Geographical Spatial Analysis in Personal Lines Territorial Ratemaking



WWW.WATSONWYATT.COM/PRETIUM

2003 CAS Ratemaking Seminar

Duncan Anderson MA FIA

Watson Wyatt LLP

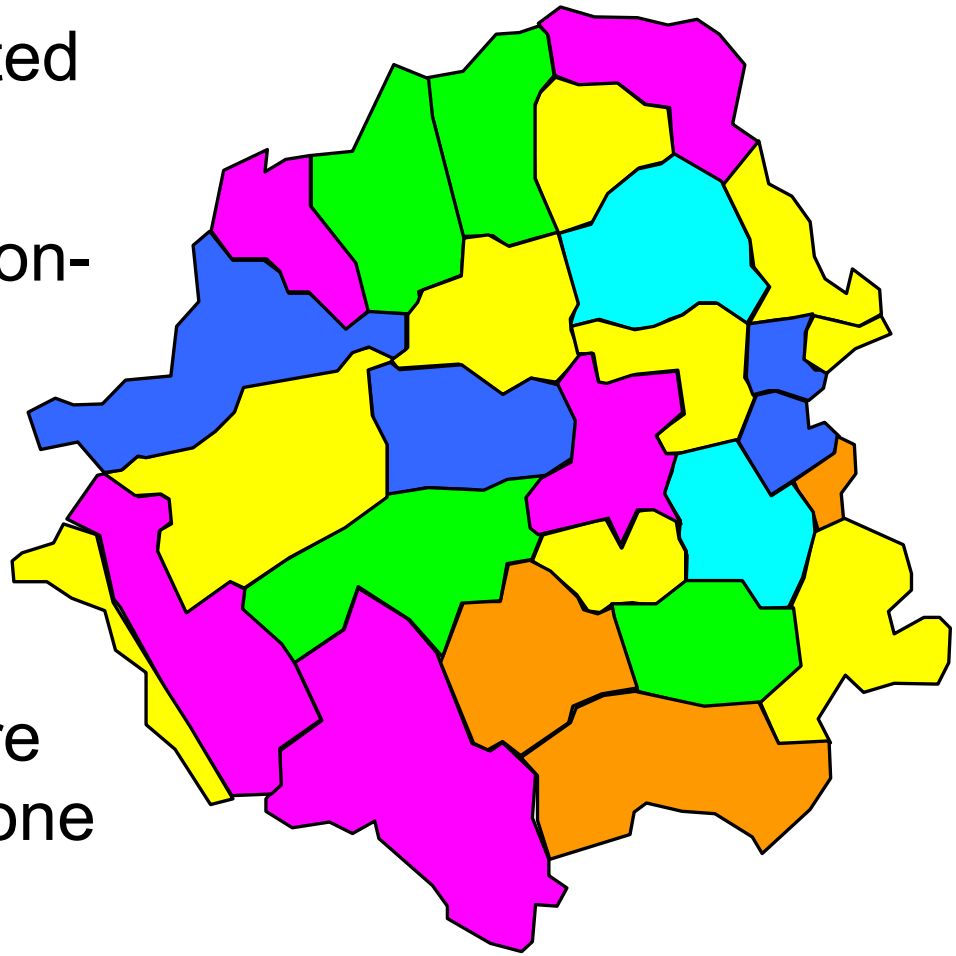
Watson Wyatt *Worldwide*

Geographical zoning

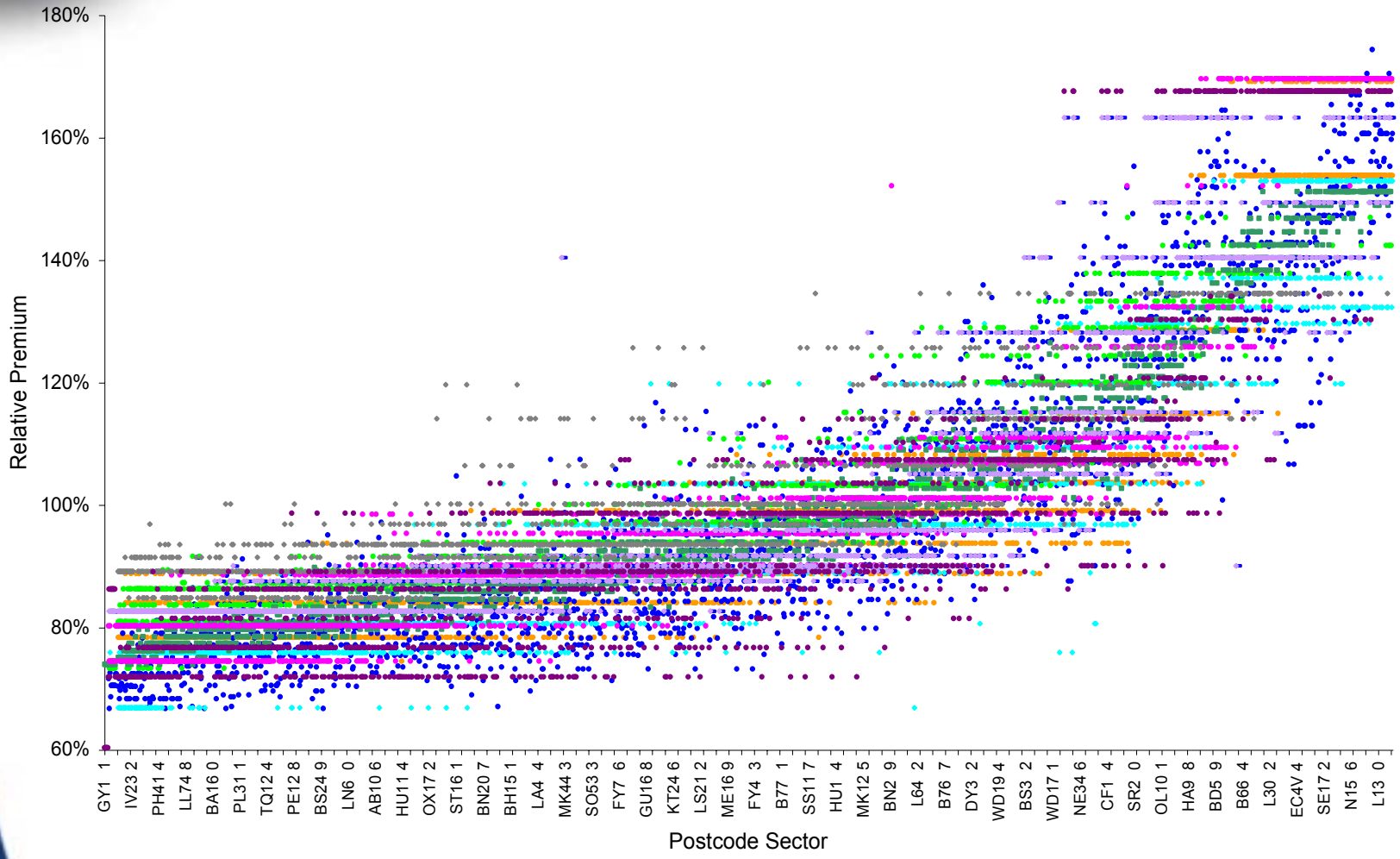
- Area is one of the main drivers of cost
- Many markets show considerable variety between insurers
- One insurer will have limited exposure in any one narrowly-defined area (eg zip code)

Geographical "zones"

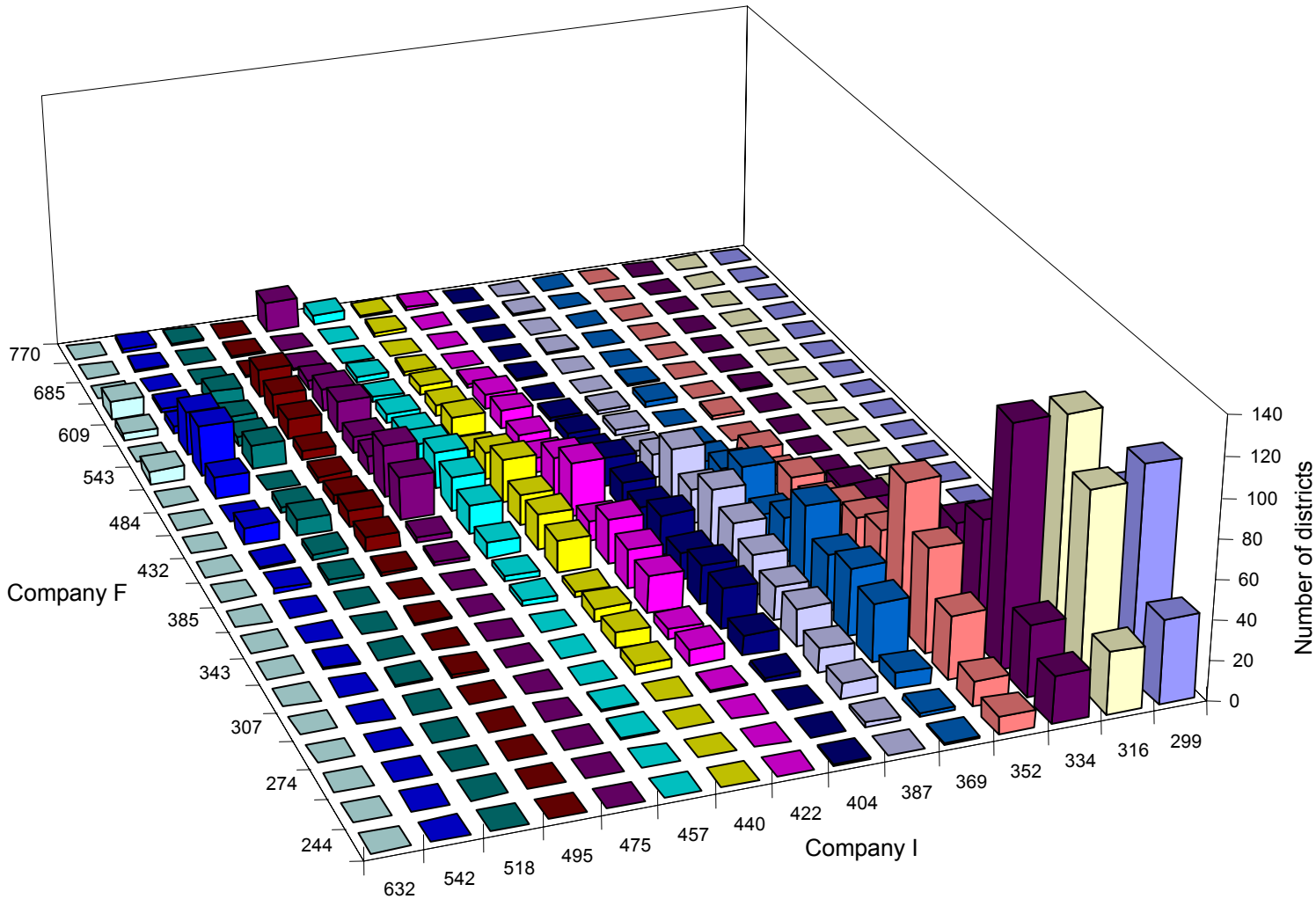
- Each zip code allocated to a zone
- Zones may contain non-contiguous zip codes
- Each zone given a premium loading
- In UK auto & homeowners there are typically around 20 zone categories, although sometimes 100+



