NANYANG THE INNOVATIVE POLYTECHNIC School of Information Technology

LAB 3: VISUALISING PROPORTIONS

Table of Contents

Learning Outcome

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

- Create bar chart
- Create a Grouped Bar Chart
- Add total to Stacked Bars
- Find the Top N within a Category
- Create pie chart with multiple labels
- Perform Pareto analysis on sales and profit

Data Preparation

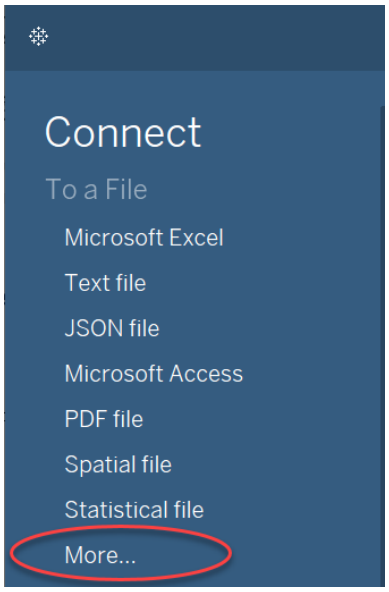
Create a Tableau workbook that connect to the **Sample - Superstore Subset (Excel) Tableau data source**.

We are going to use this data source for all the tasks of this lab.

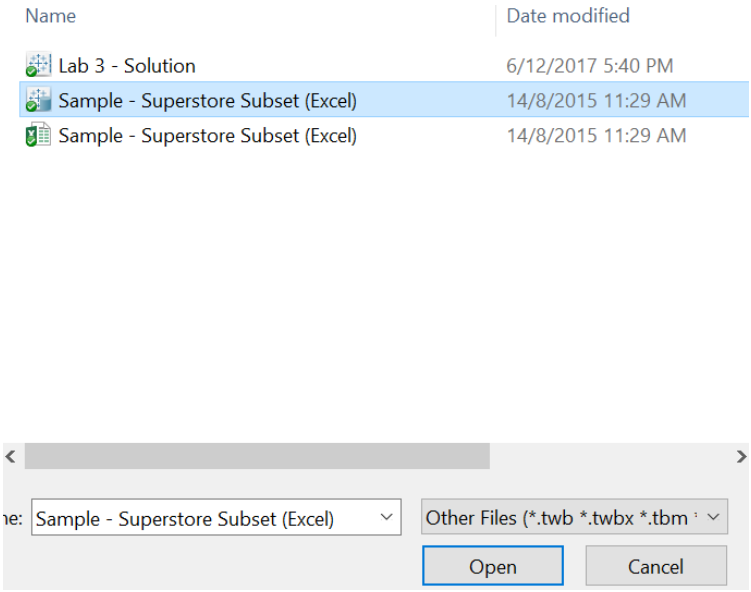
Step 1. Launch Tableau. Under Connect To a File, select **More…**

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

Step 1



Step 2

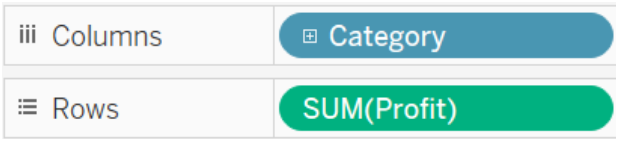


Task 1: Creating Bar Charts

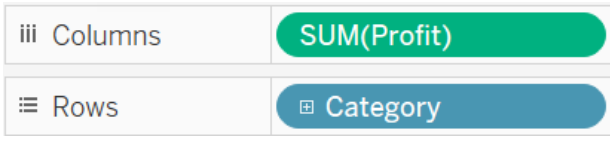
Bar charts are a great way to compare data across categories. In Tableau, you typically create a bar chart by placing a dimension on the **Rows** shelf, and a measure on the **Columns** shelf, or vice-versa.

A bar chart uses the bar mark type. Tableau automatically selects this mark type when the data view matches one of the two field arrangements shown below, provided the Marks type is set to **Automatic**. You can add additional fields on these shelves, and Tableau will not change the chart type.

Creates Vertical Bars



Creates Horizontal Bars



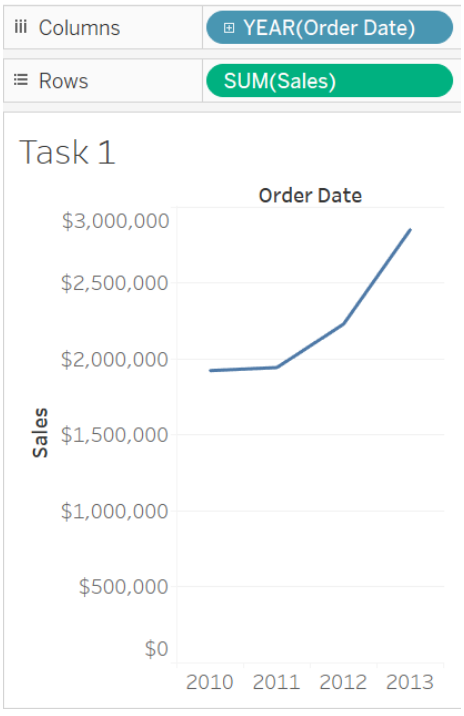
Alternatively, you can manually select **Bar** from the Marks drop-down list for any data view. For more information about bar marks, see Bar Mark.

To create a bar chart that displays total sales over a four-year period, follow these steps:

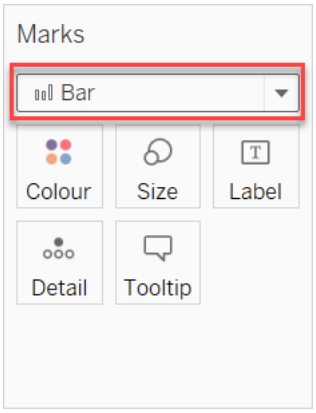
1. Open a new sheet, name it **Task 1**. Make sure the data source is connected to the **Sample - Superstore Subset (Excel)**.

2. Drag **Order Date** to the **Columns** shelf. The data is automatically aggregated by year and column headers are displayed.

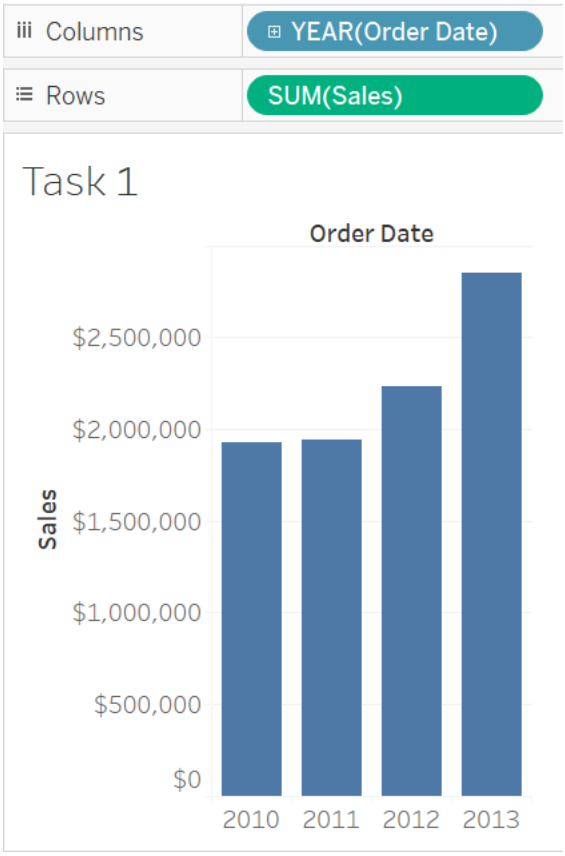
3. Drag **Sales** to the **Rows** shelf. The measure is automatically aggregated as a sum and an axis is created. The column headers are moved to the bottom of the view. Tableau automatically selects Line as the mark type because of the date dimension.



