

# Spatially indexed distribution analysis as applied to house prices in the UK

---

**Luke A. Barratt** and John A. D. Aston  
Statistical Laboratory, DPMMS  
University of Cambridge, UK

EcoSta 2025  
Tokyo, Japan  
23 August 2025

# Background

---



# House Prices in the UK

- Ipsos 2022:
  - 69% of Britons believed there was a national housing crisis in the UK.
  - 50% disagree vs 22% agree that there is enough affordable housing locally.
- YouGov August 2025:
  - housing as important an issue as the environment, defence and security, or tax.
  - among 18–24-year-olds it is the second most important issue behind ‘the economy’.
- Labour Party Manifesto, 2024:
  - Housing is ‘one of the country’s biggest barriers to growth’.
  - They aim build 1.5 million homes, including cheaper/social housing, over the parliament.

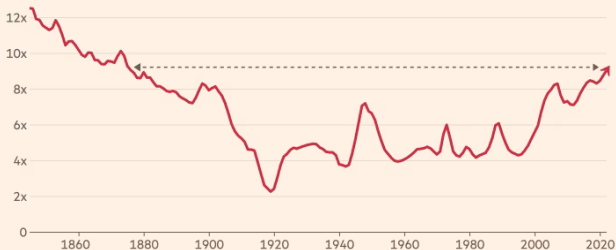
# Existing Work

## 1. Main points

- In 2024, the median average home in England, at £290,000, cost 7.7 times the median average earnings of a full-time employee (£37,600); in Wales the average home (£201,000) was 5.9 times annual earnings (£34,300).
- Affordability in England and Wales in 2024 has returned to its pre-coronavirus (COVID-19) pandemic levels after a sharp increase between 2020 and 2021 (worsening affordability); median house sales prices have increased by 1% since 2021, while average earnings have increased by 20%.
- In 2024, 9% of local authorities (LAs) (27) had homes bought for less than five times workers' earnings on average and were therefore deemed affordable; this is the highest proportion since 2015, but well below that at the start of the series in 1997 (88% of areas).
- Housing affordability improved in 289 of the 318 LAs in England and Wales (91%) and worsened in 28 (9%) since 2023.
- The most affordable LAs in 2024 were Blaenau Gwent (with a ratio of 3.8), Burnley (3.9) and Blackpool (3.9); the least affordable was Kensington and Chelsea (27.1, which was about seven times less affordable).
- Between 2019 and 2024, 4 of the 10 largest increases in affordability ratios (worsening affordability) were in the East Midlands, while the 10 largest decreases in affordability ratios have all occurred in LAs in London.

## The last time UK house prices were this expensive relative to earnings was in 1876

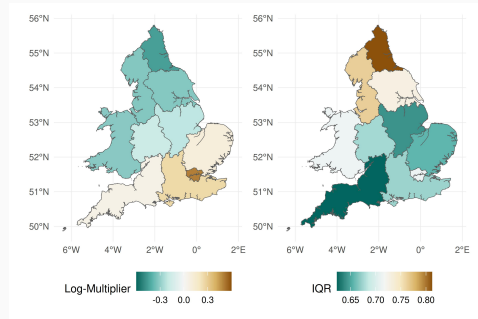
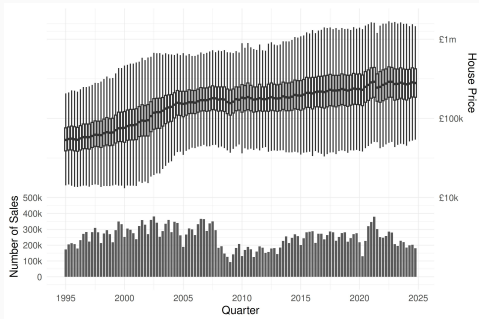
Average house price as a multiple of average earnings



Source: Schroders 607178 / Duncan Lamont  
FT graphic: John Burn-Murdoch / @jburnmurdoch  
© FT

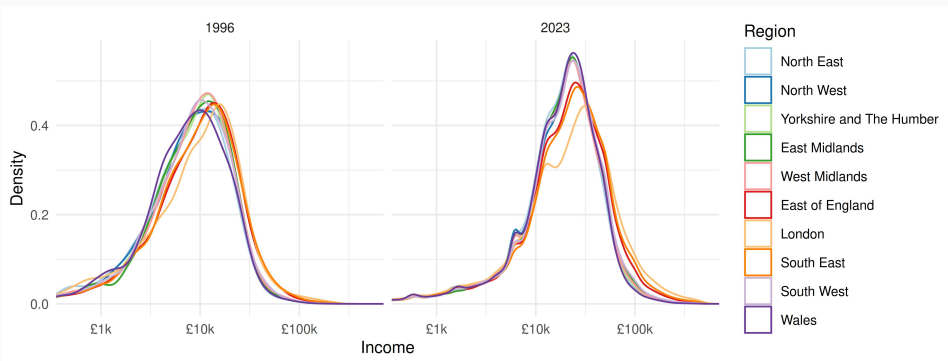
# House Price Data

- HM Land Registry Price Paid Data Set.
- 1 Jan 1995–: over 30 million single residential property sales for value, referenced by address and date, and marked by price recorded.



# Income Data

- HM Revenue and Customs' Survey of Personal Incomes.
- 6 Apr 1995–: a weighted sample of those who may be liable for income tax, referenced by region and tax year, and marked by total income.



- Prices paid:

$$p_i \sim P(s_i, t_i), \quad i \in [n]. \quad (1)$$

- Incomes earned:

$$y_i \sim Y(r_i, \tau_i), \quad i \in [m], \quad (2)$$

with weightings  $w_i$ .

- Aim is to estimate these distribution fields  $P$  and  $Y$ , then to interpret them through a regression of the quantile functions:

$$P^{-1}(s, t; q) = \theta(s, t; q)Y^{-1}(s, t; q). \quad (3)$$

- The parameter  $\theta$  can be interpreted for each  $q \in [0, 1]$ , or by taking functionals.

# Methodology

---

# House Price Distributions

- We estimate  $P$  on a spatial grid  $s \in S$  and for each tax year  $\tau \in T$ .
- We apply a Nadaraya–Watson type smoother to the observations:

$$\hat{P}(s, \tau; p) : \propto \sum_{i \in [n]: t_i \in \tau} K_h(d(s, s_i)) \mathbb{I}(p_i \leq p), \quad (4)$$

where  $K_h$  is a Gaussian kernel of bandwidth  $h$ .

- The bandwidth  $h$  is chosen locally to ensure the effective sample size is fixed:

$$\sum_{i \in [n]: t_i \in \tau} \left( K_h(d(s_i, s)) \right)^2 \bigg/ \left( \sum_{i \in [n]: t_i \in \tau} K_h(d(s_i, s)) \right)^2 := \frac{1}{n^*} = \frac{1}{100}. \quad (5)$$

# Income Distributions

- We estimate  $Y$  on the same grid  $(s, \tau) \in S \times T$ .
- For each region  $r$  we estimate a regional income distribution:

$$\hat{Y}(r, \tau; y) : \propto \sum_{i \in I(r, \tau)} w_i \mathbb{I}(y_i \leq y), \quad (6)$$

where  $I(r, \tau) := \{i \in [m] : \tau_i = \tau, r_i = r\}$ .

- We suppose there is a population density  $\rho$ , and then smooth by a Nadaraya–Watson type smoother to these estimates:

$$\hat{Y}(s, \tau; y) : \propto \sum_r \left( \int_r K_h(d(s, \tilde{s})) \rho(\tilde{s}) d\tilde{s} \right) \hat{Y}(r, \tau; y). \quad (7)$$

- The bandwidth is chosen again to ensure a fixed effective sample size:

$$\sum_r \left( \int_r (K_h(d(s, \tilde{s})))^2 (\rho(\tilde{s}))^2 d\tilde{s} \right) \left( \sum_{i \in I(r, \tau)} w_i^2 / \left( \sum_{i \in I(r, \tau)} w_i \right)^2 \right) / \left( \sum_r \int_r K_h(d(s, \tilde{s})) \rho(\tilde{s}) ds \right)^2 := \frac{1}{m^*} \quad (8)$$



- Smoothing has occurred in the estimation of distributions, so all that remains is a regression analysis of the distributions in time and space.
- Spacetime-varying concurrent regression of house price quantiles on income quantiles can be rearranged as a functional additive model:

$$\log \frac{P^{-1}(s, t; q)}{Y^{-1}(s, t; q)} = \alpha(q) + \beta(s; q) + \gamma(t; q) + \delta(s, t; q). \quad (9)$$

- This can be fitted by the backfitting algorithm to obtain functional parameter estimates  $\hat{\alpha}$ ,  $\hat{\beta}$ ,  $\hat{\gamma}$  and  $\hat{\delta}$ .

# Regression Analysis

- One can examine the parameters directly as functions of  $q \in [0, 1]$  directly, but finite-dimensional summaries of them can also be useful for interpretation.
- Here, a Legendre polynomial expansion is utilised:

$$\hat{\alpha}(q) = \sum_{k=1}^{\infty} \hat{\alpha}_k \phi_k(q), \quad \text{where} \quad \hat{\alpha}_k = \langle \hat{\alpha}, \phi_k \rangle, \quad \text{etc.}, \quad (10)$$

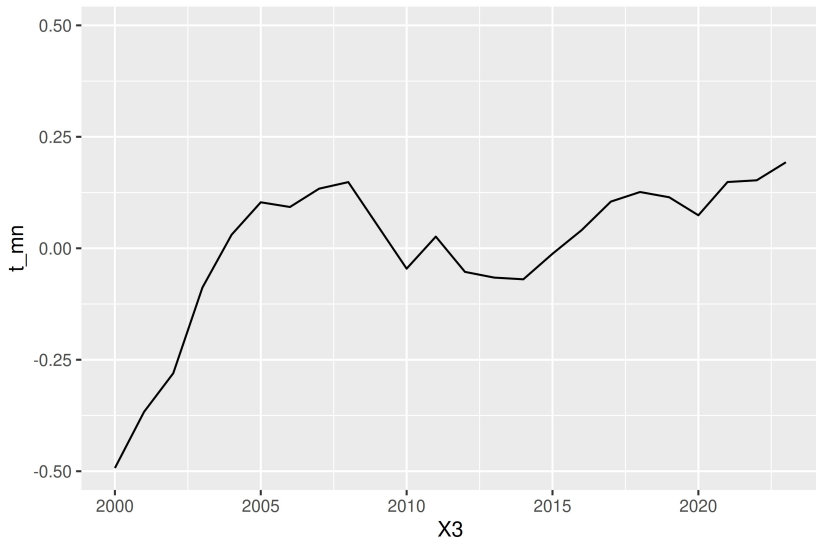
where  $\phi_k$  is the  $k$ -th order Legendre polynomial transformed to  $[0, 1]$  and normalised.

- The first three components are readily interpretable:
  1. the relative expense of housing across the income spectrum;
  2. how uniform house prices are compared to the income distribution;
  3. a skew towards either higher or lower priced homes.

