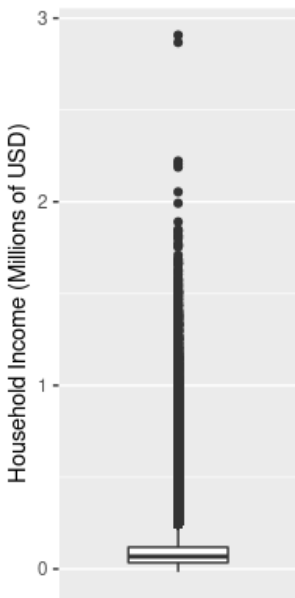The 2019 Public Use Microdata Sample (PUMS) from US Census Bureau provides data about household income for 1,276,716 households. This distribution is depicted to the right.

1. Which of the following accurately describes the distribution?

(A) The distribution is left skewed.

(B) The distribution is right skewed.

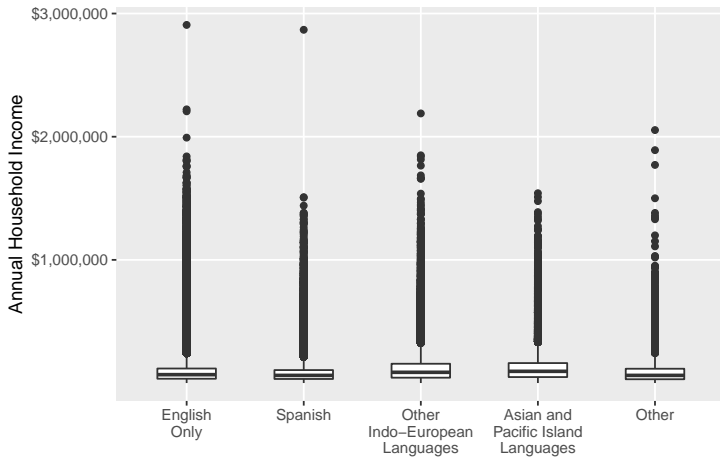(C) The distribution has no skew.

(D) None of the above.

The same data set by the US Census Bureau also tells us the household language for all of the 1,276,716 households.
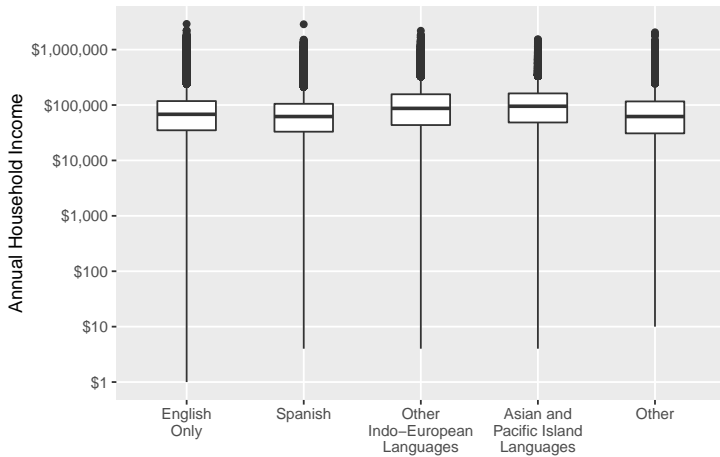
2.   If we wanted to compare the distribution of household income based on household language, we might use a...

(A) Scatter plot.

(B) Side-by-side bar plot.

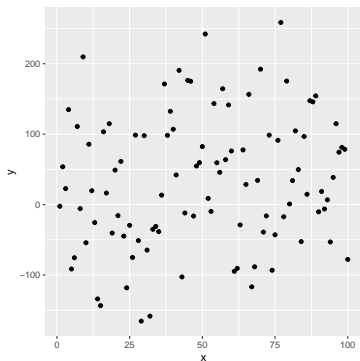(C) Mosaic plot.

(D) None of the above.
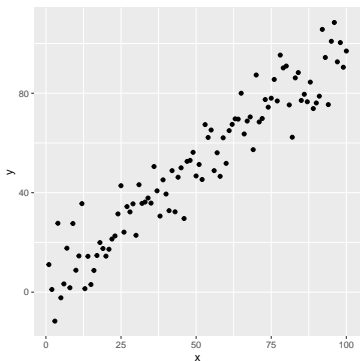
If you're curious...

A slightly clearer visualization...?
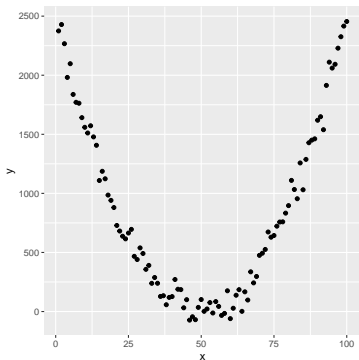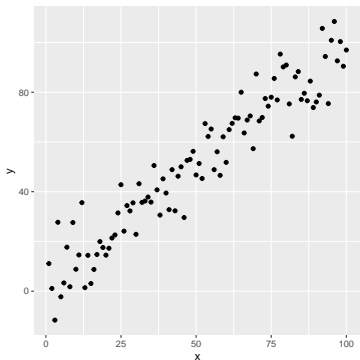
5. True or False?

