Iris Lab – Tutorial

Introduction

The Iris dataset is a classic dataset from the 1930s. The data contains 150 rows (observations) of irises; 50 from each of 3 species: setosa, versicolor and virginica. There are 6 columns:

- observation – a unique number
- petal length (in cm)
- petal width
- sepal length
- sepal width
- species

Here is an example of each species



Specific Learning Objectives

- Build a scatter chart
- Explore the formatting options

Datasets

- irisdata.csv

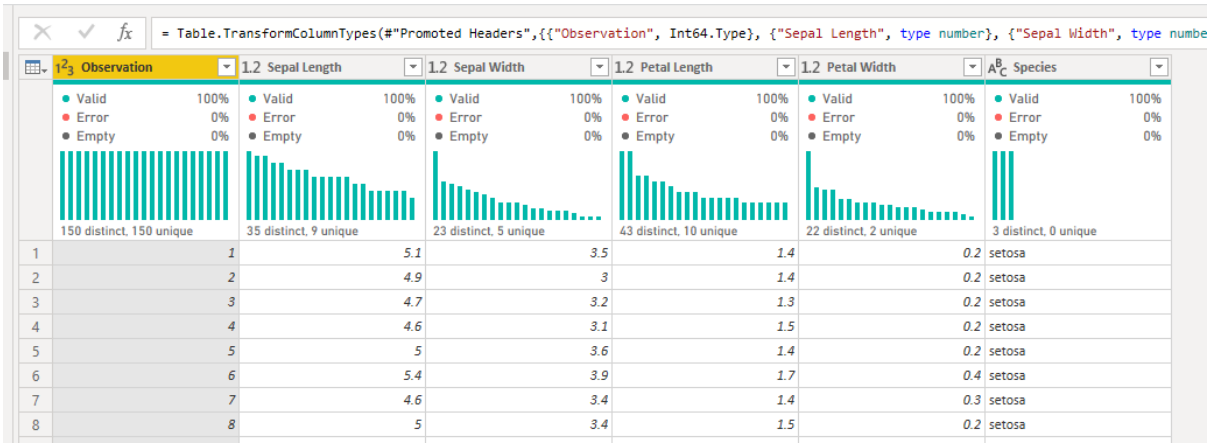
Instructions

1. Launch Power BI Desktop
2. Import the iris data CSV file
3. Build a scatter plot of petal length vs petal width and use colour to differentiate between species
4. Explore different formatting options to make a better chart; e.g. colour, shape, fill point, axes start and end values

Optional

You may want to consider the following

While in Query Editor check the Column Quality and Column distribution boxes in the View pane. What does this tell you about the data?



Set the default summarisation of the Observation column to “Don’t summarize”

Create a measure named to count the number of Irises. Use the DAX below.

```
Number of Irises = COUNTROWS(irisdata)
```

Use http://colorbrewer2.org to pick a palette of three qualitative colours and replace Power BI’s default colours with these

Choose helpful labels for both axes and a good title for the chart

Add labelled images of irises to the page or even as a background to the chart?

Add an overall trend line or a trend line for each species. Is this helpful?

Optional (advanced) Use the inbuild clustering algorithm to group into three clusters. How well has the cluster algorithm done. How many irises were put in the wrong cluster?