Assignment 2

Submitted to: Dr. Ankur Jain

Ajay Singh
20BAI10300
CSA3004
C11+C12..

1. List out the various features of good data graphics. Apply them to the following Data snippet.

Ans –

various features of good data graphics :

Content –

In certain situations, data visualization in the traditional sense is not the most effective way to display information. Something shorter and clearer then detailed reports is needed. A simple table can be used to present the most important points as efficiently as possible. With dear heading and minimal rows and columns, the reader will have all the information they need at a glance.

Design –

Adding design elements to Data Visualization can help enhance the message in subtle but important ways. It can not only help catch reader's eye but also add to the professional look and feel of the presentation. Design can also bring visual cues that add relevance for the readers.

Construction –

Line: Line can be used for a wider range of purposes, connecting context, creating patterns, etc.

Colour: Colour is used to generate emotions, define importance, create visual interest and unify branching.

Size: In graphic design, size is used to convey importance, attract attention and create contrast.

Space: It is vital part of any good graphics design. It can be used to separate or group information. Using it effectively, give the eye a rest and defines importance.

Context –

The more points that are represented in a Data Visualization, the more compelling the overall picture can become. Trends look more established and outliers look more exceptional. It is important to present the data clearly. A straightforward message shouldn't be sacrificed simply to display an impressive amount of data.

