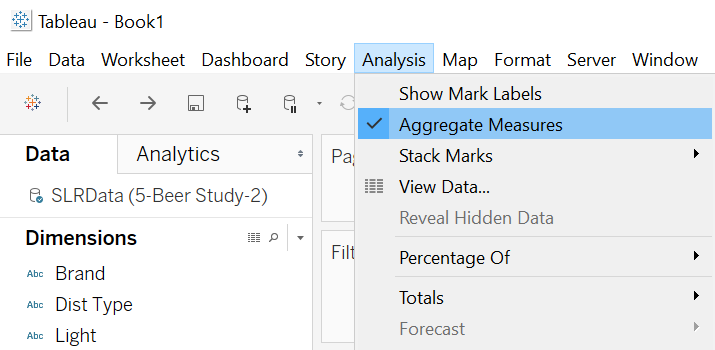
# Advanced Data Visualization – Lab #4

|  |  |
| --- | --- |
| Provided file(s): | * Lab04.docx * SLRData (5-Beer Study-2).csv |
| Submission file(s): | * Lab04.docx * Lab04\_1.twb * Lab04\_2.twb |

Work in your groups. Only one member of your group should submit. Only the last submission submitted before the deadline will be marked.

## **Part I: Scatterplot**

1. Open Tableau Desktop.
   1. Choose Open > File and select “SLRData (5-Beer Study-2).csv”.
   2. Drag “Calories” to columns and “Pct Alcohol” to rows shelves. You should get just one point.
   3. From the Analysis menu, deselect “Aggregate Measures” (as in the image). You should see a scatterplot now.

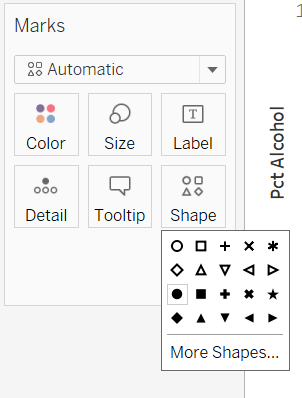


* 1. Change the title to

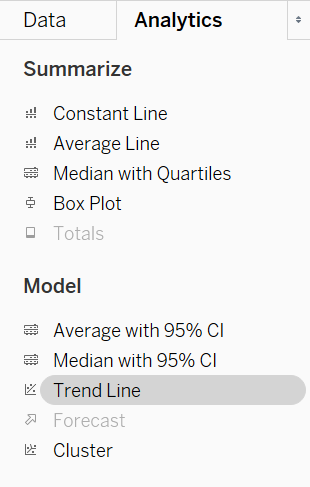
**Group #:**

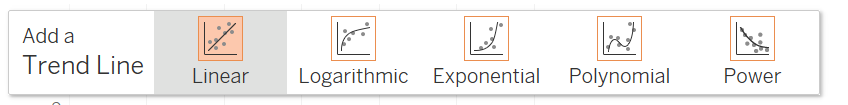
**Do Calories Affect Alcohol Content in Beer?**

This simple scatterplot shows the linear relationship between calories and percent of alcohol.



* 1. Use the Shape card (from Marks card) to change the shape of data points to filled circles.

* 1. From the Analytics tab, drag “Trend Line” onto the worksheet. You should see the different trend lines. Drop onto “Linear”.



The resulting trend line is a best fit linear model, using the method of simple linear regression.

* 1. Right click the trend line and choose “Describe Trend Line”. Copy the information displayed and then paste it below:

**P-value: < 0.0001**

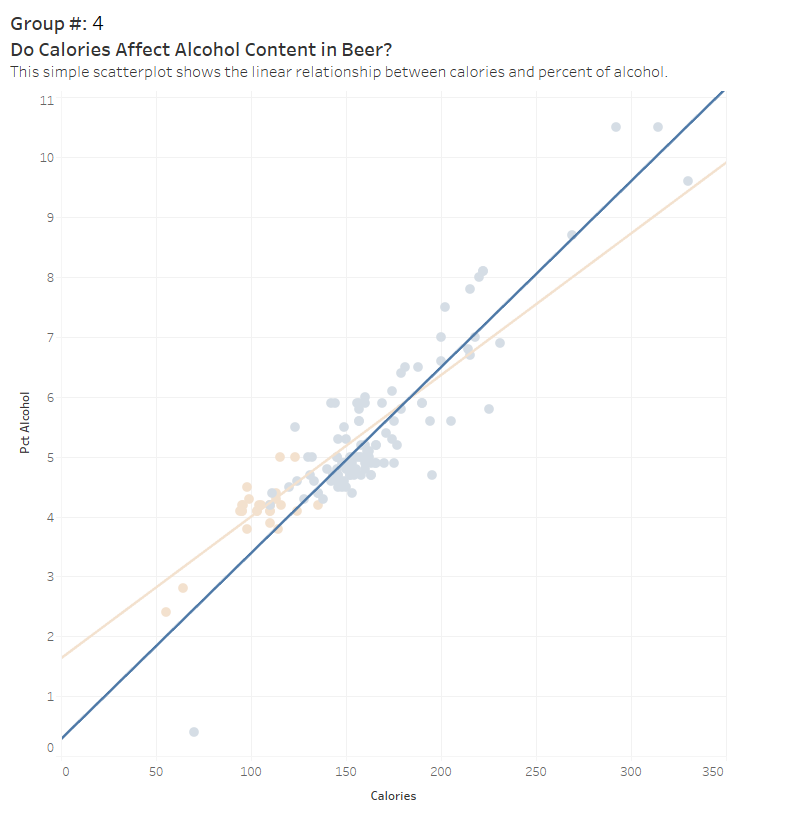
**Equation: Pct Alcohol = 0.0275028\*Calories + 0.952598**

**Coefficients**

**Term Value StdErr t-value p-value**

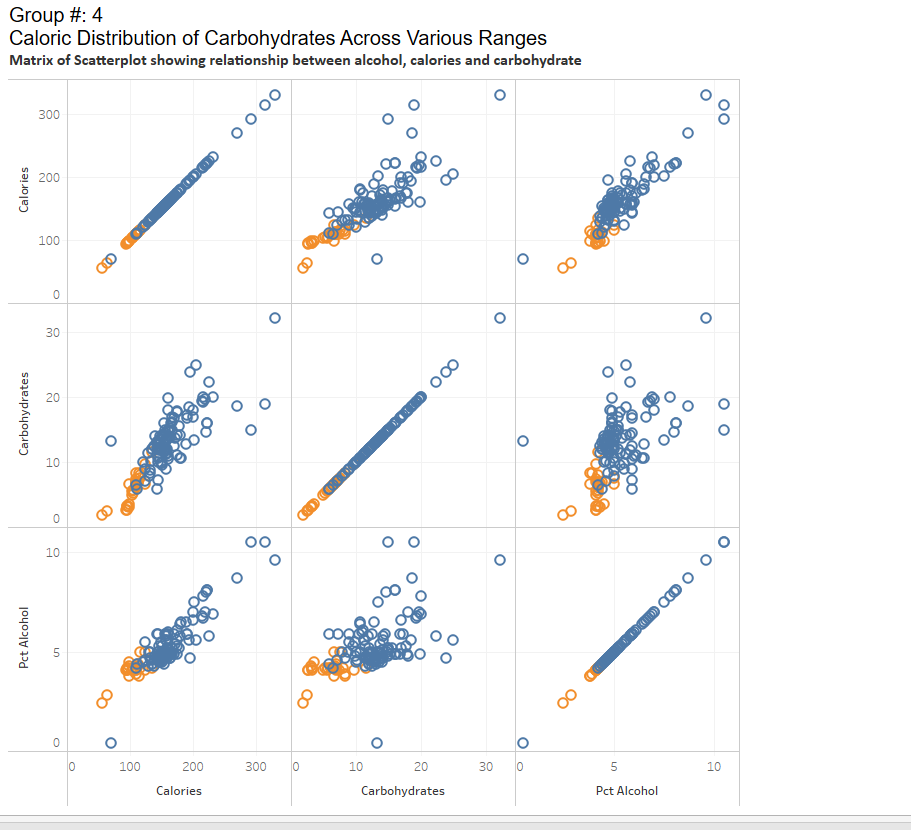
**Calories 0.0275028 0.0011233 24.4849 < 0.0001**

