**Batch:** H2-4 **Roll No.:**16010122257

**Experiment 04**

|  |
| --- |
| **Title:** Introduction to various charts |

# Objective:

# 1. To learn how to use various charts for visualizing preprocessed data

# Pie chart

# Donut chart

# Lollypop chart

# Boxplot

# Bubble chart

# Bump Chart

# Funnel chart

# Water Chart

# Bar chart /column chart / Stacked chart

# Area Chart

# Pareto chart

# Tree maps

# Scatter plot

# Gantt chart

# Word map

# 2. To learn following concepts

# Parameters

# Combine,Sets

# Label, Title and caption

# Working with metadata

# Working with filter

# Font, Size, colour, and Border,

# Calculated field, Number function, string function

# Conditional formatting, logical function

# Exporting the charts

# 3. To learn interpretation of each charts/plots

# Course Outcome:

# CO1: Learn how to locate and download datasets, extract insights from that data and present their findings in a variety of different formats.

# CO3: Apply data visualization best practices

# Books/ Journals/ Websites referred:

1.[*https://lms-kjsce.somaiya.edu/pluginfile.php/85875/mod\_folder/content/0/visual-data-storytelling-with-tableau\_LindyRyan.pdf?forcedownload=1*](https://lms-kjsce.somaiya.edu/pluginfile.php/85875/mod_folder/content/0/visual-data-storytelling-with-tableau_LindyRyan.pdf?forcedownload=1)

2.<https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization>

3.[https://www.tableau.com/learn/whitepapers/which-chart-or-graph-is-right-foryou](https://www.tableau.com/learn/whitepapers/which-chart-or-graph-is-right-for-you)

4.<https://www.kaggle.com/uciml/adult-census-income>

5.<https://archive.ics.uci.edu/ml/datasets/adult>

# Resources used:

1. <https://www.kaggle.com/uciml/adult-census-income>
2. <https://archive.ics.uci.edu/ml/datasets/adult>
3. <https://www.tableau.com/>

# Theory (About various charts explored):

# Data visualization is the process of turning your data into graphical representations that communicate logical relationships and lead to more informed decision-making.In short,data visualization is the representation of data in a graphical or pictorial format.

# It allows key decision-makers to see complex analytics in a visual layout, so they can identify new patterns or grasp challenging concepts.

# Common types of charts:

# Bubble chart:1. Comparison of Three Variables: Bubble charts are valuable when you want to compare three different variables simultaneously. The x-axis and y-axis help you compare the relationships between two variables, while the size of the bubbles provides insight into the magnitude or intensity of a third variable. This makes it easier to identify trends, patterns, and outliers.

# 2. Visualizing Multivariate Data: Bubble charts can be used to visualize multivariate or complex datasets where more than two variables are involved. For example, in financial analysis, you might use a bubble chart to show the risk, return, and market capitalization of different stocks.

# 

# Symbol maps:

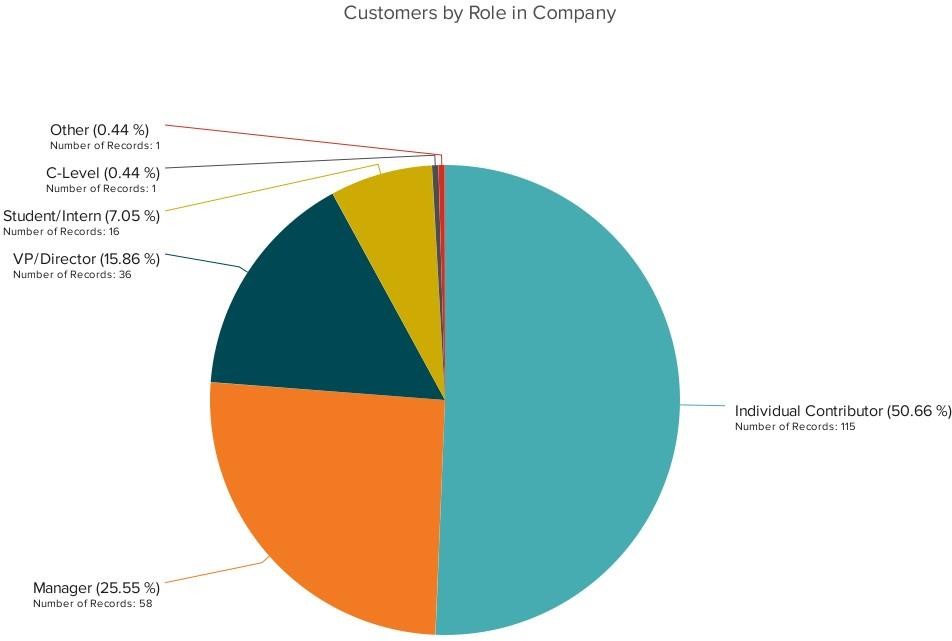
# A symbol map is a type of data visualization used in geographic information systems (GIS) and data analytics. It's designed to represent and display data on a map by using symbols or icons to represent specific locations, features, or attributes. Symbol maps are particularly useful when you want to visualize data that is associated with geographic or spatial information.

