

Handling Data Cycle

Specimen Task 4: House Prices in Mepton Bridge

Introduction

In this project I am investigating house prices in Mepton Bridge, the town where I live. My sister has just got married and she and her husband are looking to buy their first home. I am going to look at some of the factors that determine house price. The asking price for a house will depend on a number of things, most obviously the size of the plot, the location it is in, the condition of the building, the standard of the decoration, the number of fixtures and fittings, the size of the garden, and so on. I cannot hope to include all of these things in my investigation so I have narrowed it down to these three factors:

1. the number of bedrooms in the house;
2. the total floor space in the house;
3. the distance of the house from the town centre.

My sister is planning to start a family fairly soon so the number of bedrooms will be important to her. On a simple level, I think that the more bedrooms a house has, the larger the house will be, and larger houses tend to cost more, so I expect the number of bedrooms to be one of the factors that determines the asking price for a house.

Hypothesis 1

'There is positive correlation between house price and the number of bedrooms, i.e. the more bedrooms a house has, the more it will cost.'

My sister would like as large a house as possible, room for all the family when the children come along. As I said above, larger houses tend to cost more, so this should be reflected in my next hypothesis.

Hypothesis 2

'There will be positive correlation between house price and the total floor floor space.'

Mepton Bridge is an expanding town with a lot of young families. As it has grown, people have tended to move out from the town centre to the suburbs and I expect this to be reflected in the prices of the houses. Sadly, some parts of the town centre have become run down, with increases in petty crime, and this explains why some people choose to move out to the suburbs.

Hypothesis 3

'There will be positive correlation between house price and the distance from the town centre.'

Plan of Action

To test my hypothesis, I will use primary data. I will collect information by visiting estate agents in Mepton Bridge and collecting leaflets on the houses that they currently have for sale. Some estate agents use the internet to advertise their houses so I will use that as an additional source of information. I know that my data will be somewhat biased because of the time period during which I will collect my data, but there is little I can do to avoid that. My data will also be affected by my choice of estate agents.

I will try to make sure that I get a fair reflection of the town as a whole by choosing 10 different estate agents. From each one, I will use 10 houses, covering the range of prices from the houses they supply. As I have no way of checking the data personally - room sizes for example - I will have to assume that the information that the estate agents supply is accurate. This will give me a sample of 100 houses, selected reasonably randomly from across the town, and this should be enough to test my three hypotheses.

I will construct a pie chart to show the availability of 2, 3, 4 and 5 bedroom houses. I will also draw scatter diagrams to test my three hypotheses. As my sister is looking for a family home, she will need at least 3 bedrooms, and it will be useful if I can give her an indication of how much a 3-bedroom house might cost, or a 4-bedroom house in case she has twins. The scatter diagram for Hypothesis 1 should enable me to do this.

I am also going to look at the distribution of houses prices in Mepton Bridge in order to find out what the housing market is like at the moment. I will draw a histogram for house price - to see, overall, if houses in Mepton Bridge are relatively cheap, reasonably priced or expensive. I will draw a second histogram for distance from the town centre - to see where the houses for sale tend to be situated. I will construct cumulative frequency diagrams for house price and distance from town centre. From these I will determine the medians and quartiles, which I will then use to construct box and whisker diagrams in order to compare the variability in these two factors.

Data Collection

To try to reduce the bias that would result from fluctuations in house prices, I collected my data over a period of just one week. I went to 7 estate agents. I visited 2 on my way home from school one day, and I got details from 5 more when I went to the town centre the following weekend. That weekend, I located 3 more local estate agents with internet websites. From each estate agent I got the details for 10 houses.

Data Processing

I then entered the data for each house in a database that I set up. I recorded the length and width (in metres) for the lounge, the hall, the kitchen, each of the bedrooms, and for any extra (utility) room; the information provided by the estate agents gave each of these measurements to 1 decimal place (e.g. 3.7 m). I then set up a spreadsheet function to work out the area (in m^2) of each room

and to add them together to get the total floor space of the house; the total areas were to 2 decimal places, which seemed unnecessarily accurate, so I rounded them to the nearest m². For each house I also recorded the asking price, the number of bedrooms and the distance from the town centre. As I didn't have time to visit them all, I assumed that the houses were located at the middle of their road, then used the scale on an A-Z map of Mepton Bridge to estimate the distances from the town centre. This was bound to be somewhat inaccurate because some streets are longer than others, but it was the best I could do.

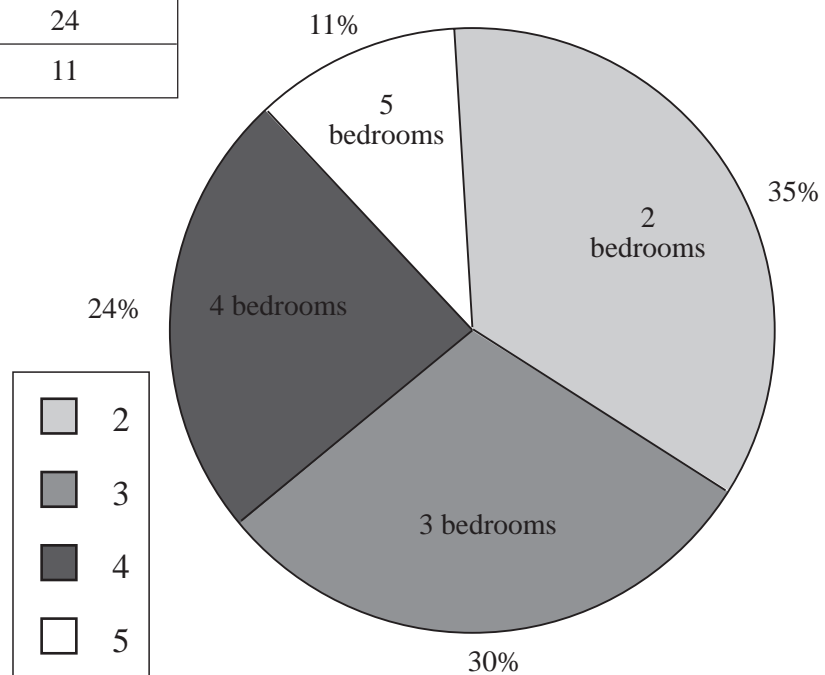
The Data Appendix (pages 1 and 2) at the back of this document show the data for the first 20 houses from my database.

Data Representation

The first thing that I looked at was how many bedrooms the houses in Mepton Bridge have. I used the COUNTIF function in EXCEL to work out how many houses there were in my sample with 2, 3, 4 and 5 houses. I then used EXCEL to construct this pie chart.

