VELAGAPUDI RAMAKRISHNA

**SIDDHARTHA ENGINEERING COLLEGE**

(AUTONOMOUS)

III/IV B.Tech DEGREE EXAMINATION, FEBRUARY, 2022

Sixth Semester

**INFORMATION TECHNOLOGY**

17IT2605B DATA VISUALIZATION

(CBCS)

**Max.Marks:70**

**PART-A**

**10\*1=10M**

**1. a. What are visualization skills for the masses?**

**Ans** The design methodology presents a sequence of important analytical and design tasks and decisions that need to be handled effectively. The skill of data storytelling is removing the noise and focusing people's attention on the key insights

**b. Define role of Data scientist in Visualization.**

**Ans.** This is basically characterized as data miner who basically the miner’s hat. Their basic role is to acquire, handle prepare and examine the data.

**c. Define Tree hierarchy in Data Visualization**

**Ans.** Hierarchical Visualizations or Trees are collections of items with each item having a link to one parent item (except the root). Items and the links between parent and child can have multiple attributes. These can be applied to items and links.

**d. What is Arrangement in Visualization Design Option?**

**Ans.** Intention with the arrangement and architecture of our design is to deliver as intuitive an experience as possible.

**e. What is Design solution in Data Visualization?**

Ans Focusing on the final tasks involved in constructing solution is Design solution in Data Visualization

**f. What is statistical analysis tool?**

**Ans** Statistical analysis tools are designed to analyze, describe, summarize and compare data. These tools can be relatively simple packages, such as Excel, suites that utilize specialized software and algorithms to deliver more in-depth insight and create visual reports.

**g. What is a Dimension?**

**Ans** Dimensions contain qualitative values (such as names, dates, or geographical data). Dimensions are used to break down the calculated values into specific categories in a chart or table and help to focus on different groups of data.

**h. Define Independent Axes in Tableau.**

**Ans** An Individual Axis in Tableau is obtained by adding measure(s) into Rows or Columns Shelf. Independent axis ranges for each row or column. Makes the axis range independent. The axis range varies for each pane in the view, depending on the range of data in each pane.

**i. Define Pie Chart.**

**Ans** A pie chart is a circular graph divided into various sectors to illustrate numerical proportions. The pie chart is a pictorial representation of data that makes it possible to visualize the relationships between the parts and the whole of a variable.

**j. What is Discrete Data?**

**Ans** Discrete data is a count that involves integers — only a limited number of values is possible. This type of data cannot be subdivided into different parts. Discrete data includes discrete variables that are finite, numeric, countable, and non- negative integers.

**PART-B**

**4\*15=60M**

**UNIT-1**

**2. a. Explain the data visualization methodology and discuss the role of cognitive scientist in data visualization. 7M**

**b. Discuss about eight hats of data visualization and the key factors surrounding a visualization project. 8M**

**a. data visualization methodology 5M**

**cognitive scientist 2M**

**b. eight hats of data visualization 4M**

**key factors surrounding a visualization project 4M**

**Ans:**

**a)**

The design methodology presents a sequence of important analytical and design tasks and decisions that need to be handled effectively. Adopting this methodology is about recognizing the key stages, considerations, and tactics that will help to navigate smoothly through visualization project.

• Design is rarely a neat, linear process and indeed some of the stages may occasionally switch in sequence and require iteration.

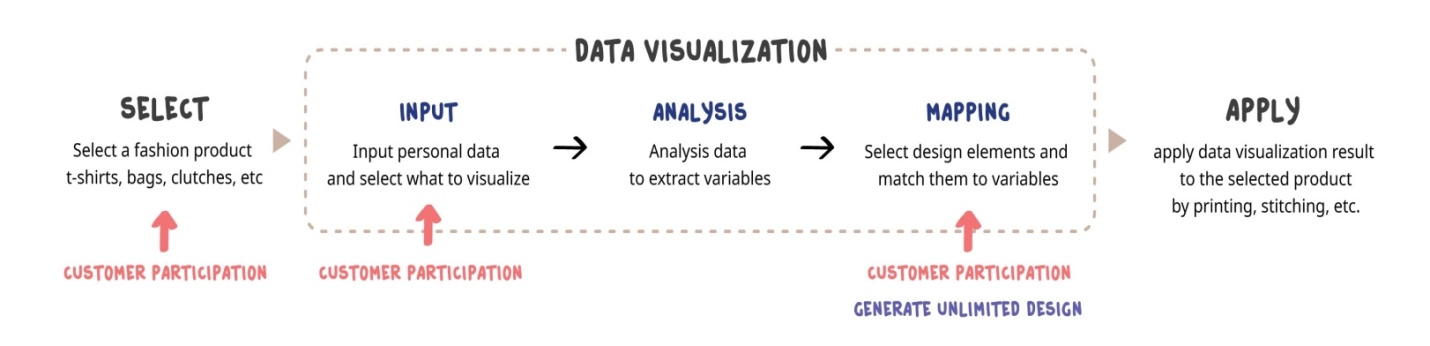
• It is natural that new factors can emerge at any stage and influence alternative solutions, so it is important to be open-minded and flexible.

