Scores on the Doors

An analysis of Food Hygiene Ratings across the UK

MMathStat Project

School of Mathematics, Statistics and Physics
Newcastle University

James Salsbury

May 7, 2021

This report looks at the distribution of Food Hygiene Ratings in England. We have been able to visualise the establishment data on a map, by grouping establishments by postcode districts. Using this idea of visualisation, we have created an interactive Shiny app (which can be accessed here: `https://jamesalsbury.shinyapps.io/interactiveshinymap/`) to view summary statistics and individual establishments in each postcode district. Trends of higher/lower rated establishments are seen in different areas of the country so we have investigated the potential reasons for this. To do so, we looked at multiple factors such as the type of establishment, whether the establishment is a food chain, deprivation data and the local authority. We also looked at whether there are clusters of similar rated establishments in city centres and the potential reasons for this.

Contents

1 Introduction

1.1 Background

Every food establishment in the UK (England, Northern Ireland, Scotland and Wales) is required, by law, to have a food hygiene inspection[1]. There are three main reasons for this:

- to ensure establishments are complying with the current guidelines;

- to ensure the premises are fit for purpose; and

- to obtain a Food Hygiene Rating, which is made available to the general public.

Establishments in England, Northern Ireland and Wales are given a Food Hygiene Rating of between 0 and 5, with 0 being the worst and 5 being the best[1] (whole integers only). Scotland has a different system for rating establishments[2], here establishments are not given a numeric score, they are only given a rating of *Pass* or *Improvement Required*.

Inspections are carried out by Environmental Health Officers (EHOs) which are employed by local authorities. During an inspection, EHOs mark each establishment on the following criteria: hygiene, structural and management[1].

- **Hygiene**: an indicator of how well the food is being stored, prepared and cooked.

- **Structural**: the layout of the premises - inspectors are looking for cleanliness, ventilation and pest control.

- **Management**: the standard of the paperwork and training; how confident the inspectors are that the standards seen will be maintained after the inspection.

Table 1.1.1 shows how the EHOs quantify the inspections. The **better** the establishment performs in each category, the **fewer** marks the establishment is awarded. From the table, we see that the scores in each of the categories range from 0 to 25/30, they increase in increments of either five or ten and 0 is the best score in each category.

To calculate a Food Hygiene Rating the scores from the three categories are added together to give an overall score for that establishment. From this, we deduce that the best overall score an establishment could obtain is 0 and the worst score is 80 $(25 + 25 + 30)$. How these overall scores are mapped onto the ratings is given in Table 1.1.2.

Also, from Table 1.1.2 we see that it is not just the overall score that is considered when calculating the rating. Establishments also need to ensure they do not exceed

Criteria	Score					
Hygiene	0	5	10	15	20	25
Structural	0	5	10	15	20	25
Management	0	5		10		20	30
	Better			$\rightarrow$		**Worse**

Table 1.1.1: Scoring system used when inspecting establishments. [3]

Score	Value					
Overall score	0-15	20	25-30	35-40	45-50	> 50
Highest permitted score	5	10	10	15	20	-
Rating	5	4	3	2	1	0
	Better		$\longrightarrow$			**Worse**

Table 1.1.2: How establishment ratings are calculated. [3]

certain scores in any one category, otherwise they will be excluded from obtaining the rating their overall score is originally mapped to. For example, an establishment which scores 0 for *hygiene*, 0 for *structural* and 15 for *management* will obtain an overall score of 15. As seen in Table 1.1.2, this score is low enough for the establishment to be given a rating of 5, but due to the management score being 15, the rating obtained will be 2. This highlights how important it is for establishments to be consistent with standards throughout their business.

In Northern Ireland and Wales it is a legal requirement for establishments to display their rating sticker[1], as found in Figure 1.1.1, in a prominent position. However, this is not the case in England where it is at the discretion of the establishment owners as to whether they display their rating.

According to research carried out in 2017,[4] only 49% of establishments in England display their rating sticker, compared to over 80% in both Northern Ireland and Wales. In England, the rate of display is directly linked to the rating of the establishment; 67% of establishments that have a rating of 5 display their rating compared to only 28% when their rating is between 0 and 3 inclusive.

1.2 The establishment data

1.2.1 Obtaining the data

The establishments and all their corresponding data, including ratings, are stored in a database by the *Food Standards Agency*[5]. The data is available to be downloaded using an Application Programming Interface (API). An API is a request made to a server by a user, which then provides the information that was requested[6].

Figure 1.1.1: Food Hygiene Rating stickers

An example of when an API is used is when searching for plane tickets online. There are multiple options available: departure city, destination city, number of seats required, dates of travel etc. When all these variables are entered into the search engine, an API searches the many different airlines databases to find flights matching the request. Once the API has collated all of this information, the findings are presented in a user-friendly way. This example is found in the flowchart in Figure 1.2.1.
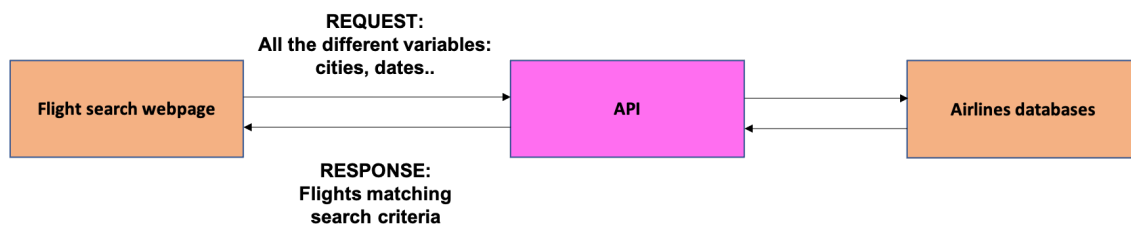

Figure 1.2.1: How an API works for the plane tickets example.

In the case of Food Hygiene Ratings, we wanted to obtain hygiene ratings for all UK establishments. The *Food Standards Agency* API allowed us to programmatically retrieve this data. This API allowed us to specify a number of parameters, including: local authority, type of establishment, establishments within x radius of a certain postcode and name of establishment.

In this context, each local authority has a *local authority code*; an integer between 1 and 409. To retrieve the full data set, we called the API for each individual local authority. At each iteration, the API returned a JavaScript Object Notation (JSON) file. A JSON file is a human-readable text file which is used to store and transmit data between a browser and a server. There were 37 columns in this JSON file, containing information about the establishments, from the addresses and ratings to the metadata. The data was then combined into a single resource. As this was a time-consuming process, we took a snapshot of the data on 2020-10-19. This data is used for all analysis in this report.

1.2.2 What data will be considered?

We have built statistical models to try and explain the distribution of ratings across the country, this is seen in Chapter 3. One of the covariates we used is the deprivation score of the postcode of the establishment. Deprivation data is available for England, Northern Ireland and Wales. However, each country compiles their own deprivation data and therefore they are not directly comparable. For example, the 10% most deprived areas in England are not necessarily equal to the 10% most deprived areas in Northern Ireland. Also, as mentioned previously, in Scotland establishments only obtain a *Pass* or *Improvement needed* rating.

As a result, we have only considered establishments in England from now on. This ensures that our findings are robust and we have consistency throughout the project.

In the full downloaded data set, there are a total of 551,690 establishments, of which, 82% (452,823) are in England. For each English establishment, we have the following data: name, type, address, postcode, longitude, latitude, score for hygiene, score for structural, score for management, authority number, authority name and Food Hygiene Rating. For example, Table 1.2.1 gives the data associated with McDonald's on Northumberland Street, Newcastle upon Tyne.

Name	Type	Address	Postcode	Longitude	Latitude
McDonald's	Restaurant/Cafe/Canteen	64-68 Northumberland Str.	NE1 7DF	-1.61218	54.97587

Hygiene score	Structural score	Management score	Authority num.	Authority name	Rating
5	5	5	416	Newcastle upon Tyne	5

Table 1.2.1: Data for McDonald's, Northumberland Street, Newcastle.

Establishment type

There are fourteen establishment types, see Table 1.2.2. Some establishment types are ambiguous, for example "Retailers - other" is used for shops which prepare and sell food such as bakeries, cake shops and butchers. This will also include shops on the high street which are selling pre-packaged food that may need to be refrigerated. Another type of establishment is "Other catering premises", these consist of religious sites, farm shops and sports clubs which may prepare buffet-style food in their clubhouses.

1.3 Rating and overall score of establishments

We also examine the different values for the ratings, see Table 1.3.1. From the table we observe that 27,049 establishments were exempt from having a rating. There are two groups of establishments that are exempt[1]:

- establishments that are low-risk to public health, for example; newsagents, chemist shops or visitor centres selling pre-wrapped goods that do not require refrigeration; and

Type	Count	Type	Count
Restaurant/cafe/canteen	102,592	Retailers - other	87,784
Other catering premises	53,585	Takeaway/sandwich shop	47,592
Pub/bar/nightclub	44,759	Caring premises	34,592
School/college/university	27,109	Mobile caterer	20,105
Hotel/bed & breakfast/guest house	13,727	Retailers - supermarkets/hypermarkets	11,528
Manufacturers/packers	6,244	Distributors/transporters	2,070
Farmers/growers	767	Importers/exporters	369

Table 1.2.2: Counts of different types of establishments.

- establishments offering caring services at home.

The establishments that are exempt are still inspected to ensure they are complying with the criteria set out by the Food Standards Agency, but they are not given a Food Hygiene Rating. The establishments that were awaiting inspection, awaiting publication or were exempt account for $\approx 14\%$ (63,559) of establishments. When using the establishment ratings for analysis, we have used the numerical values (either the ratings themselves or the broken down category scores) but here, we don't have any ratings to use to perform statistical inference. As a result, we discarded them and we now have 389,264 establishments in the data set.

Rating	Count
0	513
1	6,317
2	7,023
3	27,216
4	63,223
5	284,972
Awaiting inspection	36,498
Awaiting publication	12
Exempt	27,049

Table 1.3.1: Breakdown of ratings.

In Figure 1.3.1, we see a bar chart of the remaining establishments. From the plot, we see that the data is strong positively skewed, with $\approx 73\%$ (284,972) of establishments rated 5, and $\approx 89\%$ (348,195) of establishments rated either 4 or 5.

When analysing Food Hygiene Ratings, one of the consequences of a large proportion of establishments obtaining high ratings is that it is harder to distinguish differences between them. This could be troublesome when we are fitting statistical models to the data as there is so little to choose between establishments. One of the ways that we can identify differences between establishments is by looking at the scores for the
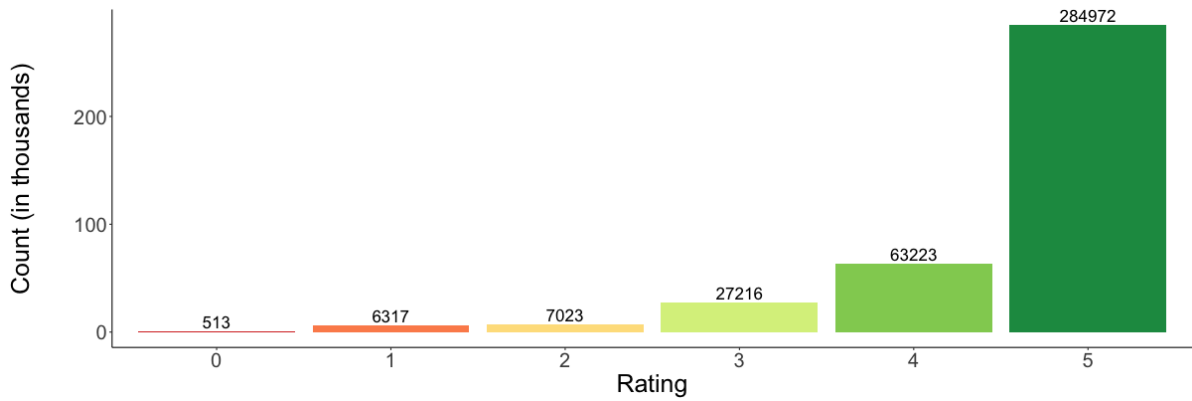
Figure 1.3.1: Ratings of the establishments that have a numeric rating.

individual categories. As explained in Section 1.1, if we combine the scores from the three categories (hygiene, structural and management) we obtain an overall score, which we will also refer to as the **raw overall score**. There are certain circumstances when using these raw overall scores for analysis is more appropriate than using the ratings, which we will see in Section 3.5.

