Unit 2

Creating visual representation

Visualization Reference Model:

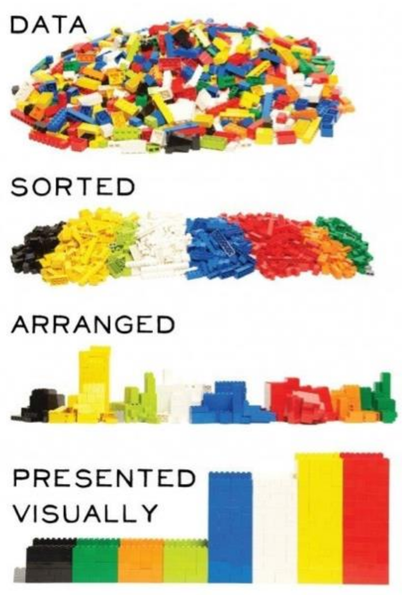
It is a conceptual model that describe the different components involved in data visualization and act as a guide to design and development of visualization system to help users to understand to interpret the result and use visualization in effective way. It helps data analysts, designers and stakeholders to ensure that the visualizations are clear informative and aligned with the goals of the project. It provides set of guidelines and design principle that can be used to create visually appealing and informative visualization eccentrics like charts, graphs, line charts, scatter plot etc. Following are the common element that a visual reference model includes:

- Data: represents the elements that are being visualized and can be structure or unstructured or quantitative or qualitative.
- Interaction: it shows how user can interact with visualization or in what ways user can interact. This includes the features like zooming, panning, tooltips and filtering.
- Color: specify a consistent color to use in visualization which makes the output more visually appealing. This includes guidelines for primary colors, accent color, how to use color to represent different data categories or values.
- Evaluation: this includes the way that the effectiveness of the visualization is measured as well as the way the visualization is tailored to the needs of the users.
- Layout: determine the overall layout of the visualization including the placement of axes, legends and data element. This includes margins, padding and spacing to create a balanced composition.
- Data encoding: define how variables should be mapped to visual properties such as position, size, shape and color. It ensures that the encoding choices align with the goals of the visualization and make it easy for viewer to interpret the data.
- Annotations and labels: specify guidelines for adding annotations, titles and labels to clarify and contextualize the dat. Annotation can include trend line, labels or text.
- Legends and scales: define how legends and scales should be presented when using color, size or other visual encodings. Legends helps to interpret the visualization.
- Consistency: this includes consistent use of colors, fonts and design element. It helps to maintain minimum deviation between actual result and result from visualization tools.

Reference Model

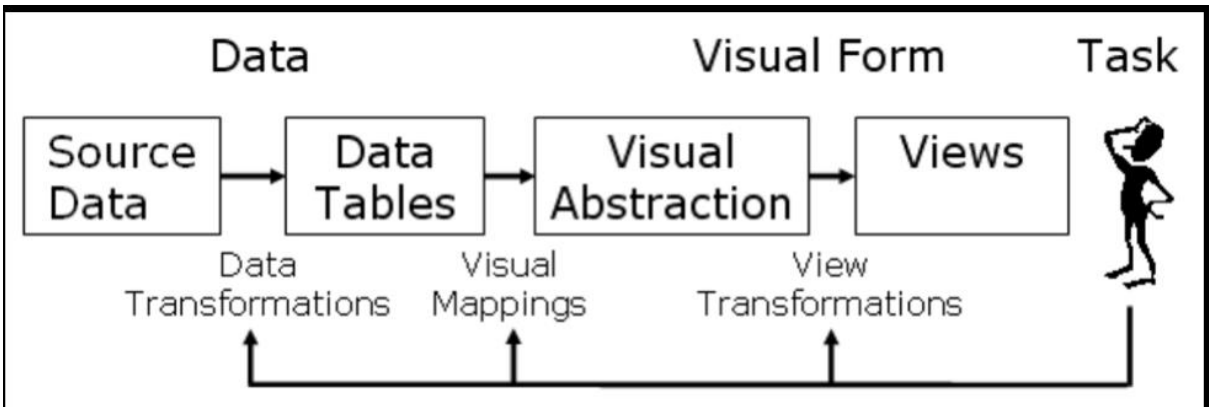
- an abstract framework or domain-specific ontology consisting of an interlinked set of clearly defined concepts produced by an expert or body of experts to encourage clear communication.

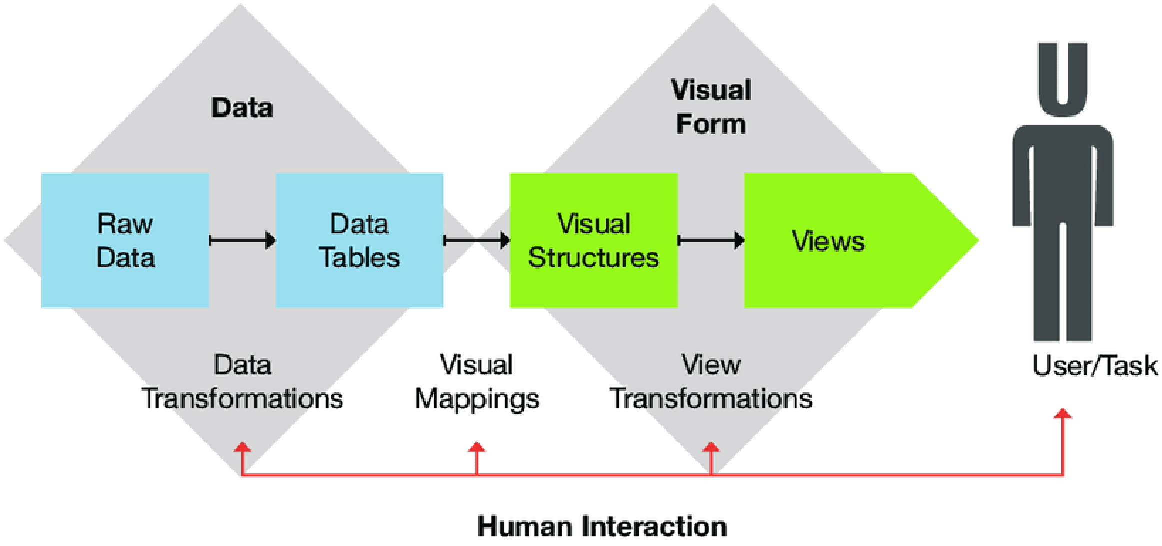
what is visualization reference model?



