Introduction

Description of your dataset
The Canadian National Fire Database (CNFDB) is a comprehensive collection of forest fire locations and perimeters compiled from Canadian fire management agencies across provinces, territories, and Parks Canada. The database is standardized into a common format for Canada-wide analysis, though it contains only fires that agencies have chosen to report and map. The dataset contains information pertaining to forest fires from 1930-2023, for the purposes of our analysis we are focusing on years 2000-2023.

Including the source, how/why the data was gathered/generated
The CNFDB is maintained through a collaborative effort involving all Canadian fire management agencies, with data compilation coordinated at the national level (accessible through https://cwfis.cfs.nrcan.gc.ca). Fire data is collected by individual provincial, territorial, and Parks Canada agencies, including field personnel, fire crews, pilots, and photo interpreters document fire locations and perimeters during and after fire events. This raw data is then processed by digitizers and analysts before being standardized and integrated into the national database. The database serves multiple critical purposes including national-scale mapping, statistical and spatial analysis, and research on natural disturbance patterns in Canadian forests. The compilation was partially funded by several Canadian government programs focused on energy, climate change, and environmental research (ENFOR, Program on Energy Research and Development, Climate Change Action Fund, and Action Plan 2000), reflecting its importance for understanding fire patterns in relation to broader environmental and resource management concerns.

Introduce your team members, their backgrounds and their interests in the dataset

- Xinghao Huang: background in statistics and data science, interested in the severity of recent fires across Canada, and how the overall fire severity varies by different causes.

Intended Audience: Who is your intended audience and what do you expect they will get from this project (what is the motivation)?

Our intended audience is the general Canadian public. These are people who are curious about wildfires in their country or even concerned citizens wanting to understand fire patterns in their region. We expect this project will help Canadians visualize and understand where and when major wildfires have occurred across the country, turning complex national fire data into meaningful and interpretable plots. By presenting this information in an engaging, interactive format, we aim to increase public awareness of Canada's fire history and patterns. This understanding is particularly relevant given increased public attention to wildfire seasons and their impact on communities, air quality, and ecosystems.

About the Data

Data Abstraction

Attribute Name	**Attribute Type**	**Data Semantics**	**Cardinality**
NFDBFIREID	Nominal	Constructed ID combining province-year-fire_id	436,564
SRC_AGENCY	Nominal	Province/Territory/parks code	13 (BC, AB, SK, MB, ON, QC, NS, NB, NL, YT, NT, PEI, PC
NAT_PARK	Nominal	National park identifier (if applicable)	48 unique values
FIRE_ID Recurring fire?	Ordinal	Constructed ID of year-fire_id. Fire identifier	329,627 unique values
FIRENAME	Nominal	Name or location description of fire (if applicable)	28,913 unique values
YEAR	Quantitative/Temporal	Calendar year of fire occurrence	80 unique years [1950-2024]
MONTH	Quantitative	Month of fire occurrence	12 [1-12]
DAY	Quantitative	Day of fire occurrence	~31 [1-31]
LATITUDE	Quantitative	Geographic coordinate (North-South)	Continuous [-16.03 - 90.00]
LONGITUDE	Quantitative	Geographic coordinate (East-West)	Continuous [-180.00 - 180.00]
SIZE_HA	Quantitative	Fire size in hectares	[0.001-44.5+]
CAUSE	Nominal	Specific fire cause category	5 unique values [H, H-PB, N, RE, U] H = Human H-PB = Human Prescribed Burn N = Natural RE = U = Undetermined
FIRE_TYPE	Ordinal ? Nominal	Type of fire event	27 unique values [Fire, IFR, PB, 4, 5,

	• Different from severity or no? If includes severity then ordinal		Grass, Forest, Other, Wildfire, Mutual Aid, Dump, 99, Type 3, Type 5, Type 4, Type 1, Type 2, Ground, Surface, Crown, Req For Assist, Request for Assist, 1, 2, 3, Prescribed Burn, OFR]
RESPONSE	Nominal/Ordinal??? • Might be ordinal as response can escalate in priority	Response type to fire	10 unique values [FUL, MON, MOD, MNP, MDP, No Action, Actioned, Limited Action, SUP, PRO] - maybe we can split these into another ordinal column??
PROTZONE	Nominal	Protection zone classification	37 unique values
PRESCRIBED Filtered version of FIRE_TYPE	Ordinal	Whether fire was a prescribed burn	3 unique values [N/A, No, PB]
REP_DATE	Temporal	Date fire was reported (if applicable)	18,212 unique values
OUT_DATE	Temporal	Date fire was declared out	12,393 unique values
ATTK_DATE	Temporal	Date fire suppression began	60 unique values?

