Assessing regional disparities in the relationship between population size and facilities present in English villages

Matthew Sharpe
Oxford University
Word count: 4231
matthew.sharpe@oii.ox.ac.uk

Abstract

In this paper we use OpenStreetMap data to explore the relationship between the population size of rural communities in England and the facilities present. Specifically, we derive a relationship between the population and the number of pubs, schools and places of worship in a given rural parish. Additionally, we collect the area of a parish and its average (and population weighted) index of multiple deprivation. We compare these across the regions of England to build an understanding of how rural communities vary by region. Finally we build a logistic model to understand whether there are differences between the probability of a parish having certain facilities by region of England.

1 Introduction

Analysis of the Domesday book suggests that more than 90% of the population of (much of) England lived in the countryside(Robinson, 2021) whereas recent estimates state that in 2019, 82.9% of England's population lived in urban areas(for Science, 2021). The impact of this significant movement has been studied, with special attention being given to the impact of (and challenges faced by) newcomers to an existing village community(Neal and Walters, 2008)(Miller, 2016)(Burchardt, 2011). Previous studies of rural life have typically taken the form of ethnographies, with researchers spending prolonged periods of time in a rural community in order to learn about sociological phenomena in village life(Miller, 2016)(Beaumont and Brown, 2015)(Bagley and Hillyard, 2015)(Jenkins, 1999). Arguments have been made detailing challenges with this approach - specifically around the reliability of the disclosures to non-local researchers(Neal and Walters, 2006); the limitations this could cause researchers (along with opportunities to research close to home!) are also discussed in academic literature(Heley, 2011).

There have been a number of advances in the study of towns and cities. Mobile phone location data has been used to understand how people travel and interact(Ratti et al., 2006), social network data has been used to understand which areas of a city enable people from disparate social groups to meet(Hristova et al., 2016), and to assess community flow and its impact on community well-being(Lathia et al., 2012). All of this work could be termed "Urban Social Data Science" but that likely begs the question "could we better study rural communities using similar techniques?" Indeed, given the number of rural communities in England and the geographic distance between them, it's likely that data science techniques are able to offer more benefits for the analysis of rural communities than it does for urban communities.

Travel writer John Hillaby once said of England that

> "Few things are more pleasant than a village graced with a good church, a good priest and a good pub."

The village pub has historically been at the heart of village life and has warranted much academic study(Markham and Bosworth, 2016)(Cabras, 2011) (interestingly, the village church and school are much less studied(Hillyard, 2020) - later in this paper I will show why that should likely not be the case!). The decline of the pub industry(Keenan, 2020) and of church attendance(Bruce and Glendinning, 2010) is well-documented, and the death of what are often regarded as the cornerstones of village life continues to generate concern. In the aftermath of the COVID-19 pandemic and subsequent lock downs, a number of employers transitioned to allow its employees to work from home. This in turn led to an 126% increase in the number of housing searches for village homes(Rightmove, 2020).

In this paper we aim to compare the facilities present in rural communities in England and see how the number of facilities varies based on the population size of a community. This will allow us to identify villages with an unusually high (or low) number of certain types of facilities (perhaps indicating good areas to move to, or good areas for developers to build in). It will also allow us to understand whether there are clear relationships between the size of a community and the facilities it has. Given an expected lag between population numbers and facilities, it could be that discrepancies indicate either the growth or decline of a rural community. Finally, regional disparities may be found that could be interesting from an historic or economic perspective.

2 Data Collection

In order to perform our analysis we need accurate information on:

1. The population of parishes in England
2. A classification as to whether a parish is rural or urban
3. The facilities present in parishes in England
4. The geographic size of a parish

5. The indices of multiple deprivation (IMD) for a parish

The Office for National Statistics publishes a guide to the different geographies that are used in creating and compiling the statistics that they publish(for National Statistics, 2022b). The key geographies that we traverse in this analysis are (for National Statistics, 2022a):

- Output Areas (OAs) - the lowest level of geography, representing around 300 people.
- Lower layer Super Output Areas (LSOAs) - groups of four to five OAs.
- Parishes - the smallest type of administrative area in England. A very old form of spatial unit with limited current function - often limited to providing facilities including community centres and playgrounds.
- County - a larger administrative, historic and cultural area with populations ranging from around 500,000 to 5,500,000.
- Region - the highest tier of sub-national division in England, established in 1994 and abolished in 2011.

Given that each of the measures we use (IMD, population size, area, urban/rural classification) are only available at some geographical resolutions and not others, we will need to be comfortable navigating between hierarchies.

In 2021, the UK Government ran a census (ONS, 2021) and made a number of results available through the NOMIS web service(ONS, 2022a). Using this service we downloaded the population of every Output Area as of the time of the 2021 census. Given that each output area has a population of roughly 300, we are unable to study areas of differing populations using only the output areas. Using the Open Geography Portal(ONS, 2022b) we are able to obtain a lookup between output areas and parishes in England, and hence group output areas into their parishes. Using the same source we can also locate parishes within their counties and regions and also obtain approximate look ups between the 2021 output areas and the 2011 output areas.

