BUSA 603

Module 3 Supplement
Data Visualization & Some Useful Charts

Brian Weikel

brian.weikel@franklin.edu

FRANKLIN
UNIVERSITY
Makes It Possible

Business Analytics
Franklin University

Spring 2024

Know Your Audience

The first step in trying to interpret data is often to visualize it in some way. Data visualization can be as simple as creating a summary table, or it could require generating charts to help interpret, analyze, and learn from the data.

▶ Data visualization is very helpful for identifying data errors and for reducing the size of your data set by highlighting important relationships and trends.

▶ Data visualization is also important in conveying your analysis to others. Although business analytics is about making better decisions, in many cases, the ultimate decision maker is not the person who analyzes the data.

▶ When preparing for presentations that summarize data, always know your audience. This will ensure that you are conveying the proper content that does not "insult the intelligence" of the audience.

Determining If a Table or Chart is More Effective

The first decision in displaying data is whether a table or a chart will be more effective. In general, charts can often convey information faster and easier to readers, but in some cases a table is more appropriate. Tables should be used when:

▶ The reader needs to refer to specific numerical values.

▶ The reader needs to make precise comparisons between different values and not just relative comparisons.

▶ The values being displayed have different units or very different magnitudes.

Data-Ink Ratio

Tufte (2001) recommends using the **data-ink ratio** idea when creating effective designs.

1. The data-ink ratio measures the proportion of what Tufte terms "data-ink" to the total amount of ink used in a table or chart.

   ▶ **Data-ink** is the ink used in a table or chart that is necessary to convey the meaning of the data to the audience.

   ▶ **Non-data-ink** is ink used in a table or chart that serves no useful purpose in conveying the data to the audience.

2. This idea implies one should avoid using vertical lines in a table unless they are necessary for clarity.

Horizontal and Vertical Line Recommendations

3. Horizontal lines are generally necessary only for separating column titles from data values or when indicating that a calculation has taken place.

4. In large tables, vertical lines or light shading can be useful to help the reader differentiate the columns and rows.

5. Since there may be audience members who are color-blind, use colors only when other data and text distinguishing options – mark type, line size, line type, italic font, bold font . . . – have been exhausted.

Graph Recommendations for Effective Presentations

Anderson (2013), Lortie (2017), and Naegle (2021), to name just a few, give advice on creating presentations with visualizations.

- ▶ Build presentation slides around good visualizations. Remember, a picture is worth a thousand words. However, on the flip side, do not muddy the point of the slide by putting too many complex graphics on a single slide.

- ▶ Thus, use simple visuals.

- ▶ Repeat critical messages (at least) twice using different visuals.

- ▶ Use the principle of parsimony in explanations of visualizations.

A Few Chart Types

Scatter Chart A chart that has two value axes: a horizontal (x) and a vertical (y) value axis. It combines x and y values into single data points and shows them in irregular intervals, or clusters. These charts are typically used for showing and comparing numeric values, like scientific, statistical, and engineering data.[1]

Line Chart A chart where category data is distributed evenly along the horizontal axis, and all value data is distributed evenly along the vertical axis. This chart can show continuous data over time on an evenly scaled axis, so they are ideal for showing trends in data at equal intervals, like months, quarters, or fiscal years.

[1]This Microsoft page has information about chart types available in their software solutions.

FRANKLIN
UNIVERSITY
Makes It Possible

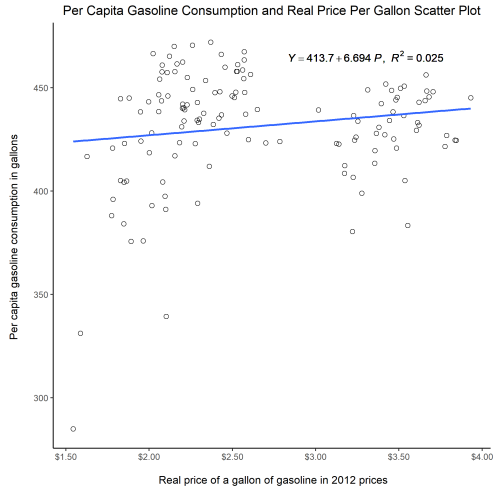Figure 1: A Petroleum Distillate Demand and Real Price Scatter Chart
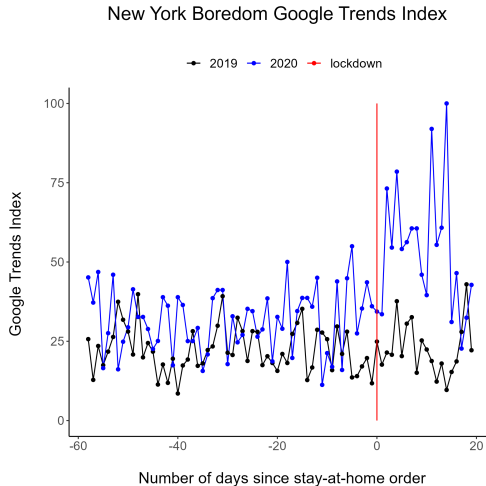
Figure 2: A Line Chart with Google Trends Data