On the Marks card, from drop-down list of view types, select **Bar**.



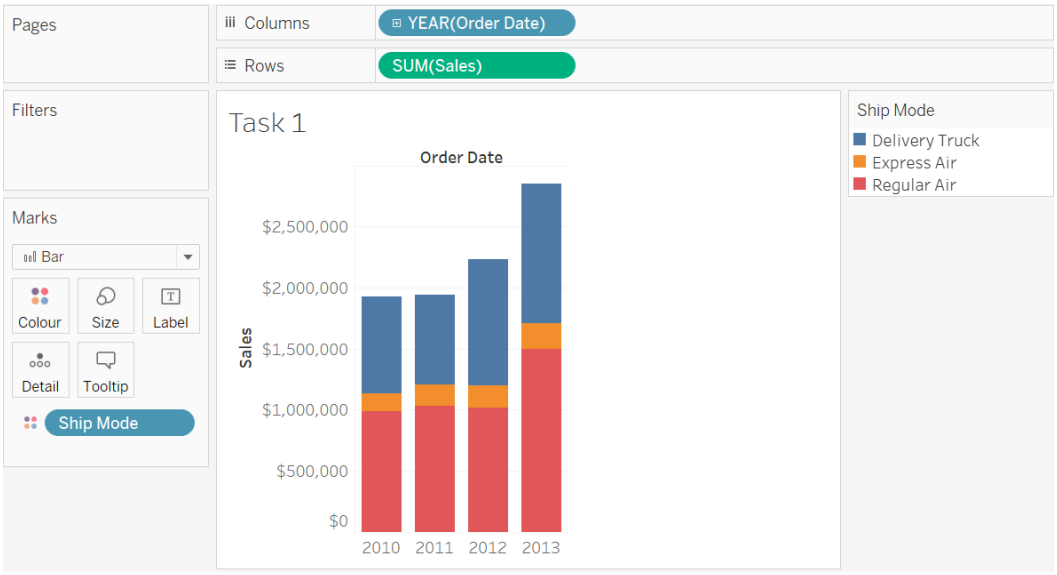
This changes the view to a bar chart.



The marks (bars) are vertical because the axis is vertical. The length of each mark represents the sum of the sales for that year. For example, the sum of the sales in 2013 is $2,852,360, which you can verify by hovering the mouse cursor over that column.

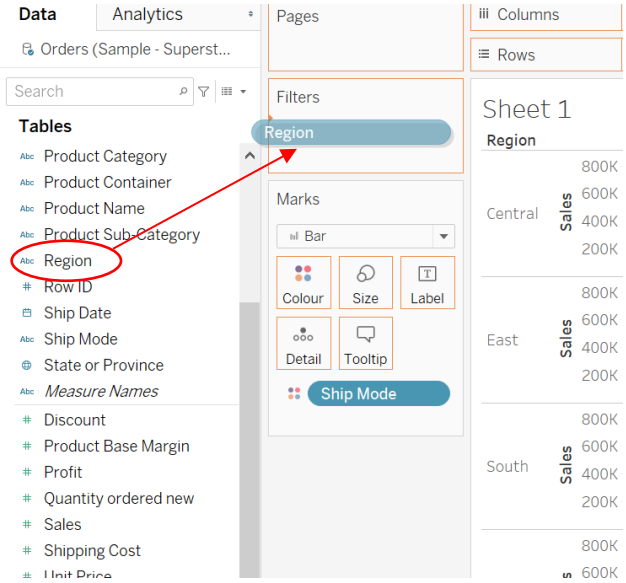
4. Drag **Ship Mode** to **Colour** on the Markscard.

   The view shows how different shipping modes have contributed to total sales over time. The ratios seem consistent from year to year:



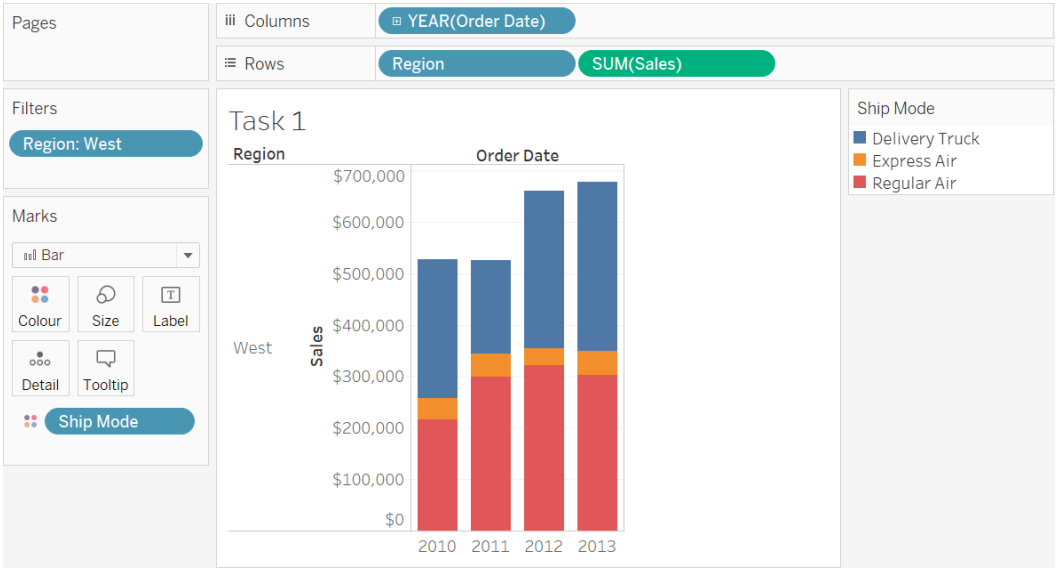
5. Drag **Region** to the **Rows** shelf and drop it to the left of **Sales** to produce multiple axes for sales by region.

6. If you want to view data in the West region only, you can filter out the other regions, either temporarily or permanently. To do this, drag the **Region** from the dimension pane to the **Filters** shelf.



7. This opens the Filter [Region] dialog box. Clear the **Central**, **East**, and **South** check boxes, and then click **OK**:

Your final view should look like this:



This view gives you enough information to answer some questions about your data— for example, how ship mode has changed in the West over the four-year period. But answers only lead to more questions, and you can go on developing the view for as long as you like.

Task 2: Creating a Grouped Bar Chart

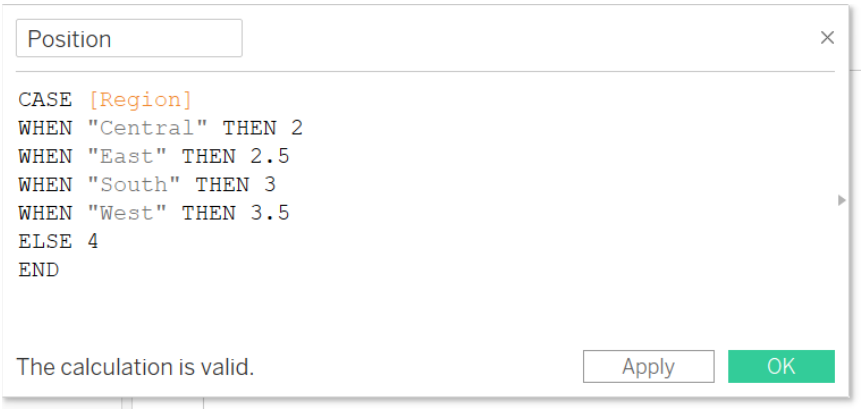
A grouped bar chart shows comparisons vividly, in a way that can be taken in at a glance.

