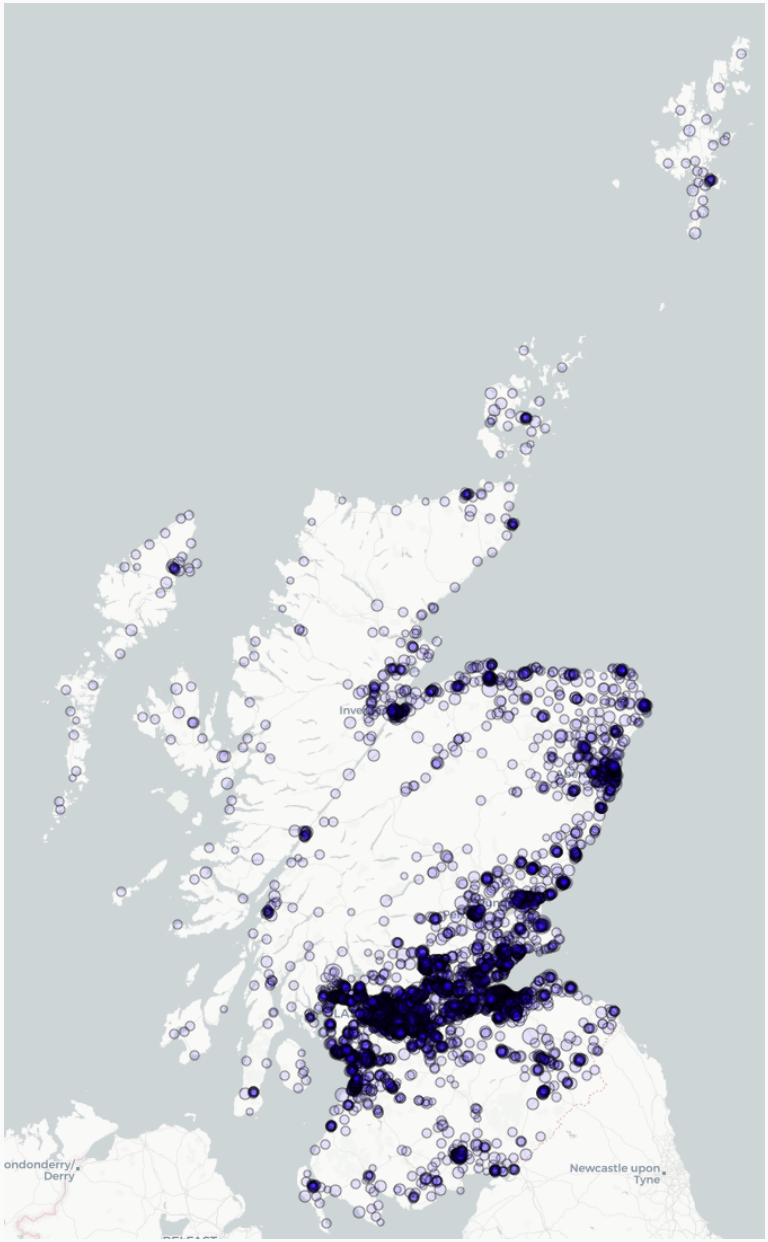
Explaining the spread of Omicron in Scotland through Deprivation, Geography and Demographics

