MATHS COURSEWORK

AIM:

The aim of this work is to find out if there is a correlation between the temperature and the humidity in Cambridge. I will collect my data from points in and around Cambridge

HYPOTHESIS

I predict that as the humidity increases the temperature will decrease.

PLAN

I have decided to collect my data in Cambridge using two accurate instruments, a digital thermometer and a digital hygrometer. These will provide me with accurate samples of the temperature and humidity at the different survey points. I also believe that the height of the buildings (i.e. 1 story, 2 story …etc) and the type of land (grass, concrete … etc) will affect at least one of the 2 probables in my data. I suspect that collecting more than 2 sets of data will give me a wide range of things to discuss. As well as this I believe having this wide range of data may help to explain any anomalies that may occur.

The sixty survey points I am going to collect data from will be decided for me when I go on my geography trip. This is not an ideal situation however these survey points are wide ranged and have a variety of different surroundings so they should give me a wide range of results.

The theoretical ideal situation would be to use a random number generator which would give 60 points on a map creating a sample which was random. However this would result in a problem as the random generator would quite often give me a co-ordinate that is inaccessible. Even though my data is not random I am collecting primary data.

I believe collecting data from different heights will change the temperature as well as the humidity so I will record at different heights around Cambridge; this is because every 10 metres climbed the temperature drops by 1%. Pedestrian density may also affect the results due to respiration; this will make the data more biased. Cars will also affect the humidity and temperature; this will also cause a problem as the more densely populated area will have a higher temperature and humidity compared to the other survey points.

I will try to take the readings at as simultaneously as possible. I will use a digital thermometer which is accurate to 0.1%, and a digital hygrometer which is accurate to every 1%. All data will also be collected in the shade so the equipment will not be in direct sunlight.

After I have collected my primary data I will collect secondary data to further test my hypothesis.

RAW DATA

Table of Results and Readings - Comparison Readings

Survey Point	Temperature (ºC)	Humidity(%)	Land use	Building Height (Stories)
1	23.7	43	t	2
2	23.3	42	g	0
3	22.8	23	g	0
4	23.5	42	g	0
5	25.8	40	g	0
6	26.1	36	t	3
7	24.3	38	t	2
8	24.6	41	t	3
9	26.3	37	g	0
10	24.2	39	t	3
11	22.6	43	t	3
12	21.3	51	g/w	0
13	20.9	51	g/w	0
14	21.2	51	g/w	0
15	21.0	52	g/w	0
16	21.5	49	g/w	0
17	22.7	45	t	4
18	23.5	53	t	2
19	24.0	41	t	2
20	23.5	42	t	2
21	25.5	38	g	0
22	25.7	38	t	3
23	24.3	39	g	0
24	24.6	38	g	0
25	24.2	35	t	3
26	23.9	36	t	3
27	24.2	36	t	3
28	23.7	32	t	3
29	24.4	36	t	4
30	24.7	37	t	4
31	24.6	36	t	4
32	24.3	40	t	3
33	25.2	39	t	3
34	23.9	41	g	0
35	25.5	39	t	5
36	25.3	38	t	3
37	25.3	38	t	3
38	25.1	39	t	3
39	24.9	37	t	3
40	24.0	40	t	3
41	24.2	37	t	3
42	24.7	39	t	3
43	25.2	38	t	3
44	25.9	39	t	3
45	25.9	42	t	4
46	25.5	38	t	4
47	25.0	36	t	5

48	25.3	35	t	5
49	24.9	37	t	4
50	24.8	37	t	4
51	25.0	36	t	4
52	24.6	38	t	3
53	24.7	38	t	3
54	24.5	38	t	3
55	24.4	40	t	2
56	24.6	39	t	2
57	24.1	41	t	2
58	23.4	43	g	0
59	23.0	45	g	0
60	23.5	42	g	0

The letters in the third column describe the type of land around the reading point. This is the key to those letters.

```
Tarmac/Paving/Buildings -
t
Water -
w
Grass -
G
```

