lec :5 **Bio statistic**

Graphical presentation of data

It is important to use the appropriate graph for each data type that clearly delivers the meaning.

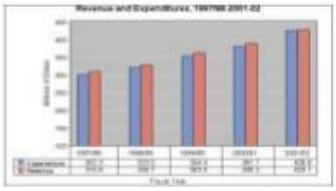
ADVANTAGES:

- ✓ They are attractive
- ✓ They give a bird's eye-view of the data
- ✓ They can be easily understood by common men
- ✓ They facilitate comparison of various characteristics
- ✓ The impression created by them are long lasting
- ✓ Theorems and results of statistics can be visualized using graphs

RULES FOR DRAWING GRAPHS AND DIAGRAMS:

- First choose the form of diagrams /graphs which is capable of representing the given set of data.
- Title- gives information of diagrams or graphs contain.
- Scale – selection of scale should be neither too small or too large. The scale should also specify the size of unit and what it represents. (eg: No. of persons in thousands).
- Neatness
- Attractive – different types of lines or shades, colours etc can be used to make the pictures more attractive.
- Originality – helps the observer to see the details with accuracy
- Simplicity –good diagram depends upon ease with which the observer can interpret it.
- Economy – cost and labour should be exercised drawing a diagram.

Limitations:



✓They are visual aids. They cannot be considered as alternatives for numerical data.

✓Though theories and results could be easily visualized by diagrams and graphs, mathematical rigour cannot be brought in

✓Diagrams and graphs are not accurate as tabular data. Only tabular data can be used for further analysis.

✓By diagrammatical and graphical misrepresentation observers can be misled easily. It is possible to create wrong impressions using diagrams and graphs.

Difference between Graphs and Diagrams:

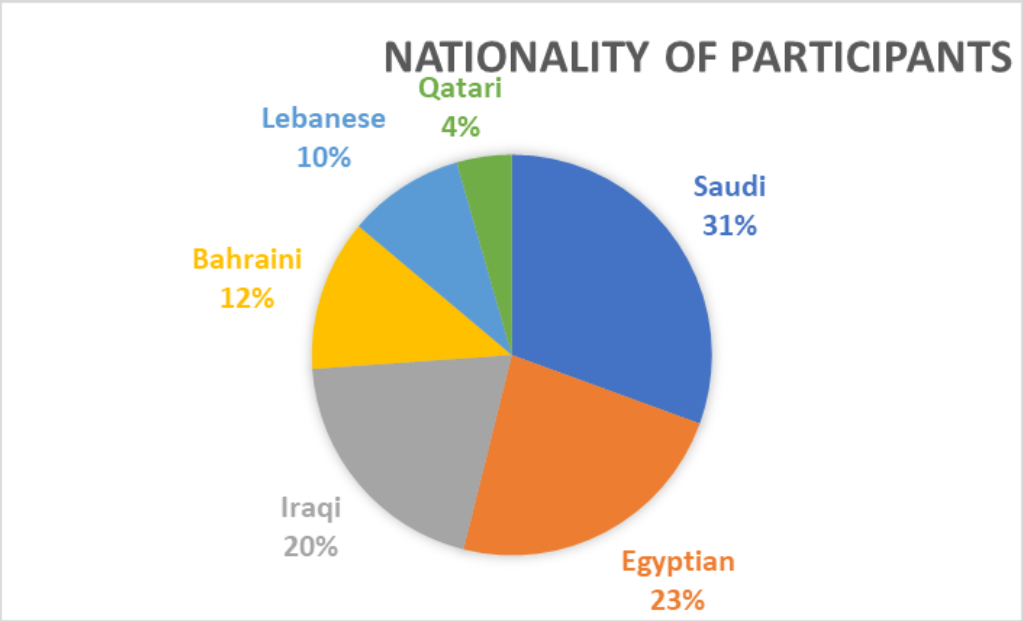
- To construct a graph, graph paper is generally used whereas a diagram is constructed on a plain paper.
- A graph represents mathematical relationship between two variables where as a diagram does not.
- Graphs are more appropriate than diagrams to represent frequency distributions and time series. Diagrams are not at all used for representing frequency distributions.
- Diagrams are more attractive to the eyes and as such are better suited for publicity and propaganda.
- Diagrams do not add anything to the meaning of the data and hence they are not helpful in analysis of data.
- Graphs are very much used by the statisticians and the research workers in their analysis.

