Scores on the Doors: An Analysis of Food Hygiene Ratings across the UK

Author: James Salsbury, Supervisor: Dr. Colin Gillespie
School of Mathematics, Statistics & Physics, Newcastle University

Motivation

Aim: Are Food Hygiene Ratings randomly scattered across the country, and if not, can we find any factors that explain the distribution of ratings?

Introduction: Food Hygiene Ratings

- Every UK food establishment is required to have a food hygiene inspection[1]
- The inspectors give marks (the more marks the worse) for:
  - **Hygiene**: cleanliness, storage of food, ...
  - **Structural**: layout of the kitchen, ventilation, ...
  - **Management**: paperwork, training, ...
- Marks are combined to give an **overall score**

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: Overall Scores mapped to Food Hygiene Ratings.

The Data

- Data was downloaded from the Food Standards Agency API
- Only considered establishments in England
- There are ~ 400,000 establishments with ratings

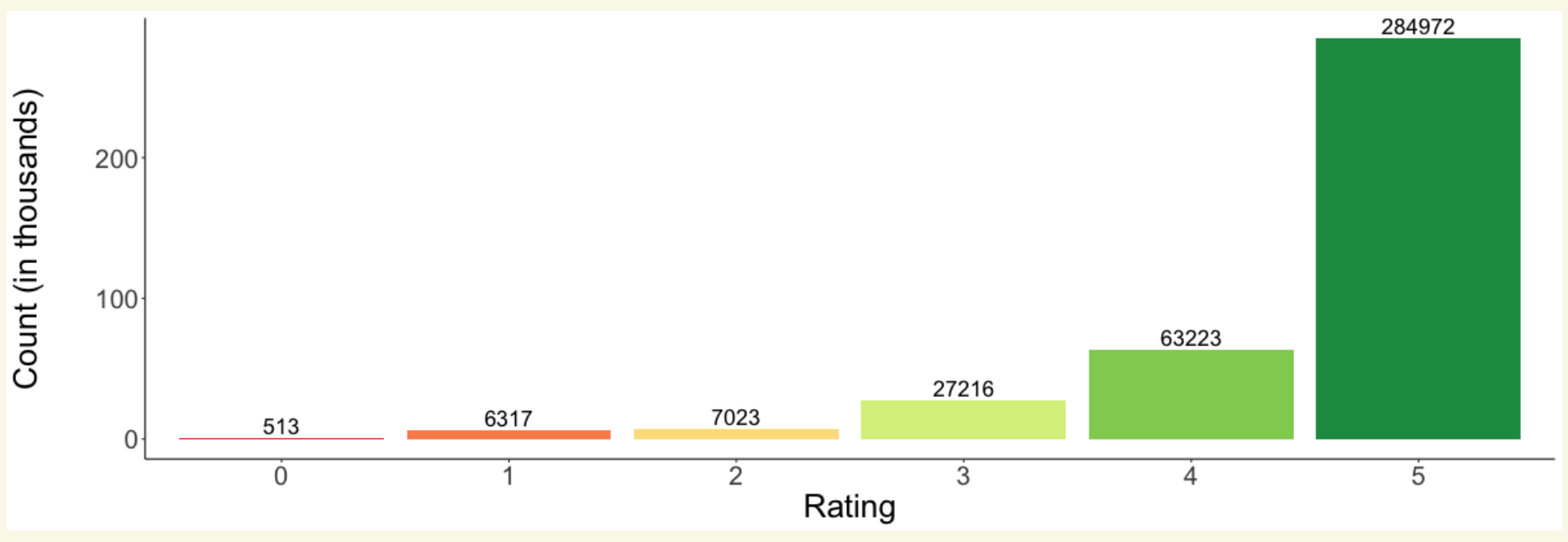


Figure 1: Ratings of the establishments.

- Fourteen types of establishment - Table 2

Type	Count	Type	Count
Restaurant/cafe/canteen	102,600	Retailers - other	87,800
Other catering premises	53,600	Takeaway/sandwich shop	47,600
Pub/bar/nightclub	44,800	Caring premises	34,600
School/college/university	27,100	Mobile caterer	20,100
Hotel/bed & breakfast	13,700	Retailers - supermarkets	11,500
Manufacturers/packers	6,200	Distributors/transporters	2,000
Farmers/growers	800	Importers/exporters	400

Table 2: Different types of establishments (to nearest 100).