BIASES

I believe there may well be a bias in the height in the buildings as I explained above so I want to have a look at my data to see if my data is biased as I believe the data will be affected by the nearby buildings. The bar chart below shows the amount of recordings with buildings of 1 stories, 2 stories … etc.
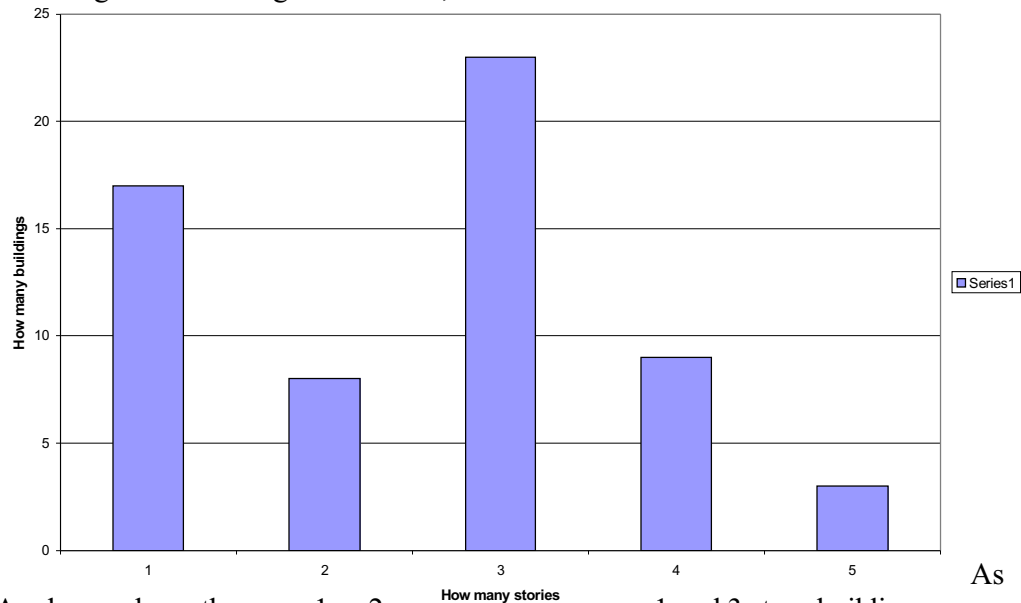
As shown above there are 1 or 2 more 1 and 3 story buildings surrounding the survey points than the others.

As shown below there is a table below which shows the mean humidity and temperature for each land use / building height, and the mean humidity divided by the mean temperature.

|				
Grass	12	39.167	24.2	1.6185
Water	5	50.8	21.18	2.3984
2 stories	8	42.125	24.0125	1.7542
3 stories	23	38	24.6565	1.5411
4 stories	9	38.2	24.72	1.5453
5 stories	3	36.67	25.267	1.4513

Grass and water = 0/1

The first thing I will do with this data is to put the data into a scatter graph to find out if there is a correlation in my results.
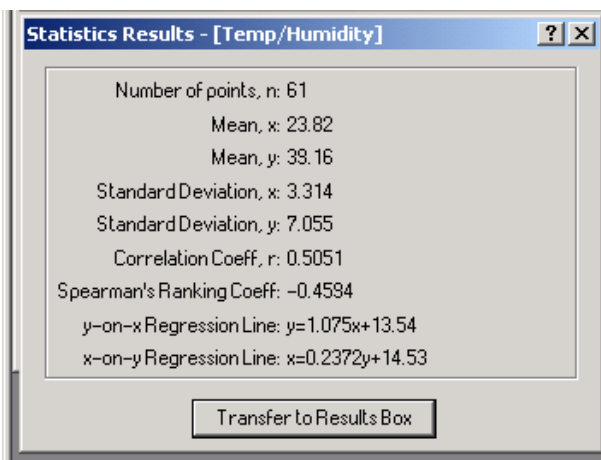
SCATTER GRAPH

In all scatter graphs in this paper the y axis measures humidit y, and the x axis measures temperature.

