KGML Patch Demo

adding `ORG` trial to Kernza Grain Machine Learning (KGML) dataset

Jesse Bealsburg

2025-01-29

Retrofitting datasets for new analyses using a patch

Ideally, all of our data or at least core data (yield) is all in a database in the same format and can be pulled across trials. Reality is that each experiement exists as a silo.

Master data

Master data is often structured so that each row is a unique site-year-plot with columns for every measurement from that plot and in a format that makes sense for collecting data in the field and processing across multiple steps in the lab.

Master data does not have a consistent format across experiments but is designed for the needs of the specific experiment

Management data

Data like the planting date and fertilizer rate is kept in the fieldwork timeline. This data is then pulled from the fieldwork timeline into columns in the master document as needed. Most master files contain little to no management data

Patches

Patches are files that allow for joining Master datasets with Management datasets.

A patch is like a key. It solves the problem of having to copy and paste identical data across reps/blocks which is both labor intensive and prone to error.

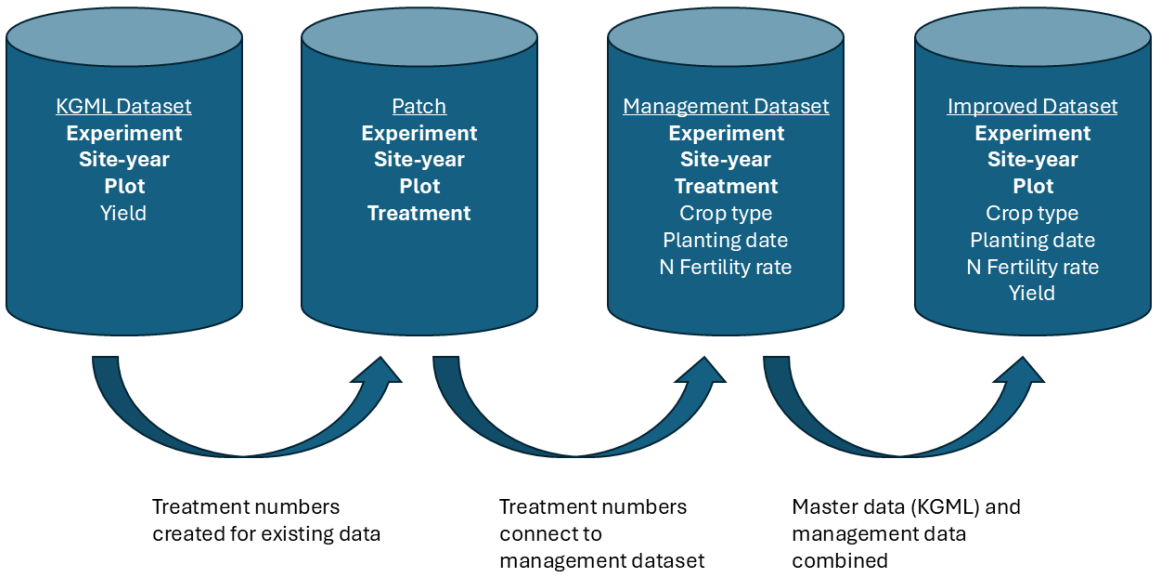

Figure 1: Using patches, we can combine management data with master datasets into a dataset better structured to for analysis across experiments.

Code

```
library(tidyverse)

read.csv("kgml_data_29Jan.csv") -> masterDat
# master data, experiment-site-year-plot-yield

# csv didn't retain filter from google sheets, so filtering just for simplicity
masterDat %>%
  filter(project == "ORG") -> masterDat

read.csv("kgml_managementData_29Jan.csv") -> managementDat
#management data, experiment-site-year-treatment-croptype

read.csv("kgml_treatmentPatch_29Jan.csv") -> trtPatch
# patch, allowing treatment numbers to be applied to plot numbers for a given experiment-site


masterDat %>%
  # distinct(year)
  mutate(site = fct_recode(location,
                           "st paul" = "St. Paul",
                           "lamberton" = "Lamberton",
                           "rosemount" = "Ros R54-55")) %>%
  rename(experiment = project) -> masterDat2
# changing column names and factor values so identical terminologies used

trtPatch %>%
  mutate(site = fct_recode(site,
                           "st paul" = "St. Paul",
                           "lamberton" = "Lamberton",
                           "rosemount" = "Ros R54-55")) -> trtPatch2
# changing column names and factor values so identical terminologies used
```

```
managementDat %>%
  filter(information == "Crop type") %>%
  rename(crop = value) %>%
  select(experiment, site, year, crop, treatment)  -> managementDat2
# changing column names for simplicity and clarity.

masterDat2 %>%
```

```
  select(experiment,site,year,plot) -> masterDat3

masterDat3 %>%
  glimpse()
```

```
Rows: 71
Columns: 4
$ experiment <chr> "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "OR~
$ site       <fct> st paul, st paul, st paul, st paul, st paul, st paul, st pa~
$ year       <int> 2019, 2019, 2019, 2019, 2018, 2018, 2018, 2018, 2019, 2019,~
$ plot       <int> 102, 205, 306, 406, 102, 205, 306, 406, 102, 204, 303, 402,~
```

```
managementDat2 %>%
  glimpse()
```

```
Rows: 18
Columns: 5
$ experiment <chr> "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "OR~
$ site       <chr> "st paul", "rosemount", "lamberton", "st paul", "rosemount"~
$ year       <int> 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2018, 2019,~
$ crop       <chr> "soybean", "soybean", "soybean", "soybean", "soybean", "soy~
$ treatment  <int> 1, 1, 1, 6, 6, 6, 2, 2, 2, 1, 1, 1, 6, 6, 6, 2, 2, 2
```

```
trtPatch2 %>%
  glimpse()
```

```
Rows: 108
Columns: 5
$ experiment <chr> "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "ORG", "OR~
$ site       <fct> lamberton, lamberton, lamberton, lamberton, rosemount, lamb~
$ year       <int> 2019, 2019, 2019, 2019, 2019, 2018, 2018, 2019, 2019, 2020,~
$ treatment  <int> 1, 1, 1, 1, 2, 2, 6, 2, 2, 1, 1, 1, 1, 1, 1, 2, 6, 1, 2, 2,~
$ plot       <int> 106, 203, 301, 405, 102, 102, 105, 102, 105, 106, 106, 203,~
```

```
masterDat3 %>%
  left_join(trtPatch2) %>%
  left_join(managementDat2) %>%
  arrange(site,year)
```

