Editing file

Feedback

Tubular and graphical presentation of data

Objectives:

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

Color Index:

- Boys' Slides
- Girls' Slides
- Doctors Notes
- Golden Notes
- Important
- Extra

Tubular of and graphical presentation

Investigation:

Data Collection

Data Presentation
- Tabulation
- Diagrams
- Graphs

Descriptive Statistics
- Measures of Location
- Measures of Dispersion
- Measures of Skewness & Kurtosis

Inferential Statistics
- Estimation Hypothesis Testing
- Point estimate
- Interval estimate

Univariate analysis / Multivariate analysis Multiple variables. To adjust the confounder variables we do multivariate analysis

Frequency Distributions:

"A Picture is Worth a Thousand Words"

It's the distribution of your data, how much frequency is in the data.
what is the Shape and Range of the data .

Data distribution:
- Pattern of variability
- The center of a distribution
- The shapes
- The range

Has 2 Types :

01 Simple frequency distributions
We are not grouping them ,we are counting them and putting the frequency . **Small** sample size 30,40,50, ...

02 Grouped frequency distributions
We are putting them into class intervals into different classes .
Large sample size 100,300, ...
Can only be done for quantitative not qualitative

Simple Frequency Distribution: Consider as ungrounded data

- The number of times that score (for what we looking to) occurs
- Make a table with highest score at top and decreasing for every possible whole number
- For example: How many students in class get A+ in research exam
- N (total number of scores) always equals the sum of the frequency

$$\sum f = N$$

Tubular of and graphical presentation

Categorical / 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 and hair color*, etc.

Example: Grouped data help in large data

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

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

→

Class (Blood type)	Frequent, 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.

Example: (ungrouped)

- 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.

Remember this is a small sample size but if we have n=160 it is large so you have to group them and put in a class interval

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.

Tubular of and graphical presentation

Relative Frequency Distributions:

- 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
- It give us percentage if we multiply by 100
- Gives us a frame of reference

Class (pulse Rate)	Frequency (F)	**Relative frequency**
53	1	1/16= 0.0625
54	3	3/16= 0.1875
55	2	2/16= 0.1250
56	2	2/16= 0.1250
57	4	1/16= 0.2500

Class (pulse Rate)	Frequency (F)	**Relative frequency**
58	2	2/16= 0.1250
60	1	1/16= 0.0625
64	1	1/16= 0.0625
Total	n= 16	1.0000

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

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
- Advantage: answer questions such as :how many families have 6 children and above ? 8
- How many families have 5 and less children? 25-8=17 families

Example: (ungrouped) Convert to group by making them in interval

- No. of children in 25 families (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)

There're 3 families they have 9 children

No.of children	9	8	7	6	5	4	3	2	1	
No. Of families	3	+ 2	+ 2	1	4	4	3	3	3	∑f= 25 (No. of families)
rel f	0.12	0.08	0.08	0.04	0.16	0.16	0.12	0.12	0.12	∑ rel f=1.0
cf	3	3+2= 5	5+2= 7	8	12	16	19	22	25	
rel. cf	.12	.20	28	.32	.48	.64	.76	.88	1.0	Same as cf but summation for rel f

What's the percentage to found 8 or more children? 0.2*100= 20%

Tubular of and graphical presentation

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

Tubular of and graphical presentation

Steps for making a table

- Step 1 Find Minimum (9.1) & Maximum (15.7)
- Step 2 Calculate difference 15.7 – 9.1 = 6.6 Why? To decide the width of intervals
- 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
- General rules:
  - 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 they must be mutually exclusive
  - We can create a table with more than 1 variable called bi-variable or tri-variable.

