KGML Patch Demo

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

Jesse Bealsburg

2025-01-31

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 experiment 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.
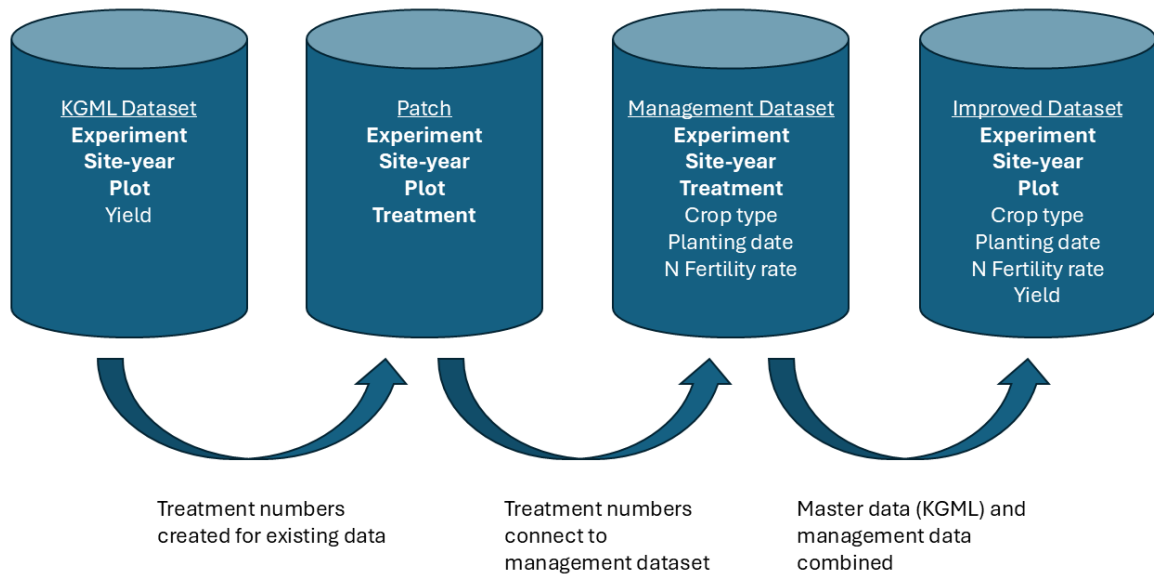
Figure 1: Using patches, we can combine management data with master datasets into a dataset better structured to for analysis across experiments. KGML dataset could be swapped with orei_manure dataset or any other master dataset.

Code

ORG

```r
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
```

```r
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,seedyld) -> masterDat3

masterDat3 %>%
  glimpse()
```

```
Rows: 71
Columns: 5
$ 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,~
$ seedyld    <dbl> 15253.800, 19731.100, 15172.500, 15393.800, 634.053, 959.46~
```

```
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(plot,site,year) %>%
relocate(seedyld,.after=crop) %>%
slice(1:37)
```

```
   experiment        site year plot treatment      crop     seedyld
1          ORG lamberton 2018  102         2       iwg     813.800
2          ORG lamberton 2019  102         2       iwg     680.800
3          ORG rosemount 2018  102         2       iwg     473.600
4          ORG rosemount 2019  102         2       iwg     183.000
5          ORG    st paul 2018  102         1   soybean     634.053
6          ORG    st paul 2019  102         1      corn   15253.800
7          ORG lamberton 2018  105         6   soybean    3013.500
8          ORG lamberton 2019  105         6      corn   13337.500
9          ORG rosemount 2018  105         6   soybean    1891.800
10         ORG rosemount 2019  105         6      corn   14084.100
11         ORG    st paul 2018  105         2       iwg     297.100
12         ORG    st paul 2019  105         2       iwg      61.900
13         ORG lamberton 2018  106         1   soybean    2264.000
14         ORG lamberton 2019  106         1      corn   12251.100
15         ORG rosemount 2018  106         1   soybean    1793.900
16         ORG rosemount 2019  106         1      corn   10783.100
17         ORG    st paul 2018  106         6   soybean     798.900
18         ORG    st paul 2019  106         6      corn   13481.800
19         ORG    st paul 2018  202         6   soybean    1395.100
20         ORG    st paul 2019  202         6      corn   16336.600
21         ORG lamberton 2018  203         1   soybean    2474.700
22         ORG lamberton 2019  203         1      corn   13475.600
23         ORG rosemount 2018  203         1   soybean    1308.800
24         ORG rosemount 2019  203         1      corn   14167.200
25         ORG    st paul 2018  203         2       iwg      58.100
26         ORG lamberton 2018  204         2       iwg     396.100
27         ORG lamberton 2019  204         2       iwg     688.900
28         ORG rosemount 2018  204         2       iwg     376.700
29         ORG rosemount 2019  204         2       iwg     111.700
30         ORG lamberton 2018  205         6   soybean    2466.400
31         ORG lamberton 2019  205         6      corn   12615.400
32         ORG rosemount 2018  205         6   soybean    1673.500
33         ORG rosemount 2019  205         6      corn   15099.300
34         ORG    st paul 2018  205         1   soybean     959.460
35         ORG    st paul 2019  205         1      corn   19731.100
36         ORG lamberton 2018  301         1   soybean    2715.200
37         ORG lamberton 2019  301         1      corn   12535.200
```

