Which list is which

October 11, 2011

To group the data into Charlie's data and Alison's data, I use three approaches. I am interpreting each data set - a, b, c, d, e, f - as a vector - A, B, C, D, E, F $\in R^{40}$.

In the first apprach I plot every vector against A. I look for the same pattern in the plots to group the data sets.
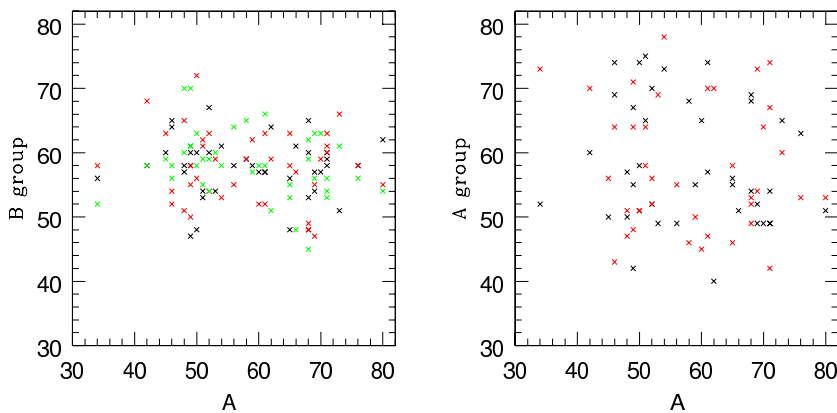

Figure 1: In the left plot I plot vectors B, C and F against A, in black, red and green respectively. In the right I plot D and E against A in black and red respectively. I find that the vectors B,C,F are in the same group because the plots of these have the same pattern to each other but different pattern to the plots of D and E.

In this approach I plot the data values against the index of that data point.
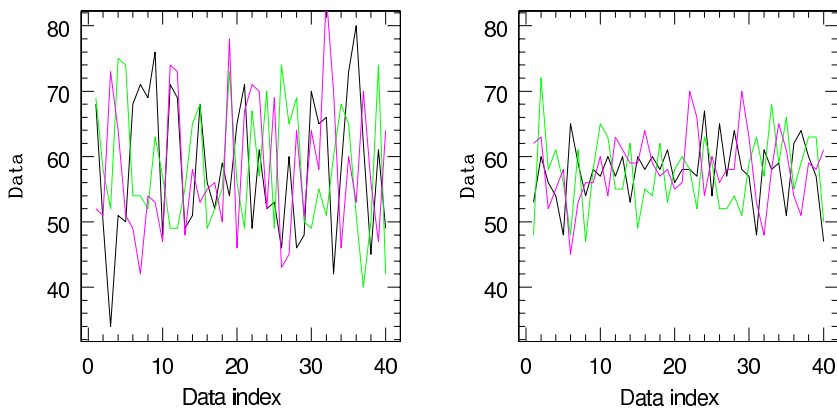

Figure 2: In the left A, D, E are plotted against index vector in black, green, magenta respectively. On the right B, C, F are plotted against index in black, green and magenta respectively. Alison's vectors (on left) have means of 58.40000153, 58.34999847, 58.29999924 and standard deviations of 10.60141468, 9.467180252, 10.6476516. Charlie's vectors (right) have Means of 57.72499847, 57.57500076, 58.20000076 and Standard deviations of 4.444026947, 5.713525772, 5.129327774. So the two groups have diffeent spreads.

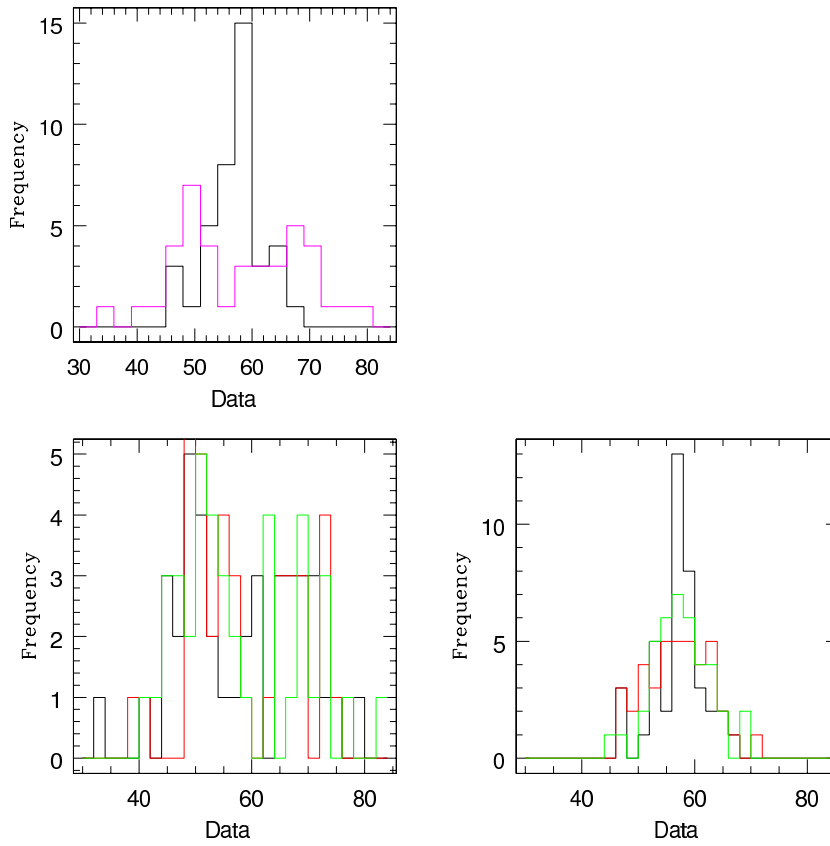In this approach I find the frequency distribution of each data set.

Figure 3: In the top left, I plot frequency distribution of a in magenta and that of b in black. In the bottom left I plot frequency distributions of a, d and e in black, green and red respectively. In the bottom right panel I plot b, c and f inb black green and red respectively. I notice that the spread of a, d and e is higher than that of b, c and f. I alao notice that the distrubution of a, d and e has more modes than that of f, b and c.