1. Open a new sheet and rename it as **Task 2**.

2. Create a new **calculated field**. A calculated field provides the environment you need to create this view. For a grouped bar chart, you need a **continuous field** that anchors the horizontal location of each bar, forming the group.

    a. Select **Analysis → Create Calculated Field**.
    b. In the Calculated Field dialog box, complete the following steps.
        i. For Name, type Region **Position**.
        ii. In the Formula box, build the following formula:

```
CASE [Region]
WHEN "Central" THEN 2
WHEN "East" THEN 2.5
WHEN "South" THEN 3
WHEN "West" THEN 3.5
ELSE 4
END
```



    c. Confirm that the status message indicates that the formula is valid, and then click OK.

3. Drag **Product → Category** onto the **Columns** shelf.

4. Right click **Position** and select **Convert to Dimension**.

5. Right click **Position** and select **Convert to Continuous**.

6. Drag **Position** to the **Columns** shelf.

7. Drag **Sales** onto the **Rows** shelf.

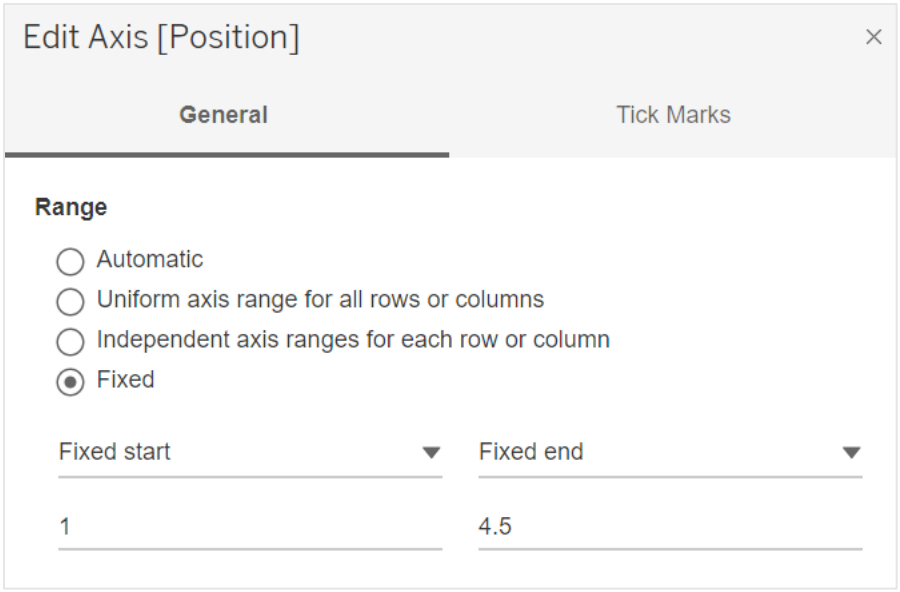
8. On the **Marks** card, on the drop-down list of **view types**, select **Bar**.

9. Drag **Region** to **Colour** on the **Marks** card.

10. On the **Marks** card, click **Size**, and then drag the slider all the way to the right to widen the bars.

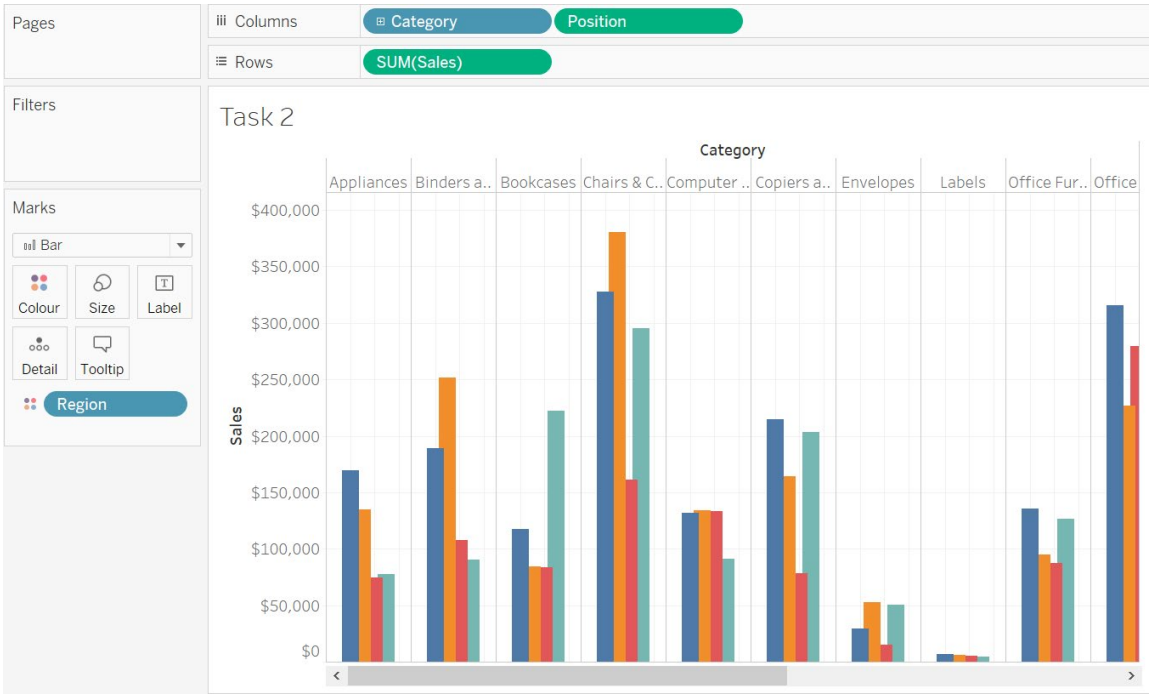
11. Right-click the x-axis and select **Edit Axis**.

12. In the Edit Axis dialog box, make the following changes:
    a. Under **Range**, select **Fixed**.
    b. Set **Fixed Start** to **1** and **Fixed End** to **4.5**.



    c. Click **OK**.

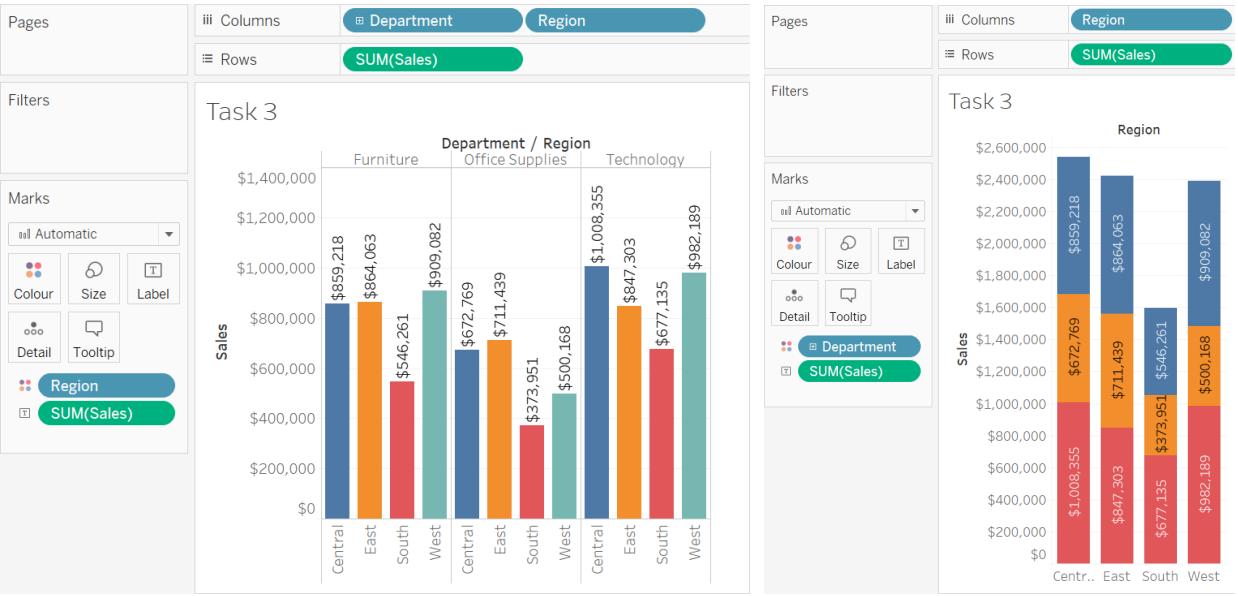
13. Right-click the x-axis and click **Show Header** to clear the selection.



