

# LAB 5: VISUALISING CORRELATION



## Table of Contents

Learning Outcome.....	2
Data Preparation for Task 1 & 2 .....	2
Task 1: Using Pearson Correlation to Make Prediction .....	3
Task 2: Build a Control Chart.....	7
Data Preparation for Task 3 .....	13
Task 3: Create a Market Basket Analysis .....	15
Data Preparation for Task 4 .....	18
Task 4: Analyse Survey Data .....	18

## Learning Outcome

At the end of this session, learners will be able to:

- Use Pearson Correlation to Make Prediction
- Build a Control Chart
- Create a Market Basket Analysis
- Analyse Survey Data

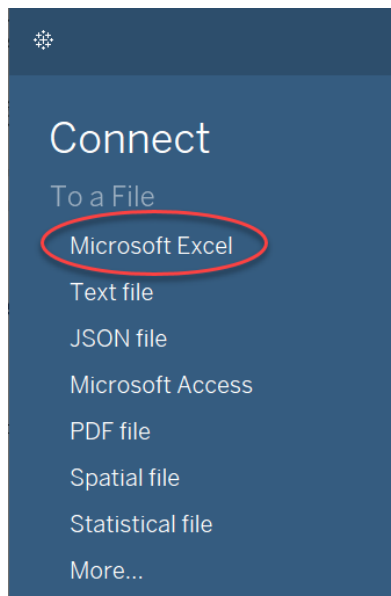
## Data Preparation for Task 1 & 2

Create a Tableau workbook that connect to the **Sample - Superstore Subset (Excel)** data source.

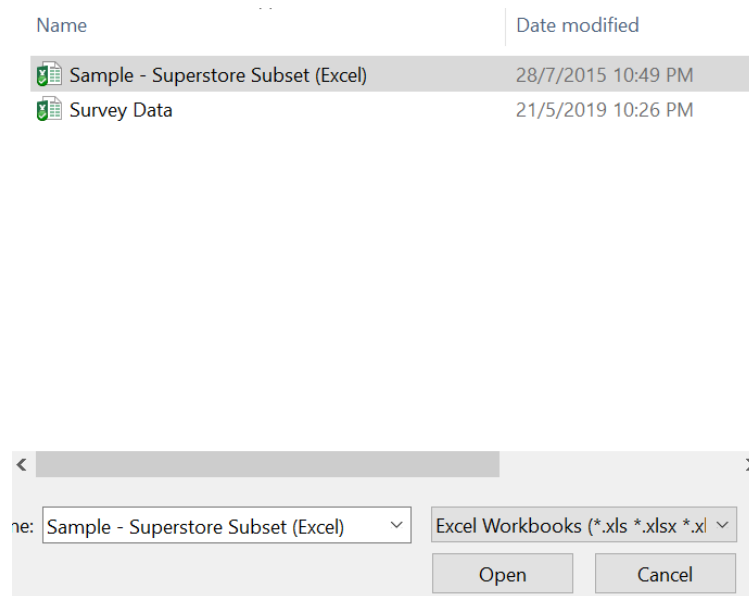
Step 1. Launch Tableau. Under Connect To a File, select **Microsoft Excel**.

Step 2. From the file open window, select the **Sample - Superstore Subset (Excel)** excel file.

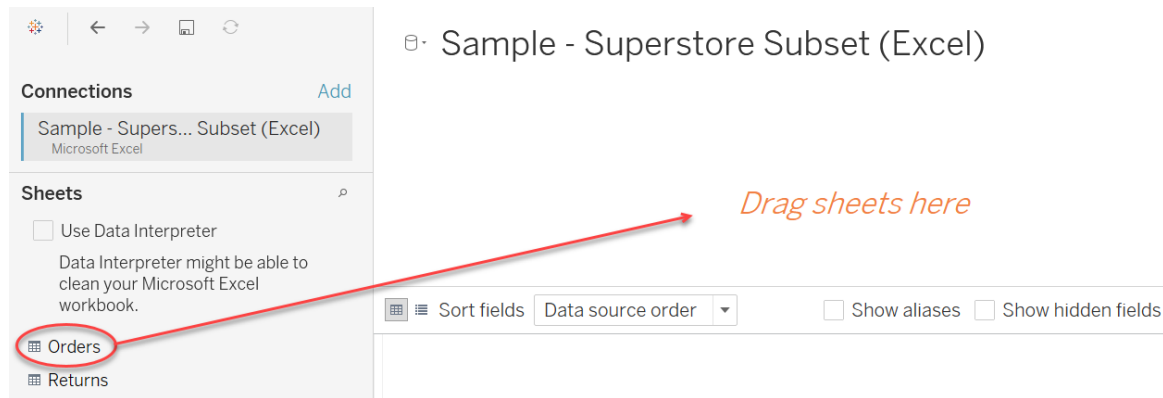
### Step 1



### Step 2



Step 3: Drag **Orders** to the data sheet.



## Task 1: Using Pearson Correlation to Make Prediction

The purpose of finding correlations is to allow us to make predictions about one measure based on what we know about another measure. For example, there is a correlation between income and education. We find that people with higher incomes have more years of education. Knowing this, if we know a person's income, we can predict their years of education.

### Pearson correlation

There are a few different types of correlations, but the one we will use is the Pearson correlation, developed by Karl Pearson in the 1880's. A correlation,  $r$ , is a single number that represents the degree of relationship between two measures. The correlation coefficient is a value such that  $-1 \leq r \leq 1$ .

A positive correlation indicates a relationship between  $x$  and  $y$  measures such that as the values of  $x$  increases, the values of  $y$  will also increase.

