2.19 a. Whether a person supports the smoking ban or not is a categorical variable.
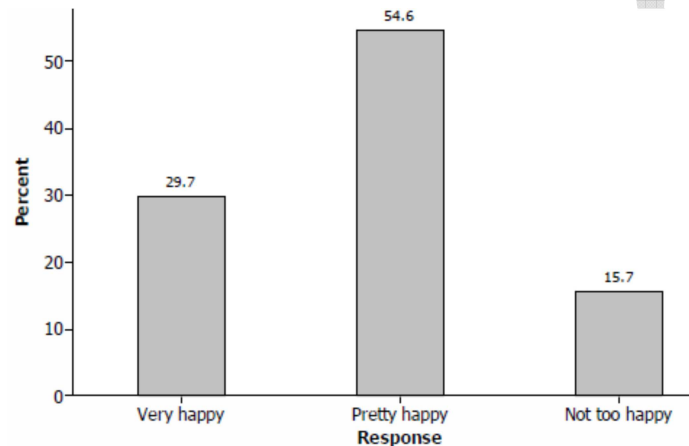**b.** Gains on verbal and math SATs are both quantitative variables.

2.28 a.

Response	Frequency	Relative frequency
Very happy	599	599/2015 = .297 (29.7%)
Pretty happy	1100	1100/2015 = .546 (54.6%)
Not too happy	316	316/2015= .157 (15.7%)
Total	2015	1 (100%)

b.

c. 29.7% + 54.6% = 84.3%

2.39 a. The center for the females is at a greater percentage than it is for the males. For females the center looks to be somewhere around 27%. For males, the center looks to be a bit less than 18%.
b. The data are more spread for the females.
c. The greatest two female percentages are set apart from the bulk of the data. The values are about 65% and 72%.

**2.43 a.** The dataset looks approximately symmetric and bell-shaped.
b. There are no noticeable outliers.
**c.** The most frequently reported value for sleep was 7 hours.
d. Roughly 14 or so students said they slept 8 hours the previous night.

2.44 a. The histogram is shown below. Notice that both the frequency(axis on the left) and the corresponding relative frequency/percentage (axis on the right) are shown here.
**b.** A majority of the sentences have between 16 and 30 words. The percentage of sentences with between 16 and 30 words is (113+107+109)/600 ≈55%. With regard to shape, the data are skewed to the right.
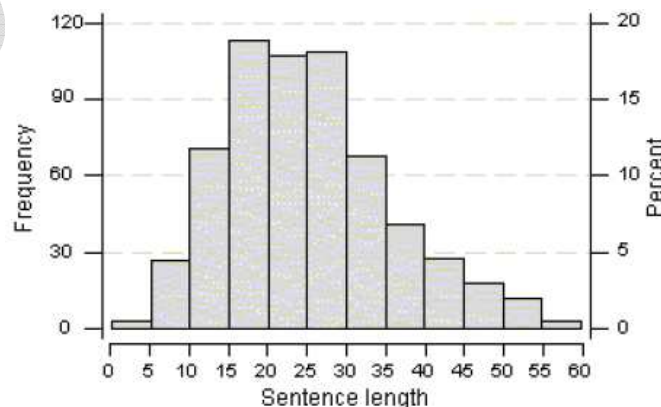
2.47 b.
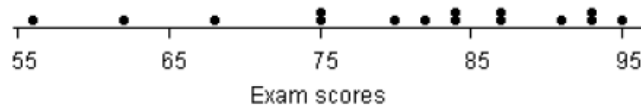
2.52 a. The answer will vary due to the flexibility possible for deciding on the endpoints of intervals. The histogram shown here is based on 5-year age groups and the age under the left edge of a bar is included in that interval while the age under the right edge is not.
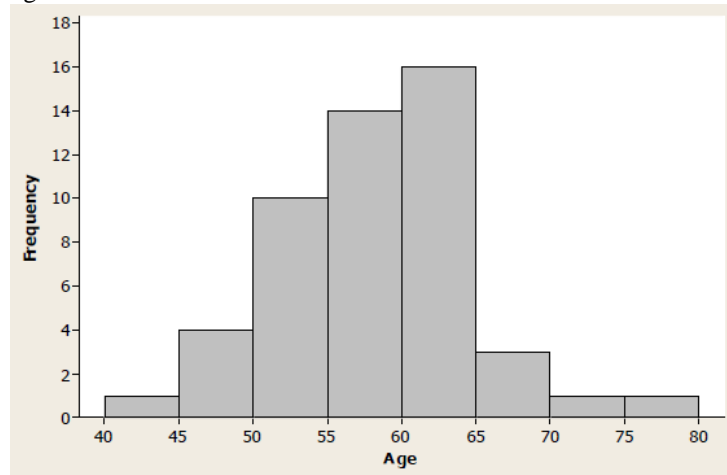
2.62 a. 12 letters.
b. 13 letters.
c. The IQR for males is $17 – 10 = 7$ while for females it is $15 -10 = 5$. The IQR is larger for males.
d. $23 – 6 = 17$ letters.
e. $23 – 6 = 17$ letters.

