Informing Non-Response Bias Model Creation in Social Surveys with Visualization

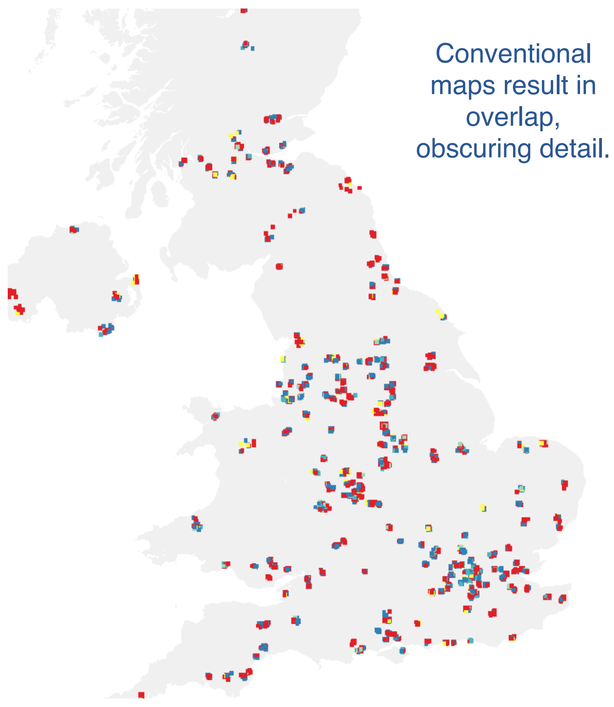
http://openaccess.city.ac.uk/12333

Kaisa Lahtinen, Sarah Butt, Rory Fitzgerald - CCSS, City University London
Jason Dykes, Aidan Slingsby - giCentre, City University London

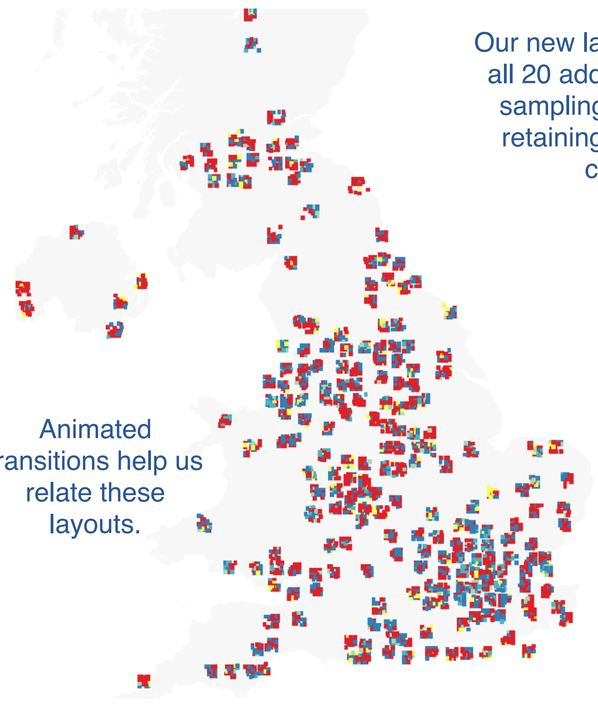
Non-response bias in social surveys is one of the most challenging issues facing social science research. We are building a model to try and predict non-response bias for the European Social Survey. Selecting variables for use in the model is difficult as the pool of variables is large, geographic a unfamiliar. A collaborative design process involving social scientists and visualization is developing solutions.

CREATING A SUITABLE LAYOUT

The 226 sets of 20 tightly clustered household locations are challenging to display in a way that both retains geographical relationship and avoids occlusion. Our carefully-designed layout preserves this two-level hierarchy, in a non-occluding manner.

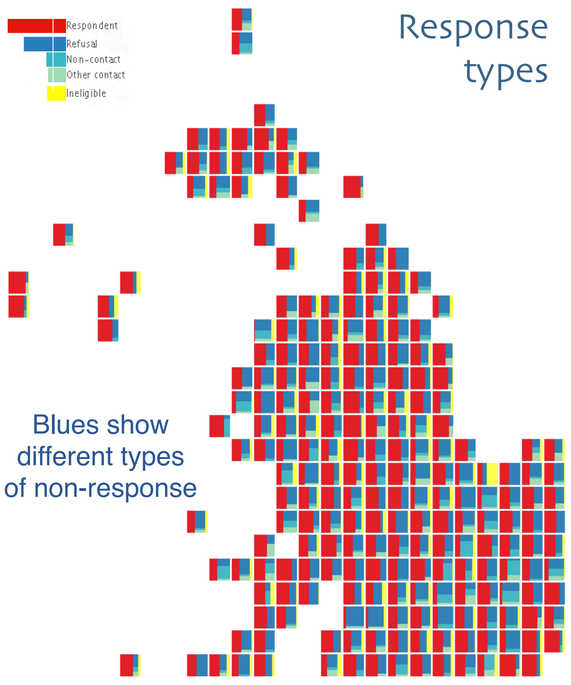


Conventional maps result in overlap, obscuring detail.



