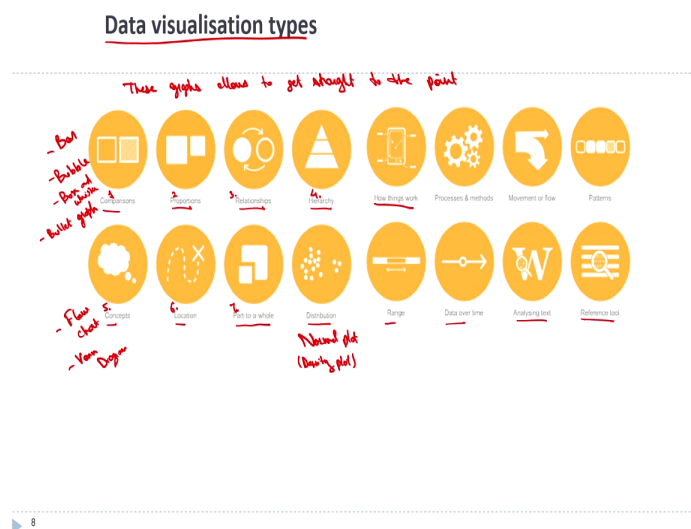


Computer Aided Decision Systems Industrial Practices using Big Analytics
Professor Deepu Philip
Department of Industrial and Management Engineering
Indian Institute of Technology Kanpur
Professor Amandeep Singh
Department of Imagineering Laboratory
Indian Institute of Technology Kanpur
Lecture 52
Data Visualization (Part 2 of 2)

So, let us continue our lecture on Data Visualization. We discussed the Data Visualization Framework, the upsides and downsides of it and what are the various tips when you are trying to present a data when we can present it just for presentation itself, we can present it for the interaction with the audience.

We can present it for Interactive Storytelling, where the audience can ask a question or it incites the audience to have more information about it, while clicking at various points or various figures or various pictures in the website page that you have developed. Now, the different kinds of data charts or graphs or plots types are there, tables are there, how do we present the data for Data Visualization, we will discuss in this lecture.

(Refer Slide Time: 00:59)



When we talk about Data Visualization Types, there could be certain targets, there could be certain basic or primary functions that it has to fulfill. It could be Comparisons, it could be

Proportions, it could be Relationships, or Hierarchies only to be shown concepts, location, Part to a Whole then, Distribution. If I talk about these graphs, these graphs allow you to get straight to the point a simple graph can tell a story of 1000 words as I said before.

So, these graphs allow us to get straight to the point. This also helps us to understand what are the trends, patterns, exceptions in the data. When I say exceptions, I can see the outliers which are there. Then, different kinds of the charts are there that could be used for Comparisons.

- If I talk about the Comparison Chart only, one can visit the website which is given in the reference here, it is data via catalog dot com. Where in Comparisons, we can have a Bar Charts, we can have a Bubble Chart, maybe Box and Whisker Chart. Certain other Charts like Line graphs also are a kind of a Comparison, when we are trying to compare the sales in year one and year two, or we are trying to have a Scatter plot. Even if we were trying to compare one thing with another, we are trying to have a correlation between them, these all could be drawn. Then, an important chart or plot here is a Bullet graph. Because, when I say exceptions, exception means outliers.

The Bullet graph helps us to understand what the exception is. So, these Charts generally have an x and y axis, which I mentioned. There could be certain Charts without an axis like the Donut Chart or maybe dot metrics or Heatmap. Pie Chart itself does not even have an axis. So, when an axis is there, and when the axis is not there, what are the advantages and disadvantages of these, that I will also discuss in the coming slides here in this lecture itself.

- So, these are the certain Charts, then, we have Proportions. When I talk about Proportions, Proportions could be the comparison again, in one way. It is a Visualization method that uses size or area to show differences or similarities between different values or parts to a whole system. The Bubble plot is also one that shows proportion, then, we have a word cloud that shows proportion when we try to take a big word.

For example, if we put a big word Data Analytics, that could be a big color in red. The color of different words related to it would be very close to it. The size of the word would be telling, maybe how often this word is being used in Data Analytics. The word could be data, it could be information, it could be Big Data, it could be interactive tools, it could be computer, certain words in the word cloud could come.

