Student ID: 7038293

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, Data Analysis, 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 [1], [2]. 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 [3]. For animals that are already under immense pressure due to anthropogenic and climate threats, roadkill could push some of these species to extinction.

Globally, road networks were an estimated 21.6 million kilometres in 2018, with this figure expected to increase by 14-23% by 2050 [4], [5]. In the UK, where road density is among the highest in Europe [6], the risk of roadkill to wildlife populations is of particular conservation concern. Especially when half of all UK grassland species are found along the roadside verges [7].

This report investigates the complex dynamics of roadkill across the UK, focusing specifically on mammals due to their ecological significance and the disproportionate impact roadkill has 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.

A key component of this report is the contribution citizen science projects play in wildlife conservation. In recent years, citizen science initiatives have run where members of the public record and share data on roadkill incidents, providing an expansive and detailed dataset that would not be possible through traditional surveying methods. These projects not only improve our understanding of roadkill trends but also encourage public engagement and awareness about wildlife conservation.

By utilising data collected by citizen scientists, the three main questions I am hoping to answer are:

1. What are the seasonal trends in mammalian roadkill numbers?
2. Where are the mammalian roadkill hotspots across the UK?
3. Do temperature or rainfall play a role in mammalian roadkill prevalence?

Using advanced data science techniques, including spatial visualisation and temporal analysis, this study provides a detailed examination of roadkill patterns. Through these analyses, it aims to shed light on the interconnected nature of ecological processes and human activities, emphasising the need for coexistence between humans and wildlife.

Methodology

All analyses were conducted using R Statistical Software (v4.4.2)[8], with the aim of addressing three key research questions through a combination of data wrangling, statistical modelling and data visualisation.

The datasets used for this analysis, roadkill and environmental data (rainfall and temperature), were sourced from reliable national citizen science programmes and databases. These data were cleaned to remove inconsistencies and missing values. Relevant variables were standardised, ensuring compatibility across datasets. The resulting datasets were prepared for analysis by structuring them based on the research question being answered.

Seasonal trends in roadkill

To investigate seasonal variations in mammalian roadkill, negative binomial generalised linear models (GLMs) were applied. These models accounted for overdispersion in the count data and included season and month as explanatory variables. This analysis was used to highlight periods of increased roadkill prevalence.

Where are the roadkill hotspots?

The roadkill records were divided into regions of the UK to show which county has the highest roadkill prevalence. To validate the spatial data and check assumptions about its distribution, normality was assessed within regions using Q-Q and diagnostic plots. This showed that the data were not normally distributed and overdispersed, justifying the use of a negative binomial GLM. An interactive map was created to visualise the spatial patterns dynamically, with a static version included in this report to summarise the findings and highlight regional disparities in roadkill prevalence.

Impacts of weather on roadkill

To explore the influence of environmental factors on roadkill counts, a GLM was used. This incorporated temperature and rainfall as fixed effects and region as a random effect to account for geographical variability. During preliminary analysis, linear models were considered however diagnostic plots showed significant deviations from the model's assumptions. Q-Q plots of residuals indicated non-normality and high-leverage points were also identified, showing the need for a different model. The GLM addressed these limitations, providing an insight into the relationship between roadkill counts and weather conditions while accounting for regional differences.

Key R Packages I used:

- **Data wrangling:** tidyverse, dplyr[9]
- **Data visualisation:** tmap[10], flextable[11], ggplot2[12]
- **Data Analysis:** MASS[13], emmeans [14]

Data

Roadkill Data

The data I will use is from The Road Lab (formerly 'Project Splatter')[15], 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 (*Sciurus vulgaris*) 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 (*Rattus rattus*) are also very rare so any records were assumed to be Brown Rats (*Rattus norvegicus*). I also grouped all mouse, bat, vole and hare records into their respective group rather than species level. This cleaned the dataset ready for the analysis in this report and future directions also.