Adapted from http://kb.tableausoftware.com/articles/knowledgebase/grouped-bar-chart

Task 3: Adding Totals to Stacked Bars

Adding totals to the tops of bars in a chart is sometimes as simple as adding the value to the Label Marks. However, when the bars are broken down by colour or size, each segment is labelled, rather than the total for the bar as shown below. So, in this task, we will learn the steps to add the totals to the stacked bars.



1. Open a new sheet and rename it as **Task 3**.

2. Drag the **Product → Department** and the **Region** to the **Columns** shelf.

3. Drag **Sales** to the **Rows** shelf. Drag **Region** to the **Colour** Marks card.

4. Drag **Sales** to the **Label** Mark card. You should see the bar chart like above left image.

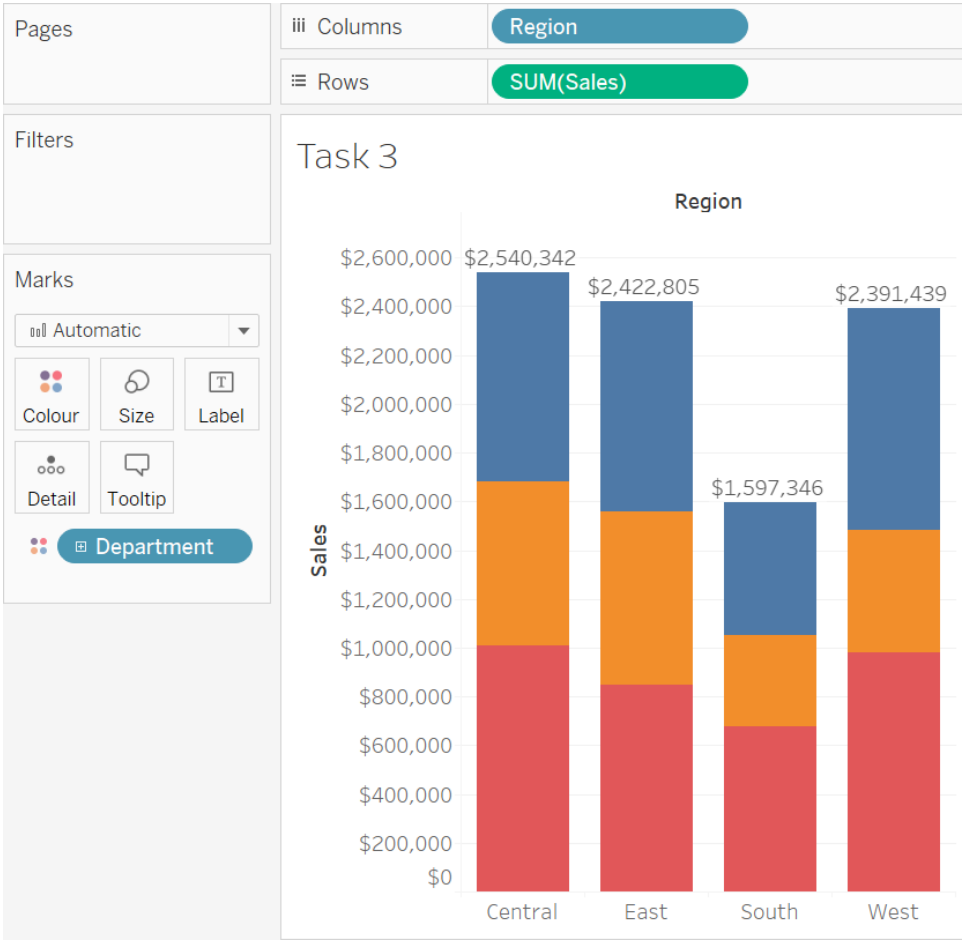
5. Under **Show Me**, change the chart to **Stacked Bar Chart**.

6. Drag **Sales** to the **Label** Marks card to display the sales amount in each segment. If the orientation of the label is unsuitable, you can change it by clicking on the Label shelf and select the appropriate Direction. You should see your chart like the above right one.

Next few steps, we shall add the total to the top of every stacked bar.

7. Add a Reference Line.
   You can use a reference line to automatically calculate the totals.

a. Right-click the y- axis and select **Add Reference Line**
b. In the **Add Reference Line** dialog box, select **Line** as the reference line type.
c. Under **Scope**, select **Per Cell**.
d. In the **Label** list, select **Value**.
e. Under **Formatting**, in the **Line** list, select **None**.
f. When finished, click **OK**.



8. Align the totals.

   The total is now displayed for each bar, but the numbers are left aligned.

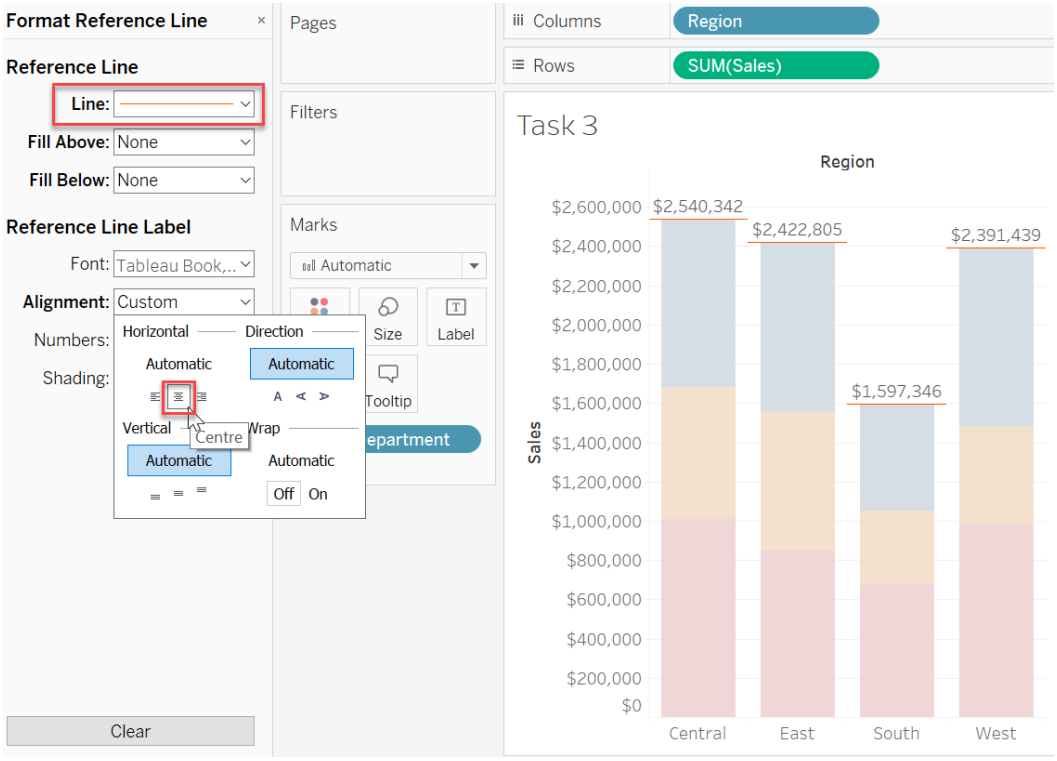
   a. Right-click any of the totals on the bar chart and select **Format**.

   b. In the **Format Reference Line** pane, under **Reference Line**, under **Line**, change the colour to **Red**.

   c. Under **Reference Line Label**, click the list arrow for **Alignment**.

   d. Under **Horizontal**, click the middle icon to center the total over the bar.

