Problem A (see associated R script)

$\min(x) = 11$

$\max(x) = 53$

$\text{sum}(x) / \text{length}(x) = 34.07 = \bar{x} = \frac{\sum x_i}{n}$

$$s^2 = \frac{\sum (x-\bar{x})^2 f_i}{\bar{n}-1}$$

where f_i is the associated freq of the data x_i.

where $\bar{n}$ is the # of grouped data $= 15$.

In this case it's easier to use R.

once all data is loaded let:

$$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$$

n = length(x) = 100

xbar = mean(x);

xm = (x - xbar)^2.

var = sum(xm)/(n-1) $\simeq$ 62.67 $= s^2$

s = sqrt(var) $\approx$ 7.92

If the population formulas are used, the mean remains the same but for the variance:

$$\sigma^2 = \frac{\sum (x-\bar{x})^2}{n}$$

and so σ^2 is a bit smaller than s^2 and σ a bit smaller than s. Using R, we get that:

$\sigma^2 \simeq 62.04$ (see script)

$\sigma \approx 7.88$

To plot a histogram we must decide on class limits and width.

We have $\min(x) = 11$ and $\max(x) = 53$

Let's use about 7 bars (personal choice, usually between 5 and 10 is a good number).

$$\Rightarrow \text{class width} = \frac{53-11}{7-1} = \frac{42}{6} = 7$$

Let's define the following classes

class 1: 10 – 16.9   boundaries: 9.95, 16.95

class 2: 17 – 23.9   boundary: 23.95

class 3: 24 – 30.9   boundary: 30.95

class 4: 31 – 37.9   boundary: 37.95

class 5: 38 – 44.9   boundary: 44.95

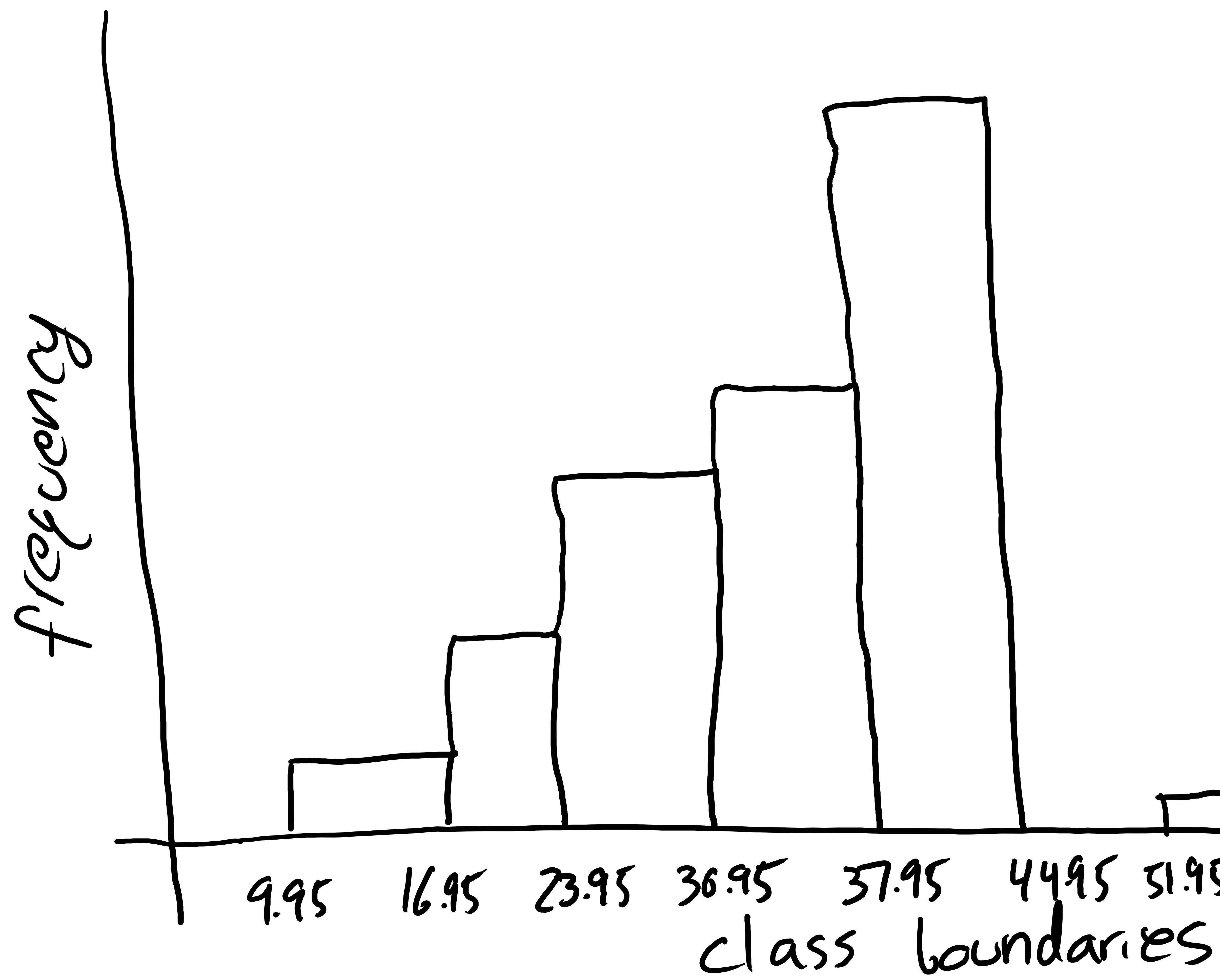
class 6: 45 – 51.9   boundaries: 51.95, 58.95

