Lecture 27: Data Visualization: An Overview, History and Skills for Researchers

Dear Learners, Welcome to the session on data visualizations as part of the NPTEL course, and also the anatomy of visiting the research gaps in research. As part of the introduction, data visualization is an important aspect of research activities and it plays a vital role in data analysis and also representation of data research results. The proper data visualizations provide the insights, better patterns, and can help readers to understand the research results better. It establishes the communication between the data and patterns and also increases the research impact and engagement with readers. See in academic research and standards, require the researchers to use the data visualizations to understand the relationship, to reason about the scientific models, and to communicate the data to others and it can support it for the decision making in research by helping the researcher to identify the critical trends and patterns in the data. Let us understand quickly about what is data, what is the visualization, and what is data visualizations.

Data are the facts and numbers, letters and symbols that describe an object, idea, conditions, situations, and other factors. The data is a source of new knowledge and helps identify the novelty of a research. Then it is available in the both qualitative and quantitative format including numerical, textual, audio-visual, and multimedia, etc. So there are researchers who undertake the research by generating either qualitative and quantitative research data.

What is the meaning of the visualizations? See the visualization refers to a graphical representation such as charts, graphs, maps to present the research results and communicate the information. It has been an effective way to communicate both abstract and ideas of information. And what is the meaning of data visualizations? It is anything that converts the data source into a visual representation. It represents the data in graphical or pictorial formats such as charts, graphs, and maps. It is emerging discipline, data visualizations, and it involves the various fields such as information technology, natural science, statistical analysis, and graphics, and interactions, and data management, and perceptions, expressions, and geographic information, and information technologies means the data information is involved with the like the presentations and is also sourced at processing and display.

It can be part of the diagrams and it can be used in science and is also the visual communications and object usability means we can say that data visualization is multidisciplinary. It can be used for every discipline research. The data visualizations combine the three umbrella terms of data visualizations. The one is information visualizations and scientific visualizations and visual analytics are the infographics. Information visualization is the process of representing the data visually and

meaningfully so that the user can better understand it and it plays an important role in making the data digestible and also turning the raw data into actionable insights.

There are a number of information visualizations, the one is cartograms, concept mapping, information visualization, reference model, and graph drawings, heat map, hyperbolic three, multidimensional scaling, parallel coordinates, and three mapping etc. If you can see the examples it shows the real time internet how much data generated from the social networks like see here you have information here you have a data symbol of social networks and it is also giving that the icon of social networks means it present that real time information with the visual graphics. One can easily look at it and understand that each social network determines how much data is generated regularly. One of the best examples that we see is related to information visualizations and the second umbrella term for data visualization is scientific visualizations. See scientific visualization is a reference to a methodology of quickly and effectively displaying the scientific data.

The use of graphical techniques and methodologies to represent the scientific information and its many applications. It is a process of displaying scientific data using graphical techniques. This kind of visualization can be are using the in the academic research and the researchers have been creating the scientific visualization graphics using their data collected from the various methodologies and some of the tools of the scientific visualization are the computer animations and computer simulations, surface rendering, volume rendering and volume visualizations. The third information visualization umbrella is the infographics. The visual representation of information, data or knowledge intended to present information quickly and clearly.

It can be used to tell a story and clarify the complex information with evidence-based information and the research findings. So using innovative designs and research targeted the audience easily. Here is another example of science storytelling. See if you can look at this science storytelling image, our brain is actually clearly good at storytelling when it comes to the marketing and we are like the excitement, we trust interesting narratives more than simple messaging. Means in this the infographic is presented how storytelling affects the brain and how faster this and also the anatomy of usage of the top marketing assets and also formula for smarter contest.

But this way the infographics can be used for visualizing the stories of research. Let us look at the history of data visualizations. The history of data visualizations can be traced back to the ancient times when the people were using visual representations such as maps, charts, the diagram to show the cities and also the moons, sun and other planets. Although we may think that data visualization as a model is innovative, it has been visually recorded since the writing was invented. It started like a tick mark on the tables and K was as shown in the image, each group of tick marks with a slash representing a set of 5 like that making the total easier to sum.