We will illustrate each type of data variable with the appropriate graphs that can represent it.

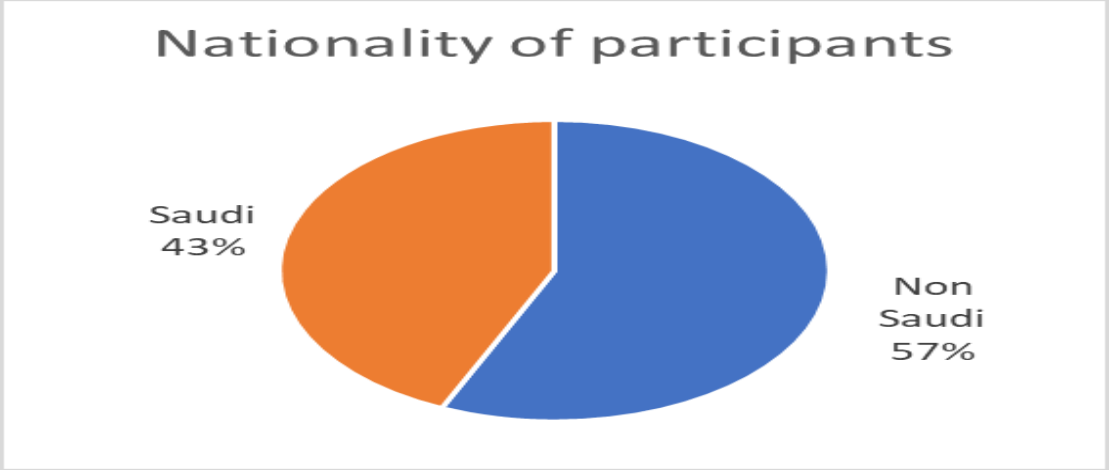
Nominal Variables

• Nominal variables: Pie chart

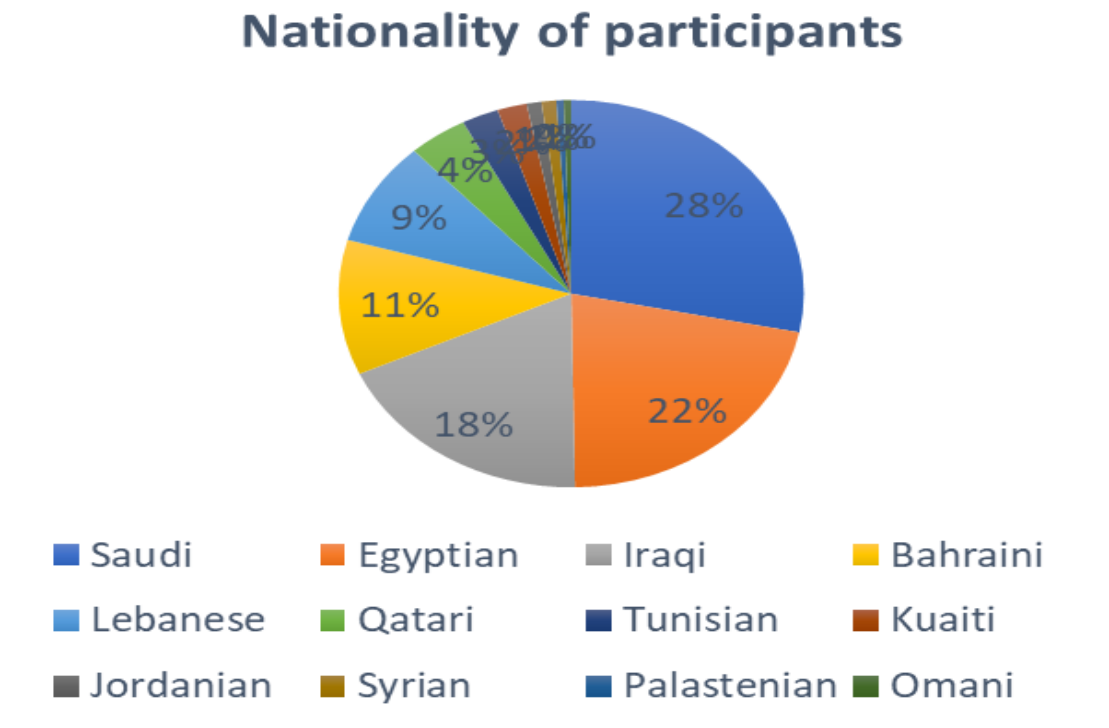
The pie (circle) represents 100% of the variable and is divided into sectors. The area of each sector represents the frequency of each category in the variable it represents as follows:



The pie chart is less commonly used in scientific papers due to its limitations. It can present only one variable.
If the categorical variable is binary (dichotomous), it will not be that informative and if the number of categories is large, the graph will not be that clear as follows:
A pie graph of a binary variable:



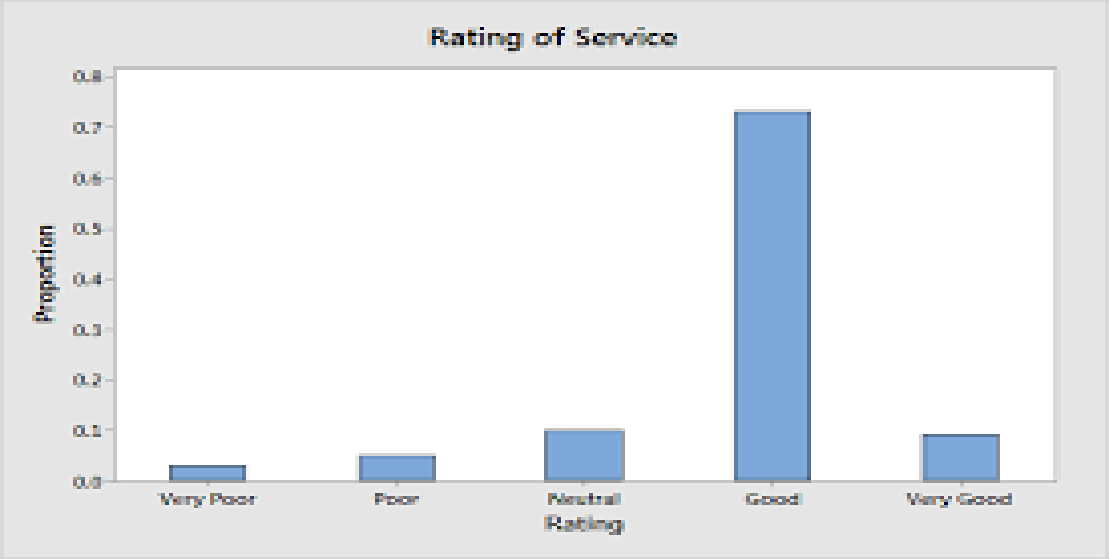
A pie graph of a variable with a large number of categories that is not clear



• Pie charts are not commonly used in scientific papers, they are usually used for presentations

- **Nominal variables: Bar graph**

Bar graphs are more commonly used to represent categorical variables. It can be vertical or horizontal graphs and can show the frequency or the percentage of each category.



As a general concept, we should use the graph the best demonstrates our data.

