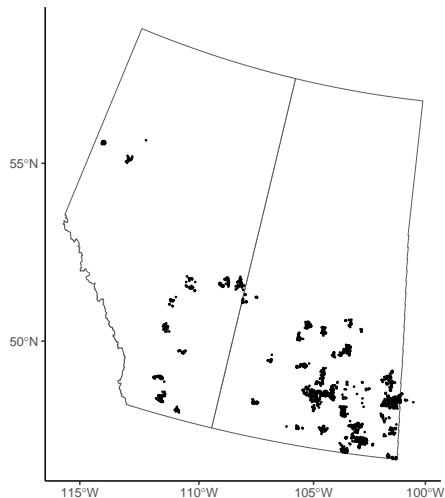# PPSN Meeting, Fall 2023
Update on yield data processing
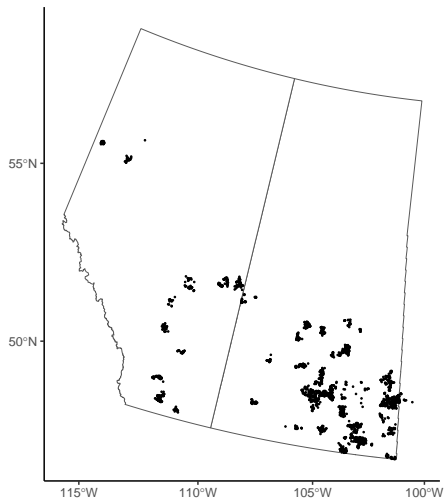
Samuel Robinson, Ph.D.

Oct 18, 2023

Update on yield data collection
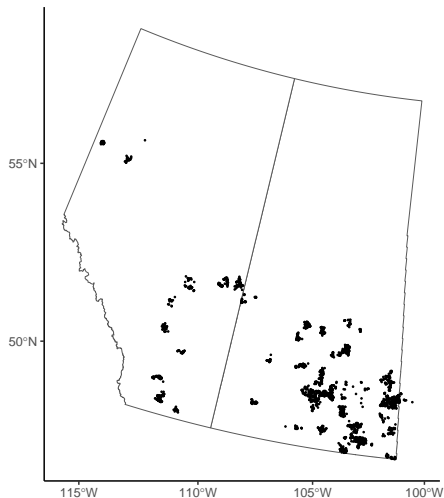
- 49 growers' datasets loaded

Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset

Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
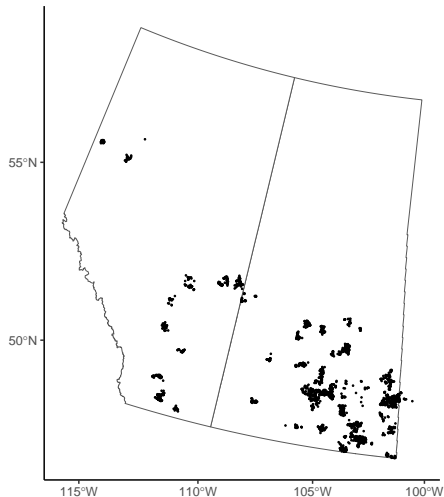
Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
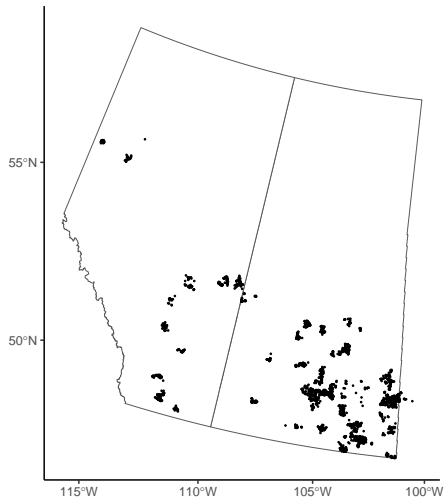- 10 years of data (2012-2022)

Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
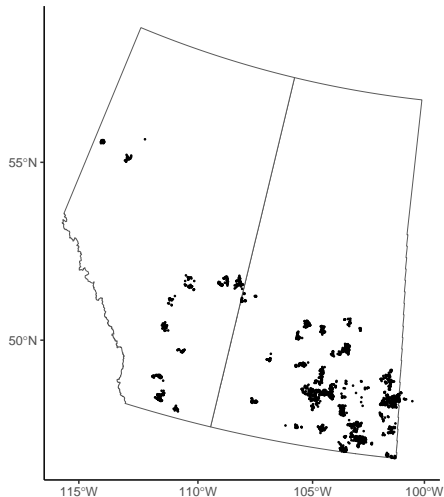- 10 years of data (2012-2022)
- 15 crop types

Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
- 10 years of data (2012-2022)
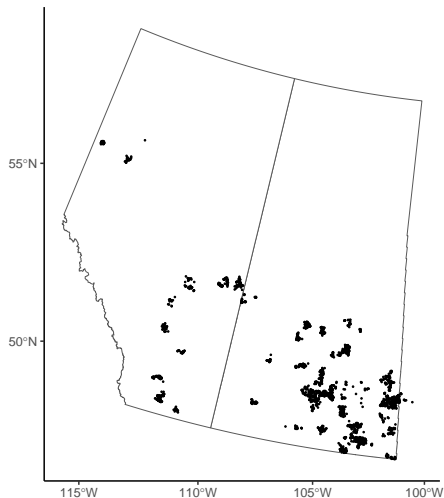- 15 crop types
- Field areas (acres):

Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
- 10 years of data (2012-2022)
- 15 crop types
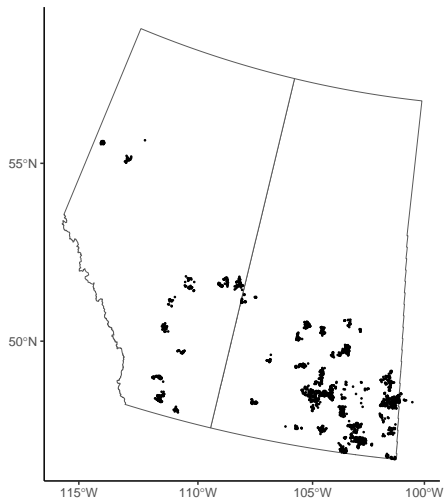- Field areas (acres):
  - Total = 795181

Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
- 10 years of data (2012-2022)
- 15 crop types
- Field areas (acres):
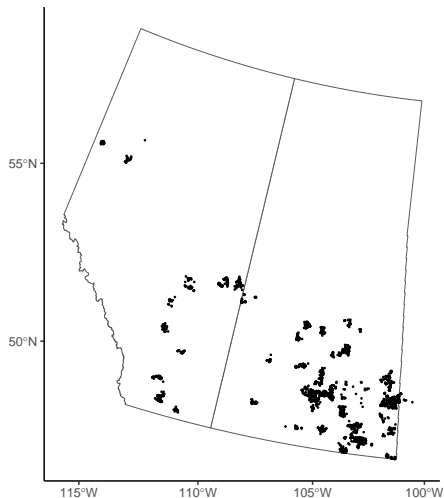 - Total = 795181
  - Mean = 330.9

Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
- 10 years of data (2012-2022)
- 15 crop types
- Field areas (acres):
  - Total = 795181
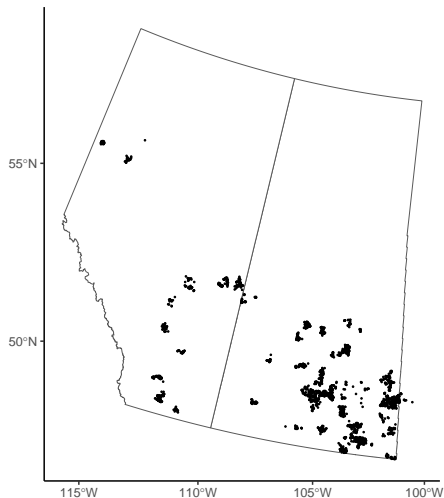 - Mean = 330.9
  - Median = 271.4

