BUSA 603

Module 3 Supplement
Data Visualization & Some Useful Charts

Brian Weikel

brian.weikel@franklin.edu

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.

A Scatter Chart Example

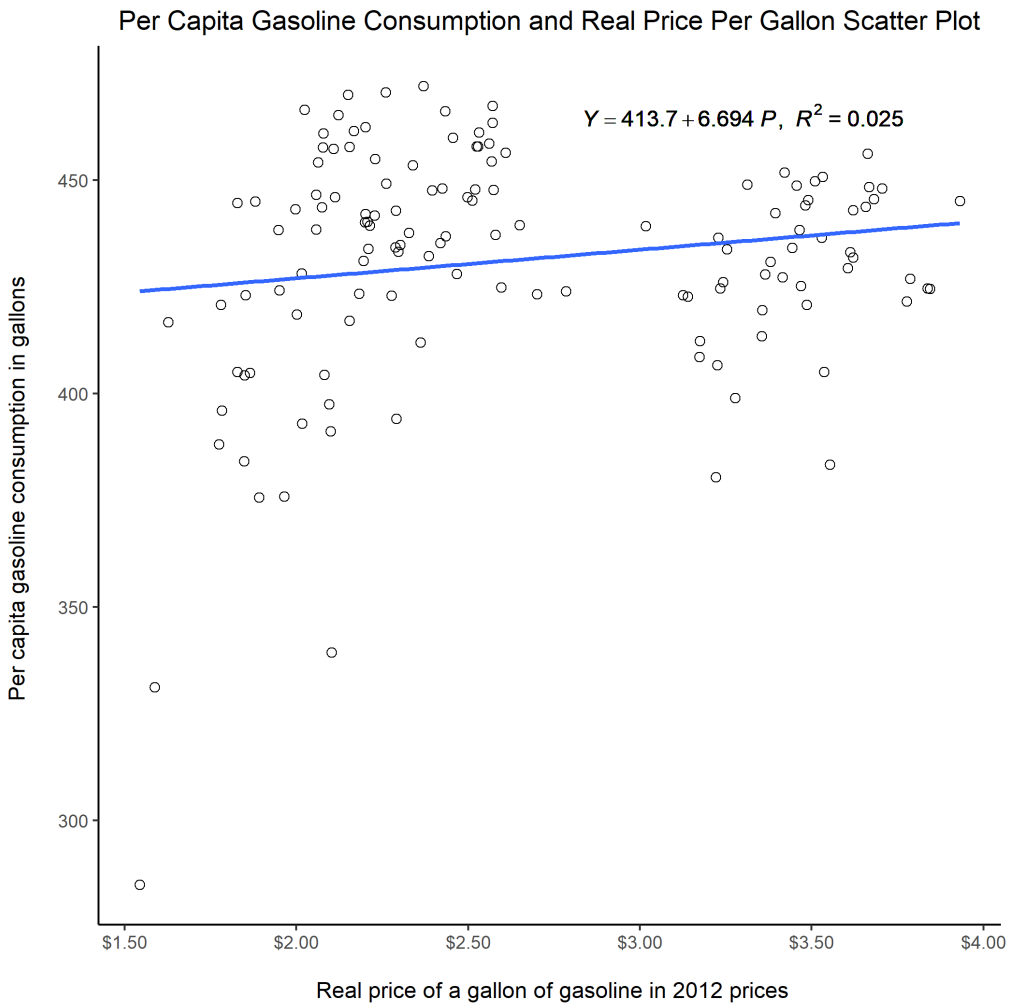

