Learning Goal: Use a scatterplot to display the relationship between two quantitative variables. Describe the overall pattern and striking deviations from the pattern.

Specific Learning Objectives:

- Identify explanatory and response variables;
- Read and interpret scatterplots;
- Identify direction, strength and form in scatterplots.

Explanatory variable: In a situation where we have gathered information about two variables for each individual in a study, the explanatory variable may explain (or even cause) a change in another variable. For example, years of education may explain annual income. If an explanatory variable is quantitative explanatory variable it is often used to predict the value of the response variable. The explanatory variable is also known as the independent variable.

Response variable: This is the variable that measures the outcome of a study. It responds to, or may be dependent on, the explanatory variable. The response variable is also known as the dependent variable.

1) Identify the explanatory and response variables in each scenario. Jot down a few notes to capture your thinking.

   a) For a group of students, a teacher compares homework grades to exam grades.

   b) For a group of adults, a sociologist uses a survey to measure stress level and compares this to each person's credit card debt.

   c) For a group of people at a bar, a researcher measures blood alcohol content (BAC) and the number of alcoholic drinks consumed within an hour prior BAC measurement.

2) The explanatory variable is plotted on the horizontal axis (x-axis). The response variable is plotted on the vertical axis (y-axis). Identify the variable that should be plotted on the x-axis in each scenario:

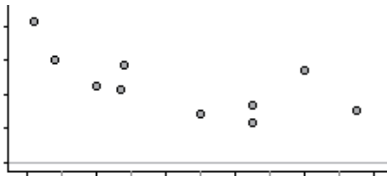
   a) Using census data, we plan to investigate the relationship between years of education and annual salary for a random sample of U.S. adults.

   b) Using the Movie data set from the Module 2 Project, we want to investigate whether the number of Oscars won by a movie is related to its opening weekend Box Office Sales.

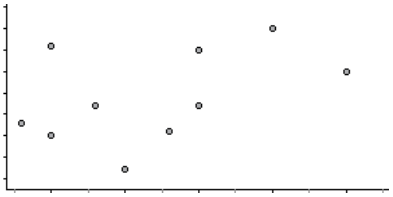
   c) Using data on breakfast cereals, we want to determine if sugar content is a good predictor of fat content.

3) The scatterplots below differ in the DIRECTION of the association. The direction is described as positive association or negative association (or neither).
   a) Match descriptions to scatterplots and jot down notes to capture your reasoning.

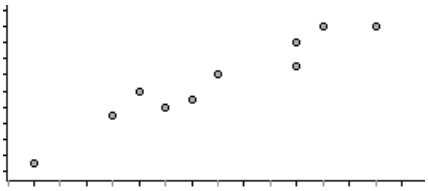
Scatterplot 1



Scatterplot 2



Scatterplot 3



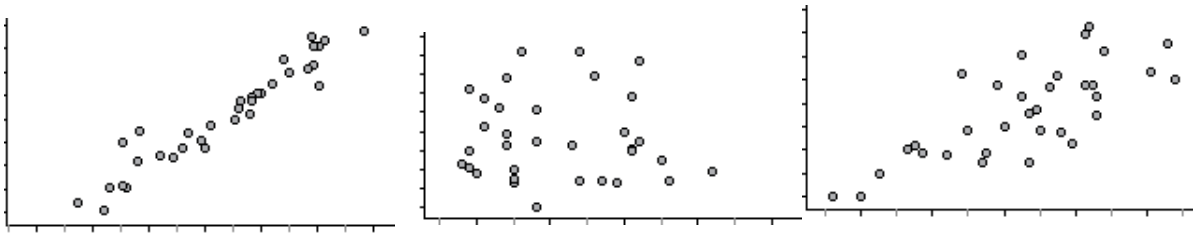
   i. x = city miles per gallons and y = highway miles per gallon for 10 cars
   ii. x = sodium (mg/serving) and y = Consumer Report rating for 10 brands of tomato soup
   iii. x = price ($) and y = sodium (mg/serving) for 10 brands of vegetable soup

   b) Based on the scatterplots, label each association as a positive association, a negative association, or neither.
   - The association between city mpg and highway mpg for a sample of cars
   - The association between sodium in a serving and Consumer Report rating for different brands of tomato soup
   - The association between price and sodium in a serving for different brands of vegetable soup