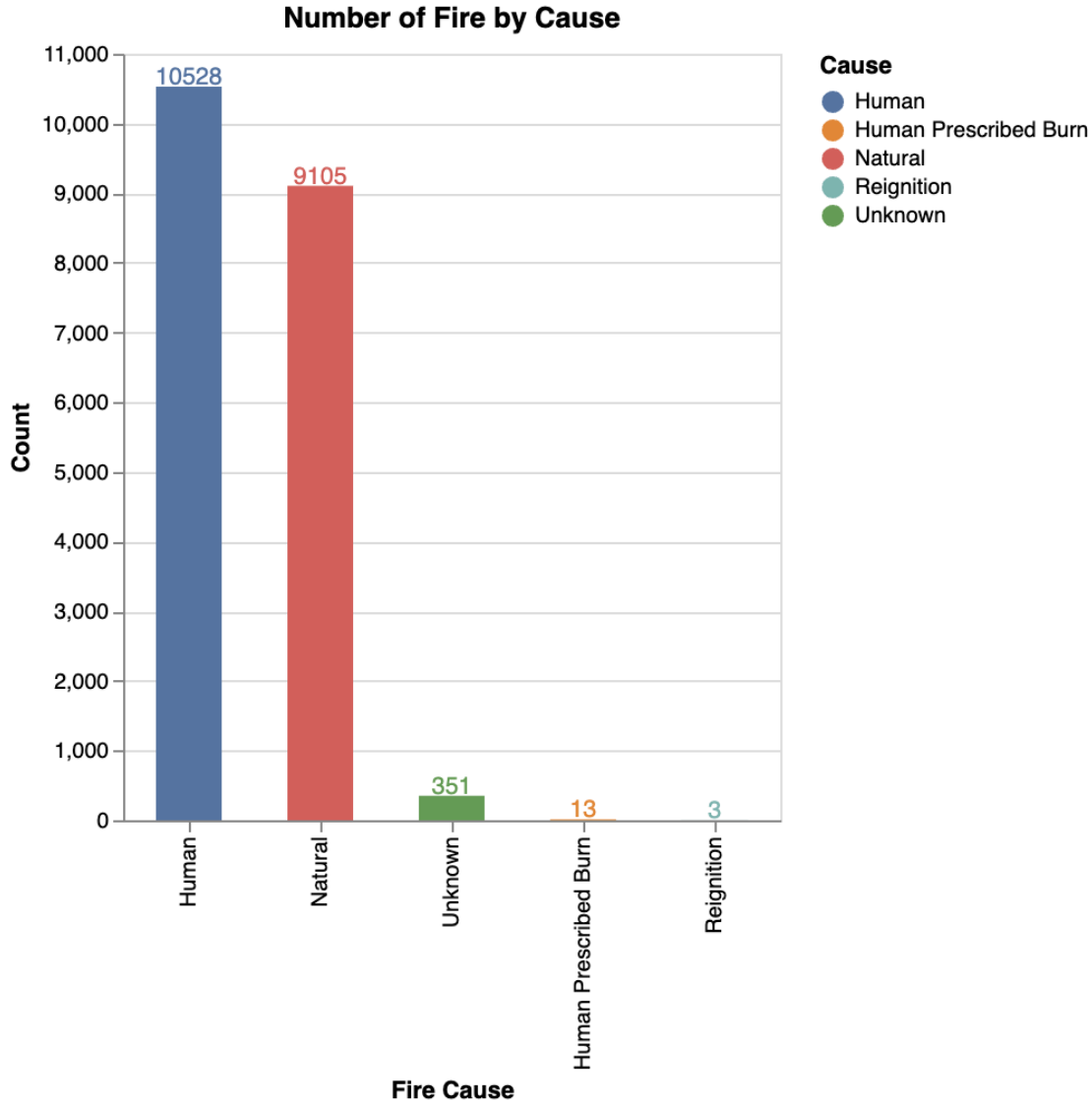
Note: our original dataset contained over 400,000 rows. After discussing this with Dr. K she advised us to take a random sample of 20,000 rows for our analysis. We all used the same sample when performing our EDA and will continue to use it throughout the rest of the project.