Applying on snippet:
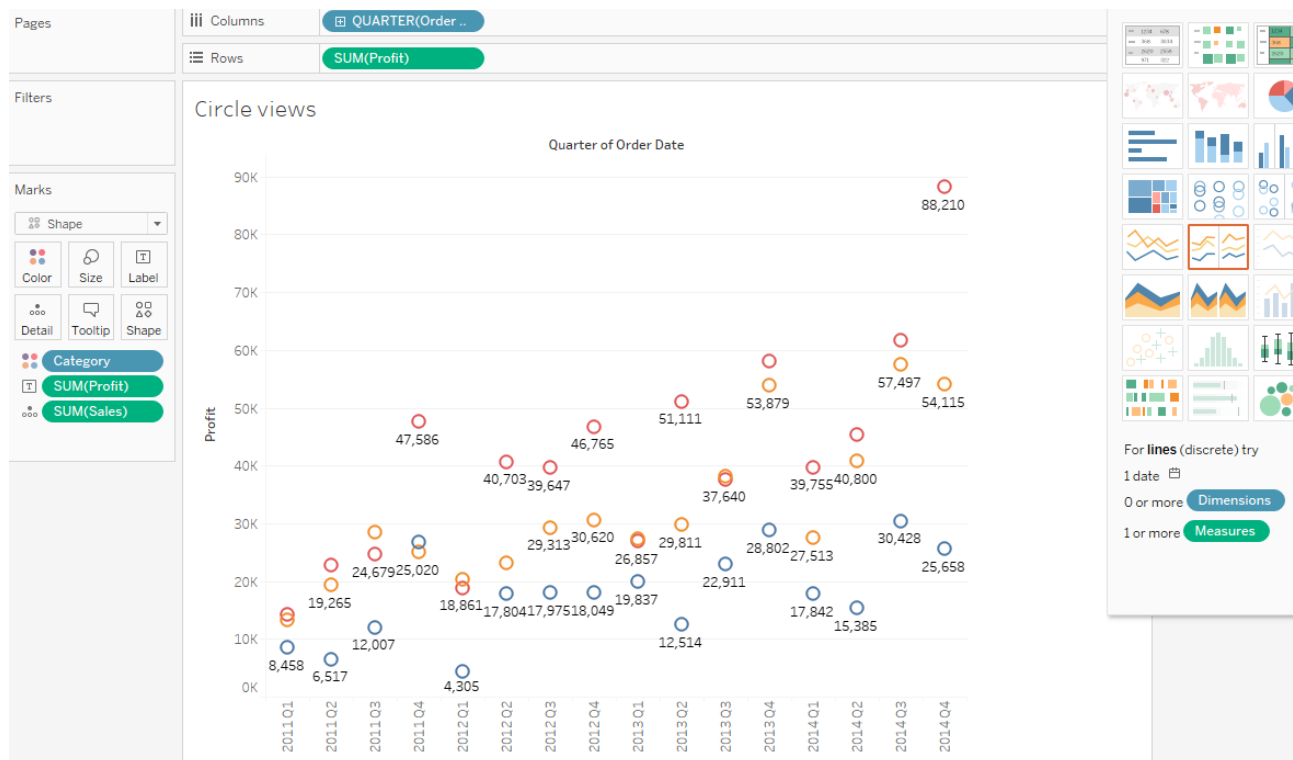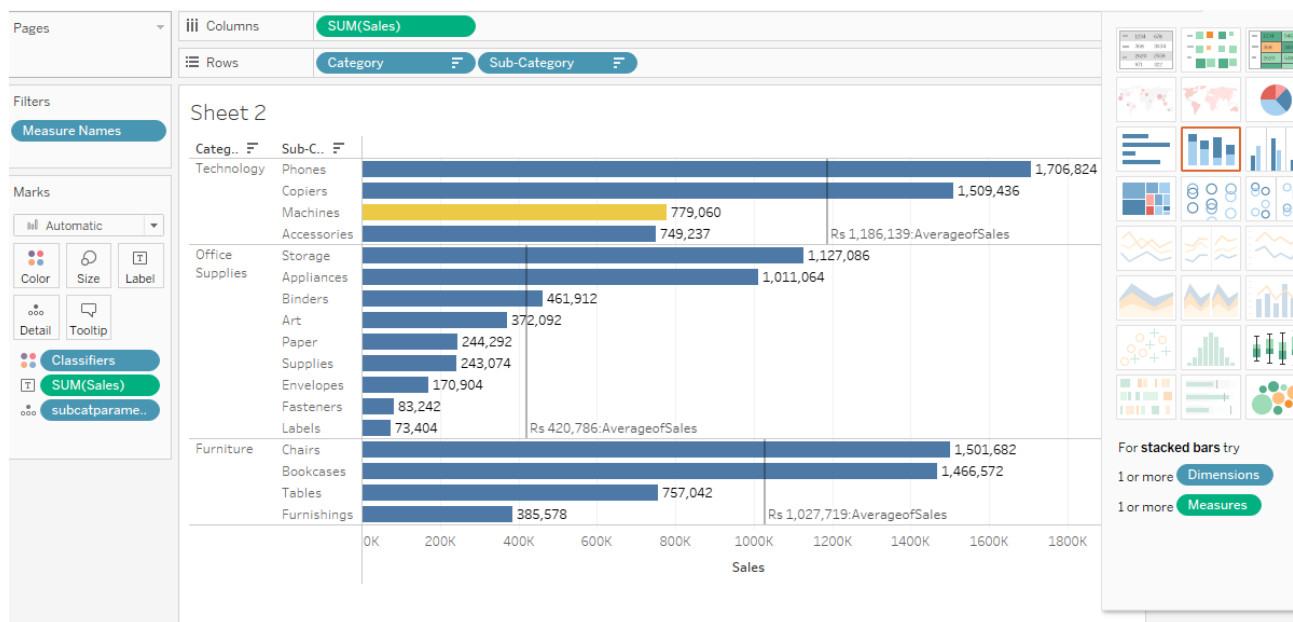
2. Compare presentation graphics and exploratory graphics.

Presenting data involves deciding what information you want to convey and drawing a display appropriate for the content and join the intended audience. You have to think about how the plot might

be perceived and whether, it will be understood as you wish. Plots which are common in are kind of publication may be unfamiliar to the readers of the another. There may be available in print for a very long time, so great care should be taken in preparing the most appropriate display.

Exploring data is much more individual matter, using graphics to find information and generate ideas. Many displays may be drawn, they can be changed at will or discarded and new versions prepared, so generally no one plot is especially important, and they have a shorter span. Clearly principles and guidelines for good presentation graphics have a role to play in exploratory graphics but personal taste and individual working style also play important roles. The same data may be presented in many alternate way and tastes and customs differ as to what is regarded as a good presentation graphic.

