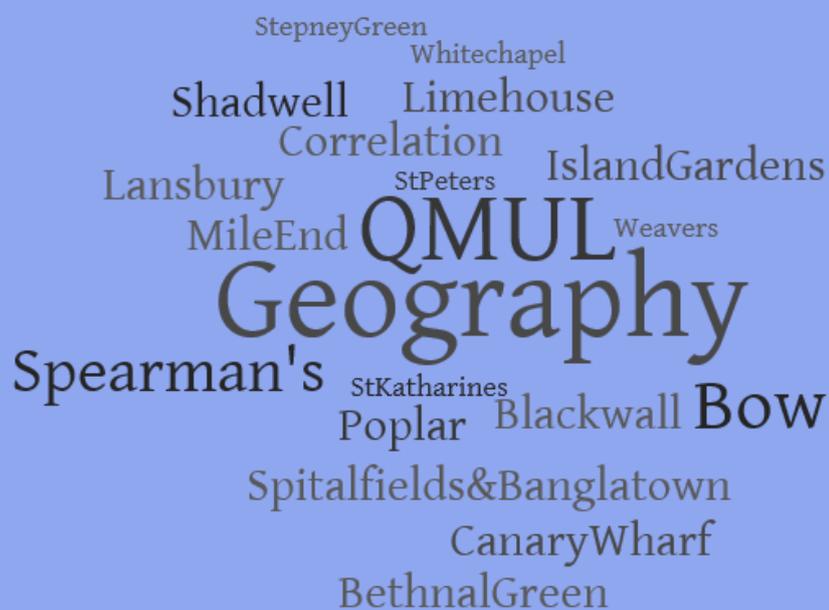

GEOGRAPHICAL TECHNIQUES



| | | | |
|---|-------------------------|---|----------------------|
| ✓ | Using quantitative data | | Using primary data |
| | Using qualitative data | ✓ | Using secondary data |

The Spearman's Rank Correlation Test



Introduction

The Spearman's rank correlation coefficient (r_s) is a method of testing the strength and direction (positive or negative) of the correlation (relationship or connection) between two variables.

As part of looking at *Changing Places* in human geography you could use data from the 2011 census online e.g. from <http://ukcensusdata.com> (and paper copies from local libraries for previous censuses in 2001, 1991 and 1981) to look at the relationships between pairs of data across an area such as a London borough. You can also use the site <https://data.london.gov.uk/dataset/ward-profiles-and-atlas> which is very easy to use.

Data on gender, age, health, housing, crime, ethnicity, education and employment is easily accessed at ward level. So, if you are looking to understand and analyse a) differences between nearby places and b) how those places have changed over time, a Spearman's Rank test could be a good way of testing the strength of your sets of data. This could help you to draw conclusions in your investigation.

You could first begin by plotting your data on a scatter graph to see if there appeared to a correlation between the two sets of data. This would help you to see quickly if the relationship was a positive or negative one. You would need to do a separate Spearman's rank correlation test for each set of pairs of data and for each time period. For example