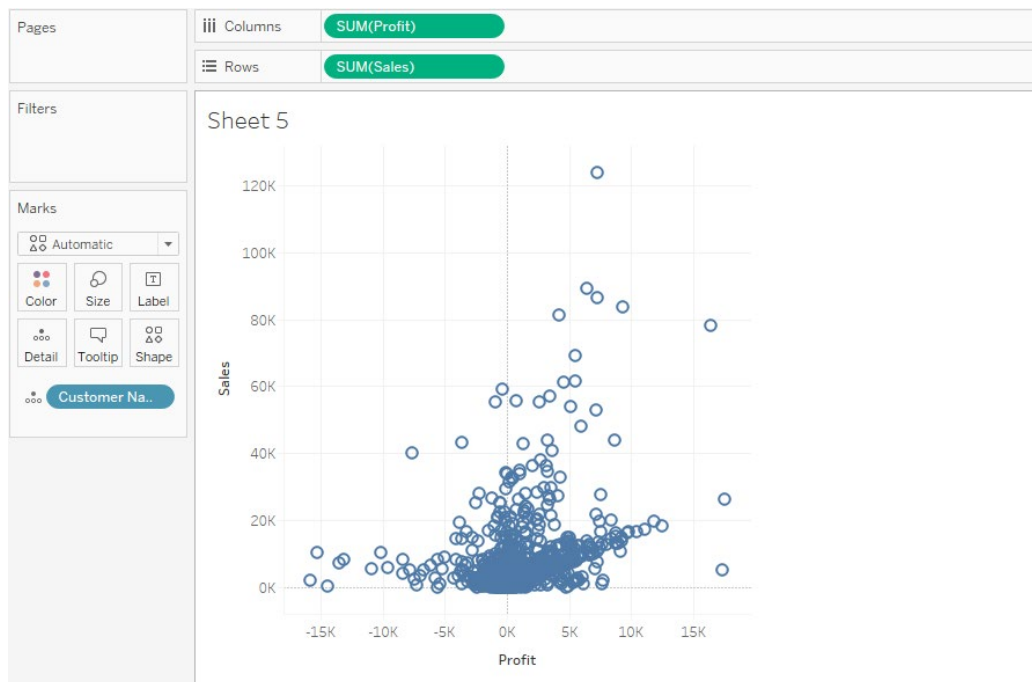
As a general rule,  $r$  between:

- +0.70 or higher Very strong positive relationship
- +0.40 to +0.69 Strong positive relationship
- +0.20 to +0.39 Moderate positive relationship
- 0.19 to +0.19 No or weak relationship
- 0.20 to -0.39 Moderate negative relationship
- 0.40 to -0.69 Strong negative relationship
- 0.70 or lower Very strong negative relationship

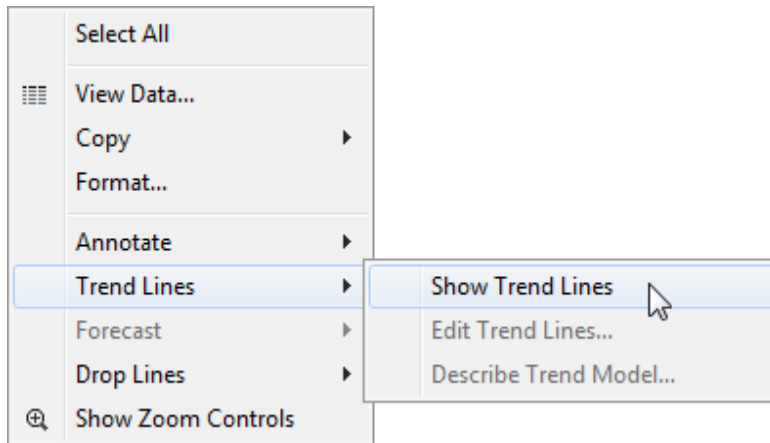
This exercise will show you how to find the correlation of two measures in Tableau.

## Create a scatter plot view

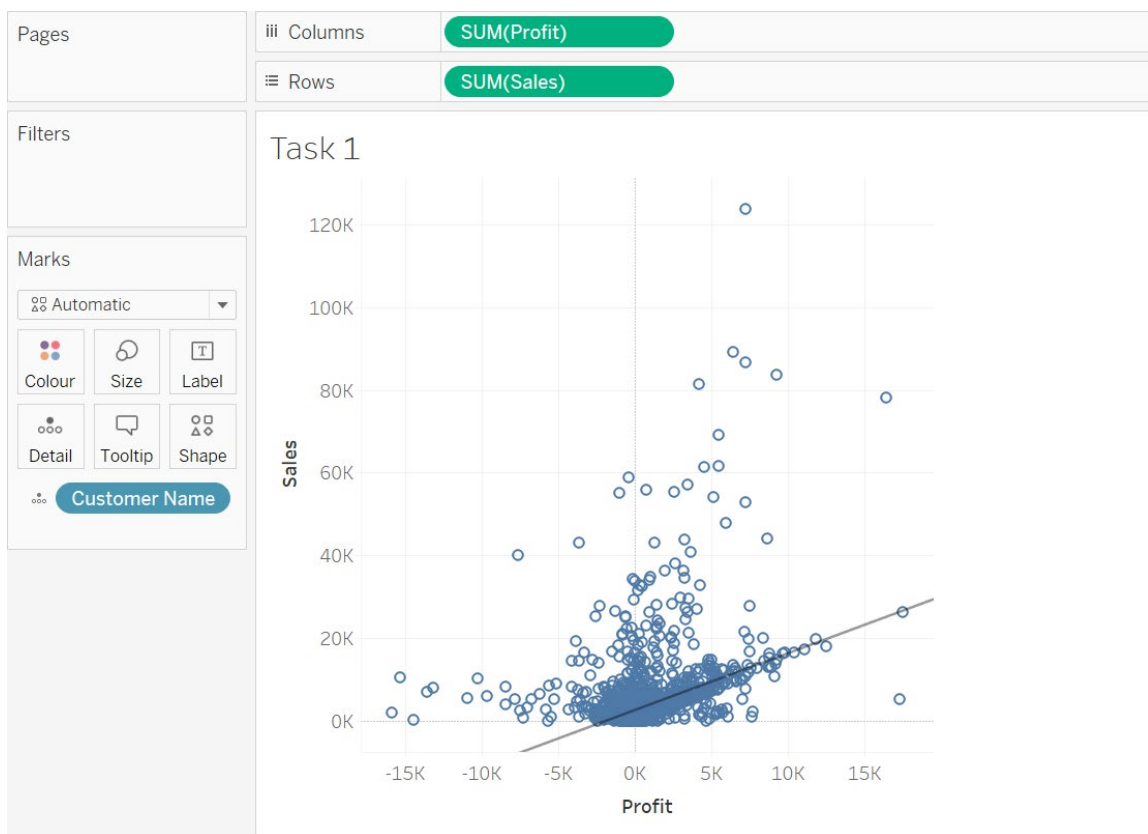
1. Open a new sheet and rename it as **Task 1**.
2. Make sure **Sample - Superstore Subset (Excel)** data source is selected.
3. Drag the **Profit** field to the **Columns** shelf and **Sales** field to the **Rows** shelf.
4. Drag a Dimension such as **Customer Name** to **Detail** shelf.



