Helping Business Picture Public Data with R

Environs

Dr Savas Hadjipavlou – Director, Justice Episteme Ltd

7th September 2022

www.justice-episteme.com
www.environs.app

Justice Episteme Ltd has developed an application, Environs, which can consolidate and map publicly available data to a variety of boundaries, to create a holistic picture of individual area statistics. Environs has been used to identify for each region in England, the Lower Super Output Areas ('LSOAs') that have the highest levels of crime.

Disclaimer - This analysis is a non-exhaustive; and does not seek to draw any conclusions about the LSOAs identified. Rather, the intention of this analysis is to demonstrate the functionality of the Environs application and that users have a powerful tool at their disposal for consolidating and interrogating data, which can direct their research, allow them to form hypotheses and support subsequent evidenced backed conclusions from the data.

Justice Episteme

- We are a small group of experts focused on data analytics.

- We use R to consolidate and combine real-world data, to create digital models which our customers use to make forecasts or to inform their organisation's strategic decisions.

- Our background is in law enforcement and we work mostly - but not exclusively - with public sector organisations, charities, policy think tanks.

- We are focused on applying scientific principles and techniques to questions of policy or practice.

- Our aim is to ensure strategic thinking, business investment and development uses the best available evidence – the numbers!

- *'A picture is worth a thousand words'*: Today I want to tell you, through some examples, how we are doing some of this, with public data, using the power of R.

  - There is a free version of Environs, an R Shiny application on our website.

  - This brings together a wide variety of publicly available socio-economic data, on service provision and other information to give a holistic picture at various geographic levels.

The Challenge

- Organisations need coherent data about demand, performance and outcomes.

- A wide range of information is available. Government publishes tons of it!

- **BUT** not usually in a form that can be easily joined together, key features understood, and from which specific and general points can emerge, and comparisons possible. It is in silos

- Wide geographical variation means we need to understand distributions.

    Tails of distributions pick out problem areas or exceptional results.
    Averages -> inadequate problem analysis -> poor performance

- Socio-economic indicators and other data published in different form, cover different geographies etc. Can achieve better picture by integrating using a common base.

- Want to be able to explore, to breakdown silos:

    - Look at a variety of metrics individually and together.

    - Identify distributions (joint distributions) and pick out top/bottom % (high/low service availability; good/poor performance).

    - Compare – areas, metrics etc.

    - Export selected data/analysis for use elsewhere.

Also A Word About the *glue* …UK Geography

- Super Output Areas are an ONS geographic hierarchy designed to improve the reporting of small-area statistics. (>200,000) – data most often published at higher levels as below:

- The Lower Super Output Areas and Data Zones list contains 42,619 areas of the following constituent geographies:

  - 32,844 Lower Super Output Areas in England.
  - 890 Super Output Areas in Northern Ireland.
  - 6,976 Data Zones in Scotland.
  - 1,909 Lower Super Output Areas in Wales.

  - Other Data reported at LA, PFA, CCG, Region, CSP, ….

And sometimes we will want to define our own depending on the problem.

- Environs Mapping algorithms can be flexible, joining up datasets at the possible geographical level.

- We also plan to accommodate census 2021 data.

Ref: https://webarchive.nationalarchives.gov.uk/ukgwa/20160106185615/https://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/index.html

What's good for? Getting insight to Real-World issues

Cost of living: Schools forced to swap hot meals for sandwiches due to food price inflation and funding cuts

i News

By Chloe Chaplain
Political Reporter

May 31, 2022 6:00 am (Updated 4:53 pm)

https://inews.co.uk/news/education/cost-of-living-schools-swap-hot-meals-sandwiches-food-price-inflation-funding-cuts-1659703

UK energy bills forecast to hit £3,850 as Russia cuts gas supply further

Consumers also warned bills of more than £3,500 a year could last 'well into' 2024

Q Search   The Guardian   International edition ˅
News website of the year

Rob Davies
🐦 @ByRobDavies
Wed 27 Jul 2022 17.25 BST

https://www.theguardian.com/money/2022/jul/27/uk-energy-bills-forecast-to-hit-3850-pounds-russia-cuts-gas-supply-further-europe-pipeline#:~:text=Energy%20bills-,UK%20energy%20bills%20forecast%20to%20hit%20%C2%A3,Russia%20cuts%20gas%20supply%20further&text=British%20households%20face%20being%20told,further%20squeezed%20Europe's%20gas%20supplies.