3. Write a short note on the trellis plot.

A trellis plot (also known as a lattice plot or panel plot) is a group of smaller plots arranged in a grid. Each small plot represents a different condition or item, and all small plots share the same scales. Trellis plots make complex or high-dimensional data easier to interpret due to the common scales of small plots aligned in a grid, with each small plot containing a single difference compared to the others. Trellis plots are based on the idea of conditioning on the values taken on by one or more of the variables in a data set. In the case of a categorical variable, this means carrying out the same plot for the data subsets corresponding to each of the levels of that variable. Trellis plots, Trellis charts or Trellis graphs, are a means of graphing multivariate data, as an array of M X N panels. As a simple example, two "primary variables" might be chosen to represent a common set of axes (repeated in each panel), against which all combinations of two categorical variables (termed "conditioning variables") are plotted. Each combination of categorical values is plotted as a separate panel. In Origin terms, when your data includes a column of categorical data, or multiple columns characterizing some sort of nested sub-grouping of data, you might gain from plotting the data as a Trellis

There are various ways to represent trellis plot
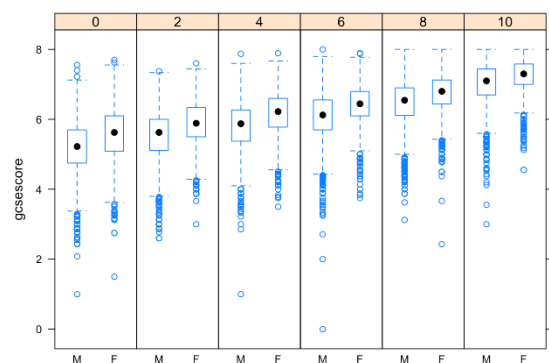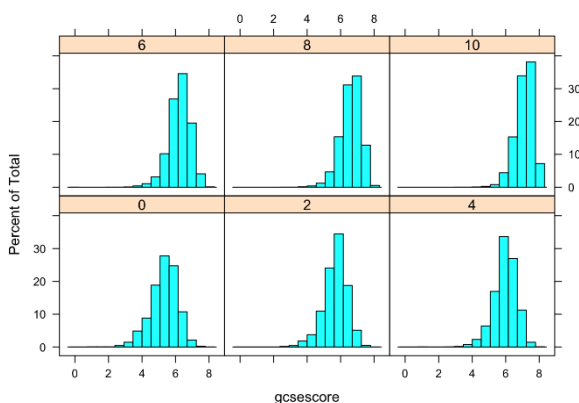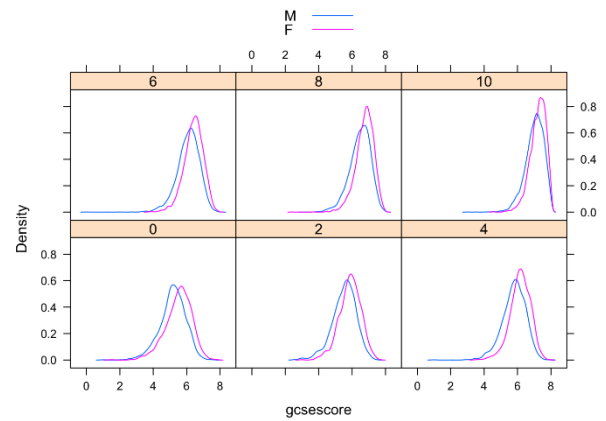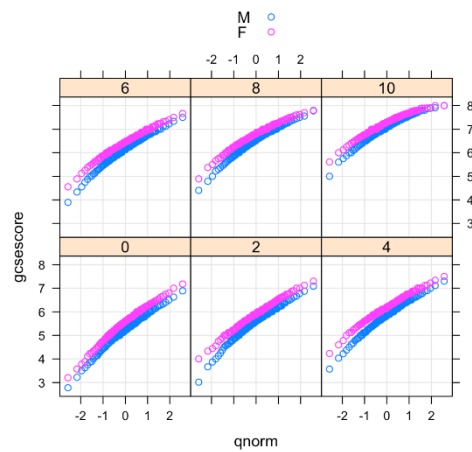
4. Design a graphical layout to show the relationship between the X and Y parameter and apply the features of the complete plot.