While there has been no attempt to classify the output areas defined in 2021 as being rural or urban, there was in 2011(ONS, 2022b). Using the aforementioned look up between 2011 output areas and 2021 output areas we are able to classify each 2021 output area as being urban or rural.

Indices of multiple deprivation (IMD) are an attempt by the UK Government to rank areas based on their relative deprivation across 7 different measures (Payne and Abel, 2012). It was most recently carried out in 2019 and data was published using the 2011 Lower Super output areas (LSOA). Using our previous look ups, we can determine which LSOA the 2021 output areas belongs to and hence find its most recent deprivation index.

OpenStreetMap offers a free API that allows a user to retrieve all the amenities within a given area. As such, it has been used extensively in academic research. Studies have sought to research alcohol(Bright et al., 2018), understand socio-economic indicators(Feldmeyer et al., 2020) and even to predict crime(Bappee et al., 2018) using this data. A number of studies have sought to assess the accuracy of the data(Kounadi, 2009) (Sehra et al., 2013) (Mooney et al., 2017) and most recently, Grinberger et. al concluded that "OSM stores significant amounts of geospatial data with relatively high data quality.". Again, using the Open Geography Portal we are able to download a list of the geographic boundaries of each output area. Using these, we are able to iterate over the list of output areas in England and determine the number of different facilities present in each output area. A procedure for doing so (and all the code used for analysis) is available here.

Ultimately, these steps give us a data set made up of an output area (with population circa. 300), a size, the type and number of facilities (ranging from postboxes all the way up to airports) and their location, the index of multiple deprivation, and the ability to move between geographical hierarchies. This data set then allows us to answer a range of questions, ranging from the micro (which is the biggest parish without a post box?) through to the macro (which region has the highest density of schools per person?).

3 Methods

3.1 Data Validation

Given the number of sources from which we drew and joined data, we wanted to ensure that our data was representative and that we didn't lose any important data that could introduce bias. As an example, failure to properly consider non-ASCII characters could lead to us improperly handling some Welsh place names (such as Ynys Môn) and could lead to their accidental exclusion from data sets. Additionally, we are using output areas and populations from the 2021 Census, 2022 parish boundaries and facilities available, and 2011 urban rural classifications and indices of multiple deprivation. These discrepancies of date are unfortunate and force us to use different geographical hierarchies and occasional approximate look ups.

Another potential cause of data pollution comes from interaction with the OpenStreetMap API. In order to not overload publicly available APIs we have to ensure that we leave sufficient time between queries. However, there are 188,881 output areas in England for which we'd like to obtain data. Given a five second gap between queries, we needed over 10 straight days of querying in order to obtain a data set comprising the entirety of England. The technical complications associated with running a data extraction process over such a long time period offer ample opportunity for data corruption.

In order to assure ourselves that our analysis isn't skewed by possibly systematic causes of failure (for instance, if our queries are more likely to fail for output areas that have many amenities, then any attempt to impute values or exclude the output areas in question will bias the results significantly), we require that a region will only be included in this analysis if every one of its output areas is present in our amenities data file.

Given that parishes can (and do) consist of multiple output areas and each output area is classified as belonging to one of 10 different urban or rural types, it's wholly possible for each parish to contain different types of output area. As an example, the make-up of Brampton in Chesterfield (Derbyshire) is given

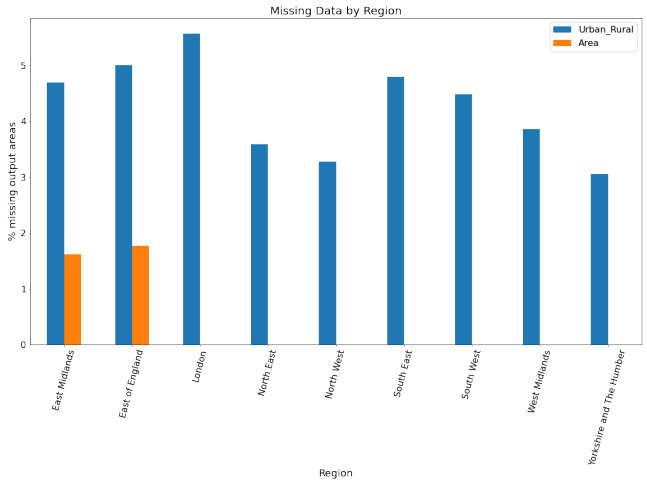

Figure 1: Missing data by region

in table 1.

Output area type	Population
Urban City and Town	253
Rural hamlets and isolated dwellings	366
Rural hamlets and isolated dwellings	318
Urban City and Town	251

Table 1: Population and classification of output areas in Brampton, Chesterfield

In this analysis, we exclude parishes that consist of mixed urban and rural output areas (though we do keep parishes with a variety of different types of rural output area) or wholly urban parishes.

3.1.1 Data Visualization

Figure 1 shows the missing data by region. Given that no region is missing more than 5% of any class of data, we feel confident progressing with the analysis. Issues with finding parishes (given output areas) are responsible for the small losses associated with the area of a parish. Issues with matching between 2021 output areas and 2011 LSOAs are responsible for the missing urban/rural (and IMD) classifications.