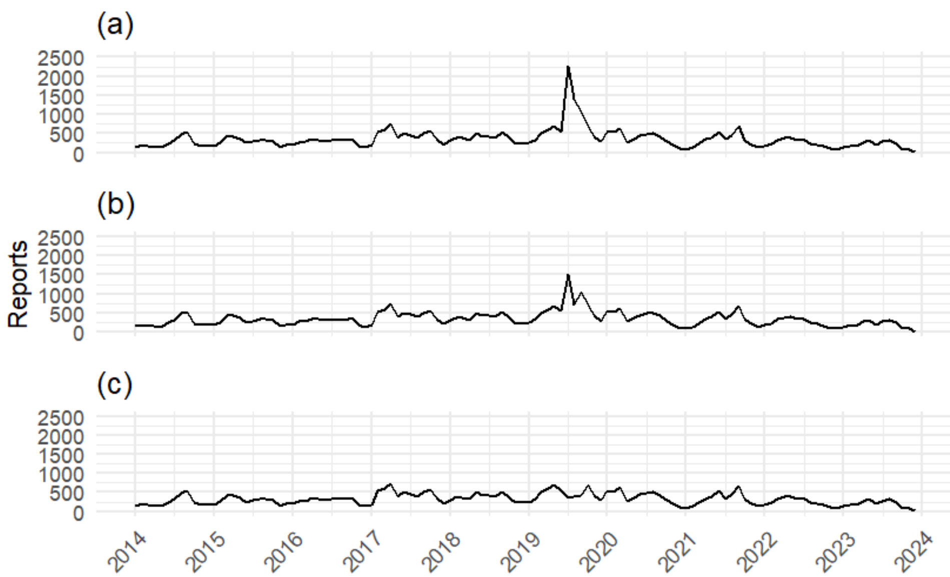

Figure 1: Graph showing total monthly roadkill reports from January 2014 to January 2024, showing (a) the spike in reports in July - September 2019. (b) after scaling and sub-sampling the inflated reports from those months. And (c) adjusted data after random subsampling to align with typical monthly trends.

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 [16]. I followed the methods from Raymond et al’s [3] 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 39230 entries. I also created a separate dataset that combined the reports for each month to create a monthly count of roadkill reports for some of the plots and analyses.

I also added a column to the dataset encompassing season. The seasons were categorised based on the UK Meteorological Office meteorological season definitions [17]: March, April and May are were classified as “Spring”; June, July and August as “Summer”; September, October and November as”Autumn” and December, January and February as “Winter”.

Weather Data

The weather data used in this analysis were obtained from the Meteorological Office’s UK and Regional Series datasets [18], which provide time-series data on monthly, seasonal and annual climate values. Specifically, the year-ordered data includes mean monthly temperature (°C; figure 2a) and total monthly rainfall (mm; figure 2b) for the entire UK. These variables were chosen to investigate their potential influence on roadkill counts. The data spans decades, allowing for an examination of temporal trends combined with roadkill data.

To prepare the data for analysis, some wrangling was required. First, I filtered the datasets to only contain the years 2014 to 2024 and then I mutated the tables from wide-format into a long format. I then combined the weather data into the roadkill dataset, based upon the date and season.