<i>Number of bedrooms</i>	<i>Number of bedrooms</i>
2	35
3	30
4	24
5	11

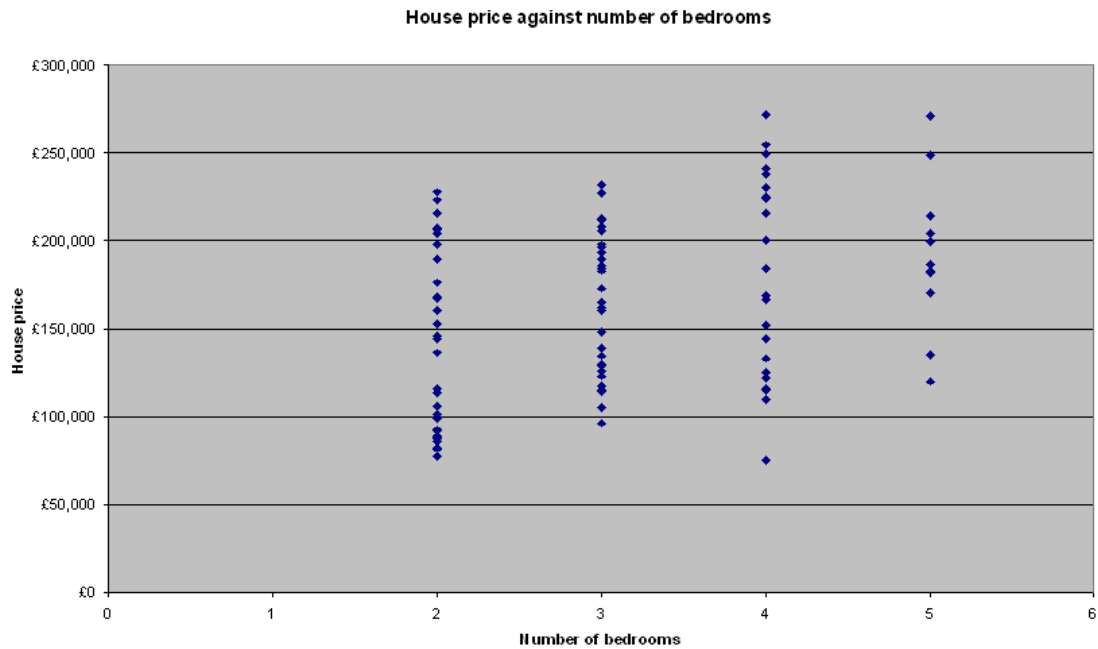
Pie Chart showing number of bedrooms



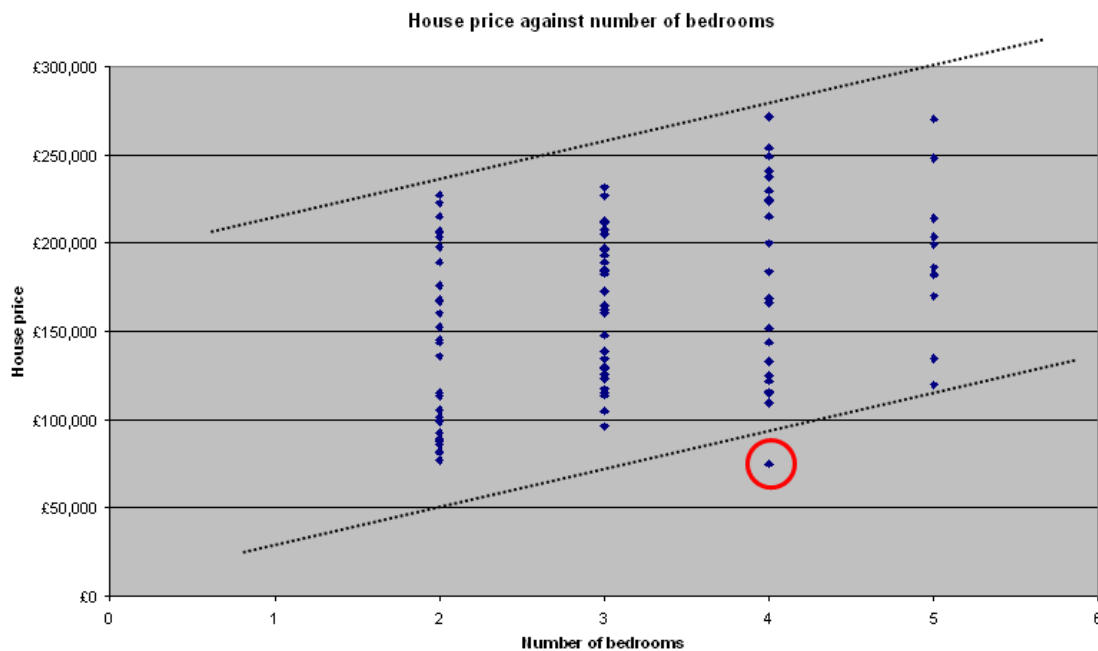
The pie chart shows that, with 35% of the share, 2-bedroomed is the most common type of house on the market at the moment in Mepton Bridge. If my sister wants a 3- or 4-bedroomed house then she has just over half the market (54%) to choose from.

I then went on to investigate my first hypothesis that there is positive correlation between house price and the number of bedrooms, i.e. the more bedrooms a house has, the more it will cost.

In order to do this, I used EXCEL to construct the scatter diagram below showing house price plotted against number of bedrooms.

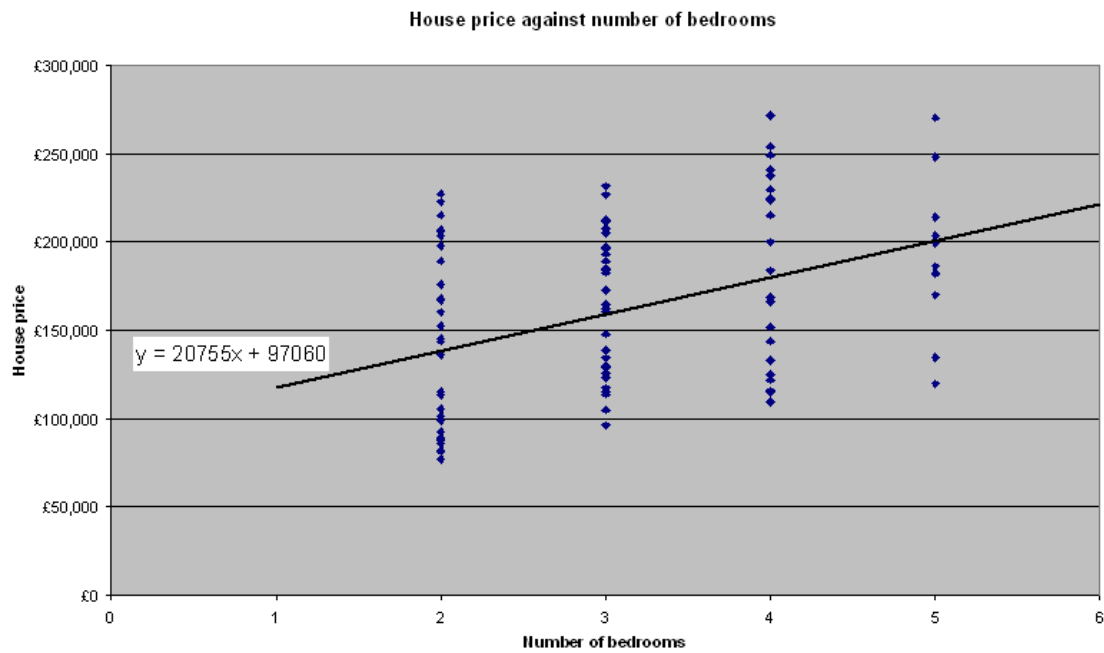


Although the evidence is not clear-cut, there is some indication that houses with more bedrooms tend to cost more. You can see this on the second version of this scatter diagram where I have drawn a broad band to show the positive correlation between house price and number of bedrooms.