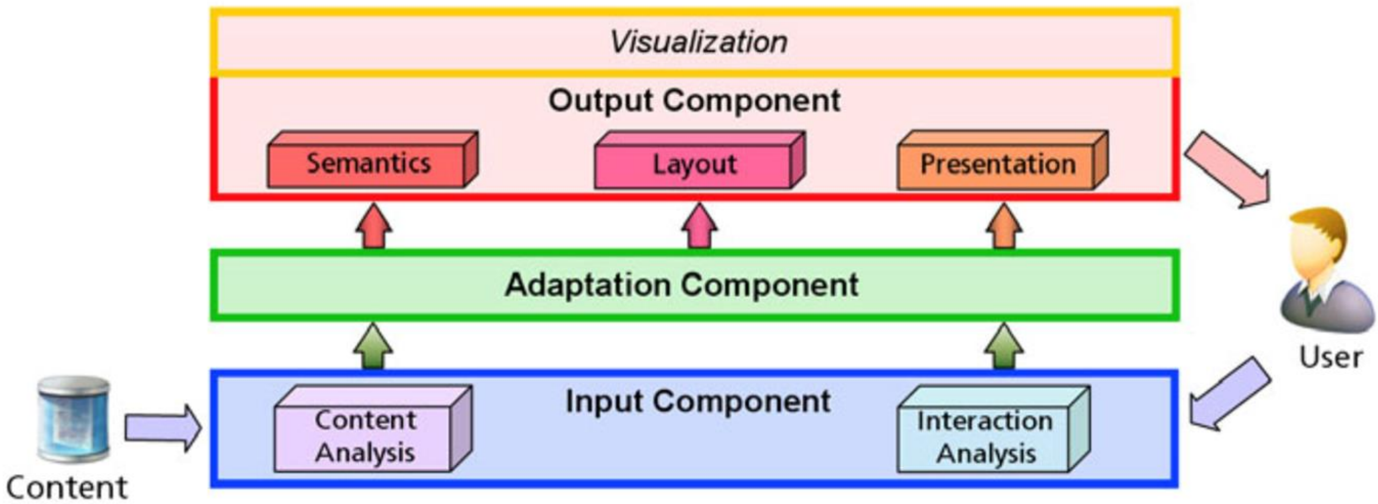
Source: http://www.hotbutterstudio.com/

THE INFOVIS REFERENCE MODEL
aka infovis pipeline, data state model [Chi99]



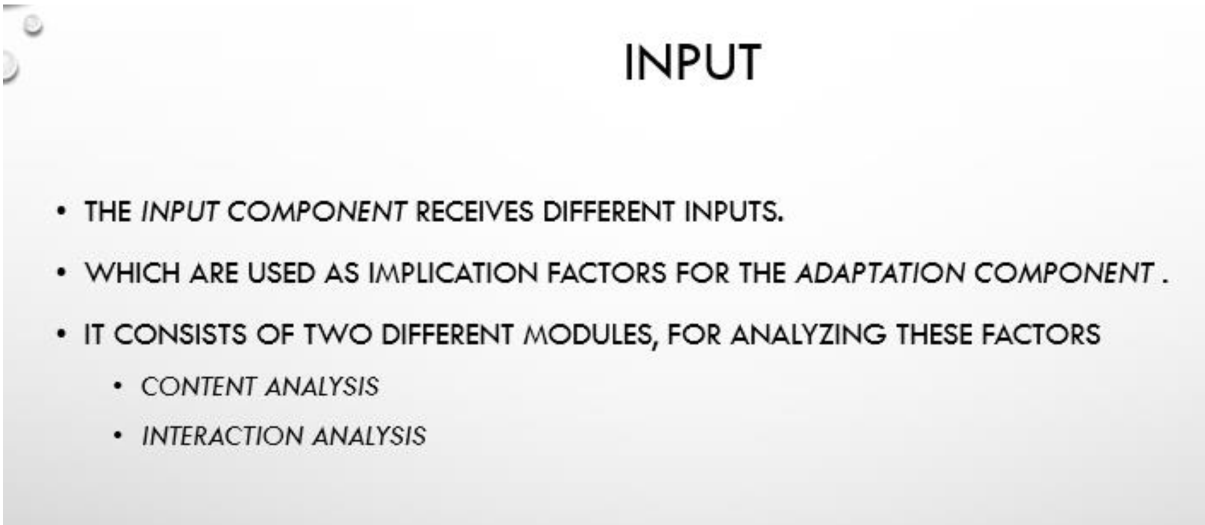


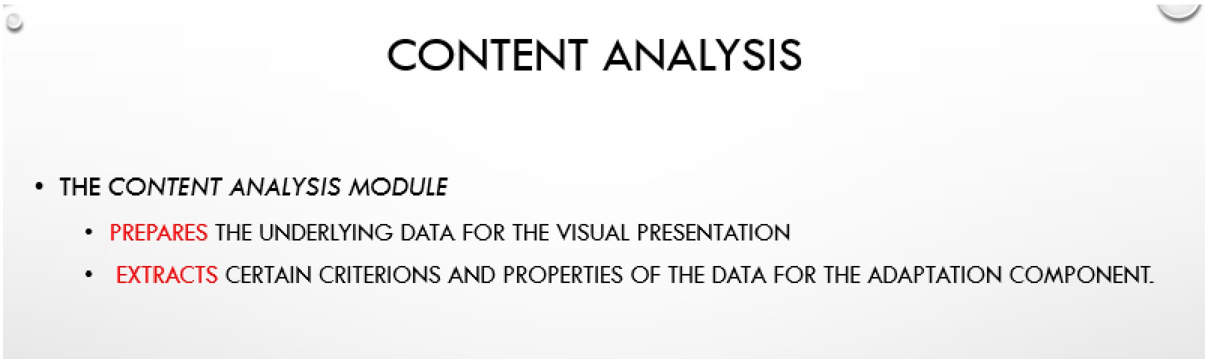
Reference model for adaptive Visualization system