Complete Plot consisting of

• axis

• labels

• data symbols or lines

• Entire plot with a single command

• Sensible locations

• Sensible Default choices

Circle views

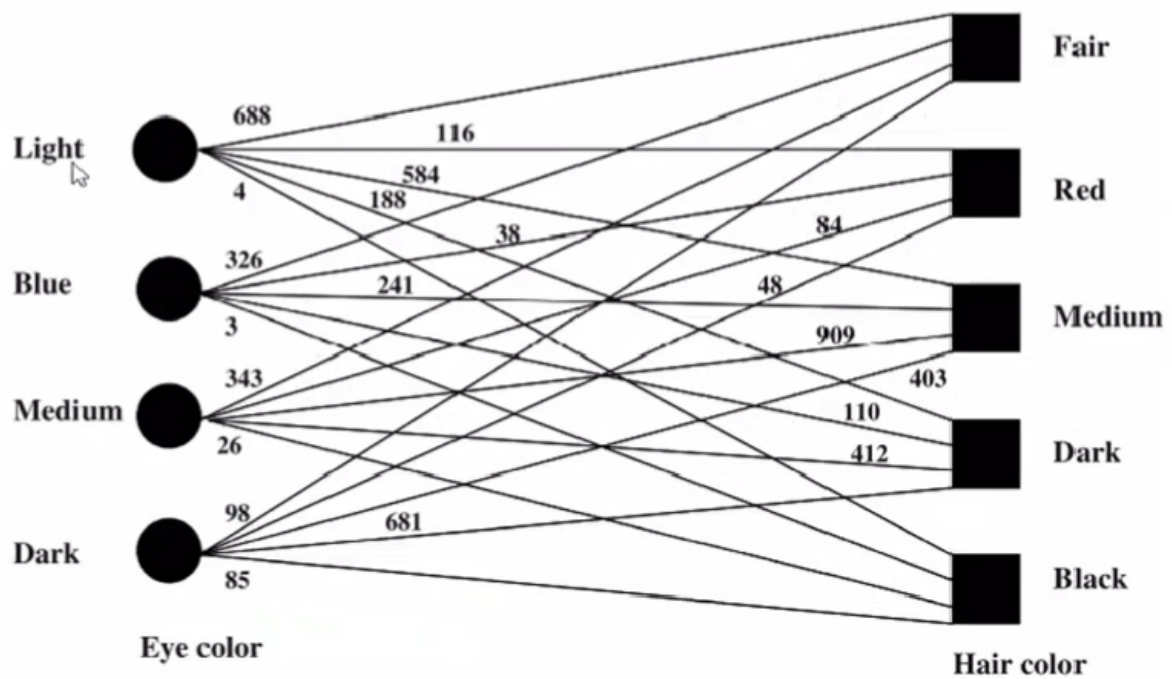Category
- Furniture
- Office Supplies
- Technology
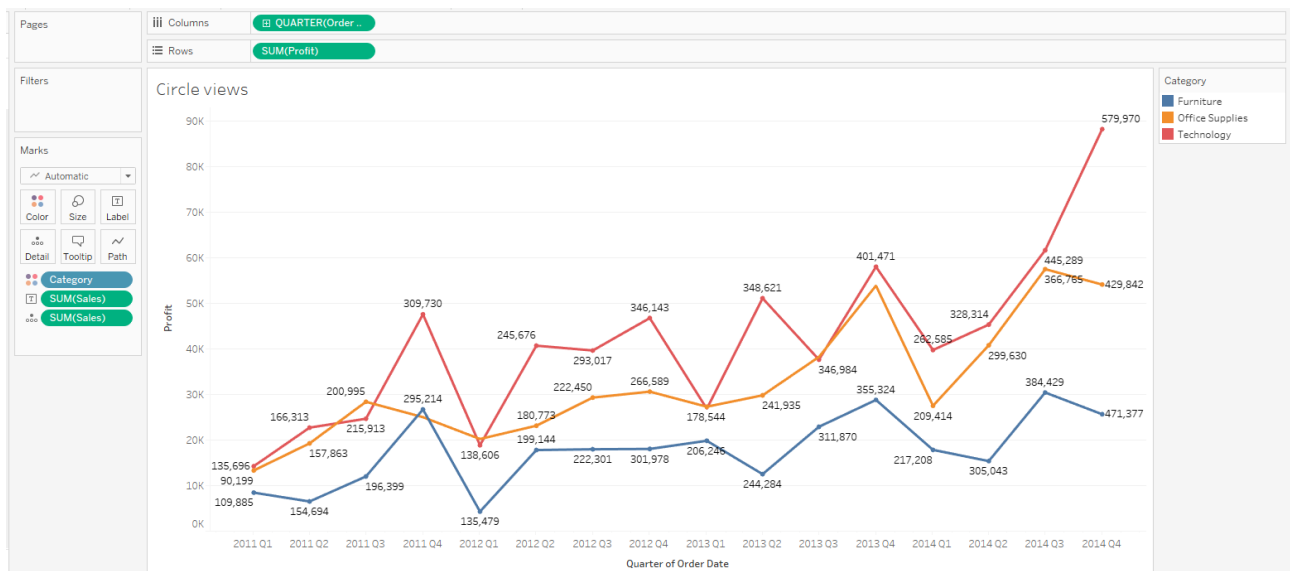
Here, we have take data of a shop to design a graphical layout. we have x axis parameter as order date and y axis we have profit . this parameter has relationship as of profit gained with respect to the time and date.

