Scatter Plot & Exploring Correlations in World Development Indicators

Ritwik 202411067 , Shubham 202411066 , Jaydeep 202411029 , Khelan 202411025

1 Part-1 : Scatter Plot

1.1 Why the first plot had one mark

Tableau aggregates measures by default, so with SUM(Sales) on Columns and SUM(Profit) on Rows and no dimensions on the view, the worksheet collapses all rows into a single total, producing one mark.
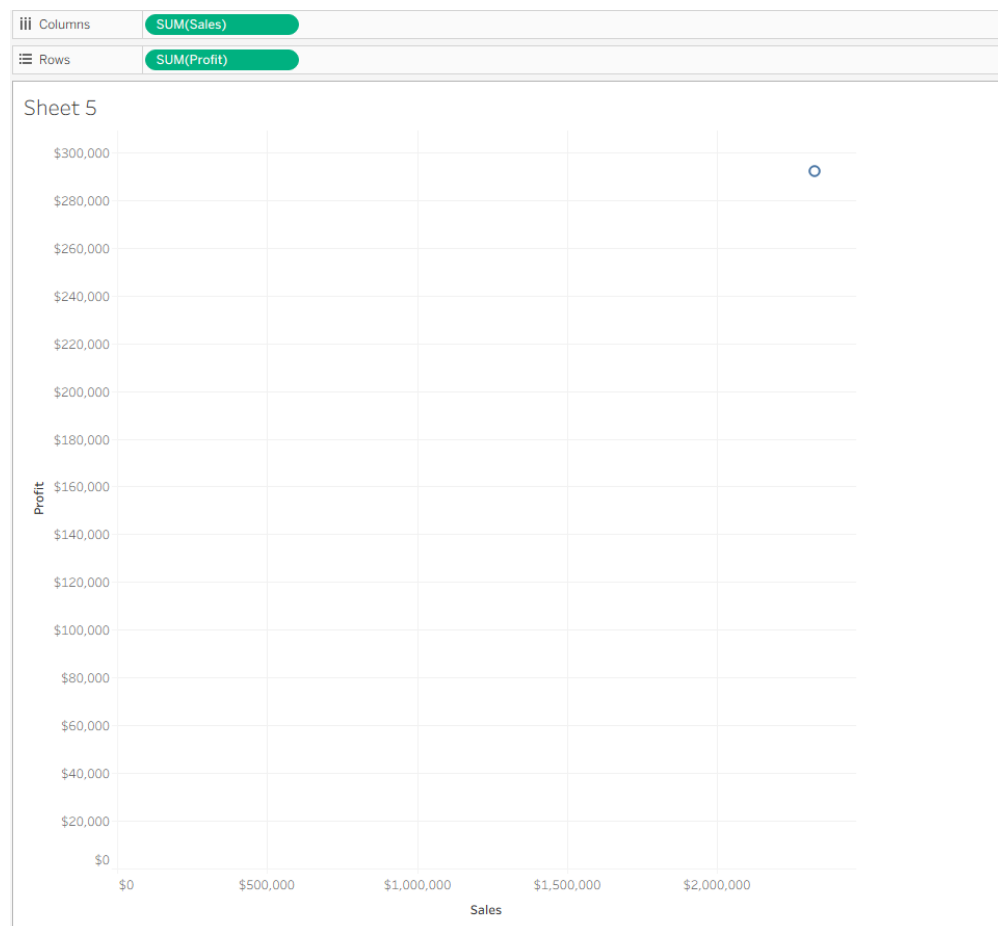

Figure 1: The single-mark scatterplot with SUM(Sales) on Columns and SUM(Profit) on RowsS

