1. (a) tally(region ~ sex, data = survey) or tally(~ region | sex, data = survey)

(b) $(25+89+19)/268 = 133/268$

(c) $25/268$

(d) $(23+95+17+19)/268 = 154/268$, or $\frac{135}{268} + \frac{36}{268} - \frac{17}{268} = \frac{154}{268}$

(e) $25/48$

(f) $(25+89)/(25+89+19) = 114/133 = 6/7$

(g) These events are not independent, since

$$\Pr(\text{female} \mid \text{rural}) = 25/48 \doteq 0.521, \text{ but } \Pr(\text{female}) = \frac{133}{268} \doteq 0.496.$$

2. (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.

3. (a) Estimating $Q_1 = 40$ and $Q_3 = 53$, we get $IQR = 53 - 40 = 13$.

(b) (c) $\}$

2(a) gf_boxplot(score ~ group, data = prob3)

2(b) gf_props(~ group | sex, data = prob3)

2(c) gf_point(y ~ x, data = prob3)

4. (a) qnorm(0.95, 70, 12)

(b) Scores = read.csv("http://alldat.com/Scores.csv")

(c) 1 - pexp(10, 0.1)

(d) filter(iris, Petal.Length ≥ 5)

5. (a) $0.5 = F(x) = \frac{1}{16}(12x - x^3)$.

(b) $\Pr(0.5 < x < 1) = F(1) - F(\tfrac{1}{2}) = \frac{1}{16}\left[12 - 1 - (6 - \tfrac{1}{8})\right] = 481/768$

(c) $E(X) = \int_0^2 x \cdot \frac{3}{16}(4 - x^2)\,dx = \frac{3}{16}\int_0^2 (4x - x^3)\,dx = \frac{3}{16}\left[2x^2 - \frac{1}{4}x^4\right]_0^2 = \frac{3}{4}$.

6. (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.

7. We have

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

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

This answer can be obtained from

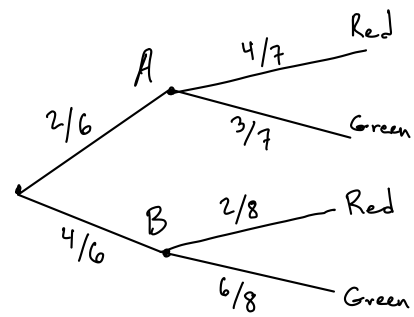
sd(~ c(11, 9, 3, 13))

8. (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.

9. (a)

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



(b)
$$\Pr(A \mid \text{Red}) = \frac{\Pr(A \text{ and Red})}{\Pr(\text{Red})}$$
$$= \frac{(2/6)(4/7)}{5/14} = \left(\frac{14}{5}\right)\left(\frac{1}{3}\right)\left(\frac{4}{7}\right) = \frac{8}{15} \doteq 0.533.$$