Visualising the Data

- The establishments were grouped by their postcode district then a mean Food Hygiene Rating was calculated for each district
- Postcode districts were plotted onto a choropleth map to show trends/variation

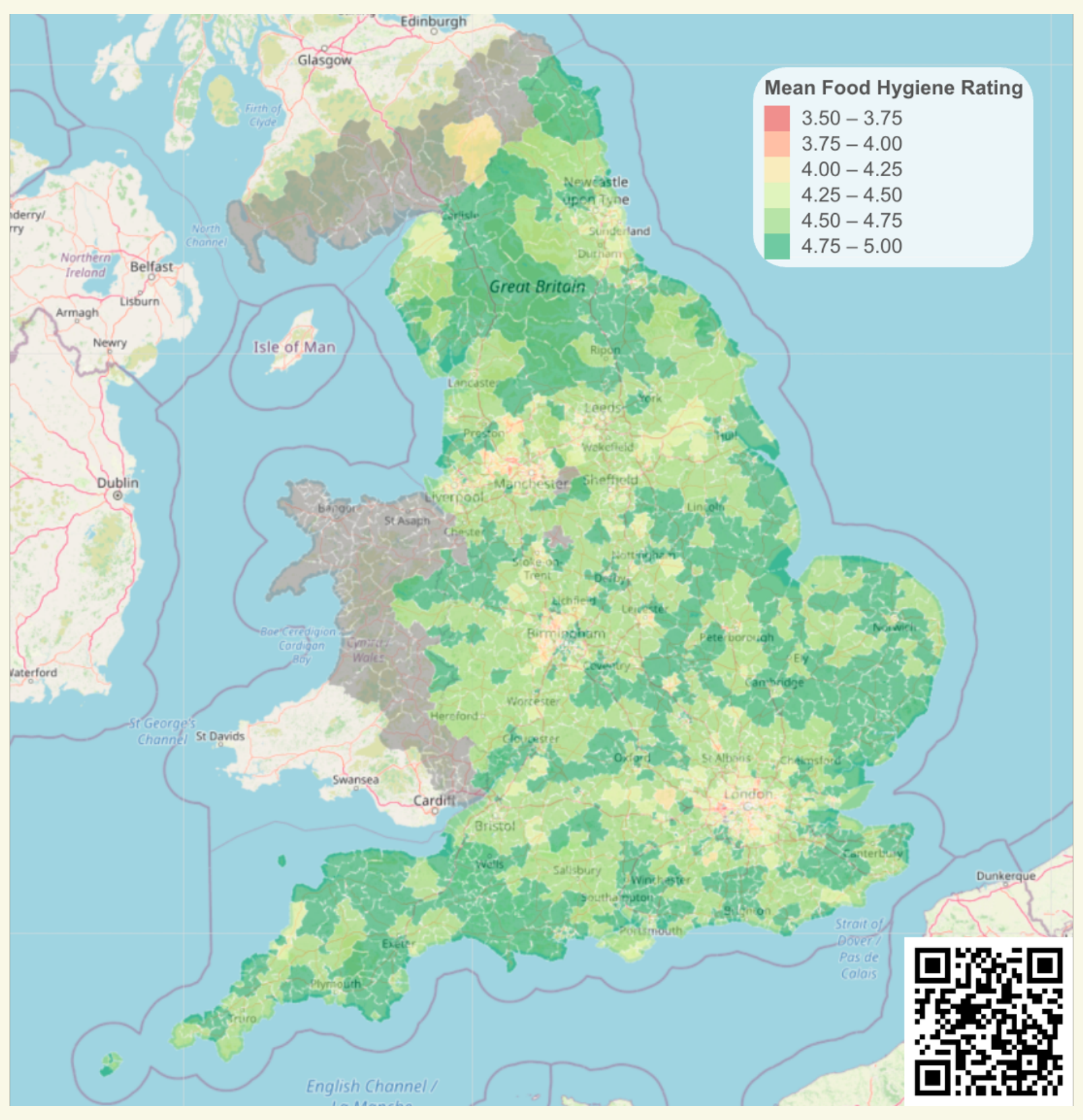


Figure 2: Mean Food Hygiene Ratings by postcode district.

