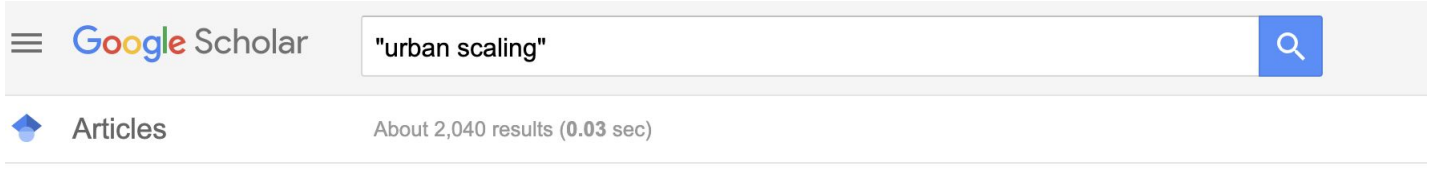
Pubs, Schools & Churches: Scaling Laws in Rural England

Matthew Sharpe

Oxford University, Spotify

Why rural?

- “For most of human history, populations lived in very low-density rural settings”[1]
- Approximately 45% of the world’s population lives rurally.[1]
- In England and Wales, around 80% of the land is classed as rural.



1 Our world in data

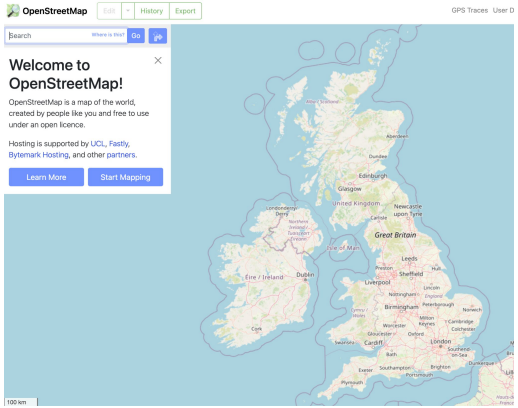
Building a dataset

Need to know

- Urban/Rural classification
  - UK Gov (2011)
- Geographic boundaries
  - UK Gov (2021)
- Population
  - UK Census (2021)
- Facilities
  - OpenStreetMap (2023)

Want to know

- Indices of deprivation
  - UK Gov (20??)
- Area
  - UK Gov (2021)
- House sales & prices
  - UK Gov (1995 - 2023)

