UNIVERSITY OF LEICESTER

A Data Exploration Exercise: Engagement with the Natural Environment

eHealth Leicester, Environmental Health and BRISSKit Workshop, 7 October 2015
Zach Collins, 3rd year Mathematics student, student member RSS
Josh Vande Hey, Physics department

What is the data?

- An open dataset from Natural England
- The data contains information about the ways that people engage with the natural environment e.g. visiting the countryside, green spaces in towns and cities, watching wildlife and volunteering to help protect the natural environment.
- The data collected about how people use the natural environment, includes the:
- type of destination
- duration
- mode of transport
- distance travelled
- expenditure
- main activities
- motivations
- barriers to visiting

Data can be found here: https://www.gov.uk/government/collections/monitor-of-engagement-with-the-natural-environment-survey-purpose-and-results

How was it collected?

- Interviewers visit people's homes and interview them about the activities they have done in the last 7 days
- Around 800 people are interviewed per week

Problems with using this data

The survey questions were mostly qualitative e.g.

2) Which of the following best describes where you spent most of your time on this visit?
SHOW SCREEN. RANDOM ORDER. SINGLE CODE.

- In a town or city
- In a seaside resort or town
- Other seaside coastline (including beaches and cliffs)
- In the countryside (including areas around towns and cities)

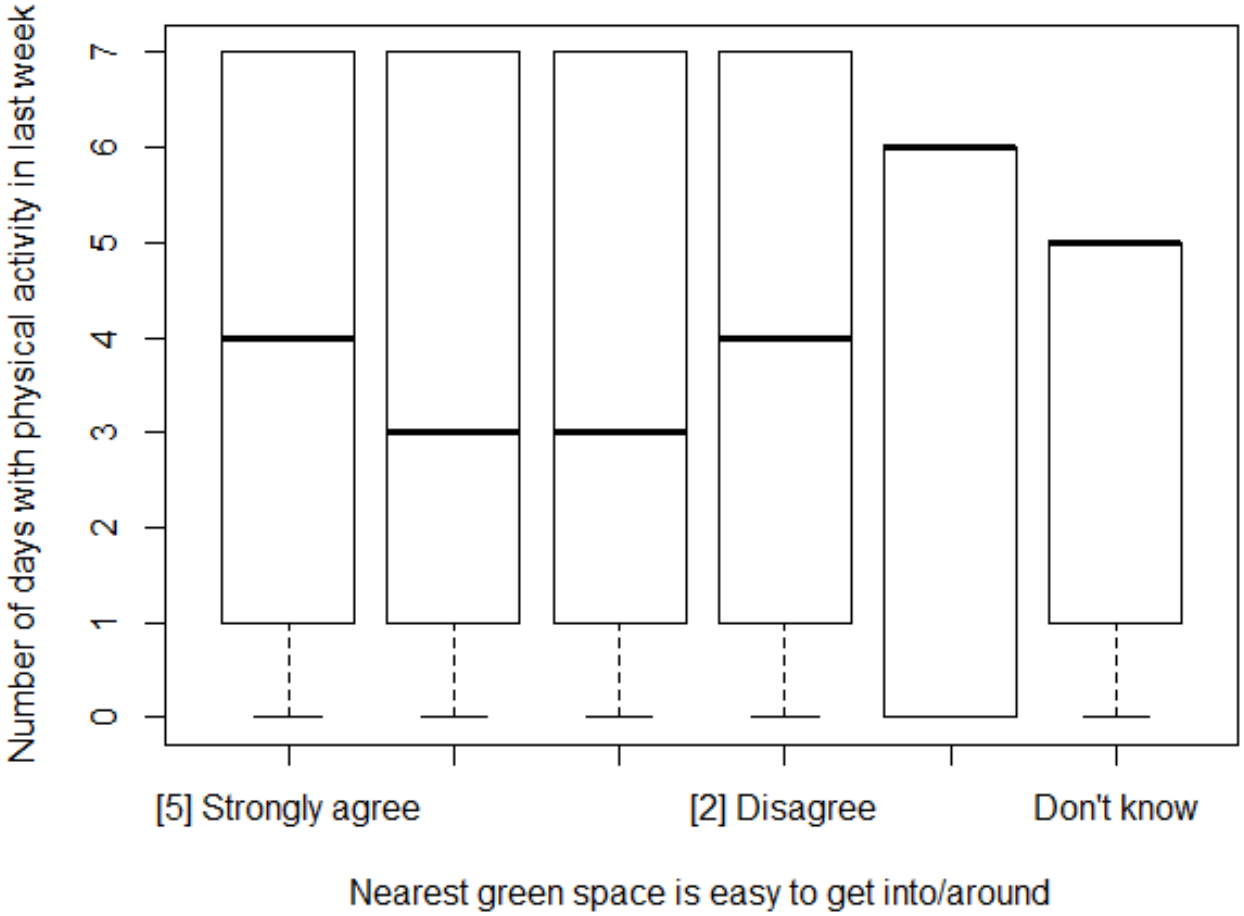
Or in ranges:

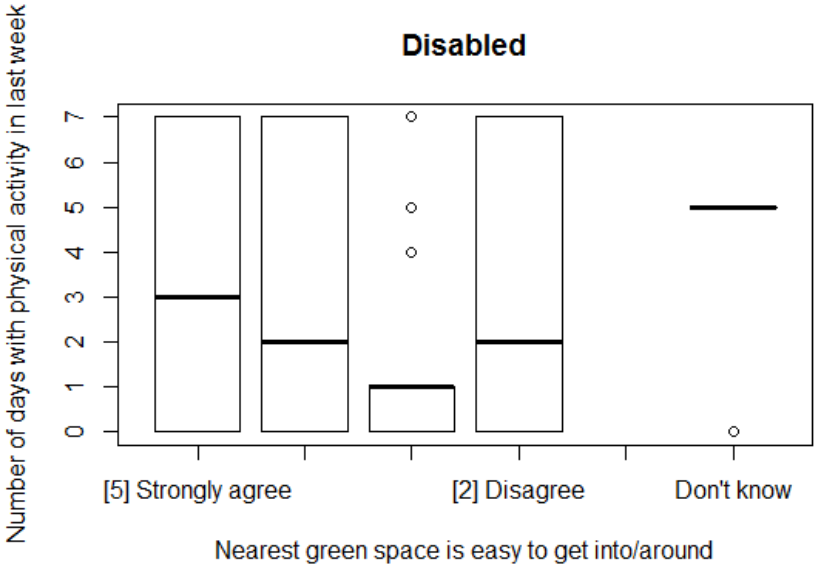
8) Approximately how far, in miles, did you travel to reach this place? By that I mean the **one way distance** from where you set off to the place visited.

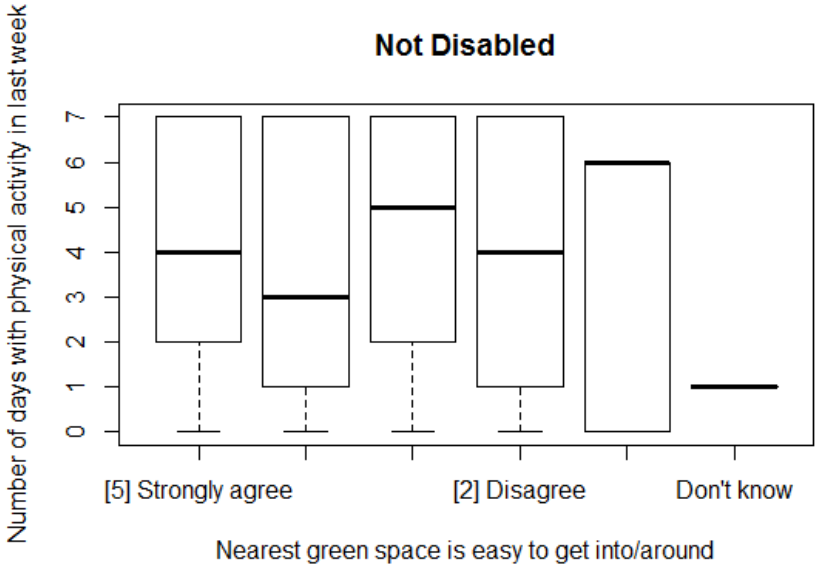
SHOW SCREEN. DO NOT RANDOMISE. SINGLE CODE.
Less than 1 mile
1 or 2 miles
3 to 5 miles
6 to 10 miles
11 to 20 miles
21to 40 miles
41to 60 miles
51to 80 miles
81to100 miles
More than 100 miles

This limits what can be done with the data.

Tried finding a correlation between some usable variables, didn't draw any conclusions