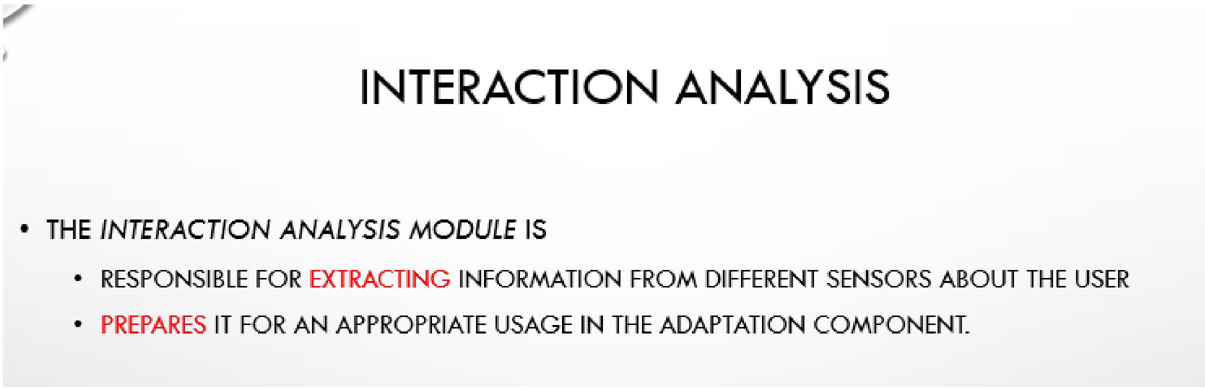


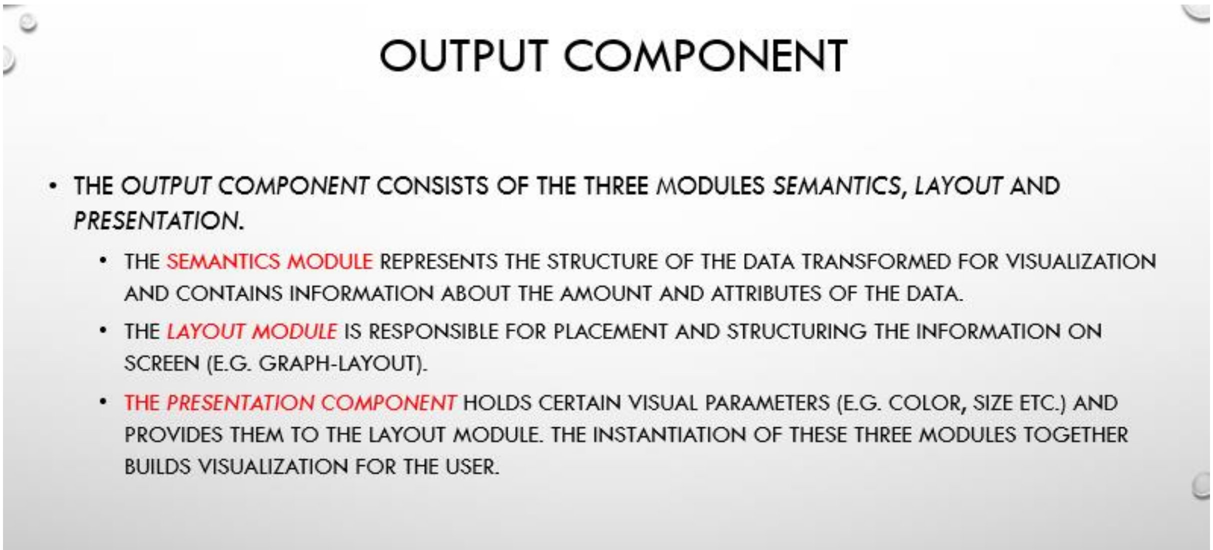
Three Basic Components

- INPUT
- ADAPTATION
- OUTPUT









Visual Mapping:

It refers to the process of assigning visual attributes and properties to data elements in order to represent specific information and patterns effectively. That is, it is the process of representing data in visual perceptive using different visualization eccentrics like map, chart, graph etc. it is a crucial step to transform raw data into meaningful and informative visual representation and involves thoughtful decisions about encoding data using visual variables to create clear and insightful visualization. It helps viewer to interpret data quickly and accurately by understanding human perception and cognition. It includes making decision about encoding data variables, using visual channels such as position, size, color shape and texture. When choosing a visual mapping technique, following factor should be considered:

- The type of data that are to be visualized
- The message to be communicate with viewer
- The target viewer
- The available tools and resources

Some of the examples of visual mappings are bar chars that uses bars to represent different values and comparison between them, line charts to show trends or change over time, scatterplot that uses dot to show correlation between variables, pie chart that uses slices of pie to show relative proportion of different values, map that uses geographic features to show relation between different data points. The choice of mapping depends on the data type, the message to be conveyed and the target audience. Following are the key aspect of visual mapping in data visualization:

- **Visual variables:** are the attributes that can be used to encode data like position (x axis and y axis to create scatterplots, bar charts etc.), size (represents data value by varying the size of graphical elements like points, bar etc.), color (to distinguish categories, numerical values or emphasize specific data points), shape (circle, square, triangles etc. to encode categories or data attributes), texture (to differentiate data elements when color is not suitable or to add additional encoding), opacity/transparency (to show overlapping data points or emphasize specific data regions).
- **Mapping data types:** data types like nominal, ordinal, interval ratio require different visual mappings. Nominal data may be mapped using distinct colors or shapes while ordinal data might use gradients of color or size to indicate order.
- **Data attributes:** includes specific attributes like quantitative, categorical, temporal or spatial which may require different mapping strategy.
- **Scale and range:** determine the range of values for each visual variables in which different color can be mapped using continuous gradient or discrete color palette.
- **Normalization**: while comparing multiple data attributes or datasets it is necessary to normalize or standardize the data to ensure fair and meaningful visual comparisons.
- Interactivity: this includes effects like zooming, filtering or highlighting data points through user interaction
- **Accessibility:** ensures that the chosen visual mappings are accessible to all users and should provide alternative text for visual elements and consider using proper color palettes.
- **Feedback and interaction:** to refine visual mapping it is necessary to gain feedback from user to ensure that the visual mappings effectively communicate the intended message

For the effective and efficient visual mapping following terms should be consider:

- Using consistent color scheme will help viewer to quickly and easily understand the data
- Using clear and concise labels on data will help the viewer to understand what the data represents.
- Using limited number of visual elements in visualization will help to avoid visual clutter and make the visualization easier to understand.

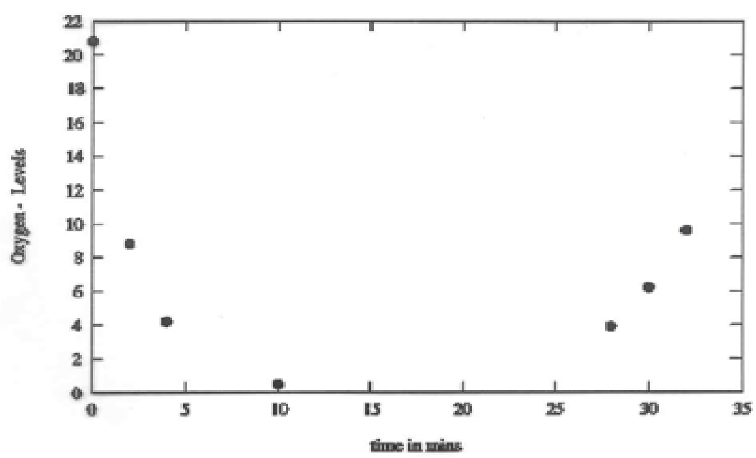
While creating data visualization various elements are employed to communicate data, relationships and insights clearly. Some essential elements of visual mapping are: position, size, color, shape, texture, line, symbols, labels, opacity, legends etc.

