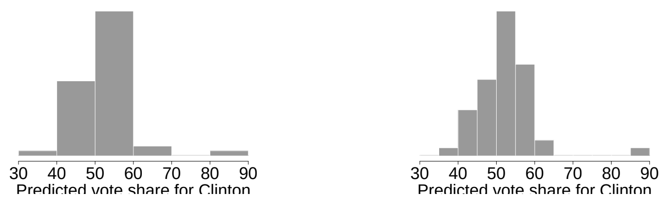

Figure B.10 *Histograms of the forecast proportion of the two-party vote for Bill Clinton in 1992 in each of the 50 states and the District of Columbia, displayed with two different choices of bin width: (a) the bin width automatically assigned by R, (b) the bin width set manually with the R command* `hist(y,breaks=seq(30,90,5))`.

an idea of the variability in the histogram itself. Do not use the default bin width in R (see Figure B.10).

General advice

Plot numerical data and inferences as graphs, not as tables. A good example is the multilevel logistic regression of vote preference on demographic and geographic predictors, with graphs on pages 306–307 that show coefficient estimates and standard errors, along with curves of the fitted model and data. Or, for a simpler example, Figure 15.9 on page 337 graphs the inference from a simple regression.

Multiple plots per page. A graph can almost always be made smaller than you think and still be readable. This then leaves room for more plots on a grid, which then allows more patterns to be seen at once and compared.

Don't plot the index numbers. For example, Figure 14.9 on 312 plots estimates for the 50 states versus average state income, rather than simply listing the states in alphabetical order. For another example, the dogs in Figure 24.1 are ordered by the time of their last shock, rather than by their ID numbers, which turn out to have no meaning in this problem.

Never display a graph you can't explain. Give a full caption for every graph (as we try to do in this book). This explains to yourself and others what you are trying to show and what you have learned from each plot. Avoid displaying graphs that have been made simply because they are conventional. For example, regressions are commonly equipped with quantile-quantile plots of residuals, but for most applications the information in such a plot is irrelevant, and a distraction from the more relevant results that could be presented.

B.4 Bibliographic note

For statistical graphics in R, the book by Murrell (2005) is an excellent overview and starting point. Fox (2002) is also helpful in that it focuses on regression models.

On the topic of statistical graphics more generally, much of the most important and influential work has appeared in books, including Bertin (1967, 1983), Chambers et al. (1983), Cleveland (1985, 1993), Tufte (1983, 1990), and Wainer (1984, 1997).

There are various systematic ways of studying statistical graphics. One useful approach is to interpret graphs as model checking (for example, if residuals are not independent of x, then there is some model violation), as we have discussed in Chapter 24. Another approach is to perform experiments to find out how well