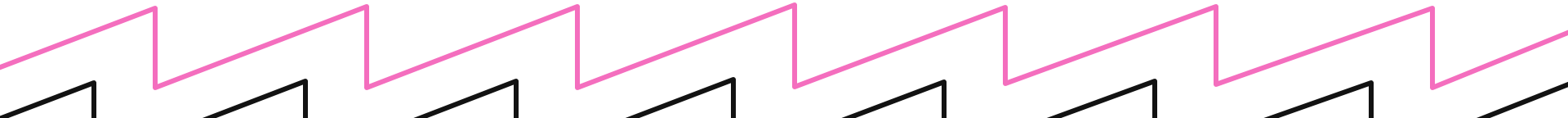


Office for National Statistics

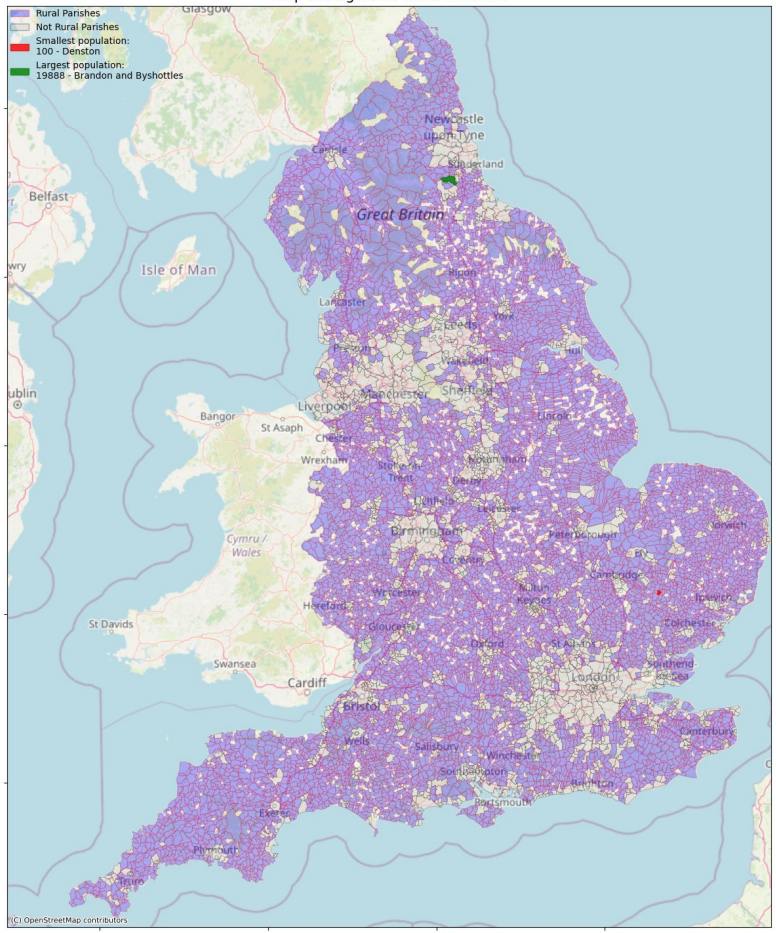
Open Geography Portal Boundaries

First, some housekeeping

- We're only going to look at England
  - Different datasets for Scotland/Northern Ireland, and indices of multiple deprivation are different for England and Wales.
- We're going to look at parish level
  - Parish most closely corresponds to the idea of a "village"
- We're going to consider a parish as being rural if all output areas contained in it are rural
  - Different levels of geographies are being worked with.
- We're going to be a little loose with our considerations of time.
  - All of our datasets are taken at slightly (or not so slightly) different points in time. Pubs open and close. Populations change. Deprivation may wax and wane.



Map of England's Rural Parishes



Stat	Rural	Urban
Number	7,781	918
Median population	527	7,581
Median size (square km)	10.1	6.6s

