Consumer Data Research Centre

An ESRC Data Investment

NOKIA BELL LABS

Newcastle University

University of Nottingham
UK | CHINA | MALAYSIA

UKRI Engineering and Physical Sciences Research Council

Analysing Prescription Outcomes Through Spatial Signatures

Author: *Christopher Clive Drowley*

Supervisors:

Newcastle University
Dr. Clement Lee
Professor Stuart Barr

Nokia Bell Labs
Dr. Sanja Šćepanović
Dr. Daniele Quercia

University College London
Dr. Stephen Law

Datasets:

MedSat[1]

Spatial Signatures of Great Britain[2]

1 Introduction

Asthma is a major public health concern and there are known associations between urban configuration, land use, and pollution. Understanding these associations is difficult at the individual level yet increasing data availability enables modelling at low-level geographical units. Modern data science practices, focusing on reproducible and reusable data products, provided the foundation for **this study**, which **predicts asthma prescription quantities for Lower Layer Super Output Areas** (LSOAs) **with a focus on model interpretability**.

2 Objectives

- **A** Merge MedSat and Spatial Signature datasets to a common geographic granularity.
- **B** Develop interactive visualisations for Spatial Exploratory Data Analysis (EDA).
- **C** Apply models to forecast Asthma prescription outcomes.
- **D** Employ ML explainability approaches to identify significant factors.

3 Methodology

Study Area: *33,755 LSOAs across England.*

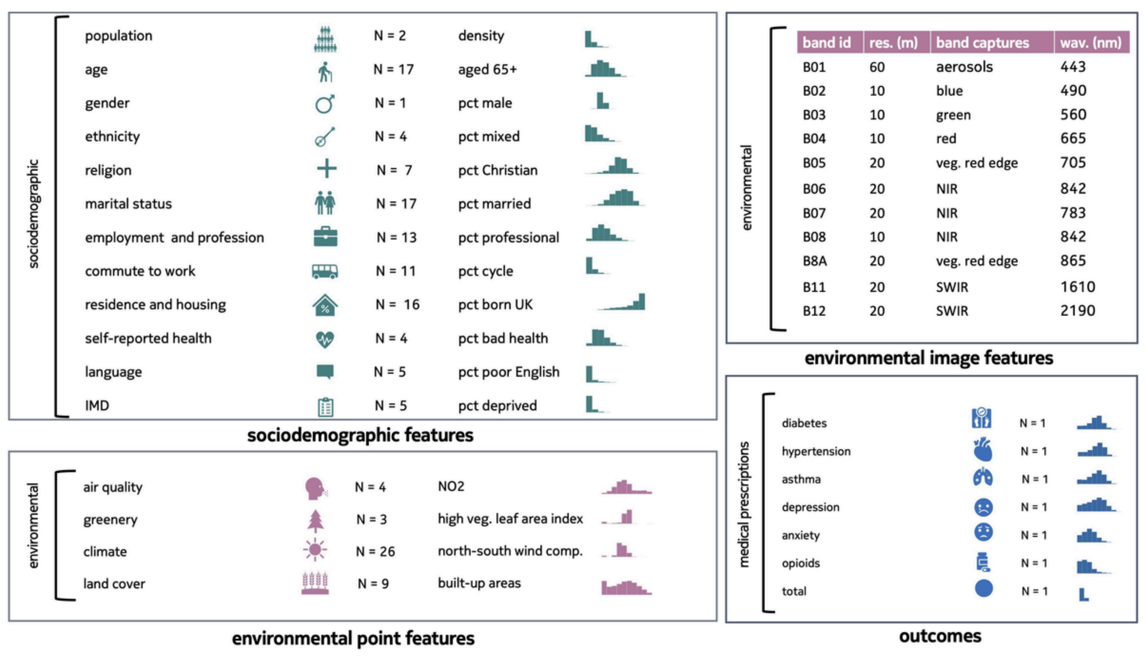
Extract, Transform, Load: *MedSat and Spatial Signatures (SS) data.*