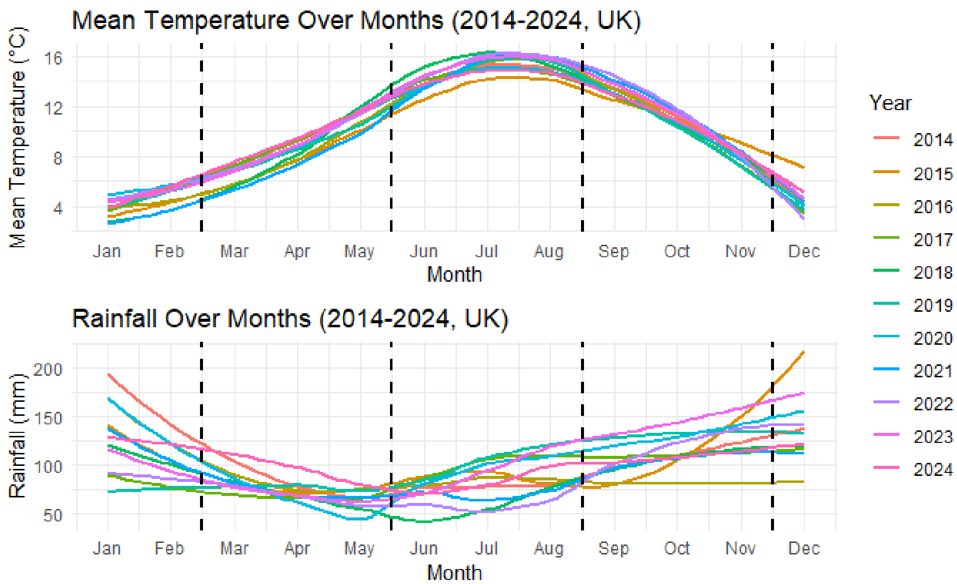

Figure 2: Plot of (a) mean monthly temperature (°C) and (b) mean monthly rainfall (mm) across the UK. Colour coded by year, demonstrating inter-annual variations in climate variables over the study period.

Results

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. To determine the best-fitting model for roadkill counts, I conducted likelihood ratio tests (Chi-squared tests) between nested models. These tests assess whether adding predictors or including interaction terms significantly improves the model fit. This approach allows for a step-wise comparison of the models to identify the simplest model that adequately explains the variability in 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, χ^2= 1.48, p = 0.69). These results suggest that Season is a significant predictor of roadkill counts, while Year and the Year × Season interaction do not provide additional explanatory power. The Season-only model provides the best balance of simplicity and fit.

Table 1: Model comparison for seasonal effects on roadkill counts using negative binomial generalized linear models.

Model	AIC	LogLik	DF	LR stat	P Value
Year	591.84	-292.92	3	N/A	N/A
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 chosen negative binomial model suggested 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 3).

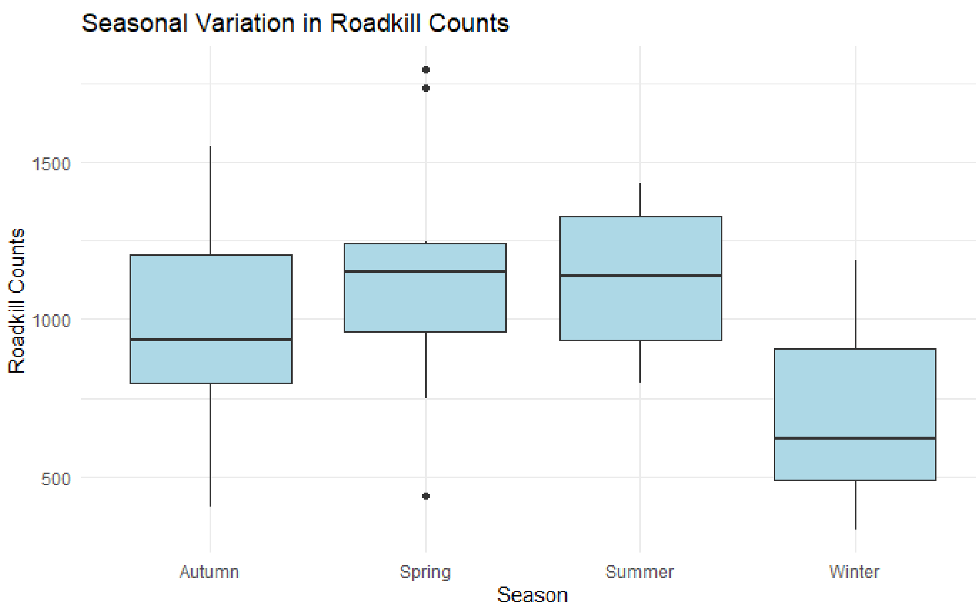

Figure 3: Boxplot of seasonal variation in roadkill reports

Where are the roadkill hotspots?

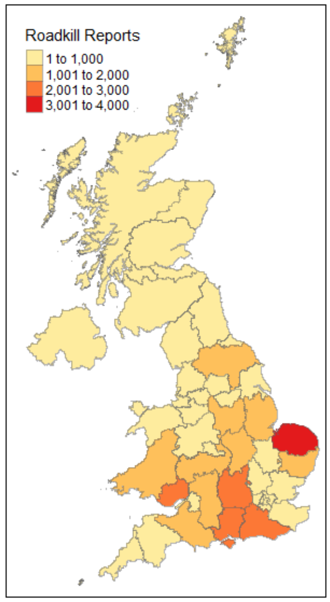

Figure 4: Static map showing roadkill counts per region in the UK.Darker colours represent higher roadkill report counts.

