6	27	11.8
8	12.8	18	9.1	28	15.1
9	13.2	19	12.9	29	13.4
10	11.2	20	14.6	30	13.1

Steps for making a table

- Step1: Find Minimum (9.1) & Maximum (15.7)
- Step 2: Calculate difference 15.7 – 9.1 = 6.6
- Step 3: Decide the number and width of the classes (7 c.l) 9.0 -9.9, 10.0-10.9,----
- Step 4: Prepare dummy table – Hb (g/dl), Tally mark, No. patients

why 7? based on the difference
There is NO strict mathematical rule. it's subjective

7 class intervals, width: 9-9,9

General Rule :

- class interval:

you shouldn't have more than 10 class intervals and not less than 5 class intervals. why?
If it's more than 10, the data will be scattered. it wont give any information
If it's less than 5, the data will be condensed , it wont give any information

- width: shouldn't overlap

DUMMY TABLE

Hb (g/dl)	Tally marks	No. patients
9.0 – 9.9 10.0 – 10.9 11.0 – 11.9 12.0 – 12.9 13.0 – 13.9 14.0 – 14.9 15.0 – 15.9		
Total		

Tally Marks TABLE

Hb (g/dl)	Tally marks	No. patients
Width 9.0 – 9.9 10.0 – 10.9 11.0 – 11.9 12.0 – 12.9 13.0 – 13.9 14.0 – 14.9 15.0 – 15.9	l lll llll 1 llll llll lll lll ll	1 3 6 10 5 3 2
Total	-	30

Table Frequency distribution of 30 adult male patients by Hb

Hb (g/dl)	No. of patients
9.0 – 9.9 10.0 – 10.9 11.0 – 11.9 12.0 – 12.9 13.0 – 13.9 14.0 – 14.9 15.0 – 15.9	1 3 6 10 5 3 2
Total	30

Table Frequency distribution of adult patients by Hb and gender

In the Same table, you can put 2 columns → called: " Bi-variable" (2 variables, Hb and Gender)
You can put upto 3 variables

Hb (g/dl)	Gender		Total
	Male	Female	
<9.0	0	2	2
9.0 – 9.9	1	3	4
10.0 – 10.9	3	5	8
11.0 – 11.9	6	8	14
12.0 – 12.9	10	6	16
13.0 – 13.9	5	4	9
14.0 – 14.9	3	2	5
15.0 – 15.9	2	0	2
Total	30	30	60

Elements of a Table

Ideal table should have

- Number
- Title
- Column headings
- Foot-notes

Number

- Table number for identification in a report

Title, place Time period

- Describe the body of the table, variables, (What, how classified, where and when)

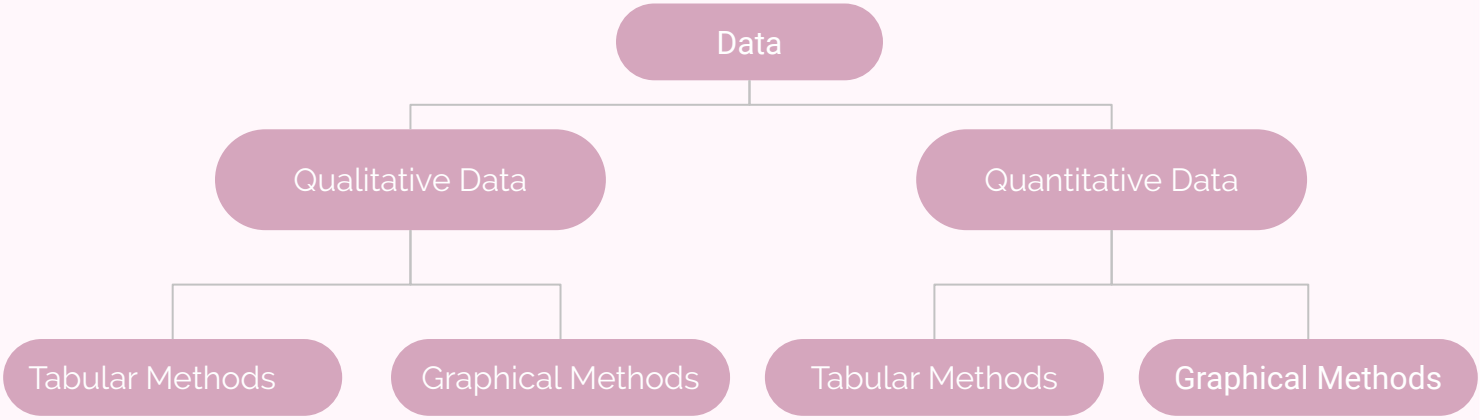
Column Heading

- Variable name, No. , Percentages (%), etc.,

Foot-note(s)

