

Workshop – Data Visualization: Methods & Tools



SESSION 1: Data Visualization for Strategic Marketing decision

INSTRUCTOR: Prof. Satyabhusan Dash

Prof. Dash gave an initial introduction about Data Visualisation where he mentioned “Data visualization is the presentation of data in a pictorial or graphical format. It enables decision makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns.” One can take the concept a step further by using interactive visualization technology to drill down into charts and graphs. He brought out the importance of Data Visualizations

as below:

- ✓ Identify areas that need attention or improvement
- ✓ Clarify which factors influence customer behavior
- ✓ Helps in understanding which products to place where
- ✓ Predict sales volumes

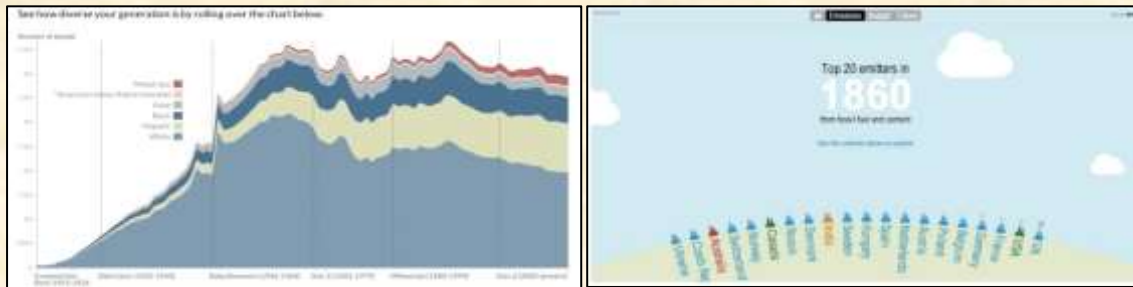
Data visualization gives business users the ability to use information intuitively, without deep technical expertise. One does not need to be an expert to draw inferences from the data, as before. Data visualization facilitates easy understanding of data values and drawing important concepts and ideas.



Prof. Dash, having been said the importance of Data visualization, further illustrated it with the help of two case studies using **Marketing Engineering software**. The first case study “Kirin Case” unfolded the understanding of marketing segmentation, targeting and positioning through data visualization. And the second case “G20 case” explained the usage of data visualization for explaining perceptual and preference mapping in a more comprehensible manner. With the help of these two cases, it was explained how data

visualization can result in simplification of available data points, promote the understanding of them through graphs and hence allowing an easy drawing of inferences from the raw data.

Some of the current practical application of data visualization was presented by Prof. Dash. Google music data visualization shows the change in number of listeners of different kinds of music in the last six decades. Second example illustrates, pictographically, the increasing diversity in the American population across generations.



Another example takes the data from the World Resources Institute's Climate Analysis Indicators Tool and the Inter-governmental Panel on Climate Change and illustrates graphically the top 20 emitter countries of greenhouse gases in 1860 from fossil fuels and cement.

Visualizations are the single easiest way for our brains to receive and interpret large amounts of information. Data visualization gives business users the ability to use information intuitively, without deep technical expertise.

