LECTURE 28 : Data Visualization : types, tools and technologies

 Dear learners, welcome back to the second session on data visualization as part of the NPTEL course, science communication, research productivity and data analytics using open source software. We have seen in the last session an overview of data visualization and what is data visualization is and also we have covered the why data visualization is important in the research and elements of data visualization and the data visualization skills that are required for the researchers. In this session, we will be covering the data visualization types, tools and technologies and also we want to look at how we can evaluate the data visualization tools and data visualization workflow in the research activities. Data visualization is a field that encompasses a wide variety of formats and types of images. All the information visuals can be divided into 4 distinct categories which falls within the 2 into 2 grids. As source in the image, there is a conceptual and declarations idea illustration is one of the type of data visualizations.

 In this type of visualizations are used commonly in visual metaphors to illustrate the complex ideas and releases. For example, we might have drawn the 3 diagrams to show how family members are related or show the hierarchy of an organization. The aim of this type of visuals to simplicity, they are used to make a complex concept or set of relationship more simplistic and easier to digest. The another type of data visualization is conceptual exploratory idea generation.

 This type of data visualizations are often informal, sketched out as a draft to explore the connections or the relationship of different elements, ideas, relationship and patterns. You may develop the several visuals to explore the different ways in which data might show the trends, sketched out different perspectives of same main ideas. The another type of data visualization is data driven to declarative. See this type of data visualizations we encounter every day in the life. In newspapers and magazines and the internet and PowerPoint presentations, we used to use this kind of data visualizations.

 These are typically line and bar charts, supply charts and scatter plots that tell us the story of specific data sets. They are often based on the survey or pulling off and also research results. And this type of charts are typically straightforward and factual and self-explanatory in the nature. The fourth type of data visualization is data driven to exploratory. This type of visuals frequently combines the lots of different data sets.

 They are often interactive and allowing the viewer to make the adjustment to different part of the data and then evaluate the impact of those adjustments. There are several numbers of data visualizations when I have gone through the literatures and reports and also the book called the better data visualizations, a guide for scholars, researchers and works. The book is that probably more than 80 visualization data types which can broadly categorize these seven types like comparing categories, times, distributions, geospaticals, relationship and part to all and qualitative. Let us explore each type of data visualizations and how many

types of data visualizations can be created under this broadly categorized types. The first category is comparing charts.

 When we say that comparing category, it is used to compare the values between the groups. When you have a data with multiple variables and comparing categories, charts can be used to distinguish the differences between the variables. These charts are used to compare the part as a whole and visually in percentage with all segments equal to the other percent. There are number of types of visualization under these categories like the bar charts, parallel charts and stocks bar, dwelling bar, dot plot and marimoko and masko charts, unit, ischikko type, wafla charts, heat map, gate charts and bullet charts and bubble comparisons and also sankya and wall profile. Bar charts is one of the most familiar data visualizations.

 The simple bar chart is perfect for making the comparisons across the categories like comparing the populations across the country. This is one source that total population in the 10 countries as source in the image. You can look at that how we can compare the different countries of total population of the countries. The second chart is parallel chart. The simple bar chart is perfect for making the comparison across the categories.

  It is like comparing the populations across the countries. But if you show the comparison not just across also within the countries, the parallel chart is one of the good options. The parallel chart and easy to read and also share the baseline make the easy to make the comparisons. This example here is given the men is in a different color and women is in a different color. The parallel charts within the country you can compare what men is and also what women are familiar with the most readers.

  The another variation can bar is stacked bar charts. The while the parallel bar chart shows to have more values of each categories. This chart subdivides the data within the each categories. This categories could sum the same total says 100 percent so that total length of the bar is the same of every groups like it is shown in the image the social expenditure for under OCD countries. You can see that in a single bar you will find that within the country health old age and others.

 The another chart is differing bar. Divering bar chart is particularly useful when we want to highlight and compare the positive and negative values within the data sets. This kind of bars can show the both positive and negative values. These are how often used call that the liquid scales as shown in the image. Image is showing the responsibility of the government to reduce the defect in income between the people with the high and low countries.

 As it shows it can cover the strongly disagree, agree and strongly agree and disagree means within the image one can look at the positives and also the negatives details. The another one is dot plot. The dot plot is an easy way to compare the categories especially the many categories. See the dot plot is places a dot for each data points and connect them with the line like the examples called the program for international student assessments. We already

can see that how the maths and the readings have been connected with the dots with the different countries.

 Another chart bar is the Marimoko and Moesko charts. This type of charts is useful when we want to make the comparison between the two variables. One comparing the categories and one showing how they sum of your totals. See these are used to visualize the categorical data over the pair of variables and this kind of scales with the width of the bars in a bar chart corresponding to another variables. The color can be used to highlight the specific values.

