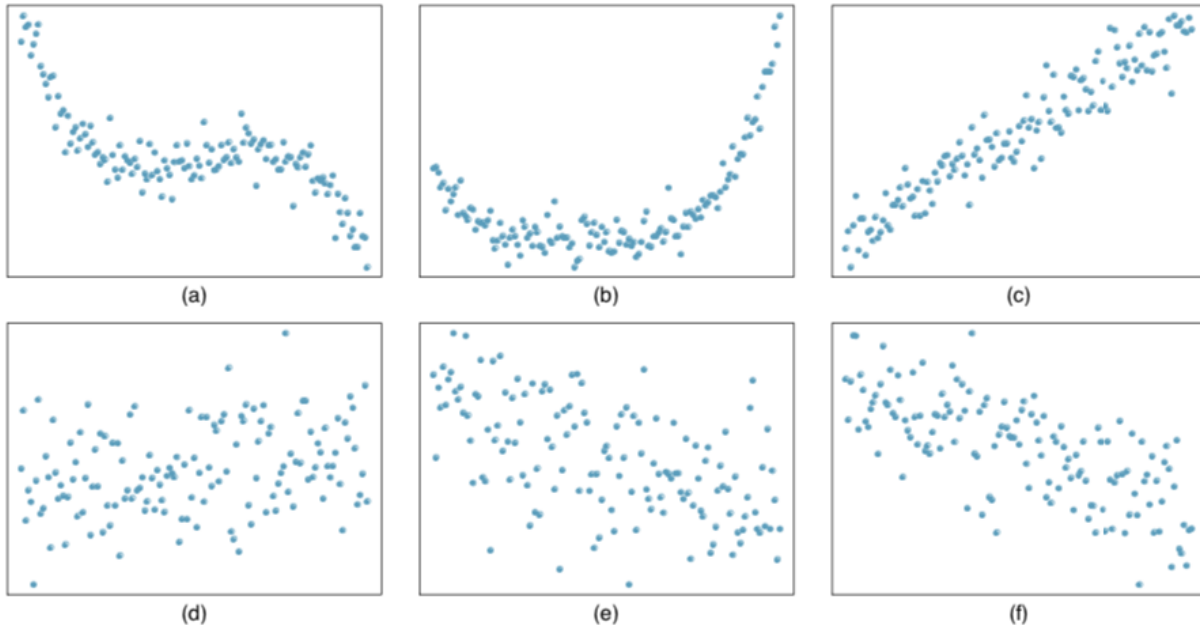


Note: “OIS” refers to OpenIntro Statistics, the textbook for this class.

Problem 1: OIS Exercise 8.4, page 313 (6 pts)

For each of the six plots, identify the strength of the relationship (e.g., weak, moderate, or strong) in the data and whether fitting a linear model would be reasonable.



