Supplementary exercises 1.42 and 1.72 of IPS7e

Exercise 1.42

For this exercise, we may either work with the stemplot provided in the problem, or construct our own plot in Minitab. As shown below, it looks exactly the same, apart from the extra Minitab redundancies.

```
MTB > Stem-and-Leaf 'over65'.
```

Stem-and-Leaf Display: over65

Stem-and-leaf of over65 N = 50

1	5	7
1	6	
1	7	
2	8	5
5	9	679
6	10	6
14	11	02233677
(13)	12	0011113445789
23	13	00012233345568
9	14	034579
3	15	36
1	16	
1	17	6

Leaf Unit = 0.1

(a) Because the stems are whole percents and the leaves are tenths of a percent, we can approximate the values to within one decimal by reading off the numbers from the stemplot. A stem value of 17 and a leaf value of 6 corresponds to 17.6% (Florida). In the same way, the approximate value for Alaska equals 5.7%.

(b) The distribution appears as unimodal and left-skewed, centered at 13% and with a fairly narrow spread: almost all the states have values between 11% and 15%.

Exercise 1.72

We prefer to use statistical software to compute numerical summaries for the distribution, despite that the wording of the problem suggests the 5-number summary to be computed from the stemplot.

```
MTB > Describe 'over65'.
```

Descriptive Statistics: over65

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
over65	50	0	12.538	0.269	1.905	5.700	11.675	12.750	13.500	17.600

The 5-number summary is: $5.7 - 11.7 - 12.8 - 13.5 - 17.6$. For the $1.5 \times \text{IQR}$ rule we further calculate:

$$\begin{aligned} \text{IQR} &= 13.5 - 11.7 = 1.8, \\ \text{lower} &= Q1 - 1.5 \times \text{IQR} = 11.7 - 2.7 = 9.0, \\ \text{upper} &= Q3 + 1.5 \times \text{IQR} = 13.5 + 2.7 = 16.2. \end{aligned}$$

Both of the two extreme states (Alaska and Florida) are beyond this range, so they would indeed be identified as suspected outliers. Another state with a value around 8.5% would however also be identified as a suspected outlier. This may seem unnatural because that value does not differ that much from several values between 9% and 10%. In a left-skewed distribution, the rule will tend to identify too many outliers in the left tail, with the converse happening in a right-skewed distribution. This is because the rule is symmetrical and does therefore not account for one tail being naturally larger than the other one.

In practice we would always have to use our judgement to decide whether "suspected outliers" should be regarded as real outliers.