Update on yield data collection

- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
- 10 years of data (2012-2022)
- 15 crop types
- Field areas (acres):
    - Total = 795181
    - Mean = 330.9
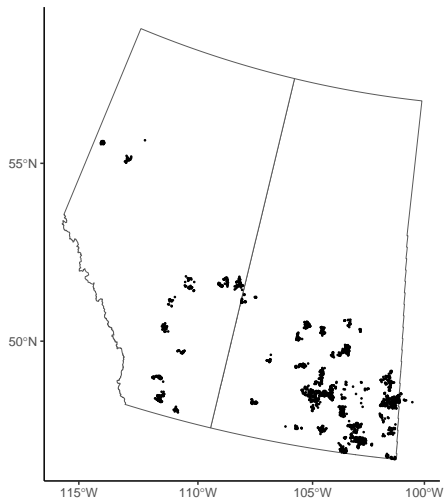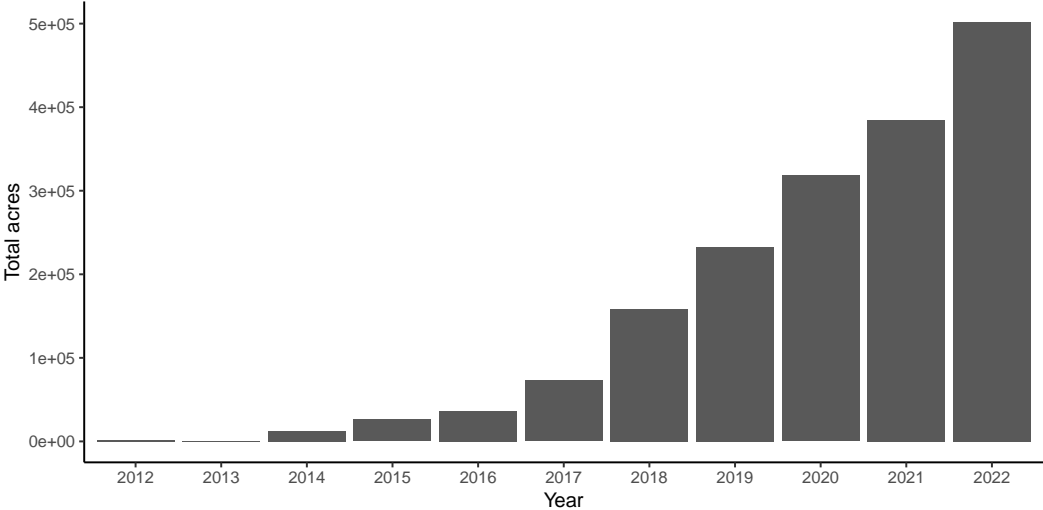   - Median = 271.4
    - Min = 2.1

Update on yield data collection
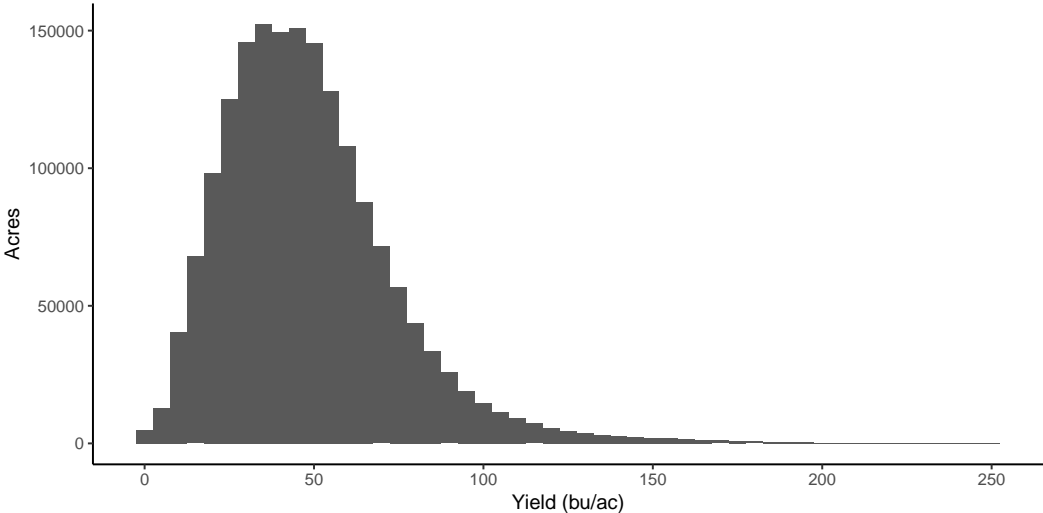
- 49 growers' datasets loaded
- 4 datasets in progress: require custom coding for each dataset
- 6523 field-years, 2016 individual fields
- 10 years of data (2012-2022)
- 15 crop types
- Field areas (acres):
  - Total = 795181
  - Mean = 330.9
  - Median = 271.4
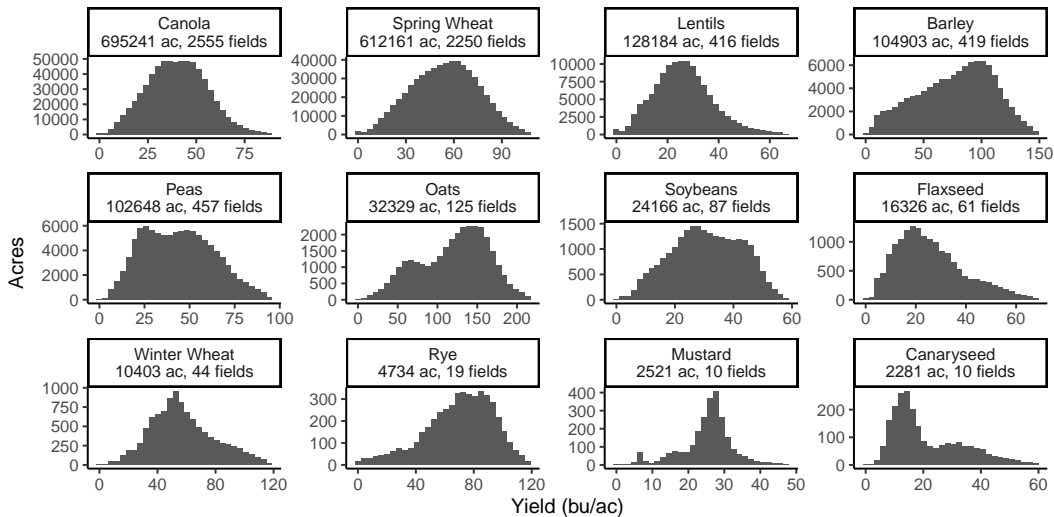 - Min = 2.1
  - Max = 2651.9