👑 GOV.UK

Press release
Better mental health support for people in crisis

Patients to benefit from new mental health ambulances and improvements to crisis care backed by £150 million as draft Mental Health Bill moves to next stage.

From: Department of Health and Social Care, Gillian Keegan MP, and The Rt Hon Sajid Javid MP
Published 27 June 2022

https://www.gov.uk/government/news/better-mental-health-support-for-people-in-crisis

POLITICS
Nadhim Zahawi to ban drug dealers from road

THE ⚜ TIMES

Matt Dathan, Home Affairs Editor

Tuesday July 12 2022, 9.57am, The Times

https://www.thetimes.co.uk/article/nadhim-zahawi-to-ban-drug-dealers-from-road-qqndkc6g6

Cost-of-living payments branded insufficient as energy bills soar

By Kevin Peachey
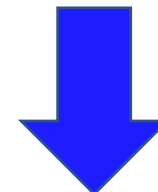Personal finance correspondent, BBC News

🅱🅱🅲

🕓 26 July   💬 Comments

NEWS

https://www.bbc.co.uk/news/business-62291652

So to motivate the approach let's look at some examples ...

Example: Health services

- The NHS is facing unprecedented challenges.

- How can we use health data to explore how people access services, e.g. GP services?

- Chart shows where people are registered with a GP.

- Daily commute means that a substantial fraction of residents in Kent have their GP in London.

- Other possibilities: use GP prescription data to characterise health need in communities.
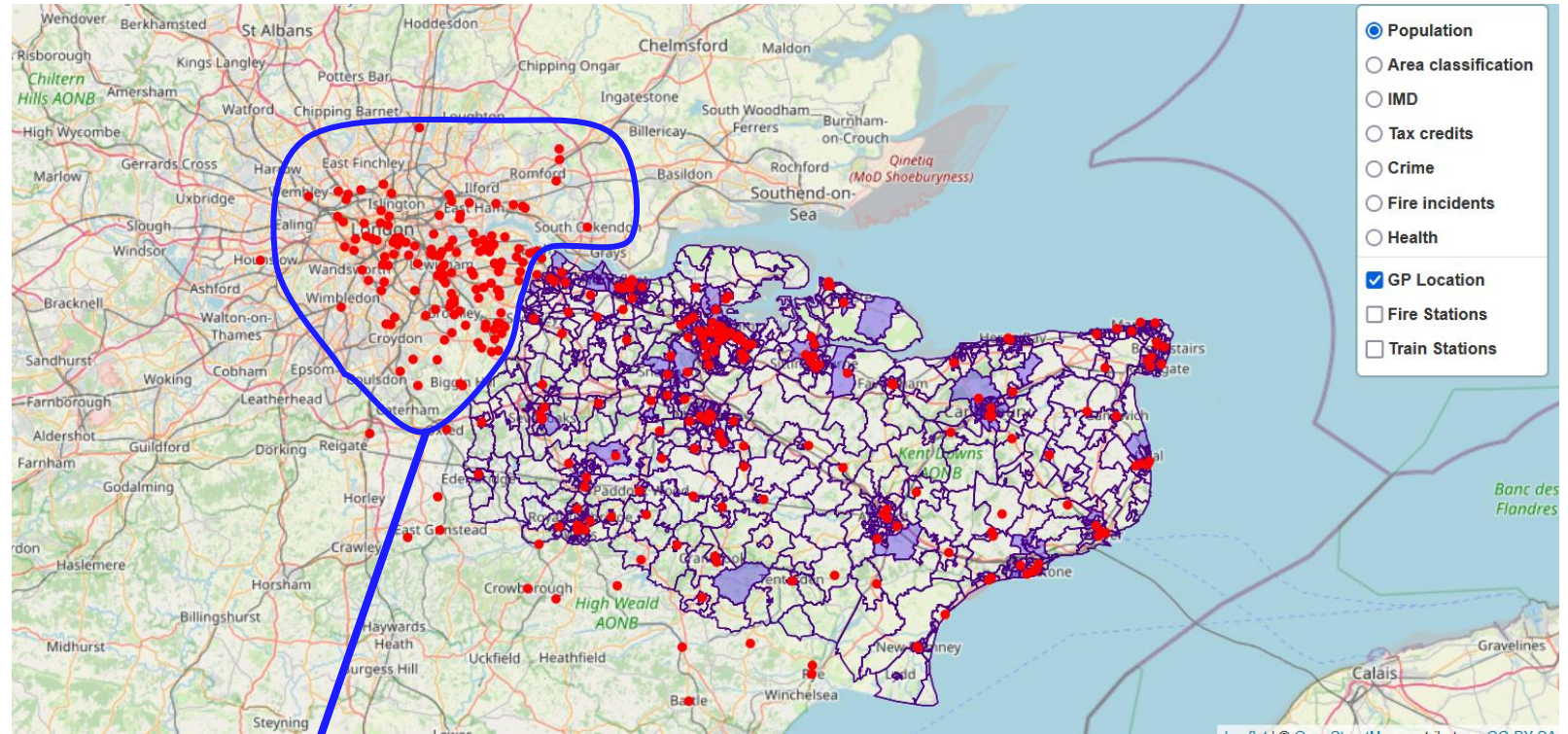
~47% of GP practices used by people living in Kent are out-of- area

