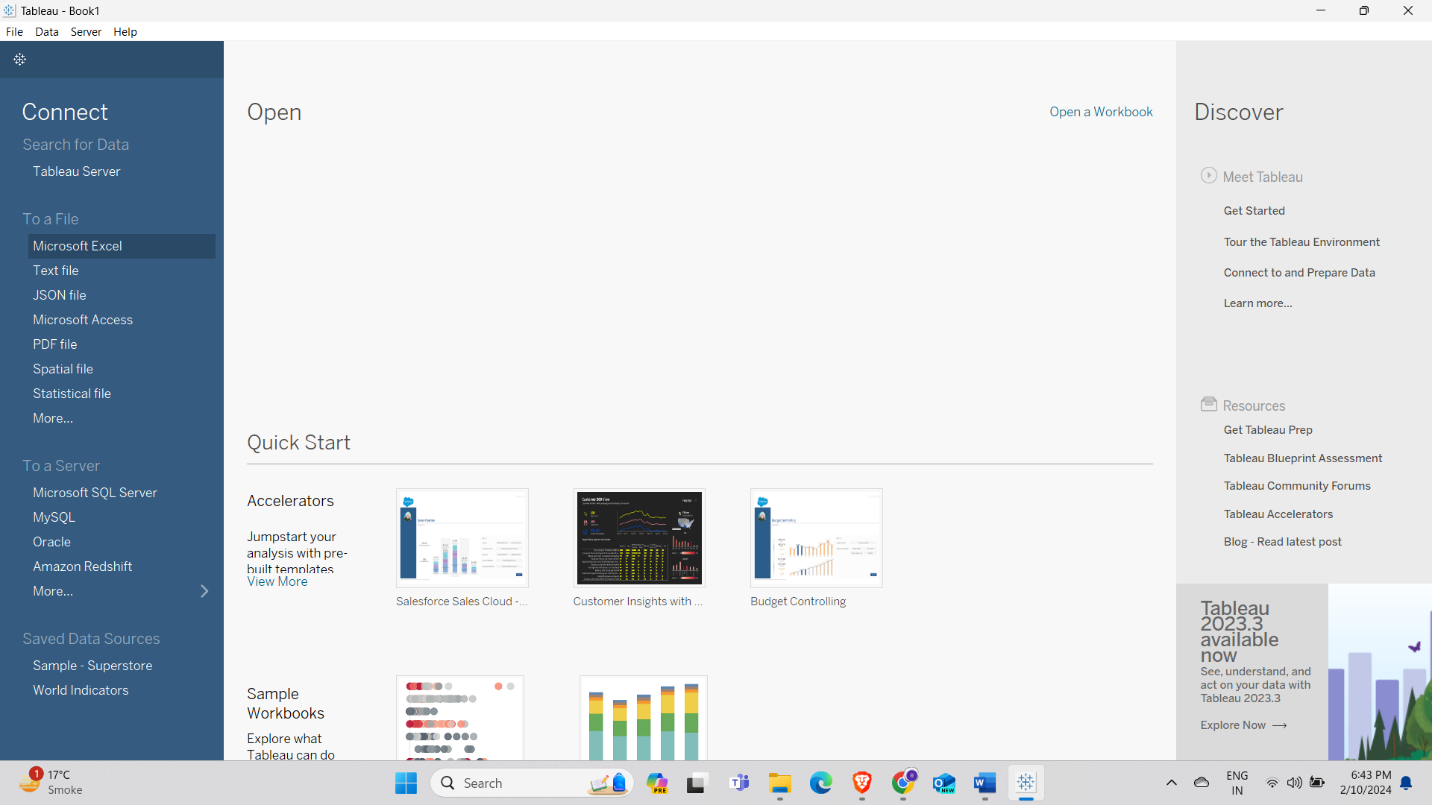
**Lab-1**

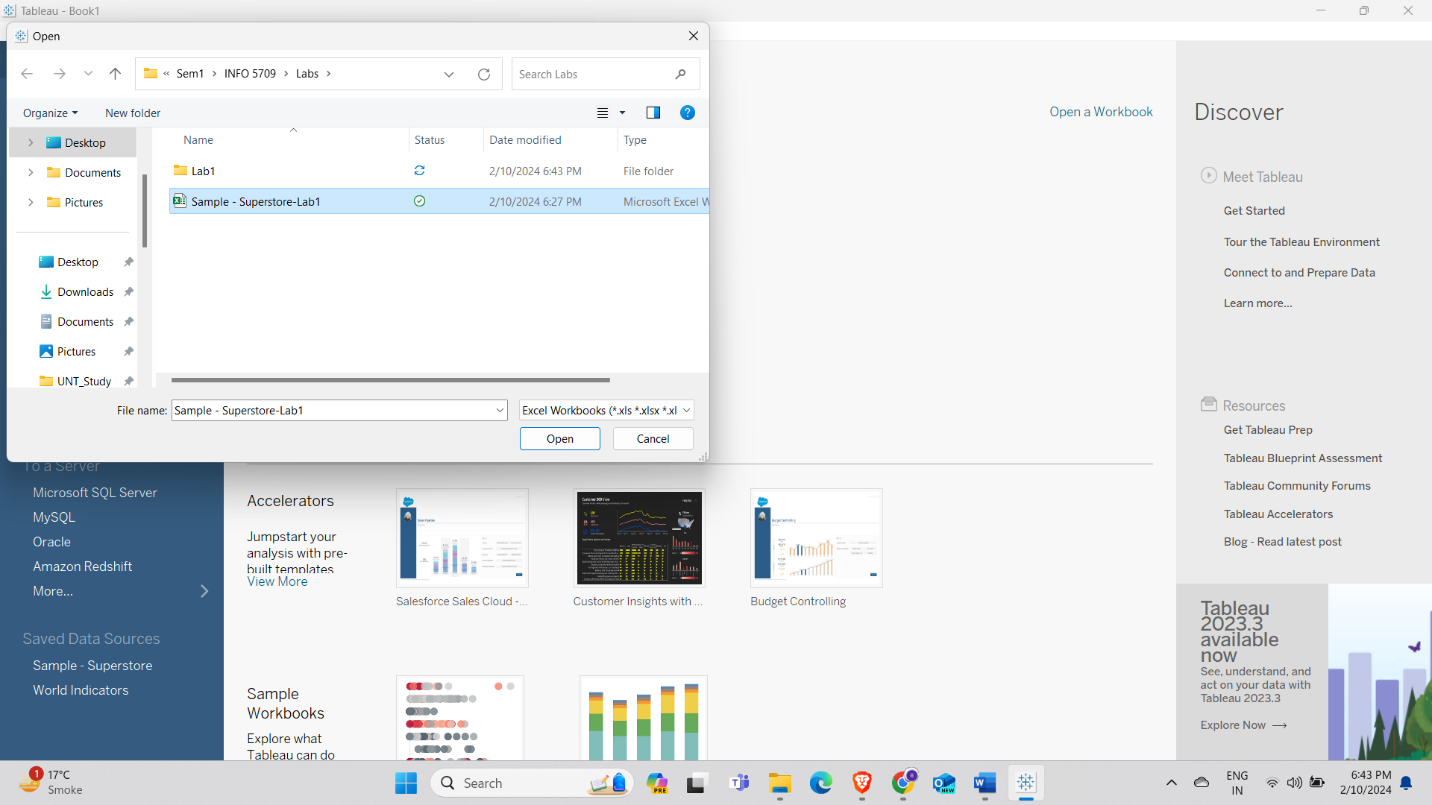
**Q1: Randomly choose one public data set and load it into tableau.**