Figure 2 shows the percentage of parishes in each region that meet our definition of rural. Given the sharp drop-off, we choose to exclude data from London and the North East as they don't have enough rural parishes for us to draw firm conclusions.

Figure 3 shows the distribution of population size of the selected rural parishes. There are 5,066 exclusively rural parishes in England with populations ranging from 100 (Denston in Suffolk) through to 12,162 (Heathfield and Waldron in East Sussex).

In order to determine which facilities to focus on, we show the most common facilities present in rural parishes in England overall (according to the OpenStreetMap data) in table 2, alongside the most over-represented facilities in wholly rural parishes and in non-wholly rural parishes in tables 3 and 4 respectively.

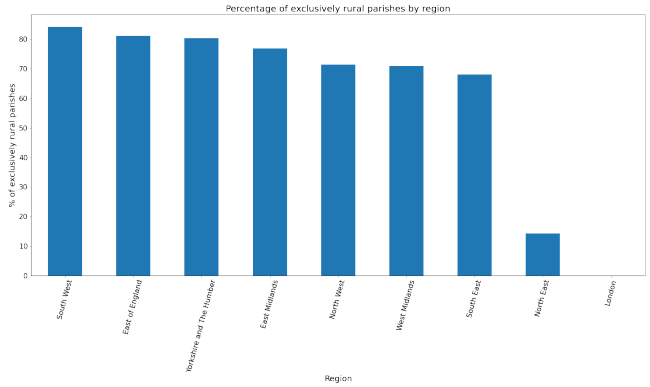

Figure 2: Percentage of English parishes that are wholly rural

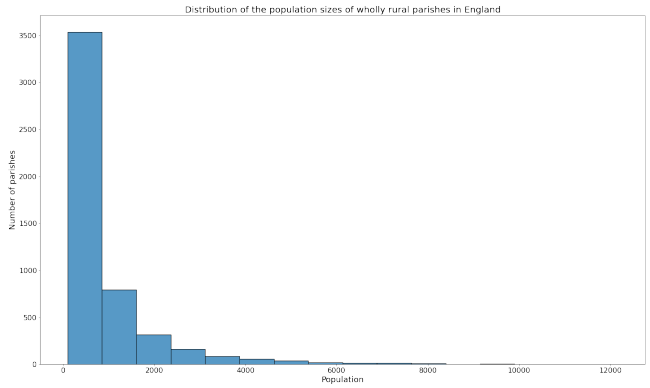

Figure 3: Population distribution of rural parishes in England

Given these lists, we choose to focus on pubs, schools and places of worship.

3.2 Relationship between population and number of facilities

Our primary question of interest is that of the relationship between the population of a parish and the probability of finding a certain type of facility in a parish. Figure 4 shows the mean number of facilities per population (after grouping population to the nearest 100 people). The relative popularity of facilities can be immediately distinguished, and a varying gradient implies that different facilities will have different dependencies on population sizes.

Given the nature of the problem, one may be tempted to model the number of facilities a parish has as a Poisson distribution. Assuming a null hypothesis (H_0) that the data is Poisson distributed, we can calculate the expected values of the Poisson distribution with the mean (μ_i) calculated as the mean of the value of the samples we have collected (for our pre-defined places of interest) and perform a Chi-squared test in order to assess the likelihood of the null hypothesis. The test statistic (χ^2) and dispersion statistic (σ_p) of the data set (for pubs, places of worship and schools) are shown in table 5. The p-values indicate that we are confident in rejecting the null hypothesis and can conclude that the number of facilities is not adequeuetly modelled through a Poisson process. Given the large degree of overdispersion (all > 10), as per (Payne et al., 2018), we determined that a negative binomial distri-

Amenity	% of overall amenities
Parking	21.3%
Bench	12.8%
Post Box	12.6%
Place of worship	6.6%
Pub	5.4%
Graveyard	4.7%
Parking space	3.4%
Waste basket	3.1%
School	2.8%

Table 2: Most common rural facilities in England

Amenity	% rural over-representation
Post Box	4.2%
Graveyard	3.3%
Bench	2.9%
Place of worship	2.8%
Hunting Stand	1.6%
Pub	1.6%
Parking	0.84%
Community Centre	0.74%

Table 3: Most over-represented facilities found in wholly rural parishes in England

bution is also not an appropriate model to use and instead decided to focus on the probability that a parish contains a given facility (rather than the number of facilities per parish). Figure 5 shows the probability that a parish of a given size (again, grouped to the nearest 100 people) has at least one of a facility.

Given the (exponential decay style) distributions of the area and population we took the logarithm of those value when constructing our models.