Example: Children in Need in London - Tax credits

- Information on tax credits allows us to explore the level of need, and in which areas.

- Services, e.g. school meals, can then be planned and better targeted.

Top 10% by number of children supported by tax credits Areas

In these areas between 50-100% of resident children receive tax credits
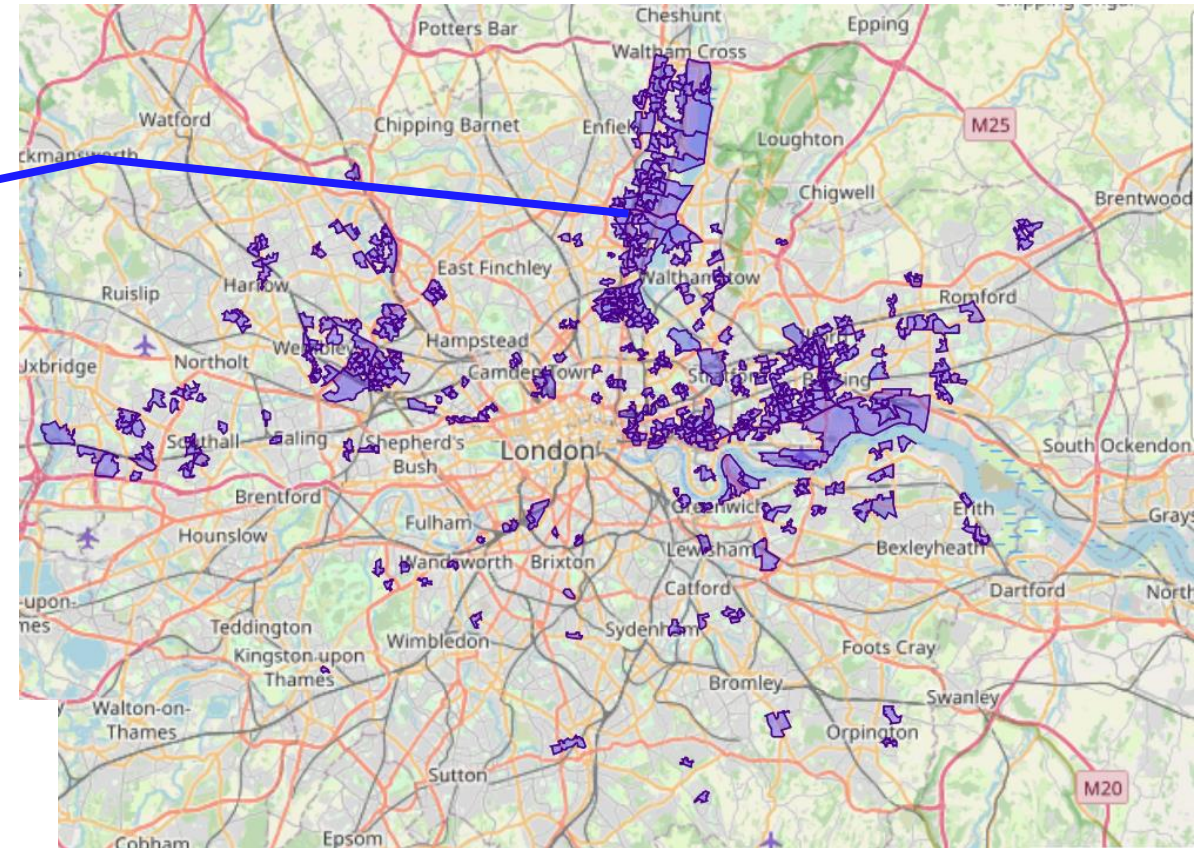
No of Children supported through Tax credits compared to resident population 0-17 years

