

The following is a brief summary of an example I use with students of Geography and also students of Applied Statistics which can be interpreted as radical in two ways. Firstly they are an example of the use of John Tukey's ideas from his book Exploratory Data Analysis which is one of the most original books to appear in recent years. Unfortunately the presentation in the book leaves much to be desired (F Mosteller and J Tukey, Data Analysis and Regression is more readable) but underneath the terminology he generates - stem leaf, hinges, box and whisker plots, median polishing - there are many useful ideas. His departure from the mean back to the good old median is in itself an unorthodox approach. These techniques of data analysis have been incorporated in our syllabuses here in the Data Analysis course we teach alongside the Probability and Modelling courses in the first year.

Tukey himself, an advisor to successive US administrations and a professor at Princeton University could hardly be described as politically radical but it is in the political sense of the word that I am more interested. Our students of statistics are treated to a well mixed concoction of theory with applications in Industrial Statistics and operations research and to a lesser extent social statistics. They are too often unaware, even by the end of their course, of the conservative nature of the subject, of the reactionary role of statistics, of issues of privacy and confidentiality and the role of statistics in control of the individual. And they don't see enough examples of statistics in pointing to inequalities in society, in health, in housing or in education.

Part of the reason for this is that statisticians in teaching are not themselves in sympathy with, for want of a better description, radical statistics. But those of us who are in sympathy are short of good examples which illustrate statistical techniques and applications and also illustrate these other issues as well. It would be useful to co-operate with others in producing materials for teaching in higher education which have these things in mind. I hesitate to suggest the reformation of the RADSTATS TEACHING GROUP but would be very interested in correspondence and exchange of materials with like minded individuals, with a view to producing case studies. Please write to Dave Drew, Department of Mathematics, Statistics & Operational Research, Sheffield City Polytechnic, Pond Street, Sheffield, S1 1WB.

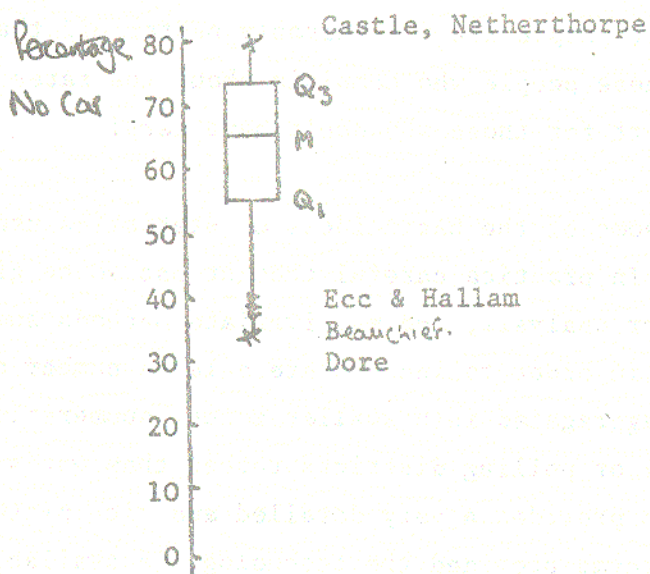
Example taken from a social area analysis of Sheffield

The following data is in the form X(a,b) where X is the Sheffield ward, a is the percentage of households with no inside WC and b is the percentage of households with no car.

| | | | | | |
|-------------|------|------|---------------|------|------|
| Attercliffe | 79.2 | 74.6 | Heeley | 44.4 | 65.2 |
| Beauchief | 6.3 | 37.8 | Hillsborough | 22.7 | 54.7 |
| Birley | 1.8 | 50.6 | Intake | 4.2 | 55.8 |
| Brightside | 26.2 | 64.5 | Manor | 6.6 | 73.6 |
| Broomhill | 17.2 | 54.6 | Mosborough | 21.2 | 57.4 |
| Burngreave | 45.9 | 72.0 | Nether Edge | 17.7 | 58.4 |
| Castle | 9.5 | 80.3 | Nethershire | 4.9 | 64.6 |
| Darnall | 53.6 | 65.9 | Netherthorpe | 27.4 | 79.0 |
| Dore | 2.6 | 35.1 | Owlerton | 23.7 | 67.2 |
| Ecclesall | 11.1 | 40.3 | Park | 10.1 | 74.0 |
| Firth Park | 16.8 | 71.0 | Sharrow | 58.7 | 73.3 |
| Gleadless | 2.1 | 58.0 | Southey Green | 2.0 | 72.6 |
| Hallam | 9.6 | 40.4 | Walkley | 56.4 | 65.5 |
| Handsworth | 14.2 | 52.0 | | | |

