

Spring Wheat

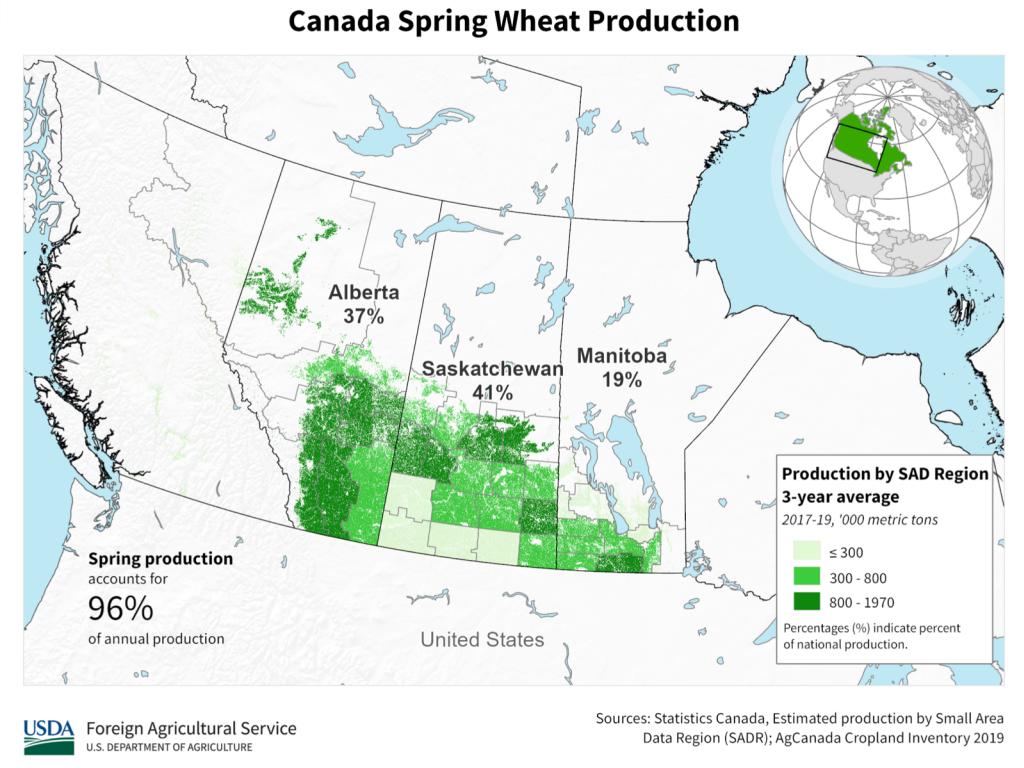
Providing insights into spring wheat yield data across Saskatchewan

Presentation by Vineetha Warriyar, Oct 2023



Objective

Can we provide some insights into patterns and grouping within spring wheat yield data across Saskatchewan based on the rural municipality information provided in the dataset?



PS: Soft wheat offers various options for addressing food allergies and intolerance.

Data Structure

- ▶ Yield data from 1938 to 2022.
- ▶ 299 Rural Municipalities.
- ▶ 16 Crop Types: Winter Wheat, Canola, Spring Wheat, Mustard, Durum, Sunflowers, Oats, Lentils, Peas, Barley, Fall Rye, Canary Seed, Spring Rye, Tame Hay, Flax and Chickpeas.
- ▶ A total of 25312 entries

The dataset consists of the following variables:

- ▶ year
- ▶ RM
- ▶ Yield for each crop (in bushels)