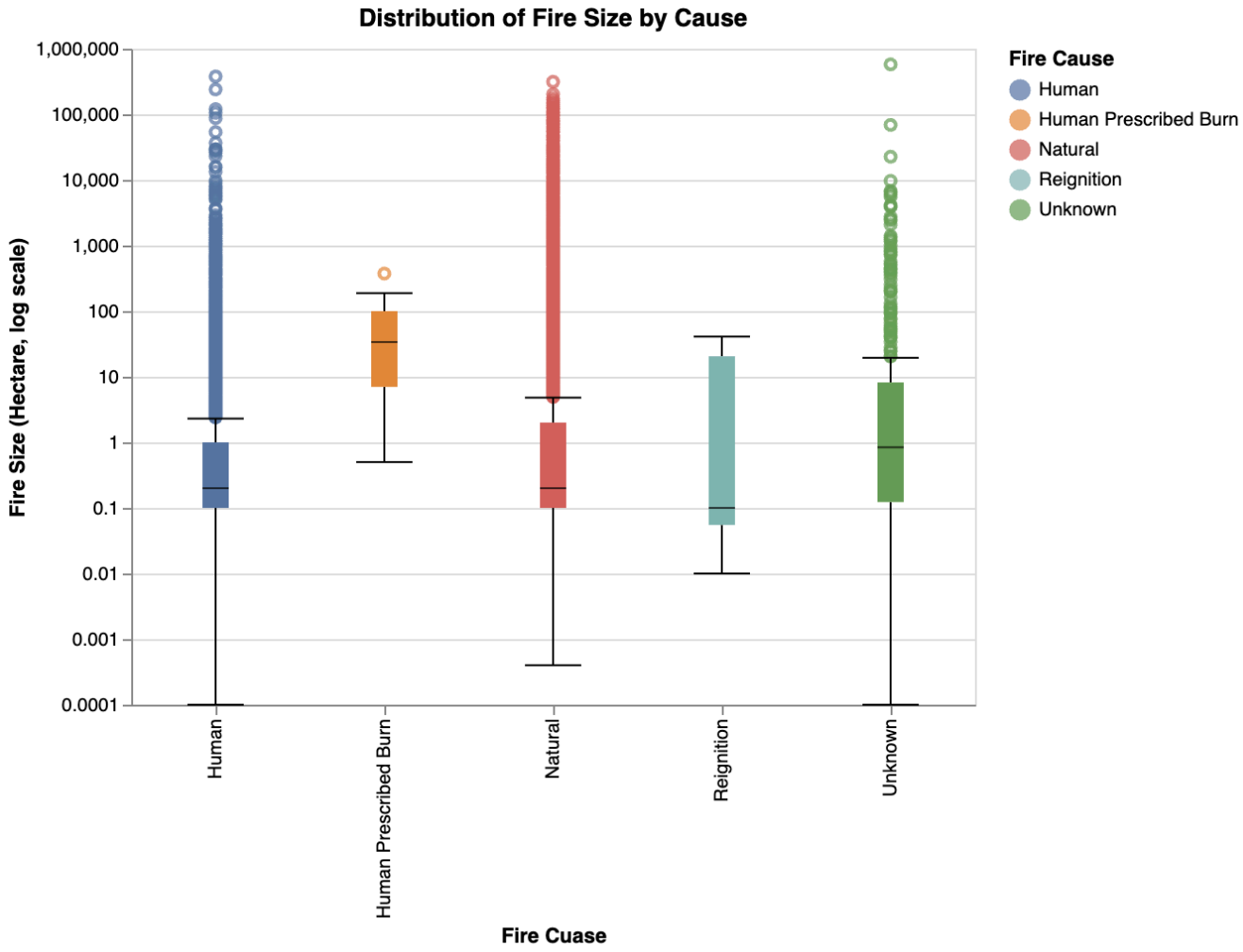
Exploratory Data Analysis

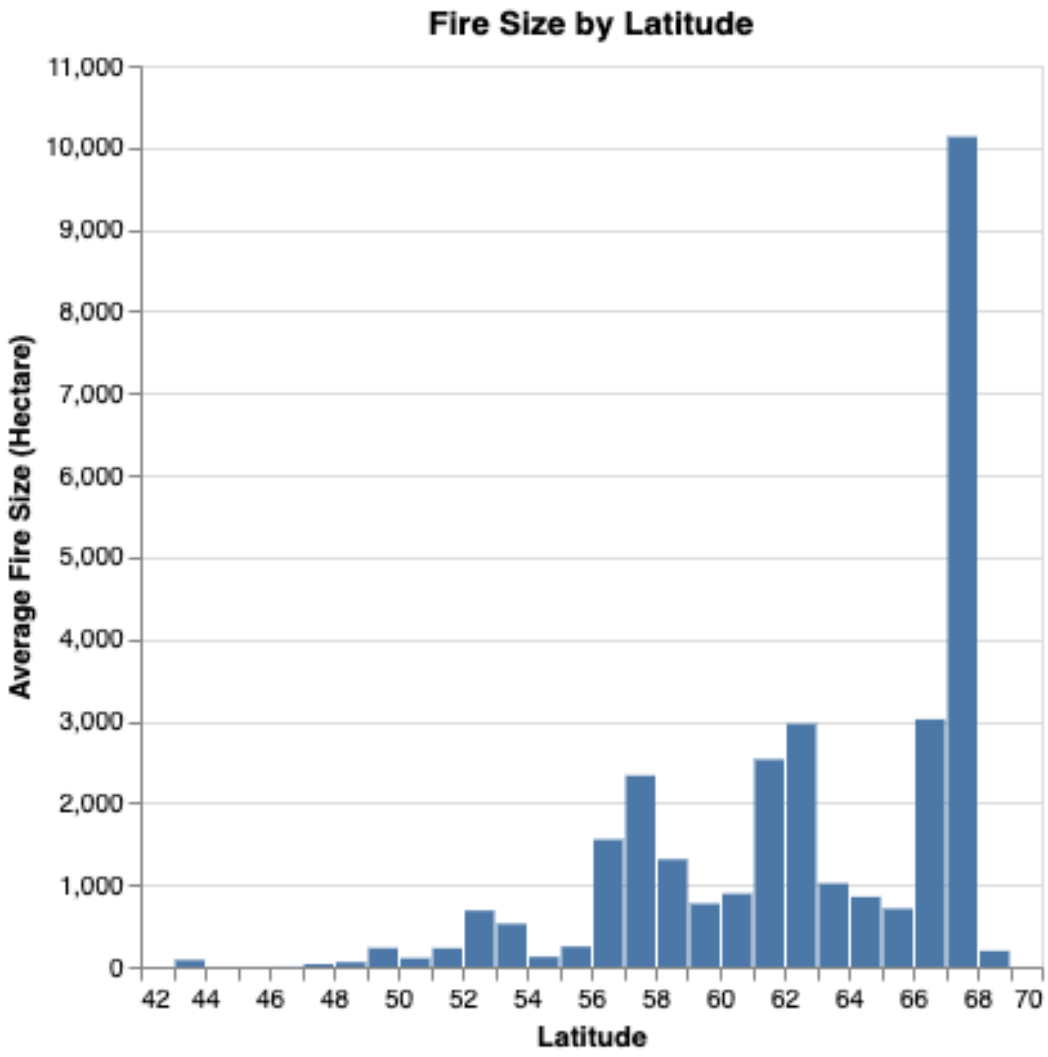
Xinghao Huang EDA Plots:

1.What is the frequency for each fire cause?