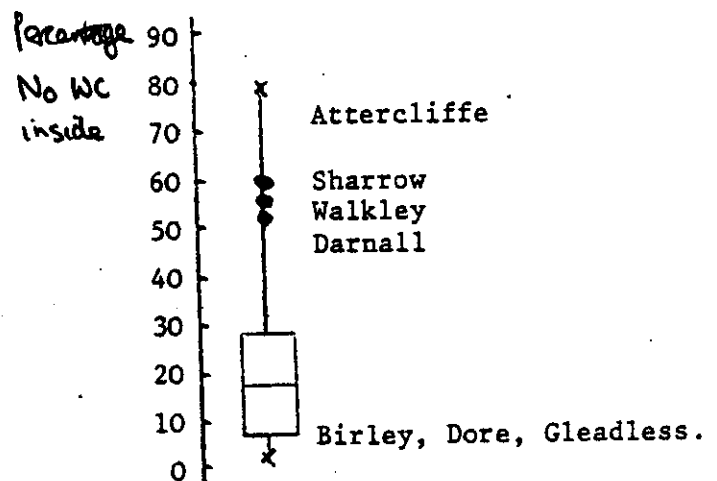
Source: 1971 Census : Sheffield Ward Data

Each variable can be represented as a stem and leaf display, a 5 value summary and a box and whisker plot (see below)



Car ownership in Sheffield is lower than in the rest of the country. The areas which are better off are Dore, Ecclesall, Hallam and Beauchief, areas on the south west side of the city, bordering the Peak District. These areas are predominantly owner occupied. In Castle and Netherthorpe, by contrast, car ownership is low and these are areas predominantly of local authority housing.

The percentage of households with no inside WC is higher for Sheffield than nationally. The distribution is highly positively skewed with Attercliffe clearly the worst area with 79% with no inside WC. This is a very high figure! This area has been largely cleared since 1971, the year when this data was collected. Sharrow, Walkley and Darnall are the other areas with large numbers of unimproved houses. These are the decaying inner city areas where there is a high percentage of private rented unfurnished accommodation.



The percentage of households with no car can be used as a proxy for income (no income data is collected in the British Census). Building council houses helps the least well off, providing decent living conditions compared with renting privately (all council houses have inside toilets) but car ownership shows that those living in council houses are amongst the least well off. Sheffield has a few areas like Dore and Ecclesall in which people are considerably better off than average. It is usually these people who complain about the rates to pay for the public transport for those who don't have cars!

This example illustrates some of the basic ideas of statistics used in social area analysis. In practice careful thought has to be given to the variables chosen for analysis, and multivariate methods such as cluster analysis are used in order to incorporate a large number of variables. Areas are disaggregated into smaller zones, enumeration districts (500 population) or polling districts rather than wards (20,000 population). This provides a very detailed and rich picture of an area in socioeconomic terms provided the technology is available ie. computer and graphics facilities, to handle such large volumes of data.

Five Value Summary

| | | No Car | No inside WC |
|--------------------|----------------|--------|--------------|
| Extreme | * | 35 | 2 |
| Lower Quartile | Q ₁ | 55 | 6 |
| Median | M | 65 | 17 |
| Upper Quartile | Q ₃ | 73 | 27 |
| Extreme | * | 80 | 79 |
| Mean | | 61.4 | 22.1 |
| Standard deviation | | 12.7 | 20.8 |
| National Mean | | 47.3 | 13.0 |

For the box and whisker plot the box is bounded by the quartiles and split by the median and the whiskers go out to the extremes.