Lab 3 (Aug 16, 2024)

Deadline – Aug 26, 11:59 PM

The assignment has two parts. Only 2nd part will be evaluated.

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/viewparts_marks_marktypes.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.

2. Trend Lines

A trend line provides a statistical definition of the relationship between two numerical values. To add trend lines, both axes must contain a field that can be interpreted as a number—similar to a scatterplot

To add a trend line, switch to the Analytics tab.

Under the Model heading, grab the **Trend Line** label and drop it onto the type of model you would like Tableau to fit to your data. First, use a simple **Linear** model.

This will produce three trend lines, one for each of the **Product Category**.

Each of the lines has two fainter lines surrounding it. These are called **confidence bands**, and they indicate how sure you are about the position of the line. You can turn them off to declutter the visualization. To remove the confidence bands: right-click on any of the trend lines and choose **Edit All Trend Lines**.

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. Start by downloading one of the sample data files available at https://public.tableau.com/app/resources/sample-data

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.