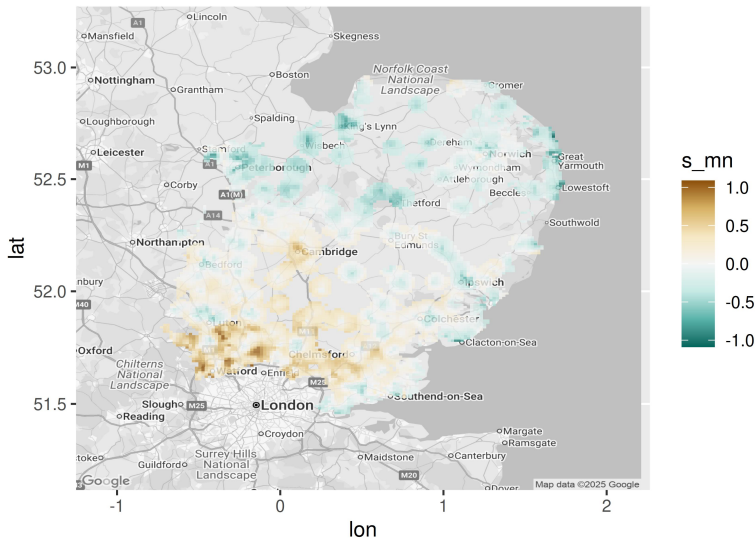
# Application

---

## East of England: $\hat{\gamma}_1$

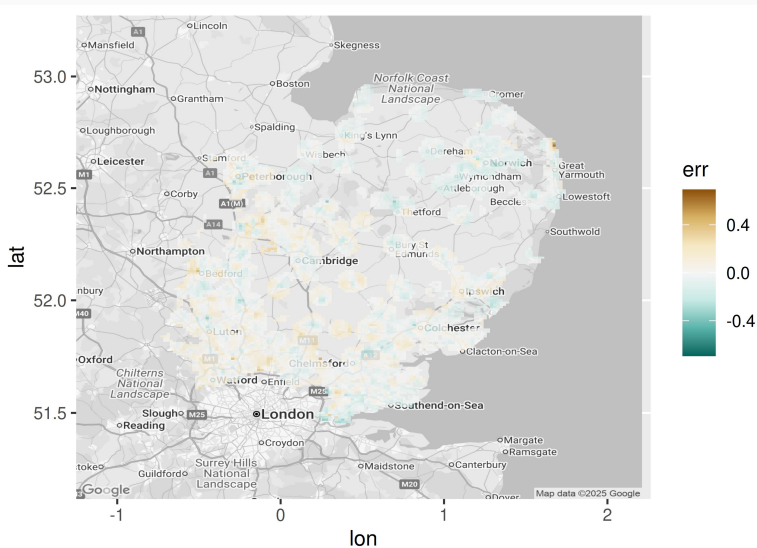


# East of England: $\hat{\beta}_1$



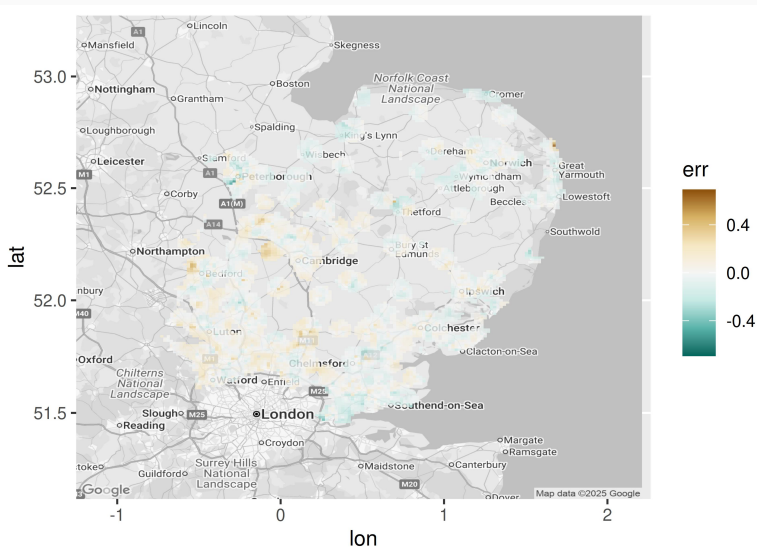
# East of England: $\hat{\delta}_1$

2000



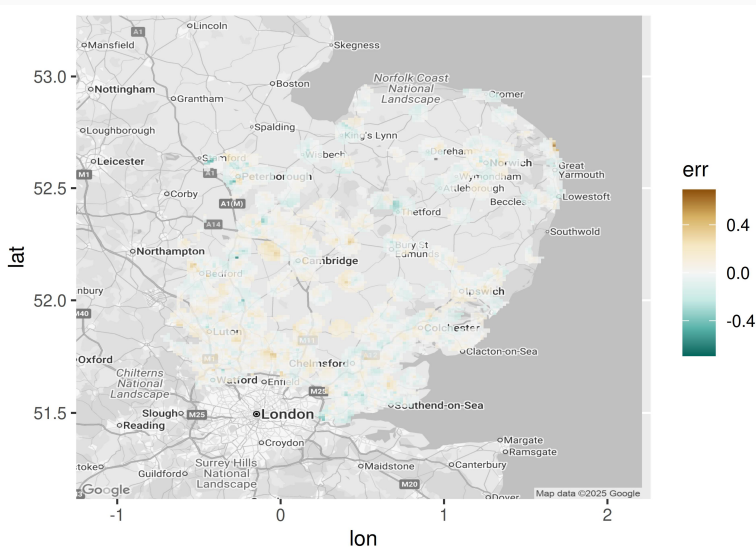
# East of England: $\hat{\delta}_1$

2001



# East of England: $\hat{\delta}_1$

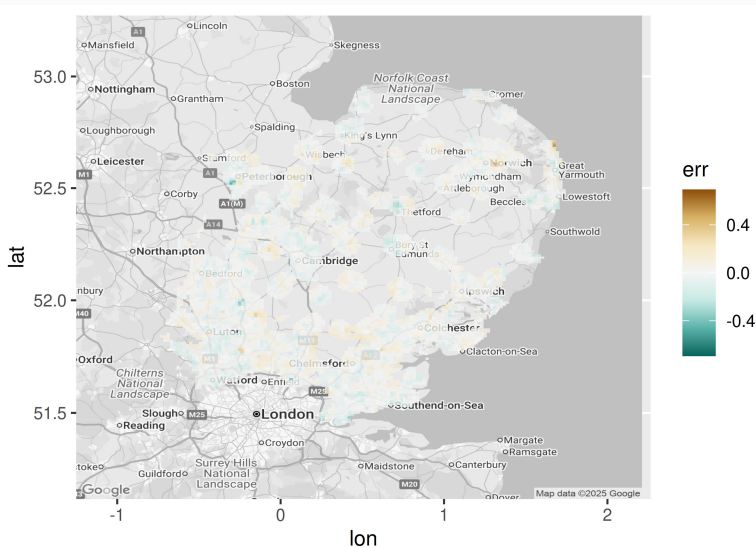
2002





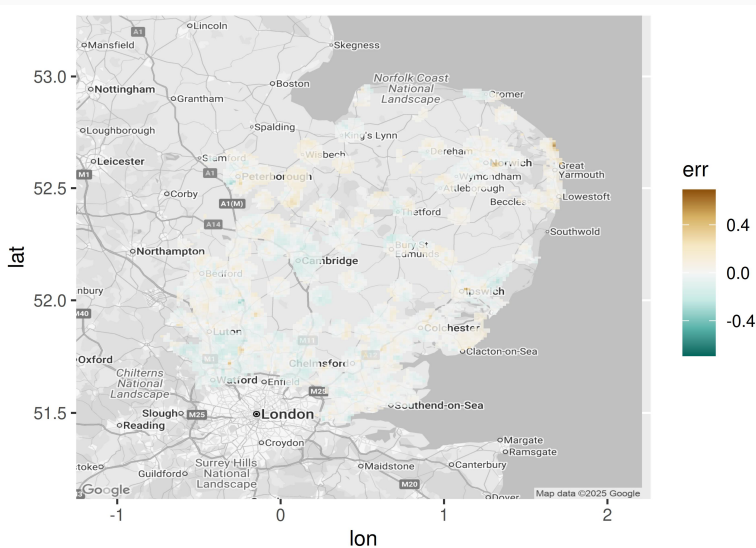
# East of England: $\hat{\delta}_1$

2003



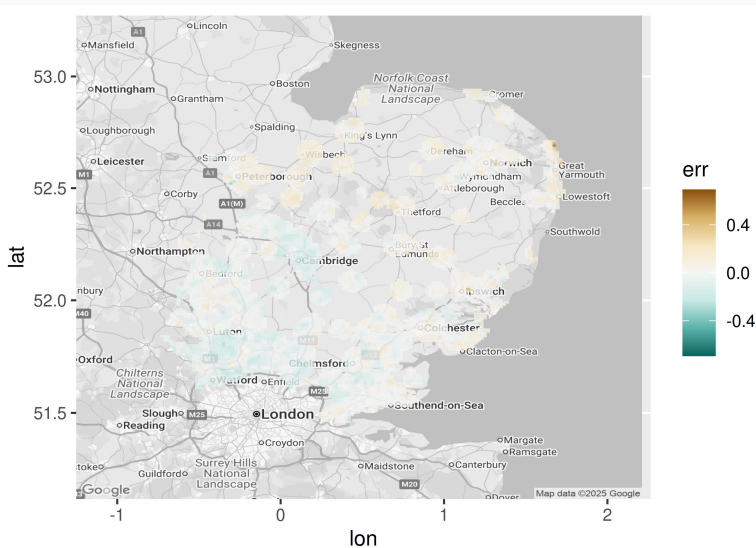
# East of England: $\hat{\delta}_1$

2004



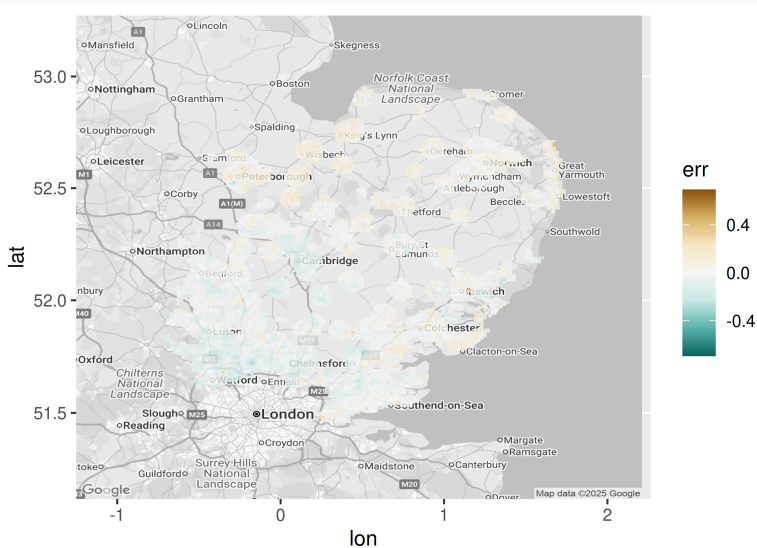
# East of England: $\hat{\delta}_1$

2005



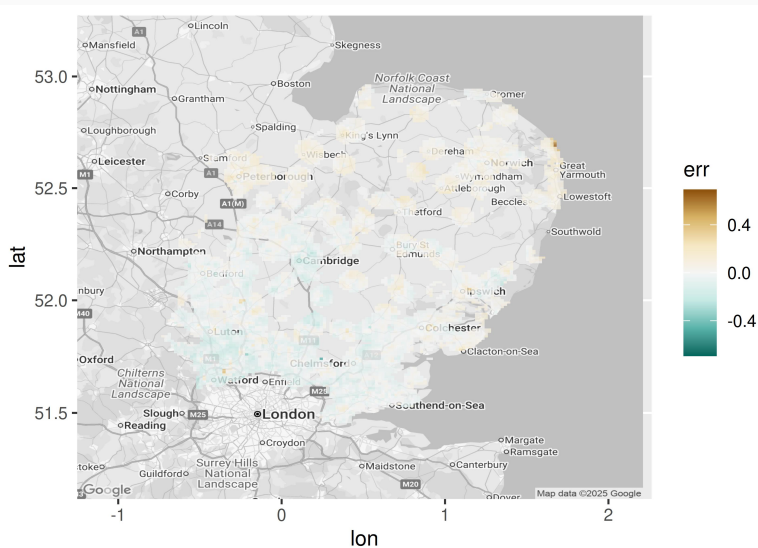
# East of England: $\hat{\delta}_1$

2006



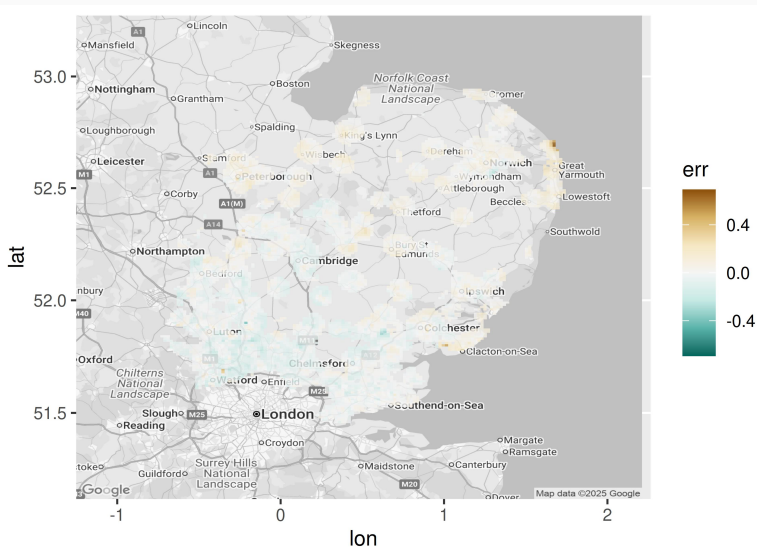
# East of England: $\hat{\delta}_1$

2007



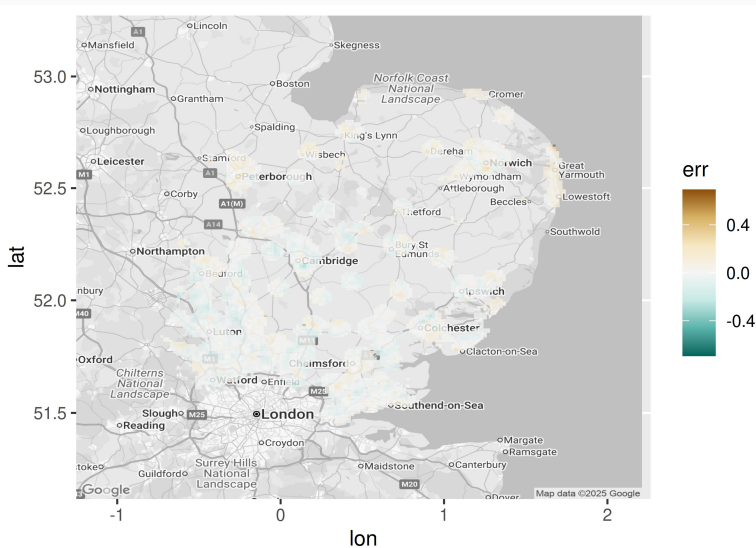
# East of England: $\hat{\delta}_1$

2008



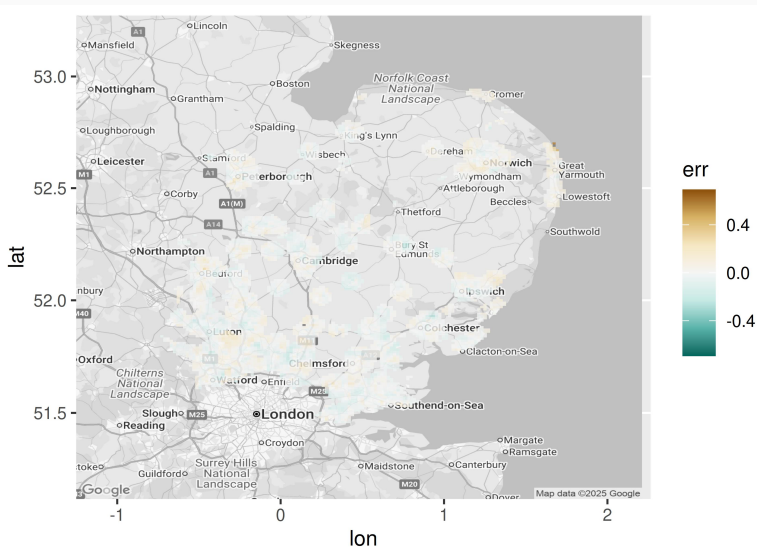
# East of England: $\hat{\delta}_1$

2010



# East of England: $\hat{\delta}_1$

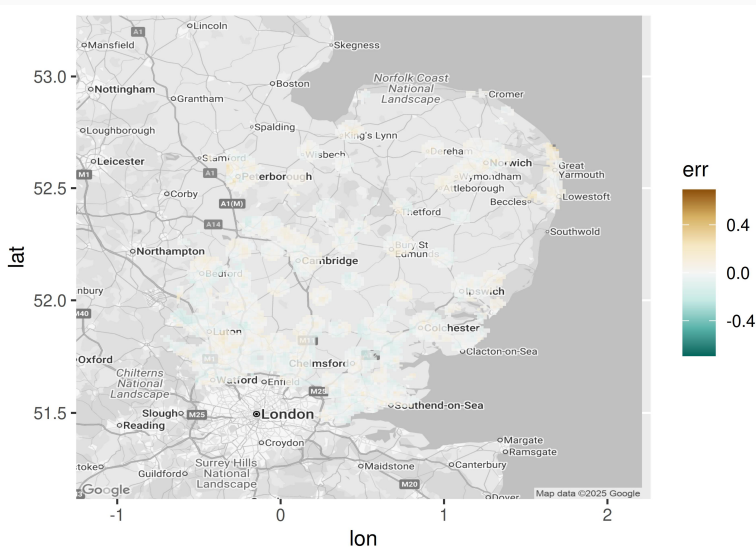
2011





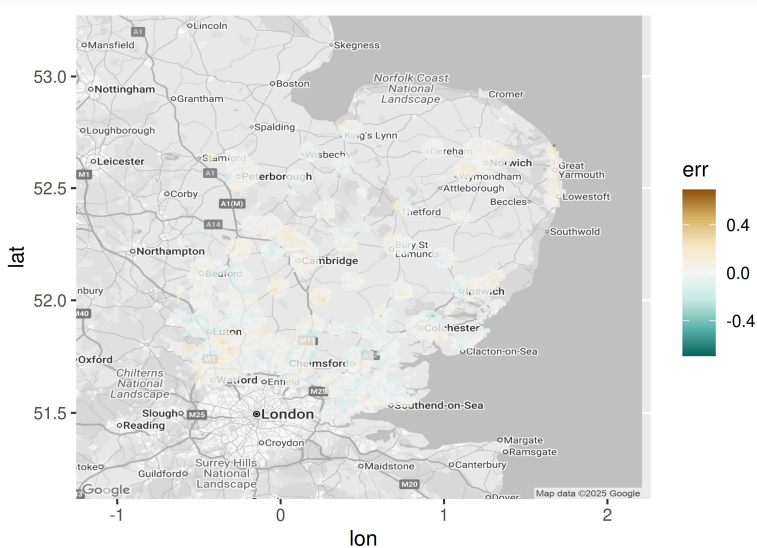
# East of England: $\hat{\delta}_1$

2012