The broad trend shown in this scatter diagram does not include one point, which I have circled. This point is an outlier. It represents a property that is very cheap, even though it has a large number of bedrooms. It may be that this is because the property is in poor condition so needs a lot of repair and re-decoration. Alternatively, it may be in a less desirable location and have been on the market some time, so has had its price reduced in order to try to get it sold.

Since I have some evidence of correlation, I removed the outlier from the scatter diagram and got EXCEL to draw a trendline (see diagram below). I used the EXCEL option to display the trendline with its equation.



The line of best fit (i.e. the trendline) has equation $y = 20755x + 97060$, i.e.
 House Price = £[(20755 × Number of Bedrooms) + 97060]

This means that,

for a 3-bedroomed house, my sister should expect to pay

$$£20755 \times 3 + 97060 = £159325$$

i.e. about £160 000;

for a 4-bedroomed house, she should expect to pay

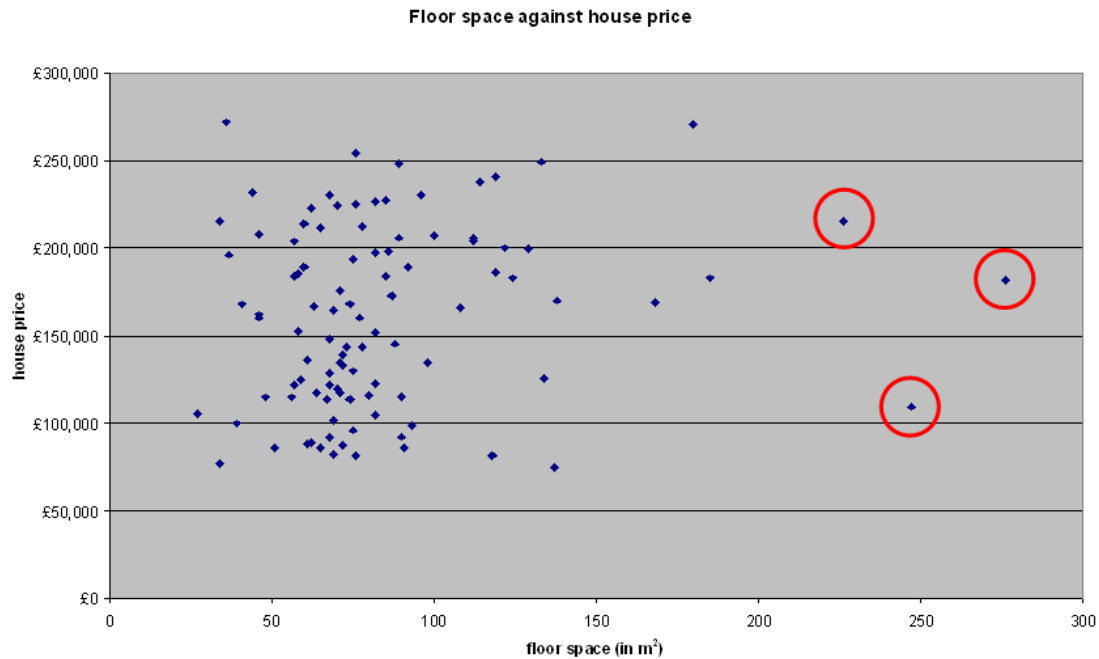
$$£20755 \times 4 + 97060 = £180080$$

i.e. about £180 000.

Summary Table

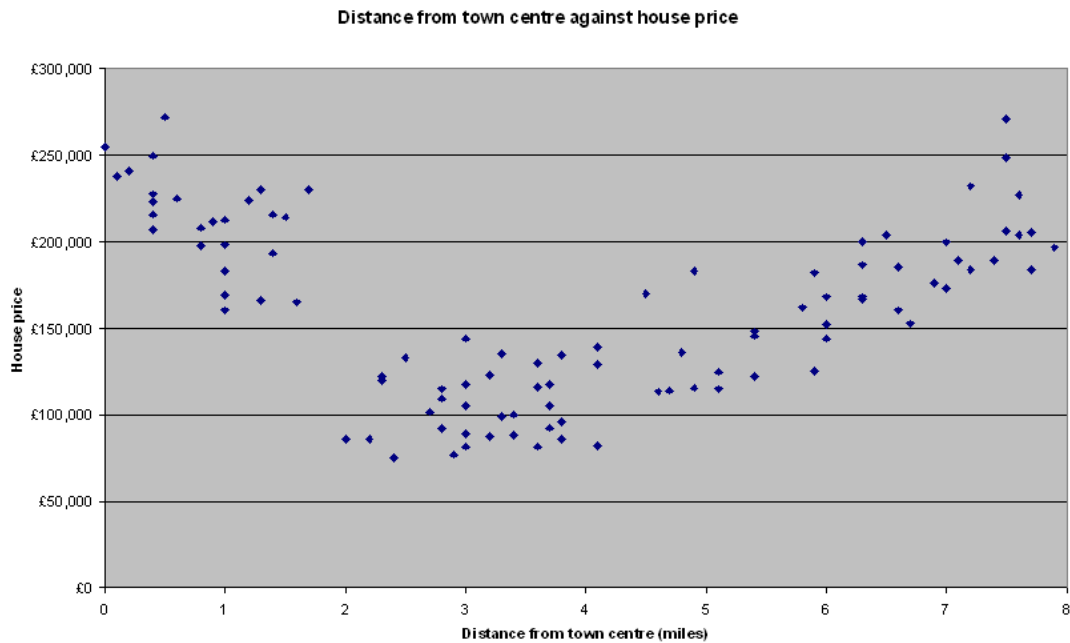
<i>Number of Bedrooms</i>	<i>Typical price</i>
2	£140 000
3	£160 000
4	£180 000
5	£200 000

Having found some correlation between house price and the number of bedrooms, I confidently expected to find stronger correlation between house price and total floor space because more bedrooms should mean more floor space. I was therefore most surprised when I got the scatter diagram below using my EXCEL database.