The correlation of x and y in the scatterplot on the left is larger than the correlation of x and y in the scatterpoint on the right.

6. True or False?

The correlation of x and y in the scatterplot on the left is larger than the correlation of x and y in the scatterpoint on the right.
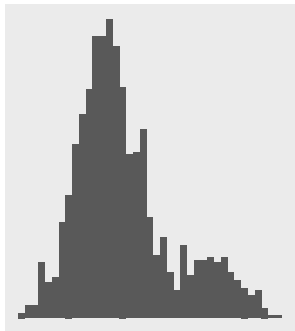
A histogram of a numerical variable x is depicted to the right.
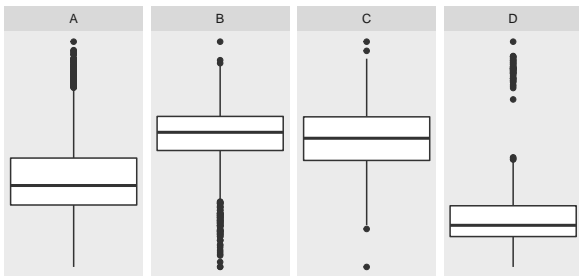

7.   Which of the following is true about x?

(A) The mean is greater than the median.

(B) The median is greater than the mean.

(C) Can't say / None of the above.

A histogram of a numerical variable x is depicted to the right.

8. Which of the box plots below might be a box plot of x?

9.   Suppose that, at a certain firm, men make 25% more than women do in any given position. Then the best fit line relating women's salaries to men's salaries has slope...

(A) 0.25

(B) 1.25

(C) 4

(D) None of the above

10.   Suppose that, at a certain firm, men make $5000 more than women do in any given position. Then the best fit line relating women's salaries to men's salaries has slope...

(A) 0

(B) 1

(C) 5000

(D) None of the above

11. Starbucks lists the calorie content of food menu items but not carbohydrate content. Suppose we gather data on some Starbucks food items and construct a best fit line to predict the carhobydrate content (in grams) using the calorie content. Then the slope of the best fit line has units...
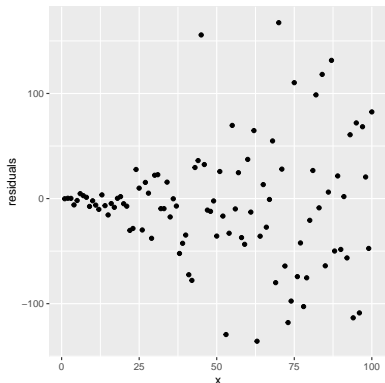
(A) g

(B) cal

(C) g/cal

(D) None of the above

Follow-up. Use the units to give an explicit interpretation of the slope of this line.

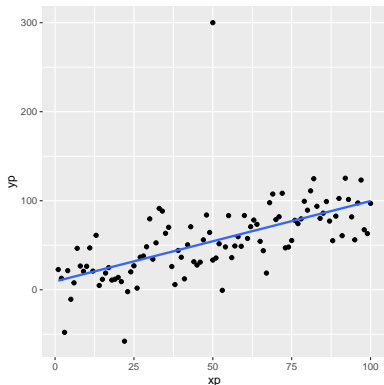The plot to the right represents the *residuals* after we fit a least squares regression to some data.

12.   Should we have concerns about applying least squares regression to this data?

(A) Yes

(B) No

(C) I don't know

13.   Which of the following is true about the depicted outlier?

(A) It doesn't have high leverage.

(B) It has high leverage but isn't influential.

(C) It has high leverage and is influential.
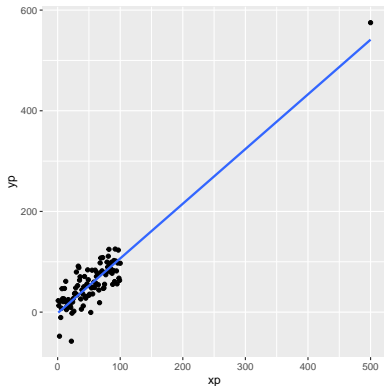
14. Which of the following is true about the depicted outlier?

(A) It doesn't have high leverage.

(B) It has high leverage but isn't influential.

(C) It has high leverage and is influential.

15. Which of the following is true about the depicted outlier?

(A) It doesn't have high leverage.

(B) It has high leverage but isn't influential.

(C) It has high leverage and is influential.