Section 5.2
Correlation

The strength and direction association between two variables can be measured. This measure is called the **correlation coefficient** and is denoted by *r* (for a sample and R for a population).

Facts about Correlation:

1. $-1 \le r \le 1$
2. Positive *r* indicates positive association. Negative *r* indicates negative association.
3. The closer $|r|$ is to 1, the stronger the association. A weak association will have an *r* close to 0.
4. Correlation is strongly influenced by outliers. *Recall: Outliers are observations that are "distant" from the rest of the data.*

Example 1: Recall the following example from Section 5.1, Example 2:
The bivariate data given below relate the high temperature reached on a given day and the number of water bottles sold from a particular vending machine.

Temperature (in degrees)	Bottled Water (16 oz)
90	30
91	32
88	29
93	33
92	31
89	29
90	30
91	31
92	32
94	34

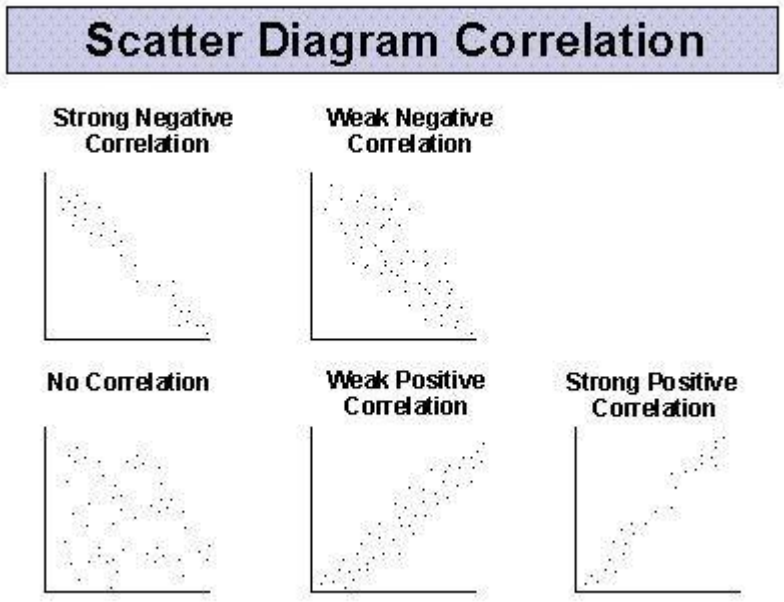
a. Find the correlation for the data.

In R, we created two lists:
>temp=c(90,91,88,93,92,89,91,92,94)
>bottles=c(30,32,29,33,31,29,30,31,32,34)

The command for correlation is cor(*x*,*y*).

Command: Answer:

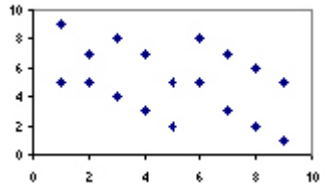
b. Does the data have a strong positive linear association?



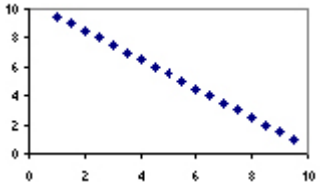
Example 2: Estimate the correlation coefficient for each scatter plot. Choose from:

$r = -1$ $r = -0.43$ $r = 0$ $r = 0.71$

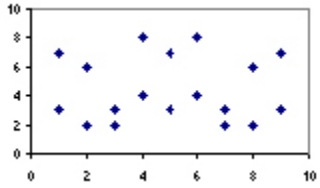
a.



b.



c.



d.