**intercept 0.952598 0.177656 5.36203 < 0.0001**

* 1. From the Data tab, drag and drop “Light” from Dimensions onto the Color card. You should now get two trend lines of different colours, for the different types of beer. Paste a screenshot of your graph below: 

## **Part II: Matrix of Scatterplots**

1. Add a new worksheet to the file you were working on in Part I.
2. Drag “Calories”, “Pct Alcohol” and “Carbohydrates” onto both rows and columns and deselect “Aggregate Measures” (as you did in section (c) of Part I).
3. Drag the “Light” dimension onto the Color card.
4. Give the graphic a descriptive title, then paste a screenshot of your visualization below:



1. Save your tableau file as **Lab04\_1.twb**.

## **Part III: Analysis**

Write a detailed analysis of the Beer data below, based on your visualizations:

**The Scatterplot of calories and alcohol presents clear relationship indicating the higher-calorie beers tend to have a higher alcohol percentage.**

**The second visualization is scatterplot matrix which represents relations between calories, carbohydrates and alcohol content.**

**Calories vs Alcohol:**

**Represents beer with higher alcohol content tend to have more calories.**

**Calories vs. Carbohydrates:**

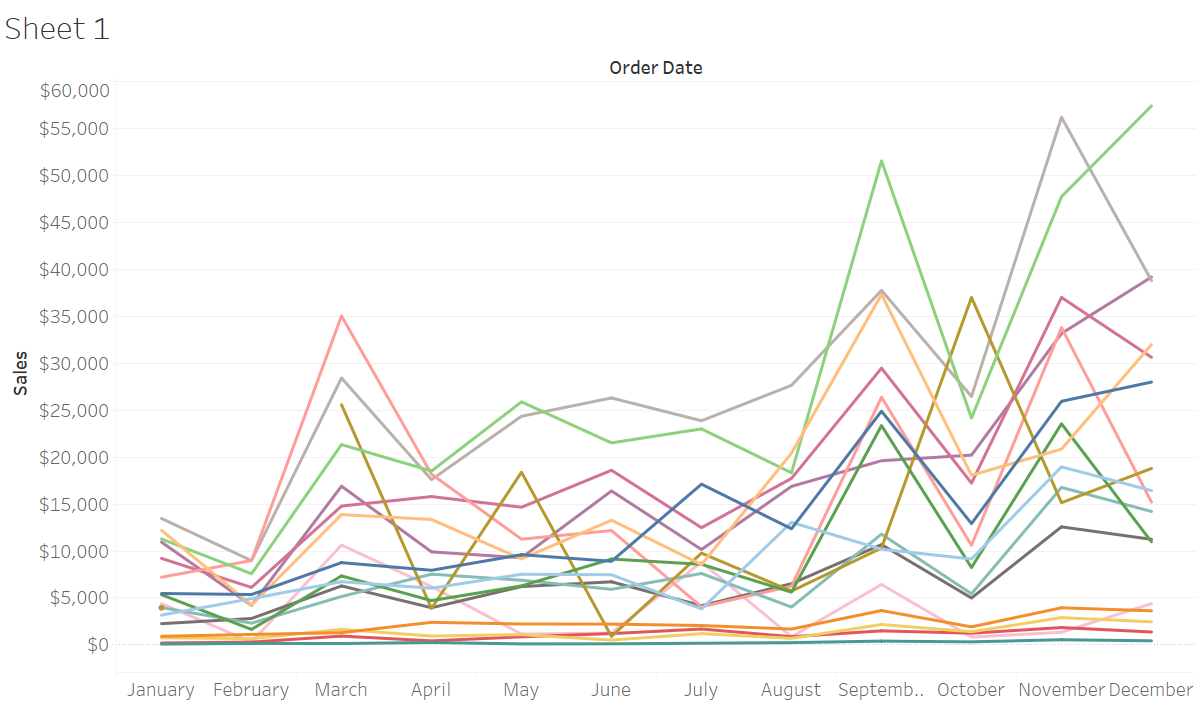
**Represent beer with more calories tend to have more carbohydrate.**