RAW DATA
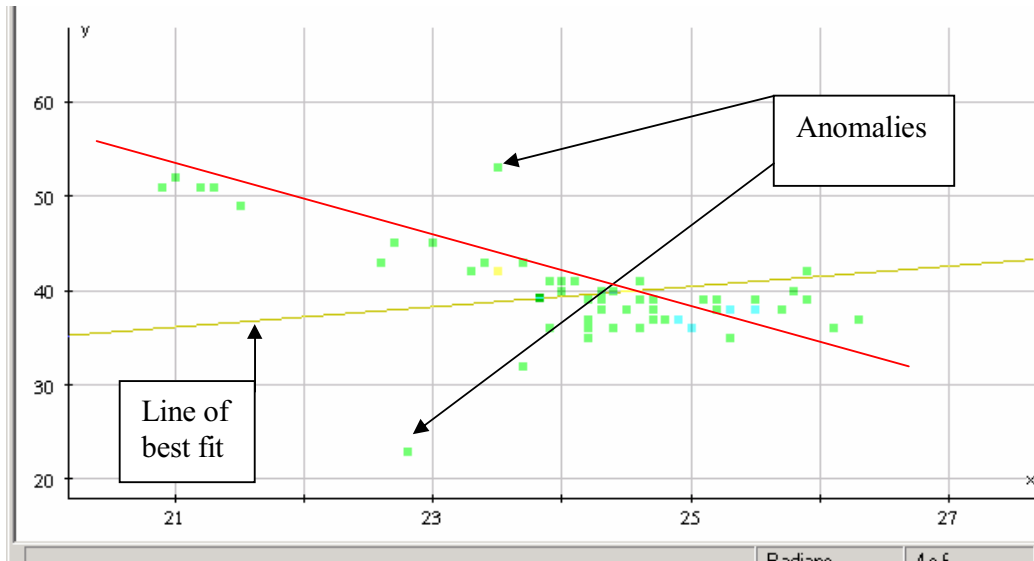
As shown above there are quite obviously 2 anomalies here which have disturbed the line of best fit. The red line I have added into the picture above is quite obviously closer to the line of best fit that autograph has put in for me. This means that the anomalies have affected the line of best fit so I will remove them for my next graph to get a more accurate result.
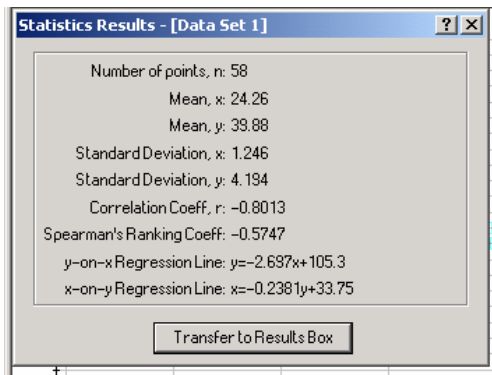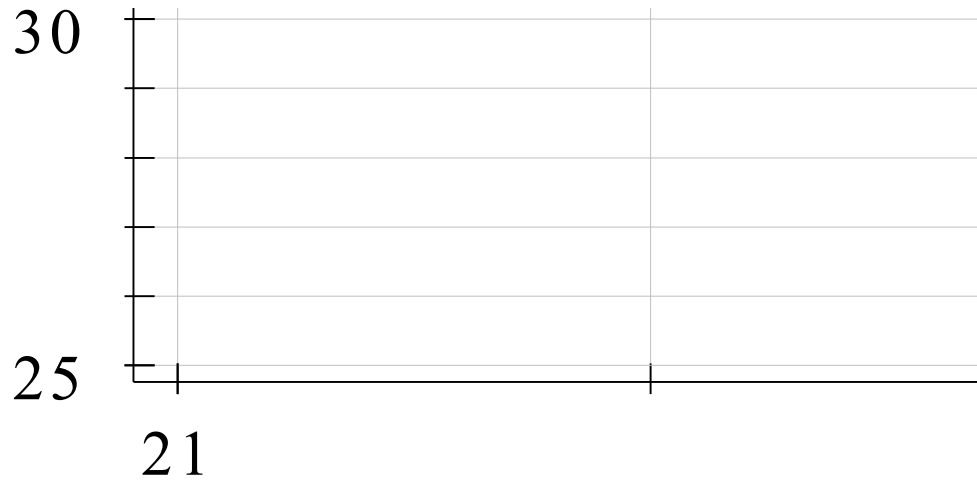

