1. (a) There is no indication that researches imposed drinking habits. They merely asked about existing habits and experiences. The study is observational.

(b) There are 3135 participants classified as "frequent bingers". So

$$\Pr(\text{frequent binger}) \doteq \frac{3135}{11160} = \frac{209}{744} \doteq 0.2809.$$

(c) There are 3320 participants who have missed class for this reason:

$$\Pr(\text{has missed class}) = \frac{3320}{11160} = \frac{83}{279} = 0.2975.$$

(d)

$$\Pr(\text{nonbinger or has not missed}) = \Pr(\text{nonbinger}) + \Pr(\text{has not missed})$$
$$- \Pr(\text{nonbinger and has not missed})$$
$$= \frac{5063}{11160} + \frac{7840}{11160} - \frac{4617}{11160} = \frac{8286}{11160} \doteq 0.742$$

(e) This probability was used once already, in part (d):

$$\Pr(\text{nonbinger and has not missed}) = \frac{4617}{11160} \doteq 0.414$$

(f) This is the conditional probability

$$\Pr(\text{has missed class} \mid \text{Occasional binger}) = \frac{\Pr(\text{has missed class and Occasional binger})}{\Pr(\text{Occasional binger})}$$
$$= \frac{915/11160}{2962/11160} = \frac{915}{2962} \doteq 0.309$$

2. (a) Perhaps "Is there an association between an ICU patient's systolic blood pressure and whether that patient survives?"

(b) This one works

$$\text{gf\_boxplot}(\text{Systolic} \sim \text{status, data} = \text{ICUAdmissions})$$

(c) The numbers R gives (and what our eyes are trying to estimate) are

$$36, \quad 89, \quad 126, \quad 140, \quad 256$$

(d) The outlier around 50 is more than $1.5 \times$ IQR ($\approx 63$) below $Q_1$, the first quartile. The one near 225 is more than $1.5 \times$ IQR above $Q_3$.

(e) The possibility of being flagged as an outlier exists only for the upper one — the one near 225. We can only know for certain by adding $1.5 \times$ IQR $\approx (1.5)(50) = 75$ to 140. If estimated correctly, any observation above 215, which seems to include this one, would be flagged.

3. (a) Since the conditions under which measurement of words (counts of them) are taken are dictated by the researcher and comprise the very thing being studied as the explanatory variable (it appears), this is an experiment.

(b) There is a definite pattern of higher $y$'s paired with higher $x$'s, lower $y$'s paired with lower $x$'s. This is a positive association.

(c) The actual correlation coefficient for this data (what our eyes attempt to estimate) is 0.764.

(d) The line, as drawn by most people, will give the largest residual to subject # 8, the one whose point is (164, 215).

4. (a) The mean is around 21. The standard deviation is about 6.

(b) Using my numbers, the $Z$-score is
$$Z = \frac{29 - 21}{6} \doteq 1.667.$$

5. (a)
$$Pr(\text{two green tokens}) = \left(\frac{5}{8}\right)\left(\frac{5}{8}\right) = \frac{25}{64} \doteq 0.391.$$

(b)
$$Pr(\text{two green tokens}) = \left(\frac{5}{8}\right)\left(\frac{4}{7}\right) = \frac{5}{14} \doteq 0.357.$$