SESSION 2: Using the R programming language for Data Visualization

INSTRUCTOR: Prof. Sameer Mathur

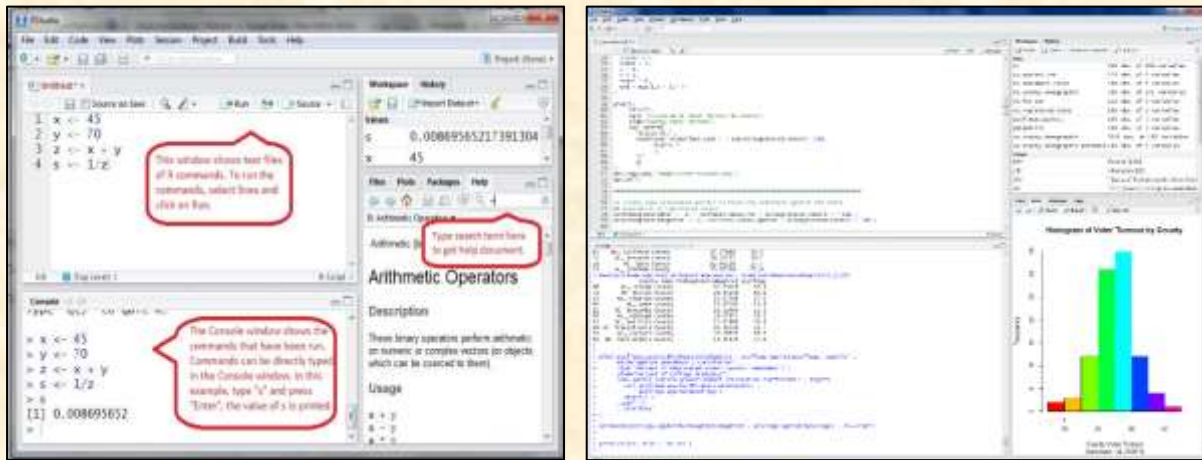


Prof. Mathur gave an overview of the R programming language and its usage in data visualization. R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, etc.) and graphical techniques, and is highly extensible. One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. It is used across the industry. It's an open source and free. Its syntax is similar to other programming language like C, Java and Pearl. It's a supercharged version of Microsoft's Excel spreadsheet software.

The capabilities of R are extended through user-created *packages*, which allow specialized statistical techniques, graphical devices (ggplot2), import/export capabilities, reporting tools ([knitr](#), [Sweave](#)), etc. These packages are developed primarily in R and sometimes in Java, C, C++, and Fortran.

RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. RStudio is also available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server or RStudio Server Pro (Debian/Ubuntu, RedHat/CentOS, and

SUSE Linux). The RStudio team contributes code to many R packages and projects. It enables R users to do some of the most innovative and important work in science, education, and industry. An R studio worksheet appears as below.



Prof. Mathur illustrated graph visualization using R. He explained how to visualize information using continuous and discrete variables. With the help of R codes and commands, comparison among groups using tables and statistical tests were observed. Prof. Mathur explained how R could be beneficial for obtaining, analysing and visualising simple statistical measures to complex linear models.

R is a cutting edge tool which helps in explaining its quick acceptance by the statisticians, engineers and scientists. They can improve the software's code or write variations for specific tasks. Packages written for R add advanced algorithms, coloured and textured graphs and mining techniques to dig deeper to database.

SESSION 3: Text Data to Insights – Creating a storyline with social Data

INSTRUCTOR: Mr Manas Ranjan Kar

Mr. Manas talked story line during a of data different parts. explained elements storytelling – posts on reviews like followed by the the story. He those topics or audience and are ideation of the

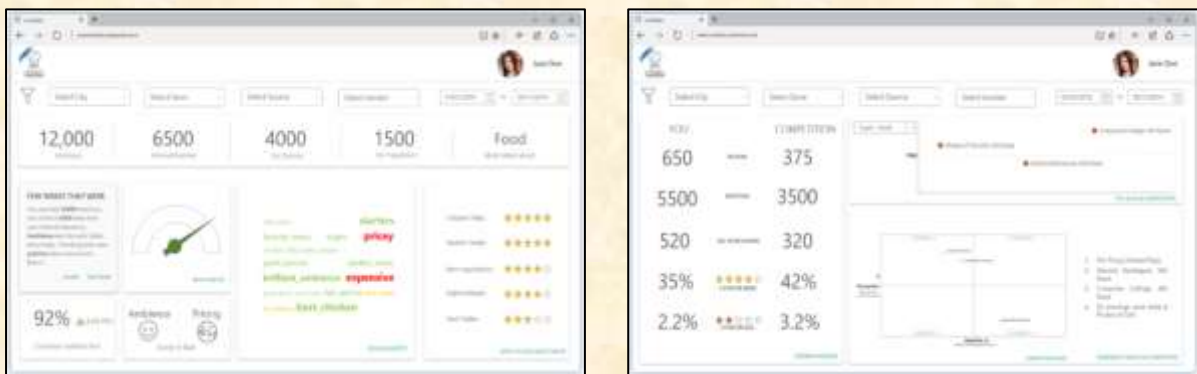


about the art of creating a presentation with the help visualization in five First part is where he the source of data for internal reviews, social FB/Twitter, Email, external Zomato, Tripadvisor etc. - process of ideation upon the emphasized on focusing on stories that may interest the relevant to them. A proper story requires double