• Things might need to be revisited, decisions reversed, and directions changed. What we are trying to do, where possible, is find the best path through the minefield of design choices.

The design challenges involved in data visualization are predominantly technology related;

• the creation and execution of a visualization design will typically require the assistance of a variety of applications and programs.

• However, the focus of this methodology is intended to be technology-neutral, placing an emphasis on the concepting, reasoning, and decision-making.



* **The cognitive scientist:**  
  The cognitive scientist is basically the thinker in terms of appreciating the work of both computer scientist and designer. They basically have knowledge about how the eye and the brain work most effectively and efficiently.
* The cognitive scientist is the thinker in terms of appreciating the science behind the effectiveness of the technical and designed solutions.
* They have the visual perception knowledge about how the eye and the brain work most effectively and efficiently.
* They also have deep knowledge about concepts such as the Gestalt Laws, communication theories, color theories, and human-computer interaction principles.
* Additionally, they are able to inform the design process in relation to the complexities of how the mind works in terms of memory, attention, decision-making, and behavioral change.

**b. eight hats of data visualization:**

**8 hats of data visualization** are basically the important persons and their roles that are basically required to carry out data visualization are as follows:

* The initiator
* The data scientist
* The journalist
* The computer scientist
* The designer
* The cognitive scientist
* The communicator
* The project manager

**Let us discuss these hats one by one...**

* **The data scientist:**  
  This is basically characterized as data miner who basically the miner’s hat. Their basic role is to acquire, handle prepare and examine the data.
* **The journalist:**  
  The role of journalist is to build up the narrative approach to the visualization problem. Working with both initiator and data scientist they are able to build up the story for the data visualization.
* **The computer scientist:**  
  The Computer Scientist is basically the executer who brings the project alive because they are only the person who is going to get the solution.
* **The designer:**  
  The designer is creative person who in harmony with the computer scientist will bring out the solution. They basically have the eye for visual detail and a flavor of innovation.
* **The cognitive scientist:**  
  The cognitive scientist is basically the thinker in terms of appreciating the work of both computer scientist and designer. They basically have knowledge about how the eye and the brain work most effectively and efficiently.
* **The communicator:**  
  The communicator is basically concerned with the communication part of the project. These basically work as a presenter in front of client.
* **The project manager:**  
  At last comes the role of project manager whose basic job is to bring the work of all other hats together and manage and supervise their work.

**(or)**

**3. a. Describe about the designer, initiator, cognitive scientist and computer scientist in Data Visualization. 8M**

**b. Write about Establishing intent-the visualization’s tone, when the function to explore and when the function is to exhibit data 7M**

**Ans**

**a designer, initiator, cognitive scientist and computer scientist 4\*2=8M**

**b Establishing intent-the visualization’s tone 7M**

**a)**

**The designer**

The designer is the creative, the one, who, in harmony with the computer scientist, will deliver the solution. They have the eye for visual detail, a flair for innovation and style and are fully appreciative of the potential possibilities that exist. However, they also have the necessary discipline to follow the message established by the initiator and taken on by the journalist.

**The initiator**

The initiator is the leader, the person who is seeking a solution to the task as per the brief or self-initiated curiosity. The hat is that of an explorer; they want to explore data and different design avenues to find answers to problems or evidence to serve their researcher mindset.

**The cognitive scientist**

The cognitive scientist is the thinker in terms of appreciating the science behind the effectiveness of the technical and designed solutions. They have the visual perception knowledge about how the eye and the brain work most effectively and efficiently. They also have deep knowledge about concepts such as the Gestalt Laws, communication theories, color theories, and human-computer interaction principles.

**The computer scientist**

The computer scientist is the executor, the person who brings the project alive. With their critical technical capability they are ultimately the ones who will construct the solution. They will also bolster the data scientist with their technical know-how to most effectively and efficiently handle the data gathering, manipulation, and pre-production visualization activities.

