MEI Introduction to Data Science

Lesson 2: Pre-processing and cleaning data
Activity 1

This activity explores the main differences in weather between 1987 and 2015 at different the locations in the dataset.

Checkpoint 1

How many records are there in the Heathrow 2015 data set?

How many fields are there in the Heathrow 2015 data set?

Which numerical fields have been imported as *floating point* numbers and which have been imported as integers?

Checkpoint 2

The command `heathrow_2015_data['Daily Mean Temperature'].describe()` gave you some summary values.

The summary value 'count' indicates the number of valid data records in the data set. What about the summary value 'mean'?

Go through all the summary values and say what you think each one tells you about the field 'Daily Mean Temperature'.

Checkpoint 3

What differences can you identify between the temperature for the two years? Is this what you expected to see?

Checkpoint 4

Which value would you use to replace the string 'tr' in the rainfall field? Give your reasons.

Checkpoint 5

Use the statistics and charts produced to answer the initial problem:
What were the main differences in weather between 1987 and 2015 at the locations in the dataset?