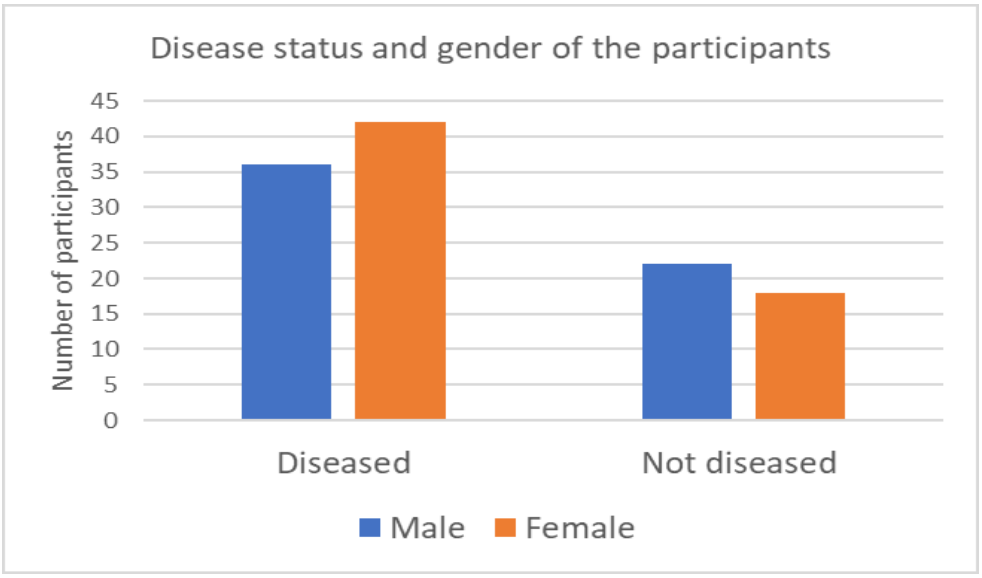
Two Categorical Variables

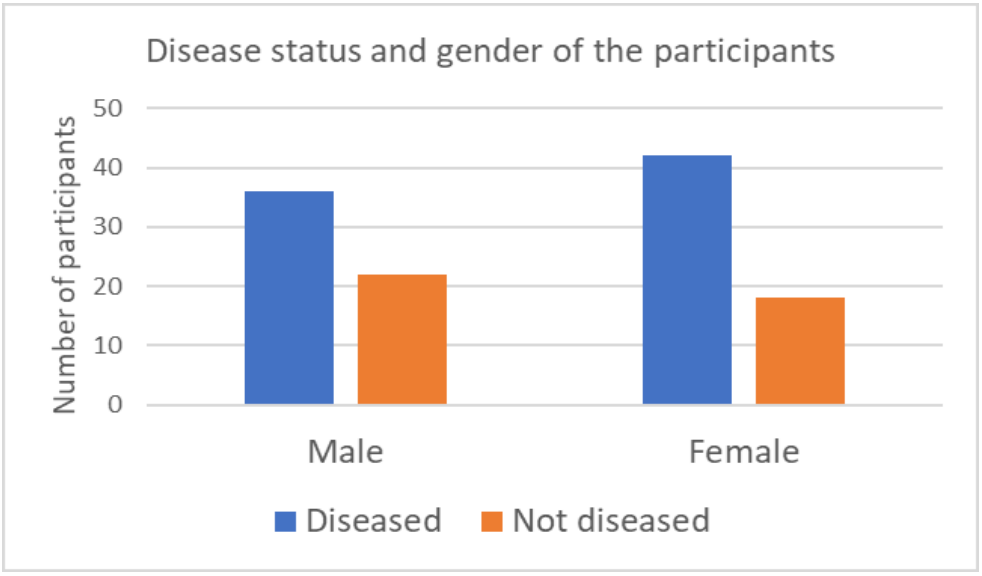
• Two categorical variables: Bar graphs

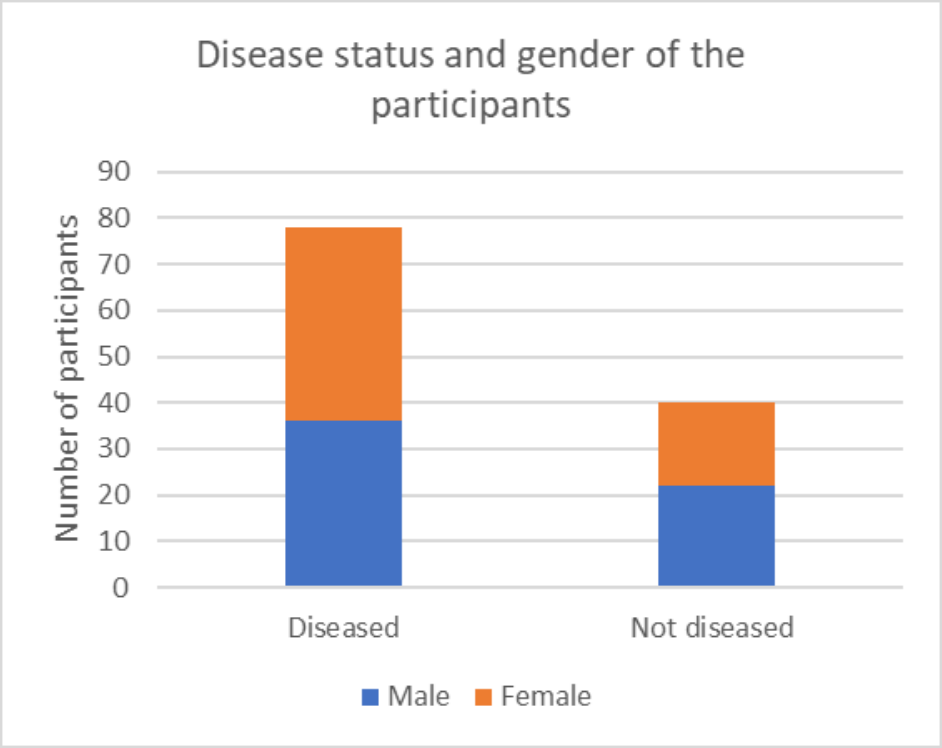
Presenting two categorical variables in the same chart