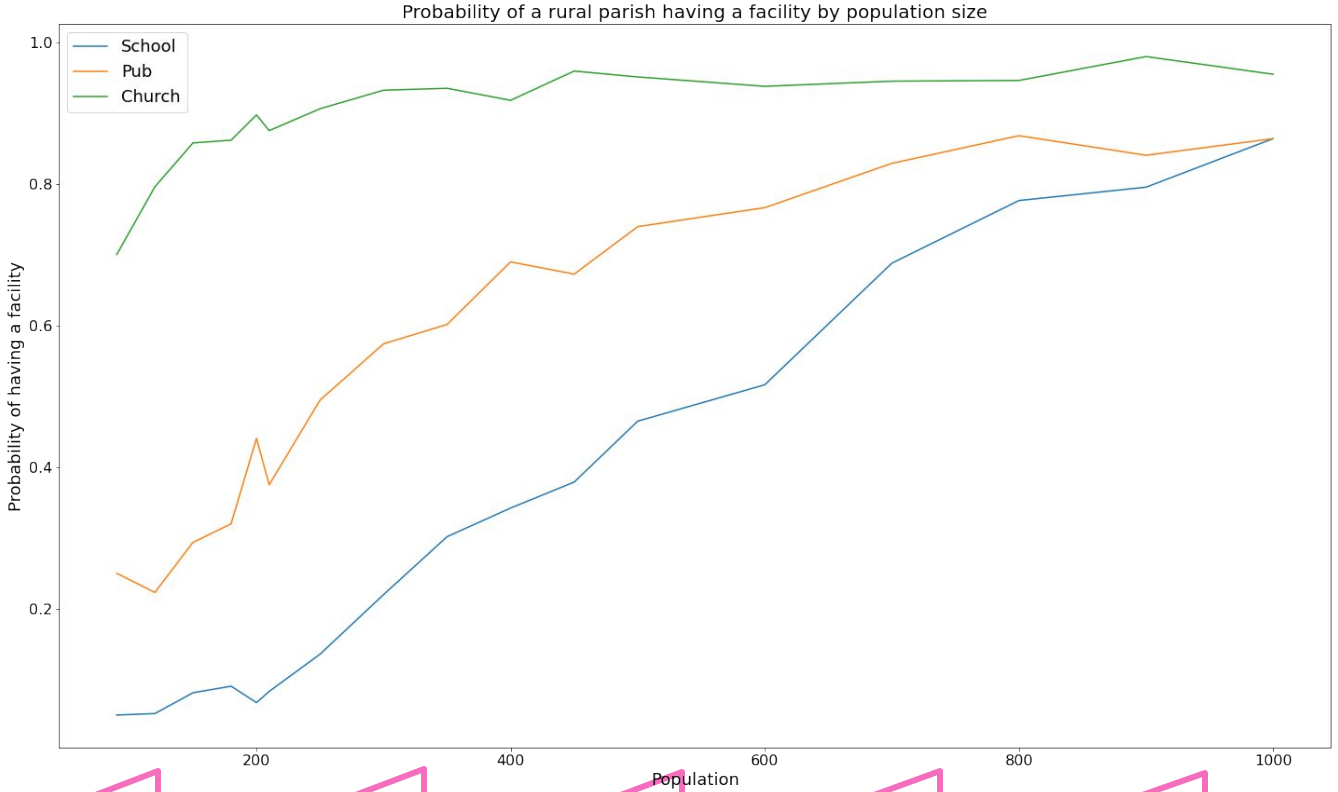
Starting simple

Rural Facilities

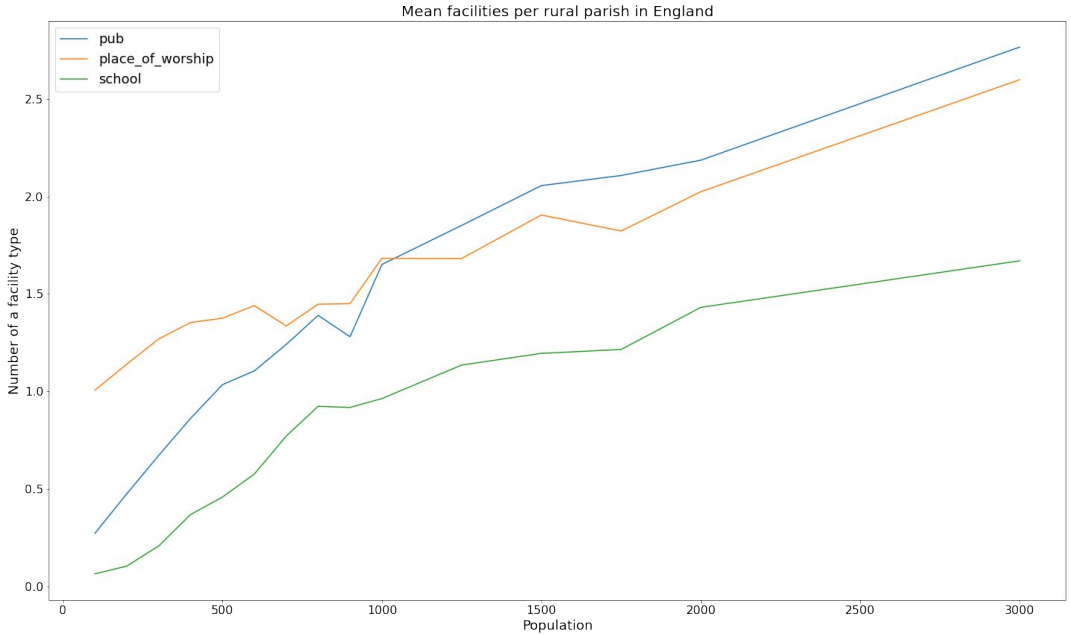
1. Post box
2. Graveyard
3. Bench
4. Place of worship
5. Hunting stand

Urban Facilities

1. Bike parking
2. Parking space
3. Fast food
4. Restaurant
5. Waste basket



Scaling laws



What's a scaling law?

Simply put, that there's a relationship between the population of a village and the number of facilities that follows the following pattern:

$$Y = Y_0 N^{\alpha}$$

Where Y is the number of facilities, Y_0 is a constant, N is the population and alpha is the scaling constant

(Really) Small scale scaling laws

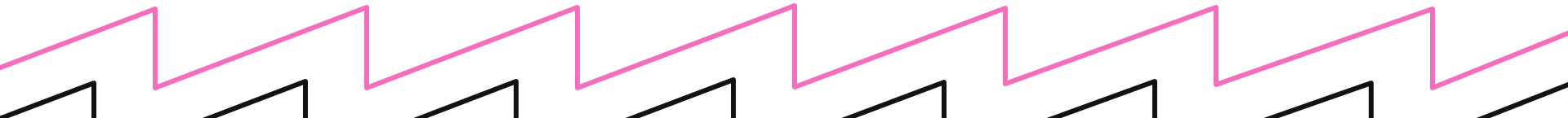
Gomez-Lievano et. al[2] proposed a method of calculating scaling laws using Bayes' theorem which may prove useful in smaller data settings.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Unfortunately, it doesn't really work for our case.