1.2　What changed after disaggregation

Turning off Analysis → Aggregate Measures displays record level marks, revealing the spread of individual transactions. This exposes outliers (very high Sales with negative Profit), heteroscedasticity (wider spread at higher Sales), and any non linear patterns.
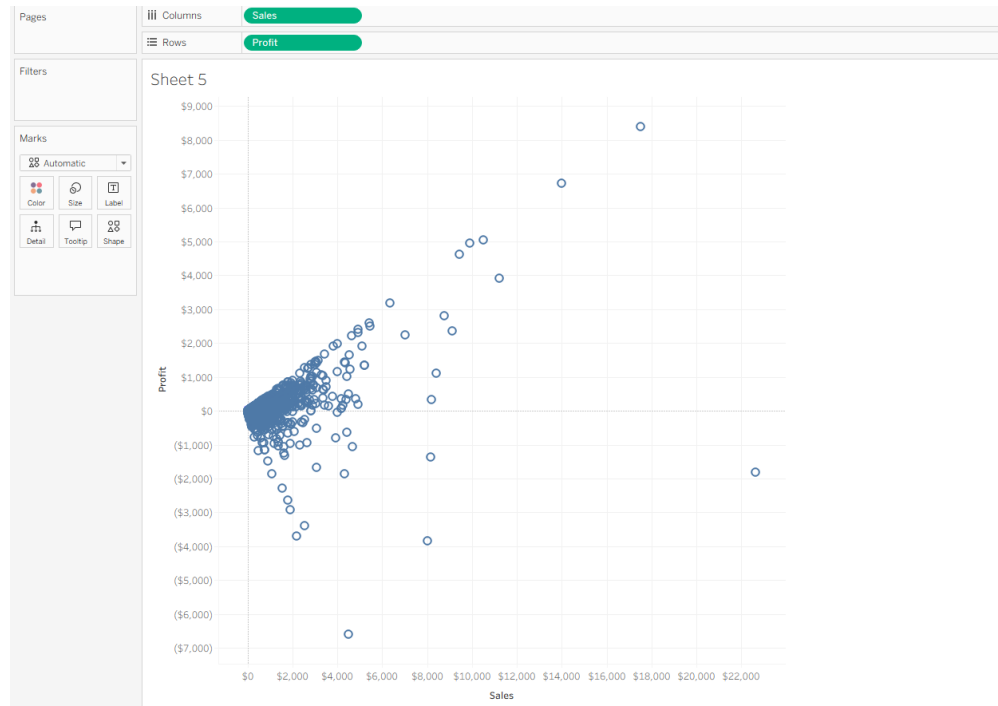

Figure 2: The single-mark scatterplot with SUM(Sales) on Columns and SUM(Profit) on RowsS

1.3　What coloring by Segment revealed

the Sales Profit space. Look for clusters, density differences, and whether any segment contributes disproportionately to negative profit points. If you enabled per segment trend lines, compare slopes and fit.
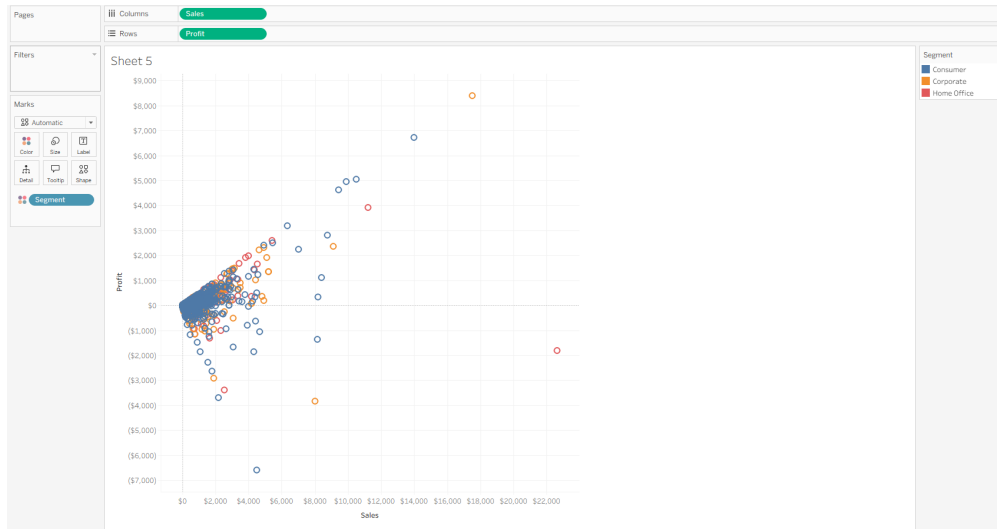
Figure 3: Disaggregated scatter colored by Segment.

