Name _____   Date _____

1. The Kentucky Derby is a horse race held each year. The following scatter plot shows the speed of the winning horse at the Kentucky Derby each year between 1875 and 2012.
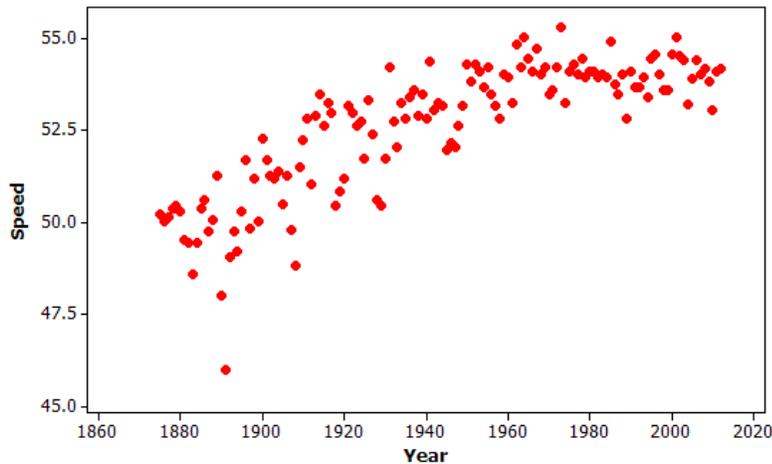
a. Is the association between *speed* and *year* positive or negative? Give a possible explanation in the context of this problem for why the association behaves this way considering the variables involved.

b. Comment on whether the association between *speed* and *year* is approximately linear and then explain in the context of this problem why the form of the association (linear or not) makes sense considering the variables involved.

c. Circle an outlier in this scatter plot and explain, in context, how and why the observation is unusual.

2. A group of students were asked to report their gender and how many times a day they typically wash their hands. Of the 738 males, 66 said they wash their hands at most once a day, 583 said two to seven times per day, and 89 said eight or more times per day. Of the 204 females, two said they wash their hands at most once a day, 160 said two to seven times per day, and 42 said eight or more times per day.

   a. Summarize these data in a two-way table with rows corresponding to the three different frequency-of-hand-washing categories and columns corresponding to gender.

   b. Do these data suggest an association between *gender* and *frequency of hand washing*? Support your answer with appropriate calculations.

3. Basketball players who score a lot of points also tend to be strong in other areas of the game, such as number of rebounds, number of blocks, number of steals, and number of assists. Below are scatter plots and linear models for professional NBA (National Basketball Association) players last season.

$$y = 21.54 + 3.833x$$

$$y = 294.9 + 22.45x$$

$$y = 98.03 + 9.558x$$

$$y = 166.2 + 2.256x$$
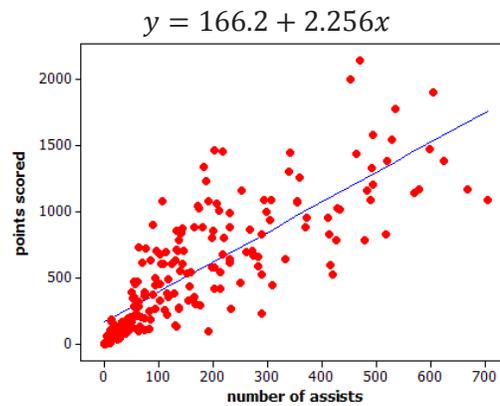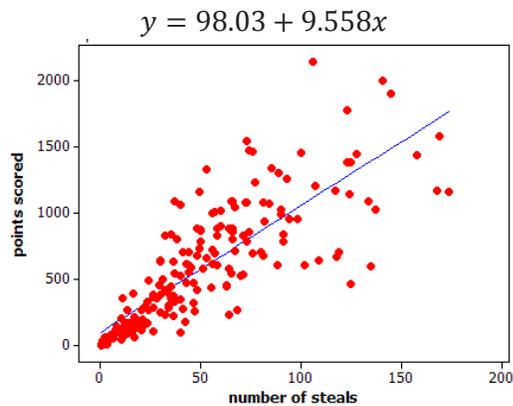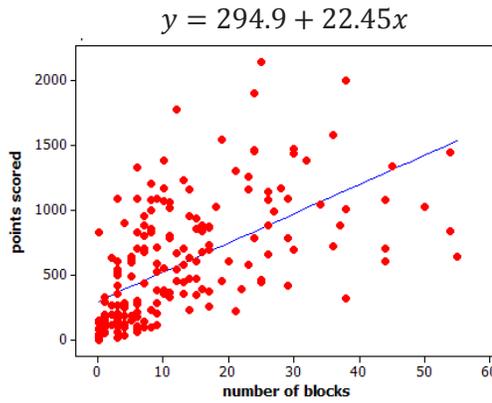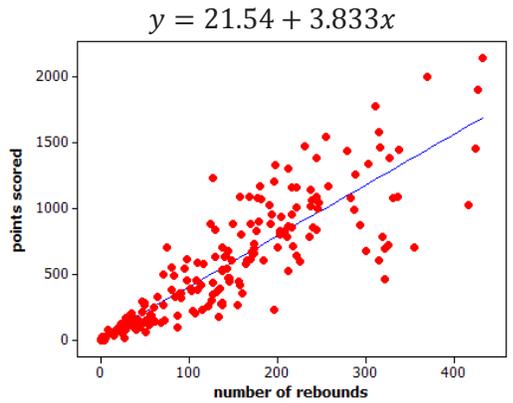
a. The line that models the association between points scored and number of rebounds is $y = 21.54 + 3.833x$, where $y =$ points scored and $x =$ number of rebounds. Give an interpretation, in context, of the slope of this line.

b. The equations above all show $y =$ number of points as a function of the other variables. An increase in which of the variables (rebounds, blocks, steals, and assists) tends to have the largest impact on the predicted points scored by an NBA player?

c. Which of the four linear models shown in the scatter plots above has the worst fit to the data? Explain how you know using the data above.