An Anatomy of Data Visualization

Abhishek Kaushik† and Sudhanshu Naithani††,
Kiel University of Applied Sciences† Kurukshetra University††

Summary
As data is being generated each and every time in the world, the importance of data mining and visualization will always be on increase. Mining helps to extract significant insight from large volume of data. After that we need to present that data in such a way so that it can be understood by everyone and for that visualization is used. Most common way to visualize data is chart and table. Visualization is playing important role in decision making process for industry. Visualization makes better utilization of human eyes to assist his brain so that datasets can be analyzed and visual presentation can be prepared. Visualization and Data Mining works as complement for each other. Here in this paper we present anatomy of Visualization process.

Key words:
Information Visualization, Scientific Visualization, Decision Making, Graph, Chart, Xmdv tool.

1. Introduction
In simple worlds Visualization is a process to form a picture in order to make it easily imaginable and understandable for other people. With Visualization, process of Data Mining and Human Computer Interaction provides better results for visual data analysis. Initially visualization was of two types - Information Visualization and Scientific Visualization. Scientific Visualization used to work for scientific data with spatial component while Information Visualization used to work for abstract and non-spatial data [12]. Presently visualization is facing problems like mapping, dimensionality, and design tradeoff [13]. Visualization helps to understand patterns, trends and relationship between different components in a dataset. In words of David McCandless [32] (author, data journalist, and information designer) :- “By visualizing information, we turn it into a landscape that you can explore with your eyes, a sort of information map. And when you’re lost in Information, an information map is kind of useful.”

Figure 1 shows the general steps in the process of Visualization. For visualization data is collected from all the available sources. Then possible aggregate meaning is generated. After that data is analyzed. After it, graphical interpretation of analyzed data takes place. And at last step user interacts with graphical interpretation.

Here are basics to generate best possible visualization [27] for any given data:

a) Try to understand size and cardinality of the data given.
b) Determine kind of information which is to communicate.
c) Process visual information according to targeted audience.
d) Use the visual portraying best and easiest form of given data for audience.

2. Classes of Data Visualization Techniques
The most common classes [2] of data visualization techniques are:

a) Describing Data
b) Viewing Relationship
c) Picturing Data (Icons, Glyphs, Color Coding)
d) Temporal Visualization
e) Spatial Visualization
f) Spatio-Temporal Visualization

Class (a) tells about the dataset. Class (b) describes relationship between observations and between variables. Class (c) maps data items into easily recognizable shapes. Class (d) describes visualization of temporal data which changes over time. Line graph is most suitable in this case. Class (e) describes spatial datasets which come from various domains that relate data to a certain landscape [2].
Map is utilized for this type of data. Class (f) has both kinds of properties i.e. spatial and temporal such as analysis of biomedical data.

3. State of the Art


4. Methods of Visualization (with examples)

4.1 Arc Diagram

Arc diagram is usually used to visualize complex data within string such as text, music, compile code. In structure of string there are repetitions of sub-string most of the time, which is a good thing as point of view of visualization because these repetitions can be used as prediction units for the visualization process. For example in any given article there will be repetition of words and phrases. Martin Wattenberg [6] described arc diagram visualization to process string by using pattern matching algorithm to find repeated substring and further representing them visually as translucent arcs. Most significant utilization of arc diagram is in the field of music which is to reveal structure in compositions of music [6]. Other utilization fields for arc diagrams are web pages, compiled codes, and nucleotide sequence from DNA etc. In future other pattern matching algorithms can be used for arc diagrams to get unknown insights of this method.

4.2 Flow Maps

As name suggests flow maps show the flow of any process i.e. how particular process is flowing. For example when people migrate from one country to another, a flow map can show this very easily. Doantam Phan [9] presented a method to generate flow map using hierarchical clustering which is inspired by graph layout algorithm. In a hand written flow map intelligent distortion of positions, merging of edges that share destination and intelligent edge routing are most common characteristics [9]. To achieve intelligent distortion, Doantam used layout adjustment algorithm. For merging edges and intelligent edge routing, hierarchical clustering is used. To implement this system an algorithm was used with following steps:-

a) Layout Adjustment  
b) Primary Hierarchical Clustering  
c) Rooted Hierarchical Clustering  
d) Spatial Layout  
e) Edge Routing  
f) Multiple –Layer Issues (when there are multiple layers in the system)

4.3 Graph Analytics

It is common and interesting topic of visualization and analytics. The main aim of graph analytics research is to meet real life challenges. It follows technology-application pair i.e. success is measured by application and not by algorithmic criteria [11]. In the applications it always turn lesson learned into lesson applied. The main applications (real life challenges) of graph analytics are listed following [11]:-

a) Electric-Power-Grid Analytics  
b) Social-Network and Citation Analytics  
c) Text and Document Analytics  
d) Knowledge Domain Analytics

4.4 Voronoi Treemaps

For visualization of attributed hierarchical data Treemaps are best method. Treemaps normally has problem of rectangular shapes limitation, which is removed by using Voronoi Treemaps. It also enables arbitrary shape visualization [37]. Michael Balzer [37] presented Treemaps based on the subdivision in arbitrary polygons which eliminate rectangle limitation because of recursive pattern. In this system following steps are repeated again and again:-

a) According to top hierarchy level polygonal subdivision of display area is created.  
b) output is a set of polygons representing the nodes of the top hierarchy level
4.5 Geometric Projection

It is a technique used for multidimensional multivariate visualization. It can map Cartesian plane as well as arbitrary space [13]. This technique is good to detect outliers and handle large datasets.