1.4 What changing to Category revealed

Using Category (dimension) instead of Segment helps compare Furniture vs. Office Supplies vs. Technology. You may observe, for example, that some categories have more negative-profit transactions or different Sales ranges. Faceting into small multiples makes these contrasts clearer by removing overplotting across categories.
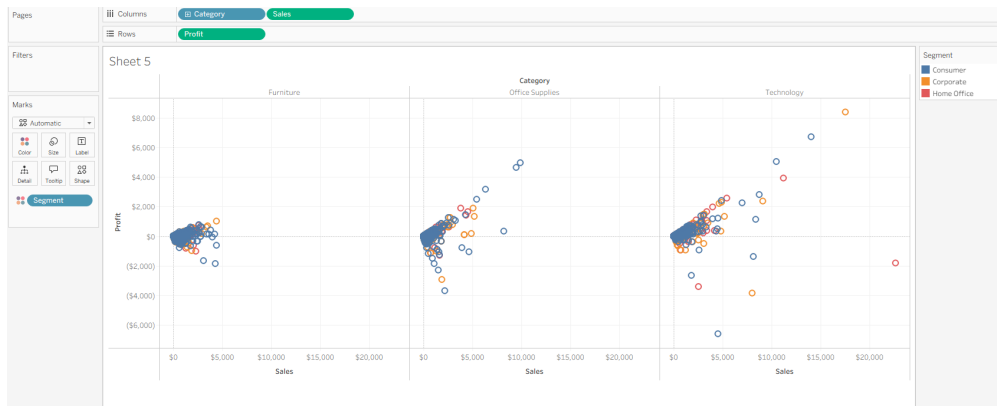

Figure 4: Disaggregated scatter colored by Category

2 Part-2 :Exploring Correlations in World Development Indicators

The goal of this analysis is to explore relationships between development indicators using the World Development Indicators dataset. We focus on GDP per Capita, Life Expectancy, Population, CO_2 Emissions, Literacy Rate, and Internet Users. Scatterplots and a correlation matrix (built in Tableau) are used to identify systematic relationships and outliers.

2.1 GDP per Capita vs Life Expectancy

A scatterplot comparing GDP per Capita with Average Life Expectancy (Figure 5) shows a clear positive relationship: countries with higher GDP per Capita tend to have higher life expectancy. The relationship is not perfectly linear, with diminishing returns at higher income levels.
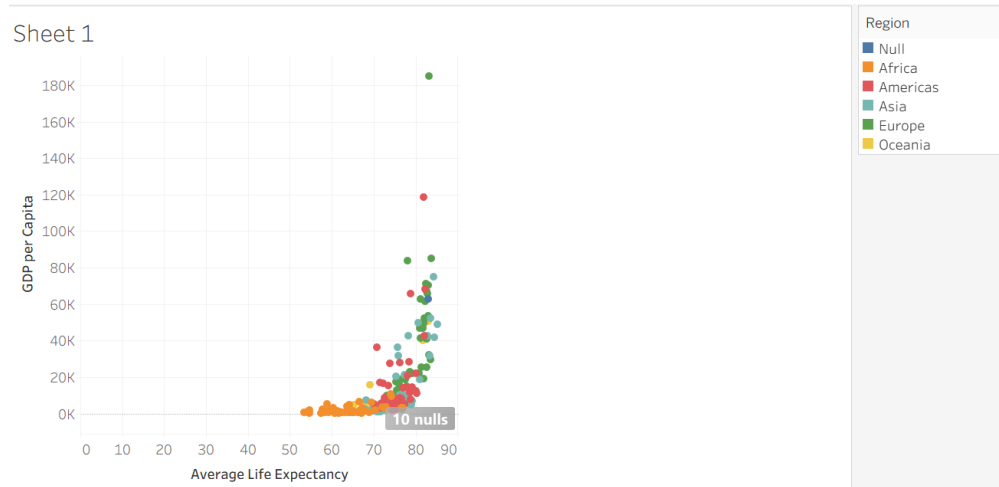