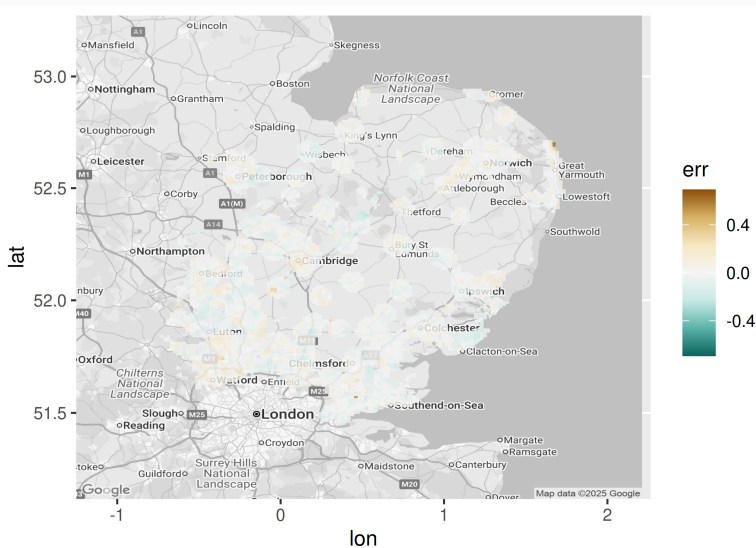
# East of England: $\hat{\delta}_1$

2013



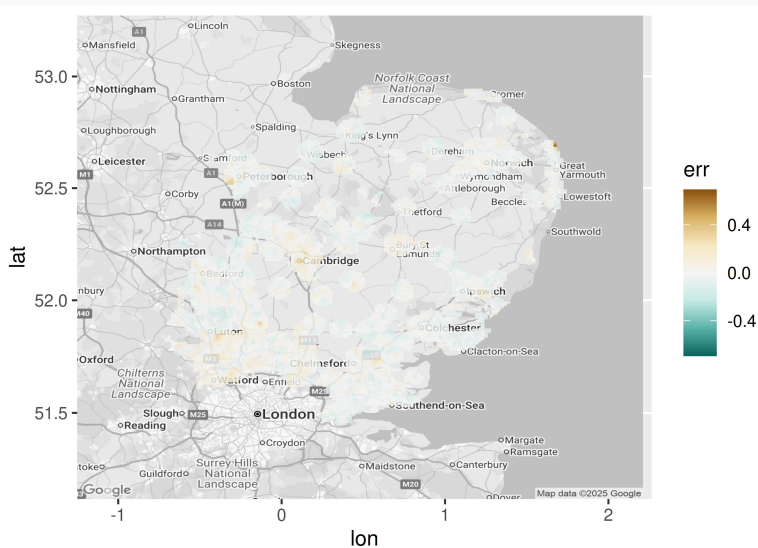
# East of England: $\hat{\delta}_1$

2014



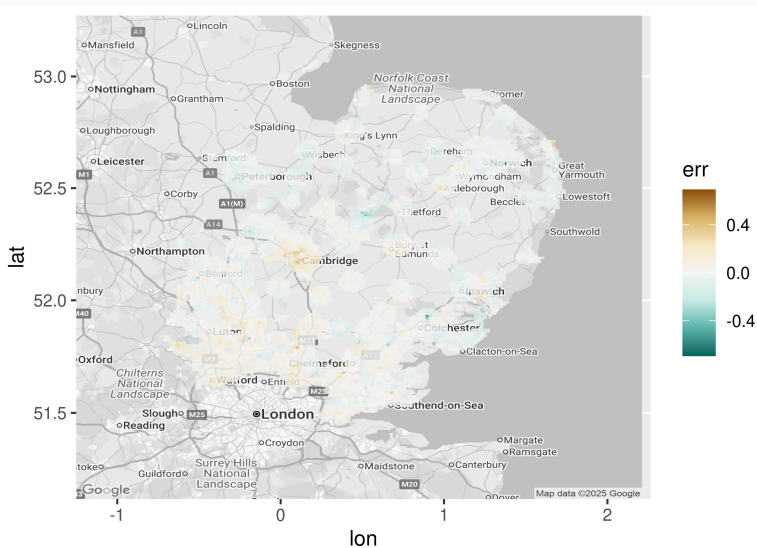
# East of England: $\hat{\delta}_1$

2015



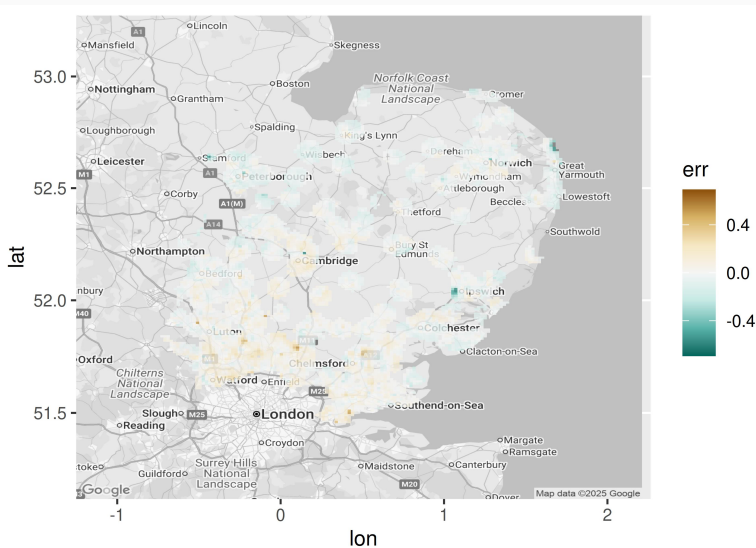
# East of England: $\hat{\delta}_1$

2016



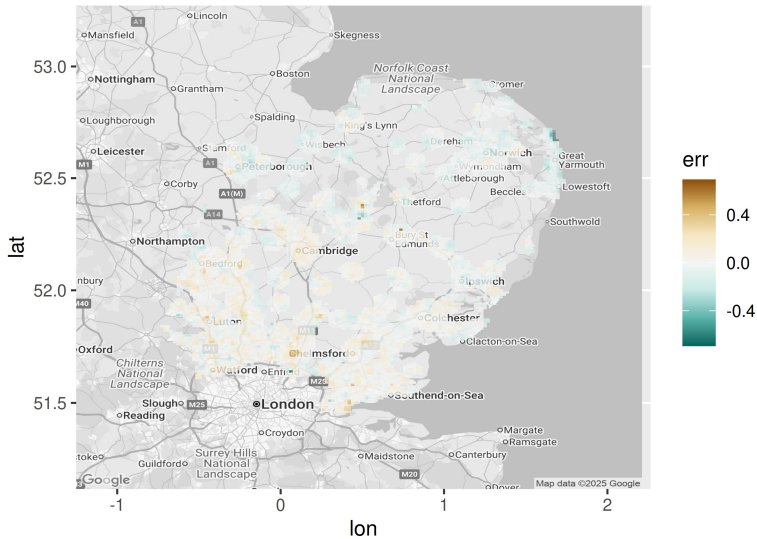
# East of England: $\hat{\delta}_1$

2017



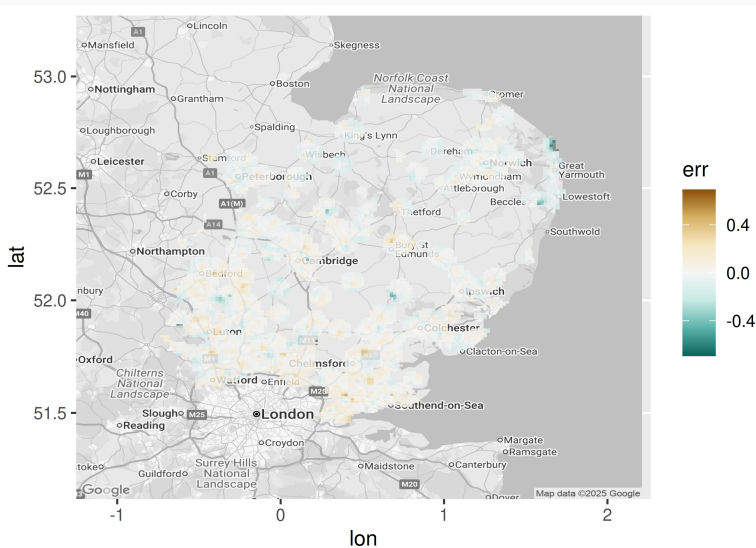
**East of England:**  $\hat{\delta}_1$

2018



# East of England: $\hat{\delta}_1$

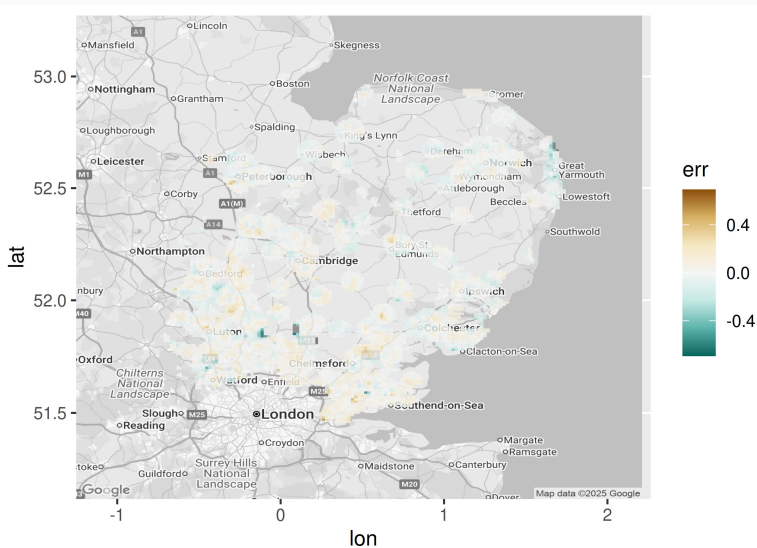
2019





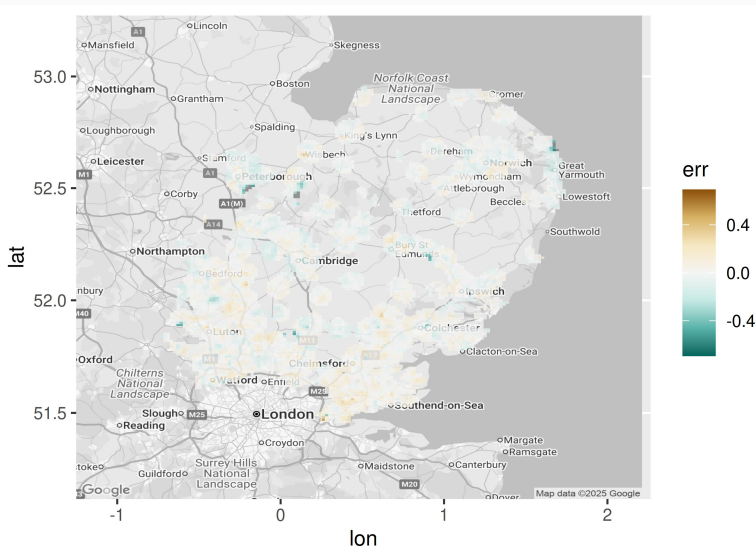
# East of England: $\hat{\delta}_1$

2020



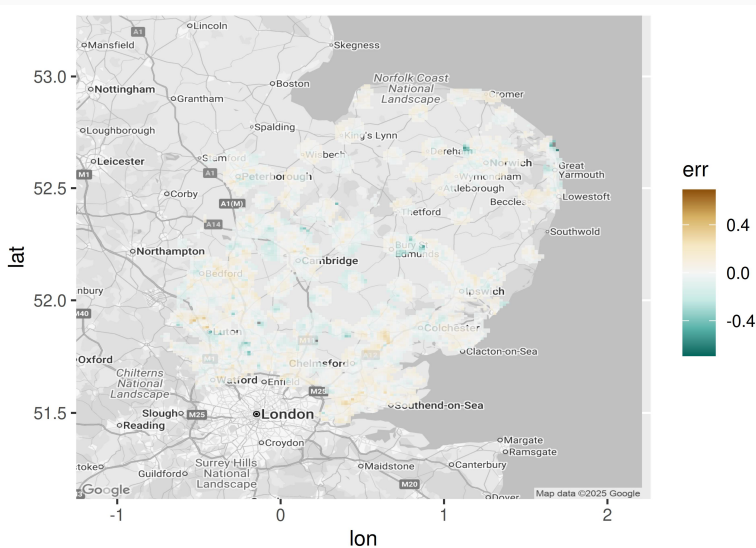
# East of England: $\hat{\delta}_1$

2021



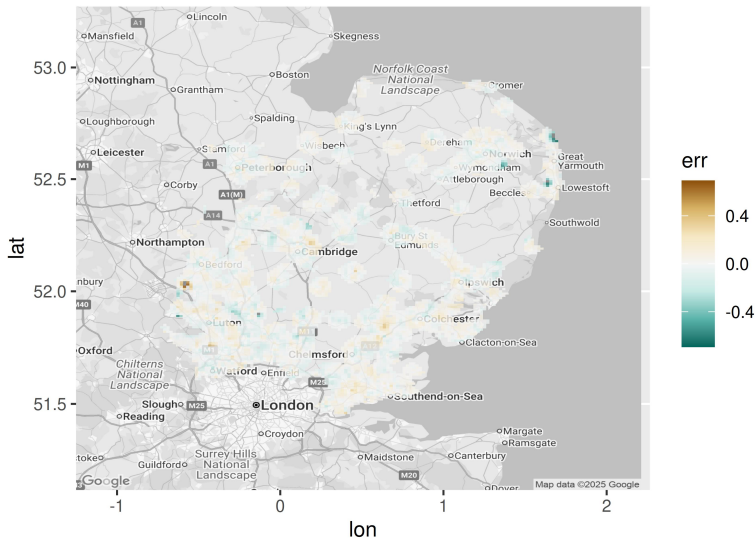
# East of England: $\hat{\delta}_1$

2022

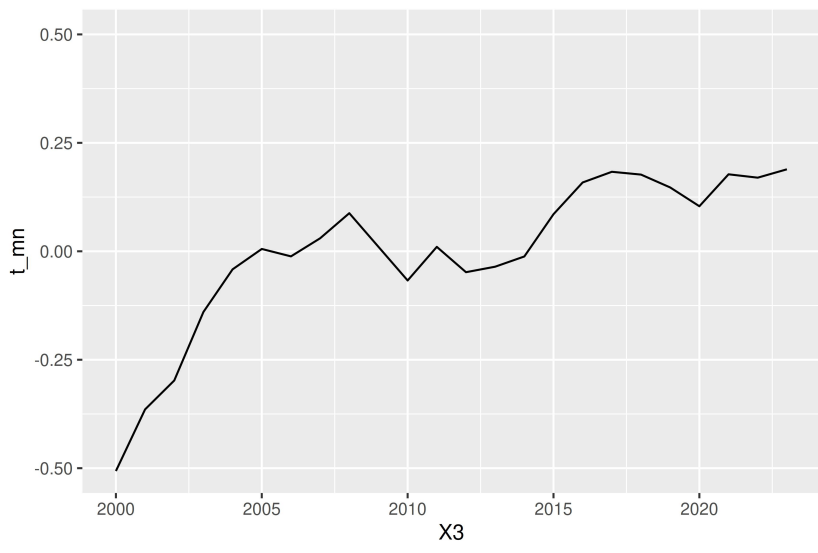


# East of England: $\hat{\delta}_1$

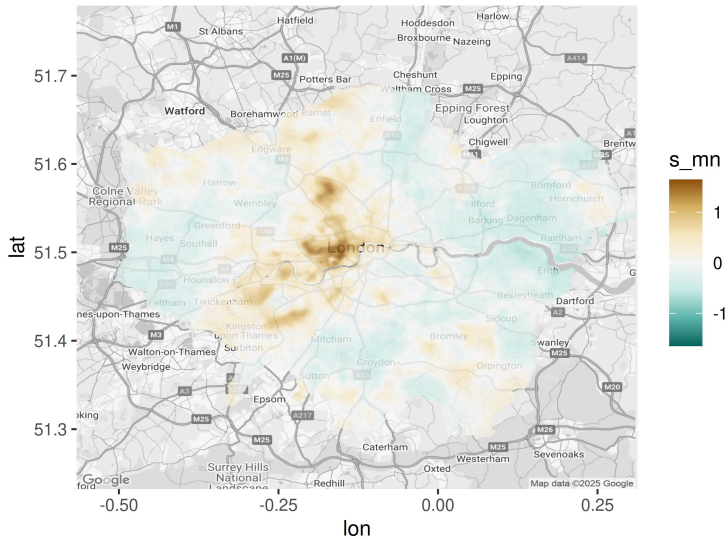
2023



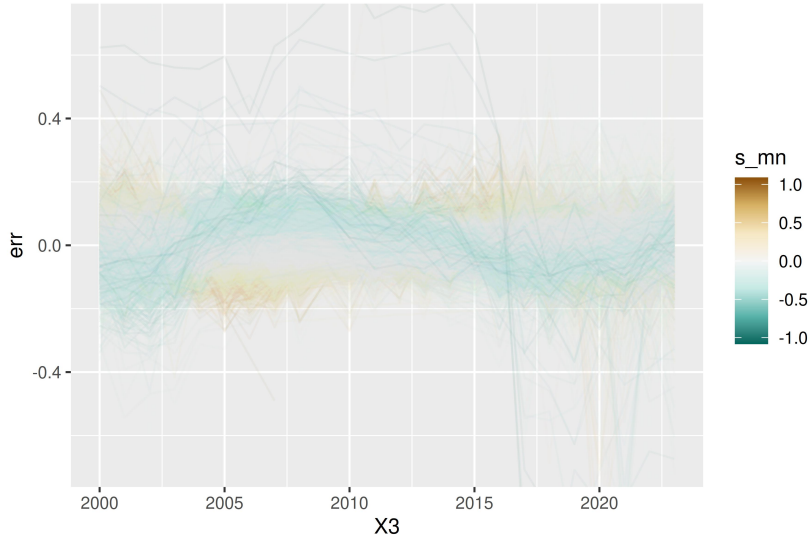
London:  $\hat{\gamma}_1$



London:  $\hat{\beta}_1$

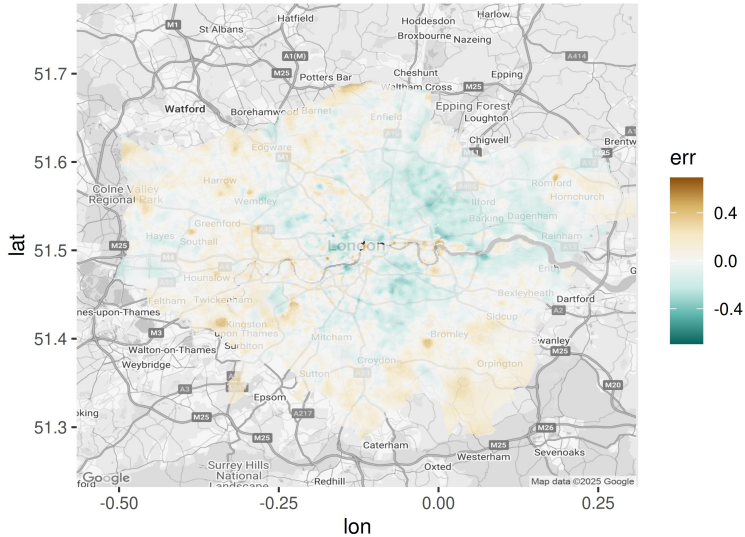


# East of England: Gentrification



London:  $\hat{\delta}_1$

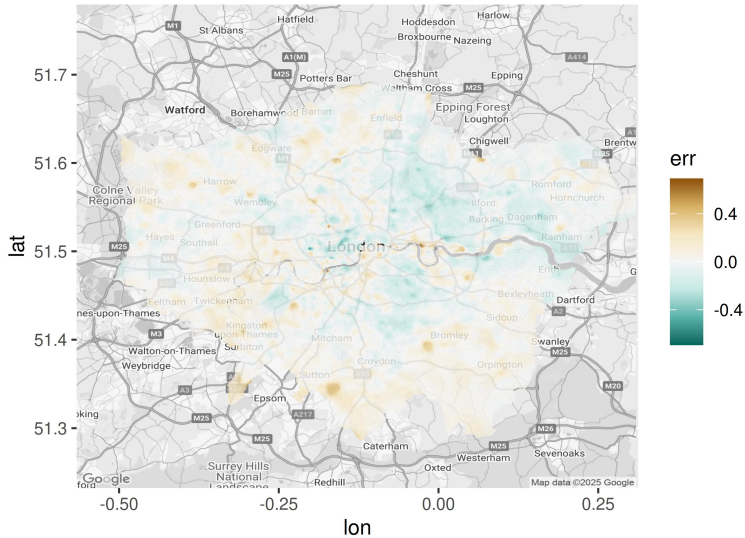
2000



