Analysing Roadkill Trends in the UK: Identifying Spatial and Temporal Patterns

"I have familiarised myself with the academic misconduct and plagiarism guidelines in the Academic Honesty and Plagiarism module and the MTHM601 (Fundamentals of Data Science) ELE site's Assessment Information tab. This report constitutes my own work, and I have explicitly referenced and acknowledged those parts that draw on the literature, online sources, and the support of others. This includes acknowledgment of any use of Artificial Intelligence tools such as ChatGPT"

Introduction

Topics: Data Wrangling, Advanced Visualisation, RMarkdown, Hierarchical Modelling, Spatial Visualisation, Temporal Trends, GitHub

With the ever-changing climate, British wildlife face an uncertain future, with a major decline in biodiversity being found all around the United Kingdom. But another thing that could be adding to the decline in British wildlife is road-traffic-accidents; collisions with vehicles is one of the major causes of wild animal death in the UK [1]. For animals that are already under immense pressure due to anthropogenic and climate threats, roadkill could push some of these species to extinction.

This report aims to provide a comprehensive exploration of roadkill trends across the UK, leveraging data science techniques to address questions about its spatial, temporal and environmental dynamics, with a specific focus on mammals.

Focusing on mammals is particularly relevant due to their ecological importance and the disproportionate impact of roadkill on their populations. Many mammal species are wide ranging with large territories that are increasingly being split up by roads, making them more vulnerable to collisions. By narrowing the scope to mammals, this study aims to provide actionable insights for mitigating roadkill impacts on vulnerable mammalian species.

The three main questions I am hoping to answer are:

1. What are the spatial and seasonal trends in mammalian roadkill numbers?

2. Do temperature or rainfall play a role in mammalian roadkill prevalence?

3. Where are the mammalian roadkill hotspots across the UK?

Objectives/Methodology

Talk about QQ plots

Q1

Spatial - using tmap package in r - create a map of regional roadkill reports

q2

QQ plots, leverage, etc of the linear model are not good. Trying GAM instead.

Q3

Data

Roadkill Data

The data I will use is from The Road Lab (formerly 'Project Splatter') [https://www.theroadlab.co.uk/], it is a citizen science project with 57 columns and 68,212 rows of data. The dataset includes mammals, birds, amphibians and reptiles but, as this report is only interested in mammals, all rows corresponding with other Classes will be removed. I also removed any rows that had blank entries in any of my columns of interest.

Data was collected between 01/01/2014 until 30/09/2024 but as there was only one entry and it was not the complete year, data from 2024 was removed. There were several recordings of "indet. Deer", "rabbits and hares" and a "Polecat-Ferret" which were also removed as they were not to a species level. Some records were grouped under one standardised species name, "Eastern Grey Squirrel" and "Squirrel" became "Grey Squirrel" because Red Squirrels (LATIN) are rare in most parts of the UK so these records were separated from those of "Red Squirrel", the same goes for "Brown Rat" and "Rat spp." because Black Rats (LATIN) are also very rare so any records were assumed to be Brown Rats (LATIN). I also grouped all mouse, bat, vole and hare records into their respective group rather than species level.

It is important to note that there is a large spike in reports in July and August 2019 (figure 1a), I have decided try to scale it in line with reports from other years, as it is not likely to reflect an actual massive increase in roadkill but it likely a result of increased news coverage on the project [https://www.bbc.co.uk/news/science-environment-48886673]. I followed the methods from Raymond et al's [1] paper for this but I feel that it still left a spike too large to be representative (figure 1b) of the actual roadkill numbers, ultimately leading to the removal of the reports from July and August 2019. I then calculated the mean number of roadkill reports for July, August and September separately, not including 2019 values. Once I had the means, I randomly selected that number from the 2019 reports for each month and removed the rest from those months (figure 1c). This was the only solution I could think of that kept the majority of the data without skewing the seasonal trends in roadkill.

Once the data was cleaned and all unneeded data was removed, I was left with 39130 entries.

Seasonal/Weather Data

This data will then be combined with data on the mean monthly temperature and total rainfall per month from the Met Office. [https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-and-regional-series]

According to the met office, the seasons will be split into months. The months of March, April and May are now "Spring", June, July and August are "Summer", September, October and November are "Autumn" and December, January and February are "Winter" months.