**b) Establishing intent-the visualization’s tone**

The challenge of enhancing visualization literacy is not just about a wider variety of charts and graphs.

– It is also about having greater cleverness to recognize what outcomes certain visualizations are attempting to complete and indeed what they are capable of achieving.

Importance of establishing ‘tone’ is one of the key dimensions of intent.

– Tone concerns the type of experience you are trying to facilitate with your visualization and the intended outcome you hope to achieve.

– In certain contexts there will be a desire to facilitate the ‘reading’ of data, whereas in others there will be greater emphasis on the outcome of ‘feeling’ data.

Tone can be describes as:

– pragmatic or analytical

– emotive or abstract concept

***1) Pragmatic or analytical:***

* "A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other."
* This quote perfectly captures the priority and intent behind pragmatic or analytical visualizations
* If visualization is designed to visually represent data, and to gain new insights into that data, it shall be called a pragmatic(practical) visualization.

***2) Emotive and abstract:***

• Emotive visualization is about type of emotional connection with the design.

**Example:**

Visualize the global airline transportation network consisting of all commercial flights worldwide

– The routes highlighted are those flights in and out of Toronto Pearson airport. The project was designed to assess the threat of infectious diseases.

– The design does not intend to offer an analytical summary of air travel statistics.

– The design is a visual interface to establish a greater sense of how interconnected the world is through air travel.

– It causes us to imagine just how easy it could be for diseases to spread across the globe in a short period of time.

**UNIT-II**

**4. a. Discuss in detail about the choosing the correct visualization method with appropriate example. 8M**

**b. Explain about data visualization design is all about choices with example. 7M**

**Ans**

**a. Explanation choosing the correct visualization method 4M**

**example 4M**

**b. data visualization design 4M**

**example 3M**

**a) Choosing the correct visualization method:**

The first matter is to determine the choice of visualization method.

Commit just yet to a specific chart or graph type, though we might have some in mind. Rather, this is about the general family or collection of chart types as defined by their primary storytelling method.

For example, a bar chart serves the function of comparing categories of values; a line chart, by contrast, enables us to show changes of values over time, geo-spatial data can often be best displayed over a map.

There are a number of ways of classifying the variety of methods for visualizing data but here is a suggested taxonomy:

• Comparing categorical values

• Assessing hierarchies and part-of-a-whole relationships

• Showing changes over time

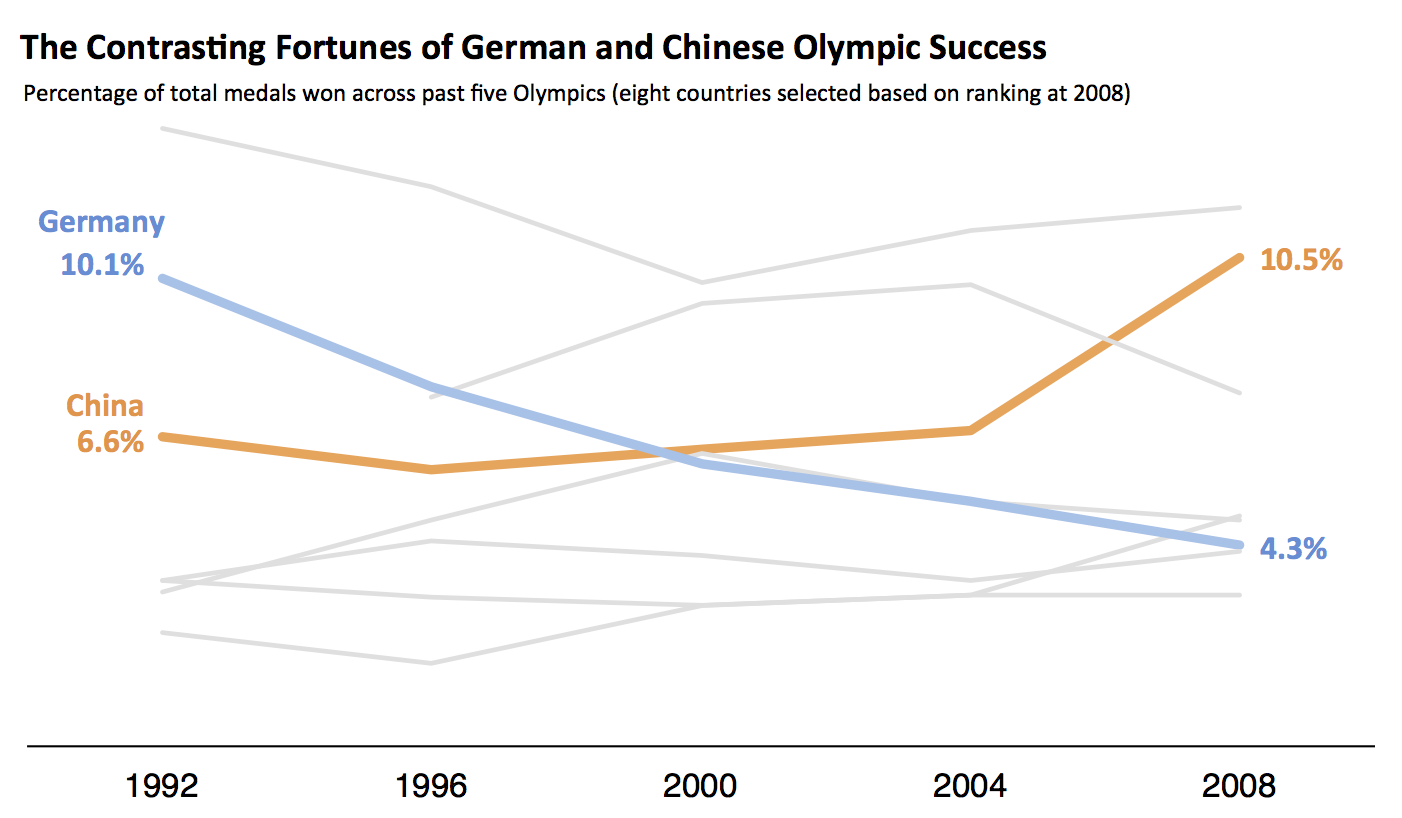
• Mapping geo-spatial data

• Charting and graphing relationships



Example:

"how have the medal-winning performances of China and Germany compared over the past five events?"



**b) Data visualization design is all about choices:**

**The process of building visualization can proceed via four basic steps:**

1.**Design**: the designers must decide how to visually represent the data and which interactions are needed.

2.**Build**: given this design, they then need to actually build the tool.

3.**Use**: the tool is given to end users to analyze their data.

4.**Iterate**: with use, various refinements are usually needed, such as collecting different data, changing the visual representation, or adding additional user interaction capabilities.

Variety of ways in which people interpret visualization design challenges:

1)The first thing to say is that there is never a single path towards a "best" solution.The inherent creativity and individualism of design work ensures that.

2) The second key observation is to remark that the very moment we take on a visualization challenge, and start our journey towards a design solution; we are commencing a unique creative route formed by numerous permutations of choices.

**(or)**

**5. Briefly discuss about choosing appropriate chart types and comparing categories. 15M**

**List and explanation of appropriate chart types 8M**

**Comparison 7M**

**Ans.**

To determine which chart is best suited for each of those presentation types, first you must answer a few questions:

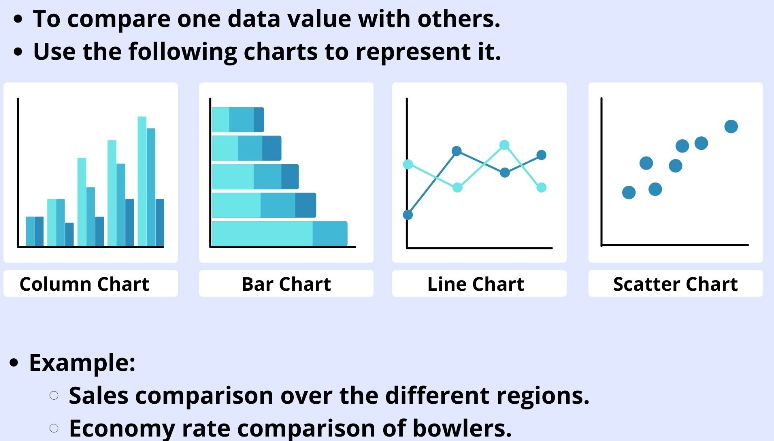
• How many variables do you want to show in a single chart? One, two, three, many?

• How many items (data points) will you display for each variable? Only a few or many?

• Will you display values over a period of time, or among items or groups?

Bar charts are good for comparisons, while line charts work better for trends. Scatter plot charts are good for relationships and distributions, but pie charts should be used only for simple compositions — never for comparisons or distributions.