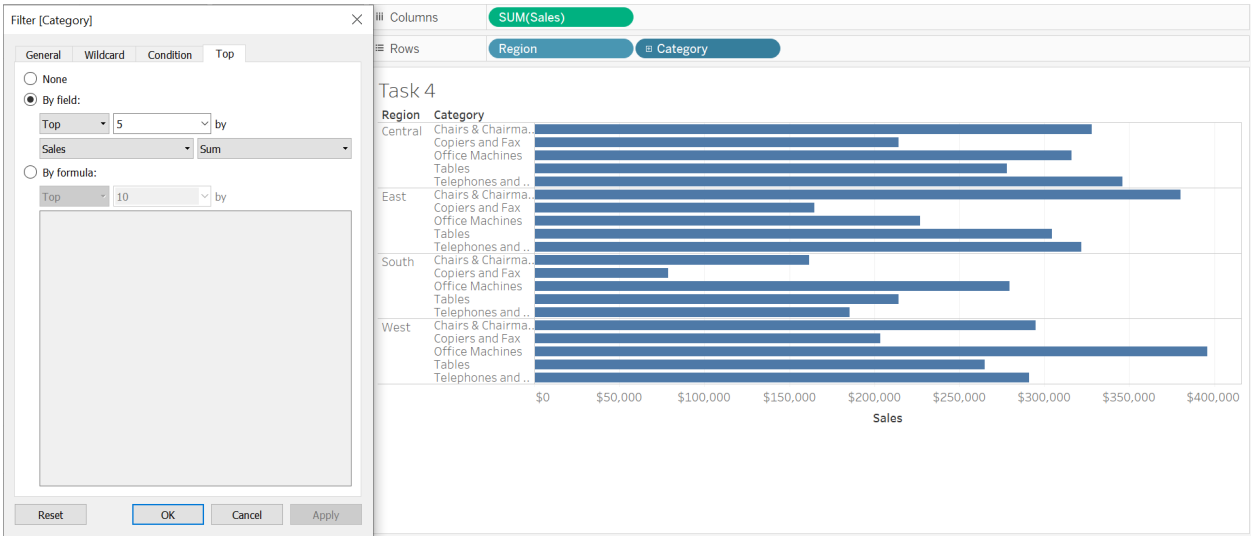
   e. The totals are now centered.



Adapted from http://kb.tableausoftware.com/articles/knowledgebase/add-totals-stacked-bars

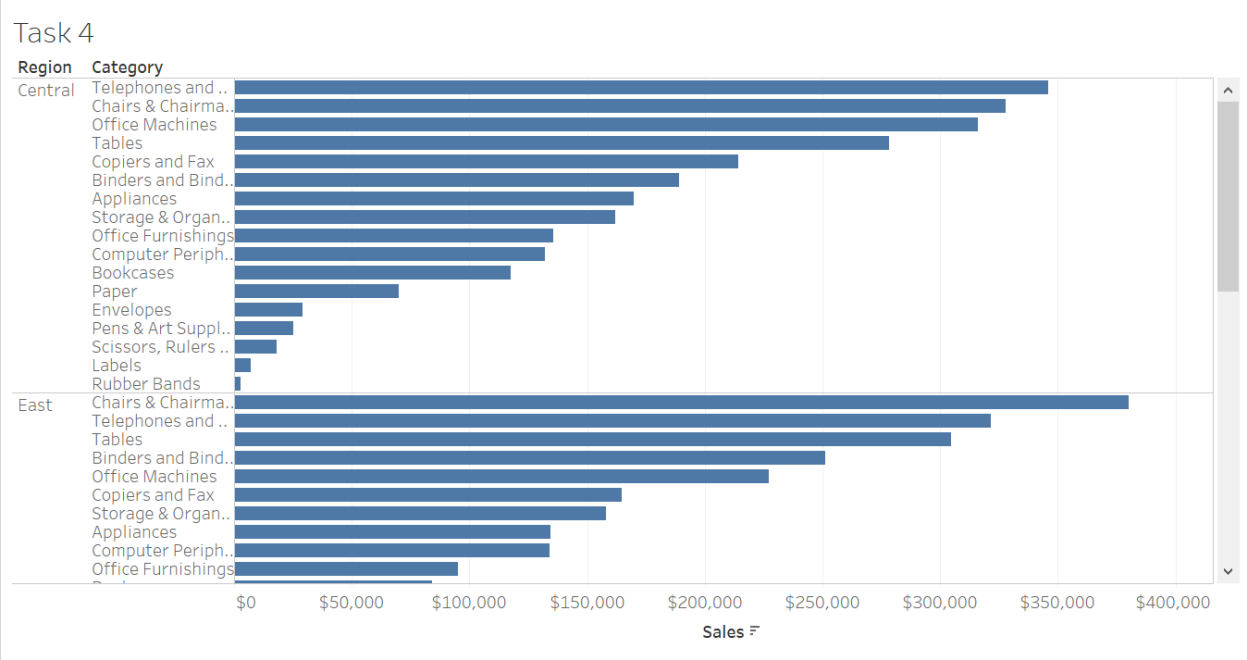
Task 4: Finding the Top N within a Category

The Top tab in the Filter dialog box allows you to define a filter to show the Top N items based on a specific measure. For example, you can use this type of filter to show the top 5 categories by sales in each region



This type of filter is computed across the entire data source. You can see that the ***same 5 categories*** are shown within each region because those are the top 5 selling categories sold across all regions. Instead of computing across the entire data source, you can calculate the **Top N within a category** by combining fields (i.e., creating a set) and then manually filtering the values instead.

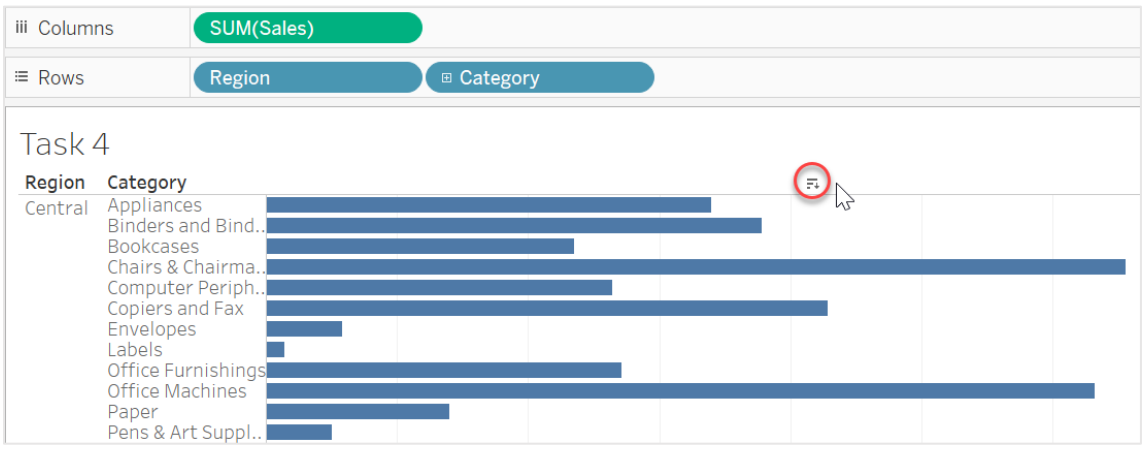
The following example shows you how to create a nested sort that looks like the following, and how to create a calculation that acts as a filter to compute the top 5 selling categories **within each region**.