Code for an interactive version of this map is available in my GitHub **(here).**

There is large regional variation in roadkill counts across the UK (figure 4), the region with the highest roadkill count is Norfolk (n=3760), indicated by the dark red colouration, and the region with the lowest was Inner London - West (n=8). To assess the significance of variation among regions, I wanted to use a GLM. To determine the correct family I first checked normality, I checked for over-dispersion using this dispersion statistic:

$$\text{Dispersion} = \frac{\text{Residual Deviance}}{\text{Degrees of Freedom}}$$

Any values above 1 show over-dispersion, for a Poisson GLM using this data, the dispersion statistic was infinite, meaning it was completely over-dispersed. Meaning a negative binomial GLM was required instead, and after checking the models assumptions (figure 5), the Residuals vs Fitted shows no strong pattern, the Scale-Location plot suggests mild heteroscedasticity, the Q-Q plot shows that the residuals are approximately normal and the Residuals vs leverage plot shows that outliers do not have a large influence on the model.

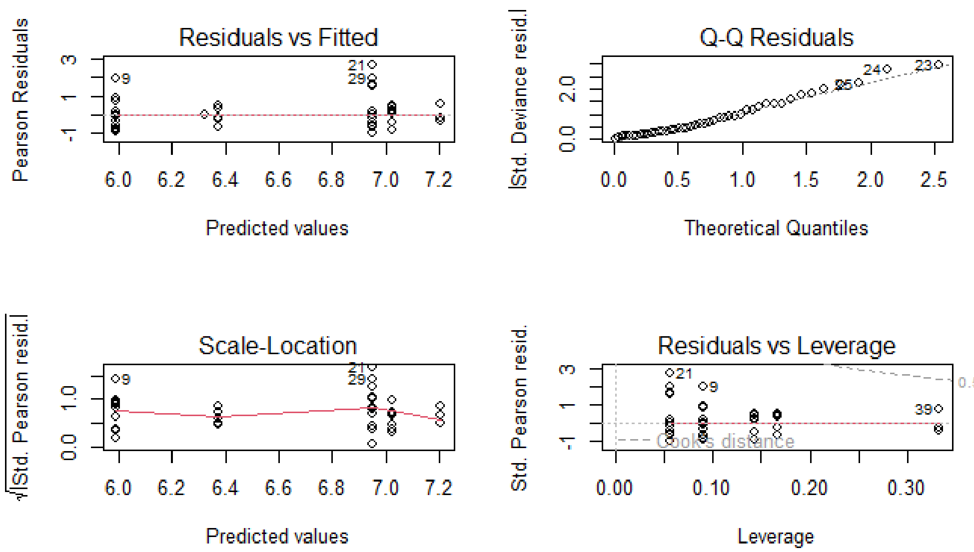

Figure 5: Diagnostic plots for the negative binomial GLM looking at variation of roadkill numbers based on region.

Regions were grouped into North, South, Midlands, Scotland, Wales and Northern Ireland for analysis. The negative binomial GLM revealed no significant differences in roadkill counts across most of these groups, except for one. Relative to the midlands, the North had significantly fewer roadkill counts (β = -1.034, p = 0.025). Using estimated marginal means, the predicted mean roadkill counts for each region group (back-transformed from the log scale) were computed (table 2). The Midlands had the highest predicted mean count and the North had the lowest predicted.

Table 2: Predicted mean roadkill counts for grouped regions based on the negative binomial GLM. The Midlands region shows the highest predicted counts, while the North shows the lowest. Confidence intervals reflect back-transformed log scale predictions.

Region Group	Predicted Count	Standard Error	df	95% CI (Lower)	95% CI (Upper)
Midlands	1125.71	406.16	Inf	555.03	2283.18
North	400.18	115.28	Inf	227.54	703.82
Northern Ireland	558.00	532.92	Inf	85.84	3627.26
Scotland	587.00	228.86	Inf	273.38	1260.39
South	1043.67	234.83	Inf	671.49	1622.14
Wales	1347.33	742.49	Inf	457.50	3967.88

Pairwise comparisons between the grouped regions were performed to assess specific differences. After applying Tukey adjustments for multiple comparisons, the difference between the North and Midlands are marginally significant (p = 0.092) but there were no other significant comparisons between any of the other groups.