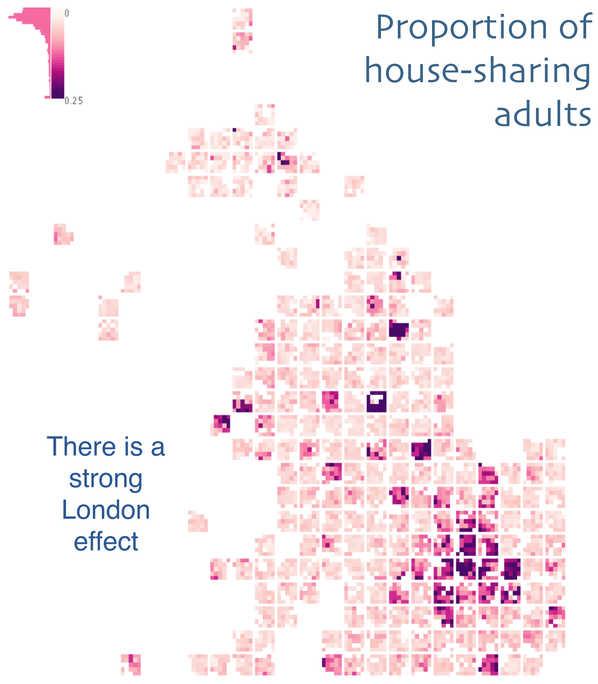
Animated transitions help us relate these layouts.

Our new layout lets us see all 20 addresses in each sampling group, whilst retaining some spatial context.



Response types

Blues show different types of non-response



Proportion of house-sharing adults

There is a strong London effect

Sports facilities within walking distance

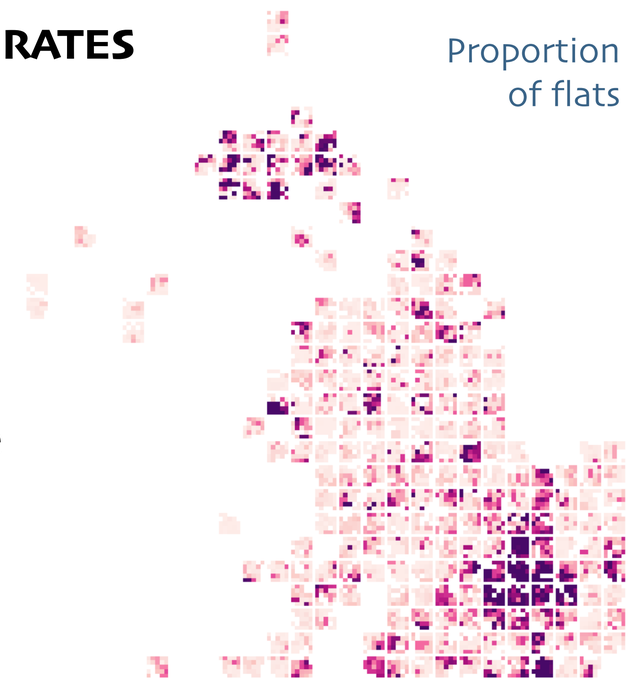
A more uniform distribution

UNDERSTANDING DATA DISTRIBUTIONS

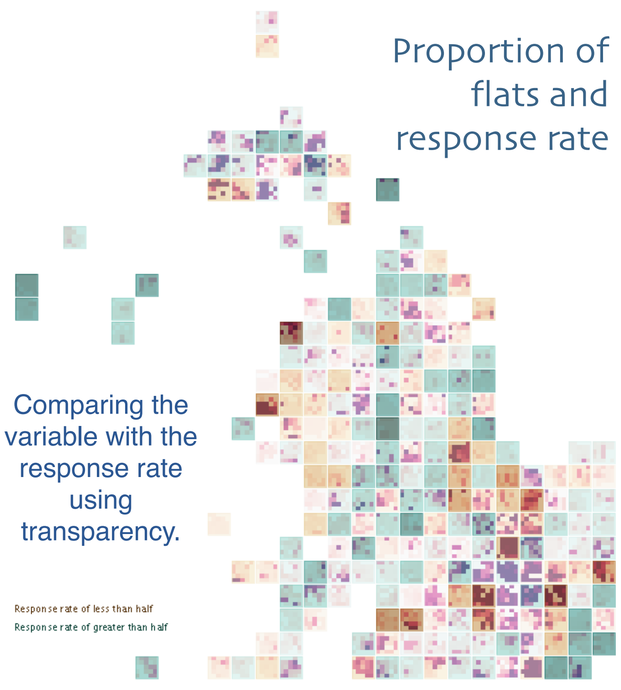
This layout enables us to quickly flick through hundreds of variables, helping us choose those variables with suitable distributions. For example, London-centric density of house-sharing adults makes this a potentially problematic variable for inclusion in a national non-response model.

COMPARING TO RESPONSE RATES

Since we are modelling non-response, we want to know how our candidate variables relate to non-response. Our diverging colour scale emphasises both low (brown) and high (turquoise) response rates. Overlaying it on our variable maps with varying degrees of transparency helps us relate variable to response as we consider local models



Proportion of flats



Comparing the variable with the response rate using transparency.

Response rate of less than half
Response rate of greater than half

Proportion of flats and response rate

Turquoise indicates high and brown indicates low response rate.

Response rate of less than half
Response rate of greater than half

Response rate

Collaborative co-design is helping us sift through hundreds of variables that might help us predict non-response, leading to a better understanding of factors affecting non-response and more informed models.

CITY UNIVERSITY LONDON
EST 1894

European Social Survey

giCentre