A parish can span multiple LSOAs and yet the indices of multiple deprivation are given on an LSOA basis. Because of this, we will need to consider an approach to producing a singular IMD value for each parish. Given the lack of variation in population size it is probably reasonable to approximate the IMD by taking the mean of the IMDS for output areas in a

Amenity	% non-rural over-representation
Bicycle parking	5.4%
Parking space	3.4%
Fast food	2.7%
Restaurant	2.0%
Waste basket	1.6%
Cafe	0.75%
ATM	0.71%
Pharmacy	0.57%

Table 4: Most over-represented facilities found in non wholly rural parishes in England

Measure	Pub	School	Worship
Degrees of Freedom	6	6	6
χ^2	157	85	891
p-value	$6e^{-30}$	$6e^{-187}$	$5e^{-15}$
σ_p	19.7	111.3	10.6

Table 5: Chi-squared test of the modelling of the number of facilities per parish as a Poisson distribution

parish. However, given little extra complexity we decided to create a population-weighted value of the IMD per parish, whereby we weight the value of the IMDs of a given OA by the percentage of the population who live in that OA.

Combining the above, we are going to build a simple linear (logistic) model describing the relationship between the probability that a parish has a given facility and the population size of a parish, the area spanned by a parish, and the weighted IMD of the parish.

3.3 Differences in the relationships between population and number of facilities by region

There are a number of ways of determining whether there are regional differences in the likelihood of facilities by region and each approach answers a slightly different question. To take pubs for example, if I want to answer the question "if I take a random rural parish in region A, and a random rural parish in region B, which is more likely to have a pub?" then comparing

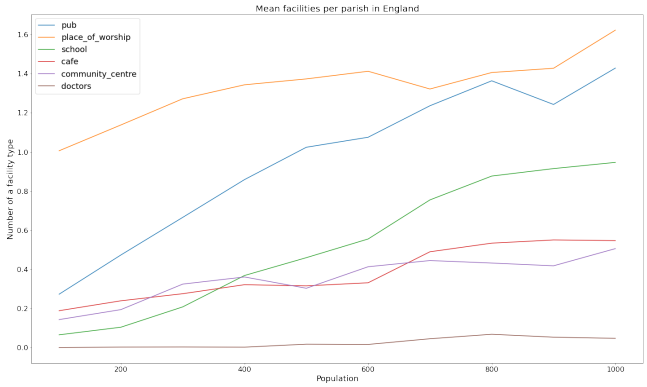

Figure 4: Average number of facilities in rural English parishes by population

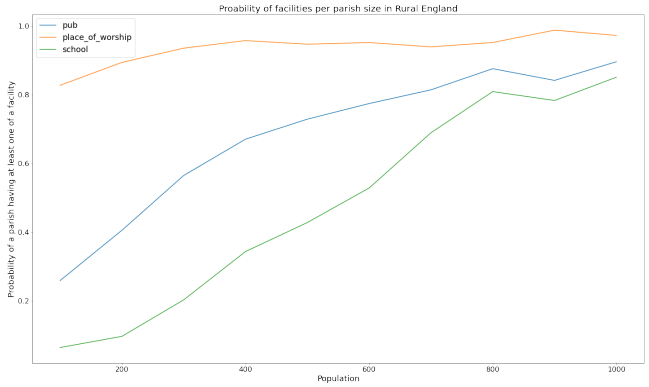

Figure 5: Probability of at least one instance of different facilities in wholly rural parishes in England

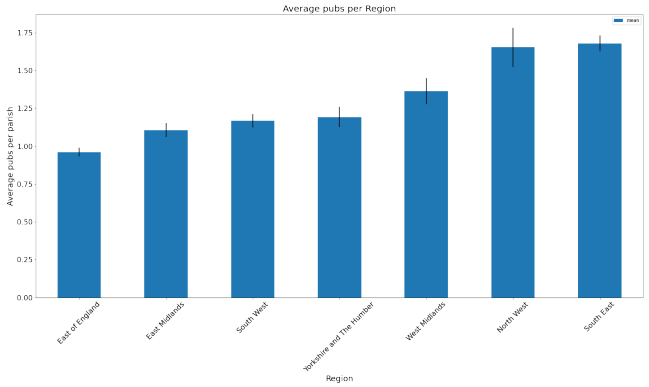

Figure 6: Average number of wholly rural parishes with at least a pub in it, per region

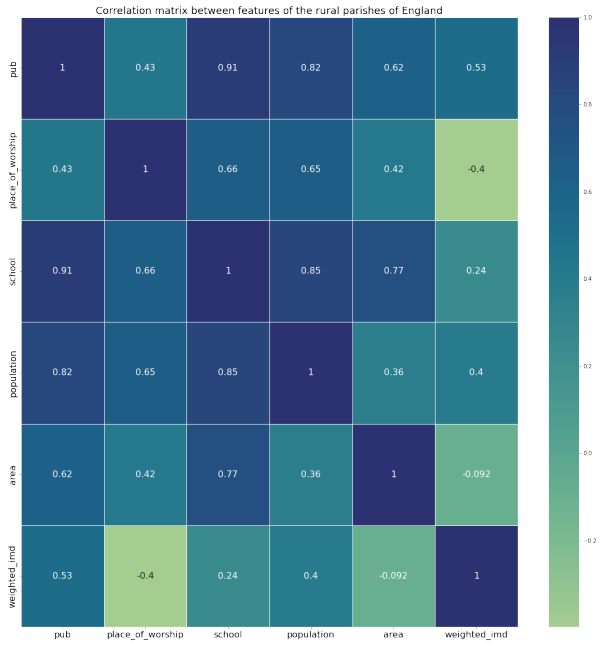

Figure 7: Correlations between the number of certain facilities and measures of a parish

the distribution of pubs in both of those regions is appropriate. This exact example is shown in figure 6. However, given that there are clear differences between the average population sizes in rural parishes (see appendix A), and there are clear differences in the number of pubs based on population sizes (shown in figure 4), it's to be expected that there will be differences in the probability of a parish having a pub in different regions. The correlations between our variables of interest are shown in figure 7.