- to describe some column/row headings, special cells, source, etc.,

Tabular and Graphical Procedures



- Frequency Distribution
- Rel. Freq. Dist.
- % Freq. Dist.
- Cross-tabulation

- Bar Graph
- Pie Chart

- Frequency Distribution
- Rel. Freq. Dist.
- Cum. Freq. Dist.
- Cum. Rel. Freq. Distribution
- Cross tabulation

- Histogram
- Freq. curve
- Box plot
- Scatter Diagram
- Stem-and-Leaf Display

DIAGRAMS/GRAPHS

Quantitative data (discrete & continuous)

- Histogram
- Frequency polygon (curve)
- Stem-and –leaf plot
- Box-and-whisker plot
- Scatter diagram

Qualitative data (Nominal & Ordinal)

- Bar charts (one or two groups)
- Pie charts

Example Data

Data of 60 patients (Age)

68	63	42	27	30	36	28	32
79	27	22	28	24	25	44	65
43	25	74	51	36	42	28	31
28	25	45	12	57	51	12	32
49	38	42	27	31	50	38	21
16	24	64	47	23	22	43	27
49	28	23	19	11	52	46	31
30	43	49	12				

Histogram

Make (Age) as continues by putting class intervals
The Rectangles are attached to each other because of the continuous scale (Age)

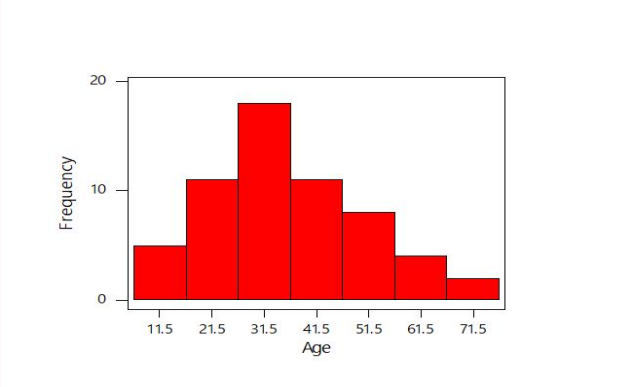
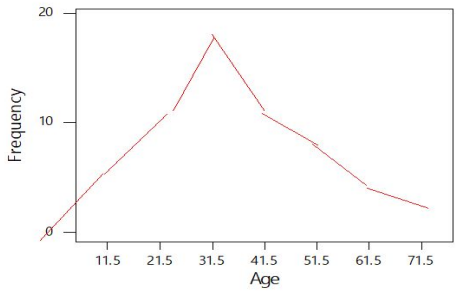


Figure 1 Histogram of ages of 60 subjects

Height of the rectangle represent the frequency
The tallest rectangle represents the highest frequency
the shortest represent the lowest frequency

Advantage :
- Minimum ,Maximum and the shape of the data can be seen

Polygon



How to draw it?
take the midpoint of each rectangle in the histogram
- Attach the midpoints with the scale → polygon
- If you draw with a smooth hand curve → curve
Both are the same

Advantage :
- Minimum ,Maximum and the shape of the data can be seen

Stem and leaf plot

```
Stem-and-leaf of Age      N = 60
Leaf Unit = 1.0
Frequency  Stem  Leaf
A 6        1     122269
 19        2     1223344555777788888
 11        3     00111226688
 13        4     2223334567999
  5        5     01127
  4        6     3458
  2        7     49
```

How to read this data?
A-there are 6 patients whose ages are 11,12,12,12,16,19
[Attach the two numerical value "stem" (1) with "leaf" (1, 2, 2,2, 6,9)]

Advantage :
- Whole data can be seen "Raw data"
- Minimum ,Maximum and the shape of the data can be seen

