14.12. The data set used to make the scatterplot shown in the text (Figure 14.12, Moore and Notz, pp 333) used data only up through 2016. The manatee data we looked at in the notes went up through 2018.

(a) Let's interpret the scatterplot in terms of our four characteristics:

- Form: There is a linear (straight-line) relationship between the number of boats registered and the number of manatee deaths.
- Direction: The relationship is positive.
- Strength: The linear relationship is strong.
- Deviations: There are no outliers.

(b) The point A has

$$\begin{aligned} x &= \text{number of boats registered (in thousands)} \approx 800 \\ y &= \text{number of manatees deaths} \approx 70. \end{aligned}$$

(c) If there was an observation at point B, then this would be an outlier. The number of manatee deaths would be unusually low for that many boats registered.

14.19. (a) There is no obvious distinction of the explanatory variable and the response variable in this example. There is no strong preference between asking how

- the homicide rate depends on the suicide rate
- the suicide rate depends on the homicide rate.

(b) For the no-caution only counties, the scatterplot is on the next page (top).
(c) There is a moderate negative linear relationship between the homicide rate and the suicide rate for the no-caution counties. The correlation is $r = -0.47$.

```
> cor(no.caution$homicide,no.caution$suicide)
[1] -0.47
```

(d) The scatterplot of all 26 counties is on the next page (bottom). I used different plotting symbols for the

- no caution counties (colored circle)
- caution counties (circles that are not colored in).

Now, it looks like there is weak-to-moderate positive linear relationship between the two variables. In fact,

```
> cor(ohio$homicide,ohio$suicide)
[1] 0.23
```

Indeed, the correlation between the two variables is now positive ($r = 0.23$) when including all the counties. Interesting!