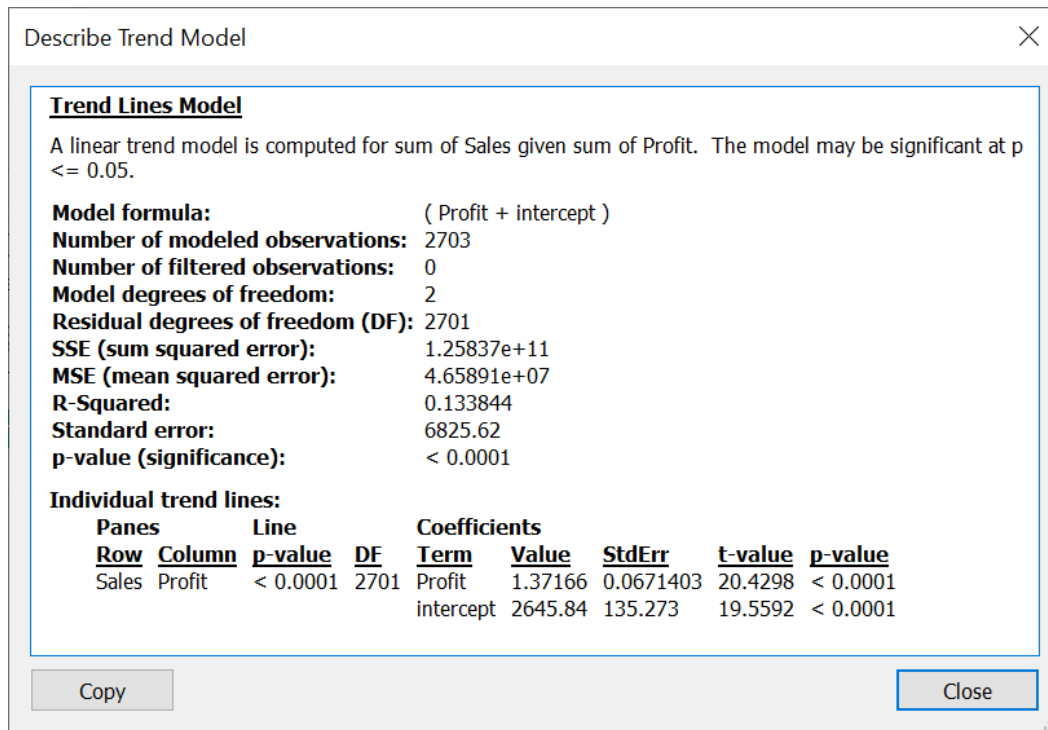
5. Right click the view and select **Trend Lines** → **Show Trend Lines**.



6. The trend line displays in the view.



7. Right click the view again and select **Trend Lines** → **Describe Trend Model**.



## Calculating the correlation

Locate the R-squared. In this example, it is 0.133844. Using a calculator or program, take the square root of it. Rounded up, the value is 0.37. This is our correlation,  $r$ .

Notice the trend line has a positive slope, this indicates that our correlation is also positive. If, instead the trend line had a negative slope, our correlation would be -0.37. This is a key importance to note. This .37 correlation tells us that there is a moderate positive correlation between sales and profit.

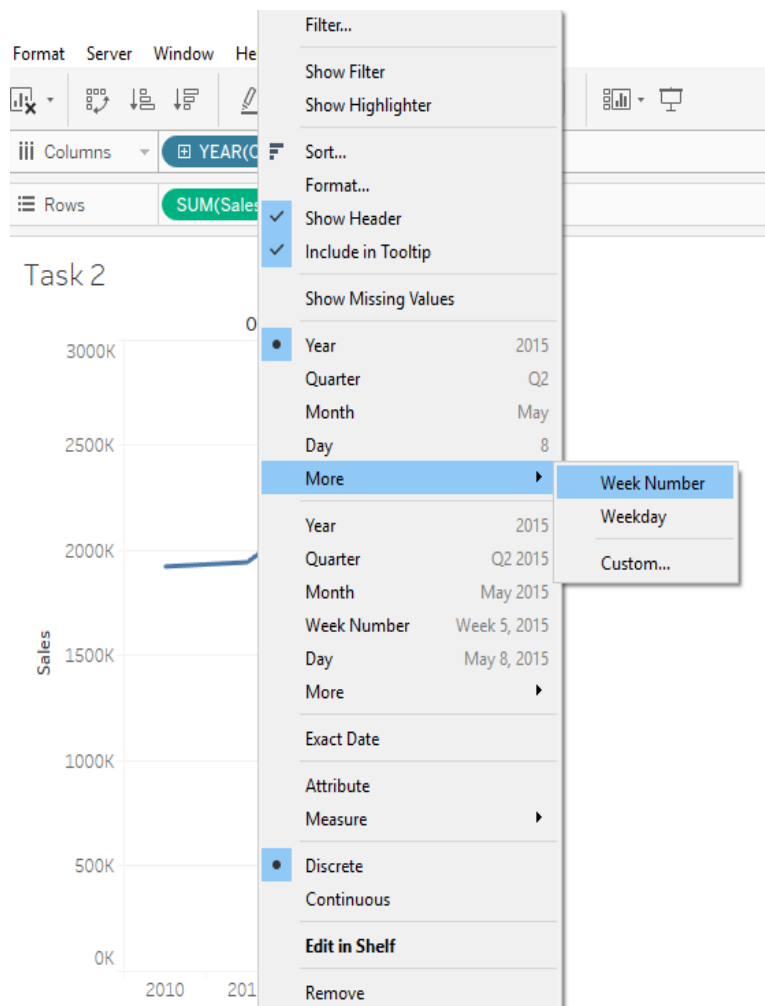
Adapted from <http://kb.tableau.com/articles/howto/finding-the-pearson-correlation>

© 2003-2013 Tableau Software. All Rights Reserved

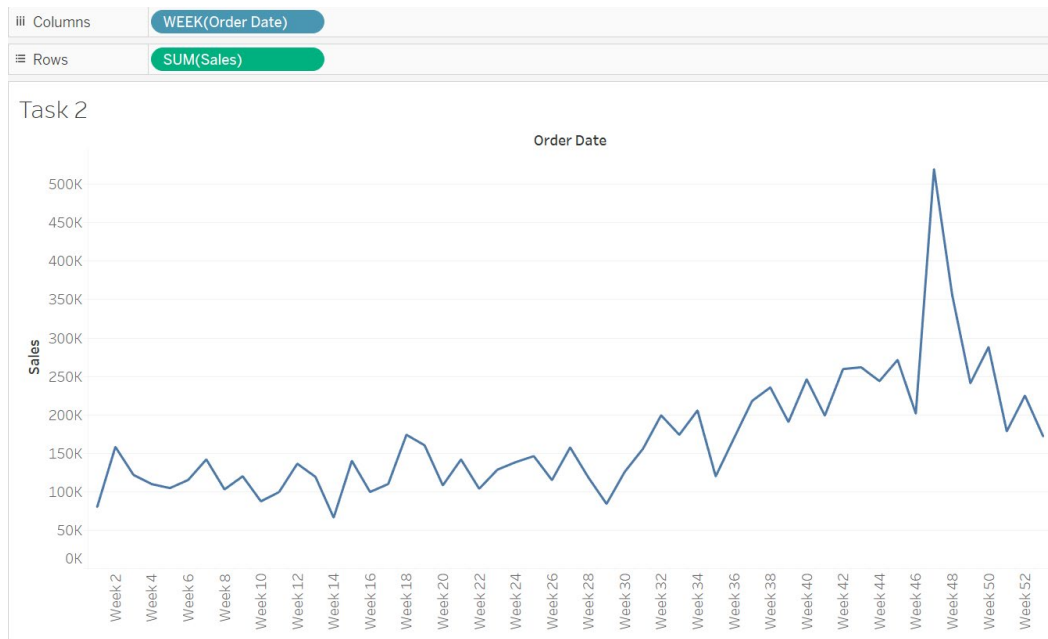
## Task 2: Build a Control Chart

A control chart is a visualization used to determine whether a process is in a state of statistical control. It involves displaying a line graph with an average reference line and typically three standard error reference lines above and below that average line. This article explains how to build a control chart.

