

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)

Activity

- We saw description of some basic plots. Select the suitable plots for below requirements
 - Comparison
 - Distribution
 - Correlation
 - Change over time
 - Part to whole

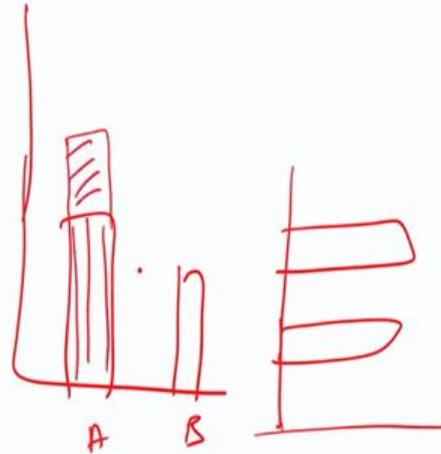


So, let us start with the activity. We saw the description of some basic plots like Bar chart, Pie chart, Stack bar chart or Histogram, Box plot something like that. Now, given this knowledge of basic charts. Select the suitable charts for the below requirements. For example, what chart you will select 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)

Activity

- Comparison
 - Bar Chart, Stacked Bar chart
- Distribution
 - Histogram
 - Line Chart
 - Box Plot
- Correlation
 - Scatter Plot
 - Line and Stacked Column chart



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)

Activity

- Change over time
 - Line Chart,
 - Area Chart
- Part to Whole
 - Pie Chart
 - Bar Chart



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.

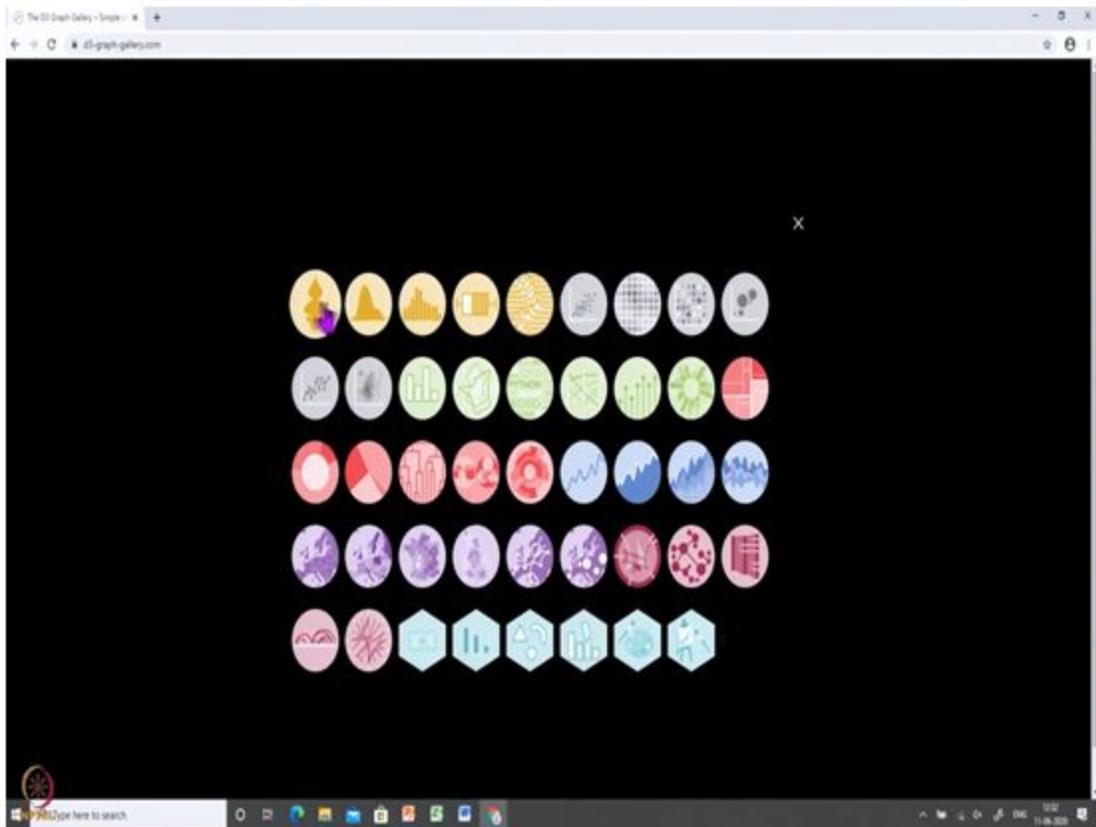
(Refer Slide Time: 04:06)

Which Chart to Select?