 As you can see in the examples the high populations and high poverty you can see that the how it is been compared with the different colors like it is a sum up with the individual countries. The another category charts are units, isotypes and wafla charts. The units charts are shown the count of variables and each symbols can represent an observation of a number of unit. We can see unit charts to show the percentage dollars or any other numbers. Isolate charts are the subclasses of unit charts that are used the images or icons instead of simple shapes.

 And wafla charts are another sub class of unit charts. They are especially good for visualizing a part to all the relationship. Another category of charts are heat map. Heat map is say table with a color coded cells. They also often used to visualize the high frequency data or when seeing the general patterns it is more important than the exact values.

 It is often used to visualize the complex data sets and identify the patterns trends and variations in the data. You always see the example given in the here. It is like the composition of the total incomes. Like you have a countries and also the income from the different sources the earnings and self income and capitals means every individual type of income it has given the different colors with the boxes and gauge and bullet charts. See gauge charts is look like a speed motors and this kind of charts are familiar and easy to read.

 Usually to display the single values of data in the qualitative way. The bullet charts are often used to show how the performance of a metric and compared to predefined targets. A bullet chart is a bar marked with the extra encoding to show the progress towards a goal and performance against the reference line. And another types of iterations using the compare category how a bubble comparison chart is a way to present the complex data sets using the bubbles of different size to represent the different data sets. A bubble comparison charts can be engaged and interested but it can also be hard to design the values.

 Sankhya diagrams this type of charts are especially used for showing how categories compared to one another and flow into each states and categories. It particularly useful for illustrating the distributions of resources within a system and understanding the relative importance of different components. Waterfall charts. A waterfall chart shows a basic

mathematical equations adding or subtracting the values from some initial values to produce their final amount. It is also shows how sequential values affect the initial stage values.

Let us look at the other type of iterations under the time series. See the time series is again another type of the relation that we can create for comparing the data during the period. It shows the data over the period of the two display. It is also used to identify the trends and changes over the time. In the time series each data point is associated with the specific time period which allows the chronological organizations of the data.

When you look at the number of types of data visualizations under the times are the one is line charts and circular line charts, slope charts, spark lines, dump charts, cycle charts, area charts and stacked area charts, steam graph, horizon chart, ground chart and also flow charts and timelines and connected scatter plots. Let us look at the how we can use these charts in our research activities. When we say that the line charts which is commonly used and it is also referred as a line graph or line plot connected the series of data points to using a line. This chart of time present the sequential values to help you identify the trends. The line chart is easily to be clear and it is a representation and easily done.

The data values are connected by lines to show the values over the continuous period and also tracking the trends and patterns. As you can look at the images it is given by the time like we have the data year to year change in USK spendings. You have to see that this is how you always see the year is given like 2001 to 2015, here 2001 to 2015 and this is how you can track that how year to year change in the US healthcare spendings. Another charts under the time series are circular line charts. It represent the information in the circular or a radial format and can be useful for plotting cycle data series.

It can help highlight the cycle patterns and also the repetitions or absence of repeated in the time. It can be used for displaying the cyclical data of several series. Here you can also see the example of percentage of year visits. It is given like they are given here in the example you have a month and you have a center for this is the percentage of year and another example percentage year visits in the flows during the 2014 and 2017.

Slope charts. Slope charts are simple graphs that quickly and directly show the transactions and changes over the time the values and rankings. We can easily see that relative values of each data points it can also be used to show a before and after story of different values based on comparing their values at different point of time. Here is an example how we can slope chart can show that. This is the biggest changes in the unemployment rates. You can always see that from January 2000 to January 2018 how we can quickly compare each kind of data and before and after.

Sparklines. A sparkline is a tiny chart in a worksheet that provide the visual representation of data. Use the sparklines to show the trends in a series of values such as seasonal increase

and decrease or health spending data and economical cycle are also highlighted maximum and minimum values. There is an example given in the healthcare spending in selected countries. You see that how that series of values of different countries being presented in this given graph like Australia, Canada and Japan.

It is a duration during the 2000 to 2015. Another chart is bump charts. See the bump chart is used to show the progression values at dimensions with respect to another dimensions over the period of time. It follows the ranking systems where it changes in the rank of particular field values are noted over the time. They are often useful for exploring the changes in the rank of value over the time dimensions or place dimensions or some of the dimensions relevant to the analysis.

Another chart is cycle charts. See cycle graphs are the charts that typically compare the small unit of the time such as a weeks or the months across the multiple EFM. See they are most commonly used to display among the seasonal trends and also it is known as a cycle diagrams and it is a visual representations of a repeating process. The cycle plots are small comparison charts that consolidated the charts using the same scale. Another type of chart is area chart. The area chart is known as the mountain charts.