Column, Bar, Pie and Bubble Charts

Column Chart Data that is arranged in columns or rows on a worksheet can be plotted in this type of chart. This chart display categories along the horizontal (category) axis and values along the vertical (value) axis.

Bar Chart A chart that is a column chart where the categories are organized along the vertical axis, and the values along the horizontal axis.

Pie Chart A chart that shows the size of items in one data series, proportional to the sum of the items. The data points in a such a chart are shown as a percentage of the entire area.

Bubble Chart A scatter chart where a third column is included that specifies the size of the bubbles/circles.

Parallel-Coordinates Plot and Treemap

Parallel-Coordinates Plot This maps each row in a data table as a line, or profile. Each attribute of a row is represented by a point on the line. While these plots are similar in appearance to line charts, the way data is translated into a plot is substantially different.

Treemap This chart displays categories by color and proximity and can easily show lots of data which would be difficult with other chart types. The chart can be plotted when empty (blank) cells exist within the hierarchical structure. This chart may be used to compare proportions within the hierarchy.

Scatter Plot Matrix, Heat Map and Sparkline

Scatter Plot Matrix This illustration is a grid (or matrix) of different charts, usually scatter charts, used to visualize bivariate relationships between combinations of variables. Each scatter plot in the matrix visualizes the relationship between a pair of variables, allowing many relationships to be explored in one chart.

Heat Map A table that depicts values for a main variable of interest across two axis variables as a grid of colored squares.

Sparkline A very small line chart, typically drawn without axes or coordinates. It presents the general shape of the variation (typically over time) in some measurement, such as temperature, stock market price, sales, profits, . . . in a simple and highly condensed way.

Figure 3: A Scatter Matrix
Bivariate Scatter Plots and Pearson Correlation Coefficients
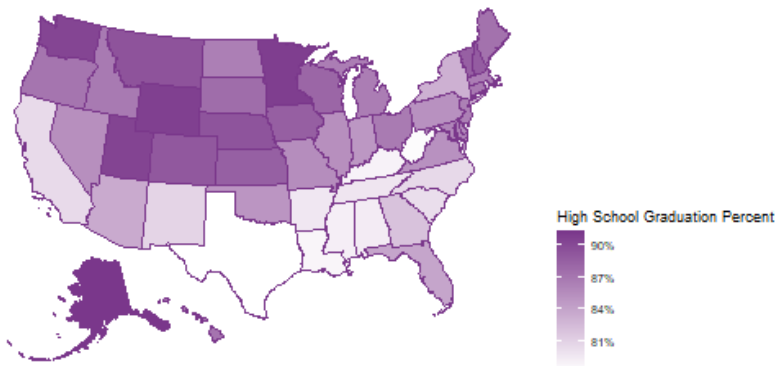
A Geographic Heat Map

Figure 4: A Geographic Heat Map

The Excel file `daily_store_sales.xlsx` contains 2013 and 2014 daily store gross value sales in Euros, daily store customers, promotional information, store descriptions and other variables for the retailer Dirk Rossmann GmbH, commonly referred to as Rossmann.[2]

The retailer is one of the largest drug store chains in Europe with more than 45,000 employees and more than 4,000 stores in Germany, Poland, Hungary, Czechia, Turkey, Albania, Kosovo, and Spain.

▶ The tab `ReadMe_First` of the Excel file has **metadata**, a set of data that describes and gives information about other data, for the tab `daily_store_sales`.