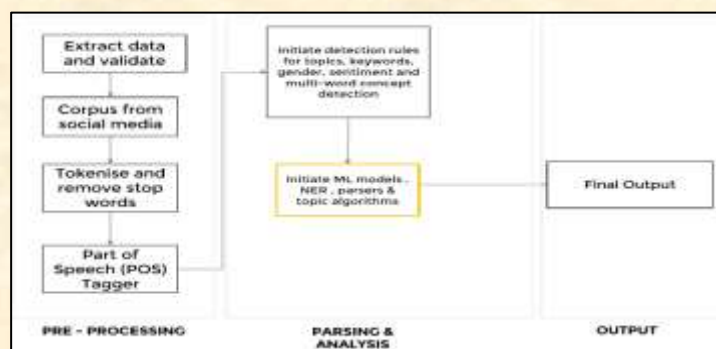
checking of the data that supports the story, focussing on one or two key statistics from the research. He advised to call out and highlight the data and to reduce texts while presenting a story.

Second part is about creating data driven narrative. Mr. Manas mentioned that typically these stories focus on how something is rising or falling over time. However, even a flattening trend can be a major story. Next logical step is to ask a question “why” such a trend? One needs to compare this trend with the trends of the peer and competitors. He mentioned that Tables are the worst kind of visualisation on a client desk.

Part three consisted of web crawling- drawing the data. The first dashboard should be a summary of the story, encapsulating the entire idea in few sentences. It should not just be all data but an automated text which can prove beneficial for layman. E.g. dash board of a restaurant review page www.lunchbox.subcortext.com.



Fourth part was focused on scraping data from the web. Acrawler is a program that visits Websites and reads their pages and other information in order to create entries for a search engine index. The major search engines on the Web all have such a program, which is also known as a "spider" or a "bot." He introduced two tools - import.io and webscraper.io – and demonstrated the real time working of the two tools. Fifth part dealt with text analytics. He explained the process of text analytics as follows:



Finally, Mr. Manas also demonstrated how to create one’s own free social media listening dashboard using Caravel. Caravel's main goal is to make it easy to slice, dice and visualize data. It empowers users to perform analytics at the speed of thought. It provides a quick way to intuitively visualize datasets by allowing users to create and share interactive dashboards. It is a rich set of visualizations to analyse the data, as well as a flexible way to extend the capabilities.

SESSION 4: Interactive Graphical Data Analysis through Tableau software

INSTRUCTOR: Prof. Manoj Dash



Prof. Dash brought out the difference between data analysis and data analytics. Data analysis is about estimation and data analytics is for visualization. Data analytics depends on big data. He described four types of Analytics:

1. Descriptive analytics- What is happening?
2. Predictive Analytics- What is likely to happen?
3. Diagnostic Analytics- Why did it happen?
4. Prescriptive analytics- What should I do about it?

Prof. Dash explained that Big Data is all about better analytic on a broader spectrum of data, and therefore represents an opportunity to create even more differentiation among industry peers. It helps the management professional by providing insights and depth understanding of various phenomenon, cause and effect, dimension and measures.



He defined data visualization as a graphical, animation, or video presentation of data and the results of data analysis. It gives us the ability to quickly identify important trends in corporate and market data which can provide competitive advantage. One can check the



magnitude of trends by using predictive models that provide significant business advantages in applications that drive content, transactions, or processes. As a data visualization tool, he introduced Tableau software to the class. Tableau is 10-100x faster than existing solutions. It's designed to support how people think. Tableau Desktop is a product that everyone can use to ask new questions, spot trends, identify opportunities, and make data-guided decisions with confidence. This Tableau session provided a deeper understanding of Tableau Architecture and concepts of Filters, Parameters, Graphs, Maps, Table Calculation and Dashboards. Tableau enables users to quickly see and understand their data by automatically turning data into pictures.



Tableau worksheet is easily navigable, even by a non-expert. Prof. Dash, then demonstrated the functioning of tableau with the help of various data sets. Sometimes Tableau may identify a field with a data type that is incorrect. For example, a field that contains dates may be identified as an integer rather than a date. One can change the data

type in Tableau by right-clicking the field in the Data window, selecting Change Data Type, and then selecting the appropriate data type. One can create parallelized dashboards, quick filters and calculations by using Tableau.