# 

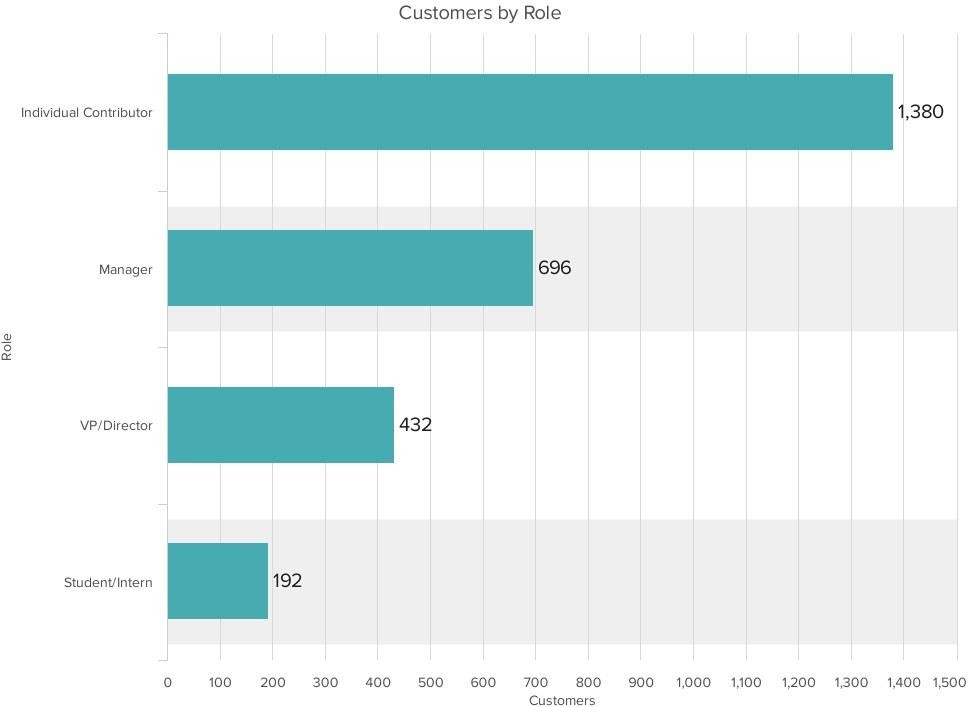
# Histogram:A histogram is a graph used to represent the frequency distribution of a few data points of one variable. Histograms often classify data into various “bins” or “range groups” and count how many data points belong to each of those bins.

# 

**Pie chart:**Pie charts are powerful for adding detail to other visualizations. Alone, a pie chartdoesn’tgivetheviewerawaytoquicklyandaccuratelycompareinformation.Sincetheviewer has to create context on their own, key points from your data are missed. Theyare used to drill down on other visualizations. A pie chart shows a static number andhow categories represent part of a whole -- the composition of something. A pie chartrepresents numbers in percentages, and the total sum of all segments needs to equal100%.