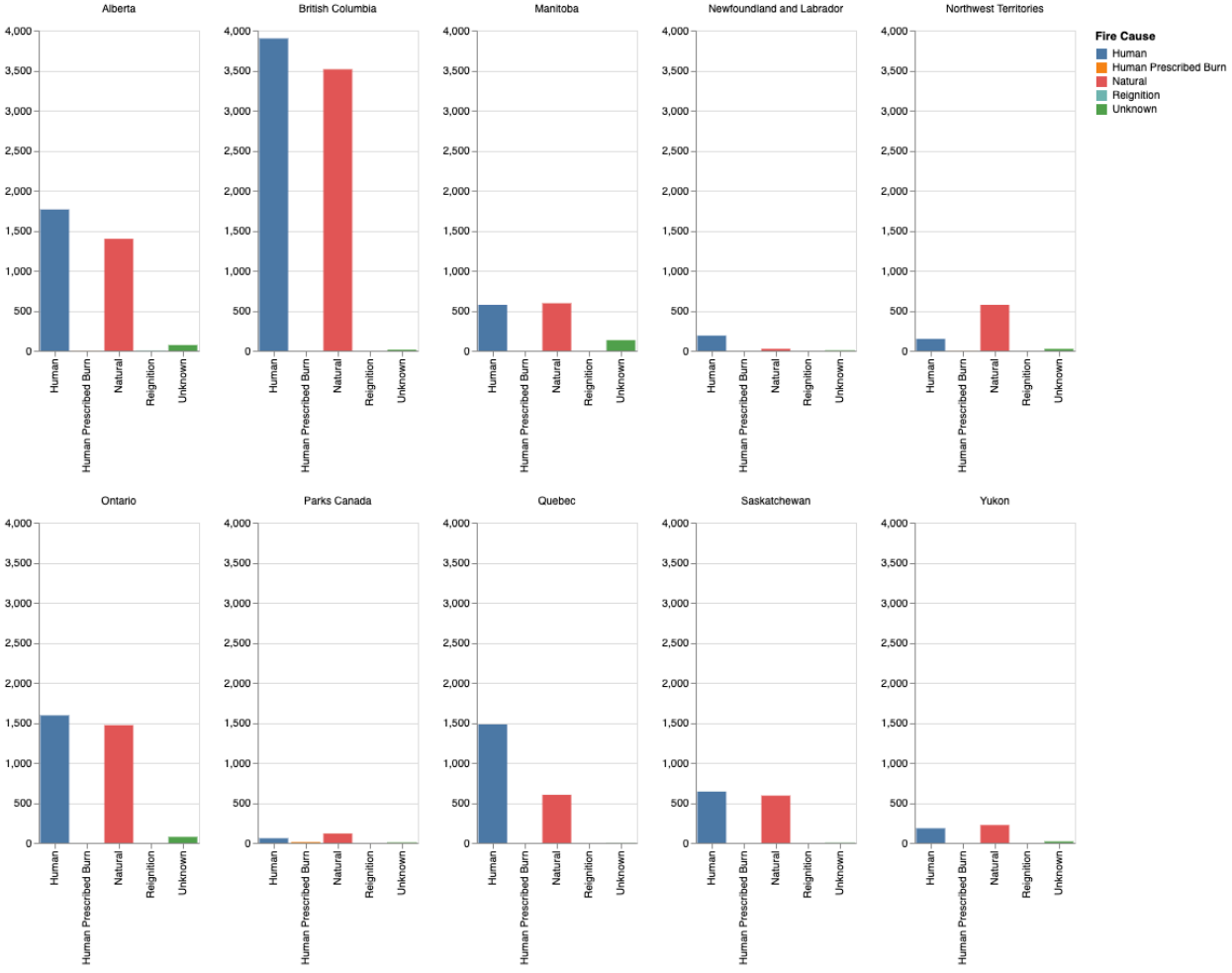
2.How does the fire severity vary by different fire causes?
3.How does fire size vary by latitude?
4.How does the frequency of each fire cause vary across provinces?
5.How does the average fire size differ across provinces?

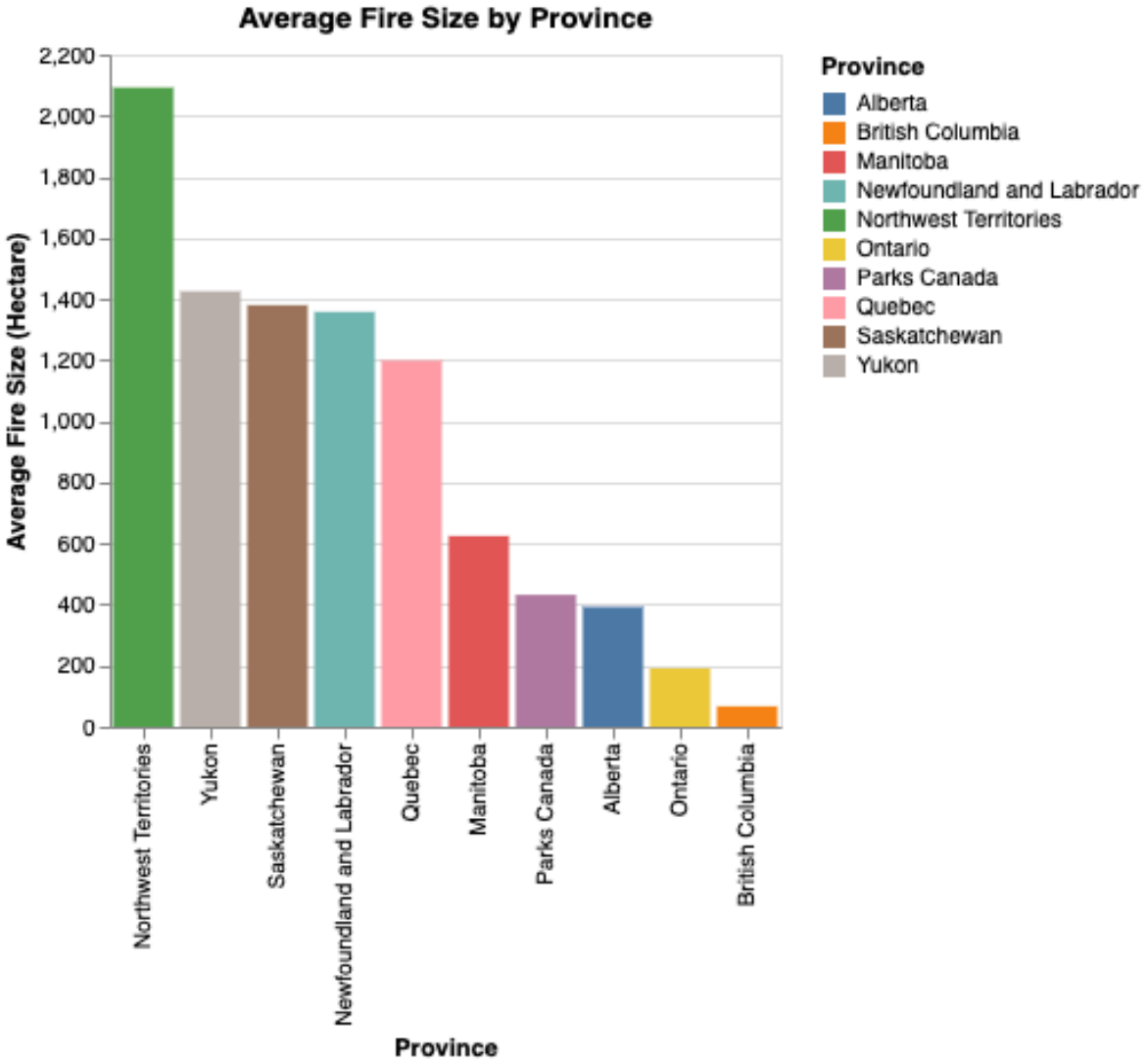



Distribution of Fire Size by Cause
Fire Size (Hectare, log scale)
1,000,000
100,000
10,000
1,000
100
10
1
0.1
0.01
0.001
0.0001
Human
Human Prescribed Burn
Natural
Reignition
Unknown
Fire Cuase
Fire Cause
Human
Human Prescribed Burn
Natural
Reignition
Unknown


