M9L9.2 Data Distributions & Outliers

Objective: We will be able to use a data set to make a dot plot, identify outliers and the shape of a data distribution.

Vocabulary

Dot Plot- is a data representation that uses a number line and Xs, dots, or other symbols to show frequency.

Outliers- is a value in a data set that is much greater or much less than most of the other values in the data set. Outliers are determined by using the first or third quartiles and the IQR.

How to Identify an Outlier

A data value x is an outlier if $x < Q_1 - 1.5(IQR)$ or if $x > Q_3 + 1.5(IQR)$.
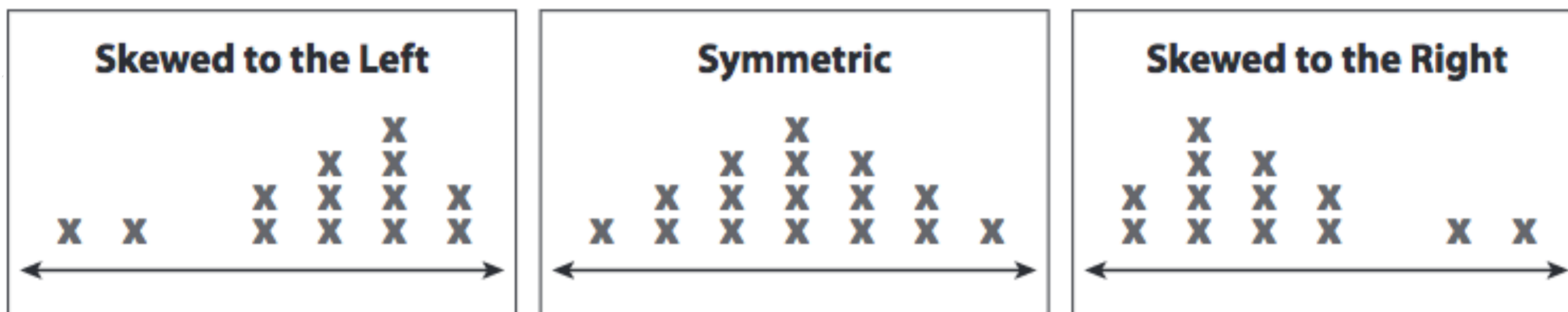
Vocabulary

Statistics- numbers that characterize a data set, such as measures of center and spread.

Symmetry- is a type of distribution where the left side of the distribution mirrors the right side

Skewed to the left- the long tail is on the left hand side.

Skewed to the right - the long tail is on the right hand side.

Skewed to the Left	Symmetric	Skewed to the Right

Twelve employees at a small company make the following annual salaries (in thousands of dollars): 25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, and 60.

