1) See the HW3 R script.

Note that the correlation coefficient indicates degree of linear correlation between variables x and y (it does not indicate causation).

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

(Let x=w). We have (from R):

$\bar{x} = 9.07$

$s^2 = \frac{1}{20-1}\left[(1-9.07)^2 + (5-9.07)^2 + \ldots + (1-9.07)^2\right]$

≈ 122.48

$\Rightarrow s \approx 11.07$

five number summary:

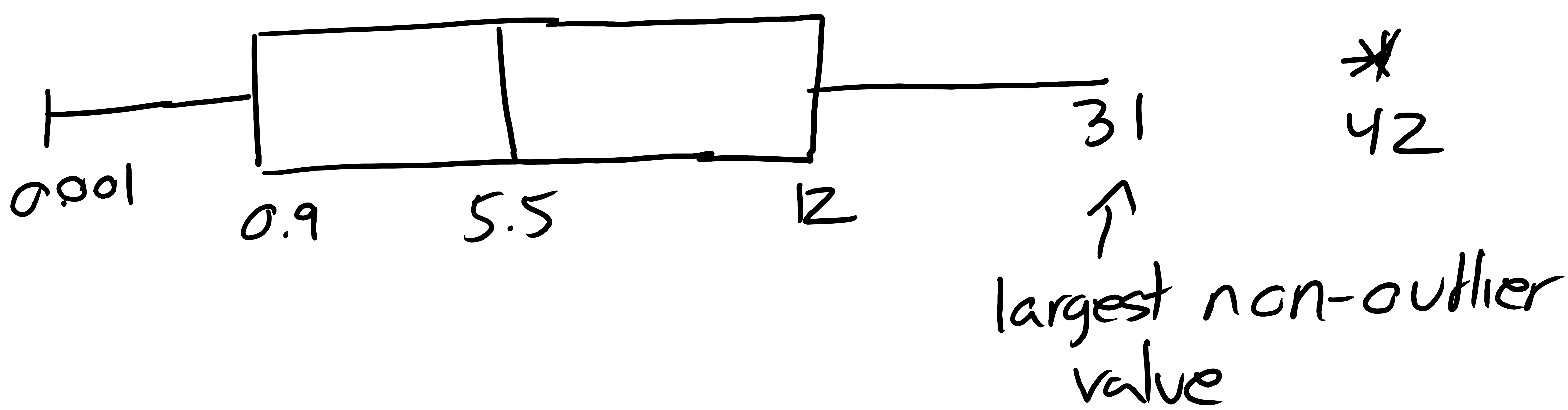
$\{\min, Q_1, Q_2, Q_3, \max\} = \{0.001, 0.9, 5.5, 12, 42\}$

$IQR = Q_3 - Q_1 = 11.1$

$(Q_1 - 1.5\,IQR,\ Q_3 + 1.5\,IQR)$
$= (-15.75,\ 28.65)$

Note that the 0.001 eb fish is not an outlier but the 42 eb fish is.

Modified boxplot:



$$z_{0.001} = \frac{0.001 - \bar{x}}{s} = -0.82$$

(only 0.82 std dev's below the mean)

$$z_{0.5} = \frac{0.5 - \bar{x}}{s} = \frac{0.5 - 9.07}{11.07} \approx -0.77$$

$$z_{31} = \frac{31-\bar{x}}{s} \approx 1.98$$

$$z_{42} = \frac{42-\bar{x}}{s} \approx 2.97$$

Notice that $x_i = 42$ value is an outlier both by the IQR and z-score characterization. $x_i = 0.001$ is not an outlier under either.

