education equals

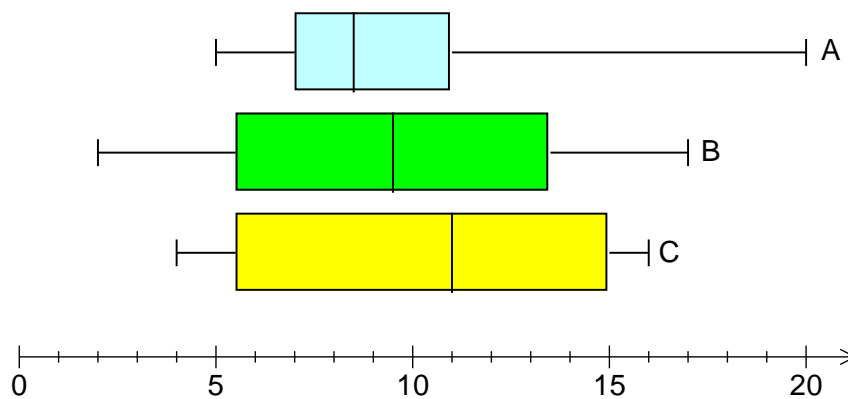**Topic: Comparing Data Sets**

Time: 45 mins   Marks: /45 marks

Calculator Assumed

Question One: [3, 3, 3: 9 marks]
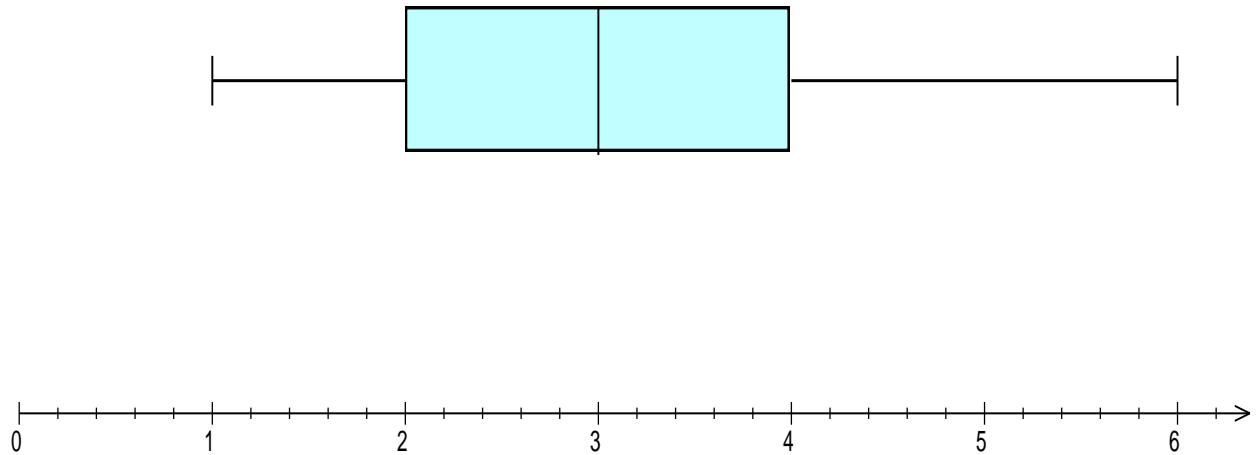
Consider the following boxplots A, B and C.

a)   Describe the data displayed in boxplot A.

b)   Do any of the sets of data have an outlier? Justify your answer mathematically.

c)   Which set of data is the most consistent? Use statistics to justify your answer.

Question Two: [1, 2, 2, 5, 2: 12 marks]

Data is collected and analysed following a medical study. The study involved two separate groups. The people in the one group received a trial medicine to relieve cold and flu symptoms and the other group received a placebo.

The box plot below is for the group who received the medicine and shows the number of days each person showed cold and flu symptoms for.

a)   State the median number of days the people in the group showed symptoms.

b)   Calculate the inter-quartile range and the range of the number days people in this group showed symptoms.

