MEI Introduction to Data Science

Lesson 5: Exploring association
Activity 1 – Suggested student responses

This activity explores which features of cars are linked to higher emissions.

Checkpoint 1

Do the range of emissions look appropriate for the different propulsion type? Give some evidence to back up your answer.

Ignoring gas/petrol as there appears to be only one car then electric/petrol cars lower emissions overall and this is what we would expect for these cars. Petrol cars tend to have higher carbon emissions than diesel as they are not as efficient whereas diesels have higher NOx. See http://www.air-quality.org.uk/26.php This is borne out by the boxplots, taking into account the high number of upper outliers for diesel NOx emissions.

Is there any further cleaning of the data that would be helpful?

There appears to be petrol car with CO2 emissions of 0 which seems unlikely and should probably be removed.

Checkpoint 2

Do cars with larger engine sizes have higher emissions on average? How do the emissions vary as engine size increases?

For CO2 there is a positive association with engine size and the range of emissions increases very slightly as engine size increases.

Do heavier cars have higher emissions on average? How do the emissions vary as mass size increases?

For CO2 there is no obvious association between mass and emissions.

Are your answers to the above two questions similar for the different types of emissions: CO2, CO and NOX?

For CO, there is very little association with either mass or engine size. For NOx, there is very little association with either mass or engine size. There is some grouping on the scatter plot which suggests we may want to investigate whether to break the data up into these separate groups.

Checkpoint 3

How are mass and engine size collectively linked to emissions?

Plotting the CO2 emissions against engine size shows a positive association. The heavier cars are distributed across this diagram indicating that mass is not associated with CO2 emissions. This confirms the answers to Checkpoint 2.

MEI Mathematics ® Education Innovation

MEI Introduction to Data Science

Which of the diagrams you have produced is the most helpful to show this?

The plot of emissions against engine size with the mass as the shading shows this well (other answers are possible).

Checkpoint 4

How does the propulsion type affect the link between mass and emissions or engine size and emissions?

The association between CO_2 and engine size is evident for petrol and diesel cars. Petrol cars have higher CO_2 emissions on average.

Are the links between mass/engine size and emissions similar for the cars registered in 2002 and 2016?

The association between CO_2 and engine size is evident for cars registered in 2002 and 2016. Cars registered in 2002 have higher CO_2 emissions on average.

Checkpoint 5

Use the charts and values calculated to answer the initial problem: Which features of cars are linked to higher emissions?

It would appear that engine size is the biggest factor in determining CO_2 emissions and this is more so with more recently registered cars. Mass is only weakly associated with emissions and this in part may be due to engine size anyway. There does not appear to be a strong determining factor for CO. Petrol cars tend to have higher CO_2 emissions than diesel and diesel cars have higher NOX emissions.