1. Open a new sheet. Rename it as **Task 4**

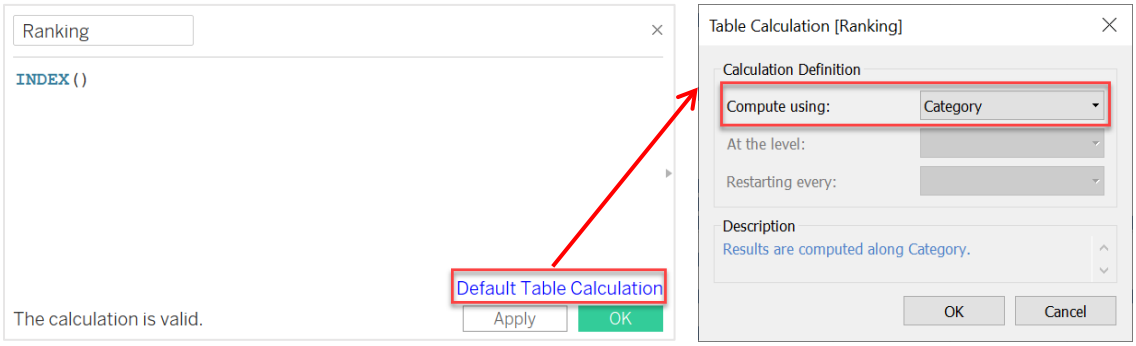
2. Drag the **Region** and **Product → Category** fields to **Rows** shelf and drag the **Sales** field to **Columns** shelf.

3. Move your cursor on top of the bar chart in the chart area to see the sort icon as shown in the red circle. Double click it to sort it in descending order.



4. Select **Analysis → Create Calculated Field**.
   In the Create Calculated Field dialog box, do the following:
   a. In the **Name** text box, type a name **Ranking** for the calculation.
   b. In the **Formula** text box, type the following: **INDEX()**
   c. Click **Default Table Calculation** at the bottom right corner of the dialog box.
   d. In the Table Calculation dialog, select **Category** from the "Compute using:" dropdown list.

   When finished, click **OK**.



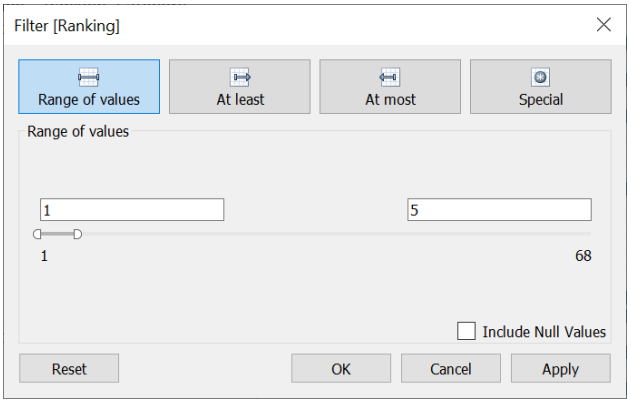
5. Right click the **Ranking** field, and select **Convert to Discrete**.

6. Place the **Ranking** field onto the **Rows** shelf, in between the **Region** and **Category** fields.

7. Drag **Ranking** to the **Detail** Marks card.

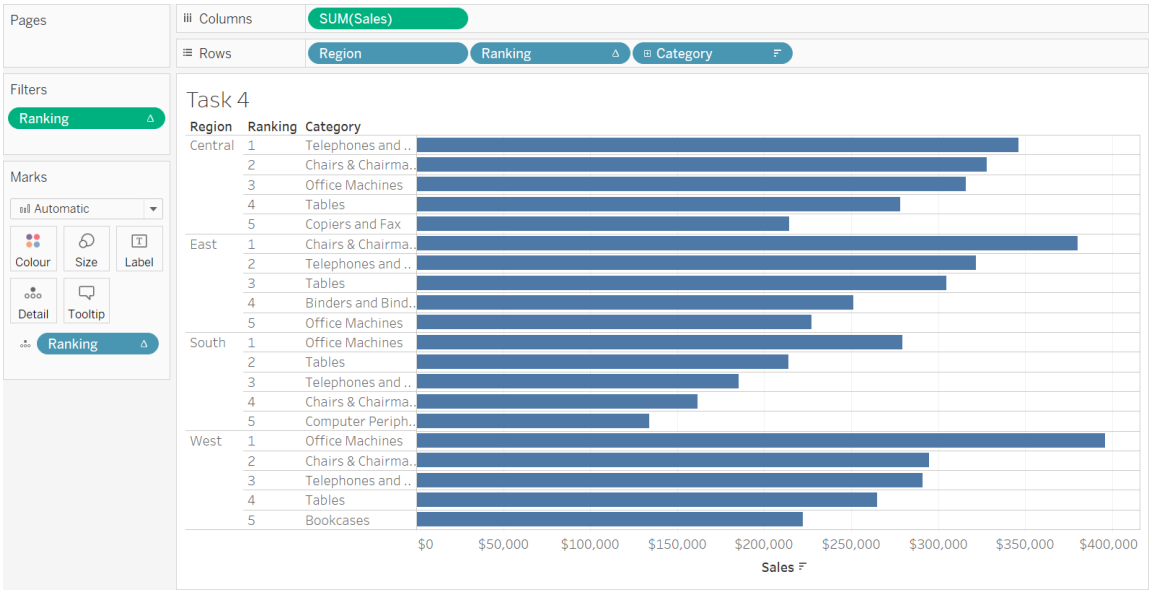
8. Right click the **Ranking** field, select **Continuous**, and move the field to the **Filter** shelf.

9. In the Filter dialog box, type **1** in the first text box and **5** in the second text box, and then click **OK**.



10. In the Fit box, select **Entire View**.

Now you can see the top 5 selling categories within each region.



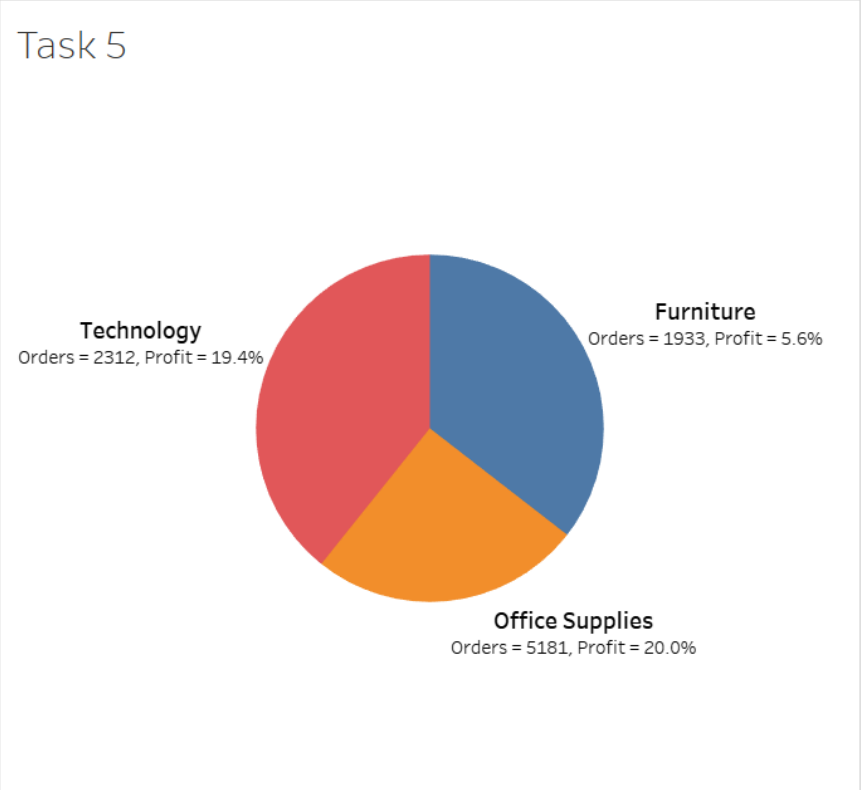
Notes:

You can hide the values in the Ranking field by right clicking the field in the Rows shelf and select Show Header to remove the selection.