Given the extremely strong correlations between all pairs of variables (except IMD and area), in this analysis we choose to focus on answering the following question:

"Given parishes of approximately the same area and indices of multiple deprivation, how does the probability of a facility vary by region and population?"

In order to answer this question, we will build the logistic model (as described above) except we will also include a (one-hot encoded, drop-one) feature for the region of England in which the parish resides.

Finally, having fitted a logistic model we will calculate the average weighted IMD and average area of a parish and plot the varying probabilities of a parish of size N having a church, pub and place of worship for different values of N.

4 Results

Table 6 shows the results of running a logistic regression model as described in the methods.

Variable	Pub	School	Worship
$Log(population)$	1.28^{***}	2.25^{***}	0.69^{***}
$Log(area)$	0.39^{***}	0.50^{***}	0.79^{***}
IMD	$3.39e^{-5***}$	$3.10e^{-5***}$	$-1.96e^{-6}$
Intercept	-13.9	-22.7	-13.8
Observations	5064	5064	5064
R^2	0.19	0.39	0.085

$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

Table 6: Probability of facilities per parish

Carrying out the same test but included a one-hot encoded version of the region creates the set of results found in table 7.

Variable	Pub	School	Worship
$Log(population)$	1.28^{***}	2.25^{***}	0.69^{***}
$Log(area)$	0.39^{***}	0.50^{***}	0.79^{***}
IMD	$3.39e^{-5***}$	$3.10e^{-5***}$	$-1.96e^{-6**}$
East Midlands	0.21^{**}	0.54^{***}	-1.27^{***}
Yorkshire & The Humber	0.50^{***}	0.53^{***}	-2.35^{***}
North West	0.055	0.68^{***}	-2.75^{***}
South East	0.81^{***}	0.03	-0.52^{*}
South West	0.22^{**}	0.42^{***}	0.27
West Midlands	0.35^{**}	-0.05	-0.99^{***}
Intercept	-14.1	-22.7	-13.8
Observations	5064	5064	5064
R^2	0.20	0.40	0.20

Note: we have dropped 'East of England' as a region and so the coefficients of the other regions are shown relative to the East of England.

Table 7: Probability of facilities per parish

5 Discussion

Whilst not the focus of this study, there are a number of interesting conclusions that we can draw from the data collected (and likely even more that haven't been considered!). Even simple descriptive statistics such as the proportion of wholly rural parishes per region/county or the proportion of the population who live in wholly rural parishes can give a flavour of the different regions of England. Neither of these have been correlated with political, educational or health markers in this study and yet all would surely yield interesting results.

That parking is the most common amenity in England surely says something about how our public spaces have become dominated by storage for a mode of transport. That bicycle

parking is the most over-represented urban amenity hints either at changes in urban environments or perhaps, the slower rate of change in rural environments. That fast food, restaurants and cafes are more common in urban environments and post boxes, grave yards and places of worship are more common in rural environments will come as little surprise. But it does provide a data-driven backing for commonly held assumptions.

That the number of facilities isn't adequately modelled by either a Poisson distribution or negative binomial distribution is noteworthy. One of the key assumptions of these distributions is that of independence - the likelihood of a new instance of a facility must be independent of any previous instances. We have concrete evidence that that is untrue for pubs, schools, and places of worship. This likely follows our intuition; whether commercial, governmental or ecumenical, the laws of supply and demand are followed.

Studying the results of our various investigations, it is clear that places of worship belong to a completely different category to pubs and schools. Graphically, it's clear to see that the probability of a parish having a place of worship is much less dependent on population size, with a number of small parishes having places of worship. Of course, a large part of this will be due to the history of parishes and their intimate relationship with the church. Statistically, the model fitted (excluding the region) demonstrated the poorest R^2 value and hence had the least amount of variance explained by the model. Given the strong inverse correlation between IMD and places of worship, the increased R^2 score of the model fitted with regions, yet the non-significant relationship between IMD and the probability of a place of worship on the latter model, inferences could be made that the key explanatory factor for why a parish might not have a church stems almost exclusively from regional differences. Yorkshire and The Humber and the North West are the stand-out regions in terms of being less likely to have a place of worship. The could be viewed as especially surprising given that the North West has the highest number of places of worship per parish - yet correcting for population, area and IMD tells a different story. The reason for this under-indexing is currently unknown to the author. The inverse correlation between IMD and probability of a place of worship is interesting, and implies that while parishes that are less deprived are more likely to have pubs and schools, they are less likely to have places of worship.