In Figure 1.3.2, we see a bar chart of the *raw overall scores* for the 389,264 establishments in the data set, coloured by rating. We see that the data is still skewed (negatively this time as low raw overall scores correspond to higher ratings) but we are now able to discriminate between establishments more here, with only $\approx 19\%$ (74,996) of establishments obtaining the lowest raw overall score of 0.
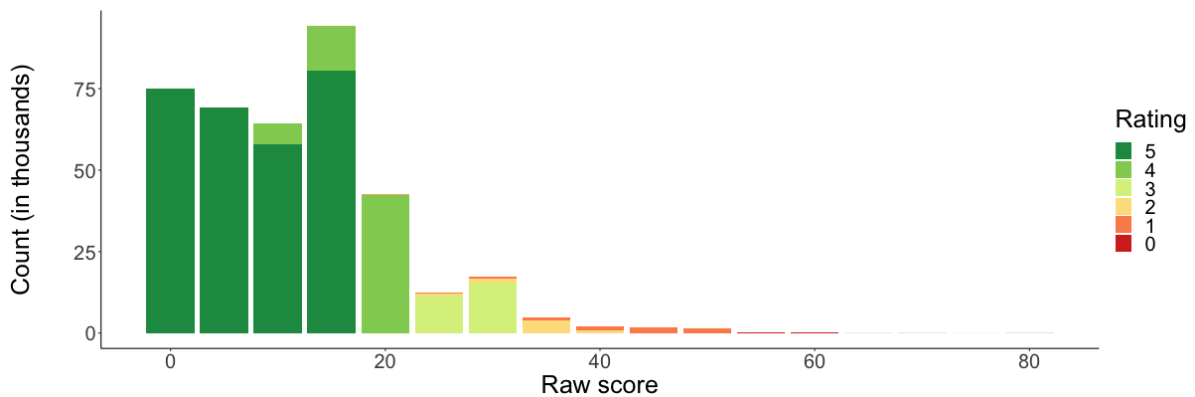

Figure 1.3.2: Raw overall scores of the establishments, filled by ratings.

In Figure 1.3.3, we see bar charts for the scores obtained by establishments in the three different categories: hygiene, structural and management. We see that the bar charts for structural and management look very similar, with five being the mode score. However, the plot for hygiene is a little different, with zero being the mode score. This is reflected in Table 1.3.2 where we see the mean score for the three categories. When considering these mean values we need to account for the fact that the scores given in

6

the management category are a little higher than the other two, as outlined in Table 1.1.1. However, in general we see that the scores given for hygiene seem to be lower than for the other two categories.

A possible reason for establishments obtaining better scores for hygiene could be that establishments potentially find it easier to adapt their hygiene processes, (e.g. handwashing and storing of food) compared to changing the layout of their premises, improving ventilation or ensuring the paperwork is of the standard required by the regulators.

Criteria	Mean score
Hygiene	3.35
Structural	4.10
Management	4.27

Table 1.3.2: Mean score of the three different categories.

	Hygiene	Structural	Management
Hygiene	1.00	0.53	0.60
Structural		1.00	0.53
Management			1.00

Table 1.3.3: Correlation matrix showing the correlations between the three categories.

Table 1.3.3 shows the correlations between the three categories. We see that all of the correlations are similar; the correlation between hygiene and structural is the same as the correlation between structural and management (both 0.53). However, the correlation between hygiene and management is slightly higher at 0.6. As the correlation coefficients are positive, this indicates that establishments score similarly in the three categories,
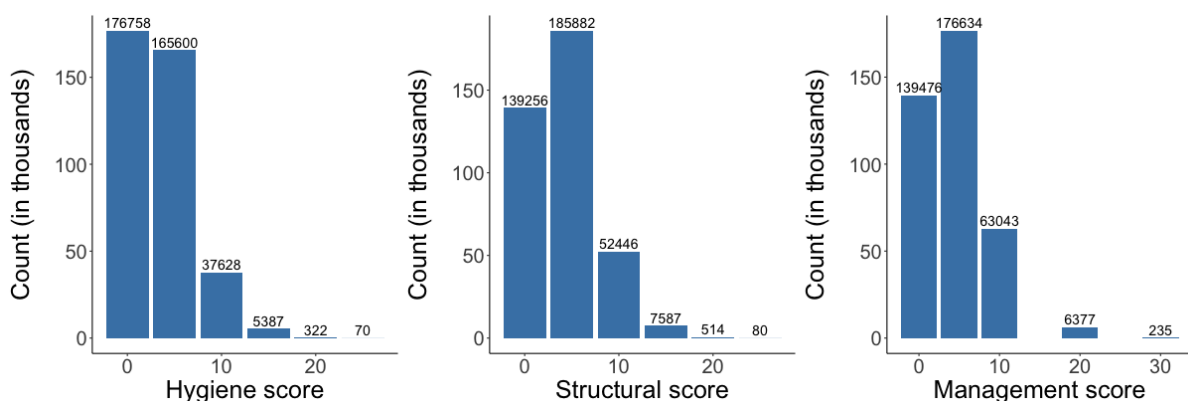

Figure 1.3.3: Scores for the 3 different criteria: (a) hygiene, (b) structural, (c) management.

which is what we would expect. In fact, we calculate that nearly half (44%, 170,923) of establishments score **exactly** the same in the three categories.

2 Visualising the Data

2.1 Introduction

The data has a spatial component. As there are $\approx$ 400,000 establishments in England with ratings, it is not practical or informative to simply plot all the establishments at once on the same map as we would become overwhelmed by the data and we would not be able to identify geographical trends.

One of the variables supplied (98% of establishments) is the postcode. Of the establishments that do not have a postcode attached, some of these are mobile caterers who, by definition, have no fixed location.

A postcode is made up of several constituent parts; most postcodes are of the form `LLNN NLL`, where an `L` represents a letter and an `N` represents a number[7]. The first set of letters represent the *postcode area*, for example all postcodes in **Ne**wcastle upon Tyne and its surrounding areas have the postcode area NE and all postcodes in **S**heffield and its surrounding areas have the postcode area S. Each postcode area is further broken down into *postcode districts* which are identified by the first set of numbers in the postcode. For example, NE**1** is the postcode district for the city centre of Newcastle upon Tyne and NE**20** is the postcode district for the town of Ponteland in Northumberland.

2.2 Why use postcode districts?

In England, there are 104 postcode areas and $2,356$ postcode districts in which there are establishments. From Section 1.3, we have 389,264 establishments in England with a Food Hygiene Rating. This gives around 165 establishments per postcode district. As such, it is useful to use postcode districts to visualise the data set on a map. This approach is sensible as we have a large number of districts so we are able to see geographical trends across the country. We also have enough establishments within each postcode district such that the findings within the postcode districts are robust and will not be skewed by a few outliers.

Additionally, postcode areas and districts are already, by definition, normalised per population. On average, there are about half a million residents per population area[8]. We assume that establishments are spread across the country broadly proportional to the population, which makes sense as more people in an area usually means more establishments.

2.3 Spatial data in R

We established that viewing the establishments by postcode district is a good idea, therefore we needed some way of actually doing so. There are a few R packages specifically designed for working with spatial data, so we make use of them here. Two of them are: {sf}[9] (which stands for simple features) and {sp}[10] (spatial, abbreviation), both are used when manipulating data with spatial properties or plotting spatial data on a map.

The simple features package, {sf}, is so called because when working with spatial data, there are different geometrical objects that are used, which are also called features. These include Point, Linestring and Polygon. In the context of spatial data, we immediately see what these different objects might be used for. Points are used for singular objects with a longitude and latitude, such as trees or postboxes, Linestrings are used for objects such as roads and rivers and Polygons are used for two-dimensional geometrical shapes. When working with spatial data, it is helpful to consider what types of features the objects you are considering will be classified as. In our case, we treat all of the establishments as Points as they exist as a single point in space (here, we are ignoring the dimensions of the actual restaurant, factory etc.) with a longitude and latitude. The postcode districts we use to view geographical trends of ratings will be Polygons as they are two-dimensional geometrical shapes.

A Coordinate Reference System (CRS) refers to the way in which spatial data represents the earth's surface, this is known as the *projection*. Different map projections cause the underlying map to look different and therefore any spatial features plotted on the map will also look different.

Files containing simple features are called shapefiles. Shapefiles can consist of one type of simple feature or they can have many different features in the same file. For our objective of viewing the establishments by postcode district we needed to import the appropriate shapefiles into R. We required shapefiles consisting of postcode districts in England, Scotland and Wales (some of the establishments are in postcode areas that span England and Wales or England and Scotland, for example LL is primarily a North Wales postcode area but there are 4 establishments in Shropshire within the postcode area LL14. As a result, there are certain "Welsh" or "Scottish" postcode areas we needed to consider).

The postcode district shapefiles are freely available on GitHub[11] as GeoJSON files. A GeoJSON file is a type of shapefile and it inherits all of the properties of a shapefile. The files on GitHub classify all of the postcode districts as Polygons and they are labelled so we are able to identify which polygon corresponds to which postcode district. These Polygons are connected to the WGS84 Coordinate Reference System so this is what we will use whilst viewing any spatial maps.

When importing the GeoJSON files into R, we found issues with four postcode areas (out of the 104); BS (Bristol), CV (Coventry), EX (Exeter) and LL (Llandudno). R could not read the contents of the file correctly which meant that the whole postcode area was unable to be used. Upon further investigation, we discovered that this was because there were multiple polygons which represented the same postcode district. This is highlighted by Figure 2.3.1 as we see that there are two polygons which represent the same postcode
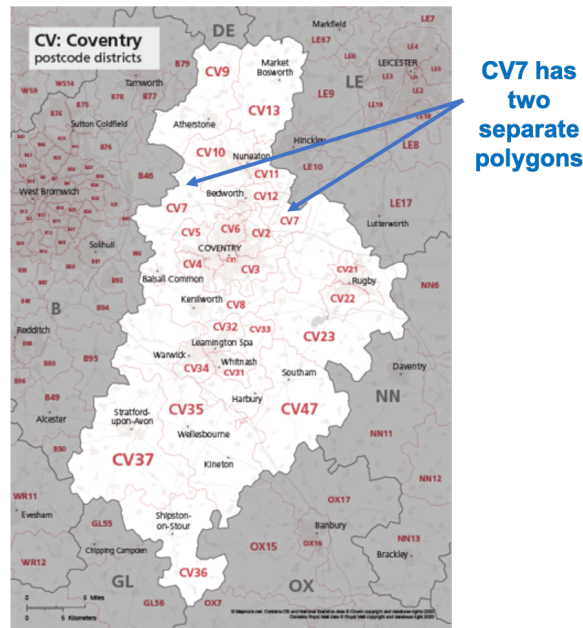
district CV7.

Figure 2.3.1: CV (Coventry) postcode districts - notice the two distinct CV7 polygons.

When these files are read into R, an error occurs as the two polygons both have the label "CV7", and both Polygons cannot represent the same element. A way of overcoming this was to change the raw file by renaming the Polygons, for example: CV7a and CV7b. As only four out of the 104 postcode areas had this problem, it was possible to fix all issues.

Once these files have been fixed, we now have the 104 postcode areas that have English establishments in them, they are all merged together to form a SpatialPolygonsDataFrame. A SpatialPolygonsDataFrame is a data frame which is able to represent polygons and its associated data. All of the postcode districts are then plotted onto a map using another spatial R package, {Leaflet}[12]. {Leaflet} is used for creating interactive spatial maps, it is written in JavaScript and it creates a base map of the world upon which different layers are added. This map containing all of the postcode districts is seen in Figure 2.3.2.

2.4 Mean Food Hygiene Ratings by postcode district

In our establishments data set, a new column was created called *postcode area* using the given postcode. This column was created using regular expressions;[13] we extracted the postcode area from the postcode by extracting characters that are capital letters until we encountered a character that is not a capital letter. This was implemented using the regular expression: str_extract(postcode, "[A-Z]+"). Regular expressions were also used to extract the postcode district, we did this by extracting numbers until we encountered
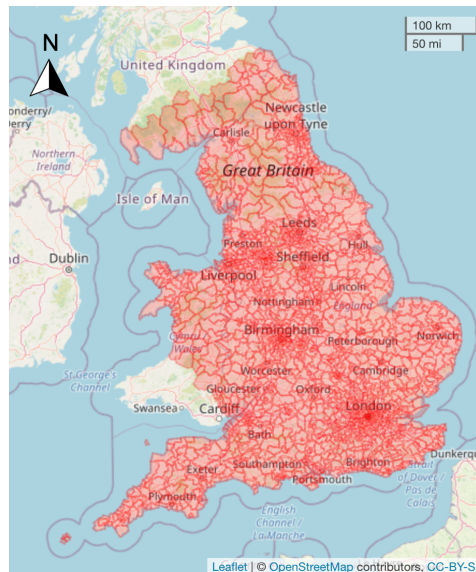
Figure 2.3.2: Postcode districts in England (some Scotland and Wales too).

a space. This was implemented using the regular expression: `str_extract(postcode, "\\w+"))`, this new column was also added to the data set.

From here, data was grouped by their postcode district and then a mean rating and mean raw overall score was calculated for each group. Once these had been calculated, these data were attached to its appropriate polygon in the `SpatialPolygonsDataFrame`. This was done by matching the postcode district to the name of the polygon, NE1 for example.

This worked correctly for the majority of the polygons - but we encountered problems when considering the polygons that we fixed by hand (e.g. CV7a from Section 2.3). This was because the postcode district did not match up with the name of any polygon. This was fixed by matching up the data to the appropriate polygon explicitly.

