SASKATCHEWAN RYE FORECASTING & PEA CLUSTERING

Alex D'Ippolito

July 31, 2023

Problem Statement

- Is it possible to predict Fall Rye yield based solely on historical data?

- Can we find clusters of similar RMs for Pea yield in Saskatchewan?
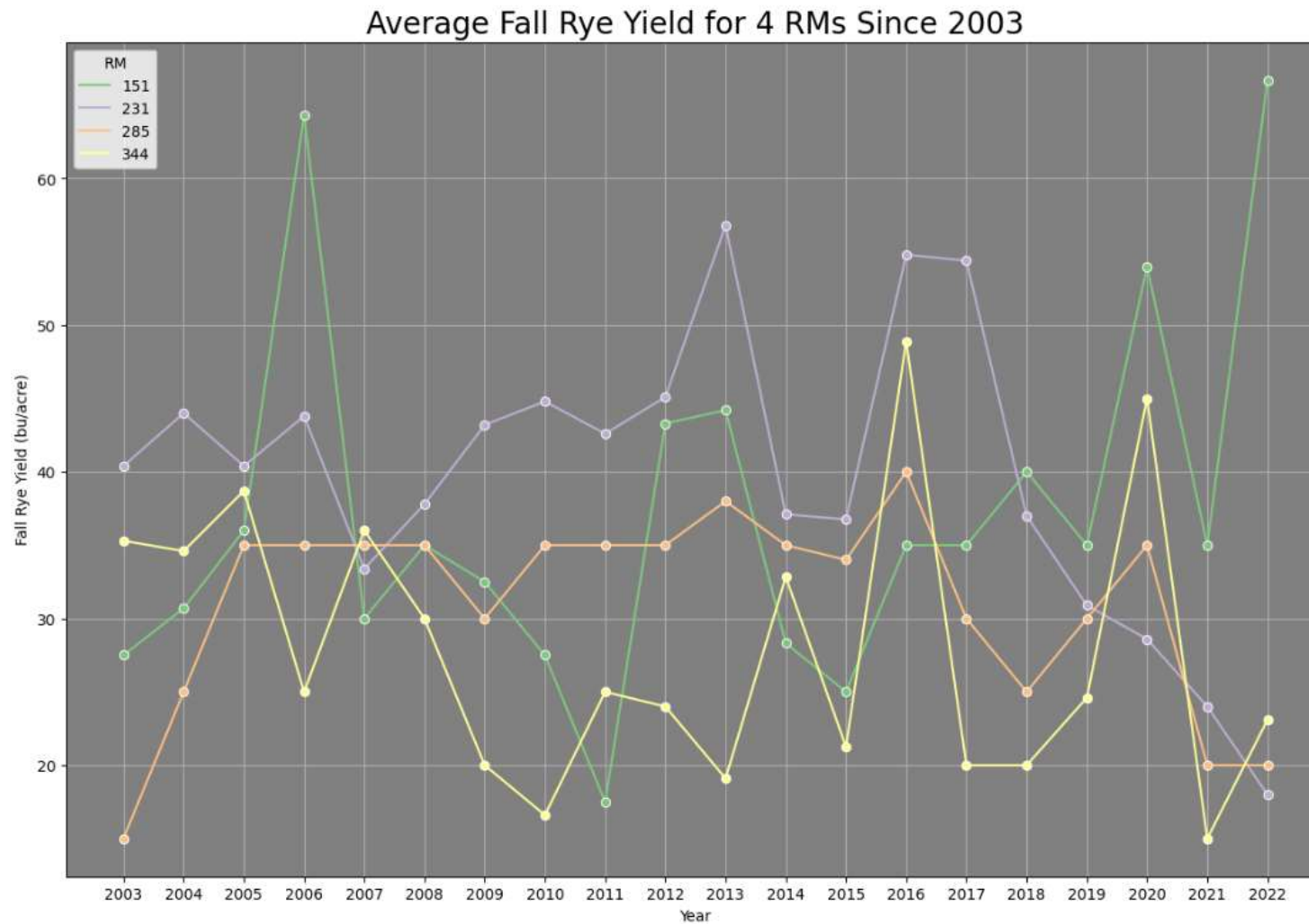
Data Collection and Preprocessing

- Historical RM yield data for 16 crops from 1938 -2022 https://dashboard.saskatchewan.ca/agriculture/rm-yields/rm-yields-data

- Geographical boundaries of Saskatchewan Rural Municipalities from https://github.com/ruhidm/Palette_Cohort_3_Data
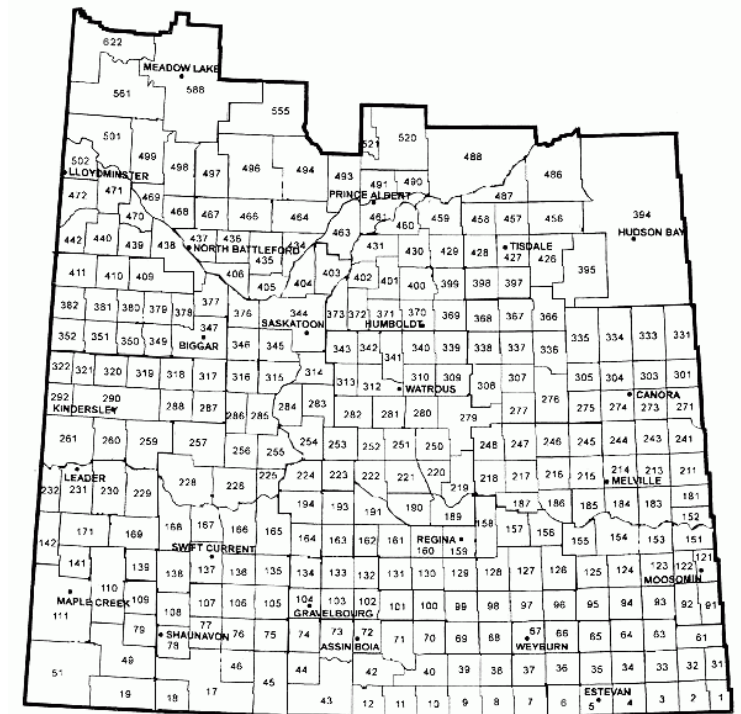
Exploratory Data Analysis