Acreage by year

Overall yield

Crop specific yield

Not shown: corn, chickpeas, soybeans

Overall profit (2022 only)
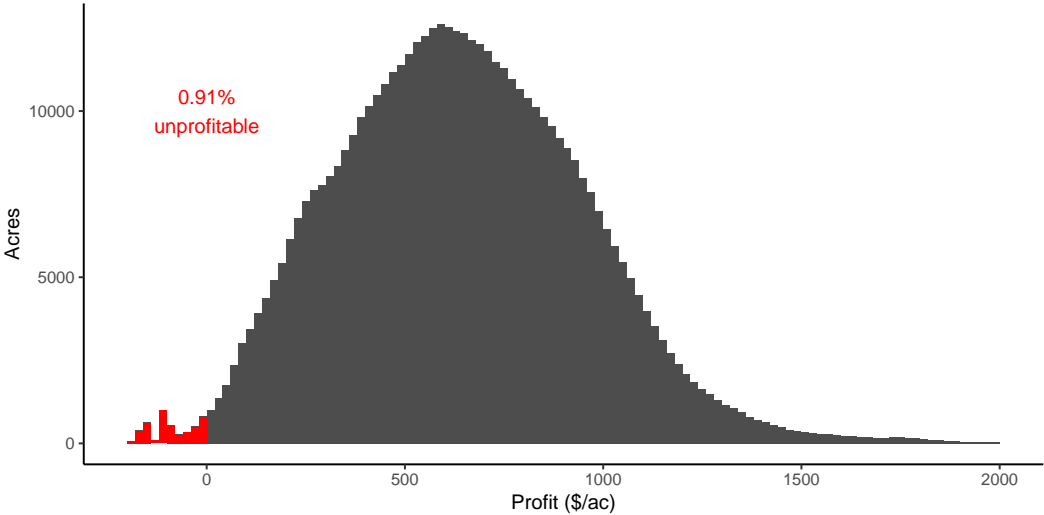
Crop-by-crop profit (2022 only)
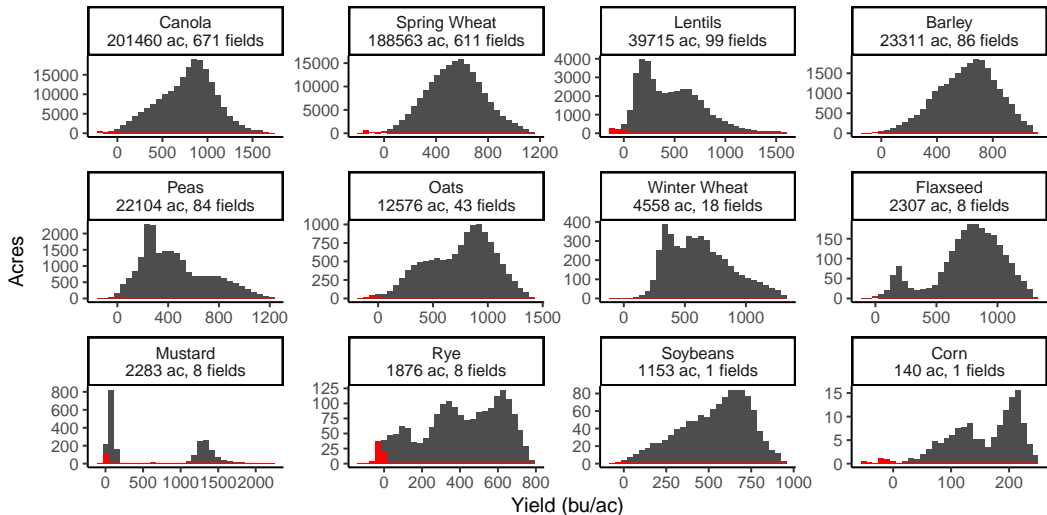
Challenge: missing data or zero yield?

