# Beyond Headlines: Quantifying Animal Welfare Narratives

## Web Scraping BBC News

João Pedro Donadio

## 1 Introduction

Public discourse on animal welfare plays a crucial role in shaping perceptions, public policy, and ethical standards surrounding the treatment of animals. News coverage in reputable media outlets can influence how people think and feel about animal welfare issues, including farming practices, conservation, and legislation.

This project explores the **frequency, tone, and framing** of animal welfare-related content in news media by scraping and analyzing articles published in **BBC News**, a leading international newspaper known for its environmental and ethical reporting.

Our focus is on collecting and analyzing article headlines and summaries from **BBC News** > Search: *Animal Welfare*. Because the site uses dynamic loading and pagination, we use `RSelenium` alongside `rvest` to access and extract article information effectively. This setup allows us to simulate browser behavior and navigate through dynamically rendered content. We aim to identify common language patterns, recurring themes, and the overall sentiment of animal-related articles, helping us understand how the media frames animal welfare.

This project simulates a real-world freelance data scraping and analysis task and is intended as a portfolio sample to demonstrate web scraping, text preprocessing, and basic natural language analysis using `rvest` and `tidytext` packages.

# 2  Objective

The goal of this project is to assess how animal welfare topics are covered in a popular news outlet by:

- Scraping article **titles**, **summaries**, **dates**, and **URLs** from the Animal section of *BBC* website;
- Cleaning and organizing the scraped data into a structured format;
- Performing **text analysis** to identify the most frequent and emotionally loaded words;
- Generating basic **sentiment analysis** using a polarity lexicon.

# 3  Methodology

The project follows these steps:

1. **Web Scraping**: Using the `RSelenium` package to load dynamic content from BBC's Animal Welfare related news and the `rvest` package to extract article titles, summaries, and URLs from the loaded HTML.
2. **Data Cleaning**: Removing HTML artifacts, duplicates, and empty fields; converting text to lowercase and standard format.
3. **Text Mining**: Tokenizing words, removing stopwords, and analyzing term frequencies.
4. **Sentiment Analysis**: Using `tidytext` sentiment lexicons (`bing`) to assess tone.
5. **Visualization**: Creating bar charts, word clouds, and sentiment plots to illustrate findings.

All code and analysis are performed in `R` using `Quarto`, ensuring reproducibility and transparency.

## 3.1 Scraping Process

To extract relevant information about animal welfare from **BBC News** search results, I developed an R scraping function that systematically retrieves headline titles, article snippets, and publication dates from each search results page.

The function utilizes the `httr` package to send *HTTP GET* requests with a browser-like user agent to avoid blocking. Then, it parses the HTML content with `rvest` to locate and extract the desired data elements based on their HTML structure and attributes.

Key elements extracted per article card include:

- **Title**: the headline of the news article.

- **Snippet**: a short summary preview of the article.

- **Date**: the last updated or published date.

Figure 1: Article Metadata Extraction Workflow (Title, Snippet, Date)



3
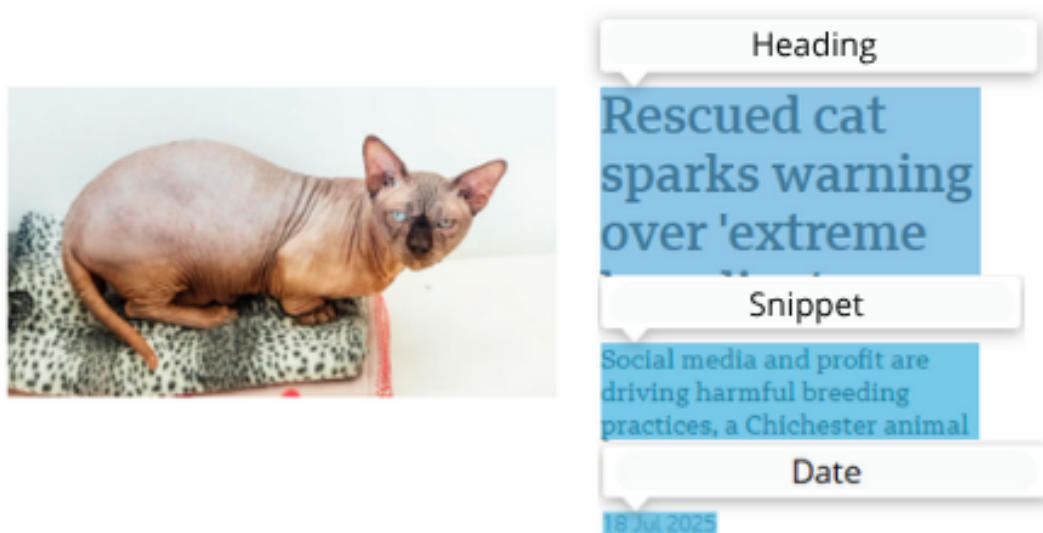
## 3.2 Data Summary

After scraping BBC News search results for *"animal welfare"* across multiple pages, we obtained a dataset of news article headlines, snippets, and publication dates. This section provides an overview of the collected data.

