Learning Analytics Tool

Professor. Ramkumar Rajendran

Department of Educational Technology

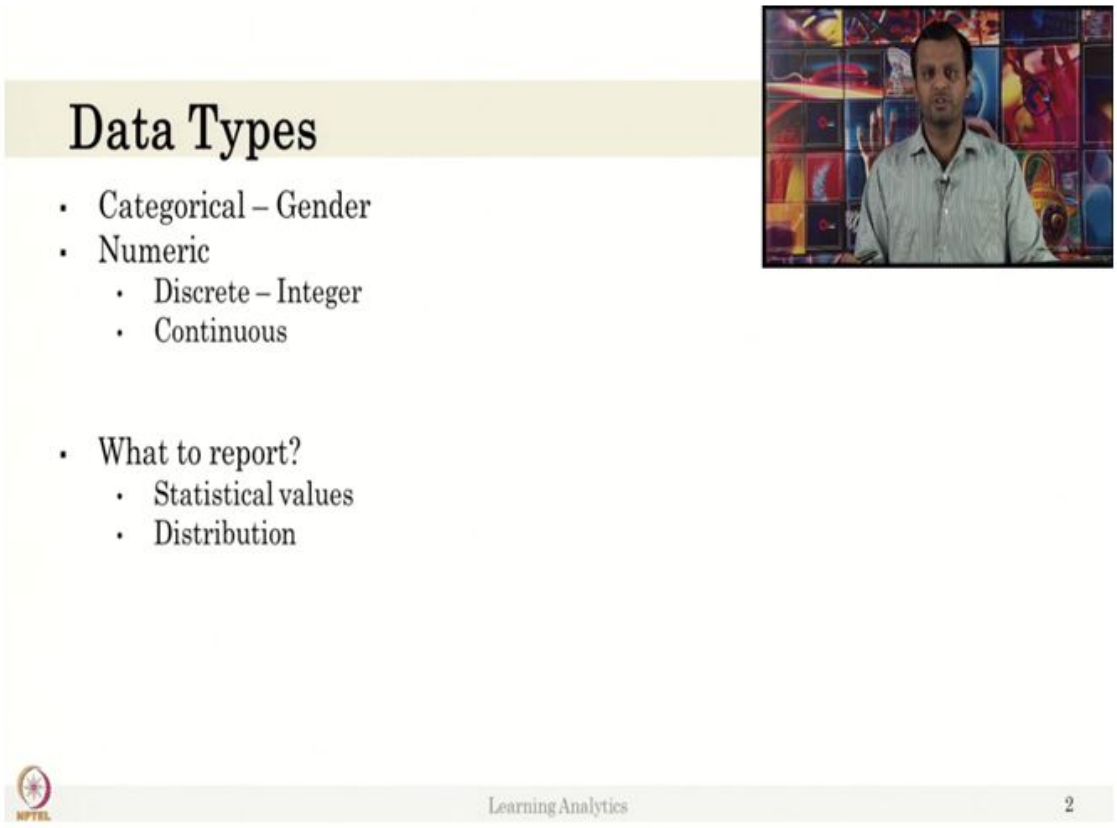
Indian Institute of Technology, Bombay

Lecture No. 4.1

Descriptive Analytics

Welcome back to Learning Analytics Tool Course. This week we will talk about Descriptive Analytics. Before we jump into what is Descriptive Analytics, let us look at the data types. There are different data types like Categorical, Ordinal, Ratio.

(Refer Slide Time: 00:33)

