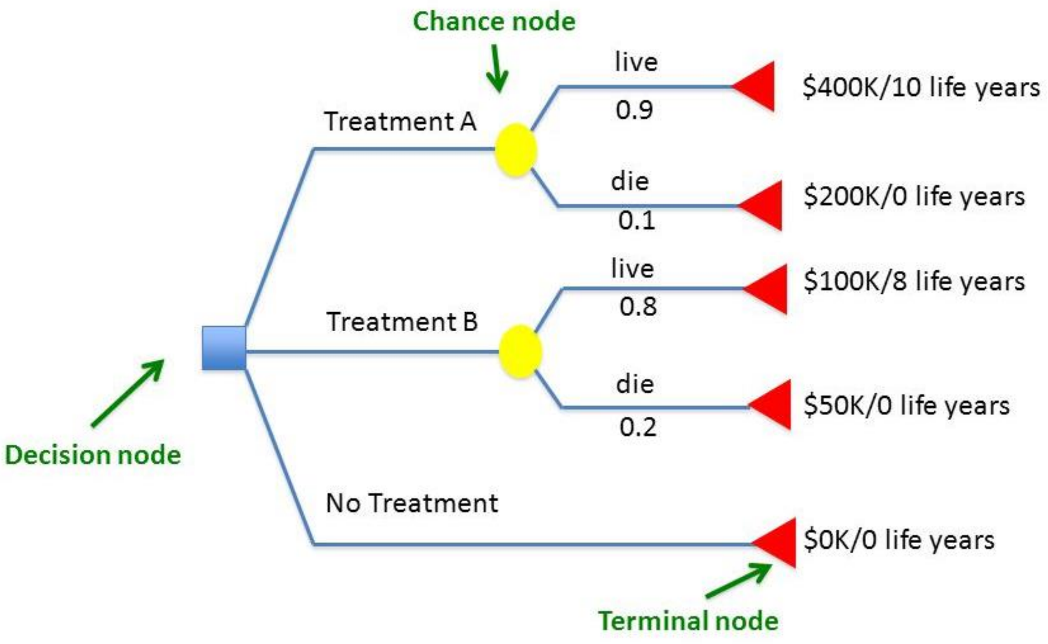
R for Decision Science

Robert Smith
Public Health Economics & Decision Science
ScHARR

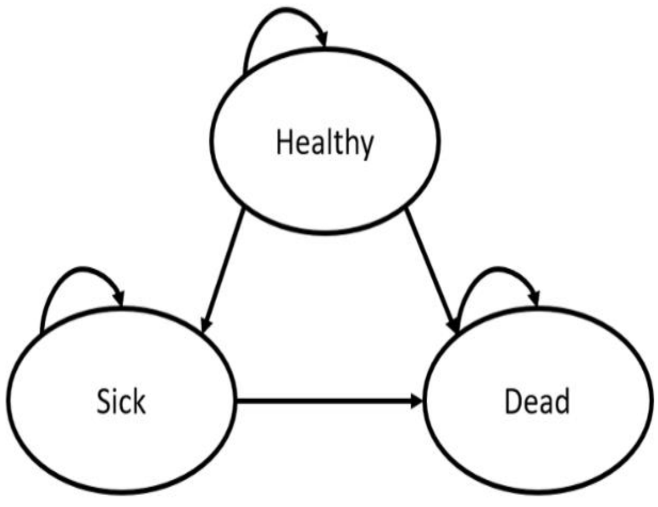
Red Deer Pub Presentation

Decision Tree

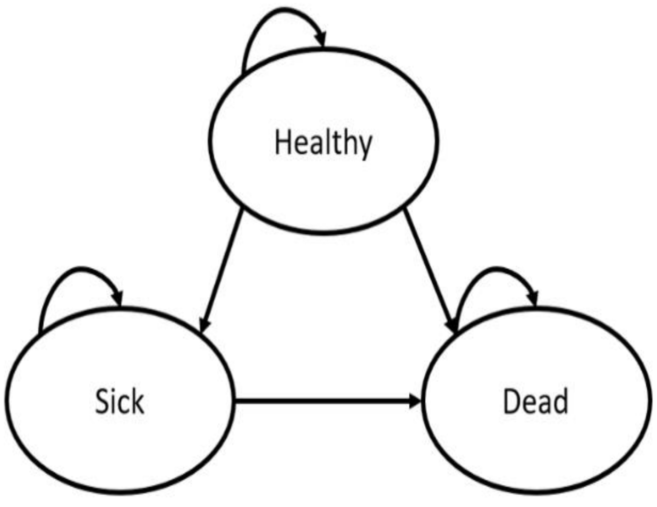


Markov Model

Baseline



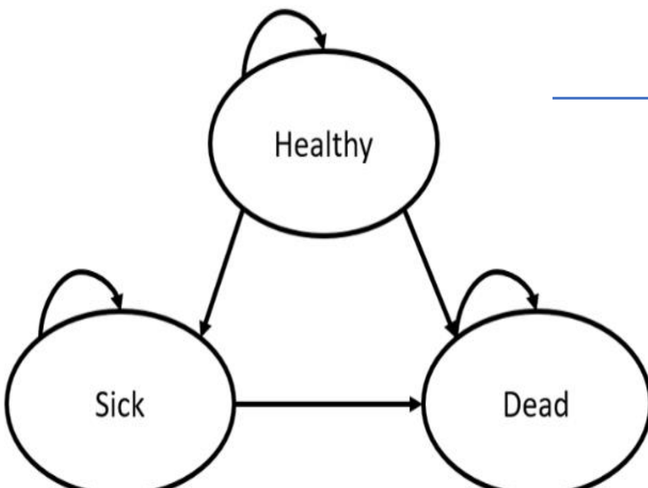
Intervention



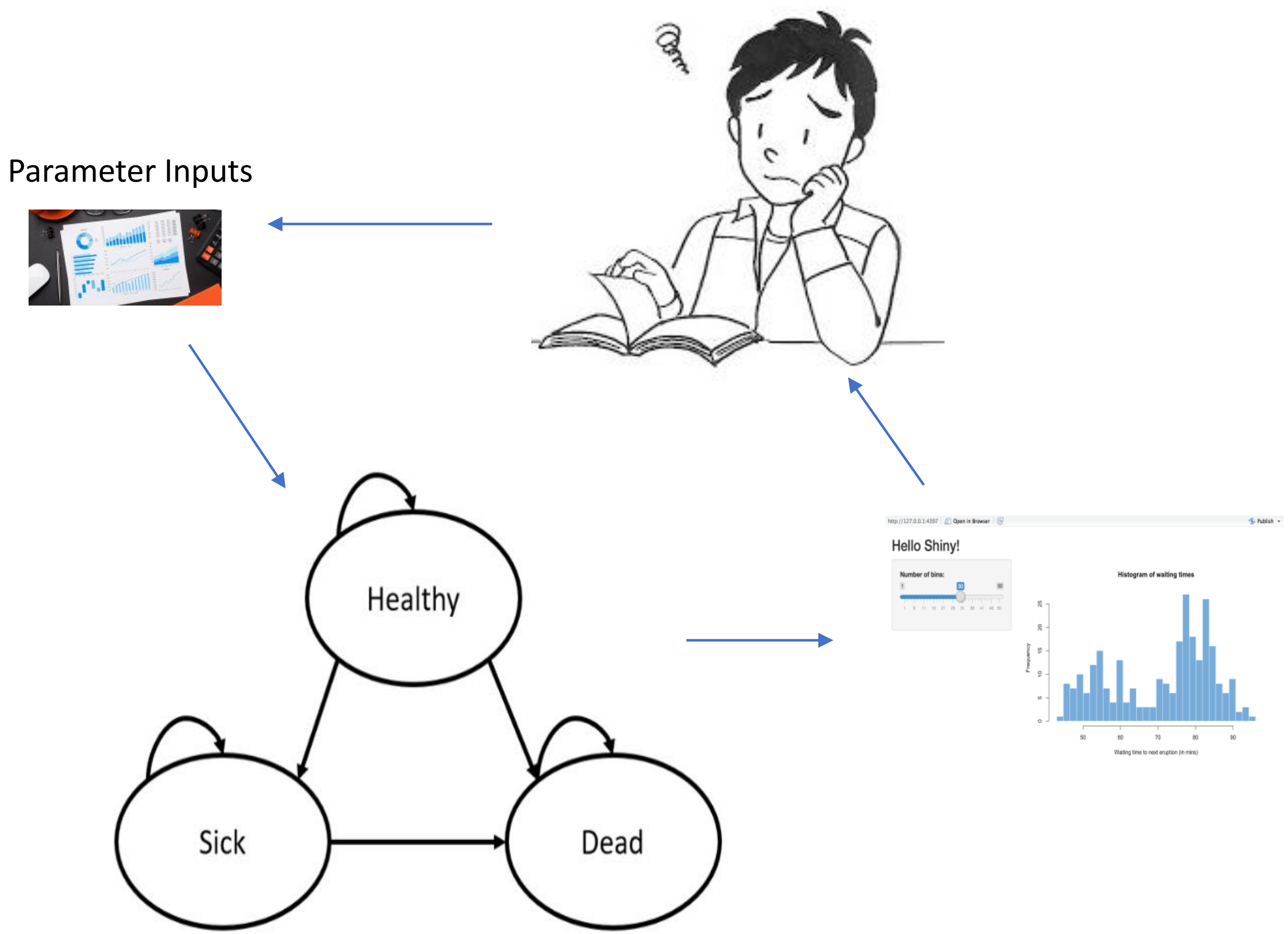
Decision Making Process