Descriptive statistics report: Boxplot

(Box and whisker plot)

Advantages :
1- Gives all the descriptive statistics of the data :

- minimum score
- maximum score
- lower quartile
- upper quartile
- median
- mean

2- use a Huge data →put it in one data

- The skew of the distribution

positive skew: mean > median & high-score whisker is longer

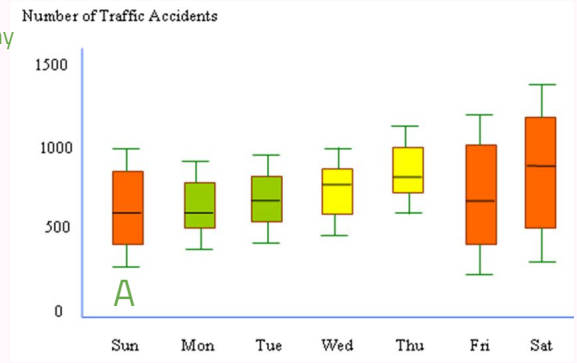
negative skew: mean < median & low-score whisker is longer

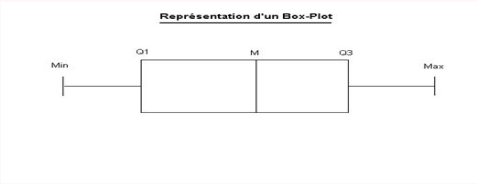
Application of a box and Whisker diagram

data of number of traffic accidents over a period of one year

A-this data represent on Sunday how many minimum and maximum accidents occurred, the mean, median, lower and upper quartile

In this diagram,where is the maximum accident?
Fri and sat

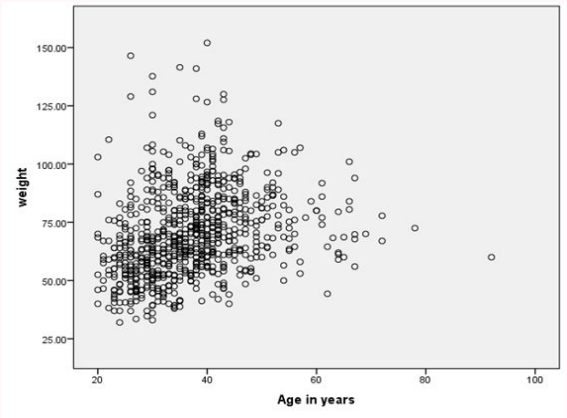




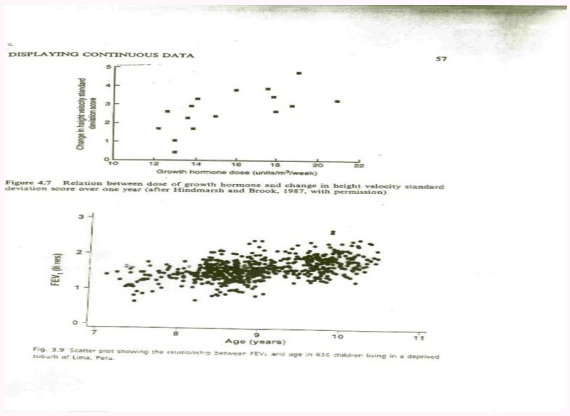
Scatter diagram

Used for :
Understanding the relationship between two quantitative variables

* each dot represent two variables of one subject

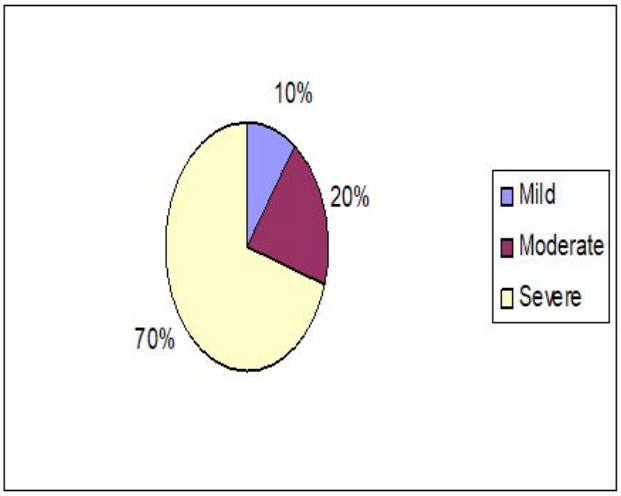


pattern:
positive linear relationship
As age increase → FEV increase



Pie Chart

- Circular diagram – total -100%
- Divided into segments each representing a category
- Decide adjacent category
- The amount for each category is proportional to slice of the pie