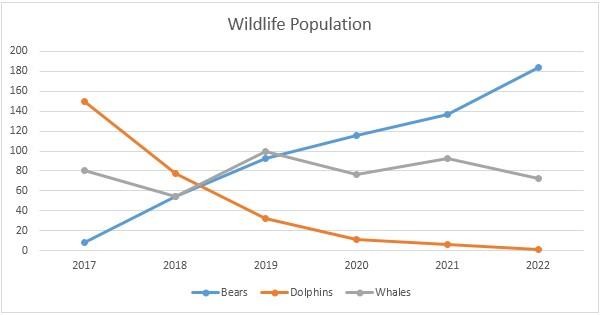


# Bar chart:A bar graph, basically a horizontal column chart, should be used to avoid clutter when one data label is long or if you have more than 10 items to compare. This type of visualization can also be used to display negative numbers. You can use them to quickly compare data across categories, highlight differences, show trends and outliers, and reveal historical highs and lows at a glance. Bar charts are especially effective when you have data that can be split into multiple categories.



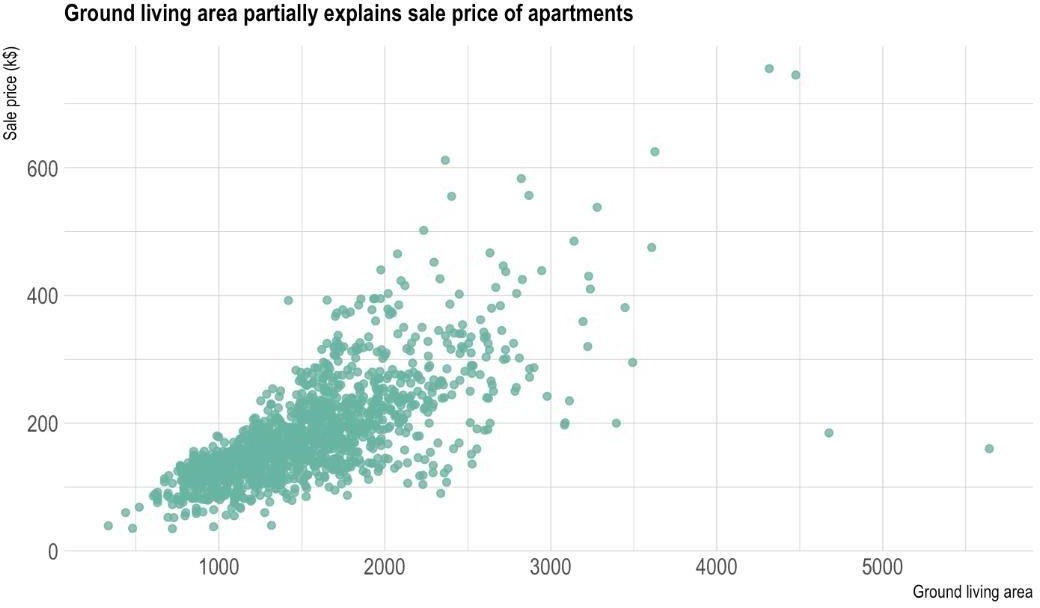
# LineChart

# The line chart, or line graph, connects several distinct data points, presenting them as one continuous evolution.Line charts are used to view trends in data,usually overtime(like stock price changes over five years or website page views for the month). The result is a simple, straightforward way to visualize changes in one value relative to another.