A SIMPLE EXAMPLE

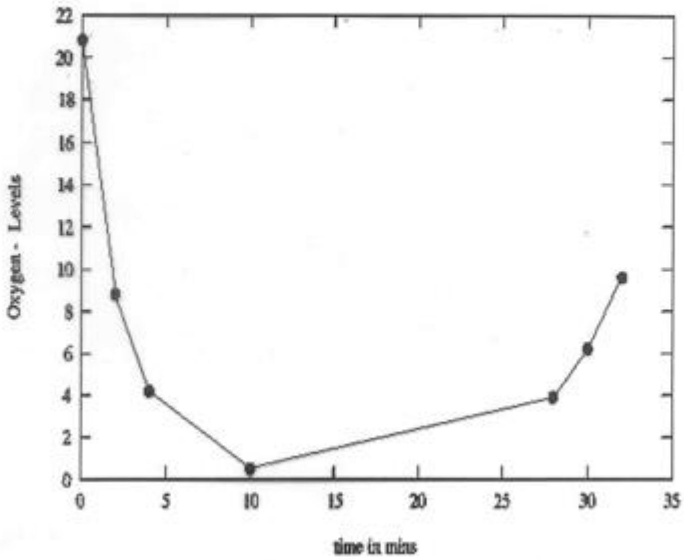
This table shows the observed oxygen levels in the flue gas, when coal undergoes combustion in a furnace

TIME (mins)	0	2	4	10	28	30	32
OXYGEN (%)	20.8	8.8	4.2	0.5	3.9	6.2	9.6

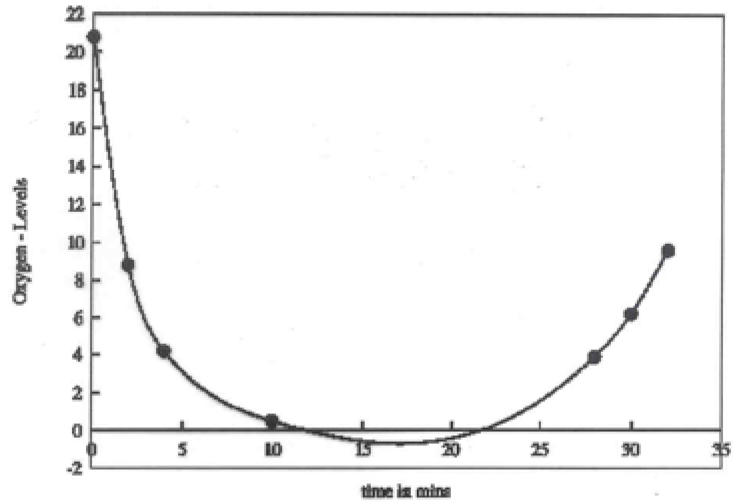
VISUALIZING THE DATA - BUT IS THIS WHAT WE WANT TO SEE?



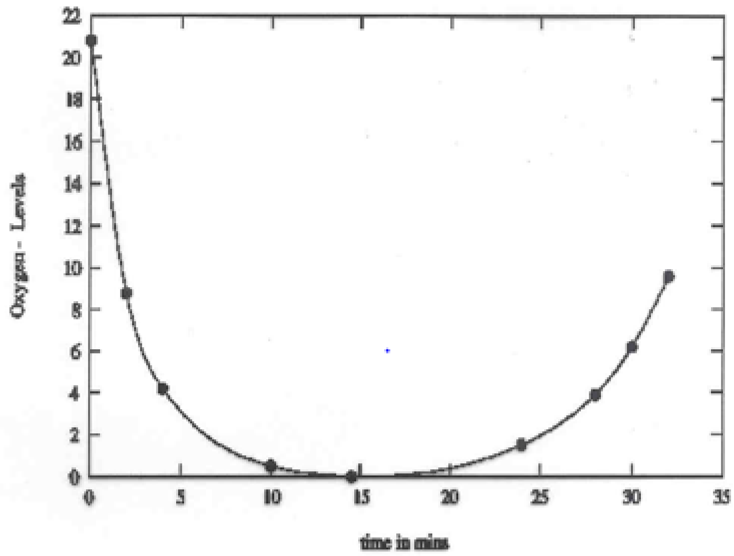
ESTIMATING BEHAVIOUR BETWEEN THE DATA - BUT IS THIS BELIEVABLE?



NOW IT LOOKS BELIEVABLE... BUT SOMETHING IS WRONG



AT LEAST THIS IS CREDIBLE..

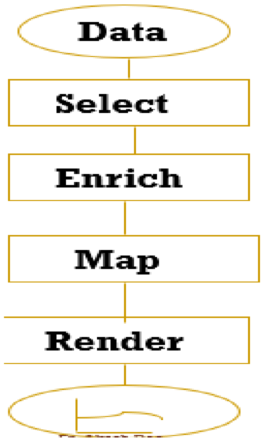


- It is not only the data that we wish to visualize - it is also the bits in between!
- The data are samples from some underlying 'field' which we wish to understand
- First step is to create from the data a 'best' estimate of the underlying field - we shall call this a MODEL
- This needs to be done with care and may need guidance from the scientist

DATA ENRICHMENT

- This process is sometimes called 'data enrichment' or 'enhancement'
- If data is sparse, but accurate, we INTERPOLATE to get sufficient data to create a meaningful representation of our model
- If sparse, but in error, we may need to APPROXIMATE

BACK TO THE SIMPLE EXAMPLE



Extract part of data we are intereste

Interpolate to create model

Select a line graph as technique and create line segments from enriched data

Draw line segments on display in suitable colour, line style and width

THE VISUALIZATION PROCESS

- Overall the Visualization Process can be divided into four logical operations:
  - DATA SELECTION: choose the portion of data we want to analyse
  - DATA ENRICHMENT: interpolating, or approximating raw data - effectively creating a model
  - MAPPING: conversion of data into a geometric representation
  - RENDERING: assigning visual properties to the geometrical objects (eg colour, texture) and creating an image

CLASSIFICATION OF MAPPING TECHNIQUES

- The mapping stage is where we decide which visualization technique to apply to our 'enriched' data
- There are a bewildering range of these techniques - how do we know which to choose?
- First step is to classify the data into sets and associate different techniques with different sets.

BACK TO THE SIMPLE EXAMPLE

- The underlying field is a function F(x)
  - F represents the oxygen level and is the DEPENDENT variable
  - x represents the time and is the INDEPENDENT variable
- It is a one dimensional scalar field because
  - the independent variable x is 1D
  - the dependent variable F is a scalar value

GENERAL CLASSIFICATION SCHEME

- The underlying field can be regarded as a function of many variables: say

$$F(x)$$

where F and x are both vectors:

$$F = (F_1, F_2, \dots F_m)$$

$$x = (x_1, x_2, \dots x_n)$$

- The dimension is n
- The dependent variable can be scalar (m=1) or vector (m>1)

A SIMPLE NOTATION

- This leads to a simple classification of data as:

$$E_n^{S/V}$$

- So the simple example is of type:

$$E_1^{S}$$

- Flow within a volume can be classed as:

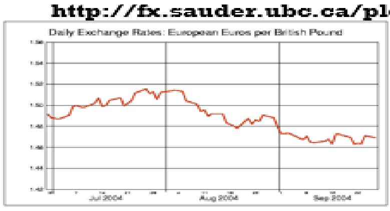
$$E_3^{V3}$$

$E^S{}_1$

- The humble graph!
- How can we represent errors in the data?