Statistics Results – [Temp/Humidity]

Number of points, n: 61
Mean, x: 23.82
Mean, y: 39.16
Standard Deviation, x: 3.314
Standard Deviation, y: 7.055
Correlation Coeff, r: 0.5051
Spearman's Ranking Coeff: –0.4594
y-on-x Regression Line: y=1.075x+13.54
x-on-y Regression Line: x=0.2372y+14.53

Transfer to Results Box

As you can see the correlation is poor as it is 0.50501 so I will rid the graph of the anomalies.

(see graph on following page)

NO ANOMOLIES

This is a scatter graph for all of the data I collected excluding the anomalous results (23.7, 32), (23.5, 53) and (22.8, 23).

30

25

21

Statistics Results - [Data Set 1] ? X

Number of points, n: 58
Mean, x: 24.26
Mean, y: 39.88
Standard Deviation, x: 1.246
Standard Deviation, y: 4.194
Correlation Coeff, r: −0.8013
Spearman's Ranking Coeff: −0.5747
y-on-x Regression Line: y=−2.697x+105.3
x-on-y Regression Line: x=−0.2381y+33.75

Transfer to Results Box

The diagonal line is the line of best fit given to me by "autograph".

I have got rid of the anomalies from the previous graph and come out with a much better line of best fit than of the previous graph. This is a much more accurate graph as the coefficient correlation is -0.8013 this is much closer to the 1 which is an absolutely perfect correlation graph.
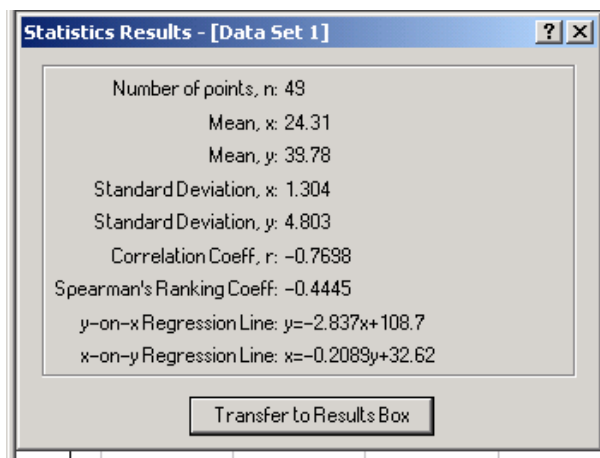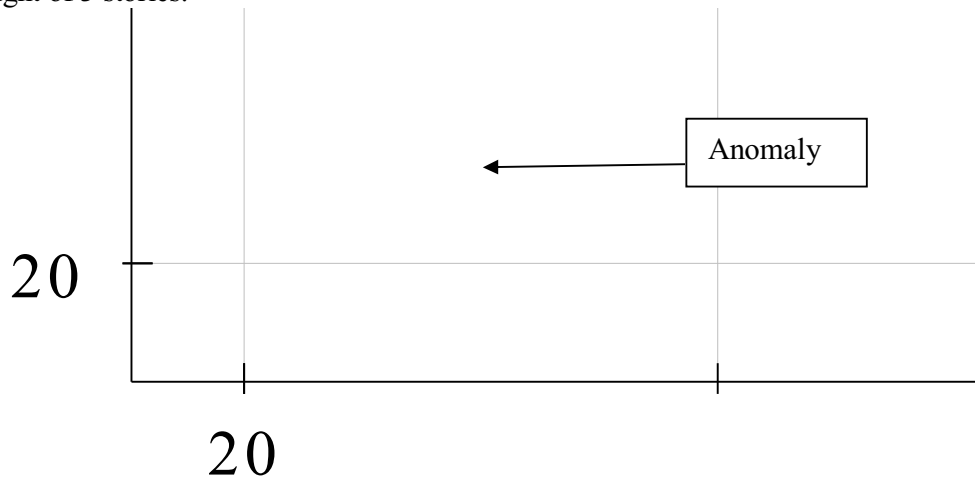
OTHER SCATTER GRAPHS

Since I have found a closely coefficient graph I will now try to find if my data can be changed to find if there is an even closer coefficient correlation than that which I already have.