2.86 a. Mean =20; Standard deviation = 1.581.
b. Mean =20; Standard deviation = 0.
c. Mean = 20; Standard deviation = 33.09.
[Enter the data into a list in your calculator and use STAT / Calc / 1-VarStats]

Web problem - IMDB ratings
Note that the IMDB ratings are updated daily, therefore small variations can occur depending on the day the IMDB website was consulted. www.imdb.com/title/tt1628841/ratings   www.imdb.com/title/tt1289401/ratings

a. While the two movies have similar average ratings, their ratings distributions are drastically different.
-The distribution of ratings for *Independence Day Resurgence* is single-peaked, roughly symmetric, with a center around 6. This is reflected in the median of 6 and mean of 5.4.
-The distribution of ratings for *Ghostbusters* is multiple-peaked, with one sharp peak at 1, another sharp peak at 10, and another, wider peak around roughly 7. Therefore, the mean and median (5.5 and 5, respectively) do not reflect the properties of the ratings distribution for this movie, and are misleading measures of center.

b. The ratings distributions for male and for female raters of the movie *Independence Day Resurgence* are fairly similar. In contrast, the ratings distributions for male and for female raters of the movie *Ghostbusters* are radically different (nearly polar opposites). The distribution is bimodal for male raters, with a sharp peak at 1 and a wider peak around 7. The distribution is left-skewed to mildly bimodal for female raters, with mostly strong ratings.

c. The two movies have similar average ratings while having radically different ratings. Raters were rated *Independence Day Resurgence* pretty consistently as a fairly mediocre movie, and this was not influenced by the sex of the rater. Ratings for *Ghostbusters* varied greatly to the point of being polarized, with male and female raters driving to a large extend this extreme variability. It is also worth noting that, for both movies, there were many more

ratings by male than by female raters. Therefore, average ratings (which are the only rating information reported each the dedicated movie page) can mask great disparities in ratings distributions and tend to represent male raters more than female raters.
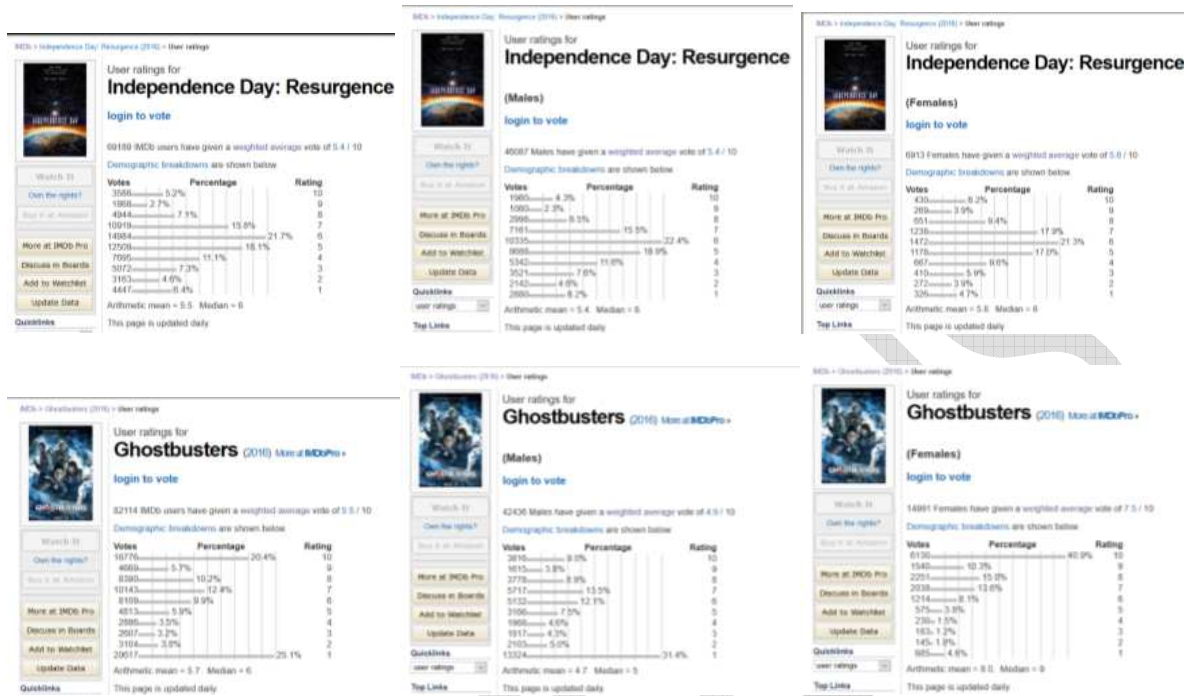
[Note: It is also important to note that raters are self-selected (not randomly selected) and therefore they are not representative of any larger population beyond themselves.]