It is a type of data visionations that combines the appearance of a line chart and the bar chart. It displays the quantitative data using the colored area to represent the cumulative values of a data over the time or the across categories. Stacked area chart. Stacked area chart build on the typical area chart by showing the multiple data series simultaneously. See it display the evaluation of values of several groups and the same graphic, stream graphs.

A stream graphs is a type of the stacked area chart that display the evaluation of numerical values for multiple groups. It displays the data in a flowing organic shapes to represent the change in values over the times or the across the categories. The other chart is origin chart. The origin chart is a two dimensional data visualizations that display the quantitative data over the continuation interval and this kind of charts are used for identify the trends and extreme values within the large data sets. Origin charts are commonly used for in the financial analysis, stock market, data visualizations and other fields where tracking changes over the time is most important.

If you can look at that image is showing the example is change in the health spending as a percentage of GDP. If you can look at the in the particular of year we can track the GDP here which is minus 2.0 percent to plus 2.5 percent. You know different countries like the Denmark, Canada and France, Germany, Japan this how we can use this origin chart to to track the how GDP is growing means if you have any data which is going to be a track like the financial or budgeting data those kind of data if you would like to track and the original can be used.

So another important chart and commonly we as a research are used for the grand chart.

The grand chart is again the horizontal bar chart that shows a project schedule to task over the time and it's been used when we as a researcher we can also use this grand chart to schedule the research process. It is widely used tool in the project management and grand chart after use it for schedule tracking devices for example tracking the different phases of the research progresses and also projects and budgets. Another type of data visualization is the flow charts and timeline. See again the flow charts and timelines are two examples of array of visuals that can show the changes over the time or different kind of processes sequences and ever keep the timeline source the when certain events are taken a place.

Again the flow chart shows that the hierarchy of the management the structure of any organizations and how the workflow from the top to down. If you can look at the example here we are given the one for the with timeline the another is like in the flow chart during the time you can look at the how supplemental nutrition assistance the programs during 2010 and another is see the workflow the how does disability insulin system works means would like to show the your concepts of framework in a in a workflow or in everything time period how you are developing your research progress this flow chart and timeline will going to help you to present those data in a good manner. Another type of visualization is connected as scatter plots it's a combination elements of both scatter plots and line graphs. This type of visualization is often used to show the relationship and trends between the two variables over the time. It display the evaluation of numerical values and it represent the data points with the dots and connect them with the line segments.

You can see that the example what we are given here see the one is a life expectancy in the South Africa and during the 96 to 2016 you can you can see that you know connected spotlights like these are the dots are which are connected the how that the life expectancy in South Africa being over the period like you can also see the example of per capita GDP in the South Africa here is a connected plots over the period to look at the GDP growth we have seen that types of data visualizations under the categories and also time series and now let us look at the what are the types of data visualizations under the geospatial visualizations. See geospatial visualizations focus on the relationship between the data and it's a physical locations to create the insights and geospatial visualizations highlight the physical connections between the data points it represent the importance of locations data by providing the visual context and the maps are more primary focus for geospatial visualizations and regularly use the maps to do the visual representation of space. I mean this kind of maps when you go to focus the geographical locations are highlight the number of within the country avoid you can differentiate by the country wise and geospatial visualizations can help you for creating such a visualizations. Certain types of visualizations under the geospatial visualization category the one is the cholaplet maps another is the cartograms and propositional symbols and dark destiny maps and flow maps. See when you say that cholaplet maps it is commonly used and most familiar and extremely popular and probably the most common thematic maps used recently.

See it maps use the colors and sets and patterns on the geographical unit to show the

proportionate quantities and magnitudes and it's provides the way to visual values over the geographical area which can show the variations and patterns. There are number of examples we are just given the one example here it is like in a Google map of the world if you have any data related to look at that you know if you are conducting this survey then you have to identify you have to visualize this survey the respondent from where they were geographically located means you got me some survey responded from the India you got some of them are USA you got some of them are United Kingdom or you got some of them are some other Europe countries. See when you wanted to highlight this you can also use this kind of cholaplet map to visualize those content and display in your survey respondents. The another kind of geospherical visualization is cartographs these are popular techniques for representing the social data and it's also called as a thematic map in which mapping variables such as a travel time populations or GNP is subtitled for the land area and distance. The types of cartographs see when we say the types of cartographs there is a continuous cartograms and non continuous cartograms and graphical cartograms and a graded cartograms and topology.