$$\bar{x} \pm 2s = (-13.06,\ 31.20)$$

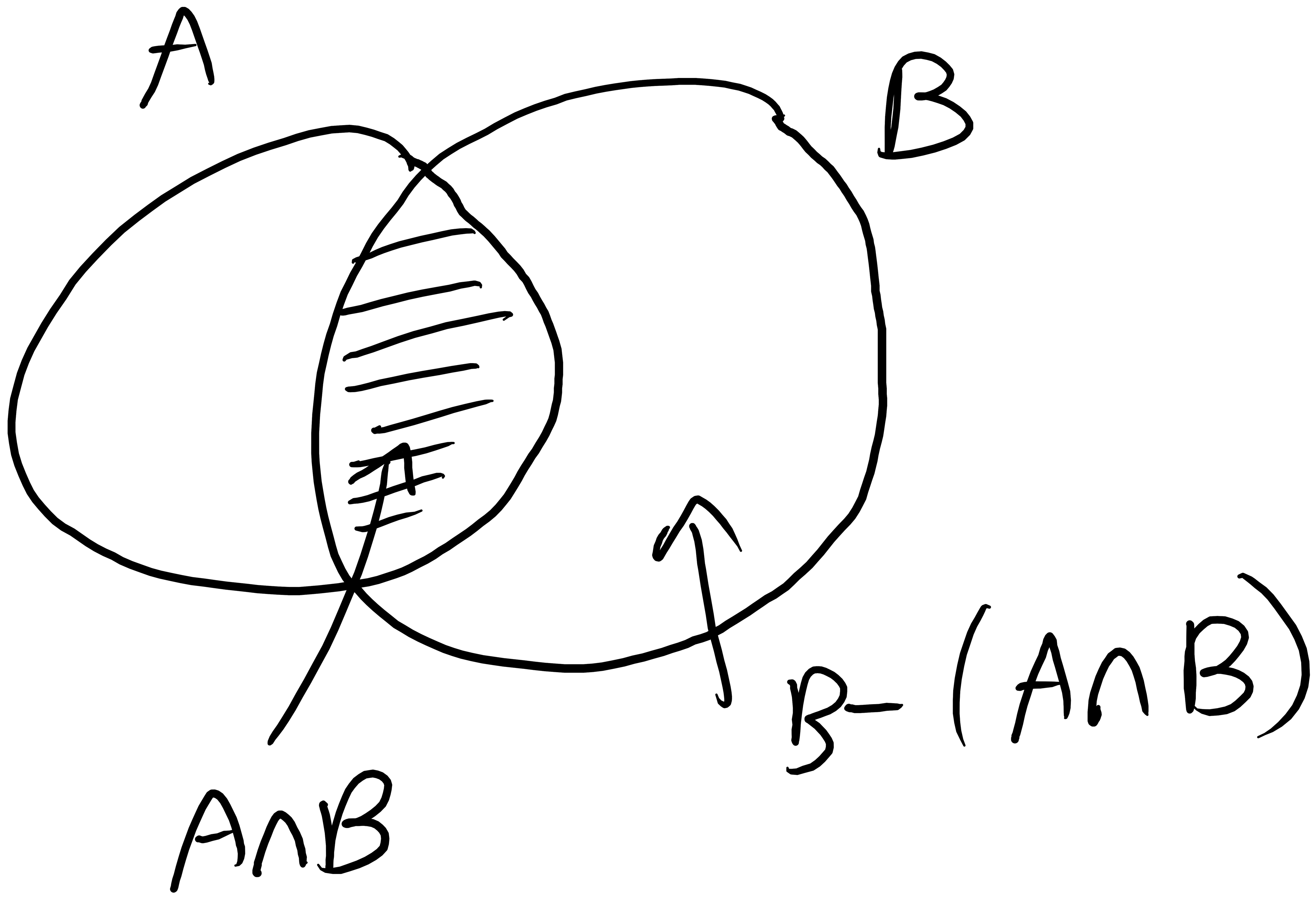
All but one value ($x_i = 42$) lie in this interval. Thus $19/20 \times 100\% = 95\%$ of values lie in this interval. This is exactly as predicted by Empirical rule for normal (bell curve fitted) data.

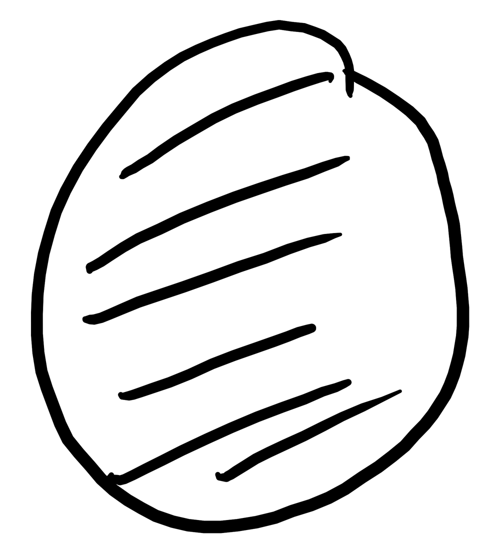
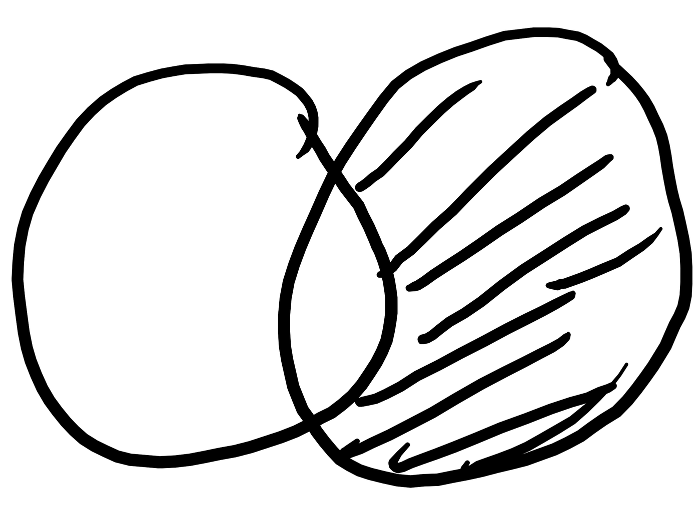
Note that plotting a quick histogram in R with hist(x) shows that the data is <u>not</u> approx normal but is skewed to the right.