So, these Proportions help us to understand, depending on the size. In the Bubble plot itself, it is a three-dimensional plot that I will also discuss. If two dimensions are there in the Scatter plot, the third dimension could be put in a Bubble plot. With the size of the bubble tells here, the specific proportion of the specific parameter that we are trying to talk about.

For example, the size of the bubble if it is presenting the age of the person, the larger bubble group presenting the person who is more aged than the person with the lower bubble size. Then, we have certain other plots, such as Stacked Bar Charts, Stacked Bar graphs, these are also Comparisons.

- Then, comes the Relationships (we just number them too). Relationships means when we are trying to connect between the different parameters and how the variables are correlated. There could be certain examples, it could be a simple Venn diagram that you see, it could be a Tree diagram, it could be a simple Bubble plot or Scatter plot. Heat Map is also one of the examples, where we are trying to see what are the Relationships between the variables. Variables could be two or more. If two variables are there, the Scatter plot could be used. For example, if three variables are there, the Scatter plot turns into a Bubble plot or we can have certain other plots. Maybe a Heat map could be used. Then, Arc diagrams, Courts diagrams are all used for different relationship presentations.
- Then, comes Hierarchy, a simple Hierarchy when we try to see top management, middle management, lower-level management, it is a simple Tree diagram. There is one director at IIT Kanpur, below him we have a deputy director, we have different deans, different dean's report to the deputy director only then, different staff in that dean office reports to a dean. Then, there is a separate category of two, just below the director, which are the departments in a department. We have different departments. In different departments, we have heads of departments, this is a tree diagram. Other than the Tree diagram, we can even have a Tree map or maybe circle parking. So, these are all Visualization methods, their different objects are ranked, and audit together in an organized system.
- Now, we have Concept notes as well here. Concept notes, when I say Concept graphs, it means we are trying to brainstorm something or we are trying to illustrate something. There could be multiple ways we try to put a concept. It could be Venn diagrams. For example, a simple Flow Chart itself is a concept diagram. Then, we have a Venn diagram.

Maybe for brainstorming itself, if you use different kinds of the word cloud that also could be a concept.

- Other than that, we have Location graphs, when it says Location graphs, it only shows the data over different geographical regions. This could again be the map in the presenting in a different form, it could be a Bubble Map that occurs in a specific bubble. The people who are taking the courses on Data Analytics in India. There are more people in Uttar Pradesh, there are lesser number of people in Gujarat, there are even lesser number of people in Tamil Nadu or so.

This could be a Bubble plot or double Bubble Map, where we are trying to see the location of the person, location of the specific map or it could be a Flow map even like the people from IIT Kanpur have flown to different parts of the world. We can show a Flow map from here to Australia, here to Canada, here two different kinds of companies in the US or so. So, it could also be a Flow map of location.

- Then, to Part to a Whole is also one of the categories of the Data Visualization where one part is connected to the overall graph itself. So, here, there are Donut Charts, the Stacked bars, one part is itself showing something, the Pie Chart itself is a Part of a Whole. One Chart or one part of the Pie, one Pie that you are taking, that is presenting one specific category, one specific parameter here. So, this is one kind of Chart.
- Then, we have Distributions, this is very important. When we talk about Distribution, there are different kinds of Distribution, it could be Density plot, simple Density plot that we call it as when we say a normal plot, it is a normal plot, technically, it is just a Density plot. Then, a Tally Chart is also a Distribution, Timeline is also Distribution, Box and Whisker is also Distribution. So, this different kind of Distribution could be given. In the geographical data itself the Dot map, the Flow map, the Bubble Map are also the Distribution Charts as well.

There are certain other categories where we can see how the things work, this is the demonstration or the label diagram of something how the android form is working.

Then, processes and methods of something, different kinds of the pattern could be used, the range, data over time, analyzing tests reference tools, multiple kinds of the Data Visualization Charts and graphs are available. But, still there are simple charts, or simple graphs, or simple tools, which are more important to understand. I will take only three of them, which are most prominently used.