The historical point of view about the data visualization states that philosophers have relied on visual metaphors to analyze the ideas and explain their theories. The data visualization has been used historically to identify the locations and the representations of quantities, distribution of geographical resources etc. However the modern concept of data visualization as we know it today began to take a shape in the 19th century with the evident new technologies such as computers and software tools and applications. In the 20th century, data visualization has been increasingly used to analyze and communicate the data in the various fields such as academic, business, finance, healthcare and science with modern various tools and technologies. Now why data visualization is important.

So when we say that research is a process of identifying the novelty at the end of the research, researchers used to communicate the research findings to the end user in a better way. The data visualization can say a lot more than words. The proper data visualization can provide better insights, detect the patterns and help the readers to understand the research findings in a better way. Data visualization is a powerful tool for cleaning the data, exploring the data structure, detecting the outliers and unusual groups, identifying the trends and clusters, sorting the local patterns and evaluating the modeling outputs and projecting the research results. The most important data visualization's information destiny.

Having the large amount of data in recent times, the data visualization is required for the information destiny. The amount of information present in the given space is the time by summarizing the large amount of information. And second important is attractiveness. Data visualization can visually appeal to catch the reader's attention and the specific research results. Sometimes the researcher also would like to use the dual coding through the data visualizations which can refer to a cognitive theory suggesting the information is a better processing and understanding when it is shown in the verbal and visual formats simultaneously.

It helps the brain to make quick connections. As shown in the example of a dual coding theory, see here we have a test in the dog, here we have a picture of the dog. When you are going to present the graphics using the dual code theory, the one can quickly understand the visual representation of the image. The other important data visualization in the research is the sketch noting. See sketch noting is a combination of simple illustrations with a small amount of test.

So when you are going to present a small amount of graphics, the sketch is an important way to communicate or present the research. See here is an example of sketch noting. If you can see that here we have a test in the blended learning at the same way the related graphics are given for the blended learnings and also the classrooms, the flipping the classrooms and this related graphics and as we move to the online course, we have related graphics means the sketch noting is another way to represent the simple information in

the graphical way with the both textual and graphical manner. Data visualization can help with the knowledge retention by making the data easier to understand and highly important information. The visualization can improve the decision quality by transferring the complex and abstract data into the clear and engaging graphics.

There are few examples given in the screens here. See this is how scientific research has been presented. Now we have seen the importance of data visualization in the research. Now let us look at the elements. See when we create the data visualizations, how it should be and what are the elements the one can look at for creating the good data visualizations.

A good data visualization comes from the design, analytics, tools and technologies. It should be clear and also correct and attractive. You may have the data but it may be the complex data set but if it is boring, it is not a good data visualization but if you have the complicated data odds, so again it is also not part of the good data visualizations. Since if you can only focus on the beautiful but if wrongly present the information, again it is not good data visualizations means you have to look at the presenting the data visualizations for creating the good data using the right functions and also the using the right design tools. It often involves the use of algorithms and software and technologies to turn the data into something that can be seen as such as a graph or the chart and animations and it is always called data visualizations should be elegant.

It might not be beautiful and in the era of information, visual and data literacy are necessary to produce the data visualization and communicate the research results. Therefore, creating a visually appealing and effective data visualization requires the design skills and understanding design principles, layouts, typography and color theory. So let us look at what are the data visualization skills that are required for the researchers. The researcher should know the data visualization best practices including technical proficiency, the visualization designs and principles, data analysis and interpretations, programming languages, statistical languages and storytelling with the data, data communication skills, effective presentation skills, etc. So let us look at what are the major skills that are required under each category.