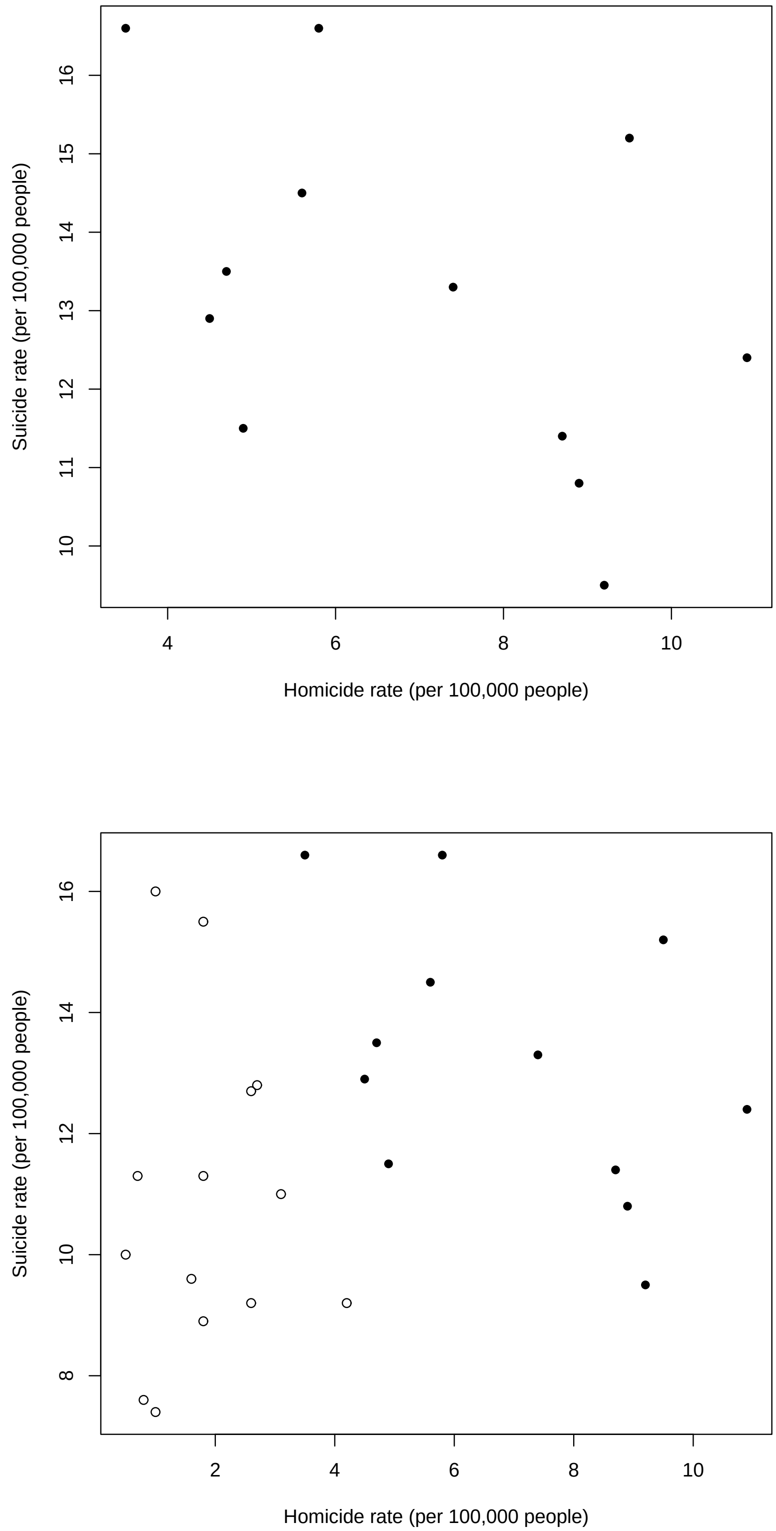


Figure 1: Ohio data. Top: Suicide rates and homicide rates for no-caution only counties. Bottom: All counties are included.

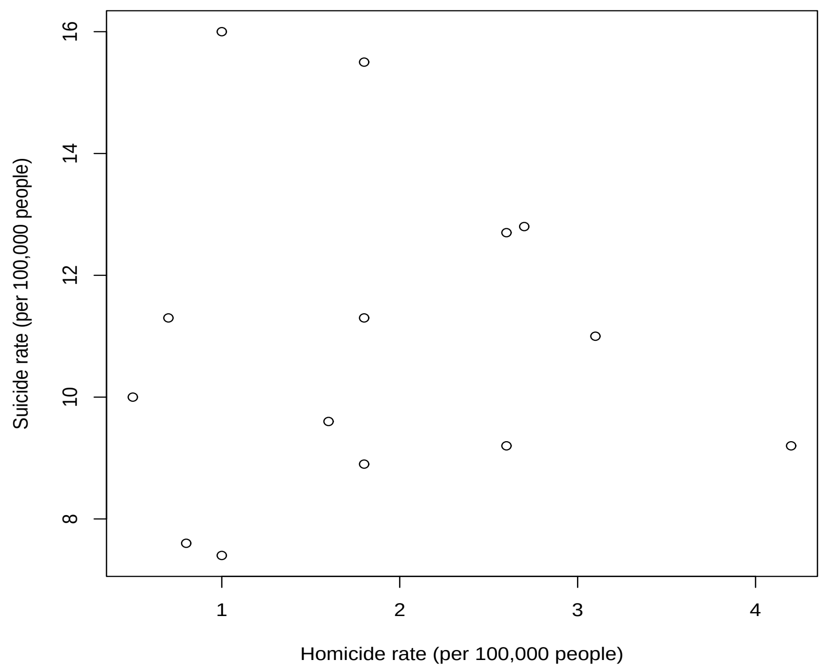


Figure 2: Ohio data. Suicide rates and homicide rates for caution-only counties.

