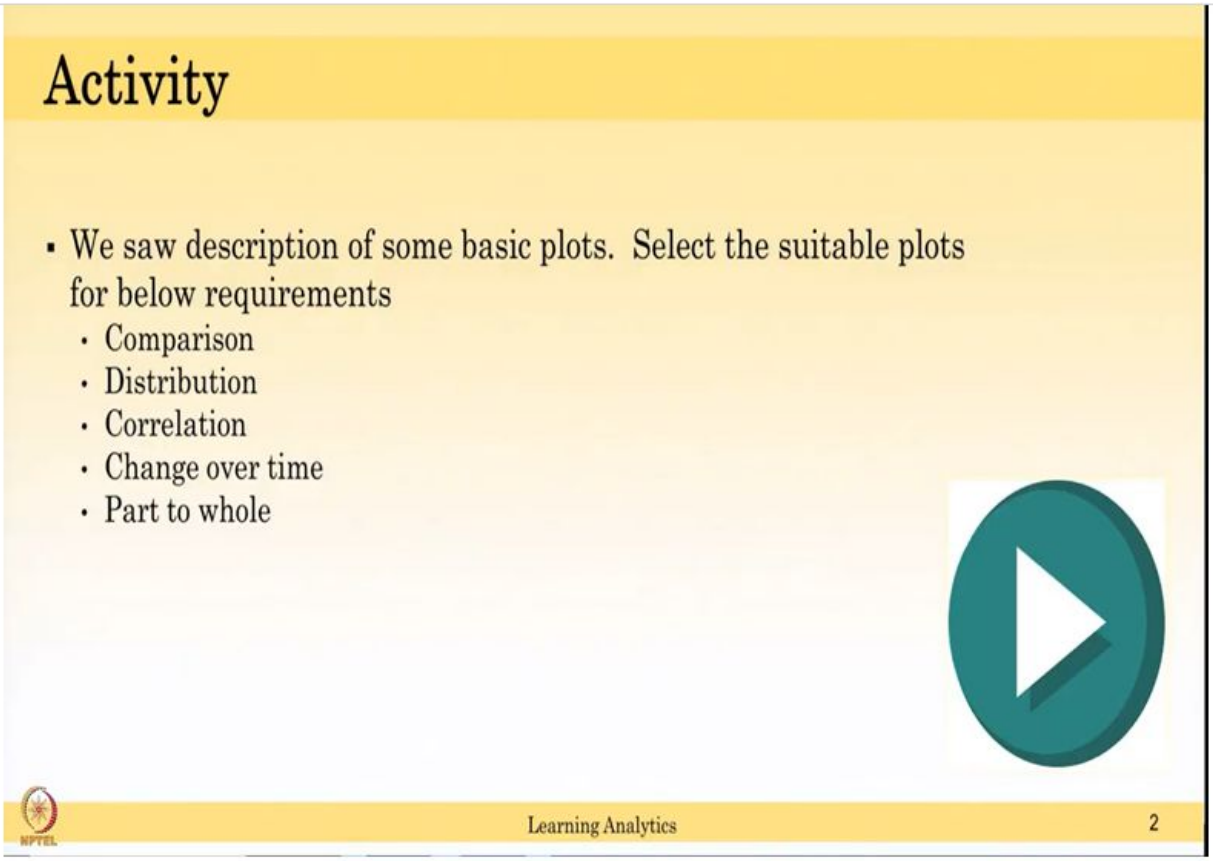
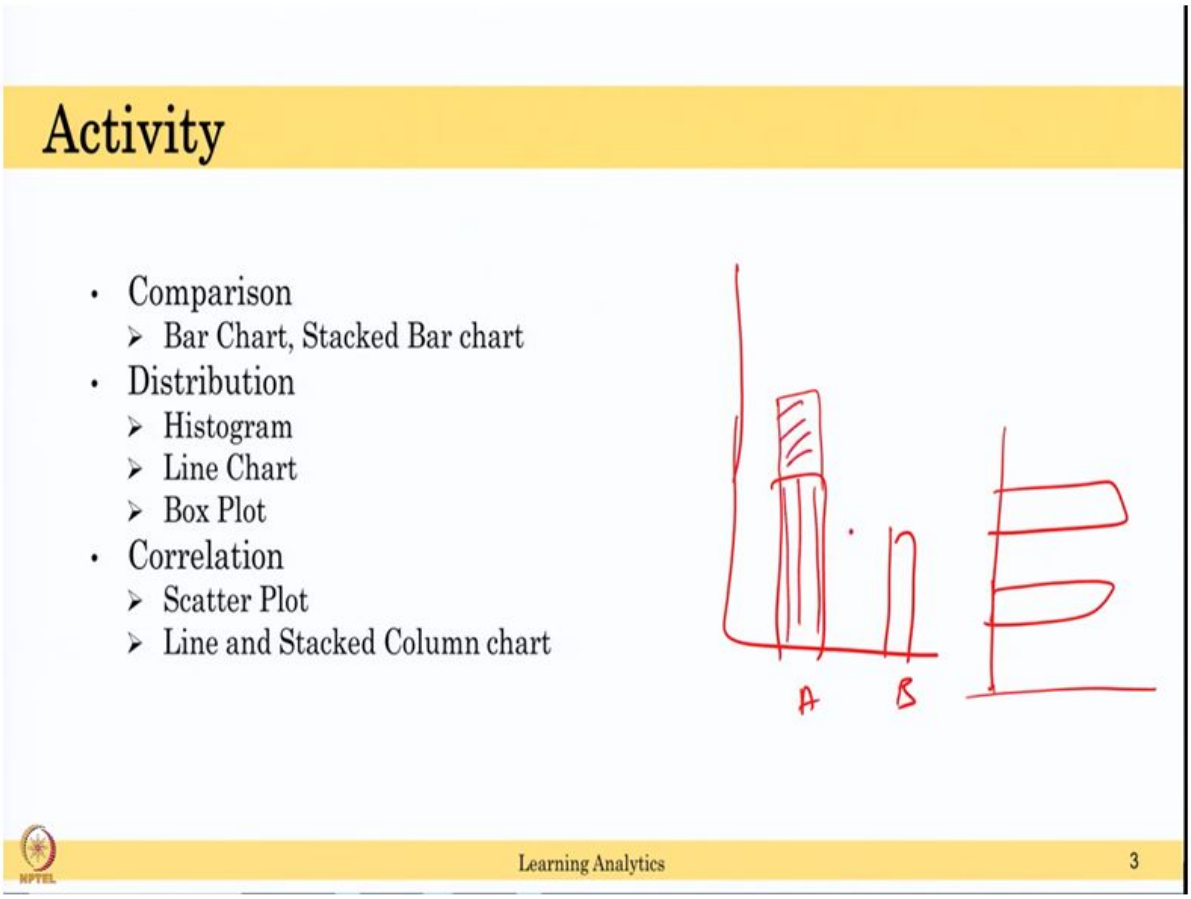
Learning Analytics Tools
Professor Ramkumar Rajendran
Educational Technology
Indian Institute of Technology, Bombay
Lecture no. 4.6
Comparing Charts