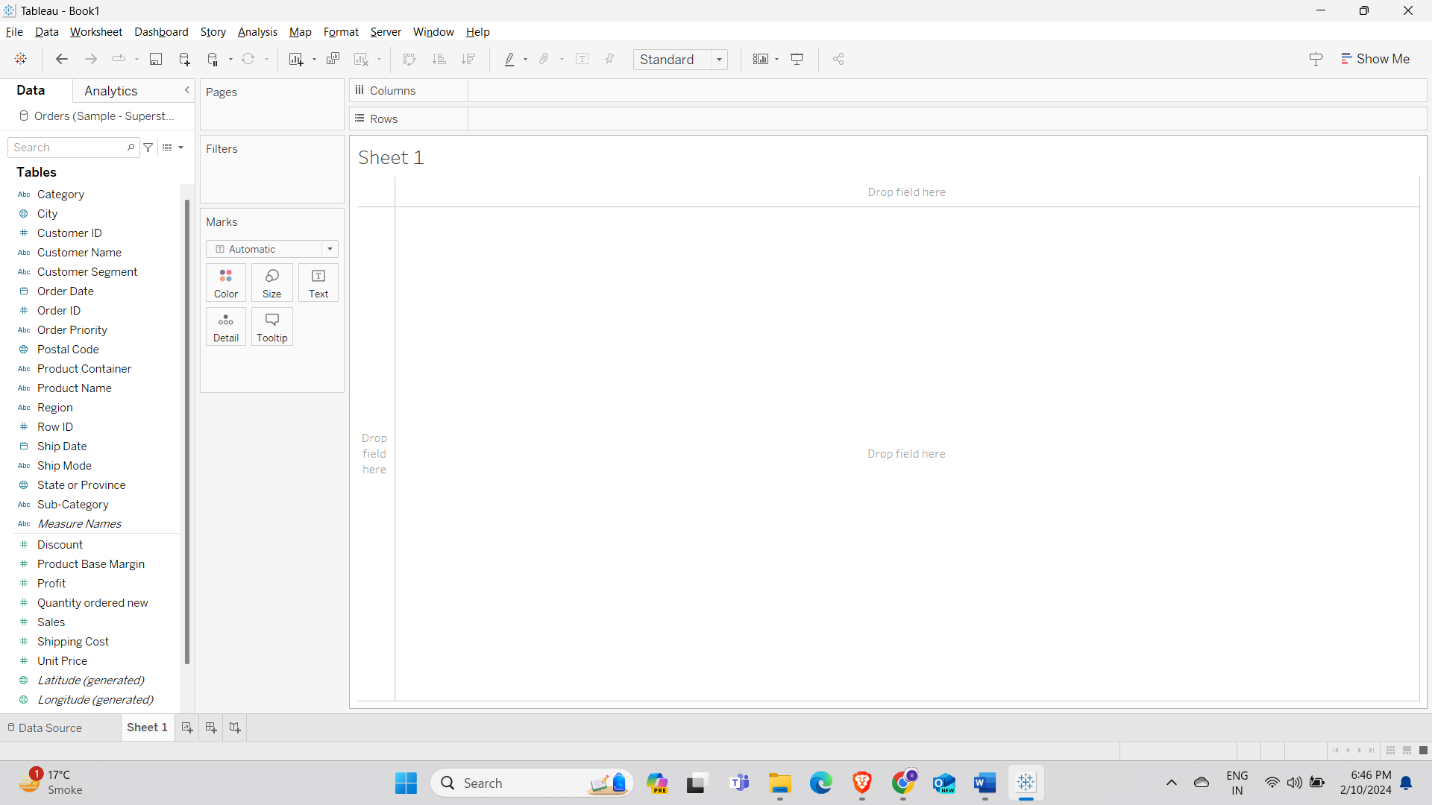
Ans: Here for this question, I have loaded Sample - Superstore-Lab1 dataset, and I will be working on that dataset for this lab.

There are steps to load dataset in tableau.

1. Open tableau.
2. In the “Connect” section, choose “Microsoft Excel”.



1. Choose the dataset you want to work with, for this lab we are choosing Sample - Superstore-Lab1.
2. Click on the sheet you want to work with.