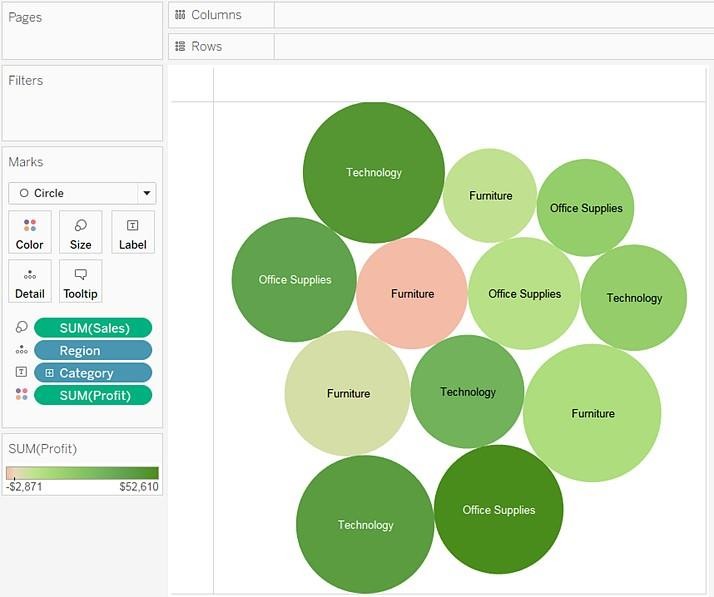
**Scatter plots:**

Scatter plots are an effective way to investigate the relationship between different variables, showing if one variable is a good predictor of another, or if they tend to change independently. A scatter plot presents lots of distinct data points on a single chart.Thechart can then be enhanced with analytics like cluster analysis or trendlines.A scatter plot or scattergram chart will show the relationship between two different variables or it can reveal the distribution trends. It should be used when there are many different data points,and you want to highlight similarities in the dataset.This is useful understanding the distribution of your data.



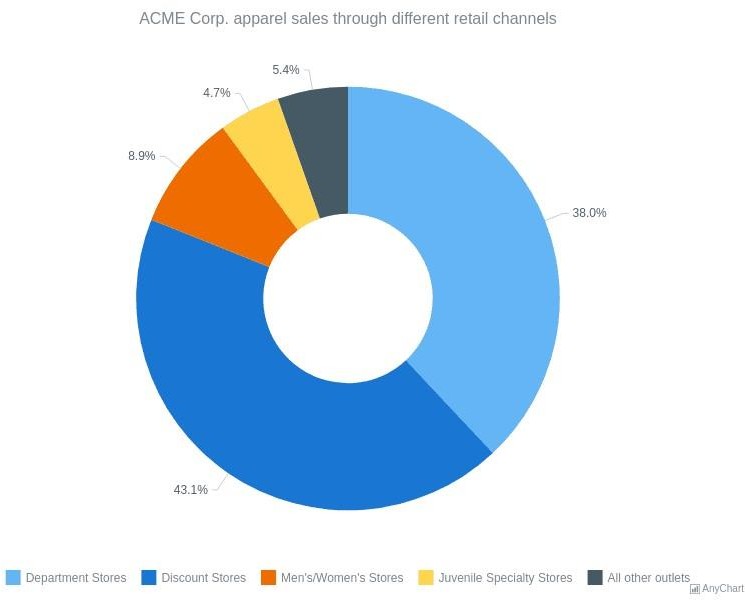
**PackedBubbleChart**

A bubble chart is similar to a scatter plot in that it can show distribution or relationship.There is a third data set, which is indicated by the size of the bubble or circle. In this variation, however, you’ll replace the data points with bubbles. You’ll also vary the sizes of the bubble to represent a third data set.As with a scatter chart, a bubble chart does not use a category axis.Rather,you’ll plot the datasets asX-values,Y-values and now,Z-values (bubble size).



**DonutChart**

The donut chart is a variant of the pie chart, with a hole in its center, and it displayscategoriesasarcsratherthanslices.Bothmakepart-to-wholerelationshipseasytograspat a glance. They differ from line charts, area charts, column charts, and bar graphs inthat they can't show changes over time. Also, Donut Charts are more space-efficientthan Pie Charts because the blank space inside a Donut Chart can be used to displayinformation insideit.



# Following points should be written by students

# Different charts explored for data visualization:

# Packed bubble chart.

# Box-and-whisker plots

# Scatter plot

# Histogram

# Pie chart

# Bar chart

# Working: Using tableau software

# Exported of each chart output with variations explored on your dataset

# Interpretation (what each chart signifies)

# Packed bubble chart:

# In this one I took 1 dimension and 2 Measures,so packed bubble was suitable for it. A packed bubble chart is suitable for this sales data dataset because it can show the relative quantity of each customer in a visual and compact way.Here,it’s comparing a customer with the highest quantity of products purchased/the ones with the highest availability.

# 

# Box and whisker plots:

# A box and whisker plot is suitable for this sales data dataset because it can show the distribution and variation of the quantity for each state in a simple and compact way.If the boxes are of different lengths or if the whiskers vary significantly, it indicates differences in the variability of sales quantities, which might be of interest in a sales analysis.Outliers in the sum(quantity) for a specific native state can be spotted as individual points beyond the "whiskers" of the box plot. This can be valuable information, indicating unusually high or low sales in specific states.