Parameter Inputs

Base: P(HS) = 0.15
Int: P(HS) = 0.05



Decision Making Process



Examples

Parkrun – Location of New Events

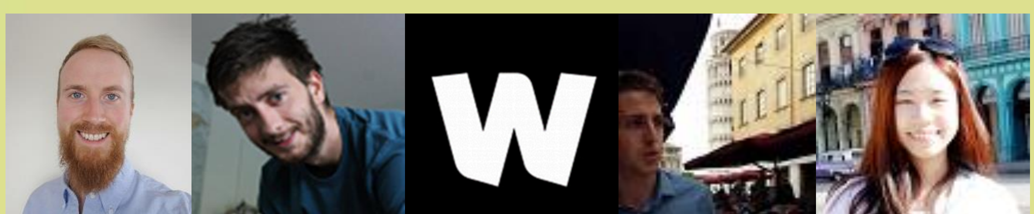
Mystery – To be revealed

Cystic Fibrosis Adherence – Markov Model

parkrun Access, Participation and Equity

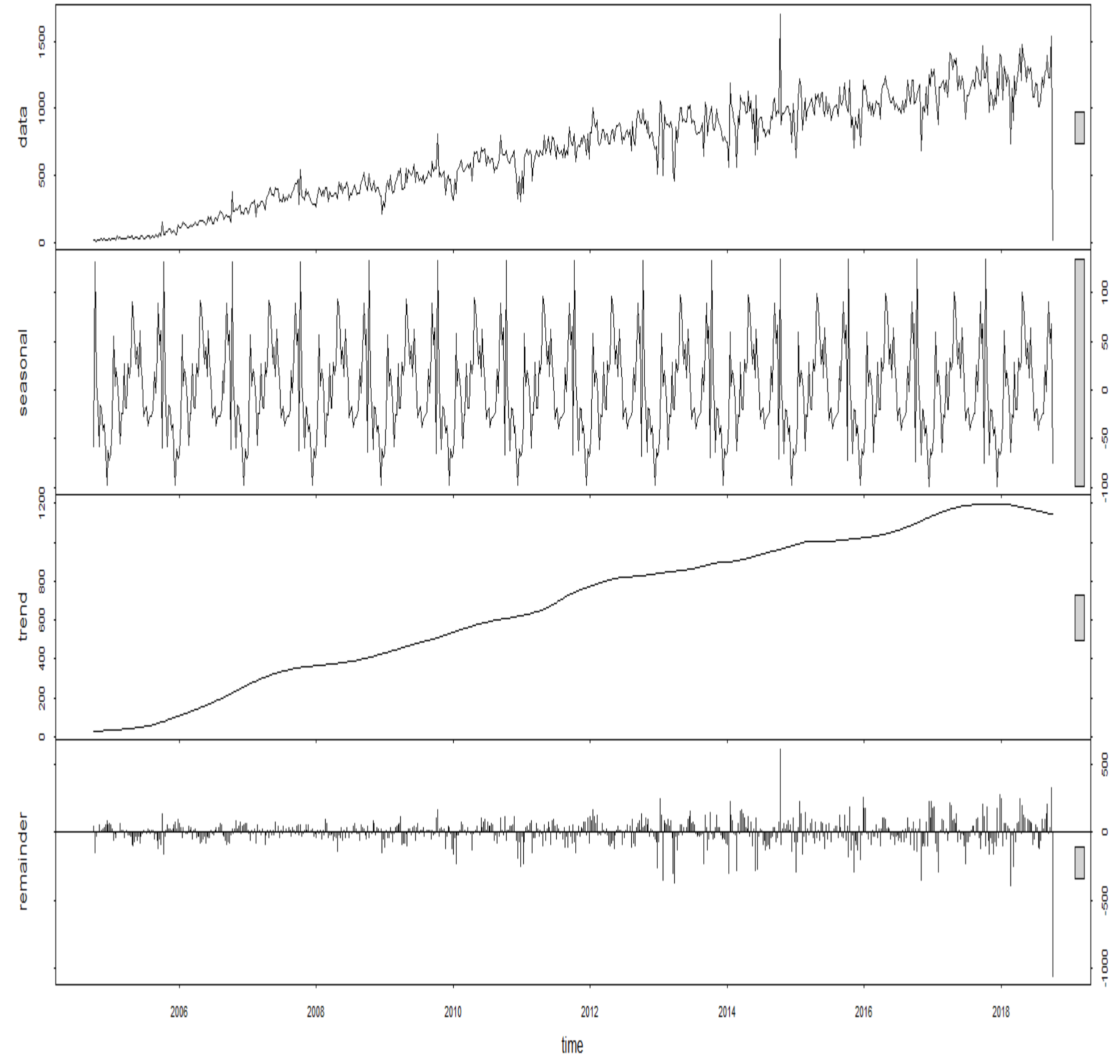
Robert Smith

Public Health Economics & Decision Science, ScHARR

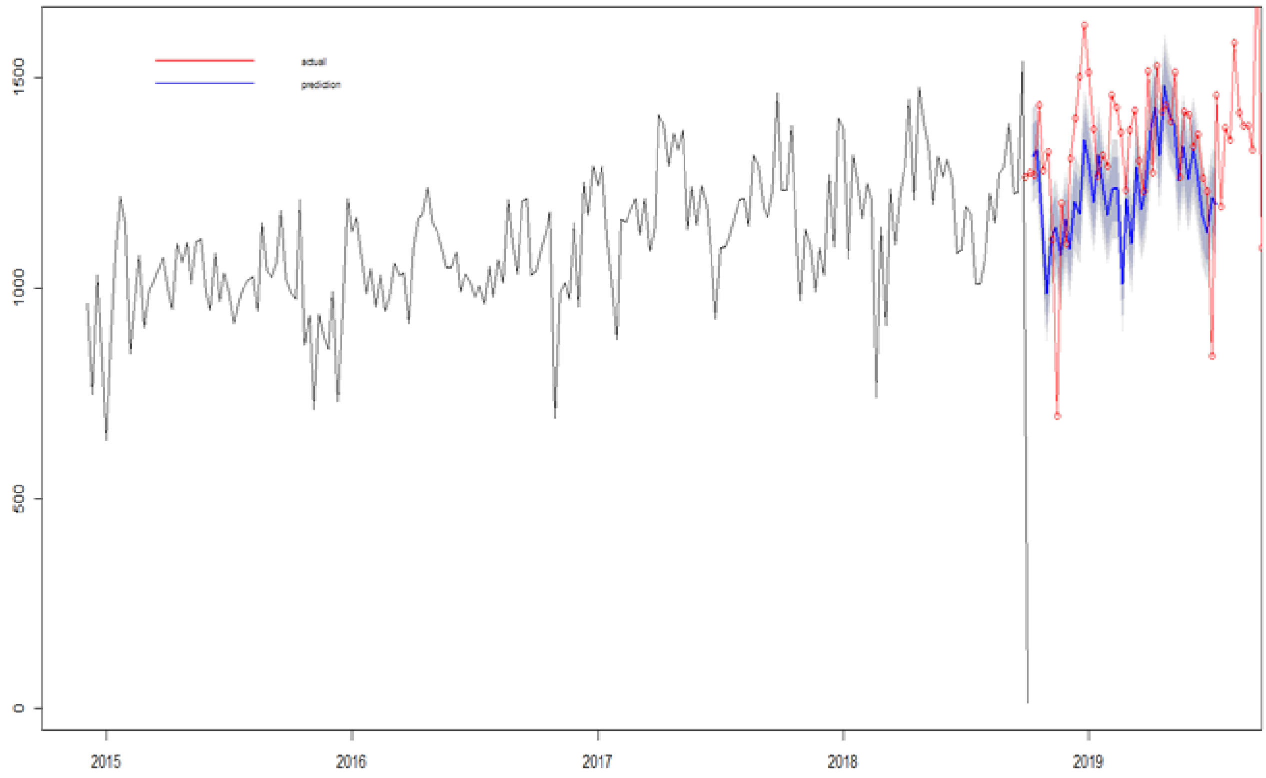


