

Data Visualisation

Chapter 2

Advanced Visualisation

DR. TONY LYONS

AUTUMN SEMESTER 2019

B Sc (Hons) Multimedia Applications Development

B Sc (Hons) Creative Computing

WATERFORD INSTITUTE OF TECHNOLOGY

Contents

1	Introduction	3
1.1	Ordering clustered bar charts	5
1.2	Stacked bar charts	7
2	Comparison of clustered and stacked bar charts	10

1 Introduction

Like bar charts, clustered bar chart can be used to represent the frequency of data in a given class. Unlike bar charts, in clustered bar charts one also arranges the bars themselves into secondary classes which we call clusters. This means the clustered bar charts can be useful when representing and comparing data structures with multiple features on a single plot.

Example 1.1: Customer Survey Data

A mobile phone company carried out a survey on a sample of 1000 customers to determine their customer satisfaction rating. The survey asked a series of **satisfied/dissatisfied** questions in relation to five areas of customer experience

- | | |
|----------------------|-----------------------|
| (1) Customer Support | (4) Contract Duration |
| (2) Pricing | |
| (3) Coverage | (5) Loyalty Rewards |

The data obtained during the survey are displayed in the table below:

Experience	Satisfied	Dissatisfied
Support	511	489
Pricing	684	316
Contract	329	671
Coverage	848	152
Rewards	215	785

Using this data answer the following:

- (i) Identify the data types given.
- (ii) Identify the clusters.
- (iii) Draw a clustered bar chart to represent the data, with correct labels and legends.
- (iv) From the data set, determine if customers are satisfied or dissatisfied overall with the company.

Solution.

- (i) The data type given is **nominal data**, since customer experience and customer satisfaction are non-numerical categories.
- (ii) The clusters in this case are the five areas of **customer experience**.
- (iii) A clustered chart for this data is shown in Figure 1.