Task 5: Adding Multiple Labels to a Pie Chart

Labels help call out each mark and make it more understandable. When building a pie chart, there are certain scenarios that may require more than one field to be identified in the label. For one single field, you can easily add labels to your pie marks by dragging the desired field to **Label**.

However, you can show multiple labels on a mark by creating a custom calculation.



1. Open a new sheet and rename it as **Task 5**.

2. Holding the CTRL key down, select the **Department** and **Sales** fields in the Data window, and then select the **Pie Chart** option using Show Me.

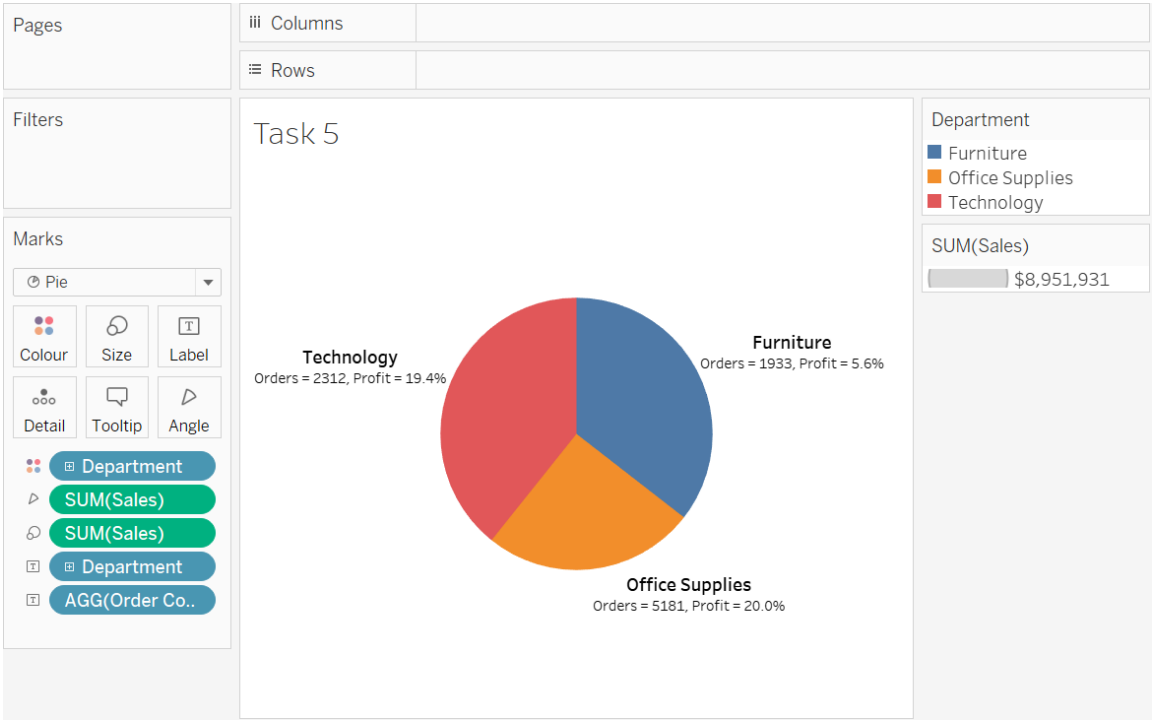
3. Drag the **Department** field to the **Label** shelf on the **Marks** card.

4. Create a calculated field named **Gross Margin %**
   Select **Analysis → Create Calculated Field**. In the Create Calculated Field dialog box, do the following tasks:
   a. In the Name text box, type **Gross Margin %**.
   b. In the Formula text box, type the formula below, and then click **OK**.
   SUM([Profit])/SUM([Sales])

5. Create Calculated Field named **Order Count & Profit %**.
   Select **Analysis → Create Calculated Field**. In the Create Calculated Field dialog box, do the following tasks:
   a. In the Name text box, type **Order Count & Profit %**.
   b. In the Formula text box, type the formula below, and then click **OK**.
   "Orders = " + STR(COUNT([Order ID]))+", Profit = " +(STR(ROUND([Gross Margin %]*100,1))+ "%")

6. Drag the calculated field **Order Count & Profit %** to the **Label** shelf on the **Marks** card.

7. In the Fit box, select **Entire View**. You can click on the **Label** shelf to format the label.



Adapted from http://kb.tableausoftware.com/articles/knowledgebase/adding-multiple-labels-to-pie-chart

Task 6: Using Pareto Analysis

The Pareto principle states that, for many events, roughly 80% of results come from 20% of the causes. For example, 80% of profits come from 20% of the products offered. In Tableau, you can apply a table calculation to sales data to create a chart that shows the percentage of total sales that come from the top products. You can also see which products account for 80% of total profit.

Preparing for the analysis

Before starting your analysis, decide what questions you want answered. These questions determine the category (dimension) and number (measure) on which to base the analysis. Below are two common questions about sales data, along with the fields that would need to be analyzed to answer each question.

Q: How is the profit distributed among the different products?

- **Dimension:** Product Name
- **Measure:** Profit

Q: How are sales distributed among sales representatives?

- **Dimension:** Sales Representative
- **Measure:** Sales

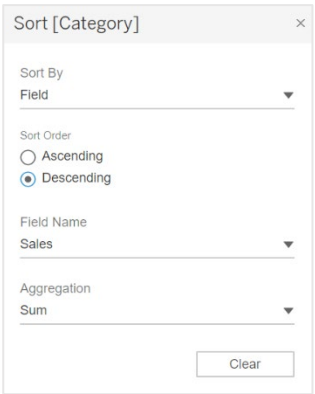
The following steps use the Superstore Subset (Excel) sample to show how to create a view that answers the first question.

1. Open a new sheet and rename it as **Task 6**.

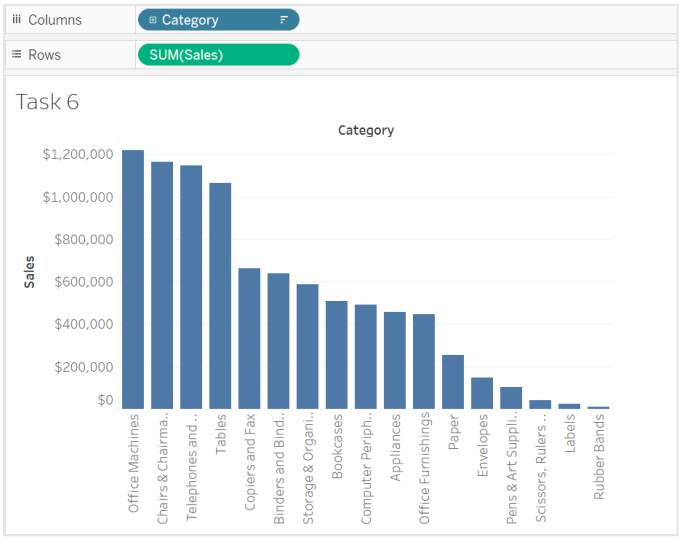
2. Drag the following fields to the view:
   a. Under the Product group, drag **Category** to the **Columns** shelf.
   b. Drag **Sales** to the **Rows** shelf.