- We would expect to see random scatter of Food Hygiene Ratings if there was no correlation between ratings and geographical location of establishments
- However, from Figure 2, we see that this is not the case
- In general, there are clusters of lower rated districts in city centres

Deprivation Data

- The government publish deprivation data every four to five years
- England is split up into 32,844 Lower Layer Super Output Areas (LSOAs) for purposes such as the census and deprivation data
- Each LSOA is given a score (the higher the more deprived) for each of the following seven criteria:
  - Income
  - Employment
  - Education
  - Health
  - Crime
  - Barriers to Housing
  - Living Environment
- An overall deprivation score is calculated by combining the seven above statistics. We see the score vs rank for every LSOA in England in Figure 3

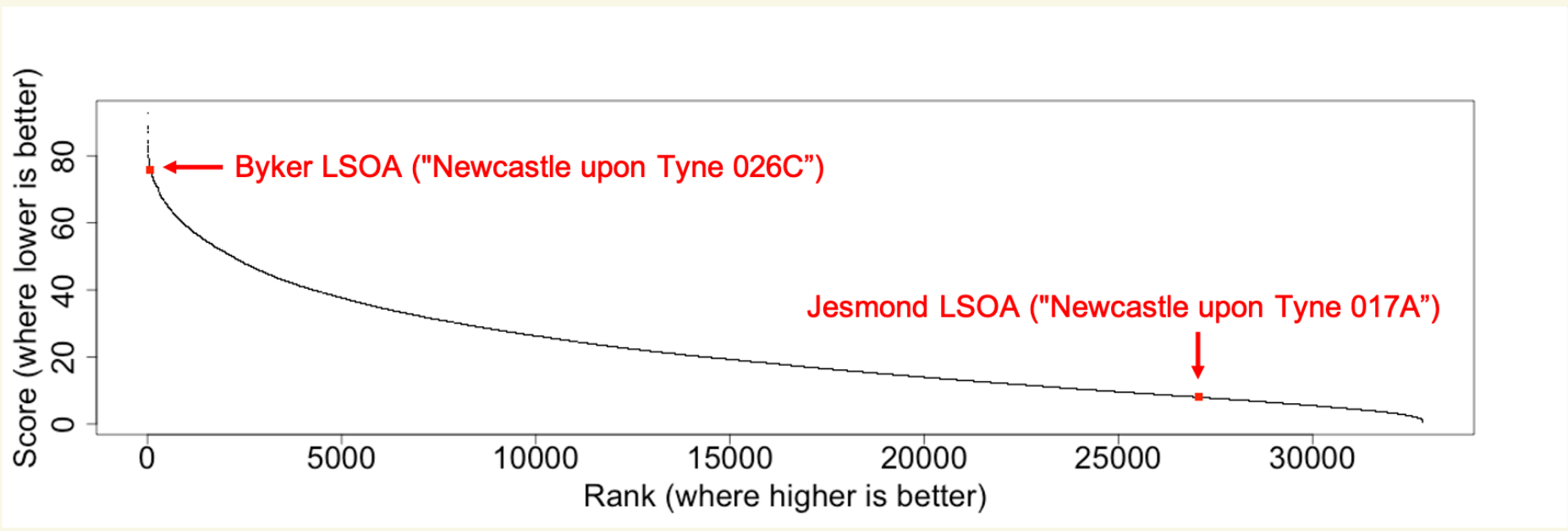
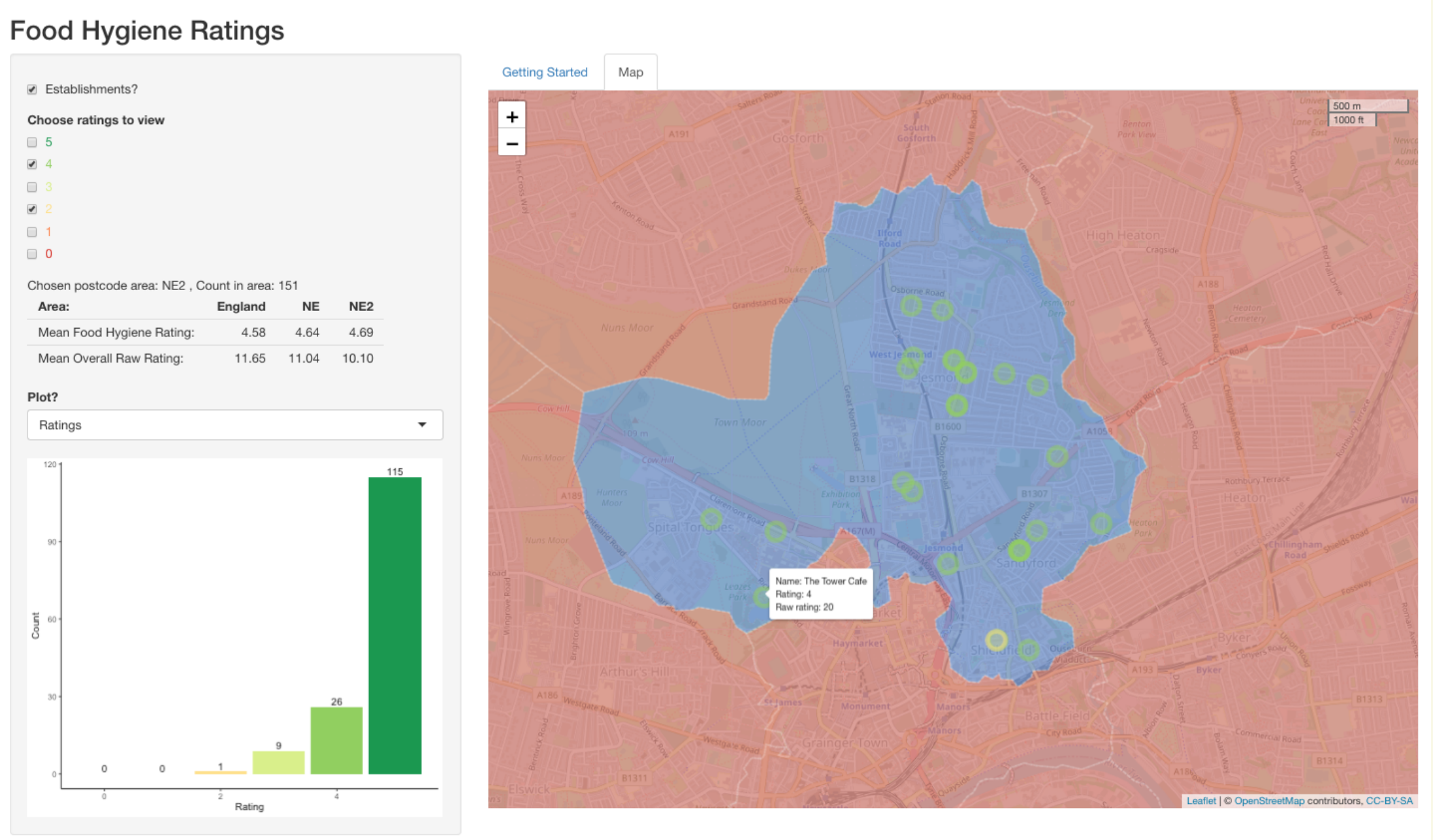