In this video, we will discuss how to pick the right chart for the right values.
(Refer Slide Time: 00:23)



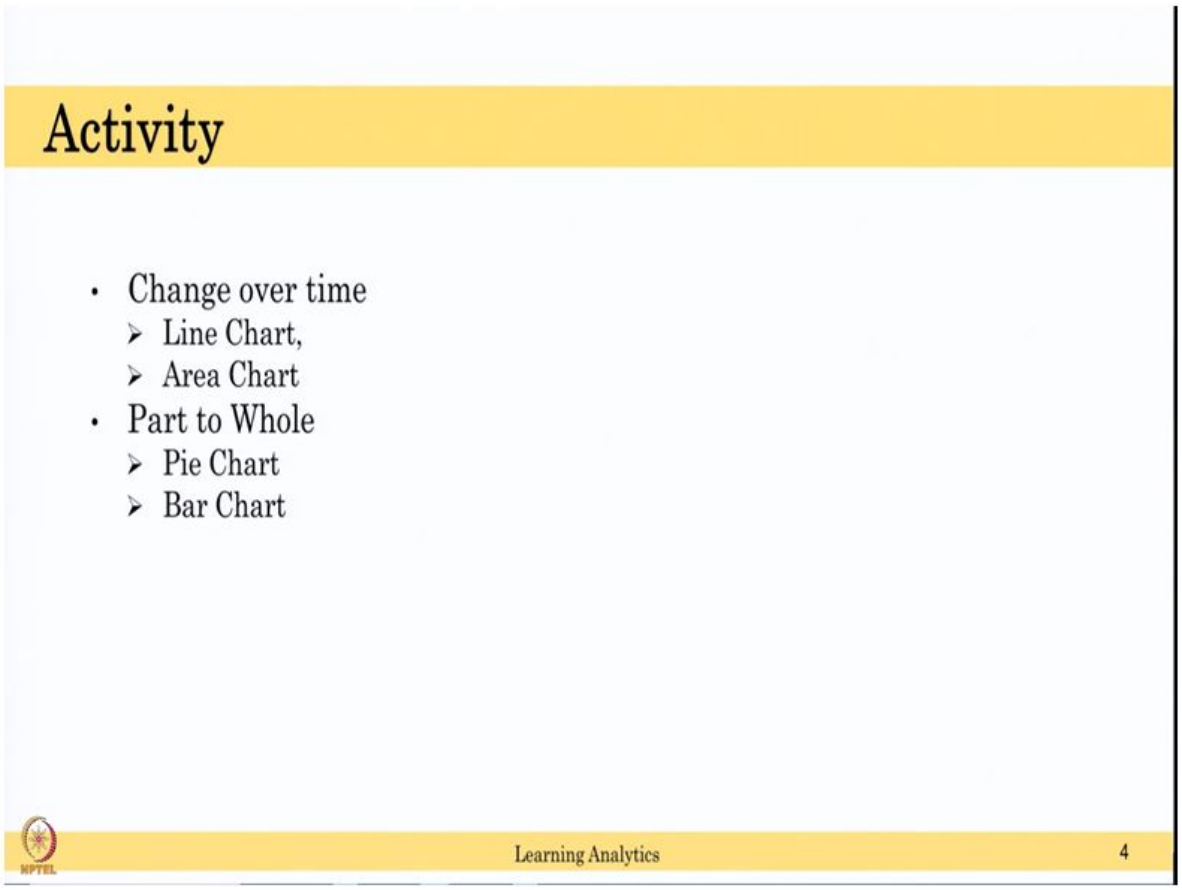
So, let us start with the activity. We saw the description of some basic plots like Bar chart, Pie chart, Stack bar chart or Histogram, Box plot something like that. Now, given this knowledge of basic charts. Select the suitable charts for the below requirements. For example, what chart you will select for comparison. What chart you will select for showing distribution and if you want to compute the correlation between A and B what particular chart you will use it?
And if you want to represent a variable which changes over time What kind of chart you use? Please pause this video and write down your answers for each of these values and these all requirements and resume the video to continue.

(Refer Slide Time: 01:14)



For comparison, we know that we can use a bar chart or stacked bar chart. We saw that you can compare with the bar chart easily. For distribution we can use histogram. Also line chart and box plot can be used for this purpose.