02/10/2004

Trends at Bushy Parkrun



Predicting Attendance at Bushy Parkrun



Background

PARTNERSHIP WITH PARKRUN WORTH £3M

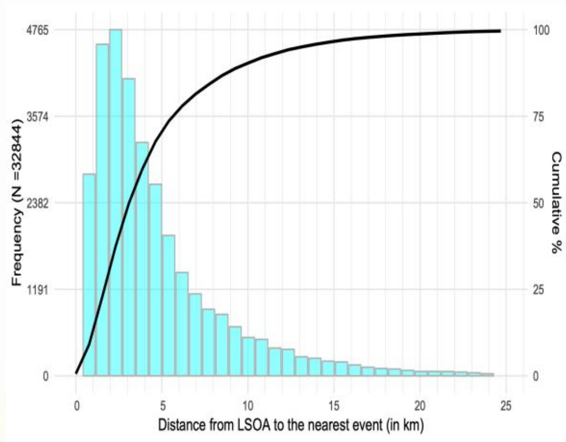
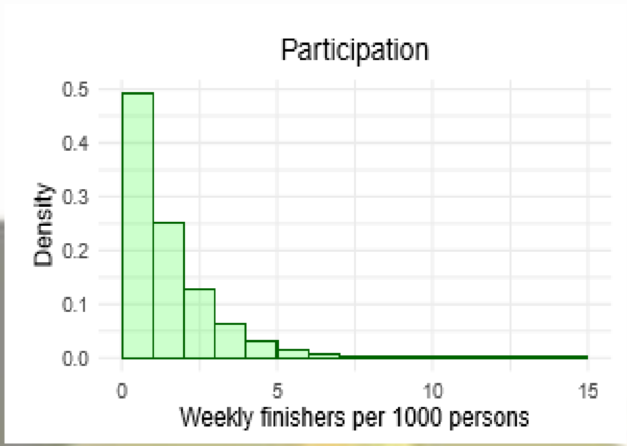
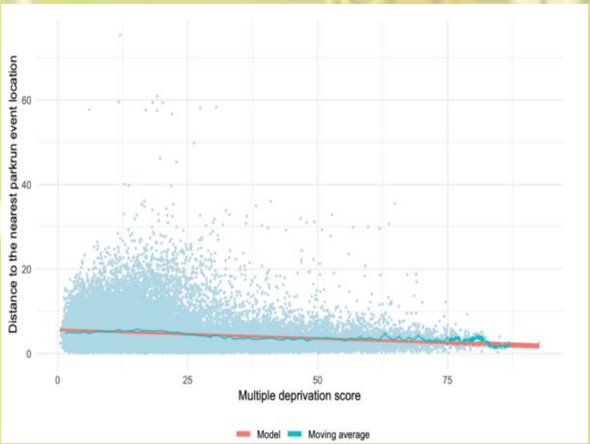
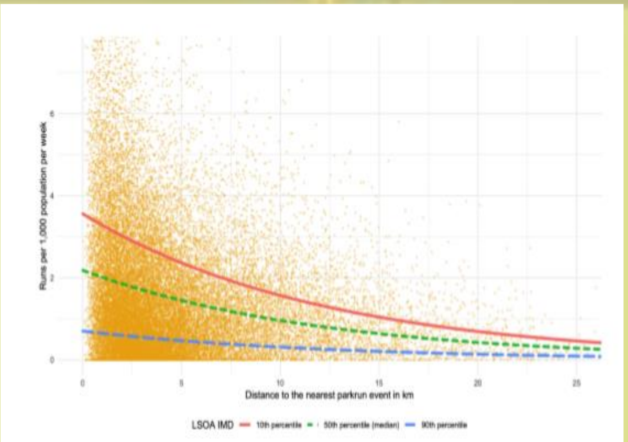
Collaboration aims to create 200 new events and boost participants from under-represented groups