The chart is a visual representation of data in a clear and simple manner. This type of visualization is often used to increase the level of comprehension, especially in cases of patterns, trends, and interconnections. Building chart is an integral part of data analysis. It helps to decide how to analyze data, how it is distributed, etc.



**Comparing categories:**

Answer whether you have one or more variables. In case you have only one variable move to the left. Then, decide whether this variable is ordered or not. So, if you have only one variable, you can choose your perfect chart among the following types:

Dot plot

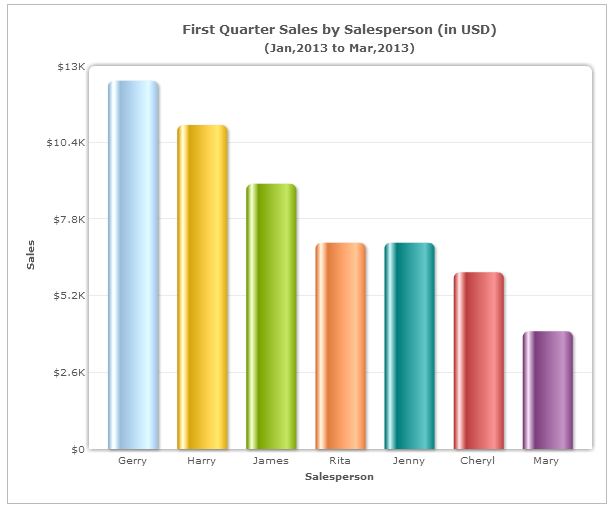
Data variables: 2 x categorical, 1 x quantitative.

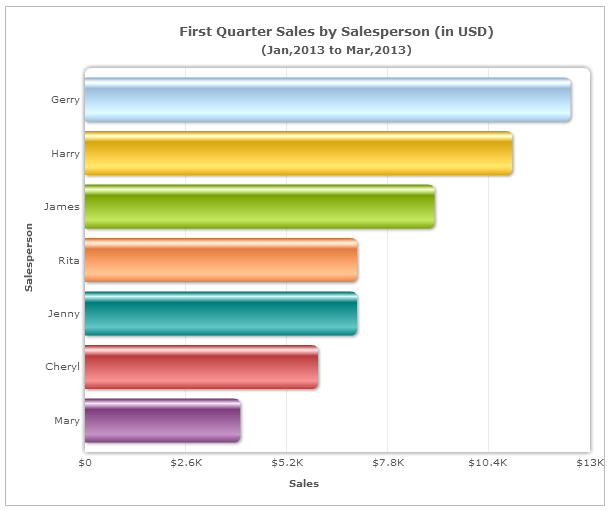
* Visual variables: Position, color-hue, symbol.

Description: A dot plot compares categorical variables by representing quantitative values with a single mark, such as a dot or symbol. The use of sorting helps you to clearly see the range and distribution of values. You can also combine multiple categorical value series on to the same chart distinguishing them using color or variation in symbol. Beyond two series things do start to get somewhat cluttered and hard to read.

* Bar chart (or column chart)

Both the Bar and the Column charts display data using rectangular bars where the length of the bar is proportional to the data value. Both the charts are used to compare two or more values. ... A bar chart is oriented horizontally whereas the column chart is oriented vertically. A bar chart is oriented horizontally whereas the column chart is oriented vertically. Although alike, they cannot be always used interchangeably because of the difference in their orientation.





**UNIT-III**

**6. a. Write in brief about programming languages and applications in design process. 7M**

**b. Describe about programming environments in constructing and evaluating your design solution 8M**

**a. programming languages 4M**

**applications explanation 3M**

**b. programming environments in constructing 4M**

**evaluating design solution 4M**

**Ans**

**a)**

**Introduction:**

• analyze both the opportunities and the pitfalls that emerge when working with technology to visualize data.

• connect relevant development theories to the technological strategies and tools

1) **Visualization software, applications, and programs:**

Data visualization software provides the conversion of textual and numeric data into visual charts, figures and tables. It is used as a means to create application/system performance or operational dashboards by bringing in important data to a central interface.

Data visualization software is typically integrated with the underlying system. It captures, analyzes and reports on key performance and/or or operational statistics from the software/system. The captured data is visually presented in a preconfigured visual data mode.

For example, the monthly number of visitors on a website may be presented in bar charts where each bar represents the number of visitors in that particular month.