(Refer Slide Time: 09:42)

Data visualisation examples, Tables

1. Tables:

No. of kids by sex and age

Sex	Age			Total
	12-13	14-15	16-17	
Girls	5	6	2	13
Boys	6	8	3	17
Total	11	14	5	30

Source: Sample survey, Fitness Club Other notes and comments

3 variables

Sex	Average		
	Height (cm)	Weight (kg)	Age (years)
Girls	148.8	48.2	13.8
Boys	163.9	56.4	14.2

Statistics

Sex	Weight		
	Min	Average	Max
Girls	32	48.2	81
Boys	36	56.9	83

Percentages:

No. of kids by sex and age, percent

Sex	Age			Total
	12-13	14-15	16-17	
Girls	16.7%	26.7%	6.7%	49.8%
Boys	20.0%	26.7%	10.0%	56.7%
Total	36.7%	46.7%	16.7%	100.0%

No. of kids by sex and age, row percent

Sex	Age			Total
	12-13	14-15	16-17	
Girls	38.5%	46.2%	15.4%	100%
Boys	35.3%	47.1%	17.6%	100%

No. of kids by sex and age, column percent

Sex	Age			Total
	12-13	14-15	16-17	
Girls	45.5%	42.9%	40.0%	
Boys	54.5%	57.1%	60.0%	
Total	100.0%	100.0%	100.0%	

So, let me start with a very simple data presentation tool which is Tables. When I say Tables, Tables means where we are trying to present the data in a matrix form. It is x, y we have the number of rows, number of columns.

So, here are certain ingredients or characteristics of the table where we need to have a table title. The table presents the number of kids by sex and age. We need to have a column title, that is between 12 to 13, 14 to 15, 16 to 17 or so, then, the total number of students which are there or the kids which are there. These are the column titles. Then, we have the row titles: girls, boys, total.

In certain tables, we also have Footnotes, that tell some specific information which is not put in the cells: the sample survey, fitness club, other notes and comments. This comes as a Footnote here. Cells are most important here, girls aged between 12 to 13, there are 5 girls, girls aged between 14 to 15 there are 6 girls, girls between 16 to 17 years of age or 2 total number of girls

are 13. These are the cell points. These very important tables present the data the grouping with data we can have a comparison between two different kinds of the forms.

For example, if I need to compare two different softwares depending upon number, point number one is speed: high, low, then, we have the availability: available, not available, then, we have the cost: low, high. A table could present this thing as well.

So, tables are very simple or directly or very widely used. Data Visualization tools I am taking as the Data Visualization tools because the data is presented in the form of a table all the times that is there, if the table itself we are trying to generate the graphs. So, tables that are also sometimes presented if you see certain posters as well, certain tables are also presented here.

So, there is the second table that presents the average between three variables. It is height, weight, age. There is a third table where several statistics are given. For example, minimum, average, maximum. So, I would say statistics or the parameters are given here: minimum, average, and maximum. Another kind of the tables that were present here are percentages. This is a starting point of Data Visualization. Simple tables have only numbers here.

Now, here I am trying to present tables with percentages. When I say percentages, the frequencies which were given in the first set are not converted into percentages, a sample percentage are directly comparable to the population percentages. So, this is a basic rule.

So, percentages to be used with caution, because when the sample is small, it does not mean anything. If I say out of three, one is successful, 33 percent is a success rate that does not mean much. If the sample is high out of 300, if 100 have grown, I will okay 33 percent people have grown. So, if the sample size is smaller percentages are not well targeted upon.

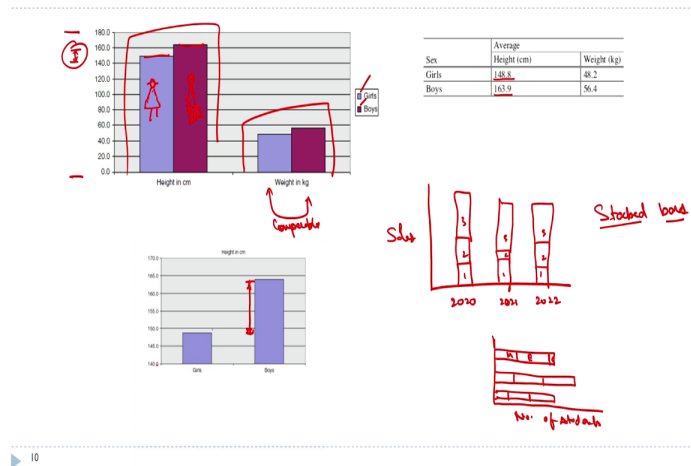
So, percentages, one may also perceive one dimension as a cause, another dimension has an effect in the percentages even. The reason why we use row percent or column percent is often a particular interest in the Distribution of one dimension, while the other dimension is only seen as a grouping only.