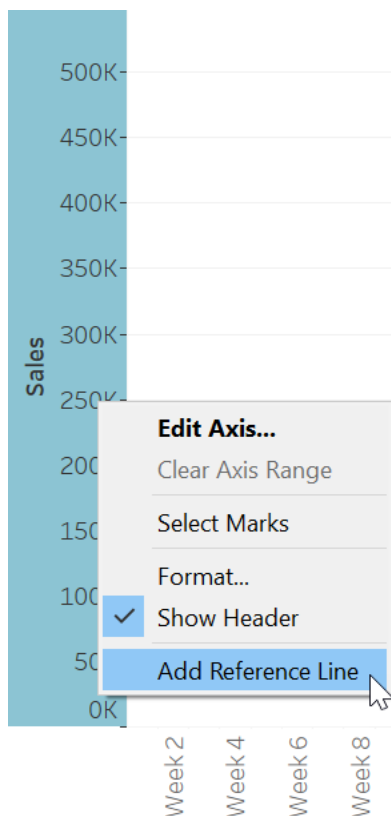
1. Open a new sheet and rename it as **Task 2**.
2. Make sure **Orders (Sample - Superstore Subset (Excel))** data source is selected.
3. Put a dimension (**Order Date in week number**) on the **Columns** shelf and a measure (**Sales**) on the **Rows** shelf.







- Right click the **Sales** axis and select **Add Reference Line**.



- In the **Add Reference Line** dialog box, the default settings are correct.

Add Reference Line, Band or Box

Line Band Distribution Box Plot

Scope  
☐ Entire Table ☒ Per Pane ☐ Per Cell

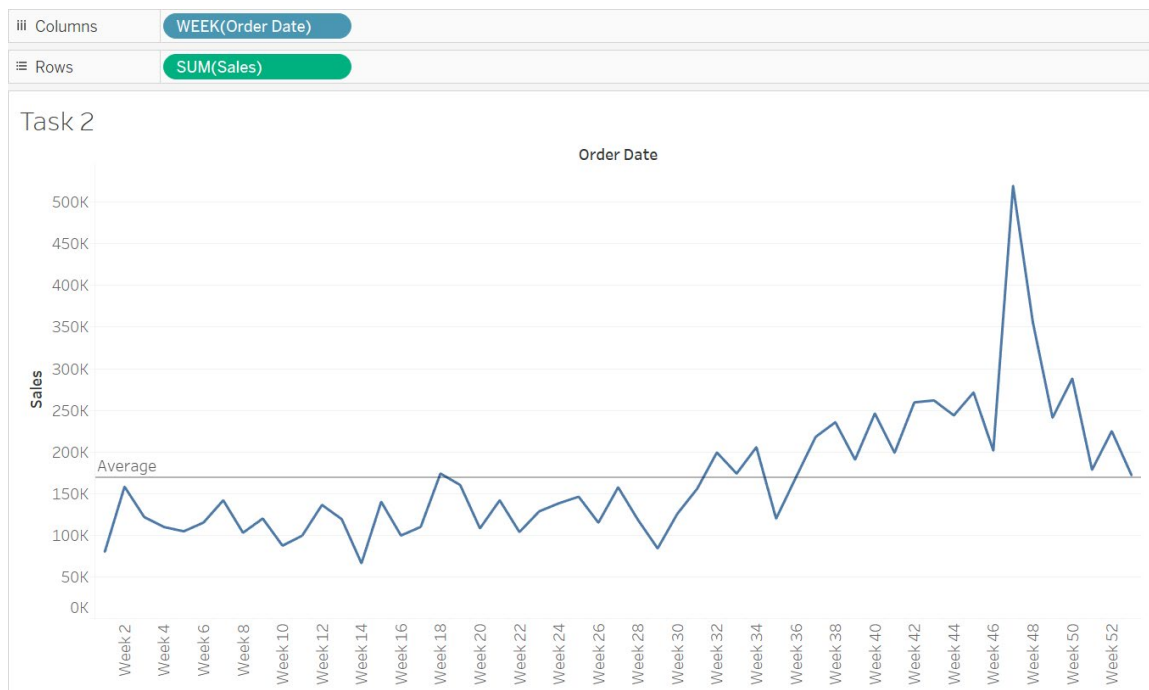
Line  
 Value: SUM(Sales) Average  
 Label: Computation  
 Tooltip: Automatic  
 Line only 95