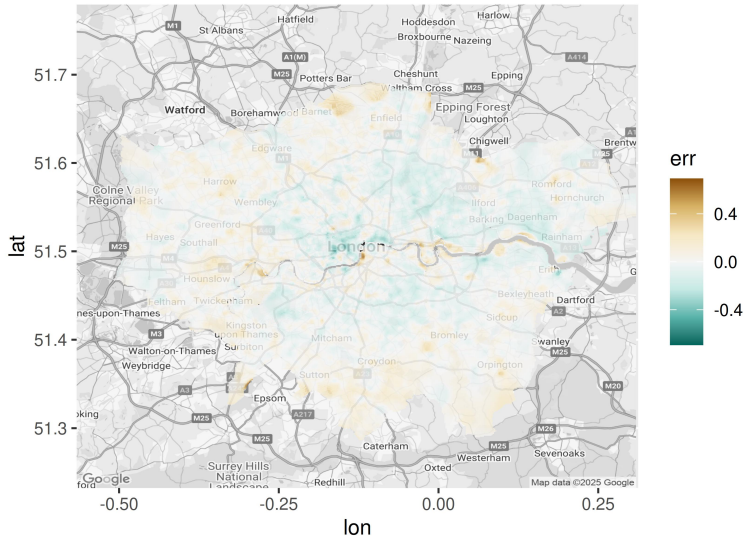


London:  $\hat{\delta}_1$

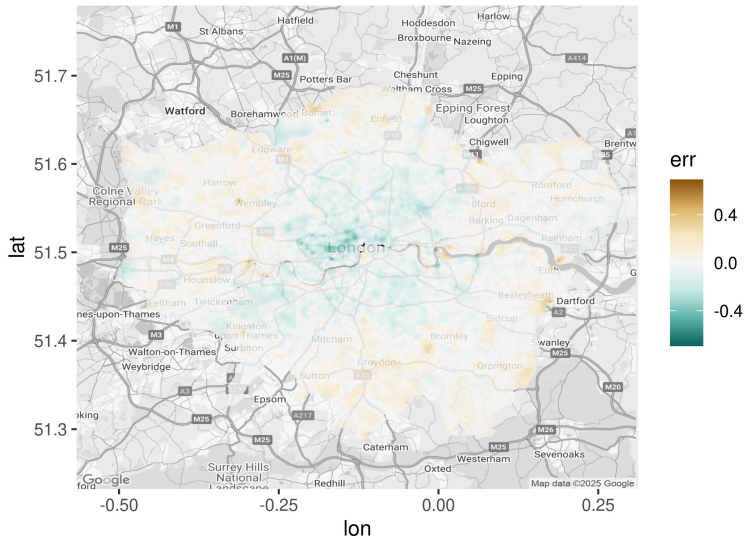
2001



2002

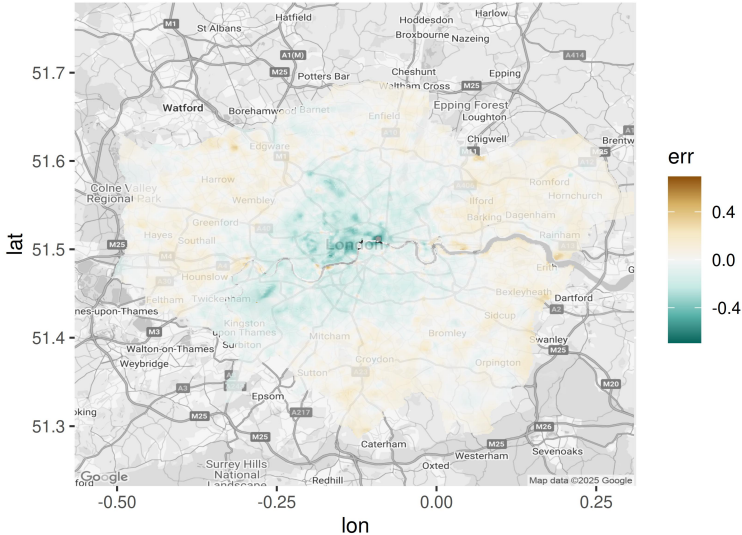


2003

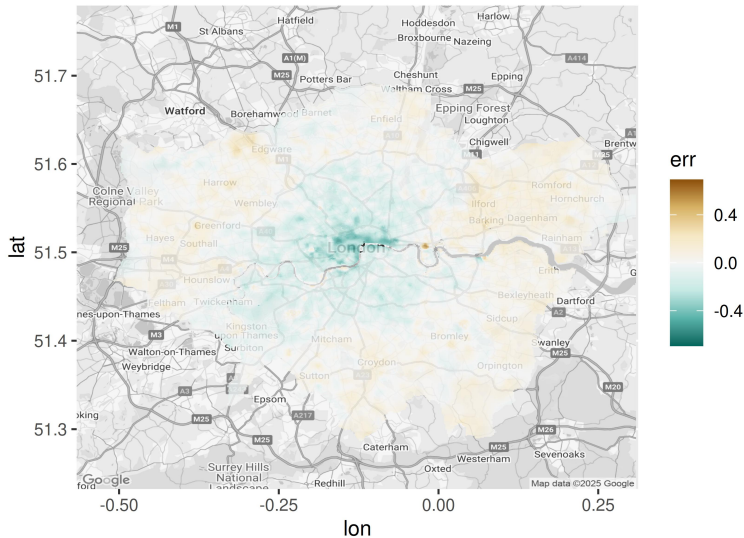


**London:**  $\hat{\delta}_1$

2004

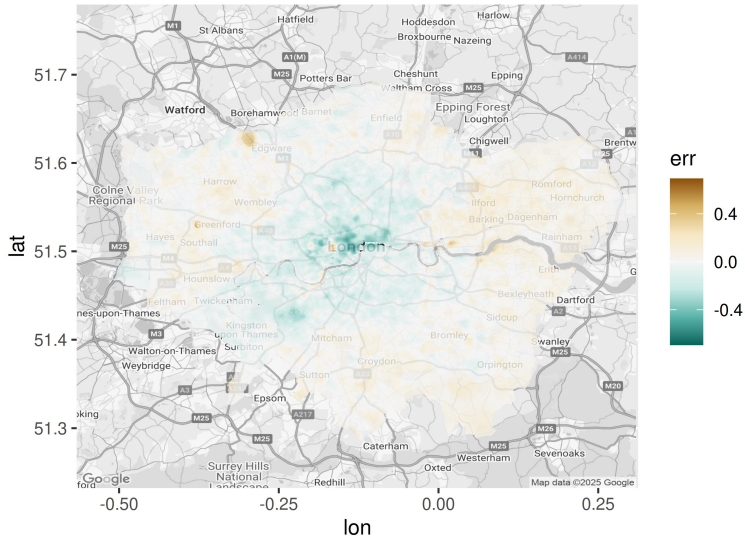


2005



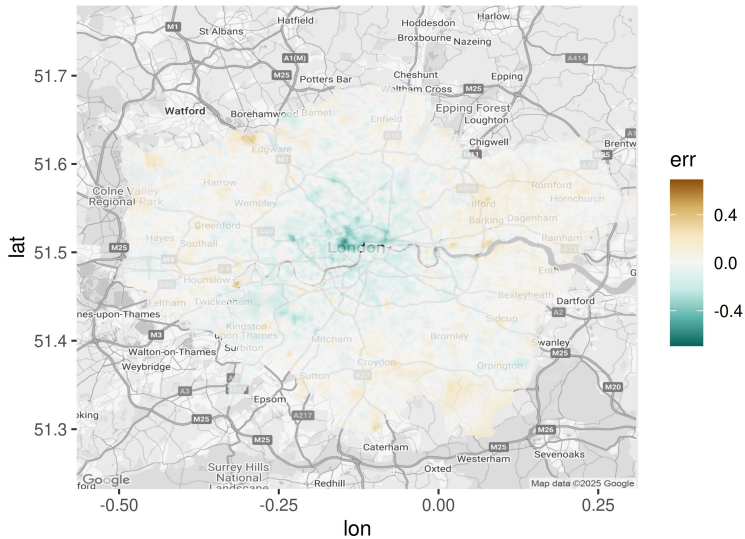
London:  $\hat{\delta}_1$

2006

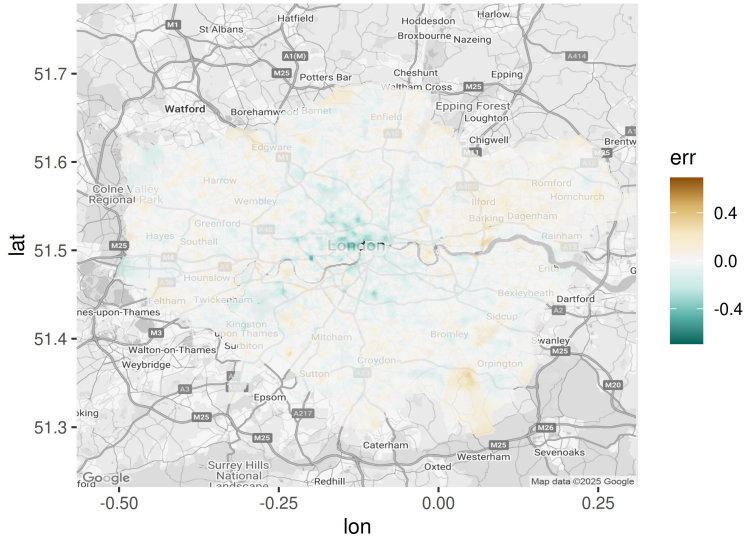


London:  $\hat{\delta}_1$

2007

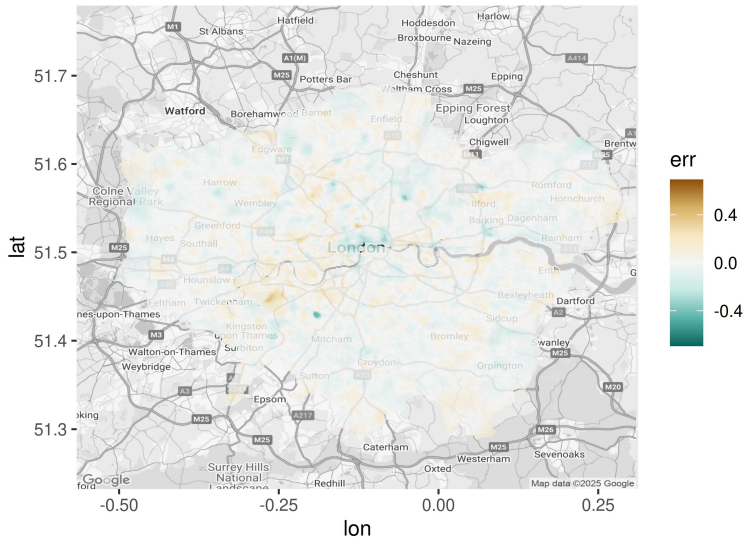


2008

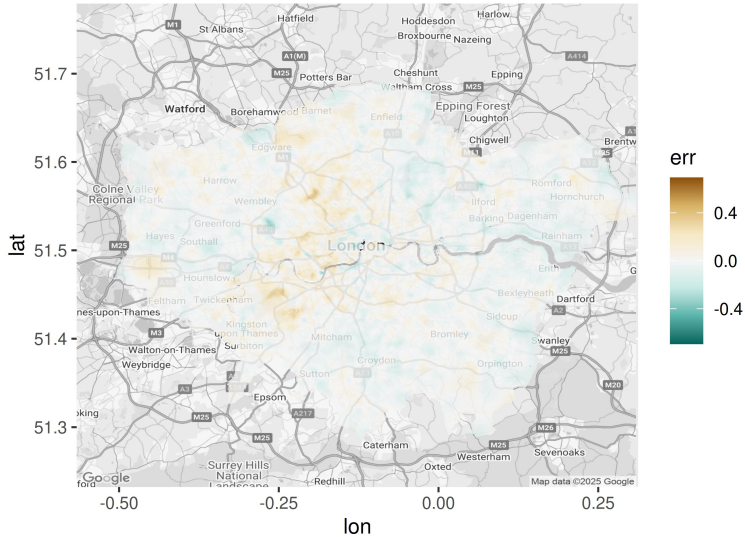




2010

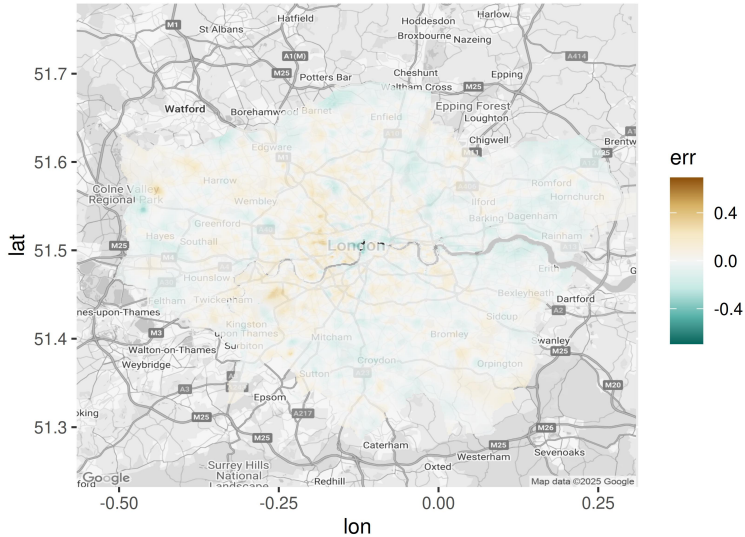


2011



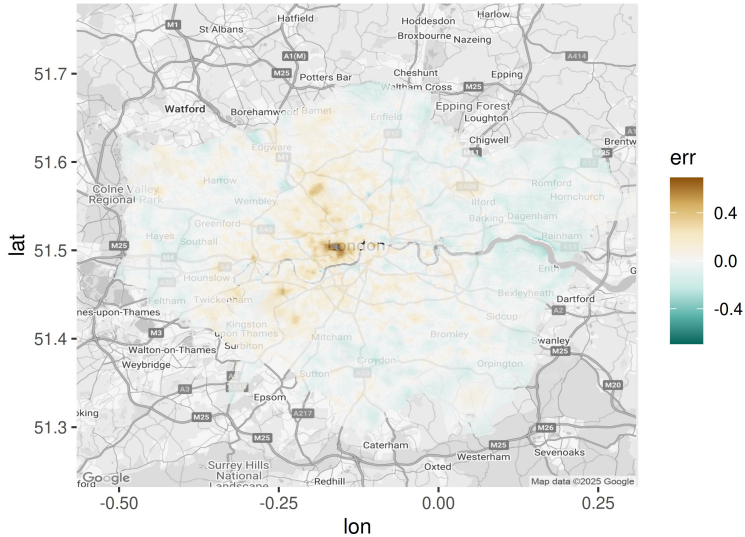
London:  $\hat{\delta}_1$

2012



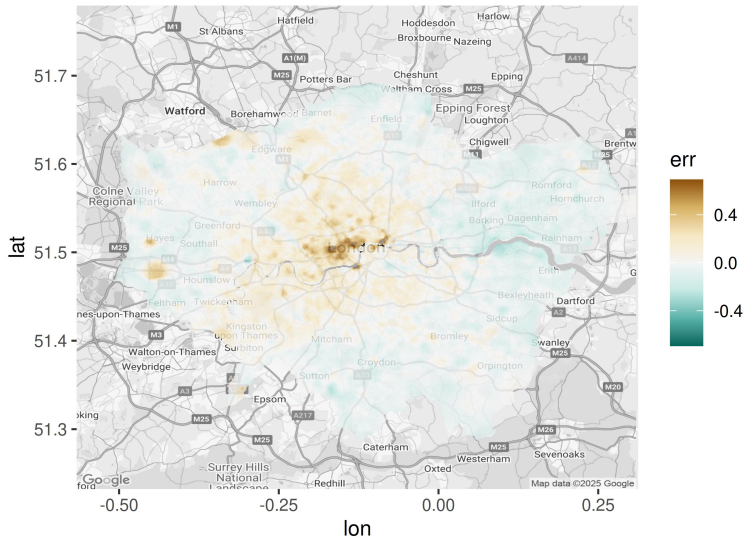
London:  $\hat{\delta}_1$

2013



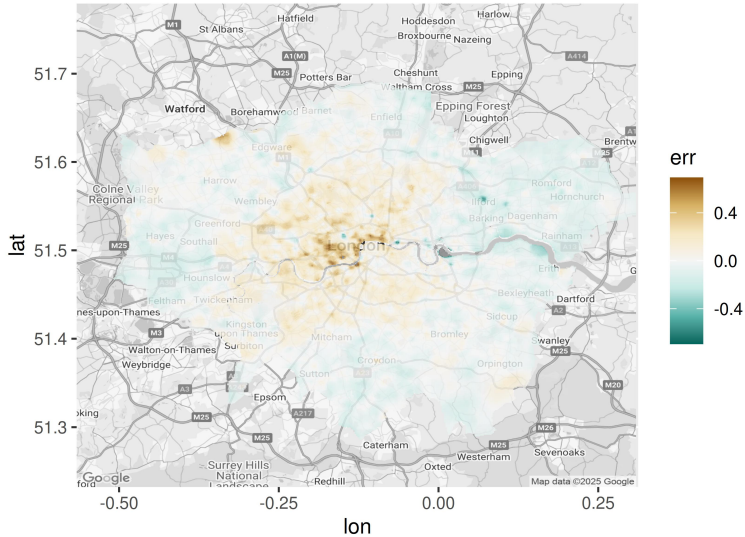
London:  $\hat{\delta}_1$

2014



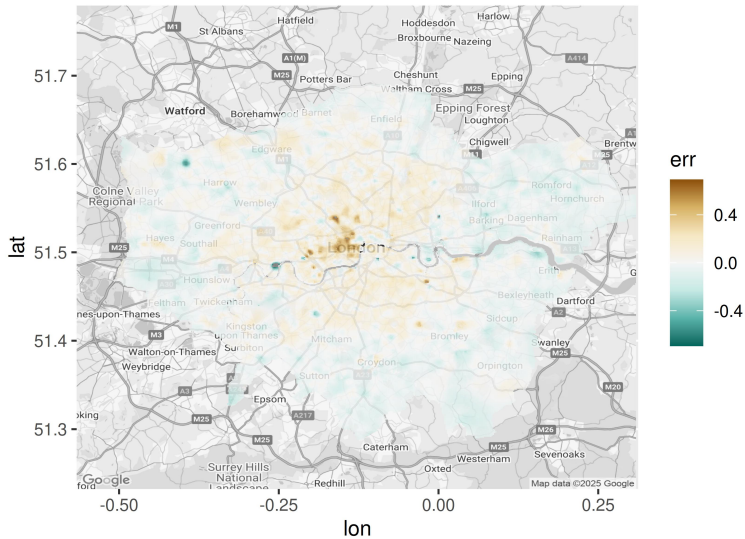
London:  $\hat{\delta}_1$

2015



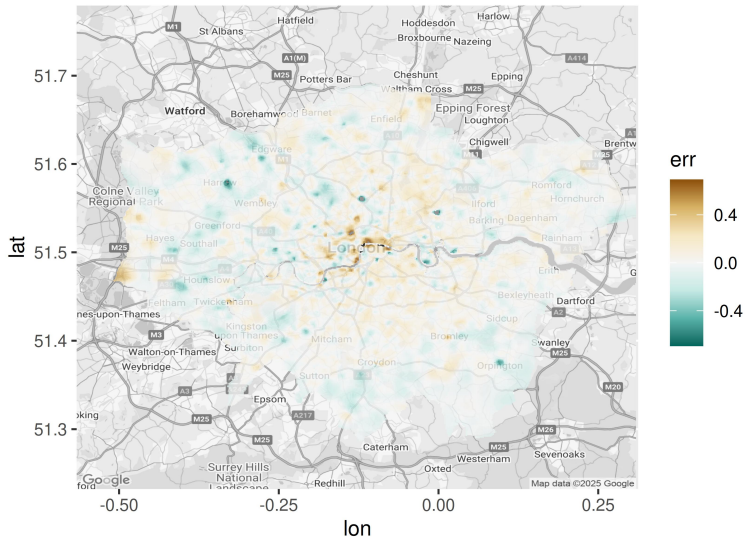
London:  $\hat{\delta}_1$

2016



London:  $\hat{\delta}_1$

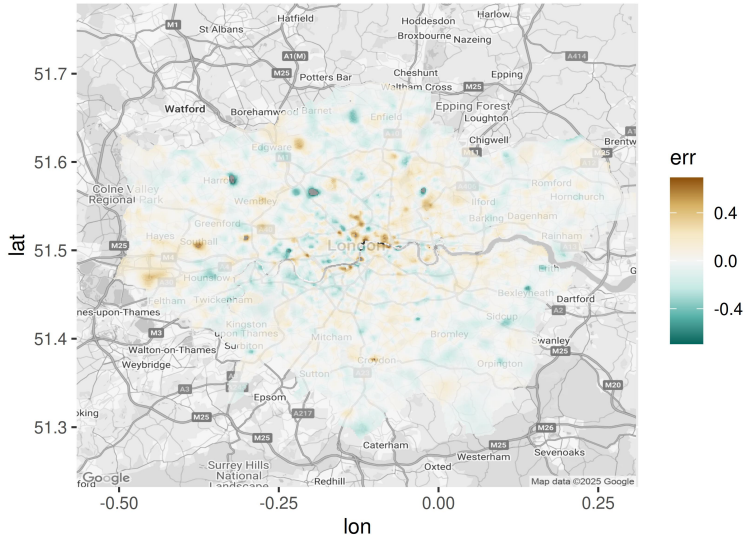
2017





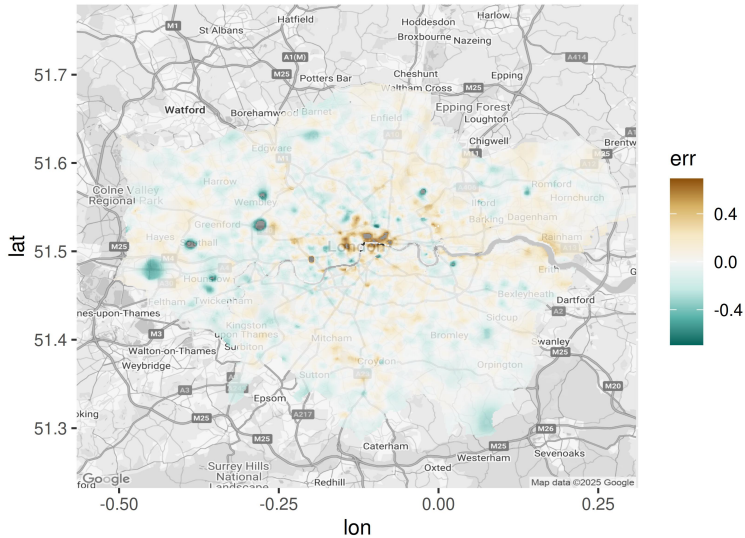
London:  $\hat{\delta}_1$

2018



London:  $\hat{\delta}_1$

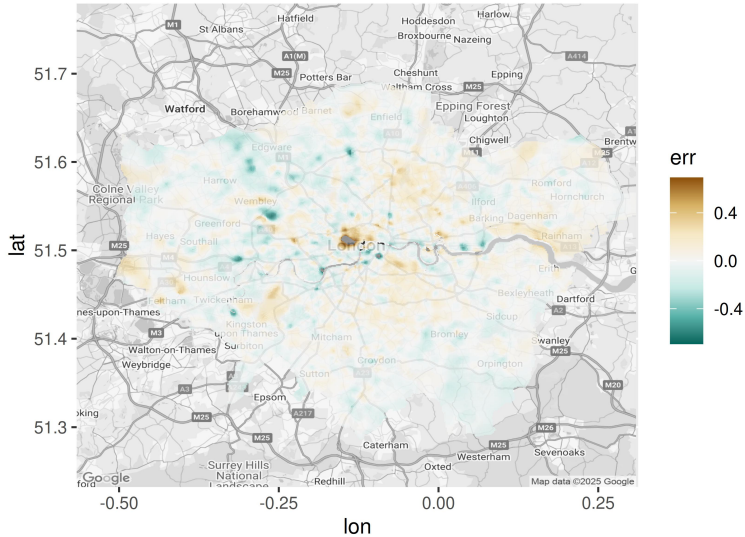
2019