[2]With the usual caveats about Wikipedia, here is the Wikipedia English language page about Rossmann. The comma delimited file `daily_store_sales.csv` has the same data as the second tab of the aforementioned Excel file.

- Excel's `Pivot Table` functionality may be used to summarize the data, where the resulting data may be used to create charts.

Suppose the data is maintained in marketing staging and production databases.[3] These databases are different than the SAP financial database used by the Rossmann finance department.

At a meeting yesterday afternoon the CMO was discussing some sales charts and tables with the CFO. Upon a careful review of these charts and tables, the CFO told the CMO that the data he presented may not be correct. As you may expect, the CMO was not particularly happy with this revelation.

---

[3]A staging database is a parallel data warehouse database that stores data temporarily while it is loaded into the appliance that maintains the production database. The primary benefit of a staging database is to reduce table fragmentation, and hence results in faster queries.

The CMO has requested your and your fellow marketing department colleagues to investigate the CFO's claim.

▶ Specifically, he has asked you to populate the charts on the next 5 slides using daily_store_sales. The charts and their input data will be compared and contrasted to like information begin obtained by the Rossmann Information Technology department. If you notice any anomalies in the data, If you notice any anomalies in the data, or insightful marketing analytic insigts, you are to mention it.

▶ While you and your colleagues may use Excel, Python, R, . . . to create the charts, the data used to create the charts must be made available in either Excel or Google Sheets.

▶ Time is of the essence since in a few weeks the marketing database will be undergoing its annual lockdown development period.[4]

---

[4]The marketing production and staging databases will not have any table, view, production queries . . . , revisions created, tested, and deployed. The lockdown ensures resources are readily available to meet end of fiscal year reporting needs.

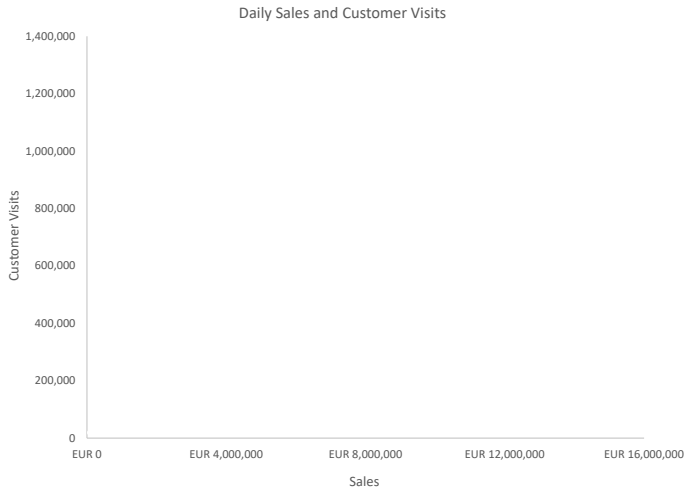# Total Sales and Total Customer Visits Scatter Plot

Figure 5: Daily Sales and Customer Visits Scatter Plot Template

Monthly Sales and Customers for 2013 and 2014

Figure 6: Monthly Sales and Customer Visits for 2013 and 2014 Time Series Chart Template

FRANKLIN
UNIVERSITY
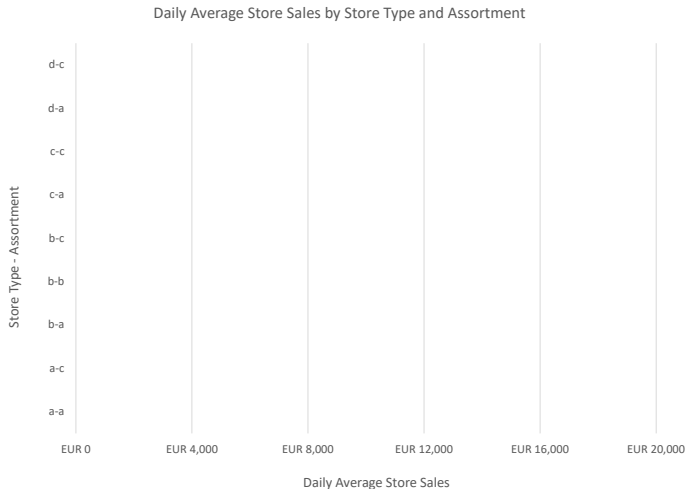*Makes It Possible*

Daily Average Store Sales by Store Type and Assortment

Figure 7: Daily Average Store Sales by Store Type and Assortment Bar Chart Template

Store Type and Assortment Average Daily Store Sales, Average Daily Customer Visits, and Standard Deviation of Daily Store Sales

Figure 8: Store Type and Assortment Average Daily Store Sales, Average Daily Customer Visits, and Standard Deviation of Daily Store Sales Bubble Chart Template

The standard deviation of daily store sales are used for bubble size.