Missing Values in Yield Dataset Sorted by Year
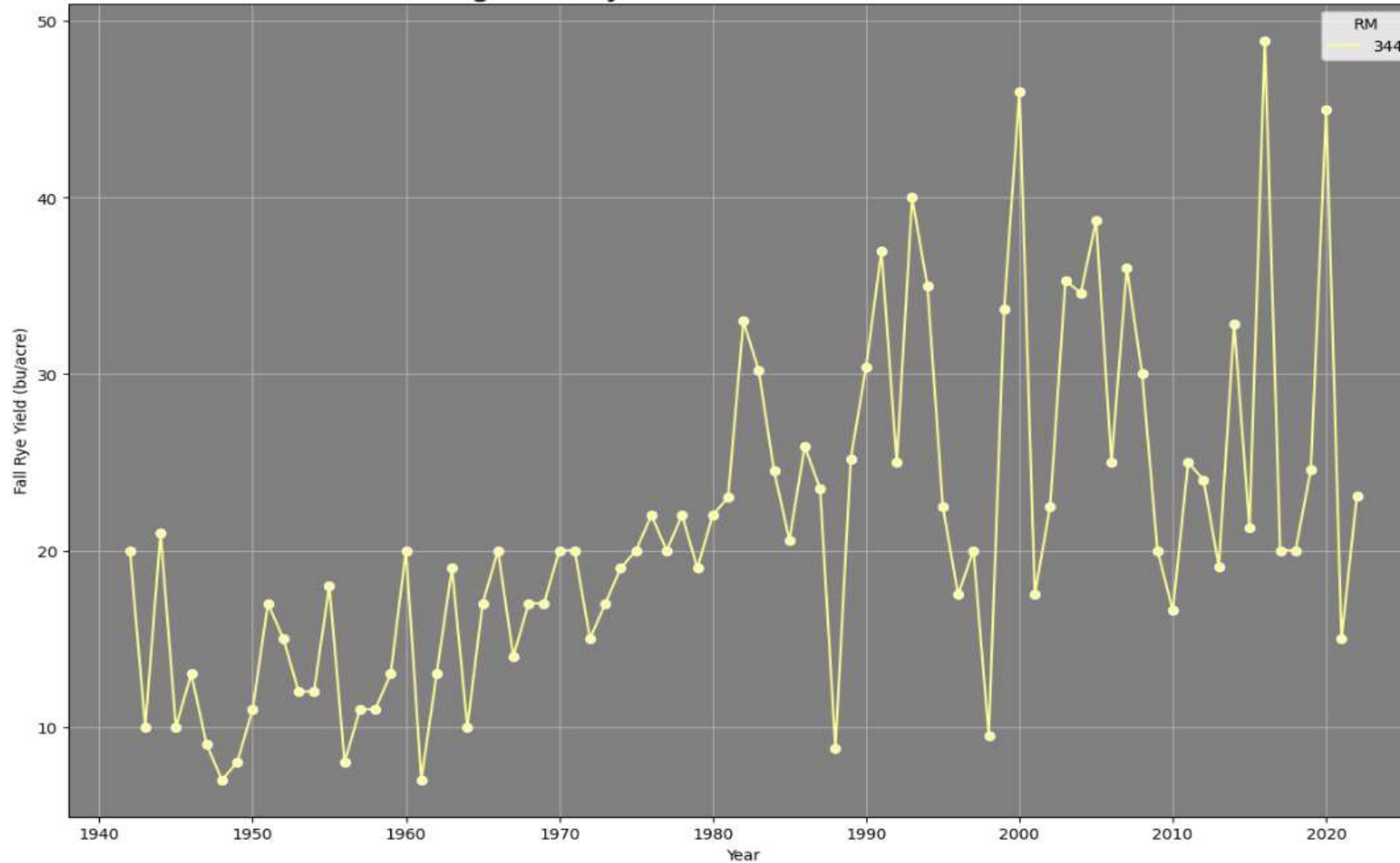
Exploratory Data Analysis
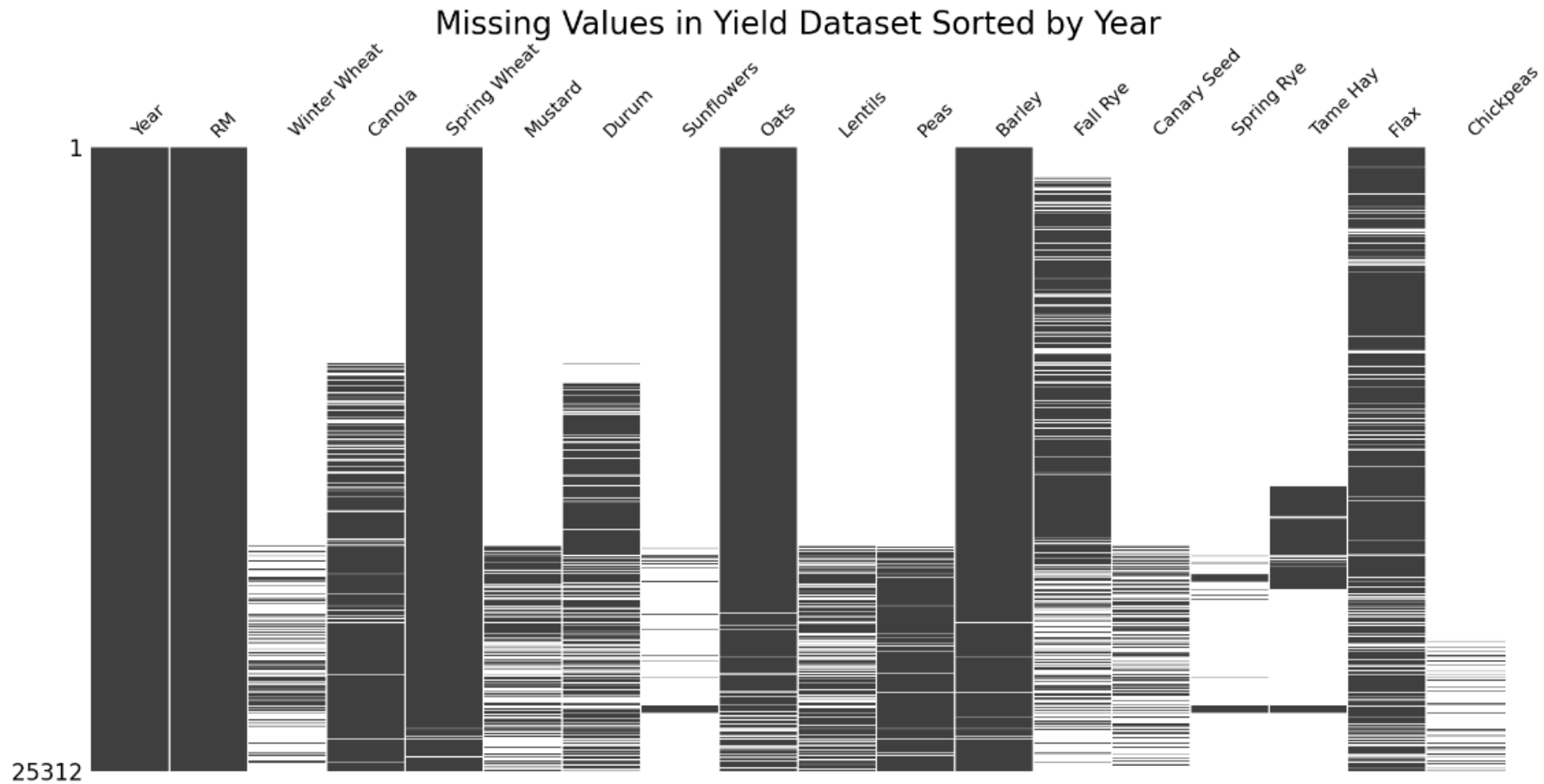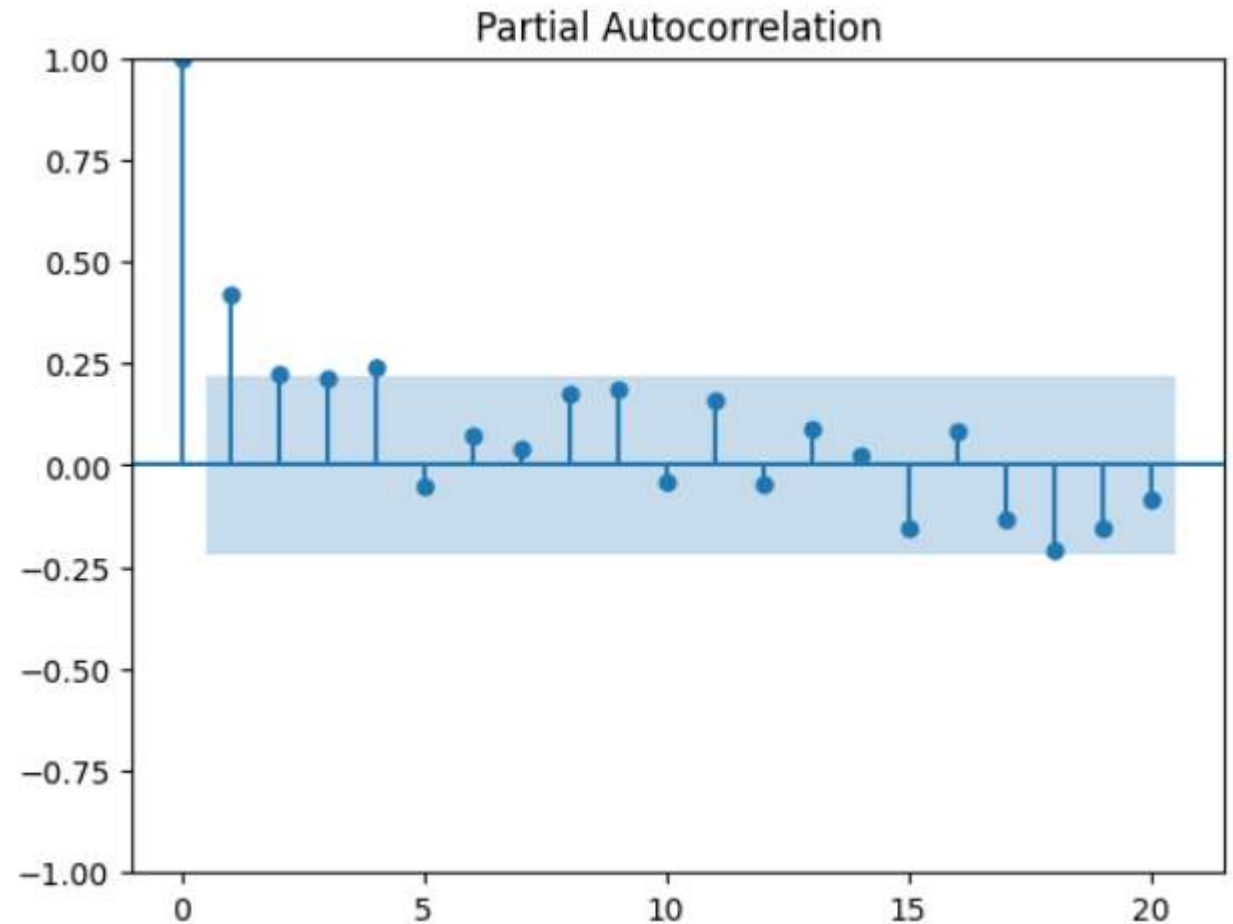