Formatting  
 Line:   
 Fill Above: None  
 Fill Below: None

☒ Show recalculated line for highlighted or selected data points

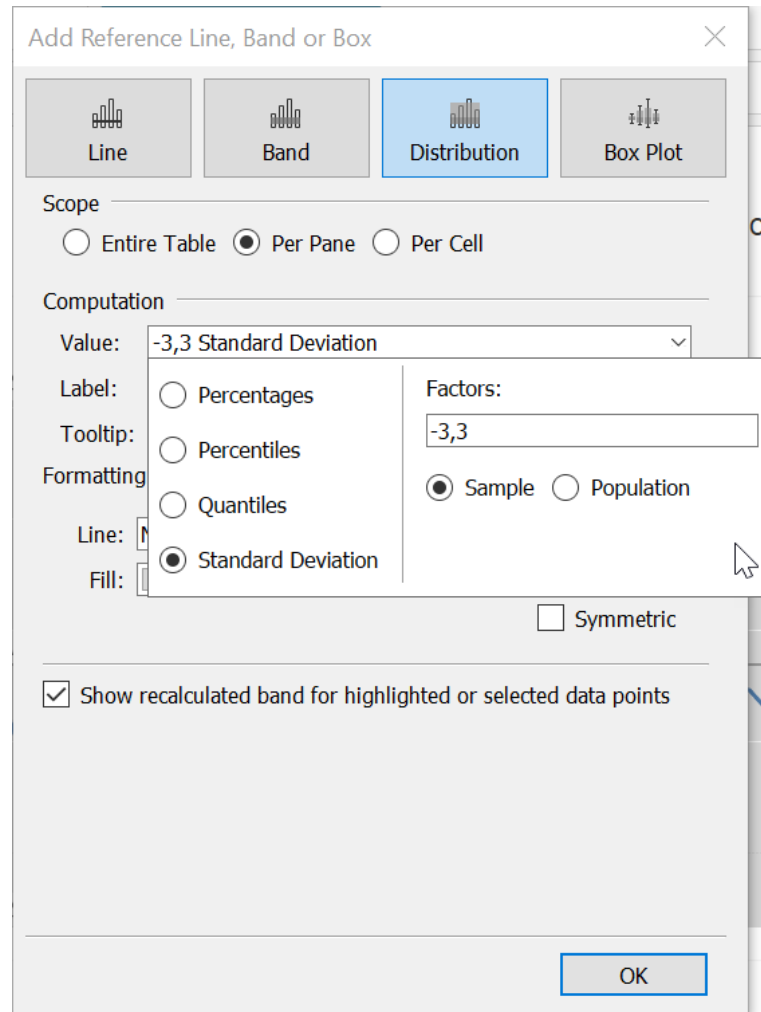
OK

- Click **OK**.
- The **Average** reference line appears in the view.



- Again, right click the **Sales** axis and select **Add Reference Line**.

9. In the **Add Reference Line** dialog box, select **Distribution**.
10. Under **Computation**, in the **Value** list, click the list arrow.
  - a. Select the **Standard Deviation** option.
  - b. In the **Factors** text box, type **-3,3**.
  - c. Keep the default selection of **Sample**.
  - d. Click outside the dropdown window to close it.



11. Under **Formatting**, in the **Line** list, select different colour for the line.

Add Reference Line, Band or Box

Line Band **Distribution** Box Plot

Scope

☐ Entire Table ☒ Per Pane ☐ Per Cell


Computation


Value: -3,3 Standard Deviation

Label: Computation

Tooltip: Automatic

Formatting

Line:  ☐ Fill Above

Fill:  ☐ Fill Below

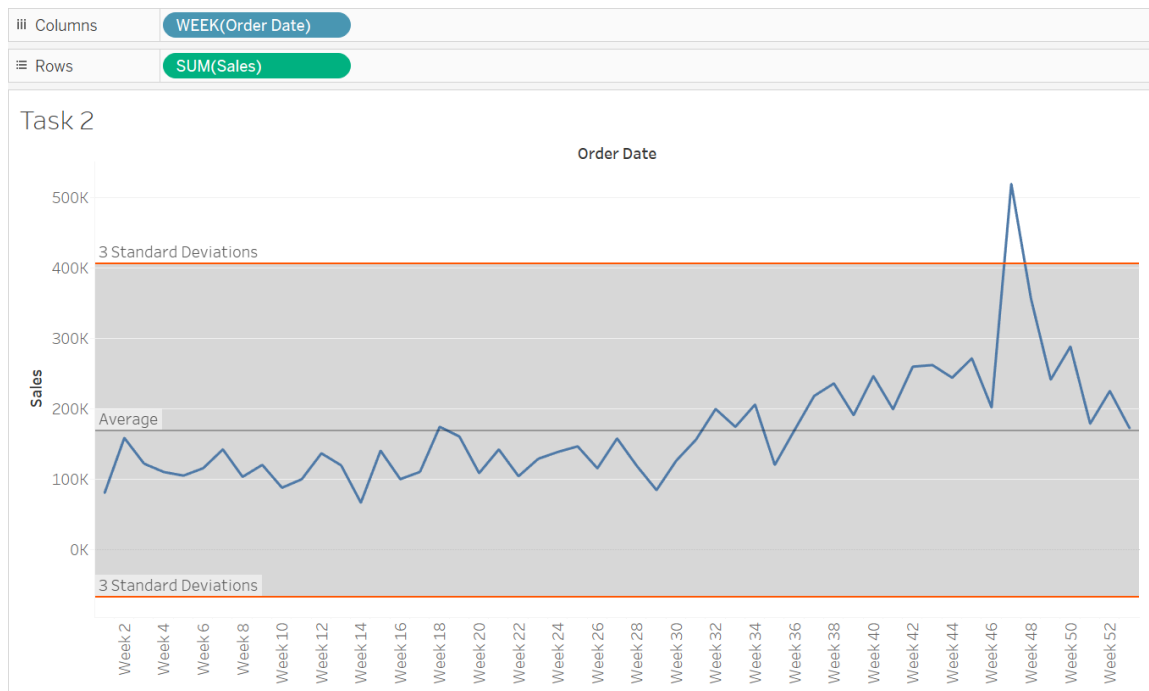
☐ Symmetric

☒ Show recalculated band for highlighted or selected data points

OK

12. When finished, click **OK**.

13. Choosing the sample standard deviation gives you the standard error. With factors of  $-3, 3$  you get a  $6\sigma$  spread, where the top line is the UCL (Upper Control Limit) and bottom line is the LCL (Lower Control Limit).




Adapted from <http://kb.tableausoftware.com/articles/knowledgebase/control-charts>

© 2003-2013 Tableau Software. All Rights Reserved

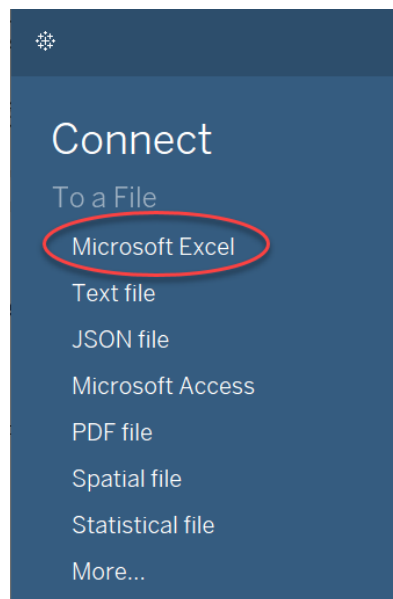
## Data Preparation for Task 3

Before you can do the market basket analysis, you need to set up the data source to facilitate it. The remaining sections of this article use the Superstore sample data source to join the table to itself in a self-join, and then create table and heat map views for the market basket analysis.