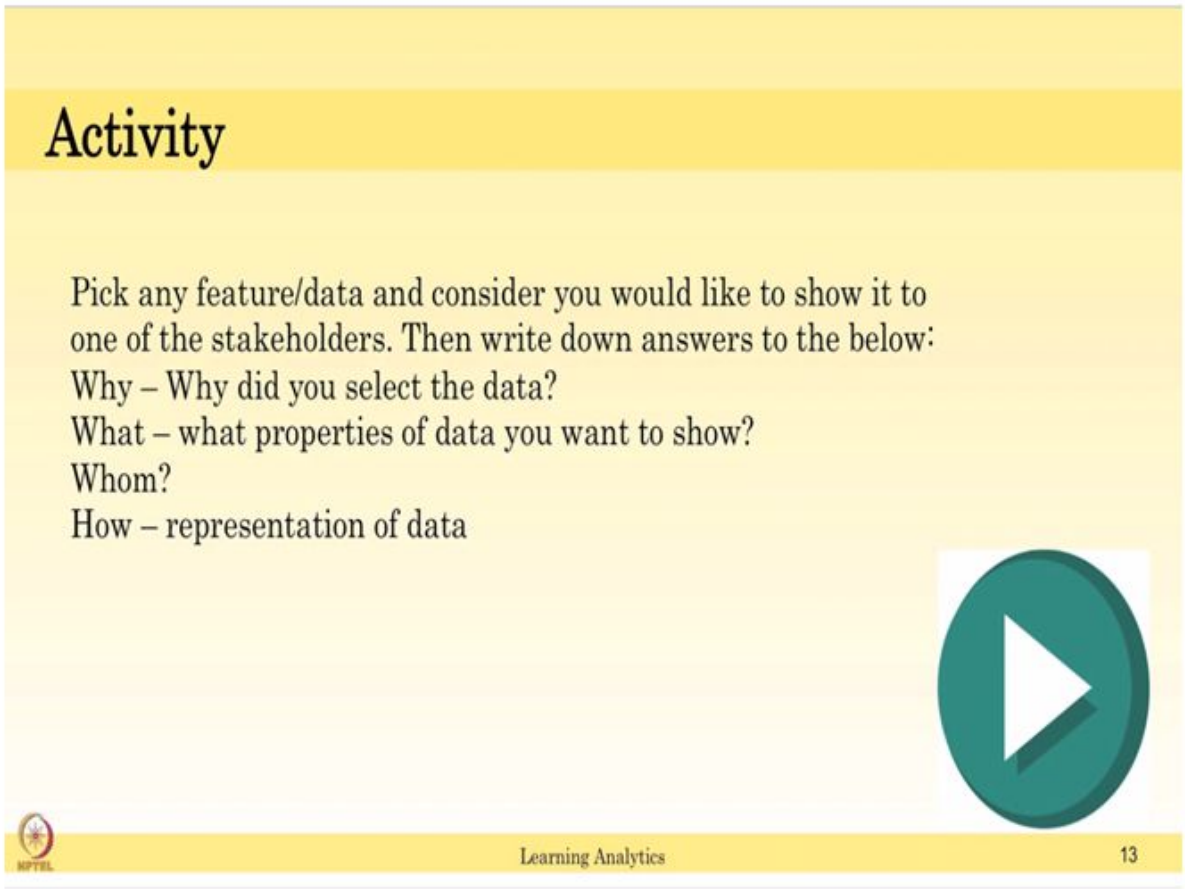


But let us use only the two types here say, Categorical and Numeric. What is Categorical data? Categorical data is the data which can be divided into groups i.e. either this or that group/category. For example, if you collect gender information, It is female or male. There is no order or something here, or if you collect a response to the survey questionnaire, the response is Yes or No, then you can have this Categorical Data.

So, most of the data we will collect in qualitative analysis is categorical. But let us look at the numerical data, which we collect from the student's interaction with the learning environment. This can be again classified into two: Discrete or Continuous. Discrete is an integer value, say students performance in the Math's exam 75, 76 that value.

So, the important thing is what to report. These are data types which we can collect from the students learning environments. So, what do we put? We have to report the statistical values just to make sense of data, what is the minimum value what is the max, what is the average, median or mode or you can say skewness something like that, you can report this statistical value of data in particular form of representation. Also, we can report the data distribution, distribution in the sense of how the particular type of variable changes over a period of time or within the student, how particular user behaviour is changing, how the student's affective state is distributed across one session, how the students, all the students in the class emotional states has been distributed for a particular art, something like that.

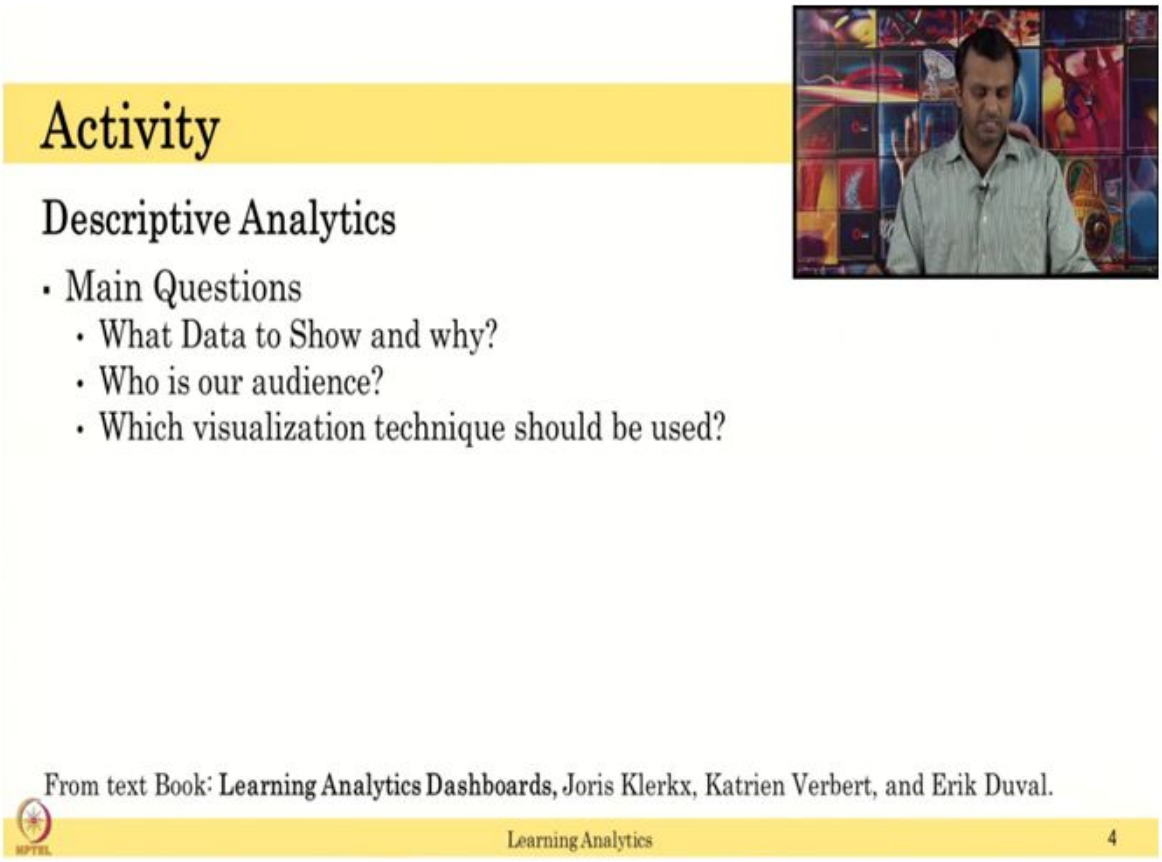
(Refer Slide Time: 02:26)