SS Interpolation: *LSOA 'Dominant Typology' by highest areal overlap.*

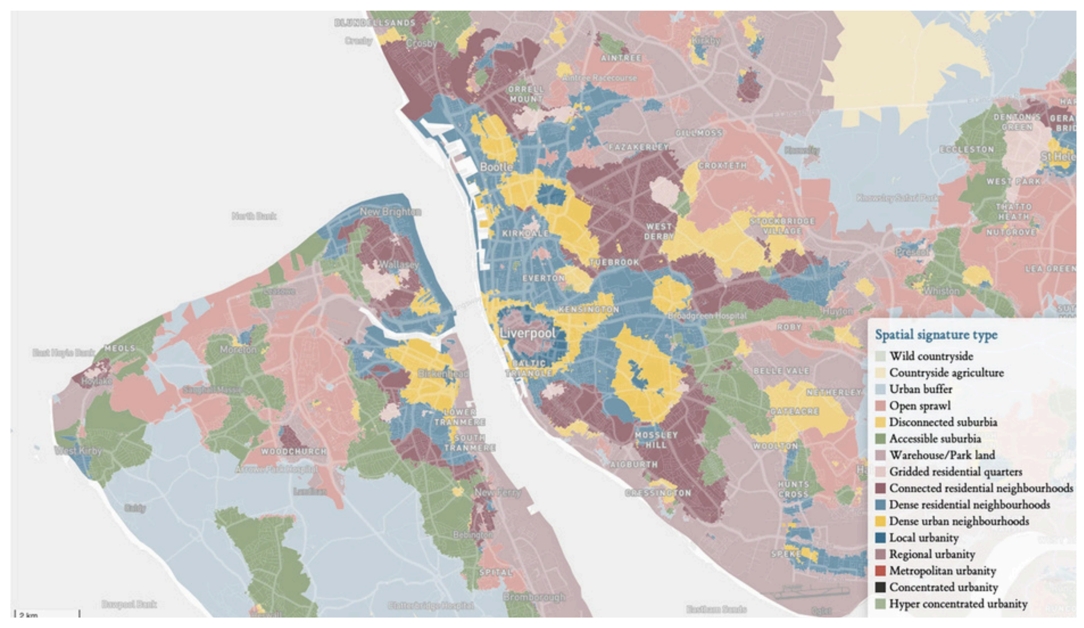
Exploratory Data Analysis: *Review distributions and relationships with asthma prescriptions, informing feature selection for modelling.*

Cross-Validation (CV): *Spatial block folds with custom class balancing CV assessed against random CV.*

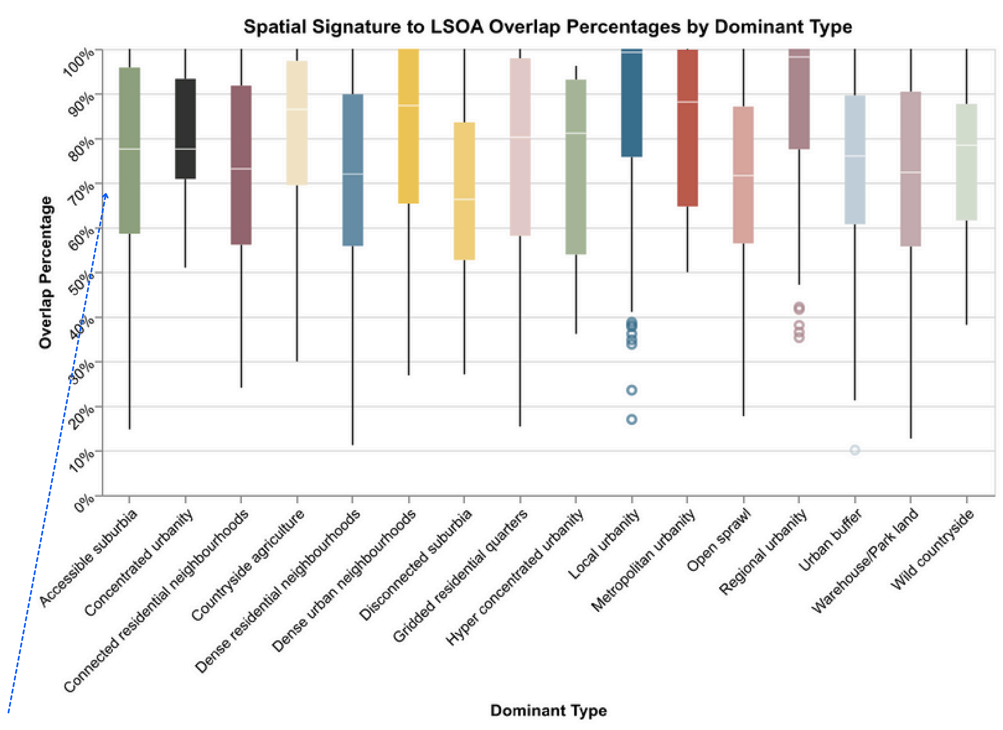
MedSat Dataset Summary [1]



