

**Batch: H2-1**

**Roll No.: 16010122151**

**Experiment 04**

**Grade: AA / AB / BB / BC / CC / CD / DD**

**Title:** Tableau- Introduction to charts

---

**Objective:** To learn how to use various charts for visualizing preprocessed data in Tableau

---

**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.

**Books/ Journals/ Websites referred:**

1. Data Visualization made simple New York: Routledge - Kristen Sosulski, First edition, 2019
2. Sosulski, K. Data Visualization Made Simple: Insights into Becoming Visual, First edition, 2018
3. <https://blog.hubspot.com/marketing/types-of-graphs-for-data-visualization>
4. <https://www.tableau.com/learn/whitepapers/which-chart-or-graph-is-right-foryou>
5. <https://www.kaggle.com/uciml/adult-census-income>
6. <https://archive.ics.uci.edu/ml/datasets/adult>
7. [https://ori.hhs.gov/education/products/n\\_illinois\\_u/datamanagement/dctopic.html](https://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dctopic.html)
8. A review of research process, data collection and analysis - Surya Raj Niraula

**Resources used:**

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

## Theory (About various charts explored):

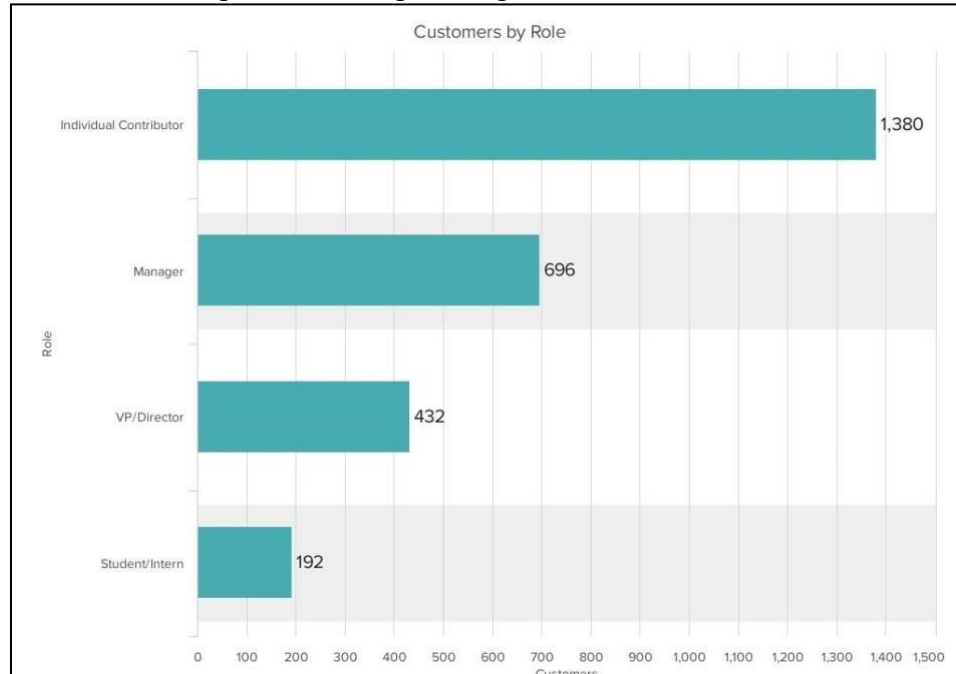
Data visualization is the process of turning your data into graphical representations that communicate logical relationships and lead to more informed decision-making. In short, data visualization is the representation of data in a graphical or pictorial format.

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:

### 1. 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.

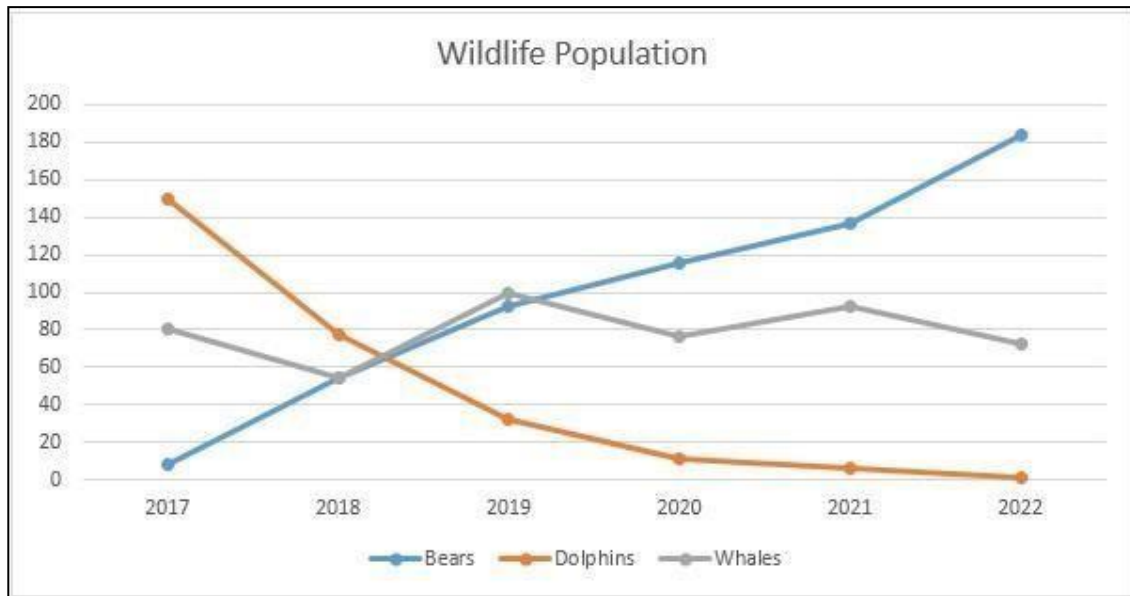


### 2. Line Chart

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 over time (like stock price changes over five years or website page views for the month). The

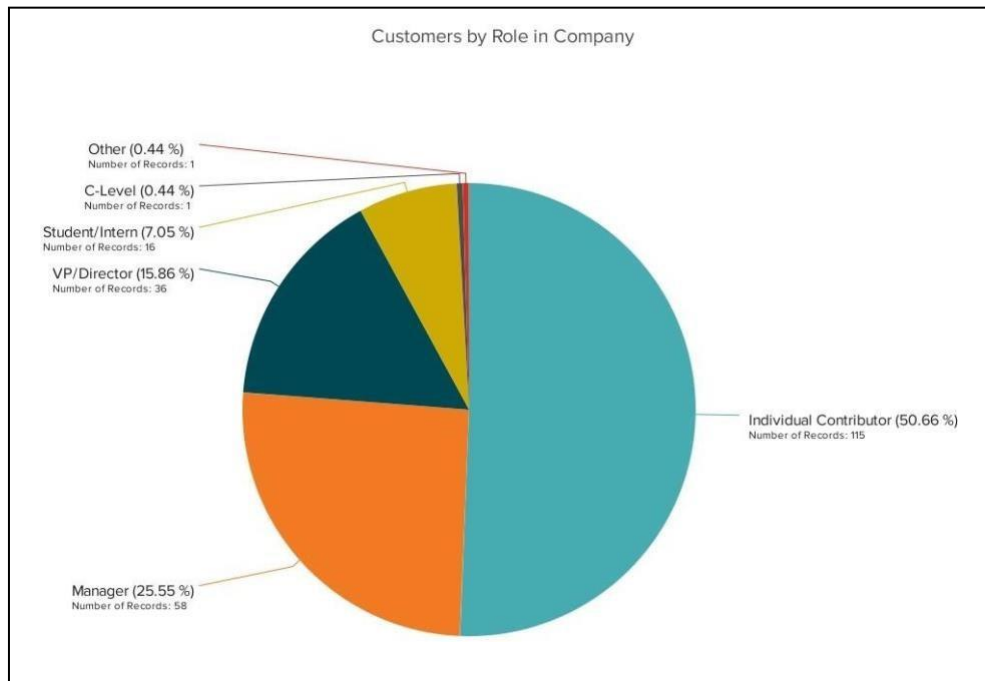
---

result is a simple, straightforward way to visualize changes in one value relative to another.