Impacts of weather on roadkill

Generalised Linear Model with a poisson family was used due to the response variable being count data, with mean temperature (°C) and rainfall (mm) as predictor variables. But once checked for over-dispersion, the dispersion statistic was 63, a negative binomial GLM was used instead. A 1°C increase in mean temperature is associated with a 3.3% increase in roadkill counts (figure 6), with the model indicating a significant positive relationship between mean temperature and roadkill counts ($z - Value$ = 3.163, p-value = 0.0016). This could be due to an increase in animal activity during warmer weather that may increase their interaction with roads.

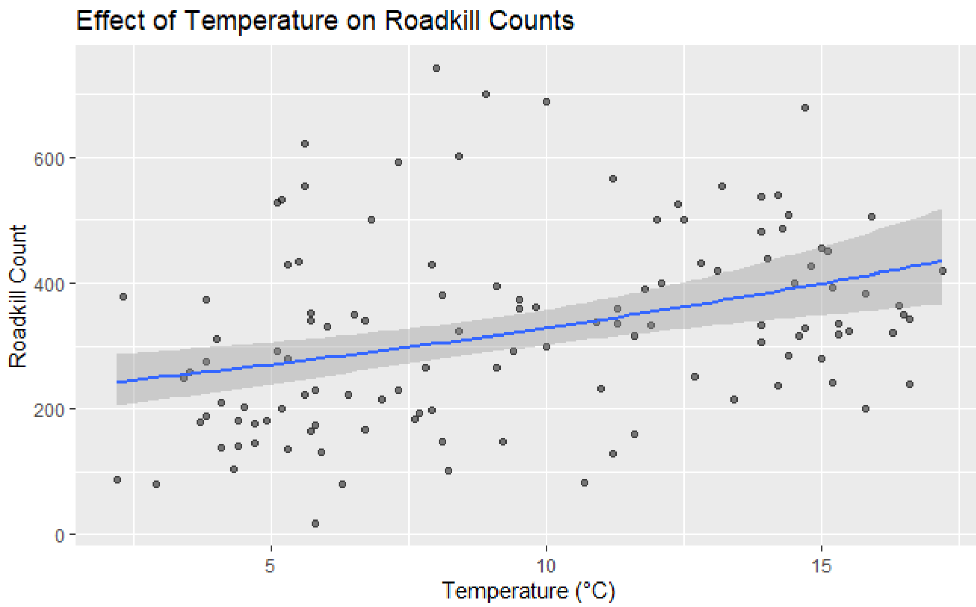

Figure 6: Effect of mean temperature on roadkill counts. The regression line represents the fitted negative binomial GLM, showing the positive relationship between temperature and roadkill counts.

Similarly, a 1mm increase in rainfall is associated with a 0.29% decrease in roadkill counts (figure 7). The model shows a significant negative relationship between rainfall and roadkill ($z - value$ = -2.752, p-value = 0.0059), which could be explained by a reduction in animal movement during heavy rain or reduced vehicle speeds, giving drivers more time to avoid animals on the road.