5. Design a bipartite graph for the following contingency table. Also, describe the graphical elements.

Table 4.1. Contingency table of 5387 school children form Caithness, Scotland, classified according to two categorical variables, hair and eye color (Fisher, 1938)

Eye color	Hair color				
	Fair	Red	Medium	Dark	Black
Light	688	116	584	188	4
Blue	326	38	241	110	3
Medium	343	84	909	412	26
Dark	98	48	403	681	85

Figure 4.2. Weighted graph representation of a contingency table

The examples deal with a contingency table, where nodes correspond to the categories and the weighted links represent the frequencies.

Graphical Elements -

Two independent sets to represent the data values of vertices.
The bipartite graph is a weighted graph which consists of vertices, edges and weights which are its graphical elements.
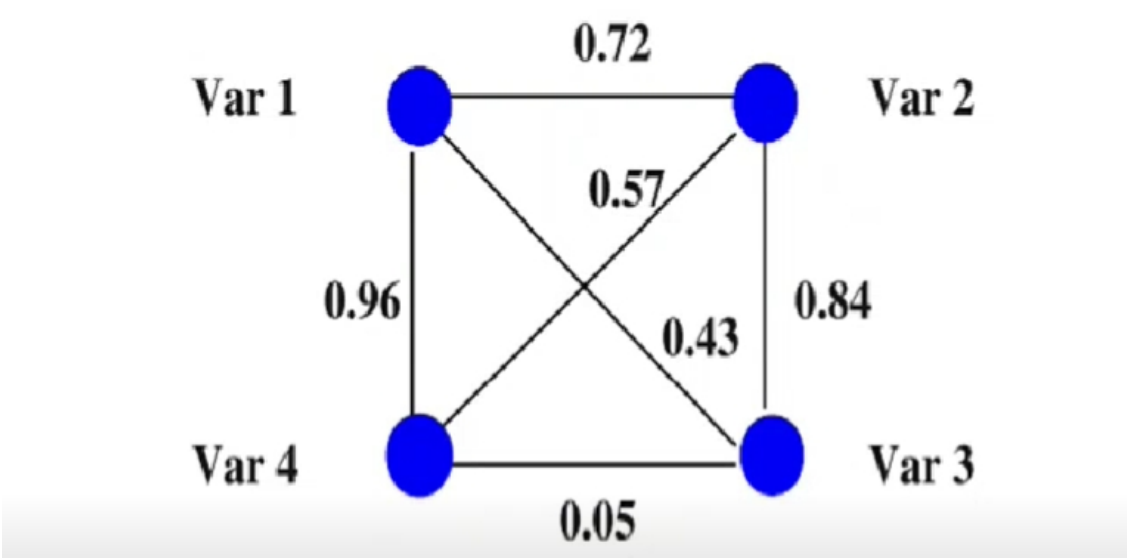
For eg - Some web pages on the Web form a directed graph, also called the web graph, where each web page is a vertex, and each hyperlink is an edge indicating from a source page to a destination page. Cluster analysis on the web graph can acknowledge communities, discover hubs and authoritative web pages, and identify web spams.