For instance, here the girls and boys are taken as grouping. These are the different age groups we are saying 16.7 percent girls are there, or 20 percent of girls that are between 14 and 15 or 6.7

percent girls are there between 16 and 17 age group, to total 43.3 percent are the girls and 56.7 percent are the boys. The percentages are very important because the percentages helped us to also build up different kinds of charts. For instance, the Pie Chart is built up through the percentages only.

(Refer Slide Time: 14:17)

Data visualisation examples, Bar Graph



Let me talk about an important graph that is known as the Bar graph. When I say Bar graph, this is a bar presenting the heights in centimeters. For girls, this is a key, the purple color and the dark pink color is for boys. So, height in centimeters is 148 for girls and boys it is 163, this is height and weight. Two different Bar graphs are put. So, these two are comparable.

The Bar graphs are familiar to most people in general. So, in this case, this kind of the Chart can be constructed to cheat with the scales as well. For instance, if we just need to say the height between the girls and boys, this is the difference. This is showing that the boys are this higher than girls. It is just a presentation. These are our girls, and, these are the boys, directly we can see the difference between the heights here.

Now, we can cheat with the scales here. For instance, if I do not start my scale from the origin, that is why I showed you the Charts, starting with the region, without the region. The start that has a region can be used to cheat with scale. I am saying used to because you only need to take

advantage of it, but this also comes as one of the demerits, when the scale is there for cheating, we should not be cheated upon.

For instance, here you can say, it started from 0 to 180, you can see the rise between 148 to 163, it was only 15 points here, these rise here. Here the scale starting from 140 it looks like the boys are quite taller than girls. So, cheating with scales is there. So, here, Bar Charts and the Line Diagram itself, cheating with scales do happen.

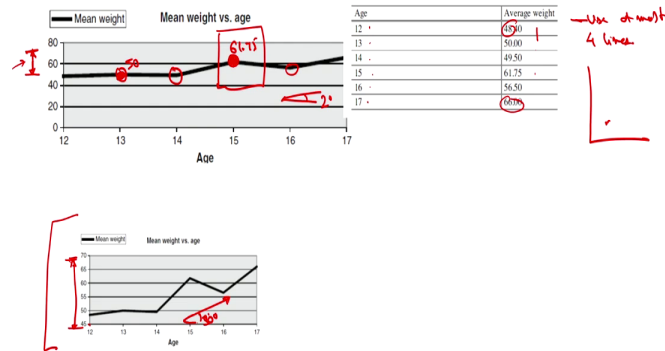
So, it is important to be aware of the axis when studying the Bar Charts, as well as when constructing them it is important to be very careful about it. So, Bar Charts are very commonly used, these are the common bars, there could even be Stacked bars, just hide cannot be stacked. For instance, if I say the sales in one year, it could be like this. Example this is product 1, 2, 3, 1, 2, 3, 1, 2, 3. So, this is the year 2020, 2021, 2022, so, these are sales. These are known as Stacked bars.

It could be vertically stacked, it could also be horizontally Stacked bars, to fit examples that could be put for them. There is a bar which is stacked horizontally as well. Certain examples could be the number of students.

There could be different branches here, maybe mechanical engineering, then, the second one is electrical engineering and so on, it could be computer engineering. See per year, or what are the different ranks of the students. So, this could be horizontally and vertically Stacked bars.

(Refer Slide Time: 18:18)

Data visualisation examples, Line Chart



Next comes the Line Chart similar to Bar Charts. Bar Charts is when we have something in a discrete form, we try to use Bar Charts. Line Charts are there. Line Charts used to illustrate a trend. You can see here is an increasing trend.

Bar Charts is only a comparison between the two or multiple data points. That is okay this is high, this is low, when we need to say that, this high and low also gives us a trend. It is an increasing trend, decreasing trend.

Then, we come up with a Line Chart, this is the mean weight of the students. You can see the age of the students or the kids. An average weight is given when we have plotted that this is the average weight and this is the mean weight which is plotted here. This is our Line Chart.