Store Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Sparkline
a													
b	For Conditional Formatting of the cells, use the Green-Yellow-												
c	Red option in Excel.												
d													

Table 1: Store Type Monthly Sales Percentage Change Heat Map
Template

Creating the CMO Requested Charts

The following steps may be used to create Table 5 in Excel.

1. Highlight the cells of the `daily_store_sales` tab (i.e., worksheet) by clicking on the `Green Triangle` left of column `A` and above row 1.

2. Choose `Pivot Table` under `Insert` of the ribbon. Select `Table/Range`. For Table/Range of the pop-up box, enter `daily_store_sales!$A$1:$S$679630`. Ensure the `New Worksheet` button is selected. Click `OK`.

3. A new tab opens. Click in the empty pivot table. This invokes the `PivotTable Fields` pane.

4. Select `Date` for `Rows`.[5] Ensure only the variable `Date` appears under `Rows` (i.e., no entry for `Years`).

---

[5]A drag and drop of `Date` to `Rows` may be required.

B. Weikel, Franklin University

5. Select `Sales` for Σ `Values`. Ensure `Sum` is selected from the `Summarize value field by` options of `Value Field Settings`. For the field `Custom Name`, enter `Total Sales`.

6. Select `Customers` for Σ `Values`. Ensure `Sum` is selected from the `Summarize value field by` options of `Value Field Settings`. For the field `Custom Name`, enter `Total Customer Visits`.

7. Under `PivotTable Analyze` of the ribbon, select `Options`, which is in the far left box below the ribbon. Then select `Totals & Filters`, followed by deselecting `Show grand totals for rows` and deselecting `Show grand totals for columns`. Click `OK`.

8. Right click in any populated cell of Column A. Click on `Ungroup`.

9. Rename the sheet to Pivot_Table_1. While ideally we could use this table directly to create a scatter chart using Recommended Charts under Insert, Excel does <u>not</u> permit one to complete such a task.

   Thus click on ⊕ at the bottom of the workbook to create a new tab. Rename this sheet Pivot_Table_1_Scatter. Copy columns A through C of Pivot_Table_1 and paste them to Columns A through C of Pivot_Table_1_Scatter as Values using Paste Special. Highlight Column A, and then right click and select Format Cells. Under Category select Date, and under Type select *3/14/2012. Ensure Locale (location) is English (United States).

10. In Pivot_Table_1_Scatter rename Row Labels to Date.

11. In Pivot_Table_1_Scatter highlight column B, and then right click and select Format Cells. Under Category select Currency, and under Symbol select EUR. Ensure Decimal places is 0.

12. In Pivot_Table_1_Scatter highlight column C, and then right click and select Format Cells. Under Category select Number. Enure Decimal places is 0, and the Use 1000 Separator (,) is selected.

13. For all empty rows above the row with column names in Pivot_Table_1_Scatter, delete them.

14. Select columns B and C of Pivot_Table_1_Scatter by highlighting them. Under Insert of the ribbon, select Recommended Charts. Then select All Charts followed by X Y Scatter. If two illustrations appear, choose the one that is farthest right. Click OK.

15. Rename the chart's title to Daily Sales and Customer Visits.

16. Click on the chart and then click ⊞. Ensure the `Axis Titles` box is checked. For the y-axis place `Customer Visits` and for the x-axis place `Sales`.

17. Right click on chart and select `Move Chart`. Ensure the `New sheet` button is selected. Name the sheet `Scatter_Chart` .