Results

Exploratory Analysis

Graph showing the

- Reports per season
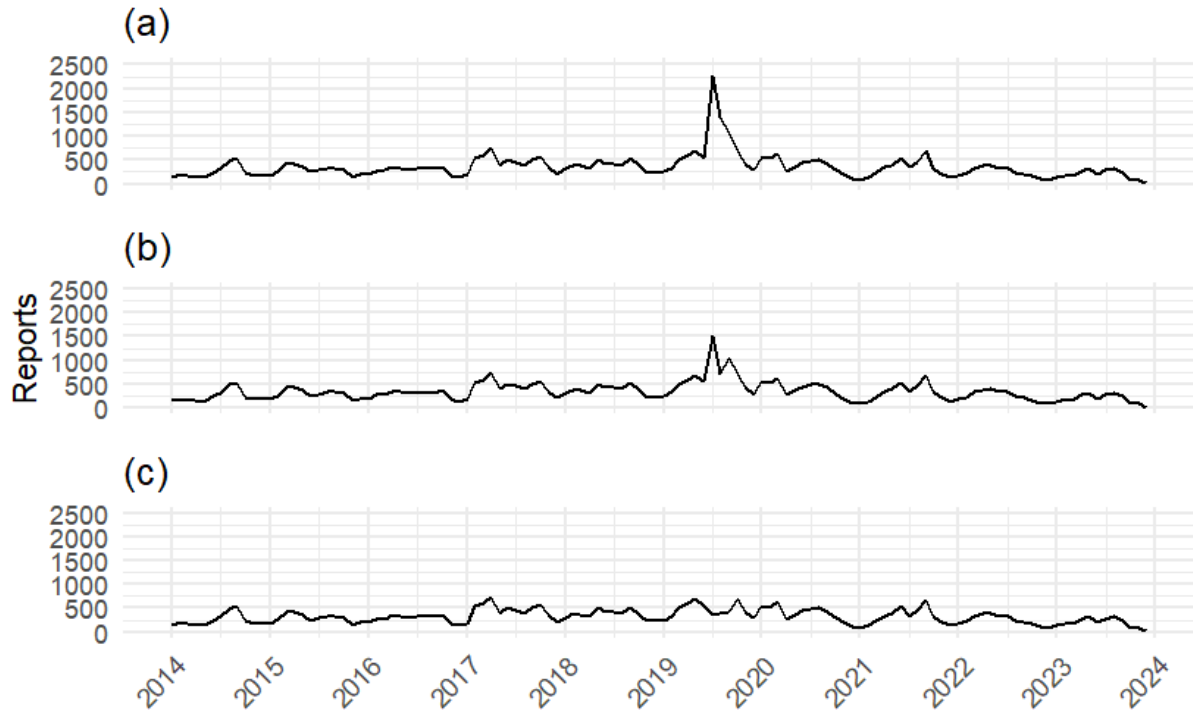- Multiple maps of the UK showing reports per season/year

Figure 1: (a) Graph showing total monthly roadkill reports from 2014 to 2024, showing the spike in reports in July and August 2019. (b) Graph showing total monthly reports of roadkill from 2014 to 2024 after scaling and subsampling the inflated reports from July and August 2019. (c) Graph showing total monthly reports of roadkill from 2014 to 2024 after taking a random sample of the reports from July, August and September that were equal to the means from those months in the other years.

Seasonal trends in roadkill

Four negative binomial Generalised Linear models were fitted to evaluate the effects of year, season and their interaction on roadkill counts. The Year-only model had the highest AIC (591.84), indicating poor fit compared to models including Season. Adding Season significantly improved the model, as evidenced by a lower AIC (584.73) and a significant Likelihood Ratio Test ($\chi^2 = 11.11$, p < 0.01; Table 1). The Year + Season model had a slightly higher AIC (585.91) and did not significantly improve fit over the Season-only model ($\chi^2 = 0.82$, p = 0.36). Similarly, the interaction model was more complex but failed to improve fit (AIC = 590.42, $\chi^2 = 1.48$, p = 0.69). These results suggest that Season is a significant predictor of roadkill counts, while Year and the Year $\times$ Season interaction do not provide additional explanatory power. The Season-only model provides the best balance of simplicity and fit.

Model	AIC	LogLik	DF	LR_Stat	P_Value
Year	591.84	-292.92	3		
Season	584.73	-287.36	5	11.11	**<0.01**
Year + Season	585.91	-286.95	6	0.82	0.36
Year * Season	590.42	-286.21	9	1.48	0.69

The negative binomial model indicated that winter had significantly lower roadkill counts compared to the reference season, Autumn (p = 0.0237). No significant differences were observed for Spring (p = 0.2993) or Summer (p = 0.3419) relative to Autumn. This suggests that roadkill counts are relatively stable across spring, summer, and autumn, but decrease significantly in winter (figure 2).