There are three houses here that do not seem typical; I have ringed these. These houses seem to be relatively cheap given their large floor space. It may be that these are medium priced, despite their size, because they are in less desirable locations. Even if I removed these 3 houses, there would still be no evidence of correlation between house price and floor space. My second hypothesis was therefore wrong. I suspect this may be because people consider location to be a more important factor than house size.

My third scatter diagram plotted house price against distance from town centre (see below). Overall, there was again no correlation, but there was a pattern in the house prices. There seemed to be some negative correlation for houses close to the town centre, and positive correlation - as I predicted in Hypothesis 3 - for houses some way out from the town centre. From the scatter diagram, the cut off distance seems to be about $1\frac{1}{2}$ to 2 miles. I think the negative correlation may reflect the attempt by the council to regenerate the town centre by converting a number of houses there into up-market dwellings for professional people. The positive correlation, I think, reflects the common wish that people have to live in the countryside, away from run down areas, but near enough to be able to drive into town to shop and to use the various amenities (cinemas, theatres, restaurants, etc.).



I decided to investigate the two types of correlation separately, using 2 miles as the cut off point. I sorted my data into order by distance from the town centre, and then separated the data into two groups: 'under 2 miles from town centre' and '2 or more miles from the town centre'. The diagram for houses close to the town centre did show some negative correlation (see diagram on next page) and gave a line of best fit equation of $y = 239265 - 30973x$ This means that the equation

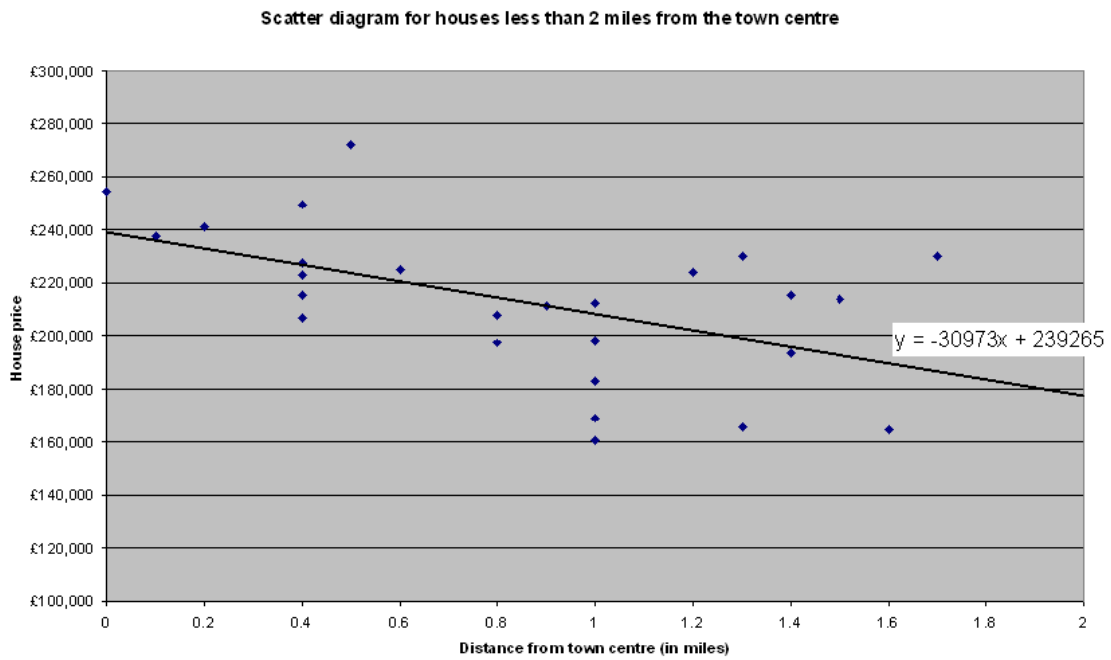
$$\text{House Price} = \text{£}[239265 - (30973 \times \text{distance from town centre})]$$

can be used to estimate the price of a house less than 2 miles from the town centre.

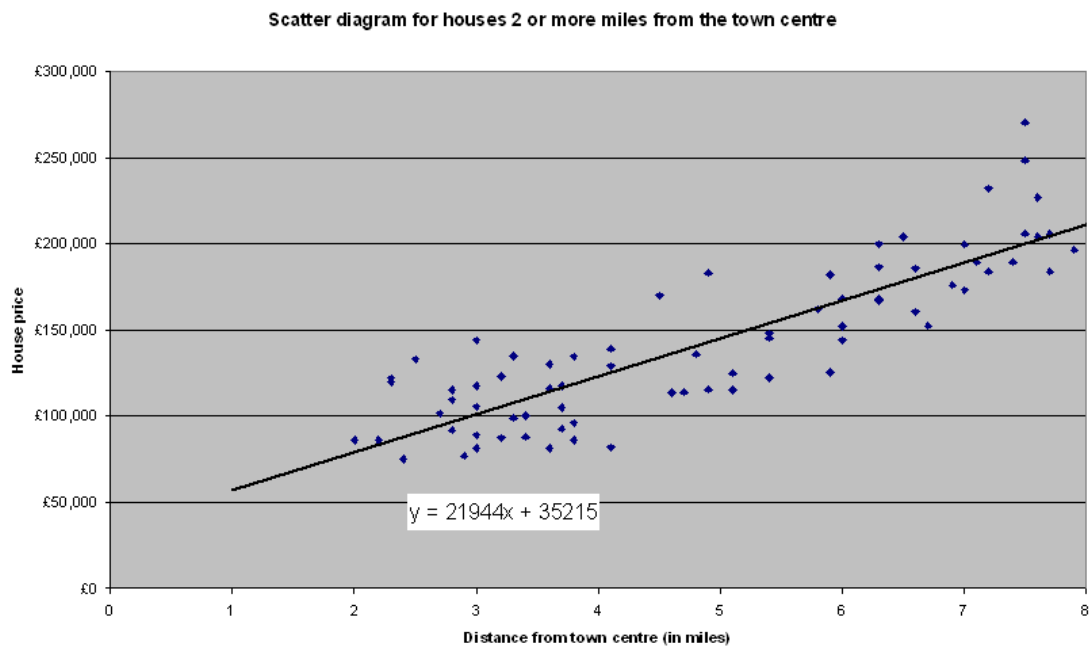
So a typical town centre house $\frac{1}{2}$ mile from the centre will cost

$$= \text{£}[239265 - (30973 \times 0.5)] = \text{£}223778.50, \text{ i.e. about } \text{£}224\,000.$$

The constant term in the equation also gives a price of about £240 000 for a house right in the centre of town.



The scatter diagram for houses 2 or more miles from the town centre showed strong positive correlation. I again got EXCEL to display the line of best fit and to give the equation for this trendline.