**b) programming environments in constructing and evaluating your design solution**

Data visualization solutions do all that work for you. The process is automatic, fast, and 100% accurate. Thanks to effective programming, it all happens behind the scenes.

The end result is data visualization tools that are accessible to everyone.

This means you can view and analyze your critical business data at any time, allowing instant access to the information you need. This has several benefits for the success of your business.

Many of today’s computer programming languages are suitable for data visualization outcomes.

• The ultimate capability in visualization design is to have complete control over the characteristics and behavior of every mark, property, and user-driven event on a chart or graph.

• The only way to fundamentally achieve this level of creative control is through the command of one or a range of programming languages.

There are different programming languages available : **Some are:**

**a) Adobe Flash:**

A powerful and creative environment for animated and multimedia designs. Flash was behind many prominent interactive visualization designs in the field.

Adobe Flash is a deprecated multimedia software platform used for production of animations, rich Internet applications, desktop applications, mobile applications, mobile games and embedded web browser video players.

**b) D3.js:**

• D3 stands for Data-Driven Documents.

• A web based visualization library that features a lot of APIs to handle the heavy lifting of creating advanced, dynamic and beautiful visualization content on the web.

• D3.js is a JavaScript library for manipulating documents based on data. D3.js helps you bring data to life using HTML, SVG, and CSS. D3’s emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

o DOM is an acronym stands for Document Object Model. It defines a standard way to access and manipulate documents. The Document Object Model (DOM) is a programming API for HTML and XML documents. It defines the logical structure of documents and the way a document is accessed and manipulated.

o Scalable Vector Graphics (SVG) are an XML-based markup language for describing two-dimensional based vector graphics. As such, it's a text-based, open Web standard for describing images that can be rendered cleanly at any size and are designed specifically to work well with other web standards including CSS, DOM, JavaScript, and SMIL. SVG is, essentially, to graphics what HTML is to text.

Example:

a data visualization that shows the distribution of mobile phone users in China

**(or)**

**7. a. Analyze about statistical analysis tools and charting productivity tools for powerful visualization capabilities 7M**

**b. Discuss about final stage of construction process and post launch of visualization. 8M**

**a) Any one tool explanation**

**statistical analysis tools 3.5M**

**charting productivity tools 3.5M**

**b) final stage of construction process 4M**

**post launch of visualization 4M**

**a.**

**Charting and statistical analysis tools:**

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs and maps, data visualization tools provide an accessible way to see and understand trends, outliers and patterns in data.

Data visualization tools provide data visualization designers with an easier way to create visual representations of large data sets. When dealing with data sets that include hundreds of thousands or millions of data points, automating the process of creating visualization, at least in part, makes a designer’s job significantly easier.

These data visualizations can then be used for a variety of purposes: dashboards, annual reports, sales and marketing materials, investor slide decks, and virtually anywhere else information needs to be interpreted immediately.

a) **Microsoft Excel:**

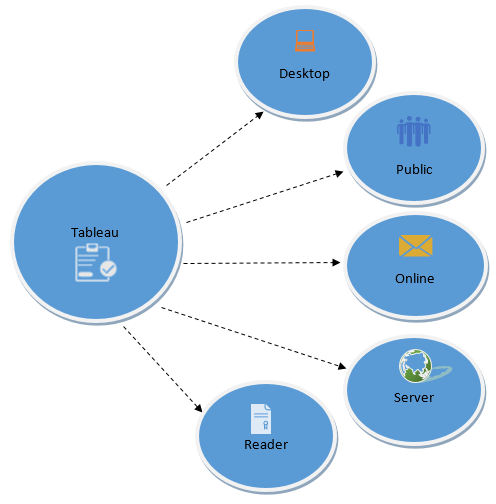
• In Excel, charts are used to make a graphical representation of any set of data.

• A chart is a visual representation of the data, in which the data is represented by symbols such as bars in a Bar Chart or lines in a Line Chart.

• Excel provides you with many chart types and you can choose one that suits your data or you can use the Excel Recommended Charts option to view charts customized to your data and select one of those.

b) **Tableau:**

Tableau is a powerful and fastest growing data visualization tool used in the Business Intelligence Industry. It helps in simplifying raw data into the very easily understandable format.

Data analysis is very fast with Tableau and the visualizations created are in the form of dashboards and worksheets. The data that is created using Tableau can be understood by professional at any level in an organization. It even allows a non-technical user to create a customized dashboard.