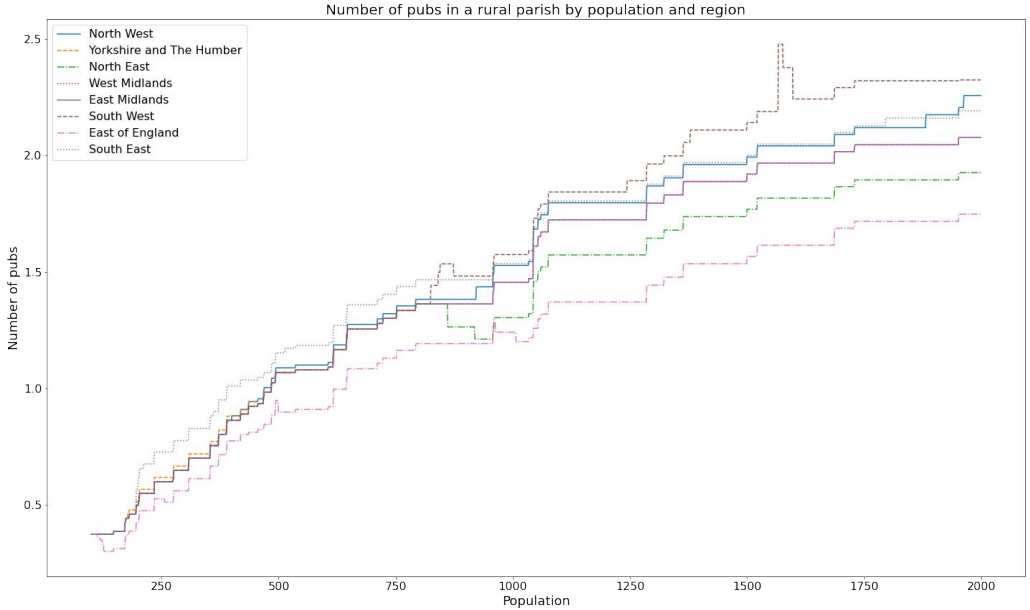
This is primarily due to a lack of independence between facilities. A village with 300 people and one school is **very** unlikely to get another one. The probability of a village getting a facility is heavily dependent on it already having one.

2 The Statistics of Urban Scaling and Their Connection to Zipf's Law



Throwing model(s) at it

Let's just predict the number of pubs/schools/churches and then examine what our model is telling us about how it's making those predictions...



OLS Regression Results

Dep. Variable:	school	R-squared:	0.228
Model:	OLS	Adj. R-squared:	0.227
Method:	Least Squares	F-statistic:	228.9
Date:	Fri, 14 Jul 2023	Prob (F-statistic):	0.00
Time:	21:51:23	Log-Likelihood:	-4636.3
No. Observations:	7781	AIC:	9295.
Df Residuals:	7770	BIC:	9371.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0866	0.018	4.862	0.000	0.052	0.121
population	0.0001	3.13e-06	40.258	0.000	0.000	0.000
area	4.902e-09	4.42e-10	11.089	0.000	4.04e-09	5.77e-09
population_weighted_imd	1.065e-05	9.15e-07	11.639	0.000	8.85e-06	1.24e-05
region_East Midlands	0.0167	0.013	1.293	0.196	-0.009	0.042
region_East of England	0.0115	0.011	1.006	0.315	-0.011	0.034
region_North East	-0.0536	0.025	-2.137	0.033	-0.103	-0.004
region_North West	0.0419	0.017	2.441	0.015	0.008	0.076
region_South East	0.0477	0.013	3.620	0.000	0.022	0.074
region_South West	0.0566	0.011	5.066	0.000	0.035	0.079
region_West Midlands	-0.0067	0.015	-0.456	0.649	-0.035	0.022
region_Yorkshire and The Humber	-0.0276	0.015	-1.888	0.059	-0.056	0.001

Omnibus:	41308.880	Durbin-Watson:	1.933
Prob(Omnibus):	0.000	Jarque-Bera (JB):	821.091
Skew:	0.147	Prob(JB):	5.04e-179
Kurtosis:	1.436	Cond. No.	1.78e+22

TL;DR?

- There are clear (non-linear, different, but positive) relationships between population of a parish and the number of pubs, schools and churches.

- There are small (but statistically significant) regional differences in the probability of a parish having a pub, school and church.

- Similarly, the indices of deprivation suggest that affluence brings with it (or perhaps measures?) an increased likelihood of pubs and schools.

Questions

p.s. while we wait on this slide, just to let you know that all the data and analysis is available here:

https://github.com/Kali89/rural_geographic_analysis

That includes, for every output area and parish in England the:

- Population
- Number of house sales (and value of them) in the last 1, 3 and 5 years.
- Every amenity listed on OpenStreetMap
- The indices of multiple deprivation
- The area
- A label saying whether it's rural, urban or mixed

JUST THINK of all the cool things you could do with that!