Pick any feature data and consider you would like to show it to one of the stakeholders. So once you are the researcher here, consider you want to report it to other stakeholders like the one who funding or the students or the teacher or you want to show it to the someone who knows developing educational learning environments.

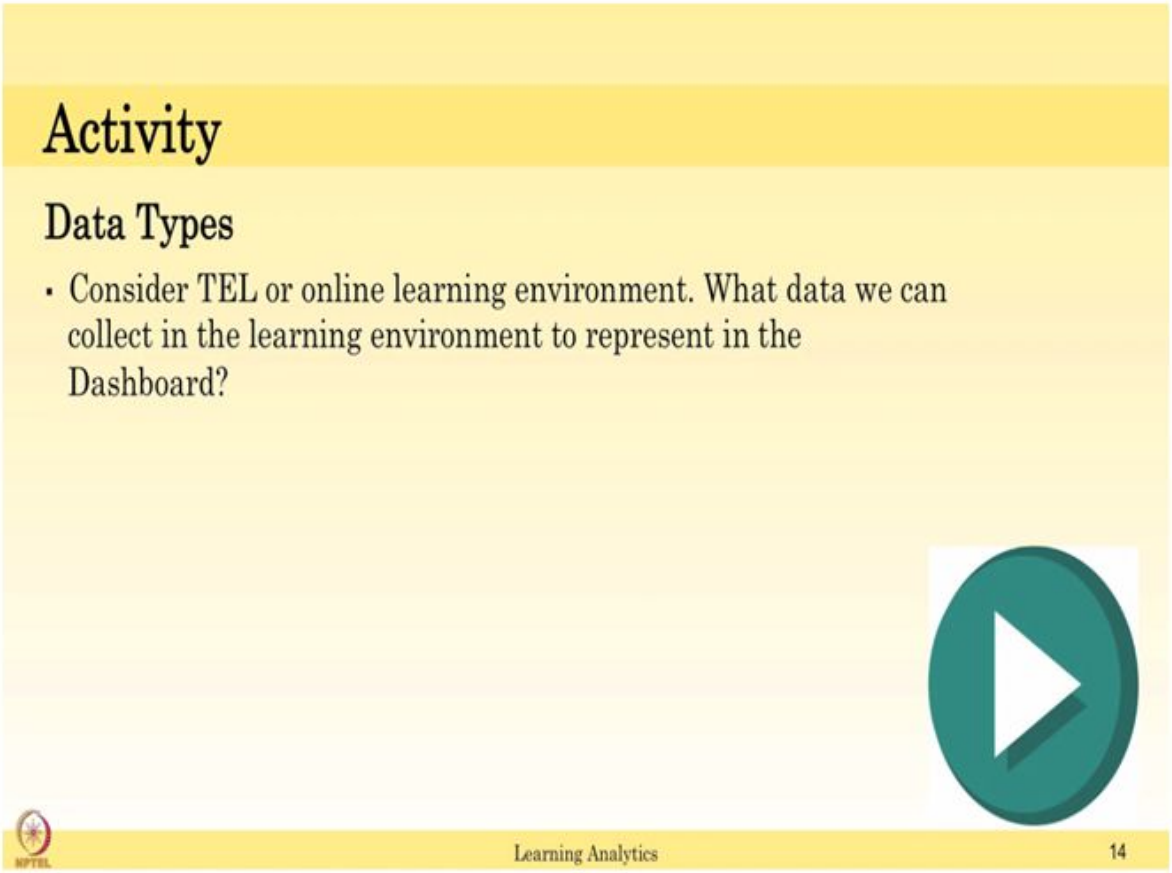
Consider any stakeholder and write down the answers to the below question. Why did you select this particular data? Why do you want to show this data to that particular stakeholder and whom you are considering? Which stakeholder you are picking, because based on the stakeholder the data you want to show will change. So, what properties of data you want to show and how do you want to represent that like a table, text format, or you wanted to present the data in some plots or graphs. So list down answers for these questions. After listing it down resume the video to continue.

(Refer Slide Time: 03:28)



Descriptive analytics is the first type of analytics in learning analytics we discussed. the Descriptive analytics deals with answering these main questions. So, the reference book for this particular part of this course is the Learning Analytics Dashboard by Joris Klerkx, and Katrein from.