Average Fall Rye Yield for 4 RMs Since 2003

Exploratory Data Analysis

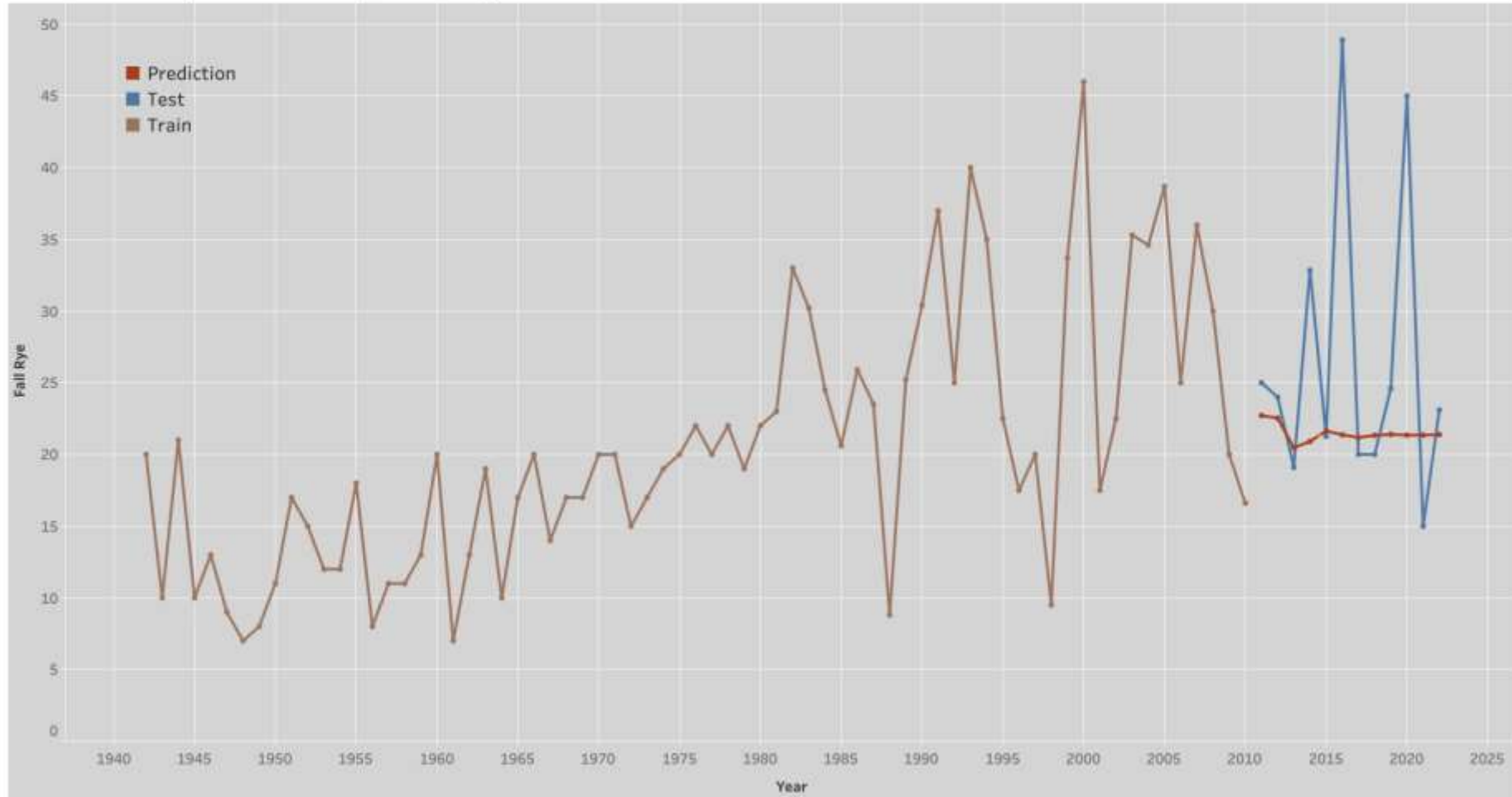Average Fall Rye Yield for RM 344 Since 1942
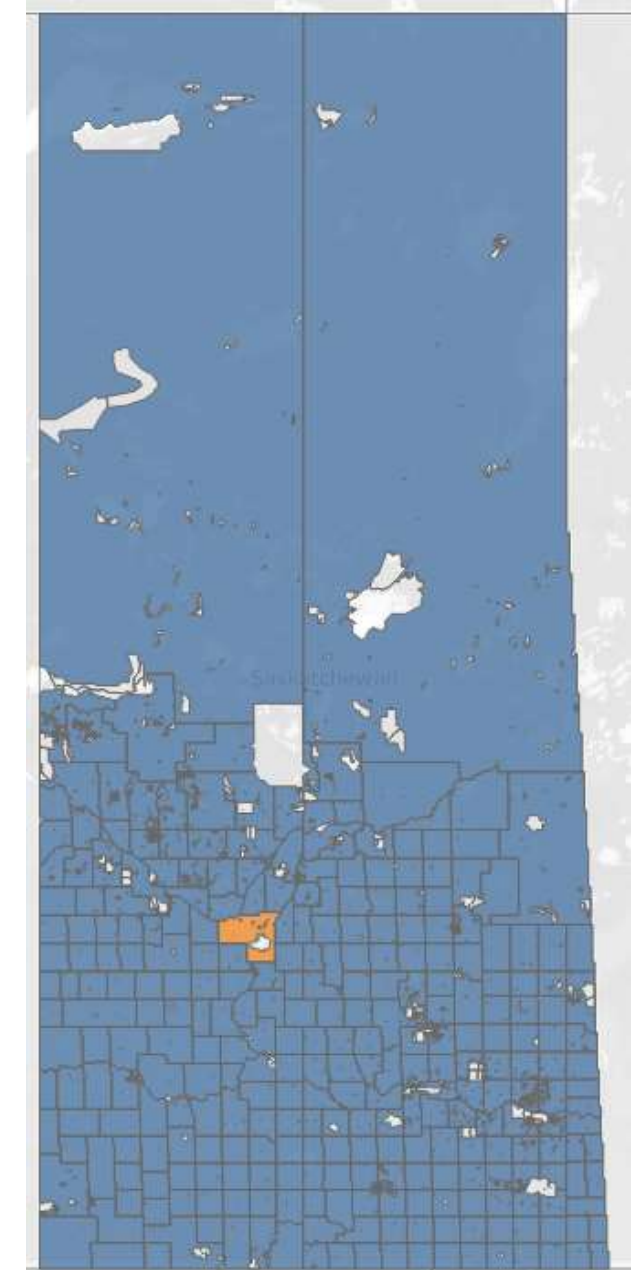
Time Series Analysis