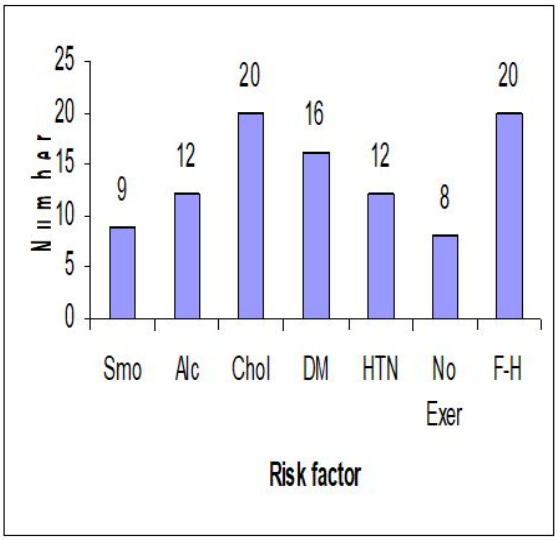


The prevalence of different degree of Hypertension in the population

Bar Graphs

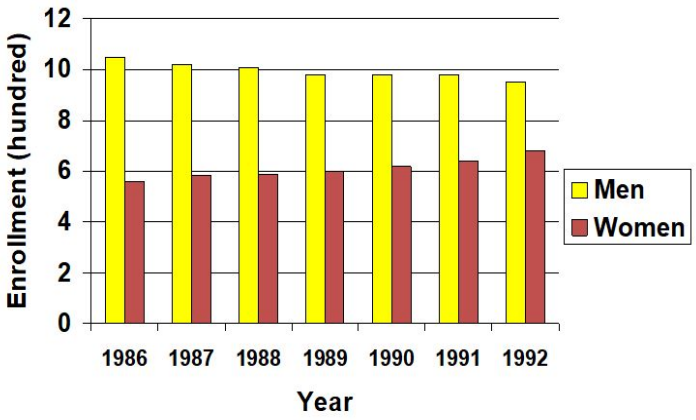
- Heights of the bar indicates frequency
- Frequency in the Y axis and categories of variable in the X axis
- The bars should be of equal width and no touching the other bars

Differences between bar graph and histogram
- In histogram there is continuity (Continuous data). With No gaps



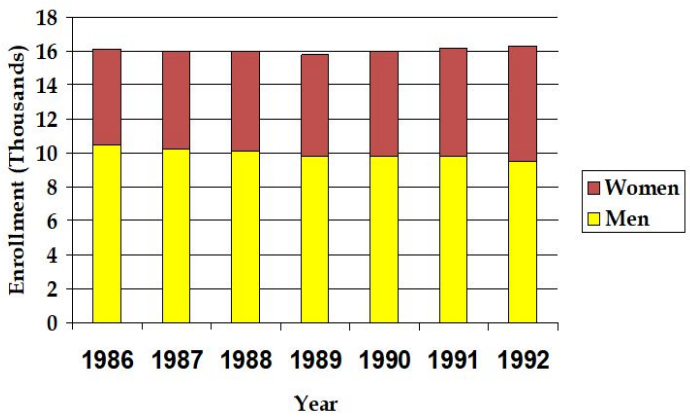
The distribution of risk factor among cases with Cardiovascular Diseases

HIV cases enrolment in USA by gender **Multiple Bar Chart**



Multiple Bar chart (more than one rectangle)

HIV cases Enrollment in USA by gender **Stocked bar chart**



Same, but instead of using two rectangles. one is used and each variable with a different color

General rules for designing graphs

- A graph should have a self-explanatory legend
- A graph should help reader to understand data
- Axis labeled, units of measurement indicated
- Scales important. Start with zero (otherwise // break) If not put // on the the x-axis
- Avoid graphs with three-dimensional impression, it may be misleading (reader visualize less easily

Title:
Table → on the top
graph → down