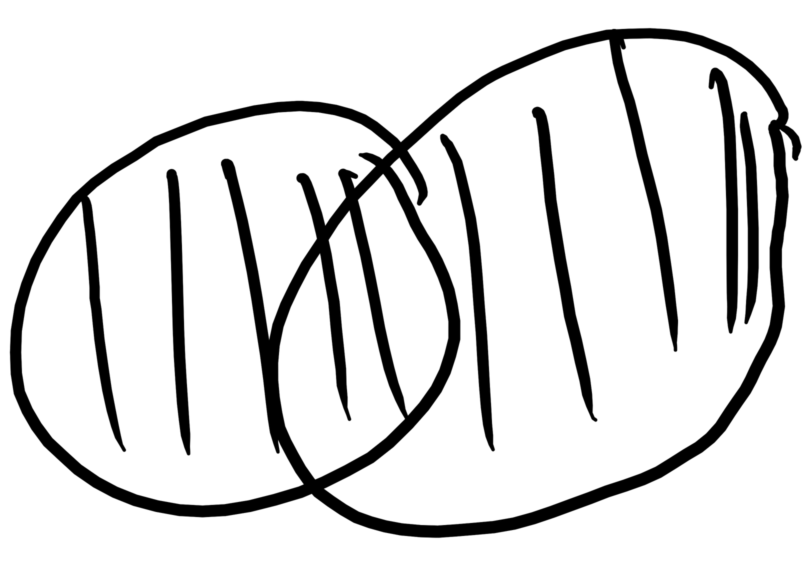
3) We show that

$$A \cup B = A \cup (B - (A \cap B))$$

using a Venn diagram below.



$A \cup (B - (A \cap B))$ is  + 

which is  $= A \cup B$

Note that since A and $B-(A\cap B)$ are mutually exclusive (since they have no elements in their intersection),

$$P(A\cup B) = P(A) + P[B-(A\cap B)]$$

$$= P(A) + P(B) - P(A\cap B).$$

4) $P(\text{ace}) =$

C: get a club
H: get a heart
S: get a spade
D: get a diamond

$= P(\text{ace} \cap H \text{ or } \text{ace} \cap S \text{ or } \text{ace} \cap D \text{ or } \text{ace} \cap C)$

$$= \frac{1}{52} + \frac{1}{52} + \frac{1}{52} + \frac{1}{52} = \frac{1}{13}$$

Notice that since only one card is drawn, $\text{ace} \cap H$, $\text{ace} \cap S$, etc are mutually exclusive.

$P(\text{Jack of hearts}) = \frac{1}{52}$ (single card)

$= P(J \cap H)$

$P(\text{3 of clubs or 6 of diamonds})$

$= P(3 \cap C) + P(6 \cap D) = \frac{1}{52} + \frac{1}{52} = \frac{1}{26}$

mutually exclusive events

$P(\text{heart}) = P(1 \cap H \text{ or } 2 \cap H \text{ or } 3 \cap H \text{ or} \ldots \text{or } 13 \cap H)$

$$= \sum_{i=1}^{13} \frac{1}{52} = \frac{13}{52} = \frac{1}{4}$$

$P(\text{any suit except spades})$

$= 1 - P(\text{spades}) = 1 - \frac{1}{4} = \frac{3}{4}$

P(neither a four nor a club)

$= P(4' \cap C') = P((4 \cup C)')$ (use de Morgan's Law)

$P(4 \cup C) = P(4) + P(C) - P(4 \cap C)$

$= \frac{1}{13} + \frac{1}{4} - \frac{1}{52}$

$P(4' \cap C') = 1 - P(4 \cup C)$

$= 1 - \left[\frac{1}{13} + \frac{1}{4} - \frac{1}{52}\right] = \frac{9}{13}$