# Here,it’s comparing the state(region) with the highest sum of quantity.

# 

# Scatter plot(with fully coloured dots):

# A scatter plot is a type of chart that displays data using dots that represent the coordinates of two variables1. A scatter plot is ideal for comparing the product and the total in Tableau because it can:

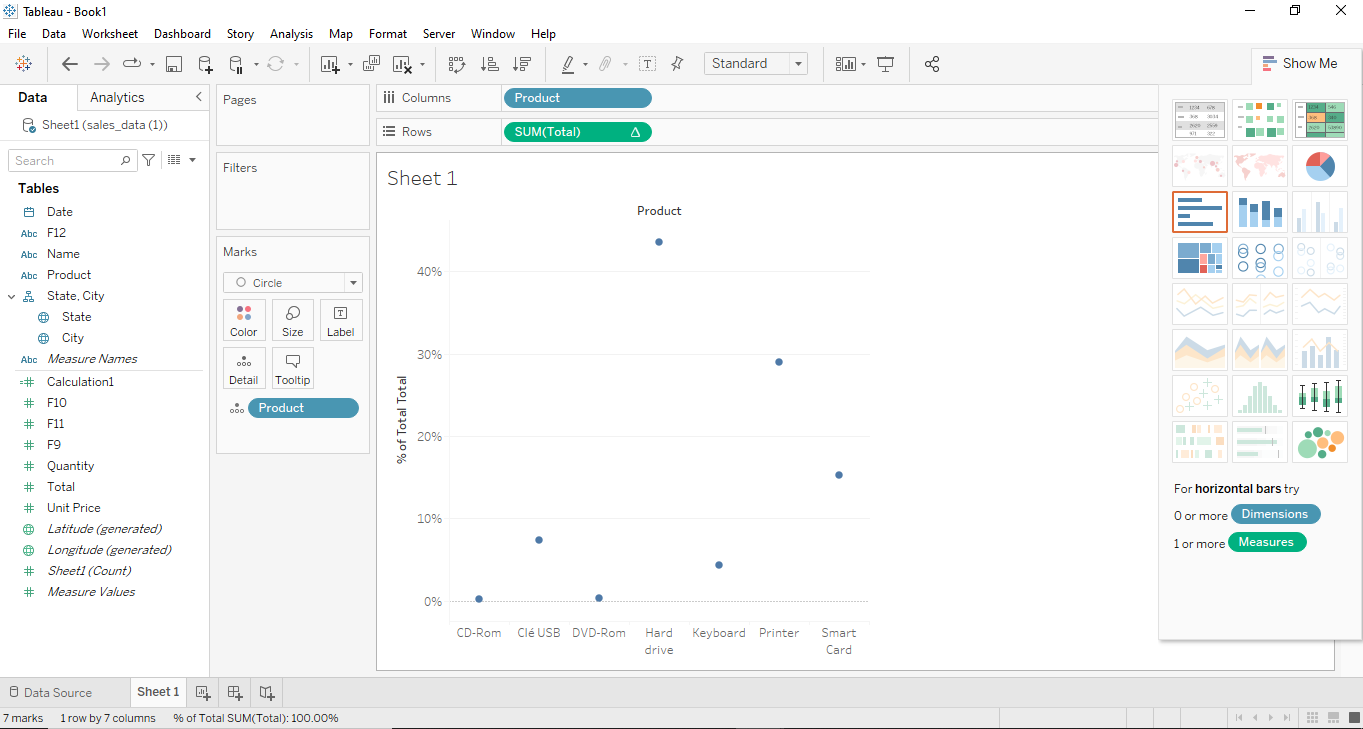
# Show the relationship between the product and the total by using the position of the dots. For example, you can see if there is a positive, negative, or no correlation between the two variables2.

# Show the variation of the product and the total by using the spread of the dots. For example, you can see if there is a lot of diversity or similarity in the data2.

# Show the outliers or anomalies in the data by using the dots that are far away from the rest of the data. For example, you can see if there are any products that have unusually high or low totals2.

# Show the clusters or groups in the data by using the color or shape of the dots. For example, you can see if there are any patterns or trends in the data based on different categories or attributes3.