class 7: 52 – 58.9

Note that 10 was an arbitrary starting point below (but close to) the mean. In fact, the last class goes past the max a bit too much. But for a rough histogram the above is fine.

We get the frequency counts using R (see the script):

class #	frequency
1	3
2	9
3	18
4	24
5	45
6	0
7	1



histogram is skewed to the left.
Not approx bell shaped.

How many measurements fall within 1 and 2 std deviations of the mean?

$(\bar{x}-s, \bar{x}+s) = (26.15, 41.99) = I_1$

$(\bar{x}-2s, \bar{x}+2s) = (18.24, 49.90) = I_2$

we use R to get % of measurements in intervals I_1 and I_2 above:

71% in I_1, 94% in I_2

This is close to what the empirical rule tells us, even though the data is not bell shaped.

By empirical rule, 68% should be in

I_1 and 95% in I_2.

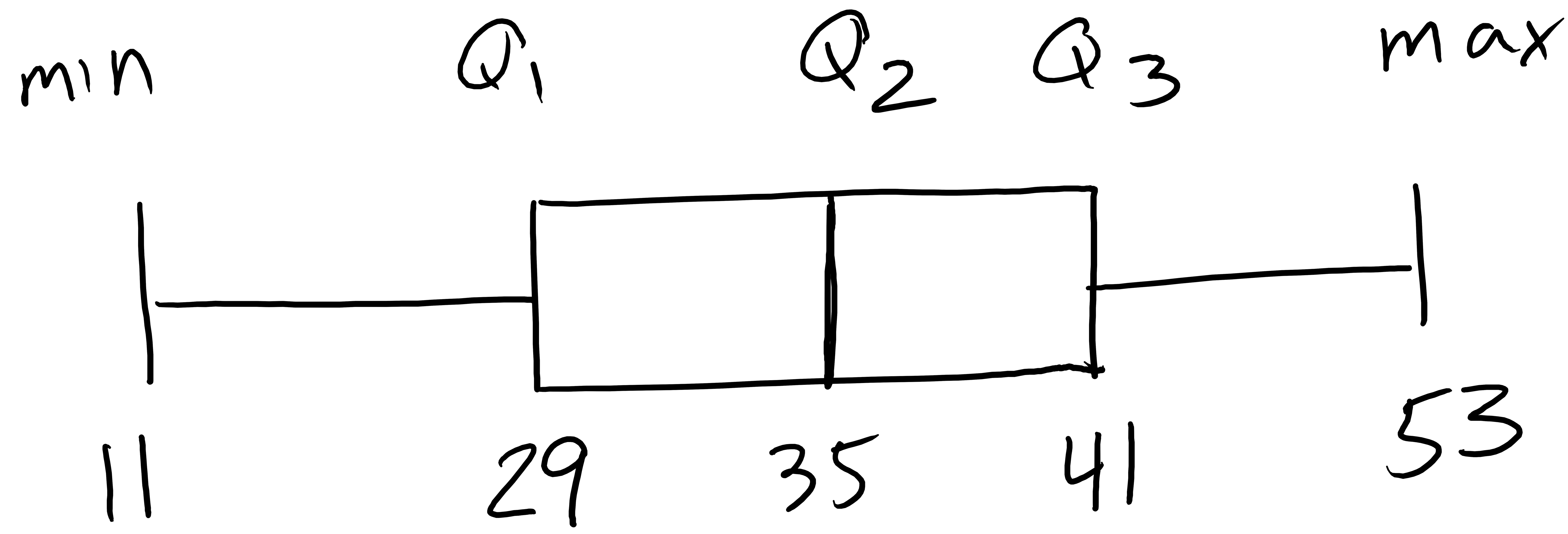
Chebyshev's rule applies only to I_2. At least $\left(1-\frac{1}{2^2}\right)\times 100\%$ of data lies in I_2.

$\Rightarrow$ at least 75% of data lies in I_2 which is consistent with our result.

For the boxplot, no outliers are present (no numbers $< Q_1 - 1.5IQR$ or $> Q_3 + 1.5IQR$). See R script.

Five # summary:

$(\min, Q_1, Q_2, Q_3, \max) = (11, 29, 35, 41, 53)$



boxplot (no outliers)