can be done using bar graphs. Either **segmented bar charts** or **side-by-side bar charts** can be used.

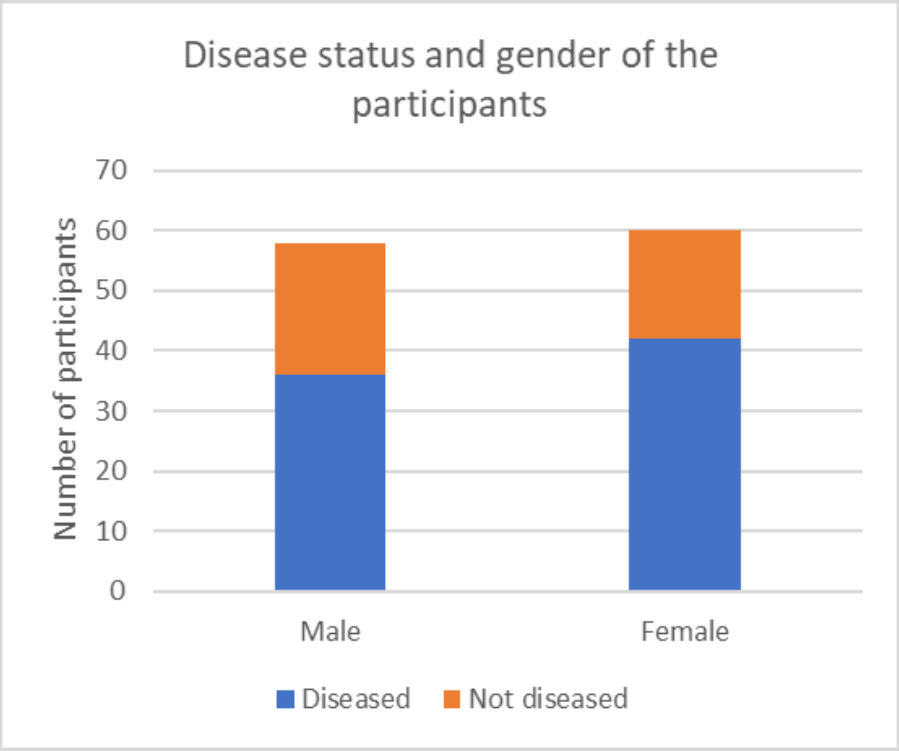
The following four graphs present the same data of the two variables, gender and disease status.