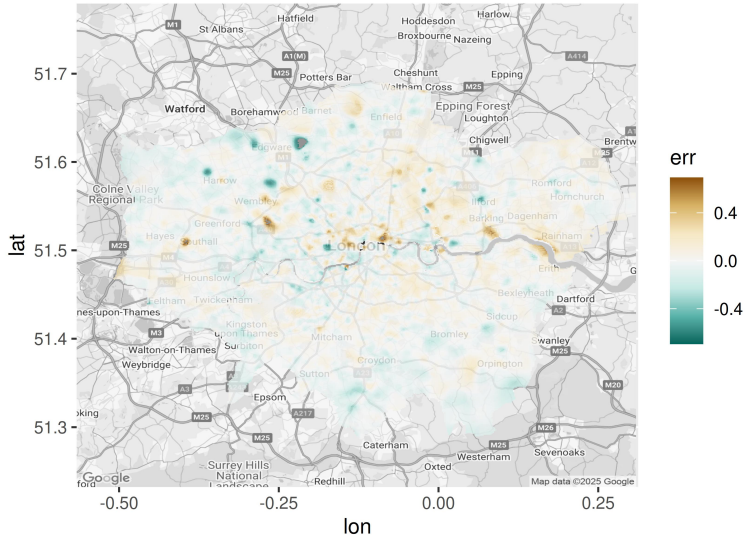
London:  $\hat{\delta}_1$

2021



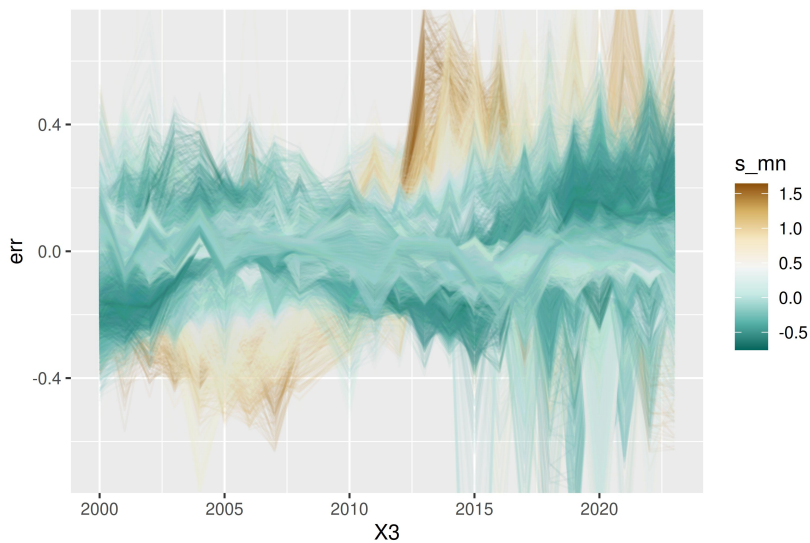
London:  $\hat{\delta}_1$

2022

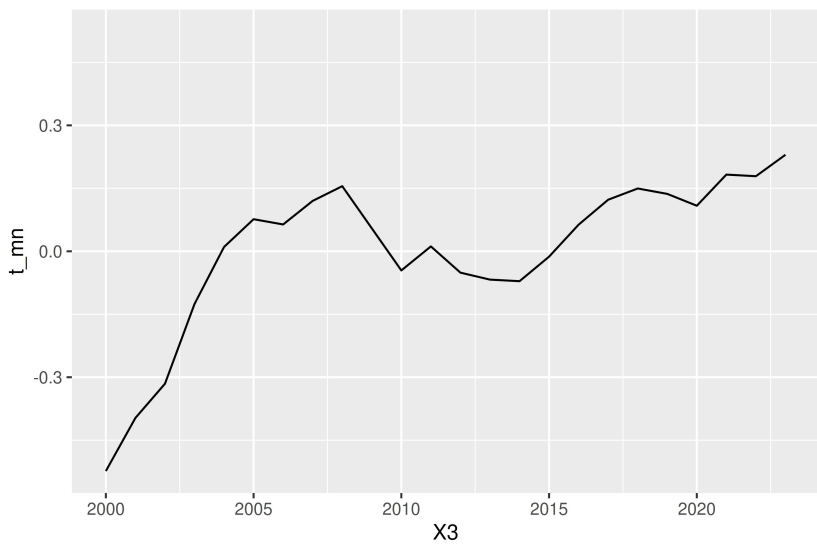




# London: Gentrification

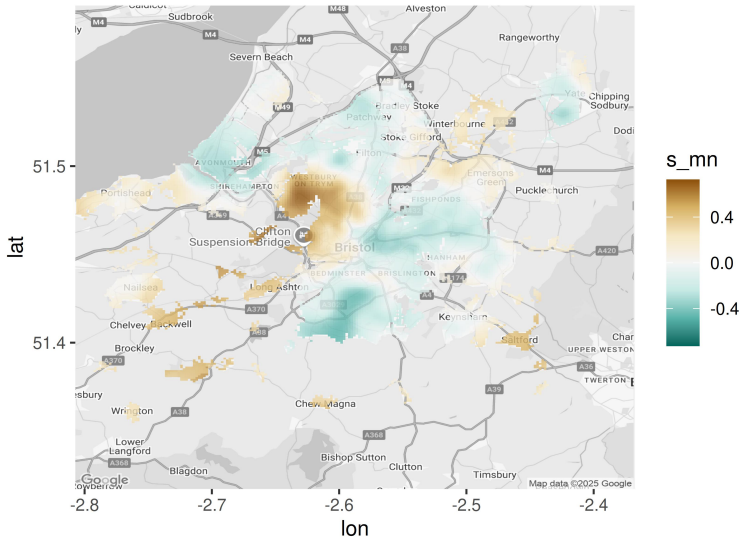


## Bristol and Surroundings: $\hat{\gamma}_1$



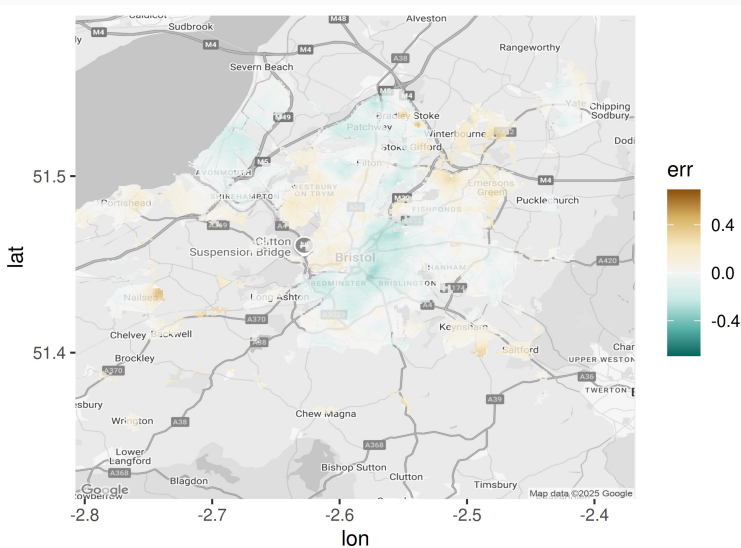


# Bristol and Surroundings: $\hat{\beta}_1$



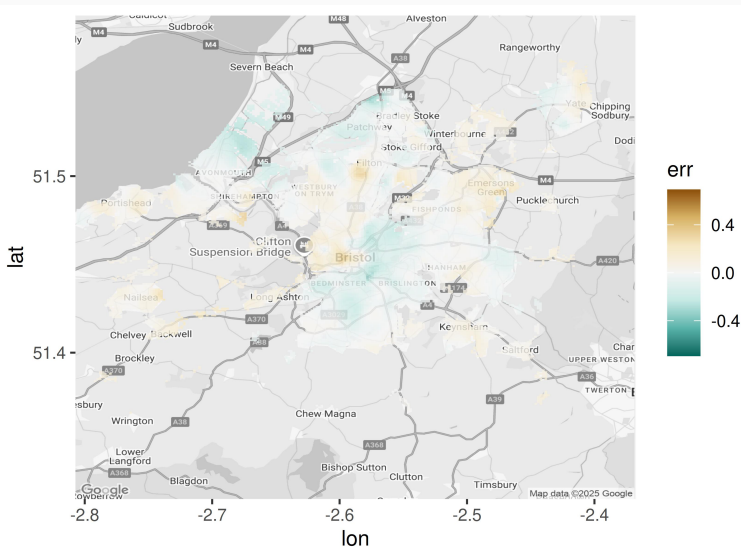
# Bristol and Surroundings: $\hat{\delta}_1$

2000



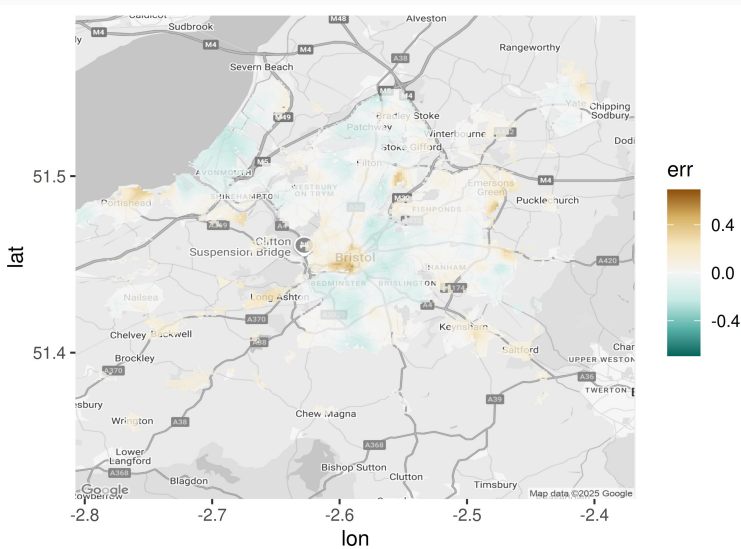
# Bristol and Surroundings: $\hat{\delta}_1$

2001



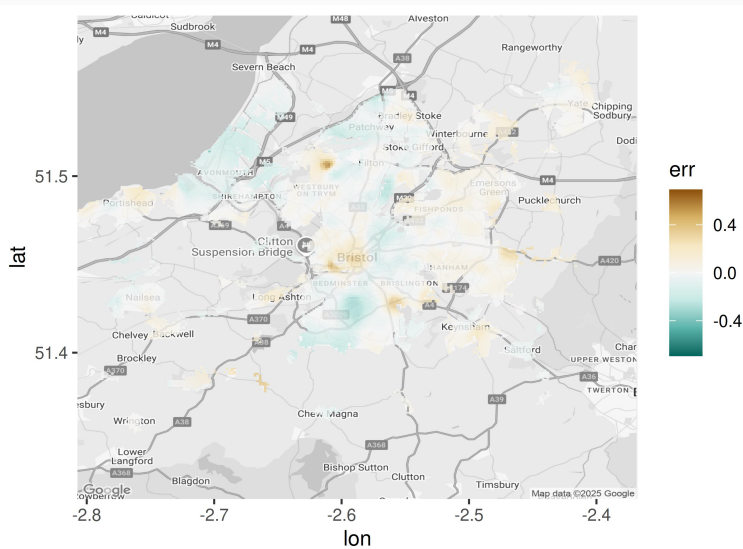
# Bristol and Surroundings: $\hat{\delta}_1$

2002



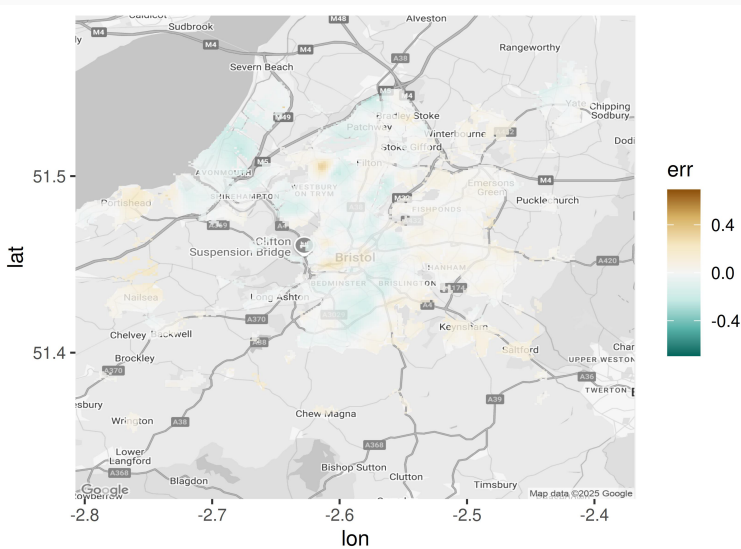
# Bristol and Surroundings: $\hat{\delta}_1$

2003



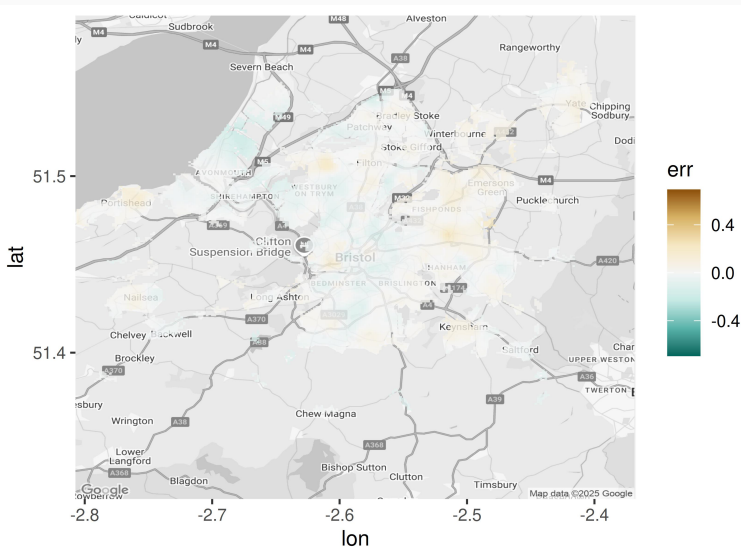
# Bristol and Surroundings: $\hat{\delta}_1$

2004



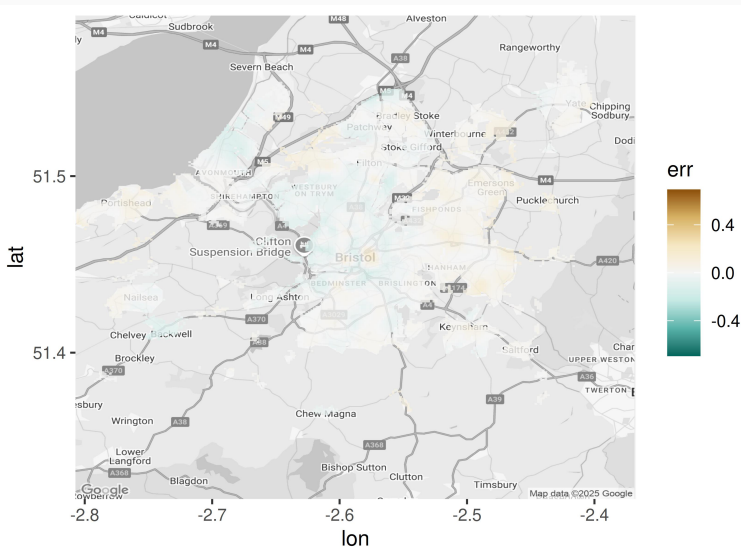
# Bristol and Surroundings: $\hat{\delta}_1$

2005



# Bristol and Surroundings: $\hat{\delta}_1$

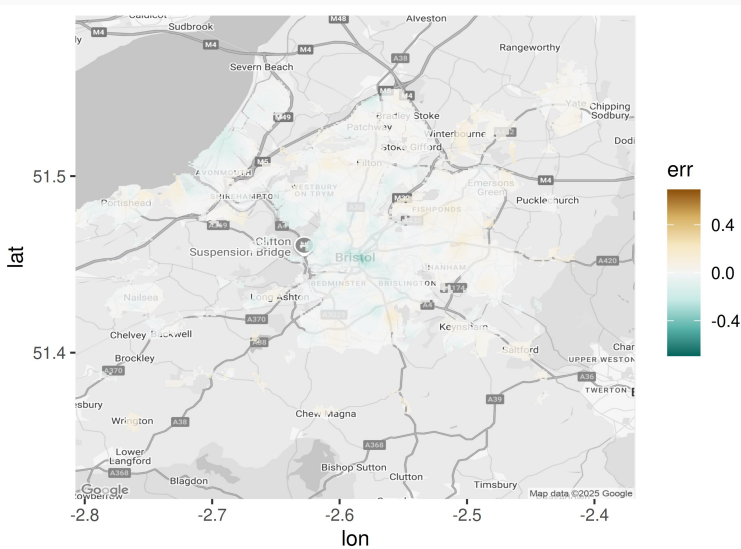
2006





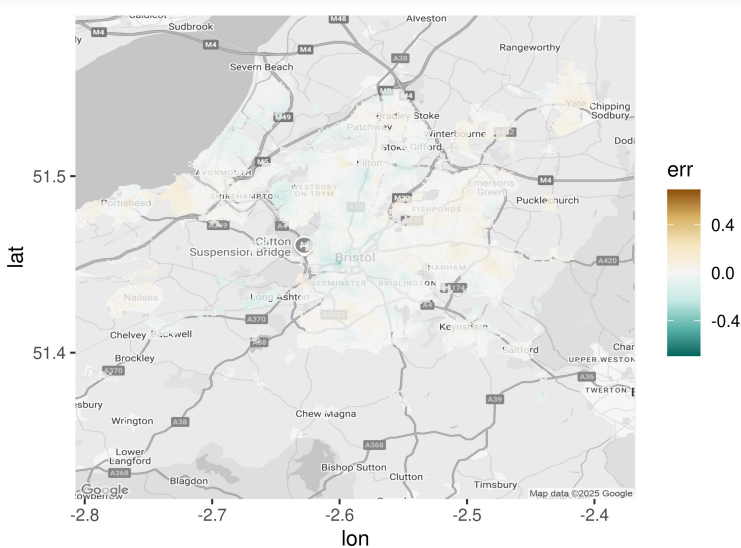
# Bristol and Surroundings: $\hat{\delta}_1$

2007



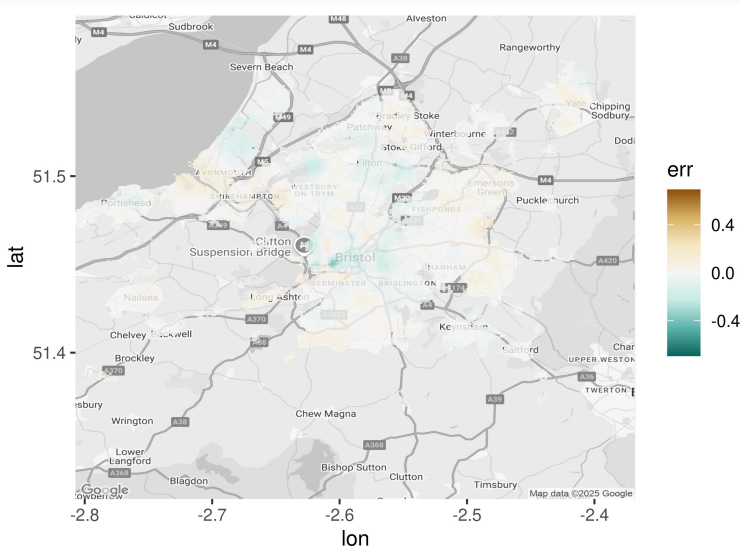
# Bristol and Surroundings: $\hat{\delta}_1$