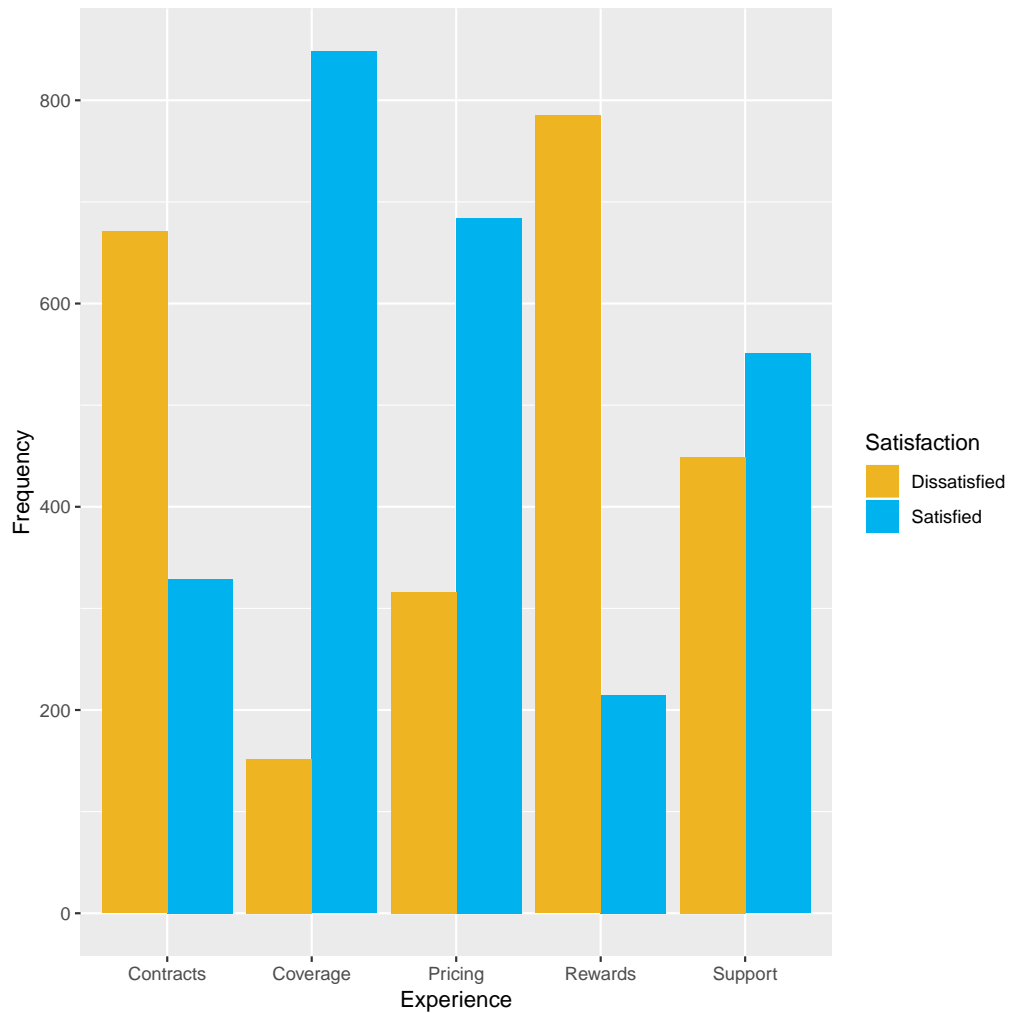


Figure 1: Sample solution for Example 1 part (iii).

- (iv) Comparing colours in the clustered bar chart in Figure 1, it appears there is somewhat of a balance between customer satisfaction and dissatisfaction.

Remarks about the clustered bar chart

Colours Colours were chosen for this plot so that there is a high contrast between satisfied and dissatisfied customer data.

Cluster spacing The clusters are spaced, while the bars within each cluster are not spaced.

Plot legend A legend is placed on the plot to highlight which colour represents satisfied and which represents dissatisfied.

Axes labels As with every plot the axes should be labeled clearly. The horizontal axis on this plot has been labeled by the classes (i.e. customer experience), with each cluster of the plot labeled clearly also. If the labels overlap, then they should be rotated to make them legible.

1.1 Ordering clustered bar charts

We recall from an earlier example of a bar plot used to represent data related to cars in a car park, that ordering of the bins may help to improve the overall appearance of the plot, and convey the information in a clearer way. The same is often true for clustered bar charts also. In this example, we could organise our clusters starting with highest satisfaction and ending with lowest. The clusters will then automatically arrange with dissatisfaction in the opposite order.

Figure 2 shows a clustered bar chart with levels of **satisfaction** arranged from highest to lowest. It is clear from the plot that levels of dissatisfaction are then arranged from lowest to highest. It is also clear, at least in this particular example, that re-ordering the data does improve the overall clarity of presentation. In particular, it should be more apparent that overall satisfaction is slightly larger than overall dissatisfaction, that is to say there is a little more blue than yellow in the plot.