I first tried just grass and water land type as we did not take the.
The Correlation = 0.09644

I then tried all the buildings/tarmac … etc.
The Correlation = 0.1085

As the correlation in both cases was so low I am not showing these graphs.

The following graph takes account of all land types (grass, water, tarmac … etc) up to a height of 3 stories.

Anomaly

20

20

Statistics Results - [Data Set 1]

Number of points, n: 49
Mean, x: 24.31
Mean, y: 39.78
Standard Deviation, x: 1.304
Standard Deviation, y: 4.803
Correlation Coeff, r: –0.7698
Spearman's Ranking Coeff: –0.4445
y–on–x Regression Line: y=–2.837x+108.7
x–on–y Regression Line: x=–0.2089y+32.62

Transfer to Results Box

The blue line on the graph is the line of best fit given to me by "autograph".

There appears to be only one anomaly.
However the coefficient correlation was only 0.7698 which although it is good it is not as good as the graph on page 7. (This is because in the graph on page 7 I have rid the graph of anomalies and in this graph I have not)

STEM AND LEAF

HUMIDITY

```
0:
10:
20:    3
30:    2 5 5 6 6 6 6 6 6 6 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9
40:    0 0 0 0 1 1 1 1 2 2 2 2 2 3 3 3 5 5 9
50:     1 1 1 2 3
60:
70:
80:
90:
100:
```

Here the data is spread out over a fairly small area but the majority is in the late 30 and early 40 section there are a few anomalous results such as 23 and 53 (the spread therefore is 30).

Mean humidity: 39.82%

TEMPERATURE

```
0:
10:
20:      0.9 1 1.2 1.3 1.5 2.6 2.7 2.8 3 3.3 3.4 3.5 3.5 3.5 3.5 3.7 3.7 3.9 3.9 4 4 4  .1 4.2
         4.2 4.2 4.2 4.3 4.3 4.3 4.4 4.4 4.5 4.6 4.6 4.6 4.6 4.6 4.7 4.7 4.7 4.8 4.9 4.9 5 5
         5.1 5.2 5.2 5.3 5.3 5.3 5.5 5.5 5.5 5.7 5.8 5.9 5.9 6.1 6.3
30:
40:
50:
60:
70:
80:
90:
100:
```

The problem with this stem and leaf diagram is that all data falls under 1 group (20)
so I have redone but with every 1 instead of 10. I have done this because all the data is
so closely related.

```
20:      .9
21:      .0 .2 .3 .5
22:      .6 .7 .8
23:      .0 .3 .4 .5 .5 .5 .5 .7 .7 .9 .9
24:      .0 .0 .1 .2 .2 .2 .2 .3 .3 .3 .4 .4 .5 .6 .6 .6 .6 . 7 .7 .7 .8 .9 .9
25:      .0 .0 .1 .2 .2 .3 .3 .3 .5 .5 .5 .7 .8 .9 .9
26:      .1 .3
27:
28:
29:
```

As you can see this makes the data much more widely spread so I can point out
groups such as the 20.9 to 21.5 section which is along way off the next at 22.6.
Although there is quite a large gap between the first small group and the main group
the ret are pretty close together.

I have found out the mean for each section of my data.
Mean temperature: 24.22°C

HISTOGRAMS

I have placed histograms directly after my stem and leaf section as they are so closely related to each other.