So, there are certain tips for it when we try to put a Line Chart here, use at most four lines because there are so many complex lines if you try to put, so many lines here, in the Line Chart. It will not give you a great message. Try to use at most fourth lines and try to connect them with a solid line. For example, this is the toilet connection. If I try to say this is a line and try to have a disk dotted point, this is not recommended.

And, Line graphs with a label itself are also important. For example, this is a Line graph where I am presenting something. I can also put a label here. 13 years this value I can say this is 50 or I can say this value here for 15 years of age, this value is 61.75 or so. For example, if my focus of interest is only 15 years of aged girls, I can even highlight this. So, these are certain tips for the Line graphs itself.

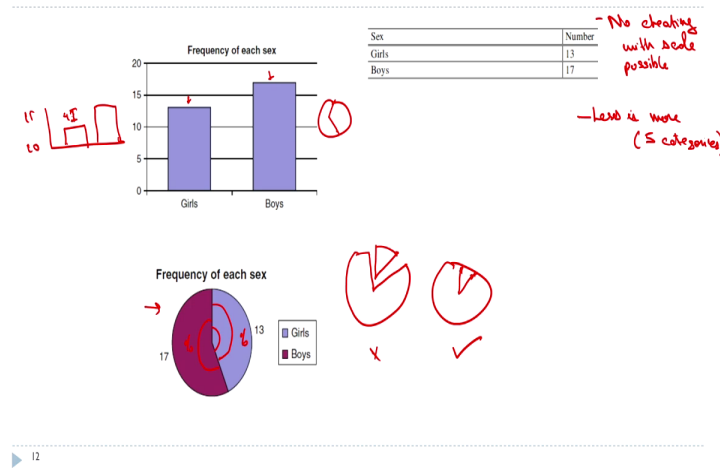
Yes, cheating of scales, when Line graphs are also there, you can see here because the axis starts from 0 to 80, we can only see the trend increasing from around 48 to 66, there is an increasing trend here 48 to 66. So, this trend is if it needs to be shown higher, we can choose a different scale starting from 45 to 70 only the scale is chosen differently here.

So, what was given a specific this height with a scale starting from 0 is now shown in this height scale starting from 45 to 70. Here it looks like that trend angle is higher. You can see if this angle is this angle, let me say if it is 30 degrees, the angle of the above graph is only around maybe 2 degrees only.

But it is given the same data and same information. However, how do you read it again to keep the axis or choose the axis is an important point when we talk about any plot that starts from a region. So, cheating with an axis is to be taken care of.

(Refer Slide Time: 21:32)

Data visualisation examples, Pie Chart



Next comes the Pie Chart. Pie Chart is a very important Chart, there is no axis, no cheating is possible here. I would say no cheating with scales possible. So, Pie Chart here you can only see, the girls and boys, number of girls, number of boys, it could put in a histogram itself or a Bar graph here, number of girls and boys.

But, here the cheating with scales could happen if I start this scale from 10 to 15, I can show here these many girls are there and these many boys are there. It looks like the number of boys is doubled here. But, only a 4 is the rise in the number. So, there could be no cheating of scales and Pie Charts are very widely used.

So, there are certain characteristics of Pie Charts even. It is perfectly used for displaying Proportions and percentages in a part of a whole relationship. So, here, the very important tip is, less is more. For example, we can further divide this Pie into the Sunburn Chart, we can have certain Distributions or we can even have a fan plot of it, but Pie Chart in itself is less. It is trying to present the categories; I would say five categories at most are recommended for here.

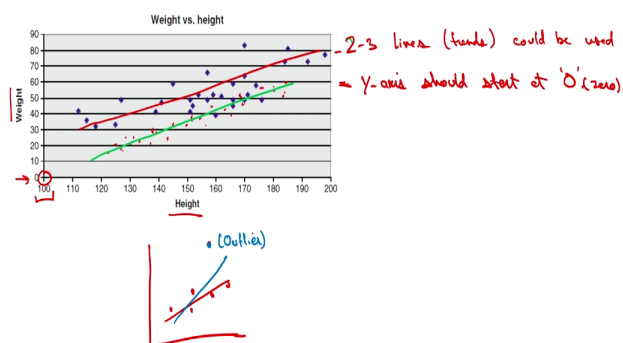
So, clearly labeling of the system also has to be there. For instance, it is only given 13 and 17. It is only a quantitative figure, if I need to convert it into percentages, I can give percentages even here, there is an important characteristic.

