

MEI Introduction to Data Science

Lesson 3: Cleaning, formatting and grouping data

Activity 1 – Suggested student responses

This activity explores whether people more likely to cycle to work or walk to work in different parts of the country.

Checkpoint 1

Which fields of the dataset would it be appropriate to convert to numerical values?

The only ones that are not numbers are the first three fields: The Geography Code, Region and LA. The remainder could be converted to numerical values.

Checkpoint 2

When would you use the raw number of people using a type of transport and when would you use a percentage?

If comparing within the same LA it would be appropriate to use raw number. If comparing across LAs, use percentages, to take into account the different populations.

Give an example of a sentence comparing the number of people who cycle and walk to work using raw numbers.

For example, in County Durham 21 thousand people walk to work, nearly ten times the number who cycle which is just over 2 thousand.

Give an example of a sentence comparing the number of people who cycle and walk to work using percentages.

For example, a greater proportion of people walk to work in Darlington (13%) compared to County Durham (9%).

Checkpoint 3

Which regions have the most variation in the proportion of people cycling or walking to work?

Analysing by region does not reveal a great difference with the percentage walking varying from 1% to 3% and those cycling from 9% to 12%. There are slight differences with London and the South West seeming to have more of both and the West Midlands having the fewest.

What local factors might affect the proportion of people cycling or walking to work? For example the geography, types of people who live there or other factors.

The biggest factors are probably the proximity of people to their work, the type of work that they do and the availability of public transport, parking, dedicated cycle lanes and walkways. These are more likely to vary by type of LA than by Region. Patterns of commuting will be very different in a city compared to a rural area.