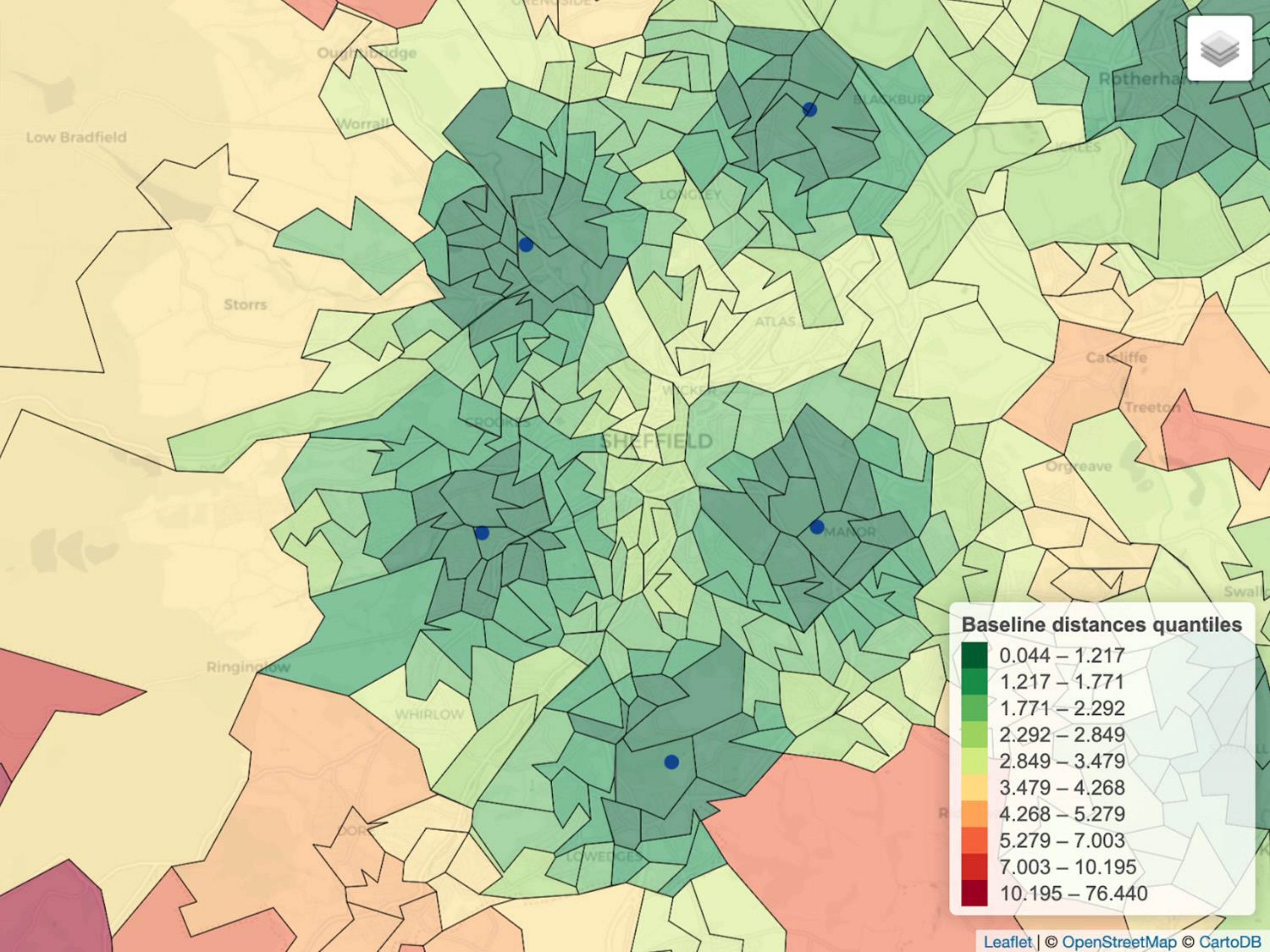
12 December 2018 News Funding

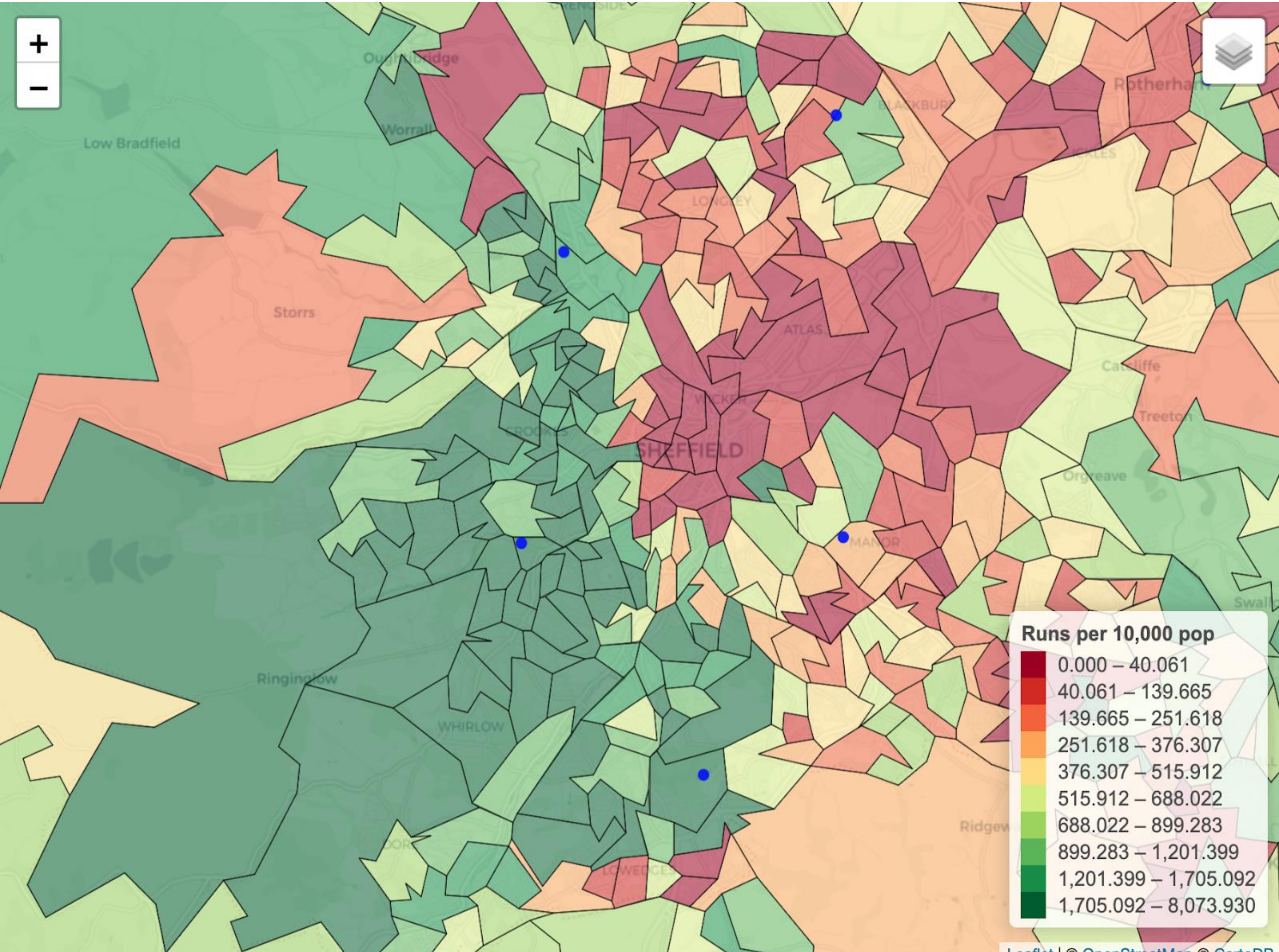
Research Questions

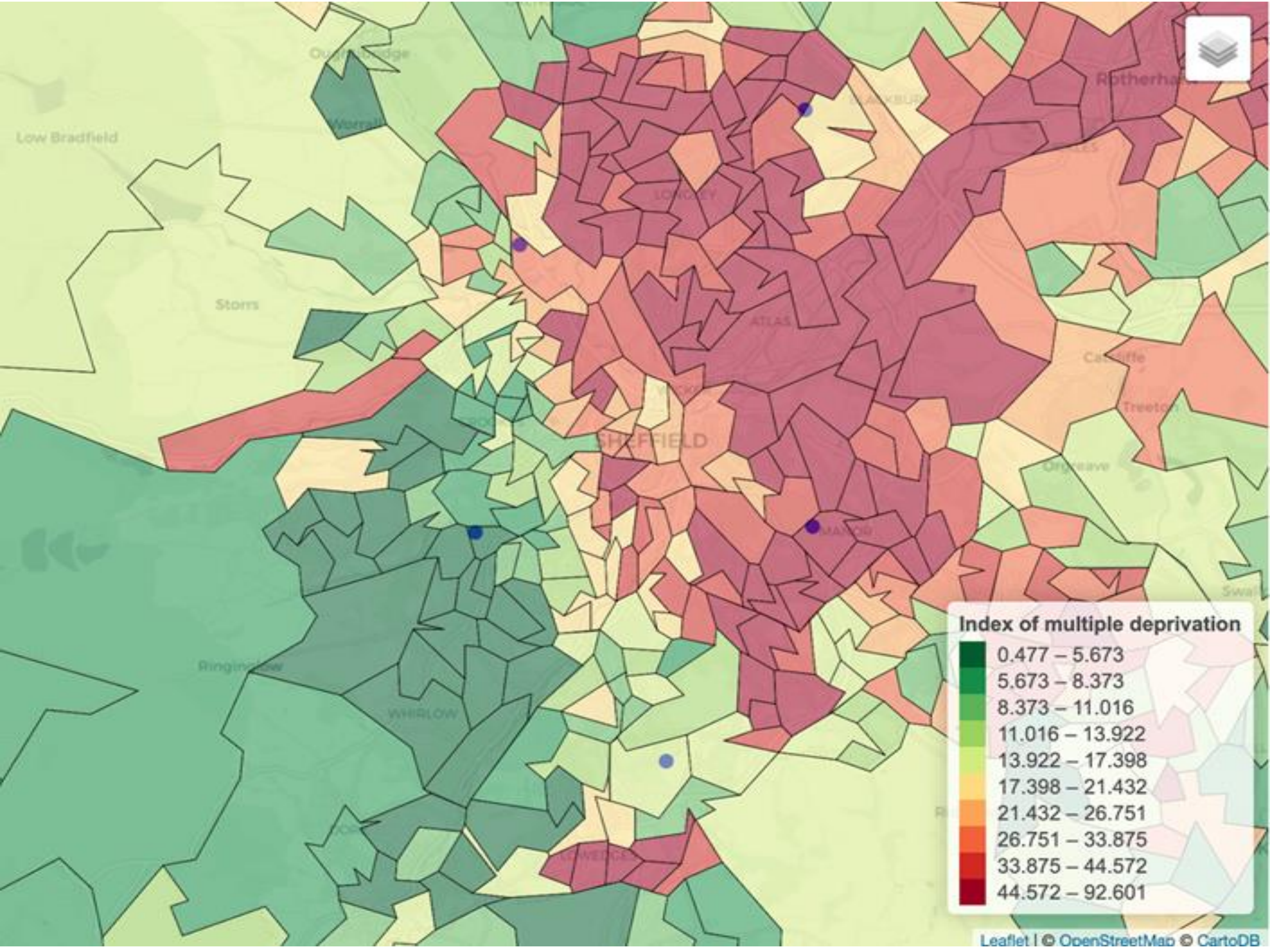
Where should parkrun locate 200 new parkrun events?

	Access	Participation
Efficiency	1. Maximize overall access.	3. Maximize overall participation.
Equity	2. Maximize deprivation weighted access.	4. Maximize deprivation weighted participation.