HUMIDITY

1

20

Temperature frequency on the y axis.
Humidity on the x axis
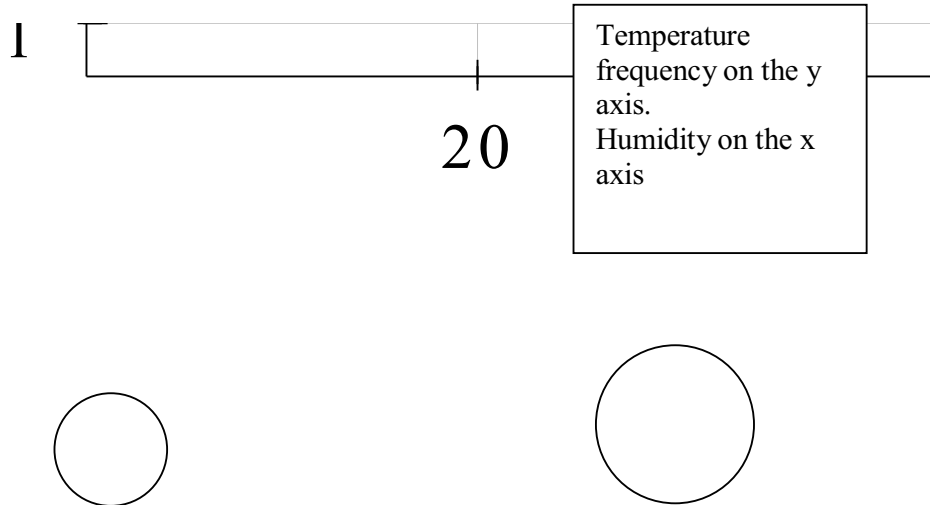
The bar chart above displays the data from the stem and leaf diagram on page 8. The items circled on the left and right of the graph represent the anomalous results shown in the stem leaf diagram on page 8.

TEMPERATURE

0

19

Humidity frequency on the y axis
Temperature on the x axis.
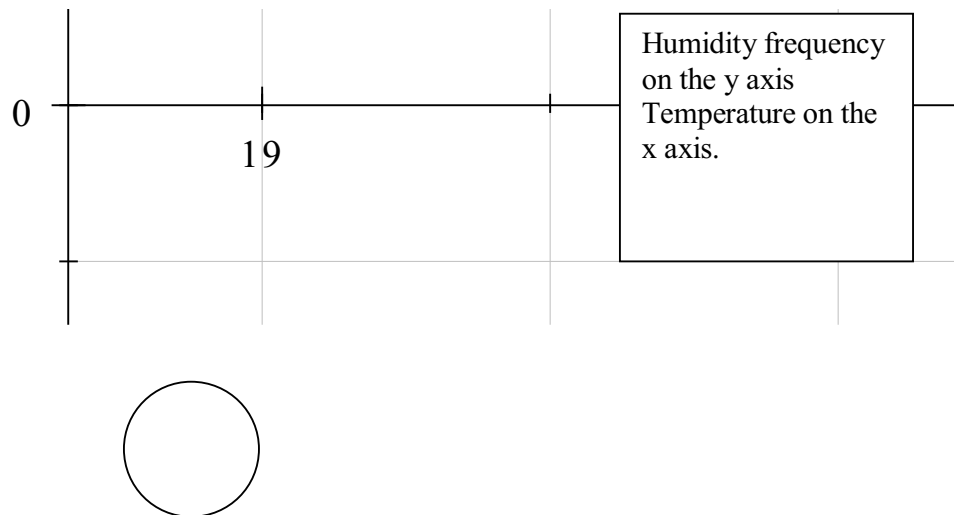
The bar chart above consists of 2 groups which I pointed out earlier in my stem and leaf diagram. The first group, which is obviously a group of anomalies is the one encircled.
The rest of the data seems to go up steadily and down steadily which indicates that the data in my stem and leaf diagram is accurate.