So, there are certain points that sometimes we say specific our area of interest in a specific Pie Chart could be given as a separate entity. But I do not recommend this thing. I would like to present it as just a part of itself. So, this separation of the Pie does not mean anything.

Pie Charts could also be given in a three-dimensional form or we can also present a Pie Chart here itself with the bar diagram itself. Because Charts give us the broad information of what our Chart or whatever data is trying to tell and histogram or the Bar Charts could give us the detailed information, what are the different kinds of the plot, different kinds of the heights that we are trying to present. The Pie Chart is an important Chart that does not have an origin.

(Refer Slide Time: 24:22)

Data visualisation examples, Scatter Plot



▶ 13

Let us have a look over a few more Charts such as Scatter plots. Scatter plot is best used when we compare a variable with another. When two variables are both in the interval or ratio scale that means that we have quantitative data. For instance, we have weight and height here, we can compare that weight and height how it is related. And, definitely when we have a Scatter plot, we can always draw a trend.

Scatter plot is majorly one that helps us to understand the relationship between two variables. We can have an understanding of the trend or what is the trend between the two variables. And, also, we can have multiple trends. For instance, if this is the height and the weight of the batch of the students in a class that is now current, we can even have a comparison between the batch of the students in the class 5 years back. For instance, in other data plots could be plotted here maybe

and the second line of trend could be something like this. So, we can say a similar trend is here or maybe one degree slope difference is there.

So, Scatter plot is used for this specific purpose. So, to show the relationship between two variables, it is perfect to use the relationship for large populations. So, we use lines to show trends and relationships. And obviously, as few as possible lines are used, two to three lines could be used. But it is always recommended to use one Scatter plot for one kind of data plot, or one kind of the specific period of the time that we are talking about.

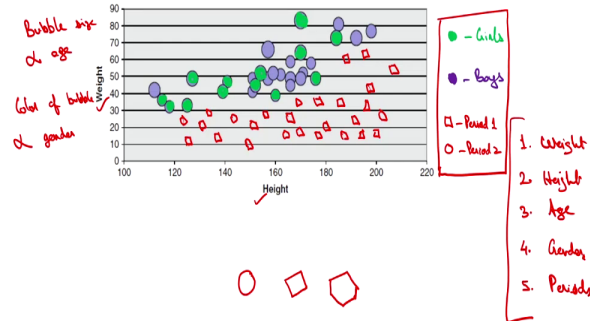
To use as few lines as possible is the first tip, the second tip would be y axis should always be started from 0, you can see the x axis has started from 100 here. So, the y axis always started from 0 because we start from the origin itself without even cheating with scales.

We start with the 0 itself. What is the trend, how the trend varies from the datum itself 0? I will write it here, y axis to start at 0. So, this helps us to understand whether we really need the correlation between the two variables, which is height and weight here or not, or on other hand one cannot by any statistical methods or by studying graphs determine whether there is a cause and effect. So, causation is always missed in the Scatter plot or in the correlation plot even. Another tip here would be if there are outliers, those are very important.

For instance, if this is Scatter plot something like this and the trend is like this, if there is an outlier there, this definitely becomes our course or the area of interest if the outlier would be here, the trend would go something like this. This outlier would be influencing the trend, because the average is going towards this side here. So, this is outlier. So, this becomes our area of interest, there are certain other things in the Scatter plot that we can add.

(Refer Slide Time: 27:59)

Data visualisation examples, Bubble Plot



▶ 14

Data visualisation examples, Tables

1. Tables:

No. of kids by sex and age

Sex	Age			Total
	12-13	14-15	16-17	
Girls	5	6	2	13
Boys	6	8	3	17
Total	11	14	5	30

Source: Sample survey, Fitness Club Other notes and comments

Percentages:

No. of kids by sex and age, percent

Sex	Age			Total
	12-13	14-15	16-17	
Girls	6.7%	20.0%	6.7%	43.3%
Boys	20.0%	26.7%	10.0%	56.7%
Total	36.7%	46.7%	16.7%	100.0%

Summary:

Average

Sex	Average		
	Height (cm)	Weight (kg)	Age (years)
Girls	148.8	48.2	13.8
Boys	163.9	56.4	14.2

Statistics:

Sex	Weight		
	Min	Average	Max
Girls	32	48.2	81
Boys	36	56.9	83

No. of kids by sex and age, row percent

Sex	Age			Total
	12-13	14-15	16-17	
Girls	38.5%	46.2%	15.4%	100%
Boys	35.3%	47.1%	17.6%	100%

No. of kids by sex and age, column percent

Sex	Age		
	12-13	14-15	16-17
Girls	45.5%	42.9%	40.0%
Boys	54.5%	57.1%	60.0%
Total	100.0%	100.0%	100.0%

▶ 9

For example, if the third dimension is here with height and weight itself. If age is also added. Here that plot is known as Bubble plot. The bubble size is proportional to age. Here you can see for the same height and weight where fitness data is there for the tables that we saw in the previous slides here.

The fitness data is there, you can say there were 13 girls and 17 boys, total 30 candidates were there. So, these 30 candidates are plotted here in the Bubble plot with the ages as well. The age is also given in this data, the ages range from 12 to 13 years for few then, 14 to 15 years and then, for 16 to 17 years, they are the categories or classes of the age here.

In this Bubble plot, each data point has its age and within that age group that lies. The use of the Bubble plot is that while using different size or color of the bubbles, the bubbles can show relationship between data in a very clear manner and that means the third or fourth variable could also be introduced here already three variables are there, one is weight, second is height, third is age that is the size of the bubble also I can put the color of the bubble. Color of the bubble employs the category or I will say gender.

For instance, I can even color the bubbles here let me say a color. There are 13 girls. There are 13 bubbles which are of green color. The size of the bubble is anywhere representing the age, putting more girls on the downside because it does not have this specific data this is just for an example. Because the girls are considered to be of a lower weight and the height might also be lower. So, there are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13. So, here green color I can put a key green color are girls and the purple color are boys.

So, we ended up with four variables:

- 1) Weight
- 2) Height
- 3) Age
- 4) Gender

In the Bubble plots the shape for the bubble that is given in a circle only. We can also pick different shapes. A circle is a very simple shape, there could be a shape of the hexagon even, it could be the shape of a diamond, it could be the shape of a square. It depends upon what kind of the shapes or what kinds of graphs you are showing. If you are trying to show different products you can pick different kinds of the bubbles.

Here, the fifth variable could be the age groups now or the class students now or the previous, we can say different periods. Example ages now, I can even plot another I will call it bubble even but it is of a different shape. This might be something of the previous year, 30 points you can take.

So, we have five different variables in a single plot where a simple Scatter plot was here, where only two variables were their height and weight, we had it age as a bubble size, we added color of the bubble as the gender, we added the shape of the bubble as another variable, for different

period. So, is period one and is period two, so on. These are 30 points. For the same people the height that was taken in the previous period or the weight was taken by the previous period, that is how it is different. That is a weight if it was lower in the previous period, we can have interpretations like those. So, this is a Bubble plot.

Now, an important point in the Bubble plot is that the key should be given in a very clear manner. The color, the size and what are the different variables it is presenting.

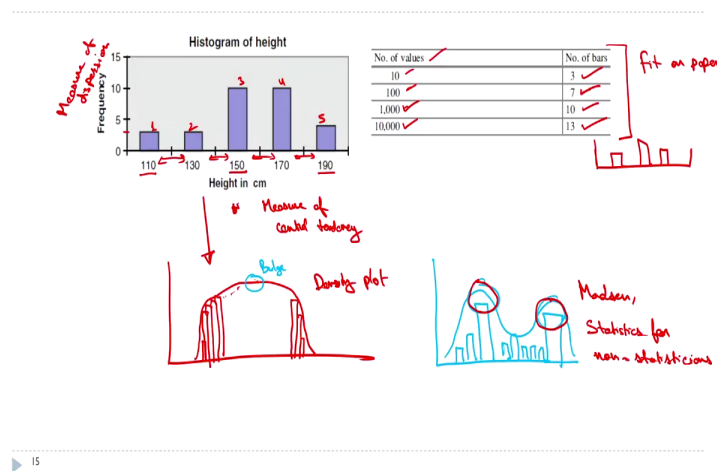
Number two, the fairest way is the size of the bubble should be exactly proportional to the data size or data point size that we are trying to present. For instance, if the age is 10 years, the size of the bubble could be maybe 10 mm, if the age is 8 years it could be 8mm I am talking about diameter.