One means technical proficiency. See under the technical proficiency researcher should know the ability which also should have the ability to effectively create design and communicate the research results through the various representation of data. Means they should know data visualization tools such as Tableau's or Power BI or online tools of data visualizations such as Canva or the OOS viewers etc. And the visualization design principles. Knowledge of data visualization design principles such as color theory, typography and layout is important for creating visually appealing visualizations.

The other most skills that are required for the researcher are data analysis and interpretation skills. See proficiency in analyzing and understanding the underlying data is very crucial. Say every researcher does collect research data but the very few researchers only know how they can interpret and analyze the data using the various statistical techniques. This also involves data cleaning, organizing and summarizing the data to extract the meaningful insights. The other skills that are required for the researcher are programming languages.

Some of the data visualization tools may not be required for such a programming language but there are advanced data visualization tools which have the good features required for the programming languages. So tools like Python, Java, R are common and require the minimum of programming skills for the researcher. And another skill for researchers is statistical knowledge. See statistical knowledge and understand the concept that they are creating meaningful visualizations. This includes the knowledge of measures of central tendency and the dispersion, correlations, regressions etc.

Researchers should know how to data storytelling. The ability to tell the compiling story through the data is a very key skill. This involves not just creating the visualizations but also interpreting and communicating the insights effectively. And finally the skills that researchers should have available to the data presentations. The ones after creating the data graphics and presentations they should know how they can evaluate that.

See means it is a process of collecting and analyzing feedback and the data presentations using the quantitative and qualitative methods. These allow assessing the effectiveness and the efficiency of the data presentations, identify the strengths and weaknesses and discover the opportunities and challenges for implementation. With these skills researchers can create meaningful and impactful visualizations that effectively communicate the research findings. Going forward to the data visualization skills and interesting data visualization topologies that we found through this literature as part of the data visualization literature should know the visualizations that have been classified with the insights, insight needs and data scale, analysis and visualization and graphic symbols and graphic variables and interactions. Means here as a researcher under the insights needs they should know the what are the categories and clusters are available for creating the data visualizations and also they should know the orders, rank and slot, distributions, the types, distribution types of data visualizations and comparison, trends, geospherical and positions and correlations.

These are the major types of data visualizations. We will be discussing this data visualization in the later sessions and again the researchers also know what are the data scales. It may be the nominal, ordinal or interval and ratio and when you go for the analysis they should have the data visualization literacy and that the analysis are statistical analysis and temporal analysis and geospherical analysis, topical and rational

analysis. See after you analyze the data then you have to move from the visualization means how we are going to visualize your data using a table method or you can use the charts or one can create the graphs, maps and tree and network and also graphical symbols. So you can use the geometric symbols, maybe the points to identify the values of the test and value of the research and you can use the lines or the areas, surface and volumes and also you can use the linguistic symbols.

It may be the test and numerical and functional marks and there are pictorial symbols. Maybe the images, icon, graphs means researchers should have available the various types of graphic symbols that can be used for presenting the research graphs. And another literacy skill are the graphic variables. The graphic variables include the spectacle, maybe the positions where we have and retinal and forms, colors, optics and motions and after you create that the research also should have the awareness about the view and search and locate and filter, details and design and history, extract and link, projection and destruction. Means these are the topologies the researchers should have available before creating the data visualizations. Now let us look at some guidelines for better data visualizations.

Our ability to visualize scientific data has been involved significantly over the past 40 years. However, this advancement does not necessarily include many common fit walls in visualization for scientific research. To address these issues the few guidelines would help to create the better data visualizations for scientific research or show the data, clear data findings, the researchers should have the clear data findings and they clean the data before visualizing the data and when you go for visualizing data reduce the cluster, avoid the too many details, effectiveness, decoration in the images and integrate it with the graphics and test. Graphics should be both tested and also be the graphics. It means that you know it doesn't mean that your graphic or your tables are the graph.