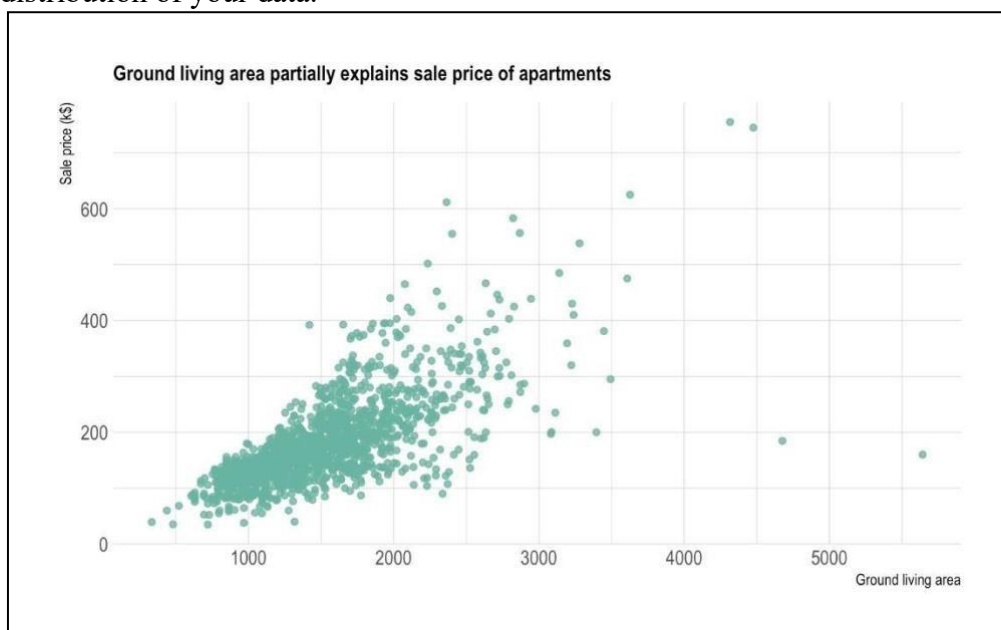
### 3. Pie Chart

Pie charts are powerful for adding detail to other visualizations. Alone, a pie chart doesn't give the viewer a way to quickly and accurately compare information. Since the viewer has to create context on their own, key points from your data are missed. They are used to drill down on other visualizations. A pie chart shows a static number and how categories represent part of a whole -- the composition of something. A pie chart represents numbers in percentages, and the total sum of all segments needs to equal 100%.