We can choose any of them based on which presents our results the best.


Disease status and gender of the participants
Number of participants
45
40
35
30
25
20
15
10
5
0
Diseased
Not diseased
Male
Female


Disease status and gender of the participants
Number of participants
50
40
30
20
10
0
Male
Female
Diseased
Not diseased





Numerical Variables

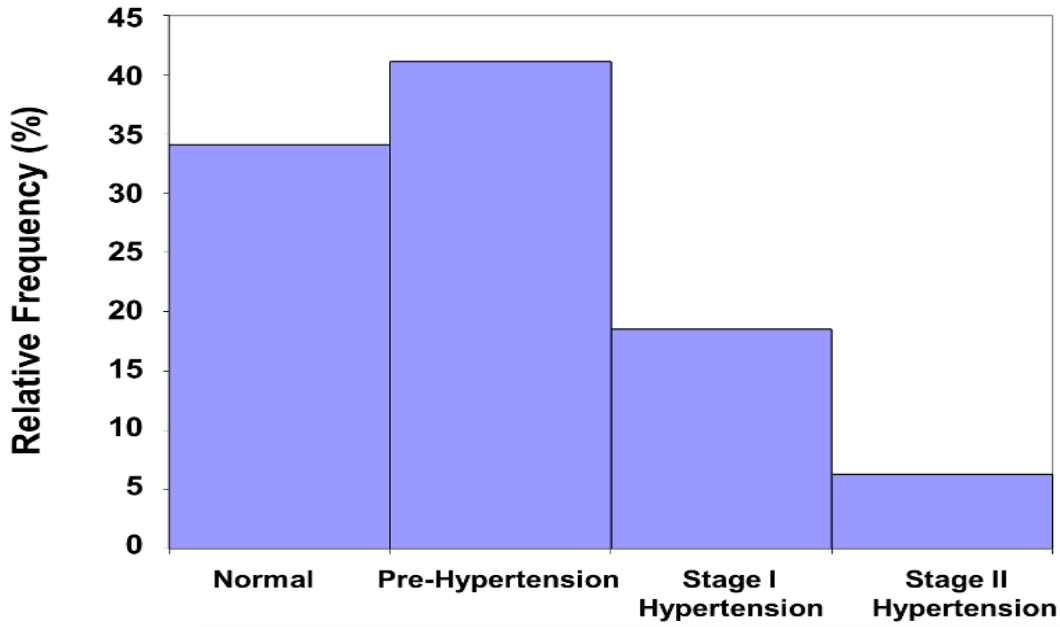
- **Numerical variables: Histogram**

It is similar to the bar chart, but there are no gaps between the bars as the variable is continuous.

The width of each bar of the histogram relates to a range of values for the variable, but in most cases, the width is kept the same.

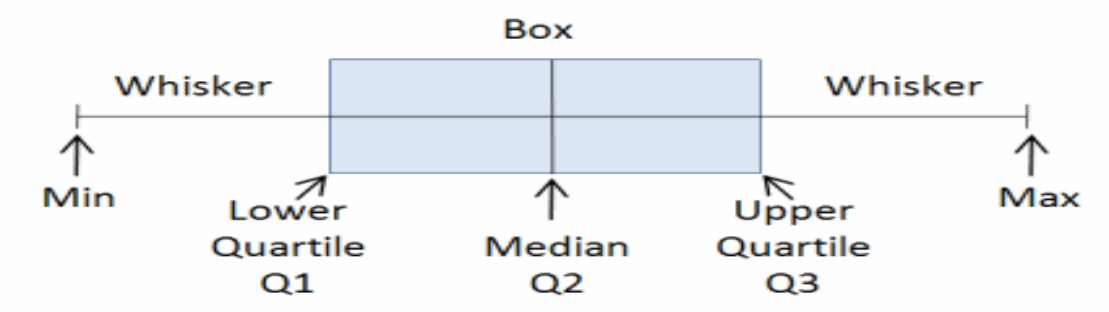
For example, a numerical variable as the birth weight in grams can be presented in the following groups (2000-2499, 2500-2999, 3000-3499, 3500-3999, and 4000-4500) with each group represented by a column.