 Another type of geospherical data visualization is the propositional symbols and dosh destiny maps see the propositional symbol maps is a type of the map that indicates the symbols of the change in the size based on the values they represent. The propositional symbols of the maps are typically used to show the counts and the totals that can be used to visualize the classic or unclassic data and different shapes of objects like lines, arrow, points, circles and more can be placed on the objects and sized according to the data values. See here we have example and per capita GDP in Europe we can see that how that the dots been used to identify the per capita in the Europe's like you have box with kind of dots here and you have a circle kind of dots here. It's also this kind of dots also depend upon the size means if it's the bigger size of dots the per capita of the continuous are more if it's a lesser size of dots it's the per capita GDP of the countries are very less. The another type of maps this is a flow maps the flow map geographically shows the movement of information object from the one location to another.

 There are different type of flow maps including the radial flow maps and the distributive flow maps. Their sequence of chain of events or processes are mapping the relationship between the stage and sub stage because this kind of flow maps we use the regularly like it's like the Google map what we use the regularity go to the one place to another places means a Google map direction is one of the kind type of the flow maps. Another type of data visualization is distributions. Distributions is a method of display frequency and how data spread out over an interval and is a group. This data charts are used to illustrate the correlations between the qualitative data points and distribution analysis in identifying the values such as a mean, medium, range, outliers etc.

 The distribution shows the relationship of a single variables over a set of range and categories. There are certain types of charts can be created under the distributions are histograms, pyramid charts, visualizing statistical uncertainty with the charts, box, whisper,

candlestick charts and valium charts, visualizing uncertainty by showing the data, steam and leaf float. Let us look at the what is mean by the histograms. The histogram is one of the most basic graph type for visualizing a distributions and it divides the entire sample into the intervals and it also known as a bins. The eye of the bin shows the number of observation within it and it is a type of the bar chart that display the frequency or count of a data points within specified intervals or the bins.

See here is an example it shows in the screen it is showing that the two histograms which is the distribution of earnings of men, the men's earning distribution 2016 and women's earning distribution 2016. The second one is pyramid charts, it is a subcategory of developing bar charts the most often to show the changes in the population based on metrics such as a birth rates and ages or overall population levels. The pyramid charts puts two groups on either side of a central vertical axis. The advantage of pyramid chart is that we can assess the overall shape of the distributions because of the both group sit on the same vertical baseline. See again another example shows here the distribution of age in United States and Japan in 2016 in both graphics the women are represented on the left of the vertical branch and men on the right.

The another visualization type is the visualization of statistical uncertainty with the charts. There are a lot of types of uncertainty in data and statistics. Statistical uncertainty can be visually visualized with the charts and graphs in a several ways. The type of chart usually depends on the data type and messages. Then we can think that the term of uncertainty in the two main ways the one uncertainty from a randomness another one is uncertainty from the unknowns.

There are some way of visualizing the uncertainties like the error bars confidence intervals and fan charts and hand on looks etc. The visualizing uncertainty and variability can help to the audience to better understanding and interpret the data sample results. The box and risk per plot the versionally called as semantic plot it is used in a box and line markers to show the specific percentage values within the distributions. Then we can also add the markers to show the outliers and other interesting data points and values in such kind of visualizations. It is a compact summary of the data distributions though it display the less details than the histograms or the variant charts.

See this kind of data charts as consists the rectangle and two lines and also the flow from the top and bottom of the box and dots for outliers and other specific data points. The another type of chart is the candlesticks or stock charts look like a box, whisker plots but they are visualizing the different content. Candlestick charts visualize the changes in the prices of stocks or bounds and securities and commodities over the time. Candlestick charts are one of the most popular component of technical analysis and enabling the traders to interpret the price information quickly and information just a few price bars. Violin charts it is also known as a violin plot and it's a statistical graphic that compare the probabilities distributions.

These charts are widely used in a different fields such as scientific research, finance and industry. They can be used to visualize the data in the various domains like biology, economics and social science and etc. And the red line plot is a series of histograms are destiny the plot shows that different groups aligned along with the same horizontal axis and presented with a slight overlap along with the vertical axis. The example shows here the earning distribution across the 13 different countries like the horizontal axis is a shared across the 13 countries and distribution sometime overlap along with the vertical dimensions.

Visualizing uncertainty by showing the data. See, visualize the uncertainty in data can help the people understand and interpret the data's sample results. There are different types of charts and graphs can be used to visualize the uncertainty and variability depending upon the data type and messages. That techniques for visualizing the uncertainty data include the script plots and rank load plot and base warm plot with sun plots etc. The steam and leaf plot is the table that shows the place value of the each data values. And these are especially used for small to moderately sized data sets providing a quick and easy to way to visualize and distributions and identify the patterns or outlines.