# Here,hard drive product’s quantity is the highest total.



**Histogram:**

It is appropriate for the following reasons:

1. Visualization of Distribution:

Histograms provide a clear visual representation of the distribution of a quantitative variable (quantity in this case). By binning the quantities into intervals and representing the count of occurrences within each interval as the height of the bars, histograms give an immediate sense of how the quantities are distributed across different ranges.

2. Understanding Frequency Counts:

The height of each bar in the histogram corresponds to the count of occurrences (frequency) within that particular quantity interval. This allows for a direct comparison of how many sales fall into each quantity range, providing insights into the most common sales quantities and their frequencies.

3. Identification of Patterns:

Histograms can reveal patterns and trends in the data. For instance, you might observe a normal distribution (bell-shaped curve) or identify multiple peaks indicating different customer segments or product categories with distinct sales quantities. These patterns are essential for understanding customer behavior and preferences.

4. You can adjust the bin sizes to see the data from different perspectives. Fine-tuning the bin width can reveal detailed patterns in the sales quantities, while broader bins might highlight broader trends. This flexibility allows for exploration and deeper understanding of the data.

# 

# Pie chart: A pie chart would be good for comparing the different products and sum(total) proportionately of each because:A pie chart can display relative proportions of multiple classes of data. For example, you can see how much each product contributes to the total sales, profit, or quantity by using the size of the slices1.

# A pie chart can summarize a large data set in visual form. For example, you can see the overall picture of the data without having to look at the individual numbers or calculations1.

# A pie chart can be easily understood by the audience due to its widespread use in business and media. For example, you can communicate your data effectively and clearly with a simple and familiar chart type2.

# 

# Bar chart: This bar chart would be a good choice for comparing name, product, city with measure values in this sales data dataset because:

# It can show the relative magnitude of each name, product, or city by using the length of the bars. For example, you can see which name has the highest or lowest sales, profit, or quantity.

# It can allow you to sort, filter, or group the data by different dimensions or measures. For example, you can sort the bars by ascending or descending order, filter the data by a specific product or city, or group the data by name and product.

# It can enable you to add labels, tooltips, or annotations to the bars to provide more information or context about the data. For example, you can show the exact values of each bar, display additional details when hovering over a bar, or add notes to highlight important points.

# Here,customer with name ‘Brown’ for product ‘Hard drive’,in San Francisco,has the highest measure values.

# 

# Conclusion (Students should write in their own words):

# In this experiment we explored our dataset and tried representing it using different charts/visualizations. We also understood the data through the graphical representations.

**Date: \_\_\_\_\_\_\_\_\_\_\_\_\_ Signature of faculty in-charge**

# Post Lab Question:

# What are different data types supported in Tableau software?

# Tableau can import:

# • Microsoft Excel

# • Text file

# • JSON file

# • PDF file

# • Microsoft Access

# • Spatial file

# • Statistical file

# • Import through Different database

# Here are some common data types supported in Tableau:

# String: This data type is used for text and alphanumeric characters.

# Integer: Integer data types are used for whole numbers without decimal places.

# Float: Float data types are used for numbers with decimal places, also known as floating-point numbers.

# Date: Date data types are used to represent dates without a specific time component.

# Datetime: Datetime data types are used to represent both date and time information.

# Boolean: Boolean data types are used to represent true or false values.

# Geographic: These data types are used for geographical coordinates and spatial data.

# Currency: Currency data types are used to represent monetary values.

# Percentage: Percentage data types are used to represent values as percentages.

# Duration: Duration data types are used to represent time intervals or durations.

# URL: URL data types are used to store web addresses or hyperlinks.

# Image: Image data types allow you to store and display images in your visualizations.

# Custom: Tableau also allows you to define custom data types using calculations or data transformations.

# These data types are essential for correctly interpreting and formatting your data when creating visualizations and calculations in Tableau. Depending on the data source and the specific needs of your analysis, you may need to work with different data types to ensure accurate results and meaningful visualizations.

# Which chart is most appropriate for visualizing your preprocessed dataset? Justify.

# Bar chart and histogram/scatter plot seemed to be most appropriate(along with pie chart) for visualizing the processed data as all the labels and and data were clearly mentioned in the graph. We could get a clear understanding of the different sales categories and their quantities,distribution using these charts.

# In case of packed bubble chart,few of the labels couldn’t be displayed and the space in chart could not be efficiently utilized,thus making visualization difficult.