

## In21-S4-CS3121 - Introduction to Data Science

### Activity - Descriptive Data Analysis - In Class Activity

Group K

# Descriptive Data Analysis

Attribute Type	Categorical		Metric	
	Nominal	Ordinal	Discrete	Continuous
Pie Chart	Yes	Yes	No	No
Tag Cloud	Yes	Yes	Yes	No
Bar Chart	Yes	Yes	Yes	No
Clustered/Stacked Bar Chart	Yes	Yes	Yes	No
Step Chart	No	Yes	Yes	No
Box Plot	No	No	Yes	Yes
Histogram	No	No	Yes	Yes
Cumulative Histogram	No	No	Yes	Yes

### 1. Pie Charts

The circular shape of a pie chart inherently implies that the slices represent parts of a whole. However, metric data often does not represent a whole itself. In such cases, pie charts should not be used to represent metric data.

### 2. Tag Cloud

Tag Cloud visualizes data by representing keywords in a font size proportional to the frequency of the data relative to other keywords. So with the definition, We can say that continuous metrics can't be represented with Tag Clouds.

### **3. Bar Chart**

The width of each bar in a bar chart implies a specific data range, but with continuous data, there's no inherent meaning to the bar width. This can lead viewers to misinterpret the actual magnitude of differences between categories. So it is not recommended to use bar charts for representing continuous data.

### **4. Clustered/Stacked Bar Chart**

This type of bar chart is used to compare different items at the same time and show the composition of each item. These are excellent to show the relationship between variables but only limited to use with Nominal, Ordinal and discrete data.

### **5. Step Chart**

Step charts are unsuitable for nominal or continuous data because they require a natural order or discrete intervals to be accurately represented. However, it is appropriate for discrete and ordinal data because it efficiently visualizes changes or transitions between different values or categories, highlighting trends and patterns in the data.

### **6. Box Plot**

Box plots are not appropriate for nominal or ordinal data since they require numerical values to calculate quartiles and generate the plots. However, it is appropriate for discrete and continuous data since it efficiently summarizes the distribution of numerical values, revealing the data's central tendency and variability.

### **7. Histogram**

Histograms can effectively represent ordinal, discrete, and continuous data by displaying the frequency distribution within intervals. However, they are not suitable for nominal data, which lacks a specific order.

### **8. Cumulative Histogram**

Cumulative histograms display the accumulation of frequency within intervals, making them useful for ordinal, discrete, and continuous data. However, they're not suitable for nominal data, which lacks a specific order.

## **Group Members**

210518H Ranasinghe K.S

210450P Pathirana L.P.T.R

210483T Prabashwara D.G.H

210588U Senarathna L.P.S.U.K

210460V Perera I.T.M