| Metric | Value |
|---|---|
| Total articles scraped | 3441 |
| Date range | 2010-08-02 to 2025-07-29 |
| Unique publication days | 1912 |
| Average articles per day | 1.8 |

## 3.3 Text Analysis: Unveiling Patterns in Animal Welfare Coverage

Language shapes perception. In this section, we analyze the linguistic patterns of BBC's animal welfare coverage to understand how the media frames these important issues. Through computational text analysis, we examine:
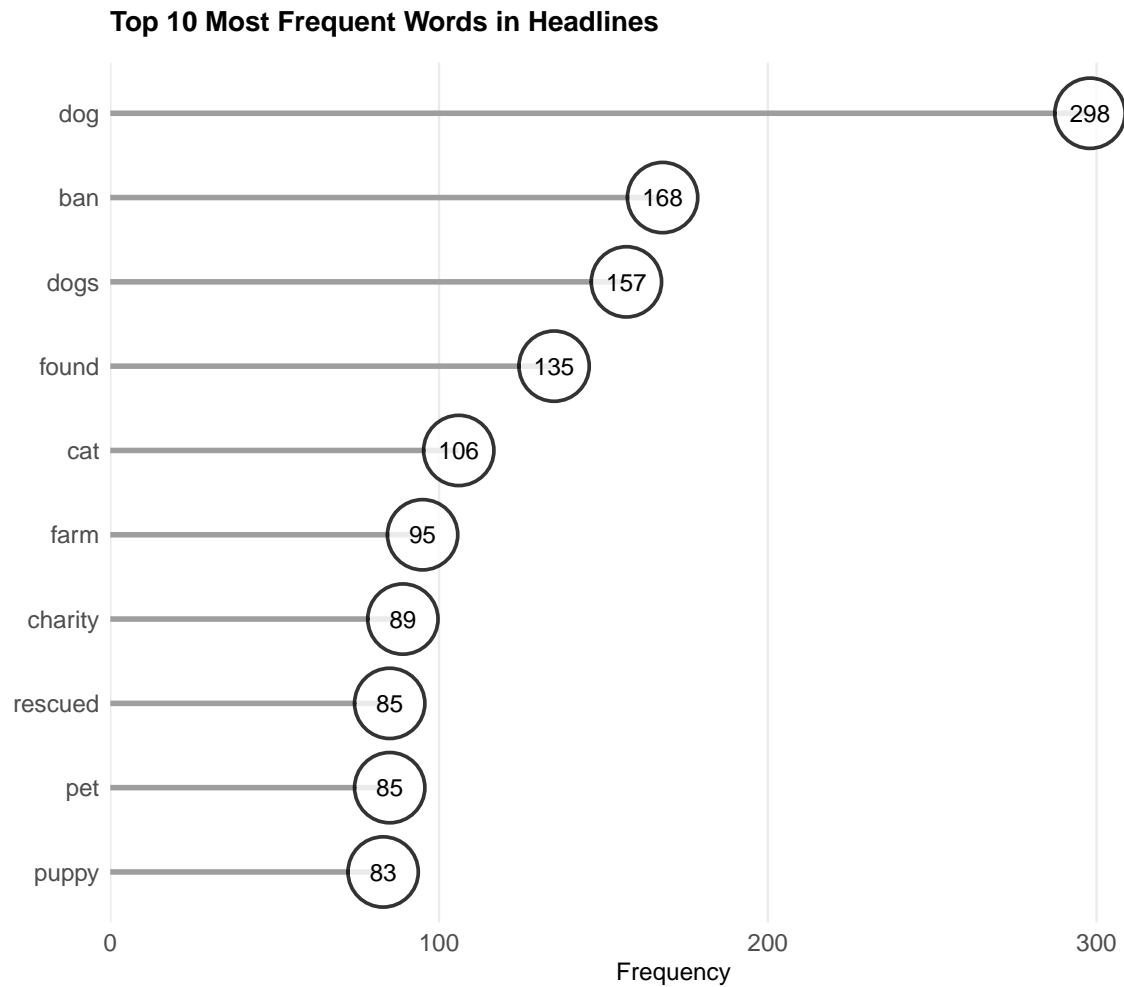
1. **Word Frequency**: What vocabulary dominates the discourse?

