Datasheet for 'Wildlife Hunting' data[*]

Samantha Barfoot

April 23, 2024

This datasheet describes the motivations and composition of the moose and black bear hunting activity and harvest data sets obtained from the Ontario Data Catalogue. It also describes collection process, cleaning process, uses, distribution, and maintenance of the data sets.

Motivation

1. *For what purpose was the data set created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*

   - The data set was created to enable analysis of Ontario moose and black bear hunting activity. Specifically, the number of active hunters and harvests of both moose and black bears in each Wildlife Management Unit for each year. The data were created to better understand and help manage the moose and black bear population in Ontario with regard to hunting activities.

2. *Who created the data set and on behalf of which entity?*

   - The creator of the original data set are unknown but they were created on behalf of the Ministry of Resources and Forestry. However, the data sets have been updated and maintained by the Science and Research Branch, specifically by Allen Kwan, Bianca Sayan, Hao Qin, Emily Hill, Eric Giang, Stephanie Whyley, Ryder Leblanc, Cody Schacter, and Razieh Faraji since 2019.

3. *Who funded the creation of the dataset?*

   - The creation of the original data set was funded by the Government of Ontario.

Composition

1. *What do the instances that comprise the dataset represent? Are there multiple types of instances?*

---

[*]Code and data are available at: https://github.com/SamanthaBarfoot/wildlife_hunting.git

- The instances that comprise the moose data set represent the number of calf, bull, and cow harvests collected in a WMU within a certain year. Additionally, the number of active hunters in that WMU and year is also recorded. The instances that comprise the moose data set represent the number of black bears harvested and the number of active hunters in a WMU within a certain year.

2. *How many instances are there in total?*

   - In total, there are 1229 instances in the moose data set and 1056 instances in the black bear data set.

3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*

   - TBD

4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

   - For moose, each instance consists of the WMU, year, number of active hunters, number of each harvest type (cow, calf, or bull), and the total harvest. For black bears, each instance consists of the WMU, year, number of active hunters, and the total harvest.

5. *Is there a label or target associated with each instance?*

   - There is no label or target associated with each instance.

6. *Is any information missing from individual instances?*

   - There are no values missing from individual instances.

7. *Are there recommended data splits?*

   - There are no recommended data splits for moose or black bear data.

8. *Are there any errors, sources of noise, or redundancies in the data set?*

   - There are no errors, sources of noise, or redundancies in the data set

9. *Is the dataset self-contained, or does it link to or otherwise rely on external resources?*

   - The data sets can be obtained from the Ontario Data Catalogue website at the following links: (https://data.ontario.ca/dataset/moose-hunting-activity-and-harvests) and (https://data.ontario.ca/dataset/black-bear-hunting-activity-and-harvests). These can be downloaded for free by anyone and come with no restrictions and are availble as long as the Goverment of Ontario provides and maintains them.

10. *Does the dataset contain data that might be considered confidential?*

    - The data sets do not contain data that might be considered confidential and can be acessed by the general public.

11. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?*

    - The data sets contain information about animal hunting that certain people might find offensive but the data contains only numbers and years and not detailed information about hunting.

12. *Does the dataset identify any sub-populations?*

    - The moose data set identifies three sub populations, cow, calf, and bull harvests. Each instance contains harvests for each sub-population as well as a total harvest obtained from the sum of the other three harvests.

13. *Is it possible to identify individuals*

    - No it is not possible to identify individuals as the only data related to people is the number of people, there are no names or other identifying traits used. All the data is anonymous.

14. *Does the dataset contain data that might be considered sensitive in any way?*

    - Since the data is about hunting and the number of animals harvested each year some individuals may find this offensive due to their personal beliefs about hunting and hurting animals.

Collection process

1. *How was the data associated with each instance acquired?*

    - The data was acquired by hunters self reporting the harvests they collected. All licences hunters with tags must do this. Therefore, not all instances can be validated. The number of active hunters is based on how many licences hunters are registered with the Government of Ontario and can thus be validated.