- Selected Age Band
- Tax Credits - Children
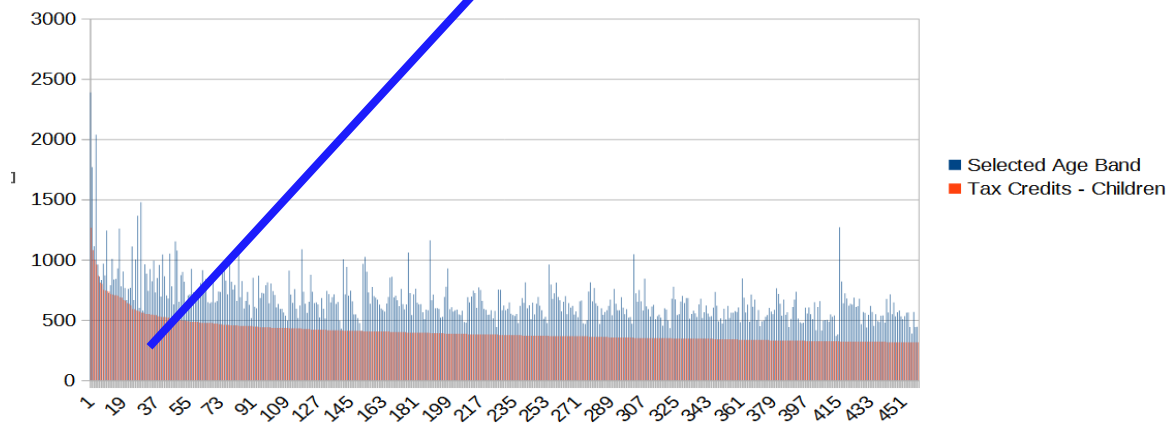
Example: Employment & Crime in Tower Hamlets

15.4% of the population in London Borough of Tower Hamlets (~332k) live in areas with highest (in area top 10%) crime rates; 16% of the population are children.