SESSION 5: Data Visualization and Data Mining using SAS

INSTRUCTOR: Prof Ashwani Kumar

Prof. Ashwani brought Mining Techniques by examples. The session Data Visualization and participants in the critical role of Data phases of Data Mining intelligence from



Data mining is a process (unknown) patterns in knowledge from large data mining techniques regression, market

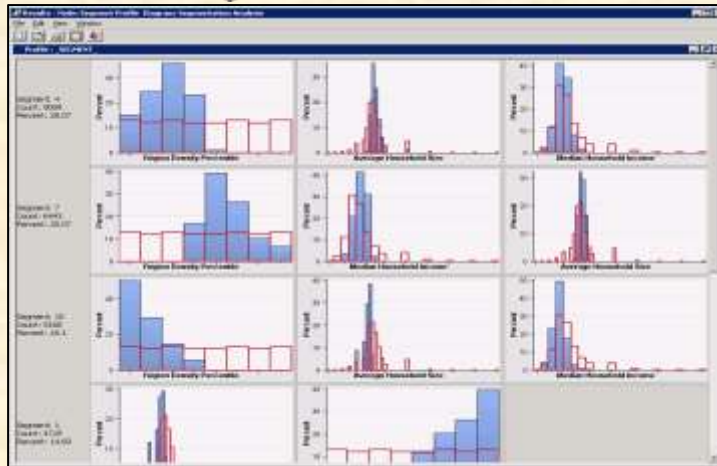
rules, clustering, neural networks etc.). SAS have their own data mining process known as SEMMA- Sample, Explore, Modify, Model, and Assess. Many of the steps in the SEMMA process directly correlate with steps in the CRISP-DM methodology.

Sample — identify input data sets (identify input data; sample from a larger data set; partition data set into training, validation, and test data sets). Explore — explore data sets statistically and graphically (plot the data, obtain descriptive statistics, identify important variables, perform association analysis). Modify — prepare the data for analysis (create additional variables or transform existing variables for analysis, identify outliers, replace missing values, modify the way in which variables are used for the analysis, perform cluster analysis. Model — fit a predictive model (model a target variable by using a regression model, a decision tree, a neural network). Assess — compare competing predictive models (build charts that plot the percentage of respondents, percentage of respondents captured, lift, and profit).

Data Visualization plays a critical role in Data Understanding and Data Preparation (CRISP-DM) and Sample, Explore and Modify phases in SAS SEMMA Process. The Final Step (Assessment/ Evaluation) is highly dependent on the Data Visualization techniques for visualizing and communicating the results. Bad/inappropriate technique may result in misunderstanding. Misunderstanding may cause an incorrect decision. It is important to consider that the Data Mining process is useless if the results are not understandable. A SAS output appears as below:

out the applications of Data using SAS software and case illustrated the integration of Data Mining and helped the developing an appreciation for Visualization tools in various Process for extracting business data.

of discovering and interpreting data that leads to actionable datasets through one or more (such as decision trees, logistic basket analysis/ association



He explained a decision making process by taking Grocery case study illustrating it through SAS.

SESSION 6: Visual storytelling: Using data and storytelling to get a message across

INSTRUCTOR: Mr. Naveen Gattu

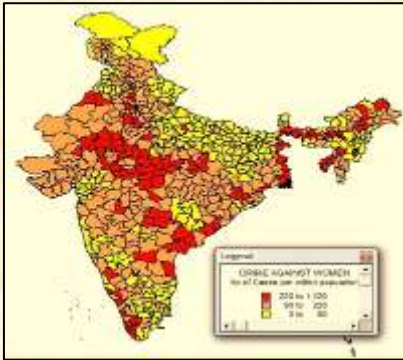
Mr. Gattu brought Science in drawing from an enormous introduction to He explains seeking, story- are three entities the consumption of and Consumer. structured and information for



out the relevance of Data understanding and meaningful inferences pile of data. He gave a Visual representations. humans are pattern-telling animals. There potentially involved in data- Data, Designer Data is processed, interpreted into consumers. Designers