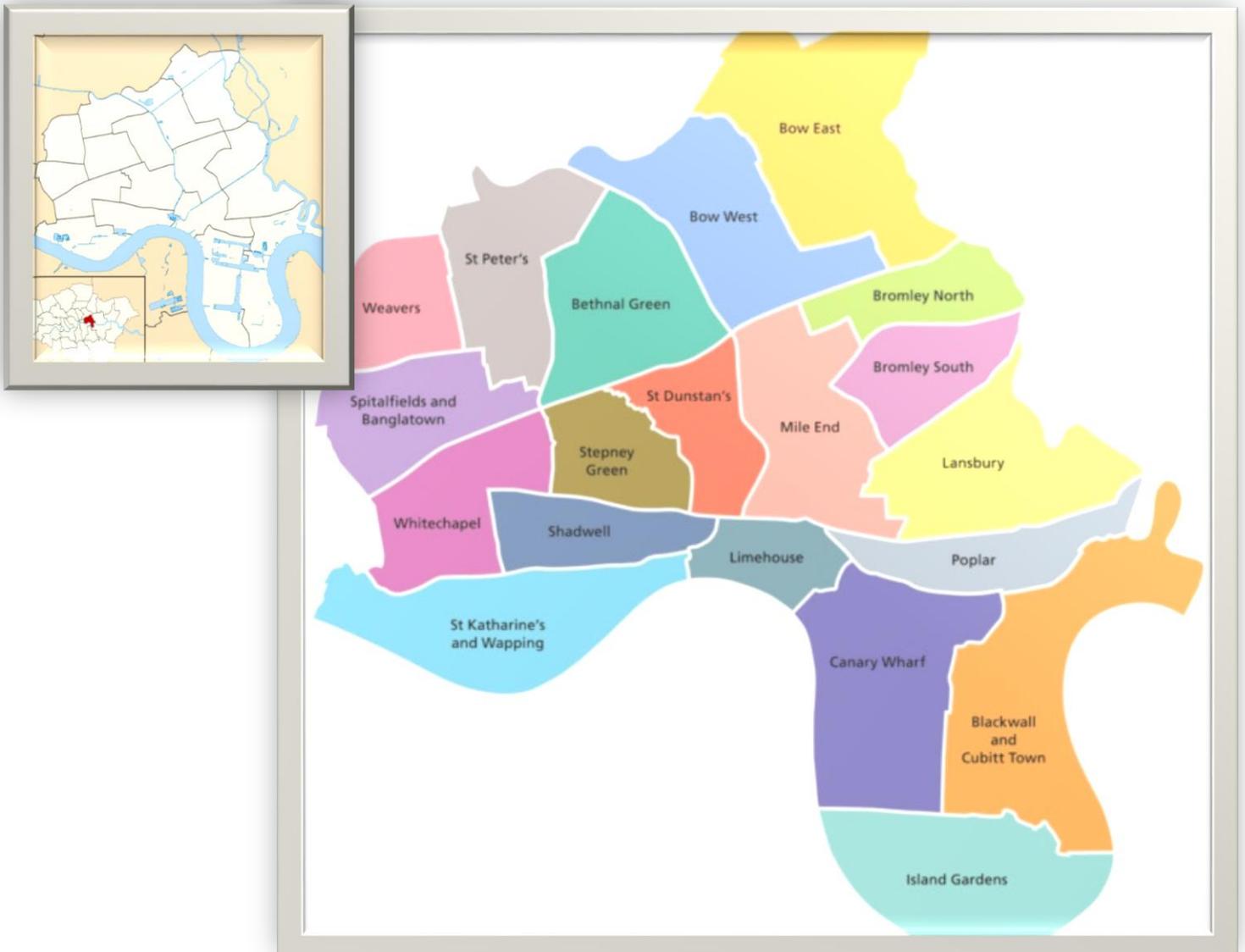
| Area of investigation | I could collect secondary data sets on | And correlate it with ... |
|--|--|--|
| I am interested in exploring how my place has changed in terms of education opportunities over time* | <ul style="list-style-type: none"> Ofsted grades for schools Exam results at age 11 or 16 Highest levels of qualifications % of population with a degree | <ul style="list-style-type: none"> Income data Index of multiple deprivation data Socio economic class data |
| I am interested in exploring how my place has changed in terms of health outcomes over time* | <ul style="list-style-type: none"> Disease rates smoking rates General health data | <ul style="list-style-type: none"> Life expectancy |
| I am interested in exploring how my area has changed in terms of employment over time * | <ul style="list-style-type: none"> Levels of unemployment | <ul style="list-style-type: none"> Highest levels of qualifications % of population with a degree Income data |
| * If you wanted to look at change from one time period to another you need to collect and compare data for two time periods. You would need to do a separate Spearman's rank test for each time period | | |

NB: You really need **at least 10 pairs of data** to be able to be sure that the end result is a robust or **significant** result. By using plenty of pairs of data, i.e. more than 10, you can be reasonably sure that the result you get comes from a true correlation and not just from chance. You can then be **confident** in your result.



A. Getting the data

The borough of Tower Hamlets is divided into 19 wards. See the map below.



This means that by using data for each ward in your test you will automatically have 19 pairs of data, so that is ideal for producing a correlation in which you can have confidence.

Data from the 2011 census is available online at <http://ukcensusdata.com> or <https://data.london.gov.uk/dataset/ward-profiles-and-atlas>

You will need to enter the name of the ward to see the full range of data available. Alternatively use the local authority website, in this case the Tower Hamlets council website, as their statisticians will have done some of the work for you pulling together the key statistics in user friendly formats called **ward profiles**.

See the ward profiles at

http://www.towerhamlets.gov.uk/lgnl/community_and_living/borough_statistics/borough_statistics.aspx.

B. Let's do a worked example together– Is there a correlation between rates of unemployment and rates of adults with no qualifications in Tower Hamlets?