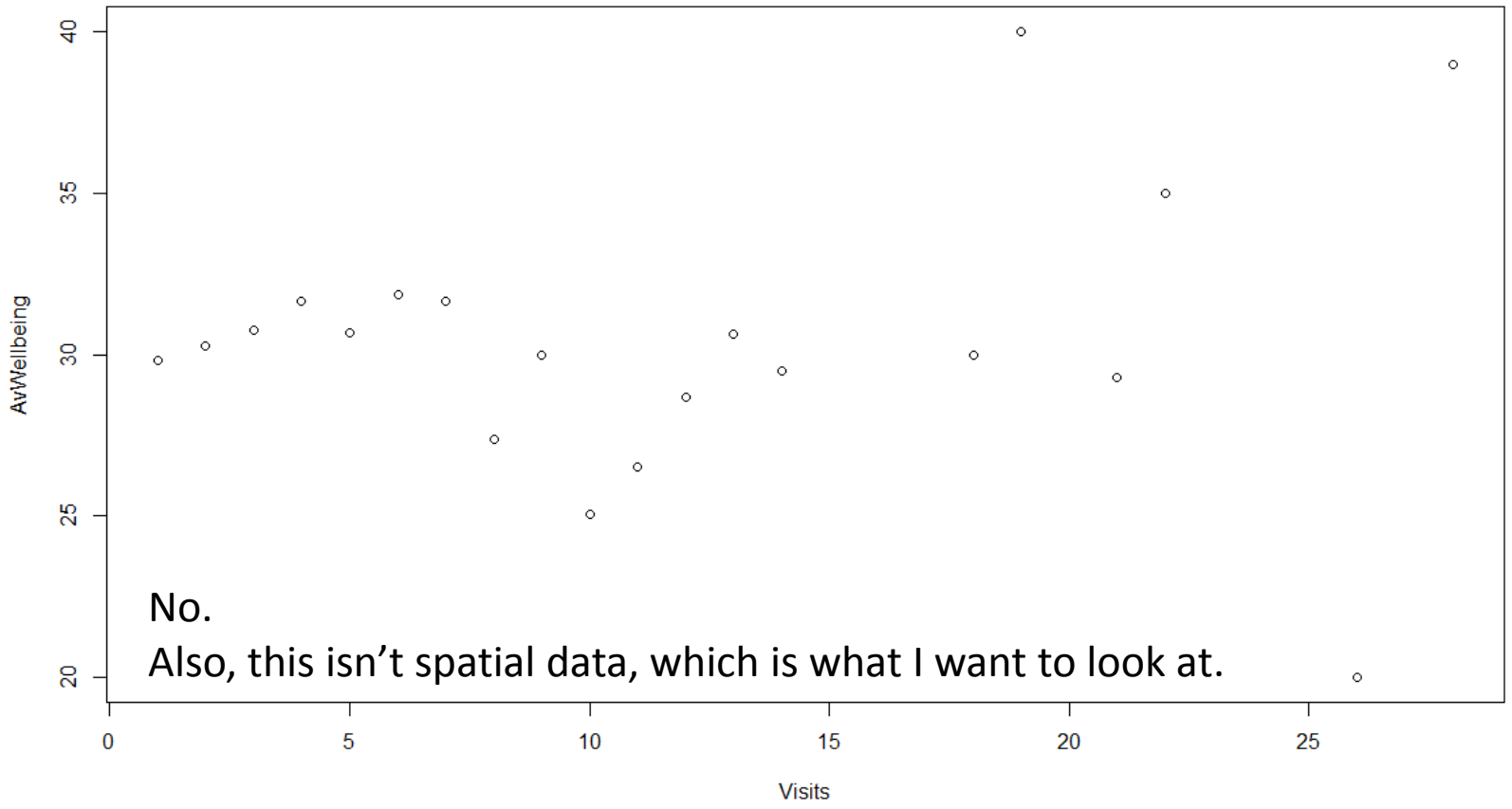




Still no conclusions, nothing interesting, not much data for disabled people.

Is there correlation between number of visits and wellbeing?

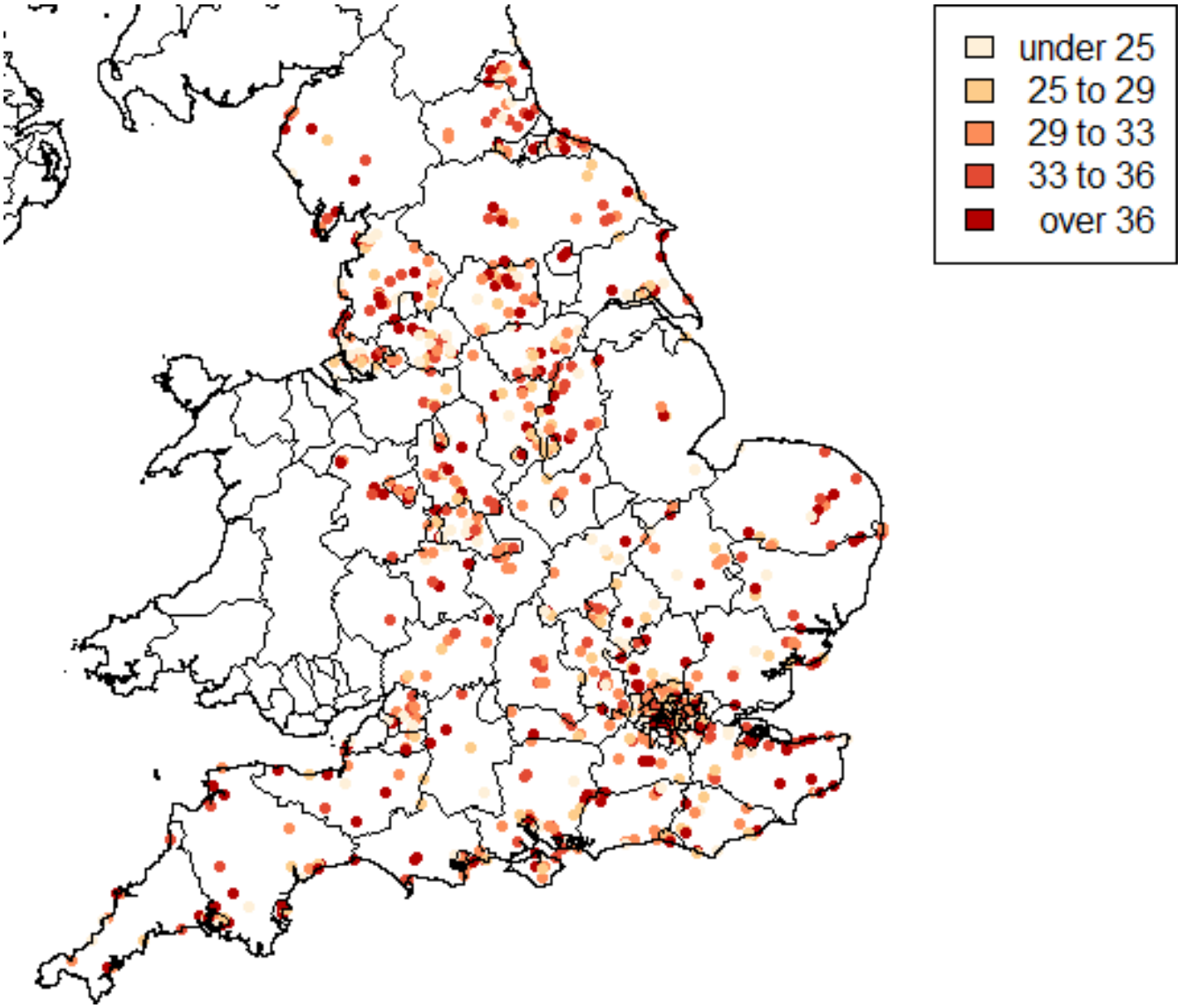
People were asked to rate on a scale of 1 to 10 how happy/worthwile/ satisfied/anxious they felt the previous day. Added the 4 scores up to give a total out of 40.