Impacts of weather on roadkill

Generalised Linear Model with a poissonn family was used due to the response variable being count data.
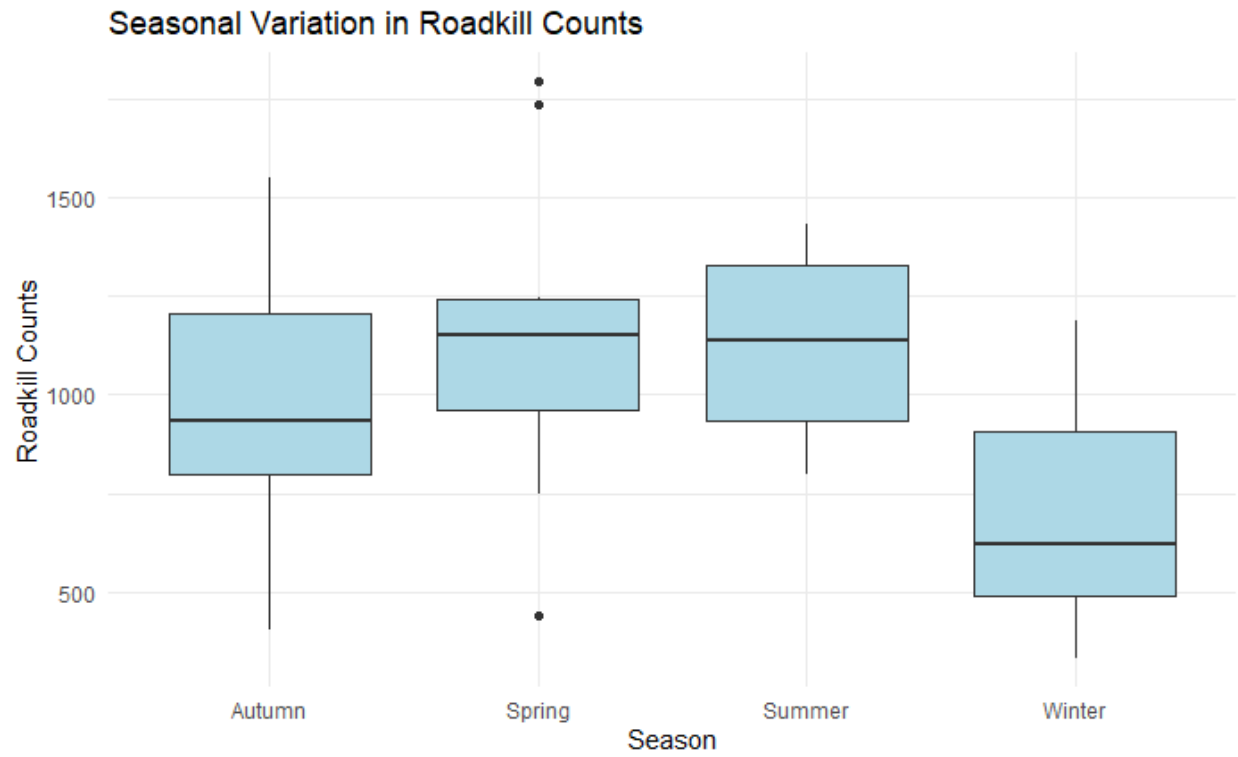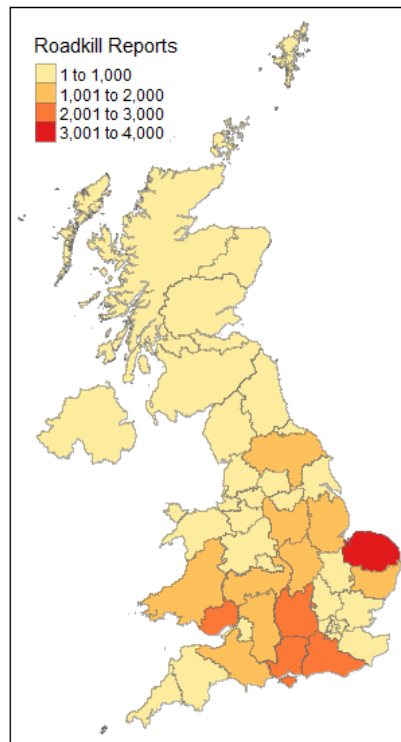
Figure 2: Boxplot of seasonal variation in roadkill reports

Where are the roadkill hotspots?

View an interactive version of this map here

Limitations

Limitations of the data

- Larger species are more likely to be seen and reported

- Only covers animals that die immediately and not those that get out of the road before dying

- Citizen science always comes with limitations

- Bias in sampling - massive increase after coverage in the news but only lasted for 2 months

- However, previous studies in South Africa [2] and California have shown the identification data to be reliable [3]

Limitations of the analysis

Correlation does not mean causation!

My data is not completely normal so some analyses may not be completely accurate

Potential to look into species trends but wasnt within the scope for this report

Conclusions

Reproducibility

All code and data files used for the analysis in this report can be found in a GitHub repository here:

References

[1]   S. Raymond, A. L. W. Schwartz, R. J. Thomas, E. Chadwick, and S. E. Perkins, "Temporal patterns of wildlife roadkill in the UK," *PLOS ONE*, vol. 16, no. 10, p. e0258083, Oct. 2021, doi: 10.1371/journal.pone.0258083.

[2]   S. Périquet, L. Roxburgh, A. le Roux, and W. J. Collinson, "Testing the Value of Citizen Science for Roadkill Studies: A Case Study from South Africa," *Frontiers in Ecology and Evolution*, vol. 6, Feb. 2018, doi: 10.3389/fevo.2018.00015.

[3]   D. P. Waetjen and F. M. Shilling, "Large Extent Volunteer Roadkill and Wildlife Observation Systems as Sources of Reliable Data," *Frontiers in Ecology and Evolution*, vol. 5, p. 89, Aug. 2017, doi: 10.3389/fevo.2017.00089.