Problem Statement

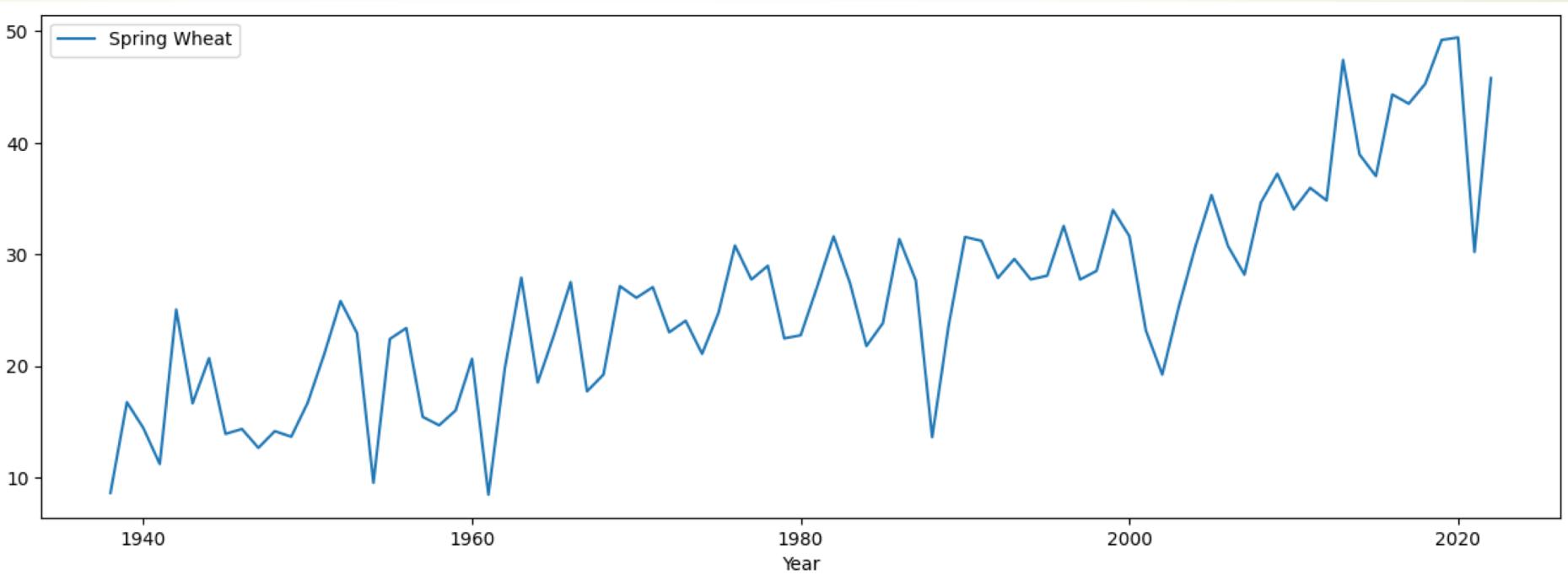
- ▶ Main Focus:
 - ▶ Spring Wheat yield across Saskatchewan
 - ▶ Rural Municipalities in Saskatchewan
- ▶ This analysis aims to partition spring wheat yield into clusters or groups with minimum within-cluster variation.
- ▶ This will provide insights into patterns and groupings within the spring wheat yield data.

Data Collection and Preprocessing

- ▶ Data Source:
 - ▶ Government of Saskatchewan
- ▶ Yield is represented in bushels.
- ▶ There are 99 missing observations in the spring wheat yield.
 - ▶ Imputed via mean imputation
- ▶ One extreme value corresponding to RM = 314 in 2018.
 - ▶ Excluded from analysis
- ▶ Adding geospatial information to yield data provides 296 rural municipalities.

