Blood proteins in children from Papua New Guinea. C-reactive protein (CRP) is a substance that can be measured in the blood. Values increase substantially within 6 hours of an infection and reach peak within 24 to 48 hours after. In adults, chronically high values have been linked to an increased risk of cardiovascular disease. In a study of apparently healthy children aged 6 to 60 months in Papua New Guinea, CRP was measured in 90 children. The units are in milligrams per liter (mg/l). Here are the data from a random sample of 40 of these children.

a. Find the five number summary.

<table>
<thead>
<tr>
<th>Column</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP</td>
<td>5.085</td>
<td>0</td>
<td>73.2</td>
<td>0</td>
<td>9.47</td>
</tr>
</tbody>
</table>

IQR = 9.47 – 0 = 9.47
Upper fence = 9.47 + 1.5(9.47) = 23.675 (this value is not to be graphed)

b. Make a boxplot.

c. Make a histogram.

Distribution of CRP in 40 children from Papua New Guinea
d. The distribution is certainly right skewed. The explanation above mentions that in adults the value of CRP increases substantially for adults at the onset of an infection within 6 hours. Since all the children were considered healthy is not a surprise to see that most have low CRP values. The boxplot shows that the first 25% of values recorded are very tightly grouped, mainly the Q1 score of 0 indicates that the lower 25% consistent of a CRP value of 0. The median, being just above 5, shows that the second 25% ranged from 0 to a measurement of about 5. The third 25% has a range similar to the second 25%, but the last 25% has a very large spread. Could these large values be indicators of an incoming infection? Is it possible to have such a high CRP value without an infection? If so we can see that it is rare to see such a high value in children considered healthy.

1.61 College tuition and fees. Figure 1.16 (page 25) is a histogram of the tuition and fees charged by the 56 four-year colleges in the state of Massachusetts. Here are the charges (in dollars), arranged in increasing order.

5-Number Summary: \{4123, 15717, 20072, 27958, 29875\}

Lower fence = 15717 – 1.5(27958 – 15717) = -2644, no outliers exist according to the 1.5IQR rule.

The upper 25% of the data, is tightly grouped, and the lower 25% is more spread out. However, you look at the nine or so values in the lower 25% you see and increase in frequency. It would be interesting to see what institutions are represented here and how that compares to the upper 25%; the question being should this group be included with the rest of the group.
1.63 Tornadoes and property damage. Table 1.5 (page 25) shows the average property damage caused by tornadoes over a 50-year period in each of the states. The distribution is strongly skewed to the right.

a) Give the five-number summary. Explain why you can see from these five numbers that the distribution is right-skewed.

5-Number Summary: {0, 2.14, 10.64, 40.96, 88.6}

b) A histogram or stemplot suggests that a few states are outliers. Show that there are no suspected outliers according to the 1.5IQR rule. You see once again that a rule is not a substitute for plotting data.

\[ \text{IQR} = 40.96 - 2.14 = 38.82. \]

If there are any outliers, it must occur in the upper portion where the variation is largest.

\[ 40.96 + 1.5(38.82) = \text{upper fence} \]

99.19

The IQR = 40.96

- 2.14 = 38.82.

If there are any outliers, it must occur in the upper portion where the variation is largest.

40.96 + 1.5(38.82) = upper fence

99.19

= 

c) Find the mean property damage. Explain why the mean and median differ so greatly for this distribution.

Summary statistics:

<table>
<thead>
<tr>
<th>Column</th>
<th>n</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>damage</td>
<td>51</td>
<td>21.952942</td>
<td>10.64</td>
</tr>
</tbody>
</table>

The mean is not a resistant measure, thus the infrequent but extreme values serve to increase the value of the mean, while the median does not take into account the actual value of a number but rather its position.