We have three pieces of information about each postcode district: mean rating, mean raw overall score and number of establishments. The mean rating and mean raw overall scores were used to create a choropleth map. Choropleth maps are used to represent statistical data through shading patterns on geographic areas. They are good for representing variability of the desired measurement, across a region[14]. We see these choropleth maps in Figure 2.4.1. Both of these maps are interactive - when the user hovers their cursor over a postcode district a label with the district name, count of establishments and mean rating/raw overall score is shown - and are accessed here:

https://rpubs.com/jamesalsbury/765878
https://rpubs.com/jamesalsbury/765880

Figure 2.4.1(a) shows the postcode districts coloured by their mean rating. In Figure 2.4.1(b), we see the same postcode districts, this time coloured by their mean raw overall score (remembering that the lower the score the better in this case). There are a few postcode districts that are coloured in grey, in general these are postcode districts in
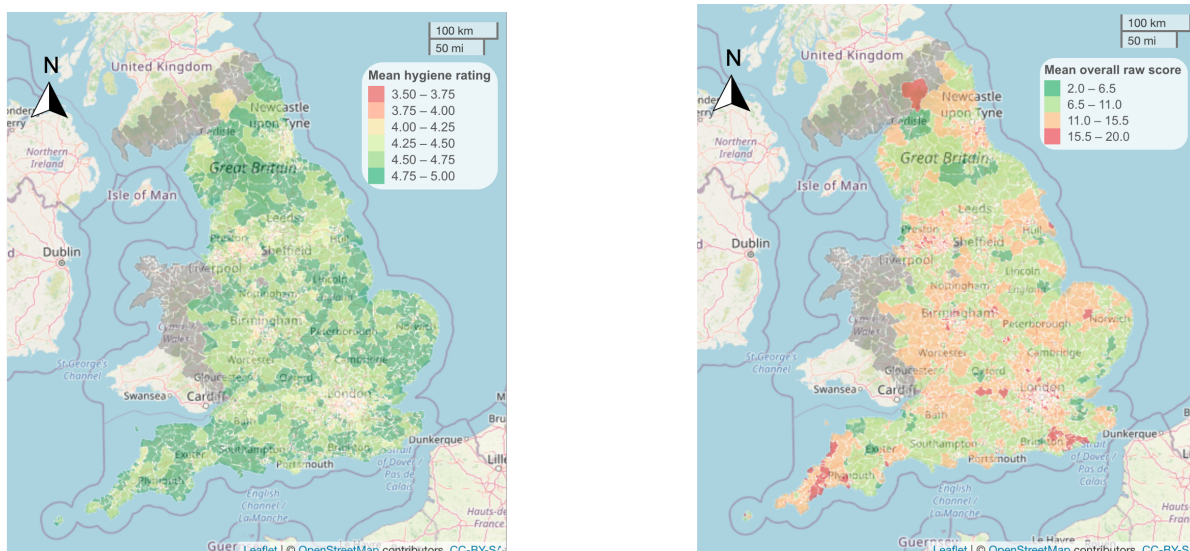
Figure 2.4.1: (a) Mean Food Hygiene Rating for postcode districts in England. (b) Mean overall score for postcode districts in England.

Scotland and Wales for which we do not have data in. However, there are a few in England (near Manchester and Stoke-on-Trent) which are grey too. This is because when the local authority inputted the establishments and their associated ratings into the database, they did not include a postcode so we are unable to determine which postcode district they lie in. These account for a very minimal number of establishments so we use the data set as it is.

If there was no correlation between geographical location of the establishments and the ratings obtained then we would see random scatter of mean ratings across the country. From both the plots in Figure 2.4.1 we see this is clearly not the case. This is especially clear in (b) as we see very large clusters of the same colour. This indicates that as we move across the country, the underlying scores of the establishments change somewhat. We will investigate a few potential reasons for this.

Looking closer at (b), the locations of clusters of higher scoring (worse) establishments normally coincide with city centres. We see that Sunderland, Liverpool, Manchester, Birmingham, and London all have "hot-spots" of higher scoring postcode districts, compared to more rural areas. Again, there are a few potential reasons for this: city centres may attract different types of establishments, for example there are probably more takeaways in Newcastle upon Tyne city centre than rural Northumberland. And then in turn, takeaways may perform worse in food hygiene inspections.

Another reason for higher raw overall scores (lower ratings) in urban areas could be deprivation; 12% of people living in urban areas live in an area that is in the top 10% most deprived areas, this drops to only 1% of people when we consider rural areas[15]. This indicates that we may be able to find a link between Food Hygiene Ratings and deprivation data, which we have looked at extensively in Chapter 3.

Also, in some sense, the owners of an establishment may feel like they only need

to have similar ratings than their neighbouring establishments. If an establishment is surrounded by establishments which have ratings of 3 or 4 then there is potentially less incentive for the owners of the establishment to spend all their resources - both time and money - to obtain a rating of 5. On the other hand, if all the neighbouring establishments are rated 5 then there could be more pressure on the establishment to not stand out as being rated lower - as this could reduce custom somewhat. This is something we have looked at in Section 3.5.

As visualising the data in this way is clearly very informative we have developed a Shiny app. The app allows the user to explore a map of England - broken down into postcode districts - with summary statistics and bar charts for every postcode district. Users are also able to view all the establishments in a postcode district by their ratings, so they see where the higher/lower rated establishments are found. The process of development and the mechanisms behind the app is found in Chapter 4.

3 Modelling

3.1 Introduction

The primary purpose of this research is to determine whether Food Hygiene Ratings are randomly scattered around the country or whether there are clusters of better/worse performing establishments. Also, if this is the case, we must ask what the potential reasons could be for this. One method of analysis we have implemented is to create models to predict the rating of establishments, which include covariates that could help to explain the distribution. This is discussed in Sections 3.2, 3.3 and 3.4.

3.2 Deprivation data

3.2.1 Why look at deprivation data?

As discussed at the end of Section 2.4, it makes sense to consider deprivation data as a covariate because we saw the differences in ratings in rural/urban areas and we know that deprivation can be associated with being in an urban area. The way deprivation data may directly affect ratings is if an establishment is in an area of higher deprivation then potential customers may have less disposable income to spend in/at restaurants. As a result of this, these establishments make less money and cannot afford to spend money on ways to increase their Food Hygiene Rating.

3.2.2 Indices of deprivation data

To build a statistical model that includes deprivation data, we needed to import the data and match it up with our establishment data set. From Section 2.1 we know that the majority of establishments have a postcode. Therefore, it made sense to obtain deprivation data for every postcode in England and then match this up with the postcode of the establishment.

The government publish reports on deprivation every four to five years. We have used the data published in 2019[16]. England is broken up into small areas called Lower Layer Super Output Areas (LSOAs) for purposes such as the census and more importantly in our case, deprivation data. There are 32,844 LSOAs in England, and several deprivation statistics are collected on each area. An overall deprivation score is calculated by linearly combining seven of these statistics, weighted as follows:

- Income Deprivation (22.5%)

- Employment Deprivation (22.5%)

- Education, Skills and Training Deprivation (13.5%)

- Health Deprivation and Disability (13.5%)

- Crime (9.3%)

- Barriers to Housing and Services (9.3%)
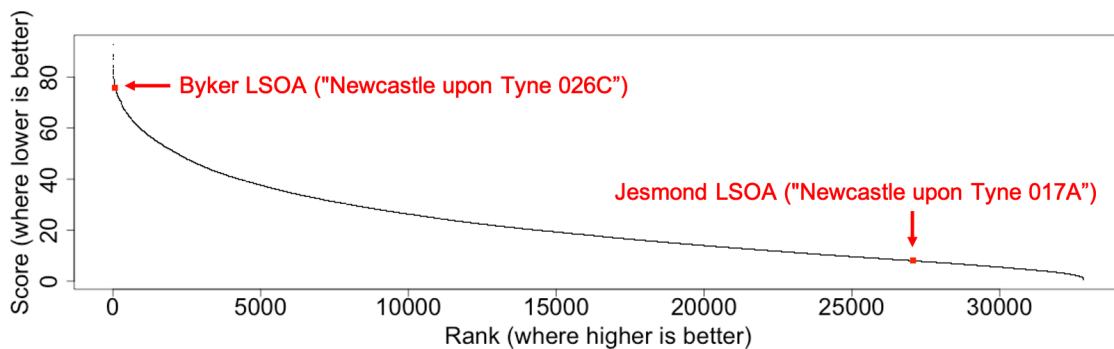
- Living Environment Deprivation (9.3%)

Figure 3.2.1: Score vs rank for all the LSOAs in England.

For each of the seven criteria, LSOAs are given higher scores for performing worse. This means that the higher the overall score for the LSOA, the more deprived it is. The LSOAs are then sorted by score and then given a rank in comparison to all the other LSOAs. This is highlighted in both Figure 3.2.1 and Table 3.2.1, where we visually see that the higher the score, the lower the LSOA ranks and the more deprived the LSOA is.

Also from Figure 3.2.1 we see that the line seems to decline very sharply at the start and then decreases more consistently. This is reflected in the fact that 91.5% of LSOAs obtain scores of less than 50% of the maximum (92.7/2=46.4), this indicates that the majority of LSOAs obtain fairly similar scores.

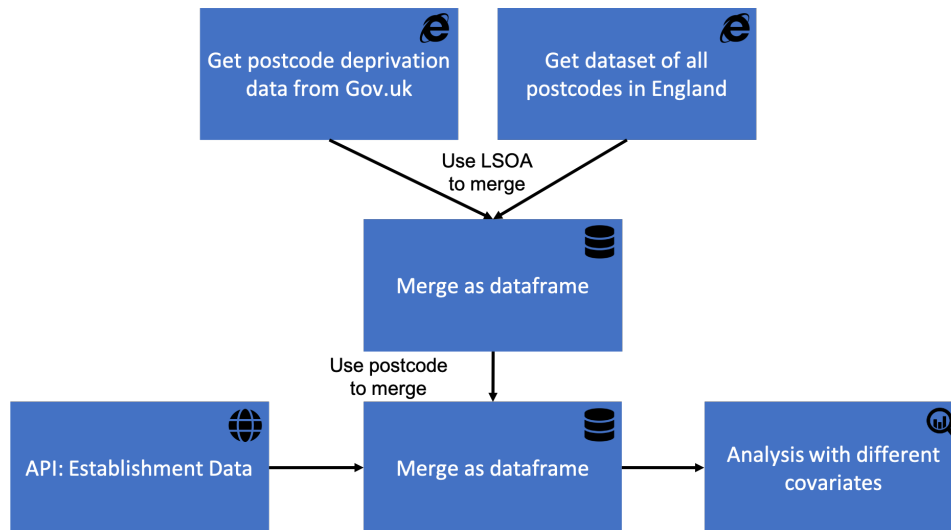| Criteria	Deprivation		
	Least	$\longrightarrow$	Most
Deprivation score	0.541	$\longrightarrow$	92.735
Deprivation rank	32,844	$\longleftarrow$	1
LSOA	Chiltern 005E	$\longrightarrow$	Tendring 018A

Table 3.2.1: Establishment ratings.

16

Figure 3.2.2: The process to merge establishment and deprivation data sets.

3.2.3 Combining postcodes and deprivation data

To obtain deprivation data for each postcode in England, firstly we needed to obtain a data set containing every postcode in England. This data set was provided by The Office for National Statistics, we used the data set published in August 2020[17] (which was updated November 2020).

In the postcode data set, one of the columns is `LSOA` so for every individual postcode in the data set we attached its corresponding deprivation data. This resulted in a data set with 2,201,092 rows and 70 columns.

3.2.4 Combining deprivation data to establishments

Once we had the data set as described in Section 3.2.3, we merged this with our establishments data set from Chapter 1. We merged the two data sets together by joining by the `postcode` column. This resulted in a data set with 382,018 rows and 85 columns. This process of obtaining deprivation data for every establishment is seen in Figure 3.2.2.

3.3 Ordinal regression models

3.3.1 Introduction

As mentioned in Section 3.1, we have created models to look at the distribution of ratings across the country. By the end of the analysis, we have built a model which includes the following covariates: deprivation data, type of establishment, local authority of establishment and an indicator variable to determine whether the establishment is a food chain.

As we have ordered outcomes, (the ratings are clearly ordered: 0, 1, 2, 3, 4 and 5) we used a method called **ordinal logistic regression**. *Logistic regression* is used when the outcome is binary, whereas ordinal logistic regression is used when we have multiple outcomes. We can use ordinal logistic regression even if we do not know how big the differences are between the outcomes, just the order of them.

3.3.2 Example

There are many different examples of where using an ordinal logistic regression model is appropriate, ranging from psychology (answers from questionnaires are ordered by construction) to business (a company sells a product in small, medium & large and wants to determine the factors which cause the customer to buy a certain size).

A familiar example is when lecturers ask students to complete a survey at the end of a course. A standard question is "*How much did you enjoy this module?*", with the options being: "*Very much*", "*Somewhat*", "*Neutral*", "*Disliked*" and "*Hated*". It is clear that these outcomes are ordered, but it is not necessarily clear that the differences between the outcomes are the same. For example, the difference between "*Somewhat*" and "*Very much*" may be larger than the difference between "*Disliked*" and "*Neutral*". Some of the covariates that the lecturer could include in this example are: gender, age, overall grade and average attendance. An ordinal logistic regression model would be fitted with these covariates to determine which (if any) are significant in explaining how much the students enjoyed the module.