They are sometimes used to transportation schedules like the trains and others. See now let's look at the what are the types of visualizations how we can visualize the relationship data using the different types. See, relationship visualizations methods that shows the relationship and connected between the data to show the correlation between two or more variables. This kind of charts are used to illustrate the relationship or correlation between the two or more variables. And these are good for identifying the outliers and trends in the cluster of data.

And these are also highlighting this similarity across the variables and smaller data sets. The number of visualizations can be created under the relationship visualization category such as scatter plots, parallel coordinate plots and weather plots, code diagrams, arc diagrams, correlation matrix, network diagrams and TV diagrams. Scatter plot. The scatter plot is perhaps the most common visualization to illustrate the correlations and these are useful for visually assessing the relationship between the two variables, identifying the patterns and spotting outliers. And scatter plots are commonly used in various fields such as statistics, data analysis, scientific research to explore and understand the relationship between the different set of data.

As an example of scatter plot showing in the image, it shows the association between the net immigration and per capita GDP using either single site transports or different colors of religions of world. So you can observe that it's also giving the positive relationship between immigration rate and per capita GDP. And also it's given that in the other way, the positive relationship between the net immigration rate and per capita GDP. The parallel coordinate plots. The parallel coordinate plots shows the relationship between the two or more

variables across the multiple vertical axis.

 And these are similar to the line charts, but the way the data is translated into the plot is different. And they allow for the exploration of relationship, trends, variations that might be hidden in the raw data. These are more commonly found in the academic and scientific communities than the business and customer data visualizations. Rather charts. These are charts like parallel coordinate charts, but the line warmth around the circle instead of being the arranged parallel to one another.

 It shows the multivariate data of three and more quantitative variables mapping into an axis. And these are charts that are particularly useful when we would like to compare the relative strengths and weaknesses of different entities based on the multiple attributes. And it is also known as a spider charts, the web charts and star charts, polar charts etcetera. Chord diagrams. The Chord diagrams is another way to show the association or relationship between the observations already in a circle.

 It is perhaps the best used to show the how observations are share statistics. It is a graphical method of displaying a relationship between the data in a matrix format. It is an example of Chord diagram shows that migration flow between the major regions of the world in 2017. Each region is placed along with the circumference of the circle and the branch and the branch emanating from each corresponding to numbers of migrants entering and leaving each regions. And arc charts.

 The arc chart is like a Chord diagram as tentated along a size of horizontal axis. This kind of charts is typically used in a graph theory or network analysis. An arc diagram displays the nodes as a points alongside a straight lines with the arc or curve connected pairs of nodes of representative relationship. The example of this arc diagram shows the same migration flow between the regions of the world as the above diagram we have shown in the Chord diagrams. The correlation matrix. A correlation matrix is a table that variables listed along with the horizontal and vertical axis.

 The basic correlation matrix is the table with the numbers that shows the strength of the relationship between the observations. It uses the same layout but instead of numbers it uses the shapes after circles to show the strength of the correlations and sometimes colors and also sets to organize the tables. Network diagrams. See network diagrams is a visual representation of the network architectures. These diagrams used to show the hierarchies and connections within and across the groups and systems.

 And these are commonly used to various fields of academic research including co-authors, networking, collaborations, keyword networking etc. See here is an example. You can see that the network diagram relationship among the evolutions, indices and scholars about the importance of logistics and enterprises success. The another diagram under the relationship or relation is three diagrams.

The T diagram show the levels of hierarchy in a systems or a group. The T diagrams are often used to show the family relations and descent. In the taxonomy the practice and science of classification in evolution of science to show the region of spaces in computers, sciences and mathematics in business and organizations for managerial purposes. And the another type of visualization is part to whole. See part to whole data visualization is a type of the data representations that illustrate the relationship between the whole and it's a part of source that the variables of its total. Particularly this type of chart are useful for showing the contributions of each component to the overall entity.

It helps to understand the percentage or the propositions each part contributes to the total. It is often used to show the something that dividing up to should be some of the 100% that basically this kind of charts uses to accomplish with two things. One maybe to compare the size of different parts and allow the users to see and what kind of bigger is that and second is to show the relative contributions of each part of the total. See there are number of charts can be created under this part to whole visualization. The one is the pie charts and other than the tree map and sunburst diagrams and night angle charts and the diagrams. The pie charts are is a distribution of important goods and is a simple chart that we can regularly use for creating the visualizations.

Again here is an example source that the pie charts are given so the distribution of important goods to goes to the United States in a seven area around the world. The tree map, the tree map are ideal for displaying large amount of erratically structured data and the tree map divides the sections of a square or the rectangles into a group of group to illustrate the hierarchy or the part to relationship. A tree map are useful for visualizing the large amount of hierarchical data in a compact space. They provide the clear overview of the structured and the relative sizes of different categories within the data.