18. Resize the axes, axis titles, and title to make the chart more visually appealing. You may also need to change the the major tick marks of each axis.
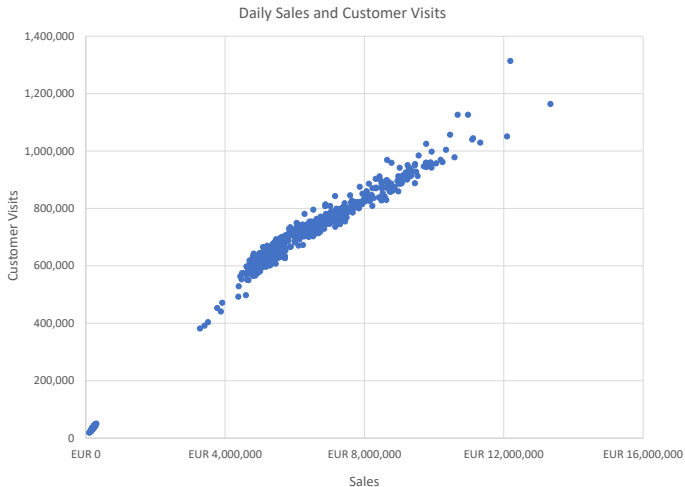
Figure 5: Daily Sales and Customer Visits Scatter Plot

The following steps may be used to create Table 6 in Excel. When using time series data, the chart is more frequently referred to as a **time series chart**.

1. Right click on Pivot_Table_1, and select Move or Copy. In the pop-up box, ensure To book is daily_store_sales.xlsx. For Before sheet, highlight (move to end). Ensure the Create a copy button is selected. Rename the tab to Pivot_Table_2.

2. In Pivot_Table_2, right click in column A. Select Group. Ensure Months and Years are selected. Click OK.

3. In the search box at the top of the workbook, type Report Layout. Select Show in Outline Form.

4. Under PivotTable Analyze of the ribbon, select Options, which is in the far left box below the ribbon. Then select Totals & Filters, followed by deselecting Show grand totals for rows and deselecting Show grand totals for columns.

5. Highlight the cell that contains 2013. Right click and deselect Subtotal ``Years''.

6. In the PivotTable Fields pane under Σ Values, left click Sum of Sales, and select Value Field Settings. For the field Custom Name, enter Total Sales.

7. In the PivotTable Fields pane under Σ Values, left click Sum of Customers, and select Value Field Settings. For the field Custom Name, enter Total Customer Visits.

8. Click on ⊕ at the bottom of the workbook to create a new tab. Rename this sheet Pivot_Table_2_Line_Chart. Copy columnsA through D of Pivot_Table_2 and paste them to Columns A through D of Pivot_Table_2_Line_Chart as Values using Paste Special. Delete the rows that has Years with a non-empty cell and Date with an empty cell. For all empty rows above the row with column names in Pivot_Table_2_Line_Chart, delete them.

9. In `Pivot_Table_2_Line_Chart` insert a column between column B and column C by first highlighting column C, followed by selecting `Insert`. In the new column in the row with column names, type `Month-Day`. In the cell directly below the row with column names, which we assume is row 2, type `=DATEVALUE(CONCATENATE(B2,"-",A2))`. Copy and paste this formula through row 25.

10. Highlight Column C of `Pivot_Table_2_Line_Chart`, and then right click and select `Format Cells`. Under `Category` select `Date`, and under `Type` select `Mar-12`. Ensure `Locale (location)` is `English (United States)`.

11. In `Pivot_Table_2_Line_Chart` highlight column D, and then right click and select `Format Cells`. Under `Category` select `Currency`, and under `Symbol` select `EUR`. Ensure `Decimal places` is 0.

12. In `Pivot_Table_2_Line_Chart` highlight column E, and then right click and select `Format Cells`. Under `Category` select `Number`. Enure `Decimal places` is 0, and the `Use 1000 Separator (,)` is

13. Inside the chart, left click on the time series `Total Customer Visits`. In the `Format Data Series` pane option `Series Options`, ensure the `Secondary Axis` button is selected.

14. Rename the chart's title to `Monthly Sales and Customer Visits for 2013 and 2014`.

15. Click on the chart and then click ⊞. Ensure the `Axis Titles` box is checked. For the x-axis place `Month and Year`, for the primary y-axis place `Sales`, and for the secondary y-axis place `Customer Visits`,

16. Right click on chart and select `Move Chart`. Ensure the `New sheet` button is selected. Name the sheet `Line_Chart` .