3.3.3 Definition

Ordinal logistic regression is also called **ordered logit** and was first considered by McCullagh, 1980[18]. Let Y be an ordinal outcome with J categories. Then $\Pr(Y \leq j)$ is the cumulative probability of Y being less than or equal to a specific category, with $j = 1, \ldots, J - 1$. The odds of being less than or equal to a particular category can be written as:

$$\frac{\Pr(Y \leq j)}{\Pr(Y > j)},$$

for $j = 1, \ldots, J - 1$. As we have cumulative probabilities, we can write $\Pr(Y > j) = 1 - \Pr(Y \leq j)$. If we take the log of the odds fraction, this is known as the **logit**, so we have the following:

$$\log\left(\frac{\Pr(Y \leq j)}{\Pr(Y > j)}\right) \equiv \text{logit}[\Pr(Y \leq j)].$$

To use the ordinal logistic regression model in `R`, we use the `clm()`[19] function, which re-parameterises the model as:

$$\text{logit}[\Pr(Y \leq j)] = \beta_{j0} - \eta_1 x_1 - \ldots - \eta_p x_p$$

where β_{j0} is the intercept term for the j'th category of interest and η_i is the coefficient of the i'th covariate x_i.

3.3.4 Example using our data

To illustrate how an ordinal logistic regression model can be fitted to our establishment data, we have added a new column to the data. This new column consists of a binary variable which indicates whether the establishment is either one of the twenty most popular food chains in the UK[20] or one of the "big 4" supermarkets[21] (Tesco, Sainsbury's, Asda and Morrisons). These establishments are identified using regular expressions, we identify $\approx 15{,}000$ establishments to be a food/supermarket chain.

As we only have a single covariate, the model we are fitting in R is:

$$\text{logit}[\Pr(Y \leq j)] = \beta_{j0} - \eta_1 x_1,$$

where x_1 indicates whether the establishment is a food chain (1 if true). When we fit the model, we obtain the following output:

```
formula: fct_rev(rating) ~ chain
data:    EstDepMerged

 link  threshold nobs   logLik     AIC        niter max.grad cond.H
 logit flexible  338168 -302656.99 605325.98 7(0)  3.56e-09 2.6e+02


Coefficients:
          Estimate Std. Error z value Pr(>|z|)
chainTRUE  1.430      0.02842    50.3  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Threshold coefficients:
     Estimate Std. Error z value
0|1 -6.49      0.044937   -144.4
1|2 -3.89      0.012488   -311.5
2|3 -3.16      0.008850   -356.8
3|4 -1.99      0.005396   -369.3
4|5 -0.86      0.003852   -224.4
```

There are five intercept threshold coefficients. These are the β_{j0} terms and when we use the model for prediction, we only use one of them. The interpretation of these intercept terms is that they are the cumulative probability of the establishment having that rating. For example, 0|1 is the probability of receiving a 0 rating compared to a 1, 2, 3, 4 or 5 rating, and 3|4 is the probability of receiving a 0, 1, 2 or 3 rating compared to a 4 or 5 rating. We cannot use them directly, we need to include them in our predictive model to make sense of their meaning. Secondly, notice that the chain covariate has an estimate, standard error etc. Again, we cannot use this value straightaway, we need to include it in a predictive model, but we can use its sign as an indicator for how it affects the ratings. If the sign is positive then it means that when we increase the covariate the probability of obtaining a lower rating is lower, and vice versa if the sign is negative.

We will use the threshold of 3|4 in this example, which is probability of having a rating of **3 or less**. The predictive model in this case is therefore:

$$\text{logit}[\Pr(Y \leq \textbf{rating of 3})] = -1.99 - 1.43x_1,$$

Using this predictive model, we see the differences between when an establishment is identified to be a food chain ($x_1 = 1$) or not ($x_1 = 0$).

Food Chain

Here we have $x_1 = 1$ so we substitute this into our equation above and then solve for $\Pr(Y \leq \textbf{rating of 3})$.

$$\text{logit}[\Pr(Y \leq \textbf{rating of 3})] = -1.99 - 1.43x_1$$
$$\text{logit}[\Pr(Y \leq \textbf{rating of 3})] = -1.99 - 1.43 \times 1$$
$$\text{logit}[\Pr(Y \leq \textbf{rating of 3})] = -3.42$$
$$\Pr(Y \leq \textbf{rating of 3}) = \exp(-3.42)$$
$$\Pr(Y \leq \textbf{rating of 3}) = 0.033$$

Not a Food Chain

Here we have $x_1 = 0$ so we substitute this into our equation above and then solve for $\Pr(Y \leq \textbf{rating of 3})$. When we do so, we find that $\Pr(Y \leq \textbf{rating of 3}) = 0.137$.

Analysis

So we see that when the establishment **is** a food chain, the probability of having a rating of less than 3 is 0.033. When we compare this to when the food chain **is not** a food chain, the probability of having a rating of less than 3 increases to 0.137. This means that if the establishment **is not** a food chain, we are **4.15x** (0.137/0.033) more likely to have a rating of less than 3. As we have the parallel lines assumption, this relationship does not change when we move the threshold of rating. This means that for any intercept that we choose, we will always be **4.15x** more likely to be in a lower rating group if the establishment **is not** a food chain.

From Figure 3.3.1 we see why this is the case. When the establishment has been identified to be a food chain (a) then the ratings are a far higher than when the establishment has not been identified to be a food chain (b).

3.4 More complex ordinal regression models

We use the method outlined above to build more complex statistical models. We do not have to limit ourselves to one explanatory variable, just as in ordinary linear regression. This is powerful as we can see which covariates are significant in explaining the distribution of ratings and if they are significant, we see whether they have a positive or negative effect on the ratings.
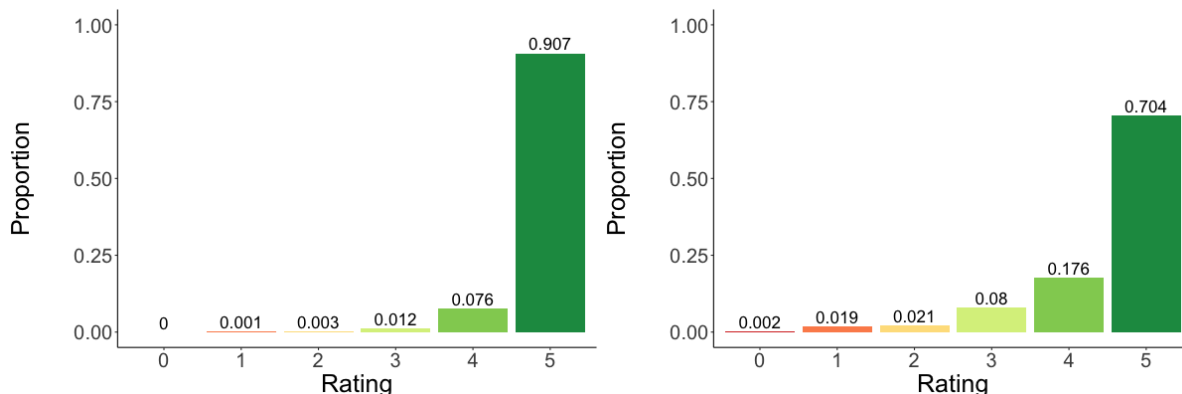
Figure 3.3.1: (a) proportion of ratings if the establishment **is** a food chain, (b) proportion of ratings if the establishment **is not** a food chain.

3.4.1 Modelling deprivation

Returning to deprivation data, we fit an ordinal logistic regression model with the **score** of the LSOA that the establishment lies in as a covariate. From Section 3.2.2 we know that the deprivation score is between 0.5 and 92.7, with the higher the score the more deprived the LSOA is. It makes more sense to use the score of the LSOA compared to the rank as the score gives a clearer indication of the actual performance of the LSOA. Therefore, the model we fit in R is:

$$\text{logit}[\Pr(Y \leq j)] = \beta_{j0} - \eta_1 x_1,$$

where x_1 is the LSOA score.

When we do so, η_1 (coefficient for LSOA score) is -0.011, as this is negative we infer that as we increase the score (more deprived) the probability of being in a lower rating group increases. In other words, being in an area of higher deprivation lowers your Food Hygiene Rating. We use the threshold of 3|4 again and we calculate that in the most deprived LSOA, compared to the least deprived LSOA, we are **2.75x** more likely to be in a lower rating group.

Again, we observe this trend visually in Figure 3.4.1. Here we see establishments in the 100 most deprived LSOAs ($n = 2,662$) and the 100 least deprived LSOAs ($n = 515$) with their proportions of ratings. There is a clear difference in the underlying distributions of the two bar charts and we see why establishments are more likely to be in a lower rating group if the establishment lies in an LSOA that has a higher deprivation score.

3.4.2 More informative models

It can sometimes be useful to only fit one covariate in the model but when we do this, we are not able to determine whether the covariate itself is explaining the variation in the response variable or whether there is some underlying causal link. For example, if the "true" reason for the difference between ratings is that establishments that are food
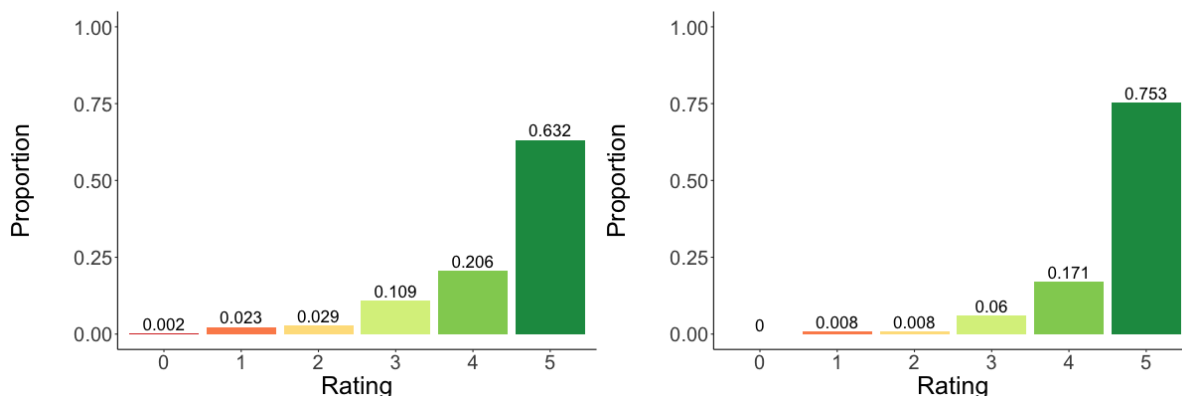
Figure 3.4.1: (a) proportion of ratings for the establishments in the 100 **most** deprived areas, (b) proportion of ratings for the establishments in the 100 **least** deprived areas.

chains are rated higher, but there happens to be double the amount of food chains in areas of lower deprivation then this underlying causal link will be carried through in the analysis of the deprivation model.

As a result of this, it is always more informative if we include multiple covariates in the model at the same time so we can determine the effect of each covariate in turn whilst "fixing" the other covariates. Therefore, we will fit a model with the deprivation score, food chain indicator and type of establishment as covariates.

Therefore, the model we fit in R is:

$$\text{logit}[\Pr(Y \leq j)] = \beta_{j0} - \eta_1 x_1 - \eta_2 x_2 - \eta_3 x_3 - \ldots \eta_{16} x_{16},$$

where x_1 is the deprivation score, x_2 is the presence of a food chain and x_i, $i = 3, ..., 16$ is the presence of the type of establishment (1 if true, 0 otherwise).

However, when we include the type of establishment covariate into the model, we set one η_i, $i > 2$ equal to zero. We do this so we are not over-parameterising the model and the model has a baseline with which to compare the other types of establishments to. In this case, we set the coefficient corresponding to "caring premises" equal to zero.

When we fit the model in R we find that all the coefficients are significant at the 99.9% level apart from the coefficients corresponding to "Farmers/growers" which is not significant at any level. We see the estimates for the coefficients for the types of establishments in Figure 3.4.2, remembering that these are compared to "caring premises" which is set to zero. The error bars ($\pm 2 \times$ Std.Error) are also plotted, we see that the types that correspond to the largest error bars are "Farmers/growers" and "Importers/exporters". These are the types that have the fewest number of establishments so that could be why we are seeing large error bars, the model has not got enough data for these types to be confident about the findings.

Again, we notice the sign and the magnitude of the estimates. They are all negative apart from "School/college/university", which indicates that, in general, the only types of establishments associated with higher ratings than caring premises are education

settings.