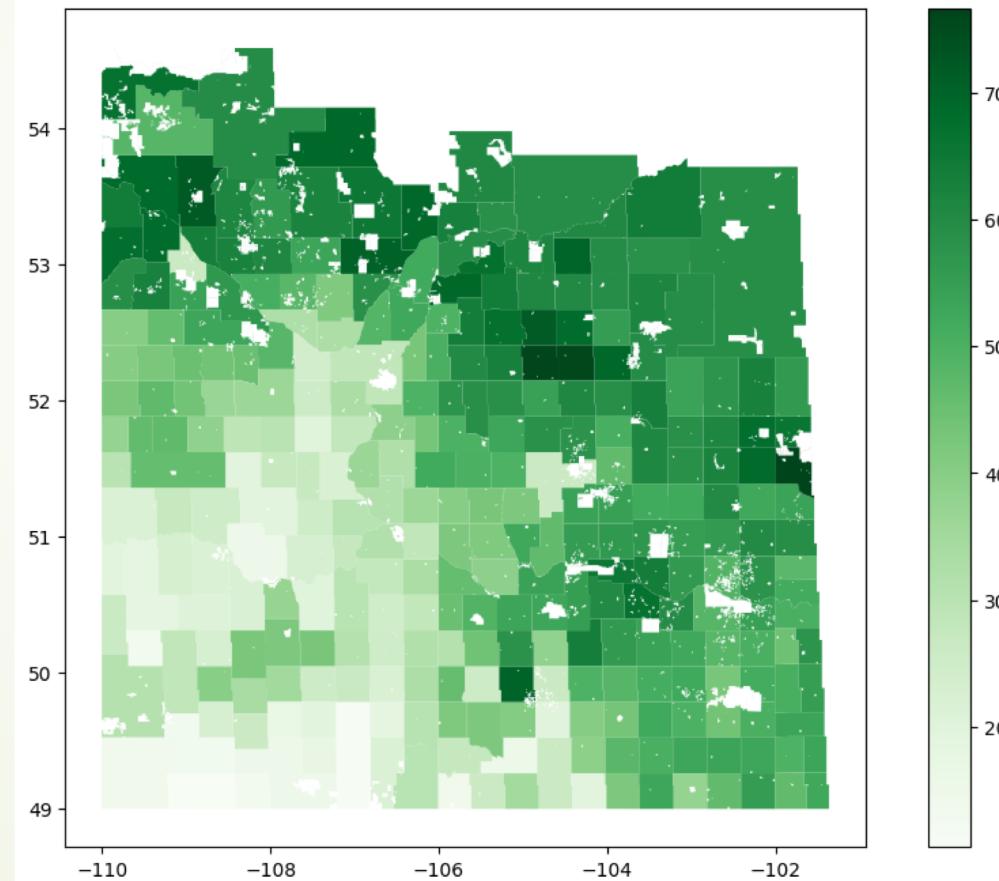
Exploratory Data Analysis

Yearly average yield from 1938 to 2022



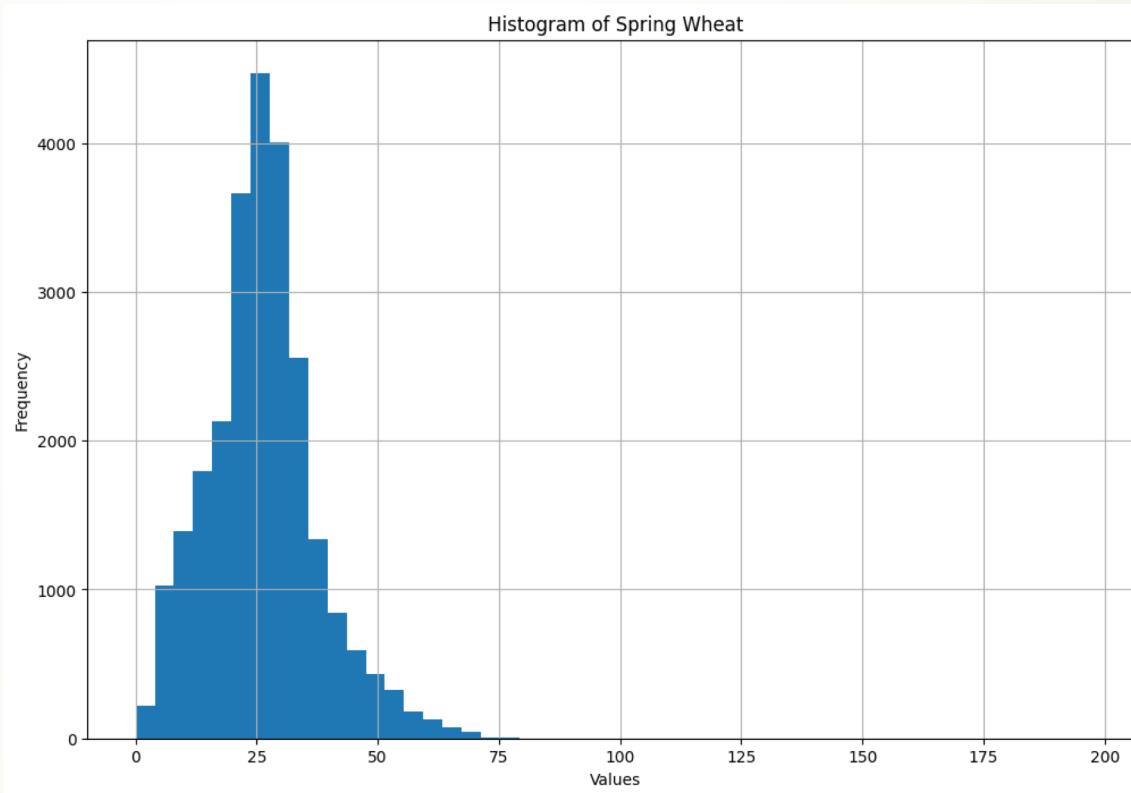
Exploratory Data Analysis

Spring wheat yield for year 2022 in various RMs - GIS Visualization