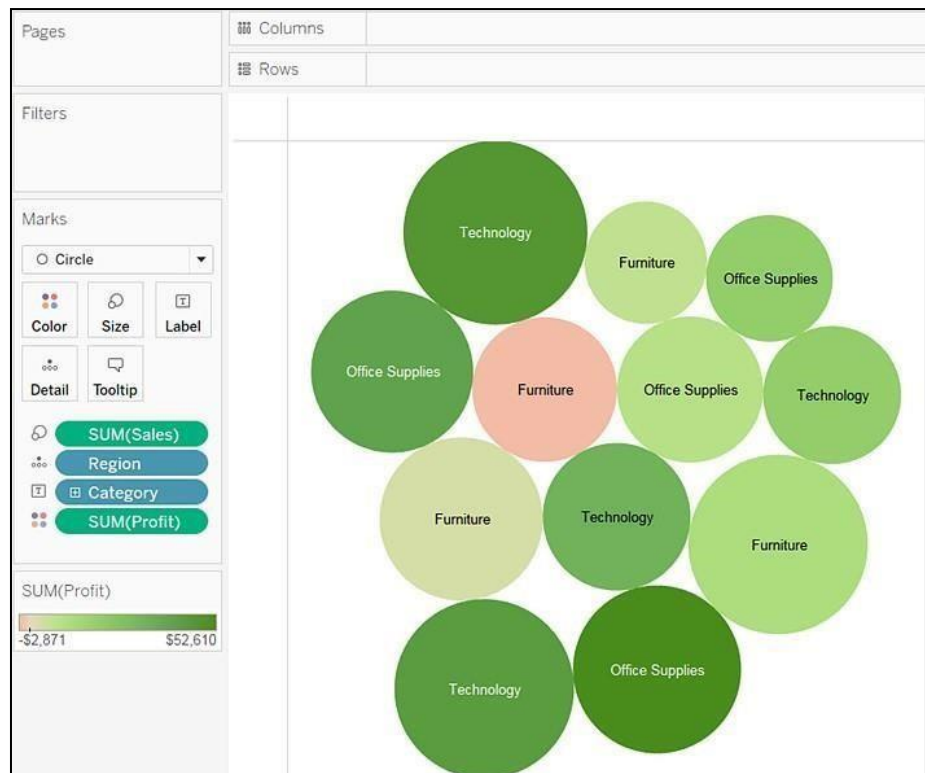
#### 4. Scatter Plot

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. The chart can then be enhanced with analytics like cluster analysis or trend lines. 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 data set. This is useful understanding the distribution of your data.



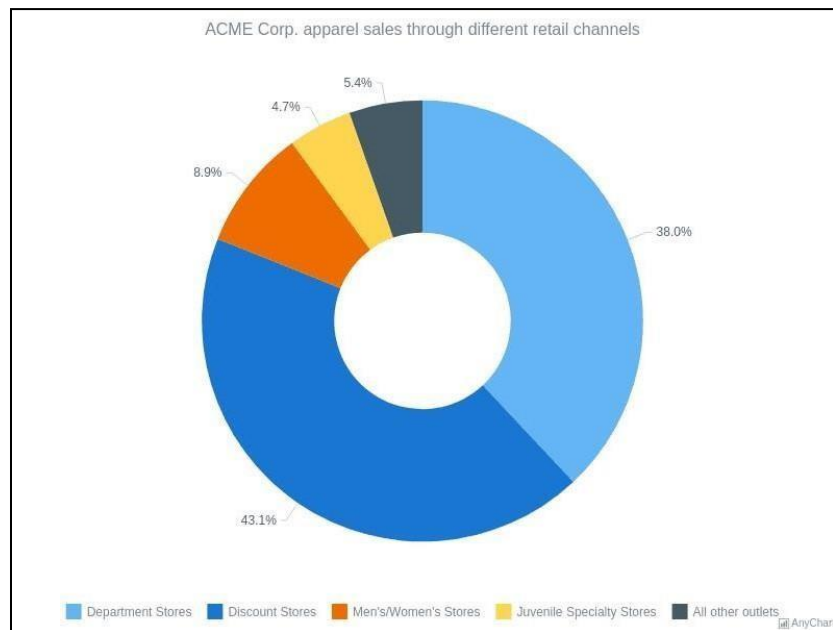
## 5. Packed Bubble Chart

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 data sets as X-values, Y-values and now, Z-values (bubble size).



## 6. Donut Chart

The donut chart is a variant of the pie chart, with a hole in its center, and it displays categories as arcs rather than slices. Both make part-to-whole relationships easy to grasp at a glance. They differ from line charts, area charts, column charts, and bar graphs in that they can't show changes over time. Also, Donut Charts are more space-efficient than Pie Charts because the blank space inside a Donut Chart can be used to display information inside it.



### Following points should be written by students

Different charts explored for data visualization:

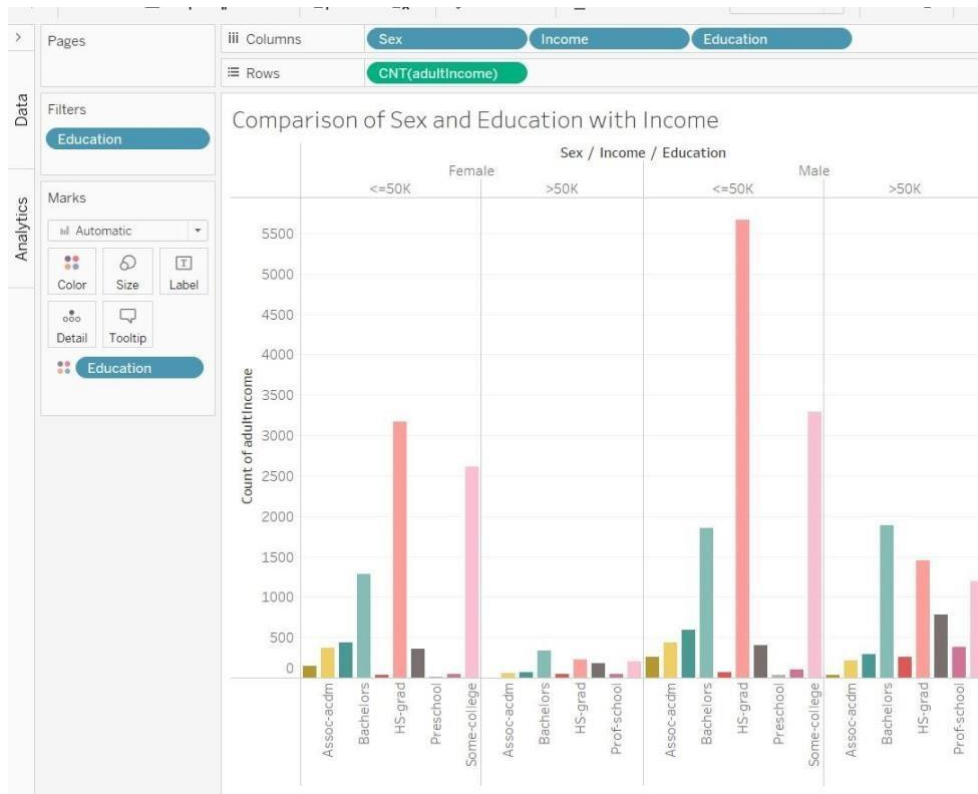
1. Bar Chart
2. Line Chart
3. Pie Chart
4. Scatter Plot
5. Packed Bubble Chart
6. Donut Chart

Platform used by the student: Tableau

Working: Using tableau software

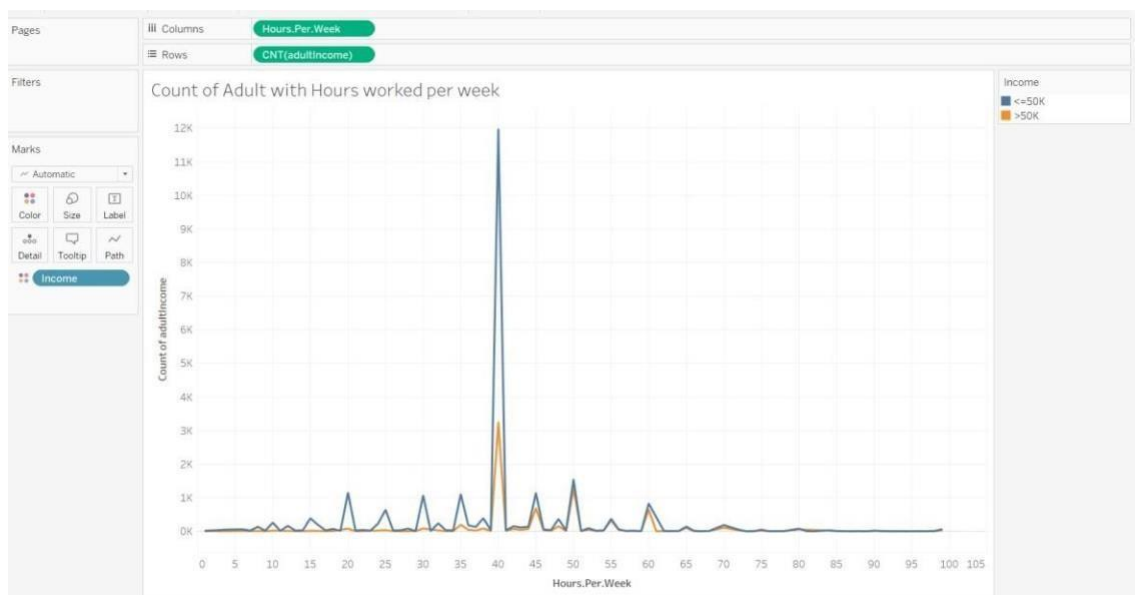
1. Screenshots of each chart output with variations explored on your dataset
2. Interpretation (what each chart signifies)

