1. Dataset Sourcing

The dataset was retrieved from the Open Government Portal (https://open.canada.ca/en), an initiative by the Government of Canada aimed at promoting dialogue, transparency, and accountability through an open, searchable database. Four filters were applied to refine the search: Open Data, Open Maps, resource type as Dataset, and format as GeoTIFF (a file type that contains a graphical image, such as a map, along with geo-referenced data including coordinates and high color depth). The dataset represents a geospatial map of the North American region (covering Canada, the USA, and part of Mexico), showing 30-year monthly temperature averages for each month in high color depth, Agriculture and Agri-Food Canada (2019).

2. Application Description

According to the Government of Canada, Agriculture and Agri-Food Canada (2024), temperature data is a critical tool in agriculture, as changes in temperature significantly impact water supply, soil conditions, harvests, and pesticide use. Warmer temperatures can extend growing seasons and make new areas suitable for crops, but they also increase evaporation rates, leading to drought and the need for better water management. Understanding how temperature changes affect factors like crop suitability, water management, and livestock is essential for maintaining consistent production. By analyzing current and expected temperature shifts, farmers can better manage risks and prepare for potential crop losses.

3. Data Transformation and Preprocessing

For data transformation, the GeoTIFF temperature files were processed using Python's rasterio library. The raster data was converted to a tabular format, where each grid cell's temperature value was mapped to its corresponding geographic coordinates (longitude and latitude) using an affine transformation. NoData placeholders (e.g., -3.4e+38) were masked to ensure only valid temperature data remained. Additionally, the year range and month information were extracted from the file names and added as variables in the dataset. Due to the large size of the data, resampling was applied by a factor of 4 to reduce the resolution, making the dataset manageable on local hardware. In the data cleaning and preparation process, the main steps involved filtering out invalid temperature values, transforming pixel coordinates to geographic locations, resampling the data to reduce size, and appending metadata from the file names. The result was a clean, structured dataset that included longitude, latitude, temperature, month, and year range, ready for geospatial and temporal analysis.

4. Single-variable Analysis: Temperature

Guiding question: What is the distribution of average monthly temperatures across the two-year ranges (1961-1990 and 1991-2010)?

Based on this plot, both 30-year ranges, 1961-1990 and 1991-2010, do not follow a normal distribution but exhibit similar patterns. It is visually clear that the first period (1961-1990) has more extreme peaks on the left, representing colder temperatures, compared to the second period. There is also a slight difference on the right, with warmer temperatures more frequent in the second 30-year period, indicating an increase in warmer conditions.

Observable differences, especially in the tails, show that during colder seasons, the first 30 years had more negative temperatures, with a denser left tail compared to the second period. Temperature variability appears relatively consistent across both periods and seasons, though extreme cold temperatures were more common in the earlier period, while extreme warm temperatures were more frequent in the later period, especially during sum

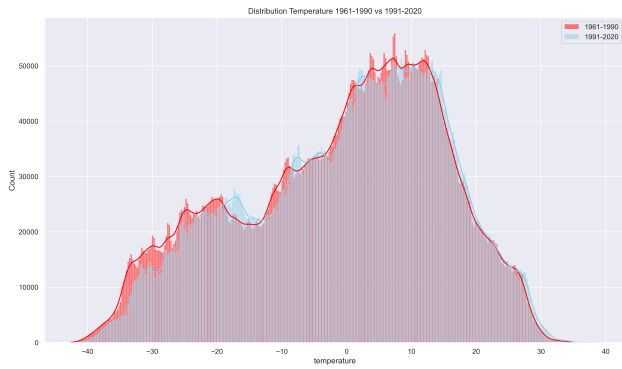

Figure 1: The Distribution of Temperature Compared with Two 30-Year Range

Guiding question: Is there any significant difference between seasons, Sarac (2023)?

During all four seasons, the mean temperatures in the second 30-year period are slightly higher than in the first. In the winter, the first 30 years show a higher standard deviation, indicating more extreme lower temperatures compared to the second period. The spring follows a similar pattern, with the first 30 years showing more volatility and extreme temperatures, while the second period shows warmer temperatures. In the summer, there are significant outliers in both periods, though the second 30 years have slightly more extreme temperatures. The fall confirms that lower temperatures were more consistent in the first 30 years, with more frequent lows and extremes than in the second period.

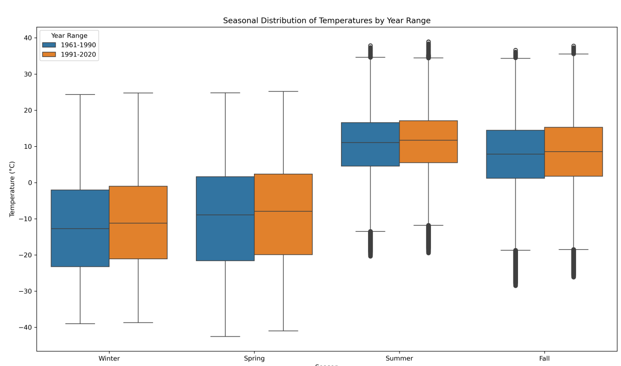

Figure 2: Seasonal Comparison of Two 30-year Range

5. Multi-variable Analysis: Temperature vs. Latitude

Guiding question: What is the relationship between these two variables (Full Data)? Using Spearman's correlation table, which measures the strength and direction of the association between Latitude and Temperature, the two variables show a negative correlation of -0.64. Spearman's method is particularly important here because it doesn't assume the variables follow a normal distribution. This result confirms expectations, as higher latitudes represent locations closer to the North or South Poles, where temperatures are generally lower.

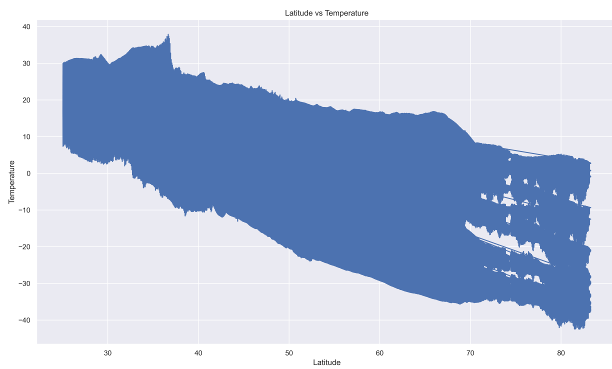

Figure 3: Temperature Chang Across Latitude