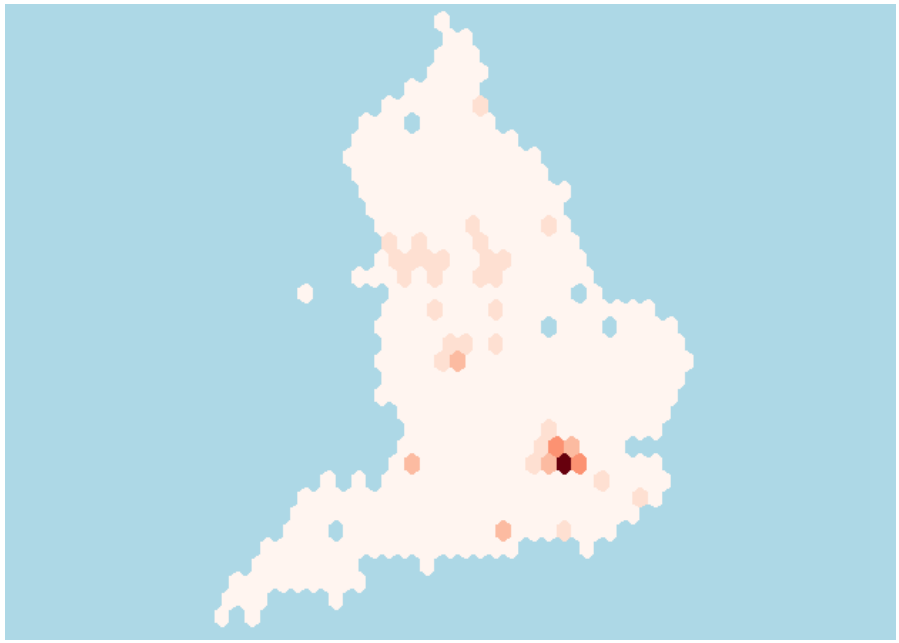
No.

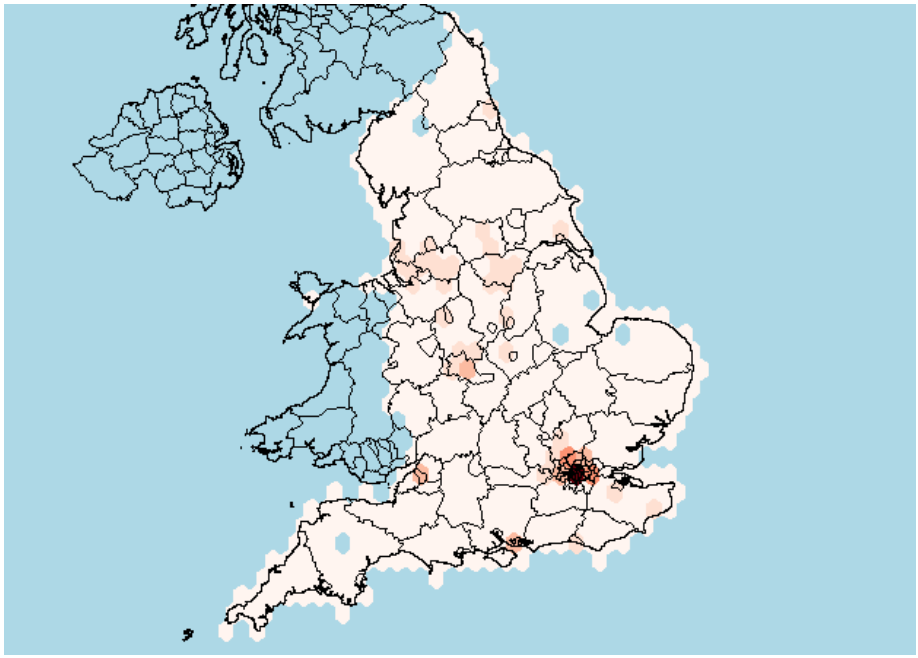
Also, this isn't spatial data, which is what I want to look at.

