Algorithm for stem and leaf displays:
1) Select leading digits, after that, the digits after the initial leading digits remaining digits are leaves.
2) List the stem values obtained in step 1 in a column.
3) Write the leaf for all data values that correspond to particular stem value.

1·3 34, 40

1·4 44, 56

11)

Stem	Leaf
6L	0 3 4
6H	6 6 7 8 8 9
7L	0 0 1 2 2 2 4 4
7H	
8L	0 0 1 1 1 1 1 2 2 3 4 4
8H	5 5 5 8 9 9
9L	0 3
9H	5 8

There are no high numbers against stems 7
So, the highlight feature in the data is that that there is no leaf against the stem 7H.

A stem-and-leaf display with only 4 stems 6, 7, 8 and 9 wouldn't give a very detailed description of the distribution of scores.

14)

Stem	Leaf
2	2 3
3	2 3 4 4 5 6 7 7 8 9
4	0 1 3 5 6 8 8 9
5	0 0 0 0 1 1 1 4 4 5 5 6 6 6 7 8 9
6	0 0 0 0 1 2 2 2 2 3 3 4 4 4 5 6 6 6 7 7 8 9 9 9 9
7	0 0 0 1 2 2 3 3 4 5 5 5 5 5 6 6 8
8	0 2 2 3 3 4 4 8
9	0 1 2 2 3 3 3 3 5 6 6 6 7 8 8
10	2 3 4 4 4 5 5 6 8 8
11	2 3 3 5 9 9 9
12	3 7
13	8
14	3 6
15	0 0 3 5
16	
17	
18	9

Stem (1.0) | Leaf tenths (0.1)

We could have divided the values into 2 groups (Low and high), but the stem and leaf display would be big and wouldn't be representative since there would be many stem values.

b) A typical, representative, value could be the median, data point that seperated the data on 2 equal parts, in this case 7.0

c) There are some data points that arent concentrated around the representative value, however, the rest of data points appear to be highly concentrated around the 7.0.

d) Data arent symmetric, positively skewed (to the right)

e) Obviously, 18.9 is far from the rest, so might be an outlier.

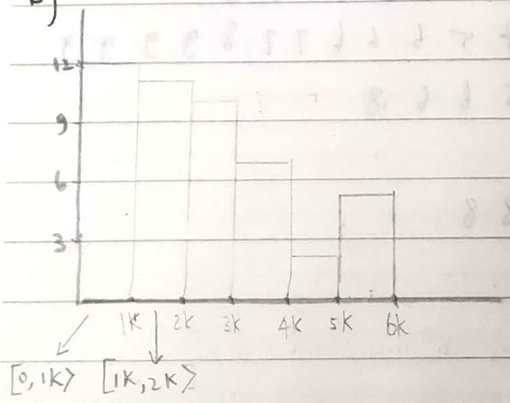
20)

a)

Stem (thousands)	Leaf (hundreds)
0	1 2 3 3 3 4 5 5 5 5 9 9
1	0 0 1 2 2 2 3 4 6 8 8
2	1 1 1 2 3 4 4 4 7 7
3	0 1 1 3 3 3 8
4	3 7
5	2 3 7 7 8

A representative could be the median, a number near 2100 (we did not take into considerations the part after hundreds). Display is bimodal (stem at 5 and at 0 would be considered mode). Positively skewed.

b)



length less than 2000 proportion
$= \frac{12+11}{47} = 0,49$ or 49%

length > 2000 && < 4k proportion
$= \frac{10+7}{47} = 0,362$ or 36,2%

Histogram looks like the stem and Leaf display. It is positively skewed and almost unimodal.

ample mean $\bar{x}$ of observations $x_1, x_2, \cdots, x_n$ is given by

$\bar{x} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

The sample median $\tilde{x}$ is obtained from ordered n observations from smallest to largest where we include the repeated values.

sample median $\tilde{x} = \begin{cases} \text{The only middle value if } n \text{ is odd} \\ \text{The average of the 2 middle value if } n \text{ is even} \end{cases}$

$\begin{cases} \left(\frac{n+1}{2}\right)^{th} & \text{, if } n \text{ is odd} \\ \text{average of } \left(\frac{n}{2}\right)^{th} \text{ and } \left(\frac{n}{2}+1\right)^{th} \text{ ordered values,} & \text{if } n \text{ is even} \end{cases}$

If we obtain a 10% trimmed mean, removing smallest 10% and the largest 10% of the sample and then averaging the rest.