For Correlation Scatter plot is very good because it will show you all the data scattered across two axes or three axes. Then you can see whether there exist a relation between value A and value B. This correlation can also be observed from Line chart and stacked column chart.
So, here is a difference between bar chart and stack column chart. I just want to clarify the difference here. This bar chart is called a column chart (see fig). The bar chart in which we have a vertical columns is called column chart. The horizontal one is called simply referred as bar chart.
So, that is two different names, so if you hear someone saying that column chart it is again a simple bar chart in a column representation. If in a bar chart, the vertical columns are absent ,it is called a bar chart. So, the stacked bar chart is this (refer pic). No need to be like 100 percent that is stacked bar chart. So, in line chart (which usually shows the average value), this value may move like that. So, you can combine a line chart and stack bar chart to show the correlation also.
(Refer Slide Time: 03:26)



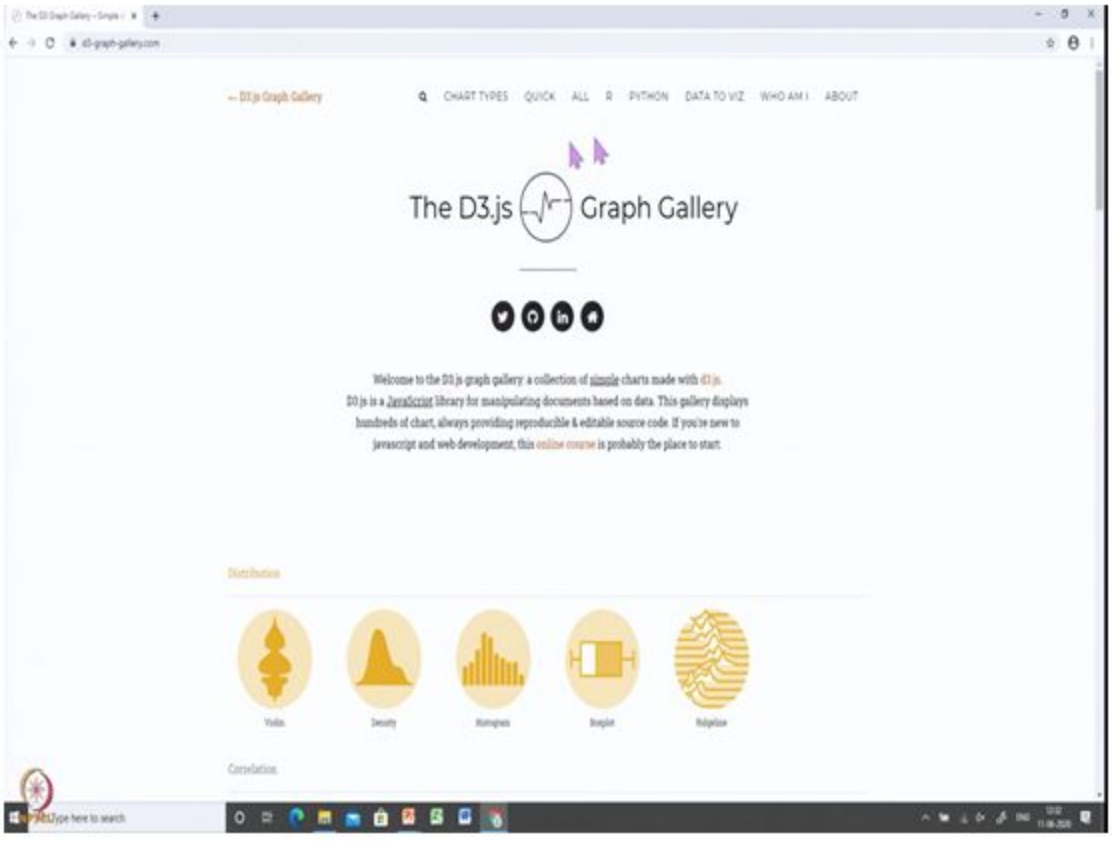
For depicting the change of overtime you can use a line chart or area chart. We have not discussed area chart but you can check that what is area chart. Area chart also can be used to show the change over time and line chart we discussed that line chart will be useful to show how the variable values change in over different time or different grades or different group or something like that.

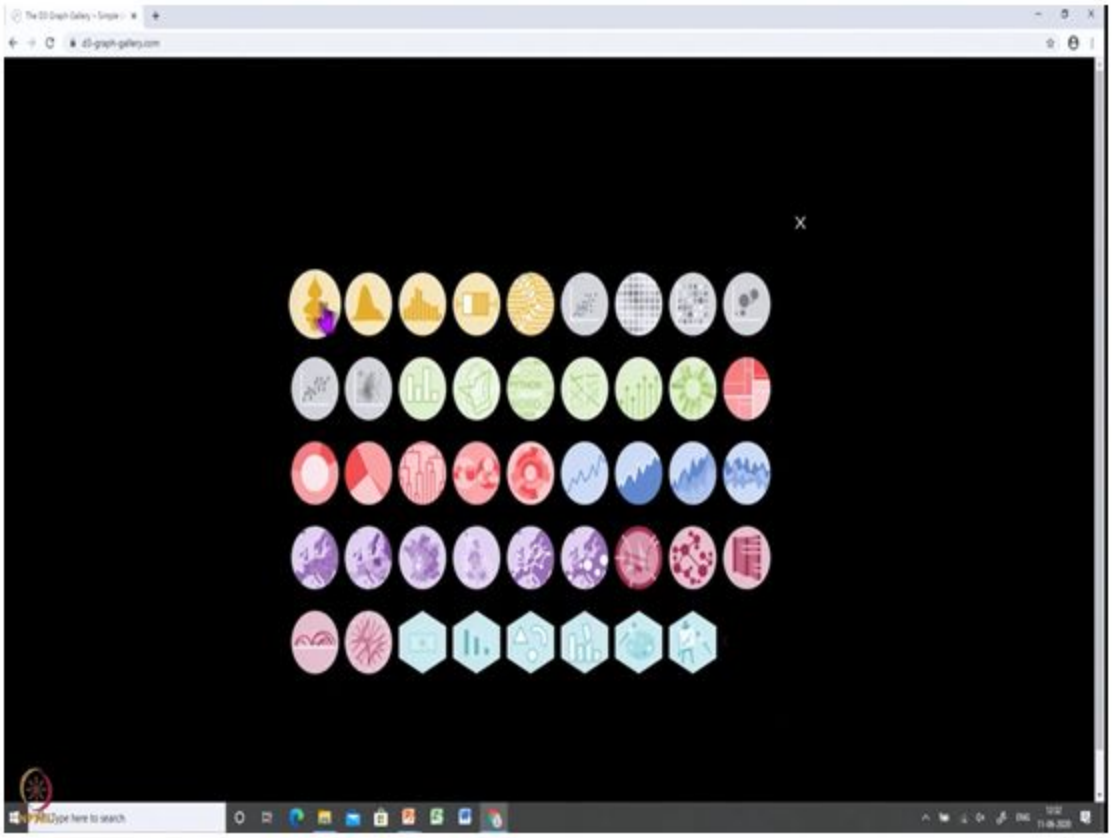
Or if you want to show the part of the whole we saw that pie chart will be useful to show the part of the whole.