**Carbohydrates vs. Alcohol Percentage:**

**Relation is not as strong but it is representing higher alcohol content beers tend to have more carbohydrates**

## **Part IV: Heat Maps**

1. Open Tableau Desktop.
   1. Open the Sample superstore dataset.
   2. Drag “Order Date” dimension onto the Columns card. Right click and choose ‘Month’.
   3. Drag “Sales” onto into Rows.
   4. Drag the Product “Sub-Categories” dimension onto the Color cards. You should get this plot: (Ctrl+Shift+B if you want to make it wider)



1. As you can see, this is a much-cluttered plot! Open a new worksheet to create a heat map to present the same information.
2. Drag “Order Date” dimension onto the Columns card. Right click and choose ‘Month’. Then right-click and choose ‘Discrete’ (it might already be selected).
3. Drag Product “Sub-Categorie s” dimension onto Rows.
4. Now drag “Sales” onto the Colors card. You should see a heat map now.
5. From the Color card, choose ‘Edit Colors’ and choose a different color palette. Note the diverging vs. sequential palettes. Experiment with both choices, then describe the difference below:

**Results from experiments show that (I) the “Sub-Category” and “Order Date” data yielded the following outcomes.**

* **In a sequential color scheme, the high-selling Phones sub-category would display a gradual color shift while the lower-selling Supplies category would maintain clear distinctions for easily tracked trends.**
* **The diverging color scheme uses a midpoint designation to make upper values stand out as “Machine” in blue and lower values as “Fasteners” in red. Key Difference**

**Sequential Palettes create an ordered sense of magnitude through their color patterns and diverging palettes establish central value distinctions by opposing color schemes. The trends required a sequential palette due to its ordered design but comparisons to bookmarks needed a diverging color scheme.**

1. Which type of color palette (diverging or sequential) is most suitable in this case (for sales)? What would be an example of an application where the other choice would be more appropriate?

**a sequential color palette is typically more appropriate for visualizing sales data like the provided “SUM(Sales) “across months or sub-categories. Sequential palettes emphasize magnitude and trends, making distinguishing lows and high sales easy. Example**

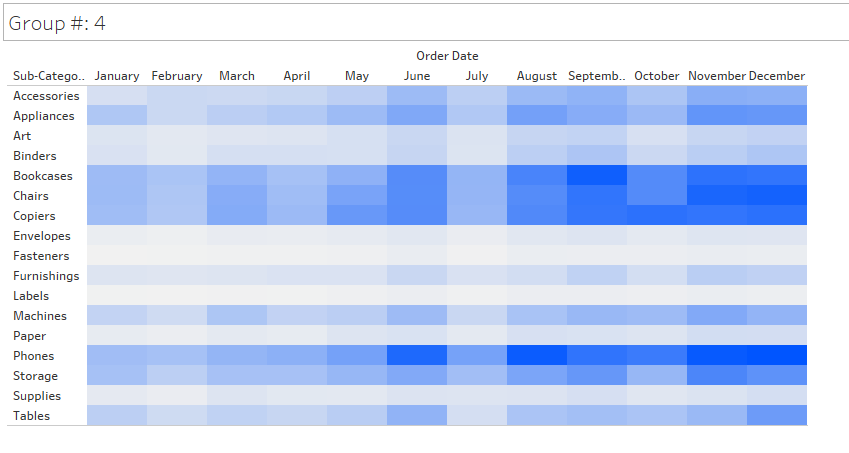
* **Using a gradient from light yellow (low sales) to dark green (high sales) to highlight which months (eg, December peaks) or sub-categories (eg: “Machines” vs “Fasteners”) perform best**

