Week 2 Monday

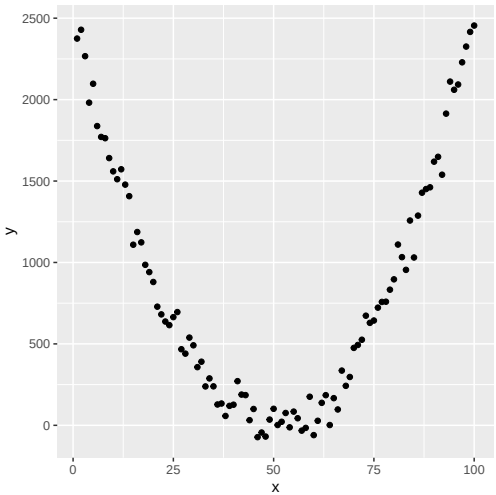
Anywhere

Turn to someone sitting near you who you haven't met before (or who you haven't talked to much before), and introduce yourself! Then discuss:

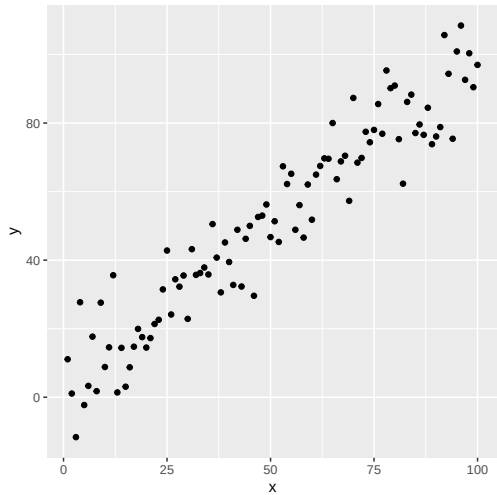
If you could live anywhere on this planet, where would you choose to live and why?

Summarizing Data 2

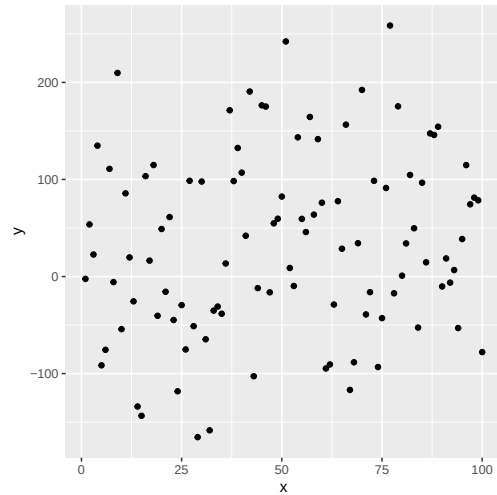
1. In which of the following three scatterplots do x and y have the greatest correlation?



(A)

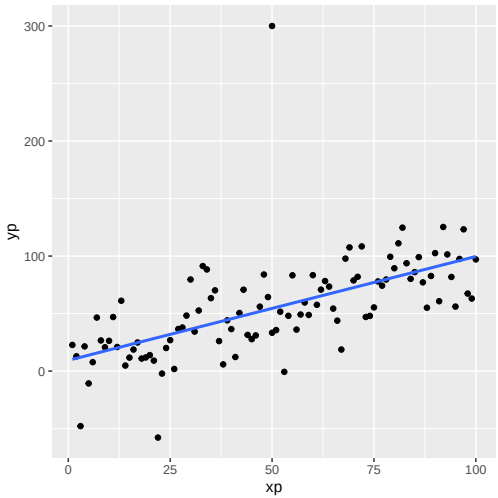


(B)

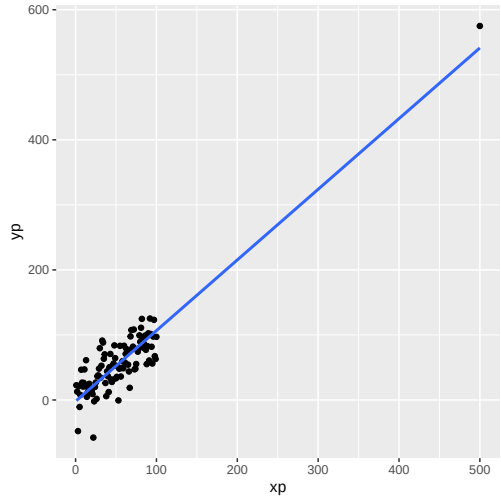


(C)

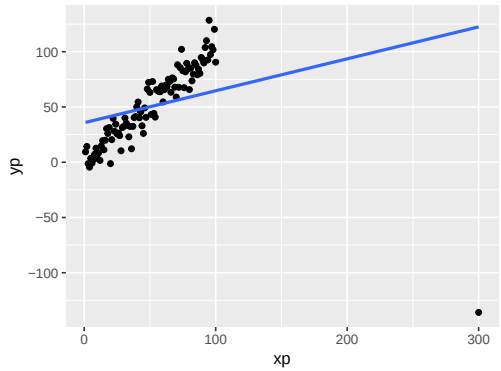
2. In each of the following scatterplots, there is an outlier.