Pubs and schools both exhibit more similarities than differences in their relationships with population, IMD and area. Pubs are more common than schools and this is especially true at higher populations where the average number of pubs rises much higher much faster than the average number of schools. The increased predictability of schools (R^2 of 0.39 versus R^2 of 0.19 for pubs) is likely explainable through the perhaps more formulaic way in which schools in England are opened and closed and reflects the triumph of central planning over decentralized market-like decision making. Equally, similarities between pubs and schools perhaps reflects the opposite - or maybe it just says that whether decision making on the presence of facilities is the preserve of local entrepreneurs or distant bureaucrats, the same considerations are likely made.

When considering the regional differences in the probability that a parish has a pub or a school, the magnitude of the purely regional differences is smaller than it is for places of worship. Points of note include the fact that there appears to be two clusters of regions that effect the probability of a school being present in a parish - with the East of England, South East and West Midlands seeing fewer schools than would be expected (controlling for population, area and IMD) and the other regions all seeing slightly more. When it comes to pubs there is a broader spread, with the South East leading the pack as the most likely region to have a pub.

5.1 Extensions

We have only scratched the surface here of the possibilities presented by the data set analysed. While the parish boundaries are used as one of the fundamental building blocks of this analysis, in reality the parish boundary doesn't provide any real demarcation for people's day-to-day lives. There are clearly opportunities to use this data to understand which areas people would be interested in living in, and to determine the best locations to build new schools, pubs, homes etc. However, rather than using parish boundaries it would likely be more appropriate to use actual geographic location (latitude and longitude) in order to determine facilities within walking, cycling and driving distance.

Additionally, we deliberately excluded all non-rural parishes and hence didn't study the similarities and differences between England's rural and non-rural areas. In doing so we lost the chance to bring a quantitative lens to the differences between village and city life, and perhaps to find rural-like areas in the urban environments and vice-versa.

Given our focus across the whole of England's rural parishes and three different amenity types, we have been unable to delve deeply into any region or facility type. There are likely fascinating explanations for the differences in prevalence of places of worship in different regions of England, tied to the history of the Church of England and the organisational decisions of the current church. It's not impossible that one could determine the approximate age of a historical parish by considering the facilities that are present (alongside its current population, IMD, area, and other descriptive statistics). In this analysis we didn't cluster parishes together but finding rural parish 'twins' across counties and regions.

The challenge with such a rich data set is not in finding interesting questions to ask, but in narrowing down the questions of interest to answer one in sufficient detail.

5.2 Weaknesses

Perhaps the biggest weakness of this study is the dependence of OpenStreetMap's API. While (as previously discussed) the data is generally of a high quality, we are completely dependent on volunteer submitted amenities and these amenities being a complete and accurate record of the amenities actually in a parish. While we have discovered no reason to doubt those data points (and have checked a number of parishes which we know), it is still an area of great uncertainty. It is especially important to note that we do not check for whether an amenity is open or not. A church that has recently closed

is likely to still be listed as present on OpenStreetMap and so our data is likely at least partially out of date.

Removing all non-rural parishes and discarding regions without sufficient number of parishes left has caused us to remove the North East and London entirely. Before we began we excluded Wales, and we didn't even consider Scotland or Northern Ireland. We have excluded a large amount of data in order to draw strong conclusions specifically about rural areas. It would likely have been prudent to have been more inclusive of semi-rural areas in order to expand our geographic remit.

Finally, we shied away from modelling the number of facilities in a given parish due to the dependence between facilities. However, the distributions of facility per person look like great examples of beta distributions and modelling that distribution could potentially have yielded interesting results.

6 Conclusion

We conclude that there is indeed a relationship between the population size of rural parishes in England and the number of facilities therein, with strong statistically significant results being shown for pubs, schools and places of worship. In determining our modelling approach, we found that neither places of worship, schools or pubs are well modelled by either a Poisson or negative binomial distribution and posit that this is likely due to the non-independence of facilities. We also determined that there are regional differences in the prevalence of those facilities, even when controlling for population, area and indices of multiple deprivation. Finally, we discovered that the average rural parish in England becomes more likely to have a pub than not at somewhere between 200 and 300 people, that the average rural parish becomes more likely to have a school than not at somewhere between 500 and 600 people and that all rural parishes in England are more likely to have a place of worship than not. That means that, unless you are looking at an especially small or deprived village in rural England with around 700 people in it, you should be surprised to not find a pub, school and place of worship. And, more importantly, if you find a village of only 150 people with a pub and a school, you might have found a gem.

References

[Bagley and Hillyard2015] Carl Bagley and Sam Hillyard. 2015. School choice in an english village: living, loyalty and leaving. *Ethnography and education*, 10(3):278–292.

[Bappee et al.2018] Fateha Khanam Bappee, Amílcar Soares Júnior, and Stan Matwin. 2018. Predicting crime using spatial features. In *Canadian Conference on Artificial Intelligence*, pages 367–373. Springer.

[Beaumont and Brown2015] Emily Beaumont and David HK Brown. 2015. 'once a local surfer, always a local surfer': Local surfing careers in a southwest english village. *Leisure Sciences*, 37(1):68–86.