This equation, $y = 21944x + 35215$, i.e.

$$\text{House Price} = \text{£}[(21944 \times \text{distance from town centre}) + 35215]$$

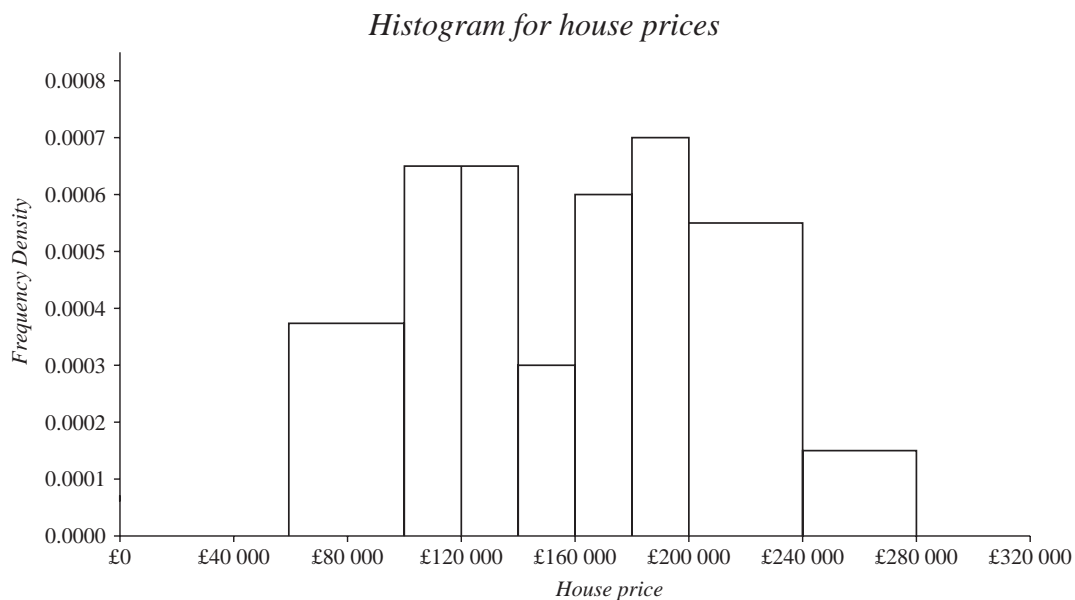
can be used to estimate the price of a house 2 or more miles from the town centre. For example, a house 6 miles from the town centre should cost around $21944 \times 6 + 35215 = \text{£}166879$, i.e. about £167 000. This is a bit more than the £160 000 I estimated earlier for a typical 3-bedroomed house.

My sister can now link my three scatter diagrams (for house price against number of bedrooms, and house price against distance from town centre) to decide how far from the town centre she needs to live to get a house in her price range and with the number of bedrooms that she wants.

I then looked at the distributions of house price and distance from the town centre. I used the COUNTIF function on EXCEL to produce a cumulative frequency table for house price, which I then used to produce this frequency table:

<i>House Price</i>		<i>Frequency</i>	<i>Frequency density</i>
£0	≤ price < £60 000	0	0
£60 000	≤ price < £100 000	15	0.000375
£100 000	≤ price < £120 000	13	0.000650
£120 000	≤ price < £140 000	13	0.000650
£140 000	≤ price < £160 000	6	0.000300
£160 000	≤ price < £180 000	12	0.000600
£180 000	≤ price < £200 000	14	0.000700
£200 000	≤ price < £240 000	21	0.000525
£240 000	≤ price < £280 000	6	0.000150

I then constructed the histogram below for this data.



This histogram surprised me. I had expected the distribution to be symmetrical, perhaps following an approximate normal distribution, with typical mid-priced houses being the most common. However, the distribution I got showed a bimodal distribution with two main price categories, namely £100 000 - £139 999 at the lower end of the price range, and £160 000 - £239 999 at the higher end of

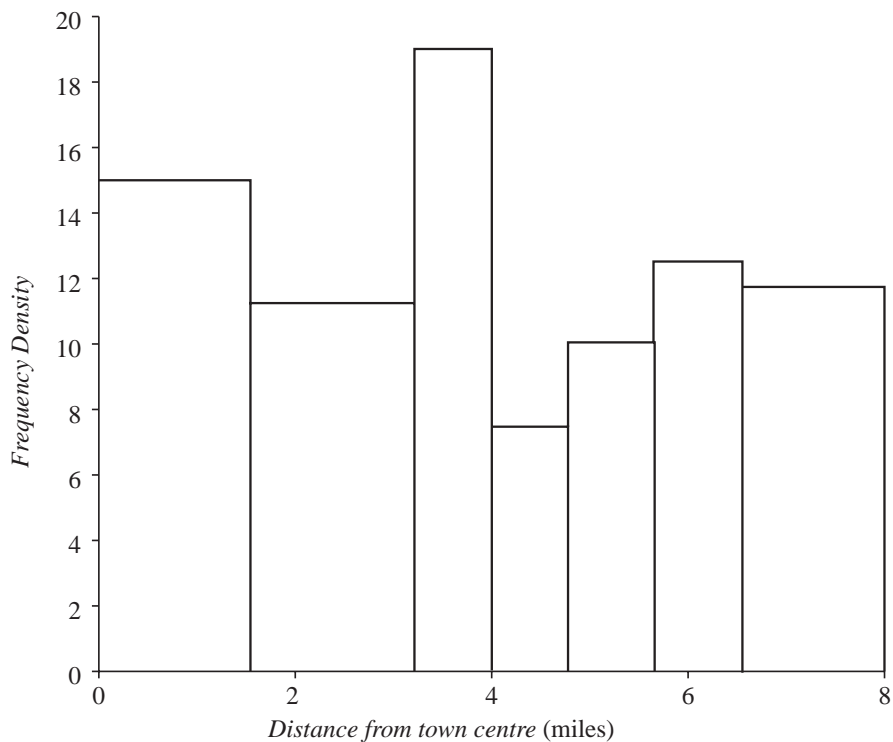
the price range. I think the lack of houses in the range £140 000 - £159 999 may reflect the council's attempt to regenerate the town centre with up-market dwellings for professional people, and the common wish that people have to live out in the suburbs, away from the run down areas, but still within reach of the town's amenities. It may be, alternatively, that there were not many houses on the market in that price range when I collected my data.

I then used the same approach to look at the distribution of distance from the town centre. This produced the following frequency table:

<i>Distance from town centre</i>	<i>Frequency</i>	<i>Frequency density</i>
0 ≤ distance < 1.6	24	15.000
1.6 ≤ distance < 3.2	18	11.250
3.2 ≤ distance < 4.0	15	18.750
4.0 ≤ distance < 4.8	6	7.500
4.8 ≤ distance < 5.6	8	10.000
5.6 ≤ distance < 6.4	10	12.500
6.4 ≤ distance < 8.0	19	11.875

Again, I constructed a histogram to represent the data.