Orei manure

Here I created management data for orei and a treatment patch, I am now connecting just the seed yield data from the master into an improved dataset which I am printing below.

```r
read.csv("thesis_prepared_data - data_main.csv") -> masterDat_orei


read.csv("jesse_kgml_dataset.xlsx - treatment patch (3).csv") -> trtPatch_orei


read.csv("jesse_KGML_dataset.xlsx - Management data.csv") -> managementDat_orei

masterDat_orei %>%
  filter(location=="MN") %>%
  mutate(site = fct_recode(location,
                           rosemount = "MN")) %>%
  mutate(
    experiment = "orei_manure",
    plot = reassigned_plot,
    seedyld = grain_yield_kg_ha_avg) %>%
  relocate(c(experiment, site, year, plot, seedyld)) %>%
  select(1:5)-> masterDat_orei2


managementDat_orei %>%
  filter(information == "Crop type") %>%
  rename(crop = value) %>%
  select(experiment, site, year, crop, treatment)  -> managementDat_orei2


masterDat_orei2 %>%
  left_join(trtPatch_orei) %>%
  left_join(managementDat_orei2) %>%
  arrange(plot,site,year) %>%
  relocate(seedyld,.after=crop) %>%
  select(-treatment) %>%
  # filter(treatment.name != "NA") %>%
  slice(1:25)
```

```
    experiment       site year plot          treatment.name crop     seedyld
1  orei_manure rosemount 2020  101 fertCommercial_cutNone  iwg   695.68966
```

```
2  orei_manure rosemount 2021  101 fertCommercial_cutNone  iwg   458.77082
3  orei_manure rosemount 2022  101 fertCommercial_cutNone  iwg   178.17178
4  orei_manure rosemount 2020  102     fertControl_cutNone  iwg   801.72414
5  orei_manure rosemount 2021  102     fertControl_cutNone  iwg   464.79595
6  orei_manure rosemount 2022  102     fertControl_cutNone  iwg    75.74453
7  orei_manure rosemount 2020  104      fertManure_cutNone  iwg   805.17241
8  orei_manure rosemount 2021  104      fertManure_cutNone  iwg   584.43788
9  orei_manure rosemount 2022  104      fertManure_cutNone  iwg    80.04819
10 orei_manure rosemount 2020  111 fertCommercial_cutFall  iwg          NA
11 orei_manure rosemount 2021  111 fertCommercial_cutFall  iwg 1740.40263
12 orei_manure rosemount 2022  111 fertCommercial_cutFall  iwg   219.48698
13 orei_manure rosemount 2020  112     fertControl_cutFall  iwg          NA
14 orei_manure rosemount 2021  112     fertControl_cutFall  iwg   592.18448
15 orei_manure rosemount 2022  112     fertControl_cutFall  iwg    30.98640
16 orei_manure rosemount 2020  114      fertManure_cutFall  iwg          NA
17 orei_manure rosemount 2021  114      fertManure_cutFall  iwg   493.20015
18 orei_manure rosemount 2022  114      fertManure_cutFall  iwg   160.09638
19 orei_manure rosemount 2020  201      fertManure_cutNone  iwg 1012.06897
20 orei_manure rosemount 2021  201      fertManure_cutNone  iwg   739.36986
21 orei_manure rosemount 2022  201      fertManure_cutNone  iwg    86.93406
22 orei_manure rosemount 2020  202 fertCommercial_cutNone  iwg   868.96552
23 orei_manure rosemount 2021  202 fertCommercial_cutNone  iwg   897.74478
24 orei_manure rosemount 2022  202 fertCommercial_cutNone  iwg    98.98432
25 orei_manure rosemount 2020  204     fertControl_cutNone  iwg   678.44828
```