Anthony Wood
Roslin Institute, University of Edinburgh

April 5 2022

anthony.wood@ed.ac.uk

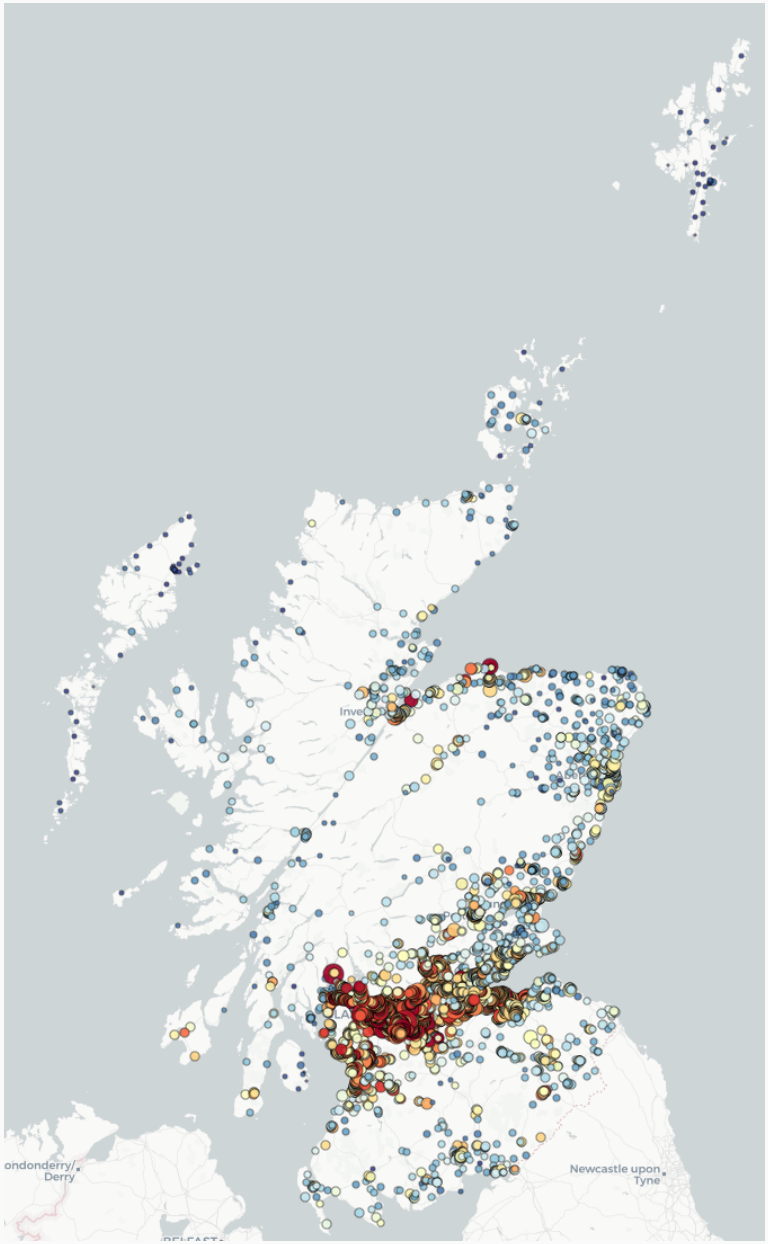
Scotland: demographics, geography and COVID



- Population: 5.5m, *central belt* contains ∼3.5m, with a few other towns/cities outside. Otherwise very sparsely populated
- Deprivation: Sharp differences, often over short distances
- COVID-19 data: from Public Health Scotland, at **individual level**. Residence given to datazone (DZ) level – populations of 500–1,000 individuals.

Why were early Omicron cases where they were?

Scotland: demographics, geography and COVID



- Population: 5.5m, *central belt* contains ∼3.5m, with a few other towns/cities outside. Otherwise very sparsely populated
- Deprivation: Sharp differences, often over short distances
- COVID-19 data: from Public Health Scotland, at **individual level**. Residence given to datazone (DZ) level – populations of 500–1,000 individuals.

Why were early Omicron cases where they were?

Understanding why cases were where they were

What affects whether you test positive

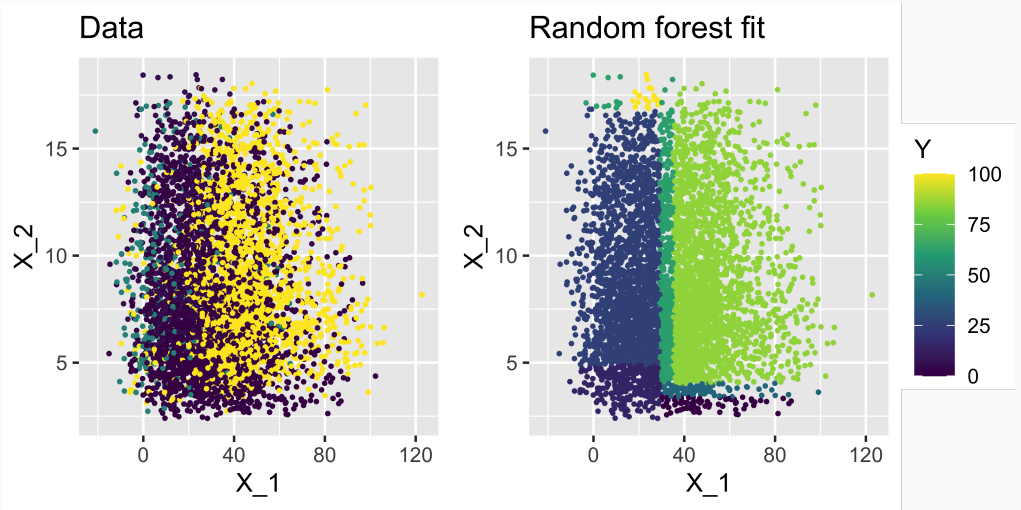
- Your interactions
- Your immunity
- Your propensity to test
  - If symptoms develop
  - If you can be bothered
  - Consequences if you need to isolate

Data available to us

- COVID data
  - Vaccination uptake
  - People testing negative
  - People with prior cases
- Census data
  - Population pyramids
  - Households
  - Deprivation

Can we use the data available to us to explain variation in the distribution of Omicron cases in the initial outbreak (15 Nov – 6 Jan)?