17. Resize the axes, axis titles, and title to make the chart more visually appealing. You may also need to change the the major tick marks of each axis.
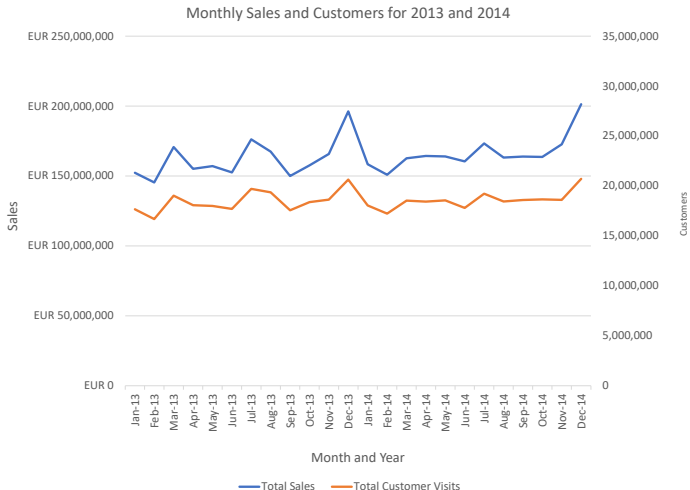
Figure 6: Monthly Sales and Customer Visits for 2013 and 2014 Time Series Chart

The following steps may be used to create Table 7 in Excel. The pivot table we create will also contain summary statistics for customers and school holidays, which will be used to create a bubble chart.

1. Highlight the cells of the daily_store_sales tab by clicking on the Green Triangle left of column A and above row 1.

2. Choose Pivot Table under Insert of the ribbon. Select Table/Range. For Table/Range of the pop-up box, enter daily_store_sales!$A$1:$S$679630. Ensure the New Worksheet button is selected. Click OK.

3. A new tab opens. Click in the empty pivot table. This invokes the PivotTable Fields pane.

4. Select StoreType for Rows. Also select Assortment for Rows.

5. Select `Sales` for Σ `Values`. Ensure `Average` is selected from the `Summarize value field by` options of `Value Field Settings`. For the field `Custom Name`, enter `Daily Average Store Sales`.

6. Select `Customers` for Σ `Values`. Ensure is selected from the `Summarize value field by` options of `Value Field Settings`. For the field `Custom Name`, enter `Daily Average Customer Store Visits`.

7. Select `Sales` for Σ `Values`. Ensure `StdDev` is selected from the `Summarize value field by` options of `Value Field Settings`. For the field `Custom Name`, enter `Standard Deviation of Daily Store Sales`.

8. Under `PivotTable Analyze` of the ribbon, select `Options`. Then select `Totals & Filters`, followed by deselecting `Show grand totals for rows` and deselecting `Show grand totals for columns`. Click `OK`. Rename the tab to `Pivot_Table_3`.

FRANKLIN
UNIVERSITY
Makes It Possible

9. In the search box at the top of the workbook, type `Report Layout`. Select `Show in Outline Form`. In the search box, type `Repeat All Label Items`.

10. Right click in a non-empty cell of the column `StoreType`. Deselect `Subtotal ''Store Type''`.

11. Click on $\bigoplus$ at the bottom of the workbook to create a new tab. Rename this sheet `Pivot_Table_3_Bar_Chart`. Copy columns `A` through `E` of `Pivot_Table_3` and paste them to Columns `A` through `E` of `Pivot_Table_3_Bar_Chart` as Values using `Paste Special`. Delete the rows that has `StoreType` with a non-empty cell and `Assortment` with an empty cell. For all empty rows above the row with column names in `Pivot_Table_3_Bar_Chart`, delete them.

12. In `Pivot_Table_3_Bar_Chart` insert a column between column B and column C by first highlighting column C, followed by selecting `Insert`. In the new column in the row with column names, type `Store Type-Assortment`. In the cell directly below the row with column names, which we assume is row 2, type `=CONCATENATE(A2,"-",B2)`. Copy and paste this formula through row 10.

13. In `Pivot_Table_3_Bar_Chart` highlight column D, and then right click and select `Format Cells`. Under `Category` select `Currency`, and under `Symbol` select `EUR`. Ensure `Decimal places` is 0.

14. In `Pivot_Table_3_Bar_Chart` highlight column E, and then right click and select `Format Cells`. Under `Category` select `Number`. Enure `Decimal places` is 0, and the `Use 1000 Separator (,)` is selected.