And we can work with this dataset as per our requirements.

**Q2: Explain all attributes of your selected dataset.**

Ans: Attributes of Sample - Superstore-Lab1 are,

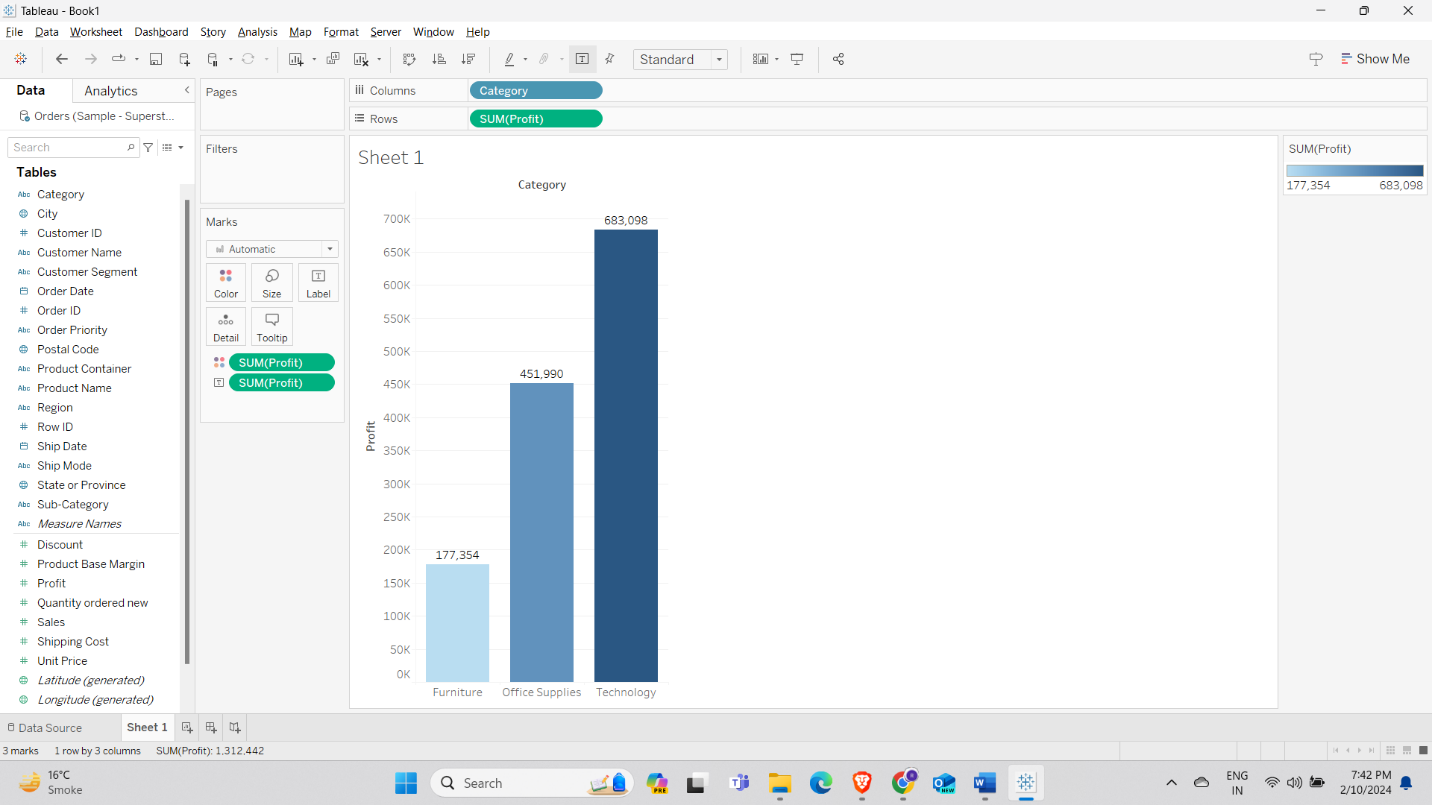
* Row ID: Represents each row with a unique value.
* Order Priority: Represents the priority of the order, using which we can sort out the order in which item should be made.
* Discount: Represent discount on each item.
* Unit price: This attribute gives the price of each item.
* Shipping Cost: Represents the cost of shipping.
* Customer ID: It gives the ID of each customer.
* Customer Name: Represents the name of the customer.
* Shipping Mode: It gives the mode of shipping i.e. Regular Air or Express Air
* Customer Segment: Represents the segment in which a customer falls either Corporate or Home Office.
* Category: It gives category of order.
* Sub-Category: It gives the type of order.
* Product Container: Represents the size of container.
* Product Name: It gives the name of the product.
* Product Base Margin: The cost of product.
* Region: Represents in which region the order needs to be delivered.
* State of Province: Delivery state of the order.
* City: Represents city in which order needs to be delivered.
* Postal Code: Postal code of the area for delivery.
* Order Date: It gives the date of the order being placed.
* Ship Date: Represents date the order is delivered.
* Profit: Profit on each order.
* Quantity Order New: It gives the quantity of order being placed as new.
* Sales: Sales of each item.
* Order ID: Represents order id of each item.

