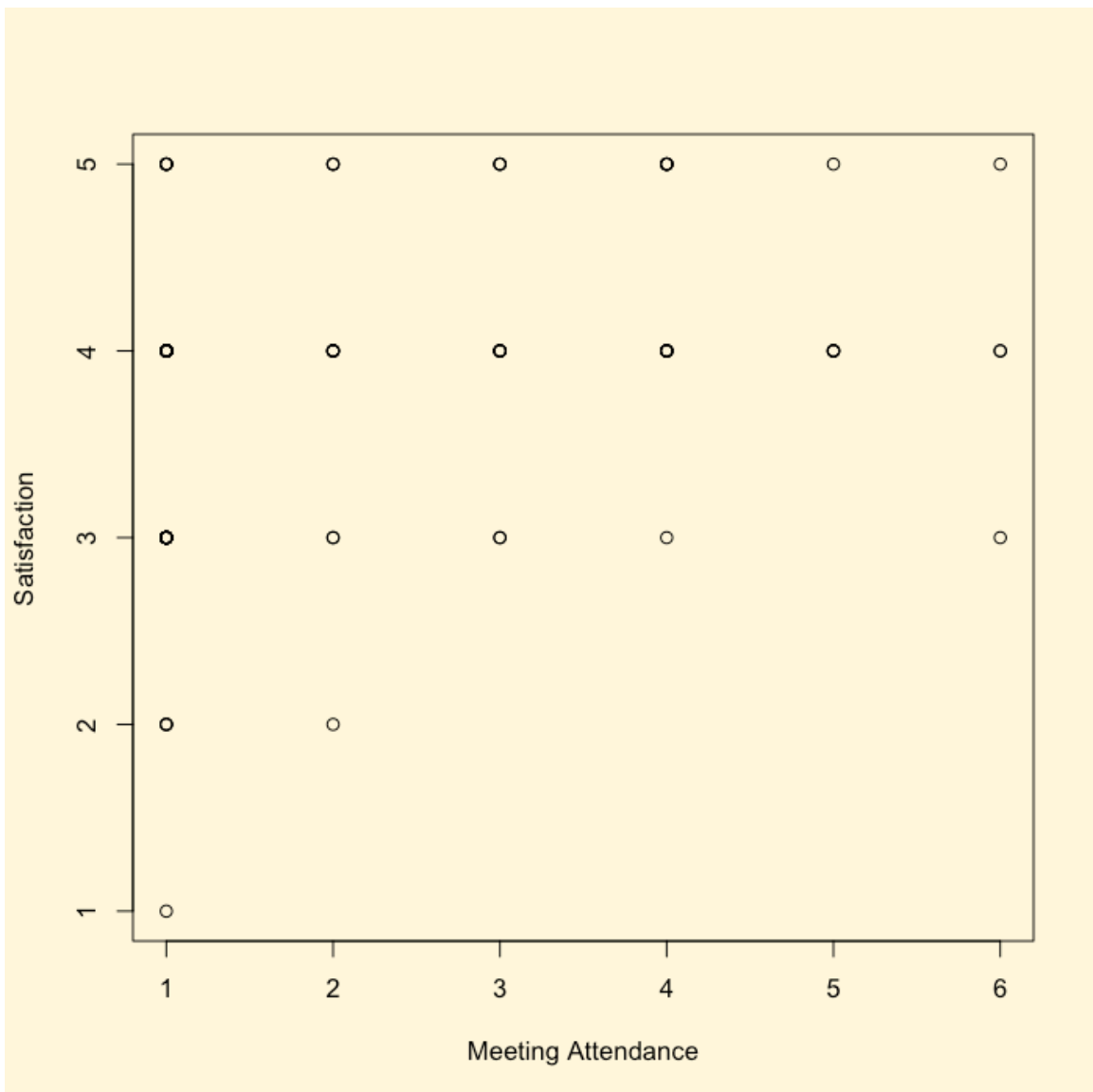


Using "Jitter plots" to Visualize Discrete Data

When analyzing the relationship between two variables, frequently the first thing we do is to make a scatter plot. This allows us to visually determine the degree and type of correlation between the variables.

This works very well with continuous variables. Unfortunately, if the data is discrete, such as survey data collected using questions scored on a Likert scale, the scatter plot leaves something to be desired.