(Refer Slide Time: 04:06)

Which Chart to Select?

- https://www.d3-graph-gallery.com/

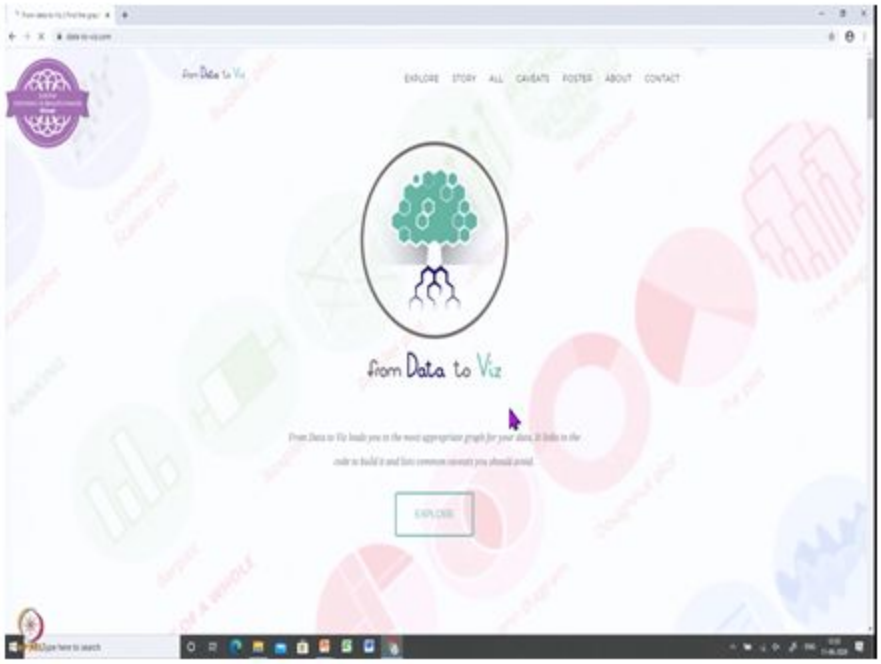


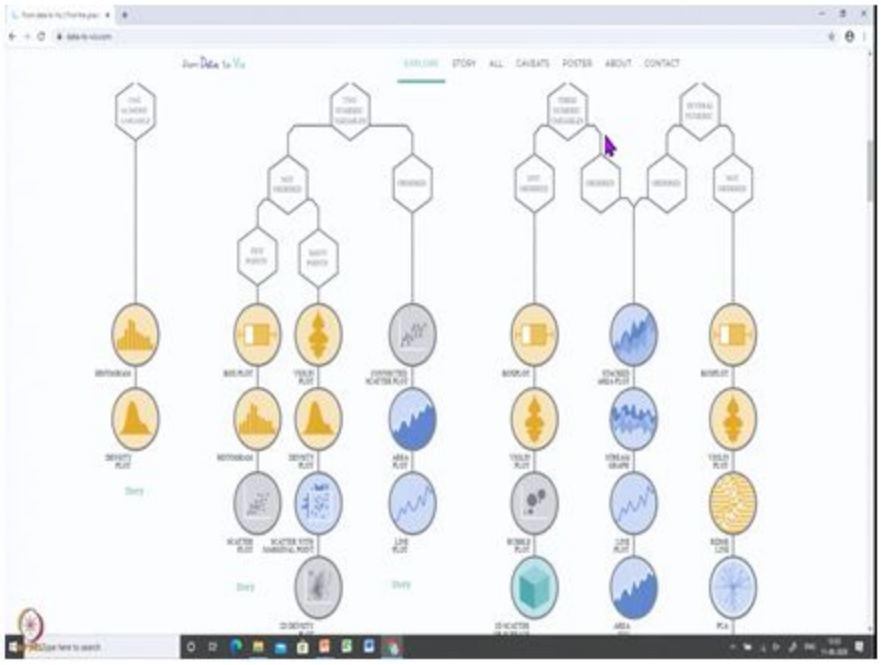


So, we discussed very basic charts and you might know all these charts already. You might have seen it. We discussed very basic charts and we discussed where these charts are used but if you want to know more about different kind of charts, not these basic charts. There are a lot of websites we talked about data visualization which data to pick.

What I recommend this website so the website is a D3.js, initially website is created for JAVA script but then the author created similar webpage for a Python and also for R. Let us look at what this website offers? So, if you look at the chart types all different chart types are given in this website like a violin chart or area chart or histogram and box plot everything. So, if you what to know what are the types, you can pick a chart based and explore is. Let us look at what is “data to visualization”?

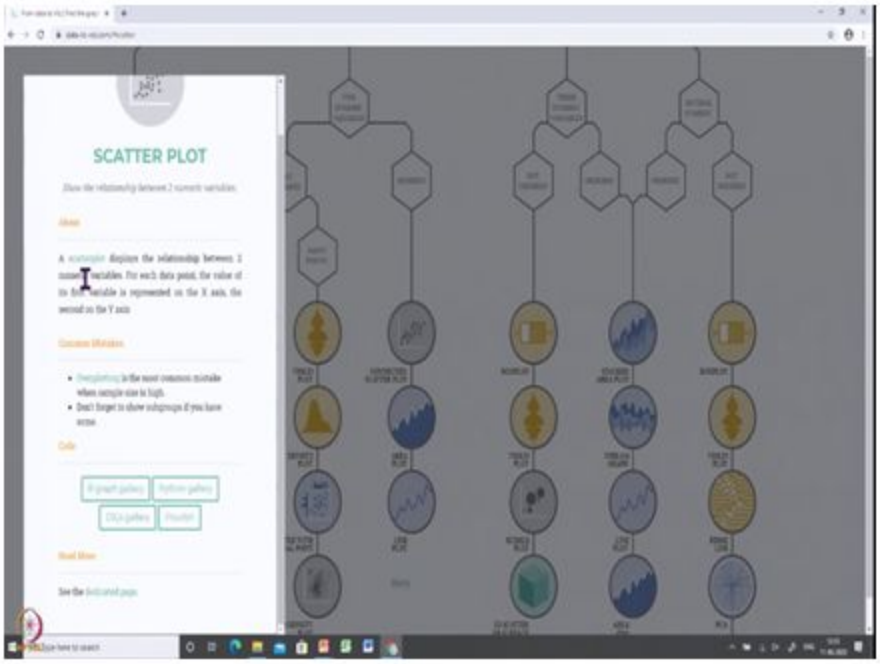
(Refer Slide Time: 05:17)



