Chapter 10

# Try it Yourself: Calculate Similarity Using Distance

Create the scatterplot for bill depth versus bill length.

Open the small\_penguins.csv dataset.

Select columns C and D with bill length and bill depth data.

Click “Insert” tab.

Select “Scatter” under “Insert Scatter (X, Y) or Bubble Chart”.

Add species as labels for the data points.

Click on the “plus” sign for “Chart Elements”.

Check “Data Labels”

Select labels in the plot, right-click, and go to “Format Data Labels...”

Check “Value From Cells” and select the values in species column:  
=small\_penguins!$A$2:$A$16

Click OK

Uncheck “Y Value”

Adjust the axes for better readability of the plot.

Calculate distances between all pairs of penguins.

Enter numbers 2 through 16 in the cells J1:X1. This will be used for cell references in the formula.

Enter the following expression for the distance between observations in cell J2:  
=SQRT(SUMXMY2(INDIRECT(“C$”&J$1&”:D$”&J$1),$C2:$D2))

Populate cells K2:X2 with the formula by dragging the square at the bottom right corner of cell J2.

Populate cells J3:X16 with the formula, starting with cells J2:X2 selected.

1. Use MAX() and MIN() functions to easily locate the extreme values in a table section of interest.