- Partial autocorrelation function to help decide how many lags make sense to use for an AR model
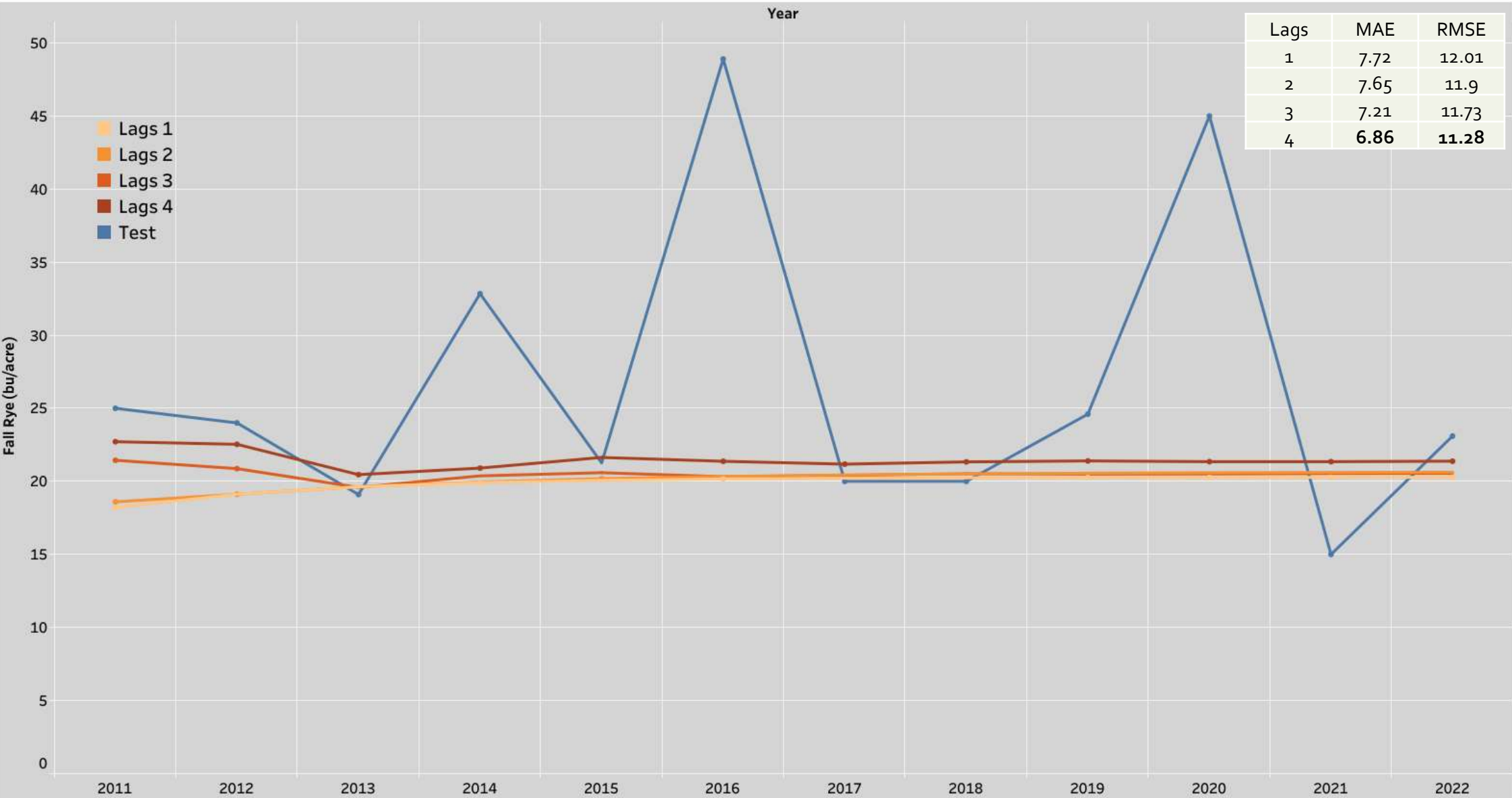
- 4 lags seems reasonable for this analysis

Partial Autocorrelation

AR Modelling

AR Model with 4 Lags for RM 344 Fall Rye Forecasting

AR Model Comparison for RM 344 Fall Rye Forecasting

Lags	MAE	RMSE
1	7.72	12.01
2	7.65	11.9
3	7.21	11.73
4	**6.86**	**11.28**

Test Data vs AR Model Comparisons

Lags	MAE	RMSE
1	7.72	12.01
2	7.65	11.9
3	7.21	11.73
4	**6.86**	**11.28**

ARIMA Model Comparison for RM 344 Fall Rye Forecasting

| Order	MAE	RMSE
4,0,1	**6.93**	10.26
4,0,2	6.96	10.24
4,1,1	7.83	**10.03**
4,1,2	9.96	11.31
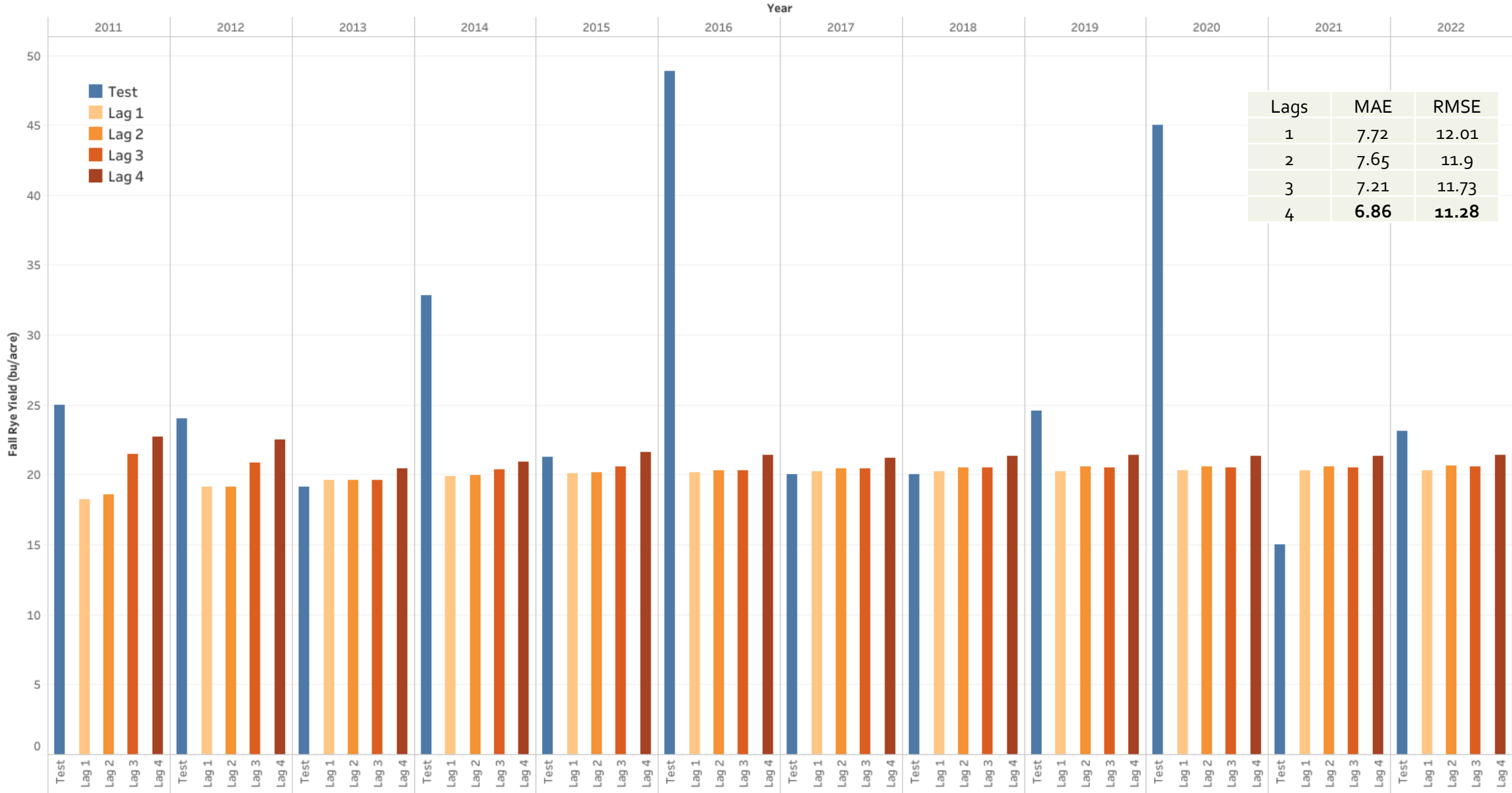
XGBoost Modelling