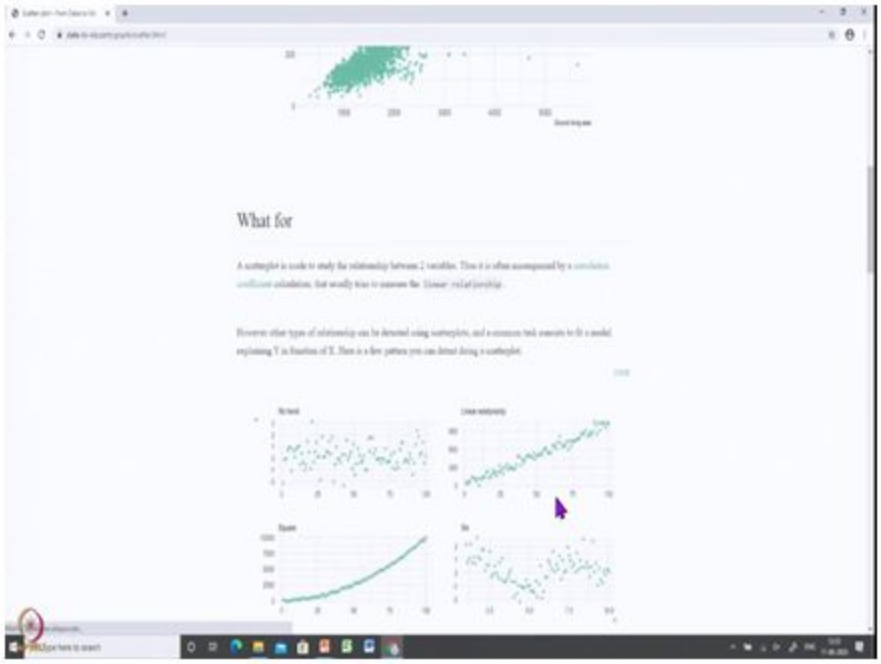


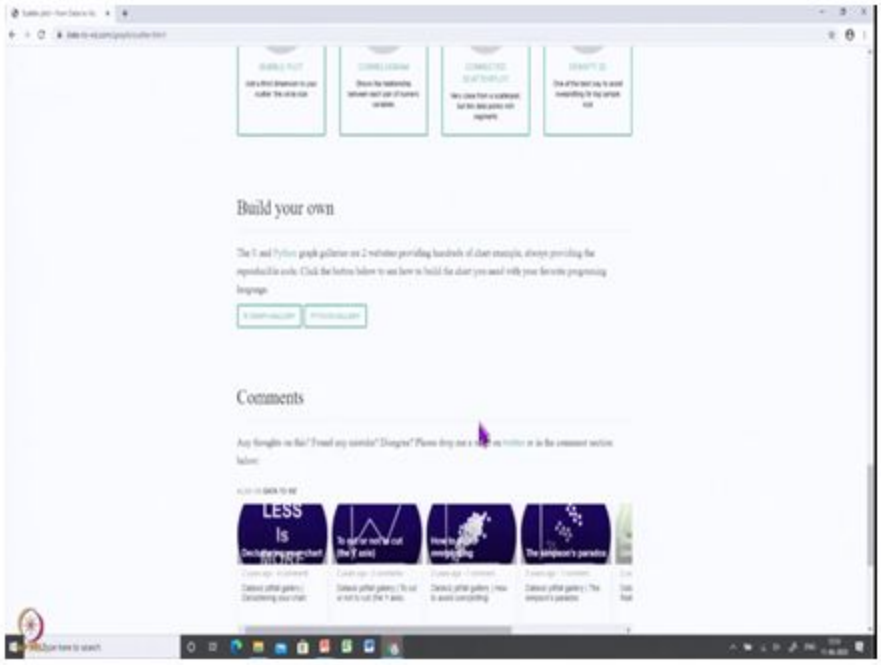
When you look at data visualization, you can explore this. You can select what type of data you have and if it is numeric data and you will see this kind of tree. And if you have one numeric value you can use a histogram or density plot. If you have two numerical value, what kind of chart you can use or if you have three numeric value what is the chart you can use? So, you can select based on this chart.