Figure 3: Score vs rank for all the LSOAs in England.

- Deprivation scores are linked to the LSOA being an urban/rural area
- 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% when we consider rural areas
- Therefore, some of the variation we see in Figure 2 could be explained by deprivation data

Shiny Application (see footer)

- As visualising the data, as seen in Figure 2, is clearly very informative we have developed a Shiny[2] 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
- Figure 4 shows a screenshot from the app - with the NE2 postcode district chosen, viewing the establishments with ratings 2 and 4. The cursor is hovering over *"The Tower Cafe"*


Food Hygiene Ratings

☑ Establishments?

Choose ratings to view

☐ 5
☑ 4
☐ 3
☑ 2
☐ 1
☐ 0

Chosen postcode area: NE2 , Count in area: 151

Area:	England	NE	NE2
Mean Food Hygiene Rating:	4.58	4.64	4.69
Mean Overall Raw Rating:	11.65	11.04	10.10

Plot?

Ratings

Name: The Tower Cafe
Rating: 4
Raw rating: 20

Figure 4: Screenshot from the Shiny App.

Ordinal Logistic Regression: Background

- Statistical models help determine which (if any) covariates are significant in explaining the distribution of ratings
- Specifically, we fit **ordinal logistic regression** models, which were first considered by Peter McCullagh in 1980 [3]
- An *ordinal logistic regression* model is appropriate when the outcome is ordered but we do not necessarily know the differences between the outcomes
- Our outcomes (the ratings) are ordered but the difference between 4 and 5 is not necessarily equal to the difference between 0 and 1
- We include covariates in the same way as in *simple linear regression* - we determine the effect of a covariate whilst "fixing" all the other covariates

Ordinal Logistic Regression: Definition

- Let Y be an ordinal outcome with $\mathcal{J}$ categories. Then $\Pr(Y \leq j)$ is the cumulative probability of Y being less than or equal to a specific category, with $j = 1, \ldots, \mathcal{J} - 1$
- The odds of being less than or equal to a particular category is:

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

  for $j = 1, \ldots, \mathcal{J} - 1$
- 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 ordinal logistic regression in `R` we use the `clm()` 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

Modelling

- We have built an ordinal logistic regression model which includes the following covariates: deprivation score of postcode of establishment, food chain indicator, type of establishment and local authority
- Establishments are:
  - **2x** more likely to have a lower rating if they are in the **most** deprived area, compared to the **least** deprived
  - **5x** more likely to have a lower rating if they **are not** a food chain, compared to if they **are** a food chain
  - **10x** more likely to have a lower rating if they are a **takeaway/sandwich shop**, compared to if they are a **school/college/university**
  - **7x** more likely to have a lower rating if they are in the **Barking and Dagenham** local authority, compared to if they are in the **Richmondshire** local authority
- These interpretations indicate that there are many different factors which influence the rating of an establishment; local authority is interesting as we would expect there to be no difference in ratings after accounting for other covariates
- Figure 5 shows the regression estimates for the types of establishment; the lower the estimate the more likely the establishments of that type will be rated lower