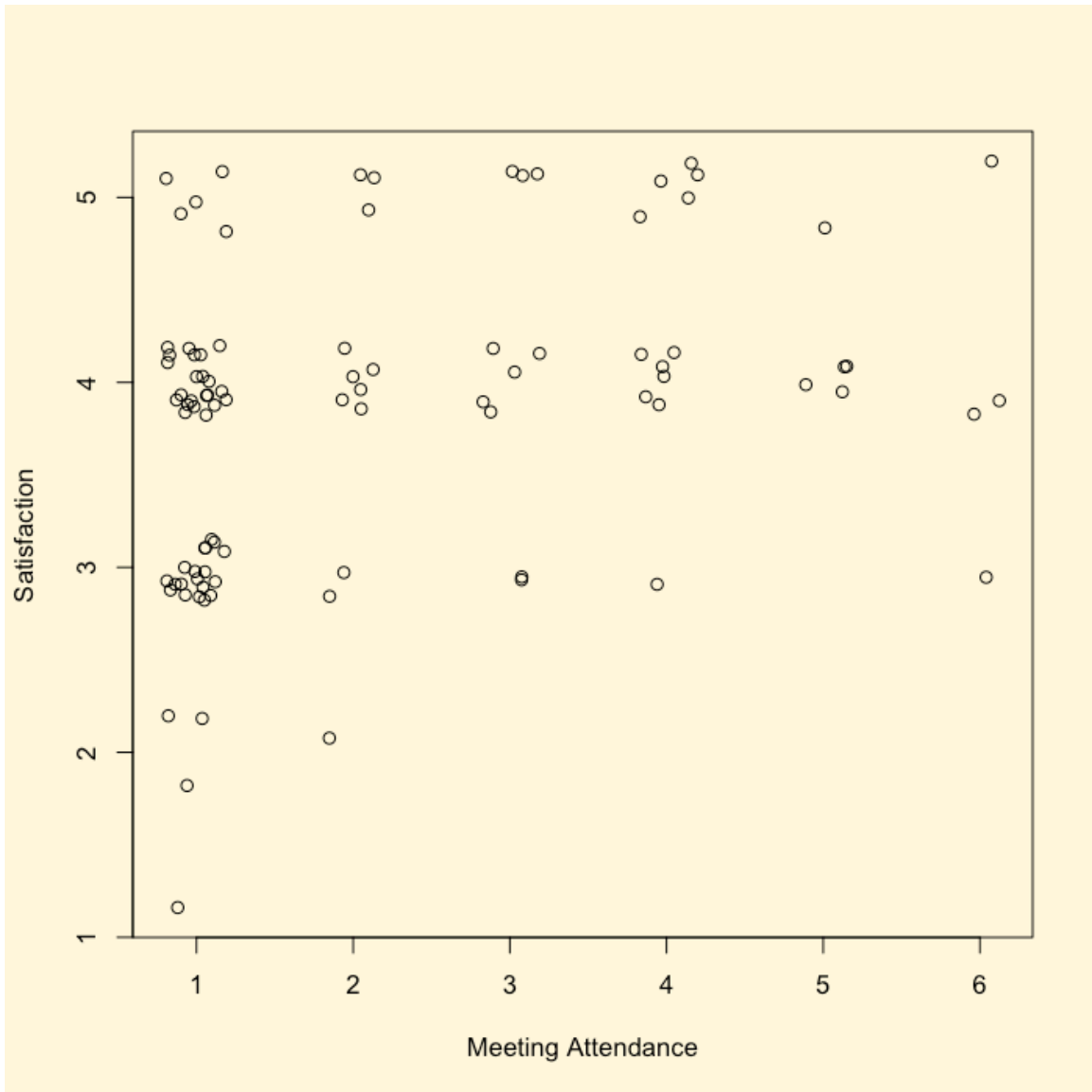
For example, here is an example of such a scatter plot. We can get an idea of the range of the data, concentrated in the upper left half of the plot. But because multiple points plot directly on top of each other, there is no way to tell which circles represent multiple samples and which represent single samples.



One way to overcome this problem is to use what I call a "jitter plot".

A jitter plot is constructed by adding a small random value to each of the x and y values in the data set. This will cause the individual data points to plot as separate points on the plot. The random values should be randomly distributed and of a such magnitude that the plotted points remain close to their actual values.

Here is the same data plotted using the "jitter plot" technique. It is now easy to see the large clusters of points near the left edge of the plot and the clustering of data along the line $y = 4$.



We can just add the fitted line to the jitter plot to complete the picture.