**A** Choose the number line with the most appropriate scale for this problem. Explain your reasoning
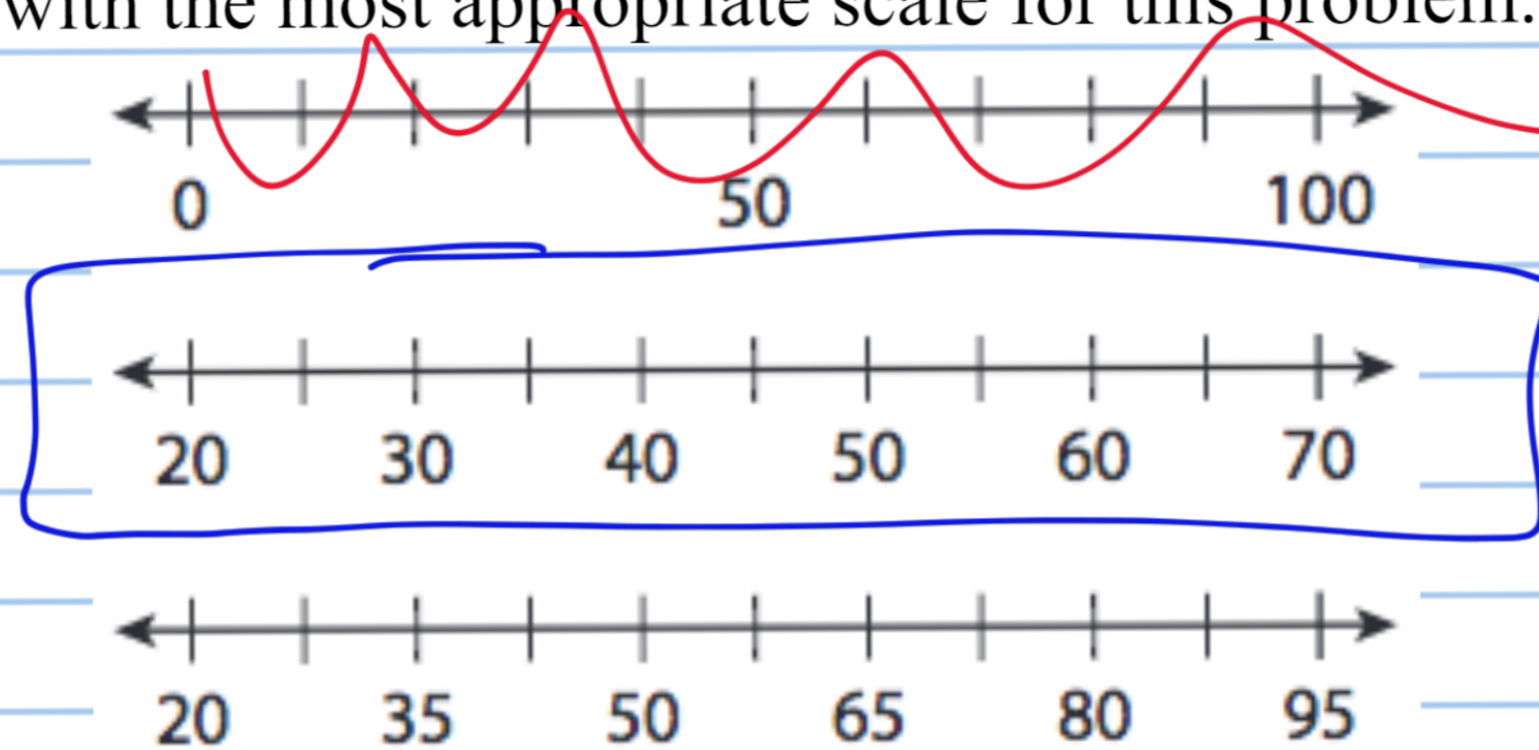
The 2nd # line has the most appropriate scale. The scale of the 1st # line include a large range of #'s that are not needed, so all the dots would be in the middle. The 3rd # line doesn't have convenient tick marks for determining there value between the labels belong.

Twelve employees at a small company make the following annual salaries (in thousands of dollars
25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, and 60.

B Create and label a dot plot of the data. Put an X above the number line for each time that
value appears in the data set.

Salary (thousands of dollars)
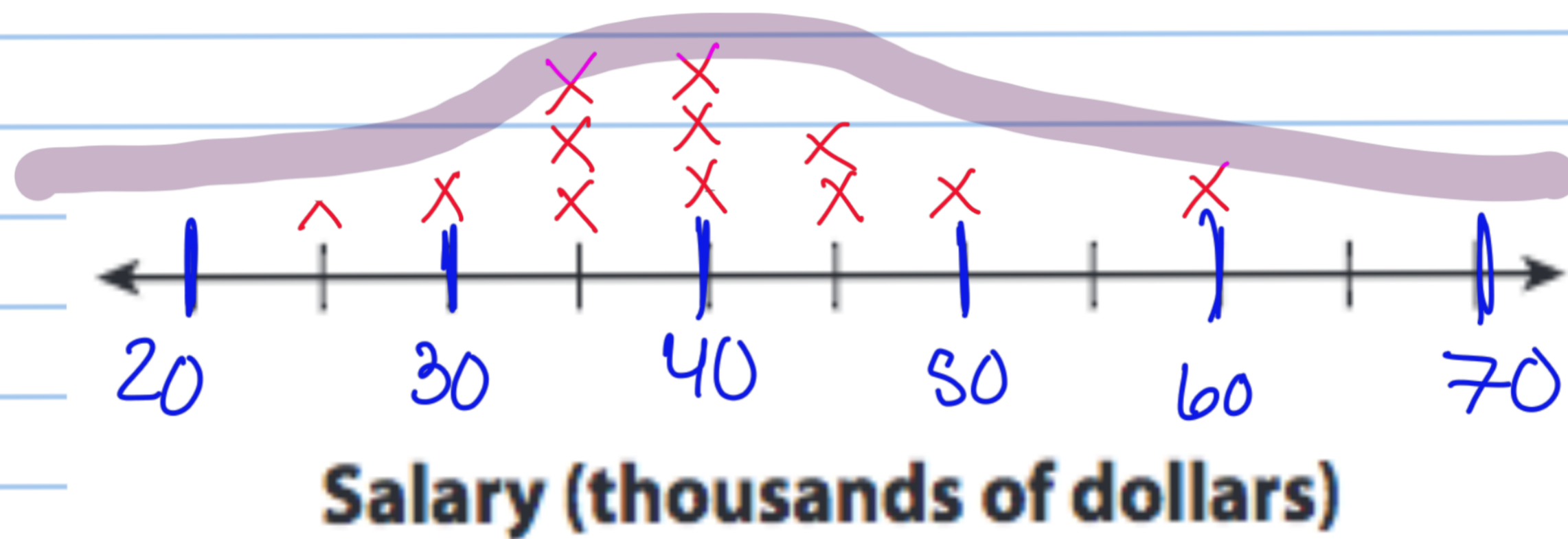
(2) Example Create a dot plot for the data set using an appropriate scale for the number line. Determine whether the extreme value is an outlier