1. Finalize the color palette with one that you think works well.

**Finalized Palette:**

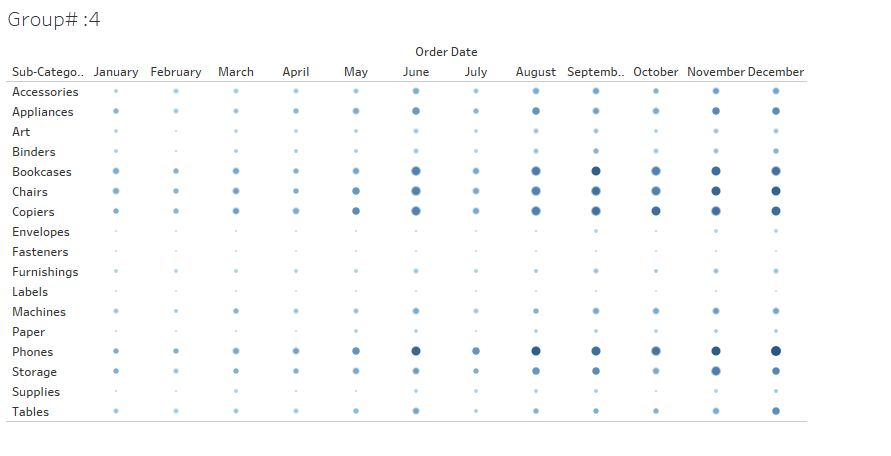
* **Name “Blue” (sequential)]**
* **Appearance: Light Blue (low) to Dark Blue (High)**
* **This choice balances readability, accessibility, and effective communication of sales trends**

1. Edit the title to something appropriate, and paste a screenshot of your heat map below:



## **Part V: More heat maps**

1. Add a new worksheet to the file you were working on in part IV.
   1. Repeat steps (f) and (g) from part IV.
   2. From Marks card, choose Circle.
   3. Drag and drop Sales onto the Color card.
   4. Now drag and drop Sales onto the Size card.
   5. Edit the title and paste a screenshot of your heat map below:



* 1. Save your tableau file as **Lab04\_2.twb** and submit.

## **Part VI: Analysis**

Write a detailed analysis of the Superstore data below, based on your visualizations:

**Bubble Chart**

The initial visualization employs a bubble chart that displays sales information through bubbles according to product sub-category monthly standings.

**Insights:**

* Each sub-category bubble shows cumulative sales reached during the specified month.
* The visual intensification between dark and big circles indicates higher sales while light and small circles indicate lower sales.
* Throughout the year, sales data reveals continuous variations that demonstrate better performance from specific product sub-categories.
* Definite yearly patterns exist since specific product classes, including Chairs and Tables, alongside Phones, record elevated sales numbers.

**Heatmap ( second image )**

The sales distribution is represented by a heatmap through a color intensity scale, beginning with light yellow and ending with green.

**Insights:**

* The sales levels in a particular area become more apparent as the colors transition from lighter to darker greens.
* Particular product subcategory sales concentrate mainly in specific months of the year.
* The visual presentation indicates which months bring in the maximum revenue.
* The visual distribution showcases smooth sales patterns, but specific months appear with higher sales volumes.

**The third heatmap visualization displays an orange-blue scale (Third Image).**

* The sales variations become more noticeable through an orange-to-red and blue color scheme in this representation.

Colors used for visualization separate the optimal sales periods between different product sub-categories.

The monthly sales patterns become easier to compare using this chart.

The sub-categories of Phones and Copiers present notable variations in sales patterns which lead to extreme highs and lows.

**3. Key Takeaways**

* The sales volume of some product groups reaches its highest points during particular months because these products usually have seasonal demand characteristics.
* Different sub-categories demonstrate steady consistency in sales performance although they do not exhibit consistent sales patterns.
* Some product groups demonstrate consistent sales records through Chairs and Tables while Phones show big differences between their highest and lowest points.

**Data Visualization Effectiveness:**

* The bubble chart proves efficient at tracking time-based proportional sales data.
* A diverging heat-map provides more efficient insights regarding performance trends and gaps than the basic lighter-colored heat-map.