2008



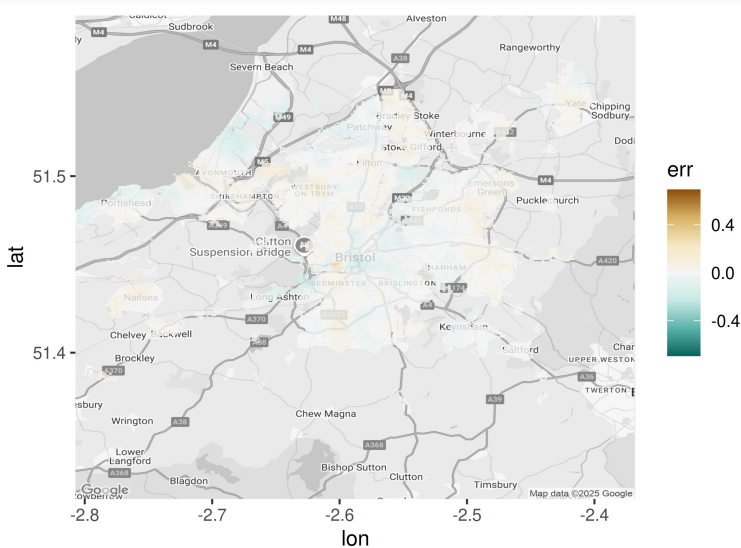
# Bristol and Surroundings: $\hat{\delta}_1$

2010



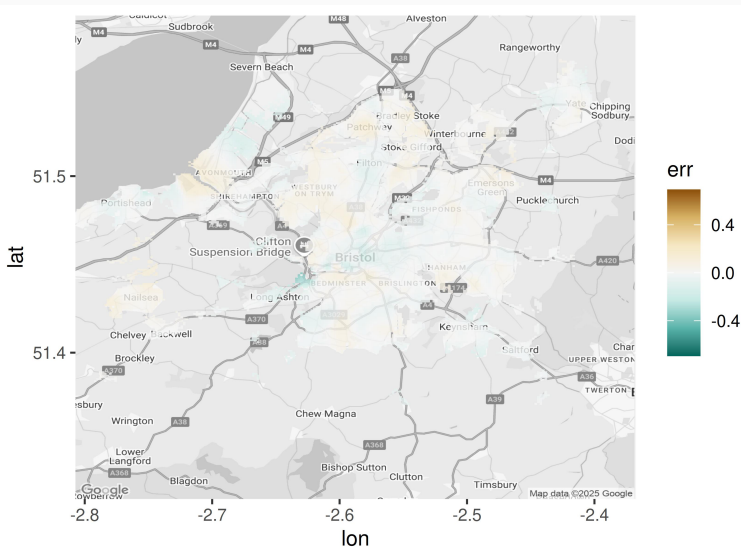
# Bristol and Surroundings: $\hat{\delta}_1$

2011



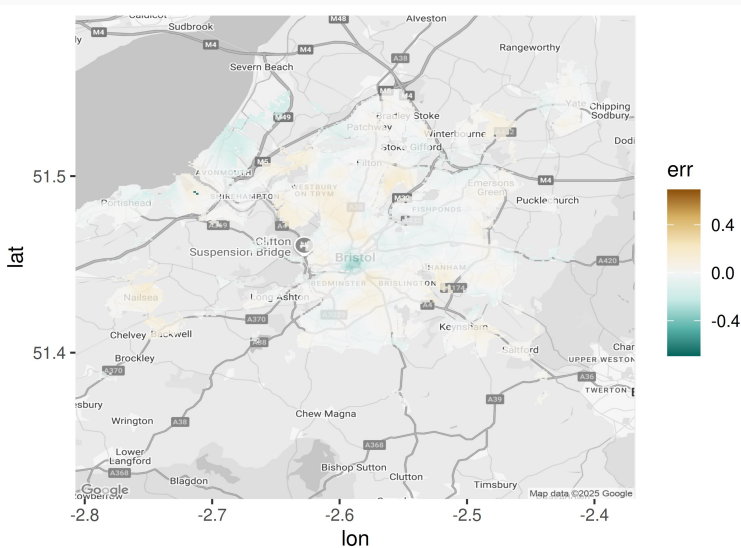
# Bristol and Surroundings: $\hat{\delta}_1$

2012



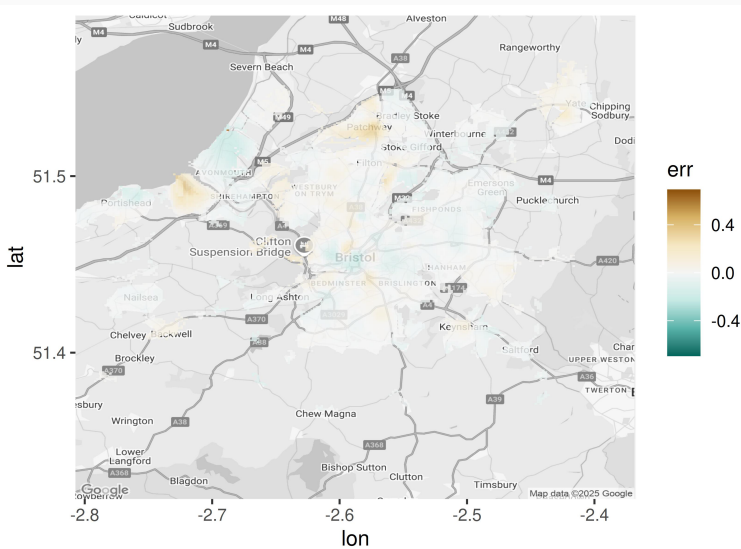
# Bristol and Surroundings: $\hat{\delta}_1$

2013



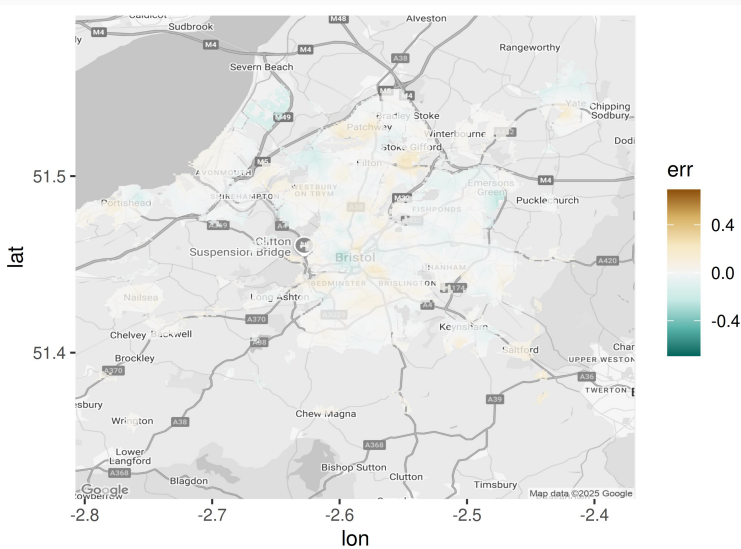
# Bristol and Surroundings: $\hat{\delta}_1$

2014



# Bristol and Surroundings: $\hat{\delta}_1$

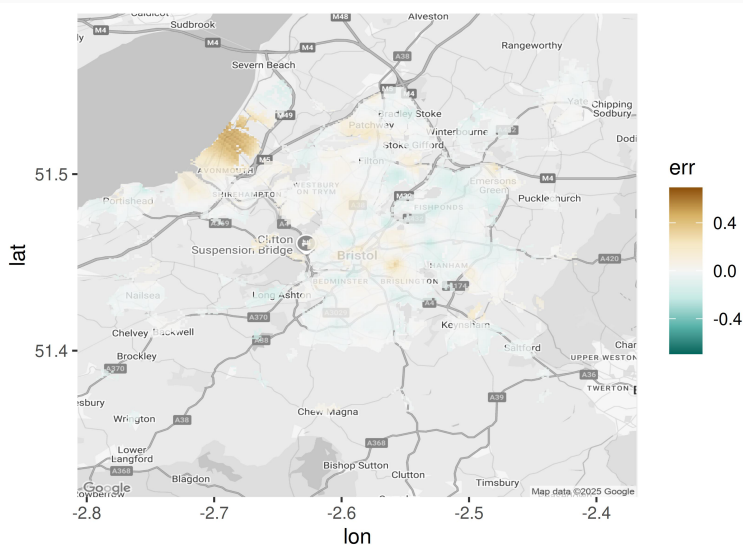
2015





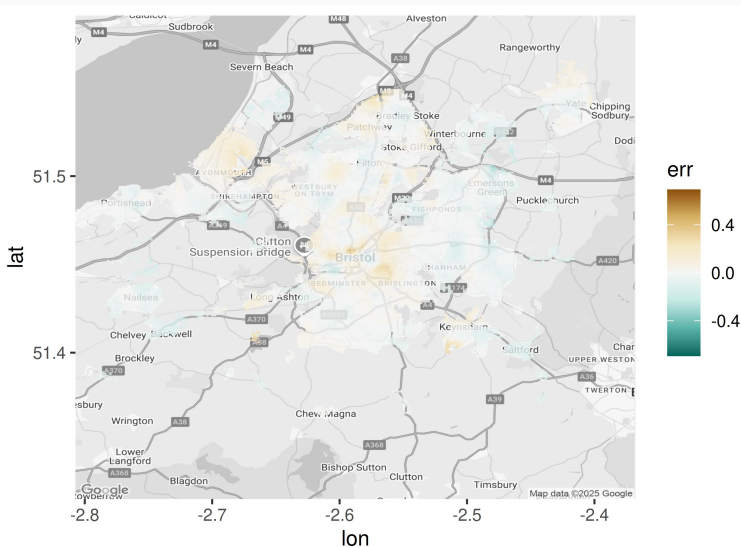
# Bristol and Surroundings: $\hat{\delta}_1$

2016



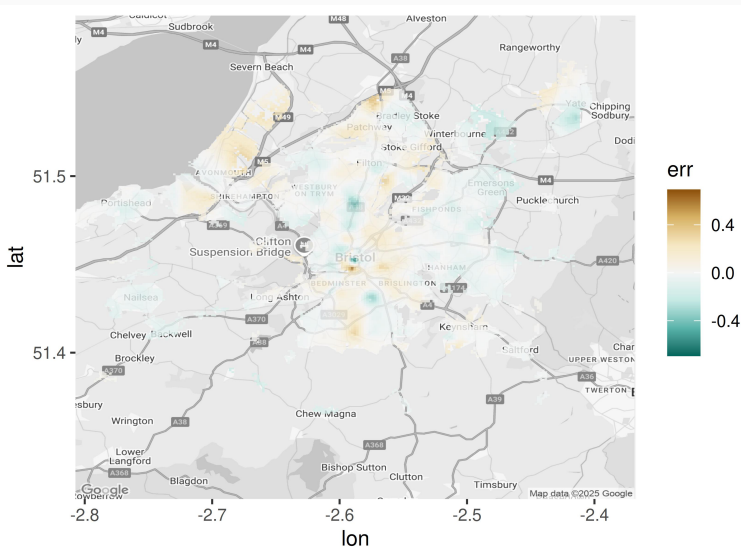
# Bristol and Surroundings: $\hat{\delta}_1$

2017



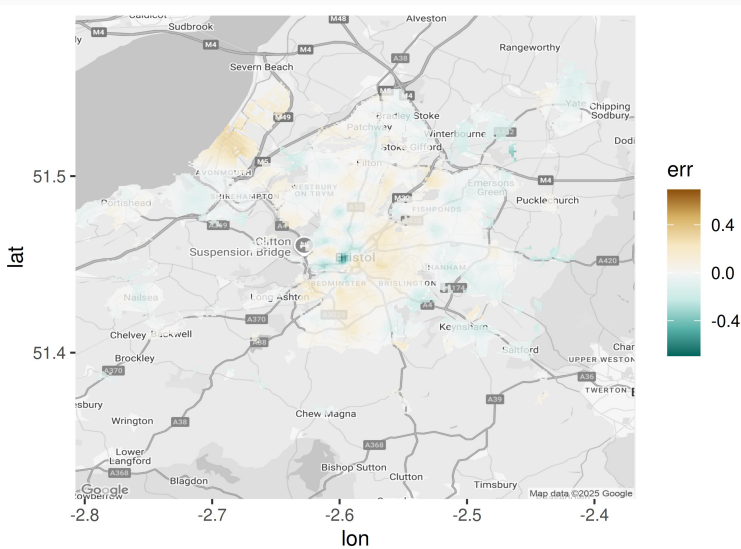
# Bristol and Surroundings: $\hat{\delta}_1$

2018



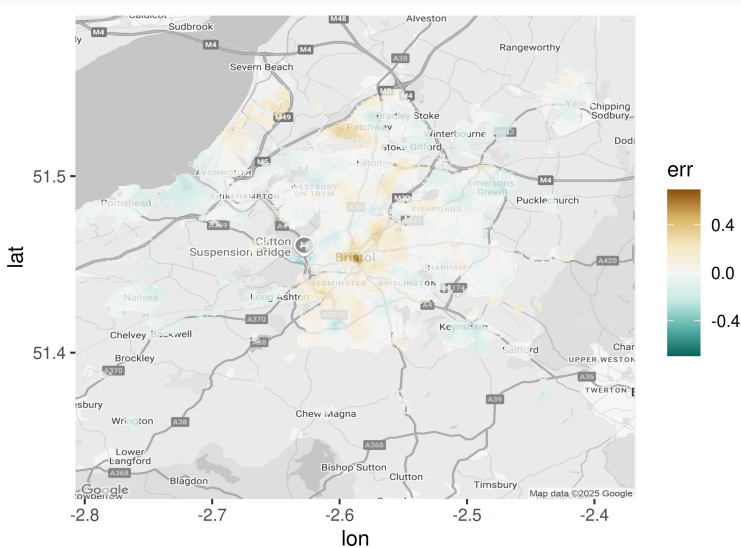
# Bristol and Surroundings: $\hat{\delta}_1$

2019



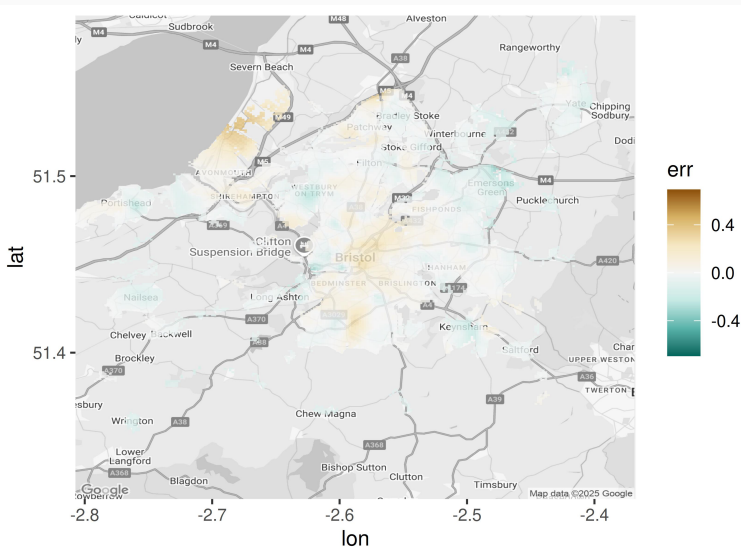
# Bristol and Surroundings: $\hat{\delta}_1$

2020



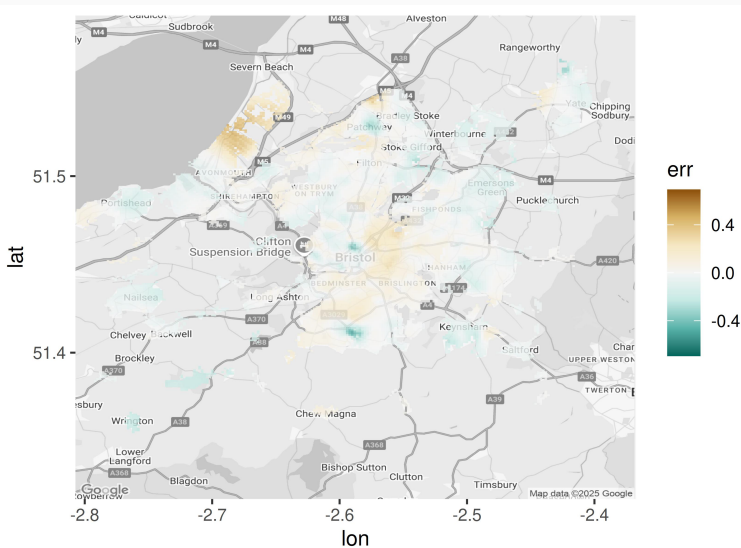
# Bristol and Surroundings: $\hat{\delta}_1$

2021



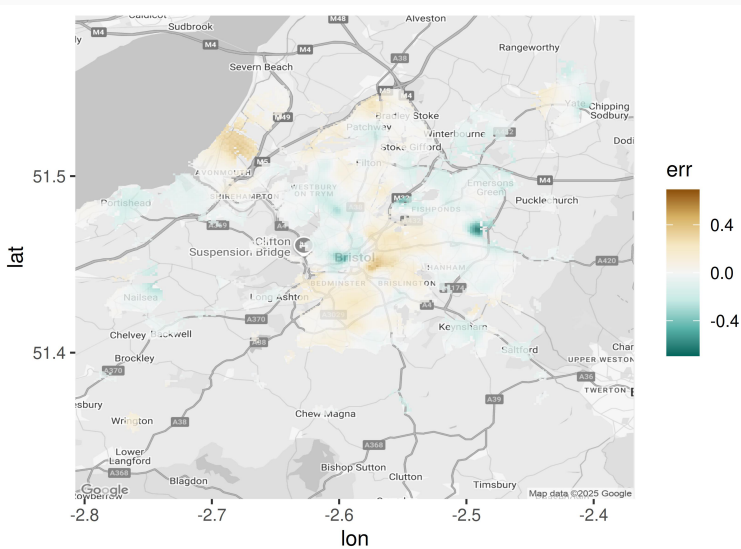
# Bristol and Surroundings: $\hat{\delta}_1$