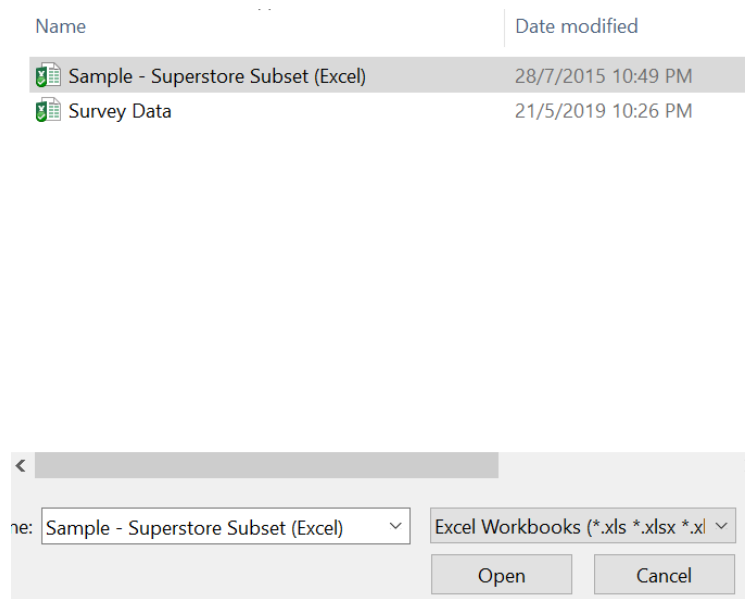
Step 1. Click on  icon to go to **Data Source** page. Under Connect To a File, select **Microsoft Excel**.

Step 2. From the file open window, select the **Sample - Superstore Subset (Excel)** excel file.

### Step 1



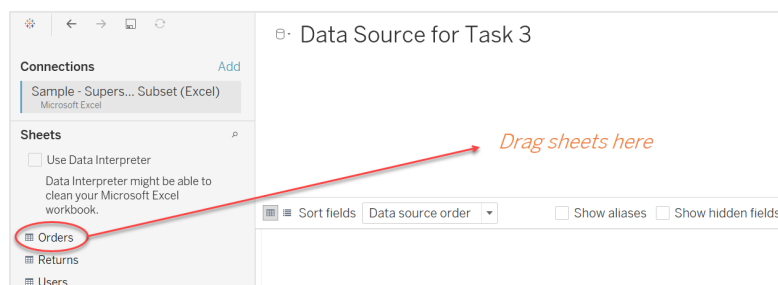
### Step 2



Step 3. Rename the data source as **Data Source for Task 3**.

## Create a joint data set

1. Drag **Orders** to the data sheet.



2. Double click Orders table in the datasheet and select duplicate.
3. Set the Join Condition as **Inner Join** with following conditions:  
**CustomerID = CustomerID (Orders1)** and  
**Product Sub-Category <> Product Sub-Category (Orders1)**

Edit Relationship

How do relationships differ from joins? [Learn more](#)

Orders

Orders1

# Customer ID = # Customer ID (Orders1)

Abc Product Sub-Category <> Abc Product Sub-Category (Orders1)

Abc City

# Customer ID

Abc Customer Name1

Abc Customer Segment

# Discount

Order Date

Edit Calculation...

<>

Abc City (Orders1)

# Customer ID (Orders1)

Abc Customer Name (Orders1)

Abc Customer Segment (Orde

# Discount (Orders1)

Order Date (Orders1)

Edit Calculation...

Close

## Task 3: Create a Market Basket Analysis

You can use *market basket* analysis to discover and understand customer purchasing behaviour. For example, you can use it to answer questions such as the following:

- How many people bought both Product A and Product B?
- People who purchased Product A also generally buy which other products?
- Students who enrolled in Course A also frequently enroll in which other courses?

In this task, we are going to analyse customer who purchased product in Category A also buy products in which other categories.

### Create a table view

To create the table view, you drag the field you want to analyse from one table to the Columns shelf, and the same field from the other table to the Rows shelf, as demonstrated in the following example.

1. Open a new sheet and rename it as **Task 3**.
2. Make sure **Data Source for Task 3** data source is selected.
3. From the Dimensions pane, drag the following fields to the view:
  - a. Drag **Product Sub-Category** to the **Columns** shelf.
  - b. Drag **Product Sub-Category(Orders\$1)** to the **Rows** shelf.

The screenshot shows the Tableau interface with a table view titled 'Task 3'. The Columns shelf contains 'Product Sub-Category' and the Rows shelf contains 'Product Sub-Category(Orders\$1)'. The table displays a grid of 'Abc' values for various product sub-categories.

Product Sub-Cat...	Appli...	Binde...	Book...	Chair...	Comp...	Copie...	Envel...	Labels	Offic...	Offic...	Paper	Pens...	Rubb...	Sciss...	Stora...	Tables	Telep...
Appliances		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Binders and Bind...	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Bookcases	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Chairs & Chairma...	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Computer Periph...	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Copiers and Fax	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Envelopes	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Labels	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Office Furnishings	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Office Machines	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc	Abc
Paper	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc	Abc
Pens & Art Suppl...	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc	Abc
Rubber Bands	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc	Abc
Scissors, Rulers ...	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc
Storage & Organ...	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc	Abc
Tables	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc
Telephones and ...	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	

4. From the **Measures** pane, drag **Orders(Count)** to **Text** on the **Marks** card.



5. The view should resemble the following image.

Pages	iii Columns	Product Sub-Category
	Rows	Product Sub-Categor...
Filters	Task 3	
Marks	Product Sub-Category	
Automatic	Product Sub-Cat... Appli... Binde... Book... Chair... Comp... Copie... Envel... Labels Offic... Offic... Paper Pens... Rubb... Sciss... Stora... Tables Telep...	
Colour	Appliances	
Size	Binders and Bind...	
Text	Bookcases	
Detail	Chairs & Chairma...	
Tooltip	Computer Periph...	
SUM(Number ...)	Copiers and Fax	
	Envelopes	
	Labels	
	Office Furnishings	
	Office Machines	
	Paper	
	Pens & Art Suppl...	
	Rubber Bands	
	Scissors, Rulers ..	
	Storage & Organ...	
	Tables	
	Telephones and ..	

