2. The following are the founding dates of wineries in New Zealand:

1975	1980	1973	1974	1943	1902	1947	1942	1937	1983
1934	1936	1935	1902	1937	1961	1979	1969	1976	1925
1922	1937	1973	1933	1948	1944	1904	1981	1872	1918
1860	1960	1876	1980	1967	1974	1977	1979	1981	1978;

MARKS

8

for these data: $\sum_{j=1}^{n} y_j = 77{,}874$; $\sum_{j=1}^{n} y_j^2 = 151{,}651{,}440$; $n = 40$.

(2, 2, 2, 2)

(a) Construct a stemplot (with *ordered* leaves) of these data, using the grid(s) below.

(b) Find the average, median, standard deviation and IQR for the data.

(c) Explain briefly how the numerical summaries in (b) are useful in describing important features of the data.

(d) On the basis of your answers to (a), (b) and (c), outline (*in 100 words or fewer*) what these data indicate about the times at which wineries were founded in New Zealand.

(a)

186	0											(1)
187	6											(1)
188												(-)
189												(-)
190	2	2	4									(3)
191	8											(1)
192	2	5										(2)
193	4	6	7	5	3	7	7					(7)
194	3	8	4	7	2							(5)
195												(-)
196	0	7	1	9								(4)
197	5	3	3	4	4	9	7	9	6	2	8	(11)
198	0	0	1	1	3							(5)
199												**(40)**

186	0										
187	6										
188											
189											
190	2	2	4								
191	8										
192	2	5									
193	3	4	5	6	7	7	7				
194	2	3	4	7	8						
195											
196	0	1	7	9							
197	2	3	3	4	4	5	6	7	8	9	9
198	0	0	1	1	3						
199											

(b)

1947 (b)	1954 (b)	33 years (b)	41 years (b)
Average	Median	Standard deviation	IQR

Average = 77,874/40 = 1946.85 ≃ 1947; s.d. = $\sqrt{\frac{1}{39}[151{,}641{,}550 - 77{,}874^2/40]}$ = 32.98916 ≃ 33 years;

median = (1948 + 1960)/2 = 1954; IQR = (1975 + 1976)/2 – (1934 + 1935)/2 = 1975½ – 1934½ = 41 years.

(c) The average and the median provide information about the 'centre' of the distribution – here, the 'centre' is around 1950, with the average and median having fairly similar values.

The standard deviation and the IQR provide information about the 'spread' of the distribution – here, the 'spread' is around 37 years, with the s.d. and IQR again having fairly similar values.

For data which consist of calendar years with a remote origin, the 'spread' is of little interest in relation to the 'centre'; rather, we note that the 'spread' here is appeciable in relation to the *range* of 123 years.

(d) The stemplot with ordered leaves (on the right above) shows the following features of these data.

- A small 'early' cluster of two wineries founded 16 years apart, an 'intermediate' cluster of 18 wineries founded over 46 years, and a 'late' cluser of 20 wineries founded over a shorter period of 23 years.
  - These clusters suggest an increasing incentive to found wineries over the 123 year period of the data, likely motvated by commercial factors.
- There were two periods (totalling 36 years) from 1877 to 1901 and from 1949 to 1959 when no wineries were founded.