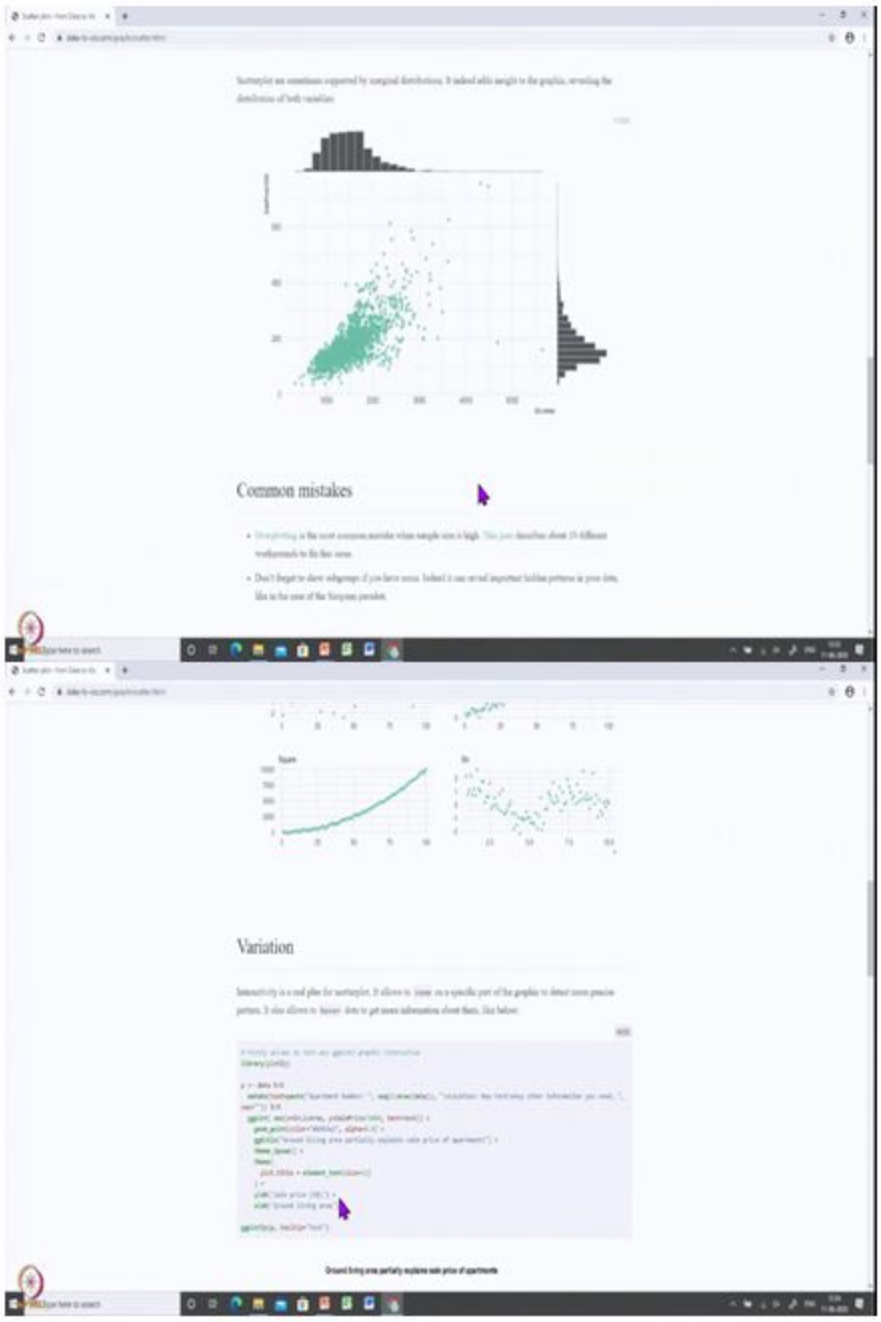
(Refer Slide Time: 05:48)


SCATTER PLOT
Show the relationship between 2 numeric variables.
About
A scatterplot displays the relationship between 2 numeric variables. For each data point, the value of its first variable is represented on the X axis, the second on the Y axis.
Common Mistakes
Overplotting is the most common mistake when example size is high.
Don't forget to show subgroups if you have some.
Code
R graph gallery
Python gallery
D3.js gallery
Plotly
Need More
See the dedicated page.


Definition
A scatterplot displays the relationship between 2 numeric variables. For each data point, the value of its first variable is represented on the X axis, the second on the Y axis.
Here is an example considering the price of 1460 apartments and their ground living area. This dataset comes from a Kaggle machine learning competition. You can read more about this example here.
Ground living area partially explains sale price of apartments
What for