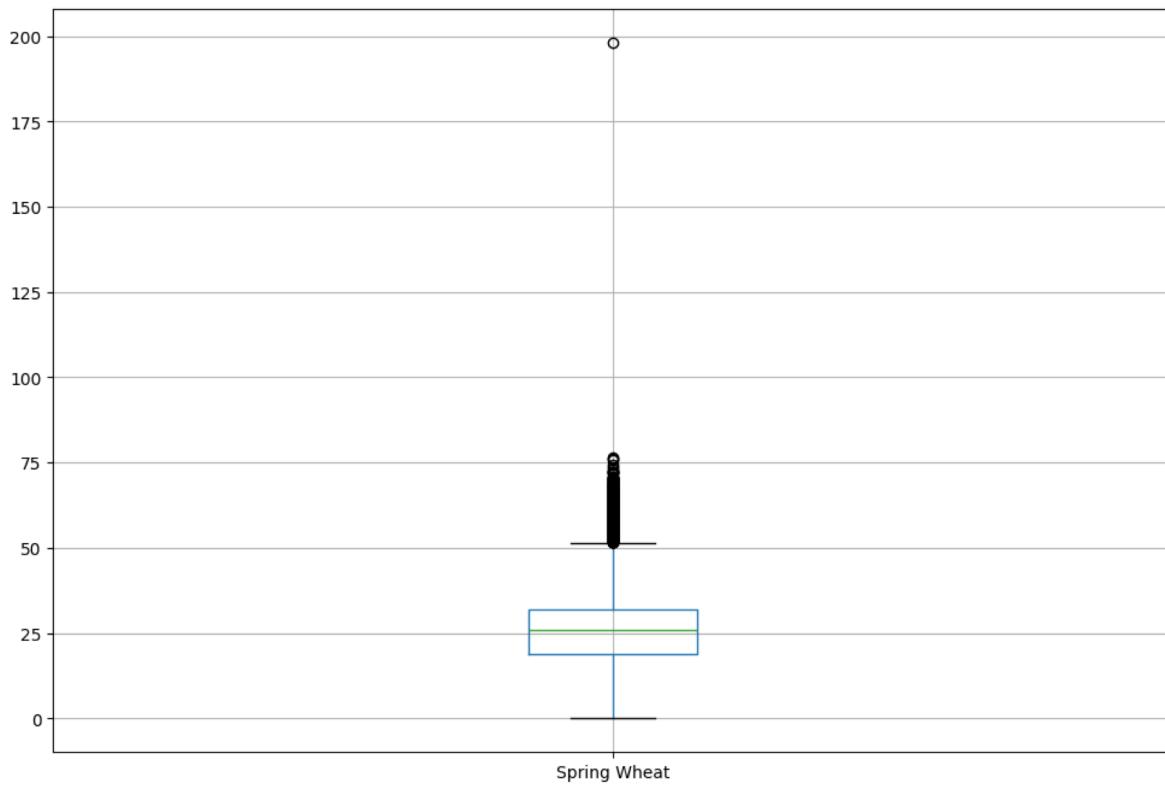
Exploratory Data Analysis

Histogram – spring wheat yield data distribution



Exploratory Data Analysis

Boxplot – Are there any outliers or extreme values?