Figure 5: GDP per Capita vs Average Life Expectancy

2.2 Population vs CO$_2$ Emissions

Figure 6 plots Population against total Greenhouse Gas Emissions (proxy for CO$_2$). The relationship is roughly proportional: larger populations generally correspond to higher total emissions. China and the United States stand out as high emitters; India emits less relative to its population, whereas the U.S. emits more than expected.
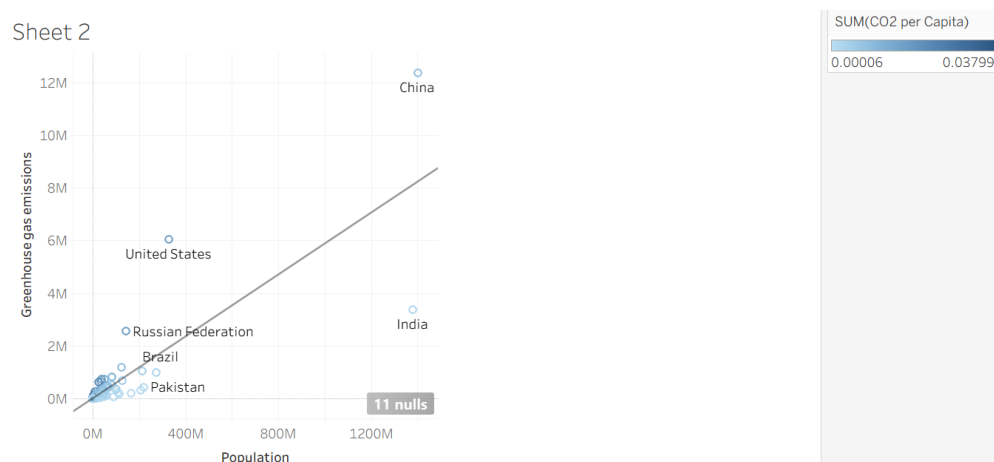

Figure 6: Population vs Greenhouse Gas Emissions

2.3 Literacy Rate vs Internet Users

Figure 7 shows a strong positive relationship between Literacy Rate and Internet Use: countries with higher literacy typically have greater internet penetration, consistent with literacy enabling digital adoption.
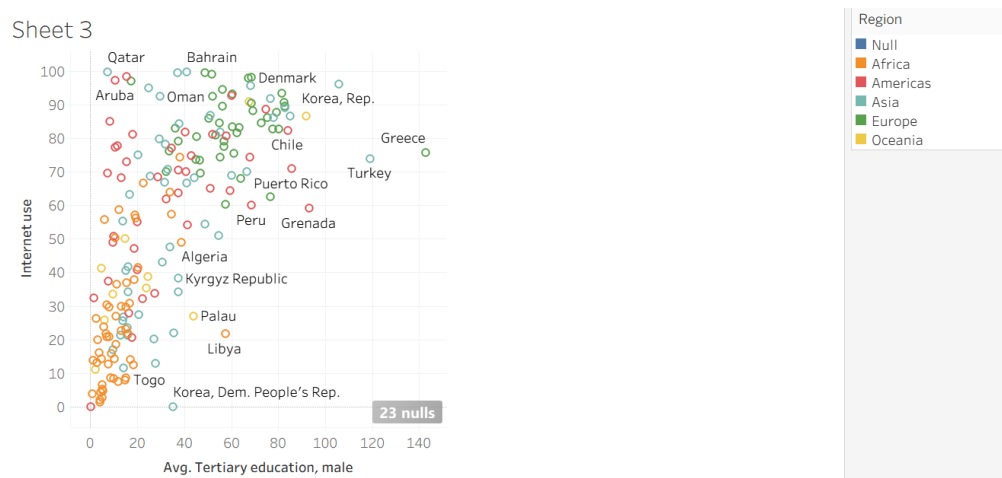

Figure 7: Literacy Rate vs Internet Use

2.4 Urban Population vs GDP per Capita

Urbanization is positively associated with GDP per Capita (Figure 8). Countries with a larger urban share often exhibit higher income levels, consistent with urban agglomeration effects. Small, high-income economies (e.g., Liechtenstein, Bermuda) appear as outliers.

