In this chapter we extend our analyses of the examples in Chapter 1 in order to display further features of the forward search. We use our analysis of the Swiss heads data to exemplify the properties of bivariate boxplots for data analysis. As a preparation for material on transformations of data in Chapter 4 we compare analyses of the data on national track records for women when the response is the time for the race and also its reciprocal, speed. This transformation leads to an appreciably simpler analysis. Our further analysis of the data on municipalities in Emilia-Romagna focuses on the last sixteen units to enter the forward search. For part of our analysis we reduce the data to five selected variables that explain much of the structure of the outliers. The last example is the data on Swiss bank notes. We analyse all 200 observations together and also look at the two groups separately. Forward plots of individual Mahalanobis distances, calibrated by plots of a large number of units of known origin, are shown to be a powerful tool for determining group membership.

3.1 Swiss Heads

To find the initial subset for the forward search we fit a robust ellipse to each bivariate scatterplot, scale the ellipse and then take the observations in the intersection of all scaled robust ellipses as our starting point (§2.13.3). Figure 3.1 shows a scatterplot matrix of the data with the ellipses superimposed. The inner ellipse is the robust ellipse containing exactly 50% of
the data. The outer ellipse is the same ellipse scaled using $\theta = 0.92$, which corresponds to a theoretical value of 60% of the data in a single boxplot. Since the content of the ellipses is similar, they are hard to distinguish in the plot, even though we have used different types of line for the two of them. There are exactly 25 observations inside all the 60% ellipses; these defined the initial subset for the search in the first chapter.

Figure 3.2 replots Figure 3.1, except that the coefficient for the outer threshold is now $\theta = 4.71$. This larger ellipse gives some indication of whether there will be any outliers, in so far as bivariate plots are enough to establish this. Units 111 and 104 were the last to enter the search in