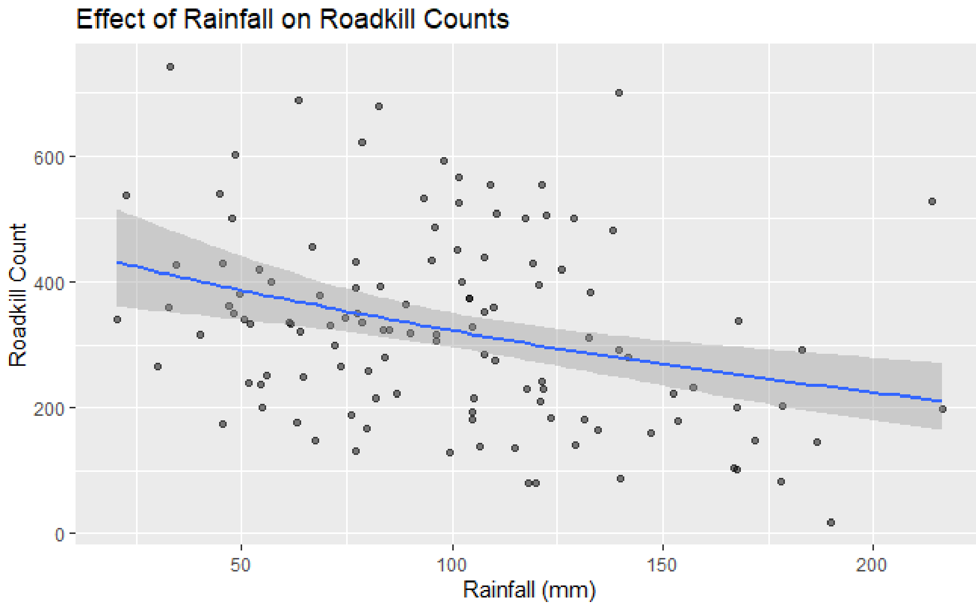

Figure 7: Effect of rainfall on roadkill counts. The regression line represents the fitted negative binomial GLM, showing the negative relationship between rainfall and roadkill counts.

Limitations

Limitations of the data

The data used in this analysis comes with several limitations that need to be addressed. One significant issue is the inherent visibility bias, where larger species are more likely to be seen and therefore reported, leading to a potential under-representation of smaller species in the dataset. It also only incorporates animals that die instantly on the road, and excludes those that leave the road before succumbing to their injuries. As with any citizen science project, there are limitations associated with variability in reporting effort and observer expertise. A notable example is the bias introduced by an increase in reporting following media coverage, which only lasted for a few months and created an uneven temporal sampling pattern.

There is also the chance that species may be misidentified but previous studies in South Africa [19] and California [20] have shown the identification data to be reliable. It is also likely that data collection will not be uniform across all regions, leading to gaps in areas with lower participation and inflated figures in areas of high participation. Furthermore, observers may be less likely to report roadkill of very common species, resulting in under-representation in the dataset.

Limitations of the analysis

The data analysis itself comes with its own limitations. It is important to emphasise that correlation does not imply causation, while they may suggest relationships, they cannot confirm underlying causative factors. The dataset is not completely normally distributed, which may affect the accuracy of some statistical analyses.

Several variables that could influence roadkill rates, such as traffic volume and habitat type, were not incorporated in analysis. Although there is potential to assess species specific trends in roadkill, this analysis did not explore them due to limited scope of this report., Similarly, looking at spatial differences across regions is difficult when there is large variation in size, which is not something that I accounted for in my analysis.

Conclusions

Seasonal Trends

Roadkill reports are significantly lower in the Winter than any of the other seasons, with report numbers being relatively stable across Spring, Summer and Autumn. This could be due to a number of ecological and environmental factors. During winter, many mammal species reduce their activity due to colder temperatures and limited food availability, some species enter hibernation or even migrate to other regions, further reducing their presence in areas where roadkill incidents typically occur. Environmental factors such as icy road conditions might also contribute to reduced speeds, giving drivers more time to react to animals on the road. Collectively, these factors likely explain the seasonal dip in roadkill counts during the winter months. However, shorter daylight hours in winter may result in more people traveling during darkness, which could limit their ability to see and report roadkill incidents, leading to under-reporting.

Where are the roadkill hotspots?

The county with the highest roadkill count was Norfolk , while the region with the lowest count was Inner London - West, which potentially shows the difference between urban and rural roadkill patterns. These findings show significant differences in roadkill counts between regions, these differences could be due to variations in road density, traffic volume and wildlife distribution. These findings show important regional disparities in roadkill counts, which are likely due to both ecological and anthropogenic factors. The high counts in Norfolk and the Midlands may reflect higher wildlife densities and increased road exposure. Conversely, the low counts in Inner London could highlight a reduction in wildlife distribution and reduced interaction between wildlife and roads.