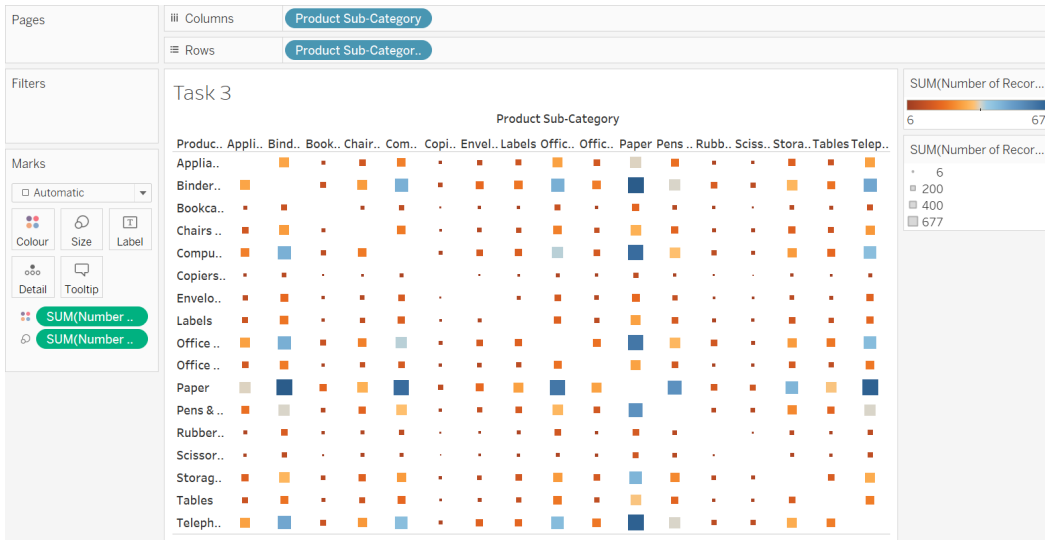
In the table, each number represents the number of individuals who have purchased something from the category shown on each corresponding axis (not necessarily at the same time). For example, in the image above, 92 people have purchased items from both the Bookcases and the Binders and Binder Accessories categories.

Notice that the numbers reflect each other along a diagonal axis, where the same two categories are represented.

## Create a heat map view

Here's another way to display the market basket analysis, continuing from the table view you created at the end of the previous section.


1. On the Marks card, drag **Orders(Count)** to **Colour** (this also removes it from Text).
2. From the **Measures** pane, drag another instance of **Orders(Count)** to **Size** on the Marks card.
3. From the **Marks** card list of views, select **Square**.
4. Click **Size**, and then drag the slider to the right to enlarge the squares to the size you want.
5. Click **Colour** to change other attributes.



Adapted from <http://kb.tableausoftware.com/articles/knowledgebase/market-basket-analysis>

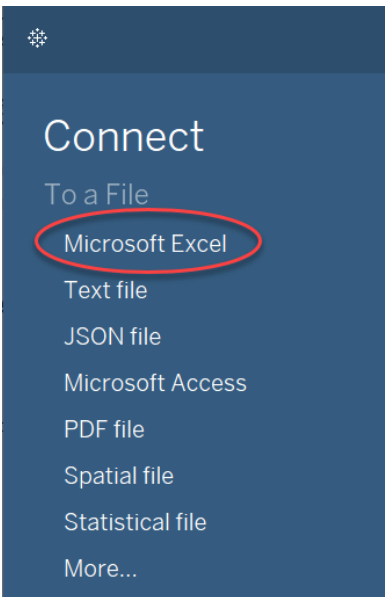
© 2003-2013 Tableau Software. All Rights Reserved

## Data Preparation for Task 4

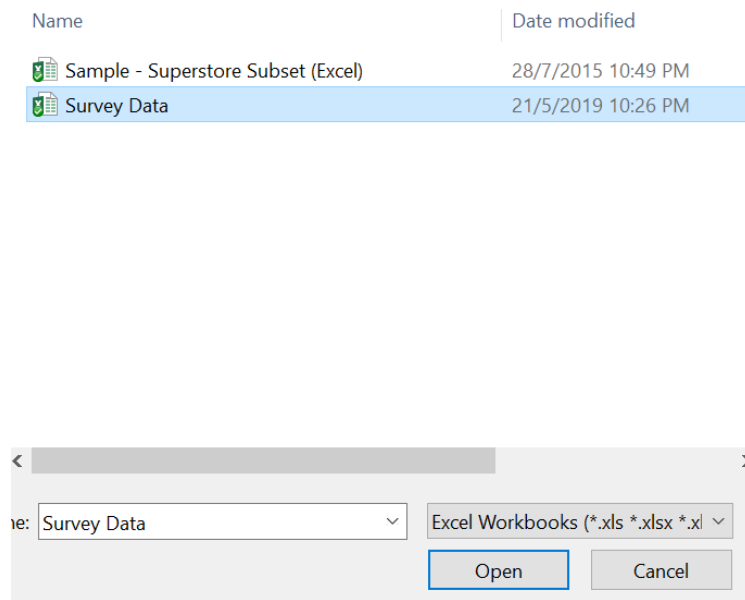
Step 1. Click on  icon to go to **Data Source** page. Under Connect To a File, select **Microsoft Excel**.

Step 2. From the file open window, select the **Survey Data** excel file.

### Step 1



### Step 2



## Task 4: Analyse Survey Data

Using Tableau, you can analyze survey data. Regardless of whether you use spreadsheets completed by hand or survey software packages, you may find that the data needs reshaping before you can use it with Tableau's full analytical power. One example of this situation is grid questions. Typically grid questions are presented with a matrix of options, and the person being surveyed selects the most appropriate box in each row.

For example, suppose you have the following grid question and survey.

I prefer this beverage:

	(1) Always	(2) Sometimes	(3) Never
(a) Coffee	X		
(b) Tea		X	
(c) Water		X	

## Survey Data

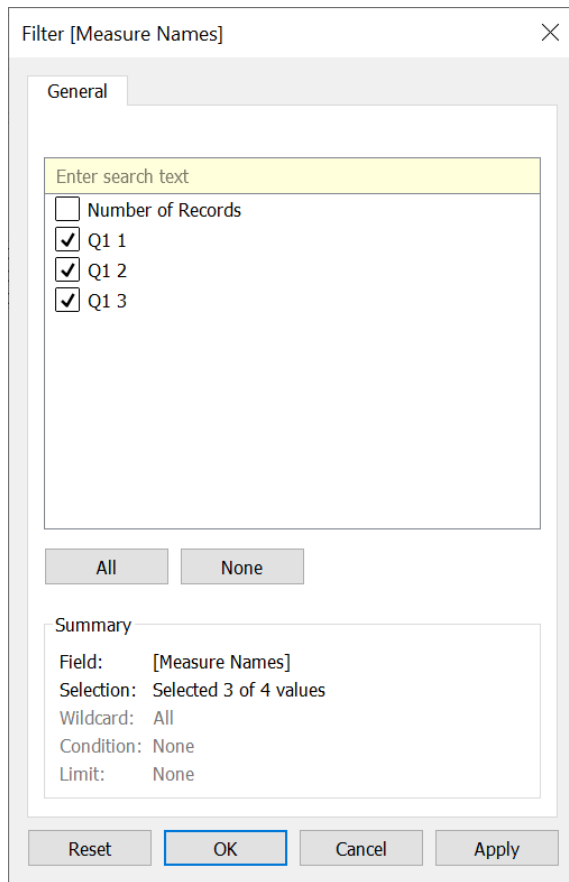
Survey ID	Row	Q1_	Q1_	Q1_
10001	Coffe	1	0	0
10002	Coffe	0	0	1
10003	Coffe	0	1	0
10004	Coffe	1	0	0
10005	Coffe	1	0	0
10001	Tea	0	1	0
10002	Tea	0	0	1
10003	Tea	1	0	0
10004	Tea	0	1	0
10005	Tea	0	1	0
10001	Wate	0	1	0
10002	Wate	1	0	0
10003	Wate	1	0	0
10004	Wate	0	0	1
10005	Wate	0	0	1

From this data, certain survey software gives the results in a tabular format. A "1" signifies a check mark and a "0" represents an empty box, like in the example above.