This is an example given for the tree maps. I have shown the breakdown of total import from the specific countries of the United States in 2016. You may find this the easier to read within the pie charts because the rectangles are most easily compared or because it's an unfamiliar graph type you might have find it slower to navigate or more difficult and sunburst diagrams. See this known as a wing charts or the multi-level pie charts or radial charts. This typically used to visualize the hierarchical data structure and it is the visualization techniques that display the hierarchical data in a circular format and these are most useful for illustrating relationship and erratically in the various fields such as a data analysis, information architectures and organization structure and it's consistently the inner circle surrounded by the rings of hierarchical levels.

The another charts is the night angle charts. It is also known as a the polar area charts or it's a the circular graphs the combinational elements of radar charts or and a column graph. The chart consists the series of wedges and each representing a different category of variables and the and the area of each weights corresponding to quality and its represents

and this kind of charts can visualize the all sort of data. And another diagram is the OEI diagrams. It's a mathematical techniques used to divide the space into regions based distance to a specified set of points and these three maps are diagrams. It's a data vision techniques used to represent the hierarchical and categorical informations in a structured and space feeling manners and this charts the particular useful in the various fields including data visualizations as it helps to analyze and representing the particular informations and patterns.

Till now we have explored the most of data visualization types which are going to communicate the quantitative data but there are also gaps that communicate the qualitative data and non-numerical information collected through means of observations and interviews, focus groups, surveys and other methods. Let's explore the what are the qualitative data visualization types are there. When we say that qualitative visualizations are used to communicate the qualitative data or non-numerical informations that does not involve the measurements or the quantities. Understanding the qualitative data is very crucial and it to creating effectiveness visualizations that communicate the insights. Difficulty in converting the qualitative data into numerical values and lack of standardizations in visualizing the qualitative data are the major challenges.

There are number of visualization types for creating the qualitative data visualizations such as the icons one can create the icons and the word clouds and specific words at word trees, specific codes, coloring places and metrics and list. See icons is like icon iconography iconography especially can help visualize and qualitative data. I can be purely decorative that can be repetitive data or they can guide the reader from visual to the nest. It can also be into a part of the emojis and these are themselves as a visual language. They can simply and communicate the ideas and feelings that are otherwise difficult to express and another one is the word clouds.

Most of them are knows what is made by the word clouds and the word clouds are the places that most popular and familiar the way to visualize the qualitative data. When you want to display the words in the visualizing sites with the size of each word indicating the frequency and importance of given the test. The word cloud of the collections or the cluster or the word of depicted in a different sites and the bolder the bigger bolder word appears more often it is means that this has the more importance on particular test. The another method for qualitative data presentation is the word trees.

It is a visualizing test they examine the consistently which different the word appears. The tree structure this source in the tree structure source the combinations in which way the word appears within the test and also word are sized according to their frequency. The font size of words represent the weights and it usually the visual the visually branching structure to show how the pre-selected words is connected to the other words. The word trees are graphically version of traditional the keyword in context method they enable the rapid quiring and explorations of bodies of the test and specific words. The another way to

visualize the qualitative test data is to combine the individual words within a quantitative matrix just like a specific data points can be shown in the scatter plots and specific words can be included in the data visualizations. The another one is a course see sometime if contest is more important the visualization into the words is not the most useful way to show your qualitative data.

  Sometime we need to show the full quotations like the using the specific course faces can be a powerful way to communicate the qualitative data when we want to quote the entire data of the entire quotation by the people or entire data of the record. The another visualization method is matrix and the leads if showing an entire qualitative data set is impossible we can simply and organize the data into the groups or categories especially this approach creates a table of qualitative data so that readers can more easily see the arguments and narratives. Within this method we can see the leads to make it easier for our readers to navigate through what might be the very dense of data sets. The matrix are often used to represent the two dimensional data of both qualitative information and quantitative data and it can be employed to create the heat maps where each cell in the matrix represent the data points and the colors are the cell indicates the values. We have seen that what are the types of data visualizations right from the beginning now we can look at it the tools for data visualizations.

  For creating the various types of data visualizations we should have a knowledge of using the tools of data visualizations and there are number of tools in the both commercial and open access platforms. One can use those tools based on the type of data and also research we can use the right tools for data visualizations. Again data visualization tools are the platform to allow the researcher to perform the various types of data visualizations by using the visual elements like charts, graphs, maps data visualization tools provide the accessible way to see the and understand the trends and outliers and patterns in data. There are number of web-based software tools and program based on the tools are available to perform the data visualizations. We have listed in the screen I have also highlighted like some of the open sources like the MS Excels, the Apache SuperScat and Google charts and also Tableau Public and R Studio and also we have some of the online platforms like Canva and Infograps and Vislo and Wingages.