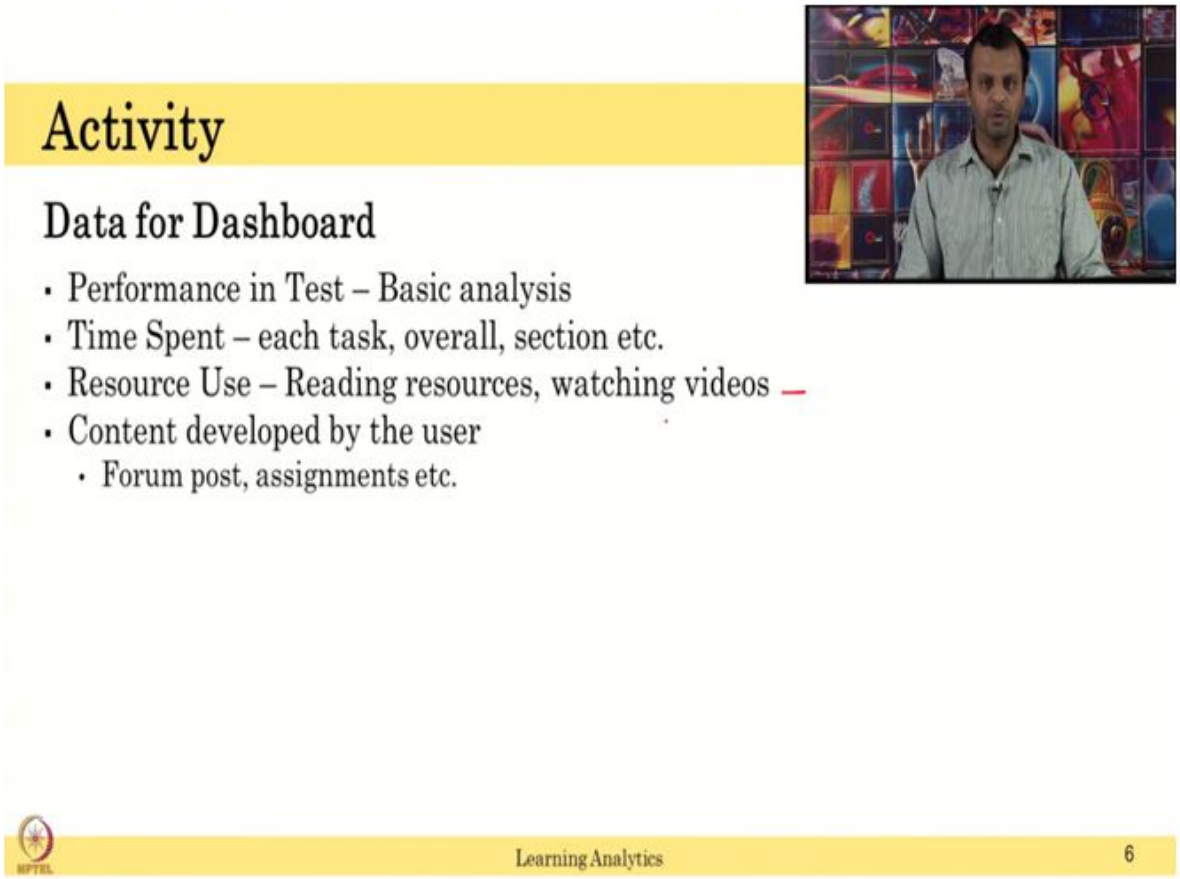
(Refer Slide Time: 03:49)



Let us start with another activity. Consider TELE or any other online learning environment, what data we can collect in the environment to represent in the Dashboard. So, imagine there is a dashboard you want to create for an environment or TEL environment you are creating. And you want to represent students interaction with TELE, what you would like to collect and show it to students or teachers?

So can you think for a moment, and list down what are the data you would like to collect from student's interaction which should be shown in the dashboard?

(Refer Slide Time: 04:29)



So in a Dashboard, to the students, we can show these kinds of data. For example, you can take the students performance in the test, you can show performance in test or performance in their final semester marks or you can show performance in subtopics. All this information can be shown. Or you can also show the time spent on each task, and overall how much time they spend on each session, and how they time spent on particular tasks compared to the other students.

That you can compare and show it. Also, you can list the data that is reading resources, are they spending time on reading resources? Are they spending more time watching videos? This data can be shown to students. Also to teachers that either your students are mostly spending time on watching videos and not reading. So, the teacher can go on motivating the class saying that you should spend more time on reading.

Similarly, for Content Developer, the user, that is how many forums post they are doing, how many assignments are submitting, this can be shown to the teachers. For times spent, this will help both students also teachers, for students, they will understand that I am putting more effort