[Bright et al.2018] Jonathan Bright, Stefano De Sabbata, Sumin Lee, Bharath Ganesh, and David K Humphreys. 2018. Openstreetmap data for alcohol research: Reliability assessment and quality indicators. *Health & place*, 50:130–136.

[Bruce and Glendinning2010] Steve Bruce and Tony Glendinning. 2010. When was secularization? dating the decline of the british churches and locating its cause 1. *The British journal of sociology*, 61(1):107–126.

[Burchardt2011] Jeremy Burchardt. 2011. Rethinking the rural idyll: The english rural community movement, 1913–26. *Cultural and Social History*, 8(1):73–94.

[Cabras2011] Ignazio Cabras. 2011. Industrial and provident societies and village pubs: exploring community cohesion in rural britain. *Environment and Planning A*, 43(10):2419–2434.

[Feldmeyer et al.2020] Daniel Feldmeyer, Claude Meisch, Holger Sauter, and Joern Birkmann. 2020. Using openstreetmap data and machine learning to generate socioeconomic indicators. *ISPRS International Journal of Geo-Information*, 9(9):498.

[for National Statistics2022a] Office for National Statistics. 2022a. England. https://www.ons.gov.uk/methodology/geography/ukgeographies/administrativegeography/england#parishes-and-communities. Accessed: 2023-01-05.

[for National Statistics2022b] Office for National Statistics. 2022b. Statistical geographies. https://www.ons.gov.uk/methodology/geography/ukgeographies/statisticalgeographies. Accessed: 2023-01-05.

[for Science2021] Government Office for Science. 2021. Trend deck 2021: Urbanisation. https://www.ft.com/content/88338ca6-7b0c-4277-9e6a-8a39dd825f5f. Accessed: 2023-01-05.

[Heley2011] Jesse Heley. 2011. On the potential of being a village boy: An argument for local rural ethnography. *Sociologia Ruralis*, 51(3):219–237.

[Hillyard2020] Sam Hillyard. 2020. The enduring insignificance of a school for its village: An english case study. *Journal of Rural Studies*, 80:618–625.

[Hristova et al.2016] Desislava Hristova, Matthew J Williams, Mirco Musolesi, Pietro Panzarasa, and Cecilia Mascolo. 2016. Measuring urban social diversity using interconnected geo-social networks. In *Proceedings of the 25th international conference on world wide web*, pages 21–30.

[Jenkins1999] Timothy Jenkins. 1999. *Religion in English everyday life: an ethnographic approach*, volume 5. Berghahn books.

[Keenan2020] Liam Keenan. 2020. Financialization, securitization and the decline of pubs in britain. *Journal of Economic Geography*, 20(6):1293–1311.

[Kounadi2009] Ourania Kounadi. 2009. Assessing the quality of openstreetmap data. *Msc geographical information science, University College of London Department of Civil, Environmental And Geomatic Engineering*, page 19.

[Lathia et al.2012] Neal Lathia, Daniele Quercia, and Jon Crowcroft. 2012. The hidden image of the city: sensing community well-being from urban mobility. In *International conference on pervasive computing*, pages 91–98. Springer.

[Markham and Bosworth2016] Claire Markham and Gary Bosworth. 2016. The village pub in the twenty-first century: Embeddedness and the "local". In *Brewing, Beer and Pubs*, pages 266–281. Springer.

[Miller2016] Daniel Miller. 2016. *Social media in an English village*. ucl Press.

[Mooney et al.2017] Peter Mooney, Marco Minghini, et al. 2017. A review of openstreetmap data.

[Neal and Walters2006] Sarah Neal and Sue Walters. 2006. Strangers asking strange questions? a methodological narrative of researching belonging and identity in english rural communities. *Journal of Rural Studies*, 22(2):177–189.

[Neal and Walters2008] Sarah Neal and Sue Walters. 2008. Rural be/longing and rural social organizations: Conviviality and community-making in the english countryside. *Sociology*, 42(2):279–297.

[ONS2021] ONS. 2021. About the census.

[ONS2022a] ONS. 2022a. official census and labour market statistics.

[ONS2022b] ONS. 2022b. Open geography portal.

[Payne and Abel2012] Rupert A Payne and Gary A Abel. 2012. Uk indices of multiple deprivation-a way to make comparisons across constituent countries easier. *Health Stat Q*, 53(22):2015–2016.

[Payne et al.2018] Elizabeth H Payne, Mulugeta Gebregziabher, James W Hardin, Viswanathan Ramakrishnan, and Leonard E Egede. 2018. An empirical approach to determine a threshold for assessing overdispersion in poisson and negative binomial models for count data. *Communications in Statistics-Simulation and Computation*, 47(6):1722–1738.

[Ratti et al.2006] Carlo Ratti, Dennis Frenchman, Riccardo Maria Pulselli, and Sarah Williams. 2006. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and planning B: Planning and design*, 33(5):727–748.

[Rightmove2020] Rightmove. 2020. Why are so many buyers escaping to the country? `https://www.rightmove.co.uk/news/articles/property-news/escape-to-the-country`. Accessed: 2023-01-05.