Step one

I want to explore education and employment to see if there is a relationship between them. I could start by suggesting that there will be no relationship between employment and qualifications. This will be called my **Null Hypothesis**.

To begin my r_s test I collected data from the ward profiles. I collected two sets of data for each of the 19 wards. Rates of unemployment and rates of adults with no qualifications. To avoid working with decimal points I turned all the figures from /00 to /000. I did this so that I would only be dealing with whole numbers. For example 15.1% becomes 151/000

I have entered data for each ward which I got from the ward profile. Because I don't like working with decimal points, I have turned all the numbers into /000

I have entered all the wards here in alphabetical order

Add 2 columns here but leave them blank for now

| Ward name | / 000 adults with no qualifications | Rank R ₁ | Unemployment per 1000 of economically active population age 16- 64 | Rank R ₂ | | |
|-----------------------------|-------------------------------------|---------------------|--|---------------------|--|--|
| Bethnal Green | 151 | | 139 | | | |
| Blackwall | 96 | | 82 | | | |
| Bow East | 157 | | 102 | | | |
| Bow West | 151 | | 104 | | | |
| Bromley North | 193 | | 168 | | | |
| Bromley South | 218 | | 166 | | | |
| Canary Wharf | 89 | | 82 | | | |
| Island Gardens | 107 | | 84 | | | |
| Lansbury | 256 | | 188 | | | |
| Limehouse | 90 | | 66 | | | |
| Mile End | 190 | | 156 | | | |
| Poplar | 67 | | 157 | | | |
| Shadwell | 203 | | 160 | | | |
| Spitalfields and Banglatown | 148 | | 138 | | | |
| Stepney Green | 207 | | 151 | | | |
| St Katharines | 68 | | 49 | | | |
| St Peters | 165 | | 122 | | | |
| Weavers | 156 | | 110 | | | |
| Whitechapel | 118 | | 101 | | | |

Step two

I need to rank each set of data. I need to do this separately for each of the pairs of data. Look for the lowest values in a column set to start, then rank from lowest to highest for the entire column.

For the *adults with no qualifications* column below the lowest value is 67/000 for **Poplar ward**. This will be rank 1.

The highest value is 256/000 in **Lansbury ward** so this will be rank 19

| Ward name | / 000 adults with no qualifications | Rank R ₁ | Unemployment per 1000 of economically active population age 16- 64 | Rank R ₂ | | |
|-----------------------------|-------------------------------------|---------------------|--|---------------------|--|--|
| Bethnal Green | 151 | | 139 | | | |
| Blackwall | 96 | | 82 | | | |
| Bow East | 157 | | 102 | | | |
| Bow West | 151 | | 104 | | | |
| Bromley North | 193 | | 168 | | | |
| Bromley South | 218 | | 166 | | | |
| Canary Wharf | 89 | | 82 | | | |
| Island Gardens | 107 | | 84 | | | |
| Lansbury | 256 | 19 | 188 | | | |
| Limehouse | 90 | | 66 | | | |
| Mile End | 190 | | 156 | | | |
| Poplar | 67 | 1 | 157 | | | |
| Shadwell | 203 | | 160 | | | |
| Spitalfields and Banglatown | 148 | | 138 | | | |
| Stepney Green | 207 | | 151 | | | |
| St Katharines | 68 | | 49 | | | |
| St Peters | 165 | | 122 | | | |
| Weavers | 156 | | 110 | | | |
| Whitechapel | 118 | | 101 | | | |

Sometimes you might have **two or more values which are the same** (this is known as tied values) so how do you rank them? Well, you add together the ranked values of the positions they would cover e.g. 3rd, 4th, 5th and divide by the number of values (in this case 3) and allocate each the same rank score. So $3 + 4 + 5 = 12$. 12 divided by the 3 places they occupy is 4 so all three values are given **the same rank** of 4 on the list.