6. Design a graph for the following correlation matrix. Also, describe the graphical elements.

Correlation matrix of four table:

	Var 1	Var 2	Var 3	Var 4
Var 1	1.00			
Var 2	0.72	1.00		
Var 3	0.43	0.84	1.00	
Var 4	0.96	0.57	0.05	1.00

correlation matrix using Weighted Graph

OR

chart data

Year	Indicator	Country	Country Code	Weighted Score
2018	1 Connectivity	Denmark	DK	19.6205
2018	1 Connectivity	Sweden	SE	18.9963
2018	1 Connectivity	Finland	FI	16.5132
2018	1 Connectivity	Netherlands	NL	20.2830
2018	1 Connectivity	Luxembourg	LU	20.0132
2018	1 Connectivity	Ireland	IE	16.2796
2018	1 Connectivity	United Kingdom	UK	17.2119
2018	1 Connectivity	Belgium	BE	18.7821
2018	1 Connectivity	Estonia	EE	16.0279
2018	1 Connectivity	Spain	ES	16.1769

| Columns | Indicator |
| Rows | Indicator |

Sheet 1

Indicator

Indicator	1 Connectivity	2 Human Capital	3 Use of Intern..	4 Integration o..	5 Digital Public..
1 Connectivity	Abc				
2 Human Capital		Abc			
3 Use of Internet			Abc		
4 Integration of Digital Technology				Abc	
5 Digital Public Services					Abc

| chart data | | chart data1 |

Join ✕

	Inner	Left	Right	Full Outer

Data Source		chart data1
Country	=	Country (chart data1)
Year	=	Year (chart data1)
Add new join clause		

Sort fie

| # chart data Year	Abc chart Indi						⊕ chart data1 Country
2018	1 Co						Sweden
2018	1 Connectivity	Finland	FI	16.5132	2018	5 Digital Public S...	Finland
2018	1 Connectivity	Finland	FI	16.5132	2018	4 Integration of ...	Finland
2018	1 Connectivity	Finland	FI	16.5132	2018	3 Use of Internet	Finland

CORR(

{ FIXED [Indicator],[Country],[Year] : SUM([Weighted Score])}
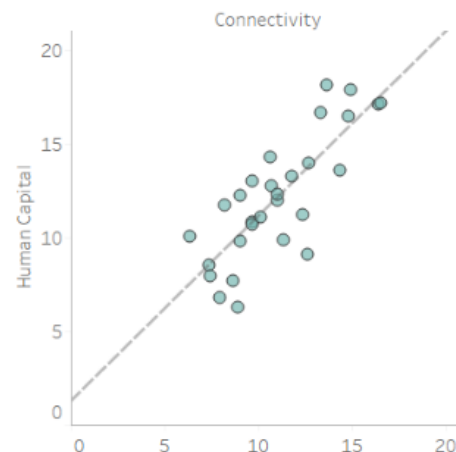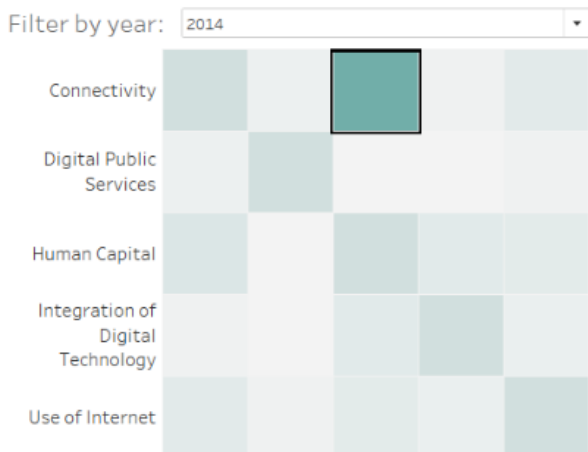
,

{ FIXED [Indicator (chart data1)],[Country (chart data1)],[Year (chart data1)] : SUM([Weighted Score (chart data1)])}

