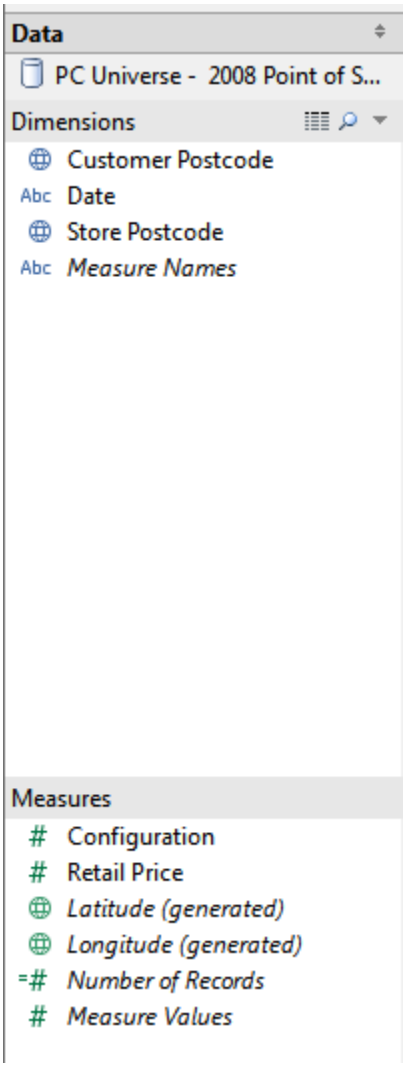
Assignment 1 - Loading data, merging data, creating visualizations, getting oriented



Practice the following actions:

Load data in Tableau: Loading data in Tableau can be done in two different ways: 1. Data>Connect to Data, or 2. Copy/Paste the data

Once the data are opened in Tableau, the software will try to guess the data types for each variable. Each column will be mapped as either a **Dimension** or **Measure**. Measures are columns with numerical values. Dimensions are columns with categorical values (text, dates). See more here.

Each column is also automatically classified as a particular data type (Number/Date/Date & Time/String).

- **Load "PC Universe - 2008 Point of Sale Data.txt".**

Here is the data dictionary for these data.

Tableau classifies *Configuration* incorrectly as a Measure of Data Type=Number. This will cause problems in creating appropriate visualizations using Configuration. Correct this by right-clicking and choosing the appropriate options.

Tableau also incorrectly classifies Date as a String. Look at the original Date column in the data (you can click the View Data icon to the right of Dimensions) and modify the Date type accordingly in Tableau.

Take a look at Tableau's toolbar. The undo and redo buttons are very useful for exploration!

The easiest way to create a visualization is to select (highlight) the variables of interest and then click the Show Me button. You will see all the charts that are "legal" to use for the types of variables that you selected.

Note the variables "Measure Names", "Number of Records" and "Measure Values" that can be used in visualizations just like other columns.

You should submit all your answers in homework submission file. PLUS copy one of your answers for Task 1 to this week's discussion board (in the thread "Task 1: Post Here").

Complete the following three tasks (fourth task optional):

Task 1: Create a bar chart showing the volume of transactions by Configuration.

Step 1: Create the chart

- Choose Configuration and Number of Records; Click Show Me; Click Bar chart
- Note how Configuration and Number of Records now appear in the Row and Column shelves

Step 2: Interact with this bar chart:

- Change axes on the chart: Instead of bars showing # of records, show %. Right-click on Sum(Number of Records) and in Quick Table Calculation choose Percent of Total.
- To better compare amounts sold of each configuration, it is useful to sort the bars by length/height: hover with your mouse near the axis label and you will see a small sort icon appear. Click once to sort ascending, twice for descending, and again for original sequence.
- To compare this bar chart across different stores, you can add Store Postcode. Double-click on Store Postcode. It will be added to the Columns or Rows area (try both!) – look at the result.
- To view more information about a particular bar, hover over it and you'll see a tooltip appear. In addition to the information displayed in the tooltip, there are also a few operations that you can perform from the tooltip such as filtering, filtering-out, viewing the data, etc.

Step 3: Export/share a visualization from Tableau

After creating a visualization in Tableau, we can save or export it in two modes: static or interactive.

- For a static screenshot, you can use Worksheet>Export Image in Tableau, or simply use PrintScreen or Windows Snipping Tool.
- For an interactive version, save the file as a Tableau packaged workbook (.twbx), which includes also the data. You can then send this file to anyone who has Tableau desktop or the free Tableau reader.
- If your data are not confidential, you can share your visualization to the Web so that anyone can interact with it from any browser (no need for Tableau software). You can even embed that visualization into a blog post or your website. To do this, go to Server > Tableau Public > Save to Web. (you'll have to open a free account on Tableau Public)

Share on the discussion board one of your bar chart variations, with a sentence describing one insight that can be seen from the chart. You can either post a static screenshot or if you've posted it to the Web, share the link.