Impacts of weather on roadkill

Roadkill counts are significantly affected by both temperature and rainfall but in opposite ways, an increase in temperature leads to an increase in roadkill reports, whereas an increase in rainfall leads to a decrease in roadkill reports. Warmer temperatures may exacerbate roadkill incidents, likely due to increased wildlife or human activity. Whereas, rainfall may reduce roadkill incidents, potentially by reducing animal activity or altering driver behaviour. These insights could be used to inform mitigation strategies such as targeted wildlife crossings or even speed restrictions in the warmer months.

Future directions

Future analyses should incorporate environmental and anthropogenic factors such as road density or type, habitat and traffic volume to better understand the variation in roadkill patterns. Also breaking analysis down into species level will enable targeted conservation measures for particularly vulnerable species, and doing the same for spatial analysis, using finer-scale spatial data might highlight local scale hotspots. Using advanced statistical methods and machine learning techniques could be used to predict roadkill hotspots or high-risk periods with greater accuracy.

The results could be used to inform conservation and mitigation strategies. Either through recommending locations for wildlife crossing structures or policy and road planning recommendations. They could also be used to enhance public awareness and engagement in other citizen science projects.

Reproducibility

All code and data files used for the analysis in this report can be found in a GitHub repository **(here)**.

References

[1] "Biodiversity in the UK: bloom or bust?" Available: https://publications.parliament.uk/pa/cm5802/cmselect/cmenvaud/136/136-report.html

[2] M. Lambertini, "8 BILLION REASONS TO SAFEGUARD NATURE."

[3] 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.

[4] L. J. Moore *et al.*, "Demographic effects of road mortality on mammalian populations: A systematic review," *Biological Reviews*, vol. 98, no. 4, pp. 1033–1050, Aug. 2023, doi: 10.1111/brv.12942.

[5] J. R. Meijer, M. A. J. Huijbregts, K. C. G. J. Schotten, and A. M. Schipper, "Global patterns of current and future road infrastructure," *Environmental Research Letters*, vol. 13, no. 6, p. 064006, May 2018, doi: 10.1088/1748-9326/aabd42.

[6] "Road network 2020 – ERF." Available: https://erf.be/statistics/road-network-2020/

[7] A. Perrett, C. Barnes, M. Schofield, L. Qie, P. Bosilj, and J. M. Brown, "DeepVerge: Classification of roadside verge biodiversity and conservation potential," doi: 10.48550/arXiv.2206.04271.

[8] R Core Team, *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, 2024. Available: https://www.R-project.org/

[9] H. Wickham, R. François, L. Henry, K. Müller, and D. Vaughan, *Dplyr: A grammar of data manipulation.* 2023. Available: https://CRAN.R-project.org/package=dplyr

[10] M. Tennekes, "tmap: Thematic maps in R," *Journal of Statistical Software*, vol. 84, no. 6, pp. 1–39, 2018, doi: 10.18637/jss.v084.i06.

[11] D. Gohel and P. Skintzos, *Flextable: Functions for tabular reporting.* 2024. Available: https://CRAN.R-project.org/package=flextable

[12] H. Wickham, *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York, 2016. Available: https://ggplot2.tidyverse.org

[13] W. N. Venables and B. D. Ripley, *Modern applied statistics with s*, Fourth. New York: Springer, 2002. Available: https://www.stats.ox.ac.uk/pub/MASS4/

[14] R. V. Lenth, *Emmeans: Estimated marginal means, aka least-squares means.* 2024. Available: https://CRAN.R-project.org/package=emmeans

[15] "The Road Lab UK." Accessed: Jan. 12, 2025. [Online]. Available: https://registry.nbnatlas.org/public/show/dp205

[16] "Badgers lead mammal roadkill list," *BBC News*, Jul. 2019, Available: https://www.bbc.com/news/science-environment-48886673

[17] "When does spring start?" Available: https://www.metoffice.gov.uk/weather/learn-about/weather/seasons/spring/when-does-spring-start

[18] "UK and regional series," *Met Office*. Accessed: Jan. 12, 2025. [Online]. Available: https://www.metoffice.gov.uk/research/climate/maps-and-data/uk-and-regional-series

[19] 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.

[20] 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, Aug. 2017, doi: 10.3389/fevo.2017.00089.