[Robinson2021] B Robinson. 2021. From domesday to wfh: the lure of english village life. `https://www.ft.com/content/88338ca6-7b0c-4277-9e6a-8a39dd825f5f`. Accessed: 2023-01-05.

[Sehra et al.2013] Sukhjit Singh Sehra, Jaiteg Singh, and Hardeep Singh Rai. 2013. Assessment of openstreetmap data-a review. *arXiv preprint arXiv:1309.6608*.

A Appendix A: Descriptive graphs of the rural parishes of regions of England

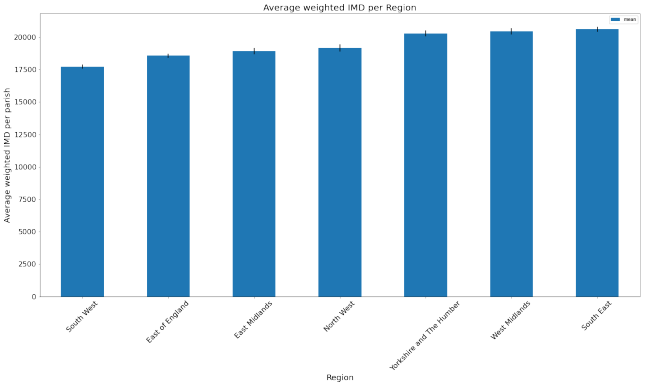

Figure 8: Weighted IMD of rural parishes by region

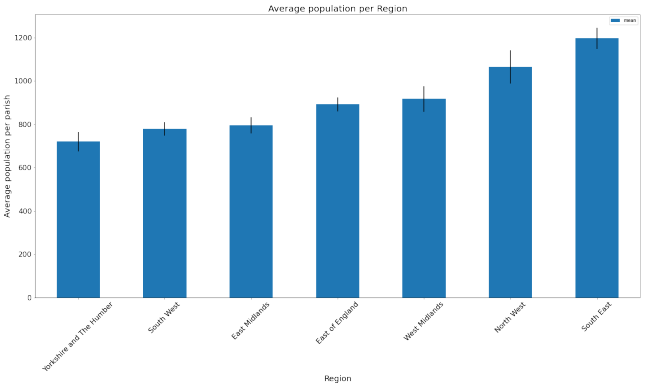

Figure 9: Average population of rural parishes by region

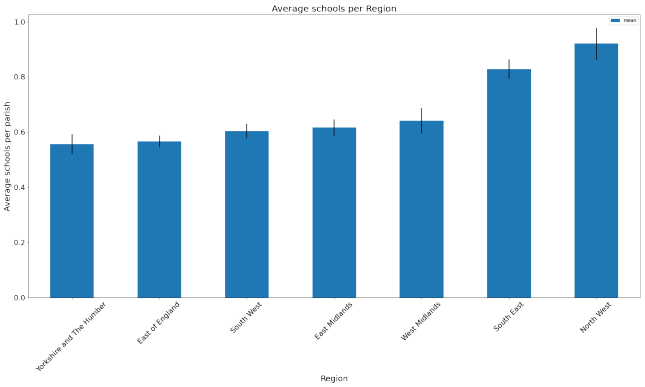

Figure 10: Average number of schools of rural parishes by region

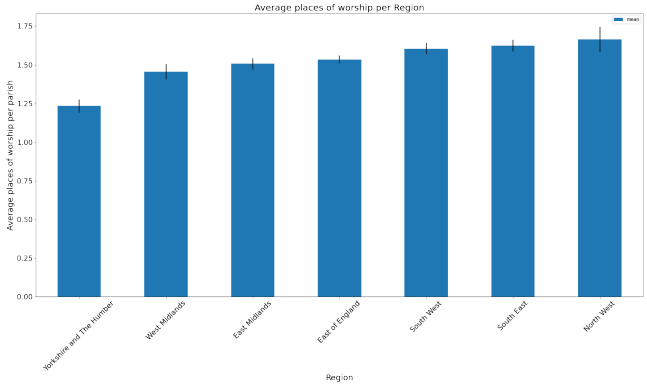

Figure 11: Average number of places of worship of rural parishes by region

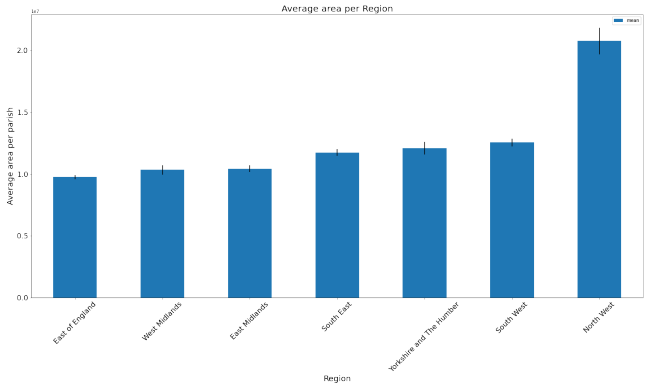

Figure 12: Average area of rural parishes by region