A nice example of web-based visualization….

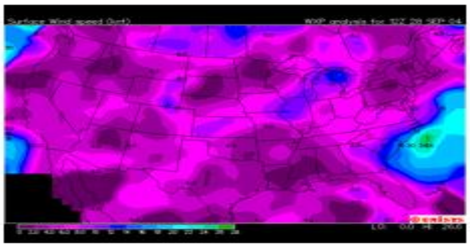
http://fx.sauder.ubc.ca/plot.html



E^S_2

- Here we see a contour map of wind speed over the USA (28-Sep-04)
- What can you observe?
- Can you use an E^S_1 technique for this sort of data?

http://weather.unisys.com/surface/



E^S_3

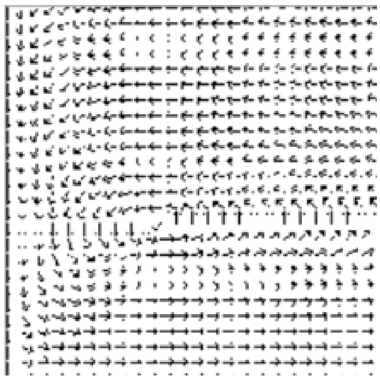
- As dimension increases, it becomes harder to visualize on a 2D surface
- Here we see a lobster within resin.. where the resin is represented as semi-transparent
- Technique known as volume rendering

E^S_3

- Corresponding to contours for E^S_2, we can generate isosurfaces
- What are the limitations of this approach compared with volume rendering?

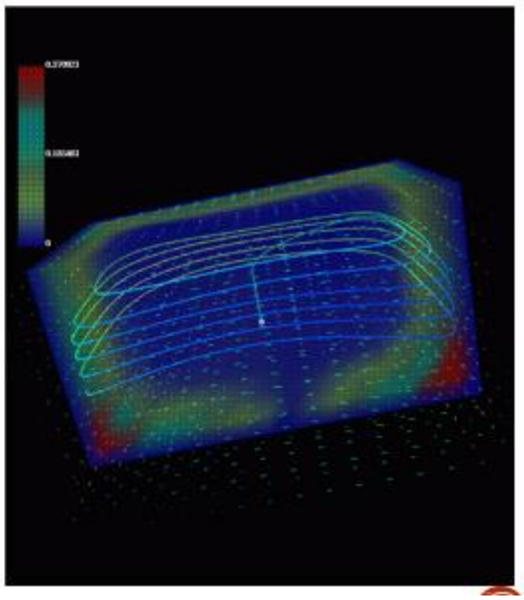
$E^{V2}{}_2$

- This is a flow field in two dimensions
- Simple technique is to use arrows..
- What are the strengths and weaknesses of this approach?
- During the module, we will discover better techniques for this



$E^{V3}{}_3$

- This is flow in a volume
- Arrows become extremely cluttered
- Here we are tracing the path of a particle through the volume



Visual analytics:

It is an approach to data analysis that combines interactive data visualization with advance analytical techniques which aims to helps users gain insight from complex data by using visualization, statistical and human cognition. It combines data visualization, data mining and interactive techniques to helps users explore and understand large and complex datasets. It is a tool for gaining effective insights from data that is difficult or impossible to gain from traditional methods. Therefore, visual analytics help users to explore, interact with data an to explore relationships and patterns. Visual analytics is widely used in various domains including business intelligence, healthcare, finance and scientific research etc. Some of the tools used for visual analytics are power bi, tableau, ggplot, matplotlib etc. Following are the advantage of visual analytics in data visualization:

- Help user to understand complex data more easily
- Help user to identify patterns and trends that are difficult to analyze from spreadsheet or table
- Can be used to share data with others and to collaborate on analysis
- Helps to drill down on specific areas of interest

Definition

- In "Illuminating the Path" [39], Thomas and Cook define visual analytics as the science of analytical reasoning facilitated by interactive visual interfaces.
- Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets.

Visual analytics involves several key components and principles:

i. Interactive visualizations: includes filter, zoom, drill down and manipulate visual representations to uncover patterns, trends and outliers.

ii. Data integration: visual analytics involves integration of data from multiple sources and formats i.e. includes structured data from databases, unstructured data from text sources and real time data which enable more comprehensive analysis.

iii. Visual data exploration: includes charts, graphs, maps and dashboard to visually explore complex data which helps users to understand and interpret such complex data and their relationship. It includes bar chart, line chart, scatter plots and geographic maps.

iv. Data transformation: visual analytics tools provide data transformation capabilities, including data cleaning, transformation capabilities, data cleaning to prepare data for analysis.

v. Data security and privacy: for the sensitive data, visual analytics platforms often incorporate security and privacy features to ensure data protection and compliance with regulation.

vi. Customization: using visual analytics tools users can customize the visualizations to suit their specific needs which may includes adjusting color, fonts and different layouts.

Importance of Visual Analytics

i. Enhanced data understanding:

   Visual analytics involves in representing complex data in visual form that allows uses to see patterns, trend and outlies which may not be seen in raw data or text-based reports.

ii. Effective communication:

   Visualization conveys information quickly and efficiently which enable clear and concise communication of data finding to users. Visual analytics helps to bridge the gap between data experts and non-experts.

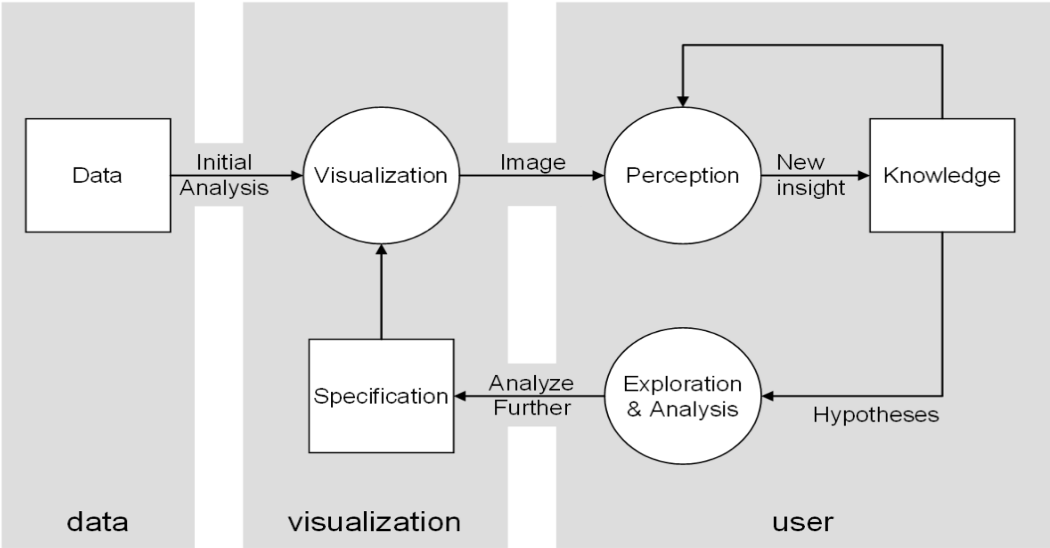
iii. Decision making: visual analytics empowers decision makers to base their choices on data driven insights rather that intuition or evidence which leads to more informed and strategic decision.

