3. (a) `sum(sample(c(rep(1, 13), rep(0, 39)), size = 5)`

(b) `gf_boxplot(~ sqFt, data = houses)`

(c) `lm(gasMileage ~ weight, data = cars)`

(d) `names(houses)`

(e) `cor(gasMileage ~ weight, data = cars)`

(f) `filter(houses, sqFt > 2000)`

4. Standard deviation, correlation and mean are all sensitive to outliers.

5. (a) The distribution is unimodal, left-skewed, with a possible outlier out far in the left tail.

(b) The mean is $<$ (less than) the median.

6. If you took the mean as stated, $\bar{x} = 10$, your calculation should be

$$s = \sqrt{\frac{1}{3}\left[(11-10)^2 + (9-10)^2 + (3-10)^2 + (13-10)^2\right]} = \sqrt{\frac{1}{3}(1+1+49+9)}$$

$$= \sqrt{20} \doteq 4.472.$$

If you took $\bar{x} = 9$, as it actually is, then

$$s = \sqrt{\frac{1}{3}\left[(11-9)^2 + (9-9)^2 + (3-9)^2 + (13-9)^2\right]} = \sqrt{\frac{1}{3}(4+0+36+16)}$$

$$= \sqrt{56/3} \doteq 4.320.$$

7. (a) As the description says, data is kept for each Michigan city. Michigan cities are the cases.

(b) What variable is measured on cities is the proportion (paid on time)/(tickets assigned). As these denominators are highly variable, this is a continuous variable.

8. (a) The correlation is close to 1.

(b) It is the point near (28, 25).

(c) $\hat{y} = (0.4497)x + 16.8964$

(d) At $x = 22.1$, the predicted value is $\hat{y} = (0.4497)(22.1) + 16.8964 \doteq 26.835$.

So, the residual is $y - \hat{y} = 27.9 - 26.835 \doteq 1.065$.

9. (a) The score distribution is the same whether looking at Group A or Group B. Knowing group does not help in predicting score, so the two variables — group and score — have no association.

(b) While affiliation with Group C seems unaffected by sex, the same cannot be said of affiliation with Groups A and B. So, there is an association between sex and group.

(c) There is no discernible pattern — nothing that would make knowledge of the x-value useful for predicting the corresponding y-value. x and y are not associated.

10. Estimating $Q_1 = 40$ and $Q_3 = 53$, we get $IQR = 53 - 40 = 13$.

11. (a) $Pr(X \geq 4) = Pr(X = 4) + Pr(X = 6) = 0.4 + 0.2 = 0.6$

(b) $$Pr(X \geq 4 \mid X \geq 2) = \frac{Pr(X \geq 4 \text{ and } X \geq 2)}{Pr(X \geq 2)} = \frac{Pr(X \geq 4)}{Pr(X \geq 2)} = \frac{0.6}{0.25 + 0.4 + 0.2} =$$

$$= \frac{12}{17} \doteq 0.706$$

(c) $$E(X) = (0)(0.15) + (2)(0.25) + (4)(0.4) + (6)(0.2) = 3.3$$

12. $$Pr(\text{Red}) = Pr(\text{Red and A}) + Pr(\text{Red and B})$$
$$= Pr(\text{Red} \mid A)\, Pr(A) + Pr(\text{Red} \mid B)\, Pr(B)$$
$$= \left(\frac{2}{6}\right)\left(\frac{4}{7}\right) + \left(\frac{4}{6}\right)\left(\frac{2}{8}\right)$$
$$= \frac{5}{14} \doteq 0.3571$$