2. *What mechanisms or procedures were used to collect the data?*

    - It is unknown what mechanisms or procedures were used to collect the data. However, since hunters self report the number of harvests it could be assumed that they submitted reports either by mail, phone, or online.

3. *Who was involved in the data collection process?*

    - It is unknown who exactly was involved in the data collection process, active hunters that reported their harvests remained anonymous as does the original creator of the data sets.

4. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances?*

   - The data was collected between 2006 and 2023 for moose data and 2012 and 2023 for black bear data. The data sets are also updated each year. These timeframes match the timeframe of the instances as the data.

5. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources?*

   - The data was collected from the Ontario Data Catalogue website.

6. *Were the individuals in question notified about the data collection?*

   - They were not notified as the data is freely accessible to the general public.

7. *Did the individuals in question consent to the collection and use of their data?*

   - The data is available for anyone to use and thus they did not need to be notified as it falls under the Open Government Licence.

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done?*

   - Cleaning of the data was done to group the data into three larger regions: Norhtern, Southeastern, and Southwestern Ontario. Harvest and active hunter data were also combined to find the total across all regions for each year. The WMU values were also cleaned so that they did not contain extra characters such as zeros and letters.

2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data?*

   - The raw data was saved and can be found here: https://github.com/SamanthaBarfoot/wildlife_hunti

3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*

   - R studio was used to clean the data, cleaning scripts can be found here: https://github.com/SamanthaBarfoot/wildlife_hunting/tree/main/scripts

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*

   - No other tasks that used this data set have been found but some possible tasks that might have used it include other papers and news articles as well as government reports.

2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*

- No there is not a repository that links all paper or systems that use the data set.

3. *What (other) tasks could the dataset be used for?*

   - The data set could also be used for research papers that analyze bear and moose hunting trends and population trends. Additionally, they could be used to inform news articles and other government reports about hunting activity.

4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? Is there anything a dataset consumer could do to mitigate these risks or harms?*

   - All data from the data set is anonymous and only contains counts, no sensitive information. Therefore, there is not anything that might impact future users.

5. *Are there tasks for which the dataset should not be used?*

   - No, as long as the data is not being used to harm someone.

Distribution

1. *Will the data set be distributed to third parties outside of the entity on behalf of which the data set was created?*

   - While the data was created by the Ontario Government for monitoring hunting activity, the data was shared publicly online for free and thus third parties can use it.

2. *How will the data set be distributed? Does the dataset have a digital object identifier (DOI)?*

   - The data set was distributed directly on the Ontario Data Catalogue Website as a csv file. There is no DOI

3. *When was the data set distributed?*

   - The data sets were first distributed 31-12-2019.

4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? Provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*

   - The data sets are licensed under an Open Government Licence - Ontario (https://www.ontario.ca/page/open-government-licence-ontario)

5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances?*

   - No third parties imposed IP based or other restrictions on the data associated with the instances

6. *Do any export controls or other regulatory restrictions apply to the data set or to individual instances?*

   - The only restriction is that the source must be acknowledged and you cannot acess personal information (who the hunters are). (https://www.ontario.ca/page/open-government-licence-ontario)

Maintenance

1. *Who will be supporting/hosting/maintaining the data set?*

   - The Science and Research Branch of the Ministry of Natural resources and Forestry maintains the data.

2. *How can the owner/curator/manager of the data set be contacted?*

   - They can be contacted online by a form or by email.

3. *Will the data set be updated? How often, by whom, and how updates will be communicated to data set consumers?*

   - The data is updated every year after each hunting season with updates being put on the Open Data Catalogue website. Each update is recorded so users know how up to date the data is.

4. *If the data set relates to people, are there applicable limits on the retention of the data associated with the instances?*

   - The data set does not directly relate to people, only counts of people and thus there are not limits to retention of data as there is no personal information included.

5. *Will older versions of the data set continue to be supported/hosted/maintained?*

   - Older versions of the data set are not maintained after new updates are released.

6. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?*

   - There is no mechanism for others to contribute to the data set as it is maintained by the Government of Ontario and having outside parties contribute to it could reduce the reliability of the data.