2022



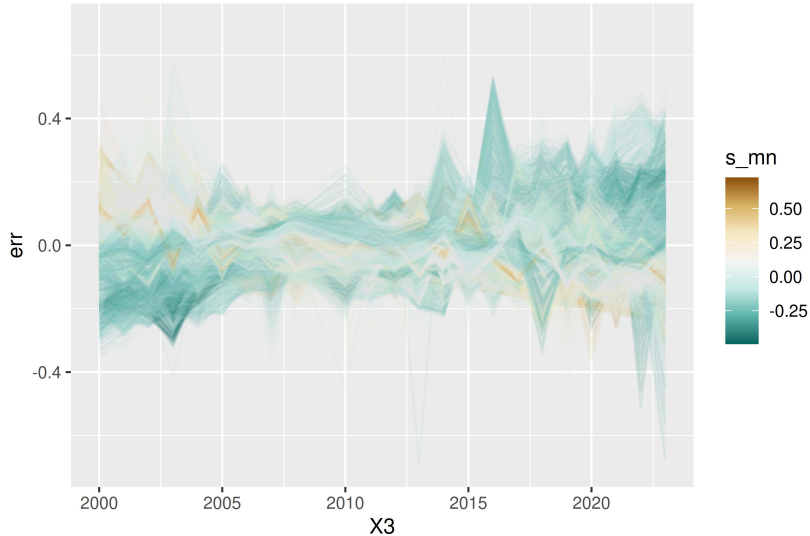
# Bristol and Surroundings: $\hat{\delta}_1$

2023



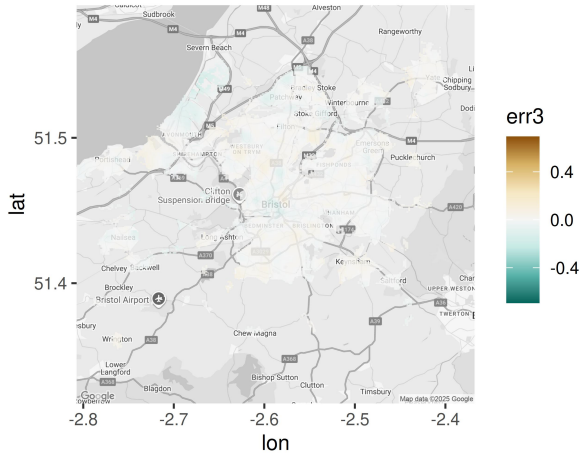
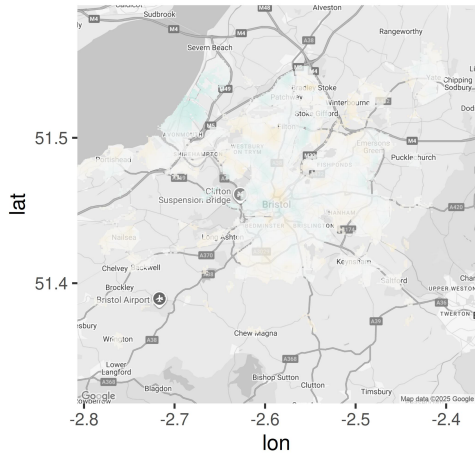


# Bristol and Surroundings: Gentrification



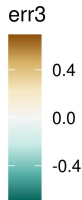
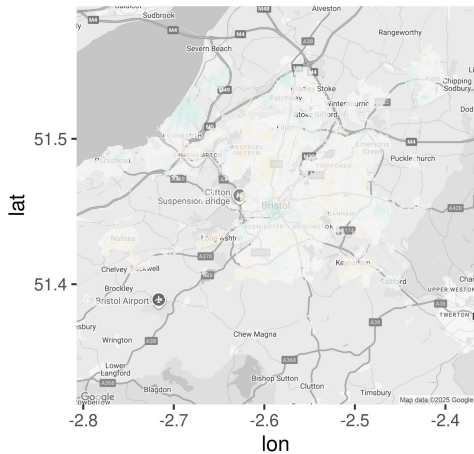
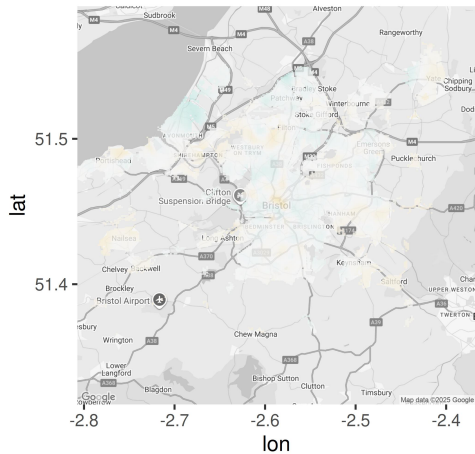
# Further Components

2000



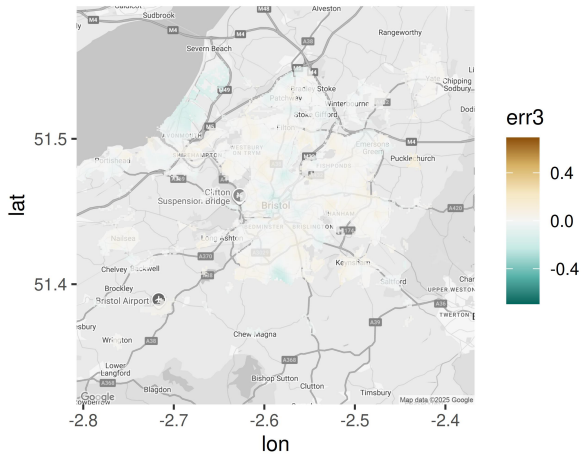
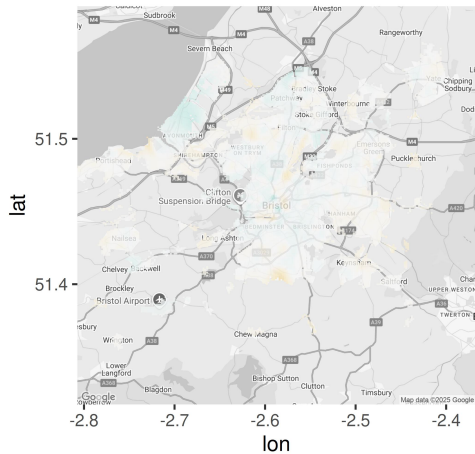
# Further Components

2001



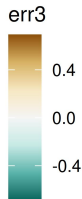
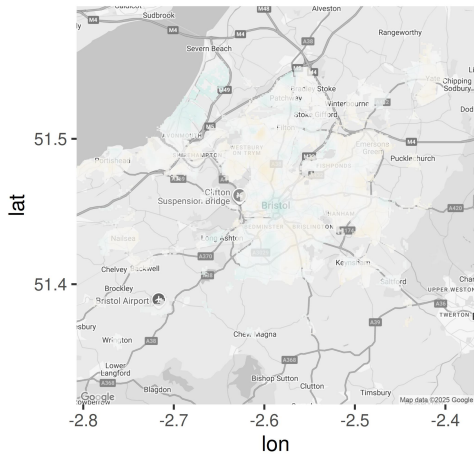
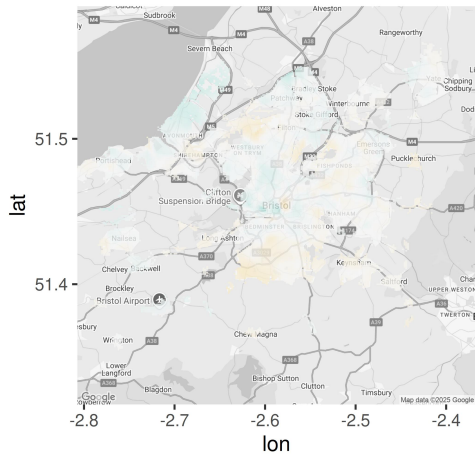
# Further Components

2002



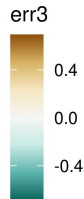
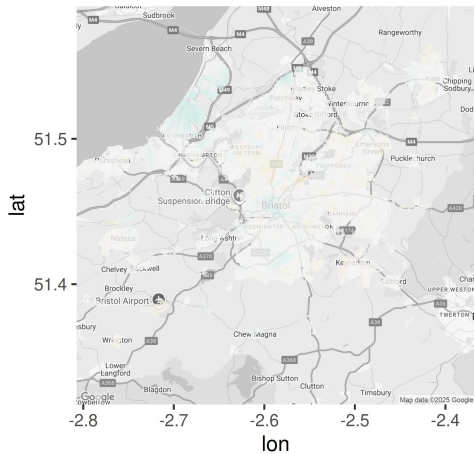
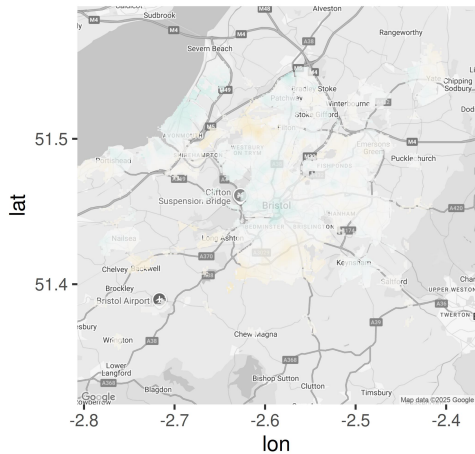
# Further Components

2003



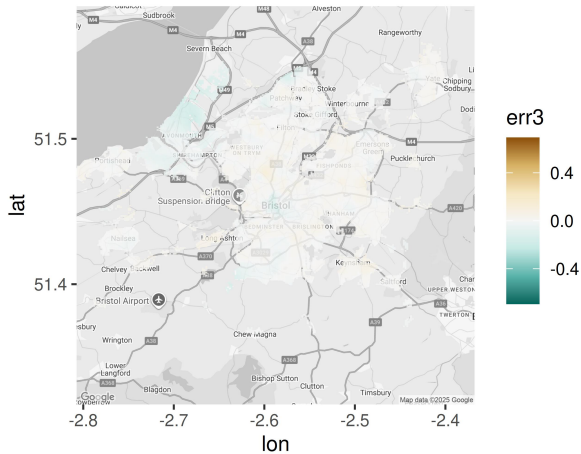
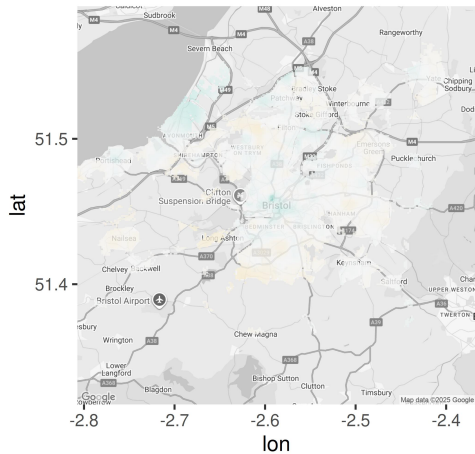
# Further Components

2004



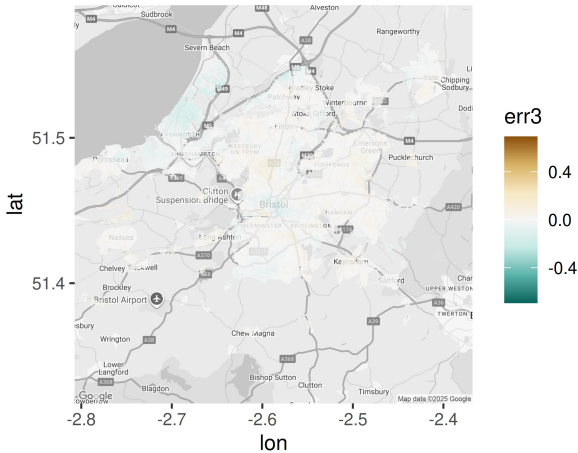
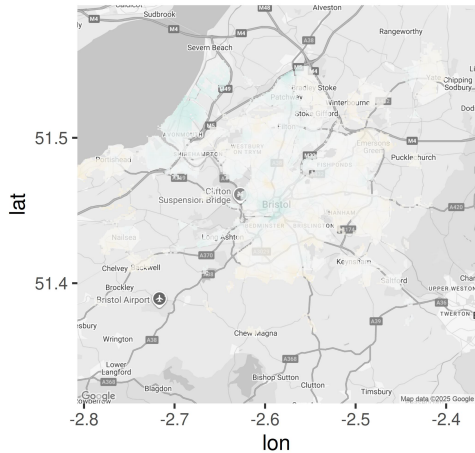
# Further Components

2005



# Further Components

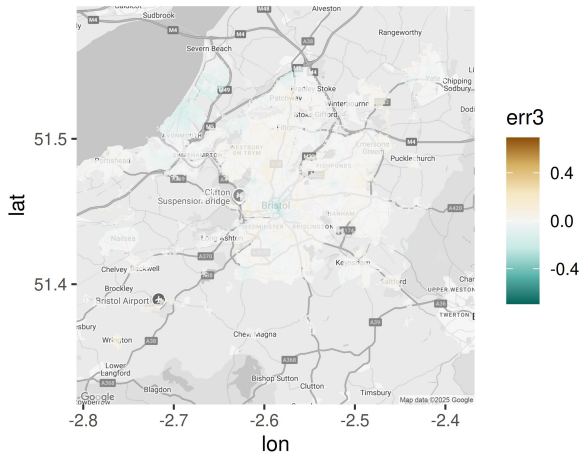
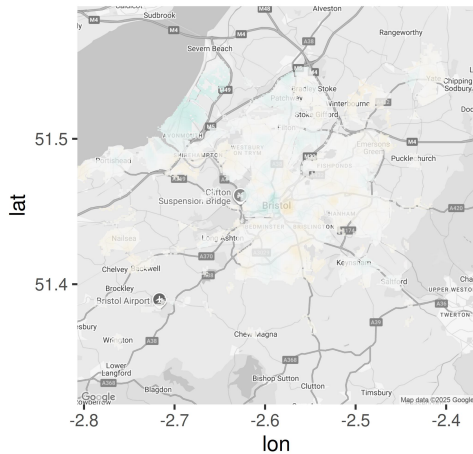
2006





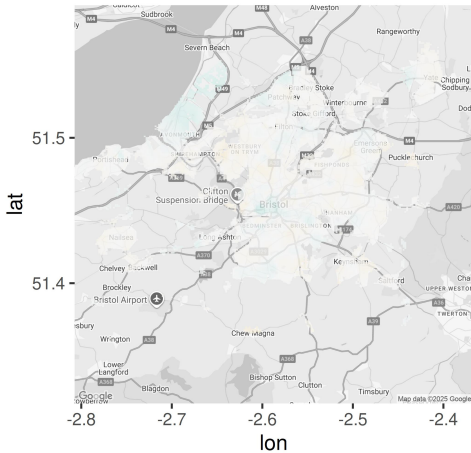
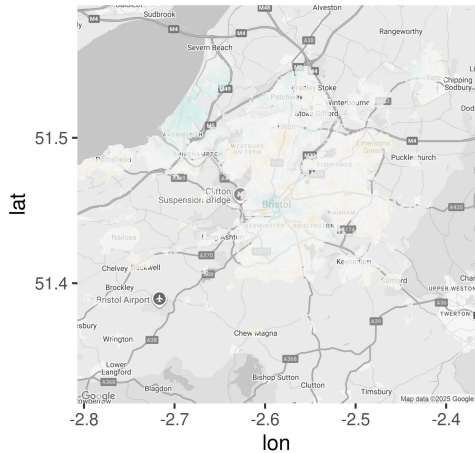
# Further Components

2007



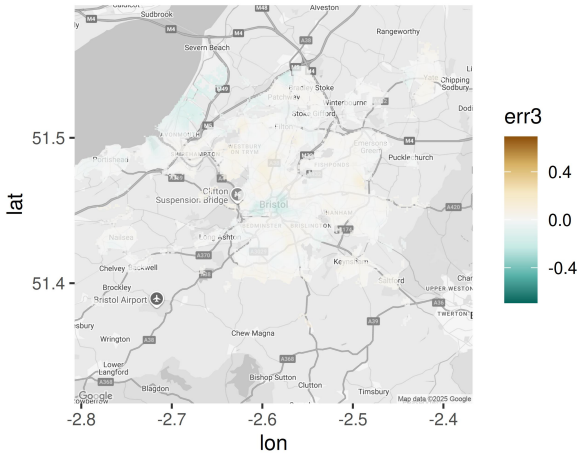
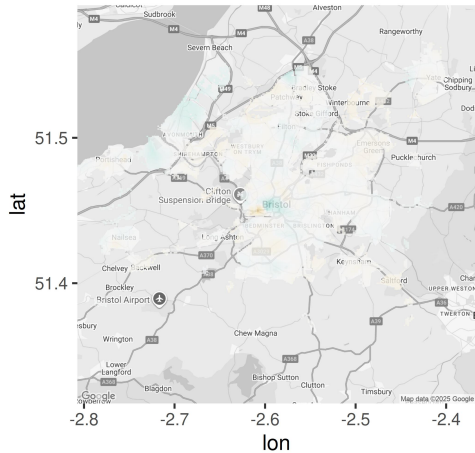
# Further Components

2008



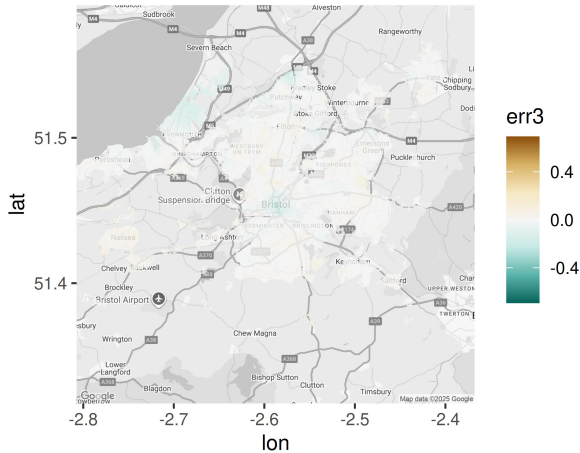
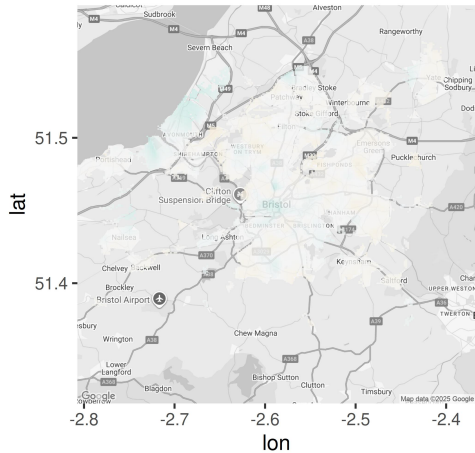
# Further Components

2010



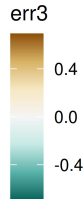
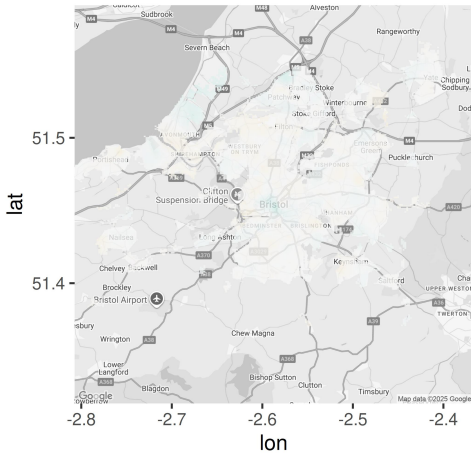
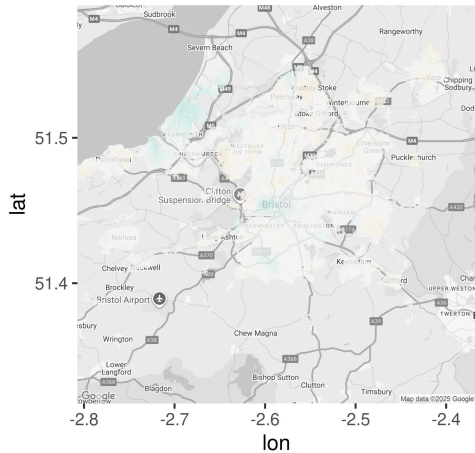
# Further Components