Dummy Table

Hb (g/dL) (Width)	Tall marks	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		
Total		

Tall Marks Table

Hb (g/dL)	Tall marks	No. Of patients
9.0 – 9.9	I	1
10.0 – 10.9	III	3
11.0 – 11.9	IIIII 1	6
12.0 – 12.9	IIII IIII	10
13.0 – 13.9	IIII	5
14.0 – 14.9	III	3
15.0 – 15.9	II	2
Total	-	30

Table Frequency distribution of 30 adult male patients by Hb

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

Table Frequency distribution of adult patients by Hb and gender

Bi-variable (2 variables, Hb and Gender)

Hb (g/dL)	Gender		No. Of patients
	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

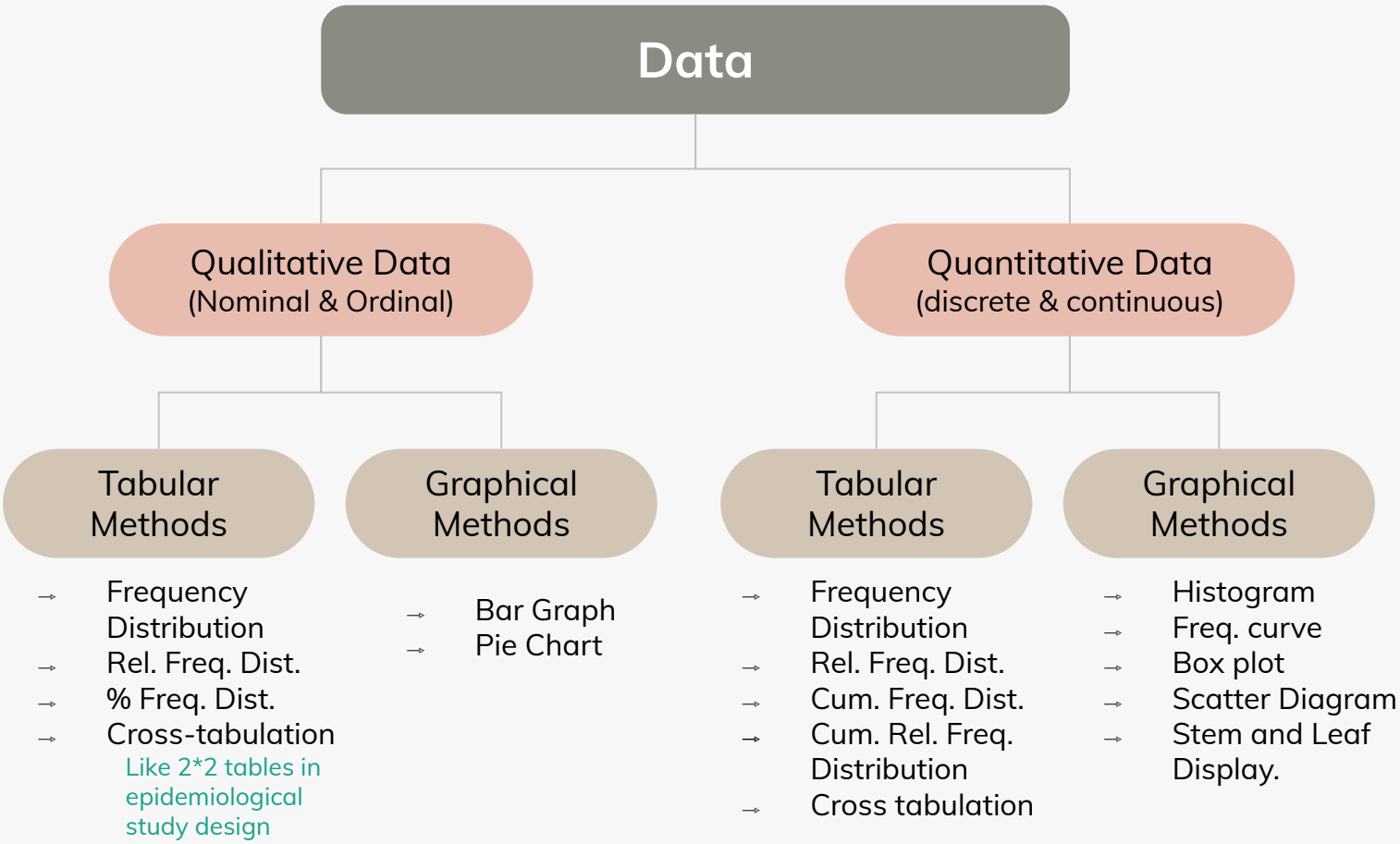
Tubular of and graphical presentation

Elements of a Table:

Ideal table should have

- **Number** → Table number for identification in a report
- **Title** (Title, place) → Describe the body of the table, variables.
- **Time period** → (What, how classified, where and when)
- **Column headings** → Variable name, No. , Percentages (%), etc.
- **Footnotes** → Describe some column/row headings, special cells, source, etc.