Histogram for distances from town centre

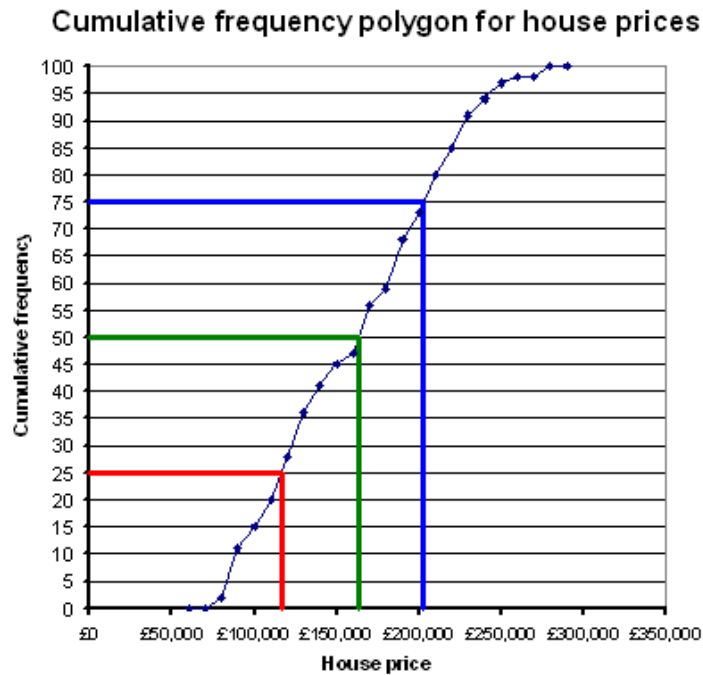


Like the previous histogram, this one also had a 'dip' in the middle, although less pronounced than before. It also had the same style of bimodal distribution. The histogram shows that the majority of housing in Mepton Bridge is within 4 miles

of the town centre, though there is still a substantial quantity of housing 5 to 9 miles from the town centre. The 'dip' around 4 to 5 miles may be due to industrial estates and shopping precincts being located at that sort of distance from the centre, although it may be that there were few houses on the market at that distance in the week that I collected my data.

To extend my investigation I looked at the variation in both house price and distance from town centre. For house price, I used the QUARTILE function on EXCEL to find the minimum, lower quartile, median, upper quartile and maximum house price. For safety, I constructed a cumulative frequency polygon (see diagram on the next page) and checked the values by taking approximate readings for the lower quartile, median and upper quartile. Since there were 100 houses in my sample, these were the house prices corresponding to cumulative frequency 25, 50 and 75 respectively. I also checked the minimum and maximum house prices from the column for that date on my original spreadsheet. I recorded these values on a table so I could show the variation in house price on a box and whisker diagram.

<i>Price under</i>	<i>Cumulative frequency</i>
£60 000	0
£70 000	0
£80 000	2
£90 000	11
£100 000	15
£110 000	20
£120 000	28
£130 000	36
£140 000	41
£150 000	45
£160 000	47
£170 000	56
£180 000	59
£190 000	68
£200 000	73
£210 000	80
£220 000	85
£230 000	91
£240 000	94
£250 000	97
£260 000	98
£270 000	98
£280 000	100

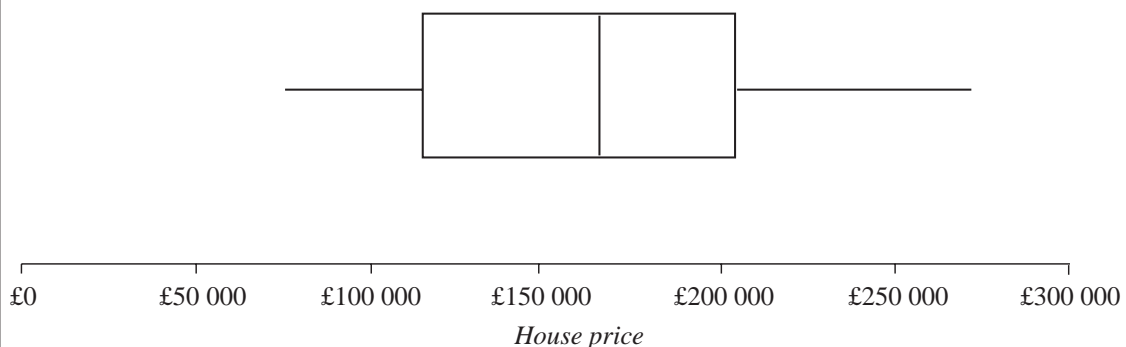


The values that I got are recorded in the following table. I rounded the values off to the nearest £5000 to make it easier to draw the box and whisker plot. This loss in accuracy will not be particularly significant since I am only looking at the distribution and variation in house price.

	<i>Using QUARTILE function</i>	<i>From cumulative frequency polygon</i>	<i>Rounded values used for box and whisker diagram</i>
Minimum	£75 000	£75 000	£75 000
Lower quartile	£115 838	£116 250	£115 000
Median	£163 475	£163 333	£165 000
Upper quartile	£203 963	£202 857	£205 000
Maximum	£271 950	£271 950	£270 000

I put this information on a box and whisker diagram.

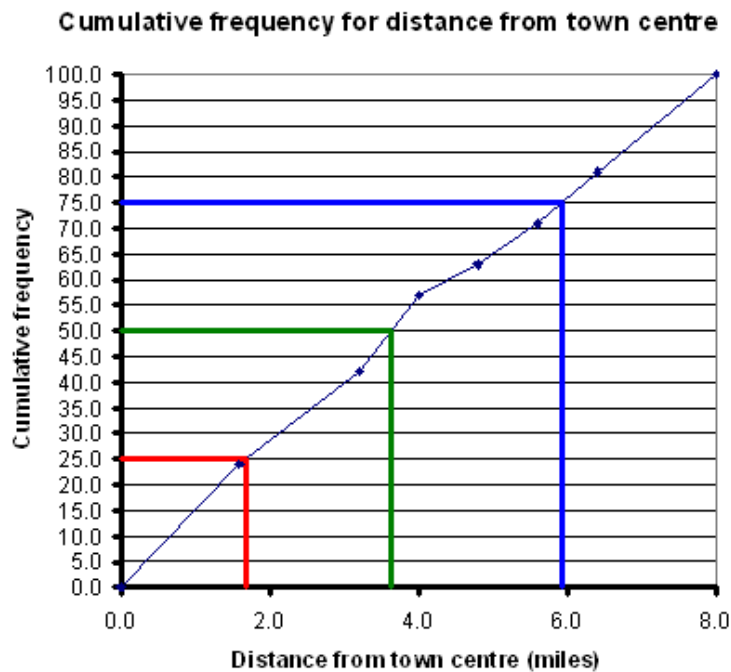
Box and whisker diagram for house prices



The box and whisker diagram showed that house prices in Mepton Bridge are distributed reasonably symmetrically, with slightly more houses at the upper end of the price range.