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



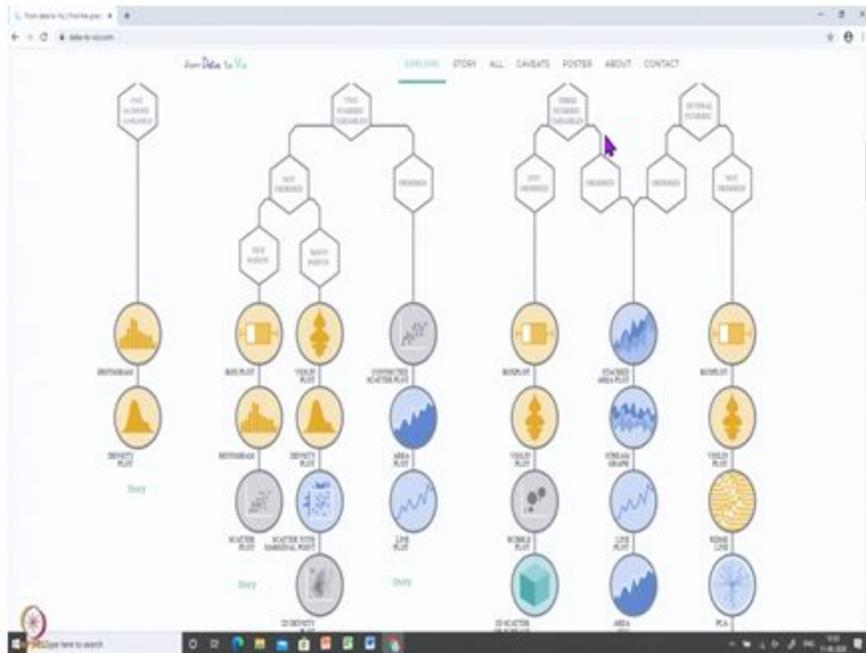
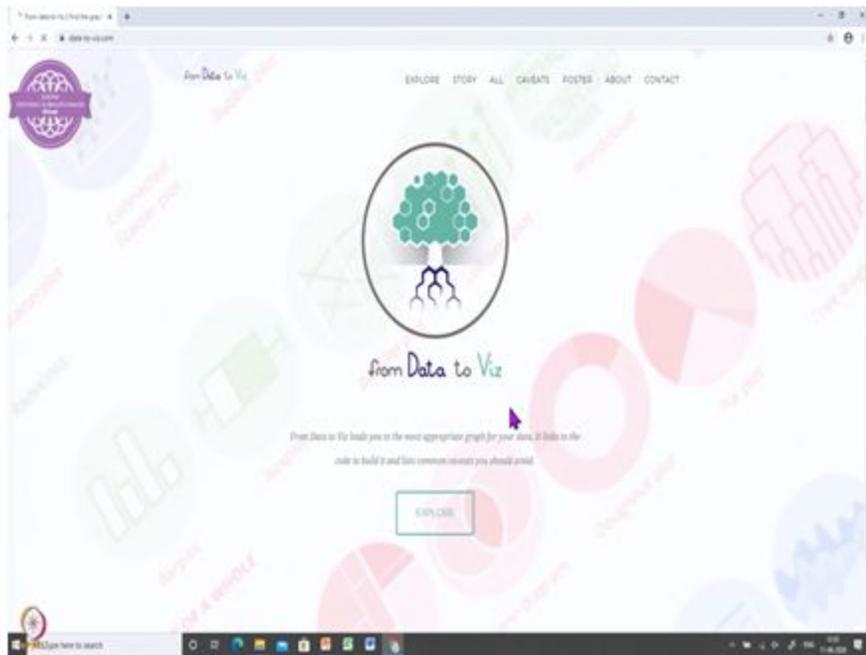
The screenshot shows a web browser displaying the D3.js Graph Gallery website. The browser's address bar shows the URL <https://www.d3-graph-gallery.com/>. The website has a navigation menu with links for CHART TYPES, QUICK, ALL, PYTHON, DATA TO VIZ, WHO AM I, and ABOUT. The main heading is "The D3.js Graph Gallery" with a logo featuring a circle and a pulse line. Below the heading are social media icons for Twitter, YouTube, LinkedIn, and Facebook. A welcome message reads: "Welcome to the D3.js graph gallery a collection of simple charts made with D3.js. D3.js is a JavaScript library for manipulating documents based on data. This gallery displays hundreds of chart, always providing reproducible & editable source code. If you're new to javascript and web-development, this online course is probably the place to start." Below this is a "Distribution" section with five icons representing different chart types: Violin, Density, Histogram, Boxplot, and Ridgeplot. At the bottom, there is a "Correlation" section. The Windows taskbar is visible at the bottom of the browser window.