Spatial Signatures Typologies around Liverpool [2]

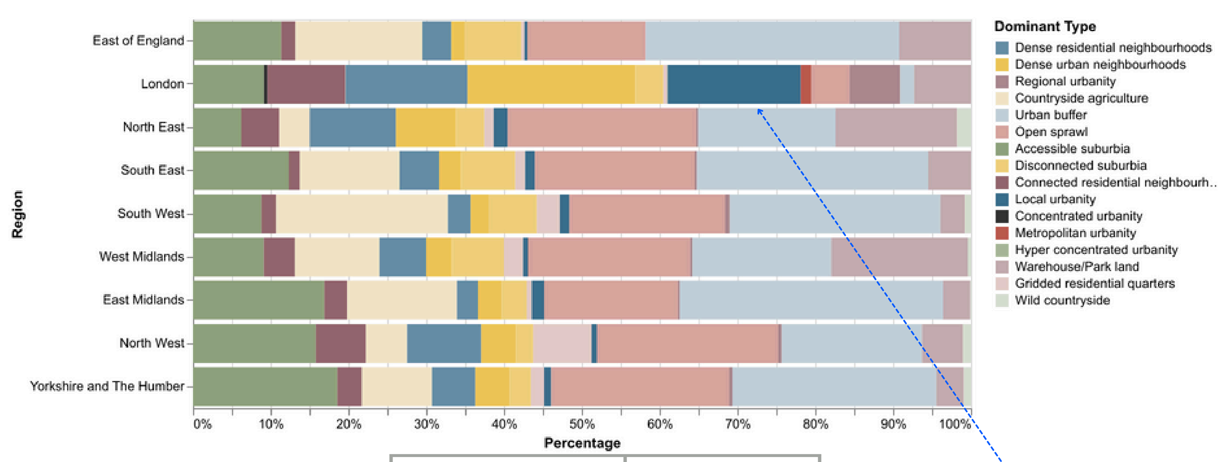


A Merging Datasets - LSOA Interpolation Results



"Most Spatial Signature typologies have a median areal overlap above 70%."

LSOA Interpolation Results - Spatial Signature Types by Region



"LSOA assigned to a Spatial Signature typology if >70% area covered reducing modelled LSOAs by ~50%."

Areal Overlap Threshold	LSOA Count
50%	25,658
70%	**16,852**
90%	8,351
99%	3,523

"London stands out with a high share of High Urbanity and Dense Neighbourhood typologies."

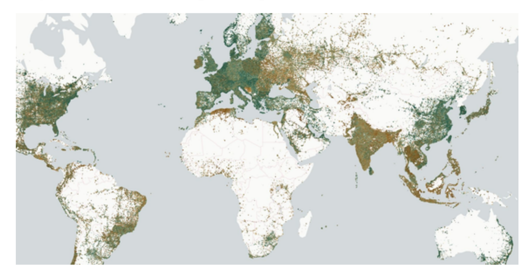
B Spatial EDA & processing with modern Geo-technology