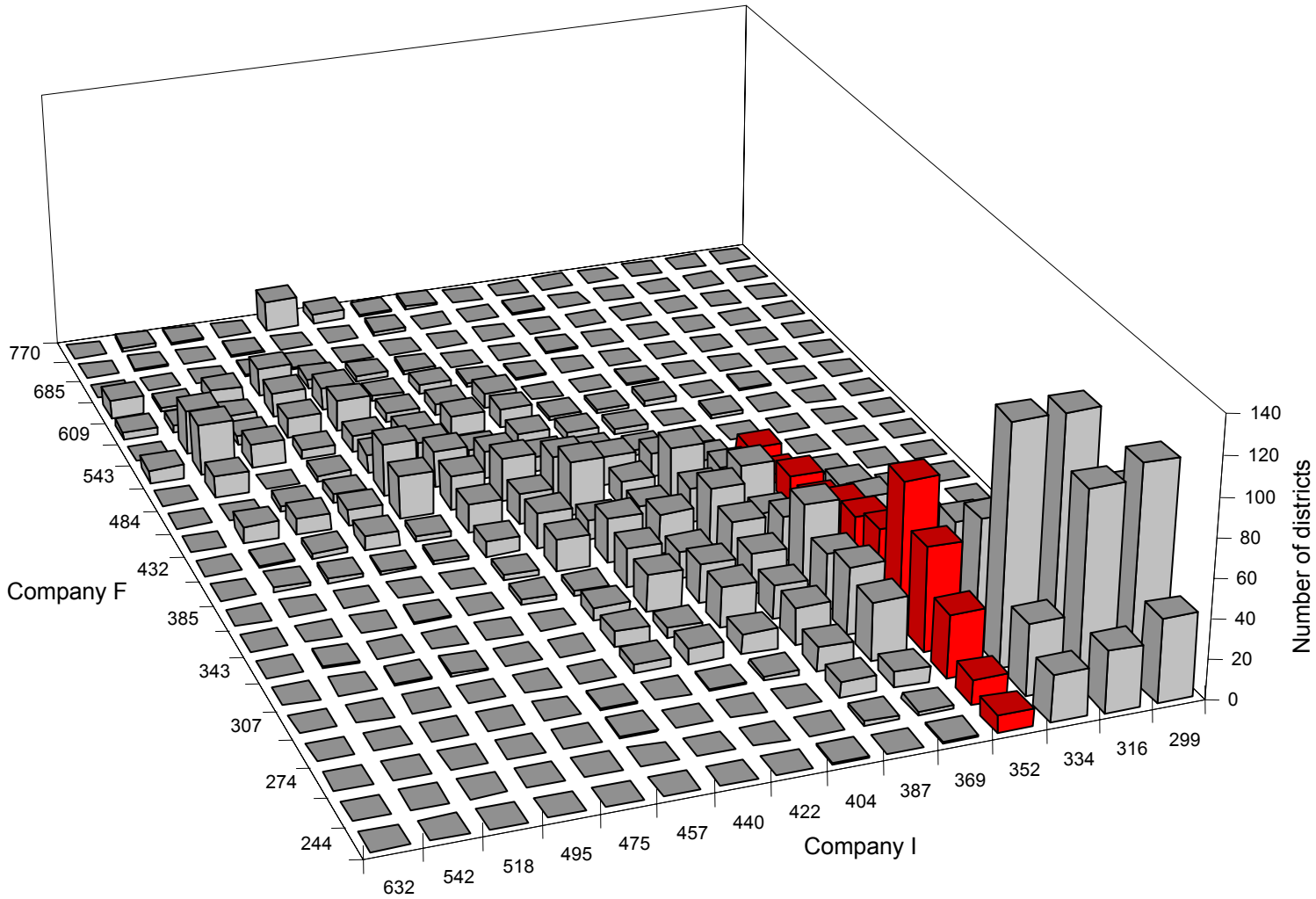
UK auto postcode loadings



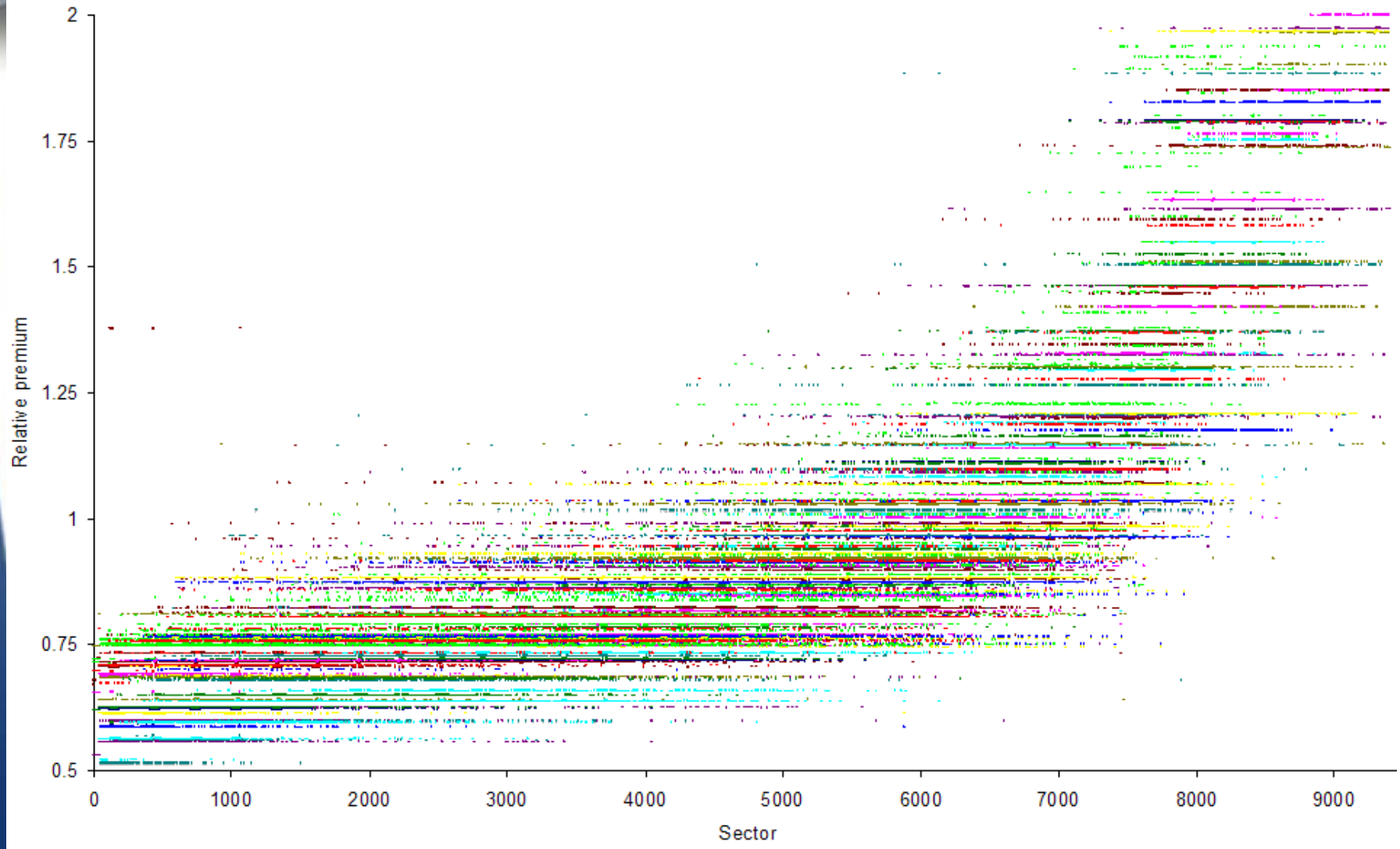
Comparison of UK auto categorisations



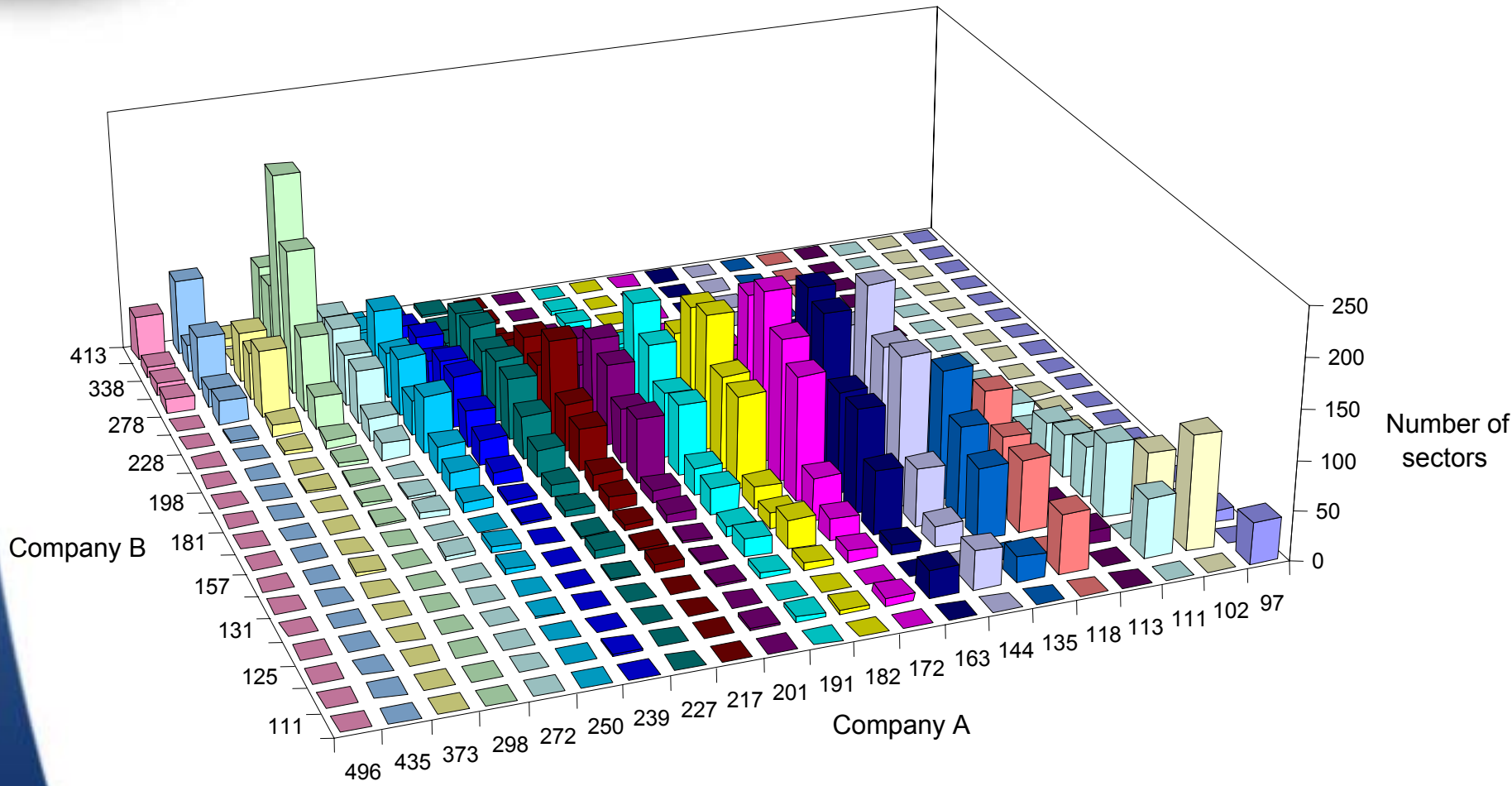
Comparison of UK auto categorisations



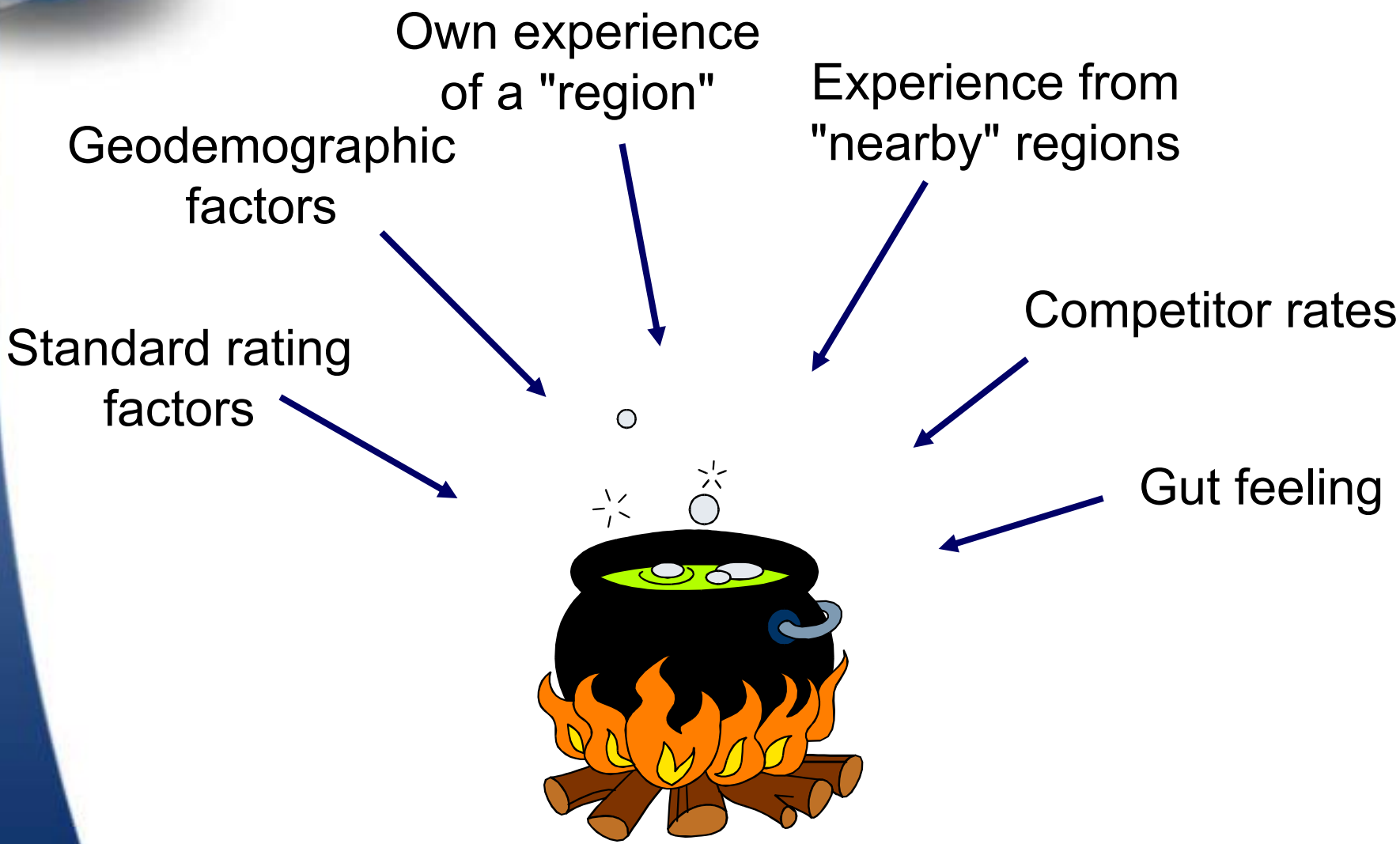
UK homeowners (contents) postcode loadings



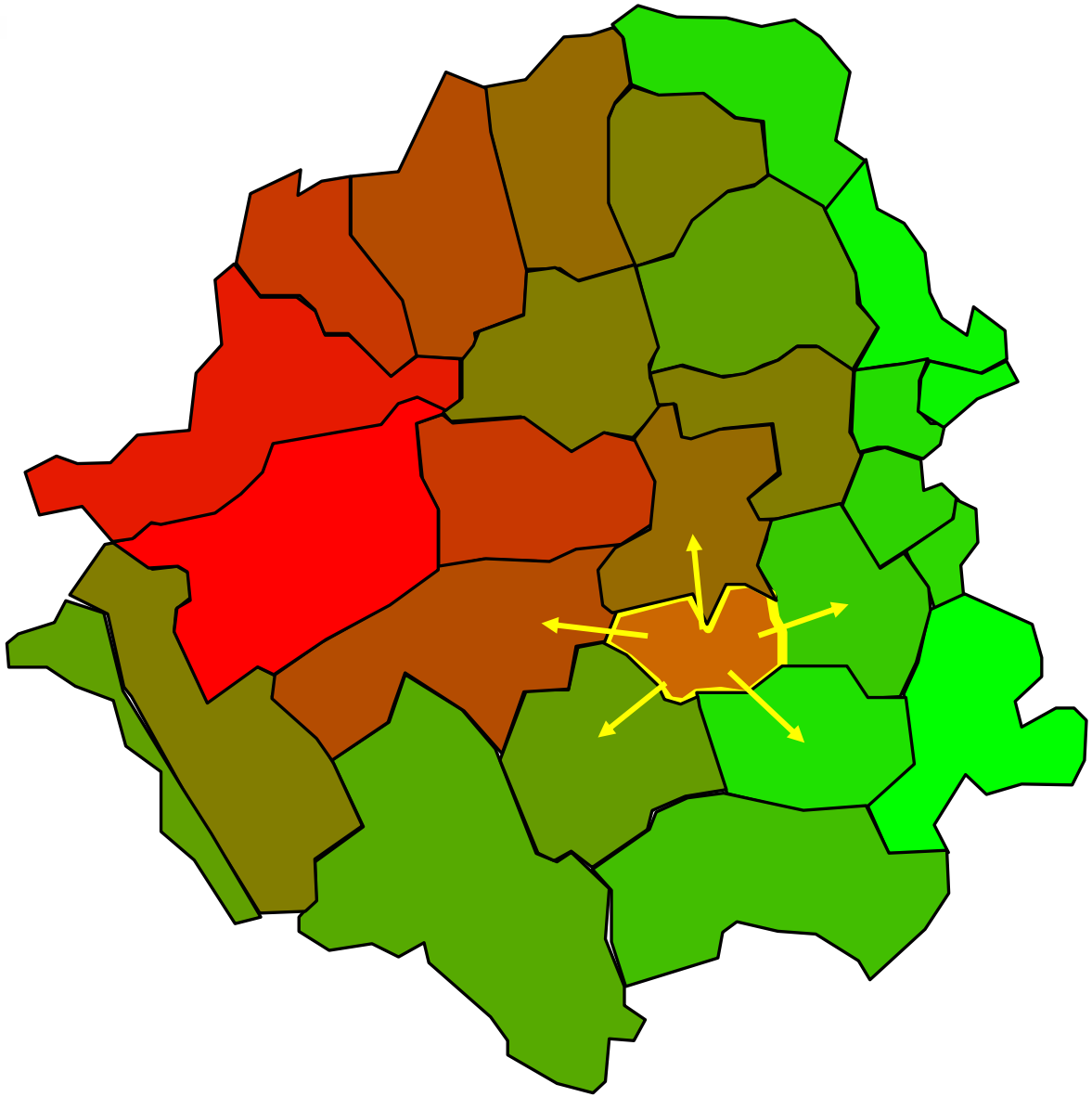
Comparison of UK homeowners contents zones



Ingredients for a solution

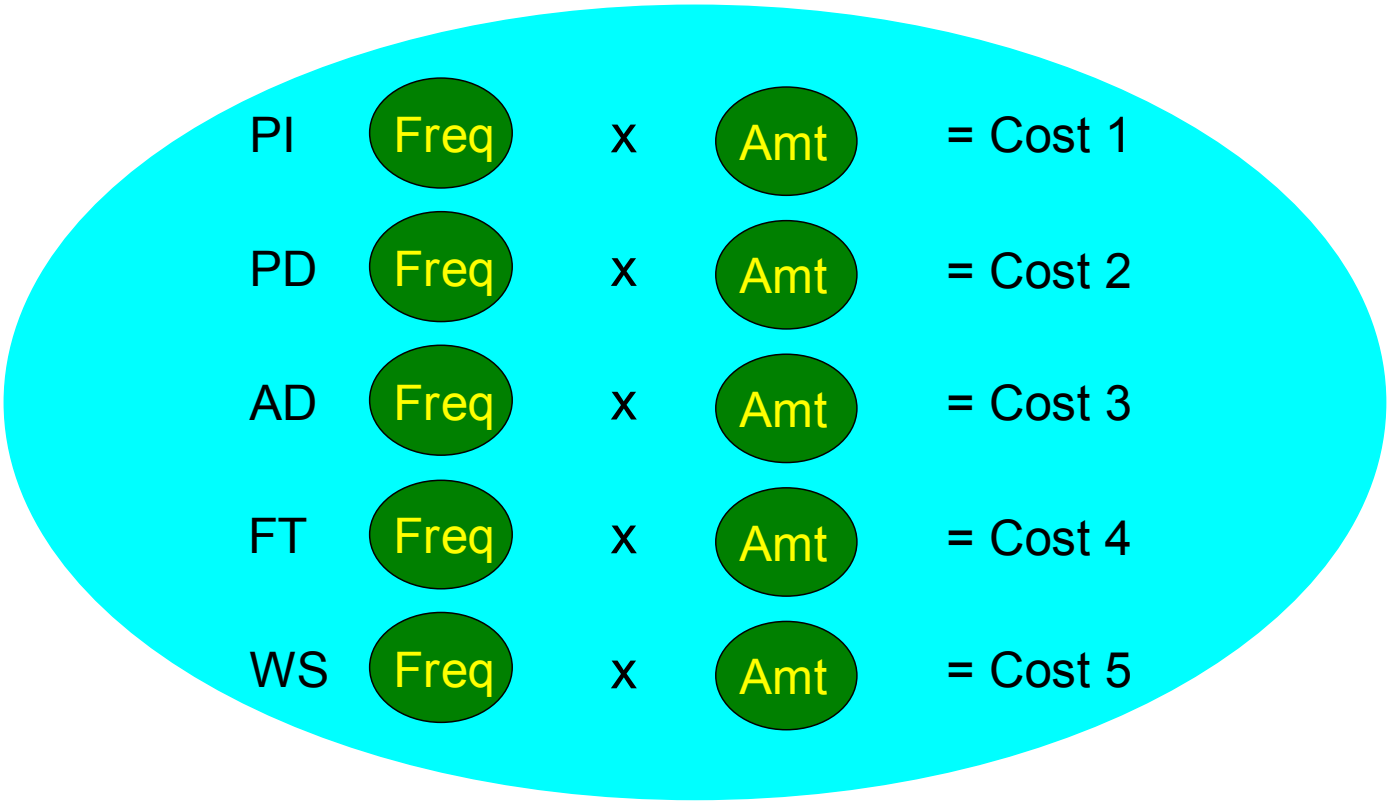


Proximity

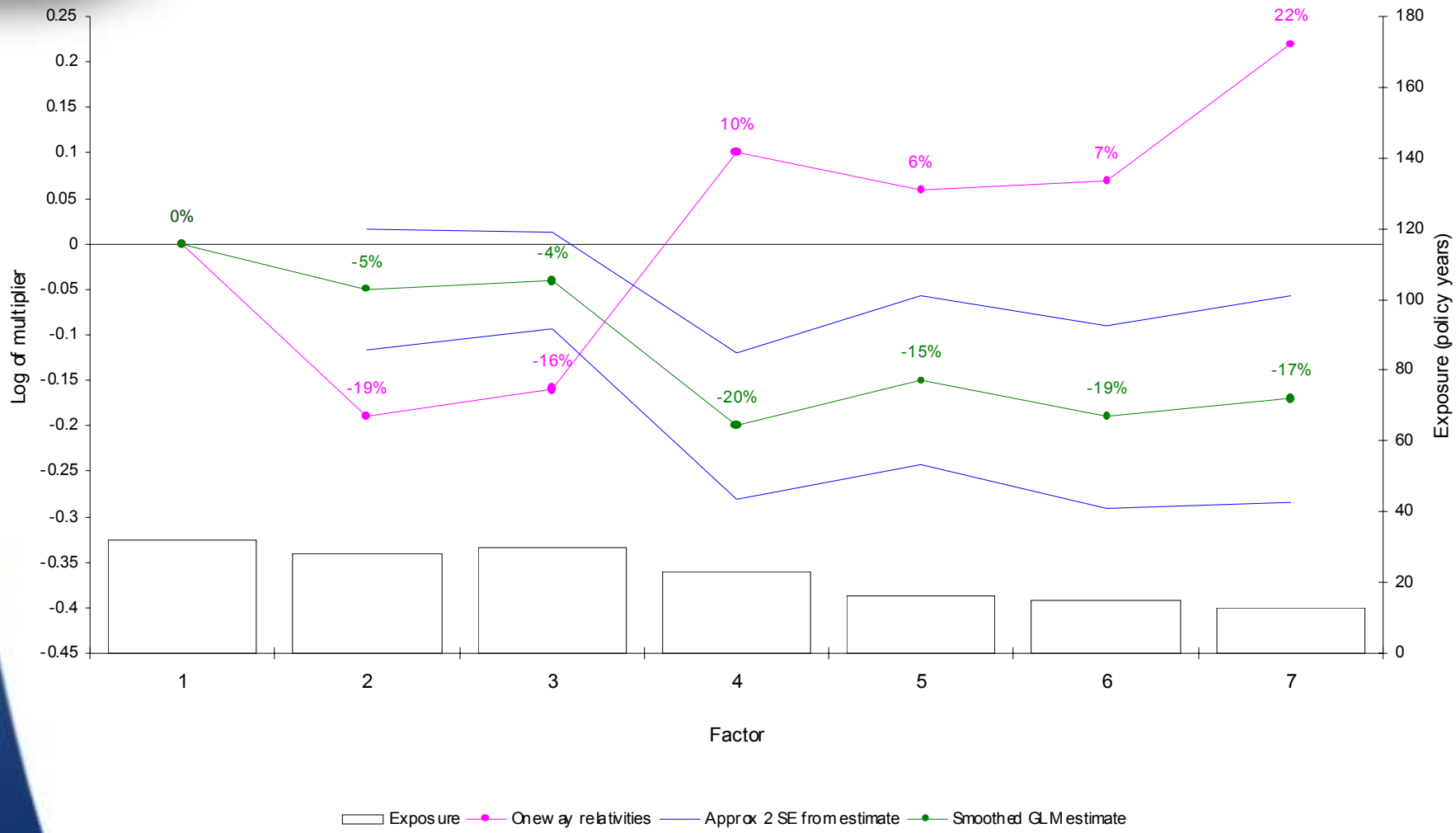


The general approach

- Select which element of experience to model



Generalized linear models



The general approach

- Do not wish to attribute to any region experience which can be explained by other rating factors
- Standardise for other factors by fitting a GLM (excluding current zones)
- Consider "residual" risk by "region"
- Seek to make this residual risk more predictive
- Then categorise into zones to derive appropriate loadings

A model form

$r_i^* = Z.r_i + (1 - Z) .$ neighboring experience

where

r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

A model form

$r_i^* = Z.r_i + (1 - Z) .$ neighboring experience

where

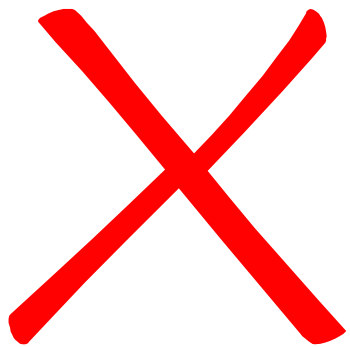
r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

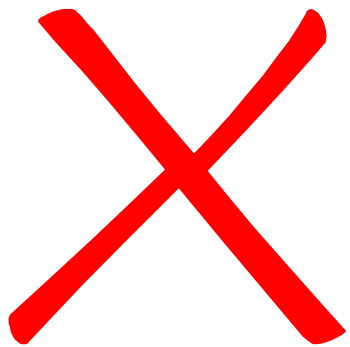
What is the "residual"?

- Ideally measure "residual" by fitting region as a GLM factor with thousands of levels
- One region with no claims will cause GLM not to converge
- If model does converge, definition of region probably too large for these techniques



A / E ?