3. On the **Columns** shelf, from the **Category** drop-down menu, select **Sort**.

4. In the **Sort** dialog box, make the following selections:
   a. Under **Sort order**, select **Descending**.
   b. Under **Sort by**, select **Field**, and leave **Sales** as the selected field.
   c. For **Aggregation**, leave Sum selected.
   d. Close the dialog.



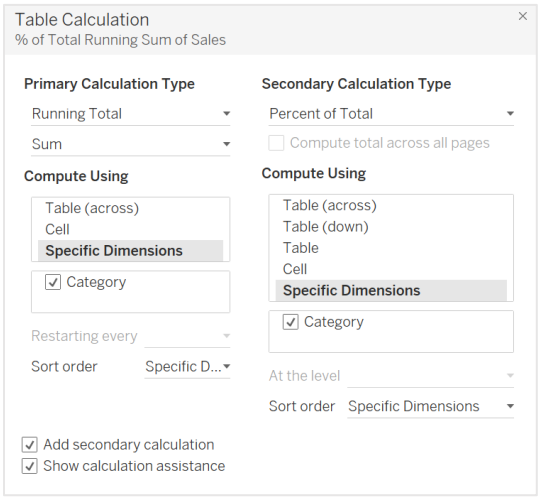
In the view, the products are sorted from highest sales revenue to lowest.



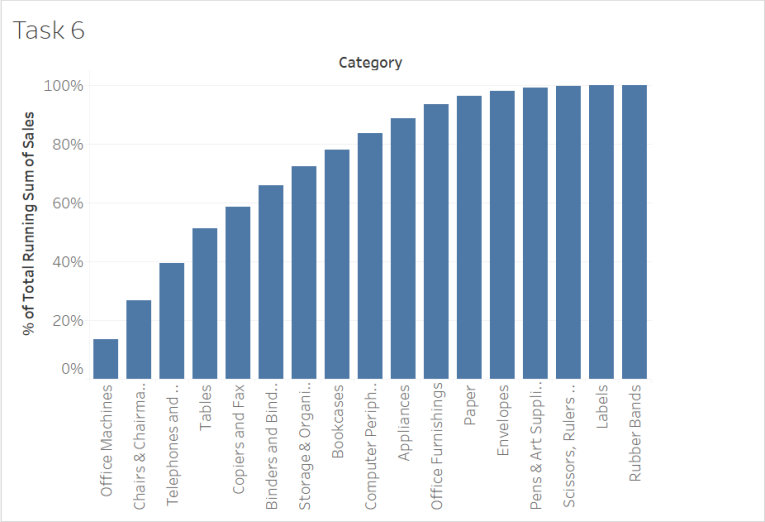
Next, create a running total of sales and normalize the cumulative total to add up to 100%.

5. On the **Rows** shelf, from the **SUM(Sales)** drop-down menu, select **Add Table Calculation**.

6. In the **Table Calculation** dialog box, complete the following steps:
   a. In the **Calculation Type** list, select **Running Total**.
   b. In the **Compute Using** list, select **Specific Dimensions**. Check **Category**.
   c. Select the **Add secondary calculation** check box.
   d. In the **Secondary Type** list, select **Percent of Total**.
   e. In the **Compute Using** list, select **Specific Dimensions**. Check **Category**.
   f. Close the dialog.



The view reflects a standard Pareto curve.

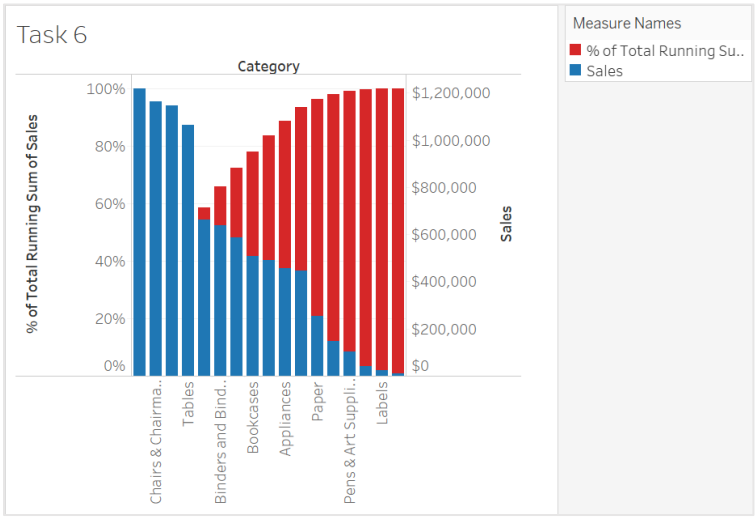


Compare the running total percentage with actual sales

Now, we add in a dual axis that shows the actual sales alongside the running total percentage of sales by complete the following additional steps.

7. Drag another instance of **Sales** to the right side of the view until a vertical dotted line appears, and then release the mouse button to create a dual axis. Tableau might change the view type at this point.

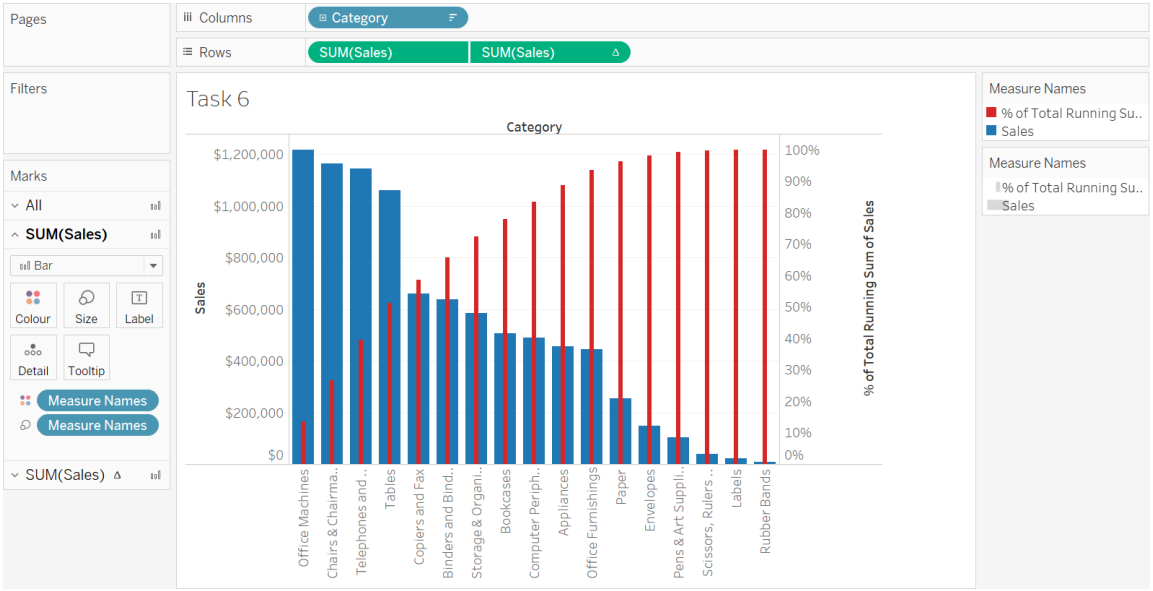
8. On the **Marks** card titled **All**, select **Bar** from the drop-down list of views. Edit the Measure Names colour.



9. Hold the Ctrl key, and on the **Rows** shelf, select both measures. While both measures are selected, right-click one of them and select **Swap**.

   Sales and Percentage of Running Total of Sales swap positions.

10. Drag another instance of **Measure Names** to **Size**.



Additional information

For additional tips on how you would compare the percentage of sales with the percentage of products, or draw reference lines that help make the 80-20 principles more apparent, see the Tableau On-Demand Training video Pareto.