6.1: Sourcing Open Data

1- Data Source

1.1 Summary of data source and why I've chosen this dataset.

Since 1990, "wildland fires" across Canada have consumed an **average of 2.5 million hectares** a year. These fires occur in forests, shrub lands and grasslands. Some are uncontrolled wildfires started by lightning or human carelessness. A small number are prescribed fires set by authorized forest managers to mimic natural fire processes that renew and maintain healthy ecosystems.

Wildland fires present a **challenge for forest management** because they have the potential to be at once harmful and beneficial. They can threaten communities and destroy vast amounts of timber resources, resulting in costly losses. However, wildland fires are a natural part of the forest ecosystem and important in many parts of Canada for maintaining the health and diversity of the forest. In this way, prescribed fires offer a valuable resource management tool for enhancing ecological conditions and eliminating excessive fuel build-up. Not all wildland fires should (or can) be controlled. Forest agencies work to harness the force of natural fire to take advantage of its ecological benefits while at the same time limiting its potential damage and costs [1].

This makes **fire control strategies a vital component of forest management and emergency management in Canada.** Increasingly accurate assessments of the fire situation across Canada are now helping land managers use forest science to reduce fire risk and optimize the benefits.

Facts about wildland fires in Canada

- Canada has about 9% of the world's forests. Each year over the last 25 years, about 7,300 forest fires have occurred. The total area burned varies widely from year to year, but averages about 2.5 million hectares annually.
- Only 3% of all wildland fires that start each year in Canada grow to more than 200 hectares in area. However, these fires account for 97% of the total area burned across the country.
- Fire suppression costs over the last decade in Canada have ranged from about $800 million to $1.5 billion a year [1].

Citation: [Canadian Wildland Fire Information System Datamart](#) , Last update: 12/31/2021

Available [on Kaggle](#)

1.2 Data Sourcing: Wildfires in Canada (1950-2021)

The CFS[1] has been involved in fire research for decades. The CFS works with partners across the country to increase the knowledge base about wildland fires, and to improve the ability of authorities to predict and manage risks and benefits.

---

[1] Canadian Forest Service

The dataset is an **external** but a **reliable** data source. We will use the dataset provided by **Natural Resources Canada** which is available on **Kaggle** to download as an open dataset. As government data we can verify this as a **trustworthy** data source.

1.3 Wildfires: Data Collection

CNFDB[2] fire point data is a collection of forest fire locations as provided by Canadian fire management agencies including provinces, territories, and Parks Canada. In fact, it is a collection of forest fire data from various sources, captured from a satellite image with high infrared intensity. These data include fire locations (point data) and fire perimeters (polygon data). To create the Canada-wide product, the data collected from each agency are projected into a common format and combined with data from other agencies; attribute fields are standardized; agency specific attribute fields are removed; and polygon areas are calculated using GIS[3] [2].

We can assume a **complete** and **accurate** data collection from government (as it has been done over multiple decades). The one caveat, however, is that not all fires can be identified from satellite imagery, either because the fires are too small or because cloud cover obscures the satellite's view of the ground.

1.4 Wildfires: Data Contents

Data includes the annual wildfires' specifications happened in Canada from 1950 to 2021.

1.5 Wildfires: Data Relevance

Objective: Determine the potential fire spread rate in each location.

We can use the historical data and trends to predict the wildfire behavior for upcoming year and prioritize the locations to have an informed forest and emergency management.

1.6 Wildfires: Data Limitations and Ethics

Measurement bias might occur since:

- The data contained in the CNFDB are not complete nor are they without error. Not all fires have been mapped, and data accuracy varies due to different mapping techniques.
- This collection includes only data that has been contributed by the agencies. Data completeness and quality vary among agencies and between years [2].
- Not all fires can be identified from satellite imagery, either because the fires are too small or because cloud cover obscures the satellite's view of the ground.
- There is currently no official national standard that has been applied to Protection Zone attribute. This caused too many blank values in PROTZONE column.

---

[2] The Canadian National Fire Database
[3] Geographic information system (GIS) is a system that connects to a map, and creates, manages, analyzes, and maps all types of data.