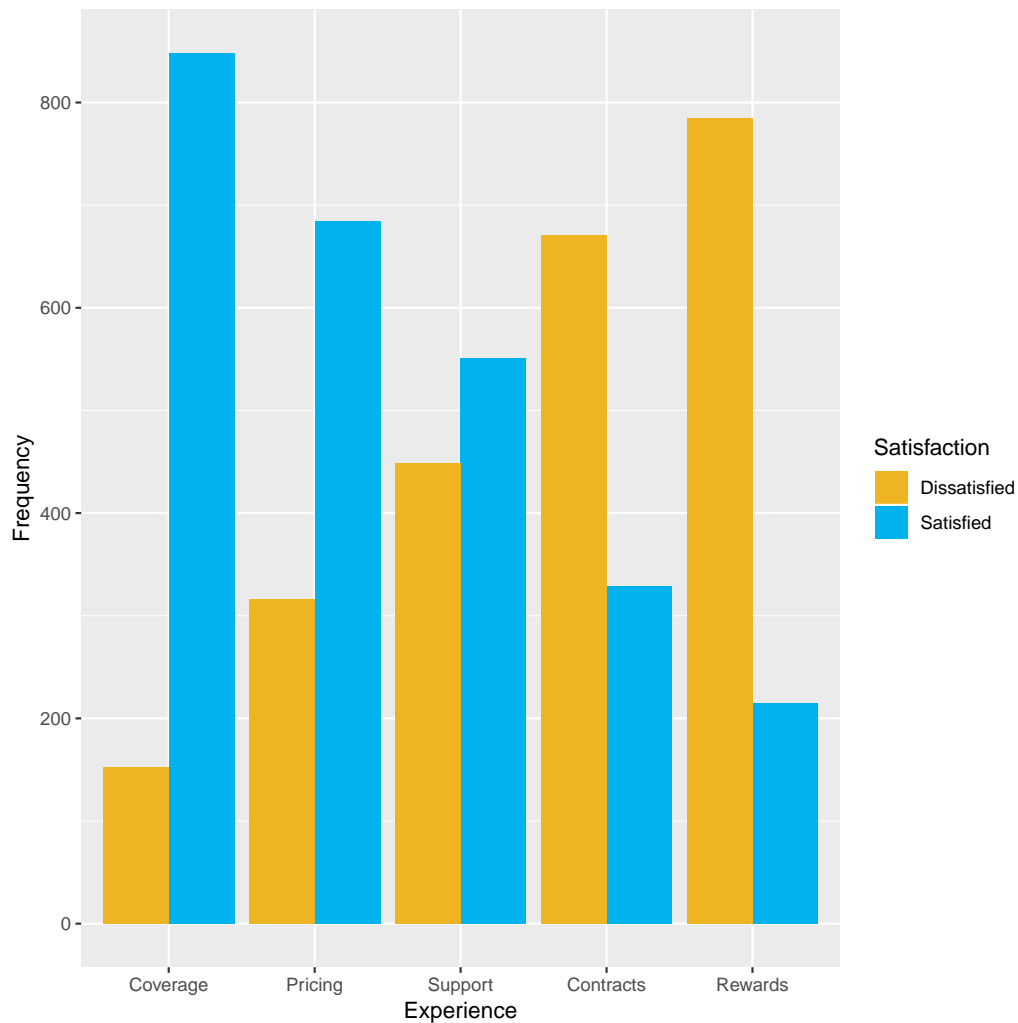


Figure 2: An **ordered** clustered bar chart representing the customer survey data in Example 1.1.

Question 1.1: Ordering of clusters

Explain why arranging the clustered bar chart from highest to lowest levels of satisfaction automatically ensures the bar chart is also arranged with dissatisfaction arranged from lowest to highest?

While ordering the bar chart in this particular case has improved the

presentation of the data, it will not always be the case that a clustered bar chart will improve in appearance by rearranging the clusters.

1.2 Stacked bar charts

Stacked bar charts are another means of representing data structure with multiple features on a single plot. In fact, the data from the previous example may also be represented on a stacked bar chart, which will result in an even clearer presentation of the data. Stacked bar charts are particularly useful when the overall frequencies of each cluster is of fixed size, but the division(s) within the clusters change. This is precisely the case we have in Example 1.1, since there are 1000 customers surveyed overall, and the levels of satisfaction/dissatisfaction within each cluster (i.e. Experience) varies.

Example 1.2: Stacked bar chart

Using the customer survey data in Example 1.1, construct a **stacked bar chart** to represent the levels of satisfaction/dissatisfaction with each area of customer experience.

Solution. In this case each category of customer experience is represented with **one** vertical bar, while that customer satisfaction/dissatisfaction are plotted different colours within each bin. A stacked bar chart representing the data in this survey is shown in Figure 3.

Remarks

- In this case all vertical bars have the same height, since there are 1000 customers surveyed in relation to each aspect of customer service.
- Again, each axis is labelled clearly, and the labels of each bin are also legible. If this were not the case, then the bin labels should be rotated to make them clearer.
- A legend is placed on the plot to indicate the meaning of each colour within the bars.

As with the clustered bar chart in Example 1.1, the appearance of the plot may be improved if the bins are rearranged, which might possibly make the data in the plot more apparent. We now rearrange the classes of the plot