Dollars
per acre

- >$400
- $300–400
- $200–300
- $100–200
- $0–100
- <$0

NA
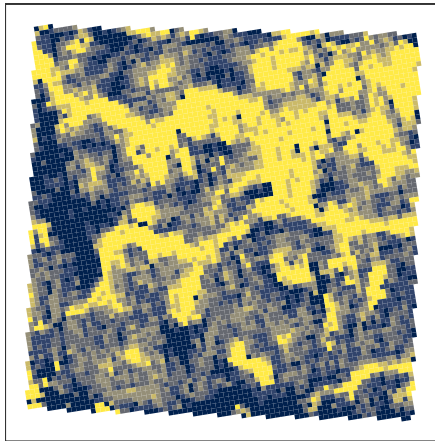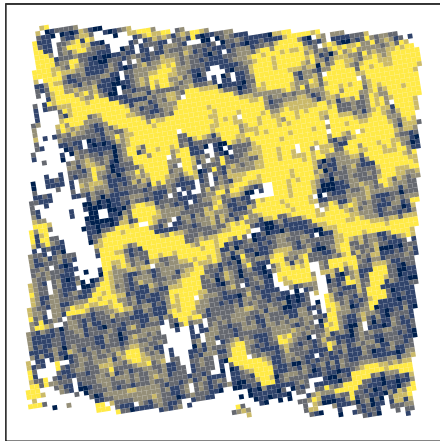
Challenge: missing data or zero yield (cont.)
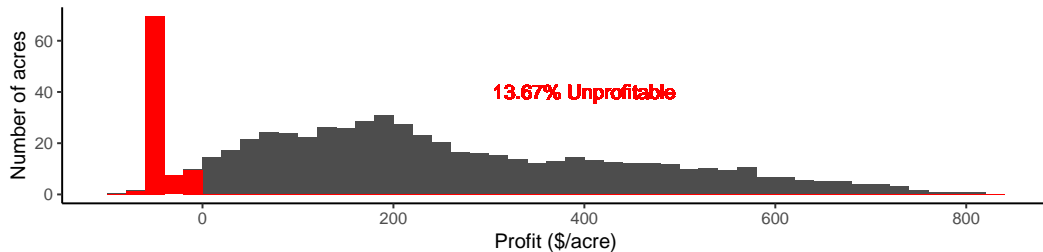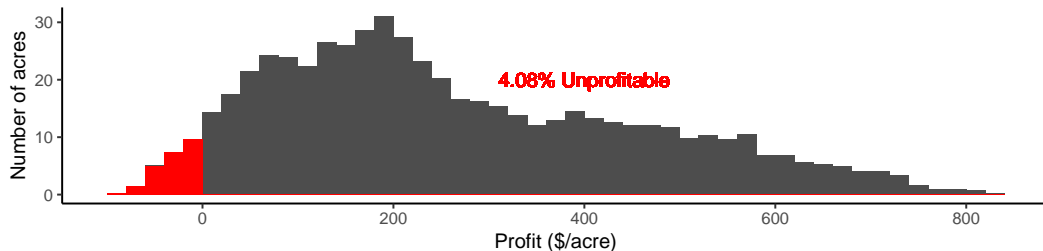
Data coverage (per field)
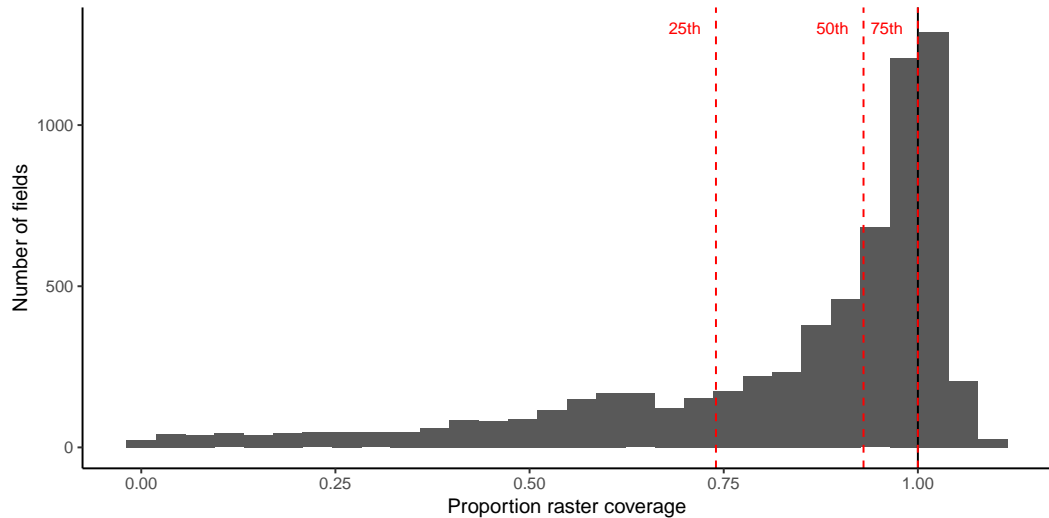
Other challenges