- Bar Chart



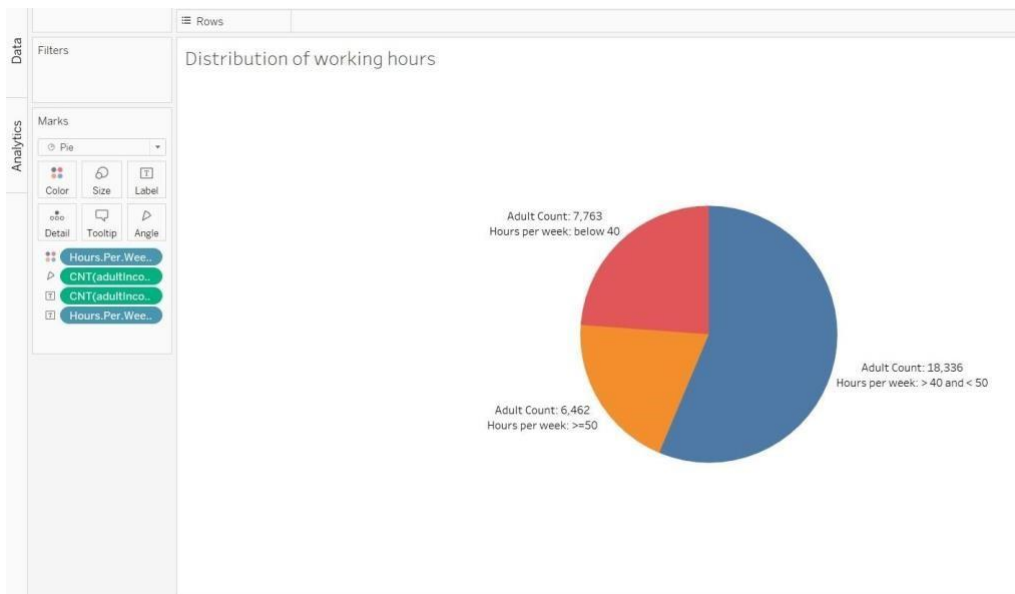
This bar chart represents the comparison of Sex and Education with income of people. We can clearly see columns showing distinction between the income. Max people seem to be HS Grad and earn below 50K. There is a small population earning above 50K. This shows how greater qualifications fetch more income.

- Line Chart



The line chart shows the hours worked but different number of people. Maximum people work for 40 hours in a week on an average. This is independent of the income earned as 30-30 hours seem to be the common category for work.

