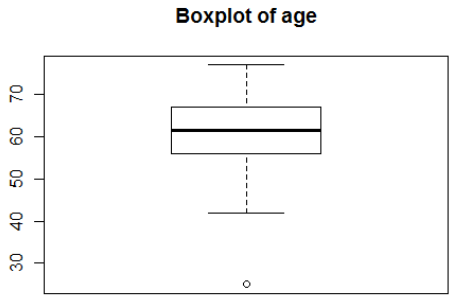
Solution to Homework 1

Question 1

(a). Boxplot by hand. Get the median 61.5, 1st and 3rd quartiles to be 56.25 and 66.75.



(b).

Group	Sample Mean	Sample Standard Deviation
Osteo= 0	0.94	0.11
Osteo= 1	0.83	0.13

Table 1: Summary statistics for the DPA measure of the bone density within osteoporosis group= 0, 1

Stem plots for osteoporosis group= 0:

```
The decimal point is 1 digit(s) to the left of the |

 7 | 2
 8 | 014478
 9 | 0125677799
10 | 3447
11 |
12 | 1
```

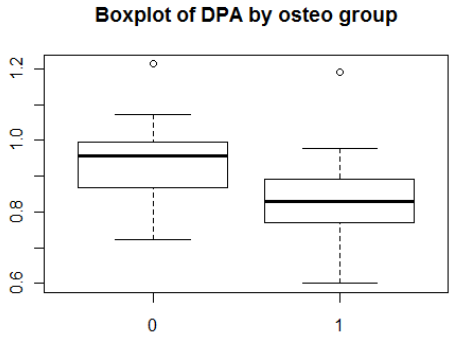
Stem plots for osteoporosis group= 1:

```
The decimal point is 1 digit(s) to the left of the |

 6 | 05906899
```

```
 8 | 02344681228
10 | 9
```

(c). This question asks you to read from the boxplot.



The middle line goes through the box is the median. The two edges of the box are the 1st and 3rd quartiles. We have

Group	Sample Median	IQR
Osteo= 0	0.96	0.12
Osteo= 1	0.83	0.11

Table 2: Meidans and IOR for the DPA measure of the bone density within osteoporosis group= 0, 1

(d). The DPA has a high median and mean in the osteo 0 group than in the osteo 1 group. They have similar standard deviations. The distribution for osteo 0 is a little bit more skewed than the osteo 1 group.

Question 2

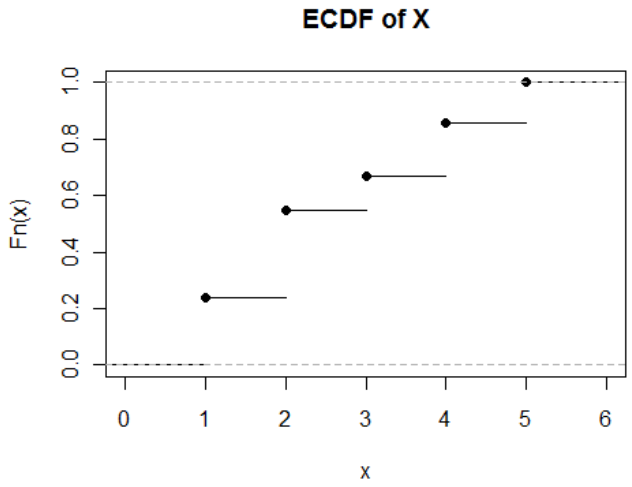
(a).

X	1	2	3	4	5
P(X)	0.238	0.310	0.119	0.190	0.143

(b). mean of X is 2.69

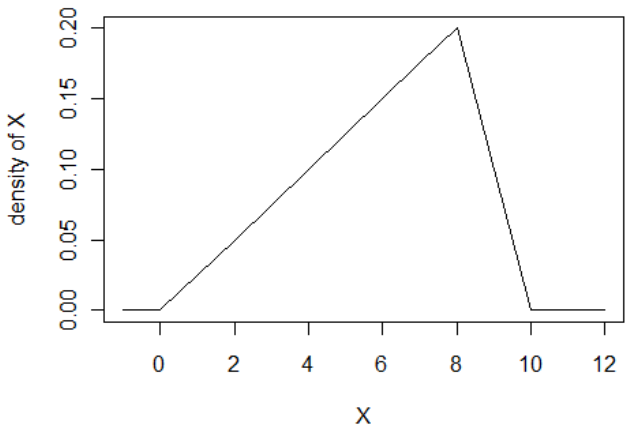
(c). Cumulative distribution function of X

X	1	2	3	4	5
F(X)	0.238	0.548	0.667	0.857	1.000



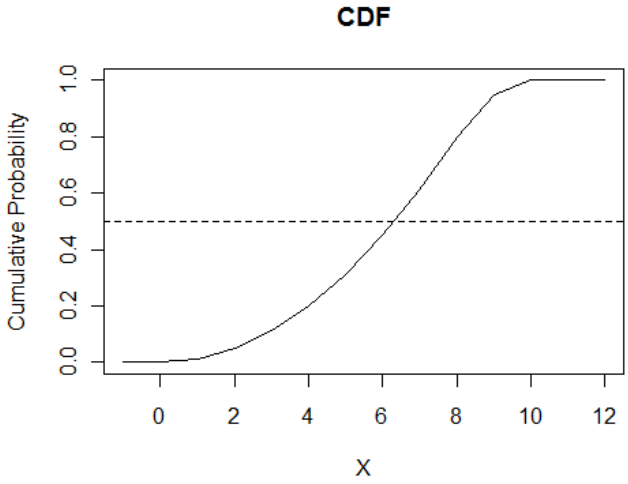
(d). This is a quite open question which is about the debate of categorical measurement versus continuous measurement. The categorical measurement is easier to process and matches human cognition process: dichotomy. However, it loses information. Which one to choose depend on the characteristics of the variable one are measuring, the signal-to-noise ratio and the purpose of the analysis, etc.

Question 3



(a).

$$
\begin{aligned}
F_X(x) &= \int_{-\infty}^{x} f_X(t)dt \\
&= 0, && \text{if } x < 8 \\
&= .0125x^2, && \text{if } 0 \le x \le 8 \\
&= -0.05x^2 + x - 4, && \text{if } 8 \le x \le 10 \\
&= 1, && \text{if } x \ge 10
\end{aligned}
$$



(b). $P(X > 7) = 1 - P(X \leq 7) = 1 - F_X(7) = .3875$

(c). Mean $\mu_X = EX = \int_0^{10} t f_X(t) dt = 6$; the median m_X satisfies that $F_X(m_X) = 0.5$, so $m_X = 6.32$.

(d). Draw a horizontal line at level 0.5. Find the intersection point that the horizontal line cross with the curve of F_X. Draw a vertical line from the intersection point. The projected value on the x-axis is the median $Q_X(.5)$. Shown in Figure (a).

(e). Not satisfied. From (e), the probability of exceeding the safety level is about .38. Even though the mean is below 7, the distribution of X is quite spread. There is still a high risk of being dangerous.