Wellbeing scores by location



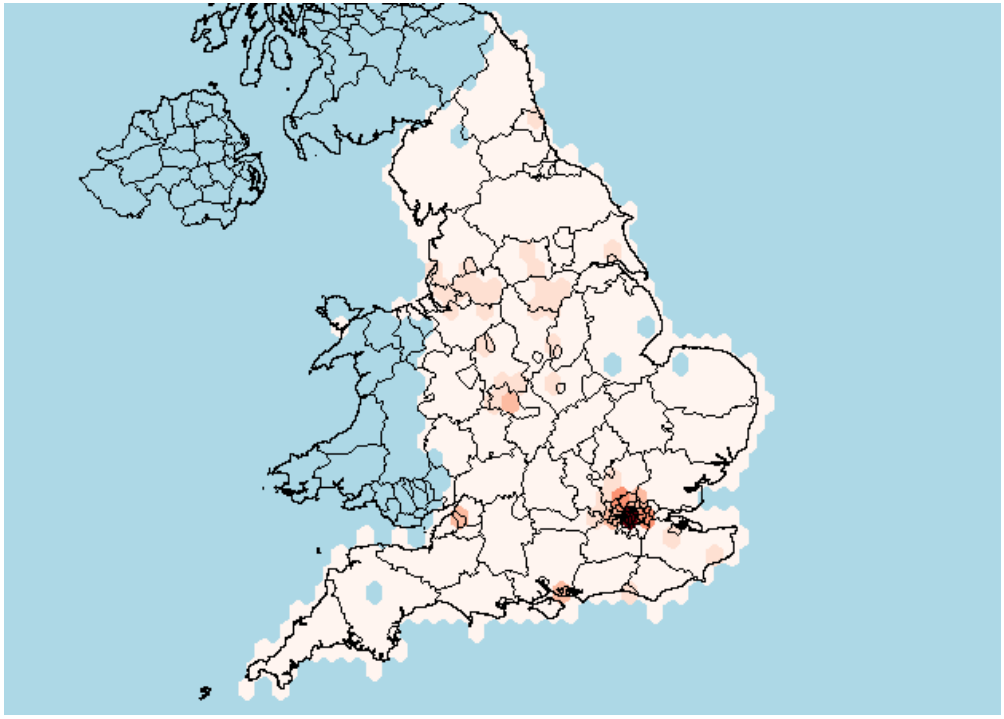
Most expensive destinations

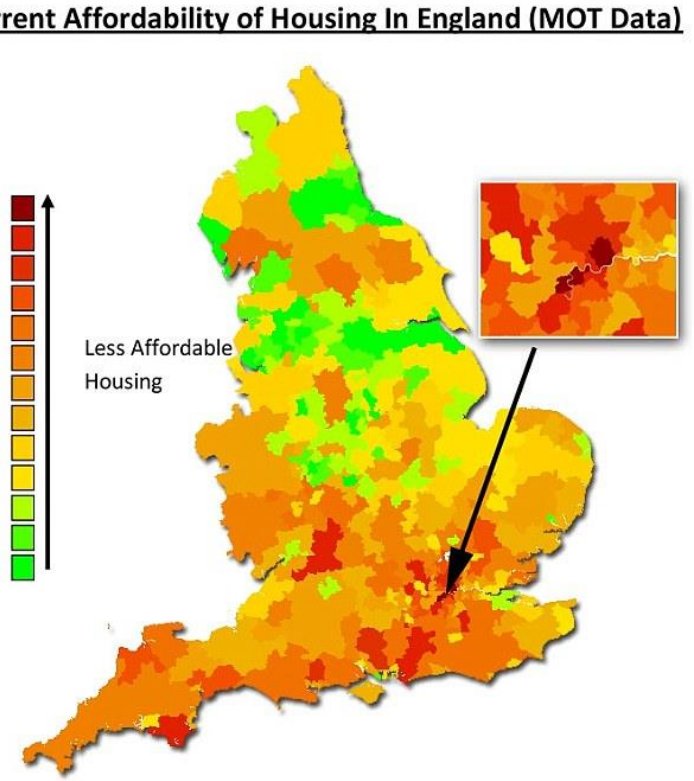




UK outline and counties from http://www.gadm.org/download

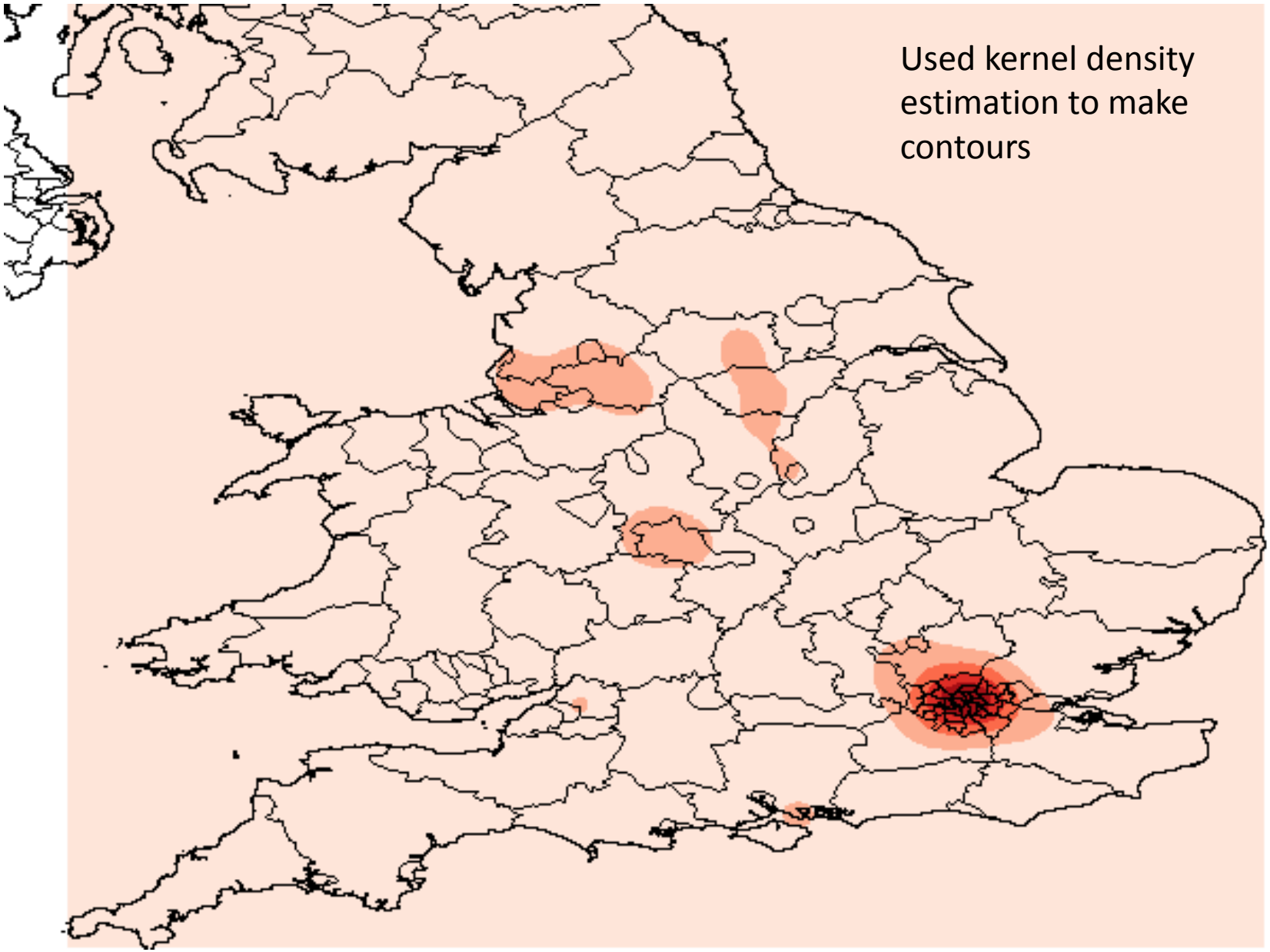
Compare with property price map as measure of wealth of area