Fire Size by Latitude
Average Fire Size (Hectare)
11,000
10,000
9,000
8,000
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
42
44
46
48
50
52
54
56
58
60
62
64
66
68
70
Latitude

Regional Frequency of Fire Cause





My EDA focused on understanding the general patterns in fire causes and fire sizes to find out relationships related to fire severity. In our sample dataset with 20,000 observations, I found that approximately 98% of all fire events were caused by human activity and nature factors, with 10,528 cases due to humans and 9,105 to natural causes. The remaining events were linked to human prescribed burns, reignition, or undetermined causes. In terms of severity, human prescribed burns had the largest median fire size of about 33.8 hectares, but with a relatively narrow spread, indicating that these fires were intentional and controlled. Other causes had smaller median fire sizes (below 1 hectare), although several extreme outliers were observed among human-caused, natural-caused, and unknown fires, implying the occurrence of a few exceptionally large fire events. For example, the largest human-caused fire was 378,328 hectares, an unusually extreme case. When analyzing geographic patterns, I initially assumed that regions with higher latitude would experience greater fire risks due to climatic conditions. The data shows that average fire size increased with latitude, suggesting that northern regions tend to experience larger fires. On the other hand, regional comparisons revealed an interesting contrast. British Columbia had the highest number of fire events (around 7,000), which was far more than the other provinces. However, it recorded smaller average fire sizes, making it

frequent but less severe compared to regions like the Northwest Territories or Yukon, where fires were less frequent but much larger on average. This may be due to outliers of fire events occurring in these regions, which largely affected the average value.

Focus of Inquiry and Preliminary Sketches

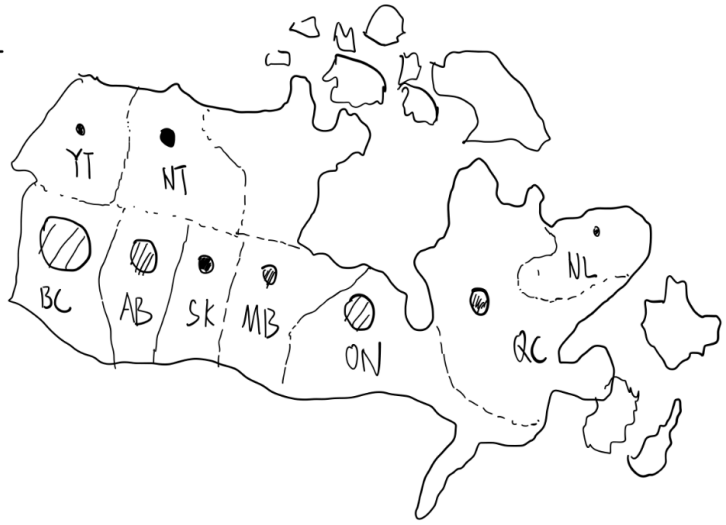
Xinghao Inquiry Theme: "**Understanding Fire Severity Across Canada**"

(Using CAUSE, SIZE_HA, SRC_AGENCY, LATITUDE, temporal attributes will be used for the future stages, e.g., when including interactions)

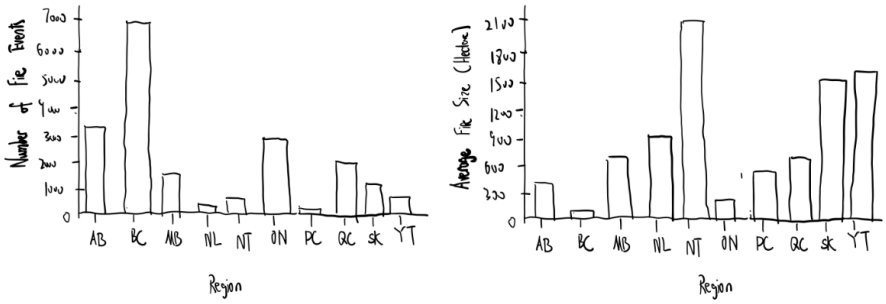
Analytic Question 1: "How does fire severity vary across regions in Canada, and what is the relationship between the frequency of fire events and their severity?"

a. **Sketch Ideas:**
    - A map of Canada where each province/territory is represented by a point whose size indicates the number of fire events and whose color saturation reflects the average fire severity.
    - A horizontal concatenated chart with fire frequency by region on the left, and fire severity by region on the right.
    - A radical chart with radius represents the frequency of fire events and color represents the fire severity. Each province/territory gets one wedge.
b. **Analytic Tasks:**
    - **Characterize Distribution:** How does fire severity distribute across different regions of Canada?
    - **Compare:** Compare fire severity across different regions.
    - **Correlate:** Examine the relationship between fire event frequency and fire severity.