Extreme value:

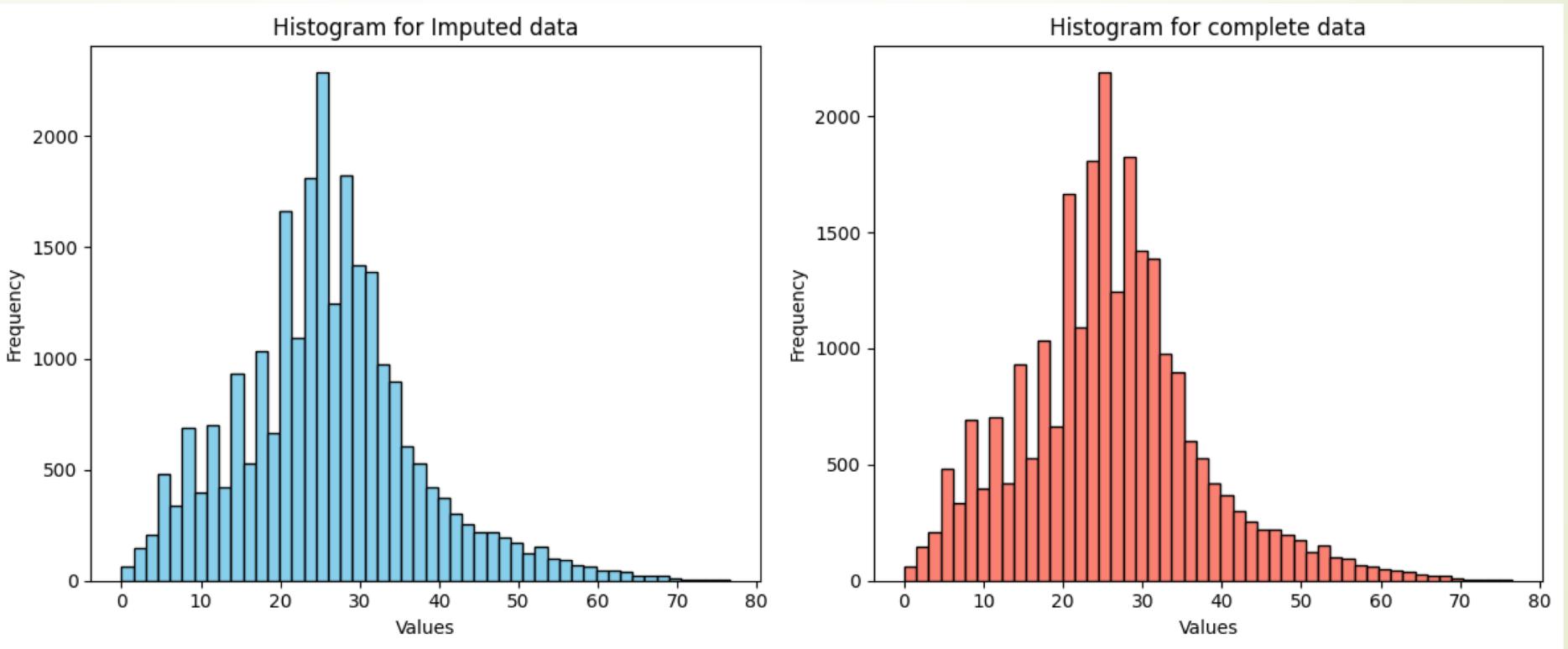
- Could be a typo?
- Rest of the years data varies from 4 to 45.
- Excluded.
- Further investigation needed.

Outliers:

- Seems to be relevant.
- High yield for corresponding RMs.
- Could do more data gathering.
- Included.