The height of the column represents the frequency of cases in this group.



- **Numerical variables: Box plot**

The boxplot (also called Box and whisker plot) is used to summarize numerical variables based on the five-number summary.
Those five numbers are minimum, maximum, median, upper quartile, and lower
quartile.
median = horizontal line in the box
• upper quartile = top edge of the box
• Lower quartile = lower edge of the box
• maximum = top of 'whisker'
• minimum = bottom of 'whisker'

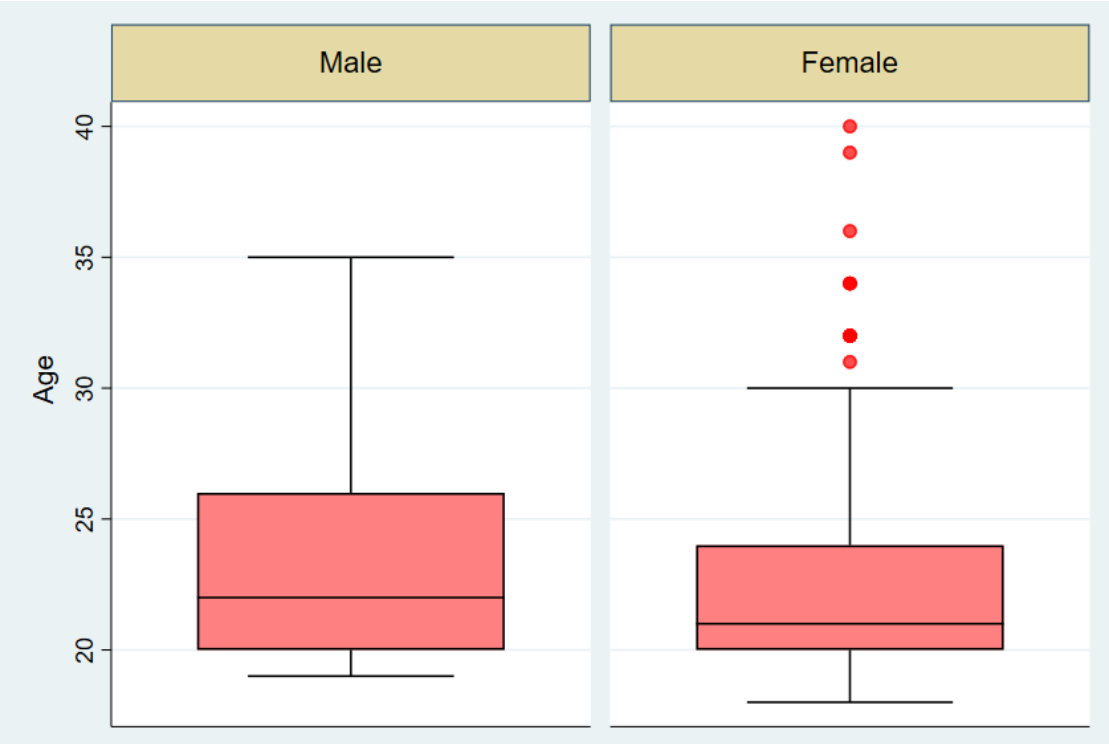


The whiskers are limited to outside 1.5 times the interquartile range above the upper quartile and below the lower quartile (Q1 - 1.5 * IQR or Q3 + 1.5 * IQR).

Boxplot is useful in showing the outliers (presented as dots outside the limits of the whiskers).

It is useful in comparing the same numeric variable across different groups as comparing a score between men and women.

The following graph shows a boxplot for men and a boxplot for women allowing us to compare the same variable (age) between the two groups.



Two Numerical Variables

• Two numerical variables: Scatter plot

- If we have two variables that are numerical (or ordinal), the relationship between them can be illustrated using a scatter diagram.
- It plots one variable against the other in a two-way diagram.

- One variable is represented on the horizontal axis and the other is plotted on the vertical axis with each dot representing one case.
The following scatter plot represents the relationship between weight and height.