- Theoretically aligned with multiplicative GLM
- If there are no claims (A=0), the residual is zero regardless of the value of E, thus losing information
- If there is so much data that A is never 0, you can probably fit a GLM using region anyway

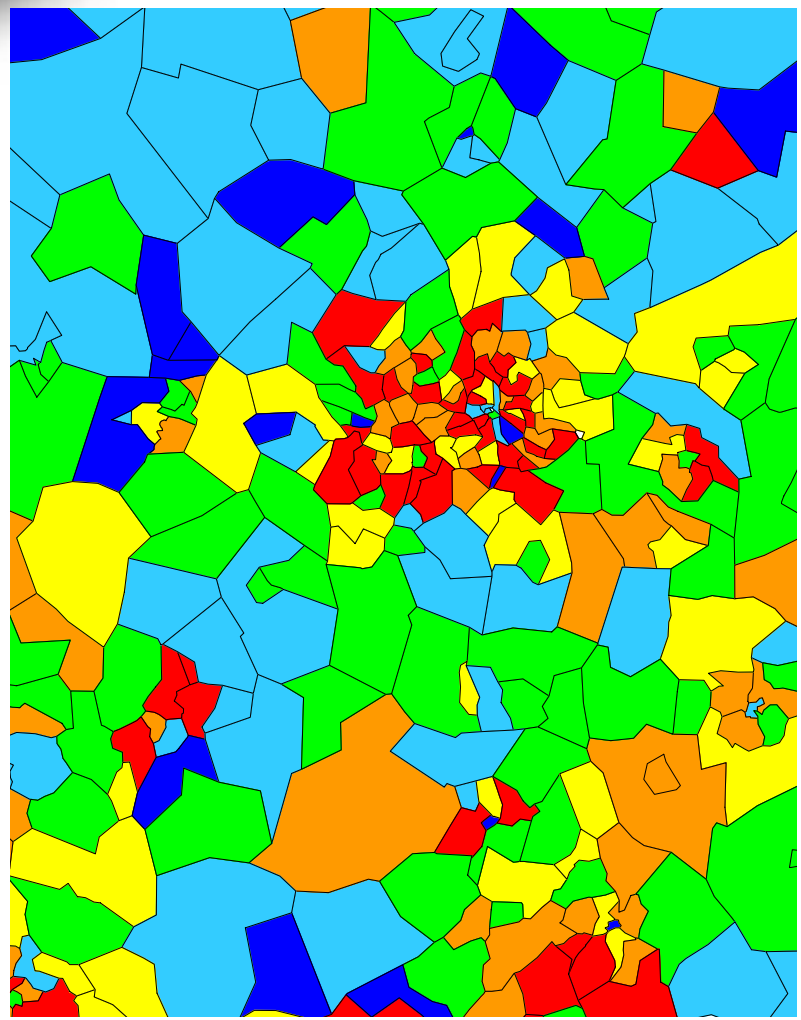


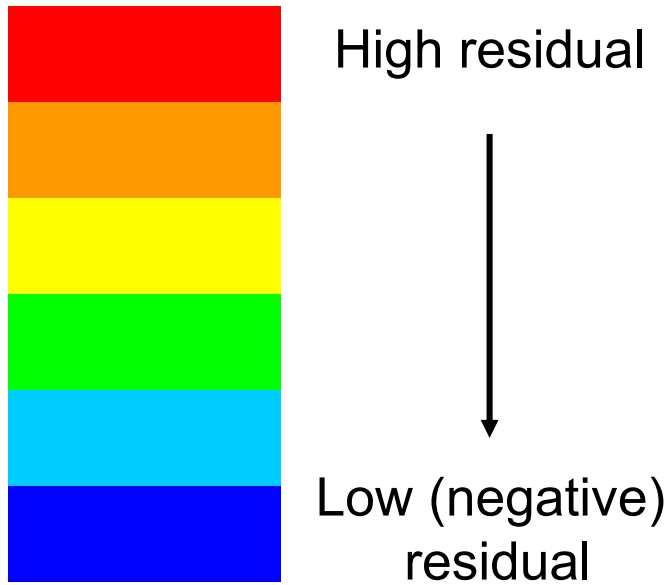
A - E

- Works for A=0
- Not aligned with multiplicative GLM
- Simple, robust and easy to calculate
- Eg for claim frequency consider

(Actual number - Expected number) / Exposure

Example residual risk
UK household contents theft frequency





A model form

$r_i^* = Z.r_i + (1 - Z) .$ neighboring experience

where

r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

$Z(e_i) = \{ e_i / (e_i + a) \}^m$, e_i = exposure in region i

A model form

$r_i^* = Z.r_i + (1 - Z) .$ neighboring experience

where

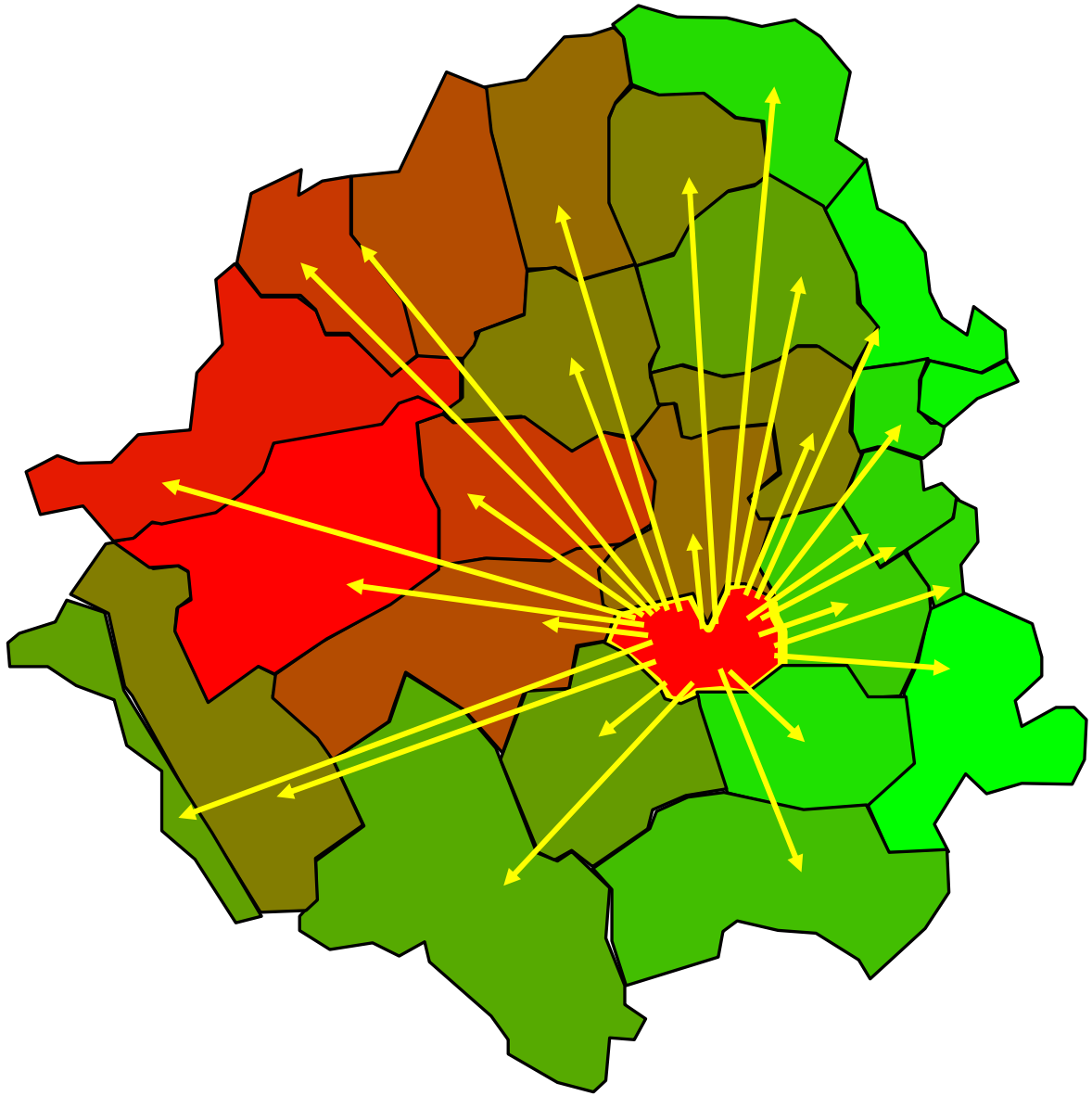
r_i^* = smoothed residual risk

r_i = unsmoothed residual risk

Z = credibility function

$Z(e_i) = \{ e_i / (e_i + a) \}^m$, e_i = exposure in region i

Definitions of "neighboring"



Model

$$r_i^* = Z(e_i).r_i + (1 - Z(e_i)) \sum_j e_j.r_j.f(d_{ij}) / \sum_j e_j.f(d_{ij})$$

where

r_i^* = smoothed residual r_i = unsmoothed residual

$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 \}^{½}$$

$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n.d_{ij})$ etc

Parameters

$$r_i^* = Z(e_i).r_i + (1 - Z(e_i)) \sum_j e_j.r_j.f(d_{ij}) / \sum_j e_j.f(d_{ij})$$

where

r_i^* = smoothed residual r_i = unsmoothed residual

$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 \}^{½}$$

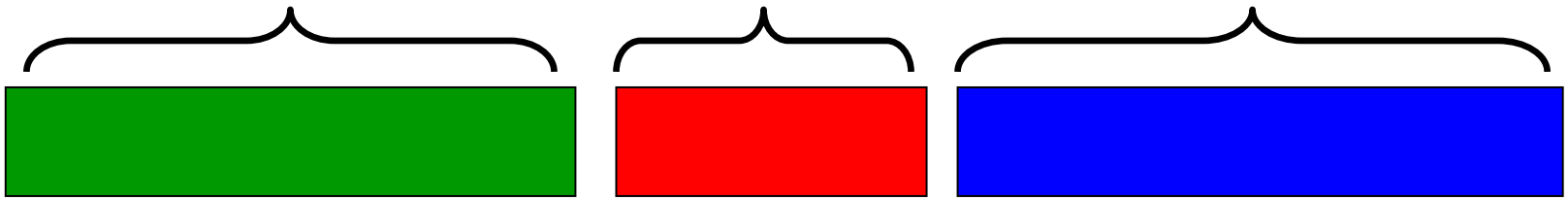
$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n.d_{ij})$ etc

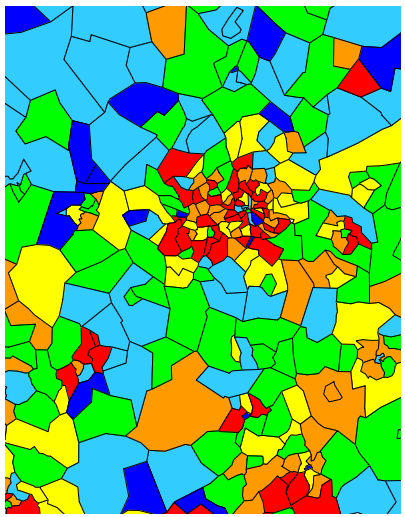
Finding the parameters

Calculate residuals

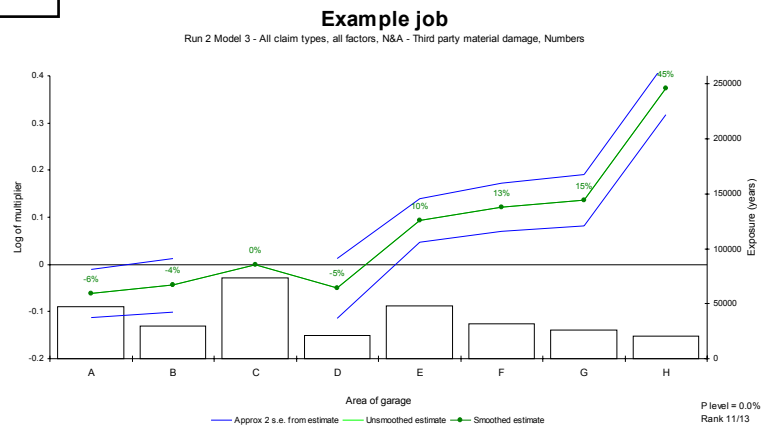
Seek parameters which minimize error

Save for determining zoning relativities



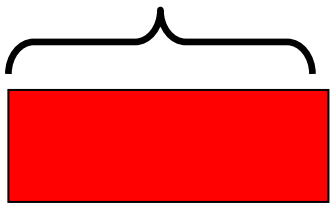


a, m, n, b



Finding the parameters

Seek parameters which minimize error



a, m, n, b

Error =

$$\sum (r_i^* - r_i)^2 * e_i$$

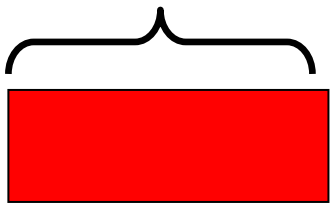
or

$$\sum \ln \{ 1 + (r_i^* - r_i)^2 \} * e_i$$

etc

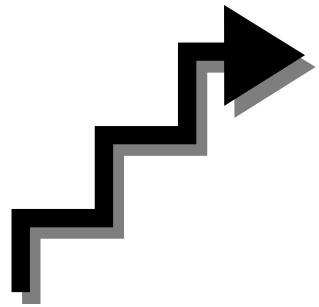
Finding the parameters

Seek parameters which minimize error

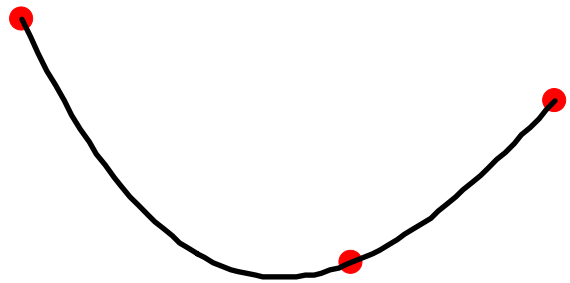


a, m, n, b

- Simple search



- Golden search

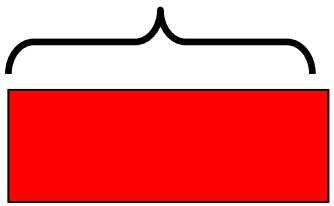


- Newton-Raphson

$$x' = x - \frac{f'(x)}{f''(x)}$$

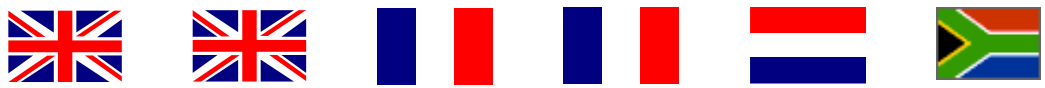
Finding the parameters

Seek parameters which minimize error



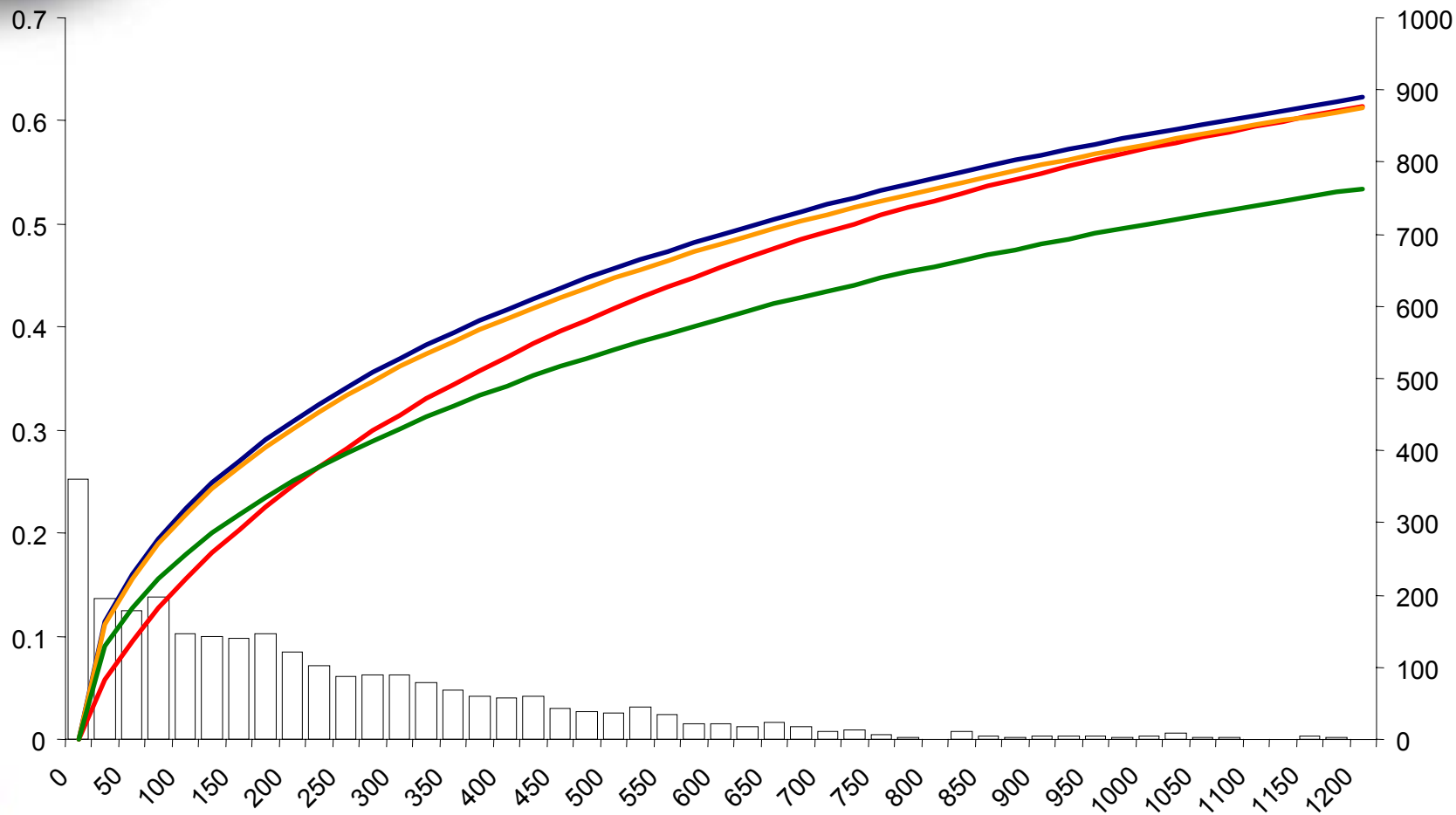
a, m, n, b

Real examples with $f(d_{ij}) = 1/d_{ij}^{n}$



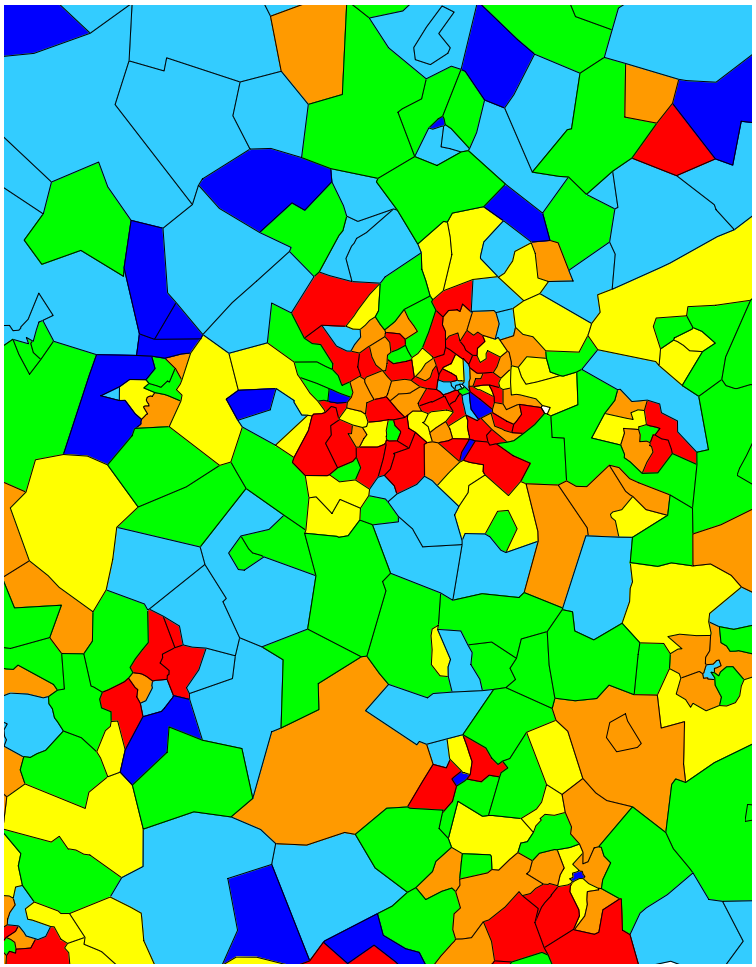
n	**1.9**	**2.2**	**1.9**	**2.0**	**1.8**	**2.2**
a	**1100**	**1150**	**3500**	**2500**	**1800**	**1000**
m	**0.75**	**0.75**	**0.5**	**0.5**	**0.47**	**0.75**