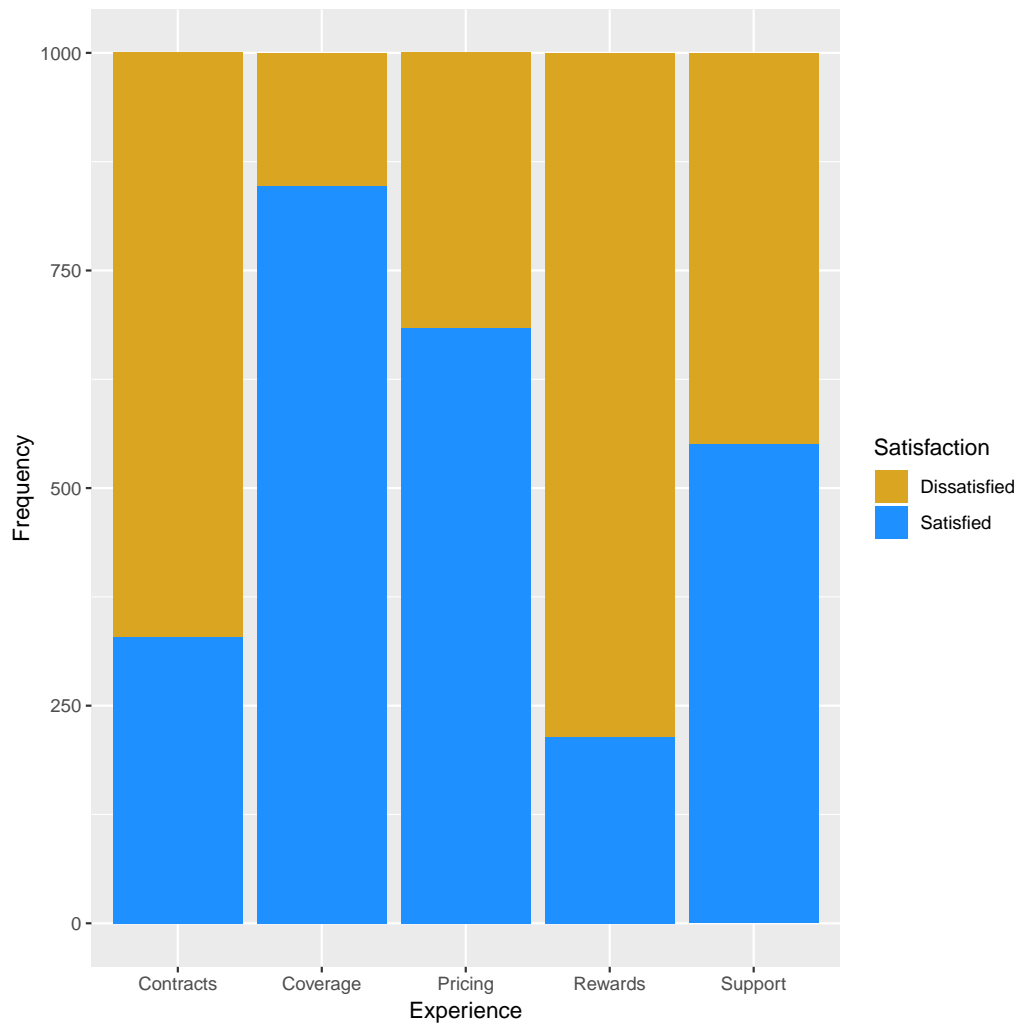


Figure 3: A stacked bar chart representing the data collected in a customer satisfaction survey.

so customer satisfaction is decreasing from left to right, as shown in Figure 4. It is again clear that by rearranging the order of the bars, the content of the data from the survey is more apparent. Again, it is easier to discern that customer satisfaction is slightly higher than customer dissatisfaction.