14.23. This is a continuation of Problem 14.19. The scatterplot of the homicide rate and the suicide for the caution-only counties is shown above. There is no linear relationship between homicide rate and suicide rate for caution-only counties. In fact, the correlation is $r = 0.02$.

```
> cor(caution$homicide,caution$suicide)
[1] 0.02
```

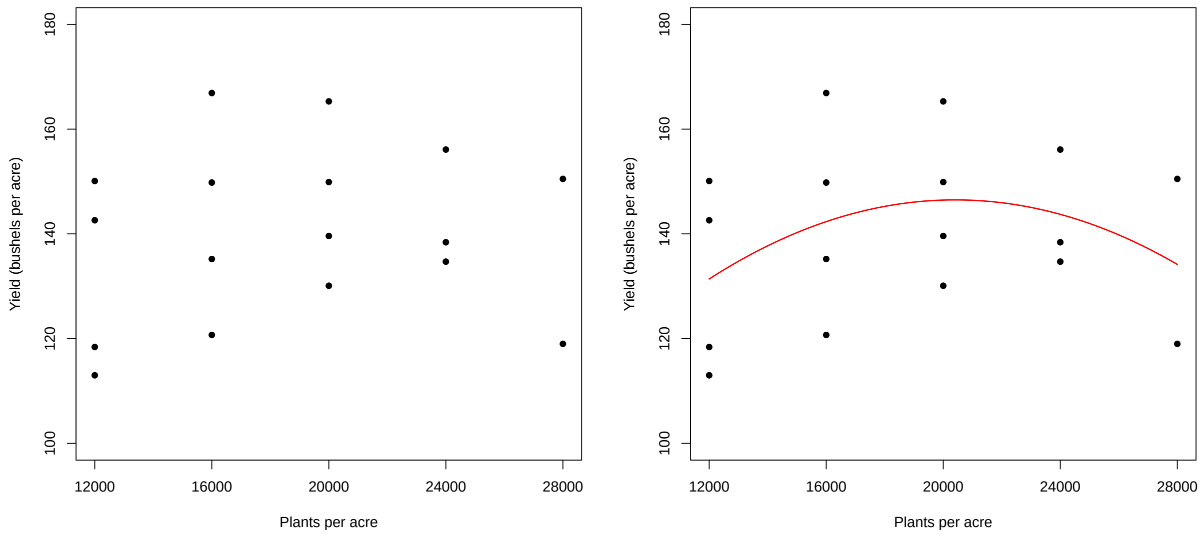
This correlation is very close to zero, which means no linear relationship between the variables.

In Problem 14.19, we calculated $r = -0.47$ for the no-caution counties. Here, $r = 0.02$ for the caution counties. *This means the level of caution is a lurking variable.* The strength and direction of the linear relationship between homicide rate and suicide rate depends whether you are looking at counties with no caution or counties with caution.

14.27. The professor's statement means that research productivity and teaching ratings are not linearly related to each other. The school's newspaper writer doesn't understand correlation! S/he is misinterpreting the professor's statement as "research productivity and teaching ratings are negatively related."

14.28. There are errors in each statement. Here they are:

(a) The correlation is only for two quantitative variables. Car manufacturer is categorical.



(b) The correlation cannot exceed +1.

(c) The correlation is unitless. It does not have units attached to it. It would just be $r = 0.53$.

14.34. (a) The yield is the response variable (y) and the plants per acre is the explanatory variable (x). It makes sense to think about hot the yield depends on the plants per acre.
(b) The scatterplot of the plants per acre and the yield is shown above (left).
(c) The relationship between the two variables does not appear to be a straight line. In fact, there is more of a curved relationship between them. I emphasized this curved relationship on the right. Because the relationship is curved, I suspect the correlation between the two variables is close to zero:

```
> cor(plants,yield)
[1] 0.13
```

Therefore, the correlation is pretty close to zero (I thought it would be closer). Remember, the correlation does not describe curved relationships.

It looks like increasing the number of plants per acre (x) does increase yield, but the yields at 24000 and 28000 are not as high as one would expect if this relationship applied "forever." This makes sense. If the acre plots are saturated with too many plants, this may extract valuable resources from the soil—leading to a lower yield (too many plants are competing on the same plot). The scatterplot above suggests this may be happen when the plots contain 24000 and 28000 plants.