Credibility curves
Some UK examples

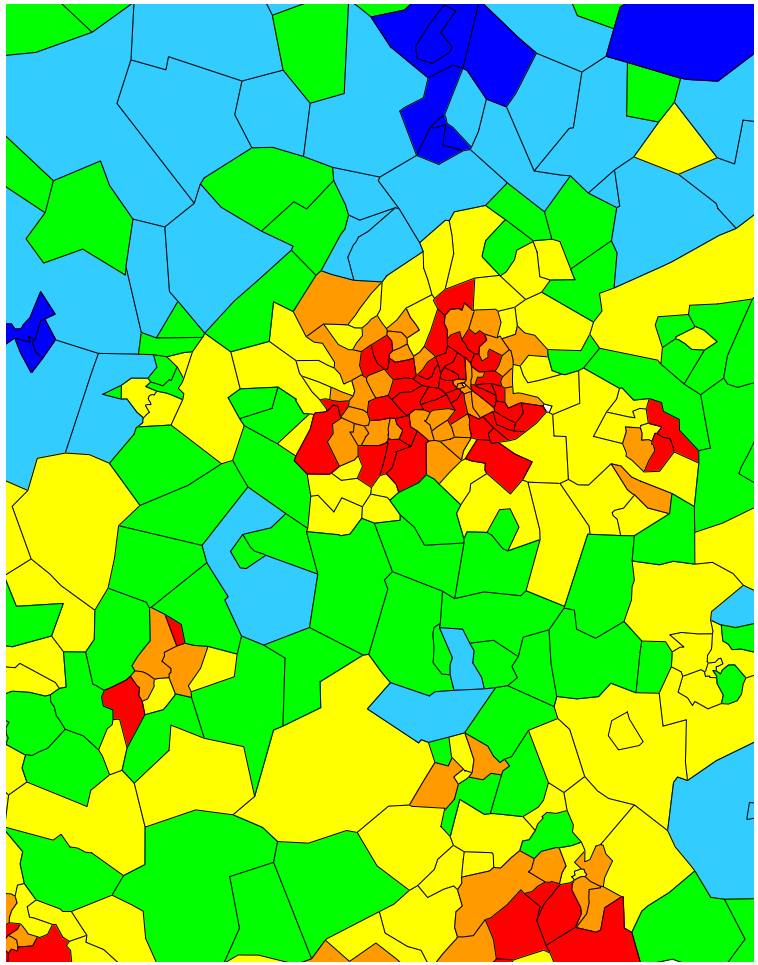


Example results

Unsmoothed residuals



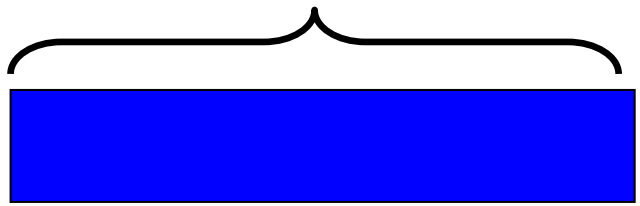
Smoothed residuals

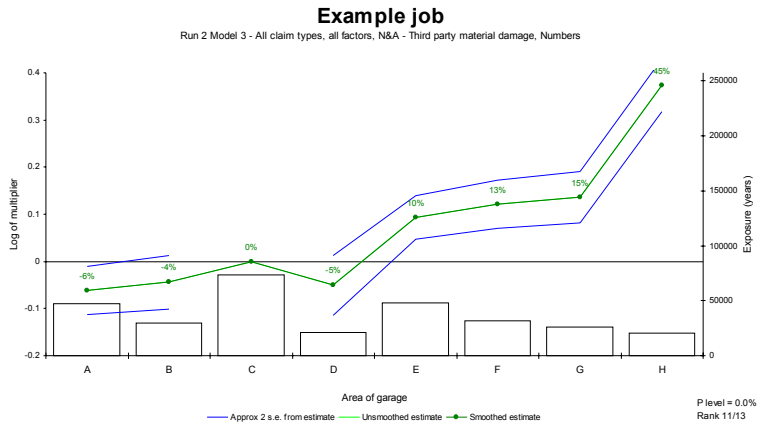


Finding the parameters

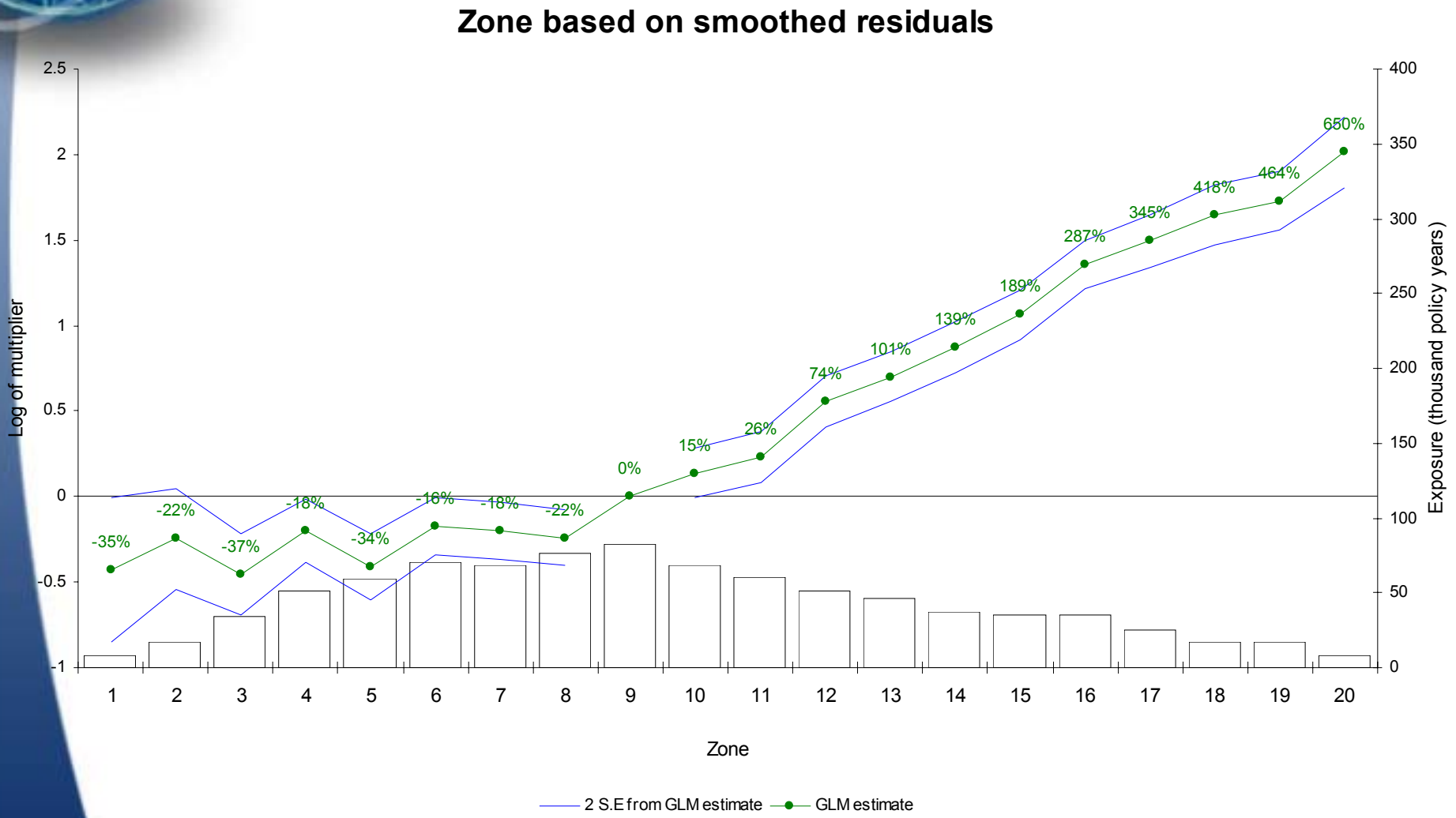
- Remodeling is necessary to "expand" the "squashed" smoothed residuals
- Fresh data required to avoid self-fulfilling prophesies

Save for determining zoning relativities



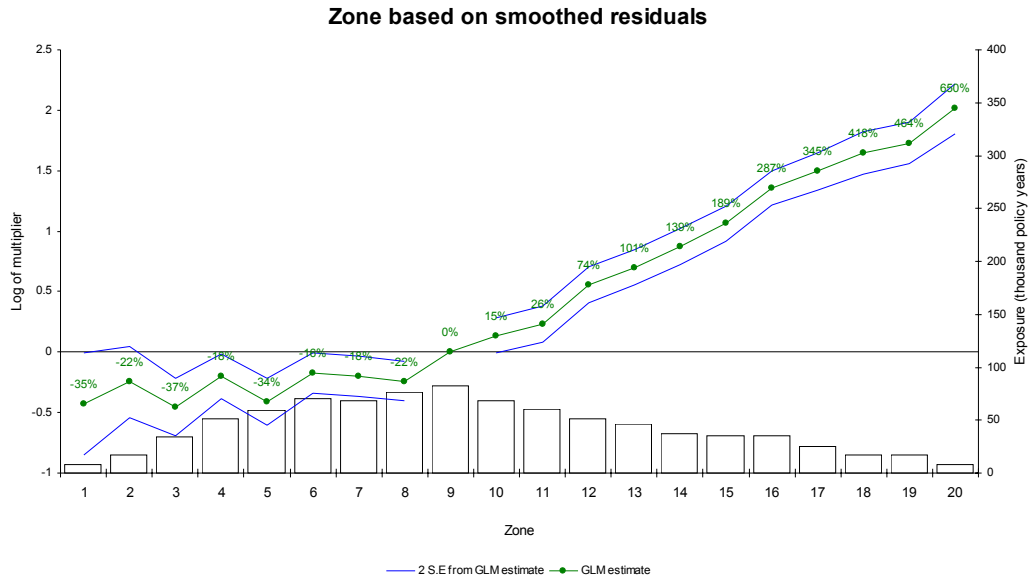


Finding the parameters
Effect of smoothed residual zone on fresh data



Finding the parameters
Effect of smoothed vs unsmoothed residual zone

Zone based on smoothed residuals

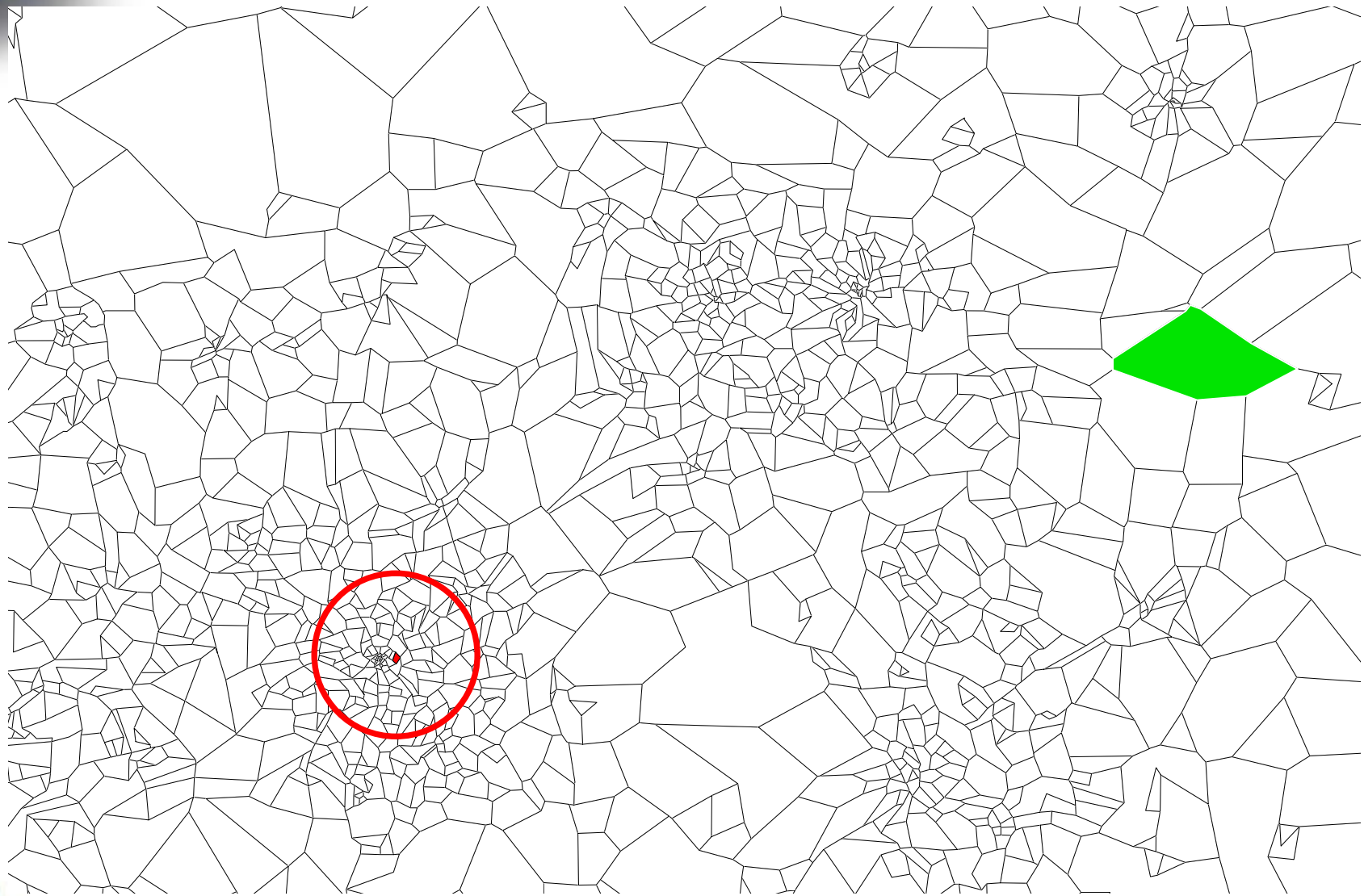


Zone based on unsmoothed residuals

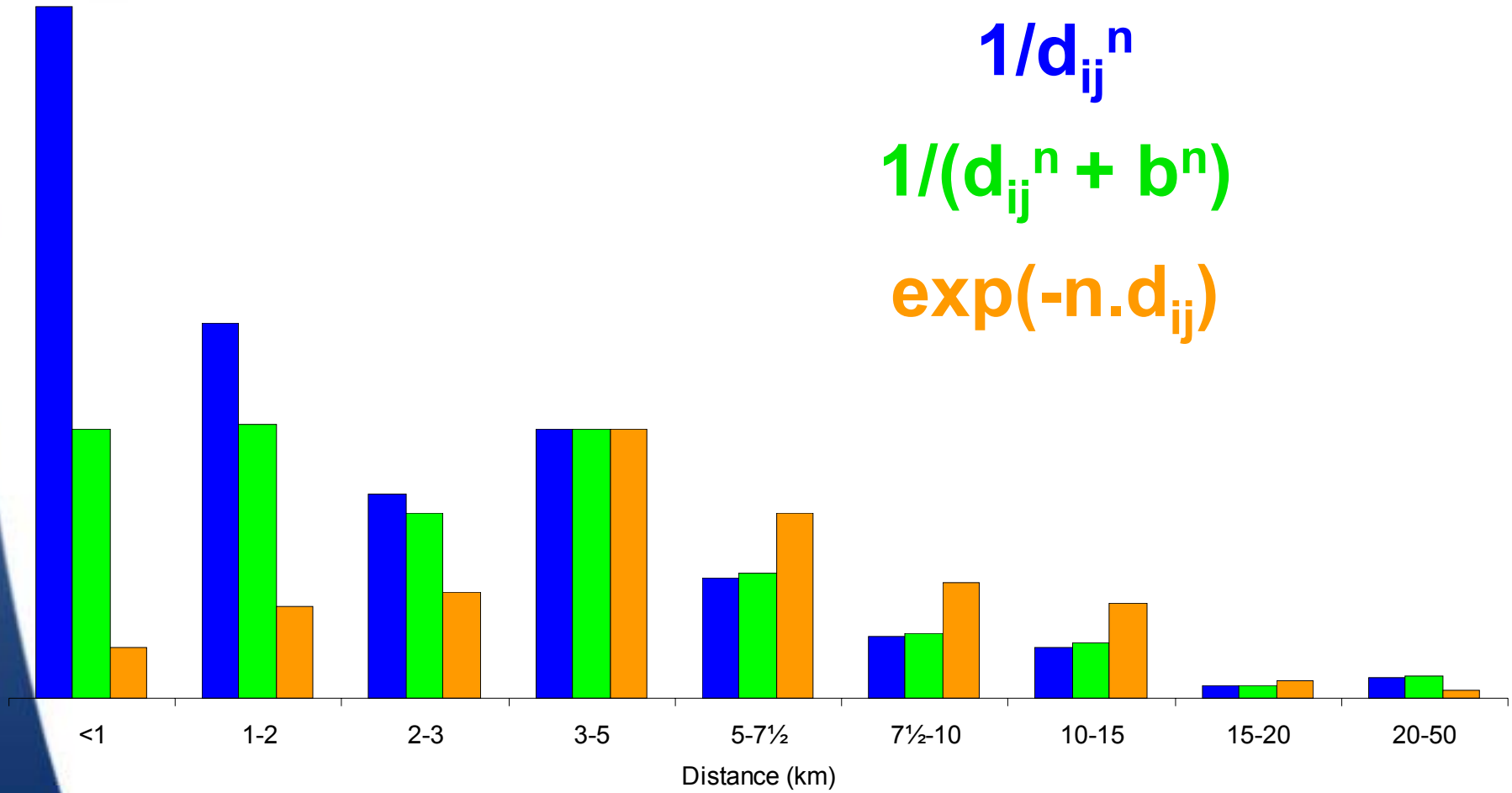
More details…

- Different
  - weighting functions
  - metrics
  - splits of the data
- Performance
- What to do when there is
  - no boundary data
  - no zip codes
- Geodemographic factors

Different weighting functions



Different weighting functions
Influence of neighbors in total - urban area

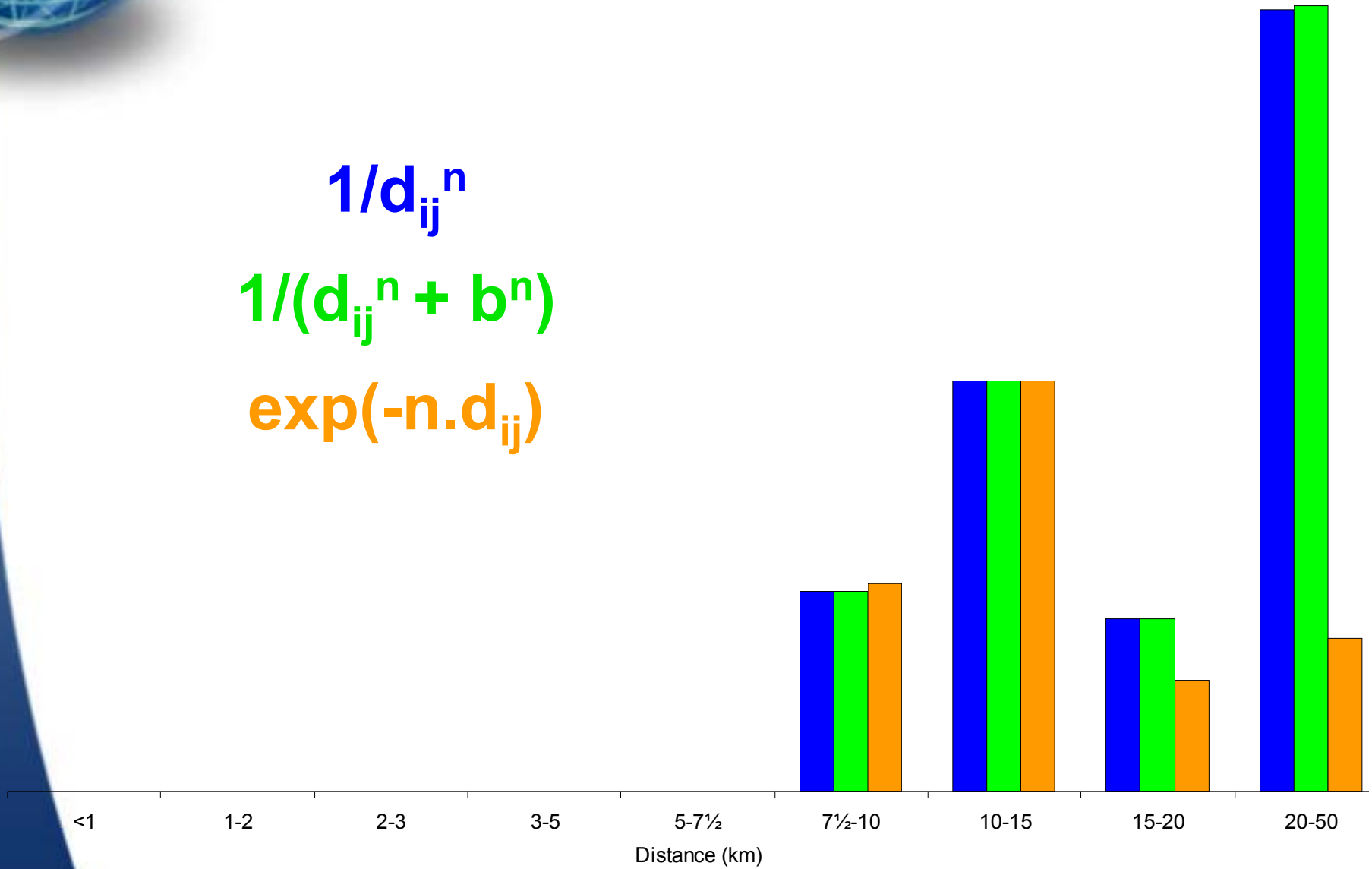


Different weighting functions
Influence of neighbors in total - rural area

$1/d_{ij}^{n}$

$1/(d_{ij}^{n} + b^{n})$

$\exp(-n.d_{ij})$



Different metrics

$$r_i^* = Z(e_i).r_i + (1 - Z(e_i)) \sum_j e_j.r_j.f(d_{ij}) / \sum_j e_j.f(d_{ij})$$

where

r_i^* = smoothed residual r_i = unsmoothed residual

$Z(e_i) = \{ e_i / (e_i + a) \}^m$ e_i = exposure in region i

$$d_{ij} = \{ (x_i - x_j)^2 + (y_i - y_j)^2 + (s.q_i - s.q_j)^2 \}^{½}$$

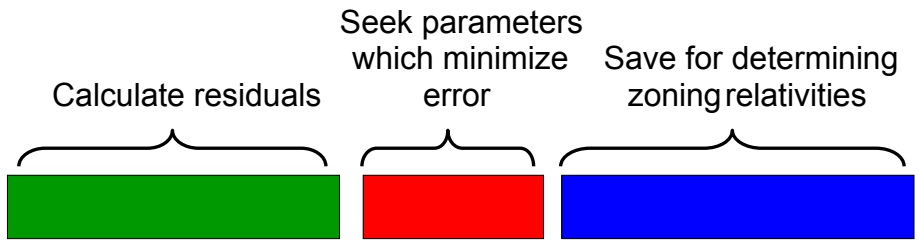
$f(d_{ij}) = 1/d_{ij}^n$ or $1/(d_{ij}^n + b^n)$ or $\exp(-n.d_{ij})$ etc