Tabular and Graphical Procedures:



Tubular of and graphical presentation

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) [Quantitative Data]

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

1- Histogram

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

Figure 1 Histogram of ages of 60 subjects

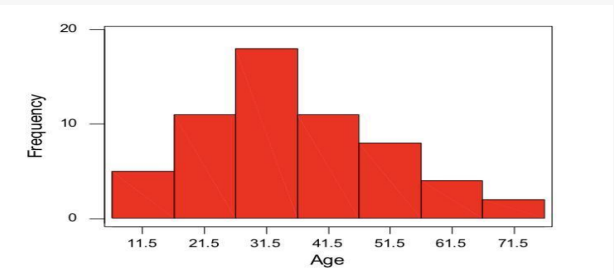
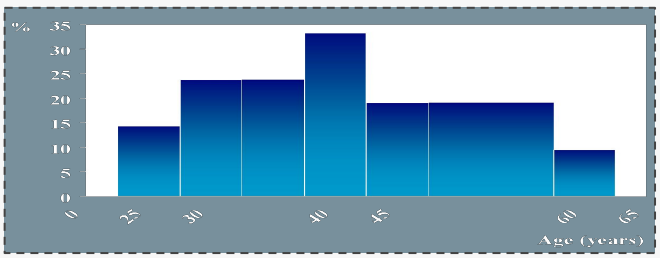


Figure (2): Distribution of 100 cholera patients at (place) , in (time) by age



Distribution of a group of cholera patients by age

Age (years)	Frequency	%
25-	3	14.3
30-	5	23.8
40-	7	33.3
45-	4	19.0
60-65	2	9.5
Total	21	100

2- Frequency polygon

Data of 60 patients (Age) [Quantitative Data]

The numbers are at the top of this page **(Example of data)**

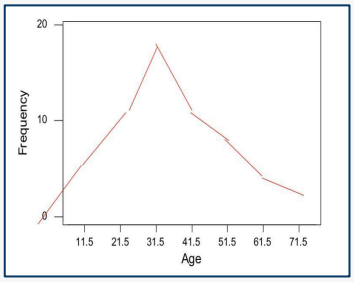
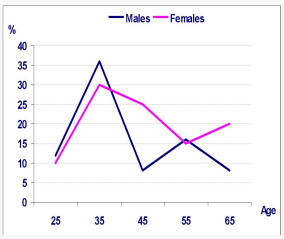
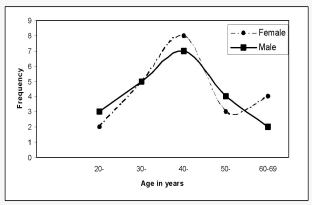


Figure (2): Distribution of 45 patients at (place) , in (time) by age and sex



Age	Sex		M-P
	M	F	
20-	(12%)	(10%)	25
30-	(36%)	(30%)	35
40-	(8%)	(25%)	45
50-	(16%)	(15%)	55
60-70	(8%)	(20%)	65

Frequency curve



- 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 it is a frequency curve not a polygon
- Advantage : - Minimum ,Maximum and the shape of the data can be seen

Tubular of and graphical presentation

3- Stem and leaf plot

Data of 60 patients (Age) [Quantitative Data]

The numbers are at the top of previous page **(Example of data)**

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

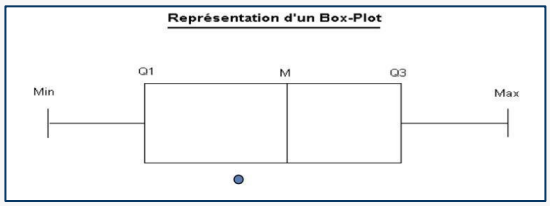
- How to read A data? there are 6 patients whose ages are 11,12,12,12,16,19 [Attach the two numerical value "stem" (1) with "leaf" (1, 2, 2,2, 6,9)]
- Advantage : - Whole data can be seen "Raw data" - Minimum ,Maximum and the shape of the data can be seen

