Suggested readings and practice problems from <u>OpenIntro Statistics, 3rd edition</u> (a free online introductory statistics textbook co-authored by Dr. Cetinkaya-Rundel) for this week:

**Suggested reading:** Chapter 7, Section 7.1, 7.2

Suggested exercises: (End of chapter exercises from OpenIntro Statistics)

- Relationship between two numerical variables: 7.1, 7.3, 7.7, 7.9, 7.11, 7.13, 7.15

(Reminder: the solutions to the end of chapter exercises are at the end of the *OpenIntro Statistics* book)

Test yourself:

1. A teaching assistant gives a quiz. There are 10 questions on the quiz and no partial credit is given. After grading the papers the TA writes down for each student the number of questions the student got right and the number wrong. What is the correlation of the number of questions right and wrong? Hint: Make up some data for number of questions right, calculate number of questions wrong, and plot them against each other.

2. Suppose you fit a linear regression model predicting score on an exam from number of hours studied. Say you've studied for 4 hours. Would you prefer to be on the line, below the line, or above the line? What would the residual for your score be (0, negative, or positive)?

3. Someone hands you the scatter diagram shown below, but has forgotten to label the axes. Can you calculate the correlation coefficient? Or do you need the labels?
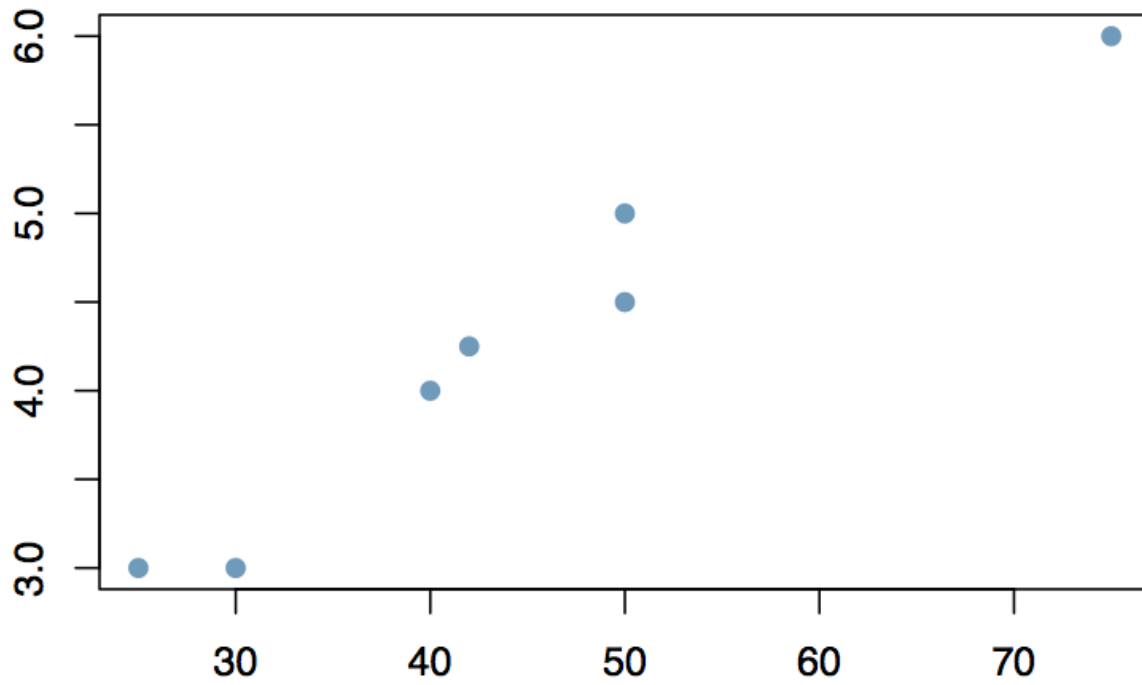
4. We would not want to fit a least squares line to the data shown in the scatterplot below. Which of the conditions does it appear to violate?
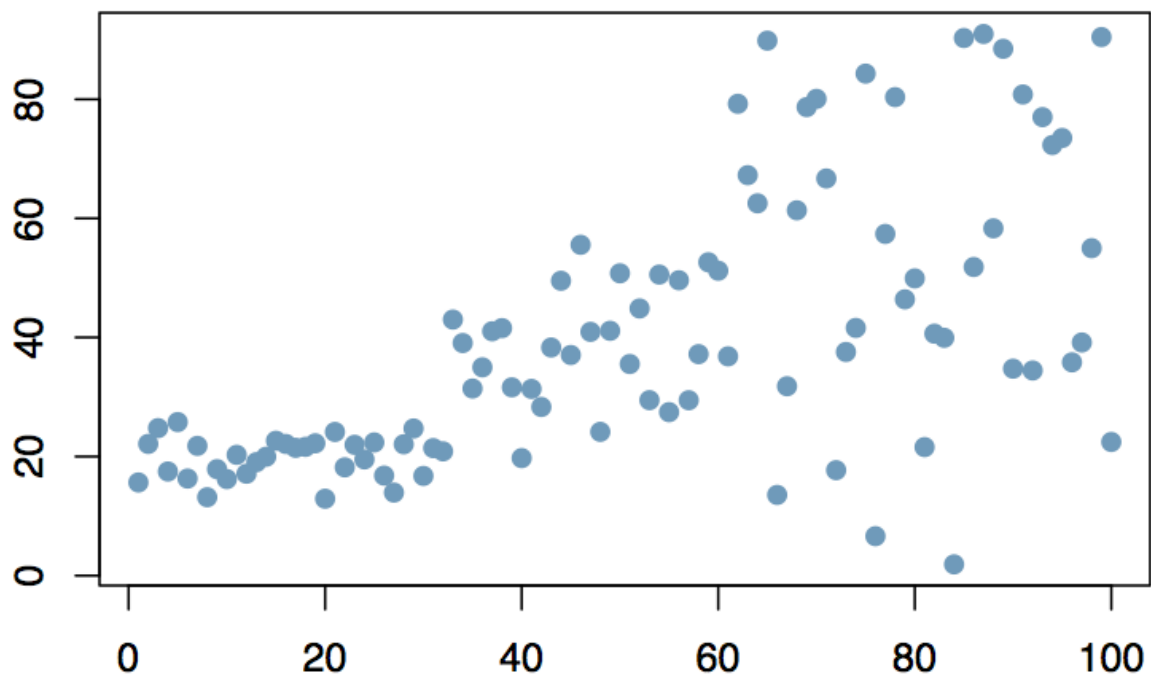
5. Derive the formula for b_0 given the fact that the linear model is $\hat{y} = b_0 + b_1 \times x$ and that the least squares line goes through $(\bar{x}, \bar{y})$.

6. One study on male college students found their average height to be 70 inches with a standard deviation of 2 inches. Their average weight was 140 pounds, with a standard deviation of 25 pounds. The correlation between their height and weight was 0.60.

Assuming that the two variables are linearly associated, write the linear model for predicting weight from height.

7. Is a male who is 72 inches tall and who weighs 115 pounds on the line, below the line, or above the line?

8. Describe what is an indicator variable, and what levels 0 and 1 mean for such variables.

9. If the correlation between two variables y and x is 0.6, what percent of the variability in y does x explain?

10. The model below predicts GPA based on an indicator variable (0: not premed, 1: premed). Interpret the intercept and slope estimates in context of the data.

$$\widehat{gpa} = 3.57 - 0.01 \times premed$$

Mark as completed