WHISKER AND BOX PLOT

In this graph I have decided to compare my humidity and my temperature to look at the difference of spread and average between the 2.
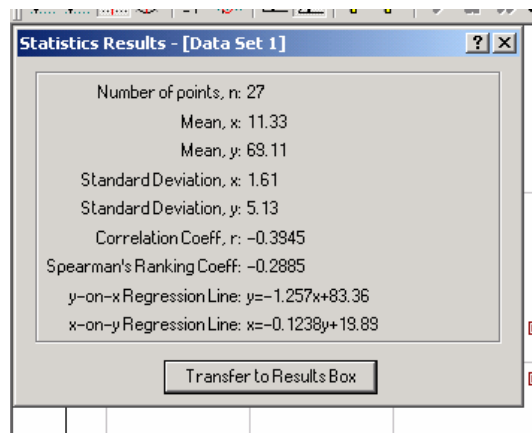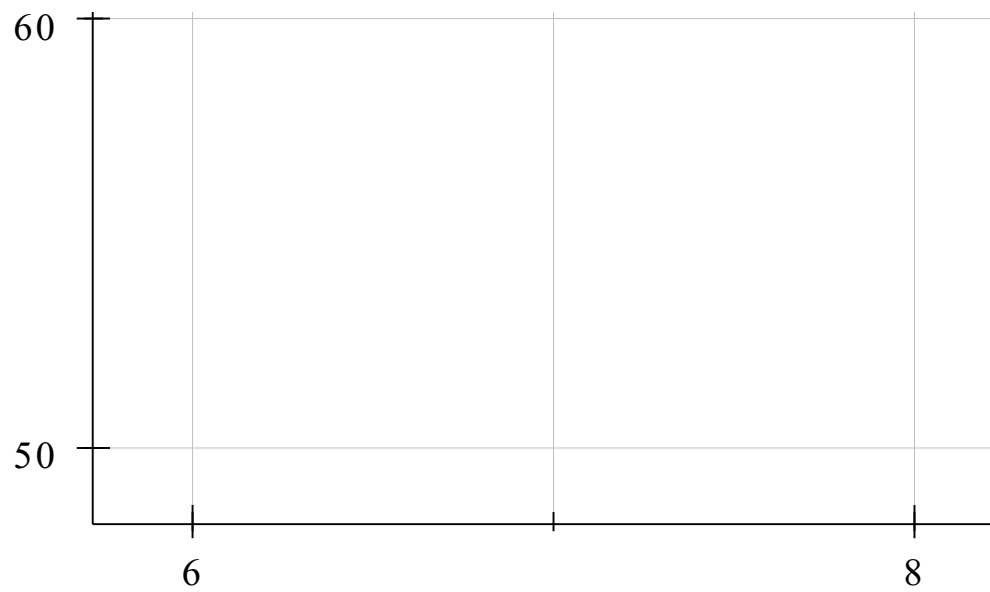
0

20

The spread of data is completely different and the averages are completely different, so I from this graph I would say there is no correlation between the 2 box and whisker plots, however I have already proved that there is a correlation between them in my scatter graphs on pages 5, 6 and 7.

SECONDARY DATA

I collected this data off the internet and it is based on points around England in December.

| | Secondary data | |
	Temperature	humidity
1	10	71
2	11	68
3	9	79
4	13	64
5	10	66
6	13	72
7	11	65
8	11	69
9	10	71
10	10	62
11	14	62
12	10	68
13	9	82
14	13	76
15	10	64
16	14	62
17	11	70
18	10	67
19	13	74
20	12	66
21	13	68
22	12	70
23	11	65
24	13	70
25	13	66
26	8	78
27	12	71

I did not collect as much data as in my primary experiment however there is enough data to make accurate graphs.

SCATTER GRAPH (secondary data)

60

50

6　　　　　　　　　　　8

Statistics Results - [Data Set 1]   ? X

Number of points, n: 27
Mean, x: 11.33
Mean, y: 69.11
Standard Deviation, x: 1.61
Standard Deviation, y: 5.13
Correlation Coeff, r: –0.3945
Spearman's Ranking Coeff: –0.2885
y-on-x Regression Line: y=–1.257x+83.36
x-on-y Regression Line: x=–0.1238y+19.89

Transfer to Results Box

The coefficient correlation of this graph is -0.3945.
This graph is of negative slant indicating that my hypothesis is correct.

STEM AND LEAF (secondary data)

HUMIDITY

```
0:
10:
20:
30:
40:
50:
60:    2 2 2 4 4 5 5 6 6 6 7 8 8 8 9
70:    0 0 0 1 1 1 2 4 6 8 9
80:    2
90:
100:
```

All the above data is well grouped.

TEMPERATURE

```
0:     8 9 9
10:    0 0 0 0 0 0 0 1 1 1 1 1 2 2 2 3 3 3 3 3 3 3 4 4
20:
30:
40:
50:
60:
70:
80:
90:
100:
```

This data is all within 6 degrees of each other so this data does need to be separated however it is hard to do as I only have 2 figures to work with.