2. Data Understanding

Column	Description	Data Type	Time Variant
FID	Fire ID	Quantitative	No
SRC_AGENCY	Agency (province, territory, parks) from which the fire data has been obtained. *Provinces and territories*: **BC** - British Columbia, **AB** – Alberta, **SK** – Saskatchewan, **MB** – Manitoba, **ON** – Ontario, **QC** – Quebec, **NS** - Nova Scotia **NB** - New Brunswick, **NL** - Newfoundland & Labrador, **YT** – Yukon, **NWT** - Northwest territories, **PC** - Parks Canada *National Parks*: **PC-BA** - Banff (National Park) **PC-BP** - Bruce Peninsula (National Park) **PC-BT** - Batoche (National Park) **PC-CB** - Cape Breton (National Park) **PC-CH** - Cypress Hills (National Park) **PC-CT** - Chilkoot Trail (National Park) **PC-EI** - Elk Island (National Park) **PC-FO** - Forillon (National Park) **PC-FR** - Fort Rodd NHS (National Park) **PC-FU** - Fundy (National Park) **PC-FW** - Fort Walsh (National Park) **PC-GB** - Georgian Bay Is. (National Park) **PC-GF** - Gulf Islands (National Park) **PC-GH** - Gwaii Haanas (National Park) **PC-GI** - Grosse-ile (National Park) **PC-GL** - Glacier (National Park) **PC-GM** - Gros Morne (National Park) **PC-GR** - Grasslands (National Park) **PC-JA** - Jasper (National Park) **PC-KE** - Kejimkujik (National Park) **PC-KG** - Kouchibouguac (National Park) **PC-KL** - Kluane (National Park) **PC-KO** - Kootenay (National Park) **PC-LL** - Lake Louise (National Park) **PC-LM** - La Mauricie (National Park) **PC-LO** - Louisbourg (National Park) **PC-MI** - Mingan (National Park) **PC-NA** - Nahanni (National Park) **PC-PA** - Prince Albert (National Park) **PC-PE** - Prince Edward Island (National Park) **PC-PP** - Point Pelee (National Park) **PC-PR** - Pacific Rim (National Park) **PC-PU** - Pukaskwa (National Park) **PC-RE** - Mount Revelstoke (National Park) **PC-RM** - Riding Mountain (National Park) **PC-RO** - Rocky Mountain House (National Park) **PC-SL** - St. Lawrence Islands (National Park) **PC-SY** - Saoyú-ʔehdacho (National Historic site) **PC-TI** - Thousand Islands (National Park) **PC-TN** - Terra Nova (National Park) **PC-VU** - Vuntut (National Park) **PC-WB** - Wood Buffalo (National Park) **PC-WL** - Waterton Lakes (National Park) **PC-WP** - Wapusk (National Park) **PC-YO** - Yoho (National Park)	Nominal	No
FIRENAME	Agency fire name	Nominal	No
LATITUDE	Latitude	Quantitative	No
LONGITUDE	Longitude	Quantitative	No
REP_DATE	Date of fire reported by individual agencies	Quantitative	Yes
SIZE_HA	Fire size in hectares	Quantitative	No
CAUSE	Cause of fire **U**: Unknown cause **L**: Lightning caused fire **H**: Human caused fire **H-PB**: Prescribed burn (human caused) **Re**: Reburn	Nominal	No
PROTZONE	Protection Zone as indicated by source agency. There is currently no official national standard that has been applied to this attribute.	Nominal	No
ECOZ_NAME	Eco district associated with fire point location.	Nominal	No

3. Data Cleaning

- Removed duplicates (3230 records)
- Removed missing values (~ 1% of data).
- Revised mixed type columns.
- Renamed columns to keep them clear and consistent.
- Excluding data before 1950.

Wrangled data frame shape: 418867 rows x 9 Columns

4. Key Questions

We would like to help CFS by analyzing and studying wildland fire behaviour including:

- How fuel ignites, flame develops and fire spreads.
- Assessing current fire activity by monitoring forest characteristics (e.g., cause, location), keeping track of current fires, and evaluating the risk of new fires starting.
- Forecasting potential fire spread rate for future years.
- Checking other variables like weather conditions (wind, temperature) to see if there is any relationship between those conditions and wildfire happened in a specific location (subject to the availability of data for weather conditions)

5. Resources

[1] Natural Resources Canada
[2] Canadian Wildland Fire Information System | Canadian National Fire Database (CNFDB) (nrcan.gc.ca)