Figure 1: A Petroleum Distillate Demand and Real Price Scatter Chart

A Line Chart Example

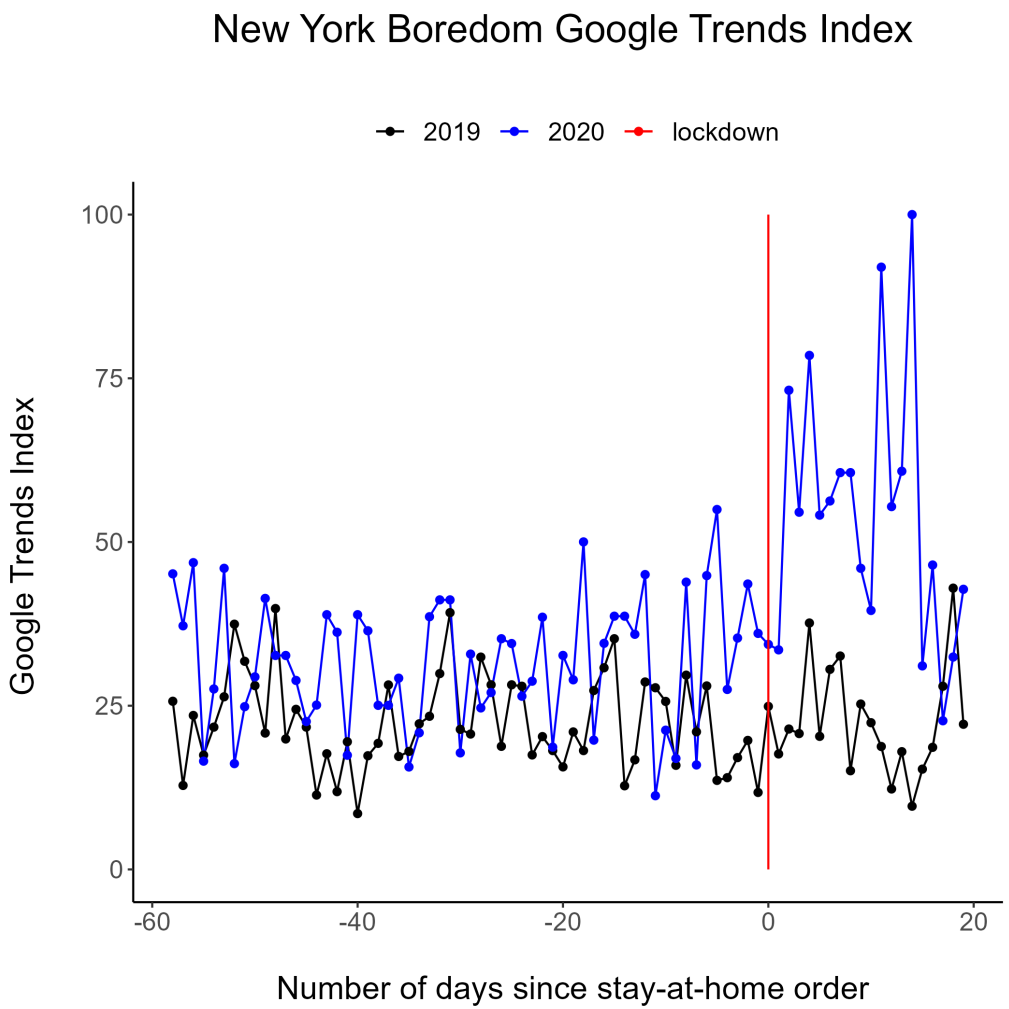

Figure 2: A Line Chart with Google Trends Data

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.