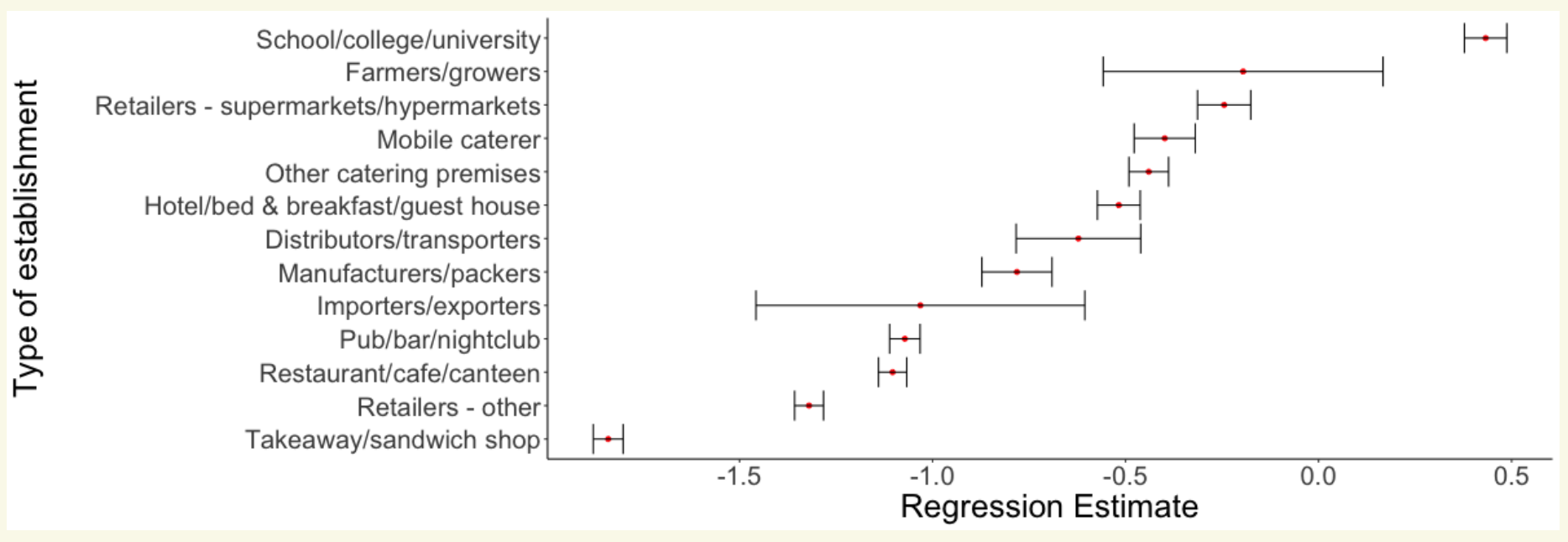


Figure 5: Regression estimates and error bars (±2× Std.Error) for the type of establishment (with caring premises as the baseline).

- Figure 6 shows the regression estimates for the local authorities; again, the lower the estimate the more likely establishments in the local authority will be rated lower

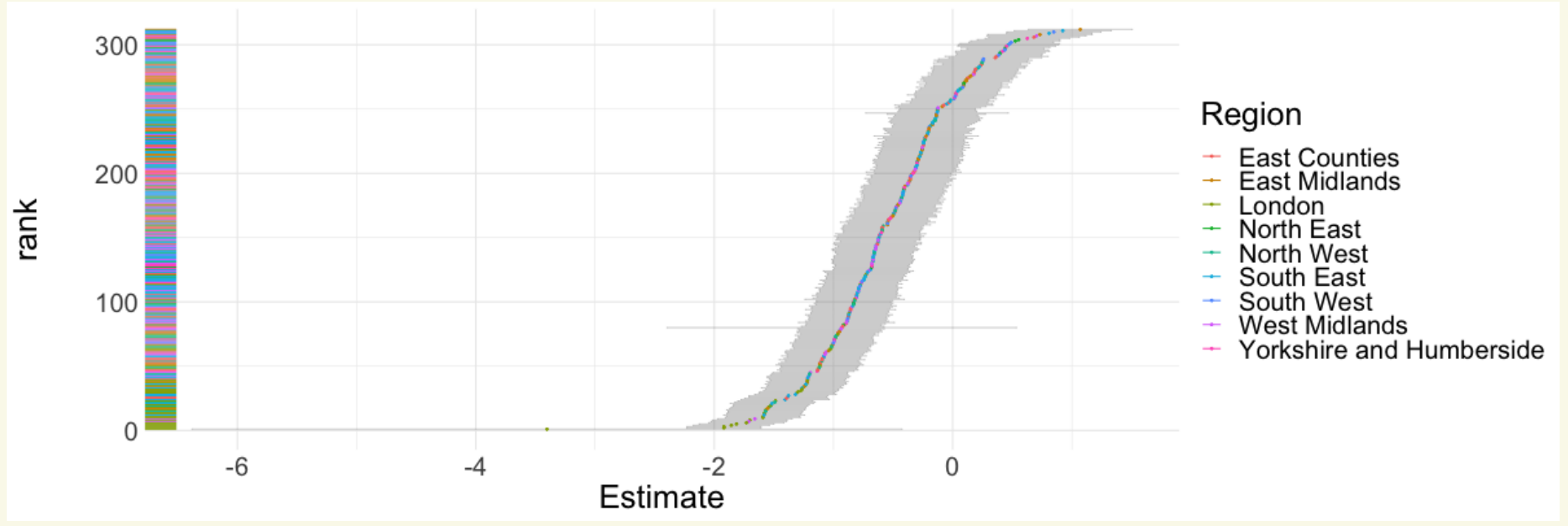


Figure 6: Regression estimates and error bars (±2× Std.Error) for the local authority (with Adur, West Sussex as the baseline).

Conclusion

- We were able to visually see that Food Hygiene Ratings are not randomly scattered across England
- We were able to find covariates which explain the distribution of ratings, such as: deprivation data, food chain, type of establishment and local authority
- We were able to develop a Shiny App (see footer) which allows users to view a map of England by postcode district

References

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

[2] 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.

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

Newcastle University

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

Newcastle University