Splitting the data

- Generally a random split is best, otherwise policy characteristics distort results

- Weather related perils are a notable exception - here a time split may be more appropriate

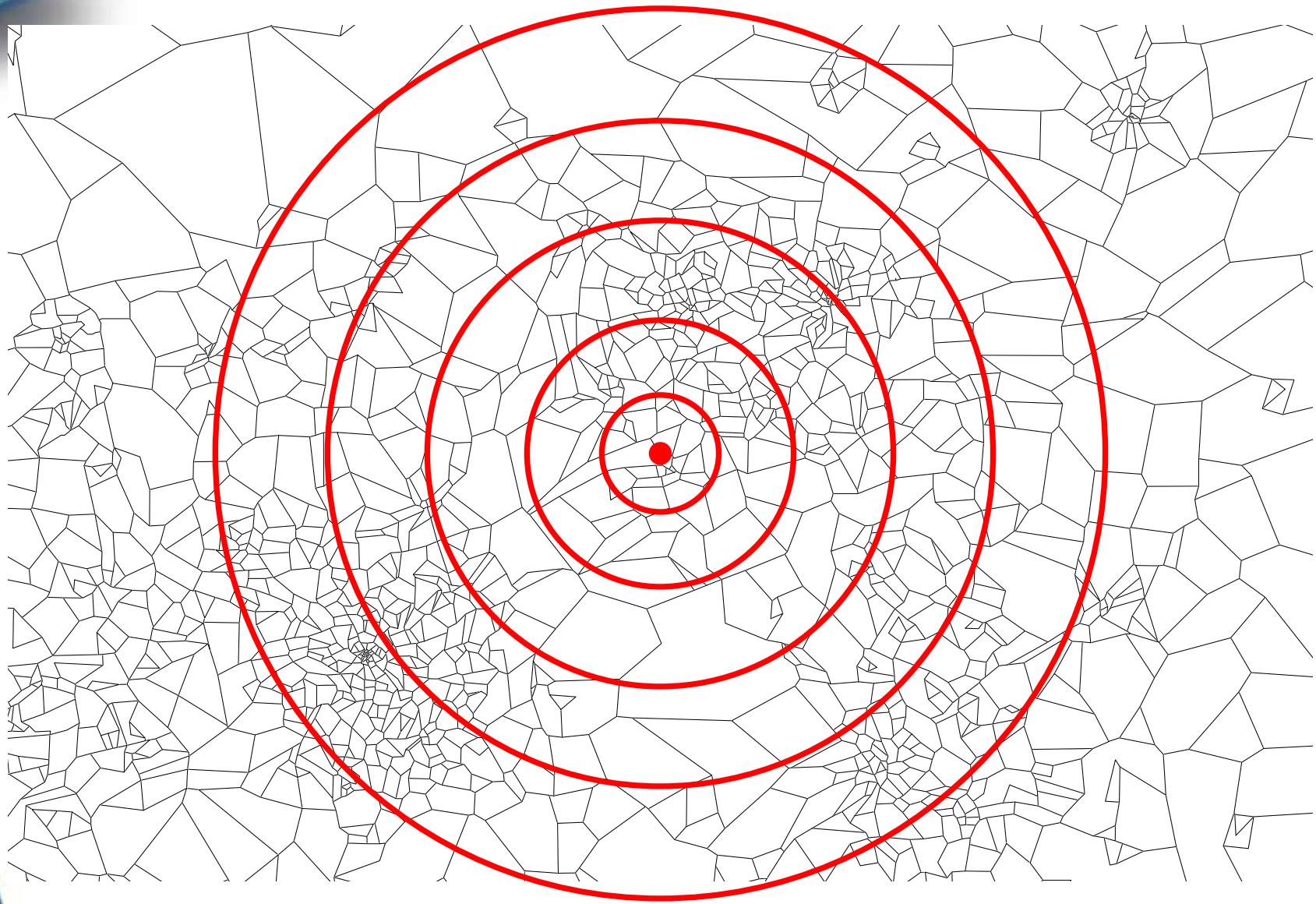
(not appropriate to model large weather events with this method)



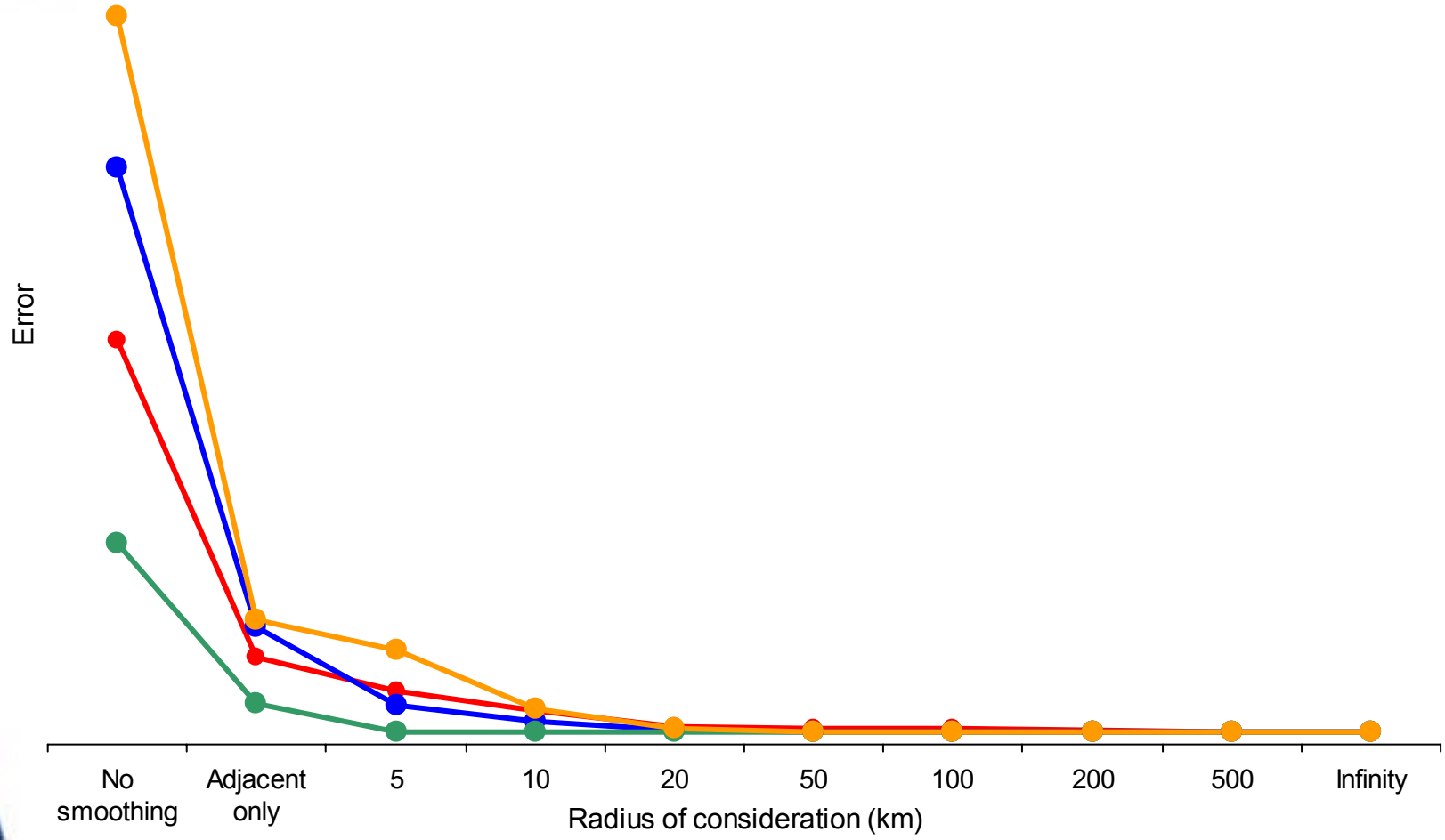
More details…

- Different
  - weighting functions
  - metrics
  - splits of the data
- Performance
- What to do when there is
  - no boundary data
  - no postcodes at all
- Geodemographic factors

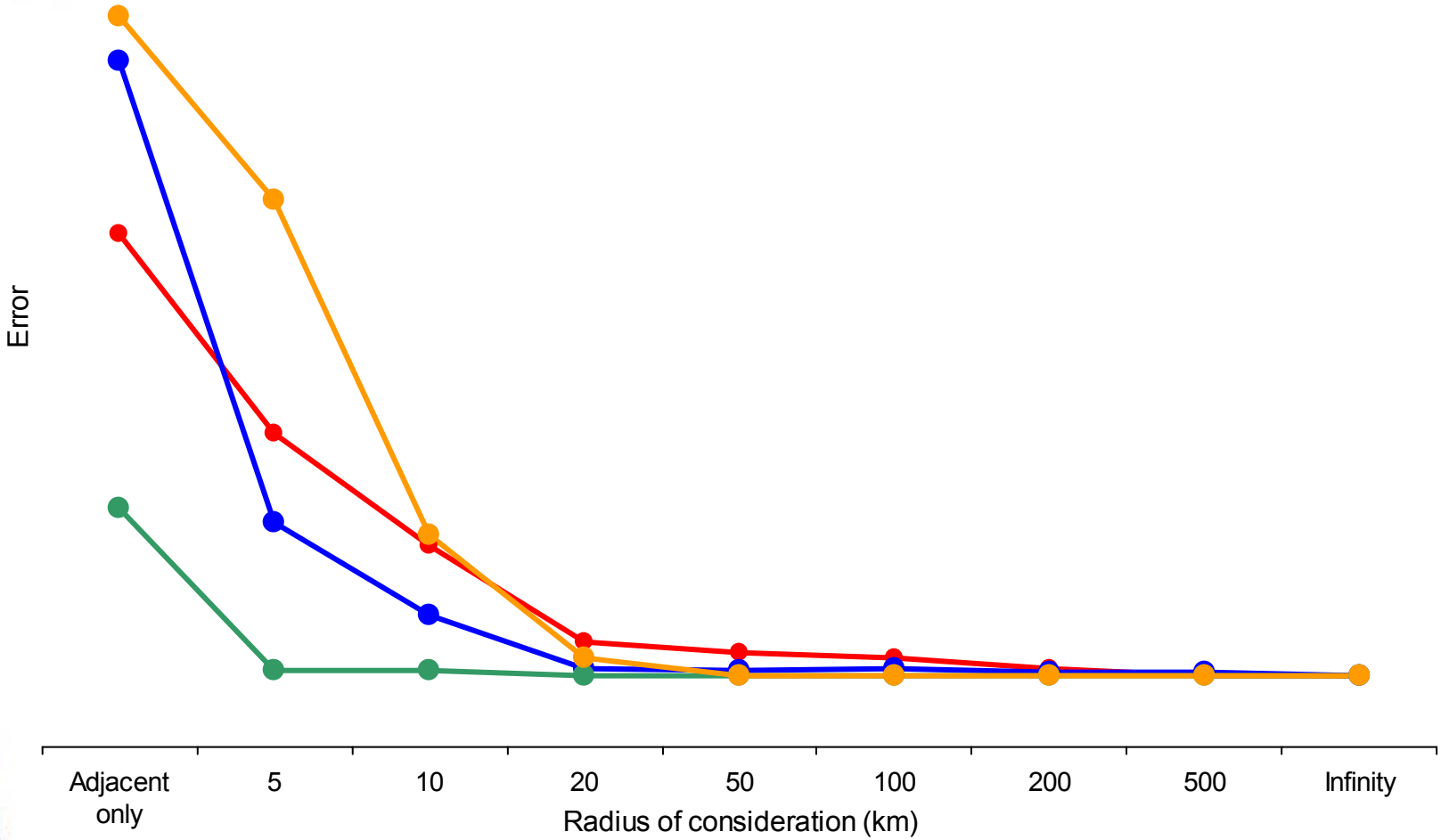
Limiting the definition of "neighboring"



Error with differing radii of consideration

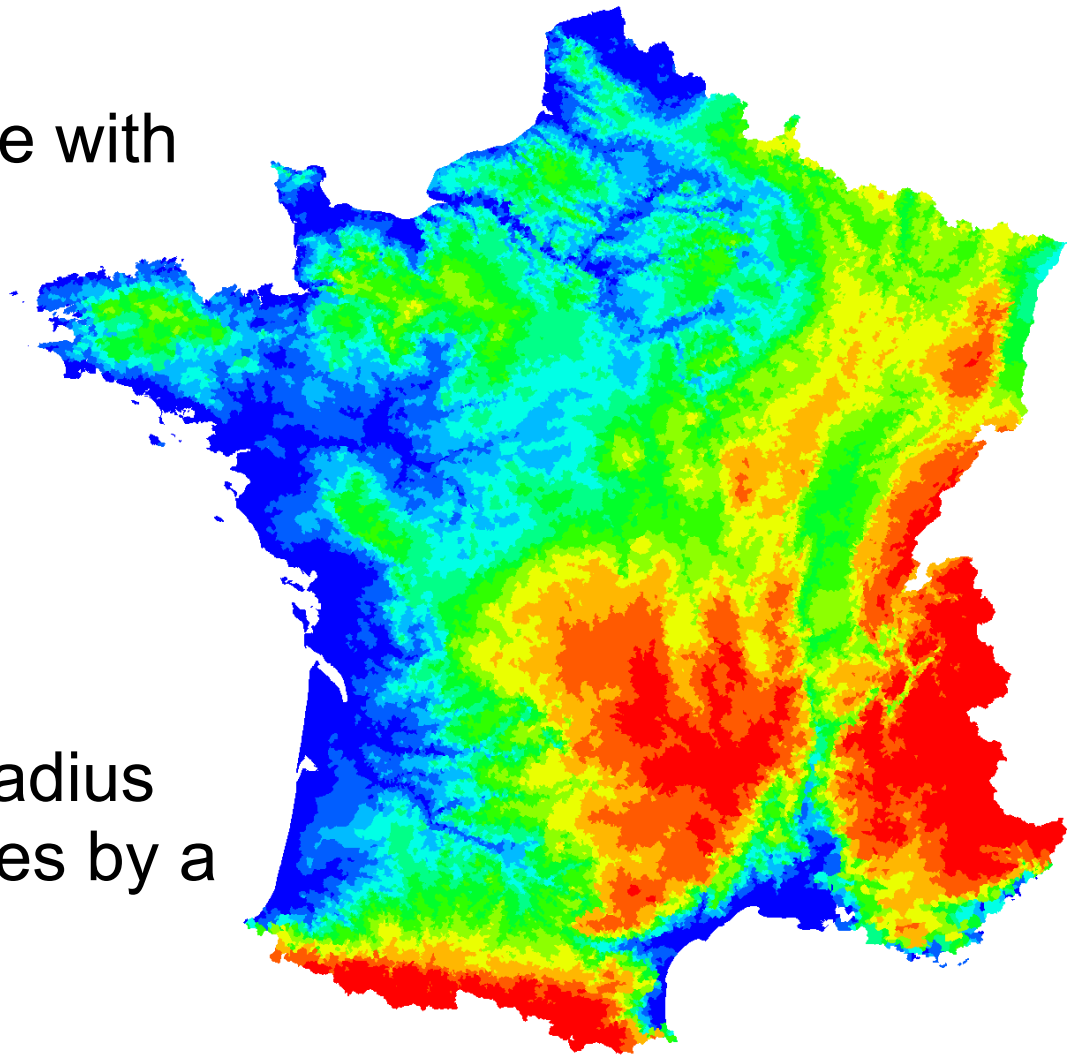


Error with differing radii of consideration



Computational short cuts

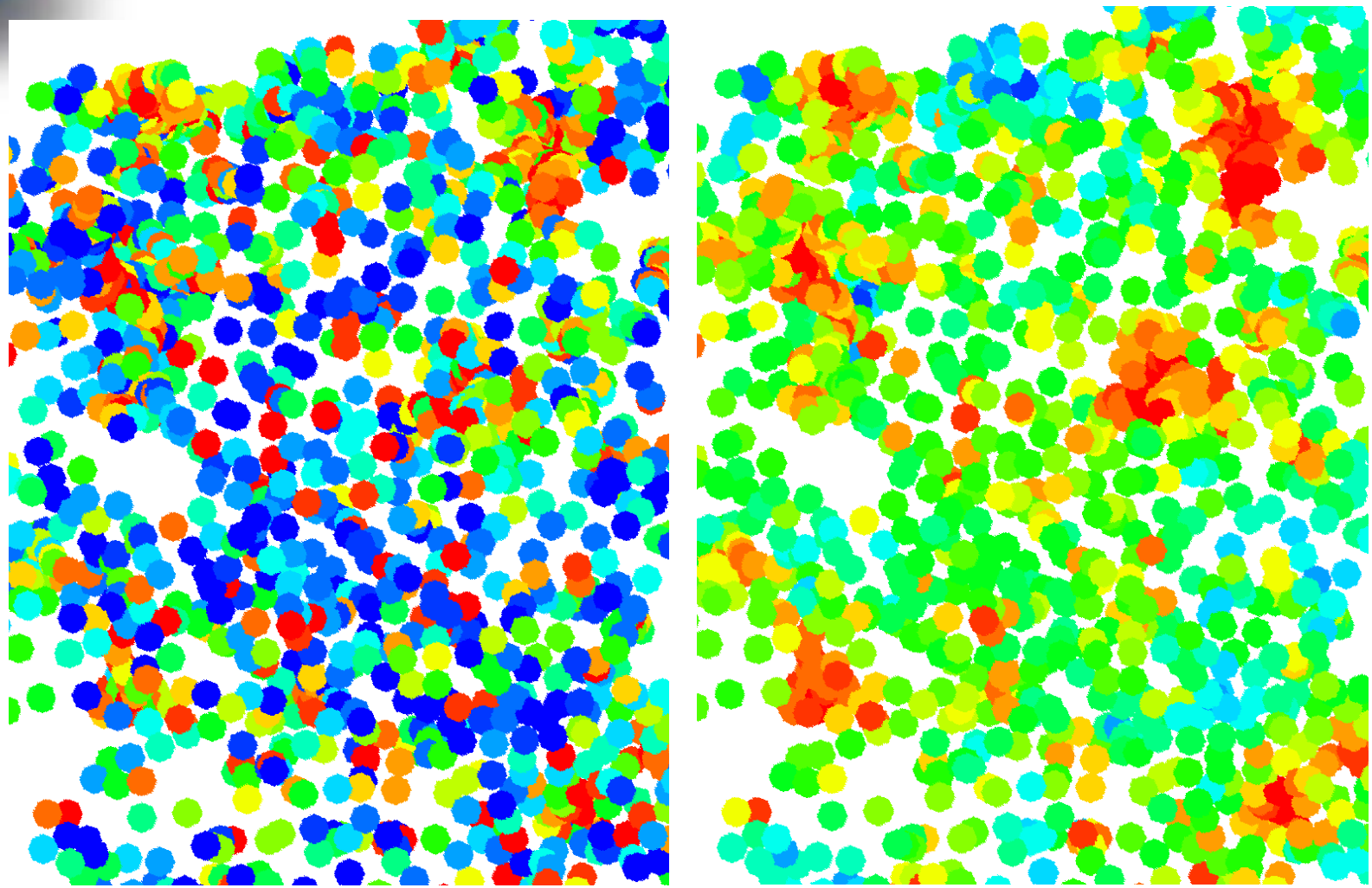
- Run times increase with (# regions)^2
- There are 36,500 communes in France, ie 1 billion calculations per iteration!
- Limiting to 50km radius decreases run times by a factor of 6