If you like using Excel you could put the data onto a spreadsheet and get Excel to rank the data for you into value order which is what I have done below here.



| Ward name | /000 adults with no qualifications | Rank R1 | Unemployment per 1000 of economically active | Rank R2 |
|------------------------------|------------------------------------|---------|--|---------|
| Bethnal Green | 151 | 9.5 | 139 | 12 |
| Blackwall | 96 | 5 | 82 | 3.5 |
| Bow East | 157 | 12 | 102 | 7 |
| Bow West | 151 | 9.5 | 104 | 8 |
| Bromley North | 193 | 15 | 168 | 18 |
| Bromley South | 218 | 18 | 166 | 17 |
| Canary Wharf | 89 | 3 | 82 | 3.5 |
| Island Gardens | 107 | 6 | 84 | 5 |
| Lansbury | 256 | 19 | 188 | 19 |
| Limehouse | 90 | 4 | 66 | 2 |
| Mile End | 190 | 14 | 156 | 14 |
| Poplar | 67 | 1 | 157 | 15 |
| Shadwell | 203 | 16 | 160 | 16 |
| Spitalfields and Bangalatown | 148 | 8 | 138 | 11 |
| St Katharines | 68 | 2 | 49 | 1 |
| St Peters | 165 | 13 | 122 | 10 |
| Stepney Green | 207 | 17 | 151 | 13 |
| Weavers | 156 | 11 | 110 | 9 |
| Whitechapel | 118 | 7 | 101 | 6 |
| TOTAL | | | | |

Step three

You have two empty columns to the right hand side of your table. We are now going to use those columns. Label the first of your two columns $d (R_1 - R_2)$. In this column, for each pair of data, subtract the rank r_2 from the rank r_1 to give d . **d stands for difference.**

In the table below I have filled in the first three rows for d .

Note that some figures will be minus (-) and some will be plus (+) figures.

| Ward name | /000 adults with a degree | Rank R_1 | / 000 adults with no qualifications | Unemployment per 1000 of economically active population age 16- 64 | Rank R_2 | $d (R_1 - R_2)$ | d_2 |
|---------------|---------------------------|------------|-------------------------------------|--|------------|-----------------|-------|
| Bethnal Green | 375 | 9.5 | 151 | 139 | 12 | -2.5 | |
| Blackwall | 569 | 5 | 96 | 82 | 3.5 | 1.5 | |
| Bow East | 454 | 12 | 157 | 102 | 7 | 5 | |

Step four

We will now complete the last column in the table. Simply square the value of d (To square something is to multiply a number by itself. For example 2 squared (2^2) is 4 and 12 squared (12^2) is 144).

We do this procedure to get rid of all the negative values (a minus value multiplied by a minus value is a plus value).

| Ward name | /000 adults with a degree | Rank R_1 | / 000 adults with no qualifications | Unemployment per 1000 of economically active population age 16- 64 | Rank R_2 | $d (R_1 - R_2)$ | d^2 |
|---------------|---------------------------|------------|-------------------------------------|--|------------|------------------|-------|
| Bethnal Green | 375 | 9.5 | 151 | 139 | 12 | -2.5 | 6.25 |
| Blackwall | 569 | 5 | 96 | 82 | 3.5 | 1.5 | 2.25 |
| Bow East | 454 | 12 | 157 | 102 | 7 | 5 | 25 |

Step five

You should now have a go at completing the whole table here to make sure you understand the process before you go on to step six.

| Ward name | / 000 adults with no qualifications | Rank R_1 | Unemployment per 1000 of economically active population age 16- 64 | Rank R_2 | $d (R_1 - R_2)$ | d^2 |
|-----------------------------|-------------------------------------|------------|--|------------|------------------|-------|
| Bethnal Green | 151 | 9.5 | 139 | 12 | -2.5 | 6.25 |
| Blackwall | 96 | 5 | 82 | 3.5 | 1.5 | 2.25 |
| Bow East | 157 | 12 | 102 | 7 | 5 | 25 |
| Bow West | 151 | 9.5 | 104 | 8 | | |
| Bromley North | 193 | 15 | 168 | 18 | | |
| Bromley South | 218 | 18 | 166 | 17 | | |
| Canary Wharf | 89 | 3 | 82 | 3.5 | | |
| Island Gardens | 107 | 6 | 84 | 5 | | |
| Lansbury | 256 | 19 | 188 | 19 | | |
| Limehouse | 90 | 4 | 66 | 2 | | |
| Mile End | 190 | 14 | 156 | 14 | | |
| Poplar | 67 | 1 | 157 | 15 | | |
| Shadwell | 203 | 16 | 160 | 16 | | |
| Spitalfields and Banglatown | 148 | 8 | 138 | 11 | | |
| St Katharines | 68 | 2 | 49 | 1 | | |
| St Peters | 165 | 13 | 122 | 10 | | |
| Stepney Green | 207 | 17 | 151 | 13 | | |
| Weavers | 156 | 11 | 110 | 9 | | |
| Whitechapel | 118 | 7 | 101 | 6 | | |
| TOTAL (Σ) | | | | | | |