  There are large number of commercial data visualization platform data visualization tools. The one can look at it if they are they would like to be explored those are available under the commercial platforms. Now we have seen that you know as being a researcher there are certain number of tools which are available to visualize the scientific research. These tools are used to conduct the most often studies like the bibliometrics or cytometrics or we want to analyze the impact of the research. Let's look at what are those tools and these are tools are available freely. These tools are like the one is the OSWU and OSWU is a software tools that's constructing and visualizing the bibliometric networks.

  These networks may for instance including the journals and the researchers and individual

publications and they are often constructed based on the citations, bibliometric coupling, co-citations and co-author relationship and we will have a practical session with OSWU and we want to look at it how we can create the visual networks and various data visualization using this OSWU. The another tool for scientific visualization research is bibliometrics. Bibliometrics as part of the R tool again it's a comprehensive science mapping analysis and bibliometric analysis tools. It is used for evaluating scientific research productivity and impact. So with the bibliometric tools researchers can analyze the bibliometric data, generate the citation network and our visual trends in the scientific literature and it gives the all the instruments to pursue the complete bibliographic analysis and to the science mapping workflow.

Again another tools we have on scientific research analysis like site space. Again this is another freely available java basis applications for visualizing and analyzing the trends and patterns in the scientific literature. A site space supports the structure and the temporal analysis of a variable of networks and delivered from the scientific publications including the collaboration networks, co-authors citation networks and documents, co-citation networks. Another tool is a SCIMAP, again science mapping analysis tools. It's an open source software tools that perform the science mapping analysis and it includes the methodologies algorithm and measures for all steps in the science mapping workflow from the reprocessing to visualizing the results.

Another one is called BIBexcel. It is especially used for analyzing the bibliometric data including the citation patterns and co-citation networks. It can be used to identify the key authors and influence publications like thematic clusters within the given field and another site net explorer is analyzing the citation pattern in the scientific literature platform. It is visualizing software tools for visualizing and analyzing the citation network and scientific publications. The tool allows these citation network can be imported directly from the web of science databases. It is integrated with the web of science databases and the citation network can be explored interactively for instance by drilling down into a network by identifying the clusters of closely related publications.

It powers the interactive visualizations that make it easier for users to navigate and interpret the complex citation networks. Gephi is another powerful open source software for visualizing and analyzing the large network graphs. There is another CR explorer, Cited References Explorer. It uses the data from the web of science as a input and is focused on the analyzing the cited references in particular and the reference publication year and the CR explorer visualize the spectrograms and creates the cited references.

And another tool for called PEZEC. It is another open source software for analysis and visualization of large networks. It uses to analyze and visualize the structure of networks such as social networks and biological networks and citation networks etc. The scholar meter is a citation pattern that extends the Google scholar of facilitation scholar citation analysis and this is an extension available on the even Google scholar also. One can install

the extension of crowdsource the field annotations and measure the universal citation impact. Means if you have the any Google scholar profile it's in a Google scholar extension if you can once install that you will also come to know that what are your profile research areas or what are the major research impact creating from your profile.

 There's a platform called Publish or Parrish and it's a software program that with you and analyze the academic citations. It uses the variety of data sources to obtain the raw citations. Then analyzing the this analyzing the data and present the present a range of citation metrics including number of papers total citations and also the h index. The last one we have citation gecko. It's a browser based tool that helps the researchers to find the relevant papers and present in a graphical and networking manner. Means what you have seen that there are specific data visualization tools which you would like to create the data visualization and a different way you can also use the means the all Google charts or Apache superscript what you have seen in the last slide and also there are specific software tools especially open source software tools.

 If you would like to visualize the scientific data if you like to evaluate the scientific visualization data most probably for undertaking the bibliometric and scientific studies and these are the software tools which are going to be useful for our researcher. Now see we have seen that a number of tools of the both data visualization tools and also the scientific research data visualization tools. Now if you are going to look at the in its means multiple tools the one can want to know the how we can evaluate the how we can identify the best one among the tools what you are finding from the internet because the most of them are openly available you should also know that if you want to select the open source software which one is the best to select or if you go for the commercial that visualization tools and which one is the best for it. For looking for the evolutions this is one of the most complex task to find the best software tool for creating the best data with the lessons and for evaluating this software tools we have given some of the criteria's how the you one can evaluate the software the tools before using them for data visualization. See here some of the key criteria's that can be applied for selecting the best data visualization tools and perform the data visualizations.