We can compare the estimates for "School/college/university" to "Takeaway/sandwich shop" (the type with the lowest estimate) to determine how much of a difference the type of establishment has on the ratings. When we do this, we set the deprivation score to be the median score of 17.648, we assume the establishment is not a food chain and we use the 3|4 threshold again. When we calculate this, we find that we are **9.7x** more likely to be in a lower rating group if the establishment is a takeaway compared to an education setting. This seems to be a significant finding as in this model we have accounted for other factors which may influence the ratings such as deprivation data and whether the establishment is a chain.
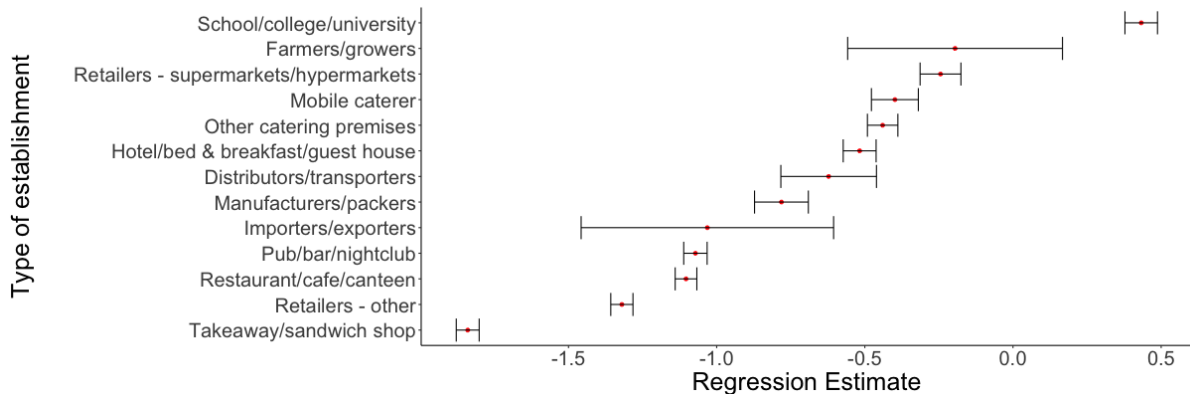

Figure 3.4.2: Regression estimates and error bars ($\pm 2 \times$ Std.Error) for the type of establishment model fitted in Section 3.4.2 (with caring premises as the baseline).

3.4.3 Local authorities

From Section 1.1 we know that inspections are carried out by EHOs employed by the local authorities. As such, we look at the effect of local authorities in the regression models to see if there are any differences between them. We would expect to see no difference between local authorities (whilst controlling for types of establishments, deprivation etc.) as the EHOs have guidance they have to follow when marking establishments[3]. However, different EHOs could interpret the guidance slightly differently or have different thresholds for what they deem to be acceptable.

Therefore, to investigate this, we look at a model containing: deprivation data, type of establishment, food chain indicator variable and local authority.

When we do so, we obtain a model with 332 covariates (5 intercept terms, 1 for deprivation score, 13 for type of establishment, 1 food chain indicator and 312 local authorities). There are 313 local authorities (different from the 409 we saw in Section 1.2.1 as we are only considering English data now) but as in the establishment types analysis, we have a baseline, which in this case is Adur, West Sussex. Similarly to the types analysis, we extract the coefficients and standard errors (which gives us error bars)

which is found in Figure 3.4.3. The points on the plot are coloured by the region that the local authority lies in.
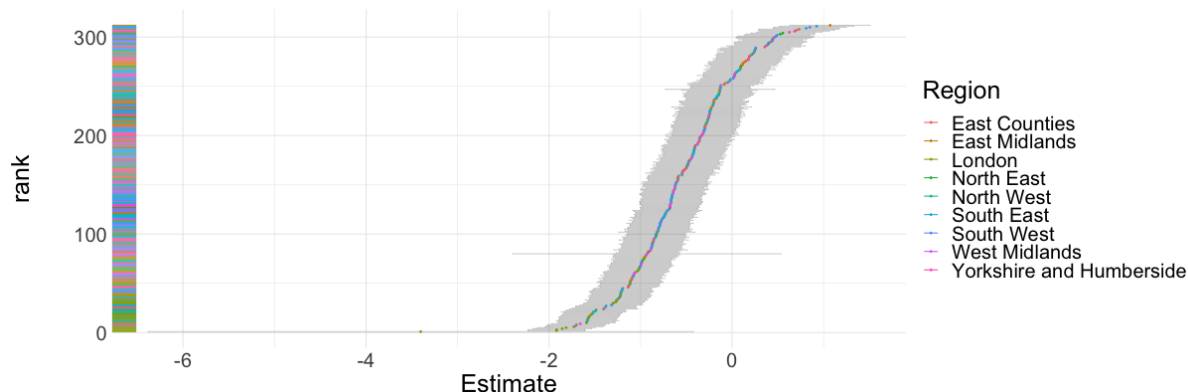

Figure 3.4.3: Regression estimates and error bars ($\pm 2\times$ Std.Error) for the local authorities of establishments model fitted in Section 3.4.3 (with Adur, West Sussex as the baseline).

The plot has the same interpretation as in Figure 3.4.2, with the more negative the estimate the more likely the local authority will have a lower rating. We see that the estimates range from about -2 to 1. This indicates that Adur (which is the baseline at 0) is in the top third of the local authorities. When we do this kind of kind of plot, we will always see some kind of difference between local authorities, so we investigate how big that difference may be.

As there will always be outliers, especially when fitting ≈ 300 terms in the model, we will look at the local authority which is ranked at 5% and compare to the local authority ranked at 95%. Barking and Dagenham, London is the local authority at 5% and has an estimate of -1.54. Richmondshire, Yorkshire is the local authority ranked at 95% and has an estimate of 0.44. We calculate the differences between these local authorities in the same way that we calculated the differences for the different types of establishments. For simplicity, we will assume that the establishment is a caring premise, it is not a food chain, deprivation data is the median score and we are using the 3|4 threshold again. When we calculate this, we find that establishments in Barking and Dagenham are **7.3x** more likely to be in a lower category of ratings than establishments in Richmondshire. Again, this seems like a hugely significant finding, as even after controlling for all other covariates there is still a very large difference in the ratings in the different local authorities.

Regional effects

In Figure 3.4.3 we see the estimates are coloured by the region that the local authority lies in. We immediately see that the estimates for London seem to be lower than other areas. To determine how much of an effect each region has on the distribution of ratings, we calculate the average estimate for each region, this is seen in Figure 3.4.4 (the error

bars seen are the average lower and upper error bars). We see that apart from London, there does not seem to be too much difference between the regions, with the other eight regions being fairly contained within the error bars. This indicates that, potentially, there are inspectors which are harsher in some local authorities than others but in general, when we average over the whole region or the whole country, these differences are not as big as initially seen. This is what we would expect, as the inspections should be consistent throughout the country, after accounting for deprivation data, types of establishments and food chains.
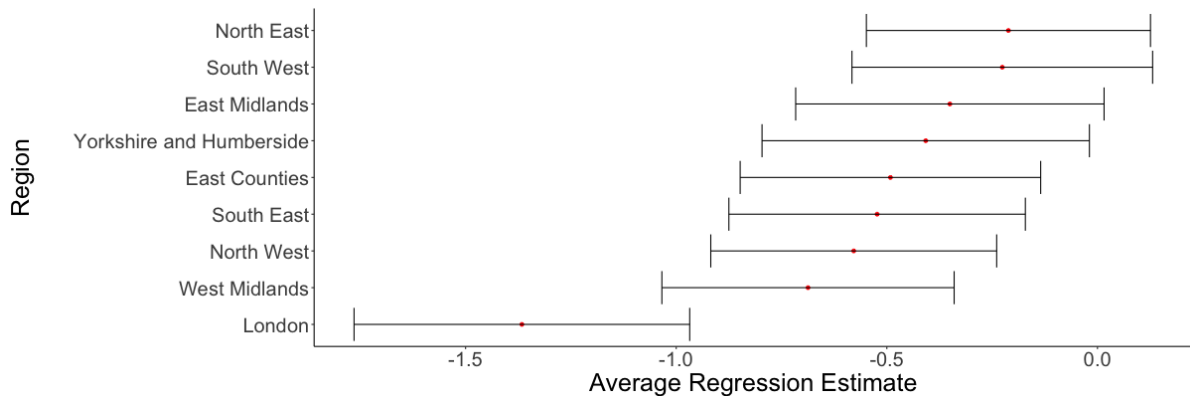

Figure 3.4.4: Average regression estimates and average error bars ($\pm 2 \times$ Std.Error) for the regions of the local authorities (from the model in Section 3.4.3).

3.5 Clustering

3.5.1 Background

As mentioned at the end of Section 2.4, we may expect clusters of similar rated establishments. This is because establishments in the same vicinity will almost definitely be inspected by the same local authority (as discussed in Section 3.4.3), the deprivation score will be very similar (if not equal), the same types of establishments may be found in the same areas (i.e. an area of care homes, an area of takeaways) and potentially, the owners might feel pressure to have similar ratings to their neighbours.

So, to determine whether clustering is present, we explored two city centres: Sheffield and Newcastle upon Tyne. City centres are good areas to look for clustering as there are many establishments in a small area so our findings should be more robust than say, exploring a 20-mile radius in Northumberland as we would be including different levels of deprivation etc.

3.5.2 Sheffield

First, we considered Sheffield city centre. To do so we needed to identify establishments in this area, we did so by filtering all the establishments that are in the S1 postcode

district. We found that there are 396 establishments here with a rating and a longitude/latitude.

The method of clustering we have implemented is to consider each establishment in turn, find all other establishments within 500 metres of it (not including itself) and then calculate the mean rating and mean raw overall score of these neighbouring establishments. Therefore, each establishment in the data set has two new columns: **mean rating of establishments within 500m** and **mean raw overall score of establishments within 500m**.

From here, we grouped the establishments by their own raw overall score and then calculated the mean rating of the mean neighbouring ratings for each group. This is seen in Figure 3.5.1.
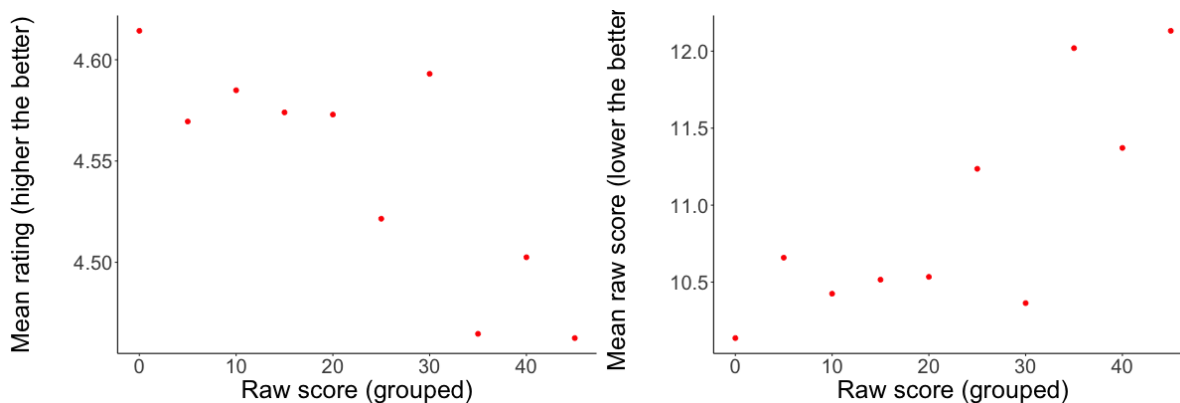

Figure 3.5.1: (a) Raw overall score (which is grouped) vs the mean rating for the neighbouring establishments, (b) Raw overall score (which is grouped) vs the mean raw overall score for the neighbouring establishments.

In Figure 3.5.1(a) we see a plot with the raw overall scores of the establishments on the x-axis, the groups of raw overall scores are from 0 to 45. On the y-axis we see the corresponding mean rating for all of the establishments with that raw overall score. For example, there are 86 establishments in the S1 postcode district with raw overall score 0, and for each of these 86 establishments, we found the mean rating of the establishments within 500m. Then, we obtain a mean of all these mean ratings which is 4.61, which is the left-most point seen in the plot. We repeat this process for each of the raw overall scores to obtain the plot in (a). We see that there seems to be a weak but observable negative trend in the plot. This indicates that as the raw overall scores for the establishments go up (they score worse) the ratings of their neighbouring establishments also becomes worse (as they go down). This is a sign that there appears to be clusters of establishments that obtain roughly similar scores. The plot in (b) is very similar, but instead of the mean ratings on the y-axis we now have the mean raw overall score. The trend is now a weak positive one (because as the raw overall score increases the worse the establishment is) so this is very much in line with what we saw in (a), that there does appear to be clusters of similar rated establishments nearby to each other.