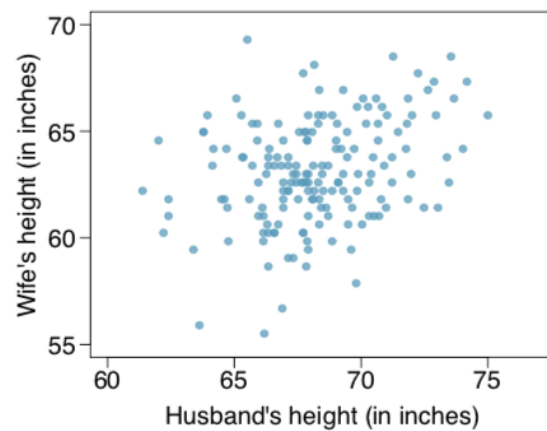
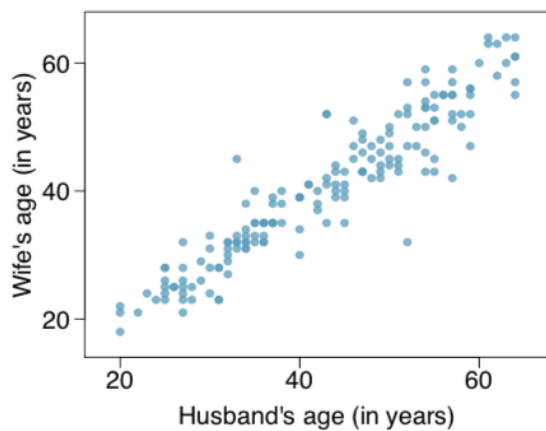
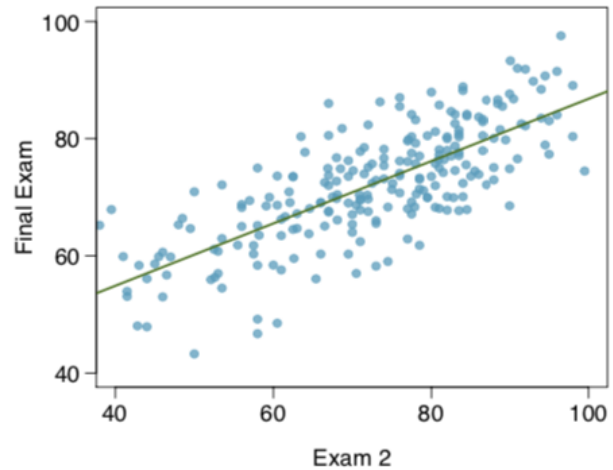
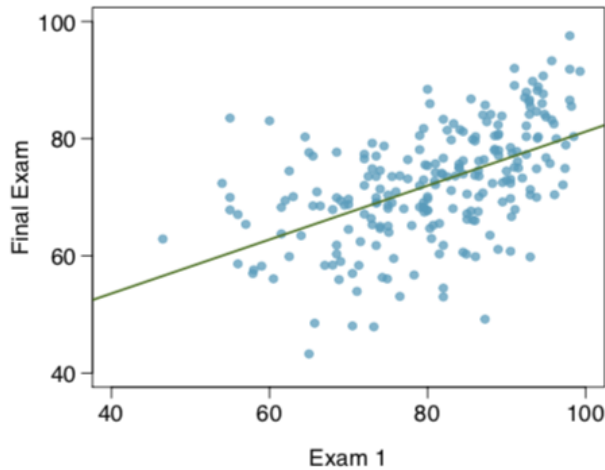
Problem 2: OIS Exercise 8.5, page 313 (2 pts)

The two scatterplots below show the relationship between final and mid-semester exam grades recorded during several years for a Statistics course at a university.

- (a) (1 pt) Based on these graphs, which of the two exams has the strongest correlation with the final exam grade? Explain.
- (b) (1 pt) Can you think of a reason why the correlation between the exam you chose in part (a) and the final exam is higher?

Problem 3: OIS Exercise 8.6, page 313 (4 pts)

The Great Britain Office of Population Census and Surveys once collected data on a random sample of 170 married couples in Britain, recording the age (in years) and heights (converted here to inches) of the husbands and wives.⁵ The scatterplot on the left shows the wife’s age plotted against her husband’s age, and the plot on the right shows wife’s heights plotted against husband’s height.



- (a) (1 pt) Describe the relationship between husbands' and wives' ages.
- (b) (1 pt) Describe the relationship between husbands' and wives' heights.
- (c) (1 pt) Which plot shows a stronger correlation? Explain your reasoning.
- (d) (1 pt) Data on heights were originally collected in centimeters, and then converted to inches. Does this conversion affect the correlation between husbands' and wives' heights?

Problem 4: OIS Exercise 8.20, page 325 (2 pts)

Suppose we fit a regression line to predict the number of incidents of skin cancer per 1,000 people from the number of sunny days in a year. For a particular year, we predict the incidence of skin cancer to be 1.5 per 1,000 people, and the residual for this year is 0.5. Did we over or under estimate the incidence of skin cancer? Explain your reasoning.

Problem 5: OIS Exercise 8.24, page 326 (6 pts)

Exercise 8.13 introduces data on shoulder girth and height of a group of individuals. The mean shoulder girth is 107.20 cm with a standard deviation of 10.37 cm. The mean height is 171.4 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.67.

- (a) Write the equation of the regression line for predicting height (2 pts).
- (b) Interpret the slope and the intercept in this context (2 pts).
- (c) A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model (1 pt).
- (d) The student from part (c) is 160 cm tall. Calculate the residual, and explain what this residual means (1 pt).

Problem 6 (5 pts)

The `babies` dataset includes data on all pregnancies between 1960 and 1967 among women in the Kaiser Foundation Health Plan in the San Francisco East Bay area.¹ The dataset includes the variables `bwt`, describing each baby's birthweight in ounces, and `gestation`, describing the length of the pregnancy in days.

```
babies <- read.csv("https://www.openintro.org/data/csv/babies.csv")
```

- (a) Using `lm()`, fit a linear regression model predicting birthweight from gestation (3 pts).
- (b) Write the equation of the line using mathematical notation (2 pts).

¹<https://www.openintro.org/data/index.php?data=babies>