I then repeated this process for the distances of the houses from the town centre.

<i>Distance from town centre under</i>	<i>Cumulative frequency</i>
0.0	0
1.6	24
3.2	42
4.0	57
4.8	63
5.6	71
6.4	81
8.0	100

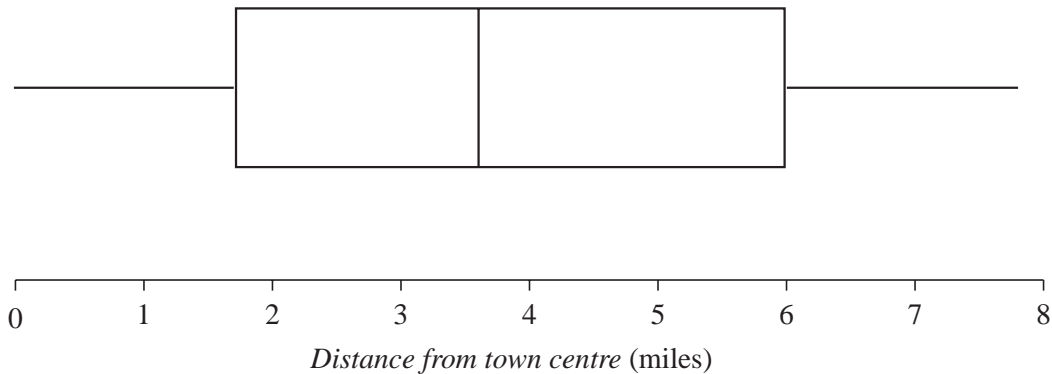


As before, I used the QUARTILE function on EXCEL to find the minimum, lower quartile, median, upper quartile and maximum distance from the town centre. Once again I checked my values by constructing a cumulative frequency polygon (see diagram above) and reading from it. My values are recorded in the next table.

	<i>Using QUARTILE function</i>	<i>From cumulative frequency polygon</i>	<i>Values used for box and whisker diagram</i>
Minimum	0.0	0.0	0.0
Lower quartile	1.7	1.7	1.7
Median	3.6	3.6	3.6
Upper quartile	6.0	5.9	6.0
Maximum	7.9	7.9	7.9

I then constructed the box and whisker diagram below for distance from town centre.

Box and whisker diagram for distance from town centre

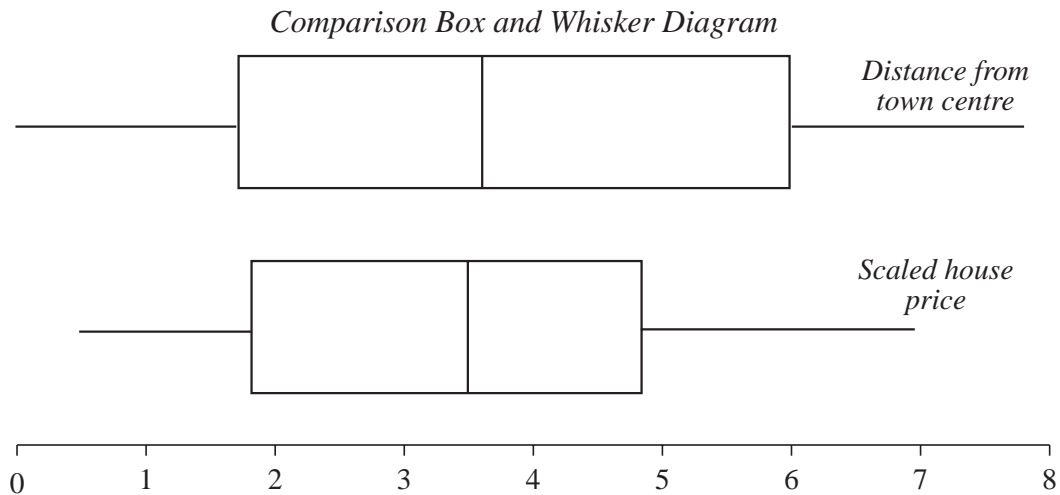


This box and whisker diagram was less symmetrical than the first, giving some indication of positive skew, which means that the house distances tend to occur more towards the lower end of the range. This can be illustrated further by looking at the median and mid-range. The median is 3.6 miles but the mid-range [the half way point from the minimum distance (0.0 miles) to the maximum distance (7.9 miles)] is 3.95 miles. This means that more than 50% of distances are below that mid-range value.

For the final part of my investigation, I decided to compare the two box and whisker plots. However, the range of values for house price was very different from the range of values for distance from the town centre, so an immediate comparison was impossible. The house prices went (approximately) from £60 000 to £300 000, a difference of £240 000. The distances from the town centre went (again approximately) from 0 to 8 miles. I realised that I could make a fair comparison if I scaled the house prices by subtracting £60 000 and dividing the result by £30 000,

$$\text{i.e. Scaled House Price} = \frac{\text{House Price} - \text{£60000}}{\text{£30000}}$$

This allows 'distance from town centre' and 'scaled house price' to both be represented on a common scale of 0 to 8 (see diagram at the top of the next page), which will then make the comparison a fair one.



The widths of the 'boxes' on the box and whisker diagrams show that the house prices are far less varied than the distances from the town centre. Putting the diagrams one above the other also emphasised the greater symmetry of the distribution of house prices and the positive skew of the distances from the town centre.

FINAL CONCLUSIONS

Overall, my investigation of the housing market in Mepton Bridge has reached the following conclusions:

- For houses 2 or more miles from the town centre, there is evidence of positive correlation between house price and distance from the town centre, i.e. the further you are from the town centre, the higher the price of housing {though clearly this pattern would change significantly once you were well out into the countryside}.
- For houses closer to the town centre than 2 miles, there is evidence of negative correlation between house price and distance from the town centre, i.e. the nearer you are to the town centre, the more houses cost.
- This difference is probably a result of the town council's attempts to regenerate the town centre by encouraging up-market housing there.
- There is some evidence of positive correlation between the price of a house and the number of bedrooms, though the evidence is not clear cut. This means you should expect to pay more for a house with more bedrooms when you are considering two properties in the same location.
- Surprisingly there is no evidence of correlation between house price and total floor space. It may be that people consider location (as measured here by distance from town centre) and number of bedrooms as more significant factors when deciding to buy a house. They may look at overall floor space later to decide between properties of comparable price, location and number of bedrooms. Alternatively, it may be that people do not consider overall floor space as being particularly significant; for example, some houses might have a large floor space overall but actually consist of a large number of

small, pokey rooms {as is the case in housing areas occupied by college and university students}.