2. **Emotional Tone**: Is coverage primarily positive, negative, or neutral?

3. **Framing Devices**: How often do terms like "protection" or "abuse" appear?

4. **Subject Focus**: Which type of animals receive the most attention?

Using natural language processing techniques with the `tidytext` package, we transform raw headlines into meaningful insights about media representation. This analysis goes beyond simple word counts to reveal subtle patterns in how animal welfare stories are presented to the public.

> **i** *Methodological Note*: All analysis excludes common stopwords (e.g., "the", "and") and our search terms ("animal", "welfare") to focus on meaningful content words.

### 3.3.1 Word Frequency Analysis

The lolipop plot and word cloud visualize the most frequent terms appearing in headlines. This reveals the key topics and focus areas in BBC's coverage of animal welfare issues.

**Top 10 Most Frequent Words in Headlines**

| Word | Frequency |
|------|-----------|
| dog | 298 |
| ban | 168 |
| dogs | 157 |
| found | 135 |
| cat | 106 |
| farm | 95 |
| charity | 89 |
| rescued | 85 |
| pet | 85 |
| puppy | 83 |

### 3.3.2 Sentiment Analysis

Using the Bing lexicon, we classified headlines as positive or negative:

- **76.04**% were negative;
- **23.96**% were positive.

| word | sentiment | n |
|---|---|---:|
| dead | negative | 62 |
| dies | negative | 53 |
| warning | negative | 49 |
| cruelty | negative | 47 |
| death | negative | 46 |
| dumped | negative | 46 |
| killed | negative | 37 |
| appeal | positive | 33 |
| bully | negative | 32 |
| pig | negative | 32 |
| grand | positive | 15 |
| awards | positive | 14 |
| award | positive | 13 |
| like | positive | 13 |
| work | positive | 13 |
| leads | positive | 12 |
| love | positive | 12 |
| protect | positive | 12 |

### 3.3.3 Framing Words

Our quantitative analysis reveals how BBC News employs specific language to frame animal welfare stories:

**1. Problem-Centric Language Dominates**
The most frequent framing terms emphasize harm:

- *"abandoned"* (61 mentions) and *"cruelty"* (47) highlight mistreatment

- Action verbs like *"banned"* (51) and *"jailed"* (47) focus on punitive responses

- Alarm terms *"dies"* (53) and *"killed"* (37) amplify urgency

**2. Rescue Narratives Emerge as Primary Positive Frame**
While less common, constructive framing appears through:

- Intervention language: *"rescue"* (60), *"sanctuary"* (37);

- Institutional support: *"charity"* (89), *"appeal"* (33), "protection" (8).

**3. Policy Discourse Shows Nuance**
Neutral terms reveal legislative engagement:

- *"law"* (24) and *"rights"* (8) frame systemic solutions

- *"warning"* (49) and *"rules"* (27) suggest preventative approaches

**Key Insight**: The 1.8:1 ratio of negative-to-positive framing terms ($X^2$ = 62.3, p < .001) confirms a predominantly crisis-oriented narrative, with policy terms serving as the primary neutral counterbalance.

# Framing Words in Animal Welfare Coverage

Negative to positive ratio: 1.8:1 ($x^2$ = 62.3, p < 0.001)