- Utilized 4 Lags as Features

- Built 3 Different XGBoost Models

Model #	n_estimators	max_depth	learning
1	100	3	0.1
2	500	4	0.1
3	1000	5	0.1
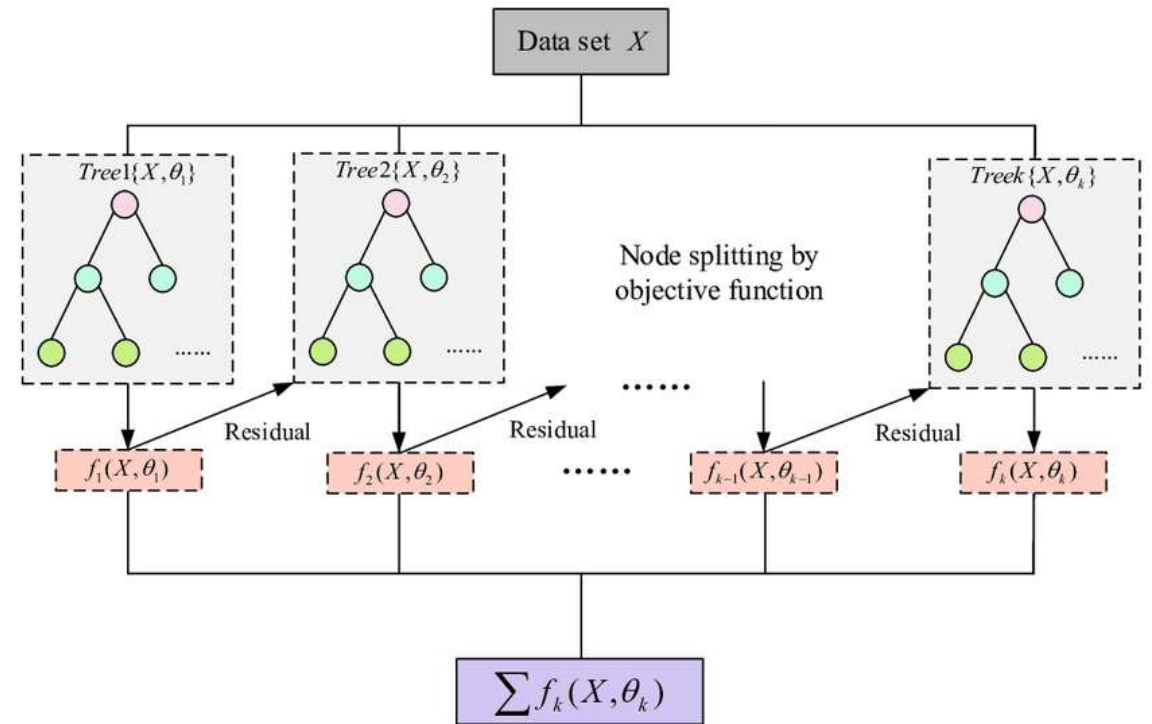
XGBoost Model Comparison for RM 344 Fall Rye Forecasting

Model #	n_estimators	max_depth	learning
1	100	3	0.1
2	500	4	0.1
3	1000	5	0.1

| Model	MAE	RMSE
1	9.66	12.6
2	9.55	**12.02**
3	**9.47**	12.12
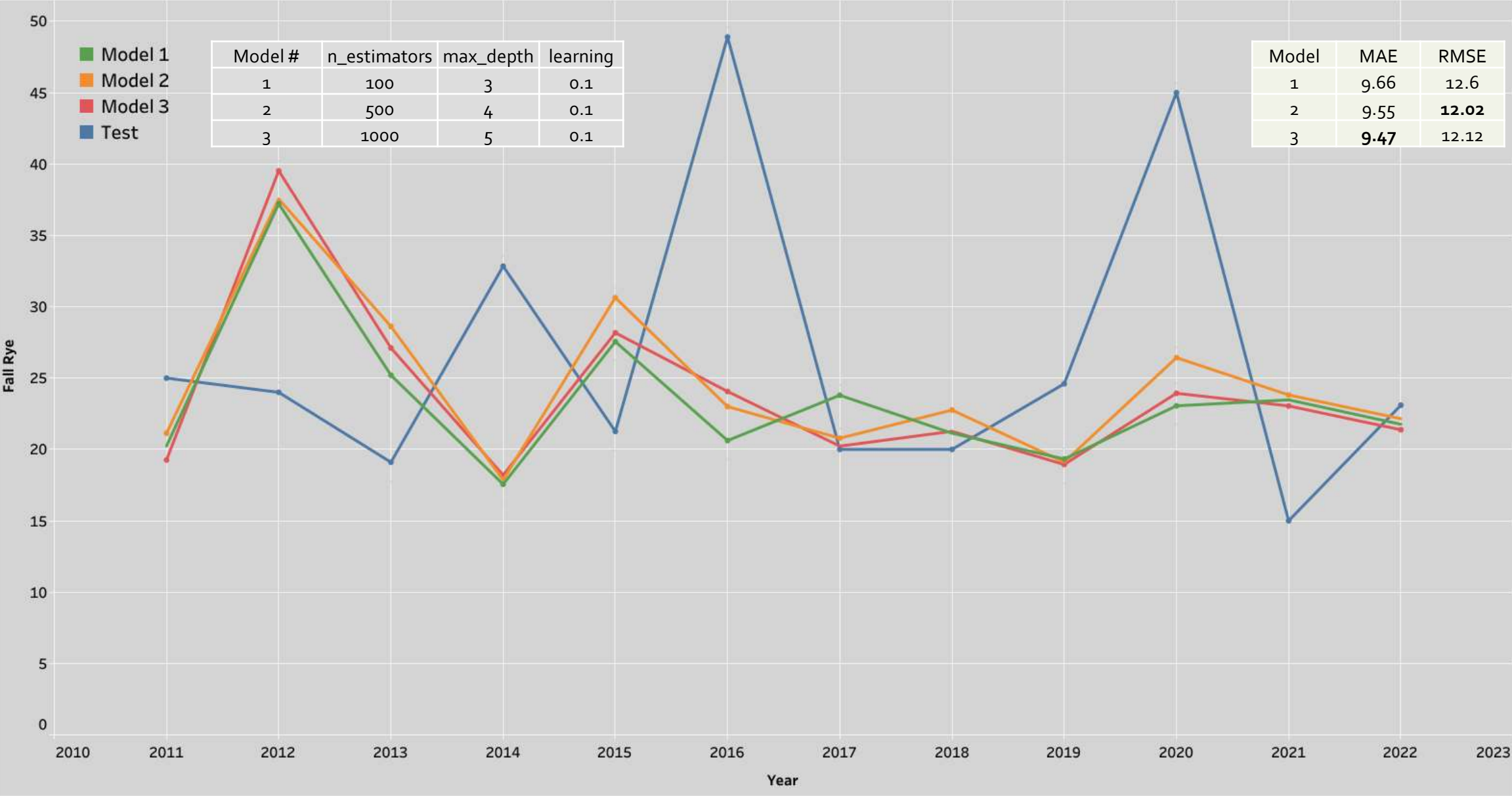
Results Summary and Next Steps