What for

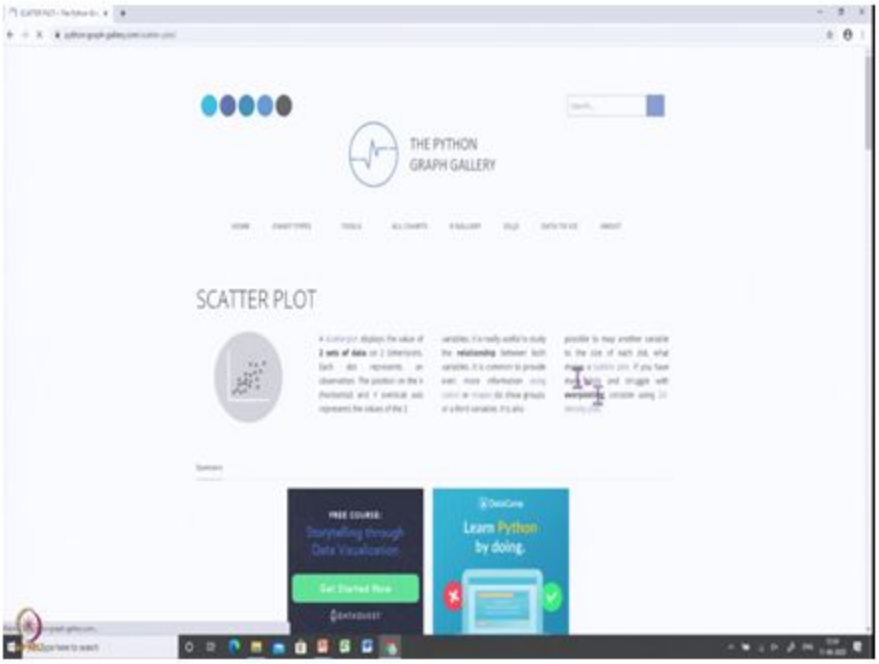

Build your own
Comments
LESS Is

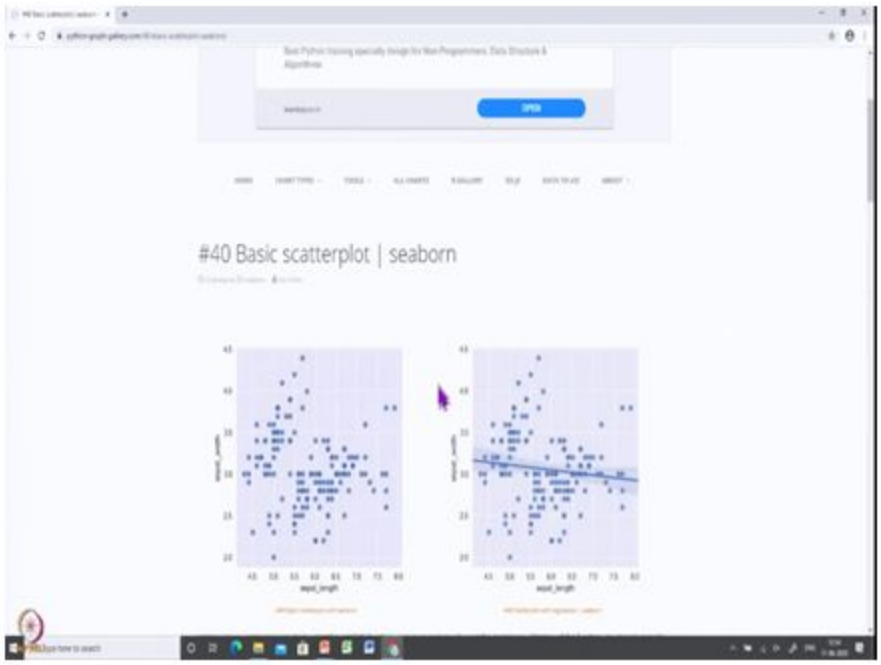


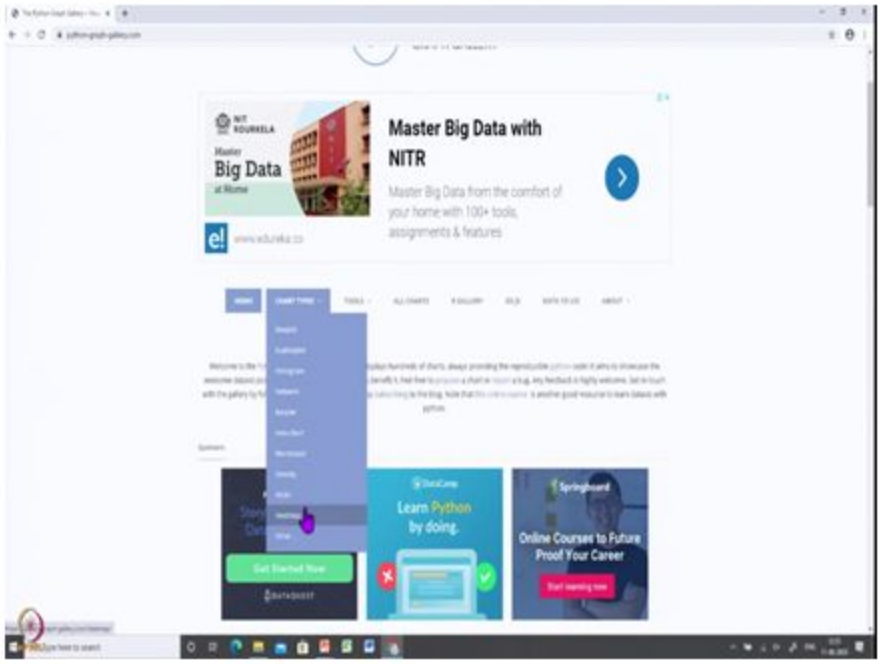
For example, if a select scattered plot, it tells about the scattered plot in detail. Lets see what is scattered plot? It gives you the scattered plot and what it is used for and the different variation of

scattered plot. What are the mistakes people make when they use the scattered plot? Interestingly when you look at this video the code is given. The code is generally for Java script.

(Refer Slide Time: 06:25)

