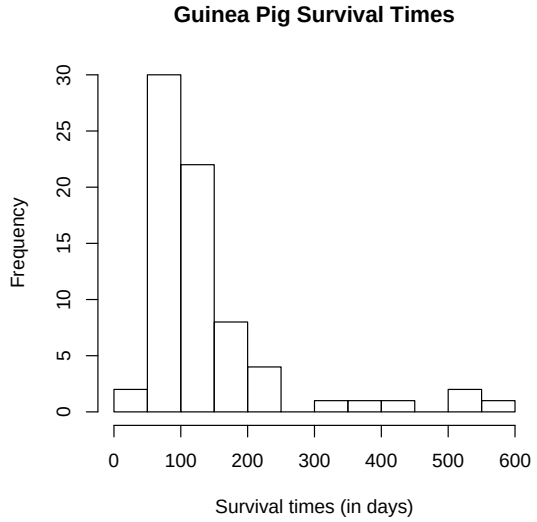
Stat 160 Fall 2008
Solutions to Assignment #2

#2.34 (a) In order to graph the distribution, it is fine to use either a histogram or a stemplot.

A histogram of the survival times (in days) of the 72 guinea pigs under consideration is shown below. As noted on Assignment #1, in order to eliminate any ambiguity in the histogram, it is crucial to include the class definitions. For this particular histogram, the classes are "open on the left, closed on the right" as indicated below.



class	count
0 <survival time≤ 50	2
50 <survival time≤ 100	30
100 <survival time≤ 150	22
ETC.	

From the histogram, the expected right-skew is clearly evident. The main peak occurs from 50 to 150 days—the guinea pigs that lived more than 500 days seem to be outliers.

A stemplot of the survival times (in days) of the 72 guinea pigs under consideration is shown below. It is rounded to the nearest tens and drawn with split stems to emphasize the skewness by showing the gaps.

0	4
0	5 5 6 6 6 6 7 7 7 7 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9
1	0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 2 2 2 3 3 4 4 4 4
1	5 5 6 6 7 8 8 8 9
2	0 1 1 4
2	5
3	3
3	8
4	0
4	
5	1 2
5	
6	0

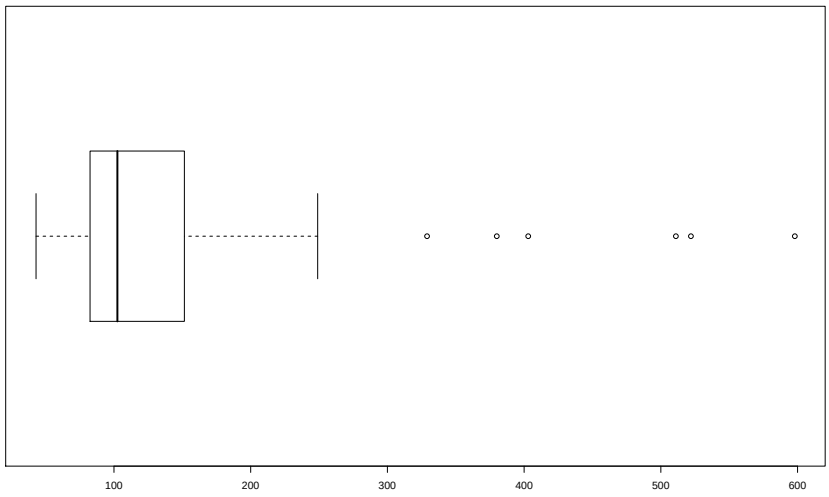
As with the histogram, the stemplot indicates that the main peak occurs from 50 to 150 days, and the guinea pigs that lived more than 500 days seem to be outliers.

#2.34 (b) Following the *Choosing a Summary* box on page 50, we choose the five-number summary since the distribution is skewed. In this case we find

$$(\min, Q_1, M, Q_3, \max) = (43,\ 82.5,\ 102.5,\ 151.5,\ 598)$$

all measured in days. The difference between Q_3 and the maximum is relatively much larger than the other differences between successive numbers. This indicates a large spread among the high observations. That is, it shows that the data are skewed to the right.

A boxplot (not required) also shows this right-skew.



#2.36 Here is a brief solution following the method described on pages 53–55.

State: Is bone mineral loss greater among the breast-feeding women?

Formulate: We need to compare the distributions, including appropriate measures of centre and spread.

Solve: Shown below are two histograms for the percent change in mineral content. The one on the left is for breast-feeding women, while the one on the right is for other women.

In both cases, the classes are "open on the left, closed on the right" as defined in the solution to Exercise 2.34.

Both distributions appeared to be slightly skewed to the right, and so it is appropriate to summarize both distributions with the five-number summary. (However, it is possible that with a different choice of histogram class width or class definition, a distribution will appear more symmetric, in which case the mean and standard deviation are the appropriate summary statistics. In either situation, it is imperative that you describe the skewness/symmetry and then choose the appropriate summary. Again, refer to the *Choosing a Summary* box on page 50 for details.) In any case, these seven numbers are given below for each group of women.

	$\overline{x}$	s	min	Q_1	M	Q_3	max
Breast-feeding women	-3.6%	2.5%	-8.3%	-5.3%	-3.8%	-2.1%	2.2%
Other women	0.3%	1.3%	-2.2%	-0.4%	-0.05%	1.1%	2.9%

Conclude: Both the histograms and the numerical summaries suggest that breast-feeding women lose bone mineral when compared with women of a similar age who are neither pregnant nor lactating.

#2.44 (a) The five-number summary (in 1999 US dollars) is

$$(\min, Q_1, M, Q_3, \max) = (0,\ 2.14,\ 10.64,\ 40.96,\ 88.6).$$

The evidence for the skew is in the large gaps between the higher numbers; that is, the differences $Q_3 - M$ and $\max - Q_3$ are large compared to $Q_1 - \min$ and $M - Q_1$.

#2.44 (b) We calculate the interquartile range as $IQR = Q_3 - Q_1 = 38.82$ so that $1.5 \times IQR = 58.23$. Thus, by the 1.5IQR rule, a data point is called an outlier if it is smaller than $Q_1 - 58.23 = -56.09$ or greater than $Q_3 + 58.23 = 99.19$. Since no data lies outside the range $[-56.09,\ 99.19]$, there are no outliers.

#2.44 (c) The mean is 21.95 (in 1999 US dollars) which is much greater than the median of 10.64. The mean is pulled in the direction of the skew which, in this case, is to the right, making it larger. (Note that this is consistent with the observation in the box *Comparing the Mean and Median* on page 41. Again, this is the reason for following the instructions in the *Choosing a Summary* box on page 50.)