- AR model with 4 lags had the lowest MAE

- ARIMA model with order (4,1,1) had the lowest RMSE

- Look into improving XGBoost models through more thorough hyperparameter tuning (GridSearch or Optuna)

- Consider adding more features
  - Yields for nearby RMs
  - Weather forecast data
  - Soil property data

- Try forecasting for other RMs and compare results

Model	MAE	RMSE
AR 1	7.72	12.01
AR 2	7.65	11.9
AR 3	7.21	11.73
AR 4	**6.86**	11.28
ARIMA (4,0,1)	6.93	10.26
ARIMA (4,0,2)	6.96	10.24
ARIMA (4,1,1)	7.83	**10.03**
ARIMA (4,1,2)	9.96	11.31
XGBoost #1	9.66	12.6
XGBoost #2	9.55	12.02
XGBoost #3	9.47	12.12

Unsupervised Learning Analysis
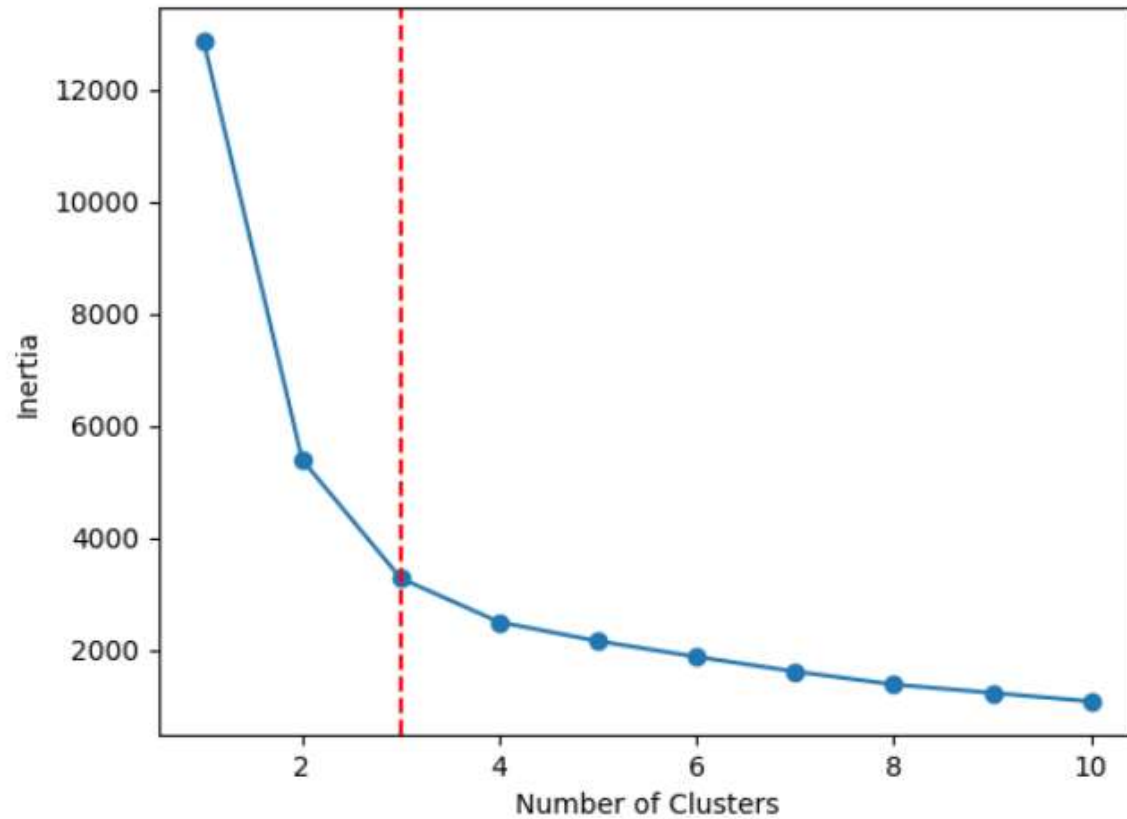

Saskatchewan RM Pea Yield Means from 2013-2022
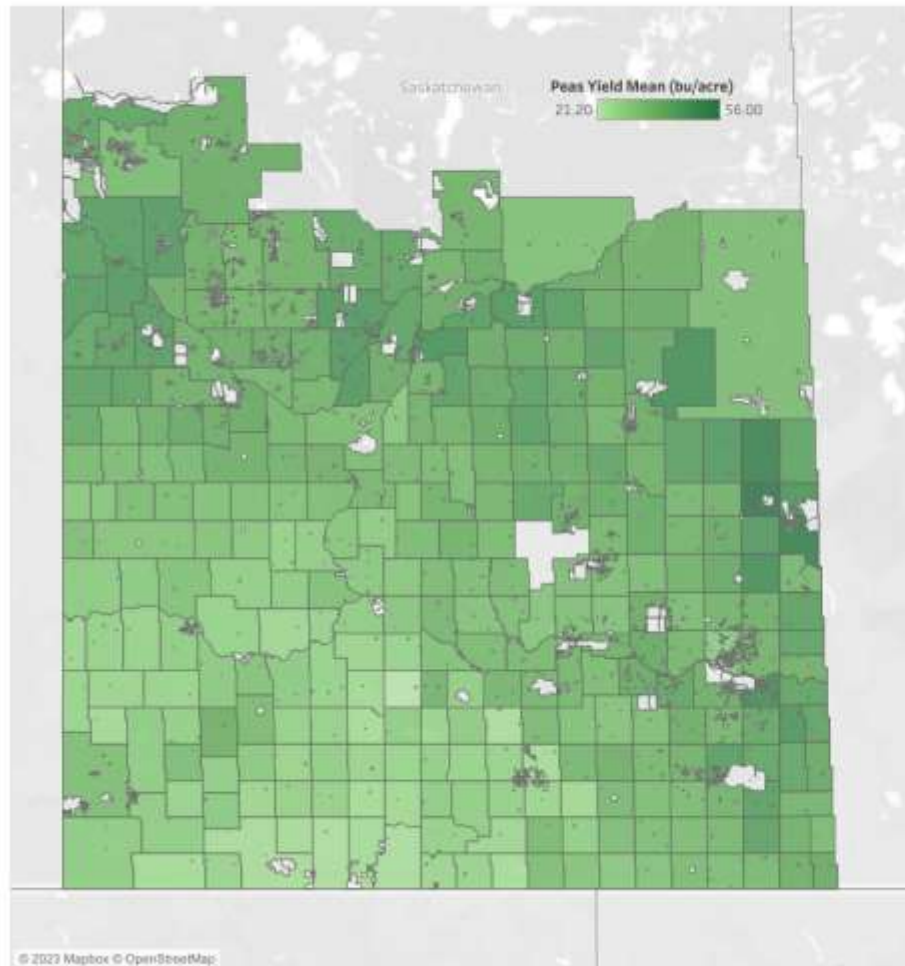
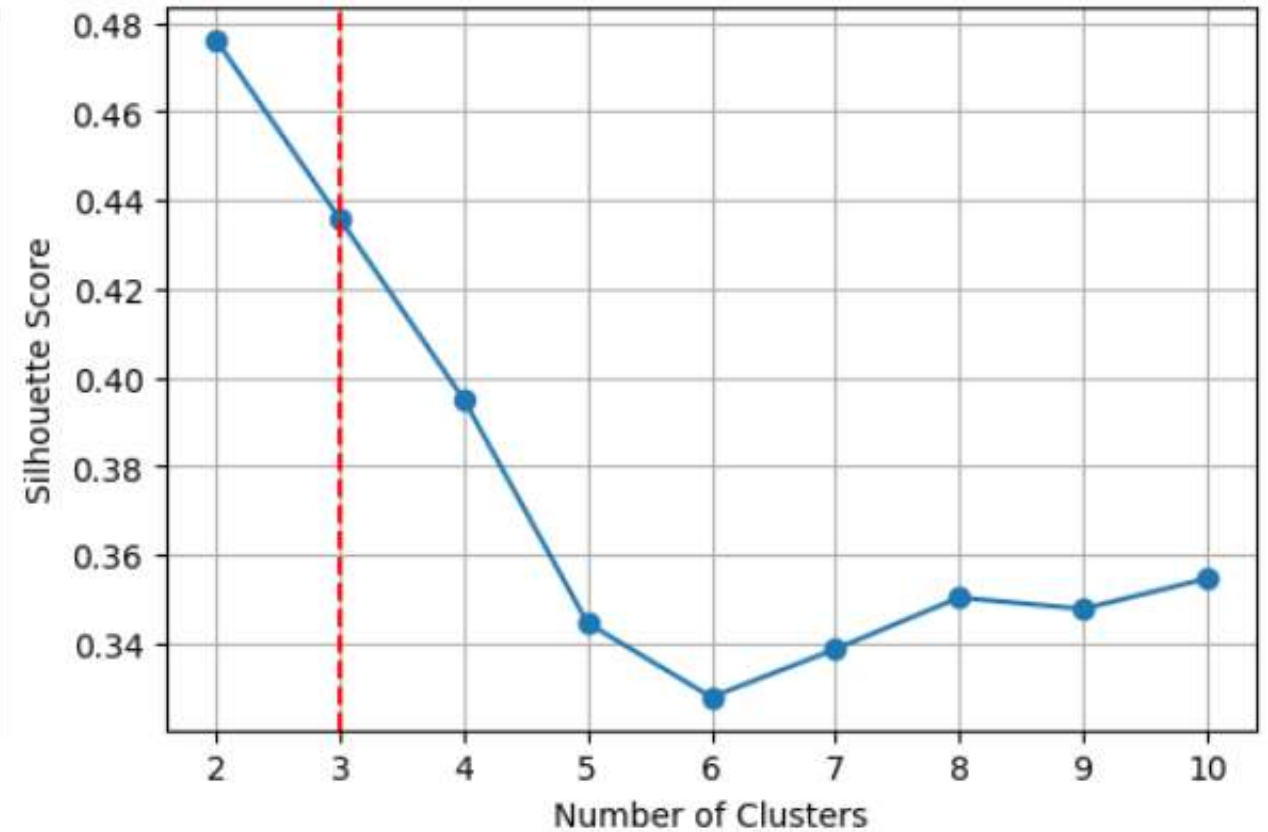Saskatchewan RM Pea Yield Standard Deviations from 2013-2022
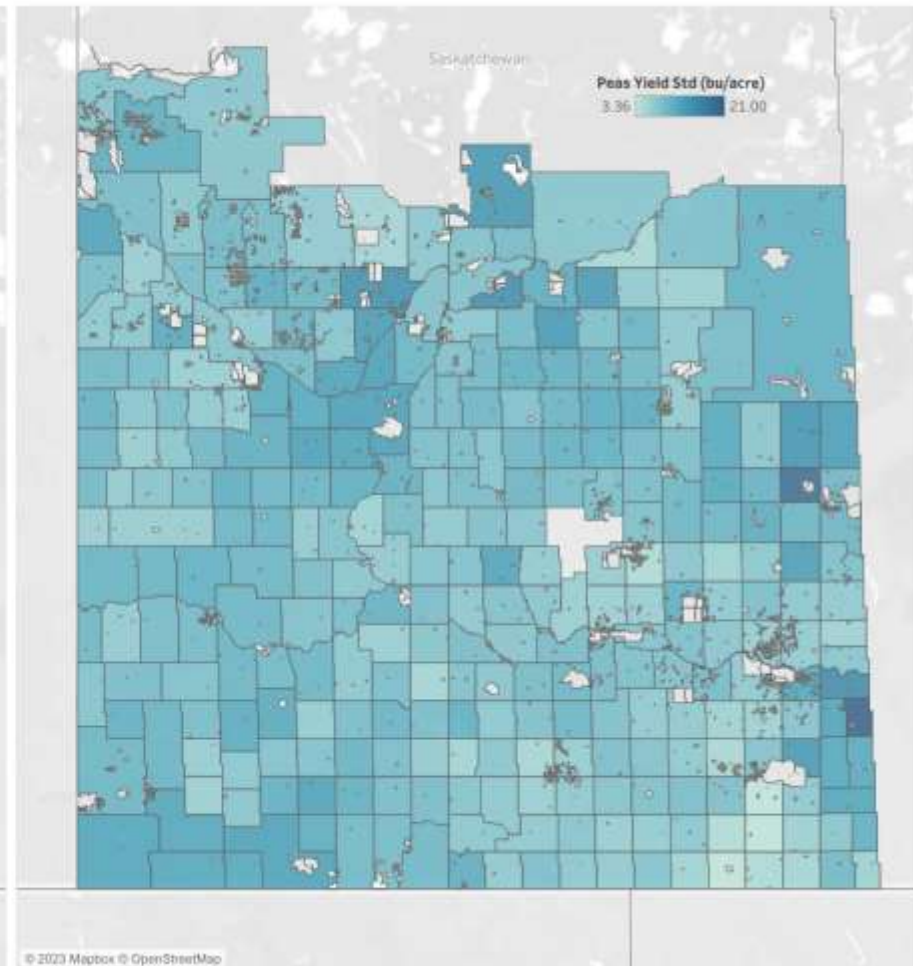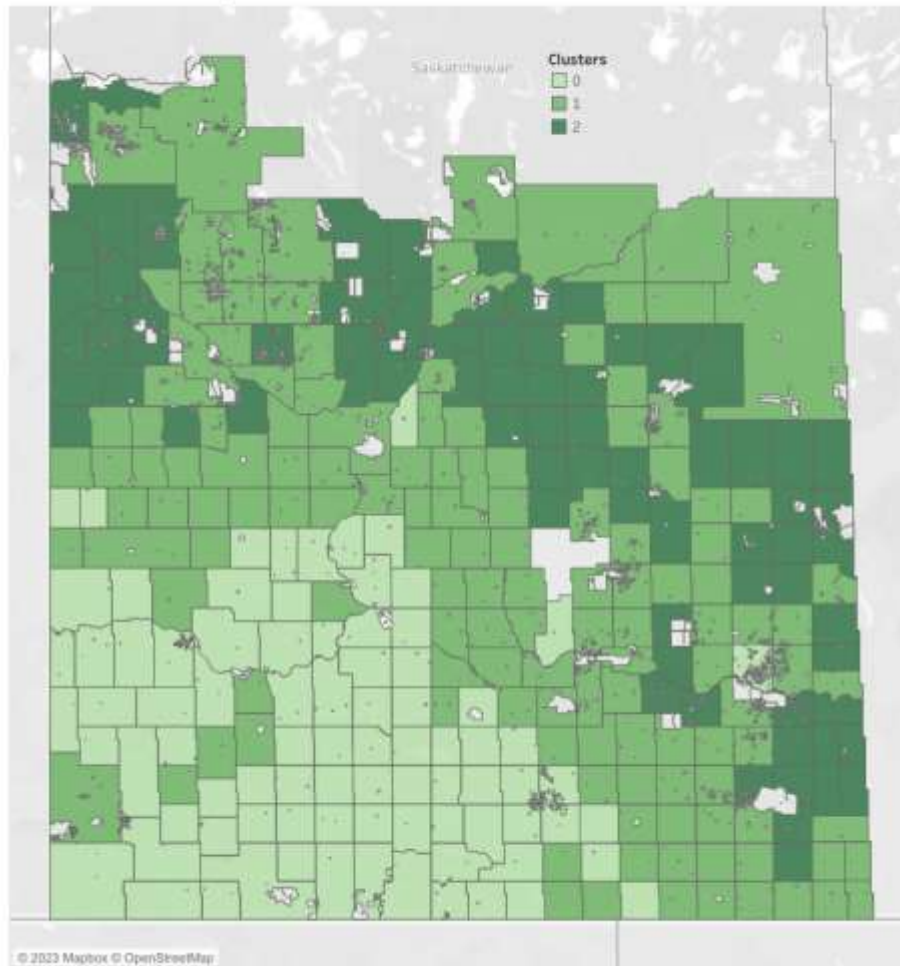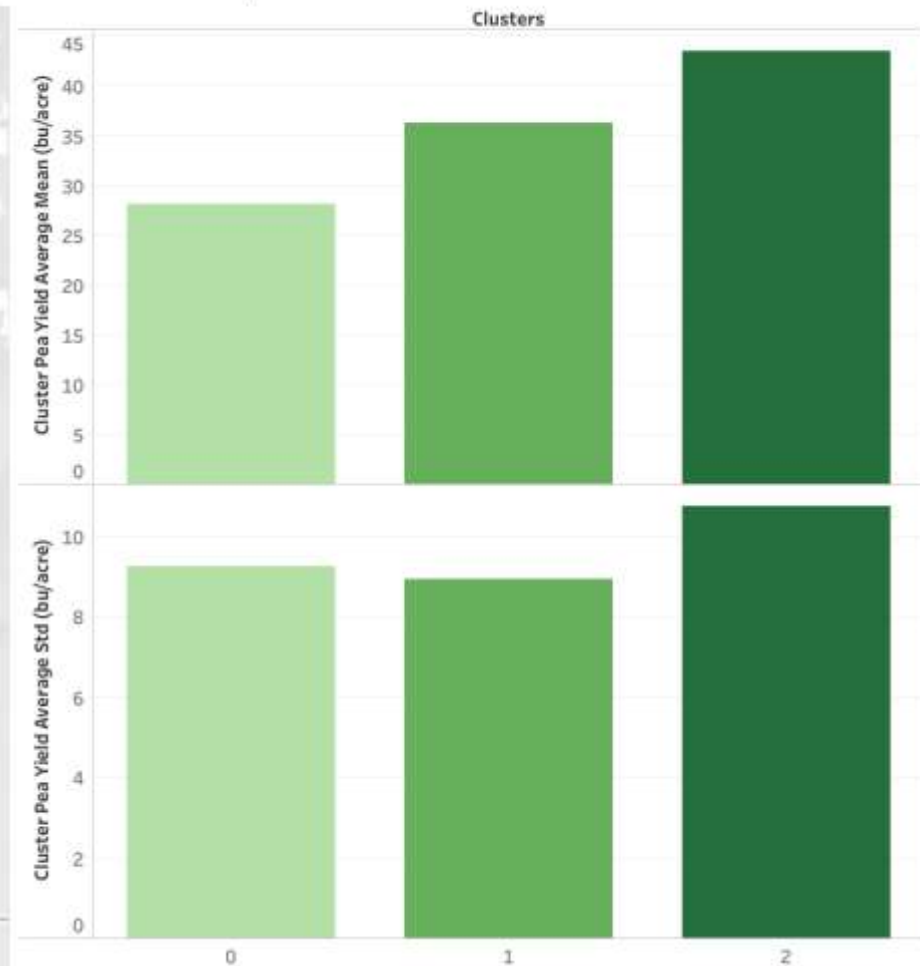
K Means Clustering