Scatterplot is used to show joint variation of 2 data items at x-y axes of Cartesian coordinates. It supports grouping. In case of 3 or more measures a matrix named scatter plot matrix is produced which is a series of scatter plots to display possible pairing of measures that are assigned to visualization [27].

Parallel Coordinates technique is used where attributes are represented by parallel vertical axes linearly scaled within their data range [13]. Coordinates also utilized to study correlations among attributes by locating points of intersection [13]. Here limited space is available for a parallel axis. It has mainly two types i.e. circular and hierarchical.

4.6 Pixel-Oriented Technique

It is also used to visualize multivariate data where an attribute is represented by colored pixel. In n-dimensional dataset to represent a data item n colored pixel will be used.

Recursive Pattern can influence data arrangements by using generic recursive process. It is a query independent. Pixel Bar Chart does not aggregate data values but presents them directly. These are derived from regular bar charts. Multi-pixel bar charts are used for high dimensional data [13].

4.7 Hierarchical Display

These techniques are mainly concerned about hierarchical data where data space is subdivided first and then subspaces are presented in hierarchical way. Dimensional Stacking which is also known as general logic diagrams is a technique which is result of modification in hierarchical axis. It separates the data space into 2D stacked subspaces [13].

Treemap partitions the screen into several regions on behalf of value of attribute by using hierarchical partitioning.

4.8 Iconography

It maps multidimensional data item to an icon and also known as icon-based techniques. The visual features vary depending on the data attribute values [13].

Chernoff Faces which is most popular technique of Iconography can visualize data items in a limit. It maps dimensional positions of a face and its properties like mouth, eyes and nose etc [13].

Star Glyph is one of the many variants of glyph family and is most widely used. Here star glyphs are used to present data items. It is not suitable when no of data items are on increase. It can also be used to encode additional information by combining with other glyphs [13].

Shape Coding uses very small array pixels to visualize data. An array is used to represent one item of data.
4.9 Chart and Graph

These are most common, widely used and easily understandable ways to visualize information for the audience. Here are some of them:-

**Line Graphs** is also called line chart and shows relationship of a variable to other variable. It is used for trend tracking comparison of items within same period of time [27].

**Bar Charts** compare qualities of two or more groups. To show values bars, which can be either vertical or horizontal, are used. When there are large no of bars and same time bars are close together, it is not possible to detect differences between bars. That’s why different colors are used for representing bars [27]. It works better when bars are having different range.

**Pie Charts** are subjects of discussion because their angles and areas cannot be easily interpreted by eyes. It is very useful when additional information (e.g percentage) is provided [27] and not ideal for developing dashboards for small screens.

5. Applications

Here are some of the applications of data visualization:-

1. **Business Decision Making Process**: - There are applications of visualization in the business decision making process. It enables the top level management to examine vast amount of data, find current markets trends, take the decision and make strategic changes if required. Common forms of visualization used in business decision making are basic charts, status indicators, scatter graphs, bubble charts, spark line charts, geographical maps, tree maps, Pareto charts etc [29].

2. **Other Areas Related to Decision Making Process**: -
   a) **Uncertainty Visualization** - Uncertainty in the information is capable of influencing decision making. There are lots of techniques for uncertainty visualization.
   b) **Risk Visualization** - Some problems also have risk in order to make decision to solve them. According to Lipkus & Hollands, (1999), users might wish to extract the following information [25] regarding risk:
      1) Risk magnitude (i.e., how large or small the risk is);
      2) Relative risk (i.e., comparing the magnitude of two risks);
3) Cumulative risk (i.e., observing trends over time);
4) Uncertainty (e.g., estimating amount of uncertainty and variability or range of scores);
5) Interactions among risk factors.

Risk visualization uses static diagrams mostly.

c) Sensitivity analysis visualization- It uses graphs, charts, surface etc. There are very few techniques which can be applied to sensitivity analysis. Tornado diagram is a graphical approach which displays outcomes of local sensitivity analysis [25].

3. Manipulate and Interact Directly with Data: -
Visualization enables users to directly interact and manipulate data unlike 1D table and chart which can only be viewed [38]. Real time visualization helps to figure out reasons for low performs of organization and can compare it with its rivals. And then most helpful changes can be made.

4. Foster a New Business Language: -
Visualization tells all the things through data. Performance indicator does not tell about growing and shrinking category of business and the reasons behind it [38]. While visualization shows performance category wise and enable user to find reasons for it by further digging the data.

5. Identify and Act on Emerging Trends Faster: -
Companies gather lot of data about their user by surveys, data mining and opinion analysis. Visualization is able to track [38] emerging trends and new opportunities for business related to those trends.

6. Xmdv Tool

Xmdv is one of the popular open source tools which are used for visualization process. It supports mainly 5 methods [39] listed below-
1. Scatterplots
2. Star Glyphs
3. Parallel Coordinates
4. Dimensional Stacking
5. Pixel-oriented Display

Applications areas of Xmdv tool includes remote sensing, financial, geochemical, census, and simulation data [39].

Fig. 8 Snapshot of Xmdv Tool.

References
[4] Zhao Kaidi, “Data Visualization”.
[16] Sandro Boccuzzo and Harald C. Gall, “Software Visualization with Audio Supported Cognitive Glyphs”.
[17] Pak Chung Wong, Paul Whitney and Jim Thomas “Visualizing Association Rules for Text Mining”.
[33] Fernanda B. Viégas and Martin Wattenberg, “Artistic Data Visualization Beyond Visual Analytics”.