Another person who took part in the study, who was not included in the box and whisker plot above, showed symptoms for 6 days.

c)   What effect will this new data have on the above box plot?
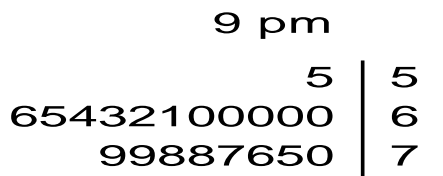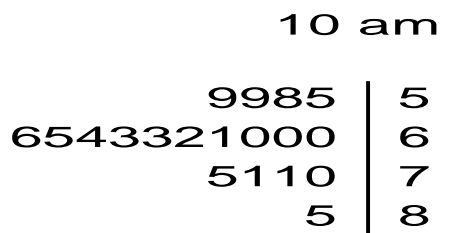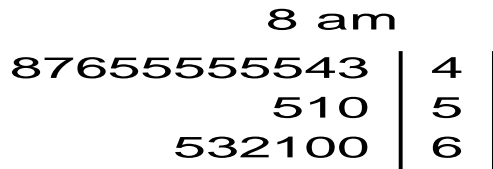
For the other group who only received the placebo medicine, the number of days they each showed cold and flu symptoms are represented in the table below.

Number of Days with Cold and Flu Symptoms	Frequency
1	0
2	1
3	2
4	5
5	6
6	6
7	1

d)   For this second group, calculate correct to 2 decimal places:

   i)   the mean

   ii)   the standard deviation

   iii)   the median

   iv)   the inter-quartile range

   v)   the range

e)   Upon checking the records it is found that one person showed symptoms for 13 days but their information was accidentally recorded as 3 days instead! Once this has been corrected, describe the effect of the change on the statistics calculated in part d.

Question Three: [2, 2, 6, 2: 12 marks]

Consider the following stem-and-leaf plots displaying the speeds in km per hour of the first 20 cars to drive on a particular stretch of road at three different times of the day. The speed limit on this stretch of road is 60km/hr.

8 am

87655555543	4	
510	5	
532100	6	

10 am

9985	5	
6543321000	6	
5110	7	
5	8	

9 pm

5	5	
65432100000	6	
99887650	7	

a) Which time of day has the smallest range in speeds?

b) Which time of day has the smallest median?

c) What percentage, of the first 20 cars to drive on the stretch of road at each time of day that the data was collected, are driving above the speed limit?

d) During which time of day would you conclude that drivers are more likely to drive above the speed limit? Justify your answer with statistics.

Question Four: [3, 3: 6 marks]

Three sets of data were used to create the following boxplots and dot plots.