K Means Clustering

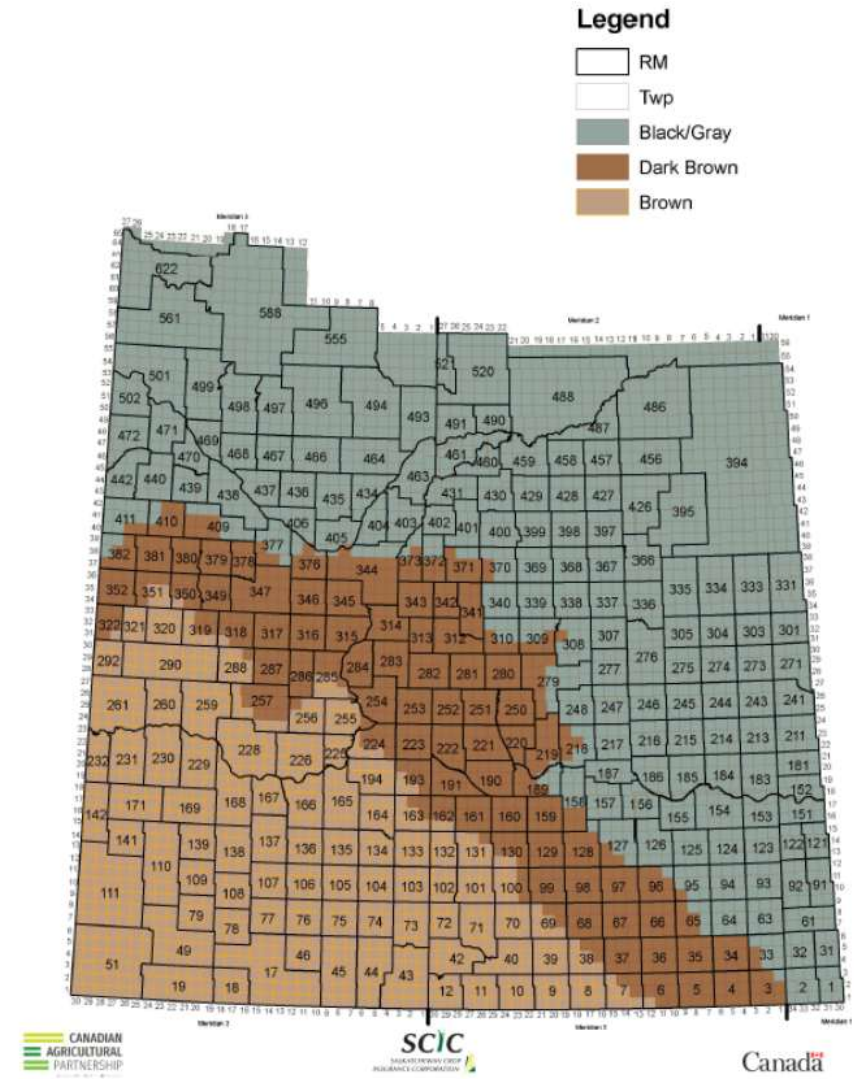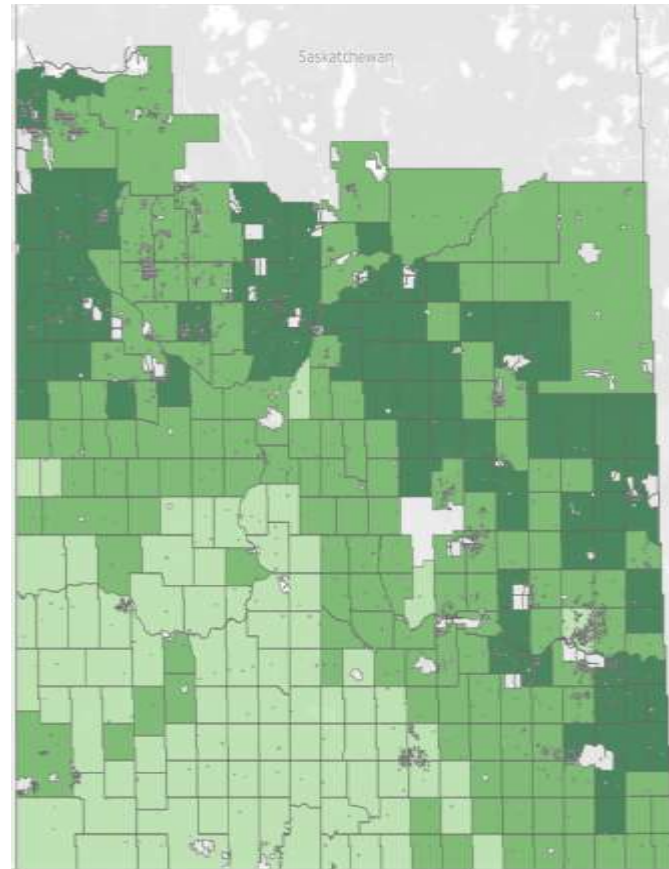Pea Yield Clustering for Saskatchewan RMs

Pea Yield Cluster Properties

Summary and Future Work

- Able to find 3 distinct clusters of pea yields in Saskatchewan

- Experiment with adding soil data, weather data, etc.

- Try K Means clustering with other pulses and compare results

Thank You