3.5.3 Newcastle upon Tyne

We implemented exactly the same method above, but for Newcastle upon Tyne, more specifically the NE1 postcode, which again, is the city centre. Here, we have $n = 601$, the results are seen in Figure 3.5.2. We see a very weak association between ratings of establishments and neighbouring ratings.
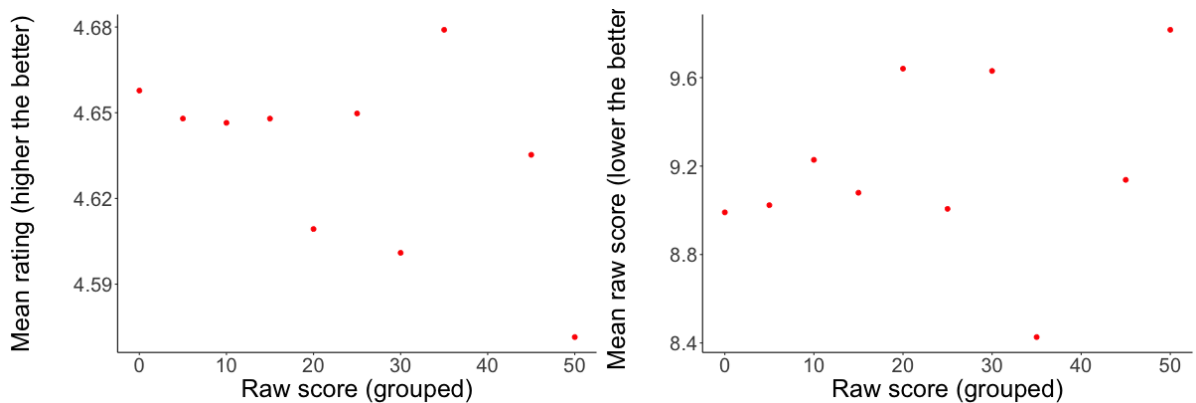

Figure 3.5.2: (a) Raw overall score (which is grouped) vs the mean rating for the neighbouring establishments, (b) Raw overall score (which is grouped) vs the mean raw overall score for the neighbouring establishments.

3.5.4 Findings

So, from our analysis of clustering in the two city centres we can say that the ratings of establishments does seem to be somewhat correlated with ratings of the neighbouring establishments, even if the correlation is weak in some cases. There could be a number of reasons for this as mentioned in Section 3.5.1, but an additional variable is to look at food chains again. We also know being a food chain has a big impact on the rating of an establishment and we would expect a city centre to have a higher proportion of food chains than normal which is indeed the case. In England, 4.4% of establishments are identified as being a food chain, compared to 7.8% in the S1 postcode district and 9.3% in the NE1 postcode district.

It is hard to determine the main factor when looking at ratings of neighbouring establishments, whether it is the establishment owners themselves "settling" for a rating or whether simply the type of establishment plays such a huge role that it is taken out of their hands. It would be informative to send a questionnaire to establishment owners to see how they feel about neighbouring establishments and whether the neighbouring ratings have any impact on the rating that they obtain.

3.6 Time series

3.6.1 Introduction

Since the start of data collection, 2020-10-19, we have been downloading the data set on a weekly basis to determine whether there have been any differences in the establishments over the course of the project. Unfortunately, during the project there were two national "lockdowns" due to the spread of COVID-19. This meant that the majority of restaurants/takeaways were forced to temporarily close so there is only limited analysis we can do with these data but here are some of our findings.

3.6.2 Types of establishments

We looked at the different types of establishments that are in the data set at different points in time. This be seen in Figure 3.6.1. From the plot we see that the number of establishments stayed very consistent throughout, the only type of establishment that had a clear increase is "Other catering premises". There are 23,367 more establishments in the final data set of our analysis (2021-04-19) than the first downloaded data set. Only two types have fewer establishments at the end - Pub/bar/nightclub and School/college/university.
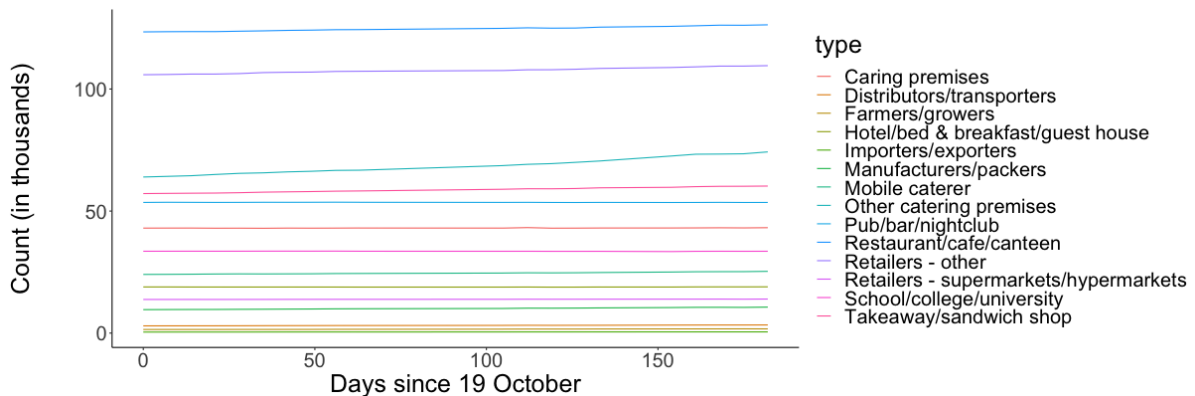

Figure 3.6.1: Counts of types of establishments throughout the project.

3.6.3 Ratings of establishments

We also looked at the ratings of the establishments at different points in time, this is seen in Figure 3.6.2. Again, there are no noticeable differences in ratings over the course of the project.

Table 3.6.1 shows counts of ratings at the start of the project (2020-10-19) - and the end of the project (2021-04-19). These figures are from England, Northern Ireland & Wales and only include the establishments with numeric ratings. These findings are
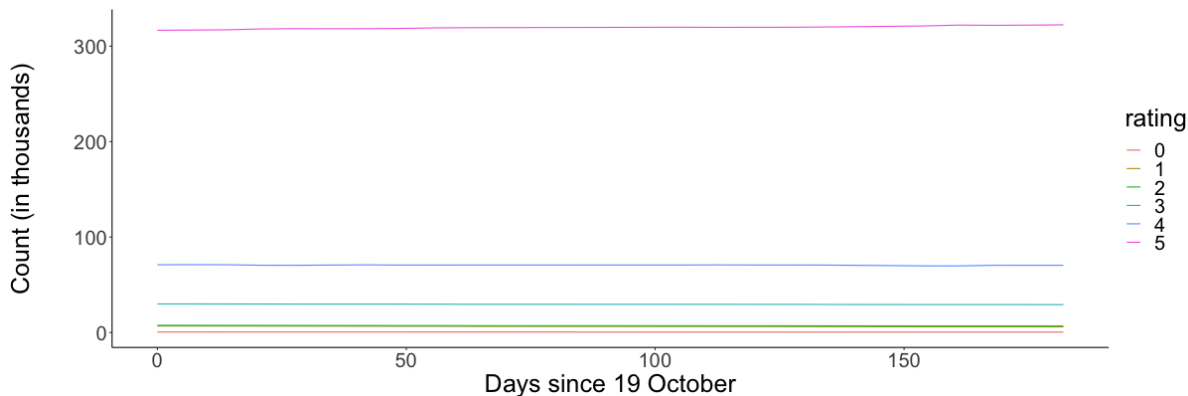
Figure 3.6.2: Counts of ratings of establishments throughout the project.

	Number of Establishments		
Rating	2020-10-19	2021-04-19	Difference
0	534	472	-62
1	6,856	5,972	-884
2	7,549	6,817	-732
3	29,919	29,200	-719
4	71,051	70,406	-645
5	316,704	322,369	5,665

Table 3.6.1: Counts of establishments at the start and end of project.

interesting as we see that the only rating with more establishments at the end is 5. There are two potential reasons for this:

- Establishments with lower ratings at the start of the project have all got substantially better, thus leading to decreases in lower ratings and an increase in the number of establishments with a rating of 5; or

- Establishments with lower ratings have closed, either due to the pandemic or because they have been forced to close by the local authority because of their low rating

The second reason seems more plausible as, in general, we would expect establishments with a higher rating to have more income so they would be more likely to be able to keep themselves afloat during a pandemic, compared to establishments who are rated lower that may not be able to survive after being closed for months. One more thing to note is that at the start of the project 39,383 establishments were "awaiting inspection", this increases to 57,914 at the end of the project. We know that establishments which score lower are inspected more often, so the majority of these establishments that are awaiting inspection may be rated lower.

4 Shiny App

4.1 Introduction

{`Shiny`}[22] is an `R` package that provides an elegant and powerful web framework for building web applications using `R`. A Shiny app usually requires some input from the user and presents output in an accessible way. We have developed a Shiny app in which the user can browse a map of England, broken down by postcode districts, and then see different summary statistics for the chosen postcode district. There is also an option to view all the establishments in a postcode district, by rating. The application can be accessed at

<div align="center">

`https://jamesalsbury.shinyapps.io/interactiveshinymap/`

</div>

The main principle behind a Shiny App is reactivity: when the user changes an input, such as a drop-down option from a list, this signals to the Shiny server that a change has been made. This triggers a chain-reaction of reactive events, where only the code associated with the changed input needs to be re-run and the output is updated accordingly.

4.2 Shiny example

When developing a Shiny app, there are two core components: the `UI` (user interface) and the `server`. The *UI* is the layout of the app, where the input/output boxes/labels/plots are found and the *server* is the underlying code manipulating the inputs to give the desired output.

We already know a very familiar example of reactivity: Microsoft Excel works in exactly this way. We have inputs (the cell values), the server (the underlying formulas associated with cells) and outputs (the cell values). We know that whenever we change the value of a cell, the underlying code recognises this and re-runs the formulas to give different outputs.

However, it is also useful to consider a Shiny example to understand how Shiny combines the UI and server to create a functioning app. In this example we will assume we have two data sets: data about the twenty English Premier League teams for the 2019-20 season and data about the twenty Spanish La Liga teams for the 2019-20 season. We have the same data for both data sets: position in the table after x match-days, top goalscorers, number of goals scored etc. We wish to compare the data from one team from the Premier League with the data from one team from La Liga. To do so, we

require two inputs from the user: the Premier League team and the La Liga team. To obtain these inputs, we have two drop-down lists in the UI which has the twenty teams from the league.

For simplicity, we will just assume the only output is two plots: position in the table vs matchday number for each team. We see an example of one of these plots in Figure 4.2.1, this shows Newcastle United's position in the table after x matchdays in the 2019-20 season.
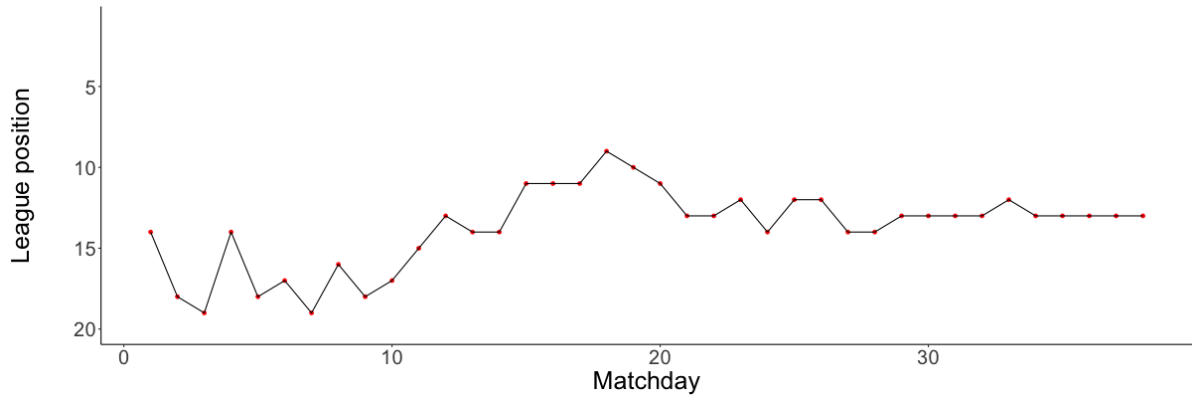

Figure 4.2.1: Newcastle United 2019-20 season, league position vs matchday.

Therefore in the server we have two functions, one for each plot. These functions will take the team as the input, filter the corresponding data set by this team and then render a plot as seen in Figure 4.2.1. This plot will then be passed to the UI which will display it accordingly. This whole process is seen schematically in Figure 4.2.2.
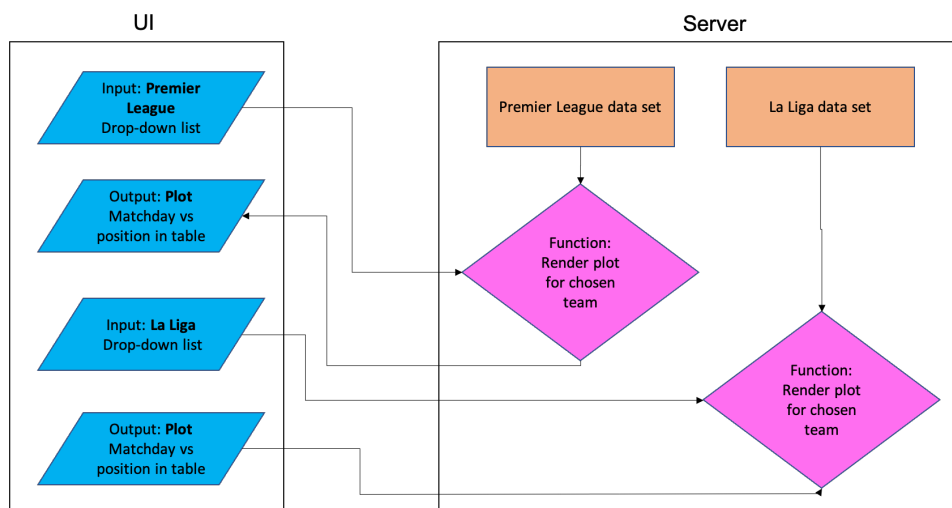

Figure 4.2.2: Shiny football example flowchart.