Yes, an important point is the bubble size could be very as per diameter or as per area. Area of bubble or bad bubble, both could present. It depends upon how variable the data is. If the data variation is low, we can use the diameter of the bubble only as the variation of the data.

If the data variation or variability is too heavy, one bubble could come very small, another bubble come very big, we can take the area. In that case that variation would be looking lower. So, it depends upon what we are trying to present in the Bubble plot.

(Refer Slide Time: 33:59)

Data visualisation examples, Histogram



Next comes another kind of the product known as Histogram. Histogram is a special or specific type of Bar Charts only where you can see the bars are there, bar 1 2 3 4 and 5, which are placed at a specific distance. So here, the frequency of the data values is given and how do we visualize the Distribution of data value that is given. So, it is the central location of the data and how large spread of the data, both of things are given.

Here the central height is 150 that is given here, it varies from $150 - 40 + 40$ that is between 110 to 190. So, measure of central tendency, a measure of dispersion both are given in this Histogram. This x axis gives me the measure of central tendency and y axis gives me what is the number of the students in the class who are having height 110. Number of students in the class having height 130, when I say 130 it is a range between maybe 120 to 140, which is given the value 130 here. So, you can see these values 10, 100, 1000, 10,000 how the number of bars should be stacked. This is the value.

So, we cannot have 200 bars in one graph, because we need the data to be meaningful. We cannot have very complex diagrams that no one understands. It is given yes, if the number of values you have are 10, 3 graphs could be used. 10 values are there using 3 graphs. Some number of values or 100 use 7 graphs. If the number of values or 1000 use 10 graphs, if the number of values is 10,000 use 13 graphs. Because this is being presented as discrete data as of now.

At most 13 bars are only suggested as given in the book by Madsen. There is a book where all the graphs are taken from that book, only the title of the book is "Statistics for non-statisticians". The reference of the book is given in the reference slide of this PPT.

On the left-hand side in the y axis, we are having the frequency that gives me the measure of dispersion. In any way, even the normal graph is also built from the Histogram itself when the number of Histograms very large, we are trying to have it as a continuous data in this kind of the discrete data only maximum 13 bars are suggested, when we try to have a continuous data, we can club the large number of bars into a continuous line and if Histograms are pleased in such a way, so on. These can be plotted as a what is called a Density plot, that is taken from the Histogram itself.

So, the general considerations that you have in mind for having these many numbers of bars is only that the bar should fit on the paper. If you try to print it, then, you should be able to accommodate all the observations. To accommodate all the observations only, if 10,000 observations are there, 13 bars could be put there. We divided into the equal number of classes accordingly.

It is important that the data basic meaning is always covered. For example, there are multiple bulges here, there is only one bulge here in the data, this is a bulge. There could be some data points in Histogram itself where it could be more than one bulge. Here, it goes something like this.

In this case, we cannot limit it to the 13 only because the actual information of the data, actual information or this specific knowledge that it is to be taken out from these two bulges has to come out. So, intervals must be defined clearly so that we have a good number of the bars. Good means, the right number of the bars that should fit to your paper that should not make the data unclear. This is Histogram, Histogram is very widely used. We will select the interval length we round off the result to the approximate number and we try to plot it to have the normal graph even.

So, these were all the major graphs which I discussed, which are very widely used. The kinds of the graphs which are other than these are very widely also used in the website designs such as Box and Whisker plot, there could be plots like different kinds of trees are there for a decision tree itself the Tree plot is there.

There could be bubbles which are intersecting which are also known as the Venn diagram itself. So, multiple plots or other multiple charts and graphs are there which are used in Data Visualization. I have only discussed very widely used Data Visualization plots here. In the previous lecture, Professor Deepu Philip only discussed, how do you translate from HTML to PHP then, to try to display the data.

With this the course content is over. We have covered majorly the Data Visualization techniques in this week. And in the previous weeks what is discussed that I will discuss in the last lecture where I will try to summarize the course. Thank you.