1. (a) tally(class ~ oncampus, data = survey) or tally(~class | oncampus, data = survey)

(b) $(16 + 26) / 268 = 42/268$

(c) $16/268$

(d) $(2 + 82 + 16 + 87 + 6)/268 = 193/268$, or $\frac{191}{268} + \frac{84}{268} - \frac{82}{268} = \frac{193}{268}$

(e) $16/42$

(f) $(16 + 87 + 6)/(82 + 16 + 87 + 6) = \frac{109}{191}$

(g) These events are not independent, since
$\Pr(\text{on campus} \mid \text{junior}) = \frac{16}{42} \doteq 0.381$, different from $\Pr(\text{on campus}) = \frac{191}{268} \doteq 0.713$.

2. (a) The group distribution is the same whether looking at females or males. Knowing sex does not help in predicting group, so the two variables — sex and group — have no association.

(b) There is a noticeable pattern in the scatterplot — an arc trend you would exploit to make your highest y-predictions for middling x-values, your lowest y-predictions for the highest x-values. x and y have an association.

(c) You would predict a higher score for a Group B case than for a Group A case. Group and score have an association.

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

(b), (c)

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

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

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

(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

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

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

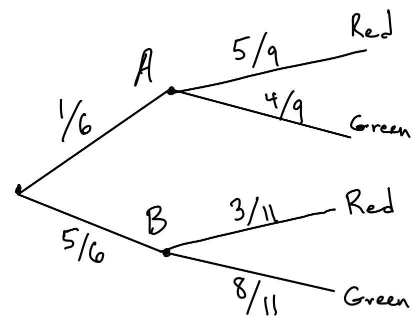
This answer can be obtained from

```
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)



$$\Pr(\text{Red}) = \Pr(\text{Red and A}) + \Pr(\text{Red and B})$$

$$= \Pr(\text{Red} \mid A)\Pr(A) + \Pr(\text{Red} \mid B)\Pr(B)$$

$$= \left(\frac{1}{6}\right)\left(\frac{5}{9}\right) + \left(\frac{5}{6}\right)\left(\frac{3}{11}\right)$$

$$= \frac{190}{594} \doteq 0.320$$

(b) $$\Pr(A \mid \text{Red}) = \frac{\Pr(\text{A and Red})}{\Pr(\text{Red})}$$

$$= \frac{(1/6)(5/9)}{190/594} = \left(\frac{5}{54}\right)\left(\frac{594}{190}\right) = \frac{11}{38} \doteq 0.289$$