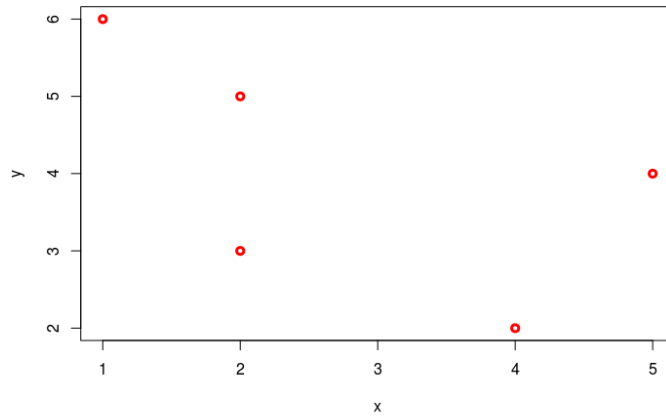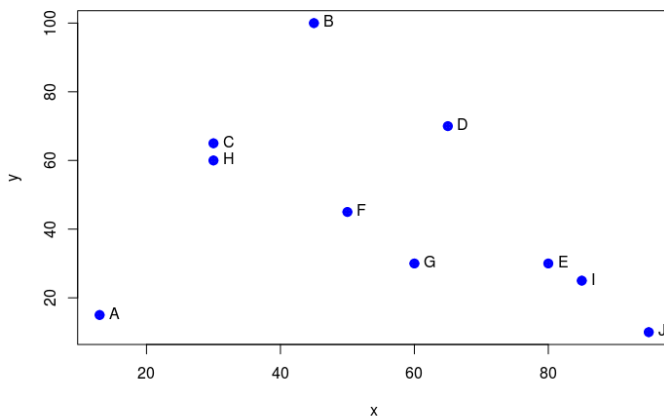1. (**Scatter Plot Basic**) Below is a scatter diagram for certain data set. Fill in the blanks.

| x	y
2	3
4	2
1	
	4

2. (**Scatter Plot interpretation**)Students named A, B, C, D, E, F, G, H, I and J took a midterm and a final in a certain course. A scatter diagram for the scores is shown below:

(a) Which student scored the same on the midterm as on the final?

(b) Which students scored higher on the final?

(c) Was the average score on the final around 25, 50 or 75?

(d) For the students who scored over 50, was the average score on the final around 30, 55 or 75 ?

(e) True/False: on the whole, students who did well on the midterm also did well on the final.

(f) True/False: there is strong positive relation between midterm scores and final scores.

3. As done in the Sketch a line that fits the data given in the following graphs:
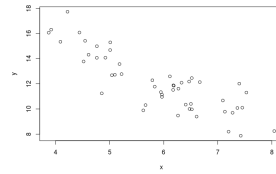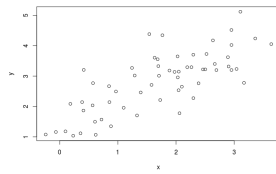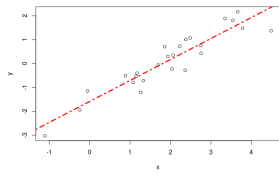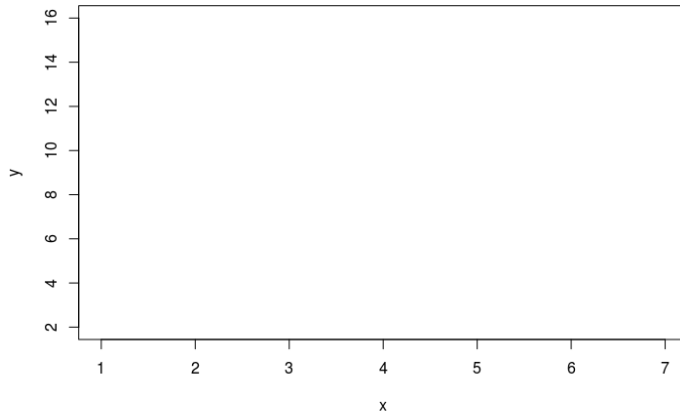

4. (**Finding the best line fit from Scatter Plot**)
   Consider the following data:

x	y
2	6
4	8.5
1	2.5
7	15
5	11

   (a) Make a scatter plot of (x, y) in the graph below:

   (b) It is believed that $y = ax + b$ is the true relationship. Below will help us find the best a, b.

   i. Using the scatter plot, *the predicted value of y for x_i is defined as $\hat{y}_i = ax_i + b$.* Write down $\hat{y}_i = ax_i + b$ for $i = 1, 2, 3, 4, 5$.

   ii. The *prediction errors* are defined by $d_i = y_i - \hat{y}_i$. Write down $\hat{y}_i = ax_i + b$ for $i = 1, 2, 3, 4, 5$

   iii. In the above scatter plot draw a line for some $a > 0$ and $b > 0$ and mark the d_i on the graph for the line that you drew.

   iv. Find the values of a and b which minimizes the error sum of squares which is $\sum_{i=1}^{n} d_i^2$.