	Access	Participation
Efficiency		
Equity		


Oughtibridge
Worrall
Low Bradfield
Storrs
LONGLEY
ATLAS
Rotherham
Catcliffe
Treeton
Orgreave
SHEFFIELD
CROOKES
MANOR
Ringinglow
WHIRLOW
LOWEDGES
Baseline distances quantiles
0.044 – 1.217
1.217 – 1.771
1.771 – 2.292
2.292 – 2.849
2.849 – 3.479
3.479 – 4.268
4.268 – 5.279
5.279 – 7.003
7.003 – 10.195
10.195 – 76.440
Leaflet | © OpenStreetMap © CartoDB


Oughtibridge
Worrall
Low Bradfield
Storrs
Rotherham
BLACKBURN
LONGLEY
ATLAS
Catcliffe
WICKER
SHEFFIELD
CROOKES
Treeton
Orgreave
MANOR
Ringinglow
WHIRLOW
LOWEDGES
Ridgeway
Runs per 10,000 pop
0.000 – 40.061
40.061 – 139.665
139.665 – 251.618
251.618 – 376.307
376.307 – 515.912
515.912 – 688.022
688.022 – 899.283
899.283 – 1,201.399
1,201.399 – 1,705.092
1,705.092 – 8,073.930


Low Bradfield
Storrs
SHEFFIELD
Ringinglow
WHIRLOW
Catcliffe
Treeton
Orgreave
Index of multiple deprivation
0.477 – 5.673
5.673 – 8.373
8.373 – 11.016
11.016 – 13.922
13.922 – 17.398
17.398 – 21.432
21.432 – 26.751
26.751 – 33.875
33.875 – 44.572
44.572 – 92.601
Leaflet | © OpenStreetMap © CartoDB

Methods

More formally, we define that for any candidate green space location j, the objective function $f(j|E)$ provides the sum of parkrun runs r_i over all LSOA i, weighted by the squared IMD score w_i^2, given the set of established parkrun event locations $E = \{e_1, e_2, ..., e_{455}\}$:

$$f(j|E) = \sum_{i=1}^{32844} w_i^2 * r_{ij}$$

In the absence of causal estimates, we use the Poisson regression model specified above to predict the expected number of runs r_{ij} for LSOA i based on its IMD score w_i, its (linear) distance to the nearest parkrun event d_{ij}, and its population p_i. The functional form is given below.

$$E(r_{ij}|w_i, d_{ij}, p_i) = \exp(\beta_0 + \beta_1 * w_i + \beta_2 * d_{ij} + \ln(p_i) + \in)$$

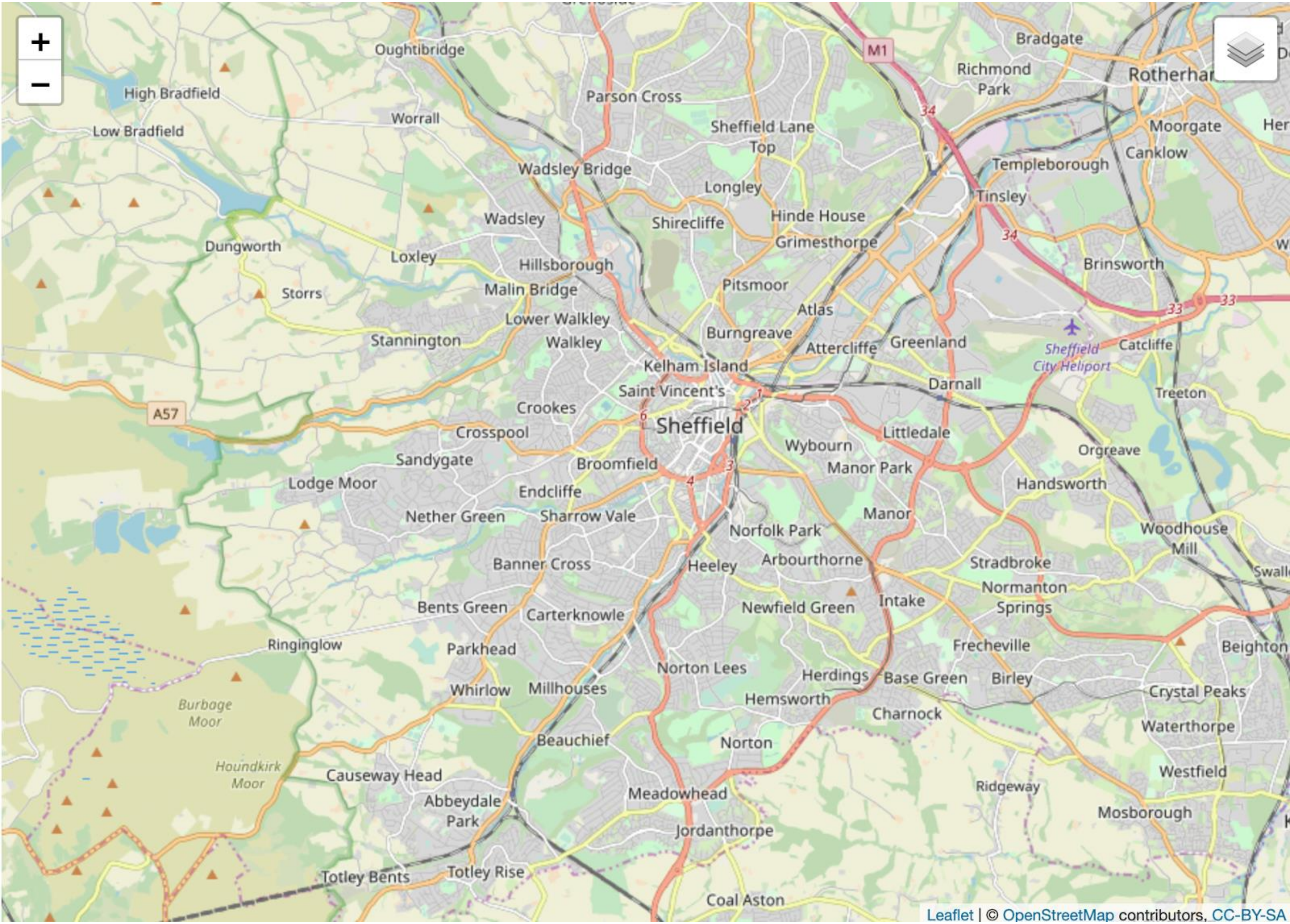
Filling-in the parameter coefficients (see table 3), we derive the following formula:

$$\hat{r_{ij}} = \exp(-5.402 - 0.048 * w_i - 0.082 * d_{ij} + \ln(p_i))$$

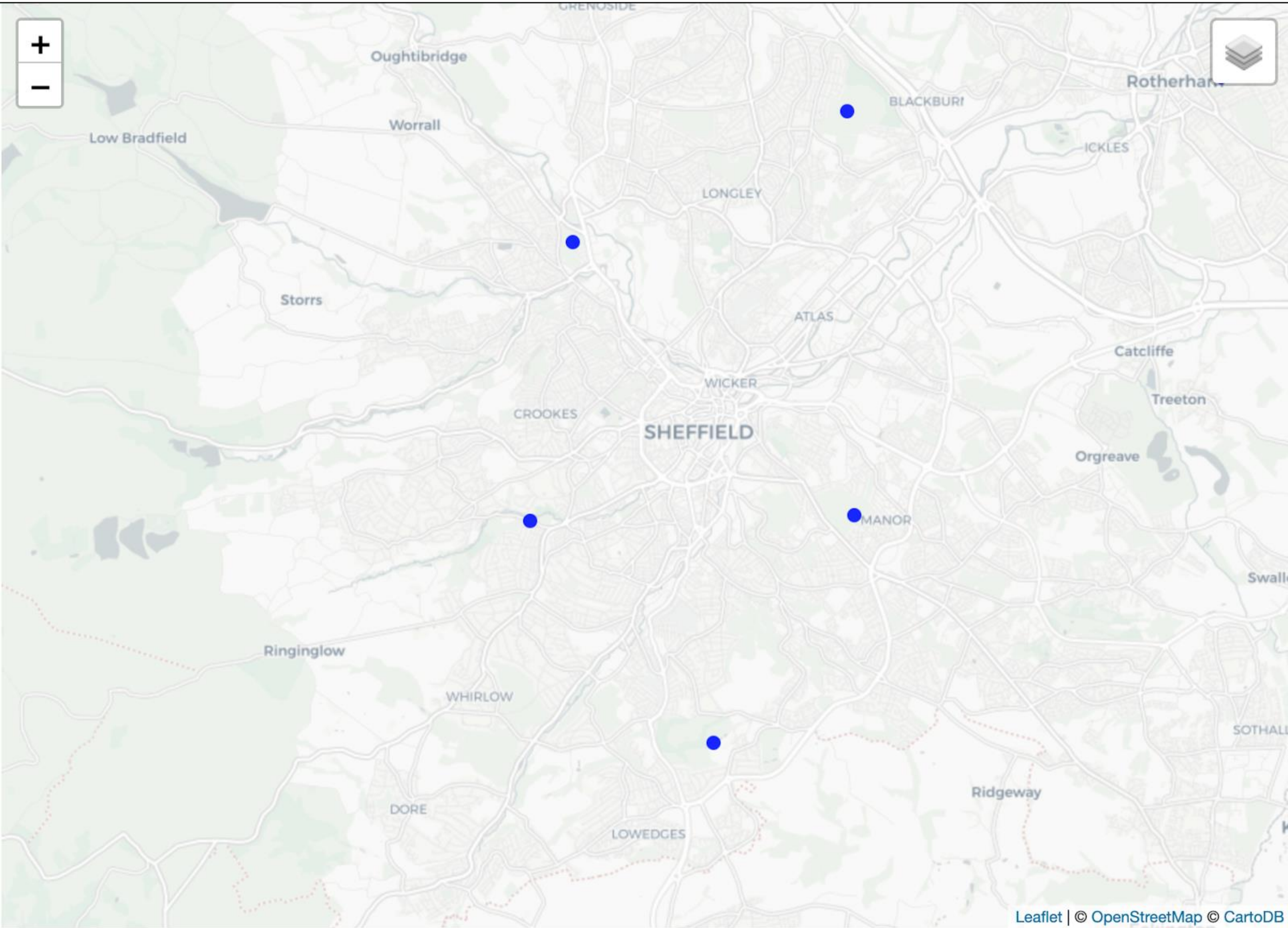
Note that j can have an effect on r_{ij} through d_{ij}: setting up a new event at location j will reduce the distance to the nearest event for some LSOA i. This means, we evaluate the distances from LSOA *is* location l_i to all established parkrun event locations $\{e_1, e_2, ..., e_{455}\} \in E$, denoted $\overline{l_i e_1}, \overline{l_i e_2}, \ldots, \overline{l_i e_{455}}$, and to the candidate green space location j, denoted $\overline{l_i j}$, and then take the minimum value, i.e. $d_{ij} = \min(\overline{l_i j}, \overline{l_i e_1}, \overline{l_i e_2}, \ldots, \overline{l_i e_{455}})$.