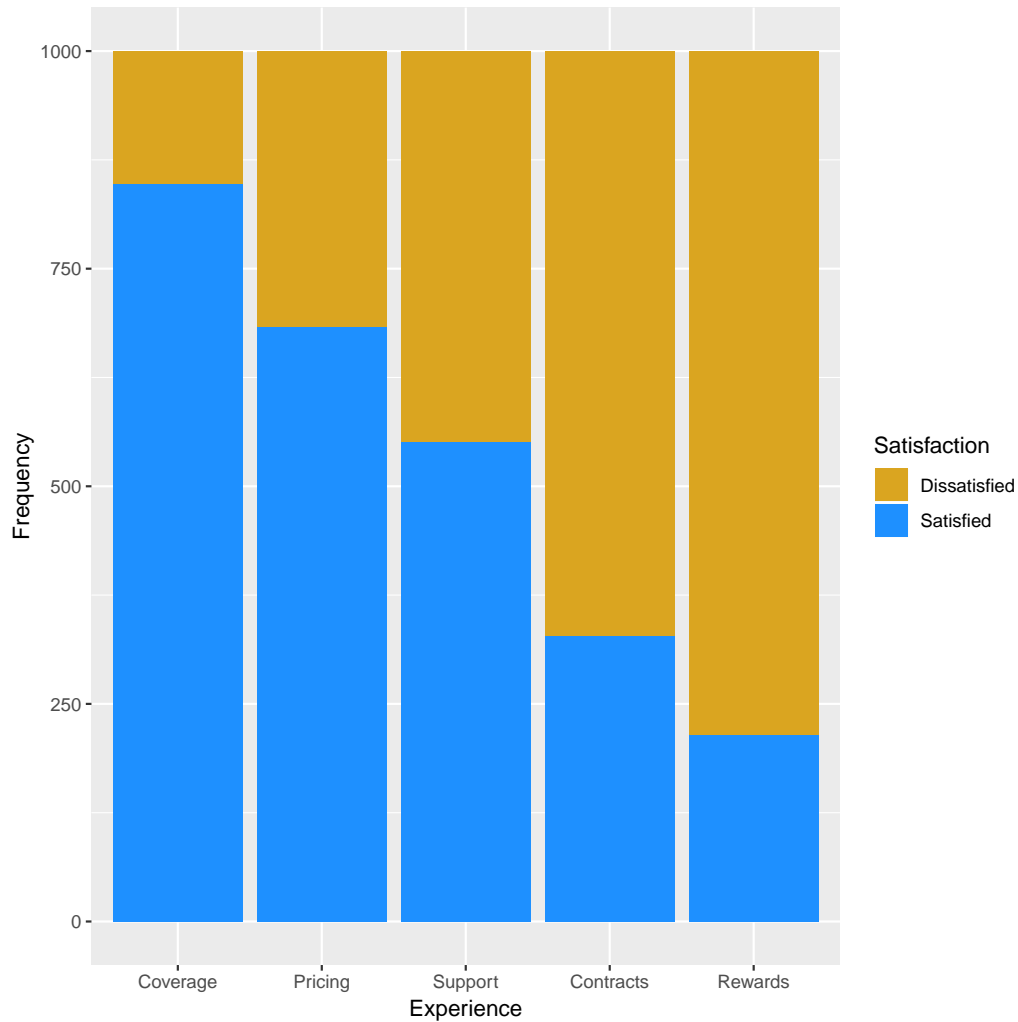


Figure 4: An ordered stacked bar chart representing the data from a customer satisfaction survey.

2 Comparison of clustered and stacked bar charts

As with most topics in data visualisation, there are no strict rules as to which type of bar chart should be used to represent a data set. In some cases it is better to try both and see which presents the data in a better format, especially since programs such as **R** can create such plots with relative ease.

Example 2.1: Online vs. Store sales

A phone retailer is collecting its monthly sales and is categorising this data according to **phone make** and **point of sale**. The data collected are given in the following

Phone make	Store sales	Online sales	Total sales
Apple	28	34	62
HTC	25	21	46
Samsung	29	37	66
Sony	19	19	38

Create a **clustered** and **stacked** bar chart to represent this data. Rearrange the clusters and bars to improve the clarity of the data being represented.

Solution. In this example the classes for the data are the **phone make** and so the data is nominal. The overall height of the bars in the stacked bar chart will vary by phone make (since it represents the total height), however this does not exclude the possibility that it may be the best form to represent the data.

In Figures 5 the clustered bar chart is shown. In Figure 6 and ordered clustered bar chart is shown. Again, it can be seen that the presentation of the data is slightly neater when the data can be ordered in some way. In Figures 7 the clustered bar chart is shown. In Figure 8 and ordered stacked bar chart is shown. This figure seems to convey the data in the clearest and neatest way, and so this is the most appropriate chart to represent this sale data. Nevertheless, it is clear from all the figures that online sales significantly outstrip store sales in these monthly sales figures.