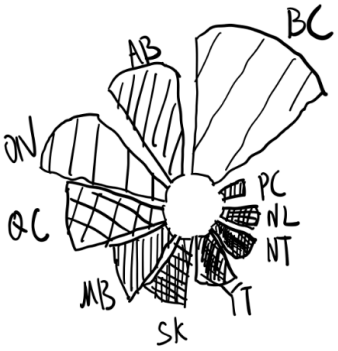
Sketch 1



Sketch 2



Sketch 3



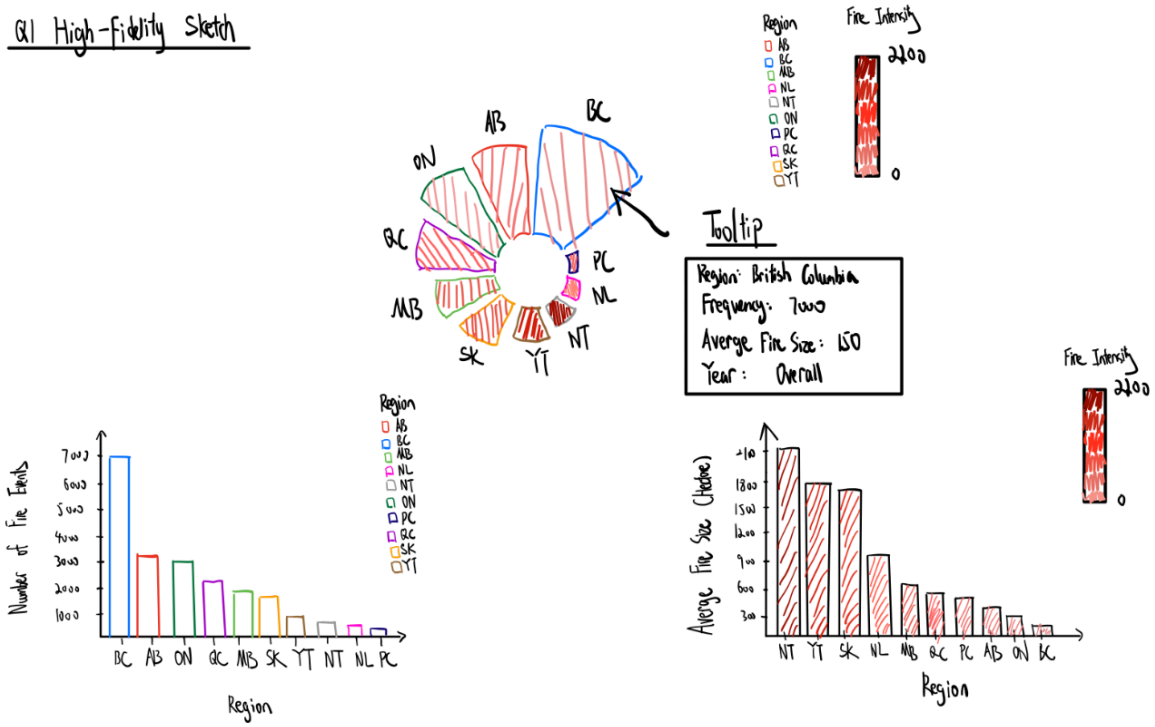
c. Critique of the plots:

1. The map enhances engagement and allows viewers to explore fire data across regions. However, using point size to encode the frequency of fire events and color intensity to represent average fire size may create visual confusion, since both encoding channels compete for attention. Regions with fewer fire events but relatively large average fire sizes could be overlooked because their points appear very small, making severe cases less noticeable at first glance.
2. The second sketch uses a more conventional and intuitive approach by separating the number of fire events and fire size into two bar charts, making

each variable easy to interpret. However, because the two charts have independent scales, it becomes difficult to connect frequency and severity at once. The viewers must compare values back and forth between plots, which reduces efficiency in identifying regions that experience both frequent and severe fires.

3. The third sketch combines the focus of each previous plot by showing both fire frequency and severity in one radial chart, making it visually engaging and informative. But the circular layout can cause visual confusion and make it difficult to compare regions accurately.

d. Finalized sketch:



Analytic Question 2: "How do different causes of fires influence the severity and scale of fire events across Canada?"

a. Sketch Ideas:

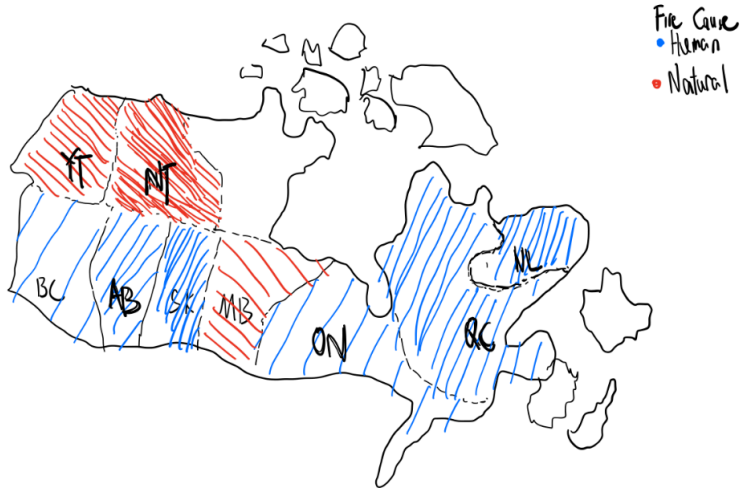
- A choropleth map with each region shaded based on the most frequent fire cause with different intensity indicating the severity.
- A strip plot with fire size on the x and fire cause on the y.
- A stacked bar chart with province/territory on the x and number of fire events on the y stacked by different fire cause with color intensity represents the severity

b. Analytic Tasks:

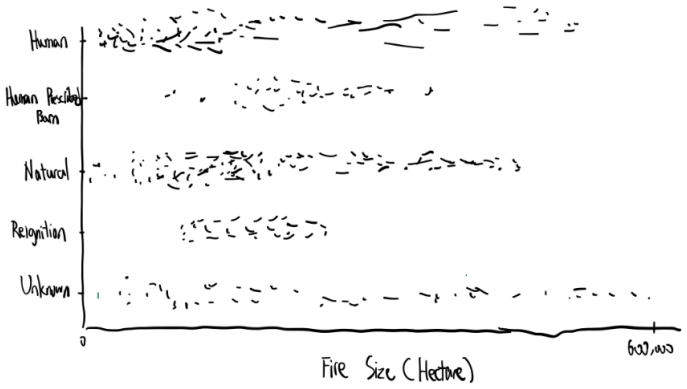
- **Compare:** Which fire causes tend to produce more severe fires?