Descriptive statistics report: [Quantitative Data]

A) Boxplot Box and whisker plot

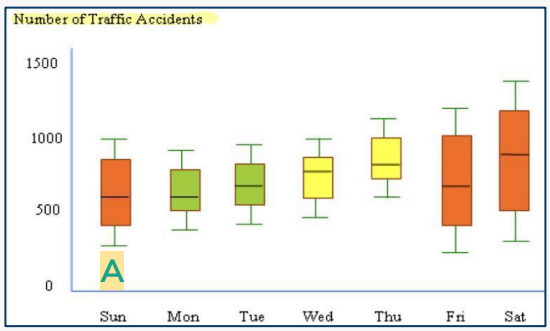
Advantages : 1- Gives all the descriptive statistics of the data :

- minimum score
- The skew of the distribution
- maximum score
- lower quartile
- upper quartile
- median
- mean



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

- 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

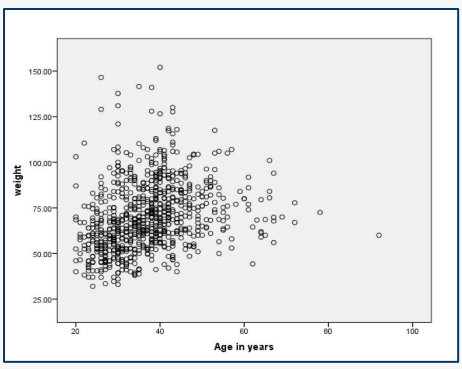


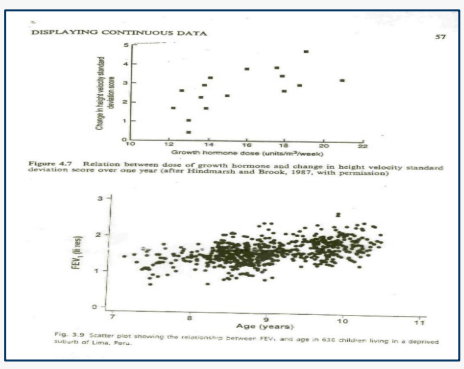
- The skew of the distribution
- Positive skew: mean > median & high-score whisker is longer
- Negative skew: mean < median & low-score whisker is longer

B) Scatter diagram

Used for : Assess the relationship between two quantitative variables
* each dot represent two variables (one variable on x-axis and the other one on y-axis) of one subject

pattern: positive linear relationship
As age increase → FEV increase



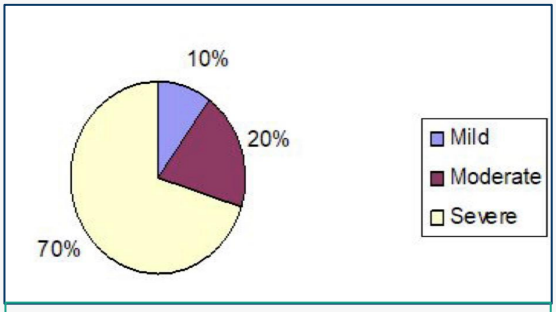


Tubular of and graphical presentation

Tabular and Graphical Procedures: [Qualitative Data]

A) Pie Chart

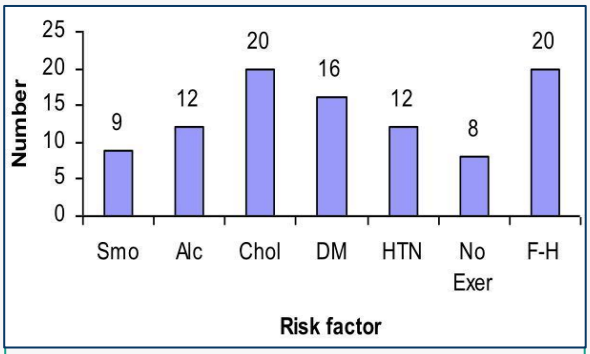
- Circular diagram – total -100%
- Divided into segments each representing a category
- Decide adjacent category
- The amount for each category is proportional to slice of the pie