From Figure 4.2.2 we see that whenever either of the input changes there is a chain reaction, leading to a different plot being generated by the server. However, we also

notice that the two chain-reactions are independent from each other, therefore if only one of the inputs is changed, only one of the plots is re-rendered. This feature is called **reactivity** and underlines the usefulness and flexible nature of Shiny apps, the app works seamlessly with whatever the user inputs.

4.3 Food Hygiene Ratings app

As discussed in the example above, the main feature of a Shiny app is the fact it is reactive: the app keeps track of any input updates from the user, runs the corresponding code and then displays the output accordingly. As such, it is important to consider all the avenues that the user could go down and write the appropriate code for every eventuality. Therefore, a flowchart, which is seen in Figure 4.3.1, is useful in this situation, to make clear what should be available for the user to click on and what the user should see after they do click on an area or change any inputs.
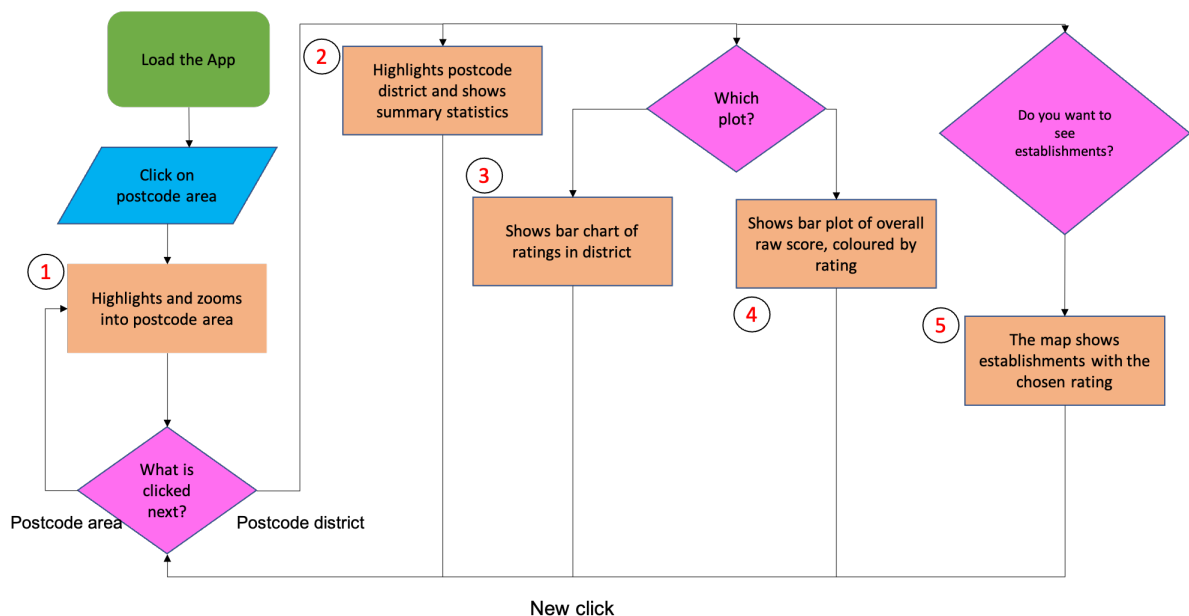

Figure 4.3.1: Flowchart for the Shiny App.

The flowchart is numbered 1-5, these numbers correspond with different input/output that the app handles, we will walk through each number in turn. At each number, the app shows the appropriate data on the map and then limits what the user can/cannot click next.

4.3.1 Flowchart component 1

What does the user see?

From the flowchart in Figure 4.3.1, we see that the first input, and the only input possible at this stage, is the postcode area. The user moves their cursor around the map and when they hover over a postcode area, the area code (e.g. NE, SW) is shown where their cursor is so they know what area they are hovering over before they click on it.

When the user clicks on a postcode area, the map zooms and centres into that area. It breaks the chosen postcode area down into its districts, which are all now clickable. It also highlights the chosen postcode area in red. The above process is shown in Figure 4.3.2, where the chosen postcode area is "NE".
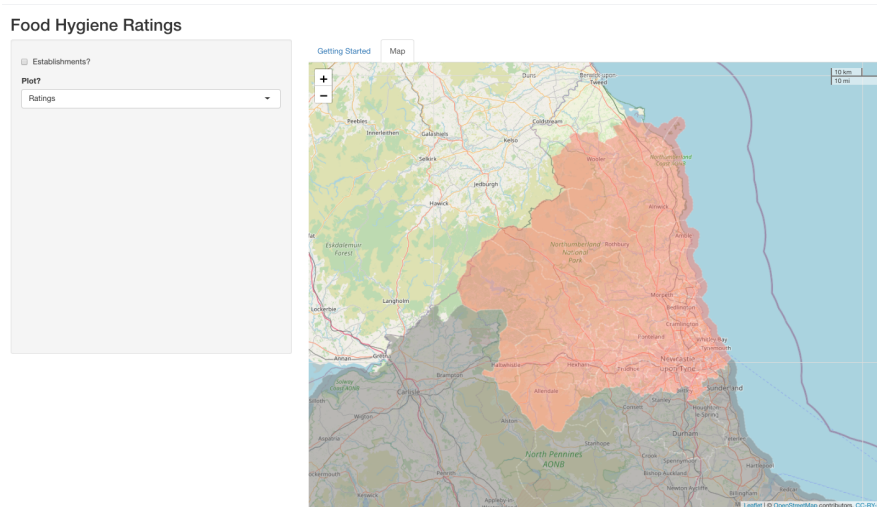

Figure 4.3.2: Shiny app when the "NE" postcode area is clicked (number 1 in flowchart).

What is the underlying code?

When the map is loaded for the first time, the postcode areas are added as layers on top of the map. There are 104 postcode areas, and each area has a `LayerID` corresponding to the alphabetical number of the postcode area (e.g. AL is 1, NE is 57). As a result of this, the app knows which polygon corresponds which postcode area. So when the user hovers over a postcode area, the app shows the area code (e.g. NE) and when the user clicks on a postcode area, the `LayerID` input is passed to a function which determines where on the map (longitude/latitude coordinates) the app should zoom into.

When the postcode area has been clicked on, the postcode districts are then added as layers, in the same way that postcode areas were added as layers when the app loaded for the first time. Again, each postcode district now has a `LayerID`, in this case, the `LayerID` is simply the name of the postcode district, for example, "NE2".

As the user is able to either click on another postcode area or a postcode district here, we use the `LayerID` to determine which one was clicked. If the `LayerID` that is clicked is an integer (e.g. 1, 57, from before) then we know we need to load the appropriate corresponding postcode area. If this option occurs, the app returns to number 1 in the flowchart and loads the new postcode area in exactly the same way. However, if the `LayerID` chosen is a string - usually in the format `LLNN` - then we know the user would like to view a postcode district. When this option is chosen, numbers 2, 3, 4 and 5 in the flowchart in Figure 4.3.1 all occur simultaneously.

4.3.2 Flowchart component 2

What does the user see?

Number 2 in the flowchart in Figure 4.3.1 occurs when the user clicks on a postcode district inside the chosen postcode area in number 1. In Figure 4.3.3 we see how the app looks when we click on the postcode district "NE2". The chosen postcode district is now highlighted in blue, with the surrounding postcode area still highlighted in red. Summary statistics are seen in a table in a panel on the left-hand side of the app. The summary statistics that are shown are the mean Food Hygiene Rating for: England, the postcode area in which the postcode district lies and the postcode district itself. It also shows the same for the raw overall scores.
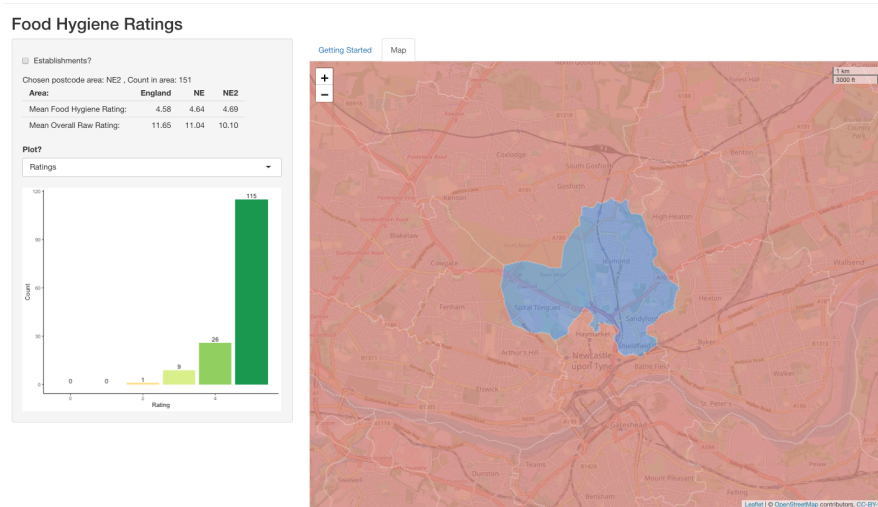

Figure 4.3.3: Shiny app when the "NE2" postcode district is clicked (numbers 2 and 3 in flowchart).

From the panel on the left, we see that the chosen postcode district is NE2, there are 151 establishments in the district and the summary statistics are shown in a table underneath. From looking at the mean Food Hygiene Rating in the summary statistics, we see that NE has a higher average rating than England, and NE2 specifically has a

higher average than both the postcode area NE and England as a whole. This is reflected in the raw overall scores (where lower values are better).

What is the underlying code?

As explained above, the app knows that a postcode district has been clicked on, rather than another postcode area as the click input will be a string rather than an integer. When this occurs, the app performs numbers 2, 3, 4 and 5 in the flowchart simultaneously. We will focus on the mechanism of number 2 here.

So, as the click on the map is a postcode district area, the app uses the `LayerID` to change the colour of the chosen postcode district. This is seen in Figure 4.3.3 as the chosen postcode district is now blue.

As described above, summary statistics are also shown as text and in a table. This is done by using the `LayerID` (here, "NE2") to filter the establishment data set (so we are left with establishments only in "NE2") and then calculating the mean rating and mean raw overall score. We use regular expressions[13] to extract the `LL` part of the `LayerID` (which will be "NE" in this case), filter the establishment data set again and then calculate mean rating and mean raw overall score. For the first column, England, we can just take a mean of all of the establishments in the data set.

4.3.3 Flowchart component 3

What does the user see?

Also in Figure 4.3.3, we see one of the two possible plots underneath the summary statistics. This plot shows a bar chart of the ratings found in NE2. The total number of establishments (adding up the size of all the individual bars) is the same as the number of establishments found above the summary statistics. If you were to use the bar chart to calculate the mean rating in the postcode district then this would also give the same result as in the summary statistics table.

What is the underlying code?

This is done in a very similar way to number 2; the `server` filters the establishment data set by `LayerID`, groups by the ratings and then takes the count of each group. From this data, a bar chart is produced.

4.3.4 Flowchart component 4

What does the user see?

The other plot that the user is able to view is seen on the left-hand panel in Figure 4.3.4. This is a bar chart of the raw overall score for the establishments in NE2, filled by ratings.
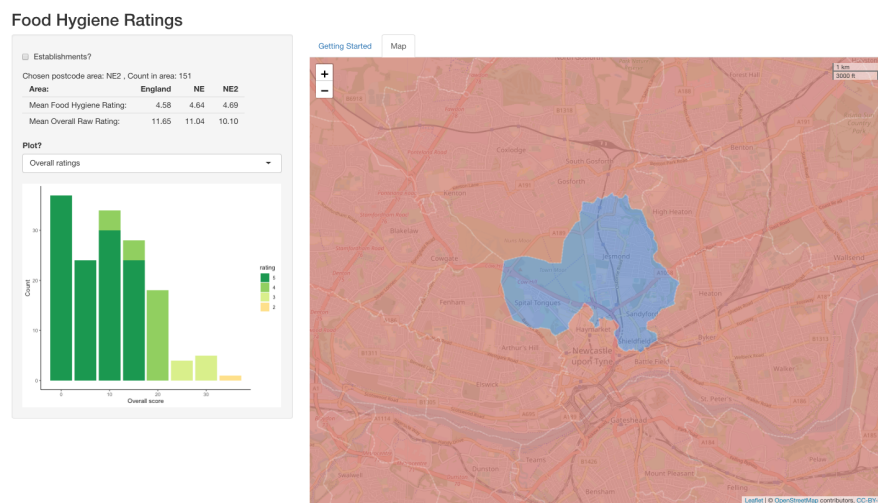
Figure 4.3.4: Shiny app when the "NE2" postcode district is clicked - raw overall plot (number 4 in flowchart).

What is the underlying code?