- In Mepton Bridge there is a reasonable spread of housing prices, from £115 000 to £205 000 {the 25th and 75th percentiles respectively}.
- A typical house in Mepton Bridge costs £165 000 {the median price}, will have 3 bedrooms and be situated about 6 miles from the town centre.
I calculated these figures for the median house by re-arranging my formulae

$$\text{Formula 1 House Price} = \text{£}[(20755 \times \text{Number of Bedrooms}) + 97060]$$

$$\text{Formula 2 House Price} = \text{£}[239265 - (30973 \times \text{distance from town centre})] \text{ (#)}$$

$$\text{Formula 3 House Price} = \text{£}[(21944 \times \text{distance from town centre}) + 35215] \text{ (*)}$$

(#) for houses less than 2 miles from the town centre

(*) for houses 2 or more miles from the town centre

Because of the context, I rounded the calculated values {3.27342809 bedrooms and 5.914372949 miles} to the nearest whole number.

N.B. Formula 2 did not predict a house at the median price within 2 miles of the town centre.

- Towards the higher end of the price range, a typical house in Mepton Bridge costing £205 000 {the 75th percentile}, will have 5 bedrooms and be situated either about 1 mile, or 8 miles, from the town centre.

At the cheaper end of the price range, a typical house in Mepton Bridge costing £115 000 {the 25th percentile}, will have 1 bedroom and be situated about $3\frac{1}{2}$ miles from the town centre.

ADDITIONAL COMMENTS AND IDEAS FOR FURTHER DEVELOPMENT

My investigation was purely into housing in Mepton Bridge. It was based on a reasonably large sample, and I did my best to select that sample fairly in order to try to reflect the town's housing stock overall. It is possible that my conclusions have been influenced by the estate agents that I selected, and by the properties that were on the market at the time I gathered my data. If I were to extend this investigation further then I would obtain a larger database by using all of the town's estate agents, thereby ensuring coverage of the whole town, and by taking data on more properties from each agent.

I would also try to compare the housing market in the town of Mepton Bridge with that in one of the outlying small villages, Holland-on-the-Wold for example. I would expect to find a very different pattern to the house prices in a village. There, distance from the village centre is likely to be much less of a factor in determining house price than number of bedrooms, overall floor space, size of garden and whether the house is detached.

DATA APPENDIX

HOUSE NUMBER	Lounge		Hall		Kitchen		Bedroom 1		Bedroom 2		Bedroom 3		Bedroom 4	
	Width	Length	Width	Length	Width	Length	Width	Length	Width	Length	Width	Length	Width	Length
1	2.7	5.2	0.9	2.7	3.0	4.6	2.1	4.0	2.7	2.7	0.0	0.0	0.0	0.0
2	4.6	5.5	1.2	1.5	1.5	2.7	5.2	6.4	2.4	3.4	0.0	0.0	0.0	0.0
3	4.6	7.0	1.2	2.1	3.4	4.0	5.2	5.2	4.9	4.9	0.0	0.0	0.0	0.0
4	3.7	4.6	0.9	3.4	1.2	2.1	4.3	6.4	1.5	2.1	1.8	2.7	0.0	0.0
5	3.4	4.3	0.6	3.4	4.0	4.6	4.0	4.9	1.5	4.0	1.2	2.4	0.0	0.0
6	4.0	4.9	1.2	1.8	2.4	3.7	2.1	4.9	2.7	4.9	0.0	0.0	0.0	0.0
7	2.4	5.2	1.2	2.1	1.8	2.7	4.0	5.8	4.0	5.2	0.0	0.0	0.0	0.0
8	3.0	4.3	1.2	1.8	2.7	4.0	4.3	6.1	2.7	3.0	1.2	3.7	0.0	0.0
9	2.1	3.0	1.2	3.7	3.0	4.3	3.0	4.0	3.7	5.5	0.0	0.0	0.0	0.0
10	3.7	4.0	0.6	0.6	4.0	5.2	3.7	6.4	3.7	5.5	0.0	0.0	0.0	0.0
11	2.1	2.4	0.6	3.4	1.8	3.0	3.0	5.2	2.4	2.7	0.9	1.5	1.5	4.3
12	4.6	4.6	1.2	3.4	3.0	3.0	2.1	2.4	1.5	1.8	2.1	4.6	0.0	0.0
13	2.4	3.0	1.2	3.0	2.1	3.0	5.5	7.0	2.1	4.0	0.0	0.0	0.0	0.0
14	2.7	3.4	0.9	1.2	1.8	2.1	3.7	3.7	3.4	5.5	0.0	0.0	0.0	0.0
15	1.8	4.0	0.6	0.9	1.2	1.2	4.9	7.3	3.4	3.4	0.0	0.0	0.0	0.0
16	4.6	5.2	0.9	3.7	3.0	4.0	3.0	5.8	2.4	3.0	0.0	0.0	0.0	0.0
17	2.1	3.4	0.6	2.4	1.5	1.5	4.9	5.5	1.2	1.8	2.4	2.7	0.9	1.5
18	2.4	5.2	1.2	1.5	3.0	3.7	2.7	2.7	1.8	3.4	3.0	4.6	4.3	4.3
19	3.7	5.2	0.6	3.4	3.4	4.3	5.5	6.1	4.6	5.8	4.3	6.4	0.9	2.7
20	2.7	4.9	0.9	0.9	1.8	2.1	2.1	4.3	1.5	3.7	0.0	0.0	0.0	0.0

DATA APPENDIX

HOUSE NUMBER	Extra room		Number of Bedrooms	Floor space	Rounded floor space	Distance from town centre	Advertised price
	Width	Length					
1	3.7	5.8	2	67.42	67	4.6	£113 500
2	3.7	4.9	2	90.72	91	2.2	£86 000
3	3.7	4.9	2	117.50	118	3.0	£81 500
4	3.7	6.4	3	81.81	82	3.2	£123 000
5	1.8	4.6	3	71.82	72	4.1	£139 000
6	1.8	2.1	2	57.94	58	6.7	£152 500
7	1.8	2.7	3	68.72	69	1.6	£164 950
8	1.8	3.4	3	70.75	71	3.7	£117 500
9	3.7	4.0	2	70.79	71	6.9	£176 000
10	1.8	4.3	2	87.73	88	5.4	£145 500
11	6.1	8.8	4	96.04	96	1.7	£230 000
12	3.7	6.4	3	75.32	75	3.8	£96 000
13	0.0	0.0	3	64.00	64	3.0	£117 500
14	0.0	0.0	2	46.43	46	6.6	£160 500
15	0.0	0.0	2	56.51	57	7.6	£204 000
16	1.8	3.0	2	69.25	69	2.7	£101 500
17	2.4	4.9	5	59.53	60	1.5	£214 000
18	0.0	0.0	5	71.08	71	3.3	£134 950
19	2.4	4.6	4	137.12	137	2.4	£75 000
20	1.8	3.7	2	39.06	39	3.4	£100 000