Areas with highest
crime rate (in area top 10%)
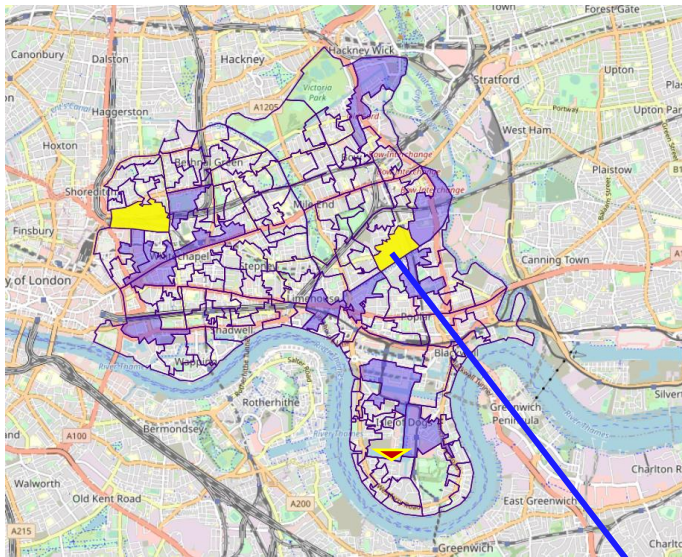
Areas with highest unemployment
(in area top 10%)
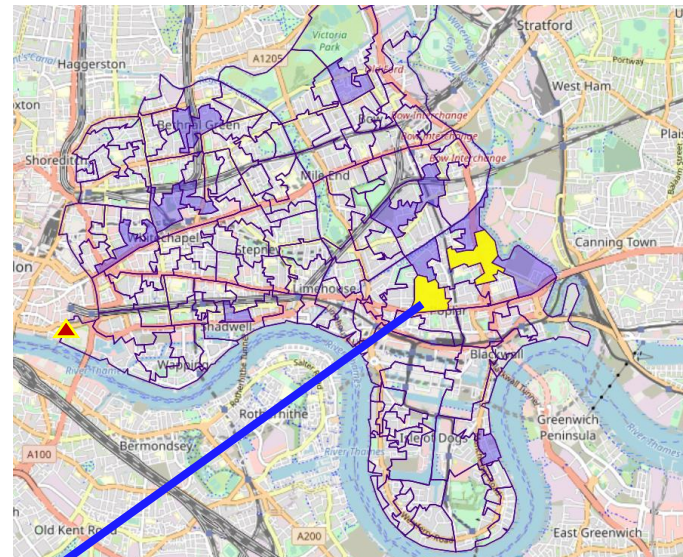
Explore strength of relationships through correlograms

Top 2 scores
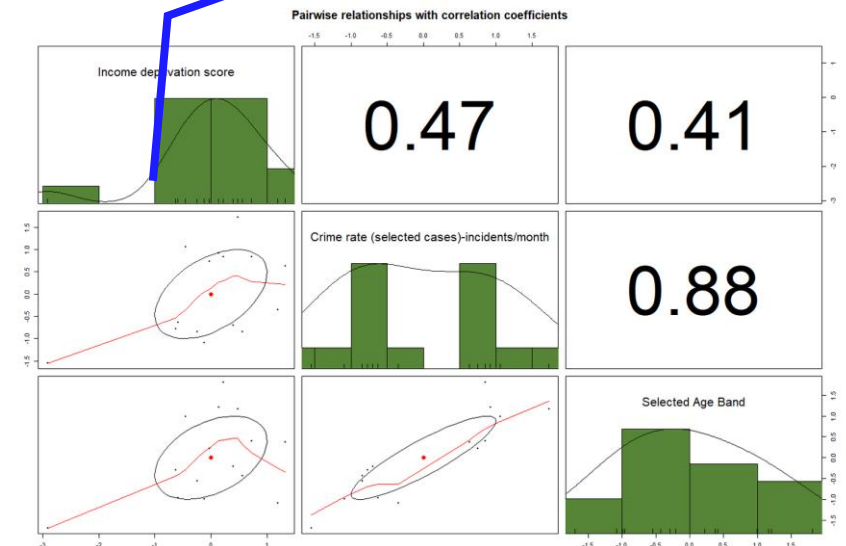
Environs allows bespoke configurations …