Step six

This is what your table should now look like

| Ward name | / 000 adults with no qualifications | Rank R1 | Unemployment per 1000 of economically active population age 16- 64 | Rank R2 | d (R1-R2) | d ² |
|-----------------------------|-------------------------------------|---------|--|---------|---------------|----------------|
| Bethnal Green | 151 | 9.5 | 139 | 12 | -2.5 | 6.25 |
| Blackwall | 96 | 5 | 82 | 3.5 | 1.5 | 2.25 |
| Bow East | 157 | 12 | 102 | 7 | 5 | 25 |
| Bow West | 151 | 9.5 | 104 | 8 | 1.5 | 2.25 |
| Bromley North | 193 | 15 | 168 | 18 | -3 | 9 |
| Bromley South | 218 | 18 | 166 | 17 | 1 | 1 |
| Canary Wharf | 89 | 3 | 82 | 3.5 | -0.5 | 0.25 |
| Island Gardens | 107 | 6 | 84 | 5 | 1 | 1 |
| Lansbury | 256 | 19 | 188 | 19 | 0 | 0 |
| Limehouse | 90 | 4 | 66 | 2 | 2 | 4 |
| Mile End | 190 | 14 | 156 | 14 | 0 | 0 |
| Poplar | 67 | 1 | 157 | 15 | -14 | 196 |
| Shadwell | 203 | 16 | 160 | 16 | 0 | 0 |
| Spitalfields and Banglatown | 148 | 8 | 138 | 11 | -3 | 9 |
| St Katharines | 68 | 2 | 49 | 1 | -1 | 1 |
| St Peters | 165 | 13 | 122 | 10 | 3 | 9 |
| Stepney Green | 207 | 17 | 151 | 13 | 4 | 16 |
| Weavers | 156 | 11 | 110 | 9 | 2 | 4 |
| Whitechapel | 118 | 7 | 101 | 6 | 1 | 1 |
| TOTAL | | | | | | 287 |

Now apply the Spearman's rank equation using the figures in your table.

$$R = 1 - \frac{6 \sum d^2}{n(n^2 - n)}$$

N= number of pairs of data

Here is how this would be worked through:

$$R = 1 - \frac{6(287)}{19(19^2 - 19)} \rightarrow R = 1 - \frac{6(287)}{19(19^2 - 19)} \rightarrow R = 1 - \frac{6(287)}{19(19^2 - 19)}$$

$$R = 1 - 0.2650 \leftarrow R = 1 - \frac{1722}{6498} \leftarrow R = 1 - \frac{1722}{19(342)} \leftarrow R = 1 - \frac{1722}{19(361 - 19)}$$

$$R = 0.735$$



Step seven: Is the correlation a positive or negative one?

The value will always come out between -1.0 and +1.0. What does this mean?

Negative correlations are those between -1 and zero. Values of less than zero, i.e. those with a **minus value** show a negative correlation. This means that, as one value rises, the other value falls. An example of negative correlations in geography might be:

| | |
|---|---|
| Life expectancy increases as % of population smoking decreases |  |
| Number of plant species in an area next to a footpath decreases as amount of trampling increases |  |
| A perfect negative correlation would be -1. A figure near to zero (e.g. -0.3) would be considered to be weak. As a guide we hope for figures of between -0.7 and -1 to indicate a strong negative correlation | |

Positive correlations are those between +1 and zero. Values of more than zero, i.e. those with a plus value, show a positive correlation. This means that as one value rises the other value rises or as one value falls the other value falls. An example of positive correlations in geography might be:

| | |
|--|---|
| The larger the town the greater the number of retail outlets it has |  |
| The greater the % of sand in the soil the greater the rate of infiltration |  |
| The lower the GDP the lower the life expectancy |  |

In our case, for this worked example, as our R value is + 0.735 we can see that it is a **positive correlation** and that correlation is considered to be **strong**.

Step eight: Can we trust our result?

When you are using statistics you need to be very careful to check that you have a meaningful result and not one which is just down to chance.

There are two checks you should do:

1. Have you used more than 10 sets of data? If not your sample may not be representative.
2. Check to see if your final figure registers as at least good enough to be **confident** of, on the **significance table**. Look for the column shown as 95% confidence (or sometimes listed as 0.5 level or 5%) If your result comes out as higher than the number shown for the number of data sets used then you can be confident the result is not down to random chance. 95% is a pretty high rate of confidence, so that is good enough.

