

# MEI Introduction to Data Science

## Lesson 3: Cleaning, formatting and grouping data

### Activity 1: Formatting fields and creating derived fields

*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?

#### Checkpoint 2

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

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

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

#### Checkpoint 3

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

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.