Outliers

- Outliers are cases that do not fit well with the regression model (cases with large residuals).

- Like correlation, regression models are sensitive to outliers so it makes sense to remove them to improve the model.

- Bear in mind that even if you remove them they may be theoretically interesting.

- Influential cases can be detected easily via Cook's distance (part of one of our diagnostic plots).

Multiple Regression in R

We are interested in whether house prices in 250 regions of the UK can be predicted by:

(a) Population size

(b) Crime rate (per 10,000 people)

(c) Average age of people in the region

(d) Average household income in the region.

Including our predictor and a column identifying our Regions, our datasets consists of 6 variables.