A tale of two cities? Manchester & Birmingham city centres

Manchester city centre
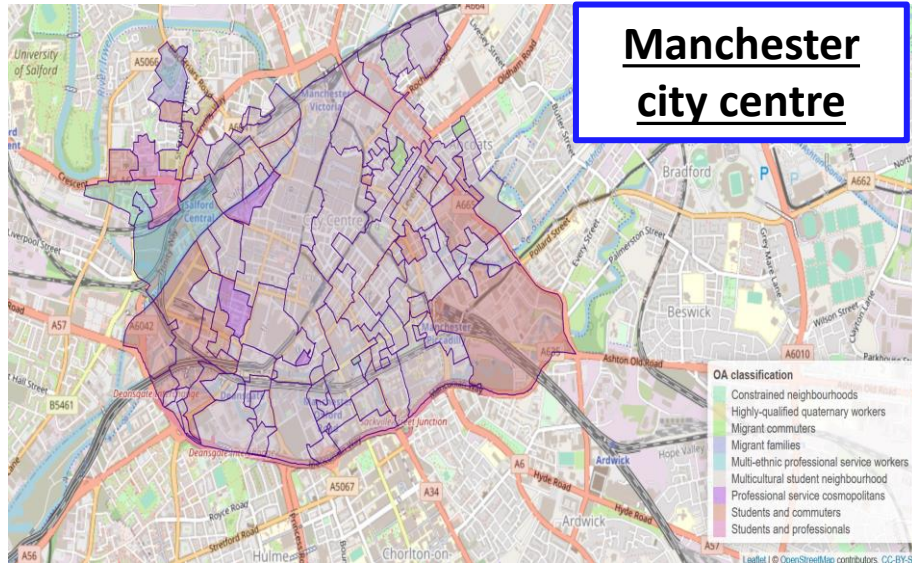
Subgroup	No of OAs
Multicultural student neighbourhood	77
Students and commuters	13
Professional service cosmopolitans	3
Constrained neighbourhoods	2
Migrant commuters	2
Highly-qualified quaternary workers	1
Migrant families	1
Multi-ethnic professional service workers	1
Students and professionals	1

Population Pyramid of selected area

View Pyramid for Year: 2020

Birmingham city centre
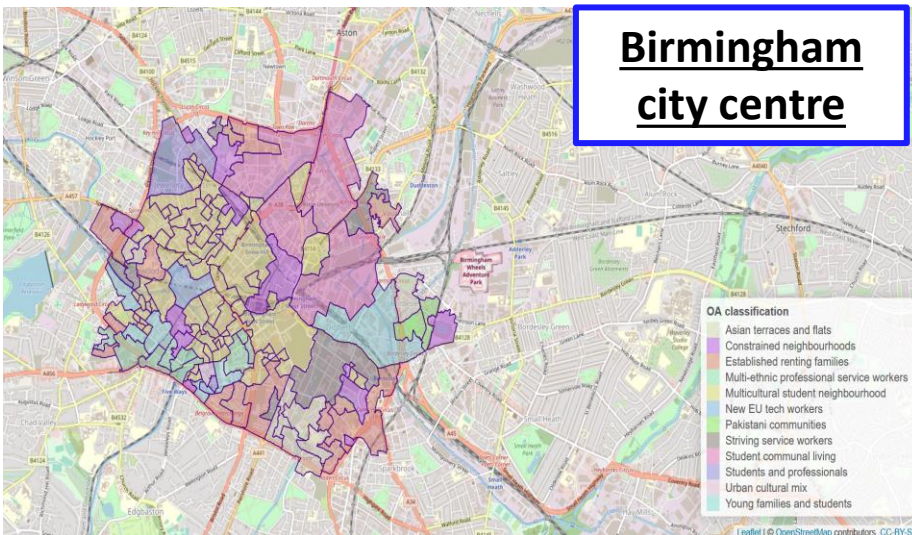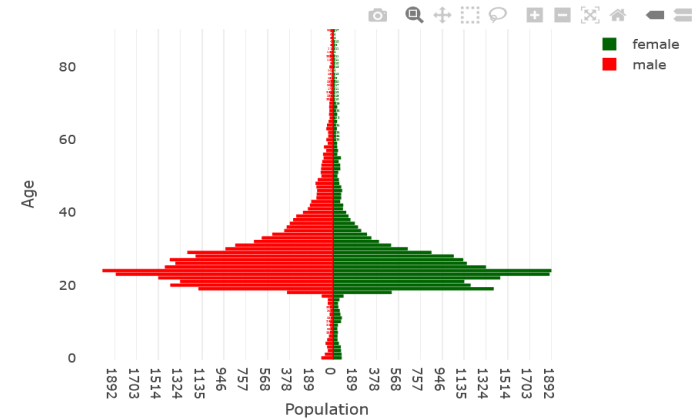
Subgroup	No of OAs
Multicultural student neighbourhood	67
Constrained neighbourhoods	13
Established renting families	13
Asian terraces and flats	8
Multi-ethnic professional service workers	8
Striving service workers	6
Young families and students	6
Student communal living	4
New EU tech workers	2
Pakistani communities	2
Students and professionals	2
Urban cultural mix	1

