

Box 1.3 A Special Note on Weights

The states and world datasets are unweighted. In analyzing unweighted data, you do not need to adjust for sampling bias, because each state or country is equally and adequately represented in the dataset. For example, to calculate the average percentage of women in parliaments of the world (recorded in the variable `world$women09`), you would ask R to sum the percentages for each country and divide by the number of countries.

By contrast, the `gss` and `nes` datasets must be weighted. Why is this? In unweighted form, these datasets contain sampling bias—that is, some groups are over- or under-represented when compared with the overall population of adults. So, for example, if you wanted to calculate the average age of respondents in the `nes` dataset, the unweighted average would be distorted, because not all age groups are equally and adequately represented in the dataset. To correct for this bias, survey designers provide sampling weights. Therefore, in order to obtain accurate results from the two survey datasets, `gss` and `nes`, you will need to weight your analyses by the appropriate sampling weight. For `nes`, the weight variable is `nes$wt`; for `gss`, it is `gss$wtss`.

Most of the base R functions do not permit sampling weights. Fortunately, the extra packages you installed in this chapter contain procedures that can be used with weighted data (such as `gss` and `nes`) or unweighted data (such as `states` and `world`). On rare occasion, however, you will learn separate procedures, one for weighted data and one for unweighted data.