Figure 8: Urban Population vs GDP per Capita

2.5 Correlation Matrix

A correlation matrix was constructed for GDP per Capita, Population, CO_2 Emissions, Life Expectancy, Literacy Rate, and Internet Users (Figure 9). Key findings:

- **Strongest positive correlations:** Internet Use with Life Expectancy; Literacy Rate with Life Expectancy; Internet Use with Literacy Rate; GDP per Capita with Life Expectancy/Internet Use.

- **Population & CO_2:** strong positive (larger countries emit more in total).

- **Weak/negative:** Population vs GDP per Capita is weakly negative (large-population countries often have lower per-capita income).
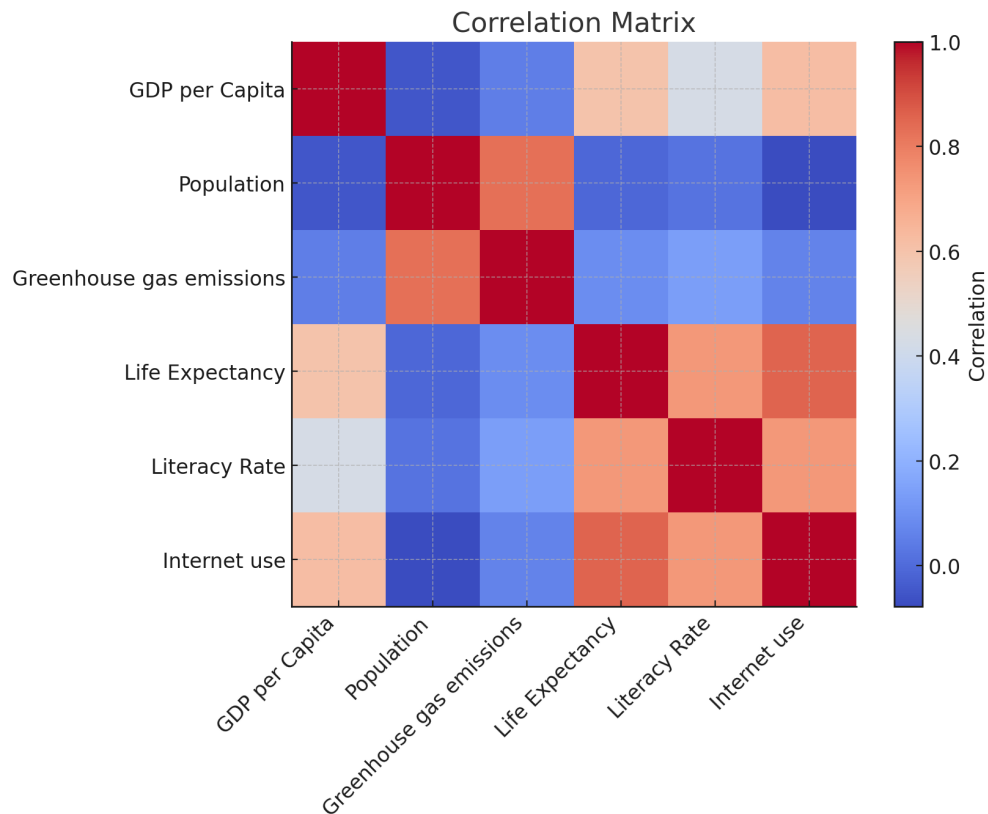

Figure 9: Correlation matrix across the six indicators

3 Conclusion

The analysis highlights: (i) economic prosperity is closely linked with life expectancy; (ii) literacy strongly supports digital adoption; (iii) urbanization aligns with higher income levels; and (iv) population size is a primary driver of total emissions, with notable policy/efficiency outliers. These relationships can guide policy aimed at sustainable development.