# Getting started hints - for students

This section provides a few brief hints for the student in how to begin thinking about analysing the data.

## Project 1 - Events over short distances

This is essentially a regression problem, where the response variable `Time` is regressed against `Date`; R can be made to recognise the latter as a date variable and will then use it appropriately in most of the modelling. When fitting a GAM, however, you might have to express the date in an alternative format such as the elapsed time from an arbitrary starting point (say, 1 January 1900).

Altitude is known to have an effect on athletic performance. In events over short distances, performance is typically enhanced at higher altitudes. The 1968 Olympic Games were held at high altitude in Mexico City so you should check that any world records set at those games do not distort your models.

A simple linear model is most unlikely to fit and even simple transforms might be expected to fail for modelling most events. When fitting a GAM, take care not to overfit the data.

## Project 2 - Events over middle distances

This is essentially a regression problem, where the response variable `Time` is regressed against `Date`; R can be made to recognise the latter as a date variable and will then use it appropriately in most of the modelling. When fitting a GAM, however, you might have to express the date in an alternative format such as the elapsed time from an arbitrary starting point (say, 1 January 1900).

Altitude is known to have an effect on athletic performance. In events over middle to long distances, performance is typically poorer at higher altitudes. The 1968 Olympic Games were held at high altitude in Mexico City so you should check that any world records set at those games do not distort your models.

A simple linear model is most unlikely to fit and even simple transforms might be expected to fail for modelling most events. When fitting a GAM, take care not to overfit the data.

## Project 3 - Events over longer distances

This is essentially a regression problem, where the response variable `Time` is regressed against `Date`; R can be made to recognise the latter as a date variable and will then use it appropriately in most of the modelling. When fitting a GAM, however, you might have to express the date in an alternative format such as the elapsed time from an arbitrary starting point (say, 1 January 1900).

Altitude is known to have an effect on athletic performance. In events over middle to long distances, performance is typically poorer at higher altitudes. The 1968 Olympic Games were held at high altitude in Mexico City so you should check that any world records set at those games do not distort your models.

A simple linear model is most unlikely to fit and even simple transforms might be expected to fail for modelling most events. When fitting a GAM, take care not to overfit the data.