- Grower records

Other challenges

- Grower records
  - Need standardization/name matching

Other challenges

- Grower records
  - Need standardization/name matching
  - Mismatches between yield data, grower records, satellite data

Other challenges

- Grower records
  - Need standardization/name matching
  - Mismatches between yield data, grower records, satellite data
- Data Filtering

Other challenges

- Grower records
  - Need standardization/name matching
  - Mismatches between yield data, grower records, satellite data
- Data Filtering
  - Parameters (i.e. "throw out when $X = \dots$")

Other challenges

- Grower records
  - Need standardization/name matching
  - Mismatches between yield data, grower records, satellite data
- Data Filtering
  - Parameters (i.e. "throw out when X = . . . ")
  - Order of filtering pipeline

Other challenges

- Grower records
    - Need standardization/name matching
    - Mismatches between yield data, grower records, satellite data
- Data Filtering
    - Parameters (i.e. "throw out when $X = \ldots$")
    - Order of filtering pipeline
    - Combine back-correction smoothing

Other challenges

- Grower records
  - Need standardization/name matching
  - Mismatches between yield data, grower records, satellite data
- Data Filtering
  - Parameters (i.e. "throw out when $X = \ldots$")
  - Order of filtering pipeline
  - Combine back-correction smoothing
  - Not trivial: affects grower reports, final models, and grower trust/buy-in

Other challenges

- Grower records
  - Need standardization/name matching
  - Mismatches between yield data, grower records, satellite data
- Data Filtering
  - Parameters (i.e. "throw out when $X = \ldots$")
  - Order of filtering pipeline
  - Combine back-correction smoothing
  - Not trivial: affects grower reports, final models, and grower trust/buy-in
  - **GARBAGE IN, GARBAGE OUT**

Other challenges

- Grower records
  - Need standardization/name matching
  - Mismatches between yield data, grower records, satellite data
- Data Filtering
  - Parameters (i.e. "throw out when X = . . . ")
  - Order of filtering pipeline
  - Combine back-correction smoothing
  - Not trivial: affects grower reports, final models, and grower trust/buy-in
  - **GARBAGE IN, GARBAGE OUT**
- Further profit mapping

Other challenges

- Grower records
  - Need standardization/name matching
  - Mismatches between yield data, grower records, satellite data
- Data Filtering
  - Parameters (i.e. "throw out when $X = \ldots$")
  - Order of filtering pipeline
  - Combine back-correction smoothing
  - Not trivial: affects grower reports, final models, and grower trust/buy-in
  - **GARBAGE IN, GARBAGE OUT**
- Further profit mapping
  - Require economic data prior to 2022