a)   Match each boxplot with its corresponding dot plot.
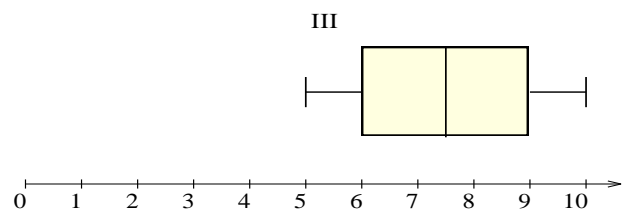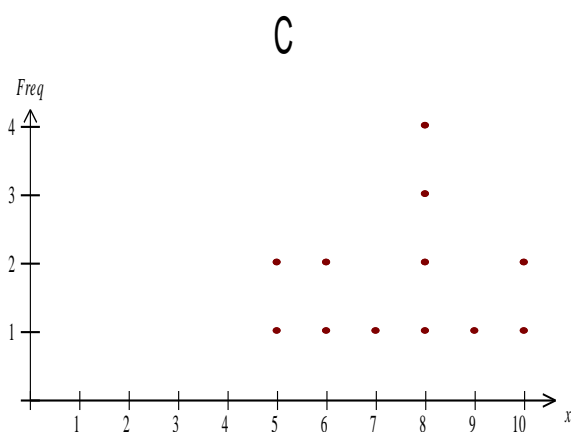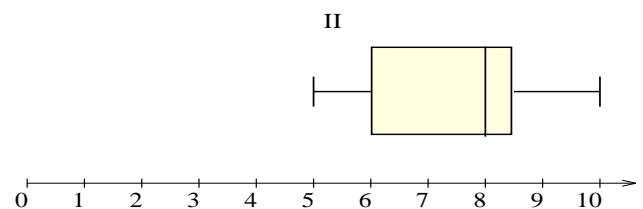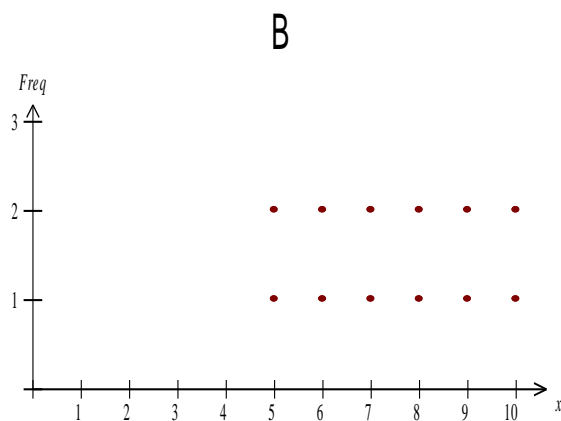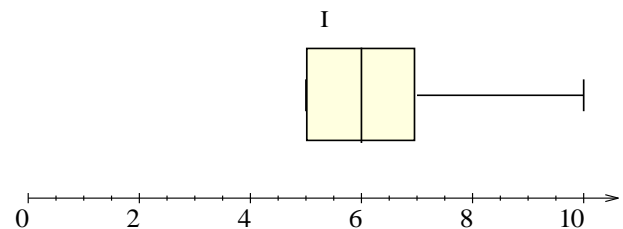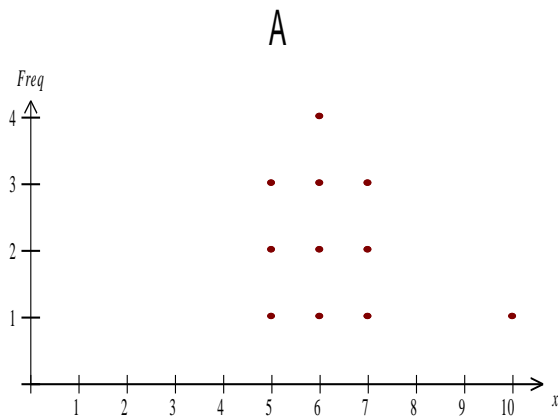
A matches with _____

B matches with _____

C matches with _____

b)   Using the boxplot labeled I, calculate the smallest possible positive integer which would be considered an outlier.

Question Five: [2, 2, 2: 6 marks]

A real estate agent is discussing the value of property in a particular suburb with a client.

The real estate agent is quoted saying, "Recent sales in your area indicate that the mean house price is close to $760 000 so we could list your house to sell for that amount."