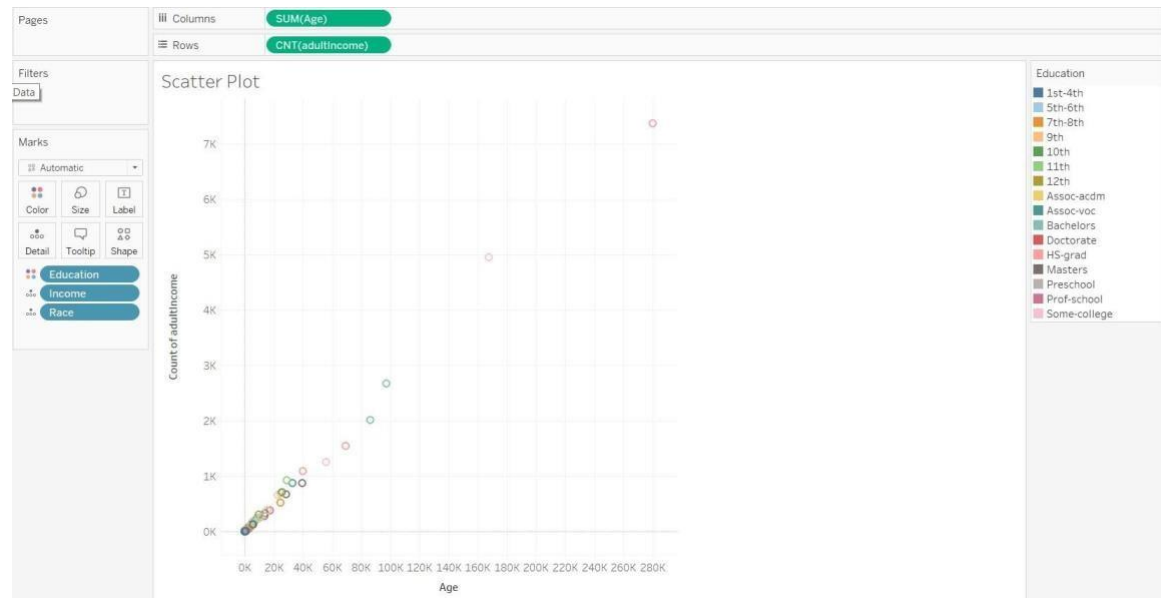
- Pie Chart



The pie chart gives the adult count on basis of hours worked per week. Maximum people work for 40-50 hours. The pie chart seems to be better than the line chart as shown above as it gives a clear indication of the people working for different hours and is visually appealing.

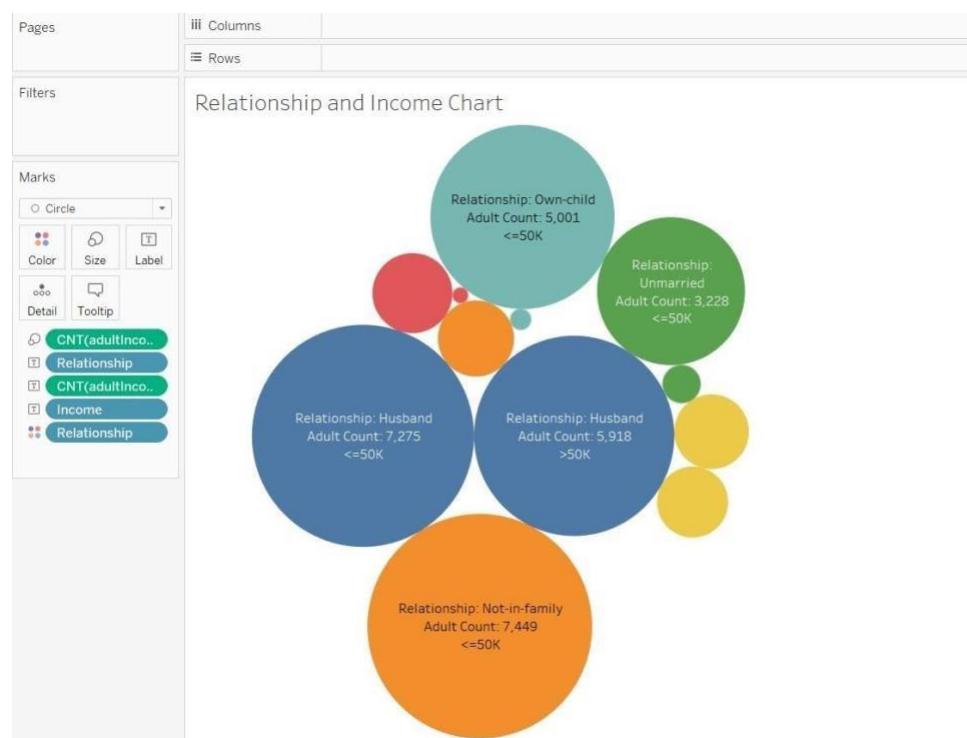
- Scatter Plot





A scatter plot presents lots of distinct data points on a single chart. This scatter plot is representation of Age and Adult count. The chart does not give clarity about the dataset, it shows the relationship between age and adult count.

- Packed Bubble Chart



Packed Bubble chart shows the link between the income, adult count and relationship of people. It shows that husbands and not in family category include max population and max income as well.

### Conclusion:

Through this experiment we explored our dataset and tried to represent it using different charts. We also interpreted the charts and understand the data through the graphical representations.

**Date: 29-08-2023**

**Signature of faculty in-charge**

**Post Lab Question:**

**1. 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

Also these are the different data types that Tableau can support:

- String values
- Number (Integer) values
- Date values
- Date & Time values
- Boolean values
- Geographic values
- Cluster or mixed values

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

Bar Chart and Pie chart seemed to be the best for visualizing the processed data as all the label and data was clearly mentioned in the graph. We could get a clear idea about the income of various categories 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. In case of Scatter plot, the chart looked busy and we couldn't draw any conclusion. The scatter plot didn't give any information as the measure involved sum of ages and adult count which didn't give any sensible analysis.

---

---