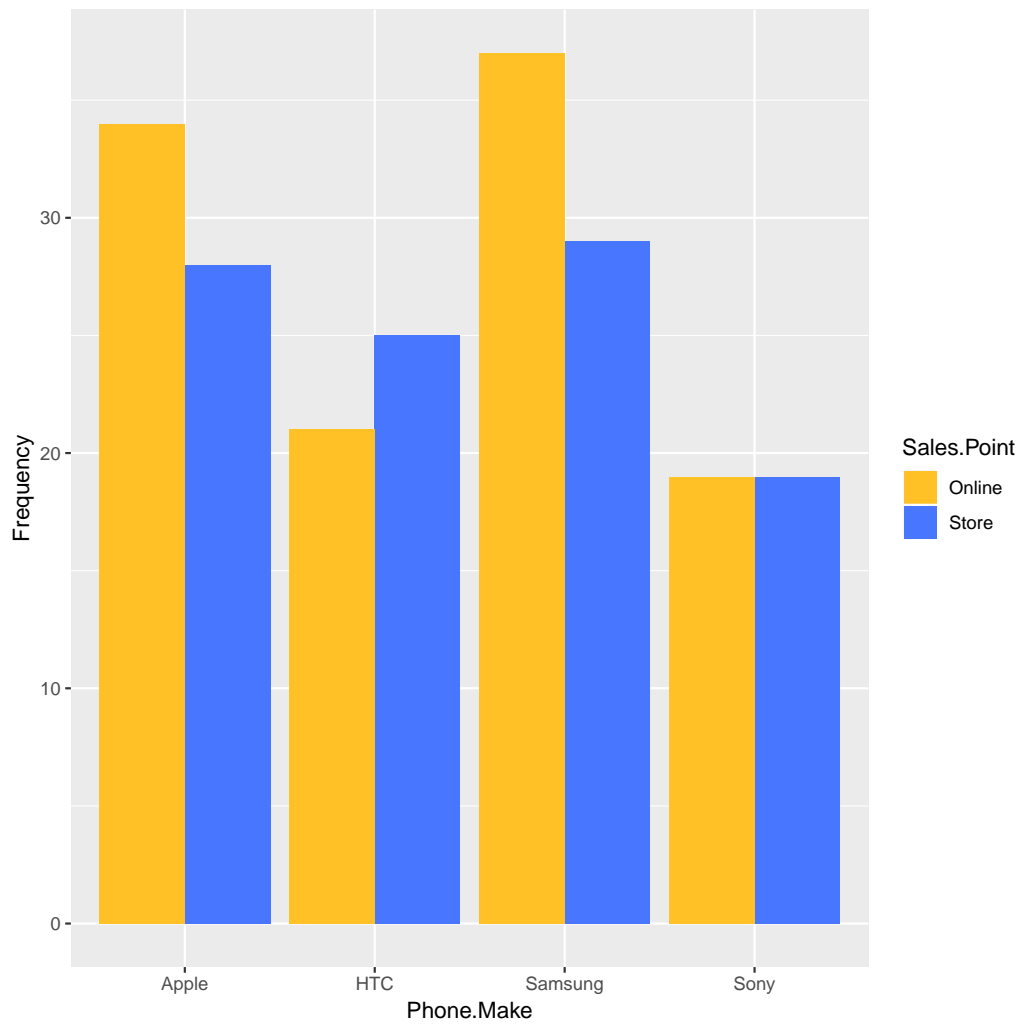


Figure 5: A clustered barchart representing the phone sales data.