Now with bringing in more than just crop information

```
managementDat_orei %>%
  filter(information == "Crop type" |
           information == "N fertilizer rate") %>%
  pivot_wider(names_from = information,
              values_from = value) %>%
  select(experiment, site, year, `Crop type`,
         `N fertilizer rate`, treatment)  -> managementDat_orei3

masterDat_orei2 %>%
  left_join(trtPatch_orei) %>%
  left_join(managementDat_orei3) %>%
  arrange(plot,site,year) %>%
  relocate(seedyld,.after=`N fertilizer rate`) %>%
  select(-treatment) %>%
  slice(1:25)
```

```
    experiment          site year plot          treatment.name Crop type
1  orei_manure rosemount 2020  101 fertCommercial_cutNone       iwg
2  orei_manure rosemount 2021  101 fertCommercial_cutNone       iwg
3  orei_manure rosemount 2022  101 fertCommercial_cutNone       iwg
4  orei_manure rosemount 2020  102    fertControl_cutNone       iwg
5  orei_manure rosemount 2021  102    fertControl_cutNone       iwg
6  orei_manure rosemount 2022  102    fertControl_cutNone       iwg
7  orei_manure rosemount 2020  104     fertManure_cutNone       iwg
8  orei_manure rosemount 2021  104     fertManure_cutNone       iwg
9  orei_manure rosemount 2022  104     fertManure_cutNone       iwg
10 orei_manure rosemount 2020  111 fertCommercial_cutFall       iwg
11 orei_manure rosemount 2021  111 fertCommercial_cutFall       iwg
12 orei_manure rosemount 2022  111 fertCommercial_cutFall       iwg
13 orei_manure rosemount 2020  112    fertControl_cutFall       iwg
14 orei_manure rosemount 2021  112    fertControl_cutFall       iwg
15 orei_manure rosemount 2022  112    fertControl_cutFall       iwg
16 orei_manure rosemount 2020  114     fertManure_cutFall       iwg
17 orei_manure rosemount 2021  114     fertManure_cutFall       iwg
18 orei_manure rosemount 2022  114     fertManure_cutFall       iwg
19 orei_manure rosemount 2020  201     fertManure_cutNone       iwg
20 orei_manure rosemount 2021  201     fertManure_cutNone       iwg
21 orei_manure rosemount 2022  201     fertManure_cutNone       iwg
22 orei_manure rosemount 2020  202 fertCommercial_cutNone       iwg
23 orei_manure rosemount 2021  202 fertCommercial_cutNone       iwg
24 orei_manure rosemount 2022  202 fertCommercial_cutNone       iwg
25 orei_manure rosemount 2020  204    fertControl_cutNone       iwg
   N fertilizer rate     seedyld
1              70   695.68966
2              70   458.77082
3              70   178.17178
4               0   801.72414
5               0   464.79595
6               0    75.74453
7              88   805.17241
8              88   584.43788
9              88    80.04819
10             70          NA
11             70  1740.40263
12             70   219.48698
13              0          NA
14              0   592.18448
15              0    30.98640
16             88          NA
```

```
17              88   493.20015
18              88   160.09638
19              88 1012.06897
20              88   739.36986
21              88    86.93406
22              70   868.96552
23              70   897.74478
24              70    98.98432
25               0   678.44828
```