

# Lab 4

## Part-1:

### 1. Scatterplot

Goal: To help visualize relationships between quantitative (numerical) variables.

In Tableau, you create a scatterplot by placing at least one measure on the **Columns** shelf and at least one measure on the **Rows** shelf.

A scatterplot can use several **mark types**. You should make an informed choice about the marks based on discussions during the lectures. For more information →

[https://help.tableau.com/current/pro/desktop/en-us/buildexamples\\_scatter.htm](https://help.tableau.com/current/pro/desktop/en-us/buildexamples_scatter.htm)

Step 1: Take the **Superstore** dataset:

Step 2: Drag the **Sales** measure to **Columns**.

Step 3: Drag the **Profit** measure to **Rows**.

Tableau will aggregate the measure as a **sum**, creating a horizontal and vertical axis. **measures** contain continuous numerical data.

Now you have a one-mark scatterplot -- A single mark showing the sum for all values for the two measures. Not a very interesting plot, right? Figure out why this happened.

Step 4: What if we disaggregate the data?

To do this, go to **Analysis >Aggregate Measures** and de-select.

Step 5: Separate the data according to whether it was in the **Consumer**, **Corporate**, or **Home Office** segments. Do that by dragging the **Segment** dimension onto one of the visual dimensions in the Marks card → use color.

Step 6: Use a different measure - **Product Category**.

Add the screenshot of the plots and share your understanding.

## Part 2:

### Goal: Exploring correlation on real data

Correlation analysis in Tableau compares two or more quantitative variables to see if values in one vary systematically with values in another. For example, as height in men increases, so typically does weight.

1. <https://www.kaggle.com/code/tombutton/world-development-indicators-v2>
2. Load the dataset into Tableau and create scatterplot/s that highlights something interesting in your dataset.
3. Explore the correlations between various measures. You can create a correlation matrix to have a better understanding
4. In your report, briefly describe what you found supported by different visualizations while following the principles learned during the lectures. Do not add too many unnecessary plots in the report.

#### GDP per Capita vs Life Expectancy

Is there a positive relationship—that is, do countries with higher GDP per capita also tend to have longer life expectancy?

#### Population vs CO<sub>2</sub> Emissions

Is CO<sub>2</sub> emissions roughly proportional to population size?

Are there countries that significantly over-emit relative to their population?

#### Literacy Rate vs Internet Users

Do higher literacy rates correspond with greater internet usage per capita?

#### Urban Population vs GDP per Capita

Is there a positive correlation between urbanization and GDP per capita?

#### Correlation Matrix

Build a matrix using *GDP per Capita*, *Population*, *CO<sub>2</sub> Emissions*, *Life Expectancy*, *Literacy*

*Rate, and Internet Users.*

Identify which variable pairs show the strongest positive or weakest (possibly negative) correlation.