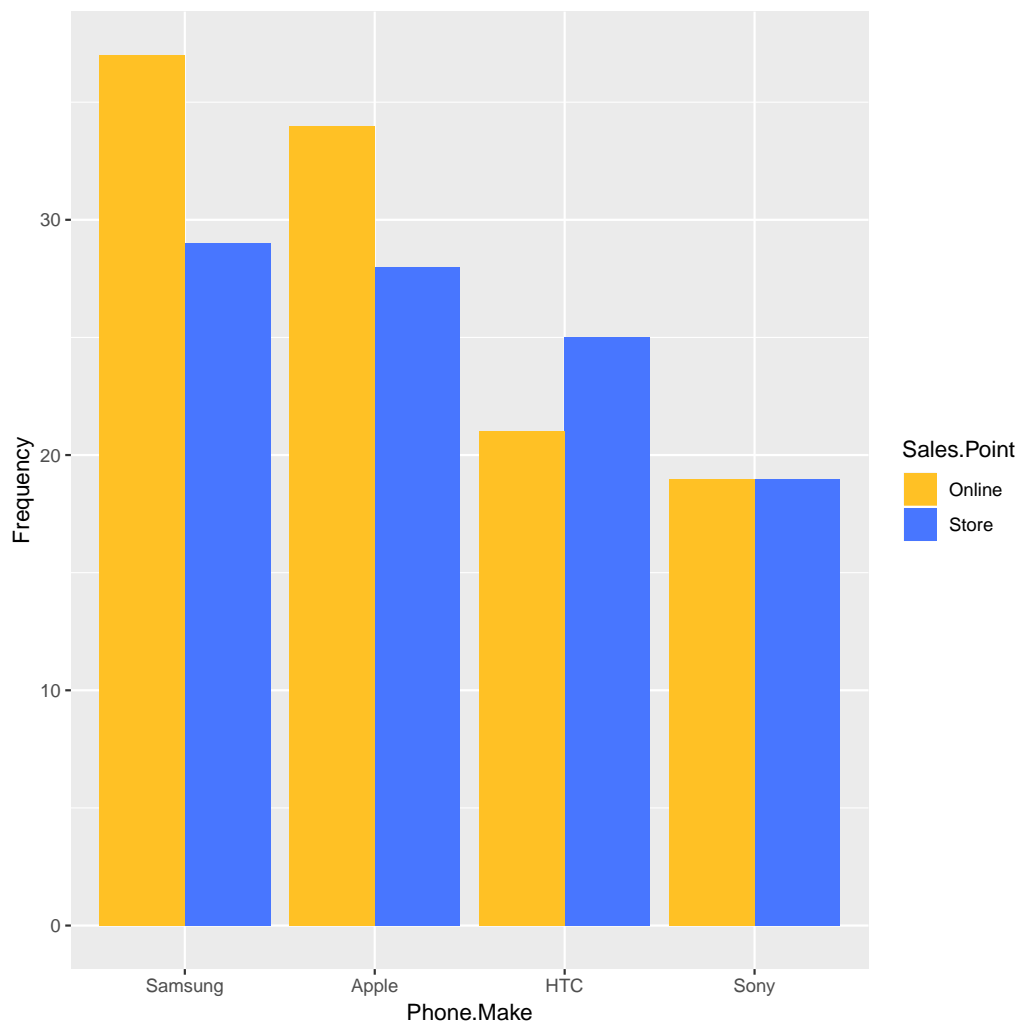


Figure 6: A ordered clustered barchart representing the phone sales data.