It's filled up with only tests or filled up with only graphics. It's a mix between both graphics and the test and also the rough, avoiding the charts that refer to the type of diagrams and the many overlapping and the taglines. You should just avoid the overlapping taglines. And color combination is one of the most important aspects of the data visualizations. Just look at the colors that better visualize for your research results and most people being used for the gray that's been better and also depending upon the what the research results and how research would like to be present.

At the end you should have the better tool and the right platform. Nowadays you will have many data visualization tools and also the platforms especially online open excel platforms. Now you will find many platforms. Those platforms and tools we are going to be explaining in the next session. If you can know those platforms and the tools, you will be creating better data visualizations without having much statistical skills or much programming skills.

Some guidelines we have given for the anatomy of presenting research graphs in the research. At the end of the research you will be presenting your graph in the research findings. Maybe in the TCS or in an article. Now how the new research graphs at the table should be? You know you should look at that overall dimension.

I mean which dimensions you wanted to be put in. Sometimes we have seen the example of putting the images in a very small space. We can't clearly look at the images showing that. Now we have to look at the overall dimension which should be the minimum dimensions of presenting the graphs in the research and also put the figure number. Now most of the researchers should know the API, the guideline for presenting the figure numbers and also the table numbers and also give the titles, the graphic titles. It may be the qualitative data or it may be the responses.

There may be some findings. It may be some of the results. It should have the titles. If that has the subtitle also, you can also have the subtitle. And the graphics always can give you that in a manner that the access tables and labels. You should know that when you're presenting the graphics, what should be the titles and how that label would be presented like X, Y.

So if you have the graphic in a manner, you should have some labels here. You should have some labels here. You should have some labels here. And the Alex lines are tick marks and grid lines. You should have – you can put this on the grid lines and sources of – and notes.

If you are going to borrow the image from a particular exposure, it is always good to have the source. These sources see this is for these images. And some of the researchers also use that it's a researcher, so the one source is primary data. And also put the notes. If you want to explain about the images, you can also put the notes.

So researchers can look at the notes. That will also help the researchers to understand what this research graphics is about. And other things, it's a logo and legends and data markers and data labels and data series. And this early autonomy is not like we can use the early arrays for a single image. Most of the maximum like overall dimensional figure number, title, subtitle, tables, labels, and lines, grid lines, source, some of them are the mandatory researcher has to be used.

And the others are dependent. Others are the depends like logo, legends, data labels, data series, data makers. Some of them are dependent. It's also depending upon what a researcher would like to present in their graphics. So some of this session, what we have covered in this session are what's meant by the data visualizations. And it's right beginning from what's meant by data, what's meant by the visualizations and data visualizations.

And there are three types of data visualizations, that is infographics or information visualizations and scientific visualizations. And also we have briefly looked at these history of data visualizations. It's given the clear indication that it's data visualization, it's not an emerging area. Though people consider this in the emerging area by using the recent tools and technologies, it's been used in the ancient time because the ancient time for the most of indicators and identifying the locations have been used with the visual maps. And we also look at what is the importance of data visualization in the research.

Because in modern scientific research, data visualization has a highly impact when you're presenting the research findings. Because the research is the one impactful insight that people may not understand if you're not presenting those research findings in a meaningful and graphical manner. So therefore, data visualization has a high impact, creating impact for your research, and better data visualizations. And at the end, we also see some of the researcher's skills. So being a researcher, we may be expert in the particular exercise of the subject because someone is expert from physics, someone is expert from science, someone is expert from chemistry, someone is expert from the social sciences and arts and humanities.

But data visualization skills are multidisciplinary. Every researcher should have those skills, at least minimum skills to analyze their data and visually present their data in a better way. And we also give some guidance. If a researcher would like to present their research data visualization in a better way, what are the minimum guidelines? So with this, we're going to conclude this session as of now. In the coming session, we will be covering the data visualization types, tools for data visualization and evaluation of data tools and also data visualization workflow. Thank you. .