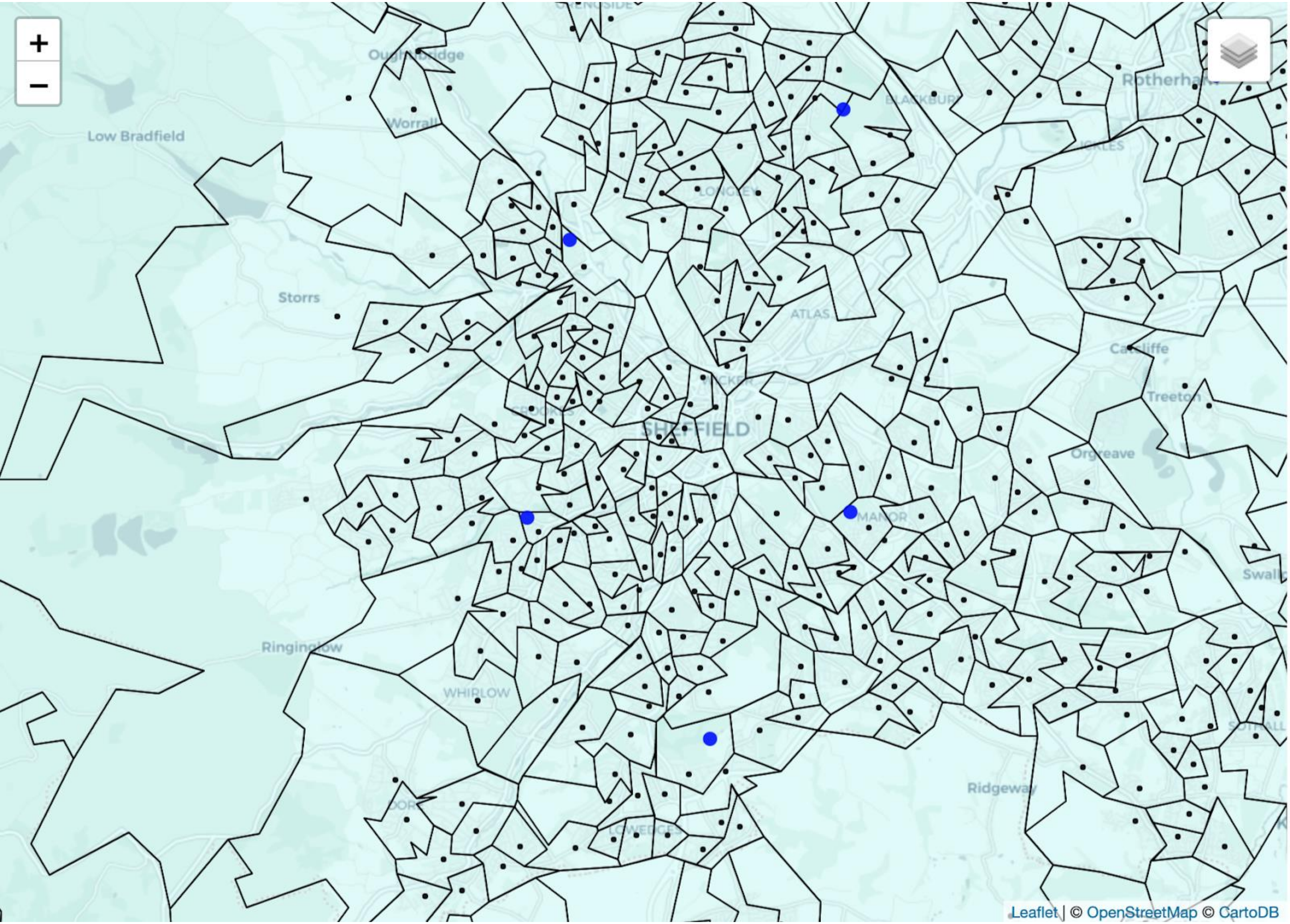
The expected change in the objective function is computed for all candidate locations j in the set of the available green spaces $C = \{c_1, c_2, \ldots, c_{2842}\}$, and the location with the maximum value is selected. The selection function is expressed in the following formula:

$$\arg\max_{j \in C} f(j|E)$$

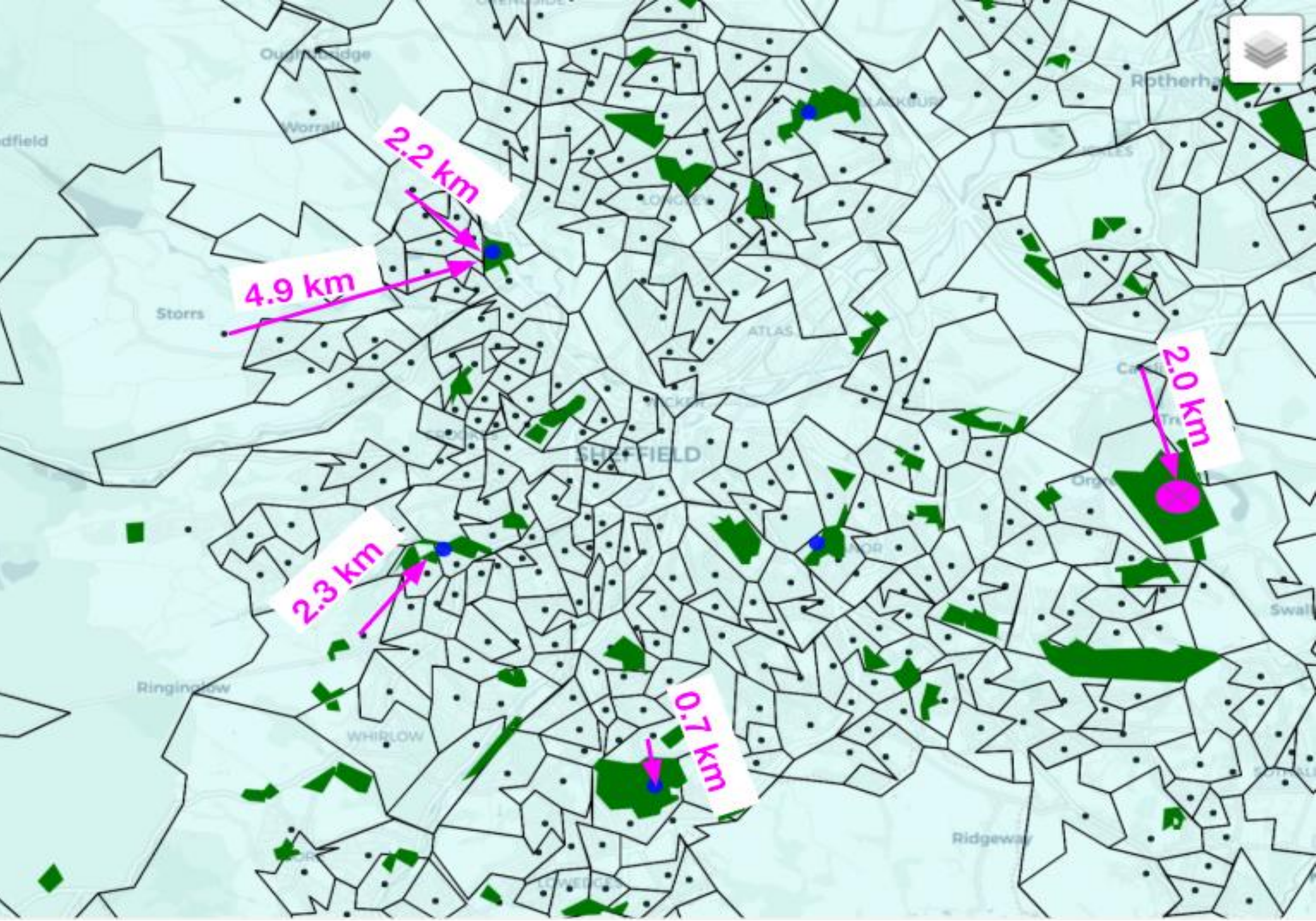


Oughtibridge
High Bradfield
Low Bradfield
Worrall
Parson Cross
Sheffield Lane
Top
M1
Bradgate
Richmond
Park
Moorgate
Canklow
Templeborough
Tinsley
Wadsley Bridge
Longley
Wadsley
Shirecliffe
Hinde House
Grimesthorpe
Dungworth
Loxley
Hillsborough
Brinsworth
Malin Bridge
Pitsmoor
Storrs
Atlas
Lower Walkley
Burngreave
Stannington
Walkley
Attercliffe
Greenland
Sheffield
City Heliport
Catcliffe
Kelham Island
Darnall
Saint Vincent's
Treeton
A57
Crookes
Sheffield
Crosspool
Littledale
Wybourn
Orgreave
Sandygate
Broomfield
Manor Park
Lodge Moor
Handsworth
Endcliffe
Nether Green
Sharrow Vale
Manor
Norfolk Park
Woodhouse
Mill
Banner Cross
Heeley
Arbourthorne
Stradbroke
Normanton
Bents Green
Carterknowle
Newfield Green
Intake
Springs
Ringinglow
Parkhead
Fretcheville
Beighton
Norton Lees
Herdings
Base Green
Birley
Whirlow
Millhouses
Crystal Peaks
Burbage
Moor
Hemsworth
Charnock
Waterthorpe
Beauchief
Norton
Houndkirk
Moor
Causeway Head
Westfield
Ridgeway
Abbeydale
Park
Meadowhead
Mosborough
Jordanthorpe
Totley Bents
Totley Rise
Coal Aston
Leaflet | © OpenStreetMap contributors, CC-BY-SA


GRENOSIDE
Oughtibridge
Worrall
Low Bradfield
BLACKBURN
Rotherham
ICKLES
LONGLEY
Storrs
ATLAS
Catcliffe
WICKER
Treeton
CROOKES
SHEFFIELD
Orgreave
MANOR
Ringinglow
WHIRLOW
SOTHALL
Ridgeway
DORE
LOWEDGES
Leaflet | © OpenStreetMap © CartoDB


Worrall
Low Bradfield
LONGLEY
Storrs
ATLAS
Treeton
SHEFFIELD
Orgreave
MANOR
Ringinglow
WHIRLOW
Ridgeway
LOWEDGES
Leaflet | © OpenStreetMap © CartoDB

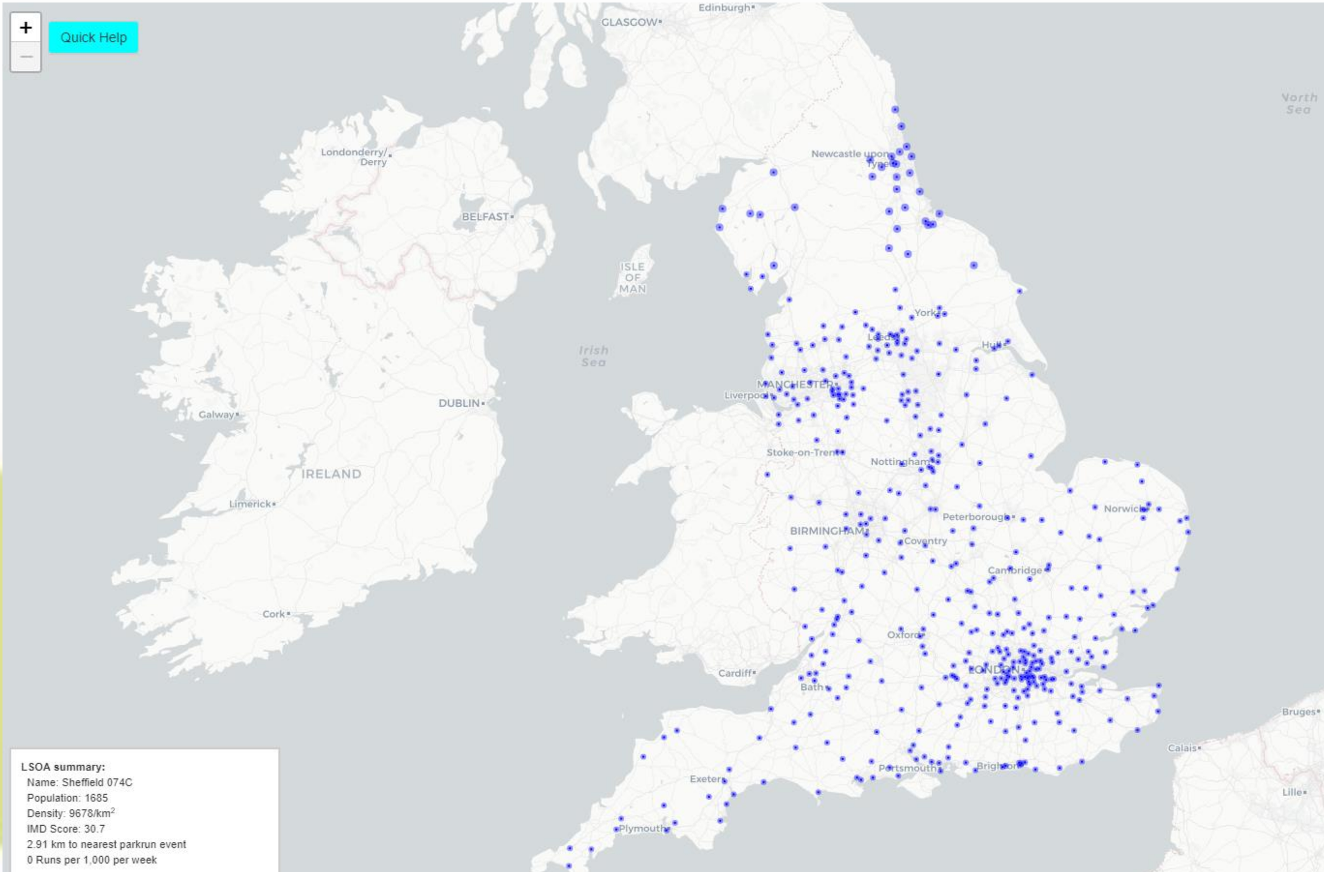
Access = { 2.2 * 1,220 + 4.9 * 2,351+ 2.3*1,915 + 0.7 * 1,844 + 5.1 * 1,530 } = 27,702.2



Access = { 2.2 * 1,220 + 4.9 * 2,351+ 2.3*1,915 + 0.7 * 1,844 + 2.0 * 1,530 } = 22,959.2

Results

Where should parkrun locate 200 new parkrun events?



https://github.com/bitowaqr/iol_map http://iol-map.shef.ac.uk/

Acknowledgements

Co-authors on pre-print Access & Participation Paper (Attachment 1):

Schneider PP[1], Bullas A[2], Bayley T[1], Haake SSJ[2], Brennan A[1], Goyder E[1]

Collaborators on conditional probability project (Attachment 2):

Chang J, Schneider PP[1], Brennan A[1], Goyder E[1]

Collaborators on the determinants of participation paper:

Schneider PP[1], Bullas A[2], Haake SSJ[2], Brennan A[1], Goyder E[1]

With special thanks to team at Parkrun Global Wellbeing:

Chrissie Wellington OBE, Rowan Ardill, Tom Mason.