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.

Notes

A **scatterplot** displays the relationship between 2 numeric variables. For each data point, the value of the first variable is represented on the X axis, the second on the Y axis.

Common Mistakes

- Don't plot on the most relevant variable when category size is high.
- Don't forget to show subgroups if you have more.

Guides

Read More

See the full content page.

Definition

A scatterplot displays the relationship between 2 numeric variables. For each data point, the value of the first variable is represented on the X axis, the second on the Y axis.

Here is an example considering the price of 1400 apartments and their ground living area. This dataset comes from a Kaggle machine learning competition. You can read more about this example here.

Searching area partially explains sale price of apartments

What for

Extension: Two Variables

What for

A scatterplot is used to study the relationship between 2 variables. This is often accompanied by a correlation coefficient calculation, that usually tries to measure the linear relationship.

However other types of relationship can be detected using scatterplots, and a regression tool attempts to fit a model explaining Y as function of X. Here is a line pattern you can detect using a scatterplot.

Build your own

The 3 and 7 colors graph gallery are 2 relations providing hundreds of chart examples, always providing the spreadsheet file. Click the buttons below to see how to build the chart you need with your favorite programming language.

3 Colors Gallery 7 Colors Gallery

Comments

Any thoughts on this? Found any number? Disagree? Please drop me a message or write in the comment section below.

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)

The screenshot shows the Python Graph Gallery website. The main heading is "THE PYTHON GRAPH GALLERY". Below it, there is a navigation menu with links: HOME, SCATTER PLOT, TABLE, ALL CHARTS, EXAMPLE, HELP, DATA SETS, ABOUT. The "SCATTER PLOT" section is highlighted. It features a circular icon with a scatter plot and a brief explanation: "A scatter plot displays the value of 2 sets of data on 2 dimensions. Each dot represents an observation. The position on the horizontal axis represents the value of the 1st variable. It is really useful to study the relationship between both variables. It is common to provide each more information using color or size of the data points of a 2nd variable. It is also possible to map another variable to the size of each dot, what this is useful for if you have a 3rd variable and struggle with overlapping markers using size, color, etc." Below this, there are two promotional banners: "FREE COURSE: Storytelling through Data Visualization" and "Learn Python by doing".

The screenshot shows the "Input Format" and "Output" sections of the Python Graph Gallery website. The "Input Format" section contains two tables. The first table is for "2 numerical variables" and the second is for "2 numerical variables and 1 categorical".

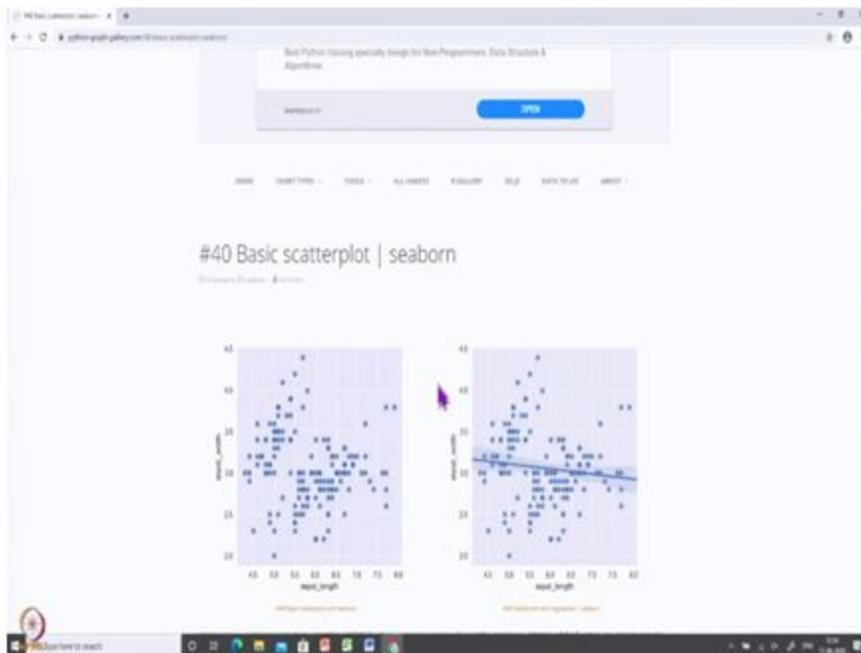
Variable 1	Variable 2
2.3	1.1
8.4	4.3
2.9	2.1
5.5	9.4
2.2	7.8