 The first one is the license and our development community. See when we say that license if you are using the open source software it's a freely available so you don't need to more look at about the like what does more about the license but when you go for the property one commercial one you have to look at the license and development community maybe the you have to look at the license how you have to look at that the other trial version is there and also if you're going to pay for it you know what are the standard fee and also the configurations and community engagement those things can one can look at it. And second is usability and accessibility. See if your software you want to select you have to look at the how that going to be used and whether that's accessible and how that speed is going to be for creating utilization how that going to interact with you means the speed of those applications and third you can also look at that supported data sets and file type because

there are certain tools you will also find that they are going to support for the all the data sets all the files support but before you go for visualizing the data you have to look at whether those software tools will going to support in a various type of the file formats especially because somehow we may need the file format in a different format not in ordinary pngl jpg we also sometime we need in other format also for evaluating the data visualization tools is graphic support you have to look at how that graphic supports how the all the elements for creating the graphs and also it has the export facilities or sharing facilities or also one can look at it we can work as a collaborative multiple researcher under the single data visualization platforms and server and software dependent whatever the data visualization tools we have seen in earlier slides the most of them are web-based tools the very few you can find it has server-based tools which depends on the server right but when you deciding that for selecting the software tools you have to decide the tools which is more convenient to you you would like to go for the web-based or you would like to select the server-based tools and another evaluation criteria look at the advanced features of software tools means advanced features like you the one can have the all the features but you should have advanced features whether it can also be a API support for creating the graphics and also direct communicated with the other tools if it whether one tools can be communicated to other tools means in a such a way you have created the graphs with using the xyz tools but those graphs you wanted to edit with the another xyz tools whether that is going to be allowed with the tools that you're going to select right those kind of features have to look at it and also the advanced access features whether you know if you are going with the web-based tools you always have an account before creating the visualizations and whether that's a platform can give you the search and filter options if you would like to have those and you'll also find for visualizing the some of the templates right whether those player templates can be searchable those templates can be responsive where you can edit or we can take the some of the elements some of the elements can insert and you you can also look at whether those the the graphics created with using these platforms are the zoom facilities we can you can zoom by narratives uh looking the the minute test you are given and the test and also you have to look at the browser compatibility whether it can be the comb or the mozilla when you're going to purchase for the commercial one also if you go for the open source ones you have to look at the minimum features before you go and creating the visualizations now another part of this session is data visualization workflow now we have seen that see since uh last two sessions we have been seeing that what data visualization is and how it is and how it is has an impact on the research activities and how it is grown uh over the period and what are the types one can create and also the tools that what we have seen recently and how we get to evaluate but when it comes to the research uh there's a workflow for performing the data visualizations see you have a data there at the end you have to visualize in between you have to be prepared for yourself how you can convert you into raw data into the visualization format there are three workflow what we thought for creating the data visualizations for raw data is the one is uh the cleaning up the data uh cleaning up the data see you have a raw data after we collecting the research you might have collected from the survey uh quantitative data or from the interviews or qualitative data has some field research but to ensure that to more data visualization

process data data sets what we whatever data we we have collected it must be cleaned and standardized using the data cleaning platforms see there are certain platforms for using for data cleanings like the open refine and also someone can use the the Microsoft excel for cleaning the data means before performing the data visualizations the data cleaning is the one we have to be standardized the keywords of the data before performing the data visualizations see like maybe the when we say that cleaning it may be the formatting the inconsistencies are the typos and we have to identify the relevant data responses and fix the structural errors and somehow we have to look at the duplicates if it's a duplicate what you can do and filter the unwanted and missing data and validate the data accuracy and the flow that that that needed for your research and maintain the data uniformity and consistency the overall for cleaning data you have to standardize the data in the structure before you performing the data workflow once you're cleaning the data then you have to create the charts like we have shown that the number of applications number of data visualization platform that we have shown so using those platforms you can create the chart but you have to know that before creating the chart at least which techniques and which way you have to be present your data that is the one and also the improve the designs after you creating the charts and then how we are going to be improve that because we have seen in the one slide in the guideline for presenting the better data graphs that includes the elements naming and figure number means you can improve the charts after one creating that to conclude this data visualization session what we had uh in the from last one and two sessions see data visualizations offer the way to transfer the data into visual representations allowing the users to spot the patterns and trend more easily and effective data visualization is making sure that the right chart and right graph for scientific data and proper or the visualization of data can provide the insights directive patterns can help the readers to understanding your research uh better and it can play the significant role in creating the visuals and representing the research results in a better way see visuals are easy to understand compared to the numbers so one can easily understand what we are trying to say when we have this research in the data visualization format see visual can help to understanding the data quickly and clearly the tools what we have shown that data vision tools are very helpful during the process of creating the data visualizations once you create the data visualizations that will helping users to easily they can make the decisions thank you