34)

a) Urban, where $n = 11$

$\bar{x}_1 = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{1}{11}(6+5+11+33+4+5+80+18+35+17+23) = \frac{1}{11}\cdot 237$

$= 21,545$

Farm, where $n = 15$

$\bar{x}_2 = \frac{x_1+x_2+\cdots x_n}{n} = \frac{1}{15}(4+14+11+9+9+8+4+20+5+8.9+21+9.2+3+2+0.3)$

$= \frac{1}{15}(128.4) = 8.56$

Sample of urban home has more than 2 times the average of farm homes.

b) ordered data of sample 1: 4 5 5 6 11 17 18 23 33 35 80

" " " " 2: 0.3 2 3 4 4 5 8 8.9 9 9 9.2 11 14 20 21

$n_1 = 11$ (odd), $\left(\frac{n+1}{2}\right) = \left(\frac{11+1}{2}\right)$: 6th ordered value, so $\tilde{x}_1 = 17$

$n_2 = 15$ (odd), $\left(\frac{n+1}{2}\right) = \frac{15+1}{2} = 8^{th}$ " " , so $\tilde{x}_2 = 8.9$

The median for urban sample is almost 2 times that of the farm sample.

The mean of urban samples differ from median, because the larger observations (like 80) raises the mean, but not the median.

c) Urban, delete smallest (4, 5) and biggest (80), gives trimmed mean:

$\bar{x}_{1,trim} = \frac{1}{9}(5+5+6+11+17+\cdots+35) = 17$

Correspon ding trimming percentage is $\frac{1}{11} \times 100\% = 9.09\%$.

the trimmed mean $\bar{x}_{1,\text{trim}}$ (17) is less than entire sample mean $\bar{x}_1$ (21,545)

" " " $\bar{x}_{1,\text{trim}}$ (17) is the same as " " median $\tilde{x}_1$ (17)

Farm, delete smallest (0.3) and biggest (21), gives trimmed mean:

$\bar{x}_{2,\text{trim}} = \frac{1}{13}(2+3+\cdots+20) = \frac{1}{13} \times 107.1 = 8,24$

corresponding trimming percentage is $\frac{1}{15} \times 100\% = 6,67\%$

The trimmed mean $\bar{x}_{2,\text{trim}}$ (8,24) is less than entire sample mean $\bar{x}_2$ (8.56)

" " " $\bar{x}_{2,\text{trim}}$ (8,24) " " " median $\tilde{x}_2$ (8.9)

40) Data of ex 27:

11	14	20	23	31	36	39	44	47	50
59	61	65	67	68	71	74	76	78	79
81	84	85	89	91	93	96	99	101	104
105	109	112	118	123	136	139	141	148	158
161	168	184	206	248	263	289	322	388	513

Sample mean $\bar{x} = \frac{1}{50}(11+14+\cdots+513) = \frac{1}{50} \cdot 5963 = 119,26$

Sample median $\tilde{x}$ for $n = 50$:

$(\frac{n}{2})^{th} = 25^{th}$ and $(\frac{n}{2}+1)^{th} = 26^{th}$

$\tilde{x} = \frac{25^{th} + 26^{th}}{2} = \frac{91+93}{2} = 92$

25% trimmed mean:

$25\% \times 50 = 12,5$, so eliminate 12 values from both sides

$\bar{x}_{tr(24)} = \frac{1}{26}(65+67+\cdots+141) = \frac{1}{26} \cdot 2480 = 95,38$

10% trimmed mean:

$10\% \times 50 = 5$, so eliminate 5 values from both sides

$\bar{x}_{tr(10)} = \frac{1}{40}(36+39+\cdots+248) = \frac{1}{40} \cdot 4089 = 102,23$

a) Sample range = largest sample value - smallest sample value = 49.3 - 23.5 = 25.8

b) Sum of sample data $\sum_{i=1}^{n} x_i$ = 29,5 + 49.3 + 30.6 + 28.2 + 28 + 26.3 + 33.9 + 29.4 + 23.5 + 31.6 = 310.3

Sample mean $\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{310.3}{10} = 31,03$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
29,5	-1.53	2,34
49.3	18.27	333,79
30.6	-0.43	0,18
28.2	-2.83	8,01
28.0	-3.03	9.18
26.3	-4.73	22,37
33.9	2.87	8.24
29.4	-1.63	2.66
23.5	-7.53	56,7
31.6	0.57	0.32
		$\sum_{i=1}^{n} (x_i - \bar{x})^2 = 443.80$

Sample variance $S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{443.8}{9} = 49.31$

c) Sample standard deviation $S = \sqrt{S^2} = \sqrt{49,31} = 7,02$

d) Using shortcut method to calculate sample variance S^2:

$\sum_{i=1}^{n} x_i^2 = 29.5^2 + 49.3^2 + \cdots + 31.6^2 = 10072,41$

$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} = 10072,41 - \left(\frac{310.3}{10}\right)^2 = 443.8$

Sample variance $S^2 = \frac{S_{xx}}{n-1} = \frac{443.8}{10-1} = 49.31$

5b)

Sample mean $\bar{x} = \frac{1}{35}(16,35 + 19,08 + \cdots + 22,25) = 19,257$

$n = 35$ (odd), so median $\hat{x} = \left(\frac{n+1}{2}\right)^{th} = \left(\frac{35+1}{2}\right)^{th} = 18^{th}$ ordered value = 19,2

Using shortcut method for sample variance S^2:

$\sum_{i=1}^{n} x_i^2 = 15,3^2 + 16,2^2 + \cdots + 23,78^2 = 13093,77$

$\sum x_i = 674,01$

$(\sum x_i)^2 = (674,01)^2 = 454289,48$

$S_{xx} = \sum x_i^2 - \frac{1}{n} \cdot (\sum x_i)^2 = 13093,77 - \frac{1}{35} \cdot 454289,48 = 114.07$

Sample variance $s^2 = \frac{1}{n-1} \cdot S_{xx} = \frac{1}{34} \times 114,07 = 3,355$

Sample standard deviation $s = \sqrt{s^2} = \sqrt{3,355} = 1,832$

low fourth = median of the smallest half

upper fourth = " " " largest "

Fourth spread f_s = upper fourth - lower fourth.

Split the list in 2 equal halves, because n is odd, the middle value appears in both half.

Each half consist of 18 elements

First half: 15.3 16.2 16.35 17.15 17.48 17.73 17.75 17.85 18 18.68 18.82 18.85
19.03 19.07 19.08 19.17 19.2 19.2

Second half: 19.2 19.33 19.37 19.45 19.48 19.5 19.58 19.6 19.62
19.9 19.97 20 20.05 21.22 22.25 22.75 23.25 23.78

upper fourth $\tilde{x}_1 = \frac{19.62 + 19.9}{2} = 19.76$

lower " $\tilde{x}_2 = \frac{18 + 18.68}{2} = 18.34$

fourth spread $f_s = \tilde{x}_1 - \tilde{x}_2 = 19.76 - 18.34 = 1.42$

The boxplot is based on 5 numbers:

smallest x_i, lower fourth, median, upper fourth and largest x_i

Observations that are $> 1.5\ f_s$ from closest (upper / lower) fourth is an outlier

" " " > 3 " " nearest fourth " " extreme outlier

" " " ≤ 3 " " " " " " mild "

$1.5\ f_s = 1.5 \times 1.42 = 2.13$

$3\ f_s = 3 \times 1.42 = 4.26$

Limits for mild and extreme outliers:

$\tilde{x}_1 + 1.5\ f_s = 19.76 + 2.13 = 21.89$ for mild outliers

$\tilde{x}_2 - 1.5\ f_s = 18.34 - 2.13 = 16.21$ " " "

$\tilde{x}_1 + 3\ f_s = 19.76 + 4.26 = 24.02$ " extreme "

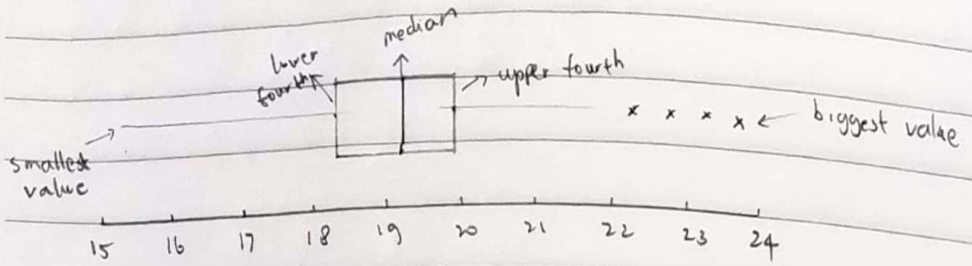
$\tilde{x}_2 - 3\ f_s = 18.34 - 4.26 = 14.08$ for " "

Mild outliers are all values between 14.08 and 16.21 as well as between 21.89 and 24.02.

The extreme outliers are values < 14.08 and > 24.02.

In our case, there is no extreme values.

Mild outliers are : 22, 75 , 22.25, 23.25, 23.78 , 15.3 and 16.2



A