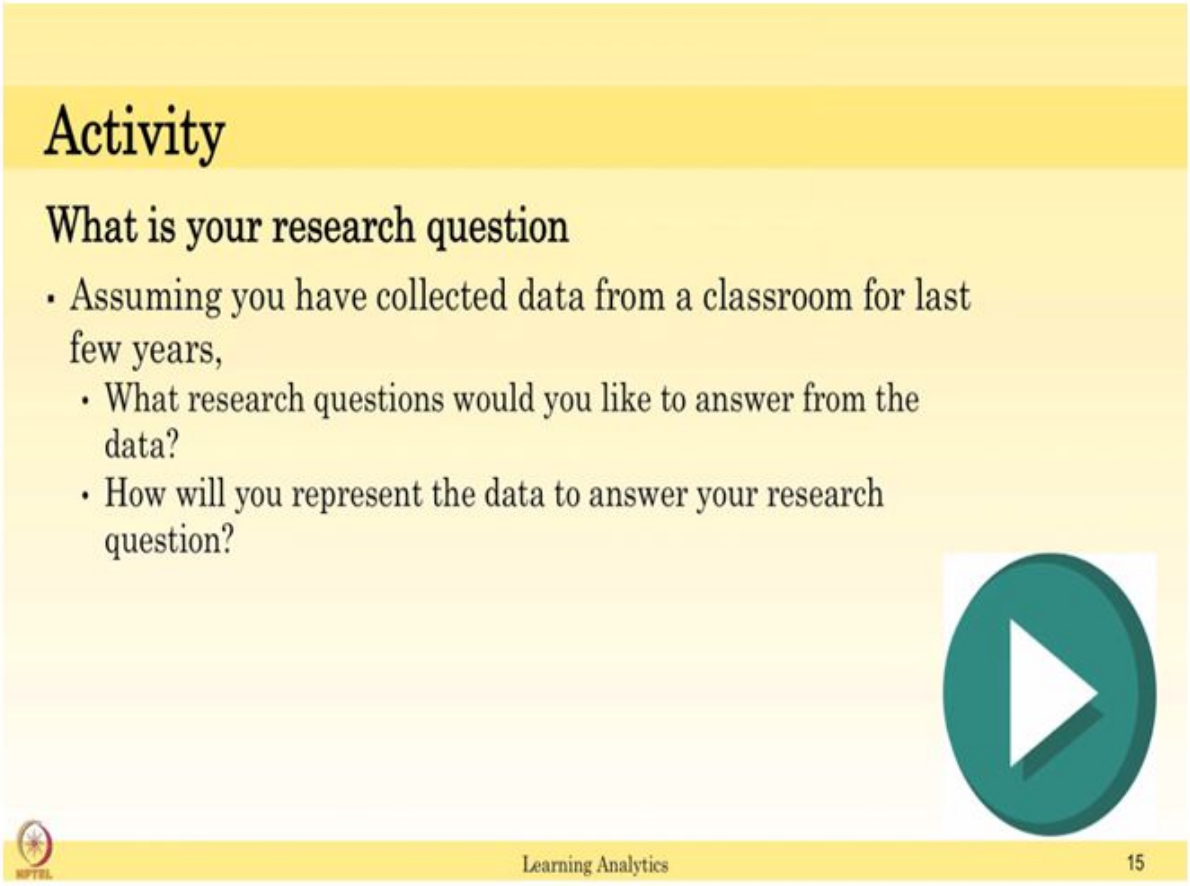
on particular task instead of other tasks, for teachers, it might give indications that in a particular task in the course students are having a problem that is why they spend more time on that particular task.

So, by creating descriptive analytics we can infer a lot of information like which particular task students are having trouble. Our students can self evaluate or compare themselves with others in the class or they can do self-reporting like they are reading too much of resources, not spending any time on watching videos or not spending time on creating a complex pot of solution or assessing the solutions.

So, this is self-evaluation happens, the teacher can infer which task or which particular concept is students are struggling with. All this information can be obtained through Descriptive analytics. So, this Dashboard can be for each individual student. Also for teachers or teachers can look at all the student's data. If you consider this dashboard to the stakeholders at a higher level. For example, this school principal or the district head, you may not need to show all this information, you might abstract this information and offer a different set of information.

Please consider that when you create a Dashboard or collect data for Dashboards. you have to think for whom you are making this dashboard.

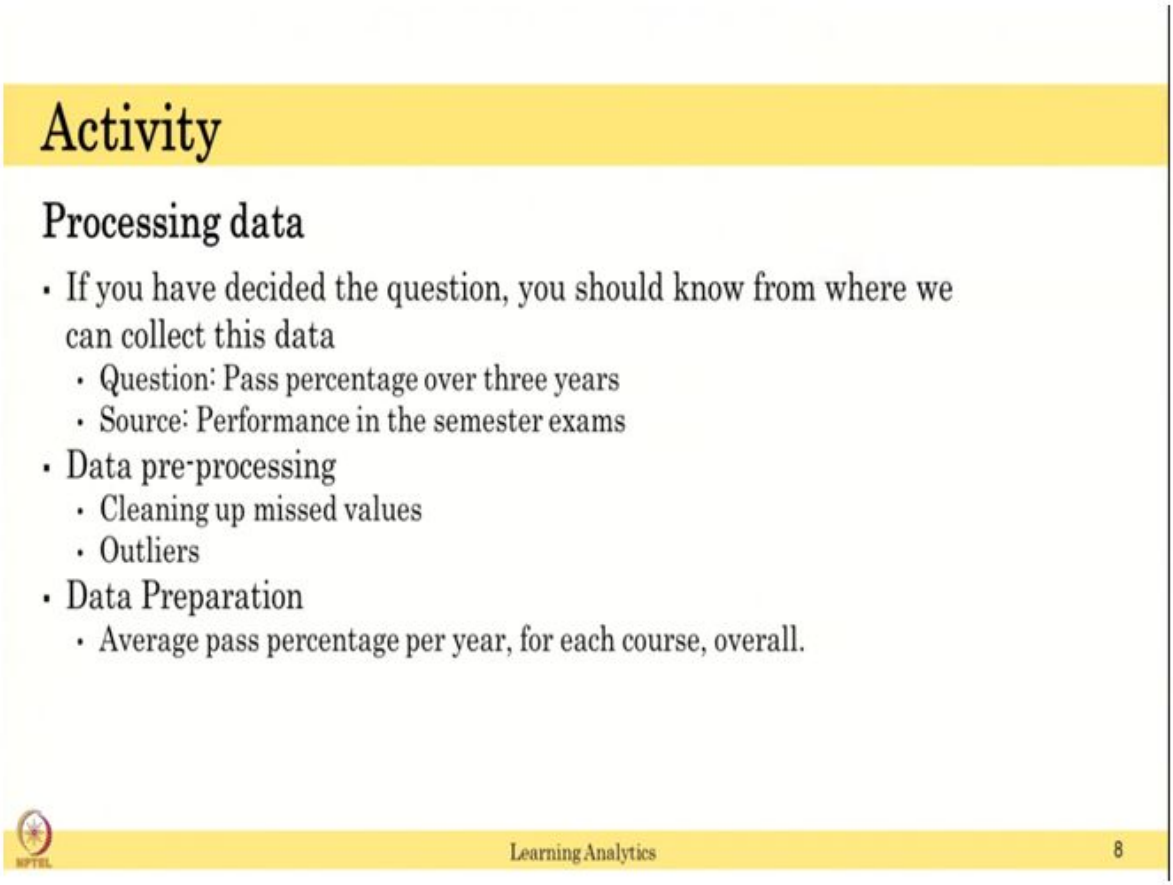
(Refer Slide Time: 07:17)