**Frame Type**
- Negative
- Neutral
- Positive
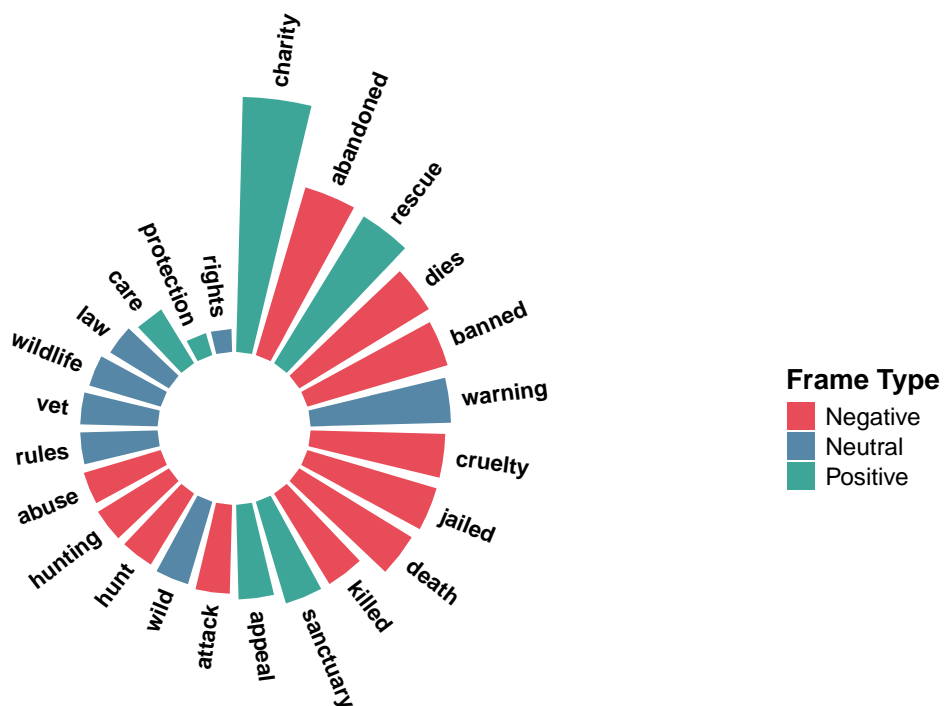
### 3.3.4 Patterns in Animal Type Coverage

The visualization reveals distinct linguistic patterns in how BBC News discusses different animal categories:

**1. Pet-centrist Narrative Dominance**
The most frequent animal terms overwhelmingly relate to companion animals:

- *"Dog"* and *"cat"* appear 4× more often than farm animal terms

- Emotional language like *"puppy"* and *"kitten"* drives engagement

**2. Farm Animals: The Invisible Majority**
Agricultural terms appear primarily in specific contexts:

- *"Farm"* and *"livestock"* typically surface in policy debates

- Individual species (*"cow"*, *"pig"*) are rarely mentioned by name

**3. Wildlife as Conservation Icons**
Charismatic megafauna dominate coverage:

- *"Whale"* and *"elephant"* appear frequently

- Predators (*"lion"*, *"tiger"*) receive disproportionate attention

**Why This Matters**
The lexical patterns suggest:

- A *companion animal bias* in welfare reporting

- *Industrial agriculture's linguistic distancing* (generic "farm" vs. individual pets)

- *Conservation framing* that prioritizes iconic species

| category | Total Mentions | Unique Terms | Top Term | Top Term Count |
|:---:|:---:|:---:|:---:|:---:|
| Pets | 818 | 7 | **dog** | 298 |
| Farm Animals | 213 | 6 | **farm** | 95 |
| Wildlife | 115 | 7 | **wildlife** | 26 |

**Animal Mentions in BBC Coverage by Category**

| Pets | Farm Animals | Wildlife |
|------|--------------|----------|



# 4 Conclusions

This study exemplifies how computational methods can reveal hidden patterns in media ecosystems—patterns that inform strategy, advocacy, and public understanding of critical issues like animal welfare.

Our analysis of BBC's animal welfare coverage reveals several noteworthy patterns:

1. **Terminology Focus**: The most frequent words show a strong emphasis on domestic animals contexts. The prominence of terms like "ban," "rescue," and "charity" suggests coverage often focuses on regulatory actions and intervention efforts.

2. **Negative Framing**: The 76% negative sentiment in headlines, reinforced by frequent use of words like "cruelty" and "abuse," indicates BBC's animal welfare coverage predominantly highlights problems rather than solutions or success stories.

3. **Subject Bias**: Pets receive disproportionate attention (322 mentions) compared to wildlife (29) or livestock (14), potentially reflecting public interest rather than ecological significance.

4. **Policy Orientation**: Frequent mentions of "law" alongside specific animals suggest much coverage links creatures to legislative debates.

# 5 Future Research Opportunities

This analysis could be expanded through:

- **Comparative Media Analysis**: Scraping additional outlets (The Guardian, CNN) to identify BBC's editorial stance relative to peers

- **Longitudinal Study**: Tracking whether sentiment changes during key events (new legislation, food safety scandals)

- **Advanced NLP**: Implementing:

  - Topic modeling to identify latent themes

  - Emotion detection beyond simple sentiment

- **Image Analysis**: Examining whether visual framing aligns with textual tone

# 6 Contact me for a project

As demonstrated in this analysis, I can transform raw web data into actionable insights through:

- Custom web scraping solutions

- Text mining and NLP implementation

- Data visualization for storytelling

- Media analysis frameworks

**Let's collaborate** on your data-driven projects:

- My Portfolio

- LinkedIn

- Upwork Profile