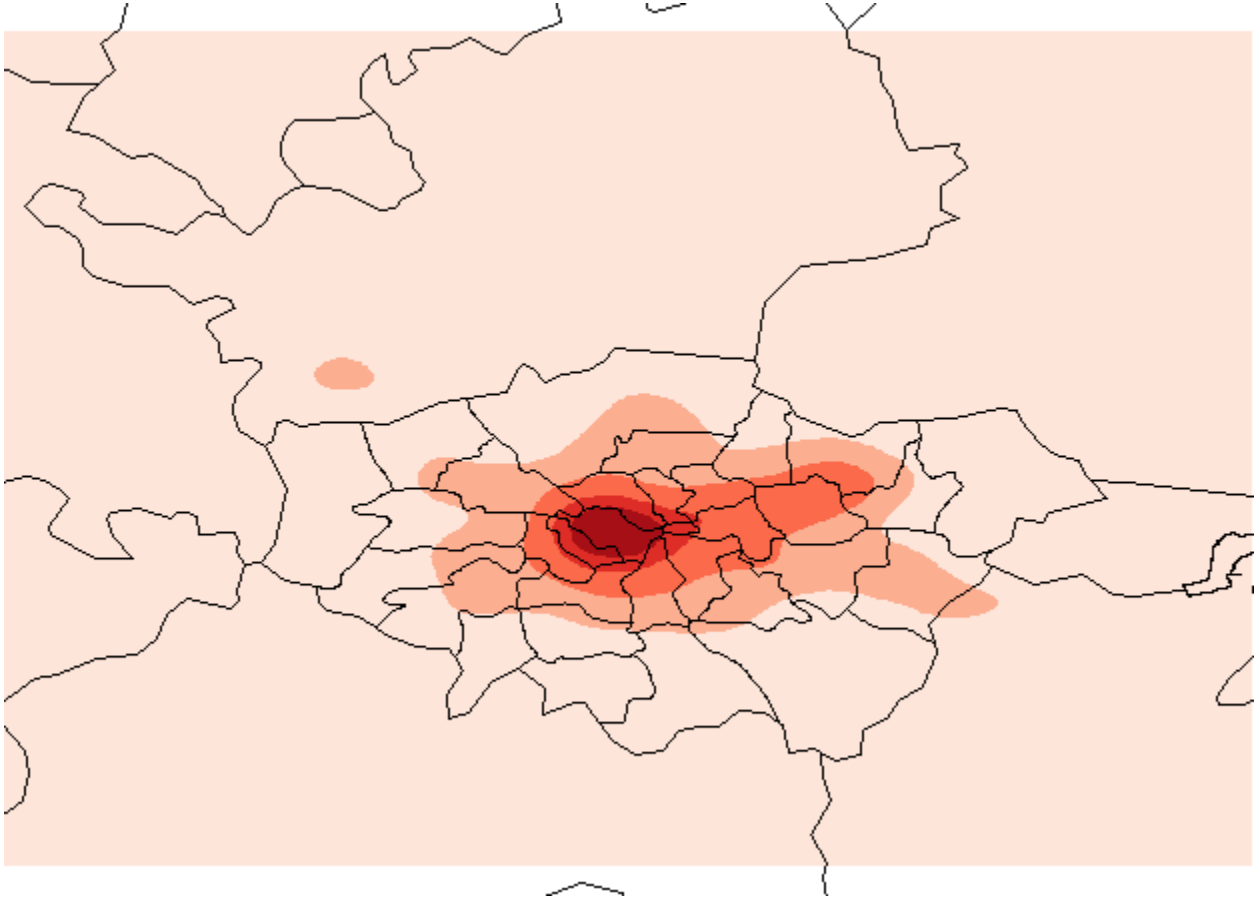
http://www.thisismoney.co.uk/money/mortgageshome/article-2295295/Property-affordability-map-Most-reasonable-house-prices-England.html 2013

Location of visits to Natural Environment



Zoom in over London area

Default h_x and h_y: `[1] 0.1639311 0.0829882`



How is default value of h calculated?

```
> kde.points
function (pts, h, n = 200, lims = NULL)
{
    xy = coordinates(pts)
    p4s = CRS(proj4string(pts))
    if (missing(h))
        h = c(bandwidth.nrd(xy[, 1]), bandwidth.nrd(xy[, 2]))
    if (is.null(lims)) {
        lims = c(range(xy[, 1]), range(xy[, 2]))
    }
    else {
        lims = t(bbox(lims))
    }
    kd = kde2d(xy[, 1], xy[, 2], h, n, lims)
    temp = SpatialPoints(expand.grid(kd$x, kd$y))
    temp = SpatialPixelsDataFrame(temp, data.frame(kde = array(kd$z,
        length(kd$z))))
    proj4string(temp) = p4s
    temp
}
<environment: namespace:GISTools>
```

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\hat{\sigma}n^{-1/5}$$

$\hat{\sigma}$= standard deviation of samples

```
> bandwidth.nrd
function (x)
{
    r <- quantile(x, c(0.25, 0.75))
    h <- (r[2L] - r[1L])/1.34
    4 * 1.06 * min(sqrt(var(x)), h) * length(x)^(-1/5)
}
<bytecode: 0x02f763d8>
<environment: namespace:MASS>
```

https://en.wikipedia.org/wiki/Kernel_density_estimation#Bandwidth_selection

Function that makes it easier to change h

```
> kde.points
function (pts, h, n = 200, lims = NULL)
```