So, let us go with the next activity. Assuming you have collected data from a classroom for the last few years. Say you are teaching the same course for the last 5 years, we started the same problem in the first week of this course. Consider you have collected the data for the last few years. And now you have the freedom to select any research question you would like to answer from this data. For example, how students performance and attendance are correlating over the years something like that.

So you can select any such question from the data. And if that is a question, how would you like to represent this data to answer your research question? So take a moment, think about it. The task is to come up with a research question. And list down what data is needed to answer the research question and how do you represent this data?

(Refer Slide Time: 08:21)



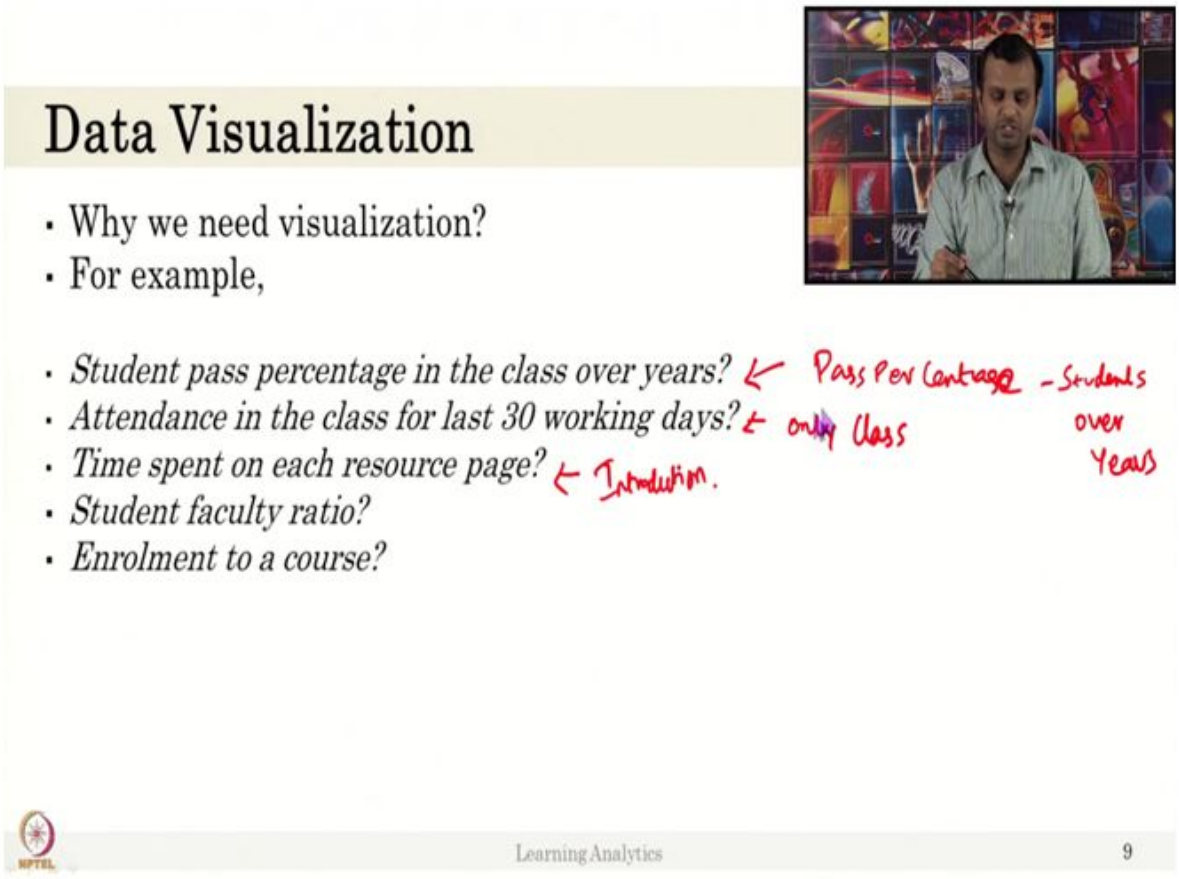
So processing data, like what data to collect? How to represent is very, very important. If you have decided the research question, you should know from where you can collect this data. For example, I said that if I want to know how students attendance and the performance are correlated, or like a hypothesis that higher attendance leads to higher marks in the exam, in that sense, I need to collect data such as their attendance and performance over the years, that data I have to collect.