)
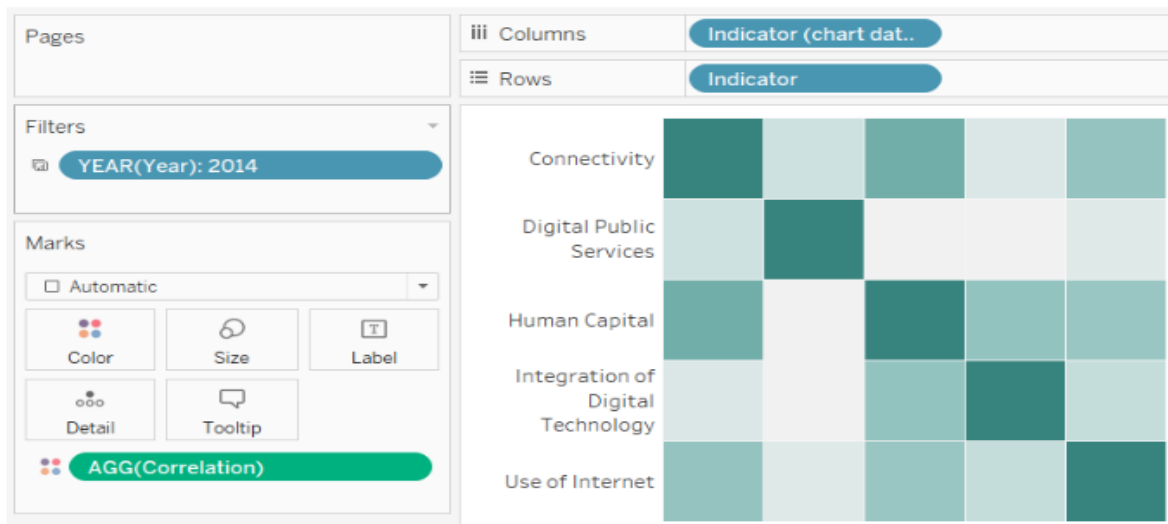
Indicator correlations

7. Compare the following graph layout techniques.

1. Force-directed technique, 2. Multidimensional scaling, 3. Bipartite graph.

1. Force directed:-

-used to visualize the connection between objects

in a network.

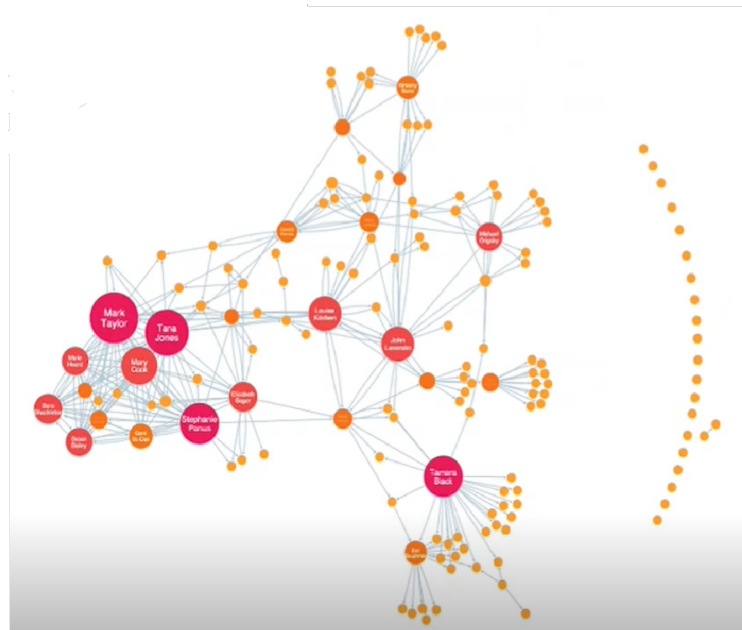-usually results in a symmetrical and clustered structure.

-The technique is based on a physical model

-The model determines the forces acting on the nodes iteratively and moves them in forces to come as close as possible to an equilibrium where the position

the nodes remains stable.

-Used in:

- Network visualization
- Large graph Visualization
- Mesh viz.
- System management
- knowledge representation

2. Multi-dimentional scaling

- Visual representation of distances or dissimilarities between Sets of objects

-Displaying the relative position of a number of objects.