The histogram below shows the data of the recent house sales in this client's area.
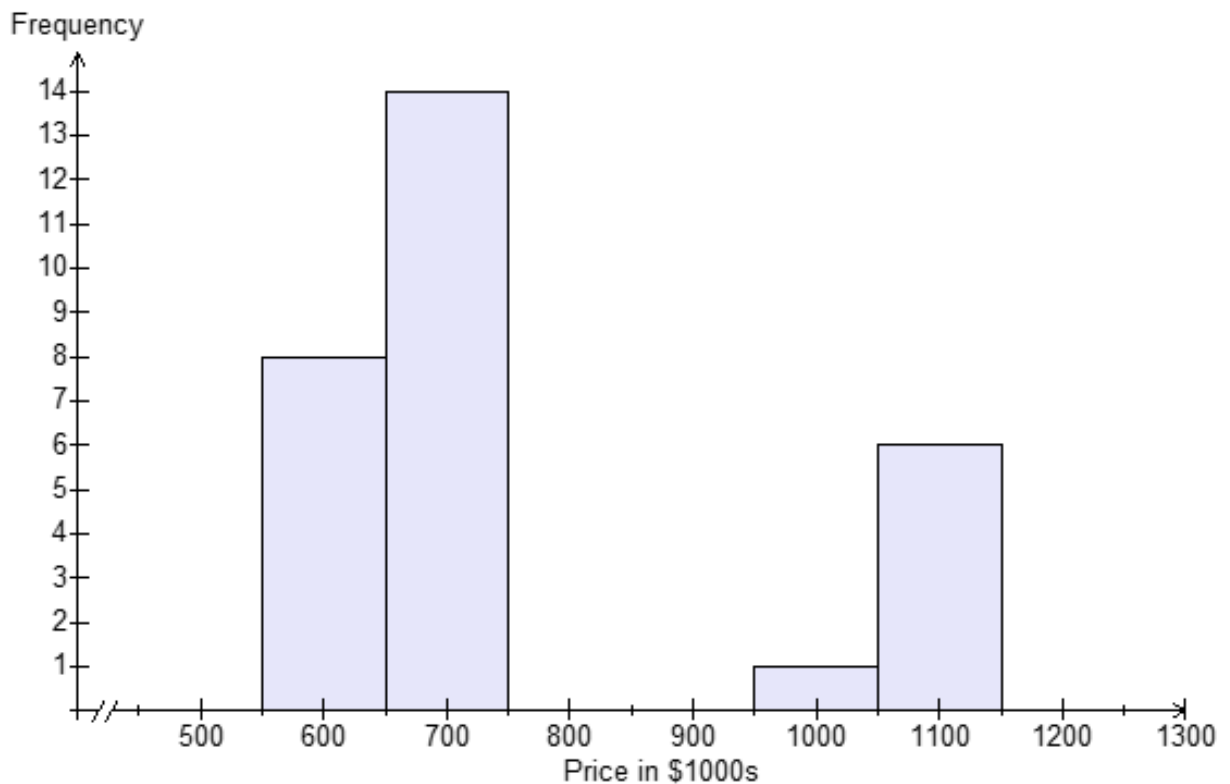
a)   Calculate the estimated mean house price.

b)   Calculate the class median house price.

c)   Using statistical measures as reasoning, discuss whether or not the real estate agent's comments are fair.
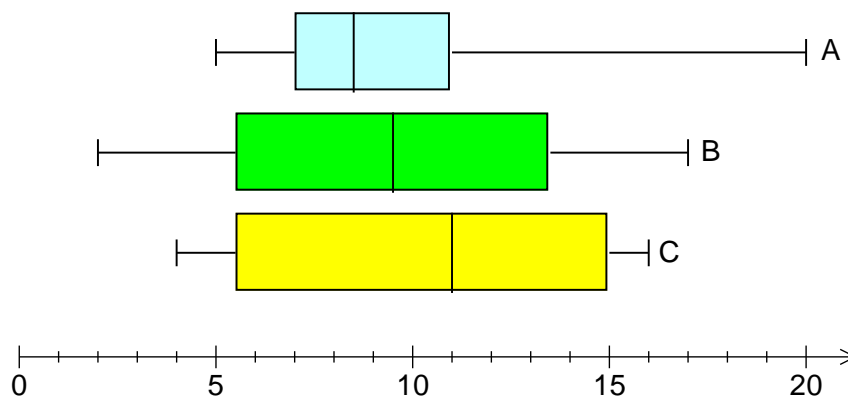
www.educationequals.com

education equals

empower engage enlighten enrich explore

Question One: [3, 3, 3: 9 marks]

Consider the following boxplots A, B and C.

a)   Describe the data displayed in boxplot A.

There is a large range. ✓

From the min – Q3 is a range of 6 and the max-Q3 is a range of 9, indicating the largest spread in Q3. ✓ ✓

b)   Do any of the sets of data have an outlier? Justify your answer mathematically.

$A: IQR = 4 \quad 1.5 \times 4 = 6$ ✓

$Outlier \; 17 +$

$The \; max = 20 \; and \; therefore \; is \; an \; outlier$ ✓

$B: no \; outliers$ ✓

$C: no \; outliers$

c)   Which set of data is the most consistent? Use statistics to justify your answer.

C → low IQR and low range ✓ ✓ ✓

Question Two: [1, 2, 2, 5, 2: 12 marks]

Data is collected and analysed following a medical study. The study involved two separate groups. The people in the one group received a trial medicine to relieve cold and flu symptoms and the other group received a placebo.