Population Pyramid of selected area

View Pyramid for Year: 2020
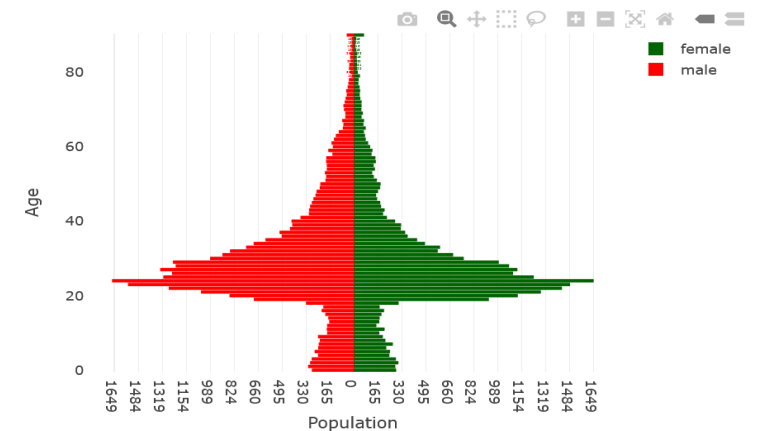
What Data Are Available? (Illustrative List)

'Geographical' data
Granular area-based data
Geographic and/or by administrative boundary
(*e.g. Police Force, Clinical Commissioning Groups, Local Authority, Parliamentary Constituency*)

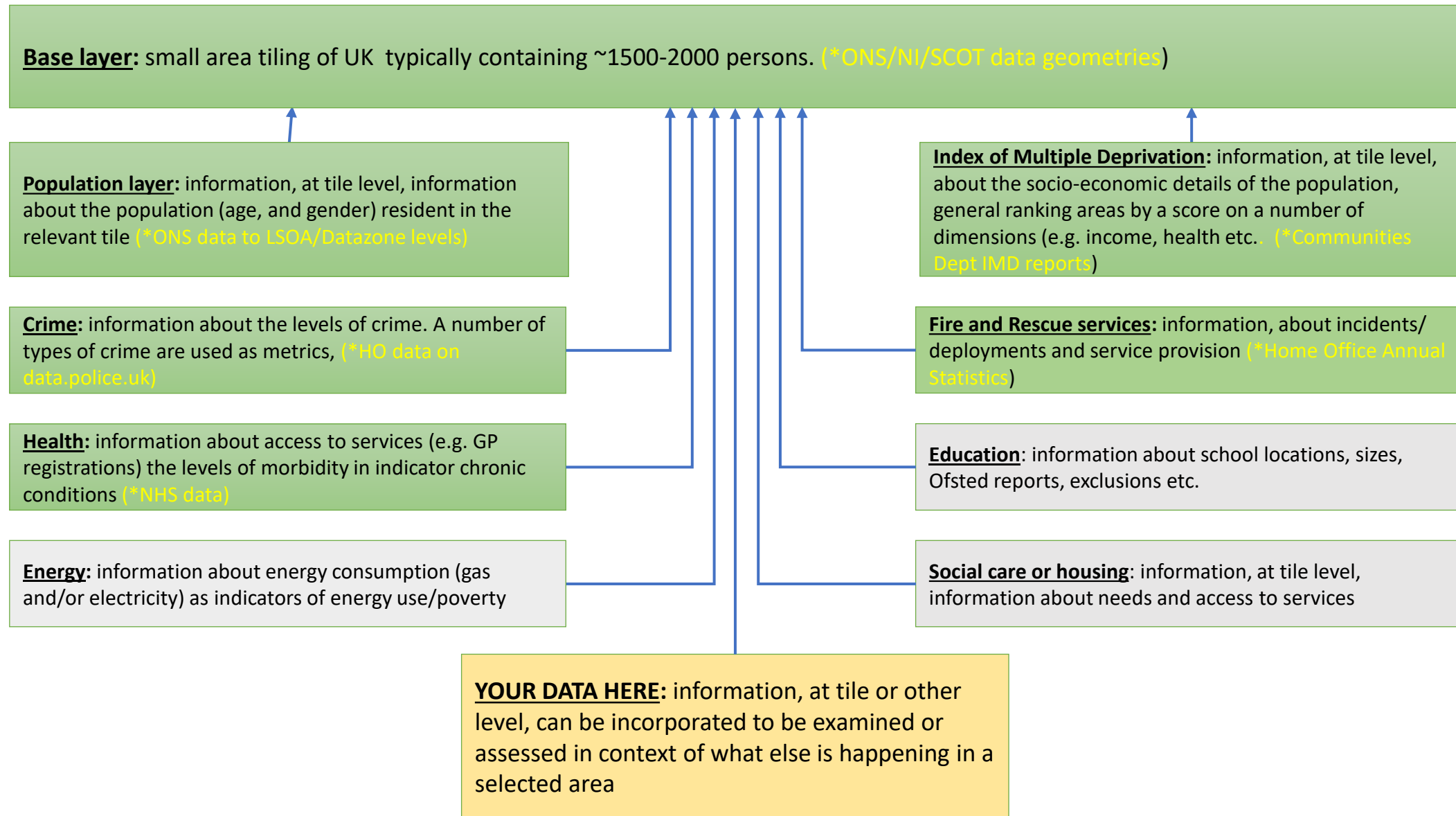|						
Census Data Socio-demographic data (e.g. age, sex/gender, nationality)	**Tax** (e.g. tax credits)	**Emergency Services Data** (e.g. Police, Fire, Ambulance)	**Education Data** (e.g. school type, school population size, free-school meals, languages spoken)	**Health Data** (e.g. GP data, Hospital data, mental health data, covid, prescription)	**Housing Data** (e.g. rental vs. ownership, energy usage, occupancy)	**Transport** (e.g. city 'day' Population, vehicle purchase data)