So based on your research question, you will know which data to collect and how to collect this data. If it is pass-percentage and the performance for three years I can collect data from the mark sheets or from the performance in the semester or mid semesters exams. After collecting data, as, as usual, step two will be processing the data that cleaning up the missed values, remove the outliers, make sure there are no errors in the data.

Then you have to prepare the data to represent it i.e. you have the data now you have to compute average pass percentage per year for each course so that you can show this data overall. For example, you collected all the data in Excel sheets now you have to go and compute average passed percentage for the year 2015, 2016, 2017, so that you can show that in the graph. So, that is important.

So not just data collection, also this data collection is connected to the research question you are asking. Similarly, hope you have answers for such question and ideas about how do you represent it.

(Refer Slide Time: 10:10)



So the most important part is why we need Data Visualization. Why? For example, if I want to show the students passed percentage in last 4 years, we can show that using a trend graph like a simple bar graph or line chart to show that a student's passed percentage in a class from 2015 to

2020 or attendance in the class for last 30 working days. You can plot all the working days in x-axis and attendance.

You can see the attendance varies or attendance reducing or attendance increasing. This kind of trend shown here. More detailed or fine-grained is that time spent on each resource page. For example, the first one, here we talked about more abstract level, for example, passed percentage. Here in this passed percentage, we are considering the group of students not a single student.

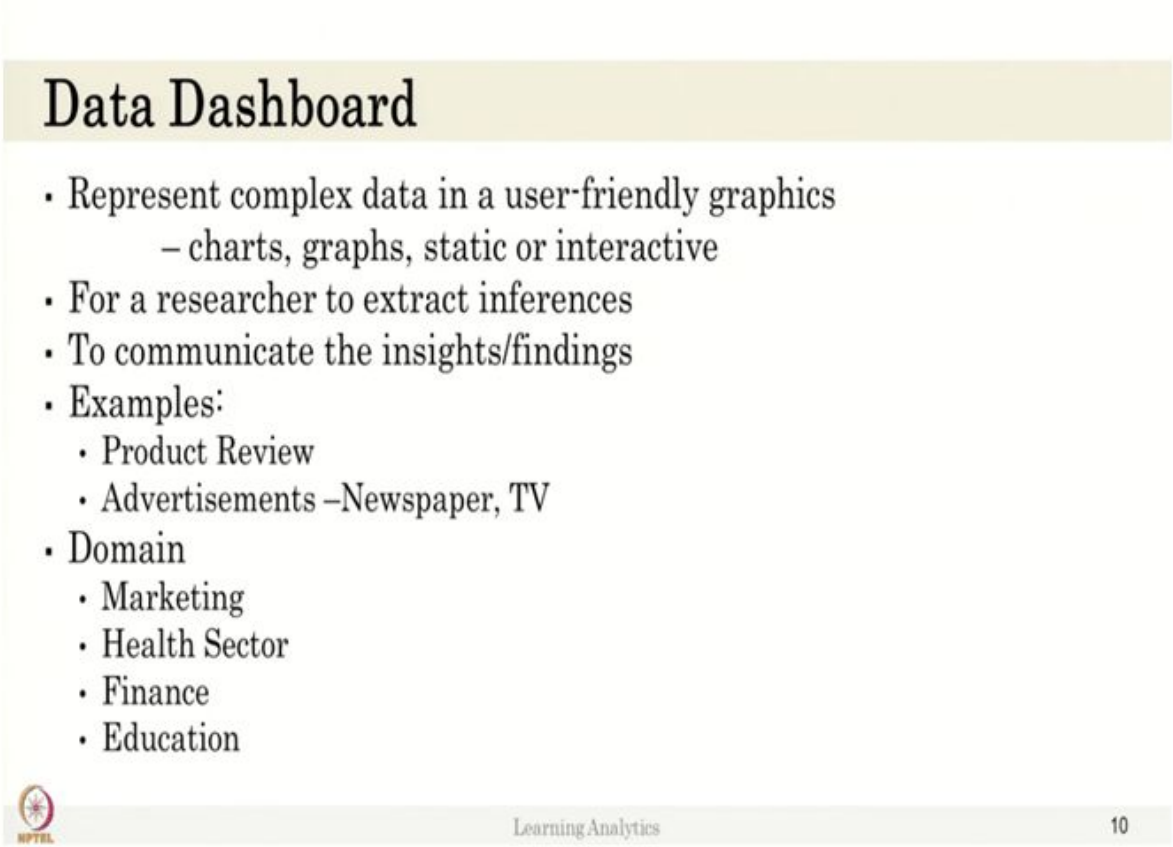
So, we are considering students over years. N will be more. In this, it is showing a particular class say your class and you have 50 students data over 30 days. So one class and N can be, say less compared to this. Now you are coming back to one particular resource page. For example here for resource page, say the PDF on introduction, what is average time all the students, students spend on this introduction PDF?

This data can be collected from the TEL environment or a MOOC or Moodle. So, this is introduction one particular page time, this also can be collected for each student. So, the data can be collected at different levels. And, or if you want to go for academic analytics, like student-faculty ratio, it might help the institutes so data is to make a decision.