The box plot below is for the group who received the medicine and shows the number of days each person showed cold and flu symptoms for.
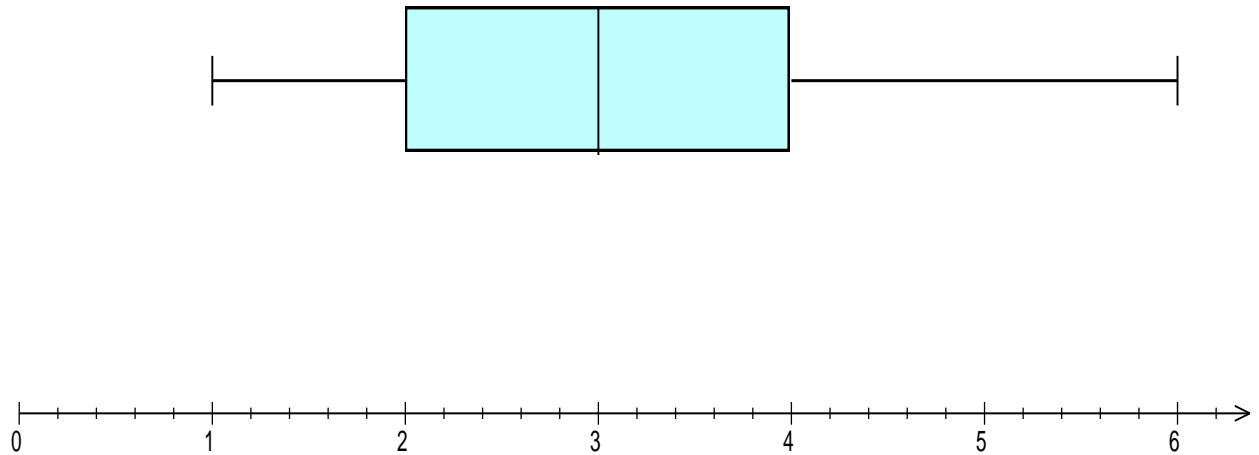
a)   State the median number of days the people in the group showed symptoms.

   3 days   ✓

b)   Calculate the inter-quartile range and the range of the number days people in this group showed symptoms.

   2 days   ✓ ✓

Another person who took part in the study, who was not included in the box and whisker plot above, showed symptoms for 6 days.

c)   What effect will this new data have on the above box plot?

   Will not change the range.   ✓

   May change the median.   ✓

For the other group who only received the placebo medicine, the number of days they each showed cold and flu symptoms are represented in the table below.

Number of Days with Cold and Flu Symptoms	Frequency
1	0
2	1
3	2
4	5
5	6
6	6
7	1

d)   For this second group, calculate correct to 2 decimal places:

  i)   the mean

       4.81 days   ✓

  ii)   the standard deviation

       1.22 days   ✓

  iii) the median

       5 days   ✓

  iv)   the inter-quartile range

       2 days   ✓

  v)   the range

       5 days   ✓

e)   Upon checking the records it is found that one person showed symptoms for 13 days but their information was accidentally recorded as 3 days instead! Once this has been corrected, describe the effect of the change on the statistics calculated in part d.

   Mean increases, standard deviation increases, median would remain the same range is increased, IQR is increased   ✓ ✓

Question Three: [2, 2, 6, 2: 12 marks]

Consider the following stem-and-leaf plots displaying the speeds in km per hour of the first 20 cars to drive on a particular stretch of road at three different times of the day. The speed limit on this stretch of road is 60km/hr.

8 am

87655555543	4
510	5
532100	6

10 am

9985	5
6543321000	6
5110	7
5	8

9 pm

5	5
65432100000	6
99887650	7

a) Which time of day has the smallest range in speeds?

8am ✓ ✓

b) Which time of day has the smallest median?

8am ✓ ✓

c) What percentage, of the first 20 cars to drive on the stretch of road at each time of day that the data was collected, are driving above the speed limit?

$8am \ \frac{4}{20} = 20\,\%$ ✓ ✓ $9am \ \frac{12}{20} = 60\,\%$ ✓ ✓ $9pm \ \frac{14}{20} = 70\,\%$ ✓ ✓

d) During which time of day would you conclude that drivers are more likely to drive above the speed limit? Justify your answer with statistics.