Exploratory Data Analysis

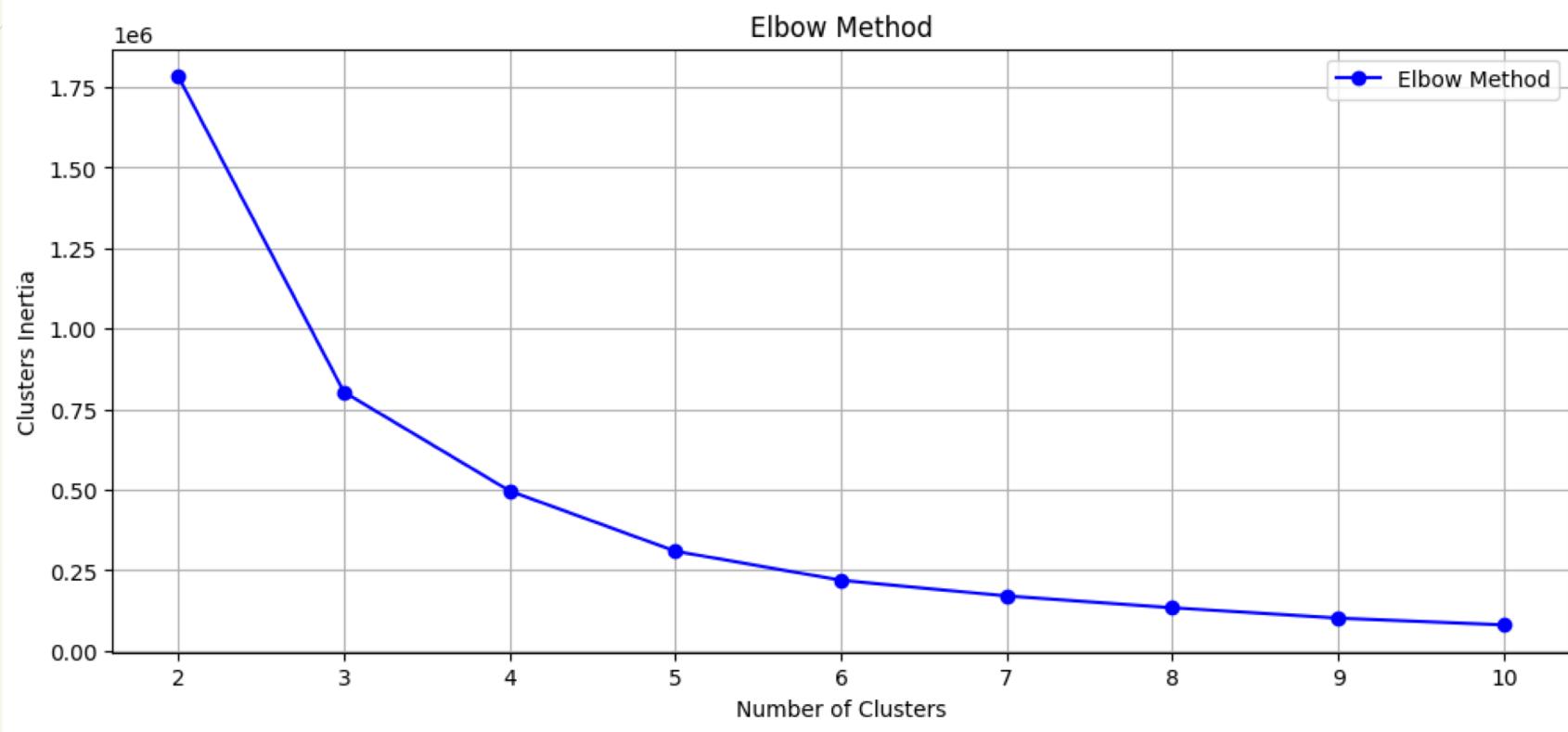
Sensitivity analysis for the 99 imputed values – both distribution looks similar



Methodology: K-Means clustering

- ▶ Calculate mean and standard deviation corresponding to each RM.
- ▶ Group yield data using unsupervised machine learning technique called K-means clustering based on rural municipalities.
- ▶ Elbow method to find the optimal number of clusters by minimizing within-cluster sum of squares.