4) The scatterplots below differ in the STRENGTH of the association. The strength is how closely the data follow a pattern.

a) These scatterplots show body measurements for 34 of adults who are physically active. Match each description to a scatterplot. Briefly explain your reasoning.

Scatterplot 1 *Scatterplot 2* *Scatterplot 3*



i. x = forearm girth (cm), y = bicep girth (cm)
ii. x = calf girth (cm), y = bicep girth (cm)
iii. x = age (years), y = bicep girth (cm)

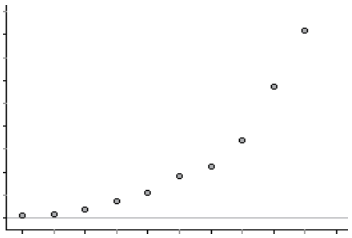
b) Based on the scatterplots, label each association as a strong association, a fairly strong association, no association (or a very weak association.)

- The association between forearm girth and bicep girth
- The association between calf girth and bicep girth
- The association between age and bicep girth

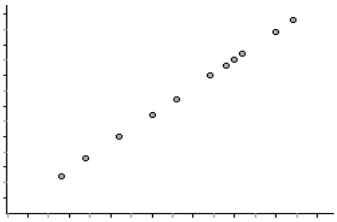
5) These scatterplots differ in FORM. Some have a linear pattern; some have a curved pattern (also called curvilinear).

a) Match each set of measurements to a scatterplot. Briefly explain your reasoning.

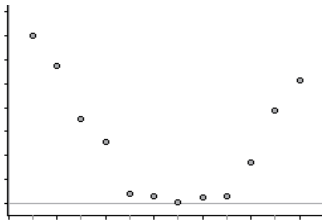
Scatterplot 1



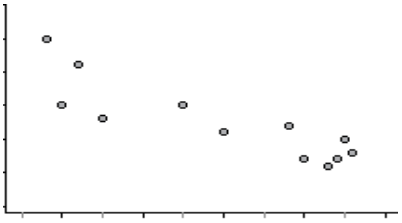
Scatterplot 2



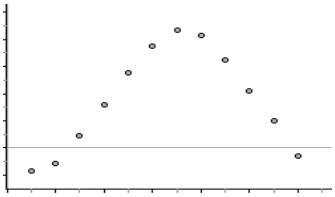
Scatterplot 3



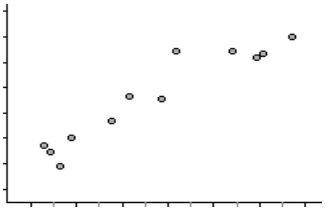
Scatterplot 4



Scatterplot 5



Scatterplot 6



i. x = month number (January = 1) and y = rainfall (inches) in Napa CA. Napa has several months of drought each summer.

ii. x = month number (January = 1) and y = average temperature in Boston MA. Boston has cold winters and hot summers.

iii. x = year (from 1965) in 5-year increments and y = Medicare expenditures ($). Medicare costs have increased over time. In addition, the increases in costs have themselves been increasing over time. (Small increases at first, followed by larger increases as time passed.)

iv. x = average temperature (°C) and y = average temperature (°F) each month in San Francisco

v. x = chest girth (cm) and y = shoulder girth (cm) for a sample of men.

vi. x = engine displacement (in liters) and y = city miles per gallon for a sample of cars. Engine displacement is roughly a measurement of the size of the engine. Large engines use more gas.

b) Label each scatterplot as linear pattern or curvilinear pattern.