The best feature Tableau are

• Data Blending

• Real time analysis

• Collaboration of data

Tableau Product Suite

The Tableau Product Suite consists of

• Tableau Desktop

• Tableau Public

• Tableau Online

• Tableau Server

• Tableau Reader

**b) final stage of construction process and post launch of visualization.** The construction process:

• After selection the tools you'll need to build your design and you are now well in to the execution stage.

• This is the part of the design process where stresses and strains emerge—the

ill-timed bugs, dataset problems, functional failures, unwanted interference. During

this stage it is important that you keep your cool and see your tasks through as

efficiently as possible.

• Through the construction process, it is important to focus on getting the functional elements of your solution working first before spending too much time achieving your desired aesthetic or incorporating technical flair.

• create a worthy project without the need for iteration: a methodological approach to this challenge gives you structure and a neat framework of concerns to work through, iteration gives you the creative breathing room to allow different ideas to blossom and influences to take hold.

• to avoid long iterative cycles but smaller ones can really help you explore, clarify, and refine your potential solutions. It may be that you end up following two or three parallel options to quite an advanced stage and then see which emerges as the strongest. Indeed, some clients will state a need for evidence of alternatives before committing themselves. For these client-based projects, you need to maintain open dialog throughout to avoid any inconsistency in interpretation from either party.

Example:

Olympics project that was formed out of thumbnail images of all the historical

event posters; this image is shown here:

When it came to incorporating this title into the final piece, it was clear that it drew

too much attention away from the rest of the visualization.

So drop it and find a simpler solution.

****

**UNIT-IV**

**8. a. Interpret about view the number of records and aggregation in Tableau with example 6M**

**b. Describe about Marks cards,encoding and level of detail with example 9M**

**a) aggregation in Tableau 3M**

**example 3M**

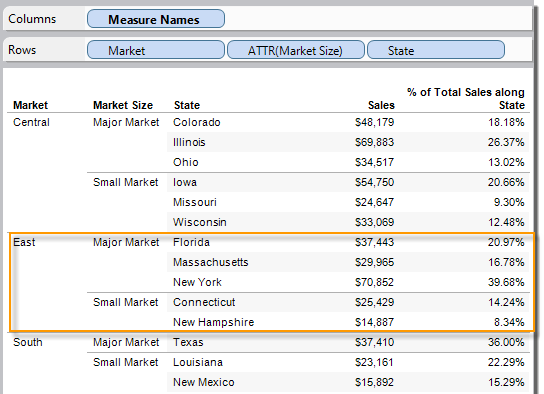
**b) Marks cards,encoding and level of detail 5M**

**example 4M**

**Ans**

**a)**

* In Tableau, you can aggregate measures or dimensions, though it is more common to aggregate measures. Whenever you add a measure to your view, an aggregation is applied to that measure by default. The type of aggregation applied varies depending on the context of the view.
* When you add a measure to the view, Tableau automatically aggregates its values. Sum, average, and median are common aggregations;
* The current aggregation appears as part of the measure's name in the view. For example, Sales becomes SUM(Sales). Every measure has a default aggregation which is set by Tableau when you connect to a data source. You can view or change the default aggregation for a measure



**b) Marks cards,encoding and level of detail:**

* The Marks Cards in Tableau provide some of the most powerful functionality in the program because they allow you to modify a view’s design, visualization type, user experience, and granularity of analysis all in one place.
* The Marks Shelf, there are six different Marks Cards: Color, Size, Label, Detail, Tooltip, and Shape. Note that the Shape Marks Card is not available for every view, but appeared because we are creating a scatter plot in this example. To help explain what each of these cards does, discuss the Detail Marks Card first.
* The Marks card is a key element for doing visual analysis in Tableau. When you drag measures and dimensions to the Marks card, you encode your data with color, size, shape, text, and other properties that add context, detail, and meaning to the marks in the view

A graphic depicting a quantitative
axis and column headers in a view.A graphic depicting a quantitative
axis and column headers in a view.