iv. Real time monitoring: visual analytics supports different tool for time series data which enables organization to respond quickly to changing conditions and make timely intervention or adjustment.

v. Pattern recognition: our visual perception is more suited for recognizing different patterns and anomalies. Visual analytics helps to simplify this character by allowing users to identify irregularities, trends and outliers more effectively than with traditional data analysis methods.

Goal

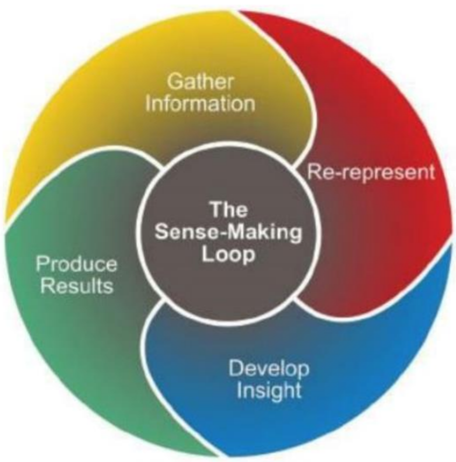
- The goal of visual analytics is the creation of tools and techniques to enable people to:
  - Synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data.
  - Detect the expected and discover the unexpected.
  - Provide timely, defensible, and understandable assessments.
  - Communicate assessment effectively for action.
- From Shneiderman’s celebrated mantra “Overview first, Filter and zoom, Details on demand”.
- Recently, Keim adjusted the mantra to bring its focus toward Visual Analytics: “Analyze first, Show the Important, Zoom, filter and analyze further, Details on demand”. In other words, this mantra is calling for astute combinations of analytical approaches together with advanced visualization techniques.



- In the Engineering domain, Visual Analytics can contribute to speed-up development time for products, materials, tools and production methods by offering more effective, intelligent access to the wealth of complex information resulting from prototype development, experimental test series, customers' feedback, and many other performance metrics.

- Visual analytics does the "heavy lifting" with data, by using a variety of processes — mechanical, algorithms, machine learning, natural language processing, etc
- to identify and reveal patterns and trends. It prepares the data for the process of data visualization, thereby enabling users to examine data, understand what it means, interpret the patterns it highlights, and help them find meaning and gain useful insights from complex data sets.

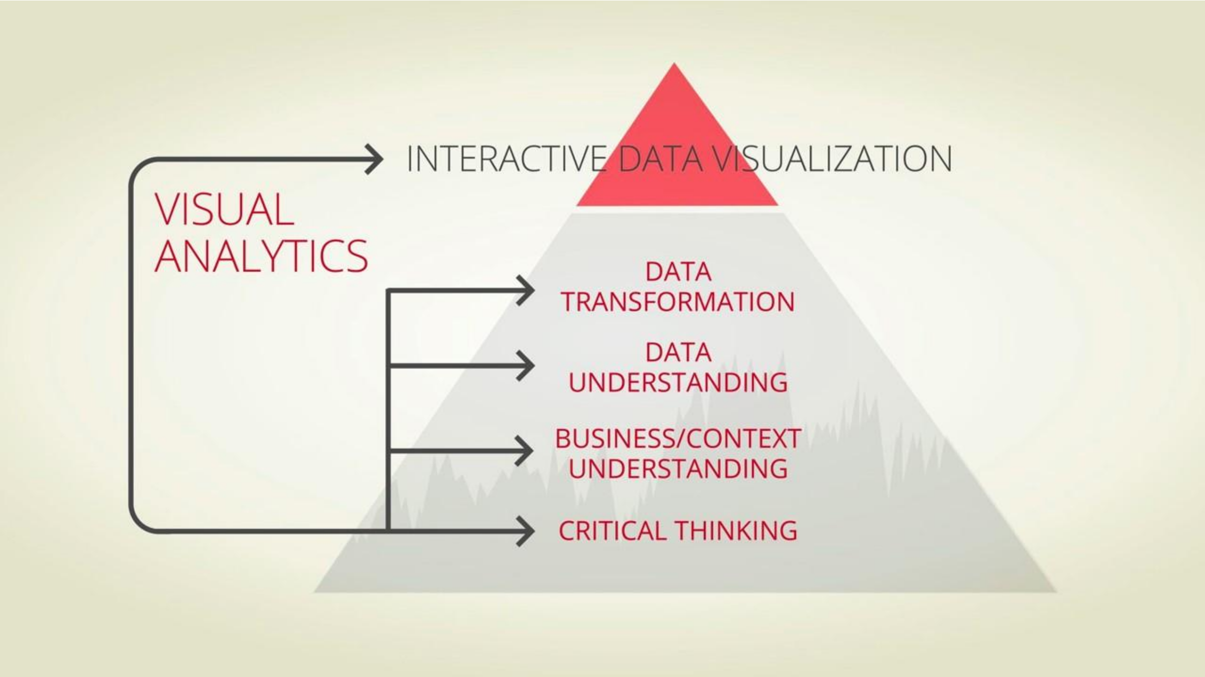
- The relationship between data visualization and visual analytics is symbiotic.
- Good data visualization enables visual analytics to be more effective and to show users better insights and better insights make for more compelling visualizations.



The role of visualizations in analytics

Data visualization can either be static or interactive. Static visualizations provide users with a single view of what's in front of them. Interactive visualizations enable users to drill down into data and extract and examine various views of the same dataset, selecting specific data points that they want to see in a visualized format.

Data visualization is what provides clarity to data-driven insights and it's what enhances understanding throughout an organization.



In this diagram, visual analytics is shown to be the foundation for interactive data, thereby demonstrating how the two are connected. Analytics acts as the source for data visualization and contributes to the health of any organization by identifying underlying models and patterns and predicting needs.

Broadly, there are three types of analytics: **descriptive, prescriptive, and predictive**. The simplest type, **descriptive analytics**, describes something that has already happened and suggests its root causes.

