/\*\*\*\*\*\*\*\*\*Example 1\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*two variable clustering\*\*\*\*\*\*\*\*/

data cluster;

length student$ 24; //the $ is SAS for character data type

input student$ exam hw;

cards;

1 58.33333333 94

2 61.66666667 49.875

3 88.66666667 94.5

4 88.66666667 95.625

5 90.66666667 90.875

6 79 55

7 94 97.75

8 66 24.75

9 87 98.5

10 52 92.75

11 93 96.75

12 84.33333333 98.375

13 88 83.375

14 95.33333333 91.875

15 88 96.625

16 94.66666667 97

17 93.66666667 95.125

18 73.5 57

19 88.33333333 78.5

20 94 85.375

21 21.33333333 10.42857143

22 77.33333333 60.375

23 74.66666667 71.375

24 85.33333333 94.5

25 82 99.375

26 79 80.5

27 65.33333333 68.125

28 14 53.28571429

29 76 81.5

30 95.66666667 67.25

31 78.33333333 98

32 57.66666667 37.375

33 96.66666667 100

34 66 79

35 74 83.75

36 74 98.25

37 7.333333333 28.57142857

38 94 95.75

39 62.66666667 71.75

40 96 50

41 93 97.125

42 81.66666667 74

43 85 98

44 20.66666667 37.5

45 59.66666667 75.875

;

run;

proc fastclus data=cluster maxclus=3 out=clus; //the procedure fastclus clusters data here into 3 groups

var hw exam; //you tell SAS which data you want to cluster

run;

quit;

proc print data=clus noobs;

run;

//this code is SAS for plotting a graph for these data

data labels;

length function text $ 32;

retain function 'label' xsys ysys '1';

set clus;

x=exam; y=hw;

text=student;

run;

title'';

proc gplot data=Clus;

symbol1 c=blue v=square h=3;

symbol2 c=black v=circle h=3;

symbol3 c=red v=dot h=1;

plot hw\*exam=Cluster/annotate=labels;

label exam='Average Exam Score' hw="Average Homework Score";

run;

quit;