Technology Review

Team Wolf

Project Background

The goal of this project is to perform a statistical analysis on the impact of the Gray Wolf population in Wisconsin in order to support or refute commonly held concerns about the impacts of Gray Wolves

Questions of Interest:
1. How has the growth in the Wisconsin Gray Wolf population impacted the Wisconsin deer population?
2. Has the growing wolf population led to an increase in the number of interactions between people and wolves?
3. Time permitting, how have the lives and deaths of wolves changed over the reporting period?

3 project components - PDF data parser, data checker, statistical analyzer

Background and Use Case

Wisconsin Department of Natural Resources (DNR) publishes data on wolf and deer in the state in PDFs and does not provide access to the underlying data.

Transcribing the data from the PDFs creates opportunity for data errors and would inhibit others from expanding on the work by adding data from other years or other states

We plan to utilize a software package to extract the data tables from the PDFs programmatically and then analyze the data.

Package Options

tabula - py:

Python wrapper of tabula-java. A library for extracting tables from PDF files and converting them into a pandas DataFrame.

Author: Aki Ariga (tabula-py) and tabula-java (Manuel Aristaran)

Requires Java 8+ and Python 3.7+

MIT License

Camelot - py:

Camelot is a PDF table extraction library that was made to offer complete control over the extraction process.

Author: Dimiter Naydenov and Vinayak Mehta

Requires ghostscript and tkinter (for Windows)

MIT License

Tabula-py Testing

- Easy to install and run
- Able to export data in numerous formats, but defaults to a pandas DataFrame.
- Results will require substantial cleaning and validation:
  - Difficulty processing tables with multi-line headers or rows.

Table 1. Total miles tracked, average miles tracked per survey, and the number of surveys completed by wolf harvest zone and personnel type during the 2021-2022 winter tracking season.

Wolf Management Unit	Total Miles Tracked	Average Miles per Survey	Number of Surveys Completed
Zone 1	5,689	35	163
Zone 2	3,764	30	128
Zone 3	2,189	38	57
Zone 4	903	26	35
Zone 5	1,219	30	41
Zone 6	3,015	33	92
Volunteer	4,165	29	142
DNR	10,902	33	332
USDA	1,457	47	31
Tribal	129	26	5
Military	126	21	6
Total (Statewide)	16,779	32	516

Index	Unnamed: 0	Total Miles	Average Miles	Unnamed: 1	Number of Surveys	Unnamed: 2
0	Wolf Management Unit	nan	nan	nan	nan	nan
1	nan	Tracked	per Survey	nan	Completed	nan
2	Zone 1	5,689	nan	35	nan	163
3	Zone 2	3,764	nan	30	nan	128
4	Zone 3	2,189	nan	38	nan	57
5	Zone 4	903	nan	26	nan	35
6	Zone 5	1,219	nan	30	nan	41
7	Zone 6	3,015	nan	33	nan	92
8	Volunteer	4,165	nan	29	nan	142
9	DNR	10,902	nan	33	nan	332
10	USDA	1,457	nan	47	nan	31
11	Tribal	129	nan	26	nan	5
12	Military	126	nan	21	nan	6
13	Total (Statewide)	16,779	nan	32	nan	516

Camelot Testing

Pros

- Easy to use
- Multi-platform support
- Can read skewed and rotated tables

Cons

- Slow on large/complex PDF files
- Difficult download process depending on OS
- Less documentation

Table 6. Wolf depredation management in Wisconsin, 15 April 2021 to 14 April 2022.

	Wolf Harvest Zone						State
	1	2	3	4	5	6	Total
Livestock Cases							
Confirmed Depredation Incidents	28	2	6	1	5	8	50
Confirmed Threat Incidents	13		3	1		1	18
Chronic Farms Affected	18	1	4		1		24
Total Farms Affected	24	1	5	2	5	6	43
Cattle Killed	27		6	1	4	3	41
Cattle Injured	5						5
Captive Deer Killed		4					4
Captive Deer Injured							0
Sheep Killed	1					16	17
Sheep Injured							0
Goats Killed							0
Alpacas Killed							0
Alpacas Injured							0
Horses Killed	1				1	1	3
Horses Injured							0
Poultry Killed							0
Non-Livestock Cases							
Confirmed Depredation Incidents	17	1	3	0	1	0	22
Confirmed Threat Incidents	4	0	0	1	0	0	5
Dogs Killed While Actively Engaged in Hunting Activities	12		2		1		15
Dogs Injured While Actively Engaged in Hunting Activities	2		1				3
Dogs Killed While Not Engaged in Hunting Activities	5						5
Dogs Injured While Not Engaged in Hunting Activities			1				1

Final Decision

- Camelot
- Camelot's sample outputs provided better dataframes
- Speed is no issue (research project)
- Still potential issues with installing based upon OS
- Not perfect but will do

	A	B	C	D
1	0	1	2	3
2	Cause of E	Wolf Man	State	% of
3		1	Total	
4	Human			
5	Agency Cc	52b	65	70.70%
6	Vehicle Cc	2	15	16.30%
7	Illegally K	4a*	9	9.80%
8	Capture R	0	0	0.00%
9	Total	58		96.70%
10	Natural			
11	Disease /	0	0	0.00%
12	Intra-	1	2	2.20%
13	Euthaniz	0	0	0.00%
14	Total	1	2	2.20%
15	Unknown	0	1	1.10%
16	Total	59	92	100.00%

Cause of Death	Wolf Management Unit						State	% of Total
	1	2	3	4	5	6	Total	
Human Caused Mortality								
Agency Control	52b	0	9	0	2	2	65	70.7%
Vehicle Collision	2	4a	0	0	5	4	15	16.3%
Illegally Killed	4a*	1	2a	0	1	1	9	9.8%
Capture Related	0	0	0	0	0	0	0	0.0%
Total Human Caused	58	5	11	0	8	7	89	96.7%
Natural Mortality								
Disease / Injury	0	0	0	0	0	0	0	0.0%
Intra-specific Aggression	1	0	1	0	0	0	2	2.2%
Euthanized (non-control)	0	0	0	0	0	0	0	0.0%
Total Natural Causes	1	0	1	0	0	0	2	2.2%
Unknown Causes	0	0	0	0	1a	0	1	1.1%
Total Detected Mortality	59	5	13	0	9	7	92	100.0%