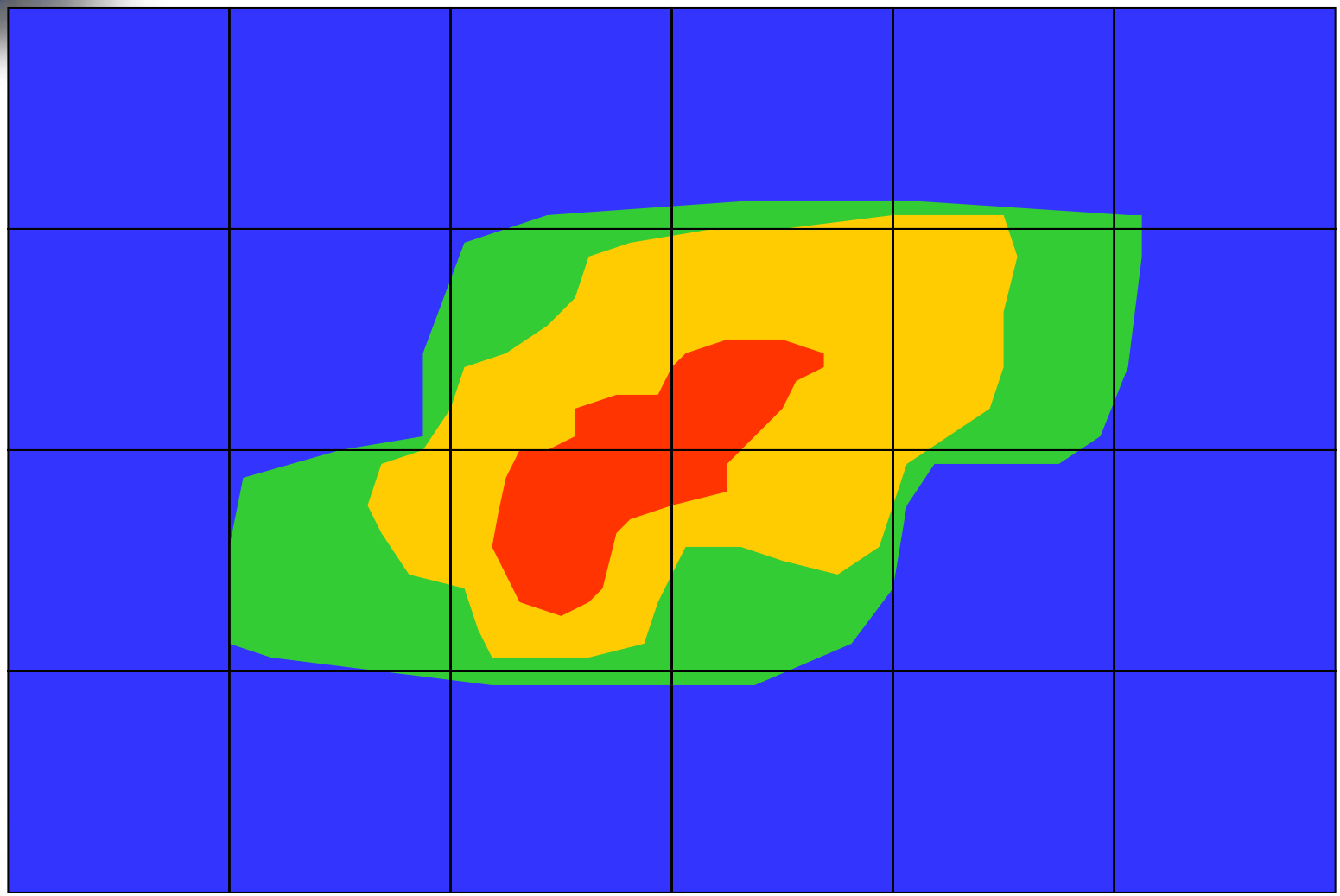
More details…

- Different
  - weighting functions
  - metrics
  - splits of the data
- Performance
- What to do when there is
  - no boundary data
  - no zip codes
- Geodemographic factors

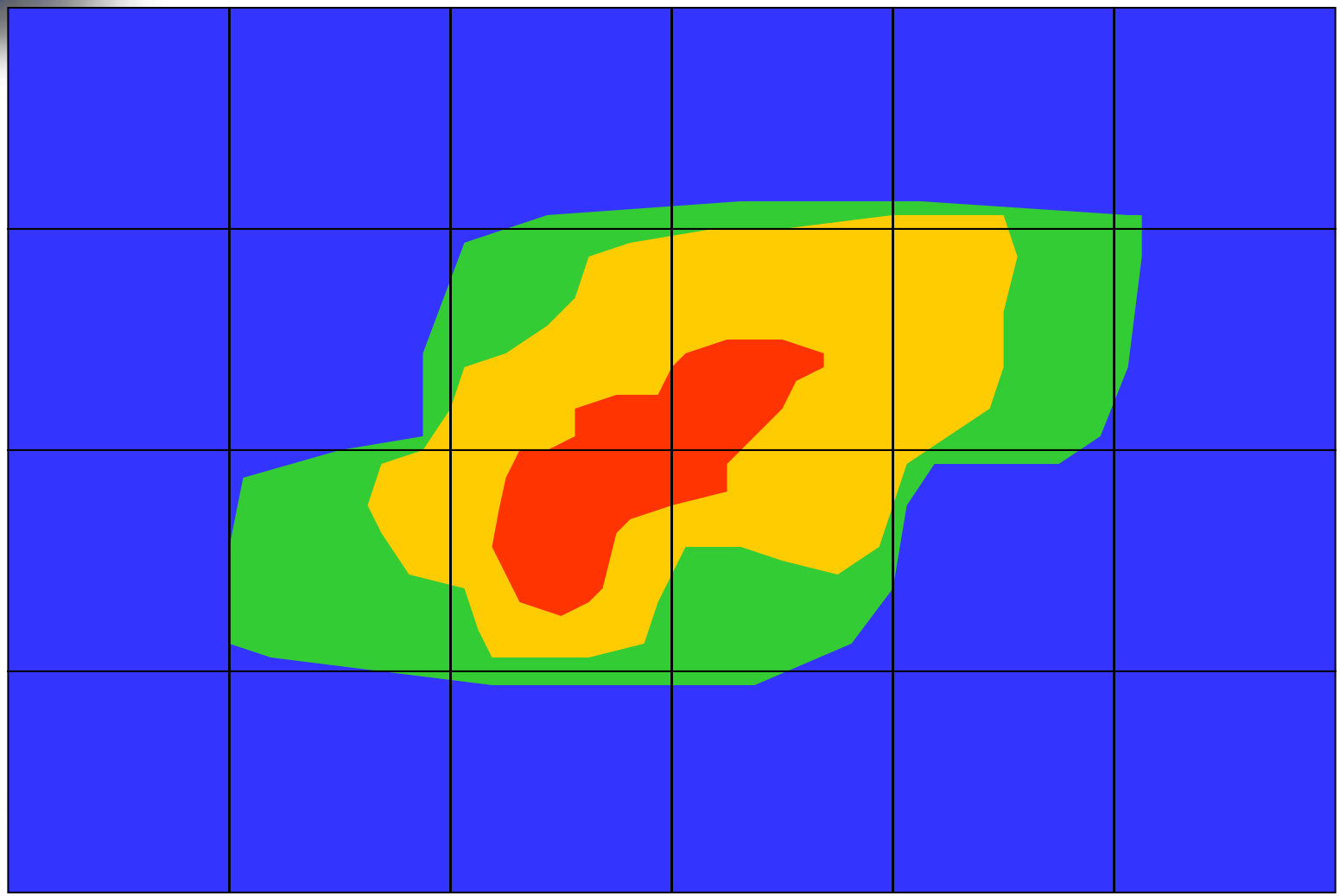
When no boundary data is available (but x, y is)…