Can already change h here, but don't know what automatic h is as a number, so it's difficult to know if you're increasing or decreasing it without calculating it first

```
#Function to create KDE contour plot
#x=data frame
#f=factor of automatically generated h bandwidths
contour.kde<-function(x,f){
  h<-c(bandwidth.nrd(coordinates(x)[, 1]), bandwidth.nrd(coordinates(x)[, 2]))
  data.dens<-kde.points(x,f*h)
  level.plot(data.dens)
  return(data.dens)
}
```

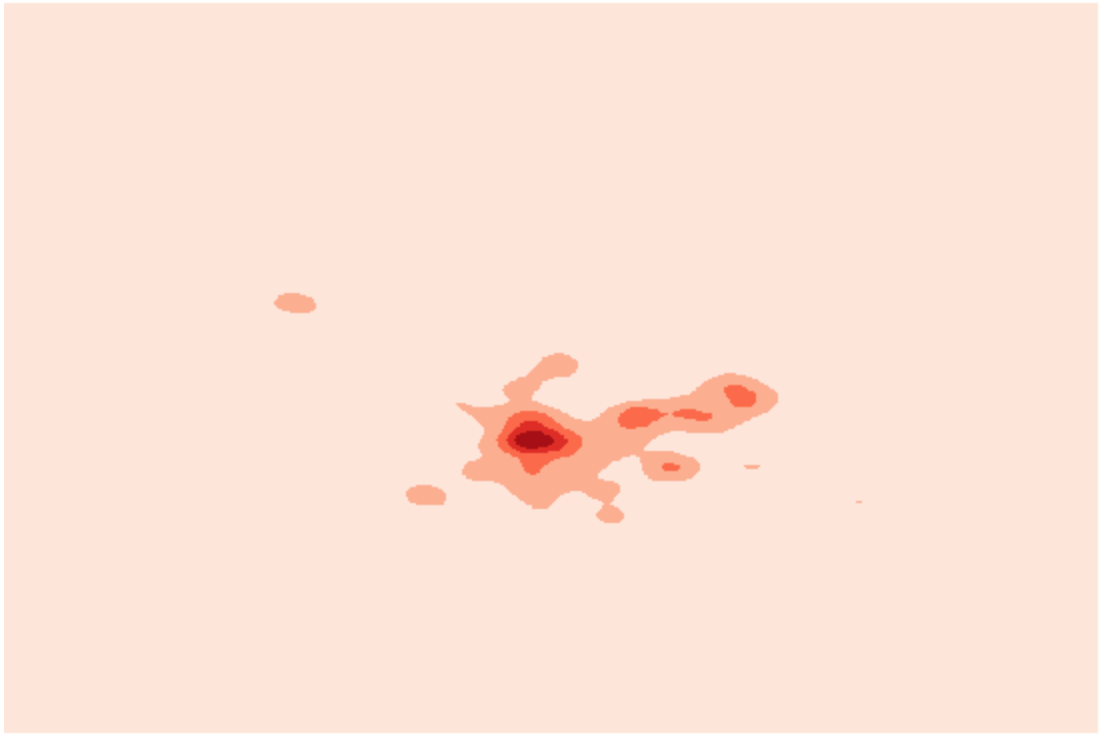
Function allows you to define a factor to multiply h by. Also plots automatically

Different visualizations of the data

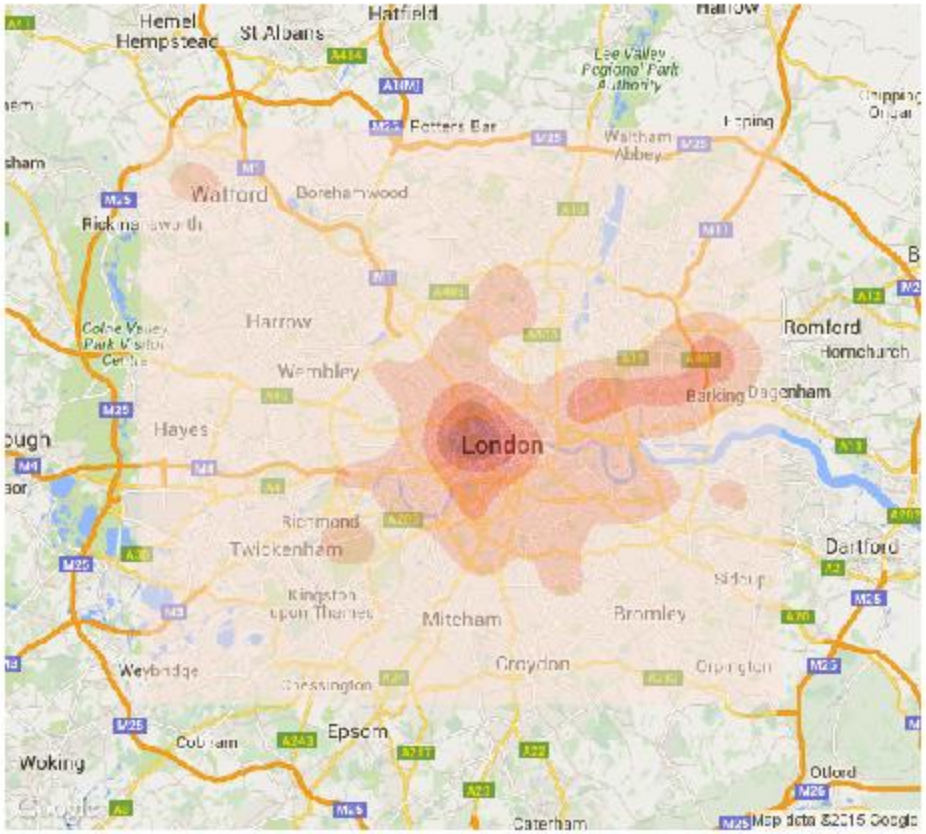
- Using different fractional values of default h