GeoParquet

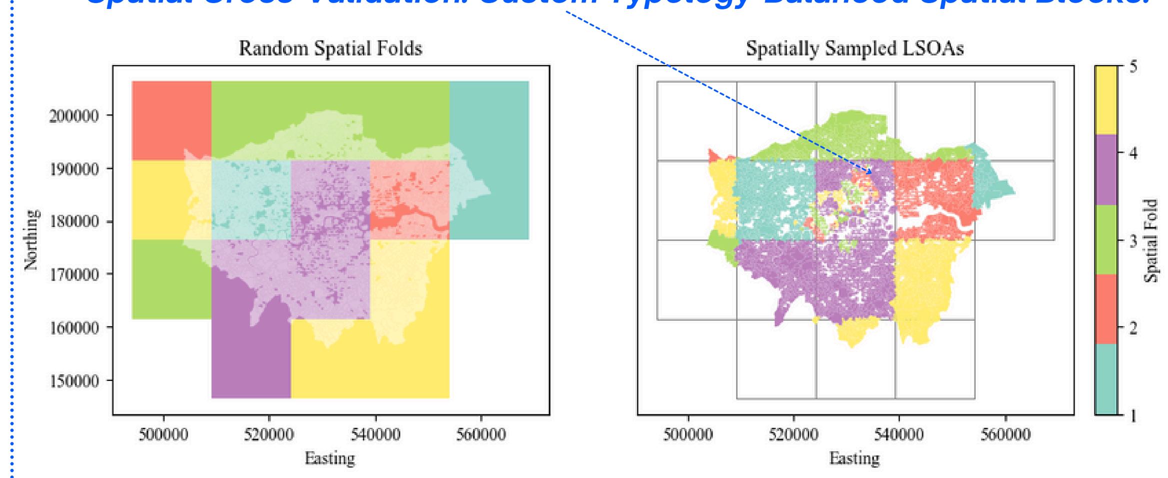
Geospatial data in Parquet

GeoArrow

developmentseed.org/blog/2023-10-23-lonboard



"Spatial Cross-Validation: Custom Typology-Balanced Spatial Blocks."



4 Analysis

Model Comparison: *Linear regression and CatBoost were used to predict asthma prescription quantities per capita.*

Model Performance: *CatBoost outperformed linear regression with minimal fine-tuning, extensive fine-tuning yielded limited gains.*

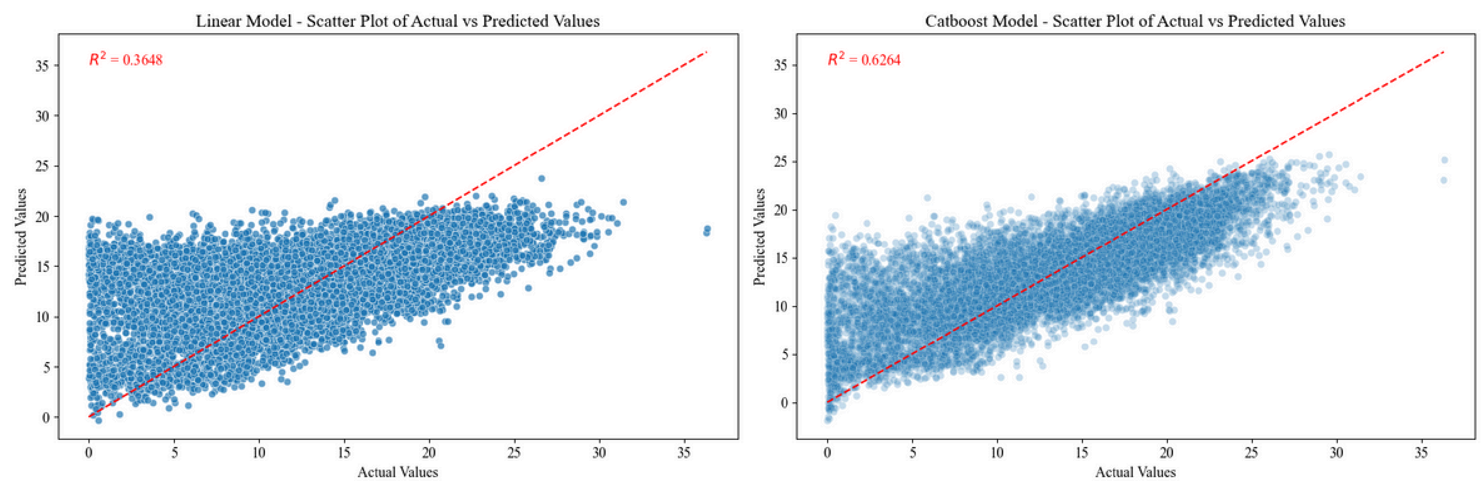
Model Interpretability: SHapley Additive exPlanations (SHAP) highlights key features impacting predictions and feature interactions.