Environs – How is this done?

Base layer: small area tiling of UK typically containing ~1500-2000 persons. (*ONS/NI/SCOT data geometries*)

Population layer: information, at tile level, information about the population (age, and gender) resident in the relevant tile (*ONS data to LSOA/Datazone levels*)

Index of Multiple Deprivation: information, at tile level, about the socio-economic details of the population, general ranking areas by a score on a number of dimensions (e.g. income, health etc.. (*Communities Dept IMD reports*)

Crime: information about the levels of crime. A number of types of crime are used as metrics, (*HO data on data.police.uk*)

Fire and Rescue services: information, about incidents/ deployments and service provision (*Home Office Annual Statistics*)

Health: information about access to services (e.g. GP registrations) the levels of morbidity in indicator chronic conditions (*NHS data*)

Education: information about school locations, sizes, Ofsted reports, exclusions etc.

Energy: information about energy consumption (gas and/or electricity) as indicators of energy use/poverty

Social care or housing: information, at tile level, information about needs and access to services

YOUR DATA HERE: information, at tile or other level, can be incorporated to be examined or assessed in context of what else is happening in a selected area

Environs can be used in an exploratory mode …

Environs also has 'in tool' analysis, looking at distributions supporting comparisons

Experience Using R

- Functionality in R is spread across the language and a vast range of packages that specialise in particular areas. These packages can be exceptionality powerful, and are well curated/maintained. And open source. They provide a solid basis on which to build.

- We made particular use of those packages that help manipulate data frames and other more complex list structures: to shape and join the available open source data into relevant and useful combinations.

- Mapping: the 'leaflet' package handles geographic information superbly!

- R Shiny provides the framework for organising the exploration and analysis work flow of the application organise the information.

- A key challenge in a project like this is deciding how much to build for yourself, and how to stitch together the functionality of ready built solutions! Not always straightforward

- All in all R is a fun language to work with (with the occasional surprise!)

Summary

- We are excited about the potential of **Environs** to gain value and provide insights. What started as a proof of concept is turning into a commercial product. And we are continuing to develop it!

- Just as important, an enterprise can choose to maximise the utility of its own data by adding that in. So bespoke configurations are possible.

- The ability of R to combine and process data, do statistics and visualise the results, all under one roof, has definitely been a key enabler.

- Questions?