The prevalence of different degree of Hypertension in the population

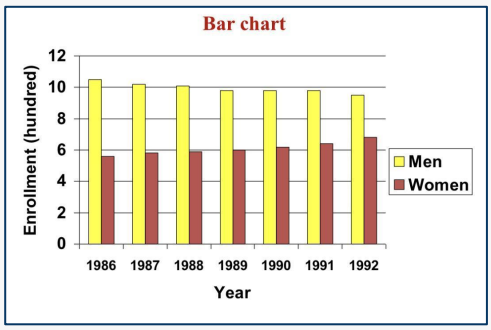
B) Bar Graphs

- 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
- Can use bar chart in 2 variable by using 2 ways : multiple bar chart , stocked bar chart



The distribution of risk factor among cases with Cardiovascular Diseases

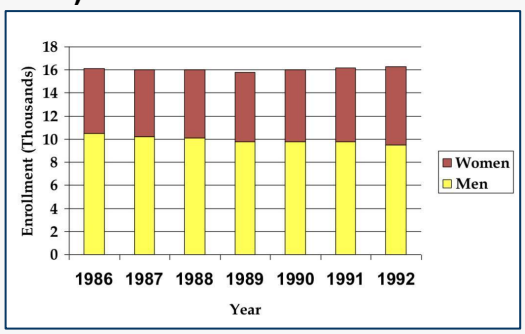
1) Multiple Bar Chart



HIV cases enrolment in USA by gender

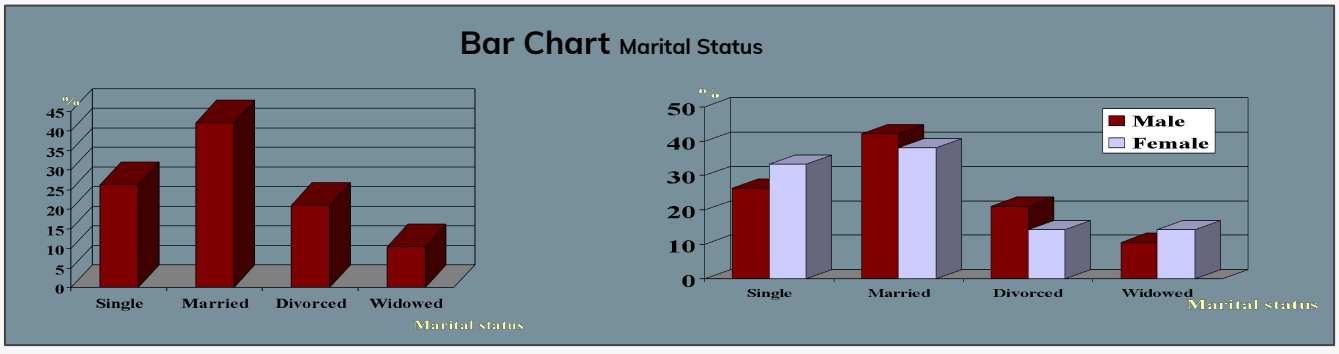
Multiple Bar chart (more than one rectangle)

2) Stocked bar chart



HIV cases enrolment in USA by gender

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



Tubular of and graphical presentation

By using charts and graphs

- **What do you lose?**
  - Ability to examine numeric detail offered by a table
  - Potentially the ability to see **additional** relationships within the data
  - Potentially **time**: often we get caught up in selecting colors and formatting charts when a simply formatted table is sufficient

- **What do you gain?**
  - Ability to **direct readers' attention** to one aspect of the evidence.
  - Ability to **reach readers** who might otherwise be apprehensive by the same data in a tabular format.
  - Ability to focus on **bigger picture** rather than perhaps minor technical details.

General rules for designing graphs

- A graph should have a self-explanatory legend Title: Table → on the top. graph → down
- 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)

Research Team

RESEARCH TEAM
438
Keep it simple, Keep it focused

Thank you for checking our work!

Leaders:

Shuaa Khdary Sarah AlQuwayz

Abdulrhman Alsuhaibany

Contact us:

Research4390@gmail.com