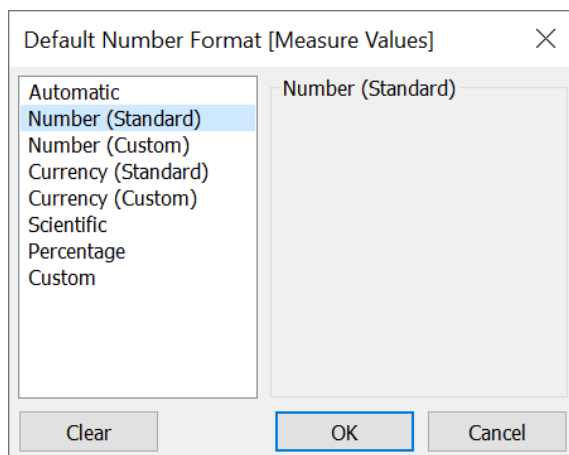
## Create a text table of your data

1. Open a new sheet and rename it as **Task 4**.
2. Make sure **Survey (Survey Data)** data source is selected.
3. From **Dimensions**, drag **Row** to the **Rows** shelf.
4. From **Dimensions**, drag **Measure Names** to the **Columns** shelf.

5. Drag **Measure Names** to the **Filters** shelf to show only the three answer columns.



6. Format the number format of Measure Values to remove the decimal places. Right click on the **Measure Values** and go to **Default Properties** → **Number Format**. Select **Number (Standard)** as the default number format.



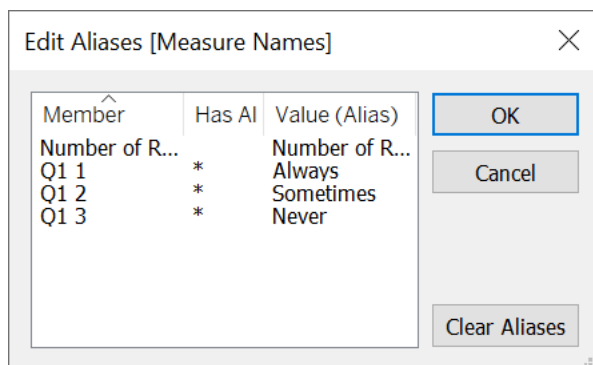
7. Drag and place **Measure Values** on the **Text Marks** card.

The screenshot shows the Tableau Desktop interface. On the left, the 'Marks' card is selected, and 'Measure Values' is added to the 'Text' mark. The 'Columns' shelf contains 'Measure Names' and the 'Rows' shelf contains 'Row'. The main view displays a table titled 'Task 4' with the following data:

Row	Q1 1	Q1 2	Q1 3
Coffee	3	1	1
Tea	1	3	1
Water	2	1	2

## Rename and Sort the measures

1. In the **Dimensions** area of the Data window, right click **Measure Names** → Aliases. In the Edit Aliases [Measure Name], rename the alias for Q1 1, Q1 2, Q1 3 to **Always**, **Sometimes**, and **Never** respectively.



2. In the table, click and drag the renamed headers to reorder them as necessary. Right click on the Row header select **Rotate Label** to the following format. The result is a text table of your survey with data on the number of people who responded to your question, and how they responded.

Pages

iii Columns
Measure Names

Rows
Row

Filters
Measure Names

Marks
Automatic

Colour

Size

Text

Detail

Tooltip

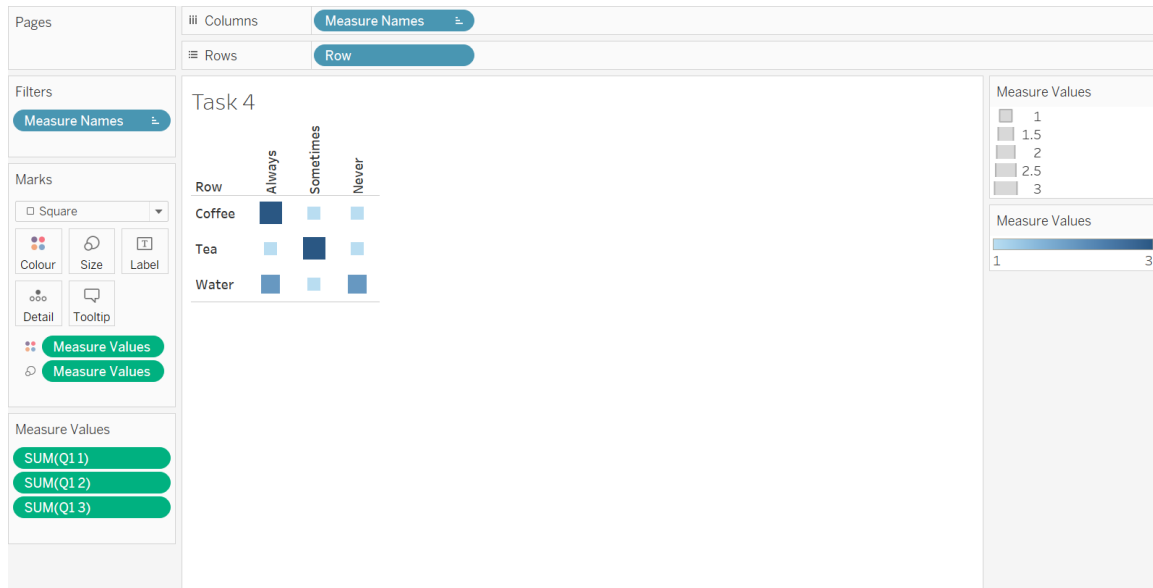
Measure Values

Measure Values
SUM(Q1 1)
SUM(Q1 2)
SUM(Q1 3)

### Task 4

Row	Always	Sometimes	Never
Coffee	3	1	1
Tea	1	3	1
Water	2	1	2

- You can change this view to a Heat Map by changing the mark type to **Square**, removing Measure Values from the **Text** shelf, and placing it on the **Size** Marks card and the **Colour** shelf.



Adapted from <http://kb.tableausoftware.com/articles/knowledgebase/survey-data-grid-questions>  
© 2003-2013 Tableau Software. All Rights Reserved