- objects can let colours, faces, map coordinates etc.

-It is a flex family of statistical methods that focus on Creating mapping of items based on distance

-Methods Metrics MDS

　　　　-Non-metaics MPS

　　　　- Individual differences scaling

　　　　-Multidimensional analyze of performance

-More technically, MDS refers to a set of related ordination techniques used in information visualization, in particular to display the information contained in a distance matrix. It is a form of non-linear dimensionality reduction.
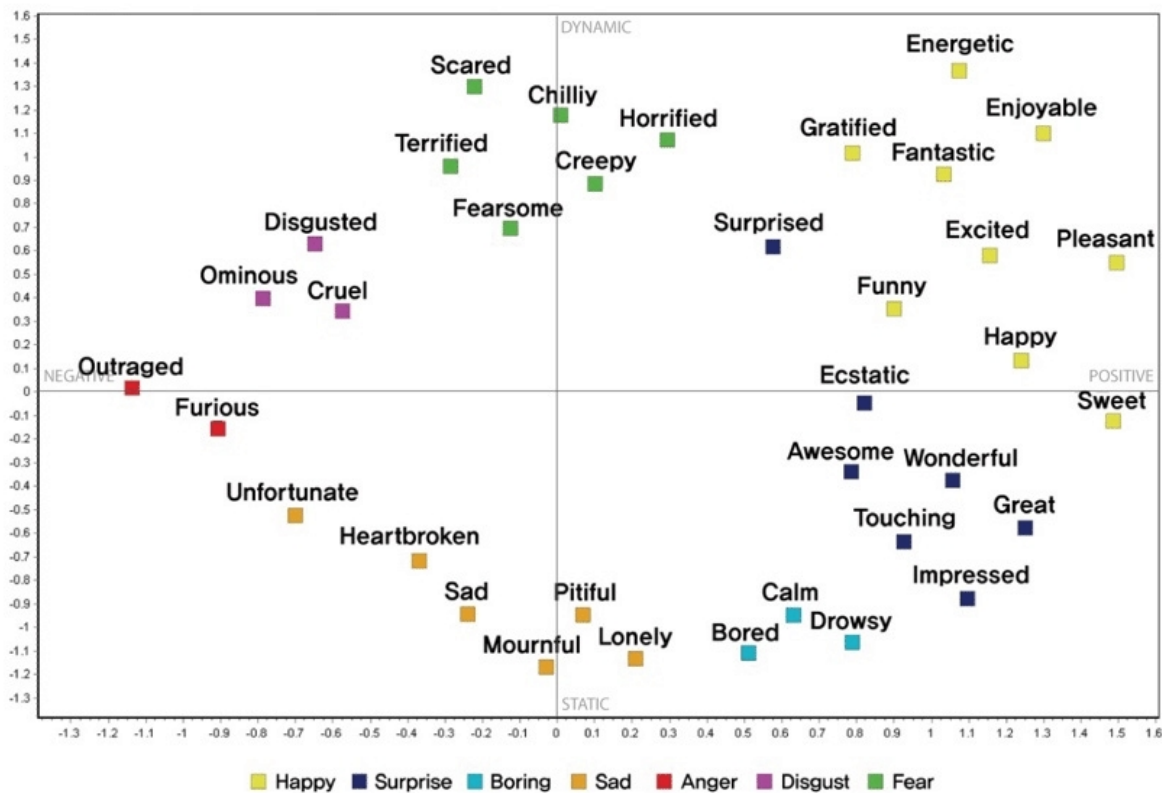
3) Bipartite graph-

-Whose vertices can be divided into two independent sets.

-Graph coloring is possible using two colors.

-Park lc-nodes family where k is layers of nodes.

-BP means 2, K=2

-Set graph vertices decompose into the disjoint sets Such as no two graph vertices within the same set are adjacent.

-The illustration above shows some bipartite graphs, with vertices in each graph coloured based on to which of the two disjoint sets they belong. Bipartite graphs are equivalent to two-colourable graphs. All acyclic graphs are bipartite. A cyclic graph is bipartite if all its cycles are of even length