Our result of 0.735 is above the figures shown for 18 sets of data at the 95% confidence level (0.475) so we can trust that this result is significant. We can be confident it is not down to chance.

| Number of pairs of data (n) | 10% chance 90% confident | 5% chance 95% confident | 2% chance 98% confident | 1% chance 99% confident |
|-----------------------------|-----------------------------|----------------------------|----------------------------|----------------------------|
| 5 | 0.9 | 1 | 1 | 1 |
| 6 | 0.829 | 0.886 | 0.943 | 1 |
| 7 | 0.714 | 0.786 | 0.893 | 0.929 |
| 8 | 0.643 | 0.738 | 0.833 | 0.881 |
| 9 | 0.6 | 0.683 | 0.783 | 0.833 |
| 10 | 0.564 | 0.648 | 0.746 | 0.794 |
| 12 | 0.506 | 0.591 | 0.712 | 0.777 |
| 14 | 0.456 | 0.544 | 0.645 | 0.715 |
| 16 | 0.425 | 0.506 | 0.601 | 0.665 |
| 18 | 0.399 | 0.475 | 0.564 | 0.625 |
| 20 | 0.377 | 0.45 | 0.534 | 0.591 |
| 22 | 0.359 | 0.428 | 0.508 | 0.562 |
| 24 | 0.343 | 0.409 | 0.485 | 0.537 |
| 26 | 0.329 | 0.392 | 0.465 | 0.515 |
| 28 | 0.317 | 0.377 | 0.448 | 0.496 |
| 30 | 0.306 | 0.364 | 0.432 | 0.478 |

There are lots of significance tables available online but this one here is courtesy of the FSC website at <http://www.field-studies-council.org>.

Step nine: So what does our result tell us?

Our result proves to be a strong correlation, upheld by a significance level of 95%. We have shown that in 2011 in Tower Hamlets there is a strong correlation in the borough between rates of unemployment and rates of lack of qualification. Places with high rates of adults with no qualifications have high rates of unemployment and places with low rates of lack of qualification have low rates of unemployment. However **correlation is not the same thing as cause**. Just because there is a relationship between two sets of variables **it does not mean that one thing causes the other**. The picture is likely to be more complicated. It is interesting to see that though the correlation at +0.7 counts as 'strong' it is only just over the 0.7 level so we are only just counting this as a strong correlation.

Are there **anomalies**? Which wards show the biggest difference in their rankings between the two columns and what could this tell us?

What **other indicators** of education might help us look further at the correlation between unemployment and education?

How have things changed over time? We used the figures for 2011. You could use the figures for 1981, for example, to see if lack of qualifications correlated with unemployment 40 years ago too.

STUDENT TASK:

- Carry out the Spearman's Rank Coefficient test using the figures provided below.
- Explain what the test shows.
- Explain what this result, together with the worked example above, tell us about education and employment in Tower Hamlets in 2011.

| Ward name | /000 adults with a degree | Rank R1 | Unemployment per 1000 of economically active population age 16-64 | Rank R2 | d (R1-R2) | d ² |
|---------------------------|---------------------------|---------|---|---------|-----------|----------------|
| Bethnal Green | 375 | | 139 | | | |
| Blackwall | 569 | | 82 | | | |
| Bow East | 454 | | 102 | | | |
| Bow West | 467 | | 104 | | | |
| Bromley North | 329 | | 168 | | | |
| Bromley South | 299 | | 166 | | | |
| Canary Wharf | 601 | | 82 | | | |
| Island Gardens | 545 | | 84 | | | |
| Lansbury | 251 | | 188 | | | |
| Limehouse | 614 | | 66 | | | |
| Mile End | 499 | | 156 | | | |
| Poplar | 322 | | 157 | | | |
| Shadwell | 312 | | 160 | | | |
| Spitalfields & Banglatown | 449 | | 138 | | | |
| St Katharines | 659 | | 49 | | | |
| St Peters | 431 | | 122 | | | |
| Stepney Green | 313 | | 151 | | | |
| Weavers | 469 | | 110 | | | |
| Whitechapel | 333 | | 101 | | | |

For further resources and CPD opportunities see our webpages at <http://www.geog.qmul.ac.uk/teachers> or contact the Schools Liaison Officer in the School of Geography 020 7882 2614 or k.amis@qmul.ac.uk