C

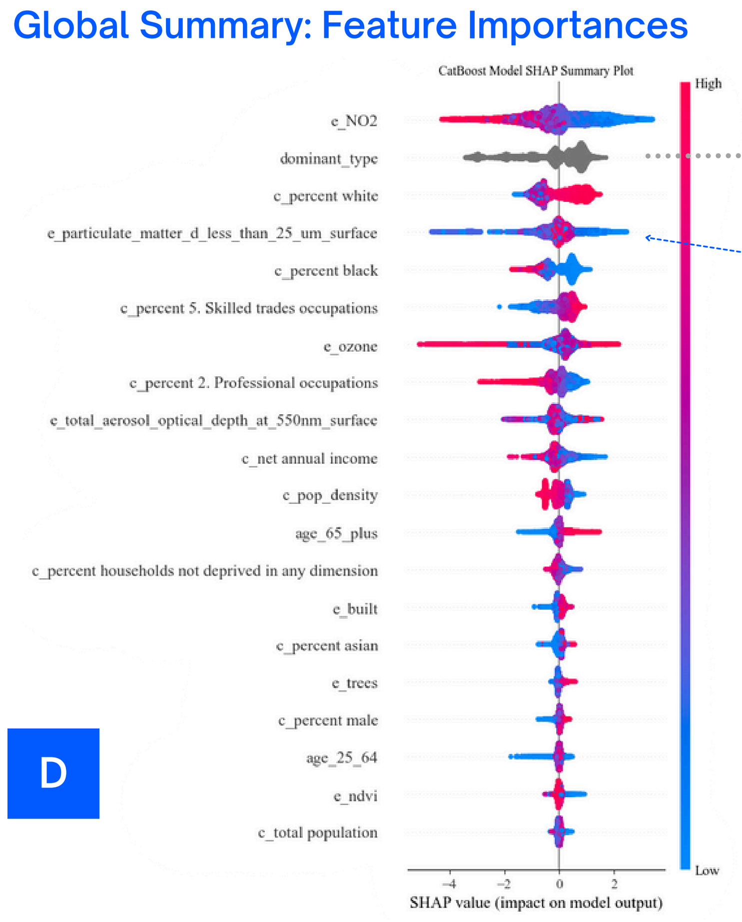
"Heavy clustering near (not including) zero for both models."

"Both models over-predict at lower prescription quantities and vice versa."

Linear & Catboost Model - Scatter Plot of Actual vs Predicted Values



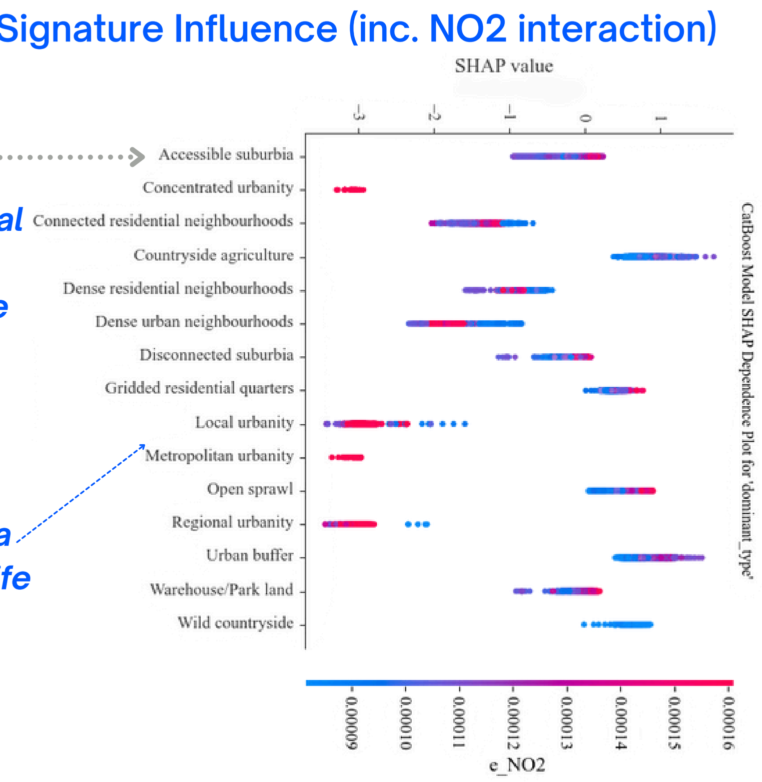
Global Summary: Feature Importances



"Key Predictors: NO2, spatial signature typology, white population, and particulate matter."