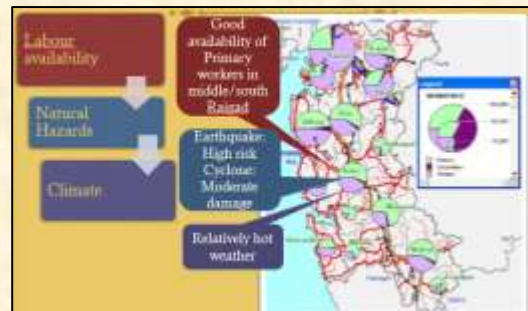
create infographics to persuade or explain a message to consumers. Designers are increasingly exploring data for its own sake, creating data art. He then explained principles of design. He views that charts are not monolithic entities but it is a mapping of data to visual encodings. It is the science of visual encodings that forms Cognitive principles of visual influence. Visual encoding vary in their effectiveness. The properties of data as follows: Quantitative (numeric), Categorical (grouped), Ordinal (ranked, time), Relational (hierarchy, influence)

Mr. Gattu explains that the key components of an engaging story is formulated by putting together the data narrative. By taking various Industry examples and Case study, he showed the power of data visualization through www.gramener.com. Gramener transforms the data into concise dashboards that make the business problem & solution visually obvious. It helps to find insights quickly, based on cognitive research, and acts as visualisations guide towards actionable decisions. He explained how Gramener has shaped tonnes of data to give meaningful information through data visualization which can be interpreted by even a layman. E.g. Amul's Advertisement analysis.



He also explained how to operate QGIS software - QGIS Software > Vector/Raster files > Databases for visualization. In another example, he showed the inter relationship between Labor force and Environment.

In addition he explained the usage of both proprietary as well as open source software with live cases studies to demonstrate visualisation applications. The talk also briefly covered the usage of multiple data sources for integration including operational/transaction data as well as secondary data.





A GLIMPSE OF SPEAKER'S SESSION

<https://www.youtube.com/watch?v=Q8GeMgvEia8> (Mr. Naveen Gattu taking session on Visual Storytelling)

<https://www.youtube.com/watch?v=WK0DLMpceb0> (Mr. Mohan Krishnan taking session on Geographic Data Visualization)

<https://www.youtube.com/watch?v=Ejzlh3rYyco&t=28s> (Prof. Sameer Mathur taking session on Data Visualization using R programming language)

<https://www.youtube.com/watch?v=CBZmiChBj4I>
(Mr. Sandeep Saxena, Director General, MRSI)

<https://www.youtube.com/watch?v=vq7G8xiYwNE>
(Mr. Omkumar Krishnan, IIM Kozhikode)

<https://www.youtube.com/watch?v=GqwOsS1qljg>
(Ms. Stuti Khullar, Hansa)

<https://www.youtube.com/watch?v=XdgatKHI6gk>
(Ms. Sakshi Mishra, GPI)

Feedback from Workshop participants



The workshop has not only proved to be beneficial in academic and industrial linkages but has updated us about the latest tools & techniques used in the industry.

Dr. Omkumar Krishnan, Indian Institute of Management, Kozhikode



The tools & techniques that have been exposed are really interesting and would surely help us when we start implementing them on data.

Ms. Stuti Khullar, Hansa Research Group



The workshop has been very interesting and would certainly be helpful in programming tasks effectively within lesser time frame.

Ms. Sakshi Mishra, GPI

WORKSHOP SNAPSHOTS





CME
Centre for Marketing
in Emerging Economies

Your suggestion (s) is important for us.



cmeerc@iiml.ac.in
Send us your feedback



[Facebook](#) Click here to join



[LinkedIn](#) Click here to join



[You Tube](#) Click here to join



Follow us on [Google+](#)



Follow us on [Twitter](#)

**NEWSLETTER CONCEPTUALIZED AND
DESIGNED BY:**

Ms. Ginni Bharti (Research Associate, CME)

**Ms. Swati Gupta (Research Coordinator,
CME)**

Visit: www.iimlcme.org

**CME, IIM Lucknow (Noida
Campus)**

**B-1, Institutional Area, Sector-
62, Noida, Uttar Pradesh-201307**

Landline: 0120-6678483