2h

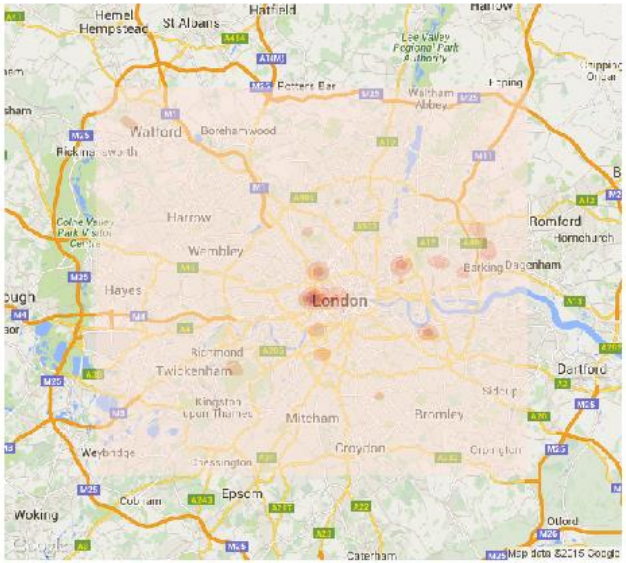
0.5h



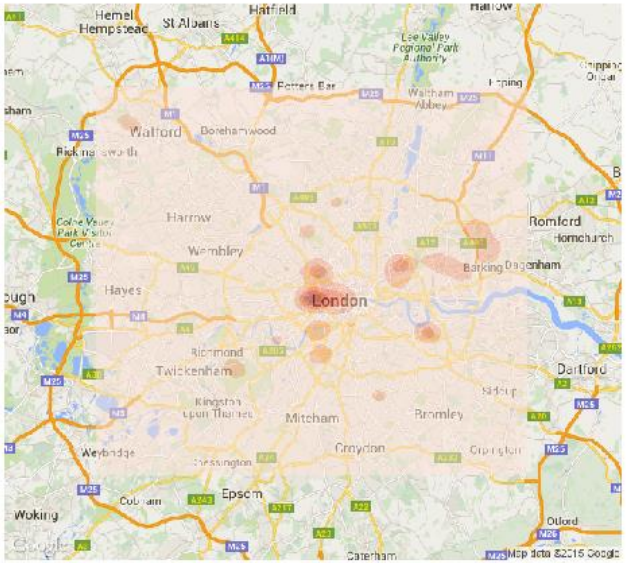
Zoomed in further over London automatic h value



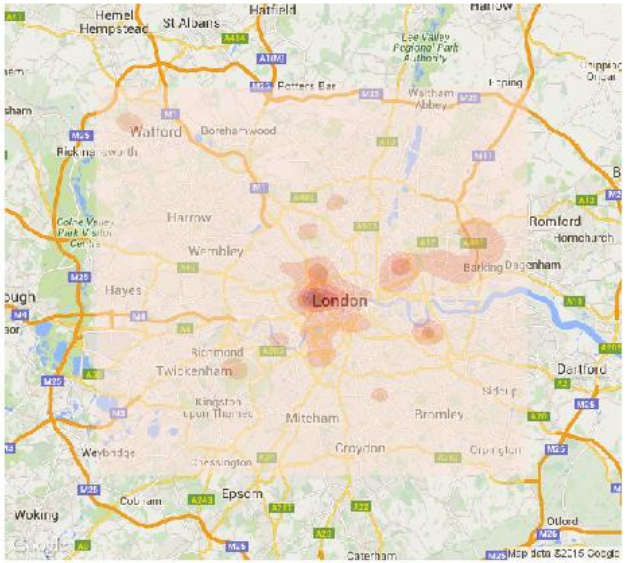
0.4h


Hemel Hempstead
St Albans
Hatfield
Harrow
Lee Valley Regional Park Authority
Potters Bar
Waltham Abbey
Epping
Watford
Borehamwood
Rickmansworth
Harrow
Wembley
Romford
Hornchurch
Barking
Dagenham
Hayes
London
Richmond
Twickenham
Dartford
Kingston upon Thames
Mitcham
Bromley
Sidcup
Weybridge
Chessington
Croydon
Orpington
Epsom
Cobham
Woking
Oxted
Caterham
Map data ©2015 Google

0.5h


Hemel Hempstead
St Albans
Hatfield
Harrow
Lee Valley Regional Park Authority
Potters Bar
Waltham Abbey
Epping
Watford
Borehamwood
Harrow
Wembley
Romford
Hornchurch
Barking
Dagenham
Hayes
London
Richmond
Twickenham
Dartford
Kingston upon Thames
Mitcham
Bromley
Sidcup
Weybridge
Chessington
Croydon
Orpington
Epsom
Cobham
Woking
Oxted
Caterham
Map data ©2015 Google

0.6h


Hemel Hempstead
St Albans
Hatfield
Harrow
Lee Valley Regional Park Authority
Potters Bar
Waltham Abbey
Epping
Watford
Borehamwood
Harrow
Wembley
Romford
Hornchurch
Barking
Dagenham
Hayes
London
Richmond
Twickenham
Dartford
Kingston upon Thames
Mitcham
Bromley
Sidcup
Weybridge
Chessington
Croydon
Orpington
Epsom
Cobham
Woking
Oxted
Caterham
Map data ©2015 Google

0.5h

Hampstead Heath
Victoria Park & Queen Elizabeth Olympic Park
The Regent’s Park
Green Park
Hyde Park
Battersea Park
Richmond Park
Clapham Common
Greenwich Park
London
Watford
Hayes
Richmond
Croydon

“Conclusions”

- Not surprisingly, most of the visits were around park areas
- On first look, the information content in the data appears to be limited
- A finer scale analysis of the data from London might allow for a more subtle understanding of how people interact with the natural environment in urban areas.