15. In `Pivot_Table_3_Bar_Chart` highlight column D, and then right click and select `Format Cells`. Under `Category` select `Currency`, and under `Symbol` select `EUR`. Ensure `Decimal places` is 0.

16. Select columns C and D of Pivot_Table_3_Bar_Chart by highlighting them. Under Insert of the ribbon, select Recommended Charts. Then select All Charts followed by Bar.[6] Click OK.

17. Rename the chart's title to Daily Average Store Sales by Store Type and Assortment.

18. Click on the chart and then click ⊞. Ensure the Axis Titles box is checked. For the x-axis place Daily Average Store Sales, and for the y-axis place Store Type - Assortment,

19. Right click on chart and select Move Chart. Ensure the New sheet button is selected. Name the sheet Bar_Chart .

20. Resize the axes, axis titles, and title to make the chart more visually appealing. You may also need to change the the major tick marks of each axis.

---

[6]Selecting Column instead of Bar will produce a column chart.
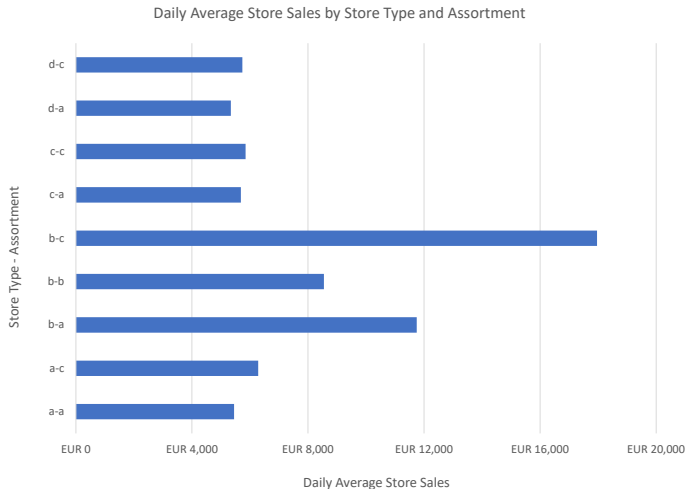
Daily Average Store Sales by Store Type and Assortment

Figure 7: Daily Average Store Sales by Store Type and Assortment Bar
Chart

We now will use Pivot_Table_3_Bar_Chart data to create Table 8.

1. Highlight cells D2 through F10 of Pivot_Table_3_Bar_Chart. Under Insert of the ribbon, select Recommended Charts. Then select All Charts followed by X Y (Scatter). For the chart options at the top of the pop-up box, scroll over the illustrations and select Bubble. Select the upper right-hand illustration. Click OK.

2. Rename the chart's title to Store Type and Assortment Average Daily Store Sales, Average Daily Customer Visits, and Standard Deviation of Daily Store Sales.[7]

3. Click on the chart and then click ⊞. Ensure the Axis Titles box is checked. For the x-axis place Daily Average Store Sales, and for the y-axis place Daily Average Customer Store Visits. Also ensure the Data Labels box is checked. Deselect Gridlines.

---

[7]If one were using this illustration in a presentation, a shorter title would be used.

4. Left click on any bubble, right click and select `Format Data Labels`. Check the `Value From Cells` box and in the pane that materializes, place `=Pivot_Table_3_Bar_Chart!$C$2:$C$10`. Click `OK`. Deselect `Y Value`. For the `Label Position` offer, choose `Above`,

5. Right click on chart and select `Move Chart`. Ensure the `New sheet` button is selected. Name the sheet `Bubble_Chart`

6. Resize the axes, axis titles, and title to make the chart more visually appealing. You may also need to change the the major tick marks of each axis. You may wish to change the colors of the bubbles, their transparency, and other characteristics. I recommend moving the labels so they are not overlapping.