2011



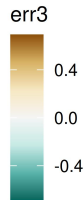
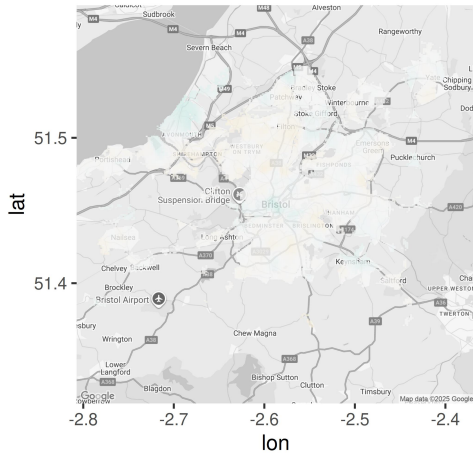
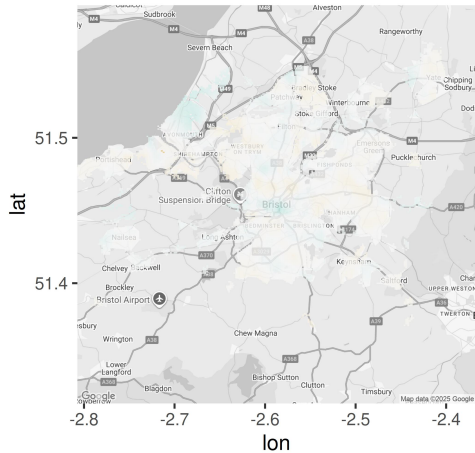
# Further Components

2012



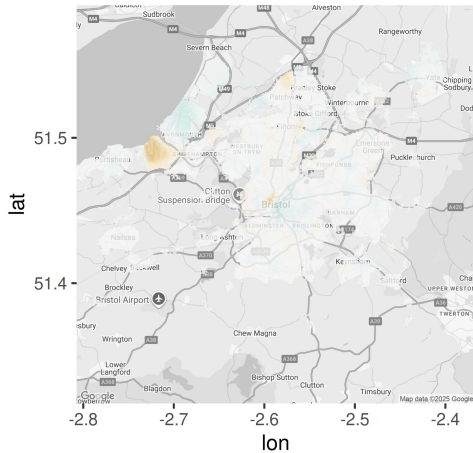
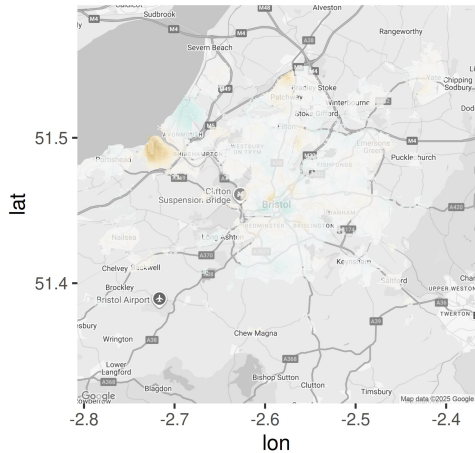
# Further Components

2013



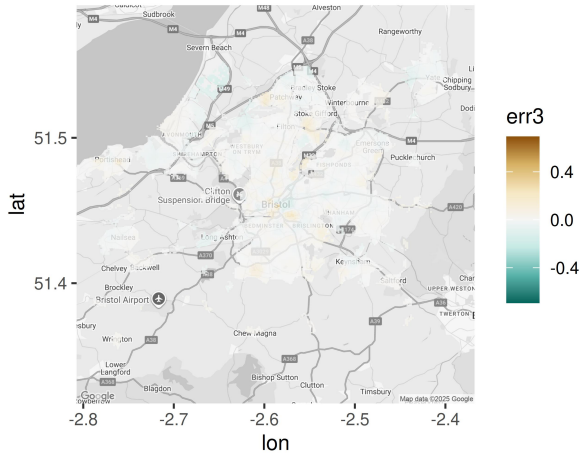
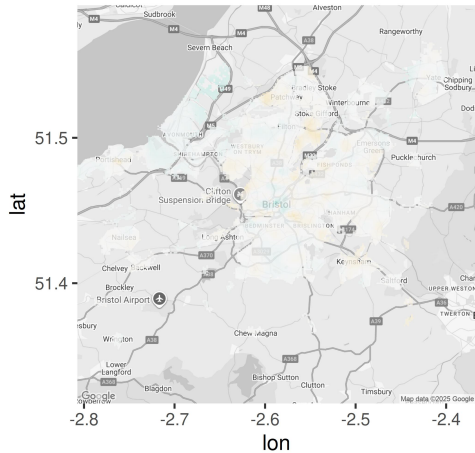
# Further Components

2014



# Further Components

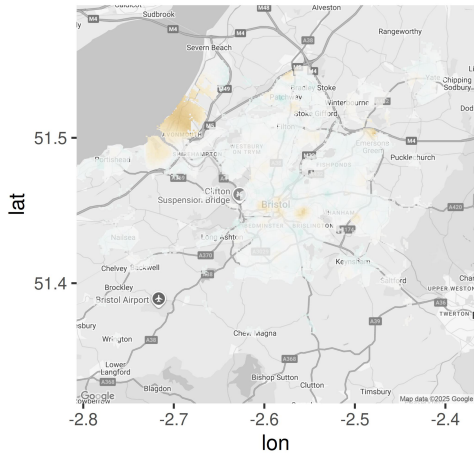
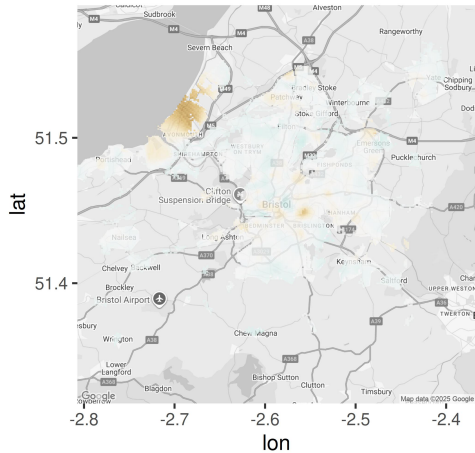
2015





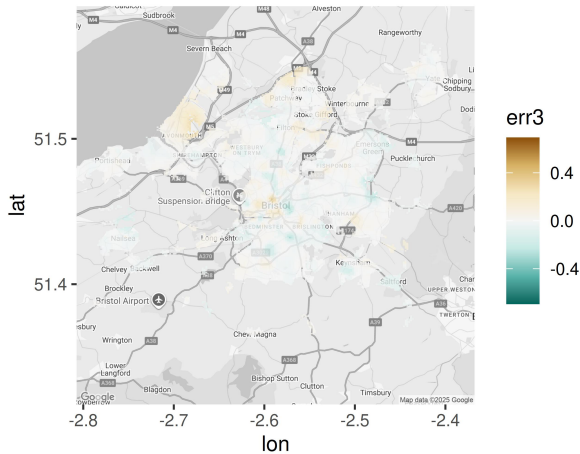
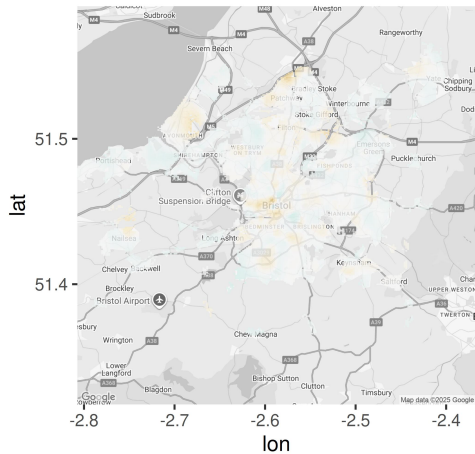
# Further Components

2016



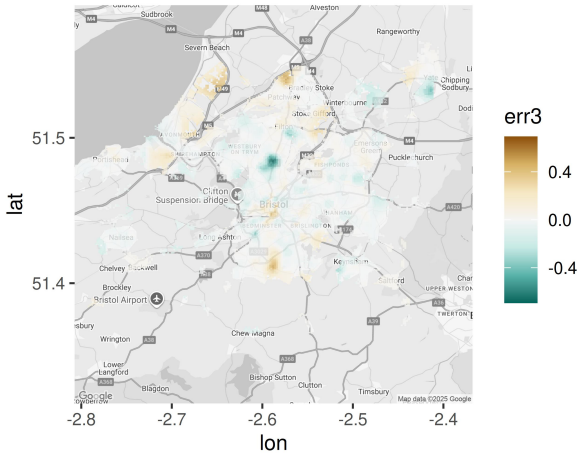
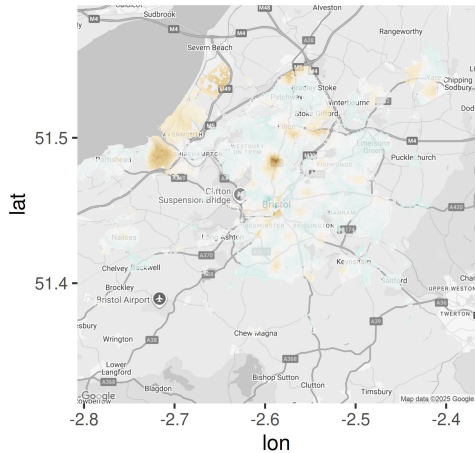
# Further Components

2017



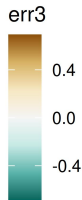
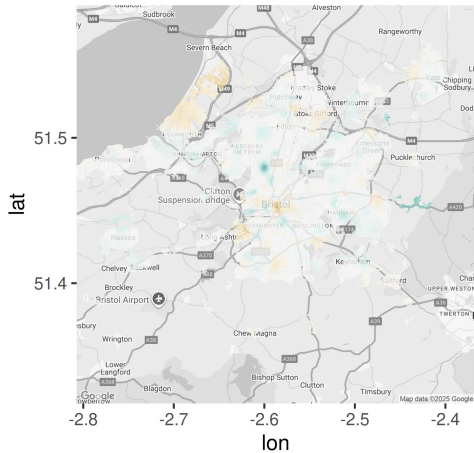
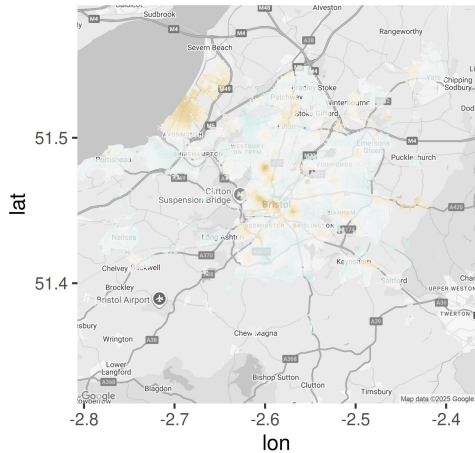
# Further Components

2018



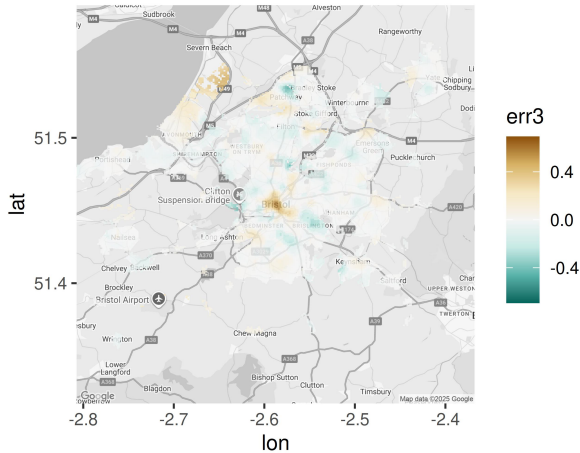
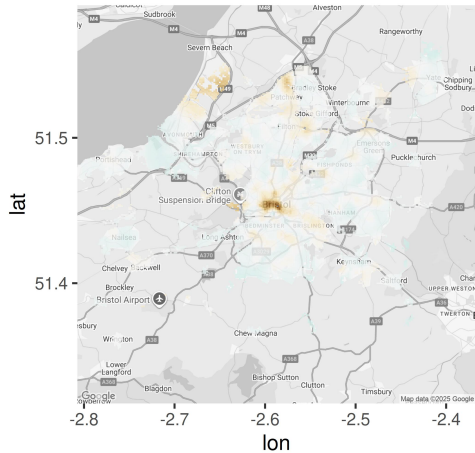
# Further Components

2019



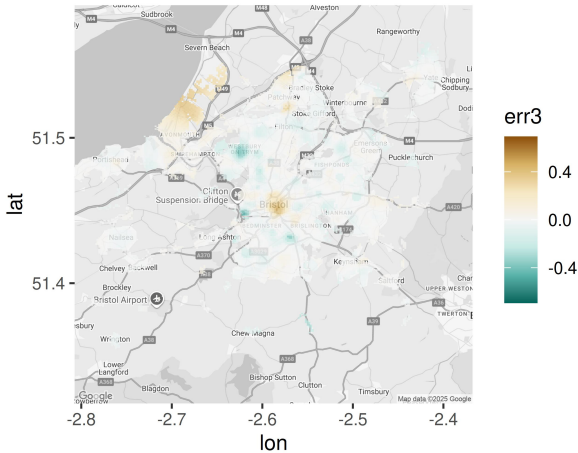
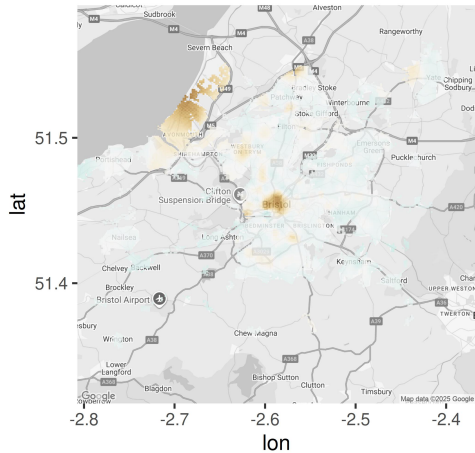
# Further Components

2020



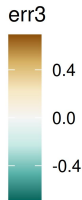
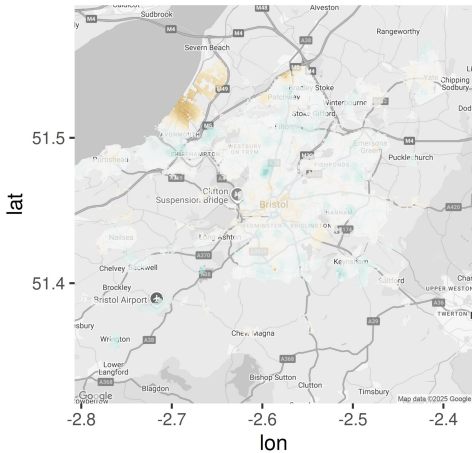
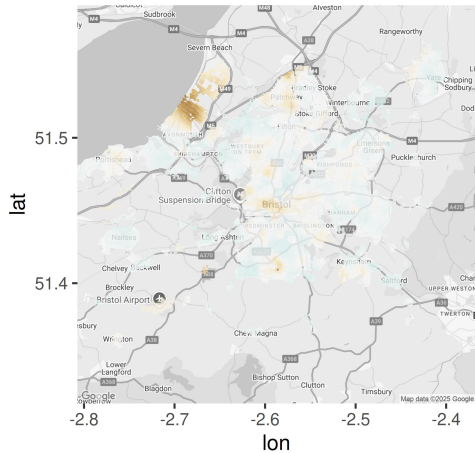
# Further Components

2021



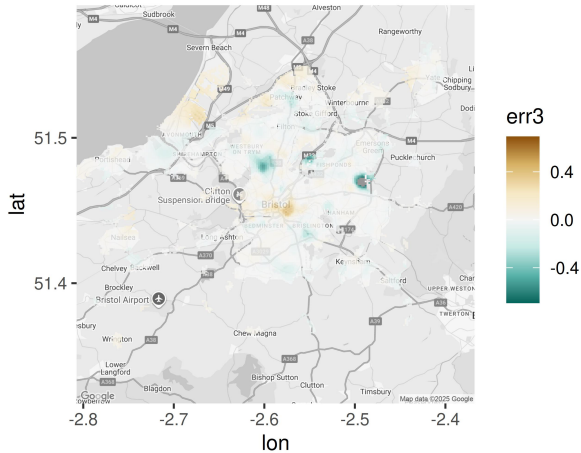
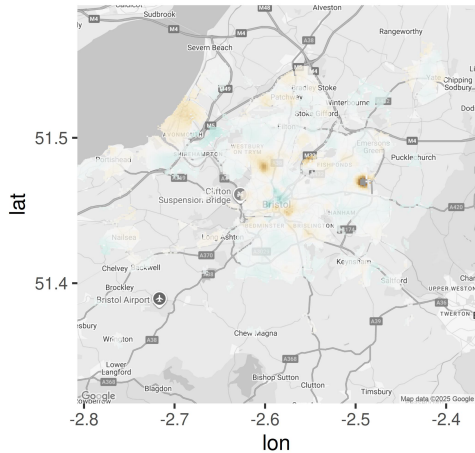
# Further Components

2022



# Further Components

2023





# Discussion

---

# Open Questions

- Has the smoothing been chosen/implemented appropriately?
- What's the best way to incorporate uncertainty quantification into the methodology?
- How can we implement an inferential framework to identify regions of interesting change over time?
- How can you factor in rental prices?
- Can this output be used to measure impact of public policy?

## Other Work

- ‘Exploring Covid-19 Spatiotemporal Dynamics: Non-Euclidean Spatially Aware Functional Registration.’ Luke A. Barratt & John A. D. Aston. July 2025. *Annals of Applied Statistics*.
- ‘A Nonparametric and Functional Wombling Methodology.’ Luke A. Barratt & John A. D. Aston. February 2025. [arXiv:2502.07740](https://arxiv.org/abs/2502.07740).
- ‘Spatially Indexed Distribution Analysis as Applied to House Prices in the UK.’ Luke A. Barratt & John A. D. Aston. August 2025. [arXiv: coming soon](https://arxiv.org/abs/2508.12345).
- ‘School Inspections and House Prices: Distribution-on-Function Regression for Spatial Data.’ Luke A. Barratt & John A. D. Aston. Working paper—email [lab85@cam.ac.uk](mailto:lab85@cam.ac.uk) for details.