- **Correlate:** Examine the relationship between fire cause and fire severity.
- **Cluster:** Identify groups of fire causes or regions based on their severity and frequency.

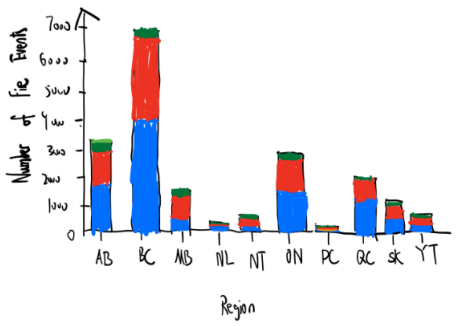
Sketch 1

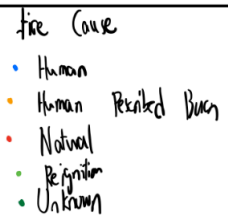


Sketch 2



Sketch 3

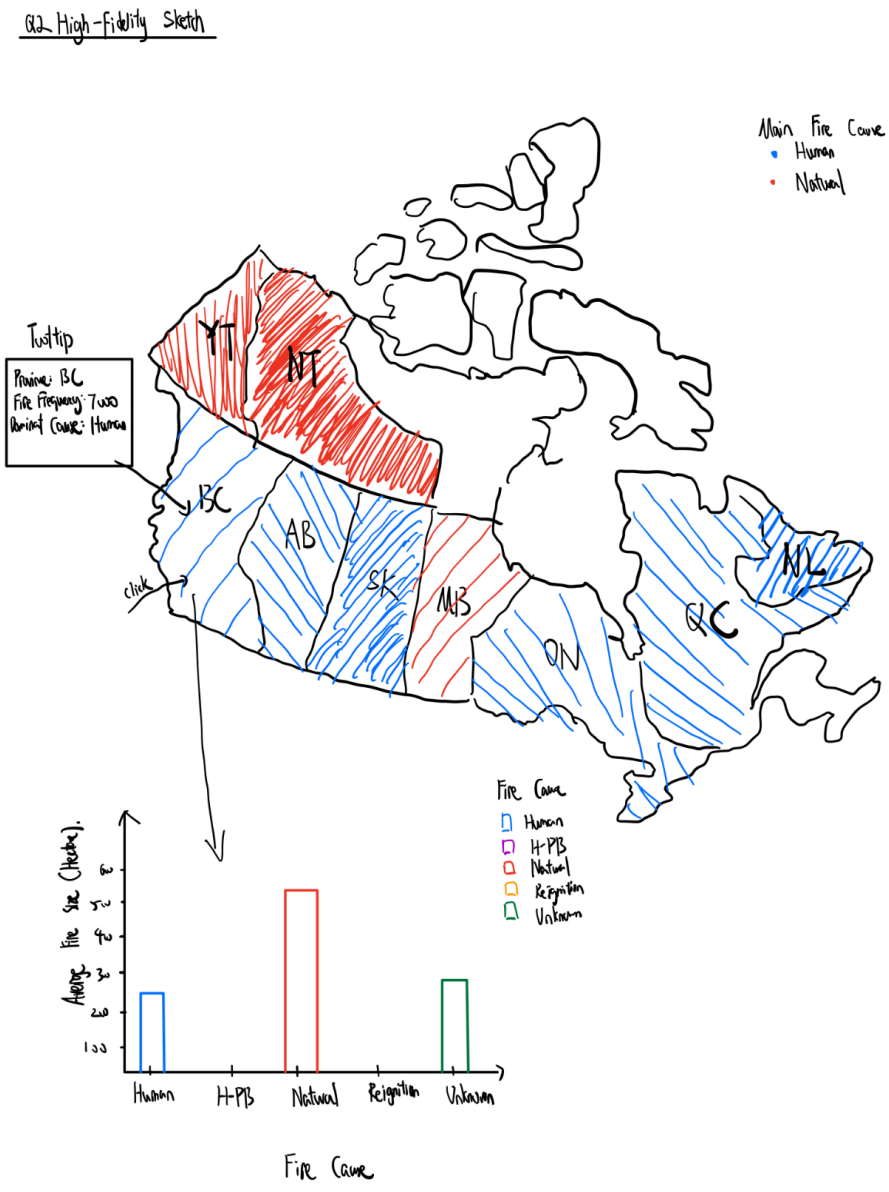




c. Critique of the plots:

1. The map clearly highlights which cause dominates each region. However, it does not show how many total observations exist within each region, and viewers may ignore less frequent but still important fire causes.
2. The strip plot effectively presents the overall distribution and number of observations for each fire cause, but it ignores regional details and does not capture spatial variation.
3. The stacked bar chart combines regional details, total fire counts, and dominant causes. However, it is hard to compare smaller categories across regions due to different alignments and limited sample sizes. And, it doesn't have details about fire size.

d. Finalized sketch:

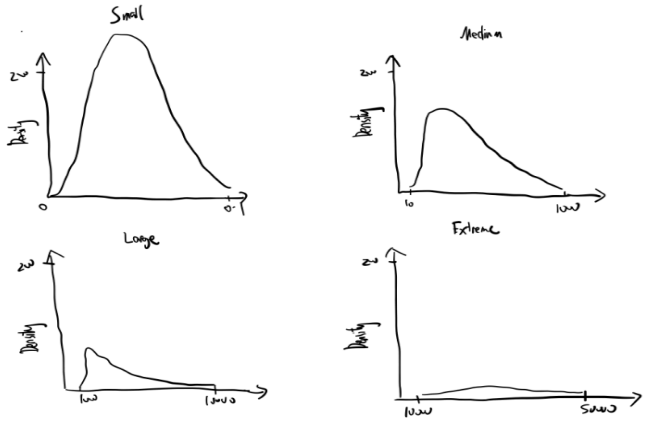


Analytic Question 3: “Are there any extreme fire events that deviate from national severity trends, and what might these outliers reveal about contributing conditions or causes?”