[1]School of Health and Related Research, University of Sheffield, Sheffield, UK.
[2]Advanced Wellbeing Research Centre, Sheffield Hallam University, Sheffield, UK.
[3]Parkrun Global Health and Wellbeing, ParkrunUK, London.





Where should new parkrun events be located? Modelling the potential impact of 200 new events on geographical and socioeconomic inequalities in access and participation.

Schneider PP[1,*], Smith RA[1], Bullas AM[2], Bayley T[1], Haake SSJ[2], Brennan A[1], Goyder E[1]

[1]School of Health and Related Research, University of Sheffield, Sheffield, UK.
[2]Advanced Wellbeing Research Centre, Sheffield Hallam University, Sheffield, UK.

Abstract

Background
parkrun, an international movement which organises free weekly 5km running events, has been widely praised for encouraging inactive individuals to participate in physical activity. Recently, parkrun received funding to establish 200 new events across England, specifically targeted at deprived communities. This study aims to investigate the relationships between geographic access, deprivation, and participation in parkrun, and to inform the planned expansion by proposing future event locations.

Methods
We conducted an ecological spatial analysis, using data on 455 parkrun events, 2,842 public green spaces, and 32,844 English census areas. Poisson regression was applied to investigate the relationships between the distances to events, deprivation, and parkrun participation rates. Model estimates were incorporated into a location-allocation analysis, to identify locations for future events that maximise deprivation-weighted parkrun participation.

https://www.medrxiv.org/content/10.1101/19004143v1

Further Analysis – role of Ethnicity

Does ethnic density influence community participation in mass participation physical activity events?: a case of parkrun

Smith R[1], Schneider PP[1], Bullas A[2], Bayley T[1], Haake SSJ[2], Brennan A[1], Goyder E[1]

[1]School of Health and Related Research, University of Sheffield
[2]Sheffield Hallam University Advanced Wellbeing Centre

Keywords: parkrun, Participation, Physical Activity, Deprivation, Ethnic density.

Intended Journal: Journal of Racial & Ethnic Health Disparities, Sports Medicine Open, Behavioural and Social Sciences, Journal Of Epi & Comm Health.

Thumbnail Sketch

What is already known on this subject?

parkrun organise weekly 5km running and walking events at parks and green spaces across the world. Recent research has shown that despite equitable geographical access to parkrun events in England, participation is much lower in more deprived areas.

What this study adds?

This study uses regression modelling techniques to better understand the relative influence of geographical access, deprivation and ethnic density on parkrun participation rates in local communities. It finds that areas with higher ethnic density tend to have lower participation rates, even when controlling for deprivation.

Policy implications

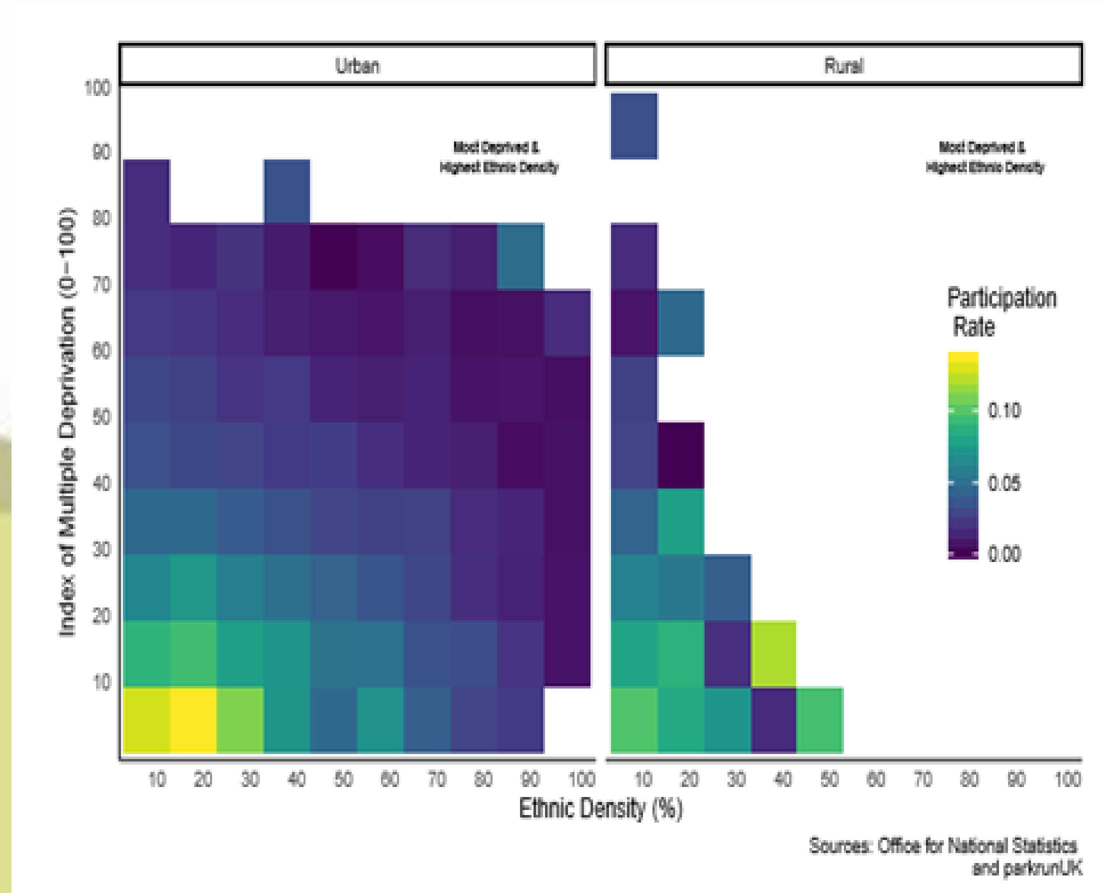
Identifying why particular communities are less likely to engage in parkrun, and finding ways to improve participation from these commmunities is likely to both improve overall population health and reduce inequalities.

Abstract

parkrun has been successful in encouraging people in England to participate in their weekly 5km running and walking events. However, there is substantial heterogeneity in parkrun participation across different communities in England: after controlling for travel distances, deprived communities have significantly lower participation rates.

This paper expands on previous findings by investigating ethnic disparities in parkrun participation. We combined geo-spatial data available through the ONS with participation data provided by parkrun, and fitted multivariable Poisson regression models to study the effect of ethnic density on participation rates at the Lower layer Super Output Level.

We find that areas with higher ethnic density have lower participation rates. This effect is independent of deprivation. An opportunity exists for parkrun to engage with these communities and reduce potential barriers to participation.



https://wellcomeopenresearch.org/articles/5-9

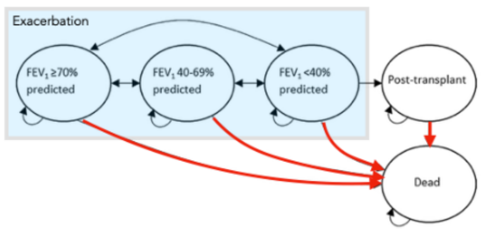
https://github.com/bitowaqr/DoPE

Cystic Fibrosis Adherence

Background & Data description

Just a simple markov model, based on a paper by Tappenden, Sadler & Wildman, about an early evaluation of the cost-effectiveness of an adherence intervention to improve outcomes for patients with cystic fibrosis. The results show that the intervention clearly dominates the baseline treatment.

For details, see their Journal article.



Problems & Questions

The model was originally implemented in Excel. I translated it into R as part of a research attachment on VOI. The code is neither very tidy nor fast, but it might serve as a sample model, for example to create some plots, build a shiny-application or play around with parallelisation.

Possible techniques & approaches

ggplot2, shiny, foreach, evppi

Any additional access restrictions that need to be incorporated for the data

The code can be distributed, re-used, and modified.

Contact

Paul Schneider
University of Sheffield
p.schneider@sheffield.ac.uk
https://bitowaqr.github.io/

https://github.com/HealthEconomicsDataDive/shiny_s5