A Scatter Matrix Example

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

A Geographic Heat Map

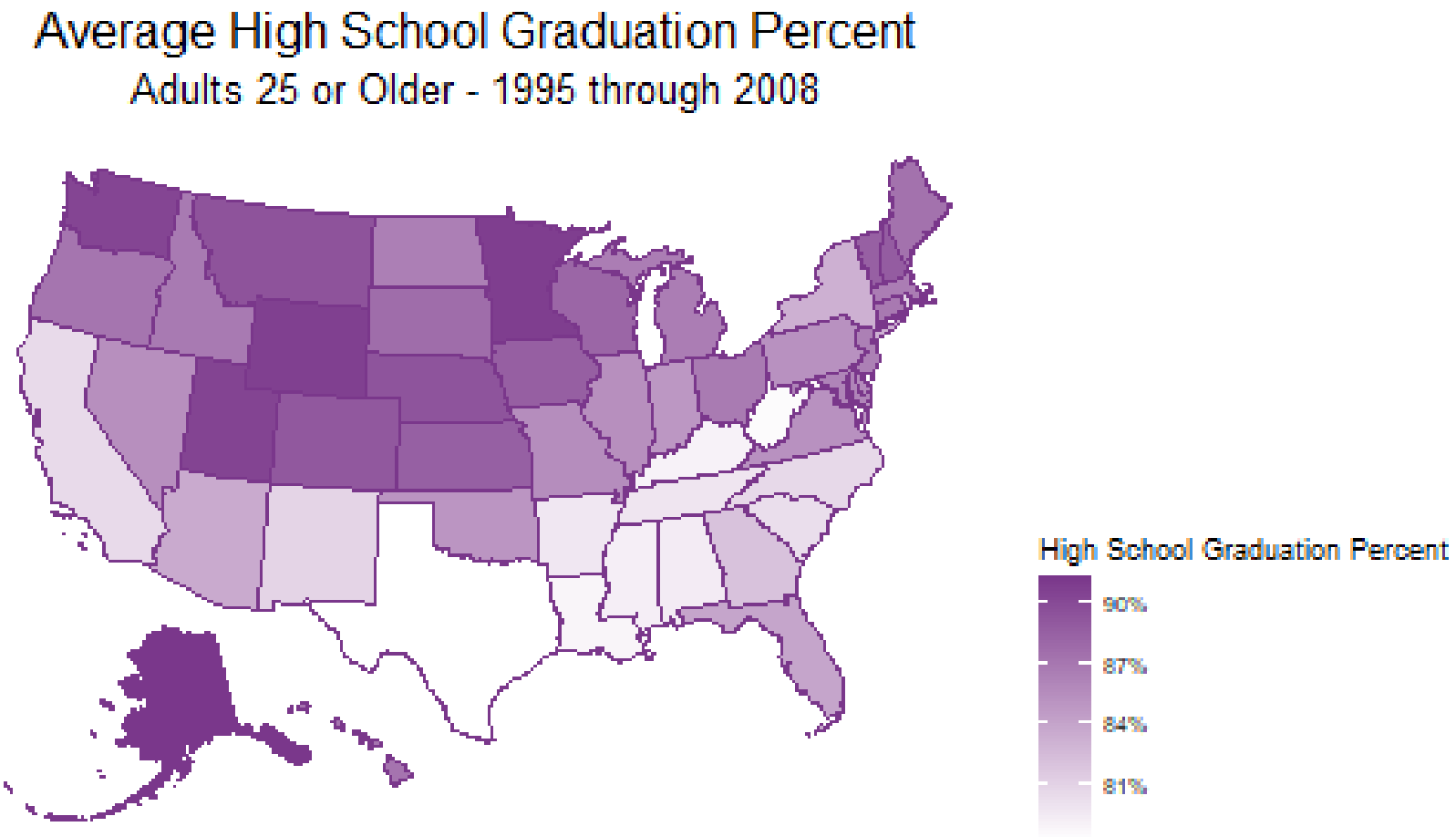

Figure 4: A Geographic Heat Map

Rossmann

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.

Data Differences

- 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.

A Call to Action!

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, 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