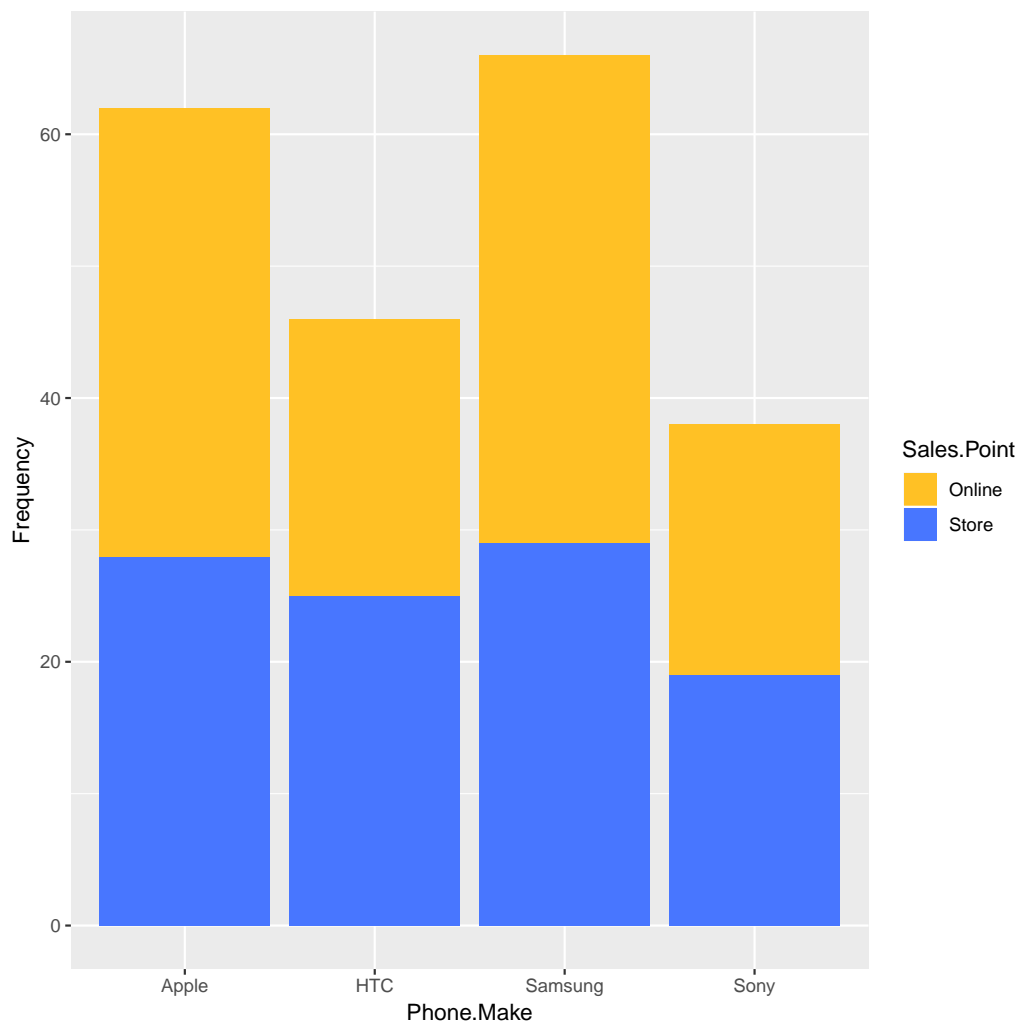


Figure 7: A stacked barchart representing the phone sales data.

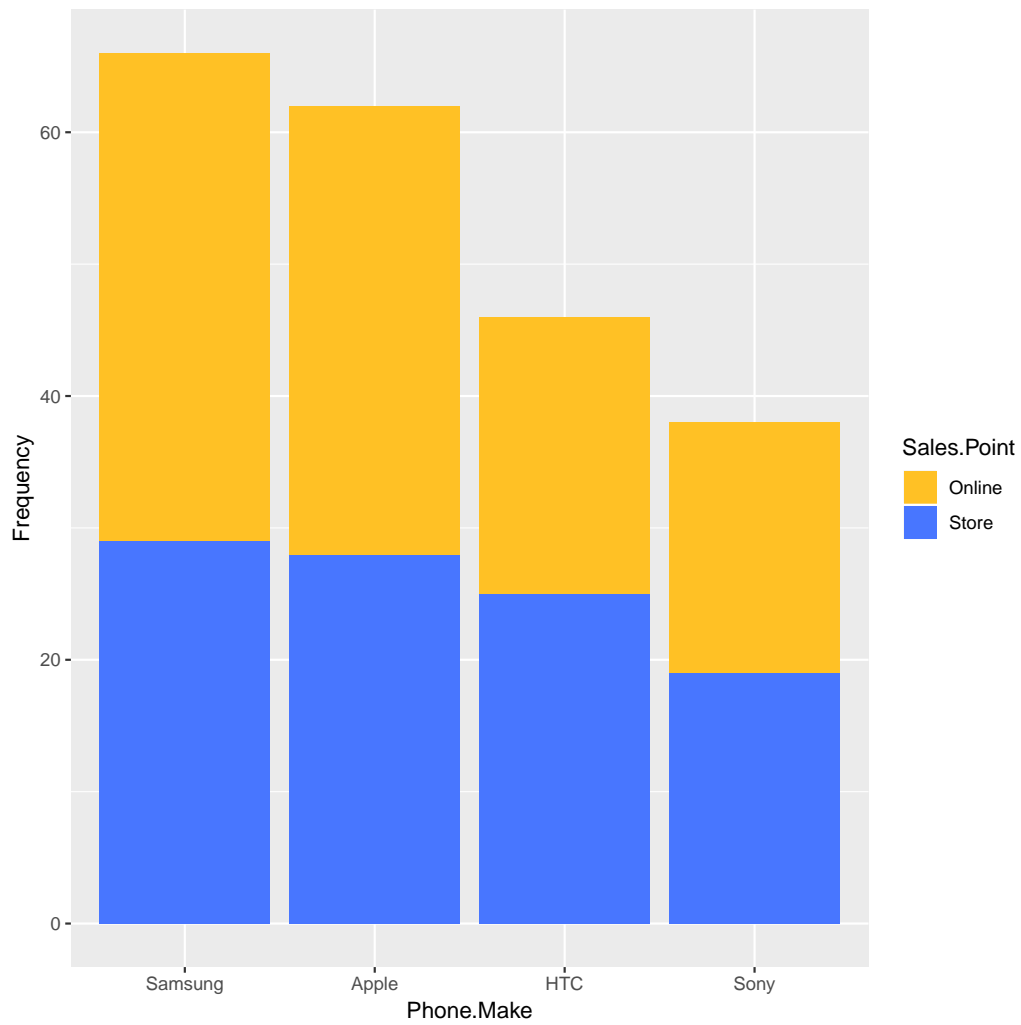


Figure 8: A ordered stacked barchart representing the phone sales data.