**Course: Advanced Bioinformatics**

**Module title: Data Visualization**

**Module no. : 16**

Data visualization is viewed by many disciplines as a modern equivalent of visual communication. It is not owned by any one field, but rather finds interpretation across many (e.g. it is viewed as a modern branch of descriptive statistics by some, but also as a grounded theory development tool by others). It involves the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information".

A primary goal of data visualization is to communicate information clearly and efficiently to users via the statistical graphics, plots, information graphics, tables, and charts selected. Effective visualization helps users in analyzing and reasoning about data and evidence. It makes complex data more accessible, understandable and usable. Users may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic (i.e., showing comparisons or showing causality) follows the task. Tables are generally used where users will look-up a specific measure of a variable, while charts of various types are used to show patterns or relationships in the data for one or more variables.

Data visualization is both an art and a science. The rate at which data is generated has increased, driven by an increasingly information-based economy. Data created by internet activity and an expanding number of sensors in the environment, such as satellites and traffic cameras, are referred to as "Big Data". Processing, analyzing and communicating this data present a variety of ethical and analytical challenges for data visualization. The field of data science and practitioners called data scientists has emerged to help address this challenge.

Common inputs for data visualization include 3D cubes, distribution charts, curves, surfaces, link graphs, image frames and movies, parallel coordinates. For output Results we can have Pie charts, scatter plots, box plots, association rules, parallel coordinates, dendograms, temporal evolution etc.

What is need for Visualization? Why do we need it at all?

A famous quote: ”A picture is worth than a million words”. In current times we have huge amounts of information but we have limited display capacity for output devices.

The power of automatic calculations and the capabilities of human processing should be copied for data visualization. Human perception offers phenomenal abilities to extract structures from pictures.

**Data Visualization Methods**

Data can be

* Univariate
* Bivariate
* Multivariate

**Univariate Data**: Measurement of single quantitative variable. It can be represented using

* Histogram
* Pie Chart

**Bivariate Data:** Constitutes of paired samples of two quantitative variables. Variables are related to each other in one or the other way. It is represented using following methods

* Scatter plots
* Line graphs

**Multivariate Data:** Multi-dimensional representation of multivariate data. It is represented using

* Icon based methods
* Pixel based methods
* Dynamic parallel coordinate system

Some Visualization Tools

* Cn3D - uses MMDB-Entrez's structure database: http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml
* RasMol: http://www.umass.edu/microbio/rasmol/
* Protein Explorer: http://www.umass.edu/microbio/rasmol/rotating.htm
* World Index of Molecular Visualization: http://molvis.sdsc.edu/visres/index.html