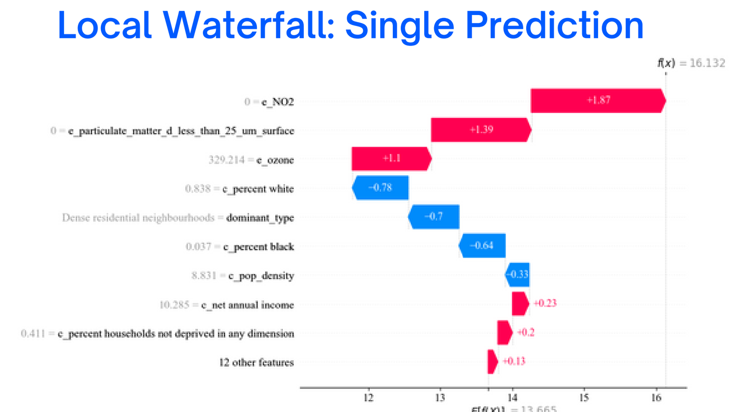
"Highly polluted and urban areas showed lower asthma prescriptions, aligning with life course risk patterns."

Dominant Spatial Signature Influence (inc. NO2 interaction)

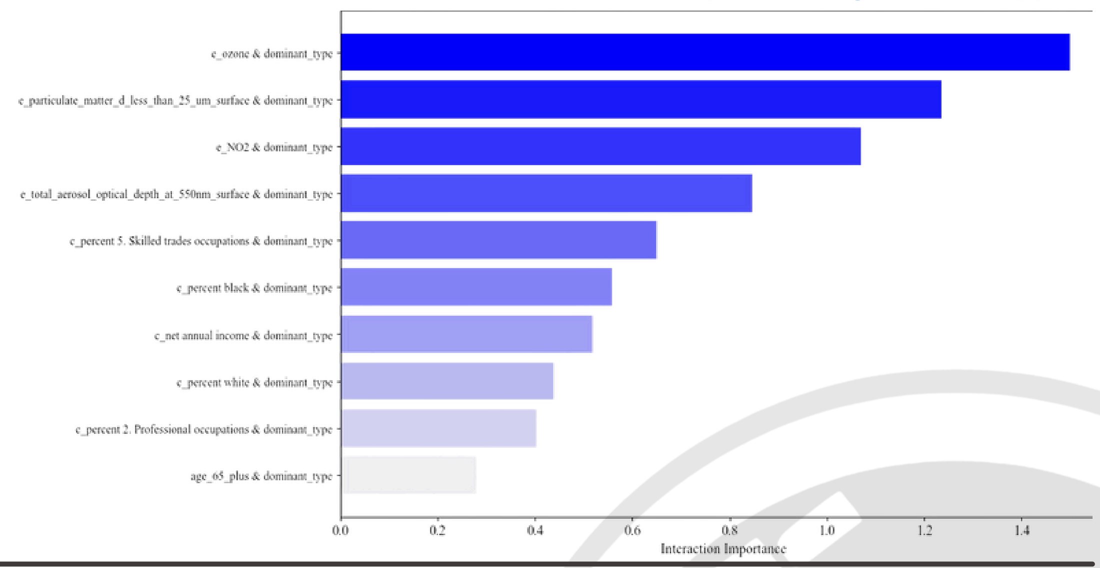


D

Local Waterfall: Single Prediction



Top Feature Interactions with the Dominant Spatial Signature Type



5 Discussion & Future Work

Urban areas with younger populations had lower asthma prescriptions despite higher pollution compared to rural, older areas. Considering population movement could help account for confounding variables and enhance understanding of which elements of urban form and function impact asthma and public health more generally.

- **Modelling:** CatBoost outperformed linear models and avoided extensive feature engineering, though the linear model served as a useful benchmark and EDA tool.
- **Interactive Tools:** Lonboard/GeoArrow enable fast spatial EDA. Work is ongoing to build a drop-in framework for any tabular data.
- **Spatial Signatures:** Classifications may benefit from hierarchy or ordering. Future work should test sensitivity to data changes and census geometry impacts.
- **Spatial Interpolation:** Future work should enumerate and evaluate various methods of spatial interpolation across a range of datasets to measure signal loss.
- **Spatial Cross Validation:** Future work should assess the utility of spatial versus traditional cross-validation techniques.

Related literature

[1] Scepanovic et al. (2023). MedSat: A Public Health Dataset for England. NeurIPS 36, 77810–77851. https://doi.org/10.14459/2023mp1714817

[2] Arribas-Bel & Fleischmann (2022). Spatial Signatures: Understanding Urban Spaces. Habitat International, 128, 102641. https://doi.org/10.1016/J.HABITATINT.2022.102641