(A) Suppose that the list of salaries from the Explore is expanded to include the owner's salary of $150,000. Now the list of salaries is 25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, and 150. To choose an appropriate scale, consider the minimum and maximum values, 25 and 150. A number line from 20 to 160 will contain all the values. A scale of 5 will be convenient for the data. Label tick marks by 20s. Plot each data value to see the distribution

Salary
(thousands of dollars)
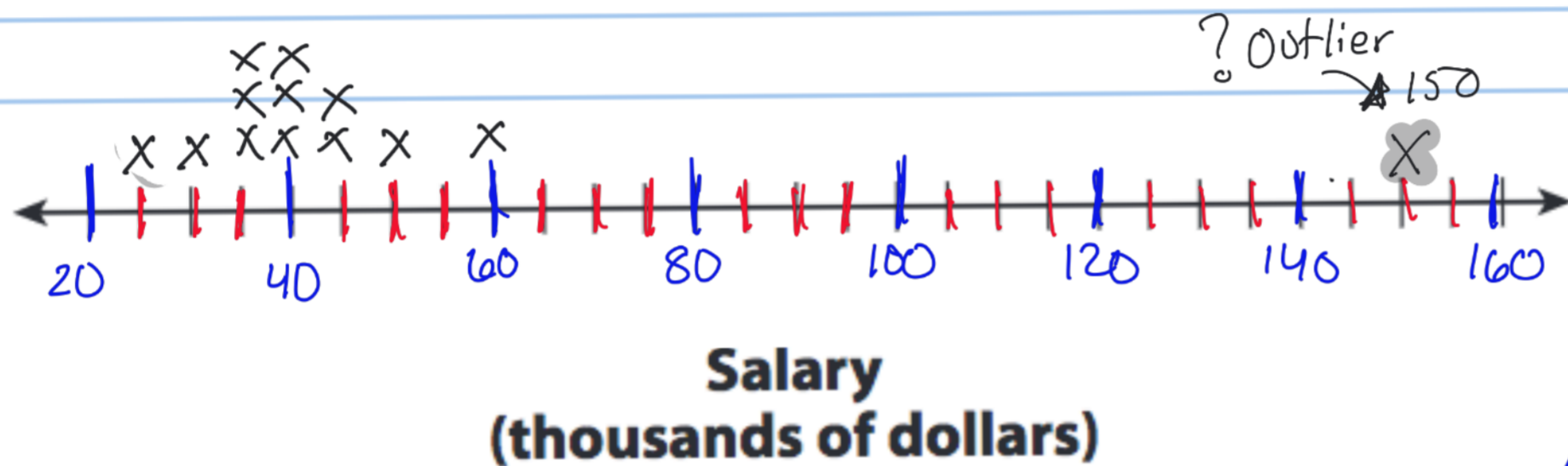
(B) Find the quartiles and the IQR to determine whether 150 is an outlier.

$$X > Q_3 + 1.5(IQR)$$

$$150 > 47.5 + 1.5(12.5)$$

$$150 > 47.5 + 18.75$$

$$150 > 66.25$$

$\boxed{\text{True}}$ / False

25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, 150

Q_2

$$Q_1 = \frac{35+35}{2}$$

$$Q_1 = 35$$

$$Q_3 = \frac{45+50}{2}$$

$$\boxed{Q_3 = 47.5}$$

$$IQR = Q_3 - Q_1$$
$$= 47.5 - 35$$
$$= 12.5$$

$\boxed{150 \text{ is an outlier}}$
b/c it was a true statement