✓ 9pm: the greatest mean and median and 70% of drivers driving over the speed limit of 60km/hr. ✓

Question Four: [3, 3: 6 marks]

Three sets of data were used to create the following boxplots and dot plots.

a)   Match each boxplot with its corresponding dot plot.

A

I

B

II

III

C
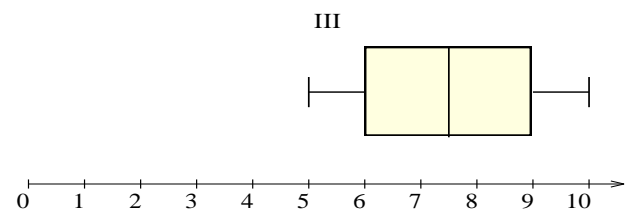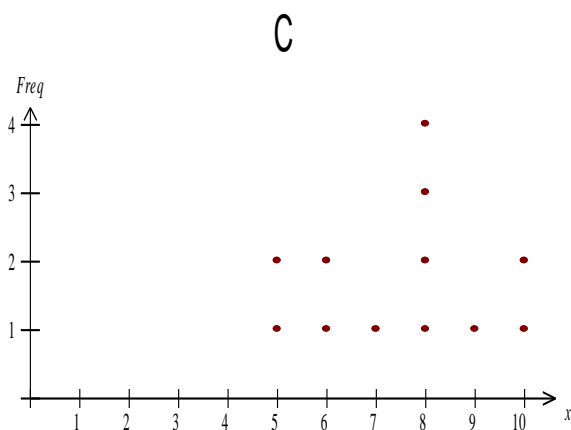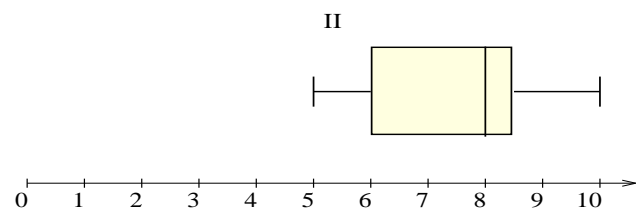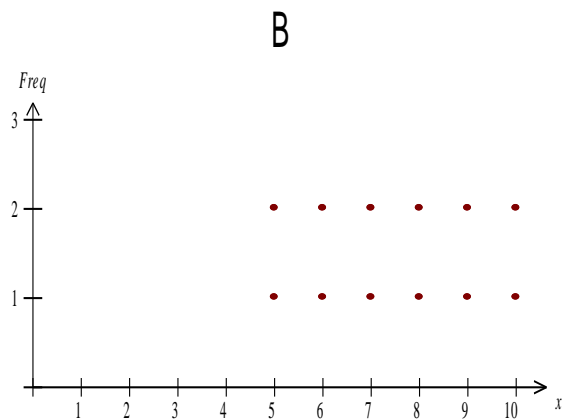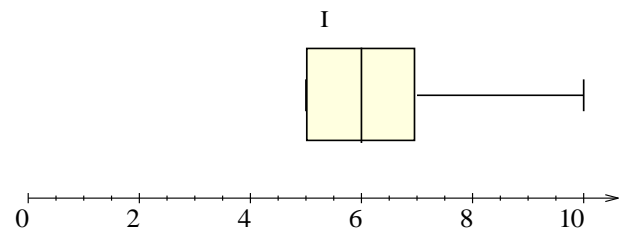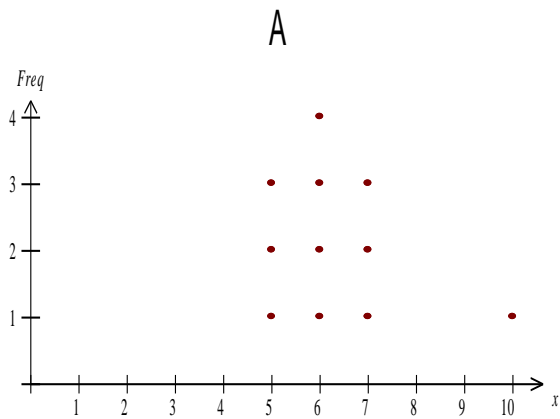
A matches with ___I___ ✓

B matches with ___III_ ✓

C matches with ___II___ ✓

b)   Using the boxplot labeled I, calculate the smallest possible positive integer which
would be considered an outlier.