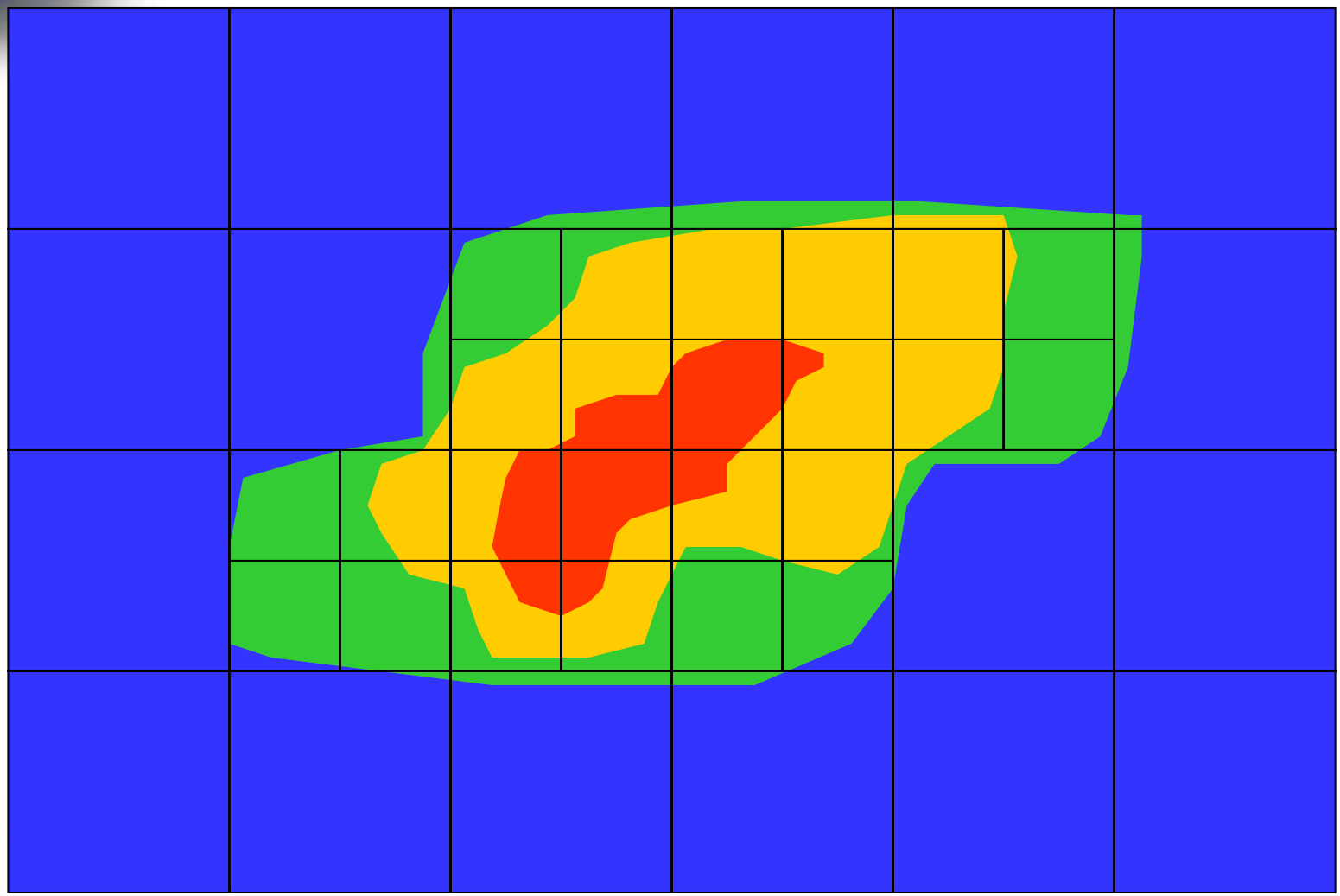
When no zip codes used…



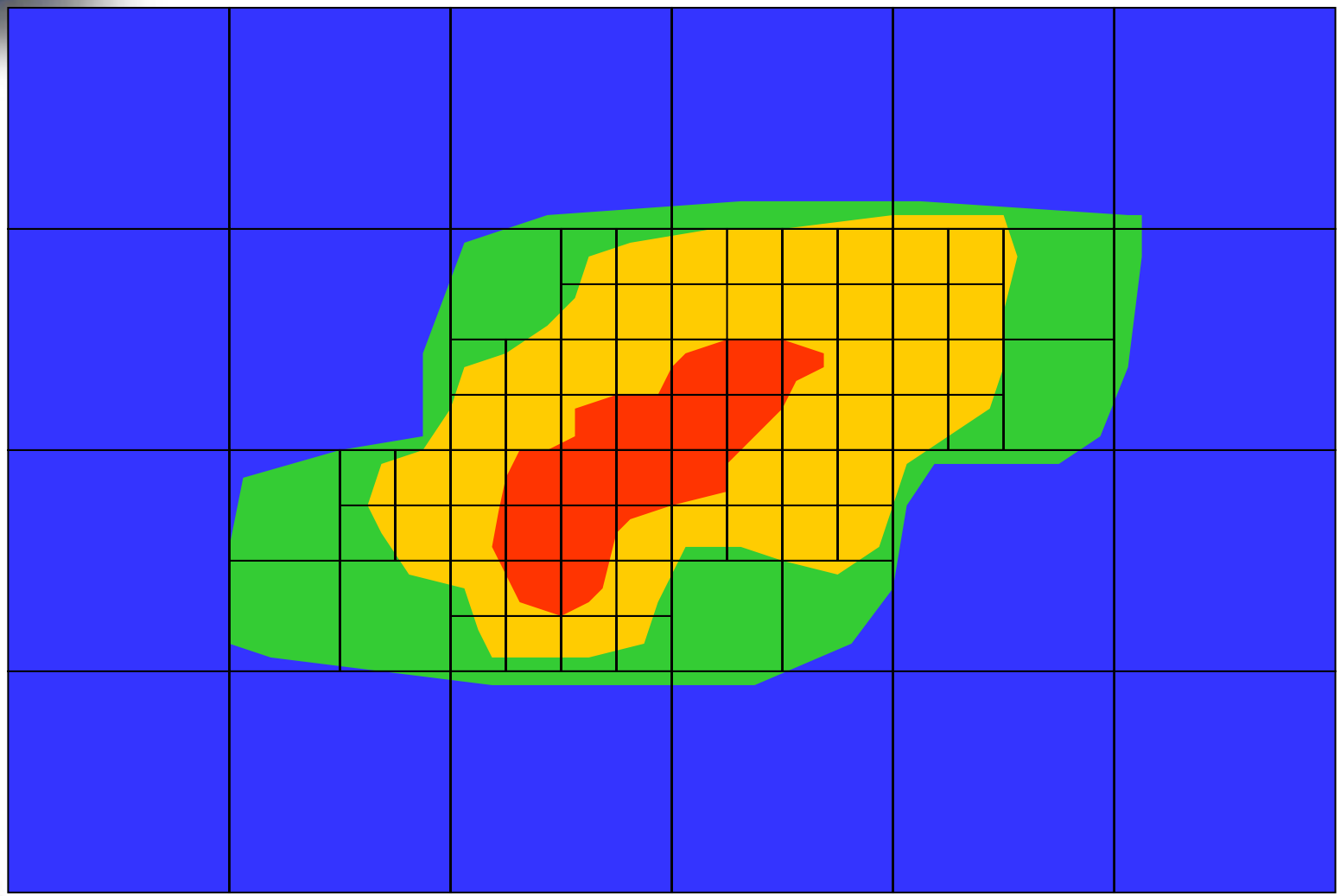
When no zip codes used…



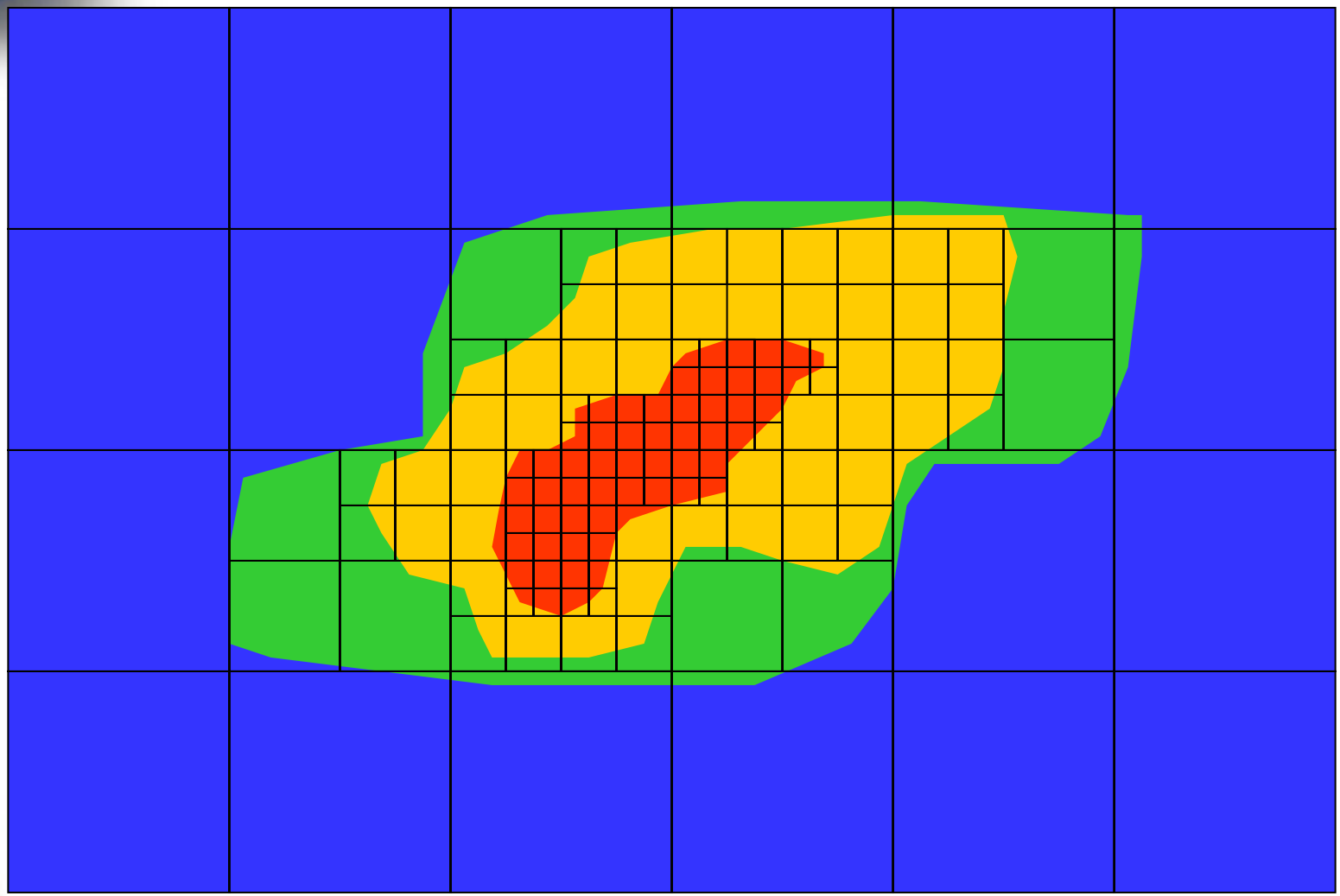
When no zip codes used…



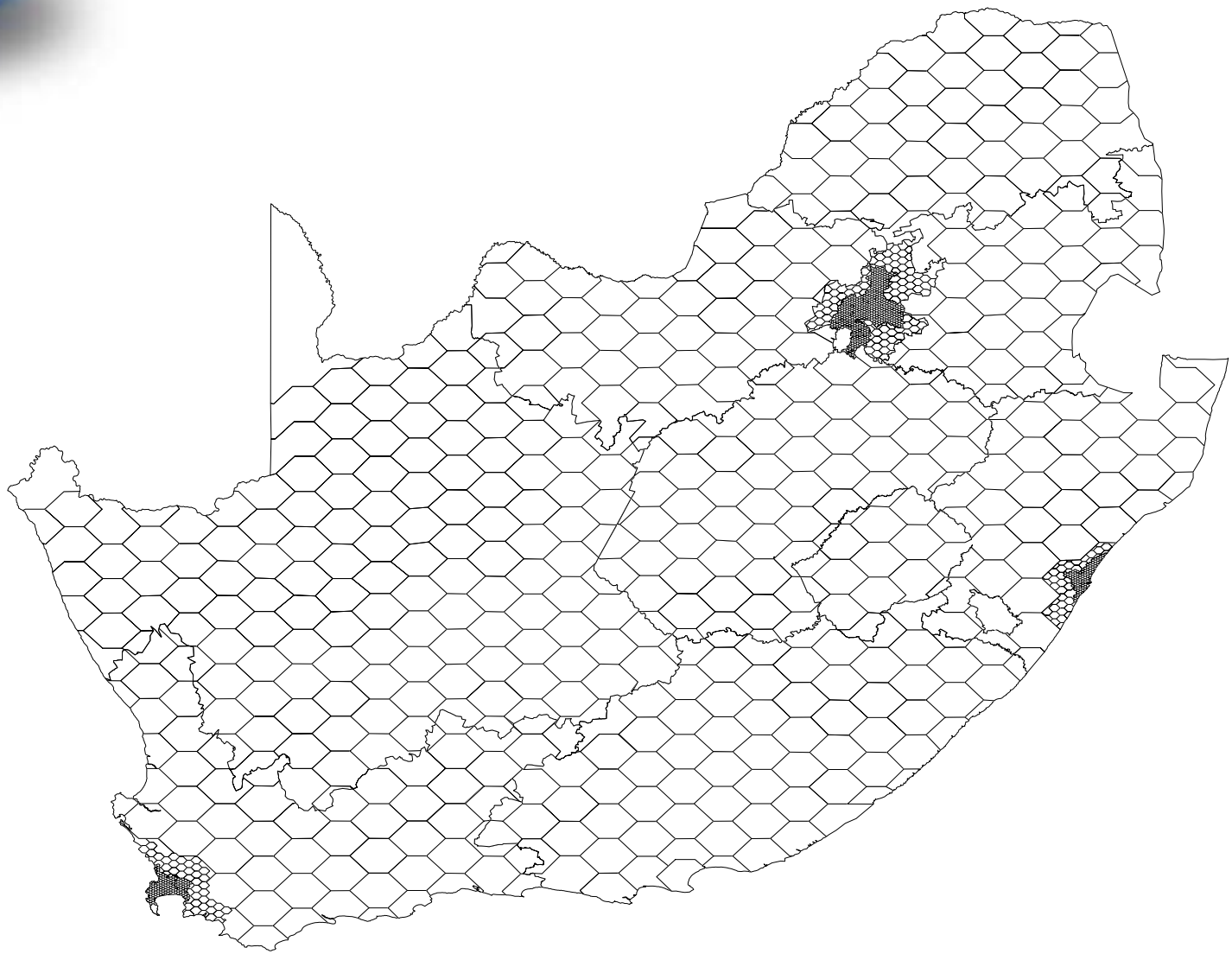
When no zip codes used…



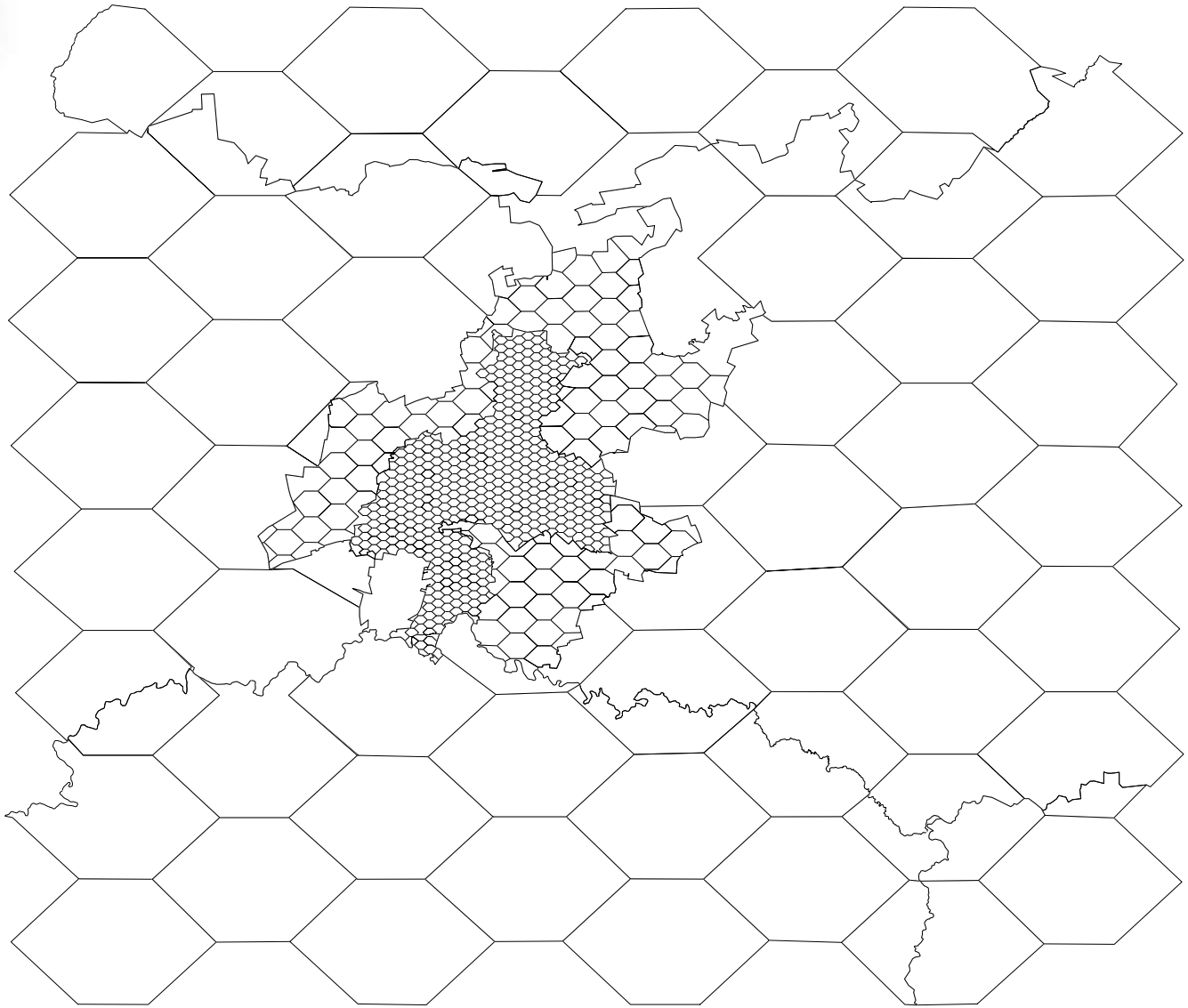
When no zip codes used…



When no zip codes used....



When no zip codes used…



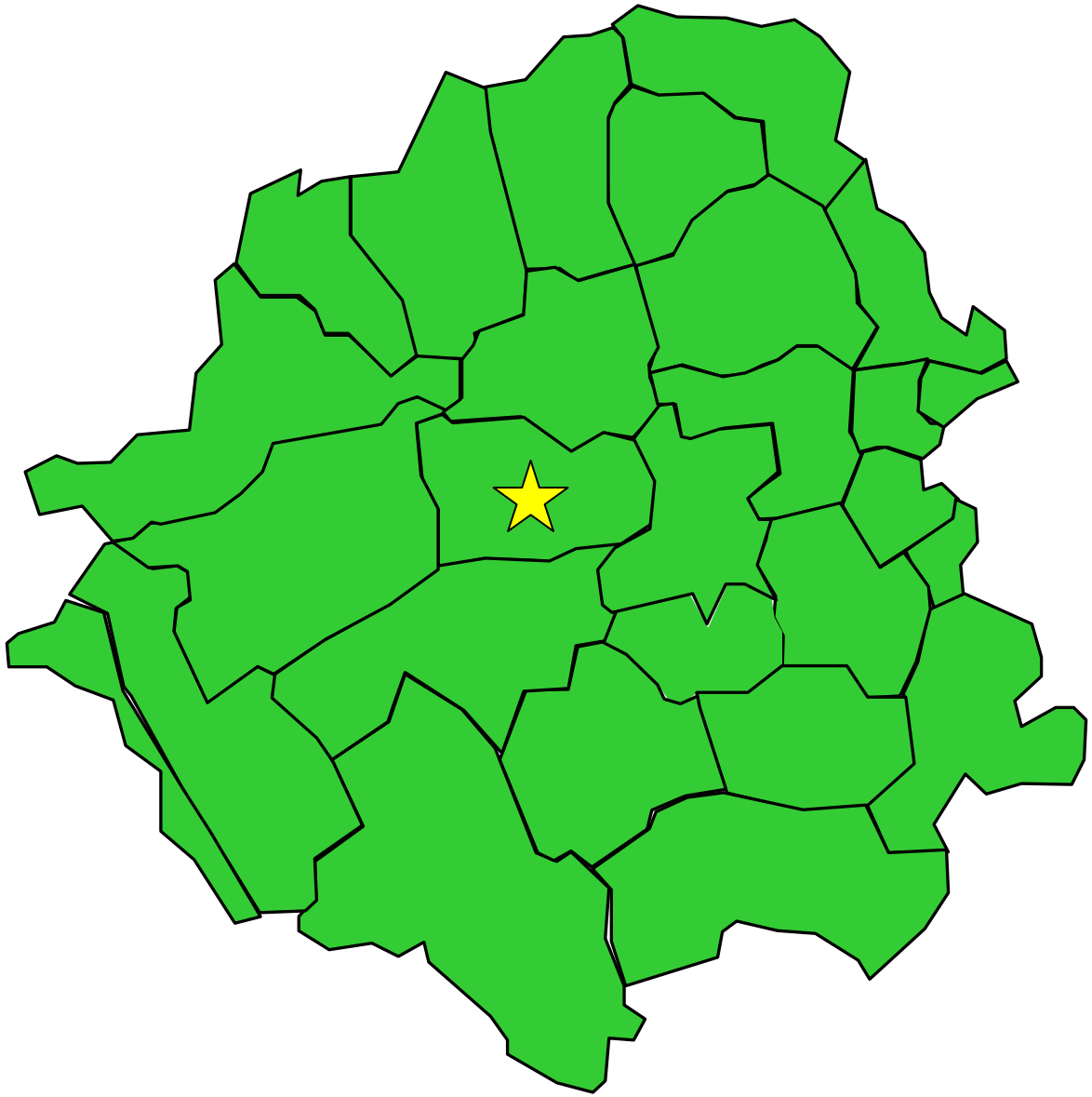
More details…

- Different
  - weighting functions
  - metrics
  - splits of the data
- Performance
- What to do when there is
  - no boundary data
  - no zip codes
- Geodemographic factors

Geodemographic factors

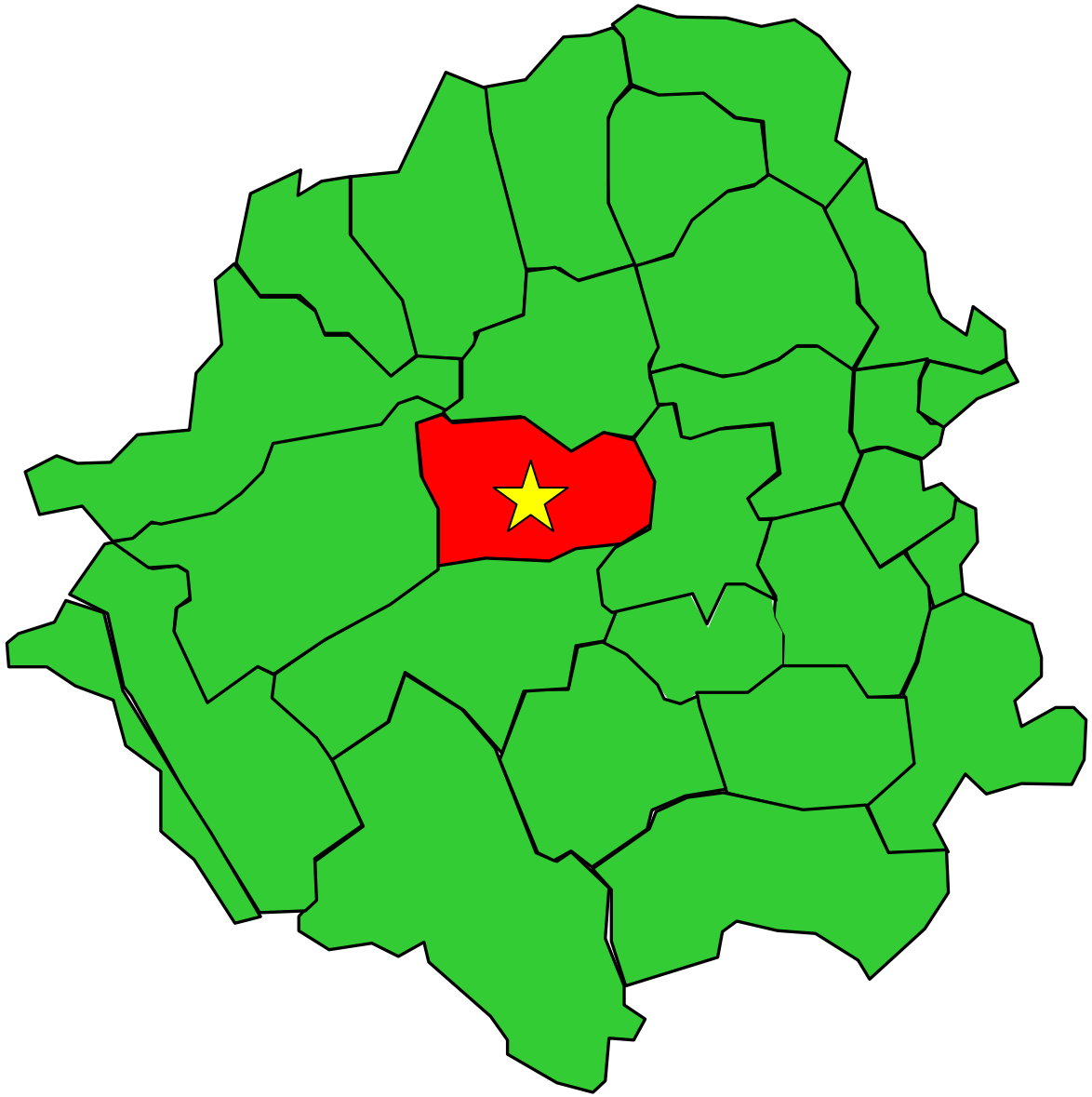
- Can be very predictive
- Various sources
- Even simple measures of urban density can be interesting