COMPARING SPREAD OF DATA

DATA	HUMIDITY SPREAD	TEMPERATURE SPREAD
Primary	30	5.4
Secondary	20	6

I believe that the data spread has no correlation as my primary data is a much more localised resource; this means that the data is not averaged so it is much more precise.
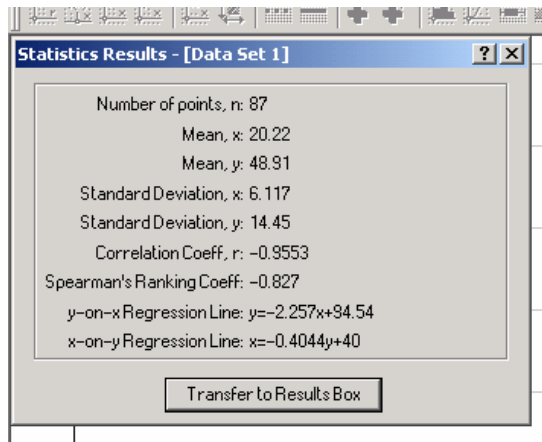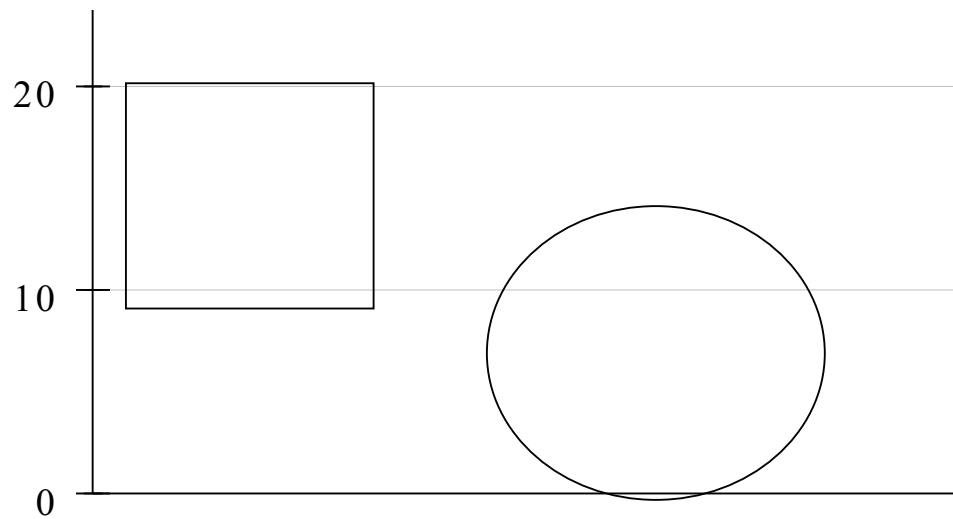
WHISKER AND BOX PLOT

I have decided not to do a whisker and box plot for my secondary data or indeed to compare the box plots with my earlier ones as they are based in different seasons, this would have no use at all except to find the spread of humidity and temperature which I have already worked out from the stem and leaf diagram.(as shown on page 14)

SCATTER GRAPH (combined data)

I have combined my primary and secondary data into one scatter graph. The hottest temperatures out of all the data have the lowest humidity and the lowest temperatures have the highest humidity.

Statistics Results - [Data Set 1]

Number of points, n: 87
Mean, x: 20.22
Mean, y: 48.91
Standard Deviation, x: 6.117
Standard Deviation, y: 14.45
Correlation Coeff, r: -0.9553
Spearman's Ranking Coeff: -0.827
y-on-x Regression Line: y=-2.257x+94.54
x-on-y Regression Line: x=-0.4044y+40

Transfer to Results Box

The primary data is encircled.
The secondary data is surrounded by a square.
The complete combined data when put together create a coefficient correlation of −0.9553.

The samples in this study are small (based in one country at only 2 separate days during the year) however there appears to be a close correlation between temperature and humidity.
If I was to study this further I would want take samples all around the globes at separate times during the year.
However from the data presented in this paper it would appear that the original hypothesis holds true.