**(or)**

**9. a. Analyze about view the underlying data and connecting to data in Tableau 8M**

**b. Discuss about five ways to create a bar in Tableau 7M**

**a) underlying data in Tableau 4M**

**connecting to data in Tableau 4M**

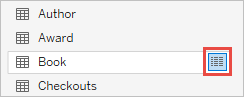
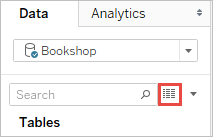
**b) five ways to create a bar in Tableau 4M**

**example 3M**

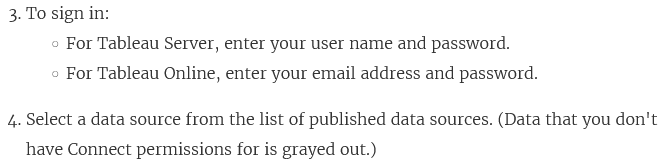
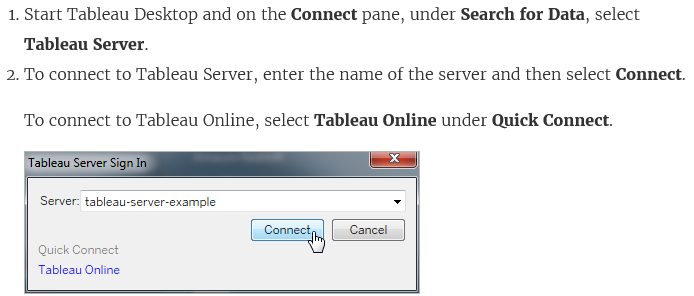
**Ans**

**a) underlying data in Tableau:**

* View Data allows you to inspect your data in a spreadsheet-like layout. You can view data either for the data source as a whole, or to see the underlying data for an individual mark or a group of marks. In a worksheet, the rows that you see in the View Data window are always scoped to the current selection or the current view.
* The View Data window displays as much of the data as possible by default, up to 10,000 rows. Field names are shown as column headers and can be dragged and dropped to change their display order. Click a column header to sort the values in that column.

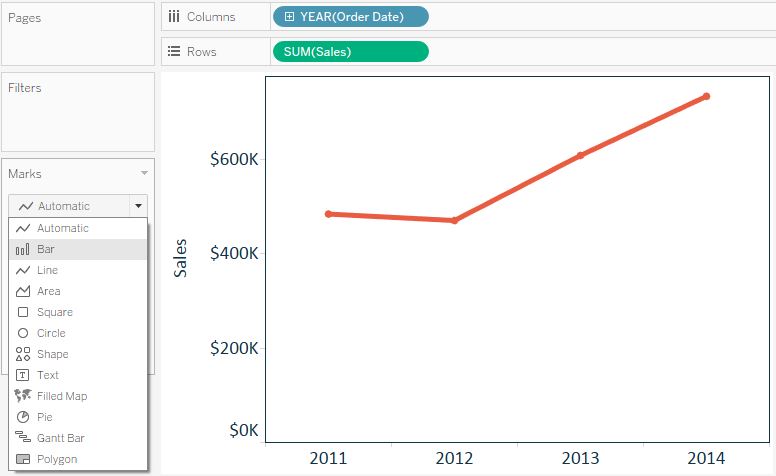
 

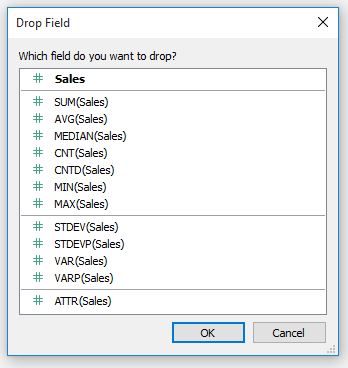
**connecting to data in Tableau:**

**b) five ways to create a bar in Tableau :**

1. The easiest way to start a bar chart in Tableau is to simply double-click on the measure you want to visualize from the Measures Shelf. By default, this will place a continuous pill for Sales on the Rows Shelf, which creates a vertical bar.
2. You could have got to this same place by left-clicking and dragging the Sales measure from the Measures Shelf to the Rows Shelf.
3. ‘Pre-select’ the Sales measure by clicking on it, then click ‘horizontal bars’ in the Show Me options. This creates a different orientation than the first two approaches because the Sales measure is placed on the Columns Shelf instead of the Rows Shelf. If you prefer the vertical orientation, you can click the Swap icon (pictured below), use the Ctrl+W shortcut, or drag and drop the Sales measure from the Columns Shelf to the Rows Shelf.
4. You can change the mark type on an existing view to Bar. Let’s say you are looking at the Sales measure by Year of Order Date as a line graph. You can convert the line graph to a bar chart by changing the mark type on the Marks Shelf from Automatic (line) to Bar.
5. Similar to approach 2 above, but if you right-click and drag the Sales measure from the Measures shelf to the Rows Shelf, you will be presented with the option to choose the data aggregation before the bar chart is created.





**Course Coordinator**

**Module Coordinator Program Coordinator HOD:IT**