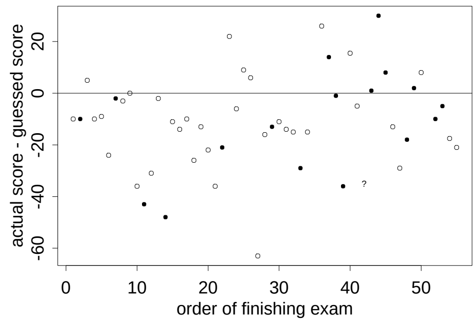

Figure B.7 *Difference between actual and guessed midterm exam scores, plotted against the order of finishing the exam. The exact order is only relevant for the first 20 or 25 students, who finished early; the others all finished within five minutes of each other at the end of the class period. Each symbol represents a student; empty circles are men, solid circles are women, and ? has unknown sex. The horizontal line represents perfect guessing. The students who finished early were highly overconfident, whereas the other students were less biased in their predictions.*

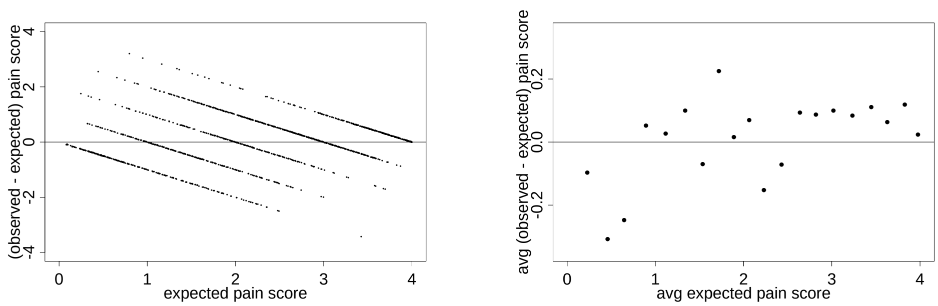

Figure B.8 *(a) Residuals (observed - expected) of pain relief scores versus expected pain relief scores from Figure B.6. (b) Average residuals versus expected pain scores, with measurements divided into 20 equally sized bins defined by ranges of expected pain scores. The average prediction errors are relatively small (as can be seen from the scale of the y-axis), but with a consistent pattern that low predictions are too low and high predictions are too high.*

As with calibration plots, it is generally a good idea to bin the points in a residual plot if the outcomes are discrete (see Figure B.8).

B.3 Miscellaneous tips

We conclude with some suggestions derived from our experiences using graphs in data analysis, first presenting a few ideas that have proved generally useful, then going through a variety of specific techniques through a series of examples.

A display of several time series of opinion polls

Each subgraph of Figure B.9 shows a time series of the support in the polls for the Republican candidate for U.S. president, as a proportion of the two-party support, for a given election year, in the months leading up to the election.

Tip: Put many little graphs on the same page. Do it with a slick graphics package if possible; otherwise, use scissors, tape, and a reducing copy.