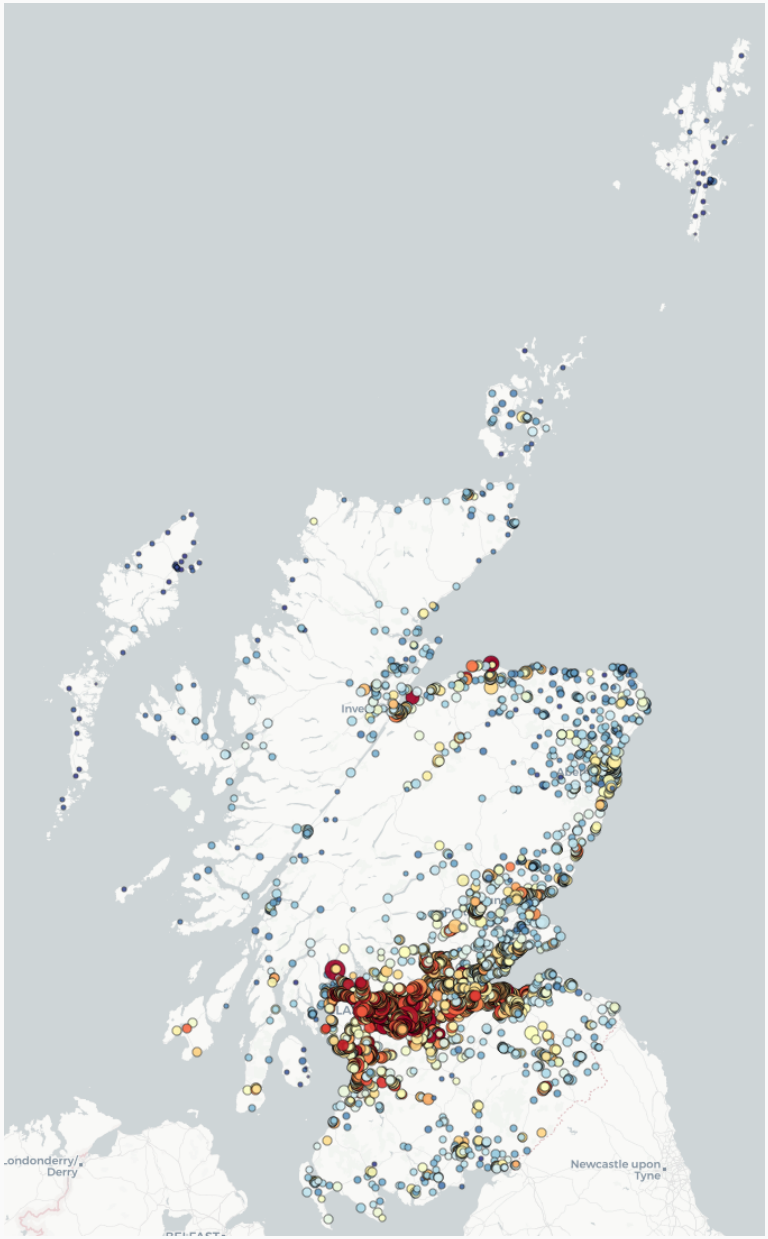
Random forest methods



- Predict an outcome Y, given data on predictors $\vec{X}$, and other outcomes.
- Unstructured fit, useful for when dealing with large numbers of predictors, when underlying pattern/interactions not obvious.

Predictors X_1 = population, X_2 = age, X_3 = sex, X_4 = vaccine uptake, X_5 = testing, X_6 = rurality...
Outcome Y = Number of Omicron cases Nov 15 – Jan 6, for some age/sex/DZ slice.