Prescriptive analytics takes things a stage further: In addition to helping organizations understand causes, it helps them learn from what's happened and shape tactics and strategies that can improve their current performance and their profitability. A simple example would be the analysis of marketing campaigns.

Predictive analytics is the most beneficial, but arguably the most complex type. It helps users to identify patterns that suggest future situations and behaviors. Using predictive analytics, organizations can plan for forthcoming scenarios, anticipate new trends, and prepare for them most efficiently and cost-effectively. Predicting forthcoming trends sets the stage for optimizing the benefits your organization takes from them.

Using visualizations to make smarter decisions
The data drawn from power visualizations comes from a variety of sources: Structured data, in the form of relational databases such as Excel, or unstructured data, deriving from text, video, audio, photos, the internet and smart devices. This data is gathered into either on-premises servers or increasingly into cloud data warehouses and data lakes. They are transformed into data visualizations and shared via dashboards and analytic apps so that users can make smarter, data-driven decisions.

Data teams and business and analytics teams are tasked with choosing and developing the best way to visualize data and to build well-organized dashboards in order to help end-users make smarter decisions. Dashboards need to be clear, quick to interpret, and easy to drill into to find the deeper insights when needed.

To achieve this successfully, you need a data and analytics platform that offers a powerful combination of visual analytics and data visualizations; with the capacity to handle huge volumes of data either stored on-premises, in the Cloud, or both; with the flexibility to integrate data from any source; and with the scalability for future growth.

Design of visualization applications:

It involves planning and consideration of various aspects likes user needs, data complexity, interactivity and usability. Throughout the design and development process, it is necessary to consider involving users and stakeholder for feedback and validation. Collaboration and user centered design are essential for creating a successfully visualization application that meets both functional and user experience requirements. **Following are the key design principles:**

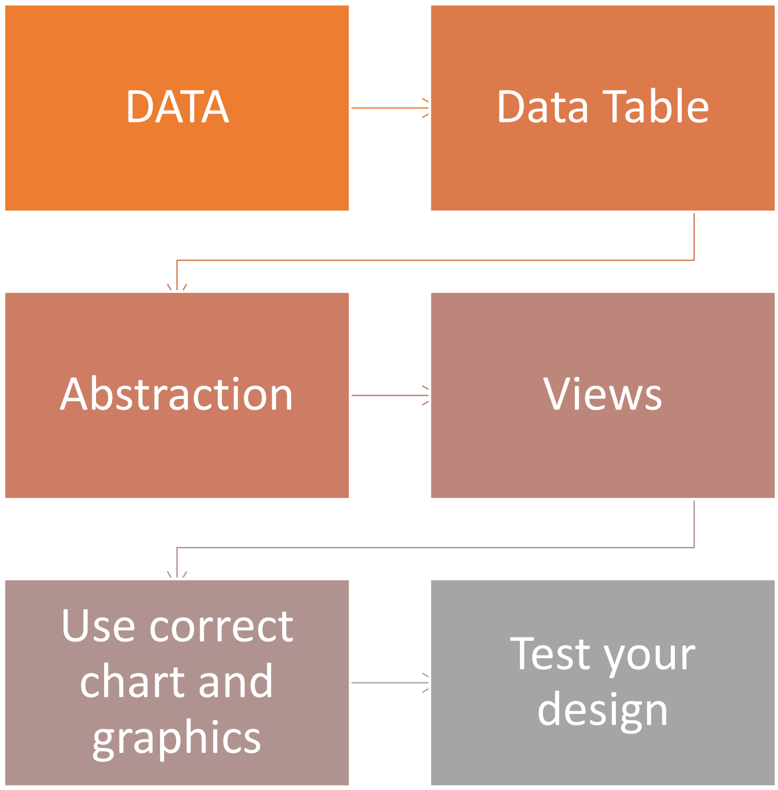
- Define objectives and users' goals
- Understand audience
- Select proper visualization: choose right types of visualization eccentrics for data like bar chart, line chart, scatter plot based on nature of data and user requirements.
- Plan data preparations: ensure data is cleaned, structured and stored in a way that is suitable for visualization. Use data transformation and aggregation before visualizing data.
- Visual consistency: maintain visual consistence across application i.e. use appropriate and uniform color scheme, typography and layout.
- Testing and iteration: conduct testing with real users to identify points and areas for improvement and iteratively refine the design based on user feedback.

- Continuous maintenance: plan for updates to keep application secure and responsive to user needs.

Following factors should be consider while designing visualization application:

- Use a consistent visual style throughout the application which will helps to understand the data and to make sense of visualization.
- Use clear and concise labels for the data
- Use a limited number of visual elements in visualization

- Designing a visual application requires a combination of technical skills, creativity, and an understanding of user experience. Understanding the Data and its presentation is crucial.
- Here are some steps you can take to design a successful visual application:



Define your goal

- Before you start designing your visual application, you should define what you want to achieve with it.
- Think about the purpose of the visualization, who the target audience is, and what features it should have.

Create a Visual Mapping

- A wireframe is a basic outline of the application's layout and functionality.
- It will help you visualize the structure of your application and how users will interact with it.

Choose a design style

- Your visual application should have a cohesive design style that reflects your brand and appeals to your target audience.
- Consider the color scheme, size, orientation, graph type, typography, and overall aesthetic of your application.

Use a grid system

- A grid system can help you create a balanced and organized layout for your visual application.
- It will also make it easier for users to navigate and find the information they need.

Test Your Design

- Once you have created your visual application, it's important to test it with users to ensure that it is easy to use and meets their needs.

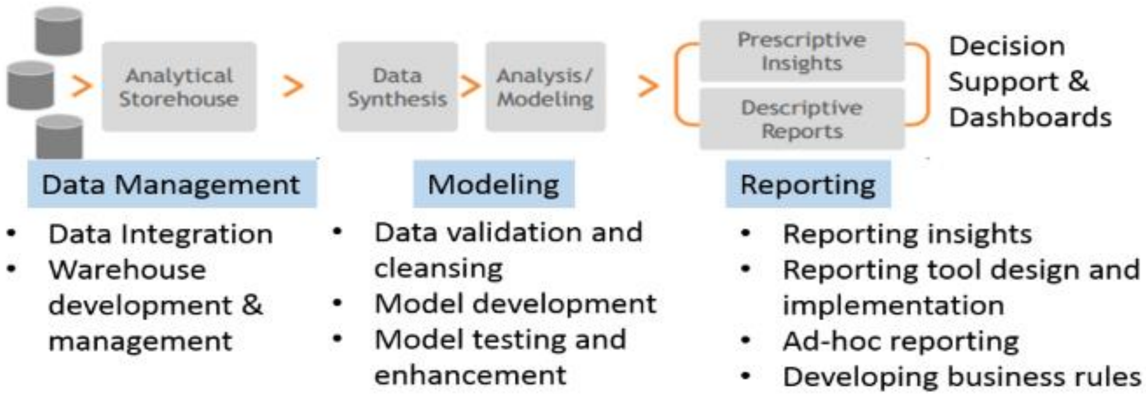
Designing a visual application is an iterative process, and you may need to make changes and adjustments along the way to create the best possible user experience.