	experiment	site	year	plot	treatment	crop
1	ORG	lamberton	2018	102	2	iwg
2	ORG	lamberton	2018	204	2	iwg
3	ORG	lamberton	2018	303	2	iwg
4	ORG	lamberton	2018	402	2	iwg
5	ORG	lamberton	2018	106	1	soybean
6	ORG	lamberton	2018	203	1	soybean
7	ORG	lamberton	2018	301	1	soybean
8	ORG	lamberton	2018	405	1	soybean
9	ORG	lamberton	2018	105	6	soybean
10	ORG	lamberton	2018	205	6	soybean
11	ORG	lamberton	2018	304	6	soybean
12	ORG	lamberton	2018	404	6	soybean
13	ORG	lamberton	2019	102	2	iwg
14	ORG	lamberton	2019	204	2	iwg
15	ORG	lamberton	2019	303	2	iwg
16	ORG	lamberton	2019	402	2	iwg
17	ORG	lamberton	2019	106	1	corn
18	ORG	lamberton	2019	203	1	corn
19	ORG	lamberton	2019	301	1	corn
20	ORG	lamberton	2019	405	1	corn
21	ORG	lamberton	2019	105	6	corn
22	ORG	lamberton	2019	205	6	corn
23	ORG	lamberton	2019	304	6	corn
24	ORG	lamberton	2019	404	6	corn
25	ORG	rosemount	2018	102	2	iwg
26	ORG	rosemount	2018	204	2	iwg
27	ORG	rosemount	2018	303	2	iwg
28	ORG	rosemount	2018	402	2	iwg
29	ORG	rosemount	2018	106	1	soybean
30	ORG	rosemount	2018	203	1	soybean
31	ORG	rosemount	2018	301	1	soybean
32	ORG	rosemount	2018	405	1	soybean
33	ORG	rosemount	2018	105	6	soybean
34	ORG	rosemount	2018	205	6	soybean
35	ORG	rosemount	2018	304	6	soybean
36	ORG	rosemount	2018	404	6	soybean
37	ORG	rosemount	2019	102	2	iwg
38	ORG	rosemount	2019	204	2	iwg
39	ORG	rosemount	2019	303	2	iwg
40	ORG	rosemount	2019	402	2	iwg
41	ORG	rosemount	2019	106	1	corn
42	ORG	rosemount	2019	203	1	corn

43	ORG	rosemount	2019	301	1	corn
44	ORG	rosemount	2019	405	1	corn
45	ORG	rosemount	2019	105	6	corn
46	ORG	rosemount	2019	205	6	corn
47	ORG	rosemount	2019	304	6	corn
48	ORG	rosemount	2019	404	6	corn
49	ORG	st paul	2018	102	1	soybean
50	ORG	st paul	2018	205	1	soybean
51	ORG	st paul	2018	306	1	soybean
52	ORG	st paul	2018	406	1	soybean
53	ORG	st paul	2018	105	2	iwg
54	ORG	st paul	2018	203	2	iwg
55	ORG	st paul	2018	305	2	iwg
56	ORG	st paul	2018	404	2	iwg
57	ORG	st paul	2018	106	6	soybean
58	ORG	st paul	2018	202	6	soybean
59	ORG	st paul	2018	304	6	soybean
60	ORG	st paul	2018	403	6	soybean
61	ORG	st paul	2019	102	1	corn
62	ORG	st paul	2019	205	1	corn
63	ORG	st paul	2019	306	1	corn
64	ORG	st paul	2019	406	1	corn
65	ORG	st paul	2019	105	2	iwg
66	ORG	st paul	2019	305	2	iwg
67	ORG	st paul	2019	404	2	iwg
68	ORG	st paul	2019	106	6	corn
69	ORG	st paul	2019	202	6	corn
70	ORG	st paul	2019	304	6	corn
71	ORG	st paul	2019	403	6	corn