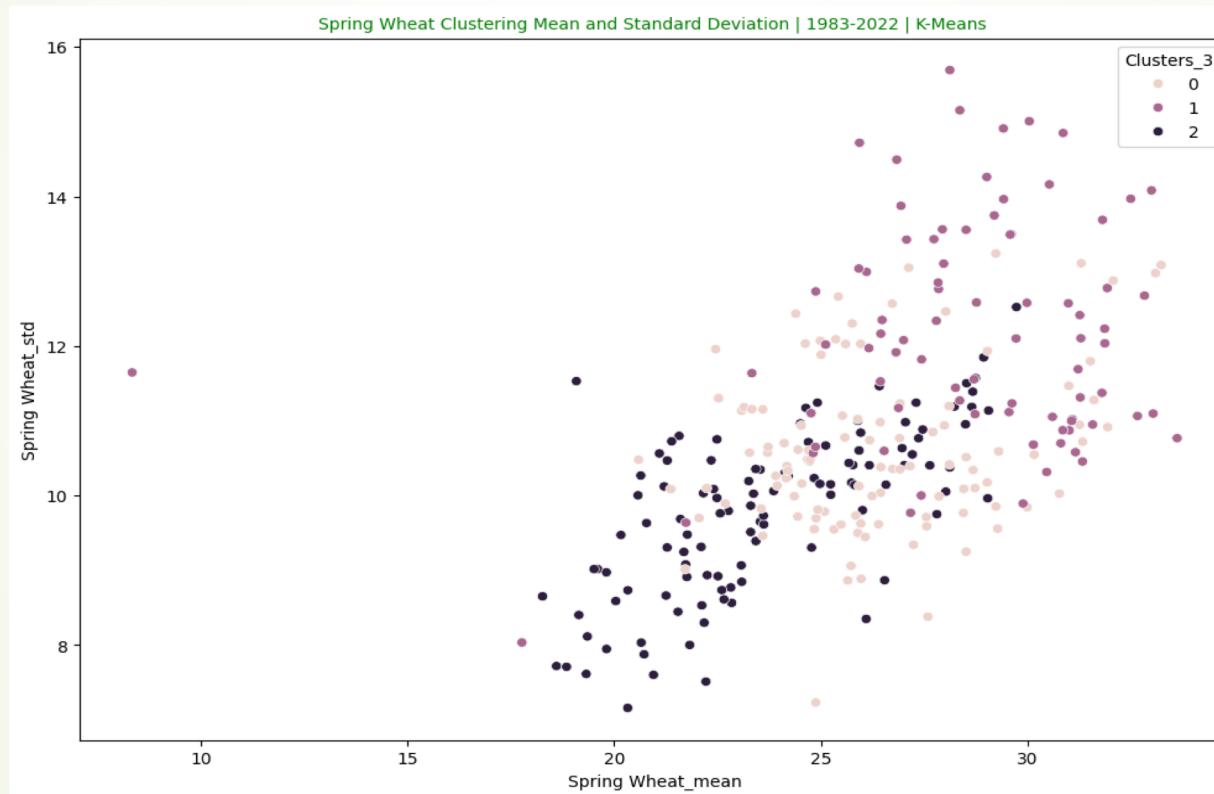
Results



The plot shows sudden change at 3 and 4. So it is reasonable to choose optimal cluster size as 3.

Results

Scatter plot with optimal cluster size, k = 3.



Results

- ▶ Provided clusters of rural municipalities with similar yield and variability.
- ▶ Spatial patterns of high and low spring wheat yields are displayed via interactive maps.
- ▶ Generated various visualization plots to get insights into spring wheat yield data.
 - ▶ Yearly average yield
 - ▶ Boxplot
 - ▶ Histogram
 - ▶ HTML Interactive files for exploration

Conclusion

- ▶ Presented a thorough view of spring wheat yield data in Saskatchewan's rural municipalities.
- ▶ Future Work
 - ▶ Add more features
 - ▶ Predictive modeling using supervised machine learning methods



Thank you!