# Identifying Patterns and Outliers in Census Data

Prof Chris Brunsdon
University of Glamorgan

## Overview

- Townsend Scores
- Weighting Schemes
- Visualisation
- Examples
- Conclusion



## 'Townsend Scoring'

- Four deprivation indicators
  - Non-Owner Occupation
  - Crowding
  - Unemployment
  - Carless Households
- Intended to be indicators of *material* deprivation
- Townsend, P., Phillimore, P. and Beattie, A. (1988) *Health and Deprivation: Inequality and the North*. Routledge, London.



## The Scoring Procedure

- Transform variables
  - Log(x+1) for all except no home ownership - sqrt(x) in that case
  - Take z-scores
  - Add them together
- Implicitly each variable is weighted equally
- Is that sensible?

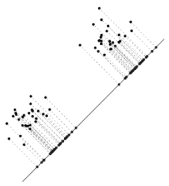


## Alternative views on Weighting

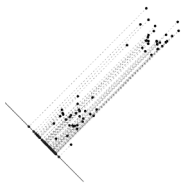
- Subjective - survey based - ie Jarman Index
- Based on interviews with London-based GPs
- Based on predicting some other variable
  - Weights are regression coefficients
  - Very good for predicting specific variables
  - Only really useful if predictor variable not readily available



## Alternative view - the geometrical projection



## Different projections tell different stories…





## Projection Pursuit

- Clearly some projections are more telling than others
- Mathematically, projections are equivalent to weighting schemes for deprivation indicators
- Choosing a 'good' weighting scheme equivalent to choosing an interesting projection



## So what is 'interesting'?

- According to Friedman, non-normality
- Bell Curves are not interesting features compared to bi- or multi-model distributions, or even non-symmetrical distributions

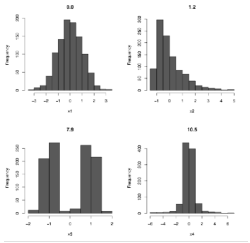


## Measuring 'Non Normality'

Friedman's Index

Higher values Suggest more non-normality

Choose a projection that maximises the index…





## One Final Technicality

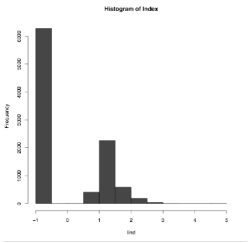
- Some projections are equivalent to negative weights
- In some situations that is fine, but
- For an index of deprivation it isn't sensible - it might imply an increase in unemployment gives a decrease in deprivation!
- Therefore we add the constraint that all weights must be positive



## Result of Projection Pursuit

| Variable | Weighting |
|---|---|
| Unemployment | 0.17 |
| No Car Households | 0.12 |
| Non home ownership | 0.00 |
| Crowding | 1.43 |

## Distribution of Index





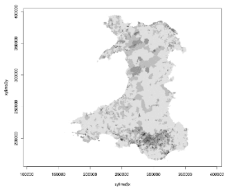

## Is non-normality the only thing to look for?

- As a geographer, spatial patterns are important
- Could seek to optimise the spatial autocorrelation
  - 'Spatial Autocorrelation' - a measure of how similar nearby places are. High levels indicate that areas with similar degrees of deprivation tend to group together geographically.



## Maximising Autocorrelation

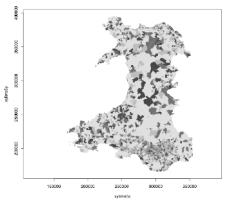
| Variable | Weight |
|---|---|
| Unemployment | 0.10 |
| No Car Households | 1.25 |
| No Home Ownership | 0.00 |
| Crowding | 0.15 |





## Minimising Autocorrelation

| Variable | Weight |
|---|---|
| Unemployment | 0.00 |
| No Car Households | 0.00 |
| No Home Ownership | 0.00 |
| Crowding | 1.47 |





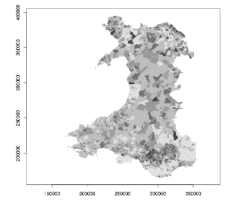
## Another Example - Educational Attainment

- None
- Level 1
  1+ 'O' level passes, 1+ CSE/GCSE any grades, NVQ level 1, Foundation GNVQ.
- Level 2
  5+ 'O' level passes, 5+ CSEs (grade 1), 5+ GCSEs (grades A-C), School Certificate, 1+ 'A' levels/'AS' levels, NVQ level 2, Intermediate GNVQ.
- Level 3
  2+ 'A' levels, 4+ 'AS' levels, Higher School Certificate, NVQ level 3, Advanced GNVQ.
- Level 4/5
  First degree, Higher degree, NVQ levels 4 and 5, HNC, HND, Qualified Teacher Status, Qualified Medical Doctor, Qualified Dentist, Qualified Nurse, Midwife, Heath Visitor.

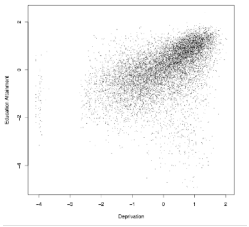


## Results

| Attainment | Weight |
|---|---|
| None | 0.75 |
| Level 1 | 0.82 |
| Level 2 | 0.65 |
| Level 3 | 0.00 |
| Level 4/5 | 0.00 |



## Relationship between the two indices






## Conclusions

- Useful to look at more than one projection
  - Multivariate relationships are important
- Useful to look at several scales
  - Eg OAs, Wards etc
- Visualisation an important tool - look at maps and histograms