**Q3: Use tableau to generate one “Comparison” visualization graphics.**

Ans:

Comparison Visualization: Comparison visualization graphics are used for to collate the different attributes of a dataset.

Example:



As we can see from the graph which contains two attributes- “Category” and “Profit”- which compares the profit of different items, in the graph we can see that furniture is the one with the lowest profit with profit of 177,354 and hence is of lightest color while the technology category is the highest profit earner with profit of 683,098 and is represented with the darkest color.

**Q4: Use tableau to generate one “Distribution” visualization graphics.**

Ans:

Distribution Visualization: Distribution visualization can be defined as the representation of data distribution.

Example: A screenshot of a computer

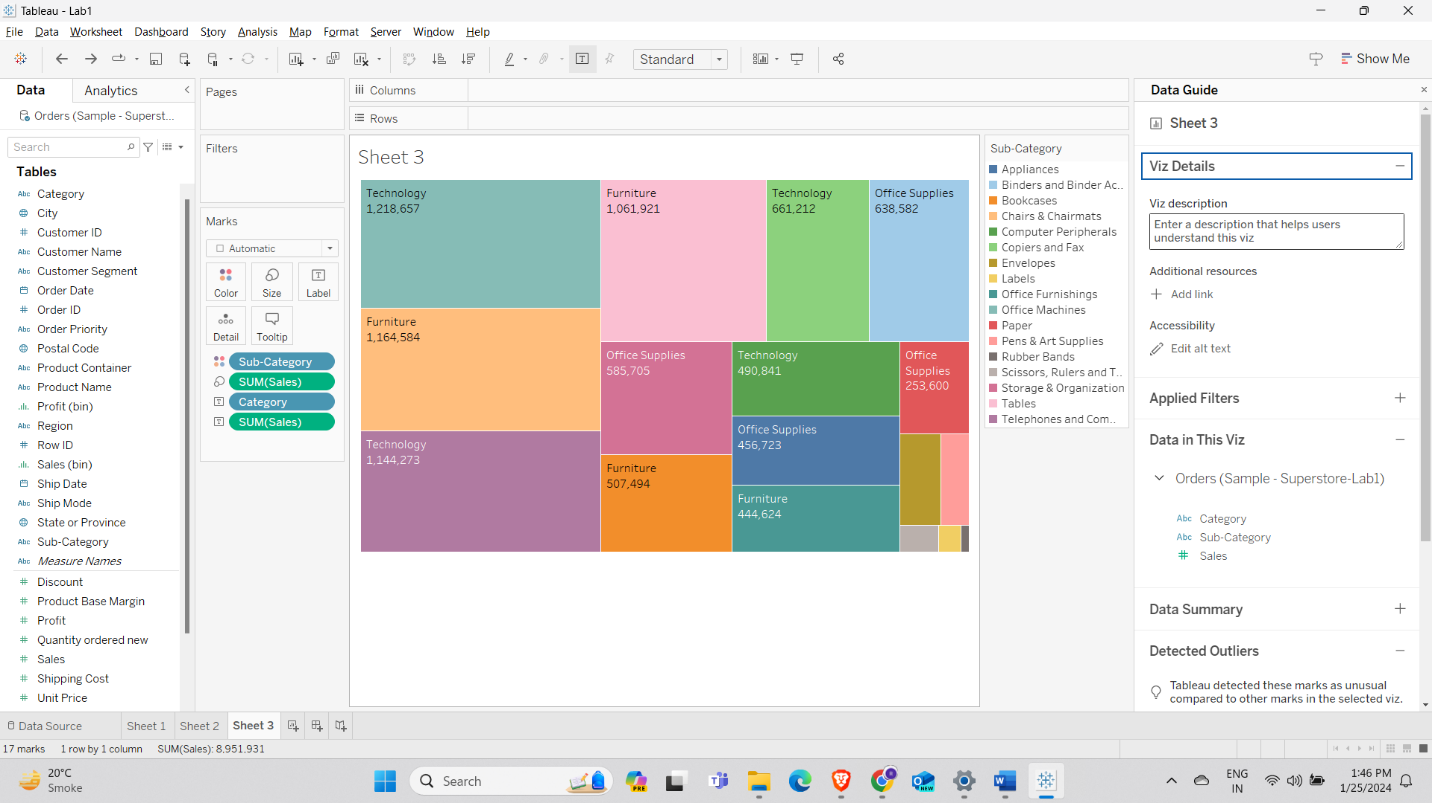
Description automatically generated

In this representation, histogram is used. Histogram displays data by dividing data into bins. Here we have taken “Sales” as entity. For the other entity the bin each of 5K had been made.

**Q5: Use tableau to generate one “Composition” visualization graphics.**

Ans:

Composition Visualization: Composition visualization is used when the data given is in the form of relative proportions such as sales of category.

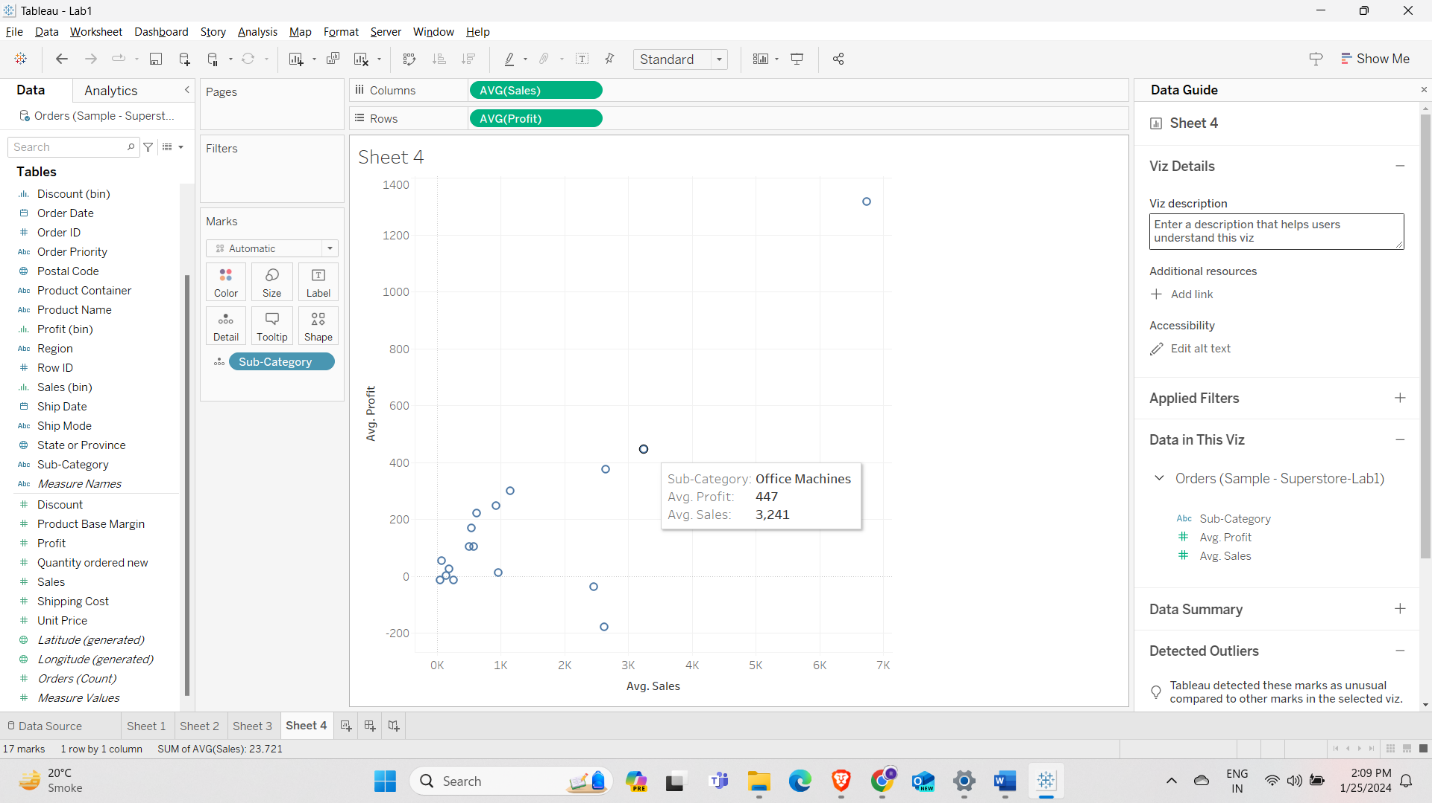
Example:

We have used a tree map for composition visualization, using tree map we have given category, subcategory, and sales attributes. Tree map shows total sales as the size of rectangle and with smaller rectangle inside representing sales of each subcategory.

**Q6: Use tableau to generate one “Relationship” visualization graphics.**

Ans:

Relationship Visualization: It is used to represent relationships between entities of the data set.

Example:

For this example, we are getting a relationship between sales and profit on average. On x-axis we have average sales and on y-axis we have average profit, and this kind of graph is called scattered plot.

**Q7: For Comparison vs. composition explain the rational why you are not selecting alternate graphics.**

Ans:

**Comparison:**

Comparison visualization is used to represent similarities or differences between different entities of data.

We have used comparison for question 3 as we wanted to find how much profit each category makes, using bar chart we can easily plot the comparison between profit and category. We can use this bar chart to find the maximum and minimum profit made by each category. To conclude we have used comparison for this question as we wanted to collate category and profit.

**Composition**:

Composition visualization is understanding of the parts that makes the whole and their relative contribution. In conclusion composition visualization is used to show how different components make total.

In question 5, to represent complete sales of each sub-category we have used the tree-map, tree-map will represent the whole sales as a rectangle and each smaller rectangle inside of that rectangle will be appear for sales of each sub-category. To conclude, we have used composition for this question because we wanted to represent a complete data for sales

**Q8: For Distribution vs. Relationship explain the rational why you are not selecting alternate graphics.**

**Distribution:**

Distribution visualization is used to represent how data points are distributed across different values.

Histogram is an ideal way to display the frequency distribution of continues data and dividing them into bins.

In question 4, as we are looking for the sales number, we are using histogram to represent that data, now histogram will take “Sales” as an entity and we need to make another entity called bin, using these, the peak and drop sales can be easily found.

**Relationship:**

Relationship visualization is used to represent the dependencies of two variables in dataset, that is, how changing one can affect the other variable.

In question 6, the use of scatter plot is a great way to show the dependence of two variables, profit and sales. Using the graph we can easily identify the patterns and correlations. Any other visualization will not give us the relationship between these two entities.