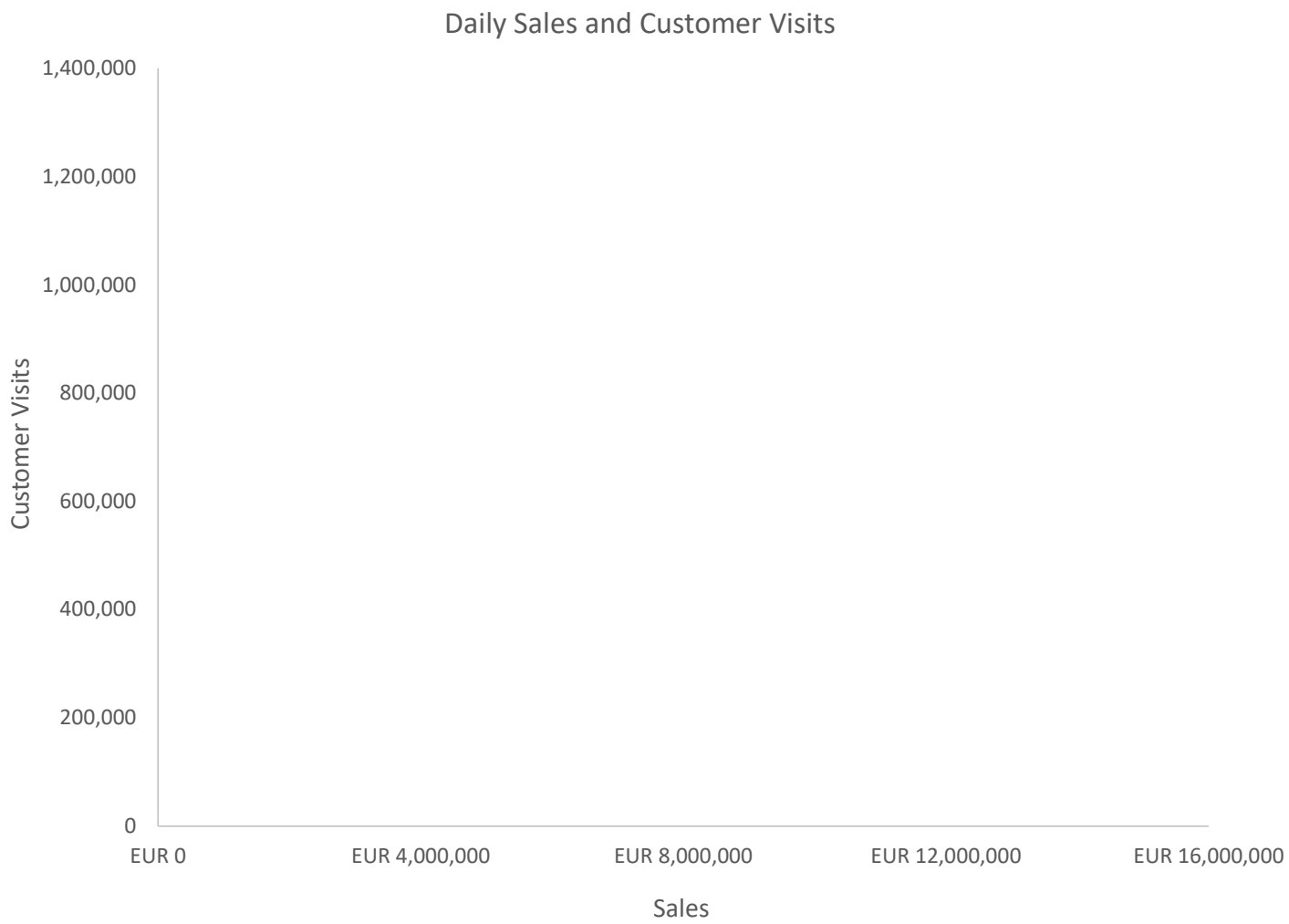

Figure 5: Daily Sales and Customer Visits Scatter Plot Template

Sales and Customer Visits Time Series Chart

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

Store Type and Assortment Bar Chart

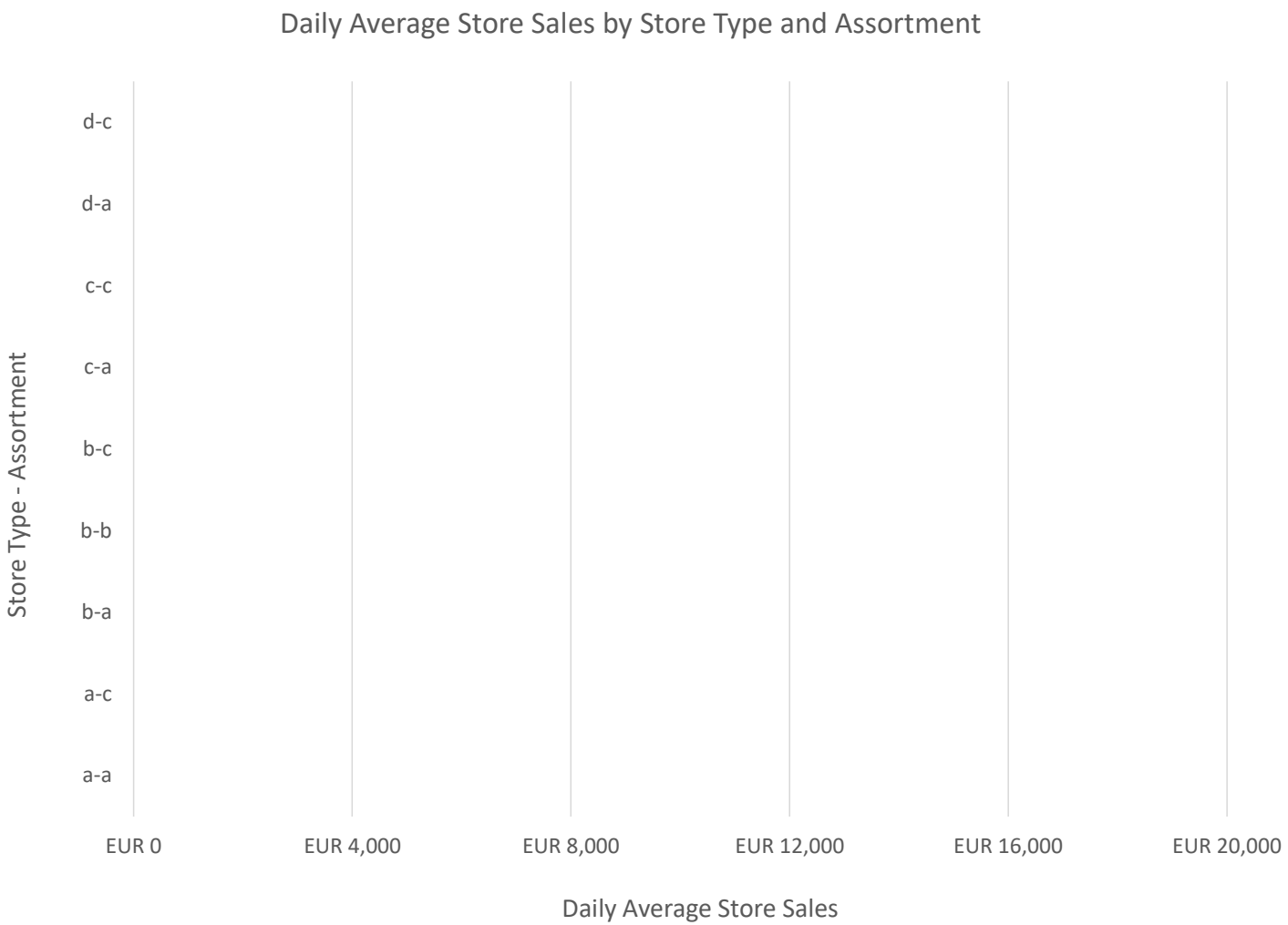

Figure 7: Daily Average Store Sales by Store Type and Assortment Bar Chart Template

Store Type and Assortment Bubble Chart

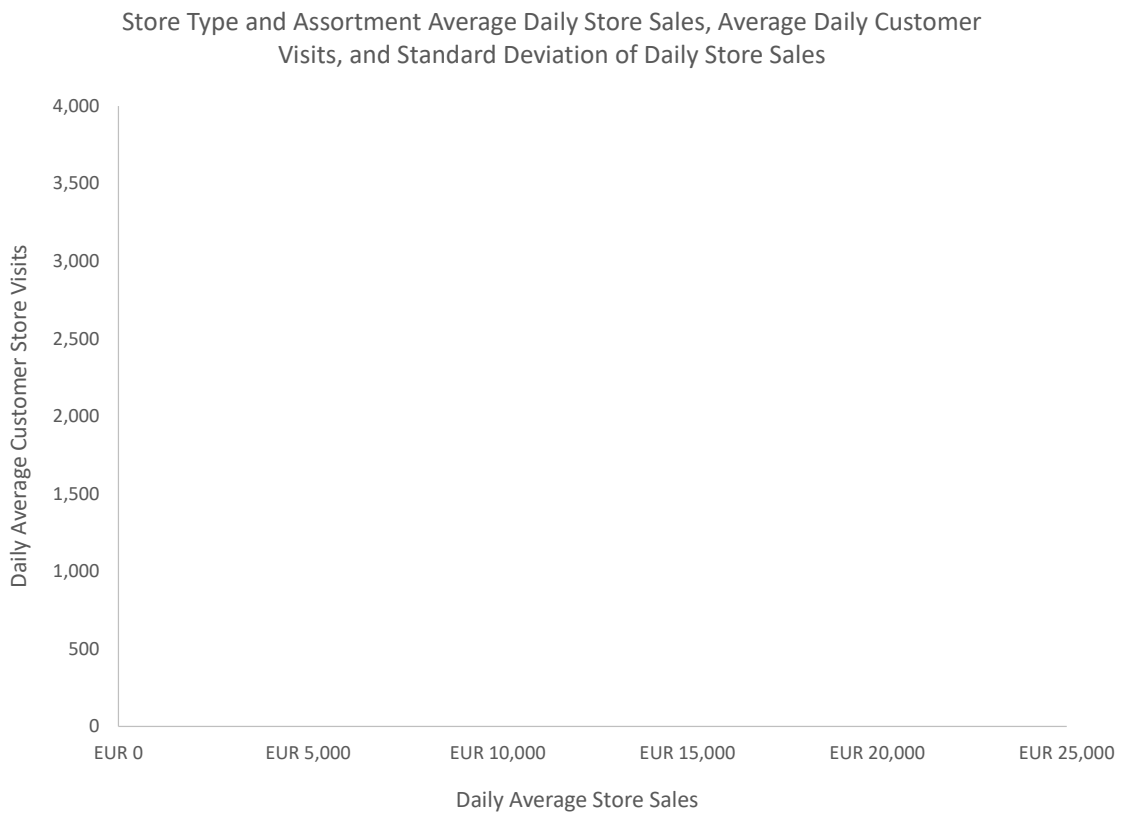

Figure 8: Store Type and Assortment Average Daily Store Sales, Average Daily Customer Visits, and Standard Deviation of Daily Store Sales Bubble Chart Template

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

Store Type Monthly Sales % Change Heat Map

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

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

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.