Applications

- **Healthcare Industries**
- **Business intelligence**
- **Military**
- **Finance Industries**
- **Data science**
- **Marketing**
- **Food delivery apps**
- **Real estate business**
- **Education**
- **E-commerce**

Process To Design

- Data Understanding
- Data Cleaning and Preparation
- Data Modelling
- Data Mapping
- Choosing a right chart.
- Visual Design
- Presentation



Step 1: Define the question(s) you want to answer

The first step is to identify **why you are conducting analysis** and **what question or challenge you hope to solve**. At this stage, you'll take a clearly defined problem and come up with a relevant question or hypothesis you can test. You'll then need to identify what kinds of data you'll need and where it will come from.

For example: A potential business problem might be that customers aren't subscribing to a paid membership after their free trial ends. Your research question could then be "What strategies can we use to boost customer retention?"

Step 2: Collect the data

With a clear question in mind, you're ready to **start collecting your data**. Data analysts will usually gather structured data from primary or internal sources, such as CRM software or email marketing tools.

They may also turn to secondary or external sources, such as **open data sources**. These include government portals, tools like **Google Trends**, and data published by major organizations such as UNICEF and the World Health Organization.

Step 3: Clean the data

Once you've collected your data, you need to get it ready for analysis—and this means **thoroughly cleaning your dataset**. Your original dataset may contain duplicates, anomalies, or missing data which could distort how the data is interpreted, so these all need to be removed. **Data cleaning** can be a time-consuming task, but it's crucial for obtaining accurate results.

Step 4: Analyze the data

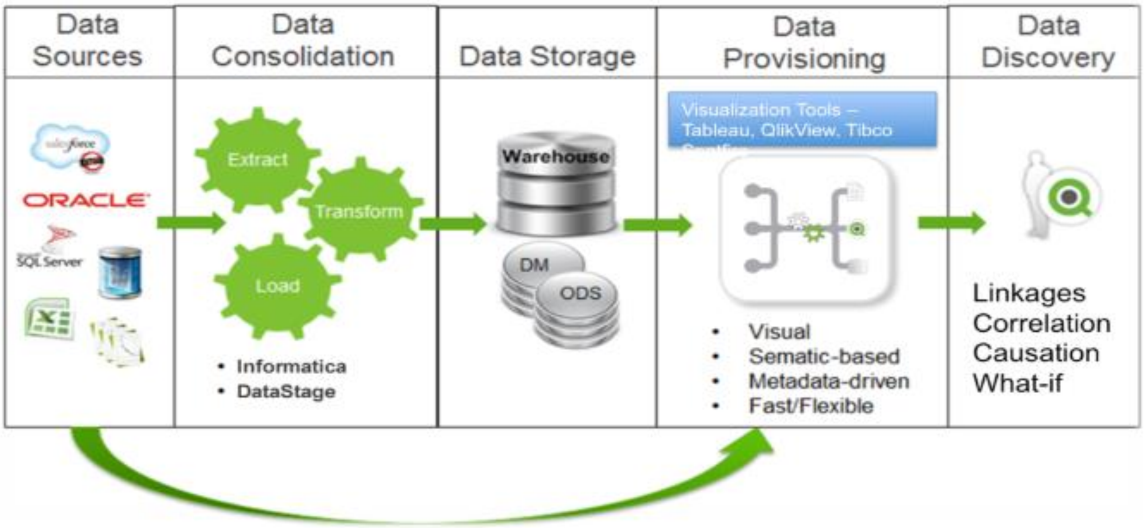
Now for the actual analysis! How you **analyze the data** will depend on the question you're asking and the kind of data you're working with, but some common techniques include regression analysis, cluster analysis, and time-series analysis (to name just a few).

We'll go over some of these techniques in the next section. This step in the process also ties in with the four different types of analysis we looked at in section three (descriptive, diagnostic, predictive, and prescriptive).

Step 5: Visualize and share your findings

This final step in the process is where **data is transformed into valuable business insights**. Depending on the type of analysis conducted, you'll present your findings in a way that others can understand—in the form of a chart or graph, for example.

At this stage, you'll demonstrate what the data analysis tells you in regards to your initial question or business challenge, and collaborate with key stakeholders on how to move forwards. This is also a good time to highlight any limitations to your data analysis and to consider what further analysis might be conducted.



Data Visualization Overview

Data visualization experts use large data sets to craft visual representations that display facts, patterns, and other relevant information. They're able to identify trends in large data sets and separate them from extraneous information. They then streamline their valuable insights into charts, graphs, infographics, or other visual representations that make the information easier to understand.

Data Analytics Overview

While data visualization and data analytics are different fields, individuals who work in these disciplines often work together. Data analytics experts focus on technology. These computer and programming professionals know how to manage and interpret large data sets for a number of different purposes. Data analysis experts might work in descriptive analytics, where they examine data over a specific period of time. For example, they could analyze sales for a company during a given quarter. They may also work in diagnostic analytics, which emphasizes finding causes for certain events, such as a drop in sales. Predictive analytics and prescriptive analytics are other possibilities. These two complex modes of analysis attempt to predict the future based on detailed data from the past.

Based on	Data Visualization	Data Analytics
Definition	Data visualization is the graphical representation of information and data in a pictorial or graphical format.	Data analytics is the process of analyzing data sets in order to make decision about the information they have, increasingly with specialized software and system.
Benefits	Identify areas that need attention or improvement Clarity which factors influence customer behavior Helps understand which products to places where Predict sales volumes	Identify the underlying models and patterns Acts as an input source for the Data Visualization, Helps in improving the business by predicting the needs Conclusion
Used for	The goal of the data visualization is to communicate information clearly and efficiently to users by presenting them visually.	Every business collects data; data analytics will help the business to make more-informed business decisions by analyzing the data.
Relation	Data visualization helps to get better perception.	Together Data visualization and analytics will draw the conclusions about the datasets. In few scenarios, it might act as a source for visualization.
Industries	Data Visualization technologies and techniques are widely used in Finance, Banking, Healthcare, Retailing etc	Data Analytics technologies and techniques are widely used in Commercial, Finance, Healthcare, Crime detection, Travel agencies etc
Tools	Plotly, DataHero, Tableau, Dygraphs, QlikView, ZingCHhart etc.	Trifecta, Excel /Spreadsheet, Hive, Polybase, Presto, Trifecta, Excel /Spreadsheet, Clear Analytics, SAP Business Intelligence, etc.
Platforms	Big data processing, Service management dashboards, Analysis and design.	Big data processing, Data mining, Analysis and design