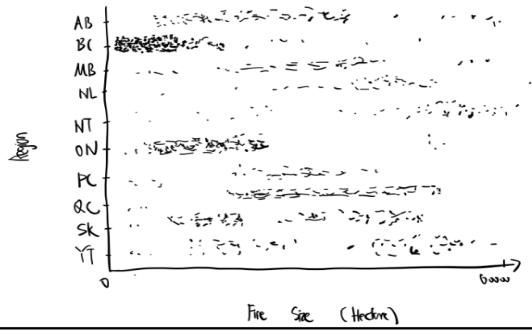
a. Sketch Ideas:
    1. Faceted density charts of fire size to show the overall distribution of each fire severity group across Canada.
    2. A strip plot with fire size on the x and province on the y.
    3. A boxplot of fire size by province with annotated outliers.

b. **Analytic Tasks**:
    - **Find Anomalies:** Identify the extreme fire events.
    - **Cluster:** Group fire events based on their size.
    - **Characterize Distribution:** Visualize the frequency and distribution of severe and extreme fires across Canada.

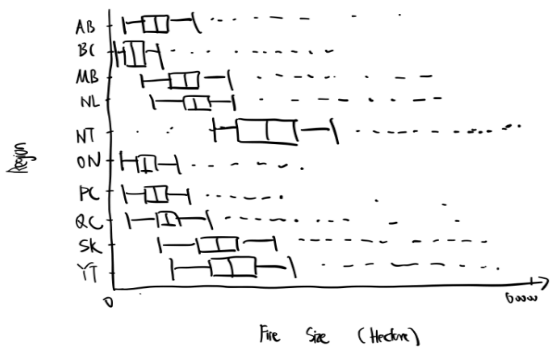
Sketch 1



Sketch 2



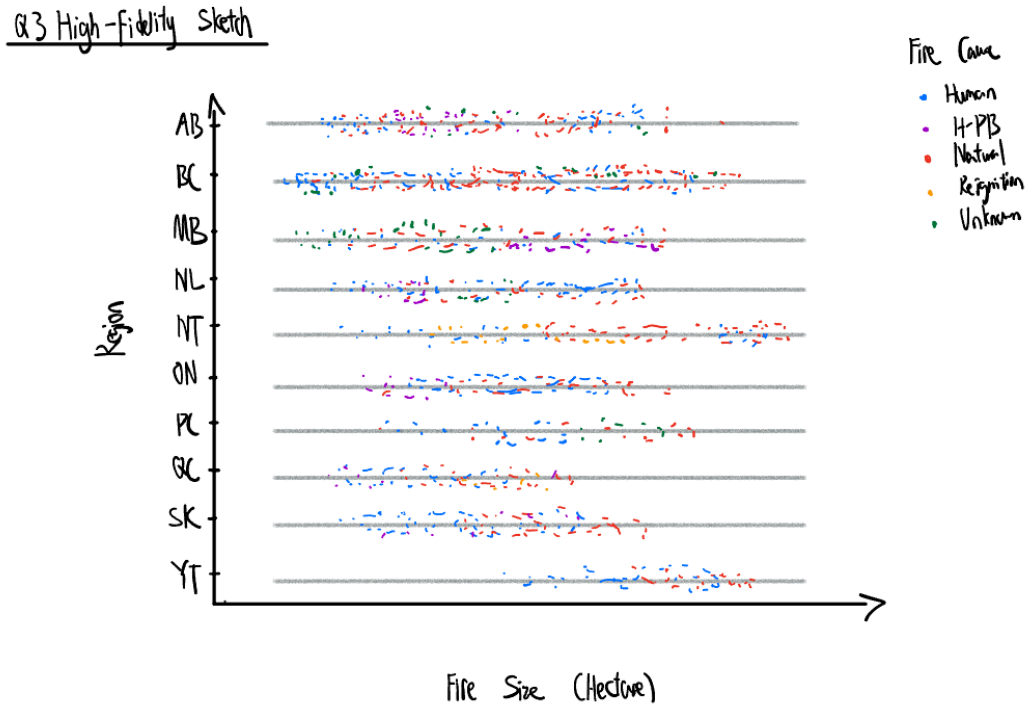
Sketch 3



c. Critique of the plots:

1. The faceted density plots effectively show the central tendency and shape of each fire severity group. However, they make it difficult to extract numerical information such as the number of observations or specific summary statistics.
2. The strip plot displays both the number of observations and the overall distribution well. However, it is difficult to interpret exact statistical summaries since only approximate comparisons can be made visually.
3. The boxplot clearly summarizes the overall distribution and provides precise statistical summaries, but it also hides the detailed shape of each distribution (e.g., multimodality) and does not show the number of observations.

d. Finalized Sketch:



Next Steps

At the end of the report, you should add a detailed, actionable plan with a timeline, and identified tasks for each group member. Feel free to use tables or point form - you will be assessed on the thoughtfulness and specificity of your plans.

- This week (Nov 10th) - finalize data cleaning, agree on color palette. Discuss the general flow of the dashboard, and what visualizations we want to implement.
    - Meeting on Wednesday November 12th (virtually, or in person after lecture)
- Next week (Nov 17th) - Implement final sketch visualisations, discuss extra details or further interactive elements we can implement. Each person will be responsible for their

section of the project based on what sections we took here. (Full overview graphs, detailed view graphs, etc.)
    - Meeting on Wednesday November 19th (virtually, or in person after lecture)
- Following week (Nov 24th) - finalize report/website and video.
    - Meeting on Wednesday November 26th (virtually, or in person after lecture)
    - Video = group effort. Each person will record a demo of the section they worked on. Combine all videos together (whoever has video editing software).
    - If time permits, we will create a website. Blake can assist with that further.
    - The report will be a group effort to write, we will work together to determine the flow and structure.
- Everyone works on their own inquiry focus according to the timeline