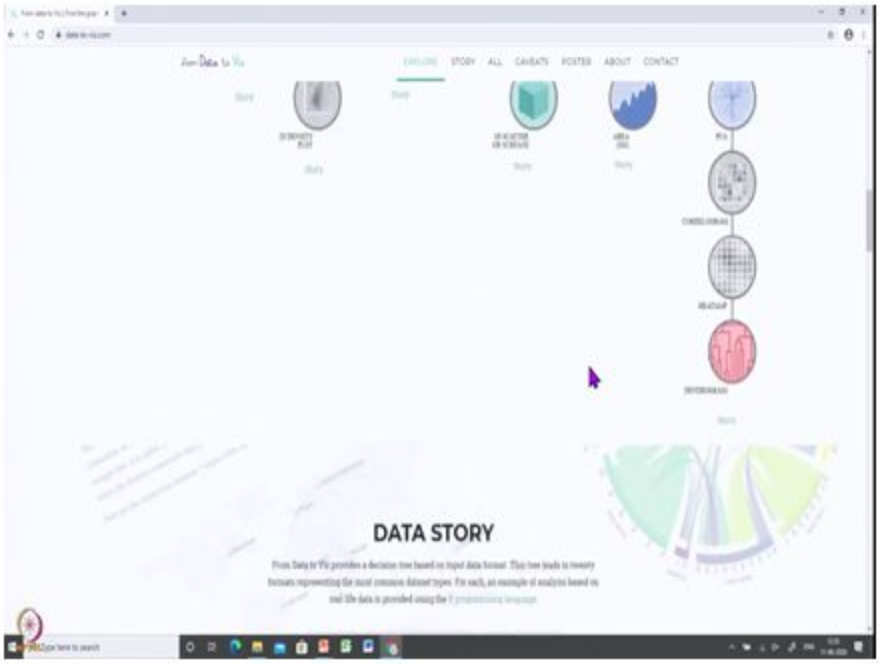


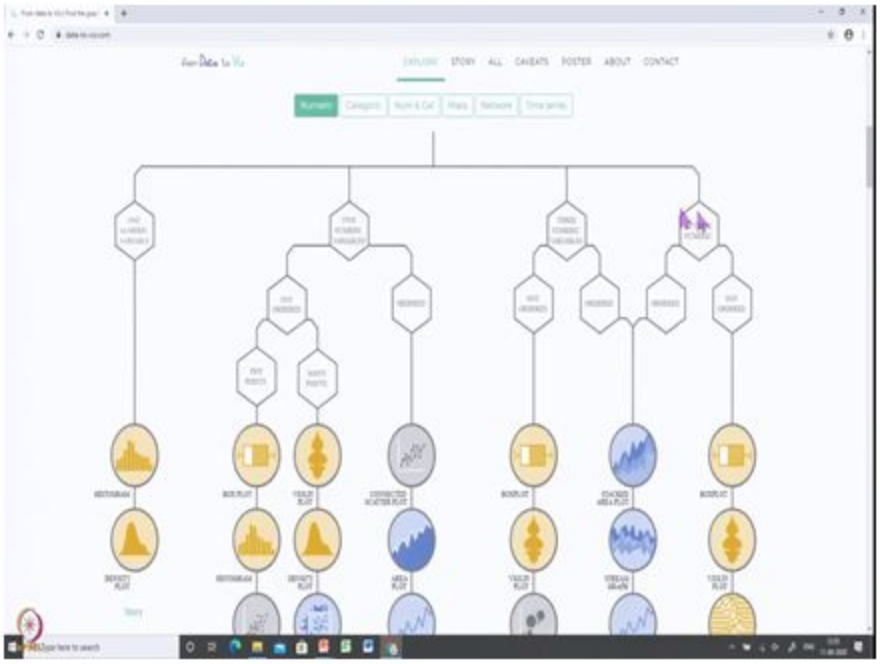


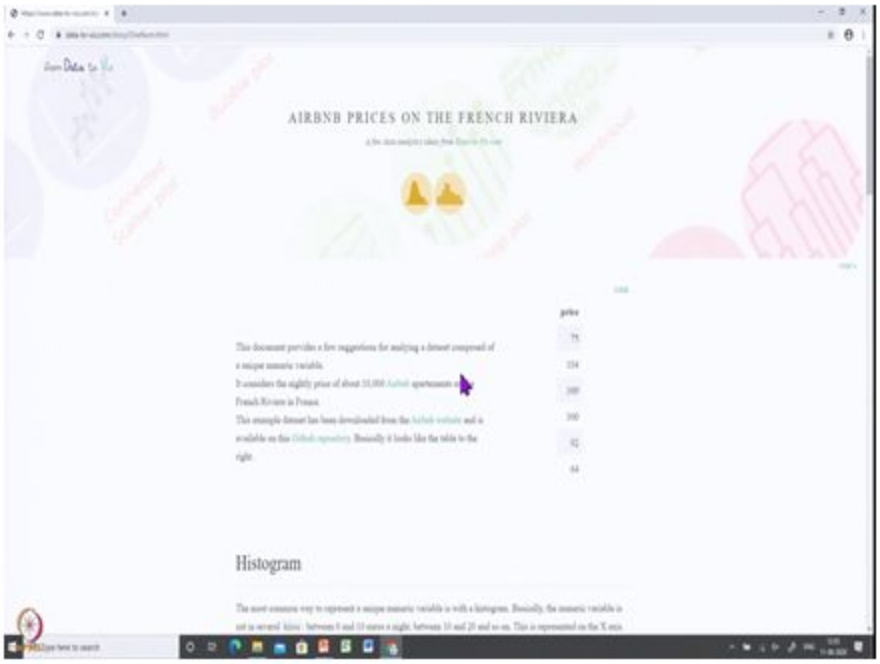


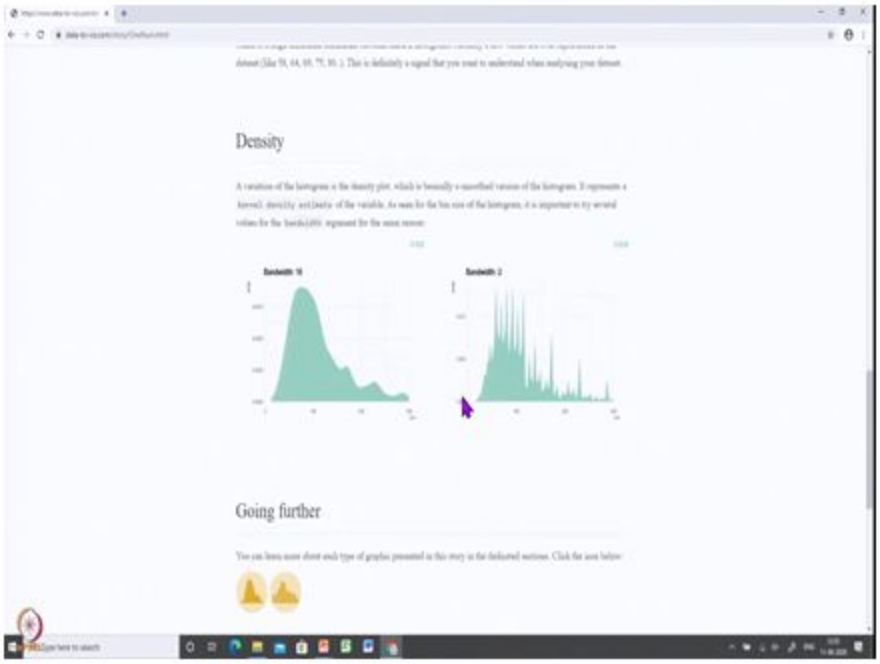
But if you want to go and use the code for Python, so let us go and look at the Python gallery and let us look at those charts and scattered plot. And in Python, the author used multiple libraries so when you Seaborn library Pandas library. So, you can use one of this library if you do not want to use Seaborn you can use another library as script is given for everything.

So here, the script which can be used to create this kind of plots I mentioned that you can use excel to create basic plots. If you want better-looking plots or no more modern plots you can go and use these scripts from the websites.
(Refer Slide Time: 07:24)









And let us look at the one Python gallery here are other chart types scattered plot, histogram, bar plot density and others in the gallery. I would suggest you to check the “data to visualization” tab and pick the chart based on your data type and requirement and you can look at the script. If you can use the script. Your script language can be Javascript or Python or R.

Or you can see the given examples of how we compare different values like A’s price and B’s price in France. We use different plots to show that difference between these two values. So, I request you do check this website and understand more about charts and check the scripts if

you want to see it, If you do not want to see the scripts its fine at least understand what is the chart is about? Which data type should be used for which.
(Refer Slide Time: 08:18)

So in this video, we talked about which chart is you should select for your data. We talked very basic plots and how to use basic plots to show correlation and distribution. And also, I request everyone to go and check the website I showed in the last slide. Thank you.