CODE: 2 numerical variables

Variable 1	Variable 2	Group
2.3	1.1	A
8.4	4.3	A
2.9	2.1	A
5.5	9.4	B
2.2	7.8	B

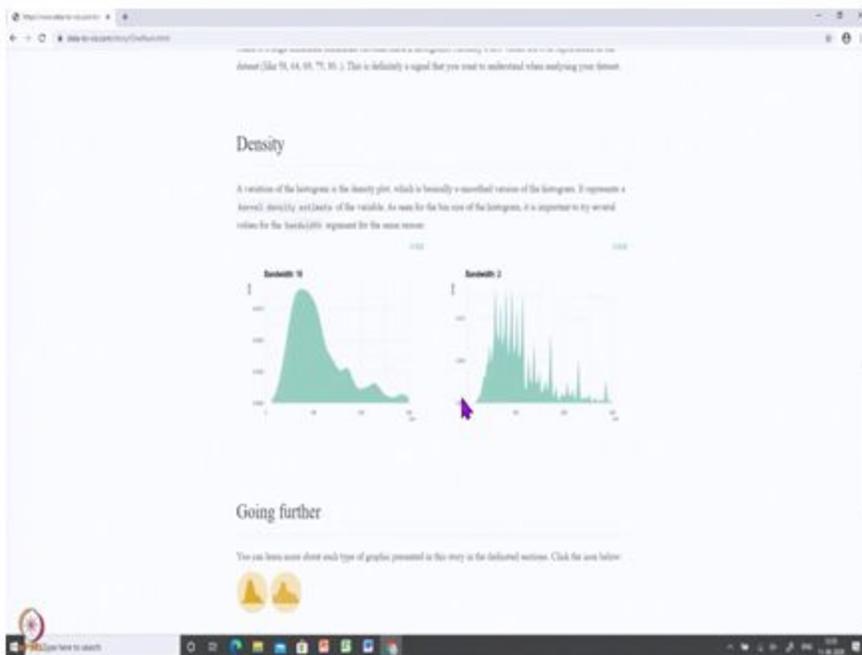
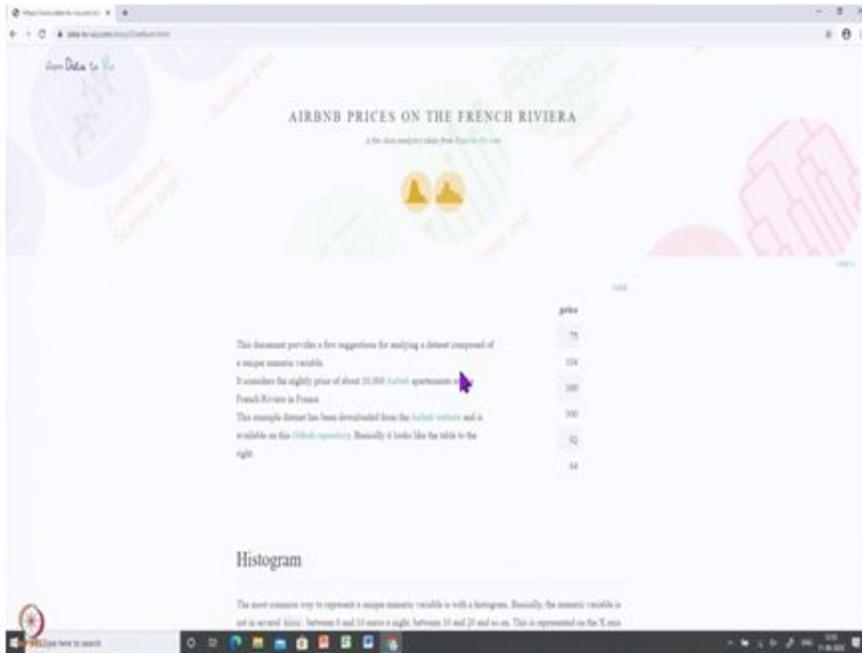
CODE: 2 numerical variables and 1 categorical

The "Output" section displays a grid of scatter plots. The first row shows four plots: "All observations in one color", "All observations in different colors", "All observations in one color", and "All observations in one color". The second row shows four plots: "All observations in one color", "All observations in one color", "All observations in one color", and "All observations in one color".



The screenshot shows a web browser window displaying a promotional banner for "Master Big Data with NITR". The banner includes the NITR logo, a building image, and the text: "Master Big Data with NITR. Master Big Data from the comfort of your home with 100+ tools, assignments & features." Below the banner is a navigation menu with options like Home, About, Contact, etc. There are also several promotional cards for learning Python and big data.

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.



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.