Store Type and Assortment Average Daily Store Sales, Average Daily Customer Visits, and Standard Deviation of Daily Store Sales
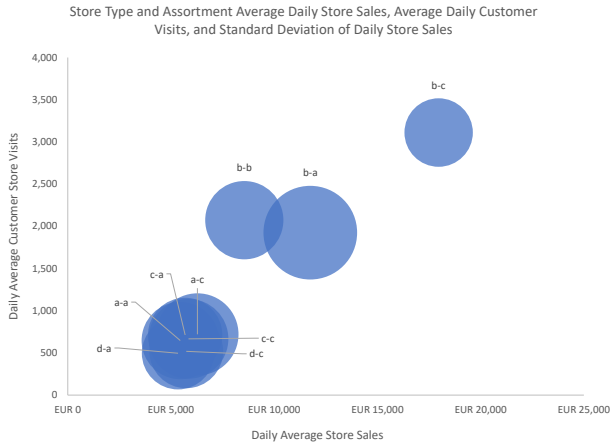
Figure 8: Store Type and Assortment Average Daily Store Sales, Average Daily Customer Visits, and Standard Deviation of Daily Store Sales Bubble Chart

We will use Pivot_Table_1 data to create a sales month-year-ago percent change heat map with sparklines. That is, the following steps may be used to create Table 9 in Excel.

1. Right click on the tab Pivot_Table_1 and select Move or Copy. Ensure To book is daily_store_sales.xlsx. Then check the box for Create a copy. Click OK. Rename this tab to Pivot_Table_4.

2. Left click on the column cell with the entry Total Customer Visits. The right click and select Show Field List. In the PivotTable Fields pane that materializes, remove Total Customer Visits from Σ Values. Then move Years and Date to the Columns window. Then move StoreType to the Rows window. Close the PivotTable Fields pane.

3. Under PivotTable Analyze of the ribbon, select Options. Then select Totals & Filters, followed by deselecting Show grand totals for rows and deselecting Show grand totals for columns. Click OK.

4. Left click the cell 2013 Total. Then right click and deselect
   Subtotal ``Years''.

5. Click on ⊕ at the bottom of the workbook to create a new tab.
   Rename this sheet Pivot_Table_4_Heat_Map. Copy columns A
   through Y of Pivot_Table_4 and paste them to Columns A through
   Y of Pivot_Table_4_Heat_Map as Values using Paste Special.
   Delete all rows above the first row that has a populated cell.

6. Rename the cell Row Labels to Store Type.

7. In cell A10, enter =A3. Copy and paste this formula to cells A11
   through A14.

8. In cell B10, enter =B3. Copy and paste this formula to cells C10
   through M10.

9. In cell B11, enter =N4/B4-1. Copy and paste this formula to cells
   B11 through M14.

10. Highlight cells B11 through M14. Right click in one of the cell and choose `Format Cells`. In the pop-up in the `Number` tab and `Category` option, choose `Percentage`. Set `Decimal places` to 1.

11. Highlight cells B11 through M14. Under the ribbon option `Home`, choose `Conditional Formatting`. Then for `Color Scales`, choose the option `Green - Yellow - Red Color Scale`.

12. Left click on cell N11. Under the ribbon option `Insert`, choose `Sparklines`. Then select `Lines`. In the pop-up box for `Data Range` enter `B11:M11`. Copy and paste N11 to cells N12 through N14.

13. In cell N10, type `Sparkline`.

14. Then format the heat map by (say) removing gridlines, adding a border, centering text, using bold font for the Months and Store Types, ...

Store Type Monthly Sales % Change Heat Map

Store Type	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Sparkline
a	2.2%	2.4%	-4.9%	4.3%	3.4%	4.4%	-2.7%	-3.2%	9.0%	2.8%	3.5%	1.5%	
b	14.1%	12.6%	3.9%	12.4%	4.5%	10.3%	7.1%	7.1%	5.4%	4.6%	6.6%	2.7%	
c	2.4%	1.8%	-8.9%	4.7%	3.3%	2.8%	-1.8%	-3.0%	7.0%	3.4%	3.1%	3.2%	
d	7.5%	6.6%	-2.9%	9.2%	7.0%	7.4%	-0.2%	-1.8%	11.5%	6.0%	5.8%	4.7%	

Table 1: Store Type Monthly Sales Percentage Change Heat Map

References

Anderson, C. (2013). How to give a killer presentation. *Harvard Business Review*. [Online; accessed 28-Sept-2022].

Lortie, C. (2017). Ten simple rules for short and swift presentations. *PLOS Computational Biology*, 13(3):1–6.

Naegle, K. M. (2021). Ten simple rules for effective presentation slides. *PLOS Computational Biology*, 17(12):1–6.

Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphic Press, Cheshire, CT.