The bar plot is produced in a very similar way to the bar plot seen in number 3. Again, the `LayerID` is used to filter the establishment data set and then the remaining establishments are grouped by their raw overall scores. These groups are plotted on a bar chart, whilst being coloured by their rating. We see that the colours of the different ratings are the same colours found in the bar plot in number 3.

However, alongside the reactive nature of the map (a postcode area or district can be clicked on at any time), there is another layer of reactivity associated with numbers 3 and 4 in the flowchart. This reactivity is the choice of which bar plot to view, by default the user will be viewing number 3. The choice is a drop-down list consisting of "Ratings" and "Overall ratings". The user is able to choose to view either plot at any time so the app needs to keep track of which plot is being shown and if the other plot is chosen, then to run the appropriate code for the other plot and display the graphic.

4.3.5 Flowchart component 5

What does the user see?

The final event to occur when a postcode district is clicked is the option to view the establishments in the postcode district, this is seen in Figure 4.3.5. We see that at the top of the left-hand panel, there is a checkbox next to "Establishments?", when this is checked a further drop-down list of checkboxes are shown, with ratings 5 to 0. When any of these individual rating checkboxes are checked the app will show the establishments with the corresponding rating on the map.

The establishments are shown as coloured circles, with the colours corresponding to
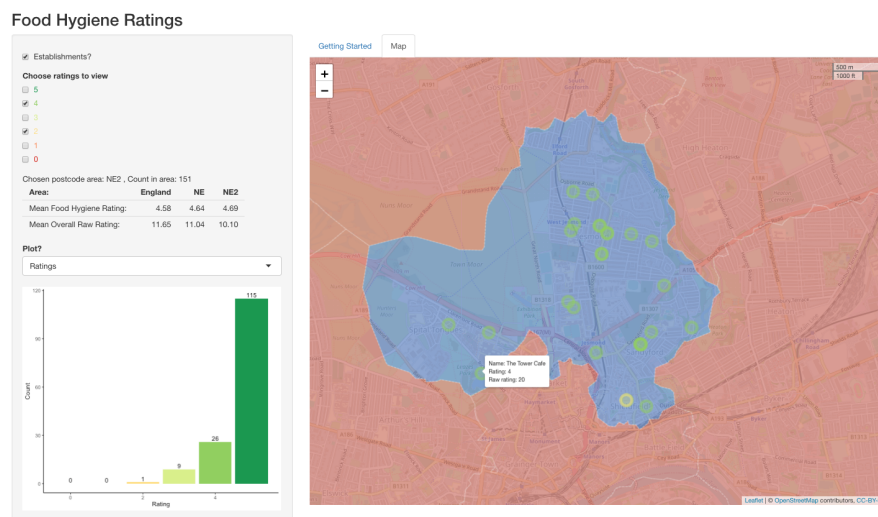
Figure 4.3.5: Shiny app when "NE2" is chosen and establishments are viewed (number 5 in flowchart).

the colours found in the text of the checkboxes and the colours of both the available bar charts. For example, we see in Figure 4.3.5 that there are three checkboxes checked; the first "Establishments" checkbox and then ratings 4 and 2. On the map we see the 26 establishments with rating 4 and the 1 establishment with rating 2 (same values as seen in the bar chart). When the user hovers their cursor over an establishment shown on the map, the map will show the name, the rating and the raw overall score of the establishment. This is highlighted in Figure 4.3.5 as we see in the South-East corner of NE2 there is "The Tower Cafe", which has a rating of 4 and a raw overall score of 20. It is coloured in green as this is the colour that corresponds to a rating of 4.

What is the underlying code?

In this section, we have a few layers of reactivity. The first layer is the "Establishments?" checkbox, when this is unchecked then the user is not able to view any establishments on the map or able to check/see any of the drop-down ratings. However, when the "Establishments?" checkbox is checked, the app needs to show the drop-down checkboxes and then look to see whether any of those are checked. If they are checked then the corresponding establishments need to be shown on the map. This process is seen in Figure 4.3.6.

When the "Establishments?" checkbox is checked, a string containing the numerical values of the ratings is passed to a function. This function considers each value of the string in turn and then displays the establishments corresponding to this rating as required. Every time the string containing the chosen ratings changes, the function is called and the "new" string of chosen ratings is considered.

The labels are added to the map in the same way as the postcode area labels were added in number 1. When the establishments are added to the map, the name, rating
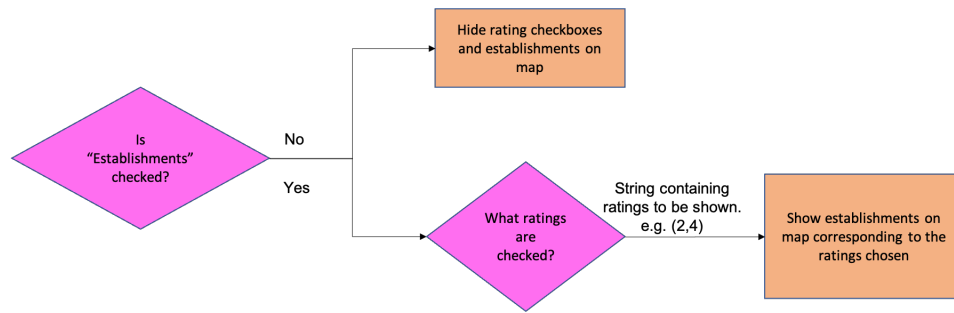
Figure 4.3.6: Reactive process for number 5 - adding establishments to map.

and raw overall score are taken from that establishment's row in the data set and paired up to the corresponding circle on the map.

5 Conclusion and Further Work

5.1 Conclusion

The aim of this project was to determine whether Food Hygiene Ratings are randomly distributed in the UK or whether there are underlying factors that influence the ratings of establishments.

In Chapter 2 we were able to see visually that the ratings are not randomly distributed and that higher ratings were more generally seen in rural areas and lower ratings were more generally seen in urban areas.

In Chapter 3 we were able to build models which explained the regional differences in the ratings by including different covariates. The covariates of interest were: deprivation data, type of establishment, a food chain indicator variable and the local authority of the establishment. We found that all of the covariates were significant when explaining ratings, even when we account for all other covariates. The effects the covariates had on the ratings are: establishments in areas of higher deprivation were more likely to be rated lower and there are certain types of establishments which are associated with lower ratings. For example, takeaways and sandwich shops were nearly 10x more likely to have a lower rating than education premises. Establishments that were identified as being a food chain were more likely to be have higher ratings and different local authorities are associated with higher/lower ratings, although when we average these estimates within the regions, we find that the differences may not be as significant as they initially seem (with the exception of London).

In Chapter 4 we were able to use all the data downloaded from an API provided by the *Food Standards Agency* to build a Shiny App. This app allows users to explore for themselves the distribution of ratings across the country. It has several functions including allowing users to view summary statistics for all postcode districts in England and to view establishments in postcode districts coloured by their rating.

5.2 Further work

There is more work that has been unexplored in this project, some of the possibilities are given below.

Other covariates

We have found two external sources of covariates; whether the establishment is a food chain and deprivation data. However, there are other covariates that could be consid-

ered for these data such as whether the establishment is on third party home delivery service (e.g. Just Eat, Deliveroo). The hypothesis would be that establishments that are considered to be on these kind of websites would be rated higher, as the home delivery services would have their own criteria when considering establishments to be on their website. For example, the Deliveroo website outlines that any establishment must have a rating of at least 2 to be considered to be on their platform [23].

Another covariate could be mean time between inspections. We know that establishments that obtain lower ratings are inspected more often, to try and increase their rating and to ensure the establishment is not posing a threat to the public. If establishments obtain a rating of 0 or 1 then local authorities will re-inspect them sooner than establishments which score 4 or 5. Therefore, in general, the longer between inspections, the higher we would expect the establishment to be rated. The actual time may vary in each local authority so we could fit an interaction term between local authority and time between inspections to make the model more accurate.

Predict future establishments

Another potential avenue of investigation would be to build a model to predict the ratings of future establishments. This could be done with our data using machine learning techniques. We could split the data in half, using one half of the data as a *training set* and the other half as a *test set*. The model could be fine-tuned with different parameters, some of which we have already considered such as deprivation data, local authority and type of establishment. This model could potentially be useful for the Food Standards Agency as they could analyse the distribution of establishments in a certain area and if the ratings differ hugely from the predictive ratings, they could investigate further to determine the cause of these differences.

Increase functionality with the Shiny App

The Shiny App is useful for looking at the distribution of ratings and individual establishments in local areas, but there is extra functionality that could be added into the app. For example, there could be another tab which acts as more of a database, allowing the user to search by different variables, filtering by: postcode, rating, type, local authority etc.

Continue with longitudinal study

As we only have data from 6 months it is hard to see long-term trends, especially in 2020/21 as a lot of businesses would have been hugely affected by COVID-19. It would be good to continue downloading the data set weekly over a number of years to gain deeper insights. For example, we could see whether establishments that have recently opened have higher ratings than their type/local authority would suggest or look at other factors, such as seasonal effects: e.g. do more establishments open in the summer? How many bad ratings does an establishment have before being closed down? Exactly

how much more often are lower rated establishments inspected compared to higher rated ones?

Bibliography

[1] Food Standards Agency: Food Hygiene Rating Scheme. `https://www.food.gov.uk/safety-hygiene/food-hygiene-rating-scheme`. Accessed: 7-Oct-2020.

[2] Food Standards Scotland: About the Food Hygiene Information Scheme. `https://www.foodstandards.gov.scot/consumers/food-safety/buying-food-eating-out/food-hygiene-information-scheme/about-the-food-hygiene-information-scheme`. Accessed: 7-Oct-2020.

[3] Food Standards Agency: The Food Hygiene Rating Scheme - Guidance for local authorities. `https://www.food.gov.uk/sites/default/files/media/document/The%20Food%20Hygiene%20Rating%20Scheme%20Guidance%20for%20Local%20Authorities%20on%20implementation%20and%20operation%20-%20Brand%20Standrad_2.pdf`. Accessed: 8-Oct-2020.

[4] Food Standards Agency: Display of food hygiene ratings. `https://www.food.gov.uk/research/research-projects/display-of-food-hygiene-ratings-in-england-wales-and-northern-ireland-2017-wave-of-research`. Accessed: 8-Oct-2020.

[5] Food Standards Agency: Food Hygiene Rating Scheme API. `https://api.ratings.food.gov.uk/Help`. Accessed: 12-Oct-2020.

[6] API Friends: What is an API? `https://apifriends.com/api-management/what-is-an-api`. Accessed: 13-Oct-2020.

[7] Ideal postcodes: The UK Postcode Format. `https://ideal-postcodes.co.uk/guides/uk-postcode-format`. Accessed: 26-Oct-2020.

[8] Nomis: 2011 Census Data. `https://www.nomisweb.co.uk/census/2011`. Accessed: 26-Oct-2020.

[9] Edzer Pebesma. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439–446, 2018.

[10] Roger S. Bivand, Edzer Pebesma, and Virgilio Gomez-Rubio. *Applied spatial data analysis with R, Second edition*. Springer, NY, 2013.

[11] GitHub: UK Postcode Polygons. `https://github.com/missinglink/uk-postcode-polygons/tree/master/geojson`. Accessed: 2-Nov-2020.

[12] Joe Cheng, Bhaskar Karambelkar, and Yihui Xie. *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*, 2019. R package version 2.0.3.

[13] Hadley Wickham. *stringr: Simple, Consistent Wrappers for Common String Operations*, 2019. R package version 1.4.0.

[14] ArcGIS: Choropleth Map. `https://www.arcgis.com/apps/MapJournal/index.html?appid=75eff041036d40cf8e70df99641004ca`. Accessed: 3-Nov-2020.

[15] GOV.UK: Index of multiple deprivation. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/854569/Deprivation_2019.pdf`. Accessed: 3-Nov-2020.

[16] GOV.UK: English indices of deprivation 2019. `https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019`. Accessed: 9-Nov-2020.

[17] Geoportal: Postcode to Lower Layer Super Output Area Lookup in the UK. `https://geoportal.statistics.gov.uk/datasets/postcode-to-output-area-to-lower-layer-super-output-area-to-middle-layer-super-output-area-to-local-authority-district-august-2020-lookup-in-the-uk`. Accessed: 9-Nov-2020.

[18] Peter McCullagh. Regression models for ordinal regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):109–142, 1980.

[19] R. H. B. Christensen. ordinal—regression models for ordinal data, 2019. R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal.

[20] YouGov: The Most Popular Dining Brands (Q4 2020). `https://yougov.co.uk/ratings/food/popularity/dining-brands/all`. Accessed: 16-Nov-2020.

[21] Statista: Market share of grocery stores in Great Britain from January 2017 to December 2020. `https://www.statista.com/statistics/280208/grocery-market-share-in-the-united-kingdom-uk`. Accessed: 16-Nov-2020.

[22] Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. *shiny: Web Application Framework for R*, 2021. R package version 1.6.0.

[23] Deliveroo: Deliveroo food hygiene rating policy. `https://help.deliveroo.com/en/articles/3198655-deliveroo-food-hygiene-rating-policy`. Accessed: 19-Apr-2021.