Measures of urban density



Measures of urban density

- Consider density of region

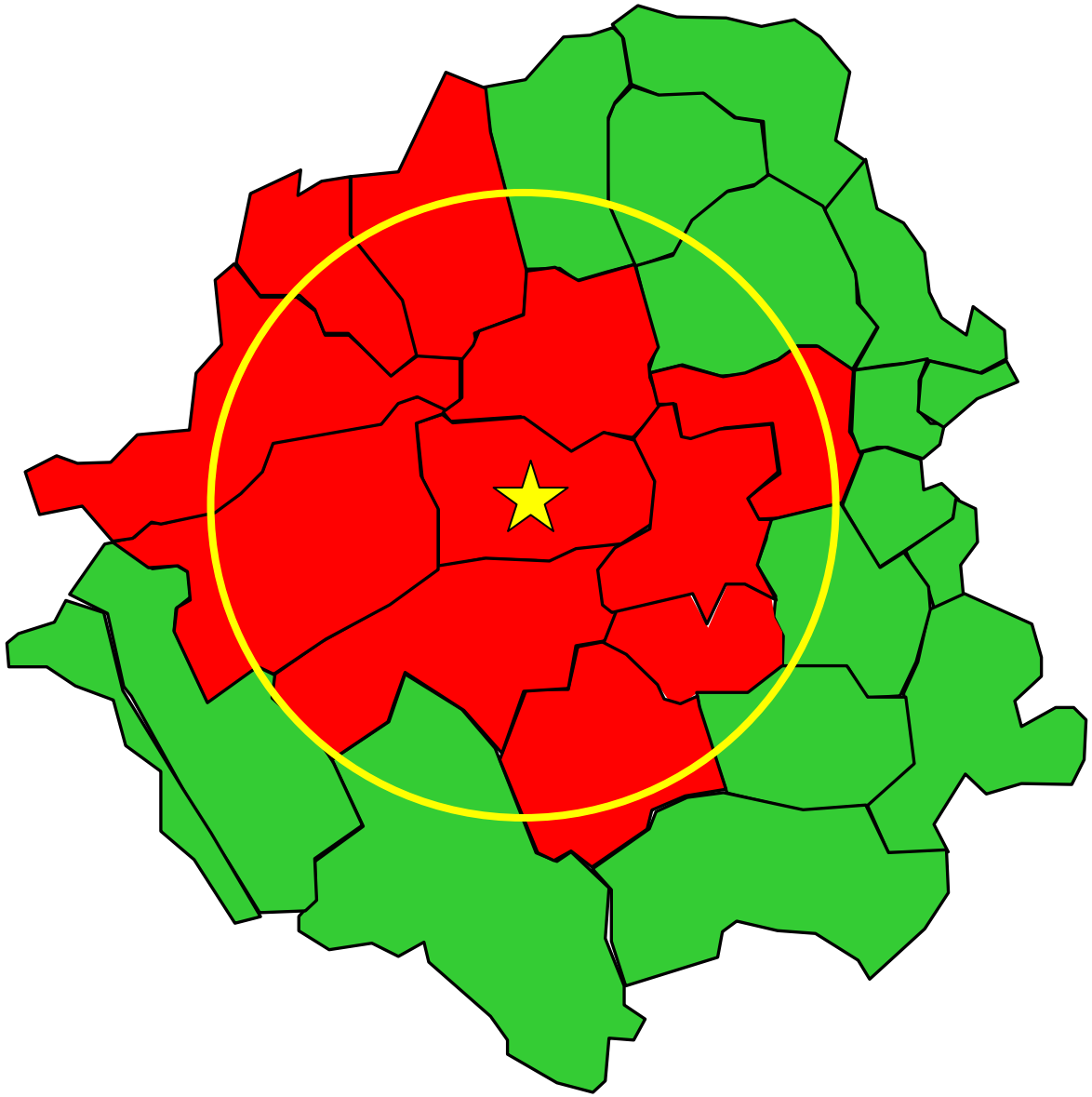


Measures of urban density

- Consider density of a wider post code region

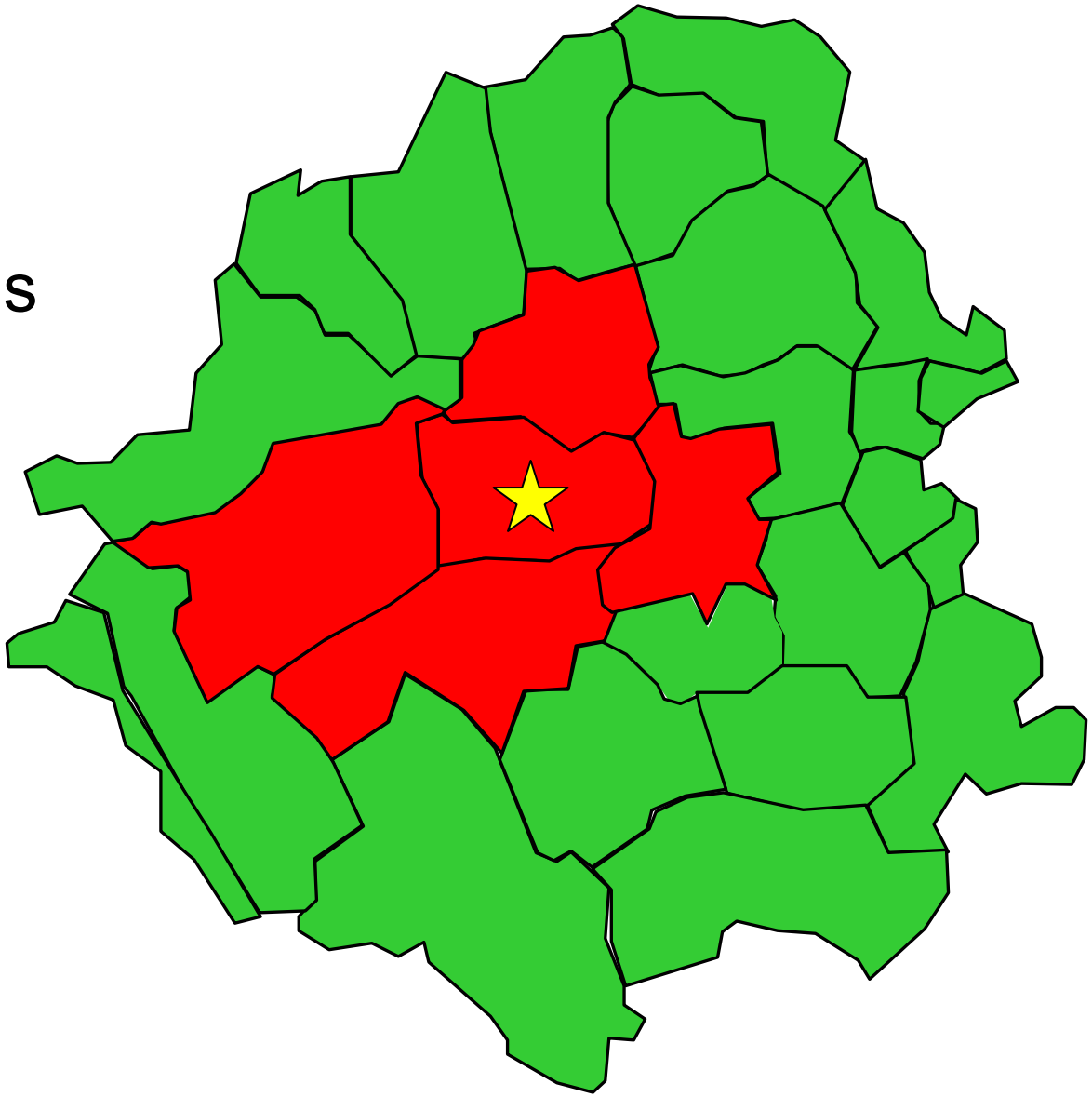
Measures of urban density

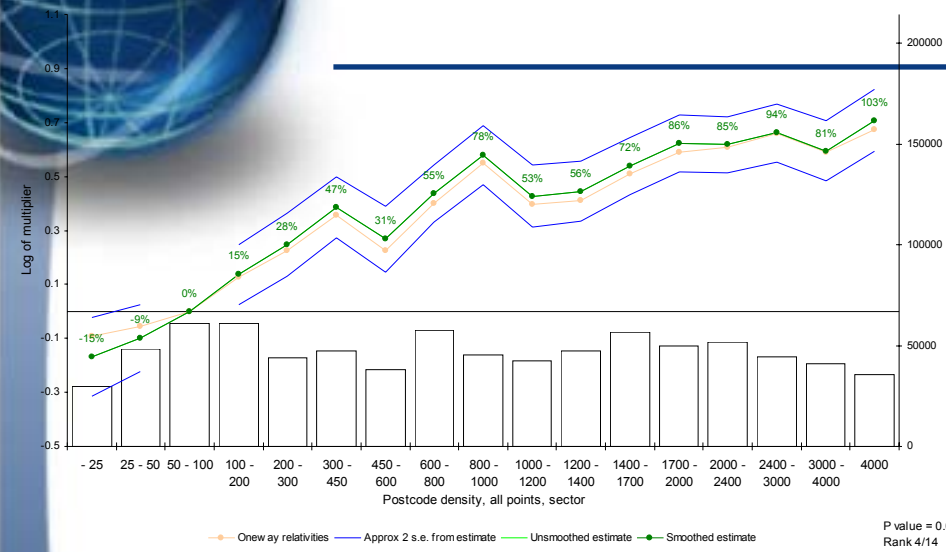
- Consider density of regions within a radius
- Useful when boundary data not available



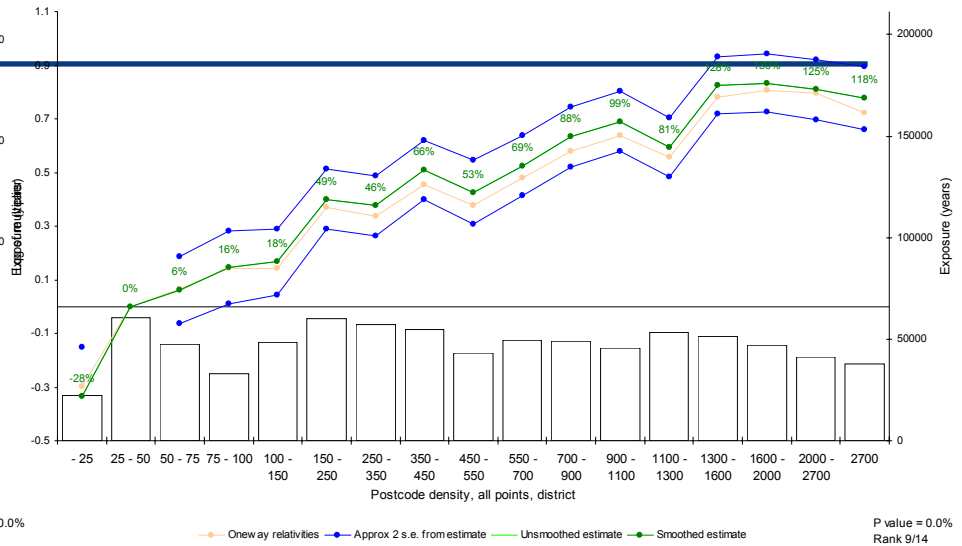
Measures of urban density

- Consider density of region plus adjoining regions

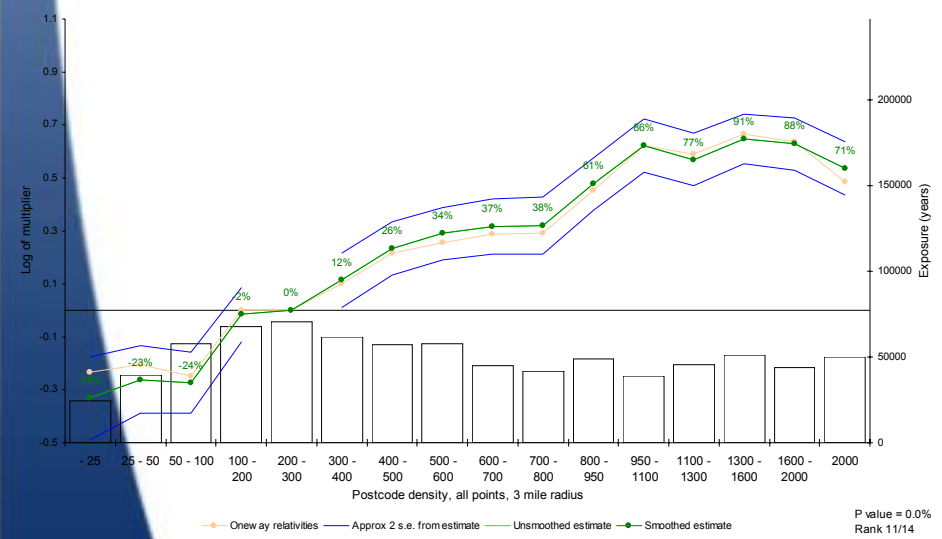




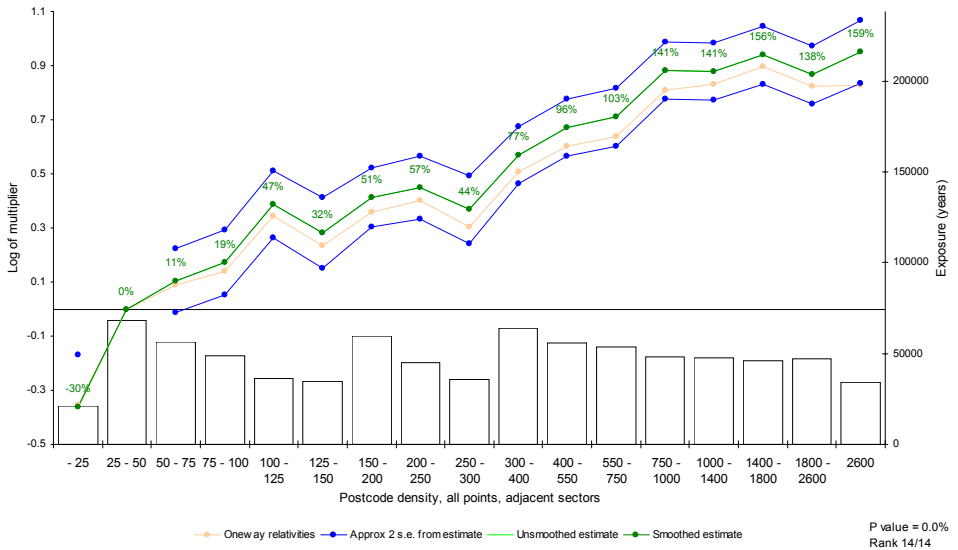
P value = 0.0%
Rank 4/14



P value = 0.0%
Rank 9/14



P value = 0.0%
Rank 11/14



P value = 0.0%
Rank 14/14

Geodemographic factors

- Can be used

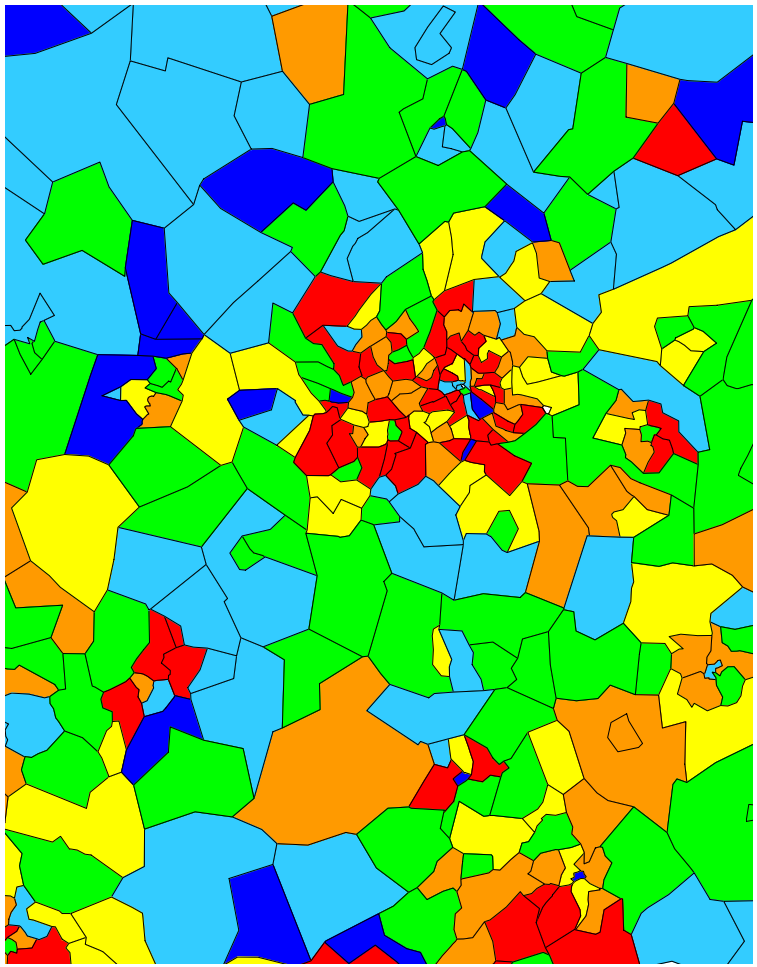
  (a) alongside zones derived as above

  (b) to standardize experience prior to smoothing

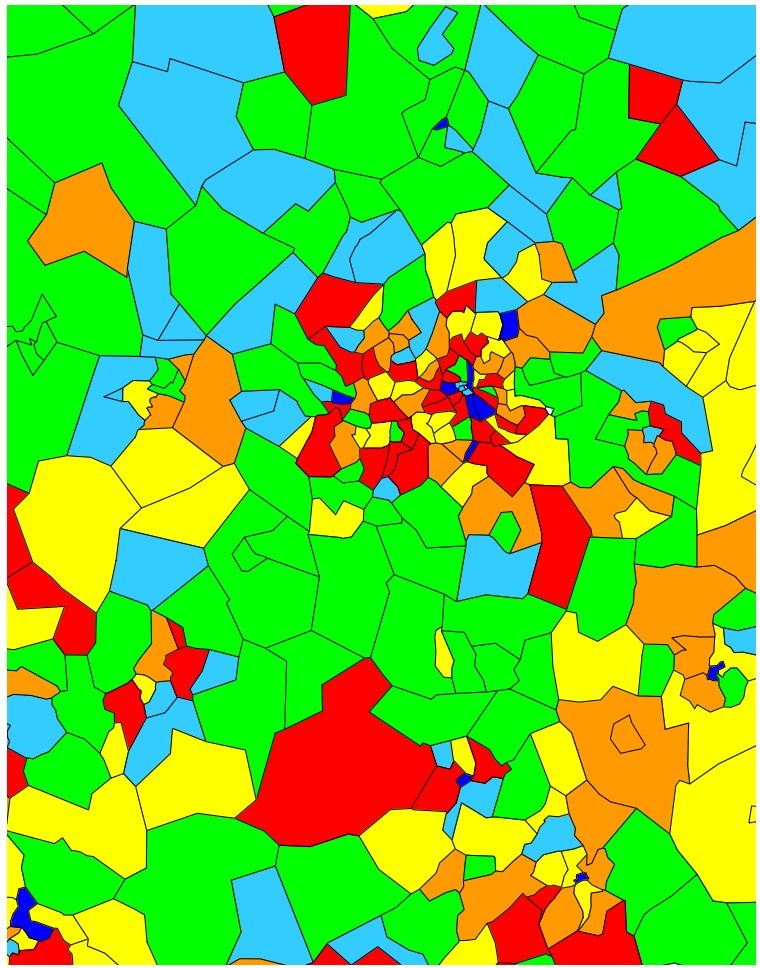
- Investigate which yields most predictive zone
- Generally speaking, seek to standardize for factors which yield inherently smoother residuals

Unsmoothed residuals

Density not in standardizing GLM

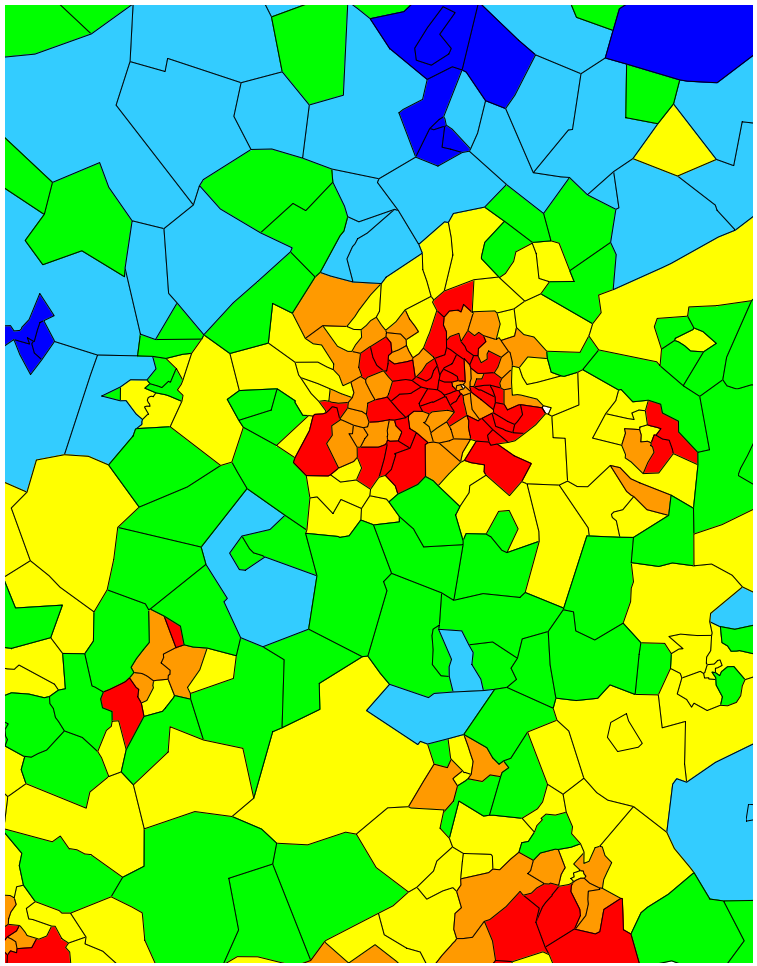


Density in standardizing GLM

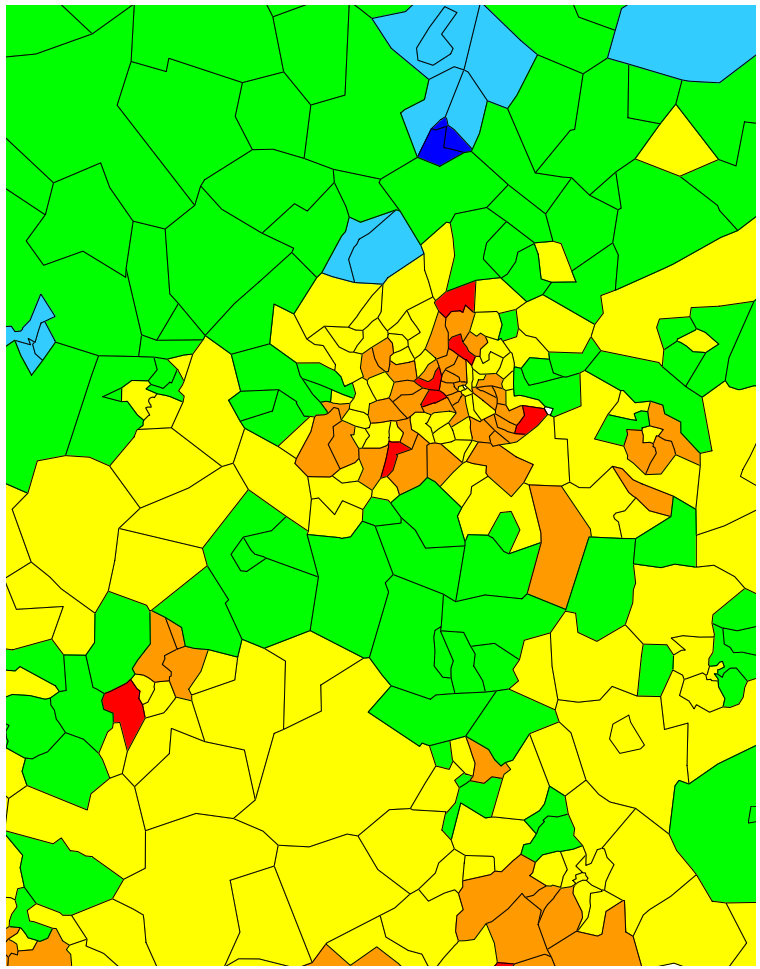


Smoothed residuals

Density not in standardizing GLM



Density in standardizing GLM



Geographical Spatial Analysis in Personal Lines Territorial Ratemaking



WWW.WATSONWYATT.COM/PRETIUM

2003 CAS Ratemaking Seminar

Duncan Anderson MA FIA

Watson Wyatt LLP

Watson Wyatt *Worldwide*