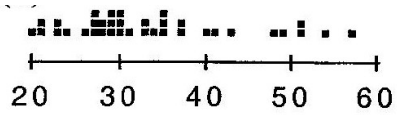
Homework 1

2.7) We use split stems in the following stem-and-leaf diagram because the sample size is large relative to the range of the data. A split stem allows us to see more features of the data set.

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

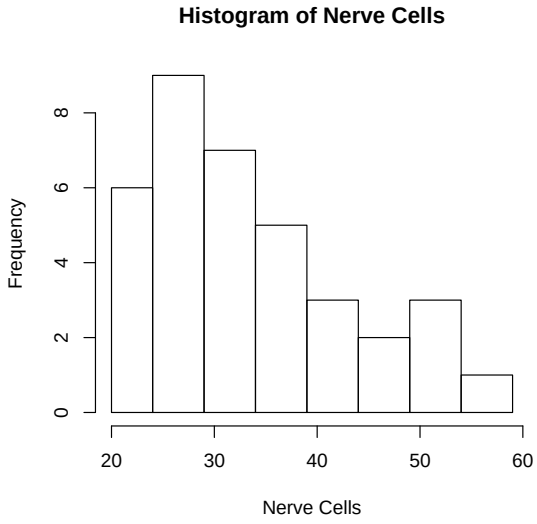
Key 2|0 = 20 branches



2.8) Many different answers are possible, depending on the bin size chosen.

One possibility is:

Branches	Frequency
20-24	6
25-29	9
30-34	7
35-39	5
40-44	3
45-49	2
50-54	3
55-59	1
Total	36



2.36) The histogram has a minimum near 12 and a maximum near 36, so boxplots (a) and (c) are eliminated because one or both whiskers extend to far or fall short of those numbers. The histogram is fairly symmetrical, with a slight skew to the left, and centered around 25. Boxplot (c) shows a large skew to the left and is centered closer to 30, so it is also eliminated. Boxplot (b) meets all of the characteristics of the histogram and is therefore the right choice.

2.46) (a) $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{7}(6.8 + 5.3 + \cdots + 6.2) = 6.343$

$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$
$= \sqrt{\frac{1}{7-1}\left[(6.8 - 6.343)^2 + (5.3 - 6.343)^2 + \cdots + (6.2 - 6.343)^2\right]} = .702$

(b) 5.3 5.9 6.0 6.2 6.8 6.8 7.4 The median is the middle value if n is odd or the average of the two middle values if n is even.
Median = 6.2
$IQR = Q_3 - Q_1$ = 6.8-5.9 = .9

(c) The coefficient of variation is the standard deviation expressed as a percentage of the mean.
coefficient of variation = $\frac{s}{\bar{y}} * 100\% = \frac{.702}{6.343} * 100\% = .111 * 100\% = 11.1\%$

(d) 5.3 5.9 6.0 6.2 6.8 6.8 10.2
new $\bar{y} = 6.77$, new $s = 1.68$, new median = 6.2, new IQR = .9
The median and IQR display resistance, meaning they are not heavily influenced by outliers, in that they do not change. The standard deviation changes greatly, showing its lack of resistance. The mean changes a modest amount.

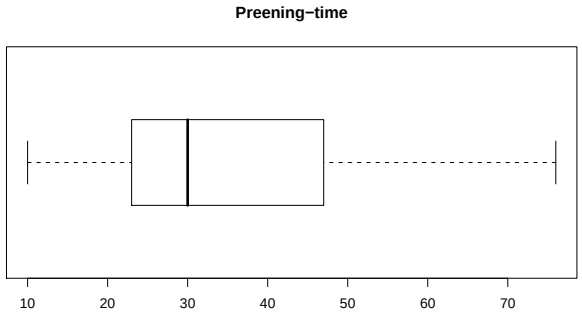
2.63)

10	16	18	19	22	24	24	25	26	29
31	32	33	34	46	48	48	52	57	76

(a) median = average of 10^{th} and 11^{th} numbers = $\frac{29+31}{2} = 30$
Q_1 = average of 5^{th} and 6^{th} numbers = $\frac{22+24}{2} = 23$
Q_3 = average of 15^{th} and 16^{th} numbers = $\frac{46+48}{2} = 47$

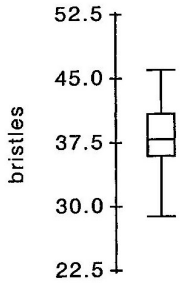
(b) IQR = $Q_3 - Q_1 = 47 - 23 = 24$

(c) Upper fence = $Q_3 + 1.5 * IQR = 47 + 1.5 * 24 = 83$
Lower fence = $Q_1 + 1.5 * IQR = 23 + 1.5 * 24 = 13$



2.73) (a) n = 119, so the median is the 60th largest observation. There are 43 observations less than or equal to 37 and 61 observations less than or equal to 38. Therefore, the median is 38.

(b) The first quartile is the 30^{th} largest observation. There are 21 observations less than or equal to 35 and 32 observations less than or equal to 36. Therefore, $Q_1 = 36$.
The first quartile is the 90^{th} largest observation. Therefore, $Q_3 = 41$.
(c)



(d) $\bar{y} = 34.45$ and $s = 3.20$ Thus, the interval $\bar{y} \pm s = 38.45 \pm 3.20$ which is (35.25, 41,65). The number of flies with 36 to 41 bristles is 11+12+18+13+10+15 = 79. Therefore, the number of observations that fall within one standard deviation of the mean is 79/119*100% = .664*100% = 66.4%.