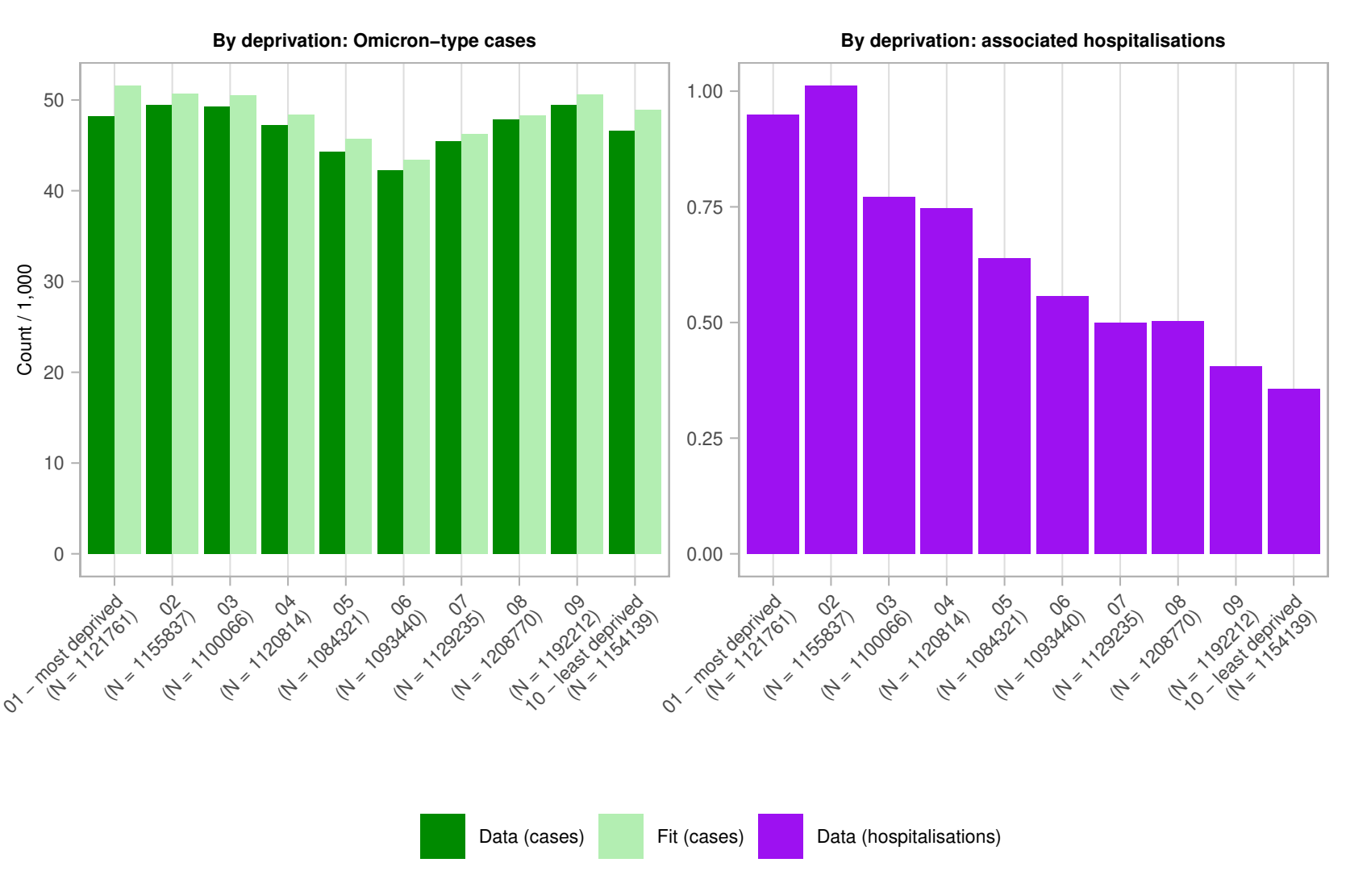
Understanding why cases were where they were



- Single-predictor models **fail** ($R^2 \sim 0 - 0.4$)
- A model combining several predictors performs much better ($R^2 \sim 0.75$), and reproduces finer-scale spatial variation
- Age dominates at individual level, but does not explain anything spatially.

- Spatial variation in hospitalisations?

Deprivation anomaly



Testing data suggest this is not entirely an inherent **health** discrepancy.

Conclusions

- A multitude of predictors are traditionally linked with high cases (age, deprivation, uptake, rurality, students, density, testing propensity). We show **no single predictor can explain spaitial patterns**. To do so adequately needs a combination of several predictors, with the resultant trends complex.

- High variation in LFD testing suggests **testing propensity may be deprivation-dependent**, and exacerbating differences in case-hospitalisation rates.

- Plenty of scope to go further – counterfactuals, understanding variation in vaccine uptake, testing, severe COVID-19 outcomes.