(I)



(II)



(III)

Which of the following is accurate?

(A) The outlier in (I) has high leverage.

(B) The outliers in (II) and (III) have high leverage.

(C) The outlier in (II) is influential.

(D) None of the above OR more than one of the above.

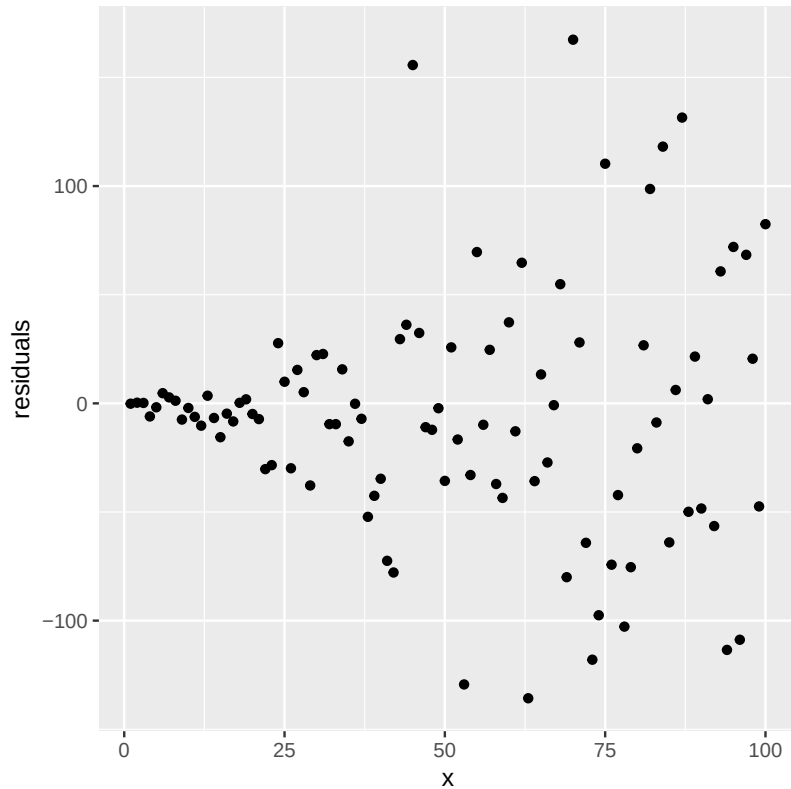
The plot to the right represents the *residuals* after we fit a least squares regression to some data.

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

(A) Yes

(B) No

(C) I don't know



4. The alien inhabitants of a distant planet have two arms, but each alien's left arm is roughly 25% longer than their right arm.

If we make a scatterplot of arm lengths in centimeters, with right-arm-length on the x-axis and left-arm-length on the y-axis, the slope of the best fit line will be approximately:

(A) 0.25

(B) 1.25

(C) 4

(D) None of the above

5. The alien inhabitants of a distant planet have two arms, but each alien's left arm is roughly 15 cm longer than their right arm.

If we make a scatterplot of arm lengths in centimeters, with right-arm-length on the x-axis and left-arm-length on the y-axis, the slope of the best fit line will be approximately:

(A) 0

(B) 1

(C) 15

(D) None of the above

6. Suppose we fit a best fit line to predict the daily number of surfing injuries in San Diego County using the daily dollar amount of ice cream sales.

The slope of the regression line has has units:

(A) injuries

(B) dollars

(C) injuries/dollar

(D) None of the above

7. Suppose we fit a best fit line to predict the daily number of surfing injuries in San Diego County using the daily dollar amount of ice cream sales.

The slope of the regression line has has units:

(A) injuries

(B) dollars

(C) injuries/dollar

(D) None of the above

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

8. Suppose we fit a best fit line to predict the daily number of surfing injuries in San Diego County using the daily dollar amount of ice cream sales.

For April 1st, 2023, the regression predicts the 10.5 surfing injuries, and the residual for was 2.5. Which of the following is accurate?

(A) The regression line overestimated the number of surfing injuries.

(B) The regression line underestimated the number of surfing injuries.

(C) The regression line predicted the correct number of surfing injuries.