Management dataset creation

orei_manure trial for Kernza Grain Machine Learning (KGML) dataset

Jesse Bealsburg

2025-01-31

What I am doing

Management data is stored in the fieldwork log. This includes planting date, fertilizer rate etc.

If this management data is only used once, like by a masters student writing up their thesis, creating a management database is unnecessary. However, if this data will need to be accessed multiple times (like by people doing meta-analyses), then a management database is useful especially if the people that need the data are unfamiliar with navigating the fieldwork timelines.

To create a management database, we need to transfer the data from the fieldwork log into a queriable format to access information like planting date. It's hard to anticipate all the management data an analyst may want. Here we are targeting just what is wanted for the KGML project

How I am doing it

creation of management data requires the addition of treatment factors

planting date may be 1 value for the variety trial

planting date may have 3 values for fert (corn, soybean, iwg), these will have to be separated by treatment

fertilizer date may have 1 value for variety trial

fertilizer date may have 5 values for N rate timing

When adding values, these are most easily added in wide form by adding columns as the treatment, then switching the treatment to long form.

start by experiment

for a given year, fill info for…

- crop
- planting date
- fertilizer date
- fertilizer type

this information is typically where there will be multiple values depending on treatment. Treatments are split out by column. For example, treatment 2 and 6 were corn and remaining treatments were iwg. Then treatment 2 and 6 have a seperate column from the other treatments. However, if they

Fieldwork timeline restructuring

Ideally, all of our fieldwork timelines have the same format and are in one document.

This is possible, all that is needed is for…

- date to be unmerged in older fieldwork logs so that every observation has a date
- experiment names and site names be consistent i.e. st paul SP Saint Paul = st paul
- all management activities are in the fieldwork log (sometimes management activities were done by stations or were done by lab members but not entered into the log)

Pivot longer

```
library(tidyverse)

read.csv("jesse_KGML_Dataset.xlsx - Sheet6 (1).csv") -> dat

dat %>%
  glimpse()
```

```
Rows: 36
Columns: 8
$ year                   <int> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020,~
$ managementInformation  <chr> "Planting date", "Crop type", "Seed harvesting ~
$ fertControl_cutNone    <chr> "2019-fall", "iwg", "2020-08-03", "2020-08-03",~
$ fertManure_cutNone     <chr> "2019-fall", "iwg", "2020-08-03", "2020-08-03",~
$ fertCommercial_cutNone <chr> "2019-fall", "iwg", "2020-08-03", "2020-08-03",~
$ fertControl_cutFall    <chr> "2019-fall", "iwg", "2020-08-03", "2020-08-03",~
$ fertManure_cutFall     <chr> "2019-fall", "iwg", "2020-08-03", "2020-08-03",~
$ fertCommercial_cutFall <chr> "2019-fall", "iwg", "2020-08-03", "2020-08-03",~
```

```
dat %>%
  pivot_longer(cols = 3:8,
               names_to = "treatment name") %>%
  mutate(experiment = "orei_manure",
         site = "rosemount",.before=year
         ) %>%
  relocate(`treatment name`, .before = managementInformation) -> dat2

dat2 %>%
  glimpse()
```

```
Rows: 216
Columns: 6
$ experiment            <chr> "orei_manure", "orei_manure", "orei_manure", "or~
$ site                  <chr> "rosemount", "rosemount", "rosemount", "rosemoun~
$ year                  <int> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, ~
$ `treatment name`      <chr> "fertControl_cutNone", "fertManure_cutNone", "fe~
$ managementInformation <chr> "Planting date", "Planting date", "Planting date~
$ value                 <chr> "2019-fall", "2019-fall", "2019-fall", "2019-fal~
```

```
dat2 %>%
  write.csv("oreiManure_managementData.csv")
```