Also, the enrollment will help to decide whether which course to offer next year which course not to offer next year. So, the data can be at different levels based on what is the research question and whom you are going to share the data with. Also, it all depends on whether you want to collect individual students data and you want to provide recommendations to individual students.

All these data to identify why the students are not able to attend the classes or identify the reason for the increase in pass percentages in one particular year. What was the teaching strategy used in that particular year? So, these kinds of questions can be answered by collecting this kind of data.

(Refer Slide Time: 13:23)



So, what is a Data Dashboard? I use this word Dashboard a couple of slides ago. Data Dashboard is not new to us. We are interacting with the dashboard or seeing this everywhere in the newspapers, or in a TV advertisement. It actually is a form of representing complex data in user-friendly graphics so that you can understand it easily.

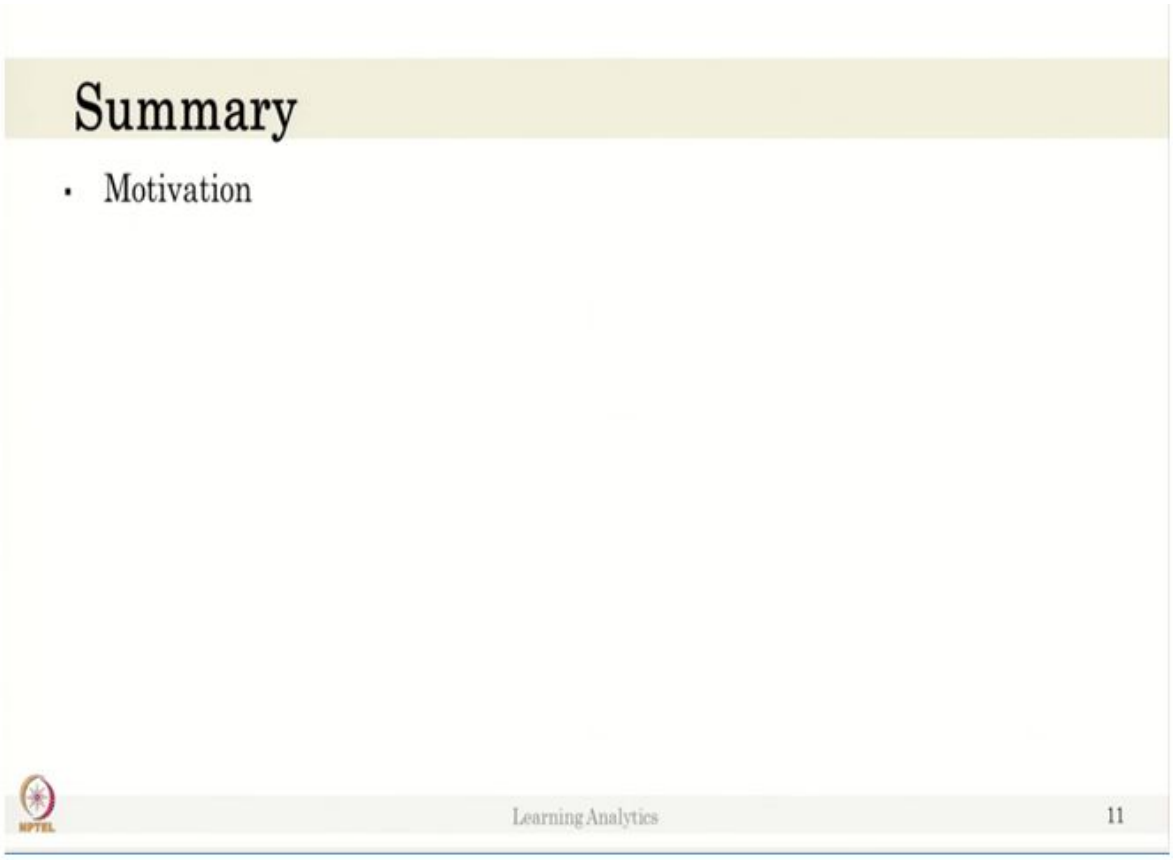
You typically use charts, graphs. These can be static or it can be interactive graphs if you use a system online based or computer-based data graphics. More recently, Data Visualization is differentiated by from infographics. We are not worried about infographics or what type of infographics has to be used in this course.

In this course, our aim is to understand what is the data or which plot to use so that we can infer from that data. For the researcher, the purpose of the Dashboard is to extract inferences from the data. When you have the plot, the data, and charts. The researchers aim is to extract the information.

Also, it is to communicate your insights and findings to other stakeholders or to publish to the general public. Examples of Dashboard is product reviews, which we see in newspapers and TV magazines, or websites. A lot of advertisements talk about these dashboards comparing the two different data to promote their product. It is mainly in the domain of marketing infographics and also in the Media, Health Care, Finance also in Education.

So, Data Dashboard is not new to us. What we are trying to see in this course is- Can we create a simple Dashboard based on the data we collect in the learning environment, such as TEL or MOOC or classroom environment?

(Refer Slide Time: 15:21)



So, this particular video is just to motivate. Like, we have to look at the data in a graphical way. And you have to start considering also that It is not just collecting data and finding the research question, also it is to represent the data to other stakeholders, or creating a Dashboard such that it will be useful for researchers and other stakeholders to make inferences. Thank you.