Task 2: Merge the laptop configuration information with the sales data.

You have already opened the file "PC Universe - 2008 Point of Sale Data.txt." Now, merge the laptop configuration data by adding a table and linking it.

- In the Data menu, click on **PC Universe - 2008 Point of Sale Data.txt** and choose Edit Tables.
- Click **Add Table** and choose Laptops.txt.
- Click the **Join** tab to make sure the two tables are joined by Configuration (it should use Join Type=left so that the laptop details from Laptops.txt are added to each row in PC Universe - 2008 Point of Sale Data.txt.

Submit a screenshot of the data area (the left part of the screen) from your Tableau file. You can use Window's Snipping Tools or PrintScreen.

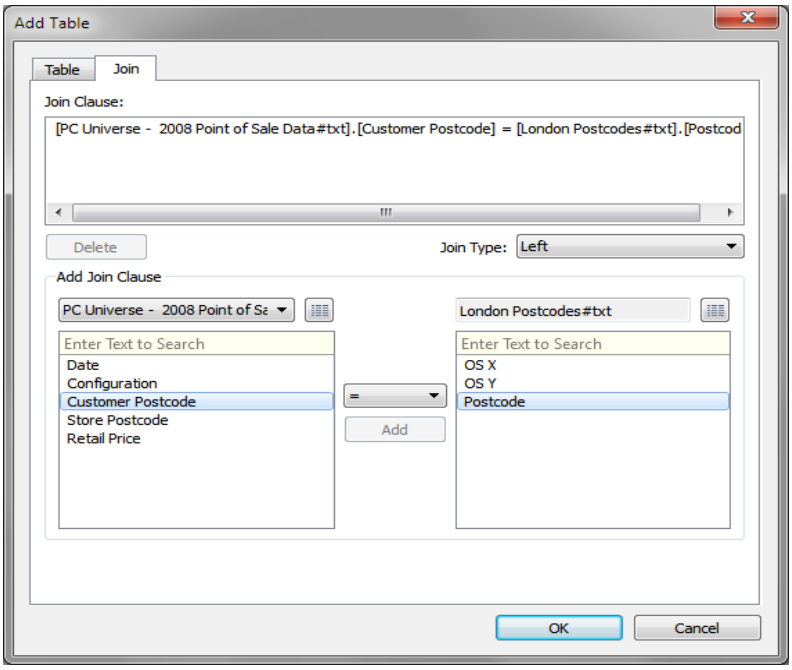
Task 3: Visualize buyers and store locations

Step 1: Merge the needed data

Add the third table from "London Postcodes.txt", which contains a mapping between a postcode and the corresponding X and Y coordinates. Our goal is to add columns for the X and Y coordinates of the customer location, and then for the store locations.

Hint: First merge the London Postcodes by joining according to the Customer Postcode (see screenshot below). Then, rename the resulting two coordinate columns (OS X and OS Y) "X Customer" and "Y Customer". Now repeat this entire process, this time joining by Store Postcode.

Make sure the new coordinate columns are set as **dimensions**, **continuous** and don't have a geographical role.



Step 2: Create a new worksheet (Worksheet > New Worksheet). Note that in Tableau, each visualization is created on a separate worksheet. You can also name your sheets by right-clicking on the sheet name (at the bottom) and choosing Rename Sheet.

Step 3: Create a scatterplot showing the locations of the customers, using the Customer X and Y coordinate columns.

- Choose the two columns and click on Show Me. Choose the scatter plot option. If the scatter plot option is unavailable, your chosen columns are not properly set (remember that they must be continuous dimensions with no geographical role).
- Make sure the Customer X column is on the x-axis. If the axes are flipped, switch them around.
- We want each marker to denote a customer postcode. To do this, drag Customer Postcode into the Detail shelf.
- Now color-code the markers in the scatterplot by store (showing the store in which each of the customers purchased the computer). This is done by dragging the Store Postcode to the Color shelf.

Submit a screenshot of your Tableau worksheet and the Data area, with a sentence about what you learned from this visualization.

Task 4 (optional): Advantages of interactive software over ordinary tools

Excel is by far the most popular tool for generating plots. Why not stick with it for data exploration? To see the advantages of using an interactive visualization tool, compare the process needed to create a color-coded scatterplot in Tableau and in Excel.

Try to create a similar scatterplot to the one from Task 3, this time using Excel

- You can either try to do all the data merging in Excel (if you have the knowledge), or else, you can export the merged data from Tableau (in the worksheet with the scatter plot, choose Worksheet > Copy > Data; see Export Data in Tableau's help for other options)
- To see how complicated things can get if you stick to Excel graphics, see Galit Shmueli's blog posts "Creating color-coded scatterplots in Excel: a nightmare" and "How to create a histogram in Excel".