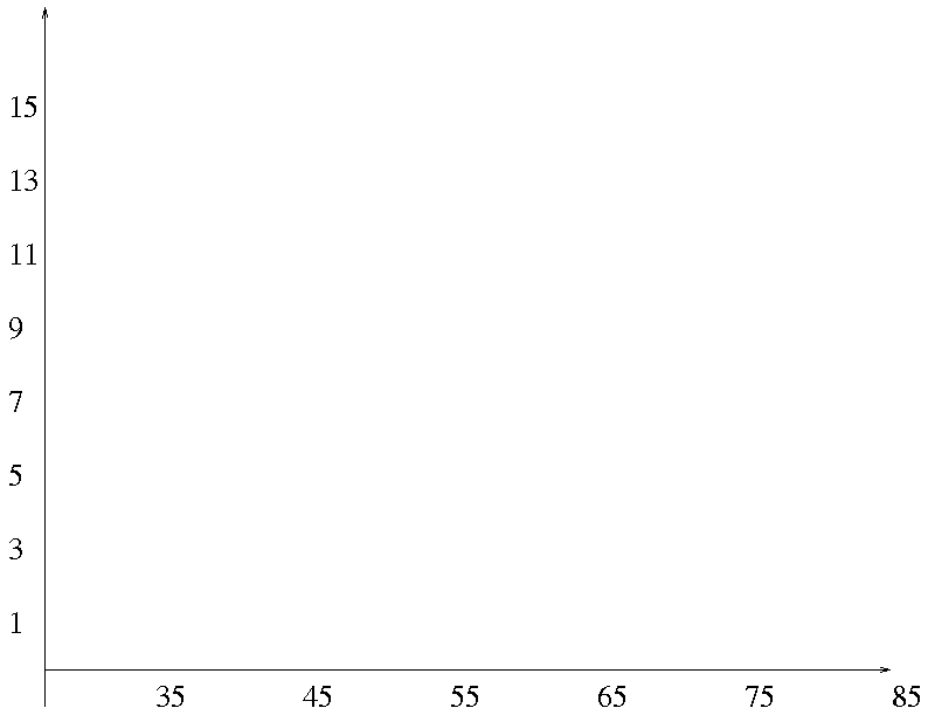
Question 1: Consider the following data set:

$$84, 49, 61, 40, 83, 67, 45, 66, 70, 69, 80, 58, 68, 60, 67, 72, 73, 70, 57, 63, 70, 78, 52, 67, 53, 67,$$
$$75, 61, 70, 81, 76, 79, 75, 76, 58, 31.$$

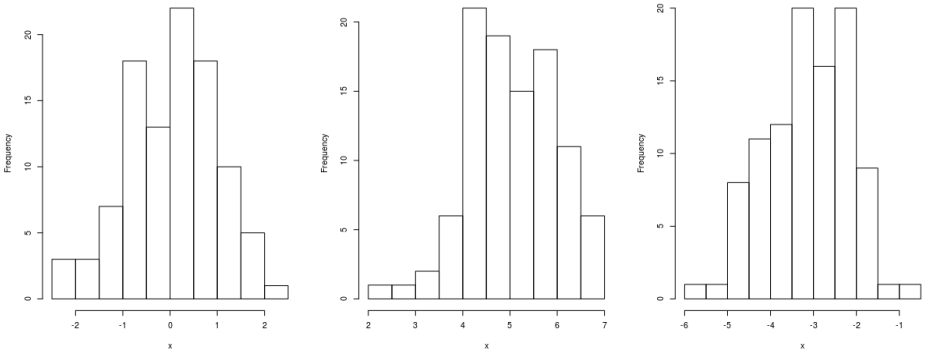
1. Find the mean, median and standard deviation of the data.
2. Fill in the following frequency table for the above data set:

Classes	Frequency
0-35	
36-45	
46-55	
56-65	
66-75	
76-85	

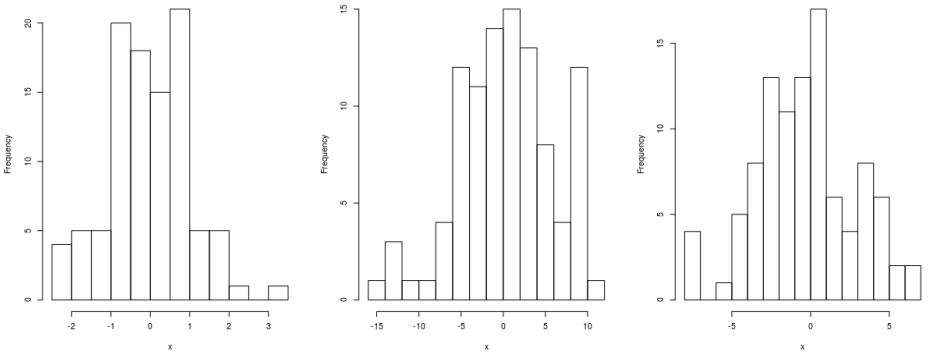
3. Draw the histogram of the above data on the graph below:



Question 2: (a) Three histograms are given below. Fill in the blank for each of the histogram: The average is around _____. Options : $-1, 0, 4, 5, -3$



(b) Three histograms are given below, all with mean 0. Which of these has maximum and minium variance?



Question 3: Let $\{(x_i, y_i) : 1 \leq i \leq n\}$ be a set of points on the plane. Let $a, b \in \mathbb{R}$, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

1. Show
$$\sum_{i=1}^{n}(ax_i + b - y_i)^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 + a^2\sum_{i=1}^{n}(x_i - \bar{x})^2 - 2a\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + n(\bar{y} - a\bar{x} + b)^2.$$
2. Identify b that minimizes $\sum_{i=1}^{n}(ax_i + b - y_i)^2$
3. Proceed to find a that minimizes $\sum_{i=1}^{n}(ax_i + b - y_i)^2$.

Question 4: 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).

1. Suppose the line that gives a best fit is:

$$y = ax + b$$

then the predicted value of y for x_i is given by $\hat{y}_i = ax_i + b$, and the *prediction errors* are defined by $d_i = y_i - \hat{y}_i$. Using answer found in Question 3, find the values of a and b which minimizes the error sum of squares which is $\sum_{i=1}^{n} d_i^2$.

(b) For the values of a and b obtained, draw the line $y = ax + b$ on the graph above and fill in the following table:

x	y	$\hat{y}$	d^2
2	6		
4	8.5		
1	2.5		
7	15		
5	11		

(c) Estimate the value of y when $x = 8, 3$.