$IQR = 2$ ✓

$1.5 \times 2 = 3$ ✓

$Q_1 = 5\ 5 - 3 = 2 \therefore 1\ would\ be\ the\ smallest\ possible\ integer.$ ✓

$IQR = 2$ ✓

$1.5 \times 2 = 3$ ✓

Question Five: [2, 2, 2: 6 marks]

A real estate agent is discussing the value of property in a particular suburb with a client.

The real estate agent is quoted saying, "Recent sales in your area indicate that the mean house price is close to $760 000 so we could list your house to sell for that amount."

The histogram below shows the data of the recent house sales in this client's area.
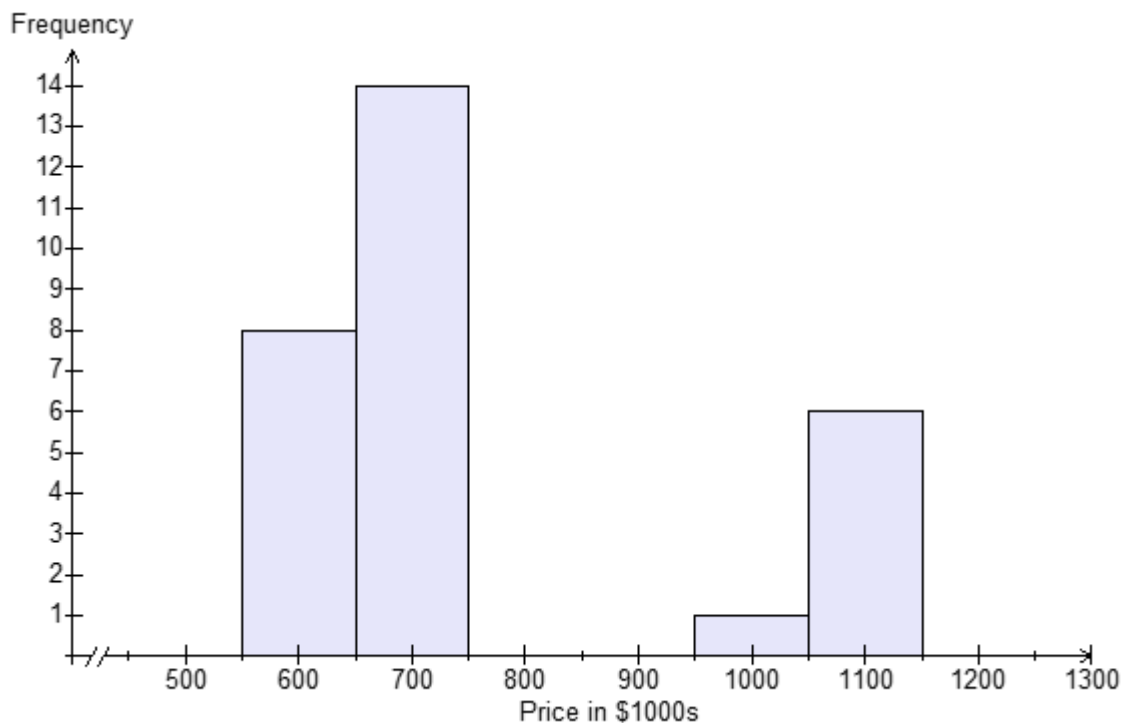
a)   Calculate the estimated mean house price.

$\bar{x} \approx 765\,517$

✓ ✓

b)   Calculate the class median house price.

$Median\ class\ =\ \$650\,000 - \$750\,000$ ✓ ✓

c)   Using statistical measures as reasoning, discuss whether or not the real estate agent's comments are fair.

It is true for the mean but the mean is skewed by a few very high sale prices. The reality is that the majority of sales are actually less than $750 000.

✓ ✓